2011
Asunción - Paraguay

Guía Metodológica para el Análisis Científico de la Información en Salud

Tomo I:

El Análisis Descriptivo Aplicado - Nociones del Análisis Inferencial

Dr. Alexis Moreno

Consultor CIRD

Guía Metodológica para el Análisis Científico de la Información en Salud

Tomo I: El Análisis Descriptivo Aplicado - Nociones del Análisis Inferencial

Dr. Alexis Moreno

Consultor CIRD

Marzo 2011

Asunción – Paraguay

Programa para el Mejoramiento de la Calidad de la Información en Salud (PMCIS)

Eje Estratégico II: Supervisión Capacitante

Componente: Análisis de la Información

Guía Metodológica para el Análisis Científico De la Información en Salud

Tomo I: El Análisis Descriptivo Aplicado - Nociones del Análisis Inferencial

Propuesta elaborada en el marco del **Proyecto Mejoramiento del Sistema de Información en Salud (MSIS)** desarrollado por el **Centro de Información y Recursos para el Desarrollo (CIRD)** con financiamiento de **USAID.**

FICHA TÉCNICA

Título de la obra
Guía Metodológica para el Análisis Científico de la Información en Salud.
Tomo I: El Análisis Descriptivo Aplicado - Nociones del Análisis Inferencial.

Como citar la obra
Moreno A. Guía Metodológica para el Análisis Científico de la Información en Salud. Tomo I: El análisis descriptivo aplicado – Nociones de estadística inferencial. Asunción. USAID – CIRD. 2011

Editor responsable
Centro de Información y Recursos para el Desarrollo (CIRD)
Mariscal López 2.029 esquina Acá Carayá.
Telefax: (+595 21) 212-540 / 226-071.
www.cird.org.py cird@cird.org.py

Autor
Dr. Alexis Moreno Orué
amoreno@cird.org.py

Agradecimientos a los revisores de la obra
Dr. Carlos Rodríguez
Dra. Mirtha Mongelós
Dra. Gloria Aguilar
Dr. Arturo Moreno
Dr. Fidel Moreno

Marzo
2011.
Asunción, Paraguay.

Contenido Temático

❖ Elaboración y Ordenación de Datos

Siguiendo con las etapas del Método Estadístico. . .
1- Recogida de datos
2- Ordenación y Elaboración
(Pasos de la etapa de Elaboración):
2. 1) Revisión y Corrección de los datos recogida.
2. 2) Clasificación y Computación de los datos.
2. 3) Presentación por Cuadros y Gráficos.
3- Análisis de la información
4- Conclusiones
5- Decisiones finales

1. RECOGIDA DE DATOS

- Recoger datos consiste en anotar los valores observados de la variable de un determinado fenómeno.

- Estos datos recogidos constituyen lo que se llama **Tabla directa de datos.**

TABLA DIRECTA DE DATOS

4	9	5	3	15
8	7	0	4	12
11	4	2	4	7
10	3	2	9	3

- En la tabla directa de datos, éstos no están ordenados, en cuanto a su valor, sino que figuran según el orden en que se han realizado las observaciones.

2. ORDENACIÓN

- Los diversos aspectos que, dentro de la etapa de ordenación hemos estudiado en esta lección, aparecen recogidos en el siguiente esquema.

ORDENACIÓN DE DATOS
- **Creciente o decreciente**
- **En intervalos**
  - **Límites**
  - **Marca de Clase**
  - **Tamaño**
  - **Construcción**

- La ordenación, creciente o decreciente, consiste en colocar los datos de menor a mayor, o de mayor a menor respectivamente.

DATOS ORDENADOS

En orden creciente

0	2	2	3	3	3
4	4	4	4	5	7
7	8	9	9	10	11
12	12	15	15	15	16

En estudiantes es más frecuente ordenar los datos clasificándolos por grupos, denominados **intervalos de clase.**

Intervalos de clase: Es cada una de las partes en que se divide el recorrido para el estudio.

Un ejemplo de la ordenación de datos en intervalos de clase, partiendo de la Tabla directa de datos anterior, es el siguiente:

Intervalo de Clase

–1	–	1
2	–	4
5	–	7
8	–	10
11	–	13
14	–	16

Limites del intervalo: Son los valores extremos de dicho intervalo.

- Al valor máximo del intervalo se llama límite superior.
- Al valor mínimo del intervalo se le llama límite inferior.

Por ejemplo, en el intervalo 8 – 10. El límite inferior es 8. El límite superior es 10.

Un intervalo de clase viene determinado por sus límites inferior y superior.

Límites exactos del intervalo: Con objeto de evitar que un dato no encaje en ninguno de los intervalos de clase, se hace coincidir el límite inferior de un intervalo con el límite superior del anterior, de manera que no haya un salto entre el límite superior de un intervalo y el inferior del siguiente.

LÍMITES DE LOS INTERVALOS		
−6	–	3
4	–	13
14	–	23
24	–	33
34	–	43

LÍMITES EXACTOS DE LOS INTERVALOS		
−6,5	–	3,5
3,5	–	13,5
13,5	–	23,5
23,5	–	33,5
33,5	–	43,5

De esta manera obtenemos unos nuevos límites del intervalo que se denomina **límites exactos.**

- Límite exacto inferior. (Li)
- Límite exacto superior. (Ls)

Se dice que un intervalo está definido por sus **límites exactos** cuando:

- Su límite inferior coincide con el límite superior del intervalo anterior.
- Su límite superior coincide con el límite inferior del intervalo siguiente.

Veamos un ejemplo:

LÍMITES EXACTOS DE LOS INTERVALOS		
3	–	9
9	–	15
15	–	21
21	–	27
27	–	33

Consideremos el intervalo 21 – 27.

- Su límite inferior coincide con el límite superior del intervalo anterior.
- Su límite superior coincide con el límite inferior del intervalo siguiente.

- Cuando al realizar un estudio estadístico alguno de los datos observados coincida con el límite exacto superior de un intervalo, se clasificará en el intervalo siguiente.
- Cuando una tabla de intervalos venga definida por sus límites no exactos, se puede trasformar en otra en que dichos intervalos estén definidos por sus límites exactos.
- Para calcular el límite exacto superior de un intervalo, se suma el límite superior de dicho intervalo con el límite inferior del siguiente y se divide por dos.

$$\textbf{Límite exacto superior} = \text{(de un intervalo)} \quad \frac{\textbf{Límite superior} \text{ (de ese intervalo)} + \textbf{Límite inferior} \text{ (del intervalo siguiente)}}{\mathbf{2}}$$

Veamos un ejemplo:

LÍMITE DE LOS INTERVALOS
0 – 9
10 – 19
20 – 29
30 – 39
40 – 49

LÍMITES EXACTOS DE LOS INTERVALOS
– 0,5 – 9,5
9,5 – 19,5
19,5 – 29,5
29,5 – 39,5
39,5 – 49,5

- El límite exacto superior del primer intervalo es:

$$\frac{9 + 10}{2} = 9,5$$

- El límite exacto superior del segundo intervalo es:

$$\frac{19 + 20}{2} = 19,5$$

De igual forma se calculan los límites exactos superiores del resto de los intervalos, menos el del último.

El **límite exacto inferior de un intervalo coincide** con el **límite exacto superior del intervalo anterior.**

Cuando el tamaño del intervalo es constante, para calcular el límite exacto inferior del primer intervalo y el límite exacto superior del último intervalo, tendremos que:

- El límite exacto inferior del primer intervalo se calculará restando al límite exacto superior del primer intervalo el tamaño del mismo.

$$Li = Ls - c$$

En el ejemplo que venimos estudiando será: $Li = Ls - c = 9,5 - 10 = -0,5$

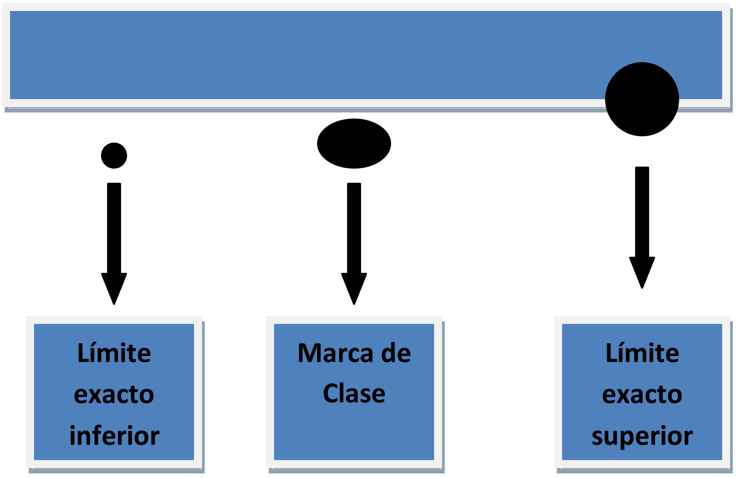
- El límite exacto superior del último intervalo se calculará sumando al límite exacto inferior del último intervalo, el tamaño del mismo.

$$Ls = Li + c$$

$$Ls = Li + c = 39,5 + 10 = 49,5$$

Cuando el tamaño del intervalo no es constante, para calcular el límite exacto inferior del primer intervalo y el límite exacto superior del último intervalo, tomaremos como valores de **c** los correspondientes al segundo y penúltimo intervalo respectivamente.

MARCA DE CLASE (X): Es el punto medio de un intervalo.



Para obtener la marca de clase de cada intervalo, se suman los límites exactos de dicho intervalo y el resultado se divide por dos.

$$\text{Marca de Clase} = \frac{\text{Límite exacto inferior} + \text{Límite exacto superior}}{2}$$

Es decir:

$$\boxed{MC = \frac{Li + Ls}{2}}$$

Observa la siguiente Tabla en la que se indican los límites exactos de los intervalos y su marca de clase:

LÍMITES EXACTOS DE LOS INTERVALOS	MARCA DE CLASE (x)
3 – 9	6
9 – 15	12
15 – 21	18
21 – 27	24
27 – 33	30

- La marca de clase del intervalo 3 – 9 se ha obtenido de la siguiente forma:

$$X = \frac{Li \quad + \quad Ls}{2} = \frac{3 \quad + \quad 9}{2} = 6$$

- La marca de clase del intervalo 9 – 15 se ha obtenido:

$$X = \frac{Li \quad + \quad Ls}{2} = \frac{9 \quad + \quad 15}{2} = 12$$

- De la misma manera se han obtenido las marcas de clase correspondientes al resto de los intervalos.

A efectos de cálculos posteriores, se puede considerar que todos y cada uno de los valores que entran dentro de un intervalo se pueden sustituir por un mismo valor: **el de su marca de clase.**

TAMAÑO DEL INTERVALO: Es la diferencia entre los límites exactos del intervalo y se representa por la letra **c.**

Es decir que: $\boxed{c = Ls - Li}$

Dado el intervalo de límites exactos 3 – 9, el tamaño del intervalo será:

$$c = Ls - Li = 9 - 3 = 6$$

El tamaño de los intervalos suele ser normalmente constante, aunque no necesariamente.

Conocidos el tamaño del intervalo y uno de los límites exactos del mismo, se puede conocer el otro límite exacto.

- Dado un intervalo cuyo tamaño es c =6, y su límite exacto inferior es 9 su límite exacto superior será:

$$Ls = Li + c = 9+6 = 15$$

- Dado un intervalo cuyo tamaño es c= 12, y su límite exacto superior es 25, su límite exacto inferior será:

$$Li = Ls - c = 27 - 6 = 21$$

CONSTRUCCIÓN DE INTERVALOS: Dado una Tabla directa de datos si queremos agrupar dichos datos en intervalos de clase, necesitamos determinar:

- El número de intervalos a utilizar, de tamaño constante.
- El límite exacto inferior del primer intervalo.

Para ello partimos e la siguiente expresión: $\boxed{\frac{R}{c} + 1}$

En ella:

- **R** es el recorrido.
- **c** es el tamaño constante de intervalo utilizado.

Al desarrollar esta expresión obtendremos, en general, una parte entera y una fraccionaria.

Veamos un ejemplo:

Supongamos que tenemos un recorrido R = 36 y tomamos un tamaño de intervalo c = 8.

La expresión quedará:

$$\frac{R}{c} + 1 = \frac{36}{8} + 1 = \frac{36 + 8}{8} = \frac{44}{8}$$

Este quebrado expresado en forma de número mixto será:

$$\frac{44}{8} = 5 + \frac{4}{8}$$

En definitiva, mediante el desarrollo de esta expresión, se ha obtenido:

- Una parte entera ("**n**") que es 5.
- Una parte fraccionaria ("**r/c**") que es $\frac{4}{8}$.

$$\boxed{\frac{R}{c} + 1} \longrightarrow \boxed{1 + \frac{n}{c}}$$

"n" = determina el n° de intervalos a usar.

"r/c" = determina las condiciones del **Li** del 1er. Intervalo.

Numero de intervalos:

Se aconseja por convención, que el n° de intervalos a usar esté comprendido entre 5 y 20. (En nuestro ej. n = 5; o sea, 5 intervalos).

Li del 1er. Intervalo:

Es el punto de partida para construir los intervalos, de forma que:

- El menor valor observado se halle en el 1er. intervalo.
- El mayor valor observado se halle en el último intervalo.

Para ello el **Li del 1er intervalo** debe cumplir unas **condiciones**:

- Ser **menor o igual** que **"a"** (menor valor observado).
- Ser **mayor** que **"a + r - c"** (**r** = numerador de la parte fraccionaria; **c** = tamaño constante del intervalo).

Condiciones del Li del 1er intervalo

- Ser **menor o igual** que **"a"** (menor valor observado).
- Ser **mayor** que **"a + r - c"**

a + r - c < **Li del 1er. intervalo** ≤ **a**

Siguiendo con nuestro ej.: **"r/c"** = 4/8; **"a" =** 5

5 + 4 - 8 = 1 < **Li del 1er. intervalo** ≤ **5**

Límites Exactos (Ls = Li + c)

1,5 - 9,5

9,5 - 17,5

17,5 - 25,5

25,5 - 33,5

33,5 - 41,5

Construcción de intervalos por la regla de Sturges

$$\boxed{K = 1 + 3{,}322\ (\log_{10} n)}$$

K (número de intervalos de clase); **n** (tamaño muestral)

De la tabla directa de la página 6:

K = 1 + 3,322 (log 20) K = 1 + 3,322 (1,3010) = **5,32**

El valor resultante (ej. K=5) debe considerarse como una guía y puede variar según convenga al estudio.

La amplitud del intervalo **(c)** $= \frac{R}{k} = \frac{15}{5} = \mathbf{3}$ (R = recorrido)

Las amplitudes deberían ser igual en las clases, aunque esto es a veces, imposible.

Tarea: Ordenar en intervalos de clases:

Valores de glicemia en ayunas

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

1. Presentación Estadística

Método Estadístico: Etapas desde la perspectiva del SIECS[1]

- Planificación
- Recolección de datos
- Elaboración y Ordenamiento
- Presentación
- Análisis propiamente dicho
- Conclusiones y sugerencia (Informes de las UARs)

El método estadístico que constituye el motivo conductor natural de la estadística aplicada.

En esta guía pretendemos darle un enfoque particular en relación al sistema de información que ha sido diseñado como ideal, el SIECS.

Si bien podemos apreciar que hay una etapa específica de análisis; ya en la etapa de elaboración y ordenamiento de los datos, y en particular, en la presentación estadística de los datos e indicadores constituyen una forma inicial de analizar los datos presentarlos de una manera resumida y lógica.

Existen dos formas clásicas de presentar estadísticamente datos e indicadores: la presentación tabular o cuadros; y la presentación gráfica.

La presentación tabular resulta imprescindible en un reporte estadístico o en un trabajo de investigación; los gráficos pueden complementar a los cuadros, pero no siempre son necesarios.

Es importante además recordar los criterios de escogencia sobre el tipo más apropiado de cuadros o gráficos, según las características particulares del tipo y variable de estudio. Es muy frecuente ver, sobre todo en quien no maneja conceptos metodológicos de estadística, usar bonitos gráficos desde una perspectiva artística, pero poco útiles para resumir una información, y aun más, darnos una interpretación equivocada de la realidad por una mala praxis en elección del gráfico.

[1] Sistema de la información para la Evidencia y el Conocimiento en Salud (SIECS).

1.1. La presentación tabular o cuadros

a) Características generales:

La finalidad de los cuadros estadísticos es presentar en forma resumida e inteligible determinado material numérico.

Aunque la del cuadro variará de acuerdo a los datos que intenta resumir, hay algunos principios comunes que deben tenerse en cuenta.

En todo cuadro debe considerarse:

Cuadro estadístico: es la presentación de la expresión numérica de los datos recogidos, por medio de una tabla o cuadro para resumir y hacerla más comprensible.

Todo cuadro debe tener:

1- Título: permite entender el contenido del cuadro. Debe ser:

a) completo: para ello debe contestar: ¿QUE universo se estudia?, ¿CÓMO se clasificarán los elementos según sus características? ¿DONDE? A qué lugar se refieren los datos. ¿CUANDO? La fecha del estudio.
b) breve: lo más conciso y claro posible, pero no debe sacrificarse la claridad por la concisión.

Título no recomendable:

"*Cuadro que muestra la distribución de las defunciones habidas en el Hospital de Clínicas, durante el año 2000, clasificadas de acuerdo con la edad y sexo de los fallecidos".*

Título Correcto:

"Defunciones por edad y sexo. Hospital de Clínicas, 2000".

2- Cuadro o tabla propiamente dicho:

El cuerpo del cuadro consta de un conjunto de casillas o celdas, dispuestas en columnas (las verticales) y filas (las horizontales).

La primera columna y la primera fila tienen una finalidad diferente a las restantes, porque en ellas irán las diferentes subdivisiones de la clasificación que se adopte, o los encabezamientos que indiquen a que se refiere los datos numéricos inscriptos.

La primera fila, es la de los ENCABEZAMIENTOS, los cuales indican a qué se refieren los datos que van inscritos en las celdas subyacentes. Ellos al igual que los títulos, deben ser breves, pero suficientemente explícitos. Así por ejemplo, en vez de poner simplemente Edad, es preferible poner Edad en años o Edad en meses, etc.

La primera columna conocida como Columna Matriz se destina a asentar las diferentes clases de la escala de clasificación utilizada. Cuando las observaciones se clasifican de acuerdo a una única escala, digamos edad, las subdivisiones de esta, deben ir en esta columna.

3. Notas explicativas:

Con el fin de que no haya dudas sobre el contenido del cuadro, éste se acompaña a veces de notas explicativas, que pueden ir en su parte superior o inferior. Convencionalmente las notas colocadas en la parte superior, afectan todo el contenido del cuadro, mientras que aquellas que se colocan en la parte inferior, solo se refieren a cifras de determinadas celdas o de una fila o columna en particular, lo cual se indicara con un pequeño número o letra.

Morbilidad según edad y causas. Lugar. Año

Edad (años)	**Examinados**	**Neumonía** (3)	**Tbc** (1)	**Chagas** (2)	**Diarreas**
< de 1 a	200	0	4	1	90
1 - 2 a	380	7	12	6	250
3 - 4 a	400	50	30	8	300
5 y + a	700	90	90	17	450
TOTAL:	1680	147	136	32	1090

(1)Tuberculina "+" (2) prueba de aglutinación "+" (3) radiología y bacteriología +

b) Diferentes clases de cuadros:

De acuerdo a su finalidad los cuadros estadísticos pueden dividirse en 2 categorías:

1. Cuadros de propósito general.

Sirven de base para la construcción de los segundos, son cuadros extensos, de resúmenes, frecuentemente destinados a presentar material básico a otros investigadores y de ahí que cuando se publican se acompañan de extensas notas explicativas y de cuidadosa mención de los procedimientos y métodos utilizados en la recolección de los datos

2. Cuadros de propósito especial

Presentación tabular de las distribuciones de frecuencia, de las series cronológicas, de los datos de asociación.

Son cuadros generalmente elaborados con propósitos analíticos. Habitualmente se intercalan en la presentación de trabajos y monografías originales y destinadas a mostrar determinadas relaciones sobre las cuales el autor quiere llamar la atención y que constituyen el núcleo de las conclusiones que de la investigación se derivan.

2.1. Presentación tabular de las Distribuciones de Frecuencia

Llamaremos distribución de frecuencias al conjunto de clases junto a las frecuencias correspondientes a cada una de ellas.

Si los individuos se clasifican de acuerdo a una única escala, el cuadro podrá hacerse como el que sigue abajo:

Defunciones por accidentes, por grupos de edad. Paraguay 2000

Años de edad	N° de defunciones
0 – 4	501
5 – 14	453
15 – 24	605
25 – 44	931
45 – 64	499
65 – 84	218
total	3.207

En los cuadros de este tipo, se acostumbra poner una columna más, con la distribución porcentual de los casos, lo cual facilita grandemente las comparaciones.

Muertes Por Causas En Infantes. Lugar. Tiempo (año)

Causas	Nº muertes	Porcentaje
Diarreas	5870	27.4
IRA	2500	11.7
Sepsis	2000	9.3
Congénitas	1555	7.3
Otros	9500	44.3
TOTAL:	21425	100.00

Finalmente, se pueden agregar dos columnas más, correspondientes a las frecuencias acumuladas tanto absolutas como relativas.
De esta manera, la forma general de un cuadro de distribución de frecuencias quedaría estructurado como sigue:

Variab	Frec.	Frec.	Frec. Abs. Acumu.	Frec. Rel. Acumu.
x	n	f	N	F

Para presentar distribución de frecuencias en escala cuantitativa se debe dividir el rango de valores de las observaciones en una serie de intervalos distintos que no se superpongan.

Si hay demasiados intervalos, el resumen no tiene grandes ventajas respecto a los datos sin procesar. Si hay muy pocos, se pierde gran cantidad de información.
Los intervalos se construyen de tal manera que tengan amplitudes iguales, para facilitar las comparaciones entre clases; pero pudieran existir clases de amplitudes diferentes.

Niveles de colesterol en hombres de entre 25 y 34 años. Laboratorio Central del MSP y BS. Enero a diciembre del 2010.

Nivel de colesterol (mg/100 ml)	Número de hombres
80-119	13
120-159	150
160-199	442
200-239	299
240-279	115
280-319	34
320-359	9
360-399	5
Total	**1067**

*La tabla muestra como se distribuyen los valores del nivel de colesterol a través de los intervalos. Se observa que las mediciones se localizan entre **80 y 399** mg/100 ml con relativamente pocas mediciones en los extremos del intervalo y una gran proporción de valores entre **120 y 279** mg/100 ml. El intervalo **160-199** mg/100 ml contiene el mayor numero de observaciones. La tabla permite una mejor comprensión de los datos de lo que haría una lista de 1067 lecturas de niveles de colesterol.*

2.2. Presentación tabular de las series Cronológicas

Cuando la escala de clasificación es el tiempo, mostrando como varía un fenómeno en relación a él. La elaboración del cuadro es muy semejante al caso anterior.

Sin embargo, como tales cuadros sólo pretenden mostrar la variación de un fenómeno de una época a otra, en ellos se omiten los totales y lógicamente, al no existir éstos, será imposible el cálculo de la respectiva columna de porcentajes.

No obstante, si el cuadro se refiere a lo ocurrido en una población cuyo número de habitantes ha variado a través de los años, es conveniente colocar una última columna que señale el número de veces que ocurrió el fenómeno estudiado por cada 1.000, 10.000 ó 100.000 habitantes. En otras palabras: las cifras absolutas se deben acompañar de los coeficientes o tasas respectivas, con lo cual se facilitará la comparación de los datos.

Defunciones por accidentes y tasas por 100.000 habitantes. Lugar. Año

Años	N° de Defunciones	Defunciones por 100.000 hbts.
1957	**2.872**	**43.3**
1958	**3.255**	**47.3**
1959	**3.390**	**47.6**
1960	**3.217**	**43.7**
1961	**3.223**	**42.4**

2.3. Presentación Tabular de los Datos de Asociación

Si los individuos se clasifican simultáneamente de acuerdo a dos escalas por ejemplo: edad y sexo (datos de asociación), una escala irá en la vertical y otra en la horizontal. El que una u otra vayan en la vertical o en la horizontal, no cambia el significado del cuadro.

Sin embargo, es conveniente poner en la vertical, aquella escala que presente más subdivisiones, ya que el ojo humano compara más fácilmente, números dispuestos en columnas de arriba abajo, que arreglados unos al lado de otro, en filas horizontales.

