Laboratorio 3 (Respuestas) - Inferencia Estadística

Distribuciones Muestrales

Laboratorista: Héctor Lira Talancón

Ago-Dic 2017

1. Una muestra aleatoria es:

   a) Cualquier conjunto de variables aleatorias con la misma distribución.

   b) Un conjunto de variables aleatorias independientes con la misma distribución.

   c) Aquella que siempre da a lugar a un estimador insesgado.

   d) Ninguna de las anteriores.

2. Un parámetro es:

   a) Un valor fijo que nos permite estimar el verdadero valor de la población.

   b) Una función de variables aleatorias.

   c) Un valor fijo y desconocido que se desea estimar.

   d) Una variable aleatoria cuyo valor es calculado a partir de los datos de la muestra aleatoria.

3. Considera las siguientes afirmaciones:

   A: Una estadística es cualquier función de las variables aleatorias que se observaron en la muestra, que inclusive puede contener parámetros.

   B: Un parámetro es una caraterística numérica de la distribución poblacional de una variable de interés.

   Entonces:

   a) A y B son falsas.

   b) A es falsa pero B no.

   c) B es falsa pero A no.

   d) Ninguna de las anteriores.

4. Se lleva a cabo una auditoría por muestreo aleatorio simple (MAS) a las 10350 operaciones llevadas a cabo por una empresa. Se tomó una muestra de 120 operaciones las cuales se analizaron detenidamente. X es el ingreso total de la operación. Sea σ_X^2 la varianza poblacional de X. Considera las siguientes afirmaciones:

   A. El MAS justifica que $Var(\bar{X}) = \sigma_X^2/120$.

   B. El MAS justifica que $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2$.

   C. El MAS es un elemento que justifica la normalidad aproximada del promedio muestral.

   Entonces:

   a) A y B son apropiadas

   b) A y C son apropiadas

   c) A y B no son apropiadas

   d) B y C son apropiadas

5. La distribución de muestreo de cualquier estadístico

   a) Es aproximadamente normal para un tamaño de muestra grande.

   b) Tiene un valor esperado igual al parámetro que se desea estimar.

   c) Especifica cómo se comportan los diferentes valores del estadístico.

   d) Todas las anteriores.

6. La varianza de un estimador es:

   a) El cuadrado de la diferencia entre el estimador y la esperanza de dicho estimador.

   b) La diferencia entre el error cuadrático medio del estimador y su sesgo.

   c) Un promeido ponderado de los posibles valores del estimador.

   d) El valor esperado del cuadrado de la diferencia entre el estimador y su valor esperado.

7. Cierto artículo tiene un precio de oferta en el mercado igual a p_o, el cual se distribuye normalmente con media de 50 pesos y desviación estándar de 5 pesos. El precio máximo que están dispuestos a pagar los consumidores (precio de demanda), también es una variable aleatoria p_d, cuya distribución es normal con media de 45 pesos y desviación estándar de 2.50 pesos. Calcule la probabilidad de que tenga lugar una transacción (sugerencia: las transacciones tendrán lugar si y solo si el precio de oferta es menor o igual al de demanda).

8. Sea $Y_1, ..., Y_{10}$ una muestra aleatoria de una población con distribución $N(0,1)$. Sea $U = \sum_{i=1}^{10} Y_i^2$ y $Q = \frac{3Y_{10}}{\sqrt{U}}$. El intervalo que contiene $P(|Q| < 2)$ es:

   a) $(0.95, 0.98]$

   b) $(0.05, 0.1)$

   c) $(0.90, 0.95)$

   d) Ninguno de los anteriores.

9. Sean $X_1, ..., X_m$ y $Y_1, ..., Y_n$ dos muestras aleatorias independientes tal que $X \sim N(\mu_x, \sigma^2)$ y $Y \sim N(\mu_y, \sigma^2)$. Sean $S_X^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X})^2$ y $S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$.

   a) Encuentra el valor esperado de $S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$.

   b) Encuentra la varianza de S_p^2.

c) Encuentra dos números $0 < a < b$ tales que $a < P(\frac{S_X^2}{S_Y^2} > 20) < b$ si $m = 5$ y $n = 4$.

10. Marque con una 'x' las afirmaciones que sean falsas.

☐ Si $Z \sim N(0,1)$ y $Y \sim \chi^2_{(n)}$ son independientes, entonces $T = \frac{Z}{(Y/n)} \sim t_n$.

☐ Si $X_1, ..., X_n$ son n variables aleatorias independientes tales que $X_i \sim N(0,1)$, para $i = 1, ..., n$ entonces $S^2 = \sum\limits_{i=1}^{n} X_i^2 \sim \chi^2_{(n-1)}$.

