



**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO**

**ESCUELA DE POSGRADO**

**MAESTRÍA EN ESTADÍSTICA**

**TESIS**

**EVALUACION Y ANÁLISIS DEL RENDIMIENTO DE 36  
GENOTIPOS DE TRIGO HARINERO UTILIZANDO EL DISEÑO  
DE BLOQUES COMPLETOS AL AZAR, DISTRITO DE ZURITE –  
2020**

**PARA OPTAR EL GRADO ACADÉMICO DE MAESTRO EN  
ESTADÍSTICA**

**AUTOR**

**Br. RIGOBERTO ESTRADA ZUNIGA**

**ASESOR:**

**Dra. NATALIE VERÓNKA RONDINEL  
MENDOZA**

**CODIGO ORCID: 0000-0001-6867-5782**

**CUSCO – PERÚ**

**2023**

## INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro.CU-303-2020-UNSAAC)

El que suscribe, asesor del trabajo de investigación/tesis titulado: **EVALUACION Y ANÁLISIS DEL RENDIMIENTO DE 36 GENOTIPOS DE TRIGO HARINERO UTILIZANDO EL DISEÑO DE BLOQUES COMPLETOS AL AZAR, DISTRITO DE ZURITE – 2020**, presentado por: **RIGOBERTO ESTRADA ZUNIGA** con Nro. de DNI: 23864292, para optar el título profesional/grado académico de... **MAESTRO EN ESTADÍSTICA**. Informo que el trabajo de investigación ha sido sometido a revisión por UNA vez, mediante el Software Antiplagio, conforme al Art. 6° del *Reglamento para Uso de Sistema Antiplagio de la UNSAAC* y de la evaluación de originalidad se tiene un porcentaje de 3%

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No se considera plagio.	X
Del 11 al 30 %	Devolver al usuario para las correcciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y **adjunto** la primera hoja del reporte del Sistema Antiplagio.

Cusco, 12 de abril de 2023



Firma

Post firma **Dra. Natalie Verónica Rondinel Mendoza.**

Nro. de DNI. 23989672

ORCID del Asesor: 0000-0001-6867-5782

**Se adjunta:**

1. Reporte generado por el Sistema Antiplagio.
2. Enlace del Reporte Generado por el Sistema Antiplagio: **Identificación de reporte de similitud: oid:27259:188345700**

NOMBRE DEL TRABAJO

**EVALUACION Y ANÁLISIS DEL RENDIMIENTO DE 36 GENOTIPOS DE TRIGO HARINERO UTILIZANDO EL DISEÑO DE BLOQ**

AUTOR

**RIGOBERTO ESTRADA ZUNIGA**

RECUENTO DE PALABRAS

**38429 Words**

RECUENTO DE CARACTERES

**183486 Characters**

RECUENTO DE PÁGINAS

**177 Pages**

TAMAÑO DEL ARCHIVO

**4.6MB**

FECHA DE ENTREGA

**Dec 12, 2022 12:33 PM GMT-5**

FECHA DEL INFORME

**Dec 12, 2022 12:36 PM GMT-5****● 3% de similitud general**

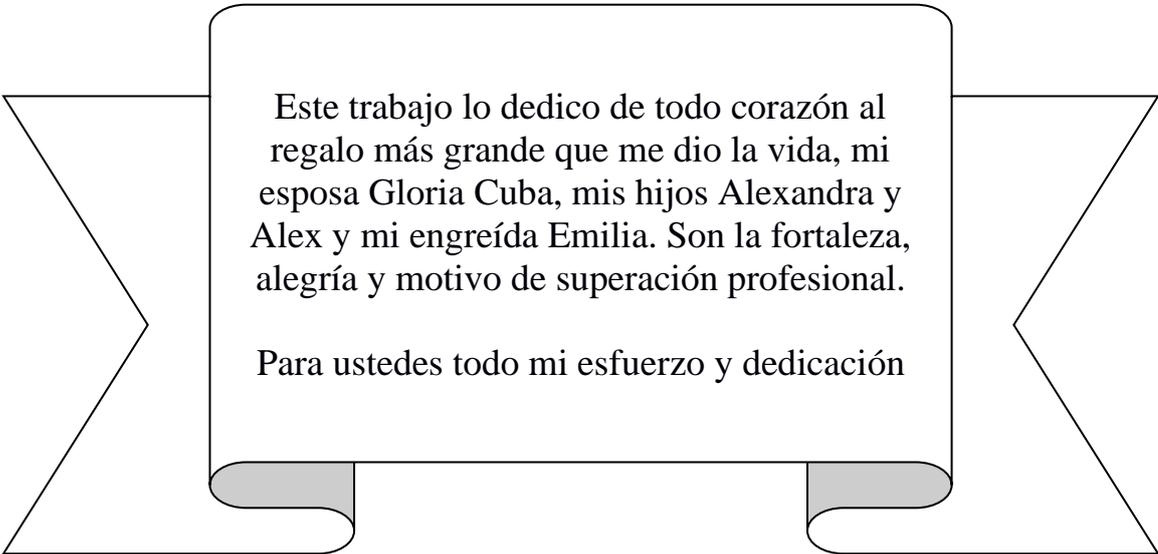
El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos

- 3% Base de datos de Internet
- Base de datos de Crossref
- 2% Base de datos de trabajos entregados
- 0% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

**● Excluir del Reporte de Similitud**

- Material bibliográfico
- Material citado
- Coincidencia baja (menos de 25 palabras)

## DEDICATORIA



Este trabajo lo dedico de todo corazón al regalo más grande que me dio la vida, mi esposa Gloria Cuba, mis hijos Alexandra y Alex y mi engreída Emilia. Son la fortaleza, alegría y motivo de superación profesional.

Para ustedes todo mi esfuerzo y dedicación

## AGRADECMIENTOS

A mi esposa Gloria por su paciencia, confianza y motivación para lograr mi superación profesional.

A la Dra. Natalie Verónica Rondinel Mendoza, tutora de la investigación mis agradecimientos por darme la oportunidad de recurrir a su capacidad, experiencia y conocimiento científico que con mucha paciencia y tolerancia me acompañó y brindó sus orientaciones hasta culminar el presente trabajo.

Mi reconocimiento a los Dr. Cleto De La torre Dueñas y al Dr. Arturo Zuñiga Blanco por sus aportes y sugerencias durante el dictamen de la tesis cuya experiencia contribuyó para mejorar el trabajo de investigación.

Mis agradecimientos a los docentes de la Maestría en Estadística de la Universidad Nacional de San Antonio Abad del Cusco que día a día nos brindaron nuevos conocimientos y motivaron a culminar con el propósito que nos trazamos al iniciar los estudios de una de las ciencias más importantes para todo científico que es la Estadística

Mis agradecimientos al Programa Nacional de Innovación Agraria del INIA, por su patrocinio del trabajo de investigación que permitirá mejorar nuestra contribución en el aporte científico y el desarrollo del sector agrario en el país.

A mis compañeras y compañeros de la maestría en estadística que gracias a su amistad y apoyo logaron aportar para alcanzar este objetivo en mi formación profesional.

A Víctor, Hernán y Julián, compañeros de trabajo por su apoyo constante en el desarrollo del trabajo de investigación.

## RESUMEN

Utilizar diseños experimentales en el mejoramiento genético de trigo harinero son fundamentales para generar la información exacta e interpretable con alto rigor científico que permita contribuir con variedades potenciales en rendimiento, calidad comercial y como consecuencia disponibilidad de alimentos para una población cada vez más creciente; con este motivo en la investigación se planteó utilizar el diseño de bloques completos al azar con cuatro repeticiones para evaluar 36 genotipos de trigo harinero con los objetivos de determinar el nivel de influencia de la altura de planta, el número de granos por espiga, el tamaño de espiga, el peso de mil granos y la madurez fisiológica en el rendimiento. Entre los principales resultados se obtuvieron en los análisis de clúster tres grupos de los tratamientos en estudio con características similares para las variables evaluadas. Para determinar el nivel de influencia de cada una de las variables en estudio en el rendimiento, se registró correlaciones estadísticamente significativas del rendimiento con altura de planta ( $r=0.589$ ), con el número de granos por espiga ( $r=0.932$ ), con el tamaño de espiga ( $r=0.860$ ), con el peso de mil granos ( $r=0.969$ ) y madurez fisiológica ( $r=0.932$ ) y el modelo explica que la proporción de la variabilidad total del rendimiento es mayor al 73% a excepción de la altura de planta que fue del 34%. Los hallazgos muestran como conclusión que existe influencia de las variables antes descritas en el rendimiento de trigo harinero, el diseño utilizado fue eficiente para evaluar el comportamiento de los 36 genotipos de trigo, las pruebas estadísticas desarrolladas permiten seleccionar a los genotipos TH-014, TH-034 y TH-001 superiores en rendimiento a los testigos de comparación y buenas características agronómicas para el mejoramiento genético.

## PALABRAS CLAVES

Rendimiento, trigo harinero, diseño de bloques completos al azar.

## ABSTRACT

Using experimental designs in the genetic improvement of bread wheat are essential to generate exact and interpretable information with high scientific rigor that allows contributing with potential varieties in yield, commercial quality and consequently food availability for an increasingly growing population; For this reason, in the investigation it was proposed to use the randomized complete block design with four repetitions to evaluate 36 genotypes of bread wheat with the objectives of determining the level of influence of the plant height, the number of grains per spike, the size of ear, the weight of thousand grains and the physiological maturity in the yield. Among the main results, three groups of the treatments under study with similar characteristics for the variables evaluated were obtained in the cluster analysis. To determine the level of influence of each of the variables under study on yield, statistically significant correlations of yield with plant height ( $r=0.589$ ), with the number of grains per spike ( $r=0.932$ ), with the spike size ( $r=0.860$ ), with the thousand grain weight ( $r=0.969$ ) and physiological maturity ( $r=0.932$ ) and the model explains that the proportion of total yield variability is greater than 73% except for the plant height that was 34%. The findings show in conclusion that there is an influence of the variables described above on the yield of bread wheat, the design used was efficient to evaluate the behavior of the 36 wheat genotypes, the statistical tests developed allow the selection of the TH-014, TH -034 and TH-001 superior in yield to the comparison controls and good agronomic characteristics for genetic improvement.

## KEYWORDS

Yield, bread wheat, randomized complete block design.

## CONTENIDO

RESUMEN .....	iv
ABSTRACT.....	v
<b>Contenido</b> .....	vi
PRESENTACION.....	xiv
INTRODUCCIÓN .....	xv
CAPÍTULO I.....	1
PLANTEAMIENTO DEL PROBLEMA .....	1
1.1. Situación problemática.....	1
1.2. Formulación del problema .....	4
a. Problema general.....	4
b. Problemas específicos.....	4
1.3. Justificación de la investigación.....	5
1.4. Objetivos de la investigación .....	6
a. Objetivo general .....	6
b. Objetivos específicos .....	7
CAPÍTULO II .....	8
MARCO TEÓRICO CONCEPTUAL .....	8
2.1. Bases teóricas .....	8
2.1.1. Análisis de Clúster .....	8
2.1.2. Diseños experimentales .....	11
2.1.3. Unidad Experimental .....	12
2.1.4. Tratamiento .....	12
2.1.5. Variables .....	12
2.1.6. Factores controlables .....	13
2.1.7. Factores no controlables .....	13
2.1.8. Principios del diseño experimental. ....	13
2.1.9. Estadísticos .....	14
2.1.10. Análisis de Regresión lineal simple .....	17
2.1.11. Medidas de Bondad de ajuste de la regresión .....	19
2.1.12. Análisis de varianza .....	24
2.1.13. Análisis de varianza de modelos estadísticos lineales .....	25
2.1.14. Verificación del modelo adecuado.....	28
2.1.15. Diseños de bloques completos al azar.....	29
2.1.16. Supuestos del modelo.....	32
2.1.17. Pruebas para la normalidad y homogeneidad de varianzas.....	34

2.1.18.	Estimación por máxima verosimilitud de los parámetros del modelo lineal de un diseño de bloques completos al azar .....	36
2.1.19.	Propiedades de los estimadores de máxima verosimilitud.....	39
2.1.20.	Comparaciones de pruebas o rangos múltiples .....	42
2.1.21.	Trigo .....	43
2.1.22.	El cultivo de trigo y su origen .....	44
2.1.23.	Clasificación del trigo .....	44
2.1.24.	El rendimiento del trigo y sus componentes .....	45
2.1.25.	El rendimiento de trigo y el número de granos .....	46
2.1.26.	El mejoramiento por rendimiento y otros caracteres de interés .....	47
2.1.27.	Variabilidad y diversidad genética .....	47
2.2.	Marco conceptual .....	48
2.2.1.	Experimento .....	48
2.2.2.	Componentes de rendimiento .....	48
2.2.3.	Peso de mil granos .....	48
2.2.4.	Madurez Fisiológica.....	49
2.2.5.	Peso unitario de grano.....	49
2.2.6.	Rendimiento .....	49
2.2.7.	Rendimiento potencial .....	50
2.3.	Antecedentes empíricos de la investigación.....	50
CAPÍTULO III .....		56
HIPÓTESIS Y VARIABLES .....		56
3.1.	Hipótesis .....	56
a.	Hipótesis general.....	56
b.	Hipótesis específicas .....	56
3.2.	Identificación de variables e indicadores .....	56
3.3.	Operacionalización de variables .....	58
CAPITULO IV.....		58
METODOLOGÍA.....		58
4.1.	Ámbito de estudio: localización política y geográfica.....	58
4.2.	Tipo y nivel de investigación .....	59
4.3.	Diseño experimental y croquis de distribución.....	59
4.4.	Tratamientos en estudio .....	61
4.5.	Unidad de análisis .....	64
4.6.	Población de estudio .....	64
4.7.	Tamaño de muestra .....	64
4.8.	Técnicas de selección de muestra .....	65

4.9. Técnicas de recolección de información.....	65
4.10. Técnicas y herramientas estadísticas de análisis de los datos.....	67
CAPÍTULO V.....	70
RESULTADOS Y DISCUSIÓN.....	70
RESULTADOS.....	70
5.1. Análisis exploratorio de los datos.....	70
5.1.1. Análisis de Conglomerados o clúster.....	72
5.2. Determinación de la influencia de la altura de la planta en el rendimiento de trigo harinero.....	79
5.2.1. Análisis de la regresión lineal para la altura de planta y el rendimiento.....	79
5.2.2. Análisis de varianza para la altura de planta.....	82
5.2.3. Comparación de medias de altura de la planta de los tratamientos.....	86
5.3. Determinación de la influencia del número de granos por espiga en el rendimiento de trigo harinero.....	89
5.3.1. Análisis de regresión lineal para el número de granos por espiga y el rendimiento de trigo harinero.....	89
5.3.2. Análisis de varianza para el número de granos por espiga.....	93
5.3.3. Comparación de medias del número de granos por espiga de los tratamientos en estudio.....	97
5.4. Determinación de la influencia del tamaño de espiga en el rendimiento de trigo harinero.....	99
5.4.1. Análisis de regresión lineal para el tamaño de espiga y el rendimiento de trigo harinero.....	99
5.4.2. Análisis de varianza para el tamaño de espiga.....	102
5.4.3. Comparación de medias para el tamaño de espiga.....	106
5.5. Determinación de la influencia del peso de mil granos en el rendimiento de trigo harinero.....	109
5.5.1. Análisis de regresión lineal para el peso de mil granos y el rendimiento de trigo harinero.....	109
5.5.2. Análisis de varianza para el peso de mil granos.....	113
5.5.3. Comparación de medias de peso de mil granos de los tratamientos.....	117
5.6. Determinación de la influencia de la madurez fisiológica en el rendimiento de trigo harinero.....	119
5.6.1. Análisis de regresión lineal para la madurez fisiológica y el rendimiento de trigo harinero.....	119
5.6.2. Análisis de varianza para la madurez fisiológica.....	123
5.6.3. Comparación de medias de los tratamientos para la variable madurez fisiológica.....	127
5.7. Análisis de varianza comparación de medias de los tratamientos para el rendimiento de los genotipos de trigo.....	129

DISCUSIÓN DE LOS RESULTADOS .....	136
CONCLUSIONES .....	143
RECOMENDACIONES.....	146
BIBLIOGRAFÍA.....	148
ANEXOS .....	154

## INDICE DE TABLAS

<b>Tabla 1.</b> Estructura del análisis de varianza para el modelo con un solo factor de efectos fijos .....	27
<b>Tabla 2.</b> Esquema típico para un experimento en bloques completos al azar.....	30
<b>Tabla 3.</b> Esquema típico para un ANOVA de bloques completos al azar .....	31
<b>Tabla 4.</b> Variables e indicadores .....	57
<b>Tabla 5.</b> Operacionalización de variables .....	58
<b>Tabla 6.</b> Características del campo experimental.....	61
<b>Tabla 7.</b> Relación del material genético utilizado en el trabajo de investigación.....	61
<b>Tabla 8.</b> Estadísticos descriptivos de las variables en estudio .....	70
<b>Tabla 9.</b> Matriz de correlaciones de Pearson para las variables en estudio .....	71
<b>Tabla 10.</b> Matriz de proximidades de los tratamientos en estudio a través de la distancia de Mahalanobis.....	72
<b>Tabla 11.</b> Características de cada conglomerado .....	78
<b>Tabla 12.</b> Análisis de la regresión lineal aplicada a la altura de la planta y el rendimiento ...	80
<b>Tabla 13.</b> Criterios de bondad de ajuste para la altura de la planta.....	86
<b>Tabla 14.</b> Comparación de tratamientos frente a la altura de la planta (cm) .....	87
<b>Tabla 15.</b> Aplicación del análisis de regresión lineal para el número de granos por espiga y el rendimiento .....	90
<b>Tabla 16.</b> Criterios de bondad de ajuste para el número de granos por espiga .....	96
<b>Tabla 17.</b> Valor medio y error estándar por tratamiento para el número de granos por espiga. ....	98
<b>Tabla 18.</b> Análisis de regresión lineal aplicado al tamaño de espiga y el rendimiento .....	100
<b>Tabla 19.</b> Análisis de varianza para el tamaño de espiga (cm).....	105
<b>Tabla 20.</b> Comparación de medias para el tamaño de espigas (cm) a través de la prueba de Tukey .....	108
<b>Tabla 21.</b> Análisis de regresión lineal aplicado al peso de mil granos y el rendimiento .....	110
<b>Tabla 22.</b> Criterios de bondad de ajuste para el peso de mil granos (gramos).....	116
<b>Tabla 23.</b> Comparación del valor promedio de los tratamientos y error estándar para el peso de mil granos (g). ....	118
<b>Tabla 24.</b> Análisis de regresión lineal aplicado a la madurez fisiológica y el rendimiento..	121
<b>Tabla 25.</b> Análisis de varianza para la madurez fisiológica.....	126
<b>Tabla 26.</b> Comparación de medias para la madurez fisiológica a través de la prueba Tukey .....	128
<b>Tabla 27.</b> Análisis de varianza para el rendimiento .....	132
<b>Tabla 28.</b> Comparación de medias para el rendimiento (t/ha) a través de la prueba de Tukey .....	134
<b>Tabla 29.</b> Matriz de consistencia para el desarrollo del trabajo de investigación.....	154

## INDICE DE FIGURAS

<b>Figura 1.</b> Determinación de la región de no rechazo (región no sombreada) con valores calculados para la prueba de independencia (X,Y) utilizando el estadístico T de Student. ....	24
<b>Figura 2.</b> Mapa de ubicación del distrito de Zurite donde se desarrolló la investigación. ....	58
<b>Figura 3.</b> Croquis de ubicación y distribución del diseño experimental.....	60
<b>Figura 4.</b> Diagrama de cajas para las variables evaluadas .....	71
<b>Figura 5.</b> Dendograma de los tratamientos en estudio que muestra la formación de tres grupos.....	75
<b>Figura 6.</b> Diagrama de los tratamientos que integran el grupo 1 del dendograma .....	76
<b>Figura 7.</b> Diagrama de los tratamientos que integran el grupo 2 del dendograma .....	76
<b>Figura 8.</b> Diagrama de los tratamientos que integran el grupo 3 del dendograma .....	77
<b>Figura 9.</b> Representación gráfica de la dispersión de los datos y la recta ajustada para el rendimiento y la altura de la planta.....	80
<b>Figura 10.</b> Representación gráfica de los residuos de la variable rendimiento frente a la altura de la planta. ....	81
<b>Figura 11.</b> Representación gráfica de la dispersión de datos de altura de planta para los tratamientos por bloques. ....	83
<b>Figura 12.</b> Gráfico de la distribución normal de los residuos de la variable altura de la planta. ....	85
<b>Figura 13.</b> Gráfico de los residuos y valores ajustados para la altura de la planta. ....	85
<b>Figura 14.</b> Comparación de altura de la planta (cm) de los tratamientos en estudio.....	88
<b>Figura 15.</b> Representación gráfica de la dispersión de los datos y la recta ajustada para el rendimiento y el número de granos por espiga. ....	90
<b>Figura 16.</b> Representación gráfica de los residuos de la variable rendimiento frente al número de granos por espiga.....	92
<b>Figura 17.</b> Gráfico de la dispersión del número de granos por espiga de los tratamientos por bloques. ....	94
<b>Figura 18.</b> Gráfica de la distribución normal de los residuos de la variable granos por espiga. ....	95
<b>Figura 19.</b> Gráfica de residuos y valores ajustados para el número de granos por espiga. ....	96
<b>Figura 20.</b> Comparación de medias para el número de granos por espiga de los tratamientos en estudio. ....	97
<b>Figura 21.</b> Gráfico de la dispersión de los datos y la recta ajustada para el rendimiento y el tamaño de espiga.....	100
<b>Figura 22.</b> Gráfica de los residuos de la variable rendimiento frente al tamaño de espiga. .	101
<b>Figura 23.</b> Análisis gráfico de la dispersión de los tratamientos por bloque para el tamaño de espiga. ....	103
<b>Figura 24.</b> Gráfico de la probabilidad normal de los residuos del tamaño de espiga. ....	104
<b>Figura 25.</b> Gráfica de residuos y valores ajustados para el tamaño de espiga.....	105
<b>Figura 26.</b> Comparación de tratamientos para el tamaño de espiga. ....	107

<b>Figura 27.</b> Representación gráfica de la dispersión de datos y la recta ajustada para el rendimiento y el peso de mil granos. ....	110
<b>Figura 28.</b> Representación gráfica de los residuos de la variable rendimiento frente al peso de mil granos.....	112
<b>Figura 29.</b> Análisis gráfico de la dispersión de los tratamientos por bloque para el peso de mil granos.....	114
<b>Figura 30.</b> Gráfico de la probabilidad normal para los residuos del peso de mil granos.....	115
<b>Figura 31.</b> Análisis de residuos y valores ajustados para el peso de mil granos. ....	116
<b>Figura 32.</b> Comparación de medias para el peso de mil granos de los tratamientos en estudio. ....	117
<b>Figura 33.</b> Representación gráfica de la dispersión de los datos y la recta ajustada para el rendimiento y la madurez fisiológica.....	120
<b>Figura 34.</b> Representación gráfica de los residuos de la variable rendimiento frente a días a madurez fisiológica.....	122
<b>Figura 35.</b> Comportamiento de la variable días a madurez fisiológica según tratamientos por bloque.....	124
<b>Figura 36.</b> Gráfica de la probabilidad normal de los residuos de días a madurez fisiológica. ....	125
<b>Figura 37.</b> Gráfica de residuos y valores ajustados para los días a madurez fisiológica. ....	125
<b>Figura 38.</b> Comparación de medias para la madurez fisiológica de los tratamientos en estudio. ....	129
<b>Figura 39.</b> Comportamiento de la variable rendimiento según tratamientos por bloques. ....	130
<b>Figura 40.</b> Gráfica de la probabilidad normal de los residuos para el rendimiento.....	131
<b>Figura 41.</b> Gráfica de los residuos y valores ajustados para el rendimiento.....	132
<b>Figura 42.</b> Comparación de medias para el rendimiento de los tratamientos en estudio.....	135
<b>Figura 43.</b> Área destinada al trabajo de investigación e instalación del ensayo.....	159
<b>Figura 44</b> Desarrollo vegetativo de la parcela experimental y evaluación para el control de malezas.....	159
<b>Figura 45.</b> Vista panorámica de la parcela experimental en fase de madurez fisiológica. ....	160
<b>Figura 46.</b> Muestreo de espigas para la evaluación del número de granos por espiga. ....	160

**SIGLAS**

CIMMYT	Centro Internacional de Mejoramiento de Maíz y Trigo
DBCA	Diseño de Bloques Completos al Azar
EEAAC	Estación Experimental Agraria Andenes Cusco
INIAF	Instituto Nacional de Investigación Agropecuaria y Forestal
FAO	Organización de las Naciones Unidas para la Alimentación y la Agricultura
INEI	Instituto Nacional de Estadística e Informática
INIA	Instituto Nacional de Innovación Agraria
INTA	Instituto Nacional de Tecnología Agropecuaria
MINAGRI	Ministerio de Agricultura y Riego
U.A.	Unidades Agropecuarias
UNALM	Universidad Nacional Agraria La Molina
UPOV	Unión Internacional para la Protección de las Obtenciones Vegetales
ANVA	Análisis de Varianza
ANOVA	Análisis of variance

## PRESENTACION

Ante nuestras autoridades de la Escuela de Posgrado de la Universidad Nacional de San Antonio Abad del Cusco, la Maestría en Estadística y miembros de jurado examinador.

En cumplimiento a lo establecido en el reglamento de grados y títulos se ha ejecutado el presente trabajo de investigación titulado “EVALUACIÓN Y ANÁLISIS DEL RENDIMIENTO DE 36 GENOTIPOS DE TRIGO HARINERO UTILIZANDO EL DISEÑO DE BLOQUES COMPLETOS AL AZAR, DISTRITO DE ZURITE – 2020” para optar el grado académico de Maestro en Estadística, con la finalidad de demostrar la eficiencia del diseño experimental aplicado en la evaluación de genotipos de trigo que sea de utilidad a las instituciones de investigación que ejecutan programas de mejoramiento genético con pruebas en condiciones de campo.

La presentación del trabajo de investigación ha sido estructurada en los siguientes capítulos: I) descripción del planteamiento del problema, objetivos, justificación y delimitación del espacio temporal del problema; II) desarrollo del marco teórico, conceptual de los diseños experimentales, mejoramiento del cultivo de trigo para el rendimiento y antecedentes de la investigación; III) hipótesis y variables con la identificación y operacionalización respectiva; IV) detalla la metodología de ejecución del trabajo desde su ubicación, el tipo de investigación, tratamientos en estudio, y las técnicas y herramientas estadísticas utilizadas; V) detalla los resultados por objetivos con análisis de regresión lineal, correlación, análisis de varianza y comparación de medias de los tratamientos y discusiones de la investigación; para culminar con las conclusiones, recomendaciones y la bibliografía utilizada que pongo en consideración.

## INTRODUCCIÓN

Los diseños de investigación son fundamentales e importantes en los procesos científicos porque contribuyen a desagregar las estrategias para generar información exacta e interpretable, entre estos se tiene a los diseños experimentales. Existen definiciones de numerosos autores entre ellos Montgomery (2004), quién señala que el experimento es una o varias pruebas donde se realizan cambios en las variables independientes para observar e identificar los cambios en las variables dependientes de un proceso.

En la actividad agrícola la investigación es una necesidad en el país y es de gran relevancia para la generación de tecnologías orientadas a la seguridad alimentaria, llevando en consideración el crecimiento de la población mundial y la necesidad de alimentos en cantidad y calidad son los desafíos actuales que serán posibles lograr con investigación y desarrollo de nuevas variedades que exige el mercado. Por esta razón, la experimentación aplicada y el uso de los diseños experimentales basados en la estadística inferencial es cada vez más relevante porque logra describir y proponer cambios con principios científicos en el mejoramiento genético vegetal.

Por otro lado, el cultivo de trigo es una de las especies de importancia económica en el Perú y especialmente en el Cusco, principalmente entre las poblaciones que integran la agricultura familiar y pequeños productores que orientan la producción a la seguridad alimentaria y el mercado nacional que amerita su atención por las entidades de innovación en la generación de variedades comerciales competitivas en el mercado que lleguen a adoptar y utilizar los productores en su cédula de cultivo, considerando que es una especie que mitigará los efectos del cambio climático por su respuesta a condiciones de baja precipitación y en cultivo de secano ya que las economías del poblador rural se verán afectados severamente por efectos de la

sequías según afirma el Ministerio del Ambiente (2015), sobre todo cuando el trigo es uno de los alimentos importantes en el mundo con consumo per cápita de 67.8 kg/persona/año, ubicado entre los cuatro principales alimentos Juárez (2019).

Bajo estas consideraciones, se planteó el presente trabajo de investigación donde se propone aplicar el diseño experimental de bloques completos al azar en la evaluación de la influencia de las variables agronómicas en el rendimiento del cultivo de trigo y seleccionar genotipos de alta productividad para los procesos de mejoramiento genético y desarrollo de nuevas variedades.

## **CAPITULO I**

### **PLANTEAMIENTO DEL PROBLEMA**

#### **1.1. Situación problemática**

Según información de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) la producción de trigo en el 2019 fue de 763.1 millones de toneladas siendo 4.2% más con referencia al 2018 principalmente en la Unión Europea, China, India, Estados Unidos, Canadá y Argentina que son los países que incrementaron la producción de trigo en el mundo. En el año 2019 en América Latina y el Caribe la producción de trigo fue de 31.8 millones de toneladas y en América del Sur para el mismo año fue de 28.5 millones de toneladas que en comparación al 2018 se observó una disminución de 3.3 millones de toneladas siendo Argentina y Brasil los principales países productores de trigo en esta región del mundo FAO (2020).

En el Perú, según el registro del compendio estadístico del Instituto Nacional de Estadística e Informática (INEI, 2018), se ha desarrollado el análisis de la producción de trigo y se destaca que en los últimos 17 años el incremento del rendimiento por unidad de área en el país para el 2017 fue de 1.572 t/ha, y en el mismo periodo, la Región Cusco registró también una notable variación del rendimiento que oscila entre los 1.680 a 1.790 t/ha. El rendimiento promedio del cultivo de trigo registrado en la Región Cusco, es superior al promedio nacional ratificándose que las condiciones agro climáticas son favorables para su producción que se puede incrementar con la disponibilidad de nuevas tecnologías (p.267).

El trigo forma parte del consumo básico de la población peruana, pero la producción es deficitaria para el consumo nacional registrándose en el año 2016 un

consumo per-cápita de 43 kg/persona/año, concentrándose el 97% de la superficie cultivada en la sierra y el 3% en la costa, dedicándose a este cultivo 87,218 Unidades agropecuarias (U.A) en el Perú y en Cusco 8,721 U.A. Específicamente en el ámbito del proyecto se registra 2,788 U.A, que representa el 32% del total de los productores de la Región Cusco que se dedican al cultivo de trigo.

Con respecto al mejoramiento genético del trigo, el Centro Internacional de Mejoramiento en Maíz y Trigo (CIMMYT) reporta que la evaluación de germoplasma experimental a nivel internacional en diversos ambientes es muy importante para agilizar el proceso de mejoramiento y desarrollo de variedades mejoradas de trigo en ambientes de diferente altitud, clima, presencia de plagas y enfermedades, dada la diversidad de zonas de producción en el mundo CIMMYT (1986).

En la actualidad, los planes de mejoramiento que se vienen implementando en el cultivo de trigo son desarrollados a nivel de investigación básica y aplicada por los centros de investigación, entre ellos en el Perú por el Instituto Nacional de Innovación Agraria (INIA) que tiene dependencia del Ministerio de Agricultura y Riego según el Decreto Ley 1060\_MINAGRI (DL-01060, 2008).

En este proceso de observación de los planes de mejoramiento existen evidencias de haber desarrollado variedades mejoradas en base al material genético procedente del Centro de Investigación en Mejoramiento de Maíz y Trigo (Kohli et al., 2003) y que actualmente son cultivados en el Perú, entre ellas según el reporte del INIA encontramos en la Región Cusco las variedades de trigo Andino INIA, Andenes INIA, INIA 403- Moray, INIA 405-San Isidro e INIA 418-San Francisco generados por diversos procesos de investigación desde poblaciones segregantes hasta lograr una variedad comercial. Sin embargo, el Ministerio de Agricultura señala que es escasa la disponibilidad de variedades, semillas certificadas y asistencia técnica que garanticen

la calidad y productividad del trigo considerando que esta especie es importante para la seguridad alimentaria y la generación de ingresos de la familia rural.

Estrada et al. (2020), señalan que en los procesos de mejoramiento genético para el desarrollo de nuevas variedades de trigo según los protocolos de desarrollo de tecnologías que implementa el Instituto Nacional de Innovación Agraria en el Cusco se encuentran diferentes fases de investigación con el fin de evaluar el material genético procedente de cruzamientos que pasan por ensayos de rendimiento, sanidad, comparación de las líneas seleccionadas en campo de agricultores para su validación técnica y económica; sin embargo, se ha venido observando que la precisión al momento de levantar la información de campo debe ser mejorada por lo que consideramos que es necesario plantear propuestas de uso de diseños experimentales que se adapten a las características del material genético en evaluación con el fin de disminuir los procesos de mejoramiento y lograr la mayor confiabilidad de los resultados a través de la eliminación de los errores durante las mediciones de los parámetros en evaluación.

En el presente estudio se propuso el uso de los diseños de bloques completos al azar (DBCA), por ser un diseño que permite llevar en consideración factores experimentales que llevan en cuenta las características genéticas y fenotípicas que se adapten al ambiente.

El presente estudio se desarrolló en la Estación Experimental Agraria Andenes Cusco (EEAAC) ubicado en el distrito de Zurite de la Provincia de Anta, Región Cusco a 3405 msnm, 13°45'47'' latitud sur y 72°25'56'' longitud oeste.

Analizando la situación actual descrita líneas arriba, con el presente trabajo de investigación se pretende aplicar los métodos estadísticos apropiados para el análisis de los datos que en el proceso de investigación y selección de material genético de

trigo harinero se logre el mayor rigor científico y técnico en la planificación y análisis de resultados obtenidos de ensayos de campo para disponer de variedades con mejor respuesta de adaptación a las condiciones ambientales donde se genere la tecnología optimizando principalmente el rendimiento.

Las acciones que se implementaron para resolver el problema fue el uso del diseño de bloques completos al azar en la planificación e implementación en campo para ser evaluados y analizados adecuadamente.

## **1.2. Formulación del problema**

### **a. Problema general**

¿Es posible determinar la variación del rendimiento del comparativo de 36 genotipos de trigo harinero a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020?

### **b. Problemas específicos**

b.1. ¿Cómo influye la altura de planta en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020?

b.2. ¿Cómo influye el número de granos por espiga en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020?

b.3. ¿Cómo influye el tamaño de espiga en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis

utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020?

b.4. ¿Cómo influye el peso de mil granos en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020?

b.5. ¿Cómo influye la madurez fisiológica en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020?

### **1.3. Justificación de la investigación**

Si bien es cierto que en la investigación agrícola, particularmente en cultivos anuales como el trigo diferentes instituciones dedicadas al desarrollo de nuevas variedades vienen utilizando diseños experimentales y otras técnicas estadísticas que permitan dar el rigor científico en la presentación de nuevas variedades comerciales, la pertinencia de la presente investigación se fundamenta en adaptar, observar y analizar el uso del diseño experimental por bloques completos al azar en la evaluación de poblaciones diversas de genotipos de trigo procedentes de diversos cruzamientos y la importancia de su aplicación bajo condiciones de campo de la Región Cusco.

En la medida en que se documente y se logre compartir el conocimiento del uso apropiado del diseño experimental antes mencionado en la evaluación de líneas avanzadas o genotipos de trigo harinero en diferentes condiciones de campo será de mayor relevancia social para los investigadores del sector agrario aplicar en zonas geográficas como el Cusco donde encontramos una gran diversidad de factores de la

producción entre ellos: las características geográficas, los tipos de suelo, la disponibilidad de riego o producción en secano, la variación climática por altitud entre otros factores que influyen en el comportamiento de las variables que se evalúan en este caso las líneas promisorias de trigo para su desarrollo como nuevas variedades y por consiguiente la respuesta de la variable dependiente como el rendimiento de trigo.

Ello implica ser consistentes con el uso del diseño propuesto para la investigación y que el modelo del experimento incorpore los parámetros necesarios para dar cuenta de la estructura de parcelas esperando que metodológicamente la utilidad de los resultados de la investigación no solo sean tener un modelo con muchas variables de evaluación sino un experimento más preciso.

El diseño experimental en la investigación agropecuaria, es considerado parte del proceso científico cuyo valor teórico permite demostrar una de las formas en que aprendemos acerca del funcionamiento de los sistemas o procesos para generar datos y utilizar esta información del experimento para establecer comparaciones de material genético para el desarrollo de nuevas variedades de trigo, por consiguiente estamos convencidos que los resultados del presente estudio contribuirán en los procesos de generación de variedades de trigo del Instituto Nacional de Innovación Agraria en la Región Cusco y del país a partir de la evaluación y selección de nuevos genotipos.

#### **1.4. Objetivos de la investigación**

##### **a. Objetivo general**

Determinar la variación del rendimiento del comparativo de 36 genotipos de trigo harinero a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020.

**b. Objetivos específicos**

- b.1. Determinar la influencia de la altura de planta en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020.
- b.2. Determinar la influencia del número de granos por espiga en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020.
- b.3. Determinar la influencia del tamaño de espiga en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020.
- b.4. Determinar la influencia del peso de mil granos en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020.
- b.5. Determinar la influencia de la madurez fisiológica en el rendimiento de trigo harinero del comparativo de 36 genotipos a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020.

## CAPITULO II

### MARCO TEÓRICO CONCEPTUAL

#### 2.1. Bases teóricas

##### 2.1.1. Análisis de Clúster

Sobre el análisis de clúster, Ramirez et al. (2017) señalan que es una técnica multivariante, exploratoria y descriptiva conocido también como análisis de conglomerados que permite agrupar variables en grupos que sean similares entre sí y distintos del resto. Por consiguiente, el análisis de clúster, es una técnica de amplio uso de análisis de datos en diversas ciencias del conocimiento entre ellas la biología, la botánica para identificar genotipos similares a partir de la caracterización de sus variables.

Una definición similar es mencionada por Montanero (2008) señalando que el análisis de clúster es una técnica multivariada de agrupación de datos tomado en consideración su afinidad sobre un vector  $Y$  de  $p$  variables observadas y para ello es necesario conocer las variables, medir la afinidad de los datos, utilizar el método apropiado de aglomeración y realizar la valoración de los conglomerados. Las variables para el análisis de conglomerados o clúster pueden ser cualitativas (ordinales o nominales) o cuantitativas (discretas o continuas).

Montanero (2008), También indica que para el análisis de conglomerados se debe disponer información que describan las relaciones entre individuos, a través de una medida apropiada de asociación que puede ser de distancia o de similaridad. Entre las más utilizadas para formar conglomerados son:

- Distancia euclídea: denotada por  $d_e(\mathbf{x}_1, \mathbf{x}_2)$  se define mediante la siguiente ecuación para determinar la distancia entre dos vectores de datos  $\mathbf{x}_1 = (x_{11}, \dots, x_{1p})'$  y  $\mathbf{x}_2 = (x_{21}, \dots, x_{2p})'$ , donde:

$$d_e(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|^2 = \sqrt{\sum_{j=1}^p (x_{1j} - x_{2j})^2}$$

- Correlación de Pearson: Denotada por  $r_{x_1, x_2}$  que trata de permutar el papel que juegan los datos y las variables, dados dos vectores de datos  $\mathbf{x}_1$  y  $\mathbf{x}_2$ , se calculan la media ( $\bar{x}_i$ ) y la desviación estándar ( $S_{ik}$ )

$$\bar{x}_i = \frac{1}{p} \sum_{j=1}^p x_{ij}, \quad S_{ik} = \frac{1}{p} \sum_{j=1}^p (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k), \quad i, k = 1, 2.$$

Entonces se considera en el coeficiente

$$r_{x_1, x_2} = \frac{S_{12}}{\sqrt{S_{11}S_{22}}} \quad (1)$$

- Distancia absoluta: Denotada por  $d_a(\mathbf{x}_1, \mathbf{x}_2)$ , se define mediante la ecuación

$$d_a(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^p |x_{1j} - x_{2j}|$$

- Distancia de Mahalanobis. Denotada por  $d_m^2(\mathbf{x}_1, \mathbf{x}_2)$ , se define mediante la ecuación

$$d_m^2(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \mathbf{x}_2)' \mathbf{S}^{-1} (\mathbf{x}_1 - \mathbf{x}_2) \quad (2)$$

Donde  $\mathbf{S}$  denota la matriz de covarianzas correspondiente a las  $p$  variables

### 2.1.1.1. Método de análisis clúster.

Montanero (2008) menciona que en función de la medida de afinidad definida se forman el clúster y para ello existen los métodos: a) jerárquico, b) no jerárquicos, cada uno de ellos con sus respectivos procedimientos. Para el

presente caso considerando la naturaleza de la investigación consideramos de relevancia teórica el método jerárquico.

**a. Método jerárquico.**

Donde cada dato se considera como un conglomerado unitario y partiendo de ello los pasos siguientes consistirán en la unión de conglomerados más cercanos y conformar uno más grande, repitiéndose este proceso hasta lograr un solo conglomerado del total de datos. Esto permite analizar diferencias grandes interconglomerados y diferencias pequeñas intraconglomerados mediante un gráfico denominado dendograma. Al ser los conglomerados conjuntos se debe determinar una medida de proximidad entre conjuntos partiendo de la medida  $d$  de proximidad entre puntos seleccionados, para ello existen las siguientes opciones:

- Vecino más cercano donde:  $d(A, B) = \min_{i \in A, j \in B} d(i, j)$
- Vecino más lejano donde:  $d(A, B) = \max_{i \in A, j \in B} d(i, j)$
- Promedio del grupo donde:  $d(A, B) = \frac{1}{n_A \cdot n_B} \sum_{i \in A, j \in B} d(i, j)$
- Centroide donde:  $d(A, B) = d(\bar{X}_A, \bar{X}_B)$

De la Fuente (2018) señala que existen diversas formas de medir la distancia entre clúster, con diversos criterios de selección del algoritmo dependiendo del propósito de la investigación. También indica que de existir más de dos objetos la distancia entre ellos se calcula con la ecuación:

$$d(C, A + B) = \delta_1 d(C, A) + \delta_2 d(C, B) + \delta_3 d(A, B) + \delta_4 |d(C, A) - d(C, B)|,$$

Donde  $\delta_i \equiv$  constante ponderación,  $i = 1, 2, \dots, n$

- Método Ward. De la Fuente (2018) señala que este método al determinar los niveles de agrupamiento es más excluyente y tiene las mismas ventajas del método de K-medias, también menciona que resultados de investigaciones de Fisher muestran que este método es más asertivo con la clasificación óptima que otros métodos como la mínima, máximo, media y centroide, siendo más recomendable su uso en el análisis de conglomerados.

#### b. Método no jerárquico

Cuadras (2019) señala que el método consiste en disponer  $h$  grupos idénticos y diferentes de los  $n$  objetos en concordancia a la matriz de datos  $X$ .

Alaminos et al., (2015) indica que los métodos no jerárquicos resultan apropiados si los casos son en número elevado, entre estos métodos el más tradicional, es el método de k-medias, para ello se debe indicar el número de clústeres a estimar y se aplica a niveles de medición de razón o de intervalo. Otro método se basa en el empleo de las medianas para formar los clústeres. Cabe precisar que con el método no jerárquico se inicia con un número definido de conglomerados y el procedimiento por la construcción de los clústeres varía en base al conocimiento del valor del centro de los grupos, o si por el contrario los centros deben de ser estimados, en ambos casos, siempre partiendo de un número prefijado de clústeres

#### 2.1.2. Diseños experimentales

Respecto a los diseños experimentales, Montgomery (2004) señala que los métodos estadísticos constituyen un enfoque objetivo de análisis si los datos que incluye el problema están sujetos a errores experimentales, por consiguiente el diseño experimental y el análisis estadístico de los datos están estrechamente relacionados

porque el segundo depende del diseño utilizado. Por lo tanto, los diseños estadísticos consideran el proceso de planeamiento del experimento para recoger los datos apropiados y lograr analizarlos con métodos estadísticos para conseguir conclusiones objetivas.

### **2.1.3. Unidad Experimental**

La unidad experimental se considera a la muestra a utilizar para la generación de un valor representativo del resultado de la prueba o experimento. Al disponer un diseño de experimento es de suma importancia definir cuidadosamente la unidad experimental (Gutiérrez y De la Vara, 2008 p. 7).

### **2.1.4. Tratamiento**

Es un valor de una variable, cantidad de un material o un método, se le aplica a una unidad experimental para estudiar el efecto que produce en una variable. En experimentos con cultivos un tratamiento podría ser una variedad de un cultivo (llamada nivel de un factor genotipo), una dosis de un fertilizante nitrogenado, una dosis de un insecticida, una dosis de un fungicida, una dosis de un nematocida, etc. (Di Rienzo et al., 2008 p.179).

### **2.1.5. Variables**

Una variable es una característica, propiedad o atributo, que en base a ello los componentes de una población difieren; en estadística, las mediciones y/o observaciones sobre los elementos de una población vienen a constituir la materia

prima con la que se trabaja y estas están o pueden ser expresadas, en términos numéricos (Di Rienzo et al., 2008 p. 179).

#### **2.1.6. Factores controlables**

Son características de los materiales experimentales que pueden ser fijados en un determinado nivel. Usualmente algunos de ellos se controlan normalmente durante el proceso, y pueden distinguirse ya que existen mecanismos para cambiar el nivel de operación de cada uno de ellos. Esto último es lo que hace posible que se pueda experimentar con ellos. Estos factores denominados controlables son denominados también como variables de entrada, condiciones de proceso, variables de diseño, parámetros del proceso Gutiérrez y De la Vara, (2008 p.8).

#### **2.1.7. Factores no controlables**

Son características de materiales y métodos que durante el experimento no se pueden controlar. Podemos citar entre ellos a la variabilidad climática de temperatura, precipitación, humedad, factores que pueden ser no controlables durante el experimento Gutiérrez y De la Vara, (2008 p.8).

#### **2.1.8. Principios del diseño experimental.**

La evaluación de experimentos en agricultura está expuesto a la variabilidad local de las condiciones externas de los genotipos en evaluación lo que origina considerar los principios básicos de la experimentación para disminuir el error estándar e incrementar la precisión en el uso de la media de la muestra para estimar la media de la población. Al respecto, Burgueño et al. citado por Kohli et al. (2003 p.

247) señala que en el mejoramiento de plantas a partir de la selección de líneas genéticas se debe de cumplir con los siguientes principios experimentales:

**Repetición:** con el objetivo de tener una medida de variabilidad y estimación del error experimental y al considerar un mayor número de repeticiones se tendrá mayor precisión en las estimaciones.

**Aleatorización:** con el objetivo de eliminar el sesgo en la estimación de medias, dar validez a la estimación del error experimental y generar independencia de los mismos. Asimismo, la aleatorización es una condición para utilizar la prueba de significación obtenida con el análisis de varianza.

**Control local:** Con el objeto de evitar efectos extraños que pueden interferir con el objetivo del estudio realizado mediante la construcción de bloques que sean lo más homogéneos posibles que contribuyen a incrementar la precisión del experimento comparando los tratamientos casi en idénticas condiciones experimentales.

