Inferencia de información para dos o más poblaciones

Contrastes de hipótesis para dos poblaciones y comparación de grupos mediante ANOVA

Blanca de la Fuente y Ángel A. Juan

PID_00161060

Universitat Oberta de Catalunya
www.uoc.edu

Índice

Introducción

En los módulos anteriores se introdujeron los conceptos básicos de estimación y de contraste de hipótesis relacionados con una población. En la práctica cotidiana, sin embargo, es fácil encontrarse con situaciones en las que se dispone de dos o más grupos de individuos o poblaciones y, en tal caso, el interés radica a menudo en ser capaz de discernir si dichos grupos o poblaciones se pueden considerar como semejantes –desde un punto de vista estadístico– o si, por el contrario, son grupos o poblaciones que muestran diferencias significativas entre ellos. Así, por ejemplo, puede ser conveniente comparar las calificaciones medias de dos grupos de estudiantes en función de si han hecho o no uso de una metodología docente innovadora, comparar los porcentajes de recuperación de dos o más grupos de enfermos según el tratamiento recibido, comparar las calidades medias de diferentes accesos a Internet en función de la empresa proveedora, comparar los precios medios de los servicios de obtención de documentos en función de la institución que los ofrezca, etc.

Cuando se consideran dos grupos o poblaciones, las técnicas que se usan para comparar las respectivas medias o proporciones son muy similares a las utilizadas en el caso de una población: contrastes de hipótesis basados en el uso de la distribución normal (cuando se comparan dos proporciones) o de la *t*-Student (cuando se comparan dos medias). En el caso de la comparación entre dos medias de grupos distintos, hay que distinguir si se trata de dos grupos independientes (por ejemplo, cuando se comparan los resultados de un test realizados a dos grupos distintos de individuos) o bien si se trata de dos grupos dependientes (por ejemplo, cuando se están considerando los resultados de un test previo con los resultados de un test posterior, ambos realizados al mismo grupo de individuos).

Finalmente, en el caso de que se deseen comparar más de dos grupos o poblaciones, los contrastes anteriores ya no sirven y resulta necesario recurrir a las técnicas ANOVA basadas en la distribución *F*-Snedecor. El uso de estas técnicas posibilita discernir si las medias correspondientes a un conjunto de tres o más grupos son todas aproximadamente iguales o si, por el contrario, se puede establecer que existen diferencias significativas entre algunas de ellas (y, por consiguiente, entre los grupos asociados).

Objetivos

Los objetivos académicos del presente módulo se describen a continuación:

1. Comparar dos poblaciones utilizando procedimientos similares a los vistos para una sola población.

2. Aprender a formular una hipótesis sobre la naturaleza de las dos poblaciones y la diferencia entre sus medias o proporciones.

3. Conocer el método para comparar las varianzas de dos poblaciones. Para realizar estos contrastes se introduce la distribución *F*.

4. Entender la importancia práctica de las técnicas ANOVA a la hora de discernir si existen diferencias significativas entre más de dos grupos o poblaciones.

5. Aprender a usar los tests *F* de ANOVA y saber interpretar adecuadamente los resultados que ofrecen.

6. Comprender la lógica que subyace a la metodología ANOVA.

7. Conocer las hipótesis que se han de satisfacer para poder aplicar las técnicas ANOVA con garantías.

8. Aprender a usar software estadístico y/o de análisis de datos como instrumento básico en la aplicación práctica de los conceptos y técnicas estadísticas.

1. Contrastes de hipótesis para dos poblaciones

En este módulo se presentan métodos para contrastar las diferencias entre las medias o proporciones de dos poblaciones y para contrastar varianzas.

Para comparar las medias o las proporciones poblacionales, se extrae una muestra aleatoria de las dos poblaciones y la inferencia sobre la diferencia entre ambas medias o proporciones se basa en los resultados muestrales. El método apropiado para analizar la información depende del procedimiento empleado al seleccionar las muestras. Consideramos las dos posibilidades siguientes:

a) Muestras dependientes (**datos pareados**): en este procedimiento, las muestras se eligen por pares, una de cada población. La idea es que aparte de la característica objeto del estudio, los elementos de cada uno de estos pares deben estar relacionados, de manera que la comparación pueda establecerse directamente. Por ejemplo, supongamos que queremos medir la eficacia de un curso de lectura rápida. Una manera de abordar el problema sería tomar nota de las palabras leídas por minuto por una muestra de alumnos antes de tomar el curso y compararlas con los resultados obtenidos *por los mismos alumnos* una vez completado el curso. En este caso, cada par consistiría en medidas de velocidad de un mismo alumno realizadas antes y después del curso, se podría averiguar si existen pruebas contundentes de la eficacia del curso de lectura rápida.

b) Muestras independientes: en este método se extraen muestras independientes de cada una de las dos poblaciones, de manera que los miembros de una muestra no tienen necesariamente relación con los miembros de la otra. Por ejemplo, se realiza un estudio para evaluar las diferencias en los niveles educativos entre dos centros de capacitación, se aplica un examen común a personas que asisten a cada centro. Las calificaciones del examen son uno de los factores principales para evaluar diferencias de calidad entre los centros.

1.1. Contrastes de hipótesis para la diferencia de medias

Contraste de hipótesis para la diferencia entre las medias de dos poblaciones: muestras independientes.

En este apartado presentaremos los procedimientos para contrastar las hipótesis acerca de la diferencia de medias de dos poblaciones.

Se supone que se dispone de muestras aleatorias independientes de n_1 y n_2, observaciones procedentes de dos poblaciones normales con medias μ_1 y μ_2 y varianzas conocidas σ_1^2 y σ_2^2 respectivamente. Se desea contrastar la hipótesis nula (H_0) que afirma que los valores de las medias de las dos poblaciones son iguales: $\boldsymbol{H_0}$**:** $\boldsymbol{\mu_1 - \mu_2 = 0}$ frente a cualquiera de las hipótesis alternativas:

Nota

A veces en lugar de:
H_0: $\mu_1 - \mu_2 = 0$
escribiremos:
H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 - \mu_2 \neq 0$, H_1: $\mu_1 - \mu_2 < 0$, H_1: $\mu_1 - \mu_2 > 0$. Se fija un nivel de significación α para realizar el contraste.

El estadístico de contraste será:

$$z^* = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

donde $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sigma_{\bar{x}_1 - \bar{x}_2}$ es el error estándar.

Es una observación de una distribución $N(0,1)$.

Recordad

$\bar{X}_1 \rightarrow N(\mu_1, \sigma_1)$ y
$\bar{X}_2 \rightarrow N(\mu_2, \sigma_2)$

La variable diferencia de medias muestrales:

$(\bar{X}_1 - \bar{X}_2) \rightarrow N\left(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$

En el caso de que no se pueda asegurar que las muestras provienen de poblaciones normales, sólo podremos contrastar la diferencia de medias si los tamaños de las muestras son superiores a treinta.

El teorema central del límite dice que si tenemos un grupo numeroso de variables independientes y todas ellas siguen el mismo modelo de distribución (cualquiera que éste sea), la suma de ellas se distribuye según una distribución normal estándar.

Por lo tanto el estadístico de contraste:

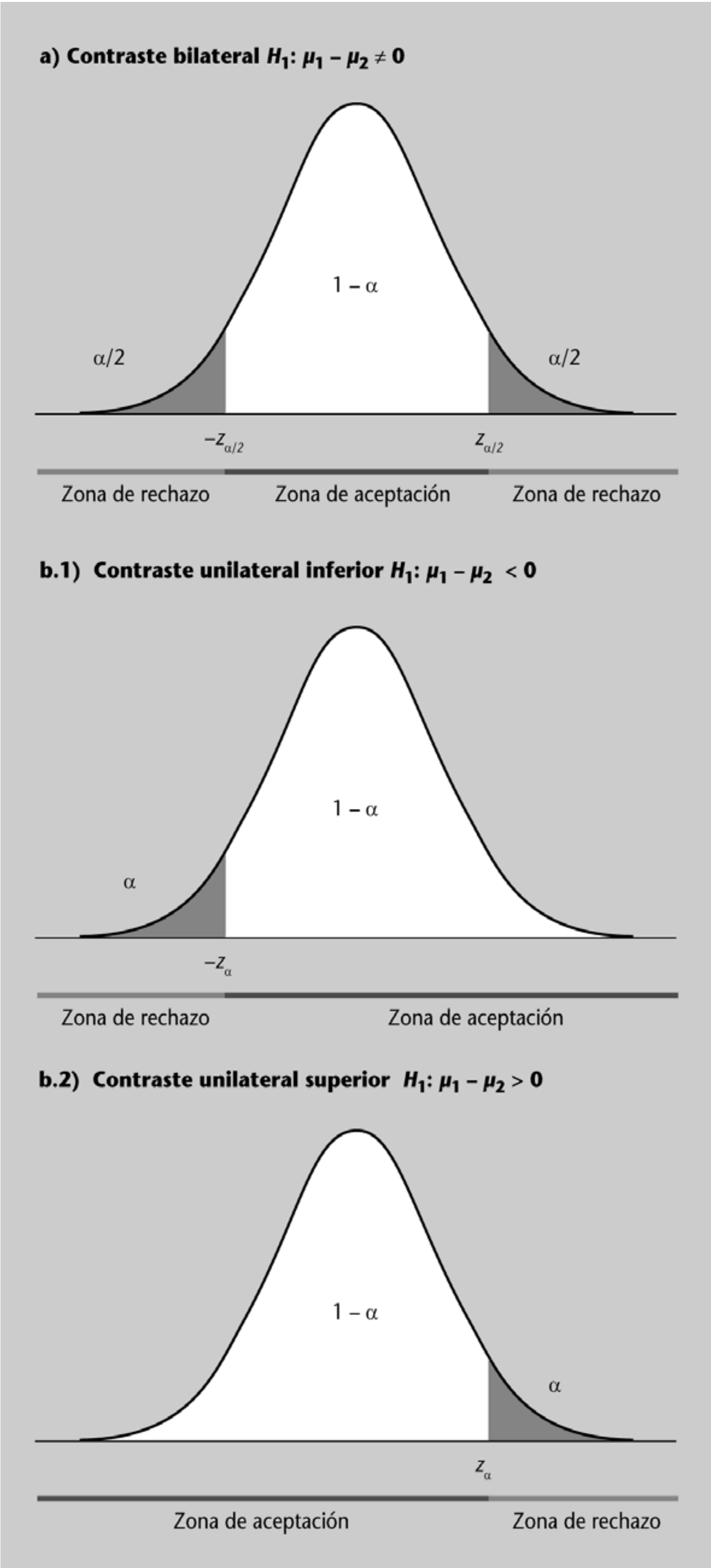
$$z^* = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

es una observación de una variable aleatoria que se distribuye aproximadamente como una $N(0,1)$.

Regla de decisión del contraste de hipótesis

Las regiones de rechazo de la hipótesis nula H_0: $\mu_1 - \mu_2 = 0$ son:

Figura 1. Regiones de rechazo para contrastes de las diferencias de medias



Varianzas poblacionales conocidas

Se puede actuar de dos maneras:

1) A partir del ***p*-valor** según sea H_1:

- p-valor = $P(|Z| > |z^*|)$
- p-valor = $P(Z < z^*)$
- p-valor = $P(Z > z^*)$

2) Si p-valor $\leq \alpha$ **se rechaza H_0** a partir de los valores críticos según sea H_1:

- Si $|z^*| > z_{\alpha/2}$ se rechaza H_0
- Si $z^* < -z_\alpha$ se rechaza H_0
- Si $z^* > -z_\alpha$ se rechaza H_0

donde

z_α es tal que $P(Z > z_\alpha) = \alpha$ y

$z_{\alpha/2}$ es tal que $P(Z > z_{\alpha/2}) = \alpha/2$

Una vez que se ha calculado el valor del estadístico de contraste, se debe determinar el *p*-valor. El *p*-valor depende de la hipótesis alternativa planteada.

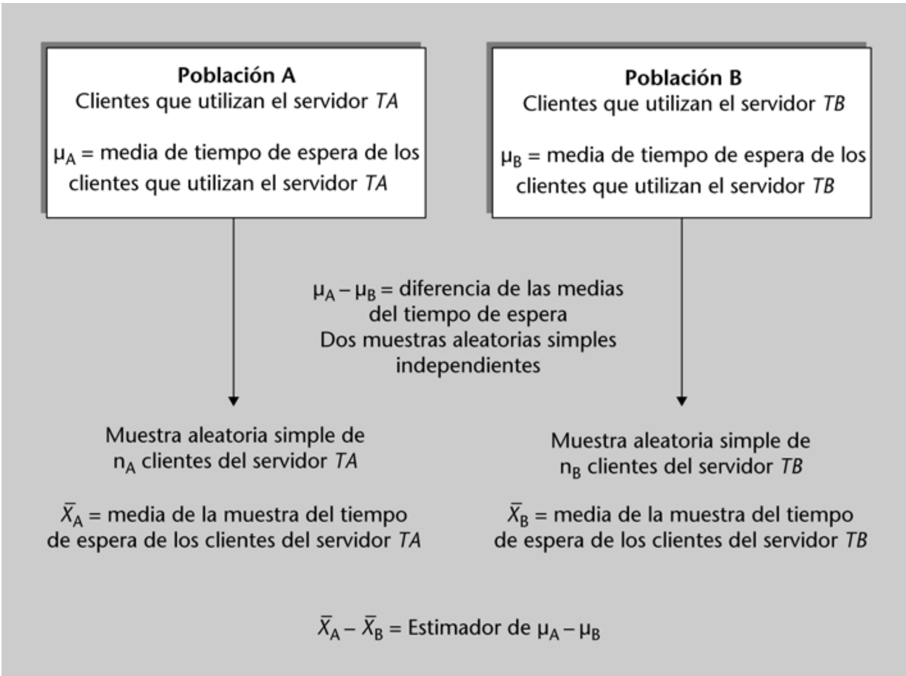
- Si $H_1 : \mu_1 - \mu_2 \neq 0$, entonces $p = 2P(Z < |z|)$
- Si $H_1 : \mu_1 - \mu_2 < 0$, entonces $p = P(Z < z)$
- Si $H_1 : \mu_1 - \mu_2 > 0$, entonces $p = P(Z > z)$

Los *p*-valores de estos contrastes son la probabilidad de obtener un valor al menos tan extremo como el estadístico de contraste obtenido.

> Si el *p*-valor es significativo se rechaza la hipótesis nula si es menor que el nivel de significación α fijado.

Ejemplo 1. "Comparación de las medias del tiempo de respuesta de dos servidores".

Figura 2. Estimación de la diferencia entre las medias de dos poblaciones



En una empresa informática se desea medir la eficiencia de dos servidores web. Para ello, miden el tiempo de espera del cliente entre la petición que hace y la respuesta que le da el servidor. En la tabla 1 vemos los tiempos de espera (en milisegundos) de ambos servidores (*TA* y *TB*) para cincuenta peticiones son:

Tabla 1. Datos del ejemplo 1. "Comparación de las medias del tiempo de respuesta de dos servidores"

Tiempo de espera para el servidor *A*				**Tiempo de espera para el servidor *B***			
9,67	10,01	8,08	10,01	6,45	6,94	12,11	10,31
9,62	10,55	9,98	9,96	9,64	10,47	12,55	10,83
9,50	11,26	10,30	9,28	8,53	8,47	7,98	8,41

Tiempo de espera para el servidor *A*				**Tiempo de espera para el servidor *B***			
10,88	10,64	7,05	10,30	9,20	7,42	10,20	9,15
8,94	10,23	11,79	11,08	4,55	7,48	11,28	7,06
10,59	11,63	9,59	10,05	8,51	11,01	6,53	8,04
9,81	8,91	10,88	9,74	12,11	9,56	8,14	11,70
9,46	10,27	9,83	11,14	7,65	6,80	8,99	10,56
9,26	9,49	10,92	9,44	8,85	8,99	10,01	7,82
9,02	8,99	10,98	9,17	8,45	7,48	8,14	6,01
8,61	10,09	9,54	10,86	8,80	12,57	9,69	8,82
9,42	9,11	10,17		8,82	7,97	7,03	
10,86	9,47	10,32		9,85	8,62	8,59	

Supongamos que las muestras aleatorias de los tiempos de espera son independientes. La empresa quiere saber si el servidor *A* es menos eficiente (más lento) que el servidor *B* con un nivel de confianza del 99%.

Para contestar a estas preguntas se hará un contraste para comparar dos medias. Dado que el enunciado nos pregunta "si el servidor *A* es menos eficiente que el servidor *B*", considerando que un servidor es menos eficiente si es más lento, entonces hemos de contrastar si la media del tiempo de espera del servidor *A* es más grande que la media del tiempo de espera del servidor *B*. Así pues, tenemos que plantear una **hipótesis alternativa unilateral**.

- Las hipótesis nula y alternativa son: $\begin{array}{l} H_0 : \mu_A - \mu_B = 0 \\ H_1 : \mu_A - \mu_B > 0 \end{array}$
- Fijamos $\alpha = 0{,}01$.
- No podemos asegurar que las poblaciones sean normales, pero como hemos mencionado anteriormente, al tratarse de muestras grandes (superiores a treinta observaciones) el estadístico de contraste será:

$$z^* = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Es una observación de una variable aleatoria que **se distribuye aproximadamente como una *N*(0,1)**.

