UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

"LA ESTADISTICA EN LA ANTROPOLOGIA FISICA"

T E S I S

Que para obtener el Título de:

A C T U A R I O

P r e s e n t a

HECTOR BENJAMIN CISNEROS REYES

México, D. F. 1994

UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

CIUDAD UNIVERSITARIA

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS
División de Estudios Profesionales
Exp. Núm. 55

M. EN C. VIRGINIA ABRIN BATULE
Jefe de la División de Estudios Profesionales
Universidad Nacional Autónoma de México.
P r e s e n t e .

Por medio de la presente, nos permitimos informar a Usted, que habiendo revisado el trabajo de tesis que realiz 6 el pasante HECTOR BENJAMIN CISNEROS REYES con número de cuenta 7415171-8 con el título: "LA ESTADISTICA EN LA ANTROPOLOGIA FISICA"

Consideramos que reune los méritos necesarios para que pueda continuar el trámite de su Examen Profesional para obtener el título de ACTUARIO.

GRADO NOMBRE Y APELLIDOS COMPLETOS	FIRMA
M. EN C. GUILLERMO GOMEZ ALCARAZ Director de Tesis	
M. EN C. ALBERTO CONTRERAS CRISTAN	
MAT. ANA LUISA SOLIS GONZALEZ-COSIO	
ACT. MARIA PILAR ALONSO REYES Suplente	
M. EN C. MARIA GUADALUPE ELENA IBARGUENGOITIA GONZALEZ Suplente	

Ciudad Universitaria, D.F., a 2 de septiembre de 199 4

LA ESTADISTICA EN LA ANTROPOLOGIA FISICA

Héctor Benjamín Cisneros Reyes.

Noviembre de 1994

Dedico éste trabajo a mi queridísima esposa Guadalupe y a mis hijos Raymundo (q.e.p.d.), Mauri, Héctor y Mariana por su incondicional apoyo y permanente cariño.

A mi mamá y a mis hermanos ejemplo de constancia y tenacidad.

A todos los Gonzalitos, gracias a mi tía Lupe, Lourdes, Ale, Gaby, Luis, Irene, Miguel Angel, Lupita, Irma, Sergio, ... uff.

A todos y cada uno de los miembros de mi familia.

INDICE

INTRODUCCION

En la investigación moderna, principalmente en aquellas áreas de la ciencia en que se pretende obtener un conocimiento para establecer las tendencias generales o interrelaciones fundamentales que presentan sus características medibles (numérica o en categorías) es extremadamente útil el uso de la Estadística. Esta ciencia busca más objetividad en la evaluación de los resultados numéricos obtenidos, aunque nunca podrá hacerse totalmente objetiva esta valoración.

Cuando se hace uso de la estadística como apoyo para lograr los objetivos planteados por los científicos de cualquier área (por ejemplo la Antropología Física) debe de haber una adaptación mutua entre los criterios, métodos, modelos y supuestos de la estadística, con aquellos de la(s) ciencia(s) que fundamentan los elementos de estudio, por lo que Estadística se conviérte en una metodología de investigación.

La Estadística al igual que el llamado método científico, provee de guías sobre el proceso de investigación. Estas guías conjuntas de la Estadística y del método científico deben considerarse en todas las fases de la investigación y no únicamente en la fase final, una vez concluido el trabajo "de campo" o de "laboratorio" y obtenidos los datos. Esta última consideración errónea, es decir, que la Estadística "se usa" sólo en la última fase (en el análisis), es muy común y frecuentemente produce investigaciones con deficiencias metodológicas fuertes. Estas deficiencias afectan fundametalmente la validez externa, la interna o ambas; introducen factores de confusión entre mediciones o elementos de estudio que son difíciles o imposibles de "corregir" con los métodos Estadísticos. Además, frecuentemente, por desconocimiento y excesiva confianza en la Estadística, ésta se aplica o se interpreta erróneamente.

Es deber de un buen investigador conocer los aspectos de la metodología y de la Estadística que más impacto tienen en su investigación. Para el caso de la filosofía y metodología conocer los alcances y limitaciones del método científico; y para la Estadística un conocimiento tipo "caja negra", esto consiste en saber como usar una técnica Estadística, en qué casos se puede usar y cómo interpretar los resultados, sin conocer a fondo los fundamentos de ella. Esta última consideración es el motivo del

presente trabajo.

La investigación es un proceso de creación de conocimientos acerca de la estructura, el funcionamiento o el cambio de algún aspecto de la realidad, este proceso es complejo y requiere una cuidadosa consideración de sus objetivos, con base en el marco de referencia, de su diseño, y de acuerdo con los recursos de que se dispone.

Este trabajo da un panorama general de como la Estadística se inserta en el proceso de investigación social, y de su aplicación en la Antropología Física en particular, por tanto, es necesario establecer los objetivos y mecánismos de ésta ciencia.

En primer lugar se da la definición de Antropología y después la de Antropología Física.

Se entiende por antropología:

> " La ciencia del hombre o más bien la ciencia comparativa del hombre que trata de sus diferencias y causas de las mismas, en lo referente a estructura, función y otras manifestaciones de la humanidad, según el tiempo, variedad, lugar y condición".

Con esa amplitud y a medida que se han acumulado datos al respecto, la Antropología ha ido dividiendose en distintas ramas, llegando a construir ciencias independientes como son: Arqueología, Etnología y Etnografía, Linguistica, Antropología Física, Paleoantropología, etc.

La Antropología Física se define como:

> "historia natural del género Homo"

y más concretamente se tienen otras definiciones de la misma:

> "Ciencia que tiene por objeto el estudio de la humanidad considerada como un todo, en sus partes y en sus relaciones con el resto de la naturaleza".
>
> "Ciencia que estudia las variaciones humanas"
>
> "Estudio comparativo del cuerpo humano y de sus funciones inseparables"
>
> "Tratado de las causas y caminos de la evolución humana, transmisión y clasificación, efectos y

tendencias en las diferencias funcionales y orgánicas"

"Ciencia que se dedica al estudio de la variabilidad y evolución orgánica del ser humano y sus determinantes culturales y comportamentales."

Es decir, la Antropología física aborda lo referente al agrupamiento cronológico, racial, social, y aún patológico de los núcleos humanos.

Así, la Antropología estudia el origen del hombre, su herencia, crecimiento, Somatología, Biotipología, Craneología, Osteología, Paleoantropología, etc.; sin olvidar la relación que hay entre los grupos humanos y el medio ambiente que lo rodea.

Puede resumirse que las definiciones coinciden en que hay que estudiar al hombre como parte de un todo y no como una pieza aislada por lo que una (por supuesto no la única) herramienta útil para lograr el cometido de esta ciencia es por medio de la Estadística.

Para establecer la manera en que la Estadística se incorpora al proceso de investigación en general y a la investigación Antropológica en particular es necesario considerar los diferentes diseños de investigación que hay en las Ciencias Sociales.

I. TIPOS DE INVESTIGACION EN LAS CIENCIAS SOCIALES

I.1 TIPOS DE INVESTIGACIÓN.

La investigación social es un proceso en el que se vinculan diferentes niveles de abstracción, se cumplen determinados principios metodológicos y se cubren diversas etapas lógicamente articuladas, apoyado dicho proceso en teorías, métodos, técnicas e instrumentos adecuados y precisos para poder alcanzar un conocimiento objetivo (en lo posible). Por lo que es necesario dar una guía general de este.

Para elaborar un protocolo de investigación, entre otras cosas, se deben adecuar los recursos con los objetivos y donde ambos tienen que determinar los métodos. Por lo que el diseño (estructura y estrategia) de la investigación se puede clasificar en:

1) De acuerdo al periodo en que se capta la información el estudio es:

   a) Retrospectivo.
      La información fue captada en un periodo anterior con fines, generalmente diferentes, a los de la investigación actual.

   b) Retrospectivo parcial.
      Parte de la información es tomada de otras investigaciones anteriormente realizadas y completada con información captada al inicio de la nueva investigación.

   c) Prospectivo.
      Toda la información es recogida por el investigador de acuerdo con sus criterios y objetivos planteados previamente en la planeación de la investigación.

2) De acuerdo con la evolución del fenómeno, se puede clasificar en:

a) Longitudinal.
Se sigue a los elementos en estudio, durante un periodo de tiempo y se mide a las características involucradas en varias ocasiones.

b) Transversal.
Se mide una sola vez la o las variables involucradas y se hace el análisis con esta información, sin pretender evaluar la evolución de esas unidades.

3) De acuerdo con la comparación de las poblaciones, el estudio es:

a) Descriptivo.
Estudio que sólo cuenta con una población la cual se pretende describir en función de un grupo de variables y respecto de la cual no existen hipótesis centrales.

b) Comparativo
Se tienen dos o más poblaciones y donde se quieren comparar algunas variable para contrastar una o varias hipótesis centrales.

4) De acuerdo con la interferencia del investigador en el fenómeno que se analiza, el estudio es:

a) Observacional.
Sólo se observa al fenómeno sin intervenir en los diferentes factores que puedan cambiar el resultado del fenómeno estudiado.

b) Experimental.
Se controlan los factores que puedan influir en los resultados de los fenómenos.

De los cuatro diseños generales de investigación se forman 8 diferentes protocolos.

I.2. PROTOCOLOS DE INVESTIGACION.

Como se mencionó anteriormente combinando las 4 formas básicas de investigación se tienen 8 protocolos, para implemetar cada uno de estos, deberán de hacerse diferentes formatos.

PROTOCOLO 1

NOMBRE: ESTUDIO DESCRIPTIVO

a) Estudio Descriptivo Retrospectivo (observacional, retrospectivo, transversal y descriptivo).

b) Estudio Descriptivo Prospectivo (observacional, prospectivo, transversal y descriptivo).

OBJETIVO

En este protocolo se estudia una población y únicamente se pretende describir la situación de ésta en un momento determinado, de acuerdo con algunas variables. No se tiene una hipótesis central, aunque el estudio puede servir para sugerir hipótesis que se contrasten después.

FORMATO PARA EL ESTUDIO DESCRIPTIVO.

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
   c) Proceso de captación de la información.
      (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.

7) Referencias.

PROTOCOLO 2

NOMBRE: ESTUDIO COMPARATIVO.

a) Estudio Comparativo Retrospectivo (observacional, retrospectivo, transversal y comparativo).
b) Estudio Comparativo Prospectivo (observacional, prospectivo, transversal y comparativo).

OBJETIVO

Se tienen dos o más poblaciones y se pretende comparar algunas variables en una ocasión única, con objeto de contrastar una o varias hipótesis.

FORMATO PARA EL ESTUDIO COMPARATIVO

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.
   d) Hipótesis.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
      vi) Comparabilidad de las muestras de poblaciones.
   c) Proceso de captación de la información.
      (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.
7) Referencias.

PROTOCOLO 3

NOMBRE: REVISION DE CASOS

Revision de casos (observacional, retrospectivo, longitudinal y descriptivo).

OBJETIVO:

Se pretende concocer la evolución del fenómeno estudiado en el pasado en relación con ciertas variables que se miden en diversas ocasiones en los sujetos (unidades) correspondientes a una sola población. La información que se utiliza se recogió con fines ajenos a esta investigación y se encuentran en documentos, expedientes, archivos, etc.

FORMATO PARA REVISION DE CASOS

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
      iii) Criterios de eliminación.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
   c) Proceso de captación de la información.
      (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.
7) Referencias.

PROTOCOLO 4

NOMBRE: CASOS Y CONTROLES

Casos y controles (retrospectivo, longitudinal, observacional, comparativo, de efecto causa).

OBJETIVO:

El propósito del estudio es conocer sí el o los grupos de casos tienen una mayor incidencia del posible factor (causa) que el o los grupos control. Para este tipo de estudios se forman uno o más grupos de elementos que presentan un determinado resultado y uno o más grupos que no presentan dichos efectos. Se recaba información acerca de la exposición pasada a uno o más factores considerados como riesgo, determinandose la proporción o grado de exposición a este factor entre los grupos de casos y de controles. La información se obtiene de expedientes, archivos, etc.

FORMATO PARA CASOS Y CONTROLES.

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.
   d) Hipótesis.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
      iii) criterios de eliminación.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
      vi) Igualación de atributos.
   c) Proceso de captación de la información.
      (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.
7) Referencias.

PROTOCOLO 5

NOMBRE: PERSPECTIVA HISTORICA

Perspectiva histórica (retrospectivo, longitudinal, observacional, comparativo de causa a efecto).

OBJETIVO:

El propósito de este estudio es conocer qué grupo, ya sea el expuesto al factor considerado como riesgo o el no expuesto, tuvo mayor incidencia en cuanto al efecto; se forman muestras con sujetos que estuvieron expuestos o no a uno o más factores considerados como riesgo. Después se recaba información acerca del efecto.

FORMATO PARA PERSPECTIVA HISTORICA

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.
   d) Hipótesis.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
      iii) criterios de eliminación.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
      vi) Igualación de atributos
   c) Proceso de captación de la información.
      (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.
7) Referencias.

PROTOCOLO 6

NOMBRE: UNA COHORTE

Una Cohorte (prospectivo, longitudinal, observacional y descriptivo).

OBJETIVO:

Se considera cohorte al grupo de sujetos que se estudian longitudinal y prospectivamente, y que tuvieron alguna experiencia en común o que comparten alguna característica especifica (raza, nivel socio-económico, edad, etc.) En este estudio se cuenta con una sola población, de la cual se hace labor de seguimiento para conocer su evolución.

FORMATO PARA UNA COHORTE.

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.
   d) Hipótesis.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
      iii) criterios de eliminación.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
   c) Proceso de captación de la información.
      (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.
7) Referencias.

PROTOCOLO 7

NOMBRE: VARIAS COHORTES

Varias Cohortes (prospectivo, longitudinal, observacional y comparativo).

OBJETIVO:

Al igual que en el protocolo de una cohorte se estudian a dos o más grupos de sujetos que tuvieron alguna experiencia en común. En este estudio se cuenta con dos o más poblaciones, de la cual se hace labor de seguimiento para conocer su evolución.

FORMATO PARA VARIAS COHORTES.

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.
   d) Hipótesis.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
      iii) criterios de eliminación.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
      vi) Igualación de atributos
   c) Proceso de captación de la información.
      (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.
7) Referencias.

PROTOCOLO 8

NOMBRE: EXPERIMENTO

Experimento (prospectivo, longitudinal, experimental, comparativo).

OBJETIVO:

Un experimento se caracteriza por la elección de las variantes del factor causal que se quieren investigar. Las unidades experimentales se asignan aleatoriamente a dichas variantes. El experimento en ciencias sociales es un estudio en el cual se ponen a prueba dos o más métodos, tratamientos o programas con fines diagnósticos.

FORMATO PARA EXPERIMENTO.

1) Definición del problema.
   a) Título.
   b) Antecedentes.
   c) Objetivos.
   d) Hipótesis.

2) Definición de la población objetivo.
   a) Características generales.
      i) criterios de inclusión.
      ii) criterios de exclusión.
      iii) criterios de eliminación.
   b) Ubicación espacio-temporal.

3) Diseño estadístico
   a) Especificación de variables y escalas de medición.
   b) Diseño Muestral.
      i) Cuándo muestrear.
      ii) Qué muestrear
      iii) Cómo muestrear
      iv) Métodos de muestreo.
      v) Determinación del tamaño de muestra.
      vi) Comparabilidad de las muestras de poblaciones.
      vii) Diseño de las maniobras o tratamientos (factor causal)
   c) Proceso de captación de la información. (Selección de instrumentos para recopilar la información)
   d) Análisis e interpretación de la información.

4) Recursos.
5) Logística.
6) Etica del estudio y procedimientos peligrosos.
7) Referencias.

Como puede observarse cada uno de los formatos presenta varios de los siguientes elementos:

1) Diseño estadístico de muestreo.
2) Especificación de variables y escalas de medición.
3) Proceso de captación de la información.
4) Análisis e interpretación de la información.
5) Cálculo del tamaño de la muestra.
6) Comparación de poblaciones.
7) Hipótesis.

Todos estos puntos de una manera u otra son propósitos de la Estadística. Que según diferentes autores es:

> " Técnica matemática que se utiliza para la colección, resumen, análisis e interpretación de información ".

Otra forma de clasificarla es de acuerdo con las variables que analiza de manera simultánea, si sólo lo hace con una o dos variables se le conoce como Estadística Univariada, si lo hace con más de dos variables entonces se le llama Estadística Multivariada. La Estadística Descriptiva y/o Exploratoria es empleada para descripción y exploración de información. Si su fín es la generalización, entonces se utilizan técnicas inferenciales, estas pueden ser Paramétricas, No Paramétricas o Bayesianas.

II. DISEÑO ESTADÍSTICO

II.1. ESPECIFICACION DE VARIABLES Y ESCALAS DE MEDICION.

Las variables son las características medibles en las unidades de estudio; deben seleccionarse en relación con los objetivos planteados. En ocasiones las variables resultan complejas tanto en su medición como en su definición, por lo que se recomienda aclararlas mediante definiciones operacionales y estableciendo el tipo de escalas que se utilizarán para medirlas. Debe de precisarse la utilidad de cada una de las variables antes de captarlas, de preferencia determinando el tipo de análisis estadístico en que intervendrán. La información captada puede ser numérica o no numérica, a la primera se le conoce como cuantitativa y a la segunda como cualitativa.

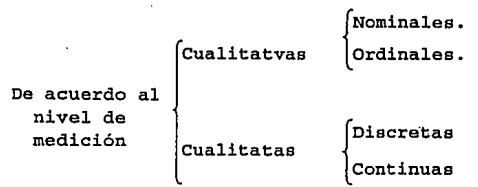
El tipo de información proporciona una idea acerca de la técnica de análisis estadístico a emplearse (no a la inversa).

La información cualitativa puede provenir de diferentes fuentes, una de ellas son las encuestas (más adelante se hace una revisión) en las que puede haber preguntas abiertas y/o cerradas, cuando son abiertas la información es analizada por medio de técnicas de análisis de contenido, cuando las respuestas de preguntas cerradas son muy concretas, se pueden categorizar, habiendo métodos de análisis estadístico para este tipo de información. Los datos cuantitativos pueden provenir de conteos o de mediciones directas de las características de interés.

Para fines estadísticos las variables se clasifican en:

i) CUALITATIVA, que se subdividen en Nominales y Ordinales

ii) CUANTITATIVAS (NUMERICAS) este tipo de variables a su vez se subdividen en variables discretas y contínuas, las discretas son el resultado de conteos y las contínuas de alguna medición.



La forma de incorporar la información necesaria, depende de los objetivos de la investigación, por lo que se debe de establecer la escala de "medición" de cada una de las características de interés, por esto, existe una clasificación de acuerdo al tipo de la escala de medición; por lo que de una manera análoga se tienen medidas de tipo cuantitativo (de intervalo y de razón) y de tipo cualitativo (nominales clasificatorias categóricas). Y se ordenan de acuerdo a su nivel de propiedades:

NOMINALES O CLASIFICATORIAS O CATEGORICAS
ORDINALES
DE INTERVALO
RAZON.

Las escalas NOMINALES o CLASIFICATORIAS son aquellas donde los datos caen dentro de una y sólo una categoría previamente establecida, y que la unión de estas categorías da el total de las posibles clasificaciones, por ejemplo: en un estudio de poblaciones se puede hacer una clasificación de hombres, mujeres. niños, adultos, estudiantes, no estudiantes, trabajadores de la construcción, de servicios, financieros, de educación etc. Dichas clasificaciones se hacen de acuerdo al objetivo de la investigación, y puede ser una clasificación natural o artificial.

Matemáticamente esta partición no tiene orden. Sólo es una clasificación por lo que no se pueden hacer operaciones aritméticas con dichos datos, unicamente conteos de los mismos.

En este tipo de información existen los llamados DICOTOMICOS, que son aquellos en donde el conjunto de datos sólo tiene dos categorias (masculino, femenino o si se prefiere "0" ó "1", tiene o no tiene tal característica) para estos datos existe una gran variedad de técnicas Estadísticas.

En las escalas ORDINALES se cumplen con las condiciones anteriores, pero además existe un orden (matemático), es decir, se pueden clasificar en "mayor que", "igual a", "menor que". No se pueden hacer operaciones aritméticas con ellos.

Las escalas de INTERVALO cumplen con lo anterior, pero además se pueden hacer operaciones matemáticas, sumas, restas, diviciones, multiplicaciones, etc. en este tipo de escalas no se tiene un "cero" absoluto por lo que no se pueden hacer comparaciones de intervalo.

Los datos de RAZON Y PROPORCION son aquellos, que cumpliendo con las condiciones anteriores, tienen un "0" absoluto; es decir, siempre se podrá hacer referencia a este punto en cualquier momento. Con esta escala se pueden hacer cualquier tipo de operaciones.

De todo lo anterior se deduce que una técnica que se pueda aplicar a un grupo de datos de razón se puede aplicar a todos los tipos anteriores, pero no a la inversa.

Otras clasificaciones según el punto de vista:

- Metodológico
  - Dependientes.
  - Independientes
- Teórico Explicativo
  - Estímulo
  - Respuesta
  - Intermediaria
- Manipulativo
  - Activas
  - Asignadas
  - Atribuidas
- De sistemas
  - Endógenas
  - Exógenas

Por esto, se tienen que hacer otras consideraciones, para decidir sobre que método de análisis estadístico utilizar son:

a) El tipo de variables.
b) La independendencia o dependencia entre variables.
c) El número de variables que de manera simultánea se pretende analizar, en el caso de una sola variable se

utilizan métodos univariados, cuando hay más de una variable en el análisis, se utilizan métodos multivariados.

d) El tipo y objetivo de la investigación (descripción, inferencia, comparación, experimento, etc.).

¿Cuantas varibles conviene medir?

El principio de parsimonia indica:

"Conviene medir tantas como sea necesario y el menor número posible"

II.2. TECNICAS PARA LA RECOLECCION DE INFORMACION.

II.2.1. INTRODUCCION

Una vez determinadas las características que sirven para conseguir los objetivos se debe de determinar la forma de recolección de información, esta forma está determinada por cada uno de los protocolos de investigación; es necesario definir a la población en estudio (población objetivo), delimitarla, cuando es posible, en espacio y en tiempo y establecer las fuentes principales de generación de información.

Existen varias fuentes (no necesariamente ajenas) principales para la generación de información, entre otras se pueden señalar las siguientes:

i) Archivos, reportes.
ii) Encuestas.
iii) Entrevistas.
iv) Exámenes.
v) Observación directa.

Los archivos y reportes pueden ser médicos, administrativos, económicos, etc. proporcionan información captada en momentos pasados con fines diferentes a los objetivos de la investigación que en ese momento se esté llevando a cabo, por lo que es necesario tener cuidado de la calidad e integridad de los mismos.

Las encuestas son cuestionarios que se aplican a la persona en estudio, dependiendo de lo que se quiera saber, estos cuestionarios pueden ser de diversas formas y tipos.

Las entrevistas son instrumentos de recolección de información que permiten detectar el lenguaje del cuerpo y de las emociones. Los exámenes son un tipo particular de cuestionarios.

En Antropología Física La observación directa de las características de interés se da muy a menudo, principalmente en

lugares y sitios de los cuales sólo hay vestigios de la población que ahí existió, en los entierros de zonas arqueológicas, en ruinas de conventos y otros lugares, por señalar algunos ejemplos.

II.2.2. ENCUESTAS

La finalidad de una encuesta es obtener información para satisfacer una necesidad definida. La necesidad de recopilar datos surge en todo campo de la actividad humana por lo que es importante establecer un esquema general.

ESQUEMA GENERAL DE ACTIVIDADES.

1) DISEÑO
   - DISEÑO CONCEPTUAL
   - DISEÑO ESTADISTICO
   - DISEÑO ADMINISTRATIVO

2) LEVANTAMIENTO
   - ORGANIZACION DEL TRABAJO DE CAMPO
   - RECLUTAMIENTO DE PERSONAL
   - LEVANTAMIENTO RECOLECCION DE INFORMACION

3) PROCESAMIENTO
   - REVISION CRITICA DE LA INFORMACION
   - CODIFICACION
   - CAPTURA

4) ANALISIS DE RESULTADOS
   - ANALISIS
   - INTERPRETACIÓN
   - DIFUSION

DISEÑO CONCEPTUAL
- REQUERIMIENTOS DE LAS ENCUESTAS
- TABULACION
- DISEÑO DE MANUALES
  - DEL ENTREVISTADOR
  - DEL SUPERVISOR

DISEÑO ESTADISTICO
- ESPECIFICACION DE VARIABLES Y ESCALAS DE MEDICION
- CAPTACION DE LA INFORMACION
- ANALISIS DE LA INFORMACION

DISEÑO ADMINISTRATIVO
- PRESUPUESTO
- MANUAL DE ORGANIZACION
- PERFILES DE RECURSOS HUMANOS
- RECURSOS

El interés de las encuestas se ha centrado en cuatro características del universo de la población en estudio, estas son:

i) Población Total
ii) Media de la población.
iii) Proporción de la población.
iv) Tasa de la población.

II.2.2.1. LA ENCUESTA SOCIAL

La encuesta social son un conjunto de técnicas destinadas a recoger información de las unidades en estudio, principalmente información Demográfica, Socio-económica, de Conducta y Actividad, de Opiniones y Actitudesés, etc.

Exísten varios tipos de encuestas, de acuerdo con la comparación de las poblaciones, las hay, DESCRIPTIVAS y XPLICATIVAS, de acuerdo con la evolución o periodo de referencia del estudio son LONGITUDINALES y TRANSVERSALES.

La versatilidad de la encuesta yace no sólo en la variedad de poblaciones a las cuales puede ser aplicada o en la elección de diseños disponibles, sino también en los muy distintos tipos de datos que pueden ser recogidos. Cualquier hecho o característica de la cual las personas estén dispuestos a informar a un entrevistador puede ser objeto de una encuesta.

Mediante la técnica de la encuesta, en general, se pueden recoger cuatro tipos de información:

a) Características demográficas.
b) Características socioeconómicas.
c) Conductas y actividades.
d) Opiniones y actitudes.

II.2.2.2. DISEÑO DE CEDULAS Y CUESTIONARIOS.

Los cuestionarios y las cédulas de entrevistas son instrumentos destinados a recolectar la información requerida por los objetivos de una investigación. Estos instrumentos se usan, algunas veces, para distinguir las situaciones en las cuales un entrevistador formula determinadas preguntas y llena, con las respuestas, el formulario correspondiente.

Las preguntas de un cuestionario pretenden alcanzar la información que permita cumplir con los objetivos de la invetigación, mediante las respuestas proporcionadas por las personas del universo o la muestra a la cual se refieren aquéllos. En el diseño de cédulas y cuestionarios es importante tomar en cuenta los siguientes aspectos:

a) CONTENIDO DE LAS PREGUNTAS.
b) REDACCION DE LAS PREGUNTAS.
c) TIPO DE PREGUNTAS.
d) ORDEN DE LAS PREGUNTAS.

II.2.2.3. TECNICAS PARA LA CONSTRUCCION DE CUESTIONARIOS DE ACTITUDES

INTRODUCCION

Una de las formas más importantes de colección de datos para un investigador social son las entrevistas y los cuestionarios.

Los cuestionarios son instrumentos de recolección de información, son más rápidos y menos costosos que las entrevistas, pero tienen la desventaja de no detectar el lenguaje del cuerpo ni de las emociones que surgen al momento de ser contestados. Sería muy benéfico, para la investigación poder aplicar ambos instrumentos (aunque ésto no es posible por costo y tiempo).

Existen cuestionarios abiertos y cerrados, en los primeros el sujeto puede responder todo lo que quiera y cuanto se le venga en mente y en los cerrados el sujeto elige una respuesta de un conjunto de opciones previamente definidas.

También existen cuestionarios mixtos y son los que tienen ambos tipos de preguntas.

Para construir un cuestionario se recomienda tener en cuenta lo siguiente:

i) Tener a la vista las hipótesis de investigación.

ii) Elaborar las áreas que debe de abarcar el cuestionario.

iii) Generar tópicos de las áreas, elaborando algunas palabras, frases, etc. que den una pista que deberán conformar el cuestionario final.

iv) clasificar los tópicos en áreas, para distribuir correctamente las ideas.

v) Formular las afirmaciones que se creen formarán parte del cuestinario.

vi) Revisar si todas las afirmaciones tienen que ver TODAS ELLAS con la hipótesis (cuando hay) de la investigación.

vii) Revisar la redacción y ortografía de CADA PREGUNTA.

viii) Verificar la validez concurrente y de apariencia.

ix) Generar para el cuestionario piloto al menos el doble o el triple de campos (preguntas o items) que inicialmente se habían calculado para el cuestionario final.

II.2.2.4. CONFIABILIDAD Y VALIDEZ EN LOS CUESTIONARIOS

La confiabilidad podría entenderse como la congruencia, presición objetividad y constancia de una investigación, también se puede afirmar que la confiabilidad es la capacidad para dar resultados iguales al ser aplicada en condiciones iguales, dos o más veces, a un mismo grupo de objetos.

a) Congruencia, porque las variables y sus indicadores deberán medir la misma cosa.

b) Precisión, porque uno mismo deberá de reproducir varias veces la investigación y deberá de obtener los mismos resultados.

c) Objetividad, porque varios experimentadores deberán realizar la misma investigación y llegar a las mismas conclusiones.

d) Constancia, porque la forma de medición del objeto no debe de alterar los resultados.

La validez podría entenderse como la relación lógica entre las definiciones y las construcciones (ítems, afirmaciones, preguntas, etc) así como la relación empírica del objeto medido con la hipótesis, es decir, es el grado en que el cuestionario mide lo que previamente se propuso medir.

En general, la confiabilidad y la validez son la medición correcta en la forma precisa.

Cuando se realizan estudios éstos pueden no ser válidos ni confiables o pueden serlo. De una u otra forma se debe de reportar la confiabilidad y la validez por dos razones:

i) Se evidencia qué tan efectiva es la investigación.

ii) Se dejan de ocultar las limitaciones reales a las que se enfrentó la investigación.

En general, los estudios concuerdan en que hay seis formas de obtener confiabilidad de una prueba o escala:

a) Antes y después (Test y Retest):
La confiabilidad se consigue, en este caso aplicando la misma prueba a los mismos sujetos en dos ocasiones distintas. Se deben de mantener constantes todas las condiciones y variar únicamente el momento en el que se aplica la prueba, se recomienda dejar transcurrir de uno a 6 meses. Cuando se tienen ambas mediciones se calcula el coefieciente de correlación de Pearson:

$$r = \frac{N \Sigma X Y - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}}$$

r coeficiente de correlación de Pearson
N número de sujetos
X puntuación de la prueba "antes de " (TEST)
Y puntuación de la prueba "después de" (RETEST)

b) Formas paralelas (Paralell Forms): La confiabilidad se consigue en este caso, aplicándole a los mismos sujetos dos pruebas distintas, consideradas como paralelas porque miden los mismo.

