Parcial 1 - Estadística II

ITAM, Primavera 2020 10/02/2021

Instrucciones

El examen es para resolver en casa. Se debe contestar individualmente y entregarse a más tardar a las 23:59 del jueves 11 de febrero. La entrega será via email a la dirección: salvador.garcia.gonzalez@itam.mx (la misma que aparece en comunidad). El examen cuenta dos secciones, la segunda con preguntas a desarrollar. Se debe cuidar la formalidad al escribir los resultados, ya que es parte de la calificación del problema. En caso de no tener el desarrollo de la pregunta, o bien se llegué a la respuesta sin una justificación se podrá anular la respuesta. Cualquier práctica fraudulenta será sancionada de acuerdo al reglamento del departamento.

Seccion A: Preguntas de teoría (45 pts)

1. **20 - Propiedades varianza y esperanza**

   Sea X una variable aleatoria con E(X) = μ y V(X) = σ^2. Y se distribuye igual a X. Se toma n muestras de X y m de Y. Calcule las siguientes expresiones:

   $E(X^2) = V(X) + E(X)^2 = \sigma^2 + \mu^2$
   $E(\bar{X}^2) = V(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2$

   1. Usando X calcule la esperanza de la media muestral: $E(\bar{X})$

      Solución $E(\frac{\sum_i X_i}{n}) = \frac{1}{n}E(\sum_i X_i) = \frac{1}{n}\sum_i E(X) = \frac{1}{n}nE(X) = \mu$

   2. Usando X calcule la esperanza de la varianza muestral (sesgada): $E(S_n^2)$

      Solución $E(\frac{\sum_i X_i^2 - n\bar{X}^2}{n}) = \frac{E(\sum_i X_i^2) - nE(\bar{X}^2)}{n} = \frac{\sum_i E(X_i^2) - nE(\bar{X}^2)}{n} = \frac{n\sigma^2 + n\mu^2 - n\frac{\sigma^2}{n} - n\mu^2}{n} = \frac{(n-1)\sigma^2}{n}$

   3. Usando X calcule la esperanza de la varianza muestral (insesgada): $E(S^2)$

      Solución $E(\frac{\sum_i X_i^2 - n\bar{X}^2}{n-1}) = \frac{E(\sum_i X_i^2) - nE(\bar{X}^2)}{n-1} = \frac{\sum_i E(X_i^2) - nE(\bar{X}^2)}{n-1} = \frac{n\sigma^2 + n\mu^2 - n\frac{\sigma^2}{n-1} - n\mu^2}{n} = \sigma^2$

   4. Usando X calcule la var'ianza de la media muestral: $V(\bar{X})$

      Solución $V(\frac{\sum_i X_i}{n}) = \frac{1}{n^2}V(\sum_i X_i) = \frac{1}{n^2}\sum_i V(X) = \frac{1}{n}\sigma^2$

   5. Usando X y Y calcule la esperanza de la diferencia de medias muestrales: $E(\bar{X} - \bar{Y})$

      Solución $E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = \mu - \mu = 0$

   6. Usando X y Y calcule la varianza de la suma de medias muestrales: $V(\bar{X} + \bar{Y})$

      Solución $V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = \frac{\sigma^2}{n} - \frac{\sigma^2}{m}$

2. **15 - Teoría** Explique los siguientes puntos:

   1. Describe la diferencia entre un estimador y un estadístico y como se relacionan

      Solución: Un estadistico está dado en función de una muestra, mientras que un estimador es un estadístico que busca aproximar a un parámetro poblacional. Todo estimador es un estadístico, pero no todo estadístico es un estimador.

   2. Describe la diferencia entre un parámetro y un estimador y como se relacionan

      Solución: Parámetro es una característica o medida de la población y un estimador es un estadístico formado con la muestra. El estimador busca aproximar al parámetro poblacional.

