Propuesta Final - Estadística I

ITAM, Primavera 2022 21/05/2022

Instrucciones

Poisson y geométrica

1. (5 pts) El Centro de Aprendizaje Redacción y Lenguas estudió el número de errores ortográficos que cometen los estudiantes del ITAM al escribir un reporte de una cuartilla (X). Encontraron que la media de errores por cuartilla es de 3.

    1. ¿Qué distribución se puede emplear para modelar los errores (X)?. Determine su media y su varianza
       $X \sim Pois(3)$,
       $E(X) = 3$,
       $V(X) = 3$
    2. Determine ¿Cuál es la probabilidad que un estudiante cometa al menos 4 errores en la cuartilla? ¿Cuál es la probabilidad que un estudiante cometa exactamente 4 errores en la cuartilla?
       $P(X >= 4) = 1 - P(X <= 3) = 1 - 0{,}647 = 0{,}353$
       $P(X = 4) = P(X <= 4) - P(X <= 3) = 0{,}815 - 0{,}647 = 0{,}168$
    3. La calificación de la cuartilla se puede modelar como la función $C(X) = 100 - \frac{X^2}{6}$. ¿Cuál es la calificación esperada de esta cuartilla?
       $E(X^2) = V(X) + E(X)^2 = 3 + 9 = 12$
       $E(C) = E(100 - \frac{X^2}{6}) = 100 - \frac{E(X^2)}{6} = 100 - \frac{12}{6} = 98$
    4. Se ha revisado una proporción de la cuartilla y se sabe que hay al menos 2 errores. ¿Cuál es la probabilidad que se tengan al menos 3 errores en la cuartilla completa?
       $P(X >= 3|X >= 2) = \frac{P(X>=3, X>=2)}{P(X>=2)} = \frac{P(X>=3)}{P(X>=2)} = \frac{1-P(X<=2)}{1-P(X<=1)} = \frac{0{,}577}{0{,}801} = 0{,}7203$

2. (5 pts) El número de llamadas que recibe el conmutador del ITAM por hora se puede modelar como una variable Poisson (X) con media de 15. Debido a condiciones del conmutador se conoce que a lo máximo atienda 20 llamadas por hora y deja las demás sin atender.

    1. ¿Cuál es la probabilidad que se dejen llamadas sin atender?
       $P(X > 20) = 1 - P(X <= 20) = 1 - 0{,}917 = 0{,}083$
    2. El ITAM incurre en un costo por llamada (tanto por las atendidas como por las no atendidas). El costo se puede modelar como: $C(X) = 100 + 10X$. Encuentre el Valor Esperado y la varianza del costo.
       $E(C) = E(100 + 10X) = 100 + 10E(X) = 100 + 10(15) = 250$
       $V(C) = V(100 + 10X) = 100V(X) = 100(15) = 150$

3. Se sabe que se han recibido al menos 10 llamadas en una hora, ¿Cual es la probilidad de que se tengan más de 20 llamadas en esa hora?
   $P(X > 20|X >= 10) = \frac{P(X>20,X>=10)}{P(X>=10)} = \frac{P(X>20)}{P(X>=10)} = \frac{1-P(X<=20)}{1-P(X<=9)} = \frac{1-0{,}917}{1-0{,}07} = \frac{0{,}083}{0{,}93} = 0{,}0892$
4. Basándose en el inciso anterior, cual es el valor esperado

3. (9 pts) Determine si las siguientes afirmaciones son verdaderas o falsas. Justifique
   1. La media de un conjunto de datos no considera el valor máximo y minimo de los datos para su cálculo
   2. La moda puede tener más de tres valores
   3. Es imposible que la moda sea igual a la mediana y a la media sean iguales
   4. El coeficiente de variación es considerado un parámetro de centralidad
   5. El IQR es considerado un parámetro de dispersión
   6. El primer cuartil es menor o igual al mínimo de los datos
   7. La mediana debe ser igual al menos a uno de los valores observados
   8. La moda debe ser igual al menos a uno de los valores observados
   9. La media debe ser igual al menos a uno de los valores observados
4. (6 pts) Determine el **tamaño** del espacio muestral (cuantos elementos tiene, sin enumerarlos) para cada uno de los experimentos:
   1. Se lanzan 10 veces una moneda de dos caras
   2. Una urna contiene cinco bolas rojas, seis blancas, una azul y tres amarilla. Una segunda contiene cinco pelotas blancas y tres azules. Una bola es seleccionada de cada urna.
   3. Las urnas del ejercicio anterior son mezcladas en una urna y se extraen dos de ellas con reemplazo