El cálculo estadístico es en realidad el mismo que en el caso anterior, con el coeficiente de correlación de Pearson.

Sólo hay que sustituir en X el resultado de la primera prueba y en Y el resultado de la segunda prueba.

c) Forma general del modelo ALPHA (α):

El modelo alpha está basado en el cálculo de la α de Cronbanch y es quizá el coeficiente de confiabilidad más utilizado por los investigadores.

$$\alpha = \frac{K}{K-1}\left\{1 - \frac{\sum_{i=1}^{k} S_i^2}{S_T^2}\right\}$$

donde :

K = número de ítems (preguntas, afirmaciones, reactivos, etc)

S_i^2 = la varianza del i-ésimo ítem.

S_T^2 = La varianza de la suma de los K-ítems.

d) Forma general del modelo ALPHA estandarizado (α_s)

Cuando se quieren homogenizar las unidades de medición se recomienda estándarizar a los datos obtenidos del cuestionario, por lo que la α de Cronbanch se denomina α_S estándarizada.

$$\alpha_S = \frac{K\,\bar{r}}{1 + [(K-1)\,\bar{r}]}$$

Donde:

K = número de ítems
r = la media de las correlaciones entre los campos.

e) División por mitades (Split-Half):

La confiabilidad se consigue correlacionando una mitad de

los reactivos del cuestionario con la otra mitad.

Dependiendo de la igualdad o diferencia Estadística de las varianzas de cada mitad y de la igualdad o diferencia entre los tamaños de cada mitad, se tienen diferentes coeficientes para medir el grado de asociación entre los grupos.

f) Método de Kuder Richardson 20 (KR):
Es utilizado cuando se desea saber si la varianza de un reactivo afecta significativamente los resultados de la prueba.

II.2.2.5. TAMAÑO DE LOS CUESTIONARIOS.

Para calcular el tamaño de un cuestionario es necesario tomar en cuenta la confiabilidad de la prueba, con ésta se puede determinar el número de reactivos que se necesitan para lograr una mejor o peor confiabilidad.

Con el coeficiente de confiabilidad se puede estimar el tamaño del cuestionario:

$$P = \frac{Cd\ (\ 1-\ conf)}{Conf\ (1-Cd)}$$

$$N = Pn$$

donde:

P= la proporción de preguntas que se deben de implantar.

Cd= la confiabilidad deseada (puede variar desde 0.70 hasta 0.99) se recomienda, cuando menos 75%.

Conf=la confiabilidad obtenida en el cuestionario a través de cualquier método.

n= número de ítems del cuestionario o prueba.

N= longitud de la prueba (en ítems) que se necesitan para alcanzar la confiabilidad deseada (Cd)

II.2.2.6. VALIDEZ DE LOS CUESTIONARIOS.

Existen cuatro tipos de validez (además de la interna y externa)

i) Concurrente
ii) Predictiva
iii) De Construcción
iv) De Contenido

II.2.2.7. CUESTIONARIO DE ACTITUDES

Un cuestinario de actitudes es un instrumento de recolección de información, que cumple con las características de un cuestionario convencional en cuanto a su elaboración pero además está enfocado para medir la predisposición hacia un objeto y/o una situación en particular (no miden lo que hace el sujeto sino la predisposición hacia esa conducta). Tal medición se basa principalmente en la teoría del JUICIO COMPARATIVO de Thurstone con lo que se establece la posibilidad de cuantificar toda la experiencia subjetiva. Los principales métodos para medir las actitudes son:

i) Método de comparaciones apareadas de Thurstone.

ii) Método de intervalos aparentemente iguales de Thurstone.

iii) Método de intervalos sucesivos de Thurstone.

iv) Método de rangos sumarizados de Likert.

v) Método de diferencial semántico de Osgood.

vi) Método del escalograma de Gutman.

Siendo el de rangos sumarizados de Likert uno de los más empleados.

La escala de medición para los métodos señalados es la escala de intervalo. Para construir un cuestionario de actitudes se siguen los lineamientos señalados para un cuestionario general, pero además dependiendo del método empleado se deberá de tomar en cuenta lo siguiente:

i) Evitar frases que se refieran al pasado en vez del presente

ii) Evitar frases que puedan ser interpretadas en más de un sentido.

iii) Evitar palabras que no tengan relación con el objeto psicológico medido.

iv) Evitar palabras en las que casi nadie o todos estarían de acuerdo.

v) Evitar frases que contengan universalidad como: todos, siempre, nunca, ninguno, etc.

vi) Evitar palabras que puedan provocar equívocos.

vii) Evitar el empleo de frases negativas complejas

viii) Utilizar un lenguaje claro, simple y directo.

ix) Los reactivos deben de ser cortos, de no más de 20 palabras.

x) Cada reactivo debe contener UNA SOLA IDEA.

II.2.2.7.1. METODO DE RANGOS SUMARIZADOS DE LIKERT

Es una técnica para medir actitudes que cumple con las siguientes características:

i) Se construye un cuestionario piloto con un mínimo de 70 preguntas por cada dimensión (objeto medido, el cual puede contener varias áreas, por ejemplo: la dimensión personalidad puede incluir las áreas de manía, depresión, hipocondriasis, etc.) 35 favorables, 35 desfavorables al objeto medido.

ii) Se trabaja con 5 alternativas que son:
a) Totalmente de acuerdo (TA)
b) Acuerdo (A)
c) Indiferente (I)
d) Desacuerdo (D)
e) Totalmente Desacuerdo (TD)

Lo anterior no indica que no se puedan variar las alternativas, por ejemplo que puedan ir de Muy perfecto a Imperfecto.

iii) La medición de las opciones anteriores, se hace por medio de una escala al menos ordinal, pero respetando que el valor más bajo para una pregunta desfavorable sea el más alto para una favorable.

iv) Al inicio de cada cuestionario (como primer hoja) se anexan las instrucciones (claras) para la forma en que los sujetos van a contestar el cuestionario.

v) Las preguntas (favorables y desfavorables) deberán de quedar en orden aleatorio.

vi) En el cuestionario piloto, el área de datos generales se pone antes de las afirmaciones que miden la actitud.

vii) Para saber si cada pregunta por separado es confiable se utilizan pruebas de t con cada ítem, cuando se rechace la hipótesis nula de cada pregunta, ésta deberá de ser incluida en el cuestionario final.

viii) El cuestionario final debe de contener entre 20 y 25 afirmaciones (la mitad favorables y la otra mitad desfavorables), es decir que de las 70 preguntas iniciales, finalmente deben de quedar 20 o 25 , las cuales tienen puntajes t altos (mayores a 1.75).

ix) Se debe de detectar la confiabilidad del instrumento y calcular la validez predictiva, concurrente, de construcción y cuidar la validez de apariencia.

II.2.2.7.2. METODOS DE INTERVALOS APARENTEMENTE IGUALES DE THURSTONE

Es una escala que permite clasificar, a los sujetos en estudio, con base en propiedades que se distinguen en ellos y se utiliza para medir actitudes; vale decir, predisposiciones individuales a actuar en favor o en contra de personas, organizaciones, objetos, etc.(en favor o en contra de la iglesia, de algún grupo étnico, etcétera).

Tiene tres formas para hacerlo:

a) Método de comparaciones apareadas.

Se presentan un número considerable de afirmaciones, las cuales serán calificadas por los jueces, en cuanto al grado de favorabilidad de los reactivos hacia el objeto medido. Posteriormente se calculan los valores Z y se presentan a los sujetos (se necesita una escala de medición de intervalo).

b) Método de intervalos aparentemente iguales.

Se forman 11 grupos, que van desde totalmente desfavorable (1) hasta totalmente favorable (11) hacia el objeto actitudinal medido. Se calculan los valores escalares y rangos intercuartilares para seleccionar los mejores reactivos. Finalmente se presentan a los sujetos (se necesita una escala de medición al menos ORDINAL).

c) Métodos de intervalos sucesivos.

Es similar al anterior, con la diferencia que se calculan las frecuencias con que los reactivos se asignaron a los 11 grupos como base para evaluar la distancia entre ellos (se necesita una escala de medición de intervalo).

Como el método de intervalos aparentemente iguales tiene la más alta confiabilidad respecto a los otros dos métodos, por lo que se explicará la manera en la que funciona:

Se seleccionan al azar 100 jueces a los que se les presentarán las afirmaciones (entre 50 y 60, previamente elaboradas), que miden la actitud hacia un cierto objeto, los jueces otorgarán un rango a cada afirmación (entre 1 y 11), donde el 1 corresponde al más desfavorable y el 11 al más favorable. Se hacen los cálculos de rangos intercuartílicos considerando a la mediana como medida de tendencia central (localización central)

II.2.2.7.3. METODO DEL DIFERENCIAL SEMANTICO DE OSGOOD (DS)

El diferencial semántico mide las reacciones de los individuos a objetos semánticos; sin embargo, Osgood, Tanenbaum y Suci (1957) definieron su posición con respecto a la adaptación del diferencial semántico a la medición de actitudes.

Se puede aplicar con cualquier clase de estímulos: adjetivos, verbos, grupos étnicos, autoimagen, láminas de pruebas proyectivas, figuras, nombres de personas, etc.

Ventajas:

Es fácil de preparar, aplicar y codificar.

La confiabilidad es alta y la validez se considera bastante razonable.

Maneja una Estadística sencilla

Desventajas:

Cuando los sujetos están muy involucrados en determinado asunto y desean dar respuestas deseables socialmente, es conveniente utilizar otra técnica que no sea la del Diferencial Semántico.

El DS mide el significado connotativo de diversos estímulos (colores, objetos, dibujos, etc.), pero básicamente de estímulos verbales. Presenta tres supuestos básicos.

a) El resultado de la evaluación o juicio puede concebirse como el lugar en que el estímulo ocupa en un continuo experiencial definido por dos términos (adjetivos bipolares).

b) Muchos de los continuos experienciales son esencialmente equivalentes, y por tanto se pueden representar unidimensionalmente.

c) Un espacio semántico (número limitado de continuos que miden cualquier estímulo) contiene básicamente tres factores importantes:

i) Factor evaluativo (E)
ii) Factor potencia (P)
iii) Factor actividad (A)

A los tres factores se les denomina EPA

La elaboración de una escala DS se inicia con la selección de los estímulos que aparentemente midan lo que plantean las hipótesis.

Es necesario seleccionar los estímulos que mejor midan la variable, de tal manera que es conveniente descartar aquellos estímulos (frases) que menos tienen que ver con el objeto medido.

Esto se puede lograr por medio de la técnica descrita por Thurstone, en la que se pone a disposición a gente conocedora del tema y esta escoge las frases que más se tienen que ver con la variable medida. Una vez que los jueces asignan los rangos de menor a mayor se hacen las tablas de distribución porcentual y se calculan los valores escalares y rangos intercuartílicos de las frases, desechándose las que tengan el valor escalar y rango intercuartílico más bajo.

Se deben de seleccionar entre 4 y 7 frases, aunque ésto depende de las áreas que mida el instrumento. Es importante que a los jueces se les presente el doble de estímulos que realmente se necesiten.

Una vez escogidas las frases, se seleccionan los adjetivos bipolares que deben de llevar todas ellas, estos se escogen de la estructura EPA y se recomienda poner tres o cuatro de cada factor, finalmente se diseña el formato del cuestionario con sus respectivas instrucciones.

Una de las más importantes del DS es la validación conceptual por lo que se recomienda correlacionar cada estímulo con el área general.

II.2.2.7.4. CUESTIONARIOS DE OPCION MULTIPLE

Para este tipo de cuestionarios se calculan los coeficientes de confiabilidad y de validez y se realiza un estudio detallado de las opciones, la clave (respuesta correcta), el índice de dificultad y del índice de discriminación, las etapas de elaboración de un cuestionario de opción múltiple no varían de las señaladas anteriormente.

Para elaborar pruebas objetivas se debe de tomar en cuenta lo siguiente:

i) Se calculan los porcentajes y números absolutos de cada ítem, analizando específicamente la frecuencia absoluta y porcentajes de cada opción, contrastándola con el número total de sujetos presentados.

$$\text{Porcentaje de sujetos que escogen una opción} = \frac{\text{Frecuencia}}{\text{Total de sujetos presentados al examen}} \times 100$$

ii) Cuando hay una clave (no siempre sucede) se calcula el porcentaje para cada una de ellas por cada 20% de sujetos, los que irán clasificados desde los puntajes más bajos hasta los más altos.

iii) Se calcula el INDICE DE DISCRIMINACION (ID) (utilizado para ver la efectividad de cada campo (ítem)

$$ID = \frac{\left(\begin{array}{l}\text{Número de sujetos del grupo superior que}\\ \text{contestó correctamente el ítem}\end{array}\right) - \left(\begin{array}{l}\text{Número de sujetos del grupo inferior que}\\ \text{contestó correctamente el ítem}\end{array}\right)}{\left(\begin{array}{l}\text{(Total de sujetos del grupo superior +}\\ \text{total de sujetos del grupo inferior) /2}\end{array}\right)}$$

El ID debe de ser mayor o igual a 0.40 y si es igual a 1 se dice que es perfectamente discriminatorio.

El INDICE DE DIFICULTAD (DIF) indica el grado de dificultad de un ítem.

$$DIF = \frac{\text{Número de sujetos que contestaron correctamente a la clave}}{\text{Total de sujetos}}$$

El DIF debe de oscilar entre 0.20 y 0.80 considerando como ideal 0.50 Un valor DIF =1 indica que el ítem es totalmente fácil y un valor de DIF= 0 indica que el campo es totalmente difícil

La varianza de un ítem se utiliza para detectar el grado de variación y su ecuación es:

$$S_i^2 = p\ q$$

donde:

S_i^2 es la varianza de cada ítem.

p = Proporción de gentes que contestan correctamente el ítem

q = 1-p

III. MUESTREO.

III.1. CONCEPTOS FUNDAMENTALES.

Una de las características de algunas investigaciones antropológicas es que la(s) población(es), en estudio ya no existen y la única manera de estudiarlas es a través de la permanencia de sus "huellas". Estas huellas son muestras muy particulares, por lo que es muy importante que el estudioso de estas áreas conozca técnicas de selección, descripción, exploración e inferencia de la información disponible. Hay también interés en conocer poblaciones actuales por lo que se tiene la necesidad de recolectar la información necesaria para estos fines, se puede reunir todos los datos de cada una de las unidades que la forman (censo) o se puede optar por recoger sólo una parte (muestra); cada una tiene ventajas y desventajas, sin embargo lo que determina el uso de una u otra técnica es el objetivo, los costos, tiempo, recursos humanos, etc.

Las desventajas de un censo son, entre otras, un costo muy superior al de un muestreo, mayor tiempo y mayor necesidad de personal especializado, una desventaja del muestreo es su cobertura, es decir, sólo se tiene una parte de la población, por lo que es importante mencionar que al realizar un muestreo los resultados que se obtengan serán diferentes a los valores que se obtendrían al efectuar el censo.

La búsqueda del conocimiento es tarea incansable de los investigadores, los antropólogos, como se ha señalado previamente, buscan describir poblaciones humanas, sean antiguas o actuales, a travéz de su variabilidad y evolución orgánica, esta descripción es de grupos y no de individuos aunque se base en los mismos para poder generalizar lo encontrado.

En el vocabulario estadístico a las características de la población se les conoce como parámetros poblacionales, y una forma de estimarlos es generalizando lo encontrado en la muestra de la misma población, la diferencia entre el valor obtenido por medio de la muestra y el valor que se obtendría por el censo se conoce como error de muestreo, los errores que se cometen al medir cada una de las características se les conoce como errores de no muestreo y si los tamadores de información los repiten cada vez que toman un dato se afirma que se ésta cometiendo un error en forma sistemática; esto indica que los errores sistemáticos serán mayores mientras más grande sea la muestra, en otras palabras los errores de no muestreo serán mayores en los censos que en las muestras, pero lo importante de esto es que estos errores no se pueden cuantificar, por lo que no se sabrá su influencia en los resultados finales.

Si la selección de una parte de las unidades de la población se hace mediante técnicas probabilísticas entonces se dice que se esta empleando muestreo probabilístico; algunas ventajas de ésta técnica de selección son:

i) Se reduce el costo
ii) Los recursos humanos especializados son más accesibles.
iii) Se reduce el tiempo de levantamiento y de proceso.
iv) Mayor exactitud de la información recabada.
v) Se puede estimar de manera confiable el margen de error.
vi) Se sabe con que probabilidad entra cada elemento muestral.

El principio fundamental del diseño de selección es, que de todos los procedimientos de selección y estimación, se preferirá el que de mayor precisión a un costo mínimo; entendiendo como precisión el mayor acercamiento posible al parámetro.

También hay técnicas de selección que no estan basadas en la teoría de probabilidad algunos ejemplos de muestreo no probablístico son:

Muestras casuales

En este tipo de muestreo se aplican los cuestionarios, entrevistas, mediciones, etc. a las unidades muestrales que se vayan presentando en forma totalmente casual, sin ningún conocimiento acerca de la probabilidad de selección de la unidad muestral. En antropología, es muy común este tipo de muestreo ya que, por ejemplo, al escabar un sitio se van presentando materiales de diversos tipos como: tepalcates, material lítico, obsidiana, etc.

Muestreo por expertos

Los expertos de una área deciden que unidades muestrales entrarán a la muestra, también esto es sin ningún esquema de muestreo, es unicamente sobre el conocimiento subjetivo del experto. En la Antropología Física sucede, por carencias de definiciones operacionales, a la hora de elegir a un sujeto Maya, Otomí, etc. sin que se sepa previamente si es el hablante de una lengua, con ciertas costumbres, nacido en alguna región específica, la combinación de lo anterior, etc.

Muestreo por juicio

Los que miden y registran la información en el instrumento de captación de información (cédulas, cuestionarios, etc) deciden por cuenta propia si incluyen o no al elemento en estudio. (puede ser un caso de muestreo por expertos).

Muestreo por cuotas.

Se solicita a los recolectores de información cumplan con una determinada cuota de elementos, al igual que las anteriores, no existe ningún otro criterio para la selección únicamente cumplir con la cuota.

La selección debe hacerse de una manera tal que la fracción represente lo más real posible al todo. Se debe, por lo tanto, ser cuidadoso en los métodos de muestreo empleados.

Una manera de ampliar el grado de generalidad de los valores hallados en la muestra (validez externa) es utilizando el muestreo probabilístico es entonces cuando se puede confiar en la muestra para que proporcione estimaciones dignas de confianza y no sesgadas de la población.

III.2. CRITERIOS GENERALES DEL MUESTREO.

Criterios estadísticos generales.

Los criterios técnicos y metodológicos mínimos que deben de ser considerados en el diseño e implantación de alguna técnica de muestreo son los siguientes:

a) MARCO POBLACIONAL.

Como todo muestreo está basado en la población, el primer paso metodológico es definirla, la muestra tiene que ser de una población específica y las inferencias sólo serán válidas para la población que se obtuvo la muestra; esta especificación de la población se llama MARCO POBLACIONAL

Evidentemente este marco, tiene que ser exhaustivo y concreto, ya que de otro modo no es posible "saber" a qué o a quién se hace referencia en las inferencias derivadas.

b) DISEÑO MUESTRAL

En este paso metodológico, se tiene que considerar específicamente la naturaleza propia de una encuesta por muestreo, y es que debido a que la inferencia se obtiene a través de una muestra si se sacara otra muestra los resulatdos serían diferentes, pero se desea que no varien mucho, en otras palabras, se trata de que los resultados fundamentales no dependan de la muestra seleccionada, y por lo tanto el criterio esencial es minimizar la variabilidad muestral.

Dependiendo del problema, en el diseño muestral, hay varios criterios que se pueden aplicar como son:

i) La estratificación de la población (Dividir la población en subconjuntos lo más homogeneos posibles, y seleccionar una submuestra de cada uno de ellos en forma independiente).

ii) La dispersión o concentración de la muestra.

iii) El tipo de estimador que se va a utilizar (promedios simples o ponderados, de razón, etc).

iv) El método de selección de la muestra (Aleatorio simple, sistemático, con probabilidades desiguales, etc.)

La lógica de cada uno de estos criterios es diferente para cada problema específico, pero todos tienden a minimizar la variación muestral y lograr que los estimadores de los parámetros sean los "mejores" es decir con poca variabilidad de muestra a muestra.

c) EL TAMAÑO DE LA MUESTRA.

En cualquier aplicación el tamaño de la muestra en una encuesta por muestreo depende fundamentalmente de:

i) Los paramétros que se desea estimar (Totales, promedios, proporciones, etc.)

ii) De la precisión necesaria y

iii) Del nivel de significación para el problema específico, ya que la variabilidad muestral inherente a la metodología, exige aceptar explícitamente un rango de variabilidad razonable de la estimación.

Contrario a la percepción intuitiva, el tamaño de la muestra no depende del tamaño de la población, sino de los elementos mencionados en el párrafo anterior. El tamaño de la muestra depende del tipo de parámetro a estimar, de la precisión y del nivel de significaión en la estimación, pero no del tamaño de la población.

d) SELECCION DE LA MUESTRA.

Como ya se dijo, la muestra se obtiene del marco muestral, mediante mecanismos aleatorios, con probabilidad conocidas, y estas son condiciones necesarias pero no suficientes para que la muestra sea represenativa de la población de referencia.

Para que la muestra sea representativa de la población, hay que obtenerla siguiendo los criterios señalados y verificando que no haya sesgos. Tipicamente en la actualidad esto se hace utilizando computadoras que tienen mecanismos que generan los números aleatorios, y por lo tanto normalmente el marco muestral estará disponoble en medios magnéticos, para poderle aplicar los mecanismos de selección aleatoria.

e) TRABAJO DE CAMPO

Aun cuando haya muestreos de población muy diferente, por ejemplo, de documentos, de personas, etc., la actividad de recabar la información pertinente al muestreo específico genéricamente se llama trabajo de campo.

En este caso, lo relevante es que quienes lo llevan acabo, sigan meticulosamente las instrucciones derivadas del diseño muestral, lo que con frecuencia exige altos niveles de supervisión. Se debe especificar la logística de este trabajo de campo, señalando las acciones a realizarse, y el flujo de la información. Procurando minimizar la ocurrencia de errores en esta etapa de trabajo. Por consiguiente, el trabajo de campo tiene que estar definido específicamente en función del diseño muestral, y ser supervisado en forma muy acuciosa.

f) PROCESAMIENTO DE LA INFORMACION MUESTRAL.

Es necesario distinguir dos fases del procesamiento de información muestral:

i) La aplicación de todos los criterios de procesamiento de cualquier tipo de información, es decir, la veracidad con respecto a la fuente original, la integridad de la información, su estructura, etc, fase en la que se deben de seguir los criterios informáticos estándares.

ii) La incorporación de los criterios estadísticos derivados del diseño muestral para poder hacer inferencias aplicando los estimadores derivados de él.

Esta es una actividad en la que los estadísticos y los informáticos tienen que trabajar en conjunto, para garantizar que los resultados son consistentes con lo diseñado. En el procesamiento de los resulatdos de las encuestas por muestreo es indispensable el trabajo conjunto de los especialistas en informática y en Estadística.

III.3 TECNICAS DE SELECCION DE MUESTRAS.

III.3.1. INTRODUCCION.

El método de muestreo que puede dar estimaciones confiables de la población se conoce técnicamente como Método de Muestreo Probabilítico, está basado en leyes de probabilidad, y en él cada una de las unidades de la población tiene una probabilidad predeterminada c igual de caer dentro de la muestra. El método puede compararse con el barajeo de las cartas, donde cada una de las 52 cartas tiene la misma probabilidad de ser incluida en una mano de 13. Es este método de muestreo, un método de seleccionar la muestra, que proporciona resultados dignos de confianza, y también proporciona una predicción de la magnitud de la incertidumbre a la cual el resultado de la muestra está sujeto.

El muestreo probabilístico es el muestreo estadístico y descansa en los conceptos de aleatoriedad y probabilidad. El concepto de aleatoriedad se refiere a que cualquier muestra para tener calidad científica y poder someterse a tratamiento estadístico debe de ser aleatoria. Es decir que en su selección no deben de intervenir los juicios del investigador sino debe de seleccionarse mediante procedimientos que garanticen la aleatoriedad de la selección; cuando se conoce el número total de elementos de la población a muestrear se selecciona al azar la muestra de interés, esto se puede hacer utilizando los números aleatorios de la calculadora de bolsillo, la computadora, la lista de números aleatorios de cualquier texto estadístico etc.

Es importante señalar que la aleatoriedad y probabilidad no garantizan la representatividad de la muestra la cual por puro efecto del azar puede resultar sesgada y hacer que a sólo determinados miembros de la población les toque la suerte de estar en la muestra.

Entonces para llevar adelante cualquier encuesta, es necesario tener un margen de seguridad de que este procedimiento es el más adecuado, que la información que se desea obtener es el dominio directo y personal de los entrevistados y que sus datos serán consistentes y precisos, no reduciéndose a expresiones vagas o a conjeturas. Esto se puede determinar por medio de los estudios piloto y de premuestreo que anteceden a cualquier encuesta.

Lo anterior sirve también para determinar las preguntas concretas que deberán hacerse a los entrevistados o que los propios encuestadores deben llenar, para lo cual es indispensable plantear escuetamente el problema y condensarlo en uno o dos postulados básicos. A este respecto un autor dice "Cada investigación por muestreo es un recipiente que no puede contener

más información que la que su capacidad le permita y a veces mucho menos, por lo cual no hay que llenarlo con otras cosas superfluas.

Por último, el muestreo, para ser utilizado, necesita representar un ahorro de tiempo y dinero, teniendo a la vez que poder alcanzar mayor amplitud y seguridad.

III.3.2. LAS ESTIMACIONES DE TOTALES.

Como el objetivo de los métodos estadísticos que trabajan con muestras es el de estimar los parámetros o las características de una población a partir de una muestra, lo primero que se necesita es determinar:

1.- La población.
2.- La unidad de muestreo.
3.- La característica que se va a medir.
4.- La variable crítica, o sea, la medida misma.

Por ejemplo: Si se tiene un grupo escolar de 50 alumnos de nivel universitario y se obtiene aleatoriamente una muestra de 10 de ellos para aplicarles una prueba de razonamiento con fines de Normalización de la misma, entonces:

1.- La población es el grupo de 50 alumnos.
2.- La unidad de muestreo es cada alumno
3.- La característica a medir es el razonamiento.
4.- La variable crítica es la puntuación obtenida en la prueba

Como aquí se trata de un grupo que se supone homogéneo, pues ya cursa estudios avanzados, al hacer el análisis estadístico de la muestra se estimarían la media y la varianza del grupo, así como los límites de confianza de estas estimaciones, pero no se podría ir más allá de hacer comparaciones con otros grupos.

Sin embargo, supóngase que en lugar de aplicarles la prueba psicológica a esos 10 estudiantes se les pregunta que digan cuánto dinero lleva cada uno en el bolsillo. En esta situación no sólo se puede sacar la media, la varianza, etc., para la muestra, sino que si se multiplica el promedio de estos 10 alumnos, por 50 que son los que forman el grupo, así que se podrá estimar dentro de sus límites de error, cuánto dinero en total hay en ese momento en el salón de clase.

En este caso:

1.- La población: 50 alumnos.
2.- La unidad de muestreo: cada alumno.
3.- La característica a medir. cantidad de dinero.
4.- La variable crítica: pesos y centavos.

Esta última situación es propiamente lo que se concoce por MUESTREO ESTADISTICO y consiste básicamente en poder hacer estimaciones totales a partir de una muestra.

Es así como con una muestra de unas cuantas parcelas se estima la cosecha total de una gran área, o con el número de habitantes de cuatro o cinco casas de una manzana se calcula el total de personas que habitan en ella, al multiplicar la media de esas casas por el total que de las que forman la manzana.

III.3.3. LOS TIPOS DE MUESTREO PROBABILISTICO.

Existen varios tipos de Muestreo Probabilístico, cada uno de los cuales tiene fórmulas matemáticas y aplicaciones específicas y particulares, siendo su error variable según sea el problema o los recursos técnicos, humanos o económicos de que se disponga.

Entre ellos se pueden citar los siguientes, de acuerdo con la clasificación de Ackoff:

1.- Irrestricto Aletorio donde cada unidad de muestreo de la población tiene un número único e igual de ser elegido.

2.- Sistemático en el cual se utilizan listas, tarjetas, expedientes, etc., y los elementos de la muestra se extraen con determinado intervalo según sea su tamaño.

3.- Aleatorio Multietápico en donde se usa el irrestricto aleatorio para cada etapa del muestreo de las cuales, por lo menos, debe de haber dos.

4.- Estratificado. Tipo de muestreo donde la población se divide en estratos según determinadas características, por lo que, la estimación de la variabilidad queda reducida al eliminar la que existe entre los estratos. Sus modalidades principales son el muestreo estratificado proporcional y con afijación óptima.

5.- Desproporcionado, igual que el estratificado, pero el tamaño de la muestra lo determinan consideraciones analíticas o de covarianza.

6.- Por conglomerados donde las unidades últimas son grupos. Estos se seleccionan al azar y se hace un censo de cada uno.

7.- El muestreo doble en el que setoma una pequeña muestra y si sus resultados son decisivos, ya no se necesita más pero si no, se toma otra muestr. Los resultados de la primera proporcionarían casi siempre las estimadas necesarias, por lo cual la segunda muestra sería planeada muy económicamente, ni muy grande ni muy chica, para obtener elementos definitivos y suficientes para una decisión.

8.- Muestreo secuencial. Su característica es la de que el número de observaciones no se determina de antemano. La decisión de continuar el experimento depende, en cada etapa, de los resultados de la anterior. Su mérito, en lo referente a pruebas de hipótesis, es que pueden hacerse con un número menor de observaciones.

9.- Muestreo secuencial de grupos, que se usa cuando el tratamiento de los datos es relativamente simple y las muestras adicionales son fáciles de obtener o están preparadas de antemano.

