ESTADISTICA

PROGRAMA:

OBJETIVO: Medir los conocimientos básicos y generales de la estadística y la probabilidad basados en los programas de estudio del nivel bachillerato más representativos.

El alumno comprenderá y aplicará los diferentes procedimientos en la obtención de estadísticos generales.

INTRODUCCIÓN

1.- Conceptos básicos y operaciones elementales en la teoría de conjuntos.

2.- Estadística y probabilidad (definición y objetivos).

3.- Variables (definición, usos y clasificación: aleatoria, continua, discreta).

4.- Población y muestreo (definición y tipos de muestreo

PROBABILIDAD

1.- Conceptos básicos de Probabilidad (experimento, espacio muestral, punto muestral y evento).

2.- Enfoques de la Probabilidad (subjetivo, frecuencial, clásico).

3.- Fenómenos deterministas y fenómenos aleatorios.

MEDIDAS Y ANÁLISIS

1.- Métodos de recolección de datos.

2.- Distribuciones de frecuencias: acumuladas y relativas acumuladas (clases, intervalos, límites, límites reales, marca).

INTRODUCCIÓN.

Ciencia matemática que se basa en la recopilación sistemática de datos numéricos relativos a fenómenos económicos, científicos, culturales demográficos y otros para, analizarlos y llegar a conclusiones que permitan tomar decisiones.

Muchos consideran que el fundador de la ciencia estadística fue Gottfried Achenwail , quien le dio el nombre actual (1748) los estudios de población que se iniciaron en esa época se complementaron luego al desarrollarse la teoría matemática de las probabilidades que se aplico al actuariado y a otros aspectos de las ciencias sociales, (natalidad, mortalidad, enfermedades, criminalidad, educación, etc)a las ciencias naturales, (física, meteorología)y finalmente a la administración industrial, a la física atómica, etc.

La sistematización y el perfeccionamiento de los métodos matemáticos es en realidad cosa del siglo XX pero ya la estadística ha entrado en todos los campos de la actividad humana como instrumento indispensable para las decisiones que deben tomarse en una sociedad de estructura cada vez más compleja.

ESTADISTICA = Sinónimo de datos numéricos

DEFINICION DE LA ESTADISTICA: Es el campo del estudio referente a la organización y resumen de datos y la extracción de inferencias acerca de un grupo de datos, cuando solo se observan una parte de datos.

SE DIVIDE EN: Estadística descriptiva y en Inferencia estadística.

Estadística descriptiva Conceptos y métodos necesarios para organizar y resumir datos.

Inferencia estadística.- tomar decisiones acerca de un volumen de datos, examinando solo una pequeña parte.

EL MÉTODO ESTADÍSTICO CONSTA DE LAS SIGUIENTES ETAPAS.

1. PLANIFICACIÓN DEL ESTUDIO.

2. RECOLECCIÓN DE LA INFORMACIÓN.

3. ELABORACIÓN DE LA INFORMACIÓN.

4. ANÁLISIS E INTERPRETACIÓN.

1.- PLANIFICACION DEL ESTUDIO.

Es esencial no solo para calcular el tiempo que durara la investigación, el personal que se requiera y el presupuesto necesario sino con el fin de que se realice con metas perfectamente definidas, evitando improvisaciones durante el desarrollo de la misma.

LOS PASOS DE LA PLANIFICACIÓN SON:

a) Planteamiento del problema.
b) Búsqueda y evaluación de la información existente.
c) Formulación de una hipótesis.
d) Verificación de la hipótesis.
e) Análisis e interpretación de los resultados obtenidos.

a) ***Planteamiento del problema.***
En este paso es en donde contestamos las interrogantes de Quien, Porque, Cual, Donde, Cuando, Como, Cuantos, Que.

b***) Búsqueda y evaluación de la información existente.***
La evaluación global del material estudiado nos permite estudiar, qué tipo de material es el que nos va a servir y cual tendremos que desechar.

c) Formulación de una hipótesis.
La formulación de la hipótesis no es más que una explicación provisional de los hechos que no es más que anticiparse aquellos fenómenos que pueden ocurrir.

d***) Verificación de la hipótesis.***
En la verificación de la hipótesis constituye la investigación propiamente dicha.

d) Análisis e interpretación de los resultados obtenidos.
Ejecutado el estudio se considera si fue realizado conforme se había planificado y con resultados a la vista se concluirá.

