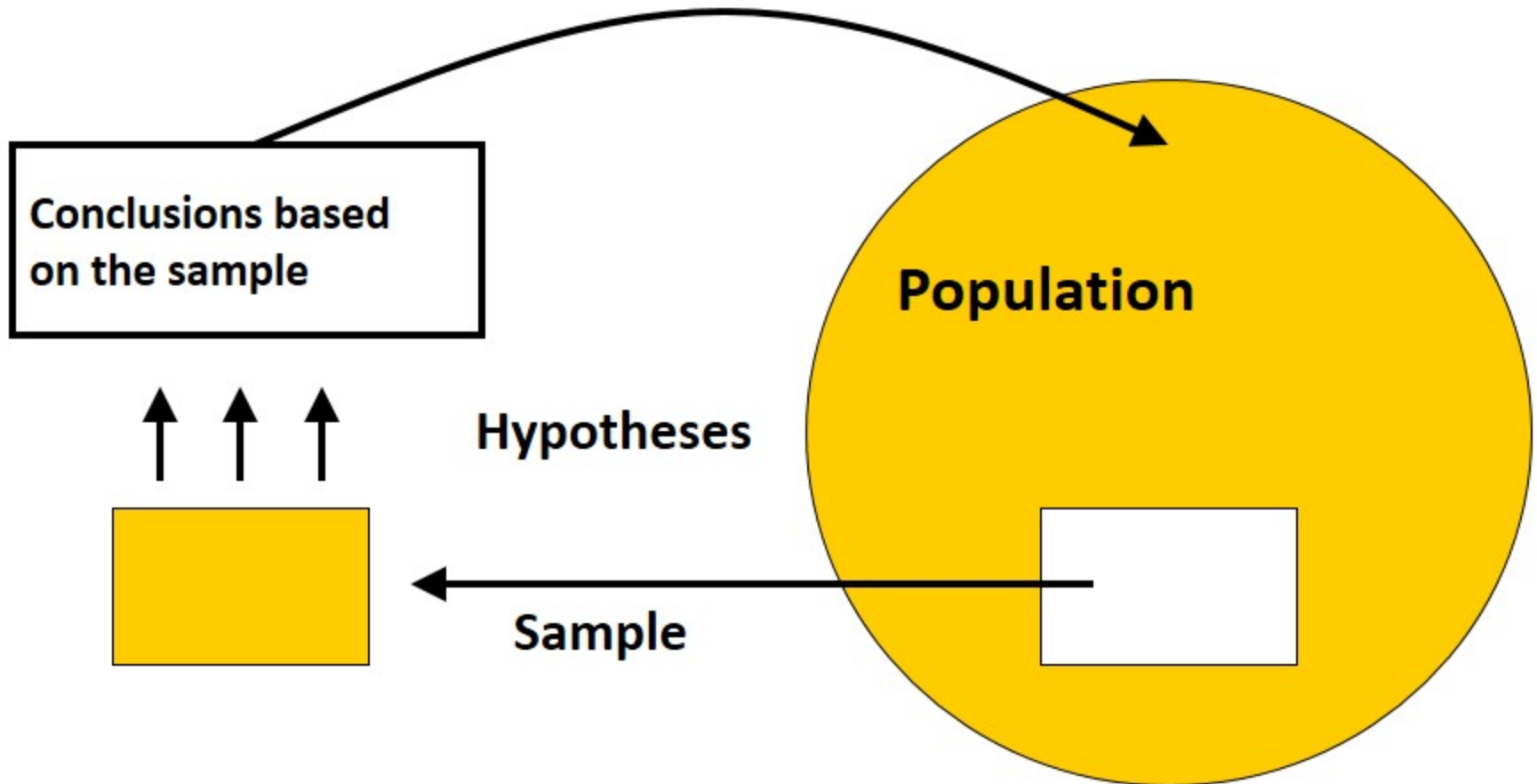


# The idea of statistical inference

Generalisation to the population

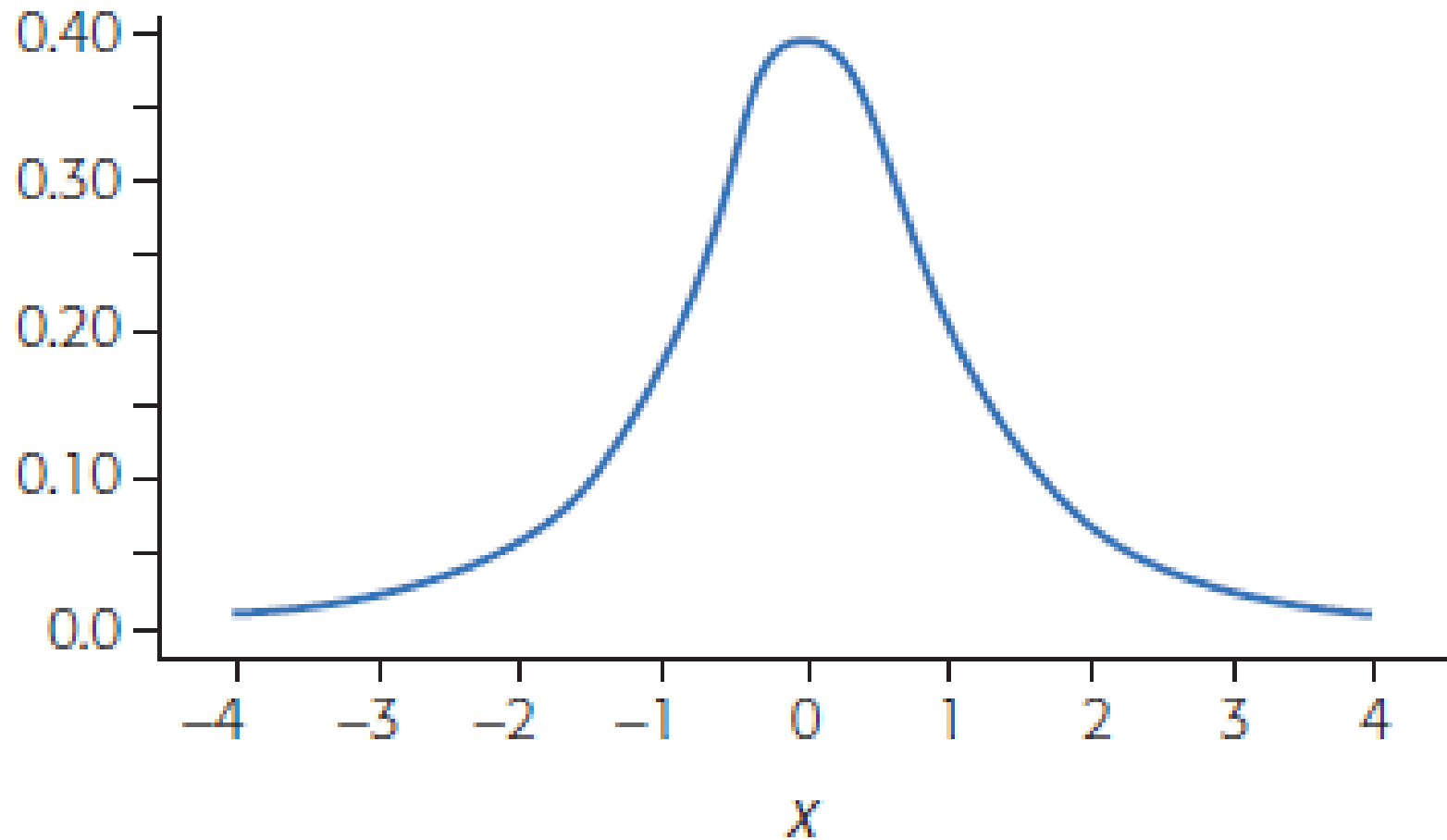


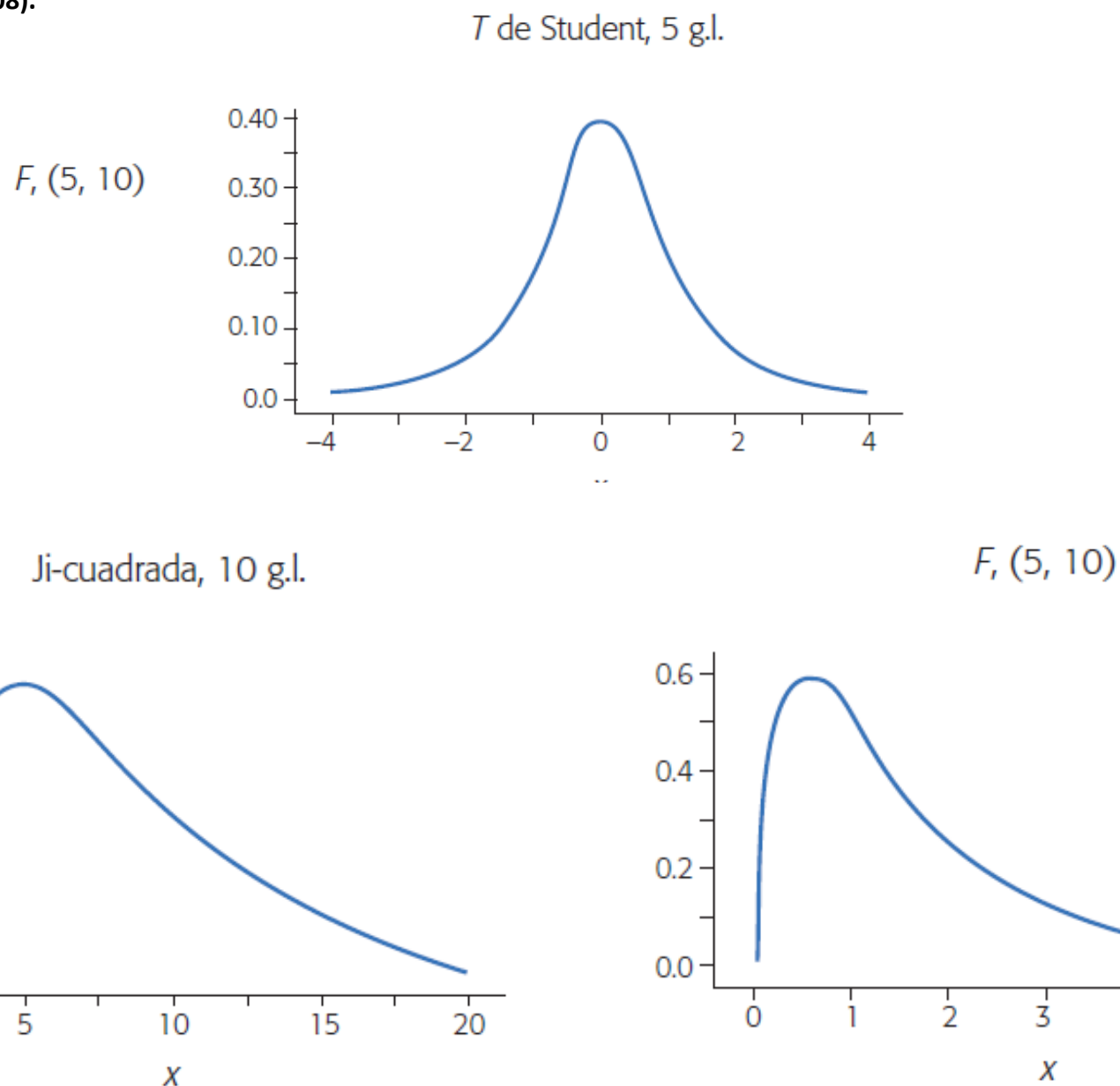
# Distribuciones de probabilidad e inferencia

- La *distribución de probabilidad* o *distribución de una variable aleatoria*  $X$  relaciona el conjunto de valores posibles de  $X$  (rango de  $X$ ), con la probabilidad asociada a cada uno de estos valores y los representa a través de una tabla o por medio de una función planteada como una fórmula.
- Las distribuciones de probabilidad que más se usan en intervalos de confianza y pruebas de hipótesis son las distribuciones: normal,  $T$  de *Student*, ji-cuadrada y F.
- La distribución normal está completamente definida por sus parámetros, que son la media,  $\mu$ , y la desviación estándar,  $\sigma$ .
- Por ejemplo: la distribución normal con  $\mu=0$  y  $\sigma=1$ , que se simboliza con  $N(0, 1)$  y se conoce como la distribución normal estándar.



## Normal estándar





**Figura 2.2** Muestra de las distribuciones de probabilidad de mayor uso en inferencia.

# **Estimación puntual y por intervalo**



# Estimación puntual

Un *estimador puntual* de un parámetro desconocido es un estadístico que genera un valor numérico simple, que se utiliza para hacer una estimación del valor del parámetro desconocido; por ejemplo, tres parámetros sobre los que con frecuencia se desea hacer inferencia son:

- La media  $\mu$  del proceso (población).
- La varianza  $\sigma^2$  o la desviación estándar  $\sigma$  del proceso.
- La proporción  $p$  de artículos defectuosos.

Los estimadores puntuales (estadísticos) más recomendados para estimar estos parámetros son, respectivamente:

- La media muestral  $\hat{\mu} = \bar{X}$
- La varianza muestral  $\hat{\sigma}^2 = S^2$ .
- La proporción de defectuosos en la muestra,  $\hat{p} = x/n$ , donde  $x$  es el número de artículos defectuosos en una muestra de tamaño  $n$ .



Estimar un parámetro poblacional mediante un estadístico que predice el valor de dicho parámetro.

$$\mu = \frac{\sum_{i \in N} x_i}{\sum_{i \in N} 1}$$

$$P_A = \frac{\text{casos de categoría A en } N}{N}$$

$$\sigma^2 = \frac{\sum_{i \in N} (x_i - \mu)^2}{N}$$

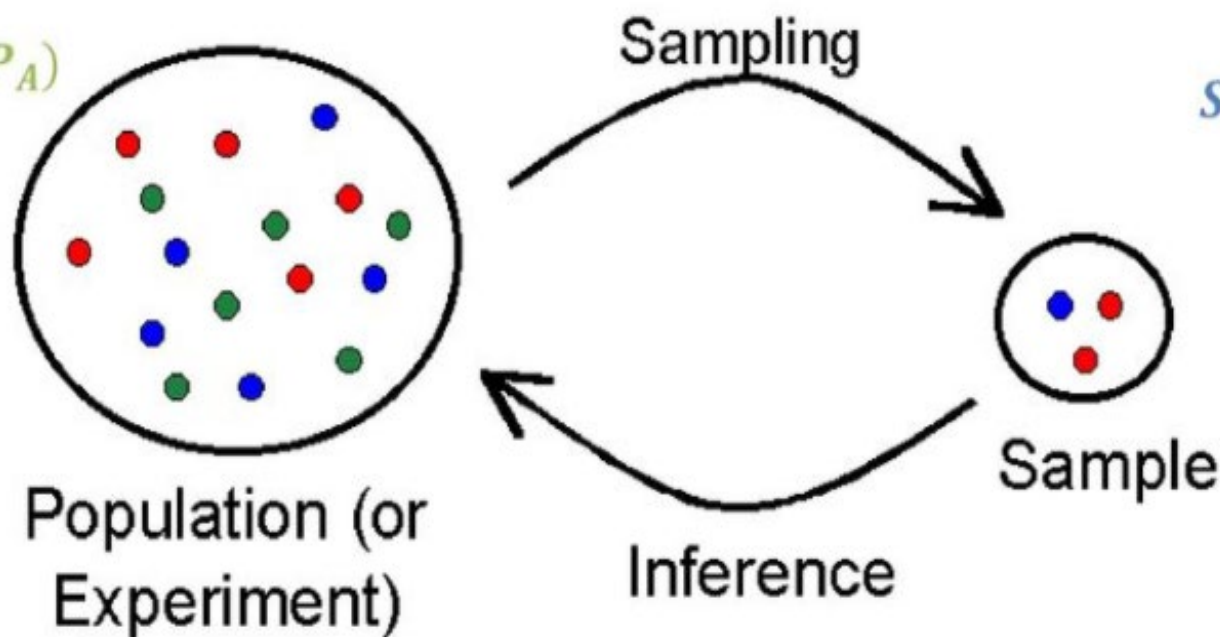
$$\sigma^2 = P_A(1 - P_A)$$

$$\bar{x} = \frac{\sum_{i \in n} x_i}{\sum_{i \in n} 1}$$

$$p_A = \frac{\text{casos de categoría A en } n}{n}$$

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

$$s^2 = \frac{n\hat{p}_A(1 - \hat{p}_A)}{n - 1}$$



# Estimación por intervalo

- La estimación puntual de un parámetro se genera a través de un estadístico, y como el valor de éste es aleatorio porque depende de los elementos que fueron seleccionados en la muestra, entonces la estimación que se hace sobre el parámetro dependerá y variará de una muestra a otra.
- De esta forma, cuando se quiere tener mayor certidumbre sobre el verdadero valor del parámetro poblacional, será necesario obtener la información sobre qué tan precisa es la estimación puntual.





# Estimación por intervalo

- Así, la estimación puntual dirá poco sobre el parámetro cuando la variación entre una estimación y otra es muy grande.
- Una forma de saber qué tan variable es el estimador, consiste en calcular la desviación estándar o error estándar del estadístico, visto como una variable aleatoria.
- La longitud del intervalo de confianza es una medida de la precisión de la estimación.
- **De aquí que es deseable que la longitud de los intervalos sea pequeña y con alto nivel de confianza.**
- El ancho de los intervalos es mayor a medida que sea mayor la varianza de la población y el nivel de confianza exigido. El ancho del intervalo es menor si se incrementa el tamaño de la muestra.