Para resolverlo manualmente calcularemos primero los valores muestrales como hemos expuesto en los módulos anteriores:

Tiempo de espera para el servidor *A*	Tiempo de espera para el servidor *B*
$n_A = 50$	$n_B = 50$
$\bar{x}_A = 9{,}94$	$\bar{x}_B = 8{,}90$
$s_A = 0{,}90$	$s_B = 1{,}75$

Las varianzas muestrales s_A^2 y s_B^2 para estimar las varianzas poblacionales y calcular el estadístico z^*:

$$z^* = \frac{(9,94 - 8,90)}{\sqrt{\frac{0,90^2}{50} + \frac{1,75^2}{50}}} = 3,75$$

Ahora se puede calcular el p-valor $p = P(Z > 3,75) = 0,00$

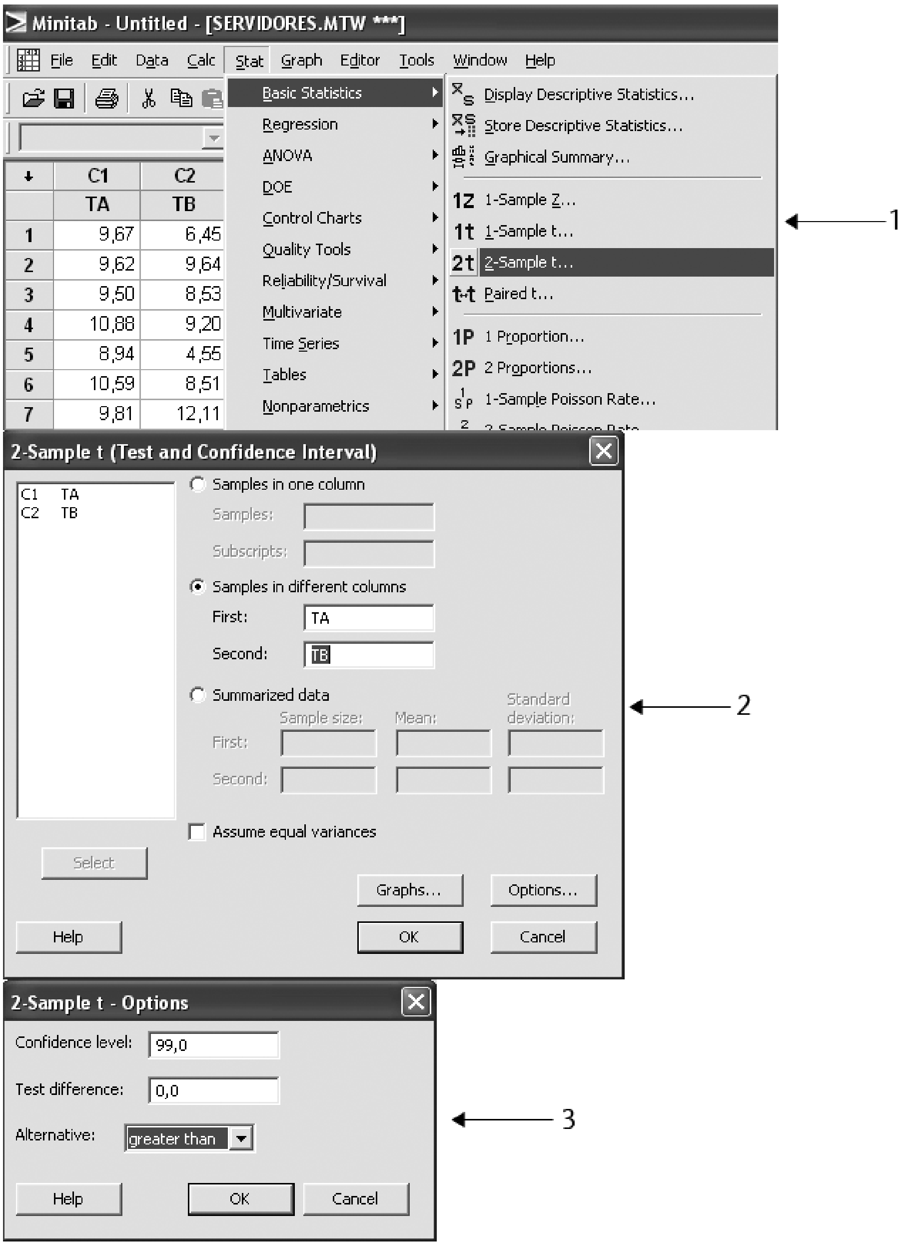
Puesto que el p-valor es menor que $\alpha = 0,01$, se rechaza la hipótesis nula a favor de la alternativa. Así, el tiempo medio de espera del servidor *A* es mayor que el del *B*. Luego el servidor *A* es menos eficiente que el *B*.

Ejemplo con Minitab: si el ejemplo anterior se resuelve con Minitab, se observa que el programa no ofrece la opción de usar la distribución normal. De todas formas, dado que las muestras son muy grandes, sabemos que la distribución *t* de Student se acerca a la normal a medida que aumenta el número de grados de libertad. Por tanto, los resultados que da Minitab no serán similares por la aproximación a lo normal.

Los resultados de la figura 3 muestran el p-valor = 0,000 < 0,001. Esto indica que podemos rechazar la hipótesis nula concluyendo que las medias de tiempos de espera del servidor *A* es mayor que las del *B*. Luego el servidor *A* es menos eficiente que el *B*.

Los grados de libertad (*DF*) del estadístico *t* aumentan si las poblaciones tienen distribución aproximadamente normal pero las varianzas poblacionales no son iguales.

Figura 3. Pasos para realizar un contraste de hipótesis para la diferencia de medias para muestras independientes



Pasos a seguir

Una vez introducidos los datos en el programa, se sigue la ruta ***Stat*** > ***Basic Statistics*** > ***2-Sample t (1)***, y se seleccionan las variables en la ventana correspondiente ***(2)***. En el cuadro de dialogo ***Options*** se completan los campos ***Confidence level***: 99,0 y el tipo de hipótesis alternativa ***Alternative:*** *greater than* ***(3)***.

Seleccionad ***OK*** para obtener el contraste.

Observad

En el paso *(2)* no presuponemos que las varianzas sean iguales.

Figura 4. Resultados del contraste de hipótesis

Two-Sample T-Test and CI: TA; TB

```
Two-sample T for TA vs TB

     N   Mean  StDev  SE Mean
TA  50  9,935  0,900     0,13
TB  50   8,90   1,75     0,25


Difference = mu (TA) - mu (TB)
Estimate for difference:  1,032
99% lower bound for difference:  0,371
T-Test of difference = 0 (vs >): T-Value = 3,71
 P-Value = 0,000  DF = 73
```

Contrastes para muestras con varianzas poblacionales desconocidas pero iguales

El procedimiento que utilizamos se basa en la distribución t con $n_1 + n_2 - 2$ grados de libertad.

El estadístico de contraste será:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde s es la desviación típica común.

Ejemplo 2. "Estudio sobre la producción científica".

El director de una escuela universitaria quiere comparar dos departamentos, *A* y *B*, de tamaño similar, por lo que se refiere al "número total de publicaciones o ponencias de calidad que puedan aportar mejoras a la actividad docente de la escuela". Se considerará que una publicación es de calidad cuando se haya publicado en una revista indexada o la haya publicado una editorial de prestigio internacional; se considerará que una ponencia es de calidad cuando se haya desarrollado en un congreso internacional con proceso de selección; para determinar si la publicación o ponencia puede aportar mejoras a la actividad docente se ha constituido un tribunal de expertos independientes.

Se ha tomado una muestra aleatoria formada por seis profesores del departamento *A* y se ha hallado el valor de la variable "número total de publicaciones o ponencias de calidad para cada uno de dichos profesores". Se ha hecho lo propio con otra muestra aleatoria formada por ocho profesores del departamento *B*. Los resultados se presentan a continuación:

Tabla 2. Datos del ejemplo 2. "Estudio sobre la difusión científica"

Dep. *A*	5	8	7	6	9	7		
Dep. *B*	8	10	7	11	9	12	14	9

Para un nivel de significación alfa = 0,05, ¿puede afirmarse que la producción media de ambos departamentos (según los criterios establecidos) es significativamente distinta?

Para realizar el estudio partiremos del supuesto de que no hay diferencias en el "número total de publicaciones o ponencias de calidad de ambos departamentos". Por consiguiente, en términos de la media del número total de publicaciones o ponencias de calidad, la hipótesis nula es que la diferencia de medias es cero. Si la evidencia de la muestra conduce al rechazo de esta hipótesis, llegaremos a la conclusión de que las medias de calidad son distintas para las dos poblaciones, lo que indica que hay diferencia en las publicaciones de calidad de los dos departamentos, y eso induciría a encontrar las razones de esa diferencia.

En este estudio hay dos poblaciones: una de los profesores del departamento *A*, y otra de los profesores del departamento *B*. Suponemos que ambas poblaciones son normales y que sus varianzas son iguales pero desconocidas.

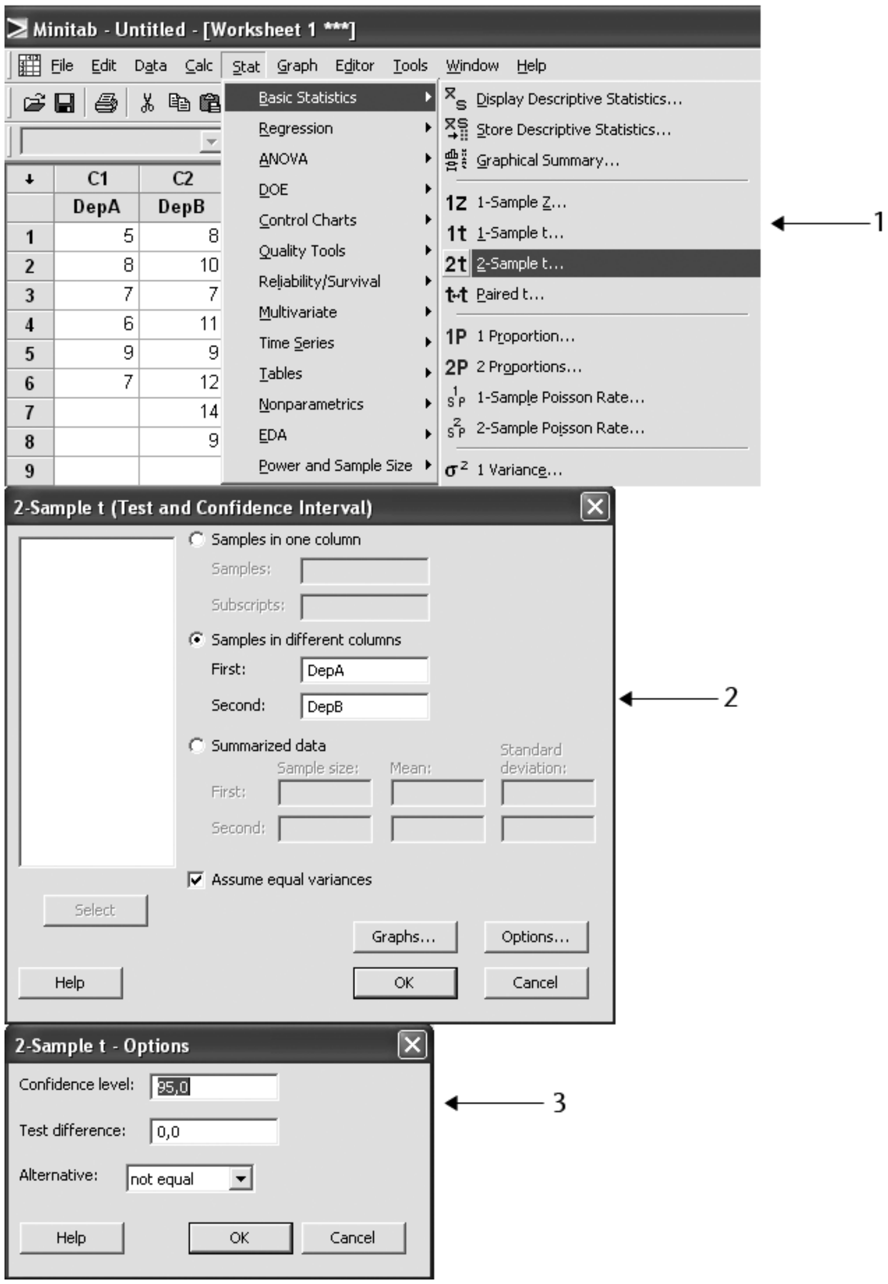
Considerando el número de publicaciones y ponencias, las medias de población son: μ_A y μ_B, se plantean las hipótesis de trabajo:

$$H_0\text{: } \mu_1 - \mu_2 = 0 \text{ (ambas medias son iguales)}$$
$$H_1\text{: } \mu_1 - \mu_2 \neq 0 \text{ (ambas medias son distintas)}$$

Se trata de un contraste de **hipótesis bilateral** sobre la media de dos poblaciones independientes.

Se empleara Minitab para probar las hipótesis acerca de la diferencia entre las medias de dos poblaciones (figura 5).

Figura 5. Pasos para realizar un contraste de hipótesis para la diferencia de medias para muestras con varianzas poblacionales desconocidas



Pasos a seguir

Una vez introducidos los datos en el programa se sigue la ruta ***Stat*** > ***Basic Statistics*** > ***2-Sample t (1)***, y se seleccionan las variables en la ventana correspondiente ***(2)***.En el cuadro de dialogo ***Options*** se completan los campos ***Confidence level***: 95,0 y el tipo de hipótesis alternativa ***Alternative***: *not equal* ***(3)***.

Seleccionad ***OK*** para obtener el contraste.

Observad

En el paso ***(2)*** suponemos que las varianzas son desconocidas pero iguales y marcamos la casilla correspondiente.

Obtuvimos los resultados de la figura 6. Aparecen los valores muestrales de ambos departamentos. El estadístico de contraste es un valor $t = -2{,}84$ con 12 grados de libertad (*DF*) y el *p*-valor *P-Value* = 0,015.

Figura 6. Resultados del contraste de hipótesis

```
Two-Sample T-Test and CI: DepA; DepB

Two-sample T for DepA vs DepB

      N   Mean  StDev  SE Mean
DepA  6   7,00   1,41     0,58
DepB  8  10,00   2,27     0,80


Difference = mu (DepA) - mu (DepB)
Estimate for difference:  -3,00
95% CI for difference:  (-5,30; -0,70)
T-Test of difference = 0 (vs not =): T-Value = -2,84
P-Value= 0,015  DF = 12
Both use Pooled StDev = 1,9579
```

Como *p*-valor = 0,015 < 0,05, se puede rechazar la hipótesis nula con $\alpha = 0{,}05$. Así, la producción media de ambos departamentos (según los criterios establecidos) es significativamente distinta en los departamento *A* y *B*. Observad que la información de Minitab para el intervalo de confianza del 95% en la figura 5 tiene como extremos los valores –5,30 y –0,70 (observad que el 0 no está incluido en dicho intervalo). Esto también nos indica que debemos rechazar la hipótesis nula y aceptar la alternativa (las medias son distintas).

Así, los resultados permiten que el director de la escuela universitaria concluya que existen diferencias significativas entre ambos departamentos en el "número total de publicaciones o ponencias de calidad".

Aplicando **Microsoft Excel** al ejemplo 2. "Estudio sobre la difusión científica".

Análisis de datos

Para realizar contrastes de hipótesis con **MS Excel** es necesario instalar previamente un complemento llamado "Análisis de datos". Para instalar las herramientas de análisis de datos haced clic en ***Herramientas*** > ***complementos***, en el cuadro de diálogo activar ***Herramientas para análisis***.

Para ejecutar una prueba *t* de dos muestras independientes para datos no apareados haced clic en (*t-Test: Two Simple* > *Assuming Equal Variants*) "prueba *t*: dos muestras suponiendo varianzas iguales" y especificad las dos columnas que contengan los datos.

La figura 7 muestra el correspondiente *output* que ofrece **Microsoft Excel**.

Figura 7. Resultados ejemplo 2. "Estudio sobre la difusión científica". Excel

	A	B	C
1	Prueba *t* para dos muestras suponiendo varianzas iguales		
2			
3		*DepA*	*DepB*
4	Media	7	10
5	Varianza	2	5,14285714
6	Observaciones	6	8
7	Varianza agrupada	3,83333333	
8	Diferencia hipotética de las medias	0	
9	Grados de libertad	12	
10	Estadístico *t*	-2,8371975	
11	*P*(*T*<=*t*) una cola	0,0074872	
12	Valor crítico de *t* (una cola)	1,78228755	
13	*P*(*T*<=*t*) dos colas	0,01497439	
14	Valor crítico de *t* (dos colas)	2,17881283	

Como observamos, el p-valor = 0,0149, al ser menor que el valor de α, se puede rechazar la hipótesis nula con $\alpha = 0{,}05$.

Contraste de hipótesis para la diferencia entre las medias de dos poblaciones: muestras dependientes (datos pareados)

Muestras dependientes

Muestras dependientes significa que tenemos **una muestra** de observaciones de dos variables.

La media de la muestra es:

$$\bar{d} = \frac{\sum_i^n d_i}{n}$$

La desviación estándar es:

$$s_d = \sqrt{\frac{\sum_i^n (d_i - \bar{d})^2}{n-1}}$$

La notación d es para recordar que la muestra apareada produce datos de *diferencia.*

Disponemos de una muestra aleatoria de n pares de observaciones de distribuciones con medias μ_A y μ_B. Denotamos por $\bar{d}$ y s_d la media muestral y la desviación típica observadas para las n diferencias (x_A-x_B) y sea $\mu_d = \mu_A - \mu_B$ media de las diferencias para la población.

Si la distribución poblacional es normal podemos realizar los siguientes contrastes para un nivel de significación α:

la hipótesis nula: $H_0 = \mu_d = 0$

la hipótesis alternativa (H_1) puede ser bilateral: $H_1 : \mu_d \neq 0$

o unilateral H_1: $\mu_d > 0$ o $H_1 : \mu_d < 0$

En este tipo de contraste se usa la misma metodología usada para el contraste de la media para una sola población que vimos en el módulo anterior.

