

Introducción a la Estadística y Ciencia de Datos

Práctica 5 - Test de hipótesis

Tests exactos para una muestra

1. Una empresa vende dos variedades de soja. La variedad 1 tiene un rendimiento por hectárea que puede considerarse una variable aleatoria con distribución $\mathcal{N}(37, 25)$, y la variedad 2 tiene un rendimiento por hectárea que puede considerarse $\mathcal{N}(40, 25)$. Un cliente realizó una compra de semillas de la variedad 2 y antes de continuar comprando a esta empresa, quiere asegurarse de que las semillas que le enviaron realmente pertenecen a esa variedad. Con ese fin, cultiva 10 parcelas de 1 hectárea y obtiene los siguientes rendimientos:

37 39,5 41,7 42 40 41,25 43 44,05 38 38,5

El cliente quiere que la probabilidad de seguir comprando a esta empresa cuando las semillas no son de la variedad pedida sea 0,05.

- a) Hallar el test más potente de nivel α para este problema. ¿Qué decisión toma?
 - b) Hallar el valor p . ¿Se hubiera rechazado H_0 para $\alpha = 0,01$?
 - c) Calcular la probabilidad de cometer un error de tipo II.
 - d) Determinar el número n de parcelas a cultivar para que el error de tipo II tenga probabilidad menor o igual que 0,05.
2. Se mide el grado de impurezas de un producto químico. El método de medición está afectado por un error que se supone $N(0, \sigma^2)$, σ^2 desconocido. Además, los errores correspondientes a diferentes mediciones son independientes entre sí. Se hicieron 12 observaciones obteniendo que el promedio es 0,85 con un desvío estándar muestral de 0,05.
 - a) A partir de estos datos, ¿hay evidencia significativa para decir que el grado de impurezas del producto es distinto de 0,7 a nivel 0,05? Hallar el p -valor.
 - b) Consideremos las hipótesis $H_0 : \sigma = 0,03$ vs. $H_1 : \sigma > 0,03$. ¿Hay evidencia para rechazar H_0 a nivel 0.05? Calcular el p -valor.
 3.
 - a) Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución $\mathcal{N}(\mu_0, \sigma^2)$ con μ_0 conocido. Encontrar el test UMP de nivel α para $H_0 : \sigma^2 \geq \sigma_0^2$ vs $H_1 : \sigma^2 < \sigma_0^2$.
 - b) Para medir la concentración de una sustancia en una solución se conoce un método cuyo error es una variable aleatoria con distribución $\mathcal{N}(0, 1)$. Se propone un nuevo método cuyo error también es normal con media cero pero varianza σ^2 desconocida. Se adoptará este nuevo método si es más preciso que el anterior. Para decidir si se adoptará el nuevo método se tomaron 21 mediciones y se obtuvo $\frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2 = 0,60$.
 - i. Se quiere que la probabilidad de cambiar de método si el nuevo en realidad es menos preciso sea a lo sumo del 1 %. ¿Adoptaría o no el nuevo método?
 - ii. Calcular el p -valor.

- iii. Calcular la probabilidad de quedarse con el viejo método de medición cuando la varianza del nuevo es en realidad 0.80.
4. Consideremos un procedimiento para medir el contenido porcentual de manganeso en un mineral. A este procedimiento se lo ha usado muchas veces y se sabe que las mediciones siguen una distribución normal cuya desviación estándar (poblacional) es 0.15. Se está estudiando si el método tiene un error sistemático (es decir, si su media es distinta de lo que debería ser).
- a) Se hacen 8 mediciones de un mineral preparado que tiene 7% de manganeso y se obtienen los siguientes resultados:
- 6,90 7,10 7,20 7,07 7,15 7,04 7,18 6,95 ($\bar{x} = 7,074$)
- ¿Cuál es su conclusión si desea que la probabilidad de equivocarse al decir que el método tiene error sistemático cuando en realidad no lo tiene sea de 0,05?
- b) Si el método tiene un error sistemático de 0,10 (o sea, si la media de las mediciones es 7,10), ¿cuál es la probabilidad de cometer un error de tipo II?
- c) Se quiere aplicar un test estadístico de modo que, al igual que en el inciso a), la probabilidad de decir que hay error sistemático cuando no lo hay sea 0,05. Además, se desea que si hay un error sistemático de 0,10, la probabilidad de detectarlo sea mayor o igual a 0,80.
- i. El test del inciso a) ¿cumple con este requisito?
- ii. En caso contrario, ¿cuántas mediciones habría que hacer como mínimo para que cumpla con el requisito?
5. (Para resolver en clase, con R) Sean X_1, \dots, X_n variables aleatorias i.i.d. con distribución $N(\mu, \sigma^2)$ con $\sigma = 20$ y $n = 16$, $\mu_0 = 100$. Para las siguientes hipótesis

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

- a) Escribir el test de nivel $\alpha = 0,05$ en términos del promedio \bar{X}_n .
- b) Graficar las funciones de densidad de \bar{X}_n para $\mu_0 = 100$ y $\mu = 110$ en colores distintos y colorear el nivel del test y la potencia.
- c) ¿Qué pasa si tomamos $n = 36$? Repetir (b) superponiendo las densidades al gráfico anterior.
6. Una asociación de consumidores, preocupada por el porcentaje de grasa contenida en una marca de hamburguesas, envía a un laboratorio una muestra aleatoria de 12 hamburguesas para su análisis. El porcentaje de grasa en cada una de las hamburguesas de la muestra resultó:

21 18 19 16 18 24 22 19 24 14 18 15

El fabricante afirma que el contenido medio de grasa de este tipo de hamburguesas es a lo sumo del 18%.

