Laboratorio 11 - Inferencia Estadística

Pruebas de Hipótesis pt. 3 - Lema de Neyman-Pearson y Prueba UMP

Laboratorista: Héctor Lira Talancón

Ago-Dic 2017

1. Se tiene una muestra aleatoria de tamaño n de una densidad Geométrica $\big(X \sim G(p)\big)$con parámetro p, con $0 < p < 1$, dada por

$$P(X = x) = \begin{cases} p(1-p)^x & \text{si } x = 0, 1, 2, ... \\ 0 & \text{e.o.c.} \end{cases}$$

Se desea probar $H_0 : p = \frac{1}{2}$ vs. $H_1 : p = \frac{3}{4}$

a) Utilizar el Lema de Neyman-Pearson para encontrar una mejor región de rechazo de H_0 para esta prueba.

b) Si $n = 100$ y sabemos que si $X \sim G(p)$, entonces $E[X] = \frac{1-p}{p}$ y $Var[X] = \frac{1-p}{p^2}$. Utilizar el inciso a) para encontrar la constante necesaria para dar una mejor región de rechazo para la prueba al nivel $\alpha = 0.05$.

2. Se tiene una muestra aleatoria de tamaño n de una densidad Poisson $\big(X \sim Po(\lambda)\big)$ con parámetro $\lambda > 0$, dada por

$$P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{si } x = 0, 1, 2, ... \\ 0 & \text{e.o.c.} \end{cases}$$

Se desea probar $H_0 : \lambda = 2$ vs. $H_1 : \lambda = \lambda_1$ con $\lambda_1 > 2$.

a) Utilizar el Lema de Neyman-Pearson para encontrar una mejor región de rechazo de H_0 uniformemente más potente para esta prueba.

b) Si $n = 10$ utilizar el inciso a) para encontrar la constante necesaria para dar una mejor región de rechazo para la prueba al nivel $\alpha = 0.05$ o lo más cercano posible.

3. Se tiene una muestra aleatoria de tamaño n de una densidad Normal $\big(X_i \sim N(0, \sigma^2)$, para $i = 1, 2, ..., n\big)$.

Se desea probar $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 = \sigma_1^2$ con $\sigma_1^2 > \sigma_0^2$.

a) Utilizar el Lema de Neyman-Pearson para encontrar una mejor región de rechazo de H_0 uniformemente más potente para esta prueba.

b) Si $n = 20$, $\sigma_0^2 = 1$ utilizar el inciso a) para encontrar la constante necesaria para dar una mejor región de rechazo para la prueba al nivel $\alpha = 0.005$ o lo más cercano posible.

c) Si $n = 20$, $\sigma_0^2 = 1$, y $\alpha = 0.005$, proporcione la potencia de la prueba cuando $\sigma_1^2 = 2$.

4. Si $X_1, X_2, ..., X_n$ constituyen una muestra aleatoria de tamaño n de una población tal que $X_i \sim Gamma(1/2, \beta)$, y se plantea la prueba de hipótesis $H_0 : \beta = 2$ vs. $H_1 : \beta = 4$:

   a) Determine la región de rechazo uniformemente más potente para la prueba utilizando el criterio de Neyman-Pearson, para la cual se cometa un Error Tipo I de 0.05 con $n = 10$.

   b) Si se tiene una muestra de tamaño 10 con valores dados por: $1.7, 2.1, 1.8, 1.6, 1.5, 1.4, 1.9, 2.0, 1.6, 1.7$. Obtenga el nivel de significancia descriptivo de la prueba (p-value) y concluya.

5. Sea $Y_1, Y_2, ..., Y_n$ una muestra aleatoria de una densidad Bernoulli, es decir, $f(y; p) = p^y (1-p)^{1-y} \mathbb{1}_{\{0,1\}}(y)$

   a) Elabora la prueba uniformemente más potente para probar $H_0 : p = 0.5$ vs. $H_1 : p = 0.8$ de tamaño aproximado α en función de $W = \sum_{i=1}^{n} Y_i$.

   b) Calcula la región de rechazo correspondiente al inciso anterior si $n = 20$ y se desea $\alpha = 0.05$.

   c) Calcula la potencia de la prueba obtenida en el inciso b).

6. Sea $Y_1, Y_2, ..., Y_n$ una muestra aleatoria de una población cuya distribución de frecuencias tiene densidad $N(1, \sigma^2)$ con la varianza desconocida.

   a) Encuentra la prueba uniformemente más potente para probar $H_1 : \sigma^2 = 4$ vs. $H_1 : \sigma^2 = 9$.

   b) Calcula la región de rechazo para una muestra de tamaño 2 y 5% de nivel de significancia.