2.- RECOLECCION DE LA INFORMACION.

Aunque en la mayoría de las investigaciones la información debe recibirse directamente de su fuente de origen, en muchas ocasiones suele aprovecharse los datos previamente recolectados por otros individuos, en el primer caso se considera que es una fuente primaria en el otro una fuente secundaria

LOS PRINCIPALES MÉTODOS DE RECOLECCIÓN:

a) ***LAS ENCUESTAS***.- se obtienen mediante el interrogatorio.
b) **LOS EXPERIMENTOS.-** los datos no existen y es necesario provocar su aparición.
c) **LOS CENSOS.-** información recolectada periódicamente, se diferencia con la encuesta en la naturaleza de la información que se busca, en el censo habitualmente se buscan datos generales y en la encuesta datos sobre un tema concreto.
d) **SISTEMA DE REGISTROS.-** se recolecta la información continuamente a medida que se va produciendo.
e) **UNIVERSO Y MUESTRAS**.- se entiende por universo o población la totalidad de individuos o elementos en los cuales puede presentarse determinada característica de ser estudiada. Una muestra a su vez es una parte o grupo de universo.
f) **FORMULARIOS.-** es toda forma impresa destinada a la recolección de datos tal como las historias clínicas, las fichas epidemiológicas, o tarjetas especiales, debe permitir obtener dos clases de datos los administrativos o de identificación, dependiendo de lo que se trate.

3.- ELABORACION DE LA INFORMACION.

Una vez recolectada la información es necesario revisarla cuidadosamente y luego resumirla y presentarla convenientemente antes de que sea posible analizarla.

4. ANÁLISIS E INTERPRETACIÓN.

Posterior a la elaboración de la información se requiere analizar los datos que nos permitan tomar decisiones.

1.- CONCEPTOS BÁSICOS Y OPERACIONES ELEMENTALES EN LA TEORÍA DE CONJUNTOS.

DEFINICIÓN DE TEORÍA DE CONJUNTOS: Herramienta útil en matemáticas.

CONJUNTO: es la colección de objetos definidos y distintos.

ELEMENTOS O MIEMBROS DEL CONJUNTO: son los objetos que constituyen un conjunto.

Para designar un conjunto se usaran letras mayúsculas, puede describirse

- Enumerando todos los elementos de un conjunto.
- Describiendo el tipo de elementos de que esta comprendido.

CARACTERÍSTICA A LOS CONJUNTOS.

1. ***Conjunto unitario***: conjunto compuesto de un solo elemento.
2. ***Conjunto vació o nulo:*** cuando no consta de elementos.
3. ***Conjunto universal: c***onjunto de elementos por los que se tiene interés
4. Si un conjunto tiene elementos y se relaciona con otro se dice que este es subconjunto del otro.
5. Por definición el conjunto nulo es subconjunto de cualquier otro conjunto.
6. Dos conjuntos son iguales si y solo si contienen los mismo elementos.

TÉCNICAS DE CONTEO, PERMUTACIONES Y COMBINACIONES

Es un arreglo ordenado de objetos, de cuantas maneras de pueden cambiar y utilizar.

Ejemplo si tenemos A, B, C, D, cuales serian las combinaciones posibles

A: B, C, D
B: A, C, D
C: A, B, D
D: A, B, C

COMBINACIÓN: es un arreglo de objetos sin tomar en cuenta el orden

En el ejemplo anterior, las permutaciones de cuatro cosas tomadas dos a la vez consistieron de los 12 arreglos siguientes,

AB	BC	CD
AC	BD	DA
AD	CA	DB
BA	CB	DC

Nótese que en la lista anterior hay varias parejas de arreglos que son semejante excepto por el orden en que aparecen las letras.

AB	AC	AD	BC	BD	CD
BA	CA	DA	CB	DB	DC

Si el orden no cuenta diríamos que se tienen 12 combinaciones

Pero si no tendríamos 6 permutaciones.

2.- ESTADÍSTICA Y PROBABILIDAD (DEFINICIÓN Y OBJETIVOS).