Los distintos tipos de muestreo probabilístico se pueden resumir en el siguiente cuadro:

- Aleatorio simple
  - con reemplazo
  - sin reemplazo
- estratificado
  - proporcional
  - no proporcional
- Agrupado (conglomerados)
  - una etapa
  - dos etapas
  - etapas múltiples

En sus diversas modalidades el muestreo probabilístico exige una metodología precisa y única. Aunque en el curso de una encuesta en gran escala pueden combinarse distintos métodos, cada uno tiene sus propias fórmulas matemáticas y, desde luego sus ventajas y desventajas, según sea el problema. Esto implica que sólo es posible hacer modificaciones a juicio del investigador, cuando este conoce a fondo las bases matemáticas en que descansa cada método.

Como se ha mencionado los datos que se estudian pueden ser cuantitativos o cualitativos, o sean de medición o de enumeración, respectivamente. Entre los primeros se tienen algunos, tales como superficies cultivadas, producción de una parcela, ingreso en pesos, número de miembros por familia, cabezas de ganado, etc.; entre los segundos, condiciones tales como soltero o casado, agricultor o alfarero alfabeto o analfabeto, etc. La naturaleza de estos datos requiere de fórmulas matemáticas específicas para cada caso.

POBLACIÓN Y MUESTRA.

A diferencia de otros procedimiento estadísticos donde las poblaciones se consideran infinitas y las muestras finitas, utilizandose respectivamente letras griegas y latinas para designar sus parámetros, las primeras y sus estimaciones las segundas, el MARCO O CAMPO se conoce, pues se tiene el total de parcelas, casas, manzanas, obreros, ejidatarios, etc., y la muestra es sólo parte de estos conjuntos.

El marco, campo, lista maestra, etc., son los términos con los que se designa a la población. Entre el marco y ésta pued no haber exacta coincidencia, sea por omisiones o porque la lista o la relación de las unidades de muestreo no esté exactamente al día. Aquí no hay manera, desde el punto de vista estadístico, de llenar esta diferencia. Es entonces en el juicio del investigador donde se apoya la decisión de considerar si tal o cual marco representa a la población.

III.3.4. EL MUESTREO IRRESTRICTO ALEATORIO CON DATOS CUANTITATIVOS O CONTINUOS.

Es el tipo de muestreo por excelencia y la base de todos los demás, os cuales a su vez, en alguna de sus etapas tienen que recurrir a su empleo. Es asi mismo el más simple ya que sólo basta con asignar un número a las unidades de la población y usando alguna técnica aleatoria para integrar la muestra.

Notación:

N tamaño de la población.

Y_i valor de la variable Y en el elemento i-ésimo de la población.

$Y = \sum_{i=1}^{N} Y_i$ es el valor total de la variable

$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$ El promedio poblacional

$\sigma_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{N}$

$S_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{N - 1}$

La notación para la muestra es:

n tamaño de la muestra

y_i valor del i-ésimo elemento de la variable y

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ el promedio muestral.

$s_y^2 = \frac{\sum \left[y_i - \bar{y} \right]^2}{n-1} = \sum y_i^2 - n \bar{y}^2$ varianza muestral

La estimación del promedio de la población $\bar{Y}$ es:

$$\hat{Y} = \bar{y} = \text{media muestral.}$$

Esta estimación tiene un error estandar:

$$S_{\bar{y}} = \frac{S}{\sqrt{n}} \sqrt{\left(1 - \frac{n}{N}\right)}$$

Los límites de confianza para la media poblacional con un un nivel de confianza 1-α es:

$$\bar{Y} \in \bar{y} \mp Z_{1-\alpha/2} \frac{S}{\sqrt{n}} \sqrt{\left(1 - \frac{n}{N}\right)}$$

El total de la población se estima mediante:

$$\hat{Y} = N \bar{y} = N \frac{\sum_{i=1}^{n} y_i}{n}$$

El error estandar de esta estimación es:

$$S_{\hat{Y}} = \sqrt{N^2 \left(1 - \frac{n}{N}\right) \frac{s^2}{n}} = s\, N \sqrt{\frac{\left(1 - \frac{n}{N}\right)}{n}}$$

El intervalo de confianza estimado para el verdadero valor del total de la población con un nivel 1-α de confianza es:

$$Y \in \left(N \bar{y} \mp Z_{1-\alpha/2} \; N \sqrt{\frac{\left(1 - \frac{n}{N}\right)}{n}} \right)$$

El tamaño de la muestra con una confianza de 1-α y una precisión δ;

$$n_0 = \frac{Z^2_{1-\alpha/2}\, s^2}{\delta^2}$$; s^2 es la desviación estandar que se conoce por experiencias previas o por una muestra piloto.

$$\therefore \quad n = \frac{n_0}{1 + \left(\frac{n_0}{N} \right)}$$

III.3.5. EL MUESTREO IRRESTRICTO ALEATORIO CON DATOS CUALITATIVOS O DISCRETOS.

Se ha mencionado que otro de los parámetros de interés son las proporciones poblacionales, supongase que la población esta dividida en dos partes, en una de ellas esta los que presentan la característica C y tiene un tamaño A y la otra por la que carecen de ella con un tamaño A', por lo que la proporción de unidades muestrales con la característa C es:

$$P = \frac{A}{N}$$

para la muestra

$$p = \frac{a}{n} \quad ; \quad q = 1-p = \left(1 - \frac{a}{n}\right)$$

en este caso la media aritmética es la proporción:

$$\bar{y} = \frac{a}{n} = p$$

El problema entonces es estimar A y P, es decir, el total y la media o proporción.

La estimación muestral de P es p

El error estándar de esta estimación es:

$$s = \sqrt{\frac{N - n}{(n-1)} p \ q}$$

El intervalo con un nivel α de confianza para P es:

$$P \in \left(p \mp Z_{1-\alpha/2} \sqrt{\frac{N - n}{(n-1)} p \ q} \right)$$

La estimación muestral de A es Np o $N \frac{a}{n}$

El error estándar de esta estimación es:

$$s = \sqrt{\frac{N \ (N - n)}{(n-1)} p \ q}$$

El intervalo con un nivel α de confianza para el total A es:

$$A \in \left[Np \mp Z_{1-\alpha/2} \sqrt{\frac{N\,(N-n)}{(n-1)}\,p\,q} \right]$$

Para obtener el tamaño se debe de preestablecer una precisión δ, un nivel α de confianza así como dar una idea de la variación poblacional (por medio de experiencias previas o de muestras piloto) con esto se obtiene un tamaño de muestra aproximado:

$$n_0 = \frac{Z^2_{1-\alpha/2}\; p\,q}{\delta^2}$$

esta primera aproximación del tamaño de muestra se puede corregir por medio de:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

III.3.6. MUESTREO ESTRATIFICADO.

Cuando la población esté constituida por unidades hetereogéneas y se pueda tener una idea previa de los grupos (a través de encuestas censos u otros medios que se hayan realizado con anterioridad, a esta información se le conoce como complementaria) el metódo de estratificación asigna a las unidades de la población a grupos o estratos de acuerdo con la información sobre x. Se trata de volver más homogéneos los estratos colocando en el mismo estrato unidades que sean similares, es decir, unidades más homogéneos entre sí, por lo que los estratos son subconjuntos de la población que agrupan unidades homogéneas, aunque sean heterogéneas entre estratos.

Cada estrato se muestrea por separado y se obtienen los estimadores de parámetros (totales, medias, proporciones) para cada estrato. Se supone que se conoce el número de unidades en cada estrato N_h .

A manera de resumen se tiene que la estratificación se recomienda cuando:

1) la población es heterogénea
2) Se conoce el tamaño de cada uno de los estratos (# de unidades)

Notación:

L número de estratos
h= estrato
i= unidad
N_h = número de elementos en el h-ésimo estrato
n_h = número de elementos del estrato h en la muestra
y_{h_i} = valor de la característica a medir en la unidad i-ésima del h-ésimo estrato; donde h=1...L.

$\bar{Y}_h$ =media poblacional del estrato h-ésimo
Y_h =total de la población del estrato h-ésimo
S_h^2 = varianza poblacional del estrato h-ésimo.

Cálculos poblacionales.

$$N = \sum_{h=1}^{L} N_h \quad \text{total de unidades en la población}$$

$$Y_h = \sum_{i=1}^{N_h} Y_{h_i} \quad \text{total de la población en el estrato h-ésimo}$$

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (Y_{h_i} - \bar{Y}_h)^2}{N_h - 1} \quad \text{varianza de la población en el estrato h}$$

$$n = \sum_{n=1}^{L} n_h \quad \text{tamaño de la muestra}$$

III.3.6.1. MUESTREO ALEATORIO ESTRATIFICADO CON DATOS CUALITATIVOS (PROPORCIONES).

Como se mencionó, uno de los supuestos para aplicar esta técnica es que se conoce el tamaño de cada estrato y el número de estratos que se forman. Para calcular el tamaño de la muestra es necesario conocer la variabilidad dentro de cada estrato, por lo que se recomienda hacer una muestra piloto o investigar en experiencias previas. Una vez que se ha determinado el tamaño de muestra existen varias formas de asignar elementos de cada estrato a la muestra.

Asignación de elementos en la muestra.

Esta asignación se puede hacer por medio de asignación proporcional, asignación igual, asignación óptima (Neyman), asignación óptima de costos variables.

Asignación igual.

$\frac{n}{L} = n_h$; se divide el tamaño de la muestra entre el número de estratos, el resultado es el número de elementos que deberán de ser seleccionados por cada estrato.

Asignación proporcional al tamaño del estrato.

$$n_h = k\, N_h$$

$$\sum n_h = k \sum N_h$$

$$k = \frac{\sum n_h}{\sum N_h} = \frac{n}{N}$$

$$\Rightarrow n_h = \frac{n}{N}\, h = n\, \frac{N_h}{N}$$; el número de elementos, con los que contribuirá cada estrato para formar la muestra final, es el resultado de multiplicar el tamaño de la muestra final deseada por la proporción que represente cada estrato del total.

Asignación Optima (NEYMAN):

$$n_h = \frac{n\, N_h\, S_h}{\sum_{h=1}^{L} N_h\, S_h}$$;

Asignación Optima (De costos variables entre estratos).

$$n_h = \frac{\frac{n\, N_h S_h}{\sqrt{C_h}}}{\sum \frac{N_h\, S_h}{\sqrt{C_h}}}$$

La estimación dependerá de la forma en que se halla asignado la muestra a los estratos, por lo que sólo se dará la estimación para la asignación proporcional al tamaño del estrato.

ESTIMACION PARA DATOS CUALITATIVOS (PROPORCIONES).

ESTRATO	TAMAÑO DEL ESTRATO	p_h	q_h	$N_h\ p_h\ q_h$
1	N_1	p_1	q_1	$N_1\ p_1\ q_1$
2	N_2	p_2	q_2	$N_2\ p_2\ q_2$
3	N_3	p_3	q_3	$N_3\ p_3\ q_3$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
L	N_L	p_L	q_L	$N_L\ p_L\ q_L$
total	N			

Para tener una idea de las proporciones con la característica deseada en cada estrato se recurre a una muestra piloto o a información de experiencias previas.

Para determinar el tamaño de la muestra es necesario saber de antemano la forma en que se va a asignar a los diferentes estratos, además de contar con información sobre p y q (por medio de información complementaria o de una muestra piloto), también es necesario establecer la varianza deseada V. Por lo que el cálculo del tamaño de la muestra (para asignación proporcional) está dado por:

$$n_0 = \frac{\sum \frac{N_h}{N}\ p_h\ q_h}{V} \quad ; \quad n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Una vez que se ha determinado el tamaño de la muestra el siguiente paso es asignar la misma en forma proporcional a los diferentes estratos.

$$n_h = \frac{n}{N} N_h = n \frac{N_h}{N}$$

ESTIMACION.

La estimación de la proporción total es:

$$\hat{P} = \sum_{h=1}^{L} \frac{N_h}{N} P_h$$

La magnitud del error que hay en esta estimación es:

$$S_{\hat{P}} = \sqrt{\frac{1}{N^2} \sum \left(N_h \ (N_h - n_h) \ \frac{p_h \ q_h}{n_h} \right)}$$

El intervalo con un 95% de confianza para la proporción:

$$P \in \left(\hat{P}_h \mp 1.96 \sqrt{\frac{1}{N^2} \sum \left(N_h \ (N_h - n_h) \ \frac{p_h \ q_h}{n_h} \right)} \right)$$

III.3.6.2. ESTIMACION DEL PROMEDIO Y EL TOTAL POBLACIONAL CON DATOS CUANTITATIVOS

Al igual que en el caso de las proporciones, para poder estimar el promedio y el total poblacional, se debe de contar con información complementaria, esta, como ya se dijo puede provenir de una muestra piloto o de experiencias previas. Antes de estimar los parámetros poblacionales se debe proceder al cálculo del tamaño de la muestra:

Para el caso de la media la estimación del tamaño de muestra se hace conociendo la variación en la población o la variación en cada estrato, la precisión deseada y la confianza requerida.

En el primer caso el tamaño de muestra es:

$$n = \frac{N \sum N_h S_h^2}{N^2 D^2 + \sum N_h S_h^2}$$

$$D = \frac{\delta}{Z_{1-\alpha/2}}$$

δ es la precisión deseada

$Z_{1-\alpha/2}$ es la confianza establecida

o por:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Por lo que se tiene:

ESTRATO	TAMAÑO DEL ESTRATO	s_h^2	$N_h s_h^2$
1	N_1	s_1^2	$N_1 s_1^2$
2	N_2	s_2^2	$N_2 s_2^2$
3	N_3	s_3^2	$N_3 s_3^2$
.	.	.	.
.	.	.	.
.	.	.	.
L	N_L	s_L^2	$N_L s_L^2$
Total	N		$\sum \left(N_h s_h^2 \right)$

Así la asignación proporcional por estrato es:

$$n_h = \frac{n}{N} N_h = n \frac{N_h}{N}$$

Una vez calculado el tamaño de muestra y decidido la forma de asignación se procede a estimar los parámetros poblacionales.

La estimación del total es:

$$\hat{Y} = \sum_{h=1}^{L} N_h \bar{y}_h \quad ; \quad \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{h_i}}{n_h}$$

$$Var(\hat{Y}) = \sum \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) s_h^2$$

$$s_h^2 = \frac{\sum (y_{h_i} - \bar{y}_h)^2}{n_h - 1}$$

La estimación del promedio es:

$$\hat{\bar{Y}} = \frac{\hat{Y}}{N} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} \; ; \; \bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{h_i}}{n_h}$$

$$\text{Var}\left(\bar{y}_h\right) = \sum W_h^2 \frac{S_h^2}{n_h} \left(1 - f_h\right) \; ; \; f_h = \left(1 - \frac{n_h}{N_h}\right) \; ; \; W_h = \frac{N_h}{n_h}$$

$$S_{y_h}^2 = \frac{\sum_{i=1}^{n} \left(y_{hi} - \bar{y}_h\right)^2}{n_h - 1}$$

IV. EJECUCION

Una vez que se ha diseñado y dado el formato de investigación, el paso natural es la ejecución que se lleva se lleva a cabo mediante los siguientes lineamientos:

1 Recolección de la información.
2 Recuento de la información
3 Presentación de la información.
4 Descripción y análisis estadístico.
5 Informe técnico.

5.1 Título.
5.2 Autor.
5.3 Resumen.
5.4 Introducción.

5.4.1 Antecedentes.
5.4.2 Problema.
5.4.3 Hipótesis (cuando hay).

5.5 Método.

5.5.1 Variables.
5.5.2 Diseño.
5.5.3 Procedimiento.
5.5.4 Recursos.
5.5.5 Cronograma.

5.6 Resultados.
5.7 Discusión.
5.8 Referencias.

IV.1. PROCESAMIENTO DE LA INFORMACION.

Continuando con el proceso de investigación y recordando el esquema general del diseño estadístico:

Diseño estadístico
a) Especificación de variables y escalas de medición.
b) Diseño Muestral.
i) Cuándo muestrear.
ii) Qué muestrear
iii) Cómo muestrear
iv) Métodos de muestreo.
v) Determinación del tamaño de muestra.
vi) Comparabilidad de las muestras de poblaciones.
vii) Diseño de las maniobras o tratamientos (factor causal)
c) Proceso de captación de la información.

d) Análisis e interpretación de la información.

Especificadas las variables con sus escalas de medición y habiendo terminado el diseño muestral se procedió, por medio de los intrumentos de recolección, a la toma de datos.

IV.1.1. CODIFICACION

Con los datos asentados en las cédulas, cuestionarios, archivos, etc, y con los objetivos perfectamente claros se procede a la codificación de los mismos. La codificación consiste en asignar números iguales a respuestas iguales o a características iguales de las unidades en estudio, hay ítems del instrumento de recolección que se pueden precodificar y otros no.

IV.1.2. PROCESAMIENTO

Cuando se tiene la información codificada con su respectivo manual se crean los archivos en computadora, éstos deberán contemplar varias estructuras de datos, sin olvidar, que deben llevar una clave de acceso para saber a que elemento de estudio pertenecen y poder depurar la información de manera satisfactoria.

El o los programas de computación de análisis estadístico que se utilicen permiten establecer los tipos de estructuras de datos necesarios, así mismo se debe de contemplar previamente los alcances y límites de los diferentes programas como los requerimientos de software y hardware, entre otros, indispensables para un buen funcionamiento de los mismos.

Es importante señalar que uno de los criterios para decidir sobre cuál paquetería utilizar es la transportabilidad de los datos tanto de entrada como de salida, tipos de gráficas, de edición, selección, clasificación, etc., así como las diferentes posibilidades de análisis estadístico; no es raro que se tenga necesidad de utilizar varios paquetes y que por lo mismo se debe de contemplar la necesidad de tener diferentes archivos en diversos formatos por lo que hay que ser muy cuidadoso en la forma que se sistematiza y se respalda la información de la investigación.

Por otra parte hay que recordar que parte de la información es generada a partir de información primaria (índices, grupos, tasas, coeficientes, etc.) y por tal motivo la estructura y codificación cambiará, por lo que hay que rehacer los manuales de códigos con sus respectivas estructuras.

La información se puede guardar en el formato del paquete de captura acompañandola de documentación, también se puede guardar en formato ASCII esto permitirá su uso posterior sin ninguna

dificultad. Se debe guardar una impresión de buena calidad de la información, acompañandola con toda la documentación posible, fechas, tipos, fórmulas, etc..

Teniendo los respaldos de la información necesarios y la seguridad de la calidad de los datos se procede al análisis de la información.

V. ANALISIS E INTERPRETACION DE DATOS.

La información recabada con los diferentes instrumentos difícilmente podría ser manejada en su presentación original, por esta razón, es necesario sintetizar la información fuente, esto es, reunir, clasificar, organizar y presentar la información en cuadros, gráficas o relaciones de datos con el fin de facilitar su análisis e interpretación.

Estas etapas se encuentran estrechamente ligadas, por lo cual suele confundírseles. El análisis consiste en separar los elementos básicos de la información y examinarlos con el propósito de responder a las distintas cuestiones planteadas en la investigación. La interpretación es el proceso mental mediante el cual se trata de encontrar un significado más amplio de la información empírica recabada.

La primera fase del análisis de la información es la exploración y descripción de los datos, esto es necesario para encontrar concentraciones, dispersión o variabilidad, tendencias, distribuciones, relación entre variables, etc.

Antes de iniciar con este tema se darán algunas definiciones que serán de utilidad en la presentación de la información.

V.1. RAZONES, PROPORCIONES, PORCENTAJES.

Estas son los principales estadísticos que corresponden al nivel de medición de las escalas nominales. La eficiencia de los mismos depende de la clasificación y la clara definición de las categorías. Cada caso observado debe pertenecer a una y sólo una, y éstas no deben de yuxtaponerse en ningún punto.

RAZONES

Por razón se entiende la relación cuantitativa entre dos magnitudes similares, determinadas por el número (entero o fracciónal) de veces que una contiene a la otra. De ahí que la razón no dependa del tamaño entre ambos. Por ello las razones de 3 a 2; de 75 a 50; de 10920 a 7286, son todas ellas iguales a 1.5

Una razón es un cociente que resulta de dividir dos cantidades, el resultado es la relación que hay de un número con respecto a otro.

PROPORCIONES

Las proporciones son un tipo especial de razones en las que el denominador es el número total de casos (N) y el numerador (n) una cierta parte de éste. En las razones, por lo general, se hace referencia a casos en los que las categorías están separadas y no pertenecen al mismo universo.

Así se tiene que $\frac{125}{25} = 5$

lo que se expresa como 5:1, y quiere decir que por cada 5 personas que tienen sangre tipo B hay 1 de sangre tipo AB.

Las razones pueden dar como resultado valores mayores a 1.

Ejemplo:

Si se tienen 4 tipos de sangre A, B, AB, O y se tiene una muestra con tamaño n= 1250 y se hacen 4 grupos de acuerdo al tipo sanguineo quedando de la siguiente forma:

$$n_A = 362,\ n_B = 125,\ n_{AB} = 25,\ n_O = 738;$$

$$n_A + n_B + n_{AB} + n_O = 1250$$

por lo que la proporción de personas en cada grupo es:

$$p_A = \frac{362}{1250} = 0.29,$$

$$p_B = \frac{125}{1250} = 0.10$$

$$p_{AB} = \frac{25}{1250} = 0.02,$$

$$p_O = \frac{738}{1250} = 0.59$$

La suma de las proporciones de las categorías que integran un todo es la unidad.

$$\frac{n_A + n_B + n_{AB} + n_O}{n} = \frac{n}{n} = 1$$

PORCENTAJES

Cuando una proporción se múltiplica por 100 se obtiene un porcentaje, asi el 29% de la muestra tiene sangre tipo A, el 10% sangre tipo B, el 2% AB, y el 59% tipo O; la suma de estos porcentajes da el 100% del grupo estudiado.

Una de las formas gráficas para representar los porcentajes son los diagramas de circulares (de pastel o pays) donde cada "rebanada" es la proporción del subgrupo en cuestión.

TASAS.

Una diciplina de suma importancia para la Antroplogía Biológica es la Demografía que es el estudio del tamaño, distribución territorial, composición, dinámica, y componentes de dicha dinámica de la población, las Estadísticas Demográficas son registros cuantitativos de las características biológicas y sociales de una población, referidas en un momento dado o bien, a cambios continuos que ocurren durante un periodo dado o, bien a cambios continuos que ocurren durante un periodo de tiempo. La materia prima del análisis demográfico está constituido por los datos estadísticos obtenidos a partir del censo, o bien, del registro civil, que informa cuántas personas, o cuántos acontecimientos sucedieron en un periodo de tiempo, respectivamente.

En el caso de estudios antropológicos en los que la población objetivo no se puede delimitar en tiempo y algunas veces ni en espacio, los datos demográficos se tienen que "estimar" a partir de datos no oficiales como fuentes históricas o con datos que los mismos encuestados, entrevistados, etc. pueden dar.

Una tasa expresa la proporción numérica que existe entre dos series de datos. En demografía, las "tasas vitales" -eventos de estado civil- que acontecen en una población x, son razones cuya característica es que invariablemente se expresan por cien, por mil; se distinguen, por el tipo de fuente de la que provienen -estadísticas vitales-, y porque toman en consideración un periodo estándar, generalmente de un año. Desde este punto de vista, las "las tasas vitales" son un tipo especial de "razones".

Así, por ejemplo, la Tasa Bruta de Natalidad (TBN) que establece el número de nacidos vivos por cada 1000 habitantes se expresa por:

TASA BRUTA DE NATALIDAD (TBN):

$$TBN = \frac{\text{No. de nacidos vivos durante el año } x}{\text{Población total estimada a mitad del año } x} \quad (1000)$$

TASA DE MORTALIDAD GENERAL (TM_g):

$$TM_g = \frac{\text{No. de Defunciones durante el año x}}{\text{Población total estimada a mitad del año x}} \quad (1000)$$

TASA DE MORTALIDAD INFANTIL (TM_i):

$$TM_i = \frac{\text{No. de Defunciones infantiles menores de un año}}{\text{Número de nacidos vivos}} \quad (1000)$$

V.2. DESCRIPCION Y EXPLORACION DE LA INFORMACION.

Para lograr los objetivos de la estadística descriptiva y exploratoria, es necesario que los datos se resuman en tablas y se representen por medio de gráficos lo que dará una visualización rápida y clara, por tanto, es necesaria ordenarla de alguna forma, ya sea por presencia o ausencia de características, por el "orden" de importancia que el investigador le de a la información o, por la naturaleza de los datos, de menor a mayor.

ESTADISTICOS DE ORDEN

Las variables ordinales y de intervalos se pueden ordenar de menor a mayor o de mayor a menor sin perder su valor.
Por ejemplo si se tienen:

X_1	X_2	X_3	X_4	X_5	X_6	X_7
17	28	30	10	45	22	33

$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$
10	17	22	28	30	33	45

en donde:

X_i representa al i-ésimo valor muestral

$X_{(i)}$ representa a la i-ésima Estadística de orden.

Así que el 10 representa la primer Estadística de orden, 17 la segunda etc.

V.2.1. MEDIDAS DE TENDENCIA CENTRAL

MEDIA ARITMETICA ($\bar{x}$) muestral se define como la suma de todos los valores divididos entre n, es decir:

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

Se dice que una Estadística (función que está en términos de la muestra) es resistente si el estadístico tiene la capacidad para "amortiguar" los cambios bruscos o arbitarios que un conjunto de datos puede sufrir, la media aritmética no es resistente, por lo que se propone a la mediana como estadístico central.

La MEDIANA es un estadístico que da el valor medio (centro) de un conjunto de datos ordenados, es decir, es el valor en el que el 50 % de los datos son menores a él y el otro 50% son mayores.

La manera de calcularla es:

Si se tiene a N como la CARDINALIDAD (tamaño o número de elementos) del conjunto, y si ésta es impar (de la forma N=2k+1) para k =0,1,2,3.... entonces:

$$\text{MEDIANA} = y_{(N+1)/2} \quad \text{Estadística de orden}$$

Si N es par (de la forma N=2k) para k=0,1,2,..... entonces la mediana es igual al promedio aritmético de las siguientes estadísticas de orden $y_{N/2}$, $y_{(N/2+1)}$

Otra medida de centralización es la media podada que se define como:

$$\frac{1}{4}(\text{Primer cuartil}) + \frac{1}{2}(\text{Segundo cuartil}) + \frac{1}{4}(\text{Tercer cuartil})$$

V.2.2. MEDIDAS DE DISPERSION.

Mínimo, máximo, rango, cuartiles, rango intercuartílico, varianza, desviación estándar.

MINIMO Es el valor más pequeño entre $\left[X_{(1)}, X_{(n)}\right]$ donde las $X_{(1)}, X_{(n)}$ representan la primer y última estadísticas de orden.

MAXIMO Es el valor más grande entre máximo $\left[X_{(1)}, X_{(n)}\right]$ donde las $X_{(1)}, X_{(n)}$ representan la primer y última estadísticas de orden.

CUARTILES. Haciendo una extensión de la idea anterior los valores que dividen a nuestros datos en cuatro partes iguales sè les llama cuartiles, los denotaremos Q1, Q2, Q3, primero, segundo y tercer cuartil respectivamente, donde Q2 es la mediana.

RANGO Es la diferencia entre la N-ésima y la primera estadística de orden, por lo que: $RANGO = X_{(N)} - X_{(1)}$

RANGO INTERCUARTILICO:

Es la diferencia entre el tercer y el primer cuartil.

$RI = Q_{(3)} - Q_{(1)}$

DESVIACION ABSOLUTA CON RESPECTO A LA MEDIANA

Adicionalmente a las estadísticas de dispersión se define la desviación absoluta con respecto a la mediana:

$DAM = |X_i - \text{mediana}|$ para $i=1,2,\ldots,n$

VARIANZA muestral se define como:

$$S^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n-1} \quad ; \; i = 1 \ldots n$$

La desviación estándar es la raíz cuadrada de la varianza:

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n-1}} \quad ; \; i = 1 \ldots n$$

A la primera y última Estadística de orden junto con el primero, segundo y tercer cuartil se les conoce como diagrama de valores de las literales, este diagrama sirve a manera de cuadro resumen de la información y también para construir los diagramas de caja que se verán más adelante.

Los cinco valores (mediana,mínimo, máximo, primer, tercer cuartil) hasta este momento calculados representan el primer cuadro resumen de estadísticas básicas del análisis exploratorio.

V.2.3. COEFICIENTE DE VARIACION

Una de las estadísticas más utilizada para describir la variación de un grupo de datos es el Coeficiente de Variación y éste se calcula de la siguiente manera:

$$C.V. = \frac{100 \; (\; S \;)}{\bar{x}}$$; S es la varianza y $\bar{x}$ la media del grupo.

V.3. DISTRIBUCION DE FRECUENCIAS.

Cuando se dispone de un gran número de datos, es útil el distribuirlos en clases o categorías y determinar el número de casos que pertenecen a cada clase. No existen reglas universales para establecer el número de intervalos ni la amplitud de los mismos, pero se recomienda que sean entre 5 y 12, que el primer intervalo parta del valor mínimo de los datos y el último intervalo termine con el valor máximo, el ancho de cada clase se determina en la forma siguiente:

$$\frac{\text{Valor máximo - Valor mínimo}}{\text{número de intervalos deseado}} = \frac{\text{rango o recorrido}}{\text{número de intervalos deseado}}$$

Una vez que se ha determinado el número de intervalos y que se ha obtenido el ancho de clase, se oberva en la información cuál es la frecuencia que tiene cada dato y se anota en el lugar correspondiente. El procedimiento que da una ordenación tabular de los datos en clases con las correspondientes frecuencias a cada una, se le conoce como una distribución de frecuencias o tablas de frecuencias.

Es importante que los intervalos queden de la misma longitud, que la amplitud de los mismos no sea tanta, que provoque una mayor inexactititud, ya que no se ponen los valores originales, sino, sólo las frecuencias. A los intervalos se les conoce como CLASES y a sus puntos medios, como MARCAS DE CLASE, al valor inicial y final del intervalo se les llama limite inferior y superior del mismo.

Para construir las tablas de frecuencias se procede de la siguiente manera:

En la primer columna de la tabla se ponen los diferentes intervalos, en la segunda columna sus respectivas frecuencias y para asegurarse que estan todas se suman y se anota el total al pie de la segunda columna debiendo corresponder esta suma de frecuencias con el tamaño de la muestra estudiado. En la tercera columna se anota el porcentaje que corresponde a la frecuencia múltiplicada por 100 y dividida entre el tamaño de la muestra, a esta columna se le conoce como Porcentajes Relativos, la cuarta columna corresponde a la frecuencia absoluta acumulada que es el resultado de ir sumando las frecuencias absolutas de cada intervalo, la quinta columna se procede de manera similar pero con los porcentajes relativos.