# Intervalo de confianza para una media

Por definición de *intervalo de confianza* se trata de encontrar dos números  $L$  y  $U$ , tales que el parámetro  $\mu$  se encuentre entre ellos con una probabilidad de  $1 - \alpha$ .

$$P(L \leq \mu \leq U) = 1 - \alpha$$

El cual sigue una distribución  $T$  de *Student* con  $n - 1$  grados de libertad. Por lo tanto, en la tabla de esta distribución o en su gráfica se pueden ubicar dos valores críticos  $t_{\alpha/2}$  y  $-t_{\alpha/2}$ , tales que:

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

$$P\left(-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2}\right) = 1 - \alpha$$

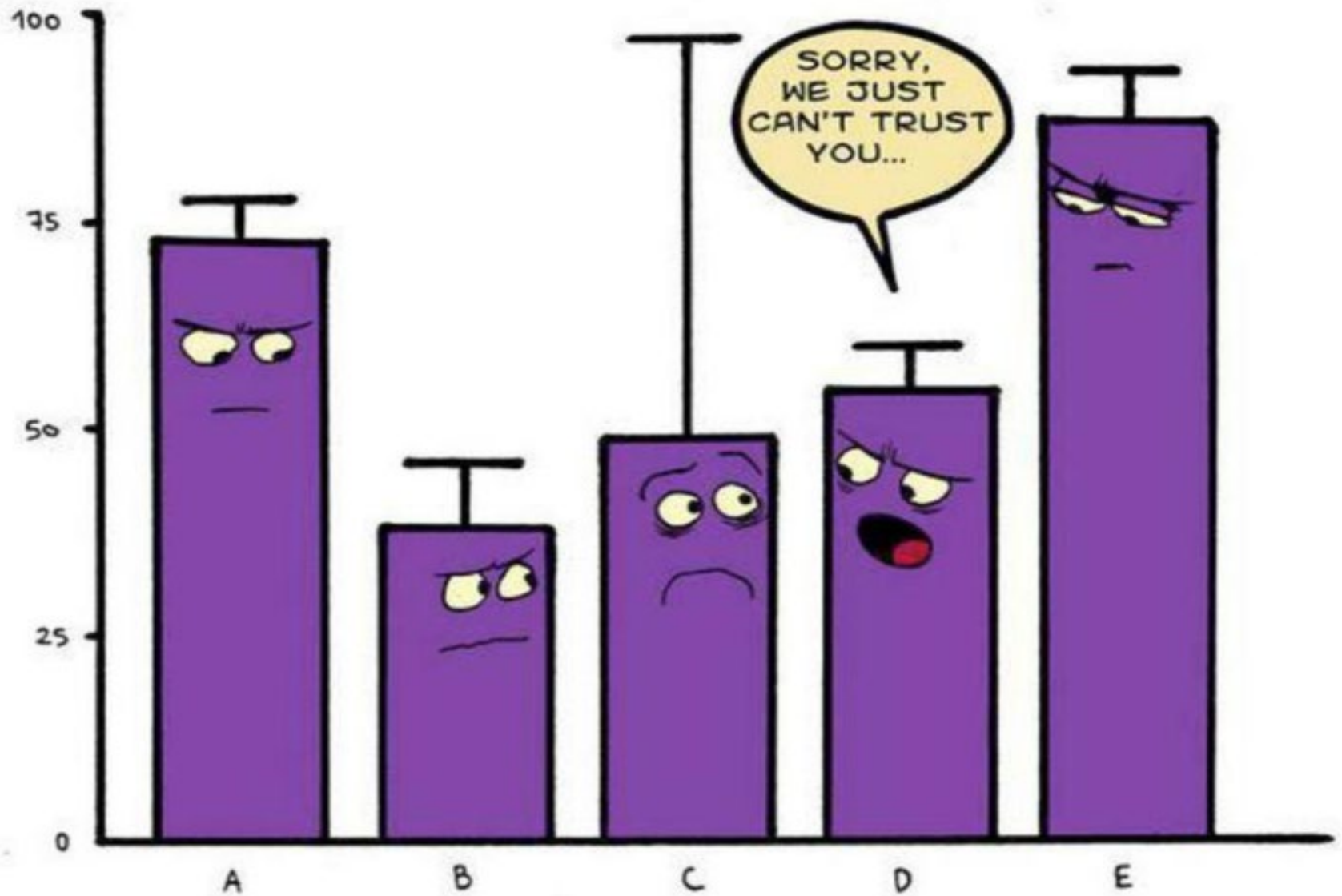
$$P\left(\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

# Intervalo para la varianza

En particular, para construir un intervalo de confianza para la varianza  $\sigma^2$ , la distribución de referencia es una ji-cuadrada con  $n - 1$  grados de libertad, ya que bajo el supuesto de que la variable de interés tiene una distribución normal con media y varianza desconocidas, el estadístico  $(n - 1)S^2/\sigma^2$  sigue la distribución ji-cuadrada con  $n - 1$  grados de libertad.

$$\frac{(n - 1)S^2}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq \frac{(n - 1)S^2}{\chi^2_{1-\alpha/2, n-1}}$$

# Precisión



# **Conceptos básicos de prueba de hipótesis**



Un estudio experimental o una investigación, por lo general tiene como último objetivo, responder en forma segura ciertas preguntas y/o tomar decisiones. En este contexto, el experimentador tiene *a priori* ciertas creencias o hipótesis que desea comprobar. Por ejemplo:

- El porcentaje de un ingrediente “x” afecta el resultado de la mezcla.
- Una fruta cultivada en diferentes pisos ecológicos tiene diferente contenido de un metabolito secundario.
- Si aumentamos la cantidad de reactivo “y” se mejora la extracción de un compuesto bioactivo.



# **Planteamiento de una hipótesis estadística**



- Una *hipótesis estadística* es una afirmación sobre los valores de los parámetros de una población o proceso, que es susceptible de probarse a partir de la información contenida en una muestra representativa que es obtenida de la población.
- Por ejemplo, la afirmación “este proceso produce menos de 8% de defectuosos” se puede plantear estadísticamente, en términos de la proporción  $p$  desconocida de artículos defectuosos que genera el proceso, como se hace a continuación.

$H_0 : p = 0.08$  (la proporción de defectuosos es 0.08)

$H_A : p < 0.08$  (la proporción es menor a 0.08)



- En general, la estrategia a seguir para **probar una hipótesis es suponer que la hipótesis nula es verdadera**, y que en caso de ser rechazada por la evidencia que aportan los datos, se estará aceptando la hipótesis alternativa.
- Así, en el caso de las proporciones, la afirmación que se desea probar se aceptará como cierta, sólo en caso de rechazar la hipótesis nula.

$H_0 : p = 0.08$  (la proporción de defectuosos es 0.08)

$H_A : p \neq 0.08$  (la proporción es diferente a 0.08)

# Estadístico de prueba



- Probar una hipótesis consiste en investigar si lo afirmado por la hipótesis nula es verdad o no.
- La estrategia de prueba parte del supuesto de que  $H_0$  es verdadera, y si los resultados de la investigación contradicen en forma suficiente dicho supuesto, entonces se rechaza  $H_0$  y se acepta la hipótesis alternativa.
- En caso de que los resultados de la investigación no demuestren claramente la falsedad de  $H_0$ , ésta no se rechaza.
- Es decir, *la hipótesis nula es verdadera mientras no se demuestre lo contrario.*



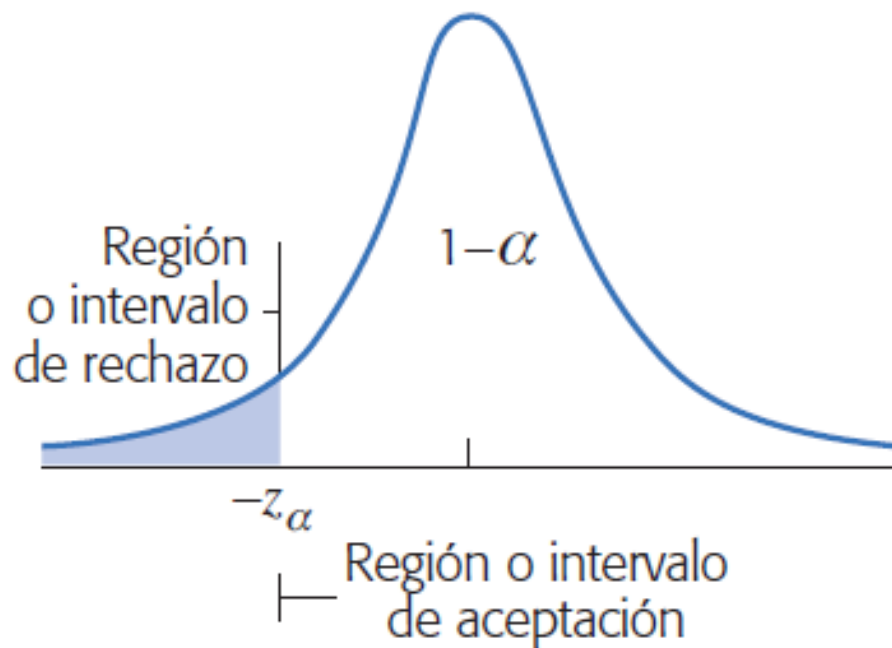
# **Criterio de rechazo**



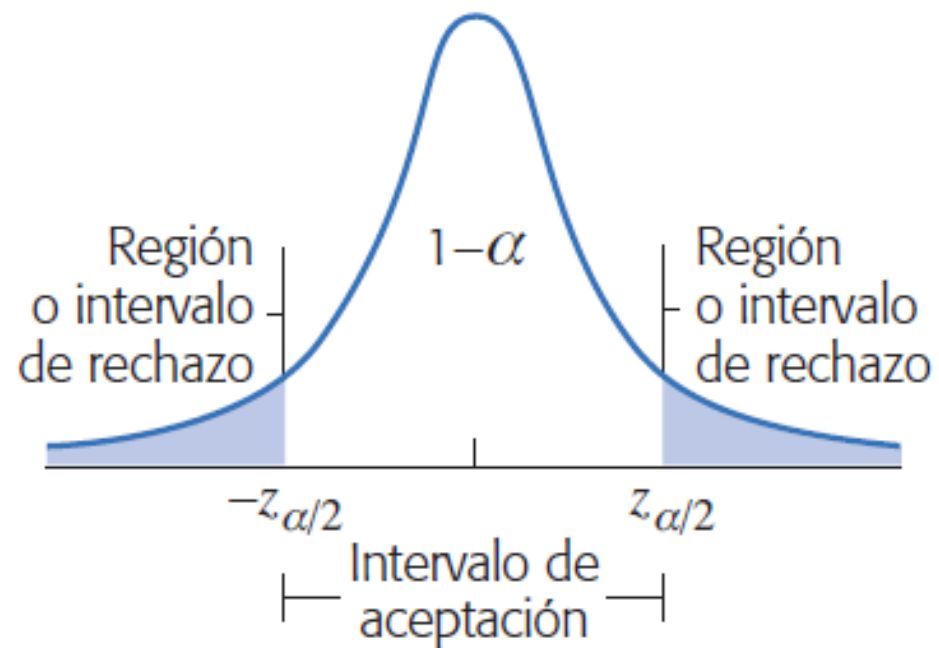
- El estadístico de prueba, construido bajo el supuesto de que  $H_0$  es verdad, es una variable aleatoria con distribución conocida.
- Si efectivamente  $H_0$  es verdad, el valor del estadístico de prueba debería caer dentro del rango de valores más probables de su distribución asociada, el cual se conoce como *región de aceptación*.
- Si cae en una de las colas de su distribución asociada, fuera del rango de valores más probables (en la región de rechazo), es evidencia en contra de que este valor pertenece a dicha distribución.
- **De aquí se deduce que debe estar mal el supuesto bajo el cual se construyó, es decir,  $H_0$  debe ser falsa.**



$H_0 : p = 0.08$   
 $H_A : p < 0.08$



$H_0 : p = 0.08$   
 $H_A : p \neq 0.08$



**Figura 2.3** Hipótesis unilateral y bilateral, regiones de aceptación y rechazo.

## ***El riesgo de una decisión equivocada:*** errores tipo I y tipo II

$\alpha = P\{\text{error tipo I}\} = \text{probabilidad de rechazar } H_0 \text{ siendo verdadera}$   
 $\beta = P\{\text{error tipo II}\} = \text{probabilidad de aceptar } H_0 \text{ siendo falsa}$

A “ $\alpha$ ” también se le conoce como la *significancia dada de la prueba* y es la probabilidad de la región o intervalo de rechazo; su valor se especifica por parte del investigador desde que planea el estudio.

Por lo general se utilizan los valores “ $\alpha$ ” = 0.05 o 0.01, dependiendo del riesgo que se quiera admitir en la conclusión.

- Mientras más pequeño es el valor de “ $\alpha$ ” se requiere más evidencia en los datos para rechazar  $H_0$ .
- Por ejemplo, si la acción a tomar después de rechazar  $H_0$  implica una inversión fuerte de recursos, se recomienda utilizar  $\alpha = 0.01$  para tener mayor confianza de que la decisión será la adecuada.
- Utilizar  $\alpha = 0.05$  significa que por cada 100 veces independientes que se aplica el procedimiento y se rechaza  $H_0$ , se espera que en un promedio de 95 veces, tal decisión sea la correcta.





- La probabilidad de *error tipo II* se controla de manera indirecta con el tamaño de la muestra, ya que a más datos  $\beta$  será menor.
- En otras palabras, con una muestra grande es mayor la *potencia de la prueba*, 1 es decir, se incrementa la probabilidad de rechazar  $H_0$  si ésta es falsa.



### Error tipo I

Es cuando se rechaza una  $H_0$  que es verdadera.



### Error tipo II

Es cuando se acepta una  $H_0$  que es falsa.



### Potencia de la prueba

Es la probabilidad de rechazar  $H_0$  cuando es falsa.

# Types of errors

		Truth	
		No diff $H_0$ to be not rejected	Diff $H_0$ to be rejected ( $H_1$ )
Decision based on the $p$ value	$H_0$ not rejected No diff	Right decision $1-\alpha$	$\beta$ <b>Type II error</b>
	$H_0$ rejected ( $H_1$ ) Diff	$\alpha$ <b>Type I error</b>	Right decision $1-\beta$

- $H_0$  is “true” but rejected: Type I or  $\alpha$  error
- $H_0$  is “false” but not rejected: Type II or  $\beta$  error

- **Por lo tanto, el error Tipo I (falso positivo) es peor que un error Tipo II (falso negativo).**
- La razón se reduce a la idea que si se mantiene el *status quo* o asunción por defecto, al menos no se estará haciendo las cosas peor.



**¿Qué es “peor”: un inocente preso o un culpable libre?;**





The Person is

Innocent



Guilty



The Judge Says



Innocent

No Error

Type 2 error



Guilty

Type 1 error

No Error



# Fórmula de Blackstone

- En el derecho penal la **fórmula de Blackstone** (también conocida como **ratio de Blackstone**) es un principio que establece que: “*es mejor que diez personas culpables escapen a que un inocente sufra*”.





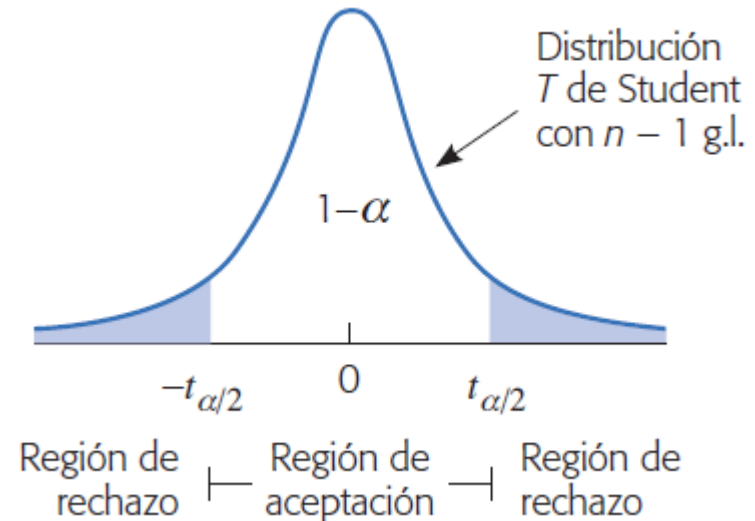
Nula	Error tipo I: $H_0$ verdadero, pero rechazado	Error tipo II: $H_0$ falso, pero no rechazado
Medicina A no alivia la Condición B.	Medicina A no alivia la Condición B, pero no se elimina como opción de tratamiento.	Medicina A alivia la Condición B, pero es eliminada como opción de tratamiento.
Consecuencias	Los pacientes con Condición B, que reciben la medicina A no se alivian. Pueden experimentar empeoramiento y/o efectos secundarios hasta incluso morir. Posible litigio.	Un tratamiento viable permanece inaccesible a pacientes con Condición B. Se pierden los costes del desarrollo. Provecho potencial eliminado.



- La *significancia predefinida* que se denota con “ $\alpha$ ”, es el riesgo máximo que el experimentador está dispuesto a correr por rechazar  $H_0$  indebidamente (error tipo I).
- Mientras que la *significancia observada* o *calculada*, también conocida como *p-value* o *valor-p*, es el área bajo la distribución de referencia más allá del valor del estadístico de prueba.



$$\text{valor-}p = P(T < -t_0) + P(T > +t_0)$$





- En la práctica **suele ser más delicado cometer el error tipo I que el error tipo II**, debido a que en la mayoría de las hipótesis el rechazar  $H_0$  implica objetar algo que se acepta de manera convencional.
- **No rechazar  $H_0$  implica, en muchos casos, seguir como “hasta ahora”.**
- Por lo anterior, es común que se controle sólo el error tipo I, mientras que el error tipo II se deja libre como si su magnitud no importara.
- Lo cierto es que **el error tipo II también importa y la magnitud de su probabilidad debe ser pequeña** (se recomienda  $\beta = 0.10$ ).
- El problema es que controlar a  $\beta$  tiene varios problemas; por ejemplo, muchas veces se requieren **grandes tamaños muestrales o se deben realizar muchas repeticiones en el experimento.**



**Significancia observada  
frente a significancia  
predefinida**



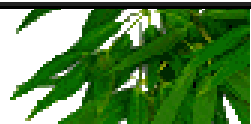
- La *significancia predefinida* que se denota con “ $\alpha$ ”, es el riesgo máximo que el experimentador está dispuesto a correr por rechazar  $H_0$  indebidamente (error tipo I).
- Mientras que la *significancia observada* o *calculada*, también conocida como *p-value* o **valor-p**, es el área bajo la distribución de referencia más allá del valor del estadístico de prueba.
- De lo anterior se desprende que  $H_0$  se rechaza si la **significancia observada** es menor que la **significancia dada**, o sea, si  $\text{valor-p} < \alpha$ .





- Con este criterio se puede ver como la probabilidad o evidencia a favor de  $H_0$ , **por lo tanto, representa una medida de la contundencia con la que se rechaza o no la hipótesis nula.**
- Por ejemplo, si la significancia observada o valor-p es igual a 0.0001, entonces sólo hay una probabilidad a favor de  $H_0$  de 0.0001, por lo que se rechazaría la hipótesis nula con un riesgo tipo I de 0.0001, **que es menor del que se está dispuesto a admitir, típicamente  $\alpha = 0.05$ .**

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	] — HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	] — SIGNIFICANT
0.049	
0.050	] — OH CRAP. REDO CALCULATIONS.
0.051	] — ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	] — HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	] — HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
$\geq 0.1$	



- En el mundo ideal, podríamos definir una muestra "perfectamente" aleatoria, la prueba más adecuada y una conclusión definitiva. Simplemente esa situación no podemos lograrla.
- Lo que podemos hacer es intentar optimizar todas las etapas de nuestra investigación para minimizar las fuentes de incertidumbre.
- Al presentar los valores de P, a algunos grupos les resulta útil utilizar el sistema de clasificación de asteriscos además de citar el valor de P:

**$P < 0.05$  \***

**$P < 0.001$  \*\*\***

**$P < 0.01$  \*\***

**$P > 0.05$  n.s.**



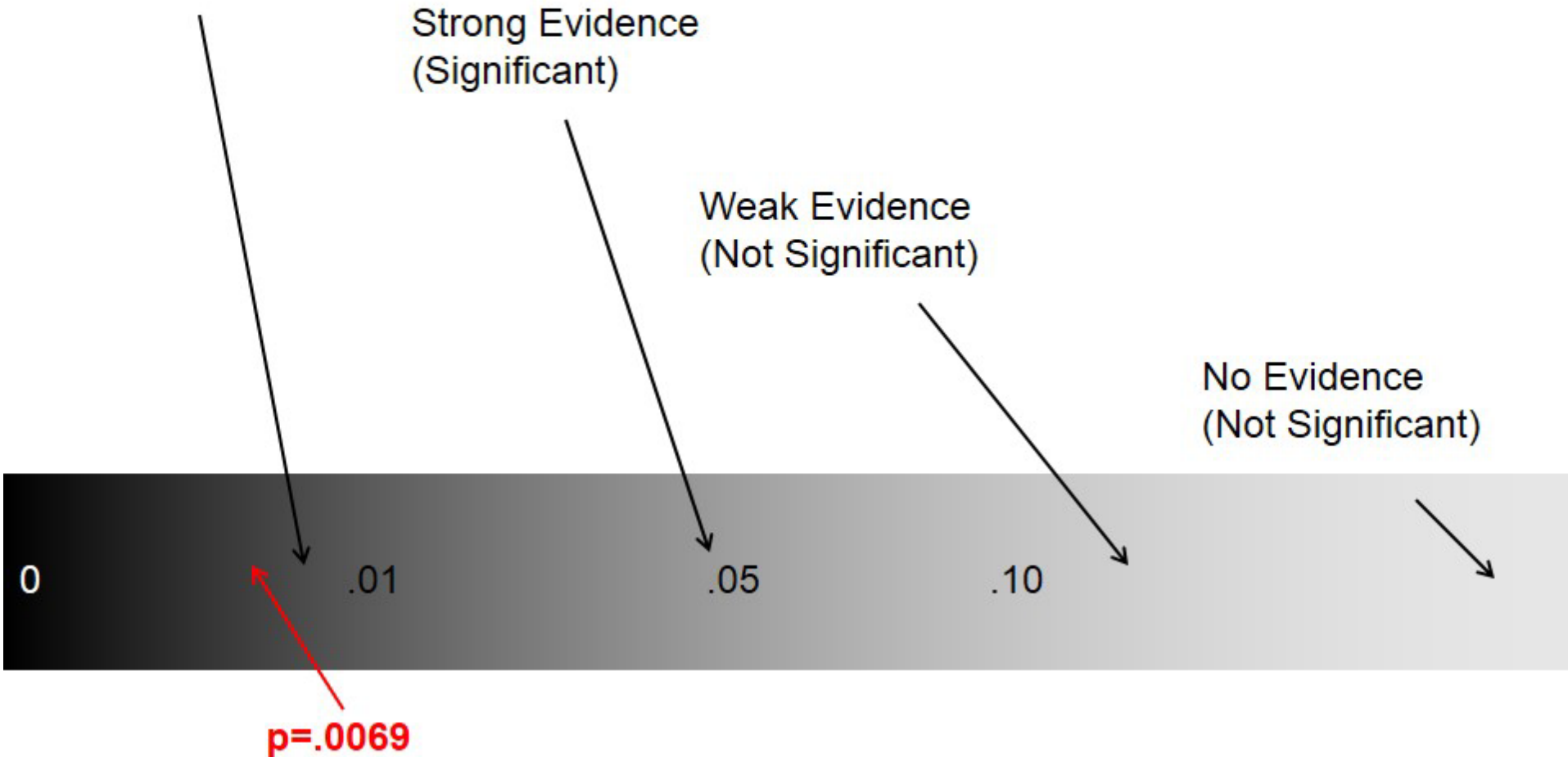
# Interpreting the p-value...