- a) La asociación de consumidores quiere saber si tiene motivos para decir que la afirmación del fabricante es falsa. Asumiendo que el contenido porcentual de grasa de cada hamburguesa de esta marca es una variable aleatoria con distribución normal, proponer un test para resolver este problema. Calcular el p -valor. ¿Rechazaría la hipótesis nula a nivel 0.08?
- b) Intentar calcular la potencia del test cuando el contenido medio de grasa es 20 %. Respecto del ejercicio anterior, ¿qué dificultad aparece cuando uno intenta hacer esta cuenta?

Sugerencia: usar que si $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, entonces

$$\sqrt{n} \frac{\bar{X} - \mu}{s} \sim t_{n-1}$$

donde t_{n-1} es una t de Student con $n - 1$ grados de libertad y $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

7. a) Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución $\mathcal{E}(\lambda)$. El objetivo del ejercicio es encontrar un test UMP de nivel α para las hipótesis $H_0 : \lambda \geq \lambda_0$ vs $H_1 : \lambda < \lambda_0$. Lo haremos paso a paso.
- Encontrar el test de Neyman – Pearson de nivel α basado en la muestra para $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda = \lambda_1$, para un valor de λ_1 fijo, con $\lambda_1 < \lambda_0$.
 - El test obtenido en el ítem (a)[i.], ¿depende del valor que toma λ_1 ? ¿Se puede extender a las hipótesis $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda < \lambda_0$? ¿Este test resulta ser UMP para estas hipótesis?
 - Extender el test obtenido en el ítem (a)[ii.] a las hipótesis $H_0 : \lambda \geq \lambda_0$ vs $H_1 : \lambda < \lambda_0$, estudiando la función de potencia, (a mano o con R).
- b) El tiempo de duración (en días) de cierto tipo de lamparitas es una variable aleatoria con distribución $\mathcal{E}(\lambda)$. La empresa garantiza que el tiempo medio de vida es mayor que 50 días y quiere asegurarse que la producción satisface este requerimiento antes de sacarla a la venta. Para ello toma una muestra de 10 lamparitas y observa un tiempo promedio de duración de 41 días.
- Si se quiere tener un 95 % de seguridad de no vender cuando no se satisfacen los requerimientos, ¿qué decisión se toma en base a estos datos?
 - Calcular el p -valor.
 - ¿Cuál es la probabilidad de no sacar la producción a la venta si el tiempo medio de vida es de 57 días?

Tests asintóticos para una muestra

8. Se ha tomado una muestra aleatoria de tamaño 30 de los días que demora una empresa en instalar una línea telefónica a partir del momento de su solicitud. Los datos obtenidos fueron los siguientes:

3, 4, 5, 3, 2, 7, 6, 4, 6, 4, 7, 2, 3, 4, 6,
7, 3, 5, 6, 6, 4, 5, 5, 7, 3, 2, 2, 4, 3, 5.

La compañía, por contrato, debe tener un valor medio de demora no mayor que 4 días. El gobierno le inició juicio a la compañía porque cree que ésta no cumplió con el contrato. El juez considera que su veredicto debe ser hecho de manera tal que la probabilidad de fallar contra la compañía cuando ésta ha cumplido con lo estipulado en el contrato sea aproximadamente 0,05.

- a) Suponiendo que la distribución de los tiempos de espera es $\mathcal{E}(\lambda)$, encontrar el test asintótico usando el TCL en el cual el juez debe basar su decisión y el veredicto al que arribaría con la muestra dada.
 - b) Expresar la función de potencia del test en términos de una función de distribución conocida.
 - c) ¿Qué tamaño de muestra debería utilizarse si se quiere que la probabilidad de condenar a la compañía sea mayor o igual a 0,99 cuando el valor medio de la demora es de 5 días?
9. Se tiene una muestra X_1, \dots, X_n de tamaño n de una variable aleatoria con distribución $Bi(1, p)$. Se quiere testear $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$ a nivel asintótico 0,05.

- a) Se considera el test con estadístico

$$W = \sqrt{n} \frac{\bar{X} - p_0}{\sqrt{p_0(1 - p_0)}}$$

Hallar el test de nivel asintótico $\alpha = 0,05$ basado en este estadístico.

- b) Se considera el test con estadístico

$$T = \sqrt{n} \frac{\bar{X} - p_0}{\sqrt{\bar{X}(1 - \bar{X})}}$$

Hallar el test de nivel asintótico $\alpha = 0,05$ basado en este estadístico.

- c) Hallar el test del cociente de máxima verosimilitud para este problema, de nivel asintótico $\alpha = 0,05$. ¿Coincide con alguno de los dos tests anteriores?
- d) Consideremos el siguiente test:
 - Hallar el intervalo de confianza de nivel asintótico $\alpha = 0,95$ basado en la muestra X_1, \dots, X_n .
 - Rechazar H_0 si p_0 no pertenece al intervalo de confianza recién calculado.
 Verificar que este test tiene nivel asintótico α . ¿Coincide con alguno de los tests anteriores?
- e) Usando aproximación normal, graficar en R las funciones de potencia aproximada de ambos tests en un mismo gráfico para $n = 50$ y $p_0 = 0,3$. ¿Difieren mucho? ¿Con cuál se quedaría?
- f) (optativo) Calcular mediante una simulación las funciones de potencia $\pi(p)$ en los valores $p = 0,03, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 0,97$ de ambos tests y graficarla interpolando en los valores de p indicados.

