



INSTITUTO DE ADMINISTRACIÓN PÚBLICA DEL ESTADO DE CHIAPAS, A.C.

MATERIA: ESTADISTICA ADMINISTRATIVA.

ACTIVIDAD No: 3

TEMA: EJERCICIOS DE ESTADÍSTICA DESCRIPTIVA.

NOMBRE DEL ALUMNO: COSME HERNANDEZ LOPEZ

CATEDRÁTICO: DR. ENRIQUE ANTONIO PANIAGUA MOLINA

Tapachula, Chiapas; a 21 de Septiembre del año 2015.

EJERCICIOS DE ESTADÍSTICA DESCRIPTIVA

Ejercicio 1: Si la señora López compra una de las casas anunciadas para su venta en un diario de TGZ, **T** es el evento de que la casa tiene tres o más baños, **U** es el evento de que tiene una chimenea, **V** es el evento de que cuesta más de \$ 100 mil pesos y **W** es el evento de que es nueva.

- Describa (con palabras) cada uno de los siguientes eventos:

$T' = U, V, W$ (una chimenea, cuesta más de cien mil pesos, casa nueva)

$U' = T, V, W$ (Casa de 3 o más baños, cuesta más de cien mil pesos, casa nueva)

$V' = T, U, W$ (Casa de 3 o más baños, una chimenea, casa nueva)

$W = T, U, V$ (casa de 3 o más baños, una chimenea, cuesta más de cien mil pesos)

$T \cap U = \text{Vacío}$ (No tiene nada en común)

$T \cap V = \text{Vacío}$ (No tiene nada en común)

$U' \cap V = \text{Vacío}$ (Cuesta más de cien mil pesos)

$V \cap U = \text{Vacío}$ (Cuesta más de cien mil; casa nueva)

$V' \cap W = \text{Vacío}$ (Casa de 3 o más baños, una chimenea, casa nueva)

$T \cap U = \text{Vacío}$ (Casa de 3 o más baños, una chimenea)

$T \cap V = \text{Vacío}$ (Casa de 3 o más baños, cuesta más de cien mil pesos)

$V \cap W = \text{Vacío}$ (No tiene nada en común)

Ejercicio 2: Un dado está arreglado de manera que cada número impar tiene el doble de probabilidad de ocurrir que un número par. Encuentra $P(B)$, donde B es el evento que un número mayor que 3 ocurra en un solo tiro del dado.

- Espacio muestral $S = \{1, 2, 3, 4, 5, 6\}$
- Sub conjunto B $B = \{4, 5, 6\}$
- Probabilidad

- Si x es la probabilidad que ocurra un número par, $2x$ sería la probabilidad que ocurra un número impar.

- Entonces, encontramos que: $2x + x + 2x + x + 2x + x = 1$
 $S = (1, 2, 3, 4, 5, 6)$

$$S = 2x + x + 2x + x + 2x + x = 1$$

$$9x = 1$$

$$x = 1/9$$

$$2/9 + 1/9 + 2/9 + 1/9 + 2/9 + 1/9 = 1$$

$$9/9 = 1 \quad 1 = 1$$

- Esto se debe al postulado 2

*La $P(B)$ sería: $4/9$

$$B = (4, 5, 6)$$

$$P(B) = 1/9 + 2(1/9) + 1/9 = 4/9$$

$$P(B) = 0.444$$

Ejercicio 3: Calcula la muestra para una población desconocida con un 96% de confianza y 10% error. Para una prevalencia de .5 y .7

FÓRMULA

$$n = \frac{Z_{\alpha}^2 \cdot p \cdot q}{i^2}$$

$$Z^2_{\alpha} = 1.7506$$

$$p = .5$$

$$q = (1 - p) = (1 - 0.7) = 0.3$$

$$i = .10$$

$$n = 76.615 \approx 77$$

Tapachula, Chiapas; a 21 de Septiembre del año 2015.