Seccion B: Preguntas a desarrollar (65 pts)

1. (10 pts) Sean A y B dos eventos tal que $P(A) = 0{,}5$ y $P(B) = 0{,}3$. Calcule las siguientes probababilidades: $P(A \cap B)$, $P(A \cup B)$, $P(A|B)$ y $P(A \cup B^c)$ si:
   1. A y B son eventos mutuamente excluyentes
   2. A y B son eventos independientes
2. (15 pts) Una persona compra 10 boletos de una rifa. En total hay 60 boletos y hay 10 boletos con premios iguales.
   1. ¿Cuál es la probabilidad de que esta persona gane un premio?
   2. ¿Cuál es la probabilidad de que esta persona gane al menos un premio?
3. (15 pts) Dos contratos de construcción son asignados de manera aleatoria a tres compañías, A, B y C. Cada compañía puede manejar ninguno, uno o los dos contratos y cada contrato es asignado a una compañia solamente. ¿Cuál es la probabilidad de que:
   1. Todos los contratos vayan a compañias diferentes

2. Entre la compañia A y B tengan todos los contratos, es decir que la compañia C no tenga contratos?

4. (15 pts) Se lanzan 3 monedas que no son honestas, es decir sea $P(A)$ la probabilidad de aguila, la moneda A tiene una $P(A) = 2/3$, la moneda B una $P(A) = 1/2$ y la moneda C una $P(A) = 2/5$.

   1. (5 pts) Escriba el espacio muestral del experimento aleatorio
   2. (5 pts) Obtenga la probabilidad de cada resultado del espacio muestral
   3. (5 pts) ¿Cuál es la probabilidad de observar al menos dos **soles**?
   4. (5 pts) ¿Cuál es la probabilidad que el último lanzamiento sea águila?

5. (10 pts) Un mecánico desea realizar pruebas con un nuevo sistema de frenos para automóviles. El menciona que la probabilidad de falla es de 0.09, mientras que su jefe tiene la idea de que la probabilidad de falla es de 0.02. Los dos están de acuerdo en que la probabilidad de que exista un choque con otro auto dado que fallaron los frenos es de 0.65, por otra parte, si no fallan, la probabilidad de un choque con otro auto es de solo 0.05. Se realiza una prueba y el auto cocha con otro automoviol:

   1. De acuerdo con el mecánico, ¿Cuál es la probabilidad de que haya fallado los frenos dado el accidente?
   2. De acuerdo con el mecánico, ¿Cuál es la probabilidad de que haya fallado los frenos dado el accidente?

Seccion C: Análisis Exploratorio de Datos (15 pts)

Resuelve una de las siguientes dos preguntas:

1. (15 pts) En una fábrica de textiles se aplicó una prueba para medir como aumentaban el número de productos dañados de acuerdo al tiempo promedio que tardaban en producirse. Se cree que entre más rápido realicen los textiles, más productos dañados tendrán. Para esto se tomó la **muestra** de 15 dias de producción con los respectivos tiempos promedio y el número de productos dañados:

Dia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Tiempo	25	20	40	28	39	38	10	30	50	22	27	33	34	40	45
Productos Dañados	4	5	2	4	3	3	10	3	2	5	4	3	3	2	2

   1. (5 pts) Construya un diagrama de dispersión con la información de la tabla
   2. (5 pts) ¿Qué concluye del diagrama?
   3. (5 pts) Proponga una medida que indique si existe una asociación lineal entre el tiempo y el número de productos dañados

2. (15 pts) Se tienen los siguientes datos que representan los ingresos semanales (X) de una **muestra** de 15 personas en una empresa manufacturera de automóviles: 1000, 1200, 1300, 1500, 1500, 1600, 1700, 1700, 1700, 1800, 2000, 2200, 2200, 2300, 3300. Adicional, se tiene que $\sum X_i = 27000$ y $\sum X_i^2 = 52,960,000$

   1. (5 pts) Construya un diagrama de tallo y hoja, donde los tallos indiquen miles y las hojas cientos.

2. (5 pts) Calcule la media, mediana, varianza y coeficiente de variación del ingreso semanal proporcionado
3. (5 pts) Construya un diagrama de caja y brazos (boxplot) para los datos observados