Observe que como hay dos escalas, la tabla debe tener dos totales. Estos suelen ponerse en la última columna y en la última fila, pero si se prefiere pueden colocarse en la primera columna y en la primera fila.

Téngase en cuenta también, que es posible presentar en el mismo cuadro tanto las cifras absolutas como los porcentajes o tasas correspondientes. Debe evitarse que el cuadro quede con demasiadas columnas, pues en tal caso su interpretación se haría difícil. En tales ocasiones, es preferible presentar la información en 2 ó más cuadros distintos (lo explicado puede verse en el cuadro siguiente).

Defunciones por accidente, por sexo y grupos de edad, Paraguay, 2000

Edad (años)	Sexo		Total
	Hombres	Mujeres	
0 – 4	275	226	501
5 – 14	288	165	453
15 – 24	519	86	605
25 – 44	835	96	931
45 – 64	400	99	499
65 – 84	115	103	218
Total	**2.432**	**775**	**3.207**

Si los individuos se clasifican al mismo tiempo de acuerdo a tres escalas, como ser edad, sexo y causa de la defunción, el cuadro aparecerá de la manera ilustrada más abajo.

Defunciones según Causa; Sexo y grupo Etario. Paraguay. 2000

Edades	SEXO			CAUSAS DE ACCIDENTE					
	M	F	Total	T	S	C	V	Otras	Total
15 años	563	391	954	272	169	63	63	387	954
15 – 44	1.354	182	1.536	786	219	110	29	392	1.536
45 y más	515	202	717	310	48	169	9	181	717
Total	2.432	775	3.207	1.368	436	342	101	960	3.207

c) Errores en la presentación tabular

Entre los errores que se cometen al elaborar un cuadro estadístico, deben evitarse especialmente los siguientes:

1. Disposición incorrecta de los datos:

Contrástese el cuadro anterior que es correcto, con el que aparece a continuación, el cual ilustra un error cometido frecuentemente por los principiantes. El error consiste en que el cuadro no clasifica a cada individuo de acuerdo a tres escalas. En realidad son dos cuadros diferentes colocados uno al lado del otro.

No indica cuántos hombres (o mujeres) murieron en cada grupo etario según causas, lo cual si puede ser determinado en el cuadro anterior.

2. Títulos y encabezamientos incompletos o inadecuados:

El cuadro debe comprenderse fácilmente, sin necesidad a recurrir al texto que acompaña, lo cual será imposible si los títulos o encabezamientos son incompletos.

3. Cuadros que muestran solamente porcentajes:

Por lo general un cuadro no debe mostrar solamente porcentajes sin indicar las cifras de donde proceden, pues un porcentaje de 50% puede significar 1 caso en 2; 10 casos en 20; 100 casos en 200, y como es obvio mientras menor es el número de casos, menor valor tendrá el porcentaje.

4. Cuadro sobrecargados:

Aquellos que intentan mostrar muchos datos a la vez, resultan confusos e inadecuados, es preferible hacer varios cuadros separados.

d) Manera de leer un cuadro estadístico

Leer cuidadosamente el título: La lectura cuidadosa del título es necesaria con el fin de comprender perfectamente **a que se refiere el cuadro**. Definir la unidad de análisis temporal y espacialmente, así como características específicas el estudio (distribuidos de acuerdo a su edad, sexo, etc.). Nos puede indicar también el método de análisis utilizado (porcentajes, tendencias centrales y de dispersión, etc.).

Leer notas explicativas: Las notas explicativas le acompañan al cuadro, permiten a menudo su **mejor compresión**. En primer lugar determinar que no se estudia a todo el universo, sino solamente **una muestra**. Esto es importante, pues ya sabemos que los resultados de las muestras están sometidos al **error del muestreo**. En segundo lugar la nota indica si el estudio se refiere a un grupo particular como las tasas específicas. Se pueden definir conceptualmente las variables y sobre todo patologías. Por ejemplo: TBC, definida como baciloscopía (+) o como mantoux (+). Menciona la fuente usada.

Averiguar les unidades de medidas utilizadas: El encabezamiento de la primera columna explica la unidad de medida y escala de clasificación.

Fijarse en el promedio o porcentaje general del grupo: En el cuadro se muestra que el 20.5 %, es decir 1 de cada 5 personas, murió por alguna de las causas consideradas.

Ej. N° 2: 1151 x 100 / 28240 = 5,49 %

Relacionar el promedio general del grupo con cada una de las variedades que se estudian: Las variables presentadas en el cuadro anterior son: edad, sexo, causas; y deben analizarse separadamente.

Mortalidad x Sexo: (mujeres) = 1033 x 100 /18030 = 5,73 % (> 5,49 %)

El grupo < 1 año muere 342 x 100 / 1780 = 19,21 %; El grupo 1 a 4

años 371 x 100 / 8750 = 4,24 %; El grupo 5 y más años 320 x 100 /

7500= 4,27 %

El grupo - 1 año: su mortalidad es > la mortalidad general (19,21 y 5,49

%).

Los otros dos grupos tienen un porcentaje (4,24 y 4,27) inferior al general (5,49).

Mortalidad x Sexo: (hombres) = mueren 5,07 % < al porcentaje general: 5,49 %

El grupo - 1 año muere en un 17,24 %; El grupo 1 a 4 años muere en un 2,71 %; El grupo 5 y + = 1,57 %

La mortalidad del grupo - 1 año (17,49 %) es > al general (5,49 %); los

Otros grupos etarios < al general.

Relacionar entre si los promedios o porcentajes de las variables que se estudian: Esto es necesario porque puede haber alguna interacción entre las variables que cause diferencias observables.

Sexo: La mortalidad femenina (5,73) es > que la masculina (5,07).

Relacionar sexo por grupos de edades: Mujeres: mueren más los < 1

año: 19,21 %, le siguen los de 5 + años (4,27); y los de 1 a 4 años (4,24).

Hombres: mueren mas los - de 1 año: 17,24 %, le siguen los de 1 a 4 años (2,71) y los de 5 + años (1,57).

Relacionar los grupos de edad por sexo: < de 1 año: mueren más mujeres: 19,21% que los homb: 17,24 %; 1 a 4 años: mueren más las mujeres 4,24 % que los varones; 5 y más años: mueren más mujeres 4,27 % que los varones.

Relacionar entre si los grupos de edades: Mortalidad en < de 1 año: =

635 x 100 / 3480 = 18,24 %; 1 a 4 años = 3,51 %; 5 y más años = 4,09 %

La mortalidad más elevada corresponde al grupo < de 1 año: 18,24 %.

Le siguen los grupos 5 y más años, y 1 a 4 años con 4,09 y 3,51 %

respectivamente.

Para los datos de muerte por diarreas, sepsis, Ca. y otras causas

corresponde hacer el mismo tipo de lectura desde el punto 5 hasta el 9 inclusive.

Buscar irregularidades en los datos

Datos anómalos e inconsistentes

Conclusiones finales

. a) La mortalidad es mayor en el primer año de vida.

b) El sexo femenino es factor en la mortalidad aumentándola en los grupos 1 a 4 años; 5 y más años.

c) La diarrea es la principal causa de muertes en todos los grupos de edad, y principalmente en el menor de 1 año.

d) La mortalidad por sepsis aumenta con los años, siendo más elevada en las mujeres.

2.2. La presentación Gráfica

Características generales:

La representación gráfica consiste en un artificio eficiente para obtener rápidamente una impresión visual de conjunto de una serie de datos y hacer resaltar sus relaciones, que es en muchos casos la primera etapa de un análisis.

La tabla estadística resume los datos que disponemos de una población analizarla de una manera más sistemática y resumida . Para una percepción rápida de las características de la población resulta útil el uso de gráficos.

Constituyen ayuda para interpretar los datos contenidos en las tablas y no un método substitutivo.

Los gráficos son más fáciles de interpretar, y como éstos, tienen título, el gráfico propiamente dicho, y las notas explicativas.

Escalas del gráfico y errores en su empleo:

- Los gráficos generalmente se inscriben en las “coordenadas rectangulares”, formada por la intersección de dos líneas en ángulo recto.
- La línea horizontal o abscisa se destina generalmente para las diferentes clases de escalas que se utilizan (años, meses, sexo, raza, vacunas, etc.).
- La línea vertical u ordenada generalmente se usa para anotar la frecuencia ó número de veces que se observa el fenómeno en estudio, según la indicación de la abscisa.

Recordar que:

a) Las escalas inscriptas en las abscisas y ordenadas deben tener la misma longitud o algo mayor la horizontal, no sobre pasando al doble de la vertical. Así si la ordenada mide 5 cms., la horizontal debe medir 5, tolerándose un máximo de 10 cms.
b) Las dos escalas deben comenzar en cero, corresponde al ángulo de unión de la línea vertical con la horizontal. Sin embargo, cuando los valores se presentan son elevados y tienen poca fluctuación, se hace uso de un artificio: “partir” el gráfico para poner de manifiesto las fluctuaciones.
c) el título debe ser claro y conciso y estar escrito debajo.
d) las escalas utilizadas deben ser claras: años, meses, minutos, kilos, gramos, etc.
e) los gráficos no dan idea de exactitud matemática, para eso están los cuadros.
f) las escalas que se utilizan deben ser por lo general números enteros.
g) si los años o meses se inscriben debajo de las ordenadas, que se las pintó en rayas de puntos como en los gráficos 4 - 5, el punto que representa determinada frecuencia se hará sobre ellas.

Si los años o meses se inscriben entre las ordenadas, la inscripción del punto se hará generalmente en la mitad de dicho espacio.

Partes del gráfico
Por homología con las tablas, los gráficos están constituidos por tres partes:

El título
El gráfico propiamente dicho
Las notas explicativas (optativo).
Aquí es aplicable la conceptualización que se hiciera con respecto a las partes de la presentación tabular

Tipos de Gráficos
Los objetivos trazados en esta Guía Metodológica hacen que no sea pertinente un listado exhaustivo de los tipos de gráficos, sino más bien, centrarnos en algunos que son representantes por antonomasia.

Primero estableceremos los criterios de escogencia de los distintos gráficos, para pasar a las especificaciones de los tipos de gráficos más frecuentemente usados.

Gráficos según el tipo de datos y la naturaleza de la variable.

Tipos de Datos	Naturaleza de la variable	Tipo de Gráfico
Distribuciones de Frecuencias	Cualitativa	Barras Segmentadas Barras simples Sectorial o circular Pictogramas
	Cuantitativa Discreta o Discontinua	Barras Sectorial Pictogramas
	Cuantitativa Continua	Histograma Polígono de Frecuencias Gráfico de correlación (con 2 datos)
Datos de Asociación	dependiente/independiente	Gráfico de correlación
Series Cronológicas	variaciones tiempo / edad	Barras Diagrama lineal

Gráficos para variables cualitativas

Los gráficos más usuales para representar variables de tipo nominal son los siguientes:

Gráfico de barras:
Se representamos en el eje de ordenadas las frecuencias absolutas o las frecuencias relativas, y en abscisas las modalidades o clases de la variable. Si, mediante el gráfico, se intenta comparar variables y/o poblaciones entre sí, existen variaciones (barras dobles, compuestas). Cuando los tamaños de las dos poblaciones son diferentes, es conveniente utilizar las frecuencias relativas, ya que en otro caso podrían resultar engañosas.

Gráfico de barras simples para una variable cualitativa

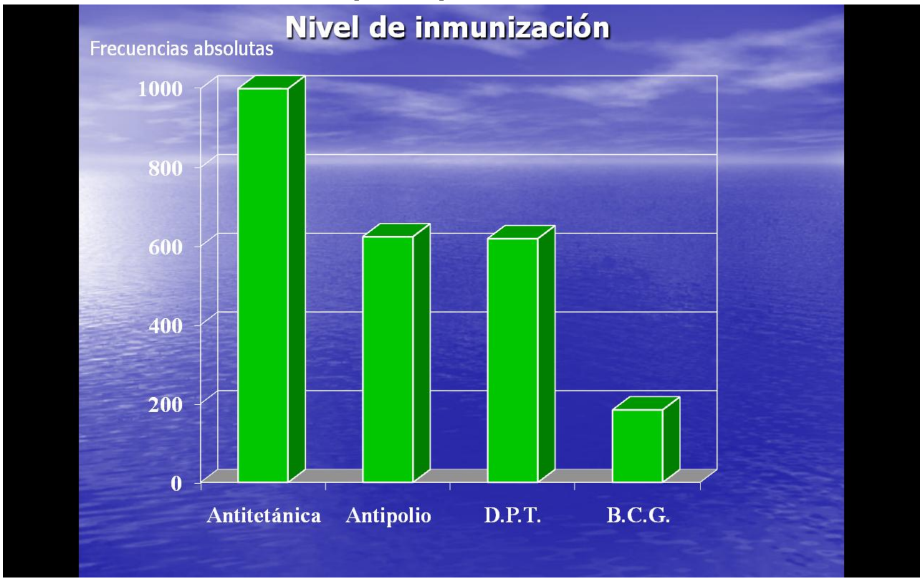
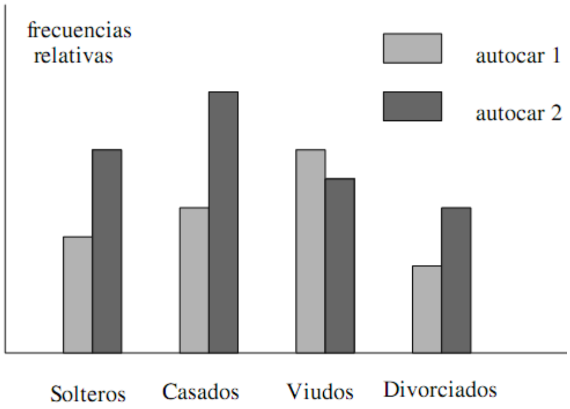
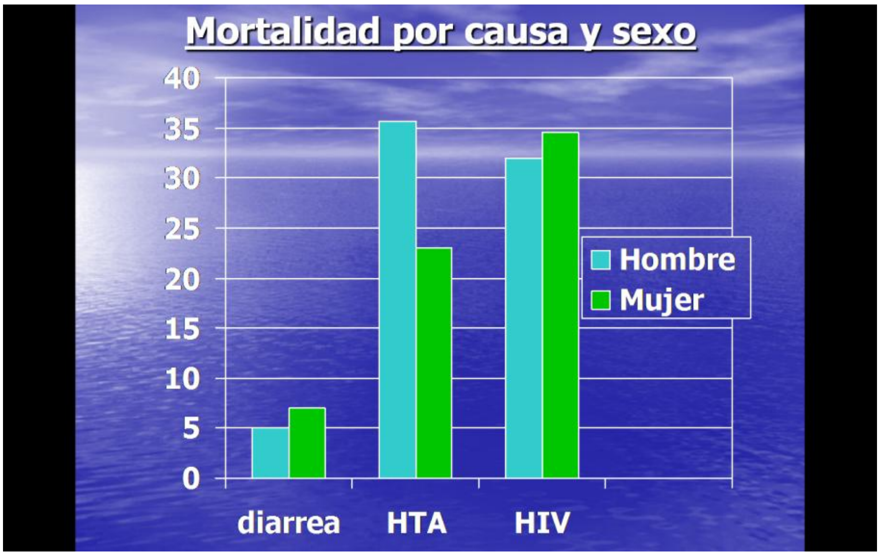


Gráfico de barras dobles

Se usa para representar datos de asociación, cuyas dos escalas sean cualitativas. También se la usa cuando se quiere comparar dos distribuciones de frecuencia con relación al tiempo.

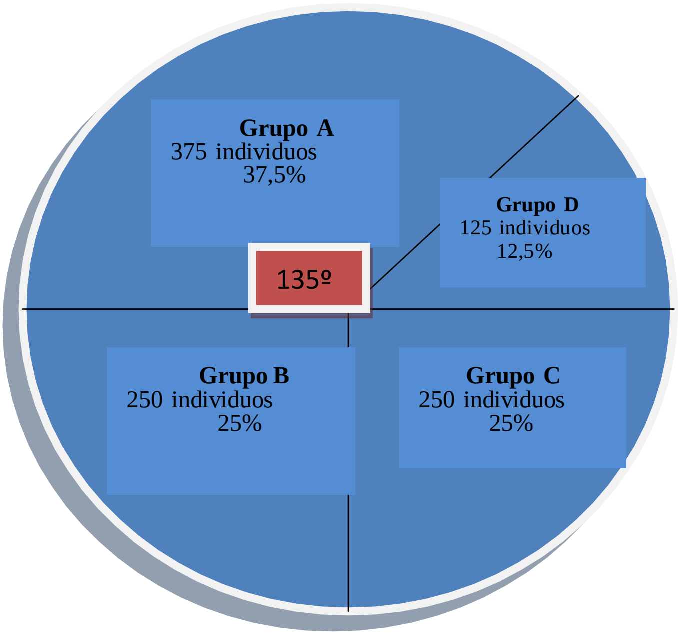
Gráfico de barras dobles para comparar una variable cualitativa en diferentes poblaciones.





Diagramas de sector (también llamado de torta).

Consiste en una circunferencia en la cual se escriben en cifras absolutas o relativas la proporción en que participan los diversos componentes del objeto en estudio. El porcentaje que corresponde a una variable se multiplica por 3,6 y se obtiene el N° de grados que le corresponde en el círculo.



El arco de cada porción se calcula usando la regla de tres:

$$\begin{array}{lll} n \longrightarrow & 360^\circ = \dfrac{360 \cdot f_1}{n} = \text{grados para } f_1 \\ f_1 \longrightarrow & ? \end{array}$$

$$\begin{array}{lll} 1000 \longrightarrow & 360^\circ = \dfrac{360 \cdot 375}{1000} = \mathbf{135^\circ} \\ 375 \longrightarrow & ? \end{array}$$

Como en la situación anterior, puede interesar comparar dos poblaciones. En este caso también es aconsejable el uso de las frecuencias relativas (porcentajes). Otra posibilidad es comparar las 2 poblaciones usando para cada una de ellas un diagrama semicircular (ver más abajo). Sean n1 ≤ n2 los tamaños respectivos de las dos poblaciones. La población mas pequeña se representa con un semicírculo de radio r1 y la mayor con otro de radio r2.

La relación existente entre los radios, es la que se obtiene de suponer que la relación entre las áreas de las circunferencias es igual a la de los tamaños de las poblaciones respectivas, es decir:

$$\frac{r_2^2}{r_1^2} = \frac{n_2}{n_1} \Longleftrightarrow r_2 = r_1 \cdot \sqrt{\frac{n_2}{n_1}}$$

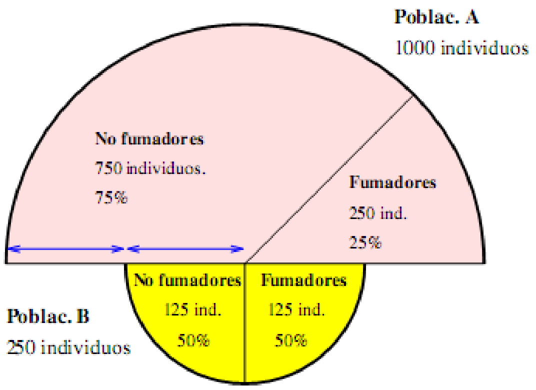


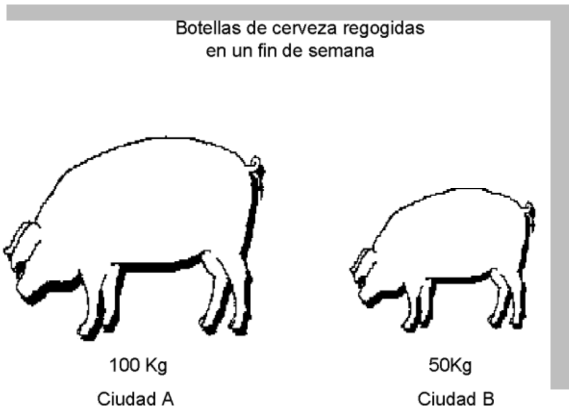
Diagrama de sectores para comparar dos poblaciones

Pictogramas

Expresan con dibujos (generalmente relacionado al tema de estudio), las frecuencias de las clases de la variable. Estos gráficos se hacen representado a diferentes escalas un mismo dibujo, como vemos más abajo.

El escalamiento[2] de los dibujos debe ser tal que cada uno de ellos sea proporcional a la frecuencia de la clase que representa. Este tipo de gráficos suele usarse en los medios de comunicación, para que sean comprendidos por el público no especializado, sin que sea necesaria una explicación compleja.
Las áreas son proporcionales a las frecuencias.

Pictograma



Gráficos para variables cuantitativas

Existen dos tipos de variables cuantitativas: discretas y continuas.

[2] Es un error hacer la representación con una escala tal que el perímetro del dibujo sea proporcional a la frecuencia, ya que a frecuencia doble, correspondería un dibujo de área cuádruple, lo que da un efecto visual engañoso.

Gráficos para variables discretas

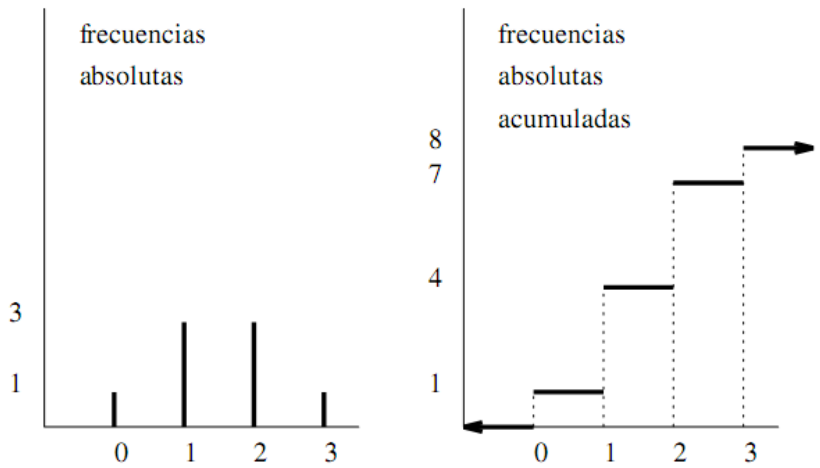
Cuando representamos una variable discreta, usamos el gráfico de barras. Se recomienda que las barras sean estrechas para representar el que los valores que toma la variable son discretos. El diagrama integral o acumulado tiene, por la naturaleza de la variable, forma de escalera.

Se lanzan tres monedas al aire en 8 ocasiones y se contabiliza el número de “cruces”, obteniéndose los siguientes resultados: 2,1,0,1,3,2,1,2

En primer lugar observamos que la variable “x” es cuantitativa discreta, presentando las clases: 0, 1, 2, 3. Ordenamos a continuación los datos en una tabla estadística.

x	n_i	f_i	N_i	F_i
0	1	1/8	1	1/8
1	3	3/8	4	4/8
2	3	3/8	7	7/8
3	1	1/8	8	8/8
	$n = 8$	1		

x = clases de la variable “x”; n_i = frecuencias absolutas; f_i = frecuencias relativas; N_i = frecuencias absolutas acumuladas; F_i = frecuencias relativas acumuladas.



Familias Clasificadas por su N° de hijos

N° de hijos (x_i)	1	2	3	4
Frecuencias (n_i)	1	3	5	3

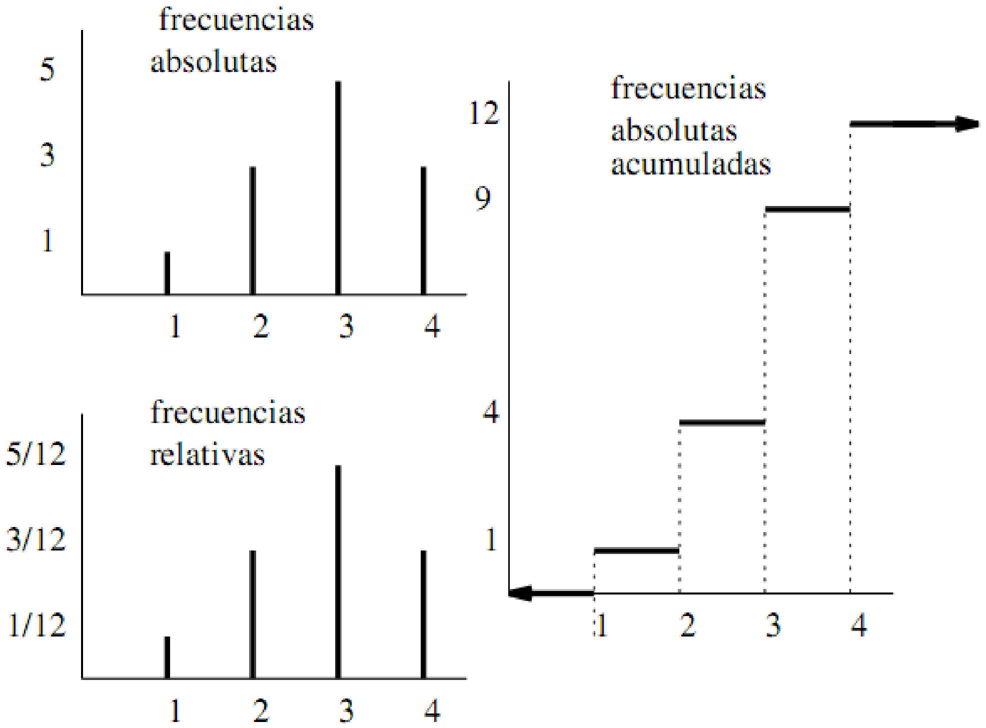
Gráfico (barras) e integral para una variable discreta. El diagrama acumulado (integral) contabiliza el número de observaciones de la variable inferiores o iguales a cada punto del eje de abscisas.

