

**TRANSFORMACIONES USADAS
EN EL ANALISIS ESTADISTICO DE DATOS**

Universidad Autónoma Agraria
"ANTONIO NARRO"

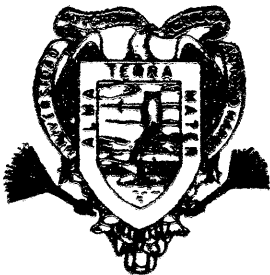


ALBERTO RODRIGUEZ HERNANDEZ

BIBLIOTECA

T E S I S

**PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL**



**Universidad Autónoma Agraria
Antonio Narro**

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

FEBRERO 1995

TRANSFORMACIONES USADAS
EN EL ANALISIS ESTADISTICO DE DATOS

ALBERTO RODRIGUEZ HERNANDEZ

TESIS

Presentada como requisito parcial

para obtener el grado de

Maestro en Ciencias

En Estadística Experimental

UNIVERSIDAD AUTONOMA AGRARIA

ANTONIO NARRO

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

Febrero de 1995

Tesis elaborada bajo la supervisión del comité particular
de asesoría y aprobada como requisito parcial, para optar
al grado de

MAESTRO EN CIENCIAS EN
ESTADISTICA EXPERIMENTAL

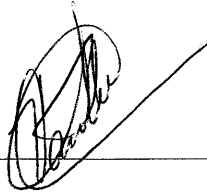
COMITE PARTICULAR

Asesor principal:



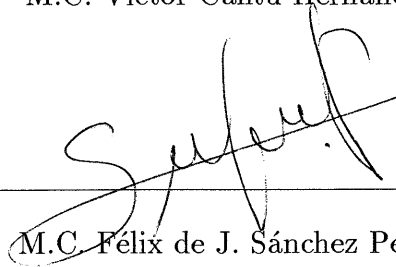
M.C. Mario Cantú Sifuentes

Asesor:

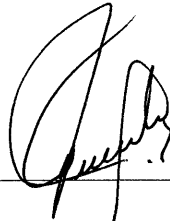


M.C. Víctor Cantú Hernández

Asesor:



M.C. Félix de J. Sánchez Pérez



Dr. Jesús Fuentes Rodríguez

Subdirector de Postgrado

Buenvista, Saltillo, Coahuila. Febrero de 1995

AGRADECIMIENTOS

A la Universidad Autónoma Agraria Antonio Narro, en especial al Departamento de Estadística y Cálculo, por brindarme la oportunidad de realizar mis estudios de Postgrado.

De una manera muy especial mi patente agradecimiento al M.C. Mario Cantú Sifuentes, quien siempre me alentó a superarme. Mi más profundo reconocimiento a su labor.

Al M.C. Victor Cantú Hernández, por su amistad, dedicación y ayuda durante mi estancia en esta Universidad.

Al M.C. Félix de J. Sánchez Pérez, por su apoyo moral y científico para la culminación de mis estudios.

Al Dr. Rolando Cavazos Cadena, por haberme inculcado sus conocimientos científicos y su filosofía. Un reconocimiento a la planta de maestros del Postgrado de Estadística Experimental.

A mis amigos y compañeros Sergio Sánchez M., José T. Rodríguez E., José A. Perales P., J. Manuel Saucedo E., Ismael Hernández B., Edgar E. Guzmán M. y Félix Rodríguez D., porque a pesar de tantos problemas nos mantuvimos juntos.

DEDICATORIA

A mis Padres:

Por su cariño y apoyo constante en todas las actividades que realizo.

A mis hermanos:

Por el cariño que siempre nos ha unido.

A Isabel, Javier Alejandro, Laura Alicia y Pedro.

A Alejandra:

Por el amor que me inspira.

COMPENDIO

Transformaciones usadas en el análisis estadístico de datos.

POR

ALBERTO RODRIGUEZ HERNANDEZ

MAESTRIA

ESTADISTICA EXPERIMENTAL

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. FEBRERO 1995.

M.C. Mario Cantú Sifuentes - Asesor -

Palabras claves: normalidad, independenciam, aditividad, homogeneidad de varianzam, transformaciones.

El presente estudio plantea el problema cuando los supuestos del análisis de varianzam no se cumplen. Asimismo, se proponen métodos para satisfacer dichos supuestos, finalizando con la aplicación de transformación a los datos.

Aunque esencialmente no se obtienen nuevas técnicas, se hace un planteamiento general del problema para dar solución al mismo. Los principales resultados se ilustran con ejemplos.

ABSTRACT

Transformations used in the statistics Analysis of data.

BY

ALBERTO RODRIGUEZ HERNANDEZ

MASTER OF SCIENCE

EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. FEBRUARY 1995.

M.S. Mario Cantú Sifuentes - Advisor -

Key words: normality, independence, additivity, homogeneity of variance, transformations.

In this study we establish the problem in case of the assumptions of analysis of variance are not carrying out. We also present those methods that can work according to the assumption, we finish this study with the application of the transformation of the data.

Although essentially no new techniques were obtained, but we try to show a general view about this problem in order to get a solution. Examples are given to illustrate the main results.

INDICE DE CONTENIDO

	Página
INDICE DE CUADROS	ix
INDICE DE FIGURAS	xii
1. INTRODUCCION	1
2. NORMALIDAD	3
CONCEPTOS PRELIMINARES	3
HISTOGRAMA DE LOS RESIDUOS	4
GRAFICA DE PROBABILIDAD NORMAL DE LOS RESIDUOS	12
PRUEBA DE BONDAD DE AJUSTE	17
PRUEBA DE SHAPIRO AND WILK	23
PRUEBA DE KOLMOGOROV-SMIRNOV	25
3. INDEPENDENCIA	31
CONCEPTOS PRELIMINARES	31
GRAFICA DE RESIDUOS CONTRA SECUENCIA DE TIEMPO	31
4. ADITIVIDAD Y LINEALIDAD	34
CONCEPTOS PRELIMINARES	34
PRUEBA DE TUKEY DE ADITIVIDAD	35
5. HOMOGENEIDAD Y VARIANZA CONSTANTE	41
CONCEPTOS PRELIMINARES	41

GRAFICA DE RESIDUOS CONTRA	
EL VALOR AJUSTADO (\hat{y}_{ij})	43
PRUEBA DE BARTLETT	
PARA HOMOGENEIDAD DE VARIANZA	45
PRUEBA DE BARTLETT Y KENDALL PARA	
HOMOGENEIDAD DE VARIANZA	47
PRUEBA DE BURR Y FOSTER PARA	
HOMOGENEIDAD DE VARIANZA	49
6. TRANSFORMACION DE DATOS	52
CONCEPTOS PRELIMINARES	52
TRANSFORMACION RAIZ CUADRADA	53
TRANSFORMACION ANGULAR O ARCO SENOS	60
TRANSFORMACION LOGARITMICA	68
SELECCION DE UNA TRANSFORMACION	
PARA IGUALAR VARIANZAS	72
SELECCION EMPIRICA DE α	75
DETERMINACION ANALITICA	
DE LA TRANSFORMACION	79
CONCLUSIONES	85
LITERATURA CITADA	86
APENDICE	88

INDICE DE CUADROS

Cuadro No.	Página
2.1. Altura total de plantas de rábano	9
2.2. Residuos de altura total en plantas de rábano	9
2.3. Resultados de un experimento de obtención de penicilina	11
2.4. Residuos del experimento de obtención de penicilina	11
2.5. Datos del experimento de resistencia a la tensión <i>lib/pulg²</i>	14
2.6. Residuos de los datos del experimento de resistencia a la tensión . . .	15
2.7. Residuos ordenados y puntos de probabilidad	15
2.8. Tiempo de desintegración de las tabletas para los cuatro tipos . . .	19
2.9. Cálculo de \bar{x} y s para los tiempos de desintegración	21
2.10. Frecuencias esperadas para los tiempos de desintegración	22
2.11. Cálculo de \bar{x} y s para las alturas totales de las plantas	28
2.12. Frecuencias esperadas de las alturas totales de las plantas	28
2.13. Cálculo de la prueba de kolmogorov-Smirnov	28
2.14. Cálculo de la prueba de kolmogorov-Smirnov	30
3.1. Orden en que se recolectaron los datos del cuadro 2.5	33
4.1. Ejemplo ilustrativo	37
4.2. Cálculo para la prueba de aditividad	39
4.3. Análisis de varianza y prueba de aditividad	39
5.1. Análisis de homogeneidad de $\log s^2$	48

Cuadro No.	Página
5.2. Análisis de varianza de $\log s^2$	48
5.3. Cálculo de las varianzas	49
6.1. Número de huevecillos de <i>Heliothis zea</i>	54
6.2. Análisis de varianza de los valores originales del cuadro 6.1	55
6.3. Valores del cuadro 6.1 transformados a \sqrt{x}	55
6.4. Análisis de varianza de los valores del cuadro 6.3	56
6.5. Porcentaje de avanamiento del aquenio	56
6.6. Análisis de varianza de los valores del cuadro 6.5	57
6.7. Valores originales del cuadro 6.5 transformados a \sqrt{x}	57
6.8. Análisis de varianza de los valores del cuadro 6.7	57
6.9. Número de <i>Iygu</i> s por 50 barridas	58
6.10. Análisis de varianza de los valores del cuadro 6.9	59
6.11. Datos transformados del <i>Iygu</i> s	59
6.12. Análisis de varianza de los valores del cuadro 6.11	60
6.13. Número de semillas de lechuga germinadas en muestras de 50	63
6.14. Análisis de varianza de los datos originales	64
6.15. Datos transformados del número de semillas de lechuga	65
6.16. Análisis de varianza de los datos transformados	65
6.17. Porcentajes de mazorcas enfermas en un experimento de maíz	66
6.18. Análisis de varianza de los valores del cuadro 6.17	67
6.19. Valores transformados del cuadro 6.17	68
6.20. Análisis de varianza del cuadro 6.19	68
6.21. Número estimado de cuatro clases de plancton	70
6.22. Transformación logarítmica del cuadro 6.21	71
6.23. Análisis de varianza de los valores del cuadro 6.22	71

Cuadro No.	Página
6.24. Transformaciones para estabilizar las varianzas cuando $\sigma_y \propto \mu^\alpha$. . .	75
6.25. Datos del gasto máximo	76
6.26. Análisis de varianza para el gasto máximo	76
6.27. Análisis de varianza de datos transformados $Y = \sqrt{y}$ del cuadro 6.25 .	79
6.28. Datos del gasto máximo para $\lambda=0.50$	81
6.29. Análisis de varianza para el gasto máximo para $\lambda=0.50$	82
6.30. Datos del gasto máximo para $\lambda=0$	83
6.31. Análisis de varianza para el gasto máximo para $\lambda=0$	83

INDICE DE FIGURAS

Figura No.	Página
2.1. Histograma de puntos de todos los residuos	10
2.2. Gráfico de todos los puntos de los residuos	12
2.3. Gráfica de probabilidad normal de los residuos	16
3.1. Modelo aproximado de la gráfica	32
3.2. Gráfica de residuos contra tiempo	33
5.1. Forma de la gráfica contra valores ajustados	44
5.2. Gráfica de residuos contra valores ajustados	44
6.1. Transmisión de errores de y a $y = y^\lambda$	74
6.2. Gráfica de residuos contra $\hat{y}_{ij} = \bar{y}_i$	77
6.3. Gráfica de $\log s_i$ contra $\log \bar{y}_i$. para el ejemplo 6.7	78
6.4. Gráfica de los residuos de los datos transformados contra \hat{Y}_{ij}	78
6.5. Gráfica de $SC_E(\lambda)$ contra λ	84

1. INTRODUCCION.

Un investigador que se conforma con aprender las “recetas” para llevar a cabo un análisis de varianza, sin buscar el dominio y la comprensión de los principios inherentes al mismo, puede encontrarse con serios problemas. Sea que los comprenda o no, el investigador hará ciertas suposiciones acerca de sus datos cuando realice un análisis de varianza. Si los datos no concuerdan con estas suposiciones, dicho análisis puede dar lugar a que el investigador llegue a conclusiones que no tienen justificación. Asimismo, el investigador puede descuidar conclusiones importantes que se alcanzarían si los datos fuesen analizados adecuadamente.

El análisis de varianza fue ideado por Sir Ronald A. Fisher en la década de 1930 y es esencialmente un procedimiento aritmético que descompone una suma total de cuadrados en componentes asociados con fuentes de variación reconocida. Se ha usado con provecho en todos los campos de investigación en los que los datos se miden cuantitativamente.

Los principales propósitos del análisis de varianza son:

- 1) Estimar diferencias de tratamientos que son de interés.
- 2) Obtener idea de la precisión de nuestro experimento.
- 3) Llevar a cabo pruebas de hipótesis.

Los principales supuestos básicos para un análisis de varianza son

- 1) Distribución aleatoria y normal de los términos de error. Se entiende por distribución aleatoria el hecho de que tanto la asignación del material experimental

como el orden en que se realizan las pruebas individuales o ensayos se determinan aleatoriamente.

2) Homogeneidad de las varianzas. Esto es, los errores experimentales deben tener varianza común.

3) Independencia de varianzas y medias. Esto es, la probabilidad de que los errores de cualquier observación que tenga un valor particular no dependa de los valores de los errores de otra observación.

4) Aditividad de los efectos principales. Esto es, debido a que los términos del modelo matemático (Modelo lineal aditivo) se suman.

Cuando estos supuestos son seriamente violados, el análisis de varianza no es válido. A fin de satisfacer los supuestos, se hacen transformaciones que corrijan las deficiencias de los datos. Las transformaciones más comunes son la raíz cuadrada, la logarítmica, la recíproca y la angular o transformación arco-seno.

El propósito de esta tesis es considerar cada suposición y discutir las consecuencias cuando los supuestos no se satisfacen. Además, se recomiendan procedimientos encaminados a satisfacer los supuestos a fin de dar validez a los datos.

2. NORMALIDAD.

Conceptos Preliminares.

Es bien sabido que en los modelos estadísticos los errores se suponen normal e independientemente distribuidos con media cero y varianza comun σ^2 .

Si estas suposiciones son válidas, el procedimiento del análisis de varianza constituye una prueba exacta para la hipótesis de igualdad en las medias de tratamientos. Por lo general, estas suposiciones no se cumplen con exactitud en la práctica. Por lo tanto, no podemos depender del análisis de varianza hasta que se hayan comprobado estos supuestos.

Hay una pérdida en la eficiencia del análisis, debido a que cuando los errores no son normales, la media de los valores observados para un tratamiento no es, en general, la estimada más precisa de la media poblacional correspondiente para este tratamiento. Si la forma matemática de la distribución de frecuencia de los errores se conociera, se podría desarrollar un análisis más eficiente. Esta forma de ataque pocas veces se intenta en la práctica, probablemente debido a que la distribución exacta de los errores normales rara vez se conoce y el análisis más elegante sería más complicado. Sin embargo, un aspecto de la no normalidad en las distribuciones es que la varianza muchas veces está relacionada a la media (Snedecor y Cochran, 1975).

Cuando la distribución de los errores experimentales es decididamente asimétrica, la componente del error de un tratamiento tiende a ser función de la media de tratamiento. Esto produce una heterogeneidad en el término de error.

Si se conoce la relación funcional, se puede encontrar una transformación que dé errores que se distribuyan de manera más cercana a la normal. De esta forma se puede efectuar un análisis de la varianza con los datos transformados de modo que el término de error sea esencialmente homogéneo.

En esta sección se analizan e ilustran métodos para comprobar la suposición de normalidad, y se presentan algunas soluciones que a menudo resultan útiles cuando esta suposición no se cumple.

Histograma de los Residuos.

Una forma de comprobar la suposición de normalidad consiste en hacer un histograma de los residuos. Si la suposición de los errores son $NID(0, \sigma^2)$ se satisface, esta gráfica debe ser semejante a la de una muestra extraída de una distribución normal centrada en cero (Box *et al.*, 1988).

En estadística, un modelo que describa la naturaleza de una observación consta de una media más error. Este es un modelo lineal aditivo. En la media puede estar sólo el parámetro μ , o puede estar compuesta de una suma de parámetros. Un supuesto es que los errores son aleatorios, esto es, que la población de las Y_{ij} se muestrea aleatoriamente. El modelo lineal aditivo más simple es

$$Y_{ij} = \mu + \epsilon_{ij},$$

Y_{ij} es una observación de la media μ , pero está sujeta a un error de muestreo, ϵ_{ij} .

Un residuo es un estimador del error aleatorio ϵ_{ij} , dado que

$$\epsilon_{ij} = Y_{ij} - \mu,$$

el residuo correspondiente denotado por e_{ij} , se define como

$$e_{ij} = y_{ij} - \bar{y}_i.$$

donde

$$i = 1, 2, \dots, k, \quad j = 1, 2, \dots, n.$$

Esto es, tenemos que para el diseño completamente al azar el modelo es

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

donde

$$i = 1, 2, \dots, k, \quad j = 1, 2, \dots, n,$$

donde se suponen que los errores son independientes y están normalmente distribuidos con media cero y varianza constante σ^2 .

El residuo de la observación j del tratamiento i se define mediante

$$e_{ij} = y_{ij} - \hat{y}_{i.},$$

en donde $\hat{y}_{i.}$ es una estimación de la observación y_{ij} correspondiente calculada por

$$\begin{aligned} \hat{y}_{ij} &= \hat{\mu} - \hat{\tau}_i \\ &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) \\ &= \bar{y}_{i.} \end{aligned} \tag{2.2.1}$$

la ecuación (2.2.1) muestra un resultado que se intuye fácilmente, ya que la estimación de cualquier observación del i -ésimo tratamiento es igual al promedio del tratamiento correspondiente (Montgomery, 1991). Por lo que

$$e_{ij} = y_{ij} - \bar{y}_{i.}$$

Para un diseño en bloques al azar tenemos que el modelo es

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

donde

$$i = 1, 2, \dots, k, \quad j = 1, 2, \dots, b,$$

μ es la media general, τ_i es el efecto del i -ésimo tratamiento, β_j es el efecto del j -ésimo bloque y ϵ_{ij} es el término usual $NID(0, \sigma^2)$ del error aleatorio. Los residuos son

$$e_{ij} = y_{ij} - \hat{y}_{i.},$$

en donde $\hat{y}_{i.}$ es una estimación de la observación y_{ij} correspondiente calculada por

$$\begin{aligned} \hat{y}_{ij} &= \hat{\mu} - \hat{\tau}_i + \hat{\beta}_j \\ &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ &= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..} \end{aligned} \tag{2.2.2}$$

la ecuación (2.2.2) se demuestra usando el método de mínimos cuadrados. Por lo tanto

$$e_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$$

Desafortunadamente, a menudo ocurren fluctuaciones considerables cuando las muestras son pequeñas, por lo que una desviación moderada aparente de la normalidad no necesariamente implica una violación seria a las suposiciones. Desviaciones mayores de la normalidad son potencialmente graves y requieren un análisis más profundo.

Cuando aparecen anomalías muy fuertes, hay que buscar, sin embargo, las posibles causas.

Ejemplo 2.1. El conjunto de datos que se incluyen en esta sección proceden de un estudio que fue presentado al Laboratorio de Estadística de la Facultad de Ciencias de la Universidad Nacional Autónoma de México en el año de 1982, por el profesor J. Soto de la Facultad de Química de la misma Universidad.

El grupo de investigación bajo la dirección del profesor Soto abordó el problema de evaluar las posibilidades de promover en forma extensiva los cultivos

hidropónicos. De acuerdo a la información proporcionada por este equipo de investigadores, la hidroponía se puede definir como el cultivo de plantas en ausencia de suelo, en el que se provee a las plantas de los nutrientes que requieren para su crecimiento, no por medio de su habitáculo natural que es el suelo sino a través de una solución de agua y sales minerales.

La importancia de efectuar el cultivo hidropónico resulta patente si se observa que en regiones donde el suelo es, por naturaleza, pedregoso o delgado, usualmente el agricultor insiste en sembrar invirtiendo cantidades considerables de mano de obra y recursos en general. Como alternativa, un método planificado de cultivos sin suelo puede ser un auxiliar valioso de la agricultura, particularmente si se establecen las condiciones en las que este tipo de cultivo resulta rentable.

Bajo estas circunstancias y concretamente en relación al proceso experimental, se decidió investigar los efectos que la adición de composta, una fuente de nutrientes relativamente barata, tiene en los cultivos hidropónicos. Específicamente se seleccionaron dos hortalizas (rábano y lechuga) que por su facilidad de manejo y el corto período de tiempo en que se desarrollan, resultan apropiadas para este tipo de estudio. En cada caso se utilizaron semillas de un lote que se consideró homogéneo y de cualquier manera, se practicaron determinaciones de pureza, germinación y viabilidad con la finalidad de corroborar la homogeneidad y representatividad del lote.

Tanto para el caso de la lechuga como el del rábano y considerando que se trata de cultivos comestibles, se registraron las siguientes variables de respuesta que están asociadas a la utilidad del producto obtenido:

- 1) Altura total.
- 2) Peso seco.

- 3) Porcentaje de proteína.
- 4) Porcentaje de humedad.
- 5) Concentración de fósforo.

Finalmente, y de acuerdo a la disponibilidad de espacio en los invernaderos, a la cantidad de materiales necesarios y al trabajo involucrado en las determinaciones requeridas, se decidió utilizar cinco repeticiones; esto es, cinco determinaciones independientes y bajo las mismas condiciones experimentales de cada variable de respuesta para cada tratamiento.

En el caso de este estudio el análisis de los dos cultivos se realizó por separado. Así, para cada variable de respuesta en cada cultivo, se propuso un modelo del siguiente tipo:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

donde $i = 1, 2, \dots, k$; $j = 1, 2, \dots, n$, que recibe el nombre de modelo de diseño con un criterio de clasificación o completamente al azar y en donde

y_{ij} representa el valor registrado de la variable de respuesta bajo estudio en la j -ésima repetición del i -ésimo tratamiento.

μ representa una media global de la variable respuesta (en este caso, altura total).

τ_i representa el efecto debido al i -ésimo tratamiento.

ϵ_{ij} representa un término de error aleatorio sobre el cual se hacen las suposiciones de normalidad, media cero, varianza constante e independencia. En el caso de este ejemplo $k=5$ (hay cinco tratamientos) y n_1, n_2, n_3, n_4, n_5 (hay cinco repeticiones).

En particular, para el caso de la variable altura total en plantas de rábano, se obtuvieron los resultados del cuadro 2.1.

En el cuadro 2.2 se muestran los residuos denotados por

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_i.$$

donde, $i = 1, 2, \dots, k$, $j = 1, 2, \dots, n$.

CUADRO 2.1. Altura total (en centímetros) de plantas de rábano.