Para ilustrar el diseño con muestras emparejadas ilustrar el ejemplo siguiente:

Ejemplo 3. "Puntuaciones de un test de actitud"

A un grupo de personas se les propuso un test de actitud acerca de un tema polémico y obtuvimos unos resultados. Luego el grupo asistió a la proyección de una película favorable al tema y acto seguido se les propuso de nuevo el test de actitud, del que se obtuvieron otros resultados. En la tabla 3 aparecen los datos acerca de las puntuaciones del test realizado a once personas. Cada persona da un par de valores, uno para antes de asistir a la proyección de la película y otro después de asistir a la proyección. Se quiere verificar la hipótesis de que la proyección de una película favorable hace que cambie la actitud desfavorable hacia el tema.

Tabla 3. Datos del ejemplo 3. "Puntuaciones de un test de actitud"

Persona	Puntuación del test antes de ver la película	Puntuación del test después de ver la película	Diferencia de puntuaciones del test (d_i)
1	24	16	8
2	20	18	2
3	24	20	4
4	28	24	4
5	30	24	6

Persona	Puntuación del test antes de ver la película	Puntuación del test después de ver la película	Diferencia de puntuaciones del test (d_i)
6	20	22	-2
7	24	20	4
8	22	18	4
9	18	10	8
10	18	8	10
11	24	20	4

$$\sum_{i=1}^{11} d_i = 52$$

Observad que la última columna de la tabla 3 contiene la diferencia entre las puntuaciones antes y después de ver la película. La clave para analizar el diseño con muestras apareadas es tener en cuenta que sólo se considera la columna de las diferencias. Verificaremos la hipótesis de investigación a un nivel de significación del 1% ($\alpha = 0{,}01$). Sea $\mu_d =$ la media de las diferencias para la población de personas.

Las hipótesis serán:

$$H_0\colon \mu_d = 0$$
$$H_1\colon \mu_d \neq 0$$

Se trata de un contraste bilateral. Si se rechaza H_0 se llega a la conclusión de que las medias de las puntuaciones del test son distintas al nivel de significación del 1%. En el módulo 2 se vio que si se puede suponer que la población tiene una distribución normal, el estadístico de contraste es una ***t*-Student** con $n - 1$ grados de libertad, para probar la hipótesis nula acerca de la media poblacional, si no conocemos la varianza de la población como en este ejemplo.

Con datos de diferencia se calcula el estadístico de prueba para la hipótesis nula

$$H_0\colon \mu_d = 0 \text{ es:}$$

$$\text{como } \bar{d} = \frac{52}{11} = 4{,}72 \text{ y } s_d = \sqrt{\frac{106{,}18}{10}} = 3{,}26$$

$$t^* = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{4{,}72 - 0}{3{,}26 / \sqrt{11}} = 4{,}80$$

Con $\alpha = 0{,}01$ y $n - 1 = 10$ grados de libertad ($t_{0{,}01/2} = t_{0{,}005} = 3{,}169$), la regla de rechazo para la prueba bilateral es:

$$\text{Rechazar } H_0 \text{ si } t^* < 3{,}169 \text{ o } t^* > 3{,}169$$

En vista de que $t^* = 4,80$ está en la región de rechazo, se rechaza H_0 y se acepta H_1 y podemos afirmar con un 99% de confianza que la película influyó en la actitud de las personas.

Con los resultados de la muestra podemos definir un intervalo de confianza de diferencia entre las dos medias de la población, con la metodología para población única del módulo 2 los cálculos son:

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} = 4,72 \pm 3,169\left(\frac{3,26}{\sqrt{11}}\right) = 4,72 \pm 3,12 = [1,60;\ 7,84]$$

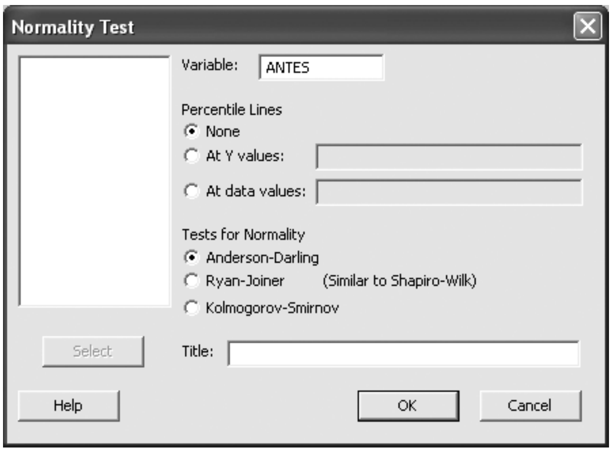
En consecuencia, el intervalo de confianza de 99% de la diferencia de medias entre las medias de las dos puntuaciones del test es de 1,6 hasta 7,84 observamos que el intervalo no incluye el valor cero, luego, como hemos visto en el contraste, podemos rechazar H_0.

Emplearemos Minitab para este ejemplo 3. “Puntuaciones de un test de actitud”.

La figura 8 muestra los pasos básicos necesarios para realizar el contraste de hipótesis.

En primer lugar comprobaremos el supuesto de que las poblaciones siguen una distribución aproximadamente normal:

Figura 8. Pasos para realizar un test de normalidad. Minitab



Pasos a seguir

Una vez introducidos los datos en el programa se sigue la ruta ***Stat*** > ***Basic Statistics*** > ***Normality Test***. Y rellenamos los campos en la ventana correspondiente.

En el cuadro de diálogo se selecciona el test de ***Anderson-Darling***.

En los gráficos resultantes (figuras 9 y 10) se observa que no hay indicios para dudar de que se cumpla el supuesto de normalidad, ya que los puntos se encuentran muy próximos a las respectivas rectas. Los gráficos nos proporcionan también el *p*-valor asociado al **test de normalidad de Anderson-Darling**, siendo dicho *p*-valor suficientemente grande en ambos casos para no descartar la hipótesis nula de este contraste: que los datos siguen una distribución normal.

Figura 9. Test de normalidad. Minitab

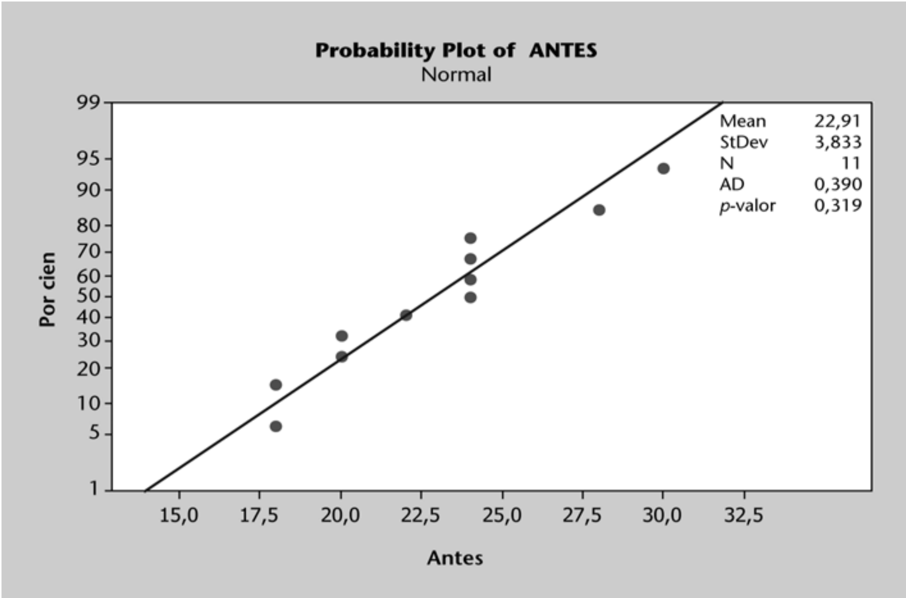
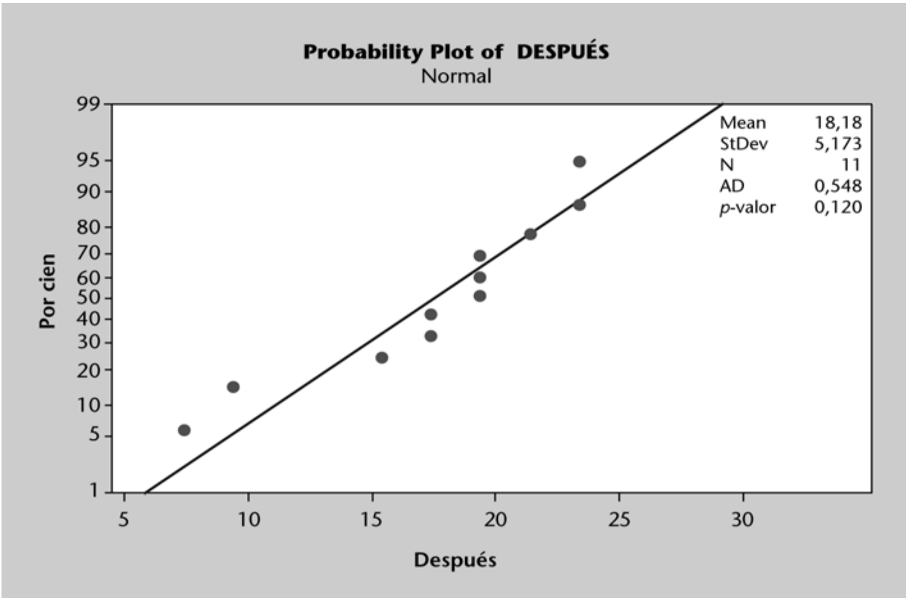
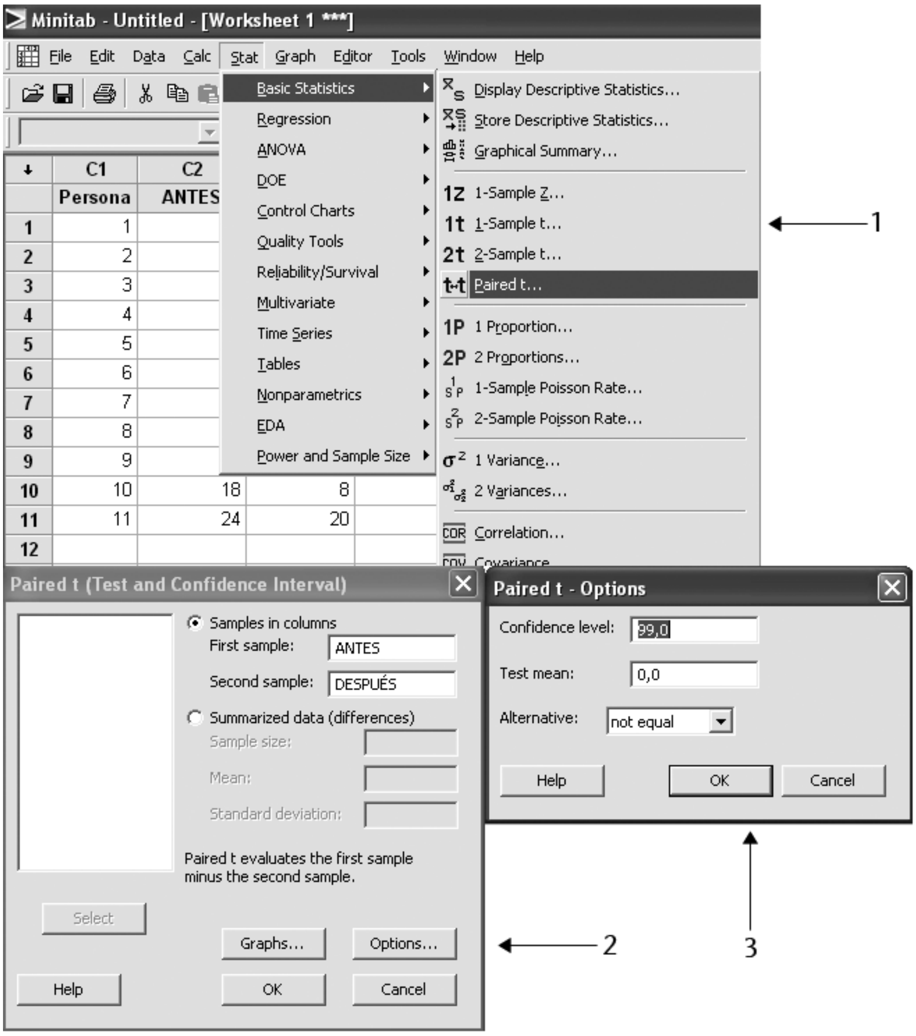


Figura 10. Test de normalidad. Minitab



Pasamos, pues a realizar las inferencias ya comentadas sobre μ_d.

Figura 11. Pasos para realizar un contraste de hipótesis para la diferencia de medias para muestras dependientes



Pasos a seguir

Se sigue la ruta ***Stat*** > ***Basic Statistics*** > ***Paired t (1)*** y se rellenan los campos en la ventana correspondiente ***(2)***. En el cuadro de diálogo ***Options*** se completan los campos ***Confidence level***: 99,0 y el tipo de hipótesis alternativa ***Alternative***: *not equal* ***(3)***.

Seleccionad ***OK*** para obtener el contraste.

Los resultados obtenidos en la figura 12 que, en base a las observaciones registradas, hay una probabilidad de 0,99 de que μ_d sea un valor del intervalo (1,613; 7,841). Además, con un p-valor de 0,001 también podemos afirmar que hay indicios suficientes para rechazar la hipótesis nula. Por lo tanto, podemos concluir que la película influyó en la actitud de las personas.

Figura 12. Resultados del contraste de medias para dos muestras dependientes. Minitab

```
Paired T for ANTES - DESPUÉS

             N   Mean  StDev  SE Mean
ANTES       11  22,91   3,83     1,16
DESPUÉS     11  18,18   5,17     1,56
Difference  11  4,727  3,259    0,982


99% CI for mean difference: (1,613; 7,841)
T-Test of mean difference = 0 (vs not = 0): T-Value = 4,81
P-Value = 0,001
```

También puede ejecutar una prueba *t* por pares utilizando **Excel**. Desde *Herramientas* > *Análisis de datos*, haced clic en *Prueba t para medias de dos muestras emparejadas* y especificad las dos columnas que contienen los datos por pares. Este comando no calcula el intervalo de confianza, de modo que tenéis que calcularlo mediante las fórmulas que aparecen en este módulo.

La figura 13 muestra el correspondiente *output* que ofrece **Microsoft Excel**.

Figura 13. Resultados del contraste de medias para dos muestras emparejadas. Excel

B13 | f_x 0,000711223615253786

	A	B	C
1	Prueba *t* para medias de dos muestras emparejadas		
2			
3		*ANTES*	*DESPUÉS*
4	Media	22,90909091	18,18181818
5	Varianza	14,69090909	26,76363636
6	Observaciones	11	11
7	Coeficiente de correlación de Pea	0,777564218	
8	Diferencia hipotética de las media	0	
9	Grados de libertad	10	
10	Estadístico *t*	4,811515866	
11	*P*(*T*<=*t*) una cola	0,000355612	
12	Valor crítico de *t* (una cola)	2,763769458	
13	*P*(*T*<=*t*) dos colas	0,000711224	
14	Valor crítico de *t* (dos colas)	3,169272672	

Al ser el p-valor = 0,0007 < α(0,01), se rechaza H_0.

1.2. Contrastes de hipótesis para la diferencia de proporciones

Al estudiar la diferencia entre dos proporciones poblacionales, el estimador es $\hat{p}_1 - \hat{p}_2$. Como hemos visto en casos anteriores, la distribución del estimador de las muestras es un factor clave para determinar los intervalos de confianza y probar las hipótesis de los parámetros.

Supongamos que disponemos de dos muestras aleatorias simples e independientes de n_1 y n_2 observaciones. Las proporciones muestrales de éxitos son respectivamente: $\hat{p}_1$ y $\hat{p}_2$.

La distribución de la variable diferencia de proporciones muestrales $\hat{p}_1 - \hat{p}_2$ se puede aproximar con una distribución $N(0,1)$.

Bajo el supuesto de la hipótesis nula cierta (H_0: $p_1 - p_2 = 0$), tenemos que el estadístico de contraste es:

$$z^* = \frac{\left(\hat{p}_1 - \hat{p}_2\right)}{\sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n_1} + \dfrac{\hat{p}\ \left(1-\hat{p}\right)}{n_2}}}$$

donde ($\hat{p}_1 - \hat{p}_2$) es la diferencia de las proporciones muestrales.

El valor $\hat{p}$ es el valor estimado común de la proporción poblacional, que podemos estimarlo a partir de las dos muestras:

Recordad

Si los tamaños de las muestras son grandes:

$$\hat{p}_1 \to N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right)$$

y

$$\hat{p}_2 \to N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right)$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Regla de decisión del contraste de hipótesis

Una vez que se ha calculado el valor del estadístico de contraste, se debe determinar el *p*-valor. El *p*-valor depende de la hipótesis alternativa planteada.

- Si $H_1 : p_1 - p_2 \neq 0$, entonces $p = 2P\ (Z < |z|)$
- Si $H_1 : p_1 - p_2 < 0$, entonces $p = P\ (Z < z)$
- Si $H_1 : p_1 - p_2 > 0$, entonces $p = P(Z > z)$

Si el *p*-valor es significativo se rechaza la hipótesis nula si es menor que el nivel de significación α fijado.

Nota

A veces, en lugar de:
H_0: $p_1 - p_2 = 0$
escribiremos:
H_0: $p_1 = p_2$

Se utilizará el ejemplo del apartado 1.1, tabla 1. Datos del ejemplo 1. "Comparación de las medias del tiempo de respuesta de dos servidores".

En una empresa informática se desea medir la eficiencia de dos servidores web. Para ello, miden el tiempo de espera del cliente entre la petición que éste hace y la respuesta que le da el servidor. Los tiempos de espera (en milisegundos) de ambos servidores (*TA* y *TB*) para cincuenta peticiones están en la tabla 2.