Comparar los diagramas de barras para frecuencias absolutas y relativas. Realizar el diagrama acumulativo creciente.

Solución: En primer lugar, escribimos la tabla de frecuencias en el modo habitual:

Variable	F. Absolutas	F. Relativas	F. Acumuladas
1	1	0,083	1
2	3	0,250	4
3	5	0,416	9
4	3	0,250	1
	1	1	

Con las columnas relativas a xi y ni realizamos el diagrama de barras para frecuencias absolutas, lo que se muestra en la figura de abajo

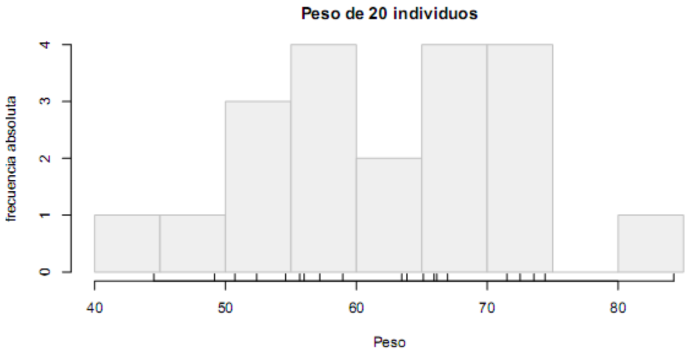


Gráficos para variables continuas

Cuando las variables son continuas, utilizamos por defecto el histograma y el polígono de frecuencias. En particular, si se parte de una tabla ordenada en intervalos de clase.

Justamente, el histograma se construye a partir de la tabla estadística, representando sobre cada intervalo, un rectángulo que tiene como base la amplitud de dicho intervalo de clase. El criterio para calcular la altura de cada rectángulo es el de mantener la proporcionalidad entre las frecuencias absolutas (o relativas) de cada intervalo para que sea el área de los mismos, la que determina el escalamiento (a diferencia de la altura del gráfico de barras).

Histograma para una variable continua (peso en kg).

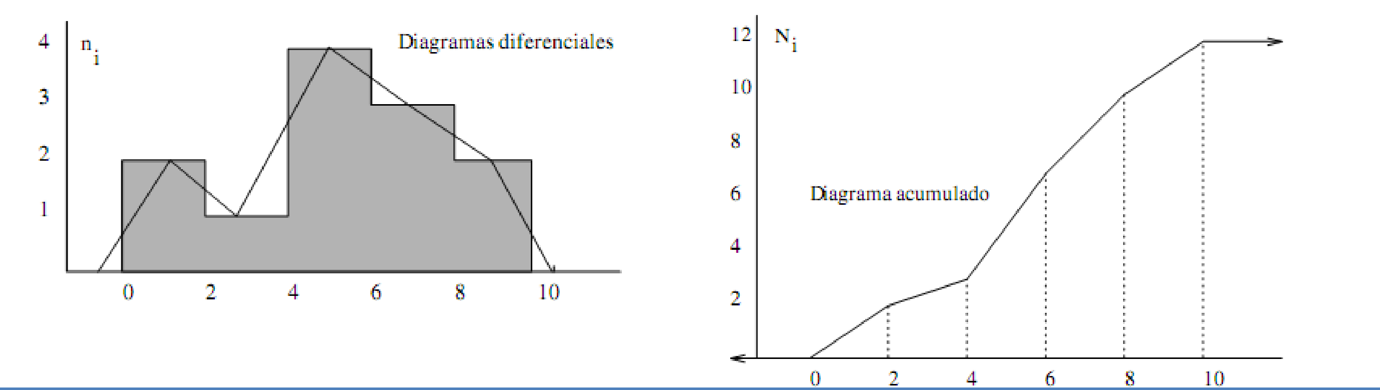


El polígono de frecuencias se construye fácilmente si tenemos representado previamente el histograma, ya que consiste en unir mediante líneas rectas los puntos del histograma que corresponden a las marcas de clase. Para representar el polígono de frecuencias en el primer y último intervalo, suponemos que adyacentes a ellos existen otros intervalos de la misma amplitud y frecuencia nula, y se unen por una línea recta los puntos del histograma que corresponden a sus marcas de clase. Existe, pues total correspondencia de estos dos tipos de gráficos, en general, se prefiere el histograma, por ser de una interpretación más intuitiva; siendo sólo la comparación de dos o más series donde tiene elección el polígono de frecuencias, dada la imposibilidad gráfica de superponer dos o más histogramas.

El diagrama integral para una variable continua se denomina también polígono de frecuencias acumulado, y se obtiene como la poligonal definida en abscisas a partir de los extremos de los intervalos en los que hemos organizado la tabla de la variable, y en ordenadas por alturas que son proporcionales a las frecuencias acumuladas. Dicho de otro modo, el polígono de frecuencias absolutas es una primitiva del histograma.

Intervalos	c_i	n_i	N_i
0 — 2	1	2	2
2 — 4	3	1	3
4 — 6	5	4	7
6 — 8	7	3	10
8 – 10	9	2	12
		12	

Diagramas diferenciales e integrales para una variable continua.



Pasos para elaborar un histograma

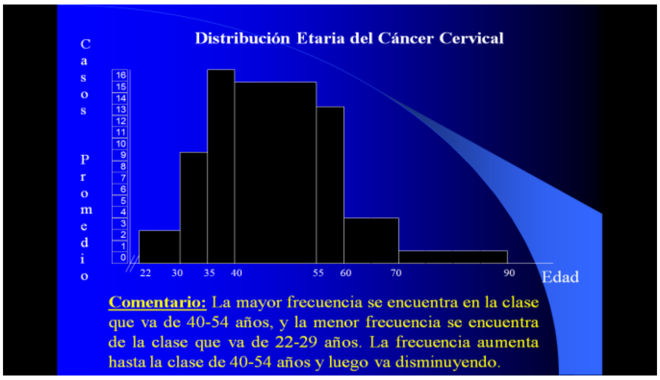
Tenemos la siguiente distribución de frecuencias: Muestra de distribución etaria del Ca cervical: 22-29 (18 casos); 30-34 a. (45 c.); 35-39 a. (79 c.); 40-54 a. (225 c.); 55-59 a. (63 c.); 60-69 a. (45 c.); 70-90 a. (13 c.).
Como se trata de una variable cuantitativa continua y ordenada en intervalos de clases, el gráfico de elección es un histograma.

Pasos:

1. Realizar la presentación tabular correspondiente, como se ve más abajo. La amplitud (c), corresponde al rango de valores de cada clase, así la primera por ejemplo (22 – 29), incluye 22, 23. . . . . 29 años = 8 valores. La amplitud está representado por la anchura de cada rectángulo (en la abscisa). Para la altura, **NO** se debe usar directamente las frecuencias absolutas[3], sino el resultado de la razón "f/c" porque se necesita un promedio unidad de la escala usada. Aquí sería casos de Ca Cervical por año de edad.

2. Al trazar el sistema de coordenadas, el ángulo debe comenzar en "0"; por tanto para guardar una coherencia debe cuidarse que si la primera clase no comienza en "0" se debe dejar un espacio correspondiente; o "partir" el eje de abscisas como se observa en la figura. De igual modo cuidar la proporcionalidad delas amplitudes de cada clase. Recordar que el inicio de cada clase, se corresponde con la finalización de la que le precede. Por convención en el gráfico se escriben los inicios de las clases.

Edades	Frecuencia (f)	Amplitud (c)	F /c
22-29	18	8	2,25
30-34	45	5	9
35-39	79	5	15,8
40-54	225	15	15
55-59	63	5	12,6
60-69	45	10	4,5
70-90	13	21	0,6



[3] Un caso especial de histograma es cuando las amplitudes de todas las clases son iguales. En este caso se puede usar directamente "f" para el escalamiento de la ordenada.

La siguiente distribución se refiere a la duración en minutos de reuniones de trabajo de integrantes de las UARs

Duración en min.	Personal UARs
300-500	50
500-700	150
700 -1.100	275
1.100	25
	Total 500

- Representar el histograma de frecuencias relativas y el polígono de frecuencias.
- Trazar la curva de de frecuencias relativas acumuladas
- Determinar el número mínimo de integrantes que se reunieron menos de 900 minutos.

Solución:

Siempre que se tenga un rango tan amplio de valores resulta más conveniente agruparla en intervalos de clases. La consecuencia es una ligera pérdida de precisión.

El último intervalo está abierto por el límite superior. Dado que en él hay 25 observaciones puede ser conveniente cerrarlo con una amplitud "razonable". Todos los intervalos excepto el tercero tienen una amplitud de 200 min, luego podríamos cerrar el último intervalo en 1.300 min[4].

Antes de realizar el histograma conviene hacer una observación importante, el histograma representa las frecuencias de los intervalos mediante áreas y no mediante alturas. Sin embargo es mucho más fácil hacer representaciones graficas teniendo en cuenta estas últimas; por eso si todos los intervalos tienen la misma amplitud no es necesario diferenciar entre los conceptos de área y altura, pero en este caso el tercer intervalo tiene una amplitud doble a los demás, y por tanto hay que repartir su área en un rectángulo de base doble (lo que reduce su altura a la mitad).

Así será conveniente añadir a la habitual tabla de frecuencias una columna que represente a las amplitudes a_i de cada intervalo, y otra de frecuencias relativas rectificadas, (f'_i) para representar la altura del histograma.

Intervalos	a_i	n_i	f_i	f'_i	F_i
300 — 500	200	50	0,10	0,10	0,10
500 — 700	200	150	0,30	0,30	0,40
700 — 1.100	400	275	0,55	0,275	0,95
1.100 — 1.300	200	25	0,05	0,05	1,00
		n=500			

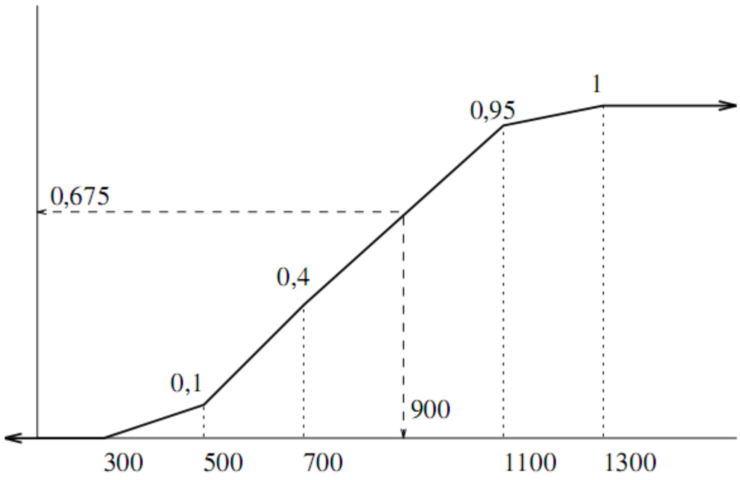
[4] Cualquier otra elección para el límite superior del intervalo que sea de "sentido común" sería válida..

En el gráfico de arriba se ve que sumando frecuencias relativas, hasta las 900 min de duración hay 0,10 + 0,30 + 0,275 = 0,675 = 67,5 % de los participantes de las UARs.

Esta cantidad se obtiene de modo más directo viendo a qué altura corresponde al valor 900 en el diagrama de frecuencias acumuladas.

Como en total son 500 participantes del as UARs, la cantidad de personas que trabajaron 900 horas o menos es 0,675 x 500= 337,5. Redondeando, 338 personas.

Diagrama acumulativo de frecuencias relativa



El gráfico de dispersión o correlación, se usa para presentar datos asociados en escala cuantitativa. Se lo estudia más adelante en el capítulo de herramientas de análisis para datos de asociación.

3. Herramientas y Métodos de Análisis

Métodos Estadísticos de Análisis

El análisis de todo estudio debe comenzar con una evaluación de la información disponible:

- Se cumplió con lo planificado y los objetivos del estudio.
- La metodología de recolección de datos.
- El rigor científico y la fidelidad de la información.

Los factores que determinan el método de análisis son:

Las técnicas de análisis estadístico y su escogencia varían según:

- El número de individuos estudiados
- El tipo de información recogida
- La escala de clasificación utilizada
- El propósito del estudio

1- Número de individuos estudiados

- Series no agrupadas: cuando el número de individuos es bajo.
- Series agrupadas: cuando incluye una cantidad apreciable de individuos es necesario clasificar previamente en grupos o clases según la característica estudiada.

2- Tipo de información recogida

- Distribución de frecuencia
- Datos de asociación
- Series cronológicas

3-Escala de clasificación utilizada

El número de escalas depende de la cantidad de variables consideradas en el estudio determina un:

- Análisis univariante; bivariante; o multivariante.

  La característica de la variante determina:

  - Escala cualitativa o nominal
  - Escala cuantitativa: se pueden ponderar o medir.

    Discretas: Series Ordinales

    Continuas: Series interválicas y proporcionales.

4- El propósito del estudio

- *Estudios Descriptivos*: busca resumir y destacar las características más importantes del grupo en estudio. Consiste en el análisis de la descripción realizada de los datos cualitativos y/o cuantitativos en estudio que se encuentran resumidos en cuadros y gráficos.
- *Estudios Comparativos*: busca si hay o no diferencias reales entre los grupos estudiados y las razones para explicarlas (significancia estadística).

Esta clasificación es más bien aparente y sería mejor considerarla como pasos del proceso inferencial.

Técnicas de análisis de los estudios descriptivos

1. Distribuciones de Frecuencia

1.1. En escala cualitativa: *Las frecuencias relativas:* porcentajes, tasas, etc.
1.2. En escala cuantitativa: *a) las tendencias centrales:* (promedio, mediana, modo). *b) medidas de dispersión* (desviación estándar, etc.).

2. Datos de Asociación

2.1. Ambas escalas son cualitativas: se resumen en frecuencias relativas.
2.2. Una escala es cuantitativa y la otra es cualitativa: (según la finalidad del estudio), se podrá escoger cualquier medida ya mencionada.
2.3. Ambas escalas son cuantitativas: se utiliza el coeficiente de correlación o el coeficiente de regresión.

3. Series Cronológicas

Se resumen por tendencias calculadas, cambios porcentuales, y técnicas de regresión.

3.1. Distribuciones de frecuencia en escala cualitativa

***1-Presentación tabular*:**

1ª columna: subdivisiones de la escala de clasificación utilizada. Ej. Sexo (masculino / femenino)	**2ª columna: nº de sujetos con dichas características (frecuencia absoluta).**	**3ª columna: frecuencias relativas (porcentajes).**

2-Presentación gráfica:

Barras (si tiene muchas subdivisiones) o de sector. Se los puede presentar en cifras absolutas o relativas.

3- Análisis mediante frecuencias relativas:

Análisis de los datos numéricos relacionados a otra cantidad que se usa como base de comparación.
Importancia: Evidencian la intensidad de la relación de 2 o más datos estudiados.
Los más comunes son: la razón, porcentaje, la proporción, las tasas, índices, coeficientes.

✓ ***Razones**:*
Relaciona frecuencias absolutas de 2 categorías distintas. Ej. Hay 297 hombres y 99 mujeres. La relación hombres / mujeres es: 297/99 = 3. (relación 3 x 1).

✓ ***Proporciones**:*
Relaciona frecuencias absolutas de 1 categoría con el total del grupo. Ej. 297 hombres + 99 mujeres = 396. La proporción de hombres es: 297/396 = 3/4 = 0,75.

Hay 3 hombres por cada mujer en este grupo.

- ✓ ***Porcentajes***:
  Es la frecuencia absoluta de 1 categoría multiplicado por 100 y dividido por el total del grupo; o sea, una proporción x 100. Ej. 297/396 x 100 = 75 %.
  Cuando la serie tenga 2 categorías pueden usarse cualquiera de estas FR; si tiene 3 o más se usan proporciones y porcentajes.

Ventajas de los porcentajes:

1- Mejor comparación de 2 o más series con totales diferentes (son reducidos a 100).
2- Resumen la probabilidad de ocurrencia del fenómeno estudiado.

Variable (sexo)	Grupo A	Grupo B
Hombre	297	255
Mujer	99	85
Total	***396***	***340***

A pesar de la aparente disparidad ambos grupos tienen el mismo porcentaje de hombres (75 %).
Hay una probabilidad de un 75 % de que la persona sea varón y un 25 % que sea mujer en estas series.

Frecuencias absolutas y relativas de colesterolemia de 2294 hombres[5]. Laboratorio Central del MSP y BS. Enero – diciembre del 2010.

Nivel de colesterol (mg/100 ml)	Edades 25-34		Edades 55-64	
	Número de hombres	Frecuencia relativa (%)	Número de hombres	Frecuencia relativa (%)
80-119	13	1.22	5	0.41
120-159	150	14.06	48	3.91
160-199	442	41.42	265	21.60
200-239	299	28.02	458	37.33
240-279	115	10.78	281	22.90
280-319	34	3.19	128	10.43
320-359	9	0.84	35	2.85
360-399	5	0.47	7	0.57
Total	**1067**	**100**	**1227**	**100**

Debido a que hay más hombres en los grupos de mayor edad, no es apropiado comparar las columnas de frecuencias absolutas de los 2 conjuntos de individuos. Interpretación:

- Los hombres mayores tienen niveles superiores de colesterol que los más jóvenes.
- Los más jóvenes tienen una gran proporción de observaciones inferiores a 200 mg/ml, mientras que los hombres mayores tienen una proporción mayor superior de este valor.

[5] *La tabla presenta la frecuencia absoluta y relativa de los niveles de colesterol de 1067 individuos de 15 a 34, así como las frecuencias correspondientes a un grupo de 1227 individuos de 55 a 64 años de edad.*

✓ ***Tasas***

Accidentes automovilísticos según sexo del conductor.

Sexo del conductor	Casos	Porcentajes
Hombre	400	80 %
Mujer	100	20 %
Total	**500**	***100 %***

Los porcentajes calculados señalan un 80 % de probabilidad de accidentes para un hombre pero sería un absurdo inferir en base a esta información que los hombres tiene mayor peligro de accidentes.

Las tasas son frecuencias relativas usadas cuando las relaciones del fenómeno estudiado son proyectas a su población correspondiente.
Miden la probabilidad de ocurrencia de un fenómeno en relación a la población en la que puede acontecer.

$$\text{Tasa} = \frac{\textit{N° de veces que ocurrió el fenómeno estudiado}}{\textit{Población en la cual ocurrió el fenómeno}} \times 10^n$$

Una **TASA** es simplemente un quebrado. El numerador, indica el número de veces que ocurrió determinado fenómeno en un área limitada y en un período de tiempo definido.

El denominador indica el número de habitantes de la población en la cual puede ocurrir el fenómeno descrito en el numerador.

Como el numerador nunca será mayor que su denominador, el resultado será menor que la unidad y para evitar el uso de decimales, los resultados multiplican por un amplificador (100, 1000, etc.)

Tasas crudas y específicas

Tasa cruda: intenta medir la probabilidad de que un hecho ocurra en la población total.

$$TC = \frac{\text{nº de hechos ocurridos en la población en un tiempo.}}{\text{Población total a mitad del mismo periodo.}}$$

Tasa específica: se refieren a partes o características particulares de la población: edad, sexo, raza, etc.

$$T.E. = \frac{\text{nº de hechos ocurridos en el grupo específico de la población en un periodo.}}{\text{Población de ese grupo específico a mitad del mismo periodo.}}$$

Tasas importantes en medicina

- ***Tasas de mortalidad**:* expresan el riesgo de morir.
- ***Tasas de morbilidad**:* expresan el riesgo de adquirir determinadas enfermedades.
- ***Tasas de natalidad**:* miden el crecimiento de las poblaciones.
- ***Tasas de letalidad**:* indican cuan graves son las enfermedades.

Estas tasas pueden calcularse para toda la población o para partes (segmentos o grupos) de la misma y se llaman tasas específicas.

Recomendaciones para construir frecuencias relativas

- Especificar que tipos de fenómenos se están comparando y cual de ellos fue usado como base.
- El valor del cociente no refleja ninguno de los fenómenos por separado, sino la relación de magnitud entre ambos.
- Con frecuencias absolutas poco numerosas, no se deben sacar frecuencias relativas, pues inducen a error.
- Todo cociente puede expresarse por su valor real o multiplicado por un factor de amplificación.
- En las tasas habrá que tener en cuenta la población expuesta al riesgo, para que el denominador sea correcto.

3.2. Distribución de frecuencia en escala cuantitativa

Presentación tabular igual a las otras distribuciones de frecuencias.

Presentación gráfica*:* si la escala es continua se hace en histogramas (por defecto) ó polígonos de frecuencia. Si la escala es discontinua se utiliza el gráfico de barras.

Análisis

Los fenómenos biológicos no suelen ser constantes, por lo que es necesario que junto a una medida que indique el valor alrededor del cual se agrupan los datos, se asocie una medida que haga referencia a la variabilidad que refleje dicha fluctuación.

Es decir, dado un grupo de datos organizados en una distribución de frecuencias (o bien una serie de observaciones sin ordenar), pretendemos describirlos mediante dos o tres cantidades sintéticas..
También pueden analizarse mediante porcentajes y porcentajes acumulados, pero se prefieren analizar usando:

a) **Constantes centrales** *(promedio aritmético, mediana y modo):* señalan cifras alrededor de las cuales se da la mayoría de las observaciones.
b) **Medidas de dispersión** *(desviación estándar, el espacio intercuartilar)*: señalan como se distribuyen las observaciones con respecto a las constantes centrales.
c) **Porcentajes**, menos frecuentemente.

El tipo de análisis a utilizarse depende de la finalidad del estudio, ya que la información brindada por ellas es diferente. Suelen utilizarse simultáneamente.

Análisis mediante frecuencias relativas: se utiliza el porcentaje y porcentaje acumulado.

Distribución de pacientes según el sobrepeso

sobrepeso Peso en grs.	**Número de pacientes**	**Porcentaje (%)**	**Porcentaje acumulado**
800 - 999	3	2,5	2,5
1000 - 1199	7	5,8	8,3
1200 - 1399	30	24,8	33,1
1400 - 1599	40	33	66,1
1600 - 1799	27	22,3	88,4
1800 - 1999	12	9,9	98,3
2000 - 2199	2	1,7	100
TOTAL	**121**	**100**	

En este caso, el simple promedio aritmético $\frac{\text{suma de peso}}{\text{Nº de casos}}$ = 1506,6 gramos;

oculta que el 33.1 % (porcentaje acumulado) de los pacientes tienen menor sobrepeso que 1400 grs.; teóricamente considerado como mínimo aceptable y que el estudio de los porcentajes nos aclara que: 2.5 % pesan de 800 a 999 grs.; 83 % de 1000 a 1199 grs. y 24.8 % de 1200 a 1399 grs.

