

# Contraste de hipòtesis

Carles Rovira Escofet

P08/75057/02308



Universitat Oberta  
de Catalunya

[www.uoc.edu](http://www.uoc.edu)



# Índice

## Sesión 1

<b>Introducción al contraste de hipótesis .....</b>	<b>5</b>
1. Hipótesis nula y alternativa .....	5
2. El nivel de significación .....	8
3. El concepto de estadístico de contraste .....	9
4. El $p$ -valor .....	11
5. Resolver un contraste de hipótesis .....	12
6. Relación con los intervalos de confianza .....	12
7. Resumen .....	13
Ejercicios .....	14

## Sesión 2

<b>Contrastes sobre la media y sobre la proporción .....</b>	<b>17</b>
1. Contrastes sobre la media .....	17
1.1. Caso de la varianza conocida .....	17
1.2. Caso de la varianza desconocida .....	18
1.3. Caso con muestras no normales .....	20
1.4. Datos aparejados .....	21
2. Contrastes sobre la proporción (muestras grandes) .....	22
3. Resumen .....	24
Ejercicios .....	25



## Introducción al contraste de hipótesis

En esta sesión presentamos una nueva técnica estadística: el contraste de hipótesis. Primero veremos algunas situaciones en las que podemos utilizarla.

*Test o contraste son dos maneras de llamar a la misma técnica.*

Tiramos al aire una misma moneda cien veces. En principio se trata de una moneda normal, de manera que esperamos que nos salga aproximadamente el mismo número de caras que de cruces. Esperamos, por tanto, una relación del orden 50-50 (50 caras, 50 cruces), o 48-52. Pero, ¿qué sucede si la relación que obtenemos es del orden 45-55 o, todavía más, de 40-60? ¿Podemos considerar que el resultado obtenido nos indica que la moneda no es perfecta? ¿O tenemos que decir que el resultado es sólo fruto de la casualidad?

Otro caso: imaginad que compramos una máquina que fabrica chips y que, según el fabricante, hace como máximo un 1% de defectuosos. Después de fabricar trescientos chips, los revisamos y descubrimos que hay cinco defectuosos (según el fabricante deberían haber sido tres). ¿Este aumento del número de chips defectuosos indica que la máquina no funciona bien, o podemos pensar que sólo es casualidad?

Una última situación. Un estudio afirma que la media de las alturas de los chicos catalanes de 17 años es de 178 cm. Visitamos un instituto y escogemos a diez chicos al azar; su altura media es de 171 cm. ¿Podemos pensar que los muchachos de este instituto tienen una altura diferente de la del conjunto de chicos de Cataluña?

Fijaos en que las tres situaciones anteriores son similares. En los tres casos suponemos cierta hipótesis y debemos decidir, a partir de nuestras observaciones, si tenemos suficientes evidencias para poder rechazarla. ¿Cómo lo haremos? Utilizaremos probabilidades.

### Importancia del test

El test será la herramienta que nos permitirá extraer conclusiones a partir de la diferencia entre las observaciones y los resultados que se deberían obtener si la hipótesis de partida fuese cierta.

### 1. Hipótesis nula y alternativa

El primer paso cuando queremos estudiar este tipo de problema es determinar bien las hipótesis. Es preciso que determinemos dos, ya que la resolución del problema consistirá, en cierta medida, en escoger una entre estas dos.

La **hipótesis nula**, que indicaremos por  $H_0$ , es la hipótesis de partida. Debe recoger el hecho que queramos someter a prueba.

La **hipótesis alternativa** es la que, como su nombre indica, ofrecemos como alternativa a la nula. Esta hipótesis, que denotamos por  $H_1$ , representa que se ha producido un cambio con respecto a la situación descrita por la hipótesis nula.

Para poder plantear correctamente las hipótesis, deberemos entender bien la estructura estadística que se encuentra tras los datos y, por tanto, deberemos conocer qué distribución siguen. Las hipótesis se expresarán normalmente en términos de algún parámetro de la distribución de los datos que estudiamos.

Volvamos al ejemplo de las alturas, pero ahora añadamos un poco más de información. Sabemos, por ejemplo, que la variable “altura” sigue una distribución normal. Por tanto, una altura es una observación de una variable  $N(\mu, \sigma^2)$ . Supondremos también que  $\sigma$  es conocida y es igual a 3. Entonces, cuando decimos que los chicos de 17 años tienen una altura media de 178 cm, en realidad proponemos que la hipótesis nula expresada en términos del parámetro  $\mu$  es ésta:

$$H_0: \mu = 178$$

Para detectar si las alturas no siguen esta media, debemos coger como hipótesis alternativa:

$$H_1: \mu \neq 178 \text{ (hipótesis bilateral)}$$

Pero si quisiéramos detectar si los chicos son más altos de 178 cm, la hipótesis alternativa debería ser ésta:

$$H_1: \mu > 178 \text{ (hipótesis unilateral)}$$

Mientras que si quisiéramos detectar si son más bajos de 178 cm, debería ser:

$$H_1: \mu < 178 \text{ (hipótesis unilateral)}$$

La hipótesis alternativa puede ser bilateral, si el parámetro es diferente del valor correspondiente a la hipótesis nula, o unilateral, si sólo lo compara en una dirección.

#### Observación

Con la misma hipótesis nula podemos estudiar varias hipótesis alternativas.

#### Ejemplo de la moneda

Consideremos el ejemplo de lanzar una moneda y ver si es perfecta. Podemos representar el lanzamiento de una moneda por una variable de Bernoulli de parámetro  $p$ ,  $B(p)$ , donde  $p$  indica la probabilidad de sacar una cara.

Si queremos “contrastar” el hecho de que la moneda sea perfecta con respecto a que no lo sea, la hipótesis nula  $H_0$  corresponde a “la moneda es perfecta”, y la hipótesis alternativa  $H_1$ , a “la moneda no es perfecta”, que se puede representar en términos del parámetro  $p$  de la manera siguiente:

$$H_0: p = \frac{1}{2}, \text{ contra } H_1: p \neq \frac{1}{2}$$

Fijaos en que cogemos como hipótesis nula el supuesto “la moneda es perfecta”, que es la situación habitual; la hipótesis alternativa es la que indica que ha habido cambios.

#### Observad que...

... las hipótesis se expresan en términos del parámetro de los datos que utilizan.

Observad que podríamos plantear el caso de manera que la hipótesis nula correspondiera a “la moneda es perfecta” y la alternativa, a “la probabilidad de cara es más alta que la de cruz”. Si lo escribimos en función del parámetro, resulta:

$$H_0: p = \frac{1}{2}, \text{ contra } H_1: p > \frac{1}{2}$$

Tenemos la misma hipótesis nula, pero la hipótesis alternativa es diferente. Fijaos en que el primer caso era una alternativa bilateral, mientras que en el segundo es unilateral.