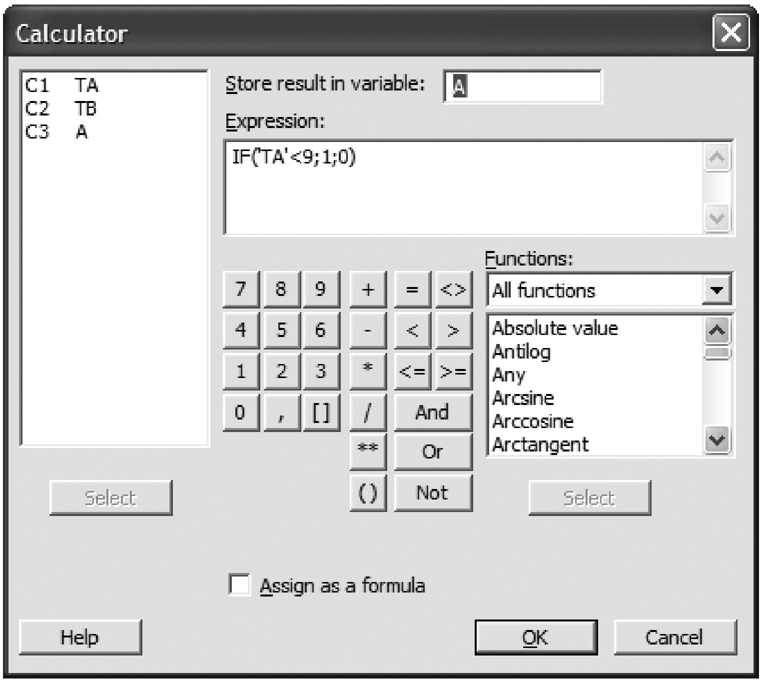
Diremos que el tiempo de espera es aceptable si es menor que 9 milisegundos. ¿Podemos decir que la proporción de peticiones con tiempo de espera aceptable es distinta para los dos servidores?

Para contestar esta pregunta debemos hacer un contraste de diferencia de proporciones que resolveremos con Minitab.

Lo primera operación es calcular para cada tipo de servidor la proporción de tiempo inferior a 9 milisegundos. Para ello, creamos una nueva columna de nombre A donde pondremos un 1 si la observación de tiempo de espera del servidor *A* es inferior a 9 y 0 en caso contrario. Después sumaremos los valores de la columna y obtendremos el número de observaciones de tiempo del servidor *A* inferior a 9 milisegundos.

En la figura 14 se indican los pasos a seguir:

Figura 14. Pasos a seguir para recalcular una variable nueva

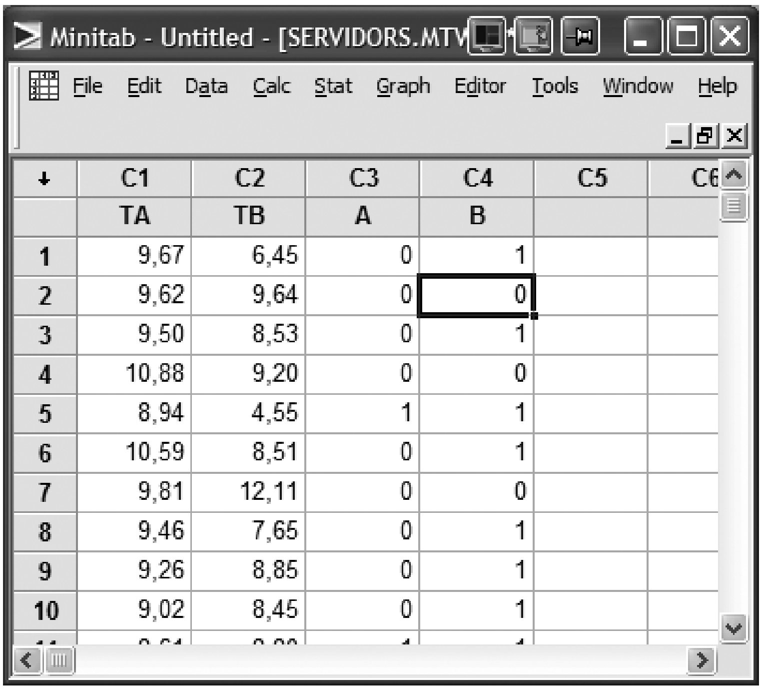


Indicación

Para hacer este ejercicio primero calcularemos una nueva variable, que valga **1** si el tiempo de espera es menor que 9 milisegundos y **0** en caso contrario. Para calcular esta variable, podemos utilizar la instrucción **IF** de Minitab.

Hacemos lo mismo para el tiempo del servidor *B*, creamos una columna de nombre *B* con 1 si el tiempo es inferior a 9 y 0 en caso contrario.

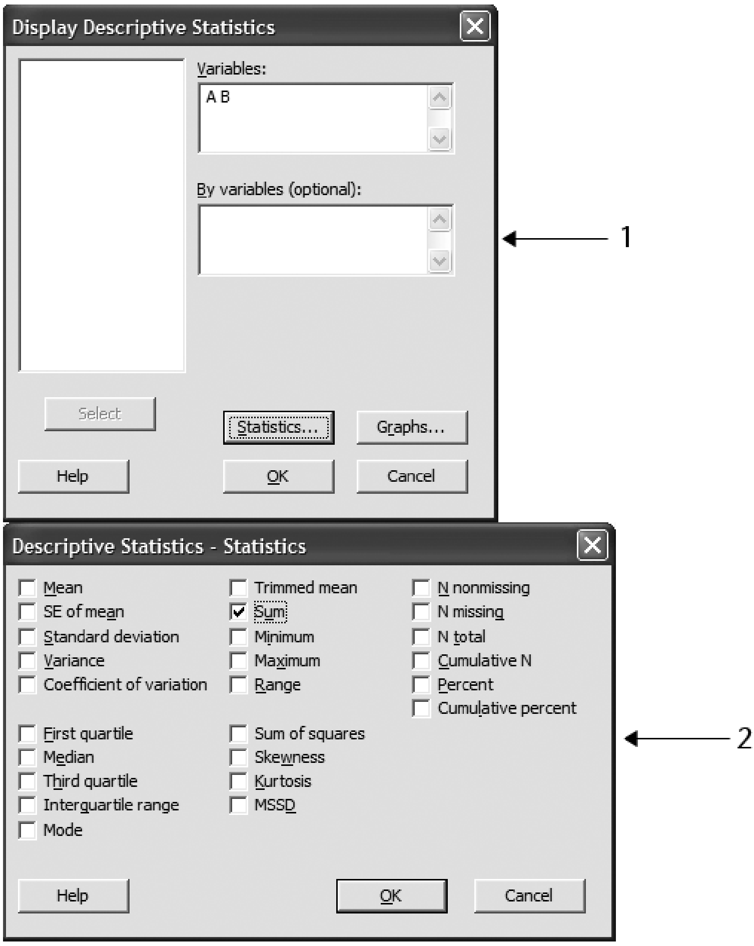
Figura 15. Datos



↓	C1	C2	C3	C4	C5	C6
	TA	TB	A	B		
1	9,67	6,45	0	1		
2	9,62	9,64	0	0		
3	9,50	8,53	0	1		
4	10,88	9,20	0	0		
5	8,94	4,55	1	1		
6	10,59	8,51	0	1		
7	9,81	12,11	0	0		
8	9,46	7,65	0	1		
9	9,26	8,85	0	1		
10	9,02	8,45	0	1		

Una vez tenemos estas dos nuevas columnas, calculamos la suma de cada una de ellas y así tendremos para cada servidor el número de observaciones de tiempo inferior a 9:

Figura 16. Pasos a seguir para obtener el valor suma



Pasos a seguir

Se sigue la ruta ***Stat*** > ***Basic Statistics*** > ***Display Descriptive Statistics***. ***(1)***, y se rellenan los campos en la ventana correspondiente.

En el cuadro de dialogo ***Statistic*** se marca ***Sum (2)***.

Seleccionad ***OK*** para obtener el contraste.

Figura 17. Resultados

Descriptive Statistics: A; B

Variable	Sum
A	6.0000
B	31.0000

Para el servidor *A* hay seis observaciones con un tiempo de espera inferior a 9 milisegundos y para el servidor *B* el número de observaciones menores de 9 milisegundos es treinta y una.

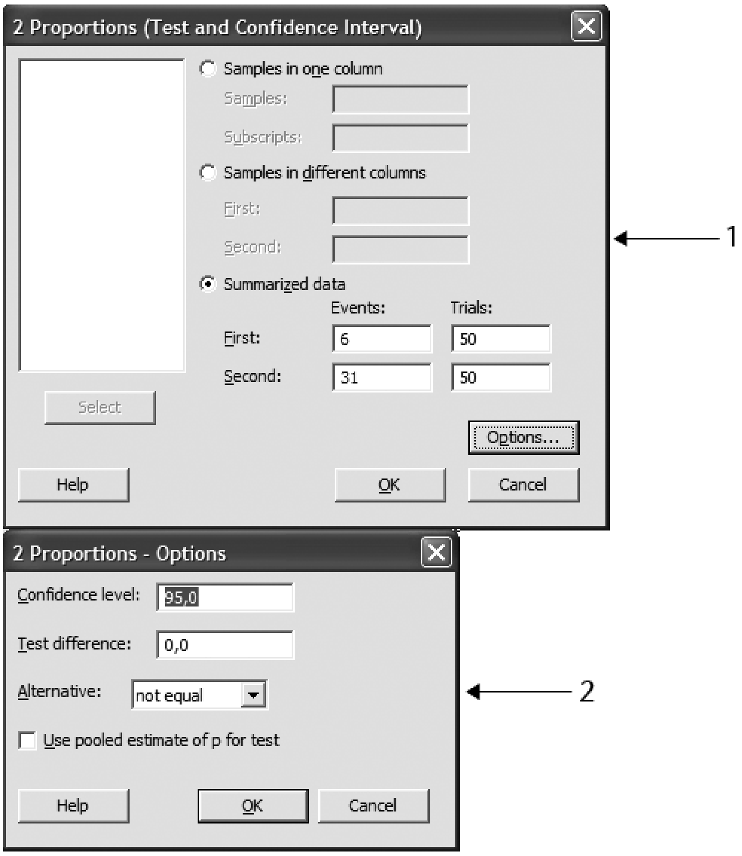
Plantearemos el siguiente contraste:

$$H_0 : p_A - p_B = 0$$
$$H_1 : p_A - p_B \neq 0$$

Fijamos $\alpha = 0{,}05$.

La figura 18 muestra los pasos a seguir para realizar el contraste de la diferencia de proporciones.

Figura 18. Pasos para hacer un contraste de hipótesis para la diferencia de proporciones



Pasos a seguir

Se sigue la ruta ***Stat*** > ***Basic Statistics*** > ***2-Proportions*** y se rellenan los campos en la ventana correspondiente.

Summarized data (1)

First: *Events:* 6 *Trials*: 50

Second: *Events*: 31 *Trials*: 50

Seleccione ***Options (2)*** y se rellenan los campos:

Confidence level: 95,0

Alternative: not equal

Los resultados de la figura 19 muestran el *p*-valor = 0,000 < 0,05. Esto indica que podemos rechazar la hipótesis nula y concluimos que la proporción de peticiones con tiempo de espera aceptable es diferente para los dos servidores.

Figura 19. Resultados del contraste de diferencia de proporciones. Minitab

```
Test and CI for Two Proportions

Sample   X   N  Sample p
1        6  50  0,120000
2       31  50  0,620000

Difference = p (1) - p (2)
Estimate for difference:  -0,5
95% CI for difference:  (-0,661908; -0,338092)
Test for difference = 0 (vs not = 0):  Z = -6,05
P-Value = 0,000
```

1.3. Contrastes de hipótesis de comparación de varianzas

Uno de los contrastes desarrollados en el apartado 1.1 para la comparación de medias poblacionales depende del supuesto de igualdad de las dos varianzas poblacionales. Aunque en muchas aplicaciones prácticas este es un supuesto razonable, conviene usar los datos disponibles para contrastar su validez.

En este apartado consideramos el caso de dos muestras aleatorias independientes de poblaciones normales y contrastaremos la igualdad de varianzas poblacionales.

Sea s_1^2 la varianza muestral de una muestra de n_1 observaciones de una población normal con varianza σ_1^2, y s_2^2 la varianza muestral de una muestra independiente de n_2 observaciones de una población normal con varianza σ_2^2. Siempre que las dos varianzas poblacionales sean iguales ($\sigma_1^2 = \sigma_2^2$). La distribución de la relación de las dos varianzas de las muestras s_1^2 / s_2^2 está definida por el estadístico F que sigue una **distribución *F* de Snedecor** con $n_1 - 1$ grados de libertad para el numerador y $n_2 - 1$ grados de libertad para el denominador,

$$F_{n_1-1;n_2-1} = \frac{s_1^2}{s_2^2}$$

Contraste de igualdad de varianzas de dos poblaciones normales

Ahora nos interesa contrastar la hipótesis nula que asegura que las varianzas de las poblaciones son iguales $\sigma_1^2 = \sigma_2^2$, es decir, la varianza de la población 1 es igual a la varianza de la población 2. Primero fijaremos el nivel de significación α del contraste.

Nota

A veces, en lugar de:

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

escribiremos:

H_0: $\sigma_1^2 / \sigma_2^2 = 1$

Hipótesis alternativa, puede ser:

- Bilateral: $\boldsymbol{H_1}$: $\sigma_1^2 \neq \sigma_2^2$, las varianzas de las dos poblaciones son distintas.
- Unilateral: $\boldsymbol{H_1}$: $\sigma_1^2 > \sigma_2^2$, la varianza de la población 1 es mayor que la varianza de la población 2.
- Unilateral: $\boldsymbol{H_1}$: $\sigma_1^2 < \sigma_2^2$, la varianza de la población 1 es menor que la varianza de la población 2.

Bajo el supuesto de la hipótesis nula cierta H_0: $\sigma_1^2 = \sigma_2^2$, el estadístico de contraste es:

$$F^*_{n_1-1;n_2-1} = \frac{s_1^2}{s_2^2}$$

Regla de decisión del contraste de hipótesis

Una vez que se ha calculado el valor del estadístico de contraste, se debe determinar el p-valor. El p-valor depende de la hipótesis alternativa planteada.

- Si H_1: $\sigma_1^2 \neq \sigma_2^2$, entonces $p\text{-valor} = 2P\left(F_{n_1-1,n_2-1} > F^*\right)$

- Si H_1: $\sigma_1^2 < \sigma_2^2$, entonces p-valor $= P\left(F_{n_1-1,n_2-1} < F^*\right)$
- Si H_1: $\sigma_1^2 > \sigma_2^2$, entonces p-valor $= P\left(F_{n_1-1,n_2 1} > F^*\right)$
- Si p-valor $\leq \alpha$ se rechaza H_0

Ejemplo 4. "Variabilidad de procesadores de texto".

Queremos comparar dos tipos de procesadores de textos: el LaTeX y el OpenOffice. Para hacerlo, consideramos textos más o menos de la misma longitud y contamos la variabilidad del espacio que deja cada procesador entre las palabras. En el caso del LaTeX, consideramos diez textos y obtenemos que la desviación estándar muestral del espacio que deja es de 2,5 mm, mientras que para el OpenOffice consideramos quince textos y obtenemos que la desviación estándar muestral del espacio que deja es de 3,5 mm. Suponiendo normalidad, ¿podemos afirmar que los dos procesadores de textos tienen la misma variabilidad en el espacio que dejan entre palabras?

Para contestar a la pregunta hemos de realizar un contraste de igualdad de varianzas.

El contraste de hipótesis es: $\begin{array}{l} H_0 : \sigma^2_{LaTeX} = \sigma^2_{OpenOffice} \\ H_1 : \sigma^2_{LaTeX} \neq \sigma^2_{OpenOffice} \end{array}$

Fijamos el valor de $\alpha = 0{,}05$.

El estadístico de contraste vale: $F^* = \dfrac{s^2_{LaTeX}}{s^2_{OpenOffice}}$. Los valores de s^2_{LaTeX} y $s^2_{OpenOffice}$ son, respectivamente, $s^2_{LaTeX} = 6{,}25$ y $s^2_{OpenOffice} = 12{,}25$.

El estadístico F sigue la distribución F de Fisher-Snedecor con 9 y 14 grados de libertad. El valor del estadístico de contraste será:

$$F^* = \frac{6{,}25}{12{,}25} \approx 0{,}51.$$

Los valores críticos serán:
$F_{1-\alpha/2,9,14} = F_{0{,}975,9,14} \simeq 0{,}265$ y $F_{\alpha/2} = F_{0{,}025,9,14} \simeq 3{,}21$.

Para calcular los valores críticos utilizaremos la tabla F o mediante un software estadístico.

Como $F_{0,0,25,9,14} < F^* < F_{0,975,9,14}$ aceptamos la hipótesis nula y concluimos que las varianzas son iguales. Luego los dos procesadores tienen la misma variabilidad en el espacio que dejan entre palabras. Si quisiéramos realizar el contraste con el *p*-valor, éste valdría: $p = 2 \cdot p(F_{9,14} < 0,51) \approx 0,312$. Como es mucho mayor que 0,05, aceptamos la hipótesis nula y llegamos a la misma conclusión.

En el **ejemplo 1. "Comparación de las medias del tiempo de respuesta de dos servidores"**, cuando realizamos el contraste de diferencia de medias con Minitab **no** presupusimos que las varianzas fueran iguales. Ahora realizaremos un contraste para comparar las dos varianzas y ver si son iguales.

