INSTITUTO SUPERIOR DE CIENCIAS MEDICAS

DR. "SERAFIN RUIZ DE ZARATE DUIZ"

FACULTAD DE ESTOMATOLOGIA

BIOESTADISTICA PARA ESTOMATOLOGOS

SEGUNDA PARTE

ESTADISTICA INFERENCIAL

AUTORES:

MsC. Juan Miguel Mirabal Díaz

MsC. Pedro Artiles González

PROLOGO

Las Ciencias Médicas actualmente tienen un considerable números de aspectos de actividad dentro de las cuales se incluyen: la prevención, curación, la atención comunitaria, el diseño, planificación, ejecución y evaluación de programas educativos que permitan una sociedad de hombres sanos física y espiritualmente.

Este conjunto de acciones de los profesionales de la salud, necesitan del uso de indicadores que nos permitan conocer, evaluar y comparar nuestra realidad en los distintos grupos etéreos, étnicos, en los diferentes unidades asistenciales y otros países del mundo.

Lo planteado anteriormente plantea la necesidad de aplicación de métodos investigativos y métodos estadísticos. Esto lleva a la necesidad de preparar a los futuros profesionales de la salud para cumplir esta vital función investigativa en el pregrado, en postgrado y en su vida profesional futura.

Este material docente tiene como objetivo principal lograr una adecuada preparación de los profesionales de la salud vinculando los métodos estadísticos con la aplicación de las nuevas tecnologías de la Informática y las Comunicaciones mediante el uso de paquetes estadísticos profesionales y el mismo esta divido en dos partes que son:

Primera Parte: Estadística Descriptiva.

Segunda Parte:Estadística Inferencial

El colectivo de autores pertenecientes al Instituto Superior de Ciencias Médicas “Dr. “Serafín Ruiz de Zarate Ruiz” agradece cualquier critica o sugerencia que permita mejorar o ampliar estos materiales.

Colectivo de Autores.

CONTENIDOS

Capítulo1:Nociones sobre la teoría de probabilidades.

1.1. Experimentos aleatorios.
1.2. Concepto de Probabilidad.
1.3. Probabilidad Condicionada.
1.4. Teorema de Bayes
1.5. Variable aleatoria.
1.6. Modelos teóricos de distribución.
1.7. La distribución normal.
1.8. Ejercicios y problemas del capítulo.

OBJETIVOS:

1. Explicar que es Probabilidad y la relación que guarda con las Ciencias Médicas y en especial con la Estomatología.
2. Identificar y aplicar el concepto de Variable aleatoria.
3. Identificar los modelos teóricos de distribución.
4. Clasificar las variables en estudio en una investigación

La probabilidad y la estadística son, sin duda, las ramas de las Matemáticas que están en mayor auge en este siglo, y tienen una tremenda aplicabilidad en todos los aspectos y ciencias incluyendo en las Ciencias de la Salud, puesto que aquellas variables que incluyen en dichas ciencias, demográficas, suelen tener carácter aleatorio, es decir, no son deterministas, y se fundamentan en predicciones a partir de datos conocidos. Todo aquello que implique predicción nos lleva al terreno de la probabilidad.
El concepto de probabilidad no es extraño para quienes trabajan en las ciencias de la salud, y suele encontrarse en la comunicación cotidiana. Por ejemplo, puede escucharse decir a un médico que un paciente tiene 50% de probabilidad de sobrevivir a una cierta operación. Otro médico puede decir que está un 95% seguro de que un paciente tiene una determinada enfermedad. Una enfermera puede decir que 9 de 10 veces un cierto paciente cancela una consulta.
Para conocer el concepto de probabilidad es necesario antes estudiar los llamados experimentos aleatorios.

1.1- Experimentos Aleatorios.

En todos los aspectos de la vida a veces nos encontramos con acontecimientos predeterminados, es decir, tales que podemos decir el resultado de dichos acontecimientos antes de que finalice o incluso antes de que comience. Tal es el caso de:
1. Preparar una vacuna (sabemos que tipo de vacuna se creará).
2. Calentar un jarro de agua a 100^0C y a presión normal (sabemos que el agua pasa de estado líquido a estado gaseoso).
Tales acontecimientos o experimentos de los que podemos predecir el resultado antes de que se realicen se denominan experimentos deterministas.
Sin embargo, analicemos otro tipo de experimentos, mucho más interesantes desde el punto de vista matemático:
Imaginemos que lanzamos un dado al aire (normal, de 6 caras y no trucado). Podemos predecir el resultado que vamos a obtener? Evidentemente no. Este es un experimento que no es determinista.
A este tipo de experimentos, en los cuales no se puede predecir el resultado antes de realizar el experimento se les denomina **experimentos aleatorios.**

Otros ejemplos de experimentos aleatorios pueden ser:

Observar el número de caries de un paciente, Observar el número de consultas a la que asiste un paciente, Determinar la longitud de un incisivo, etc.

La teoría de probabilidades se ocupa de asignar un cierto número a cada posible resultado que pueda ocurrir en un experimento aleatorio, con el fin de cuantificar dichos resultados y saber si un suceso es más probable que otro o relaciones parecidas. Con este fin, introduciremos algunas definiciones.

Si realizamos un experimento aleatorio, llamaremos *espacio muestral* del experimento al conjunto de todos los posibles resultados de dicho experimento.

Al espacio muestral lo representaremos por E (o bien por la letra griega omega.).

A cada elemento que forma parte del espacio muestral se le denomina *suceso elemental*.

Ejemplo #1:

1¿Cuál es el espacio muestral asociado al experimento de lanzar un dado normal al aire y observar la cara que queda hacia arriba?

En nuestro caso:

Experimento aleatorio:" Lanzar un dado"

Evidentemente, en este caso hay 6 posibles resultados (6 sucesos elementales) y el espacio muestral estará formado por: E={1,2,3,4,5,6}.

2. Y en el caso de observar el número de caries de un grupo de pacientes?

En este otro caso:

Experimento aleatorio: "Observar el número de caries"

Entonces E={0,1,2,3,4,5,6,7,8}

Llamaremos *suceso aleatorio* a cualquier subconjunto del espacio muestral. El concepto de suceso es fundamental en probabilidad. Dicho de forma simple, un suceso de un experimento aleatorio es cualquier cosa que se nos ocurra afirmar sobre dicho experimento. Así, si observamos el sexo de dos pacientes, serían sucesos todos los siguientes:

1. Los dos son masculino.
2. El primero es masculino y el segundo femenino.
3. El primero es femenino y el segundo masculino.
4. Los dos son femenino.

Llamaremos *suceso imposible* al que no tiene ningún elemento y lo representaremos por Φ.

Llamaremos *suceso seguro* al formado por todos los posibles resultados (es decir, al espacio muestral) E.

Llamaremos *espacio de sucesos* y lo representaremos por S, al conjunto de todos los sucesos aleatorios.

Ejemplo#2:

En el caso de observar el sexo de un paciente el espacio muestral es E=*{F,*M*}* , analicemos

quién es el espacio de sucesos:

- Sucesos con 0 elementos: Φ.
- Sucesos con 1 elemento: *{F}*, *{M}*
- Sucesos con 2 elementos:*{F*,M*}*

De modo que el espacio de sucesos es: S=*{*Φ,*{F}*,*{M}*,*{F*,M*}}*.

Si realizamos un experimento aleatorio y consideramos varios sucesos A, B, C, etc, asociados a dicho experimento, podemos realizar varias operaciones entre ellos. Las más importantes son:

1. **Igualdad de sucesos:** Dos sucesos A y B son iguales si están compuestos por los mismos elementos. Lo expresaremos por A = B.

2. **Intersección de sucesos:** Llamaremos *suceso intersección* de los sucesos A y B, y lo representaremos por $A \cap B$, al suceso "ocurren A y B a la vez".

Ejemplo#3: Si tiramos un dado, ya sabemos que el espacio muestral asociado es E=*{*1,2,3,4,5,6*}*.

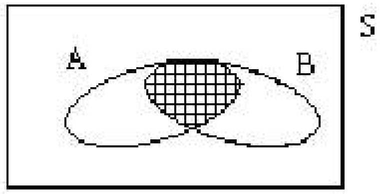
Sean los sucesos A="sacar un nº par"= *{*2,4,6*}*, y B="sacar un número entre 2 y 4 (inclusive)"= *{*2,3,4*}*.

El suceso $A \cap B$ es tal que ocurren A y B a la vez, es decir:

$A \cap B$="sacar un nº par y que esté entre 2 y 4 (inclusive)"= *{*2,4*}*.

El suceso $A \cap B$ son los elementos comunes a los conjuntos A y B (elementos que están en los dos conjuntos).

Representado en diagrama de Venn:



Intersección de sucesos: $A \cap B$

En ocasiones podremos encontrarnos con sucesos que NO tengan elementos en común. En estos casos se dice que los sucesos A y B son *incompatibles*, y su intersección se representa con el conjunto vacío:

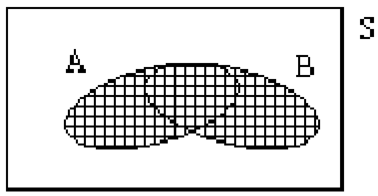
$A \cap B = \Phi$

Evidentemente, si los sucesos sí tienen intersección, diremos que son *compatibles.*

3**. Unión de sucesos:** Llamaremos *suceso unión* de los sucesos A y B y se representa por *AUB* al suceso "ocurre A o bien ocurre B o bien ocurren ambos a la vez" (también podemos decir que "ocurre alguno").

Es decir *AUB* son los elementos que están en ambos conjuntos (aunque no necesariamente en los dos a la vez).

Representado en diagrama de Venn:



Unión de sucesos: *A U B*

Ejemplo#4: En el caso anterior:

A U B="sacar un número par o un número que esté entre 2 y 4 (inclusive)"=*{*2,3,4,6*}*.

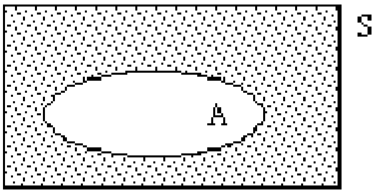
NOTA: Observemos que la intersección de dos conjuntos siempre es "menor"que la unión, de hecho es "menor" que el propio conjunto.

4. **Suceso contrario de otro:** Dado un suceso A, denominaremos *suceso contrario* de A y se representará por Ā (o bien Á o bien A^c) al suceso que tiene por elementos a todos aquellos que no pertenecen a A.

Ejemplo#5: Si tiramos un dado, ya sabemos que el espacio muestral asociado es E=*{*1,2,3,4,5,6*}*.

Como antes, los sucesos A="sacar un número par"= *{*2,4,6*}*, por tanto Ā=*{*1,3,5*}* y B= "sacar un número entre 2 y 4 (inclusive)"=*{*2,3,4*}*, de modo que $\bar{B}$ =*{*1,5,6*}*.

En un diagrama de Venn:

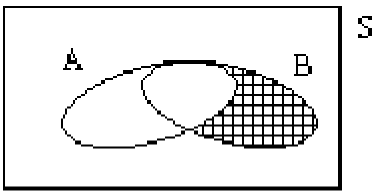


La parte punteada es Ā .(Todo lo que no está incluido en A)

5. **Diferencia de sucesos:** Si A y B son dos sucesos, llamaremos diferencia entre A y B al suceso $B - A$, que consta de los elementos que están en B pero no están en A. Por ejemplo, si A=*{*2,4,6*}*, B=*{*2,3,4*}*, tenemos que $B - A$ = *{*3*}*.

Se cumple que $B - A = B - A \cap B$, y también que $B - A = \bar{A} \cap B$.

Representado en un diagrama de Venn:



La parte rayada es $B - A$, todos los elementos de B que no estén en A

De todas formas, hemos de ser cuidadosos con esta operación: No se debe confundir con una simple resta como operación numérica, sino que es una diferencia conjuntista, quitar los elementos comunes a dos conjuntos.

Además de estas sencillas propiedades (que se demuestran fácilmente mediante un diagrama de Venn), las operaciones con sucesos tienen otras dos propiedades muy importantes:

Leyes de De Morgan: Si A y B son dos sucesos, se verifican:

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$$

$$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$$

2-1 Concepto de Probabilidad.

Hasta el momento hemos descrito lo que es un experimento aleatorio y hemos definido los conceptos básicos asociados a este experimento.

El primer concepto de probabilidad lo formuló el matemático francés Pierre Simon Laplace (1749-1827), quién enunció la regla que lleva su nombre:

Regla de Laplace:

Si realizamos un experimento aleatorio en el que hay n sucesos elementales, todos igualmente probables, entonces si A es un suceso, la probabilidad de que ocurra el suceso A es:

$$p(A) = \frac{\textit{números de casos}\text{ favorables al suceso A}}{\text{números de casos posibles}}$$

Ejemplo#6: Lanzamos un dado normal al aire. Consideramos el suceso A= “sale par”. Calcular p(A).

Casos posibles hay 6, pues E=*{*1,2,3,4,5,6*}*.

Casos favorables al suceso A=*{*2,4,6*}*.

Por tanto $p(A) = \frac{3}{6} = \frac{1}{2} = 0.5$

(Notemos que la probabilidad siempre es un número positivo y menor, o a lo sumo igual a 1). El inconveniente que plantea la definición de Laplace es que necesariamente los sucesos elementales tienen que tener la misma probabilidad de ocurrir.

Observemos un caso tan sencillo como el siguiente:

Al observar el número de piezas que se les extrajo a un grupo de pacientes, se apreció que a 26 se les extrajo 1 pieza, a 8 se les extrajo 2 piezas, a 3 se les extrajo 3 piezas y a 1 paciente se les extrajo 4 piezas Calcula la probabilidad de que la persona seleccionada se le haya extraído:

a) Una pieza

b) Dos piezas

c) Tres piezas.

d) Cuatro piezas

El espacio muestral en este caso será: E={1,2,3,4}, que consta sólo de cuatro elementos, pero sería un poco ingenuo asignar las probabilidades mediante la regla de Laplace,

$p(1) = \frac{1}{4}$ $p(2) = \frac{1}{4}$ $p(3) = \frac{1}{4}$ $p(4) = \frac{1}{4}$

porque ya intuitivamente se ve que hay más posibilidades, por ejemplo, de que extraiga una pieza a que se extraiga dos piezas, de modo que cómo asignar probabilidades?.

Este método tiene sus inconvenientes ya que no siempre es posible determinar el número de casos favorables, por lo que es conveniente usar en estos casos el método de frecuencia relativa, calculando el valor de la probabilidad después que se han observado los resultados de un número de hechos.

¿Cómo se define la frecuencia relativa?

$$\text{Frecuencia relativa} = \frac{\text{Frecuencia absoluta}}{\text{No. Total de casos}}$$

Definición: Si algún experimento se repite un gran número de veces n y si algún evento resultante con la característica A reúne m veces la frecuencia relativa de ocurrencia de A, m/n, será aproximadamente igual a la probabilidad de A.

$$P(A) = \frac{m}{N}$$

Sin embargo, debe tenerse en cuenta que, estrictamente hablando, m/n es solo una estimación de P(A).

Ejemplo#7: En el ejemplo anterior de la cantidad de piezas extraídas a un grupo de pacientes la distribución es la siguiente:

CANTIDAD DE PIEZAS EXTRAIDAS	**NUMERO**
1	26
2	8
3	3
4	1
TOTAL	38

Al calcular la frecuencia relativa de la extracción de una pieza obtenemos:

$$P(A) = \frac{26}{38} = 0{,}68$$

Teniendo definido ya ambos conceptos de probabilidad estudiaremos ahora la Ley de Regularidad Estadística, que se refiere a determinada estabilidad en la frecuencia relativa lo que se observa en fenómenos de tipo aleatorio.

Ejemplo: Queremos calcular la probabilidad de obtener una determinada cara de una moneda.

A una cara

B otra cara

P(A) = ½ probabilidad clásica

Sabemos que la probabilidad de obtener A es 0,5 pero debemos recordar que no se puede predecir de antemano cual cara se va a obtener y que además para una sola tirada de la moneda no nos va a servir de mucho.

Supóngase que se ha hecho un gran número de lanzamientos de una moneda y que dichos lanzamientos los hemos ido agrupando de la forma siguiente

Los primeros 10 lanzamientos

Los primeros 20 lanzamientos

Los primeros 30 lanzamientos

Considérese a los efectos prácticos que se van obteniéndose valores tales como sigue:

Frecuencia relativa de estrella en el primer grupo: 7/10 = 0,7

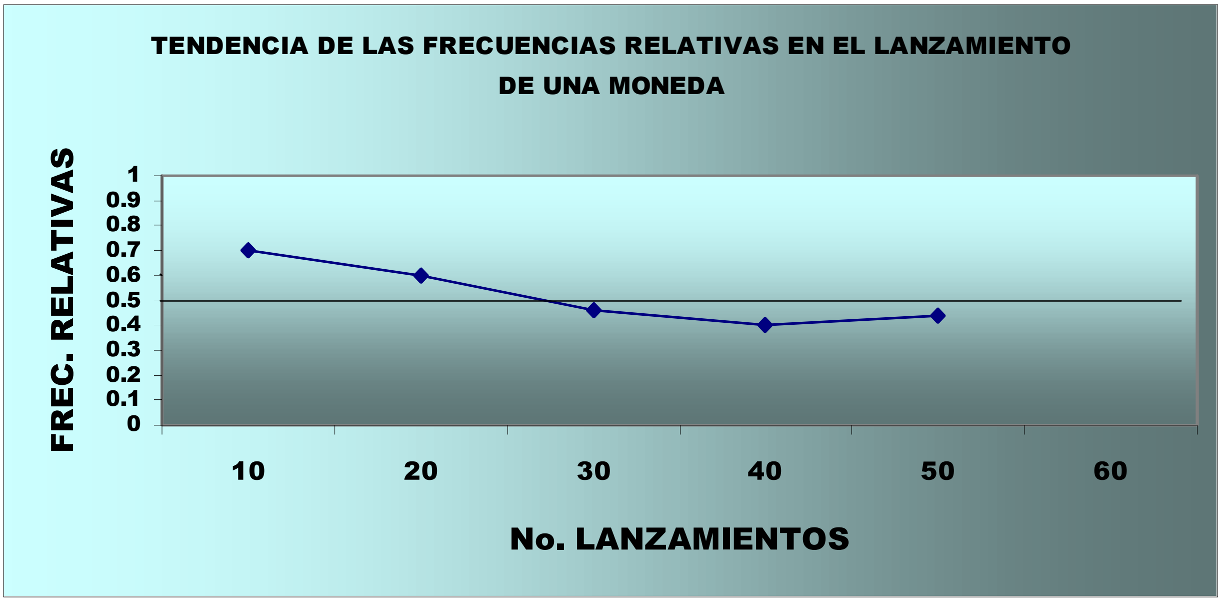
Frecuencia relativa de estrella en el segundo grupo: 12/20 = 0,6

Frecuencia relativa de estrella en el tercer grupo: 14/30 = 0,46

Frecuencia relativa de estrella en el 4to grupo: 16/40 = 0,4

Frecuencia relativa de estrella en el 5to. Grupo 22/50 =0,449

Si con estos valores se hace un gráfico nos queda como sigue:



En este gráfico se observa la tendencia de las frecuencias relativas a irse estabilizando alrededor de cierto valor a medida que aumenta el número de lanzamientos, en el caso de la moneda alrededor de 0,5 y que esta propiedad es una cualidad objetiva de estos valores.
A este valor constante para cada suceso, al cual se acerca cada vez más la frecuencia relativa, a medida que aumenta el número de realizaciones del fenómeno es a lo que se denomina como probabilidad.

Ejemplo#8: al determinar la frecuencia de nacimientos de niñas y niños, con la determinación de la probabilidad de un método clásico tenemos:

P(V) = ½ y P(H) = ½

Sin embargo a través de la frecuencia relativa se ha comprobado que P(V) = 0, 51 y P(H) = 0,49

Fue el matemático ruso Kolmogorov quién precisó el término de probabilidad

Definición axiomática de probabilidad:

Una probabilidad p es una función que asocia a cada suceso A del espacio de sucesos S , un número real p(A), es decir: $p : S \to \mathbb{R}$, y que cumple las propiedades:

1. $0 \leq p(A) \leq 1$, (es decir, cualquier suceso tiene probabilidad positiva y menor o igual que 1).
2. $p(E) = 1$ (la probabilidad del suceso seguro es 1).
3. Si A y B son incompatibles, es decir $A \cap B = \square$, entonces $p(A \cup B) = p(A) + p(B)$. (es decir la probabilidad de la unión es la suma de las probabilidades si los sucesos tienen intersección vacía).

Ejemplo#9:

Sea un experimento aleatorio cualquiera y definamos en S (espacio de sucesos) la siguiente probabilidad:

$$p(A) = \frac{\textit{números de}\ \text{elementos del conjunto A}}{\text{número total de elementos}}$$

Comprobemos que p es una probabilidad.

Para ello, comprobemos las tres propiedades:

a) Se ve que la probabilidad de cualquier suceso está entre cero y uno, puesto que cualquier conjunto que tenga elementos ya tendrá probabilidad positiva, y el número de elementos de cualquier conjunto no puede ser mayor que el número total de elementos existentes.

b) p(E) = 1, es evidente.

c) Tomemos dos sucesos A y B que no tengan elementos en común. Entonces:

$$p(AUB) = \frac{\textit{elementos que}\ \text{forman parte de A o de B}}{\text{número total de elementos}}$$

$$p(AUB) = \frac{\textit{número de}\ \text{elementos de A} + \text{número de elementos de B}}{\text{número total de elementos}}$$

$p(AUB) = p(A) + p(B)$

puesto que si A y B no tienen elementos comunes, el número de elementos de la unión es la suma de los elementos de cada conjunto por separado.

Por tanto se cumplen las 3 propiedades y **p** así definida es una probabilidad. Esta sería la definición de probabilidad que utilicemos a partir de ahora.

Ejemplo#10:

En el ejemplo de la extracción de piezas anterior, lo lógico es definir la probabilidad así: Como en total hay 38 pacientes y a 26 se les extrajo 1 pieza dental, 8 pacientes se les extrajo 2 piezas, 3 pacientes se les extrajo 3 piezas y 1 paciente se les extrajo 1 pieza, $p(1) = \frac{26}{34}$ $p(2) = \frac{8}{34}$ $p(3) = \frac{3}{34}$ $p(4) = \frac{1}{34}$.

Se puede comprobar que así definida p es una probabilidad.

Sin embargo, comprobar las propiedades de la definición de Kolmogorov es una labor larga y engorrosa, puesto que hay que verificar que se cumple para todos aquellos sucesos del espacio de sucesos S, que es ciertamente amplio en muchas ocasiones. El siguiente resultado simplifica la tarea de decidir cuando una función p sobre el espacio de sucesos es una probabilidad, basándose sólo en los sucesos elementales, es decir, aquellos que forman parte del espacio muestral. Lo enunciaremos sin demostración:

Propiedad

Si $w1, w2, \ldots, wn$ son los n sucesos elementales de un suceso aleatorio cualquiera, p una función

$p: S \rightarrow R$ de modo que cumple las propiedades:

1. $0 \leq p(wi) \leq 1 \quad \forall i \in \{1, 2, \ldots, n\}$
2. $p(w1) + p(w2) + \ldots + p(wn) = 1$

Entonces p es una probabilidad.

Ejemplo#11: Comprobar si las siguientes funciones definidas para los sucesos elementales son probabilidad, siendo E=$\{a,b,c,d\}$ el espacio muestral del experimento aleatorio:

a) $p(a)= \frac{1}{2}$ $p(b)= \frac{1}{3}$ $p(c)= \frac{1}{4}$ $p(d) = \frac{1}{5}$

Es obvio que la primera propiedad se cumple, puesto que las 4 probabilidades son números positivos menores que 1.

Para ver si se cumple la segunda, basta realizar la suma:

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} = \frac{30+20+15+12}{60} = \frac{77}{60}$$

que evidentemente NO es 1, luego p NO es probabilidad.

b) $p(a) = \frac{1}{4}$, $p(b) = \frac{1}{2}$, $p(c) = 0$, $p(d) = \frac{1}{2}$

Es obvio que la primera propiedad se cumple, puesto que las 4 probabilidades son números positivos o cero menores que 1.

Para ver si se cumple la segunda, basta realizar la suma:

$$\frac{1}{4} + \frac{1}{2} + \frac{1}{2} = 1$$

Luego p SI es probabilidad.

Consecuencias de la definición de probabilidad:

1. $p(\bar{A}) = 1 - p(A)$

En efecto, puesto que $E = A \cup \bar{A}$ y además A y Ā son incompatibles, resulta por la propiedad 3) de la definición que

$p(E) = p(A \cup \bar{A}) = p(A) + p(\bar{A})$

Y por la propiedad 2), p(E)=1, luego $1 = p(A) + p(\bar{A})$ y por tanto $p(\bar{A}) = 1 - p(A)$

2. $p(\Phi) = 0$

Como $\bar{E} = \Phi$, resulta que:

$p(\bar{E}) = p(\Phi) = 1 - p(E) = 1 - 1 = 0$

3. Si A y B son dos sucesos cualesquiera,

$p(A \cup B) = p(A) + p(B) - p(A \cap B)$

4. Si A, B y C son tres sucesos cualesquiera,

$p(A \cup B \cup C) = p(A) + p(B) + p(C) - p(A \cap B) - p(A \cap C) - p(B \cap C) + p(A \cap B \cap C)$

Ejemplo#12:

Se observa el sexo de un paciente en 3 veces. Calcular la probabilidad de observar alguna vez un paciente femenino.

Los problemas de este tipo, en los que se pide la probabilidad de observar "alguna" cosa, se suelen resolver muy bien por paso al complementario. En este caso concreto,

A = "observar algún paciente femenino".

Ā= "no observar ningún paciente femenino"= "observar 3 pacientes masculino".

Entonces, $p(A) = \frac{1}{8}$, pues hay 8 casos posibles y sólo uno favorable (MMM, 3 masculino), por tanto:

$p(A) = 1 - p(\bar{A}) = 1 - \frac{1}{8} = \frac{7}{8}$

1-3 Probabilidad condicionada.

Hasta ahora nos hemos limitado a calcular probabilidades únicamente partiendo de un experimento aleatorio, sin tener más información. Pero, .qué ocurre si conocemos alguna información adicional?.