3.3.1. Constantes centrales en Series No Agrupadas (simples)

a) **Promedio aritmético o media aritmética ($\bar{X}$):**

Es el cociente de la sumatoria de los datos observados (Σ xn) dividido el número total de los mismos (N)

$$\bar{X} = \frac{\Sigma\, xn}{N} \qquad \bar{X}^{(*)} = \frac{\Sigma\, xn.fn}{N}$$

(*) *Si hay valores de la variable que se repiten: Se modifica el numerador por la sumatoria de los productos de cada variable por su frecuencia correspondiente (Σ xn. fn)*

Ejemplo: Los pesos en Kgs. de 6 niños son: 40; 38; 42; 44; 39; 45 La media aritmética de estos valores es:

$$\frac{40+38+42+44+39+45}{6} = 41.5 \text{ Kgs.}$$

Si representamos el valor 40 como x1; 38 como x2; 42 como x3; ....Xn.
Al número de observaciones, por "n" y a la media aritmética por x tendremos:

$$\dot{X} = \frac{x1+x2+x3+............Xn}{N}$$

Ahora supongamos que sean 200 los casos y que los valores en kg. de los niños aparecen repetidos un cierto número de veces:
40 Kgs. 20 veces; 38 Kgs. 40 veces; 42 Kgs. 30 veces: 44 Kgs. 50 veces;
39 Kgs. 15 veces; 43 Kgs. 45 veces.

$$X = \frac{(40x20)+ (38x40)+ (42x30)+ (44x50)+ (39x15)+ (43x45)}{20 + 40 + 30 + 50 + 15 + 45} =$$

$$X = \frac{800 + 1520 + 1260 + 2200 + 585 + 1935}{200} = \frac{8300}{200} = 41.5$$

Si usamos los símbolos f1 para indicar la frecuencia asociada 20 de la primera variable de peso 40, y f2 a la segunda frecuencia 40 de la segunda frecuencia 30 de la tercera variable 42........fn, tendremos:

$$X = \frac{(x_1.f_1) + (x_2.f_2) + (x_3.f_3)..............x_n.f_n}{\sum f}$$

Media Aritmética ponderada: pueden los valores x1; x2; xn etc. ser asociado con ciertos pesos: w1; w2........Wn que dependen del significado o importancia asignado a los valores. En estos casos la media aritmética se calcula así:

$$Xp = \frac{x1.w1+x2.w2+....xn.wn}{\sum w} =$$

Ejemplo: si en un curso académico dividido en tres períodos, se asigna a cada período una importancia de 1, 2, 3 respectivamente y un alumno obtiene las siguientes calificaciones: 50, 60, 75, su nota promedio será:

$$Xp= \frac{50x1 + 60x2 + 75x3}{1+2+3} = \frac{395}{6} = 65.8$$

b) La mediana (Ma): es aquella cifra de observación que divide a la serie en dos partes iguales, dejando igual número de observaciones a cada lado.

Es necesario ordenar las cifras de las observaciones de menor a mayor.

Se tienen los siguientes datos: 8, 18, 8, 3, 4, 4, 5, 6, 8.

Se ordena 3, 4, 4, 5, **6,** 8, 8,8, 18.

La mediana es 6; porque es el valor de la serie que deja a ambos lados igual número de valores.

Cuando el número de cifras de observación es par, no existe una cifra que ocupa el centro de la serie. En estos casos para obtener la mediana se promedian los dos valores centrales. Ej. 5, 5, 7, **9, 11**, 12, 15, 18.

La mediana está comprendida entre los valores 9 y 11, que se promedia:

$$\frac{9+11}{2}=10$$

c) El modo (Mo): el modo de un conjunto de valores es aquel que se presenta con mayor frecuencia, es decir es el valor más común. El modo puede no existir o puede no ser único.

Ej. 3, 7, 3, 8, 5. El Mo = 3.

5, 3, 2, 4, 8, 9 No hay modo porque todos se presentan con una sola frecuencia.

1, 2, 5, 3, 3, 6, 5, 4, 4, 7, 9, 13. Los modos son 3, 5, 4, pues ellos se presentan en dos ocasiones, en tanto que las otras cifras se presentan una sola vez.

CONSTANTES CENTRALES EN SERIES AGRUPADAS

a) Promedio Aritmético: para calcularlo se presume que todos los valores que caen dentro de un intervalo de frecuencia de clase coinciden con el punto medio de la clase. En estos casos los valores X1, X2........Xn (1) corresponden a la marca de clase y f1, f2,......fn; son las frecuencias correspondientes a esos valores.
Así la fórmula para el cálculo de la media aritmética será: $X = \frac{\Sigma X.f}{n}$

Si tuviéramos los datos sin agrupar, la media se obtendría directamente de los valores no ordenados y sería más exacta, porque en el caso de las series agrupadas suponemos que todos los valores comprendidos dentro del intervalo 20 - 24 por ejemplo, toman el valor de 22, lo cual no es cierto. Pero el error cometido calculando la media a partir de los datos agrupados no es importante si la variable es continua y la distribución simétrica.

X = La suma de los productos de cada marca de clase por su frecuencia correspondiente ($\Sigma\ x_n . f_n$) dividida por el número total de observaciones (N)

$$\mathbf{X = \frac{(\Sigma\ x_n . f_n)}{N}}$$

.

Peso en Kgr. de pacientes

LUGAR XX FECHA XX

Peso en Kgs	Nº de casos (f_n)	Punto medio (x_n)	X.f
20 - 24	4	22	88
25 - 29	8	27	216
30 - 34	9	32	288
35 - 39	10	37	370
40 - 44	7	42	294
45 - 49	6	47	282
50 - 54	6	52	312
TOTAL	n= 50	$\Sigma(x_i . f_i)$=1.850	

Reemplazando según la fórmula: $X = \Sigma (x_i \, . \, f_i)/n$ **X = 1.850 / 50 = 37 kg.**

Consideraciones especiales:

1.-Cuando una escala de distribución de frecuencias tiene uno ó los dos extremos indeterminados: -20 kgs. y + de 54 kgs., no hay indicación acerca del valor que debe elegirse para representar a la clase para el cálculo de la media. Si supusiéramos que el grupo indeterminado tiene el mismo punto medio que el que le precede o sigue, daría un valor muy bajo, porque se suele dejar una clase abierta por haber unas pocas frecuencias dispersas entre escala de valores amplios.

2.- La media calculada para una distribución de frecuencia asimétrica de intervalos de clase desiguales, es solo una aproximación de la verdadera media.

3.- Si la serie es simétrica, se la puede calcular con bastante exactitud, promediando los valores extremos.

Del ej. anterior: del cuadro Nº 2 los valores extremos de los puntos medios, promediando nos dá : $\frac{20 + 54}{2}$ = 37 kg;

4.- Cuando se presentan valores aberrantes ó insólitos se lo desecha para los cálculos de media porque se considera que proceden debido a factores que lo hacen no comparables a las otras observaciones.

5.- Promedio de porcentaje: cuando se desee promediar porcentajes se deben tomar en cuenta su base como factor de ponderación.

Ejemplo: Personas Tuberculino-positivas por grupos de edad

País............... - Año.........

Edad (años)	Pruebas leídas	Nº Mantoux +	% Mantoux +
1 - 6	**167.715**	**14.591**	**8.7**
7 - 14	**235.695**	**70.708**	**30.0**
15 y +	**288.578**	**173.147**	**60.0**
TOTAL	**691.988**	**258.446**	**37.3**

El promedio de estos porcentajes **NO** es:

$$X = \frac{8.7 + 30.0 + 60.0}{3} = 32.9$$

Debe darse a cada porcentaje el peso correspondiente representado por el número de personas examinadas:

$$\frac{0.087x167.715 + 0.30x235.695 + 0.60x288.578}{691.988} = 37.3$$

Esto resulta correcto cuando lo referimos al número de examinados, pero se dimensiona más si lo queremos referir a todo el universo.

Edad	Población	% +	Nº +
1 - 6	**420.000**	**8.7**	**36.540**
7 - 14	**560.000**	**30.0**	**168.000**
15 y +	**1.020.000**	**60.0**	**612.000**
TOTAL	**2.000.000**	**40.8**	**816.540**

$X = \dfrac{816.540 \text{ x } 100}{2.000.000} = 40.8$ ó lo que es lo mismo;

$$X = \frac{0.087x420.000 + 0.30x560.000 + 0.60x1.020.000}{2.000.000} = \mathbf{40.8}$$

Resultado diferente a 37.3 encontrado sin tener en consideración la población total.

6.- Promedio de promedios: es problema similar:

Egresos y promedios de permanencia por especialidad.

Hospitales del país...........Año ....

Especialidades	Nº egresos	Promedio de estadía (días de internación)
Clínica Médica	**37.000**	**7.3**
Clínica Quirúrgica	**15.000**	**21.0**
Obstetricia	**35.000**	**3.5**
Tuberculosis	**600**	**65.0**
TOTAL	**87.600**	

El promedio general ó promedio de promedios **NO** es:

$$X = \frac{7.3 + 21.0 + 3.5 + 65.0}{4} = 24.2$$

Se lo debe calcular así:

$$X = \frac{7.3x37.000 + 21.0x15.000 + 3.5x35.000 + 65.0x600}{87.600} = 8.5 \text{ días de internación}$$

b) Mediana: Para el cálculo (ver cuadro siguiente)

Distribución de frecuencias de pesos en Kgr
Lugar XX Fecha XX

Peso en Kgs.	Nº de casos (f)	Frecuencia acumulada (Fa)
50 - 54	**1**	**1**
55 - 59	**2**	**3**
60 - 64	**11**	**14**
65 - 69	**10**	**24**
70 - 74	**12**	**36**
75 - 79	**21**	**57**
80 - 84	**6**	**63**
85 - 89	**9**	**72**
90 - 94	**4**	**76**
95 - 99	**4**	**80**
TOTAL	**80**	

1.- Se divide el total de observaciones (N) en dos partes iguales (N/2), en nuestro ejemplo 80/2 = 40, que es el lugar que ocupa la mediana.

2.- Se busca el intervalo al cual pertenece esa observación, para ello se mira la columna de frecuencias acumuladas, se observa corresponde el intervalo 75-79.

3.- Determinar el punto medio del intervalo 75 - 79, que corresponde a la observación de 40 sujetos, por medio de la siguiente fórmula:

$$Me = Li + \frac{(N/2 - Fa)}{fi} \times C$$

Me: mediana; Li: límite inferior del intervalo donde cae la mediana; fi frecuencia del intervalo donde cae la mediana; C: amplitud del intervalo donde cae la mediana.

Donde: Li = 74.5 ; N/2 = 40 ; Fa = 36 ; Fi = 21 ; C = 5

Nótese que en la fórmula la porción de intervalo que se suma a Li está dada por las frecuencias del intervalo que se acumulan hasta la mediana(N/2-Fa) divididas por el total de frecuencias del intervalo multiplicadas por la amplitud del intervalo.

Como la mediana en la observación es 40 y como hay 36 casos en el acumulativo por debajo de 74.5 (verdadero límite inferior del intervalo de clase 75 - 79), para ubicar a la mediana. Considerando que las 21 observaciones del intervalo 75 - 79 están a una misma distancia; se tomará 4/21 de la amplitud de clase (5) y se añadirá a 74.5 que es su verdadero comienzo, con el fin de obtener la mediana, así:

$$Ma = 74.5 + \frac{4}{21} \times 5 = 75.45$$

Es decir que las 21 observaciones están entre los 5 espacios (75 - 79).

Aplicando la fórmula: $$Ma = \frac{Li + N/2 - Fa}{Fi} C$$

$$Ma = 74.5 + \frac{(40 - 36)}{21} \times 5 = 75.45$$

ESCOGENCIA ENTRE EL PROMEDIO, LA MEDIANA Y EL MODO:

El promedio aritmético: tiene la ventaja de tomar en consideración la totalidad de los valores, que a su vez es una desventaja cuando existen valores anormalmente altos ó bajos, pues es influenciado por ellos.

Cuando se presentan valores aberrantes ó insólitos se los desecha para los cálculos de media porque se considera que proceden debido a factores que lo hacen no comparables a las otras observaciones.

Si la serie es simétrica, se la puede calcular con bastante exactitud, promediando los valores extremos (marca de clase, en series agrupadas). Ej. (22 + 52) / 2 = 37.

Si la distribución es simétrica pueden usarse las tres medidas (son parecidas); pero se prefiere la media.

La mediana: Si la distribución es asimétrica la media no es un valor típico y es preferible usar la mediana. En cuanto más sesgados estén los datos, más separadas entre si estarán las 3 medidas de centralización; y siendo la **M_a** menos sensible a valores extremos y atípicos, que la media, refleja mejor que esta, el lugar en donde están la gran masa de valores. Ej. Días de hospitalización de 5 pacientes = 2; 3; 4; 6; 30**. X = 9 días; M_a = 4 días.** Observar los cambios al aumentar o disminuir el sesgo.

Se debe usar la mediana cuando la primera ó última clase no tienen límites precisos (se suele dejar una clase abierta por haber unas pocas frecuencias dispersas entre escala de valores amplios). Del ej. anterior: menos de 25 kg (1ª clase); más de 50 kg (última clase).

La medida de tendencia central a utilizarse dependerá:

- Si la distribución es simétrica pueden usarse indistintamente las tres medidas.
- Si la distribución es asimétrica la media no es un valor típico y es preferible usar la mediana.
- Se debe usar la mediana cuando la primera ó última clase no tienen límites precisos.
- El promedio aritmético tiene la ventaja de tomar en consideración la totalidad de los valores, que a su vez es una desventaja cuando existen valores anormalmente altos ó bajos, pues es influenciado por ellos.
- Por regla general podemos decir que cuando la serie es más ó menos simétrica, el promedio debe ser preferido.

Algunos inconvenientes de la media

La media presenta inconvenientes en algunas situaciones:

- Uno de ellos es que es muy sensible a los valores extremos de la variable: ya que todas las observaciones intervienen en el cálculo de la media, la aparición de una observación extrema, hará que la media se desplace en esa dirección. En consecuencia;
- No es recomendable usar la media como medida central en las distribuciones muy asimétricas;
- Si consideramos una variable discreta, por ejemplo, el N° de hijos el valor de la media puede no pertenecer al conjunto de valores de la variable; Ej. $x = 1,2$ hijos.

Propiedades de la mediana

Entre las propiedades de la mediana, vamos a destacar las siguientes:

- Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas. Por ello es adecuado su uso en distribuciones asimétricas.
- Es de cálculo rápido y de interpretación sencilla.
- A diferencia de la media, la mediana de una variable discreta es siempre un valor de la variable que estudiamos (ej. La mediana de una variable N° de hijos toma siempre valores enteros).

Resumiendo….

La elección de las CC depende:

Modo: Rapidez sobre precisión; de elección con datos nominales o cualitativos; se usa más marketing que en biología.

Mediana: Escalas ordinales; series asimétricas y con datos anómalos; clases abiertas.

Media: Escalas interválicas continuas; en series simétricas. Si es aplicable, es la CC de preferida por:

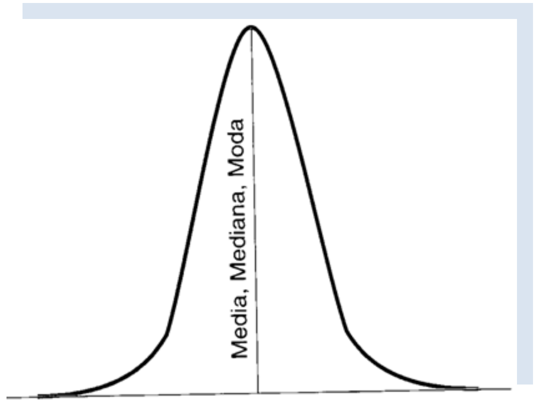
- Tiene en cuenta todos los valores de la serie.
- Es el valor más estable (en series simétricas).
- Su fórmula permite cálculos algebraicos útiles para la estadística inferencial.

Relaciones entre las CC Y MD:

Escala	CC	MD
Nominal	Moda	Rango
Ordinal	Mediana	Rq
Interválica	Media	DE
Proporcional	Media	DE

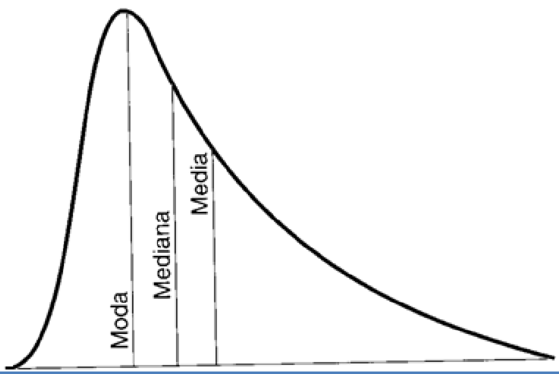
Efectos de la Simetría. Relaciones de las Constantes centrales

La serie es *Simétrica* cuando la proporción de los valores extremos o raros (altos o bajos) son aproximadamente iguales y están a los lados del grupo central.



Serie Simétrica. X = Ma = Mo

La serie es *Asimétrica o sesgada* cuando predominan uno de los valores extremos; si predominan los valores altos (hacia la derecha o +), se llama *asimetría (+).* Si es hacia los valores pequeños se llama *asimetría (-).*
La asimetría atrae el valor de la X.



Asimetría o sesgo (+)
X - Mo = 3 (X - Ma) o, Ma = 2 X + Mo

En este caso el X será mayor que la Ma; la situación se invierte en la asimetría (-)

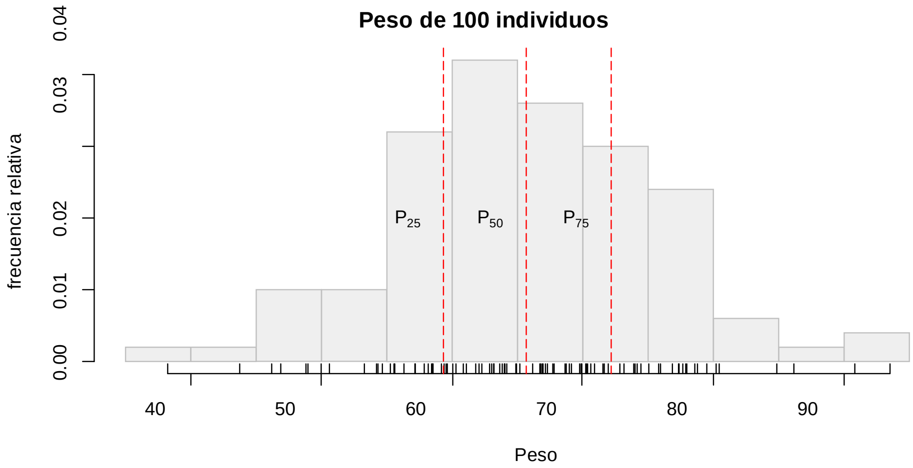
3.3.2. Estadísticos de Posición

Los estadísticos de posición van a ser valores de la variable caracterizados por superar a cierto porcentaje de observaciones en la población (o muestra). Tenemos fundamentalmente a los percentiles como medidas de posición, y asociados a ellos veremos también los cuartiles, deciles y cuartiles.

Percentiles

Para una variable discreta, se define el **percentil de orden** k, como la observación, P_k, que deja por debajo de si el k % de la población. Véase la figura de abajo. Ej.: el 10°Percentil es el valor por debajo del que están el 10 % de las observaciones. Esta definición nos recuerda a la mediana, y por extensión de la definición es evidente que

$$M_{ed} = P_{50}$$



Percentiles 25, 50 y 75 de una variable. Los que se muestran dividen a la muestra en cuatro intervalos iguales, y reciben también el nombre de cuartiles.

Cuartilos

Si los datos de la serie están ordenados según la magnitud, el valor que divide a la serie en dos partes iguales es la mediana. Aquellos valores que dividen la serie en 4 partes lo llamamos “cuartilos” y se simbolizan con:

Q1 = primer cuartil (25° percentil): valor por debajo del que están el 25 % de las observaciones.

Q2 = segundo cuartil (50° percentil): equivale a la mediana

Q3 = tercer cuartil (75° percentil) valor por debajo del que están el 75 % de las observaciones.

Intervalo Intercuartilar: es el comprendido entre el primer cuartil y el tercer cuartil; equivale al 50 % de las observaciones centrales NO AFECTADAS por las fluctuaciones extremas de una serie.

3.3.3. Medidas de Dispersión o Variabilidad

Supongamos 3 grupos de enfermos de 7 sujetos cada uno. Si la duración de la enfermedad en cada sujeto fuera la que aparece a continuación:

Duración de la Enfermedad en tres tipos de Patologías

Enfermedad	Días de hospitalización
Enteritis	1, 3, 5, 7, 9, 11, 13
Bronquitis	1, 2, 3, 7, 11, 12, 13
Conjuntivitis	1, 5, 6, 7, 8, 9, 13

Se constata que:

Las tres series tienen el mismo número de observaciones = 7 pacientes; la misma amplitud = 1 a 13; la misma mediana = 7 días; y el mismo promedio = 7 días.

Sin embargo, son muy distintas. . . Grafiquemos:

Enfermedad	Días de duración de la enfermedad												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Enteritis (Distribución uniforme)	x		x		x		x		x		x		x
Bronquitis (Se agrupan en los extremos)	x	x	x				x				x	x	x
Conjuntivitis (Se agrupan en el centro)	x				x	x	x	x	x				x

Sin embargo son muy distintas porque como vemos en la representación gráfica, la forma en que se agrupan o distribuyen las observaciones en estas tres series son muy distintas entre si:

1.- Los 7 pacientes de la enteritis se distribuyen uniformemente en duración en el lapso de 1 a 13 días.
2.- Los pacientes con bronquitis se agrupan en los extremos.
3.- Los pacientes con conjuntivitis se agrupan hacia el centro.

Esto indica que ante un grupo de observaciones no son suficientes conocer el promedio y su mediana, sino que es necesario tener una medida que indique como se distribuyen las observaciones alrededor de las medidas centrales.
Las medidas de dispersión nos indican si esas puntuaciones o valores están próximas entre sí, o por el contrario están muy dispersas.
Las más utilizadas son:

1. RECORRIDO
   - Recorrido de la variable o Rango
   - Recorrido o intervalo intercuartílico
   - Recorrido interdecílico.
   - Recorrido intercentílico.

2. DESVIACION MEDIA
3. VARIANZA
4. DESVIACION TIPICA O ESTANDAR
5. CUASIVARIANZA

Rango

Se obtiene restando el valor mas bajo de un conjunto de observaciones del valor más alto.

Propiedades del rango

- Es fácil de calcular y sus unidades son las mismas que las de la variable.
- No utiliza todas las observaciones (solo dos de ellas);
- Se puede ver muy afectada por alguna observación extrema;
- El rango aumenta con el número de observaciones, o bien se queda igual. En cualquier caso nunca disminuye.

Se utiliza la desviación estándar para medir la dispersión de las observaciones alrededor de la media aritmética y el intervalo intercuartilar para la mediana.

Recorrido o intervalo intercuartílico (RQ):

Es la diferencia entre el cuartil tercero (Q3) y el primero (Q1); y comprende el 50 % de las observaciones.

$$RQ = Q3 - Q1$$

Cuartil (Qn): Son los valores que dividen las observaciones en 4 partes iguales

x1, x2, x3, . . . x12, x13; x14, x15, . . . x25, x26; x27, x28, . . . xn, xn; xn, xn . . . xn, xn

25 % | 25 % | 25 % | 25 %

Q1 (N/4) Q2 (N/2) Q3 (3N/4)

|_______________RQ_____________|

|________________________________R________________________________|

DESVIACION TIPICA O ESTANDAR - DE (S; δ):

Es la media cuadrática de las desviaciones; o sea, es la raíz cuadrada de la varianza.

$S = \sqrt{S^2}$ o, $S = \sqrt{x_n^2 - X^2}$

Valores sin repetir: $S = \sqrt{\frac{\Sigma d_n^2}{N}} = \sqrt{\frac{\sum x_n^2 - \bar{X}^2}{N}}$

Valores repetidos y/o agrupados: $S = \sqrt{\dfrac{\Sigma d_n^2 f_n}{N}} = \sqrt{\dfrac{\Sigma x_n^2 f_n - X^2}{N}}$

Si los valores de la variable se agrupan en clases, se utilizan las marcas de clase de cada intervalo para la aplicación de la fórmula.

En distribuciones normales unimodales y simétricas; el 68,27 % de los casos está entre el X ±1DE; el 95,45 % entre el X ±2DE; y el 99,73 % de los casos entre X ±3DE.
Es una medida de dispersión que tiene la misma dimensionalidad que las observaciones.

