

Segundo examen parcial

Estadística para Ciencia de datos

Alumno 1: _____

Alumno 2: _____

Conteste las siguientes preguntas, y para todos los casos escriba o copie una imagen con las expresiones a calcular, los comandos ejecutados y sus resultados.

1. Sabemos que las manzanas Golden pesan en promedio 200g, y su distribución de probabilidad, sigue una distribución normal con una desviación estándar de 30g. Por otra parte, sabemos que las naranjas pesan en promedio 150 g, y su distribución de probabilidad, también sigue una distribución normal con una desviación estándar de 45 g.

- Cree un conjunto de datos manzanas que representen 500 manzanas elegidas al azar, siguiendo la distribución mencionada.
- Seleccione todas las manzanas que pesen entre 170 y 230 g. Calcule el porcentaje del total que representan las manzanas seleccionadas.
- Cree un conjunto de datos naranjas que representen 300 naranjas elegidas al azar, siguiendo la distribución mencionada.
- Seleccione todas las naranjas que pesen entre 105 y 195 g. Calcule el porcentaje que representan las manzanas seleccionadas.

Explique el resultado de los pasos 2 y 4 usando las características de la distribución normal.

2. Se sabe que la duración de las conexiones a un sitio de ventas de calzado tiene una distribución normal con una media de 30 minutos y una desviación estándar de 14 minutos. Debido a un rediseño de la pagina web se espera que los clientes pasen el triple de tiempo en el sitio y se ha agregado una consulta de uso de cookies al inicio de la sesión que toma 3 minutos contestarla. Los clientes no pueden ingresar al sitio si no contestan la consulta.

- Cree una muestra aleatoria de 400 datos que representen las conexiones al sitio actualmente. Luego transforme los datos para que representen como cambiará la muestra con la nueva página web.
- Calcule la media, mediana, rango de datos, IQR y sd, de ambas muestras y compárelas.
- Explique cada uno de los cambios observados.
- Si la compañía recibe 5 centavos por el despliegue de un anuncio por minuto en su página. Y el sitio recibe en promedio 50 visitas al día, calcule cuanto ingresa en promedio actualmente por la publicación de dos anuncios en la pagina web y en cuanto incrementará el monto del ingreso con la nueva página web.
- Grafique la distribución de probabilidad de la muestra original y de la muestra transformada. Explique la diferencia entre las gráficas.

3. Utilizando los datos del ejercicio uno de manzanas y naranjas, ayude a un granjero a determinar si una manzana de 240 gr es mejor que naranja de 190 g. Para comparar dos datos que siguen diferentes distribuciones normales, en lugar de comparar los valores originales, se comparan sus respectivos z-score.

- Realice los cálculos pertinentes para hacer la comparación y explique porque un z score mayor representa una mejor fruta en este caso. Puede apoyarse en la regla empírica de la distribución normal.
- Calcule cuanto deben pesar una manzana y una naranja para encontrarse en el 5% de frutas más pesadas de sus respectivas distribuciones.

4.- En la compañía de servicio técnico La PC express, solucionan fallas técnicas de computadoras, laptops y celulares que no tomen mucho tiempo. Así que las ordenes de servicios se califican como completadas o no completadas, al final del día. Cada técnico tiene una eficacia diferente para lograr reparar los dispositivos electrónicos. El Técnico 1 logra completar exitosamente 45 de 50 reparaciones. El técnico 2 logra repara exitosamente 29 de 35 reparaciones y el técnico 3 logra reparar exitosamente 31 de 40 reparaciones. Al día se le asignan 12 reparaciones a cada técnico. Conteste las siguientes preguntas, y para todos los casos escriba las expresiones a calcular, los comandos ejecutados y sus resultados.

- Grafique, para cada técnico, la probabilidad de acabar exitosamente de 1 a 12 reparaciones al día. Explique la diferencia entre las gráficas.
- Diga cuantos servicios completados exitosamente se espera de cada técnico al final del día.
- Diga de cual técnico se espera la mayor variación en el número de servicios completados exitosamente en un día. Justifique su respuesta
- Encuentre que técnico tiene la mayor probabilidad de terminar justamente 10 reparaciones exitosamente en un día. Justifique su respuesta
- Encuentre que técnico tiene la mayor probabilidad de terminar menos de 7 reparaciones al día. Justifique su respuesta.
- Si entra a trabajar un nuevo técnico y se espera que en promedio logre terminar 7 reparaciones exitosamente al día, grafique la probabilidad del nuevo técnico de acabar exitosamente 1 a 12 reparaciones al día. Explique la gráfica.

5.- En un centro de llamadas telefónicas que atiende las quejas por ventas en línea, se ha detectado que en temporada baja durante los meses de septiembre y octubre se reciben un promedio de 20 llamadas al día. Mientras que, durante las vacaciones de verano, en los meses de junio y julio, se reciben 30 llamadas al día. Finalmente, durante el mes de diciembre se reciben 40 llamadas al día. Conteste las siguientes preguntas, y para todos los casos escriba las expresiones a calcular, los comandos ejecutados y sus resultados.

- Grafique la distribución de probabilidad de que lleguen entre 5 y 50 llamadas en cada una de las diferentes temporadas. Explique la diferencia entre las gráficas.
- Encuentre en que temporada es más probable tener 35 llamadas al día. Justifique la respuesta

- Encuentre en que temporada es más probable tener 25 llamadas al día. Justifique la respuesta
- ¿En cuál temporada se tiene más variación en el número de llamadas recibidas al día? Justifique su respuesta.
- En que temporada existe la mayor probabilidad de recibir entre 25 y 35 llamadas al día. Justifique su respuesta.
- En que temporada existe la menor probabilidad de recibir 25 llamadas. Justifique su respuesta.

6.- Utilizando el dataset de worldcup del paquete faraway, cree las siguientes muestras, y calcule los datos solicitados de cada muestra.

- Cree tres muestras aleatorias de 50 jugadores, con remplazo. Calcule la media, desviación estándar, rango y cuartiles, del tiempo dentro de la cancha y tiros de los jugadores seleccionados (time, shots). Explique el comportamiento de los resultados.
- Cree tres muestras sistemáticas de 100 jugadores con intervalo de 10 entre muestras. Calcule la media, desviación estándar, rango y cuartiles del tiempo dentro de la cancha y tiros de los jugadores seleccionados (time, shots). Explique el comportamiento de los resultados.
- Cree tres muestras de 50 jugadores, utilizando el muestreo estratificado, y considerando la información de la columna Position como estratos. Calcule la media, desviación estándar, rango y cuartiles del tiempo dentro de la cancha y tiros de los jugadores seleccionados (time, shots). Explique el comportamiento de los resultados y compárelos con los resultados de las muestras aleatorias de los incisos anteriores.
- Con base a los resultados anteriores, recomiende la mejor forma de realizar un muestreo para analizar el tiempo pasado por los jugadores dentro de la cancha y la cantidad de tiros realizados. mean