   3. Explique el proceso de inferencia estadística y de un ejemplo de inferencia

      Solución: La inferencia estadística es el proceso por el cuál usamos solo la información de una muestra y buscamos extrapolarla a conclusiones de la población. Por ejemplo, si seleccionamos una muestra 100 de fumadores del ITAM, y buscamos extrapolar la conclusión a toda la población del ITAM.

   4. Justifique la razón por la cual nos interesa conocer la distribución de los estadísticos

      Solución: Por que al ser una aproximación, y al ser una función de variables aleatorias, tienen una variación inherente. Esta variación es importante para saber donde y con que probabilidad está localizado el parámetro poblacional.

   5. Explique las diferencias entre distribución muestral exacta y aproximada

      Solución La distribución muestral exacta construye todas las posibles muestras y calcula el valor del estadístico para cada una de ellas, mientras que la aproximada solo utiliza una muestra y teoremas para construir la distribución.

3. **10 - Verdadero/Falso** Justifique si las siguientes afirmaciones son verdaderas o falsas y justifique

   1. Si se tiene una población de tamaño 10 y se toman muestras de tamaño 1, la distribución de muestreo con reemplazo y sin reemplazo son iguales.

      Solución: VERDADERO, al tener solo el tamaño de muestra de 1, es como si tuvieramos la población completa en ambos casos

   2. El teorema central del limite se puede utilizar siempre y cuando tengamos muestras de tamaño pequeñas.

      Solución: FALSO, solo se utiliza para tamaños de muestra mayor o igual a 27

   3. Suponga una muestra de tamaño 40. Para encontrar la distribución de $\bar{X}$, donde cada X_i se distribuye normal, es necesario utilizar el TCL.

      Solución: FALSO No es necesario, al distribuirse normal entonces por propiedades de normales sabemos la distribución exacta

   4. La importancia del Teorema Central del Límite radica en encontrar la distribución de estadísticos solamente cuando las distribuciones son normales, exponenciales y poisson.

      Solución: FALSO se puede ocupar para cualquier distribución

Seccion B: Preguntas a desarrollar (55 pts)

1. (20 pts) Suponga que una variable aleatoria puede tomar los valores {1,3,5}. Considere muestras de tamaño 2 con reemplazo.

   1. Calcule $E(X)$ y $V(X)$

      Solución

      $E(X) = (1+3+5)/3 = 3$ $V(X) = E(X^2) - E(X)^2 = \sum(X_i^2 p(X_i)) - 9 = (1+9+25)/3 - *(9) = 2{,}6667$

   2. Obtenga distribución de muestreo de $\overline{X}$, $E(\overline{X})$ y $V(\overline{X})$

(X_1, X_2)	Probabilidad	$\bar{(X)}$	S^2
1,1	1/9	1	0
1,3	2/9	2	2
1,5	2/9	3	8
3,3	1/9	3	0
3,5	2/9	4	2
5,5	1/9	5	0

$\overline{X}$	$p(\bar{X})$
1	1/9
2	2/9
3	3/9
4	2/9
5	1/9

$E(\overline{X}) = 1(1/9) + 2(2/9) + 3(3/9) + 4(2/9) + 5(1/9) = 27/9 = 3$

$V(\overline{X}) = E(\overline{X}^2) - E(\overline{X})^2 = \sum(\overline{X}_i^2 p(X_i)) - 3^2 = (1(1/9) + 4(2/9) + 9(3/9) + 16(2/9) + 25(1/9) - (9) = 1{,}3333$

   3. Obtenga distribución de muestreo de S^2, $E(S^2)$ y $V(S^2)$

S^2	$p(S^2)$
0	3/9
2	4/9
8	2/9

$E(S^2) = 0(3/9) + 2(4/9) + 8(2/9) = 2{,}6667$

$V(S^2) = E({S^2}^2) - E(S^2)^2 = \sum({S_i^2}^2 p(S_i^2)) - 2{,}6667^2 = 0(3/9) + 4(4/9) + 64(2/9) - 2{,}6667^2 = 8{,}8888$

   4. Calcule la probabilidad que S^2 tome valores entre 1 y 3 **La probabilidad es 4/9**

   5. ¿Las distribuciones de S^2 y $\overline{X}$ son distribuciones exactas? **Ambas son distribuciones exactas, estamos construyendo la distribución de muestreo con todas las posibles muestras.**

2. (15 pts) Suponga que se tienen 5 vehiculos V_1, V_2, V_3, V_4, V_5. Los vehículos V_3 y V_4 requieren reparación.

