PROBABILIDADES Y ESTADÍSTICA (C)

PRÁCTICA 7

1. Encontrar un intervalo de confianza de nivel exacto $1-\alpha$ para la media de una población normal

   a) con varianza conocida.
   b) con varianza desconocida.

2. Se realiza a 10 pacientes un análisis de sangre con el fin de determinar el porcentaje de hemoglobina, obteniéndose $\bar{X} = 12\%$ y $s = 0.5\%$.

   a) Hallar un intervalo de confianza para la media verdadera, de nivel exacto 0.90, suponiendo que la concentración de hemoglobina se distribuye normalmente.
   b) Idem (a), suponiendo además que $\sigma = 0.6\%$.
   c) Si se quisiera que la longitud del intervalo hallado en (b) fuera a lo sumo 0.5, ¿a cuántos pacientes debería analizarse?

3. Encontrar un intervalo de confianza de nivel exacto $1 - \alpha$ para la varianza de una población normal

   a) con media conocida.
   b) con media desconocida.

4. En un aserradero se cortan varillas de madera cuya longitud es una v.a. con distribución normal. Se miden 25 varillas elegidas al azar, obteniéndose $\bar{X} = 180$ cm y $s = 10$ cm.

   a) Hallar un intervalo de confianza de nivel exacto 0.90 para la varianza verdadera, suponiendo que $\mu = 185$.
   b) Idem (a), suponiendo μ desconocida.

5. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $\mathcal{E}(\lambda)$.

   a) Probar que $2\lambda\sum_{i=1}^{n} X_i$ tiene distribución χ^2_{2n}.
   b) Hallar un intervalo de confianza para λ de nivel exacto $1 - \alpha$.
   c) ¿Cuál sería el intervalo de confianza de nivel exacto $1 - \alpha$ para $E(X_1)$? ¿Cuál sería su longitud esperada?
   d) Aplicar (b) a los datos del Ejercicio 5.(a) de la Práctica 7, con un nivel del 95%.
   e) Hallar un intervalo de nivel asintótico $1 - \alpha$ para λ.

6. Una muestra aleatoria de 1000 votantes es encuestada respecto a cierta propuesta política. Como resultado, 200 están de acuerdo con la propuesta, 600 se oponen y 200 están indecisos.

   a) Hallar un intervalo de confianza de nivel asintótico 0.90 para la proporción de votantes que se oponen a la propuesta.

   b) ¿Cuántos votantes deberían encuestarse para que la longitud del intervalo obtenido fuese menor o igual que 0.02?

7. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $Bi\,(k, \theta)$.

   a) Hallar un intervalo de confianza de nivel asintótico $1-\alpha$ para θ, siendo k conocido.

   b) Encontrar una cota superior para la longitud del intervalo hallado en el inciso anterior.

8. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $\mathcal{P}\,(\lambda)$.

   a) Hallar un intervalo de confianza de nivel asintótico $1-\alpha$ para λ.

   b) Aplicar (a) a los datos del Ejercicio 5.(b) de la Práctica 7, con $\alpha = 0.05$.

9. a) Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $\mathcal{U}[0, \theta]$. Hallar el intervalo de confianza de nivel exacto $1-\alpha$ para θ, de menor longitud. (Sugerencia: Encontrar la distribución de $\max\limits_{1\leq i\leq n} (X_i)\,/\theta$.)

   b) Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con densidad

   $$f(x;\theta) = e^{-(x-\theta)} I_{[\theta,\infty)}(x)$$

   Hallar el intervalo de confianza de nivel exacto $1-\alpha$ para θ, de menor longitud. (Sugerencia: Encontrar la distribución de $\min\limits_{1\leq i\leq n} (X_i) - \theta$.)