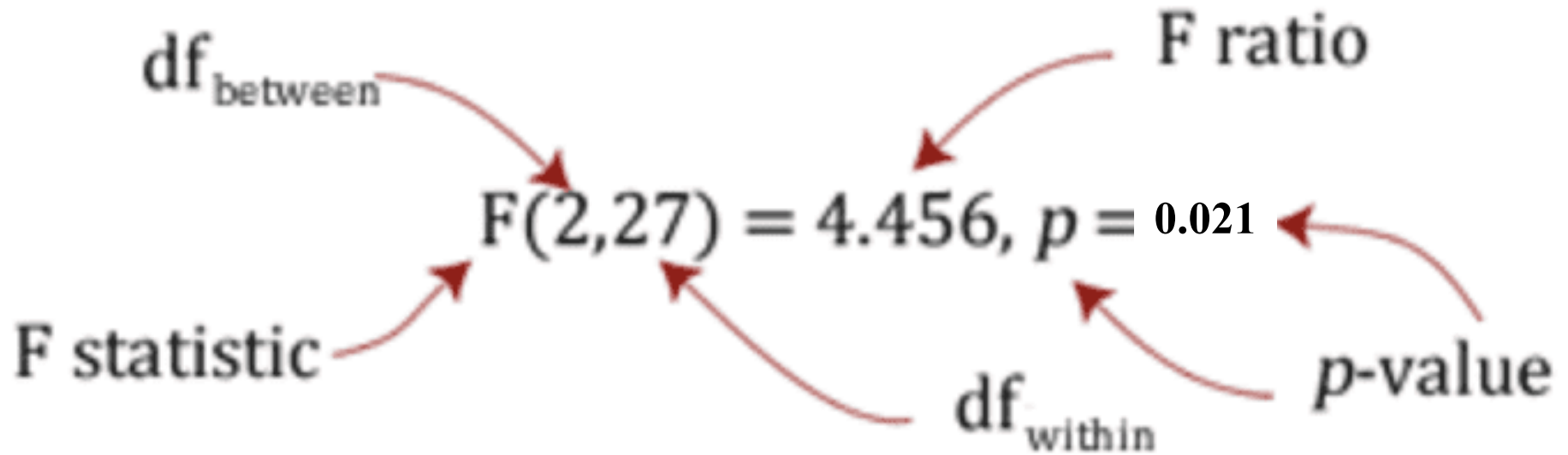
Overwhelming Evidence  
(Highly Significant)

Strong Evidence  
(Significant)

Weak Evidence  
(Not Significant)

No Evidence  
(Not Significant)





“Detectamos una diferencia estadísticamente significativa entre los grupos mediante el ANOVA de una vía ( $F(2,27) = 4,467, p = 0,021$ )”

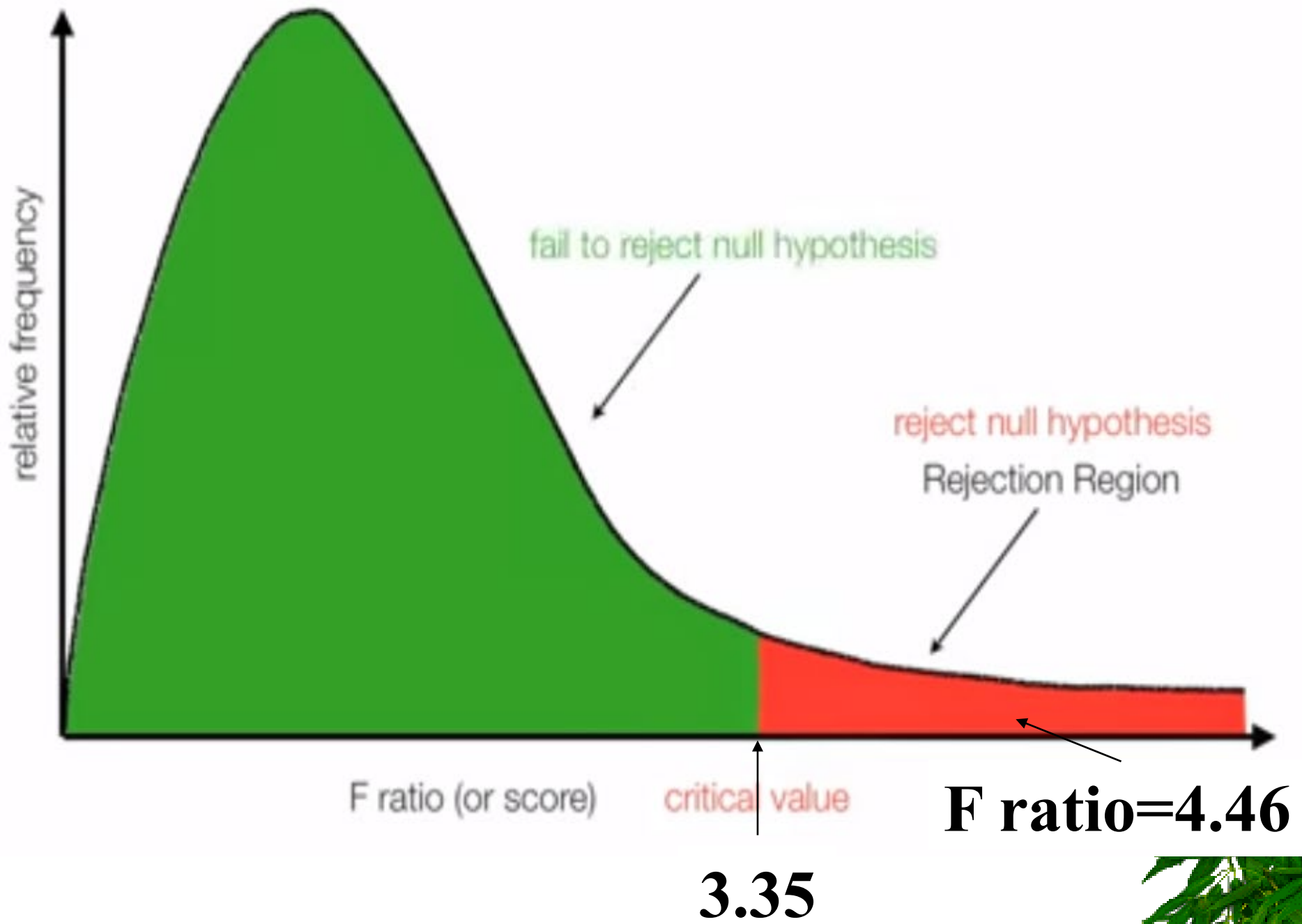
$$F_{0.05, \nu_1, \nu_2}$$

$\nu_2 \backslash \nu_1$	Grados de libertad del numerador ( $\nu_1$ )																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65



$$F_{0.05, \nu_1, \nu_2}$$

$\nu_2 \backslash \nu_1$	Grados de libertad del numerador ( $\nu_1$ )																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94													
4	7.71	6.94	6.59	6.39	6.26	6.16													
5	6.61	5.79	5.41	5.19	5.05	4.95													
6	5.99	5.14	4.76	4.53	4.39	4.28													
7	5.59	4.74	4.35	4.12	3.97	3.87													
8	5.32	4.46	4.07	3.84	3.69	3.58													
9	5.12	4.26	3.86	3.63	3.48	3.37													
10	4.96	4.10	3.71	3.48	3.33	3.22													
11	4.84	3.98	3.59	3.36	3.20	3.09													
12	4.75	3.89	3.49	3.26	3.11	3.00													
13	4.67	3.81	3.41	3.18	3.03	2.92													
14	4.60	3.74	3.34	3.11	2.96	2.85													
15	4.54	3.68	3.29	3.06	2.90	2.79													
16	4.49	3.63	3.24	3.01	2.85	2.74													
17	4.45	3.59	3.20	2.96	2.81	2.70													
18	4.41	3.55	3.16	2.93	2.77	2.66													
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.49	2.42	2.39	2.31	2.23	2.19	2.11	2.07	2.03	1.99	1.95	1.90
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65



# QuickCalcs

[1. Select category](#)

[2. Choose calculator](#)

[3. Enter data](#)

[4. View results](#)

## P Value Results

F=4.456 DF<sub>n</sub>=2 DF<sub>d</sub>=27

The P value equals 0.0213

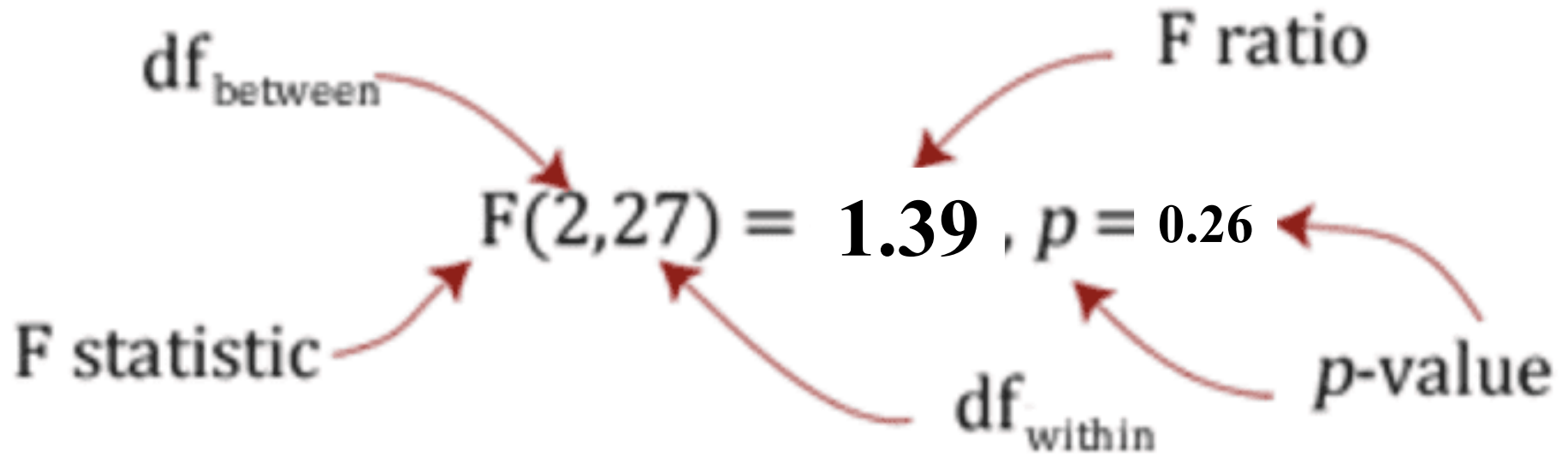
By conventional criteria, this difference is considered to be statistically significant.

Be sure you didn't accidentally mix up the two degrees of freedom. If you did so, the P value is not correct.

Adapted from Javascript written by John C. Pezzullo, PhD, Associate Professor, Pharmacology and Biostatistics Georgetown University Medical Center, and used with permission. The algorithms came from [Handbook of Math Functions](#) by Abramowitz.

FUENTE:

<https://www.graphpad.com/quickcalcs/pValue2/>



“No detectamos diferencias estadísticamente significativas entre las medias de los grupos como se determinó por ANOVA de una vía ( $F(2,27) = 1,397, p = 0,15$ ”

# Regla de decisión

**Elegimos el nivel alfa de significación que vamos a utilizar, usualmente  $\alpha=5\%$  o 0,05.**

- **Cuando el p-valor del estadístico de la prueba ANOVA es inferior al nivel alfa de significación que hemos elegido, entonces rechazamos la hipótesis nula y nos quedamos con la alternativa.**

- **Es decir, hay al menos dos medias grupales que son diferentes entre sí.**




- **En caso contrario, no podemos rechazar la hipótesis nula** y concluimos que no existen diferencias significativas entre los grupos evaluados.

- Es importante recordar que la prueba ANOVA no nos dice nada acerca de qué grupos específicos son diferentes.
- Para ello debemos realizar pruebas de comparación múltiple post hoc, que veremos más adelante.



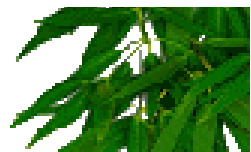
Se quiere medir el tamaño de partícula de dos sedimentos, para medir esta característica se utilizan dos centrifugadoras, y se sospecha que éstas reportan mediciones **distintas** para los mismos sedimentos analizados. Se decide hacer un estudio que permita comparar las medias y las varianzas reportadas por los dos equipos; para lo cual, se tomaron 13 lecturas de 13 sedimentos analizados de forma ordenada, con cada centrifugadora. Los resultados son los siguientes:

Centrifugadora x	4 714	4 601	4 696	4 896	4 905	4 870	4 987
	5 144	3 910	4 066	4 561	4 626	4 924	
	$\bar{X}_A =$	4 683.54		$S_A^2 =$	124 190.77		
Centrifugadora y	4 295	4 271	4 326	4 530	4 618	4 779	4 752
	4 744	3 764	3 797	4 401	4 339	4 700	
	$\bar{X}_B =$	4 408.92;		$S_B^2 =$	112 020.00		


$$H_0 : \mu_x = \mu_Y$$

$$H_A : \mu_x \neq \mu_Y$$

la cual se desea probar con un nivel de significancia de 5% ( $\alpha = 0.05$ ). Suponiendo igualdad de varianzas para el tamaño de la partícula, el estadístico de prueba calculado con las fórmulas (2.14) está dado por

$$t_0 = \frac{4\,684.00 - 4\,408.92}{344.06\sqrt{(1/13) + (1/13)}} = 2.04$$





## II. Puntos porcentuales de la distribución $t^e$

$\nu \backslash \alpha$	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690



De la tabla de distribución  $T$  de Student con  $13 + 13 - 2 = 24$  grados de libertad, se obtiene el punto crítico  $t_{(0.025, 24)} = 2.064$ . Como  $|t_0| = 2.04 < 2.064 = t_{\alpha/2}$ , no se rechaza  $H_0$ , por lo que se concluye que las centrifugadoras A y B reportan en promedio el mismo tamaño de partícula. Es decir, las centrifugadoras son estadísticamente iguales en cuanto a sus medias. Sin embargo, conviene observar que el rechazo es por un margen muy escaso, puesto que el estadístico de prueba y el punto crítico son muy similares. Al comparar la significancia predefinida  $\alpha = 0.05$  con el valor- $p = 0.053$  se concluye lo mismo (no se rechaza  $H_0$ ), pero se aprecia que si tal significancia predefinida por el experimentador fuera  $\alpha = 0.055$ , la decisión sobre la hipótesis sería rechazada.





## ANOVA Simple - valor por centrifuga

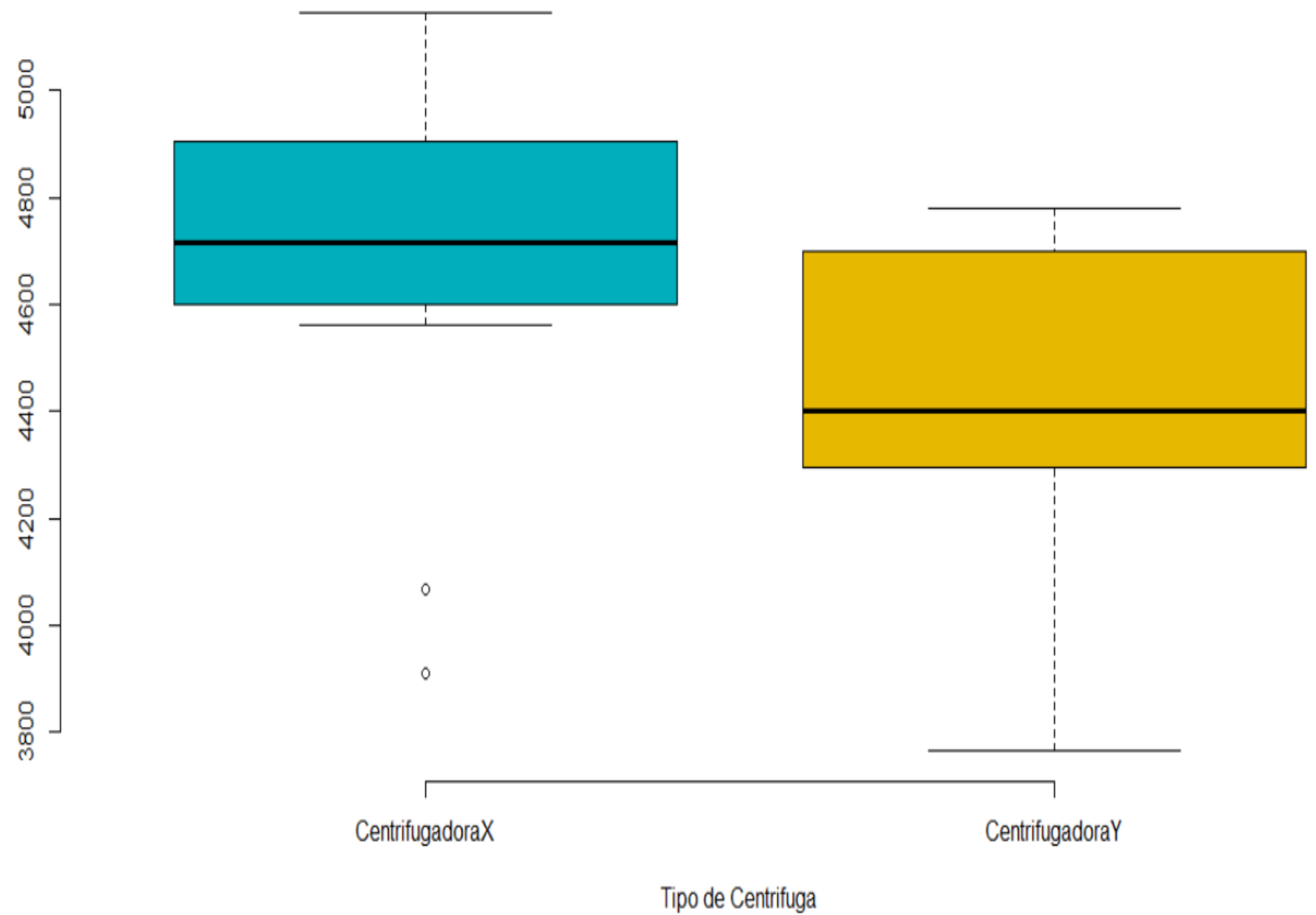
**Tabla ANOVA para valor por centrifuga**