	<u>Tratamientos</u>				
<u>Repeticiones</u>	1	2	3	4	5
1	31.33	25.66	34.16	31.33	49.00
2	35.50	31.66	32.00	34.50	42.33
3	36.66	24.00	22.50	35.50	37.00
4	33.66	36.00	27.33	37.33	38.66
5	23.66	40.00	51.33	30.33	44.66
\bar{y}_i	32.16	31.46	33.46	33.60	42.33

CUADRO 2.2. Residuos de altura total, en plantas de rábano.

	<u>Tratamientos</u>				
<u>Repeticiones</u>	1	2	3	4	5
1	-0.83	-5.80	+0.69	-2.33	+6.67
2	+3.34	+0.19	-1.46	+0.90	0
3	+4.49	-7.76	-10.9	+1.90	-5.33
4	+1.49	+4.54	-6.13	+3.73	-3.67
5	-8.50	+8.54	+17.87	-3.27	+2.33

Se realizará una inspección general dibujando los residuos en un gráfico de puntos en la figura 2.1.

Como vemos este gráfico tiene en general la apariencia de una distribución normal centrada en cero. Lo cual proporciona evidencias de que el supuesto de normalidad con media cero y varianza comun σ^2 , se cumple.

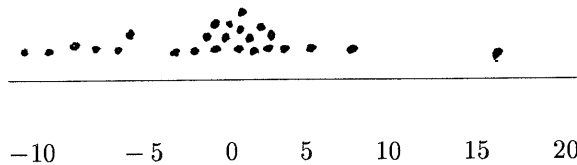


FIGURA 2.1. Histograma de puntos de todos los residuos.

Ejemplo 2.2. En el cuadro 2.3 se muestra un ejemplo de experimento en bloques al azar. Se investiga el proceso de obtención de penicilina, siendo la producción la variable de mayor interés. Hay $k=4$ variantes del proceso básico a estudiar y que denominaremos tratamientos A, B, C y D.

Se sabe que una materia prima importante, el licor de maíz, es muy variable. Afortunadamente se cuenta con suficiente licor de maíz para, llevar a cabo los $k=4$ tratamientos dentro de cada uno de los $n=5$ bloques (mezclas de licor en maíz). Se protegió el experimento de posibles sesgos extraños desconocidos realizando los tratamientos en orden aleatorio dentro de cada bloque.

En el cuadro 2.3 se exponen, además de los valores de las observaciones propiamente dichas, las medias de bloque \bar{y}_j , las medias de tratamientos \bar{y}_i y la media general $\bar{y}..$.

CUADRO 2.3. Resultados de un experimento de obtención de penicilina.

<u>Bloque</u>	<u>Tratamientos</u>				\bar{y}_j
	A	B	C	D	
1	89	88	97	94	92
2	84	77	92	79	83
3	81	87	87	85	85
4	87	92	89	84	88
5	79	81	80	88	82
\bar{y}_i .	84	85	89	86	86 = $\bar{y}..$

En la cuadro 2.4 se muestran los residuos de los resultados del experimento de la obtención de la penicilina, los residuos en un análisis de diseño en bloques al azar son

CUADRO 2.4. Residuos del experimento de obtención de penicilina.

<u>Bloque</u>	<u>Tratamientos</u>			
	A	B	C	D
1	-1	-3	+2	+2
2	+3	-5	+6	-4
3	-2	+3	-1	0
4	+1	+5	-2	-4
5	-1	0	-5	+6

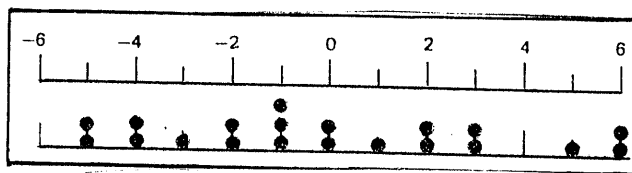


FIGURA 2.2. Gráfico de todos los puntos de los residuos.

En la figura 2.2 no se observa nada de interés, por lo que podríamos decir que no se violan las suposiciones de normalidad, sin embargo, por su forma de distribución de residuos hay que aplicar otras pruebas de normalidad para corroborar que la muestra es extraída de una distribución normal centrada en cero.

Gráfica de probabilidad normal de los residuos.

El método consiste en construir una gráfica de probabilidad normal de los residuos. Una gráfica de este tipo es la representación de la distribución acumulada de los residuos sobre papel de probabilidad normal, en otras palabras, es un papel para gráficas cuya escala de ordenadas es tal que la distribución normal acumulada sea una recta.

Para construir una gráfica de probabilidad normal, deben disponerse los residuos en orden ascendente y graficar el k -ésimo de estos residuos ordenados contra su punto de probabilidad acumulada

$$P_k = \frac{(k - \frac{1}{2})}{N}$$

sobre papel de probabilidad normal.

Si la distribución de los errores es normal, la gráfica de distribución acumulada, es aproximadamente lineal (Nater et al., 1989). Al visualizar dicha línea hay que poner más énfasis en los valores centrales de la gráfica que en los extremos (Montgomery, 1991).

Un defecto que a menudo se muestra en las gráficas de probabilidad normal es un residuo que es mucho mayor que los otros. Este residuo se denomina comúnmente residuo inusitado. La presencia de uno o más residuos inusitados puede distorsionar seriamente el análisis de varianza.

Si los errores ϵ_{ij} son $N(0, \sigma^2)$, los residuos estandarizados deben ser aproximadamente normales con media cero y varianza igual a uno.

Si es válida la suposición de que los errores tienen las mismas varianzas para todos los niveles de k , entonces una gráfica de los residuos de cada tratamiento no revelará ninguna diferencia apreciable en la dispersión de los residuos alrededor del cero. Si esta dispersión es notablemente diferente para algunos tratamientos, entonces es posible que las varianzas no sean iguales para todos los tratamientos. Para normalizar la escala de magnitudes de los residuos es preferible emplear los residuos estandarizados d_{ij} (Canavos, 1986).

Una manera informal de detectar residuos distanciados es analizar los residuos estandarizados

$$d_{ij} = \frac{e_{ij}}{\sqrt{CM_E}}$$

donde CM_E es el cuadrado medio del error del análisis de varianza correspondiente.

Si los errores ϵ_{ij} son $N(0, \sigma^2)$, los residuos estandarizados deben ser aproximadamente normales con media cero y varianza igual a uno. Por lo tanto, aproximadamente el 68 por ciento de los residuos estandarizados deben encontrarse entre los límites de ± 1 , alrededor del 95 por ciento entre ± 2 , y prácticamente

todos deben estar entre ± 3 . Un residuo a una distancia mayor que tres o cuatro desviaciones estándar del origen es potencialmente un residuo distanciado.

Frecuentemente la causa de un residuo distanciado es un error en los cálculos, en la codificación de los datos o en la transmisión de los mismos. Si esto no es la causa, deben investigarse cuidadosamente las circunstancias experimentales con relación a esta medición.

Ejemplo 2.3. Un ingeniero de desarrollo de productos está interesado en maximizar la resistencia a la tensión de una nueva fibra sintética que se empleará en la manufactura de tela para camisas de hombre. El ingeniero sabe por experiencia que la resistencia es influida por el porcentaje de algodón presente en la fibra. Además, él sospecha que elevar el contenido de algodón incrementará la resistencia, al menos inicialmente.

El ingeniero decide probar muestras (o probetas) a cinco niveles de porcentaje de algodón: 15, 20, 25, 30 y 35 por ciento. Asimismo, decide ensayar cinco muestras a cada nivel de contenido de algodón. En el cuadro 2.5 muestra los valores de los resultados de los experimentos de la resistencia a la tensión de la fibra sintética.

CUADRO 2.5. Datos del experimento de resistencia a la tensión en lb/plg^2

<u>Tratamientos</u>	<u>Repeticiones</u>					\bar{y}_i
	1	2	3	4	5	
15	07	07	15	11	09	09.8
20	12	17	12	18	18	15.4
25	14	18	18	19	19	17.6
30	19	25	22	19	23	21.6
35	07	10	11	15	11	10.8

En el cuadro 2.6 se tabulan los residuos para el análisis correspondiente de la forma

$$e_{ij} = y_{ij} - \bar{y}_i.$$

CUADRO 2.6. Residuos de los datos del experimento de resistencia a la tensión.

<u>Tratamientos</u>	<u>Repeticiones</u>				
	1	2	3	4	5
15	-2.8	-2.8	+5.2	+1.2	-0.8
20	-3.4	+1.6	-3.4	+2.6	+2.6
25	-3.6	+0.4	+0.4	+1.4	+1.4
30	+2.6	+3.4	+0.4	-2.6	+1.4
35	-3.8	-0.8	+0.2	+4.2	+0.2

En el cuadro 2.7 los residuos se organizan en orden ascendente y se calculan sus puntos de probabilidad acumulada P_k .

CUADRO 2.7. Residuos ordenados y puntos de probabilidad.

<u>Orden</u>	<u>Residuo</u>	$P_k =$	<u>Orden</u>	<u>Residuo</u>	$P_k =$
k	e_{ij}	$\frac{(k-\frac{1}{2})}{25}$	k	e_{ij}	$\frac{(k-\frac{1}{2})}{25}$
1	-3.8	0.0200	14	+0.4	0.5400
2	-3.6	0.0600	15	+0.4	0.5800
3	-3.4	0.1000	16	+1.2	0.6200
4	-3.4	0.1400	17	+1.4	0.6600
5	-2.8	0.1800	18	+1.4	0.7000
6	-2.8	0.2200	19	+1.4	0.7400
7	-2.8	0.2600	20	+1.6	0.7800
8	-2.6	0.3000	21	+2.6	0.8200
9	-0.8	0.3400	22	+2.6	0.8600
10	-0.8	0.3800	23	+3.4	0.9000
11	+0.2	0.4200	24	+4.2	0.9400
12	+0.2	0.4600	25	+5.2	0.9800
13	+0.4	0.5000			

En la figura 2.3 aparece la gráfica de probabilidad normal con los residuos graficados contra $P_k \times 100$ en la escala vertical derecha. Un examen de esta representación gráfica produce la impresión general de una distribución ligeramente sesgada, con el extremo derecho más largo que el izquierdo. La tendencia de la gráfica de probabilidad normal a desviarse hacia abajo en el lado izquierdo implica que el extremo izquierdo de la distribución del error es más reducido que lo esperado en una distribución normal; en otras palabras, que los errores negativos no son tan grandes (en valor absoluto) como se esperaba. Sin embargo, esta gráfica no es demasiado anormal, porque nuestra línea es aproximadamente rectilínea.

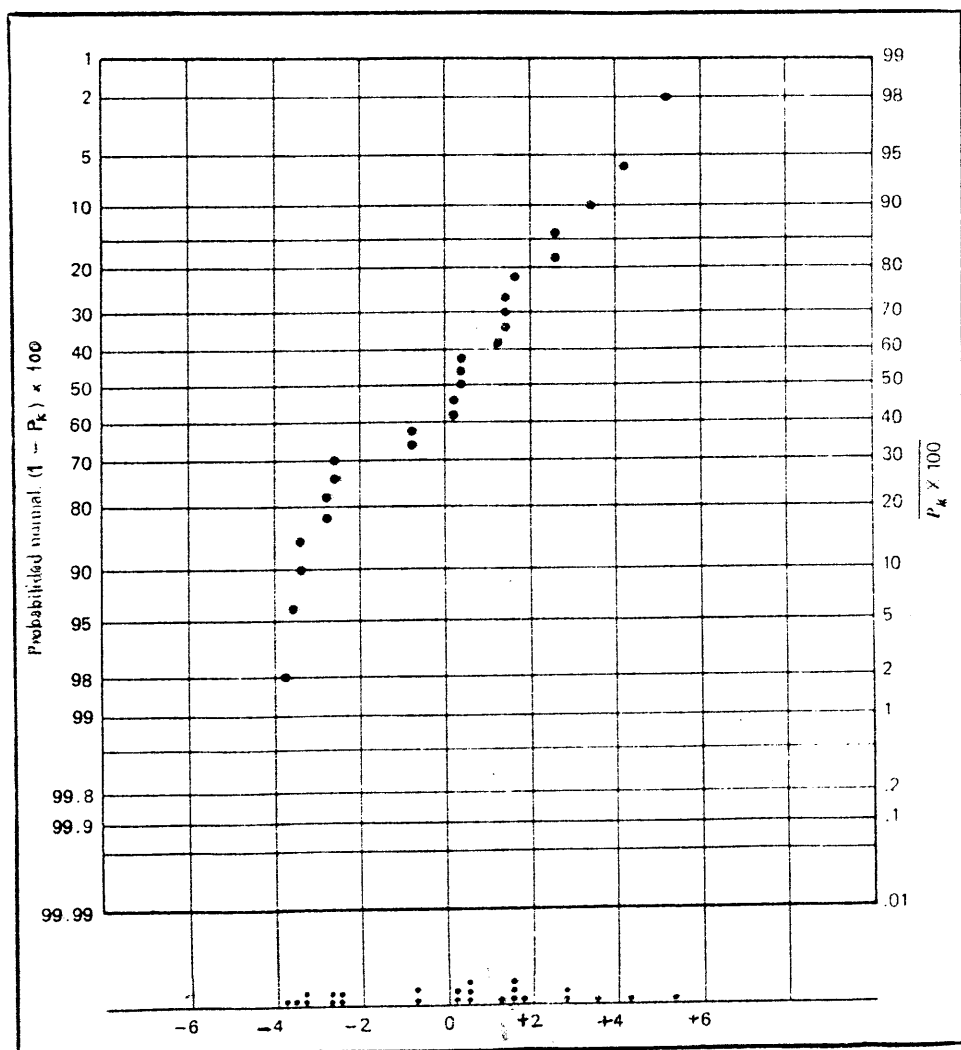


FIGURA 2.3. Gráfica de probabilidad normal de los residuos.

La gráfica de probabilidad normal no muestran algún residuo distanciado. Más aún, el residuo estandarizado máximo es

$$d_{13} = \frac{e_{13}}{\sqrt{CM_E}} = \frac{5.2}{8.06} = 1.83$$

lo cual no debe causar inquietud.

Prueba de bondad de ajuste.

Una prueba de bondad de ajuste se emplea para decidir cuando un conjunto de datos se apega a una distribución de probabilidad dada.

Esta prueba fue propuesta por Karl Pearson en 1900 antes que se desarrollase la teoría general de las pruebas de hipótesis. El criterio de esta prueba es (Mood, et al., 1974)

$$u = \sum \frac{(n_i - np_{1i})^2}{np_{1i}},$$

la cual, para muestras grandes, tiene aproximadamente la distribución ji cuadrado con $k-1$ grados de libertad, si H_0 es verdadera. Tenemos que

$u = \chi^2 =$ ji cuadrado calculado.

$n_i = f_o =$ frecuencias observadas.

$np_{1i} = f_e =$ frecuencias esperadas.

Por lo que el estadístico χ^2 calculado de los datos muestrales está dado por

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Si el χ^2 calculado es mayor que el valor tabulado de χ^2 al nivel de significación especificado y los grados de libertad, la hipótesis nula, H_0 se rechaza en favor de la hipótesis alternativa H_1 .

Los grados de libertad están dados por

$$gl = k - m - 1,$$

donde k = número de categorías; m = número de parámetros de población estimados a partir de estadísticos muestrales.

Cuando la frecuencia esperada de una categoría es menor a cinco, la categoría debe ser combinada con una adyacente.

El número de clases (k) que se emplea para clasificar los datos en un conjunto depende del total de observaciones en éste. Si el número de observaciones es relativamente pequeño, el número de clases a emplear será cercano a cinco, pero generalmente nunca menor que este valor. Si existe una cantidad sustancial de datos, el número de clases debe encontrarse entre ocho y doce y generalmente no existirán más de quince clases. Un número muy pequeño de clases puede ocultar la distribución real del conjunto de datos, mientras que un número muy grande puede dejar sin observaciones a algunas de las clases, limitando de esta forma su uso (Canavos, 1986). Hay dos métodos:

1) $k = \sqrt{n}$

2) Por medio de tabla

n	k
000-050	05-07
051-100	06-10
101-250	07-12

Ejemplo 2.4. En la manufactura de tabletas se usaron cuatro tipos de tabletas de trisilicato de magnesio los cuales son

1) Tipo A, magnesio (el tamaño de grano en la malla es 16).

- 2) Tipo B, talco en polvo (el tamaño de grano en la malla es 16).
 3) Tipo C, petróleo líquido (el tamaño de grano en la malla es 16).
 4) Tipo D, magnesio (el tamaño de grano en la malla es 20).

Fueron completamente aleatorizados en la manufactura y en las pruebas. Diez tabletas fueron probadas para el tiempo de desintegración en segundos para cada una de los cuatro tipos de tabletas. Los datos son presentados en el cuadro 2.8. Para hacer los cálculos correspondientes ver Cantú, 1980.

CUADRO 2.8. Tiempo (segundos) de desintegración de las tabletas para los cuatro tipos.

	<u>Tipos de tabletas.</u>							
	<u>A</u>		<u>B</u>		<u>C</u>		<u>D</u>	
	020	042	008	012	050	124	151	178
	028	025	010	024	067	072	125	151
	036	024	012	010	090	078	180	152
	016	031	016	019	103	070	149	161
	025	033	009	010	090	076	175	118
<u>Media</u>	028		013		082		154	
<u>Varianza</u>	059.55		026.22		430.89		436.226	

$n =$ número de datos = 40.

$x_{máx} =$ dato máximo = 180.

$$x_{\min} = \text{dato m\u00ednimo} = 8.$$

$$R = \text{rango de los datos} = x_{\max} - x_{\min} = 172.$$

Para el n\u00famero de intervalos (clase, barras o agrupamiento)

$$1) k = \sqrt{n} = \sqrt{40} = 6.3 \cong 6.$$

2) Por medio de tabla

n	k
000-050	05-07
051-100	06-10
101-250	07-12

entonces k (n\u00famero de intervalos, clases, barras o agrupamiento) = 6.

$$AI = \text{amplitud del intervalo} = \frac{R}{k} = 28.$$

$$UF = \text{unidad de frecuencia} = 1.$$

$$LI = \text{l\u00edmite inferior} = x_{\min} - UF = 7.$$

$$LS = \text{l\u00edmite superior (del rango)} = LI + AI = 35.$$

CUADRO 2.9. Cálculo de \bar{x} y s

para los tiempos de desintegración de las tabletas para los cuatro tipos.

<i>Intervalo</i>	<i>Marca</i>				
<i>Clase</i>	<i>Clase X</i>	f_o	$f_o X$	X^2	$f_o X^2$
007 – 035	022.0	19	0418	00484.00	009196.0
036 – 064	050.0	02	0100	02500.00	005000.0
065 – 093	079.0	07	0553	06241.00	043687.0
094 – 122	108.0	02	0216	11664.00	023328.0
123 – 151	137.0	06	0822	18769.00	112614.0
152 – 180	166.0	04	0664	27556.00	110224.0
<i>SUMAS</i>	562.0	40	2773	67214.00	304049.0

Para nuestra distribución calculamos \bar{x} y s .

$$\bar{x} = \frac{\sum f_o X}{n} = \frac{2773}{40} = 69.325$$

$$s = \sqrt{\frac{\sum f_o X^2 - n\bar{x}^2}{n - 1}}$$

$$s = \sqrt{\frac{304049 - (40)(69.325)^2}{39}} = 53.543841$$

CUADRO 2.10. Frecuencias esperadas para los tiempos de desintegración

de las tabletas para los cuatro tipos usando $\bar{x} = 69.325$ y $s = 53.543841$.

<i>x = clase superior</i>	$z = \frac{x-69.325}{53.5438}$	<i>Area a la izq de x</i>	<i>Area del interv. de clase</i>
≤ 035	- 0.6411	0.2611	$0.2611 \times 40 = 10.444$
064	- 0.0995	0.4641	$0.2030 \times 40 = 8.120$
093	+ 0.4422	0.6700	$0.2059 \times 40 = 8.236$
122	+ 0.9838	0.8365	$0.1665 \times 40 = 6.660$
151	+ 1.5254	0.9370	$0.1005 \times 40 = 4.020$
> 180	+ 2.0670	0.9821	$0.0451 \times 40 = 1.804$
<i>SUMAS</i>			$\approx 1.0000 \quad \approx 40.000$

entonces,

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \left[\frac{(19 - 10.444)^2}{10.444} + \frac{(02 - 8.120)^2}{8.120} + \frac{(07 - 8.236)^2}{8.236} + \frac{(02 - 6.660)^2}{6.660} + \frac{((6 + 4) - (4.020 + 1.804))^2}{4.020 + 1.804} \right] = 18.0623.$$

Nótese que las últimas dos clases de las frecuencias observadas y esperadas se combinaron porque la $f_e < 5$, y los grados de libertad son $gl = k - m - 1 = 6 - 2 - 1 = 3$. Como se estimaron dos parámetros de la población (μ y σ con \bar{x} y s , respectivamente), $m = 2$. El valor tabulado de χ^2 con $\alpha = 0.05$ y $gl = 3$ es 7.81.

Como el valor de χ^2 calculado es mayor que el valor tabulado, rechazamos H_0 . Es decir, rechazamos la hipótesis de que la muestra aleatoria de los datos de

los tiempos de desintegración de las tabletas viene de una distribución normal con $\mu = 69.325$ y $\sigma = 53.5438$.

Prueba de Shapiro and Wilk.

Esta prueba es llamada también la prueba W para normalidad (muestras completas) y fue desarrollada por Shapiro y Wilk (1965).

Para calcular el valor de W , dado una muestra aleatoria completa de tamaño n , con x_1, x_2, \dots, x_n , el procedimiento adecuado es

1) Ordenar las n observaciones para obtener una muestra ordenada como sigue

$$y_1 \leq y_2 \leq \dots \leq y_n.$$

2) Calcular

$$S^2 = \sum_{i=1}^n (y_i - \bar{y})^2.$$

3) (i) Si n es par, $n = 2k$, calcule

$$b = \sum_{i=1}^k a_{n-i+1} (y_{n-i+1} - y_i),$$

donde los valores de a_{n-i+1} son dados en el apéndice A.6. (ii) Si n es impar, $n = 2k + 1$, entonces omitimos el valor medio de y_{k+1} , y calculamos

$$b = \sum_{i=1}^k a_{n-i+1} (y_{n-i+1} - y_i) = a_n (y_n - y_1) + \dots + a_{k+2} (y_{k+2} - y_k).$$

donde los valores de a_{n-i+1} son dados en la tabla A.6.

4) Calcule $W = \frac{b^2}{S^2}$.

5) 1, 2, 5, 10, 50, 90, 95, 98 y 99 por ciento de nivel de la distribución de W son dados en la tabla A.7. Valores pequeños de W son significativos, esto es,

indican no normalidad. Es decir, la hipótesis de normalidad es rechazada cuando el valor calculado W es menor que el valor dado en la tabla A.7.