☐ Si $U \sim \chi^2_{(n)}$ y $V \sim \chi^2_{(m)}$ son dos variables aleatorias independientes, sabemos que $F = \frac{(U/n)}{(V/m)} \sim F_{(n,m)}$. Entonces $G = \frac{(V/m)}{(U/n)} \sim F_{(m,n)}$.

☐ Si $X_1, ..., X_n$ son n variables aleatorias tales que $X_i \sim N(0, \sigma^2)$, para $i = 1, ...n$, entonces $W = \frac{\sum\limits_{i=1}^{n} X_i}{n} \sim N(\mu, \frac{\sigma^2}{n})$.

11. Sean X_1 y X_2 una muestra aleatoria de tamaño 2 de una distribución $N(1,1)$. Entonces, la distribución de $(X_2 + X_1 - 2)^2/(X_2 - X_1)^2$ es:

a) $F_{(2,1)}$

b) $F_{(1,1)}$

c) $F_{(2,2)}$

d) Ninguna de las anteriores.

12. Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución $N(\mu, \sigma^2)$ y $\bar{X}$ y S^2 la media y la varianza muestral, respectivamente. Sea $X_{n+1} \sim N(\mu, \sigma^2)$ y asuma que $X_1, ..., X_{n+1}$ son independientes. Encuentre la distribución de:

$$\frac{X_{n+1} - \bar{X}}{S\sqrt{\frac{n+1}{n}}}$$

Sabemos que si $X_i \sim N(\mu, \sigma^2)$ entonces $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Analizando el numerador
$X_{n+1} - \bar{X} \sim N(0, \frac{\sigma^2}{n} + \sigma^2) \equiv N(0, \sigma^2(\frac{1}{n} + 1)) \equiv N(0, \sigma^2(\frac{n+1}{n}))$

Entonces,
$\frac{X_{n+1} - \bar{X}}{\sqrt{\frac{n+1}{n}}} \sim N(0, \sigma^2)$

Esto implica que
$\frac{X_{n+1} - \bar{X}}{\sigma\sqrt{\frac{n+1}{n}}} \sim N(0,1)$

Sabemos que
$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$

Reescribiendo la expresión original obtenemos

$$\frac{X_{n+1}-\bar{X}}{S\sqrt{\frac{n+1}{n}}} = \frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}}} \cdot \frac{1}{\sqrt{S^2}} \cdot \frac{\sigma}{\sigma} \cdot \frac{\sqrt{n-1}}{\sqrt{n-1}}$$

$$= \frac{X_{n+1}-\bar{X}}{\sigma\sqrt{\frac{n+1}{n}}} \cdot \frac{\sqrt{n-1}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}}}$$

$$= N(0,1) \cdot \frac{1}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1}}}$$

$$= \frac{N(0,1)}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1}}}$$

Por lo tanto, la expresión sigue una distribución $t_{(n-1)}$.

13. Sean $X_1, ..., X_m$ y $Y_1, ..., Y_n$ dos muestras aleatorias independientes de distribuciones $N(\mu_1, \sigma^2)$ y $N(\mu_2, \sigma^2)$, respectivamente. Sean α y β dos números reales fijos. Si $\bar{X}_m$ y $\bar{Y}_m$ denotan las correspondientes medias muestrales, encuentre la distribución de:

$$\frac{\alpha(\bar{X}_m-\mu_1)+\beta(\bar{Y}_n-\mu_2)}{\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}}\sqrt{\frac{\alpha^2}{m}+\frac{\beta^2}{n}}}$$

Analicemos el numerador de esta expresión:
$\alpha(\bar{X}_m-\mu_1)+\beta(\bar{Y}_n-\mu_2) = \alpha\bar{X}_m - \alpha\mu_1 + \beta\bar{Y}_n - \beta\mu_2 = (\alpha\bar{X}_m + \beta\bar{Y}_n) - (\alpha\mu_1+\beta\mu_2)$

donde,
$\alpha\bar{X}_m = \sum\limits_{i=1}^{m} \frac{\alpha}{m}X_i \sim N(\sum\limits_{i=1}^{m} \frac{\alpha}{m}\mu_1, \sum\limits_{i=1}^{m} (\frac{\alpha}{m})^2\sigma^2) \equiv N(\alpha\mu_1, \frac{\alpha^2\sigma^2}{m})$
y
$\beta\bar{Y}_n = \sum\limits_{i=1}^{n} \frac{\beta}{n}Y_i \sim N(\sum\limits_{i=1}^{n} \frac{\beta}{n}\mu_2, \sum\limits_{i=1}^{n} (\frac{\beta}{n})^2\sigma^2) \equiv N(\beta\mu_2, \frac{\beta^2\sigma^2}{n})$