Una vez planteadas las hipótesis nula y alternativa, debemos tomar una decisión a partir de las observaciones. Por otro lado, existen dos decisiones posibles:

- 1) Aceptar la hipótesis nula
- 2) Rechazar la hipótesis nula

Por otro lado, hay dos situaciones posibles:

- a) La hipótesis nula es cierta
- b) La hipótesis nula es falsa

Esto hace que podamos cometer dos clases de errores diferentes: podemos aceptar la hipótesis nula cuando ésta es falsa o podemos rechazar la hipótesis nula cuando es cierta, tal como se ve en la tabla siguiente:

		Decisiones	
		Aceptar $H_0$	Rechazar $H_0$
Situaciones	$H_0$ cierta	Decisión correcta	Error de tipo I
	$H_0$ falsa	Error de tipo II	Decisión correcta

#### Importancia de las observaciones

Según el conjunto de observaciones concreto, decidiremos rechazar o aceptar la hipótesis nula.

Podemos medir estos errores con la probabilidad de que se den las situaciones respectivas.

Así, podemos considerar los errores siguientes:

- El error de tipo I es la probabilidad de que rechacemos la hipótesis nula cuando ésta es cierta.
- El error de tipo II es la probabilidad de que aceptemos la hipótesis nula cuando ésta es falsa.

Ahora necesitamos una regla de decisión que nos permita determinar si debemos aceptar o rechazar la hipótesis nula a partir de las observaciones. Puesto que podemos cometer dos tipos de errores diferentes, nos interesaría tomar la decisión de que minimizara los dos. Como siempre, la solución de este problema no es fácil. Normalmente, la experiencia nos enseña que cuando establecemos una regla de decisión que reduce mucho uno de los errores, el otro aumenta.

Dado que no podemos hacer pequeños los dos errores al mismo tiempo, utilizaremos la siguiente estrategia: buscaremos reglas de decisión que nos permitan tener limitado el error de primer tipo, es decir, el error que cometemos cuando decidimos rechazar la hipótesis nula y ésta es cierta. Esta estrategia es conservadora, tendemos a no rechazar la hipótesis nula excepto si los resultados son muy poco probables con ésta.

Una buena forma de hacer pequeños los dos errores y, por tanto, de mejorar nuestros resultados, es aumentar el tamaño de las muestras que utilizamos.

### Orígenes de la estrategia

La estrategia de decisión que utilizamos aquí se basa en los resultados de Neyman y Pearson, dos importantes matemáticos del siglo XX.

## 2. El nivel de significación

El nivel de significación  $\alpha$  de un contraste es el error máximo de tipo I que estamos dispuestos a asumir.

Determinar el nivel de significación que hay que fijar cuando se comienza el estudio no es un problema estrictamente matemático, ya que depende de cada caso particular. Se suele utilizar el valor estándar de  $\alpha = 0,05$ . Otros niveles utilizados son del orden de 0,1, de 0,01 o incluso de 0,001 (si queremos aumentar la precisión).

A la hora de escoger el nivel de significación, recordemos que el error de tipo I (rechazar  $H_0$  cuando ésta es cierta) indica que detectamos un cambio con respecto a la hipótesis nula, cuando en realidad dicho cambio no existe. Por tanto, podemos tener en cuenta las consideraciones siguientes:

- Si detectar este cambio –que en realidad no existe y sólo es consecuencia del carácter aleatorio de los datos– no es demasiado grave, entonces podemos escoger un nivel de significación alto (por ejemplo, del orden de 0,1).
- Si detectar este cambio –que en realidad no existe– es un error grave, entonces debemos fijar un nivel de significación pequeño.

Sin embargo, hay que ir con cuidado. Cuanto menor sea el valor de  $\alpha$  que fijemos, más tendencia tendremos a aceptar la hipótesis nula. El caso extremo sería fijar un nivel de significación 0. En este caso aceptaremos siempre la hipótesis nula, de manera que nunca se dará el error de tipo I, pero está claro que si tenemos que aceptar siempre, el estudio que hemos llevado a cabo no nos aporta nada nuevo.

El hecho de tener acotado el error de tipo I hace que nuestros contrastes sean conservadores y tiendan a aceptar la hipótesis nula, a menos que haya evidencias muy claras de que tenemos que rechazarla.

Cuando aceptamos la hipótesis nula, no estamos seguros de que sea realmente cierta, ya que no controlamos el error de tipo II (el error que cometemos cuando aceptamos la hipótesis nula y ésta es falsa).

### Sentido del nivel de significación...

... de un contraste.

Un nivel  $\alpha = 0,05$  significa que, aunque la hipótesis nula sea cierta, los datos de cinco de cada cien muestras nos harán rechazarla. Es decir, aceptamos que podemos rechazar la hipótesis nula de forma equivocada cinco de cada cien veces.

### ¿"Aceptamos" o "no rechazamos"?

Puesto que cuando aceptamos la hipótesis nula no estamos demasiado seguros, normalmente, en lugar de decir "Aceptamos la hipótesis nula", decimos "No rechazamos la hipótesis nula".



Cuando rechazamos la hipótesis nula, estamos seguros de que tenemos que rechazarla porque tenemos acotado el error de tipo I (el error que cometemos cuando rechazamos la hipótesis nula y ésta es cierta).

### 3. El concepto de estadístico de contraste

Ya tenemos fijadas las hipótesis, así como el error de tipo I que estamos dispuestos a asumir. Para decidir si rechazamos la hipótesis nula o no, utilizaremos el llamado *estadístico de contraste*.

Un **estadístico de contraste** es una función de la muestra cuya distribución conocemos bajo la hipótesis nula.

Todo problema diferente tiene su estadístico de contraste correspondiente.

Veamos cómo utilizar el estadístico de contraste para decidir si rechazamos la hipótesis nula o no. Por ejemplo, en el caso de las alturas de los chicos, sabemos que si tenemos una muestra de alturas de  $n$  chicos escogidos al azar, bajo la hipótesis nula (es decir,  $\mu = 178$ ) podemos definir la variable:

$$z = \frac{\bar{x} - 178}{\frac{3}{\sqrt{n}}}$$

#### Recordad que...


... si tenemos una muestra de tamaño  $n$  de una distribución  $N(\mu, \sigma^2)$ , entonces la variable:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

sigue una distribución normal estándar.