10. Sean $X_1, \ldots, X_n$ v.a. continuas i.i.d. y con función de densidad dada por $f(x)$. Sea $\tilde{\mu}$ la mediana de la distribución de las X's, es decir $P(X_i \leq \tilde{\mu}) = \dfrac{1}{2}$ para todo i.

    a) Probar que

    $$P\left(\max_{1\leq i\leq n} X_i < \tilde{\mu} \quad \cup \quad \min_{1\leq i\leq n} X_i > \tilde{\mu}\right) = \left(\frac{1}{2}\right)^{n-1}.$$

    b) Deducir que $\left(\min\limits_{1\leq i\leq n} X_i,\ \max\limits_{1\leq i\leq n} X_i\right)$ es un intervalo de confianza para $\tilde{\mu}$ de nivel $1-\left(\dfrac{1}{2}\right)^{n-1}$.

11. 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 - 39.50 - 41.70 - 42 - 40 - 41.25 - 43 - 44.05 - 38 - 38.50

    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) Explicar por qué las hipótesis para este problema son

    $$H_0 : \mu = 37 \qquad H_1 : \mu = 40.$$

    b) Encontrar un test para estas hipótesis y decir qué decisión se toma en base a la muestra dada. Calcular la probabilidad del error de tipo II.

    c) Encontrar el test para el caso en que se cultiven n parcelas (con igual nivel).

    d) Determinar el número n de parcelas a cultivar para que el error de tipo II tenga probabilidad menor o igual que 0.05.

    e) Explicar por qué el test planteado en (c) sirve también para las hipótesis

    $$H_0 : \mu = 37 \qquad H_1 : \mu > 37.$$

    f) Calcular la función de potencia $\pi(\mu)$ del test (c), verificar que es creciente para $\mu \in \mathcal{R}$ y deducir que este test sirve también para testear

    $$H_0 : \mu \leq 37 \qquad H_1 : \mu > 37.$$

12. En la construcción de un edificio debe usarse un concreto que tenga una tensión media mayor a 300 psi. Para verificar si el concreto preparado a partir del cemento "Loma Blanca" cumple con este requerimiento, se realizan 15 mediciones en forma independiente de la tensión de este concreto. Se observa una media muestral de 304 psi. Se sabe que el desvío estándar de la tensión de este concreto es 10 psi. El constructor desea correr un riesgo del 5% de comprar cemento "Loma Blanca" cuando éste produce un concreto que no cumple con las especificaciones. Suponiendo que los datos están normalmente distribuídos:

    a) Plantear el test correspondiente. ¿Qué decisión se toma?

    b) Calcular el valor p. Con los datos observados, ¿para qué valores del nivel α sería rechazada la hipótesis nula?

    c) ¿Qué tamaño de muestra debería haberse tomado para que $\pi(302)$ fuera al menos 0.95?

13. Se diseñó un nuevo sistema de riego de manera que el tiempo promedio de activación sea a lo sumo 25 segundos. Se lo prueba 11 veces, obteniéndose los siguientes tiempos de activación:

    27 - 41 - 22 - 27 - 23 - 35 - 30 - 24 - 27 - 28 - 22

    Suponiendo que el tiempo de activación es una v.a. con distribución normal:

    (a) A nivel 0.05, ¿contradicen estos datos las especificaciones del sistema?

    (b) Acotar el valor p.

14. Se supone que 1 de cada 10 fumadores prefiere la marca A. Después de una campaña publicitaria en cierta región de ventas, se entrevistó a 200 fumadores para determinar la efectividad de la campaña. El resultado de esta encuesta mostró que 26 personas preferían la marca A.

    a) ¿Indican estos datos, a nivel aproximado 0.05, un aumento en la preferencia por la marca A?

    b) Calcular el valor p.

    c) ¿Cuál es la probabilidad aproximada de decidir que la campaña publicitaria no fue efectiva, cuando en realidad la proporción de preferencia por la marca A después de la campaña es 0.15?

    d) ¿Qué tamaño de muestra debería tomarse para que la probabilidad de (c) fuese a lo sumo 0.05?

15. Una compañía que vende CD's por correo está tratando de decidir entre dos nuevas grabaciones del oratorio "Elías" de Mendelsohn para agregar a su catálogo. Si ambas versiones son igualmente atractivas, ambas serán ofrecidas, mientras que si una de ellas es claramente preferida será la única ofrecida.