10. a) Para curar cierta enfermedad se emplea actualmente un tratamiento que tiene un 40 % de éxito. Un nuevo tratamiento es probado en 10 pacientes elegidos al azar y 9 de ellos se curan.
- Si se quiere que la probabilidad de adoptar el nuevo tratamiento cuando no es mejor que el actual sea 0.05, ¿qué decisión se toma?
 - Calcular el p -valor.
 - ¿Cuál es la probabilidad de no cambiar de tratamiento si el nuevo en realidad tiene una probabilidad de éxito del 45 %?
- b) Seis estudiantes se pusieron a dieta para bajar de peso, con los siguientes resultados:

Nombre	Abdul	Ed	Jim	Max	Phil	Ray
Peso antes (X)	87	95	94	91	100	94
Peso después (Y)	83	93	91	89	102	90

- ¿Puede concluirse que esta dieta es efectiva? ¿Cuál es el p -valor para estos datos? (AYUDA: La dieta es efectiva cuando $P(X > Y) > P(X \leq Y)$.)
- c) Se supone que en cierta población de insectos la cantidad de hembras es igual a la de machos. Para poner a prueba esta creencia, un entomólogo tomó una muestra de 100 de estos insectos, resultando machos 43 de ellos.
- Utilizando un test de nivel asintótico, decidir si hay suficiente evidencia a nivel 0,05 para rechazar la suposición. Calcular el valor p aproximado.
 - Si el porcentaje de hembras fuera 0,55, ¿cuál sería la probabilidad de tomar una decisión incorrecta?
11. El número de personas que llega a la boletería de una estación de trenes entre las 14:00 y 14:30 es una variable aleatoria con distribución de Poisson. El jefe de la estación está evaluando la posibilidad de abrir otra ventanilla en ese horario y considera que es necesario sólo si la media de arribos en ese intervalo de tiempo es superior a 20. Además, quiere que la probabilidad de abrir una nueva ventanilla cuando en realidad no es necesario sea a lo sumo 0,01. Durante 45 días se observó el número de personas que arriba en dicho horario, resultando un promedio de 21 individuos.
- Proponer un test asintótico de nivel α para este problem, usando el TCL. ¿Qué decisión toma en base a estos datos? ¿Cuál es el p -valor aproximado?
 - A partir del test propuesto en el ítem anterior, ¿cuál es la probabilidad de dejar una sola boletería cuando en realidad la media verdadera es de 22 personas?
 - Si se hubiera querido que la probabilidad calculada en (b) fuera a lo sumo 0,05, ¿qué tamaño muestral debería haberse tomado?
12. (Para hacer con R) A un científico le interesa analizar una cierta variable aleatoria X que tiene distribución $N(\mu, 1)$ con μ desconocido. Quiere testear $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$ a nivel $\alpha = 0,05$ (en el fondo, el científico quiere rechazar la hipótesis nula). Para una muestra X_1, \dots, X_n , considera el estadístico $T = \sqrt{n} \bar{X}$, (que bajo H_0 tiene distribución normal

estándar). El científico rechaza H_0 cuando $T > z_{0,05}$, donde z_α es el percentil $1 - \alpha$ de la normal estándar. En principio planea tomar una muestra de tamaño $n = 30$. Supongamos que en este problema en particular, las muestras se van observando secuencialmente y obtener cada una de ellas es muy caro. Por eso, para ver si puede ahorrar dinero, el científico detiene el proceso que obtiene las muestras cuando $n = 20$ y *espía* el valor del estadístico T usando estas 20 observaciones. Al mirar este valor, se da cuenta de que no supera el valor de corte $z_{0,05}$, por lo tanto continúa obteniendo las 10 observaciones restantes (si hubiera observado un valor de T mayor a $z_{0,05}$, hubiera rechazado H_0 sin necesidad de buscar más muestras). Ahora, con las 30 observaciones, vuelve a calcular el valor del estadístico T y en este caso, sí supera el valor de corte $z_{0,05}$. El científico entonces asegura que todo este procedimiento le permite rechazar H_0 a nivel 0,05.

- a) Probar empíricamente, a través de una simulación, que lo que dice el científico es falso. ¿A qué nivel puede rechazar H_0 con este procedimiento?
 - b) Hallar (a través de una simulación) la función de potencia del test correspondiente a este procedimiento, evaluada en los siguientes valores de μ : 0,5; 1; 2; 5; 10.
13. Sean $X_1, \dots, X_n \sim \varepsilon(\lambda)$. Consideramos las hipótesis $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$. Diseñar un test asintótico de nivel α para estas hipótesis usando la distribución asintótica del cociente de máxima verosimilitud.
14. Un dado se tira 600 veces y se obtienen los siguientes resultados:

Número	1	2	3	4	5	6
Frecuencia	87	96	108	89	122	98

¿Puede decirse que el dado no es balanceado? Calcular el valor p aproximado.
Sugerencia: Utilice el test del cociente de máxima verosimilitud.

