Tema 8. Contrastes paramétricos en una población

Resumen del tema

8.1. Contrastes sobre la media

8.1.1. Varianza poblacional conocida

condiciones	estadístico	contraste	región crítica
• Muestra aleatoria simple de tamaño n. • σ conocida. • Población Normal ó población cualquiera siempre que $n \geq 30$.	$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$Z < -Z_{1-\alpha/2}$ $Z > Z_{1-\alpha/2}$
		$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$Z < -Z_{1-\alpha}$
		$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$Z > Z_{1-\alpha}$

8.1.2. Varianza poblacional desconocida

condiciones	estadístico	contraste	región crítica
• Muestra aleatoria simple de tamaño n. • σ desconocida. • Población Normal ó población cualquiera siempre que $n \geq 30$.	$T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$T < -t_{n-1\,,\,1-\alpha/2}$ $T > t_{n-1\,,\,1-\alpha/2}$
		$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$T < -t_{n-1\,,\,1-\alpha}$
		$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$T > t_{n-1\,,\,1-\alpha}$

8.2. Contrastes sobre la varianza

8.2.1. Media poblacional conocida

condiciones	estadístico	contraste	región crítica
• Muestra aleatoria simple: $X_1, X_2, \ldots, X_n$. • μ conocida. • Población Normal.	$U = \dfrac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma_0^2}$	$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$U < \chi^2_{n\,,\,\alpha/2}$ $U > \chi^2_{n\,,\,1-\alpha/2}$
		$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$U < \chi^2_{n\,,\,\alpha}$
		$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$U > \chi^2_{n\,,\,1-\alpha}$

8.2.2. Media poblacional desconocida

condiciones	estadístico	contraste	región crítica
• Muestra aleatoria simple de tamaño n. • μ desconocida. • Población Normal.	$V = \dfrac{(n-1)S^2}{\sigma_0^2} = \dfrac{ns^2}{\sigma_0^2}$	$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$V < \chi^2_{n-1\,,\alpha/2}$ $V > \chi^2_{n-1\,,1-\alpha/2}$
		$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$V < \chi^2_{n-1\,,\alpha}$
		$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$V > \chi^2_{n-1\,,1-\alpha}$

8.3. Contrastes sobre la proporción

condiciones	estadístico	contraste	región crítica
• Muestra aleatoria simple de tamaño n. • $n \geq 30$, • $0'1 \leq p_0 \leq 0'9$.	$Z = \dfrac{(X - np_0) \pm 0'5}{\sqrt{np_0(1-p_0)}}$ con X=n$^{\text{o}}$ de éxitos en la muestra	$H_0 : p = p_0$ $H_1 : p \neq p_0$	$Z < -Z_{1-\alpha/2}$ $Z > Z_{1-\alpha/2}$
		$H_0 : p \geq p_0$ $H_1 : p < p_0$	$Z < -Z_{1-\alpha}$
		$H_0 : p \leq p_0$ $H_1 : p > p_0$	$Z > Z_{1-\alpha}$
El signo "±" significa "−" si $(X - np_0) > 0$ y "+" si $(X - np_0) < 0$			

Ejemplos que se van a resolver en clase

Ejemplo 8.1. Retomamos los datos del Ejemplo 7.1: En la tabla siguiente aparecen los datos de 10 bibliotecas, en las cuales se ha observado las siguientes variables: número total de títulos catalogados en un año (X), número de horas totales al año que emplea la biblioteca en catalogar sus títulos (Y) y costo, en euros, de una hora de catalogación (Z).

x_i	y_i	z_i
1550	220	15’75
1640	230	14’50
1000	140	16’40
950	135	16’70
750	110	17’10
1700	255	12’50
1650	228	14’80
1860	270	15’25
1900	280	18’50
900	130	17’30

$\sum_{i=1}^{10} z_i = 158'8$	$\sum_{i=1}^{10} z_i^2 = 2547'965$

a) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'01$, que la media poblacional del costo de una hora de catalogación es menor que 17 euros?

b) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'01$, que la desviación típica poblacional del costo de una hora de catalogación es mayor que 2 euros?

Ejemplo 8.2. En la tabla siguiente aparecen los resultados del peso, en gramos, (X) y del precio, en euros, (Y) de una muestra de 12 libros.

x_i	y_i
325	110
890	30
415	75
400	45
515	32
650	69
790	30
890	34
320	42
420	46
620	53
720	97

$\sum_{i=1}^{12} y_i = 663$	$\sum_{i=1}^{12} y_i^2 = 44589$

a) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'05$, que la muestra de datos de la variable Y es aleatoria?

b) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'02$, que la variable aleatoria Y es Normal?

c) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'01$, que la media poblacional del precio es igual a 55 euros?

d) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'01$, que la desviación típica poblacional del precio es igual a 24 euros?

Ejemplo 8.3. Deseamos conocer la postura de los bibliotecarios frente a la informatización de las bibliotecas. Para ello, interrogamos a 150 de ellos (elegidos aleatoria e independientemente) sobre este tema, obligándoles a manifestarse a favor o en contra. El resultado es que 82 se manifiestan a favor y consiguientemente, 68 en contra. ¿Es compatible este resultado con que la proporción de bibliotecarios (en el colectivo total) a favor de informatizar las bibliotecas es la misma que la proporción de bibliotecarios en contra?