Las tablas de distribución de frecuencias se hacen para cada variable y/o subgrupo natural o artifial de los datos, así se puede tener la forma de la distribución de la variable en cuestión, por ejemplo: la edad, escolaridad, ingreso, etc.

La representación gráfica de estas tablas se obtiene mediante los histogramas, polígonos de frecuencias y OJIVAS.

Un histograma es la representación de las frecuencias absolutas de una tabla de distribución de frecuencias. En el eje de las abscisas se ponen los intervalos de clase, y en el eje de las ordenadas a las frecuencias de los mismos. Las marcas de clase son los puntos medios de cada intervalo y la altura es la frecuencia de los datos de cada intervalo.

La unión de los puntos medios en la parte superior de cada intervalo se le llama polígono de frecuencias.

La representación gráfica de las frecuencias relativas acumuladas se les conoce como OJIVAS y se construye de la manera siguiente: en el eje de las abscisas del plano cartesiano se anotan los intervalos de clase y en el eje de las ordenadas se anota el valor de las frecuencias relativas acumuladas, por lo que ésta gráfica comenzará desde el valor cero y se irá acumulando en cada intervalo hasta llegar al ciento por ciento.

Otra forma de obtener la distribución de frecuencia de un grupo de datos es mediante la gráfica de tallo y hoja.

V.3.1. DIAGRAMAS DE TALLO Y HOJA.

Los diagramas de tallo y hoja son métodos gráficos que tienen como objetivo la descripción de los datos y son un método alternativo a los histogramas y polígonos de frecuencias, es decir, ayudan a visualizar algún posible patrón como puede ser:

Concentración.
Dispersión
Distribución
Simetría
Agrupación
Tendencia, etc.

Para aplicar la técnica de tallo y hoja los datos deben de estar ordenados de menor a mayor o de mayor a menor.

NUMERO DE TALLOS Y HOJAS EN LOS DIAGRAMAS.

Cuando se esta trabajando manualmente se podrá definir de acuerdo a "n" (número total de datos) cuantos tallos se necesitan, pero cuando ésto se hace con paquetes de cómputo esto ya esta previamente definido sin embargo se tienen las siguientes opciones para determinar el número de máximo de tallos (líneas).

i) $L=[\ 10 \times \log 10\ n]$

ii) $L=2 * \sqrt{n}$

ii) $L=1+LOG_2\ n$

Los corchetes indican que se debe de tomar el máximo entero que no exceda al valor resultante, si por ejemplo se tienen n=21 datos:

$$L = [10*\log_{10} 21] = [10 * 1.32] = [13.2] = 13$$

$$L = [\ 2 * \sqrt{21}\ ] = [2 * 4.58] = [9.16] = 9$$

$$L = [1 + \log_2(21)] = [1 + (4.392)] = [5.4] = 5$$

Los resultados anteriores indican que el número de líneas en el diagrama de tallo y hoja corresponden a los valores entre 5 y 13 líneas.

Por otro lado se debe de establecer un mecanismo para determinar la amplitud (intervalo) de las hojas, insistiendo que si se esta trabajando con paquetería estas formas ya están predeterminadas, por lo que se define la siguiente fórmula:

$$\text{No.hojas} = \frac{\text{rango}}{L}$$; este valor se redondea al entero más próximo.

Esta parte se finaliza presentando una tabla resumen para determinar el número de tallos, dependiendo de la n que se esté trabajando, sin olvidar que sólo es una propuesta ya que realmente esto depende de la investigación que se este realizando y del propio investigador.

n	$10\log_{10} n$	$2*\sqrt{n}$	$1+\log_2 n$
10	10	6.3	4.3
20	13	8.9	5.3
30	14.7	10.9	5.9
40	16.0	12.6	6.3
50	16.9	14.1	6.6
75	18.7	17.3	7.2
100	20.0	20.0	7.6
150	21.7	24.4	8.2
200	23.0	28.2	8.6
300	24.7	34.6	9.2

V.3.2. CASOS ABERRANTES.

Procurar cerciorarse de cuáles son los hechos que describen un conjunto de datos es un paso fundamental que en buena parte evita equívocos o falacias respecto a la interpretación de los mismos. Para ello conviene cuestionar que es un caso extremo (aberrante, outliers).

Básicamente, un caso aberrante es una observación (o un conjunto de observaciones) que se desvia de manera notoria respecto a las demás observaciones que componen a un conjunto de datos. Para el caso de un conjunto univariado (que sólo registra una variable), el conjunto de datos se define por:

$$y_1, y_2, y_3, \cdots y_n$$

que al ordenarse adquiere la forma:

$$y_{(1)}, y_{(2)}, y_{(3)}, \cdots y_{(n)}$$

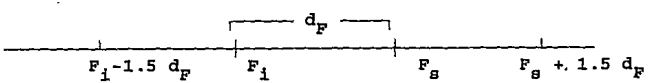
A partir de esta definición, los casos aberrantes típicamente aparecen en los límites de la distribución de y_r, como puede ser:

$$y_{(1)}, y_{(n)}, \text{ o el par } [y_{(n-1)}, y_{(n)}].$$

Cabe destacar que por definición, todo conjunto ordenado de datos posee extremos en su distribución. Sin embargo, ello no implica que se deba reaccionar automáticamente ante la presencia de los extremos de la distribución. Con cautela y empleando el criterio se debe de inspeccionar aquellos valores numéricos (altos, bajos) que notoria y visualmente se aparten de la distribución en que se concentra la mayoría de los casos.

Para identificar casos o valores aberrantes en un conjunto de datos es necesario poseer una medida que sea insensible a estos valores. Más aún, es primordial que éste índice estadístico se ocupe principalmente de analizar el comportamiento de la proporción central de los datos más que el comportamiento de los extremos. La medida que cumple con tales requisitos es el índice de dispersión cuarta (d_F), que es la diferencia entre el tercer y primer cuartil. Se dice que un dato y_i es extremo o aberrante si:

$$y_i \notin \left(F_i - 1.5\, d_F,\ F_s + 1.5\, d_F \right)$$



Es muy importante la identificación de estos valores ya que esto permite, entre otras cosas, detectar si estos casos son causa de una mala medición, transcripción o simplemente parte de la variación natural del fenómeno estudiado, se recomienda que se haga un análisis con y sin estos datos con el fin de determinar el "peso" de éstos en el resto de la información.

V.3.3. DIAGRAMAS DE CAJA.

Los diagramas de caja son una herramienta gráfica que se utilizan cuando se quieren hacer comparaciones entre grupos de una muestra o entre muestras de varias poblaciones. Permiten visualizar el comportamiento (concentración, dispersión, valores extremos, valores aberrantes, tendencias) de los subgrupos, la manera de construirlos es a través de la ayuda de los cuadros resumen: mediana, primer y tercer cuartil, rango intercuartílico, mínimo y máximo, ubican en forma gráfica a la mediana, primer y tercer cuartil así como las "cercas o bigotes" de las cajas, permiten, (cuando las escalas de medición son discretas, continuas y de razón) detectar tendencias. etc.

En el eje de las absisas se ponen los subgrupos (por sexo, por edades, por regiones, por niveles socioeconómicos, etc) a comparar y en el eje de las ordenadas el primer, segundo, tercer cuartil, el valor de los bigotes se determina de la siguiente forma:

el bigote inferior es:

$$C_i \text{ más cercano a } (F_i - 1.5\ d_F);\ C_i \in (d_F - 1.5\ d_F)$$

el bigote superior es el valor:

$$C_s \text{ más cercano a } (F_s + 1.5\ d_F);\ C_s \in (d_F + 1.5\ d_F)$$

V.4. ASOCIACION ENTRE VARIABLES.

Muchas de las hipótesis que se formulan en una investigación con el fin de buscar su verificación empírica, adoptan la forma de presuntas asociaciones entre dos o más variables, esta asociación se conoce como correlación; para poder afirmar que "existe asociación entre las variables X y Y " es necesario plantear previamente una hipótesis para, que posteriormente, se acepte o se rechace mediante una prueba estadística, algunas de las hipótesis especifican la naturaleza de esa asociación y dan alguna indicación de su magnitud.

La correlación es una medida de asociación entre variables, puede ser lineal o no lineal, se tienen diferentes medidas de correlación y éstas dependen del tipo de variables explicativas y del tipo de variables de respuesta, las técnicas Estadísticas a emplearse para detectar, cuantificar y probar significancia dependerán de los supuestos distribucionales (TÉCNICAS PARAMETRICAS Y NO PARAMÉTRICAS) que de las muestras se hagan.

Antes de establecer la posible relación entre dos o más variables, es necesario justificar teóricamente dicha sociación y no dar todo el peso a la prueba estadística. Sería aberrante decir que la variable que mide la contaminación en la ciudad de México esta significativamente correlacionada con la variable que mide el índice de natalidad de la misma región.

Dada la definición heurística de lo que es un coeficiente de asociación y hecha la consideración anterior, se recomienda como primer paso, graficar a las variables en estudio.

Cuando se tienen p-variables de respuesta, al menos ordinales, lo más común es graficar a pares, es decir, la variable 1 con la variable 2, la variable 1 con la 3,; etc. las gráficas de cada par de darán una visión del comportamiento de una con respecto a la otra.

Cuando las variables son de tipo nominal se pueden hacer tablas resumen, en las que se indique cuántos casos están en cada categoría y su respectivo porcentaje.

A continuación se darán algunos ejemplos de coeficientes de asociación.

V.4.1. CORRELACION DE PEARSON.

En el caso de las técnicas paramétricas es muy común utilizar el coeficiente de correlación de Pearson que mide la relación lineal entre dos variables de respuesta de tipo continuo (intervalo y de razón) y a diferencia de otros, es un coeficiente acotado e independiente de las unidades de medición.

Se designará a "r" como el coeficiente de correlación de Pearson y la forma de calcularlo es la siguiente:

i) con valores estándarizados

$$r = \frac{\sum_{i=1}^{n} \left(Z_{x_i} \; Z_{y_i} \right)}{n}$$

Donde

$$Z_{x_i} = \frac{X_i - \bar{X}}{S_x}$$

$$Z_{y_i} = \frac{Y_i - \bar{Y}}{S_y}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} \quad ; \quad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}} \quad ; \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

ii) con valores sin estándarizar

$$r = \frac{Cov(X,Y)}{Sx\ Sy} = \frac{n\sum_{i=1}^{n} X_i\ Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$$

Como ya se mencionó anteriormente r está acotado entre 1 y -1, es decir :

$$-1 \leq r \leq 1$$

La interpretación de r es la siguiente:

Cuando el valor de r tiende a 1 es posible que exista una relación lineal directa entre el par de variables.

En el caso de que r tiende a -1 es posible que exista una relación lineal inversa entre las variables.

La posibilidad de la relación, debe de ser confirmada mediante la prueba de hipótesis.

Si r tiende a estar muy cerca del cero, implica que no existe una relación lineal entre las variables (que no excluye la posibilidad de una relación no lineal) que no necesariamente se debe de interpretar como independencia entre variables.

PRUEBA DE SIGNIFICANCIA.

La prueba estadística más común, para la significancia del coeficiente de correlación muestral r es para determinar cuando ρ es diferente de 0, es decir, ¿Están las variables X y Y correlacionadas?

$H_0: \rho = 0 \quad H_1: \rho \neq 0$
$H_0: \rho \geq 0 \quad H_1: \rho < 0$
$H_0: \rho \leq 0 \quad H_1: \rho > 0$

La estadística de prueba es:

$$t = \frac{r - \rho}{\sqrt{\frac{\left(1 - r^2\right)}{\left(n - 2\right)}}} \quad ; \text{ con n-2 grados de libertad}$$

Los límites de confianza de ρ son:

$$P\left[Z_r - Z_{1-\alpha/2} \frac{1}{\sqrt{n - 3}} < Z_\rho < Z_r + Z_{1-\alpha/2} \frac{1}{\sqrt{n - 3}} \right] = 1 - \alpha$$

en donde $Z_r = \frac{1}{2} \ln \left| \frac{1 + r}{1 - r} \right|$ y $Z_{1-\alpha/2}$ es el valor de la variable Normal estandar.

Ejemplo:

En una muestra de familias de migrantes se obtuvo la escolaridad y fecundidad de las madres y se obtuvieron los siguientes datos:

Escolaridad años de estudio	Promedio de embarazos	Desviación estándar	Número de casos
0	6.9057	3.4377	53
1	6.8462	3.0509	13
2	5.9231	3.9678	13
3	5.5714	3.2071	7
4	4.5000	2.1679	6
5	3.8750	1.7269	8
6	3.4615	1.9415	13
7	3.3750	2.1998	8

Escolaridad años de estudio	Promedio de hijos	desviación estándar	número de casos
0	5.7736	3.0233	53
1	5.7692	2.8622	13
2	4.6923	2.7503	13
3	4.8571	3.0237	7
4	3.6667	1.6330	6
5	3.5000	1.7728	8
6	3.1538	1.7723	13
7	3.0000	1.7728	8

En el siguiente cuadro se puede observar que hay una correlación inversa entre escolaridad y fecundidad, es decir que a mayor escolaridad menor número de embarazos y de hijos vivos. Para determinar si esta correlaciòn es significativa Estadísticamente se calculó el coeficiente de correlación de Pearson entre estas variables. Se presentan los resultados en el Cuadro siguiente:

Coeficiente de Correlación entre la escolaridad, el número total de embarazos y de hijos vivos.

	ESCOLAR	TEMBARAZ	THVIVOS
ESCOLAR	1.0000	-.4165**	-.3805**
TEMBARAZ	-.4165**	1.0000	.9424**
THVIVOS	-.3805**	.9424**	1.0000

N de casos: 121 2-colas Signif: ** - .001

Interpretación:
Existe una relacion lineal directa entre el total de embarazos y el número de hijos vivos, una relación lineal inversa entre la escolaridad y el número total de embarazos esto con una confianza de 99 %.

V.4.2. COEFICIENTES DE ASOCIACIÓN PARA VARIABLES NOMINALES Y ORDINALES.

Para el caso de muestras en las que no se hacen supuestos distribucionales, se emplean medidas de correlación no paramétrica. Los tipos de variables que estén involucrados darán una idea de las técnicas a utilizar, algunos ejemplos de éstas son:

i) Coeficiente de contingencia
ii) Coeficiente de correlación de rangos de Kendall
iii) Coefiente de concordancia de Kendall
iv) Coeficiente de correlación de rangos de Spearman.

V.4.2.1. COEFICIENTE DE CORRELACION POR RANGOS DE SPEARMAN

El coeficiente de correlación de rangos de Spearman (ρ) es una medida de asociación no paramétrica y se utiliza cuando se tienen dos muestras con datos (variable de respuesta) independientes y una variable explicativa de tipo ordinal.

Como hay mediciones de dos variables para un mismo individuo se tienen n-parejas de datos (x_i, y_i)

i) Se asignan rangos de 1 hasta n a los valores de X (1 al menor, 2 al que le sigue ,... , n al mayor). Luego se asignan rangos a los valores de Y, también de menor a mayor.

ii) Se calculan las diferencias entre los rangos de X y de Y para cada pareja, se elevan al cuadrado y se suman, por último se calcula la Estadística de prueba:

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} (R_{X_i} - R_{Y_i})^2}{n(n^2 - 1)}$$

donde :
R_{X_i} representa los rangos de la muestra de la variable X

R_{Y_i} representa los rangos de la muestra de la variable Y

n es el tamaño de muestra
(la suma se realiza para los n pares de datos).

Hay que notar que este coeficiente tiene una cota inferior de -1 y una superior de +1 incluyendo al cero dentro de su variación, por lo que $-1 \leq \rho \leq 1$.

Cuando $\rho = 0$ no existe asociación entre las variables, si ρ tiende a valores cercanos a 1 entonces se interpreta como una asociación directa positiva (directamente proporcional) a diferencia de cuando los posibles valores de ρ tienden a -1 lo que implica que la relación es negativa (inversamente proporcional).

Prueba de significancia

Si se desea probar que existe una correlación positiva entre las dos variables, se utiliza una prueba de co,a superior. Para probar una correlación negativa, se utiliza una prueba de cola inferior. Para probar una desviación de independencia en cualquiera de las dos direcciones, deberá utilizarse ua prueba de dos colas. Si n es un valor entre 4 y 30 se utilizan los valores críticos siguientes:

Valores críticos de ρ para la prueba de Spearman

	Nivel de significancia de una cola, α					
	0.100	0.050	0.025	0.010	0.005	0.001
	Nivel de significancia de dos colas, α					
	0.200	0.100	0.050	0.020	0.010	0.002
n						
4	1.000	1.000				
5	0.800	0.900	1.000	1.000		
6	0.657	0.829	0.886	0.943	1.000	
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.768	0.727	0.818
13	0.385	0.484	0.560	0.648	0.703	0.791
14	0.367	0.464	0.538	0.626	0.679	0.771
15	0.354	0.446	0.521	0.604	0.657	0.750
16	0.341	0.429	0.503	0.585	0.635	0.729
17	0.328	0.414	0.488	0.566	0.618	0.711
18	0.317	0.401	0.474	0.550	0.600	0.692
19	0.309	0.391	0.460	0.535	0.584	0.675
20	0.299	0.380	0.447	0.522	0.570	0.660
21	0.292	0.370	0.436	0.509	0.566	0.647
22	0.284	0.361	0.425	0.497	0.544	0.633
23	0.278	0.353	0.416	0.486	0.532	0.620
24	0.271	0.344	0.407	0.476	0.521	0.608
25	0.265	0.337	0.398	0.466	0.511	0.597
26	0.259	0.331	0.390	0.457	0.501	0.586
27	0.255	0.324	0.383	0.449	0.492	0.576
28	0.250	0.318	0.375	0.441	0.483	0.567
29	0.245	0.312	0.369	0.433	0.475	0.557
30	0.240	0.306	0.362	0.426	0.467	0.548

si n es mayor a 30 entonces se calcula Z;

$Z = \rho \sqrt{n-1}$ que es una aproximación a la distribución Normal Estandar por lo que los valores críticos de comparación serán los correspondientes a la tabla de la Normal Estandar.

Ejemplo:

Se realizó una entrevista a 4 jefes de familia, se preguntó cuál era su opinión sobre la situación económica despues de migrar (X) y también se les preguntó cuanto tiempo (años) pensaban quedarse en el país huesped (Y), las respuestas fueron:

individuo	X (opinión)	Y (tiempo en años)
1	buena	5
2	mala	0
3	muy mala	1
4	regular	3

La hipótesis nula a probar es la no correlación entre las variables. Como se tienen dos variables X y Y cada una de ellas con escala de medición de tipo Ordinal (se pueden ordenar los valores en forma ascendente).

Codificación de X:

muy mala = 0
mala = 1
regular = 2
buena = 3

i) asignación de rangos:

Valor de X	Rango X	Rango Y	diferencia D	D^2
X_3 =0	1	2	1-2	1
X_2 =1	2	1	2-1	1
X_1 =2	3	3	3-3	0
X_1 =3	4	4	4-4	0

$$\sum D^2 = 2$$

$$\rho = 1 - \frac{6\,(2)}{4(16-1)} = 1 - 0.2 = 0.8$$

Este valor indicaría que existe una relación directa entre la situación económica actual y el tiempo que piensan quedarse en el país huesped.

Para realizar prueba de significancia de dos colas con una $\alpha = 0.05$ se checa en la tabla de Spearman la columna para el número de casos, en este ejemplo n=4 y en la columna de una cola para $\alpha = 0.05$ y se tiene una valor de 1, como $\rho = 0.8$ se acepta la hipótesis nula.

Interpretación

A pesar de que $\rho \approx 0.8$, no existe evidencia que sea estadísticamente suficiente para aceptar que existe una correlación entre situación ecónomica y tiempo posible de estancia en el país huesped, (esto puede deberse al tamaño de la muestra, o a la validéz de la prueba, se mide de forma exacta pero no existe relación entre lo medido y la hipótesis).

VI. EJEMPLOS

VI.1. INTRODUCCION

En esta parte de la tesis se dan tres ejemplos, el primero lleva por título: "Papel de la estimación de las dimensiones del esqueleto en la determinación del peso recomendable para mujeres mexicanas obesas del Hospital General de México " y es un estudio sobre nutrición, se hace un análisis sobre las características antropométricas y óseas para determinar el peso recomendable de las personas, es decir, no se basa únicamente en la edad y estatura para hacer la recomendación sobre el peso sino en la estructura del esqueleto.

El segundo ejemplo toma una muestra de adolescentes de la Ciudad de México y hace un análisis de cómo van creciendo los sujetos, se calcularon las estadísticas básicas de cada una de las variables involucradas y se hace una estimación de las tendencias de crecimiento para dos submuestras.

Por último se presenta un estudio en el que se tomaron tres muestras, cada una en tres diferentes regiones del Estado de Puebla; el objetivo de la investigación es determinar la influencia de las condiciones socio económicas y los niveles de mestizaje en el crecimineto de las poblaciones en estudio.

VI.2. EJEMPLO 1

TIPO DE ESTUDIO

Es una encuesta COMPARATIVA PROSPECTIVA (Observacional, prospectiva, transversal, comparativo).

OBSERVACIONAL:
porque en el estudio el investigador solo observará el fenómeno analizado.

PROSPECTIVO:
ya que la información se toma de acuerdo a los criterios de del investigador y para los fines específicos del investigador.

TRANSVERSAL:
dado que se mide en una sola ocasión a las variables.

COMPARATIVO:
porque el estudio comparará a dos poblaciones (premenopáusicas,menopáusicas)

DEFINICION DEL PROBLEMA:

TITULO:

"Papel de la estimación de las dimensiones del esqueleto en la determinación del peso recomendable para mujeres mexicanas obesas del Hospital General de México "

OBJETIVOS:

En este trabajo se evaluará el papel que juega la estimación antropométrica de la estructura del esqueleto para ajustar el peso recomendable en mujeres mexicanas obesas adultas, con edades cercanas a la menopausia.

General.

- "Analizar el papel que juega la estimación antropométrica de la estructura del esqueleto en el ajuste del peso recomendable de mujeres mexicanas obesas adultas, con edades cercanas a la menopausia (entre 22 y 60 años), seleccionadas del Servicio de Endocrinología del Hospital General de México, entre el Io. de junio y el 31 de diciembre de 1992 y que presentan I.M.C. superior a 30 con presión sistólica controlada, igual o menor a 100 mm de mercurio".

Específicos.

- " Analizar el margen de error de las estimaciones de peso recomendable calculadas por medio del Indice de la Masa Corporal y el Peso Relativo, en relación con el porcentaje de grasa corporal real estimado por absorciometría".

- " Evaluar la correlación entre peso recomendable corregido mediante la estructura del esqueleto y peso recomendable establecido por el porcentaje de grasa corporal real estimado por absorciometría".

- " Valorar cuál de las dos técnicas de estimación antropométrica de la estructura del esqueleto (anchura de codo o EAT) se correlaciona mejor con la medida absorciométrica de tejido óseo".

- " Identificar el cambio de la correlación entre las dos técnicas de estimación antropométrica de la estructura del esqueleto con el contenido mineral de éste y la medida del tejido óseo (bone points o puntos óseos) obtenidas por absorciometría".

- " Identificar la correlación entre las dos técnicas de estimación antropométrica de la estructura del esqueleto entre sí".

HIPOTESIS.

La medida de tejido óseo obtenida por absorciometría (bone points o puntos óseos) mostrará la técnica antropométrica (EAT, IEM, anchura de codo) más conveniente para la determinación del peso recomendable, a través del coeficiente de correlación más elevado.

En el grupo de mujeres menopáusicas, de los resultados obtenidos por absorciometría, los coeficientes de correlación serán superiores con la medida de tejido óseo, que con el contenido mineral óseo correspondiente.

DEFINICION DE LA POBLACION OBJETIVO.

La población por estudiar estará constituida por un grupo de mujeres mexicanas, pacientes obesas del Hospital General de México.

CARACTERISTICAS GENERALES.

La población por estudiar estará constituida por un grupo de mujeres mexicanas, pacientes obesas del Hospital General de México, cuya obesidad se determinará con la presencia de Indices de Masa Corporal superiores a 30; que cuenten además con edades entre 22 y 60 años y que pueden presentar hipertensión arterial controlada como único padecimiento asociado a la obesidad, con presión sistólica igual o menor de 100 mm de mercurio . Este grupo se separará para su análisis en dos grupos: aquellas que continúan presentando períodos menstruales y las menopáusicas, definidas por tener cuando menos un año de no haberlos presentado, en ausencia de embarazo.

Se considera que este grupo de obesas es el más adecuado para este trabajo debido a que es difícil demostrar en una población normal la necesidad de evaluar la composición corporal para determinar el peso recomendable.

El criterio inicial de obesidad se establecerá mediante la relación entre el peso y la estatura, a través del Indice de la Masa Corporal. Para establecer el peso recomendable se emplearán las siguientes técnicas:

a) Establecer para cada paciente el peso que corresponda a un Indice de Masa Corporal de 25.

b) Establecer para la estatura de cada paciente un peso relativo equivalente a 120%.

La estructura del esqueleto se establecerá mediante:

a) La anchura del codo, clasificando a los esqueletos como grácil es, si la anchura es menor a 5.5cm; intermedios, entre 5.5 y 6.2 cm y robustos si es mayor de 6.2 cm. Estas cifras son resultado del análisis de la distribución de la anchura del codo en 672 mujeres jóvenes mexicanas y como puntos de corte se emplearon las percentilas 15 y 85.

---

b) La técnica EAT, propuesta por Katch y Freedson y que se explica en el cuerpo de este trabajo.

Cada mujer será clasificada de acuerdo a la estructura de su esqueleto mediante los dos criterios anteriores y su peso recomendable se corregirá de acuerdo con este resultado. Para ello se ha calculado que el esqueleto influye sobre el peso corporal de una mujer de referencia (1.64 m y peso de 56.81 Kg) con 6.8 Kg, que es el 12% del peso. En tanto más grácil sea el esqueleto, menor será su contribución para el peso corporal y por el contrario, en tanto más robusto sea, su contribución será mayor. Por lo tanto,a las mujeres con esqueleto grácil les será corregido el peso recomendable, disminuyéndolo un 5%. No se harán correcciones en las de esqueleto intermedio y se aumentará un 5% en las de esqueleto robusto. Estas cifras son arbitrarias, pero permitirán probar la hipótesis de que la estructura del esqueleto influye de manera significativa sobre el peso corporal.

Para verificar si las estimaciones antropométricas del peso recomendable y de la estructura del esqueleto corresponden con la realidad en las mujeres obesas, compararemos los resultados que se obtengan mediante la absorciometría por doble emisión de rayos x, que determina de manera directa el porcentaje de grasa corporal y la cantidad de tejido óseo, así como el contenido mineral del esqueleto.

Para la mujer de referencia, la grasa corporal total equivale al 27% del peso corporal. Para el tercer cálculo del peso recomendable, tomando en cuenta la determinación absorciométrica de la grasa corporal total, se supone que las mujeres deben llegar a un máximo de 27% de grasa corporal. A este peso recomendable se le considerará como aquel con el que se compararán los establecidos por el Indice de la Masa Corporal, el peso relativo y las correciones logradas en el Indice de la Masa Corporal, mediante la estructura del esqueleto. Así se establecerá la medida de correlación entre cada uno de los resultados

obtenidos y los que se deriven de la absorciometría, para estimar el margen de error en cada uno.

En vista de que con la osteoporosis de la menopausia disminuye el contenido mineral, es necesario subdividir el grupo en mujeres premenopáusicas y menopáusicas, ya que nuestra hipótesis es que los cálculos de la estructura del esqueleto se correlacionarán mejor con la medida de tejido óseo (bone points o puntos óseos) que con la de contenido mineral, así que al separarse estos grupos de mujeres, la comparación de las medidas de correlación de la estructura del esqueleto permitirá probar esta hipótesis.

CRITERIOS DE INCLUSION:

---

a) Individuos del sexo femenino, con obesidad de tipo exógeno exclusivamente.

b) Mayores de 22 y menores de 60 años de edad, por lo que pueden ser tanto premenopáusicas, como menopáusicas.

c) Individuos con H.T.A. controlada exclusivamente y ninguna otra enfermedad asociada oculta o aparente.

d) Individuos con I.M.C. igual o mayor a 30.

e) Individuos que habiten en el área metropolitana de la Ciudad de México.

f) Individuos que acepten participar en el estudio.

CRITERIOS DE EXCLUSION:

---

a) Mujeres menores de 22 años y mayores de 60.

b) Individuos que padezcan cualquier enfermedad no mencionada en los criterios de inclusión.

c) Individuos que sean sometidos a cualquier intervención quirúrgica durante el periodo de estudio.

d) Embarazadas durante el estudio.

e) Individuos que habiten fuera del Area Metropolitana de la Ciudad de México.

UBICACION ESPACIO-TEMPORAL.

Para poder ser incluídos en este estudio, los pacientes serán seleccionados a partir de los reportes de consulta externa del Servicio de Nutrición del Pabellón de Endocrinología del Hospital General de México, en los que se indica que mensualmente acuden a consulta por primera vez, en promedio, 91 pacientes, de los cuales presentan obesidad o sobrepeso 47 por mes, según revisión de los datos presentados para 1991. Estos, además, deberán cumplir con los siguientes requisitos:

FORMACION DE LA MUESTRA DE ESTUDIO.

La selección previa del paciente obeso se efectuará en la Consulta Externa del Servicio de Endocrinología, de acuerdo al I.M.C., si éste es mayor a 30.

El grupo de estudio será evaluado clínicamente con el apoyo de la segunda parte de la cédula de recolección de datos que se propone para tal fin (anexo 1) y se formarán dos grupos de estudio, debido a que se sabe actualmente (46) que el contenido mineral del esqueleto se altera con la menopausia:

- Pacientes premenopáusicas, o que continúan presentando períodos menstruales.

- Pacientes menopáusicas, o que no presentan períodos menstruales desde hace un año.