Tests exactos para dos muestras

15. a) Sea X_1, \dots, X_n una m.a. de una distribución $\mathcal{N}(\mu_1, \sigma^2)$ y sea Y_1, \dots, Y_m una m.a. de una distribución $\mathcal{N}(\mu_2, \sigma^2)$ independiente de la muestra anterior. Si llamamos $\delta = \mu_1 - \mu_2$, probar que el test de cociente de máxima verosimilitud (CMV) de nivel α para $H_0 : \delta = 0$ contra $H_1 : \delta \neq 0$ rechaza H_0 si y sólo si

$$\frac{|\bar{X} - \bar{Y}|}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} > t_{n+m-2, \alpha/2}.$$

Sugerencia: Puede ahorrar muchas cuentas recordar lo probado en los ejercicios 7 y 12 de la Práctica 2, de EMV bajo normalidad.

- b) Un terreno se divide en 20 lotes semejantes. Se pone a prueba un nuevo fertilizante (I) colocándolo por aspersión en 10 lotes elegidos al azar. En los otros 10 lotes se utiliza como control un fertilizante de uso corriente (II). En cada parcela se planta el mismo número de plantas de tomate, y se observa el rendimiento (en kg/ha) en cada lote, obteniéndose

$$\begin{array}{lll} \bar{X}_{\text{obs}} & = & 130 \quad s_X = 10 \quad \text{para el fertilizante I} \\ \bar{Y}_{\text{obs}} & = & 120 \quad s_Y = 8 \quad \text{para el fertilizante II} \end{array}$$

Se supone que los datos provienen de muestras aleatorias con distribución normal y varianzas iguales. Se asume además que los rendimientos en los 20 lotes son independientes.

- i. Testear a nivel 0,05 si hay diferencia entre los rendimientos entre los dos fertilizantes.
 - ii. Calcular el p -valor. ¿Qué decisión se hubiera tomado a nivel 0,01?
16. Se comparó un método espectroscópico de absorción atómica de llama para determinar antimonio en la atmósfera con el método colorimétrico recomendado. Para ello se tomaron seis muestras de atmósfera urbana (porciones de atmósfera urbana extraída en distintos momentos). Cada muestra se dividió en dos partes, una se midió con el método tradicional y la otra con el método nuevo, obteniendo los siguientes resultados:

Antimonio encontrado (mg/m^3)		
Muestra	Método standard	Método nuevo
1	25.0	23.8
2	19.5	19.0
3	16.6	15.9
4	21.3	20.4
5	20.7	19.6
6	16.8	15.8

- a) ¿Difieren significativamente las medias de ambos métodos? Antes de aplicar un test, observe si es razonable hacer las suposiciones correspondientes.
 - b) En caso de que responda afirmativamente a la pregunta anterior, calcule un intervalo de confianza al 95 % para la diferencia de las medias.
17. a) Sea X_1, \dots, X_n una m.a. de una población con distribución $\mathcal{N}(\mu_1, \sigma_1^2)$ y sea Y_1, \dots, Y_m una m.a. de una población con distribución $\mathcal{N}(\mu_2, \sigma_2^2)$ independiente de la anterior. Probar que el test de CMV de nivel α para $H_0 : \sigma_1^2/\sigma_2^2 = 1$ contra $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$ rechaza H_0 si y sólo si

$$\frac{S_1^2}{S_2^2} > k_1 \quad \text{ó} \quad \frac{S_1^2}{S_2^2} < k_2$$

para ciertas constantes k_1 y k_2 adecuadas. (Por simplicidad, se toma $k_1 = f_{n-1, m-1, \alpha/2}$ y $k_2 = f_{n-1, m-1, 1-\alpha/2}$.)

- b) El riesgo de inversiones alternativas es evaluado generalmente por la varianza de las ganancias asociadas a cada inversión. La distribución de las ganancias anuales de dos inversiones alternativas A y B se supone normal. En base a la información de 10 años pasados en el caso de la inversión A y de 16 años en el caso de la inversión B, se calcularon las varianzas estimadas de las ganancias para las dos inversiones, obteniéndose $s_A^2 = 3,2$ y $s_B^2 = 7,5$. ¿Puede decirse a nivel 0,05 que la inversión B es tan riesgosa como la A? ¿Y a nivel 0,10?