Problemas propuestos

Problema 8.1. El número medio de libros por estante en una biblioteca es de $24'4$, con una desviación típica de $1'6$. Una muestra aleatoria simple de 36 estantes de dicha biblioteca tiene una media de $25'2$ libros por estante. ¿La información proporcionada por la muestra es representativa de toda la población?

Problema 8.2. El número medio recomendado de usuarios servidos semanalmente por cada miembro del personal de una biblioteca es de 100. En una muestra aleatoria simple de 81 miembros del personal de las bibliotecas de una determinada región se obtiene una media de $132'88$ usuarios servidos semanalmente, con una cuasidesviación típica de $55'19$. ¿Las bibliotecas de dicha región siguen la recomendación mencionada?

Problema 8.3. El precio medio de los libros en rústica es de $63'4$ euros, con una desviación típica de $14'8$ euros. Una muestra aleatoria simple de 61 libros en rústica con ilustraciones en color tiene un precio medio de $69'5$ euros, con una cuasidesviación típica de $16'6$ euros.

a) ¿Permiten los datos afirmar que los libros en rústica con ilustraciones en color son más caros que el resto de libros en rústica?

b) ¿La varianza del precio de los libros en rústica con ilustraciones en color es mayor que la del precio de los libros en rústica?

Problema 8.4. Se sabe que el número medio de veces que un artículo científico es citado durante los 5 siguientes años a su publicación es de $6'5$. Se eligen aleatoria e independientemente 71 artículos de medicina, obteniéndose una media de $7'8$ citas durante los 5 siguientes años a su publicación, con una cuasidesviación típica de $2'3$. ¿Se puede afirmar que durante los 5 siguientes años a su publicación se citan más los artículos de medicina que el resto de artículos científicos?

Problema 8.5. En una muestra aleatoria simple de 15 individuos que consultan bases de datos, el tiempo (en minutos) que están utilizando el ordenador para realizar esta tarea es:

22	13	17	14	15	18	19	14	17	20	21	13	15	18	17

a) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'05$, que la muestra es aleatoria?

b) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'05$, que la variable aleatoria "tiempo empleado en consultar bases de datos por ordenador" es Normal?

c) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'05$, que la media poblacional del tiempo empleado en consultar bases de datos por ordenador es mayor que 15 minutos?

d) ¿Se puede aceptar, con un nivel de significación de $\alpha = 0'05$, que la desviación típica poblacional del tiempo empleado en consultar bases de datos por ordenador es menor que 2 minutos?

Problema 8.6. En una biblioteca desconocemos la proporción de libros escritos en español. De una muestra aleatoria simple de 125 libros, 80 de ellos están escritos en español, y el resto en otros idiomas. Según estos datos, ¿se puede afirmar que la proporción de libros escritos en español en dicha biblioteca es mayor que $0'6$?

Problema 8.7. En una biblioteca escolar hay una proporción de libros prestados que se devuelven con retraso. De una muestra aleatoria simple de 250 libros, 50 de ellos se han devuelto con retraso. ¿Permiten los datos afirmar que la proporción de libros prestados que se devuelven con retraso a dicha biblioteca escolar es mayor que $0'15$?

Soluciones de los problemas propuestos

Solución del problema 8.1. Sea X=*Número de libros por estante de la biblioteca.* Aunque μ es conocida, para averiguar si la información proporcionada por la muestra es representativa de toda la población, tenemos que hacer el contraste en el que la hipótesis nula es $H_0 : \mu = 24'4$. Si aceptamos la hipótesis nula (que sabemos que es verdadera) entonces la información proporcionada por la muestra es representativa de toda la población; en caso contrario, no lo es.

Como σ es conocida, el valor del estadístico de contraste es $Z = 3$. Si tomamos un nivel de significación de $\alpha = 0'05$, entonces la región crítica es $Z < -1'96$ ó $Z > 1'96$. En consecuencia, rechazamos H_0 y, por tanto, la información proporcionada por la muestra no es representativa de toda la población. Finalmente, la respuesta a la pregunta es **NO**.

Solución del problema 8.2. Sea X=*Número de usuarios servidos semanalmente por cada miembro del personal de la biblioteca.* Hacemos un contraste sobre μ, con σ desconocida. La hipótesis nula es $H_0 : \mu = 100$. El valor del estadístico de contraste es $T = 5'3618$. Si tomamos un nivel de significación de $\alpha = 0'05$, entonces la región crítica es $T < -1'9901$ ó $T > 1'9901$. En consecuencia, rechazamos H_0 y, por tanto, las bibliotecas de dicha región no siguen la recomendación. Finalmente, la respuesta a la pregunta es **NO**.

