Estadística (M) - Segundo cuatrimestre 2019

Práctica 5 - Tests de hipótesis

Se recomienda empezar por los ejercicios marcados sin †.

A) *Conceptos básicos y tests bajo normalidad*

1. Una colonia de ratones de laboratorio tiene varios cientos de animales. El peso en gramos de los ratones adultos sigue una distribución normal con media igual a 30g. y desvío estándar de 5g. Como parte de un experimento, se les pidió a algunos estudiantes que eligieran 25 ratones adultos, sin ninguna premisa. Se asume que el peso de los ratones elegidos por estos estudiantes sigue una distribución normal con desvío estándar (poblacional) igual a 5g. y media no menor a 30g. Se quiere hacer un test de hipótesis para ver si hay evidencia de que los estudiantes tienden a elegir los ratones más pesados, es decir, si el peso medio de los ratones elegidos por los estudiantes es mayor a 30g.

   (a) Plantear un test de hipótesis de nivel 0.05 para esta situación. Recordar definir las variables involucradas, las hipótesis a testear, el estadístico del test, su distribución bajo la hipótesis nula y la región de rechazo.

   (b) Si el peso promedio de los 25 ratones elegidos fue de 33g., ¿hay evidencia a nivel 0.05 de que los estudiantes tienden a elegir ratones más pesados? Hallar el p-valor del test.

2. Consideremos un procedimiento para medir el contenido 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 error sistemático (es decir, si su media es distinta de lo que debería ser).

   (a) Se hacen 8 determinaciones 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 $(\overline{x} = 7.074)$

   ¿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 0.05?

   (b) Si el método tiene un error sistemático de 0.10 (o sea, si la media es 7.10), ¿cuál es la probabilidad de cometer 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 ≥ 0.80 (o lo que es equivalente, la probabilidad de error tipo II sea ≤ 0.20).

      i. El test del inciso a) ¿cumple con este requisito?

      ii. En caso contrario, ¿cuántas determinaciones habría que hacer como mínimo?

3. Una asociación de consumidores, preocupada por el porcentaje de grasa contenida en una marca de hamburguesas, envía a un laboratorio independiente una muestra aleatoria de 12 hamburguesas para su análisis. El porcentaje de grasa en cada una de las hamburguesas de la muestra es:

   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 menor al 18%. Basándose en la salida de R que figura más abajo, resuelva los siguientes items.

(a) La asociación de consumidores quiere saber si tiene motivos para decir que la afirmación del fabricante es falsa. Asumiendo que el contenido de grasa de cada hamburguesa de esta marca es una v.a con distribución normal con varianza deconocida y media no menor a 18%, 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 en este caso cuando el contenido medio de grasa es 20%. Respecto del ejercicio anterior, ¿qué dificultad aparece cuando uno intenta hacer esta cuenta? Opcional: pensar alguna forma computacional de hacer esta cuenta.

```
hamburguesas<-scan()
21 18 19 16 18 24 22 19 24 14 18 15
c(mean(hamburguesas),var(hamburguesas),sd(hamburguesas))
[1] 19.000000 10.545455  3.247377
```

Sugerencia: usar que si $X_1, \ldots, X_n$ son i.i.d. $N(\mu, \sigma^2)$, entonces

$$\sqrt{n}\frac{\overline{X} - \mu}{s} \sim T_{n-1}$$

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

4. 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.

   Sugerencia: usar que si $X_1, \ldots, X_n$ son i.i.d. $N(\mu, \sigma^2)$, entonces

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

   donde χ_n^2 es una Chi Cuadrado con n grados de libertad.

† 5. (Para hacer con R) Sea $X_1, \ldots, X_n$ una muestra aleatoria de distribución $N(\mu, \sigma^2)$ con μ y σ^2 desconocidos. Queremos testear $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$ a nivel $\alpha = 0.05$ y para esto se desea aplicar el test t, usado en el ejercicio **??**). Supongamos fijados ciertos parámetros n, γ y c, donde $n \in \mathbb{N}$, $0 \leq \gamma$ y $c \in \mathbb{R}$. El objetivo de este ejercicio es comprobar empíricamente cuán robusto es el nivel de significación de este test bajo la presencia de datos atípicos. Consideramos el siguiente procedimiento para la simulación:

   (a) Generar una muestra de tamaño n de una distribución $N(0, 1)$.

   (b) Agregarle a dicha muestra una proporción γ de valores atípicos, cada uno de ellos con el valor c. Es decir, agregarle a la muestra original $n\gamma$ veces el número c.

   (c) Aplicarle el test t a la muestra del item anterior (se puede usar el comando `t.test` de `R`). Rechazar H_0 si el $p-$valor es menor al nivel nominal 0.05.