Otra vuelta de rosca a Tests

18. (Para realizar con R) ¿Cómo se distribuye el p-valor?
- a) Generar $n = 25$ datos como si proviniesen de una muestra aleatoria donde cada variable tiene distribución $N(10,3)$.
 - b) Con dichos datos usando el comando `t.test` de R chequear las hipótesis $H_0 : \mu = 10$ vs. $H_0 : \mu \neq 10$ y guardar el p-valor que devuelve el test. ¿Qué espera que pase?
 - c) Repetir b) $Nrep = 1000$ veces.
 - d) Realizar un histograma de los p-valores obtenidos en c). ¿Identifica a qué distribución correspondería este histograma?
 - e) A partir de los valores obtenidos en c), estime la probabilidad de rechazar la hipótesis nula si deseáramos que el nivel de los tests fuera $\alpha = 0,05$. ¿Qué pasa si $\alpha = 0,10$? Observar que aunque todas las muestras del ítem c) se generan bajo H_0 , las probabilidades estimadas no son nulas.
19. Los estudios de asociación del genoma (Genome association studies) son en la actualidad ampliamente utilizados en la investigación médica, particularmente de cáncer. Son estudios que buscan determinar si los genes están asociados con ciertas enfermedades (u otros rasgos). Para llevarlos a cabo, usualmente se obtiene el ADN de dos (o más) grupos de personas, algunas de las cuales tienen la enfermedad en cuestión y otras no. Estos datos de ADN se componen de mediciones realizadas en miles de genes diferentes, y para cada gen se testea si existe diferencia entre los pacientes enfermos y los pacientes sanos. Los resultados pueden ser útiles en la evaluación de riesgos, diagnóstico y la búsqueda de una cura, orientando además hacia nuevas líneas de investigación.
- Uno de los estudios de asociación más famosos ¹ testó las diferencias genéticas entre pacientes con dos diferentes tipos de leucemia (leucemia mieloide aguda y leucemia linfoblástica aguda). En este estudio, los científicos recopilaron datos sobre 7129 genes diferentes para 38 pacientes con leucemia. (Ejemplo 4.26 del Lock, 2da. Edición).
- a) Si todos los tests usaron nivel de significación $\alpha = 0,01$, y si no hubiera diferencias genéticas entre los dos tipos de leucemia, ¿aproximadamente en cuántos de los tests esperaríamos encontrar diferencias estadísticamente significativas entre los dos grupos?
 - b) ¿Tenemos razones para creer que todos los genes para los que se encontraron diferencias estadísticamente significativas están realmente asociados con el tipo de leucemia?
 - c) En el estudio real, el 11 % de los tests realizados arrojaron p-valores inferiores a 0.01. ¿Tenemos razones para creer que existe alguna asociación entre genes y el tipo de leucemia?
20. (Para resolver en clase, con R) En la base de datos de nacimientos de EE.UU. que está publicada en el sitio

¹Golub, T. R., et al., *Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring*, Science, 1999, 286: 531-537

https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm figuran los datos de (casi) todos los nacimientos registrados en 1977. Usamos los datos de ese año porque en los años posteriores, los datos registran la semana en la que se produce el nacimiento, no el día exacto. En las bases anteriores está el día exacto, pero hay menos datos incorporados. Seleccionamos 5000 nacimientos al azar. En la siguiente tabla, figuran los totales por día de la semana.

lunes	martes	miércoles	jueves	viernes	sábado	domingo
747	766	749	741	714	659	624

- Podemos definir la variable aleatoria X_i como la indicadora de que el i -ésimo nacimiento se haya producido un lunes. Entonces, X_1, \dots, X_n son v.a. i.i.d. con distribución $Be(\theta)$, $n = 5000$. ¿Cuánto espera que valga θ ? Realice un test de nivel asintótico $\alpha = 0,05$ para $H_0 : \theta = 1/7$ vs. $H_1 : \theta \neq 1/7$. Escriba el estadístico del test y su valor observado. ¿Qué concluye? Calcule el p-valor.
- Repita (a) para los seis días restantes y escriba la conclusión y los p-valores en cada caso.
- Codifiquemos Lunes $\rightarrow 1$, Martes $\rightarrow 2, \dots$, Domingo $\rightarrow 7$, y θ_j = probabilidad de que un nacimiento se produzca en el día j -ésimo ($j = 1, \dots, 7$). Ahora estamos interesados en el siguiente test:

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_7 = 1/7 \text{ vs. } H_1 : \text{al menos uno de los } \theta_j \text{ es distinto de } 1/7$$

Si tomamos como test ψ al que rechaza si alguno de los siete tests definidos en (a) o (b) rechaza su respectiva hipótesis nula, ¿qué nivel tiene este test ψ ? Si puede, calcúlelo, sino acótelos. ¿Puede mejorar el nivel de ψ cambiando el nivel individual de cada test de (a) y (b) y conseguir nivel asintótico $\alpha = 0,05$ para ψ ?

- Consideremos, para cada nacimiento, el siguiente vector aleatorio $\mathbf{Y} = (Y_1, \dots, Y_7)$ definido por Y_j es la indicadora de que el nacimiento sucede el día j -ésimo de la semana. Observe que $\mathbf{Y} \sim Mult(1, \boldsymbol{\theta})$ con $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_7)$. Si sumamos los vectores de todos los nacimientos obtenemos un vector $\mathbf{W} \sim Mult(n = 5000, \boldsymbol{\theta})$. El valor de W_{obs} aparece en el enunciado de este ejercicio. Realice el test de CMV para las hipótesis del ítem (c), que también pueden escribirse como

$$H_0 : \boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_7) = (1/7, \dots, 1/7) \text{ vs. } H_1 : \boldsymbol{\theta} \neq (1/7, \dots, 1/7)$$

Realice el test asintótico de CMV de nivel 0.05 para estos datos y calcule el p-valor. Puede ser instructivo explorar las funciones `prop.test` del R y `multinom.test` de la librería `ExactMultinom`. Y quizá mirar la función `prop.test`.