Ejemplo 2.5. Tomemos los datos de los tipos de tabletas, del ejemplo 2.4. Escogemos el tipo B, los cuales son, 8, 10, 12, 16, 9, 12, 24, 10, 19, 10.

1) Ordenamos las observaciones: 8, 9, 10, 10, 10, 12, 12, 16, 19, 24.

2) Calcular

$$S^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = 236.$$

3) Como $n = 10$ es par (i), $n = 2k = 2(5) = 10$, calcule

$$b = \sum_{i=1}^k a_{n-i+1}(y_{n-i+1} - y_i),$$

donde los valores de a_{n-i+1} son dados en la tabla A.6, entonces

$$b = (0.5739)(24 - 8) + (0.3291)(19 - 9) + \dots + (0.0399)(12 - 10) = 14.0826.$$

4) Calcule

$$W = \frac{b^2}{S^2} = \frac{(14.08)^2}{236} = 0.840 \quad (\text{calculado}).$$

5) De la tabla A.7, tenemos que $W = (\alpha, n) = (0.05, 10) = 0.842$.

Entonces, como W calculado es menor que W tabulado, rechazamos la hipótesis de normalidad a un nivel $\alpha = 0.05$.

Ejemplo 2.6. Tenemos los siguientes datos

$$y_1 = 148, \quad y_2 = 154, \quad y_3 = 158, \quad y_4 = 160,$$

$$y_5 = 161, \quad y_6 = 162, \quad y_7 = 165, \quad y_8 = 170,$$

$$y_9 = 182, \quad y_{10} = 195, \quad y_{11} = 236.$$

2) Calcular

$$S^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = 6239.$$

3) Como $n = 11$ es impar (ii), $n = 2k + 1 = 2(5) + 1 = 11$, entonces

$$b = a_n(y_n - y_1) + \dots + a_{k+2}(y_{k+2} - y_k).$$

$$b = 0.56(236 - 148) + 0.33(195 - 154) + 0.23(182 - 158)$$

$$+ 0.14(170 - 160) + 0.07(165 - 161)$$

$$b = 70$$

4) Calcule

$$W = \frac{b^2}{S^2} = \frac{(70)^2}{6239} = 0.79 \quad (\text{calculado}).$$

5) En la tabla A.7 tenemos que, $W = (\alpha, n) = (0.05, 11) = 0.850$.

Entonces, como W calculado es menor que W tabulado, rechazamos la hipótesis de normalidad a un nivel $\alpha = 0.05$.

Prueba de Kolmogorov-Smirnov.

La prueba de Kolmogorov-Smirnov no necesita que los datos se encuentren agrupados y es aplicable a muestras de tamaño pequeño. Esta se basa en una comparación entre las funciones de distribución acumulativa que se observan en la muestra ordenada y la distribución propuesta bajo la hipótesis nula. Si esta comparación revela una diferencia suficientemente grande entre las funciones de distribución muestral y propuesta, entonces la hipótesis nula de que la distribución es $F_0(x)$, se rechaza.

Considérese la hipótesis nula

$$H_0 : F(x) = F_0(x),$$

en donde el modelo de probabilidad propuesto $F_0(x)$ se encuentra especificado, de manera completa, con respecto a todos los parámetros. Denótese por x_1, x_2, \dots, x_n a las observaciones ordenadas de una muestra aleatoria de tamaño n y defínase la función de distribución acumulativa muestral como

$$S_n(x) = \begin{cases} 0, & x < x_1, \\ k/n, & x_k \leq x < x_{k+1}, \\ 1, & x \geq x_n, \end{cases} \quad (2.6.1)$$

en otras palabras, para cualquier valor ordenado x de la muestra aleatoria, $S_n(x)$ es la proporción del número de valores en la muestra que son iguales o menores a x . Ya que $F_0(x)$ se encuentra completamente especificada, es posible evaluar a $F_0(x)$ para algún valor deseado de x , y entonces comparar este último con el valor correspondiente de $S_n(x)$. Si la hipótesis nula es verdadera, entonces es lógico esperar que la diferencia sea relativamente pequeña.

El estadístico de Kolmogorov-Smirnov se define como

$$D_n = \max_x |S_n(x) - F_0(x)|,$$

el estadístico D_n tiene una distribución que es independiente del modelo propuesto bajo la hipótesis nula. Por esta razón, se dice D_n es un estadístico independiente de la distribución. Lo anterior da como resultado que la función de distribución de D_n pueda evaluarse sólo en función del tamaño de la muestra y después usarse para cualquier $F_0(x)$. En la tabla A.4, se proporcionan los valores cuantiles superiores de D_n para varios tamaños de la muestra. Debe notar que los valores asintóticos de D_n que se encuentran en la parte inferior de la tabla proporcionan una adecuada aproximación para valores de n mayores de 50.

Para un tamaño α del error tipo I, la región crítica es de la forma

$$P\left(D_n > \frac{c}{\sqrt{n}}\right) = \alpha,$$

de acuerdo con lo anterior, la hipótesis H_0 se rechaza si para algún valor x observado, el valor D_n se encuentra dentro de la región crítica de tamaño α .

El estadístico de Kolmogorov-Smirnov es, en general, superior a la prueba de bondad de ajuste ji cuadrada cuando los datos involucran una variable aleatoria continua, debido a que no es necesario agrupar los datos. Además la prueba de Kolmogorov-Smirnov tiene la propiedad de ser aplicable a muestras de tamaño pequeño.

Por otro lado, el estadístico se encuentra limitado, ya que el modelo propuesto bajo H_0 debe especificarse en forma completa.

Ejemplo 2.7. Tomemos los datos del ejemplo 2.1.

CUADRO 2.1. Altura total, en centímetros de plantas de rábano.

	<u>Tratamientos</u>				
<u>Repeticiones</u>	1	2	3	4	5
1	31.33	25.66	34.16	31.33	49.00
2	35.50	31.66	32.00	34.50	42.33
3	36.66	24.00	22.50	35.50	37.00
4	33.66	36.00	27.33	37.33	38.66
5	23.66	40.00	51.33	30.33	44.66
\bar{y}_i .	32.16	31.46	33.46	33.60	42.33

Para lo cual (ver ejemplo 2.4), tenemos que

$$n = 25, x_{m\acute{a}x} = 51.33, x_{m\acute{i}n} = 22.5, Rango = 28.83, k = 5.$$

$$Amplitud = \frac{28.83}{5} \approx 5.78.$$

Luego hacemos los calculos siguientes

CUADRO 2.11. Cálculo de \bar{x} y s .

<i>Intervalo</i>	\bar{X}	f_o	$f_o \cdot X$	$f_o \cdot X^2$
22.4-28.18	25.29	5	126.45	3197.9205
28.19-33.97	31.08	6	186.48	5795.7984
33.98-39.75	36.86	9	331.74	12227.936
39.76-45.54	42.65	3	127.95	5457.0675
45.55-51.33	48.44	2	96.88	4692.8672
SUMAS	184.32	25	869.5	31371.59

Para nuestra distribución calculamos \bar{x} y s .

$$\bar{x} = \frac{\sum f_o X}{n} = \frac{869.5}{25} = 34.78$$

$$s = \sqrt{\frac{\sum f_o X^2 - n\bar{x}^2}{n-1}} = \sqrt{\frac{31371.59 - (25)(34.78)^2}{24}} = 6.86$$

y los valores correspondientes del modelo normal propuesto se obtienen estandarizando primero a $N(0,1)$ y empleando la tabla A.1, tenemos que la z es

$$z = \frac{x - \bar{x}}{s} = \frac{x - 34.78}{6.68}$$

CUADRO 2.12. Frecuencias esperadas.

Límite Sup. (x)	z	Area de x	Area de $F_0(x)$	f_e
≤ 28.18	-0.9617	0.1685	0.1685	4.2125
33.97	-0.1180	0.4522	0.2837	7.0925
39.75	0.7242	0.7673	0.3151	7.8775
45.54	1.5676	0.9418	0.1745	4.3625
> 51.33	2.4115	0.9920	0.0502	1.2550
SUMAS			≈ 1.0000	≈ 24.0000

CUADRO 2.13. Cálculo de la prueba de Kolmogorov-Smirnov.

f_o	f_o Acumululada	$S_n(x)$	$F_0(x)$	$ S_n(x) - F_0(x) $
5	05	0.20	0.1685	0.0315
6	11	0.44	0.4522	0.0122
9	20	0.80	0.7673	0.0327
3	23	0.92	0.9418	0.0218
2	25	1.00	0.9920	0.0080

Se observa que la máxima desviación es de 0.0327. De la la tabla A.4, el valor crítico de D_{25} para $\alpha=0.05$ es 0.2700. Dado que $0.0327 < 0.2700$, no puede

rechazarse la hipótesis nula al nivel de cinco por ciento de significancia, y se concluye que la muestra viene de una población de distribución normal.

Ejemplo 2.8. Se tiene una muestra de tamaño 12 y se desea probar si proviene de una distribución normal. Los valores de la muestra son

76.5, 73.9, 74.7, 76.0, 74.6, 75.2, 77.3, 76.6, 77.4, 74.2, 77.7, 76.9.

ordenando tenemos

73.9, 74.2, 74.6, 74.7, 75.4, 76.0, 76.5, 76.6, 76.9, 77.3, 77.4, 77.7.

En el cuadro 2.14 se ilustra como se calcula $|S_n(x) - F_0(x)|$.

Considérese la prueba de la siguiente hipótesis nula

$$H_0 : F(x) = F_0(x),$$

donde $F_0(x)$ la función de distribución normal con media 76.25 y desviación estándar 1.33. La función de distribución muestral se obtiene mediante el empleo de (2.6.1) para los valores ordenados. Lo anterior involucra un incremento de $1/12 = 0.0833$ al valor previo de la distribución muestral. Los valores correspondientes del modelo normal se obtienen estandarizando primero a $N(0,1)$ y empleando la tabla A.1 de la distribución normal, con

$$z = \frac{x - 76.25}{1.33}$$

CUADRO 2.14. Cálculo de la prueba de Kolmogorov-Smirnov.

x	z	$F_0(x)$	Frec. Acu.	$S_n(x)$	$ S_n(x) - F_0(x) $
73.9	-1.7669	0.0384	01	0.0833	0.0449
74.2	-1.5414	0.0618	02	0.1666	0.1048
74.6	-1.2406	0.1075	03	0.2500	0.1425
74.7	-1.1654	0.1210	04	0.3333	0.2123
75.4	-0.6391	0.2611	05	0.4166	0.1555
76.0	-0.1879	0.4247	06	0.5000	0.0753
76.5	+0.1879	0.5753	07	0.5833	0.0080
76.6	+0.2632	0.6026	08	0.6666	0.0640
76.9	+0.4887	0.6879	09	0.7500	0.0621
77.3	+0.7895	0.7852	10	0.8333	0.0481
77.4	+0.8647	0.8078	11	0.9166	0.1088
77.7	+1.0902	0.8621	12	1.0000	0.1373

Se observa que $D_n=0.2123$. De la tabla A.4, el valor crítico de D_{12} para $\alpha=0.05$ es 0.3750. Dado que $0.2123 < 0.3750$, no puede rechazarse la hipótesis nula al nivel de cinco por ciento de significancia, y se concluye que la muestra viene de una población de distribución normal.

3. INDEPENDENCIA.

Conceptos Preliminares.

La suposición de independencia de errores no correlacionados es una suposición crucial y su importancia no debe ser subestimada. Por supuesto, utilizando el artificio de aleatorización, el investigador puede hacer todo lo posible para ver que la correlación entre errores no favorecerá continuamente ningún tratamiento particular (o posterior).

Las varianzas y las medias de las distintas muestras no están correlacionadas. En algunos datos existe una relación definida entre las medias de las muestras y sus varianzas. Este es un caso especial y la causa más común de heterogeneidad de varianza. Una correlación positiva entre medias y varianzas suele encontrarse cuando existe un amplio rango de medias de la muestra. Afortunadamente, esto no sucede con frecuencia; a menudo podemos transformar los datos en forma tal que el supuesto de independencia entre varianza y media sea válido. Luego podemos proceder con el análisis de varianza de los datos transformados (Little y Hills, 1990).

Gráfica de Residuos contra la Secuencia de Tiempo.

Graficar los residuos contra el orden del tiempo en el que se recopilan los datos es útil para detectar alguna correlación entre ellos. Una tendencia a tener rachas con residuos positivos y negativos indica una correlación positiva. Esto implica que la suposición de independencia de los residuos ha sido violada. Potencialmente éste es un problema serio y difícil de corregir; por ello es importante

prevenir el problema cuando los datos son recopilados. Un paso importante para asegurar la independencia es realizar un procedimiento apropiado de aleatorización.

Algunas veces la habilidad del experimentador (o de los operadores) cambia a medida que el experimento se desarrolla, o puede ocurrir que el proceso se desvíe, o se vuelva más errático. A menudo, esto produce un cambio en la varianza del error al transcurrir el tiempo. Esta situación conduce a una gráfica de residuos contra el tiempo que exhibe mayor dispersión en un extremo que en el otro. Una varianza variable es un problema potencial serio. En la figura 3.1 aparece la forma aproximada para este modelo.

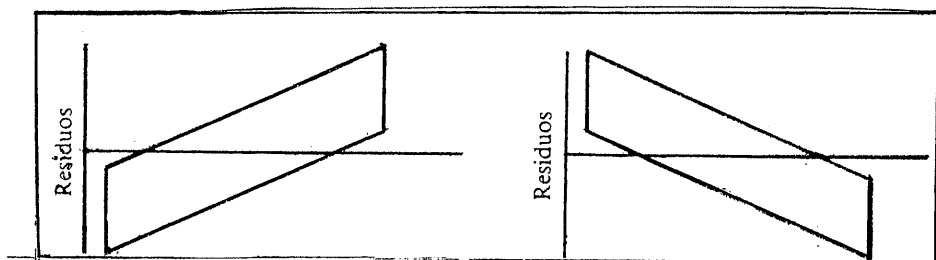


Figura 3.1 Modelo aproximado de la gráfica.

de Residuos contra la Secuencia de Tiempo.

Ejemplo 3.1. En el cuadro 3.1 se muestran los datos de la sucesión del tiempo en que fueron recopilados los datos de la resistencia a la tensión del ejemplo 2.3.

CUADRO 3.1. Orden en que se recolectaron los datos de la tabla 2.5.

<u>Tratamientos</u>	<u>Repeticiones</u>				
	1	2	3	4	5
15	15	19	25	12	06
20	08	14	01	11	03
25	18	13	20	07	09
30	22	05	02	24	10
35	17	21	04	16	23

En el cuadro 2.6 se muestran los residuos los datos de la resistencia a la
 a. Una Gráfica de los residuos en contra de el tiempo que aparece en la figura
 por lo cual no existe razón alguna para suponer que no se cumple la suposición
 dependencia o de varianza constante.

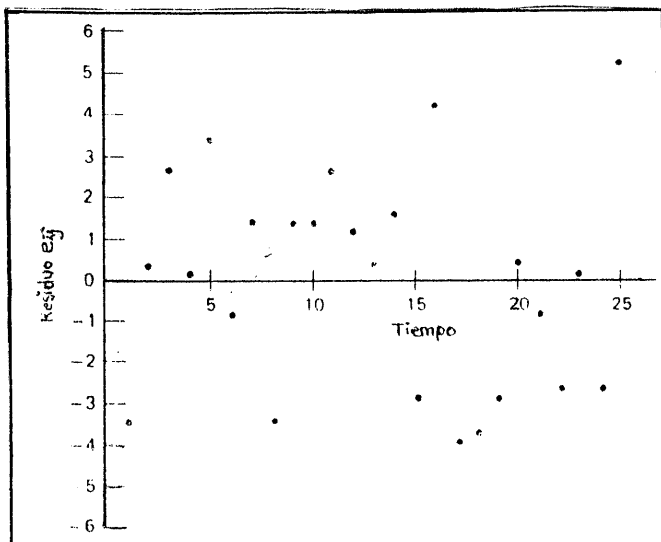


Figura 3.2. Gráfica de residuos contra tiempo.

CUADRO 3.1. Orden en que se recolectaron los datos de la tabla 2.5.

<u>Tratamientos</u>	<u>Repeticiones</u>				
	1	2	3	4	5
15	15	19	25	12	06
20	08	14	01	11	03
25	18	13	20	07	09
30	22	05	02	24	10
35	17	21	04	16	23

En el cuadro 2.6 se muestran los residuos los datos de la resistencia a la tensión. Una Gráfica de los residuos en contra de el tiempo que aparece en la figura 3.2, por lo cual no existe razón alguna para suponer que no se cumple la suposición de independencia o de varianza constante.

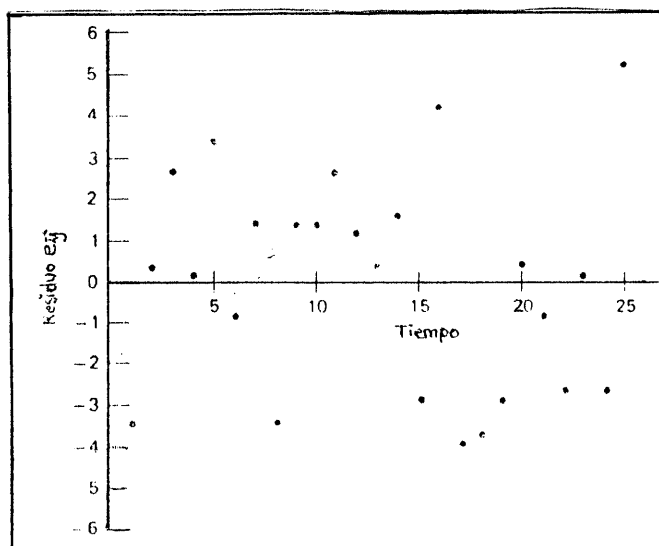


Figura 3.2. Gráfica de residuos contra tiempo.

4. ADITIVIDAD Y LINEALIDAD.

Conceptos Preliminares.

En el análisis estándar de varianza el modelo especifica que los efectos de los diferentes factores fijos (tratamientos, hileras, columnas, etc.) son aditivos, y que los errores son normales e independientes con la misma varianza.

Los efectos principales son aditivos; para cada diseño experimental existe un modelo matemático denominado Modelo Lineal Aditivo (debido a que los términos de éste se suman). El modelo para un diseño bloques al azar, por ejemplo, implica que un efecto de tratamientos es el mismo para todos los bloques y que el efecto de bloques es el mismo para todos los tratamientos, en otras palabras, si se encuentra que un tratamiento incrementa la producción de cierta cantidad promedio por encima de la media general, suponemos que éste tiene el mismo efecto en los bloques de alta producción que en los bloques de baja producción (Little y Hills, 1990).

Cuando es objetable la suposición de aditividad, el problema es un poco más complicado, esto se debe a que existen tres causas mayores de no aditividad, a saber (Ostle, 1983)

- 1) Los verdaderos efectos pueden ser multiplicativos.
- 2) Pueden existir interacciones, pero los términos que representan tales efectos no se incluyeron en el modelo supuesto.
- 3) pueden presentarse observaciones equivocadas.

La presencia de no aditividad en los datos conduce a una aparente heterogeneidad del error debido a supuestos falsos cuando no se efectúa una transformación antes del análisis (Steel y Torrie, 1986).

Prueba de Tukey de aditividad.

Tukey (1949) propone un método para aislar una suma de cuadrados del error con el propósito de probar la no aditividad; ésta tiene un grado de libertad. El método originalmente propuesto es aplicable a la clasificación de dos vías o diseño de bloques completos aleatorizados. También se ha ideado un método para diseños de cuadro latino.

Esta es muy útil en diversas maneras

- 1) Ayuda a decidir si la transformación es necesaria.
- 2) Sugiere la transformación apropiada.
- 3) informa si la transformación ha tenido éxito para producir aditividad.

Para clasificación de dos vías o diseño de bloques completos aleatorizados y si los efectos son exactamente aditivos en la escala de y , tenemos

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})$$

$$y_{ij} = \bar{y}_{..} \left[\frac{1 + \{(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})\}}{\bar{y}_{..}} \right]$$

Estamos suponiendo que los efectos de hileras y de columnas son pequeños en relación a la media. Esto implica que $\alpha_i = \frac{(\bar{y}_{i.} - \bar{y}_{..})}{\bar{y}_{..}}$ y $\beta_j = \frac{(\bar{y}_{.j} - \bar{y}_{..})}{\bar{y}_{..}}$ son también pequeñas.

Ahora $x_{ij} = y_{ij}^{\frac{1}{p}}$ donde p es la potencia y se expande por la serie usual de Taylor. Esto da

$$x_{ij} = \bar{y}_{..}^{\frac{1}{p}} [1 + \alpha_i + \beta_j]^{\frac{1}{p}}$$

$$x_{ij} = \bar{y}_{..}^{\frac{1}{p}} \left[1 + \frac{1}{p}(\alpha_i + \beta_j) + \frac{1}{p} \frac{(1-p)}{p} \frac{1}{2}(\alpha_i^2 + 2\alpha_i\beta_j + \beta_j^2) + \dots \right]$$

Ahora en la escala de x , los términos de α_i , α_i^2 representan los efectos de hilera y los términos en β_j , β_j^2 representan los efectos de columna que se suman en la expresión anterior. Estos términos son aditivos en la escala de x . El primer término no aditivo es el que tiene α_i , β_j . Anotando totalmente este término es

$$\frac{\bar{y}_{..}^{\frac{1}{p}}(1-p)(\bar{y}_i - \bar{y}_{..})}{p^2 \bar{y}_{..}^2} \quad (4.2.1)$$

Para nuestros fines es preciso anotar esta expresión en términos de x en lugar de y . Por expansiones de Taylor de término único nuevo, tenemos, puesto que $y = x^p$

$$\bar{y}_i - \bar{y}_{..} \doteq p\bar{x}_{..}^{p-1}(\bar{x}_i - \bar{x}_{..});$$

$$\bar{y}_j - \bar{y}_{..} \doteq p\bar{x}_{..}^{p-1}(\bar{x}_j - \bar{x}_{..}).$$

Por sustitución en (4.2.1) nos da el primer término no aditivo, en x_i .

$$\frac{(1-p)\bar{y}_{..}^{\frac{1}{p}}(\bar{y}_i - \bar{y}_{..})(\bar{y}_j - \bar{y}_{..})\bar{x}_{..}^{2p-2}}{\bar{y}_{..}^2}$$

Haciendo uso de $\bar{y}_{..} \doteq \bar{x}_{..}^p$, este término puede expresarse aproximadamente como

$$\frac{(1-p)}{\bar{y}_{..}}(\bar{y}_i - \bar{y}_{..})(\bar{y}_j - \bar{y}_{..})$$

Como este término representa un efecto no aditivo de hileras y columnas aparecerá en el residual de x_{ij} , cuando se ajusta un modelo aditivo en la escala de x . Las conclusiones de este razonamiento son como sigue

1) Si este tipo de no aditividad está presente en x , y \hat{x}_{ij} es el valor ajustado dado por el modelo aditivo, el residual $x_{ij} - \hat{x}_{ij}$ tiene regresión lineal en la variante $(\bar{y}_i - \bar{y}_{..})(\bar{y}_j - \bar{y}_{..})$.