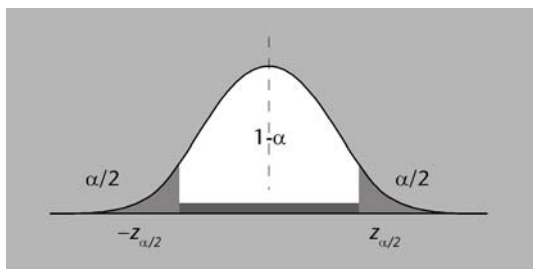
Este valor es una observación de una ley  $N(0,1)$ .

Si la hipótesis nula es cierta, el valor observado  $z$  debería estar en la zona en la que la distribución normal estándar concentra una mayor probabilidad, es decir, alrededor del cero. Si nos sale un valor muy alejado del cero, este valor será poco probable bajo la hipótesis nula, y nos llevará a decidir rechazarla, ya que pensaremos que su aparición no puede ser debida al azar, sino al hecho de que la hipótesis nula debe de ser falsa. Ahora bien, ¿hasta qué punto debe ser poco probable para decidir rechazarla? Esto vendrá dado por el nivel de significación fijado.

Así, imaginad que hacemos el contraste  $H_0: \mu = 178$  contra  $H_1: \mu \neq 178$ . Un valor del estadístico de contraste cercano a 0 es más probable bajo  $H_0$  que bajo  $H_1$ . De este modo, utilizaremos la regla de decisión siguiente: 

Ejemplo de contraste bilateral.

- Aceptaremos  $H_0$  si  $|z| \leq z_{\frac{\alpha}{2}}$
- Rechazaremos  $H_0$  si  $|z| > z_{\frac{\alpha}{2}}$



### Zona de aceptación de la hipótesis nula

La parte del gráfico con una línea más gruesa corresponde a la zona en la que queremos aceptar la hipótesis nula. Fijaos en cómo influye la hipótesis alternativa a la hora de decidir.

donde  $z_{\frac{\alpha}{2}}$  es el llamado *valor crítico*.

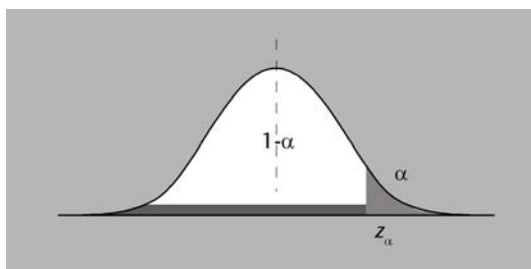
Para determinar el valor crítico  $z_{\alpha/2}$ , sólo hay que imponer que el error de tipo I (probabilidad de rechazar  $H_0$  cuando es cierta) sea menor o igual que el nivel de significación  $\alpha$ , es decir:

$$P(|Z| > z_{\frac{\alpha}{2}}) = P(Z > z_{\frac{\alpha}{2}}) + P(Z < -z_{\frac{\alpha}{2}}) \leq \alpha$$

donde  $Z$  es la distribución del estadístico de contraste. En este caso, bajo la hipótesis nula, sigue una distribución normal estándar. Por ejemplo, para  $\alpha = 0,05$  encontramos (por ejemplo, en las tablas de la normal) que  $z_{\alpha/2} = 1,96$ .

Hagamos ahora un contraste unilateral, en particular  $H_0: \mu = 178$  contra  $H_1: \mu > 178$ . En este caso un valor positivo del estadístico de contraste será más probable bajo  $H_1$  (ya que bajo la hipótesis alternativa lo más normal es que  $\bar{x}$  sea mayor que 178), mientras que los valores negativos lo serán bajo  $H_0$ . En este caso la regla de decisión será ésta:

- Aceptaremos  $H_0$  si  $z \leq z_{\alpha}$
- Rechazaremos  $H_0$  si  $z > z_{\alpha}$



Para determinar el valor crítico  $z_{\alpha}$  imponemos, como antes, que el error de tipo I sea menor que el nivel de significación  $\alpha$ , en este caso:

$$P(Z > z_{\alpha}) = \alpha$$

donde  $Z$  es la distribución del estadístico de contraste. Como antes, bajo la hipótesis nula, está distribuido como una normal estándar. Por ejemplo, para  $\alpha = 0,05$ , encontramos que  $z_{\alpha} = 1,65$ .

### Terminología

En los contrastes bilaterales, a veces se dice que la probabilidad de las dos colas (la parte que corresponde a la zona donde tenemos que rechazar la hipótesis nula) debe ser  $\alpha$ .

### Ejemplo de contraste unilateral

### Validez del método

El método es el mismo para cualquier distribución simétrica, así que también sirve si el estadístico de contraste sigue una distribución  $t$  de Student.

### Terminología

En este contraste unilateral se dice que la probabilidad de la cola de la derecha debe ser  $\alpha$ .

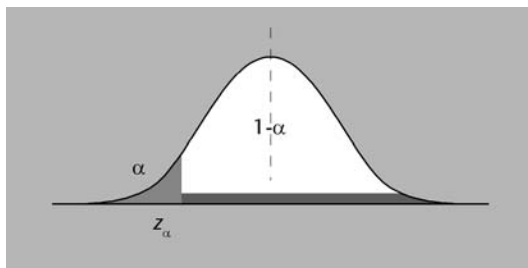
De la misma manera, vemos que para el contraste  $H_0: \mu = 178$  contra  $H_1: \mu < 178$ , la regla de decisión será la siguiente:

- Aceptaremos  $H_0$  si  $z \geq z_\alpha$ .
- Rechazaremos  $H_0$  si  $z < z_\alpha$ .

Con la condición:

$$P(Z < z_\alpha) = \alpha$$

donde  $Z$  es, como antes, la distribución del estadístico de contraste, que bajo la hipótesis nula tiene una distribución normal estándar. Para  $\alpha = 0,05$  encontramos que  $z_\alpha = -1,65$ .



La ley de esta variable  $Z$  se llama *ley del estadístico de contraste*.

#### Terminología

En este contraste unilateral se dice que la probabilidad de la cola de la izquierda debe ser  $\alpha$ .

## 4. El $p$ -valor

En muchos casos, para resolver un contraste de hipótesis no calcularemos el valor crítico, sino que utilizaremos el llamado  $p$ -valor.

El  $p$ -valor es la probabilidad del resultado del estadístico de contraste observado o de uno más alejado cuando la hipótesis nula es cierta.

El software estadístico más habitual utiliza el  $p$ -valor para hacer el contraste de hipótesis.

#### Sentido del $p$ -valor

El  $p$ -valor asociado a una observación del estadístico de contraste es el menor nivel de significación que nos permite rechazar la hipótesis nula.