DISEÑO ESTADISTICO
ESPECIFICACION DE VARIABLES Y ESCALAS DE MEDICION

Campo: EDAD	Tipo:Numérico	Ancho: 6	Dec:0	
Campo: PESO	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: TALLA	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: T_SENTAD	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:T_SENTADO
Campo: A_MUNECA	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: A_BIACRO	Tipo:Numérico	Ancho: 9	Dec:2	Nom. Ant.:A_BIACROM
Campo: A_BICRES	Tipo:Numérico	Ancho: 9	Dec:2	Nom. Ant.:A_BICREST
Campo: A_BITRO	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: A_CODO	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: C_ANTEBR	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:C_ANTEBRA
Campo: C_BRA_RE	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:C_BRA_REL
Campo: C_BRA_FL	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:C_BRA_FLE
Campo: C_MIN_AB	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:C_MIN_ABDO
Campo: C_MAX_GL	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:C_MAX_GLUT
Campo: C_MUSLO	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: C_MUNECA	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: PLI_ANTE	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: PLI_SUBE	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:PLI_SUBES
Campo: PLI_SUPR	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:PLI_SUPRA
Campo: PLI_TRIC	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:PLI_TRICE
Campo: PLI_BICE	Tipo:Numérico	Ancho: 7	Dec:2	Nom. Ant.:PLI_BICEP
Campo: BMD	Tipo:Numérico	Ancho: 7	Dec:4	
Campo: DENSITY	Tipo:Numérico	Ancho: 7	Dec:4	
Campo: TBC	Tipo:Numérico	Ancho: 6	Dec:0	
Campo: BONE_POI	Tipo:Numérico	Ancho: 9	Dec:0	Nom. Ant.:BONE_POINT
Campo: FAT_DENS	Tipo:Numérico	Ancho: 7	Dec:2	
Campo: FAT_FOR	Tipo:Numérico	Ancho:10	Dec:4	
Campo: IMC	Tipo:Numérico	Ancho: 8	Dec:3	
Campo: IEM	Tipo:Numérico	Ancho: 8	Dec:3	
Campo: HAT	Tipo:Numérico	Ancho:11	Dec:5	
Campo: PESOPROM	Tipo:Numérico	Ancho:10	Dec:2	

ESCALAS DE MEDICION.
peso(kg)
TALLA(cm)
TALLA SENTADO(cm)
ANCHURA BIACROMIAL(cm)
ANCHURA BICRESTAL(cm)
ANCHURA BICROCANTERICA*

ANCHURA del CODO (cm)
CIRCUNF. MEDIA del
brazo FLEXIONADO (cm)
CIRCUNF. MINIMA
del abdomen (cm)
CIRCUNF. MAXIMA DE
LOS GLUTEOS (cm)
CIRCUNF. MUSLO (cm).
CIRCUNF. MUÑECA (cm)
PLIEGUE SUPRAILIACO(mm)
PLIEGUE del TRICEPS(mm)
PLIEGUE del BICEPS (mm)

PROCESO DE CAPTACION DE LA INFORMACION

CEDULA DE RECOLECCION DE DATOS.

ANTROPOMETRIA - COMPOSICION CORPORAL
CARNET:_____
NOMBRE:_____ EDAD (años y meses)_____
SEXO: _____ PADECIMIENTO(S):_____

FECHA
PESO(kg)
TALLA(cm)
TALLA SENTADO(cm)
ANCHURA BIACROMIAL(cm)
ANCHURA BICRESTAL(cm)
ANCHURA BICROCANTERICA
ANCHURA DEL CODO (cm)
CIRCUNF. MEDIA DEL BRAZO FLEXIONADO (cm)
CIRCUNF. MINIMA DEL ABDOMEN (cm)
CIRCUNF. MAXIMA DE LOS GLUTEOS (cm)
CIRCUNF. MUSLO (cm).
CIRCUNF. MUÑECA (cm)
PLIEGUE SUPRAILIACO(mm)
PLIEGUE del TRICEPS(mm)
PLIEGUE del BICEPS (mm)

ANALISIS E INTERPRETACION DE LA INFORMACION.

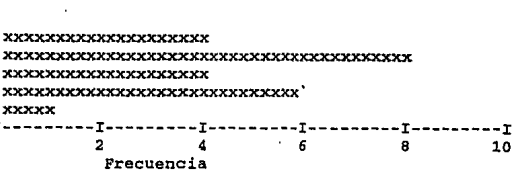
Medidas de tendencia central y medidas de dispersión.

Estadísticas descriptivas.

Variable	Media	Desv. Estandar	Varianza	Mínimo	Máximo	n
Edad	40.83	11.09	123.06	22	60	23
Peso	81.46	12.71	161.59	60.40	106.00	23
Talla	152.69	6.37	40.60	141.50	162.30	23
T. sentado	79.62	4.37	19.12	67.20	88.00	21
A. biacrom.	33.39	1.69	2.87	30.00	36.00	23
A. bicrestal	30.28	2.19	4.82	26.00	34.80	23
A. bitroc.	31.95	2.21	4.88	29.00	36.40	23
A. codo	6.12	.33	.11	5.30	6.70	23
A. muñeca	5.06	.28	.08	4.50	5.90	22
C. abdomen	102.69	12.32	151.69	85.30	129.00	23
C. glúteos	110.63	10.97	120.32	91.00	139.00	23
C. brazo rel.	35.22	3.75	14.10	29.80	43.30	23
C. brazo flex.	35.01	3.69	13.60	30.30	45.00	23
C. antebrazo	24.56	3.16	10.00	21.50	34.60	20
C. muñeca	16.79	1.37	1.89	14.70	20.20	23
C. muslo	64.33	6.04	36.52	54.30	77.90	23
P. tricip.	35.00	7.68	58.91	26.00	51.00	23
P. biceps	24.13	9.02	81.32	10.00	41.00	23
P. subescap.	49.18	8.08	65.23	30.60	63.00	23
P. suapril.	44.24	9.87	97.47	23.50	58.50	23
P. antebrazo	15.78	13.23	175.01	5.00	61.00	19
IEM	158.66	23.40	547.68	125.78	213.79	23
IMC	34.90	4.81	23.11	27.90	46.12	23
DMC	1.16	.09	.01	.98	1.31	23
Densidad	1.01	.00	.00	.99	1.01	23
Cont.Total Hueso	915.61	126.82	16083.07	660.00	1124.00	23
EAT	1646.76	67.01	4490.84	1531.01	1751.60	23
Puntos óseos	4510.04	381.61	145628.41	3850	5145	23
Grasa por dens.	45.81	4.30	18.51	37.50	52.50	23
Grasa por fórm.	41.81	2.34	5.48	37.0700	45.5500	23

PESO
TABLA DE FRECUENCIAS

Intervalo	Valor	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
(60.00 hasta 70)	1.00	4	17.4	17.4	17.4
(70.01 hasta 80)	2.00	8	34.8	34.8	52.2
(80.01 hasta 90)	3.00	4	17.4	17.4	69.6
(90.01 hasta 100)	4.00	6	26.1	26.1	95.7
(100.01 hasta 110)	5.00	1	4.3	4.3	100.0
	TOTAL	23	100.0	100.0	



Percentiles					
5.0000	10.0000	25.0000	50.0000	75.0000	90.0000
95.0000					
61.2200	65.6000	72.2000	77.3000	94.2000	99.0000
104.7800					

Prueba t

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	12	84.6125	8.926	2.577
Group 2	11	78.0136	15.584	4.699

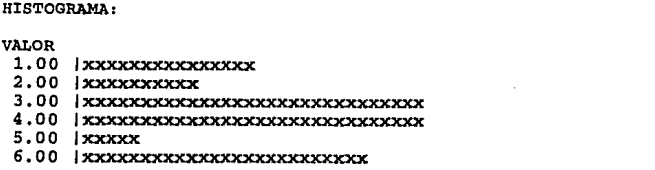
Pooled Variance Estimate				Separate Variance Estimate			
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
3.05	.081	1.26	21	.221	1.23	15.63	.236

El 52% de la muestra está entre 60 y 80 kg. mientras que el resto se encuentra entre 80 y 106 kg. La correlación entre el peso y anchura biacromial, anchura bicrestal, anchura bitrocantérica, circunferencia del brazo relajado, circunferencia de los glúteos, circunferencia del abdomen, circunferencia del muslo, grasa obtenida por densidad, grasa obtenida por fórmula, IEM, IMC, pliegue subescapular, pliegue tricipital es significativamente positiva, mientras que con densidad y edad están inversamente relacionados (ver cuadro de correlaciones). No hay diferencias entre estos dos grupos.

TALLA
TABLA DE FRECUENCIAS.

Intervalo	Valor	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
(140.00 hasta 144.00)	1.00	3	13.0	13.0	13.0
(144.01 hasta 148.00)	2.00	2	8.7	8.7	21.7
(148.01 hasta 152.00)	3.00	6	26.1	26.1	47.8
(152.01 HASTA 156.00)	4.00	6	26.1	26.1	73.9
(156.01 HASTA 160.00)	5.00	1	4.3	4.3	78.3
(160.01 HASTA 164.00)	6.00	5	21.7	21.7	100.0
	TOTAL	23	100.0	100.0	

HISTOGRAMA:



Percentiles

5.0000	10.0000	25.0000	50.0000	75.0000	90.0000	95.0000
141.7400	142.8200	148.5000	152.1000	159.6000	161.7600	162.2800

Prueba t

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	12	156.0333	4.820	1.391
Group 2	11	149.0455	5.979	1.803

Pooled Variance Estimate				Separate Variance Estimate			
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.54	.489	3.10	21	.005	3.07	19.25	.006

La talla de la muestra se presenta en un 75% entre los valores 148 y 164 cm. Por otra parte está variable está inversamente relacionada con la edad y proporcionalmente relacionada con la variable estatura acromial y trocantérica (ver cuadro de correlaciones). La prueba t indica que los dos grupos son diferentes con un nivel de significancia de 0.05.

TABLA DE CORRELACIONES

	A_BIACRO	A_BICRES	A_BITRO	A_CODO	A_MUNECA	BMD
A_BIACRO	1.0000	.4379	.5550	.2577	.2508	.3511
A_BICRES	.4379	1.0000	.8166**	.1084	-.0117	.256
A_BITRO	.5550	.8166**	1.0000	-.0568	-.0809	.4922
A_CODO	.2577	.1084	-.0568	1.0000	.2358	.0179
A_MUNECA	.2508	-.0117	-.0809	.2358	1.0000	-.4387
BMD	.3511	.2560	.4922	.0179	-.4387	1.000
BONE_POI	.0070	.2957	.2780	.2674	-.2573	.5789
C_ANTEBR	.2059	.5383	.5291	.2148	.0828	.086
C_BRA_FL	.1922	.1791	.0565	.3922	.3938	-.1634
C_BRA_RE	.2065	.3181	.2424	.2991	.2603	.021
C_MAX_GL	.4179	.6295*	.5474	.2785	.3680	.1037
C_MIN_AB	.5479	.5823	.6589*	.1768	.4160	.3072
C_MUNECA	.2750	.2066	.0454	.3792	.5468	-.3546
C_MUSLO	.2401	.5869	.6184*	.2071	-.1126	.3956
DENSITY	-.5605	-.3751	-.5474	.0357	-.1375	-.5062
EDAD	.0198	-.4744	-.2980	.0147	.1310	-.2211
FAT_DENS	.2446	.5687	.4890	.3034	.0646	.3132
FAT_FOR	.5598	.3743	.5475	-.0340	.1388	.5066
HAT	-.0977	.3024	.0944	.3493	-.1374	.2124
IEM	.6498*	.6803*	.7342**	.2486	.4032	.2562
IMC	.6922*	.6543*	.7248**	.1991	.4088	.2434
PESO	.6217*	.7518**	.7258**	.3688	.3156	.3494
PLI_ANTE	-.3519	.1084	.0096	.2143	.1292	-.3665
PLI_BICE	.3482	.4191	.5588	-.0374	.0264	.3565
PLI_SUBE	.3427	.4299	.5500	.2543	.2373	.483
PLI_SUPR	.4673	.2358	.2607	-.1679	-.0157	.2998
PLI_TRIC	.4573	.2300	.4380	.1708	.1898	.4645
TALLA	-.1418	.2647	.0480	.3439	-.1414	.1874
TBC	.2221	.3299	.4451	.1846	-.3657	.8788**
T_SENTAD	-.0323	.3364	.2395	.0875	-.0734	.0717

Número de casos:18
Prueba de dos colas, significancia: * - .01 ** -.001

RECURSOS.

RECURSOS HUMANOS:

Pasante de Lic. en Dietética y Nutrición de la E.D.N. del I.S.S.S.T.E.

Personal capacitado en el manejo del absorciómetro, por parte del Hospital de México y la U.N.A.M.

RECURSOS FISICOS:

Báscula clínica tipo romano.
Calibrador de panículo adiposo, marca C.S.T.

Cinta métrica flexible de fibra de vidrio, marca Rollfix, graduada en milímetros.

Antropómetro tipo Martin, marca Sieber y Hegner.

Absorciómetro de doble emisión de rayos x, tipo "Hologic QDR-1000

TM X-ray Bone Densitometer".

BIBLIOGRAFIA.

1.- Parizková J: Body fat and physical fitness. The Hague. Martinus Nijoff, 1977.

2.- Benke AR, Wilmore JH: Evaluation and regulation of body build and composition. 1a. ed. New York. Prentice Hall, 1977.

3.- Feilig P y col.: Endocrinología y metabolismo. 3a.ed. México. Mc Graw Hill, 1983.

4.- Ruderman N y col.: The metabolically obese, normal weight individual. Am.J.Clin.Nut, 1989; 49: 745 - 751.

VI.3. EJEMPLO 2

TIPO DE ESTUDIO

Es una encuesta DESCRIPTIVA PROSPECTIVA (observacional, prosepectiva, transversal, descriptivo)

OBSERVACIONAL:
Porque en el estudio el investigador solo observara el fenómeno analizado

PROSPECTIVO:
Ya que la información se toma de acuerdo a los criterios del investigador y para los fines específicos del investigador.

TRANSVERSAL:
Dado que se mide en una sola ocasión a las variables.

DESCRIPTIVO:
Porque el estudio describe el "crecimiento" de los subadultos.

DEFINICION DEL PROBLEMA.

TITULO:

"Terminando de crecer en México".

OBJETIVOS:

Describir la "forma" de crecer de los subadultos (edades de 14.5 a 18.5 años) en México.

DEFINICION DE LA POBLACION OBJETIVO.

Subadultos (14.5-18.5 años de edad) del D.F. en los años de 1990-1991.

FORMACION DE LA MUESTRA DE ESTUDIO

Para el presente estudio se estudiaron 1190 hombres y 1110 mujeres entre 14.5 y 18.5 años de edad de procedencia de toda la república (dividida en zonas geográficas) sin embargo el 88% de la muestra proviene del D.F.

Estas personas asistieron a la muestra de "Ciencia y Deporte" en el Museo de Ciencias y Artes de la Universidad Nacional Autónoma de México.

Cuadro 13

PESO (Kg.) EN RELACION CON LA ESTATURA (cm). Hombres

Estatura (cm.)	n	Media Peso (Kg.)	s desv. estd.	Peso estimado (Kg.)	Estatura (cm.)	n	Media Peso (Kg.)	s desv. estd.	Peso estimado (Kg.)
148	2	47.12	7.88	45.89	167	96	58.66	7.52	59.84
149	1	41.85	----	46.62	168	84	60.15	7.67	60.57
150	6	49.46	9.63	47.36	169	62	61.63	7.54	61.31
151	2	51.15	3.15	48.09	170	76	63.62	7.27	62.04
152	6	46.31	2.88	48.83	171	64	62.14	8.32	62.78
153	2	52.20	13.25	49.56	172	44	63.31	8.46	63.51
154	14	50.51	8.05	50.29	173	51	64.99	9.03	64.24
155	12	50.07	8.57	51.03	174	31	66.30	7.63	64.98
156	17	51.04	6.89	51.76	175	23	66.35	10.77	65.71
157	18	49.49	4.35	52.50	176	27	65.96	8.12	66.45
158	23	53.59	6.50	53.23	177	25	70.47	8.49	67.18
159	29	52.16	6.57	53.97	178	22	68.32	5.93	67.91
160	37	53.24	6.57	54.70	179	21	70.29	9.58	68.65
161	52	54.54	6.51	55.43	180	9	72.10	10.77	69.38
162	52	56.20	7.27	56.17	181	14	73.87	9.05	70.12
163	62	58.49	10.03	56.90	182	5	73.67	2.85	70.85
164	64	57.60	7.11	57.64	183	5	67.61	6.53	71.59
165	65	58.08	7.31	58.37	184	4	66.31	3.20	72.32
166	63	57.78	5.70	59.10					

Ecuación de regresión:

Y=(-62.7694) + (0.734181 * estatura)

Cuadro 14

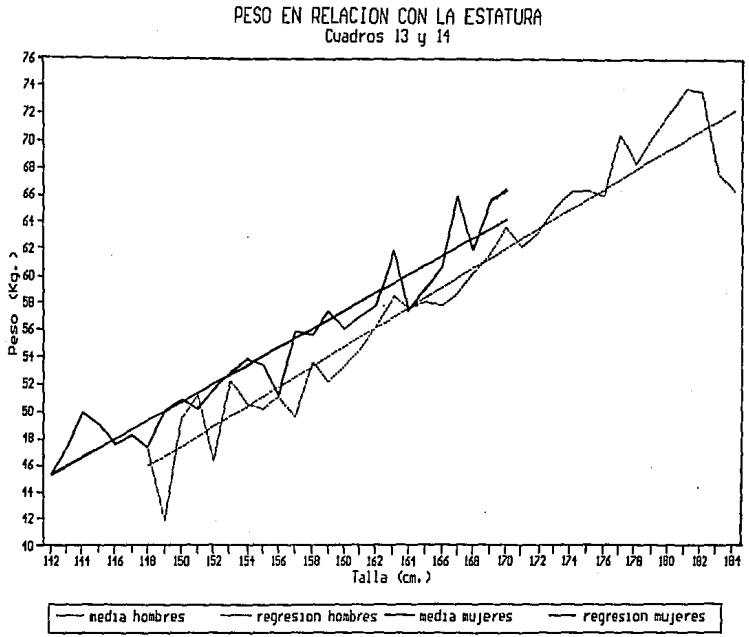
PESO (Kg.) EN RELACION CON LA ESTATURA (cm). Mujeres

Estatura (cm.)	n	Media Peso (Kg.)	s desv. estd.	Peso estimado (Kg.)	Estatura (cm.)	n	Media Peso (Kg.)	s desv. estd.	Peso estimado (Kg.)
142	5	45.25	1.70	45.24	157	74	55.83	7.48	55.38
143	7	47.21	3.79	45.92	158	62	55.66	5.96	56.06
144	4	49.85	7.25	46.59	159	53	57.32	8.23	56.73
145	16	49.00	6.46	47.27	160	46	56.00	5.94	57.41
146	20	47.51	8.48	47.94	161	47	56.95	7.57	58.09
147	21	48.17	7.25	48.62	162	38	57.80	8.33	58.76
148	32	47.32	5.85	49.30	163	33	61.90	7.35	59.44
149	32	50.01	5.28	49.97	164	28	57.46	7.39	60.12
150	50	50.80	4.48	50.65	165	29	58.97	7.85	60.79
151	50	50.09	4.94	51.32	166	11	60.63	9.88	61.47
152	57	51.63	6.02	52.00	167	7	65.88	14.08	62.14
153	50	52.83	5.80	52.68	168	13	61.97	9.08	62.82
154	72	53.79	6.24	53.35	169	10	65.62	8.70	63.50
155	77	53.42	7.27	54.03	170	2	66.40	13.65	64.17
156	63	51.20	5.92	54.71					

Ecuación de regresión:

Y=(-50.7824) + (0.676207 * estatura)

Figura 9


PESO EN RELACION CON LA ESTATURA
Cuadros 13 y 14
Peso (Kg.)
Talla (cm.)
media hombres
regresion hombres
media mujeres
regresion mujeres

Cuadro 51

LONGITUD DE LA MANO (cm). Hombres

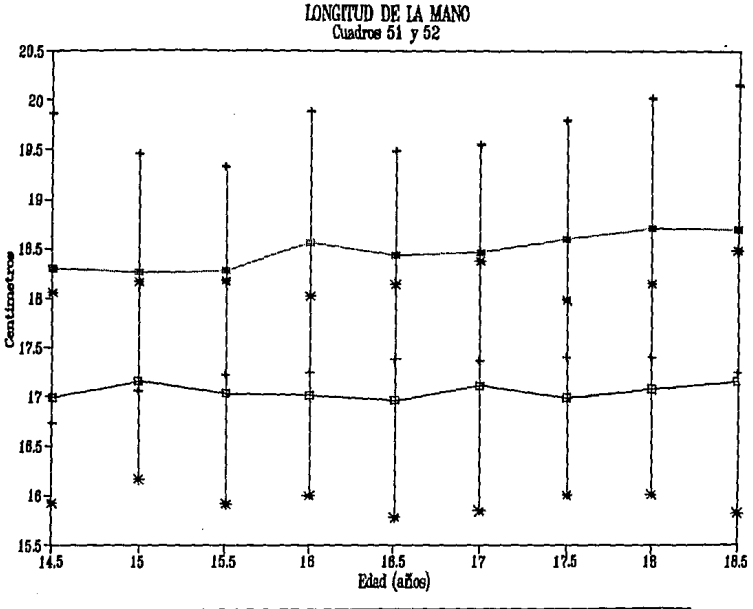
edad años	n	desv. estd. s	M-3s	M-2s	M-s	media M	M+s	M+2s	M+3s	mediana med.
14.50	125	1.57	13.59	15.16	16.73	18.30	19.87	21.44	23.01	18.30
15.00	131	1.20	14.67	15.87	17.07	18.26	19.46	20.66	21.86	18.20
15.50	121	1.06	15.12	16.17	17.23	18.28	19.34	20.40	21.45	18.30
16.00	168	1.32	14.60	15.93	17.25	18.58	19.90	21.23	22.55	18.50
16.50	156	1.05	15.28	16.34	17.39	18.45	19.50	20.56	21.61	18.45
17.00	121	1.10	15.18	16.28	17.38	18.47	19.57	20.66	21.76	18.40
17.50	137	1.20	15.02	16.22	17.42	18.61	19.81	21.01	22.20	18.60
18.00	137	1.31	14.78	16.09	17.41	18.72	20.03	21.35	22.66	18.90
18.50	71	1.46	14.34	15.80	17.25	18.71	20.17	21.62	23.08	18.70

Cuadro 52

LONGITUD DE LA MANO (cm). Mujeres

edad años	n	desv. estd. s	M-3s	M-2s	M-s	media M	M+s	M+2s	M+3s	mediana med.
14.50	96	1.06	13.80	14.86	15.93	16.99	18.06	19.12	20.18	17.10
15.00	107	1.00	14.16	15.16	16.17	17.17	18.17	19.17	20.17	17.10
15.50	135	1.13	13.64	14.77	15.91	17.04	18.17	19.31	20.44	17.00
16.00	139	1.02	13.97	14.98	16.00	17.02	18.04	19.05	20.07	17.00
16.50	141	1.19	13.39	14.58	15.77	16.96	18.15	19.34	20.53	16.90
17.00	125	1.26	13.33	14.59	15.85	17.12	18.38	19.65	20.91	17.10
17.50	116	0.99	14.01	15.01	16.00	16.99	17.99	18.98	19.97	17.00
18.00	84	1.07	13.88	14.95	16.02	17.09	18.16	19.23	20.30	16.95
18.50	57	1.33	13.17	14.50	15.83	17.16	18.49	19.82	21.15	17.10

Figura 26


LONGITUD DE LA MANO
Cuadros 51 y 52
20.5
20
19.5
19
18.5
18
17.5
17
16.5
16
15.5
Centímetros
14.5
15
15.5
16
16.5
17
17.5
18
18.5
Edad (años)
media hombres
desv. estd. hombres
media mujeres
desv. estd. mujeres

Cuadro 55

ANCHURA DE LOS HOMBROS (diam. biacromial) (cm). Mujeres

edad años	n	desv. estd. s	media M-3s	M-2s	M-s	M	M+s	M+2s	M+3s	mediana med.
14.50	96	1.96	28.48	30.44	32.39	34.35	36.31	38.27	40.23	34.70
15.00	108	1.56	29.98	31.55	33.11	34.67	36.23	37.79	39.35	34.85
15.50	135	1.96	28.57	30.53	32.49	34.46	36.42	38.38	40.34	34.60
16.00	139	1.95	28.81	30.76	32.71	34.66	36.60	38.55	40.50	34.60
16.50	141	2.44	27.06	29.50	31.93	34.37	36.81	39.24	41.68	34.60
17.00	126	2.07	29.07	31.14	33.20	35.27	37.34	39.40	41.47	35.20
17.50	117	1.54	30.12	31.67	33.21	34.75	36.30	37.84	39.38	34.70
18.00	90	1.93	29.30	31.22	33.15	35.07	37.00	38.93	40.85	35.30
18.50	62	2.05	28.43	30.48	32.53	34.58	36.63	38.68	40.73	34.55

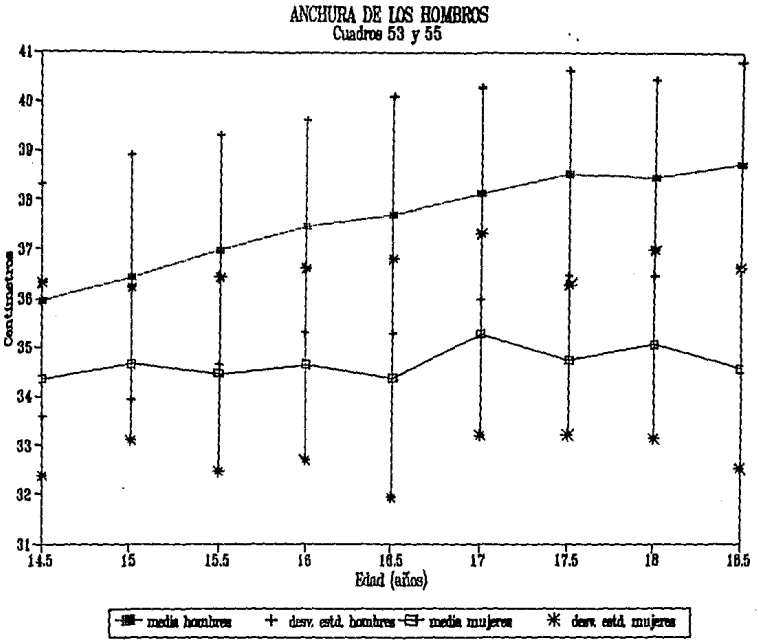
Cuadro 56

ANCHURA DE LOS HOMBROS (diam. biacromial) (cm). Mujeres

Escala percentilar

edad años	n	3	10	25	50	75	90	97
14.50	96	31.30	32.00	33.35	34.70	35.70	36.45	37.30
15.00	108	31.80	32.80	33.50	34.85	35.60	36.50	37.55
15.50	135	30.90	32.40	33.60	34.60	35.80	36.60	37.40
16.00	139	31.60	32.40	33.30	34.60	36.00	37.50	38.00
16.50	141	29.90	32.10	33.10	34.60	36.10	37.00	37.60
17.00	126	31.60	32.85	34.05	35.20	36.50	37.35	40.00
17.50	117	31.90	32.90	33.70	34.70	35.80	37.00	38.10
18.00	90	31.25	33.05	34.10	35.30	36.40	37.30	38.45
18.50	62	28.50	32.55	33.50	34.55	36.20	37.15	40.15

Figura 27


ANCHURA DE LOS HOMBROS
Cuadros 53 y 55
Centímetros
41
40
39
38
37
36
35
34
33
32
31
14.5
15
15.5
16
16.5
17
17.5
18
18.5
Edad (años)
media hombres
desv. estd. hombres
media mujeres
desv. estd. mujeres

Cuadro 95

AREA GRASA DEL BRAZO (cm^2) Hombres

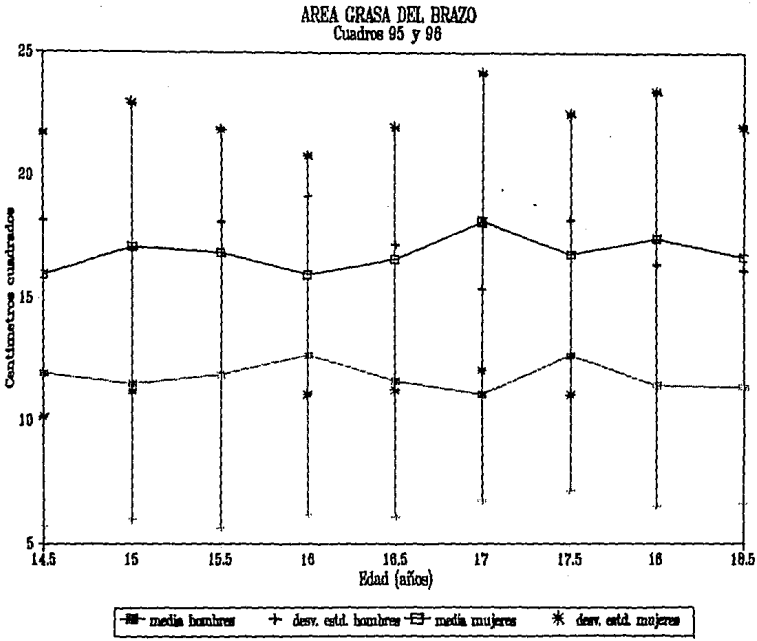
edad años	n	desv. estd. s	M-2s	M-s	M	media M+s	M+2s	mediana med.
14.50	127	6.23	-0.54	5.69	11.92	18.15	24.38	10.80
15.00	132	5.55	0.38	5.93	11.47	17.02	22.57	9.97
15.50	121	6.24	-0.63	5.60	11.84	18.07	24.30	10.04
16.00	171	6.50	-0.38	6.12	12.62	19.12	25.62	10.59
16.50	162	5.54	0.52	6.06	11.60	17.14	22.68	10.26
17.00	120	4.31	2.42	6.73	11.04	15.34	19.65	10.53
17.50	143	5.52	1.61	7.13	12.64	18.16	23.67	11.46
18.00	140	4.92	1.58	6.50	11.42	16.34	21.26	10.28
18.50	80	4.75	1.86	6.61	11.36	16.10	20.85	11.00

Cuadro 96

AREA GRASA DEL BRAZO (cm^2). Mujeres

edad años	n	desv. estd. s	M-2s	M-s	media M	M+s	M+2s	mediana med.
14.50	96	5.78	4.40	10.18	15.96	21.74	27.52	14.67
15.00	108	5.87	5.34	11.20	17.07	22.93	28.80	16.31
15.50	135	5.03	6.75	11.78	16.81	21.83	26.86	16.22
16.00	139	4.85	6.21	11.06	15.90	20.75	25.60	15.91
16.50	141	5.37	5.86	11.22	16.59	21.95	27.31	16.00
17.00	126	6.06	5.97	12.03	18.10	24.16	30.22	16.65
17.50	117	5.73	5.31	11.04	16.76	22.49	28.22	15.71
18.00	90	6.01	5.39	11.40	17.40	23.41	29.42	15.74
18.50	62	5.28	6.08	11.37	16.65	21.94	27.22	15.98

Figura 46


AREA GRASA DEL BRAZO
Cuadros 95 y 96
Centimetros cuadrados
25
20
15
10
5
14.5
15
15.5
16
16.5
17
17.5
18
18.5
Edad (años)
media hombres
desv. estd. hombres
media mujeres
desv. estd. mujeres

VI.4. EJEMPLO 3

Antecedentes:

Uno de los propósitos de la Antropología Física es el conocimiento de la variabilidad Biológicas de las poblaciones humanas. En ese contexto, se ha pretendido investigar las influencias tanto del medio ambiente como aquellas derivadas del patrimonio genético de las poblaciones. Este tipo de estudios se ha realizado especialmente en población infantil, ya que durante el crecimiento físico, el cuerpo humano es súmamente sensible a los factores del medio ambiente; tales como el estado de salud, condiciones de vida y en particular el estado nutricional de los individuos.