Solución del problema 8.3. Sea X=*Precio de los libros en rústica con ilustraciones color.*

a) Hacemos un contraste sobre μ, con σ desconocida. La hipótesis nula es $H_0 : \mu \leq 63'4$. El valor del estadístico de contraste es $T = 2'8700$. Si tomamos un nivel de significación de $\alpha = 0'05$, entonces la región crítica es $T > 1'6706$. En consecuencia, rechazamos H_0 y, por tanto, los libros en rústica con ilustraciones en color son más caros (tienen un precio medio mayor) que el resto de los libros en rústica. Finalmente, la respuesta a la pregunta es **SÍ**.

b) Hacemos un contraste sobre σ^2, con μ desconocida. La hipótesis nula es $H_0 : \sigma^2 \leq (14'8)^2$. El valor del estadístico de contraste es $V = 75'4821$. Si tomamos un nivel de significación de $\alpha = 0'05$, entonces la región crítica es $V > 79'0819$. En consecuencia, aceptamos H_0 y, por tanto, no se puede aceptar que la varianza del precio de los libros en rústica con ilustraciones en color sea mayor que la varianza del precio de todos los libros en rústica. Finalmente, la respuesta a la pregunta es **NO**.

Solución del problema 8.4. Sea X=*Número de veces que los artículos de medicina son citados durante los cinco siguientes años a su publicación.* Hacemos un contraste sobre μ, con σ desconocida. La hipótesis nula es $H_0 : \mu \leq 6'5$. El valor del estadístico de contraste es $T = 4'7626$. Si tomamos un nivel de significación de $\alpha = 0'05$, entonces la región crítica es $T > 1'6669$. En consecuencia, rechazamos H_0 y, por tanto, se citan más los artículos de medicina que el resto de artículos científicos (la media del número de citas es mayor). Finalmente, la respuesta a la pregunta es **SÍ**.

Solución del problema 8.5. Sea X=*Tiempo empleado en consultar bases de datos por ordenador.*

a) Hacemos el contraste de las rachas sobre aleatoriedad de la muestra en el que la hipótesis nula es H_0 :*La muestra de datos de la variable X es aleatoria.* El valor del estadístico de contraste es $R = 10$. Como el nivel de significación es $\alpha = 0'05$, entonces la región de aceptación es el intervalo $(3, 10)$. Por tanto, aceptamos H_0. Finalmente, la respuesta a la pregunta es **SÍ**.

b) Hacemos el contraste de D'Agostino sobre normalidad en el que la hipótesis nula es H_0 :*La variable aleatoria X es Normal.* El valor del estadístico de contraste es $D_{exp} = 0'284074$.

Como el nivel de significación es $\alpha = 0'05$, entonces la región de aceptación es el intervalo $(0'2568, 0'2858)$. Por tanto, aceptamos H_0. Finalmente, la respuesta a la pregunta es **SÍ**.

c) Hacemos un contraste sobre μ, con σ desconocida. La hipótesis nula es $H_0 : \mu \leq 15$. El valor del estadístico de contraste es $T = 2'536486$. Como el nivel de significación es $\alpha = 0'05$, entonces la región crítica es $T > 1'7613$. En consecuencia, rechazamos H_0 y, por tanto, la media del tiempo empleado en consultar bases de datos por ordenador es mayor que 15 minutos. Finalmente, la respuesta a la pregunta es **SÍ**.

d) La pregunta que se nos hace es ¿$\sigma < 2$? Esta pregunta es equivalente a ¿$\sigma^2 < 2^2$? Por tanto, hacemos un contraste sobre σ^2, con μ desconocida. La hipótesis nula es $H_0 : \sigma^2 \geq 2^2$. El valor del estadístico de contraste es $V = 28'4\,\hat{3}$. Como el nivel de significación es $\alpha = 0'05$, entonces la región crítica es $V < 6'57063$. En consecuencia, aceptamos H_0 y, por tanto, no se puede aceptar que la desviación típica (poblacional) del tiempo empleado en consultar bases de datos por ordenador es menor que 2 minutos. Finalmente, la respuesta a la pregunta es **NO**.

Solución del problema 8.6. Hacemos un contraste sobre la proporción poblacional, p=proporción de libros escritos en español en la biblioteca. La hipótesis nula es $H_0 : p \leq 0'6$. El valor del estadístico de contraste es $Z = 0'821584$. Si tomamos un nivel de significación de $\alpha = 0'05$, entonces la región crítica es $Z > 1'6449$. En consecuencia, aceptamos H_0 y, por tanto, no se puede aceptar que la proporción de libros escritos en español sea mayor que $0'6$. Finalmente, la respuesta a la pregunta es **NO**.

Solución del problema 8.7. Hacemos un contraste sobre la proporción poblacional, p=proporción de libros prestados que se devuelven con retraso. La hipótesis nula es $H_0 : p \leq 0'15$. El valor del estadístico de contraste es $Z = 2'125476$. Si tomamos un nivel de significación de $\alpha = 0'05$, entonces la región crítica es $Z > 1'6449$. En consecuencia, rechazamos H_0 y, por tanto, la proporción de libros prestados que se devuelven con retraso a la biblioteca es mayor que $0'15$. Finalmente, la respuesta a la pregunta es **SÍ**.