Inferencia Estadística

Tarea 6

14/11/2019

Escriba de manera concisa y clara sus resultados, justificando los pasos necesarios. Serán descontados puntos de los ejercicios mal escritos y que contenga ecuaciones sin una estructura gramatical adecuada. Las conclusiones deben escribirse en el contexto del problema. Todos los programas y simulaciones tienen que realizarse en `R`.

1. Las hojas de una planta se examinan buscando insectos. El número de insectos en una hoja sigue una distribución de Poisson con media μ, con la excepción de que muchas de las hojas no tienen insectos pues son inadecuadas para que se alimenten de ellas y esto no es simplemente el resultado de la variación aleatoria de la ley de Poisson.

   a) Encuentre la probabilidad condicional de que una hoja contenga i insectos, dado que contiene al menos uno.

   b) Supongamos que se observan x_i hojas conteniendo i insectos ($i = 1, 2, 3, \ldots$), con $\sum x_i = n$. Muestre que el estimador de máxima verosimilitud de μ satisface la ecuación

   $$\hat{\mu} = \overline{x}(1 - e^{-\hat{\mu}}),$$

   donde $\overline{x} = \sum i x_i / n$.

   c) Determine $\hat{\mu}$ numéricamente para el caso $\overline{x} = 3.2$. Utilice `R`.

2. Los siguiente son los tiempos (en horas) entre fallas sucesivas del sistema de aire acondicionado en un avión:

97	51	11	4	141	18	142	68	77
80	1	16	106	206	82	54	31	216
46	111	39	63	18	191	18	163	24.

   a) Suponiendo que estas son observaciones independientes de una distribución exponencial con media θ, encuentre $\hat{\theta}_{MLE}$.

   b) Haga una tabla de fecuencias para estos datos usando las clases $(0, 50]$, $(50, 100]$, $(100, 200]$ y $(200, \infty)$. Calcule el estimador de máxima verosimilitud de las frecuencias esperadas para estas clases bajo el supuesto en el inciso anterior. ¿La distribución exponencial parece ser un modelo adecuado para los datos?

3. Se hicieron 27 mediciones de los rendimientos de dos procesos industriales, con los siguientes resultados:

   Proceso 1: $n_1 = 11$ $\quad \overline{y}_1 = 6.23$ $\quad s_1^2 = 3.79$

   Proceso 2: $n_2 = 16$ $\quad \overline{y}_2 = 12.74$ $\quad s_2^2 = 4.17$

   Suponiendo que los rendimientos se distribuyen normalmente con la misma varianza, encuentre intervalos de confianza para las medias μ_1 y μ_2 y para la diferencia de medias $\mu_1 - \mu_2$.

4. Un experimento para determinar el efecto de una medicina en la concentración de glucosa en la sangre de ratas diabéticas dio los siguientes resultados:

Grupo control:	2.05	1.82	2.00	1.94	2.12	
Grupo tratamiento:	1.71	1.37	2.04	1.50	1.69	1.83

   Analiza la hipótesis de que el tratamiento no tiene efecto sobre la media de la concentración de la glucosa en la sangre. Mencione las hipótesis bajo las cuales realiza el análsis.

5. Sea Y el tiempo hasta que falla cierto componente eléctrico. La distribución de Y es exponencial con media θ/t, donde t es la temperatura a la cual el componente opera. Supongamos que n componentes se prueban de manera independiente a temperaturas $t_1, t_2, \ldots, t_n$, respectivamente, y que sus tiempos de vida observados son $y_1, y_2, \ldots, y_n$. Derive una expresión para el estimador de máxima verosimilitud de θ.

6. Diez componentes electrónicos con tiempos de vida distribuidos exponencialmente fueron probados por períodos de tiempo determinados. Tres de los componentes sobrevivieron sus periodos de prueba y los siete restantes fallaron en los siguientes tiempos.

No. Componente:	1	2	3	4	5	6	7	8	9	10
Período de prueba:	81	72	70	60	41	31	31	30	29	21
Tiempo de falla:	2	-	51	-	33	27	14	24	4	-

   Encuentre el estimado de máxima verosimilitud para la media de la distribución exponencial θ.