Cálculo de la Desviación Estándar (EN SERIES NO AGRUPADAS)
Tenemos el ejemplo de los 7 pacientes con enteritis:
1; 3; 5; 7; 9; 11; 13 = 49 Promedio 49/7 = 7 días
Se disponen los datos de la siguiente manera:

DURACION DE LA ENFERMEDAD EN 7 PACIENTES CON ENTERITIS
CALCULO DE LA DESVIACION ESTANDAR

PACIENTES (1)	DIAS DE HOSPITAL (2)	DESVIACIONES valor de c/observación menos promedio (3)	DESV. AL CUADR. (4)
Primer	1	1-7 = -6	36
Segundo	2	2-7 = -5	25
Tercer	3	3-7 = -4	16
Cuarto	7	7-7 = 0	0
Quinto	11	11-7= 4	16
Sexto	12	12-7= 5	25
Séptimo	13	13-7= 6	36
TOTAL	**49**	**0**	**154**

X = 49/7 = 7 días	a-) Obtener el promedio (X);
$S^2 = \sum D^2 / n \quad S = \sqrt{S^2}$	b-) Buscar las d_n de cada observación y el X (3);
$S = \sqrt{\sum D^2 / n} = \sqrt{154 / 7}$	c-) Elevar las d_n al cuadrado (4) y su sumatoria, dividirla por las observaciones;
$S = \sqrt{22} = 4{,}7$ días	d) Extraer la raíz cuadrada.

Cuando el grupo de observación es mayor digamos a 30, n queda igual; pero cuando es pequeño como este ejemplo de 7 pacientes, es aconsejable usar el ajuste para muestras pequeñas (n-1); es decir se debe poner 6 en lugar de 7.

DESVIACION ESTANDAR EN SERIES AGRUPADAS

Escolares de acuerdo a su peso. Cálculo de la Desviación Estándar

Peso Kgs.	Nº ind. (f)	Punto medio de clase (x)	$f_1 . x_1$ (2).(3)	Desv. (d_n) x_n -X	Desv.al cuad. (d^2)	$f_1 . d^2$ (2) . (6)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
20 - 24	4	22	88	-15	225	900
25 - 29	8	27	216	-10	100	800
30 - 34	9	32	288	- 5	25	225
35 - 39	10	37	370	0	0	0
40 - 44	7	42	294	5	25	175
45 - 49	6	47	282	10	100	600
50 - 54	6	52	312	15	225	1350
TOTAL	50		1.850			4.050

Los pasos son:

1. Calcular el promedio (X) conforme a lo estudiado.

$$X = \frac{\sum \mathbf{f1 . x1}}{\sum \mathbf{f1}} = \frac{1850}{50} = \mathbf{37\ Kg.}$$

2. Encontrar la diferencia o desvío (5) entre cada uno de los puntos medios de clases (3) y promedio.
3. Elevar al cuadrado la anterior diferencia (6).
4. Multiplicar fila a fila las cifras de las columnas 2 y 6, pues esas diferencias se refieren a un solo individuo.
5. Sumar los productos de (7), para saber la diferencia global entre todos los individuos y su promedio.
6. Dividir la suma anterior por el número de individuos estudiados (n = 50).
7. Extraer la raíz cuadrada, cuyo resultado será la D.E.

$$D.E. = \sqrt{\frac{\sum \mathbf{f1 . D^2}}{\sum \mathbf{f1}}} = D.E. = \sqrt{4050 / 50} = \mathbf{9\ kg.}$$

La expresión matemática de la distribución normal:

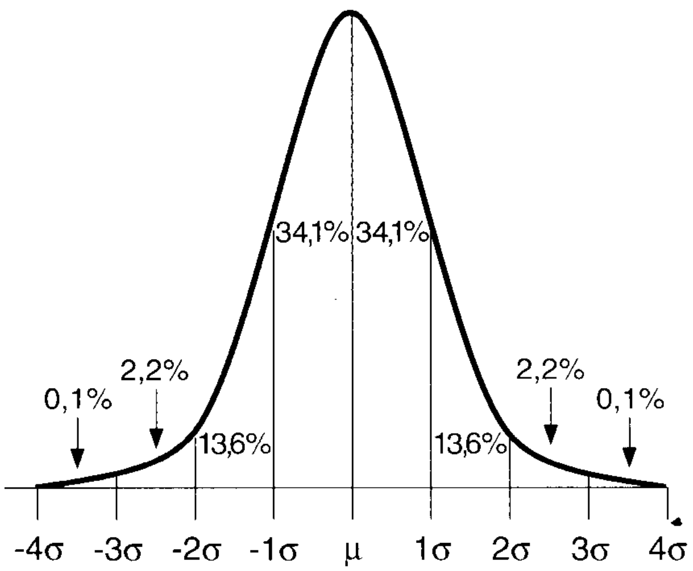
$$\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}$$

Determina un principio estadístico:

X ± 1 D.E. = 68 % del área (68,27 %);

X ± 2 D.E. = 95 % del área (95,45%);

X ± 3 D.E. = 100 % del área (99,73%).



Por tanto a partir de estas medidas de resumen de variables cuantitativas que determinan las bases para un análisis estadístico descriptivo, podemos incursionar en los principios de la inferencia estadística, al determinar intervalos de confianza con una determinada certeza.

Coeficiente de Variación de Pearson (V)

Hemos visto que las medidas de centralización y dispersión nos dan información sobre una muestra. Nos podemos preguntar si tiene sentido usar estas magnitudes para comparar dos poblaciones. Por ejemplo, si nos piden comparar la dispersión de la variable considerada de dos poblaciones diferentes.

¿Qué ocurre si lo que comparamos es la altura con respecto a su peso? Tanto la media como la desviación típica, x y S, se expresan en las mismas unidades que la variable. Por ejemplo, en la variable altura podemos usar como unidad de longitud el metro y en la variable peso, el kilogramo. Comparar una desviación (con respecto a la media) medida en metros con otra en kilogramos no tiene ningún sentido.

El problema no deriva solo de que una de las medidas sea de longitud y la otra sea de masa. El mismo problema se plantea si medimos cierta cantidad, por ejemplo la longitudes/talla, de dos poblaciones, pero con distintas unidades. Este es el caso en que comparamos talla en cm de una población de hombres con el correspondiente en micras de una población de microorganismos.

El problema no se resuelve tomando las mismas escalas para ambas poblaciones; ya que las dispersiones pueden ser falsamente nulas según sea el caso.

En los dos primeros casos mencionados anteriormente, el problema viene de la dimensionalidad de las variables, y en el tercero de la diferencia enorme entre las medias de ambas poblaciones.

El problema se resuelve con las medidas de dispersión relativa, que relacionan una medida de dispersión con su correspondiente medida de centralización. El coeficiente de variación de Pearson (V), es el mejor exponente de este grupo; su potencialidad se basa en, eliminar la dimensionalidad de las variables. Tiene en cuenta la proporción existente entre medias y desviación estándar.

Se define del siguiente modo: $V = \frac{DE}{X} \cdot 100$

Supongamos que queramos comparar rendimientos en relación a indicadores.

Para realizar comparaciones de de observaciones individuales de series diferentes, podemos recurrir al valor tipificado.

Valor tipificado (z): Es el valor que permite expresar un valor inicial (en las unidades originales), en términos de unidades DE

$$\mathbf{z = \frac{(x_1 - \bar{X})}{DE}}$$

$\mathbf{x_1}$ = valor inicial (a investigar); $\mathbf{\bar{X}}$ = promedio; **DE** = desviación estándar

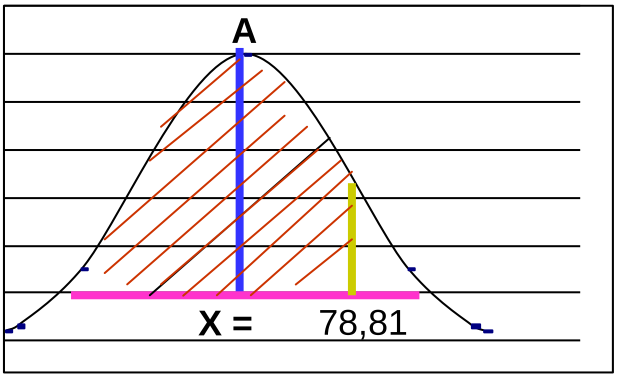
Ejemplo:

Estamos comparando el rendimiento de un indicador de cobertura (en %) en distritos (“a” y “b”) de dos Regiones Sanitarias (“A” y “B”);

“a” tuvo 84 %, el X de la RS”A” = 76 % y un DE = 10 %;

“b” tuvo 90 %, el X de la RS”B” = 82 % y un DE = 16 %.

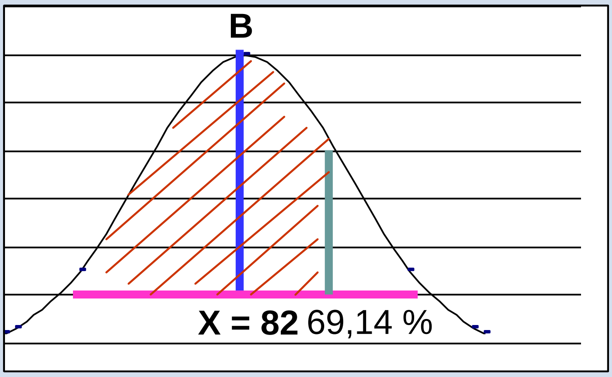
¿Cuál de los dos distritos tuvo un mejor rendimiento?



$Z_A = \frac{x - X}{DE}$ $Z_A = \frac{84 - 76}{10} = 0{,}8$

$Tabla_A$: 0,8 = 0,2881.

50% + 28,81% = 78,81 %



$Z_B = \frac{x - X}{DE}$ $Z_B = \frac{90 - 82}{16} = \mathbf{0,5}$

$Tabla_B$: 0,5 = 0,1914

50 % + 19,14% = **69,14%**

Interpretación: *"a" tuvo un mejor rendimiento más alto, porque está por encima del 78,81% de los distritos de su RS; en cambio, "b" tuvo un rendimiento más alto que sólo el 69,14% de los distritos de su RS. A pesar que "b" aparentemente fue mejor, no posee un rendimiento relativo mayor que "a" porque el rendimiento de la RS fue alto (X= 82).*

3.4. Análisis de Datos de Asociación

Se denomina datos de asociación cuando los elementos se clasifican simultáneamente en dos escalas ya sean cuantitativas o cualitativas ó de ambos.

Ambas escalas cualitativas

Presentación Tabular: en ella tomaremos filas (horizontales) y columnas (verticales), cada una con sus correspondientes totales.
De las dos escalas, generalmente corresponde ubicar en la vertical a lo que tiene más subdivisiones.
Pueden asociarse columnas ó filas con los respectivos porcentajes, en este último caso se debe inscribir el número de observaciones sobre el cual se basan los porcentajes.

CASOS DE TUBERCULOSIS POR FORMAS CLINICAS Y SEXO, PARAGUAY.
AÑO............ POBLACION : ....

Formas clínicas	Sexo		
	Hombres	Mujeres	Ambos Sexos
Pulmonar	741	735	1476
Meníngea	5	3	8
Otras	13	9	22
Total de casos	759	747	1506

Presentación gráfica: se la puede graficar en cualquiera de las formas de barras.

<u>Análisis</u>: relacionar en % los grupos entre sí ó cada grupo en sí con la población correspondiente.

El 98 % de los casos son pulmonares: $\frac{1476}{1506} \times 100 = 98\ \%$

El 0.5 % de todos los casos son meníngeos: $\frac{8}{1506} \times 100 = 0.5\ \%$

El 1.5 % de todos los casos tienen otras localizaciones: $\frac{22}{1506} \times 100 = 1.5\ \%$

El 50.4 % de todos los casos son del sexo masculino: $\frac{759}{1506} \times 100 = 50.4$

El 50.2 % de todos los casos pulmonares son diagnosticados en el sexo masculino: $\frac{741}{1476} \times 100 = 50.2\ \%$

Una Escala Cuantitativa y otra Cualitativa

Presentación tabular: la escala cuantitativa se pone en la primera columna que corresponde al inicio de las filas, salvo que la escala cualitativa tenga muchas subdivisiones, y la escala cualitativa en los encabezamientos de las siguientes columnas.

Peso al nacer de 180 niños según sexo

	Sexo		
Peso en gramos	Hombres	Mujeres	Total
2.000 - 2.499	8	12	20
2.500 - 2.999	20	25	45
3.000 - 3.499	50	35	85
3.500 - 3.999	16	14	30
Total	94	86	180

Presentación gráfica: se utiliza el polígono de frecuencia. El diagrama puede trazarse en papel de pautado aritmético, semilogarítmico ó logarítmico.

Análisis: por medio de porcentajes, tasas, promedios, mediana, moda, desviación estándar, cuartilos, según la necesidad del caso. Medidas estas que nos permitirán apreciar variaciones de los fenómenos en estudio.

Si no se está seguro de cual de ellos es el indicado, utilizar las tasas y porcentajes.

3.4.1 Ambas escalas cuantitativas

En estos casos se investigan en cada individuo dos variables cuantitativas diferentes como ser edad y peso, estatura y peso.

Al hablar de correlación y regresión simple nos referimos al análisis bivariante (dos variables); y al usar la palabra "lineal" expresamos el concepto de que la relación existente es de este tipo.

De esto intuimos que se pueden calcular relaciones de más de 2 variables (análisis multivariante); y que pueden existir otro tipos de relaciones: exponencial, parabólicas, lógicas, etc.

Aspectos Generales:

El análisis presenta dos aspectos diferentes:

Correlación: Si las dos variables están asociadas (como edad y peso) y medir hasta que punto los cambios en una pueden explicarse por los cambios que ocurren en otra.

Regresión: Cuando conocida una correlación se la cuantifica con el fin de calcular los valores de la otra variable conociendo una de ellas (la escala de peso depende de la edad).

El peso tiene una relación con la edad, es decir la escala peso depende de la escala edad.

Al subir ó bajar la edad, el peso tiende a subir ó bajar. A la escala edad se la denomina escala independiente (x), y a la escala peso: escala dependiente (y).

A veces no puede decirse que ciertos valores determinan los otros, sino que unos y otros varían conjuntamente. Las relaciones tienen la misma tendencia, aunque independientemente la una de la otra; existiendo ó no una causa común para ello, (aumento de pulsaciones y aumento de respiraciones).

En caso de no existir un trasfondo común es pura coincidencia, como sería el aumento de embarazos al aumentar el consumo de papas.

La aplicación de un método no excluye necesariamente al otro. Al estudiar la variación de cualquier característica debemos averiguar primero cuales factores pueden explicar dichas variaciones (correlación) y una vez identificado dedicarnos a medir en que grados los cambios que ellos experimentan afectan a la característica que nos interesa (regresión)

PRESENTACION TABULAR:

Si son pocos los elementos observados, especificar los valores observados al lado de cada elemento; si son numerosos se deberán agrupar en clases.

PRESENTACION GRAFICA:

Se hace mediante el gráfico de puntos.

Coeficiente de Regresión

a) Cálculo del coeficiente de regresión.

b) La línea de regresión.

c) Limitaciones en la utilización.

Para hallar el coeficiente de regresión se hace primero un cuadro, luego el gráfico que le corresponde, poniendo los valores de la variable independiente en la abscisa.

Método predictivo para calcular una variable a partir del dato conocido de otra variable con la que guarda correlación. El cálculo se basa en la ecuación de la línea o recta de regresión.

Línea de Regresión (LR): Es la recta que pasando por los datos minimiza la suma de los cuadrados de las desviaciones de los datos reales con relación a su ajuste correspondiente (minimiza la varianza); el método utilizado para calcularla se llama: Análisis de la Regresión o Método los Mínimos Cuadrados.

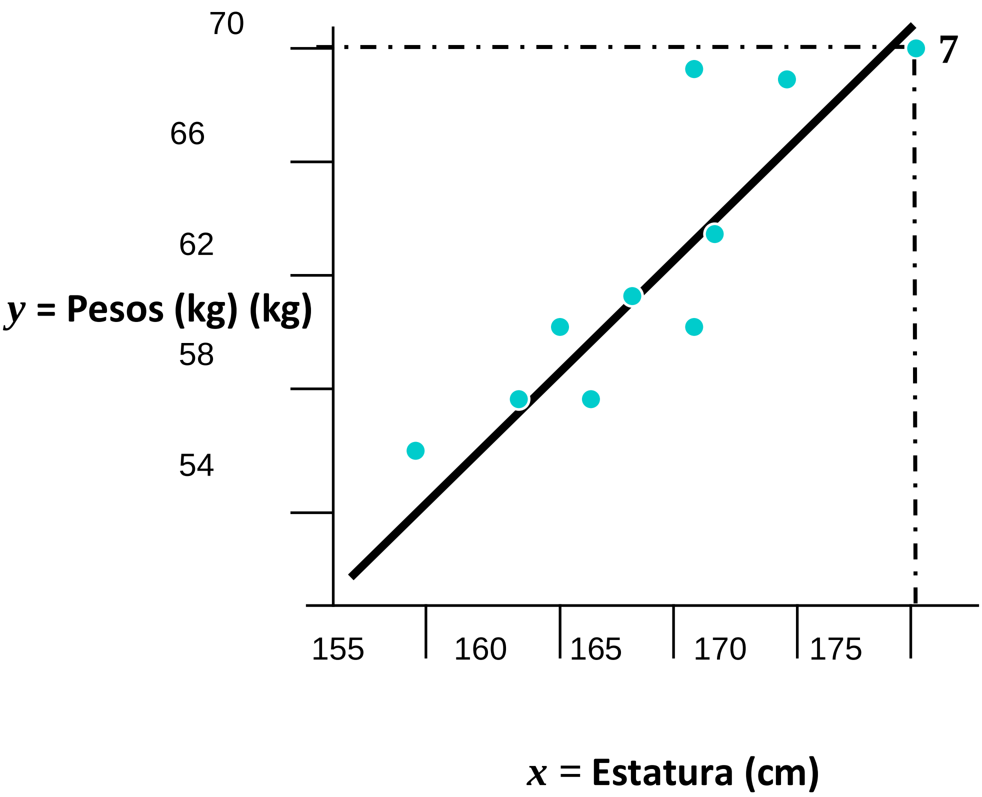
La LR pasa por las medias de la variable dependiente; y es la que mejor explica la ecuación: **y = a + b.x**

ESTATURA Y PESO DE LOS ALUMNOS DE BIOSESTADISTICA – FACULTAD….. – AÑO….

Alumnos	Estatura en centímetros	Peso en Kgm.
1	162	58
2	158	58
3	155	56
4	162	60
5	170	68
6	160	61
7	175	70
8	165	60
9	168	64
10	165	69

Diagrama de dispersión

Estatura y peso de los alumnos del curso de Estadística



La relación entre las variables (estatura, peso) puede representarse por la línea recta trazada en el gráfico.

Mientras más tiendan los puntos a caer sobre la línea de regresión, más estrecha es la relación entre las dos variables.

Si el fenómeno en estudio puede ser resumido por una línea recta, podemos resumir matemáticamente dicha relación, con lo cual nos será posible predecir los valores de la escala dependiente cuando se conocen los de la escala independiente.

Coeficiente de regresión (b) indica que los valores de la escala dependiente cambian "b" unidades por cada unidad que cambian los valores de la v. independiente. En el ejemplo presente encontramos que b=0,8, lo cual significa que por cada aumento de 1cm en la estatura se observa un aumento de 0,8 Kg. de peso.

Coeficiente de regresión puede ser positivo o negativo. Si es positivo, o aumentan o disminuyen a la vez. Si es negativo quiere decir que cuando una variable aumenta la otra disminuye, o viceversa.

Si el coeficiente de regresión fuera 0, para cualquier valor de la escala independiente, habría siempre el mismo valor en la escala dependiente.

Cálculo del coeficiente de regresión

Estatura y peso de un grupo de pacientes

(f)	Talla (x)	Peso (y)	Talla dx	Peso dy	dx2	dy2	(dx)(dy)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	162	58	-2	-4	4	16	8
2	158	54	-6	-8	36	64	48
3	155	56	-9	-6	81	36	54
4	162	60	-2	-2	4	4	4
5	170	68	6	6	36	36	36
6	160	61	-4	-1	16	1	4
7	175	70	11	8	121	64	88
8	165	60	1	-2	1	4	-2
9	168	64	4	2	16	4	8
10	165	69	1	7	1	49	7
Sumas	1.640	0	0		316	278	255
Prom.	164	62					

a) Cálculo del coeficiente de regresión

Que el cálculo del coeficiente de regresión es fácil, aunque las operaciones son muy laboriosas.

Los pasos para el cálculo del coeficiente de regresión son:

1. Disponer los datos como en el cuadro anterior.
2. Obtener el promedio para cada una de las variables en estudio.
3. Ver cuánto difiere cada observación de su promedio respectivo. Estas desviaciones se obtendrán primero para una variable y luego para la otra y se anotarán respectivamente en las columnas (4) y (5) del cuadro, teniendo cuidado de indicar si son positivas o negativas.

4. Elevar al cuadrado las anteriores desviaciones, y situándolas en las columnas respectivas (6) y (7).
5. Totalizar las dos columnas de cuadrados.
6. Buscar el producto de las desviaciones obtenidas en el punto (3), para lo cual se multiplicaron renglón a renglón las columnas (4) y (5) conservando los signos algebraicamente. Totalizar luego estos productos.
7. Calcular el coeficiente de regresión, para lo cual se divide el total de la columna de productos por el total de la columna dx2 (desviaciones cuadrados de los valores independientes).

$$b = \frac{\sum dx \,.\, Dy}{\sum dx2} = \frac{255}{316} = 0{,}80 \text{ kilos}$$

b) Línea de regresión (LR)

La línea recta trazada en el gráfico puede usarse para predecir el peso de cualquier individuo cuya estatura se conozca. El gráfico anterior presenta una inclinación de 0,8 kilos de peso por cada centímetro de estatura.

La ecuación de la LR: $y = a + b.x$; donde $a = \bar{Y} - b.\bar{X}$

$$y = (\bar{Y} - b \,.\, \bar{X}) + b \,.\, x$$

Donde: y = valor de la variable dependiente a conocer
Y ; X = promedio de las variables
b y x = coeficiente de regresión y valor de la variable independiente respectivamente.

<u>Ejemplo</u>: se quiere conocer el peso (Y) de un individuo que mide 180 centímetros de estatura (X).
Mediante el cuadro 176 sabemos que:
Y = 62
X = 164
B = 0,8 kilos
X = 180, es la estatura del individuo cuyo correspondiente peso (Y) que queremos saber:
Reemplazando: Y_{180} = 62- (0,8 x 164) + (0,8 x 180)
Calcular algebraicamente, y poner correctamente los signos:
Y_{180} = 62- 131,2 + 144,0 = 74,8 kilos.

En la misma forma, si se quiere averiguar el peso de un individuo de 158 centímetros de estatura:
Y_{158} = 62- (0,8x164) + (0,8x158) = 57,2 kilos.

Limitaciones para su uso:

- El coeficiente de regresión se calcula únicamente si el gráfico demuestra que la relación estudiada es lineal.
- El coeficiente no puede aplicarse para estimar valores que exceden los límites de la serie donde fue calculado. Así será absurdo pretender calcular por este procedimiento la estatura de una persona de 60 años.
- Si el coeficiente de regresión (b) tiene valor positivo significa que las dos variables aumentan o disminuyen al mismo tiempo.
- Si es negativo: cuando uno aumenta el otro disminuye, o viceversa.
- Si es cero, para cada valor de la escala independiente corresponde cualquier valor en la escala dependiente:

Coeficiente de correlación: "r"

En un problema de correlación no puede decirse que una variable sea dependiente y otra independiente.

Tenemos que hallar el coeficiente de regresión b´, que representará la variación de "y" con respecto a "x" y el coeficiente de regresión b´´, que representará la variación de "x" con respecto a "y".

Con el fin de reducir las dos constantes a una sola, se utiliza el coeficiente de correlación "r", que es igual a la raíz cuadrada del producto de los dos coeficientes de regresión mencionados.

$$\mathbf{r = \sqrt{b' \times b''}}$$

Coeficiente de correlación

a) Cálculo del coeficiente de correlación
b) Valores de "r"
c) Interpretación del "r".

a) **Cálculo del coeficiente de correlación**

El cálculo del coeficiente de correlación se hace fácilmente con los datos del cuadro precedente.