- ¿Se le ocurre cómo simular el test del nivel 0.05 para el estadístico CMV (de modo que el nivel sea aproximado por la simulación)? Es decir, ¿sabría cómo simular la distribución del estadístico del test bajo H_0 ? Simule una cantidad grande de veces ($Nrep = 10000$) el estadístico bajo H_0 ². Estime el cuantil 0.95 de dicha distribución a partir de la

²Para ello,

muestra generada.³ Es interesante que compare este valor simulado con el cuantil de la distribución asintótica. Dibuje una estimación de la densidad a partir de `estadsimu` y superponga el gráfico de la distribución asintótica.

21. (Otro test para diferencia de medias) Tenemos dos muestras independientes, y un mecanismo para construir IC de nivel $1 - \alpha$ para las medias de cada una de las poblaciones de las que provienen. Nos interesa desarrollar un test para la diferencia de las medias.

- Sean X_1, \dots, X_n v.a.i.i.d con $E(X_1) = \mu_1$. Notamos $\mathbf{X} = (X_1, \dots, X_n)$. El intervalo de confianza de nivel $(1 - \alpha)$ para μ_1 basado en la muestra está dado por

$$IC_1 = [a_1(\mathbf{X}), a_2(\mathbf{X})]$$

- Sean Y_1, \dots, Y_m v.a.i.i.d con $E(Y_1) = \mu_2$. Notamos $\mathbf{Y} = (Y_1, \dots, Y_m)$. El intervalo de confianza de nivel $(1 - \alpha)$ para μ_2 basado en la muestra está dado por

$$IC_2 = [b_1(\mathbf{Y}), b_2(\mathbf{Y})]$$

- a) Proponemos el siguiente intervalo de confianza para $\mu_2 - \mu_1$:

$$IC_3 = [b_1(\mathbf{Y}) - a_2(\mathbf{X}), b_2(\mathbf{Y}) - a_1(\mathbf{X})]$$

¿Qué nivel tiene IC_3 ? Acótelos usando la sugerencia. ¿Qué puede asegurar acerca del nivel de IC_3 ?

Sugerencia: Probar y luego usar que $\{\mu_1 \in IC_1\} \cap \{\mu_2 \in IC_2\} \subseteq \{\mu_2 - \mu_1 \in IC_3\}$

- b) Proponemos el siguiente test ϕ para $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$: Rechazar H_0 si $0 \notin IC_3$. Usando la dualidad entre tests e intervalos de confianza, el test ϕ ¿tiene nivel mayor o menor a α ?
- c) Pruebe la siguiente equivalencia:

$$0 \notin IC_3 \iff IC_1 \cap IC_2 = \emptyset$$

Observe que esto da una manera más intuitiva de definir al test ϕ :

$$\text{Rechazamos } H_0 \text{ si } IC_1 \cap IC_2 = \emptyset$$

-
- Genere una $\mathbf{Y} \sim \text{Mult}(n = 5000, \theta)$ asumiendo que H_0 es verdadera. Notar que es lo mismo que generar $n = 5000$ vectores aleatorios $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim \text{Mult}(\mathbf{1}, \theta)$. El comando `rmultinom` puede ser de utilidad.
 - Calcule el valor que toma el estadístico del test en la muestra generada. Guarde este valor.
 - Repita lo anterior $Nrep = 10000$ veces. Obtendrá un vector de longitud $Nrep = 10000$ con valores simulados del estadístico del test generados asumiendo H_0 verdadera. Llamaremos `estadsimu` a este vector.

³Observemos que como nos interesa simular el cuantil 0.95 de una distribución la simulación no podría realizarse con pocas replicaciones, por ejemplo con $Nrep = 90$ no bastaría. Si nos interesara la mediana, que es un percentil que no está cerca de las colas de la distribución, podríamos hacer la estimación con un número de replicaciones menor. Tomamos $Nrep = 10000$ para cubrimos. Si prefiere, aumente este valor y compare lo obtenido.

- d) A fin de ilustrar el comportamiento de la función de potencia de estos tests haremos un experimento numérico.

Para ello generar dos muestras independientes que sigan una distribución normal con $n = m = 15$. Más precisamente, generar X_1, \dots, X_n v.a.i.i.d con $X_i \sim N(\mu_1, 25)$ con $\mu_1 = 1$ y por otro lado, en forma independiente, generar Y_1, \dots, Y_m v.a.i.i.d con $Y_i \sim N(\mu_2, 25)$, para $\mu_2 = 1, 1.1, 1.5, 2, 3, 4, 5, 6$.

Aplicar el test ϕ basado en IC_3 con $\alpha = 0,05$ y el test de comparación de medias de dos poblaciones normales basado en muestras independientes con igualdad de varianzas que denotaremos ψ de nivel $\alpha = 0,05$. Observar qué decisión se toma con cada uno de ellos para cada valor de μ_2 .

Repetir este procedimiento $Nrep = 10000$, computar la frecuencia relativa de rechazo de cada test y completar para los distintos valores de μ_2 la siguiente tabla.

¿Qué observa? ¿Cuál es el comportamiento de la función de potencia empírica de estos dos tests? ¿Cuál prefiere? Observar que en la fila correspondiente a $\mu_2 = 1$ se obtiene el nivel empírico de cada uno de estos tests.