Supongamos que estamos realizando el experimento aleatorio de lanzar un dado y obtener el número que sale. Consideremos el suceso A= “sale un 4”. Evidentemente, $p(A) = \frac{1}{6}$

Pero si con antelación a esta probabilidad se ha fijado otra probabilidad de que al lanzar el dado ha salido un número par?.

Disponemos entonces de una información adicional, B=$\{2,4,6\}$.

Hemos reducido nuestro espacio muestral, que ahora sólo consta de 3 elementos y tenemos que cambiar las probabilidades asignadas.

Ahora el suceso A no tiene una posibilidad entre 6 de ocurrir, sino una entre tres, es decir, $p(A) = \frac{1}{3}$

Esta es la idea de la probabilidad condicionada: La información obtenida B, modifica la probabilidad de A. Lo expresaremos así: $p(A/B) = \frac{1}{3}$

y se lee “probabilidad de A condicionada a B” o “probabilidad de A conociendo B”.

El caso anterior es muy sencillo, pues directamente podemos calcular $p(A/B)$, pero si el espacio muestral se amplía, el problema es más complicado. La fórmula siguiente simplifica el problema.

Definición:

Sea A un suceso aleatorio asociado a un experimento aleatorio, y sea B otro suceso que sabemos que se ha realizado. Llamaremos *probabilidad de A condicionada a B* y lo expresaremos por *p*(*A/B*) a la expresión:

$$p(A/B) = \frac{p(A \cap B)}{p(B)}$$

(de idéntico modo se define *p*(*B/A*)).

Sucesos independientes

Si bien el conocer cierta información adicional modifica la probabilidad de algunos sucesos, puede ocurrir que otros mantengan su probabilidad, pese a conocer dicha información.

Por ejemplo, en el lanzamiento de un dado, consideremos los sucesos: A= “sacar un número par” y B= “sacar un número menor o igual que 2” Es claro que A= *{*2,4,6*}* y B= *{*1,2*}*.

Calculemos la probabilidad de A conociendo que se ha realizado el suceso B, es decir, *p*(*A/B*).

Utilizando la fórmula:

$$p(A/B) = \frac{p(A \cap B)}{p(B)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{3}{6} = \frac{1}{2} = 0{,}5$$

puesto que $p(A \cap B)$=p(sacar par y menor o igual que 2)= $\frac{1}{6}$ y p(B)= $\frac{1}{3}$

Pero si no conociésemos la información B, .cuál sería la probabilidad de A?.

p(A)=p(sacar par)= $\frac{3}{6}$ = 0,5, es decir que p(A/B)=p(A), y por tanto el conocer la información B no modifica la probabilidad de A.

Cuando esto ocurre es decir, cuando *p*(*A/B*) = *p*(*A*), diremos que los *sucesos A y B son independientes* (el hecho de que ocurra B no modifica la probabilidad de A).

Propiedad:

A y B son sucesos independientes . $\Leftrightarrow$ $p(A \cap B) = p(A) \cdot (B)$.

Demostración: $\Rightarrow$) Si A y B son independientes, $p(A/B) = p(A)$, y por la fórmula de la probabilidad condicionada, $p(A/B) = \dfrac{p(A \cap B)}{p(B)}$, luego $\dfrac{p(A \cap B)}{p(B)} = p(A)$, y por tanto $p(A \cap B) = p(A) \cdot p(B)$.

$\Leftarrow$) Partiendo de $p(A \cap B) = p(A) \cdot p(B)$, entonces

$$p(A/B) = \frac{p(A \cap B)}{p(B)} = \frac{p(A) \cdot p(B)}{p(B)} = p(A)$$

luego $p(A/B) = p(A)$ y por tanto A y B son independientes.

Ejemplo#13:

En el caso anterior, $p(A \cap B) = \dfrac{1}{6}$, y por otra parte $p(A) = \dfrac{1}{2}$ y $p(B) = \dfrac{1}{3}$

,luego se cumple que $p(A \cap B) = \dfrac{1}{6} = \dfrac{1}{2} \cdot \dfrac{1}{3} = p(A) \cdot p(B)$

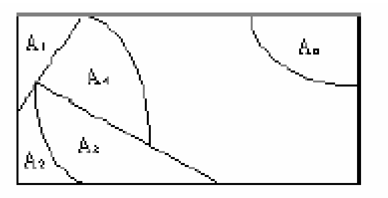
luego A y B son independientes.

1-4 El teorema de Bayes.

Teorema de la probabilidad total: Si $A1, A2, \ldots, An$ son sucesos incompatibles 2 a 2, y cuya unión es el espacio muestral ($A1 \cup A2 \cup \ldots \cup An = E$), y B es otro suceso, resulta que:

$$p(B) = p(A1) \cdot p(B/A1) + p(A2) \cdot p(B/A2) + \ldots + p(An) \cdot p(B/An)$$

Nota: El conjunto $A1, A2, \ldots, An$ que verifica la incompatibilidad 2 a 2 y que la unión de todos ellos es el espacio muestral se denomina *sistema completo de sucesos* y este sistema “divide el espacio muestral en partes que no se solapan”. Mediante representación gráfica:



Sistema completo de sucesos: $A1, A2, \ldots, An$

Ejemplo#14:

Tenemos dos grupos de segundo año de estomatología, uno con 17 estudiantes femeninos y 3 estudiantes masculinos, y otro grupo con 15 estudiantes femeninos y 5 estudiantes masculinos. Seleccionamos un estudiante. Cuál es la probabilidad de que el estudiante seleccionado sea masculino?.

Solución:

Experimento aleatorio: "Seleccionar un estudiante"

Suceso B {Observar un estudiante masculino}

Suceso A_1 {Seleccionar el grupo1}

Suceso A_2 {Seleccionar el grupo 2}.

En primer lugar, se trata de elegir un grupo, para lo cuál lanzamos una moneda. Si A1= "elegir el grupo 1" y A2= "elegir el grupo 2", es claro que p(A1)= p(A2)= $\frac{1}{2}$

Y aplicando el teorema:

$p(B) = p(A1)\, p(B/A1) + p(A2)\, p(B/A2) = \frac{1}{2} \cdot \frac{3}{20} + \frac{1}{2} \cdot \frac{5}{20} = \frac{3+5}{40} = \frac{8}{40} = 0.2$

Como consecuencia del teorema de la probabilidad total y de las propiedades de la probabilidad condicionada, resulta importante **El teorema de Bayes,** teorema que permite calcular probabilidades condicionadas.

Teorema de Bayes:

Si *A*1*, A*2*, . . ., An* son sucesos incompatibles 2 a 2, y cuya unión es el espacio muestral

($A1 \cup A2 \cup . . . \cup An = E$), y B es otro suceso, resulta que:

$$p(Ai/B) = \frac{p(A_i)\cdot p(B/A_i)}{p(A1)\cdot p(B/A1) + p(A2)\cdot p(B/A2) + ... + p(An)\cdot p(B/An)}$$

Demostración:

Por definición, $p(Ai/B) = \frac{p(A_i \cap B)}{p(B)}$

Ahora bien, recordando que $p(Ai \cap B) = p(Ai) \cdot p(B/Ai)$, debido a que

$p(B/Ai)$ = $\frac{p(A_i \cap B)}{p(A_i)}$

Por tanto, combinando los dos hechos:

$p(Ai/B)$ = $\frac{p(A_i \cap B)}{p(B)} = \frac{p(A_i) \cdot p(B/A_i)}{p(B)}$

Como por el teorema de la probabilidad total es:

$p(B) = p(A1) \cdot p(B/A1) + p(A2) \cdot p(B/A2) + ... + p(An) \cdot p(B/An)$

resulta que sustituyendo:

$p(Ai/B)$ = $\frac{p(A_i) \cdot p(B/A_i)}{p(B)}$ =

$$p(Ai/B) = \frac{p(A_i) \cdot p(B/A_i)}{p(A1) \cdot p(B/A1) + p(A2) \cdot p(B/A2) + ... + p(An) \cdot p(B/An)}$$

y el teorema queda demostrado.

Ejemplo#15:

Dado el ejemplo anterior donde tenemos dos grupos de segundo año de estomatología, uno con 17 estudiantes femeninos y 3 estudiantes masculinos, y otro grupo con 15 estudiantes femeninos y 5 estudiantes masculinos y se seleccionó un estudiante. Cuál es la probabilidad de que el estudiante seleccionado sea del grupo 1 dado que sea masculino?

Solución:

Experimento aleatorio: “Seleccionar un estudiante”

Suceso B {Observar un estudiante masculino}

Suceso A_1 {Seleccionar el grupo1}

Suceso A_2 {Seleccionar el grupo 2}.

Dado que A1= “elegir el grupo 1” y A2= “elegir el grupo 2”, ya habíamos calculado que p(A1)= p(A2)= $\frac{1}{2}$

Y aplicando el teorema total obtuvimos que P(B)= 0.2

Aplicando el teorema de Bayes debemos calcular:

$$p(A1/B) = \frac{p(A1) \cdot p(B/A1)}{p(A1) \cdot p(B/A1) + p(A2) \cdot p(B/A2)}$$

$$p(A1/B)= \frac{\frac{1}{2}\cdot\frac{3}{20}}{\frac{1}{2}\cdot\frac{3}{20}+\frac{1}{2}\frac{5}{20}} = \frac{\frac{3}{40}}{\frac{8}{40}} = \frac{3}{8} = 0{,}375$$

1.5 Variable aleatoria

Una de las herramientas de la metodología científica es la experimentación, y por lo general, toda experimentación requiere la medición de observaciones. Medimos la altura y el peso del recién nacido. Con esas observaciones, unidas a las anteriores que hemos tomado vamos formando una base de datos que en un momento dado las utilizaremos para proponer algunas conclusiones o tesis de trabajo. Incluso, en algún momento, con la sola observación de un determinado dato, de un determinado peso, por ejemplo, que esté debajo de un cierto nivel o umbral inmediatamente tomamos una decisión, un programa de alimentación especial. En rigor, podemos decir que la Estadística nos ayuda en la toma de decisiones. Para el caso particular de las Ciencias de la Salud, sin ninguna duda la Estadística ha sido fundamental. Incluso, en algún momento de la historia la propia ciencia de la salud abrió puertas para el avance del conocimiento de la Estadística.

Ahora bien, utilizaremos un lenguaje matemático para poder operar con observaciones estadísticas. Supongamos que queremos medir una serie de observaciones de un determinado fenómeno aleatorio, para realizar esto necesitaremos del concepto de variable aleatoria. ¿Qué es una variable aleatoria?

La palabra variable significa lo siguiente: "magnitud que puede tener un valor cualquiera de los comprendidos en un conjunto". En rigor es una definición ajustada al uso práctico que los estadísticos hacen de una "variable aleatoria". Es una magnitud, dice la definición, que va a tomar un valor cualquiera de los comprendidos en un conjunto, de modo que necesitamos dos entes. En primer lugar, debemos definir una letra que describirá la magnitud; y en segundo lugar, debemos definir el conjunto de los valores que asumirá la letra (la magnitud). Por lo general, de manera clásica pero en absoluto la única forma, se utiliza la letra ***X*** para designar una variable, y se denota por ***E*** al conjunto de valores que

asumirá, de alguna forma, la variable ***X***. Luego hemos dado a conocer lo que se entiende por variable. Pero ahora, ¿qué significa el apellido aleatoria?
De manera clara entonces, para definir una variable aleatoria necesitamos una letra de cualquier alfabeto conocido o más usual, siendo muy sencilla esta elección; y, necesitamos saber el conjunto de resultados E que describan, de alguna manera, todos los posibles resultados de un determinado experimento aleatorio.
Se dice que hemos definido una variable aleatoria para un experimento aleatorio cuando hemos asociado un valor numérico a cada resultado del experimento.
Sea E el espacio muestral asociado a un experimento. Se llama variable aleatoria a toda aplicación del espacio muestral E en el conjunto de los números reales (es decir, asocia a cada elemento de E un número real).

Ejemplos de variable aleatoria.

1. Lanzamos un dado, y estamos interesados en ver los resultados de la cara superior del dado al detenerse en su lanzamiento. Definamos entonces la letra ***X*** que tomará valores en el conjunto ***E*** = {1, 2, 3, 4, 5, 6}. Esta letra *X* es una variable aleatoria que describe todos los posibles resultados que pueden obtenerse al lanzar un dado.
2. Se lanzan una moneda al aire n veces, y se quiere saber el número de caras que se pueden obtener. Sea ***X*** la variable que toma valores en el conjunto ***S*** = {0, 1, ... , n}. Esta variable describe todos los posibles números de cara que se pueden obtener al lanzar n veces una moneda.
3. Deseamos observar la cantidad de caries a todos los pacientes atendidos a partir del 1 de enero del 2006 hasta el 31 de diciembre del mismo año. Entonces definimos por ***X*** como la variable que denotará la cantidad de caries de todo paciente a partir del primero de enero y hasta el 31 de diciembre del 2006. Ahora bien, ¿cuál es el espacio de todos los posibles resultados para estas observaciones?

Podemos concluir entonces, que una variable aleatoria es una variable que "tomará" un determinado valor de un conjunto de resultados que a su vez emergen de un fenómeno aleatorio, o de un fenómeno cuyos resultados son azarosos.

Existen diversas clasificaciones de variable aleatoria, y esta clasificación está en relación con el espacio E de los posibles resultados que deseamos observar.

Dentro de este tipo de variables podemos distinguir dos grupos: **discreta y continua.**

Variable aleatoria discreta (v.a.d.)

Se dice que una variable aleatoria es discreta cuando el espacio de todos los posibles resultados, constituye un conjunto discreto. Ahora bien, un conjunto es discreto cuando es contable o que se puede numerar. Piense usted en el conjunto de los números naturales o en el conjunto {0, 1, ... , n}.

Ejemplos:

1.- En el experimento aleatorio lanzar una moneda al aire E = {Cara, Escudo}.

Cara → 1 Escudo → 2

Es decir existen dos posibles resultados

2.- En el experimento aleatorio Observar el número de piezas que se extraen un grupo de pacientes E = {1, 2, 3, 4}.

Es decir el espacio muestral dependerá de la cantidad máxima de piezas que se le extraiga a un paciente. Pero si es una variable discreta puesto que estamos contando.

Supongamos que tenemos una variable aleatoria discreta X que puede tomar los valores x1, x2, x3, ... A toda función Pk, definida por Pk = P {X= xk} con k $\in \mathbb{N}$aturales, que le hace corresponder a cada valor xk se le denomina función de masa de probabilidad.

Se le llama ley de distribución de una v.a.d. X a la tabla siguiente donde se expresan los pares (xk, Pk)

X

Xk	X1	X2	X3	· · ·
Pk	P1	P2	P3	· · ·

Por ejemplo en el Experimento aleatorio : “Lanzar una moneda al aire”

X

Xk	1	2
Pk	1/2	1/2

Propiedades:

1.- $pk \geq 0 \;\; \forall\, k = 1, 2, \ldots, n$

2.- $\sum_{k=1}^{n} pk = 1$

Variable aleatoria continua (v.a.c.)

Se dice que una variable aleatoria es continua cuando el espacio de todos los posibles resultados, constituye un conjunto continuo. Ahora bien, un conjunto es continuo cuando no es contable o que no se puede numerar. Piense usted en el conjunto de los números reales. Entre dos valores, siempre es posible encontrar otro valor.

Ejemplos:

1.- En el experimento aleatorio "Observar el tamaño del incisivo de una paciente"

E = {Todos los posibles tamaños de un incisivo, existe una cantidad infinita no numerable}.

En el caso específico de las v.a.c. es imposible construir una ley de distribución ya que la probabilidad de ocurrencia de un valor resultaría muy baja, por lo que se construye la función de densidad.

Una función $y=f(x)$ es una función de densidad de una variable aleatoria continua si cumple las siguientes condiciones :

- Es positiva en todo su dominio : $0 \leq f(x) \leq 1$
- Permite obtener $p(a \leq X \leq b)$ como área bajo la gráfica entre $x = a$ y $x = b$. Verifica la fórmula $p(a \leq X \leq b) = \int_a^b f(t)dt$
- El área total entre la gráfica de f y el eje X vale 1 $\int_{-\infty}^{+\infty} f(t)dt = 1$
- Permite obtener $F(x)$ como área bajo la gráfica hasta el valor x

A la función F(x) se le denomina Función de distribución

En general, la *función de distribución* de una variable aleatoria continua X es el modelo teórico de la curva de frecuencias acumuladas que se espera obtener para X, y debe cumplir, evidentemente, estas propiedades:

Ser creciente

Tomar valores de 0 a 1

Si X es una variable aleatoria continua con valores en un intervalo [a, b], entonces F(x) será la probabilidad de que la variable X tome valores entre a y *x*. $F(x)=P(a \leq X \leq x)$.

1.6 Modelos teóricos de distribución.

Al estudiar las variables aleatorias discretas pudimos observar que la probabilidad se concentran en determinados puntos así por ejemplo en el Experimento aleatorio : “Lanzar una moneda al aire” por lo que E = {Cara, Escudo}.

Donde: Cara → 1 Escudo → 2 y Xk es que aparezca cara o escudo.

Xk	1	2
Pk	1/2	1/2

Luego, si calculamos la probabilidad de que al lanzar una moneda al aire aparezca cara entonces:

Suceso A: {Observar cara}

Donde Xk ahora es el número de caras a obtener.

Xk	0	1
Pk	1/2	1/2

Si valoramos el Experimento aleatorio : “Lanzar una moneda dos veces al aire” luego E={(Cara ,Cara),(Cara, Escudo),(Escudo ,Cara),(Escudo, Escudo) }.

Y la ley de distribución será:

Xk	0	1	2
Pk	1/4	1/2	1/4

Es decir la probabilidad de que aparezca cara una ves sigue siendo de 1/2.

Si ahora valoramos el Experimento aleatorio : “Lanzar una moneda tres veces al aire” por lo que E = {(Cara ,Cara, Cara), (Cara, Cara, Escudo), (Cara, Escudo, Cara), (Escudo ,Cara, Cara), (Cara, Escudo, Escudo), (Escudo, Cara, Escudo), (Escudo, Escudo, Cara), (Escudo, Escudo, Escudo)}.

Y la ley de distribución será:

Xk	0	1	2	4
Pk	1/8	3/8	3/8	1/8

Es decir la probabilidad de que aparezca cara una ves es igual a 3/8.

Los experimentos aleatorios anteriores tienen las siguientes características:

En cada prueba del experimento sólo son posibles dos resultados: el suceso A (éxito) y su contrario Ạ̄ (fracaso).

El resultado obtenido en cada prueba es independiente de los resultados obtenidos anteriormente.

La probabilidad del suceso A es constante, la representamos por *p*, y no varía de una prueba a otra. La probabilidad de Ā es 1- *p* y la representamos por *q*

El experimento consta de un número *n* de pruebas.

Todo experimento que tenga estas características diremos que sigue el modelo de la ***distribución Binomial***. A la variable *X* que expresa el número de éxitos obtenidos en cada prueba del experimento, la llamaremos ***variable aleatoria binomial***.

En nuestro modelo anterior con el lanzamiento de la moneda, suceso A es obtener cara por lo que Ā será no obtener cara (por ejemplo obtener escudo).

Los cálculos que realizamos anteriormente son fáciles, pero supongamos ahora el siguiente Experimento aleatorio : “Lanzar una moneda veinte veces al aire” y deseamos calcular la probabilidad de que aparezca cara una ves. Si realizamos el mismo procedimiento que hemos empleado anteriormente sería un tanto complicado ¿Inténtelo usted?

Siendo p la probabilidad de ocurrencia del suceso, entonces 1 – p =q es la probabilidad contraria o de no ocurrencia; comúnmente en los textos se denominan éxito y fracaso respectivamente desde un punto didáctico.

Cuando se dividen los sucesos de un experimento en éxitos y fracasos y conociendo sus probabilidades p y 1 – p respectivamente la probabilidad de obtener x éxitos y n – x fracasos es en algún orden dado

$$px(1-p)^{n-x}$$

$$P(x,n,p)=\binom{n}{x}p^{x}(1-p)^{n-x} \qquad \binom{n}{x}=\frac{n!}{x!(n-x)!}$$

En nuestro caso se tiene

$$P(1;20;0,5)=\binom{20}{1}(0,5)^{1}(0,5)^{19}=0,000019$$

Dado que los números $\binom{n}{x}$ se denominan coeficientes binomiales por corresponder a los que origina el binomio en desarrollo $(a+b)^{n}$ da a la distribución de estas probabilidades el nombre de distribución binomial.

Parámetros de la Distribución Binomial

Media	$\mu = np$
Varianza	$\sigma^2 = npq$
Desv.Típica	$\sigma = \sqrt{npq}$

Función de Distribución de la v.a. Binomial

$$F(x_i)=p(X\le x_i)=\binom{n}{0}p^{0}q^{n}+\binom{n}{1}p^{1}q^{n-1}+\ldots+\binom{n}{k}p^{k}q^{n-k}$$

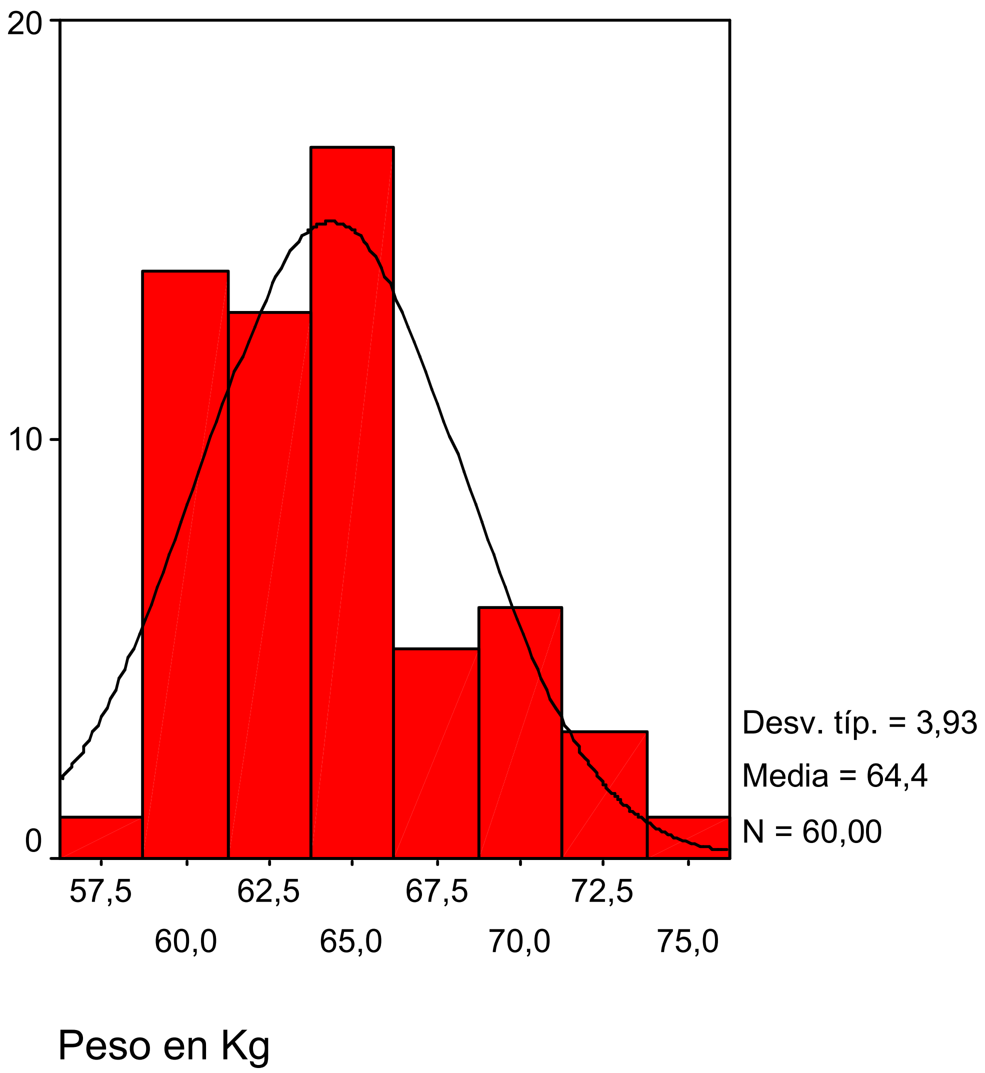
siendo k el mayor número entero menor o igual a x_i.

Esta función de distribución proporciona, para cada número real x_i, la probabilidad de que la variable X tome valores menores o iguales que x_i.

1.7 La distribución normal

Al estudiar las variables aleatorias continuas pudimos observar que la probabilidad no es posible concentrarla en determinados puntos, por lo que se construyó la función de densidad. Cuando se dispone de datos relativos a una variable aleatoria continua como la hemoglobina, la talla de un incisivo, el presión arterial de un paciente u otra el proceso para obtener sus probabilidades resulta más complicado por lo que una representación gráfica a través de un histograma para ello posibilita ciertos análisis estadísticos adecuados.

Por ejemplo si representamos gráficamente a la variable peso podemos obtener:



Si un histograma es aproximado por medio de una curva suave, la frecuencia, porcentaje o probabilidad de cualquier clase dada está representada por el área correspondiente debajo de la curva.

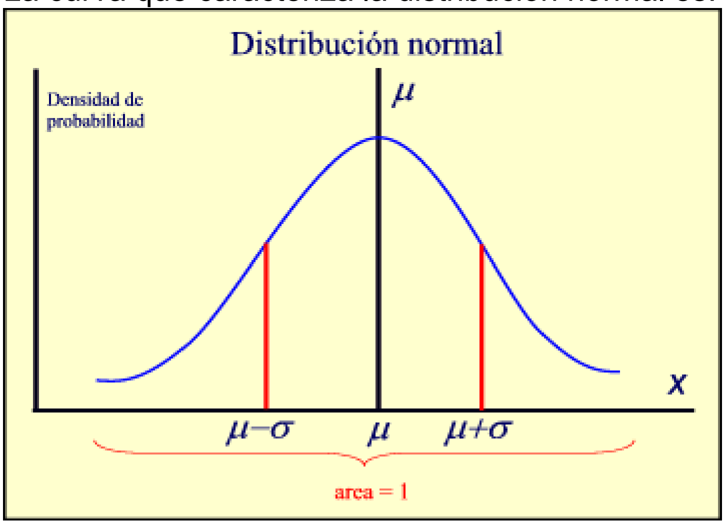
Estas curvas pueden tener formas diferentes, pero obedecen al nombre de curvas de distribución o distribuciones continuas, siempre que el área debajo de estas curvas, entre dos valores a y b donde a < b sea igual a la proporción de casos que caen entre dichos valores.