    Las hipótesis a considerar son

    $$H_0 : p = 0.5 \qquad H_1 : p \neq 0.5$$

    siendo p la proporción de suscriptores que prefieren la grabación A.

    Suponga que 10 suscriptores son elegidos al azar y a cada uno de ellos se le pide que escuche ambas grabaciones e indique su preferencia. Sea X el número de suscriptores que prefiere la grabación A.

    a) Calcular la probabilidad del error de tipo I para la región de rechazo $\{0, 1, 2, 8, 9, 10\}$. ¿El nivel de este test es exacto o asintótico?

    b) Si se observa $X = 9$, basándose en el test hallado en (a), ¿deberían ofrecerse ambas grabaciones? ¿Qué tipo de error pudo haberse cometido?

    c) Utilizando el test propuesto en (a), calcular la probabilidad del error de tipo II para $p = 0.4$, 0.6 y 0.8.

16. Tenemos un dado de seis caras. Queremos saber si el dado está cargado o no. Para ello se arroja el dado 100 veces obteniéndose los siguiente resultados

```
5 3 4 5 5 1 2 2 6 1 2 6 3 1 2 6 2 1 2 6 3 3 3 2 6 3 4 6 4 2 3 4 5
2 2 2 6 3 5 5 6 5 3 5 6 5 5 5 4 5 2 2 5 5 3 1 5 2 4 3 4 2 1 2 3 2
3 1 6 5 2 3 2 5 3 3 5 3 1 3 1 6 3 5 2 6 3 5 5 6 2 6 6 5 5 6 2 3 2
2 5 5 6 4 2 5 4 6 6 3 2 5 4 2 3 3 2 4 1 5 2 5 1 6 4 5 2 3 2 3 5 2
2 6 5 5 2 2 3 6 5 1 2 4 6 2 1 4 5 5 2 2 5 2 3 6 3 2 6 3 2 5 1 2 5
2 2 5 1 2 3 2 2 6 1 5 6 5 5 1 2 6 2 2 2 2 5 6 2 5 1 2 5 4 3 5 5 6
6 1 2 6 3 2 2 4 4 5 2 3 2 3 2 3 6 2 2 4 4 2 3 3 5 4 2 2 6 5 5 3 5
6 3 2 2 5 1 2 5 4 5 5 4 2 1 6 5 3 3 4 2 5 2 5 2 3 5 3 2 5 2 6 1 3
5 5 2 5 1 6 1 4 6 2 3 3 4 2 2 2 2 5 6 5 6 6 5 2 6 5 2 5 3 3 2 1 5
2 2 5 5 3 3 5 5 3 6 2 2 2 2 4 5 3 5 6 2 4 2 2 3 5 6 1 2 2 2 1 4 6
4 3 5 6 5 5 5 5 3 2 2 5 6 6 4 5 6 2 2 6 5 5 5 5 5 2 2 2 4 4 4 3 3
2 1 5 5 3 6 5 6 1 2 2 2 4 3 5 3 2 3 1 2 3 2 2 5 2 5 6 4 5 2 2 3 4
5 5 5 2 6 2 2 2 2 4 3 2 5 5 2 5 3 3 2 5 2 6 1 5 3 2 2 4 5 5 3 6 3
2 4 5 6 6 3 4 4 5 3 2 4 3 6 3 2 3 6 3 1 6 2 6 5 5 2 4 5 5 5 6 1 4
6 2 1 5 6 4 2 2 5 5 2 5 4 5 3 6 3 6 3 2 6 3 5 2 2 4 1 2 2 5 3 3 6
5 6 4 5 6 5 3 2 6 3 3 5 2 5 4 3 3 5 4 4 6 6 3 2 5 3 5 2 5 2 6 2 2
3 2 2 1 6 2 6 4 6 5 5 5 6 2 2 6 3 6 6 1 6 3 5 1 5 4 2 2 3 2 6 6 2
5 5 4 2 5 3 6 2 4 3 1 6 2 5 5 5 6 2 2 6 3 4 2 6 1 2 2 6 6 2 3 2 3
5 6 5 6 6 6 1 6 5 5 3 5 1 6 3 2 2 6 5 6 3 2 2 5 6 1 2 5 3 5 1 5 4
6 2 6 6 6 2 2 3 2 2 5 5 3 5 2 1 4 4 2 5 6 6 2 3 2 1 2 5 3 2 1 5 5
4 4 2 5 3 2 3 2 6 4 2 5 5 6 6 5 2 3 3 6 6 4 2 2 1 3 6 6 3 3 4 3 6
6 4 2 6 3 6 2 6 2 5 2 6 5 6 5 5 3 4 5 5 5 2 6 5 6 6 2 6 3 1 3 2 5
6 2 2 1 2 6 5 1 5 5 6 3 2 2 4 1 2 5 5 4 2 2 1 3 2 5 5 5 2 3 1 5 4
2 2 3 6 2 5 5 2 1 6 3 6 2 2 3 6 3 5 5 5 5 1 1 1 4 3 5 6 3 4 6 2 6
5 5 1 2 4 6 5 2 6 6 1 2 2 5 3 4 3 3 2 5 3 6 2 5 3 3 2 2 4 2 6 5 5
5 6 2 6 5 3 4 1 5 2 3 5 6 3 5 5 5 5 5 2 3 2 5 5 6 3 2 5 2 5 2 6 5
4 5 3 3 2 3 5 2 5 5 2 3 4 1 3 5 3 4 4 2 3 3 1 4 6 6 5 6 1 5 6 5 5
6 2 6 1 1 5 2 3 3 2 5 3 4 2 5 2 2 3 5 5 3 3 5 3 2 2 5 5 6 2 2 5 4
4 1 2 6 5 5 2 3 1 4 6 2 3 5 5 5 4 6 5 1 5 5 6 5 4 6 1 2 3 6 5 5 6
6 1 5 2 5 3 1 3 3 3 2 1 4 4 6 3 6 4 5 4 2 2 5 4 5 1 3 3 5 6 2 6 5
5 2 2 6 2 5 2 3 4 2
```