2) El coeficiente de regresión r es estimado por $\frac{(1-p)}{\bar{x}_{..}}$. Así, la potencia p a la cual x habrá de elevarse para producir aditividad, está estimada por $(1 - r\bar{x}_{..})$.

3) La prueba de Tukey se utiliza para probar hipótesis nula, de que el valor poblacional de r es cero.

La prueba está relacionada a las transformaciones de forma $y = x^p$, en que x está en la escala original, y buscamos una potencia p para x tal que los efectos sean aditivos en la escala de $y = x^p$. Así

$p = \frac{1}{2}$, representa la transformación de la raíz cuadrada.

$p = -1$, representa la transformación de recíprocas, analizando $\frac{1}{x}$ en lugar de x .

$p = 0$, representa la transformación logarítmica, porque la variable x^p se comporta como $\log x$ cuando p es pequeña.

Ejemplo 4.1. Como ejemplo, presentamos los datos que analiza Tukey (1949) en su artículo, en el cuadro 4.1, el diseño que usó es para un diseño de bloques completos al azar.

CUADRO 4.1. Ejemplo Ilustrativo.

		<u>Bloques</u>			
Tratamientos	1	2	3	4	
	1	14	2	1	2
	2	02	0	2	2
	3	02	1	5	0

El procedimiento a seguir es

1) Calcúlese $d_i = \bar{x}_i - \bar{x}_{..}$ y $d_j = \bar{x}_{.j} - \bar{x}_{..}$, de modo que ambos juegos sumen exactamente cero.

$$d_1 = 4.75 - 2.75 = +2.00,$$

$$d_2 = 1.50 - 2.75 = -1.25,$$

$$d_3 = 2.00 - 2.75 = -0.75,$$

y

$$d_{.1} = 6.00 - 2.75 = +3.25,$$

$$d_{.2} = 1.00 - 2.75 = -1.75,$$

$$d_{.3} = 2.67 - 2.75 = -0.08,$$

$$d_{.4} = 1.33 - 2.75 = -1.42.$$

2) Calcúlense

$$w_i = \sum_j x_{ij} d_{.j}$$

y anótese en la columna del extremo derecho del cuadro 4.2. Luego encuéntrese

$$N = \sum_i w_i d_i = \sum_i \sum_j x_{ij} d_i d_{.j}$$

N es el numerador de r .

$$w_{1.} = (14)(+3.25) + (02)(-1.75) + (01)(-0.08) + (02)(-1.42) = 39.08$$

$$w_{2.} = (02)(+3.25) + (00)(-1.75) + (02)(-0.08) + (02)(-1.42) = 03.50$$

$$w_{3.} = (02)(+3.25) + (01)(-1.75) + (05)(-0.08) + (00)(-1.42) = 04.35$$

Comprobar:

$$46.93 = (18)(+3.25) + (03)(-1.75) + (08)(-0.08) + (04)(-1.42).$$

Por lo que

$$N = \sum_i w_i d_i = (39.08)(02) + (03.50)(-1.25) + (04.35)(-0.75) = 70.5225.$$

CUADRO 4.2. Cálculo para la prueba de aditividad.

	01	02	03	04	x_i	\bar{x}_i	d_i	w_i
1	14	02	01	02	19	4.75	+2.00	39.08
2	02	00	02	02	06	1.50	-1.25	03.50
3	02	01	05	00	08	2.00	-0.75	04.35
x_j	+18.00	+3.00	+8.00	+4.00	33			46.93
\bar{x}_j	+06.00	+1.00	+2.67	+1.33		2.75		
d_j	+03.25	-1.75	-0.08	-1.42			0	

3) El denominador D de r es $(\sum d_i^2)(\sum d_j^2)$. Así, $r = \frac{N}{D}$.

$$\sum_i d_i^2 = (+2.00)^2 + (-1.25)^2 + (-0.75)^2 = 06.1250,$$

$$\sum_j d_j^2 = (+3.25)^2 + (-1.75)^2 + (-0.08)^2 + (-1.42)^2 = 15.6478,$$

$$D = (\sum_i d_i^2)(\sum_j d_j^2) = (06.1250)(15.6478) = 95.843.$$

4) La aportación de no aditividad a la suma de cuadrados del error de x es $\frac{N^2}{D}$, con 1 g.l.. Esto se comprueba por la prueba F con el remanente de la suma de cuadrados del error, que tiene $\{(a-1)(b-1)-1\}$ g.l.

$$S.C. \text{ para no aditividad} = \frac{N^2}{D} = \frac{(70.5225)^2}{95.843} = 51.89.$$

La prueba está hecha en la Tabla 4.3.

CUADRO 4.3. Análisis de varianza y prueba de aditividad.

Fuente Varia.	g.l.	S.C.	C.M.	F_0
Filas	2	24.50	12.20	0.866
Columnas	3	46.90	15.60	1.166
Error	6	84.80	14.10	
No Aditividad	1	51.89	51.89	7.88
Remanente	5	32.91	06.58	

Sea $F = \frac{51.89}{6.58} = 7.88$ en 1 y 5 grados de libertad es significativo a 5% de nivel. Por lo que $F_0 > F_{\alpha,1,5}$ Esto es, la hipótesis de aditividad es rechazada. Por lo que determinamos el tipo de transformación de la siguiente manera,

$$r = \frac{N}{D} = \frac{70.5225}{95.84} = 0.7358,$$

$$\hat{p} = 1 - r\bar{x}_{..} = 1 - (0.7358)(2.75) = -1.023.$$

La prueba sugiere una potencia -1 de x . Esta se comporta como x^{-1} la cual es semejante a una transformación recíproca.

5. HOMOGENEIDAD Y VARIANZA CONSTANTE.

Conceptos Preliminares.

Algunas veces uno o más tratamientos tienen varianzas que difieren del resto, aunque no hay razón para sospechar la no normalidad de errores. Cuando la varianza de error es heterogénea en esa forma, la prueba F tiende a dar demasiados resultados significativos. El disturbio es usualmente sólo moderado si cada tratamiento tiene el mismo número de repeticiones. La comparación de pares o subgrupos de medias de tratamientos puede, sin embargo, ser afectada seriamente, ya que la estimada usual de varianza de errores que abarca la varianza de todos los tratamientos, dará errores estándar que son muy grandes para algunas comparaciones y muy pequeños para otros (Snedecor y Cochran, 1975).

Las varianzas de las diferentes medias son homogéneas. Por lo general en el análisis de varianza utilizamos un promedio de n varianzas (CME) para obtener la mejor estimación de la varianza común. Pero si las varianzas dentro de tratamientos fuesen de hecho distintas, no tendríamos justificación para combinarlas ya que el promediar varianzas de tratamientos mayores y menores podría arrojar resultados engañosos. La diferencia entre dos tratamientos con varianzas grandes puede ser considerada significativa cuando en realidad ésta pudo haber ocurrido fácilmente, por casualidad. Por otro lado, la diferencia entre dos tratamientos con varianzas pequeñas puede ser declarada no significativa cuando de hecho lo era (Little y Hills, 1990).

La varianza variable también ocurre en casos cuyos datos no tienen una

distribución normal y están sesgadas, porque en las distribuciones sesgadas la varianza tiende a ser función de la media.

Si la suposición de homogeneidad no se cumple, la prueba F es afectada sólo ligeramente en los modelos balanceados de efectos fijos. Sin embargo, el problema es más serio si el diseño está desbalanceado o si una varianza es mucho mayor que las otras. En un modelo de efectos aleatorios, la desigualdad en las varianzas del error puede perturbar significativamente las inferencias sobre las componentes de varianza, aunque se use un diseño balanceado (Montgomery, 1991).

Aparte de las gráficas de residuos que frecuentemente se usan para diagnosticar la igualdad de varianza, se han propuesto algunas pruebas estadísticas. Estas son pruebas formales para la hipótesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2.$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2, \text{ para al menos una } i \neq j.$$

Anderson y McLean (1974) proponen para las pruebas de homogeneidad de varianzas con un α porcentaje de significancia se proceda como sigue

1) Si la prueba de homogeneidad es aceptada a un nivel $\alpha = 0.01$, no transformar.

2) Si la prueba de homogeneidad es rechazada a un nivel $\alpha = 0.001$, transformar.

3) Si el resultado de la prueba de homogeneidad se encuentra entre $\alpha = 0.01$ y $\alpha = 0.001$, es difícil para el investigador encontrar la distribución teórica de los datos. Si encuentra una razón práctica para transformar, adelante.

Gráfica de Residuos contra el Valor Ajustado \hat{y}_{ij} .

Si el modelo es correcto y las suposiciones se satisfacen, los residuos no deben tener algún patrón, ni deben estar relacionados con alguna otra variable, incluyendo la respuesta y_{ij} . Una comprobación sencilla consiste en graficar los residuos contra los valores ajustados \hat{y}_{ij} , recordemos que para un modelo de diseño completamente al azar,

$$\hat{y}_{ij} = \bar{y}_i,$$

y para el modelo en bloques al azar es

$$\hat{y}_{ij} = \bar{y}_i + \bar{y}_j - \bar{y}_{..},$$

y así, para los distintos modelos que se usen.

En esta gráfica no debe revelarse ningún patrón obvio. Un defecto que en ocasiones revela la gráfica es el de varianza variable. Algunas veces la varianza de las observaciones aumenta a medida que la magnitud de las observaciones lo hace. Esto resulta cuando el error es proporcional a la magnitud de las observaciones (comúnmente esto sucede en muchos instrumentos de medición; el error es proporcional a la escala de la lectura). Si éste es el caso, los residuos aumentan a medida que y_{ij} lo hace, y la gráfica de los residuos contra \hat{y}_{ij} parecerá un embudo que se ensancha o un altavoz. En la figura 5.1 se muestra la forma de la gráfica de residuos contra valores ajustados.

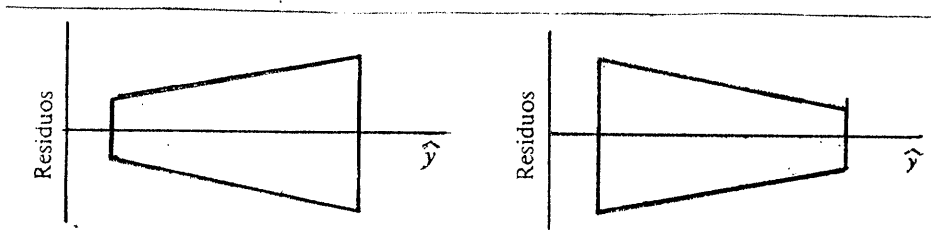


Figura 5.1. Forma de la gráfica de residuos contra valores ajustados.

Ejemplo 5.1. En la Figura 5.2 se grafican los residuos contra los valores ajustados de los datos de la resistencia a la tensión del Ejemplo 2.3. Ningún patrón inusual es evidente, por lo que no hay evidencia de varianza variable.

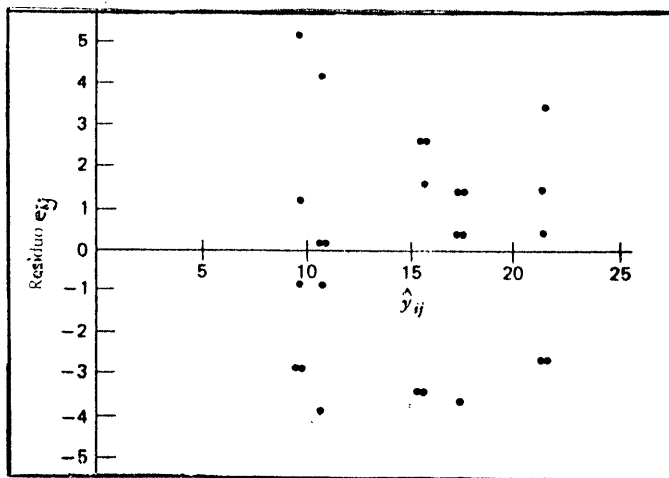


Figura 5.2. Gráfica de residuos contra valores ajustados.

Prueba de Bartlett para Homogeneidad de Varianzas.

Bartlett (1937) propone un método para probar la homogeneidad de la varianza. Esta prueba es una modificación de la prueba de razón de verosimilitud de Neyman-Pearson.

Este procedimiento consiste en calcular una estadística cuya distribución muestral es, aproximadamente, ji cuadrada con $k - 1$ grados de libertad cuando las k muestras aleatorias provienen de poblaciones normales independientes.

La estadística de la prueba es

$$\chi_{k-1}^2 = \frac{M}{C},$$

donde

$$M = 2.3026(g.l.)[k \log \bar{s}^2 - \sum \log s^2]$$

la cual

$$\bar{s}^2 = \frac{\sum s^2}{k}$$

s^2 es la varianza muestral de la i -ésima población,

$g.l.$ = grados de libertad para la varianza,

k = número de tratamientos o tipo de muestra; y

$$C = 1 + \frac{k + 1}{3(g.l.)(k)}.$$

Por lo tanto, debe rechazarse H_0 para valores grandes de χ_{k-1}^2 ; en otras palabras, se rechaza H_0 , sólo si

$$\chi_{k-1}^2 > \chi_{\alpha, k-1}^2$$

en donde $\chi_{\alpha, k-1}^2$ es el punto α porcentual superior de la ji cuadrada con $k - 1$ grados de libertad.

Varios estudios indican que la prueba de Bartlett es muy sensible a la suposición de normalidad y no debe ser aplicada cuando exista alguna duda en cuanto a esta suposición (Anderson y McLean, 1974).

Ejemplo 5.2. De los datos del Ejemplo 2.4 probamos la homogeneidad de varianza. El estadístico de prueba es

Prueba de Homogeneidad de Bartlett.

k	Tipo	s^2	$\log s^2$
1	A	059.55	1.77488
2	B	026.22	1.41863
3	C	430.89	2.63437
4	D	436.22	2.63971
	Total	952.88	8.46759

$$\bar{s}^2 = \frac{\sum s^2}{k} = \frac{952.88}{4} = 238.22,$$

$$\log \bar{s}^2 = 2.37698,$$

$$M = 2.3026(9)[4(2.37698) - 8.46759] = 21.56,$$

$$C = 1 + \frac{5}{3(9)(4)} = 1.046,$$

$$\chi_{k-1}^2 = \chi_3^2 = \frac{21.56}{1.046} = 20.6$$

y

$$\chi_{\alpha, k-1}^2 = \chi_{0.001, 3}^2 = 1.63.$$

Como el valor de $\chi_3^2 = 20.6$ es mayor que $\chi_{0.001, 3}^2 = 1.63$, por lo tanto, podemos rechazar la hipótesis nula, concluyéndose que las varianzas no son homogéneas.

Prueba de Bartlett y Kendall para Homogeneidad de Varianzas.

Esta prueba fue sugerida por Bartlett y Kendall (1946), y utiliza el $\log s^2$ como una variable. Ellos sugieren que se haga una tabla de análisis de varianza para el $\log s^2$, asumiendo que las varianzas del $\log s^2$ son homogéneas. El fundamento para usar el $\log s^2$ como la variable es que es distribuido aproximadamente normal si $n \geq 5$. El procedimiento es el siguiente

En cada tratamiento (siempre y cuando el tratamiento sea igual o mayor que cinco, $n \geq 5$), se agrupa en dos partes iguales, de modo que al primer grupo de datos se le calcule su varianza y luego se le saca el logaritmo a la varianza, después para el segundo grupo se hace lo mismo.

Luego hacemos el análisis de varianza para el $\log s^2$, de acuerdo al modelo estadístico correspondiente. Usando la hipótesis de prueba

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2.$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2, \text{ para al menos una } i \neq j.$$

Para decidir si aceptamos o rechazamos la hipótesis de homogeneidad de varianza. Por lo tanto, debe rechazarse H_0 para valores grandes de $F_{\alpha, a-1, N-a}$; en otras palabras, se rechaza H_0 , sólo si

$$F_0 > F_{\alpha, a-1, N-a}$$

en donde $F_{\alpha, a-1, N-a}$ es el punto α porcentual superior de la tabla A.2.

Ejemplo 5.3. De los datos del Ejemplo 2.4 determinamos el $\log s^2$ para cada columna de cada tipo de tabletas, porque $n \geq 5$, esto es, cada columna de cada tipo de tableta es de tamaño 5. En el cuadro 5.1 se muestran las varianzas de cada columna.

CUADRO 5.1. Análisis de Homogeneidad de $\log s^2$.

<u>Tipos de tabletas.</u>							
<u>A</u>		<u>B</u>		<u>C</u>		<u>D</u>	
<u>s^2</u>	<u>$\log s^2$</u>	<u>s^2</u>	<u>$\log s^2$</u>	<u>s^2</u>	<u>$\log s^2$</u>	<u>s^2</u>	<u>$\log s^2$</u>
059.0	1.7709	010.0	1.0000	449.5	2.6527	493.0	2.6928
052.5	1.7202	039.0	1.5911	510.0	2.7076	478.5	2.6799

Ahora hacemos los cálculos para el análisis de varianza para un diseño completamente aleatorizado.

Sea $N = 8$, $n = 2$, $a = 4$,

$$SC_{Total} = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{N} = 38.427725 - \frac{(16.8152)^2}{8} = 3.0838561.$$

$$SC_{Tratam.} = \sum_i \frac{y_i^2}{n} - \frac{y_{..}^2}{N} = \frac{76.5003}{2} - \frac{(16.8152)^2}{8} = 2.906281.$$

$$SC_{Error} = SS_{Total} - SS_{Tratam.} = 3.0838561 - 2.906281 = 0.177575.$$

En el cuadro 5.2 se muestra la tabla de análisis de varianza para $\log s^2$.

CUADRO 5.2. Análisis de Varianza de $\log s^2$.

Fuente Varia.	S.C.	g.l.	C.M.	F_0
Tipo Tabletás	2.9063	3	0.9688	21.82
Dentro Error	0.1776	4	0.0444	
Total	3.0839	7		

Tenemos que $F_{\alpha, a-1, N-a} = F_{0.01, 3, 4} = 16.69$. Como $F_0 > F_{0.01, 3, 4}$, rechazamos la hipótesis de que las varianzas son homogéneas a $\alpha = 0.01$.

Prueba de Burr y Foster para Homogeneidad de Varianzas.

Esta prueba es llamada también la Prueba Q (Burr y Foster, 1972) para igualdad de varianzas. Esta basada en un estadístico el cual es una función monotonamente creciente del coeficiente de variación de la varianza simple.

Para igual tamaño de muestra n y para cada población p , s_i^2 (para $i = 1, 2, \dots, p$) denota la i -ésima varianza muestral. Denotando el valor del estadístico de prueba por q , tenemos que

$$q = \frac{(s_1^4 + s_2^4 + \dots + s_p^4)}{(s_1^2 + s_2^2 + \dots + s_p^2)^2}$$

para desigual tamaño de muestra especificamos que cada varianza muestral s_i^2 , es calculada para ser dividida por los grados de libertad v_i , para el tamaño de muestra n_i (donde $v_i = n_i - 1$, para $i = 1, 2, \dots, p$). Sea \bar{v} el promedio aritmético de los grados de libertad. En este caso tenemos que

$$q = \frac{\bar{v}(v_1 s_1^4 + v_2 s_2^4 + \dots + v_p s_p^4)}{(v_1 s_1^2 + v_2 s_2^2 + \dots + v_p s_p^2)^2}$$

para valores grandes de q , rechazar la hipótesis de igual varianza poblacional. Los valores críticos son dados en la tabla A.5, para varios números de población p , y varias posibilidades para igual grados de libertad. Para desigual grados de libertad, \bar{v} es sustituida por v .

Ejemplo 5.4. En la cuadro 5.3 usamos los datos del Ejemplo 2.4.

CUADRO 5.3. Cálculo de las varianzas.

<i>Tipo</i>	$\underline{s^2}$	$\underline{s^4}$
A	059.55	003546.20
B	026.22	000687.49
C	430.89	185666.19
D	436.22	190287.89
<i>Total</i>	952.88	380187.77

$$q = \frac{\sum_{i=1}^4 s_i^4}{(\sum_{i=1}^4 s_i^2)^2} = \frac{380187.77}{(952.88)^2} = 0.419.$$

Con $v = n - 1 = 9$ y $p = 4$ usamos la tabla A.5, para buscar el valor crítico de $q_{crit}(\alpha)$, para lo cual el valor de

$$q_{crit}(0.01) = 0.398,$$

$$q_{crit}(0.001) = 0.468.$$

Como $q_{crit}(0.01) < q < q_{crit}(0.001)$ y la teoría indica que el tiempo de desintegración no es normalmente distribuido, rechazamos la hipótesis de que las varianzas son homogéneas y entonces procedemos a transformar la variable (ver punto 3 de los conceptos preliminares).

6. TRANSFORMACION DE DATOS.

Conceptos Preliminares.

Frecuentemente se pueden hacer transformaciones que corregirán las deficiencias de los datos, a fin de satisfacer los supuestos.

El enfoque usual para tratar varianzas variables consiste en aplicar una transformación para igualar las varianzas y volver a aplicar el análisis de varianza de los datos transformados. Debe notarse que, bajo este enfoque, las conclusiones del análisis de varianza se aplican a las poblaciones transformadas.

Si se conoce la distribución de la población principal, entonces se conoce la relación entre las medias de los tratamientos y las varianzas de los tratamientos calculados con base en los tratamientos individuales. Los datos pueden transformarse o medirse en una nueva escala de medida de tal manera que los datos transformados se distribuyan de forma aproximadamente normal. Tales transformaciones también se proponen hacer que las medias y las varianzas sean independientes, y que las varianzas resultantes sean homogéneas. Este resultado no siempre se consigue. Cuando es imposible hallar una transformación que haga que las medias y las varianzas sean independientes y la varianza estable, entonces deben usarse otros métodos de análisis.