CONCEPTOS BÁSICOS DE PROBABILIDAD.

Es la base para la inferencia estadística, este concepto no es extraño en las ciencias de la salud, para dar un pronóstico sobre alguna enfermedad, cumplimiento de una cita

Podemos medir la probabilidad de ocurrencia de un evento por medio de un número entre cero y uno, mientras más probable sea más próximo estará al uno, mientras más improbable más cerca estará al cero.

Desde punto de vista de la probabilidad:

- objetivo. esta a su vez puede ser, clásica o a priori y de frecuencia relativa o a posteriori.
- Subjetivo.

DEFINICIÓN: si un evento puede ocurrir en n maneras mutuamente exclusivas e igualmente probables y si m de estas poseen una característica E, la probabilidad de la ocurrencia de E es igual a m/N.

Formula P(E) = m/N

Ejemplo juegos al azar, dados, baraja.

TEORÍA DE LA PROBABILIDAD, PROPIEDADES.

- Dado algún proceso (experimento) con n resultados mutuamente exclusivos (llamados eventos) E1,E2....eN a la probabilidad de cualquier evento, Ei, se le asigna un número no negativos es decir deben tener una probabilidad mayor o igual a cero.
- La suma de las probabilidades de todos los resultados mutuamente exclusivos es igual a 1.
- Considérense dos eventos mutuamente exclusivos cualesquiera, la probabilidad de la ocurrencia es igual a la suma de probabilidades individuales.

Evento exclusivo: cuando no puede ocurrir simultáneamente a otro

Dos eventos son independientes cuando la probabilidad de aparición de uno de ellos no depende no se modifica por la aparición del otro (ejem. Efectos secundarios de un fármaco en dos pacientes).
P (A, B) = P (A Y B) = P(A) XP (B)

PROBABILIDAD CONDICIONADA.

La probabilidad de A condicionada a la ocurrencia de B se define considerando únicamente los casos en los que aparece B, y viendo en cuántos de estos casos ocurre el suceso de A. es la probabilidad de que suceda A, una vez que ya ha sucedido B.

Probabilidad condicional.-cuando se calculan las probabilidades con un subconjunto de un conjunto universal.

Probabilidad marginal:- cuando se ignoran uno o más criterios de clasificación al calcular una probabilidad.

Distribución de probabilidad de variable discreta: Es una tabla, grafica, formula o cualquier otro medio que se use para especificar todos los valores posibles de una variable aleatoria discreta junto con sus probabilidades respectivas.

Ejemplo: una enfermera tiene a su cargo 50 familias, construyamos la distribución de probabilidad X , el número de niños por familiar para esta

x	Frecuencia de la ocurrencia de x	P(X=x)
0	1	1/50
1	4	4/50
2	6	6/50
3	4	4/50
4	9	9/50
5	10	10/50
6	7	7/50
7	4	7/50
8	2	8/50
9	2	2/50
10	1	1/

DISTRIBUCION DE PROBABILIDAD.

Una vez que hemos escogido la muestra y recogido los datos, el siguiente paso es inferir las propiedades de la población a partir de la muestra. el instrumento conceptual que permitirá esta generalización es un modelo de la población, es decir, una representación simbólica de su comportamiento.

Encontramos tres leyes teórica que tienen la particularidad de ser seguidas por una inmensa mayoría de los fenómenos biológicos habituales:

a) Distribución Binomial.

b) Distribución de Poisson.

c) Distribución normal o de Gauss.

DISTRIBUCIÓN BINOMIAL.

Es una ley teórica de aplicación siempre que se conozca de entrada, la probabilidad de aparición de un fenómeno biológico (p). el resultado es dicotómico y puede ser evaluado como “éxito o fracaso”. La variable de interés es el numero de éxitos alcanzados en W ”n” pruebas.

La distribución binomial constituye siempre un diagrama de barras discontinuo = se aplica a variables discretas.

DISTRIBUCIÓN DE POISSON.

Constituye un caso particular de la ley binominal para aquellas situaciones en que la probabilidad de aparición de un fenómeno es muy pequeña. Es por tanto, la ley adecuada para los “sucesos raros”.

Cuando $p<0,1$ y $n \times p <5$.