También se puede partir de las proporciones a las probabilidades y llamaremos a una curva, curva de distribución continua, si el área debajo de ella entre dos valores es igual a la probabilidad de obtener un valor entre ellos.

Los datos agrupados pueden determinar formas variadas de histogramas y por tanto sus aproximaciones a curvas suaves también pueden ser diversas: estadísticamente se tienen distribuciones de varios tipos siendo las más usadas la binomial que ya estudiamos anteriormente la normal, la de Bernoulli, la T de Student y la Chi – cuadrado.

Entre las leyes probabilísticas ocupa un lugar importante la llamada distribución Normal o de Gauss. La importancia de esta distribución teórica radica en el hecho de que una gran cantidad de variables y en particular variables biológicas se rigen por este modelo.

La curva que caracteriza la distribución normal es:



Es una curva en forma de campana perfectamente simétrica y es por eso que también se le llama campana de Gauss.

La expresión matemática de esta curva es

$$f(x) = (1/\sigma\sqrt{2\pi})e^{-(x-\mu)^2 / 2\sigma^2} \qquad -\infty < x < +\infty$$

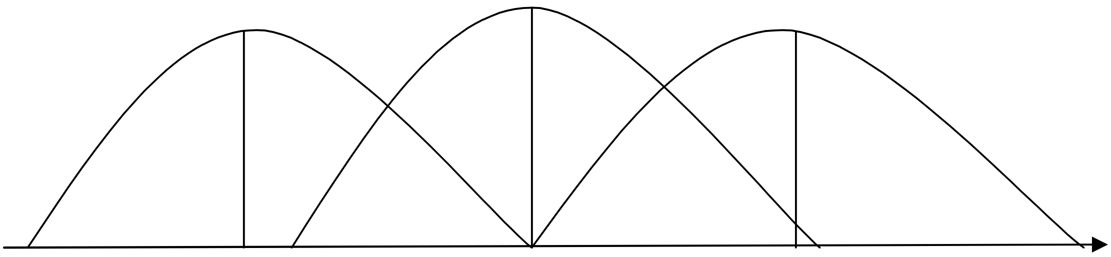
- ¿De cuántos parámetros depende esta función?

Por lo que se puede apreciar esta función depende de 2 parámetros:

La media poblacional y la desviación estándar poblacional

Propiedades de la distribución normal

1. Es simétrica respecto a su media
2. La media la mediana y la moda coinciden por lo que el 50% del área está a la derecha de la media y el 50% a la izquierda
3. El área comprendida entre la curva y el eje de las abcisas es aproximadamente igual a 1
4. La curva depende de la media poblacional y de la desviación estándar poblacional por lo que para cada par de valores de estos parámetros se obtienen diferentes curvas.
5. Sean $X_1 \sim N(\mu_1, \sigma^2)$; $X_2 \sim N(\mu_2, \sigma^2)$; $X_3 \sim N(\mu_3, \sigma^2)$



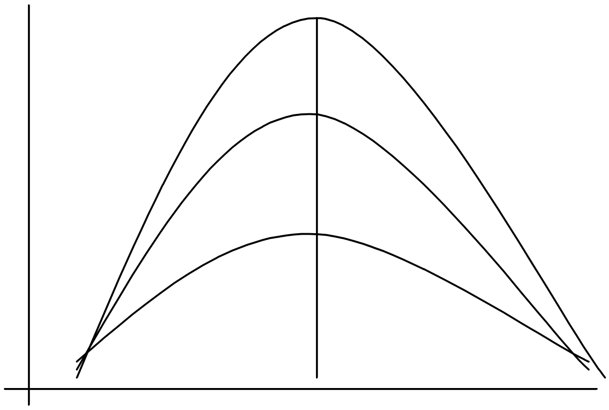
$$\mu_1 < \mu_2 < \mu_3$$

Debe notarse que la forma del gráfico (altura y extensión) coincide en todos los casos pero varía la posición del gráfico al variar su eje de simetría.

Por esta razón a la media poblacional se le denomina parámetro de posición o localización

6- Mientras mayor sea la desviación estandar poblacional más dispersa o suave es la distribución normal, los datos aparecen alejados de su media.

Sean $X_1 \sim N(\mu, \sigma^2_1)$; $X_2 \sim N(\mu, \sigma^2_2)$; $X_3 \sim N(\mu, \sigma^2_3)$



$\sigma_1 < \sigma_2 < \sigma_3$

Una propiedad muy importante en la cual nos apoyamos más adelante, es la que se representa a continuación.

Para cualquier valor de μ y σ (cualquier curva normal) se cumple que las áreas en los intervalos $(\mu-\sigma,\mu+\sigma)$, $(\mu-2\sigma,\mu+2\sigma)$ $(\mu-3\sigma,\mu+3\sigma)$ es constante, siendo respectivamente 0,6827 , 0,9545 y 0,9973.

$P(\mu-\sigma,\mu+\sigma) = 0,6827$
$P(\mu-2\sigma,\mu+2\sigma) = 0,9545$
$P(\mu-3\sigma,\mu+3\sigma) = 0,9973$

Ejemplo #16: Si la edad es una variable aleatoria con distribución normal en una población determinada con Media de 30 años y Desviación estándar 3 años, entonces:

$P(30-3<x<30+3) = P(27<x<33) = 0,6827$

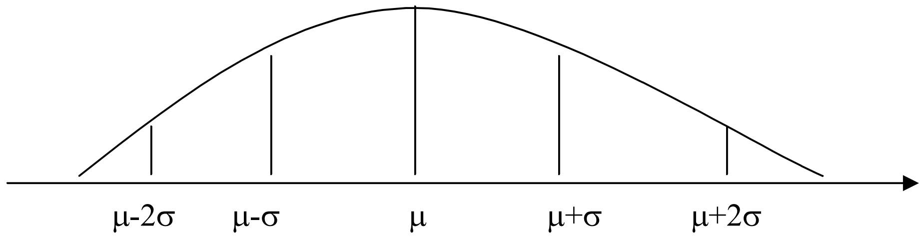
$P(30-2*3<x<30+2*3) = P(24<x<36) = 0,6827$

$P(30-3*3<x<30+3*3) = P(21<x<39) = 0,6827$

Entre las aplicaciones prácticas de esta propiedad en salud puede citarse:

Cuando se sabe que una variable aleatoria sigue una distribución normal, o aproximadamente normal suele considerarse como normal el rango comprendido entre

$(\mu-\sigma,\mu+\sigma)$, el riesgo comprendido entre $(\mu-2\sigma,\mu-\sigma)$ o $(\mu+\sigma,\mu+2\sigma)$, y patológico los que se encuentran por debajo de $\mu-2\sigma$ o por encima de $\mu+2\sigma$



Supongamos que deseamos conocer la probabilidad de que un paciente con 4 caries tenga una hemoglobina de 13 o menos. Para ello deberíamos tener el número de pacientes con 13 de hemoglobina o menos y conocer el total de pacientes con 4 caries.

Hallar la probabilidad de que un paciente tenga una hemoglobina que oscile entre a y b, equivaldría a calcular el área bajo la curva de frecuencias teóricas que media entre esos valores. Esto se expresaría como la integral de la función que describe la curva entre los puntos a y b. Esto es :

$$p(a < X < b) = \int_a^b (1/\sigma \sqrt{2\pi}) e^{-(x-\mu)^2 / 2\sigma^2} \, dx$$

Para calcular estas probabilidades es entonces necesario estandarizar la distribución normal .

Cuando μ = 0 y σ = 1, la expresión de la función que caracteriza a este modelo de distribución normal se simplifica, reduciéndose a:

$$f(x) = (1/\sqrt{2\pi}) e^{-x^2 / 2} \qquad -\infty < x < +\infty$$

Se dice entonces que la variable aleatoria sigue una distribución normal estándar.

1.8 EJERCICIOS Y PROBLEMAS DEL CAPITULO.

1. Escribir el espacio muestral asociado al experimento aleatorio seleccionar un paciente que sea femenino.

2. Escribir el espacio muestral asociado al experimento aleatorio medir el tamaño del incisivo de un paciente.

3. Escribir el espacio muestral asociado al experimento aleatorio observar el estado de salud bucodental de un grupo de pacientes.

4.- Se ha encargado a un estudiante de estomatología que de un día de práctica en la consulta observó a 12 pacientes y de ellos 4 tiene enfermedades bucodentales. Hallar la probabilidad de que paciente elegido al azar:

a) Tenga una enfermedad bucodental.

b) Esté sano.

5.- Dado el ejercicio anterior. Calcula la probabilidad de que un paciente tenga una enfermedad bucodental, si se sabe que el paciente seleccionado es masculino y conociendo que de los 12 pacientes 8 son masculinos.

6.- A una consulta de estomatología se sabe que asistieron 20 pacientes de ellos 12 femenino y 8 masculino. Se seleccionan sucesivamente 2 pacientes. Calcula la probabilidad de que:

a) El primer paciente sea femenino y el segundo masculino.

b) Ambos sean femenino.

7.- Se tienen dos consultas. En la primera asistieron 10 pacientes, 3 con enfermedades bucodentales y 7 no. En la segunda asistieron 12 pacientes de ellos 6 con enfermedades bucodentales. Se elige un paciente al azar. Calcular:

a) Probabilidad de seleccionar un paciente con enfermedad bucodental.

b) Sabiendo que el paciente tiene una enfermedad bucodental calcular la probabilidad de que provenga de la segunda consulta.

Capítulo2:Estimación de parámetros poblacionales.

2.1. Estimación puntual.
2.2. Estimación por intervalo de confianza para la media poblacional.
2.3. Estimación por intervalo de confianza para la proporción poblacional.
2.4. Tamaño de muestra.
2.5. Ejercicios y problemas

OBJETIVOS:

Al finalizar este capitulo el alumno será capaz de:

1. Enfatizar sus conocimientos sobre estimación puntual y por intervalo.
2. Estimar un intervalo de confianza para la media y la proporción poblacional.
3. Aplicar el procesador estadístico profesional SPSS a la bases de datos de forma tal que puedan realizar una estimación puntual y por intervalo de confianza para la población.
4. Determinar el tamaño de una muestra.
5. Interpretar las estimaciones puntual y por intervalo.

2.1.- Estimación puntual.

En la primera parte del libro se estudio lo correspondiente a universo o población, muestra, sobre la selección de la muestra y los diferentes tipos de muestreo, es decir los muestreos probabilísticos y no probabilísticos.

Aprendimos que una muestra se selecciona para reducir los costos del trabajo investigativo del personal que se debe emplear en el mismo. Todo lo que se busca es que los datos obtenidos de una población y que aparecen en la muestra, pueden contener toda la información que se desee de ella. De lo que se trata es de extraerle esa información a la muestra, es decir a los datos muestrales sacarle toda la información de la población.

La muestra debe tener toda la información deseada para tener la posibilidad de extraerla, esto sólo se puede lograr con una buena selección de la muestra y un trabajo muy cuidadoso y de alta calidad en la recogida de los datos.
A partir de la muestra podemos realizar entonces una estimación de los parámetros poblacionales, pero que entendemos por parámetro

Parámetro:

Una parámetro es una medida usada para describir alguna característica de una población, tal como una media aritmética, una desviación estándar o una proporción de una población.
Ahora debemos conocer que entendemos por estimación.

¿Qué es una estimación?

Cuando queremos realizar un estudio de una población cualquiera de la que desconocemos sus parámetros, por ejemplo su media poblacional o la probabilidad de éxito si la población sigue una distribución binomial, debemos tomar una muestra aleatoria de dicha población a través de la cual calcular una aproximación a dichos parámetros que desconocemos y queremos estimar. Bien, pues esa aproximación se llama estimación.

Cuando los dos nuevos términos de arriba son usados, por ejemplo, el proceso de estimación en inferencia estadística puede ser descrito como el proceso de estimar un parámetro a partir del estadístico correspondiente.

Cuales son entonces esos estadísticos para cada parámetro

Parámetro	estadístico	Población
Media aritmética	$\bar{X}$	μ
Varianza	S^2	σ^2
Desviación Standard	S	σ
Proporción	p	P

Existen dos tipos de estimación que podemos realizar a cada parámetro poblacional, la estimación puntual y la estimación por intervalo.

Estimación puntual:

Una estimación puntual del valor de un parámetro poblacional desconocido (como puede ser la media μ, o la desviación estándar σ), es un número que se utiliza para aproximar el verdadero valor de dicho parámetro poblacional. A fin de realizar tal estimación, tomaremos una muestra de la población y calcularemos el parámetro muestral asociado ($\bar{x}$ para la media, *s* para la desviación estándar, etc.). El valor de este parámetro muestral será la estimación puntual del parámetro poblacional.

¿Qué propiedades debe cumplir todo buen estimador?

La diferencia entre el resultado obtenido de una muestra (un estadístico) y el resultado el cual deberíamos haber obtenido de la población (el parámetro correspondiente) se llama el error muestral o error de muestreo. Un error de muestreo usualmente ocurre cuando no se utiliza la población completa, sino que se toma una muestra para estimar las características de la población. El resultado de la media indica la precisión de la estimación de la población basada en el estudio de la muestra. Mientras más pequeño es el error muestral, mayor es la precisión de la estimación.

Luego un buen estimador debe ser:

Insesgado.

Eficiente.

Consistente.

Suficiente.

Estimador insesgado (o sin sesgo):

Si la media de las dispersiones de muestreo con un estadístico es igual que la del correspondiente parámetro de la población, el estadístico se llamara estimador sin sesgo, del parámetro.

Ejemplo: En el caso de la media poblacional, podemos decir que el estadístico media muestral es un estimador insesgado si se cumple que $E(\bar{x}) = \mu$, es decir el valor esperado de la media muestral es igual a la media poblacional.

Estimador Eficiente:

Si tenemos dos estadísticos de un mismo parámetro poblacional y ambos tienen el mismo valor esperado (o esperanza), el de menor varianza se llama un estimador eficiente..

Si consideramos todos los posibles estadísticos de un mismo parámetro poblacional siendo todos estimadores insesgados, se considerará eficiente a aquel de varianza mínima.

Estimador Consistente:

A medida que aumente el tamaño de la muestra n y se aproxima al tamaño de la población N, la probabilidad de que el estimador (estadístico) sea cercano al parámetro poblacional tiende a 1.

Por ejemplo $P(\bar{x} \to \mu) \to 1$ cuando $n \to N$

Estimador Suficiente:

Un estadístico se considerará estimador suficiente (como por ejemplo $\bar{x}$), si es un estimador que utiliza toda la información que contiene la muestra acerca del parámetro que pretende estimarse. Al decir que $\bar{x}$, es un estimador suficiente de la media poblacional μ, esto significa que ningún otro estimador de μ, puede dar una mayor información acerca del parámetro μ.

Estudiaremos a continuación dos de los estimadores puntuales más importantes en salud.

Estimación puntual para la media:

Sea X una variable aleatoria con distribución normal X ~ N(μ, σ^2), donde el parámetro μ representa la <u>media</u> de X en la población. Si el parámetro μ es desconocido, se puede estimar el mismo a partir de una muestra aleatoria: x1,x2,...,xn, de la población, para ello utilizaremos el estadístico media muestral:

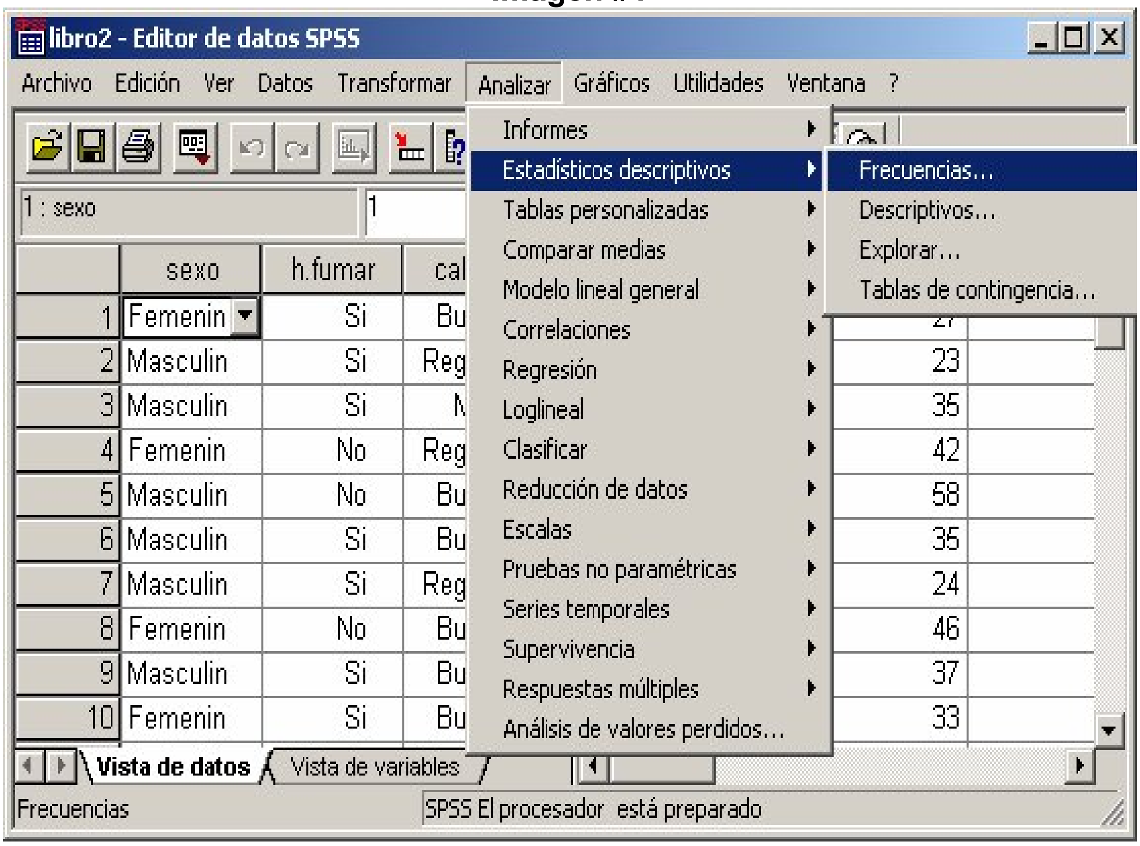
$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ejemplo #1: Se ha realizado un estudio con 20 pacientes en una consulta de estomatología. Se quiere realizar una estimación puntual de la hemoglobina promedio que poseen todos pacientes.

Solución:

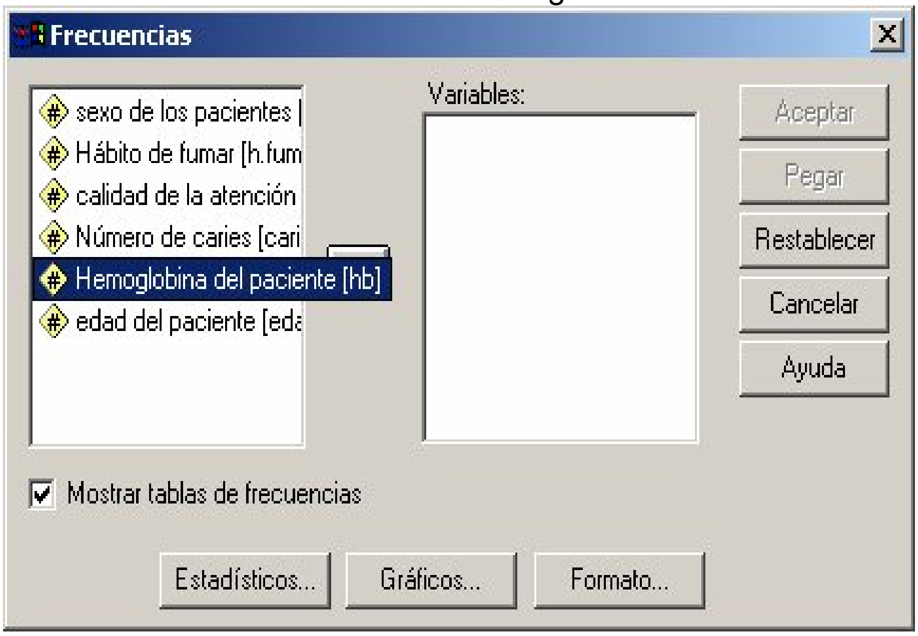
Para proceder con el SPSS para una estimación puntual se debe:

Imagen #1



Es decir en el menú seleccionar **Analizar,** luego **Estadísticos descriptivos** y de aquí la opción de **Frecuencias.** Al hacer clic en **Frecuencias** le aparecerá el siguiente cuadro de diálogo:

Imagen #2



Corresponde aquí pasar a la variable Hemoglobina para la casilla de Variables, desmarcar donde dice **Mostrar tablas de frecuencia.** A continuación seleccionar la opción de **Estadísticos** le aparecerá el siguiente cuadro de diálogo:

Imagen #3

Como estamos realizando una estimación puntual para la media, precisamente marcamos donde dice Media, que está debajo de lo referente a Tendencia central y luego teclear en Continuar y en la **Imagen #2** teclear en Aceptar.

Le debe aparecer una ventana de resultados donde le mostrará:

Estadísticos
Hemoglobina del paciente

N	Válidos	20
	Perdidos	0
Media		13,635

Por lo que la hemoglobina promedio de todos los pacientes debe ser 13,635.

Estimación puntual para la proporción:

Se quiere determinar la proporción o por ciento P de casos en la población que presentan determinada característica X (X=1, está presente la característica, X=0 está ausente). A partir de una muestra aleatoria de la población se puede estimar P.

Estimación puntual de P: El estimador puntual de la proporción poblacional es la proporción muestral p=a/n, donde n es el tamaño de la muestra y a la cantidad de sujetos de la muestra que presentan la característica.

$$\hat{P} = p = \frac{\sum_{i=1}^{n} x_i}{n}$$

Donde X_i son los elementos que poseen la característica y n la cantidad total de la muestra.

Ejemplo #2: Se ha realizado un estudio con 20 pacientes en una consulta de estomatología. Se quiere realizar una estimación puntual de la proporción de fumadores en todos pacientes.

Solución:

Para proceder con el SPSS para una estimación puntual se debe proceder igual que en la estimación puntual para la media.

Le debe aparecer una ventana de resultados donde le mostrará:

Estadísticos

Hábito de fumar

N	Válidos	20
	Perdidos	0
Media		,65

Por lo que la proporción de fumadores de los pacientes debe ser de 0,65 o lo que es lo mismo el 65 por ciento de todos los pacientes deben ser fumadores.

2.2- Estimación por intervalo de confianza para la media poblacional.

Ejemplo #3: Supongamos que se va a hacer un estudio con una población de 5 pacientes y cuyas edades son 45, 36, 54, 32, y 48 años de edad.

Si formamos todas las muestras posibles de tamaño 2 sin repetición obtendríamos:

(45, 36)	(45,54)	(45,32)	(45,48)
(36,54)	(36,32)	(36,48)	(54,32)
(54,48)	(32,48)		

Es decir, como fórmula general es n = $C_n^N = \binom{N}{n} = \dfrac{N!}{n!(N-n)!}$

En nuestro ejemplo es C_2^5 = 10

Al calcular la media de cada muestra de tamaño 2 obtenemos:

40,5	49,5	38,5	46,5
45,0	34,0	42,0	43,0
51,0	40,0		

Sin embargo al calcular la media poblacional obtenemos:

$$\mu = \frac{45 + 36 + 54 + 32 + 48}{5} = 43{,}0$$

Luego como usted puede apreciar en nuestro ejemplo solamente una muestra coincide con el verdadero valor de la media poblacional (incluso puede que en un estudio no coincida ninguno), mientras que en los otros casos el valor de la

media muestral o está por debajo o está por encima del valor de la media poblacional.
En un estudio donde la población sea con 100 individuos y supongamos que nuestra muestra sea de 20 individuos. ¿Cuántos posibles muestran diferentes se pueden formar, de tamaño 20?

Aplicando la fórmula es: $C_{20}^{100} = \dfrac{100!}{20\,!\,(100-20)\,!} = 535983370403809682970$

Por lo que usted puede apreciar resulta una cantidad muy grande, luego como podemos saber al seleccionar una muestra y realizar un a estimación puntual, que error estaremos cometiendo (error muestral). Es por ello que en vez de realizar una estimación donde se obtenga un único valor se propone todo un intervalo de valores. Asignar al parámetro un rango de valores entre los cuales se espera que pueda encontrarse con una alta probabilidad el valor real del parámetro.

Intervalo de confianza para el parámetro θ
P (θ1≤θ≤θ2) = 1-α donde θ es el parámetro poblacional
θ1 = Estadístico – error muestral y θ2 = Estadístico + error muestral. θ1 y θ2 son v.a. Sus valores cambian de unas muestras a otras, el verdadero valor del parámetro θ está en el intervalo en el (1-α)% de las veces.
Para un intervalo fijo θ estará o no estará en este intervalo. El nivel de confianza representa la proporción de intervalos que contienen el parámetro.

Intervalo de confianza para la media poblacional.
Sea X una variable aleatoria con distribución normal X ~ N(μ, σ^2), donde el parámetro μ representa la <u>media</u> de X en la población. Si el parámetro μ es desconocido, se puede estimar el mismo a partir de una muestra aleatoria: x1,x2,...,xn, de la población, de dos formas diferentes:
•Estimación puntual (como estudiamos anteriormente).
•Estimación por intervalo de confianza
<u>Estimación por intervalo de confianza:</u> Se fija un nivel de confiabilidad (γ) del 95% o del 99% y a partir de la muestra se construye el siguiente intervalo:

(I) o (II)

$$\bar{x} - k\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + k\frac{\sigma}{\sqrt{n}} \qquad \bar{x} - k\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + k\frac{s}{\sqrt{n}}$$

La primera expresión (I) se utiliza cuando σ (desviación estándar poblacional) es conocida, en caso contrario se utiliza la expresión (II) sustituyendo σ por s (desviación estándar muestral). s es el estimador puntual de σ.

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

γ representa el nivel de confiabilidad (95% o 99%)

Un nivel de confiabilidad del 95% se interpreta de la siguiente manera. Si se extrajeran 100 muestras aleatorias de tamaño fijo n, en 95 de los intervalos determinados se encontrará el parámetro μ.