Si se considera primero que la estatura, es la variable independiente, entonces, conforme ya se ha visto:

$$b' = \frac{\sum dx \cdot dy}{\sum dx^2} = \frac{255}{316} = 0{,}80 \text{ kilos}$$

Y si se considera luego que el peso es la variable dependiente, entonces se calculará b´´, para lo cual sólo variará el denominador del quebrado:

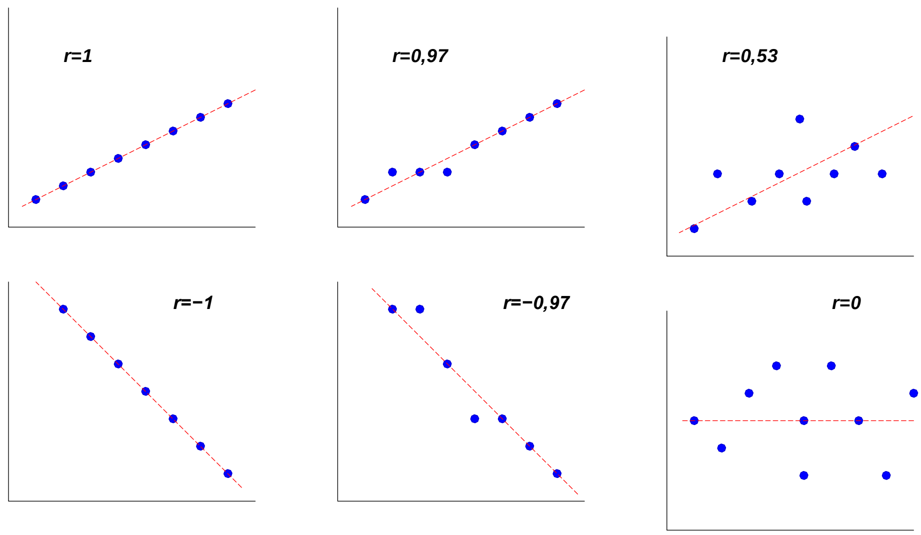
$$b'' = \frac{\sum dx \cdot dy}{\sum dy^2} = \frac{255}{278} = 0{,}91 \text{ cm.}$$

Por lo tanto:

r $= \sqrt{b' \cdot b''} = \sqrt{0{,}80 \times 0{,}91} =$ **0,86**

b) Valores de r

- Puede valer entre -1 y +1
- Valores positivos indican que ambas variables aumentan o disminuyen al mismo tiempo.
- Valores negativos indican que cuando uno aumentan, el otro disminuye, o viceversa.
- Si r es exactamente igual a –1 o +1, quiere decir que hay una perfecta asociación entre las dos variable, en el sentido de que por cada unidad que aumenta o disminuye una variable, la otra cambia siempre igual número de unidades. En dichas ocasiones los puntos en el gráfico caerían todos sobre una línea recta.
- Si r = 0, significa que no hay ninguna asociación entre las dos variables, o que de existir, no es una relación lineal.



Interpretación de la r

r = ±1 es lo mismo que decir que las observaciones de ambas variables están perfectamente alineadas. El signo de r nos indica el crecimiento o decrecimiento de la recta. La relación lineal es tanto más perfecta cuanto **r** está cercano a ±1.

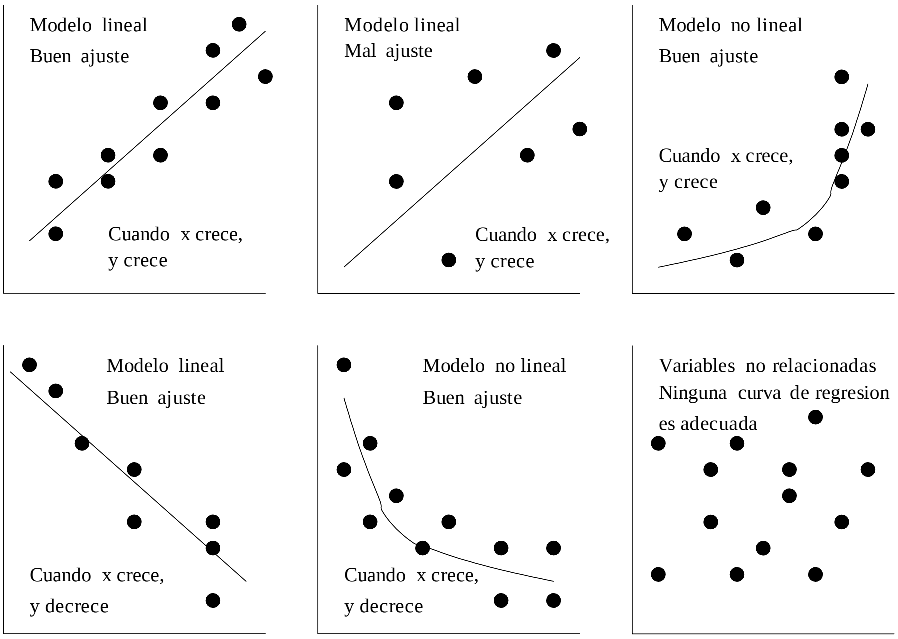
c) **Interpretación de r**

- En el ejemplo que nos sirvió para el cálculo del coeficiente de correlación encontramos que r = 0,86 Este valor no debe interpretarse como se hace corriente mente, en el sentido de que el 86% de las variaciones en el peso son “causadas” por las variaciones en la Talla.

- Para una interpretación más correcta debe tomarse el *coeficiente de determinación (r^2),* que expresa la parte proporcional de variancia de la variable dependiente relacionada con la v. independiente. El valor equivale al cuadrado del coeficiente de correlación.
  Por ej. r^2= (0,86)2 = 0,74, 2 indicar que un 74% de los cambios en el peso se explican por las variaciones de la estatura.

$$r2 = \frac{\text{variación explicada}}{\text{variación total}}$$

El coeficiente determinación r^2 es la expresión de que tan bueno es el método para explicar la influencia de la variable independencia sobre la dependiente, esto se conoce como *bondad del ajuste*.



Bondad de Ajuste del modelo lineal

- “r” menores de 0,50 son por lo general difíciles de interpretar.
- Se debe evaluar la significancia estadística en las tablas que muestran los valores críticos[6] que debe llegar la “r” de Pearson[7] para alcanzar la significancia estadística según el grado de libertad n-2.

[6] Ver Anexos

Recomendaciones finales

Al interpretar las asociaciones encontradas entre dos variables, debe tenerse en cuenta, finalmente, los siguientes principios.

a) *Correlación no es sinónimo de casualidad*, que una variable aumente o disminuya al mismo tiempo que otra, no quiere decir que los cambios en una son determinados por la otra.

b) Las asociaciones encontradas no deben aplicarse indebidamente a valores que excedan los límites de las series estudiadas.

c) El cálculo del coeficiente de correlación r sólo se justifica cuando el gráfico señala que la asociación estudiada puede ser convenientemente resumida por una línea recta.

d) La interpretación del coeficiente de correlación se hará teniendo en cuenta el número de individuos estudiados.

e) Cuando se estudia un grupo grande de individuos es necesario agrupar los datos en un número conveniente de clases, con el fin de facilitar la computación, las cuales sin embargo siguen siendo muy laboriosas y se prestan a múltiples equivocaciones cuando se carece de práctica.

f) Los métodos de correlación y regresión son métodos paramétricos de análisis por lo tanto potentes, pero que deben reunir requisitos muy estrictos para ser aplicados:

Supuestos de Correlación:

1. La metodología de muestreo para la selección de pares debe ser aleatorio e independiente.
2. Las variables "X" e "Y" deben tener una distribución normal o gaussiana
3. Debe existir una relación lineal.
4. Homocedasticidad.
5. Ambas escalas deben ser cuantitativas continuas (condición relativa).

3.5. Análisis de tendencias en series cronológicas

Líneas De Regresión En Series Cronológicas: (Válido si los datos siguen una línea recta) Dará una mejor estimación del cambio anual del fenómeno al tener en cuenta todos los valores de la serie. Se debe calcular el coeficiente de regresión (b).

Desviación Desviación 2

Años Años Tasas Año Tasa Año Tasa producto

	X	y	dx	dy	dx2	dy2	(dx) (dy)
1966	1	15,2	-2	-0,2	4	0,04	0,4
1967	2	15,3	-1	-0,1	1	0,01	0,1
1968	3	15,4	0	0	0	0	0,0
1969	4	15,5	+1	+0,1	1	0,01	0,1
1970	5	15,6	+2	+0,2	4	0,04	0,4
Suma	15	77,0	0	0	10	0,10	1,0
Prom.	3	15,4					

CR (b) = $\frac{\sum (dx)\,(dy)}{\sum dx2}$ = 1/10 = 0,1; aplicando la fórmula $y = (\bar{y} - b.\bar{x}) + b.x$

$\bar{y}$ = promedio de la variable dependiente; .$\bar{x}$ = promedio de la variable independiente)
b = CR; x = valor de conocido de la variable independiente que se relaciona o influye en el valor de y (incógnita o valor desconocido de la variable dependiente que se desea estimar)

Para saber la tasa en 1971 (6to. Año de la serie): x = 6.

Y = 15,4 – (0,1 x 3) + (0,1 x 6) = 15,7

4. Análisis estadístico aplicando las funciones estadísticas de Excel

Hasta aquí se ha presentado la forma clásica del uso de los distintos métodos de análisis descriptivo de la información; resulta además didáctico para explicar aspectos conceptuales. No obstante, la informática aplicada resulta ser una herramienta de apoyo importantísimo, dado la simplificación y celeridad en lo que hace a la estimación propiamente dicha; por supuesto que tanto uno como la otra necesita del irremplazable concurso de la interpretación humana.

A continuación se presenta una descripción de algunas de las funciones estadísticas que nos presenta Excel para analizar la información; y luego se desarrolla una suerte de"paso a paso" sobre el uso de las distintas funciones estadísticas para el análisis de la información.

Función PROMEDIO

Este artículo describe la sintaxis de la fórmula y el uso de la **función** PROMEDIO en Microsoft Office Excel.

Descripción

Devuelve el promedio (media aritmética) de los argumentos. Por ejemplo, si el rango A1:A20 contiene números, la fórmula **=PROMEDIO (A1:A20)** devuelve el promedio de dichos números.

Sintaxis

```
AVERAGE (number1, [number2],...)
```

La sintaxis de la función **PROMEDIO** tiene los siguientes argumentos:

- **número1** Obligatorio. El primer número, referencia de celda o rango para el que desea el promedio.
- **número2, ...** Opcional. Números, referencias de celda o rangos adicionales para los que desea el promedio, hasta un máximo de 255.

Observaciones

- Los argumentos pueden ser números o nombres, rangos o referencias de celda que contengan números.

- Se tienen en cuenta los valores lógicos y las representaciones textuales de números escritos directamente en la lista de argumentos.
- Si el argumento de un rango o celda de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.
- Los argumentos que sean valores de error o texto que no se pueda traducir a números provocan errores.
- Si desea incluir valores lógicos y representaciones textuales de números en una referencia como parte del cálculo, utilice la función **PROMEDIOA**.
- Si desea calcular el promedio de sólo los valores que cumplen ciertos criterios, use la función **PROMEDIO.SI** o la función **PROMEDIO.SI.CONJUNTO**.

NOTA La función **PROMEDIO** mide la tendencia central, que es la ubicación del centro de un grupo de números en una distribución estadística. Las tres medidas más comunes de tendencia central son las siguientes:

- **Promedio**, que es la media aritmética y se calcula sumando un grupo de números y dividiendo a continuación por el recuento de dichos números. Por ejemplo, el promedio de 2, 3, 3, 5, 7 y 10 es 30 dividido por 6, que es 5.
- **Mediana**, que es el número intermedio de un grupo de números; es decir, la mitad de los números son superiores a la mediana y la mitad de los números tienen valores menores que la mediana. Por ejemplo, la mediana de 2, 3, 3, 5, 7 y 10 es 4.
- **Moda**, que es el número que aparece más frecuentemente en un grupo de números. Por ejemplo, la moda de 2, 3, 3, 5, 7 y 10 es 3.

Para una distribución simétrica de un grupo de números, estas tres medidas de tendencia central son iguales. Para una distribución sesgada de un grupo de números, las medidas pueden ser distintas.

Sugerencia Cuando esté calculando el promedio de celdas, tenga en cuenta la diferencia existente entre las celdas vacías y las que contienen el valor cero, especialmente si ha desactivado la casilla **Mostrar un cero en celdas que tienen un**

valor cero en el cuadro de diálogo **Opciones de Excel**. Cuando esta opción está seleccionada, las celdas vacías no se tienen en cuenta, pero sí los valores cero.

Para encontrar la casilla **Mostrar un cero en celdas que tienen un valor cero**:

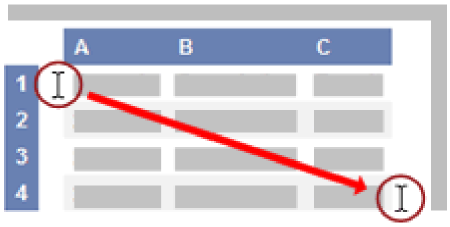
- Haga clic en el **botón de Microsoft Office** , haga clic en **Opciones de Excel** y, a continuación, en la categoría **Avanzadas**, vea **Mostrar opciones para esta hoja**.

Ejemplo El ejemplo será más fácil de entender si lo copia en una hoja de cálculo en blanco.

⊞ ¿Cómo copio un ejemplo?

1. Seleccione el ejemplo de este artículo.

IMPORTANTE No seleccione los encabezados de columna o de fila.



Seleccionar un ejemplo de la Ayuda

2. Presione CTRL+C.
3. En Excel, cree una hoja de cálculo o un libro en blanco.
4. En la hoja de cálculo, seleccione la celda A1 y presione CTRL+V.

IMPORTANTE Para que el ejemplo funcione correctamente, debe pegarlo en la celda A1 de la hoja de cálculo.

5. Para cambiar entre ver los resultados y ver las fórmulas que devuelven los resultados, presione Alt+º, o en la ficha **Fórmulas**, en el grupo **Auditoría de fórmulas**, haga clic en el botón **Mostrar fórmulas**.

	A	B	C
1	**Datos**		
2	10	15	32
3	7		
4	9		
5	27		
6	2		
7	**Fórmula**	**Descripción**	**Resultado**
8	=PROMEDIO(A2:A6)	Promedio de los números en las celdas A2 a A6.	11
9	=PROMEDIO(A2:A6;5)	Promedio de los números en las celdas A2 a A6 y el número 5.	10
10	=PROMEDIO(A2:C2)	Promedio de los números en las celdas A2 a C2.	19

Función MEDIANA

Devuelve la mediana de los números dados. La mediana es el número que se encuentra en medio de un conjunto de números.

Sintaxis

MEDIANA (**número1**; número2; ...)

Número1, número2... son de 1 a 255 números cuya mediana desea obtener.

Observaciones

- Si la cantidad de números en el conjunto es par[8], MEDIANA calcula el promedio de los números centrales. Vea la segunda fórmula del ejemplo.
- Los argumentos pueden ser números, o nombres, matrices o referencias que contengan números.
- Se tienen en cuenta los valores lógicos y las representaciones textuales de números escritos directamente en la lista de argumentos.
- Si el argumento matricial o de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.
- Los argumentos que sean valores de error o texto que no se pueda traducir a números provocan errores.

NOTA La función MEDIANA mide la tendencia central, que es la ubicación del centro de un grupo de números en una distribución estadística. Las tres medidas más comunes de tendencia central son las siguientes:

- **Promedio** Es la media aritmética y se calcula sumando un grupo de números y dividiendo a continuación por el recuento de dichos números. Por ejemplo, el promedio de 2, 3, 3, 5, 7 y 10 es 30 dividido por 6, que es 5.

[8] Si la frecuencia de la serie es impar, el valor central que divide la serie ya ordenada de manera creciente, y que deja igual cantidad de valores a cada lado, es la MEDIANA.

- **Mediana** Es el número intermedio de un grupo de números; es decir, la mitad de los números son superiores a la mediana y la mitad de los números tienen valores menores que la mediana. Por ejemplo, la mediana de 2, 3, 3, 5, 7 y 10 es 4.
- **Moda** Es el número que aparece más frecuentemente en un grupo de números. Por ejemplo, la moda de 2, 3, 3, 5, 7 y 10 es 3.

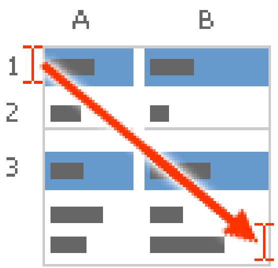
Para una distribución simétrica de un grupo de números, estas tres medidas de tendencia central son iguales. Para una distribución sesgada de un grupo de números, las medidas pueden ser distintas.

Ejemplo El ejemplo será más fácil de entender si lo copia a una hoja de cálculo en blanco.

⊞ Cómo copiar un ejemplo

1. Cree una hoja de cálculo o un libro en blanco.
2. Seleccione el ejemplo en el tema de Ayuda.

NOTA No seleccione los encabezados de columna o de fila.



Seleccionar un ejemplo de la Ayuda

3. Presione CTRL+C.
4. En la hoja de cálculo, seleccione la celda A1 y presione CTRL+V.
5. Para cambiar entre ver los resultados y ver las fórmulas que devuelven los resultados, presione CTRL+` (acento grave), o en el grupo **Auditoría de fórmulas** de la ficha **Fórmulas**, haga clic en el botón **Mostrar fórmulas**.

	A
1	**Datos**
2	1
3	2
4	3
5	4
6	5
7	6

Fórmula	Descripción (resultado)
=MEDIANA(A2:A6)	La mediana de los 5 primeros números de la lista anterior (3)
=MEDIANA(A2:A7)	La mediana de todos los números anteriores, o el promedio de 3 y 4 (3,5)

Función MODA

Devuelve el valor que se repite con más frecuencia en una matriz o rango de datos.

Sintaxis

MODA (**número1**; número2; ...)

Número1, número2... son de 1 a 255 argumentos cuya moda desea calcular. También puede utilizar una matriz única o una referencia matricial en lugar de argumentos separados con punto y coma.

Observaciones

- Los argumentos pueden ser números, o nombres, matrices o referencias que contengan números.

- Si el argumento matricial o de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.
- Los argumentos que sean valores de error o texto que no se pueda traducir a números provocan errores.
- Si el conjunto de datos no contiene puntos de datos duplicados, MODA devuelve el valor de error #N/A.

NOTA La tendencia central de medidas de la función MODA, que es la ubicación del centro de un grupo de números en una distribución estadística. Las tres medidas más comunes de tendencia central son las siguientes:

- **Promedio** Es la media aritmética y se calcula sumando un grupo de números y dividiendo a continuación por el recuento de dichos números. Por ejemplo, el promedio de 2, 3, 3, 5, 7 y 10 es 30 dividido por 6, que es 5.
- **Mediana** Es el número intermedio de un grupo de números; es decir, la mitad de los números son superiores a la mediana y la mitad de los números tienen valores menores que la mediana. Por ejemplo, la mediana de 2, 3, 3, 5, 7 y 10 es 4.
- **Moda** Es el número que aparece más frecuentemente en un grupo de números. Por ejemplo, la moda de 2, 3, 3, 5, 7 y 10 es 3.

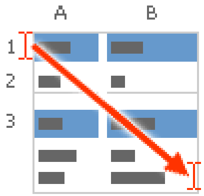
Para una distribución simétrica de un grupo de números, estas tres medidas de tendencia central son iguales. Para una distribución sesgada de un grupo de números, las medidas pueden ser distintas.

Ejemplo El ejemplo será más fácil de entender si lo copia a una hoja de cálculo en blanco.

⊞ Cómo copiar un ejemplo

1. Cree una hoja de cálculo o un libro en blanco.
2. Seleccione el ejemplo en el tema de Ayuda.

    NOTA No seleccione los encabezados de columna o de fila.



Seleccionar un ejemplo de la Ayuda

3. Presione CTRL+C.
4. En la hoja de cálculo, seleccione la celda A1 y presione CTRL+V.
5. Para cambiar entre ver los resultados y ver las fórmulas que devuelven los resultados, presione CTRL+` (acento grave), o en el grupo **Auditoría de fórmulas** de la ficha **Fórmulas**, haga clic en el botón **Mostrar fórmulas**.

	A	
1	**Datos**	
2	5,6	
3	4	
4	4	
5	3	
6	2	
7	4	
	Fórmula	**Descripción (resultado)**
	=MODA(A2:A7)	Moda, es decir, número que se repite con más frecuencia (4)

Función PERCENTIL

Devuelve el k-ésimo percentil de los valores de un rango. Esta función permite establecer un umbral de aceptación. Por ejemplo, podrá examinar a los candidatos cuya calificación sea superior al nonagésimo percentil.

Sintaxis

PERCENTIL (**matriz**;**k**)

Matriz es la matriz o rango de datos que define la posición relativa.

K es el valor de percentil en el intervalo de 0 a 1, inclusive.

Observaciones

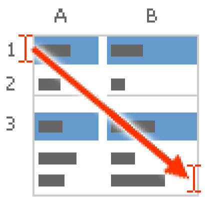
- Si el argumento matriz está vacío o contiene más de 8.191 puntos de datos, PERCENTIL devuelve el valor de error #¡NUM!
- Si el argumento k no es numérico, PERCENTIL devuelve el valor de error #¡VALOR!
- Si k es < 0 o si k > 1, PERCENTIL devuelve el valor de error #¡NUM!
- Si k no es un múltiplo de 1/(n - 1), PERCENTIL interpola para determinar el valor en el k-ésimo percentil.

Ejemplo El ejemplo será más fácil de entender si lo copia a una hoja de cálculo en blanco.

⊞ Cómo copiar un ejemplo

1. Cree una hoja de cálculo o un libro en blanco.
2. Seleccione el ejemplo en el tema de Ayuda.

NOTA No seleccione los encabezados de columna o de fila.



Seleccionar un ejemplo de la Ayuda

3. Presione CTRL+C.
4. En la hoja de cálculo, seleccione la celda A1 y presione CTRL+V.

5. Para cambiar entre ver los resultados y ver las fórmulas que devuelven los resultados, presione CTRL+` (acento grave), o en el grupo **Auditoría de fórmulas** de la ficha **Fórmulas**, haga clic en el botón **Mostrar fórmulas**.

	A
1	**Datos**
2	1
3	3
4	2
5	4

Fórmula	**Descripción (resultado)**
=PERCENTIL(A2:A5;0,3)	El percentil 30 de la lista anterior (1,9)

Función DESVEST

Calcula la desviación estándar de una muestra. La desviación estándar es la medida de la dispersión de los valores respecto a la media (valor promedio).

Sintaxis

DESVEST (**número1**; número2; ...)

Número1, número2,... son de 1 a 255 argumentos numéricos correspondientes a una muestra de una población. También puede utilizar una matriz única o una referencia matricial en lugar de argumentos separados con punto y coma.

Observaciones

- DESVEST parte de la hipótesis de que los argumentos representan la muestra de una población. Si sus datos representan la población total, utilice DESVESTP para calcular la desviación estándar.

- La desviación estándar se calcula utilizando el método "n-1".
- Los argumentos pueden ser números, o nombres, matrices o referencias que contengan números.
- Se tienen en cuenta los valores lógicos y las representaciones textuales de números escritos directamente en la lista de argumentos.
- Si un argumento es una matriz o una referencia, sólo se considerarán los números de esa matriz o referencia. Se pasan por alto las celdas vacías, valores lógicos, texto o valores de error de la matriz o de la referencia.
- Los argumentos que sean valores de error o texto que no se pueda traducir a números provocan errores.
- Si desea incluir valores lógicos y representaciones textuales de números en una referencia como parte del cálculo, utilice la función DESVESTA.
- DESVEST utiliza la fórmula siguiente:

$$\sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}}$$

donde x es la media de muestra PROMEDIO (número1;número2;…) y n es el tamaño de la muestra.

Ejemplo

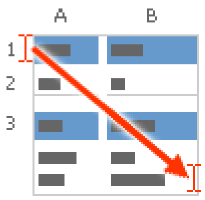
Supongamos que 10 herramientas forjadas en la misma máquina durante el mismo proceso de producción son elegidas como una muestra aleatoria y medimos su resistencia a la ruptura.

El ejemplo será más fácil de entender si lo copia a una hoja de cálculo en blanco.

⊞ Cómo copiar un ejemplo

1. Cree una hoja de cálculo o un libro en blanco.
2. Seleccione el ejemplo en el tema de Ayuda.

NOTA No seleccione los encabezados de columna o de fila.



Seleccionar un ejemplo de la Ayuda

3. Presione CTRL+C.

4. En la hoja de cálculo, seleccione la celda A1 y presione CTRL+V.

5. Para cambiar entre ver los resultados y ver las fórmulas que devuelven los resultados, presione CTRL+` (acento grave), o en el grupo **Auditoría de fórmulas** de la ficha **Fórmulas**, haga clic en el botón **Mostrar fórmulas**.