Planteámiento:

La investigación aquí planteada pretende aproximarse al conocimiento acerca de la manera en la que el proceso de crecimiento infantil se ve influida por factores del medio ambiente, tales como las condiciones socio-económicas de las familias y el estado general de nutrición en tres comunidades de la Sierra Norte de Puebla.

Objetivos:

1) Establecer las condiciones sociales y económicas en las cuáles crecen y se desarrollan los niños del estudio.

2) Describir el parecido genético de cada una de las poblaciones estudiadas, a travez del estudio de los dermatoglifos digitales y palmares.

3) Conocer el proceso de crecimiento físico en cada una de las poblaciones, en niños de 8 años de edad hasta personas de más de 18 años.

4) Conocer las diferencias entre las poblaciones, tanto de orden génetico como de caracter ambiental a partir de la comparación de las variables manejadas.

Hipótesis:

En todos los casos se parte de la hipótesis nula en el sentido de que las tres comunidades estudiadas no presentan diferencias significativas en las variables consideradas en virtud de pertenecer al mismo Universo.

1) H0: No existen diferencias significativas entre las poblaciones, con respecto a las dimensiones y proporciones corporales en los grupos de individuos de la misma edad y sexo

H1: Se esperan diferencias significativas entre las poblaciones con respecto a tamaños y proporciones corporales según grupos de edad y sexo.

2) H0: No existen diferencias significativas entre las poblaciones, con respecto a los rasgos dermopapilares según lugar de procedencia y género.

H1: Hay diferencias significativas entre las poblaciones, con respecto a los rasgos dermatoglificos, según lugar de procedencia y género.

Material y técnicas

Muestreo:

Muestra: Se tomaron como muestra (1978 - 1983) de estudio individuos de 8 a 18 o más años de edad pertenecientes a tres comunidades de la región de la Sierra Norte de Puebla. Cada una de estas comunidades presenta diferentes condiciones de desarrollo, en lo que toca a comunicación, urbanización y servicios.

Para la investigación dermatoglifica se reunieron a las unidades muestrales por sexo y lugar de procedencia. En tanto que para las variables antropometricas se hicieron grupos según edad, sexo y lugar de procedencia.

CEDULA DE COLECCION DE DATOS

LABORATORIO DE INVESTIGACIONES SOMATOLOGICAS

CEDULA INDIVIDUAL ANTROPMETRICA

CEFALOMETRIA			
DIAMETRO ANTERO_POSTERIOR	_____	ALTURA NASION_NAGTION	_____
DIAMETRO TRANVERSO MAXIMO	_____	ALTURA DE LA NARIZ	_____
ANCHURA DE LA FRENTE	_____	ANCHURA DE LA NARIZ	_____
DIAMETRO BICIGOMATICO	_____	ANCHURA DE LA BOCA	_____
DIAMETRO BIGONIACO	_____	ESPESOR DE LABIOS	_____
		PERIMETRO CEFALICO	_____

SOMATOMETRIA			
ESTATURA TOTAL	_____	PERIMETRO DEL BRAZO CONTRAIDO	_____
ALTURA AL TRAGION	_____	PERIMETRO DEL MUSLO	_____
ALTURA AL ACROMION	_____	PERIMETRO DE LA PANTORRILLA	_____
ALTURA AL PTO. RADIAL	_____	DIAMETRO ANTERO POSTERIOR DE TORAX	___
ALTURA AL ESTILION	_____	ESTATURA SENTADO	_____
ALTURA AL DACTILION	_____	ANCHURA DEL CODO	_____
ALTURA ESPINO ILIACA	_____	ANCHURA DE LA MUÑECA	_____
ALTURA YUGULAR	_____	ANCHURA DE LA MANO	_____

Metodología estadística de análisis:

El análisis de los dermatoglifos se realizó a partir de la identificación de los diseños digitales y lineas principales de la mano. En el caso de los datos antropométricos se incluyeron tres grupos principales de variables: Tamaños en longitud, anchuras y volumen (peso), con estos se pueden conocer dimensiones en tamaño y proporciones corporales.

DERMATOGLIFOS:

En el caso de los dermatoglifos, por tratarse de variables cualitativas, el análisis descriptivo se realizó a partir de la distribución porcentual de los datos. Y para la contrastación de hipótesis se emplearon métodos estadísticos no paramétricos, en este caso se hicieron tablas de contingencia y se usó la prueba X^2.

DATOS ANTROPOMETRICOS:

Para el análisis de las datos antropométricos, un primer paso consistió en la obtención de los estadísticos de tendencia centrales así como de dispersión según grupos de edad, sexo, y comunidad de origen; esta información se concentró en múltiples cuadros y se elaborarón gráficas para ilustrar las características correspondientes. Como segundo paso, se indagó si había o no diferencias significativas de las variables estudiadas entre las poblaciones consideradas; para el caso se hizo uso del análisis de la varianza.

Cuadro 1. Número de individuos estudiados según edad, sexo y lugar de origen.

Edad	Zacapoaxtla		Mecapalapa		Caxhuacan		Total
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	
7					10	8	18
8	27	28	10	4	13	14	96
9	29	35	24	20	30	28	166
10	43	35	20	26	29	17	170
11	37	56	30	39	32	19	213
12	41	57	37	41	33	31	240
13	25	30	58	39	29	20	201
14	29	27	37	30	48	22	193
15	29	26	35	38	26	19	173
16	29	25	30	27	25	15	151
17	9	4	32	12	21	7	85
18	34	29	44	16	45	33	201
Total	332	352	357	292	341	233	1907

Cuadro 2. Ocupación de los padres de familia.

Ocupación	Zacapoaxtla, Pue.			
	Padre		Madre	
	n	%	n	%
Hogar			311	83.16
Campesino	76	20.32		0.00
Albañil	51	13.64		0.00
Comerciante	40	10.70	22	5.88
Servicio doméstico	1	0.27	16	4.28
Empleado	32	8.56	14	3.74
Oficios calificados	33	8.82		
Profesor	29	7.75	11	2.94
Chofer	27	7.22		
Labores no calificadas	15	4.01		
Panadero	10	2.67		
Profesionista	6	1.60		
Mecánico	5	1.34		
Obrero	5	1.34		
Carnicero	4	1.07		
Sastre o costurero	4	1.07		
Taxista	4	1.07		
No se supo	32	8.56		
Total	374	100	374	100

Cuadro 3. Distribución ocupacional de los padres de familia según sectores de la producción.

	Zacapoaxtla, Pue.			
Sectores	Padre	%	Madre	%
Primario (1)	76	20.32		
Secundario (2)	56	14.97		
Terciario (3)	210	56.15	374	100.00
No se supo	32	8.56		
Total	374	100	374	100

(1) Actividad económica relacionada con la agricultura, ganadería, caza, selvicultura y pesca.
(2) Actividad económica relacionada con la minería, extracción del petróleo y gas, industria manufacturera, electricidad, agua y construcción.
(3) Actividad económica relacionada con el comercio, transporte, comunicaciones y servicios.

Cuadro 4. Escolaridad de los padres de familia.

	Zacapoaxtla, Pue.			
Escolaridad	Padre		Madre	
	n	%	n	%
Analfabeto (a)	50	13.37	96.00	25.67
Primaria incompleta	158	42.25	154.00	41.18
Primaria completa	64	17.11	56.00	14.97
Secundaria o técnica	25	6.68	33.00	8.82
Preparatoria o Normal	31	8.29	12.00	3.21
Licenciatura	8	2.14	2.00	0.53
No se supo	38	10.16	21.00	5.61
Total	374	100	374	100

Cuadro 5. Algunas características de la vivienda.

Zacapoaxtla, Pue.

	n	%
Viviendas		
Propias	612	49.88
Con agua entubada:		
Dentro de la vivienda	780	63.57
Fuera de la vivienda	295	24.04
Con drenaje	872	71.07
Con piso diferente a tierra	977	79.63
Con energía eléctrica	962	81.09
Con radio	995	81.09
Con televisión	694	56.55
Total	1227	

Fuente: Censo General de Población 1980

Cuadro 8. Escolaridad de los padres de familia.

Mecapalapa, Pue.

Escolaridad	Padre		Madre	
	n	%	n	%
Analfabeto (a)	147	38.38	171	44.65
Primaria incompleta	174	45.43	172	44.91
Primaria completa	34	8.88	27	7.05
Secundaria o técnica	3	0.78	1	0.26
Preparatoria o normal	2	0.52		
Licenciatura	1	0.26		
No se supo	22	5.74	12	3.13
Total	383	100	383	100

Fuente: Encuesta realizada en 1979.

Cuadro 9. Algunas características de la vivienda.

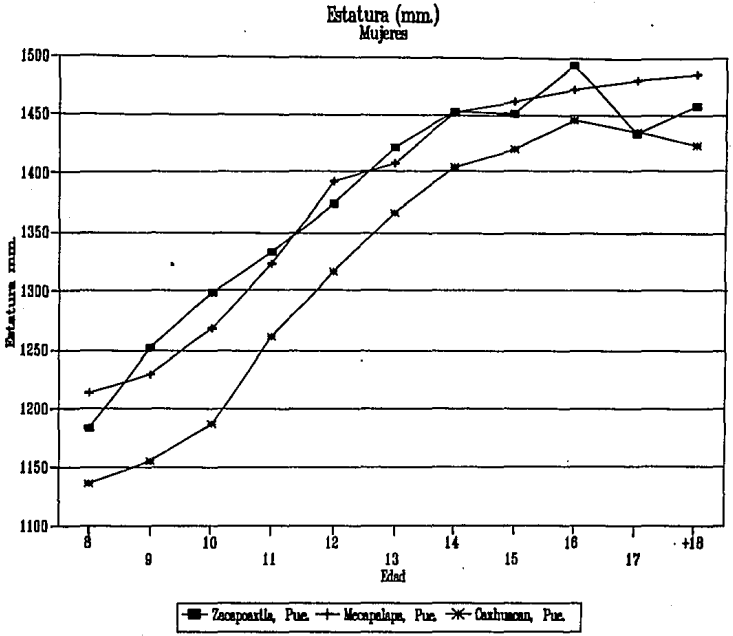
Mecapalapa, Pue.

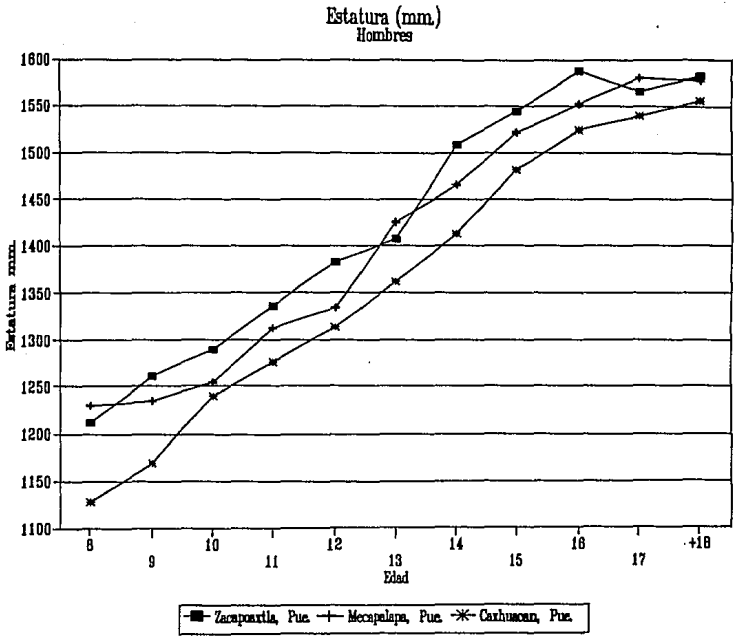
Viviendas	n	%
Propias	327	72.03
Con agua entubada:		
Dentro de la vivienda	108	23.79
Fuera de la vivienda	95	20.93
Con drenaje	123	27.09
Piso diferente a tierra	161	35.46
Con energía eléctrica	207	45.59
Con radio	289	63.66
Con televisión	101	22.25

Fuente: Censo General de Población, 1980.

Cuadro 51. Estatura (cm.)

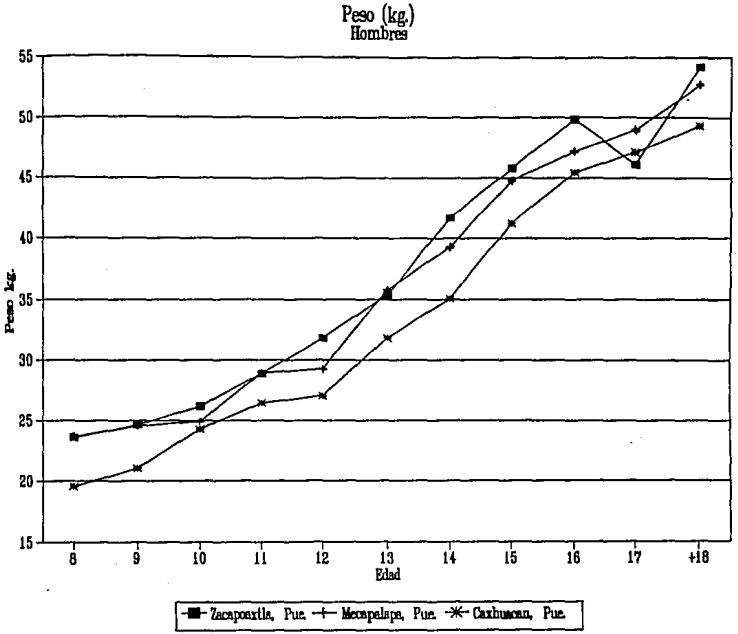
Grupos de edad	Hombres n	media	desv. estd.	Mujeres n	media	desv. estd.
Zacapoaxtla, Pue.						
8	27	121.22	5.96	28	118.34	6.82
9	29	126.13	4.43	35	125.22	6.28
10	43	129.07	5.98	35	129.95	6.22
11	37	133.65	5.00	56	133.45	6.88
12	41	138.30	6.36	57	137.49	7.71
13	25	140.77	5.70	30	142.27	6.87
14	29	150.81	7.85	27	145.34	6.18
15	29	154.45	8.05	26	145.20	5.75
16	29	158.85	6.46	25	149.33	5.02
17	9	156.50	7.31	4	143.48	4.90
+18	34	158.27	6.24	29	145.83	4.89
Mecapalapa, Pue.						
8	10	123.04	4.29	4	121.32	3.65
9	24	123.52	4.85	20	122.97	5.39
10	20	125.51	4.90	26	126.90	6.01
11	30	131.30	6.53	39	132.43	7.01
12	37	133.47	5.88	41	139.42	6.43
13	58	142.56	9.09	39	140.91	7.03
14	37	146.55	7.47	30	145.23	5.53
15	35	152.15	7.24	38	146.18	7.38
16	30	155.24	6.93	27	147.20	4.15
17	32	158.09	6.24	12	148.06	4.94
+18	44	157.74	6.18	16	148.59	6.41
Caxhuacan, Pue.						
6	3	107.30	1.14		104.20	0.30
7	10	107.61	5.44	8	106.61	3.91
8	13	112.83	5.37	14	113.61	6.19
9	30	116.84	4.91	28	115.54	6.12
10	29	123.98	6.37	17	118.72	7.71
11	32	127.66	8.28	19	126.23	8.06
12	33	131.42	7.66	31	131.71	6.56
13	29	136.23	6.83	20	136.74	8.01
14	48	141.31	9.07	22	140.65	6.29
15	26	148.15	7.43	19	142.14	5.79
16	25	152.48	4.31	15	144.65	4.86
17	21	154.00	7.73	7	143.70	2.81
+18	45	155.58	5.13	33	142.55	0.40

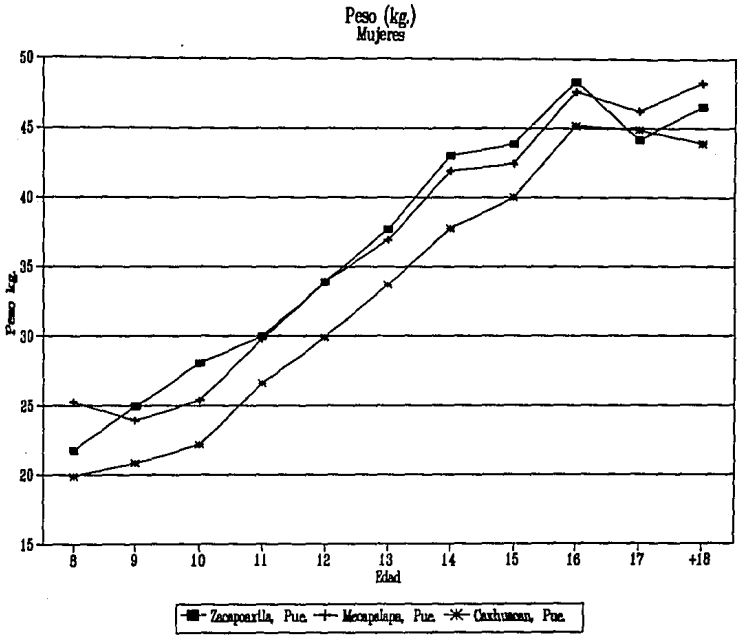

Estatura (mm.)
Mujeres
1500
1450
1400
1350
1300
1250
1200
1150
1100
Estatura mm.
8
9
10
11
12
13
14
15
16
17
+18
Edad
Zacapoaxtla, Pue.
Mecapalapa, Pue.
Cuxhuacan, Pue.


Estatura (mm.)
Hombres
1600
1550
1500
1450
1400
1350
1300
1250
1200
1150
1100
Estatura mm.
8
9
10
11
12
13
14
15
16
17
+18
Edad
Zacapoaxtla, Pue.
Mecapalapa, Pue.
Cuyuaco, Pue.

Cuadro 67. Peso (kg.)

Grupos de edad	Hombres n	media	desv. estd.	Mujeres n	media	desv. estd.
Zacapoaxtla, Pue.						
8	27	23.59	4.52	28	21.71	2.76
9	29	24.62	3.31	35	24.90	3.86
10	43	26.19	4.70	35	28.07	4.98
11	37	28.85	4.21	56	29.96	5.00
12	40	31.84	4.67	58	33.90	6.60
13	25	35.34	4.63	30	37.68	5.95
14	29	41.72	6.57	27	43.00	8.87
15	29	45.78	6.15	26	43.81	5.90
16	29	49.84	8.03	24	48.27	7.69
17	9	46.11	4.14	4	44.12	5.27
+18	34	54.18	9.79	28	46.48	6.37
Mecapalapa, Pue.						
8	10	23.65	2.23	4	25.25	2.59
9	24	24.58	3.28	20	23.90	2.90
10	20	24.95	3.75	26	25.38	3.23
11	30	28.98	4.40	39	29.76	6.32
12	37	29.23	3.54	41	33.90	6.14
13	59	35.77	7.47	39	36.91	7.13
14	36	39.25	7.49	30	41.92	9.14
15	35	44.77	7.70	38	42.38	6.09
16	30	47.17	5.57	27	47.57	6.00
17	32	48.92	6.61	12	46.17	2.67
+18	44	52.67	7.09	16	48.25	5.21
Caxhuacan, Pue.						
6	3	18.50	0.41			
7	10	18.05	1.60	8	17.38	2.18
8	13	19.58	1.99	14	19.82	2.75
9	30	21.07	2.05	28	20.82	2.77
10	28	24.27	2.35	17	22.15	3.43
11	32	26.44	5.54	18	26.58	5.19
12	33	27.05	3.71	31	29.90	5.53
13	29	31.79	4.91	20	33.67	6.89
14	48	35.06	6.77	22	37.70	5.45
15	26	41.25	5.17	19	40.00	5.73
16	23	45.46	6.68	15	45.17	5.17
17	21	47.10	7.56	7	44.86	6.66
+18	45	49.32	6.86	32	43.91	3.69


Peso (kg.)
Hombres
55
50
45
40
35
30
25
20
15
Peso kg.
8
9
10
11
12
13
14
15
16
17
+18
Edad
Zacapoaxtla, Pue.
Mecapalapa, Pue.
Cuyuacan, Pue.


Peso (kg.)
Mujeres
50
45
40
35
30
25
20
15
Peso kgs.
8
9
10
11
12
13
14
15
16
17
+18
Edad
Zacapoaxtla, Pue.
Mecapalapa, Pue.
Cuxhuacan, Pue.

VII. PROBABILIDAD

VII.1. INTRODUCCIÓN

Un fenómeno aleatorio (fortuito o al azar) es un fenómeno empírico que se caracteriza por la propiedad de que, al observarlo bajo ciertas condiciones, no siempre se obtiene el mismo resultado (de manera que no existe regularidad determinística) sino que los diferentes resultados ocurren con regularidad Estadística.

Un evento aleatorio tiene la propiedad de que la frecuencia relativa con la que aparece en una sucesión muy larga de observaciones realizadas al azar, se acerca a un valor límite estable a medida que el número de observaciones tiende a infinito: el valor límite de la frecuencia relativa se llama probabilidad del evento aleatorio.

La teoría de probabilidades estudia los métodos de análisis que son comunes en el tratamiento de fenómenos aleatorios, cualquiera que sea el área en que estos se presenten.

La probabilidad de un evento (que se entiende como la frecuencia relativa con la que ocurre un evento) es igual al cociente del número de resultados del experimento en los que se observa dicho evento entre el total de resultados posibles del experimento.

El espacio de descripciones muestrales de un fenómeno aleatorio, que usualmente se denota con la letra S, es el espacio de las descripciones (o nombres) de todos los resultados posibles del experimento.

Evento es un subconjunto del espacio de descripciones muestrales S. En particular, el espacio de descripciones S es un subconjunto de sí mismo, por lo que también constituye un evento, al que se le llama evento seguro, puesto que la manera de formularlo asegura que siempre ocurrirá.

Evento elemental es un evento que contiene exactamente una descripción.

Al estudiar un fenómeno aleatorio son de interés los eventos que pueden ocurrir (o, más precisamente, las probabilidades de que ocurran), por lo que el interés del espacio de descripciones muestrales no radica en sus elementos, sino en sus subconjuntos ¡que son los eventos!

VII.2. ALGEBRA DE EVENTOS.

Dado un evento E cualquiera, es común calcular la probabilidad de no suceda como la de que suceda, y resulta conveniente asociar con E otro evento, designado como E^c, y que se llama complemento de E; E^c es el evento de que no suceda E, y consta de todas las descripciones de S que no pertenecen a E.

Dados dos eventos E y F entonces E∩F (E intersección F) es el evento en que E y F ocurren a la vez, E∪F (E unión F) es el evento de que ocurre E u ocurre F.

Se dice que dos eventos son iguales, es decir, que E=F, si cada una de las descripciones en uno de ellos pertenece también al otro y viceversa.

El evento imposible denotado como ø, es el evento que no contiene descripción alguna, y que por tanto, no puede ocurrir.

Dos eventos E y F que no puedan ocurrir simultáneamente, es decir tales que su intersección EF sea el evento imposible, se llaman mutuamente exclusivos o excluyentes. Así pues, dos eventos E y F son mutuamente exclusivos si y sólo si EF=ø.

Definición de probabilidad como función de los eventos contenidos en el espacio de descripciones muestrales de un fenómeno aleatorio:

Dada una situación aleatoria, que queda descrita por un espacio de descripciones muestrales S, la probabilidad es una función P[·] que asigna a cada evento un número real no negativo, denotado por P[E], que se llama probabilidad del evento E. Dicha función de probabilidades debe de satisfacer los tres axiomas siguientes:

Axioma 1. P[E] ≥ 0 para todo evento E.
Axioma 2. P[S] = 1 donde S es el evento seguro.
Axioma 3. P[E∪F] = P[E] + P[F], en el caso de que EF=ø;

es decir, la probabilidad de la unión de dos eventos mutuamente exclusivos es la suma de sus probabilidades.

El espacio de descripciones muestrales $S=\{D_1, D_2, \ldots D_N\}$, tiene descripciones igualmente probables si todos los eventos elementales de S tienen la misma probabilidad, de manera que:

$$P[\{D_1\}]=P[\{D_2\}]=\ldots\ldots=P[\{D_N\}]=\frac{1}{N};$$

La suma de las probabilidades de estos eventos es igual 1.

La fórmula para calcular las probabilidades de eventos cuando el espacio de descripciones muestrales S es finito y todas las descripciones son igualmente probables: Para cualquier evento E de S se tiene:

$P[E]=\frac{N[E]}{N[S]} = \frac{\text{tamaño de E}}{\text{tamaño de S}}$, donde el tamaño de un evento E es el número de miembros de dicho evento.

VII.2.1. TEORIA ELEMENTAL DE CONJUNTOS.

CONJUNTOS, DEFINICION y NOTACION.

El término conjunto juega un papel fundamental en el desarrollo de las matemáticas modernas. El origen de este concepto se debe al matemático alemán George Cantor (1845-1918) y, surgió de la necesidad de darle rigurosidad lógica a las discusiones matemáticas con el fín de eliminar la ambigüedad del lenguaje cotidiano.

Definición intuitiva:

Un conjunto no tiene definición matemática, sin embargo, en forma intuitiva un conjunto es un agregado o colección de objetos de cualquier naturaleza con caraterísticas bien definidas de manera que se pueden distinguir todos sus elementos. A los objetos que lo componen se les llama elementos del conjunto.

NOTACION:

A los conjuntos se les denota con letras mayúsculas y a sus elementos con letras minúsculas; a los elementos se les encierra entre { } y se separan con ",". Así por ejemplo, el conjunto D cuyos elementos son los números que aparecen al lanzar un dado se escribe:

$D=\{1,2,3,4,5,6\}$

Se pueden citar infinidad de ejemplos de conjuntos, algunos de ellos son:

El conjunto de días de la semana.
El conjunto de las vocales
El conjunto de los números reales
El conjunto de entierros en un sitio arqueológico.
El conjunto de tepalcates hallados en un recorrido de superficie.

Los conjuntos se pueden escribir en forma implícita (por descripción) cuando no se enumeran o enlistan todos sus elementos, o en forma explícita (por enumeración) cuando se enlistan todos sus elementos; por ejemplo:

El conjunto de días de la semana se puede escribir en forma implícita: $S = \{ x \mid x \text{ es un día de la semana} \}$

y se debe de leer así:

S es el conjunto de los elementos x tales que x es un día de la semana.

También se puede escribir en forma explícita:
S={domingo,lunes,martes,miercoles,jueves,viernes,sábado}

CARDINALIDAD DE UN CONJUNTO:

Según la cantidad de elementos de un conjunto, estos se pueden clasificar en conjuntos finitos los que tienen un número conocido de elementos, y los conjuntos infinitos, los que tienen un número ilimitado de elementos.

El número de elementos diferentes de un conjunto se le llama cardinalidad del conjunto, y lo denotamos por n(A) ó # (A).

Propiedades:

1) $n(A) \geq 0 \ \forall \ A \subseteq U$
2) $n(\varnothing) = 0$
3) $n(A \cup B) = n(A)$, si $A \cap B = \varnothing$

Proposición:
Sea A cualquier subconjunto del conjunto universal U, entonces $n(A^c) = n(U) - n(A)$.

Proposición:
Si A y B son dos subconjuntos cualesquiera del conjunto universal U, entonces: $n(A \cup B) = n(A) + n(B) - n(A \cap B)$.

Cuando se quiere indicar que un elemento pertenece a un conjunto determinado o que es un miembro del conjunto se utiliza el símbolo $\in$ que se lee "*pertenece a*" o "*es elemento de*"; y cuando se quiere indicar que no pertenece al conjunto se utiliza: $\notin$

CONJUNTO VACIO:

Un conjunto que no tiene elementos se le llama conjunto vacío y se denota por ø ó por { }.

CONJUNTO UNIVERSAL:

El conjunto universal U es el conjunto de todos los elementos considerados en un problema o situación dada. Por ejemplo, si sólo se quiere trabajar con los números reales positivos, el conjunto universal será:

$$U = \mathbb{R}^+$$

SUBCONJUNTOS:

Un conjunto B es un *subconjunto* de un conjunto A si todos los elementos de B pertenecen a A y se escribe $B \subset A$, y se lee: " B está contenido en A " ó " B es subconjunto de A".

SUBCONJUNTO PROPIO

Sean dos conjuntos A y B, se dice que B es un subconjunto propio de A si todos los elementos de B pertenecen a A y además A contiene por lo menos un elemento que no pertenece a B.

IGUALDAD ENTRE CONJUNTOS

Dos conjuntos son iguales, si A es subconjunto de A y si B es subconjunto de A, es decir $A=B \Leftrightarrow A \subset B$ y $B \subset A$.

OPERACIONES ENTRE CONJUNTOS

Básicamente se definen cuatro operaciones con conjuntos; la unión, la intersección, el complemento y la diferencia.

UNION DE CONJUNTOS

Sean A y B dos subconjuntos del conjunto universal U. La unión de A con B denotada por $A \cup B$, es el conjunto de todos los elementos que pertenecen a A, a B o a ambos, en símbolos:

$$A \cup B = \{x \mid x \in A \text{ ó } x \in B\}$$

INTERSECCION DE CONJUNTOS

Sean A y B dos subconjuntos del conjunto universal U. La intersección de A con B denotada por $A \cap B$, es el conjunto de todos los elementos que pertenecen a A y B simultaneamente. En símbolos:

$$A \cap B = \{x \mid \in A \text{ y } x \in B\}$$

CONJUNTOS DISJUNTOS

Dos conjuntos A y B que no tienen elementos en común, es decir $A \cap B = \emptyset$ se llaman disjuntos.