$\bar{x}, s$ Media y desviación estándar muestral

Cuando σ es conocida, k toma el valor de 1.96 si γ=95% y k=2.58 si γ=99%. Estos valores representan los percentiles 97.5 y 99.5 de la distribución normal estándar.

Ejemplo #4:

En una investigación sobre enfermedad parodontal se determinó el fósforo inorgánico en le plasma sanguíneo de 50 pacientes y se obtuvo un promedio de 3,62 mg%. De estudios anteriores se ha establecido que la proporción es normal con una desviación Standard de 0,4mg%. Determine un intervalo de confianza para el promedio poblacional con un 95% de confiabilidad.

Solución:

<u>Datos:</u>

n = 50

$\bar{x}$ = 3,62

σ = 0,4

Por lo que el intervalo que tenemos que utilizar es:

$$\bar{x} - k\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + k\frac{\sigma}{\sqrt{n}}$$

γ=95% Sustituyendo obtenemos que:

$\downarrow$ $3,62 - 1,96\ \frac{0,4}{\sqrt{50}} \le \mu \le\ 3,62 + 1,96\ \frac{0,4}{\sqrt{50}}$

k = 1,96 Realizando las correspondientes operaciones obtenemos que: $3,51 \le \mu \le 3,73$

Respuesta: El promedio poblacional de fósforo inorgánico en el plasma sanguíneo está entre 3,51 y 3,73 mg% con un 95% de confiabilidad.

Cuando σ es desconocida *k* es un valor que depende del nivel de confiabilidad γ y del tamaño de la muestra (n)

Si n > 30 entonces:

k = 1,96 si g = 95 %

k = 2,58 si g = 99 %

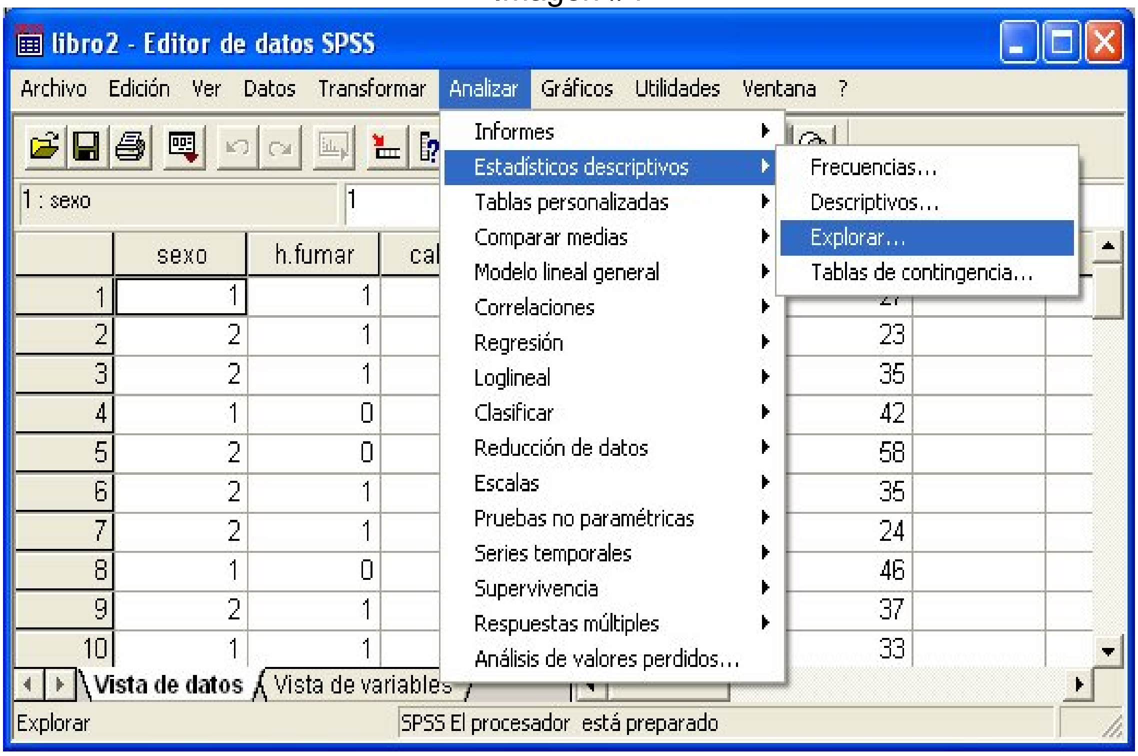
Si n ≤ 30 entonces *k* toma del valor de la variable t con distribución t de Student correspondiente al nivel de confiabilidad y grados de libertad igual a n-1 (Ver la tabla en el Anexo 3)

Ejemplo #5: Se ha realizado un estudio con 20 pacientes en una consulta de estomatología. Se quiere realizar una estimación por intervalo de la hemoglobina promedio que poseen todos los pacientes. (Ver datos en el Anexo)

Solución: (Aquí σ es desconocida, se poseen datos de una muestra)

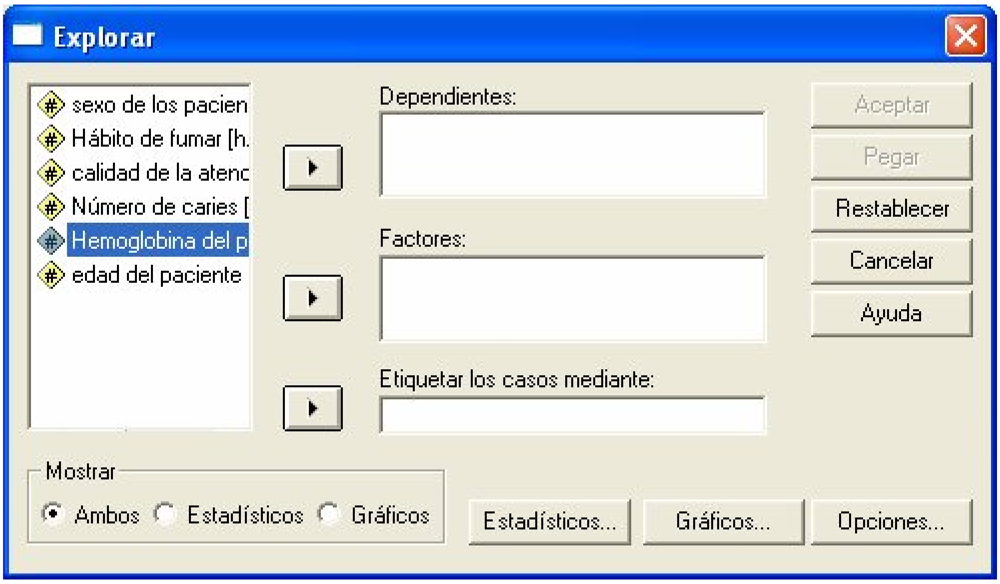
Para proceder con el SPSS para una estimación por intervalo se debe:

Imagen #4



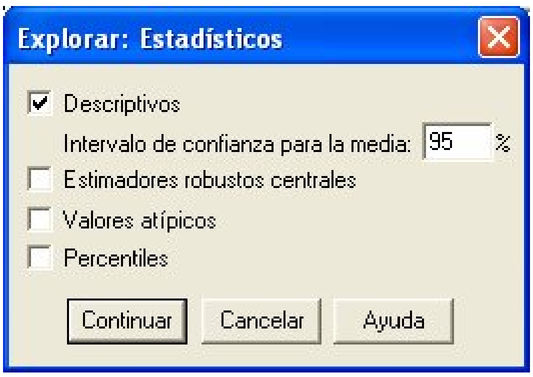
Es decir en el menú seleccionar **Analizar,** luego **Estadísticos descriptivos** y de aquí la opción de **Explorar.** Al hacer clic en **Explorar** le aparecerá el siguiente cuadro de diálogo:

Imagen #5



Corresponde aquí pasar a la variable Hemoglobina para la casilla Dependientes. A continuación seleccionar la opción de **Estadísticos** le aparecerá el siguiente cuadro de diálogo:

Imagen #6



Como usted puede apreciar si la confiabilidad es del 95% no tiene que modificarlo, pero si va a utilizar otro nivel de confiabilidad (Por ejemplo 99%), debe en la casilla marcada cambiarla. A continuación presionar la tecla Continuar y luego en la **Imagen # 5** seleccionar Aceptar.

Le debe aparecer una ventana de resultados donde le mostrará:
Descriptivos

			Estadístico	Error típ.
Hemoglobina del paciente	Media		13,635	,302
	Intervalo de confianza para la media al 95%	Límite inferior	13,004	
		Límite superior	14,266	
	Media recortada al 5%		13,650	
	Mediana		13,700	
	Varianza		1,819	
	Desv. típ.		1,349	
	Mínimo		11,6	
	Máximo		15,4	
	Rango		3,8	
	Amplitud intercuartil		2,525	
	Asimetría		-,188	,512
	Curtosis		-1,535	,992

La Media nos informa la hemoglobina promedio de todos los pacientes que es de 13,635, o sea una estimación puntual.
Debajo aparece Intervalo de confianza para la media al 95%, por lo que la hemoglobina promedio de todos los pacientes está entre 13,004 y 14,266 con 95% de confiabilidad.

2.3- Estimación por intervalo de confianza para la proporción poblacional.

Se quiere determinar la proporción o por ciento P de casos en la población que presentan determinada característica X (X=1, está presente la característica, X=0 está ausente). A partir de una muestra aleatoria de la población se puede estimar P de dos formas diferentes:
Estimación puntual de P (como ya estudiamos anteriormente).
Estimación por intervalo de confianza:
Se parte de un nivel de confianza (95% y 99%) y a partir de la muestra se construye el siguiente intervalo en el cuál deberá estar el valor exacto de P:

$$p - k\sqrt{\frac{pq}{n}} \leq P \leq p + k\sqrt{\frac{pq}{n}}$$

q= 1-p (proporción)   q=100-p (por ciento)

Ejemplo #6: Se ha realizado un estudio con 20 pacientes en una consulta de estomatología. Se quiere realizar una estimación por intervalo de la proporción de fumadores en todos pacientes.

Solución:

Para proceder con el SPSS para una estimación por intervalo se debe proceder igual que en la estimación por intervalo para la media.

Le debe aparecer una ventana de resultados donde le mostrará:

Descriptivos

			Estadístico	Error típ.
Hábito de fumar	Media		,65	,11
	Intervalo de confianza para la media al 95%	Límite inferior	,42	
		Límite superior	,88	
	Media recortada al 5%		,67	
	Mediana		1,00	
	Varianza		,239	
	Desv. típ.		,49	
	Mínimo		0	
	Máximo		1	
	Rango		1	
	Amplitud intercuartil		1,00	
	Asimetría		-,681	,512
	Curtosis		-1,719	,992

La Media nos informa la proporción de fumadores de todos los pacientes que es de 0,65, o sea una estimación puntual (65 % de fumadores).

Debajo aparece Intervalo de confianza para la media al 95%, por lo que proporción de fumadores de todos los pacientes está entre ,42 y 0,88 con 95% de confiabilidad, es decir la proporción de fumadores está entre un 42% y un 88%.

2.4- Tamaño de muestra.

Todo estudio en salud lleva implícito la determinación del tamaño muestral necesario para la ejecución de la investigación. El no seleccionar la muestra adecuada, puede llevarnos a dos situaciones diferentes: primera que realicemos el estudio sin el número adecuado de individuos, con lo cual no podremos ser precisos al estimar los parámetros y además no encontraremos diferencias significativas cuando en la realidad sí existen. La segunda situación es que podríamos estudiar un número innecesario de individuos, lo cual lleva implícito no solo la pérdida de tiempo e incremento de recursos innecesarios sino que además la calidad del estudio.

El tamaño de la muestra depende de tres factores:

1) Variabilidad del universo que se estudia, pues mientras más variable sea este, mayor ha de ser el tamaño de la muestra.
2) Precisión que se quiere en los resultados, es decir la magnitud del error que podemos tolerar. Se comprende que para afirmar que el promedio del peso de un grupo de individuos está entre 40 y 60 kilos se necesitará una muestra más pequeña, que si quisiéramos afirmar que dicho promedio está entre 50 y 51 kilos.
3) Margen de certeza (confiabilidad) que se desee obtener (95% o 99%), pues para determinada precisión mientras mayor sea la certeza que se busca, mayor debe ser el tamaño de la muestra

Para el caso de la estimación de μ, se establece que:

$\left|\mu - \bar{x}\right| < d$, el error de muestreo sea menor que un valor d . A mayor precisión de la estimación, diminuye el valor de d y aumenta el tamaño de la muestra

Si se desea una estimación de μ que diste de ella a lo sumo en d unidades:

$$n = \frac{k^2 \sigma^2}{d^2}$$

Donde:

k representa la confiabilidad.

σ^2 representa la variabilidad

d^2 representa la precisión.

Ejemplo #7: Dado el ejemplo #4. ¿Cuántos pacientes habría que seleccionar para estimar dicho promedio poblacional con una precisión de 0,1 mg?

Solución:

Datos

k = 1,96

σ = 0,4

d = 0,1

Por lo que la fórmula que tenemos que utilizar es:

$$n = \frac{k^2 \sigma^2}{d^2}$$

Sustituyendo obtenemos que:

$$n = \frac{(1{,}96)^2 (0{,}4)^2}{(0{,}1)^2}$$

Realizando las operaciones correspondientes obtenemos que:

n = 62

Respuesta: Habría que seleccionar 62 pacientes.

Si la población es finita, es decir conocemos el total de la población y desearíamos saber cuantos del total tendríamos que estudiar, la respuesta sería:

$$n = \frac{N k^2 \sigma^2}{d(N-1)^2 K^2 \sigma^2}$$

Para el caso de la estimación de P, se establece que:

$|P - p|$ <d

Si se desea una estimación de P que diste de ella a lo sumo en d unidades:

$$n = \frac{k^2 pq}{d^2}$$

p=1-q

p=100-q

Ejemplo #8: Dado el ejemplo #6. ¿Cuántos pacientes habría que seleccionar para estimar dicha proporción poblacional y no difiera del valor real en más de 0,15?

Datos:

k = 1,96

p = 0,65

q = 0,35

Por lo que la fórmula que tenemos que utilizar es:

$$n = \frac{k^2 pq}{d^2}$$

d = 0.15

Sustituyendo obtenemos que:

$$n = \frac{(1{,}96)^2 (0{,}65)(0{,}35)}{(0{,}15)^2}$$

Realizando las operaciones correspondientes obtenemos que:

n = 39

Respuesta: Habría que seleccionar 39 pacientes.

Si la población es finita, es decir conocemos el total de la población y deseásemos saber cuántos del total tendremos que estudiar la respuesta seria:

$$n = \frac{\mathrm{N}\, k^2 pq}{d(N-1)^2 K^2 pq}$$

El tamaño muestral ajustado a las pérdidas:

En todos los estudios es preciso estimar las posibles perdidas de pacientes por razones diversas (pérdida de información, abandono, no respuesta….) por lo que se debe incrementar el tamaño muestral respecto a dichas pérdidas.

El tamaño muestral ajustado a las pérdidas se puede calcular:

Muestra ajustada a las pérdidas = n (1 / 1–R)

- n = número de sujetos sin pérdidas
- R = proporción esperada de pérdidas

Ejemplo #9: Así si dado el ejemplo#7 esperamos tener un 10% de pérdidas el tamaño muestral necesario sería:

62 (1 / 1-0.10) = 69.

Respuesta: Realmente habría que seleccionar 69 pacientes.

2.5 EJERCICIOS Y PROBLEMAS DEL CAPITULO

1.- Se sabe el tiempo en minutos que tardan 30 especialistas de Prótesis en realizar cierta operación. Formar una distribución de frecuencias de 5 intervalos.

164	132	125	157	146	140	147	136	148	144
165	135	128	119	153	156	163	135	150	176
161	145	135	142	147	142	173	146	168	126

¿Cuál es el valor el tiempo promedio que tardan los especialistas de Prótesis en realizar cierta operación, suponiendo que la muestra ha sido obtenida por muestreo aleatorio simple sobre una población normal?

2.- La edad promedio de 15 mujeres embarazadas, al analizar su estado de salud bucal, fue de 25, con una desviación típica de 5.32. Obtener un intervalo de confianza para la media al 99%, suponiendo que la muestra fue extraída mediante muestreo aleatorio simple sobre una población normal.

3.- Un cardiólogo se encuentra interesado en encontrar límites de confianza al 95%, para el tiempo utilizado en la cantidad de cambios de cura en el servicio de Endodoncia de un Clínica Dental. Obtenerlos si en 50 días se obtuvo un valor de la media de 75 y de la desviación típica de 10. Suponemos que el comportamiento de la variable aleatoria es normal.

4.- En una muestra de 25 pacientes que asistieron a consultas de Estomatología en la provincia de Villa Clara en el año 2003, se obtuvo un peso medio de 62.5 Kg. De estudios anteriores se sabe que la variable peso posee distribución normal con una desviación típica de 14 Kg.

Obtener un intervalo de confianza (al 95%) para el peso medio poblacional.

¿Cuántos pacientes habría que tomar para estimar dicha media con una precisión de 15 Kg?

5.- La vida media de la actividad de cierto fármaco después de su fabricación y envasado, es normal $N(\mu, \sigma^2)$ con σ =40 días. Se desea enviar un lote de medicamentos de modo que la vida media muestral no sea inferior a 1 180 días con probabilidad de 0,95. Calcular el tamaño del lote.

6.- Se desea estimar la talla promedio de pacientes que asistieron a consultas de Estomatología con una precisión de 5 segundos. Ante la ausencia de cualquier información acerca de la variabilidad del tiempo de sangría en este tipo de individuos, se tomó una muestra preliminar de 10 pacientes, en los que se obtuvieron las siguientes tallas (en centímetros):

169, 178, 165, 155, 145, 152, 153, 145, 162, 147.

Determinar el tamaño mínimo de muestra, al 95%, para cumplir el objetivo anterior.

7.- Un investigador está interesado en estimar la proporción del tipo de anestesia GET (General endotraqueal) que se les empleó a un grupo de pacientes. Su experiencia le indica que sería sorprendente que tal proporción supere el valor de ½. ¿Qué tamaño de muestra debe tomar para estimar la anterior proporción, con una confianza del 99%, para que el valor estimado no difiera del valor real en más de 0,03?

8.- Se realizó un estudio con un grupo de pacientes para evaluar sus conocimientos sobre enfermedad periodontal; si de 74 pacientes observados 42 poseen un mal conocimiento, dar una estimación puntual y un intervalo de la proporción de los que poseen mal conocimiento. ¿Qué número de pacientes habría que observar para estimar la proporción de los que tienen un mal cocimiento con un error inferior a 0,05 y una confianza del 95%?

9.- En la provincia de Villa Clara durante el año 2003 se tomó una muestra aleatoria de 125 pacientes, de los cuales 60 fumaban.

Estímese la proporción de fumadores en dicha provincia.

Si queremos estimar dicha proporción con un error máximo del 4%, para una confianza del 95%, ¿qué tamaño de muestra debemos tomar?

10.- Se quiere estimar la incidencia de molestias con su prótesis en un estudio con pacientes del Área X en Santa Clara durante el año 2003. ¿Cuántos pacientes tenemos que observar para, con una confianza del 95%, estimar dicha incidencia con un error del 2% en los siguientes casos:

Sabiendo que en un sondeo previo se ha observado un 30% de molestias con sus prótesis.
Sin ninguna información previa.

CAPITULO 3. PRUEBAS DE HIPÓTESIS PARAMETRICAS.

3.1 Conceptos básicos

3.2 Comparación entre el promedio de una muestra y el promedio del universo

3.3 Estudios comparativos para dos muestras.

3.4 Ejercicios y problemas del capítulo.

OBJETIVOS:

1. Identificar cuando deberá usarse el método de prueba de hipótesis para una media poblacional en problemas biomédicos
2. Identificar cuando deberá usarse el método de prueba de hipótesis en una comparación para diferencias de medias entre estos valores en problemas biomédicos donde los datos del misma se reflejan a través de este parámetro estadístico.
3. Explicar el significado de los conceptos fundamentales relativos a una prueba de hipótesis cualquiera.
4. Usar un procesador de datos estadísticos profesional con vista a servirse del mismo para resumir un conjunto de datos biomédicos en forma automatizada y así obtener información necesaria par poder aplicar e interpretar el método de las pruebas de hipótesis.

3.1. CONCEPTOS BÁSICOS.

Decisiones Estadísticas: son las decisiones que tenemos que tomar a partir de las inferencias muéstrales.

Ejemplos:

- Determinar si un determinado medicamento es efectivo para curar una enfermedad.
- Se quiere evaluar entre dos tratamientos para aliviar un dolor, cual es más efectivo, para lo cual se le aplica a cada uno de los pacientes primero un tratamiento y otro después de cierto tiempo prudencial, recogiéndose cada vez si el paciente alivio o no el dolor.
- Se aplica a un determinado conjunto de niños de un círculo infantil un proyecto educativo de salud estomatológica. ¿Cómo determinar si es eficaz o no?

Hipótesis estadística: Una hipótesis estadística es una suposición relativa a uno o varios parámetros de distribución de probabilidades de una o varias variables aleatorias.

Toda hipótesis estadística está compuesta por una hipótesis nula y una hipótesis alternativa.

La hipótesis nula (Ho) generalmente representa la hipótesis que se quiere aceptar, mientras que la hipótesis alternativa (H1) representa generalmente lo contrario.

Ejemplo:

De experiencias anteriores se sabe que el tiempo promedio que invierten los especialistas en prótesis estomatológica para realizar un tratamiento es de tres horas. Se quiere saber si este tiempo promedio ha variado producto de un entrenamiento que han realizado con otras técnicas.

Luego:

Ho: El tiempo promedio no ha variado ($\mu = 3$ horas)

H1: El tiempo promedio ha variado ($\mu \neq 3$ horas)

El problema consiste ahora en aceptar o rechazar la hipótesis Ho sobre la base de un criterio certero y objetivo.

Dócima o prueba de hipótesis: Llamamos dócima de una prueba estadística a todo procedimiento cuyo resultado final implica aceptar o rechazar la hipótesis considerada sobra la base de una muestra aleatoria.

Tipos de errores: Al tomar una decisión estadística cometemos el riesgo de cometer uno de los dos posibles errores: error tipo I o error tipo II.

Se llama **error tipo I o error α,** o también **nivel de significación** a la probabilidad de rechazar la hipótesis Ho siendo cierta. Se llama **error tipo II o error β** a la probabilidad de aceptar Ho siendo falsa.

Se decide:	Cuando en realidad	
	Ho es verdadera	Ho es falsa
Aceptar Ho	Correcto	Error Tipo II
Rechazar Ho	Error Tipo I	Correcto

Nivel de significación: En la práctica es normal tomar un nivel de significación de 0,05 o de 0,01 aunque se puede tomar otros valores.

Un nivel de significación (por ejemplo) de 0,05 o 5 % al diseñar una prueba de hipótesis significa entonces que hay alrededor de cinco oportunidades sobre 100 de que rechacemos la hipótesis cuando esta debe ser aceptada, es decir tenemos un 95 % de confianza de haber tomado una decisión correcta.

Estadígrafo: es un número calculado a partir de una muestra.

Región crítica o región de rechazo: de una dócima al conjunto de valores del estadígrafo para los cuales se rechaza la hipótesis nula.

El nivel de significación debe indicarse conjuntamente con la decisión tomada para señalar que al tomar tal decisión se está corriendo un riesgo, pero en parte, controlado.

Todo problema relativo a la prueba de hipótesis estudiada deben tenerse en cuenta los siguientes pasos:

I. Identificar una(s) variable(s) aleatoria(s) que tienen una distribución conocida y en consecuencia se identifica la prueba de hipótesis a aplicar.
II. Plantear la hipótesis nula y la alternativa
III. Determinar el nivel de significación
IV. Calcular el valor del estadígrafo correspondiente
V. Tomar la decisión estadística (aceptar o rechazar Ho)
VI. Interpretación de los resultados obtenidos.

3.2. COMPARACIÓN ENTRE EL PROMEDIO DE UNA MUESTRA Y EL PROMEDIO DEL UNIVERSO.

PROBLEMA:

Se desea saber con un 99 % de certeza si el número de pulsaciones por minutos del grupo de pacientes (anexo 1) difieren significativamente del valor de 70 pulsaciones por minutos considerados como normal.

Como se puede apreciar en este problema tenemos un conjunto de pacientes al que se le tomo el número de pulsaciones por minutos al llegar al cuerpo de guardia y lo queremos comparar con un valor considerado como normal, en investigaciones anteriores, para una población.