Tenemos entonces que
$(\alpha\bar{X}_m + \beta\bar{Y}_n) \sim N(\alpha\mu_1+\beta\mu_2, \frac{\alpha^2\sigma^2}{m} + \frac{\beta^2\sigma^2}{n}) \equiv N(\alpha\mu_1+\beta\mu_2, \sigma^2(\frac{\alpha^2}{m}+\frac{\beta^2}{n}))$

De aquí, vemos fácilmente que
$(\alpha\bar{X}_m + \beta\bar{Y}_n) - (\alpha\mu_1+\beta\mu_2) \sim N(0, \sigma^2(\frac{\alpha^2}{m}+\frac{\beta^2}{n}))$

Escalando esta normal obtenemos que
$\frac{(\alpha\bar{X}_m+\beta\bar{Y}_n)-(\alpha\mu_1+\beta\mu_2)}{\sqrt{\frac{\alpha^2}{m}+\frac{\beta^2}{n}}} \sim N(0,\sigma^2)$

Esto implica que
$\frac{(\alpha\bar{X}_m+\beta\bar{Y}_n)-(\alpha\mu_1+\beta\mu_2)}{\sigma\sqrt{\frac{\alpha^2}{m}+\frac{\beta^2}{n}}} \sim N(0,1)$

Por último, definamos
$S_p^2 = \frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2} = \frac{1}{m+n-2}[\sigma^2\frac{(m-1)S_1^2}{\sigma^2} + \sigma^2\frac{(n-1)S_2^2}{\sigma^2}] = \frac{\sigma^2}{m+n-2}[\frac{(m-1)S_1^2}{\sigma^2} + \frac{(n-1)S_2^2}{\sigma^2}]$

$= \frac{\sigma^2}{m+n-2}[\chi^2_{(m-1)} + \chi^2_{(n-1)}] = \frac{\sigma^2}{m+n-2}[\chi^2_{(m+n-2)}]$

Esto nos sugiere que
$\frac{(m+n-2)S_p^2}{\sigma^2} \sim \chi^2_{(m+n-2)}$

Reescribiendo la expresión original obtenemos

$$\frac{\alpha(\bar{X}_m - \mu_1) + \beta(\bar{Y}_n - \mu_2)}{\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}}\sqrt{\frac{\alpha^2}{m} + \frac{\beta^2}{n}}} = \frac{\alpha(\bar{X}_m - \mu_1) + \beta(\bar{Y}_n - \mu_2)}{\sqrt{S_p^2}\sqrt{\frac{\alpha^2}{m} + \frac{\beta^2}{n}}}$$

$$= \frac{(\alpha\bar{X}_m + \beta\bar{Y}_n) - (\alpha\mu_1 + \beta\mu_2)}{\sigma\sqrt{\frac{\alpha^2}{m} + \frac{\beta^2}{n}}} \cdot \frac{\sqrt{m+n-2}}{\sqrt{\frac{(m+n-2)S_p^2}{\sigma^2}}}$$

$$= N(0,1) \cdot \frac{\sqrt{m+n-2}}{\sqrt{\chi^2_{(n+m-2)}}}$$

$$= \frac{N(0,1)}{\sqrt{\frac{\chi^2_{(n+m-2)}}{(n+m-2)}}}$$

Por lo tanto, la expresión sigue una distribución $t_{(n+m-2)}$.

14. Sean $Y_1, ..., Y_n$ y $Z_1, ..., Z_k$ dos muestras aleatorias independientes de distribuciones $N(\mu, \sigma^2)$ y $N(0,1)$, respectivamente, y sea $W = c\frac{(\bar{Y}-\mu)}{\sqrt{\sum_{i=1}^{k} Z_i^2}}$. El valor de c que hace que W tenga distribución t-Student debe ser:

a) $c = 1$

b) $c = \frac{\sqrt{nk}}{\sigma}$

c) $c = \frac{\sqrt{k-1}}{\sigma}$

d) Ninguna de las anteriores.

