



GUIA 1, Probabilidades (IME-276)

(Ingenierías Civiles Varias)

Tema: Técnicas de Exploración de Datos

- 1.- Revisar todos los artículos fabricados que salen de una línea de ensamble con el fin de detectar defectos, es un procedimiento costoso y lento. Un método económico y eficaz para determinar la cantidad de artículos defectuosos implica la selección y examen de una fracción de los artículos por parte de un experto en control de calidad. Se determina el porcentaje de artículos examinados que resultaron ser defectuosos y esta cifra se utiliza para estimar el porcentaje de todos los artículos fabricados en la línea que tienen defectos.
 - a. Identifique Población y muestra.
 - b. Identifique parámetro, estimador y explique la inferencia que se puede hacer en este problema.
- 2.- Los insecticidas aplicados a un cultivo, pueden ocasionar inadvertidamente contaminación del aire. *Environmental Science & Technology* (octubre de 1993) informó sobre residuos tóxicos del insecticida *chlorpyrifos* utilizado en huertos latentes en el valle de San Joaquín, California. Se recolectaron diariamente muestras de aire en uno de los huertos durante el periodo de rociado intensivo (un total de 13 días) y se midió el nivel de residuos tóxicos (ng/m^3) en los especímenes extraídos diariamente.
 - a. Identifique la población de interés para los investigadores.
 - b. Identifique la muestra.
3. Investigadores inventaron un método nuevo de preenfriamiento para preparar verduras en floridita antes de enviarlas al mercado. El sistema utiliza una mezcla de aire y agua diseñada para lograr un enfriamiento eficaz con un menor flujo de agua que el hidrogenfriamiento convencional. A fin de comparar la eficacia de los sistemas, 20 lotes de tomates verdes se dividieron en dos grupos; un grupo se preenfrió con el método nuevo, y el otro, con el método convencional. Se registro el gasto de agua (en galones) requerido para enfriar efectivamente cada lote.
 - a. Identifique la población, la muestra y el problema asociado al experimento.
 - b. ¿ Comó podrían utilizarse los datos muestrales para comparar la eficacia de enfriamiento de los sistemas.?
4. Los Ingenieros eléctricos saben que una corriente neutral elevada en los sistemas de alimentación de los computadores es un problema potencial. A fin de determinar la gravedad del problema, se efectuó un estudio de las corrientes de carga de los sistemas de alimentación de computadores en 146 instalaciones en Estados Unidos (IEEE Transactions on Industry Applications, julio/agosto de 1990). El estudio reveló que al menos el 10% de los sitios tenían valores elevados de la razón entre corriente neutral y carga completa. (Ejer 1.3 pag5; Mendenhall, Probabilidad y Estadística, cuarta edición).
 - a. Identifique la población de interés para los investigadores.
 - b. Identifique la muestra.

- c. Utilice la información de la muestra para hacer una inferencia acerca de la población.
5. Para cada uno de los siguientes situaciones, clasifique las variables según concepto y escala.
- Tiempos de llegadas de 16 ondas sísmicas reflejadas.
 - Tipos de software para computadora utilizado en un sistema de gestión de base de datos.
 - Marcas de calculadoras empleadas por 100 estudiantes de ingeniería de la Universidad de la Frontera.
 - Porcentaje de contenido de agua en maderas aserradas.
 - Kilometraje recorrido por 12 automóviles que utilizan alcohol como combustible.
 - Número de caracteres impresos por minuto por 10 marcas distintas de impresoras.
 - Tipo de defecto en la salida de una línea de producto terminado de automóviles.

6.- Considere la cantidad $\sum_{i=1}^n (x_i - \mathbf{a})^2$. ¿Para qué valor de a esta expresión es mínima?

7.- Sea $Y_i = \mathbf{a} + b x_i$, $i = 1, 2, \dots, n$, donde **a** y **b** son constantes no nulas. Encuentre la relación entre \bar{X} e \bar{Y} y entre S_x^2 y S_y^2 .

8.- Considere la muestra x_1, x_2, \dots, x_n con media muestral \bar{X} y una desviación estándar muestral **S**.
Sea $Z_i = \frac{(x_i - \bar{X})}{S}$, con $i = 1, 2, 3, \dots, n$. ¿Cuáles son los valores de la media y de la desviación estándar muestral de Z_i .

9.- Suponga que se tiene una muestra x_1, x_2, \dots, x_n y que sobre ella se ha calculado \bar{X} y S_n^2 . Después se tiene disponible una observación más $X_{(n+1)}$. Sean $\bar{X}_{(n+1)}$ y $S_{(n+1)}^2$, la media y la varianza muestral respectivamente, si para calcularlas se emplean $(n+1)$ observaciones, entonces:

- Indique cómo calcular $\bar{X}_{(n+1)}$ utilizando \bar{X}_n y $X_{(n+1)}$
- Demuestre que $n S_{(n+1)}^2 = (n-1) S_n^2 + \frac{n}{n+1} (X_{(n+1)} - \bar{X}_n)^2$

10.- Un experimento para investigar el tiempo de duración en horas de un componente electrónico consiste en colocar las partes en una celda de prueba y utilizarlas durante 100 horas bajo condiciones de temperatura elevada (este proceso recibe el nombre de prueba de duración "acelerada"). Se prueban 8 componentes y se obtienen los siguientes tiempos de falla:

75, 63, 100⁺, 36, 51, 45, 80, 90

La observación "100⁺" indica que la unidad continúa funcionando después de 100 horas. ¿Existe alguna medida de tendencia central adecuada a estos datos, la cual pueda calcularse. ¿Cuál es su valor?.