### 2.1.9. Estadísticos

Montgomery (2004) señala que un estadístico se define como cualquier función real de las observaciones de una muestra que no contiene parámetros, siendo los más usados la media y varianza muestral.

$$\text{Media muestral } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, \dots, n \quad (3)$$

$$\text{Varianza muestral } S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} ; i = 1, 2, \dots, n \quad (4)$$

#### 2.1.9.1. Método de máxima verosimilitud para la estimación de la media y la varianza poblacional

Para obtener los estimadores puntuales encontramos entre los métodos de estimación el de máxima verosimilitud como uno de los más eficientes que consiste en tomar como parámetro poblacional el valor de la muestra más probable y permite obtener estimadores consistentes y eficientes.

Sea  $x_1, x_2, \dots, x_n$ ,  $n$  observaciones de la variable aleatoria  $X$ , con función de densidad denotado por  $f(x_1, x_2, \dots, x_n)$  que dependen del parámetro  $\theta$ , se establece la función de verosimilitud denotado por  $L(\theta, X)$  y definido como:

$$L(\theta, X) = \prod_{i=1}^n f(x_i)$$

Sea  $X$  proviene de una distribución normal con parámetros  $\mu, \sigma^2$ , denotado por  $X \sim N(\mu, \sigma^2)$ , con función de densidad

$$f(x_i) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} ; -\infty < x_i < +\infty; -\infty < \mu < +\infty; \sigma^2 > 0$$

entonces la función de verosimilitud denotado por  $L(\theta, X)$ ; donde  $\theta = (\mu, \sigma^2)$  y  $X = (x_1, x_2, \dots, x_n)$

$$L(\theta, X) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \right) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$L(\theta, X) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}$$

Aplicando la función de log-verosimilitud expresado por  $l(\theta) = \log L(\theta, X)$ .

$$l(\theta) = \log L(\theta, X) = \log \frac{1}{(\sqrt{2\pi\sigma})^n} e^{-\frac{1}{2\sigma^2}\sum (x_i-\mu)^2} = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2}$$

Para encontrar el estimador máximo verosímil del vector de parámetro  $\theta$ , aplicando la derivada parcial a  $l(\theta)$  se tiene

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 .$$

### 2.1.9.2. Propiedad de insesgamiento de los estimadores de la media y varianza muestral

Entre las propiedades de estos dos estimadores podemos indicar que  $(\hat{\mu}, \hat{\sigma}^2)$  son estimadores puntuales de la media y varianza poblacional  $(\mu, \sigma^2)$ , siendo el estimador puntual una variable aleatoria que se calcula a partir de los datos muestrales donde dos de sus más importantes propiedades, debe ser insesgado y debe tener una varianza mínima. Para demostrar que  $(\hat{\mu}, \hat{\sigma}^2)$ , son estimadores insesgados de  $(\mu, \sigma^2)$  a través del valor esperado, según Montgomery (2004) se tiene:

$$\begin{aligned} E(\hat{\mu}) &= E\left[\frac{\sum_{i=1}^n x_i}{n}\right] \\ &= \frac{1}{n} E\left(\sum_{i=1}^n X\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(x_i) \\ E(\hat{\mu}) &= \frac{1}{n} \sum_{i=1}^n \mu \\ &= \mu \end{aligned}$$

luego  $\hat{\mu}$  es un estimador insesgado de  $\mu$ . Por otro lado, para la varianza se tiene:

$$\begin{aligned} E(\hat{\sigma}^2) &= E\left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}\right] \\ &= \frac{1}{n-1} E\sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n-1} E(SS) \end{aligned}$$

Donde  $SS = \sum_{i=1}^n (x_i - \bar{x})^2$  es la suma de cuadrados corregida de las observaciones  $x_i$ . Entonces

$$\begin{aligned} E(SS) &= E \left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right] \\ &= E \left[ \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right] \\ &= \sum_{i=1}^n (\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n) \\ &= (n - 1)\sigma^2 \end{aligned}$$

Por lo tanto,  $E(\hat{\sigma}^2) = \frac{1}{n-1} E(SS) = \sigma^2$ .

Entonces  $\hat{\sigma}^2$  es un estimador insesgado de  $\sigma^2$ .

### 2.1.10. Análisis de Regresión lineal simple

Tiene por objeto explicar el comportamiento de una variable respuesta (Rendimiento del trigo harinero) en función de una o más variables independientes o predictoras (Altura de planta, número de granos por espigas, tamaño de espiga, peso de mil granos, madurez fisiológica) y conocer cómo los cambios de las variables predictoras afectan a la variable respuesta; por consiguiente, tal como señala Gutierrez y De la Vara (2008 p. 340), cada observación de  $Y$  puede ser descrita mediante el siguiente modelo.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma^2) \quad (5)$$

Donde:

$Y_i$  = variable dependiente observada bajo el  $i$ -ésimo observación de  $X$ ,  $i = 1, \dots, k$

$X_i$  = valor  $i$ -ésimo de la variable independiente,  $i = 1, \dots, k$

$\beta_0$  = representa el intercepto (indica valor esperado de  $Y$  cuando  $X=0$ )

$\beta_1$  = representa la pendiente de la recta (indica el cambio de Y frente al cambio unitario en X).

$\varepsilon_i$  = variación aleatoria (no explicada por el modelo) asociada a la observación de Y bajo la observación  $X_i$ .

Los  $\varepsilon_i$  se deducen ser normales distribuidos independientemente con esperanza 0 y varianza constante  $\sigma^2$  para todo X en un intervalo donde el modelo se deduce verdadero. Esto es  $\varepsilon_i \sim N(0, \sigma^2)$

El modelo (5) se aplica para una variable independiente y ver que la esperanza de la variable dependiente es modificada con tasa constante, según aumenta o disminuye el valor de la variable independiente. Por consiguiente, para un valor esperado de Y siendo fijo un determinado valor de X y por propiedades de la esperanza matemática se tiene:

$$E(Y|X) = \beta_0 + \beta_1 x$$

donde  $\beta_0$  y  $\beta_1$  son constantes desconocidas que representan los parámetros del modelo y en la ecuación que relaciona las dos variables se debe estimar el parámetro  $\beta_0$  punto donde la línea recta intercepta al eje y, y el parámetro  $\beta_1$  pendiente de la recta y su cantidad aumenta o reduce la variable Y por cada unidad de X.

Uno de los procedimientos para ajustar la recta Y es estimar  $\beta_0$  y  $\beta_1$  mediante el método de mínimos cuadrados que consiste en despejar los errores de la ecuación (5), se eleva al cuadrado, se realiza la suma de cuadrados del error (SCE) obteniendo lo siguiente:

$$SCE = \sum_{i=1}^n (\varepsilon_i)^2 = \sum_{i=1}^n (y_i - [\beta_0 + \beta_1 x_i])^2$$

El procedimiento matemático para minimizar los errores de la ecuación anterior y encontrar los estimadores de mínimos cuadrados de  $\beta_0$  y  $\beta_1$ , consiste en derivar a SCE con respecto a  $\beta_0$ ,  $\frac{\partial SCE}{\partial \beta_0}$  y derivar también a  $\beta_1$ ,  $\frac{\partial SCE}{\partial \beta_1}$ , donde se obtienen:

$$\frac{\partial SCE}{\partial \beta_0} = \sum_{i=1}^n 2(y_i - [\beta_0 + \beta_1 x_i]) \quad (6)$$

$$\frac{\partial SCE}{\partial \beta_1} = \sum_{i=1}^n 2x_i (y_i - [\beta_0 + \beta_1 x_i]) \quad (7)$$

Al igualar a cero las ecuaciones (6) y (7) y resolverlas simultáneamente con respecto a  $\beta_0$  y  $\beta_1$ , se obtiene la siguiente solución única:

$$\hat{\beta}_1 = \frac{SCE_{xy}}{SCE_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Donde

$$SCE_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}$$

$$SCE_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}$$

Siendo  $\bar{x}$  y  $\bar{y}$ , las medias muestrales de las dos variables X y Y.

## 2.1.11. Medidas de Bondad de ajuste de la regresión

### 2.1.11.1. Coeficiente de determinación $R^2$

Montgomery (2004) señala que el coeficiente de determinación mide la proporción de la variación en Y que es explicada por la relación con X, cuyo cálculo se desarrolló a través de la siguiente expresión:

$$R^2 = 1 - \frac{SCE}{SCTotal} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (8)$$

Donde:  $Y_i$   $i$ -ésimo valor de respuesta observada,  $\bar{Y}$  respuesta media,

$\hat{Y}_i$   $i$ -ésima respuesta ajustada.

Cuando se agregan términos al modelo de regresión,  $R^2$  se incrementa siempre por lo tanto es preferible utilizar el estadístico  $R_{aj}^2$  ajustado que se define de la siguiente manera:

$$R_{aj}^2 = 1 - (1 - R^2) \left( \frac{n - 1}{n - p} \right)$$

Donde:  $p$  es el número de parámetros y  $n$  número de observaciones

El coeficiente de determinación se basa en la descomposición de la varianza total, donde:  $Y_t = \hat{Y}_t + \hat{\varepsilon}_t$ , restando en ambos miembros  $\bar{Y}$ , se tiene que

$$Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + \hat{\varepsilon}_i$$

Elevando al cuadrado ambos miembros se obtiene que

$$(Y_i - \bar{Y})^2 = [(\hat{Y}_i - \bar{Y}) + \hat{\varepsilon}_i]^2$$

Es decir,

$$(Y_i - \bar{Y})^2 = (\hat{Y}_i - \bar{Y})^2 + \hat{\varepsilon}_i^2 - 2\hat{\varepsilon}_i(\hat{Y}_i - \bar{Y})$$

Sumando ambos miembros de la expresión anterior, se tiene

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2 - 2 \sum_{i=1}^n \hat{\varepsilon}_i (\hat{Y}_i - \bar{Y})$$

En la ecuación anterior puede verse que el tercer término del segundo miembro es

$$2 \sum_{i=1}^n \hat{\varepsilon}_i (\hat{Y}_i - \bar{Y}) = 2 \sum_{i=1}^n \hat{\varepsilon}_i \hat{Y}_i - 2\bar{Y} \sum_{i=1}^n \hat{\varepsilon}_i = 0$$

Considerando que  $\sum_{i=1}^n \hat{\varepsilon}_i = 0$  y  $\sum_{i=1}^n \hat{\varepsilon}_i X_i = 0$ , entonces

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2$$

Si la expresión anterior se divide en ambos miembros entre  $n$ , se obtiene

$$\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{n} + \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n}$$

Por lo tanto, la varianza total se descompone en dos partes: varianza explicada por la regresión y varianza residual y el coeficiente de determinación está definido como uno menos la proporción de la varianza total explicada por la regresión. Siendo su expresión la siguiente

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

#### 2.1.11.2. Error Cuadrático Medio de Predicciones (ECMP)

Se utiliza para evaluar la calidad de un estimador o conjunto de predicciones por su variación y el grado de sesgo, se calcula a través de la siguiente fórmula:

$$ECMP = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (9)$$

Donde:  $\hat{Y}_i$  es el valor predicho o ajustado,  $Y_i$  es el valor observado y para efectos de la interpretación de los resultados utilizaremos raíz cuadrada del ECMP que viene a ser la desviación estándar que tiene las mismas unidades que la cantidad que se estima:  $s = \sqrt{ECMP}$ , (Di Rienzo et al.,2008).

#### 2.1.11.3. Coeficiente de Correlación de Pearson

Balzarini et al. (2011), indica que “la correlación de Pearson tiene la finalidad de buscar las relaciones existentes entre variables y llegar a cuantificar la

magnitud de estas asociaciones". Para ello, el coeficiente de correlación lineal poblacional entre las variables aleatorias  $X$  e  $Y$  se define como:

$$\rho_{x,y} = \frac{cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \quad (10)$$

Según la proposición  $Cov(U,V)^2 \leq Var(U) \cdot Var(V)$ , en la demostración podemos observar que para cualquier constante  $a$ , y para cualquier variable aleatoria  $U$  y  $V$  se tiene que

$$E[(aU + V)^2] \geq 0$$

desarrollando el cuadrado y aplicando las propiedades de la esperanza resulta:

$$a^2E[U^2] + 2aE[UV] + E[V^2] \geq 0$$

La ecuación representa una parábola que toca al eje de las abscisas máximo en un punto, siendo la ecuación,

$$a^2E[U^2] + 2aE[UV] + E[V^2] = 0$$

lo que significa que su discriminante debe ser menor o igual a cero, esto es

$$4(E[UV])^2 - 4E[U^2]E[V^2] \leq 0$$

Que es lo mismo indicar

$$(E[UV])^2 \leq E[U^2]E[V^2]$$

Si consideramos que  $U = (X - E[X])$  y  $V = (Y - E[Y])$  se demuestra la proposición

$$-1 \leq \rho_{x,y} \leq 1$$

Si  $|\rho| = 1$  entonces  $Y = aX + b$ , donde los valores de  $(X, Y)$  se disponen a lo largo de una recta y para demostrar este argumento existe una constante  $a$  tal que

$$E[(aU + V)^2] = 0$$

Ya que los términos  $(aU + V)^2$  al ser un cuadrado por la única forma de que su esperanza sea 0, siempre son iguales o mayores a cero, es que

$$aU + V = 0.$$

Por lo tanto, si  $U = X - E[X]$  y  $V = Y - E[Y]$

$$a(X - E[X]) + Y - E[Y] = 0$$

De donde

$$Y = aX - aE[X] + E[Y]$$

Si denominamos

$$b = -aE[X] + E[Y]$$

se tiene que

$$Y = aX + b$$

que demuestra que la relación es lineal entre  $X$  e  $Y$ ; por consiguiente, el coeficiente de correlación lineal muestral de Pearson denotado por  $r$ , se define como:

$$r = \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\sqrt{\left(\left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}\right)\left(\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n}\right)\right)}} \quad (11)$$

Para probar la independencia entre las variables  $X$  e  $Y$  se plantea las siguientes hipótesis estadísticas:

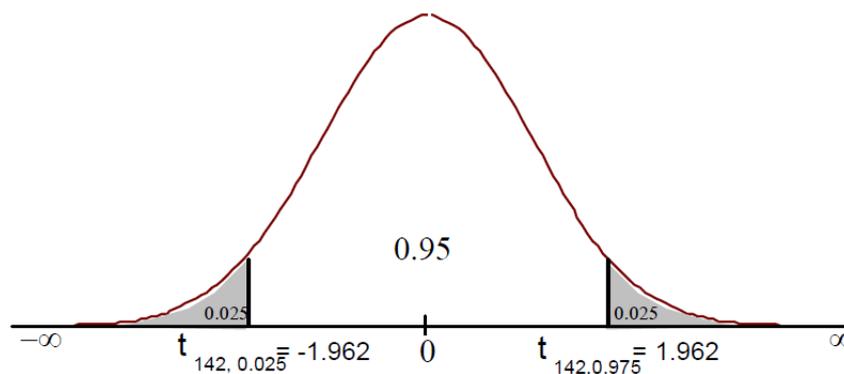
$H_0: \rho = 0$  que implica que  $X$  e  $Y$  son estadísticamente independientes

$H_1: \rho \neq 0$ , que implica que hay asociación entre  $X$  e  $Y$  pero no implica relación de causalidad.

El estadístico de la prueba:

$$T = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}, \quad (12)$$

Donde  $T$  sigue una distribución de  $t$  de Student con  $n-2$  grados de libertad, siendo  $n$  el número de pares  $(X, Y)$  con nivel de significación  $\alpha = 0.05$ , se determina la región de no rechazo de la  $H_0$  como intervalo delimitado por los cuantiles 0.025 y 0.975 con una  $t_{(n-2)}$



**Figura 1.** Determinación de la región de no rechazo (región no sombreada) con valores calculados para la prueba de independencia  $(X, Y)$  utilizando el estadístico  $T$  de Student.

### 2.1.12. Análisis de varianza

Según Gutierrez y De la Vara, (2008 p.66), el análisis de varianza (ANVA) o (ANOVA por sus siglas del inglés analysis of variance) es la técnica central en el análisis de datos experimentales que permite probar las hipótesis de igualdad de los tratamientos con respecto a la media de la variable respuesta a partir de la separación de la variación total en las partes con las que contribuye cada fuente de variación en el experimento.

### 2.1.13. Análisis de varianza de modelos estadísticos lineales

Con respecto al análisis de varianza con el criterio de clasificación de efectos fijos y aleatorios se tiene en cuenta el siguiente modelo lineal:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}; \quad \varepsilon_{ij} \sim N(0, \sigma^2) \text{ con } i = 1, 2, \dots, a; j = 1, 2, \dots, b \quad (13)$$

en que  $Y_{ij}$  es la observación  $i$ -ésimo tratamiento,  $j$ -ésima repetición,  $\mu$  es el parámetro de escala común a los tratamientos, denominado media global;  $\tau_i$ , mide el efecto del tratamiento  $i$ -ésimo y  $\varepsilon_{ij}$ , es el error aleatorio que incorpora todas las fuentes de variabilidad del experimento de aquellos factores no controlables.

La expresión (13) es conocida como el modelo de los efectos, este modelo describe dos situaciones diferentes en relación a los efectos de los tratamientos  $\tau_i$ . Según Montgomery (2004) señala: "Primero, los  $\tau$  tratamientos pudieron ser elegidos expresamente por el investigador. En este caso se prueban las hipótesis acerca de las medias de los tratamientos, aplicando las conclusiones solo a los niveles del factor considerado en el análisis. No pueden ampliarse estas conclusiones a tratamientos similares, llamándose a este, modelos de efectos fijos. Segundo, los  $\tau$  tratamientos se pueden obtener de una muestra aleatoria de una población más grande de tratamientos, en este caso las conclusiones se extienden a los tratamientos de la población, sea que se hayan considerado explícitamente en el análisis o no. En este caso se prueban hipótesis acerca de la variabilidad de los tratamientos y se intenta estimar su variabilidad".

Sea el modelo (13), entonces

$$y_{i.} = \sum_{j=1}^n y_{ij}, \quad \bar{y}_{i.} = \frac{y_{i.}}{n}; \quad i = 1, 2, \dots, a \quad (14)$$

$$y_{..} = \sum_{i=1}^a \sum_{j=1}^n y_{ij}, \quad \bar{y}_{..} = \frac{y_{..}}{N} \quad (15)$$

En que  $N = an$ , corresponde al número total de observaciones. donde el subíndice "punto" implica la operación suma sobre el subíndice que reemplaza.

La finalidad es probar la igualdad de las  $a$  medias de los tratamientos; es decir:

$$E(y_{ij}) = \mu + \tau_i = \mu_i, i = 1, 2, \dots, a$$

Las hipótesis apropiadas son:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a$$

$$H_1: \mu_i \neq \mu_j \text{ al menos para un par } (i, j)$$

En el modelo de los efectos fijos, la media  $\mu_i$  del tratamiento  $i$ -ésimo se descompone en dos componentes tales que:  $\mu_i = \mu + \tau_i$

Por lo general,  $\mu$ , se considera como una media global, donde:

$$\frac{\sum_{i=1}^a \mu_i}{a} = \mu, \text{ bajo la condición } \sum_{i=1}^a \tau_i = 0$$

Una forma equivalente de escribir las hipótesis anteriores en términos de los efectos de los tratamientos  $\tau_i$  es de la siguiente manera:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_a$$

$$H_1: \tau_i \neq 0 \text{ para al menos una } i$$

Por consiguiente, permite probar la igualdad de las medias de los tratamientos en el análisis de varianza (Ver Tabla 1).

**Tabla 1.** Estructura del análisis de varianza para el modelo con un solo factor de efectos fijos

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Prueba F
Entre los tratamientos	$SS_{Trat} = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2$	$a - 1$	$MC_{Trat}$	$F_0 = \frac{MS_{Trat}}{MS_E}$
Error (dentro de los tratamientos)	$SS_E = SS_T - SS_{Trat}$	$N - a$	$MS_E$	
Total	$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$	$N - 1$		

Donde:

$y_i$  = Suma de las observaciones del tratamiento  $i$

$\bar{y}_{..}$  = Media global o promedio de todas las observaciones

$\bar{y}_i$  = Media de las observaciones del  $i$ -ésimo tratamiento

$y_{..}$  = suma total de las  $N = n_1 + n_2 + \dots + n_k$  mediciones

$SS_{Trat}$  = Suma de cuadrado de los tratamientos

$SS_E$  = Suma de cuadrado de los residuos

$SS_T$  = Suma de cuadrado del total

$MC_{Trat}$  = Cuadrado medio de los tratamientos

$MS_E$  = Cuadrado medio de los residuos

Siguiendo con la descripción del ANOVA, según Montgomery (2004), la denominación de análisis de varianza deriva de la partición de la variabilidad total de las observaciones en sus partes componentes, por lo tanto, la suma de cuadrados del total se puede escribir como:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n [(\bar{y}_i - \bar{y}_{..}) + (y_{ij} - \bar{y}_i)]^2 \quad (16)$$

o

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (y_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 + 2 \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.}) \quad (16)$$

sin embargo, el término del producto cruzado de la ecuación (16) es cero, ya que

$$\sum_{j=1}^n (y_{ij} - \bar{y}_{i.}) = y_{i.} - n\bar{y}_{i.} = y_{i.} - n(y_{i.}/n) = 0$$

por lo tanto, se obtiene

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (y_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 \quad (17)$$

esta expresión es conocida como la ecuación fundamental del análisis de varianza y cada uno de sus términos se denomina suma de cuadrados que puede ser abreviado en la expresión:

$$SST = SS_{Trat} + SS_E$$

En que:

$$SS_{Trat} = n \sum_{i=1}^a (y_{i.} - \bar{y}_{..})^2$$

$$SS_E = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

#### 2.1.14. Verificación del modelo adecuado

Al analizar la verificación y la adecuación del modelo Montgomery (2004) indica: *“Cuando se descompone la variabilidad que está en las observaciones a través de la identidad del análisis de varianza, es una relación netamente algebraica. Por lo tanto, la utilidad de la partición para probar que no hay diferencias en las medias de los tratamientos necesita satisfacer algunos supuestos”*. De la expresión (13), se tiene que,  $\epsilon_{ij} \sim N(0, \sigma^2)$ , por consiguiente, se obtiene  $Y_{ij} \sim N(\mu + \tau_i, \sigma^2)$ .

El mismo autor señala: “Particularmente los supuestos indican que la expresión (13) describe adecuadamente las observaciones  $(Y_{ij})$ , donde los errores  $(\varepsilon_i)$  siguen una distribución normal e independiente con media cero y varianza  $\sigma^2$  constante pero desconocida. Si se satisfacen estos supuestos, el procedimiento del análisis de varianza es una prueba exacta de la hipótesis de que no hay diferencias en las medias de los tratamientos. Por lo tanto, en la práctica es común que no se satisfagan exactamente estos supuestos. Por consiguiente, no es recomendable confiar en el análisis de varianza hasta haber verificado los supuestos” (p. 76).  $\varepsilon_{ij}$ , que vienen a ser los residuos están definidos como la diferencia existente entre la respuesta observada  $(Y_{ij})$  y la respuesta predicha por el modelo  $(\hat{Y}_{ij})$ , permitiendo hacer el diagnóstico directo de la calidad del modelo donde también su magnitud describe a los datos. La verificación de los supuestos del modelo será visto en la sección 2.1.16

#### **2.1.15. Diseños de bloques completos al azar**

Respecto a los diseños de bloques completos al azar Montgomery, (2004) señala “*en cualquier experimento, la variabilidad que surge de un factor perturbador puede afectar los resultados*” (p.126).

La asignación aleatoria de tratamientos según señala Di Rienzo et al.( 2008), es una manera de ser imparciales en distribuir los tratamientos de manera igual en cada uno de los bloques; “si estas diferencias pueden ser anticipadas entonces es conveniente agruparlos en bloques y aleatorizar los tratamientos, logrando que cada bloque sea una repetición completa. A este arreglo se denomina **diseño en bloques completos al azar**. Indicamos que son completos porque en cada bloque aparecen todos los tratamientos, y al azar porque dentro de cada bloque los tratamientos son

distribuidos aleatoriamente” (p. 251). El modelo lineal para un análisis de la varianza con un factor en un diseño de bloques completos al azar es el siguiente:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad (18)$$

Donde:  $Y_{ij}$  es la respuesta del  $i$ -ésimo tratamiento en el  $j$ -ésimo bloque

$\mu$  es la media general.

$\tau_i$  es el efecto del  $i$ -ésimo tratamiento  $i=1, \dots, a$

$\beta_j$  es el efecto del  $j$ -ésimo bloque  $j=1, \dots, b$

$\varepsilon_{ij}$  es el término del error aleatorio.

El mismo autor Di Rienzo et al. (2008) señala “Del modelo lineal original (ecuación 13) para el diseño de bloques completos al azar sólo se ha agregado el término  $\beta_j$ . que puede modelar un efecto fijo o aleatorio con distribución normal independiente, con esperanza cero y varianza  $\sigma^2$  e independiente del término de error. Este término modela la variación introducida por los bloques y tiene por finalidad reducir el error experimental”.

**Tabla 2.** Esquema típico para un experimento en bloques completos al azar

Tratamientos	Bloque 1	Bloque 2	...	Bloque b	Total
1	$y_{11}$	$y_{12}$	...	$y_{1b}$	$y_{1\cdot}$
2	$y_{21}$	$y_{22}$	...	$y_{2b}$	$y_{2\cdot}$
..	..	..	.....	..	..
a	$y_{a1}$	$y_{a2}$	...	$y_{ab}$	$y_{a\cdot}$
Total	$y_{\cdot 1}$	$y_{\cdot 2}$		$y_{\cdot b}$	$y_{\cdot \cdot}$

Fuente: Di Rienzo et al.(2008 p.234)

En resumen, considerando la expresión (18) los cálculos para un diseño de bloques completos al azar bajo el modelo de efectos fijos, las hipótesis estadísticas que se

someten a prueba en un ANOVA para este diseño están establecida sobre las medias de las poblaciones relacionadas a cada tratamiento:

$$(\mu_i = \mu + \tau_i \text{ con } i = 1, \dots, a):$$

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a,$$

$H_1$ : Al menos un par de medias poblacionales difiere

se presentan en la Tabla 3.

**Tabla 3.** Esquema típico para un ANOVA de bloques completos al azar

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F
Bloques	$SCB = \sum_{j=1}^a \frac{(y_{.j})^2}{a} - \frac{(y_{..})^2}{ab}$	g <sub>lb</sub> =b-1	$CMB = \frac{SCB}{g_{lb}}$	$\frac{CMB}{CME}$
Entre tratamientos	$SCTrat = \sum_{i=1}^b \frac{(y_{i.})^2}{b} - \frac{(y_{..})^2}{ab}$	g <sub>le</sub> =a-1	$CMTrat = \frac{SCTrat}{g_{le}}$	$\frac{CMTrat}{CME}$
Dentro (Error Experimental)	SCE=SCT-SCE-SCB	g <sub>ld</sub> =(a-1)(b-1)	$CME = \frac{SCE}{g_{ld}}$	
Total	$SCT = \sum_{i=1}^a \sum_{j=1}^b y_{ij} - \frac{(y_{..})^2}{ab}$	g <sub>lt</sub> =ab-1		

Fuente: Di Rienzo et al.(2008 p.235)

$SCB$ : Suma de cuadrados de bloques.

$SCTrat$ : Suma de cuadrados de tratamientos.

$SCE$ : Suma de cuadrados de los residuos o error experimental.

$SCT$ : Suma de cuadrados del total.

$CMB$ : Cuadrado medio de bloques.

$CMTrat$ : Cuadrado medio de tratamientos.

$CME$ : Cuadrado medio de los residuos o el error experimental.

Así mismo, utilizamos el coeficiente de variación para conocer el grado de precisión del experimento a través de la información con respecto a la relación porcentual entre

la variabilidad residual y la media de los datos calculado a través de la siguiente fórmula:

$$CV = \frac{\sqrt{CMError}}{Media\ general} \quad (19)$$

Pimentel, G (1990) citado por Álvares et al., (2017) propone la siguiente clasificación del coeficiente de variabilidad en base de trabajos desarrollados en cultivos como el maíz, arroz entre otros que son especies que botánicamente pertenecen a la misma familia del trigo siendo esta la siguiente:

- Bajo: Coeficiente de variabilidad inferior al 10%
- Medio: Coeficiente de variabilidad entre 10 a 20%
- Alto: Coeficiente de variabilidad entre 20 a 30%
- Muy alto: Coeficiente de variabilidad superior al 30%

Así mismo Balzarini et al., (2011) manifiesta también que cuando el coeficiente de variabilidad del análisis de varianza es menor, la información en el estudio es de mejor calidad porque brinda la información en porcentaje sobre la relación que existe entre la variabilidad residual y la media de los datos.

#### **2.1.16. Supuestos del modelo**

En un análisis de varianza, los resultados obtenidos para ser válidos están supeditada a que se cumplan los supuestos del modelo tal como se mencionaron en la sección 2.1.15. Según Gutierrez y De la Vara (2008), señalan que estos supuestos son: normalidad, varianza constante e independencia. Suponiendo que el número de unidades experimentales para cada bloque coincide con el número de tratamientos, la variable respuesta  $Y$  puede depender de un primer factor de interés (A) y de la variable bloque (B) (p.81).

Sobre el modelo en necesario considerar las siguientes hipótesis estadísticas:

**a) Normalidad:**  $\varepsilon_{ij}$  sigue una distribución normal. Lo que es equivalente a que  $Y_{ij}$  sigue una distribución normal. Para la normalidad utilizamos la prueba de Shapiro-Wilks con las siguientes hipótesis estadísticas planteadas:

$H_0$ : Los residuos siguen una distribución normal

$H_1$ : Los residuos no siguen una distribución normal

Si el  $p$ -valor es mayor a nivel de significación  $\alpha = 0.05$  no rechazamos la hipótesis nula.

La prueba de Shapiro - Wilks será visto en la sección 2.1.17.a.

**b) Linealidad:**  $E(\varepsilon_{ij}) = 0$ , esto es equivalente a que  $E(Y_{ij}) = \mu + \tau_i + \beta_j$ .

**c) Homocedasticidad:**  $Var(\varepsilon_{ij}) = \sigma^2$ . Esto es equivalente a que  $Var(Y_{ij}) = \sigma^2$ , para la homogeneidad de la varianza de los residuos se utiliza la prueba de Bartlett con las siguientes hipótesis estadísticas.

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_\alpha^2$

$H_1$ : Al menos dos varianzas son distintas

Si el  $p$ -valor es mayor a nivel de significación  $\alpha = 0.05$  no rechazamos la hipótesis nula

La prueba de Bartlett será visto en la sección 2.1.17.b.

**d) Independencia:**  $\varepsilon_{ij}$  son independientes entre sí. Esto es equivalente a que  $Y_{ij}$  son independientes entre sí.

### Observación:

Desde el punto de vista de la evaluación de genotipos procedentes de cruzamientos es posible no cumplir en su integridad los supuestos del análisis de varianza, bajo el análisis de Cubillos citado por Kohli et al.(2003 p. 206-218), señala que al utilizar genes particulares en la selección de cultivares de alto rendimiento se percibe incorporar la noción de un mejoramiento genético para

heterogeneidad de los genotipos, especialmente utilizando métodos biotecnológicos.

Según Fernández Escobar et al. (2017) el diseño de bloques completos al azar en la investigación agrícola *“permite reducir y controlar la varianza del error experimental considerando que las unidades experimentales son relativamente homogéneas con respecto a factores que afectan la variable de respuesta y que las diferencias observadas sean específicamente debido a los tratamientos”*.

### 2.1.17. Pruebas para la normalidad y homogeneidad de varianzas

a) **Prueba de Shapiro Wilks**, es una prueba estadística que permite verificar si los datos determinados  $(X_1, X_2, \dots, X_n)$  siguen una distribución normal planteando las siguientes hipótesis.

$H_0$ : Los residuos siguen una distribución normal

$H_1$ : Los residuos no siguen una distribución normal

$$H_0 : X \sim N(\mu, \sigma^2) \text{ y } H_1 : X \neq N(\mu, \sigma^2)$$

Siendo el estadístico de prueba

$$W = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (20)$$

Donde:

$x_i$  es el número que ocupa la *i-ésima* posición en la muestra.

$\bar{x} = \frac{(x_1 + \dots + x_n)}{n}$  es la media muestral

Las variables  $a_i$  se calculan

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

Donde  $m = (m_1, \dots, m_n)^T$ , siendo  $m_1, \dots, m_n$  los valores del estadístico medio ordenado de variables aleatorias independientes e idénticamente distribuidas, muestreadas de distribuciones normales y  $V$  denota la matriz de covarianzas de ese estadístico de orden. El valor de  $W$  puede oscilar entre 0 y 1.

**b) Prueba de Bartlett**, permite contrastar si más de 2 muestras presentan igualdad de varianzas (homocedasticidad). que es de vital importancia para otros análisis estadísticos que requieren la igualdad de varianzas de los datos para poder llevarlos a cabo. Para esta prueba se plantean las siguientes hipótesis.

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_\alpha^2$$

$H_1$ : Al menos dos varianzas son distintas

Si tenemos  $k$  muestras con tamaño  $n_i$  y varianzas de las muestras  $S_i^2$  entonces el estadístico de prueba de Bartlett es:

$$X^2 = \frac{(N - k) \ln(S_p^2) - \sum_{i=1}^k (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(k-1)} \left( \sum_{i=1}^k \left( \frac{1}{n_i - 1} \right) - \frac{1}{N - k} \right)} \quad (21)$$

Donde  $N = \sum_{i=1}^k n_i$  y  $S_p^2 = \frac{1}{N - k} \sum_{i=1}^k (n_i - 1) S_i^2$  es la estimación combinada de la varianza.

El estadístico de prueba tiene aproximadamente una distribución  $X_{k-1}^2$ . En qué, la hipótesis nula se rechaza si  $X^2 > X_{k-1, \alpha}^2$  es el valor crítico de la cola superior para la distribución (donde  $X_{k-1, \alpha}^2$  es el valor crítico de la cola superior para la distribución  $X_{k-1}^2$ ).

**c) Prueba de Breusch\_Pagan**, se utiliza para determinar si la homogeneidad de varianzas está presente o no en un modelo de regresión y consiste en ajustar un modelo de regresión lineal con variable respuesta dada por residuales del modelo original al cuadrado  $e_i^2$  y como covariables las variables del modelo

original. Si se tiene  $k=2$  covariables que explican a  $Y$ , entonces el modelo de regresión para la homogeneidad de varianzas es:

$$\hat{e}_i^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 \mu \quad (22)$$

### 2.1.18. Estimación por máxima verosimilitud de los parámetros del modelo lineal de un diseño de bloques completos al azar

Wackerly et al. (2002), citado por Vianey (2015) plantean que son elegidos como estimadores que maximizan la verosimilitud a los valores de los parámetros de la muestra, indicando que estos son insesgados, consistentes y de varianza mínima por tal motivo considerando que estos estimadores poseen buenas propiedades, siendo un método muy eficiente para el análisis de modelos lineales.

Para ello Lara (1998), en el caso de los diseños de bloques completamente al azar, señala que el método de máxima verosimilitud consiste en encontrar los parámetros que maximizan una función de verosimilitud donde la hipótesis de normalidad con respecto a los términos de error lleva el hecho que las variables  $Y_{ij}$  sean normales e independientes, al construir la función de verosimilitud asociada a la muestra  $\mathbf{y} = (y_{11}, \dots, y_{1J}, \dots, y_{I1}, \dots, y_{IJ})$  indica que:

$$L(\mu, \tau_i, \beta_j, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^J [y_{ij} - \mu - \tau_i - \beta_j]^2\right) \quad (23)$$

donde:

$N = IJ$  número total de observaciones,  $y_{i.}$  es el total de observaciones bajo el  $i$ -ésimo tratamiento,  $y_{.j}$  es el total de observaciones bajo el  $j$ -ésimo bloque

Para los parámetros  $\mu, \tau_i, \beta_j, \sigma^2$ , los estimadores de máxima verosimilitud son valores donde la función de verosimilitud alcanza su máximo y es necesario obtener los

puntos críticos de la función (18) para determinarlos. Por conveniencia, se determina el logaritmo de dicha función en lugar de maximizar la función de verosimilitud.

$$\ln(L(\mu, \tau_i, \beta_j, \sigma^2)) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^J [y_{ij} - \mu - \tau_i - \beta_j]^2 \quad (24)$$

Y se hallan las derivadas parciales respecto de los parámetros del modelo que son:

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^J [y_{ij} - \mu - \tau_i - \beta_j] \\ \frac{\partial \ln L}{\partial \tau_i} &= \frac{1}{\sigma^2} \sum_{j=1}^J [y_{ij} - \mu - \tau_i - \beta_j] \quad i = 1, \dots, I \end{aligned} \quad (25)$$

$$\frac{\partial \ln L}{\partial \beta_j} = \frac{1}{\sigma^2} \sum_{i=1}^I [y_{ij} - \mu - \tau_i - \beta_j] \quad i = 1, \dots, J$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^I \sum_{j=1}^J [y_{ij} - \mu - \tau_i - \beta_j]^2$$

Haciendo la igualdad de las derivadas parciales a cero, obtenemos el sistema de ecuaciones que proporciona los estimadores máximos verosímiles, determinados por las siguientes ecuaciones

$$\hat{\mu} = \frac{\sum_{i=1}^I \sum_{j=1}^J y_{ij}}{N} = \bar{y}_{..} \quad , \quad (26)$$

$$\hat{\tau}_i = \frac{1}{J} \sum_{j=1}^J y_{ij} - \hat{\mu} = \bar{y}_{i.} - \bar{y}_{..} \quad (27)$$

$$\hat{\beta}_j = \frac{1}{I} \sum_{i=1}^I y_{ij} - \hat{\mu} = \bar{y}_{.j} - \bar{y}_{..} \quad (28)$$

Por consiguiente, se puede comprobar que

$$\sum_i \hat{\tau}_i = 0$$

$$\sum_j \hat{\beta}_j = 0$$

Donde,  $I-1$  y  $J-1$  son grados de libertad asociados a los tratamientos y bloques, respectivamente.

Por consiguiente, cambiando  $\hat{\mu}$ ,  $\hat{\tau}_i$  y  $\hat{\beta}_j$  en la ecuación (25) igualada a cero, obtenemos el estimador de máxima verosimilitud para la varianza poblacional

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^J [y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j]^2 \quad (29)$$

Por otro lado, los residuos se definen como las diferencias entre los valores observados  $y_{ij}$  y los valores previstos por el modelo  $\hat{y}_{ij}$  y los denotamos por  $e_{ij}$

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j = y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}$$

Por lo tanto, el estimador de máxima verosimilitud  $\hat{\sigma}^2$ , se puede escribir como

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^I \sum_{j=1}^J e_{ij}^2}{N}$$

Se verifica que la suma de los residuos por filas y columnas es cero, por tanto:

$$\begin{aligned} \sum_j e_{ij} &= \sum_j (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j) = J\bar{y}_{i\cdot} - J\bar{y}_{\cdot\cdot} - J\hat{\tau}_i - \sum_j \hat{\beta}_j = J\bar{y}_{i\cdot} - J\bar{y}_{\cdot\cdot} - J(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \\ &= 0; \quad i = 1, \dots, I \end{aligned}$$

$$\begin{aligned} \sum_i e_{ij} &= \sum_i (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j) = I\bar{y}_{\cdot j} - I\bar{y}_{\cdot\cdot} - \sum_i \hat{\tau}_i - I\hat{\beta}_j = I\bar{y}_{\cdot j} - I\bar{y}_{\cdot\cdot} - I(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) \\ &= 0; \quad j = 1, \dots, J \end{aligned}$$

Por lo tanto,  $(I-1)(J-1)$  residuos independientes, ya que

$$IJ - (I+J-1) = (I-1)(J-1)$$

### 2.1.19. Propiedades de los estimadores de máxima verosimilitud

Lara (1998), menciona algunas propiedades que verifican los estimadores del modelo, entre ellos: la esperanza, varianza y distribución en el muestreo.

#### Propiedades de $\hat{\mu}$

a)  $\hat{\mu}$  es un estimador centrado de  $\mu$ , en que

$$E[\hat{\mu}] = E[\bar{y}_{..}] = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^J E[y_{ij}] = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^J (\mu + \tau_i + \beta_j) =$$

$$\frac{1}{N} \left( N\mu + J \sum_{i=1}^I \tau_i + I \sum_{j=1}^J \beta_j \right) = \frac{1}{N} N\mu = \mu$$

b) La varianza de  $\hat{\mu}$  es  $\sigma^2/N$ , puesto que al ser independientes las observaciones se verifica:

$$Var[\hat{\mu}] = Var \left[ \sum_{i=1}^I \sum_{j=1}^J \frac{y_{ij}}{N} \right] = \sum_{i=1}^I \sum_{j=1}^J Var \left[ \frac{y_{ij}}{N} \right] = \sum_{ij} \frac{Var[y_{ij}]}{N^2} =$$

$$\sum_{i,j} \frac{\sigma^2}{N^2} = \frac{1}{N^2} N\sigma^2 = \frac{\sigma^2}{N}$$

c)  $\hat{\mu}$  se distribuye bajo una Normal, ya que dicho estimador es combinación lineal de las variables  $y_{ij}$  y éstas son variables aleatorias independientes con distribución Normal.

#### Propiedades de $\hat{\tau}_i$

a)  $\hat{\tau}_i$ , es un estimador centrado de  $\tau_i$ , en que

$$E[\hat{\tau}_i] = E[\bar{y}_{i.}] - E[\bar{y}_{..}] = \mu + \tau_i - \mu = \tau_i$$

En efecto

$$E[\bar{y}_{i.}] = E \left[ \frac{1}{J} \sum_{j=1}^J y_{ij} \right] = \frac{1}{J} E \left[ \sum_{j=1}^J (\mu + \tau_i + \beta_j + \mu_{ij}) \right] =$$

$$\frac{1}{J} \left[ J\mu + J\tau_i + \sum_{j=1}^J E[\mu_{ij}] \right] = \mu + \tau_i$$

b) La varianza de  $\hat{\tau}_i$  es  $(I - 1) \frac{\sigma^2}{N}$ , puesto que

$$\begin{aligned} \text{Var}[\hat{\tau}_i] &= \text{Var}[\bar{y}_i - \bar{y}_..] = \text{Var} \left[ \frac{1}{J} \sum_j y_{ij} - \frac{1}{N} \sum_{i,j} y_{ij} \right] = \\ &= \frac{1}{J^2} \sum_j \text{Var}[y_{ij}] + \frac{1}{N^2} \sum_{i,j} \text{Var}[y_{ij}] - \frac{2}{NJ} \text{Cov} \left[ \sum_j y_{ij}, \sum_{i,j} y_{ij} \right] = \\ &= \frac{1}{J^2} \sum_j \sigma^2 + \frac{1}{N^2} \sum_{i,j} \sigma^2 - \frac{2}{NJ} J \sigma^2 = \\ &= \frac{1}{J} \sigma^2 + \frac{1}{N} \sigma^2 - \frac{2\sigma^2}{N} = \frac{\sigma^2}{J} - \frac{\sigma^2}{N} = (I - 1) \frac{\sigma^2}{N} \quad (30) \end{aligned}$$

Para el modelo equilibrado se deduce con facilidad que la varianza de  $\hat{\tau}_i$  es  $(I - 1)\sigma^2/N$

c)  $\hat{\tau}_i$  tiene una distribución normal, ya que dicho estimador está expresado como función lineal de variables aleatorias con distribución normal

### Propiedades de $\hat{\beta}_j$

a)  $\hat{\beta}_j$  es un estimador centrado de  $\beta_j$ , puesto que

$$\begin{aligned} E[\hat{\beta}_j] &= E \left[ \frac{1}{I} \sum_i y_{ij} \right] = \frac{1}{I} E \left[ \sum_i (\mu + \tau_i + \beta_j + \mu_{ij}) \right] = \\ &= \frac{1}{I} \left[ I\mu + \sum_i \tau_i + I\beta_j + \sum_i E[\mu_{ij}] \right] = \mu + \beta_j \end{aligned}$$

b) La varianza de  $\hat{\beta}_j$  es  $(J - 1) \frac{\sigma^2}{N}$

c)  $\hat{\beta}_j$  se distribuye según una normal

### Propiedades de $\hat{\sigma}^2$

$\hat{\sigma}^2$  no es un estimador insesgado de  $\sigma^2$ . Para demostrarlo veamos en primer lugar que  $\frac{N\hat{\sigma}^2}{\sigma^2}$  se distribuye según una  $\chi^2$  con  $N - 1$  grados de libertad

$$\frac{N\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^J (e_{ij})^2 \rightsquigarrow \chi_{(I-1)(J-1)}^2$$

Donde los grados de libertad de la distribución  $\chi^2$  corresponden al número de residuo independientes, por tanto

$$E\left[\frac{N\hat{\sigma}^2}{\sigma^2}\right] = (I-1)(J-1) \Rightarrow E[\hat{\sigma}^2] = \frac{(I-1)(J-1)}{N} \sigma^2$$

Como  $\hat{\sigma}^2$  no es un estimador insesgado de  $\sigma^2$  con el resultado anterior se construye un estimador centrado considerando

$$\tilde{\sigma}^2 = \frac{N}{(I-1)(J-1)} \sigma^2$$

Este estimador recibe el nombre de varianza residual, se denota por  $\hat{S}_R^2$  y se expresa de la siguiente forma

$$\tilde{\sigma}^2 = \hat{S}_R^2 = \frac{\sum_{i=1}^I \sum_{j=1}^J [y_{ij} - \hat{y}_{ij}]^2}{(I-1)(J-1)} = \frac{\sum_{i=1}^I \sum_{j=1}^J e_{ij}^2}{(I-1)(J-1)} \quad (31)$$

En resumen,

$$\hat{\mu} \rightsquigarrow N\left(\mu, \sigma^2/N\right);$$

$$\hat{\tau}_i \rightsquigarrow N\left(\tau_i, (I-1)\sigma^2/N\right);$$

$$\hat{\beta}_j \rightsquigarrow N\left(\beta_j, (J-1)\sigma^2/N\right);$$

$$N\hat{\sigma}^2/\sigma^2 \rightsquigarrow \chi_{(I-1)(J-1)}^2$$

### 2.1.20. Comparaciones de pruebas o rangos múltiples

Gutierrez y De la Vara (2008) mencionan que, al rechazar la hipótesis nula en el análisis de varianza, se requiere conocer qué tratamientos son diferentes. Cuando no se rechaza la hipótesis nula  $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ , automáticamente la conclusión es que los tratamientos no son diferentes. Bajo el otro escenario donde se rechaza  $H_0$ , y aceptamos la hipótesis alterna  $H_1: \mu_i \neq \mu_j$  para  $i \neq j$ , se genera la necesidad de indagar que tratamientos son diferentes, o que tratamientos provocan la diferencia, estas interrogantes se responden probando la igualdad de todos los posibles pares de medias, para lo que existen diversos métodos, conocidos como **métodos de comparaciones múltiples o pruebas de rango múltiple**. Los métodos difieren entre sí en la potencia que tienen para detectar las diferencias entre las medias. Cuando un método es capaz de detectar diferencias más pequeñas se considera como una prueba más potente.

Existen varias pruebas como alternativa para desarrollar la comparación de medias luego de rechazar la hipótesis nula, destacando entre ellas la prueba de Tukey, Duncan, Dunnet y Fisher. Al respecto Di Rienzo et al. (2008) señala que “la prueba Tukey, está basado en el estadístico de Tukey calculando como valor crítico la identificación de las diferencias significativas, a través de una cantidad de la diferencia mínima significativa (DMS) basada en el cuantil correspondiente de la distribución de rangos estudentizados. Si los tamaños de muestra son iguales, la prueba controla la tasa de error por experimento, bajo hipótesis nulas completas o parciales” (p. 107).