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Entre grupos	494041.	1	494041.	4.17	0.0522
Intra grupos	2.84105E6	24	118377.		
Total (Corr.)	3.33509E6	25			

### El StatAdvisor

La tabla ANOVA descompone la varianza de valor en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 4.17345, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de valor entre un nivel de centrifuga y otro, con un nivel del 95.0% de confianza.





```
OneWayANOVA.R x
Source on Save
Run Source
33 ##ANOVA
34 anova(lm(VLOR ~ CENTRIFUGA, data = my_data))
35 # }
36
37 # Compute the analysis of variance
38 res.aov <- aov(VLOR ~ CENTRIFUGA, data = my_data)
39 # Summary of the analysis
40 summary(res.aov)
41
42
43
44 ##VERIFICACION DE SUPUESTOS EN FORMA GRÁFICA
45 par(mfrow=c(2,2))
46 plot(res.aov)
47 par(mfrow=c(1,1))
48
```

Console Jobs x

C:/Users/chris/Desktop/Proyeccion Social FIAL/Curso totales/2021/1. Diseños Experimentales/cursos/PRACTICAS/R/0. One way ANOVA/

```
      Df Sum Sq Mean Sq F value Pr(>F)
CENTRIFUGA 1 494041 494041  4.1734 0.05219 .
Residuals 24 2841050 118377
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # Compute the analysis of variance
> res.aov <- aov(VLOR ~ CENTRIFUGA, data = my_data)
> # Summary of the analysis
> summary(res.aov)
      Df Sum Sq Mean Sq F value Pr(>F)
CENTRIFUGA 1 494041 494041  4.173 0.0522 .
Residuals 24 2841050 118377
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Environment History Connections Tutorial

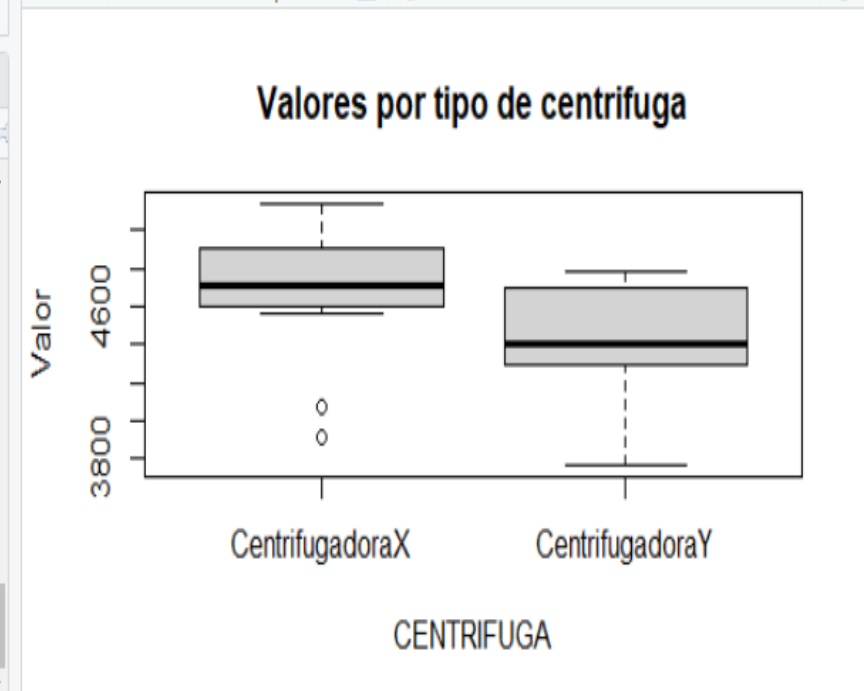
Import Dataset

Global Environment

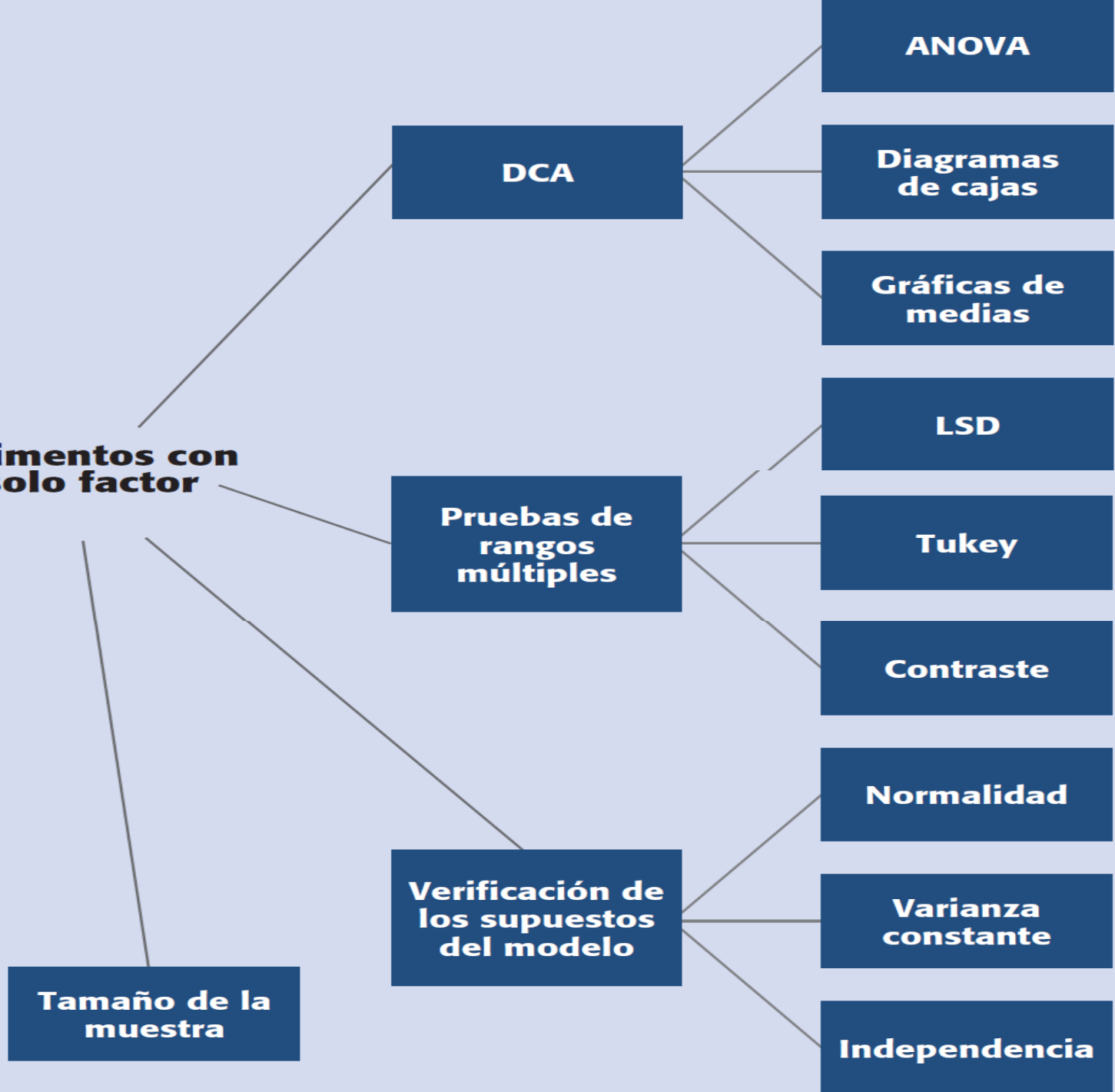
Name	Type	Length	Size	Value
my_data	data.fra...	2	1.6 KB	26 obs. of 2 variab...
res.aov	aov	13	22 KB	List of 13

Files Plots Packages Help Viewer

Zoom Export Publish



## **Experimentos con un solo factor**



- Por lo general, el interés del experimentador está centrado en comparar los tratamientos en cuanto a sus medias poblacionales, sin olvidar que también es importante compararlos con **respecto a sus varianzas**.
- Así, desde el punto de vista estadístico, la hipótesis fundamental a probar cuando se comparan varios tratamientos es:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k = \mu$$

$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j$$

- **Hipótesis nula**,  $H_0$ : las medias de los grupos son iguales.
- **Hipótesis alternativa**,  $H_1$ : alguna de las medias es distinta.
- Donde “ $\mu$ ” es la media del grupo y “ $k$ ” el número de



- El análisis de varianza (ANOVA) de una vía se utiliza para determinar si existen diferencias estadísticamente significativas entre las medias de tres o más grupos.
- En este caso utilizaremos **grupos independientes** (no relacionados) por lo que lo llamaremos **ANOVA de un factor entre-grupos**.
- Luego de realizar el ANOVA, si los grupos presentan un comportamiento distinto, podremos determinar **entre qué grupos específicos existen diferencias estadísticamente significativas** mediante **pruebas de comparación múltiple *post hoc***.
- El nombre *Analysis of Variance* (ANOVA), análisis de varianza, se basa en el enfoque en el que el procedimiento utiliza varianzas para determinar si las medias son diferentes.





- **En el caso particular del *one-way* ANOVA, “una vía” significa que tenemos una única variable explicativa o predictor, también llamada variable independiente.**
- **Esta variable debe tener tres o más niveles o categorías.**
- El ANOVA, fue desarrollado por Ronald Fisher en 1918, extiende la prueba t y la prueba z que comparan tan solo 2 grupos, quien fue un estadístico y biólogo que usó la matemática para combinar las leyes de Mendel con la selección natural, de manera que ayudó así a crear una nueva síntesis del Darwinismo conocida como la síntesis evolutiva moderna.



**DCA**



- Muchas comparaciones de tratamientos en un análisis experimental, se hacen con base en el diseño completamente al azar (DCA), que es el más simple de todos los diseños que se utilizan para comparar tres o más tratamientos, dado que sólo consideran dos fuentes de variabilidad: los *tratamientos* y el *error aleatorio*.
- También existen los diseños que consideran la influencia de otras fuentes de variabilidad (bloques).
- Este diseño se llama completamente al azar porque todas las corridas experimentales se realizan en orden aleatorio completo.



- De esta manera, si durante el estudio se hacen en total N pruebas, éstas se corren al azar, de manera que los posibles efectos ambientales y temporales se vayan repartiendo equitativamente entre los tratamientos.

## Modelo Estadístico en el DCA

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad i = 1, 2, \dots, k; j = 1, 2, \dots, n$$

El i-ésimo tratamiento en la j-ésima repetición.

La media poblacional

Efecto de i-ésimo tratamiento

El error experimental asociado a la observación  $y_{ij}$ .

# Supuestos del modelo en el DCA

## NORMALIDAD DE LOS RESIDUOS



Se debe verificar si los residuos siguen una distribución normal

$H_0$ : Los residuos se distribuyen normalmente

$H_1$ : Los residuos *no* se distribuyen normalmente

Si el P-valor es  $> \alpha$  No se rechaza  $H_0 \rightarrow$  Se cumple el supuesto de Normalidad

## HOMOCEDASTICIDAD



Se debe verificar la homogeneidad de las varianzas

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

$H_1$ : Al menos una varianza es diferente

Si P-valor  $> \alpha$  No se rechaza  $H_0 \rightarrow$  Se cumple el supuesto de Homocedasticidad

# Supuestos del modelo DCA

- Los datos deben ser descritos adecuadamente por el modelo:
- 

Donde:

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} \rightarrow N(0, \sigma^2)$$

- Los errores deben satisfacer lo siguiente:
  - Son v.a con distribución normal con media cero (**normalidad** de residuos)
  - Tienen varianza constante (**homocedásticidad**)
  - Son variables aleatorias independientes (**no autocorrelación** de residuos)

### 3.1.3. Pasos para la realización del análisis de varianza

#### 1. Obtener el Factor de Corrección (F.C)

$$F.C. = \frac{(Y_{..})^2}{r \times t}; Y_{..} = \sum_{i=1}^t \sum_{j=1}^r Y_{ij}$$

#### 2. Obtener la Suma de Cuadrados Total SC Total

$$SCTotal = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - F.C.$$

### 3. Obtener Suma de Cuadrados de Tratamientos

$$SC_{Tratamientos} = \frac{\sum_{i=1}^t (Y_i)^2}{r} - (F.C.)$$

### 4. Obtener Suma de Cuadrados del Error

$$SC_{Error} = SC_{Total} - SC_{Tratamientos}$$

### 5. Obtener Grados de Libertad

$$g.l.Tratamientos = t - 1$$

$$g.l.Error = t(r - 1)$$

$$g.l.Total = (t \times r) - 1$$



## 6. Obtener Cuadrados Medios

$$CM_{Trat} = \frac{SC_{Trat}}{g.l. Trat}$$

$$CM_{Error} = \frac{SC Error}{g.l. Error}$$

## 7. Obtener Valores de F

$$F_{trat} = \frac{CM_{trat}}{CM Error}$$

## 8. Obtener el Coeficiente de Variación

$$CV(\%) = \left( \frac{\sqrt{CM Error}}{\bar{Y}_{..}} \right) \times 100 \quad \bar{Y}_{..} = \text{Media general}$$

## 9. Obtener el coeficiente de determinación $R^2$

$$R^2 = \frac{SC_{Tratamientos}}{SC_{Total}}$$

- $R^2$  indica la proporción de la suma de cuadrados total que es explicada por la variación entre tratamientos.
- Conforme el valor de  $R^2$  se aproxima a 1.0 esto indicará que los datos analizados tuvieron un mejor ajuste del modelo lineal aditivo

## Hipótesis Estadística en el DCA

$H_0: \tau_i = 0$  (Todos los tratamientos producen el mismo efecto sobre la variable en estudio)

$H_1: \tau_i \neq 0$  (Al menos uno de los tratamientos produce un efecto diferente sobre la variable en estudio)

O equivalentemente:

$H_0: \mu_1 = \mu_2 = \dots = \mu_i = \mu$  (Todos los tratamientos tienen la misma media)

$H_1: \text{Al menos un } \mu_i \text{ es diferente}$

## Caso: Entrenamiento para empleados nuevos

Repetición	Método 1	Método 2	Método 3
1	28	18	22
2	27	17	18
3	25	19	19
4	21	16	22
5	23	22	17

Con un nivel de significación del 5%,

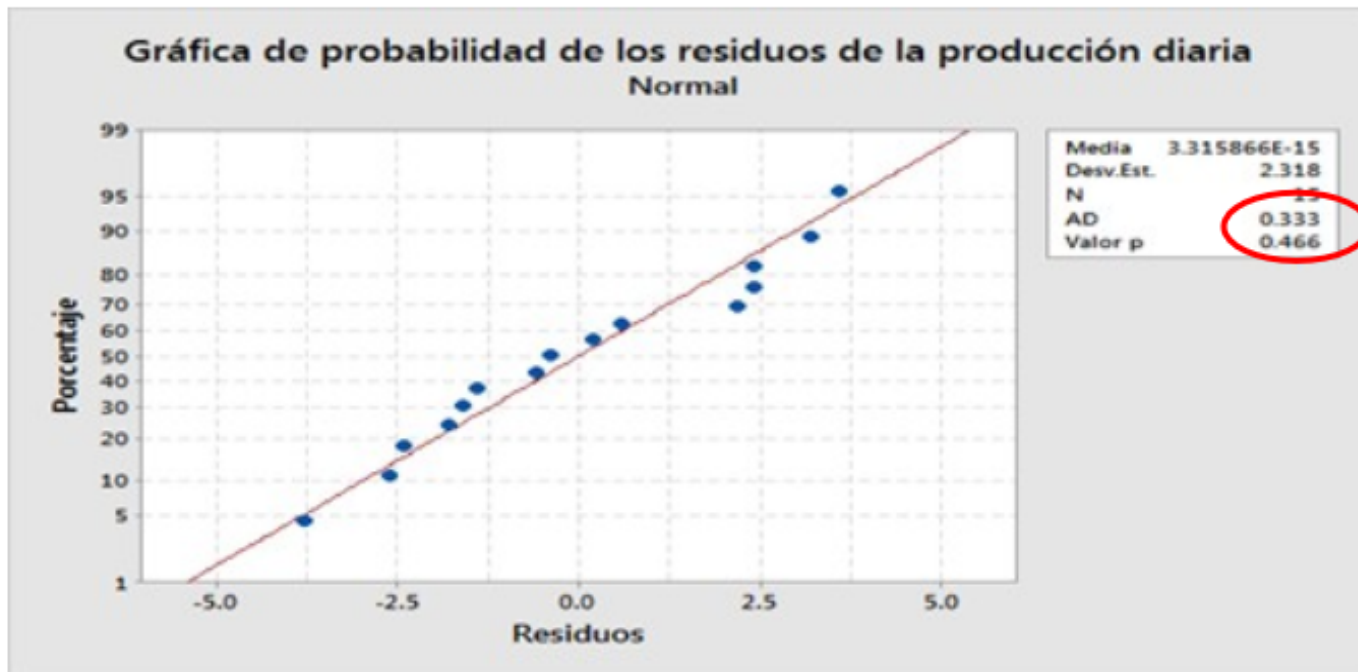