Si la forma de el cambio de varianza con la media es conocida, esta determina el tipo de transformación a usar (Bartlett, 1947). Suponga que

$$\sigma_x^2 = f(m), \quad (6.1.1)$$

donde σ_x^2 es la varianza en la escala original de medición x con la media de x igual a m . Entonces para cualquier función $g(x)$ tenemos aproximadamente

$$\sigma_g^2 = \left(\frac{dg}{dm}\right)^2 f(m), \quad (6.1.2)$$

así, que si σ_g^2 es la constante C^2 , tenemos que

$$g(m) = \int \frac{C dm}{\sqrt{f(m)}}. \quad (6.1.3)$$

Por ejemplo, si la desviación estandar σ_x tiende a ser proporcional a la media m , tenemos que $f(m)$ es proporcional a m^2 , y $g(m)$ es proporcional al $\log m$; esto es, deberíamos usar la escala logarítmica.

Sin embargo, una varianza constante no solo es una condición que debemos buscar, es necesario ser precabidos cuando efectuemos el análisis de varianza con las variables transformadas. El caso ideal es

- 1) La varianza de la variable transformada no debe estar relacionada con la media.
- 2) La variable transformada debera ser normalmente distribuida.
- 3) La escala transformada debe ser de tal forma que el promedio aritmético sea un estimador eficiente de la media verdadera para cualquier grupo particular de medición.
- 4) La escala transformada debe ser tal que los efectos reales sean lineales y aditivos.

Se ha realizado una investigación considerable sobre el tema de la selección de la transformación apropiada. Si el experimentador conoce la distribución teórica de las observaciones, es posible usar esta información para elegir la transformación. Las transformaciones realizadas para igualar las varianzas también afectan la forma

de la distribución de los errores. En la mayoría de los casos esta distribución se acerca más a la normal cuando se realiza una transformación.

Las transformaciones más comunes que se usan en el análisis de varianza son

1) Transformación Logarítmica ($Y = \log x$). Se aplica cuando (i) Los efectos verdaderos son multiplicativos (o proporcionales), o (ii) La desviación estándar es proporcional a la media.

2) Transformación Raíz Cuadrada ($Y = \sqrt{x}$ ó $Y = \sqrt{x+1}$). Se aplica cuando la varianza es proporcional a la media, por ejemplo, cuando los datos originales son muestras de una distribución de Poisson.

3) Transformación Arc sen ($Y = \arcsen\sqrt{p}$). Se aplica cuando la varianza es proporcional a la media, por ejemplo, cuando los datos originales son muestras (expresadas como proporciones o frecuencias relativas) de poblaciones binomiales.

4) Transformación Recíproca ($Y = \frac{1}{x}$). Se aplica cuando la desviación estándar es proporcional al cuadrado de la media.

Transformación de la raíz cuadrada.

Cuando los datos consisten de números enteros pequeños, por ejemplo, el número de colonias de bacterias en un recuento de placas, el número de plantas o insectos de una especie determinada en una zona dada, tales datos siguen a menudo la distribución de Poisson, en la cual la media y la varianza son iguales (ver ecuación 6.1.1). Cuando intervienen valores muy pequeños, \sqrt{x} tiende a corregir en exceso, así que el intervalo de los valores transformados que dan una media pequeña, puede ser mayor que el intervalo de valores transformados que dan una media mayor. Por esta razón se recomienda a $\sqrt{x + \frac{1}{2}}$ como transformación adecuada cuando algunos

de los valores está por debajo de 10 o aún por debajo de 15 y especialmente cuando hay ceros (Bartlett, 1947).

Cuando los datos están en *porcentajes* y donde el intervalo de porcentaje va de cero a 20 por ciento o de 80 a 100 por ciento, pero no de ambos, también pueden analizarse con la transformación raíz cuadrada \sqrt{x} . Los porcentajes entre 80 y 100 por ciento deberán restarse de 100 antes de hacer la transformación, es decir, se debe restar a 100 cada valor para obtener datos entre cero y 20. La misma transformación es útil para porcentajes en los mismos intervalos en que las observaciones están en una escala continua, ya que las medias y varianzas pueden ser aproximadamente iguales. Si el intervalo de porcentajes es de 30 a 70, es dudosa la necesidad de transformación (Steel y Torrie, 1986).

Ejemplo 6.1. En un experimento, se hicieron cortes a los estigmas de maíz, y se hizo un conteo de los huevecillos del gusano elotero encontrados. Los tratamientos fueron: un corte a 2 y 4 cm; dos cortes a 2 y 4 cm. de longitud del estigma, y testigo, sin corte. Los resultados se anotan en el cuadro 6.1.

CUADRO 6.1. Número de huevecillos de *Heliothis zea*.

<i>Días</i>	<i>2cm - 1c</i>	<i>2cm - 2c</i>	<i>4cm - 1c</i>	<i>4cm - 2c</i>	<i>Testigo</i>	x_j
1	38	31	56	67	138	330
2	39	27	65	72	138	341
3	05	05	17	03	10	40
4	13	02	20	02	15	52
5	35	21	44	13	41	152
6	93	50	94	64	171	472
7	34	28	36	18	41	157
8	30	19	27	30	45	151
9	44	31	35	34	51	195
10	24	16	20	17	24	101
11	12	13	19	14	23	81
12	07	14	04	16	12	53
x_i	374	257	437	350	709	2127
\bar{x}_i	31.17	21.42	36.42	29.17	59.10	
<i>Rango</i>	88	48	90	70	161	

1) El testigo tiene mayor media y también mayor variación, estimados por el rango.

2) La media y la variación están relacionados; no son independientes.

3) Las varianzas, estimadas por los rangos, son heterogéneas.

Se úso un análisis de varianza para un bloques al azar.

CUADRO 6.2. Análisis de varianza con los valores originales del cuadro 6.1.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Días	040804.1	11	003709.5	10.20
Tratamientos	009770.8	04	002442.7	06.70
Error	016016.0	44	000364.0	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{2127}{(5)(12)} = 35.45.$$

$$CV = \frac{\sqrt{364}}{35.45} \times 100 = 53.82\%.$$

CUADRO 6.3. Valores del cuadro 6.1 transformados a \sqrt{x} .

<i>Días</i>	<i>2cm - 1c</i>	<i>2cm - 2c</i>	<i>4cm - 1c</i>	<i>4cm - 2c</i>	<i>Testigo</i>
1	6.16	5.57	7.48	8.19	11.75
2	6.24	5.20	8.06	8.49	11.75
3	2.24	2.24	4.12	1.73	3.16
4	3.61	1.41	4.47	1.41	3.87
5	5.92	4.58	6.63	3.61	6.40
6	9.64	7.07	9.70	8.00	13.08
7	5.83	5.29	6.00	4.24	6.40
8	5.48	4.36	5.20	5.48	6.71
9	6.63	5.57	5.92	5.83	7.14
10	4.90	4.00	4.47	4.12	4.90
11	3.46	3.61	4.36	3.74	4.80
12	2.65	3.74	2.00	4.00	3.46
x_i	62.76	52.64	68.41	58.84	83.42
\bar{x}_i	5.53	4.39	5.70	4.90	6.95
<i>Rango</i>	7.40	5.66	7.70	7.08	9.92

CUADRO 6.4. Análisis de varianza de los valores del cuadro 6.3.

Fuente Varia.	S.C.	g.l.	C.M.	F ₀
Días	254.78	11	23.16	18.64
Tratamientos	045.54	04	11.39	09.17
Error	054.66	44	01.24	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{326.07}{(5)(12)} = 5.43.$$

$$CV = \frac{\sqrt{1.242}}{5.43} \times 100 = 20.5\%.$$

Los datos del cuadro 6.3 indican varianzas homogéneas e independientes de los tratamientos. El análisis de varianza del cuadro 6.2, con $CV = 53.82\%$, y el del cuadro 6.4 con $CV = 20.5\%$, hacen más confiables dichos datos.

Ejemplo 6.2. *Datos en porcentajes.* En un experimento con girasol se determinó el porcentaje de avanamiento del aquenio, en capítulos tratados con diferentes insecticidas. Los datos se presentan en el cuadro 6.5, donde M=Malatión, G=Gardona, L=Lindano, S=Sevimol, A=Abodrin, y T=Testigo.

CUADRO 6.5. Porcentaje de avanamiento del aquenio.

<u>Trata</u>	<u>Bloques</u>						x_i	\bar{x}_i	<i>Rango</i>
	I	II	III	IV	V	VI			
M	04.0	08.3	05.3	04.6	04.0	04.0	30.2	05.03	4.3
G	11.7	05.3	08.0	03.6	05.6	07.3	41.5	06.92	8.1
L	05.3	06.0	06.0	04.3	09.3	06.3	37.2	06.20	5.0
S	13.3	06.0	04.6	07.0	06.6	07.0	44.5	07.42	8.7
A	06.7	07.0	06.3	05.3	09.3	02.6	37.2	06.20	4.0
T	12.7	14.0	19.6	17.6	14.6	15.6	94.1	15.68	6.9
x_j	53.7	46.6	49.8	42.4	49.4	42.8	287.7		

CUADRO 6.6. Análisis de varianza de los valores del cuadro 6.5.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Bloques	016.039	05	03.208	0.522
Tratamientos	454.669	05	90.934	14.81
Error	153.542	25	06.142	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{284.7}{36} = 7.908.$$

$$CV = \frac{\sqrt{6.142}}{7.908} \times 100 = 31.34\%.$$

CUADRO 6.7. Valores originales del cuadro 6.5 transformados a \sqrt{x} .

	<i>Bloques</i>						x_i	\bar{x}_i	<i>Rango</i>
<i>Trata</i>	I	II	III	IV	V	VI			
M	2.00	2.88	2.30	2.14	2.00	2.00	13.32	2.22	0.88
G	3.42	2.30	2.83	1.90	2.37	2.70	15.52	2.59	1.52
L	2.30	2.45	2.45	2.07	3.05	2.51	14.83	2.47	0.98
S	3.65	2.45	2.14	2.65	2.57	2.65	16.11	2.69	1.51
A	2.59	2.65	2.51	2.30	3.05	1.61	14.71	2.45	0.54
T	3.56	3.74	4.43	4.20	3.82	3.95	23.70	3.95	0.87
x_j	17.52	16.47	16.66	15.26	16.86	15.42	98.19		

CUADRO 6.8. Análisis de varianza de los valores del cuadro 6.7.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Bloques	00.632	05	0.1264	0.662
Tratamientos	11.419	05	2.2980	12.05
Error	04.767	25	0.1907	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{98.19}{36} = 2.7275.$$

$$CV = \frac{\sqrt{0.1907}}{2.7275} \times 100 = 16.01\%.$$

Los datos son más confiables. Las varianzas estimadas por los rangos en los tratamientos, son más homogéneas y CV es más bajo.

Ejemplo 6.3. Los datos presentados en el cuadro 6.9 muestran el número de insectos *Iygu* obtenidos en cincuenta barridas en cada parcela de un experimento para probar 10 insecticidas y un tratamiento de control, repetido cuatro veces en un diseño de bloques completos al azar.

CUADRO 6.9. Número de *Iygu* por 50 barridas.

<i>Trata</i>	<i>Bloques</i>				<i>Total</i>	<i>Media</i>	s_i^2
	I	II	III	IV			
A	7	5	4	1	17	4.25	6.25
B	6	1	2	1	10	2.50	5.67
C	6	2	1	0	9	2.25	6.92
D	0	1	2	0	3	0.75	0.92
E	1	0	1	2	4	1.00	0.67
F	5	14	9	15	43	10.75	21.58
G	8	6	3	6	23	5.75	4.25
H	3	0	5	9	17	4.25	14.25
I	4	10	13	5	32	3.00	18.00
J	6	11	5	2	24	6.00	14.00
K	8	11	2	6	27	6.75	14.25

Observamos que la variación es proporcional a la media y los datos son números enteros pequeños, y aún más los datos tienen una distribución Poisson. El análisis de varianza es

CUADRO 6.10. Análisis de varianza de los valores del cuadro 6.9.

Fuente Varia.	S.C.	g.l.	C.M.	F ₀
Bloques	012.25	03	04.08	0.40
Tratamientos	380.00	10	38.00	3.70
Error	308.00	30	10.27	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{209}{(11)(4)} = 4.75.$$

$$CV = \frac{\sqrt{10.27}}{4.75} \times 100 = 67.46\%.$$

Transformando los datos mediante la aplicación de $\sqrt{x + \frac{1}{2}}$, queda

CUADRO 6.11. Datos transformados del Iygu.

<u>Trata</u>	<u>Bloques</u>				<u>Total</u>	<u>Media</u>	<u>s_i²</u>
	I	II	III	IV			
A	2.74	2.35	2.12	1.22	8.43	2.11	0.41
B	2.55	1.22	1.58	1.22	6.57	1.65	0.39
C	2.55	1.58	1.22	0.71	6.06	1.52	0.60
D	0.71	1.22	1.58	0.75	4.22	1.06	0.18
E	1.22	0.71	1.22	1.58	4.73	1.18	0.13
F	2.35	3.81	3.08	3.92	13.18	3.29	0.54
G	2.92	2.55	1.87	2.55	9.89	2.45	0.19
H	1.87	0.71	2.35	3.08	8.01	2.00	0.99
I	2.12	3.24	3.67	2.35	11.38	2.84	0.53
J	2.55	3.39	2.35	1.58	9.87	2.47	0.55
K	2.92	3.39	1.58	2.55	10.44	2.61	0.59

Notemos que la varianza no es proporcional a la media, y el análisis de varianza correspondiente es

CUADRO 6.12. Análisis de varianza de los valores del cuadro 6.11.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Bloques	00.532	03	0.177	0.36
Tratamientos	19.993	10	1.999	4.40
Error	14.841	30	0.492	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{92.78}{(11)(4)} = 2.11.$$

$$CV = \frac{\sqrt{0.492}}{2.11} \times 100 = 33.24\%.$$

El coeficiente de variación bajó, por lo tanto uno de los supuestos del análisis de varianza fue violado en los datos originales, lo cual se subsanó mediante la transformación.

En general, se puede decir que los datos que requieren la transformación de raíz cuadrada no violan los supuestos del análisis de varianza tan drásticamente como los datos que requieren una transformación logarítmica. Consecuentemente, los cambios en el análisis provocados por la transformación no son espectaculares.

Transformación angular o transformación arco seno.

La transformación del seno inverso de la raíz cuadrada

$$g(x) = \text{sen}^{-1}\sqrt{x}$$

es aplicable a datos binomiales expresados como fracciones decimales o porcentajes, y se recomienda especialmente cuando los porcentajes cubren un intervalo amplio

de valores. La mecánica de la transformación requiere fracciones decimales, pero las tablas de la transformación arco seno están generalmente en porcentajes (tabla A.8). Los valores tabulados o valores arco seno se expresan en grados o radianes. La varianza de las observaciones resultantes es aproximadamente constante, y es $821/n$ cuando los datos transformados se expresan en grados y $1/(4n)=0.25/n$ cuando están en radianes, siendo n el denominador común de todas las fracciones. Esta varianza aclara que todos los porcentajes se deben basar en un número igual de observaciones. Sin embargo, la transformación se usa a menudo cuando los denominadores son desiguales, y especialmente si son aproximadamente iguales (Steel y Torrie, 1986).

Bartlett (1947) sugiere que cero se sustituya por $25/n$ por ciento y que cien por ciento se sustituya por $100 - 25/n$, siendo n su divisor.

Esta transformación fue desarrollada para proporciones binomiales. Si a_{ij} éxitos, de un total de n , se obtienen en la j -ésima repetición del i -ésimo tratamiento, la proporción $\hat{p}_{ij} = \frac{a_{ij}}{n}$ tiene varianza $\frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n}$. Por medio de Tablas (tabla A.8), elaborada por C.I. Bliss, sustituimos \hat{p}_{ij} por el ángulo cuyo seno es $\sqrt{\hat{p}_{ij}}$. En la escala angular las proporciones cercanas a cero o a uno están esparcidas para aumentar su varianza. Si toda la varianza de error es binomial, la varianza del error en la escala angular es más o menos $821/n$. La transformación no elimina las desigualdades en varianza que surgen de valores diferentes de n . Si las n varían ampliamente, un análisis ponderado en la escala angular es muy recomendable. Con $n < 50$, una proporción cero habrá de contarse como $1/4 n$, antes de transformar a ángulos, y una proporción de $100n$ por ciento como $(n - 1/4)/n$. Cuando cae entre 30 por ciento y 70 por ciento la transformación angular de hecho no se necesita. Cuando cae en este intervalo la transformación probablemente no producirá un cambio notable en las conclusiones, a no ser que las \hat{p}_{ij} caigan entre cerca de cero a 30 por ciento o entre 70 y 100 por ciento (Snedecor y Cochran, 1975).

Por regla general, los datos en porcentajes o proporciones tienen una distribución binomial, en vez de una distribución normal. Una de las características de esta distribución es que las varianzas se hallan relacionadas con las medias. En los datos binomiales, las varianzas tienden a ser pequeñas en los dos extremos de los rangos de valores (cerca de cero y a 100 por ciento), pero mayores en el medio (alrededor del 50 por ciento) (Little y Hills, 1991).

La varianza es proporcional a la media, como por ejemplo, cuando los datos originales son muestras expresadas como proporciones o frecuencias relativas de poblaciones binomiales (Ostle, 1983).

Ejemplo 6.4. Los datos del cuadro 6.13 corresponden a un experimento con diseño completamente aleatorizado, con semillas de lechuga en el que se incluyen 24 tratamientos, cada uno de ellos repetidos tres veces. Los tratamientos se encuentran dispuestos por el orden de la magnitud de sus medias.

CUADRO 6.13. Número de semillas de lechuga que germinan, en muestras de 50.

<i>Tratam</i>	<i>Repeticiones</i>			<i>Media</i>	s_i^2	$\log s_i^2$
	1	2	3			
1	0	0	1	0.33	0.33	-0.48
2	0	1	0	0.33	0.33	-0.48
3	0	0	1	0.33	0.33	-0.48
4	0	2	0	0.67	1.33	0.12
5	2	0	0	0.67	1.33	0.12
6	0	2	3	1.67	2.33	0.37
7	7	10	7	8.00	3.00	0.48
8	11	12	15	12.67	4.33	0.64
9	13	18	18	16.33	8.33	0.92
10	22	16	13	17.00	21.00	1.32
11	24	13	18	18.33	30.33	1.48
12	23	21	29	24.33	17.33	1.24
13	24	29	29	27.33	8.33	0.92
14	37	28	27	30.67	30.33	1.48
15	42	41	40	41.00	1.00	0
16	39	41	45	41.67	9.33	0.97
17	41	45	40	42.00	7.00	0.85
18	47	41	43	43.67	9.33	0.97
19	45	42	48	45.00	9.00	0.95
20	46	42	48	45.33	9.33	0.97
21	49	46	48	47.67	2.33	0.37
22	48	49	48	48.33	0.33	-0.48
23	50	49	48	49.00	1.00	0
24	49	49	50	49.33	0.33	-0.48
<i>Total</i>					178.00	11.77

Obsérvese que hay una marcada tendencia, en las varianzas de los extremos, a ser más pequeñas que aquellas en la mitad de la distribución. Esto es clásico de los datos binomiales. Los logaritmos de las varianzas se han incluido de manera que la prueba de Bartlett pueda llevarse a cabo de la siguiente manera

$$\bar{s}_i^2 = \frac{178}{24} = 7.417,$$

$$\log \bar{s}_i^2 = 0.8702,$$

$$M = 2.3026(2)[24(0.8702) - 11.77] = 41.976,$$

$$C = 1 + \frac{25}{3(2)(24)} = 1.1736,$$

$$\chi_{23}^2 = \frac{M}{C} = 35.7668.$$

$$\chi_{23}^2(0.05) = 35.1725.$$

Este sólo es significativo en el nivel de 5 por ciento (valor requerido 35.1725). Esto es, $\chi_{23}^2 > \chi_{23}^2(0.05)$, de modo que contamos con muy buena evidencia de que las varianzas no son homogéneas.

Analizando los datos originales, obtenemos el siguiente análisis de varianza

CUADRO 6.14. Análisis de Varianza de los datos originales.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Tratamientos	25266	23	1098.52	148.12
Error	00356	48	0007.42	

Usando la tabla A.8, para transformar los datos originales quedando de la siguiente manera

CUADRO 6.15. Datos transformados del número de semillas de lechuga.

<i>Tratam</i>	<i>Repeticiones</i>			<i>Media</i>	s_i^2	$\log s_i^2$
	1	2	3			
1	0	0	5.74	1.91	10.98	1.04
2	0	5.74	0	1.91	10.98	1.04
3	0	0	5.74	1.91	10.98	1.04
4	0	8.13	0	2.71	22.03	1.34
5	8.13	0	0	2.71	22.03	1.34
6	0	8.13	9.98	6.04	36.44	1.56
7	15.34	18.44	15.34	16.37	10.26	1.01
8	19.37	20.27	22.77	20.80	3.10	0.49
9	21.13	25.10	25.10	23.78	5.25	0.72
10	27.97	23.58	21.13	24.23	12.01	1.08
11	29.33	21.13	25.10	25.19	16.82	1.23
12	28.66	27.28	32.58	29.51	7.56	0.88
13	29.33	32.58	32.58	31.50	3.52	0.55
14	37.47	31.90	31.31	33.56	11.55	1.06
15	40.40	39.29	39.23	39.64	0.43	-0.37
16	38.60	39.29	42.13	40.01	3.50	0.54
17	39.29	42.13	39.23	40.22	2.75	0.44
18	43.28	39.29	40.98	41.18	4.01	0.60
19	42.13	40.40	43.90	42.14	3.06	0.49
20	42.71	40.40	43.90	42.34	3.17	0.50
21	44.33	42.71	43.90	43.65	0.70	-0.15
22	43.90	44.33	43.90	44.04	0.06	-1.22
23	45.00	44.33	43.90	44.41	0.31	-0.51
24	44.33	44.33	44.41	44.41	0.002	-2.69
<i>Total</i>					201.50	12.007

Aparentemente eliminamos toda relación entre las varianzas y las medias.

Un análisis de varianza de los datos transformados no parece conducirnos a una conclusión distinta de la del análisis de los datos originales.

CUADRO 6.16. Análisis de Varianza de los datos transformados.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Tratamientos	60725.7	23	2640.25	102.37
Error	01237.9	48	0025.79	

Recuérdese que no se transformaron los datos para obtener resultados que sean más agradables, sino que se transforman de modo que el análisis sea válido y las conclusiones correctas.

Ejemplo 6.5. En el cuadro 6.17 se presentan datos de un experimento con fertilizantes foliares en el maíz. Los valores indican los porcentajes de mazorcas enfermas en cada unidad experimental.

CUADRO 6.17. Porcentajes de mazorcas enfermas en un experimento con maíz.