Según Gutierrez y De la Vara (2008), la prueba de Tukey examina con un mismo estadístico todas las diferencias de medias muestrales en estudio con el valor crítico dado por:

$$T_{\alpha} = q_{\alpha}(k, N - k) \sqrt{CM_E/n_i} \quad (32)$$

Donde  $CM_E$  es el cuadrado medio del error,  $n_i$  es el número de observaciones por tratamiento,  $k$  es el número de tratamientos,  $N - k$  es igual a los grados de libertad para el error,  $\alpha$  es el nivel de significancia prefijado y el estadístico  $q_{\alpha}(k, N - k)$  son puntos porcentuales de la distribución del rango estudentizado. Se declaran significativamente diferentes los pares de medias cuya diferencia muestral en valor absoluto sea mayor que  $T_{\alpha}$ . A diferencia de los métodos Least significant difference de Fisher (LSD) y Duncan, el método de Tukey trabaja con un error muy cercano al declarado por el experimentador (p.74-76).

### 2.1.21. Trigo

Etimológicamente la denominación "trigo" es proveniente del vocablo latino *Triticum*, que significa triturado o trillado, y hace mención a la acción de separar el grano de trigo de la capa que lo recubre. *Triticum* significa "el grano" que necesita trillar para su uso en el consumo. Por consiguiente, el trigo (*Triticum*) es una palabra ancestral para denominar a los cereales con referencia a su trituración o trillado.

Algunos conceptos de trigo, se tiene: i) Botánicamente especies, con espigas terminales compuestas de cuatro o más hileras de granos que al ser triturados se obtiene harina para la panificación. Se encuentran varias especies y muchas variedades, ii) Trigo es un término utilizado para denominar a los cereales de grano pequeño cultivados o silvestres del género *Triticum*, fenotípicamente son plantas anuales de la familia Poaceas, cultivadas ampliamente en el mundo.

### **2.1.22. El cultivo de trigo y su origen**

El trigo es uno de los principales alimentos de la población mundial y se consume de diversas formas, según señala Moya et al. (2007) en los trigos comerciales se diferencia en trigo harinero y trigo duro. El trigo harinero es cultivado en mayor proporción en el mundo en comparación con el trigo duro, considerando que es de mayor uso la fabricación de harina, la base de todos los panes, galletas y pasteles gracias a la elasticidad del gluten que contiene y permite la panificación como un alimento básico. Por otro lado, el trigo se considera como una fuente comercial de almidón utilizado en diversidad de productos procesados en la industria, desde el procesamiento de alimentos hasta la fabricación de papel y otros derivados.

Según De Decandolle y Vavilov citado por Moya et al. (2007) señala que la especie *Triticum* es originario de Mesopotamia, y el segundo autor señala que este género tuvo su centro de variación en La India, Turquía y Afganistán. También existen autores que señalan que es originario de Asia Menor y Afganistán con estos estudios se puede señalar que en alguno de estos lugares de climas similares el hombre primitivo encontró por primera vez el trigo silvestre.

### **2.1.23. Clasificación del trigo**

En el estudio desarrollado por Moya et al. (2007) señala que el trigo pertenece a la familia de los gramíneas, en la tribu *Triticeae*, subtribu *Triticinae* y dentro de esta última el género *Triticum* que comprende más de 30 diferencias genéticas para encontrar especies y subespecies distintas entre ellas 16 especies cultivadas el resto silvestres y de las cultivadas solo dos son los tipos comerciales que se producen en el mundo, estos son el trigo harinero (*Triticum aestivum* L.) y el trigo cristalino o duro (*Triticum durum*).

Por otro lado, Prescott y Saari (2012) señala igualmente que las especies más importantes de trigo son el trigo harinero y el trigo duro; sin embargo llegan al detalle de la subespecie, para el caso de trigo harinero (*Triticum aestivum ssp. aestivum*) y para el trigo duro (*Triticum turgidum ssp. Durum*). También señala que los estudios citológicos demuestran que el trigo duro es tetraploide (28 cromosomas) y el trigo harinero es hexaploide (42 cromosomas) (p. 9).

#### **2.1.24. El rendimiento del trigo y sus componentes**

Castañeda-Saucedo et al. (2009 p.291), señalan que en mejoramiento genético el rendimiento y sus componentes son caracteres más estudiados para desarrollar nuevas variedades entre ellas el número de espiga/m<sup>2</sup>, el número de granos por espiga, peso de mil granos entre otros y la expresión de estos caracteres está influenciado por el medio ambiente particularmente la temperatura que influye en el crecimiento, desarrollo y rendimiento.

Con respecto al rendimiento de trigo y los determinantes fisiológicos, Eslafer y Calderini (2003) citado por Kohli et al. (2003), indican que el análisis de los componentes del rendimiento están en base al número de granos por espiga, número de espigas por planta, el número de espiguillas por espiga, esta aproximación de los componentes de rendimiento es el más adecuado para hacer un análisis del cultivo (p. 16).

Moya et al. (2007) desarrollaron un estudio para analizar los efectos de la roya lineal en la fenología, el rendimiento y sus componentes en trigo harinero en dos fechas de siembra empleando el diseño experimental de alfa latice, y observó que la roya disminuyó el ciclo del cultivo y el rendimiento en grano en 41.7 y 43.3% en la primera y segunda fecha de siembra y de los componentes que generan el número

de granos por espiga fue afectado en mayor medida porque la mayor severidad de la roya ocurrió después del espigamiento (p. 7).

Las condiciones del cultivo en el inicio de su desarrollo son importantes para proyectar su potencial de rendimiento y actualmente se puede desagregar el rendimiento en sus componentes e investigar cómo la influencia de las condiciones manejo y el uso de variedades afectan los rendimientos porque son dependientes del ambiente para la producción de materia seca hasta la floración y las características genéticas de la variedad determinan ese porcentaje de la biomasa que se transfiere a los granos Mundstock citado por (Santos, Marza, Butron, y Quispe Félix, 2015 P. 59-69).

#### **2.1.25. El rendimiento de trigo y el número de granos**

En la diversidad de los cultivares de trigo que existen en el mercado según estudios desarrollados por Abbate et al citado por Santos et al. (2015) señalan que: “Ciertas variedades llevan como estrategia de incremento del rendimiento al mayor peso por grano, en lugar de un número de granos/m<sup>2</sup> demostrando que el rendimiento tiene mejor asociación con el número de granos ( $r=0.69$ ,  $gl=23$ ) que con su peso de granos( $r=-0.13$ ). También señala la existencia de diferencias claras entre variedades en el número de granos. El autor hace referencia a la variedad Oasis que produjo un alto número de granos por espiga, mientras que las variedades Puntal y Federal un gran número de espigas/m<sup>2</sup>. Sin embargo, la asociación del número de granos/m<sup>2</sup> y estos componentes entre experimentos no tuvo consistencia (-0.25 y 0.73), manteniendo siempre el número de espigas/m<sup>2</sup> una relación negativa con el número de granos por espiga entre -0.42 y -0.85 (p. 73).

### **2.1.26. El mejoramiento por rendimiento y otros caracteres de interés**

Con respecto al mejoramiento por rendimiento Braun y Payne citado por Reynolds M.P. y Pask A.J.D., (2013), manifiestan que el mejoramiento por rendimiento depende de la introducción de caracteres ampliamente adaptados a diferentes ambientes y por esta razón el fitomejoramiento en mega-ambientes es reconocido por los programas internacionales como el más eficiente para zonas específicas (p. 18).

Respecto al mejoramiento genético de plantas, Camarena, Chura, y Blas (2014), manifiestan que la investigación agrícola más allá de producir más alimentos, debe satisfacer los cambios de la demanda, conservando la diversidad genética y cuidando el medio ambiente. Por consiguiente, el mejoramiento genético puede contribuir en hacer más sostenible los sistemas de producción agraria, utilizando variedades con adaptación a condiciones ambientales diversas y las características que el mercado demanda (p.11).

### **2.1.27. Variabilidad y diversidad genética**

Al desarrollar trabajos de mejoramiento genético la diversidad genética está basada en la variabilidad heredable que ocurre en una población o entre poblaciones en condiciones naturales, por otro lado la variabilidad genética en los programas de mejoramiento es posible disponer a partir de la diversidad genética y la variación de la estructura genética y generar poblaciones con alta variabilidad a partir de los cruzamientos u otros métodos que conllevan a la concentración de genes favorables para una nueva variedad que presentarán diferencias en base a sus estructuras genéticas específicas generando la heterogeneidad en el comportamiento de los

genotipos de las poblaciones segregantes hasta lograr la estabilidad de un genotipo para la selección de un nuevo cultivar (Rimieri, 2017).

## **2.2. Marco conceptual**

### **2.2.1. Experimento**

A la modificación de las características de operación de un proceso o sistema se conoce como experimento, el cual se desarrolla con el objetivo de determinar el efecto sobre las propiedades del producto o resultado. También, el experimento ayuda a incrementar el conocimiento acerca del proceso o sistema, (Gutiérrez y De la Vara, 2008 p. 7).

### **2.2.2. Componentes de rendimiento**

El rendimiento en cualquier especie cultivada se analiza por un modelo simple que permite descomponer al mismo en partes más simples a través de sucesivas aproximaciones: rendimiento igual a número de granos/m<sup>2</sup>, peso medio de los granos igual a número de granos/espiga, número de espigas/m<sup>2</sup> igual a número de espigas/planta, así, se interpreta al rendimiento como el producto de estos componentes, por lo sencillo de la determinación de sus componentes o partes esta aproximación es utilizada ampliamente en diversos estudios de fisiología del rendimiento de cultivos de granos (Kohli et al., 2003 p. 31).

### **2.2.3. Peso de mil granos**

Existen factores que influyen en el peso de las semillas como la variedad, las condiciones particulares de cada campaña, la calidad de los lotes de producción, la presencia de enfermedades o el ataque de insectos de almacenamiento, el peso de

mil semillas puede presentar variaciones más o menos importantes y es un dato muy importante para determinar la densidad de siembra y la población de plantas por unidad de área.

#### **2.2.4. Madurez Fisiológica**

Se refiere a la etapa del desarrollo del cultivo donde se ha producido el máximo crecimiento y maduración. Generalmente está asociada con la completa madurez denominado de cosecha. Son tres fases del desarrollo de una planta (crecimiento, madurez y envejecimiento) y por lo general se dificulta distinguir las claramente porque los momentos de transición entre cada una de las etapas son muy lentas y poco diferenciadas (FAO, 1987 p. 41).

#### **2.2.5. Peso unitario de grano**

Se define el peso unitario del grano como la concentración de asimilados que fueron almacenados durante el periodo de antesis y la maduración del cultivo y son dependientes de la variedad y las condiciones ambientales durante el crecimiento (Austin y Col., 1980, citado por González, 2001 p.13).

#### **2.2.6. Rendimiento**

Según FAO (2014) el rendimiento, Es la cantidad de producto obtenido de un cultivo determinado en una unidad de área motivo del análisis, bajo las características de las prácticas de producción aplicadas, se expresa el rendimiento del cultivo en toneladas por hectárea.

### **2.2.7. Rendimiento potencial**

El rendimiento potencial es considerado como el producto entre el número de granos potencial por unidad de superficie y el peso de mil granos como componentes principales, durante el crecimiento del trigo identificaron que el período de crecimiento de las espigas previo al inicio del llenado del grano, es el más crítico para la determinación del número de granos potencial y el rendimiento potencial en el cultivo de trigo (Abbate, 2017 p. 9).

### **2.3. Antecedentes empíricos de la investigación**

Romani y Herrera (2017) desarrollaron un ensayo comparativo de rendimiento de trigo bajo riego en la campaña 2016 – 2017 en el INTA de Argentina con el objetivo de conducir los ciclos del cultivo y periodos de instalación buscando que la fase de antesis o floración ocurra lo más tempranamente posible para evitar las pérdidas de rendimientos, para lo cual utilizaron el método de investigación explicativa para evaluar 46 cultivares de trigo de cinco ciclos utilizando el diseño en bloques completamente aleatorizados bajo riego en dos fechas de siembra. Del trabajo los autores concluyeron que bajo condiciones del ámbito de estudio se recomienda manejar los ciclos y fechas de siembra para que la fase de antesis o floración ocurra lo más tempranamente posible; sin embargo, esta decisión podría incrementar el riesgo de pérdidas de rendimiento en años con heladas tardías, también recomiendan disponer de información local sobre el comportamiento del cultivar a sembrar.

Silva, Michma, Marza y Butrón (2016), realizaron un estudio durante la campaña agrícola 2013-2014 en Bolivia, con el objetivo de determinar el comportamiento productivo de diez variedades de trigo en condiciones del altiplano

central de Bolivia. La metodología que utilizaron fue la distribución de los tratamientos en un diseño de bloques completos al azar con tres repeticiones para evaluar las variables: altura de planta, número de espigas por metro cuadrado, longitud de espiga, número de espiguillas por espiga, número de granos por espiga, peso de mil granos, peso hectolítrico y rendimiento en grano. Del trabajo de investigación, los autores concluyeron que estadísticamente las variedades más rendidoras fueron: Anzaldo, Chullchuncaneño y Tiraque. La variable rendimiento presentó una correlación altamente significativa con número de espigas por metro cuadrado (0,542), con el número de espiguillas por espiga (0,444) y con el número de granos por espiga (0,444). Seleccionaron dos componentes principales, el primero que contribuyó con más del 30.5 % de la variabilidad explicada por las variables rendimiento en grano, número de espigas por metro, peso hectolítrico, reacción al desgrane y altura de planta como las características más importantes ya que estas fueron las que más aportaron en forma positiva y de forma negativa la variable tipo de grano. Como segundo componente que contribuyó con el 24.3 % de la varianza total fueron las variables: peso de mil granos, tipo de grano, número de granos por espiga, número de espiguillas por espiga y longitud de espiga que aportaron de forma positiva, mientras que reacción al desgrane de forma negativa.

Rodríguez-González et al. (2014), desarrollan un estudio de estabilidad de rendimiento de las variedades comerciales de trigo: Yécora F-70 (Y), Cachanilla F-2000 (C) y Triguenio F-2000 (TR) en el Valle de Mexicali, en México en los periodos 2010 - 2012 con el objetivo de comprobar la estabilidad de rendimiento y de otros componentes agronómicos que no han sido estudiados durante la liberación de las variedades en cuatro ambientes, desarrollaron ensayos bajo un diseño experimental de bloques completos al azar con cuatro repeticiones. Todos los genotipos fueron

clasificados como estables para rendimiento en grano, peso hectolítrico y proteína en grano, en el trabajo se evaluaron siete variables orientadas al rendimiento y la calidad de grano en harina para cada uno de los cuatro ambientes determinados en función a la fertilidad residual del suelo y la fertilización comercial en dos años consecutivos 2007 y 2008. Los autores concluyen que la interacción Genotipo x Ambiente fue altamente significativa entre las variables rendimiento de grano y sus componentes agronómicos evaluadas en cuatro ambientes, los genotipos (C), (Y) y (TR) se comportaron en forma estable para rendimiento de grano y peso hectolítrico.

Enrique et al. (2012), desarrollaron el trabajo de investigación denominado Análisis de 20 genotipos de trigo harinero en el Valle del Yaqui de Sonora en México en dos localidades, con el objetivo de evaluar los componentes el rendimiento en grano y la reacción a roya de 17 líneas de trigo harinero y tres cultivares (testigos). La investigación lo realizaron con riego en dos localidades del Valle del Yaqui en la campaña agrícola otoño-invierno de 2008, utilizaron un diseño experimental de bloques completos al azar con cuatro repeticiones por localidad, la unidad experimental fue de dos parcelas instaladas a doble hilera, de 0.76 m de ancho, y 3 m de largo. En las evaluaciones de la investigación registraron variables de rendimiento, variables de fenología del cultivo, reacción a roya. Para el análisis de los datos utilizaron el ANOVA por localidad y combinado para las dos localidades, comparación de medias de tratamientos con la prueba de Tukey ( $\alpha=0.01$ ) y al análisis de componentes principales. Los autores concluyen que los genotipos con mayor rendimiento fueron RSMF7CJ-62 (8.32 t/ha en la localidad 1), y BABAX/LR42 (7.53 t/ha en la localidad 2). También mencionan que RSMF7CJ-92 (7.78 t/ha), BABAX / LR42 (7.75 t/ha) y RSMF7CJ -78 (7.74 t/ha) fueron los genotipos más sobresalientes en rendimiento en ambas localidades. Respecto al comportamiento agronómico, los

genotipos 01W20728 y Blanca Fuerte fueron pequeñas, 05W90710, RSMF7CJ-90 y RSMF7CJ-92 fueron semi-pequeñas y los 13 restantes y los tres testigos fueron altos, destacando que existe variabilidad fenotípica entre genotipos de utilidad para incorporar en un programa de mejoramiento genético.

Huanca, Butron, Marza y Quispe (2015), realizaron el estudio sobre la estimación de la variabilidad fenotípica de trigo harinero para componentes de productividad, en la Estación Experimental San Benito – Bolivia, con el objetivo de determinar la variabilidad fenotípica de 208 líneas avanzadas de trigo harinero introducido del CIMMYT por el Instituto Nacional de Investigación Agropecuaria y Forestal (INIAF) correspondiente al vivero 32 SAWSN, durante la campaña 2014-2015, el ensayo lo establecieron en surcos de 4 metros, con una densidad de siembra de 100 kg/ha. Se registraron 10 variables cuantitativas y cualitativas (altura de la planta, longitud de espiga, densidad de la espiga, tipo y tamaño del grano, número de granos por espiga, número de espigas por metro cuadrado, peso de mil granos, peso hectolítrico y rendimiento en grano). Del trabajo concluyen los autores que se registró correlaciones altas entre el rendimiento con la longitud de espiga, número de granos por espiga, altura de la planta y peso de mil granos. Los genotipos L53, L102, L112, L131, L129, L51, L143 y L203 destacan para características de productividad. Los genotipos que destacan para características de grano y calidad fueron las líneas L205, L98, L108, L84, L106, L204 y L1 a través de la técnica de agrupación jerárquica.

Santos et al. (2015) desarrollaron el estudio de Correlación para variables agronómicas y componentes de rendimiento de trigo harinero en 304 genotipos con el objetivo de saber la existencia de la asociación de caracteres agronómicos e identificar componentes de rendimiento considerando que en mejoramiento genético es necesario identificar la relación entre variables y componentes de rendimiento. El

estudio fue desarrollado en la Estación Experimental de San Benito – Cochabamba – Bolivia entre el 2014-2015, según la metodología utilizada por los autores, el análisis de sendero permitió explicar las posibles causas de las correlaciones que se presentaron entre la variable respuesta (rendimiento) y las variables predictoras seleccionadas en el estudio llegando a concluir que el rendimiento está altamente relacionado con altura de planta ( $r=0.846$ ), con el número de espigas por metro cuadrado ( $r=0.805$ ), con el número de granos por espiga ( $r=0.676$ ), con la longitud de espiga ( $r=0.581$ ) y el peso de mil granos ( $r=0.569$ ), precisan también los autores que la correlación altamente significativa corresponde a longitud de espiga y número de granos por espiga ( $r=0,764$ ) versus rendimiento, las cuales deben ser consideradas en los procesos de selección y mejoramiento.

Chavez et al.(2015) con el objetivo de observar el comportamiento agronómico de veinte variedades experimentales de maíz (*Zea mays*) y seleccionar genotipos con resistencia a plagas y rendimiento para utilizarlo como base en el cruzamiento genético, ejecutaron el experimento en 55 campos situados en las zonas occidental y oriental de Cuba en un diseño de bloques parcialmente aleatorizado observando los valores promedios de características agronómicas y morfológicas entre ellas la longitud y el diámetro de la mazorca, número de hileras de granos, número de granos por hilera, longitud de grano, peso de 100 granos utilizaron el análisis de clúster calculando la matriz de distancia entre individuos a partir del coeficiente de similitud en la selección de genotipos de maíz para establecer el grupo de individuos que presenten características similares, entre sus conclusiones seleccionan a tres variedades recomendadas para estudios genéticos, indican que los caracteres más significativos en este tipo de estudios son la longitud de la mazorca, del grano y el

peso de 100 granos y recomiendan utilizar la distancia de Gower y el método de Ward para agrupamiento jerárquico a partir de caracteres mixtos.

Limaylla (2016) en condiciones de la sierra de Huánuco ejecutó el ensayo de identificación, adaptación y eficiencia de diez genotipos de trigo harinero para determinar la estabilidad del rendimiento de diez genotipos de trigo (*Triticum aestivum* L) en cinco localidades, utilizó el diseño de bloques completos al azar con cuatro repeticiones en cinco localidades de la Región Huánuco llegando a la conclusión que los mejores tratamientos fueron ALTAR 84 / AEGILOPS SQUARROSA (TAUS), CHEN / AE. Q //2\* OPATA /3/ BABAX /4/ JARU y WORRAKATTA / 2\* PASTOR, que presentaron rendimientos cercanos a los 3000 kg/ha y combinaciones favorables de sus parámetros de estabilidad. La variedad utilizada como testigo fue el trigo Centenario que demostró estabilidad y consistencia simultánea con un rendimiento de 2,835 kg/ha, mostrando buen indicador de amplia adaptabilidad a las cinco localidades del estudio

## **CAPÍTULO III**

### **HIPÓTESIS Y VARIABLES**

#### **3.1. Hipótesis**

##### **a. Hipótesis general**

La evaluación y análisis de la influencia de características agronómicas mediante el diseño de bloques completos al azar son eficientes para seleccionar genotipos de trigo harinero de alto rendimiento en el distrito de Zurite – 2020.

##### **b. Hipótesis específicas**

- b.1. La altura de planta tiene una influencia significativa en el rendimiento de trigo harinero.
- b.2. El número de granos por espigas tiene una influencia significativa en el rendimiento de trigo harinero.
- b.3. El tamaño de espiga tiene una influencia significativa en el rendimiento de trigo harinero.
- b.4. El peso de mil granos tiene una influencia significativa en el rendimiento de trigo harinero.
- b.5. La madurez fisiológica tiene una influencia significativa en el rendimiento de trigo harinero.

#### **3.2. Identificación de variables e indicadores**

En la Tabla 4, se mencionan las variables a ser estudiadas en el presente trabajo con sus respectivos indicadores, mientras que en la Tabla 5 se indica la debida operacionalización de las variables del presente estudio.

**Tabla 4.** Variables e indicadores

Variables	Indicadores
Líneas de trigo harinero	Altura de la planta (cm). Número de granos por espiga. Tamaño de la espiga (cm). Peso de mil granos (gramos). Días a la madurez fisiológica (días).
Rendimiento	Peso por unidad de área (t/ha).

### 3.3. Operacionalización de variables

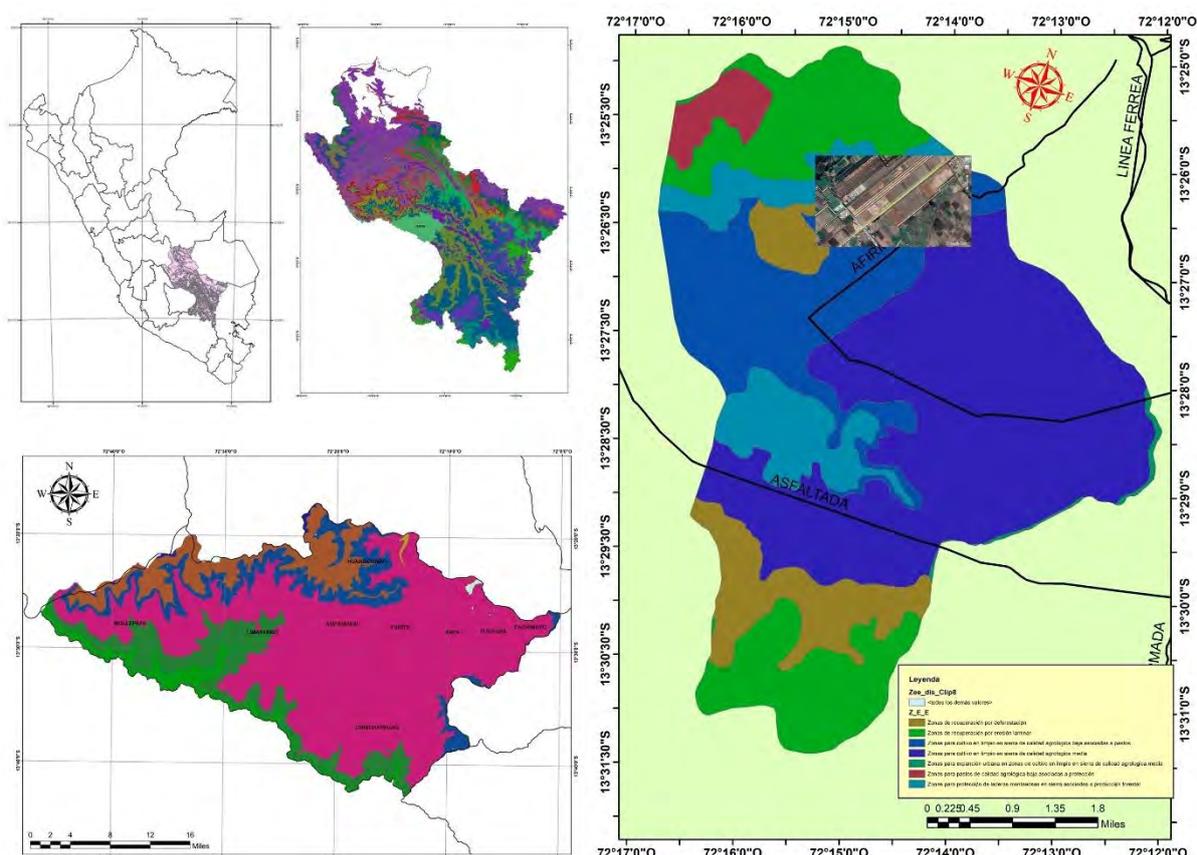
**Tabla 5.** Operacionalización de variables

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	INDICADORES
Líneas de trigo harinero	Líneas o cultivares de trigo Harinero, se define como un grupo de plantas con características distintas, uniformes y estables que presentan su propia identidad que la distingue de las demás. Los descriptores varietales que confieren la identidad de la línea o cultivar son: ciclo vegetativo, características de grano, caracteres morfológicos, reacción a enfermedades y plagas, producción de granos, patrones enzimáticos o de ácidos nucleicos. La estabilidad del cultivar es importante para una identificación de generación en generación (Camarena et al., 2014).	Las líneas de trigo harinero constituyen material genético procedente de cruzamientos para el desarrollo de nuevas variedades comerciales.	Altura de la planta (cm) N° granos por espiga Tamaño de la espiga Peso de mil granos Días a la madurez de cosecha
Rendimiento	<p>El rendimiento de grano de trigo (<i>Triticum spp.</i>) es la cantidad de producto obtenido en una unidad de área motivo del análisis, bajo las características de las prácticas de producción aplicadas y está determinado por los componentes peso y número de granos por unidad de superficie, que presenta correlación con el rendimiento (Abbate, 2017 p.9). El número de granos por unidad de superficie se establece en la fase reproductiva, que comprende el periodo entre las etapas de iniciación floral y antesis.</p> <p>El Rendimiento del cultivo de trigo está determinado por muchos factores genéticos y ambientales que interactúan en mayor o menor grado en función al material genético sean líneas promisorias o variedades motivo por el cual es necesario conocer experimentalmente los rendimientos de líneas promisorias de trigo en proceso de investigación para el desarrollo de variedades (FAO, 2014).</p>	El rendimiento del cultivo de trigo está en función de los componentes principales de rendimiento, el desarrollo del cultivo, las fases fenológicas y las condiciones climatológicas que se registran en el proceso productivo.	Peso por unidad de área (t/ha).

## CAPITULO IV

### METODOLOGÍA

#### 4.1. Ámbito de estudio: localización política y geográfica



**Figura 2.** Mapa de ubicación del distrito de Zurite donde se desarrolló la investigación.

El ámbito de desarrollo de la investigación fue en el distrito de Zurite (Figura 2) ubicado en la provincia de Anta, del departamento de Cusco, a 3405 m.s.n.m. Geológicamente presenta variadas formaciones geológicas, fuertes pendientes y desniveles en el terreno de producción agrícola. Estas características variadas de topografía y fisiografía que se encuentran en su territorio quedan determinados e influenciados por la presencia de climas variados y marcadamente diferentes que dan

lugar a diversas zonas de vida, que generan características favorables para los recursos naturales, tipos de vegetación y tierra para la explotación de la actividad agropecuaria (Gobierno Regional de Cusco, 2016).

#### **4.2. Tipo y nivel de investigación**

El trabajo de investigación fue experimental de tipo aplicada con una variable dependiente porque nuestra variable de interés es el rendimiento, al respecto Vara (2012) señala *“La investigación aplicada normalmente identifica la situación problema y busca, dentro de las posibles soluciones, aquella que pueda ser la más adecuada para el contexto específico”*(p.202).

El nivel de investigación fue explicativo de naturaleza cuantitativa, que según, Fernández y Baptista, (2014) señala que *“los estudios explicativos son más estructurados que los estudios de los demás alcances y se centran en el interés de explicar por qué ocurre un fenómeno y en qué condiciones se manifiesta”* y precisamente con la investigación se pretende explicar cómo influyen los componentes de rendimiento de trigo en grano evaluados en un diseño de bloques completamente al azar en el ámbito del estudio (p.96).

#### **4.3. Diseño experimental y croquis de distribución**

Se utilizó el diseño experimental de bloques completos al azar con cuatro repeticiones cuyas características de ubicación, dimensiones y distribución del campo experimental se detallan en la Figura 3 y Tabla 6.

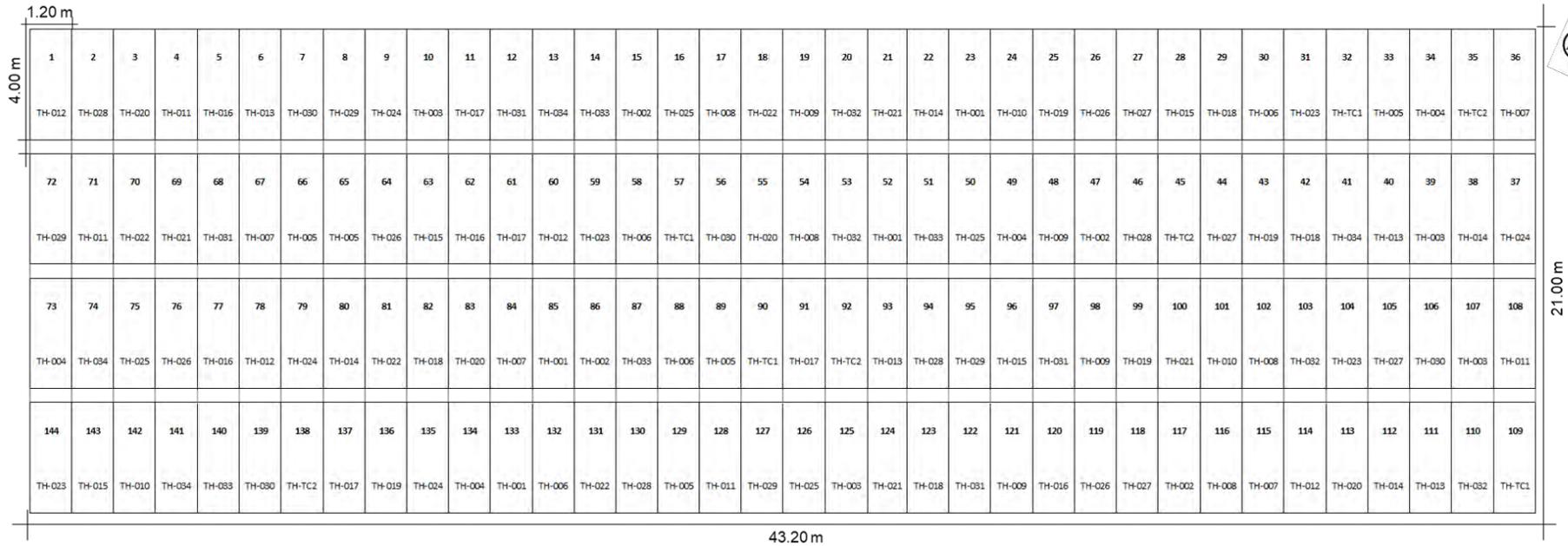


Figura 3. Croquis de ubicación y distribución del diseño experimental.

**Tabla 6.** Características del campo experimental

<b>Características</b>	<b>Dimensiones</b>
Largo de tratamiento(m)	4.00
Ancho de tratamiento (m)	1.20
Área total de tratamiento (m <sup>2</sup> )	4.80
Separación de los surcos (m)	0.30
Número total de surcos por Unidad experimental	4.00
Área neta de evaluación por Unidad experimental (m <sup>2</sup> )	2.40
Longitud de Bloque (m)	43.20
Ancho de Bloque (m)	4.00
Área de Bloque (m <sup>2</sup> )	172.80
Ancho de calle (m)	1.00
Largo de Calle (m)	43.20
Área de Calle (m <sup>2</sup> )	43.20
Largo campo experimental (m)	43.20
Ancho Campo experimental (m)	21.00
Área del campo experimental (m <sup>2</sup> )	907.20

#### 4.4. Tratamientos en estudio

Para el trabajo de investigación se utilizó líneas avanzadas de trigo harinero procedentes del Centro Internacional de Mejoramiento de Maíz y trigo (CIMMYT) y como testigos de comparación dos variedades comerciales de trigo harinero difundidos en el ámbito del departamento de Cusco que se detallan en la Tabla 7.

**Tabla 7.** Relación del material genético utilizado en el trabajo de investigación

<b>Tratamiento</b>	<b>CRUZA, LINEA Y/O VARIEDAD</b>	<b>ORIGEN</b>
TH-001	MUNAL*2/WESTONIA	25SAWYT 303
TH-002	SAUAL/3/SW89.3064//CMH82.17/SERI/4/SAUAL/5/PBW343*2/KUKUNA*2//FRTL/PIFED/6/SAUAL/KRONSTAD F2004	25SAWYT 322
TH-003	FRANCOLIN #1/YANAC/5/KIRITATI/4/2*BAV92//IRENA/KAUZ/3/HUITES	25SAWYT 307
TH-004	UP2338*2/SHAMA/3/MILAN/KAUZ//CHIL/CHUM18/4/UP2338*2/SHAMA*2/5/PBW343*2/KUKUNA*2//FRTL/PIFED	25SAWYT 321
TH-005	WHEAR//2*PRL/2*PASTOR/3/WAXBI/4/COPIO	25SAWYT 309
TH-006	OTUS//WBLL1*2/TUKURU/3/2*PBW343*2/KUKUNA*2//FRTL/PIFED	25SAWYT 310

TH-007	TACUPETO F2001/BRAMBLING//KIRITATI/5/C80.1/3*BATAVIA//2*WBL L1/3/ATTILA/3*BCN*2//BAV92/4/WBLL1*2/KURUKU/6/ROL F07/YANAC//TACUPETO F2001/BRAMBLING	25SAWYT 313
TH-008	ROLF07/YANAC//TACUPETO F2001/BRAMBLING*2/5/UP2338*2/SHAMA/3/MILAN/KAUZ/ /CHIL/CHUM18/4/UP2338*2/SHAMA	25SAWYT 315
TH-009	FRET2*2/SHAMA//PARUS/3/FRET2*2/KUKUNA*2/4/KINGBI RD #1//INQALAB 91*2/TUKURU	25SAWYT 317
TH-010	TRCH/SRTU//KACHU*2/5/UP2338*2/SHAMA/3/MILAN/KAU Z//CHIL/CHUM18/4/UP2338*2/SHAMA	25SAWYT 318
TH-011	TRCH/SRTU//KACHU*2/5/UP2338*2/SHAMA/3/MILAN/KAU Z//CHIL/CHUM18/4/UP2338*2/SHAMA	25SAWYT 319
TH-012	PBW343*2/KUKUNA//SRTU/3/PBW343*2/KHVAKI/4/VORB/ FISCAL//AKURI #1/5/PBW343*2/KUKUNA//SRTU/3/PBW343*2/KHVAKI	25SAWYT 314
TH-013	SAUAL/YANAC//SAUAL*2/3/TACUPETO F2001/BRAMBLING*2//KACHU	25SAWYT 323
TH-014	SAUAL/MUTUS/4/KACHU #1//WBLL1*2/KUKUNA/3/BRBT1*2/KIRITATI	25SAWYT 325
TH-015	BORL14//KFA/2*KACHU	25SAWYT 327
TH-016	TILILA/TUKURU/4/SERI.1B*2/3/KAUZ*2/BOW//KAUZ/5/KFA /2*KACHU	25SAWYT 328
TH-017	WAXWING/2*ROLF07//BORL14	25SAWYT 329
TH-018	BECARD/FRNCLN/3/KACHU #1/KIRITATI//KACHU	25SAWYT 330
TH-019	MUU/KBIRD/3/PRL/2*PASTOR*2//FH6-1-7	25SAWYT 333
TH-020	BECARD//ND643/2*WBLL1/4/KIRITATI//ATTILA*2/PASTOR /3/AKURI	25SAWYT 334
TH-021	ND643//2*ATTILA*2/PASTOR/3/WBLL1*2/KURUKU/4/WBLL 1*2/BRAMBLING/6/BABAX/LR42//BABAX*2/3/KUKUNA/4/C ROSBILL #1/5/BECARD	25SAWYT 335
TH-022	CNO79//PF70354/MUS/3/PASTOR/4/BAV92*2/5/HAR311/6/ PBW343*2/KUKUNA*2//FRTL/PIFED/7/CNO79//PF70354/M US/3/PASTOR/4/BAV92*2/5/HAR311	25SAWYT 337
TH-023	MUNAL #1/FRANCOLIN #1/5/KIRITATI/4/2*BAV92//IRENA/KAUZ/3/HUITES/6/BECA RD/FRNCLN	25SAWYT 339
TH-024	FRANCOLIN #1//WBLL1*2/BRAMBLING*2/3/COPIO	25SAWYT 340
TH-025	FRANCOLIN #1/YANAC*2/3/PBW343*2/KUKUNA*2//FRTL/PIFED	25SAWYT 341
TH-026	NADI/3/PBW343*2/KUKUNA*2//FRTL/PIFED/4/NADI	25SAWYT 342
TH-027	NADI/COPIO//NADI	25SAWYT 343
TH-028	SITE/MO//PASTOR/3/TILHI/4/WAXWING/KIRITATI/5/KACH U #1/KIRITATI//KACHU	25SAWYT 326

TH-029	YAV_3/SCO//JO69/CRA/3/YAV79/4/AE.SQUARROSA (498)/5/LINE 1073/6/KAUZ*2/4/CAR//KAL/BB/3/NAC/5/KAUZ/7/KRONST AD F2004/8/KAUZ/PASTOR//PBW343/9/PBW343*2/KUKUNA*2 //FRTL/PIFED/10/KIRITATI//PRL/2*PASTOR/5/OASIS/SKAU Z//4*BCN/3/PASTOR/4/KAUZ*2/YACO//KAUZ/6/KIRI	25SAWYT 345
TH-030	BABAX/LR42//BABAX/3/ER2000/4/NIGHAR	25SAWYT 346
TH-031	SOKOLL/3/PASTOR//HXL7573/2*BAU/4/PANDION//FILIN/2 *PASTOR/3/BERKUT	25SAWYT 347
TH-032	MEX94.27.1.20/3/SOKOLL//ATTILA/3*BCN/5/GK ARON/AG SECO 7846//2180/4/2*MILAN/KAUZ//PRINIA/3/BAV92	25SAWYT 348
TH-033	WHEAR/SOKOLL/8/BOW/VEE/5/ND/VG9144//KAL/BB/3/YA CO/4/CHIL/6/CASKOR/3/CROC_1/AE.SQUARROSA (224)//OPATA/7/PASTOR//MILAN/KAUZ/3/BAV92	25SAWYT 349
TH-034	SUP152/6/OASIS/5*BORL95/5/CNDO/R143//ENTE/MEXI75/ 3/AE.SQ/4/2*OCI	25SAWYT 350
TH-TC1	INIA 405 SAN ISIDRO	TESTIGO COMERCIAL
TH-TC2	INIA 419 SAN FRANCISCO	TESTIGO COMERCIAL

Durante el desarrollo del trabajo de investigación, el manejo agronómico que se dio a la parcela fue el siguiente:

- La preparación del terreno fue mecanizada utilizando un tractor agrícola con arado de discos y rastra para dejar las condiciones homogéneas del campo experimental y una surcadora experimental.
- La labor de siembra fue manual utilizando una densidad de 120 kg/ha de semilla distribuidos en surcos distanciados a 0,30 m.
- Se aplicó fertilización química en base a los resultados de análisis de suelos del campo experimental utilizando la fórmula de fertilización de 80-80-00 de N, P<sub>2</sub>O<sub>5</sub>, K<sub>2</sub>O. Aplicando a la siembra la fertilización fosforada y el 38% de nitrógeno y al macollamiento el 62% de nitrógeno restante
- El control de malezas se realizó con el herbicida de principio activo (*Metsulfuron metil*) a la dosis de 10 gramos/ha aplicado a los 35 días posteriores a la emergencia.

- La labor de cosecha y poscosecha fue semi mecanizada con el corte manual de los surcos centrales para las evaluaciones de rendimiento y en grano y la trilla utilizando una maquinaria experimental independientemente para cada tratamiento, posteriormente se realizó la limpieza y selección de granos con maquinaria experimental para eliminar los residuos de cosecha.

#### **4.5. Unidad de análisis**

La Unidad de análisis para el estudio, fue la unidad experimental que consiste en la parcela de una línea promisorio de trigo harinero de 4.8 m<sup>2</sup> que manifiesta una respuesta a la condición bajo la que se encuentra en cada bloque, respuesta que es registrada como un valor de la variable en estudio.

#### **4.6. Población de estudio**

La población en estudio, fue el total de unidades experimentales en las que se aplicó los 36 tratamientos (34 líneas avanzadas de trigo harinero procedentes del CIMMYT y como testigos de comparación dos variedades comerciales de trigo harinero difundidos en el ámbito del departamento de Cusco: estos son INIA 405 San Isidro e INIA 418 San Francisco, mencionado en la Tabla 7.

#### **4.7. Tamaño de muestra**

Vara (2012 p. 221) define “*El muestreo es el proceso de extraer una muestra a partir de una población*” para lo cual la unidad experimental fue la parcela por cada línea promisorio y se han considerado 10 plantas o partes de plantas para la evaluación de las variables en estudio a excepción del rendimiento que fue la cosecha del área central de la parcela para evitar el efecto borde.

#### **4.8. Técnicas de selección de muestra**

Arias (2012) indica: “La técnica es el conjunto de instrumentos y medios a través de los cuales se efectúa el método. Si el método es el camino, la técnica proporciona las herramientas, para recorrer ese camino” (p.65).

En el presente estudio se utilizó la técnica de selección del muestreo probabilístico aleatorio simple, al respecto Arias (2012) menciona que el *“Muestreo es un proceso en el que se conoce la probabilidad que tiene cada elemento de integrar la muestra.*

En el presente trabajo de investigación la técnica de selección de muestra para las variables altura de planta, tamaño de espiga y número de granos por espiga fue considerando la elección aleatoria de 10 plantas en competencia completa en los dos surcos centrales de cada unidad experimental para evitar los efectos de borde. Para la variable madurez fisiológica por observación directa de la fase fenológica en la unidad experimental. Para el peso de mil granos por conteo al azar de mil semillas de la producción de los dos surcos centrales de la unidad experimental se registró el peso y con respecto al rendimiento se cosecho el área central de la unidad experimental correspondiente a dos surcos centrales (2.4 m<sup>2</sup>) eliminado los surcos laterales para evitar el efecto de borde que luego de la limpieza fue registrado el peso y el contenido de humedad para calcular el rendimiento ajustado a un promedio de humedad de almacenamiento del 12%.

#### **4.9. Técnicas de recolección de información**

Las evaluaciones de las variables asociadas con el rendimiento se desarrollaron en campo y en laboratorio siguiendo los protocolos de mejoramiento disponibles en el INIA (Estrada et al, 2020), estas evaluaciones fueron las siguientes:

- Altura de planta, las evaluaciones se realizaron en 10 plantas del área central de cada tratamiento efectuando las mediciones con cinta milimétrica desde el cuello de la planta hasta las aristas.
- Tamaño de espiga, se hizo el muestreo al azar de 10 espigas del área central de la parcela donde se realizó las mediciones utilizando un vernier digital de lectura mínima de 0.01 mm.
- Número de granos por espiga, se hizo el muestreo al azar de 10 espigas del área central de la parcela donde se realizó las mediciones utilizando un vernier digital de lectura mínima de 0.01 mm.
- Días a la madurez fisiológica, la evaluación se desarrolló cuando el grano llegó a su máximo contenido de materia seca y el contenido de humedad se registraba entre el 35 a 38%.
- Peso de mil granos, para esta evaluación se realizó el conteo de mil semillas por unidad experimental registrando el peso a través de una balanza analítica de 0.01 gramo de precisión.
- Humedad de grano, se registró la humedad de grano al momento de evaluar el rendimiento utilizando un determinador de humedad digital de lectura mínima de 0.1%.
- Rendimiento, se registró el peso del grano procedente del área central de la unidad experimental utilizando una balanza digital de lectura mínima de 0.1 gramo y capacidad de cinco kilogramos. Posteriormente se realizó el cálculo del rendimiento ajustado al 12% de humedad según establecido por CIMMYT (2012) para determinar el rendimiento real a humedad de almacenamiento de los tratamientos en estudio para ello se utilizaron las siguientes fórmulas:

$$\text{Cantidad de humedad} = \text{Peso del grano total} \times \text{porcentaje de humedad}$$

$$Rdto. seco \left( \frac{kg}{ha} \right) = \frac{Rdto. total del grano - cantidad de humedad}{Area} \times 10$$

$$Rdto. al 12\% Humedad \left( \frac{kg}{ha} \right) = Rendimiento seco / 0.88$$

#### 4.10. Técnicas y herramientas estadísticas de análisis de los datos

Para el presente estudio, se utilizaron diferentes técnicas y herramientas estadísticas para realizar el análisis estadístico de los datos. Entre las herramientas estadísticas fueron utilizados los siguientes softwares: Lenguaje o entorno de programación R y el InfoStat para obtener los respectivos resultados del análisis estadístico de los datos recolectados, mientras que el Minitab se utilizó para evaluar la confiabilidad, validez y objetividad de los instrumentos de medición utilizados.

Las técnicas utilizadas para el análisis estadístico de los datos fueron las siguientes:

- a) **Análisis de conglomerados o clúster**, como técnica multivariante exploratoria con la finalidad de agrupar los tratamientos en estudio considerando su afinidad con respecto a las variables evaluadas y observar la relación existente entre los genotipos de trigo tratando de lograr la máxima homogeneidad dentro del grupo y la mayor diferencia entre los grupos para ello primeramente se desarrolló un análisis descriptivo de los datos con respecto a cada una de las variables, posteriormente al encontrarse las variables en diferentes unidades de medida se procedió a estandarizar y utilizar los datos estandarizados para la clasificación de los conglomerados aplicando la distancia de Mahalanobis según la ecuación (2) al existir correlación significativa entre las variables evaluadas según la ecuación (1) para determinar las distancias de similitud o proximidad entre los genotipos aplicando el conglomerado jerárquico con el método de Ward recomendado por Chavez et al. (2015) que es un procedimiento donde en cada etapa se unen

los dos clúster para los que se tenga el menor incremento en el valor total de la suma de los cuadrados de las diferencias dentro de cada clúster, ya que el número de tratamientos no es muy grande y nos permite analizar diferencias grandes inter conglomerados y diferencias pequeñas intra conglomerados utilizando la gráfica del dendograma.

b) **Análisis de regresión lineal simple** según la ecuación (5) para obtener la variación de la influencia de las variables independientes (altura de planta, número de granos por espiga, tamaño de espiga, peso de mil granos, madurez fisiológica) sobre la variable de rendimiento y la construcción del modelo para la estimación de la media de la variable rendimiento a partir de las variables independientes.