Donde:

$$Z^2_{\alpha}=1.7506$$

$$P=0.07$$

$$q=(1-p)=(1-.07)=0.3$$

$$i=0.10$$

Sustituyendo en la fórmula

$$n = \frac{(1.7506)^2(.7)(.3)}{(0.10)^2} = n=64.357 = n=64$$

Ejercicio 4: Calcula la muestra para una población de 350,000 familias, con un 99% de confianza y 5% error. Para una prevalencia de .5 y .7

FÓRMULA

$$n = \frac{Z_{\alpha}^2 \cdot N \cdot p \cdot q}{i^2 (N - 1) + Z_{\alpha}^2 \cdot p \cdot q}$$

Para una prevalencia de 0.5

$$Z^2_{\alpha}= 1.7506$$

$$P= .5$$

$$.q= 1-p = (1-.5) = .5$$

$$.i= .05$$

Sustituyendo en la fórmula

$$n = \frac{(2.3263)^2 (350000)(.5)(.5)}{(0.05)^2(350000-1) + (2.3263)^2(.5)(.5)} = \frac{(5.4116)(350000)(.25)}{(0.0025)(349999) + (5.4116)(.25)} =$$

Tapachula, Chiapas; a 21 de Septiembre del año 2015.

$$n = \frac{473515}{(874.9975)+1.3529} = \frac{473515}{876.3504} =$$

$$n = 540.326 \approx 540$$

Para una prevalencia de 0.7

$$Z^2\alpha = 1.7506$$

$$P = .5$$

$$q = 1-p = (1-.7) = .3$$

$$i = .05$$

Sustituyendo en la fórmula

$$n = \frac{(2.3263)^2 (350000)(.7)(.3)}{(.05)^2(350000-1) + (2.3263)^2(.7)(.3)} = \frac{(5.4116)(350000)(.21)}{(0.0025)(349999) + (5.4116)(.21)} =$$

$$n = \frac{397752.60}{(874.9975)+1.364} = \frac{397752.60}{876.1339} =$$

$$n = 453.98 \approx 454$$

Ejercicio 5: De una Población de 1,176 padres de familia de la Cd. de Tuxtla Gutiérrez. Se pretende conocer la aceptación de los programas educativos mediante caricaturas. Se pretende obtener una muestra para saber el número de entrevistas y con ello obtener información estadísticamente confiable. Se asume un error estándar de 1.5% con un nivel de confiabilidad del 90%

Datos:

N=1176

Z=1.2815

p=0.5

q= 0.5

$$n = \frac{Z_{\alpha}^2 \cdot N \cdot p \cdot q}{i^2 (N - 1) + Z_{\alpha}^2 \cdot p \cdot q}$$

i=0.015 Fórmula:

$$n = \frac{(1.2815)^2 (1176)(.5)(.5)}{(0.015)^2 (1176-1) + (1.2815)^2 (.5)(.5)} = \frac{(1.6422)(1176)(.25)}{(0.000225)(1175) + (1.6422)(.25)}$$

$$n = \frac{482.8068}{(0.2643) + 0.2643} = \frac{482.8068}{0.6746}$$

$$n = 715.48 \approx 716$$

Estaturas (X)	f	Ni	$x_i * f_i$	$x_i^2 * f_i$
1.52	1	1	1.52	2.3104
1.54	5	6	7.7	11.858
1.55	4	10	6.2	9.61
1.58	5	15	7.9	12.482
1.6	2	17	3.2	5.12
1.62	4	21	6.48	10.4976
1.64	7	28	11.48	18.8272
1.66	3	31	4.98	8.2668
1.7	5	36	8.5	14.45
1.71	8	44	13.68	23.3928
1.73	6	50	10.38	17.9574
1.74	5	55	8.7	15.138
1.77	3	58	5.31	9.3987
1.8	1	59	1.8	3.24
1.83	1	60	1.83	3.3489
Total	60		99.66	165.8978