<i>Trata</i>	<i>Bloques</i>				x_i	<i>Rango</i>	\bar{x}_i	s_i^2	$\log s_i^2$
	I	II	III	IV					
A	24	10	16	07	57	17	14.25	7.5	1.75
B	19	14	30	13	76	17	19.00	7.79	1.78
C	13	23	22	08	66	15	16.50	7.23	1.72
D	24	10	18	12	64	14	16.00	6.32	1.59
E	15	15	19	06	55	13	13.75	5.5	1.48
F	05	13	26	19	63	21	15.75	8.92	1.90
x_j	100	85	131	65	381				

Como la variación excede los límites que la experiencia de muchos investigadores ha fijado entre cero por ciento y 20 por ciento, las varianzas de los extremos tienden a ser más pequeñas que las del centro de la distribución, por lo que sus datos se asemejan a una distribución binomial, es necesario la transformación $\text{arc sen}\sqrt{\text{porcentaje}}$. Los logaritmos de las varianzas se han incluido de manera que la prueba de Bartlett pueda llevarse a cabo

$$\bar{s}_i^2 = \frac{317.83}{6} = 52.97,$$

$$\log \bar{s}_i^2 = 1.724,$$

$$M = 2.3026(3)[6(1.724) - 10.22] = 0.8584,$$

$$C = 1 + \frac{7}{3(3)(6)} = 1.1296,$$

$$\chi_5^2 = \frac{M}{C} = 0.7599.$$

$$\chi_5^2(0.99) = 0.5543.$$

$$\chi_{23}^2 > \chi_{23}^2(0.05).$$

De modo que contamos con muy buena evidencia de que las varianzas no son homogéneas.

CUADRO 6.18. Análisis de varianza de los valores del cuadro 6.17.

Fuente Varia.	S.C.	g.l.	C.M.	F ₀
Bloques	386.79	03	128.930	3.390
Tratamientos	069.78	05	013.875	0.365
Error	570.46	15	038.03	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{381}{(6)(4)} = 15.875.$$

$$CV = \frac{\sqrt{CM_{error}}}{\bar{X}} \times 100 = \frac{\sqrt{38.03}}{15.875} \times 100 = 38.85\%$$

Como se muestra en esa tabla A.8, hay hileras desde 0.0 hasta 100.0 y columnas desde desde cero hasta nueve; para transformar un porcentaje de por ejemplo 24 por ciento, al valor $\text{arc sen}\sqrt{24\%}$, se localiza en la tabla la hilera 24, y en la intersección con la columna cero se tiene

$$\text{arc sen}\sqrt{24\%} = 29.33.$$

Para un valor de por ejemplo 9.4%, se localiza la hilera nueve, y en la intersección con la columna cuatro se tiene

$$\text{arc sen}\sqrt{9.4\%} = 17.85 + .$$

El signo + indica redondear a 17.9; el signo - indica bajar el valor al redondear.

Por ejemplo

$$\text{arc sen}\sqrt{8.4\%} = 16.85- = 16.8.$$

Usando la tabla A.8, se obtuvieron los datos presentados en el cuadro 6.19.

CUADRO 6.19. Valores transformados de los porcentajes de mazorcas de maíz.

<i>Tratam</i>	<i>Bloques</i>				x_i	\bar{x}_i	s_i^2
	I	II	III	IV			
A	29.33	18.44	23.58	15.34	86.69	21.67	37.61
B	25.84	21.97	33.21	21.13	102.15	25.54	30.37
C	21.13	28.66	27.97	16.43	94.19	23.55	42.95
D	29.33	18.44	25.10	20.27	93.14	23.29	24.13
E	22.79	22.79	25.84	14.18	85.60	21.40	25.24
F	12.92	21.13	30.66	25.84	90.55	22.64	57.11
x_j	141.34	131.43	166.36	113.19	552.32		

CUADRO 6.20. Análisis de varianza de los valores del cuadro 6.19.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Bloques	245.69	03	81.896	3.230
Tratamientos	045.09	05	09.018	0.356
Error	379.88	15	25.325	

$$\bar{X} = \frac{x_{..}}{ab} = \frac{552.32}{(6)(4)} = 23.01.$$

$$CV = \frac{\sqrt{CM_{error}}}{\bar{X}} \times 100 = \frac{\sqrt{25.325}}{23.01} \times 100 = 21.19\%$$

Como se observa por el valor del CV, los valores transformados merecen más confianza.

Transformación Logarítmica.

Cuando las varianzas son proporcionales a los cuadrados de las medias de los tratamientos o las desviaciones estándar son proporcionales a las medias, la

transformación logarítmica $\log x$ equilibra las varianzas. Se usa la base diez por comodidad, si bien cualquier base es satisfactoria. Los efectos que son multiplicativos en la escala original de medida, se vuelven aditivos en la escala logarítmica, por ejemplo, si un tratamiento da una respuesta 20 por ciento mayor que otro. La transformación logarítmica se usa con números enteros positivos que cubren un amplio intervalo. No puede usarse directamente para valores cero y cuando algunos de los valores son menores que 10, es deseable contar con una transformación que opere como la raíz cuadrada para valores pequeños y como logarítmica para valores grandes. La suma del valor uno a cada número antes de tomar los logaritmos, tiene el efecto deseado. Esto es, $\log(x + 1)$ se comporta como la transformación raíz cuadrada para números hasta 10 y difiere poco de $\log x$ para valores mayores de 10 (Steel y Torrie, 1986).

Las medias obtenidas volviendo a la escala original mediante el *antilog* de las medias de los valores de $\log x$ son medias geométricas de los datos originales. Cuando se usan pruebas como la desviación media significativas con los datos transformados son equivalentes a las razones mínimas significativas de medias que han sido transformadas de nuevo a la escala original.

Para estabilizar la varianza en la escala de la raíz cuadrada en el caso de una serie de malos experimentos son un tanto inesperados. Si es considerable la heterogeneidad que está presente en los números, la varianza es a menudo establecida para ser correlacionada con la media en una escala raíz cuadrada, y pueden ser solo estabilizadas si la transformación es hecha para la escala logarítmica. La explicación natural de que una varianza es más grande que la media, es que, la misma media fluctua, es decir

$$\sigma_x^2 = m + \sigma_m^2.$$

Para poblaciones biológicas, los incrementos en los números son a menudo

proporcionales alrededor de los presentes números, dando elevada variación ellas mismas, en las medias de lugar en lugar son proporcionales. esto es, como σ_m^2 podría ser proporcional a m^2 , así que podríamos esperar que

$$\sigma_x^2 = m + \lambda^2 \sigma_m^2.$$

Para λ grande, o m grande, esta regla de la varianza implica la transformación logarítmica (Bartlett, 1947).

Los datos con valores negativos no pueden transformarse en esta forma. Si existen ceros entre los datos, afrontemos el problema de que el logaritmo de cero es menos infinito. Para evitar esta situación, se recomienda sumar uno a cada dato antes de la transformación. Los datos que contienen un gran número de ceros probablemente se manejarían mejor mediante algún otro método (Little y Hills, 1990).

Ejemplo 6.6. Las pescas de plancton del cuadro 6.21 el número estimado de cuatro clases de plancton (I, II, III, IV), pescado en seis cargas con dos redes, cedieron muy bien a la transformación logarítmica.

CUADRO 6.21. Número estimado de cuatro clases de plancton.

<i>Carga</i>	<i>Números estimados</i>			
	I	II	III	IV
1	895	1520	43300	11000
2	540	1610	32800	8600
3	1020	1900	28800	8260
4	470	1350	34600	9830
5	428	980	27800	7600
6	620	1710	32800	9650
7	760	1930	28100	8900
8	537	1960	18900	6060
9	845	1840	31400	10200
10	1050	2410	39500	15500
11	387	1520	29000	9250
12	497	1685	22300	7900
<i>Media</i>	671	1701	30775	9396
<i>Rango</i>	663	1480	24400	9440

CUADRO 6.22. Transformación logarítmica

de los número estimado de cuatro clases de plancton.

<i>Carga</i>	<i>Números estimados</i>			
	I	II	III	IV
1	2.95	3.18	4.64	4.04
2	2.73	3.21	4.52	3.93
3	3.01	3.28	4.46	3.92
4	2.67	3.13	4.45	3.99
5	2.63	2.99	4.44	3.88
6	2.79	3.23	4.52	3.98
7	2.88	3.29	4.45	2.95
8	2.73	3.29	4.28	3.78
9	2.93	3.26	4.50	4.01
10	3.02	3.38	4.60	4.19
11	2.59	3.18	4.46	3.97
12	2.70	3.23	4.35	3.90
<i>Media</i>	2.802	3.221	4.480	3.962
<i>Rango</i>	0.43	0.39	0.36	0.41

CUADRO 6.23. Análisis de varianza de los valores del cuadro 6.22.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	<i>F₀</i>
Bloques	20.2070	03	6.7357	02.99
Tratamientos	00.3387	11	0.0308	10.99
Error	00.2300	33	0.0070	

Las amplitudes originales y las medias para las cuatro clases de plancton fueron casi proporcionales, siendo las razones, de amplitud de la media, de

$$\frac{663}{671} = 0.99, \frac{1480}{1701} = 0.87, \frac{24400}{30775} = 0.79, \frac{9440}{9396} = 1.00.$$

Después de la transformación, las amplitudes fueron casi iguales y no correlacionados con las medias.

Transformando de nuevo a la original, el número medio estimado pescado para las cuatro clases de plancton son

$$\text{antilog } 2.802 = 634; \text{ antilog } 3.221 = 1663; \text{ antilog } 4.480 = 30200; \text{ antilog } 3.962 = 9162.$$

Estas son medias geométricas.

Se encontrará que las medias de los logaritmos son significativamente diferentes para las cuatro clases de plancton.

La desviación estándar de los logaritmos es $\sqrt{0.0070} = 0.084$, y el antilogaritmo de este número es 1.21. Ahora, una desviación de 0.084 en los logaritmos de cada carga significa que cada una de éstas ha sido multiplicada (o dividida) por 1.21. por lo tanto, podemos decir que una desviación estándar de los logaritmos corresponde a una desviación estándar de porcentaje, o coeficiente de variación, de 21 por ciento en cada carga.

Selección de una Transformación para Igualar Varianzas.

En la sección anterior se hizo notar que si el experimentador conoce la relación entre la varianza de las observaciones y la media, puede usar esta información como guía al seleccionar la transformación. A continuación, se profundiza en este tema y se muestra como, a partir de los datos, es posible encontrar la forma de la transformación apropiada.

Sea $E(y) = \mu$ la media de y . Supóngase que la desviación estándar es proporcional a alguna potencia de la media de y , es decir

$$\sigma_y \propto \mu^\alpha,$$

se desea determinar la transformación de y que produzca una varianza constante. Se supone que la transformación es una potencia de los datos originales, en otras

palabras, que

$$Y = y^\lambda.$$

Entonces (ver figura 6.1)

$$\sigma_Y = \theta \sigma_y \propto \theta \mu^\alpha,$$

y θ , el gradiente de la curva de Y en función de y , depende del valor medio μ de y , le denominaremos con más propiedad θ_μ . Ahora

$$Y = y^\lambda,$$

el gradiente θ_μ es proporcional a $\mu^{\lambda-1}$. Así

$$\sigma_Y = \alpha \mu^{\lambda-1} \mu^\alpha = \mu^{\lambda+\alpha-1}.$$

Así Y se elige de tal forma que σ_y no dependa de μ , esto es, la varianza de los datos transformados es constante si

$$\lambda = 1 - \alpha.$$

Tomando los límites matemáticos apropiados se puede demostrar que el valor $\lambda = 0$ corresponde a la transformación logarítmica.

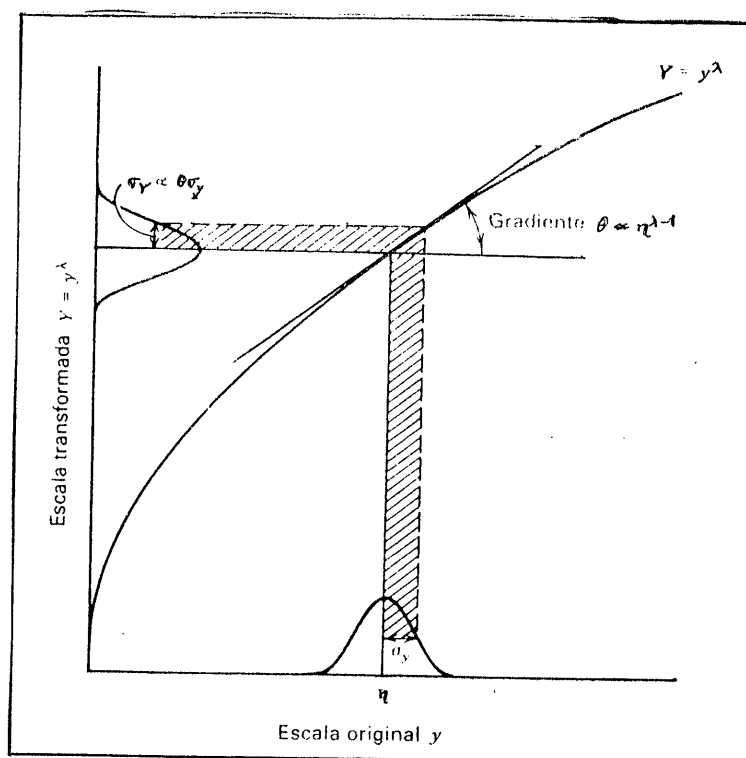


Figura 6.1. Transmisión de errores de y a $y = y^\lambda$.

Las transformaciones más comunes correspondientes a distintos valores de α se muestran en el cuadro 6.24. Estas transformaciones están acomodadas en el orden en el que potencia de la transformación aumenta.

La potencia de una transformación se refiere a la cantidad de curvatura que ésta produce. Una transformación poco potente aplicada a datos que se encuentran en un intervalo pequeño tiene un efecto pequeño, mientras que se producen efectos dramáticos si una transformación potente se aplica a datos distribuidos en un intervalo grande. A menudo las transformaciones tienen un efecto leve, a menos que la razón $\frac{y_{max}}{y_{min}}$ sea mayor que dos o tres.

CUADRO 6.24. Transformaciones para estabilizar las varianzas cuando $\sigma_y \propto \mu^\alpha$.

<i>Función</i>	α	$\lambda = 1 - \alpha$	<i>Transformación</i>
$\sigma \propto \mu^{\text{Constante}}$	0	1	Ninguna
$\sigma \propto \mu^{1/2}$	$\frac{1}{2}$	$\frac{1}{2}$	$Y = \sqrt{y}$
$\sigma \propto \mu^1$	1	0	$Y = \log_{10} y$
$\sigma \propto \mu^{3/2}$	$\frac{3}{2}$	$-\frac{1}{2}$	$Y = \frac{1}{\sqrt{y}}$
$\sigma \propto \mu^2$	2	-1	$Y = \frac{1}{y}$

Selección Empírica de α .

En muchas situaciones de diseños experimentales es posible estimar empíricamente α usando los datos. Como en la combinación del i -ésimo tratamiento

$$\sigma_{yi} \propto \mu_i^\alpha = \theta \mu_i^\alpha,$$

donde θ es la constante de proporcionalidad, puede tomarse el logaritmo para obtener

$$\log \sigma_{yi} = \log \theta + \alpha \log \mu_i,$$

en donde $\log \theta$ es una constante.

Por lo tanto, una gráfica de $\log \sigma_{yi}$ contra $\log \mu_i$ debe resultar en una recta con pendiente igual a α . Ya que σ_{yi} y μ_i son desconocidas, pueden sustituirse por estimaciones apropiadas y usar la pendiente de la recta ajustada como una estimación de α . Normalmente se usa la desviación estándar y el promedio del i -ésimo tratamiento (o más generalmente la combinación del tratamiento i -ésimo o conjunto de condiciones experimentales) para estimar σ_{yi} y μ_i .

Ejemplo 6.7. Un ingeniero civil está interesado en determinar si cuatro métodos diferentes para estimar la frecuencia de inundaciones, producen estimaciones equivalentes del gasto (o flujo) máximo cuando se aplica a una misma cuenca. Cada

procedimiento se usó seis veces en la cuenca y los resultados de gasto máximo (en pies cúbicos por segundo) aparecen en el cuadro 6.25. Se usó un análisis de varianza para una sola vía.

CUADRO 6.25. Datos del gasto máximo.

<u>Método</u>	<u>Observaciones</u>						\bar{y}_i	s_i
1	00.34	00.12	01.23	00.70	01.75	00.12	00.71	0.66
2	00.94	02.94	02.14	02.36	02.86	04.55	02.63	1.09
3	06.31	08.37	09.75	06.09	09.82	07.24	07.93	1.66
4	17.15	11.82	10.95	17.20	14.35	16.82	14.72	2.77

El análisis de varianza de los datos, mostrado en el cuadro 6.26, implica que existe diferencia en las estimaciones de la media del gasto máximo calculados por los cuatro procedimientos.

CUADRO 6.26. Análisis de varianza para el gasto máxima.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Método	708.3471	03	236.1157	76.07
Error	062.0811	20	003.1041	
Total	770.4282	23		

La gráfica de los residuos contra los valores ajustados, que aparecen en la figura 6.2, indica, por su forma de embudo que se ensancha, que viola la suposición de varianza constante.

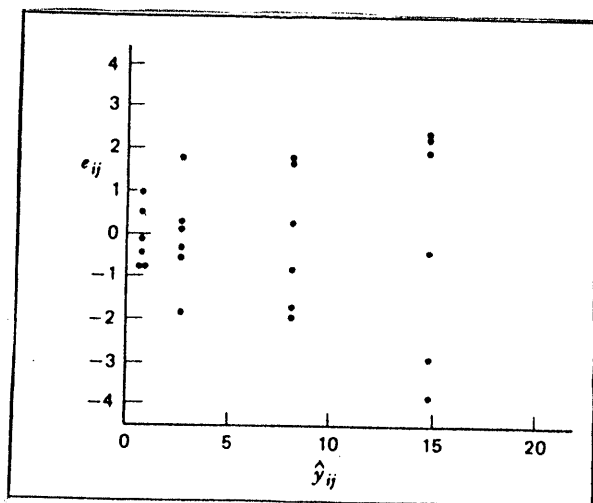


Figura 6.2. Gráfica de residuos contra $\hat{y}_{ij} = \bar{y}_i$.

Para investigar la posibilidad de usar una transformación con objeto de igualar las varianzas, en la figura 6.3, se gráfica el $\log s_i$ contra $\log \bar{y}_i$. A continuación se hacen los cálculos de los logaritmos

\bar{y}_i	$\log \bar{y}_i$	s_i	$\log s_i$
00.71	-0.3425	0.66	-0.4155
02.63	+0.9669	1.09	+0.0862
07.93	+2.0707	1.66	+0.5068
14.72	+2.6891	2.77	+1.0188

La pendiente de la recta que pasa a través de estos cuatro puntos es, aproximadamente, igual a $1/2$, esto es $\alpha = \frac{1}{2}$ (Para determinar la pendiente de los puntos de la figura 6.3 se recomienda utilizar el Análisis de Regresión para estimar la pendiente $\hat{\beta}_1$, aunque también se pueden utilizar otros métodos.).

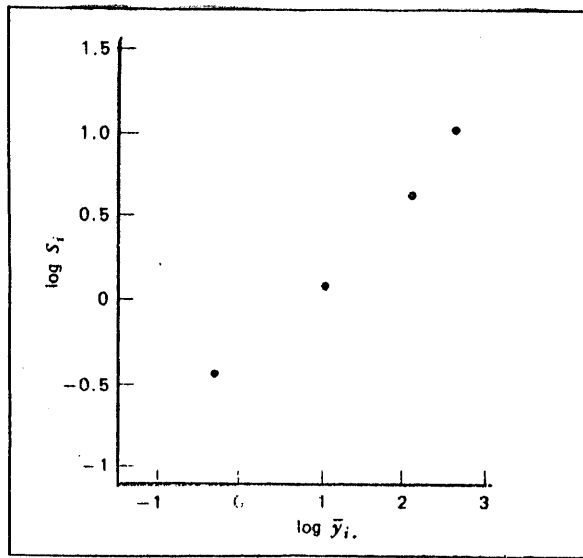


Figura 6.3. Gráfica de $\log s_i$ contra $\log \bar{y}_i$ para el Ejemplo 6.7.

El cuadro 6.24 indica que la transformación mediante la raíz cuadrada puede ser apropiada. El análisis de varianza para los datos transformados $Y = \sqrt{y}$ se presentan en el cuadro 6.27 y una gráfica de los residuos en la figura 6.4.

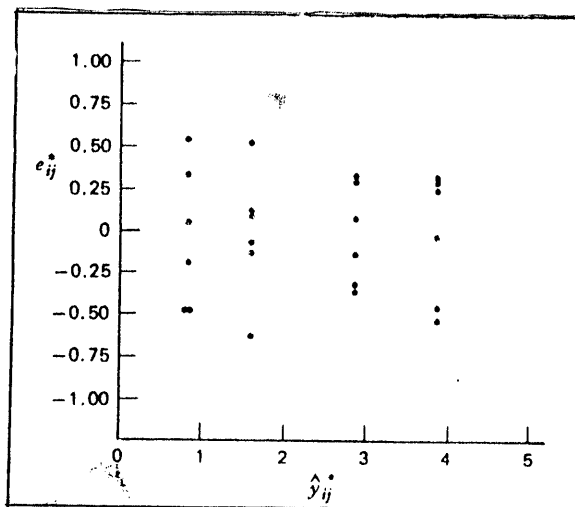


Figura 6.4. Gráfica de los residuos de los datos transformados contra \hat{Y}_{ij} del ejemplo 6.7.

Se concluye que la transformación raíz cuadrada ha sido útil porque esta

gráfica de residuos está muy mejorada en comparación con la Figura 6.2. Debe observarse que en la Tabla 6.27 se han reducido en uno los grados de libertad del error (ver Box y Cox, 1964), con el fin de garantizar el uso de los datos con objeto de estimar el parámetro de transformación α .

CUADRO 6.27. Análisis de varianza para los datos transformados $Y = \sqrt{y}$ del gasto máximo.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Método	032.6842	03	010.8947	76.99
Error	002.6884	20	000.1415	
Total	035.3726	22		

Determinación Análítica de la transformación.

Box y Cox (1964) mostraron cómo es posible estimar el parámetro λ en la transformación $Y = y^\lambda$ simultáneamente con los otros parámetros del modelo (la media general y los efectos de tratamiento), usando el método de máxima verosimilitud. Este procedimiento consiste en realizar un análisis de varianza estándar para diversos valores de λ sobre

$$y^{(\lambda)} = \frac{y^\lambda - 1}{\lambda y^{\lambda-1}}, \text{ si } \lambda \neq 0 \quad (6.7.1)$$

$$y^{(\lambda)} = \dot{y} \ln y, \text{ si } \lambda = 0.$$

donde $\dot{y} = \log_{10}^{-1} \left[\frac{\sum \log_{10} y}{n} \right]$ es la media geométrica de las observaciones, en otras palabras, la media geométrica se obtiene haciendo la media del $\log_{10} y$ y tomando el antilogaritmo del resultado de la media.

