Estadística
CURE **Práctico 3** - parte (b) Curso 2018

**EJERCICIO 1.** Sea $X \sim$ Binomial$(2, \theta)$ y consideremos el test $H_0 : \theta = 1/2$ vs. $H_1 : \theta = 3/4$.

(a) Aplique el lemma de Neyman-Pearson con $3/4 < k < 9/4$ para demostrar que el test que rechaza $H_0$ si $X = 2$ es un test UMP (uniformly most powerful) de nivel $\alpha = 1/4$.

(b) Discuta los otros casos $1/4 < k < 3/4$, $k < 1/4$ y $k > 9/4$.

(c) Repita la parte (a) para el caso $k = 3/4$.

**EJERCICIO 2.** En este ejercicio consideramos un test UMP en el caso normal. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $\mathcal{N}(\theta, \sigma^2)$. Supondremos que $\sigma^2$ es conocido.

A continuación enunciamos un corolario del Lema de Neyman-Pearson.

**Proposición.** *En las condiciones del Lema de Neyman-Pearson, suponga que $T(X)$ es un estadístico suficiente para $\theta$ y que $g(t|\theta_i)$ es la pdf o pmf de $T$ correspondiente a $\theta_i$, $i = 0, 1$. Entonces cualquier test basado en $T$ con región de rechazo $S$ es un test UMP de nivel $\alpha$ si satisface*

$$t \in S \text{ si } g(t|\theta_1) > kg(t|\theta_0), \tag{1}$$

*y*

$$t \in S^C \text{ si } g(t|\theta_1) < kg(t|\theta_0), \tag{2}$$

*para algún $k \geq 0$, donde*

$$\alpha = P_{\theta_0}(T \in S). \tag{3}$$

(a) Considere el test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, con $\theta_0 > \theta_1$. Utilizando el corolario demuestre que el test con región de rechazo $\bar{x} < c$ es un test UMP de nivel $\alpha$, donde $\alpha = P_{\theta_0}(\bar{X} < c)$. Dé una expresión para $c$.

(b) Considere el test $H_0' : \theta \geq \theta_0$ vs. $H_1' : \theta < \theta_0$ usando el tet que rechaza $H_0'$ si $\bar{X} < -\frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0$. Demuestre que este es un test UMP de nivel

$$\alpha = P_{\theta_0}\left(\bar{X} < -\frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\right).$$

**EJERCICIO 3.** Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $\mathcal{N}(\theta, \sigma^2)$. Supondremos que $\sigma^2$ es conocido.

(a) Considere el test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Demuestre que no existe un test UMP de nivel $\alpha$ para este problema.

(b) Demuestre que existe un test de nivel $\alpha$ que es UMP *en la clase de tests insesgados.*

**EJERCICIO 4.** En este problema consideramos el *Test de Exactitud de Fisher*. Sean $S_1$ y $S_2$ observaciones independientes, con $S_1 \sim \text{Binomial}(n_1, p_1)$ y $S_2 \sim \text{Binomial}(n_2, p_2)$. Considre el test $H_0 : p_1 = p_2$ vs. $H_1 : p_1 > p_2$. Sea $p$ el valor común de $p_1 = p_2$ bajo $H_0$. Demuestre que el $p$-valor condicional es $p(s_1, s_2) = \sum_{j=s_1}^{\text{mín}\{n_1,s\}} f(j|s)$, la suma de probabilidades (distribución hipergeométrica). El test definido con este $p$-valor se llama Test de Exactitud de Fisher.

**EJERCICIO 5.** Sea $X_1, \ldots, X_n$ una familia con distribución de Poisson de parámetro $\lambda$. Supongamos que $\lambda$ tiene una distribución $\text{Gamma}(\alpha, \beta)$ (la familia conjugada de la Poisson). Considere un test Bayesiano $H_0 : \lambda \leq \lambda_0$ vs. $H_1 : \lambda > \lambda_0$.

(a) Calcule expresiones para las probabilidades a posteriori de $H_0$ y de $H_1$.

(b) Si $\alpha = 5/2$ y $\beta = 2$, la distribución a priori es una distribución chi-cuadrado con 5 grados de libertad. Explique cómo una tabla de la distribución chi-cuadrado se podría utilizar para realizar un test Bayesiano.