Cuando el  $p$ -valor sea pequeño, indicará que el valor del estadístico de contraste que hemos observado tenía una probabilidad pequeña de salir bajo la hipótesis nula y, por tanto, deberemos rechazar la hipótesis nula. En cambio, cuando sea grande, indicará que era un valor bastante probable bajo la hipótesis nula y, por tanto, es lógico que aceptemos  $H_0$ .

Si el  $p$ -valor es inferior al nivel de significación  $\alpha$ , rechazaremos la hipótesis nula.

Si el  $p$ -valor es superior o igual al nivel de significación  $\alpha$ , aceptaremos la hipótesis nula.

#### Validez del método

El método es el mismo para cualquier distribución simétrica, así que también sirve si el estadístico de contraste sigue una distribución  $t$  de Student, como veremos más adelante.

### Ejemplo de aplicación del $p$ -valor

En el ejemplo de las alturas, imaginad que nuestro estadístico de contraste observado es 1,61 y denotamos por  $Z$  una variable aleatoria que tiene una distribución normal estándar, que es la ley del estadístico de contraste bajo la hipótesis nula. Supongamos, además, que hemos fijado un nivel de significación  $\alpha = 0,1$ . Así:

1) Si hacemos el contraste  $H_0: \mu = 178$  contra  $H_1: \mu \neq 178$ , entonces el  $p$ -valor es (probabilidad de las dos colas):

$$P(Z > |1,61|) = P(Z > 1,61) + P(Z < -1,61) = 2 \cdot 0,0537 = 0,1074 > \alpha \Rightarrow \text{No rechazamos } H_0$$

Fijaos en que para calcular el  $p$ -valor, calculamos la probabilidad de que la variable  $Z$  tome un valor más alejado que el valor 1,61 observado.

2) Si hacemos el contraste  $H_0: \mu = 178$  contra  $H_1: \mu > 178$ , entonces el  $p$ -valor es (probabilidad de la cola de la derecha):

$$P(Z > 1,61) = 0,0537 < \alpha \Rightarrow \text{Rechazamos } H_0$$

Fijaos en que, en este caso, el valor obtenido es muy poco probable bajo la hipótesis nula. De todos modos, si hubiéramos considerado un nivel de significación de 0,05, tendríamos que haber aceptado la hipótesis nula.

3) Si hacemos el contraste  $H_0: \mu = 178$  contra  $H_1: \mu < 178$ , entonces el  $p$ -valor es (probabilidad de la cola de la izquierda):

$$P(Z < 1,61) = 0,9463 > \alpha \Rightarrow \text{No rechazamos } H_0$$

En cambio, en este caso el valor obtenido es muy probable bajo la hipótesis nula, por eso la aceptamos.

Estas probabilidades en muchas ocasiones no se pueden calcular con tablas estadísticas, por lo que hay que utilizar un *software* estadístico.

## 5. Resolver un contraste de hipótesis

Podemos describir el procedimiento para plantear y resolver un contraste de hipótesis en cinco etapas:

- 1) Fijar las hipótesis nula y alternativa.
- 2) Fijar un nivel de significación.
- 3) Determinar el estadístico de contraste y su ley.
- 4) Calcular el  $p$ -valor asociado a nuestro estadístico de contraste calculado.
- 5) Comparar el  $p$ -valor con el nivel de significación y tomar una decisión.

### Otro procedimiento

Podemos sustituir los pasos 4 y 5 del procedimiento presentado para resolver un contraste de hipótesis por el cálculo del valor crítico.

## 6. Relación con los intervalos de confianza

Para los contrastes bilaterales, podemos hacer un contraste de hipótesis utilizando intervalos de confianza. Veámoslo con un ejemplo.

El peso en gramos de las manzanas de una explotación agraria sigue una distribución normal de desviación típica 3,5. Un estudio nos dice que el peso medio es de 155 gramos y queremos comprobar si eso es cierto. Si denotamos por  $\mu$  la esperanza de la distribución de las manzanas, queremos contrastar:

$$H_0: \mu = 155 \text{ contra } H_1: \mu \neq 155$$

Para hacerlo, hemos cogido una muestra aleatoria de 42 manzanas y hemos obtenido una media aritmética de 163,8.

Primero debemos encontrar un intervalo de confianza del 95% para la media de la distribución. Recordad que se puede obtener a partir de la fórmula siguiente:

$$\bar{x} \pm 1,96 \frac{\sigma}{\sqrt{n}}$$

que en nuestro caso nos da el intervalo (162,74, 164,85). Sabemos que de cada cien muestras que consideremos, el valor auténtico de la  $\mu$  estará, como mínimo, en noventa y cinco de los intervalos correspondientes. Por eso, si el valor que queremos contrastar a la hipótesis nula cae dentro del intervalo, aceptaremos la hipótesis nula.

Si el valor que queremos contrastar está en el intervalo de confianza, aceptamos la hipótesis nula, y si no está, la rechazamos.

En este caso, puesto que  $155 \notin (162,74, 164,85)$ , debemos rechazar la hipótesis nula.

## 7. Resumen

En esta sesión hemos presentado los conceptos fundamentales de una nueva técnica estadística: el contraste de hipótesis. Hemos introducido los conceptos de hipótesis nula, hipótesis alternativa, errores de tipo I y II, estadístico de contraste, valor crítico y  $p$ -valor. Finalmente hemos visto cómo utilizar un intervalo de confianza para hacer un contraste de hipótesis.

## Ejercicios

1. Compramos una máquina que fabrica chips que, según el fabricante, fabrica como máximo un 1% de chips defectuosos. Explicad las variables aleatorias que hay en esta situación y discutid las posibles hipótesis que podemos estudiar para comprobar si las especificaciones del fabricante son ciertas.
2. Se escoge una muestra de ochenta personas y se observa que diez tienen los ojos azules. Dad un intervalo de confianza del 95% para la proporción de personas con los ojos azules. Contrastad si se puede afirmar que la proporción de personas con los ojos azules es del 7% con un nivel de significación  $\alpha = 0,05$ .
3. Tenemos una muestra de tamaño 10 de una variable aleatoria con distribución normal de media  $\mu$  desconocida y varianza  $\sigma^2 = 4$ . Nuestras observaciones son éstas:

11,57	10,57	11,32	12,30	10,36
12,00	10,29	11,48	8,74	10,03

Contrastad, con nivel de significación de 0,1, las hipótesis siguientes:

- a)  $H_0: \mu = 11$  contra  $H_1: \mu \neq 11$
- b)  $H_0: \mu = 11$  contra  $H_1: \mu > 11$
- c)  $H_0: \mu = 11$  contra  $H_1: \mu < 11$

## Solucionario

1. En realidad se trata de hacer un control de calidad para determinar si la probabilidad de producir un chip defectuoso es inferior a 0,01. Consideremos un modelo en el que para cada chip tenemos una variable  $X_i$  con ley de Bernoulli de parámetro  $p$  desconocido, que toma el valor 1 cuando el chip es defectuoso, y el valor 0 cuando no lo es. Podemos considerar varios contrastes:

- a)  $H_0: p = 0,01$  contra  $H_1: p \neq 0,01$

En este caso la hipótesis alternativa  $H_1$  nos dice que la proporción de chips defectuosos no es del 1%. Si rechazamos la hipótesis nula, no sabremos si lo hacemos porque la proporción es mayor o menor. Es decir, podríamos rechazar una máquina que no hiciera ninguna pieza defectuosa. Este contraste no nos interesa.