- Coeficiente de determinación según la ecuación (8) para medir la proporción de la variación de la variable dependiente (Rendimiento) en relación a las variables independientes.
- Error Cuadrático Medio de Predicción según la ecuación (9) para medir la cantidad del valor predicho y el valor observado, que para efectos de los resultados serán interpretados a través de la raíz cuadrada del ECMP que viene a ser la desviación estándar.
- Coeficiente de Correlación según la ecuación (11), para evaluar el grado y el sentido de la asociación lineal que exista entre el rendimiento con las variables (altura de planta, número de granos por espiga, tamaño de espiga, peso de mil granos, madurez fisiológica) independientemente con cada una de ellos para determinar y cuantificar las relaciones existentes entre estas variables y a partir de la ecuación (10) determinar el contraste de hipótesis sobre la correlación.

c) **Análisis de Varianza**, para probar la hipótesis de posición de los tratamientos en estudio en cada una de las variables evaluadas previo a la verificación de los supuestos del modelo, se utilizó el análisis de varianza de la ecuación (18) y el esquema para la presentación de los resultados del ANOVA de la Tabla 3. para la verificación de los supuestos se utilizó: la ecuación (20) correspondiente a la prueba de normalidad a través Shapiro-Wilks para determinar la distribución de los residuos, la ecuación (21) correspondiente a la prueba de Bartlett para determinar la homogeneidad de la varianza de los residuos.

En base a los resultados del análisis de varianza al encontrar diferencias estadísticas entre los tratamientos en estudio se analizó la comparación de medias a través de la prueba de Tukey, ecuación (32).

Para los casos en los que no se cumplieron con los supuestos de homogeneidad de varianzas para realizar el ANOVA, entonces se procedió a realizar una descripción y comparación de grupos usando las medias con su respectivo error estándar. También es necesario precisar que no se aplicó la prueba no paramétrica, porque los datos del estudio parten del supuesto de seguir una distribución normal, y según señala Fernández et al. (2017) las pruebas no paramétricas se aplican cuando no se conoce la naturaleza de la distribución en la que se fundamentan los datos; cuando se utilizan categorías y cuando se trata de datos no numéricos y permiten medir las puntuaciones medias de los datos durante tres o más momentos. Por lo general, la eficiencia de los métodos no paramétricos como Friedman es alta para muestras pequeñas,  $n \leq 10$ , y disminuye conforme aumenta  $n$ , que no corresponden a la información del presente estudio.

## CAPÍTULO V

### RESULTADOS Y DISCUSIÓN

#### RESULTADOS

Para la obtención de los resultados se utilizaron herramientas estadísticas como el software R versión 4.1.2 y el paquete estadístico InfoStat versión 2018 con la finalidad de dar respuesta a los objetivos del estudio.

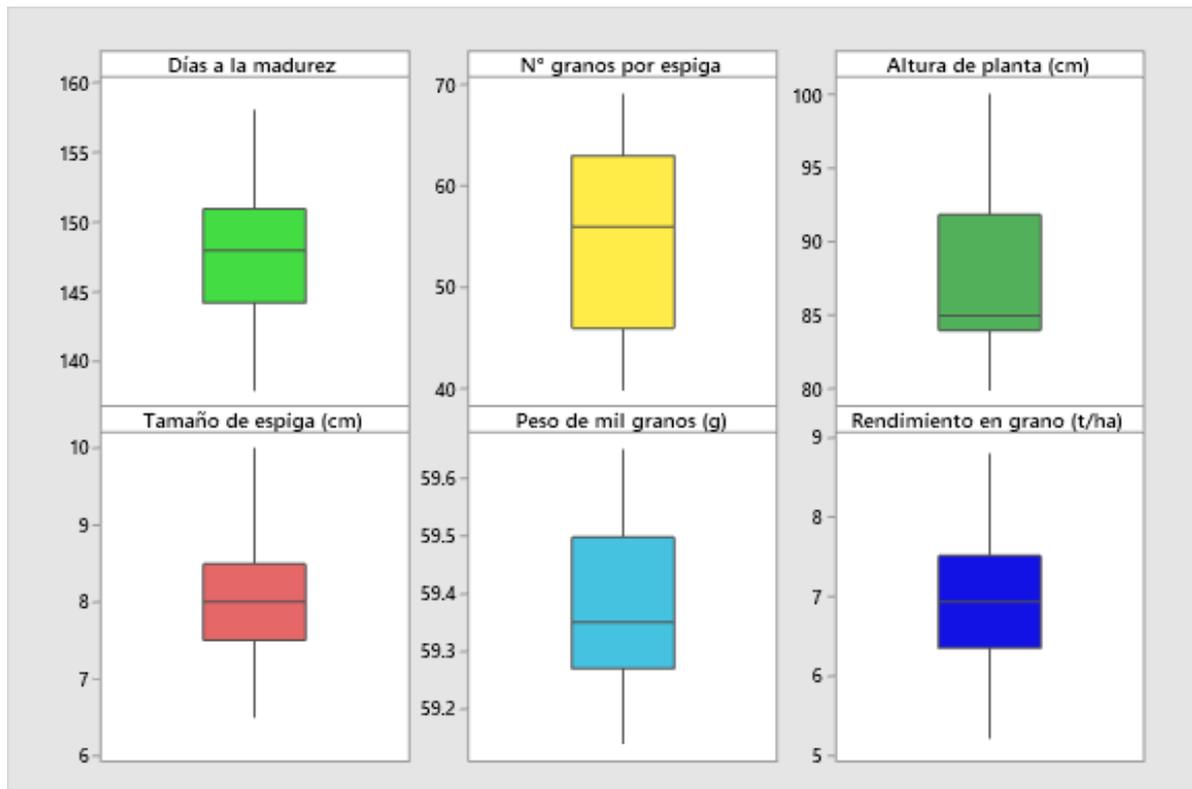
#### 5.1. Análisis exploratorio de los datos

El análisis descriptivo de la Tabla 8 refleja que los días a la madurez de los tratamientos en estudio fluctúa entre 138 a 158 días con una media de 148 días, el número de granos por espiga fluctúa de 40 a 69 y la altura de planta en promedio fue de 87.48 cm. con tamaño de espiga en promedio de 7.34 cm. siendo el peso de mil granos promedio de 59.38 gramos y el rendimiento de los tratamientos se registró entre 5.22 a 8.80 t/ha. En las variables evaluadas no se han registrado casos atípicos que puedan influenciar en los análisis de los datos como se puede observar en la Figura 4.

**Tabla 8.** Estadísticos descriptivos de las variables en estudio

Variable	Media	D.E.	Mín	Máx
Días a madurez	147	4.98	138.00	158.00
Número de granos por espiga	55	8.71	40.00	69.00
Altura planta (cm)	87.48	4.98	80.00	100.00
Tamaño de espiga (cm)	7.34	0.96	6.00	9.80
Peso de mil granos (g)	59.38	0.14	59.14	59.65
Rendimiento en grano (t/ha)	6.94	0.81	5.22	8.80

Donde: D.E.= Desviación estándar, Mín.= Valor mínimo de los datos, Máx.=Valor máximo de los datos



**Figura 4.** Diagrama de cajas para las variables evaluadas

**Tabla 9.** Matriz de correlaciones de Pearson para las variables en estudio

	Días a madurez	N° granos por espiga	Altura de la planta (cm)	Tamaño de espiga (cm)	Peso de mil granos (g)	Rendimiento en grano (t/ha)
Días a madurez	1					
N° granos por espiga	0.8925**	1				
Altura de la planta (cm)	0.5202**	0.5214**	1			
Tamaño de espiga (cm)	0.7728**	0.7739**	0.6691**	1		
Peso de mil granos (g)	0.9024**	0.9337**	0.5947**	0.8574**	1	
Rendimiento en grano (t/ha)	0.9316**	0.9324**	0.5893**	0.8600**	0.9693**	1

### 5.1.1. Análisis de Conglomerados o clúster

Con la finalidad de analizar el cumplimiento de los supuestos del análisis de clasificación se ha desarrollado el análisis de correlación entre las variables y se observa en la Tabla 9 que existe correlación significativa entre las variables en estudio por consiguiente no se estaría cumpliendo con el primer supuesto para la aplicación del análisis de clúster que señala que debe existir ausencia de correlación entre las variables y para cumplir con este supuesto se ha utilizado la distancia de Mahalanobis para compensar la correlación en lugar de utilizar la distancia euclídea.

Una vez compensado el primer requisito a través de la medida de distancia de Mahalanobis y teniendo un número de variables apropiado para el análisis de clúster en este caso seis variables; al encontrarse diferentes unidades de medida para cada variable se procedió a estandarizar las unidades recategorizando todas las variables.

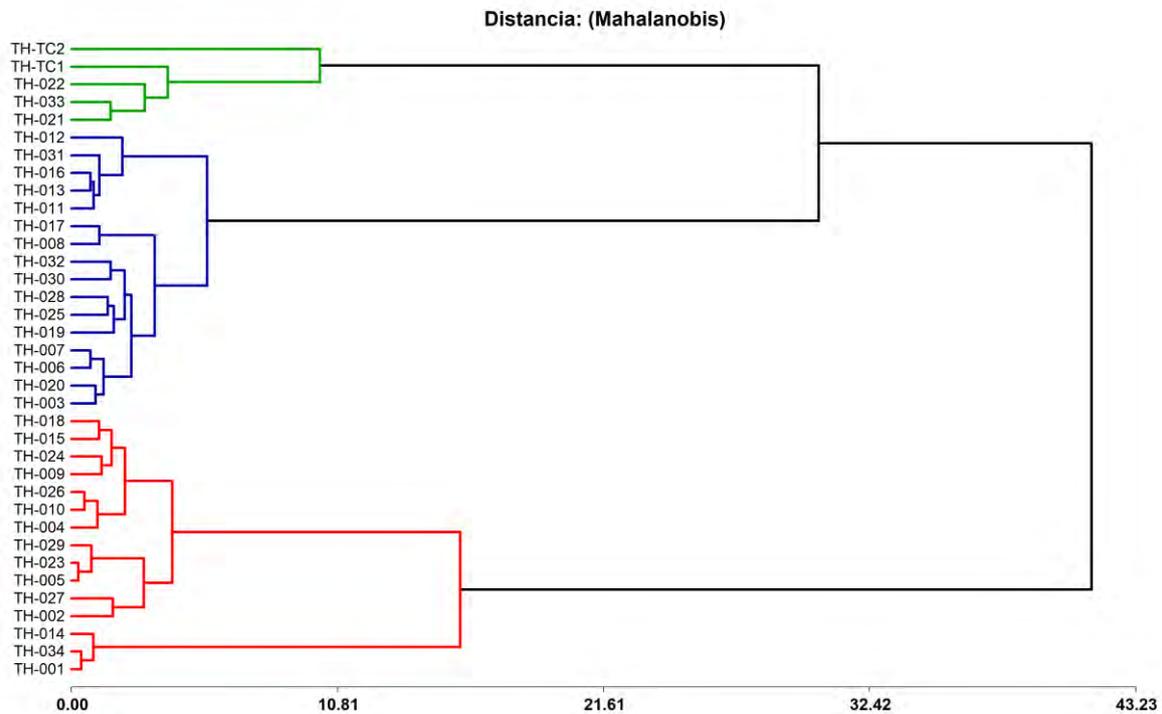
**Tabla 10.** Matriz de proximidades de los tratamientos en estudio a través de la distancia de Mahalanobis

Casos	TH-001	TH-002	TH-003	TH-004	TH-005	TH-006	TH-007	TH-008	TH-009	TH-010	TH-011	TH-012
TH-001	0.00											
TH-002	2.84	0.00										
TH-003	6.46	4.44	0.00									
TH-004	4.16	2.46	3.84	0.00								
TH-005	3.92	1.85	4.36	1.59	0.00							
TH-006	5.78	3.62	0.94	3.00	3.54	0.00						
TH-007	5.88	3.67	1.26	3.34	3.83	0.77	0.00					
TH-008	4.96	2.90	2.34	1.88	2.34	1.64	2.25	0.00				
TH-009	4.28	2.01	3.54	1.46	1.28	2.70	3.03	1.56	0.00			
TH-010	3.87	2.30	4.13	1.00	1.56	3.30	3.57	2.32	1.36	0.00		
TH-011	7.10	4.88	1.37	3.74	4.36	1.50	1.87	2.43	3.61	4.18	0.00	
TH-012	7.55	5.46	1.70	4.84	5.24	2.27	2.52	3.24	4.60	5.35	1.73	0.00
TH-013	6.59	4.46	1.22	3.49	4.17	1.18	1.46	2.27	3.51	4.01	0.88	1.66
TH-014	0.92	3.43	7.02	4.70	4.29	6.37	6.53	5.47	4.78	4.39	7.66	8.07
TH-015	4.36	2.38	3.53	1.41	1.95	2.69	2.85	2.10	1.26	0.93	3.62	4.85
TH-016	7.01	4.97	1.66	3.64	4.52	1.71	2.04	2.56	3.82	4.20	0.86	1.97
TH-017	5.22	2.95	2.30	2.03	2.30	1.56	1.99	1.15	1.46	2.10	2.33	3.47
TH-018	4.83	3.02	4.17	1.30	2.13	3.34	3.59	2.46	1.63	1.13	3.94	5.32
TH-019	5.92	3.56	2.16	3.00	3.44	1.55	1.86	1.68	2.37	3.33	2.20	2.98
TH-020	6.50	4.21	0.99	3.49	3.92	0.89	1.29	1.99	3.15	3.90	0.93	1.61
TH-021	10.99	8.35	6.19	8.28	8.41	6.27	5.94	7.23	7.78	8.70	5.79	5.50





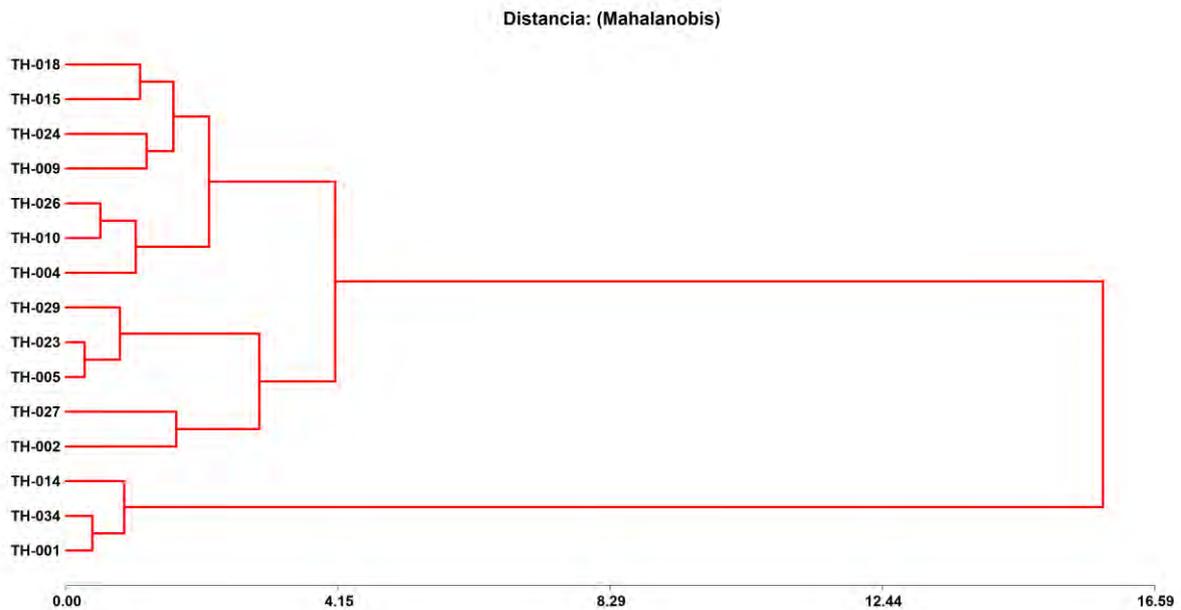
diferentes en su comportamiento respecto a las seis variables utilizadas para el análisis de conglomerados.



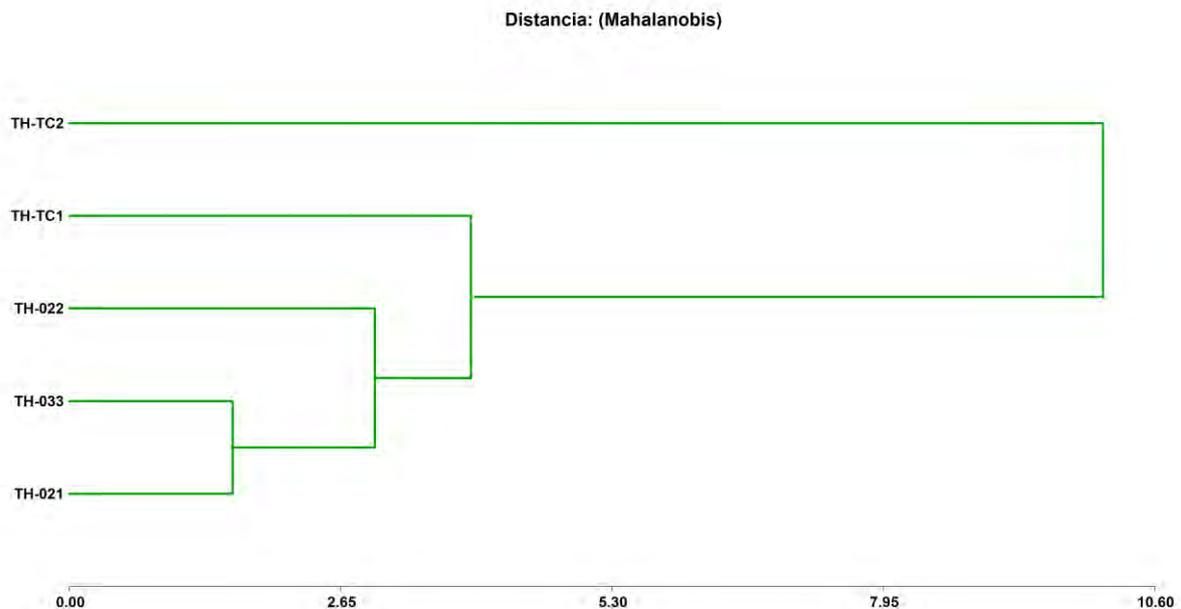
**Figura 5.** Dendrograma de los tratamientos en estudio que muestra la formación de tres grupos

Con estos datos se procedió al análisis de clúster por el método jerárquico con la medida de distancia de Mahalanobis y el método de enlace Ward que une los tratamientos buscando minimizar la varianza dentro de cada grupo como se muestra en la Tabla 10 y la Figura 5, los coeficientes obtenidos utilizando la distancia de Mahalanobis indican la distancia existente entre los tratamientos en estudio considerando las seis variables de análisis, por consiguiente un análisis detallado de la Tabla 10 y la Figura 5 muestra una gran similitud de los tratamientos TH-001 y TH-034 (con una distancia de 0.40), otros tratamientos que muestran similitud son las líneas TH-005 y TH-023 (con una distancia de 0.29), TH-005 y TH-029 (con una distancia de 0.63) y TH-004 y TH-026 (con una distancia de 0.85); Por otro lado los

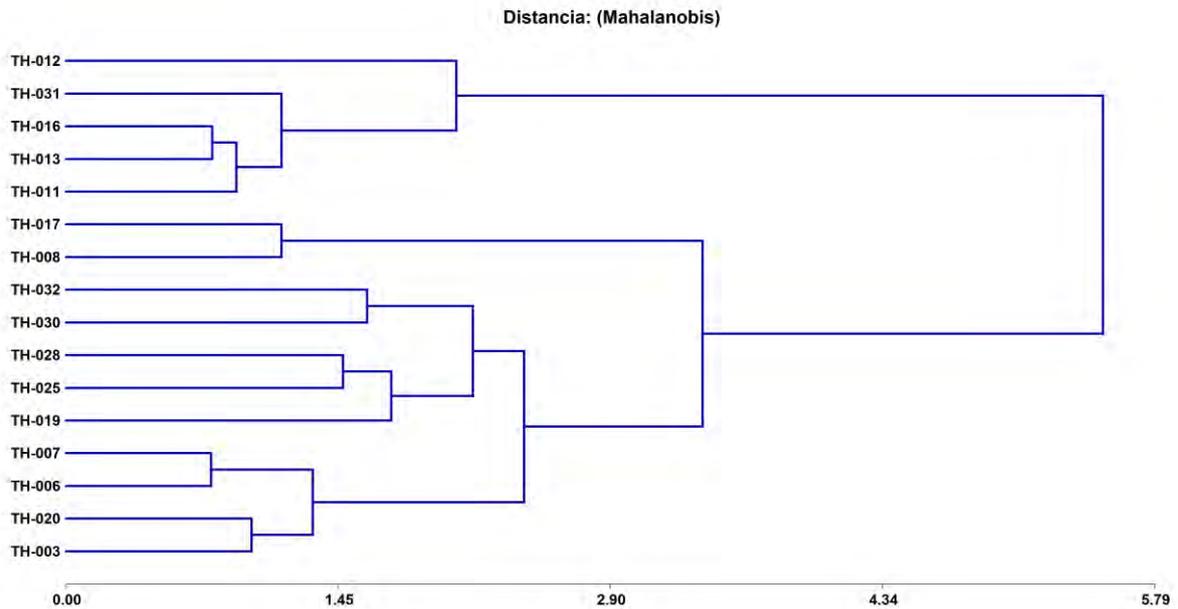
tratamientos más distantes con respecto a las variables en estudio fueron TH-001 con TH-021, TH-033 y TH-TC1 (con distancias de 10.99, 11.95 y 11.48), también el tratamiento TH-014 con TH-033, TC-TC1 Y TH-TC2 (con distancias de 12.60, 12.03 y 10.02).



**Figura 6.** Diagrama de los tratamientos que integran el grupo 1 del dendrograma



**Figura 7.** Diagrama de los tratamientos que integran el grupo 2 del dendrograma



**Figura 8.** Diagrama de los tratamientos que integran el grupo 3 del dendrograma

De las Figuras 6,7,8 que muestran la gráfica disgregada de la Figura 5 para cada uno de los clúster y la Tabla 11 los resultados muestran que el primer clúster está integrado por 15 tratamientos, caracterizándose por presentar los mayores valores promedios para las variables en estudio, el segundo clúster por cinco tratamientos que presentan los valores promedios inferiores de los tres conglomerados a excepción de la variable altura de planta y el tercer clúster por 16 tratamientos que representa en promedio a los tratamientos con menor tamaño de planta

**Tabla 11.** Características de cada conglomerado

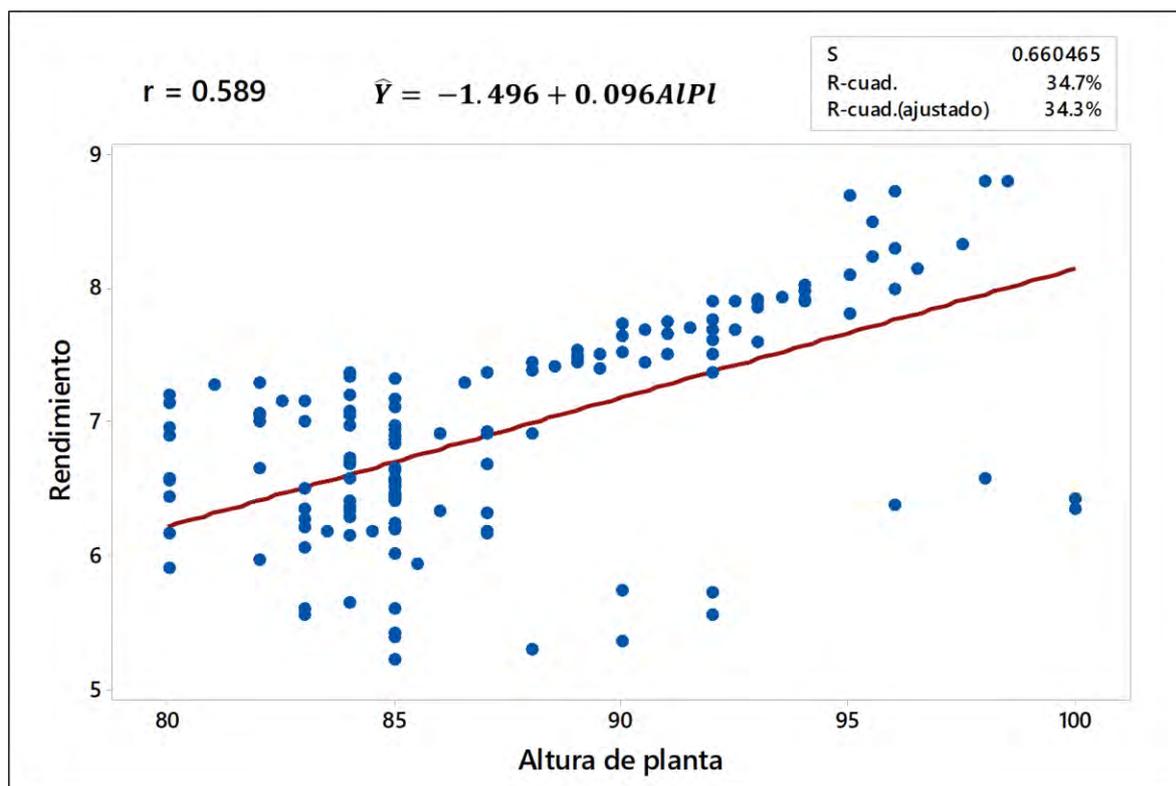
Clúster		Días a madurez	N° granos por espiga	Altura de la planta (cm)	Tamaño de espiga (cm)	Peso de mil granos (g)	Rendimiento en grano (t/ha)
1	n	15	15	15	15	15	15
	Media	152.12	63.07	90.33	8.12	59.51	7.7
	D.E.	2.8	2.83	4.94	0.92	0.08	0.45
	Mín	147	56	80	7	59.33	6.99
	Máx	158	69	98.5	9.8	59.65	8.8
2	n	5	5	5	5	5	5
	Media	140.65	44.8	88.38	6.47	59.19	5.74
	D.E.	2.91	3.52	6.12	0.31	0.04	0.4
	Mín	138	40	80	6	59.14	5.22
	Máx	148	55	100	6.9	59.3	6.58
3	n	16	16	16	16	16	16
	Media	145.61	49.77	84.52	6.88	59.31	6.61
	D.E.	2.78	6.19	2.3	0.48	0.07	0.39
	Mín	140	41	80	6	59.2	5.97
	Máx	153	65	93	9	59.58	7.91
Total	n	36	36	36	36	36	36
	Media	147	55	87.48	7.34	59.38	6.94
	D.E.	4.98	8.71	4.98	0.96	0.14	0.81
	Mín	138	40	40	6	59.14	5.22
	Máx	158	69	100	9.8	59.65	8.8

Donde: n= Número de tratamientos del grupo, D.E.= Desviación estándar, Mín.= Valor mínimo de los datos, Máx.=Valor máximo de los datos

## 5.2. Determinación de la influencia de la altura de la planta en el rendimiento de trigo harinero

### 5.2.1. Análisis de la regresión lineal para la altura de planta y el rendimiento

En la Figura 9 se muestra el comportamiento de la variable rendimiento de trigo harinero frente a la variable altura de la planta, en el que se puede observar que existe una relación positiva moderada expresada a través del coeficiente de correlación de Pearson,  $r=0.589$ . Al realizar el contraste de hipótesis para la prueba de correlación (ver expresión 12), se obtiene un valor  $T= 8.6938$  ( $p\text{-valor}<0.0000$ ) que está fuera de la región de aceptación, por lo tanto, se rechaza la  $H_0$  de correlación nula, en consecuencia, con estos resultados del estudio para estas dos variables se puede decir que bajo las condiciones experimentales y con un 5% de nivel de significancia hay una correlación lineal estadísticamente significativa entre la altura de planta y el rendimiento. En la misma Figura, se observa el modelo de regresión ajustado dado por  $\hat{Y} = -1.496 + 0.096AIP$ . Esta ecuación sugiere que la regresión del rendimiento sobre la altura de planta representa la esperanza de la distribución de los rendimientos según la altura explicando que hay un incremento del rendimiento en 0.096 toneladas por cada incremento de una unidad de la altura de planta.



**Figura 9.** Representación gráfica de la dispersión de los datos y la recta ajustada para el rendimiento y la altura de la planta.

**Tabla 12.** Análisis de la regresión lineal aplicada a la altura de la planta y el rendimiento

Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	ECMP	S
Rendimiento	144	0.347	0.343	0.453	0.673

Coeficientes estimados de la regresión e intervalos de confianza

	Estimador	Error Estándar.	Intervalo de confianza		T	p-valor
			LI (95%)	LS (95%)		
Intercepto	-1.496	0.972	-3.418	0.425	-1.539	0.126
Pendiente de la altura planta	0.096	0.011	0.075	0.118	8.694	<0.0001

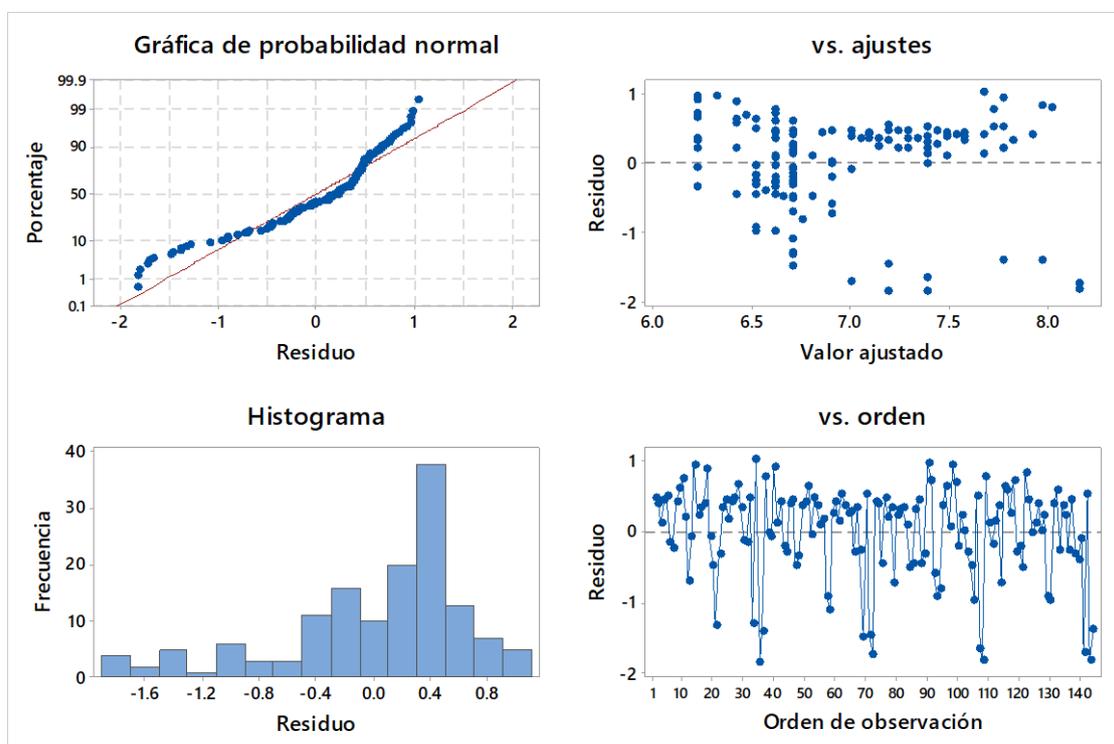
Análisis de varianza de la Regresión lineal para la altura de la planta y el rendimiento

F.V.	GL	SC	CM	F	p-valor
Regresión	1	32.970	32.97	75.583	<0.0001
Error	142	61.942	0.436		
Total	143	94.913			

Error Estándar Residual = 0.660

En la Tabla 12, se detallan los resultados del modelo de la regresión, entre estos se observan los criterios de bondad de ajuste, cómo el coeficiente de determinación  $R^2=0.347$ , indica que el modelo ajustado está explicando que el 34% de la variabilidad total presente en el rendimiento es debido a la variación observada en la altura de planta. Por otro lado, el coeficiente de la pendiente asociada a la altura de planta presenta un estimador de 0.096 t con un error estándar de 0.011 y un intervalo de confianza para el coeficiente de la pendiente con un grado de confianza del 95% [0.075, 0.117 t.].

En la misma Tabla 12, se observa el análisis de varianza para la regresión donde se muestra que la regresión lineal ajustada es estadísticamente significativa para el comportamiento de la variable rendimiento frente a la variable altura de planta ( $p\_valor < 0.0001$ , con 1 gl.), con una precisión del modelo expresado por el error estándar residual de 0.66 t.



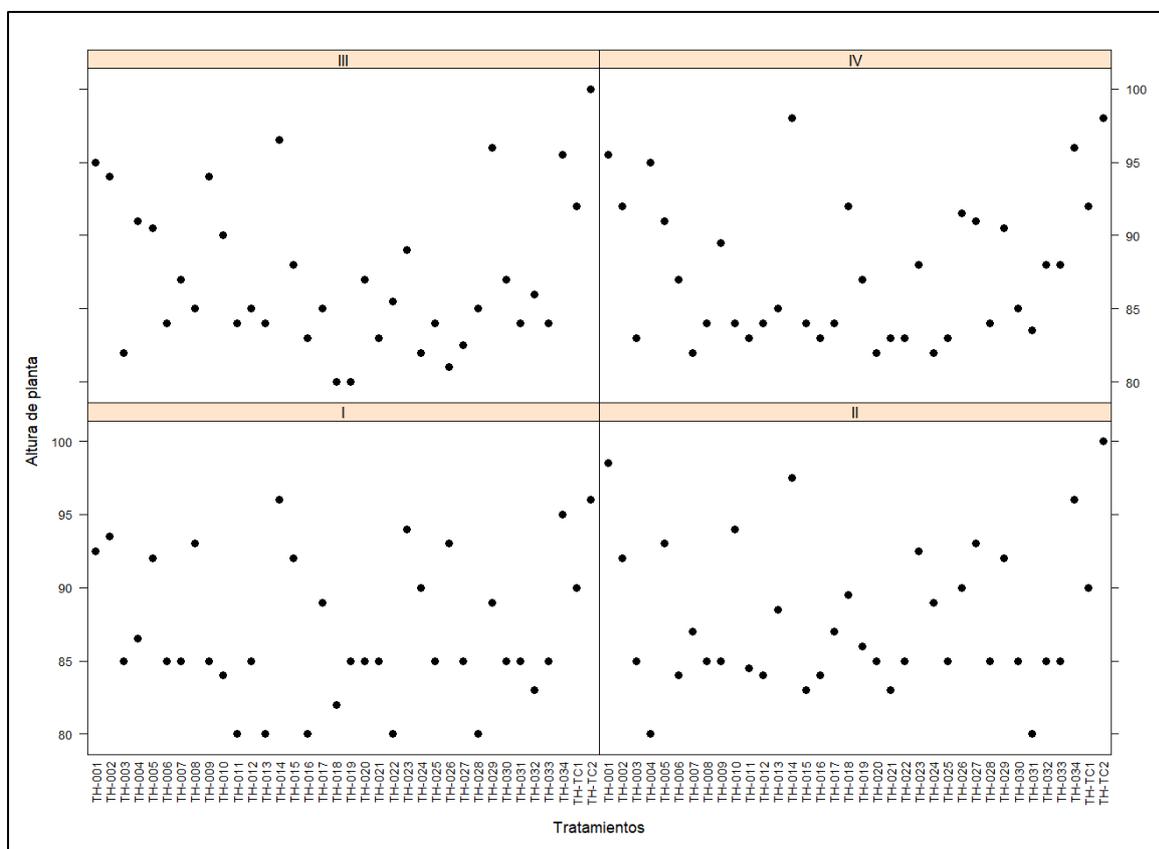
**Figura 10.** Representación gráfica de los residuos de la variable rendimiento frente a la altura de la planta.

Para validar el modelo ajustado se procedió a verificar los supuestos del modelo en función de los residuos para la variable rendimiento. Según los gráficos de la Figura 10, en caso de la normalidad: en el gráfico de la probabilidad normal, los residuos parecen ajustarse a una línea recta y en el histograma los residuos parecen seguir una distribución normal. En caso de la homogeneidad de varianzas de los residuos, el gráfico de los residuos versus los valores ajustados, no presentan un patrón sistemático y en el gráfico de residuos frente a las observaciones ordenadas, no se observa algún patrón indicando evidencias de independencia de residuos. Sin embargo, no basta con la simple visualización de los gráficos, sino se realizaron las pruebas estadísticas pertinentes para confirmar las afirmaciones. Se utilizó la prueba de Shapiro-Wilks con la finalidad de verificar la normalidad de la variable respuesta rendimiento, estableciéndose  $p\text{-valor}=0.100$  que es superior al nivel de significancia  $\alpha = 0.05$ , por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos provienen de una distribución normal. Así mismo, utilizando la prueba de Breusch\_Pagan para verificar la homogeneidad de las varianzas para la variable respuesta rendimiento encontramos un  $p\text{-valor} = 0.480$  siendo, superior al nivel de significancia  $\alpha = 0.05$  consecuentemente, bajo las mismas condiciones experimentales se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala la homogeneidad de varianzas como se puede observar en la Figura 10.

### **5.2.2. Análisis de varianza para la altura de planta**

Con el previo conocimiento de que la altura de planta influye en el rendimiento de trigo harinero, se procedió a evaluar, analizar y comparar los diferentes tratamientos de la variable altura de planta utilizando el diseño de bloques completos al azar.

La Figura 11, muestra el comportamiento de la altura de la planta según los diferentes tratamientos, distribuidos en 4 bloques, donde se puede observar que existen evidencias de mayor variabilidad de la altura de planta en algunos tratamientos bajo estudio y entre los diferentes bloques como es el caso de TH-TC1, TH-TC2, TH-003, TH-006, TH-007, TH-011, TH-012, TH-013; por consiguiente, se puede indicar que ésta variabilidad entre tratamientos es debido a la procedencia de diversos progenitores causantes del comportamiento genético.

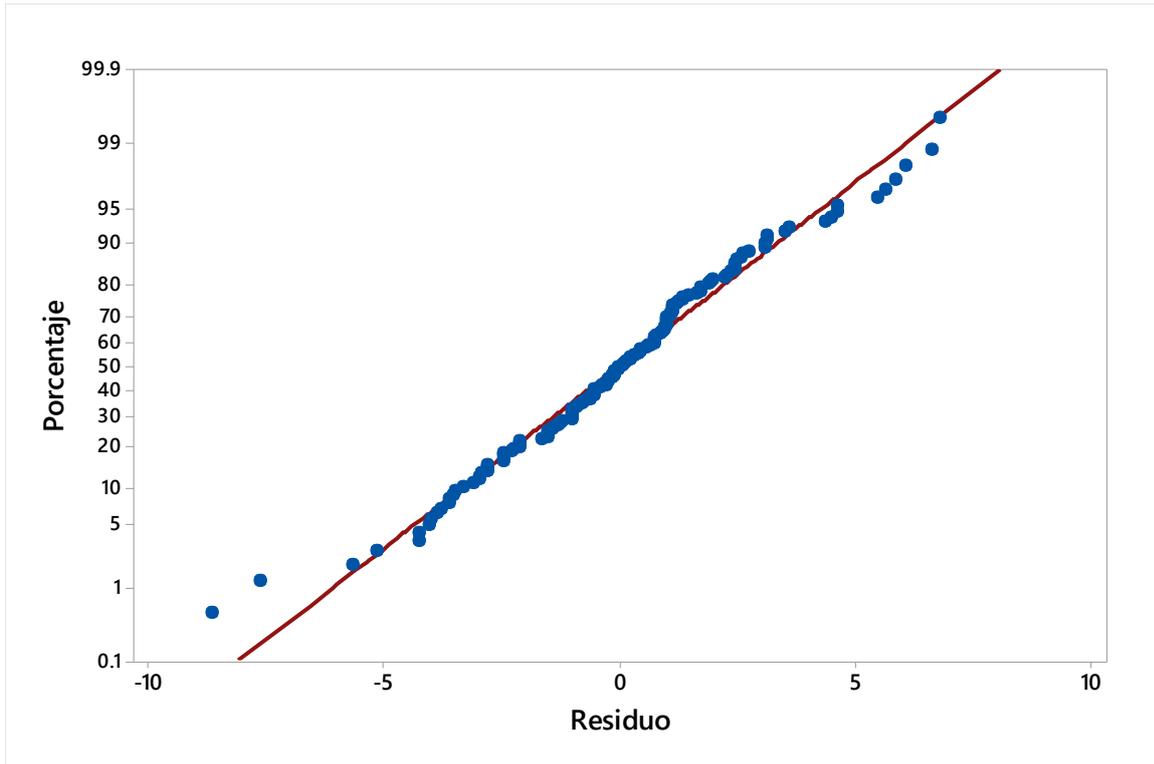


**Figura 11.** Representación gráfica de la dispersión de datos de altura de planta para los tratamientos por bloques.

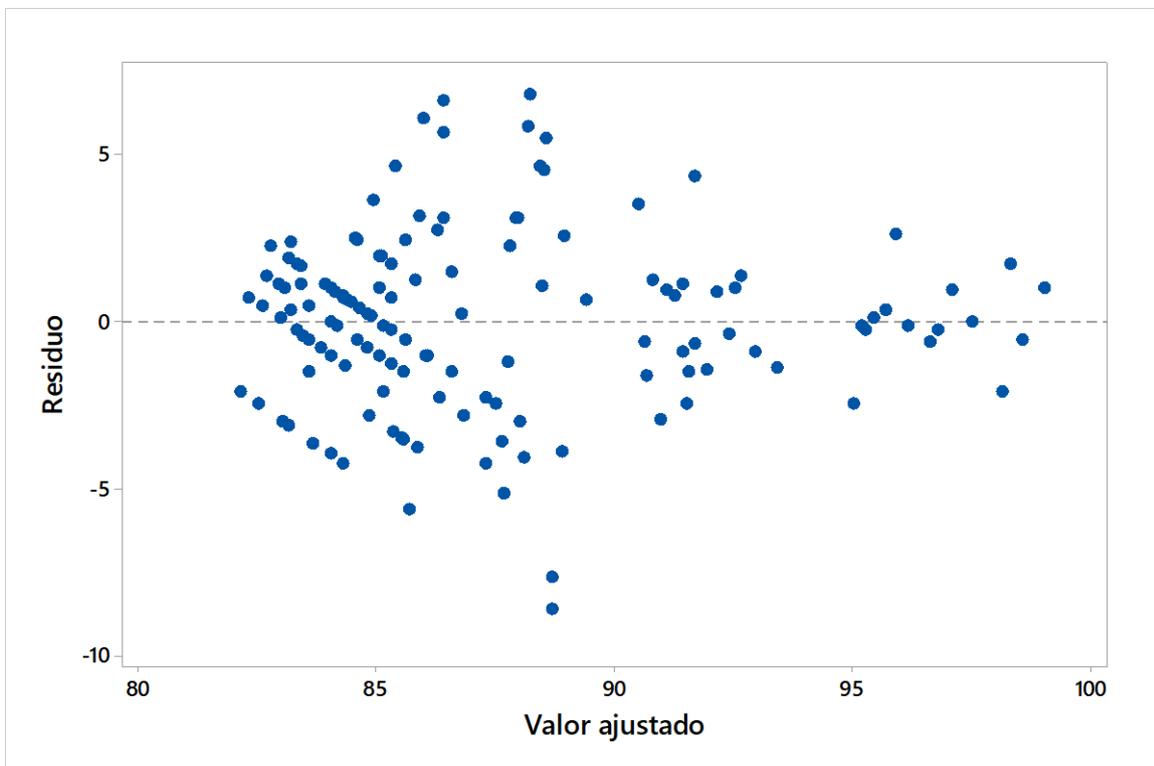
Utilizando el modelo lineal de la expresión (18) para la variable respuesta altura de la planta (*AIPi*) en función de los tratamientos y de los bloques, expresado por el siguiente modelo:

$$AlPl_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2); i = 1, \dots, 36, j = 1, 2, 3, 4$$

se procedió a efectuar el cumplimiento de los supuestos del modelo lineal para realizar un ANOVA, como la prueba de Shapiro Wilk y según la ecuación (20) se obtuvo un *p-valor* = 0.09482 superior al nivel de  $\alpha = 0.05$  que demuestra que los datos provienen de una distribución normal por tanto se tiene evidencias estadísticas para no rechazar la *H<sub>0</sub>* que señala que los datos de la altura de planta proviene de una distribución normal (Figura 11); Así mismo, utilizando la prueba de Bartlett con la ecuación (21) para el supuesto de homogeneidad de varianzas para los residuos, se obtuvo el *p-valor* = 0.0008269 bastante inferior al nivel de significación de  $\alpha = 0.05$ , por tanto, estas evidencias con llevan a aceptar la hipótesis alterna señalando que las varianzas de los residuos no son homogéneas (Figura 12), en el que se observa un claro patrón en forma de cono de la distribución de residuos de los datos originales, indicando dos situaciones fundamentales, la falta de independencia de los residuos y la heterogeneidad de la varianza de los mismos, siendo inaceptables para realizar un ANOVA. Siendo uno de los requisitos fundamentales verificar la homogeneidad de varianza de los residuos para realizar un ANOVA, se procedió a la transformación cuadrática de los datos de la variable altura de la planta, pero lamentablemente no se logró homogeneizar las varianzas, por tal motivo presentar los resultados a través de un ANOVA no procede. La solución alterna fue presentar el promedio de los tratamientos y su error estándar.



**Figura 12.** Gráfico de la distribución normal de los residuos de la variable altura de la planta.



**Figura 13.** Gráfico de los residuos y valores ajustados para la altura de la planta.

**Tabla 13.** Criterios de bondad de ajuste para la altura de la planta

Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	CV
Altura de planta	144	0.725	0.625	3.485

La Tabla 13, se da a conocer los valores del criterio de bondad de ajuste para el modelo lineal, en que el coeficiente de determinación ajustado,  $R^2 \text{ Aj}=0.625$ , explica que el 62.5% de la variabilidad presente en la altura de la planta está explicado por el modelo lineal a través del efecto de los tratamientos y de los bloques. Con un coeficiente de variación,  $CV=3.48\%$ , brinda información acerca de la precisión del experimento y de la sensibilidad del análisis (relación que existe entre la variabilidad residual y la media de los datos) siendo este valor del CV bajo según la clasificación propuesta por Pimentel citado por Álvares et al., (2017).