(d) Repetir los items anteriores $N = 10000$ veces, contando el total de veces en que se rechazó H_0. Se puede estimar el nivel de este test como la proporción empírica de rechazos, es decir, dividiendo la cantidad total de rechazos por N (notar que si no hubiera datos atípicos, esto debería dar cerca de 0.05).

Completar la tabla con los niveles de significación estimados para $n = 100$ y los valores indicados para γ y c. El test para estos modelos, ¿cuándo resulta liberal y cuando resulta conservador?

	c					
γ	1	3	6	10	15	20
0						
0.01						
0.05						
0.1						

†6. (Para hacer con el 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 $X_1, \ldots, X_n$, considera el estadístico $T = \sqrt{n}\,\overline{X}$, (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 obervado 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 $\mu : 0.5, 1, 2, 5, 10$.

B) *Teorema de Neyman-Pearson para hipótesis simples*

1. Una empresa vende dos variedades de soja. La variedad 1 tiene un rendimiento por ha. que puede considerarse una variable aleatoria con distribución $N(37, 25)$, y la variedad 2 tiene un rendimiento por ha. que puede considerarse $N(40, 25)$. 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 ha. y obtiene los siguientes rendimientos:

$$37 \quad 39.5 \quad 41.7 \quad 42 \quad 40 \quad 41.25 \quad 43 \quad 44.05 \quad 38 \quad 38.5$$

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) Plantear el test MP para este problema. ¿Qué decisión se toma?

(b) Hallar el valor p. ¿Se hubiera rechazado H_0 para $\alpha = 0.01$?

(c) Calcular la probabilidad del 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. Consideremos dos funciones de probabilidad puntual $P_0(x)$ y $P_1(x)$ dadas por la siguiente tabla:

	x_1	x_2	x_3	x_4
P_0	0.05	0.00	0.05	0.90
P_1	0.00	0.05	0.90	0.05

Se observa una v.a. X y se quiere testear $H_0 : X \sim P_0$ vs $H_1 : X \sim P_1$.

(a) Encontar los 4 tests no aleatorizados de nivel $\alpha = 0.05$ para este problema.

(b) Entre estos 4 tests de nivel α, ¿cuál es el mejor?

(c) Probar que el test hallado en el inciso anterior es el más potente de nivel α.

(d) ¿Hay algún test no aleatorizado de nivel menor que α? ¿Qué potencia tiene?

(e) Encontrar el test más potente de nivel $\alpha = 0.10$.

C) *Tests para familias con cociente de verosimilitud monótono*

1. Probar que las siguientes familias son de cociente de verosimilitud monótono:

(a) $\mathcal{U}(\theta, \theta + 1)$

(b) Las familias exponenciales a un parámetro con densidad o probabilidad

$$p(x|\theta) = a(\theta)h(x)\exp(c(\theta)t(x))$$

siempre que $c(\theta)$ sea estrictamente creciente. Luego las familias: binomial, binomial negativa, Poisson, Gamma, Beta y Normal (estas 3 últimas considerando uno de los parámetros fijos) son de CVM.

† (c) La familia doble exponencial

$$f(x|\alpha,\beta) = \frac{1}{2\beta}\exp\left\{-\frac{|x-\alpha|}{\beta}\right\}$$

con β desconocido.

† (d) La familia exponencial negativa

$$f(x|\alpha) = \exp\{-(x-\alpha)\}\,\mathrm{I}_{(\alpha,\infty)}(x)$$

2. (a) Dada $X_1, \ldots, X_n$ una m.a. de una distribución $\mathcal{E}(\lambda)$, hallar el test UMP de nivel α para $H_0 : \lambda \geq \lambda_0$ vs $H_1 : \lambda < \lambda_0$.

(b) Se sabe que el tiempo de duración de cierto tipo de lamparitas sigue una 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.

i. 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?

ii. Calcular el $p-$valor.

iii. ¿Cuál es la probabilidad de no sacar la producción a la venta si el tiempo medio de vida es de 57 días?

3. (a) Sea $X_1, \ldots, X_n$ una m.a. de una distribución $Bi\,(1, p)$. Encontrar el test UMP de nivel α para $H_0 : p \leq p_0$ vs $H_1 : p > p_0$.

(b) 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.

i. 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?

ii. Calcular el $p-$valor.

iii. ¿Cuál es la probabilidad de no cambiar de tratamiento si el nuevo en realidad tiene una probabilidad de éxito del 45%?

(c) 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 $P\,(X > Y) > P\,(X \leq Y)$.)

D) *Tests de cociente de máxima verosimilitud*

1. (a) Sea $X_1, \ldots, X_n$ una m.a. de una distribución $N(\mu, \sigma^2)$ con varianza σ^2 desconocida. Hallar el test de cociente de máxima verosimilitud de nivel α para $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.