   1. Obtenga la distribución de muestreo de la proporción de vehículos $\hat{p}$ que necesitan reparación, si se toman muestras de dos vehiculos sin reemplazo

X	$p(X)$
0 (no requiere)	3/5
1 (si requiere)	2/5

(X_1, X_2)	$p((X_1, X_2))$	$\hat{p}$
0,0	3/5*2/4 = 6/20	0
0,1	3/5*2/4 = 6/20	0.5
1,0	2/5*3/4 = 6/20	0.5
1,1	2/5*1/4 = 2/20	1

$\hat{p}$	$p(\hat{p})$
0	6/20
0.5	12/20
1	2/20

2. Calcule el valor esperado de la distribución de muestreo y la varianza de la distribución de muestreo $E(\hat{p})$ y $V(\hat{p})$

$E(\hat{p}) = 0(6/20) + 0{,}5(12/20) + 1(2/20) = 0{,}4$

$V(\hat{p}) = E(\hat{p}^2) - E(\hat{p})^2 = \sum(\hat{p}^2 p(\hat{p})) - 0{,}4^2 = 0(6/20) + 0{,}25(12/20) + 1(2/20) - 0{,}4^2 = 0{,}09$

3. (15 pts) En la torre de control de un aeropuerto se contaron cuantos aviones despegaban durante 40 periodos de una hora seleccionados al azar durante un mes. Supongase que la distribución del número de aviones (X) que despegan por hora es normal con $\mu = 50$ y desviación estándar $\sigma = 7$.

   1. ¿Cuál es la probabilidad de que la media muestral para n = 40 periodos de una hora sea mayor a 55?

      Solución: $p(\bar{X} > 55) = p(\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} > \frac{(55-50)}{\sqrt{49/40}}) = P(Z > 4{,}5175) \approx 0$

   2. Suponga que n = 5, ¿Cuál es la probabilidad que $\bar{X}$ sea mayor que 55?

      Solución: $p(\bar{X} > 55) = p(\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} > \frac{(55-50)}{\sqrt{49/5}}) = P(Z > 1{,}5972) \approx 0{,}05511$

   3. ¿La distribución presentada para n = 40 es una distribución exacta o aproximada? ¿Para n = 5 es una distribución exacta o aproximada?

      Solución: Para ambas es una distribucion exacta, ya que como tal no ocupamos TCL ya que utilizamos propiedades de las normales para la distribución

   4. ¿Cuál es la probabilidad que el número total de aviones para un periodo de 4 horas sea mayor que 180?

      Solución: $p(\bar{X} > 45) = p(\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} > \frac{(45-50)}{\sqrt{49/4}}) = P(Z > -1{,}4286) \approx 0{,}9234$

4. (15 pts) El peso de unas computadoras que se distribuye de forma normal con media 10 kgs y desviación estándar de 3.

   1. ¿Cuál es la probabilidad que un producto seleccionado al azar pese más de 12 kg?

      Solución: $n = 1$, $p(\bar{X} > 12) = p(\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} > \frac{(12-10)}{\sqrt{9/1}}) = P(Z > 0{,}6667) \approx 0{,}2525$

   2. ¿Si se toma una muestra de 9 productos, ¿Cuál es la probabilidad que la media de la muestra sea menor a 10?

      Solución: $n = 9$, $p(\bar{X} < 10) = p(\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < \frac{(10-10)}{\sqrt{9/9}}) = P(Z < 0) = 0{,}5$