Desviación estándar para datos agrupados:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{N}$$

$$\bar{x} = \frac{99.66}{60} = 1.661$$

$$\sigma = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\frac{165.8978}{60} - 1.661^2}$$

$$\sigma = \sqrt{2.764963 - 2.75892}$$

$$\sigma = \sqrt{.006043}$$

$$\sigma = 0.077736$$

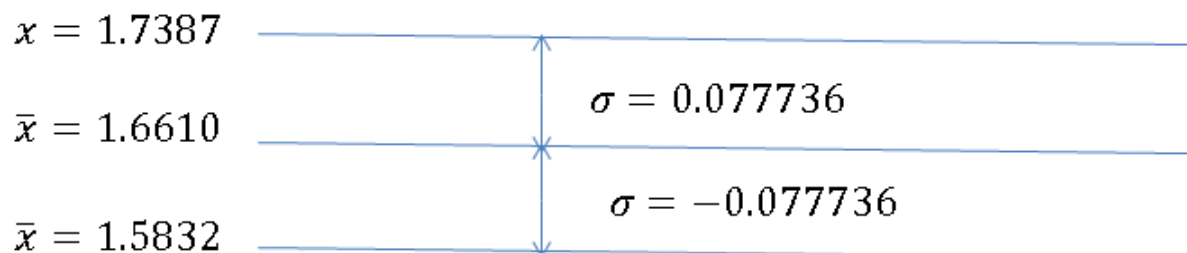
Varianza para datos agrupados:

$$\sigma^2 = \sum_{i=1}^n \frac{x_i^2 f_i}{N} - \bar{x}^2$$

$$\sigma^2 = \frac{165.8978}{60} - 1.661^2$$

$$\sigma^2 = 2.764963 - 2.758921$$

$$\sigma^2 = .006042$$



6.- Con los siguientes datos

Estaturas (x)	f
1.52	1
1.54	5
1.55	4
1.58	5
1.6	2
1.62	4
1.64	7
1.66	3
1.7	5
1.71	8
1.73	6
1.74	5
1.77	3
1.8	1
1.83	1
Total	60

- Son los resultados de preguntarle la estatura a 60 trabajadores del departamento de limpia municipal de SCLC.
- Obtén la media aritmética (para datos agrupados)
- Obtén la desviación estándar y la varianza (para datos agrupados)
- Interpreta los resultados

Estaturas (x)	f	Ni	xi * fi	xi 2 * fi
1.52	1	1	1.52	2.3104
1.54	5	6	7.7	11.858
1.55	4	10	6.2	9.61
1.58	5	15	7.9	12.482
1.6	2	17	3.2	5.12
1.62	4	21	6.48	10.4976
1.64	7	28	11.48	18.8272
1.66	3	31	4.98	8.2668
1.7	5	36	8.5	14.45

1.71	8	44	13.68	23.3928
1.73	6	50	10.38	17.9574
1.74	5	55	8.7	15.138
1.77	3	58	5.31	9.3987
1.8	1	59	1.8	3.24
1.83	1	60	1.83	3.3489
Total	60		99.66	165.8978

Desviación estándar para datos agrupados:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{N} \quad \bar{x} = \frac{99.66}{60} = 1.661$$

$$\sigma = \sqrt{\sum_{i=1}^n \frac{x_i^2 f_i}{N} - \bar{x}^2} \quad \sigma = \sqrt{\frac{165.8978}{60} - 1.661^2}$$

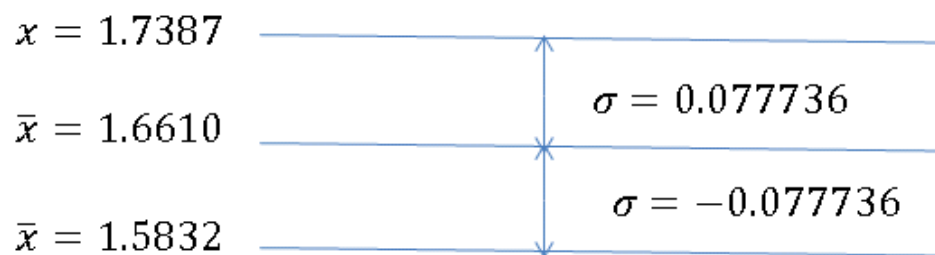
$$\sigma = \sqrt{2.764963 - 2.75892} \quad \sigma = \sqrt{.006043}$$

$$\sigma = 0.077736$$

Varianza para datos agrupados:

$$\sigma^2 = \sum_{i=1}^n \frac{x_i^2}{N} - \bar{x}^2 \quad \sigma^2 = \frac{165.8978}{60} - 1.661^2$$

$$\sigma^2 = 2.764963 - 2.758921 \quad \sigma^2 = .006042$$



ELECCIÓN DE LA MUESTRAS

Las muestras se utilizan por economía de tiempo y recursos. Para seleccionar una muestra, lo primero que hay que hacer es definir la unidad de análisis (personas, organizaciones, periódico, comunidades, situaciones, eventos, etc.). El sobre que o quienes se van a recolectar datos depende del planteamiento del problema a investigar y de los alcances del estudio. Estas acciones nos llevarán al siguiente paso, que consiste en delimitar una población.