7. Considera $X_1, X_2, ..., X_n$ una muestra aleatoria de una densidad $f(x|\theta) = \frac{x}{\theta} e^{\frac{-x^2}{2\theta}} \mathbb{1}_{(0,\infty)}(x)$ con $\theta > 0$. Encuentra el estadístico de prueba y la región de rechazo uniformemente más potente para probar $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ con $\theta_1 < \theta_0$.

8. Considera $X_1, X_2, ..., X_n$ una muestra aleatoria de una densidad $N(\mu, \sigma^2 = 4)$.

   a) Obtén la prueba uniformemente más potente para probar $H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1$ con $\mu_0 < \mu_1$.

   b) Encuentra la región de rechazo para un nivel de significancia del 10%, $n = 9$ y $\mu_0 = 5$.

   c) Calcula la potencia de esta prueba para $\mu_1 = 7$.

9. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una densidad $Exp(\beta)$.

   a) Encuentra la región UMP para probar $H_0 : \beta = \beta_0$ vs. $H_1 : \beta = \beta_1$ con $\beta_0 < \beta_1$ en función de $\hat{\beta} = \bar{X}$.

   b) Encuentra la región UMP para probar $H_0 : \beta = 2$ vs. $H_1 : \beta = 3$ si $n = 7$ con $\alpha = 0.05$ en función de $\hat{\beta}$.

   c) Obtén una cota inferior y una cota superior de la potencia de la región obtenida en b) cuando $\beta = 3$.

10. Se desea estimar la probabilidad (p) de que un encendedor funcione. Para esto se toma una muestra aleatoria de tamaño n de un lote de encendedores. Para el i-ésimo encendedor en la muestra se observa X_i, el número de ensayos hasta que falla por primera vez. Se supone el siguiente modelo $f(x|p) = (1-p)p^{x-1}\mathbb{1}_{\{1,2,...\}}(x)$ para $0 < p < 1$.

    Encuentra la región uniformemente más potente para probar $H_0 : p = p_0$ vs. $H_1 : p = p_1$ con $p_0 < p_1$.

11. Suponga que $X_1, X_2, ..., X_n$ es muestra aleatoria con $f_{X_i}(x_i|\theta) = \theta x_i^{\theta-1}\mathbb{1}_{(0,1)}(x_i), \theta > 0$. Se desea probar $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ con $\theta_1 > \theta_0$.

    a) Mediante el Lema de Neyman-Pearson, demuestre que la región de rechazo de la correspondiente prueba uniformemente más potente es de la forma $RR = \left\{ -\sum_{i=1}^{n} \ln(X_i) \leq \lambda \right\}$

    b) Demuestre que para una prueba de tamaño α, el valor crítico de la región de rechazo del inciso anterior es $\lambda = \frac{\chi^2_{2n,1-\alpha}}{2\theta_0}$ (sugerencia: primero demuestre que $-\ln(X_i) \sim Exp(\frac{1}{\theta})$ y luego recuerde que si $S \sim Gamma(\alpha, \beta)$, entonces $\frac{2S}{\beta} \sim \chi^2_{(2\alpha)}$).

    c) Suponga que X_i denota el porcentaje de daño que sufre un hotel en caso de tsunami. A partir de una muestra aleatoria de 5 hoteles del sureste asiático afectados por tsunamis se observó que los daños fueron de $60\%, 75\%, 68\%, 80\%$ y 92%. Un reportero observa estos datos y publica una nota diciendo que: "la media poblacional del daño que sufren los hoteles en caso de tsunami es mayor a 75%". Realice una prueba de hipótesis de tamaño 5% para determinar si la afirmación del reportero es correcta (sugerencia: Note que $E[X_i] = \frac{\theta}{\theta+1}$ para plantear adecuadamente sus hipótesis).

12. Se desea monitorear un proceso de pintado de cofres de autos de un modelo dado (digamos Toyota Camry). La variable de interés es X, el número de defectos de pintura detectados en el cofre por el inspector. Se supone que X tiene una distribución de Poisson con media μ.

    a) Supongamos que se tienen $X_1, X_2, ..., X_n$ variables aleatorias i.i.d. con distribución de Poisson con parámetro μ. Deduce la prueba uniformemente más potente para probar $H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1$ con $\mu_1 > \mu_0$. Justifica.

b) En el proceso productivo se toma de manera aleatoria una muestra de dos cofres de la producción de cada hora. Obtén la región de rechazo de la hipótesis nula con un nivel de significancia lo más cercano a 5% si $\mu_0 = 1.6$ defectos por cofre.

c) En la producción de las 10:00 a las 11:00 horas se observaron 2 y 3 defectos de pintura respectivamente en los cofres inspeccionados. ¿Hay evidencia de que μ aumentó?

d) Calcula la potencia de la prueba para detectar si la media del proceso cambió a

$\mu_1 = 2, 2.5, 3$. Haz una gráfica.