- 1) Verifique los supuestos.
- 2) ¿Al menos un método de entrenamiento para empleados nuevos es diferente a los demás?
- 3) ¿Qué método(s) de entrenamiento para empleados nuevos es más efectivo?

## 1) Verifique los supuestos

### NORMALIDAD DE LOS ERRORES

$H_0$ : Los errores se distribuyen normalmente

$H_1$ : Los errores no se distribuyen normalmente



**> 0.05**

AD	0.333
----	-------

P-valor	0.466
---------	-------

Decisión	No se rechaza $H_0$
----------	---------------------

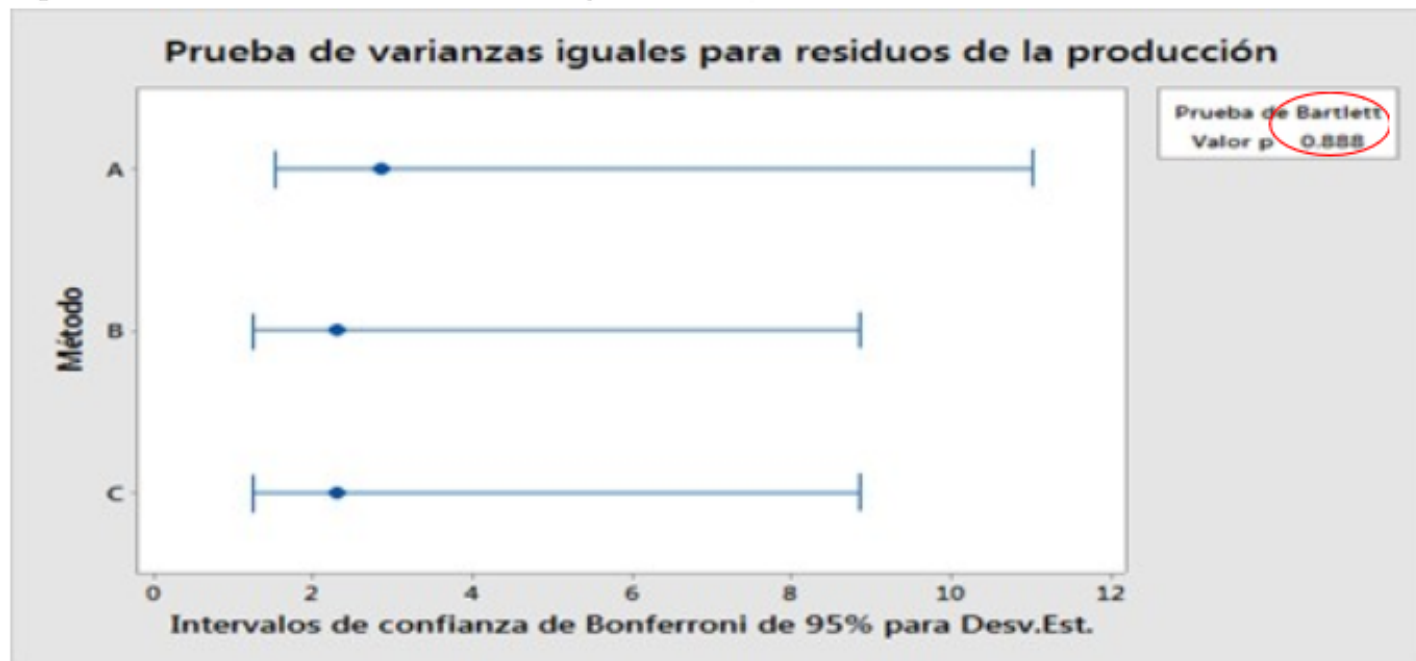
**Con un Nivel de Significación del 5%, los errores se distribuyen normalmente.**

## 1) Verifique los supuestos

### HOMOGENEIDAD DE VARIANCIAS

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

$H_1$  : Al menos un  $\sigma_i^2$  es diferente



> 0.05

Bartlett	0.24
<b>P-valor</b>	<b>0.888</b>
Decisión	No se rechaza $H_0$

Con un nivel de significación del 5%, existe homogeneidad de variancias.

## Verificación de Varianza

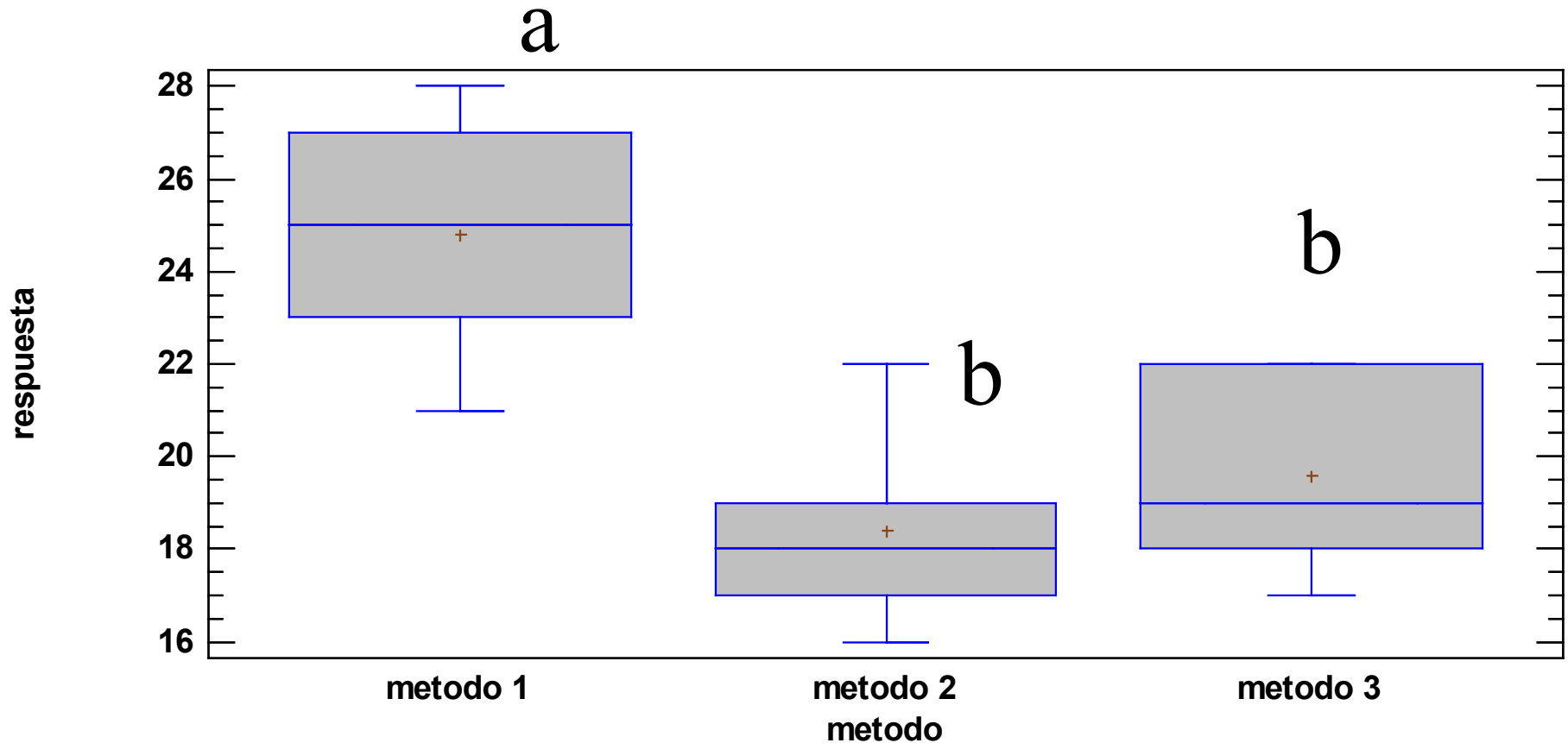
	<i>Prueba</i>	<i>Valor-P</i>
de Bartlett	1.02231	0.887688

<i>Comparación</i>	<i>Sigmal</i>	<i>Sigma2</i>	<i>F-Ratio</i>	<i>P-Valor</i>
metodo 1 / metodo 2	2.86356	2.30217	1.54717	0.6827
metodo 1 / metodo 3	2.86356	2.30217	1.54717	0.6827
metodo 2 / metodo 3	2.30217	2.30217	1.0	1.0000

El estadístico mostrado en esta tabla evalúa la hipótesis de que la desviación estándar de respuesta dentro de cada uno de los 3 niveles de metodo es la misma. De particular interés es el valor-P. Puesto que el valor-P es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre las desviaciones estándar, con un nivel del 95.0% de confianza.

La tabla también muestra una comparación de las desviaciones típicas para cada par de muestras. P-valores por debajo de 0.05, de los cuales hay 0, indican una diferencia estadísticamente significativa entre las dos sigmas al 5% de nivel de significación.

## Gráfico Caja y Bigotes



- También conocido como diagrama de caja y bigote, *box plot*, *box-plot* o *boxplot*.
- Es un método estandarizado para representar gráficamente una serie de datos numéricos a través de sus cuartiles.
- De esta manera, el diagrama de caja muestra a simple vista la mediana y los cuartiles de los datos, pudiendo también representar los **valores atípicos** de estos.

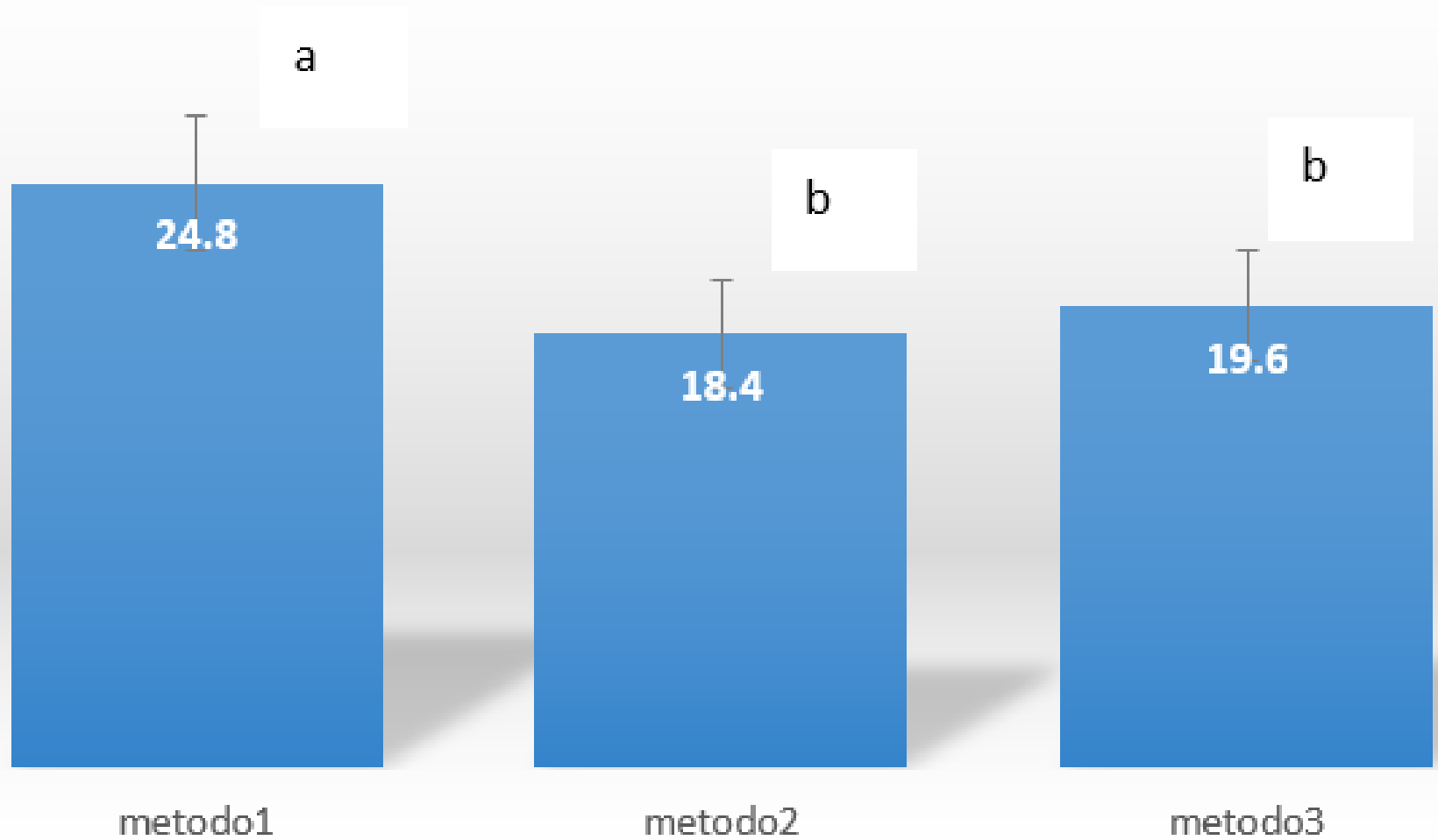




- Proporcionan una visión general de la simetría de la distribución de los datos; si la mediana no está en el centro del rectángulo, la distribución no es simétrica.
- Son útiles para ver la presencia de valores atípicos también llamados *outliers*.
- Pertenece a las herramientas de la estadística descriptiva. Permite ver como es la dispersión de los puntos con la mediana, los percentiles 25 y 75 y los valores máximos y mínimos.
- Ponen en una sola dimensión los datos de un histograma, facilitando así el análisis de la información al detectar que el 50% de la población está en los límites de la caja.



# Resultados



## 2) ¿Al menos un método de entrenamiento para empleados nuevos es diferente a los demás?

### HIPÓTESIS:

**H<sub>0</sub>:** No hay diferencia en la producción promedio diaria que obtiene un empleado nuevo cuando se le entrena con cualquiera de los tres métodos.

**H<sub>1</sub>:** Al menos un método de entrenamiento es diferente en la producción promedio diaria que obtiene un empleado nuevo.

Nivel de significación= 0.05

#### Modelo lineal general: Producción vs. Métodos

##### Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Métodos	2	115.73	57.867	9.23	0.004
Error	12	75.20	6.267		
Total	14	190.93			

**Decisión:** Se rechaza la Hipótesis Nula.

**Conclusión:** Con un nivel de significancia del 5%, se puede afirmar al menos un método de entrenamiento es diferente en la producción promedio diaria que obtiene un empleado nuevo.

# Comparaciones Múltiples

Agrupar información utilizando el método de Tukey.

$$H_0: \mu_i = \mu_j$$

$$H_a: \mu_i \neq \mu_j$$

Para todo  $i$  diferente de  $j$

Agrupar información utilizando el método de Tukey

	N	Media	Agrupación
M1	5	24.800	A
M3	5	19.600	B
M2	5	18.400	B

Las medias **que no comparten una letra son significativamente diferentes.**

Se concluye que no hay diferencias significativas entre los métodos 3 y 2, pero si hay diferencias significativas de ambos con respecto al método 1.

Por lo tanto, **el método 1:** Asignar un empleado nuevo con un trabajador experimentado **es más efectivo**, la producción promedio por día es mayor para este método con respecto a los métodos 2 y 3.



## Pruebas de Múltiple Rangos para respuesta por metodo

Método: 95.0 porcentaje Tukey HSD

metodo	Casos	Media	Grupos Homogéneos
metodo 2	5	18.4	X
metodo 3	5	19.6	X
metodo 1	5	24.8	X

Contraste	Sig.	Diferencia	+/- Límites
metodo 1 - metodo 2	*	6.4	4.23549
metodo 1 - metodo 3	*	5.2	4.23549
metodo 2 - metodo 3		-1.2	4.23549

\* indica una diferencia significativa.

### El StatAdvisor

Esta tabla aplica un procedimiento de comparación multiple para determinar cuáles medias son significativamente diferentes de otras. La mitad inferior de la salida muestra las diferencias estimadas entre cada par de medias. El asterisco que se encuentra al lado de los 2 pares indica que estos pares muestran diferencias estadísticamente significativas con un nivel del 95.0% de confianza. En la parte superior de la página, se han identificado 2 grupos homogéneos según la alineación de las X's en columnas. No existen diferencias estadísticamente significativas entre aquellos niveles que compartan una misma columna de X's. El método empleado actualmente para discriminar entre las medias es el procedimiento de diferencia honestamente significativa (HSD) de Tukey. Con este método hay un riesgo del 5.0% al decir que uno o más pares son significativamente diferentes, cuando la diferencia real es igual a 0.





## Pruebas de Múltiple Rangos para respuesta por metodo

Método: 95.0 porcentaje Duncan

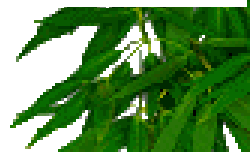
<i>metodo</i>	<i>Casos</i>	<i>Media</i>	<i>Grupos Homogéneos</i>
metodo 2	5	18.4	X
metodo 3	5	19.6	X
metodo 1	5	24.8	X

<i>Contraste</i>	<i>Sig.</i>	<i>Diferencia</i>
metodo 1 - metodo 2	*	6.4
metodo 1 - metodo 3	*	5.2
metodo 2 - metodo 3		-1.2

\* indica una diferencia significativa.

### El StatAdvisor

Esta tabla aplica un procedimiento de comparación múltiple para determinar cuáles medias son significativamente diferentes de otras. La mitad inferior de la salida muestra las diferencias estimadas entre cada par de medias. El asterisco que se encuentra al lado de los 2 pares indica que estos pares muestran diferencias estadísticamente significativas con un nivel del 95.0% de confianza. En la parte superior de la página, se han identificado 2 grupos homogéneos según la alineación de las X's en columnas. No existen diferencias estadísticamente significativas entre aquellos niveles que compartan una misma columna de X's. El método empleado actualmente para discriminar entre las medias es el procedimiento de comparación múltiple de Duncan. Con este método hay un riesgo del 5.0% al decir que uno o más pares son significativamente diferentes, cuando la diferencia real es igual a 0.



## Pruebas *a posteriori*

- Diferencia mínima significativa irrestricta LSD
- Diferencia mínima significativa restringida de Fisher
- Prueba de rangos múltiples de Duncan
- Prueba de Student-Newman-Keuls (S.N.K)
- Tukey
- Scheffé
- Intervalo de Bonferroni o corrección de Bonferroni



## Pruebas *a priori* o planeadas

- Dunnett (comparación con un control)



- **DMS.** Diferencia Mínima Significativa basada en la distribución  $t$  de Student. Este método, inicialmente propuesto por Fisher (1935), no ejerce ningún control sobre la *tasa de error*. Es decir, cada comparación se lleva a cabo utilizando el nivel de significación establecido (generalmente 0,05), por lo que la *tasa de error* para el conjunto de comparaciones puede llegar a  $1-(1-\alpha)^k$ , siendo  $\alpha$  el nivel de significación y  $k$  el número de comparaciones llevadas a cabo. (suele encontrarse en la literatura estadística con su acrónimo inglés: *LSD = Least Significant Difference*).