Las hipótesis nula y alternativa son: $\begin{array}{l} H_0 : \sigma_A^2 = \sigma_B^2 \\ H_1 : \sigma_A^2 \neq \sigma_B^2 \end{array}$

Fijamos α = 0,1. El estadístico de contraste es: $F^* = \dfrac{s_A^2}{s_B^2}$, donde s_A^2 y s_B^2 son respectivamente, las varianzas de los tiempos de espera de los servidores *A* y *B*. La distribución de *F* es la de la *F* de Snedecor con 50 – 1 = 49 grados de libertad en el numerador y 50 – 1 = 49 grados de libertad en el denominador.

Se resolverá el problema con Minitab. Los resultados de Minitab se muestran en la figura 20.

Figura 20. Resultados del contraste de varianzas. Minitab

```
Test for Equal Variances: TA; TB

90%   Bonferroni   confidence   intervals   for   standard
deviations

     N    Lower     StDev    Upper
TA  50  0,75196  0,90019  1.12176
TB  50  1.45954  1.74726  2.17731

F-Test (Normal Distribution)
Test statistic = 0,27; p-value = 0,000

Levene's Test (Any Continuous Distribution)
Test statistic = 13.14; p-value = 0,000
```

Pasos a seguir

Se sigue la ruta ***Stat*** > ***Basic Statistics*** >***2-Variances*** y se rellenan los campos en la ventana correspondiente.

En el cuadro de dialogo se completan los campos:

Samples in different colums:

First: TA

Second: TB

Seleccionad ***Options***, completad los campos:

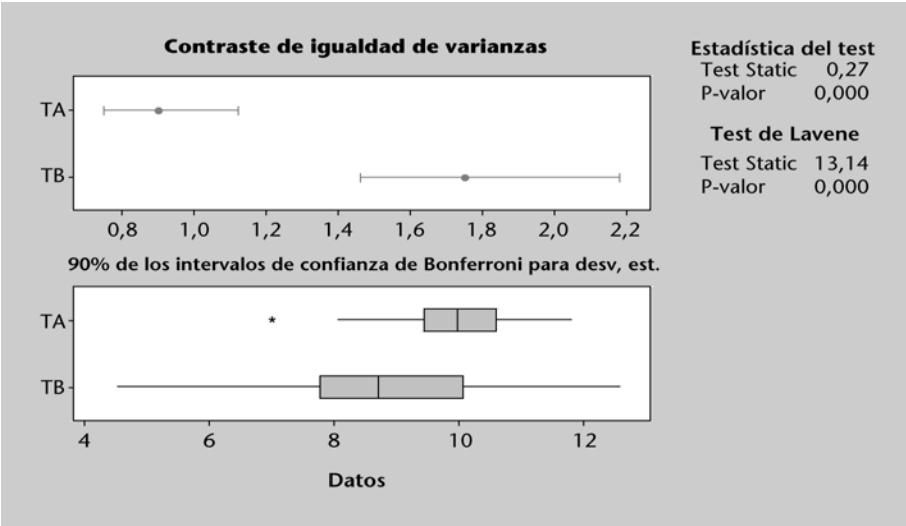
Confidence level: 90,0

Title: Contraste de igualdad de varianzas

Podemos ver que como el *p*-valor es prácticamente cero, hemos de rechazar la hipótesis nula, es decir, no podemos considerar que las varianzas sean iguales. Por esa razón cuando hicimos el contraste de diferencia de medias **no** asumimos que las varianzas fueran iguales.

Minitab también nos ha proporcionado el siguiente gráfico para el contraste de igualdad de varianzas:

Figura 21. Resultados del contraste de igualdad de varianzas. Minitab



La figura 21 presenta un gráfico con los intervalos de confianza de las varianzas de las dos poblaciones, se observa que los intervalos son distintos y no se solapan. El *p*-valor del test *F* indica que se rechaza la hipótesis de igualdad de varianzas.

En el gráfico de *boxplot* se ve claramente que la variabilidad del tiempo de espera del servidor *A* es mucho más pequeña que la del servidor *B*.

2. Comparación de grupos mediante ANOVA

En el apartado anterior se presentaron algunos de los contrastes de hipótesis que se usan habitualmente para determinar si existen diferencias significativas entre dos poblaciones o grupos de individuos. En ocasiones, sin embargo, se deseará comparar más de dos poblaciones o grupos entre sí, para lo cual se emplearán las técnicas de ***analysis of variance*** **(ANOVA)** que se introducen en este apartado.

Nota

El acrónimo **ANOVA** viene del término *analysis of variance* (análisis de la variación existente entre las distintas medias consideradas, para ver si existen diferencias significativas entre las mismas).

Así, por ejemplo, las técnicas ANOVA se podrían aplicar para dar respuestas a preguntas como las siguientes:

- ¿Existen diferencias significativas entre la duración media de los juicios según el tipo de delito cometido (homicidio, abuso sexual, robo, piratería, fraude fiscal, etc.)?

- ¿Existen diferencias significativas entre el gasto anual promedio en tecnología según la franja de edad a la que pertenezca el individuo (niño, joven, adulto, anciano)?

- ¿Existen diferencias significativas entre el número medio de alumnos y ordenadores por centro escolar entre los diferentes países de la eurozona?

- ¿Existen diferencias significativas entre el número medio de autocitas a revistas científicas según la editorial (Elsevier, Inderscience, Taylor & Francis, IGI Global, etc.)?

- ¿Existen diferencias significativas entre el consumo medio de combustible según el modelo de automóvil usado (deportivo, turismo, todoterreno, monovolumen, etc.)?

- ¿Existen diferencias significativas entre la calidad media (medida a partir de unos parámetros definidos) de los resultados de búsquedas en línea según el tipo de motor usado (Google, Microsoft Bing, Yahoo!, etc.)? (figura 22)

Observad

Los ejemplos que se presentan en este capítulo se caracterizan porque la pertenencia a una población o a otra depende de un único factor (tipo de delito, franja de edad, país, editorial, modelo de automóvil, motor de búsqueda, etc.). En estos casos, se usa ANOVA de un único factor (en inglés *one-way* ANOVA o *single-factor* ANOVA). Sin embargo, existen también técnicas ANOVA para el caso en que los grupos vengan determinados por dos factores (p. ej.: tipo de delito y solvencia económica del acusado, franja de edad y clase social, etc.).

Figura 22. ANOVA permite comparar la calidad media de diferentes servicios

2.1. Comparaciones de varias medias

Cuando se desean comparar entre sí las medias correspondientes a más de dos poblaciones o grupos de individuos, se podría pensar en comparar dichas medias dos a dos mediante un contraste de hipótesis para dos poblaciones. Así, por ejemplo, en el caso de tres poblaciones se podría pensar en realizar una serie de tests *t* de hipótesis para comparar las distintas medias entre sí: un primer test *t* para comparar las medias de las poblaciones 1 y 2, otro para comparar las medias de las poblaciones 1 y 3, y otro para comparar las medias de las poblaciones 2 y 3. Sin embargo, esta aproximación tiene un grave problema: si para cada test *t* se usa un nivel de significación α (generalmente se usa $\alpha = 0{,}05$), entonces la probabilidad de cometer un **error de tipo I** es α en cada test; en tales condiciones, se puede comprobar que la probabilidad de cometer un error de tipo I en el global de los tres tests sería de $1-(1-\alpha)^3$ (si $\alpha = 0{,}05$ dicha probabilidad sería de, aproximadamente, 0,14). En otras palabras, comparando las medias dos a dos se está realizando un test global con una probabilidad de error de tipo I mucho mayor que la prevista inicialmente para cada test individual. Para evitar este problema se pueden usar las técnicas ANOVA, que permiten realizar un único test global con una probabilidad de error de tipo I determinada (generalmente $\alpha = 0{,}05$).

Recordad

Un **error de tipo I** consiste en rechazar la hipótesis nula cuando resulta que ésta es cierta. En este caso, la hipótesis nula sería que las medias son coincidentes.

El test *F* de ANOVA

A fin de comparar las medias correspondientes a *k* poblaciones o grupos de individuos distintos ($k \geq 3$), se puede plantear el siguiente contraste de hipótesis, donde el símbolo μ_i representa la media de la población *i*-ésima para $i = 1, 2, \ldots, k$:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \text{ (todas las medias son iguales)} \\ H_a : \text{no todas las medias son iguales} \end{cases} \tag{1}$$

En otras palabras, la hipótesis nula, H_0, sostiene que no hay diferencias significativas entre las distintas medias poblacionales, mientras que la hipótesis alternativa, H_a, sostiene todo lo contrario, p. ej.: que las medias sí son significativamente distintas. Es importante observar aquí que la hipótesis nula no dice que todas las medias sean significativamente distintas entre sí, sino simplemente que no todas las medias son iguales, aunque puede haber algunas de ellas que sí lo sean (podría ocurrir, por ejemplo, que $\mu_1 \neq \mu_2$ y $\mu_2 \neq \mu_3$ pero siendo $\mu_1 = \mu_3$). Por tanto, si se concluyese que no todas las medias son iguales, cabría realizar un análisis posterior para determinar cuáles de ellas son diferentes entre sí.

Software estadístico

En la actualidad existe una gran variedad de **programas estadísticos** o de análisis de datos de gran calidad, tanto comerciales (Minitab, SPSS, MS Excel, SAS, S-Plus, etc.) como de código abierto (R, Calc de Open Office, etc.).

El contraste de hipótesis (1) se llama test *F* de ANOVA, y generalmente se recurre al uso de algún ***software* estadístico** para resolverlo, es decir, para obtener el *p*-valor asociado al test. A partir de dicho *p*-valor corresponde al investigador determinar si ha sido posible encontrar suficientes evidencias para rechazar la hipótesis nula o si, por el contrario, los datos empíricos parecen no estar en contradicción con la hipótesis nula y, por tanto, se acepta ésta como válida. Como en cualquier otro tipo de contraste estadístico, antes de resolver el test se suele fijar un valor de significación, α (por lo general $\alpha = 0{,}05$ o bien $\alpha = 0{,}01$). Una vez obtenido el *p*-valor, si $p - valor < \alpha$ se rechaza la hipótesis nula; en caso contrario no hay indicios suficientes para hacerlo y, por tanto, se aceptará la hipótesis nula como válida. La elección del valor concreto para α dependerá del nivel de confianza, $1 - \alpha$, que se desee que tenga la decisión final sobre la aceptación o no de la hipótesis nula. Así, por ejemplo, un $\alpha = 0{,}05$ implicará un nivel de confianza en la decisión final del 95%, mientras que un $\alpha = 0{,}01$ implicará un nivel de confianza en la decisión final del 99%. El problema de seleccionar niveles de confianza excesivamente elevados (superiores al 99% o, lo que es lo mismo, valores de α inferiores a 0,01) es que entonces el contraste de hipótesis se vuelve excesivamente "conservador", de manera que sólo cuando las evidencias empíricas en contra de la hipótesis nula son totalmente abrumadoras (es decir, sólo cuando las diferencias entre algunas de las medias son desproporcionadas), es posible obtener un *p*-valor más pequeño o igual que α. Por ese motivo, en la mayoría de los casos prácticos se suele usar el valor $\alpha = 0{,}05$ o bien $\alpha = 0{,}01$.

Ejemplo de aplicación de ANOVA: comparando el número medio de accesos a contenidos en línea según la posición del enlace en el portal

En un portal web de acceso a publicaciones en línea, se sospecha que la posición que ocupa el enlace a una determinada base de datos afecta al número de consultas diarias que ésta recibe. Para comprobarlo, se han seleccionado al azar un total de 13 días laborables de un mes y, para cada uno de ellos, se ha contabilizado el número de accesos recibidos. La tabla 4 muestra los valores obtenidos, los cuales han sido agrupados según la posición diaria del enlace (en el encabezado de la página, en el margen derecho o en el margen izquierdo).

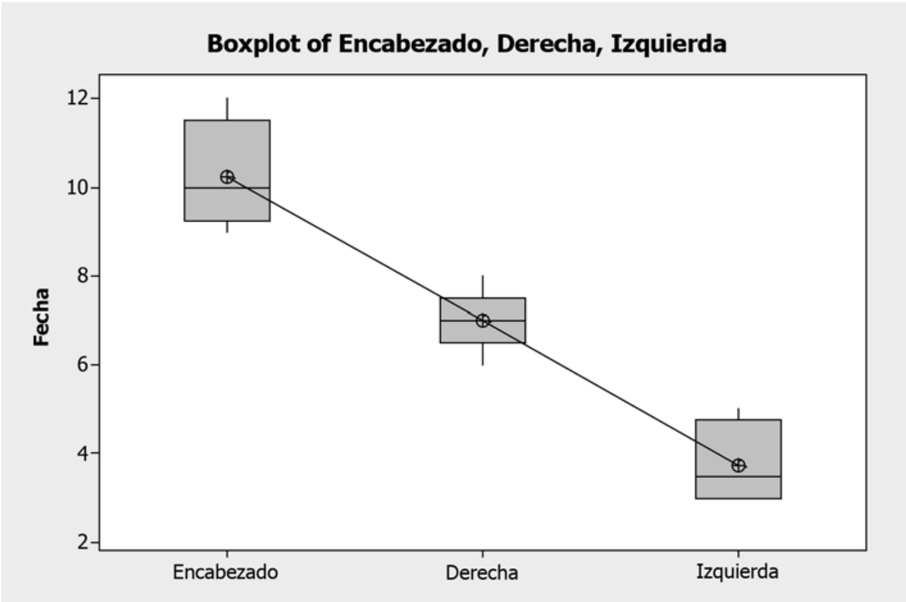
Tabla 4. Accesos a una base de datos según la posición del enlace

	Posición del enlace		
	Encabezado (1)	**Derecha (2)**	**Izquierda (3)**
	10	7	3
	12	6	3
	10	7	5
	9	8	4
		7	
Total	41	45	15
Media	$\bar{x}_1 = 10{,}25$	$\bar{x}_2 = 7{,}0$	$\bar{x}_3 = 3{,}75$

¿Se puede afirmar que hay diferencias significativas entre las distintas medias?, es decir: ¿depende el número medio de consultas diarias de la posición que ocupe el enlace?

Como primera aproximación a este problema, se puede optar por generar un diagrama de cajas y bigotes (*boxplot*) para cada uno de los grupos de datos. La figura 23 muestra dicho diagrama que incluye además una línea uniendo las respectivas medias. Se aprecian claras diferencias entre los tres grupos considerados, tanto a nivel de *boxplots* como a nivel de las respectivas medias.

Figura 23. *Boxplot* del número de consultas para cada posición



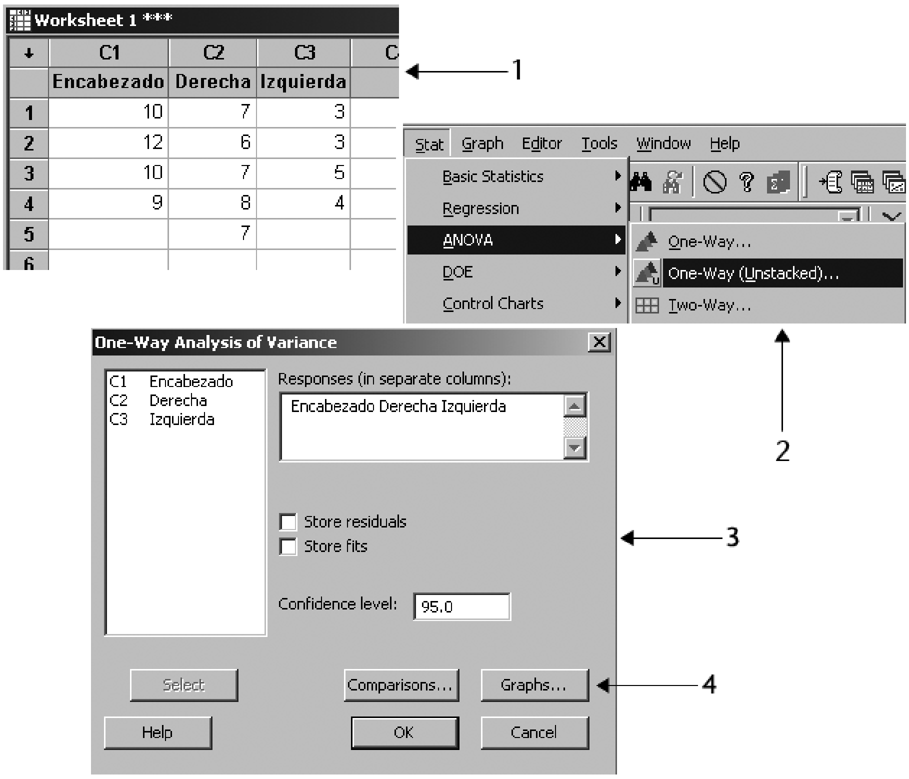
Sin embargo, para contestar de forma contundente a las preguntas anteriores, resulta necesario realizar un test *F* de ANOVA. El contraste de hipótesis se puede formular como sigue:

$$\begin{cases} H_0 : \bar{x}_1 = \bar{x}_2 = \bar{x}_3 \\ H_a : \text{no todas las medias son iguales} \end{cases}$$

Para resolver dicho contraste, se fijará un valor de significación $\alpha = 0{,}05$ y se recurrirá al uso de software estadístico para obtener el *p*-valor correspondiente a las observaciones de la tabla 4.

La figura 24 muestra los pasos básicos necesarios para generar un análisis ANOVA con el **programa Minitab**. Por su parte, la figura 24 muestra el *output* generado para los datos de este ejemplo. Se observa que el valor resultante para el estadístico del contraste es $F = 44{,}47$. El estadístico *F* es una variable aleatoria que se comporta según una distribución *F*-Snedecor con 2 grados de libertad en el numerador (*DF Factor*) y 10 grados de libertad en el denominador (*DF Error*). El *p*-valor no es más que la probabilidad de que una variable aleatoria con esas características supere el valor observado para el estadístico de contraste, p. ej.: *p*-valor = $P(F_{2,10} > 44{,}47)$. Según se observa en el *output*, en este caso se obtiene *p*-valor = 0,000. Dado que el *p*-valor es mucho menor que el nivel de significación escogido (*p*-valor = $0{,}000 < 0{,}05 = \alpha$), se concluye que los datos obtenidos parecen contradecir la hipótesis nula y, por tanto, ésta se debe rechazar. Así pues, hay indicios claros para pensar que no todas las medias son iguales, p. ej.: que el número medio de consultas diarias sí depende de la posición que ocupe el enlace.