COMPLEMENTO DE UN CONJUNTO

Sea A un subconjunto del conjunto universal U. El complemento de A, denotado por A^c o A' es el conjunto de los elementos de U que no pertenecen a A, es decir:

$$A^c = \{x | x \in U \text{ y } x \notin A \}.$$

DIFERENCIA DE CONJUNTOS

Sean A y B dos subconjuntos de U, la diferencia de A - B, es el conjunto de los elementos de A que no pertenecen a B, esto es:

$$A - B = \{ x \mid x \in A \text{ y } x \notin B \}.$$

LEYES O PROPIEDADES DE LAS OPERACIONESCON CONJUNTOS

$A \cup \varnothing = A$
$A \cap \varnothing = \varnothing$
$A \cup U = U$
$A \cap U = A$
$\varnothing \cup U = U$
$\varnothing \cap U = \varnothing$
$A \cup A = A$
$A \cap A = A$

$\varnothing^c = U$

$U^c = \varnothing$

$A \cup A^c = U$

$A \cap A^c = \varnothing$

$(A^c)^c = A$
$A \cup B = B \cup A$
$A \cap B = B \cap A$
$A \cup (B \cup C) = (A \cup B) \cup C$
$A \cap (B \cap C) = (A \cap B) \cap C$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$(A \cup B)^c = A^c \cap B^c$

$(A \cap B)^c = A^c \cup B^c$

VII.2.2. ESPACIO MUESTRAL Y EVENTO.

Experimento:

Un experimento es cualquier operación cuyo resultado no puede prdecirse.

Experimento Aleatorio es un proceso que tiene las siguientes propiedades:

1) El proceso se efectúa de acuerdo a un conjunto bien definido de reglas.

2) Es de naturaleza tal que se repite o puede concebirse la repetición del mismo.

3) El resultado de cada ejecución depende de "la casualidad" y, por lo tanto, no se puede predecir un resultado único.

Una sola ejecución del experimento se llama ENSAYO.

ESPACIO MUESTRAL

El espacio de Descripciones Muestrales S de un experimento es el conjunto de todos los resultados posibles del experimento.

Evento:

Es un subconjunto del espacio de descripciones muestrales. Cada subconjubto es un evento.

Un evento ocurre si cada uno de sus elementos es el resultado del experimento.

EVENTOS MUTUAMENTE EXCLUSIVOS.

Dos eventos A y B que no ocurren simultáneamente o que no tienen elementos en común, es decir $A \cap B = \emptyset$ se les llama eventos exclusivos o mutuamente excluyentes.

EVENTOS COMPLEMENTARIOS.

Dos eventos A y B son complementarios si $A \cup B = S$ y $A \cap B = \emptyset$, y a B se le denota por A^c.

DEFINICION DE PROBABILIDAD.

Antes de profundizar en la forma que se utilizan las probabilidades, es necesario conocer de cierta manera de donde provienen. Hay tres formas de calcular o estimar la probabilidad.

i) El enfoque clásico o a "*a priori*" proveniente de los juegos de azar o de definición clásica de Laplace que se emplea cuando los espacios muestrales son finitos y tienen resultados igualmente probables.

ii) La definición empírica, "*a posteriori*" o frecuencial que se basa en la frecuencia relativa de ocurrencia de un evento con respecto a un gran número de ensayos repetidos

iii) Por último la definición de Kolmogorov o definición axiomática o matemática de la probabilidad.

Seleccionar uno de los tres enfoques dependerá de la naturaleza del problema.

DEFINICION CLASICA DE LAPLACE O "*A PRIORI*".

Esta definición es de uso limitado puesto que descansa sobre la base de las dos siguientes condiciones:

i) El espacio muestral de todos los resultados posibles S es finito.

ii) Los resultados del espacio muestral deben de ser igualmente probables.

Bajo estas condiciones y si A es el evento formado por n(A) resultados del espacio muestral y el número total de resultados posibles es n(S), entonces:

$$p(A) = \frac{n(A)}{n(S)}$$

DEFINICION EMPIRICA

"*A POSTERIORI O FRECUENCIAL*":

La definición clásica se ve limitada a situaciones en las que hay un número finito de resultados igualmente probables. Por desgracia, hay problemas prácticos que no son de este tipo y la definición de Laplace no se puede aplicar. Por ejemplo, si se pregunta por la probabilidad de que un paciente sea curado mediante cierto tratamiento médico, o la probabilidad de que una determinada máquina produzca artículos defectuosos, entonces no hay forma de introducir resultados igualmente probables. Por ello se necesita un concepto más general de probabilidad. Una forma de dar respuesta a estas preguntas es obtener algunos datos empíricos en un intento por estimar las probabilidades. Supongase que se efectua un experimento n veces y que en esta serie de ensayos el evento A ocurre exactamente r veces, entonces la frecuencia relativa del evento es:

$$\frac{r}{n}\ ,\ \text{o sea}\ f_r(E) = f(E) = \frac{r}{n}$$

Si se continua calculando la frecuencia relativa de cada cierto número de ensayos, a medida que se aumenta n, las frecuencias relativas correspondientes serán más estables; es decir; tienden a ser casi las mismas; en este caso se dice que el experimento tiene *regularidad Estadística*.

La gran mayoría de experimentos aleatorios de importancia práctica tienen estabilidad, por esto se puede sospechar que prácticamente será cierto que la frecuencia relativa de un evento E en un gran número de ensayos es aproximadamente igual a un determinado número P(E), o sea la probabilidad del evento E es:

$$P(E) = \lim_{n \to \infty} \frac{r}{n}$$

Obsérvese que este número no es una propiedad que depende solamente de E, sino que se refiere a un cierto espacio muestra S y a un experimento aleatorio. Entonces, decir que el evento E tiene la probabilidad P(E) significa que si se efectúa el experimento muchas veces, es prácticamente cierto que la frecuencia relativa de E, f(E) es aproximadamente igual a P(E).

Cuando se usa la definición frecuencial, es importante tomar en cuenta los siguientes aspectos:

i) La probabilidad obtenida de esta manera es únicamente una estimación del valor real.

ii) Cuanto mayor sea el número de ensayos, tanto mejor será la estimación de la probabilidad; es decir, a mayor número de ensayos mejor será la estimación.

iii) La probabilidad es propia de sólo un conjunto de condiciones idénticas a aquéllas en las que se obtuvieron los datos, o sea, la validez de emplear esta definición depende de que las condiciones en que se realizó el experimento sean repetidas idénticamente.

DEFINICION AXIOMATICA O MATEMATICA DE KOLMOGOROV:

Las definiciones anteriores son netamente empíricas o experimentales, sin embargo después de establecer una forma de determinar la probabilidad experimentalmente, se pueden deducir leyes o propiedades de la probabilidad en forma lógica o computacional bajo ciertas suposiciones llamados axiomas de la probabilidad.

La probabilidad de un evento A se define como el número P(A), tal que cumple con los siguientes axiomas:

Axioma 1:

La probabilidad P(A) de cualquier evento no debe ser menor que cero ni mayor que uno.

$0 \leq P(A) \leq 1.$

Axioma 2:

La probabiliad del evento seguro es :
$P(S)=1$

Axioma 3:

Si A y B son dos eventos mutuamente exclusivos $\{ A \cap B = \varnothing \}$, entonces $P(A \cup B) = P(A) + P(B)$.

Toda la teoría elemental de la probabilidad está construida sobre las bases de estos simples tres axiomas.

VII.2.3. DETERMINACION PRACTICA DE PROBABILIDADES.

La determinación práctica de probabilidades depende del problema que se presente, si se tiene un espacio muestral finito con resultados igualmente probables, se utilizará el concepto clásico de probabilidad, ya que éste satisface los tres axiomas de la definición matemática de probabilidad.

Hay que recordar que las condiciones para el cálculo de probabilidades (clásica) son:

i) El espacio muestra de todos los resultados posibles S es finito.

ii) Los resultados del espacio muestra deben de ser igualmente probables.

Bajo estas condiciones y si A es el evento formado por n(A) resultados del espacio muestral y, el número total de resultados posibles es n(S), entonces:

$$p(A) = \frac{n(A)}{n(S)}$$

Como puede observarse en la expresión anterior está involucrado el número de elementos de los conjuntos A y S por lo que es necesario conocer técnicas de conteo.

Para responder a las preguntas de: "Cuántos" o "cuál es el número " se pueden proceder de dos formas: una es listar todos los posibles resultados (que puede ser una tarea muy laboriosa) y la otra es determinar el número sin necesidad de listar todos los posibles resultados. Para determinar el número existen técnicas de conteo que ayundan enormemente en este proceso.

VII.2.3.1. TECNICAS DE CONTEO.

Principios fundamentales en el proceso de contar:

Principio de la multuplicación.

"Si una operación o un proceso consiste de n diferentes pasos, de los cuales el primero puede ser realizado de p_1 formas, el segundo de p_2 formas, el tercero p_3 formas,....., y el k ésimo de p_k formas entonces la operación o el proceso completo se puede realizar de: p_1 p_2 p_3 p_4 ... p_k formas".

Principio de adición.

"Si los procesos $p_1\ p_2\ p_3\ p_4 \cdots p_k$ se pueden realizar de: $n_1, n_2, n_3, \ldots, n_r$ formas diferentes, entonces el proceso: p_1 ó p_2 ó p_3 ó ... ó p_r se puede realizar de $n_1 + n_2 + \ldots + n_r$ formas diferentes".

Factorial de un número entero positivo:

El producto de los n primeros enteros positivos consecutivos es llamado n-factorial o el factorial de n y se denota por n!. Si n=0, entonces 0! =1. Por lo tanto:

$n!=n(n-1)(n-2)\ldots(3)(2)(1)$.

Ordenaciones.

Las ordenaciones de n elementos tomados todos juntos o en una parte de ellos son:

"Las ordenaciones sin repetición de n-elementos tomados todos a la vez es n!"

"Las permutaciones sin repetición de n-elementos tomados de r en r son $n(n-1)(n-2) \ldots (n-r+1)$"

"El número de ordenaciones con repetición de n-elementos tomados de r en r es n^r."

En cada una de las "ordenaciones" mencionadas se entiende que el orden de los elementos es de importancia, cuando este orden no importa se tienen "combinaciones" sin reprepetición de n elementos tomadas de r en r y su expresión análitica es:

$$\binom{n}{r} = \frac{n!}{(n-r)!\ r!}$$

Si la naturaleza del experimento no señala que el número finito de resultados tenga igual posibilidad de ocurrir, o si el espacio muestral no es finito y la naturaleza del experimento no indica como subdividir el espacio muestral en un número finito de eventos igualmente probables, se deben de asignar probabilidades usando las frecuencias relativas que se observen en largas secuencias de ensayos. Esto se debe de hacer de manera que los axiomas de la probabilidad se satisfagan. De esta manera se obtienen valores aproximados.

A veces la probabilidad del evento A se reporta como:

P(A) × 100, que significa que cada 100 veces que se realice el experimento, P(A) × 100 veces se verifica el evento A, así por ejemplo, si P(A) = 0.25, se puede decir que el evento A tiene una probabilidad de 25 % o que el evento ocurre 25 % de las veces.

VII.3. PROPIEDADES DE LA PROBABILIDAD.

i) $P(\emptyset) = 0$.

ii) Si $A_1, A_2, \ldots, A_r$ son eventos mutuamente exclusivos entonces:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_r) = P(A_1) + P(A_2) + \ldots + P(A_r)$$

iii) Si A y B son eventos cualesquiera del espacio muestral S, entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

iv) $P(A^c) = 1 - P(A)$.

VII.4. FUNCION VALOR ESPERADO DE UNA VARIABLE ALEATORIA.

Cuando se requiere tener noción del valor promedio o esperado de una variable aleatoria X, y no solamente una afirmación de la probabilidad de que X tome un cierto valor o la probabilidad de que la variable aleatoria X esté en un cierto intervalo:

$$P(a \leq X \leq b).$$

o acerca de la probabilidad de que X sea mayor o igual a un cierto valor x:

$$P(X \geq x),$$

la probabilidad de que la variable aleatoria X sea menor o igual a un cierto valor x:

$$P(X \leq x), \text{ etc.}$$

Se utiliza el concepto de ESPERANZA MATEMATICA que es un valor promedio teórico, y la definición para variables aleatorias discretas es:

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) \quad i = 1 \ldots n; \text{ para el caso discreto}$$

$$E(X) = \int f(x)\, dx \text{ para el caso continuo}$$

Ejemplos:

Si se apuesta N$ 1 en un juego de azar, se puede preguntar cuánto se espera ganar o perder (en promedio) dependiendo de las reglas del juego.

Una paciente se somete a un tratamiento de reducción de peso y le pregunta al médico cuánto espera bajar de peso (en promedio).

En un vivero se cultivan peces de diversas especies, al encargado se le pregunta cuál será la población adulta (esperanza de vida) de cada especie.

Cada una de estas preguntas se refiere a un valor promedio de una variable aleatoria.

Ejemplo:

I) La probabilidad de que una casa de cierto tipo quede destruida por un incendio en cualquier periodo de doce meses es de 0.005. Una compañía de seguros ofrece vender al propietario una póliza anual de seguros contra incendio por N$ 200,000 por una prima de N$ 2000.

¿ Cuál es la "ganancia" esperada de la compañía ?

Solución:

i) Se ve cual es el espacio muestral:

$$S = \{ \text{se incendie, no se incendie} \}$$

ii) De acuerdo al experimento se define la variable aleatoria.

Se define la variable aleatoria X (una función que asigna un valor real a cada elemento del espacio muestral)

$X = -(200{,}000 - 2000)$ si se incendia
$X = 2000$ si no se incendia.

iii) Se calcula la probabilidad para cada valor de la variable aleatoria:

la probabilidad de que se incendie es de 0.005 (esto es por condiciones del problema, en la vida real se consultan las Estadísticas oficiales sobre los siniestros ocurridos)

la probabilidad de que NO se incendie es igual a:

$$1 - P(\text{incendio}) = 1 - 0.005 = 0.995$$

iv) Se calcula la esperanza de la variable aleatoria

$$E(X) = -198{,}000\ P(\text{incendio}) + 2000\ P(\text{no incendio}) =$$
$$E(X) = -198{,}000\ P(X=-198000) + 2000\ P(X=2000)$$
$$E(X) = -198000\ (0.005) + 2000\ (0.995) = 1000$$

v) Se concluye el problema.

Lo que indica que la compañía espera ganar:

N$ 1000 en promedio

VII.5. FUNCIONES DE PROBABILIDAD DISCRETAS

Prueba Bernoulli.

Una prueba Bernoulli es un experimento aleatorio que tiene dos resultados posibles, a los cuales con frecuencia se les llama "éxito" y "fracaso". Por lo que es el espacio muestral para una prueba Bernoulli es:

$$S = \{ \text{éxito, fracaso} \}$$

Ejemplos:

Se tira una moneda:
$S = \{$ sol, águila $\}$

Un alumno hace un exámen:
$S = \{$ acredita, no acredita$\}$

Se dispara un proyectil
$S = \{$ Da en el blanco, no da en el blanco$\}$

Se prueba un medicamento:
$S = \{$ cura, no cura $\}$

Se practica una operación:
$S = \{$ Tiene éxito, no tiene éxito $\}$

En general la notación a emplear es la siguiente:

$P(\text{éxito}) = p$
$P(\text{fracaso}) = P(\text{no éxito}) = 1 - P(\text{éxito}) = q$

Se debe de notar que:

$$P(\text{éxito}) + P(\text{fracaso}) = P + (1 - P) = 1$$

VII.5.1. EL CASO DE LA FUNCION DE PROBABILIDAD BINOMIAL.

Definición:

Sea X el número de éxitos en n pruebas independientes y repetidas de ensayos Bernoulli, cada ensayo Bernoulli tiene probabilidad de éxito P. A la variable aleatoria X se le llama *BINOMIAL* con paramétros n (número de repeticiones de ensayos Bernoulli) y P (probabilidad de éxito en cada ensayo).

Como el número de éxitos para esta variable pueden ser:
0, 1, 2, 3, ... , n; entonces, X es una variable aleatoria discreta, y como tal debe de tener una función de probabilidad. Cuando se dice que tiene parámetros n, p quiere decir que si se conocen los valores de n y p entonces la función de probabilidad es totalmente conocida.

La expresión analítica de la función de probabilidades es:

$$P(X=x) = \begin{cases} \binom{n}{x} p^{x} q^{n-x} & \text{para } x = 0,1,2,3,\ldots,n \\ 0 \text{ en otro caso.} \end{cases}$$

n representa el número de veces que se realiza el experimento Bernoulli
(Recuérdese que estos experimentos son independientes, es decir, que el resultado de un ensayo no afecta el resultado de otro).

x representa el número de "éxitos" en los n ensayos.
p representa la probabilidad de "éxito"
q representa la probabilidad de "fracaso" = 1-p

$$\binom{n}{x} = \frac{n!}{x!\ (n-x)!} = \frac{n(n-1)(n-2)\ \ldots(n-x+1)}{x(x-1)(x-2)\ldots(1)}$$

El valor esperado de esta función es:

$$E(X) = n\ p$$

La varianza de la Binomial es:

$$Var_X = n\ p\ q$$

La desviación Estándar es:

$$D.S._X = \sqrt{n\ p\ q}$$

Ejemplo:

1) Se conoce que un 10 % de cierta población es miope. Si se selecciona una muestra de 15 personas al azar de esta población, calcule las probabilidades siguientes:

a) 4 o menos sean miopes.
b) 5 o más sean miopes.
c) Entre 3 y 6 inclusive sean miopes.
d) Cuántos miopes se espera que haya en una muestra de 30

SOLUCION:

Primero hay que definir cuál es el éxito, en este problema se llamará éxito a que una persona sea miope.

n=15
p(éxito) = 0.10 que es la proporción de miopes de la población.

como
p(fracaso)=1-P(éxito)= 1 - 0.10 = 0.90 = q

a) 4 o menos sean miopes:

$$P(X \leq 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(x=4)$$

$$P(X=0) = \binom{15}{0} (0.10)^0 (0.90)^{15} = .2059$$

$$P(X=1) = \binom{15}{1} (0.10)^1 (0.90)^{14} = .3431$$

$$P(X=2) = \binom{15}{2} (0.10)^2 (0.90)^{13} = .2669$$

$$P(X=3) = \binom{15}{3} (0.10)^3 (0.90)^{12} = .1285$$

$$P(x=4) = \binom{15}{4} (0.10)^4 (0.90)^{11} = .0429$$

$$\therefore P(X \leq 4) = .2059 + .2669 + .1285 + .0429 = 0.9873$$

b) 5 o más sean miopes:

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - .9873$$

c) Entre 3 y 6 inclusive sean miopes:

$$\begin{aligned} P(3 \leq X \leq 6) &= P(X=3) + P(X=4) + P(X=5) + P(X=6) \\ &= .1285 + .0429 + .0105 + .0019 \\ &= 0.1838 \end{aligned}$$

d) Cuántos miopes se espera que haya en la muestra de 30:
$E(X) = np \Rightarrow E(X) = 30\ (0.10) = 3$ miopes en promedio

VII.5.2. DISTRIBUCION DE POISSON.

La distribución de POISSON es otra distribución discreta, cuyo nombre se debe al matemático francés, Simeon Denis Poisson (1781-1840), quien la introdujo en 1837. Esta distribución tiene grandes usos en muchos campos, especialmente en Biología y Medicina.

Muchos problemas consisten en observar lo que se conoce como eventos discretos en un intervalo continuo (intervalo de tiempo, espacio, volumen, etc.).

Ejemplos:

El número de autos que llegan a un estacionamiento entre las 10 y las 11 de la mañana a un estacionamiento.

El número de personas que llegan a una caja de un supermercado entre las 11 y 12 hrs. de la mañana.

El número de llamadas telefónicas que recibe un conmutador entre la 1 y 3 de la tarde.

El número de defectos en un cable de 100 metros de longitud.

El número de partículas que hay en un determinado recipiente

DEFINICION:
Los eventos discretos que se generan en un intervalo continuo (tiempo, longitud, espacio, volúmen, etc) son procesos POISSON con parámetro λ, si cumplen con:

I) Se puede tomar un intervalo suficientemente corto de longitud h, tal que:

   i) La probabilidad de exactamente una ocurrencia en el intervalo es aproximadamente λh, y

ii) La probabilidad de 2 o más ocurrencias en el intervalo es aproximadamente 0.

II) La ocurrencia de un evento en un intervalo de longitud *h* no tiene efecto en la ocurrencia o no ocurrencia de otro intervalo no traslapado de longitud *h*. (Las ocurrencias son Estadísticamente independientes entre sí)

DEFINICION

Si se observa un proceso de POISSON con parámetro λ durante S unidades de tiempo, espacio, volúmen (o cualquier intervalo que se pueda considerar como continuo) y se define a X como el número de eventos que ocurren en dichas unidades, entonces se llama a X la variable aleatoria POISSON con parámetro λS.

TEOREMA:

Si X es una variable aleatoria de POISSON con parámetro λS, entonces:

$$P(X=k) \begin{cases} \dfrac{(\lambda S)^k}{k!} e^{-\lambda s}, & k=0,1,2,\ldots \\ 0 & \text{en otro caso} \end{cases}$$

Donde λS es el promedio de ocurrencias del evento aleatorio en el intervalo.

El símbolo e es una constante cuyo valor aproximado a 4 cifras decimales es 2.7183 (base de los logaritmos naturales).

Teóricamente es posible que el evento pueda ocurrir infinitas veces en el intervalo.

La probabilidad de que ocurra un evento en un intervalo es proporcional a la longitud del intervalo.

El valor esperado de esta variable aleatoria es:

$$E(X) = \lambda S$$

La varianza de una POISSON es:

$$Var_X = \lambda S$$

La desviación estándar es:

$$D.S._X = \sqrt{\lambda S}$$

Ejemplo:

a) Suponga que el número de muertes por suicidio en una ciudad es un proceso POISSON con parámetro $\lambda=2$ por día. Sea X el número de muertes por suicidio que ocurren en una semana. Entonces X es una variable aleatoria de Poisson con parámetro: $\lambda = 2$ y $S= 7$ días (una semana, intervalo continuo), por lo que $\lambda S = 2\ (7) = 14$

$$P(X=k) = \frac{(14)^k}{k!} e^{-14}, \quad k=0,1,2,.....$$

por lo que la probabilidad de que haya exactamente 14 suicidios en una semana es:

$$P(X=14) = \frac{(14)^{14}}{14!} e^{-14} = 0.106$$

b) Suponga que el número de llamadas que llegan a un conmutador telefónico de una industria es un proceso de POISSON con parámetro $\lambda = 120$ llamadas por hora. Sea X el número de llamadas que llegan en un periodo 1 minuto. Entonces X es la variable aleatoria Poisson con parámetro $\lambda= 120$ llamadas por hora y $S = \frac{1}{60}$ (intervalo continuo)

lo que indica que $\lambda S = 120 \left(\frac{1}{60}\right) = 2$

$$\therefore P(X=k) = \frac{(2)^k}{k!} e^{-2}, \quad k=0,1,2,.....$$

La probabilidad de que haya 0 llamadas en un minuto es:

$$P(X=0) = \frac{(2)^0}{0!} e^{-2} = 0.135$$

VII.5.3. Aproximación de la distribución BINOMIAL mediante la distribución POISSON.

Cuando el número de ensayos es muy grande, los cálculos de las probabilidades BINOMIALES se vuelven tediosos y largos, por lo que es conveniente tener una aproximación de estas probabilidades, una de ellas es la aproximación de POISSON.

La aproximación de la distribución BINOMIAL mediante la distribución POISSON se usa cuando n es muy grande (número de repeticiones de ensayos BERNOULLI independientes) y p (que es la probabilidad de éxito en cada ensayo BERNOULLI) muy pequeño, de tal manera que np (que es la esperanza de una función de probabilidad BINOMIAL) permanezca constante. Lo anterior se plasma en el siguiente teorema:

TEOREMA:

Si en una función de probabilidad BINOMIAL p (la probabilidad de éxito) en un ensayo es muy cercana cero y el número de ensayos n tiende a infinito (de tal manera que la media μ = np permanezca constante) entonces la distribución BINOMIAL se aproxima a la distribución de POISSON con media λ = np.

Ejemplo:

Supóngase que un conmutador de teléfonos maneja 300 llamadas en promedio durante una hora de actividad, y que el tablero puede hacer a lo más de 10 conexiones por minuto. Estimar la probabilidad de que el tablero este sobrecargado en un minuto dado.

Solución:

Como se está dando el promedio de llamadas en una hora, es decir ya se tiene λ = np = 300 por hora, y como el parámetro de la función de probabilidad POISSON es λS, entonces S debe de ser la medida por minuto por lo que:

$S = \frac{1}{60}$ lo que indica que $\lambda S = 300 \left\{ \frac{1}{60} \right\} = 5$ llamadas por minuto.

por lo que la probabilidad de que el tablero este sobrecargado equivale a que reciba más de 10 llamadas por minuto, en otras palabras se está preguntando que:

$$P(X > 10) = 1 - P (X \leq 10)$$

$$= 1 - \sum_{x=0}^{10} P (X = x)$$

$$= 1 - \sum_{x=0}^{10} \frac{(\lambda S)^x}{x!} e^{-\lambda S}$$

$$= 1 - \sum_{x=0}^{10} \frac{(5)^x}{x!} e^{-5}$$

$$= 1 - 0.986310$$

$$= 0.013690$$

VII.6. FUNCIONES PROBABILISTICAS CONTINUAS.

VII.6.1. INTRODUCCION:

Si la variable aleatoria es continua, tiene un número incontable de valores posibles. La función de distribución acumulada será continua en el sentido de que su gráfica es "suave", es decir no hay saltos en la probabilidad en cualquiera de los valores de la variable. Debido a la continuidad, la probabilidad de X en un punto cualquiera de sus valores posibles es CERO.

Por esta razón se define el concepto de función de densidad de probabilidades de la variable aleatoria X. Por lo tanto, la función densidad es una medida de concentración de probabilidad dentro de un intervalo.

Esta probabilidad puede interpretarse como una área bajo la curva, llamada curva de densidad, limitada por las ordenadas de los extremos del intervalo.

VII.6.2. DISTRIBUCION NORMAL.

En los problemas prácticos, la función de densidad Normal es una de las leyes de probabilidad más utilizada, la gráfica de la Normal tiene forma de campana, por lo que se conoce como campana de Gauss, es simétrica con respecto a μ, esto quiere decir que una variable aleatoria dsitribuida normalmente tiene mayor probabilidad de tomar un valor cercano a μ y una menor probabilidad de tomar valores alejados de μ (a cada lado).

La distribución Normal es la distribución continua más importante por las siguientes razones:

1) Muchas variables aleatorias que aparecen en relación con experimentos u observaciones prácticas están distribuidas normalmente.

2) Otras variables están distribuidas normalmente en forma aproximada.

3) Algunas veces una variable aleatoria NO está distribuída normalmente, ni siquiera en forma aproximada, pero se puede convertir en una variable con distribución normal por medio de una transformación sencilla.

4) Ciertas distribuciones se pueden aproximar mediante la distribución normal (como el caso de las distribuciones Binomial y de Poisson).

5) En Estadística teórica, muchos problemas pueden ser resueltos fácilmente en el supuesto de una población normal. En el trabajo aplicado se encuentran que métodos más elaborados según la ley de probabilidades normal dan resultados satisfactorios, aunque no se cumpla totalmente el supuesto de una población normal.

6) Ciertas variables que son básicas para justificar pruebas Estadísticas estan distribuidas normalmente.

La definición formal de esta función (expresión análitica):

Definición: Una variable aleatoria X esta distribuida normalmente si y sólo sí su función de densidad de probabilidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{para toda x real.}$$

donde μ, σ son respectivamente, la media y la desviación estándar de la variable aleatoria X.

El parámetro μ puede ser igual a cualquier número real.
El parámetro σ debe ser un número real positivo.

La función de distribución acumulada es:

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx.$$

Esta expresión analítica aparentemente tan complicada, especialmente si se desean calcular probabilidades con ella, puesto que su función de distribución acumulada es una integral que no se puede evaluar por métodos elementales, no debe de preocupar porque se pueden calcular fácilmente usando las tablas que se han publicado para tal efecto. Pero antes de esto señalarán algunas propiedades geométricas.

Es importante familiarizarse con las propiedades geométricas de la densidad Normal:

1) La curva que representa a f(x) se denomina "*campana de Gauss*" y es simétrica con respecto a μ, X toma valores entre $(-\infty, \infty)$

2) En los valores $\mu-\sigma$, $\mu+\sigma$ ocurren puntos de inflexión.

3) El área total bajo la curva es 1.
$P(\mu-\sigma < X < \mu+\sigma) \cong 0.680$
$P(\mu-2\sigma < X < \mu+2\sigma) \cong 0.995$
$P(\mu-3\sigma < X < \mu+3\sigma) \cong 0.997$

4) La curva es asintótica con el eje X.

5) Como el exponente $-\frac{1}{2}\left\{\frac{X-\mu}{\sigma}\right\}^2$ de e es negativo, cuanto mayor es la desviación de X con relación a μ, tanto menor es la densidad de probabilidad de X, f(X).

6) Un cambio en el valor μ desplaza la distribución normal a la izquierda o a la derecha del eje X (abscisas).

7) Realmente no existe la distribución normal sino infinitas distribuciones normales, para cada pareja de valores de μ y de σ existe una distribución normal; pero se puede establecer una relación entre ellas.

Valores Normalizados Z y área bajo la curva.

Definición:

A la distribución normal cuya media es cero y cuya desviación estándar es 1 se le conoce como DISTRIBUCION NORMAL ESTÁNDAR y para distinguirla de las demás, se utiliza la letra Z para ella.

La función de densidad es:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad ; \quad -\infty < z < \infty$$

VII.6.3. APROXIMACION NORMAL A LA DISTRIBUCION BINOMIAL.

INTRODUCCION.

Se sabe que cuando n es muy grande los cálculos de las probabilidades binomiales resultan muy laboriosos debido a la obtención de los coeficientes binomiales y a las potencias de p y de q. Por ello se vió como aproximar estas probabilidades por medio de la distribución de Poisson cuando n es muy grande y p muy pequeño. Ahora se verá que la distribución normal resulta una buena aproximación de las distribuciones binomiales cuando n es grande y p no necesariamente pequeño.