- **Bonferroni.** Método basado en la distribución  $t$  de Student y en la desigualdad de Bonferroni (también conocido como método de Dunn —su promotor en 1961— o de Dunn-Bonferroni). Controla la tasa de error dividiendo el nivel de significación ( $\alpha$ ) entre el número de comparaciones ( $k$ ) llevadas a cabo. Cada comparación se evalúa utilizando un nivel de significación  $\alpha_c = \alpha/k$ .
- **Sidak** (1967). Al igual que el procedimiento de Bonferroni, se basa en la distribución  $t$  de Student, pero controla la tasa de error evaluando cada comparación con un nivel de significación  $\alpha_c = 1-(1-\alpha)^{1/k}$ . Esta solución es algo menos conservadora que la de Bonferroni (es decir, rechaza la hipótesis de igualdad de medias en más ocasiones que el método de Bonferroni).
- **Scheffé** (1953, 1959). Este método, basado en la distribución  $F$ , permite controlar la tasa de error para el conjunto total de comparaciones que es posible diseñar con  $J$  medias (una con otra, una con todas las demás, dos con dos, etc.). Utilizado para efectuar sólo comparaciones por pares, es un procedimiento muy conservador: tiende a considerar significativas menos diferencias de las que debería.



- 
- **S-N-K.** Student-Neuwman-Keuls (Newman, 1939; Keuls, 1952). Método basado en la distribución del rango estudentizado. Al igual que los métodos R-E-G-W  $F$  y  $Q$ , éste también se basa en una ordenación de las medias por su tamaño. Pero a diferencia de ellos, aquí el nivel de significación para cada conjunto de medias separadas  $r$  pasos es siempre  $\alpha$ . Cuantos más pasos existen entre dos medias, mayor es la *diferencia mínima* necesaria para considerar que esas medias difieren significativamente.
  - **Tukey** (1953). Diferencia honestamente significativa de Tukey. Equivale a utilizar el método de Student-Newman-Keuls con  $r = J = n^{\circ} \text{ de medias}$ . Por tanto, todas las comparaciones son referidas a una misma *diferencia mínima*. Es uno de los métodos de mayor aceptación.
  - **Tukey-b** (1953). Este método consiste en considerar como *diferencia mínima* el valor medio entre la *diferencia honestamente significativa* de Tukey y la *diferencia mínima* obtenida con el método de Student-Newman-Keuls para el caso de  $r = 2$ .
  - **Duncan** (1955). Prueba del rango múltiple de Duncan. Método de comparación por pasos basado en la distribución del rango estudentizado. Controla la tasa de error utilizando, para el conjunto de medias separadas  $r$  pasos, un nivel de significación  $\alpha_c = 1 - (1 - \alpha)^{r-1}$ . Cuantos más pasos existen entre dos medias, mayor es la *diferencia mínima* con la que vamos a considerar que esas medias difieren significativamente.
- 



- **R-E-G-W  $F$** . Método de Ryan (1960), Einot - Gabriel (1975) y Welsch (1977) basado en la distribución  $F$ . Se trata de un método por pasos. Tras ordenar de forma ascendente las  $J$  medias por su tamaño, se efectúan todas las comparaciones posibles entre pares de medias teniendo en cuenta el número de escalones ( $r$ ) que las separan: con  $J$  medias, la media más pequeña y la más grande están separadas  $r = J$  escalones; la media más pequeña y la segunda más grande están separadas  $r = J - 1$  escalones; la media más pequeña y la tercera más grande están separadas  $r = J - 2$  escalones; etc. Dos medias adyacentes tras la ordenación están separadas 2 escalones.
- **GT2 de Hochberg** (1974). Es un procedimiento muy similar a la Diferencia honestamente significativa de Tukey, pero se basa en la distribución del módulo máximo estudentizado. El método de Tukey suele ser más potente.
- **Gabriel** (1969). También se basa en la distribución del módulo máximo estudentizado. Con grupos del mismo tamaño, este método es más potente que el de Hochberg, pero con tamaños muy desiguales ocurre lo contrario.
- **Waller-Duncan** (1969). Utiliza la distribución  $t$  de Student y una aproximación bayesiana. Si los tamaños muestrales son distintos, utiliza la media armónica.
- **Dunnett** (1955). Sirve para comparar cada grupo con un grupo control. Por tanto, controla la tasa de error para  $k - 1$  comparaciones. Por defecto, se considera que la última categoría del factor es la que define el grupo control, pero puede seleccionarse la primera categoría. Permite efectuar tanto contrastes bilaterales como unilaterales.



# **Comparaciones o pruebas de rango múltiples**



## Diferencia mínima significativa (LSD)

$$|\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}| > t_{\alpha/2, N-k} \sqrt{CM_E \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} = LSD$$

- Es la diferencia mínima que debe haber entre dos medias muestrales para considerar que dos tratamientos son diferentes.

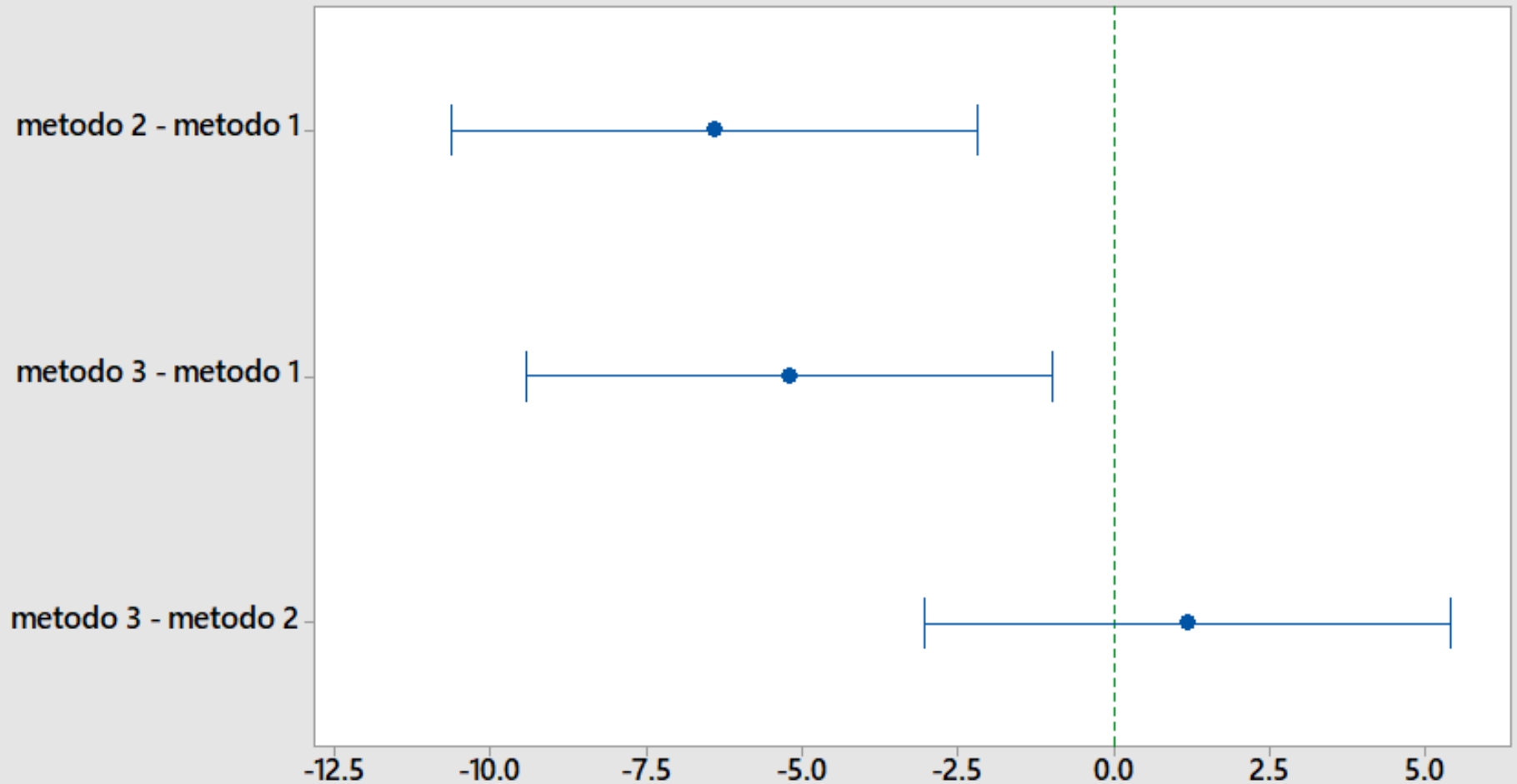
## Método de Tukey

- Un método más conservador para comparar pares de medias de tratamientos es el método de Tukey, el cual consiste en comparar las diferencias entre medias muestrales con el valor crítico dado por:
- Se declaran significativamente diferentes los pares de medias cuya diferencia muestral en valor absoluto sea mayor que  $T_\alpha$ .
- A diferencia de los métodos LSD y Duncan, el método de Tukey trabaja con un error  $\alpha$  muy cercano al declarado por el experimentador.
- Esto es congruente con el hecho de que la prueba de Tukey es menos potente que la prueba LSD, por lo que las pequeñas diferencias no son detectadas como significativas.

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



## Tukey Simultaneous 95% CIs Differences of Means for respuesta



*If an interval does not contain zero, the corresponding means are significantly different.*



## Método de Duncan

- En este método para la comparación de medias, si las  $k$  muestras son de igual tamaño, los  $k$  promedios se acomodan en orden ascendente y el error estándar.
- Si alguno o todos los tratamientos tienen tamaños diferentes, se reemplaza  $n$  por la media armónica de las  $\{n_i\}$ .
- En las comparaciones donde la diferencia observada es mayor que el rango respectivo, se concluye que esas medias son significativamente diferentes.
- Si dos medias caen entre otras dos que no son muy diferentes, entonces esas dos medias poblacionales también se consideran estadísticamente iguales.
- En general, las pruebas de Duncan y LSD tienen un desempeño similar.

$$S_{\bar{Y}_{i\cdot}} = \sqrt{CM_E / n}$$



# Comparación de tratamientos con un control (método de Dunnet)

- Una vez que se rechaza  $H_0$  con el ANOVA, en ocasiones uno de los  $k$  tratamientos a comparar es el llamado tratamiento control y el interés fundamental es comparar los  $k - 1$  tratamientos restantes con dicho control.
- En muchos casos el tratamiento control se refiere a un tratamiento estándar de referencia o también a la ausencia de tratamiento

$$|\bar{Y}_{i\cdot} - \bar{Y}_{k\cdot}| > D_{\alpha}(k-1, l) \sqrt{CM_E \left( \frac{1}{n_i} + \frac{1}{n_k} \right)}$$



# Método de Sheffé

- Este método está diseñado para probar todos los contrastes de medias que pudieran interesar al experimentador, sin el inconveniente de inflar por ello el error tipo I (detección de diferencias que no existen). Supongamos que interesa contrastar las Hipótesis

$$H_0 : 2\mu_A = \mu_B + \mu_C \quad (3.14)$$

$$H_A : 2\mu_A \neq \mu_B + \mu_C$$

donde la hipótesis nula se puede escribir alternativamente como  $H_0 : 2\mu_A - \mu_B - \mu_C = 0$ , lo cual implica que la hipótesis está definida por el contraste  $C_0 = 2\mu_A - \mu_B - \mu_C$ . De manera que el contraste estimado está dado por