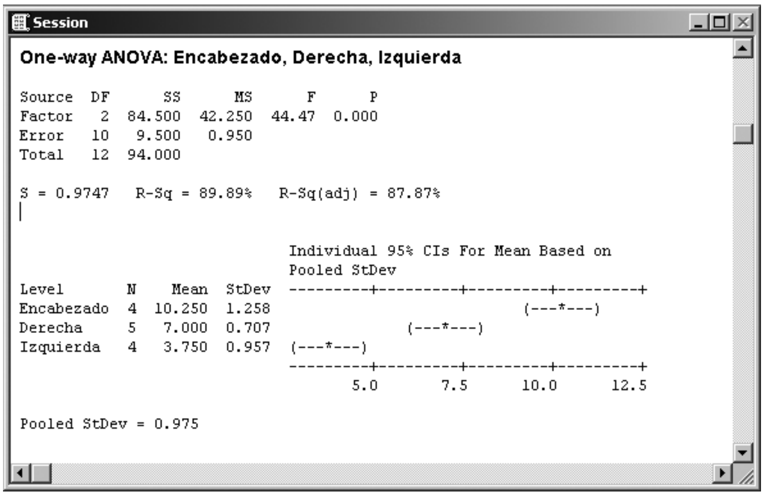
Figura 24. Pasos a seguir para realizar un análisis ANOVA en Minitab



Pasos a seguir

Una vez introducidos los datos en el programa ***(1)***, se sigue la ruta ***Stat*** > ***ANOVA*** > ***One-Way (Unstacked***) ***(2)*** y se seleccionan las variables en la ventana de ANOVA ***(3)***. Más adelante se hará uso de la opción ***Graphs*** de esta ventana ***(4)***.

Figura 25. *Output* ANOVA de Minitab para la comparativa de posiciones



En la segunda parte del *output* Minitab se representa cada una de las medias junto con su respectivo intervalo de confianza para un nivel de confianza del 95%. Se observa que los intervalos son disjuntos (no se solapan), lo que significa que las observaciones aportan evidencias de que las tres medias son significativamente distintas. En general, sin embargo, el hecho de que todas las medias no sean iguales no implicará necesariamente que todas sean distintas (es decir, podría haber intervalos que se solapasen y otros que no).

La figura 26 muestra el correspondiente *output* ANOVA que ofrece **Microsoft Excel**. Se observa el mismo valor para el estadístico F = 44,47, así como un p-valor = 1,0543E-05 (es decir, p-valor = 0,00001543 o, redondeando, p-valor = 0,000).

Figura 26. *Output* ANOVA de Excel para la comparativa de posiciones

	A	B	C	D	E	F	G
1	Análisis de varianza de un factor						
2							
3	RESUMEN						
4	*Grupos*	*Cuenta*	*Suma*	*Promedio*	*Varianza*		
5	Encabezado	4	41	10,25	1,58333333		
6	Derecha	5	35	7	0,5		
7	Izquierda	4	15	3,75	0,91666667		
8							
9							
10	ANÁLISIS DE VARIANZA						
11	*Origen de las variaciones*	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
12	Entre grupos	84,5	2	42,25	44,4736842	1,0543E-05	4,102821015
13	Dentro de los grupos	9,5	10	0,95			
14							
15	Total	94	12				

Nota

Para poder realizar ANOVA con **MS Excel**, es necesario instalar previamente un complemento llamado "Análisis de datos". Usando Google o cualquier otro buscador es fácil encontrar información detallada sobre el proceso de instalación. También existe un complemento similar para Open Office Calc.

Ejemplo de aplicación de ANOVA: comparando promedios de resultados válidos ofrecidos por un motor de búsqueda según el algoritmo empleado

Los desarrolladores de un nuevo motor de búsqueda especializado en recursos de investigación están probando tres algoritmos distintos de recuperación de la información. Para comprobar si el promedio de resultados válidos que proporciona cada algoritmo es el mismo en los tres casos, se han realizado unas pruebas aleatorias con cada uno de ellos. La tabla 5 muestra las observaciones que se han obtenido tras realizar las pruebas.

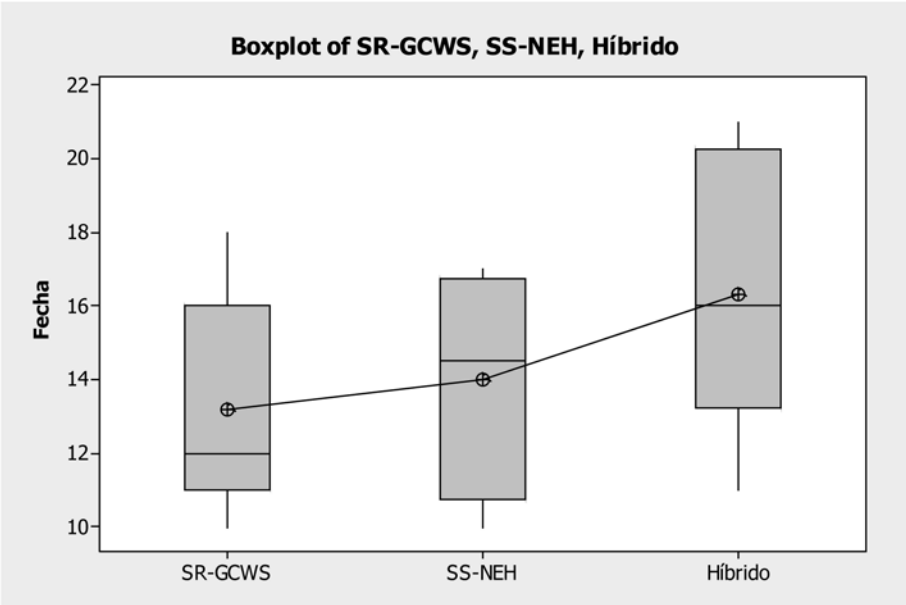
Tabla 5. Resultados válidos obtenidos con cada algoritmo

	Algoritmo		
	SR-GCWS	**SS-NEH**	**Híbrido**
	12	10	16
	10	17	14
	18	16	16
	12	13	11
	14		20
			21
Total	66	56	98
Media	$\bar{x}_1$ = 13,20	$\bar{x}_2$ = 14,00	$\bar{x}_3$ = 16,33

¿Se puede afirmar que hay diferencias significativas entre los distintos promedios?, es decir: ¿depende el promedio de resultados válidos obtenidos del algoritmo que implemente el motor de búsqueda?

Nuevamente, para responder adecuadamente a estas preguntas resulta necesario llevar a cabo un test F de ANOVA. Como paso previo, sin embargo, podemos graficar los correspondientes *boxplots*. Como se observa en la figura 27, en este caso las diferencias entre los distintos grupos no parecen ser excesivas, si bien el algoritmo híbrido parece haber proporcionado resultados ligeramente superiores al resto.

Figura 27. *Boxplot* del número de resultados válidos para cada algoritmo

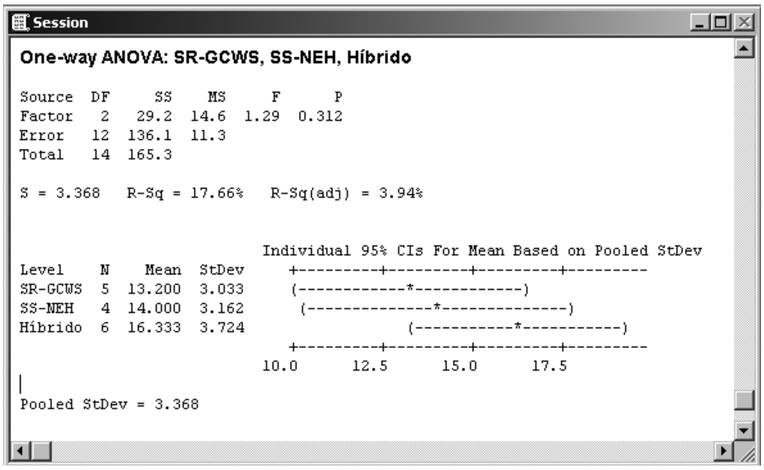


A fin de comprobar si las diferencias entre los promedios son o no estadísticamente significativas, se formula el siguiente contraste ANOVA:

$$\begin{cases} H_0 : \overline{x}_1 = \overline{x}_2 = \overline{x}_3 \\ H_a : \text{no todas las medias son iguales} \end{cases}$$

Nuevamente se hará uso de un nivel de significación $\alpha = 0{,}05$ (es decir, el nivel de confianza usado es del 95%). Las figuras 28 y 29 muestran, respectivamente, los *output* Minitab y Excel para este ejemplo.

Figura 28. *Output* ANOVA de Minitab para la comparativa de algoritmos



Session

One-way ANOVA: SR-GCWS, SS-NEH, Híbrido

```
Source  DF     SS    MS     F      P
Factor   2   29.2  14.6  1.29  0.312
Error   12  136.1  11.3
Total   14  165.3

S = 3.368   R-Sq = 17.66%   R-Sq(adj) = 3.94%


                             Individual 95% CIs For Mean Based on Pooled StDev
Level    N    Mean  StDev    +---------+---------+---------+---------
SR-GCWS  5  13.200  3.033    (------------*------------)
SS-NEH   4  14.000  3.162     (--------------*--------------)
Híbrido  6  16.333  3.724              (-----------*-----------)
                             +---------+---------+---------+---------
                          10.0      12.5      15.0      17.5

Pooled StDev = 3.368
```

Figura 29. *Output* ANOVA de Excel para la comparativa de algoritmos

	A	B	C	D	E	F
1	Análisis de varianza de un factor					
2						
3	RESUMEN					
4	*Grupos*	*Cuenta*	*Suma*	*Promedio*	*Varianza*	
5	SR-GCWS	5	66	13,2	9,2	
6	SS-NEH	4	56	14	10	
7	Híbrido	6	98	16,33333333	13,8666667	
8						
9						
10	ANÁLISIS DE VARIANZA					
11	*Origen de las variaciones*	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*
12	Entre grupos	29,2	2	14,6	1,28697356	0,311619296
13	Dentro de los grupos	136,1333333	12	11,34444444		
14						
15	Total	165,3333333	14			
16						

En ambos *outputs* se observa un valor del estadístico F = 1,29. En esta ocasión, dicho estadístico es una variable aleatoria que se distribuye según una F-Snedecor con 2 grados de libertad en el numerador (*DF Factor*) y 12 en el denominador (*DF Error*). La probabilidad de que una variable como esta alcance o supere el valor 1,29 obtenido por el estadístico es de 0,312, que es precisamente el p-valor que se observa en ambos *outputs*. Puesto que p-valor = 0,312 > α = 0,05, no parece que haya indicios suficientes como para rechazar la hipótesis nula. En otras palabras, los datos observados parecen estar en sintonía con la hipótesis nula, por lo que aceptaremos la hipótesis de que los promedios de resultados válidos son equivalentes para los tres algoritmos, sin que haya diferencias estadísticamente significativas entre ellos.

De hecho, en la segunda parte del *output* Minitab se observa que los intervalos de confianza para las tres medias se solapan parcialmente, lo que significa que para un nivel de confianza del 95% no se puede afirmar que haya diferencias significativas entre dichas medias.

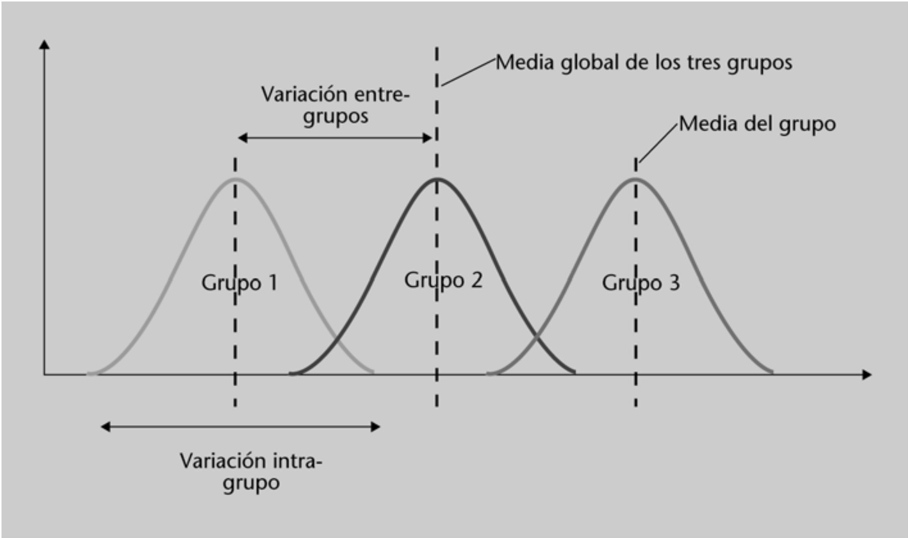
2.2. La lógica del contraste ANOVA

Cuando mediante un experimento aleatorio se recogen una serie de datos (observaciones) y estos son clasificados en varios grupos o niveles según un factor determinado (franja de edad, clase social, etc.), se pueden analizar dos tipos distintos de varianza en las observaciones (figura 30):

- Por un lado, la variación existente entre los distintos grupos o niveles (p.ej.: la variación entre las respectivas medias de cada grupo). Esta se conoce como “variación entre-grupos” o “MS Factor”.

- Por otro, la variación existente dentro de cada grupo o nivel. Esta se conoce como “variación intra-grupos” o “MS Error”.

Figura 30. Variación entre-grupos y variación intra-grupos



En el fondo, lo que hace el test ANOVA es comparar las dos medidas de variabilidad, la variación entre-grupos (MS Factor) y la variación intra-grupos (MS Error). Si ocurre que el MS Factor es significativamente mayor que el MS Error (figura 31), entonces el test concluirá que las medias de los distintos grupos no son iguales en todos los casos (lo que implica que no todos los datos pertenecen a un mismo grupo o, lo que es lo mismo, que el valor de las observaciones sí depende del factor considerado). Si, por el contrario, el MS Factor no es significativamente mayor que el MS Error (figura 32), entonces el test concluirá que no se aprecian diferencias significativas entre las medias de los distintos grupos (en otras palabras, que las observaciones parecen proceder todas de un único grupo o, lo que es lo mismo, que las observaciones no parecen depender del factor considerado).

Figura 31. La variación entre-grupos es mayor que la intra-grupos

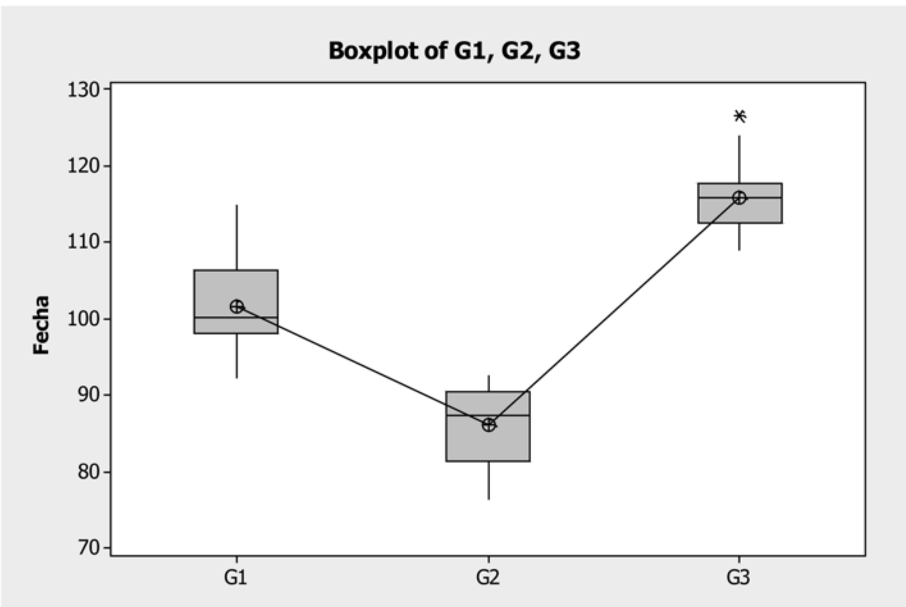
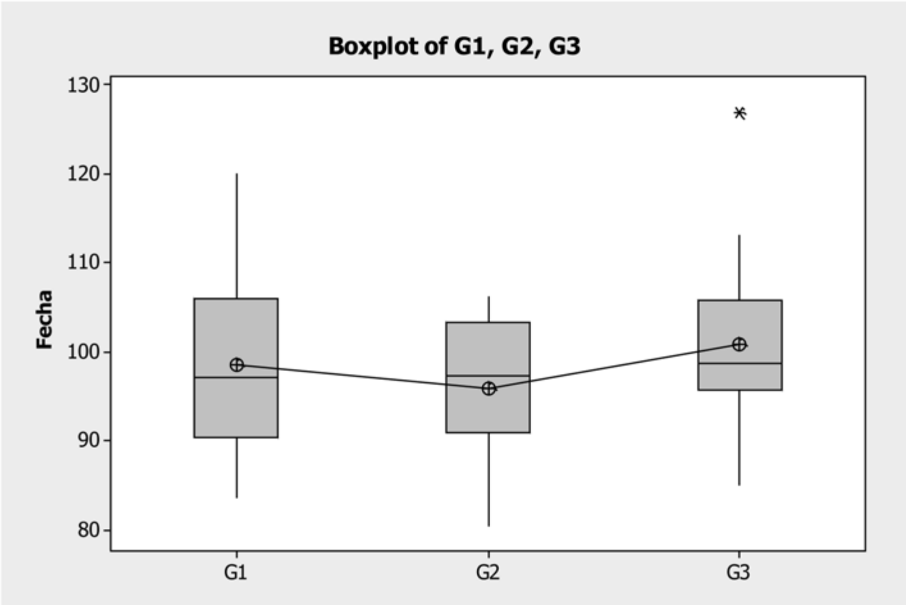