	A
1	**Resistencia**
2	1345
3	1301
4	1368
5	1322
6	1310
7	1370
8	1318
9	1350
10	1303
11	1299

Fórmula	**Descripción (resultado)**
=DESVEST(A2:A11)	Desviación estándar de la resistencia a la rotura (27,46391572)

Función COEF.DE.CORREL

Devuelve el coeficiente de correlación entre dos rangos de celdas definidos por los argumentos matriz1 y matriz2. Use el coeficiente de correlación para determinar la relación entre dos propiedades. Por ejemplo, para examinar la relación entre la temperatura promedio de una localidad y el uso de aire acondicionado.

Sintaxis

COEF.DE.CORREL (**matriz1**; **matriz2**)

Matriz1 es un rango de celdas de valores.

Matriz2 es un segundo rango de celdas de valores.

Observaciones

- Si el argumento matricial o de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.
- Si los argumentos matriz1 y matriz2 tienen un número diferente de puntos de datos, COEF.DE.CORREL devuelve el valor de error #N/A.
- Si el argumento matriz1 o matriz2 está vacío, o si s (la desviación estándar) o sus valores son cero, COEF.DE.CORREL devuelve el valor de error #¡DIV/0!
- La ecuación para el coeficiente de correlación es:

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}$$

donde x e y son las medias de muestra PROMEDIO (matriz1) y PROMEDIO (matriz2).

Ejemplo

	A	B
1	**Datos1**	**Datos2**
2	3	9
3	2	7
4	4	12
5	5	15
6	6	17
	Fórmula	**Descripción (resultado)**
	=COEF.DE.CORREL(A2:A6;B2:B6)	Coeficiente de correlación de los dos conjuntos de datos anteriores (0,997054)

Función PEARSON

Devuelve el coeficiente de correlación producto o momento r de Pearson, r, un índice adimensional acotado entre -1,0 y 1,0 que refleja el grado de dependencia lineal entre dos conjuntos de datos.

Sintaxis

PEARSON(**matriz1**;**matriz2**)

Matriz1 es un conjunto de valores independientes.

Matriz2 es un conjunto de valores dependientes.

Observaciones

- Los argumentos deben ser números o nombres, constantes matriciales o de referencia que contengan números.
- Si el argumento matricial o de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.

- Si los argumentos matriz1 y matriz2 están vacíos o contienen un número diferente de puntos de datos, PEARSON devuelve el valor de error #N/A.
- La fórmula para el coeficiente de correlación del momento del producto Pearson, r, es:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

donde x e y son las medias de muestra PROMEDIO (matriz1) y PROMEDIO (matriz2).

Ejemplo

	A	B
1	**Valores independientes**	**Valores dependientes**
2	9	10
3	7	6
4	5	1
5	3	5
6	1	3
	Fórmula	**Descripción (resultado)**
	=PEARSON(A2:A6;B2:B6)	Coeficiente de correlación del momento del producto Pearson para los conjuntos de datos anteriores (0,699379)

Función COEFICIENTE.R2

Devuelve el cuadrado del coeficiente de correlación de momento del producto Pearson mediante los puntos de datos de conocido_y y conocido_x. Para obtener más información, vea PEARSON. El valor R cuadrado puede interpretarse como la proporción de la varianza de y que puede atribuirse a la varianza de x.

Sintaxis

COEFICIENTE.R2 (**conocido_y**;**conocido_x**)

Conocido_y es una matriz o un rango de puntos de datos.

Conocido_x es una matriz o un rango de puntos de datos.

Observaciones

- Los argumentos pueden ser números o nombres, matrices o referencias que contengan números.
- Se tienen en cuenta los valores lógicos y las representaciones textuales de números escritos directamente en la lista de argumentos.
- Si el argumento matricial o de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.
- Los argumentos que sean valores de error o texto que no se pueda traducir a números provocan errores.
- Si los argumentos conocido_y y conocido_x están vacíos o contienen un número diferente de puntos de datos, COEFICIENTE.R2 devuelve el valor de error #N/A.
- Si conocido_y y conocido_x contienen sólo contienen un punto de datos, COEFICIENTE.R2 devuelve el valor de error #¡DIV/0!
- La ecuación para el coeficiente de correlación del momento del producto Pearson, r, es:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

donde x e y son las medias de muestra PROMEDIO (conocido_x) y PROMEDIO (conocido y).

COEFICIENTE.R2 devuelve r^2, que es el cuadrado de este coeficiente de correlación.

Ejemplo

	A	B
1	**Valor de y conocido**	**Valor de x conocido**
2	2	6
3	3	5
4	9	11
5	1	7
6	8	5
7	7	4
8	5	4
	Fórmula	**Descripción (resultado)**
	=COEFICIENTE.R2(A2:A8;B2:B8)	Cuadrado del coeficiente de correlación de momento del producto Pearson mediante los puntos de datos anteriores (0,05795)

Cálculo de los parámetros de la ecuación de la línea de regresión (y = a + b.x)

Función INTERSECCION.EJE

Calcula el punto en el que una línea intersecará el eje (parámetro “a” de la ecuación). El punto de intersección se basa en el mejor ajuste de la línea de regresión trazado con los valores X y los valores Y. Utilice la función INTERSECCION.EJE para determinar el valor de la variable dependiente cuando la variable independiente es igual a 0 (cero). Por ejemplo, puede emplear la función INTERSECCION.EJE para predecir la resistencia eléctrica de un metal a 0 °C si los puntos de datos se han tomado a temperatura ambiente o superior.

Sintaxis

INTERSECCION.EJE (**conocido_y**;**conocido_x**)

Conocido_y es el conjunto de observaciones o datos dependientes.

Conocido_x es el conjunto de observaciones o datos independientes.

Observaciones

- Los argumentos deben ser números o nombres, matrices o referencias que contengan números.
- Si el argumento matricial o de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.
- Si los argumentos conocido_y y conocido_x contienen un número diferente de puntos de datos o no contienen ninguno, INTERSECCION.EJE devuelve el valor de error #N/A.
- La ecuación que representa la intersección de la línea de regresión, a, es:

$$a = \overline{y} - b\overline{x}$$

donde la pendiente, b, se calcula como: $b = \dfrac{\sum (x-\overline{x})(y-\overline{y})}{\sum (x-\overline{x})^2}$

y donde x e y son las medias de muestra PROMEDIO (conocido_x) y PROMEDIO (conocido_y).

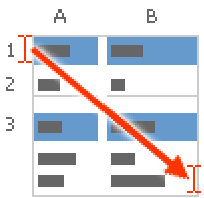
- El algoritmo subyacente utilizado en las funciones INTERSECCIÓN.EJE y PENDIENTE es diferente al algoritmo subyacente utilizado en la función ESTIMACION.LINEAL. La diferencia entre estos algoritmos puede producir resultados distintos cuando los datos son indeterminados y colineales. Por ejemplo, si los puntos de datos del argumento conocido_y son 0 y los puntos de datos del argumento conocido_x son 1:
  - INTERSECCION.EJE y PENDIENTE devuelven un error #¡DIV/0! El algoritmo de INTERSECCIÓN y PENDIENTE está diseñado para buscar una y solamente una respuesta, y en este caso puede haber más de una respuesta.
  - ESTIMACION.LINEAL devuelve un valor 0. El algoritmo ESTIMACION.LINEAL está diseñado para devolver resultados razonables para los datos colineales y, en este caso, se puede encontrar al menos una respuesta.

Ejemplo El ejemplo será más fácil de entender si lo copia a una hoja de cálculo en blanco.

⊞ Cómo copiar un ejemplo

1. Cree una hoja de cálculo o un libro en blanco.
2. Seleccione el ejemplo en el tema de Ayuda.

NOTA No seleccione los encabezados de columna o de fila.



Seleccionar un ejemplo de la Ayuda

3. Presione CTRL+C.
4. En la hoja de cálculo, seleccione la celda A1 y presione CTRL+V.

5. Para cambiar entre ver los resultados y ver las fórmulas que devuelven los resultados, presione CTRL+` (acento grave), o en el grupo **Auditoría de fórmulas** de la ficha **Fórmulas**, haga clic en el botón **Mostrar fórmulas**.

	A	B
1	**Valor de y conocido**	**Valor de x conocido**
2	2	6
3	3	5
4	9	11
5	1	7
6	8	5
	Fórmula	**Descripción (resultado)**
	=INTERSECCION.EJE(A2:A6;B2:B6)	El punto en el que una línea intersecará el eje y utilizando los valores anteriores de x e y (0,0483871)

Función PENDIENTE (parámetro “b” de la ecuación).

Devuelve la pendiente de una línea de regresión lineal creada con los datos de los argumentos conocido_x y conocido_y. La pendiente es la distancia vertical dividida por la distancia horizontal entre dos puntos cualquiera de la recta, lo que corresponde a la tasa de cambio a lo largo de la línea de regresión.

Sintaxis

PENDIENTE (**conocido_y**;**conocido_x**)

Conocido_y es una matriz o rango de celdas de puntos de datos numéricos dependientes.

Conocido_x es el conjunto de puntos de datos independientes.

Observaciones

- Los argumentos deben ser números o nombres, matrices o referencias que contengan números.

- Si el argumento matricial o de referencia contiene texto, valores lógicos o celdas vacías, estos valores se pasan por alto; sin embargo, se incluirán las celdas con el valor cero.
- Si los argumentos conocido_y y conocido_x están vacíos o contienen un número diferente de puntos de datos, PENDIENTE devuelve el valor de error #N/A.
- La ecuación para la pendiente de la línea de regresión es:

$$b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

donde x e y son las medias de muestra PROMEDIO (conocido_x) y PROMEDIO (conocido y).

- El algoritmo subyacente utilizado en las funciones PENDIENTE e INTERSECCION.EJE es diferente al algoritmo subyacente utilizado en la función ESTIMACION.LINEAL. La diferencia entre estos algoritmos puede producir resultados distintos cuando los datos son indeterminados y colineales. Por ejemplo, si los puntos de datos del argumento conocido_y son 0 y los puntos de datos del argumento conocido_x son 1:
  - PENDIENTE e INTERSECCION.EJE devuelven un error #¡DIV/0! El algoritmo de PENDIENTE e INTERSECCION.EJE está diseñado para buscar una y solamente una respuesta, y en este caso puede haber más de una respuesta.
  - ESTIMACION.LINEAL devuelve un valor 0. El algoritmo ESTIMACION.LINEAL está diseñado para devolver resultados razonables para los datos colineales y, en este caso, se puede encontrar al menos una respuesta.

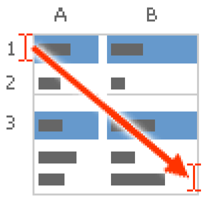
Ejemplo El ejemplo será más fácil de entender si lo copia a una hoja de cálculo en blanco.

⊞ Cómo copiar un ejemplo

1. Cree una hoja de cálculo o un libro en blanco.

2. Seleccione el ejemplo en el tema de Ayuda.

NOTA No seleccione los encabezados de columna o de fila.



Seleccionar un ejemplo de la Ayuda

3. Presione CTRL+C.

4. En la hoja de cálculo, seleccione la celda A1 y presione CTRL+V.

5. Para cambiar entre ver los resultados y ver las fórmulas que devuelven los resultados, presione CTRL+` (acento grave), o en el grupo **Auditoría de fórmulas** de la ficha **Fórmulas**, haga clic en el botón **Mostrar fórmulas**.

	A	B
1	**Valor de y conocido**	**Valor de x conocido**
2	2	6
3	3	5
4	9	11
5	1	7
6	8	5
7	7	4
8	5	4
	Fórmula	**Descripción (resultado)**
	=PENDIENTE(A2:A8;B2:B8)	Pendiente de la regresión lineal a través de los puntos de datos anteriores (0,305556)

La función predictiva del modelo de regresión

Estima el valor "y" (de la variable dependiente) que se asocia a un valor dado o conocido "x" (var. Independiente) según al ajuste del modelo de regresión.

Función PRONÓSTICO

Calcula o pronostica un valor futuro a través de los valores existentes. La predicción del valor es un valor y teniendo en cuenta un valor x. Los valores conocidos son valores x y valores y existentes, y el nuevo valor se pronostica utilizando regresión lineal. Esta función se puede utilizar para realizar previsiones de ventas, establecer requisitos de inventario o tendencias de los consumidores.

Sintaxis

PRONOSTICO(**x**;**conocido_y**;**conocido_x**)

X es el punto de datos cuyo valor se desea predecir.
Conocido_y es la matriz o rango de datos dependientes.
Conocido_x es la matriz o rango de datos independientes.

Observaciones

- Si x no es numérico, PRONOSTICO devuelve el valor de error #¡VALOR!
- Si no se ha especificado ningún valor para conocido_y o conocido_x, o si contienen un número diferente de puntos de datos, PRONOSTICO devuelve el valor de error #N/A.
- Si la varianza de conocido_x es igual a cero, PRONOSTICO devuelve el valor de error #¡DIV/0!
- La ecuación de la función PRONOSTICO es a + bx, donde: $a = \bar{y} - b\bar{x}$

  y: $$b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

  y donde x e y son las medias de muestra PROMEDIO (conocido_x) y PROMEDIO (conocido y).

	A	B
1	Y conocido	X conocido
2	6	20
3	7	28
4	9	31
5	15	38
6	21	40
	Fórmula	**Descripción (resultado)**
	=PRONOSTICO(30;A2:A6;B2:B6)	Predice un valor para y dado un valor de 30 para x (10,60725)

Introducción al análisis inferencial

Función INTERVALO.CONFIANZA

Devuelve un valor que se puede utilizar para construir un intervalo de confianza para una media de una población.

El intervalo de confianza es un intervalo de valores. La media de la muestra, x, está en el centro de este intervalo, y el intervalo es x ± INTERVALO.CONFIANZA. Por ejemplo, si x es la media de una muestra de tiempos de entrega de productos encargados por correo electrónico, x ± INTERVALO.CONFIANZA es un intervalo de medias de la población.

Para cualquier media de población μ_0 (en este intervalo), la probabilidad de obtener una media de muestra más alejada de μ_0 que de x es mayor que alfa; para cualquier media de población μ_0 (fuera del intervalo), la probabilidad de obtener una media de muestra más alejada de μ_0 que de x es menor que alfa. Es decir, suponga que utilizamos x, desv_estándar y tamaño para crear una prueba de dos colas con un nivel de importancia alfa de la hipótesis consistente en que la media de la población es μ_0.

Entonces, no rechazaremos la hipótesis si μ_0 está dentro del intervalo de confianza, y la rechazaremos en caso de que μ_0 no esté en el intervalo de confianza. El intervalo de confianza no nos permite inferir que hay una probabilidad 1 – alfa de que el tiempo de entrega del próximo paquete que encarguemos estará dentro del intervalo de confianza.

Sintaxis

INTERVALO.CONFIANZA (**alfa**;**desv_estándar**;**tamaño**)

Alfa es el nivel de significación utilizado para calcular el nivel de confianza. El nivel de confianza es igual a 100*(1 - alfa)%, es decir, un alfa de 0,05 indica un nivel de confianza de 95%.

Desv_estándar es la desviación estándar de la población para el rango de datos y se presupone que es conocida.

Tamaño es el tamaño de la muestra.

Observaciones

- Si uno de los argumentos no es numérico, INTERVALO.CONFIANZA devuelve el valor de error #¡VALOR!
- Si alfa ≤ 0 o alfa ≥ 1, INTERVALO.CONFIANZA devuelve el valor de error #¡NUM!
- Si el argumento desv_estándar ≤ 0, INTERVALO.CONFIANZA devuelve el valor de error #¡NUM!
- Si el argumento tamaño no es un entero, se trunca.
- Si el argumento tamaño < 1, INTERVALO.CONFIANZA devuelve el valor de error #¡NUM!
- Si suponemos que el argumento alfa es igual a 0,05, se tendrá que calcular el área debajo de la curva normal estándar que es igual a (1 - alfa) o 95%. Este valor es ± 1,96. Por lo tanto, el intervalo de confianza es:

$$\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

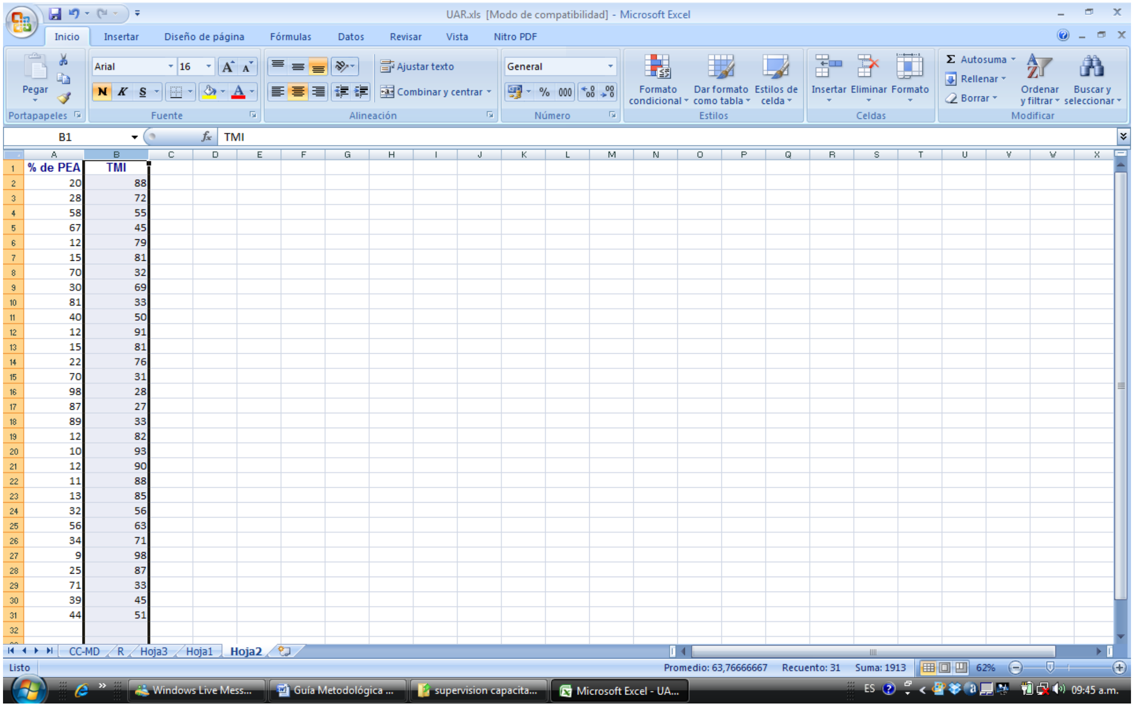
Ejemplo

Suponga que observamos que en nuestra muestra de 50 personas que realizan diariamente un trayecto, la distancia media de viaje es 30 minutos con una desviación estándar de la población de 2,5. Con alfa = 0,05, INTERVALO.CONFIANZA (0,05;2,5;50) devuelve 0,692952. El intervalo de confianza correspondiente será entonces 30 ± 0,692952 = aprox. [29,3;30,7]. Para cualquier media de población μ_0 (en este intervalo), la probabilidad de obtener una media de muestra más alejada de μ_0 que de 30 es mayor que 0,05. Asimismo, para cualquier media de población μ_0 (fuera de este intervalo), la probabilidad de obtener una media de muestra más alejada de μ_0 que de 30 es menor que 0,05.

	A	B
1	**Datos**	**Descripción**
2	0,05	Nivel de significación
3	2,5	Desviación estándar de la población
4	50	Tamaño de la muestra
	Fórmula	**Descripción (resultado)**
	=INTERVALO.CONFIANZA(A2;A3;A4)	Intervalo de confianza para la media de una población. Es decir, el intervalo de confianza para la media de la población de desplazamientos es 30 ± 0,692952 minutos, o de 29,3 a 30,7 minutos. (0,692952)

5. El "paso a paso"en la aplicación de las funciones estadísticas para el análisis de la información.

5.1. El análisis descriptivo de una base de datos



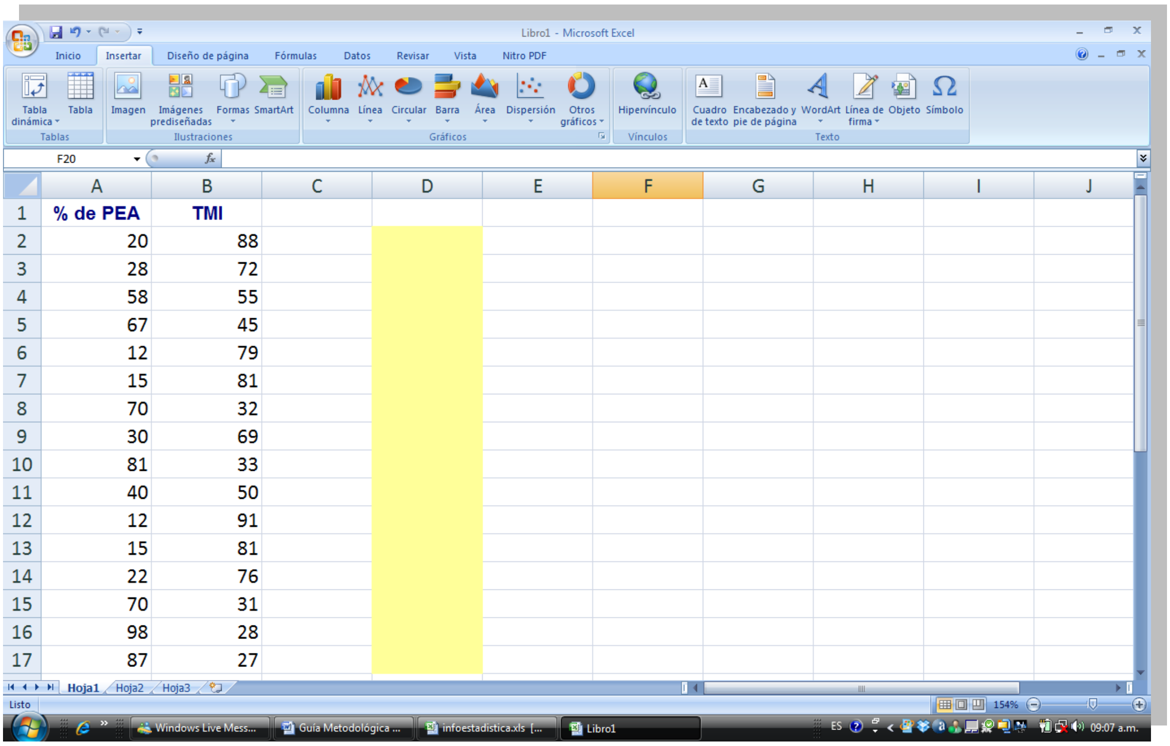
Supongamos que tenemos que analizar una serie de frecuencias con 30 TMI[9] situados en la columna "B" de la figura de arriba y realizar un análisis descriptivo.

Necesitamos medidas centrales, dispersión, posición. Determinar la variabilidad del fenómeno.

Si la situación lo justifica, realizar un análisis de tendencias como por ejemplo una serie cronológica, gracias al método de regresión.

Lo mismo que un análisis de asociación y consistencia con los métodos de correlación y regresión, si se cumplen todos los supuestos o requisitos necesarios para aplicar este estudio paramétrico.

[9] Tasa de Mortalidad Infantil: Defunciones en menores de 1 año por 1000 nacidos vivos.