# **Verificación de los supuestos del modelo**



# Supuestos que tiene el test ANOVA

- **La variable dependiente o respuesta debe ser continua.** Por ejemplo, el tiempo de revisión (medido en horas), inteligencia (medida mediante la puntuación de CI), desempeño del examen (medido de 0 a 100), peso (medido en kg), etc.
- **La variable independiente o explicativa debe estar formada por tres o más grupos categóricos e independientes.** Por ejemplo, la etnicidad (caucásico, afroamericano e hispano), el nivel de actividad física (sedentario, bajo, moderado y alto), la profesión (doctor, enfermera, dentista, terapeuta), etc.
- **La variable dependiente se distribuye normalmente en cada grupo** que se compara en el ANOVA de una vía (técnicamente, son los residuos los que necesitan ser distribuidos normalmente, pero los resultados serán los mismos). Puedes probar la normalidad usando la prueba de normalidad de Shapiro-Wilk.



# Supuestos tiene el test ANOVA

- **Hay homogeneidad de varianzas.** Esto significa que las varianzas de la respuesta en cada grupo son iguales. Puedes probar esta suposición usando la prueba de Levene para la homogeneidad de las varianzas.
- **Las observaciones son independientes.** Esto es principalmente un tema del diseño del estudio y, como tal, necesitará determinar si cree que es posible que sus observaciones no sean independientes en función del diseño del estudio (por ejemplo, si los valores han sido tomados diariamente las muestras más cercanas en el tiempo serán más similares entre sí respecto al resto de muestras -correlación temporal-).
- **No debemos tener valores atípicos (*outliers*) influyentes.** Los valores atípicos son simplemente valores dentro de sus datos que no siguen el patrón habitual. El problema con los valores atípicos es que pueden afectar el resultado del ANOVA, reduciendo la validez de sus resultados.



*Normalidad*

- Representaciones gráficas: histograma con curva normal, gráfico P-P o de proporciones, gráfico Q-Q o de cuantiles
- Valores de asimetría y apuntamiento
- Contraste de Shapiro y Wilk
- Contraste de  $\chi^2$
- Contraste de Kolmogorov-Smirnov con corrección de Lilliefors

*Homoceasticidad*

- Contraste de Bartlett
- Contraste de Lehman
- Contraste de Hartley
- Contraste de Cochran
- Contraste de Levene
- Prueba con los logaritmos de las cuasivarianzas (para ANOVA)

*Independencia*

- Coeficiente de correlación serial de separación 1 (prueba de autocorrelación)
- Contraste de rachas

- La validez de los resultados obtenidos en cualquier análisis de varianza queda supeditado a que los supuestos del modelo se cumplan. Estos supuestos son: *normalidad, varianza constante (igual varianza de los tratamientos) e independencia*.
- Esto es, la respuesta ( $Y$ ) se debe distribuir de manera normal, con la misma varianza en cada tratamiento y las mediciones deben ser independientes.
- Estos supuestos sobre  $Y$  se traducen en supuestos sobre el término error ( $\varepsilon$ ) en el modelo.



- Es una práctica común utilizar la muestra de *residuos* para comprobar los supuestos del modelo, ya que si los supuestos se cumplen, los **residuos o residuales se pueden ver como una muestra aleatoria de una distribución normal con media cero y varianza constante.**
- Los residuos,  $e_{ij}$ , se definen como la diferencia entre la **respuesta observada** ( $Y_{ij}$ ) y la **respuesta predicha** por el modelo ( $\hat{Y}_{ij}$ ), lo cual permite hacer un diagnóstico más directo de la calidad del modelo, ya que su magnitud señala qué tan bien describe a los datos el modelo.



$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (3.15)$$

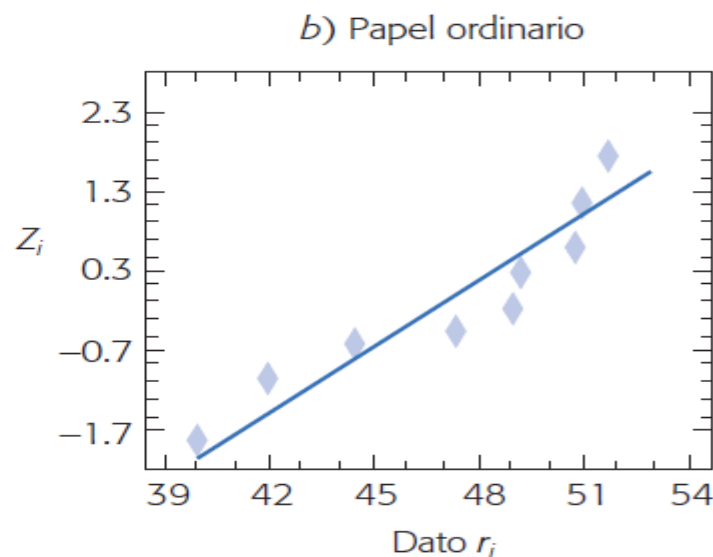
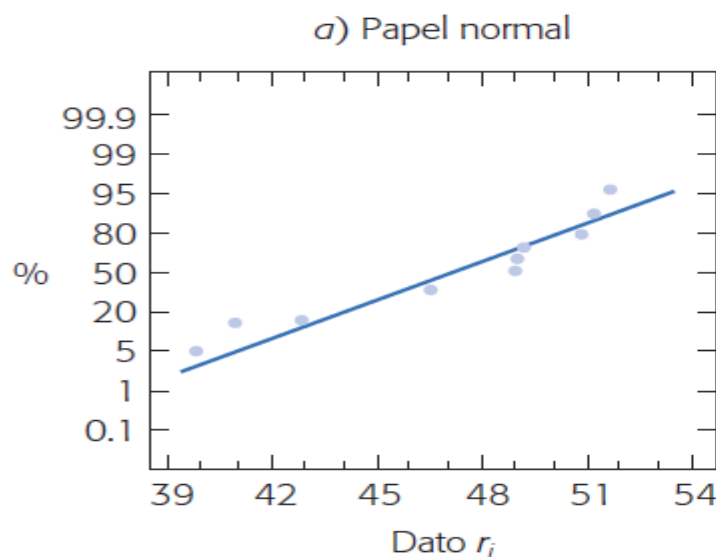
donde  $Y_{ij}$  ( $i = 1, 2, \dots, k; j = 1, 2, \dots, n$ ) es el  $j$ -ésimo dato en el tratamiento  $i$ ;  $\mu$  es la media global,  $\tau_i$  es el efecto del tratamiento  $i$  y  $\varepsilon_{ij}$  representa al error asociado con la observación  $Y_{ij}$ .

Los supuestos del modelo lineal (3.15), en términos de los residuos, son:

1. *Los  $e_{ij}$  siguen una distribución normal con media cero.*
2. *Los  $e_{ij}$  son independientes entre sí.*
3. *Los residuos de cada tratamiento tienen la misma varianza  $\sigma^2$ .*

## Prueba de Shapiro-Wilks para normalidad

- Consideremos una muestra aleatoria de datos  $x_1, x_2, \dots, x_n$  que proceden de cierta distribución desconocida denotada por  $F(x)$ .
- Se quiere verificar si dichos datos fueron generados por un proceso normal, mediante las hipótesis estadísticas:
  - $H_0$  : Los datos proceden de una distribución normal ( $F(x)$  es normal).
  - $H_A$  : Los datos no proceden de una distribución normal ( $F(x)$  no es normal).





## ***Prueba de Bartlett para homogeneidad de varianzas***

Supongamos que se tienen  $k$  poblaciones o tratamientos independientes, cada uno con distribución normal:

$$(N(\mu_i, \sigma_i^2), i = 1, 2, \dots, k)$$

donde las varianzas son desconocidas.

Se quiere probar la hipótesis de igualdad de varianzas dada por:

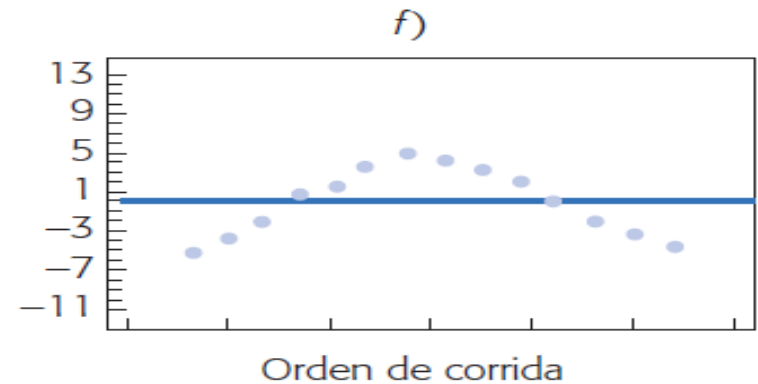
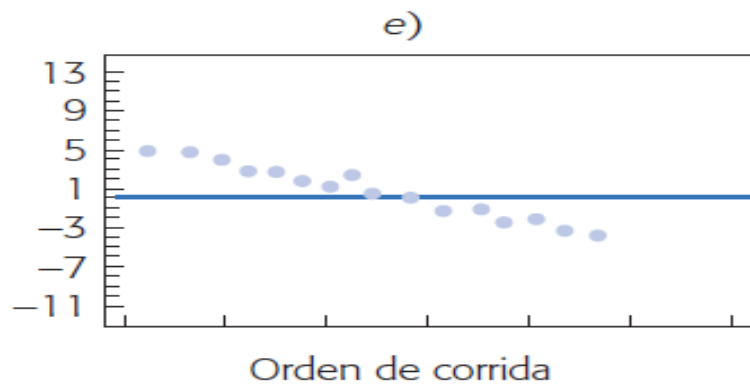
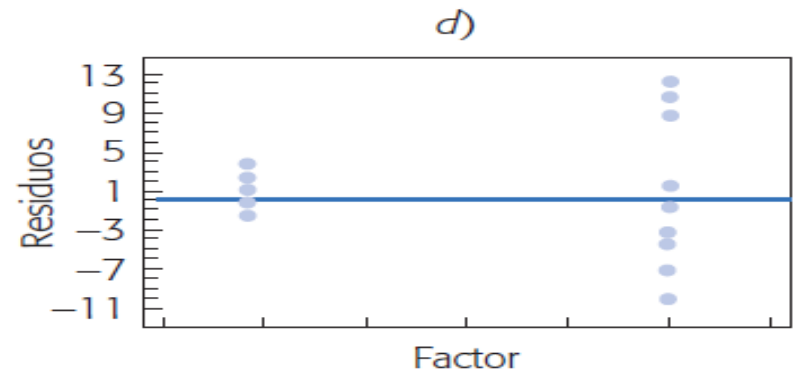
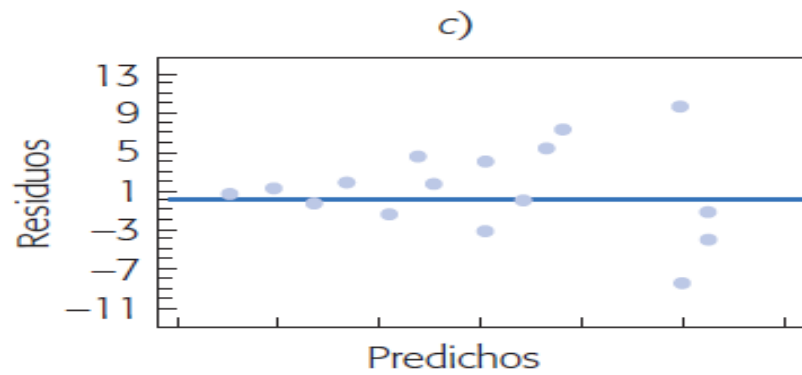
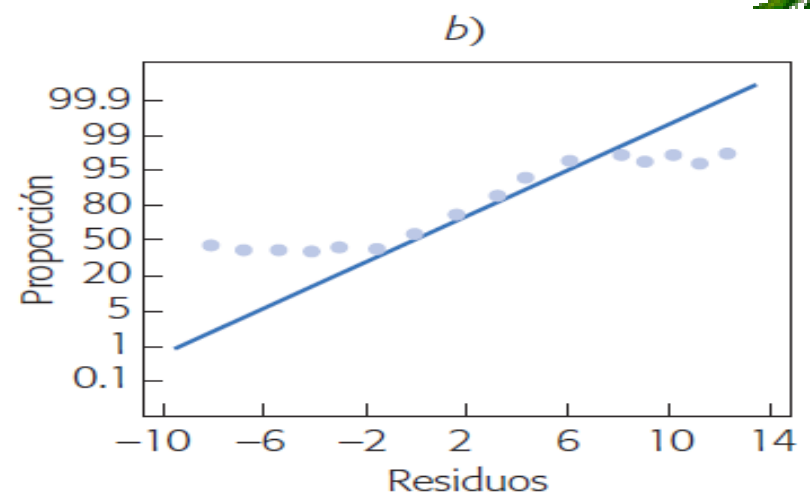
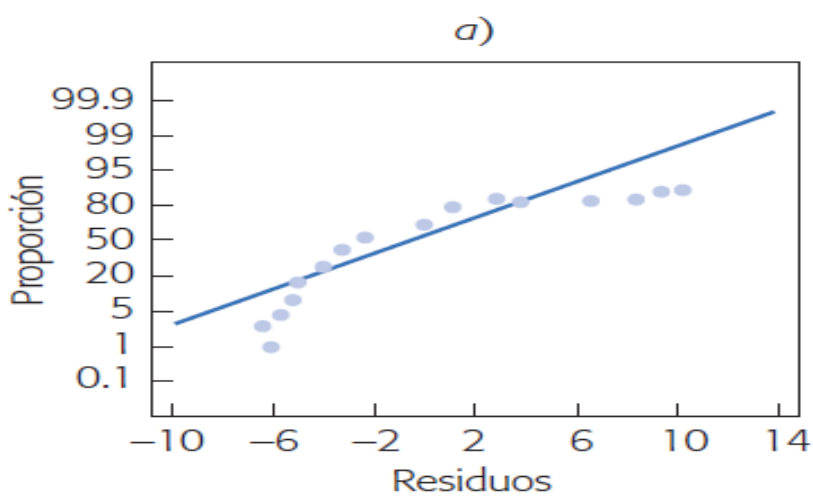
$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2$$

$$H_A : \sigma_i^2 \neq \sigma_j^2 \text{ para algún } i \neq j$$

## ***Independencia***

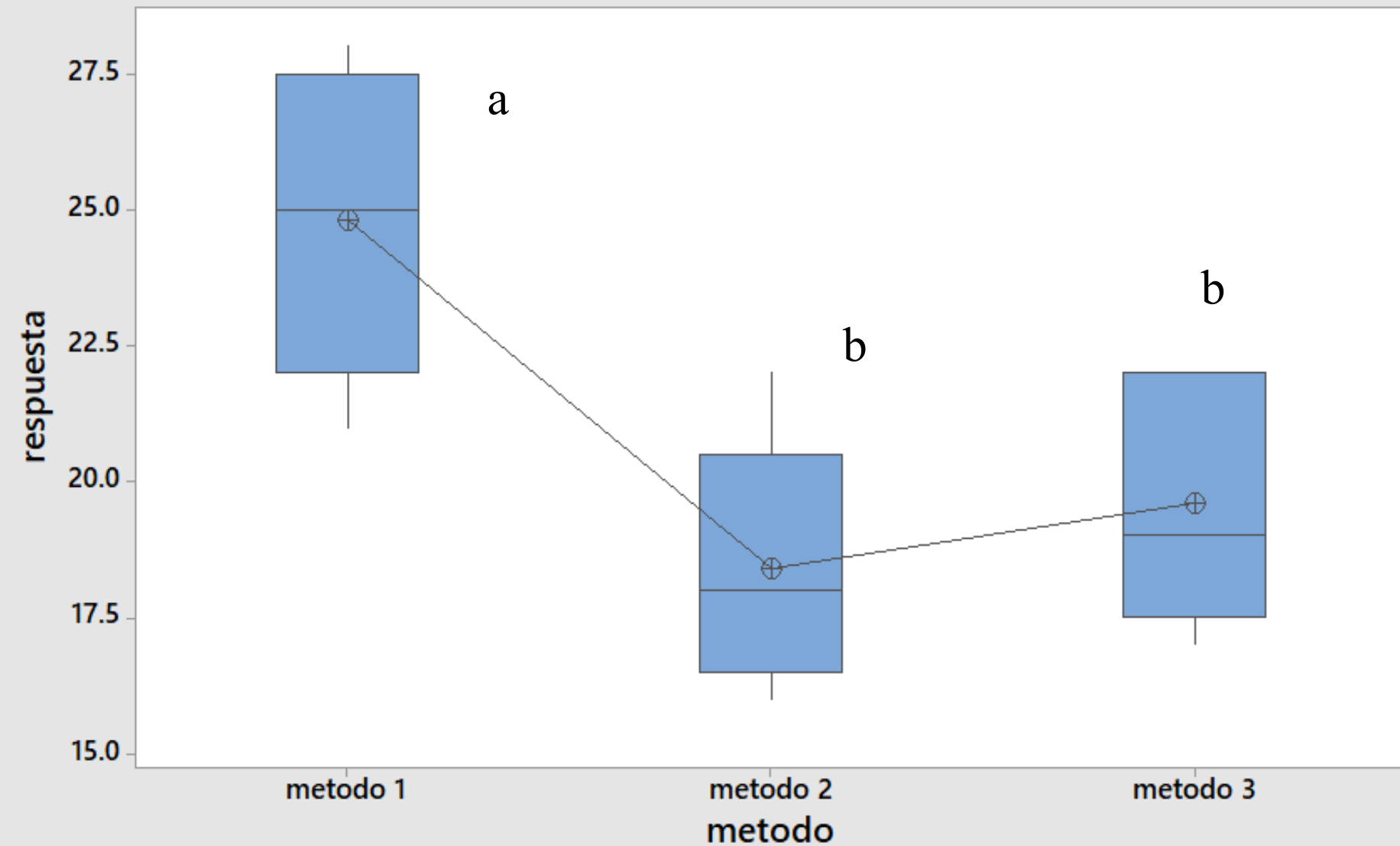
- La suposición de independencia en los residuos puede verificarse si se grafica el orden en que se colectó un dato contra el residuo correspondiente.
- De esta manera, si al graficar en el eje horizontal el tiempo (orden de corrida) y en el eje vertical los residuos, se detecta una tendencia o patrón no aleatorio claramente definido, esto es evidencia de que existe una correlación entre los errores y, por lo tanto, el supuesto de independencia no se cumple.
- ***Una prueba analítica para verificar la independencia entre residuos consecutivos es la prueba de Durbin-Watson***



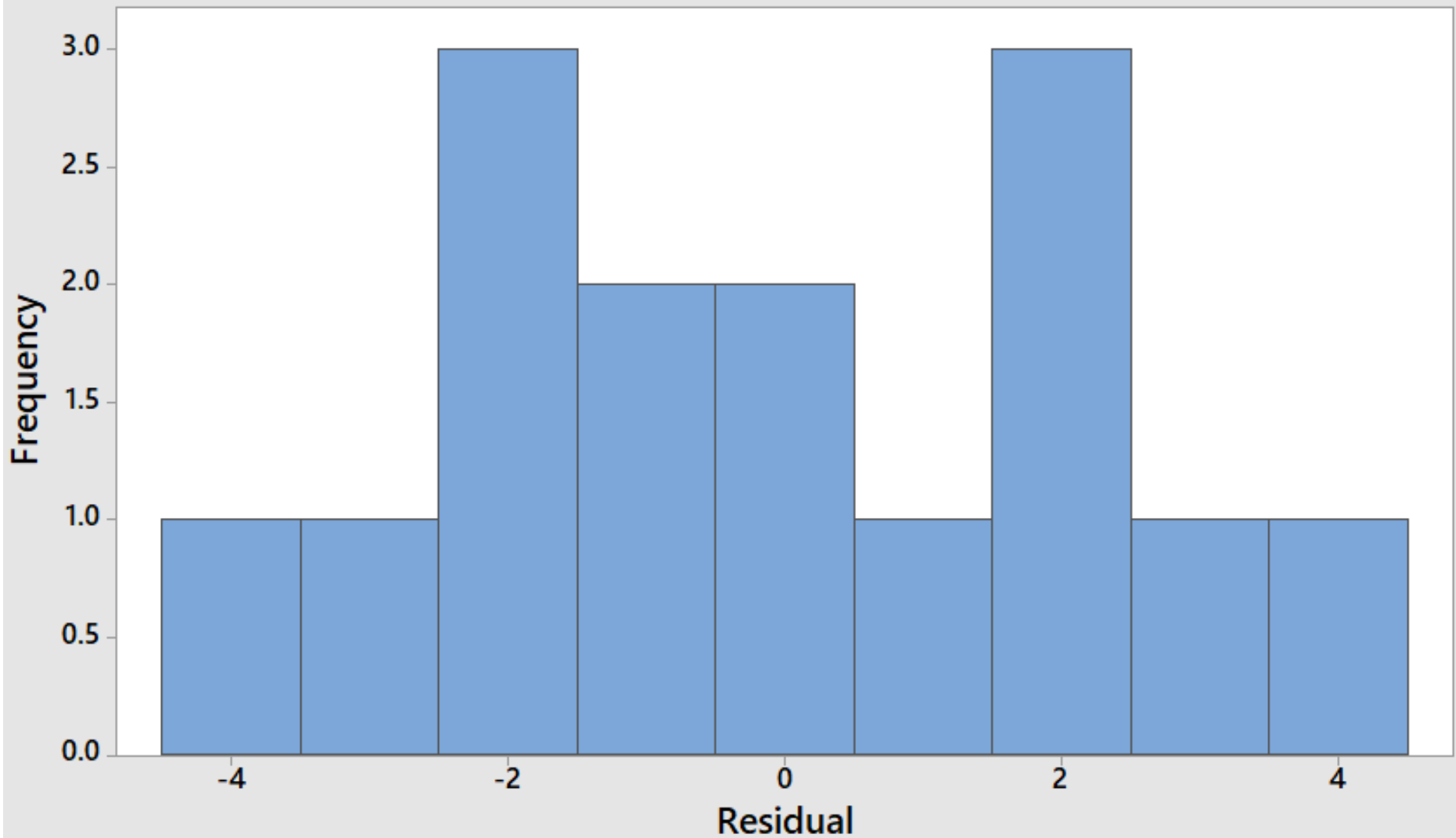


**Figura 3.6** Ejemplos de gráficas de residuos donde no se cumplen los supuestos para el ANOVA.

Boxplot of respuesta

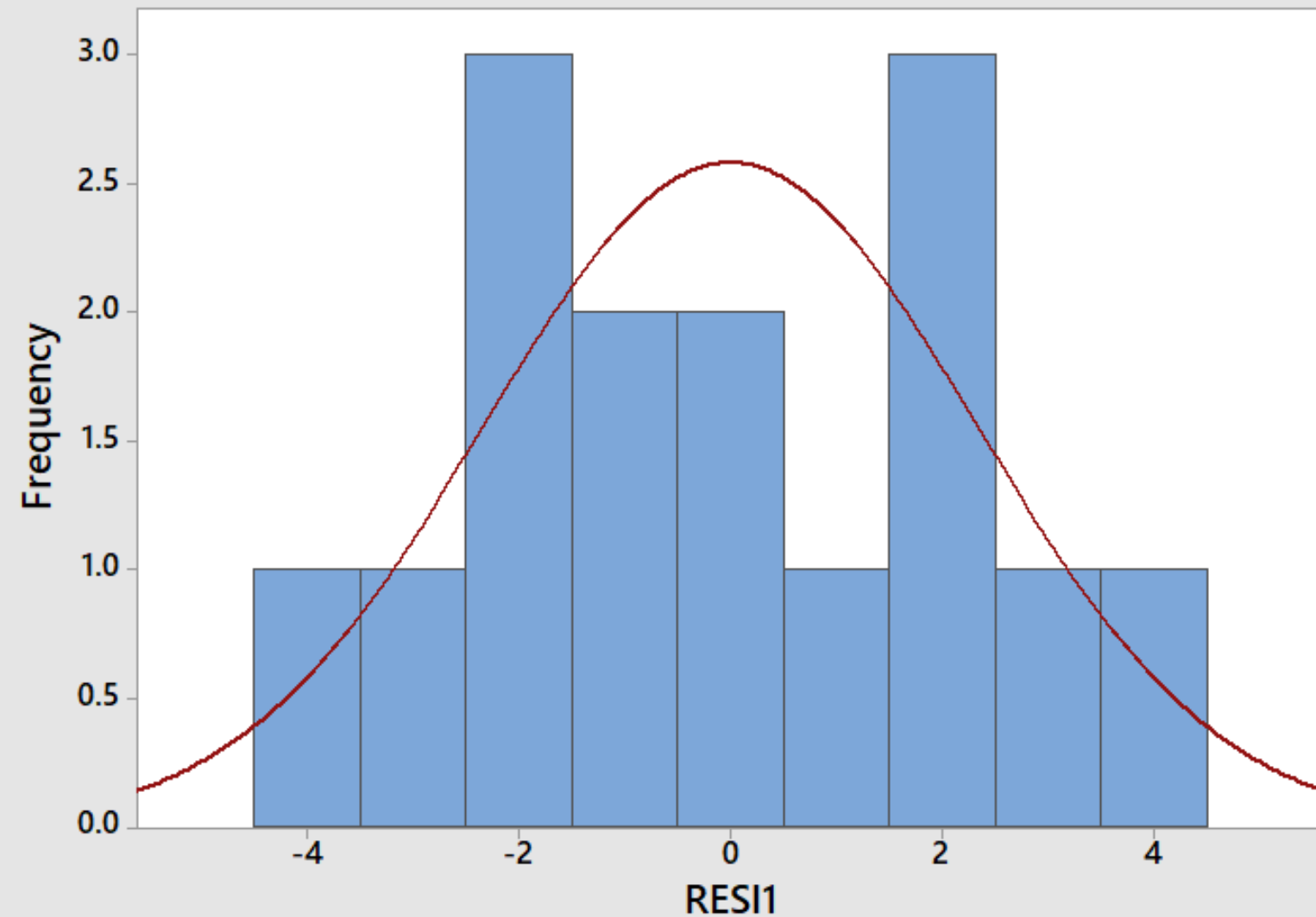


Histogram  
(response is respuesta)



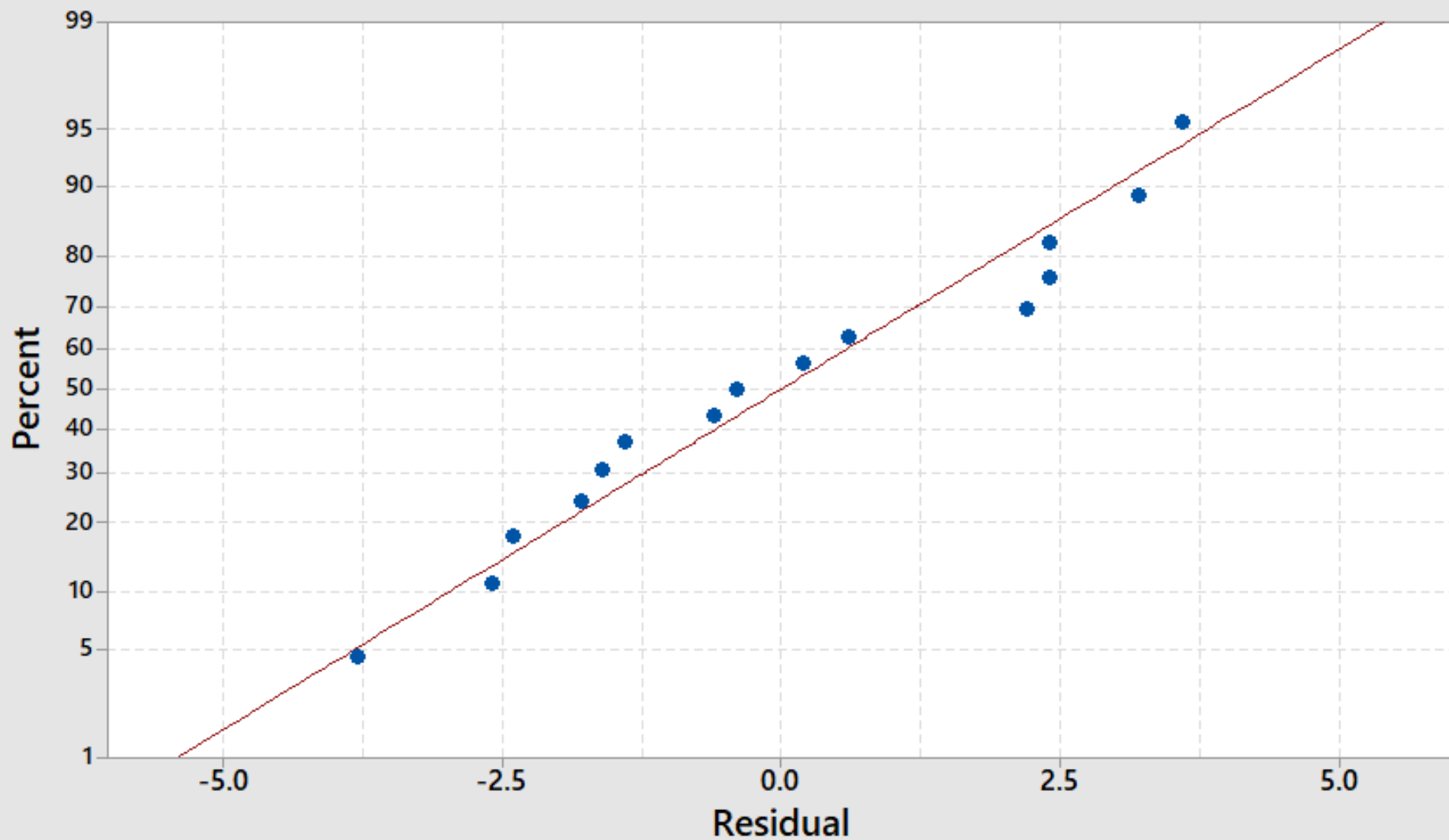
## Histogram of RESI1

Normal



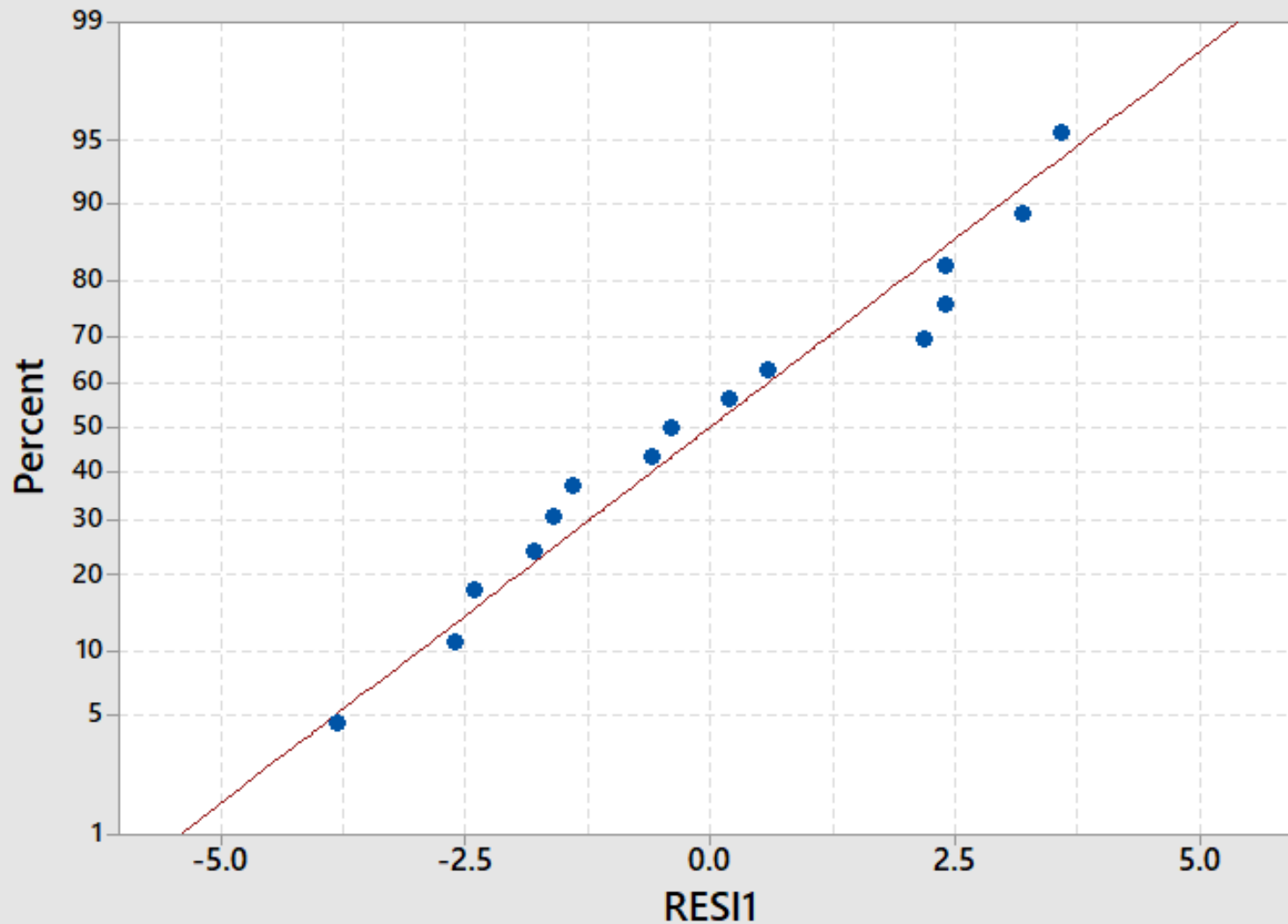
Mean	3.315866E-15
StDev	2.318
N	15

# Normal Probability Plot (response is respuesta)



## Probability Plot of RESI1

Normal

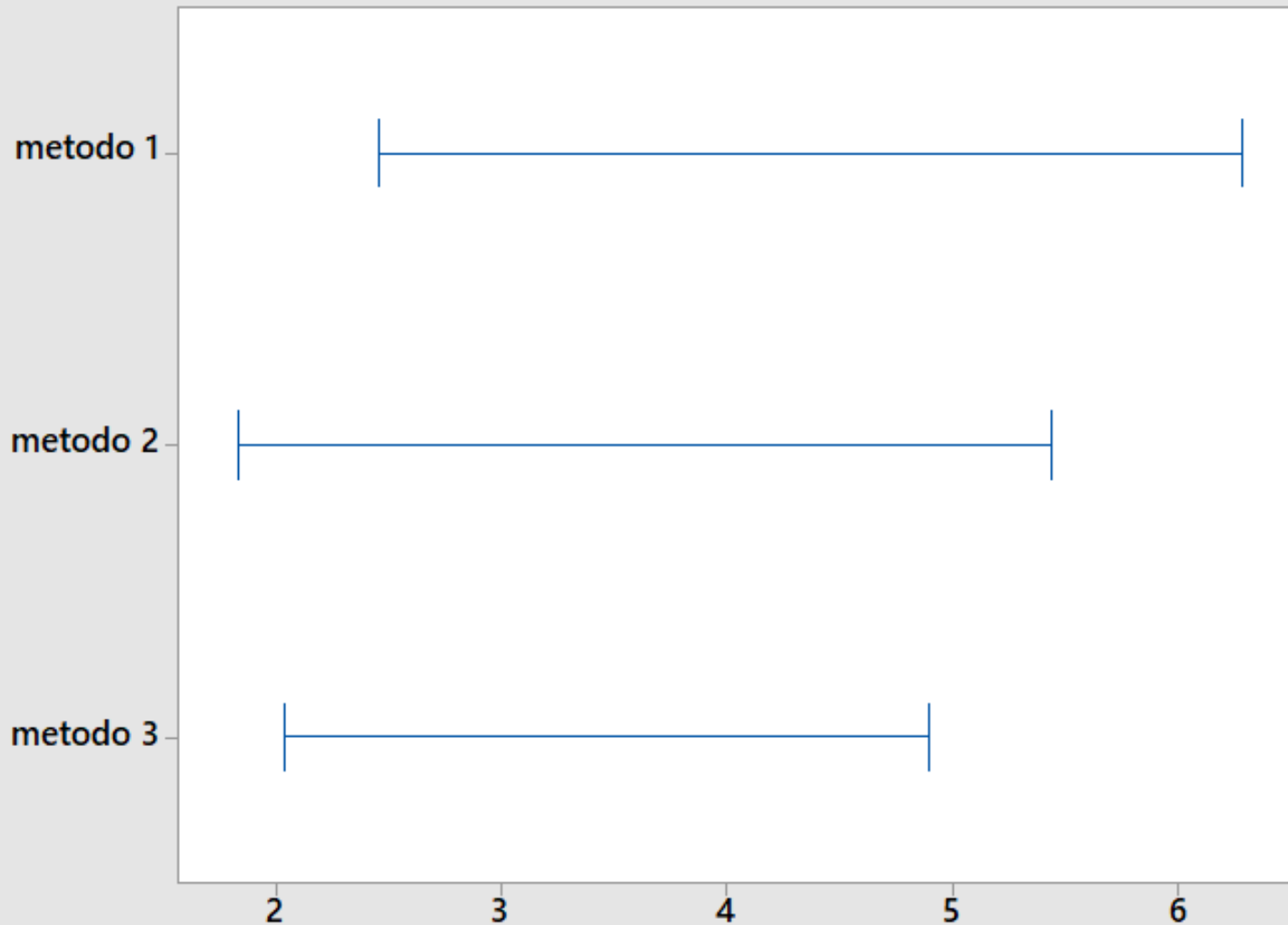


Mean	3.315866E-15
StDev	2.318
N	15
AD	0.333
P-Value	0.466



## Test for Equal Variances: RES11 vs metodo

Multiple comparison intervals for the standard deviation,  $\alpha = 0.05$



Multiple Comparisons

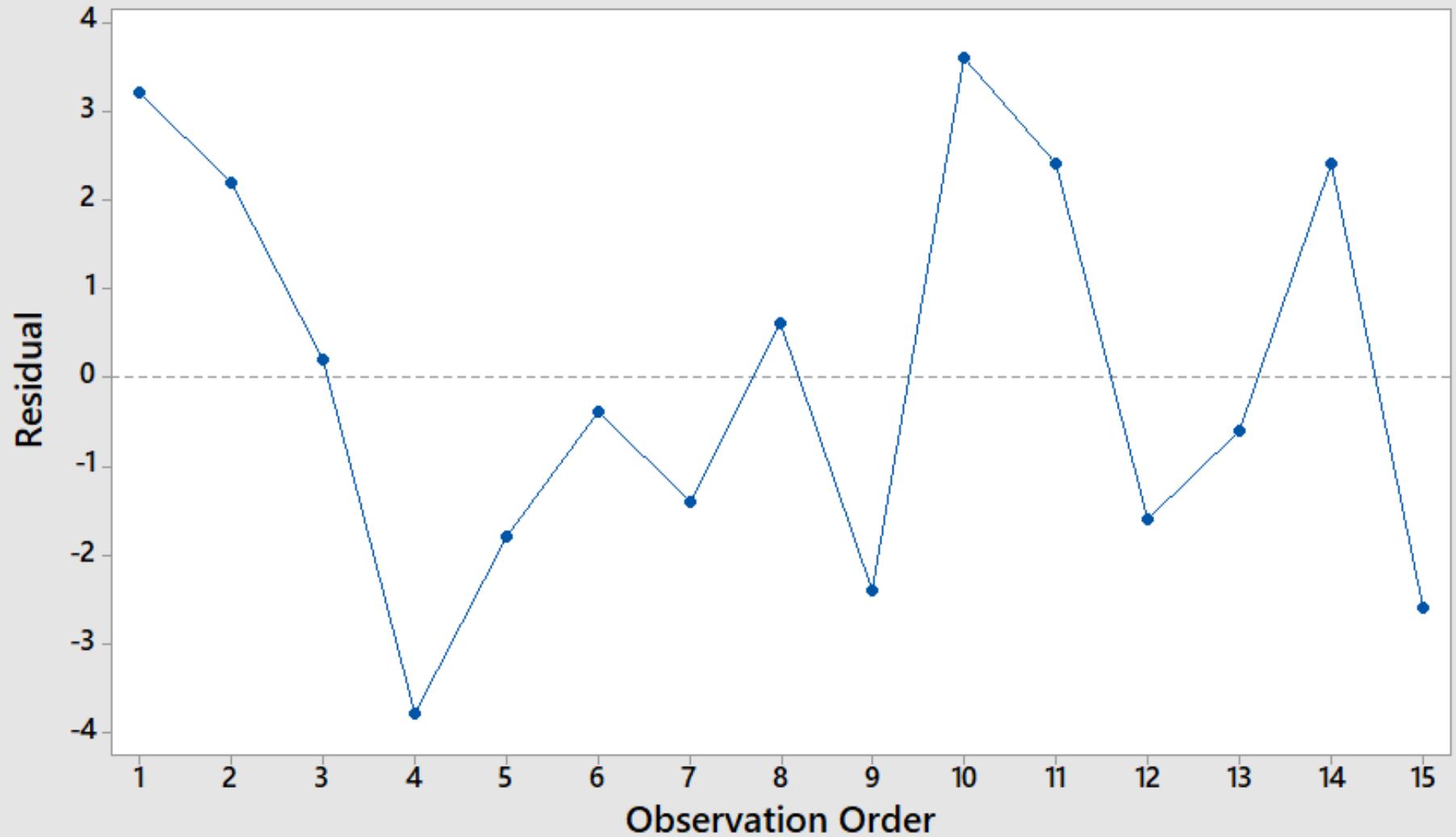
P-Value 0.839

Levene's Test

P-Value 0.801

*If intervals do not overlap, the corresponding stdevs are significantly different.*

## Versus Order (response is respuesta)



# **TRABAJO EN “R”**



DCA.R x

Source on Save

Run Source

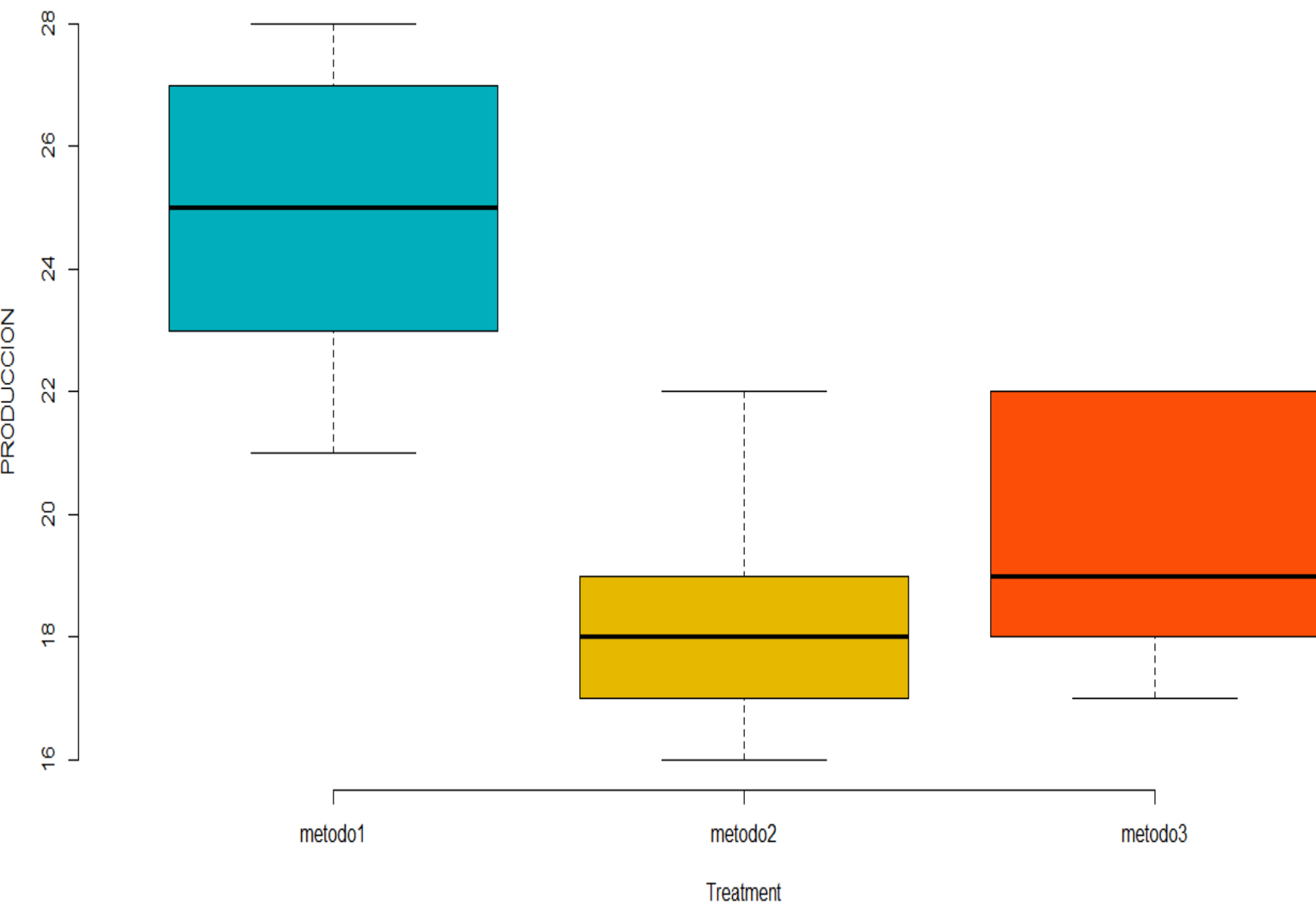
```
1 #IMPORTACION DE DATOS
2 #-----
3 #.CSV
4 # antes se fue a "session", y luego en "set working directory"
5 # y se carga "chose directory", despues en "open"
6
7 my_data=read.table("DCA.csv", header = T, sep=";", dec="." )
8 head(my_data)
9 #View(my_data)
10 str(my_data)
11 head(my_data)
12
13 my_data$METODOS <- ordered(my_data$METODOS,
14                             levels = c("metodo1", "metodo2", "metodo3"))
15
16 # show the levels
17 levels(my_data$METODOS)
18
19 # Box plots
20 # Box plot
21 boxplot(PRODUCCION ~ METODOS, data = my_data,
22          xlab = "Treatment", ylab = "PRODUCCION",
23          frame = FALSE, col = c("#00AFBB", "#E7B800", "#FC4E07"))
```

25:1 # (Untitled)

R Script

Console C:/Users/chris/Desktop/Cursos 2020-2/METODOS ESTADISTICOS 2020-2/PRACTICAS/R/DCA/