Sabemos que si $Y_i \sim N(\mu, \sigma^2)$ entonces $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$ y $(\bar{Y} - \mu) \sim N(0, \frac{\sigma^2}{n})$.
Para obtener la distribución t-Student necesitamos tener una normal estándar en el numerador y la raíz de una Ji-cuadrada dividida entre sus grados de libertad en el denominador.
Para obtener la normal estándar en el numeriador necesitamos dividir $(\bar{Y} - \mu)$ entre $\frac{\sigma}{\sqrt{n}}$, esto es, que c contenga $\frac{\sqrt{n}}{\sigma}$ (para estar dividiendo a $(\bar{Y} - \mu)$ entre su desviación estándar).
Para obtener la raíz cuadrada de una variable $\chi^2_{(n)}$ dividida entre sus grados de libertad nos fijamos primero en que $\sum_{i=1}^{k} Z_i^2$ sigue una distribución $\chi^2_{(k)}$. Esto nos indica que c debe de contener $\sqrt{k}$ (para estar dividiendo al término del denominador).
Por lo tanto, c debe de ser $\frac{\sqrt{nk}}{\sigma}$.
El inciso correcto es b).

15. Sea $X_1, ..., X_5$ una muestra aleatoria de una densidad $N(0, \sigma^2), \sigma^2$ conocida. Entonces la distribución de $Y = \sum_{i=1}^{5} \frac{X_i^2}{\sigma^2}$ es:

a) $F_{(4,1)}$

b) $F_{(5,1)}$

c) $\chi^2_{(5)}$

d) $\chi^2_{(4)}$

Sabemos que si tenemos $Z_i \sim N(0,1)$ entonces $\sum\limits_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$.

Dado que $X_i \sim N(0,\sigma^2) \Rightarrow \frac{X_i}{\sigma} = Z_i \sim N(0,1)$.

Por lo tanto, $Y = \sum\limits_{i=1}^{5} \frac{X_i^2}{\sigma^2} = \sum\limits_{i=1}^{5} (\frac{X_i}{\sigma})^2 = \sum\limits_{i=1}^{5} Z_i^2 \sim \chi^2_{(5)}$.

El inciso correcto es c).

16. Sea $X_1, X_2, ...$ una sucesión de variable aleatorias i.i.d. con densidad Bernoulli con parámetro p. Definimos $T_n = \frac{(\sum\limits_{i=1}^{n} X_i)+1}{n+2}$ como estimador de p.

a) Encuentra $Var(T_n)$.

$Var(T_n) = Var(\frac{(\sum\limits_{i=1}^{n} X_i)+1}{n+2}) = \frac{1}{(n+2)^2} Var((\sum\limits_{i=1}^{n} X_i)+1) = \frac{1}{(n+2)^2} Var(\sum\limits_{i=1}^{n} X_i) = \frac{1}{(n+2)^2} \sum\limits_{i=1}^{n} Var(X_i)$

$= \frac{1}{(n+2)^2} \sum\limits_{i=1}^{n} p(1-p) = \frac{1}{(n+2)^2} (np(1-p)) = \frac{np(1-p)}{(n+2)^2}$

b) Calcula el error cuadrático medio de T_n. ¿Cuál es el límite de esta cantidad cuando $n \to \infty$?

El error cuadrático medio se define como sigue:

$ECM(T_n) = E[(T_n - p)^2]$

Desarrollando la expresión obtenemos que

$E[(T_n - p)^2] = E[T_n^2 - 2T_n p - p^2] = E[T_n^2] - 2E[T_n]p - p^2$

$= E[T_n^2] - 2E[T_n]p - p^2 + E^2[T_n] - E^2[T_n] = E[T_n^2] - E^2[T_n] + E^2[T_n] - 2E[T_n]p - p^2$

$= (E[T_n^2] - E^2[T_n]) + (E[T_n] - p)^2$

$= Var(T_n) +$ sesgo$^2(T_n)$

Ahora bien,

sesgo$(T_n) = E[T_n] - p = E[\frac{(\sum\limits_{i=1}^{n} X_i)+1}{n+2}] - p = \frac{1}{n+2} E[(\sum\limits_{i=1}^{n} X_i + 1] - p$

$= \frac{1}{n+2}(E[\sum\limits_{i=1}^{n} X_i] + 1) - p = \frac{1}{n+2}(\sum\limits_{i=1}^{n} E[X_i] + 1) - p$

$= \frac{1}{n+2}(\sum\limits_{i=1}^{n} p + 1) - p = \frac{1}{n+2}(np+1) = \frac{np}{n+2} - (\frac{n+2}{n+2})p$

$= \frac{np+1-np-2p}{n+2} = \frac{1-2p}{n+2}$

Por lo tanto,

$ECM(T_n) = \frac{np(1-p)}{(n+2)^2} + (\frac{1-2p}{n+2})^2 \to 0$ si $n \to \infty$.