La estimación de máxima verosimilitud de λ es el valor para el cual la suma de cuadrados del error, $SC_E(\lambda)$ es mínima. Este valor λ se encuentra usualmente,

al hacer una gráfica de $SC_E(\lambda)$ contra λ y usarla para determinar el valor de λ que minimiza $SC_E(\lambda)$. Por lo regular, son suficientes entre 10 y 20 valores de λ para hacer una estimación óptima. Es posible realizar una segunda iteración usando una red más fina de valores si se requiere más precisión en la estimación de λ .

Hay que notar que no puede seleccionarse el valor de λ comparando directamente las sumas de cuadrados del error de los análisis de varianza de y^λ porque éstas se miden en una escala diferente para cada valor de λ . La Ecuación 6.7.1 reajusta las respuestas para que sean directamente comparables. Se recomienda que se use valores simples de λ porque las diferencias prácticas entre $\lambda = 0.50$ y $\lambda = 0.55$ por ejemplo, probablemente son pequeñas, siendo la primera más fácil de interpretar.

Un intervalo de confianza aproximado del $100(1 - \alpha)\%$ para λ puede determinarse primero calculando

$$SC^* = SC_E(\lambda) \left(1 + \frac{t_{\alpha/2, v}^2}{v} \right) \quad (6.7.2)$$

en donde v son los grados de libertad del error, y después, leyendo los correspondientes límites de confianza directamente de la gráfica. Si este intervalo de confianza incluye el valor de $\lambda = 1$, los datos indican que no es necesario realizar la transformación.

Ejemplo 6.8. Usaremos los datos del gasto máximo del cuadro 6.25. Calcular la media geométrica \hat{y} (de los datos originales) de la siguiente forma

$$\begin{aligned} \sum_{i=1}^n \log_{10} y &= \sum_{i=1}^{24} \log_{10} y = 12.425714 \\ \frac{\sum \log_{10} y}{n} &= \frac{12.425714}{24} = 0.5177381 \\ \hat{y} &= \log_{10}^{-1} \left[\frac{\sum \log_{10} y}{n} \right] = \log_{10}^{-1} (0.5177381) = 3.29411. \end{aligned}$$

Luego escogemos diferentes valores arbitrarios de λ , para encontrar la $SC_E(\lambda)$, por ejemplo, escogemos para $\lambda = 0.50$, calcúlmos

$$\lambda \hat{y}^{\lambda-1} = 0.5(3.29411)^{0.50-1} = 0.2754869,$$

para cada dato y original calcular

$$y^{(\lambda)} = \frac{y^\lambda - 1}{\lambda y^{\lambda-1}}.$$

Así, para $y = 0.34$, tenemos que

$$y^{(\lambda)} = \frac{(0.34)^{0.50} - 1}{0.2754869} = -1.513338,$$

para $y = 0.12$, tenemos que

$$y^{(\lambda)} = \frac{(0.12)^{0.50} - 1}{0.2754869} = -2.3724897,$$

para $y = 1.23$, tenemos que

$$y^{(\lambda)} = \frac{(1.23)^{0.50} - 1}{0.2754869} = +0.3958578,$$

hasta terminar con $y = 16.82$, que es

$$y^{(\lambda)} = \frac{(16.82)^{0.50} - 1}{0.2754869} = 11.257230.$$

Por lo que el cuadro de estos datos es

CUADRO 6.28. Datos del gasto máximo para $\lambda = 0.50$.

	<u>Método</u>			<u>Observaciones</u>		
1	-01.513338	-02.3724897	+00.3958578	-00.5929137	+01.1720182	-02.3724897
2	-00.164798	+02.5941082	+01.6802029	+01.9464778	+02.5088432	+04.1129828
3	+05.488360	+06.8718157	+07.7045369	+05.3279939	+07.7451520	+06.1372239
4	+11.402560	+08.8498674	+08.3818084	+11.4244570	+10.1207690	+11.2572300

El análisis de varianza para los datos cuando $\lambda = 0.50$ es

CUADRO 6.29. Análisis de varianza para el gasto máxima cuando $\lambda = 0.50$.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Método	430.6617	03	143.5539	81.05
Error	035.4240	20	001.7712	
Total	466.0857	23		

Por lo tanto, el valor de la $SC_E(\lambda) = 35.42$. Ahora, escogemos otro valor de λ , por ejemplo, para $\lambda = 0$, hacemos el mismo procedimiento.

Para cada dato y original calcular

$$y^{(\lambda)} = y \ln y.$$

Así, para $y = 0.34$, tenemos que

$$y^{(\lambda)} = (3.29411) \ln(0.34) = -3.55,$$

para $y = 0.12$, tenemos que

$$y^{(\lambda)} = (3.29411) \ln(0.12) = -6.98,$$

para $y = 1.23$, tenemos que

$$y^{(\lambda)} = (3.29411) \ln(1.23) = 0.68,$$

hasta terminar con $y = 16.82$, que es

$$y^{(\lambda)} = (3.29411) \ln(16.82) = 9.29.$$

Por lo que el cuadro de estos datos es

CUADRO 6.30. Datos del gasto máximo para $\lambda = 0$.

<u>Método</u>	<u>Observaciones</u>					
1	- 3.55	- 6.98	+ 0.68	- 1.18	+ 1.84	- 6.98
2	- 0.31	+ 3.55	+ 2.51	+ 2.83	+ 3.46	+ 4.99
3	+ 6.07	+ 6.99	+ 7.50	+ 5.95	+ 7.53	+ 6.52
4	+ 9.36	+ 8.14	+ 7.88	+ 9.37	+ 8.78	+ 9.29

El análisis de varianza para los datos cuando $\lambda = 0.50$ es

CUADRO 6.31. Análisis de varianza para el gasto máxima cuando $\lambda = 0$.

Fuente Varia.	<i>S.C.</i>	<i>g.l.</i>	<i>C.M.</i>	F_0
Método	461.0436	03	153.6811	33.48
Error	091.9622	20	004.5981	
Total	553.0058	23		

Por lo tanto, el valor de la $SC_E(\lambda) = 91.96$. Y se continua con los otros valores de λ que se seleccionaron para determinar la $SC_E(\lambda)$. Los valores de $SC_E(\lambda)$ para varios valores de λ son

λ	$SS_E(\lambda)$
-1.00	7922.11
-0.50	0687.10
-0.25	0232.52
0	0091.96
+0.25	0046.99
+0.50	0035.42
+0.75	0040.61
+1.00	0062.08
+1.25	0109.82
+1.50	0208.12

Una gráfica de los valores en la proximidad del mínimo aparece en la figura 6.5. De aquí puede concluirse que $\lambda = 0.52$ produce un valor mínimo aproximado de $SC_E(\lambda) = 35.00$. Un intervalo de confianza aproximado del 95 por ciento para λ se determina calculando SC^* en la Ecuación 6.7.2, como se muestra a continuación

$$SC^* = SC_E(\lambda) \left(1 + \frac{t_{0.025, 20}^2}{20} \right)$$

$$SC^* = 35.00 \left(1 + \frac{(2.086)^2}{20} \right)$$

$$SC^* = 42.61.$$

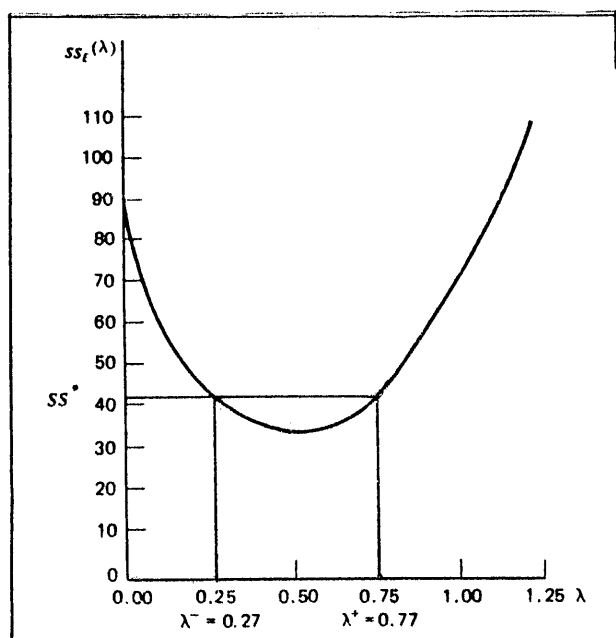


Figura 6.5. Gráfica de $SC_E(\lambda)$ contra λ .

Al ubicar este valor en la figura 6.5 y determinar los puntos en la escala de λ donde esta recta intersecta la curva, se obtienen los límites de confianza inferior y superior, $\lambda^- = 0.27$ y $\lambda^+ = 0.77$. Como este intervalo no incluye el valor de 1, es necesario utilizar alguna transformación. La transformación de acuerdo al cuadro 6.24, es la transformación de la raíz cuadrada ($\lambda = 0.5$). Notese que se usó anteriormente.

CONCLUSIONES.

Siempre es útil examinar los datos para asegurarse si existe o no correlación entre las medias de los tratamientos y sus varianzas dentro de los tratamientos. Desafortunadamente, esto no es posible para la mayoría de diseños. Si la variación muestra poco cambio, es dudoso el valor de cualquier transformación. Cuando hay cambio en la variación, no siempre es claro cuál es la mejor transformación.

Cuando hay duda sobre la transformación apropiada, a veces es útil transformar los datos de varios tratamientos, incluyendo algunos con medias pequeñas, intermedias y grandes en la escala original, y de nuevo examinar varianzas y medias para ver la posibilidad de alguna relación en la escala transformada.

Cuando se hace una transformación, todas las comparaciones de los intervalos de confianza se hace en la escala transformada. Si no se desea presentar resultados en la escala transformada, entonces las medias deben transformarse volviendo a la escala original. Cuando se presentan los resultados de un análisis de datos transformados en la escala original, ésta se le debe aclarar al investigador. No es apropiado transformar desviaciones estándar o varianzas provenientes de datos transformados, volviéndolas a la escala original.

Es así como se intentó presentar los aspectos más relevantes de los supuestos básicos del análisis estadístico de datos y las transformaciones apropiadas.

LITERATURA CITADA.

- Anderson, V.L. and McLean, R.A. 1974. Design of Experiments: A Realistic Approach. Ed. Marcel Dekker Inc. New York, United States of America. pp 16-27, 45-46.
- Bartlett, M.S. 1937. Properties of Sufficiency and statistical Test, Proc. Roy. Soc. A160, pp 268-282.
- Bartlett, M.S. and Kendall, D.G. 1946. The statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation. J. Roy. Stat. Soc. (London). 8, pp 128-138.
- Bartlett, M.S. 1947. The Use of Transformations. Biometrics, 3, pp 39-52.
- Box, P.E.G. y Cox, D.R. 1964. An Analysis of Transformations. Journal of the Royal Statistical Society, B, 26. pp 211-243.
- Box, P.E.G., Hunter, G.W. and Hunter, S.J. 1978. Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. John Wiley and Sons. New York, United States of America. pp 165-244.
- Burr, I.W. and Foster, L.A. 1972. A Test for Equality of Variances. Department of Statistics Mimeo, Purdue University. Lafayette, Indiana. Series No. 282.
- Canavos, G.C. 1986. Probabilidad y Estadística: Aplicaciones y Métodos. Primera Edición. Ed. Mc Graw-Hill. México. pp 362-419, 532-535.

- Cantú, H.V. 1980. Introducción a la Estadística. UAAAN. México. pp 1-31.
- Little, M.T. y Hills, J.F. 1990. Métodos Estadísticos para la Investigación en la Agricultura. Segunda Edición. Ed. Trillas. México. pp 125-143.
- Montgomery, C.D. 1991. Diseños y Análisis de Experimentos. Ed. Iberoamérica. México. pp 85-97.
- Mood, M.A., Graybill, A.F. and Boes, C.D. 1974. Introduction to the Theory of Statistics. Third Edition. Ed. McGraw-Hill. United States of America. pp 442-448, 508-510.
- Nater, J., Wasserman, W. and Kutner, H.M. 1989. Applied Linear Regression Models. Second Edition. Irwin, Inc. United States of America. pp 113-158.
- Ostle, B. 1983. Estadística Aplicada. Primera Edición. Ed. Limusa, S.A. México. pp 372-375, 513-514.
- Shapiro, S.S. and Wilk, M.B. 1965. An Analysis of Variance Test for Normality. Biometrika, 52, pp 591-612.
- Snedecor, W.G. y Cochran, G.W. 1975. Métodos Estadísticos. Primera Edición. Ed. Continental, S.A. México. pp 266-267, 396-417.
- Steel, G.D.R. y Torrie, H.J. 1986. Bioestadística: Principios y Procedimientos. Primera Edición. Ed. Mc Graw-Hill. pp 162-165, 226-230, 363-367.
- Tukey, J.W. 1949. One Degree of Freedom for Non Additivity. Biometrics, 5, 232-242.

APENDICE

Tabla A.1. Distribución normal estándar acumulativa.

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

z	.00	.01	.02	.03	.04	z
.0	.50000	.50399	.50798	.51197	.51595	.0
.1	.53983	.54379	.54776	.55172	.55567	.1
.2	.57926	.58317	.58706	.59095	.59483	.2
.3	.61791	.62172	.62551	.62930	.63307	.3
.4	.65542	.65910	.66276	.66640	.67003	.4
.5	.69146	.69497	.69847	.70194	.70540	.5
.6	.72575	.72907	.73237	.73565	.73891	.6
.7	.75803	.76115	.76424	.76730	.77035	.7
.8	.78814	.79103	.79389	.79673	.79954	.8
.9	.81594	.81859	.82121	.82381	.82639	.9
1.0	.84134	.84375	.84613	.84849	.85083	1.0
1.1	.86433	.86650	.86864	.87076	.87285	1.1
1.2	.88493	.88686	.88877	.89065	.89251	1.2
1.3	.90320	.90490	.90658	.90824	.90988	1.3
1.4	.91924	.92073	.92219	.92364	.92506	1.4
1.5	.93319	.93448	.93574	.93699	.93822	1.5
1.6	.94520	.94630	.94738	.94845	.94950	1.6
1.7	.95543	.95637	.95728	.95818	.95907	1.7
1.8	.96407	.96485	.96562	.96637	.96711	1.8
1.9	.97128	.97193	.97257	.97320	.97381	1.9
2.0	.97725	.97778	.97831	.97882	.97932	2.0
2.1	.98214	.98257	.98300	.98341	.98382	2.1
2.2	.98610	.98645	.98679	.98713	.98745	2.2
2.3	.98928	.98956	.98983	.99010	.99036	2.3
2.4	.99180	.99202	.99224	.99245	.99266	2.4
2.5	.99379	.99396	.99413	.99430	.99446	2.5
2.6	.99534	.99547	.99560	.99573	.99585	2.6
2.7	.99653	.99664	.99674	.99683	.99693	2.7
2.8	.99744	.99752	.99760	.99767	.99774	2.8
2.9	.99813	.99819	.99825	.99831	.99836	2.9
3.0	.99865	.99869	.99874	.99878	.99882	3.0
3.1	.99903	.99906	.99910	.99913	.99916	3.1
3.2	.99931	.99934	.99936	.99938	.99940	3.2
3.3	.99952	.99953	.99955	.99957	.99958	3.3
3.4	.99966	.99968	.99969	.99970	.99971	3.4
3.5	.99977	.99978	.99978	.99979	.99980	3.5
3.6	.99984	.99985	.99985	.99986	.99986	3.6
3.7	.99989	.99990	.99990	.99990	.99991	3.7
3.8	.99993	.99993	.99993	.99994	.99994	3.8
3.9	.99995	.99995	.99996	.99996	.99996	3.9

^aReproducido con permiso de *Probability and Statistics in Engineering and Management Science*, 3a. ed., por W. W. Hines y D. C. Montgomery, Wiley, Nueva York, 1990.

Tabla A.1. Distribución normal estándar acumulativa (continuación).

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

z	.05	.06	.07	.08	.09	z
.0	.51994	.52392	.52790	.53188	.53586	.0
.1	.55962	.56356	.56749	.57142	.57534	.1
.2	.59871	.60257	.60642	.61026	.61409	.2
.3	.63683	.64058	.64431	.64803	.65173	.3
.4	.67364	.67724	.68082	.68438	.68793	.4
.5	.70884	.71226	.71566	.71904	.72240	.5
.6	.74215	.74537	.74857	.75175	.75490	.6
.7	.77337	.77637	.77935	.78230	.78523	.7
.8	.80234	.80510	.80785	.81057	.81327	.8
.9	.82894	.83147	.83397	.83646	.83891	.9
1.0	.85314	.85543	.85769	.85993	.86214	1.0
1.1	.87493	.87697	.87900	.88100	.88297	1.1
1.2	.89435	.89616	.89796	.89973	.90147	1.2
1.3	.91149	.91308	.91465	.91621	.91773	1.3
1.4	.92647	.92785	.92922	.93056	.93189	1.4
1.5	.93943	.94062	.94179	.94295	.94408	1.5
1.6	.95053	.95154	.95254	.95352	.95448	1.6
1.7	.95994	.96080	.96164	.96246	.96327	1.7
1.8	.96784	.96856	.96926	.96995	.97062	1.8
1.9	.97441	.97500	.97558	.97615	.97670	1.9
2.0	.97982	.98030	.98077	.98124	.98169	2.0
2.1	.98422	.98461	.98500	.98537	.98574	2.1
2.2	.98778	.98809	.98840	.98870	.98899	2.2
2.3	.99061	.99086	.99111	.99134	.99158	2.3
2.4	.99286	.99305	.99324	.99343	.99361	2.4
2.5	.99461	.99477	.99492	.99506	.99520	2.5
2.6	.99598	.99609	.99621	.99632	.99643	2.6
2.7	.99702	.99711	.99720	.99728	.99736	2.7
2.8	.99781	.99788	.99795	.99801	.99807	2.8
2.9	.99841	.99846	.99851	.99856	.99861	2.9
3.0	.99886	.99889	.99893	.99897	.99900	3.0
3.1	.99918	.99921	.99924	.99926	.99929	3.1
3.2	.99942	.99944	.99946	.99948	.99950	3.2
3.3	.99960	.99961	.99962	.99964	.99965	3.3
3.4	.99972	.99973	.99974	.99975	.99976	3.4
3.5	.99981	.99981	.99982	.99983	.99983	3.5
3.6	.99987	.99987	.99988	.99988	.99989	3.6
3.7	.99991	.99992	.99992	.99992	.99992	3.7
3.8	.99994	.99994	.99995	.99995	.99995	3.8
3.9	.99996	.99996	.99996	.99997	.99997	3.9

Tabla A.2. Valores críticos de la distribución F .

gl del denominador	Probabilidad de un valor más alto de F	gl del numerador								
		1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
	.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
	.010	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
	.005	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.35	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.35	99.37	99.39
	.005	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.59	27.35
	.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51
8	.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
	.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
	.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72

Tabla A.2. Valores críticos de la distribución F (continuación).

gl del numerador											gl
10	12	15	20	24	30	40	60	120	∞	P	
60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33	.100	1
241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3	.050	
968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018	.025	
6056	6106	6157	6209	6235	6261	6287	6313	6339	6366	.010	
24224	24426	24630	24836	24940	25044	25148	25253	25359	25465	.005	
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49	.100	2
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	.050	
39.40	39.41	39.43	39.45	39.45	39.46	39.47	39.48	39.49	39.50	.025	
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50	.010	
199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5	.005	
5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13	.100	3
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	.050	
14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90	.025	
27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	.010	
43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83	.005	
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76	.100	4
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	.050	
8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26	.025	
14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	.010	
20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32	.005	
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10	.100	5
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	.050	
6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02	.025	
10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	.010	
13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14	.005	
2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72	.100	6
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	.050	
5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85	.025	
7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	.010	
10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88	.005	
2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47	.100	7
3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	.050	
4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14	.025	
6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	.010	
8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08	.005	
2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29	.100	8
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	.050	
4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67	.025	
5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	.010	
7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95	.005	
2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16	.100	9
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	.050	
3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33	.025	
5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	.010	
6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19	.005	
2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06	.100	10
2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	.050	
3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08	.025	
4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	.010	
5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64	.005	
2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97	.100	11
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	.050	
3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.89	.025	
4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	.010	
5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23	.005	
2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90	.100	12
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	.050	
3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72	.025	
4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	.010	
5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90	.005	
2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85	.100	13
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	.050	
3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60	.025	
4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	.010	
4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65	.005	
2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80	.100	14
2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	.050	
3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49	.025	
3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.13	3.09	3.00	.010	
4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44	.005	

Tabla A.2. Valores críticos de la distribución F (continuación).

gl del denominador	Probabilidad de un valor más alto de F	gl del numerador								
		1	2	3	4	5	6	7	8	9
15	.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
	.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
	.010	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
	.005	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.95
	.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	.005	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	.005	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
	.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
	.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	.005	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88
22	.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
	.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
	.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	.005	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
	.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
	.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
	.005	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.005	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
	.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
	.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	.005	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
	.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
	.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	.005	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
	.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
	.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	.005	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
28	.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
	.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
	.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
	.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
	.005	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52

Tabla A.2. Valores críticos de la distribución F (continuación).

gl del numerador											P	gl
10	12	15	20	24	30	40	60	120	∞			
2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76	.100	15	
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	.050		
3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.45	2.40	.025		
3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.04	2.95	2.87	.010		
4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.47	3.37	3.26	.005		
2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72	.100	16	
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.05	2.01	.050		
2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32	.025		
3.69	3.55	3.41	3.26	3.18	3.10	3.01	2.93	2.84	2.77	.010		
4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11	.005		
2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69	.100	17	
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	.050		
2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25	.025		
3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	.010		
4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98	.005		
1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66	.100	18	
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	.050		
2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19	.025		
3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	.010		
4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87	.005		
1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63	.100	19	
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	.050		
2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13	.025		
3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49	.010		
3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78	.005		
1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61	.100	20	
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	.050		
2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09	.025		
3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	.010		
3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69	.005		
1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59	.100	21	
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	.050		
2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04	.025		
3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	.010		
3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61	.005		
1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57	.100	22	
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	.050		
2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00	.025		
3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	.010		
3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55	.005		
1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55	.100	23	
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	.050		
2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97	.025		
3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	.010		
3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48	.005		
1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53	.100	24	
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	.050		
2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94	.025		
3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	.010		
3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43	.005		
1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52	.100	25	
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	.050		
2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91	.025		
3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	.010		
3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38	.005		
1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50	.100	26	
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	.050		
2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88	.025		
3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13	.010		
3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33	.005		
1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49	.100	27	
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	.050		
2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85	.025		
3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10	.010		
3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29	.005		
1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48	.100	28	
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	.050		
2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83	.025		
3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	.010		
3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25	.005		

Tabla A.2. Valores críticos de la distribución F (continuación).

gl del denominador	Probabilidad de un valor más alto de F	gl del numerador								
		1	2	3	4	5	6	7	8	9
28	.100	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
	.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
	.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
	.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
	.005	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	.010	7.56	5.33	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
40	.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
	.005	8.93	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
60	.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	.005	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
120	.100	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
	.050	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
	.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
	.010	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
	.005	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
∞	.100	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63
	.050	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
	.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11
	.010	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41
	.005	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

Fuente: Una parte de "Tables of percentage points of the inverted beta (F) distribution", *Biometrika*, vol. 33 (1943) por M. Merrington y C. M. Thompson y de la Tabla 18 de *Biometrika Tables for Statisticians*, vol. 1, Cambridge University Press, 1954, editado por E. S. Pearson y H. O. Hartley. Reproducido con permiso de los autores, editores, y de los fideicomisarios de *Biometrika*.

gl del numerador											
10	12	15	20	24	30	40	60	120	∞	P	gl
1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47	.100	29
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64	.050	
2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81	.025	
3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	.010	
3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21	.005	
1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46	.100	30
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	.050	
2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	.025	
2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	.010	
3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18	.005	
1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38	.100	40
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	.050	
2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	.025	
2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	.010	
3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93	.005	
1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29	.100	60
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	.050	
2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48	.025	
2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	.010	
2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69	.005	
1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19	.100	120
1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25	.050	
2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31	.025	
2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	.010	
2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43	.005	
1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00	.100	∞
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	.050	
2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.09	.025	
2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00	.010	
2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00	.005	

Tabla A.3. Puntos porcentuales de la distribución χ^2 .