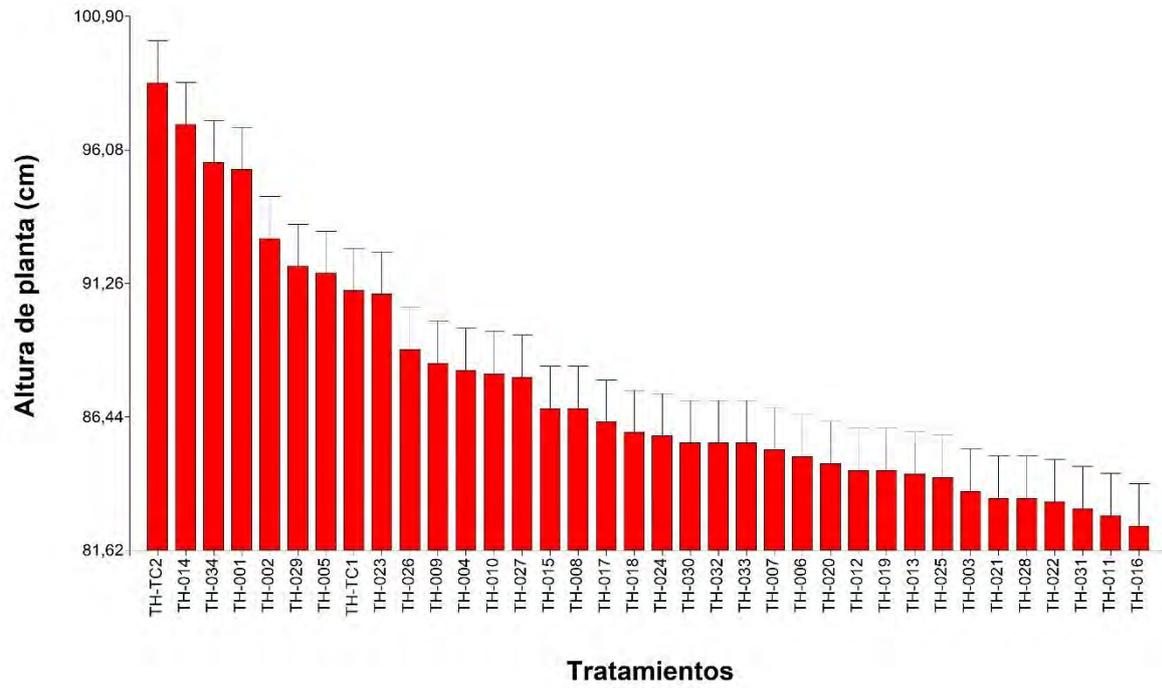
### 5.2.3. Comparación de medias de altura de la planta de los tratamientos

La Tabla 14 y Figura 14 nos muestran el valor medio y el error estándar de los tratamientos:

- Las medias muestrales, ordenadas en forma descendente, muestran que el genotipo TH-TC2, tiene el mayor desarrollo de planta ( $98.50 \pm 0.96 \text{ cm}$ ) superior a los demás genotipos evaluados, por consiguiente, muestran evidencias de ser diferente frente a los demás tratamientos. De esta manera resultados del ensayo pueden proyectarse a la población ya que se espera estabilidad de las relaciones halladas. Las medias poblacionales difieren entre los genotipos y es probable plantearse que estas diferencias agrónomicamente pueden influir con relevancia económica en grandes superficies de cultivo.

**Tabla 14.** Comparación de tratamientos frente a la altura de la planta (cm)

Tratamiento	Media (cm)		Error estándar
TH-TC2	98.50	±	0.96
TH-014	97.00	±	0.46
TH-034	95.63	±	0.24
TH-001	95.38	±	1.23
TH-002	92.88	±	0.52
TH-029	91.88	±	1.51
TH-005	91.63	±	0.55
TH-TC1	91.00	±	0.58
TH-023	90.88	±	1.42
TH-026	88.88	±	2.70
TH-009	88.38	±	2.15
TH-004	88.13	±	3.22
TH-010	88.00	±	2.45
TH-027	87.88	±	2.47
TH-008	86.75	±	2.10
TH-015	86.75	±	2.06
TH-017	86.25	±	1.11
TH-018	85.88	±	2.89
TH-024	85.75	±	2.17
TH-030	85.50	±	0.50
TH-032	85.50	±	1.04
TH-033	85.50	±	0.87
TH-007	85.25	±	1.18
TH-006	85.00	±	0.71
TH-020	84.75	±	1.03
TH-012	84.50	±	0.29
TH-019	84.50	±	1.55
TH-013	84.38	±	1.75
TH-025	84.25	±	0.48
TH-003	83.75	±	0.75
TH-021	83.50	±	0.5
TH-028	83.50	±	1.19
TH-022	83.38	±	1.25
TH-031	83.13	±	1.09
TH-011	82.88	±	1.01
TH-016	82.50	±	0.87



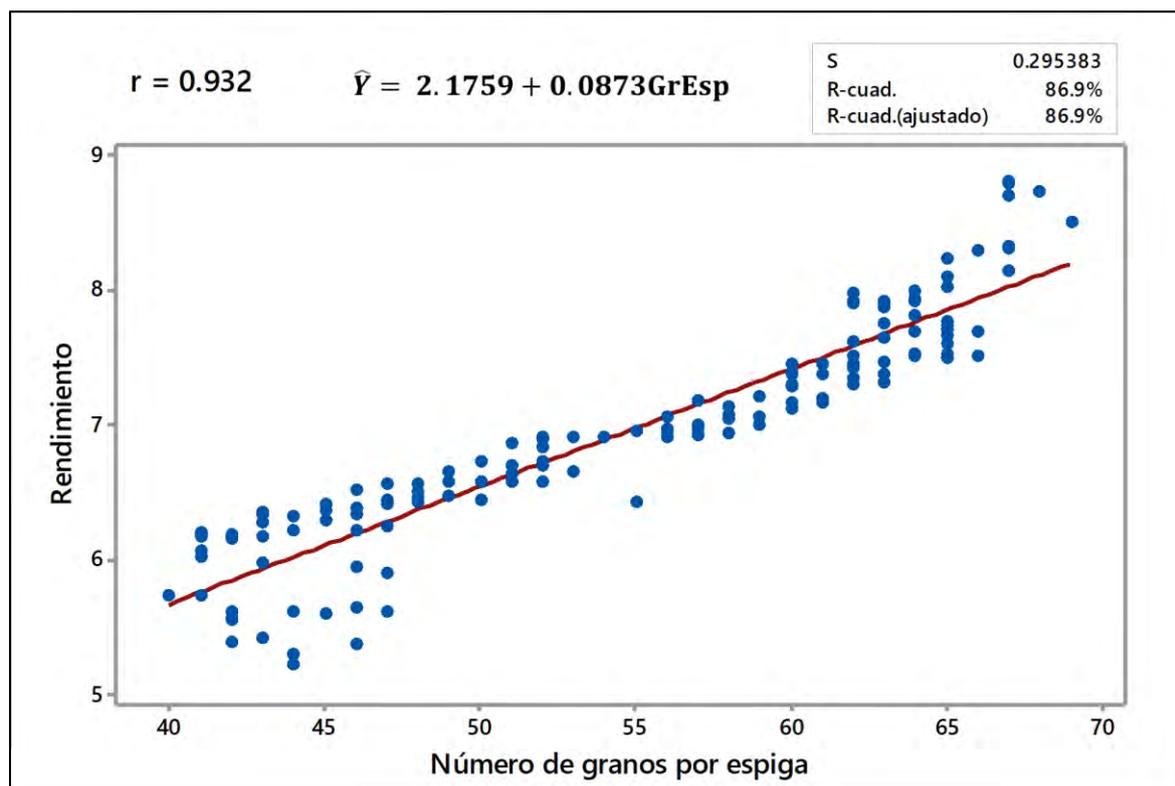
**Figura 14.** Comparación de altura de la planta (cm) de los tratamientos en estudio.

### **5.3. Determinación de la influencia del número de granos por espiga en el rendimiento de trigo harinero**

#### **5.3.1. Análisis de regresión lineal para el número de granos por espiga y el rendimiento de trigo harinero**

En la Figura 15 se muestra el comportamiento de la variable rendimiento frente a la variable número de granos por espiga donde se visualiza evidencias de que existe una relación positiva alta con un  $r=0.932$ , según la expresión (11), indicando que muestra evidencias de tener una relación lineal entre esas dos variables observadas. Al realizar el contraste de hipótesis para la prueba de correlación (ecuación 12), se obtiene un valor  $T= 30.754$  ( $p\text{-valor}=0.00001$ ) que está fuera de la región de aceptación por lo tanto se rechaza la  $H_0$  de correlación nula en consecuencia con estos resultados del estudio para estas dos variables se puede decir que hay una correlación lineal estadísticamente significativa entre el número de granos por espiga y el rendimiento.

Para obtener el modelo de regresión que describa la variación del rendimiento en función del número de granos por espiga de los tratamientos fue utilizando la expresión (5), cuyo modelo ajustado que se muestra en la misma Figura 15 está dado por  $\hat{Y} = 2.1759 + 0.0873 GrEsp$ . Este modelo sugiere que la regresión del rendimiento sobre el número de granos por espiga representa la esperanza de la distribución del rendimiento explicando que hay un incremento del rendimiento en 0.0873 toneladas por cada incremento en el número de granos por espiga.



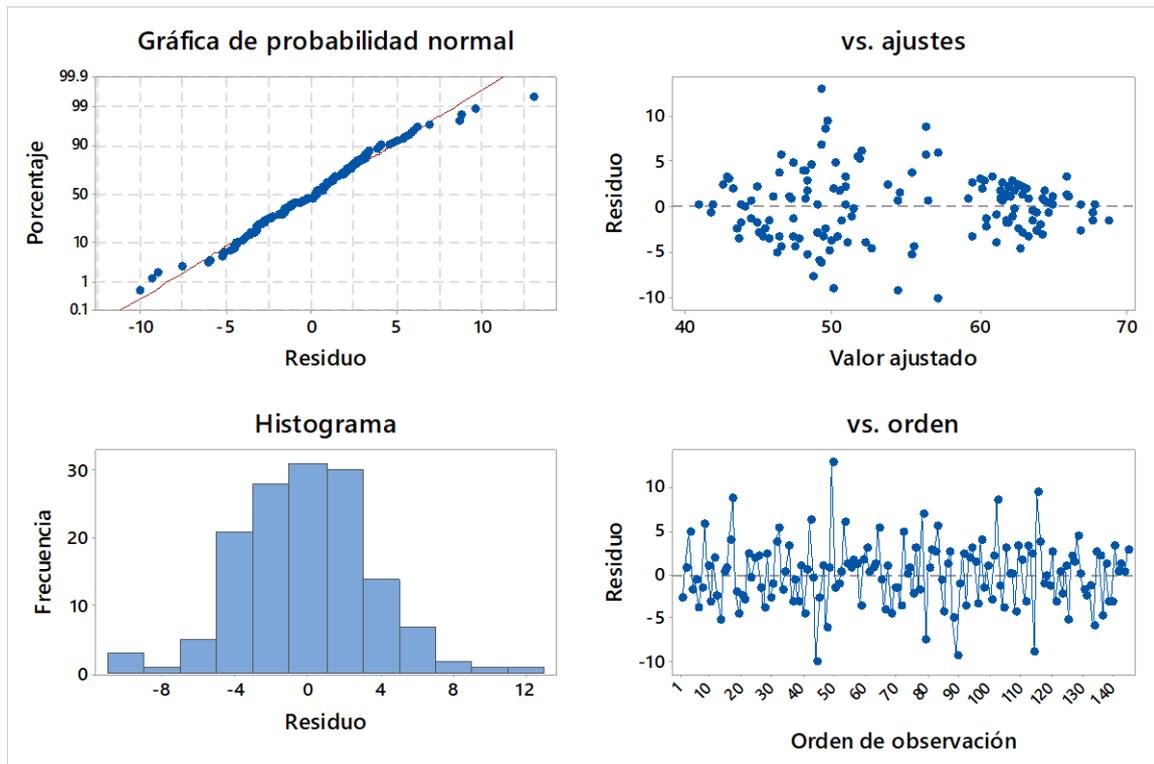
**Figura 15.** Representación gráfica de la dispersión de los datos y la recta ajustada para el rendimiento y el número de granos por espiga.

**Tabla 15.** Aplicación del análisis de regresión lineal para el número de granos por espiga y el rendimiento

Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	ECMP	s	
Rendimiento	144	0.8695	0.8685	0.0902	0.300	
Coeficientes estimados en intervalos de confianza						
Coeficiente	Estimador	Error Estándar	Intervalo de confianza (95%)		T	p-valor
			LI	LS		
Intercepto	2.176	0.157	1.866	2.486	13.867	<0.0001
Pendiente para el N° granos por espiga	0.0873	0.003	0.082	0.093	30.754	<0.0001
Análisis de varianza para la regresión lineal del número de granos por espiga y el rendimiento						
F.V.	GL	SC	CM	F	p-valor	
Regresión	1	82.523	82.523	945.813	<0.0001	
Error	142	12.389	0.087			
Total	143	94.912				

Error Estándar Residual = 0.295

Los resultados del modelo de la regresión se muestran en la Tabla 15, así el criterio de bondad de ajuste que muestra el coeficiente de determinación,  $R^2=0.8695$ , explica que el 86.95% de la variabilidad total presente en el rendimiento está siendo explicado por el modelo ajustado a través de la variación observada en el número de granos por espiga. Por otro lado, el estimador del coeficiente de la pendiente de la regresión asociado al número de granos por espiga presenta un valor estimado de 0.087 toneladas con un error estándar de 0.003 toneladas a un intervalo de confianza del 95% entre los límites de  $[0.082, 0.093 t]$ . Así mismo el p-valor  $<0.0001$  nos indica que la pendiente que relaciona el rendimiento con el número de granos por espiga es significativamente distinta de cero. Seguidamente, en los resultados del análisis de varianza para la regresión se observa el p-valor  $<0.0001$  en el que indica que la regresión lineal ajustada es estadísticamente significativa para el comportamiento de la variable rendimiento frente a la variable número de granos por espiga con un error estándar residual de 0.295 toneladas.



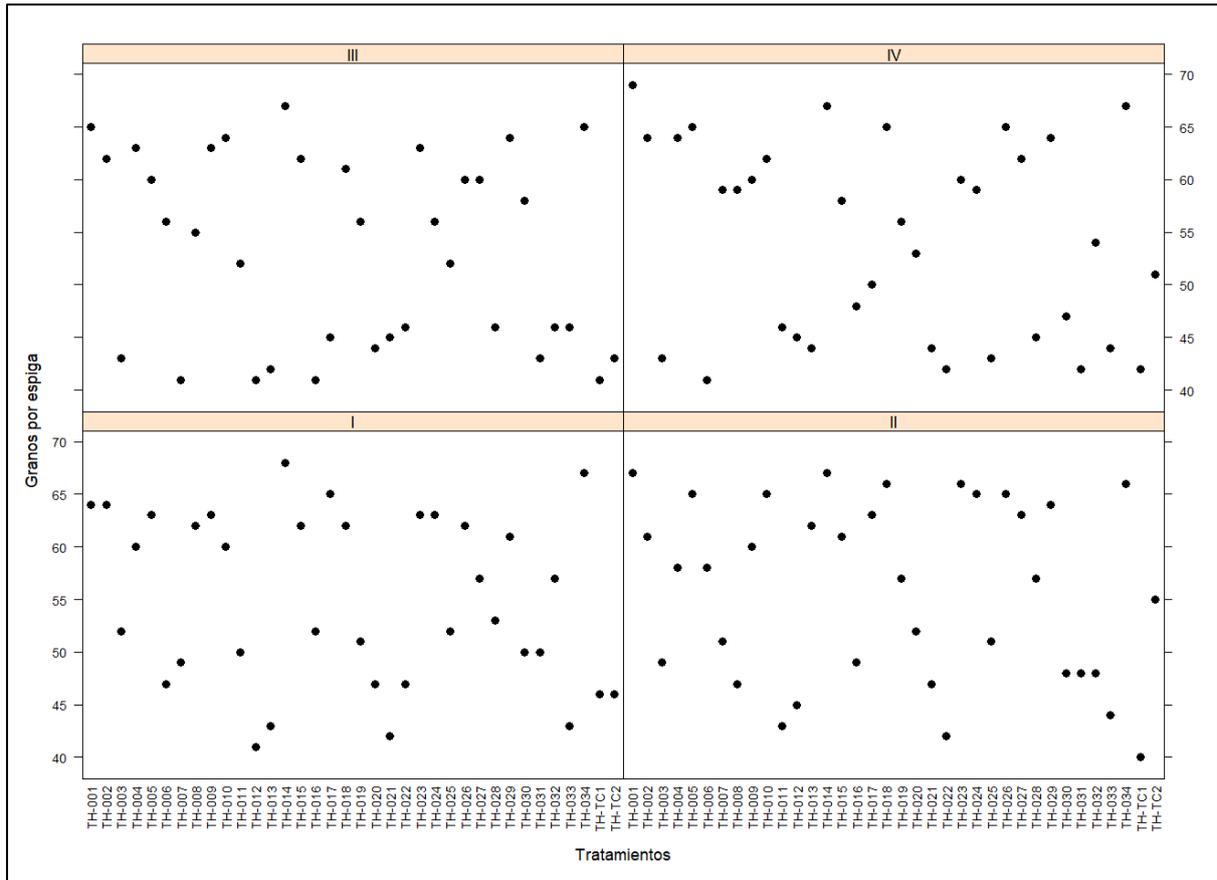
**Figura 16.** Representación gráfica de los residuos de la variable rendimiento frente al número de granos por espiga.

Para validar el modelo ajustado se procedió a verificar las condiciones del modelo a través del uso de residuos para la variable rendimiento. La Figura 16, muestra diferentes gráficas, entre estos de la probabilidad normal y el respectivo histograma de los residuos, donde se observa que los errores muestran evidencias de un comportamiento normal. Para verificar este comportamiento de los residuos de la variable respuesta rendimiento, se utilizó la prueba de Shapiro-Wilks resultando un p-valor=0.3629 que es superior al nivel de significancia  $\alpha=0.05$ ; por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos provienen de una distribución normal. Así mismo, los gráficos de los residuos frente a los valores observados y orden de observación de la Figura 16, muestran evidencias de la distribución de los residuos lo cual se observa un patrón totalmente aleatorio. Utilizando la prueba de Breusch-Pagan para verificar la homogeneidad de las

varianzas para la variable respuesta rendimiento se determinó un  $p\_valor=0.3144$  comparado con el nivel de significancia  $\alpha=0.05$  resulta ser superior y consecuentemente, bajo las mismas condiciones experimentales se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala la homogeneidad de varianzas de los residuos. Después de validar el modelo, se puede decir que la variable número de granos por espiga si influye en la variable rendimiento y cuya variabilidad está siendo explicada en un 86.95% por el modelo.

### **5.3.2. Análisis de varianza para el número de granos por espiga**

Con el previo conocimiento de que el número de granos por espiga influye en el rendimiento de trigo harinero, se procedió a evaluar, analizar y comparar los diferentes tratamientos de la variable número de granos por espiga utilizando el diseño de bloques completos al azar. La Figura 17, muestra el comportamiento del número de granos por espigas según los 36 tratamientos distribuidos en 4 bloques. Donde podemos observar mayor variabilidad de respuesta de algunos tratamientos en estudio como es el caso de TH-006, TH-017, TH-022, TH-025, TH-TC2 en los diferentes bloques; por consiguiente, se puede indicar que existe variabilidad entre tratamientos debido a su comportamiento genético debido a la procedencia de diversos progenitores.



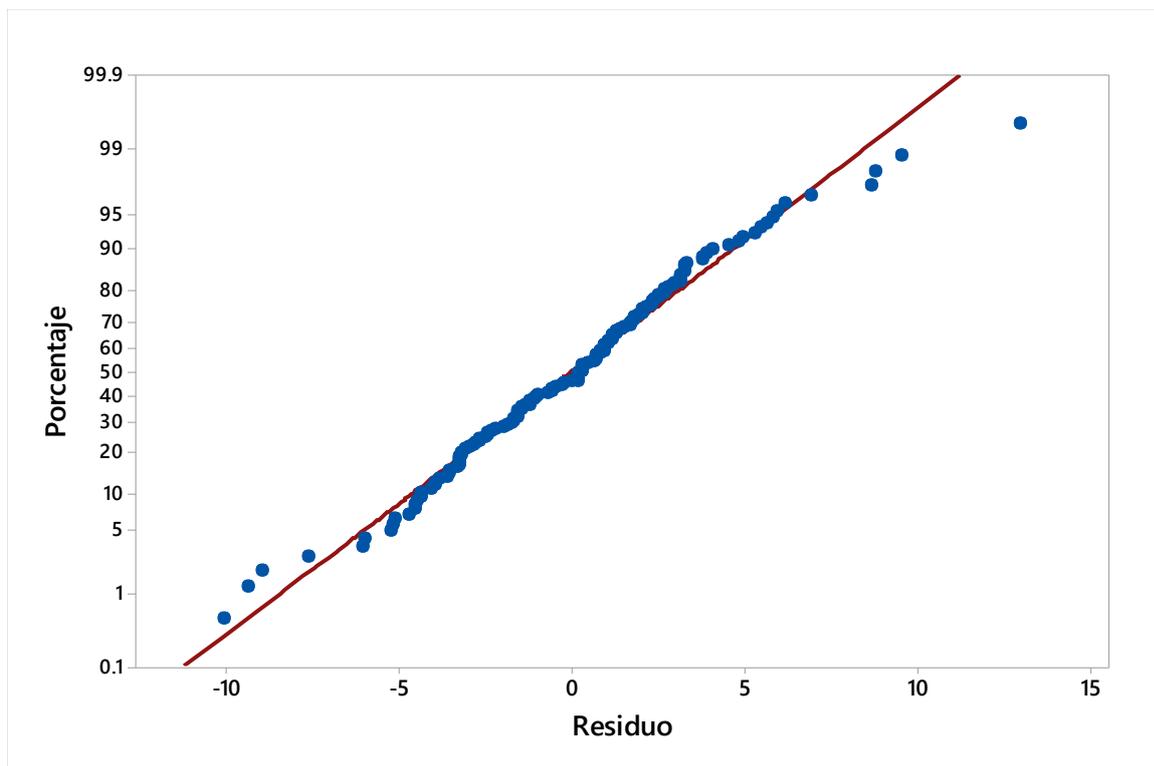
**Figura 17.** Gráfico de la dispersión del número de granos por espiga de los tratamientos por bloques.

Utilizando el modelo lineal de la expresión (18) para la variable respuesta número de granos por espiga (GrEsp) en función de los tratamientos y de los bloques, expresado por el siguiente modelo:

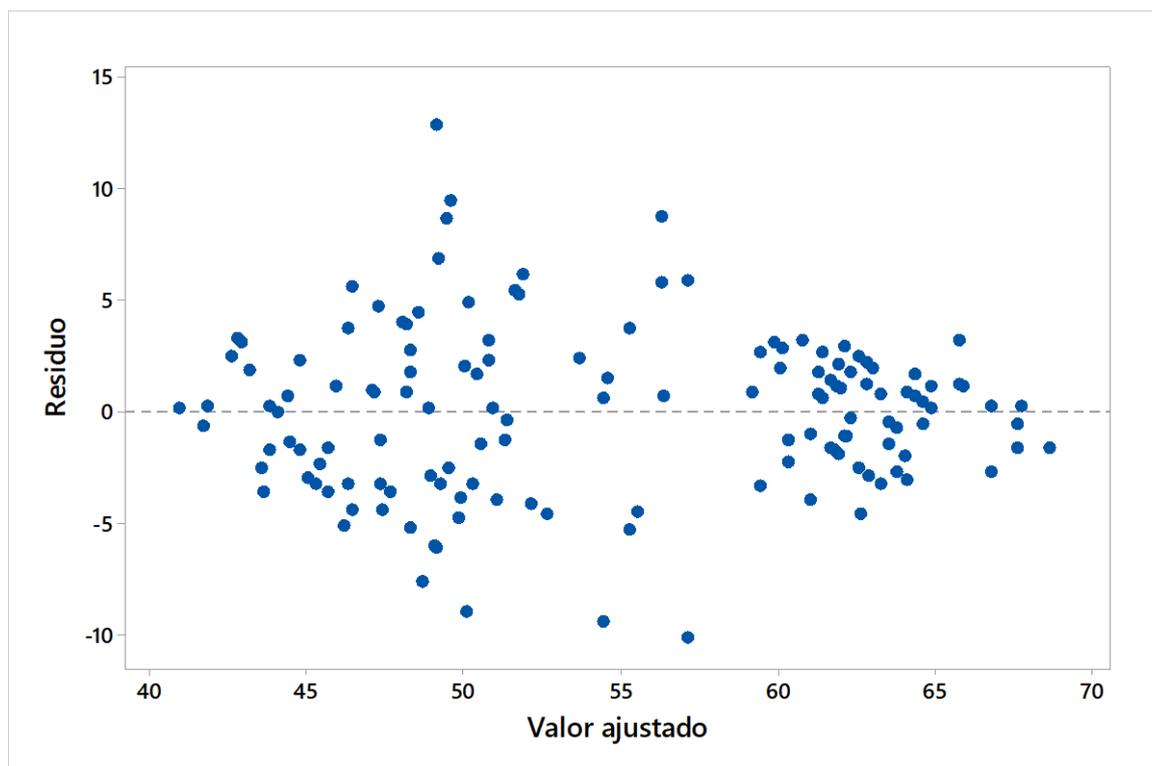
$$GrEsp_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2); i = 1, \dots, 36, j = 1, 2, 3, 4$$

se procedió a efectuar el cumplimiento de los supuestos del modelo lineal para realizar un ANOVA, como la prueba de Shapiro Wilks según la ecuación (20) se obtuvo un p-valor = 0.1376 superior al nivel de  $\alpha = 0.05$ , que demuestra que los datos provienen de una distribución normal (Figura 18); por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos del número de granos por espiga provienen de una distribución normal. Así mismo, a través de la prueba de Bartlett con la ecuación (21) para el supuesto de homogeneidad de varianzas se

obtuvo el  $p$ -valor=0.0008319 bastante inferior al nivel de significación de  $\alpha=0.05$ ; por tanto, estas evidencias llevan a aceptar la hipótesis alterna señalando que las varianzas de los residuos no son homogéneas (Figura 19), en el que se observa un claro patrón sistemático de la distribución de residuos de los datos originales, indicando dos situaciones fundamentales, la falta de independencia de los residuos y la heterogeneidad de la varianza de los mismos, siendo inaceptables para realizar un ANOVA, se procedió a la transformación cúbica de los datos de la variable granos por espiga, pero no se logró homogeneizar las varianzas, por tal motivo no se presentan los resultados a través de un ANOVA y la alternativa es mostrar el promedio de los tratamientos y su error estándar y, en un nuevo ensayo, corregir las causas que ocasionaron la heterogeneidad de varianzas.



**Figura 18.** Gráfica de la distribución normal de los residuos de la variable granos por espiga.



**Figura 19.** Gráfica de residuos y valores ajustados para el número de granos por espiga.

**Tabla 16.** Criterios de bondad de ajuste para el número de granos por espiga.

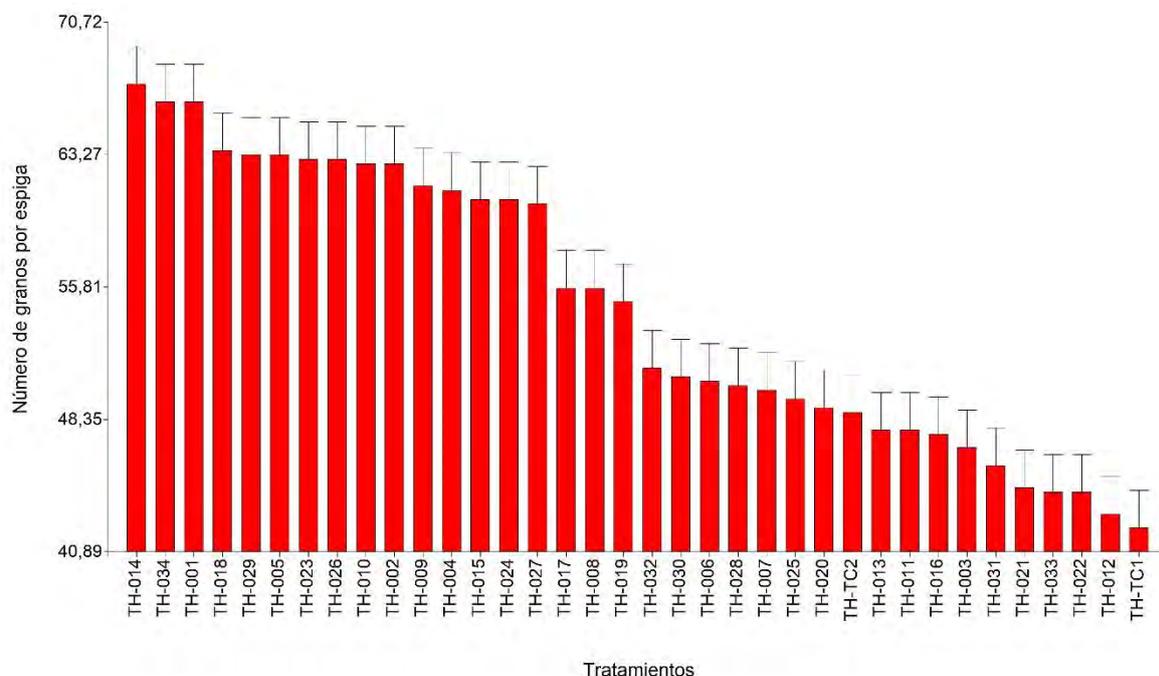
Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	CV
Nº granos por espiga	144	0.826	0.764	7.751

La Tabla 16, da a conocer los valores del criterio de bondad de ajuste para el modelo lineal, en que el coeficiente de determinación ajustado,  $R^2 \text{ Aj}=0.764$ , explica que el 76.4% de la variabilidad presente en el número de granos por espiga está explicado por el modelo lineal a través del efecto de los tratamientos y de los bloques conociendo las condiciones experimentales a los que fueron expuestos los tratamientos en estudio y con un coeficiente de variación,  $CV=7.75\%$  estableciendo la precisión del experimento como óptimo comparado al 10%, siendo este valor del CV bajo según la clasificación propuesta por Pimentel citado por Álvares et al., (2017).

### 5.3.3. Comparación de medias del número de granos por espiga de los tratamientos en estudio.

La Tabla 17 y Figura 20 nos muestran el valor medio y el error estándar de los tratamientos:

- Las medias muestrales, ordenadas en forma descendente, muestran que el genotipo TH-014, tiene el mayor número de granos por espiga ( $67 \pm 0.25$ ), superior a los demás genotipos evaluados.
- Este genotipo muestra evidencias significativas frente a los demás tratamientos por presentar mayor capacidad de desarrollo de granos por cada espiga; es decir las diferencias muestrales observadas pueden haberse dado por azar y por tanto no ser repetibles. Las medias poblacionales del número de granos por espiga difieren entre los genotipos y es probable plantearse que estas diferencias agronómicamente pueden influir con relevancia económica en grandes superficies de cultivo.



**Figura 20.** Comparación de medias para el número de granos por espiga de los tratamientos en estudio.

**Tabla 17.** Valor medio y error estándar por tratamiento para el número de granos por espiga.

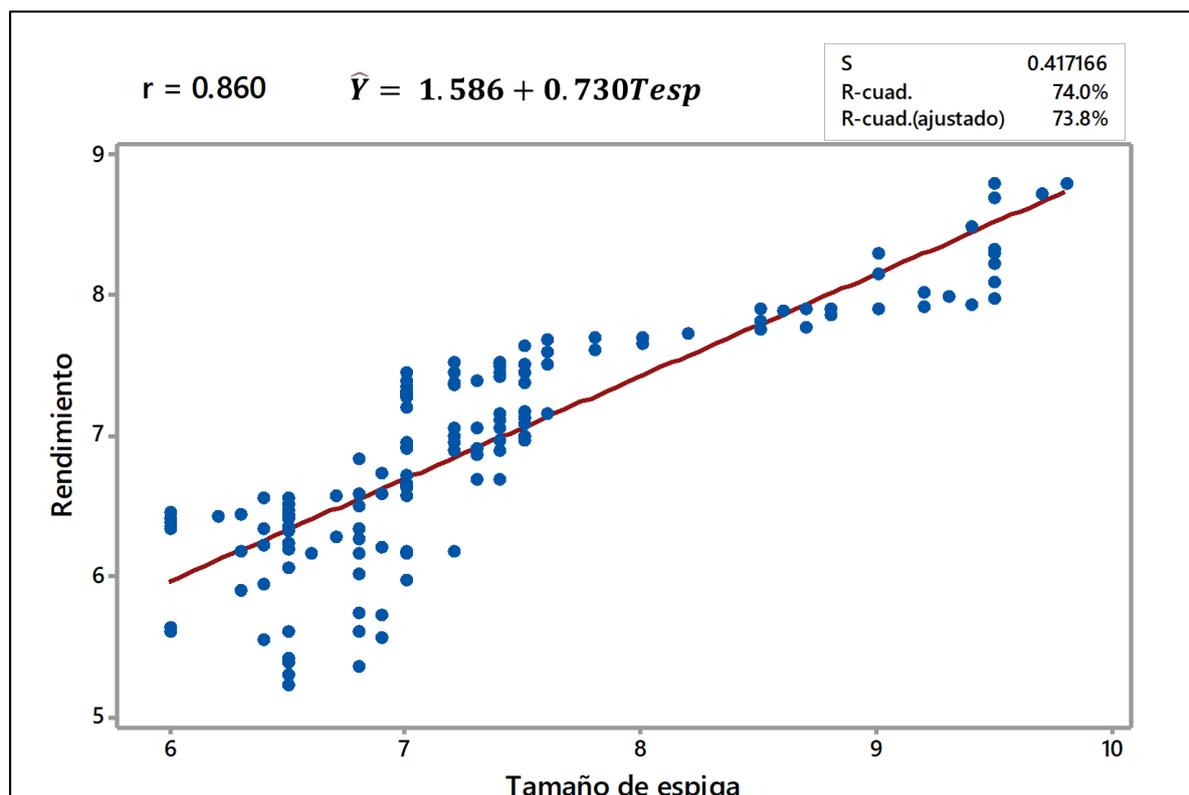
Tratamiento	Media		Error estándar
TH-014	67	±	0.25
TH-001	66	±	1.11
TH-034	66	±	0.48
TH-018	64	±	1.19
TH-005	63	±	1.18
TH-029	63	±	0.75
TH-023	63	±	1.22
TH-026	63	±	1.22
TH-002	63	±	0.75
TH-010	63	±	1.11
TH-009	62	±	0.87
TH-004	61	±	1.38
TH-015	61	±	0.95
TH-024	61	±	2.02
TH-027	61	±	1.32
TH-008	56	±	3.25
TH-017	56	±	4.89
TH-019	55	±	1.35
TH-032	51	±	2.56
TH-030	51	±	2.5
TH-006	51	±	3.97
TH-028	50	±	2.87
TH-007	50	±	3.7
TH-025	50	±	2.18
TH-020	49	±	2.12
TH-TC2	49	±	2.66
TH-011	48	±	2.02
TH-013	48	±	4.77
TH-016	48	±	2.33
TH-003	47	±	2.25
TH-031	46	±	1.93
TH-021	45	±	1.04
TH-022	44	±	1.31
TH-033	44	±	0.63
TH-012	43	±	1.15
TH-TC1	42	±	1.31

#### **5.4. Determinación de la influencia del tamaño de espiga en el rendimiento de trigo harinero.**

##### **5.4.1. Análisis de regresión lineal para el tamaño de espiga y el rendimiento de trigo harinero**

En la Figura 21 se muestra el comportamiento de la variable rendimiento frente a la variable tamaño de espiga, donde muestra evidencias de que existe una relación positiva alta establecido por el coeficiente de correlación de Pearson,  $r=0.860$ . Al realizar el contraste de hipótesis para la prueba de correlación con la ecuación (11), se obtiene un valor  $T=20.0845$  ( $p\text{-valor}=0.0000$ ) que está fuera de la región de aceptación por lo tanto se rechaza la  $H_0$  de correlación nula en consecuencia con estos resultados del estudio para estas dos variables se puede decir que hay una correlación lineal estadísticamente significativa entre el tamaño de espiga y el rendimiento.

Para obtener el modelo de regresión que describa la variación del rendimiento en función del tamaño de espiga de los tratamientos fue utilizando la expresión (5), el resultado del modelo ajustado que se muestra en la misma figura está dado por la regresión  $\hat{Y} = 1.586 + 0.730T_{esp}$ . Esta ecuación sugiere que la regresión del rendimiento sobre el tamaño de espiga representa la esperanza de la distribución de los rendimientos explicando que hay un incremento del rendimiento en 0.730 toneladas por cada incremento del tamaño de espiga.



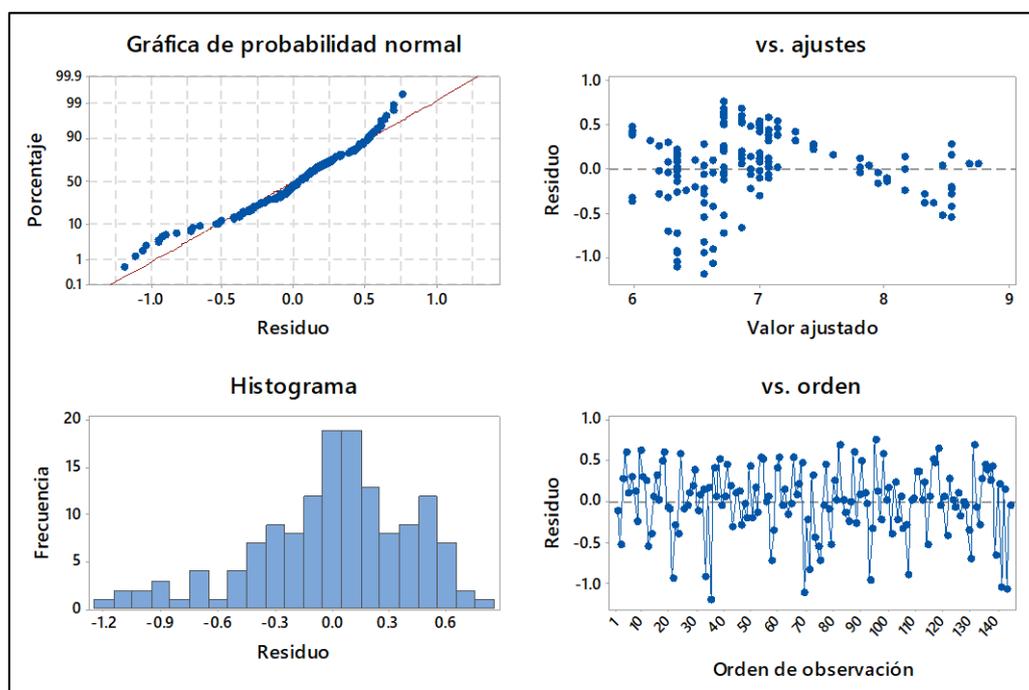
**Figura 21.** Gráfico de la dispersión de los datos y la recta ajustada para el rendimiento y el tamaño de espiga.

**Tabla 18.** Análisis de regresión lineal aplicado al tamaño de espiga y el rendimiento

Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	ECMP	s	
Tamaño espiga	144	0.74	0.738	0.178	0.422	
Coeficientes estimados en intervalos de confianza						
Coeficiente	Estimador	Error Estándar	Intervalo de confianza (95%)		T	p-valor
			LI	LS		
Intercepto	1.586	0.269	1.054	2.117	5.896	<0.0001
Pendiente para el tamaño de espiga	0.73	0.036	0.658	0.802	20.085	<0.0001
Análisis de varianza para la regresión lineal del tamaño de espiga y el rendimiento						
F.V.	GL	SC	CM	F	p-valor	
Regresión	1	70.201	70.201	403.391	<0.0001	
Error	142	24.712	0.174			
Total	143	94.913				

Error estándar residual = 0.417

La Tabla 18, muestra los resultados del modelo de la regresión con un coeficiente de determinación  $R^2=0.740$  indicando que la proporción de la variabilidad total del rendimiento, está siendo explicado por la relación lineal explícitamente a través de la variación observada en el tamaño de espiga. Por otro lado, el coeficiente de regresión de la pendiente asociada al tamaño de espiga presenta un valor estimado de 0.730 toneladas con un error estándar de 0.036 toneladas y un intervalo de confianza para el coeficiente de la pendiente con un grado de confianza del 95% [0.658, 0.802 t.]. Así mismo, la pendiente de la regresión que relaciona el rendimiento con el tamaño de espiga es significativamente distinta de cero ( $p<0.0001$ ). Seguidamente se utilizó un análisis de varianza para la regresión en el que se puede observar que el modelo de la regresión es adecuado ( $p\text{-valor}<0.0001$ ), en el que muestra que la regresión es estadísticamente significativa para el comportamiento de la variable rendimiento frente a la variable tamaño de espiga con un error estándar residual de 0.417 toneladas.



**Figura 22.** Gráfica de los residuos de la variable rendimiento frente al tamaño de espiga.

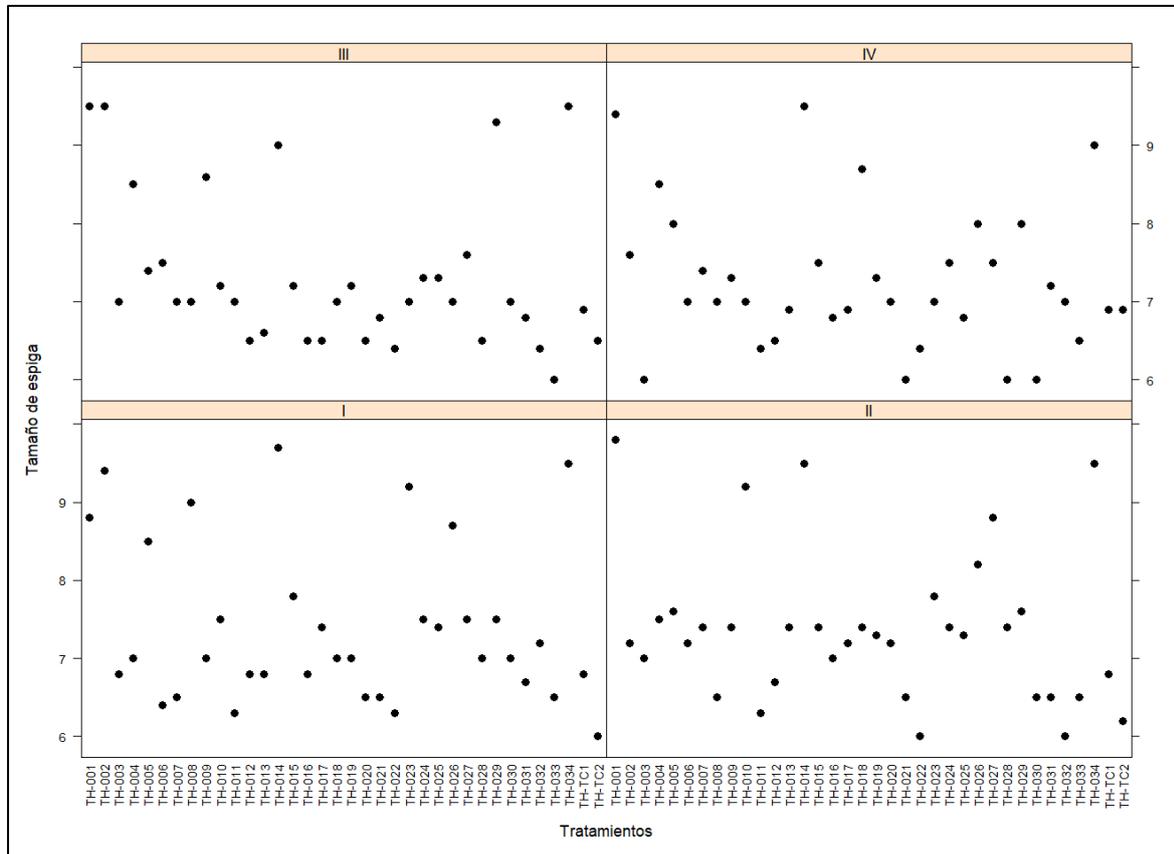
Para validar el modelo ajustado se procedió a verificar las condiciones del modelo a través de los residuos para la variable rendimiento. La Figura 22, muestra las gráficas de la probabilidad normal y del histograma de los residuos donde se observa un comportamiento casi lineal y una distribución asimétrica de los residuos, respectivamente, para lo cual se utilizó la prueba de Shapiro-Wilks con la finalidad de verificar la normalidad de los residuos de la variable respuesta rendimiento, mostrando un  $p\text{-valor}=0.0745$  que es superior al nivel de significancia  $\alpha = 0.05$ ; por tanto, se concluye que existen evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos provienen de una distribución normal. Así mismo, las gráficas de los residuos frente a los valores observados y los valores ordenados de la Figura 22, muestran un comportamiento aleatorio de los residuos en torno de cero. Utilizando la prueba de Breusch-Pagan para verificar la homogeneidad de las varianzas para la variable respuesta rendimiento encontramos un  $p\text{-valor}=0.0614$  siendo, superior al nivel de significancia  $\alpha = 0.05$  consecuentemente, bajo las mismas condiciones experimentales se concluye que se tiene evidencias estadísticas para no rechazar la  $H_0$ , señalando la homogeneidad de varianzas de los residuos de la variable rendimiento.

#### **5.4.2. Análisis de varianza para el tamaño de espiga**

Con el previo conocimiento de que el tamaño de espiga influye en el rendimiento de trigo harinero, se procedió a evaluar, analizar y comparar los diferentes tratamientos de la variable tamaño de espiga utilizando el diseño de bloques completos al azar.

En la Figura 23 se puede observar el comportamiento de los datos de la variable tamaño de espiga según los 36 tratamientos distribuidos en 4 bloques, en el que se observa que el tamaño de espiga presenta mayor variabilidad en algunos tratamientos en estudio en los 4 bloques como es el caso de TH-005, TH-008, TH-027, por

consiguiente, se puede indicar que existe variabilidad entre tratamientos debido a su comportamiento genético por causa de la procedencia de diversos progenitores.



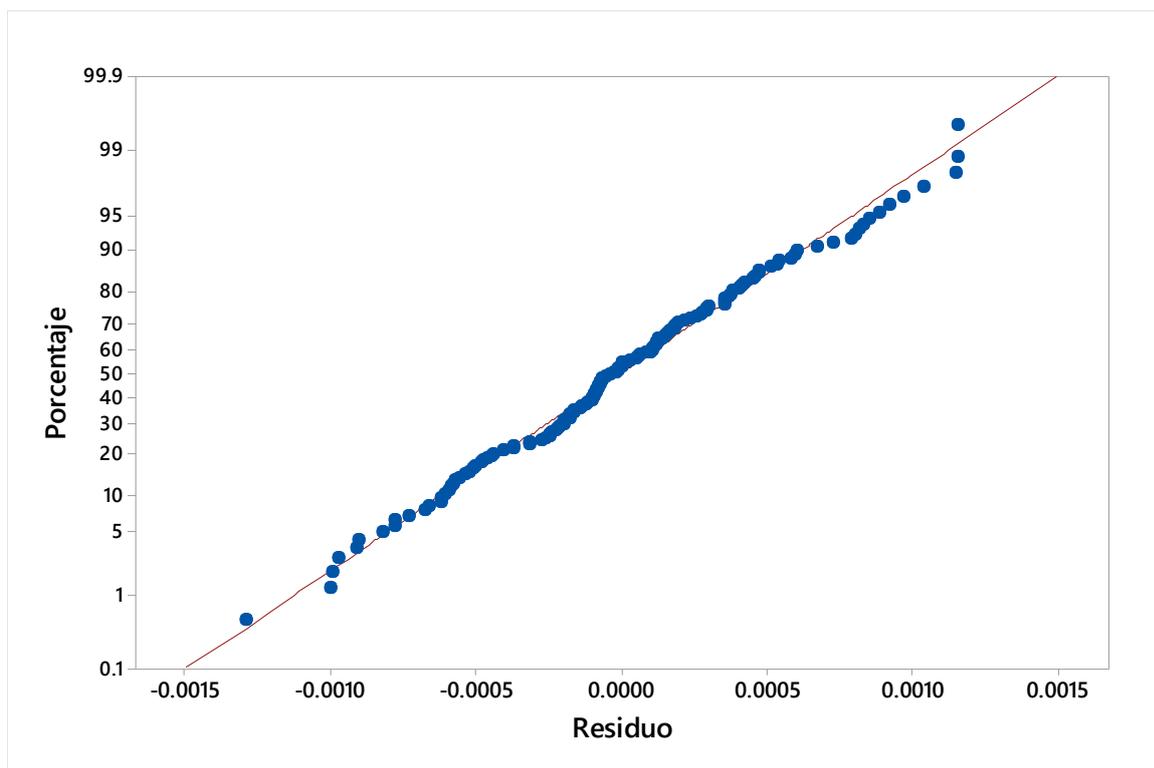
**Figura 23.** Análisis gráfico de la dispersión de los tratamientos por bloque para el tamaño de espiga.