Para el proceso cuantitativo la muestra es un subgrupo de la población de interés (sobre el cual se recolectarán datos, y que tiene que definirse o delimitarse de antemano con precisión), este deberá ser representativo de la población. El investigador pretende que los resultados encontrados en la muestra logren generalizarse o extrapolarse a la población.

Una vez que se ha definido cuál será la unidad de análisis, se procede a delimitar la población que va a ser estudiada y sobre la cual se pretende generalizar los resultados. Una población es el conjunto de todos los casos que concuerdan con una serie de especificaciones (Selltitz *et al.*, 1980). Es preferible establecer con claridad las características de la población, con la finalidad de delimitar cuáles serán los parámetros muestrales. La delimitación de las características de la población no solo depende de los objetivos del estudio, sino de otras razones prácticas. Un estudio no será mejor por tener una población más grande; la calidad de un trabajo investigativo estriba en delimitar claramente la población con base en el planteamiento del problema. Las poblaciones deben situarse claramente en torno a sus características de contenido, de lugar y en el tiempo.

Al seleccionar la muestra debemos evitar tres errores que pueden presentarse: 1) no elegir a casos que deberían ser parte de la muestra (participantes que deberían estar no fueron seleccionados), 2) incluir casos que no deberían estar porque no forman parte de la población y 3) seleccionar casos que son verdaderamente inelegibles (Mertens, 2005). El primer paso para evitar tales errores es una adecuada delimitación del universo o población. Los criterios que cada investigador cumpla depende de sus objetivos de estudio, lo importante es establecerlos de manera más específica. Toda investigación debe ser transparente, así como estar sujeta a crítica y réplica.

¿Cómo seleccionar la muestra?

La muestra es, en esencia, un subgrupo de la población. Digamos que es un subconjunto de elementos que pertenecen a ese conjunto definido en sus características al que llamamos población. En realidad, pocas veces es posible medir a toda la población, por lo que obtenemos o seleccionamos una muestra y, desde luego, se pretende que este subconjunto sea un reflejo del conjunto de la población.

Básicamente categorizamos las muestras en dos grandes ramas: las muestras no probabilísticas y las muestras probabilísticas. En las últimas los elementos de la población tienen la misma posibilidad de ser escogidos y se obtienen definiendo las características de la población y el tamaño de la muestra, y por medio de una selección aleatoria o mecánica de las unidades de análisis. En las muestras no probabilísticas, la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación o de quien hace la muestra. Aquí el procedimiento no es mecánico, ni con base en fórmulas de probabilidad, sino que depende del proceso de toma de decisiones de una persona o de un grupo de personas y, desde luego, las muestras seleccionadas obedecen a otros criterios de investigación. Elegir entre una muestra probabilística o una no probabilística depende de los objetivos del estudio, del esquema de investigación y de la contribución que se piensa hacer con ella.

¿Cómo se selecciona una muestra probabilística?

La elección entre la muestra probabilística y la no probabilística se determina con base en el planteamiento del problema, las hipótesis, el diseño de investigación y el alcance de sus contribuciones. Las muestras probabilísticas tienen muchas ventajas, quizá la principal sea que puede medirse el tamaño del error en nuestras predicciones, el principal objetivo en el diseño de una muestra probabilística es reducir al mínimo este error, al que se le llama error estándar (Kish, 1995).

Las muestras probabilísticas son esenciales en los diseños de investigación transeccionales, tanto descriptivo como correlacionales-causales, donde se pretende hacer estimaciones de variables en la población. Estas variables se miden y se analizan con pruebas estadísticas en una muestra, donde se presupone que esta es probabilística y todos los elementos de la población tienen una misma probabilidad de ser elegidos. Las unidades o elementos muestrales tendrán valores muy parecidos a los de la población, de manera que las mediciones en el subconjunto nos darán estimados precisos del conjunto mayor.