Figura 32. La variación entre-grupos es menor que la intra-grupos



En la figura 25 (*output* ANOVA de Minitab) se observan los valores MS Factor = 42,250 y MS Error = 0,950. Es decir, en este caso la variación entre-grupos (MS Factor) es mucho mayor que la variación intra-grupos (MS Error), lo que ya deja entrever que, probablemente, el test concluya que no todas las medias son iguales. Pero, ¿cómo llega el test a la conclusión final? La figura 33 ayuda a entender mejor cómo funciona el test F de ANOVA:

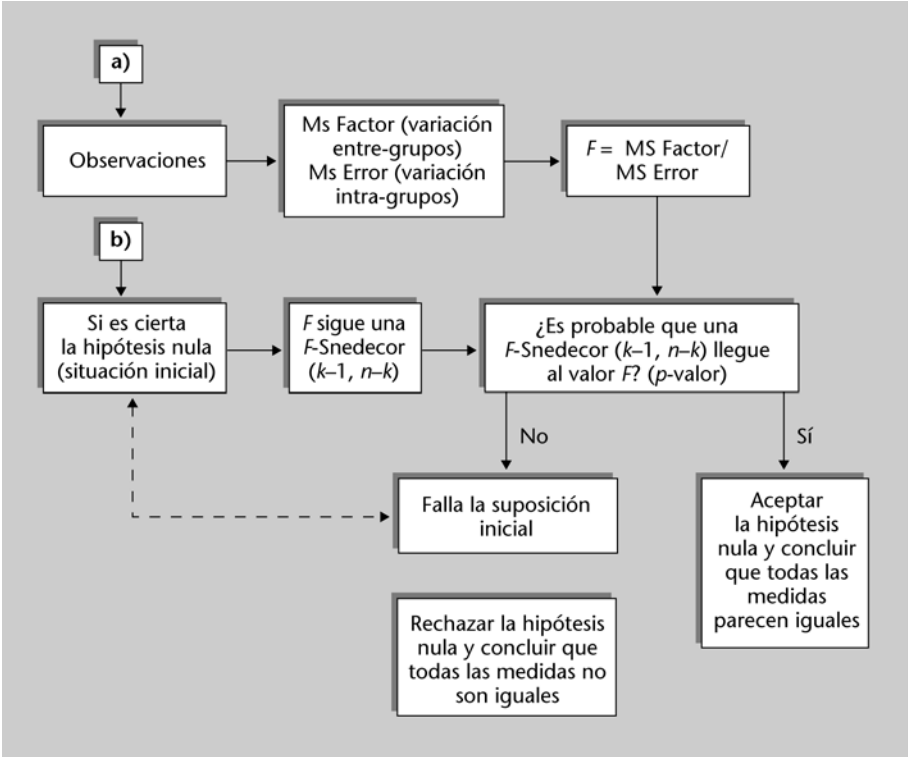
a) Por un lado, a partir de los valores obtenidos para MS Factor y MS Error se calcula el estadístico de contraste $F = \frac{MS\ Factor}{MS\ Error}$.

En este caso, $F = 44{,}47$.

b) Por otra parte, se sabe que si la hipótesis nula fuese cierta (p. ej.: si todas las medias son iguales), este estadístico F sería una variable aleatoria que seguiría una distribución F-Snedecor con $k - 1$ grados de libertad en el numerador (DF Factor), y $n - k$ grados de libertad en el denominador (DF Error), siendo k el número de grupos o niveles y n el número total de observaciones.

En el ejemplo de la figura 25, DF Factor = 2 y DF Error = 10. Ahora bien, ¿cuál es la probabilidad de que una variable aleatoria F-Snedecor (2, 10) alcance un valor como el obtenido por el estadístico de contraste F? En otras palabras, ¿es razonable pensar que una F-Snedecor (2,10) haya alcanzado un valor de 44,47? La probabilidad de que esto ocurra nos la proporciona el p-valor. De esta manera, un p-valor "pequeño" (inferior al nivel de significación α) se puede interpretar como una probabilidad demasiado baja de que una F-Snedecor (2, 10) pueda dar el valor obtenido para F, lo que pone en entredicho la suposición inicial de que la hipótesis nula era cierta. Por otra parte, un p-valor "grande" (superior al nivel de significación α) se puede interpretar como una probabilidad aceptable de que, en efecto, una *F*-Snedecor (2, 10) tome dicho valor y, por tanto, no habría evidencias para dudar de la hipótesis nula.

Figura 33. Funcionamiento interno del test F de ANOVA



Observad

que cuando el valor obtenido para el estadístico *F* a partir de las observaciones no es coherente con lo que cabría esperar de una *F*-Snedecor (k – 1, n – k), entonces lo que está fallando es la suposición inicial de que la hipótesis nula es cierta.

2.3. Las hipótesis del modelo ANOVA

Como cualquier otra técnica de inferencia estadística, el contraste ANOVA se puede usar con garantías, para comparar poblaciones o grupos, sólo si se cumplen unas determinadas condiciones de entorno o supuestos básicos:

1) Las observaciones son independientes entre sí y constituyen, para cada población o grupo, una muestra aleatoria.

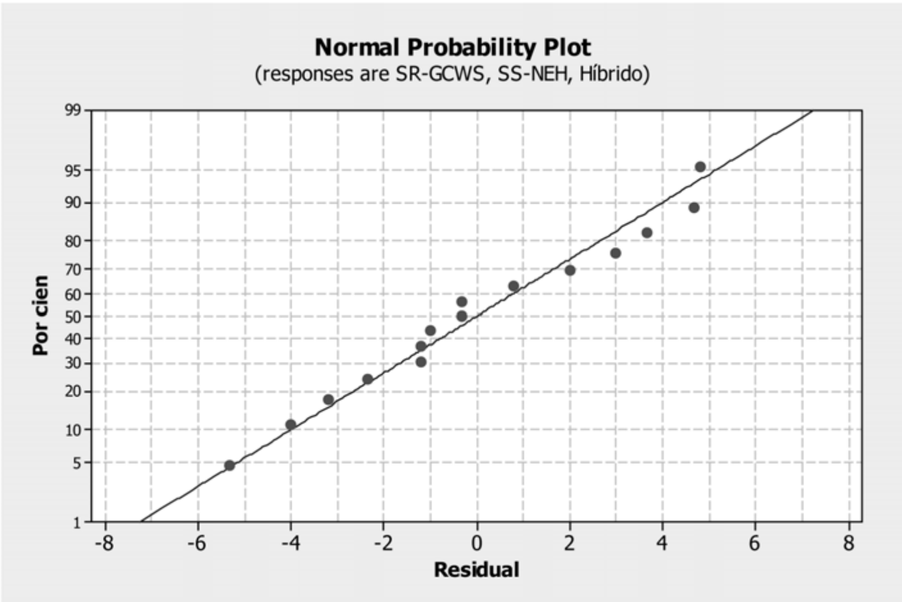
2) Las observaciones de cada población o grupo siguen una distribución aproximadamente normal.

3) Las observaciones de cada población o grupo tienen una varianza σ^2, que es aproximadamente la misma para todos los grupos.

El primer supuesto garantiza que las muestras son aleatorias e independientes, lo que es un requisito común en las técnicas de inferencia estadística. Si las muestras no fuesen aleatorias o las observaciones no fuesen independientes, la información que se generaría estaría sesgada y, por tanto, no sería válida. Es función del investigador garantizar, durante la fase de diseño del experimento y posterior recogida de datos, que se cumple este supuesto.

Por lo que respecta al supuesto segundo (normalidad de los datos), éste se suele comprobar mediante la realización de un gráfico de normalidad para el conjunto de los datos. La figura 34 muestra dicho gráfico para el ejemplo anterior de los algoritmos. Siempre que los puntos (que representan a las observaciones) estén razonablemente cerca de la línea recta (que representa a la distribución normal) y no muestren un patrón de comportamiento extraño, no hay motivos para sospechar que falla el supuesto de normalidad. Si se observase algún patrón de comportamiento anómalo (e.j.: muchos puntos excesivamente alejados de la línea o bien muchos puntos consecutivos situados al mismo lado de la línea), entonces el supuesto de normalidad quedaría en entredicho. Para el ejemplo de los algoritmos, no se observa en el gráfico nada extraño y, por tanto, se puede validar el supuesto de normalidad de los datos.

Figura 34. Gráfico de normalidad para los datos del ejemplo de algoritmos



Pasos a seguir

Este tipo de gráfico se puede obtener con Minitab sin más que marcar la casilla "Normal plot of residuals" en las opciones de Graphs de la ventana ANOVA (figura 24).

Finalmente, por lo que respecta al supuesto de varianza constante, este se suele comprobar o bien calculando las desviaciones estándar de las muestras para verificar que no hay grandes diferencias entre ellas (figura 35), o bien mediante un gráfico que permita comparar visualmente la dispersión de los datos en cada grupo (figura 36). En el caso del ejemplo de los algoritmos no se observan diferencias sustanciales entre las varianzas de los distintos grupos, lo que permite validar el supuesto de varianza constante.

Figura 35. La columna StDev permite estimar la varianza de cada grupo

Session

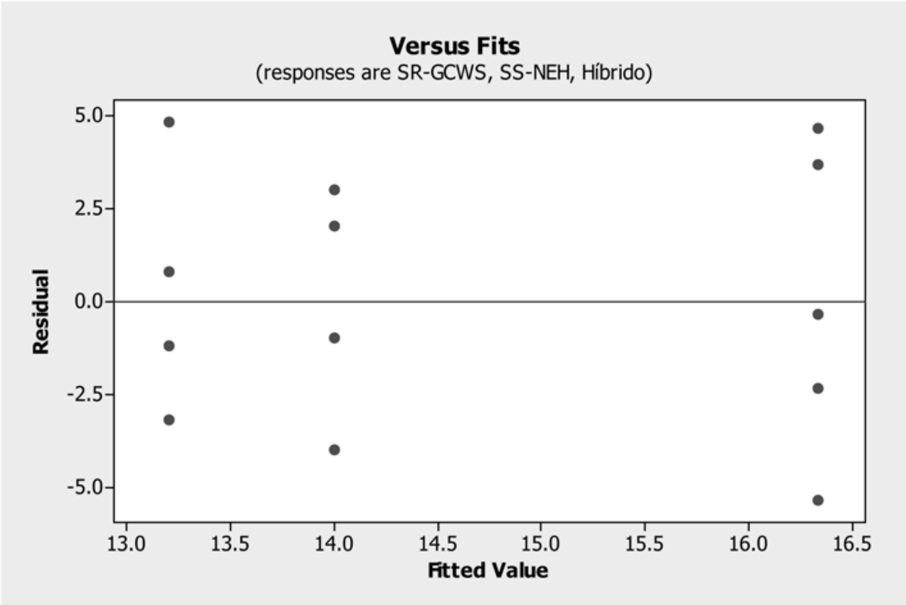
Descriptive Statistics: SR-GCWS, SS-NEH, Híbrido

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
SR-GCWS	5	0	13.20	1.36	3.03	10.00	11.00	12.00	16.00	18.00
SS-NEH	4	0	14.00	1.58	3.16	10.00	10.75	14.50	16.75	17.00
Híbrido	6	0	16.33	1.52	3.72	11.00	13.25	16.00	20.25	21.00

Recordad

La varianza, σ^2, es el cuadrado de la desviación estándar o típica, σ. Por lo general, el valor exacto de la varianza poblacional, σ^2, será desconocido, pero dicho valor se puede estimar mediante la varianza de la muestra, s^2.

Figura 36. El gráfico muestra la dispersión de cada grupo



Pasos a seguir

Este tipo de gráfico se puede obtener con Minitab sin más que marcar la casilla "Residuals versus fits" en las opciones de Graphs de la ventana ANOVA (figura 24).

Ejemplo de aplicación de ANOVA: comparando valoraciones medias en un cuestionario de escala Likert según el perfil de los encuestados

En una universidad se ha implementado recientemente un nuevo servicio online que facilita el acceso a recursos didácticos complementarios. Se desea conocer la opinión de los estudiantes sobre este nuevo servicio y, en particular, si existen diferencias significativas en la valoración media del servicio según la titulación a la que pertenezca el estudiante. Para ello, un investigador ha seleccionado al azar cinco estudiantes de cada uno de los principales estudios que se ofrecen y les ha pedido que rellenen un cuestionario de evaluación del servicio. El cuestionario usa una escala Likert entre 1 (mínima valoración) y 7 (máxima valoración). Los resultados obtenidos se muestran en la tabla 6.

Tabla 6. Valoraciones obtenidas según perfil del estudiante

Estudios				
CC. Información	**CC. Empresariales**	**Ing. Informática**	**Derecho**	**Psicología**
6	4	5	4	3
5	4	4	4	3
5	3	4	5	2
7	3	6	4	4
4	2	2	6	2

La figura 37 muestra el *output* Minitab correspondiente a los estadísticos descriptivos para cada grupo o nivel de observaciones. A simple vista parecen apreciarse diferencias considerables entre la máxima valoración media (CC. Información, con 5,4) y la mínima (Psicología, con 2,8). El *boxplot* de la figura 38 también apunta a la posibilidad de que las valoraciones medias del servicio puedan depender del perfil del estudiante, no siendo las mismas para todas las titulaciones.

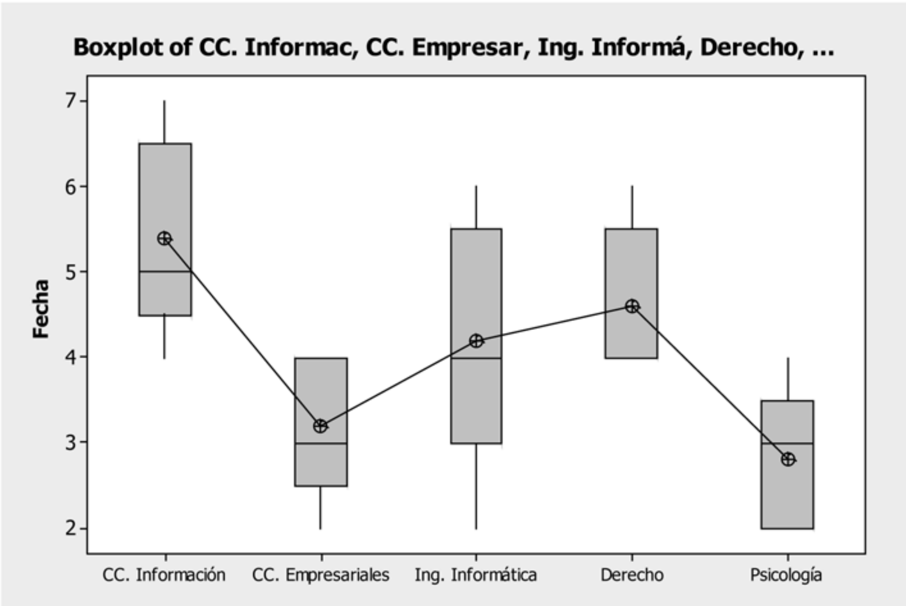
Figura 37. Estadísticos descriptivos de las valoraciones por grupo

Session

Descriptive Statistics: CC. Informac, CC. Empresar, Ing. Informá, Derecho, ...

Variable	N	N*	Mean	SE Mean	StDev	Sum	Minimum	Q1	Median
CC. Información	5	0	5.400	0.510	1.140	27.000	4.000	4.500	5.000
CC. Empresariales	5	0	3.200	0.374	0.837	16.000	2.000	2.500	3.000
Ing. Informática	5	0	4.200	0.663	1.483	21.000	2.000	3.000	4.000
Derecho	5	0	4.600	0.400	0.894	23.000	4.000	4.000	4.000
Psicología	5	0	2.800	0.374	0.837	14.000	2.000	2.000	3.000

Figura 38. *Boxplot* para las valoraciones del servicio por titulación



Para poder corroborar o desmentir estas impresiones visuales de una forma más científica, se opta por realizar un test *F* de ANOVA con un nivel de significación $\alpha = 0{,}01$ (es decir, en este caso se opta por usar un nivel de confianza del 99%).

La figura 39 muestra el *output* ANOVA de Minitab, en el que se aprecia un MS Factor = 5,54 (variación entre-grupos), un MS Error = 1,14 (variación intra-grupos) y un valor para el estadístico de contraste $F = 5{,}54 / 1{,}14 = 4{,}86$. En el supuesto de que la hipótesis nula fuese cierta, este estadístico seguiría una distribución *F*-Snedecor con 4 grados de libertad en el numerador (DF Factor) y 20 grados de libertad en el denominador (DF Error). La probabilidad de que una variable aleatoria *F*-Snedecor (4, 20) tome un valor igual o superior a 4,86 es 0,007 (*p*-valor). Esta probabilidad es extremadamente baja (más baja que el valor de significación fijado), lo cual pone en entredicho el supuesto inicial de que la hipótesis nula era cierta. En otras palabras: puesto que p-valor $< \alpha$ hay que rechazar la hipótesis nula. Así, pues, según las evidencias empíricas encontradas, se puede afirmar con un 99% de confianza que las valoraciones medias de los grupos no son todas iguales.