Utilizando el modelo lineal de la expresión (18) para la variable respuesta tamaño de espiga ( $Tesp$ ) en función de los tratamientos y de los bloques, expresado por el siguiente modelo:

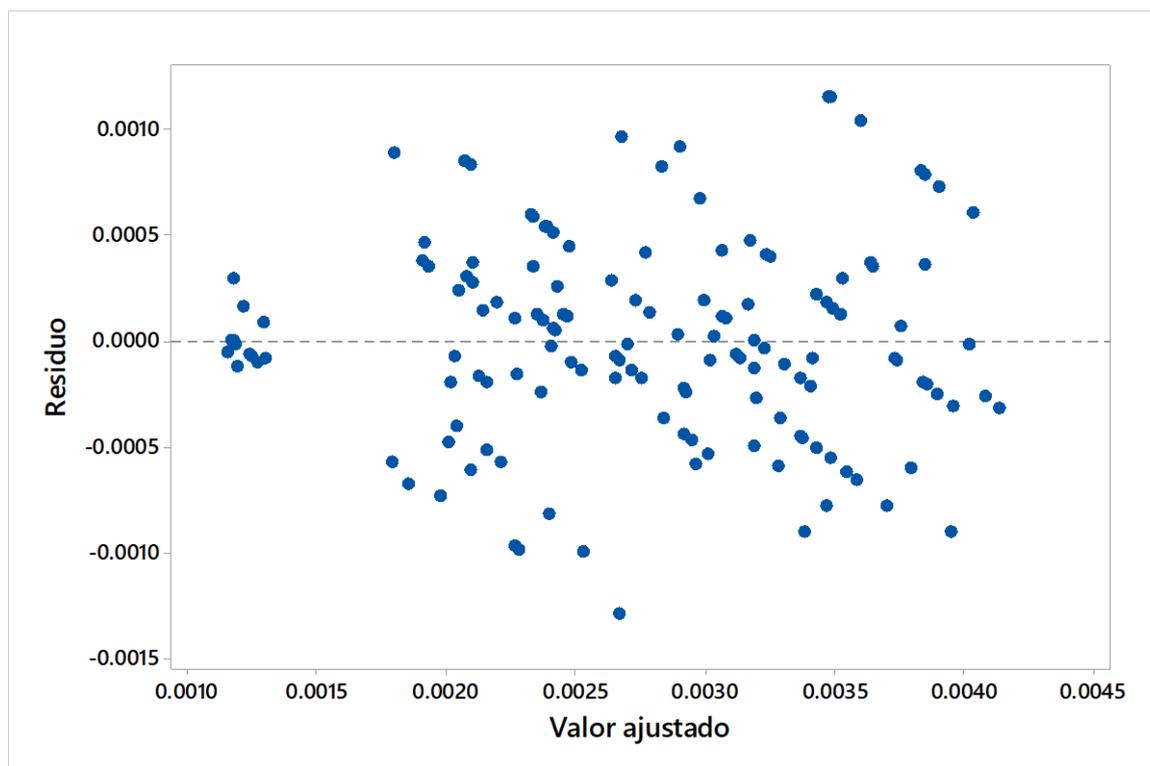
$$Tesp_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2); i = 1, \dots, 36, j = 1, 2, 3, 4$$

se procedió a efectuar el cumplimiento de los supuestos del modelo lineal para realizar un ANOVA, como la prueba, Para ello con la prueba de Shapiro Wilk según la ecuación (29) se obtuvo un  $p$ -valor = 0.3765 superior al nivel de  $\alpha = 0.05$  que

demuestra que los residuos provienen de una distribución normal. Por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos del tamaño de espiga provienen de una distribución normal, tal como se puede observar la Figura 24, en el que se muestra el comportamiento lineal de los residuos. Así mismos, utilizando la prueba de Bartlett según la ecuación (21) para el supuesto de homogeneidad de varianzas de los residuos se obtuvo un  $p\text{-valor}=0.1476$  superior al nivel de significación de  $\alpha = 0.05$ , con ello se puede indicar que se tiene evidencias estadísticas para no rechazar la  $H_0$ , por lo tanto, los residuos para la variable tamaño de espiga presentan varianzas homogéneas (Figura 25) en el que no se observa un patrón de heterogeneidad de varianzas).



**Figura 24.** Gráfico de la probabilidad normal de los residuos del tamaño de espiga.



**Figura 25.** Gráfica de residuos y valores ajustados para el tamaño de espiga.

#### 5.4.3. Prueba de comparación de medias del tamaño de espiga de los tratamientos

**Tabla 19.** Análisis de varianza para el tamaño de espiga (cm)

F.V.	GL	SC	CM	F	p-valor
Bloques	3	0.381	0.127	0.396	0.756
Tratamientos	35	97.697	2.791	8.719	<0.0001
Error	105	33.614	0.32		
Total	143	131.692			

Criterios de bondad de ajuste para el tamaño de espiga				
Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	CV
Tamaño de espiga	144	0.745	0.652	7.713

La Tabla 19 presenta los resultados del análisis de varianza para la variable tamaño de espiga frente al efecto de los tratamientos y efecto de bloques utilizando la ecuación (8) donde se observa que el coeficiente de determinación toma el valor de  $R^2=0.745$ , indicando que el 74.5% de la variabilidad total de la variable tamaño de espiga es explicado por el modelo lineal ajustado a través de los diferentes

tratamientos en estudio. Con un coeficiente de variación,  $CV=7.713\%$ , estableciendo un grado de precisión óptimo del experimento y brindan información respecto al porcentaje de relación que existe entre la variabilidad residual (no explicada por el modelo) y el promedio de los datos. Para este caso, al ser bajo el coeficiente de variabilidad la información disponible en el estudio es de mejor calidad, siendo este valor del CV bajo según la clasificación propuesta por Pimentel citado por Álvares et al., (2017).

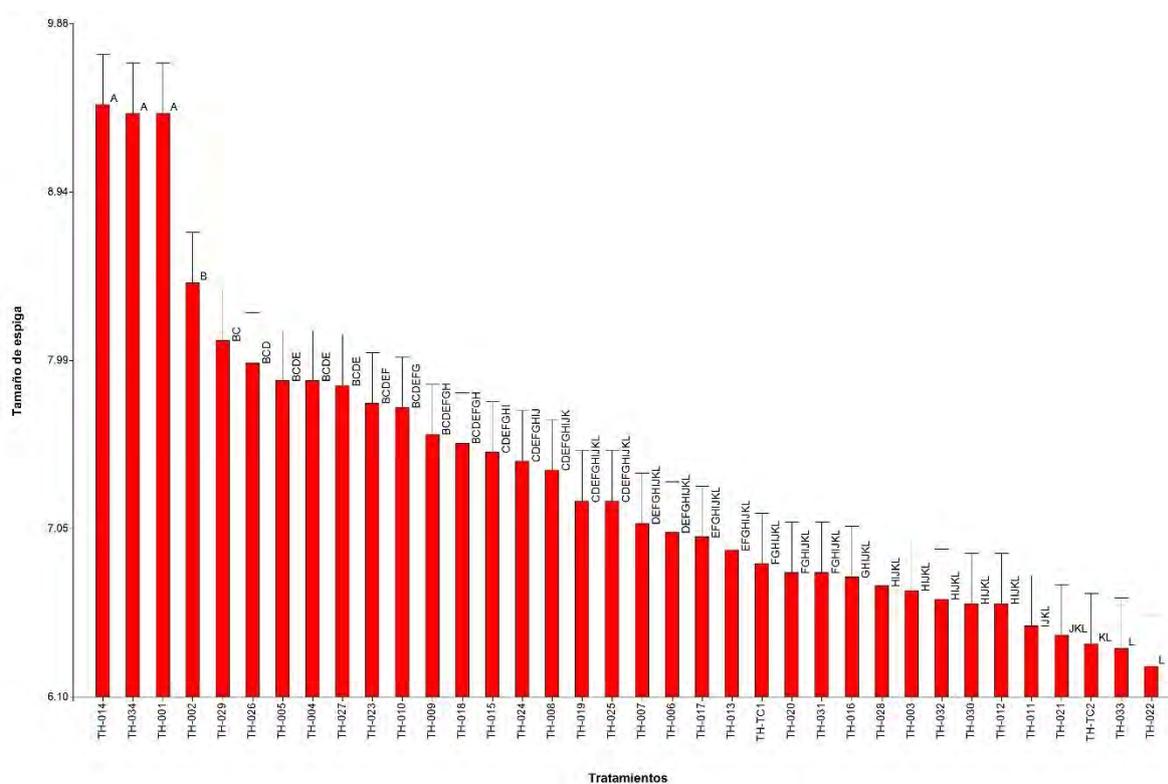
A partir de los resultados del ANOVA y con un nivel de significancia del 5%, se concluye que los tratamientos evaluados afectan la variación del tamaño de espiga, además brinda información que el factor genotipo es estadísticamente significativo ( $p<0.0001$ ) en al menos dos tratamientos. Además, los bloques no difieren de manera significativa, ya que el cuadrado medio de bloques (0.127), es pequeño en relación con el cuadrado medio del error (0.320). Existiendo bajo porcentaje de variabilidad cuya explicación se da por algunos factores que no se tomaron en cuenta en el análisis.

#### **5.4.3. Comparación de medias para el tamaño de espiga**

Al encontrar diferencia estadística significativa en el ANOVA para los tratamientos deducimos que hay al menos un genotipo que presenta el tamaño de espiga diferente a los demás, para ello a través de la prueba de comparación de medias (Tukey) en la Tabla 20 y la Figura 26 nos muestran los siguientes resultados:

- Las medias muestrales, ordenadas en forma descendente, muestran que el genotipo TH-014, tiene el mayor desarrollo de espiga (9.43 cm.), superior a los demás genotipos evaluados.
- Por consiguiente, estadísticamente diferente a los demás tratamientos; sin embargo, las diferencias muestrales observadas pueden haberse dado por azar y por tanto no ser repetibles. Por lo tanto, las observaciones de las

diferencias que resultaron estadísticamente significativas. Y estos pueden extenderse a la población. Las medias poblacionales difieren significativamente entre los genotipos y es probable plantearse que estas diferencias agronómicamente pueden influir con relevancia económica en grandes superficies de cultivo.



**Figura 26.** Comparación de tratamientos para el tamaño de espiga.

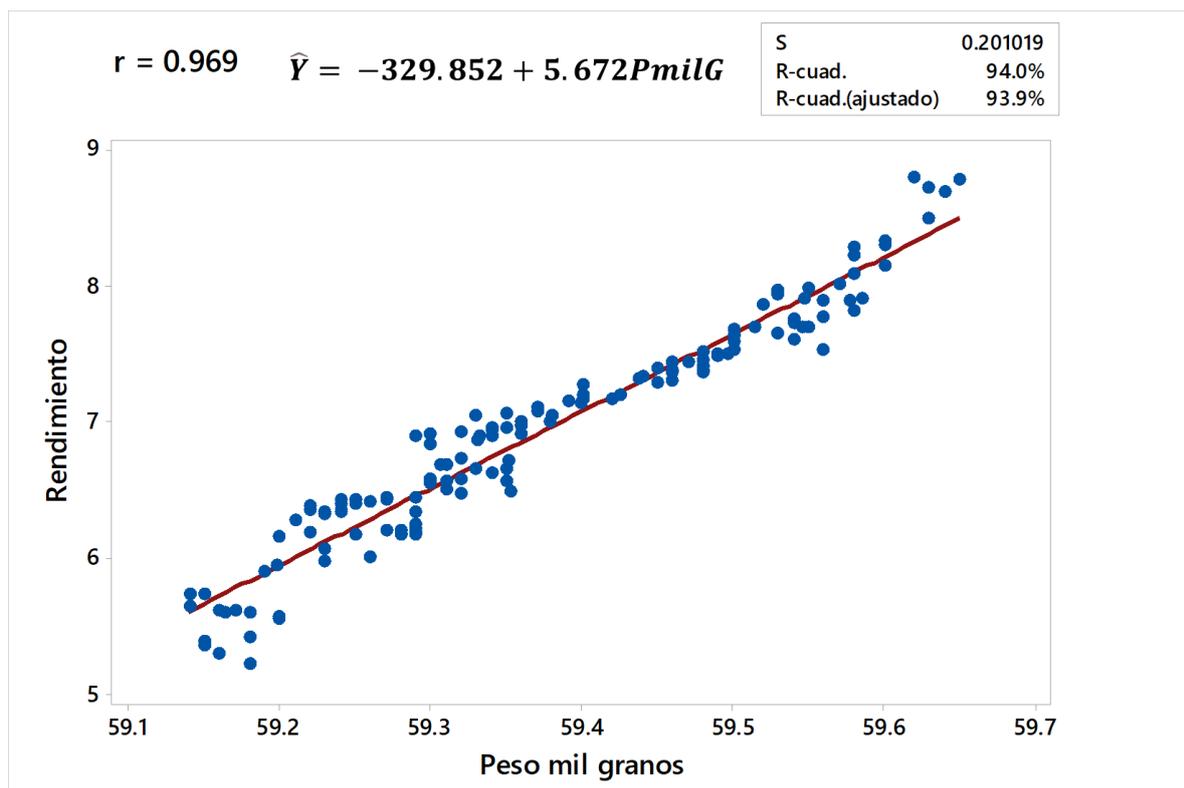
**Tabla 20.** Comparación de medias para el tamaño de espigas (cm) a través de la prueba de Tukey

Tratamiento	Tamaño de espiga	Test: Tukey Alfa=0.05 DMS=1.63443 Error: 0.3423 gl: 105
TH-014	9.43	a
TH-034	9.38	a
TH-001	9.38	a
TH-002	8.43	a b c
TH-029	8.10	a b c d
TH-026	7.98	a b c d e
TH-005	7.88	a b c d e f
TH-004	7.88	a b c d e f
TH-027	7.85	a b c d e f
TH-023	7.75	b c d e f
TH-010	7.73	c d e f
TH-009	7.58	c d e f
TH-018	7.53	c d e f
TH-015	7.48	c d e f
TH-024	7.43	c d e f
TH-008	7.38	c d e f
TH-019	7.20	c d e f
TH-025	7.20	c d e f
TH-007	7.08	c d e f
TH-006	7.03	c d e f
TH-017	7.00	c d e f
TH-013	6.93	c d e f
TH-TC1	6.85	c d e f
TH-020	6.80	c d e f
TH-031	6.80	d e f
TH-016	6.78	d e f
TH-028	6.73	d e f
TH-003	6.70	d e f
TH-032	6.65	d e f
TH-030	6.63	d e f
TH-012	6.63	d e f
TH-011	6.50	e f
TH-021	6.45	e f
TH-TC2	6.40	e f
TH-033	6.38	e f
TH-022	6.28	f

## 5.5. Determinación de la influencia del peso de mil granos en el rendimiento de trigo harinero

### 5.5.1. Análisis de regresión lineal para el peso de mil granos y el rendimiento de trigo harinero

En la Figura 27 se puede observar el comportamiento de la variable rendimiento de trigo harinero frente a la variable peso de mil granos, en el que podemos apreciar que la dispersión de los datos sigue una tendencia lineal positiva, siendo confirmada esta afirmación mediante el coeficiente de correlación de Pearson,  $r=0.969$ . Así mismo, al realizar el contraste de hipótesis para la prueba de correlación según expresión (11), se obtuvo un valor  $T= 47.0162$  (p-valor) el que está fuera de la región de aceptación, por lo tanto, se rechaza la  $H_0$  concluyendo que la correlación no es nula, en consecuencia, con estos resultados se concluye que estas dos variables presentan una correlación lineal estadísticamente significativa entre el peso de mil granos y el rendimiento de trigo harinero. También, se observa en la Figura 27 la recta ajustada expresado por  $\hat{Y} = -329.852 + 5.672 P_{milG}$ . Esta ecuación, describe la variación del rendimiento en función del peso de mil granos, explicando que hay un incremento del rendimiento en 5.672 toneladas por cada incremento del peso de mil granos.



**Figura 27.** Representación gráfica de la dispersión de datos y la recta ajustada para el rendimiento y el peso de mil granos.

**Tabla 21.** Análisis de regresión lineal aplicado al peso de mil granos y el rendimiento

Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	ECMP	s
Peso mil granos	144	0.94	0.939	0.042	0.205

Coeficientes estimados de la regresión e intervalos de confianza

Coeficiente	Estimador	Error Estándar	Intervalo de confianza (95%)		T	p-valor
			LI	LS		
Intercepto	-329.852	7.164	-344.013	-315.69	-46.043	<0.0001
Pendiente para el peso de mil granos	5.672	0.121	5.434	5.911	47.012	<0.0001

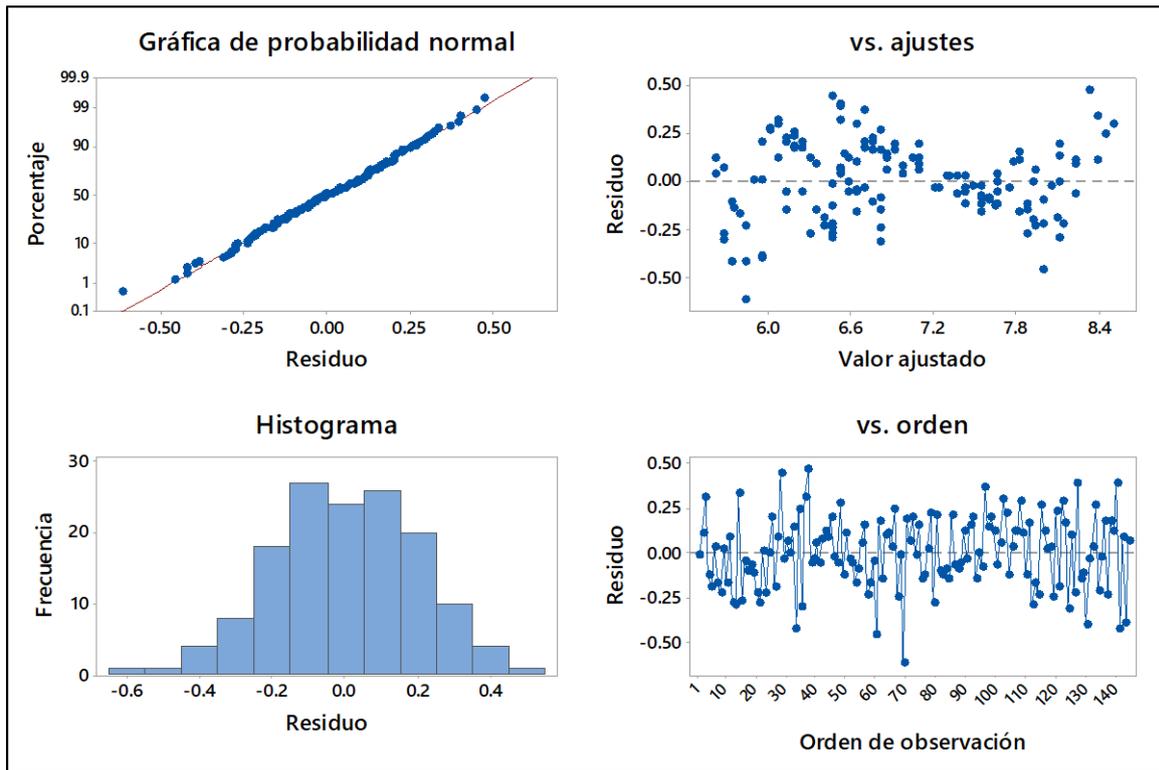
Análisis de varianza para la regresión lineal del peso de mil granos y el rendimiento

F.V.	GL	SC	CM	F	p-valor
Regresión	1	89.183	89.183	2210.16	<0.0001
Error	142	5.730	0.040		
Total	143	94.913			

Error Estándar Residual = 0.200

En la Tabla 21, se muestran los resultados del modelo de la regresión ajustada del rendimiento de trigo harinero, en función del peso de mil granos con un coeficiente de determinación  $R^2=0.940$ , es decir, el modelo ajustado está explicando que el 94% de la variabilidad presente en la variable rendimiento del trigo harinero es explicado por la variación observada en el peso de mil granos, considerando un coeficiente estimado de la pendiente asociado al peso de mil granos presenta de 5.672 toneladas con un error estándar de 0.121 toneladas y un intervalo de confianza al 95% para el coeficiente de la pendiente de la regresión [5.434, 5.911 t.]. Así mismo, con un 5% de nivel de significancia (p-valor  $<0.0001$ ) se concluye que la pendiente de la regresión asociado al peso de mil granos es distinta de cero, esto significa, que existe una relación lineal entre rendimiento de trigo harinero y peso de mil granos.

En la misma Tabla 21, se observa el análisis de varianza para la regresión donde se muestra que el modelo lineal ajustado es estadísticamente significativo para el comportamiento de la variable rendimiento frente a la variable peso de mil granos ( $p<0.0001$  con 1 gl.), con una precisión del modelo expresado por el error estándar residual de 0.200 toneladas.



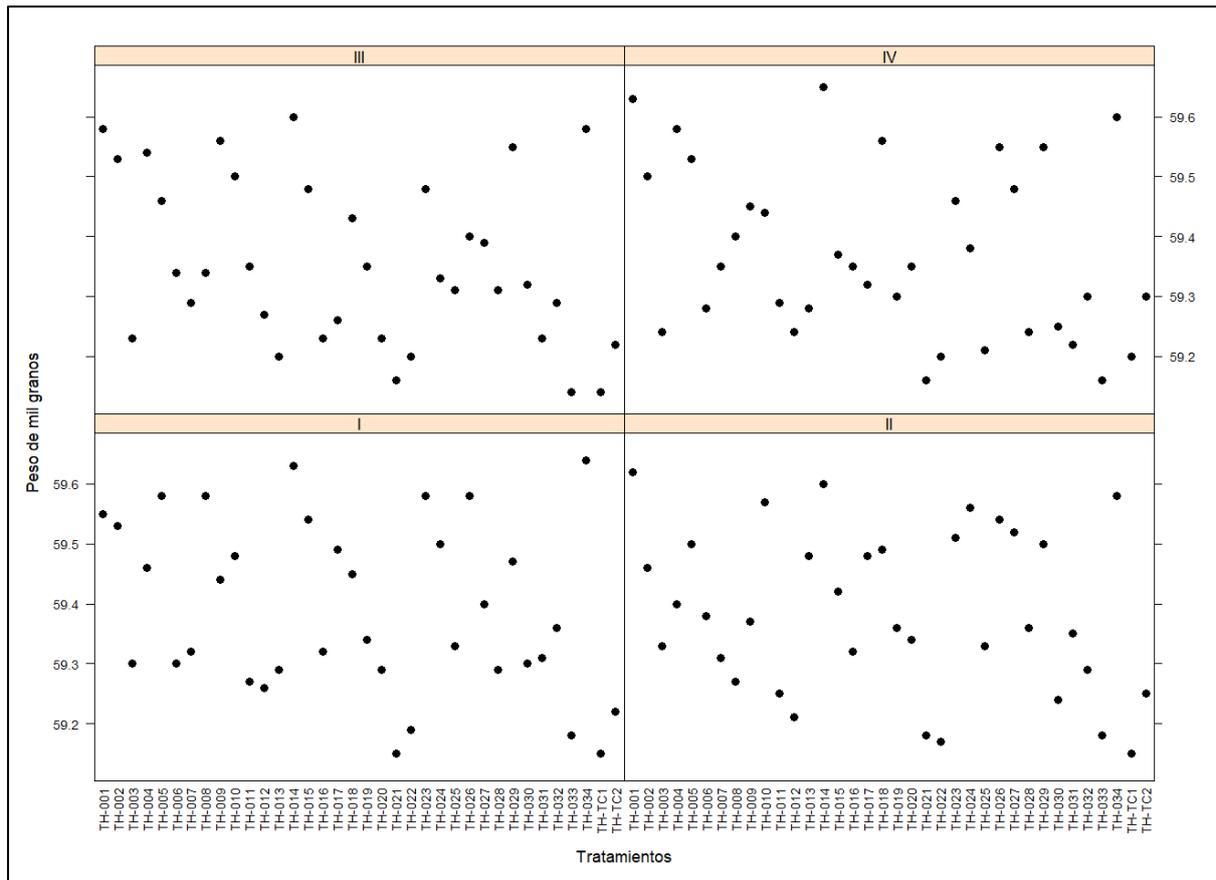
**Figura 28.** Representación gráfica de los residuos de la variable rendimiento frente al peso de mil granos.

Para validar el modelo ajustado del rendimiento de trigo harinero frente al peso de mil granos, se procedió a verificar las condiciones del modelo en función de los residuos para la variable rendimiento. En la Figura 28, se visualiza los gráficos de la probabilidad normal y el histograma en el que se observa en el primero, que los residuos siguen una tendencia lineal y en el segundo (histograma) los residuos muestran evidencias de tener una distribución simétrica, para esto procedemos a utilizar la prueba de Shapiro-Wilks y verificar la normalidad de los residuos de la variable respuesta rendimiento, estableciéndose un  $p\text{-valor}=0.9198$  siendo superior al nivel de significancia  $\alpha=0.05$ ; por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los residuos provienen de una distribución normal. Así mismo, observando los gráficos de residuos frente a los valores ajustados (homogeneidad de varianzas) y residuos versus las observaciones ordenadas

(independencia de residuos) en la Figura 28, en los dos gráficos se aprecia que los residuos siguen un comportamiento aleatorio en torno de cero dando evidencias que se cumple con la homogeneidad de varianzas, para confirmar esa afirmación visual fue utilizado la prueba de Breusch\_Pagan con un  $p\_valor=0.0569$  siendo, superior al nivel de significancia  $\alpha = 0.05$  luego, bajo las mismas condiciones experimentales se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que existe homogeneidad de varianzas de los residuos. Luego, concluimos que el modelo de la regresión lineal ajustado para rendimiento frente al peso de mil granos queda validado.

### **5.5.2. Análisis de varianza para el peso de mil granos**

Con el previo conocimiento de que el peso de mil granos influye en el rendimiento de trigo harinero, se procedió a evaluar, analizar y comparar los diferentes tratamientos de la variable peso de mil granos utilizando el diseño de bloques completos al azar. El comportamiento de las observaciones de peso de mil granos frente a los tratamientos y distribuidos en 4 bloques se muestran en la Figura 29, presenta mayor variabilidad en algunos tratamientos en estudio en los 4 bloques como es el caso de TH-004, TH-010, TH-011 y TH-027, por consiguiente, se puede indicar que existe variabilidad entre tratamientos debido a su comportamiento genético por causa de la procedencia de diversos progenitores.



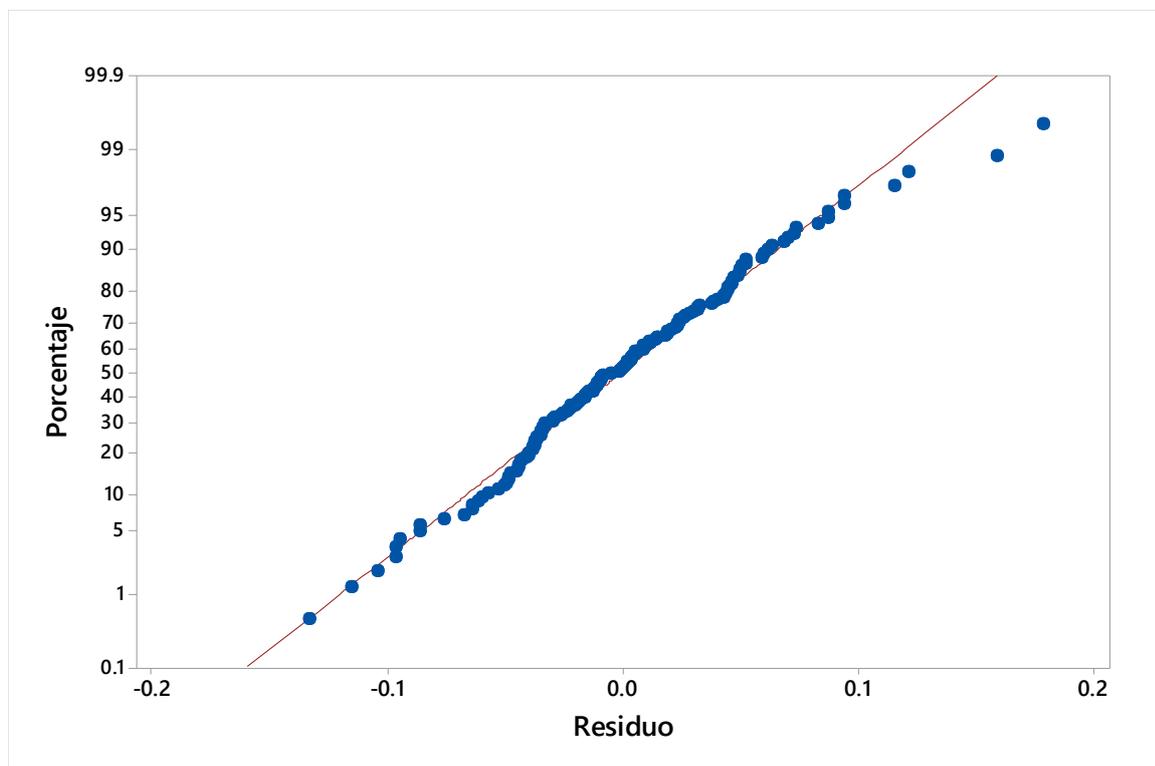
**Figura 29.** Análisis gráfico de la dispersión de los tratamientos por bloque para el peso de mil granos.

Utilizando el modelo lineal de la expresión (18) para la variable respuesta Peso de mil granos ( $PmilG$ ) en función de los tratamientos y de los bloques, expresado por el siguiente modelo:

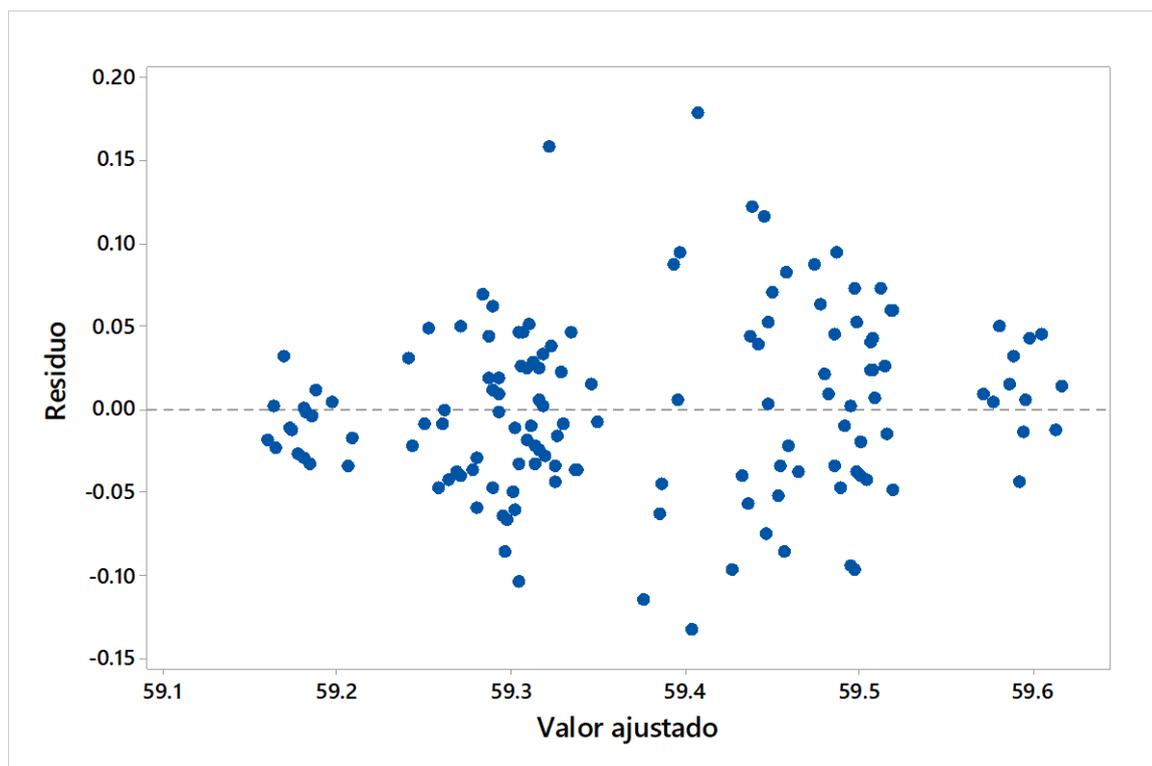
$$PmilG_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2); i = 1, \dots, 36, j = 1, 2, 3, 4$$

se procedió a efectuar el cumplimiento de los supuestos del modelo lineal para realizar un ANOVA, así se procedió a verificar la normalidad de los residuos por medio de la prueba de Shapiro-Wilks donde se obtuvo un p-valor = 0.1631 superior al nivel de  $\alpha = 0.05$  que demuestra que los residuos provienen de una distribución normal por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los residuos de las observaciones del peso de mil granos proviene de una distribución normal, tal como se puede observar la Figura 30. De la misma forma, se procedió al

uso de la prueba de Bartlett para probar la homogeneidad de varianzas de los residuos del modelo lineal establecido, en el que se obtuvo el p-valor=0.0303 bastante inferior al nivel de significación de  $\alpha = 0.05$ , ; por tanto, estas evidencias llevan a aceptar la hipótesis alterna señalando que las varianzas de los residuos no son homogéneas, en la Figura 31 se observa tal comportamiento de los residuos el que muestra un patrón no aleatorio, mismo transformando los valores de la variable respuesta peso de mil granos en términos cuadráticos no se logró homogeneizar las varianzas de los residuos, por tal motivo presentar resultados a través de un ANOVA no procede.



**Figura 30.** Gráfico de la probabilidad normal para los residuos del peso de mil granos.



**Figura 31.** Análisis de residuos y valores ajustados para el peso de mil granos.

**Tabla 22.** Criterios de bondad de ajuste para el peso de mil granos (gramos)

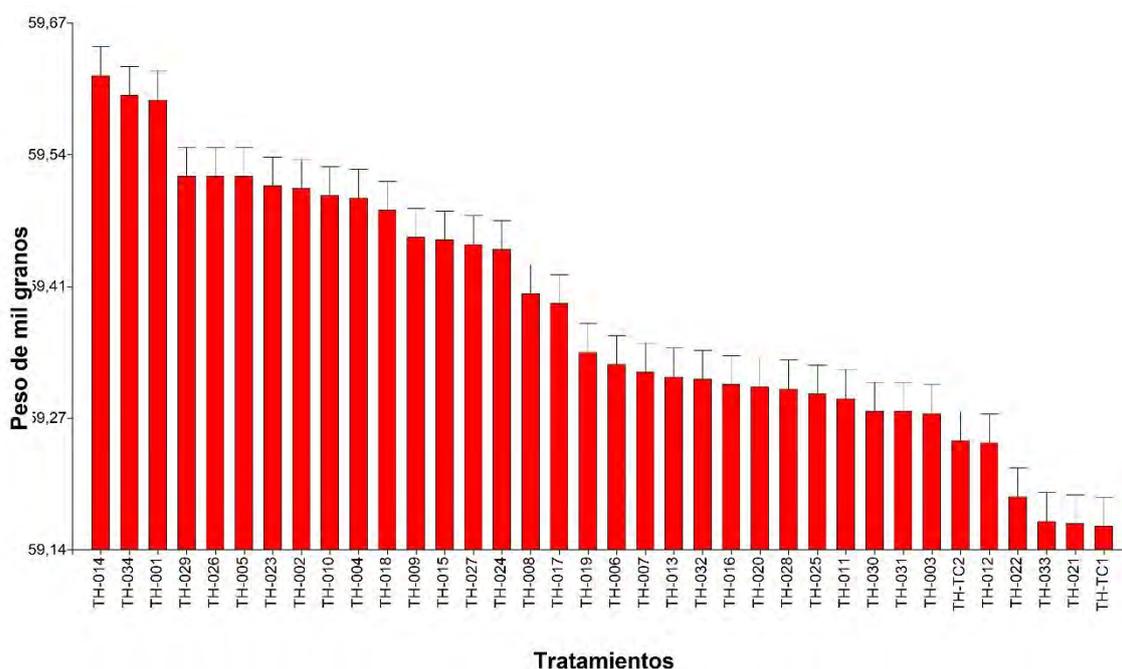
Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	CV
Peso de mil granos	144	0.868	0.820	0.099

En la Tabla 22, se muestran los coeficientes de determinación y de variación del modelo lineal ajustado de peso de mil granos frente a los efectos de tratamientos y de bloques, se observa un coeficiente de determinación ajustado,  $R^2 \text{ Aj}=0.820$ , el que explica que el 82% de la variabilidad presente en el pesos de mil granos está explicado por el modelo lineal a través del efecto de tratamientos y de los bloques, y un coeficiente de variación de 9.9% que mide el grado de precisión del experimento; es decir, la variabilidad residual existente en relación a la media de los datos, siendo este valor del CV bajo según la clasificación propuesta por Pimentel citado por Álvarez et al., (2017).

### 5.5.3. Comparación de medias de peso de mil granos de los tratamientos

La Tabla 23 y Figura 32 nos muestran los siguientes resultados:

- Las medias muestrales, ordenadas en forma descendente, muestran que el genotipo TH-014, tiene el mayor peso de mil granos ( $59.62 \pm 0.01$  gramos), superior a los demás genotipos evaluados.
- Este genotipo TH-014, por consiguiente, muestra evidencias significativas diferentes frente a los demás tratamientos, es decir las diferencias muestrales observadas pueden haberse dado por azar y por tanto no ser repetibles. Las medias poblacionales difieren entre los genotipos y es probable plantearse que estas diferencias agrónomicamente pueden influir con relevancia económica en grandes superficies de cultivo.



**Figura 32.** Comparación de medias para el peso de mil granos de los tratamientos en estudio.

**Tabla 23.** Comparación del valor promedio de los tratamientos y error estándar para el peso de mil granos (g).

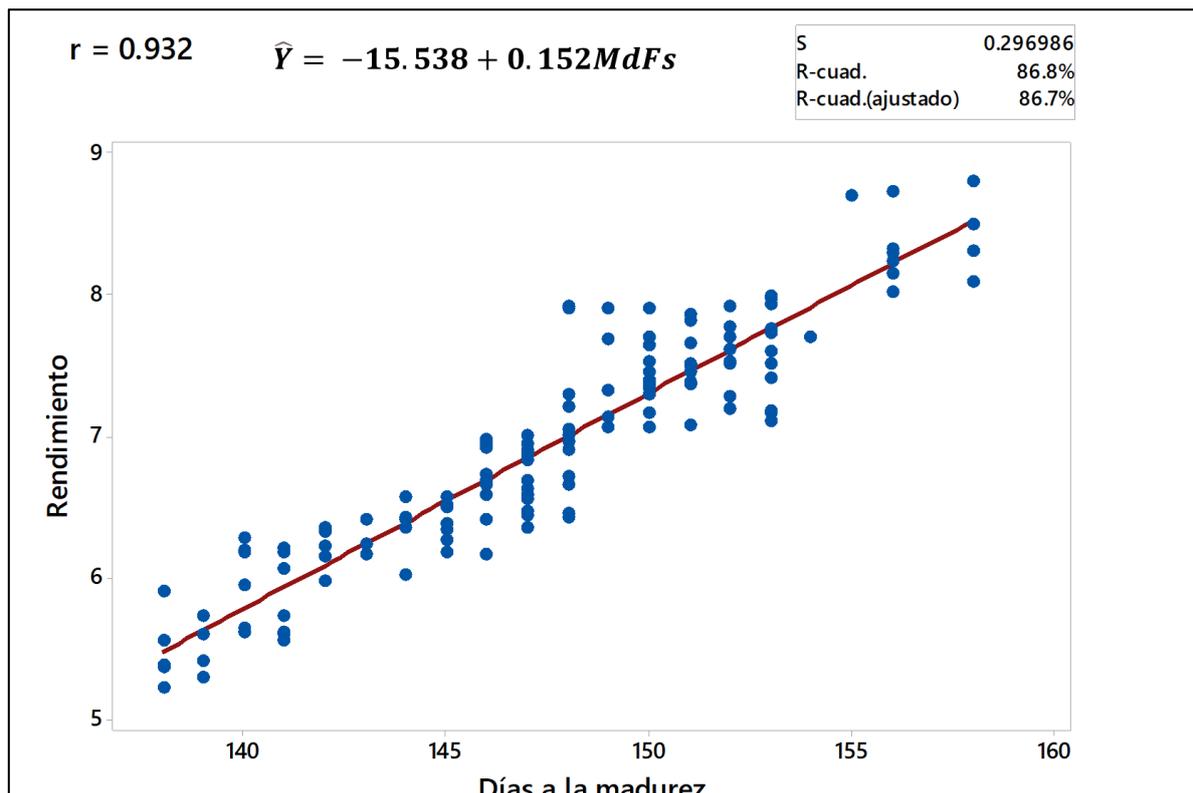
Tratamiento	Media		Error estándar
TH-014	59.62	±	0.01
TH-034	59.60	±	0.01
TH-001	59.59	±	0.02
TH-005	59.52	±	0.03
TH-026	59.52	±	0.04
TH-029	59.52	±	0.02
TH-002	59.51	±	0.02
TH-023	59.51	±	0.03
TH-010	59.50	±	0.03
TH-004	59.49	±	0.04
TH-018	59.48	±	0.03
TH-009	59.45	±	0.04
TH-015	59.45	±	0.04
TH-027	59.45	±	0.03
TH-024	59.44	±	0.05
TH-008	59.40	±	0.07
TH-017	59.39	±	0.06
TH-019	59.34	±	0.01
TH-006	59.33	±	0.02
TH-007	59.32	±	0.01
TH-013	59.31	±	0.06
TH-032	59.31	±	0.02
TH-016	59.30	±	0.03
TH-020	59.30	±	0.03
TH-025	59.30	±	0.03
TH-028	59.30	±	0.02
TH-011	59.29	±	0.02
TH-030	59.28	±	0.02
TH-031	59.28	±	0.03
TH-003	59.27	±	0.02
TH-TC2	59.25	±	0.02
TH-012	59.24	±	0.01
TH-022	59.19	±	0.01
TH-021	59.16	±	0.01
TH-033	59.16	±	0.01
TH-TC1	59.16	±	0.01

## 5.6. Determinación de la influencia de la madurez fisiológica en el rendimiento de trigo harinero

### 5.6.1. Análisis de regresión lineal para la madurez fisiológica y el rendimiento de trigo harinero

En la Figura 33, se muestra el comportamiento de la variable rendimiento frente a la variable madurez fisiológica, en el que se observa un comportamiento lineal de los datos, confirmando esta afirmación con el valor del coeficiente de correlación de Pearson,  $r=0.932$ , siendo esa relación lineal positiva y alta. Al realizar el contraste de hipótesis para la prueba de correlación, se obtiene un valor  $T= 30.5631$  ( $p$ -valor=0.0000) que está fuera de la región de aceptación; por lo tanto, se rechaza la  $H_0$  en el que la correlación poblacional es nula, en consecuencia, con estos resultados del estudio para estas dos variables se puede decir que hay una correlación lineal estadísticamente significativa entre la madurez fisiológica y el rendimiento.

También se observa en la Figura 33 la recta ajustada expresado por  $\hat{Y} = -15.538 + 0.152Mdf_s$ . Esta ecuación, describe la variación del rendimiento en función de la madurez fisiológica, explicando que hay un incremento del rendimiento en 0.152 toneladas por cada incremento de la madurez fisiológica.



**Figura 33.** Representación gráfica de la dispersión de los datos y la recta ajustada para el rendimiento y la madurez fisiológica.

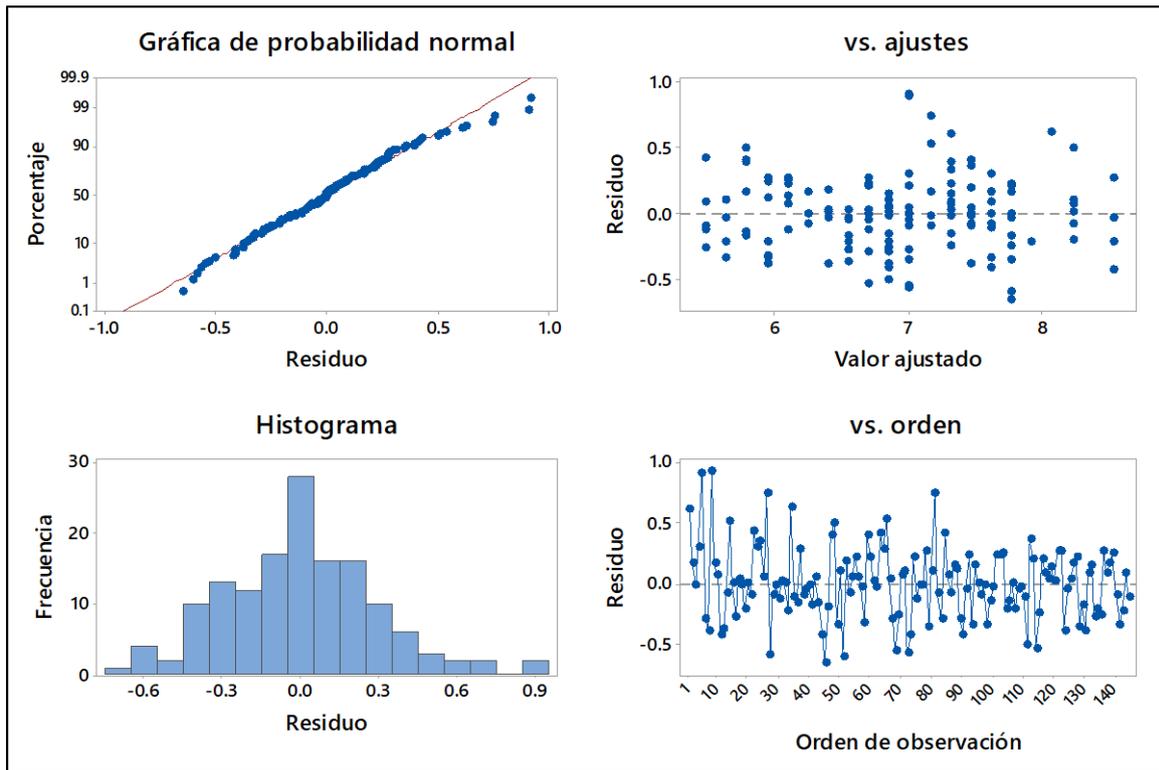
En la Tabla 24 se listan los resultados del modelo de la regresión ajustada del rendimiento del trigo harinero en función de la madurez fisiológica, con un coeficiente de determinación,  $R^2=0.868$ ; es decir, el modelo ajustado está explicando que el 86.8% de la variabilidad presente en la variable rendimiento del trigo harinero es explicado por la variación observada de la madurez fisiológica. Así mismo, el coeficiente de la pendiente asociada a la madurez fisiológica presenta un valor estimado equivalente a 0.152 toneladas con un error estándar de 0.005 toneladas y un intervalo de confianza para el coeficiente de la pendiente con un grado de confianza del 95%, [0.142, 0.162 t.]. Seguidamente, con un 5% de nivel de significancia ( $p$ -valor  $<0.0001$ ) se concluye que la pendiente de la regresión asociado a la madurez fisiológica es distinta de cero, esto significa, que existe una relación lineal entre rendimiento de trigo harinero y la madurez fisiológica. Seguidamente se

utilizó un análisis de varianza para la regresión donde el  $p$ -valor  $< 0.0001$  con 1 gl. muestra que el modelo ajustado de la regresión es estadísticamente significativo para el comportamiento de la variable rendimiento frente a la variable madurez fisiológica con un error residual de 0.297 toneladas.