```
5 metodo1      23
6 metodo2      18
> my_data$METODOS <- ordered(my_data$METODOS,
+                             levels = c("metodo1", "metodo2", "metodo3"))
> # show the levels
> levels(my_data$METODOS)
[1] "metodo1" "metodo2" "metodo3"
> # Box plots
> # Box plot
> boxplot(PRODUCCION ~ METODOS, data = my_data,
+          xlab = "Treatment", ylab = "PRODUCCION",
+          frame = FALSE, col = c("#00AFBB", "#E7B800", "#FC4E07"))
>
```



Source

Console C:/Users/chris/Desktop/Cursos 2020-2/METODOS ESTADISTICOS 2020-2/PRACTICAS/R/DCA/

```

> # Box plots
> # Box plot
> boxplot(PRODUCCION ~ METODOS, data = my_data,
+         xlab = "Treatment", ylab = "PRODUCCION",
+         frame = FALSE, col = c("#00AFBB", "#E7B800", "#FC4E07"))
> # Box plots
> # Box plot
> boxplot(PRODUCCION ~ METODOS, data = my_data,
+         xlab = "Treatment", ylab = "PRODUCCION",
+         frame = FALSE, col = c("#00AFBB", "#E7B800", "#FC4E07"))
> ## One factor ANOVA example from Dobson's book, cf. Table 7.4:
> require(stats); require(graphics)
> boxplot(PRODUCCION ~ METODOS, data = my_data, main = "Datos Crecimiento",
+         ylab = "Dried PRODUCCION of plants", col = "lightgray",
+         notch = FALSE, varwidth = TRUE)
> ## One factor ANOVA example from Dobson's book, cf. Table 7.4:
> require(stats); require(graphics)
> boxplot(PRODUCCION ~ METODOS, data = my_data, main = "Métodos de Entrenamiento",
+         ylab = "Produccion Promedio Diaria", col = "lightgray",
+         notch = FALSE, varwidth = TRUE)
> anova(lm(PRODUCCION ~ METODOS, data = my_data))
Analysis of variance Table

```

Response: PRODUCCION

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
METODOS	2	115.73	57.867	9.234	0.003733 **
Residuals	12	75.20	6.267		

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

> # Compute the analysis of variance
> res.aov <- aov(PRODUCCION ~ METODOS, data = my_data)
> # Summary of the analysis
> summary(res.aov)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
METODOS	2	115.7	57.87	9.234	0.00373 **
Residuals	12	75.2	6.27		

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

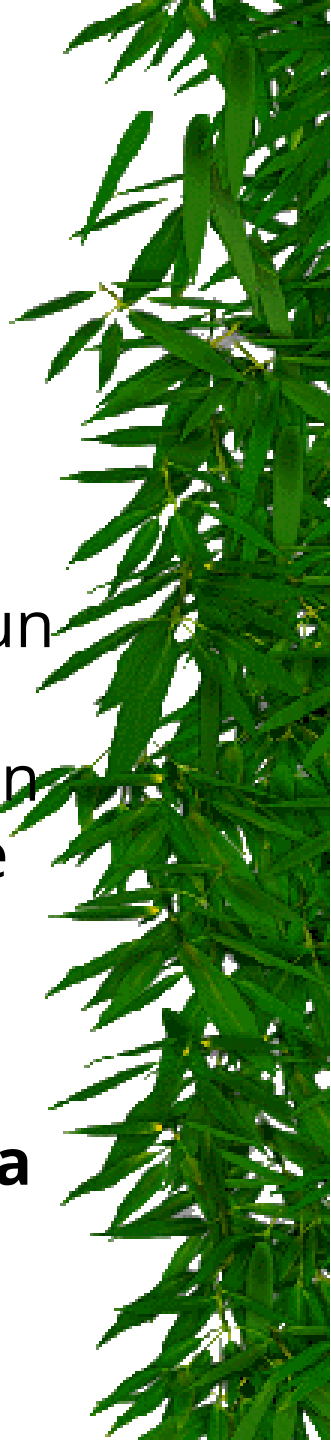
&gt;

**¿Qué sucede si mis  
datos no cumplen  
estos supuestos?**



# **El ANOVA de una vía se considera una prueba robusta frente a la falta de normalidad.**

- Esto significa que tolera las violaciones a su supuesto de normalidad bastante bien siempre que no sea demasiada.
- Puede tolerar datos que no son normales (distribuciones sesgadas o empinadas) con sólo un pequeño efecto sobre la tasa de error Tipo I.
- Sin embargo, las distribuciones aplanadas pueden tener un efecto profundo cuando los tamaños de grupo son pequeños.
- **Esto nos deja con dos alternativas: (1) transformar los datos para que la forma de la distribución sea normal o (2) elegir una prueba no paramétrica que no supone normalidad.**





- **En el caso de que no exista homogeneidad de variantes podemos aplicar el ANOVA de Welch y una prueba de Games-Howell en lugar de una prueba post hoc de Tukey.** Otra alternativa es aplicar una prueba no paramétrica.

- **La falta de independencia de los casos es la más grave.** Deberíamos pensar en otro tipo de prueba para hacer frente a este problema, por ejemplo un diseño de medidas repetidas.