ν	α								
	.995	.990	.975	.950	.500	.050	.025	.010	.005
1	0.00 +	0.00 +	0.00 +	0.00 +	0.45	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	1.39	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	2.37	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	3.36	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	4.35	11.07	12.38	15.09	16.75
6	0.68	0.87	1.24	1.64	5.35	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	6.35	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	7.34	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	8.34	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	9.34	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	10.34	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	11.34	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	12.34	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	13.34	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	14.34	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	15.34	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	16.34	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	17.34	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	18.34	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	19.34	31.41	34.17	37.57	40.00
25	10.52	11.52	13.12	14.61	24.34	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	29.34	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	39.34	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	49.33	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	59.33	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	69.33	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	79.33	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	89.33	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	99.33	124.34	129.56	135.81	140.17

ν - grados de libertad.

^aAdaptado, con permiso, de *Biometrika Tables for Statisticians*, Vol. 1, 3a. ed., por E. S. Pearson y H. O. Hartley, Cambridge University Press, Cambridge, 1966.

Tabla A.4. Valores de cuantiles superiores de la distribución
de la estadística D_n de Kolgomorov-Smirnov.

n	$1 - \alpha$				
	0.80	0.85	0.90	0.95	0.99
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Fórmula para una n mayor	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Fuente: F. J. Massey, Jr., *The Kolmogorov-Smirnov test for goodness of fit*, J. Amer Statistical Assoc. 46 (1951), 68-78. Publicado con permiso.

Tabla A.5. Puntos percentiles para la prueba Q .

PERCENTILE POINTS FOR Q-TEST, FOR EQUAL DEGREES OF FREEDOM v , AND FOR p SAMPLES								
p	$v = 1$		$v = 2$		$v = 3$		$v = 4$	
	.99	.999	.99	.999	.99	.999	.99	.999
3	*	*	.863	*	.757	.919	.684	.828
4	.920	*	.720	.898	.605	.754	.549	.675
5	.828	*	.608	.773	.512	.644	.443	.552
6	.744	.949	.539	.690	.430	.546	.369	.461
7	.671	.865	.469	.606	.372	.471	.318	.394
8	.609	.793	.412	.537	.325	.411	.276	.342
9	.576	.750	.371	.481	.287	.363	.244	.300
10	.528	.694	.333	.433	.257	.324	.218	.267
12	.448	.598	.276	.358	.211	.265	.179	.217
14	.391	.522	.234	.303	.178	.222	.151	.181
15	.365	.490	.217	.280	.165	.205	.140	.167
16	.343	.460	.202	.261	.154	.190	.130	.155
18	.304	.409	.178	.228	.135	.165	.114	.135
20	.273	.367	.158	.202	.120	.146	.101	.119
22	.246	.332	.142	.180	.108	.130	.090	.106
24	.224	.302	.129	.162	.098	.117	.082	.096
26	.206	.276	.118	.148	.090	.107	.075	.087
28	.190	.254	.108	.135	.082	.098	.069	.080
30	.176	.234	.100	.124	.075	.090	.064	.074
32	.163	.218	.093	.115	.070	.083	.060	.068
36	.143	.189	.082	.100	.062	.072	.052	.060
40	.127	.167	.072	.088	.055	.064	.047	.053
45	.111	.145	.063	.076	.048	.055	.041	.046
50	.098	.127	.056	.067	.043	.049	.037	.041
60	.080	.102	.045	.053	.035	.039	.030	.033
64	.074	.094	.042	.049	.033	.037	.028	.031

*These entries exceeded 1 using the approximating distribution. Since $Q \geq 1$, they are omitted.

Tabla A.5. Puntos percentiles para la prueba Q (continuación).

p	v = 5		v = 6		v = 8		v = 10	
	.99	.999	.99	.999	.99	.999	.99	.999
3	.631	.760	.593	.708	.539	.633	.512	.596
4	.498	.608	.461	.558	.413	.490	.383	.446
5	.399	.490	.368	.446	.328	.388	.303	.351
6	.334	.407	.307	.368	.271	.318	.250	.288
7	.284	.345	.261	.311	.230	.268	.212	.242
8	.246	.298	.226	.268	.199	.231	.184	.209
9	.217	.261	.199	.235	.176	.202	.162	.183
10	.194	.232	.178	.208	.157	.179	.145	.163
15	.123	.145	.113	.131	.101	.113	.094	.103
20	.090	.104	.083	.094	.074	.082	.069	.075
30	.058	.065	.053	.059	.048	.052	.045	.048
40	.042	.047	.039	.043	.035	.038	.033	.035
50	.033	.036	.031	.033	.028	.030	.026	.028
60	.027	.029	.025	.027	.023	.024	.022	.023

p	v = 12		v = 14		v = 16		v = 20	
	.99	.999	.99	.999	.99	.999	.99	.999
3	.486	.558	.466	.530	.451	.508	.429	.476
4	.362	.415	.347	.393	.335	.375	.319	.351
5	.287	.326	.275	.308	.265	.295	.252	.276
6	.236	.267	.227	.253	.219	.242	.209	.226
7	.201	.225	.192	.213	.186	.204	.178	.191
8	.174	.194	.167	.184	.162	.176	.154	.166
9	.154	.170	.148	.162	.143	.155	.136	.146
10	.137	.152	.132	.144	.128	.138	.122	.130
15	.089	.097	.086	.092	.083	.089	.080	.084
20	.066	.070	.063	.067	.062	.065	.059	.062
30	.043	.045	.042	.043	.040	.042	.039	.040
40	.032	.033	.031	.032	.030	.031	.029	.030
50	.025	.026	.024	.025	.024	.025	.023	.024
60	.021	.022	.020	.021	.020	.020	.019	.020

For $v > 60$, calculate $p\nu(pq-1)$ and compare with χ^2 with $(p-1)$ degrees of freedom in Appendix 7.

Tabla A.6. Coeficientes $\{a_{n-i+1}\}$ para la prueba W para normalidad.

Coefficients $\{a_{n-i+1}\}$ for the W test for normality,										
for $n = 2(1)50$										
$i \backslash n$	2	3	4	5	6	7	8	9	10	
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	
2		.0000	.1677	.2413	.2806	.3031	.3164	.3244	.3291	
3				.0000	.0875	.1401	.1743	.1976	.2141	
4						.0000	.0561	.0947	.1224	
5								.0000	.0399	
$i \backslash n$	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	.3315	.3325	.3325	.3318	.3306	.3290	.3273	.3253	.3232	.3211
3	.2260	.2347	.2412	.2460	.2495	.2521	.2540	.2553	.2561	.2565
4	.1429	.1586	.1707	.1802	.1878	.1939	.1988	.2027	.2059	.2085
5	.0695	.0922	.1099	.1240	.1353	.1447	.1524	.1587	.1641	.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7			.0000	.0240	.0433	.0593	.0725	.0837	.0932	.1013
8					.0000	.0196	.0359	.0496	.0612	.0711
9							.0000	.0163	.0303	.0422
10									.0000	.0140
$i \backslash n$	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	.3185	.3156	.3126	.3098	.3069	.3043	.3018	.2992	.2968	.2944
3	.2578	.2571	.2563	.2554	.2543	.2533	.2522	.2510	.2499	.2487
4	.2119	.2131	.2139	.2145	.2148	.2151	.2152	.2151	.2150	.2148
5	.1736	.1764	.1787	.1807	.1822	.1836	.1848	.1857	.1864	.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	.1092	.1150	.1201	.1245	.1283	.1316	.1346	.1372	.1395	.1415
8	.0804	.0878	.0941	.0997	.1046	.1089	.1128	.1162	.1192	.1219
9	.0530	.0618	.0696	.0764	.0823	.0876	.0923	.0965	.1002	.1036
10	.0263	.0368	.0459	.0539	.0610	.0672	.0728	.0778	.0822	.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12			.0000	.0107	.0200	.0284	.0358	.0424	.0483	.0537
13					.0000	.0094	.0178	.0253	.0320	.0381
14							.0000	.0084	.0159	.0227
15									.0000	.0076

Tabla A.6. Coeficientes $\{a_{n-i+1}\}$ para la prueba W para normalidad (continuación).

Coefficients $\{a_{n-i+1}\}$ for the W test for normality, for $n = 2(1)50$ (cont.)										
$n \setminus i$	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	.2921	.2898	.2876	.2854	.2834	.2813	.2794	.2774	.2755	.2737
3	.2475	.2463	.2451	.2439	.2427	.2415	.2403	.2391	.2380	.2368
4	.2145	.2141	.2137	.2132	.2127	.2121	.2116	.2110	.2104	.2098
5	.1874	.1878	.1880	.1882	.1883	.1883	.1883	.1881	.1880	.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	.1433	.1449	.1463	.1475	.1487	.1496	.1505	.1513	.1520	.1526
8	.1243	.1265	.1284	.1301	.1317	.1331	.1344	.1356	.1366	.1376
9	.1066	.1093	.1118	.1140	.1160	.1179	.1196	.1211	.1225	.1237
10	.0899	.0931	.0961	.0988	.1013	.1036	.1056	.1075	.1092	.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	.0585	.0629	.0669	.0706	.0739	.0770	.0798	.0824	.0848	.0870
13	.0435	.0485	.0530	.0572	.0610	.0645	.0677	.0706	.0733	.0759
14	.0289	.0344	.0395	.0441	.0484	.0523	.0559	.0592	.0622	.0651
15	.0144	.0206	.0262	.0314	.0361	.0404	.0444	.0481	.0515	.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17			.0000	.0062	.0119	.0172	.0220	.0264	.0305	.0343
18					.0000	.0057	.0110	.0158	.0203	.0244
19							.0000	.0053	.0101	.0146
20								.0000	.0049	.0094
$n \setminus i$	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	.2719	.2701	.2684	.2667	.2651	.2635	.2620	.2604	.2589	.2574
3	.2357	.2345	.2334	.2323	.2313	.2302	.2291	.2281	.2271	.2260
4	.2091	.2085	.2078	.2072	.2065	.2058	.2052	.2045	.2038	.2032
5	.1876	.1874	.1871	.1868	.1865	.1862	.1859	.1855	.1851	.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	.1531	.1535	.1539	.1542	.1545	.1548	.1550	.1551	.1553	.1554
8	.1384	.1392	.1398	.1405	.1410	.1415	.1420	.1423	.1427	.1430
9	.1249	.1259	.1269	.1278	.1286	.1293	.1300	.1306	.1312	.1317
10	.1123	.1136	.1149	.1160	.1170	.1180	.1189	.1197	.1205	.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	.0891	.0909	.0927	.0943	.0959	.0972	.0986	.0998	.1010	.1020
13	.0782	.0804	.0824	.0842	.0860	.0876	.0892	.0906	.0919	.0932
14	.0677	.0701	.0724	.0745	.0765	.0783	.0801	.0817	.0832	.0846
15	.0575	.0602	.0628	.0651	.0673	.0694	.0713	.0731	.0748	.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	.0379	.0411	.0442	.0471	.0497	.0522	.0546	.0568	.0588	.0608
18	.0283	.0318	.0352	.0383	.0412	.0439	.0465	.0489	.0511	.0532
19	.0188	.0227	.0263	.0296	.0328	.0357	.0385	.0411	.0436	.0459
20	.0094	.0136	.0175	.0211	.0245	.0277	.0307	.0335	.0361	.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22			.0000	.0042	.0081	.0118	.0153	.0185	.0215	.0244
23					.0000	.0039	.0076	.0111	.0143	.0174
24							.0000	.0037	.0071	.0104
25								.0000	.0035	.0068

Tabla A.7. Puntos porcentuales para la prueba W .

Percentage points of the W test for $n = 3(1)50$									
n	Level								
	0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	.687	.707	.748	.792	.935	.987	.992	.996	.997
5	.686	.715	.762	.806	.927	.979	.986	.991	.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	.730	.760	.803	.838	.928	.972	.979	.985	.988
8	.749	.778	.818	.851	.932	.972	.978	.984	.987
9	.764	.791	.829	.859	.935	.972	.978	.984	.986
10	.781	.806	.842	.869	.938	.972	.978	.983	.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	.805	.828	.859	.883	.943	.973	.979	.984	.986
13	.814	.837	.866	.889	.945	.974	.979	.984	.986
14	.825	.846	.874	.895	.947	.975	.980	.984	.986
15	.835	.855	.881	.901	.950	.975	.980	.984	.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	.851	.869	.892	.910	.954	.977	.981	.985	.987
18	.858	.874	.897	.914	.956	.978	.982	.986	.988
19	.863	.879	.901	.917	.957	.978	.982	.986	.988
20	.868	.884	.905	.920	.959	.979	.983	.986	.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	.878	.892	.911	.926	.961	.980	.984	.987	.989
23	.881	.895	.914	.928	.962	.981	.984	.987	.989
24	.884	.898	.916	.930	.963	.981	.984	.987	.989
25	.888	.901	.918	.931	.964	.981	.985	.988	.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	.894	.906	.923	.935	.965	.982	.985	.988	.990
28	.896	.908	.924	.936	.966	.982	.985	.988	.990
29	.898	.910	.926	.937	.966	.982	.985	.988	.990
30	.900	.912	.927	.939	.967	.983	.985	.988	.990
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	.904	.915	.930	.941	.968	.983	.986	.988	.990
33	.906	.917	.931	.942	.968	.983	.986	.989	.990
34	.908	.919	.933	.943	.969	.983	.986	.989	.990
35	.910	.920	.934	.944	.969	.984	.986	.989	.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	.914	.924	.936	.946	.970	.984	.987	.989	.990
38	.916	.925	.938	.947	.971	.984	.987	.989	.990
39	.917	.927	.939	.948	.971	.984	.987	.989	.991
40	.919	.928	.940	.949	.972	.985	.987	.989	.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	.922	.930	.942	.951	.972	.985	.987	.989	.991
43	.923	.932	.943	.951	.973	.985	.987	.990	.991
44	.924	.933	.944	.952	.973	.985	.987	.990	.991
45	.926	.934	.945	.953	.973	.985	.988	.990	.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	.928	.936	.946	.954	.974	.985	.988	.990	.991
48	.929	.937	.947	.954	.974	.985	.988	.990	.991
49	.929	.937	.947	.955	.974	.985	.988	.990	.991
50	.930	.938	.947	.955	.974	.985	.988	.990	.991

Tabla A.8. Transformación de la $\sqrt{\text{porcentaje}}$ arco seno.

Transformación de porcentajes binomiales, en los márgenes, a ángulos de igual información en grados. Los signos + y - que siguen a los valores angulares que terminan en 5 son la guía en la aproximación a un decimal.

%	0	1	2	3	4	5	6	7	8	9
0.0	0	0.57	0.81	0.99	1.15 -	1.28	1.40	1.52	1.62	1.72
0.1	1.81	1.90	1.99	2.07	2.14	2.22	2.29	2.36	2.43	2.50
0.2	2.56	2.63	2.69	2.75 -	2.81	2.87	2.92	2.98	3.03	3.09
0.3	3.14	3.19	3.24	3.29	3.34	3.39	3.44	3.49	3.53	3.58
0.4	3.63	3.67	3.72	3.76	3.80	3.85 -	3.89	3.93	3.97	4.01
0.5	4.05 +	4.09	4.13	4.17	4.21	4.25 +	4.29	4.33	4.37	4.40
0.6	4.44	4.48	4.52	4.55 +	4.59	4.62	4.66	4.69	4.73	4.76
0.7	4.80	4.83	4.87	4.90	4.93	4.97	5.00	5.03	5.07	5.10
0.8	5.13	5.16	5.20	5.23	5.26	5.29	5.32	5.35 +	5.38	5.41
0.9	5.44	5.47	5.50	5.53	5.56	5.59	5.62	5.65 +	5.68	5.71
1	5.74	6.02	6.29	6.55 -	6.80	7.04	7.27	7.49	7.71	7.92
2	8.13	8.33	8.53	8.72	8.91	9.10	9.28	9.46	9.63	9.81
3	9.98	10.14	10.31	10.47	10.63	10.78	10.94	11.09	11.24	11.39
4	11.54	11.68	11.83	11.97	12.11	12.25 -	12.39	12.52	12.66	12.79
5	12.92	13.05 +	13.18	13.31	13.44	13.56	13.69	13.81	13.94	14.06
6	14.18	14.30	14.42	14.54	14.65 +	14.77	14.89	15.00	15.12	15.23
7	15.34	15.45 +	15.56	15.68	15.79	15.89	16.00	16.11	16.22	16.32
8	16.43	16.54	16.64	16.74	16.85 -	16.95 +	17.05 +	17.16	17.26	17.36
9	17.46	17.56	17.66	17.76	17.85 +	17.95 +	18.05 -	18.15 -	18.24	18.34
10	18.44	18.53	18.63	18.72	18.81	18.91	19.00	19.09	19.19	19.28
11	19.37	19.46	19.55 +	19.64	19.73	19.82	19.91	20.00	20.09	20.18
12	20.27	20.36	20.44	20.53	20.62	20.70	20.79	20.88	20.96	21.05 -
13	21.13	21.22	21.30	21.39	21.47	21.56	21.64	21.72	21.81	21.89
14	21.97	22.06	22.14	22.22	22.30	22.38	22.46	22.55 -	22.63	22.71
15	22.79	22.87	22.95 -	23.03	23.11	23.19	23.26	23.34	23.42	23.50
16	23.58	23.66	23.73	23.81	23.89	23.97	24.04	24.12	24.20	24.27
17	24.35 +	24.43	24.50	24.58	24.65 +	24.73	24.80	24.88	24.95 +	25.03
18	25.10	25.18	25.25 +	25.33	25.40	25.48	25.55 -	25.62	25.70	25.77
19	25.84	25.92	25.99	26.06	26.13	26.21	26.28	26.35 -	26.42	26.49
20	26.56	26.64	26.71	26.78	26.85 +	26.92	26.99	27.06	27.13	27.20
21	27.28	27.35 -	27.42	27.49	27.56	27.63	27.69	27.76	27.83	27.90
22	27.97	28.04	28.11	28.18	28.25 -	28.32	28.38	28.45 +	28.52	28.59
23	28.66	28.73	28.79	28.86	28.93	29.00	29.06	29.13	29.20	29.27
24	29.33	29.40	29.47	29.53	29.60	29.67	29.73	29.80	29.87	29.93
25	30.00	30.07	30.13	30.20	30.26	30.33	30.40	30.46	30.53	30.59
26	30.66	30.72	30.79	30.85 +	30.92	30.98	31.05 -	31.11	31.18	31.24
27	31.31	31.37	31.44	31.50	31.56	31.63	31.69	31.76	31.82	31.88
28	31.95 -	32.01	32.08	32.14	32.20	32.27	32.33	32.39	32.46	32.52
29	32.58	32.65 -	32.71	32.77	32.83	32.90	32.96	33.02	33.09	33.15 -
30	33.21	33.27	33.34	33.40	33.46	33.52	33.58	33.65 -	33.71	33.77
31	33.83	33.89	33.96	34.02	34.08	34.14	34.20	34.27	34.33	34.39
32	34.45 -	34.51	34.57	34.63	34.70	34.76	34.82	34.88	34.94	35.00
33	35.06	35.12	35.18	35.24	35.30	35.37	35.43	35.49	35.55 -	35.61
34	35.67	35.73	35.79	35.85 -	35.91	35.97	36.03	36.09	36.15 +	36.21
35	36.27	36.33	36.39	36.45 +	36.51	36.57	36.63	36.69	36.75 +	36.81
36	36.87	36.93	36.99	37.05 -	37.11	37.17	37.23	37.29	37.35 -	37.41
37	37.47	37.52	37.58	37.64	37.70	37.76	37.82	37.88	37.94	38.00
38	38.06	38.12	38.17	38.23	38.29	38.35 +	38.41	38.47	38.53	38.59
39	38.65 -	38.70	38.76	38.82	38.88	38.94	39.00	39.06	39.11	39.17
40	39.23	39.29	39.35 -	39.41	39.47	39.52	39.58	39.64	39.70	39.76
41	39.82	39.87	39.93	39.99	40.05 -	40.11	40.16	40.22	40.28	40.34
42	40.40	40.46	40.51	40.57	40.63	40.69	40.74	40.80	40.86	40.92
43	40.98	41.03	41.09	41.15 -	41.21	41.27	41.32	41.38	41.44	41.50
44	41.55 +	41.61	41.67	41.73	41.78	41.84	41.90	41.96	42.02	42.07
45	42.13	42.19	42.25 -	42.30	42.36	42.42	42.48	42.53	42.59	42.65 -
46	42.71	42.76	42.82	42.88	42.94	42.99	43.05 -	43.11	43.17	43.22
47	43.28	43.34	43.39	43.45 +	43.51	43.57	43.62	43.68	43.74	43.80
48	43.85 +	43.91	43.97	44.03	44.08	44.14	44.20	44.25 +	44.31	44.37
49	44.43	44.48	44.54	44.60	44.66	44.71	44.77	44.83	44.89	44.94

Fuente: Esta tabla apareció en *Plant Protection* (Leningrad), 12: 67 (1937), y se reproduce con autorización del autor, C. I. Bliss.