μ_2	ϕ	ψ
1		
1.1		
1.5		
2		
3		
4		
5		
6		

Cuadro 1: Potencia empírica

- e) La Figura 1 fue extraída del artículo ⁴ En dicho artículo se comparan los IQ (resultado del test de coeficiente intelectual) para 311 niños de seis años cuyas madres tomaron, por prescripción médica pues eran epilépticas, alguna de cuatro drogas antiepilépticas distintas (carbamazepine, lamotrigine, phenytoin, o valproate) durante el embarazo. Además, algunas madres tomaron ácido fólico (*periconceptional folate*) que es un suplemento vitamínico que se indica durante el embarazo para prevenir defectos de nacimiento graves del cerebro y de la columna vertebral del bebé. La Figura 1 muestra intervalos de confianza de nivel 0.95 para el IQ medio de los niños de 6 años de cada uno de los grupos (en línea sólida si la madre tomó ácido fólico o línea punteada cuando no lo tomó). En el pie de la figura se describen los tamaños de muestra en los que se basa cada intervalo. A partir del gráfico, indique para cuáles de las cuatro drogas puede asegurar que hay diferencias estadísticamente significativas entre tomar o no el

⁴Kimford J Meador, Gus A Baker, Nancy Browning, Morris J Cohen, Rebecca L Bromley, Jill Clayton-Smith, Laura A Kalayjian, Andres Kanner, Joyce D Liporace, Page B Pennell, Michael Privitera, David W Loring, for the NEAD Study Group. Fetal antiepileptic drug exposure and cognitive outcomes at age 6 years (NEAD study): a prospective observational study. *Lancet Neurol* 2013; 12: 244–52.

ácido fólico a nivel menor o igual a 0.05. En base a la información disponible sólo puede realizarse el test ϕ presentado en el ítem (c). Si contara con todos los datos de este estudio, ¿preferiría hacer esta comparación mediante este test o elegiría algún otro de los tests para comparar las medias de dos muestras presentados en el curso?

Figura 1: Figura 3 del paper de *Lancet*.

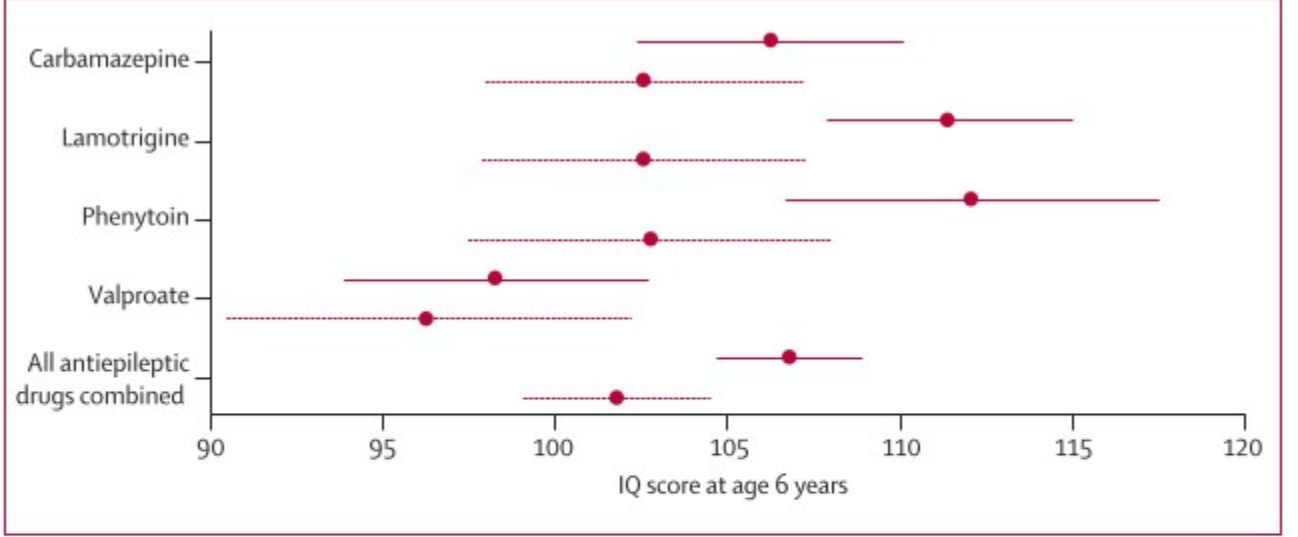


Figure 3: Child IQ at 6 years, by exposure to maternal antiepileptic drug use and periconceptional folate
Mean (95% CIs) are shown for folate (solid lines) and no folate (dashed lines). For carbamazepine, 56 children were exposed to periconceptional folate (mean IQ 106, 95% CI 102–110) and 38 children were not (103, 98–107); for lamotrigine 60 children were exposed to periconceptional folate (111, 108–115) and 40 children were not (103, 98–107); for phenytoin, 23 children were exposed to periconceptional folate (112, 107–118) and 32 children were not (103, 98–108); for valproate, 40 children were exposed to periconceptional folate (98, 94–103) and 22 children were not (96, 91–102); for all antiepileptic drugs combined, 179 children were exposed to periconceptional folate (107, 105–109) and 132 children were not (102, 99–105). Appendix p 7 shows means and analyses for dose effects. IQ=intelligence quotient.