DISTRIBUCIÓN NORMAL O DE GAUSS.

Es seguida por una inmensa cantidad de variables biológicas cuyas medidas se agrupan alrededor de un valor central, y que presentan una frecuencia cada vez menor a medida que se alejan de dicho valor medio.

CARACTERISTICAS:

- Corresponde a variables cuantitativas continuas.
- Se caracteriza por dos medidas: media y desviación típica.
- Es unimodal.
- Es simétrica alrededor de la media, por tanto, media, mediana y moda coinciden.
- Tiene forma acampanada, sin un pico excesivo.
- Va desde infinito a finito (asintonica al eje de abscisas).
- El área bajo la curva = 1.

El 50 % de las observaciones se encuentran por debajo de la media aritmética y el 50 % por encima el 68% de las observaciones se encuentran dentro del intervalo de la media.

3.- VARIABLES.

DEFINICION.- Al observar una característica se encuentra que toma valores diferentes en personas, lugares o cosas se le denomina variable.

SE CLASIFICA DE LA SIGUIENTE MANERA.

Variable cuantitativa. Es toda característica que puede medirse.

Variable cualitativa.-Característica que no puede ser medida.

Variable cualitativa nominal.- se identifica con un nombre.

Variable cualitativa ordinal.- Se identifica con un dato subjetivo.

Variable aleatoria.- cuando el resultado obtenido es resultado de factores fortuitos.

Variable aleatoria discreta.- Cuando se expresa en números enteros indivisibles.

Variable aleatoria continua.-Cuando la unidad admite ser fraccionada

.Ejemplo de las características de las variables.

No.	**Cualitativa**		**Cuantitativa**	
	Nominal	**Ordinal**	**Discreta**	**Continua**
1	**Masculino**	**Mejor**	**Pulso 80 x min**	**Peso 85.2 kgr.**
2	**Femenino**	**Delicado**	**“ 90 x min**	**72.5 kgr.**
3	**Femenino**	**Grave**	**“ 140 x min**	**63.2 kgr.**
4	**Masculino**	**Estable**	**“ 85 x min**	**68.7 kgr.**

POBLACIÓN Y MUESTREO

***Población de elementos*.-** mayor colección de elementos por los cuales se tiene cierto interés en un instante particular.

***Población de valores*.-** mayor colección de valores de una variable aleatoria por las cuales se tiene cierto interés en un instante particular.

Es finita.- cuando se tiene número fijo de valores y si tiene sucesión sin fin de valores es **infinita.**

***Muestra*.-** se define como parte de una población

TIPOS DE MUESTREO.

1.- PRUEBA DE t DE STUDENT.

La prueba t es el procedimiento estadístico más utilizado en la literatura médica, comúnmente conocida como la prueba de t de student, se utiliza cuando se quiere responder respecto a la diferencia de la media de dos grupos independientes. Se deberá tener consideración sobre las condiciones necesarias para llevar a cabo la prueba de t: 1) para que las consideraciones de cada grupo sean reales deberá cerciorarse que las muestras de cada grupo tengan una distribución normal, ya que si esto no se cumple el valor de P que se obtiene es menor a lo normal, por lo que el rechazo de la hipótesis nula sería más fácil, y decir que existe diferencia cuando en realidad no la hay. El tamaño de la muestra debe ser mayor a 302. 2) Deberá cerciorarse que las desviaciones estándar o las varianzas de ambos grupos sean iguales, conocidas como varianzas homogéneas, esto es porque la hipótesis nula establece que las dos medias son iguales es una forma de decir que las observaciones de dos grupos se realizan en una misma población, por lo tanto la hipótesis nula será verdadera cuando las deviaciones estándar en ambos grupos sean iguales, si se encuentra que las desviaciones estándar en ambos grupos sean iguales, si se encuentra que las desviaciones estándar no son iguales, no deberá hacerse caso siempre y cuando el tamaño de la muestra sea igual para ambos grupos (prueba t robusta) 3) la otra condición necesaria es que las muestras correspondan a grupos independientes, es decir que el valor de las observaciones de un grupo no muestran información sobre las observaciones de otro grupo, cuando este requisito no se cumple se deberá hacer otro tipo de análisis conocido como t para muestras dependientes (t pareada).