La aproximación será conveniente siempre y cuando:

a) $n \to \infty$ (n tiende a ser muy grande)

b) $p \to 0.5$ aún cuando n sea pequeño

c) $p < \frac{1}{2}$ y $np > 5$

d) $p > \frac{1}{2}$ y $nq > 5$

En general, para pasar de una distribución binomial a normal se usa un factor de conversión de una variable discreta a continua de 0.5

i) $$P(X=k) \cong P\left\{ \frac{k - 0.5 - np}{\sqrt{n\,p\,q}} < Z < \frac{k + 0.5 - np}{\sqrt{n\,p\,q}} \right\}$$

ii) $$P(X \geq k) \cong P\left\{ Z > \frac{k - 0.5 - np}{\sqrt{n\,p\,q}} \right\}$$

iii) $$P(X \leq k) \cong P\left\{ Z < \frac{k + 0.5 - np}{\sqrt{n\,p\,q}} \right\}$$

iv) $$P(X > k) \cong P\left\{ Z > \frac{k + 0.5 - np}{\sqrt{n\,p\,q}} \right\}$$

v) $$P(X < k) \cong P\left\{ Z < \frac{k - 0.5 - np}{\sqrt{n\,p\,q}} \right\}$$

VII.6.4. DISTRIBUCIONES MUESTRALES MUESTRALES Y EL TEOREMA CENTRAL DEL LIMITE.

VII.6.4.1. DISTRIBUCIONES MUESTRALES.

Se llama distribución muestral de un estadístico a la distribución de probabilidad de todos los valores posibles que pueden ser tomados por dicho estadístico, calculados a partir de muestras del mismo tamaño extraidos aleatoriamente de una población.

Para construir una distribución muestral se procede de la siguiente manera:

1) De una población finita discreta, de tamaño N, se extraen aleatoriamente todas las muestras de tamaño n. En general para muestreo con reemplazo, el número de muestras posibles de tamaño n de una población de tamaño N, es igual a N^n, y para muestreo sin reemplazo e ignorando el orden, el número de muestras de tamaño n son las combinaciones de $\binom{N}{n}$

2) Se calcula el estadístico de interés para cada muestra.

3) Se construye una tabla de los valores del estadístico de interés con sus respectivas probabilidades.

4) Se calculan los parámetros μ y σ^2 de la distribución.

Por ejemplo:

Considérese la población formada por los números 5,7,8 y 10 cuya media $\mu=7.5$ y varianza $\sigma^2= 3.25$, ahora determinense todas las muestras de tamaño 2 sin reemplazo e ignorando el orden de la población, por lo que se obtiene la siguiente distribución muestral para la media $\bar{x}$:

muestra	$\bar{X}$
5,7	6
5,8	6.5
5,10	7.5
7,8	7.5
7,10	8.5
8,10	9

$\bar{X}$	f	$\bar{f}(X)$
6	1	1/6
6.5	1	1/6
7.5	2	2/6
8.5	1	1/6
9	1	1/6

$\mu_{\bar{x}} = \sum \bar{x}\ f(\bar{x}) = 7.5$

$\sigma^2_{\bar{x}} = \bar{x}^2\ f(\bar{x}) - \mu^2_{\bar{x}} = 57.33 - 56.25 = 1.0833$

Se puede observar que $\mu_{\bar{x}} = \mu$; $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\ \frac{N-n}{N-1}$

Si se tomaran a todas las muestras de tamaño 2 con reemplazo se obtendría exactamente el mismo resultado para la media y para la desviación estándar se tiene: $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

Se define como *error estándar* a la desviación estándar de la distribución muestral de un estadístico, así $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ es el error estándar de la media o simplemente error estándar, cuando el muestreo se hace con reemplazo.

También se puede tener el error estándar de la varianza o de la mediana o de la proporción, etc

$$\sigma_{s^2}\ ;\ \sigma_{Md}\ ;\ \sigma_p$$

VII.6.4.2. EL TEOREMA CENTRAL DEL LIMITE.

Dada una población cualquiera con media μ, y varianza σ^2. La distribución muestral de $\bar{X}$, calculada a partir de todas las muestras aleatorias de tamaño n con reemplazo de esta población, estará distribuida en forma aproximadamente normal con media $\mu_{\bar{x}} = \mu$; y varianza $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ y la aproximación a la normal será mejor cuanto mayor sea n.

VII.6.4.3. DISTRIBUCIÓN T - STUDENT

Una de las funciones muestrales es la t-student que publicó W. S. Gosset en 1908 con el pseudónimo de "Student", el teorema que establece esta distribución es:

TEOREMA:

Si Z es una variable aleatoria que se distribuye como Normal (0,1), y si U es una variable aleatoria $\chi^2(r)$, y si Z y U son independientes, entonces:

$$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{s} ;$$

tiene una distribución t-student con n-1 grados de libertad.

Esta distribución tiene las siguientes propiedades:

i) Su media es $\mu_t = 0$.

ii) Es simétrica con respecto a μ_t

iii) En general, su varianza σ^2_t es mayor que 1, pero se aproxima a 1 cuando n es grande.

iv) La variable t toma valores de $(-\infty, \infty)$

v) La distribución t es realmente una familia de distribuciones (curvas), ya que hay una distribución diferente para cada valor de los grados de libertad v=n-1.

vii) La distribución t se aproxima a la normal cuando n-1 se aproxima a ∞ .

VII.7. INFERENCIA ESTADISTICA

Como inferencia Estadística se entiende al proceso de generalización de los resultados hallados en la muestra a la población en estudio, el grado de generalidad estará determinado por la validez externa, a menor precisión de la población en estudio mayor generalidad, a mayor precisión de la población en estudio menor grado de generalidad. La inferencia se divide en dos grandes rubros a)estimación de parámetros poblacionales, b) prueba de hipótesis; a su vez la estimación de parámetros se divide en estimación puntual y en estimación por intervalos. La inferencia Estadística se puede abordar con técnicas paramétricas, no paramétricas, bayesianas, univariadas y multivariadas.

VII.7.1. ESTIMACION

INTRODUCCION:

Supongamos que se quiere saber cuál es la estatura promedio de una población, o saber cuál es el contenido promedio de un cierto líquido en la producción del producto X, o saber cuál es el peso promedio de los niños recién nacidos, o saber cuál es la proporción de la población que tiene el tipo de sangre O RH negativo, etc.

Para resolver este tipo de problemas es necesario tomar muestras aleatorias de la población en estudio ya que no siempre es posible hacer censos (censo: medición de la característica en cada elemento de la población en estudio) algunas veces por costosos, otras por la imposibilidad de obtención de las unidades muestrales, otras por tiempo etc.

La muestra aleatoria de la población objetivo sirve para encontrar valores aproximados (parámetros) de la población, estos valores por ser de una muestra no son iguales a los valores de la población, sin embargo nos sirven para *estimar* a los verdaderos.

Por lo que al proceso de tomar una muestra aleatoria de una población objetivo y a partir de esta se puede decir cuales son los valores poblaciones se le conoce como ESTIMACION.

La ESTIMACION puede ser de dos tipos: puntual cuando sólo se da un valor y por intervalo cuando se asocia una probabilidad de que el valor de la muestra sea el valor de la población.

Un ESTIMADOR es un valor aproximado de un parámetro poblacional, determinado de los estadísticos muestrales obtenidos de la población

VII.7.1.1. ESTIMACIÓN POR INTERVALO

INTRODUCCION.

El interés es encontrar un intervalo (probabilístico) en el cual se encuentre el parámetro buscado, este intervalo tendrá asociado una probabilidad y está es la probabilidad de que el parámetro deseado este en el intervalo encontrado.

Ejemplo:

Sea una variable aleatoria X distribuida en forma NORMAL con media μ y desviación estándar σ, calcular:

a) $P(\mu - \sigma \leq x \leq \mu + \sigma)$
b) $P(\mu - 2\sigma \leq x \leq \mu + 2\sigma)$
c) $P(\mu - 3\sigma \leq x \leq \mu + 3\sigma)$

Respuesta:

a) Si $x = \mu - \sigma$ estándarizando se tiene:

$$Z = \frac{(\mu - \sigma) - \mu}{\sigma} = -1$$

Si $x = \mu + \sigma$

$$Z = \frac{(\mu + \sigma) - \mu}{\sigma} = +1$$

$\therefore P(\mu - \sigma \leq x \leq \mu + \sigma)$

$= P(-1 \leq z \leq 1) = 0.6826$

b) Si $x = \mu - 2\sigma$ estándarizando se tiene:

$$Z = \frac{(\mu - 2\sigma) - \mu}{\sigma} = -2$$

Si $x = \mu + 2\sigma$ estándarizando se tiene:

$$Z = \frac{(\mu + 2\sigma) - \mu}{\sigma} = +2$$

$\therefore P(\mu - 2\sigma \leq x \leq \mu + 2\sigma)$

$= P(-2 \leq z \leq 2) = 0.9544$

c) Si $x = \mu - 3\sigma$ estándarizando se tiene:

$$Z = \frac{(\mu - 3\sigma) - \mu}{\sigma} = -3$$

Si $x = \mu + 3\sigma$ estándarizando se tiene:

$$Z = \frac{(\mu + 3\sigma) - \mu}{\sigma} = +3$$

$\therefore P(\mu - 3\sigma \leq x \leq \mu + 3\sigma)$

$= P(-2 \leq z \leq 2) = 0.9972$

Si se a t un estadístico muestral cualquiera y si la distribución del estadístico es NORMAL, o aproximadamente NORMAL con media μ_t y desviación estándar σ_t, se tendrá de acuerdo a lo anterior que:

a) $P(\mu_t - \sigma_t \leq t \leq \mu_t + \sigma_t)$

b) $P(\mu_t - 2\sigma_t \leq t \leq \mu_t + 2\sigma_t)$

c) $P(\mu_t - 3\sigma_t \leq t \leq \mu_t + 3\sigma_t)$

Respuesta:

a) Si $t = \mu_t - \sigma_t$ estándarizando se tiene:

$$Z = \frac{(\mu_t - \sigma_t) - \mu_t}{\sigma_t} = -1$$

Si $t = \mu_t + \sigma_t$

$$Z = \frac{(\mu_t + \sigma_t) - \mu_t}{\sigma_t} = +1$$

$\therefore P(\mu_t - \sigma_t \leq t \leq \mu_t + \sigma_t) = P(-1 \leq z \leq 1) = 0.6826$

b) Si $t = \mu_t - 2\sigma_t$ estándarizando se tiene:

$$Z = \frac{(\mu_t - 2\sigma_t) - \mu_t}{\sigma_t} = -2$$

Si $t = \mu_t + 2\sigma_t$ estándarizando se tiene:

$$Z = \frac{(\mu_t + 2\sigma_t) - \mu_t}{\sigma_t} = +2$$

$\therefore P(\mu_t - 2\sigma_t \leq t \leq \mu_t + 2\sigma_t) = P(-2 \leq z \leq 2) = 0.9544$

c) Si $t = \mu_t - 3\sigma_t$ estándarizando se tiene:

$$Z = \frac{(\mu_t - 3\sigma_t) - \mu_t}{\sigma_t} = -3$$

Si $t = \mu_t + 3\sigma_t$ estándarizando se tiene:

$$Z = \frac{(\mu_t + 3\sigma_t) - \mu_t}{\sigma_t} = +3$$

$$\therefore P(\mu_t - 3\sigma_t \leq t \leq \mu_t + 3\sigma_t) = P(-3 \leq z \leq 3) = 0.9972$$

Si se requiere estimar por medio de un intervalo probabilístico al valor promedio de la población μ, tomando como base al intervalo:

$$(\mu_t - \sigma_t \leq t \leq \mu_t + \sigma_t)$$

y de este se despeja a μ_t

$$(t - \sigma_t \leq \mu_t \leq t + \sigma_t)$$

que es equivalente al intervalo anterior, por lo que puede decirse que en el 68.26% de los casos la media de la distribución de t está contenida en dicho intervalo.

A este intervalo se le llama el intervalo de confianza de la media de la distribución muestral del estadístico t al 68.26% de nivel de confianza. Los extremos del intervalo son los límites de confianza de la estimación de μ_t

Otra interpretación equivalente es se tiene una confianza del 68.26% de que la media μ_t esté comprendida dentro de los límites del propio intervalo.

Otra interpratación equivalente es que se puede asegurar que la media μ_t *siempre* estará dentro de los límites de confianza del intervalo, aceptando que en el "siempre" se estará errando en el:

(100 - 68.26)% =31.74% de las veces.

En forma semejante a lo hecho con:

$$P(\mu_t - \sigma_t \leq t \leq \mu_t + \sigma_t) \cong P(-1 \leq z \leq 1)$$
$$\cong P(t - \sigma_t \leq \mu_t \leq t + \sigma_t) = 0.6826$$

se puede hacer con los otros intervalos:

$$P(\mu_t - 2\sigma_t \leq t \leq \mu_t + 2\sigma_t) = P(-2 \leq z \leq 2)$$
$$\cong P(t - 2\sigma_t \leq \mu_t \leq t + 2\sigma_t) = 0.9544$$

$$P(\mu_t - 3\sigma_t \leq t \leq \mu_t + 3\sigma_t) \cong P(-3 \leq z \leq 3)$$
$$\cong P(t - 3\sigma_t \leq \mu_t \leq t + 3\sigma_t) \cong 0.9972$$

Procediendo en forma similar, puede encontrarse que los límites del intervalo de la μ_t al 95% y 99% de confoanza son:

$t \pm 1.96\ \sigma_t$ y $t \pm 2.58\ \sigma_t$ respectivamente.

En general, el intervalo de confianza de μ_t a cualquier nivel de confianza es:

$$(t - Z_c\ \sigma_t \leq \mu_t \leq t + Z_c\ \sigma_t)$$

en donde Z_c depende del nivel de confianza considerado. A Z_c se le llama el coeficiente de confianza o valor crítico.

INTERVALOS DE CONFIANZA DE LA MEDIA DE LA POBLACION.

Si el estadístico t que se mencionó en la sección anterior representa a la media muestral $\bar{x}$, puede aceptarse que tiene distribución NORMAL, o aproximadamente normal, siempre que la población tenga también ditribución normal o el tamaño n de las muestras sea grande, respectivamente. En cualesquiera de estos casos, y de acuerdo con $(- Z_c \sigma_t \leq \mu_t \leq t + Z_c \sigma_t)$, se tendrá que el intervalo de confianza de la media de la distribución de medias de las muestras es:

$$(\bar{x} - z_c \sigma_{\bar{x}} \leq \mu_{\bar{x}} \leq \bar{x} + z_c \sigma_{\bar{x}})$$

Pero se sabe que $\mu_{\bar{x}} = \mu$ y $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ si el muestreo es con reemplazo o la la población es infinita. Luego se obtiene:

$$\left(\bar{x} - Z_c \frac{\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{x} + Z_c \frac{\sigma_x}{\sqrt{n}} \right)$$

que es el intervalo de confianza de la media de la población para el caso en que la población es NORMAL o el tamaño de las muestras es grande, y el muestreo es con reemplazo o la población es infinita.

Ejemplo:

Supóngase que un investigador está interesado en estimar el nivel medio de alguna enzima en una cierta población, toma una muestra de 10 individuos, determina el nivel de la enzima en cada uno y calcula la media de la muestra que es igual a 22. Supóngase además que se sabe que la variable de interés está distribuida en forma aproximadamente normal con varianza σ^2 de 45. Entonces el intervalo de confianza del 95% para μ es:

$$P\left(22 - 1.96 \sqrt{\frac{45}{10}} < \mu < 22 + 1.96 \sqrt{\frac{45}{10}} \right) = 0.95$$

$$p(22 - 4.1578 < \mu < 22 + 4.1587) = 0.95$$

$$p(17.84 < \mu < 26.1578) = 0.95$$

Si la población es NORMAL o el tamaño de las muestras es grande, y el muestreo se realiza sin reemplazo de una población finita de tamaño N, el intervalo de confianza para la media de la población es:

$$\left(\bar{x} - Z_c \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq \mu_x \leq \bar{x} + Z_c \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right)$$

En cualesquiera de los casos tratados, si no se conoce la desviación estándar σ_x de la población, que es lo natural, es posible sustituirla por sus estimadores S_x o $\hat{S}_x$, siempre que n sea grande.

Si el tamaño de la muestran n no es grande, la aproximación que se obtiene al sustituir σ_x por alguno de sus estimadores es pobre y NO debe de hacerse. En este caso, y siempre que la población sea normal, puede obtenerse otra expresión para la estimación de la media de la población por medio de la distribución t de Student:

$$\left(\bar{x} - t_c \frac{S_x}{\sqrt{n-1}} \leq \mu_x \leq \bar{x} + t_c \frac{S_x}{\sqrt{n-1}} \right)$$

que es el intervalo de confianza para la media cuando la población es normal, el muestreo es con reemplazo o la población es infinita, y el tamaño de la muestra es cualquiera. (grande o pequeño).

VII.7.2. PRUEBAS DE HIPOTESIS ESTADISTICAS.

Al proceso de generalización de los resultados hallados en una muestra a la población se le conoce como inferencia Estadística, la inferencia tiene dos propósitos generales, el de estimación y el de la prueba de hipótesis.

Hipótesis Estadística es una afirmación que se hace acerca de la distribución de una variable aleatoria. En forma equivalente se puede afirmar que una hipótesis Estadística es una afirmación o conjetura acerca de una o más poblaciones.

Una prueba Estadística es una regla de decisión que permite aceptar o rechazar una hipótesis Estadística con base en la evidencia de los datos muestrales.

Al subconjunto C del espacio muestral (conjunto de todos los posibles resultados de un experimento aleatorio) que permita rechazar una hipótesis de acuerdo con la regla de decisión preestablecida se conoce como región crítica.

Nunca se sabe con absoluta certeza la verdad o falsedad de una hipótesis Estadística, a no ser que se examine a la población entera. En lugar de tomar a la población, se toma una muestra aleatoria y se utilizan los datos que contiene tal muestra para proporcionar evidencias que confirmen o no la hipótesis. La evidencia de la muestra que es inconsistente con la hipótesis planteada conduce a un rechazo de la misma, mientras que la evidencia que apoya la hipótesis conduce a su aceptación.

En las pruebas de hipótesis se distinguen dos tipos de hipótesis: la *hipótesis nula*, designada por H_0 , conocida por hipótesis de no diferencia, que se establece, en principio, con el único fin de rechazarla o anularla; *hipótesis alternativa* (H_1), que es cualquier suposición que difiera de la nula. Las hipótesis nula y alternativa deben de ser mutuamente exclusivas.

TIPOS DE PRUEBA

Hay dos pruebas, bilaterales y unilaterales; las bilaterales o de dos colas cuya forma general es:

$$H_0: \Theta = \Theta_0$$

$$H_1: \Theta \neq \Theta_0$$

Donde Θ representa cualquier parámetro y Θ_0 es un valor supuesto del parámetro.

Las pruebas unilaterales o de una cola, pueden ser: a) unilateral de cola izquierda, cuya forma general es:

$$H_o: \Theta \geq \Theta_o$$

$$H_1: \Theta < \Theta_o$$

b) unilateral de cola superior o de cola derecha de la forma:

$$H_o: \Theta \leq \Theta_o$$

$$H_1: \Theta > \Theta_o$$

VII.7.3. TIPO DE ERRORES.

El procedimiento de probar una hipótesis nula contra una alternativa sobre la base de información obtenida de la muestra, conducirá a dos tipos de errores posibles, debido a fluctuaciones al azar en el muestreo. Si la hipótesis nula en realidad es cierta, pero los datos de la muestra son incompatibles con ella y se rechaza, se comete un *error de tipo I*; es decir, se comete cuando se rechaza una hipótesis nula verdadera. Por otro lado, se comete un *error de tipo II* cuando se acepta una hipótesis nula falsa. En resumen se tiene:

Decisiones

eventos	Aceptar	Rechazar
H_o verdadera	correcta	error tipo I
H_o falsa	Error tipo II	Correcta

Las probabilidades de cometer errores de tipo I y II, pueden considerarse como los riesgos de decisiones incorrectas. La probabilidad de cometer un error tipo I se llama nivel de significancia o de significación y se designa por α; la probabilidad de cometer un error de tipo II no tiene un nombre en particular, pero se representa por β.

$$P(\text{error tipo I}) = P(\text{rechazar } H_o \mid H_o \text{ verdadera}) = \alpha$$

$$P(\text{error tipo II}) = P(\text{aceptar } H_o \mid H_o \text{ falsa}) = \beta$$

ESTADISTICA DE PRUEBA

Puesto que la decisión de aceptar H_0 ó H_1 se hace con base en la muestra, es necesario escoger una función de las n observaciones como Estadística de prueba. En general, el estadístico de prueba debe de ser uno cuya distribución muestral sea conocida en el supuesto de que H_0 sea cierta. El estadístico de prueba generalmente resulta el estimador convencional del parámetro previsto en H_0.

Los cuatro pasos básicos que se realizan en la contrastación de una hipótesis son:

1) Formulación de hipótesis nula (H_0) y alternativa (H_1).
2) Determinación de criterio, prueba y contraste
3) Obtención de los datos muestrales.
4) Toma de decisión e interpretación.

VII.7.3. PRUEBAS DE SIGNIFICANCIA.

Introducción.

Cuando se hacen investigaciones para conocer características de una población en estudio, medir cada una de las características en toda la población no en pocas ocasiones es imposible, pero aún en el caso de la posibilidad, resulta muy costoso por lo que una forma de conocer las características es a través de muestras, si estas se realizan en forma adecuada entonces los valores encontrados en las muestras pueden generalizarse a la población en estudio. A esta generalización se conoce como INFERENCIA ESTADISTICA y se realiza con pruebas de hipótesis. La inferencia es abordada en formas diferentes dependiendo de los supuestos que se hagan sobre la población objetivo.

VII.7.3.1. COMPARACION DE DOS MEDIAS MUESTRALES

PRUEBA DE T DE STUDENT

Esta prueba es paramétrica y sirve para confirmar si dos muestras son Estadísticamente diferentes.[1]

Las condiciones que deben de verificarse para la aplicación de la prueba t son que las dos muestras deben de ser tomadas en forma aleatoria ser independientes y estar idénticamente distribuidas. Las variables deben de ser continuas (intervalo y de razón) con distribución Normal.

La Estadística de prueba para la hipótesis nula dependerá de los supuestos que se hagan sobre las desviaciones estándar de cada una de las muestras.

[1] Es importante señalar que, cuando se tienen más de dos muestras la prueba t-Student no debe de aplicarse, debido a que el nivel de significancia crece de manera importante y este no puede ser controlado.

Básicamente se tienen los siguientes casos

i) La desviación estándar de la población es desconocida

ii) Las desviaciones estándares de las dos poblaciones son iguales y desconocidas.

iii) Las desviaciones estándares de las dos poblaciones son diferentes y desconocidas

Caso 1

o) Se tiene una sola muestra de una población normal con desviación estándar desconocida (cuando se conoce la desviación estándar de la población se realiza una prueba normal) y se requiere probar la hipótesis nula de que la media es igual a una constante se procede de la siguiente forma:

a) Se calcula la Estadística de prueba:

$$t = \frac{\bar{X} - c}{\frac{S}{\sqrt{n}}} \quad \text{con n-1 grados de libertad.}$$

Donde c es la constante a comparar

S es la desviación estándar de la muestra

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} \quad ; \quad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

n el tamaño de la muestra

$\bar{X}$ es la media aritmética de la muestra.

b) La significancia del estadístico de prueba se realiza comparando el valor obtenido con el valor de tablas de una t-Student con n-1 grados de libertad y un valor predeterminado de confianza (α).

Ejemplo:

Para las muestras de migrantes se desea probar que la edad promedio de las edades de adultos en las poblaciones es de 37 años, Se tomó una muestra de 45 familias de migrantes y se obtuvo que el promedio aritmético de las edades de los adultos fue de 39 años con una desviación estándar de 7 años.

Prueba:

$$t = \frac{39 - 37}{\frac{7}{\sqrt{45}}} \quad , \; gl = 45 - 1 = 44$$

$$t = 1.9166$$

Interpretación:

como t =1.9166 es menor que una t de tablas con α =0.05 de confianza y 44 g.l. = 2.41 se acepta la hipótesis de que las medias son iguales. Por lo que se afirma, que existe evidencia Estadística en los datos que apoyan la hipótesis nula de que la edad promedio de los adultos de familias de migrantes es de 37 años.

Caso 2

i) Cuando se tienen dos muestras, se pueden comparar los valores promedio (medias aritméticas) para probar la hipótesis nula de que las dos muestras pertenecen a una misma población con el supuesto adicional de que las desviaciones estándar de cada una de las poblaciones son desconocidas pero iguales.

el procedimiento para este primer caso es el siguiente:

a) Se calcula la desviación estándar común de las muestras

$$S = \sqrt{\frac{(n-1)\ S_1^2 + (n-1)\ S_2^2}{n_1 + n_2 - 2}}$$

b) Se calcula la Estadística de prueba:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

c) Se determinan los grados de libertad:

$$g.l. = n_1 + n_2 - 2 \text{ grados de libertad}$$

d) La significancia del estadístico de prueba se realiza comparando el valor obtenido con el valor de tablas de una t-Student con los grados de libertad encontrados y un valor predeterminado de confianza (α).

Ejemplo:

Se tienen dos muestras de dos grupos de familias de migrantes (35 Guatemaltecas, 28 Salvadoreñas) y se sabe que el ingreso promedio por familia es de $175000 ± 17500, $275000 ± 27500 respectivamente, se supone que las dos poblaciones tienen igual varianza (aunque se desconoce el monto de la misma). Se desea corroborar la hipótesis de que las poblaciones de donde se obtuvieron las muestras, tienen los mismos ingresos por lo que se realiza una prueba Estadística:

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 \neq \mu_2 \Leftrightarrow \mu_1 - \mu_2 \neq 0$$

se calcula el estadístico de prueba:

t=0.8474
gl= 61

Interpretación se tiene que la t de tablas con 61 g.l. y $\alpha = 0.05$ de confianza es igual a 2.003, por lo que se toma la decisión de aceptar la hipótesis de que los ingresos promedio de las familias son iguales.

Caso 3

ii) Se tienen dos muestras aleatorias y se supone que las desviaciones estándar de cada una de las poblaciones son desconocidas pero diferentes (cuando las desviaciones estándar de cada una de las poblaciones son conocidas se realiza una prueba Normal), se procede de la forma siguiente:

a) se calculan los grados de libertad

$$g.l. = \frac{\left\{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]\right\}}{\frac{\left\{\frac{s_1^2}{n_1}\right\}^2}{n_1 + 1} + \frac{\left\{\frac{s_2^2}{n_2}\right\}^2}{n_2 + 1}} - 2$$

b) Se calcula el estadístico de prueba t

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

c) La significancia del estadístico de prueba se realiza comparando el valor obtenido con el valor de tablas de una t-Student con los grados de libertad mencionados y un valor predeterminado de confianza (α).

Ejemplo:

eSe tienen dos muestras de dos grupos de familias de migrantes (35 Guatemaltecas, 28 Salvadoreñas) y se sabe que el ingreso promedio por familia es de $375000 ± 37500, $301000 ± 30100 respectivamente, se supone que las dos poblaciones tienen diferente varianza (aunque se desconoce el monto de la misma).

Se desea corroborar la hipótesis de que las poblaciones de donde se obtuvieron las muestras, tienen los mismos ingresos por lo que se realiza una prueba Estadística:

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 \neq \mu_2 \Leftrightarrow \mu_1 - \mu_2 \neq 0$$

el nivel de significancia es $\alpha = 0.05$

La Estadística de prueba es:
t= 0.016
gl= 46

Interpretación:

Checando en tablas se tiene que t con 46 g.l. y $\alpha = 0.05$ =2.01 como t=0.016 es menor a tablas se acepta la hipótesis nula de que los ingresos promedio de las familias son Estadísticamente iguales.

BIBLIOGRAFIA:

BLALOCK, HUBERT M.
1981 Estadística Social, Fondo de Cultura Económica, México.

BRIONES, GUILLERMO.
1982 Métodos y Técnicas de Investigación para las Ciencias Sociales, Trillas, México.

CAMBELL, DONALD T.
Diseños Experimentales y Cuasiexperimentales en la Investigación Social, Amorrourtu editores, Buenosa Aires.

CHAMBERS, JOHN.
1983 Graphical Methods for Data Analysis, Bell Laboratories, Belmont California.

Des, Raj.
1979 La estructura se las encuestas por muestreo, Fondo de Cultura Económica, México.

1979 Teoría del muestreo, Fondo de cultura Económica, México.

GALTUNG, JOHAN.
1973 Teoría y Métodos de la Investigación Social, Editorial Universitaria de Buenos Aires, Argentina.

GARCIA ROMERO, JAIME.
1984 Manual de estadística aplicada a la salud, Universidad Nacional Autónoma de México, México.

GRIEGO, RICARDO.
1974 Conceptos de Probabilidad-I, Fondo de Cultura Económica, México.

HOAGLIN, DAVID C.
1983 Understanding Robust and Exploratory Data Analysis, Jhon Wiley & Sons, Inc, Harvard University, U.S.A.

HOGG, ROBERT V.
1978 Introduction to Mathematical Statistics, Macmillan Publishing Company, New York, U.S.A.

Hogg, Robert V.
1989 Probability and Statistical Inference, Macmillan Publishing Company, New York.

Hurst Thomas, David.
1976 Figuring anthropology first principles of probability and sattistics, Holt, Rinehart and Wiston, New York.

Larson, Harold J.
1986 Introducción a la teoría de probabilidades e inferencia estadística, Limusa, México.

Leach, Chris.
1982 Fundamentos de estadística enfoque no paramétrico para las ciencias sociales, Limusa, México.

MARQUES, MARIA JOSE.
1988 Probabilidad y Estadistica para Ciencias Químico-Biológicas, Universidad Nacional Autónoma de México. México

Méndez Ramirez, Ignacio
1984 El protocolo de investigación, lineamientos para su elaboración y análisis, Trillas, México

Montemayor García, Felipe.
1973 Fórmulas de estadística para investigadores, Instituto Nacional de Antropología e Historia, México.

MOOD, ALEXANDER M.
1963 Introduction to the Theory of Statistics, Mc Graw Hill, U.S.A.

Nadelsticher mitranai, ABRAHAM.
1983 Técnicas para la Construcción de cuestionarios de actitudes de opción múltiple, Instituto de Cinecias Penales, México.

OLIVERA, ANTONIO.
1979 Serie de Probabilidad y Estadística, IMPOS estitores, México.

PARZEN, EMANUEL.
1979 Teoría Moderna de Probabilidades y sus Aplicaciones, Limusa, México.

PEREZ, LUIS A.
1981 Estadística Matemática, Centro de Estudios Avanzados del I.P.N., México.

RUIZ Moncayo, ALBERTO.
1973 Introducción y Métodos de Probabilidad, Trillas, México.

TUKEY, JOHN W.
1978 Exploratory Data Analysis, Addison Wesley Publishing Company, Princeton University.