Para resolver este problema aplicamos una prueba estadística llamada **T de Student para una muestra,** para poder aplicar esto debemos garantizar que:

- La variable tenga una distribución normal
- Los datos sean cuantitativos
- Se conozca el promedio poblacional

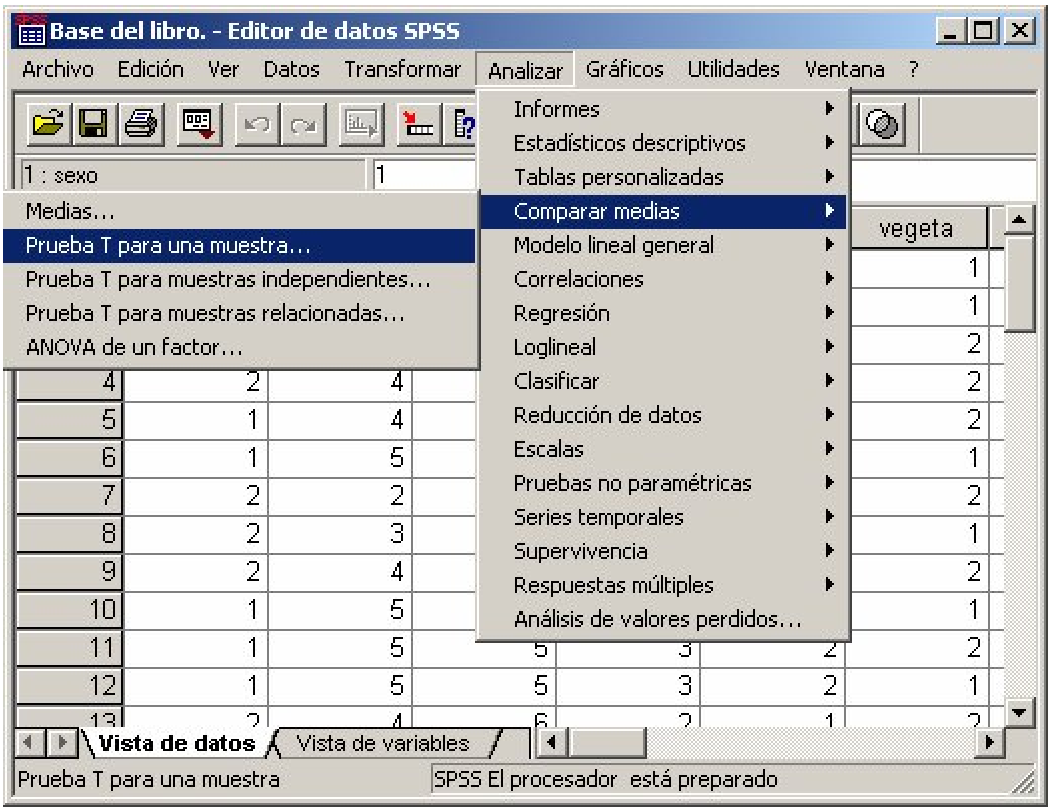
Los pasos a seguir para aplicar la prueba son:

Pasos	I	II	III
1	Ho: $\mu=\mu o$ H1: $\mu\neq\mu o$	Ho: $\mu=\mu o$ H1: $\mu>\mu o$	Ho: $\mu=\mu o$ H1: $\mu<\mu o$ Prueba unilateral

	Prueba Bilateral o de dos colas	Prueba unilateral (derecha) o de una cola	(izquierda) o de una cola
2	α=5% (0.05) o α=1% (0.01)		
3	$t = \dfrac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$ Si Ho es cierta, t tiene distribución T de Student con gl=n-1		
4	Se rechaza Ho		
	Si $p < \alpha$	Si $p < 2\alpha$ y $t > 0$	Si $p < 2\alpha$ y $t < 0$
5	En caso de no rechazar Ho, se dice que μ no difiere significativamente (α=5%) o muy significativamente (α=1%) de μo		

Esta prueba la aplicamos utilizando el procesador estadístico SPSS (versión 11). La forma de ejecutar el procedimiento es:

Analizar ⇨ **Comparar medias** ⇨ **Prueba t de Student para una muestra.** Obtenemos los siguientes cuadros de dialogo:



Seguidamente aparece el siguiente cuadro de dialogo:

Ahora Ud debe trasladar la variable a contrastar y en el valor de la prueba situar la media poblacional, a continuación en el botón opciones debe especificar en nivel de significación seguidamente da aceptar, obteniendo los resultados de la prueba aplicada

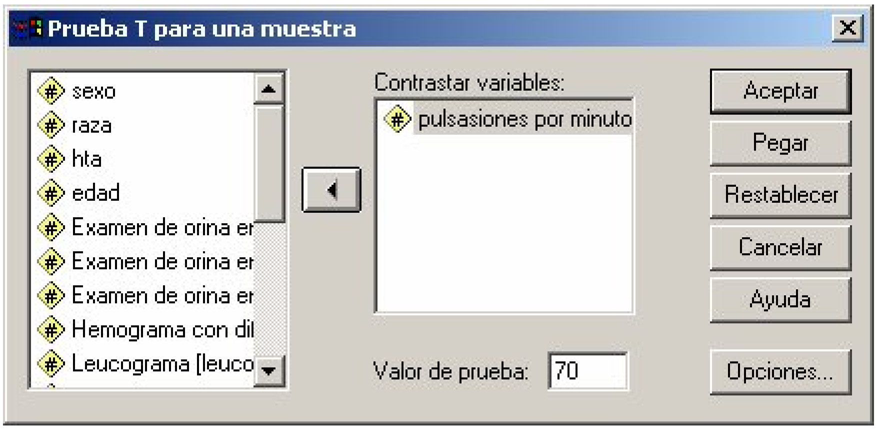
En el casos de nuestro problema práctico planteado al inicio Usted debe:

1. Plantear las hipótesis:

$$Ho: \mu = 70 \text{ pulsaciones por minutos}$$

$$H1: \mu \neq 70 \text{ pulsaciones por minutos}$$

2. Determinar el nivel de significación 1 %
3. Determinar los valores del estadígrafo, para ello: trasladar al cuadro Contrastar variables la variable pulsaciones por minutos y abajo en Valor de prueba escribir 70 Finalmente hacer clic en Aceptar y le aparecerá:



En la ventana de resultados la siguiente imagen:

Prueba T

Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
pulsasiones por minutos en el cuerpo de guardia	40	68,30	5,748	,909

Prueba para una muestra

	Valor de prueba = 70					
					99% Intervalo de confianza para la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
pulsasiones por minutos en el cuerpo de guardia	-1,871	39	,069	-1,70	-4,16	,76

$t = -1.871 \quad p = 0.069 \quad \alpha = 0.01$

4. Si $p < \alpha$ rechazamos Ho. Como: $0.069 > 0.01$ no rechazamos Ho
5. En la interpretación planteamos que el número de pulsaciones por minutos de los pacientes al llegar al cuerpo de guardia no difiere muy significativamente de 70 pulsaciones por minutos considerados como normal

3.3 ESTUDIOS COMPARATIVOS PARA DOS MUESTRAS.

Los estudios comparativos tienen como objetivo investigar si existe diferencia entre dos o mas grupos que se estudian y tratar de determinar las causas que la explican.

En las ciencias experimentales fundamentalmente se busca comparar dos tratamientos que reciben dos grupos de pacientes, dos métodos de enseñanza, etc., y para ello debemos tener un grupo de control.

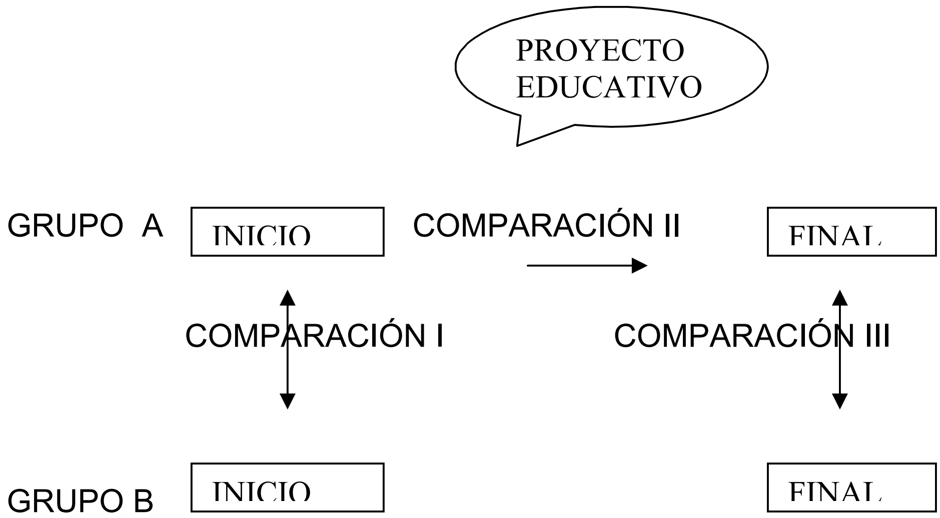
El grupo de control debe ser escogido de forma tal que sea perfectamente comparable con el grupo experimental. Este grupo de control puede ser escogido de diferentes formas, las más frecuentes son:

- **Control histórico:** los resultados observados en el grupo experimental se compara con aquellos obtenidos en tiempos pasados.
- **Control simultaneo:** en este caso el grupo de control debe tener similitud a los del grupo experimental en todas aquellas variables importantes sen relación con el problema que se estudia.
- **Individuos del grupo experimental como sus propios controles:** en este caso a cada individuo se le hacen las mismas mediciones antes y después de aplicarle un procedimiento.

PROBLEMA PRACTICO.

Ud., es un Estomatólogo General Básico que al ser ubicado en un área de salud encuentra que sus pacientes presentan diferencias en cuanto a su nivel de conocimiento sobre salud bucodental, por lo que decide aplicar un Proyecto Educativo. Con el objetivo de valorar su efectividad o no divide a la población dos grupos (A y B) de los cuales extrae una muestra aleatoria simple de 15 personas de cada grupo, al grupo A le aplica el proyecto mientras que el B solamente recibe la influencia educativa de los medios de difusión masiva, al cabo de tres meses comprueba los conocimientos en ambos grupos. Cómo proceder estadísticamente para ver si hay diferencias significativas en el grupo A en comparación con el grupo B.

Esquemáticamente lo podemos representar así:



¿QUÉ DEBE OCURRIR?

COMPARACIÓN I: Debemos garantizar que no existan diferencias significativa entre las notas promedio de los grupos A y B es decir:

Ho: La nota promedio del grupo A y B son iguales

H1: La nota promedio de los grupos A y B son desiguales

COMPARACIÓN II: Debemos observar si existen diferencias significativas entre las notas promedios del grupo A al inicio y al final, debiendo ser superior en la etapa final.

Ho: La nota promedio del grupo A al inicio y al Final son iguales

H1: La nota promedio del grupo A al inicio es inferior a las del final

COMPARACIÓN III: Debemos observar si existen diferencias significativas entre las notas promedio del grupo A y B al finalizar debiendo ser superiores en el grupo A

Ho: La nota promedio del grupo A y B son iguales

H1: La nota promedio del grupo A son superiores a las del grupo B

La comparación I y III reciben también el nombre de comparación vertical (trasversal), mientras que la comparación II recibe el nombre de comparación horizontal (longitudinal).

El grupo A es el grupo experimental mientras que el grupo B es el grupo de control.

COMPARACIÓN VERTICAL. PRUEBA T DE STUDENT PARA COMPARAR LAS MEDIAS EN DOS POBLACIONES INDEPENDIENTES:

Se desea saber con un 99 % de certeza si el número de pulsaciones por minutos de los hombres y mujeres del grupo de pacientes (anexo 1) difieren significativamente, conociendo que esta variable tiene una distribución normal.

Como se puede apreciar en este problema tenemos un conjunto de pacientes al que se le tomó el número de pulsaciones por minutos al llegar al cuerpo de guardia y lo queremos comparar tomando como criterio de comparación el sexo.

Para resolver este problema aplicamos una prueba estadística llamada **T de Student para dos muestras independientes,** para poder aplicar esto debemos garantizar que:

- La variable tenga una distribución normal
- Los datos sean cuantitativos
- Se conozca que sean dos grupos independientes, es decir que la intercepción de los elementos de ambos grupos de cómo resultados un conjunto nulo.

¿Qué nos plantea esta prueba estadística?

Sea X una variable aleatoria que se distribuye normalmente en dos poblaciones independientes A y B:

$$X \sim N(\mu_A, \sigma_A) \text{ y } X \sim N(\mu_B, \sigma_B)$$

Se desea comparar los promedios de X de ambas poblaciones (μ_A VS μ_B), partiendo de los valores de X correspondientes a dos muestras aleatorias (una extraída de la población A y la otra de B).
m-Tamaño de la muestra de la población A, n es el tamaño de la muestra de la población B

$\bar{x}_A$, S_A media y desviación standard de X en la muestra de A

$\bar{x}_B$, S_B media y desviación standard de X en la muestra de B

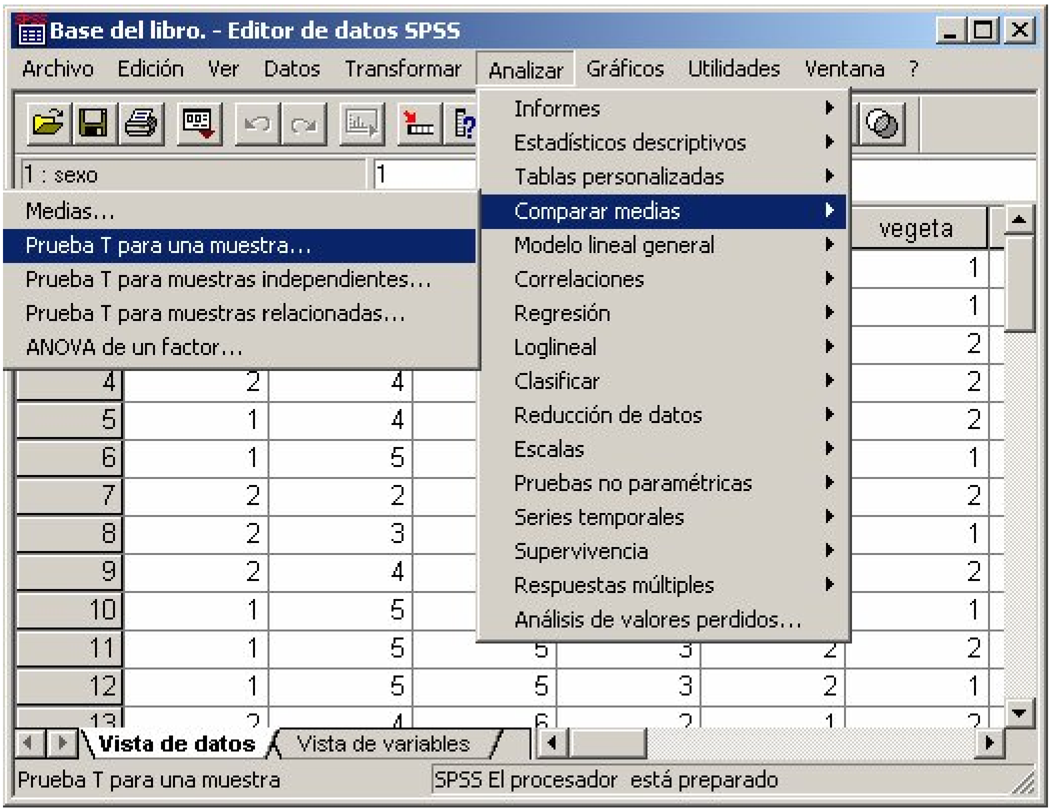
Los pasos a seguir para aplicar la prueba son:

Paso	I	II	III
1	Ho: $\mu A=\mu B$ H1: $\mu A\neq\mu B$ Prueba Bilateral o de dos colas	Ho: $\mu A=\mu B$ H1: $\mu A>\mu B$ Prueba unilateral (derecha) o de una cola	Ho: $\mu A=\mu B$ H1: $\mu A<\mu B$ Prueba unilateral (izquierda) o de una cola
2	α=5% (0.05) o α=1% (0.01)		

<table>
<tr>
<td>3</td>
<td colspan="3">
Verificar si hay homogeneidad de varianzas ($\sigma^2 A=\sigma^2 B$).
Prueba F de Fisher (F y p_F) Si $p_F<\alpha$, se rechaza la hipótesis de homogeneidad.
Homogeneidad de varianzas
$$t=\frac{\overline{X}_A-\overline{X}_B}{\sqrt{\frac{(m-1)S_A^{\ 2}+(n-1)S_B^{\ 2}}{m+n-2}}}\sqrt{\frac{mn}{m+n}}$$
Si Ho es cierta, t se distribuye T de Student con grados de libertad igual a m+n-2.
p_t es la significación de t
<u>No Homogeneidad de varianzas</u>
$$t=\frac{\overline{X}_A-\overline{X}_B}{\sqrt{\frac{S_A^{\ 2}}{m}+\frac{S_B^{\ 2}}{n}}}$$
Si Ho. Es cierta se distribuye T de Student con grados de libertad igual a:
</td>
</tr>
<tr>
<td>4</td>
<td>$p_t < \alpha$</td>
<td>$p_t < 2\alpha$
y
($t>0$) o $X_A>X_B$</td>
<td>$p_t < 2\alpha$
y
($t<0$) o $X_A<X_B$</td>
</tr>
</table>

Esta prueba la aplicamos utilizando el procesador estadístico SPSS (versión 11). La forma de ejecutar el procedimiento es:

Analizar ⇨ **Comparar medias** ⇨ **Prueba t de Student para muestras independientes.** Obtenemos los siguientes cuadros de dialogo:



Seguidamente aparece el siguiente cuadro de dialogo:

En contrastar variable usted traslada la variable cuantitativa y la variable de agrupación la que contiene los grupos que va a formar, definiendo los mismos posteriormente.

En el casos de nuestro problema práctico planteado al inicio Usted debe:

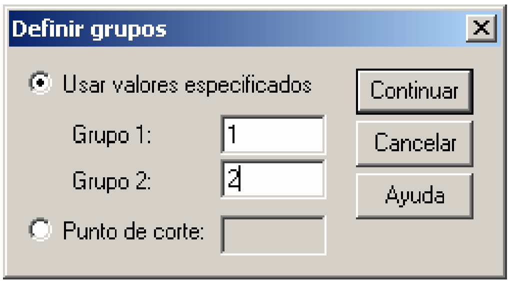
1. Plantear las hipótesis:

i. Ho: μF =μ M pulsaciones por minutos

ii. H1: μ F$\neq$ μM pulsaciones por minutos

2. Determinar el nivel de significación 99 %

3. Determinar los valores del estadígrafo, para ello: trasladar al cuadro Contrastar variables la variable pulsaciones por minutos y abajo en variable de asignación el sexo definiendo los grupos Masculinos y Femeninos



Seguidamente presiona el botón continuar y posteriormente aceptar obteniéndose el siguiente cuadro de resultados:

Estadísticos de grupo

	SEXO	N	Media	Desviación típ.	Error típ. de la media
pulsaciones por minutos en el cuerpo de guardia	Femenino	19	66,68	5,841	1,340
	Masculino	21	69,76	5,384	1,175

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias					
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	99% confi d Infer
pulsaciones por minu en el cuerpo de guarc	Se han asumido varianzas iguales	,727	,399	-1,734	38	,091	-3,08	1,775	-7,8
	No se han asumi varianzas iguales			-1,727	36,761	,093	-3,08	1,782	-7,9

$F = 0.727$

$P_{F\,=}\ 0.399$

$t = -\ 1.734$

$p = 0.091$

Interpretación:

Prueba F. De Fisher como 0.091 > 0.01 nos dice que la variabilidad inicial de ambos grupos es la misma y nos sugiere Utilizar el t-student con varianzas iguales

Si la $p < \alpha$ rechazamos Ho, pero en nuestro caso 0.091 > 0.01 por lo que aceptamos Ho es decir no hay diferencias entre el promedio de pulsaciones por minutos al llegar al cuerpo de guardia de hombres y mujeres.

COMPARACIÓN HORIZONTAL. PRUEBA T DE STUDENT PARA COMPARAR LAS MEDIAS DE UNA POBLACIÓN EN DOS MOMENTOS (ANTES Y DESPUÉS)

Se desea conocer si el promedio de pulsaciones por minuto de los pacientes (anexo 1) al llegar al cuerpo guardia y después del poosperatorio ha variado con un 99 % de certeza, conociendo que esta variable tiene una distribución normal.

Como se puede apreciar en este problema tenemos un conjunto de pacientes al que se le tomo el número de pulsaciones por minutos al llegar al cuerpo de guardia y después del postoperatorio y queremos comparar ambos promedios. Es decir estamos en presencia de dos momentos a una misma persona (antes y después).

Para resolver este problema aplicamos una prueba estadística llamada **T de Student** **para muestra dependiente o relacionadas,** para poder aplicar esto debemos garantizar que:

- La variable tenga una distribución normal
- Los datos sean cuantitativos
- Se conozca que sea un mismo grupos en dos momentos diferentes
-

¿Qué nos plantea esta prueba estadística?

Sean (X1, X2) pares de valores correspondientes a mediciones de una variable en un mismo sujeto antes y después de un período de tiempo o tratamiento, pero también pudieran representar dos variables distintas pero relacionadas entre sí y que el valor de una de ella tiene su contraparte en la otra. Sea

D=X1-X2 o D=X2-X1 una variable aleatoria con distribución normal con parámetros (μ_D,σ_D). Se quiere determinar a partir de una muestra aleatoria de X1 y X2 si μ_D es igual a 0, o lo que es lo mismo, entre las medias de X1 y X2 no hay diferencia.

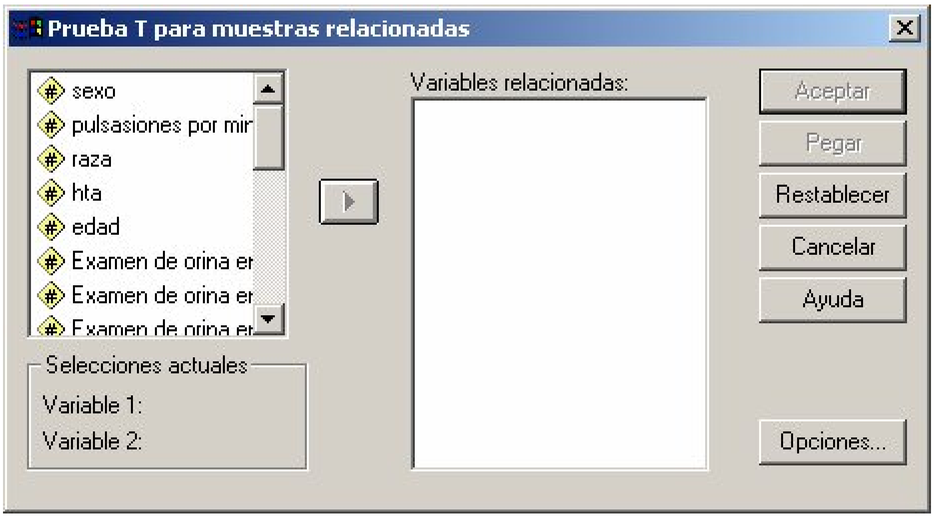
<table>
<tr><td>Paso</td><td>I</td><td>II</td><td>III</td></tr>
<tr><td>1</td><td>Ho: $\mu_D=0$
H1: $\mu_D\neq0$

Prueba Bilateral o de dos colas</td><td>Ho: $\mu_D=0$
H1: $\mu_D>0$

Prueba unilateral (derecha) o de una cola</td><td>Ho: $\mu_D=0$
H1: $\mu_D<0$
Prueba unilateral(izquierda) o de una cola</td></tr>
<tr><td>2</td><td colspan="3">α=5% (0.05) o α=1% (0.01)</td></tr>
<tr><td>3</td><td colspan="3">d rayita y S_d representan la media y la desviación standard de la variable diferencia en la muestra
Si Ho es cierta, se dice que t tiene distribución T de Student con gl igual a n-1
$$t=\frac{\bar{d}}{\frac{S_d}{\sqrt{n}}}$$</td></tr>
<tr><td rowspan="2">4</td><td colspan="3">Se rechaza Ho</td></tr>
<tr><td>Si $p < \alpha$</td><td>$p<2\alpha$
y $t>0$ o X1>X2</td><td>$p<2\alpha$
y $t<0$ o X1<X2</td></tr>
<tr><td>5</td><td colspan="3">Interpretación de los resultados.</td></tr>
</table>

Esta prueba la aplicamos utilizando el procesador estadístico SPSS (versión 11). La forma de ejecutar el procedimiento es:

Analizar ⇨ **Comparar medias** ⇨ **Prueba t de Student para muestras relacionadas.** Obtenemos los siguientes cuadros de dialogo:

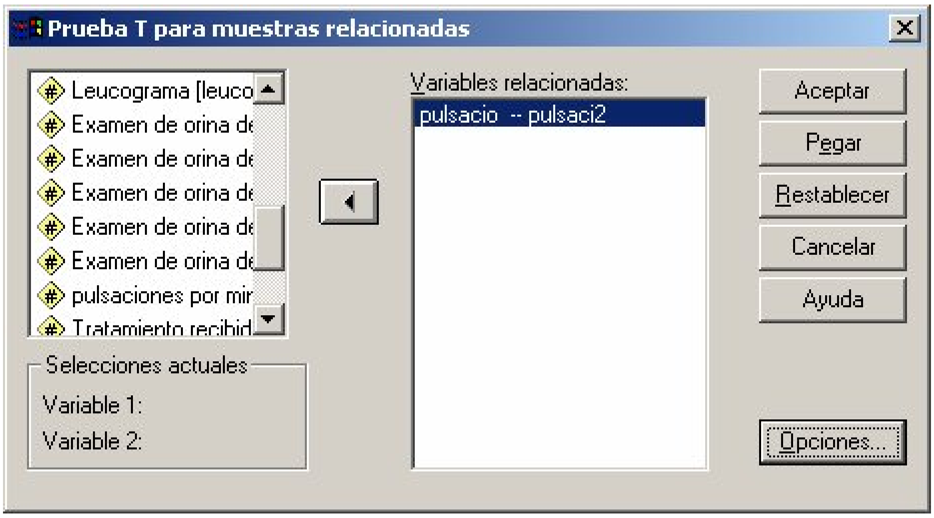


Seguidamente trasladamos las variables que estamos interesados en relacionar y al presionar el botón opciones situamos en nivel de significación solicitados.

En el caso de nuestro problema práctico planteado al inicio Usted debe:

4. Plantear las hipótesis:
   i. Ho: $\mu I = \mu F$ pulsaciones por minutos
   ii. H1: $\mu I \neq \mu F$ pulsaciones por minutos
5. Determinar el nivel de significación 99 %

Determinar los valores del estadígrafo



Y al aceptar obtenemos el cuadro de resultados siguientes:

Prueba T

Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	pulsasiones por minutos en el cuerpo de guardia	68,30	40	5,748	,909
	pulsaciones por minutos despues del posoperatorio	68,93	40	5,681	,898

Correlaciones de muestras relacionadas

		N	Correlación	Sig.
Par 1	pulsasiones por minutos en el cuerpo de guardia y pulsaciones por minutos despues del posoperatorio	40	,021	,897

Prueba de muestras relacionadas

		Diferencias relacionadas					
					99% Intervalo de confianza para la diferencia		
		Media	Desviación típ.	Error típ. de la media	Inferior	Superior	t
Par 1	pulsasiones por minutos en el cuerpo de guardia - pulsaciones por minutos despues del posoperatorio	-,63	7,996	1,264	-4,05	2,80	-,494

En este caso t= -0.494 y p = 0.624

Luego si p < α rechazamos Ho, pero como 0.624 > 0.01 aceptamos Ho es decir no hay diferencias significativas entre el promedio de pulsaciones por minutos de los pacientes al llegar al cuerpo de guardia y posterior al postoperatorio.