Es una cuestión importante determinar la normalidad de la distribución de la muestra considerando que la desviación estándar en las oblaciones es igual y que las observaciones son independientes.

2.- PRUEBA DE CHI CUADRADA.

La prueba de Chi cuadrada(X^2), se puede utilizar para comparar frecuencias y proporciones en dos o más grupos independientes, es una prueba bastante versátil y muy utilizada, los datos son sencillos de usar y fáciles de presentar en tablas. No se debe utilizar la prueba cuando los numerales en la tabla de 2 x 2, sean menores a 2, o cuando el 20 % de las celdas tenga menos de cinco elementos dentro de una de las celdas, en caso de ser así lo que se usaría con mayor certeza es la prueba exacta de Fisher. La tabla de 2 x 2 es la mejor forma de ordenar los datos y poder llevar a cabo el análisis de estos. La prueba de X^2 es tan versátil que se puede utilizar sin importar de qué modo haya sido planteada la hipótesis, utilizaremos un estudio hipotético en el cual se emplearan 200 pacientes, 100 pacientes en un grupo de tratamiento y 100 pacientes en un grupo de control de los cuales 50 pacientes respondieron de manera positiva y 150 pacientes lo hicieron de manera negativa. Para realizar el análisis de X^2 se deberán seguir los siguientes pasos:

1.- Ordenar los datos en una tabla de 2 x 2, y una vez ordenados los datos obtenidos se deberán calcular las frecuencias esperadas para ese grupo de pacientes.

MEDIDAS DE TENDENCIA CENTRAL

Un arreglo ordenado deja mucho que desear con respecto a la obtención de un resumen informativo por lo que se utilizan las medidas descriptivas.

- Cuando se calcula a partir de los datos de una muestra recibe el nombre de **ESTADÍSTICA.**
- Cuando se calcula a partir de los datos de una población recibe el nombre de **PARÁMETRO.**

Medidas de tendencia central más comunes.

- **MEDIA**
- **MEDIANA**
- **MODA**

MEDIA ARITMÉTICA.- Es la más conocida, se tiene en mente cuando se requiere sacar promedio.

Se obtiene sumando todos los valores de una población o muestra y dividiendo entre el número de valores.

$$\text{Edad media} = \frac{10+22+...+70+49}{100} = \frac{3920}{100} = 39.20$$

$$U = \frac{\sum_{i=1}^{N} Xi}{N} \qquad \bar{X} = \sum_{i=1}^{n} Xi$$

PROPIEDADES.

- **Unicidad.**
- **Simplicidad.**
- **Como cada uno y todos los valores intervienen en el cálculo esta es afectada por cada valor.**

MEDIANA. Es un conjunto finito de valores es el valor que divide al conjunto en dos partes iguales tales que el numero de valores es igual a o mayores y menores que la mediana.

Si el número es impar será el valor que está en medio.

Si el número es par no solo se tiene una observación en medio sino dos.

De la tabla anterior. 50,51 36,37 mediana es (36 +37)/ 2 = 36.5

Ejemplo.- 10,54, 21,33 y 53

MODA. En un conjunto de valores es aquel valor que ocurre con mas frecuencia, si todos los valores son diferentes no existe moda.

EJEMPLO: 20,21, 20, 20, 34, 22, 24, 27, 27, 29

EJERCICIO:

NIVELES DE GLUCOSA EN SANGRE EXTRAÍDA A 10 NIÑOS EN AYUNAS

Numero	valor	número	valor
1	56	6	65
2	62	7	65
3	63	8	68
4	65	9	70
5	65	10	72

a) la media
b) la mediana
c) la moda.

MEDIDAS DE DISPERSIÓN.

LA DISPERSIÓN.- se refiere a la variedad que exhiben los valores, si todos son los mismos no existe dispersión,

Puede ser pequeña si los valores están muy próximos entre sí, o grande si están ampliamente desparramados.

RECORRIDO.- es la diferencia entre el valor menor y el mayor en un conjunto de observaciones.

R = X L _ X s

La variancia. Se calcula restando la media de cada uno de los valores se eleven al cuadrado las diferencias y se suman

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Observación de la población, observación de la muestra.