22. **Test No Paramétrico.** Sean X_1, \dots, X_n v.a.i.i.d con distribución F . Sea $X \sim F$ y llamemos $Med_F(X)$ a la mediana de la distribución de las X 's. Fijamos un valor real a_0 . Proponemos un test, para las siguientes hipótesis.

$$(A) \begin{aligned} H_0 &: Med_F(X) = a_0 \\ H_1 &: Med_F(X) \neq a_0 \end{aligned}$$

Para $i = 1, \dots, n$, sean

$$T_i = \begin{cases} 1 & \text{si } X_i \geq a_0 \\ 0 & \text{sino.} \end{cases}$$

Luego, bajo H_0 , T_i tiene distribución $Be(\theta)$, ¿con qué valor de θ ? El test que proponemos se basa en $\bar{T}_n = \frac{1}{n} \sum_{i=1}^n T_i$. Una forma de testear las hipótesis (A) para las X 's consiste en testear las hipótesis

$$\begin{aligned} \text{(B)} \quad & H_0 : \theta = 1/2 \\ & H_1 : \theta \neq 1/2 \end{aligned}$$

para las T 's. El test δ propone rechazar H_0 si

$$Z = \frac{\sqrt{n}(\bar{T}_n - \frac{1}{2})}{\sqrt{1/4}} \geq z_{\alpha/2} \quad \text{ó} \quad Z = \frac{\sqrt{n}(\bar{T}_n - \frac{1}{2})}{\sqrt{1/4}} \leq -z_{\alpha/2}$$

con $z_{\alpha/2}$ el cuantil $1 - \frac{\alpha}{2}$ de la $N(0, 1)$. δ se conoce como el Test del signo.⁵

- a) Verificar que δ resulta ser un test de nivel asintótico α para las hipótesis (A).
- b) Si se cuenta con X_1, \dots, X_n v.a.i.i.d con distribución $N(\mu, \sigma_0^2)$ con σ_0 conocida, observemos que en este caso la media y la mediana poblacionales de las X 's coinciden. Queremos testear las hipótesis

$$\begin{aligned} \text{(C)} \quad & H_0 : \mu = a_0 \\ & H_1 : \mu \neq a_0 \end{aligned}$$

Como $\mu = \text{Med}(X)$, las hipótesis (A) y (C) coinciden en este caso, y tenemos dos tests disponibles, el test del signo (δ) y el test para la media para una población normal con varianzas conocida. Queremos comprar gráficamente la potencia de ambos.

Para ello, tenga en cuenta que si $\mu = \mu_1$, entonces

$$\mathbb{P}_{\mu=\mu_1}(T_i = 1) = \mathbb{P}_{\mu=\mu_1}(X_i > a_0) = \mathbb{P}\left(\frac{X_i - \mu_1}{\sigma_0} > \frac{a_0 - \mu_1}{\sigma_0}\right) = 1 - \Phi\left(\frac{a_0 - \mu_1}{\sigma_0}\right) = \theta_1$$

Utilice este valor θ_1 para estandarizar el estadístico del signo en el cálculo de la potencia.

- c) (Opcional) Grafique para μ_1 entre 10 y 30 en el mismo eje de coordenadas las funciones de potencia de ambos tests, para $n = 100$, $a_0 = 20$, $\sigma_0 = 4$, $\alpha = 0,05$. ¿Cuál de los tests resulta ser más potente?

23. Volvemos a los datos del estudio acerca del cáncer de páncreas correspondientes a los Ejercicios 1 y 9 de la Práctica 1 y al Ejercicio 22 de la Práctica 2.

- a) La comparación del biomarcador **LYVE1** en los tres grupos formados por los valores de la variable **diagnosis** es el objetivo de esta parte del estudio. Sin embargo, antes de realizar el análisis es recomendable explorar la distribución del biomarcador en cada grupo. A tal fin, realizar el **qqnorm** de **LYVE1** para cada nivel de **diagnosis**, ¿parece

⁵Este test es un ejemplo de lo que se conoce como “*tests no paramétricos*”, en el sentido que no asume que la distribución de las variables X 's pertenezca a una familia de distribuciones conocidas parametrizadas con uno o más parámetros como la normal, exponencial, Poisson, etc. Como el test del signo usa poca información de la distribución de las variables involucradas, esperamos que resulte menos potente que un test que tenga en cuenta toda esta estructura.

simétrica la distribución? ¿Qué concluye si realiza los boxplots? Comparar todos estos gráficos con los obtenidos para la variable transformada con el logaritmo, es decir con $\log(\mathbf{LYVE1})$.

- b)* En primera instancia, interesa comparar las medias del logaritmo de **LYVE1** del grupo con **diagnosis=1** y del grupo con **diagnosis=2** a nivel aproximado 0.01. ¿Qué decisión toma?
- c)* Con un interés más amplio, se desea comparar las medias del logaritmo de **LYVE1** de los tres grupos determinados por los tres niveles de la variable **diagnosis**. Una estrategia naive, podría ser realizar las tres comparaciones involucradas, es decir contrastar las media del grupo con **diagnosis=1** vs. la del grupo con **diagnosis=2**, por otro lado, la del **diagnosis=1** vs. la del grupo con **diagnosis=3** y por último, la medias del **diagnosis=2** vs. con la del grupo con **diagnosis=3**. Si se quisiera tomar una decisión con un nivel global aproximado 0.01, ¿con qué nivel debería tomarse cada decisión individual?
- d)* Relacionar las conclusiones a las que arribó con las realizadas en forma exploratoria en la Práctica 1.