3.4 EJERCICIOS Y PROBLEMAS DE CAPITULO.

1. A un grupo de 10 enfermos se les suministró un antidepresivo. Mediante pruebas adecuadas, se valoró en 4 el estado inicial de los enfermos, y el estado después de haber tomado el medicamento, fue el siguiente:

3 5 4,5 7 6 6,5 4 5,5 7 7.

¿Se puede aceptar que, por término medio, mejoraron? Realícese el test con nivel de significación α =0,05, en los casos:

2. Se está realizando un estudio con una muestra aleatoria de niños que reciben tratamiento en el Servicio de Ortodoncia del Hospital Universitario "Arnaldo Milián Castro" y al medirles el Índice Incisivo Superior se obtuvo los siguientes resultados:

19,7 23,3 22,5 21,4 18,3 22,1 20,4 19,8 22,5 22,2 24,1 18,3 21,7 20,8 20,3

Considerando que la variable se distribuye normal. Se plantea que el valor promedio no difiere significativamente de 21,2 mm.

3. Se tienen el tiempo empleado en realizar dos tratamientos diferentes en estomatología:

Tratamiento 1	16	19	12	23	15	18	21	17	18	22	20	19
Tratamiento 2	18	25	19	24	23	22	23	19	26	23		

Supongamos que el tiempo se distribuye normal.

¿Es el tiempo promedio en el Tratamiento 1 significativamente menor que en el tratamiento2?

4. Se desea comparar la presencia de ácido ascórbico contenido en la saliva (mg/10000 cc) entre un grupo de 6 pacientes libres de caries (Xi) y otro grupo de 7 pacientes con presencia de caries (X2):

X1	20	31	14	19	21	17	
X2	19	9	13	20	21	15	12

Suponiendo normalidad, con α = 0,01 se pide:

¿Hay diferencias significativas entre el grupo libres de caries y el grupo con presencia de caries?

¿Se puede admitir que la presencia de ácido ascórbico contenido en la saliva en el grupo de pacientes libres de caries es superior a la del grupo de pacientes con presencia de caries?

5. En un estudio sobre cierto tratamiento aplicado a inflamación gingival se estableció el PMA de 10 pacientes elegidos aleatoriamente. Dicho índice se determinó antes y después de 8 semanas de tratamiento. Los datos son los siguientes:

Inic.	**14**	**2**	**3**	**12**	**7**	**10**	**0**	**1**	**6**	**3**
Fin.	**8**	**1**	**7**	**0**	**6**	**3**	**0**	**0**	**3**	**2**

¿Es efectivo el tratamiento? Utilice un nivel de significación de 0,05. Considere que el PMA es una variable con distribución normal

6. En un estudio sobre el tiempo en minutos empleado según la cantidad de cambios de curas en el servicio de Endodoncia de un Clínica Dental se eligieron dos períodos. Los datos son los siguientes:

Período A	**40**	**50**	**60**	**55**	**45**	**50**	**70**	**60**	**55**	**45**	**50**
Período B	**60**	**70**	**80**	**90**	**75**	**75**	**65**	**80**	**50**	**70**	**60**

Según estos datos, ¿podemos afirmar que existe diferencia en cuanto al tiempo en minutos empleado para el cambio de curas según el período en que se realizó? Suponga normalidad y utilice un nivel de significación de α = 0,01.

7. Se desea comparar la edad de un grupo de 25 personas control (con una mediana de frecuencia de cepillado igual a 2) y otro de 36 personas experimental (con una mediana de frecuencia de cepillado igual a 3). Se tomó la edad de estos dos grupos de personas según una encuesta aplicada en los dos grupos de estudio. Los datos obtenidos fueron los siguientes:

Personas de control	*n*$_1$=25	$\overline{x_1} = 49{,}8$	$s_1 = 23{,}7$
Personas experimental	***n*$_2$=36**	$\overline{x_2} = 36{,}5$	$s_2 = 31{,}7$

¿Se observan diferencias significativas entre el grupo control y el grupo experimental? Supongamos que la edad es una variable con distribución normal.

8. Se está realizando un estudio sobre el nivel de glucosa efectuado a dos grupos de hombres normales en ayunas, cuyos índices periondontales estaban comprendidos entre **1,00 a 1,80 (Grupo 1)** y de **2,18 a 2,83 (Grupo 2).**

Grupo 1 Concentración de glucosa (mg/dl)		**Grupo 2 Concentración de glucosa (mg/dl)**	
90	**71**	**88**	**73**
70	**74**	**75**	**72**
82	**79**	**88**	**82**
20	**180**	**123**	**175**
76	**83**	**74**	**82**
77	**84**	**79**	**81**
82	**79**	**81**	**78**
75	**77**	**72**	**79**
73		**72**	**77**
77		**75**	**79**

¿Existen diferencias significativas entre ambos grupos de hombres en cuanto al nivel de glucosa? Utilice un nivel de significación de 0,05 y considere que la variable nivel de glucosa posee distribución normal.

9.- Al procesar una encuesta se obtuvieron los siguientes datos:

Pacientes	Edad	Sexo	Estado bucal	Número de dientes	Longitud de corona	Reducción fuerza masticatoria
1	23	F	Regular	3	6,73	Si
2	47	F	Bueno	5	8,37	Si
3	35	M	Bueno	4	7,16	No
4	52	F	Malo	2	5,92	Si
5	44	M	Regular	3	6,26	No
6	38	M	Regular	3	7,47	Si
7	32	F	Bueno	4	6,28	Si
8	26	M	Regular	3	6.63	Si
9	56	F	Malo	1	7,38	Si
10	51	F	Malo	2	5,67	No
11	31	M	Regular	3	6,78	Si
12	48	M	Bueno	4	7,53	Si
13	36	F	Regular	3	7,27	Si
14	31	M	Malo	2	6,92	Si
15	49	M	Regular	3	7,83	Si
16	27	M	Regular	3	7,92	Si
17	34	F	Bueno	4	7,28	No
18	53	M	Regular	3	8,52	Si
19	47	F	Malo	2	7,37	Si
20	39	F	Malo	1	8,47	Si

a) ¿Es mayor la longitud promedio de la corona en los hombres que entre las mujeres?

b) ¿Existen diferencias significativas en la edad promedio entre hombres y mujeres?

c) ¿Se puede afirmar que la longitud promedio de los pacientes es superior a 6,20?

d) ¿Es cierto que la edad promedio de los que si tienen reducción en la fuerza masticatoria es inferior a 40?

10.- Para comprobar si la tolerancia a la glucosa en sujetos sanos tiende a decrecer con la edad se realizó un test oral de glucosa a dos muestras de pacientes sanos, unos jóvenes y otros adultos. El test consistió en medir el nivel de glucosa en sangre en el momento de la ingestión (nivel basal) de 100 grs. de glucosa y a los 60 minutos de la toma. Los resultados fueron los siguientes:

Jóvenes:

Basal	81	89	80	75	74	97	76	89	83	77
60 minutos	136	150	149	141	138	154	141	155	145	147

Adultos:

Basal	98	94	93	88	79	90	86	89	81	90
60 minutos	196	190	191	189	159	185	182	190	170	197

a) ¿Es mayor la concentración de glucosa en sangre a los 60 minutos, en adultos que en jóvenes?

b) El contenido basal de glucosa en sangre, ¿es menor en jóvenes que en adultos?

CAPITULO 4. PRUEBAS NO PARAMÉTRICAS

OBJETIVOS:

1 Identificar cuando deberá usarse el método de prueba de hipótesis para Chi – cuadrado en problemas biomédicos.

2. Aplicar el método de prueba de hipótesis Chi- cuadrado en la toma de decisiones en problemas biomédicos. Así como explicar el significado del resultado de la prueba en relación con el problema objeto de análisis.

3. Usar un procesador de datos estadísticos profesionales con vista a servirse del mismo para resumir un conjunto de datos biomédicos en forma automatizada y así obtener información necesaria para poder aplicar e interpretar el método de prueba de hipótesis.

En este capítulo estudiaremos las pruebas no paramétricas. Para realizar un análisis no paramétrico deben partirse de las siguientes consideraciones:

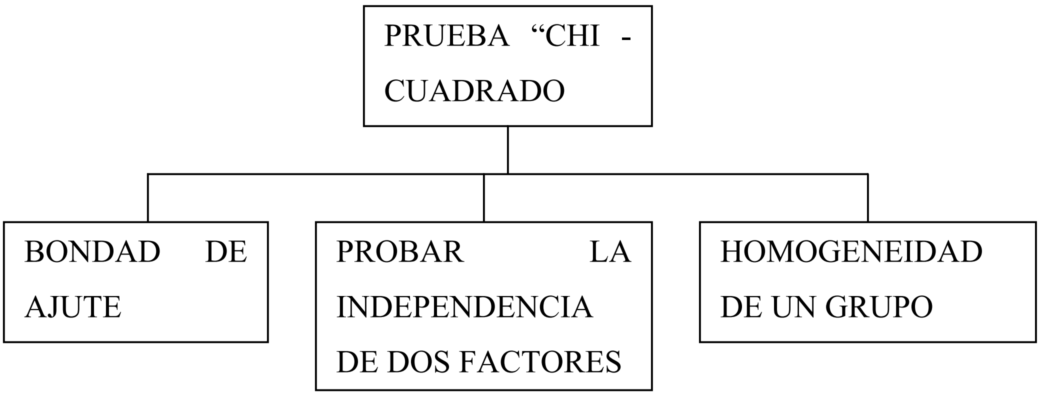
- Este análisis acepta que la distribución de la(s) variable(s) aleatoria(s) no tenga una distribución normal.
- Las variables pueden ser cualitativas nominales u ordinales.

Existen múltiples y variadas pruebas no paramétricas, las más utilizadas en las Ciencias Médicas son:

- La prueba Chi – cuadrado
- Los coeficientes de correlación
- U de Man – Whitney
- Z de kolmogorov – Smirtnov
- H de Kruskal – Wallis

En este capítulo estudiaremos la prueba "Ji cuadrado"

La prueba estadística "Chi – cuadrado" la podemos utilizar para:



PROBLEMA PRACTICA 1.

Se quiere determinar si el Municipio de Ranchuelo (Anexo 2) el por ciento de fumadores es mayor que el 30 %. A qué conclusión Ud. arribaría considerando un nivel de significación del 5 %.

A partir de los datos debemos pensar que si el 30 % fuma y se encuestan 3000 personas, debemos esperar tener 900 fumadores.

¿Cómo proceder en este caso?

Para resolver el problema debemos aplicar la prueba "Chi – cuadrado" para la bondad de ajuste, esta nos permite comparar las proporciones, por cientos o frecuencias observadas de cada valor de X con las proporciones, por cientos o frecuencias esperadas o teóricas correspondientes, partiendo del supuesto de que se cumple una hipótesis determinada.

PASOS	
1	**Ho : P1= E1, P2=E2.... Pk = Ek** **H1: Existe al menos un i tal Pi ≠ Ei**
2	**Nivel de significación**
3	$\chi^2 = n \sum_{i=1}^{k} \frac{(Pi - Ei)^2}{Ei -}$
4	**Se rechaza Ho si p < α**
5	**Interpretación**

1) Ho: P = 30% (P = Po) (Prueba de hipótesis para una proporción)
   H1: P > 30% (P > Po) Prueba de Hipótesis unilateral.

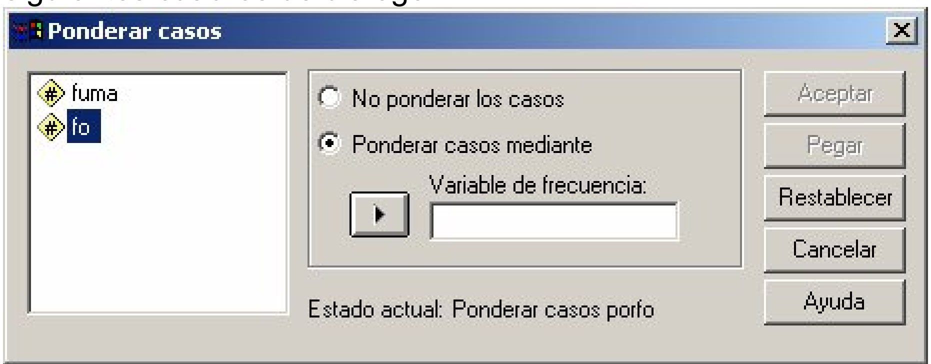
(Por ciento de fumadores en el municipio)

2) α = 0.05
3) Calcular el estadígrafo de la prueba:

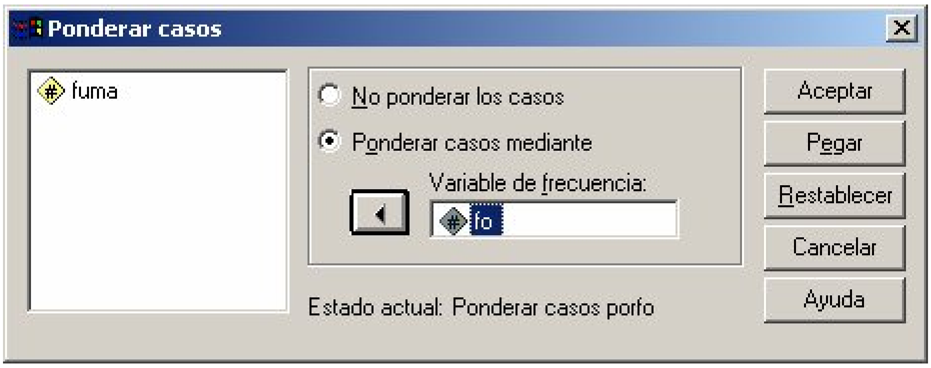
Fichero de datos:

fuma	fo
0	2000
1	1000

Datos -> Ponderar casos (variable de frecuencia: fo) obteniéndose los siguientes cuadros de dialogo:



Seguidamente señalamos poderar casos mediante y trasladamos la variable fo, obteniendo:



Presionamos aceptar.
Posteriormente procedemos a:
Analizar ⇨ Pruebas no paramétricas ⇨ Chi-cuadrado

Obteniendo el siguiente cuadro de diálogo:

Trasladamos para contrastar la variable fuma y en valores esperados 2100, añadir y 900, añadir y por ultimo aceptar:

Seguidamente obtenemos en la ventana de resultados:

FUMA

	N observado	N esperado	Residual
0	2000	2100,0	-100,0
1	1000	900,0	100,0
Total	3000		

Estadísticos de contraste

	FUMA
Chi-cuadrado[a]	15,873
gl	1
Sig. asintót.	,000

a. 0 casillas (,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 900,0.

4) Se rechaza Ho, cuando $p_{\chi^2} < 2\alpha$ y p>Po (por ciento muestral mayor que valor hipotético)

0.000< 0.10 y 33.3%>30% por lo tanto se rechaza Ho.

5) El por ciento de fumadores en el municipio de Ranchuelo, es significativamente mayor que 30.

En el problema anterior los datos nos lo dan a partir de las frecuencias absolutas, pero no siempre ocurre así, podemos tener la base de datos de los pacientes (Anexo 3), para lo cual podemos plantearnos el siguiente problema:

PROBLEMA PRACTICO:

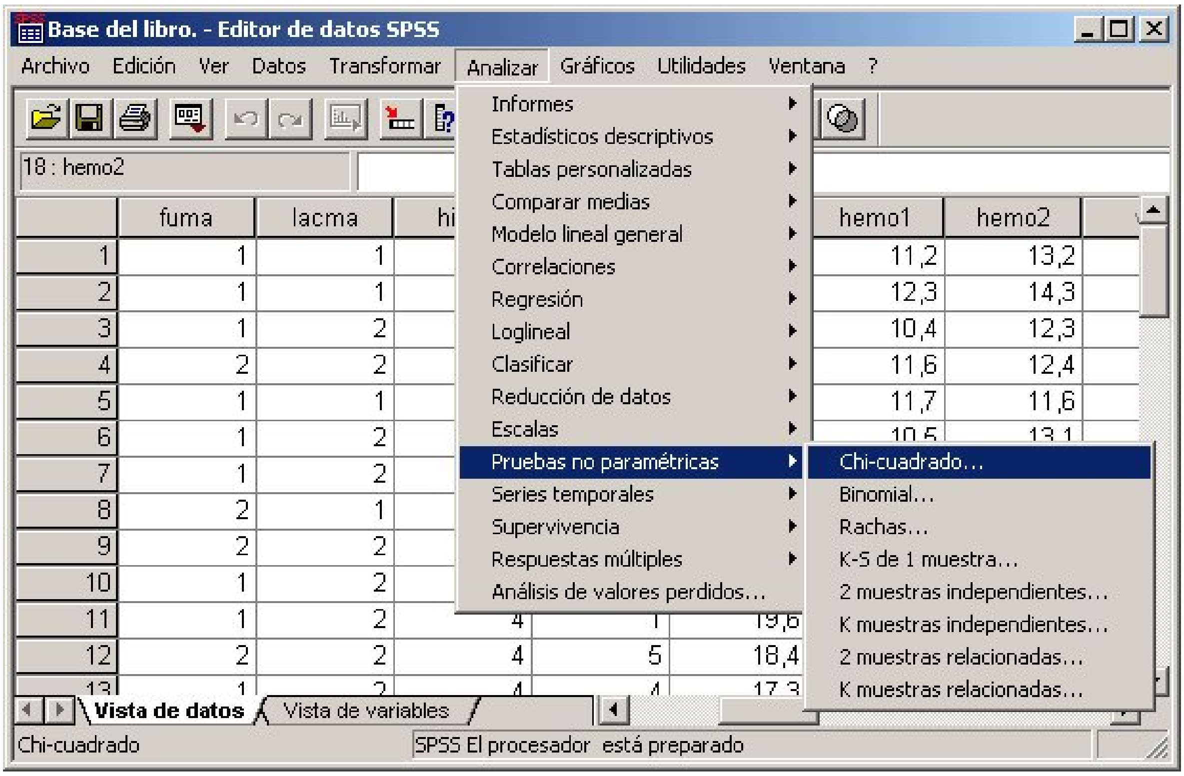
Determinar si se puede afirmar que la proporción de niños que conocen las consecuencias del hábito de fumar es igual al 70%.**(variable cualitativa en escala nominal)**

1) Ho: P= 70% (P=Po) (Prueba de hipótesis para una proporción)
   H1: P ≠ 70% (P≠ Po) (Prueba de Hipótesis bilateral).

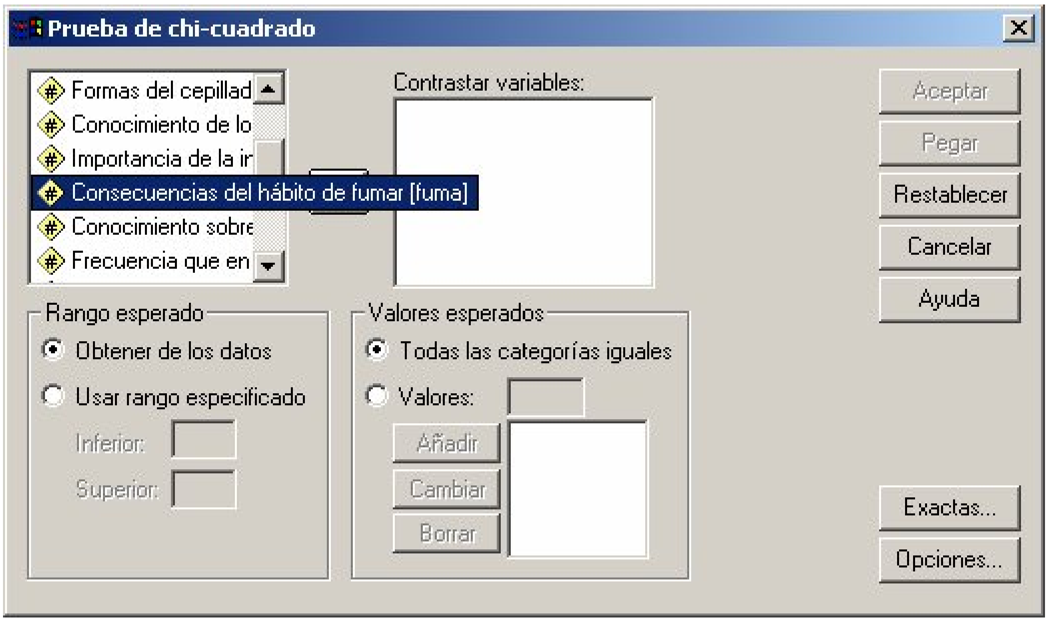
2) α = 0.05

3) Valor del estadígrafo:

¿Cómo debemos proceder para utilizar la prueba chi-cuadrado(χ^2) a través del S.P.S.S.?
Para ello debemos en el menú Analizar → Pruebas no paramétricas → Chi-cuadrado. obteniendo:



Al hacer clic en Chi-cuadrado le aparecerá el siguiente cuadro de diálogo:



La variable Consecuencias del hábito de fumar (fuma) usted la debe pasar para la derecha en Contrastar variables y en Valores esperados, si las categorías de la variables poseen el mismo valor esperado lo deja marcado (como aparece señalado por defecto), en el caso de que sea diferente, como es en nuestro ejemplo, se marca en Valores y escribe 70 es el valor del porcentaje de niños que se supone si tiene conocimientos sobre el hábito de fumar y además en la declaración de las variables en el fichero (Base del libro) aparece en esta variable como el número 1 la categoría "Si" es decir que si tiene conocimientos de las consecuencias del hábito de fumar, luego haga clic en Añadir, vuelve a Valores y escribe 30 (que es el resto de 100 – 70)(Son los que "No", que aparecen con el número 2 la categoría "No", tienen conocimientos de las consecuencias del hábito de fumar) y luego haga clic en Añadir (Este proceso se repite según sea la cantidad de categorías que posea la variable y teniendo en cuenta que la suma sea igual a100, por último hacer clic en Aceptar y le aparecerá en la ventana de resultados:

Consecuencias del hábito de fumar

	N observado	N esperado	Residual
Si	11	11,9	-,9
No	6	5,1	,9
Total	17		

Estadísticos de contraste

	Consecuencias del hábito de fumar
Chi-cuadrado	,227
gl	1
Sig. asintót.	,634

a 0 casillas (,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 5,1.

4) Se rechaza Ho, cuando $p_{\chi^2} < 2\alpha$ y p>Po (por ciento muestral mayor que valor hipotético)

0.634> 0.10 por lo tanto no se rechaza Ho.

5) El por ciento de niños que conocen las consecuencias del hábitos de fumar es del 70 %.

5.2. Prueba de hipótesis Chi-cuadrado para la independencia de dos factores.

Con mucha frecuencia el profesional de la salud le interesa saber si dos variables cualitativas (factores) están relacionadas o no. Con este fin se extrae una muestra aleatoria de la población y se *clasifican a posteriori* los sujetos en base a estas dos variables.

Problema Práctico:

En el estudio realizado en problema práctico anterior (Anexo 2) de los 1000 fumadores, 860 tenían algún grado de afección estomatológica y de los 2000 no fumadores 948 tienen afectaciones estomatológicas. Podríamos afirmar que en esa población en estudio hay dependencia o no entre el hábito de fumar y las afecciones estomatológicas.

Para resolver este problema debemos aplicar la prueba Chi-cuadrado para la independencia o dependencia de dos factores.

PASOS	
1	**Ho : P1= E1, P2=E2.... Pk = Ek** **H1: Existe al menos un i tal Pi $\neq$ Ei**
2	**Nivel de significación**
3	$\chi^2 = \sum_{i=1}^{k} \sum_{e=1}^{k} \frac{(Oij - Eij)}{Eij}$ **Oij –Número observado** **Eij – Número esperado**
4	**Se rechaza Ho si p < α**
5	**Interpretación**

Comenzamos para resolver el problema confeccionando una tabla de Contingencia.

Una tabla de contingencia es una tabla de doble entrada o dos clasificaciones como esta en que A y B son dos factores de clasificación de una muestra aleatoria en n niveles según el factor de clasificación B y m niveles según el factor de clasificación A. Es decir:

A/B	B1	B2	…….	Bn	total
A1					
A2					
. .					
An					
Total					

Aspectos a tener en cuenta para la aplicación de esta prueba:

•No puede haber ninguna frecuencia esperada menor que 1. Si esto sucede, y la tabla no es 2x2 se pueden unir filas o columnas. Si es una tabla 2x2 se puede incrementar el tamaño de la muestra o se aplica el Test Exacto de Fisher.

•En una tabla de 2x2 se dan dos valores del estadígrafo Chi-cuadrado con sus significaciones p correspondientes:

Chi-cuadrado de Pearson: Se toma cuando el tamaño de la muestra es mayor que 25

Chi-cuadrado con el Factor de Corrección: Se selecciona cuando el tamaño de la muestra es menor o igual que 25.

En nuestro caso:

	AFECCIONES ESTOMATOLOGICAS		TOTAL
FUMADORES	SI	NO	
SI	860	140	1000
NO	948	1052	2000
TOTAL	1808	1192	3000

A partir de lo anterior se quiere determinar si existe relación significativa entre el hábito de fumar y las afecciones estomatológicas.

Ho: El hábito de fumar y las afecciones estomatológica son independientes

H1: El hábito de fumar y las afecciones estomatológica son dependientes

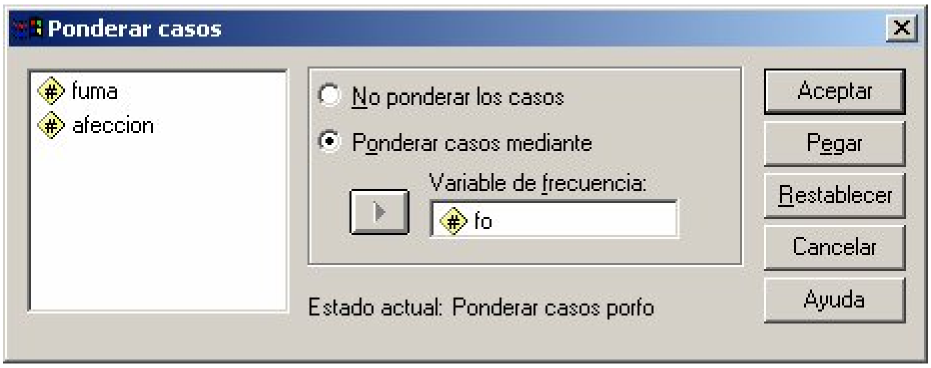
$\alpha = 0,05$

Procederíamos a partir del SPSS de la siguiente forma.