- b)  $H_0: p = 0,01$  contra  $H_1: p < 0,01$

En este caso la hipótesis alternativa  $H_1$  nos dice que la proporción de chips defectuosos es menor del 1%. Sólo rechazaremos la hipótesis nula si la proporción de chips defectuosos es claramente inferior al 1%. Este contraste nos puede

interesar si sólo queremos una máquina de la que estemos muy seguros de que su producción será muy buena. Entonces nos quedaremos la máquina cuando rechacemos la hipótesis nula. Pero está el problema de que el test tiene tendencia a aceptar la hipótesis nula, de manera que cabe la posibilidad de que rechacemos máquinas que satisfacen las especificaciones del fabricante.

c)  $H_0: p = 0,01$  contra  $H_1: p > 0,01$

En este caso la hipótesis alternativa  $H_1$  nos dice que la proporción de chips defectuosos es mayor del 1%. Sólo rechazaremos la hipótesis nula si la proporción de chips defectuosos es claramente superior al 1%. Es muy posible que nos quedemos con una máquina con una producción de chips defectuosos superior a la especificada.

2. El intervalo de confianza al 95% sobre una proporción  $p$  se calcula a partir de la expresión siguiente:

$$\hat{p} \pm z_{0,025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Puesto que  $\hat{p} = 10/80 = 0,125$  y  $n = 80$ , obtenemos el intervalo  $(0,053, 0,197)$ . Queremos hacer el contraste:

$$H_0: p = 0,07 \text{ contra } H_1: p \neq 0,07$$

Dado que  $0,07 \in (0,053, 0,197)$ , aceptaremos la hipótesis nula.

3. De la misma manera que en el ejemplo de las alturas, cogeremos el estadístico de contraste siguiente:

$$z = \frac{\bar{x} - 11}{\frac{2}{\sqrt{n}}}$$

Bajo la hipótesis nula, este estadístico corresponde a una observación de una ley  $N(0,1)$ . En nuestro caso podemos calcular:

$$\bar{x} = 10,866$$

de manera que el estadístico de contraste es éste:

$$z = \frac{\bar{x} - 11}{\frac{2}{\sqrt{n}}} = -0,21$$

a) Para el contraste  $H_0: \mu = 11$  contra  $H_1: \mu \neq 11$  (es bilateral), calculamos los valores críticos  $\pm z_{\alpha/2}$ , de manera que:

$$P(|Z| > z_{\alpha/2}) = P(Z < -z_{\alpha/2}) + P(Z > z_{\alpha/2}) = 0,1$$

Y teniendo presente la simetría de la distribución normal:

$$P(Z > z_{\alpha/2}) = 0,05$$

donde  $Z$  corresponde a una distribución normal estándar. Encontramos que  $z_{\alpha/2} = 1,65$ . Por tanto, debemos aceptar la hipótesis nula.

Por otro lado, si calculamos el  $p$ -valor, debemos buscar la probabilidad de las dos colas:

$$P(|Z| > 0,21) = 2P(Z > 0,21) = 2 \cdot 0,4168 = 0,8336 > \alpha$$

y debemos aceptar la hipótesis nula.

b) Para  $H_0: \mu = 11$  contra  $H_1: \mu > 11$  (unilateral), el valor crítico satisface:

$$P(Z > z_{\alpha}) = 0,1$$

es decir, que  $z_{\alpha} = 1,28$ . La regla de decisión es:

- Aceptaremos  $H_0$  si  $z \leq z_{\alpha}$
- Rechazaremos  $H_0$  si  $z > z_{\alpha}$

Por tanto, en este caso aceptaremos la hipótesis nula.

Si calculamos el  $p$ -valor, debemos buscar la probabilidad de la cola de la derecha:

$$P(Z > -0,21) = P(Z < 0,21) = 0,5832 > \alpha$$

y debemos aceptar la hipótesis nula.

c) En el caso  $H_0: \mu = 11$  contra  $H_1: \mu < 11$  (unilateral), el valor crítico satisfará:

$$P(Z < z_{\alpha}) = 0,1$$

encontramos que  $z_{\alpha} = -1,28$ . La regla de decisión es ésta:

- Aceptaremos  $H_0$  si  $z \geq z_{\alpha}$
- Rechazaremos  $H_0$  si  $z < z_{\alpha}$

Por tanto, también aceptaremos la hipótesis nula.

Y calculamos el  $p$ -valor correspondiente. La probabilidad de la cola de la izquierda es la siguiente:

$$P(Z < -0,21) = 0,4168 > \alpha$$

de manera que debemos aceptar la hipótesis nula.



## Contrastes sobre la media y sobre la proporción

En esta sesión estudiaremos cómo llevar a cabo un contraste de hipótesis, primero sobre medias de poblaciones y después sobre proporciones.

El paso fundamental para hacer un contraste de hipótesis es determinar el estadístico de contraste. Sabemos que si tenemos una muestra lo bastante grande de una variable aleatoria con esperanza  $k$ , entonces un estadístico del tipo:

$$\frac{\text{Media muestral} - k}{\text{Error estándar de la media}}$$

es una observación de una distribución normal estándar. También sabemos que en muestras pequeñas, si las observaciones provienen de una distribución normal, este estadístico sigue o bien una distribución normal estándar (si la varianza es conocida) o bien una  $t$  de Student (si la varianza es desconocida). Veamos cómo utilizar estos estadísticos para hacer contrastes.

Una proporción se puede ver como la media en una población binaria.

### 1. Contrastes sobre la media

Dada una muestra  $x_1, \dots, x_n$  que proviene de una distribución normal  $N(\mu, \sigma^2)$ , supongamos que queremos hacer un contraste sobre el valor de la media  $\mu$ .