Figura 39. *Output* ANOVA de Minitab para la comparativa de valoraciones

```
Session

One-way ANOVA: CC. Informac, CC. Empresar, Ing. Informá, Derecho, Psicología

Source  DF     SS    MS     F      P
Factor   4  22.16  5.54  4.86  0.007
Error   20  22.80  1.14
Total   24  44.96

S = 1.068   R-Sq = 49.29%   R-Sq(adj) = 39.15%


Level                N   Mean  StDev
CC. Información      5  5.400  1.140
CC. Empresariales    5  3.200  0.837
Ing. Informática     5  4.200  1.483
Derecho              5  4.600  0.894
Psicología           5  2.800  0.837

                     Individual 99% CIs For Mean Based on Pooled StDev
Level                  ---------+---------+---------+---------
CC. Información                        (--------*--------)
CC. Empresariales         (--------*--------)
Ing. Informática                (--------*--------)
Derecho                            (--------*--------)
Psicología              (--------*--------)
                       ---------+---------+---------+---------
                     1.5       3.0       4.5       6.0

Pooled StDev = 1.068
```

La segunda parte del *output* Minitab ofrece los intervalos de confianza, a un nivel de confianza del 99% en este caso, para cada una de las medias. Se observa cómo los intervalos más extremos, p. ej.: los correspondientes a CC. Información y Psicología, no se solapan por muy poco. Esto es lógico, puesto que el p-valor = 0,007 está muy cercano al valor de significación escogido α = 0,01. Si el p-valor hubiera sido todavía menor, ambos intervalos estarían claramente separados. Si, por el contrario, el p-valor hubiera sido mayor, ambos intervalos se solaparían parcialmente como ocurre en el resto de los casos.

Antes de dar por definitivas las conclusiones anteriores, conviene validar que se cumplen los supuestos básicos de normalidad y varianza constante de los datos. La figura 40 muestra el gráfico de normalidad correspondiente a las observaciones. No parecen observarse patrones extraños ni demasiados puntos excesivamente alejados de la recta, por lo que se aceptará como válido el supuesto de normalidad. Por su parte, la figura 41 muestra el gráfico de dispersión de cada grupo. Tampoco se observan grandes diferencias entre las dispersiones de los distintos niveles, por lo que se aceptará como válido el supuesto de varianza constante entre los distintos grupos.

Atención

Para evitar confusiones en la segunda parte del *output*, es importante fijar bien el nivel de confianza (1 – α) en la ventana ANOVA (figura 24), de manera que éste se corresponda con el nivel de significación α escogido en cada caso.

Figura 40. Gráfico de normalidad de las valoraciones registradas

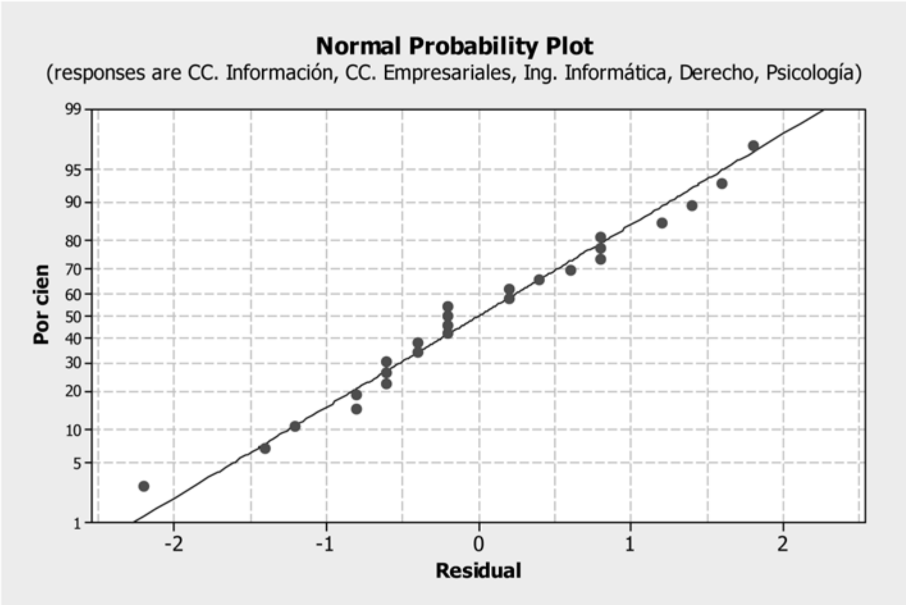
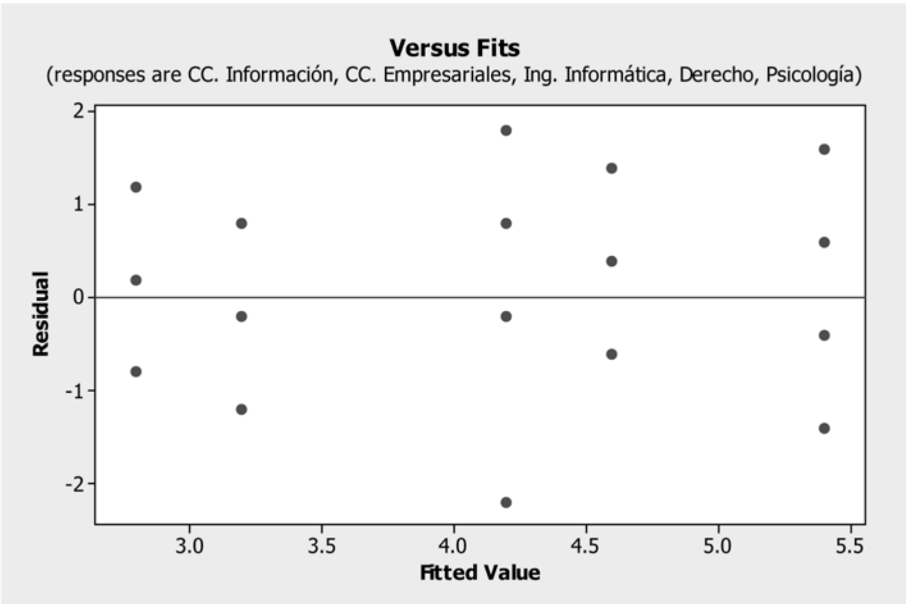


Figura 41. Gráfico de dispersión de cada grupo



Resumen

En este módulo se han presentado las principales técnicas estadísticas que permiten comparar estadísticamente dos o más grupos y discernir si existen o no diferencias significativas entre ellos. En el caso de dos grupos, se usa un contraste de hipótesis basado en la t-Student (si se están comparando dos medias) o en la normal (si se están comparando dos proporciones).Por último, se han estudiado procedimientos que se pueden aplicar para hacer inferencias acerca de varianzas poblacionales. Se presentó la distribución F, para emplearla en pruebas de hipótesis acerca de las varianzas de dos poblaciones normales.

En el caso de tres o más grupos, se usa un test ANOVA basado en la F-Snedecor.

Conviene tener siempre muy presente que lo más importante de un test de hipótesis no son los cálculos matemáticos que subyacen al mismo (en gran parte porque dichos cálculos se pueden automatizar mediante el uso de software), sino la correcta interpretación de los resultados obtenidos y la credibilidad de los mismos, que dependerá de que se cumplan o no los supuestos necesarios para poder aplicar cada una de las técnicas de inferencia vistas en este módulo. Si bien el ordenador puede ser muy útil efectuando los cálculos matemáticos con precisión y rapidez, es responsabilidad del investigador saber interpretar los resultados y comprobar la validez de los supuestos.

Ejercicios de autoevaluación

1) Se estudia el impacto que causa la reubicación forzada sobre la buena vecindad. Se entrevista a seis individuos tanto antes como después de que se les obligara a mudarse. Las entrevistas producen las siguientes puntuaciones:

Entrevistado	Antes	Después
1	2	1
2	1	2
3	3	1
4	3	1
5	1	2
6	4	1

Realizad un contraste de hipótesis al nivel de confianza del 95%.

2) De una muestra de ochenta y cinco mensajes de correo con virus que llegan al servidor de nuestra empresa, nuestro programa KILLVIRUS instalado en el servidor sólo ha detectado veinticinco. Las especificaciones del programa decían que el programa detectaba más del 40% del correo con virus. ¿Estáis de acuerdo con los resultados obtenidos con las especificaciones del programa? (considerad $\alpha = 0{,}1$). Hallad el *p*-valor del contraste.

3) Queremos comparar la eficiencia de dos compiladores de dos sistemas de indización diferentes: A y B. Para hacerlo, se diseñan ocho programas en cada uno de los dos sistemas y se mide el tiempo de ejecución que tarda cada uno de los programas para resolver ocho problemas determinados de optimización. Los resultados se muestran en la tabla siguiente:

Problema de optimización a resolver	Tiempo de ejecución usado por el ejecutable compilado con el sistema A (en segundos)	Tiempo de ejecución usado por el ejecutable compilado con el sistema B (en segundos)
P1	1,2	1,4
P2	1,3	1,7
P3	1,5	1,5
P4	1,4	1,3
P5	1,7	2,0
P6	1,8	2,1
P7	1,4	1,7
P8	1,3	1,6

¿Podemos asegurar a partir de los datos anteriores que el compilador del sistema A es más eficiente que el compilador del sistema B? (considerad $\alpha = 0{,}05$). Hallad el *p*-valor del contraste.

4) Dos empresas, A y B, quieren comprar un dispositivo de almacenamiento para realizar copias de seguridad. Antes de hacer la compra se hace un estudio de cuántos gigas necesitarían para realizar la copia. Este estudio consiste en calcular durante diez días toda la información de la empresa necesaria para la copia de seguridad. Los resultados se muestran en la tabla siguiente:

Día	1	2	3	4	5	6	7	8	9	10
Empresa A (gigas)	34	45	47	49	31	30	24	33	35	40
Empresa B (gigas)	45	47	50	42	40	51	46	59	42	46

Suponiendo normalidad y un nivel de significación del 0,05, ¿podemos afirmar que la empresa B necesita más capacidad de almacenamiento que la empresa A? Indicación: antes de nada, tenéis que realizar el contraste correspondiente para ver si las varianzas de las dos muestras son iguales a un nivel de significación de 0,05.

5) Se ha diseñado un experimento aleatorio para analizar durante cuánto tiempo es efectiva cada una de las cuatro drogas distintas que se pueden emplear para aliviar el dolor tras una operación quirúrgica. Los datos obtenidos se muestran en la tabla siguiente:

	Droga			
	A	***B***	***C***	***D***
Tiempo (horas)	8	6	8	4
	6	6	10	4
	4	4	10	2
	2	4	10	
			12	

Para un nivel de significación α = 005, contrastar la hipótesis nula de que las cuatro drogas son igualmente efectivas.

6) A la hora de descargar programas *open-source* de Internet, suele ser habitual poder optar por hacerlo desde varios servidores (*mirrors*). Generalmente, las velocidades de descarga desde cada servidor dependen de la distancia existente entre el servidor y el cliente que solicita la descarga. En este caso se desea estudiar si las velocidades de descarga desde cinco servidores distintos se pueden considerar equivalentes o no. Para cada uno de los servidores, se han seleccionado algunos ficheros al azar (todos ellos del mismo tamaño) y se han descargado en el cliente, obteniendo los tiempos de descarga (en segundos) que se muestran en la tabla siguiente:

	Servidor				
	A	***B***	***C***	***D***	***E***
Tiempo de descarga (en segundos)	3,8	6,8	4,4	6,5	6,2
	4,2	7,1	4,1	6,4	4,5
	4,1	6,7	3,9	6,2	5,3
	4,4		4,5		5,8

¿Se puede afirmar que la velocidad media de descarga es independiente del servidor seleccionado? Usar un nivel de significación $\alpha = 0{,}01$.

7) Se desean comparar los ingresos por familia (en miles de euros) correspondientes a tres provincias de una misma comunidad autónoma. A tal efecto, para cada provincia se han seleccionado 9 familias al azar y se han registrado sus ingresos. La tabla siguiente muestra las observaciones obtenidas:

	Provincia		
	A	***B***	***C***
Ingresos familiares (miles de euros)	45	32	40
	39,5	30	42
	42	37	45
	35	35	39,5
	40	28,5	40
	37	37,5	38
	44	31	51
	48,5	37,6	47,5
	50	25	41

Para un nivel de significación $\alpha = 0{,}05$, ¿se puede afirmar que los ingresos medios por familia no dependen de la provincia a la que ésta pertenezca?

8) Una universidad hace uso de tres consultorías externas que ofrecen servicios de asesoramiento técnico en línea a sus estudiantes. Para cada una de estas consultorías, se han escogido al azar seis servicios prestados durante el año en curso y se ha registrado el cambio porcentual en su precio con respecto al precio medio del año anterior. Los datos recogidos se muestran en la tabla siguiente:

	Consultoría		
	A	*B*	*C*
Cambio porcentual en el precio del servicio	3,0	4,5	1,0
	2,5	2,5	–2,5
	–1,5	7,0	–3,5
	4,0	9,0	2,0
	–1,0	1,5	4,6
	5,5	2,0	0,5

Para un nivel de significación $\alpha = 0{,}01$, se desea contrastar la hipótesis nula de que el cambio porcentual medio en el precio del servicio es el mismo para las tres consultorías.

9) Se desea comparar el nivel de innovación/originalidad de seis revistas distintas, aunque todas ellas pertenecientes a un mismo ámbito temático. A tal efecto, se han seleccionado al azar siete ejemplares de cada una de las revistas y un comité de expertos ha evaluado el nivel de innovación/originalidad de cada ejemplar, para lo cual se ha usado una escala entre 1 (mínimo) y 300 (máximo). Los datos recogidos se muestran en la tabla siguiente:

	Revista					
	A	*B*	*C*	*D*	*E*	*F*
Nivel de innovación/ originalidad	300	190	228	276	162	264
	300	164	300	296	175	168
	300	238	268	62	157	254
	260	200	280	300	262	216
	300	221	300	230	200	257
	261	132	300	175	256	183
	300	156	300	211	92	93

A partir de estas observaciones, ¿se puede afirmar que todas las revistas muestran un nivel de innovación/originalidad equivalente o, por el contrario, existen diferencias significativas entre los niveles de innovación/originalidad de las distintas revistas? Utilizar un nivel de significación $\alpha = 0{,}05$.

Solucionario

1) Se trata de un contraste de diferencia de medias para **muestras dependientes o emparejadas (e**stadístico de contraste t^* =1,49; el valor crítico es $t_{\alpha/2=0,05/2,}$ y 5 grados de libertad = 2,571. $t^* < t_{0,025;\,5}$ no se puede rechazar H_0. Podemos decir con un 95% de confianza que la buena vecindad no ha variado cuando se produce la reubicación.

2) Se trata de un contraste de **diferencia de proporciones**. El estadístico de contraste sigue aproximadamente la distribución normal $N(0,1)$ si el tamaño de la muestra es suficientemente grande como en nuestro caso. El valor del estadístico de contraste es $z^* = -1{,}993$.

El valor crítico será: $z_{0,1} \approx 1{,}28$. Como $z < z_{0,1}$, aceptamos la hipótesis nula y concluimos que las especificaciones del servidor son falsas.

3) Se trata de un contraste de **diferencia de medias dependientes**. El valor del estadístico de contraste vale: $t \approx -3{,}481$. El valor crítico vale $t_{0,05,7} \approx 1{,}895$. Como $t < -t_{0,05;7}$; rechazamos la hipótesis.

El p-valor es $p = p(t_7 < -3{,}481) \approx 0{,}0051$, valor que es menor que 0,05. Por tanto, llegamos a la misma conclusión: rechazar la hipótesis nula.

4) El resultado del Minitab para el **contraste de varianzas** es:

```
Test for Equal Variances: A; B

95% Bonferroni confidence intervals for standard deviations

    N    Lower    StDev    Upper
A  10  5,33911  8,16224  16,4359
B  10  3,60659  5,51362  11,1025
F-Test (Normal Distribution)
Test statistic = 2,19; p-value = 0,258
Levene's Test (Any Continuous Distribution)
Test statistic = 1,61; p-value = 0,221
```

Se acepta la igualdad de varianzas.

El resultado del Minitab para el **contraste de diferencia medias independientes** es:

```
Two-Sample T-Test and CI: Empresa A; Empresa B

Two-sample T for Empresa A vs Empresa B

            N   Mean  StDev  SE Mean
Empresa A  10  36,80   8,16      2,6
Empresa B  10  46,80   5,51      1,7


Difference = mu (Empresa A) - mu (Empresa B)
Estimate for difference:  -10,00
95% upper bound for difference:  -4,60
T-Test of difference = 0 (vs <): T-Value = -3,21  P-Value =
0,002  DF = 18
Both use Pooled StDev = 6,9650
```

Como p-valor < 0,05, rechazamos la hipótesis nula, por lo tanto aceptamos que la empresa B necesita más capacidad de almacenamiento que la empresa A.

5) Estadístico de contraste $F = 12{,}50$; p-valor = 0,001 < $\alpha = 0{,}05$ → Rechazar la hipótesis nula, p. ej.: no todos los grupos tienen el mismo comportamiento.

6) Estadístico de contraste $F = 31,6$; p-valor = 0,000 < α = 0,01 → Rechazar la hipótesis nula, p. ej.: no todos los grupos tienen el mismo comportamiento.

7) Estadístico de contraste $F = 13,83$; p-valor = 0,000 < α = 0,05 → Rechazar la hipótesis nula, p. ej.: no todos los grupos tienen el mismo comportamiento.

8) Estadístico de contraste $F = 2,91$; p-valor = 0,085 > α = 0,01 → No rechazar la hipótesis nula, p. ej.: todos los grupos parecen tener el mismo comportamiento.

9) Estadístico de contraste $F = 5,30$; p-valor = 0,001 < α = 0,05 → Rechazar la hipótesis nula, p. ej.: no todos los grupos se comportan igual.