1. Fichero de datos:

FUMA	AFECIONES ESTOMATOLÓGICAS	Fo
1	1	860
1	2	140
2	1	948
2	2	1052

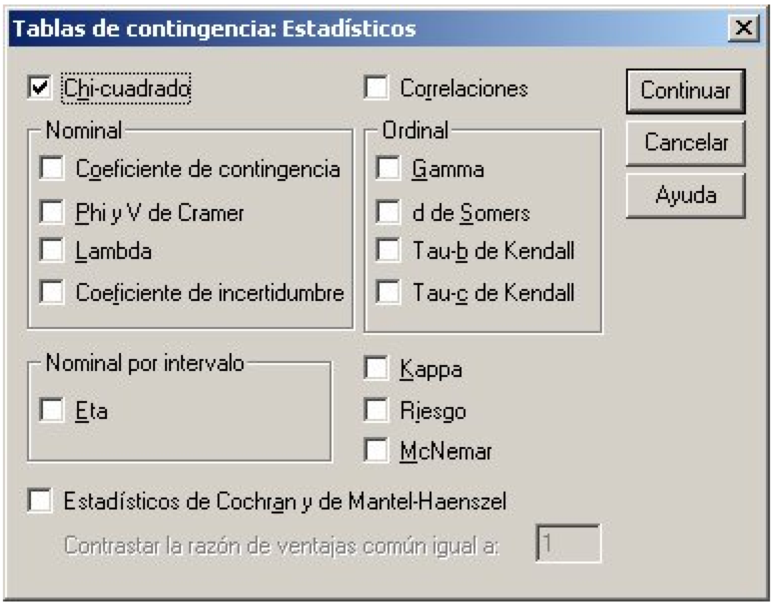
Después de confeccionada la base de datos vamos a Datos, ponderar casos y en variable de frecuencia fo obteniendo:



Seguidamente vamos al menú analizar, Estadístico descriptivos, tabla de contingencia, donde trasladamos a la fila fumadores y a la columna Afecciones estomatológicas, obteniendo:

Seguidamente presionamos el botón casillas para seleccionar frecuencias observadas y esperadas y posteriormente continuar.

Posteriormente el botón estadístico para seleccionar el estadígrafo Chi-cuadrado y posteriormente continuar.



Para posteriormente aceptar, obteniendo los resultados siguientes:

Tablas de contingencia

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
FUMA * AFECCION	3000	100,0%	0	,0%	3000	100,0%

Tabla de contingencia FUMA * AFECCION

			AFECCION		Total
			1	2	
FUMA	1	Recuento	860	140	1000
		Frecuencia esperada	602,7	397,3	1000,0
	2	Recuento	948	1052	2000
		Frecuencia esperada	1205,3	794,7	2000,0
Total		Recuento	1808	1192	3000
		Frecuencia esperada	1808,0	1192,0	3000,0

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	414,812[b]	1	,000		
Corrección por continuidad[a]	413,201	1	,000		
Razón de verosimilitud	454,388	1	,000		
Estadístico exacto de Fisher				,000	,000
Asociación lineal por lineal	414,674	1	,000		
N de casos válidos	3000				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 397,33.

Para tomar el valor de Chi-cuadrado, es necesario fijarse en las frecuencias esperadas y en el tamaño de la muestra.

Chi- cuadrado = 414.8 gl= 1 $px^2 = 0.000$

Como $px^2 < 0.05$, se rechaza Ho

Luego si existe relación significativa entre el hábito de fumar y las afecciones estomatológicas observadas en el municipio de Ranchuelo.

En el problema anterior los datos nos lo dan a partir de las frecuencias absolutas, pero no siempre ocurre así, podemos tener la base de datos de los pacientes (Anexo 3), para lo cual nos podemos plantear el siguiente problema:

PROBLEMA PRÁCTICO

Determinar si existe relación significativa entre el sexo de los niños y el conocimiento de las Consecuencias del hábito de fumarse(Anexo 3) **(Ambas variables cualitativas en escala nominal)**

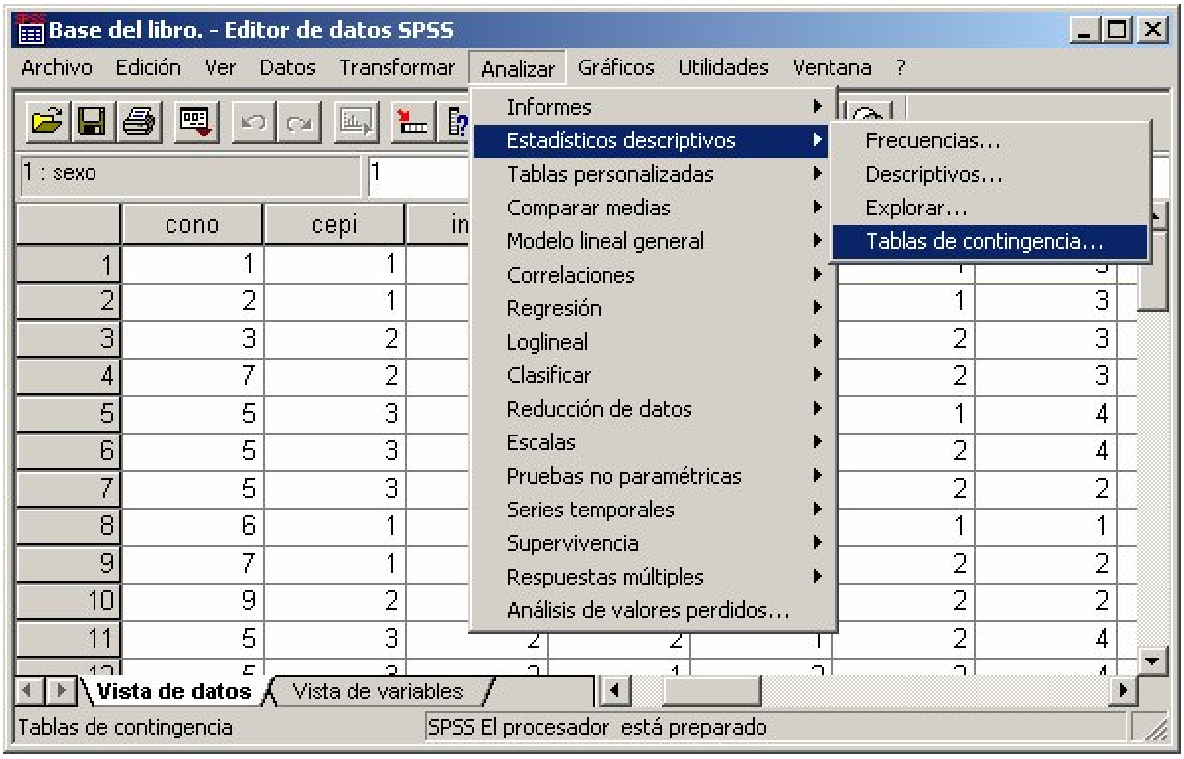
Solución.

Ho. El sexo de los niños y el conocimiento son independientes

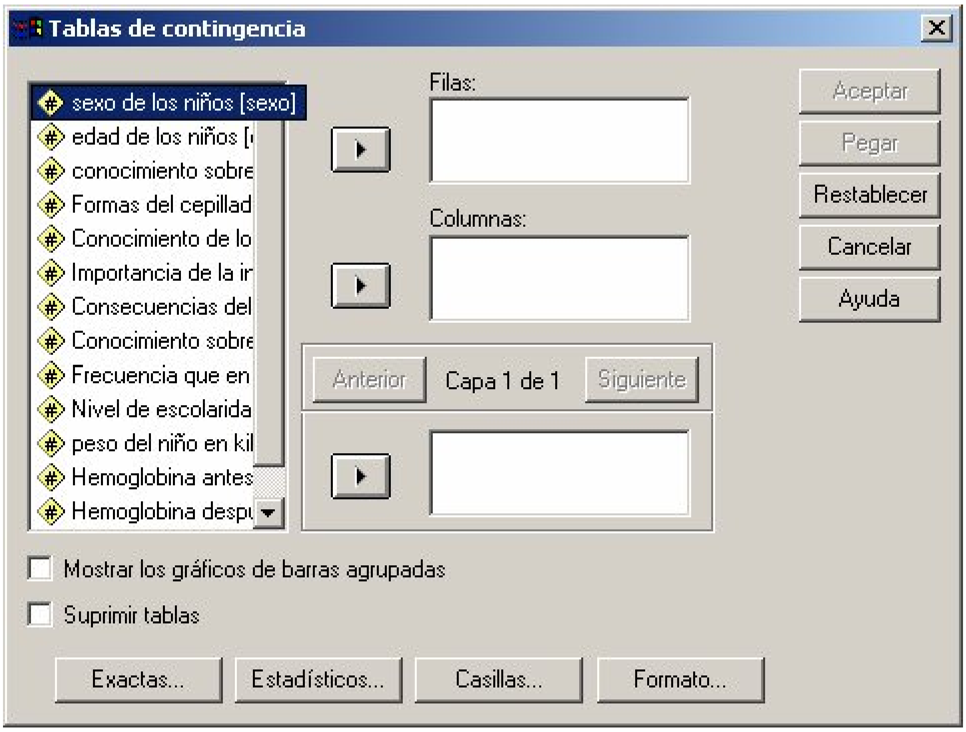
H1. El sexo de los niños y el conocimiento son dependientes

¿Cómo debemos proceder para utilizar la prueba chi-cuadrado(χ^2) a través del S.P.S.S.?

Para ello debe:

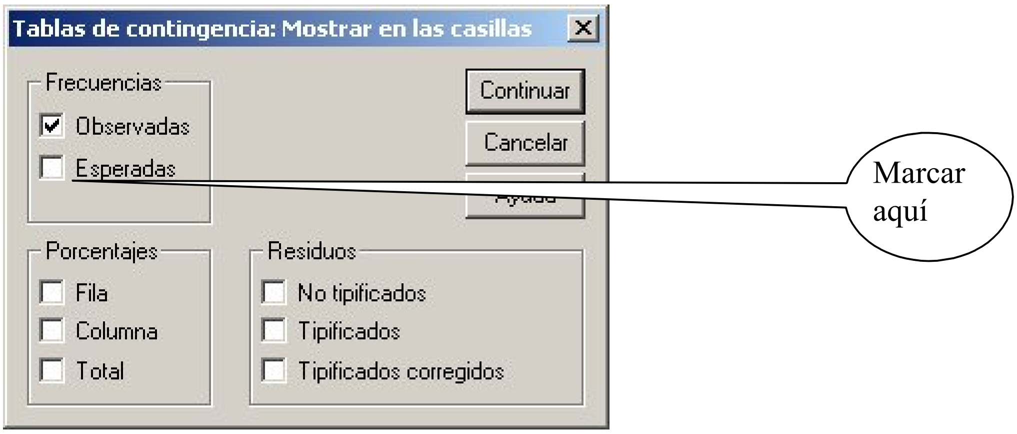


Es decir del menú Analizar → Estadísticos descriptivos → Tablas de contingencia. Al hacer clic en Tablas de contingencia le aparecerá el siguiente cuadro de diálogo:



La variable sexo usted la debe pasar para la derecha en Filas y la variable fuma para Columnas (es decir cada variable en una casilla diferente). En nuestro ejemplo se puede pasar al revés, todo depende de que para la Columnas se deba pasar siempre la que tenga más categorías. Cuando haga esto debe hacer clic en Estadísticos y le aparecerá el siguiente cuadro de dialogo:

Aquí debe marcar en Chi-cuadrado y luego haga clic en Continuar para volver al cuadro de diálogo anterior y aquí luego debe hacer clic en Casillas y le aparecerá el siguiente cuadro de dialogo:



Aquí debe marcar en Esperadas (esto es para que aparezca en la ventana de resultados los valores esperados) y luego haga clic en Continuar para volver al cuadro anterior hacer clic en Aceptar y le aparecerá en la ventana de resultados la siguiente imagen:

Tabla de contingencia sexo de los niños * Consecuencias del hábito de fumar

			Consecuencias del hábito de fumar		Total
			Si	No	
sexo de los niños	Masculino	Recuento	6	3	9
		Frecuencia esperada	5,8	3,2	9,0
	Femenino	Recuento	5	3	8
		Frecuencia esperada	5,2	2,8	8,0
Total		Recuento	11	6	17
		Frecuencia esperada	11,0	6,0	17,0

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	,032	1	,858		
Corrección por continuidad	,000	1	1,000		
Razón de verosimilitud	,032	1	,858		
Estadístico exacto de Fisher				1,000	,627
Asociación lineal por lineal	,030	1	,862		
N de casos válidos	17				

a Calculado sólo para una tabla de 2x2.

b 2 casillas (50,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,82.

P=1.0 Si p >α no se rechaza ho por tanto el sexo y el conocimientos son independientes.

4.3 Prueba Chi- cuadrado para verificar homogeneidad de grupos respecto a la distribución de una variable cualitativa

Se quieren comparar frecuencia, por cientos o proporciones de una de las categorías de la variable X en cada uno de los grupos (dados por las categorías de una variable Y). Se *fija de antemano el número de sujetos* a seleccionar en cada grupo, y después de seleccionados se *clasifican* de acuerdo a las categorías de la variable X

Problema práctico,

En el estudio (Anexo 2) se quiso además valorar la influencia del hábito de fumar sobre las afecciones estomatológicas. Con este propósito se seleccionaron tres grupos dado el nivel de severidad de la afección (maligno, pre maligno y común). Los investigadores están interesados en verificar si estos tres grupos son homogéneos respecto a la proporción o por ciento de portadores del hábito toxico, obteniéndose la siguiente tabla de contingencia:

Fumadores	**Tipo de lesión**	**Total**

	Maligna	Premaligna	Común	
Si	70	96	694	860
No	32	46	870	948
Total	102	142	1564	1808

A partir de lo anterior se quiere determinar si los grupos son homogéneos respecto a la proporción de afecciones estomatológicas.

Ho: P1 = P2 = P3

H1: Pi ≠ Pj

α = 0,05

Procederíamos a partir del SPSS de la siguiente forma.

Fichero de datos:

FUMA	AFECIONES ESTOMATOLÓGICAS	Fo
1	1	70
1	2	96
1	3	694
2	1	32
2	2	46
2	3	870

A partir de lo anterior se quiere determinar si existe relación significativa entre el hábito de fumar y las afecciones estomatológicas.

α = 0,05

Procederíamos a partir del SPSS de forma similar al ejemplo anterior, obteniéndose los siguientes resultados:

Tablas de contingencia

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
FUMA * TIPO	1808	100,0%	0	,0%	1808	100,0%

Tabla de contingencia FUMA * TIPO

			TIPO			Total
			Maligna	Pre maligna	Común	
FUMA	si	Recuento	70	96	694	860
		Frecuencia esperada	48,5	67,5	743,9	860,0
	no	Recuento	32	46	870	948
		Frecuencia esperada	53,5	74,5	820,1	948,0
Total		Recuento	102	142	1564	1808
		Frecuencia esperada	102,0	142,0	1564,0	1808,0

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	47,397[a]	2	,000
Razón de verosimilitud	48,056	2	,000
Asociación lineal por lineal	42,251	1	,000
N de casos válidos	1808		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 48,52.

Obteniéndose que:

χ^2 =47.397 gl= 2 pχ^2 = 0.0

Como pχ^2 es menor que alfa, se rechaza Ho

Los grupos dados por el estado de la afección estomatológica no son homogéneos respecto a la proporción de portadores, dicho de otra forma, la proporciones o por cientos de portadores de cada grupo se diferencian significativamente entre si .

En el problema anterior los datos nos lo dan a partir de las frecuencias absolutas, pero no siempre ocurre así, podemos tener la base de datos de los pacientes (Anexo 3), para lo cual nos podemos plantear el siguiente problema:
Problema práctico:

Determinar si los varones y las hembras son homogéneos respecto a las proporciones o por cientos de los que tienen o no cocimientos de las Consecuencias del hábito de fumar **(Sexo variable cualitativa en escala nominal)**

¿Cómo debemos proceder para utilizar la prueba chi-cuadrado(χ^2) a través del S.P.S.S.?

Para ello debe proceder igual que en el caso anterior (para probar la independencia o no de dos factores), obteniendo el siguiente resultado:

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	,032	1	,858		
Corrección por continuidad	,000	1	1,000		
Razón de verosimilitud	,032	1	,858		
Estadístico exacto de Fisher				1,000	,627
Asociación lineal por lineal	,030	1	,862		
N de casos válidos	17				

a Calculado sólo para una tabla de 2x2.

b 2 casillas (50,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,82.

4.4 EJERCICIOS Y PROBLEMAS DEL CAPITULO

1. El resultado de un experimento para investigar el efecto de la aplicación de dos tipos de anestesia en un grupo de pacientes después de realizada una intervención quirúrgica, en cuanto a los casos que resultaron positivos después de aplicada la anestesia y los casos que resultaron negativos, se muestra en la tabla siguiente:

Tipo de Anestesia	**Resultados**		**Total**
	Positivo	**Negativo**	
Anestesia A	42	12	54
Anestesia B	38	18	56
Total	80	30	110

¿Se puede afirmar que no existen diferencias significativas en cuanto a los resultados obtenidos según el tipo de anestesia aplicado para un α= 0,05?

2 .Los resultados en frecuencia del estado bucal de dos grupos de mujeres embarazadas A y B:

	Mal	**Regular**	**Bien**
Grupo A	2	10	23
Grupo B	8	12	10

¿Hay diferencias significativas entre los dos grupo de mujeres en cuanto al estado bucal? Tome α = 0,01.

3. Se quiere relacionar la formación de cálculos según el hábito de fumar y para ello se tomaron tres grupos de investigación, el primero (**Grupo A**) de pacientes no fumadores, el segundo (**Grupo B**)de pacientes que fuman menos de 10 cigarros en el día y el tercero (**Grupo C**) que fuman más de 10 cigarros en el día. A tal fin, se analiza a cada persona si posee cálculo o no, y en el, caso de que posea cálculo si es supragingival o subgingival. En una muestra de 610 personas se obtuvo:

Formación	**Grupo A**	**Grupo B**	**Grupo C**
No Cálculos	21	56	123
Cálculos supragingivales	15	75	250
Cálculos subgingivales	4	18	48

¿Se puede aceptar que hay relación entre la formación de cálculos y el hábito de fumar, si $\alpha = 0,05$?

4. Se quiere conocer si existe relación entre el grado de infección microbiana y grupo de estudio (Control o Caso). Los datos se muestran en la siguiente tabla.

Grupo de Estudio	**Grado de Infección**	
	Bajo	**Alto**
Control	41	19
Caso	32	28

5. En la tabla que se muestra a continuación aparecen los resultados de un estudio sobre la estomatitis aftosa recurrente, aftosis oral o aftas, según el grupo de edad al que pertenece. Se quiere determinar si existe relación entre la edad y la raza de los pacientes afectados.

Edad (años)	**Raza**	
	Blanca	**Negra**
2 a 5	**10**	**2**
6 a 14	**30**	**3**
Más de 15	**5**	**1**

6. En un estudio del efecto de instrucciones para la higiene dental en niños con carie dental se seleccionaron aleatoriamente 100 niños; 50 recibieron instrucciones y 50 no. Después de un periodo de 6 meses el número de nuevas caries por cada niño fueron recogidos en la siguiente tabla:

Instrucción	Número de nuevas caries		
	0 - 1	2 - 3	4 - 5
Si	30	15	5
No	20	15	15

Usando un nivel de significación del 5% ¿se puede decir que existe una asociación entre la instrucción en higiene dental y el número de nuevas caries al final de los 6 mases?.

7. En una encuesta de 200 personas sobre el cepillado de los dientes, 33 de ellos dijeron que se cepillaban los dientes 4 veces en el día. ¿Puede considerarse a partir de estos datos que en la población muestreada, la proporción de quienes se cepillan los dientes 4 veces en el día es significativamente mayor que 0.2?

8. Antes del inicio de un programa educativo sobre la higiene bucal en una ciudad una encuesta reveló que 50 integrantes de una muestra de 200 niños de primaria, no poseían una buena higiene bucal . ¿Son estos datos, compatibles

con el punto de vista de que el 50% de los niños de primaria de dicha ciudad, no poseían una buena higiene bucal? Sea $\alpha = 0.05$

9. Se tomaron 215 pacientes y se les clasificó según la ubicación geográfica (1. Urbana, 2. Rural y 3. Plan Turquino) y según la frecuencia del cepillado de los dientes en orden creciente (Obtenido mediante una encuesta). Los datos son los de la tabla siguiente:

	Ubicación		
Frecuencia	**1**	**2**	**3**
1	20	15	5
2	60	20	10
3	45	15	15
4 o más	10	5	5

¿Existe relación entre la ubicación geográfica y la frecuencia del cepillado?

10.- En un estudio diseñado para comparar un nuevo tratamiento estomatológico, 78 de los 100 individuos que recibieron el tratamiento estándar respondieron favorablemente. De los 100 individuos que recibieron el nuevo tratamiento, 90 de ellos respondieron satisfactoriamente. ¿Proporcionan estos datos evidencia suficiente que indique que el nuevo tratamiento es más efectivo que el estándar?

11. Se quiere conocer si existe o no relación entre padecer o no una enfermedad estomatológica y el hábito de fumar. Se observo que de 60 pacientes seleccionados aleatoriamente el 30% eran enfermos y de ellos 17 fumaban. De los sanos 7 no fumaban.

12. La tabla siguiente muestra la distribución de 60 pacientes seleccionados al azar para determinar si presentar o no una enfermedad bronquial esta relacionada con el hábito de fumar o no.

Fuma	Enfermo	Casos
SI	Si	17
Si	Si	3
Si	No	9
Si	No	1
No	Si	1
No	Si	9
No	No	3
No	No	17

13. Para determinar la preferencia por un nuevo analgésico para aliviar el dolor dental, se estudiaron 100 pacientes que sufrían de dolores dentales. Se le preguntó a cada paciente si preferían o no el nuevo analgésico antes del que tradicionalmente usaban. De los 100 pacientes 60 preferían el nuevo analgésico. Basados en estas respuestas ¿se puede decir que el nuevo tratamiento se prefiere antes del tradicional?

CAPITULO 5. CORRELACIÓN Y REGRESIÓN LINEAL

OBJETIVOS:

1. Explicar el significado de un diagrama de dispersión con vistas a valorar la posibilidad de establecer una relación causal de tipo lineal entre las variables que originan el diagrama.
2. Explicar el significado del coeficiente de correlación muestral y su relación con el problema fundamental de la regresión lineal.
3. Usar un procesador de datos estadísticos profesional con vistas a servirse del mismo para resumir un conjunto de datos biomédicos en forma automatizada y así obtener la información necesaria para poder aplicar e interpretar el objeto de estudio.

En los capítulos anteriores hemos realizado el análisis de los datos univariados, es decir el comportamiento de una variable cuantitativa respecto a la(s) muestra(s) de una población en estudio. Pero no hemos estudiado la asociación entre las variables.
Al estudiar la asociación entre las variables debemos valorar dos aspectos distintos pero relacionados que son:

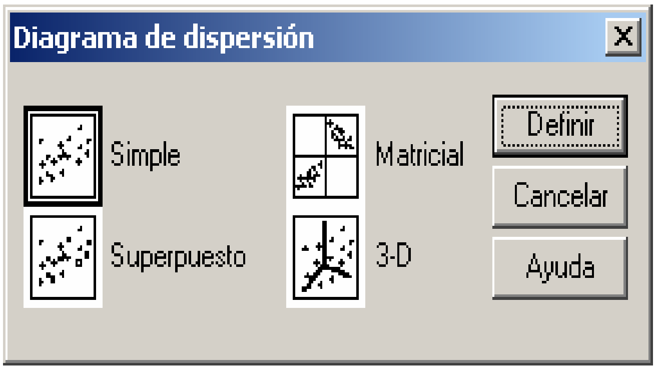
- Análisis de correlación lineal
- Análisis de regresión lineal

5.1 DIAGRAMA DE DISPERSIÓN.

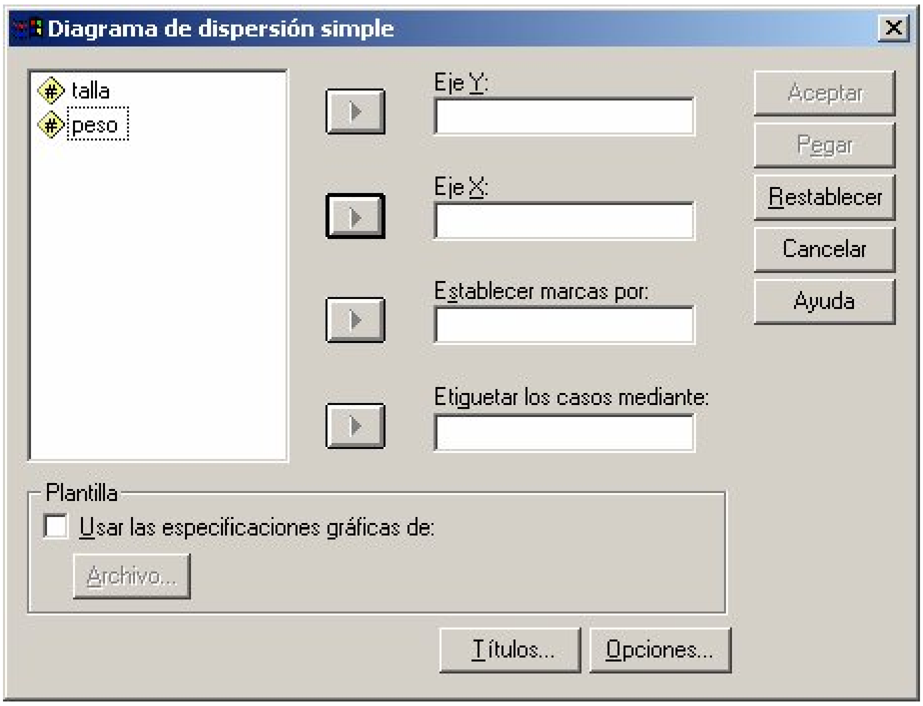
En muchos problemas de estudio se parte de consideraciones de fenómenos en que cada observación da lugar a un par de medidas ejemplo: la talla y el peso de una persona; la nota en Fisiología y Bioquímica, Edad y presión sistólica sanguínea. A este tipo de distribución se les llama distribución bidimensional.
Ejemplo práctico:
Un ejemplo de este tipo de distribución sería la tabla siguiente donde se relaciona la talla X (en centímetros) y el peso Y (en kilogramos) de 15 niños de nueve años:

Niño	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Talla	131	133	140	136	134	136	127	130	147	134	131	142	129	138	139
Peso	32	29	29	33	28	30	32	23,5	36.7	27	28	34	32	35	31.5

En este tipo de investigación el investigador trata de descubrir la relación entre dos variables (X e Y) con el objetivo de descubrir la forma aproximada de la relación, para ello representa los datos en un sistema de coordenadas, este tipo de gráfico recibe el nombre de **diagrama de dispersión.**
Para realizarlo utilizando el procesador estadístico SPSS versión 11 precederíamos de la siguiente forma (después de introducir los datos):
1ro. Seleccionamos en el menú gráfico el gráfico de dispersión y aparece una ventana como la que se muestra en la figura y elegimos el gráfico simple:



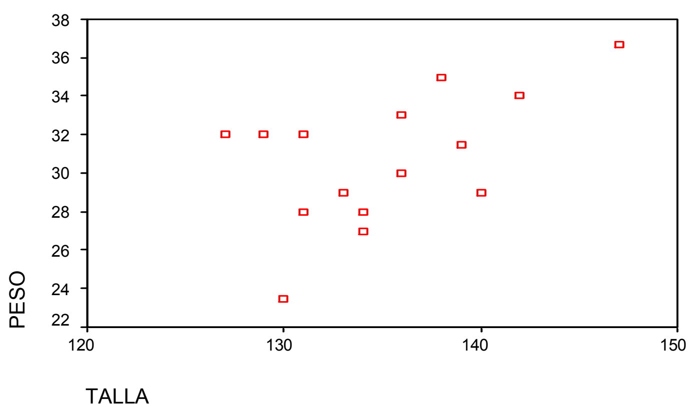
2do. Se pulsa el botón definir, y aparece una ventana donde trasladamos al eje Y la variable peso y al eje X la variable talla



Obteniéndose como resultado:

Relación entre la talla y el peso de

de niños de nueve años estudiados



Datos del problema

A observar el gráfico de dispersión podemos apreciar que:

- La forma alargada de la **nube de puntos** siguiendo una dirección creciente es decir a mayor talla en los niños, mayor peso
- Esta **nube de puntos se condesa** alrededor de una línea imaginaria

5.1 COEFICIENTE DE CORRELACIÓN LINEAL DE PERSON.

Existe una medida cuyo objetivo es reflejar a través de su valor el grado de asociación lineal existente entre las variables cuantitativas X y Y. Este valor se conoce con el nombre de coeficiente de correlación lineal.