La hipótesis nula será del tipo:

$$H_0: \mu = \mu_0$$

donde  $\mu_0$  es una constante fijada. Podemos tener dos tipos de hipótesis alternativas:

- $H_1: \mu \neq \mu_0$  (bilateral)
- $H_1: \mu < \mu_0$  o  $H_1: \mu > \mu_0$  (unilateral)

#### Atención

En esta sesión, para resolver un contraste, buscaremos el  $p$ -valor y no calcularemos el valor crítico. Sin embargo, pensad que podemos resolver los contrastes de las dos formas.

El paso siguiente consiste en determinar el estadístico de contraste. Distinguiremos dos casos, según si conocemos la varianza  $\sigma^2$  o no la conocemos.

En una situación real difícilmente conoceremos la varianza.

#### 1.1. Caso de la varianza conocida

Si la varianza es conocida, el error estándar de la media se puede calcular como

$$\frac{\sigma}{\sqrt{n}}.$$

La variable:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

bajo la hipótesis nula sigue una distribución normal estándar.

Por tanto, para hacer contrastes sobre la media con  $\sigma$  conocida, el estadístico de contraste es éste:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

que se corresponde, bajo la hipótesis nula, con una observación de una distribución normal estándar.

Una vez calculado el valor observado del estadístico de contraste, debemos determinar el  $p$ -valor. Como ya hemos visto, éste depende de la hipótesis alternativa:

- Si  $H_1: \mu \neq \mu_0$ , entonces (probabilidad de las dos colas):

$$p = P(|Z| > |z|) = P(Z > |z|) + P(Z < -|z|)$$

- Si  $H_1: \mu < \mu_0$ , entonces (probabilidad de la cola de la izquierda):

$$p = P(Z < z)$$

- Si  $H_1: \mu > \mu_0$ , entonces (probabilidad de la cola de la derecha):

$$p = P(Z > z)$$

donde recordemos que  $Z$  corresponde a una ley normal estándar.

## 1.2. Caso de la varianza desconocida

Si la varianza es desconocida, para determinar el error estándar de la media, aproximamos la desviación estándar  $\sigma$  por la desviación estándar muestral:

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

En este caso, bajo la hipótesis nula, la variable:

$$\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s}$$

sigue una ley  $t$  de Student con  $n - 1$  grados de libertad.

Por tanto, para hacer contrastes sobre la media con  $\sigma$  desconocida, el estadístico de contraste es éste:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

que se corresponde, bajo la hipótesis nula, a una observación de una  $t$  de Student con  $n - 1$  grados de libertad.

Como antes, el último paso es el cálculo del  $p$ -valor asociado a nuestro valor observado del estadístico de contraste. En este caso:

- Si  $H_1: \mu \neq \mu_0$ , entonces (probabilidad de las dos colas):

$$p = P(|t_{n-1}| > |t|) = P(t_{n-1} > |t|) + P(t_{n-1} < -|t|)$$

- Si  $H_1: \mu < \mu_0$ , entonces (probabilidad de la cola de la izquierda):

$$p = P(t_{n-1} < t)$$

- Si  $H_1: \mu > \mu_0$ , entonces (probabilidad de la cola de la derecha):

$$p = P(t_{n-1} > t)$$

donde recordemos que  $t_{n-1}$  sigue una ley  $t$  de Student con  $n - 1$  grados de libertad.

### Ejemplo de contraste sobre la media en caso de varianza desconocida

Se mide el ancho de una muestra aleatoria de 64 maderas y se obtiene una media de 35,02 y una desviación estándar muestral de 0,1224. Supongamos que sabemos que el ancho de las maderas sigue una distribución normal. Queremos contrastar la hipótesis de que la media verdadera es 35 con un nivel de significación de  $\alpha = 0,05$ .

En este caso consideramos las hipótesis siguientes:

$$H_0: \mu = 35 \text{ contra } H_1: \mu \neq 35$$

Dado que la varianza es desconocida, el error estándar es:

$$\frac{s}{\sqrt{n}} = 0,1224 \sqrt{\frac{1}{64}} = 0,0153$$

y el estadístico de contraste es:

$$\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{35,02 - 35}{0,0153} = 1,307$$

Dado que corresponde a una distribución  $t$  de Student con 63 grados de libertad y la hipótesis alternativa es bilateral, el  $p$ -valor es (utilizando el *software* estadístico):

$$P(|t_{63}| > 1,307) = 2(1 - P(t_{63} \leq 1,307)) = 2(1 - 0,9020) = 0,1960$$

Por tanto, debemos aceptar la hipótesis nula, lo que significa que con los datos que tenemos no podemos afirmar que la media no sea 35.

Si quisiéramos resolver el contraste utilizando el valor crítico, deberíamos determinar un valor  $z_{\alpha/2}$  tal que:


$$P\left(|t_{63}| \geq z_{\frac{\alpha}{2}}\right) = 0,05$$

Si buscamos este valor obtenemos (para calcular esta probabilidad, hay que utilizar el *software* estadístico):

$$z_{\frac{\alpha}{2}} = 1,9983$$

Puesto que  $1,307 \in (-1,9983; 1,9983)$ , obtenemos también que debemos aceptar la hipótesis nula.


### 1.3. Caso con muestras no normales

Si no sabemos si las muestras provienen de una distribución normal, también podemos hacer un contraste de hipótesis por la media, siempre que tengamos una muestra de tamaño mayor de 30,  $n > 30$  (el caso  $n < 30$  con muestras que no provienen de una ley normal no se estudia en esta sesión). Supondremos también que no conocemos su varianza. 

En esta situación, por el teorema del límite central, sabemos que podemos aproximar la variable:

$$\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s}$$

a una distribución  $N(0,1)$ . A partir de aquí podemos acabar el test como en el caso del contraste sobre la media con varianza conocida.

Observad el contraste sobre la media en el caso de la varianza conocida en el apartado 1.1 de esta sesión. 

#### Ejemplo de contraste sobre la media con muestras no normales

En un laboratorio se mide el tiempo de vida (tiempo de funcionamiento hasta que se estropea) de un tipo de bombillas. El experimento se repite 300 veces y se obtiene una media muestral de 5,14 y una desviación típica muestral de 0,25 (las unidades son semanas). Hacemos el contraste:

$$H_0: \mu = 5 \text{ contra } H_1: \mu \neq 5$$

Sabemos que los tiempos de vida normalmente no siguen una distribución normal porque sus valores se ajustan mejor a una exponencial. De todos modos, puesto que tenemos una muestra lo bastante grande, podemos aplicar el tipo de contraste expuesto. El error estándar de la media es éste:

$$\frac{s}{\sqrt{n}} = 0,25 \sqrt{\frac{1}{300}} = 0,0144$$

y el estadístico de contraste:

$$\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{5,14 - 5}{0,0144} = 9,722$$

que sigue una normal estándar. Tenemos una hipótesis bilateral, así que el  $p$ -valor es:

$$P(|Z| > 9,722) = 2P(Z > 9,722) = 0,000$$

y debemos rechazar la hipótesis nula, lo que significa que estamos seguros de que el tiempo de vida esperado no es 5, sea cual sea el nivel de significación escogido 0,1, 0,05 o, incluso, 0,001.