(a) En base a los resultados obtenidos. ¿Qué decisión tomaría?

Estamos interesados en testear las hipótesis H_0 : El dado es equilibrado vs. H_1 : El dado está cargado. Para ello consideramos distintos tests:

(b) Sea X = cantidad de veces que el resultado del dado es par. Determinar la región de rechazo para un test de nivel $\alpha = 0.05$ basado en X. Para el conjunto de datos obtenido, ¿cuál es el valor p?, ¿está cargado el dado?

(c) Sea Y = cantidad de veces que el resultado del dado es menor o igual que 3. Determinar la región de rechazo para un test de nivel $\alpha = 0.05$ basado en Y. Para el conjunto de datos obtenido, ¿cuál es el valor p?, ¿está cargado el dado?

(d) Sea W = cantidad de veces que el resultado del dado es exactamente 3. Determinar la región de rechazo para un test de nivel $\alpha = 0.05$ basado en W. Para el conjunto de datos obtenido, ¿cuál es el valor p?, ¿está cargado el dado?

(e) Sea U = cantidad de veces que el resultado del dado es exactamente 2. Repetir el ítem anterior.

¿Está cargado el dado?

17. En cada caso indique si la afirmacin es verdadera o falsa y justifique:

(a) El nivel de significacin de un test es igual a la probabilidad de que la hiptesis nula sea cierta.

(b) Un Error de tipo II es ms grave que un Error de tipo I.

(c) Si el p-valor es 0.3, el test correspondiente rechazar al nivel 0.01.

(d) Si un test rechaza al nivel de significacin 0.06, entonces el p-valor es menor o igual a 0.06.

(e) Si un intervalo de confianza de nivel 0.99 para la media μ de una distribucin normal calculado a partir de una muestra da como resultado $(-2.0, 3.0)$, entonces el test para las hiptesis $H_0 : \mu = -3$ vs. $H_1 : \mu \neq -3$ basado en los mismos datos rechazar la hiptesis nula al nivel 0.01.