ρ -> Coeficiente de correlación lineal poblacional

r -> Coeficiente de correlación lineal muestral.

Este coeficiente de correlación lineal cumple las siguientes propiedades de r o ρ :

•$-1 \leq r \leq 1$.

•Si $r \approx 1$ entonces la relación entre X y Y puede considerarse aproximadamente lineal positiva o directa.

•Si $r \approx 1$ entonces la relación entre X y Y puede considerarse aproximadamente lineal negativa o inversa.

•Si $r \approx 0$ se interpreta como que hay ausencia de relación lineal entre X y Y (puede haber relación no lineal).

El coeficiente de correlación lineal poblacional en general no se conoce, por lo tanto para poder tener una estimación de él, es necesario hacerlo a partir de los datos de una muestra.

X : x1 x2 ... xn

Y : y1 y2 ... yn, n es el tamaño de la muestra

Como estimador de r se utiliza el coeficiente de correlación lineal muestral que se denota por r:

$$r = \frac{\sum_{i=1}^{n}\left[(x_i - \bar{x})(y_i - \bar{y})\right]}{\sqrt{\left(\sum_{i=1}^{n}(x_i^2 - n\bar{x}^2)(\sum_{i=1}^{n}(y_i^2 - n\bar{y}^2)\right)}}$$

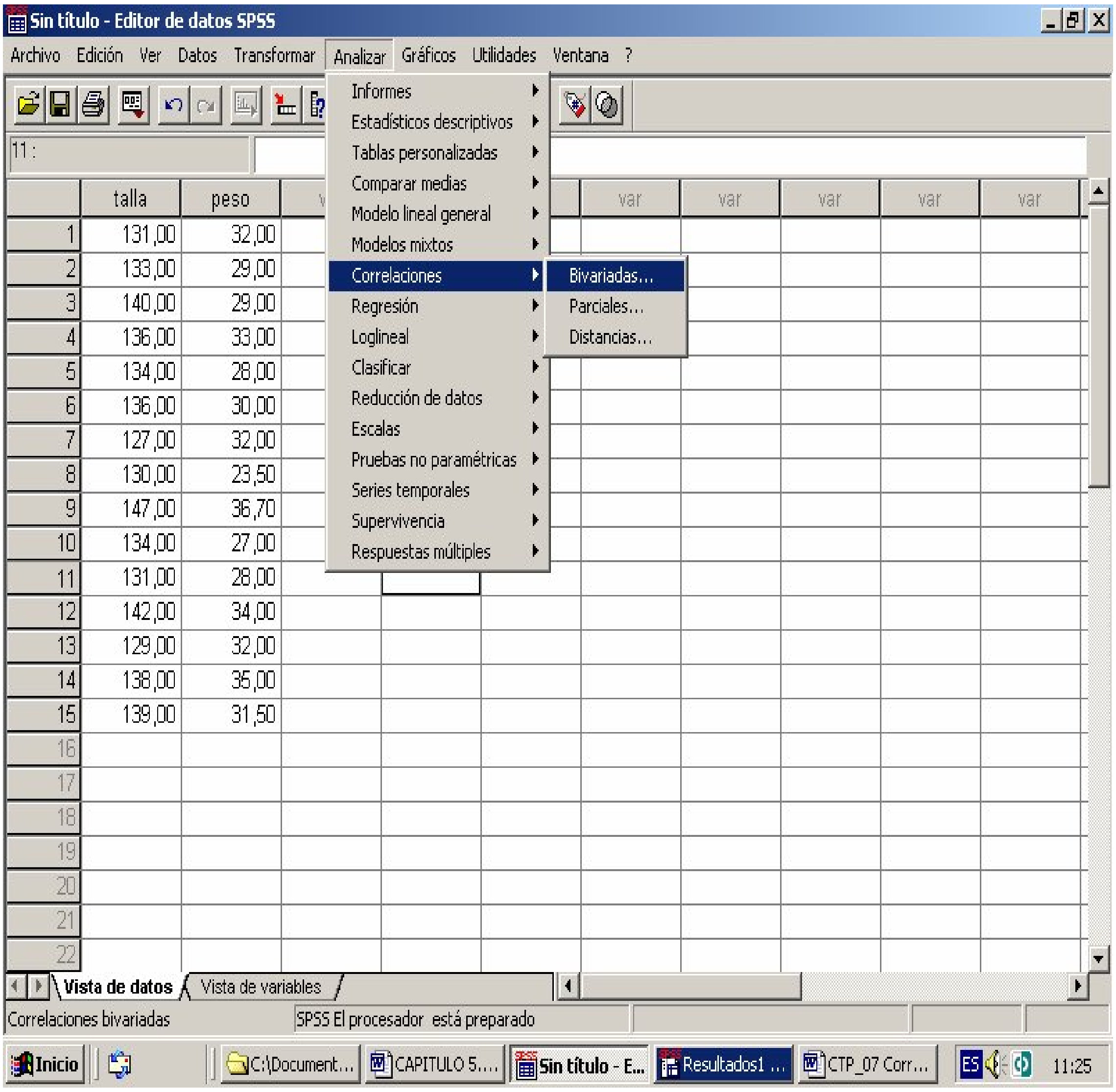
Si $|r| \geq 0.85$, se considera una fuerte asociación o relación lineal entre X y Y

Si $0.40 \leq |r| < 0.85$, se considera una asociación o relación moderada, entre X y Y

Si $|r| < 0.40$, se considera una asociación o relación débil entre X y Y.

Para realizar el cálculo del coeficiente de correlación entre dos variables utilizando el procesador estadístico SPSS, procedemos de la siguiente forma:

1ro. En el menú analizar buscamos correlaciones y en ella bivariada, obteniendo:



y a continuación trasladamos las variables Talla y Peso, obteniéndose:

Y en la ventana resultados:

Correlaciones

		TALLA	PESO
TALLA	Correlación de Pearson	1	,534
	Sig. (bilateral)	,	,040
	N	15	15
PESO	Correlación de Pearson	,534	1
	Sig. (bilateral)	,040	,
	N	15	15

* La correlación es significante al nivel 0,05 (bilateral).

El valor de r es 0,534, por lo tanto se considera a nivel de la muestra existe una correlación lineal moderada positiva y directa entre la talla y el peso de los niños de nueve años.

La significación p (p=0.040), que aparece en la tabla de resultados, es la significación asociada al estadígrafo que se utiliza en la prueba de hipótesis correspondiente al coeficiente de correlación lineal poblacional:

Ho: $\rho = 0$
H1: $\rho \neq 0$

Si p es menor que α se rechaza Ho, y se concluye que entre las variables Peso y Estatura, existe relación lineal a nivel poblacional.

5.3. REGRESIÓN LINEAL SIMPLE:

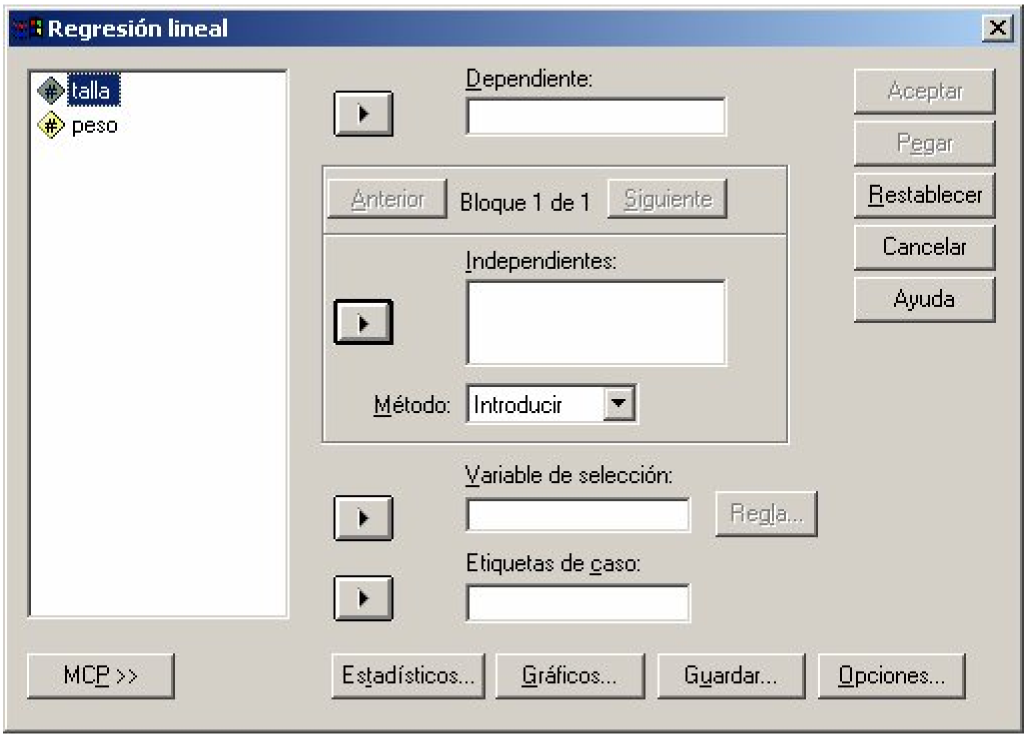
Si en el problema anterior la relación entre el peso (Y) y la estatura (X) se pudiera reflejar a través de una línea recta, esto sería conveniente, para poder predecir o inferir el peso de un individuo sin necesidad de pesarlo. La ecuación de la recta de regresión se representa como:

$y = \beta o + \beta 1x$

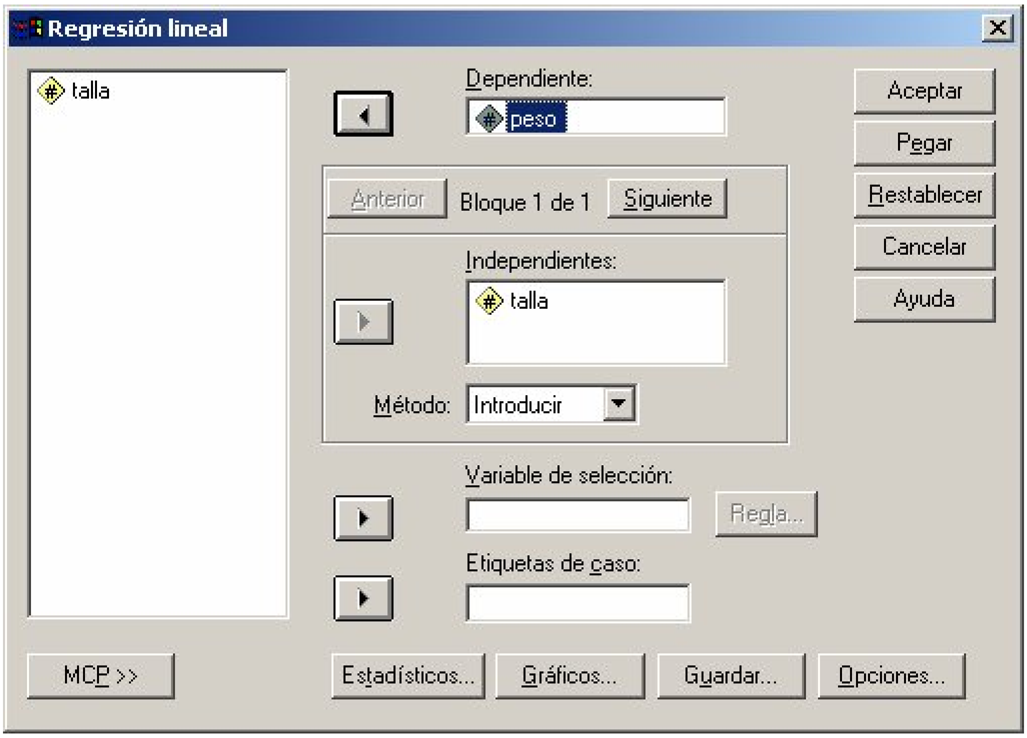
El parámetro βo representa el intercepto (INTERCEPT) de la recta con el eje de las Y y el parámetro $\beta 1$ la pendiente de la recta (SLOPE). A partir de los valores de la muestra se pueden estimar estos parámetros. Estos estimadores se denotan como bo y b1 respectivamente.

Utilizamos el procesador estadístico SPSS procederíamos de la siguiente forma:

1ro. En el menú analizar señalamos regresión y en esta lineal, obteniendo el siguiente resultado:



2do. Trasladamos como variable independiente la estatura y como variable dependiente el peso, obteniendo:



3. Y presionar aceptar en la ventana resultado obtenemos:

Coeficientes[a]

Modelo		Coeficientes no estandarizados B	Coeficientes no estandarizados Error típ.	Coeficientes estandarizados Beta	t	Sig.
1	(Constante)	-14,849	20,041		-,741	,472
	TALLA	,337	,148	,534	2,275	,040

a. Variable dependiente: PESO

Obteniéndose:

Intercepto (Bo) = -14,849

Pendiente (b1) = 0,336

Luego **la ecuación de regresión lineal simple** es:

Peso (estimado) = -14,849 + 0,336 * talla

Este modelo es válido solo cuando ya se ha demostrado que hay relación de tipo lineal entre las variables a nivel de toda la población.

5.5 EJERCICIOS Y PROBLEMAS DEL CAPITULO

1. Para los datos siguientes:

X	1	2	3	4	5
Y	2	4	5	7	8

Dibuje el diagrama de dispersión.
Halle el coeficiente de correlación lineal entre x e y. Interprete este valor.

2. La tabla siguiente expresa las edades (X) y la presencia de ácido ascórbico contenido en la saliva (mg/10000 cc) (Y) de 6 hombres:

X	20	36	22	28	42	21
Y	20	13	18	25	12	15

Halle el coeficiente de correlación lineal entre X e Y. Interprete este valor.
Suponiendo que la muestra fue correctamente seleccionada, ¿habrá correlación de tipo lineal a nivel de población?

3. Se realiza un estudio para valorar si existe relación entre la cantidad de cambios de curas y el tiempo que se emplean en el servicio de Endodoncia de un Clínica Dental. Se desea disponer de un modelo para estimar el tiempo que demorará en ejecutarse los cambios de curas, en función de la cantidad de cambios de curas, a partir de los datos obtenidos en un experimento, donde se mide el tiempo en minutos (y) y la cantidad de cambios de curas (x)

X	4	5	6	7	8	9
Y	20	25	32	36	38	40

Halle el coeficiente de correlación lineal entre X e Y. Interprete este valor.

Determine la ecuación de regresión.
Estime el tiempo que se emplearán si se realizarán 10 cambios de curas.

4. Se realiza un estudio para establecer una ecuación mediante la cual se pueda utilizar la concentración de estrona en saliva*(X) para predecir la* concentración del esteroide en plasma libre *(Y). Se extrajeron los siguientes datos de 14 varones sanos*:

X	1,4	7,5	8,5	9	9	11	13	14	14,5	16	17	18	20	23
Y	30	25	31,5	27,5	39,5	38	43	49	55	48,5	51	64,5	63	68

a) Estúdiese la posible relación lineal entre ambas variables.
b) Obtener la ecuación que se menciona en el enunciado del problema.
c) Determinar la variación de la concentración de estrona en plasma por unidad de estrona en saliva.

5. En un ensayo clínico realizado tras el posible efecto hipotensor de un fármaco, se evalúa la tensión arterial diastólica (TAD) en condiciones basales (*X*), y tras 4 semanas de tratamiento (*Y*), en un total de 14 pacientes hipertensos. Se obtienen los siguiente valores de TAD:

X	95	100	102	104	100	95	95	98	102	96	100	96	110	99
Y	85	94	84	88	85	80	80	92	90	76	90	87	102	89

¿Existe relación lineal entre la TAD basal y la que se observa tras el tratamiento?
¿Cuál es el valor de TAD esperado tras el tratamiento, en un paciente que presentó una TAD basal de 95 mm de Hg?

BIBLIOGRAFIA

1. Astruan Rodquez Dr Ma. Análisis Estadístico (en sopertte magnético). 2001
2. Colectivo de Autores del ICCMH. Bioestadística y Computación. Ministerio de Salud Publica. 1988.
3. Colectivo de Autores del CECAN. Cuaderno de ejercicios de Bioestadística. Facultad de Estomatología. 1981
4. Murria R. Spiegel. Teoría y Problemas de Estadística. Ediciones Revolucionarias. La Habana. 1987.
5. Luna del Castillo Juan de Dios y otros "Introducción al manejo del programa SPSS 12.0, (en soporte mangnético) septiembre 2005.
6. Rojo Abuin José Manuel Introducción al Paquete Estadístico SPSS. (en soporte magnético)
7. Candela Moreno Vicente, y otros "Manual de Introducción del SPSS" (en soporte magnético) 2001

ANEXO No 1

Los siguientes datos corresponden a un estudio realizado en un servicio de Cirugía Masilo Facial a un grupo de pacientes al llegar al cuerpo de guardia y a las 72 horas después del post operatorio.

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	60	1	1	43	100	70	40	1	45	160	60	1	2	70	1	1	1
2	2	70	2	1	32	110	80	45	2	60	180	70	1	2	80	1	1	2
3	2	75	3	2	23	120	90	60	2	26	185	75	2	1	70	2	1	2
4	1	68	2	2	25	156	126	70	2	18	140	68	1	2	69	2	1	2
5	1	80	2	1	47	148	160	75	1	17	100	80	2	2	65	1	2	1
6	2	82	1	1	44	159	180	68	2	15	85	82	1	2	70	2	1	2
7	2	86	1	2	45	170	185	80	2	14	90	86	2	1	65	1	1	2
8	1	60	1	2	44	125	140	82	1	13	90	60	1	2	70	2	3	1
9	2	78	2	2	34	180	100	86	2	18	120	78	2	2	68	2	3	2
10	1	50	3	2	23	200	85	60	2	35	125	50	1	1	56	2	2	2
11	2	40	3	2	14	100	90	78	1	23	160	40	2	2	70	1	2	1
12	2	45	3	2	15	110	90	50	1	32	180	45	1	2	80	1	1	1
13	1	58	1	1	17	120	120	40	2	45	126	58	2	1	70	1	3	2
14	1	40	2	2	56	156	125	45	2	60	70	40	1	1	69	2	1	2
15	1	45	3	2	32	148	160	58	2	26	80	45	2	2	65	2	1	2
16	2	60	3	1	33	159	180	40	2	18	90	60	1	2	70	1	2	2
17	2	70	2	1	35	170	126	45	1	17	126	70	2	2	80	2	1	1
18	2	75	1	1	38	125	70	60	1	15	160	75	1	2	70	1	1	1
19	1	68	2	1	36	180	80	70	2	14	180	68	2	1	69	1	1	2
20	2	80	2	2	41	200	90	75	2	13	160	80	2	2	65	2	1	2
21	1	60				100	126	68	2	18	180	60	2	2	70	2	2	2

			1	2	45													
22	2	70	1	1	60	110	160	80	2	35	185	70	1	2	65	1	1	2
23	2	75	2	2	26	120	180	82	1	23	140	75	1	1	70	2	1	1
24	1	68	3	2	18	156	185	86	2	32	100	68	1	2	68	1	3	2
25	2	80	3	1	17	148	140	60	2	33	85	80	2	2	56	2	3	2
26	1	82	3	1	15	159	100	78	2	45	90	82	1	1	70	2	2	2
27	2	86	2	1	14	170	85	50	1	60	90	86	2	2	80	2	2	1
28	1	60	2	1	13	125	90	40	2	26	120	60	1	2	70	1	1	2
29	2	78	1	2	18	180	90	45	2	18	125	78	2	1	69	1	3	2
30	1	50	1	1	35	200	120	58	1	17	160	50	1	1	65	1	1	1
31	1	40	1	1	23	100	125	40	2	15	180	40	2	2	70	2	1	2
32	2	45	2	2	32	110	160	45	2	14	126	45	1	2	80	2	2	2
33	1	58	2	1	33	120	180	60	1	13	70	58	2	2	70	1	1	1
34	2	40	3	2	25	156	126	70	1	18	80	40	1	2	69	2	1	1
35	1	45	3	2	26	148	70	75	2	35	90	45	2	1	65	1	1	2
36	2	60	3	1	25	159	80	68	2	23	126	60	1	2	70	1	1	2
37	1	70	2	2	27	170	90	80	2	32	160	70	2	2	65	2	2	2
38	2	75	1	1	29	125	126	82	2	33	180	75	1	2	70	2	1	2
39	1	68	2	2	30	180	160	86	1	34	160	68	2	1	68	1	1	1
40	2	80	2	1	31	200	180	60	1	33	180	80	1	2	56	2	3	1

Variables:

1. Sexo (1 femenino. 2 masculino)
2. Pulsaciones por minutos
3. Raza (1 blanca, 2 negra, 3 mulata)
4. Es hipertenso (1 si, 2 no)
5. Edad

6. Cloro en orina en el cuerpo de guardia
7. Sodio en orina en el cuerpo de guardia
8. Potasio en orina en el cuerpo de guardia
9. Hemograma en el cuerpo de guardia (1 normal, 2 alterado)
10. Leucograma en el cuerpo de guardia (1 normal, 2 alterado)
11. Cloro en orina 72 horas posteriores al post operatorio
12. Sodio en orina 72 horas posteriores al post operatorio
13. Potasio en orina 72 horas posteriores al post operatorio
14. Hemograma 72 horas posteriores al post operatorio (1 normal, 2 alterado)
15. Leucograma 72 horas posteriores al post operatorio (1 normal, 2 alterado)
16. Pulsaciones 72 horas posteriores al post operatorio
17. Tipo de tratamiento (1 tradicional, 2 experimental)
18. Respuesta al tratamiento (1 curado, 2 curado con secuela, 3 empeoro)

ANEXO No 2

En el municipio de Ranchuelo se encuentra una de las mayores fábricas de cigarro de la provincia de Villa Clara y en consecuencia se sospecha que en dicha población existe un numero considerable de fumadores.

Considerando las afecciones que dicho hábito produce en la salud de las personas en general y en particular las graves afecciones estomatológicas que están ocurriendo. La dirección de Estomatología considero que más del 30 % de la población era fumadora activa.

Con ese objetivo encuesta una muestra mediante un muestre aleatorio simple de 3000 personas de las cuales extrajo los siguientes datos:

- Muestra 3000
- Personas fumadoras...1000
- Personas fumadoras con afecciones estomatológicas....860
- Tipo de afección estomatológica en las personas fumadoras:
    - Malignas..........................70
    - Pre malignas................... 96
    - Común 694
- Personas no fumadoras con afecciones estomatológicas:
    - Malignas 32
    - Pre malignas 46
    - Común 870

ANEXO 3.

Los siguientes datos fueron tomados de un estudio realizado en una consulta de estomatología de la Clínica de Especialidades en enero del 2006.

No	Sexo	H. Fumar	Calidad de la atención	Número de caries	Nivel de Hb	Edad
1	Femenino	Si	Buena	2	12,5	27
2	Masculino	Si	Regular	4	14,3	23
3	Masculino	Si	Mala	6	15,1	35
4	Femenino	No	Regular	3	11,6	42
5	Masculino	No	Buena	1	12,7	58
6	Masculino	Si	Buena	3	13,4	35
7	Masculino	Si	Regular	5	14,8	24
8	Femenino	No	Buena	4	15,1	46
9	Masculino	Si	Buena	3	13,7	37
10	Femenino	Si	Buena	2	12,6	33
11	Masculino	Si	Buena	3	11,6	28
12	Masculino	No	Regular	2	15,4	43
13	Femenino	No	Mala	4	12,3	53
14	Masculino	Si	Buena	5	14,6	37
15	Femenino	Si	Buena	3	11,7	43
16	Masculino	Si	Regular	2	14,7	48
17	Masculino	No	Buena	3	15,2	41
18	Femenino	Si	Buena	4	12,5	38
19	Masculino	Si	Regular	3	13,7	45
20	Masculino	No	Mala	1	15,2	42