11. Las medias de dos muestras son $\bar{X}_1 = 100$ y $\bar{X}_2 = 110$, respectivamente. A continuación, las muestras se juntaron para formar una muestra más grande, para lo cual debía calcularse la media. ¿Cuál es la media si los tamaños de las muestras son:

- $n_1 = 30$, $n_2 = 70$?
- $n_1 = 80$, $n_2 = 20$?
- $n_1 = 50$, $n_2 = 50$?
- $n_1 = 15$, $n_2 = 15$?

12. A menudo la media aritmética es un valor representativo; pero si se tienen una gran desviación estándar, puede no serlo en absoluto. ¿En cuál de los siguientes casos la media no es representativa?

- Un matrimonio muy deportista, entre ambos trotan un promedio de 6 kilómetros diarios. La señora trota 12 kilómetros diariamente.

- b. El rendimiento de un automóvil en una autopista es de 32 millas.
- c. El costo promedio de reparaciones en un automóvil es de 250.000 pesos.
- d. El estadístico medio tiene 3,46 hijos.
- e. El ingreso percapita por persona de un país.

13. Una agencia estatal para el medio ambiente, prueba dos diferentes métodos para quemar carbón bituminoso para generar electricidad, en conexión con cuatro distintos purificadores para reducir la contaminación del aire.. El interés primordial es la emisión de partículas. Se lleva a cabo cuatro ensayos con cada purificador combinados con cada método de combustión. La emisión de partículas se mide en cada ensayo. Los datos son:

<i>método-A</i>					<i>método-B</i>				
rep\purif	purif-1	purif-2	purif-3	purif-4	rep\purif	purif-1	purif-2	purif-3	purif-4
rep-1	18.9	8.8	23.7	23.6	rep-1	24.3	24.0	9.3	18.4
rep-2	16.1	16.5	15.9	22.1	rep-2	21.1	27.1	12.1	8.6
rep-3	14.7	11.7	16.2	16.7	rep-3	18.0	22.6	15.6	15.1
rep-4	16.9	13.0	18.0	18.9	rep-4	16.2	23.1	12.4	9.9

¿ La emisión de partículas es independiente del purificador y del método ?

- a) Identifique las variables presentes en el problema.
- b) Clasifique las variables del problema según la clasificación de variables Conceptual y por Escala.
- c) Identifique Población y Muestra.
- d) Construya un archivo de datos para esta información en formato STATGRAPHICS.
- e) Utilice las herramientas graficas que conoce para responder al problema.
- f) Proponga una metodología de solución y lleve a cabo utilizando el programa STATGRAPHICS.

- 14.- En la costa de la provincia de Valdivia, se efectuó un estudio con el fin determinar, cual de 4 lugares considerados para instalar balsas de cultivo de Choros produce los mejores resultados, respecto de la longitud medida en milímetros. Los lugares elegidos para la instalación de las balsas fueron Corral, Quehule, Chaihuin y Río Colún. En cada balsa se pusieron 10 encordados.

El día que se instalaron los encordados en las balsas se tomo una muestra aleatoria de cada una (Muestreo Aleatorio Simple) de tamaño 10, y se midio la longitud de los choros en mm (L0), transcurridos 6 meses se mide nuevamente la longitud (L1) utilizando el mismo procedimiento. Los datos se muestran en la siguiente tabla.

	QUEHULE		RI O COLUN		CHAI HUI N		CORRAL	
CHORO	L0	L1	L0	L1	L0	L1	L0	L1
1	5. 2	58. 0	5. 5	63. 7	7. 0	74. 5	8. 0	31. 2
2	6. 3	56. 0	8. 3	56. 7	6. 0	64. 8	7. 0	19. 1
3	10. 5	61. 0	9. 1	54. 7	8. 5	74. 4	11. 0	26. 7
4	9. 7	53. 8	6. 7	60. 2	10. 3	64. 0	5. 0	25. 2
5	7. 0	64. 0	7. 4	46. 1	11. 0	76. 1	6. 5	16. 7
6	9. 0	65. 2	6. 0	64. 5	7. 0	56. 0	6. 9	16. 0
7	6. 0	62. 3	10. 0	59. 9	6. 2	81. 5	4. 5	16. 0
8	6. 3	54. 4	7. 9	64. 2	6. 5	65. 0	8. 3	20. 8
9	5. 9	61. 5	8. 2	56. 7	10. 0	75. 3	9. 0	22. 1
10	8. 6	64. 2	9. 7	58. 7	12. 0	73. 6	10. 2	29. 5

- a) Reconozca las variables presentes en el problema.
- b) Clasifique las variables del problema según la clasificación de variables Conceptual y por Escala.
- c) Identifique Población y Muestra.

- d)** Construya un archivo de datos para esta información en formato STATGRAPHICS.
- e)** Utilice las herramientas graficas que conoce para responder al problema.
- f)** Proponga una metodología de solución y lleve a cabo utilizando el programa STATGRAPHICS.