**EJERCICIO 6.** Para muestras de tamaño $n = 1, 4, 16, 64, 100$ de una población normal con media $\mu$ y varianza conocida $\sigma^2$, grafique la función de potencia de los siguientes LRTs. Tome $\alpha = 0.05$.

(a) $H_0 : \mu \leq 0$ vs. $H_1 : \mu > 0$.

(b) $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$.

**EJERCICIO 7.** Sea $(X_1, X_2), \ldots, (X_n, Y_n)$ una muestra aleatoria de una población con distribución normal bivariada con parámetros $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$. Estamos interesados en el test $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$.

(a) Demuestre que las v.a. $W_i = X_i - Y_i$ son iid $\mathcal{N}(\mu_W, \sigma_W^2)$.

(b) Demuestre que la hipótesis se puede testear utilizando el estadístico

$$T_W = \frac{\bar{W}}{\sqrt{\frac{1}{n} S_W^2}},$$

donde $\bar{W} = \frac{1}{n} \sum_{i=1}^{n} W_i$ y $S_W^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (W_i - \bar{W})^2$.

(c) (opcional) Demuestre que bajo $H_0$, $T_W \sim$ Student's $t$ con $n - 1$ grados de libertad.

**EJERCICIO 8.** Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población $\mathcal{N}(\mu_X, \sigma_X^2)$, y sea $Y_1, \ldots, Y_m$ una v.a. independiente de una distribución $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Estamos interesados en el test $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$, con la hipótesis que $\sigma_X^2 = \sigma_Y^2 = \sigma^2$.

(a) Calcule las LRT bajo estas hipótesis. Demuestre que el LRT se puede basar en el estadístico

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m}\right)}},$$

donde

$$S_p^2 = \frac{1}{(n+m-2)}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{m}(Y_i-\bar{Y})^2\right).$$

(b) Demuestre que bajo $H_0$, $T \sim t_{n+m-2}$. Este test se llama test $t$ a dos muestras (*two-sample t test*).

(c) Muestras de madera se obtuvieron del centro y de la periferia de cierta iglesia Bizantina. La fecha de la madera se determinó dando los datos siguientes. Utilice el test $t$ a dos muestras para determinar si el promedio de edad del centro es el mismo que el promedio de edad de la periferia.

| Centro | | Periferia | |
|---|---|---|---|
| 1294 | 1251 | 1284 | 1274 |
| 1279 | 1248 | 1272 | 1264 |
| 1274 | 1240 | 1256 | 1256 |
| 1264 | 1232 | 1254 | 1250 |
| 1263 | 1220 | 1242 | |
| 1254 | 1218 | | |
| 1251 | 1210 | | |

**EJERCICIO 9.** La hipótesis de varianzas iguales que se hizo en el ejercicio anterior (Ejercicio 8) no siempre se puede utilizar. En tal caso, la distribución del estadístico $T$ no sigue una distribución $t$. Este problema de hacer inferencia en promedios cuando las varianzas no son iguales se llama problema de *Behrens-Fisher*.

Considere la siguiente modificación del test $t$ a dos muestras: $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$, donde no asumimos $\sigma_X^2 = \sigma_Y^2$. Utilizamos el estadistico

$$T' = \frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)}},$$

donde

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2, \qquad S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2.$$

La distribución exacta de $T'$ es difícil de calcular, pero se puede utilizar la aproximación de *Satterthwaite*.

(a) Demuestre que, aproximadamente,

$$\frac{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} \sim \frac{\chi_\nu^2}{\nu}.$$

(b) La distribución de $T'$ se puede aproximar por una distribución $t$ con $\hat{\nu}$ grados de libertad. Justifique.

(c) Re-examine los datos del Ejercicio 8 utilizando el test $t$ aproximado de este ejercicio. Es decir, se considerará testeará si la edad promedio del centro es la misma que la edad promedio de la periferia utilizando el estadístico $T'$.

(d) ¿Hay alguna evidencia estadística que la varianza de los datos del centro puede ser diferente de la varianza de los datos de la periferia?