(b) Se diseñó un sistema de riego de manera que el tiempo promedio de activación sea exactamente 25 segundos. Se lo probó 11 veces, obteniéndose los siguientes tiempos de activación:

27 41 22 27 23 35 30 24 27 28 22

Se supone que el tiempo de activación es una v.a. con distribución normal.

i. A nivel 0.05, ¿contradicen estos datos las especificaciones del sistema?

ii. Calcular el $p-$valor.

2. (a) Sea $X_1, \ldots, X_n$ una m.a. de una distribución $N\,(\mu_1, \sigma^2)$ e $Y_1, \ldots, Y_m$ una m.a. de una distribución $N\,(\mu_2, \sigma^2)$ independiente de la anterior. Si llamamos $\Delta = \mu_1 - \mu_2$, probar que el test de CMV de nivel α para $H_0 : \Delta = 0$ vs $H_1 : \Delta \neq 0$ rechaza H_0 si y sólo si

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

(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

$$\begin{aligned} \bar{X} &= 130 & s_X &= 10 & &\text{para el fertilizante I} \\ \bar{Y} &= 120 & s_Y &= 8 & &\text{para el fertilizante II} \end{aligned}$$

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?

3. Se tomaron aleatoriamente 8 pares de pollos hermanos y se eligió al azar dentro de cada par cuál iba a la muestra A y cuál a la B. Los pollos de la muestra A han sido criados en encierro, mientras que los de la muestra B al aire libre. Los datos siguientes corresponden al peso de los pollos de 7 semanas:

Muestra A	255	481	360	368	425	283	311	368
Muestra B	226	425	311	311	255	340	311	284

(a) ¿existe alguna diferencia en el peso medio a las 7 semanas entre los métodos de crianza? Aplicar un test de cociente de máxima verosimilitud. ¿Para qué variable tuvo que asumir normalidad? Testear a nivel 0.01.

(b) Calcular el $p-$valor.

Sugerencia: ¿se trata de muestras independientes o apareadas?

E) *Tests con nivel de significación asintótica*

1. 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.

(a) Utilizando un test de nivel asintótico, decidir si hay suficiente evidencia a nivel 0.05 para rechazar la suposición. Calcular el valor p aproximado.

(b) Si el porcentaje de hembras fuera 0.55, ¿cuál sería la probabilidad de tomar una decisión incorrecta?

2. El número de personas que llega a la boletería de una estación de trenes entre las 14 y 14:30 hs. tiene 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 es superior a 20 individuos. 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 observa el número de personas que arriba en dicho horario y se calcula una media muestral de 21 individuos.

(a) Proponer un test asintótico para este problema usando el TCL. ¿Qué decisión toma en base a estos datos? ¿Cuál es el valor p aproximado?

(b) Para el test del item anterior, ¿cuál es la probabilidad de dejar una sola boletería cuando en realidad la media verdadera es de 22 personas?

(c) Si se hubiera querido que la probabilidad calculada en (b) fuera a lo sumo 0.05, ¿qué tamaño muestral debería haberse tomado?

3. (a) Sean $X_1, \ldots, X_n \sim \varepsilon(\lambda)$. 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.

   (b) Tenemos $k \geq 2$ muestras $X_1^j, \ldots, X_n^j$ para $j = 1, \ldots, k$, todas independientes. Asumimos que $X_i^j \sim \varepsilon(\lambda_j)$. Se quiere testear $H_0 : \lambda_1 = \cdots = \lambda_k$ vs $H_1 : \exists\, i \neq j / \lambda_i \neq \lambda_j$. Usando la distribución asintótica del cociente de máxima verosimilitud, plantear un test asintótico de nivel α para estas hipótesis.

4. 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:

$$\begin{array}{ccccccccccccccc} 3, & 4, & 5, & 3, & 2, & 7, & 6, & 4, & 6, & 4, & 7, & 2, & 3, & 4, & 6, \\ 7, & 3, & 5, & 6, & 6, & 4, & 5, & 5, & 7, & 3, & 2, & 2, & 4, & 3, & 5. \end{array}$$

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?

† 5. 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.

† 6. (Para hacer con el R) Se tiene una muestra de tamaño n de una distribución $Bi(1, p)$. Se quiere testear $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$ a nivel asintótico 0.05. Se considera el test con estadístico

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

y región de rechazo $\{|W| > z_{\alpha/2}\}$. Vamos a considerar los pares (n, p_0) con n variando en el conjunto $\{10, 70, 200\}$ y p_0 variando en el conjunto $\{0.5, 0.8, 0.97\}$ (9 pares en total). Para el test correspondiente a cada par (n, p_0), calcular mediante una simulación la función de potencia $\beta(p)$ en los valores $p = 0.03, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.97$ del test correspondiente y graficarla interpolando en los valores de p indicados. ¿Qué se puede decir sobre el nivel de significación en cada caso? ¿En que casos es buena la aproximación normal?