Las muestras probabilísticas requieren la determinación del tamaño de la muestra y de un proceso de selección aleatoria que asegure que todos los elementos de la población tengan la misma probabilidad de ser elegido.

El tamaño de una muestra también depende del número de subgrupos que nos interesan en una población, lo óptimo de una muestra depende de cuánto se aproxima su distribución a la distribución de las características de la población, esta aproximación mejora al incrementarse el tamaño de la muestra. La normalidad de las distribución de muestras grandes no obedece a

la normalidad de la distribución de una población, la distribución de diversas variables a veces es normal y en ocasiones está lejos de serlo; sin embargo, la distribución de muestras de cien o más elementos tiende a ser normal y esto sirve para el propósito de hacer estadística inferencial, sobre los valores de una población.

Las muestras no probabilísticas o dirigidas, también llamadas muestras dirigidas, suponen un procedimiento de selección informal. Se utilizan en muchas investigaciones cuantitativas y cualitativas.

En la lectura primero se delimita si en todas las investigaciones se tiene una muestra, cuando se realiza un censo, se deben incluir a todos los que tengan una influencia en el caso objeto del estudio y en caso de las personas se recomienda incluir a todos para evitar las sensaciones de exclusión, las muestras se utilizan por economía de tiempo y recursos.

Dependiendo del planteamiento de la investigación es el enfoque de “que o a quienes” (sujetos, objetos, sucesos, etc...), serán las unidades de análisis, también conviene involucrar a las partes influyentes en la acción-planteamiento, ósea a todos los que tengan injerencia y delimitar en base a parámetros de edad, sexo, raza, etc..., para obtener mejores resultados, nos muestra una tabla que plantea cuestionamientos sobre los errores y soluciones en las unidades de análisis.

Para delimitar una población, después de definir la unidad de análisis, se observa que, el conjunto de casos que concuerdan con las características que requiere la investigación de contenido, geográfica y temporales.

Para seleccionar la muestra o el tipo de muestra se debe de tomar a un sub grupo de la población, esto permite tomar un conjunto que represente a toda la población del estudio, existen básicamente 2 tipos de muestras, las no probabilísticas y las probabilísticas.

Para seleccionar una muestra probabilística se deben tomar en cuenta las necesidades de la investigación para reducir el máximo margen de error, son las muestras ideales en las investigaciones transeccionales, descriptivas y correlacionales-causales, donde se hacen cálculos de las variables en la población, se analizan con pruebas estadísticas, en donde cada elemento de la muestra tiene las mismas posibilidades de ser elegidos, se necesita determinar el tamaño de la muestra y seleccionar los elementos muestrales de manera aleatoria.

Pero para definir el tamaño de la muestra la lectura nos facilita dos fórmulas $n' = s^2/V^2$ y $n = V^2/1 + n'IN$, esta muestra probabilística tiene varias opciones en las cuales encontramos la muestra estratificada y por racimos. Para definir el proceso de selección se necesita definir el número de racimos necesario y para seleccionar los sujetos de cada racimo nos debemos basar en la aleatoriedad, se usan 3 procedimientos de selección, la tómbola y los números aleatorios.

No debemos omitir el marco de la muestra ya que este nos permite identificar de manera física los elementos de la población, la posibilidad de enumerarlos. La forma más común de estos son los listados y estos también se pueden subdividir para tener un mayor control sobre los datos. Otro ejemplo de marco muestral son los archivos ya que ayudan a la identificación y selección de la muestra, claro esto con el debido análisis previo de la viabilidad de los sujetos en las unidades, los mapas son de gran utilidad como marco de referencia en muestras de racimos; los volúmenes son una forma de archivar los compilados de información que contienen a los elementos.

Las muestras no probabilísticas o muestras dirigidas tienen un proceso de selección informal, generalmente son usadas en investigaciones cuantitativas y cualitativas. Selecciona a un grupo de sujetos con la esperanza de que sean representativos de la población base del estudio.

El mejor ejemplo son los muestreos al azar por llamada telefónica; la muestra multietapas o polietapica solo subdivide en estratos y racimos pero la selección de los mismos sigue siendo aleatoria y no representativa.