7. Una máquina de bebidas está diseñada para descargar, cuando opera apropiadamente, al menos 7 onzas de bebida por taza con una desviación estándar de 0.2 onzas. Si un estadístico selecciona una muestra aleatoria de 16 tazas para examinar el servicio al cliente y este está dispuesto a tomar un riesgo $\alpha = 0.05$ de cometer un error de Tipo I, calcule la potencia de la prueba y la probabilidad (β) de tener un error del Tipo II si la media poblacional de la cantidad despachada es:

   a) 6.9 onzas por taza.

   b) 6.8 onzas por taza.

   Puede asumir que los datos son normales.

8. Construya un intervalo de confianza aproximado del 90 % para el parámetro λ de una distribución de Poisson. Evalúe su intervalo si una muestra de tamaño 30 produce $\sum x_i = 240$.

9. Sea $X_1, \ldots, X_n \sim Uniforme(0, \theta)$ y sea $Y = \text{máx}\{X_1, \ldots, X_n\}$. Deseamos probar

$$H_0 : \theta = 1/2 \qquad vs. \qquad H_1 : \theta > 1/2.$$

   Supongamos que decidimos probar esta hipótesis rechazando H_0 cuando $Y > c$.

   a) Calcule el poder de la prueba.

   b) ¿Qué elección de c tene que tomarse para que la prueba tenga una significancia de 0.05?

   c) En una muestra de tamaño $n = 20$ con $Y = 0.48$, ¿cuál es el p-valor? ¿Qué conclusión sobre H_0 haría?

d) En una muestra de tamaño $n = 20$ con $Y = 0.52$, ¿cuál es el p-valor? ¿Qué conclusión sobre H_0 haría?

10. Sea $X_1, \ldots, X_n \sim Poisson(\lambda)$.

    a) Sea $\lambda_0 > 0$. Asumiendo que n es suficientemente grade, proponga un estadístico de prueba para examinar

    $$H_0 : \lambda = \lambda_0 \qquad vs \qquad H_1 : \lambda \neq \lambda_0.$$

    b) Establezca la región de rechazo para un nivel de significancia de $\alpha = 0.05$.

    c) Sea $\lambda_0 = 1$, $n = 20$ y $\alpha = 0.05$. Simule $X_1, \ldots, X_n \sim Poisson(\lambda_0)$ y aplique la prueba que propuso. Repita esto 10^4 veces y cuente que tan a menudo rechaza la hipótesis nula. ¿Qué tan cercana es la tasa de error tipo I de 0.05? ¿Qué hay de la tasa de error tipo II?

11. En la librería boot de `R` accesar los datos `cd4` los cuales son conteos de células CD4 en pacientes VIH-positivos antes y después de un año de tratamiento con un antiviral.

    a) Construya un intervalo de confianza bootstrap para el coeficiente de correlación entre los conteos base y los conteos después del tratamiento.

    b) Calcule el coeficiente estimado de correlación corregido por sesgo usando Jackknife.

12. Los siguientes 15 datos forman una muestra aleatoria de una distribución Gamma con parámetro de forma $\alpha = 3$ y parámetro de escala $\beta = 2$ (la media es $\alpha\beta$ y la varianza $\alpha\beta^2$)

14.18	10.99	3.38	6.76	5.56	1.26	4.05	4.61
1.78	3.84	4.69	2.12	2.39	16.75	4.19	

    Encuentre un intervalo de confianza para la mediana de la distribución.

13. Sean $X \sim Binomial(n_1, p_1)$ y $Y \sim Binomial(n_2, p_2)$, se quiere estimar $\delta = p_2 - p_1$. Use la priori $f(p_1, p_2) = f(p_1)f(p_2) = 1$, para hallar la posterior $f(\delta|x^n)$. Halle también la media posteriori y la densidad posteriori de δ.

14. Sean $X_1, \ldots, X_n$ Normal$(\mu, 1)$

    a) Simule un conjunto de datos (use $\mu = 5$) de $n = 100$ observaciones.

    b) Tome $f(\mu) = 1$ y halle la densidad posteriori. Grafique la densidad.

    c) Simule 1000 observaciones de la posteriori. Grafique un histograma y compare con la densidad del punto anterior.

    d) Sea $\theta = e^{\mu}$. Halle la densidad posteriori para θ de forma analítica y por simulación.

    e) Halle un intervalo posteriori del 95 % para θ.

    f) Halle un intervalo de confianza del 95 % para θ.

Entrega: 27/11/2019.