**Tabla 24.** Análisis de regresión lineal aplicado a la madurez fisiológica y el rendimiento

<b>Variable</b>	<b>N</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup> Aj</b>	<b>ECMP</b>	<b>s</b>	
Madurez fisiológica	144	0.868	0.867	0.091	0.302	
Coeficientes estimados en intervalos de confianza						
<b>Coeficiente</b>	<b>Estimador</b>	<b>Error Estándar</b>	<b>Intervalo de confianza (95%)</b>		<b>T</b>	<b>p-valor</b>
			<b>LI</b>	<b>LS</b>		
Intercepto	-15.538	0.736	-16.992	-14.083	-21.113	<0.0001
Pendiente para la madurez fisiológica	0.152	0.005	0.142	0.162	30.563	<0.0001
Análisis de varianza para la regresión lineal de los días a la madurez fisiológica y el rendimiento						
<b>F.V.</b>	<b>GL</b>	<b>SC</b>	<b>CM</b>	<b>F</b>	<b>p-valor</b>	
Regresión	1	82.388	82.388	934.1	<0.0001	
Error	142	12.525	0.088			
Total	143	94.913				

Desviación. Estándar del CME = 0.297



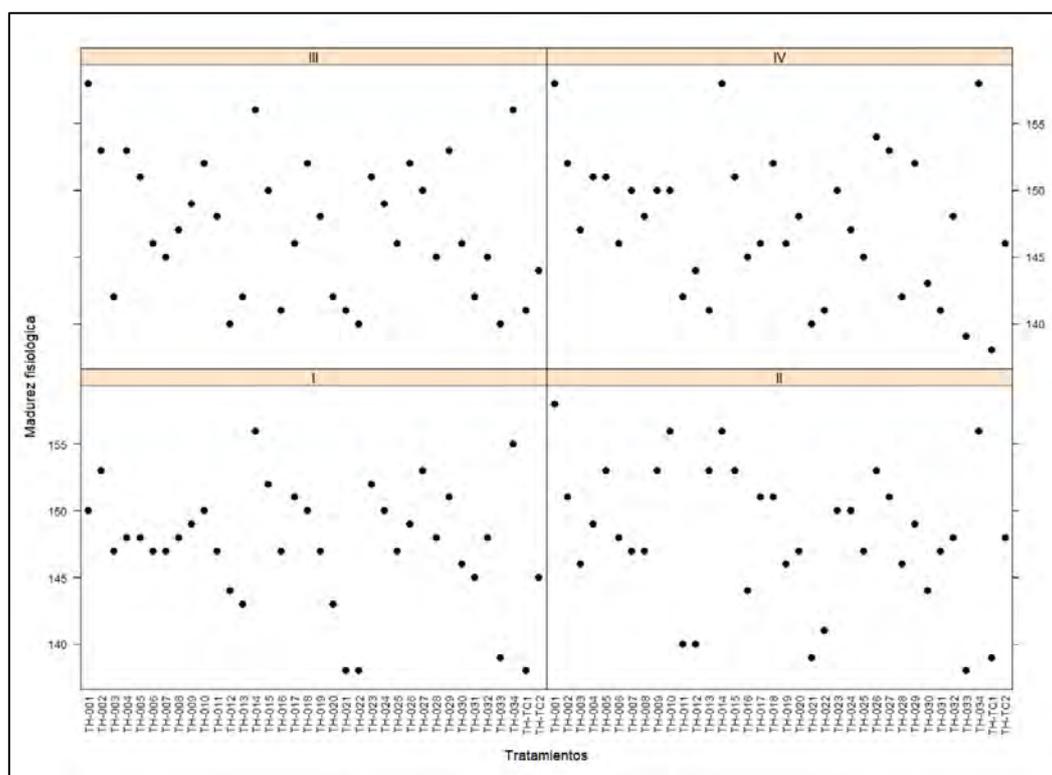
**Figura 34.** Representación gráfica de los residuos de la variable rendimiento frente a días a madurez fisiológica.

Para validar el modelo ajustado se procedió a realizar el análisis de residuos para la variable rendimiento (Figura 34), donde se utilizó la prueba de Shapiro-Wilks con la finalidad de verificar la normalidad de los residuos de la variable respuesta rendimiento, estableciéndose un  $p\text{-valor}=0.07413$ , que es superior al nivel de significancia  $\alpha=0.05$ , por tanto, no se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los residuos provienen de una distribución normal, tal comportamiento se puede observar en la Figura 34 en los gráficos de la probabilidad normal y el histograma de los residuos. Así mismo, utilizando la prueba de Breusch-Pagan para verificar la homogeneidad de varianzas de residuos para la variable respuesta rendimiento se obtuvo un  $p\text{-valor}=0.2965$  siendo, superior al nivel de significancia  $\alpha = 0.05$  consecuentemente, bajo las mismas condiciones experimentales se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala la

homogeneidad de varianzas como se puede observar un patrón aleatorio de los residuos en torno de cero en los gráficos de los residuos versus valores ajustados y residuos versus observaciones ordenadas de la Figura 33.

### 5.6.2. Análisis de varianza para la madurez fisiológica

Tomando el conocimiento previo que la madurez fisiológica influye en el rendimiento de trigo harinero, se procedió a evaluar, analizar y comparar los diferentes genotipos de la variable madurez fisiológica utilizando el diseño de bloques completos al azar. El comportamiento de las observaciones de la madurez fisiológica frente a los tratamientos y distribuidos en 4 bloques se muestran en la Figura 35, donde se puede observar que existen evidencias de presencia de variabilidad en algunos tratamientos en estudio en los 4 bloques como es el caso de TH-003, TH-004, TH-013, TH-021, TH-022 y TH-032, por consiguiente, se puede indicar que existe variabilidad entre tratamientos debido a su comportamiento genético por causa de la procedencia de diversos progenitores.

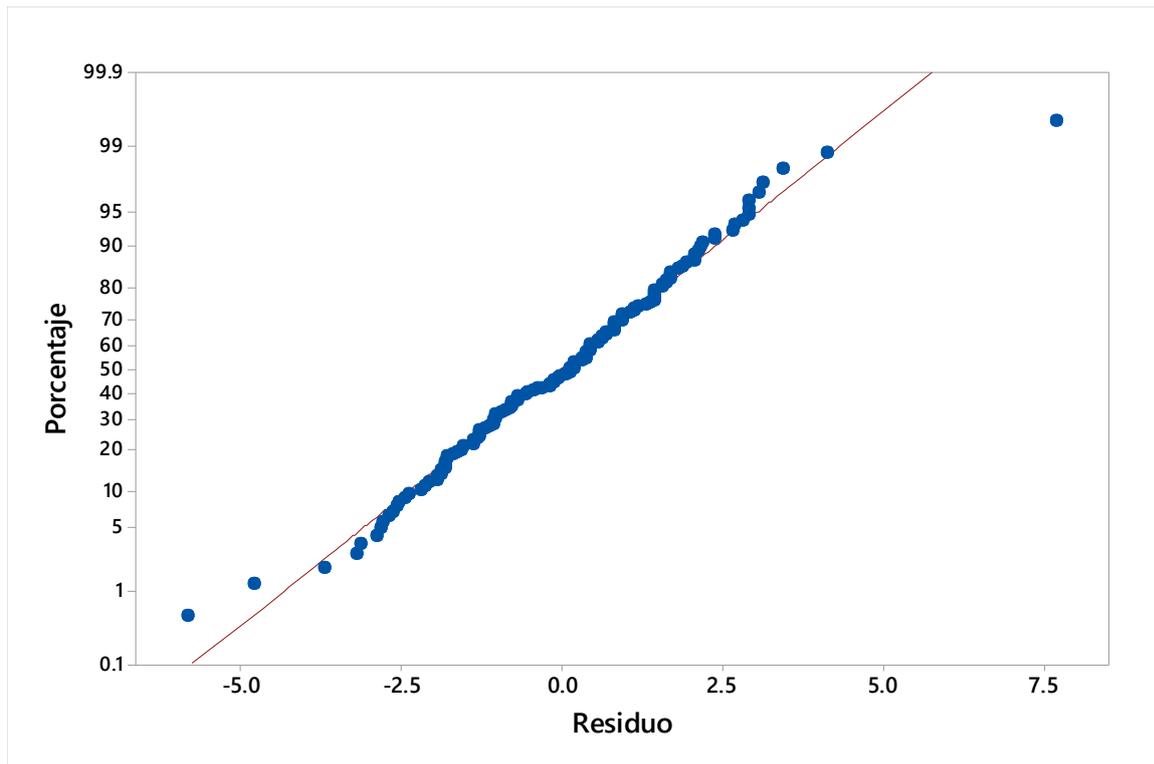


**Figura 35.** Comportamiento de la variable días a madurez fisiológica según tratamientos por bloque.

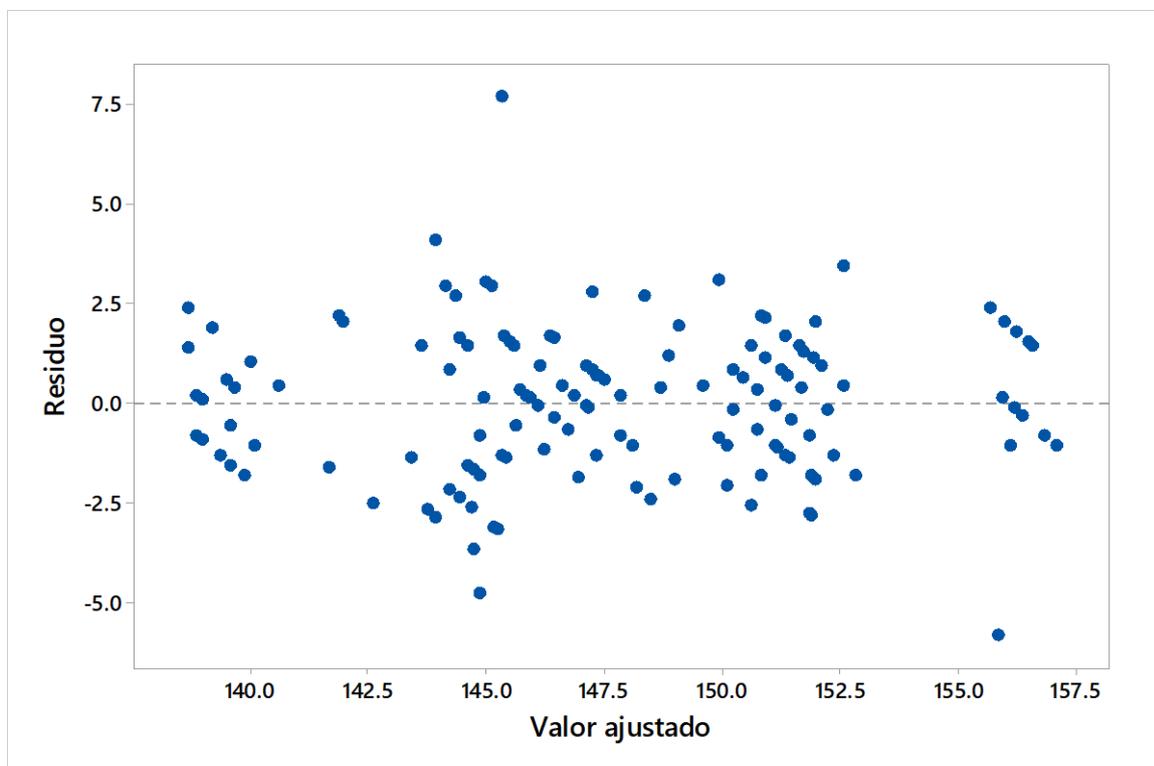
Utilizando el modelo lineal de la expresión (18) para la variable respuesta madurez fisiológica ( $MdFs$ ) en función de los tratamientos y de los bloques, expresado por el siguiente modelo:

$$MdFs_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2); i = 1, \dots, 36, j = 1, 2, 3, 4$$

se procedió a efectuar el cumplimiento de los supuestos del modelo lineal para realizar un ANOVA. La Figura 36, muestra el comportamiento lineal de los residuos de la variable respuesta madurez fisiológica, siendo confirmada esta afirmación con la prueba de Shapiro Wilks según la ecuación (20) se obtuvo un  $p$ -valor = 0.05081 superior al nivel de  $\alpha = 0.05$ , que demuestra que los residuos provienen de una distribución normal. Por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos de la madurez fisiológica provienen de una distribución normal; Así mismo la Figura 37, muestra un comportamiento aleatorio en torno de cero de los residuos de la variable en estudio siendo confirmada a través de la prueba de Bartlett según la ecuación (21) para el supuesto de homogeneidad de varianzas de los residuos, se obtuvo el  $p$ -valor = 0.05747 superior al nivel de significación de  $\alpha = 0.05$ , por lo tanto, estas evidencias nos llevan a aceptar la  $H_0$ , señalando que se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos de madurez fisiológica presentan varianzas homogéneas.



**Figura 36.** Gráfica de la probabilidad normal de los residuos de días a madurez fisiológica.



**Figura 37.** Gráfica de residuos y valores ajustados para los días a madurez fisiológica.

**Tabla 25.** Análisis de varianza para la madurez fisiológica

<b>F.V.</b>	<b>GL</b>	<b>SC</b>	<b>CM</b>	<b>F</b>	<b>p-valor</b>
Bloques	3	16.91	5.637	1.196	0.315
Tratamientos	35	3041.743	86.907	18.441	<0.0001
Error	105	494.84	4.713		
Total	143	3553.493			

<b>Criterios de bondad de ajuste para la madurez fisiológica</b>				
<b>Variable</b>	<b>N</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup> Aj</b>	<b>CV</b>
Madurez fisiológica	144	0.861	0.81	1.47

Con el modelo lineal establecido para la madurez fisiológica frente a los efectos de los tratamientos y de los bloques, la Tabla 25 muestra los resultados obtenidos del ANOVA, así se tiene que el coeficiente de determinación toma el valor  $R^2=0.861$ , por lo tanto, el modelo lineal adoptado explica que el 81% de la variabilidad total presente en la variable madurez fisiológica son explicadas por los efectos de los tratamientos que fueron expuestos en el estudio. Con un coeficiente de variación,  $CV=1.47\%$ , que brinda información que existe entre la variabilidad residual (no explicada por el modelo) y la media de los datos. Para este caso al ser bajo el CV la información disponible en el estudio es de mejor calidad según la clasificación propuesta por Pimentel citado por Álvares et al., (2017).

Así mismo, el ANOVA muestra resultados de la variabilidad entre las diferentes fuentes de variación y con un nivel de significancia del 5%, se concluye que los tratamientos evaluados afectan la variación de la madurez fisiológica, además brinda información que los tratamientos (factor genotipo) es estadísticamente significativo ( $p<0.0001$ ) en al menos dos tratamientos. Además, los bloques no difieren de manera significativa. Observando que es significativo los tratamientos en el ANOVA, se procedió a efectuar la comparación de medias múltiples para determinar el mejor comportamiento de la madurez fisiológica en base a los genotipos de trigo harinero.

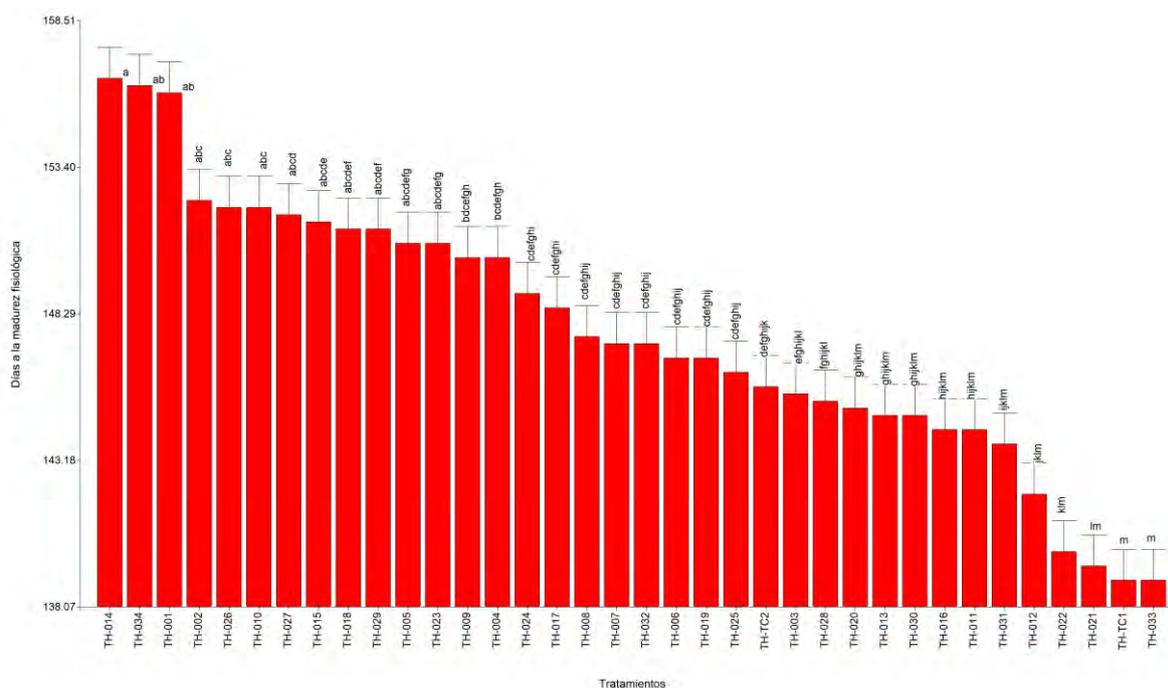
### **5.6.3. Comparación de medias de los tratamientos para la variable madurez fisiológica**

Al encontrar diferencia estadística significativa en el ANOVA para los tratamientos deducimos que existe al menos un genotipo que presenta la madurez fisiológica diferente a los demás, para ello a través de la comparación de medias aplicando la prueba de Tukey según la ecuación (32), en la Tabla 26 y Figura 38 nos muestran los siguientes resultados:

- Las medias muestrales, ordenadas en forma descendente, muestran que el genotipo TH-014, alcanza la madurez fisiológica en mayor tiempo (156 días) demorándose mayor tiempo en campo en comparación a los demás genotipos evaluados.
- Por consiguiente, el genotipo TH-014 es estadísticamente diferente a los demás tratamientos, observándose diferencias entre los genotipos con resultados estadísticamente significativa que pueden ampliarse a la población ya que se espera la existencia de estabilidad entre las relaciones halladas. Las medias poblacionales difieren significativamente entre los genotipos y es probable plantearse que estas diferencias agronómicamente pueden influir con relevancia económica en grandes superficies de cultivo.

**Tabla 26.** Comparación de medias para la madurez fisiológica a través de la prueba Tukey

Tratamiento	Madurez fisiológica		Test: Tukey Alfa=0.05 DMS=6.06492 Error: 4.7128 gl: 105
TH-014	157	a	
TH-034	156	a b	
TH-001	156	a b	
TH-002	152	a b c	
TH-026	152	a b c	
TH-010	152	a b c	
TH-027	152	a b c d	
TH-015	152	a b c d e	
TH-018	151	a b c d e f	
TH-029	151	a b c d e f	
TH-005	151	a b c d e f g	
TH-023	151	a b c d e f g	
TH-009	150	b c d e f g h	
TH-004	150	b c d e f g h	
TH-024	149	c d e f g h i	
TH-017	149	c d e f g h i	
TH-008	148	c d e f g h i j	
TH-007	147	c d e f g h i j	
TH-032	147	c d e f g h i j	
TH-006	147	c d e f g h i j	
TH-019	147	c d e f g h i j	
TH-025	146	c d e f g h i j	
TH-TC2	146	d e f g h i j k	
TH-003	146	e f g h i j k l	
TH-028	145	f g h i j k l	
TH-020	145	g h i j k l m	
TH-013	145	g h i j k l m	
TH-030	145	g h i j k l m	
TH-016	144	h i j k l m	
TH-011	144	h i j k l m	
TH-031	144	i j k l m	
TH-012	142	j k l m	
TH-022	140	k l m	
TH-021	140	l m	
TH-TC1	139	m	
TH-033	139	m	



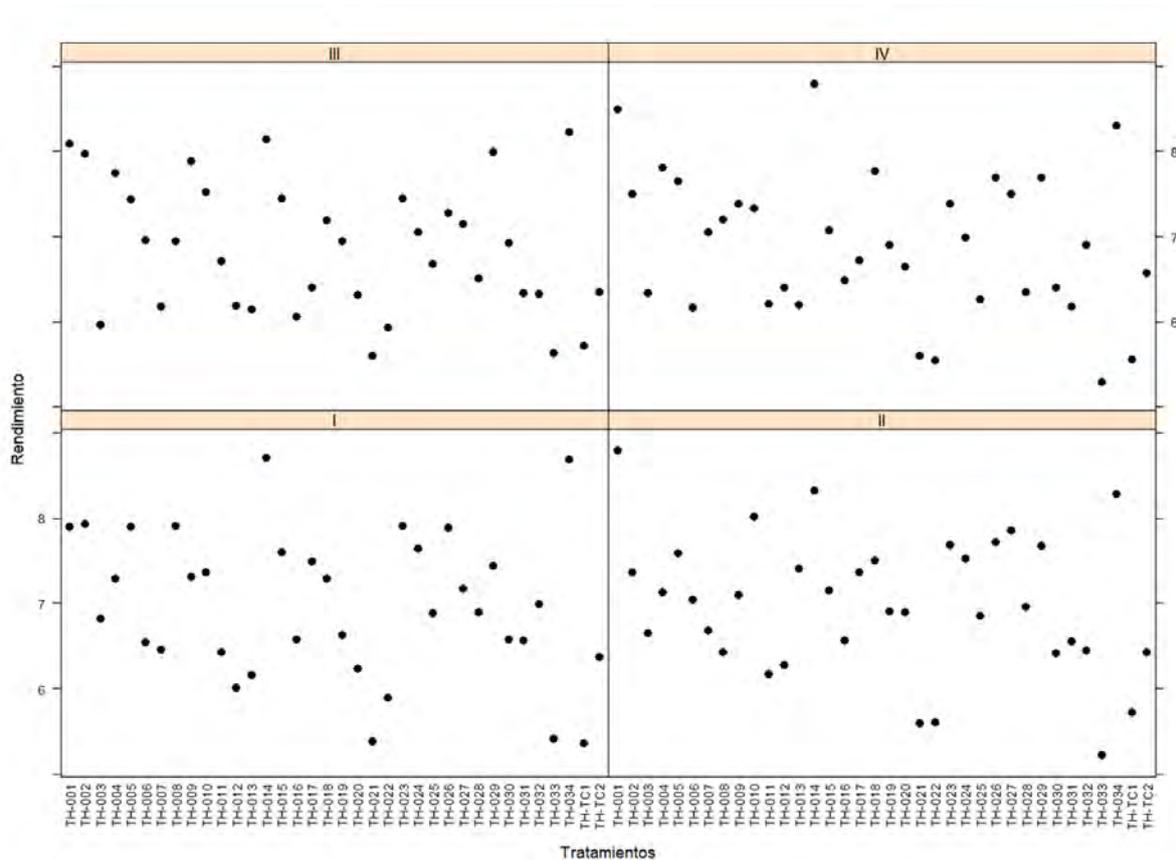
**Figura 38.** Comparación de medias para la madurez fisiológica de los tratamientos en estudio.

### 5.7. Análisis de varianza comparación de medias de los tratamientos para el rendimiento de los genotipos de trigo

Siendo el rendimiento en grano la variable respuesta principal producto de la interacción de las características cualitativas y cuantitativas de los genotipos en estudio y considerando el objetivo general de la investigación se procedió a evaluar, analizar y comparar la variable rendimiento utilizando el diseño de bloques completos al azar para determinar la variación de los 36 genotipos y encontrar aquellos que presenten un alto potencial de rendimiento para su uso en el programa de mejoramiento y el desarrollo de nuevas variedades de trigo harinero.

Para lo cual a través de la Figura 39 se muestra el comportamiento de los datos de la variable rendimiento de los 36 tratamientos distribuidos en los cuatro bloques, donde se observa que existen evidencias de variabilidad de rendimiento en algunos tratamientos en estudio entre los diferentes bloques particularmente para el caso de los tratamientos: TH-004, TH-021, TH-022, TH-028, TH-034, TH-TC1; por

consiguiente se puede indicar que esta variabilidad está asociado a la procedencia de los progenitores de los genotipos evaluados y su comportamiento genético.



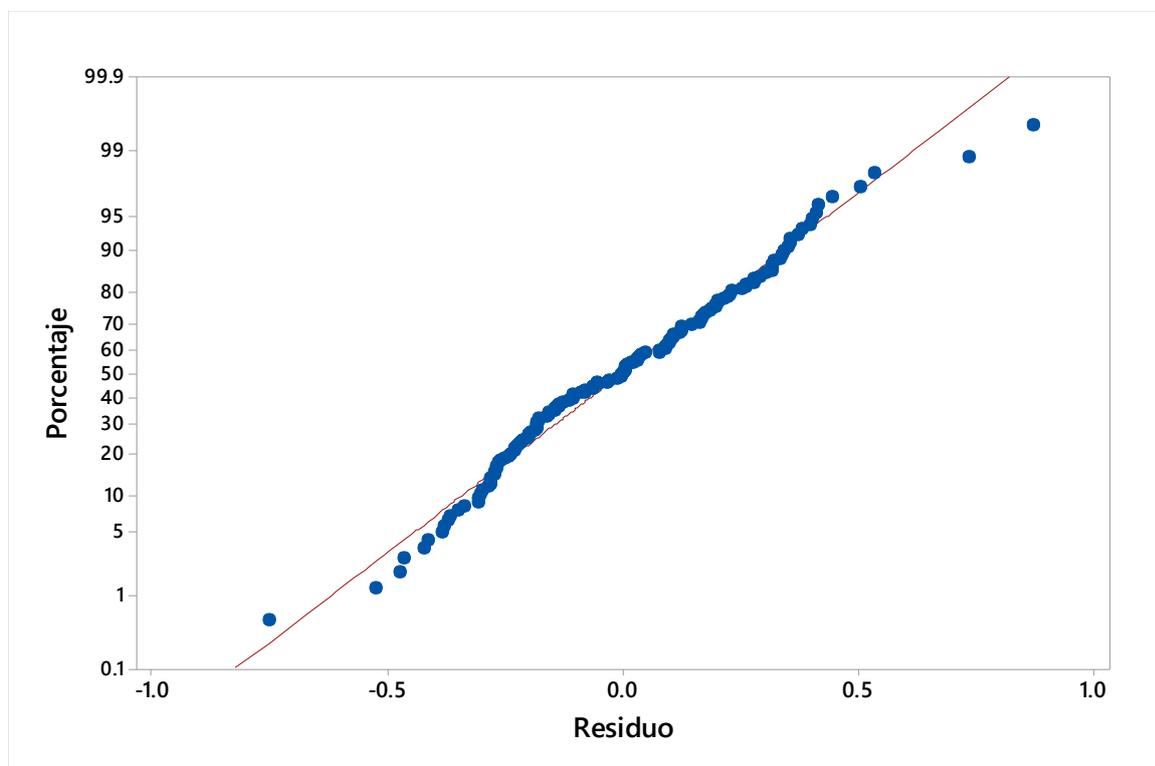
**Figura 39.** Comportamiento de la variable rendimiento según tratamientos por bloques.

Utilizando el modelo lineal de la expresión (18) para la variable rendimiento ( $Rdto$ ) en función de los tratamientos y de los bloques, expresado por el siguiente modelo:

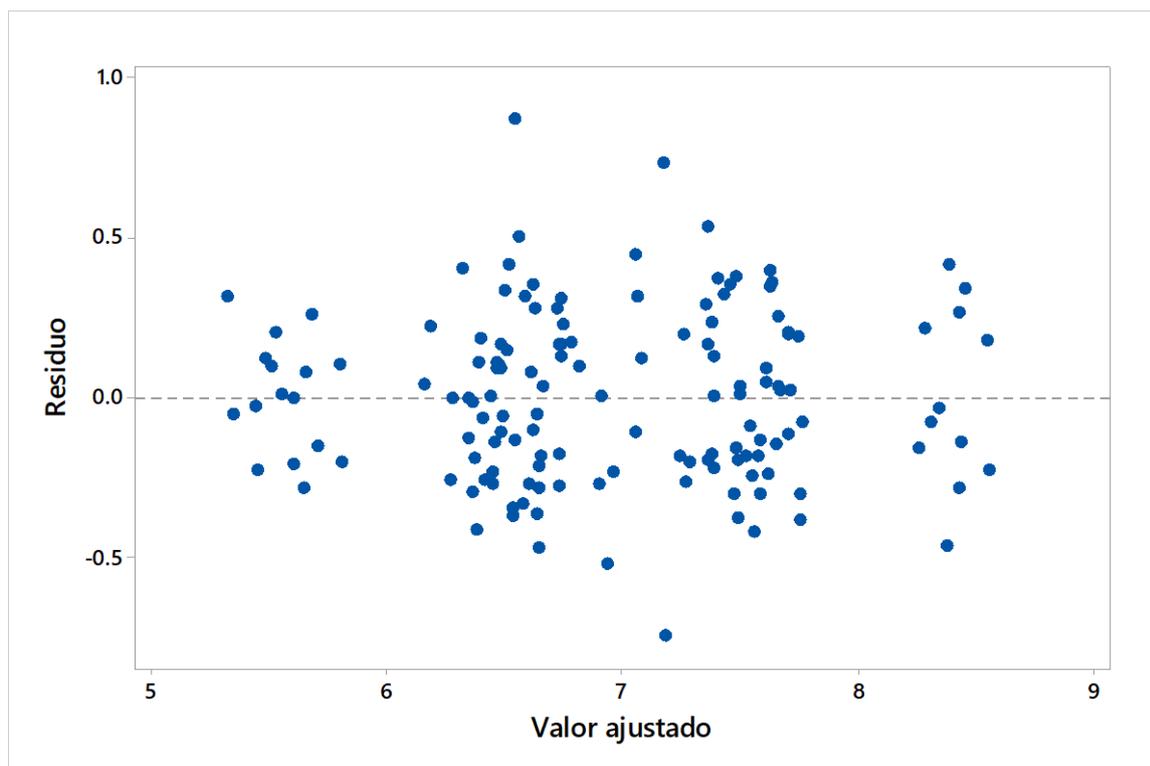
$$Rdto_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2); i = 1, \dots, 36, j = 1, 2, 3, 4$$

Se procedió a verificar el cumplimiento de los supuestos del modelo lineal para realizar un ANOVA, para ello con la prueba de Shapiro Wilk según la ecuación (20) se obtuvo un  $p$ -valor = 0.1538 superior al nivel de  $\alpha = 0.05$  que demuestra que los residuos provienen de una distribución normal. Por tanto, se tiene evidencias estadísticas para no rechazar la  $H_0$  que señala que los datos de rendimiento

proviene de una distribución normal, tal como se puede observar la Figura 40, en el que se muestra el comportamiento lineal de los residuos. Así mismos, utilizando la prueba de Bartlett según la ecuación (21) para el supuesto de homogeneidad de varianzas de los residuos se obtuvo un  $p\text{-valor}=0.6999$  superior al nivel de significación de  $\alpha = 0.05$ , con ello se puede indicar que se tiene evidencias estadísticas para no rechazar la  $H_0$ , por lo tanto, los residuos para la variable rendimiento presentan varianzas homogéneas como se observa en la Figura 41.



**Figura 40.** Gráfica de la probabilidad normal de los residuos para el rendimiento.



**Figura 41.** Gráfica de los residuos y valores ajustados para el rendimiento.

**Tabla 27.** Análisis de varianza para el rendimiento

<b>F.V.</b>	<b>GL</b>	<b>SC</b>	<b>CM</b>	<b>F</b>	<b>p-valor</b>
Bloques	3	0.4392	0.1464	1.5204	0.2135
Tratamientos	35	84.3623	2.4104	25.0303	<0.0001
Error	105	10.1112	0.0963		
Total	143	94.9128			
Criterios de bondad de ajuste para el rendimiento					
<b>Variable</b>	<b>N</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup> Aj</b>	<b>CV</b>	
Rendimiento	144	0.89	0.85	4.47	

El análisis de varianza para la variable rendimiento frente a los efectos de los tratamientos y efecto de bloques se aplicó la ecuación (18). Y en la Tabla 27 se observa el coeficiente de determinación con un valor  $R^2=0.890$  por lo tanto el modelo lineal ajustado explica que el 89% de la variabilidad total de la variable rendimiento a través de los diferentes tratamientos en estudio son explicadas conociendo las condiciones experimentales a las que fueron expuestas los tratamientos en estudio, con un coeficiente de variación  $CV=4.47\%$  que establece un óptimo grado de

precisión del experimento y nos muestra información respecto a la relación porcentual que existe entre la variabilidad residual (no explicada por el modelo) y el promedio de los datos. Para este caso, al ser bajo el coeficiente de variabilidad la información disponible en el estudio es de mejor calidad, según la clasificación propuesta por Pimentel citado por Álvares et al., (2017). A partir de los resultados del ANOVA y con un nivel de significancia del 5%, se concluye que los tratamientos evaluados presentan variación del rendimiento, además brinda información que el factor genotipo es estadísticamente significativo ( $p < 0.0001$ ) en al menos dos tratamientos. Además, los bloques no difieren de manera significativa, ya que el cuadrado medio es pequeño en relación con el cuadrado medio del error.

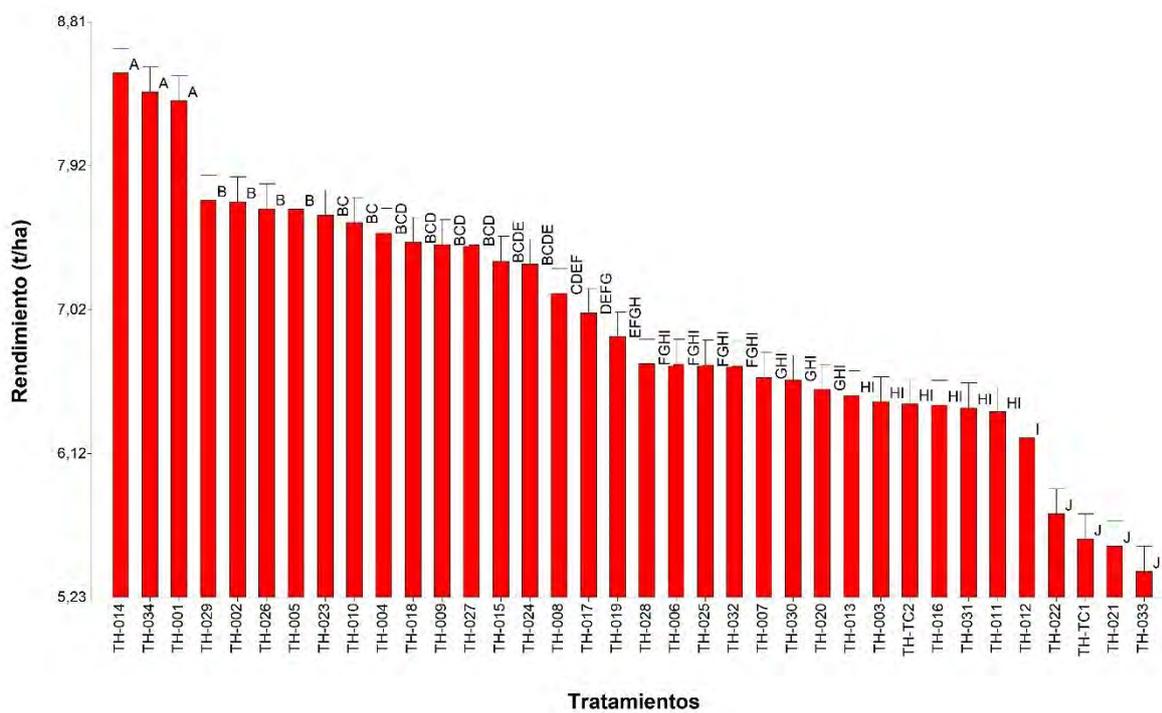
Al encontrar diferencia estadística significativa en el ANOVA para los tratamientos deducimos que hay al menos dos genotipos que presenta el rendimiento diferente, para ello a través de la comparación de medias aplicando la prueba de Tukey, en la Tabla 28 y Figura 42 nos muestran los siguientes resultados:

- Las medias muestrales, ordenadas en forma descendente, muestran que el genotipo TH-014, alcanza el mayor rendimiento (8.494 t/ha) demostrándose como el genotipo con más potencial de rendimiento en campo en comparación a los demás tratamientos evaluados.
- El genotipo TH-014, es estadísticamente diferente a los demás tratamientos; sin embargo, las diferencias muestrales observadas pueden haberse dado por azar y por tanto no ser repetibles. Por ello, se observan diferencias estadísticamente significativas. Y, estos resultados pueden ampliarse a la población ya que se espera estabilidad de las relaciones encontradas. Las medias poblacionales difieren significativamente entre los genotipos y es

probable plantearse que estas diferencias agronómicamente pueden influir con relevancia económica en grandes superficies de cultivo.

**Tabla 28.** Comparación de medias para el rendimiento (t/ha) a través de la prueba de Tukey

Tratamiento	Rendimiento t/ha.	Test: Tukey Alfa=0.05 DMS=0.86695 Error: 0.0963 gl: 105
TH-014	8.494	a
TH-034	8.377	a
TH-001	8.322	a
TH-029	7.702	b
TH-002	7.692	b
TH-026	7.649	b
TH-005	7.645	b
TH-023	7.609	b c
TH-010	7.563	b c
TH-004	7.497	b c d
TH-018	7.439	b c d
TH-009	7.426	b c d
TH-027	7.423	b c d
TH-015	7.322	b c d e
TH-024	7.303	b c d e
TH-008	7.122	c d e f
TH-017	6.998	d e f g
TH-019	6.850	e f g h
TH-028	6.682	f g h i
TH-006	6.680	f g h i
TH-025	6.675	f g h i
TH-032	6.668	f g h i
TH-007	6.597	g h i
TH-030	6.581	g h i
TH-020	6.524	g h i
TH-013	6.483	h i
TH-003	6.446	h i
TH-TC2	6.432	h i
TH-016	6.425	h i
TH-031	6.408	h i
TH-011	6.383	h i
TH-012	6.220	i
TH-022	5.747	j
TH-TC1	5.592	j
TH-021	5.547	j
TH-033	5.390	j



**Figura 42.** Comparación de medias para el rendimiento de los tratamientos en estudio.

## DISCUSIÓN DE LOS RESULTADOS

Del análisis exploratorio utilizando el análisis de clúster primeramente podemos señalar que son tres las condiciones básicas para el análisis de clúster De la Fuente (2018) i) No debe presentarse correlación entre las variables, ii) Debe ser reducido número de variables y iii) las variables no estén medidas en distintas unidades, al respecto entre las variables en estudio existe correlación significativa que no cumple con la primera condición básica y para compensar esta correlación según lo señalado por el mismo autor se ha utilizado la distancia de Mahalanobis, con respecto a la segunda condición sobre el número de variables si se cumple ya que son solo seis variables en evaluación y con respecto a la tercera condición al encontrarse las variables en diferentes unidades se ha realizado la estandarización sugerida por el mismo autor. Del análisis de clúster se obtuvieron tres grupos de los tratamientos en estudio con características similares a través del agrupamiento con el método de enlace Ward que en el estudio desarrollado por Chavez et al. (2015) con este método se dio lugar a un agrupamiento jerárquico que eliminó los problemas de encadenamiento se muestran al emplear otras técnicas. El análisis de clúster del presente estudio nos permite seleccionar a las líneas TH-001 y TH-034 como material genético muy promisorio y recomendados para el programa de mejoramiento genético y las seis variables evaluadas son representativas para este tipo de estudios, resultado similar muestra Chavez et al. (2015) en su evaluación de genotipos de maíz señalando que los caracteres más significativos en análisis de clúster son la longitud de la mazorca, del grano y el peso de 100 granos. Bajo estas consideraciones en el estudio consideramos que los días a la madurez, el número de granos por espiga, la altura planta, el tamaño de espiga, el peso de mil granos y el rendimiento fueron las variables adecuadas para el análisis de clúster.

Del cumplimiento de los supuestos del ANOVA para el diseño experimental bloques completos al azar se observa en el estudio para el caso particular de altura de planta, número de granos por espiga y peso de mil granos no se cumple el supuesto de homogeneidad de varianzas, al respecto podemos indicar que los tratamientos en estudio corresponde a líneas genéticas procedentes de cruzamientos simples y dobles de progenitores con distintas características (Tabla 8) que contribuyen genéticamente en la variabilidad del comportamiento de sus características cuantitativas asociadas al medio ambiente, este aspecto es fundamental resaltar considerando que en procesos de desarrollo de variedades a partir de líneas procedentes de cruzamientos se obtiene alta variabilidad genética. Al respecto Cubillos citado por Kohli et al.(2003) resalta que el incorporar genes particulares en el mejoramiento genético es para conseguir la heterogeneidad de los genotipos; también Rimieri (2017) señala que la concentración de genes favorables para una nueva variedad que presentarán diferencias en base a sus estructuras genéticas específicas genera la heterogeneidad en el comportamiento de los genotipos de las poblaciones segregantes hasta lograr la estabilidad de un genotipo para la selección de un nuevo cultivar. Bajo estas consideraciones en el estudio consideramos que el uso de líneas de trigo en proceso de mejoramiento ha mostrado el comportamiento de genes asociados a las características de altura planta, número de granos por espiga y peso de mil granos con menor estabilidad que generaron la falta de homogeneidad para la varianza de los residuos.

Tampoco fue posible utilizar las pruebas no paramétricas tomando en consideración lo indicado por Fernández et al., (2017) donde precisa que los procedimientos no paramétricos deben aplicarse en cuanto se desconozca la naturaleza de la distribución en la que se sustentan los datos y se realizan hipótesis sobre estas,

también su uso se da cuando se trata de dato que no son numéricos, precisando que la eficiencia de los métodos no paramétricos es alta en muestras pequeñas ( $n < 10$ ) y en el caso del presentes estudio se trabajó con 36 tratamientos y con datos numéricos.

Con respecto al uso de modelos líneas generales en el análisis de los datos estas fueron aplicadas considerando que las variables evaluadas cumplen con el supuesto de la distribución normal, según Bandera y Pérez (2018) los modelos mixtos se utiliza ampliamente en los estudios de genética en plantas cuando se cumple con el supuesto de normalidad corroborando el presente estudio, y los modelos lineales generalizados mixtos permiten ampliar los análisis con aquellas variables que por su naturaleza, no se ajustan a una distribución normal y que la distribución probabilística de la variable respuesta sea cualquiera de la familia exponencial, también cuando los datos están incompletos, desbalanceados entre otros.

Del análisis de la altura planta y el rendimiento en el trabajo de investigación los hallazgos fueron que existe una relación positiva moderada entre ambas variables estadísticamente significativa con un coeficiente de correlación de Pearson,  $r=0.589$ , el modelo de regresión ajustado explica que hay un incremento en el rendimiento en 0.096 toneladas por cada incremento en la altura de planta. Por otro lado, el modelo ajustado explica que el 34% de la variabilidad total presente en el rendimiento es debido a la variación observada en altura de planta. Así mismo, para el comportamiento de la altura de planta en función de los tratamientos y los bloques no se cumple el supuesto de homogeneidad de varianzas del modelo lineal y la comparación del valor promedio de la altura de planta y su error estándar nos muestra el testigo comercial TH-TC2 tuvo el mayor desarrollo de planta ( $98.50 \pm 0.96 \text{ cm}$ ) que muestra evidencia de ser superior en desarrollo de planta frente a los demás

tratamientos en estudio, comparado estos resultados con el estudio desarrollado por Huanca et al. (2015) sobre la variabilidad fenotípica de 208 líneas avanzadas de trigo harinero introducido del CIMMYT encontraron correlaciones moderadas entre el rendimiento con la altura de la planta ( $r=0.5$ ) con promedio de desarrollo de 54.86 cm con un máximo de 71 cm. , al respecto podemos señalar que bajo las condiciones ambientales del estudio la correlación entre estas dos variables fue relativamente superior y el desarrollo de planta es mayor, puesto que el rango de variación fue entre 82.50 a 98.50 cm lo que indica que en condiciones del ámbito de estudio y con los genotipos evaluados se logra un mayor desarrollo de planta que influye en el rendimiento.

Para el análisis del número de granos por espiga y el rendimiento existe una relación positiva alta con un coeficiente de correlación de Pearson de  $r=0.932$ , también el modelo de regresión ajustado explica que hay un incremento de 0.0873 toneladas por cada incremento en el número de granos por espiga. También el modelo explica que la proporción de la variabilidad total del rendimiento en un 86.9% en una relación lineal por la variación observada en el número de granos por espiga, Espitia Rangel et al. (2021) encontraron una correlación significativa de  $r= 0.55$  para estas dos variables resultando una mayor correlación en el presente estudio. Respecto al comportamiento del número de granos por espiga en función de los tratamientos y los bloques no se cumple el supuesto de homogeneidad de varianzas del modelo lineal y la comparación del valor promedio del número de granos por espiga y su error estándar nos muestra al genotipo TH-014 con el mayor número de granos por espiga ( $67\pm 0.25$  granos por espiga), superior a los demás genotipos evaluados, comparado estos resultados con el estudio desarrollado por Huanca et al. (2015) sobre la variabilidad fenotípica de 208 líneas avanzadas de trigo harinero introducido del CIMMYT

encontraron correlaciones altas entre el rendimiento el número de granos por espiga ( $r=0.6$ ) con promedio de 33.92 con un rango entre 24 a 48 granos por espiga; al respecto, podemos señalar que bajo las condiciones ambientales del estudio la correlación entre estas dos variables fue superior y mucho mayor puesto que el rango de variación fue entre 42 a 67 granos por espiga lo que indica que en condiciones del ámbito de estudio y con los genotipos evaluados presentan mejores condiciones que influye en el rendimiento.

Respecto a la influencia del tamaño de espiga en el rendimiento, el coeficiente de correlación de Pearson,  $r=0.860$  muestra que existe evidencias de una relación positiva alta entre estas variables y el modelo ajustado de la regresión explica que hay un incremento del rendimiento en 0.730 toneladas por cada incremento del tamaño de espiga; por otro lado, la variabilidad total del rendimiento de los tratamientos en estudio en una proporción del 74% es explicada a través de la relación lineal por la variación observada en el tamaño de espiga. Así mismo, entre tratamientos en estudio existe diferencias estadísticas significativas ( $p\text{-valor}<0.0001$ ) y a través de la comparación de medias se muestra que el genotipo TH-014, tiene el mayor desarrollo de espiga (9.43 cm.) similar a los tratamientos TH-034 y TH-001 y superiores a los demás genotipos evaluados, comparado estos resultados con el estudio desarrollado por Huanca et al. (2015) sobre la variabilidad fenotípica de 208 líneas avanzadas de trigo harinero introducido del CIMMYT encontraron correlaciones altas entre el rendimiento con la longitud de espiga, ( $r=0.48$ ) con promedio de desarrollo de 7.5 cm y un rango entre 5 a 9.5 cm., al respecto podemos señalar que bajo las condiciones ambientales del estudio la correlación entre estas dos variables fue relativamente superior y el desarrollo de la espiga es mayor puesto que el rango de variación fue entre 6.28 a 9.43 cm lo que indica que en condiciones del ámbito de

estudio y con los genotipos evaluados se logra un mayor desarrollo de espiga que influye en el rendimiento.