## 1.4. Datos aparejados

Todos los contrastes que hemos visto hasta ahora se han planteado utilizando sólo una muestra. Podemos utilizar también estas mismas técnicas cuando tenemos dos muestras con datos aparejados.

Imaginad que queremos estudiar si un rato de relajación ayuda a mejorar la productividad de los trabajadores. Podemos hacer este estudio de dos maneras:

1) Podemos coger una muestra de trabajadores que han hecho relajación y una muestra de trabajadores que no la han hecho y comparar la productividad de los dos grupos (datos no aparejados).

2) Podemos coger un único grupo de trabajadores, medir su productividad sin haber hecho relajación y volver a medirla al cabo de algunos días, durante los cuales sí han hecho ejercicios de relajación diarios. Entonces podemos estudiar la variación de la productividad (datos aparejados).

Fijaos en que en el segundo caso utilizamos a los mismos individuos para obtener las dos muestras, de manera que si observamos la diferencia individuo a individuo, podemos pasar a estudiar una única muestra, que será el incremento de la productividad.

Una muestra de **datos aparejados** es un conjunto de observaciones de dos variables diferentes observadas en los mismos individuos, de manera que por cada individuo tenemos dos observaciones.

### Datos aparejados de muestras normales

Si los datos aparejados provienen de muestras normales, al hacer la diferencia volvemos a obtener datos que provienen de una ley normal.

### Ventaja de los datos aparejados

Los datos aparejados presentan la ventaja de que eliminan la variabilidad a causa del hecho de utilizar individuos diferentes en las dos muestras.

Ahora podemos aplicar los contrastes vistos en los subapartados anteriores a la diferencia de las dos muestras.

### Ejemplo de contraste con datos aparejados

Después de un tratamiento de seis meses, la bilirrubina total de cuatro enfermos presentó la variación siguiente (medidas en mg/ml):

	Inicial	Final
Individuo 1	2,3	2,2
Individuo 2	2,0	1,9
Individuo 3	1,8	1,8
Individuo 4	2,8	2,4

Suponemos normalidad y queremos estudiar con un nivel de significación  $\alpha = 0,01$  si se puede aceptar que la bilirrubina ha disminuido a causa de este tratamiento. Puesto que se trata de pruebas que se han hecho sobre los mismos enfermos, podemos decir que se trata de muestras aparejadas.

Primero convertimos los datos de las dos muestras en los de una sola, resultado de la diferencia:

Diferencia
0,1
0,1
0
0,4

Para estos datos calculamos  $\bar{x} = 0,15$  y  $s = 0,1732$ . Formulamos las hipótesis de contraste: consideraremos que la hipótesis nula es aquella en la que no ha habido una disminución de la bilirrubina, mientras que la alternativa considera que sí que la ha habido (así, si rechazamos la hipótesis nula, estaremos bastante seguros de que el tratamiento hace disminuir la bilirrubina). Por tanto, tenemos el contraste siguiente:

$$H_0: \mu = 0 \text{ contra } H_1: \mu > 0$$

Puesto que la varianza es desconocida, el error estándar de la media es éste:

$$\frac{s}{\sqrt{n}} = 0,1732 \sqrt{\frac{1}{4}} = 0,0866$$

y el estadístico de contraste:

$$\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0,15}{0,0866} = 1,732$$

Dado que corresponde a una distribución  $t$  de Student con tres grados de libertad (el tamaño de la muestra es 4) y la hipótesis alternativa es unilateral, el  $p$ -valor es el siguiente:

$$P(t_3 > 1,732) = 1 - P(t_3 \leq 1,732) = 1 - 0,9092 = 0,0908$$

Por tanto, debemos aceptar la hipótesis nula, lo que significa que con los datos que tenemos no podemos afirmar que el tratamiento haga disminuir la bilirrubina.

## 2. Contrastes sobre la proporción (muestras grandes)

Consideremos ahora que tenemos una muestra  $x_1, \dots, x_n$ , que proviene de una distribución de Bernoulli de parámetro  $p$  y queremos hacer un contraste sobre el parámetro. Necesitaremos también que el tamaño de la muestra sea lo bastante grande, como mínimo  $n > 30$ .

### El parámetro $p$

Recordad que muchas veces  $p$  indica la probabilidad de éxito en un experimento.

La hipótesis nula será del tipo:

$$H_0: p = p_0$$

donde  $p_0$  es una constante fijada. Podemos tener dos tipos de hipótesis alternativa:

- $H_1: p \neq p_0$  (bilateral)
- $H_1: p < p_0$  o  $H_1: p > p_0$  (unilateral)

Una vez establecidas las hipótesis, hay que determinar el estadístico de contraste.

En este caso la media muestral corresponde a la frecuencia, que indicamos por  $\hat{p}$ . Puesto que la varianza de una distribución de Bernoulli de parámetro  $p_0$  es  $p_0(1 - p_0)$ , el error estándar de la media se puede calcular a partir de la expresión siguiente:

$$\frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

Por el teorema del límite central, el estadístico:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{\sqrt{n}(\hat{p} - p_0)}{\sqrt{p_0(1 - p_0)}}$$

se puede aproximar por una distribución normal estándar.

Por tanto, en el caso de las proporciones, el estadístico de contraste es éste:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

que sigue una distribución normal estándar bajo la hipótesis nula.

El cálculo del  $p$ -valor se efectúa a partir de aquí igual como se ha discutido en el caso de la varianza conocida.

Observad el cálculo del  $p$ -valor efectuado en el apartado 1.1 de esta sesión.



### Ejemplo de cálculo del $p$ -valor sobre proporciones

En una encuesta realizada a 1.000 barceloneses, 350 de los encuestados manifestaban practicar alguna actividad deportiva. ¿Podemos considerar que una tercera parte de los barceloneses practica alguna actividad deportiva contra que sean menos de una tercera parte los que practican actividades deportivas, con un nivel de significación de  $\alpha = 0,1$ ?

Fijaos en que queremos hacer el contraste siguiente:

$$H_0: p = 0,33 \text{ contra } H_1: p < 0,33$$

ya que en la hipótesis alternativa nos interesa verificar sólo si la proporción que practica una actividad deportiva es inferior al 33%.