Numero	Numero	Valor
X12	X1	10
X20	X2	54
X36	X3	21
X62	X4	33
X98	X5	53

Ecuación $\frac{10+54+21+33+53}{5} = \frac{171}{5} = 34.2$

$$S^2 = \frac{(10-34.2)^2 + (54-34.2)^2 + (21-34.2)^2 + (33-34.2)^2 + (53-34.2)^2}{4}$$

$$= \frac{1506.8}{4} = 376.7$$

Para obtener una medida de dispersión en las unidades originales simplemente se toma la raíz cuadrada de la variancia. El resultado se llama **VARIACIÓN ESTÁNDAR**

COEFICIENTE DE VARIACIÓN.- es una medida de variación relativa en lugar de una variación absoluta

$$C.V. = \frac{S}{\bar{X}} (100)$$

	Muestra 1	Muestra 2
Edad	25 años	11 años
Peso medio	72.5Kg	40 Kg.
Desviación media	5 Kg.	5 Kg.

DATOS AGRUPADOS.- como se ha visto un conjunto de observaciones puede hacerse más compresible y adquirir y adquirir un mayor significado en un arreglo ordenado si se agrupan los datos puede proporcionar una mayor síntesis,

Para agrupar se selecciona un conjunto de intervalos contiguos que no se traslapen, que pueda colocarse en uno y solo uno de los intervalos. Estos intervalos se conocen como **INTERVALOS DE CLASE.**

Se utiliza la ecuación de Sturges la formula da k = 1 + 3.322

K representa el no de intervalos de clase.

$$W = \frac{R}{K}$$

EJERCICIO

Partículas de materia suspendidas en Muestra de aire tomada en 57 ciudades

Cd	valor	Cd	valor	Cd	valor	Cd	valor
1	68	16	65	31	12	46	22
2	63	17	43	32	32	47	43
3	42	18	25	33	49	48	27
4	27	19	74	34	38	49	49
5	30	20	51	35	42	50	28
6	36	21	36	36	27	51	23
7	28	22	42	37	31	52	19
8	32	23	28	38	50	53	46
9	79	24	31	39	38	54	30
10	27	25	28	40	21	55	43
11	22	26	25	41	16	56	49
12	23	27	45	42	24	57	12
13	24	28	12	43	69		
14	25	29	57	44	47		
15	44	30	51	45	23		

$K = 1 + 3.322 (\log 57)$

$= 1 + 3.322 (1.7559)$
$= 7$

$$\frac{R}{K} = \frac{79 - 12}{7} = \frac{67}{7} = 9.6$$

EN EL EJEMPLO LOS INTERVALOS SE PUEDEN DAR DE 10 EN 10

10-19

20-29

30-39

40-49

50-59

60-69

70-79

En relación a la tabla anterior se agrupan contando los que pueden caer dentro

Tabla final

Intervalos de clase	frecuencia	frecuencia Acumulada	frecuencia relativa	frecuencia relativa acumulada
10 a 19	5	5	.0877	.0877
20 a 29	19	24	.3333	.4210
30 a 39	10	34	.1754	.5964
40 a 49	13	47	.2281	.8245
50 a 59	4	51	.0702	.8947
60 a 69	4	55	.0702	.9649
70 a 79	2	57	.0351	1.0000
	57		10000	

HISTOGRAMA una distribución de frecuencias puede presentarse gráficamente donde los valores los valores de la variable constituyen el eje longitudinal, y el eje vertical tiene escala de frecuencia.

MEDIDAS DE TENDENCIA CENTRAL CALCULADAS A PARTIR DE DATOS AGRUPADOS.

Una vez que se han agrupado los datos es posible que se desee calcular algunas de las medidas descriptivas, no siempre el investigador tiene accesos a estos datos ya que tiene que sacarlos.

Tabla para calcular la media a partir de los datos agrupados

Intervalo de clase	punto medio de la clase	frecuencia de la clase	
	mi	fi	mifi
10 a 19	14.5	5	72.5
20 a 29	24.5	19	465.5
30 a 39	34.5	10	345.0
40 a 49	44.5	13	578.5
50 a 59	54.5	4	218.0
60 a 69	64.5	4	258.0
70 a 79	74.5	2	149.0
Total	57		2086.5