PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 8

Sobre la notación: A lo largo de esta materia, utilizamos, indistintamente, $\overline{x}_n$, $\overline{x}$, $\overline{X}_{\text{obs}}$ para denotar promedios. Un abuso de notación análogo usamos para los valores observados de otros estimadores.

1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.

   (a) Encontrar un intervalo de confianza de nivel exacto $1 - \alpha$ para la media cuando la varianza es conocida.

   (b) Con cierto instrumento se realizan 20 mediciones de una magnitud física μ. Cada observación es de la forma $X_i = \mu + \epsilon_i$, donde ϵ_i denota el error aleatorio de la i-ésima medición. El promedio de los valores obtenidos es $\overline{x} = 25.01$.

   Supongamos que los errores de medición tienen distribución normal con media cero y varianza 0.36.

   i) Hallar el intervalo de confianza estimado de nivel exacto 0.95 para el valor de la magnitud física de interés.

   ii) Si se quisiera que la longitud del intervalo hallado en i) fuera a lo sumo 0.05, ¿cuántas mediciones deberían realizarse?

   iii) Suponga ahora que la varianza del error es 0.6. ¿cuántas mediciones deberían realizarse para construir un intervalo de confianza exacto de nivel 0.95 para μ cuya longitud fuera a lo sumo 0.05?

   iv) Compare los resultados obtenidos en los items ii) y iii) e interprete la diferencia.

   (c) En `https://probac2019.shinyapps.io/datos_practica8_download/` encontrará una aplicación para generar resultados simulados de este experimento. Genere un conjunto de datos de tamaño 10 y rehaga el item b) i). Luego genere un conjunto de datos del tamaño hallado en el item b) ii), calcule el intervalo de confianza estimado y verifique que su longitud es menor que 0.5.

2. Consideremos variables aleatorias $X_1, \ldots, X_n$ con distribución $N(\mu, \sigma^2)$. En este ejercicio estudiaremos la distribución de $T_n = (\overline{X}_n - \mu)\sqrt{n}/s$ para diferentes valores de n. Para ello, para diferentes valores de n, y un μ y un σ dados, simularemos 1000 conjuntos de n datos con distribución $N(\mu, \sigma^2)$ y luego calcularemos el estadístico T_n para cada conjunto de datos. Obtendremos así 1000 realizaciones de la variable $T_n = (\overline{X}_n - \mu)\sqrt{n}/s$. Finalmente, haremos boxplots, qqplots, e histogramas de las 1000 realizaciones de T_n para estudiar su densidad.

   Generar dos valores con distribución $N(1, 4)$ y computar el valor que toma el estadístico T_n para esos valores observados, $T_{n\,obs}$. Replicar 1000 veces, guardando los resultados en un vector de longitud 1000.

(a) Hacer un boxplot y un qqplot del conjunto de valores de $T_{n\,obs}$. ¿Qué se observa ?

(b) A partir de los valores replicados de $T_{n\,obs}$ realizar un histograma. ¿Qué características tiene este histograma?

(c) Al histograma del item anterior, superponerle el gráfico de la densidad normal estándar. Superponer también el gráfico de la verdadera densidad de la variable T_n. ¿Qué observa?

(d) Con el comando `density`, dibujar la densidad estimada de T_n. En el mismo gráfico, superponer la verdadera densidad de T_n y la curva normal estándar. ¿Qué observa?

Repetir para $n = 4, 20, 30, 500$ y comparar los resultados.

3. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.

   (a) Encontrar un intervalo de confianza de nivel exacto $1 - \alpha$ para la media cuando la varianza es desconocida.

   (b) Repetir la parte i) del item b) del Ejercicio 1, suponiendo que la varianza es desconocida y el desvío estándar muestral es $s = 0.76$.

4. El rendimiento anual de almendros (en kg.) en parcelas cultivadas en cierta región es una variable aleatoria con distribución normal de media μ y varianza σ^2. Se observa el rendimiento anual de $n = 5$ parcelas elegidas al azar y se obtiene una media muestral de 525 kilogramos y un desvío muestral de 10 kilogramos.

   (a) En base a los datos obtenidos, obtenga una estimación por intervalos para μ utilizando un procedimiento de nivel 0.95.

   (b) Decidir si son verdaderas o falsas las siguientes afirmaciones:

      i. Aproximadamente el 95% de las parcelas de esta región tiene un rendimiento que pertenece al intervalo hallado en a), en base a la muestra observada.

      ii. La probabilidad de que el intervalo hallado en a) contenga a μ es 0.95.

      iii. Si se extraen muchas muestras de tamaño 5 de manera independiente y para cada una se calcula el intervalo de confianza como en a), aproximadamente el 95% de estos intervalos contendrá a μ.

5. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.

   (a) Encontrar un intervalo de confianza de nivel exacto $1-\alpha$ para la varianza cuando la media es conocida. SUGERENCIA:: probar que $\sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2 \sim \chi^2_n$.

   (b) 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 las siguientes longitudes

      Hallar el intervalo de confianza estimado de nivel exacto 0.90 para la varianza verdadera, suponiendo que $\mu = 185$.

176.50	191.50	186.90	181.10	195.70
188.10	187.40	185.10	176.90	191.20
193.80	187.00	179.00	173.00	184.40
199.60	190.40	206.80	193.00	177.10
201.00	192.50	176.60	180.10	186.40

6. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal.
   (a) Encontrar un intervalo de confianza de nivel exacto $1-\alpha$ para la varianza cuando la media es desconocida.
   (b) Repetir el item b) del Ejercicio 5 suponiendo que μ es desconocida.
7. (a) Definir una función que, tenga por input un conjunto de datos $x = (x_1, \ldots, x_n)$ provenientes de una distribución normal y un nivel de confianza $1-\alpha$ y devuelva un intervalo de confianza de nivel $1-\alpha$ para la media de la normal de la que provienen los datos. ¿Cómo cambiaria la función si ahora queremos aplicarla a normales con varianza conocida? ¿Y si ahora queremos aplicarla a muestras grandes pero con distribución desconocida?
   (b) Elegir un valor de μ y un valor de σ. Generar 1000 conjuntos de datos de n valores cada uno con distribución $N(\mu, \sigma^2)$ y calcular intervalos de confianza para μ asumiendo que μ es desconocido y σ es conocido. Repetir para diferentes valores de n entre 3 y 1000. ¿Qué proporción de intervalos contiene al verdadero valor de μ?
   (c) Rehacer el item anterior calculando también, para cada conjunto de datos, el intervalo de confianza para μ con σ desconocido. Para valores grandes de n, ¿de qué otra forma podría calcula los intervalos de confianza? Calcular los tres intervalos para cada muestra y compararlos.
   (d) Rehacer los dos items anteriores pero, en lugar de datos provenientes de una normal, generar datos provenientes de una $U(0,1)$. Interpretar.
8. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $E(\lambda)$.
   (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 es su longitud esperada?
   (d) Aplicar b) a los datos del Ejercicio 4 d) de la Práctica 7, con nivel $1-\alpha = 0.95$.
   (e) Hallar un intervalo de nivel asintótico $1-\alpha$ para λ.

9. Se desea conocer la opinión de los ciudadanos de cierta población acerca de una propuesta política. Para ello, se realiza una encuesta con tres posibles respuestas: a favor, en contra o indeciso.

   (a) Hallar un intervalo de confianza de nivel asintótico 0.90 para la proporción de votantes que se oponen a la propuesta. ¿Qué supuestos debe hacer?

   (b) Se realiza la encuesta a 1000 ciudadanos, obteniéndose como resultado que 200 están a favor 600 en contra y 200 están indecisos. Hallar el intervalo de confianza estimado de nivel asintótico 0.90 para la proporción de votantes que se oponen a la propuesta. Interpretar el resultado.

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

   (d) En `https://probac2019.shinyapps.io/datos_practica8_download/` encontrará una aplicación para generar resultados simulados de este experimento. Genere un conjunto de datos de tamaño 1000 y rehaga el item (b). Luego genere un conjunto de datos del tamaño hallado en el item (c), calcule el intervalo de confianza estimado y verifique que su longitud es menor que 0.02.

10. 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 a).

11. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $P\,(\lambda)$.

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

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

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

    (c) (**Opcional**) En los casos a) y b) obtener el intervalo de menor longitud esperada.

13. Sean X e Y dos variables aleatorias con distribución normal con la misma varianza: $X \sim N(\mu_1, \sigma^2)$ e $Y \sim N(\mu_2, \sigma^2)$. Considere las siguientes observaciones para X:

$$0.44,\ -1.63,\ 2.59,\ 1.54,\ 0.45,\ -0.13,\ -2.76,\ -1.53$$

y para Y:

$$0.06,\ -0.24,\ 4.65,\ 2.27,\ 3.88,\ 2.35,\ 3.92,\ -0.73.$$

(a) Calcule el intervalo de confianza estimado de nivel 0.95 para μ_1.

(b) Calcule el intervalo de confianza estimado de nivel 0.95 para μ_2.

(c) Calcule el intervalo de confianza estimado de nivel 0.95 para $\mu_1 - \mu_2$.