Para el análisis del peso de mil granos y el rendimiento la dispersión de los datos sigue una tendencia lineal positiva, confirmada mediante el coeficiente de correlación de Pearson,  $r=0.969$ , también se observa que modelo de la regresión ajustada del rendimiento de trigo harinero, en función del peso de mil granos explica la variabilidad total del rendimiento en un 94% en una relación lineal por la variación observada en el peso de mil granos, Sobre el comportamiento del peso de mil granos en función de los tratamientos y los bloques no se cumple el supuesto de homogeneidad de varianzas del modelo lineal y la comparación del valor promedio del peso de mil granos y su error estándar nos muestra al genotipo TH-014 con el mayor valor ( $59.62 \pm 0.01$  gramos), superior a los demás genotipos evaluados, comparado estos resultados con el estudio desarrollado por Huanca et al. (2015) sobre la variabilidad fenotípica de 208 líneas avanzadas de trigo harinero introducido del CIMMYT encontraron correlaciones altas entre el rendimiento y peso de mil granos con promedio de 29.84 con variación entre 21 a 42 gramos, al respecto podemos señalar que bajo las condiciones ambientales del estudio la correlación entre estas dos variables fue relativamente baja y el peso de mil granos es mayor puesto que el rango de variación fue entre 59.16 a 59.62 lo que indica que en condiciones del ámbito de estudio y con los genotipos evaluados se logra un mayor peso que influye en el rendimiento.

Los hallazgos de la investigación sobre la madurez fisiológica y su influencia en el rendimiento nos muestran un comportamiento lineal, confirmando ésta afirmación con coeficiente de correlación de Pearson,  $r=0.932$ , siendo una relación lineal positiva y alta, por otro lado, la variabilidad total del rendimiento de los tratamientos en estudio

es en una proporción del 86%, explicada a través de la relación lineal por la variación observada en la madurez fisiológica; a través de la comparación de medias, se muestra que la madurez fisiológica se registró entre los 139 a 156 días, comparado estos resultados con investigaciones similares muestran que en condiciones de Argentina Romani y Herrera (2017b), encontraron cultivares que llegaron a la madurez fisiológica entre los 133 a 154 días por consiguiente podemos indicar que bajo condiciones similares de zonas agroecológicas de producción de trigo el periodo vegetativo está influenciado tanto por las condiciones ambientales y las características genéticas de los genotipos cultivados.

Respecto al análisis de varianza para el rendimiento en el estudio los hallazgos muestran la diferencia estadística entre tratamientos ( $p\text{-valor} < 0.0001$ ) y el modelo lineal adoptado explica el 89% de la variabilidad total de la variable rendimiento son explicadas conociendo las condiciones experimentales a las que fueron expuestas los tratamientos en estudio, por otro lado, la comparación de medias muestra que el genotipo TH-014, alcanza el mayor rendimiento (8.494 t/ha) en campo en comparación a los demás tratamientos en estudio, Al respecto Espitia Rangel et al. (2021) en la evaluación de 19 genotipos de trigo en siete ambientes en México encontraron rendimiento máximo de 5.458 t/ha y Enrique et al.(2012) en el estudio desarrollado con genotipos de trigo procedentes del CIMMYT seleccionaron genotipos con rendimientos entre 8.32 y 7.53 t/ha frente a ello podemos señalar que bajo las condiciones ambientales en nuestro ámbito de estudio encontramos genotipos de trigo con mayor potencial de rendimiento que permitan incrementar la productividad por unidad de área.

## CONCLUSIONES

En base a los resultados del trabajo de investigación según los objetivos específicos del presente estudio llegamos a las siguientes conclusiones:

### **PRIMERO.**

La influencia de la altura de la planta en el rendimiento de trigo harinero es positiva y alta determinado a través del coeficiente de correlación de Pearson ( $r= 0.589$ ), en consecuencia, presenta una relación lineal significativa en el rendimiento determinándose también que el 34% de la variabilidad total presente en el rendimiento se explica por la variación presente en la altura de la planta medido a través del coeficiente de determinación  $R^2=0.347$ ; con estos resultados se puede determinar que la característica de la altura de la planta es muy importante tomar en consideración en los programas de mejoramiento genético para la selección de nuevas variedades de trigo con mayor productividad.

### **SEGUNDO.**

La influencia del número de granos por espiga en el rendimiento de trigo harinero en el presente estudio fue positiva y alta como se observa en el coeficiente de correlación de Pearson con un  $r=0.932$ ; así mismo, los resultados demuestran que el 86.95% de la variabilidad total presente en el rendimiento es explicada por la variación observada en el número de granos por espiga, con estos resultados se ratifica la importancia priorizar en los programas de mejoramiento genético para rendimiento en grano seleccionar genotipos con mayor número de granos por espiga por ser influyente para mejorar la productividad del cultivo de trigo harinero.

**TERCERO.**

El tamaño de la espiga para la selección de nuevas variedades más productivas del cultivo de trigo harinero presenta una influencia positiva y alta con un coeficiente de correlación de Pearson,  $r=0.860$ , mostrando una correlación lineal estadísticamente significativa entre el tamaño de la espiga y el rendimiento en grano cuya variabilidad total de este último es explicada por un 74% por la variabilidad del tamaño de espiga; Así mismo, en el presente estudio se ha encontrado que el tamaño de espiga es una característica bastante diferenciada entre genotipos de trigo harinero cuya importancia de considerar en los programas de mejoramiento son de mucha utilidad para seleccionar variedades de trigo harinero con mayor productividad.

**CUARTO.**

La influencia del peso de mil granos en el rendimiento de trigo harinero es positiva y alta determinado a través del coeficiente de correlación de Pearson ( $r= 0.969$ ), en consecuencia, presenta una relación lineal significativa en el rendimiento donde el 94% de la variabilidad total presente en el rendimiento se explica por la variación presente en el peso de mil granos a través del coeficiente de determinación  $R^2=0.940$ ; con estos resultados se puede determinar que cuanto mayor es el peso de mil granos de una nueva variedad que sea seleccionada repercute altamente en obtener mejores rendimientos de trigo harinero por tanto esta características es necesario considerar en la selección por rendimiento en el desarrollo de nuevas variedades.

**QUINTO.**

La influencia de la madurez fisiológica en el rendimiento de trigo harinero en el presente estudio fue positiva y alta como se observa en el coeficiente de

correlación de Pearson con un  $r=0.932$ ; así mismo, los resultados demuestran que el 86.80% de la variabilidad total presente en el rendimiento es explicada por la variación observada en el tiempo que alcanza la madurez fisiológica los genotipos en estudio entre 139 a 157 días desde la siembra lo que implica que esta característica para la selección de nuevas variedades de trigo harinero en los programas de mejoramiento debe ser analizada principalmente considerando las zonas de producción y los sistemas de producción, por ser bastante influyente para mejorar la productividad del cultivo de trigo harinero.

### **Conclusión del objetivo general de la investigación**

Asumiendo el planteamiento del objetivo general del estudio se concluye con las evidencias de los resultados que el diseño experimental utilizado de bloques completos al azar con cuatro repeticiones utilizado en la investigación fue eficiente para evaluar y determinar la variación del rendimiento de los 36 genotipos utilizados, puesto que a través de las pruebas estadísticas utilizadas para el análisis, y el modelo lineal adoptado explica con el porcentaje superior al 75% la variabilidad existente entre los genotipos en estudio conociendo las condiciones experimentales a las que fueron expuestas siendo los coeficientes de variabilidad inferiores al 10%, la información utilizada para los análisis fue de buena calidad, no obstante existe variabilidad que es explicado por algunos factores que no fueron tomados en cuenta en los análisis entre ellos la procedencia de los progenitores causantes del comportamiento genético evaluado que corresponde a líneas procedentes de cruzamientos en proceso de selección que contribuyeron en la variabilidad del comportamiento de sus características asociados al medio ambiente ya que las estructuras genéticas de cada tratamiento generó la heterogeneidad en el comportamiento de los mismos

que se reflejaron en el supuesto de homogeneidad de varianzas para el caso particular de altura de planta, número de granos por espiga y peso de mil granos donde los resultados demostraron que existió evidencias para señalar que las varianzas de los residuos no fueron homogéneas.

Finalmente, entre los tratamientos en estudio destacaron con la mejor manifestación de sus características genéticas para las variables número de granos por espiga, tamaño de espiga, peso de mil granos en condiciones del ámbito de estudio los genotipos TH-014, TH-034 y TH-001 que superaron a las variedades comerciales utilizadas como testigos de comparación, con rendimientos de 8.494, 8.377 y 8.322 t/ha respectivamente con periodos de madurez fisiológica de 156 días considerada como semi precoz para altitudes superiores a los 3000 msnm. Considerando a estos genotipos como promisorios para el desarrollo de nuevas variedades comerciales de trigo harinero y/o su uso como progenitores en los planes de mejoramiento de trigo.

## **RECOMENDACIONES**

### **PRIMERO**

Orientado a las entidades que ejecutan programas de mejoramiento genético y el desarrollo de nuevas variedades se debe desarrollar investigaciones que permitan validar la eficiencia de los diferentes diseños experimentales con pruebas experimentales en las condiciones agroecológicas y geográficas de la zona andina en nuestra región Cusco considerando la diversidad topográfica a fin generar recomendaciones prácticas a los centros de investigación para estudios agronómicos con diseños experimentales apropiados al número de tratamientos y características topográficas de ubicación de los experimentos.

**SEGUNDO**

A los centros de investigación de las ciencias biológica, en investigaciones con cultivos como variables biológicas procedentes de recombinaciones genéticas generadas por cruzamientos se recomienda desarrollar trabajos de investigación y aplicar análisis con métodos más intensivos utilizando modelos robustos que permita llevar en consideración la variabilidad genética y la heterogeneidad de las varianzas que esté expresado en los residuos, entre ellos los modelos aditivos generalizados y los modelos aditivos generalizados mixtos que llevan en cuenta la heterogeneidad de los residuos para lograr resultados en los procedimientos de análisis.

**TERCERO**

A los centros de investigación se recomienda utilizar los genotipos seleccionados en ensayos multi-ambientes para determinar los índices ambientales y los rangos de adaptación que contribuyan a la generación de nuevas variedades comerciales de trigo harinero.

## BIBLIOGRAFÍA

- Abbate, P. E. (2017). Manual del Cultivo del Trigo Instituto Internacional de Nutrición de Plantas. *Manual Del Cultivo de Trigo, October 2017*, 33–52.
- Alaminos, A., Francés, F., Penalva, C., & Santacreu, O. (2015). *Análisis multivariante para las Ciencias Sociales I Índices de distancias, conglomerados y análisis factorial* (PYDLOS EDI).
- Álvares, G., Rodrigues, L., Solveira, M., Botelho, C., Rodrigues, C., & Avelar, F. (2017). *Propuesta para clasificación de coeficiente de variación en experimentos de mejoramiento genético de café Introducción*. Universidad Federal de Lavras.
- Arias. (2012). *El proyecto de investigación, Introducción a la metodología científica* (E. E. C.A. (ed.); sexta edic).
- Balzarini, M., Di Rienzo, J., Tablada, M., Gonzalez, L., Bruno, C., Córdoba, M., Robledo, W., & Casanoves, F. (2011). Estadística y biometría. In *Estadística y Biometría Ilustraciones del Uso de InfoStat en Problemas de Agronomía*. [http://www.agro.unc.edu.ar/~mcia/archivos/Estadistica y Biometria.pdf](http://www.agro.unc.edu.ar/~mcia/archivos/Estadistica%20y%20Biometria.pdf)
- Bandera, E., & Pérez, L. (2018). Los modelos lineales generalizados mixtos. Su aplicación en el mejoramiento de plantas. *Cultivos Tropicales*, 39(1), 127–133.
- Camarena, F., Chura, J., & Blas, R. (2014). Mejoramiento genético y biotecnológico de plantas. In Universidad Nacional Agraria La Molina (Ed.), *Perú: Vol. I* (Primero). <https://doi.org/10.1021/pr3003326>
- Castañeda-Saucedo, M. C., López-Castañeda, C., Colinas-De León, M. T. B., Moreno, J. C. M., & Hernández-Livera, A. (2009). Rendimiento y calidad de la semilla de cebada y trigo en campo e invernadero. *Interciencia*, 34(4), 286–292.
- Chavez, D., Ileana, M., Varela, M., & Lianne, F. (2015). Utilización del análisis de

- clusters con variables mixtas en la selección de genotipos de maíz. *Revista Investigación Operacional*, 30(3), 209–216.
- CIMMYT. (1986). *Instructivo para el manejo y registro de resultados de los ensayos internacionales del Programa de trigo del CIMMYT* (Primera).
- CIMMYT. (2012). *Manual de determinación de rendimiento* (p. 42).
- Cuadras, C. (2019). *Nuevos Métodos de Análisis Multivariable* (CMC Editio).
- De la Fuente, S. (2018a). Análisis de Conglomerados (Análisis Cluster). In *Técnicas de análisis multivariante de datos* (p. 14).
- De la Fuente, S. (2018b). Análisis de Conglomerados (Análisis Cluster). In *Técnicas de análisis multivariante de datos*.
- Di Rienzo, J., Casanovez, F., Gonzales, L. A., Tablada, E., Días, M. del P., Robledo, C., & Balzarini, M. (2008). *Estadística para las ciencias agropecuarias*. 372.
- Di Rienzo J.A, Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., R. C. W. (2008). Manual del Usuario InfoStat Software Estadístico. *Infostat*, 53(November 2015), 336.
- DL-01060-2008.pdf. (2008, June). *MINAGRI*, 4.
- Enrique, J., Magallanes, V., Pérez, D. J., Devi, S. R., & Melgarejo, A. B. (2012). Análisis de 20 genotipos de trigo harinero en el Valle del Yaqui , Sonora \*  
Analysis of 20 genotypes of bread wheat in the Yaqui Valley , Sonora Resumen. *Revista Mexicana de Ciencias Agrícolas*, 3, 1521–1534.
- Espitia Rangel, E., Martinez Cruz, E., Villaseñor Mir, H. E., Santa Rosa, R. H., Limón Ortega, A., & Lozano Grande, A. (2021). Variabilidad genética y criterios de selección del rendimiento y los componentes en trigos harineros de temporal. *Revista Mexicana de Ciencias Agrícolas*, 12(2), 305–315.  
<https://doi.org/10.29312/remexca.v12i2.2787>

- Estrada, R., Gutierrez, J., & Gonza, V. (2020). *Protocolo para la generación de variedades, mantenimiento e incremento de semilla de trigo y cebada* (E. Alvarez (Ed.); Primera).
- FAO. (1987). Manual para el mejoramiento del manejo poscosecha de frutas y hortalizas - Las frutas y hortalizas frescas como productos perecibles. In *Tecnología Postcosecha* (Tecnología Postcosecha 6).
- FAO. (2014). *Producción de cultivos*.
- FAO. (2020). Perspectivas de cosechas y situación alimentaria, No. 1, marzo 2020. In *Perspectivas de cosechas y situación alimentaria, No. 1, marzo 2020* (p. 48). <https://doi.org/10.4060/ca8032es>
- Fernández, C., & Baptista, P. (2014). *Metodología de la Investigación* (S. A. D. C. . McGRAW-HILL / INTERAMERICANA EDITORES (Ed.); sexta Edic, Issue 1).
- Fernández, R., Trapero Casas, A., & Domínguez Jimenz, J. (2017). *Experimentación Agraria* (D. de Santos (Ed.)).
- Gobierno Regional de Cusco. (2016). Estudio de diagnóstico y zonificación para el tratamiento de la demarcación territorial de la provincia Tahuamanu-Volumen I. In *Sub gerencia de Acondicionamiento Territorial* (p. 302).
- González, A. (2001). *Estudio de caracteres fenológicos, agronómicos, morfológicos y fisiológicos en relación con la tolerancia al estrés hídrico en cebada*. Universidad Complutense de Madrid.
- Gutierrez, H., & De la Vara, R. (2008). *www.FreeLibros.org* (M. Interamericana (Ed.); Segunda).
- Huanca, N., Butron, R., Marza, F., & Quispe, F. (2015). Estimación de la variabilidad fenotípica de trigo harinero para componentes de productividad. *Instituto Nacional de Innovación Agropecuaria y Forestal*, 77–83.

- INEI. (2018). Compendio estadístico del Perú 2018. In *Instituto Nacional de Estadística e Informática* (p. 36).  
<https://doi.org/10.1017/CBO9781107415324.004>
- Juarez, J. (2019, September 11). El mercado mundial y nacional del trigo. *El Economista*, 20.
- Kohli, M. M., Ackermann, M. D., & Castro, M. (2003). Estrategias y metodologías utilizadas en el mejoramiento de trigo: un enfoque multidisciplinario. In M. Mohan, M. Díaz, & M. Castro (Eds.), *Seminario Internacional INIA La Estanzuela, Colonia, Uruguay*.
- Lara, A. (1998). *Diseños completamente aleatorizados* (U. de Granada (Ed.); pp. 1–35).  
<http://www.ugr.es/~bioestad/guiaspss/practica7/ArchivosAdjuntos/EfectosFijos.pdf>
- Limaylla, R. (U. N. H. V. (2016). *Ensayo de identificación, adaptación y eficiencia de 10 genotipos de trigo harinero (Triticum aestivum L.) en condiciones de Huanuco*. 7–12.
- Montanero Fernández, J. (2008). *Análisis Multivariante* (U. de Extremadura (Ed.); Vol. 59).
- Montgomery, D. (2004). *Diseño y análisis de experimentos* (Editorial Limusa S.A. (Ed.); Primera).
- Moya, S., Espino, H., Mir, V., Eduardo, H., Santracruz, A., & L, W. T. (2007). *Roya lineal, fenología, rendimiento y sus componentes en trigo harinero (Triticum aestivum L.)*.
- Prescott, J., & Saari, E. (2012). Manual de Metodología de enfermedades de los cereales. In *Foreign Affairs* (Alma McNab, Vol. 91, Issue 5, p. 46).

<https://doi.org/10.1017/CBO9781107415324.004>

- Ramirez-Anormaliza, R., Fabricio, G.-V., Mayra, D. R., Rubi, P.-H., Ruth, F.-L., Freddy, B.-D., Jose, D.-M., Juan, C.-C., Omar, F.-A., Fabian, F. R.-G., Edwain, C. R., David, V.-D., Roberto, B.-Q., Oscar, V.-O., Javier, B.-P., & Jesmy, C.-G. (2017). *Análisis multivariante: Teoría y práctica de la principales técnicas* (Universida).
- Reynolds M.P., Pask A.J.D., M. D. M. y C.-D. P. N. (2013). Fitomejoramiento fisiológico I: Enfoques Interdisciplinarios para mejorar la adaptación del cultivo. In *Fitomejoramiento fisiológico I: enfoques interdisciplinarios para mejorar la adaptacion del cultivo*.
- Rimieri, P. (2017). La diversidad genética y la variabilidad genética: Dos conceptos diferentes asociados al germoplasma y al mejoramiento genético vegetal. *Journal of Basic and Applied Genetics*, XXVIII(2), 7–13.
- Rodríguez, R. E., Paz Hernández, J. J., Iñiguez Monroy, C. G., Rueda Puente, E. O., Avendaño-Reyes, L., Cruz-Villegas, M., Ail-Catzim, C. E., Stoycheva, M., Koytchev Zlatev, R., Renganathan, P., & García López, A. M. (2014). Estabilidad de rendimiento en trigo en Valle de Mexicali, México. *Phyton*, 83, 65–70.
- Romani, M., & Herrera, I. (2017a). Ensayo comparativo de rendimiento de trigo bajo riego. Campaña 2016 – 2017. Campo Experimental “La María” del INTA EEA Santiago del Estero. *Instituto Nacional de Tecnología Agropecuaria, Pret 1231205*, 7.
- Romani, M., & Herrera, I. (2017b). *Ensayo comparativo de rendimiento de trigo bajo riego. Campaña 2016 – 2017. Pret 1231205*.
- Santos, F., Marza, F., Butron, R., & Quispe Félix. (2015). Correlación y coeficientes

de sendero para variables agronómicas y componentes de rendimiento de trigo harinero. *Revista Científica de Investigación INFO-INIAF*, March 2020, 8.

Silva, E., Michma, R., Marza, F., & Butrón, R. (2016). Ensayo comparativo de rendimiento de diez variedades de trigo en condiciones de altiplano central de Bolivia. *Programa Nacional de Trigo, Instituto Nacional de Innovación Agropecuaria y Forestal (INIAF)*, 84–89.

Vara, A. ( U. S. M. de P. (2012). Siete pasos para un atesis exitosa. In *Journal of Petrology: Vol. I* (segunda, Issue 1).

<https://doi.org/10.1017/CBO9781107415324.004>

Vianey, L. (2015). *Estimación máximo versosímil y mínimos cuadrados en el modelo estadístico lineal mixto: Metodología y análisis*. Universidad Central de Venezuela.

## ANEXOS

Tabla 29. Matriz de consistencia para el desarrollo del trabajo de investigación

<b>MATRIZ DE CONSISTENCIA</b>				
<b>TITULO: Evaluación y Análisis del rendimiento de 36 genotipos de trigo harinero utilizando el diseño de bloques completos al azar, distrito de Zurite – 2020</b>				
<b>MAESTRISTA: RIGOBERTO ESTRADA ZUNIGA</b>				
<b>MAESTRIA EN CIENCIAS CON MENCION EN ESTADISTICA</b>				
<b>PLANTEAMIENTO DEL PROBLEMA</b>	<b>OBJETIVOS DE LA INVESTIGACION</b>	<b>HIPOTESIS DE LA INVESTIGACION</b>	<b>VARIABLES EN ESTUDIO</b>	<b>METODOLOGIA</b>
<b>PROBLEMA GENERAL</b>	<b>OBJETIVO GENERAL</b>	<b>HIPOTESIS GENERAL</b>	Líneas de trigo harinero	Tipo de Investigación: Aplicada  Nivel de Investigación: Explicativo de naturaleza cuantitativa  Diseño: Experimental
¿Es posible determinar los niveles de rendimiento de 36 genotipos de trigo harinero a través de la evaluación y análisis por diseño de bloques completos al azar?	Determinar la variación de rendimiento de 36 genotipos de trigo harinero a través de la evaluación y análisis por diseño de bloques completos al azar	La evaluación y análisis utilizando diseño de bloques completos al azar son eficientes para seleccionar líneas de trigo harinero de alto rendimiento		
<b>PROBLEMAS ESPECIFICOS</b>	<b>OBJETIVOS ESPECIFICOS</b>	<b>HIPOTESIS ESPECIFICAS</b>		

¿Cómo influye la altura de planta en el rendimiento de trigo harinero?	Determinar la influencia de la altura de planta en el rendimiento de trigo harinero	La altura de la planta tiene una alta influencia en el rendimiento de trigo harinero	<u>Indicadores:</u> Altura de la planta N° granos por espiga Tamaño de Espiga Peso de mil granos Días a la madurez fisiológica
¿Cómo influye el número de granos por espiga en el rendimiento de trigo harinero?	Determinar la influencia del número de granos por espiga en el rendimiento de trigo harinero	El número de granos por espiga tiene una alta influencia en el rendimiento de trigo harinero	
¿Cómo influye el tamaño de espiga en el rendimiento de trigo harinero?	Determinar la influencia del tamaño de espiga en el rendimiento de trigo harinero	El tamaño de espiga tiene una alta influencia en el rendimiento de trigo harinero	
¿Cómo influye el peso de mil granos en el rendimiento de trigo harinero?	Determinar la influencia del peso de mil granos en el rendimiento de trigo harinero	El peso de mil granos tiene una alta influencia en el rendimiento de trigo harinero	
¿Cómo influye la madurez fisiológica en el rendimiento de trigo harinero?	Determinar la influencia de la madurez fisiológica en el rendimiento de trigo harinero	La madurez fisiológica tiene una alta influencia en el rendimiento de trigo harinero	

## Librerías y comandos utilizados en R para el análisis de los datos

```
##Librerías y comandos para el análisis de regresión y correlación
```

```
library(tidyverse)
```

```
library(boot)
```

```
library(car)
```

```
library(QuantPsyc)
```

```
library(ggplot2)
```

```
attach(regr)
```

```
names(regr)
```

```
class(alpl)
```

```
class(rdto)
```

```
modelo1=lm(rdto ~ alpl, data=regr, na.action = na.exclude)
```

```
summary(modelo1)
```

```
sqrt(0.3474)
```

```
par(mfrow=c(2,2))
```

```
plot(modelo1)
```

```
shapiro.test(residuals(modelo1))
```

```
library(lmtest)
```

```
bptest(modelo1)
```

```
grafico1=ggplot(regr, aes(alpl,rdto))
```

```
grafico1
```

```
grafico1+geom_point()
```

```
grafico1+geom_point()+geom_smooth(method = "lm",colour="Red")
```

```
dwt(modelo1$residuals)
```

```
##Librerías y comandos para el análisis de varianza y la comparación de medias

##altura de planta

altura=read.delim("clipboard")

attach(altura)

str(altura)

boxplot(alpl~trt)

library(lattice)

bwplot(alpl~trt|factor(blq),data=altura,scales = list(x=list(rot=90)),ylab="Altura de
planta", xlab="Tratamientos")

#modelo lineal

modelo_lineal1=lm(alpl~blq+trt)

summary(modelo_lineal1)

names(modelo_lineal1)

residuos<-rstudent(modelo_lineal1)

residuos

plot(modelo_lineal1)

shapiro.test(residuos)

library(lmtest)

bptest(modelo_lineal1)

library(MASS)

boxcox(modelo_lineal1,lambda =seq(-5,5), ylab="altura planta")

#ANOVA

mod1=aov(alpl ~ blq+trt)

shapiro.test(residuals(mod1))

bartlett.test(alpl ~ trt)
```

```
summary(mod1)
cv.model(mod1)
#Comparacion de medias
library(agricolae)
HSD.test(mod1, "trt", console=T)
HSD.test(gra, trt, 105,9.30, alpha = 0.05, group=TRUE, main = NULL,
         unbalanced=FALSE,console=TRUE)
```



**Figura 43.** Área destinada al trabajo de investigación e instalación del ensayo.



**Figura 44** Desarrollo vegetativo de la parcela experimental y evaluación para el control de malezas.



**Figura 45.** Vista panorámica de la parcela experimental en fase de madurez fisiológica.



**Figura 46.** Muestreo de espigas para la evaluación del número de granos por espiga.

## EVALUACION Y ANÁLISIS DEL RENDIMIENTO DE 36 GENOTIPOS DE TRIGO HARINERO UTILIZANDO EL DISEÑO DE BLOQUES COMPLETOS AL AZAR, DISTRITO DE ZURITE – 2020

Estrada Zúniga, Rigoberto<sup>1</sup>  
Rondinel Mendoza, Natalie Verónica<sup>2</sup>

### RESUMEN

Utilizar diseños experimentales en el mejoramiento genético de trigo harinero son fundamentales para generar información exacta e interpretable con alto rigor científico que permita contribuir con variedades potenciales en rendimiento, calidad comercial para una población cada vez más creciente; La investigación se planteó utilizando el diseño de bloques completos al azar con cuatro repeticiones para evaluar 36 genotipos de trigo harinero con los objetivos de determinar el nivel de influencia de la altura de planta, el número de granos por espiga, el tamaño de espiga, el peso de mil granos y la madurez fisiológica en el rendimiento. Entre los principales resultados se obtuvieron tres conglomerados de los tratamientos con características similares para las variables evaluadas. Se registró correlaciones estadísticamente significativas del rendimiento con altura de planta ( $r=0.589$ ), con el número de granos por espiga ( $r=0.932$ ), con el tamaño de espiga ( $r=0.860$ ), con el peso de mil granos ( $r=0.969$ ) y madurez fisiológica ( $r=0.932$ ) y el modelo explica que la proporción de la variabilidad total del rendimiento es mayor al 73% a excepción de la altura de planta que fue del 34%. Los hallazgos muestran como conclusión que existe influencia de las variables antes descritas en el rendimiento de trigo harinero, el diseño utilizado fue eficiente para evaluar el comportamiento de los 36 genotipos de trigo, y se seleccionaron a los genotipos TH-014, TH-034 y TH-001 superiores en rendimiento a los testigos de comparación y buenas características agronómicas para el mejoramiento genético.

### ABSTRACT

Using experimental designs in the genetic improvement of bread wheat are essential to generate exact and interpretable information with high scientific rigor that allows contributing with potential varieties in yield, commercial quality for an increasingly growing population; The research was carried out using a randomized complete block design with four repetitions to evaluate 36 bread wheat genotypes with the objectives of determining the level of influence of plant height, the number of grains per spike, spike size, thousand grain weight and physiological maturity in yield. Among the main results, three clusters of treatments with similar characteristics for the variables evaluated were obtained. Statistically significant correlations of yield with plant height ( $r=0.589$ ), with the number of grains per spike ( $r=0.932$ ), with spike size ( $r=0.860$ ), with the weight of a thousand grains ( $r=0.969$ ) and physiological maturity ( $r=0.932$ ) and the model explains that the proportion of total yield variability is greater than 73%, except for plant height, which was 34%. The findings show in conclusion that there is an influence of the variables described above on the yield of bread wheat, the design used was efficient to evaluate the behavior of the 36 wheat genotypes, and the TH-014, TH-034 and TH-034 genotypes were selected. TH-001 superior in yield to the comparison controls and good agronomic characteristics for genetic improvement.

### PALABRAS CLAVE

Rendimiento, trigo harinero, diseño de bloques completos al azar

---

<sup>1</sup> Ingeniero. Agrónomo, Universidad Nacional de San Antonio abad del Cusco, [190522@unsaac.edu.pe](mailto:190522@unsaac.edu.pe)

<sup>2</sup> Doctora. Docente de la Facultad de Ciencias Físicas, Químicas y Matemáticas-Universidad Nacional de San Antonio Abad del Cusco, [natalie.rondinel@unsaac.edu.pe](mailto:natalie.rondinel@unsaac.edu.pe)

## INTRODUCCION

Los procesos de investigación científica en el sector agrario son ejes fundamentales para el desarrollo tecnológico del país, por esta razón la experimentación aplicada y el uso de los diseños experimentales basados en la estadística inferencial es cada vez más relevante, al respecto Montgomery (2004), señala que el experimento es una o varias pruebas donde se realizan cambios en las variables independientes para observar e identificar los cambios en las variables dependientes de un proceso. En la actualidad, en investigaciones en trigo la precisión en las evaluaciones y análisis de datos debe ser mejorada. Por este motivo la investigación fue con el objetivo de determinar la variación del rendimiento del comparativo de 36 genotipos de trigo harinero a través de la evaluación y análisis utilizando el diseño de bloques completos al azar en el distrito de Zurite – 2020 y responder a la hipótesis para conocer si la influencia de características agronómicas mediante el diseño de bloques completos al azar es eficiente para seleccionar genotipos de trigo harinero de alto rendimiento. Montanero (2008), indica que para el análisis de conglomerados se debe disponer información que describan las relaciones entre individuos, a través de una medida apropiada de asociación que puede ser de distancia o de similitud entre ellos la distancia euclídea o de Mahalanobis, Gutiérrez y De la Vara (2008), señala que el análisis de regresión lineal simple explica el comportamiento de una variable respuesta en función de una o más variables independientes o predictoras y respecto al Análisis de Varianza señala que es la técnica central en el análisis de datos que permite probar las hipótesis de igualdad de los tratamientos con respecto a la media de la variable respuesta a partir de la separación de la variación total en las partes con las que contribuye cada fuente de variación en el experimento. Investigaciones similares fueron desarrolladas por Silva et al (2016), Huanca et al (2014) utilizando el DBCA que encontraron asociación del rendimiento con las principales variables agronómicas en el cultivo de trigo, entre ellos con el número de granos por espiga, la altura de planta, el peso de mil granos.

La investigación fue experimental de tipo aplicada desarrollada en el distrito de Zurite utilizando el diseño de bloques completos al azar con cuatro repeticiones para evaluar 36 genotipos de trigo harinero y entre los principales resultados se registró que la influencia de la altura de la planta en el rendimiento de trigo harinero fue positiva y moderada con ( $r= 0.589$ ), respecto al número de granos por espiga en el rendimiento la influencia fue positiva y alta ( $r=0.932$ ), el tamaño de la espiga para la selección de variedades de trigo harinero presenta una influencia positiva y alta con ( $r=0.860$ ), el peso de mil granos en el rendimiento de trigo harinero la influencia fue positiva y alta con ( $r= 0.969$ ) y la influencia de la madurez fisiológica en el rendimiento de trigo harinero fue positiva y alta ( $r=0.932$ ). Los supuestos del modelo para el ANOVA del caso particular de altura de planta, número de granos por espiga y peso de mil granos no se cumplió la homogeneidad de varianzas, posiblemente porque los tratamientos corresponde a líneas genéticas procedentes de cruzamientos de diversos progenitores que contribuyeron en la variabilidad del comportamiento de sus características cuantitativas, y respecto al ANOVA para el rendimiento existió diferencia estadística entre tratamientos ( $p$ -valor $<0.0001$ ), la comparación de medias muestra que el genotipo TH-014, alcanza el mayor rendimiento (8.494 t/ha) en campo en comparación a los demás tratamientos y testigos de comparación.

## MATERIALES Y METODOS

La investigación fue experimental de tipo aplicada con dos variables Vara-Horna, (2012), explicativo de naturaleza cuantitativa (Hernández, 2014, p.95), se desarrolló en la localidad de Zurite a 3405 msnm, utilizando el diseño de Bloques Completamente al Azar con cuatro repeticiones con 36 genotipos de trigo harinero procedentes del Centro Internacional de Mejoramiento de Maíz y Trigo con un área por unidad experimental de 4.80 m<sup>2</sup> y área experimental por localidad de 907.20 m<sup>2</sup>. La técnica de selección del muestreo probabilístico aleatorio simple, Arias (2012). para las variables altura de planta, tamaño de espiga y número de granos por espiga con elección aleatoria de 10 plantas en competencia completa en dos surcos centrales de la unidad experimental, la variable madurez fisiológica por observación directa de la fase fenológica, el peso de mil granos por conteo al azar de mil semillas de la producción de dos surcos centrales y con respecto al rendimiento la cosecha del área central de la unidad experimental (2.40 m<sup>2</sup>). Las herramientas estadísticas para el análisis de los datos fueron el lenguaje o entorno de programación R y el paquete estadístico InfoStat 2018 para analizar los datos con las siguientes expresiones matemáticas:

1. Conglomerado jerárquico a través de la distancia de Mahalanobis con el enlace de Ward para determinar la similitud de los genotipos

$$d_m^2(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \mathbf{x}_2)' \mathbf{S}^{-1} (\mathbf{x}_1 - \mathbf{x}_2)$$

2. Análisis de regresión lineal simple

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma^2)$$

3. Coeficiente de determinación como medida de bondad de ajuste

$$R^2 = 1 - \frac{SCE}{SCTotal} = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

4. Error cuadrático medio como medida de bondad de ajuste

$$ECMP = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

5. Coeficiente de correlación de Pearson como medida de asociación

$$r = \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\sqrt{\left( \left( \sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n} \right) \left( \sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \right) \right)}}$$

6. Análisis de varianza

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

## RESULTADOS

Tabla 1. Matriz de correlaciones de Pearson para las variables en estudio

	Días a madurez	N° granos por espiga	Altura de la planta (cm)	Tamaño de espiga (cm)	Peso de mil granos (g)	Rendimiento en grano (t/ha)
Días a madurez	1					
N° granos por espiga	0.8925**	1				
Altura de la planta (cm)	0.5202**	0.5214**	1			
Tamaño de espiga (cm)	0.7728**	0.7739**	0.6691**	1		
Peso de mil granos (g)	0.9024**	0.9337**	0.5947**	0.8574**	1	
Rendimiento en grano (t/ha)	0.9316**	0.9324**	0.5893**	0.8600**	0.9693**	1

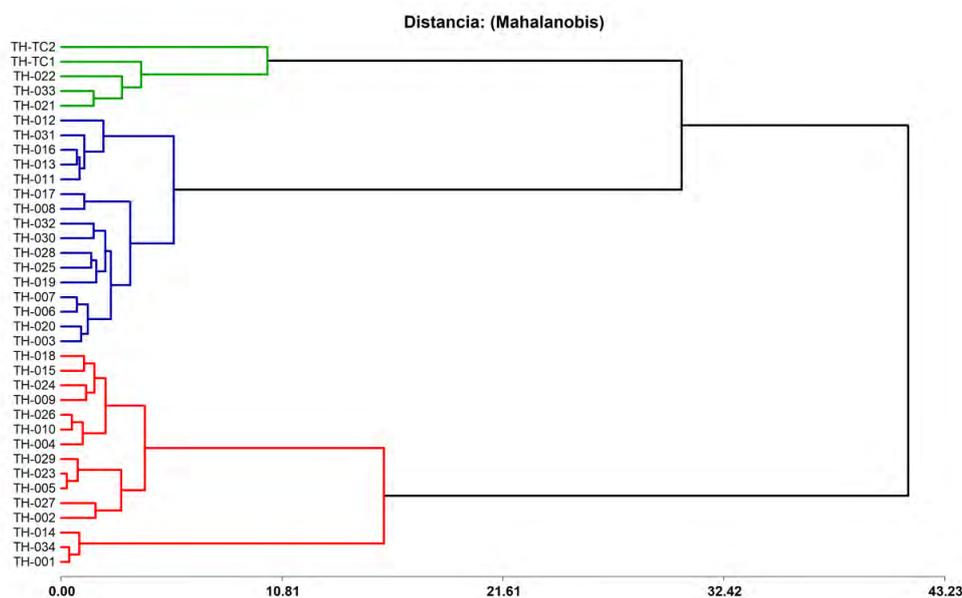


Figura 1. Dendrograma de los tratamientos en estudio que muestra la formación de tres grupos

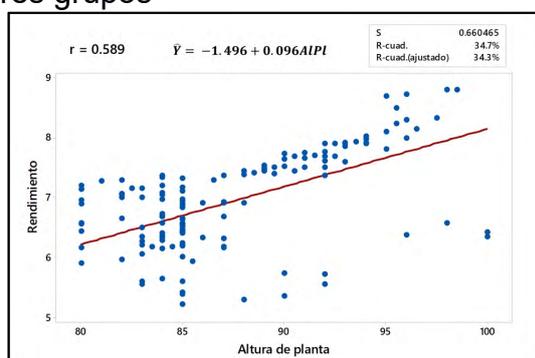


Figura 2. Recta ajustada para el rendimiento y la altura planta.

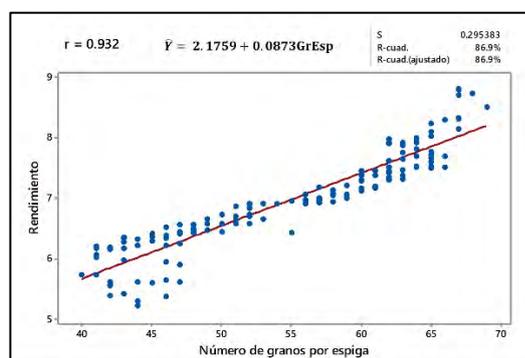


Figura 3. Recta ajustada para el rendimiento y número granos/espiga.

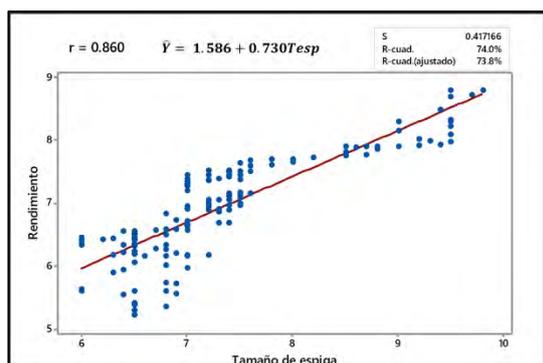


Figura 4. Recta ajustada para el rendimiento y el tamaño de espiga.

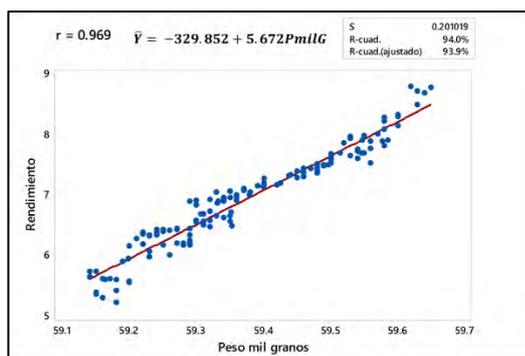


Figura 5. Recta ajustada para el rendimiento y el peso de mil granos.

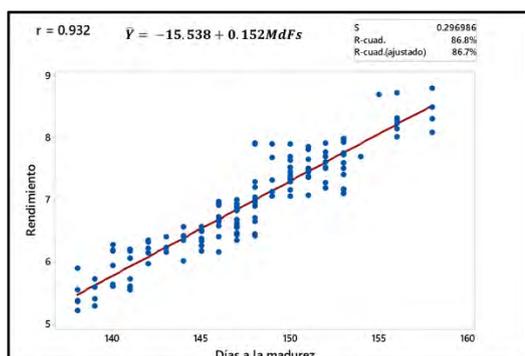


Figura 6. Recta ajustada para el rendimiento y la madurez.

**Tabla 27.** Análisis de varianza para el rendimiento

F.V.	GL	SC	CM	F	p-valor
Bloques	3	0.4392	0.1464	1.5204	0.2135
Tratamientos	35	84.3623	2.4104	25.0303	<0.0001
Error	105	10.1112	0.0963		
Total	143	94.9128			

Criterios de bondad de ajuste para el rendimiento

Variable	N	R <sup>2</sup>	R <sup>2</sup> Aj	CV
Rendimiento	144	0.89	0.85	4.47

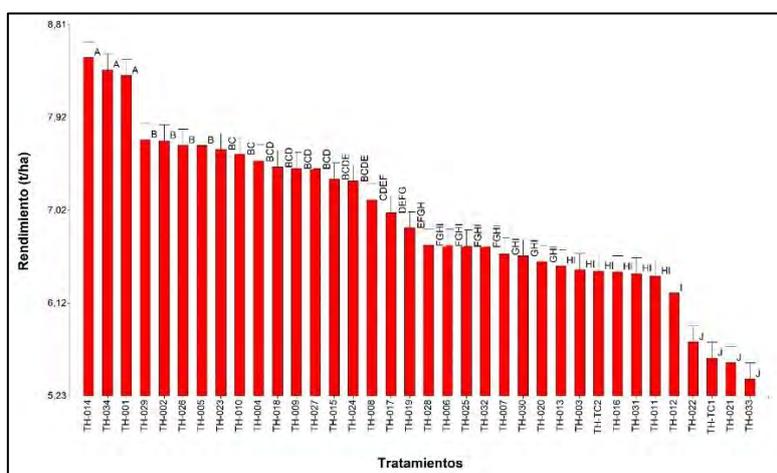


Figura 7. Comparación de medias para el rendimiento de los tratamientos

## DISCUSION

Del análisis de clúster se obtuvieron tres grupos con características similares a través del agrupamiento con el método de enlace de Ward que en el estudio desarrollado por Chávez et al. (2015) con este método se dio lugar a un agrupamiento jerárquico que eliminó los problemas de encadenamiento se muestran al emplear otras técnicas, resultado similar muestra Chávez et al. (2015) en su evaluación de genotipos de maíz señalando que los caracteres más significativos en análisis de clúster son la longitud de la mazorca, del grano y el peso de 100 granos. Bajo estas consideraciones en el estudio consideramos que los días a la madurez, el número de granos por espiga, la altura planta, el tamaño de espiga, el peso de mil granos y el rendimiento fueron las variables adecuadas para el análisis de clúster.

El uso de modelos líneas generales en el análisis de los datos fue aplicado considerando el cumplimiento de los supuestos del modelo llegando a los siguientes resultados: Del análisis de regresión lineal simple y la correlación del rendimiento con la altura planta, el número de granos por espiga, el tamaño de espiga, el peso de mil granos y la madurez fisiológica en la presente investigación se encuentran coeficientes de correlación de 0.589, 0.932, 0.860, 0.969 y 0.932 respectivamente que en comparación a los reportes desarrollados por Huanca et al. (2015), Romaní y Herrera (2017b) y Espitia Rangel et al. (2021) se observa que existe mayor correlación bajo las condiciones donde desarrollamos la investigación y al 5% de significación estadística a excepción de la correlación entre el rendimiento y la altura de planta donde es similar a lo obtenido por Huanca et al. (2015); Así mismo, el modelo lineal ajustado de la regresión muestra que la variabilidad total presente en el rendimiento es explicada a través de la variabilidad observada en las variables independientes.

Con las evidencias de los resultados se concluye que el diseño de bloques completos al azas con cuatro repeticiones en la investigación fue eficiente para evaluar y determinar la variación del rendimiento de los 36 genotipos de trigo, puesto que el modelo lineal utilizado en las pruebas estadísticas explica con más del 75% la variabilidad entre los genotipos bajo las condiciones experimentales de evaluación. La información utilizada para los análisis fue de buena calidad como detalla los coeficientes de variabilidad menores al 10% del ANOVA para cada variable, no obstante existe variabilidad que es explicado por factores que no fueron tomados en cuenta en los análisis entre ellos la procedencia de los progenitores que contribuyeron en la variabilidad del comportamiento de sus características asociados al medio ambiente donde las estructuras genéticas generó la heterogeneidad reflejado en el supuesto de homogeneidad de varianzas para el caso particular de altura de planta, número de granos por espiga y peso de mil granos. Finalmente, destacaron los genotipos TH-014, TH-034 y TH-001, con rendimientos de 8.494, 8.377 y 8.322 t/ha respectivamente que fueron seleccionados para el desarrollo de nuevas variedades comerciales de trigo harinero y/o su uso como progenitores en los planes de mejoramiento.

**REFERENCIAS BIBLIOGRÁFICAS.**

- Arias. (2012). El proyecto de investigación, Introducción a la metodología científica (E. E. C.A. (ed.); sexta edic).
- Chavez, D., Ileana, M., Varela, M., & Lianne, F. (2015). Utilización del análisis de clusters con variables mixtas en la selección de genotipos de maíz. *Revista Investigación Operacional*, 30(3), 209–216.
- Espitia Rangel, E., Martínez Cruz, E., Villaseñor Mir, H. E., Santa Rosa, R. H., Limón Ortega, A., & Lozano Grande, A. (2021). Variabilidad genética y criterios de selección del rendimiento y los componentes en trigos harineros de temporal. *Revista Mexicana de Ciencias Agrícolas*, 12(2), 305–315. <https://doi.org/10.29312/remexca.v12i2.2787>
- Gutiérrez, H., & De la Vara, R. (2008). *www.FreeLibros.org* (M. Interamericana (Ed.); Segunda).
- Huanca, N., Butron, R., Marza, F., & Quispe, F. (2015). Estimación de la variabilidad fenotípica de trigo harinero para componentes de productividad. Instituto Nacional de Innovación Agropecuaria y Forestal, 77–83.
- Montanero Fernández, J. (2008). *Análisis Multivariante* (U. de Extremadura (Ed.); Vol. 59).
- Montgomery, D. (2004). *Diseño y análisis de experimentos* (Editorial Limusa S.A. (Ed.); Primera).
- Romaní, M., & Herrera, I. (2017a). Ensayo comparativo de rendimiento de trigo bajo riego. Campaña 2016 – 2017. Campo Experimental “La María” del INTA EEA Santiago del Estero. Instituto Nacional de Tecnología Agropecuaria, Pret 1231205, 7.
- Silva, E., Michma, R., Marza, F., & Butrón, R. (2016). Ensayo comparativo de rendimiento de diez variedades de trigo en condiciones de altiplano central de Bolivia. Programa Nacional de Trigo, Instituto Nacional de Innovación Agropecuaria y Forestal (INIAF), 84–89.
- Vara, A. ( U. S. M. de P. (2012). Siete pasos para un atesis exitosa. In *Journal of Petrology*: Vol. I (segunda, Issue 1). <https://doi.org/10.1017/CBO9781107415324.004>