Calculemos ahora el error estándar de la proporción:

$$\sqrt{\frac{0,33(1 - 0,33)}{1.000}} = 0,0149$$

Puesto que la frecuencia observada es ésta:

$$\hat{p} = \frac{350}{1.000} = 0,35$$

el valor del estadístico de contraste observado es éste:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,35 - 0,33}{0,0148} = 1,345$$

Y si  $Z$  es una normal estándar, el  $p$ -valor correspondiente es éste:

$$P(Z < 1,345) = 1 - P(Z \geq 1,345) = 1 - 0,0901 = 0,9099$$

Dado que este valor es mayor que  $\alpha$ , podemos aceptar la hipótesis nula y decir que una tercera parte de los barceloneses practica alguna actividad deportiva.

### 3. Resumen

En esta sesión hemos aprendido a hacer contrastes de hipótesis para la media y para proporciones. Para la media hemos distinguido según si los datos provienen o no de una distribución normal y según si conocemos el valor de la varianza poblacional o no.



## Ejercicios

1. Para estudiar los diámetros de unos ejes cogemos una muestra de tamaño 100. La media muestral obtenida es de 38 y la varianza muestral, 4. ¿Podemos decir con un nivel de significación del 0,05 que la media de los diámetros de los ejes es de 40?
2. Las calificaciones de los exámenes de estadística y de álgebra de un grupo de estudiantes matriculados en las dos asignaturas han sido éstas:

Estudiante	Estadística	Álgebra
1	7,6	7,1
2	4,8	3,0
3	5,1	4,2
4	7,5	5,2
5	4,3	4,8
6	4,1	3,6
7	5,5	3,9
8	8,3	1,1
9	6,9	5,8

Estudios de los resultados de semestres anteriores han demostrado que las notas siguen una distribución normal.

- a) Contrastad la hipótesis nula de que la media de las notas de estadística sea 5,5 contra que sea menor que 5,5.
- b) Contrastad la hipótesis nula de que las notas de estadística y de álgebra tengan la misma media contra que sean diferentes.

3. Compramos una máquina que fabrica chips que, según el fabricante, como máximo fabrica un 1% de chips defectuosos. Después de fabricar 300 chips, los revisamos y descubrimos que hay cinco defectuosos. Contrastad si podemos decir que el porcentaje de chips defectuosos es, como mucho, del 1% con un nivel de significación  $\alpha = 0,1$ .

## Solucionario

1. Queremos hacer el contraste siguiente:

$$H_0: \mu = 40 \text{ contra } H_1: \mu \neq 40$$

En principio, no sabemos si los datos provienen de una distribución normal, pero como que la muestra es bastante grande ( $> 30$ ), podemos hacer un contraste. El error estándar de la media es éste:

$$\frac{2}{\sqrt{100}} = 0,2$$

y, por tanto, el estadístico de contraste vale:

$$Z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{38 - 40}{0,2} = -10$$

El  $p$ -valor correspondiente para nuestro contraste bilateral es ( $Z$  indica una normal estándar):

$$P(|Z| > 10) = 2P(Z > 10) \cong 0,000$$

de manera que debemos rechazar la hipótesis nula, es decir, estamos seguros de que la media no es 40.W

2.

a) En este caso consideramos el contraste siguiente:

$$H_0: \mu = 5,5 \text{ contra } H_1: \mu < 5,5$$

A partir de los datos calculamos la media y la varianza muestral:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 6,01; s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 1,57$$

Así, el error estándar es éste:

$$\frac{s}{\sqrt{n}} = 1,57 \sqrt{\frac{1}{9}} = 0,525$$

y el estadístico de contraste:

$$\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{6,01 - 5,5}{0,525} = 0,971$$

Los datos provenían de una distribución normal y la varianza es desconocida; por tanto, bajo la hipótesis nula, el estadístico de contraste corresponde a una distribución  $t$  de Student con 8 grados de libertad. Así, el  $p$ -valor es éste:

$$P(t_8 < 0,971) = 0,82$$

Por tanto, debemos aceptar la hipótesis nula.

b) Dado que tenemos datos aparejados, primero de todo calculamos las diferencias:

Estudiante	Estadística	Álgebra	Diferencia
1	7,6	7,1	0,5
2	4,8	3,0	1,8
3	5,1	4,2	0,9
4	7,5	5,2	2,3
5	4,3	4,8	-0,5
6	4,1	3,6	0,5
7	5,5	3,9	1,6
8	8,3	1,1	7,2
9	6,9	5,8	1,1

A partir de aquí podemos hacer un contraste como el del apartado a, teniendo en cuenta que ahora queremos hacer el contraste siguiente:

$$H_0: \mu = 0 \text{ contra } H_1: \mu \neq 0$$

Calculamos la media y la desviación típica muestral:

$$\bar{x} = 1,71 \quad s = 2,21$$

El error estándar es el siguiente:

$$\frac{s}{\sqrt{n}} = 2,21 \sqrt{\frac{1}{9}} = 0,73$$

y el estadístico de contraste:

$$\frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{1,71}{0,73} = 2,34$$

Como en el apartado a, el estadístico de contraste corresponde a una distribución  $t$  de Student con 8 grados de libertad. Ahora el  $p$ -valor es éste:

$$P(|t_8| > 2,34) = 2(1 - P(t_8 \leq 2,34)) = 2(1 - 0,9763) = 0,0474$$

Por tanto, con un nivel de significación de 0,05, debemos rechazar la hipótesis nula y, en consecuencia, podemos asegurar que las medias son diferentes. En cambio, si cogiéramos un nivel de 0,01, deberíamos aceptarla, es decir, con este nivel todavía no tenemos suficientes evidencias para decir que las dos medias son diferentes.

3. Queremos hacer un contraste para una proporción con una muestra bastante grande. Si llamamos  $p$  a la proporción de chips defectuosos, planteamos el contraste siguiente:

$$H_0: p = 0,01 \text{ contra } H_1: p > 0,01$$

Ahora calculamos el error estándar:

$$\sqrt{\frac{0,01(1-0,01)}{300}} = 0,0057$$

Dado que la proporción poblacional es ésta:

$$\hat{p} = \frac{5}{300} = 0,0167$$

el valor observado del estadístico de contraste es el siguiente:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,0167 - 0,01}{0,0057} = 1,1754$$

Y si  $Z$  es una normal estándar, el  $p$ -valor correspondiente es éste:

$$P(Z > 1,1754) = 1 - P(Z \leq 1,1754) = 1 - 0,8801 = 0,1199$$

Por tanto, puesto que es mayor que  $\alpha$ , podemos aceptar la hipótesis nula y decir que la proporción de chips defectuosos es del 1%.