	A	B
1	% de PEA	TMI
2	20	88
3	28	72
4	58	55
5	67	45
6	12	79
7	15	81
8	70	32
9	30	69
10	81	33
11	40	50
12	12	91
13	15	81
14	22	76
15	70	31
16	98	28
17	87	27

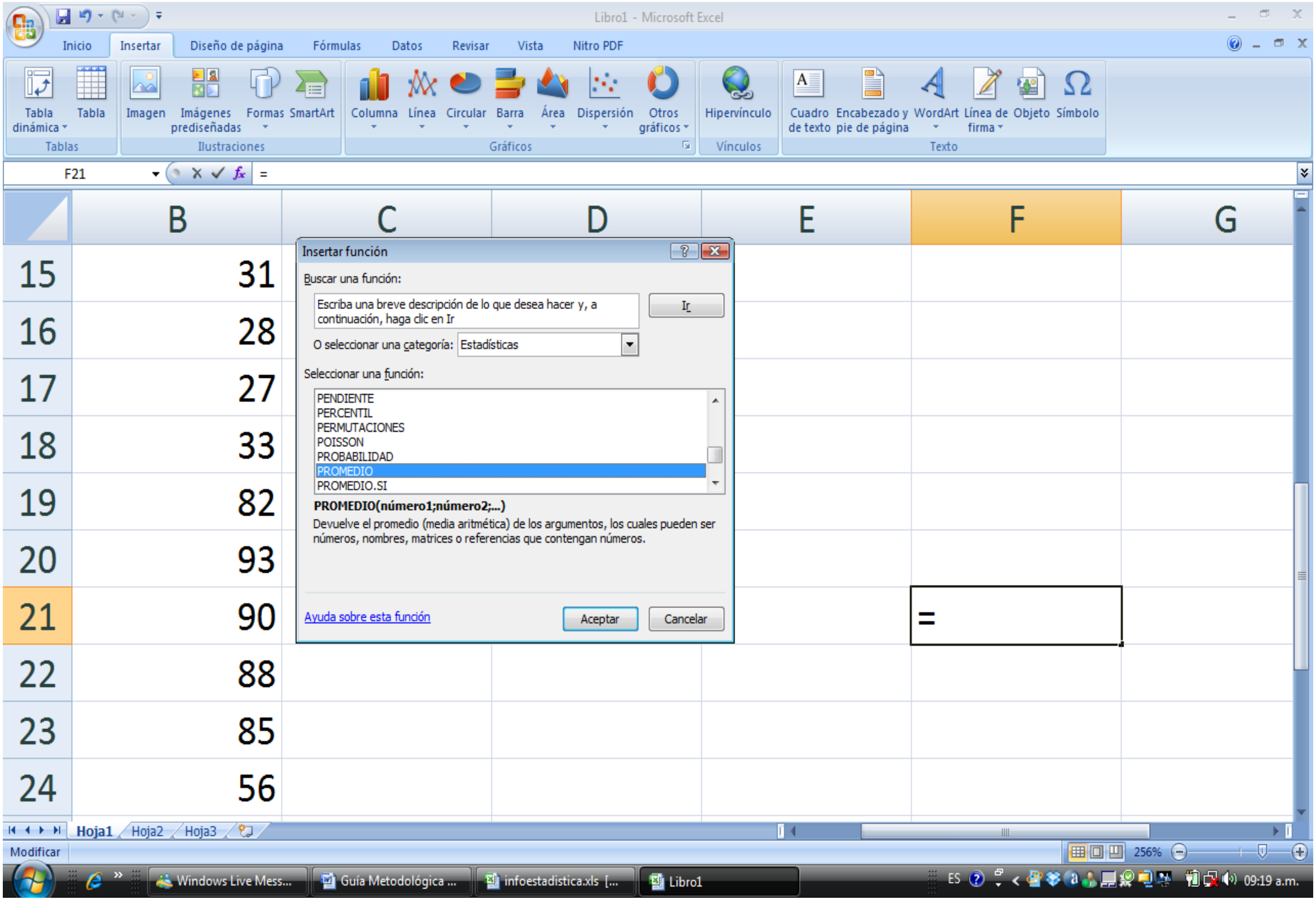
Pasos

Ir a insertar función (***f**$_x$*) y seleccionar la categoría “Estadísticas”.

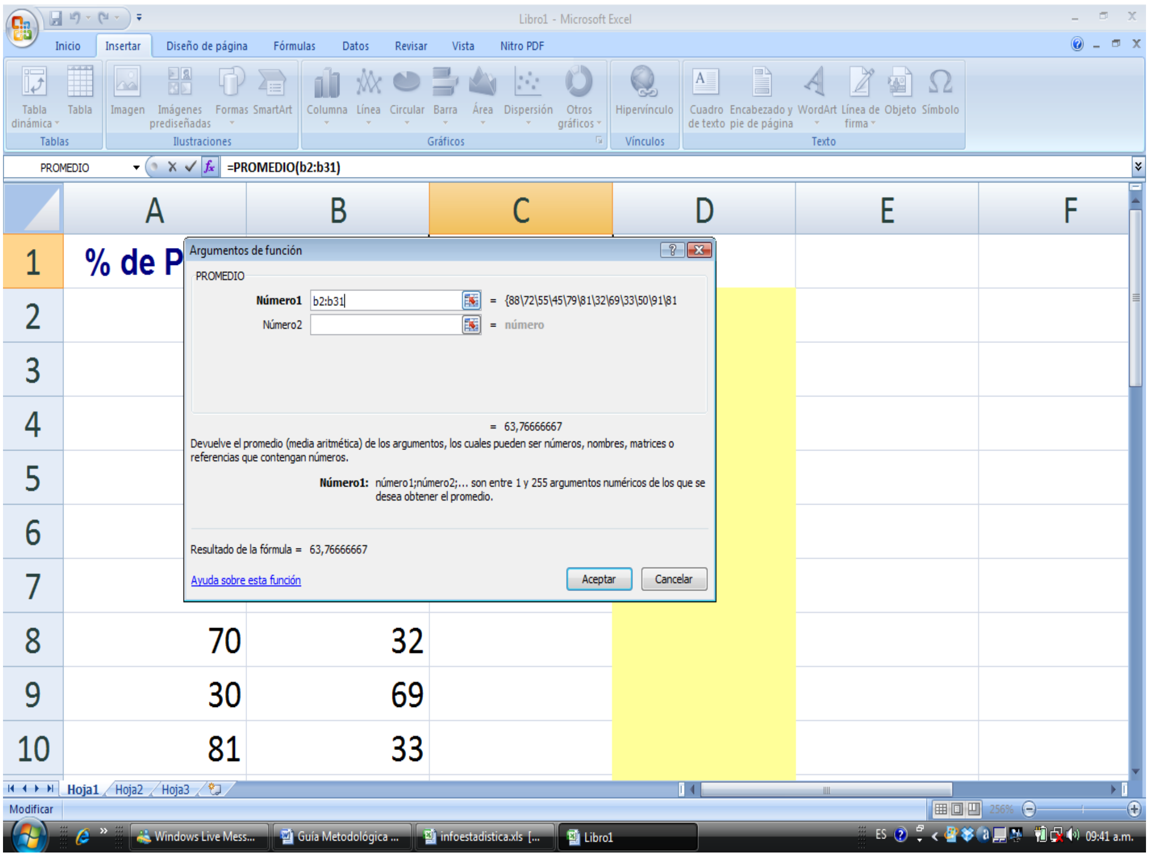
Las distintas funciones aparecen en orden alfabético.

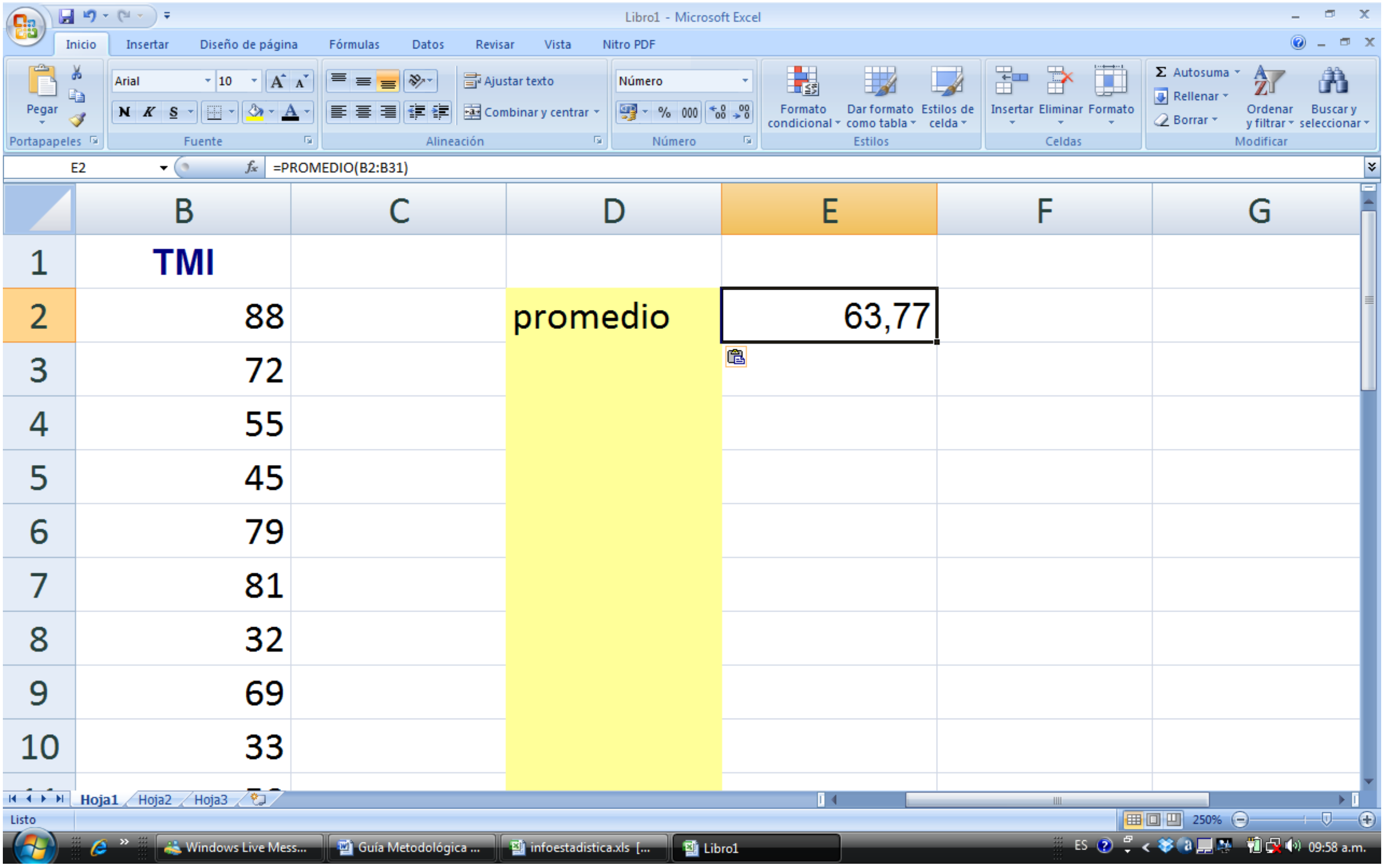
Nos ocuparemos en primer lugar de las medidas centrales (promedio, mediana, modo); las de posición (percentiles) y de dispersión.

La casilla de salida de resultados debe estar vacía.



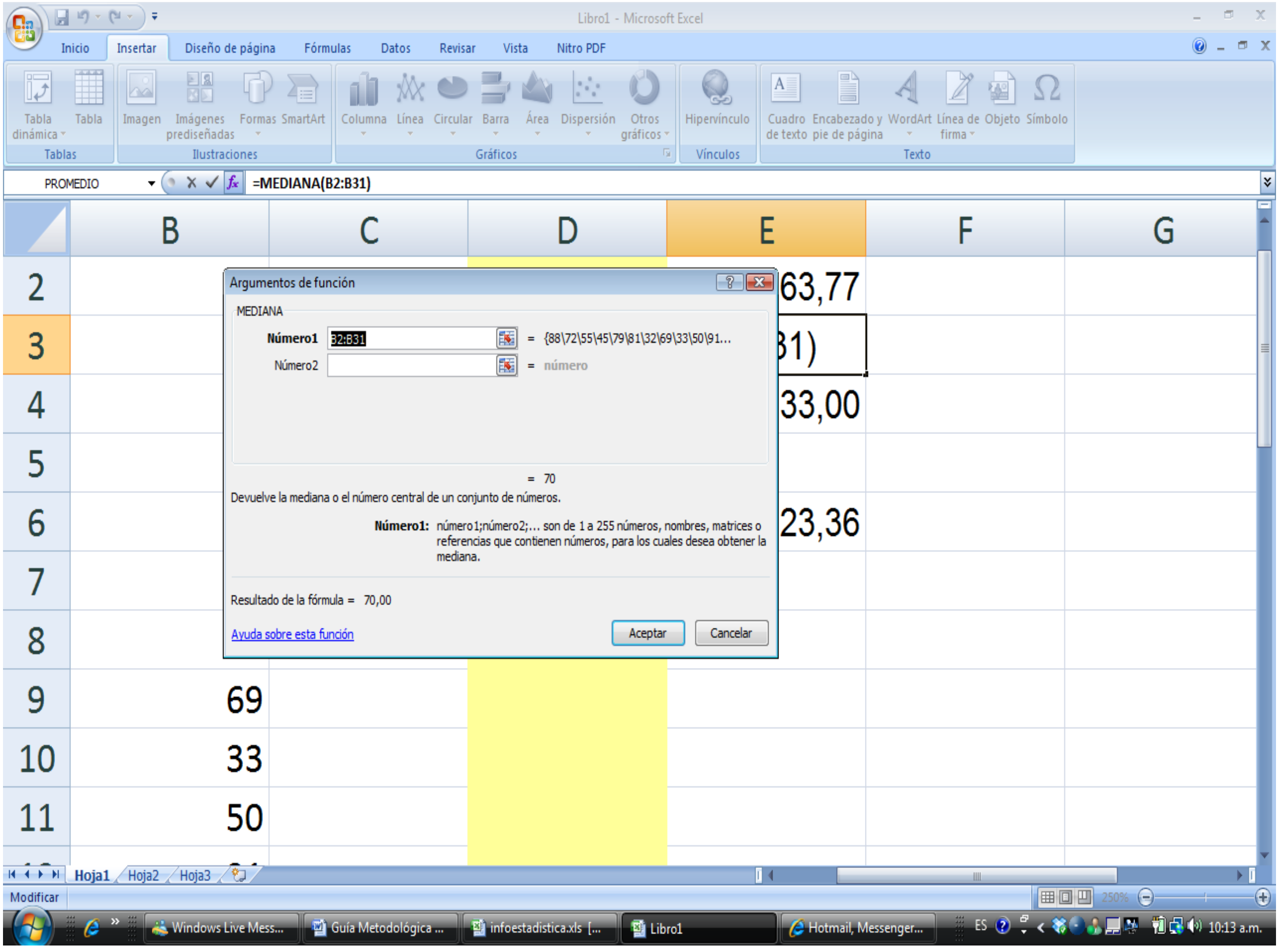
Seleccionamos la función que se desea (por ej. “promedio”) y aceptamos





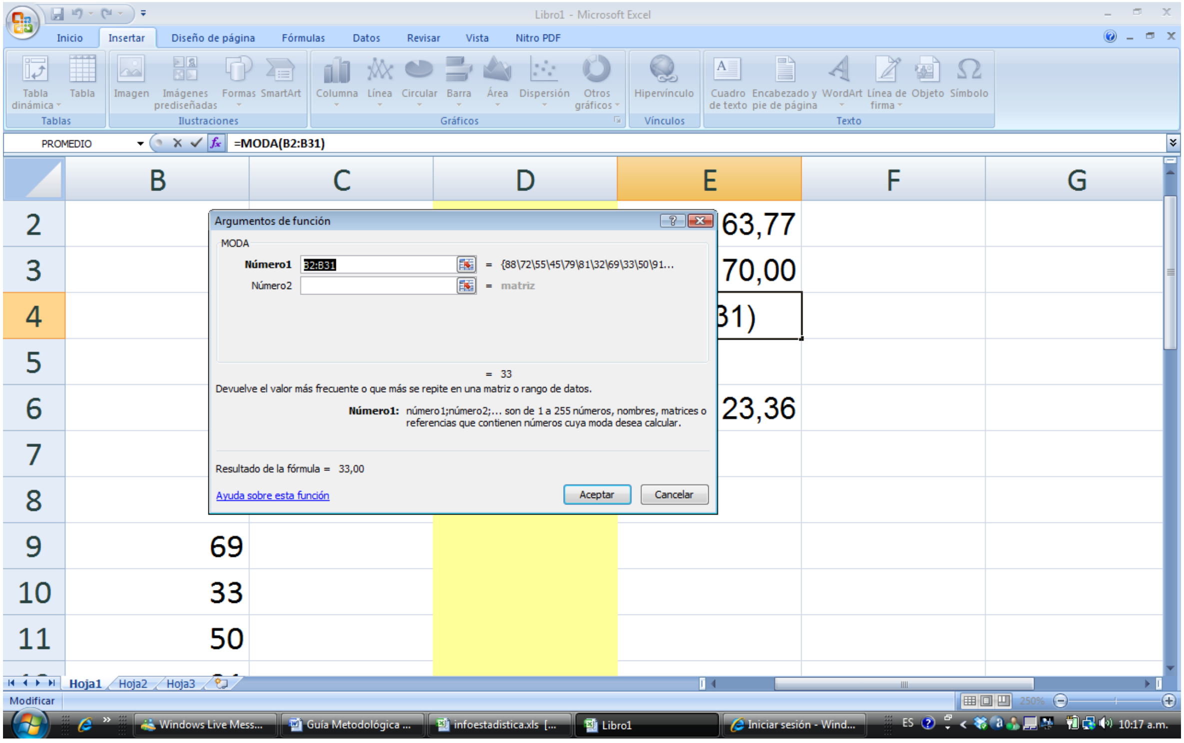
Se introduce en “Número 1” el rango de valores (**b2:b31**) que se desea analizar. Para nuestro ejemplo, los 30 datos de la variable “TMI” (tasa de mortalidad infantil”). Se da aceptar. El resultado aparecerá en la celda en donde fijamos el cursor.

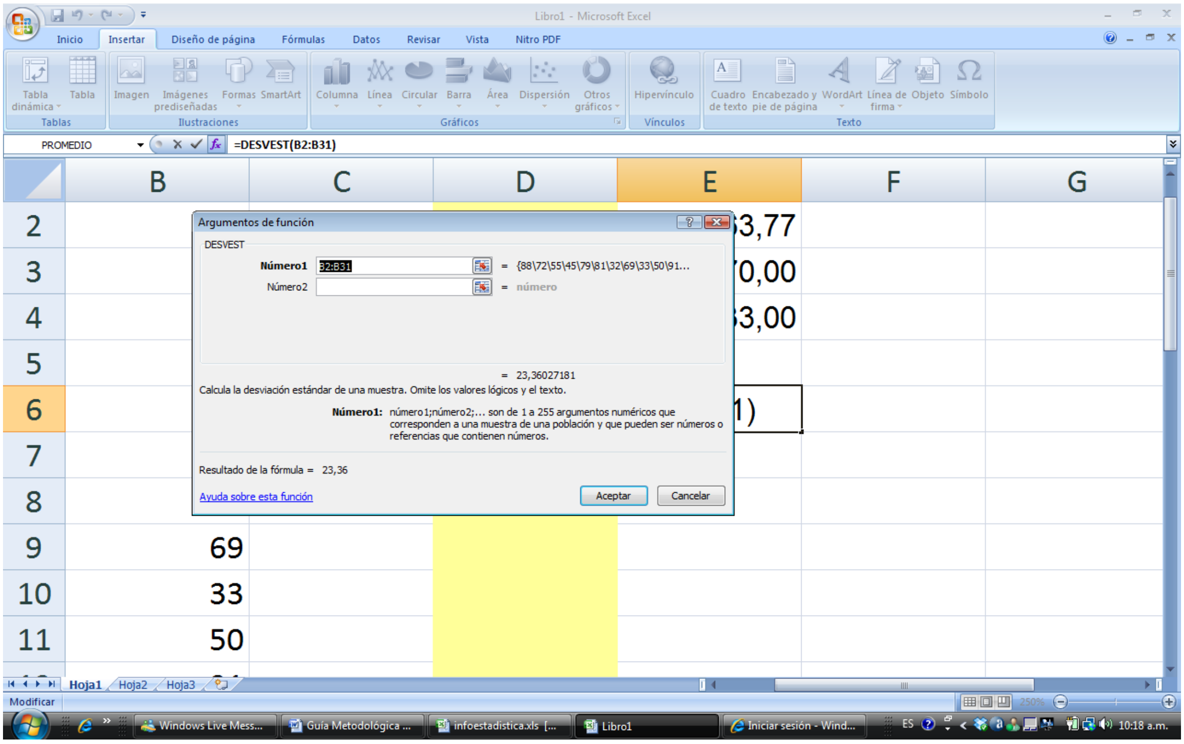
Se procede de idéntica manera con la mediana, moda, desvío estándar (se halla listado bajo la nomenclatura “desvest”).

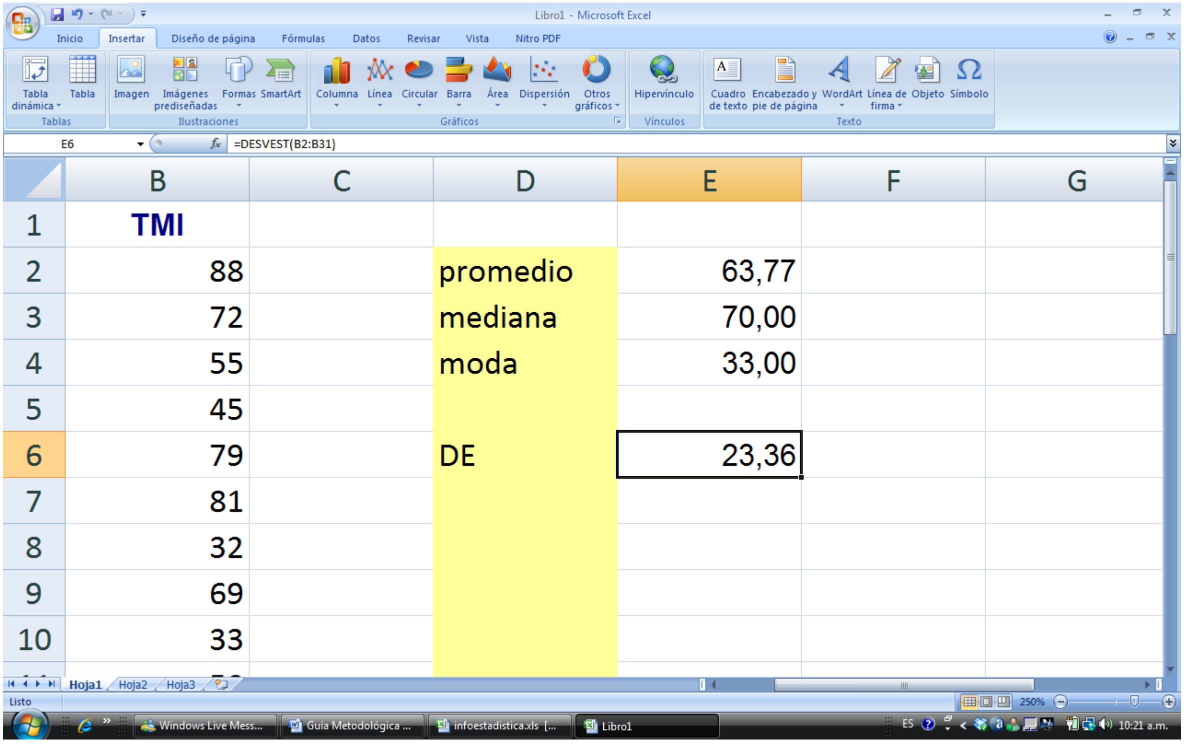


Recordemos que la mediana es a la vez una medida de centralización y de posición; cuyo valor divide la serie de frecuencias analizada en dos partes iguales, 50% para abajo y 50 % para arriba. Es ideal para analizar con mayor seguridad la centralización de valores con respecto a la media aritmética, que se halla “arrastrada” por algún dato anómalo o raro como ocurre en una serie asimétrica.

En nuestro ejemplo, la Ma corresponde a TMI = 70. Es decir, que el 50 % de las observaciones son ≤ 70.


=MODA(B2:B31)
Argumentos de función
MODA
Número1
Número2
Devuelve el valor más frecuente o que más se repite en una matriz o rango de datos.
Resultado de la fórmula = 33,00
Aceptar
Cancelar


=DESVEST(B2:B31)
Argumentos de función
DESVEST
Número1
Número2
= 23,36027181
Calcula la desviación estándar de una muestra. Omite los valores lógicos y el texto.
Resultado de la fórmula = 23,36
Aceptar
Cancelar



Interpretación:

Según lo desarrollado en los capítulos precedentes, podemos afirmar que tenemos una serie con asimetría negativa, con una TMI[10] promedio de 64, donde la mediana que se convierte en la medida de centralización mas fidedigna, nos informa que el 50 % de las observaciones son ≤ a 70.La TMI más frecuentemente hallada en la serie es 33.

Un DE de 23 y un coeficiente de variabilidad[11] (V) cercano al 37 % nos indica que le fenómeno que estamos analizando en la serie es bastante variable; por lo que las inferencias deben considerarse con más cuidado.

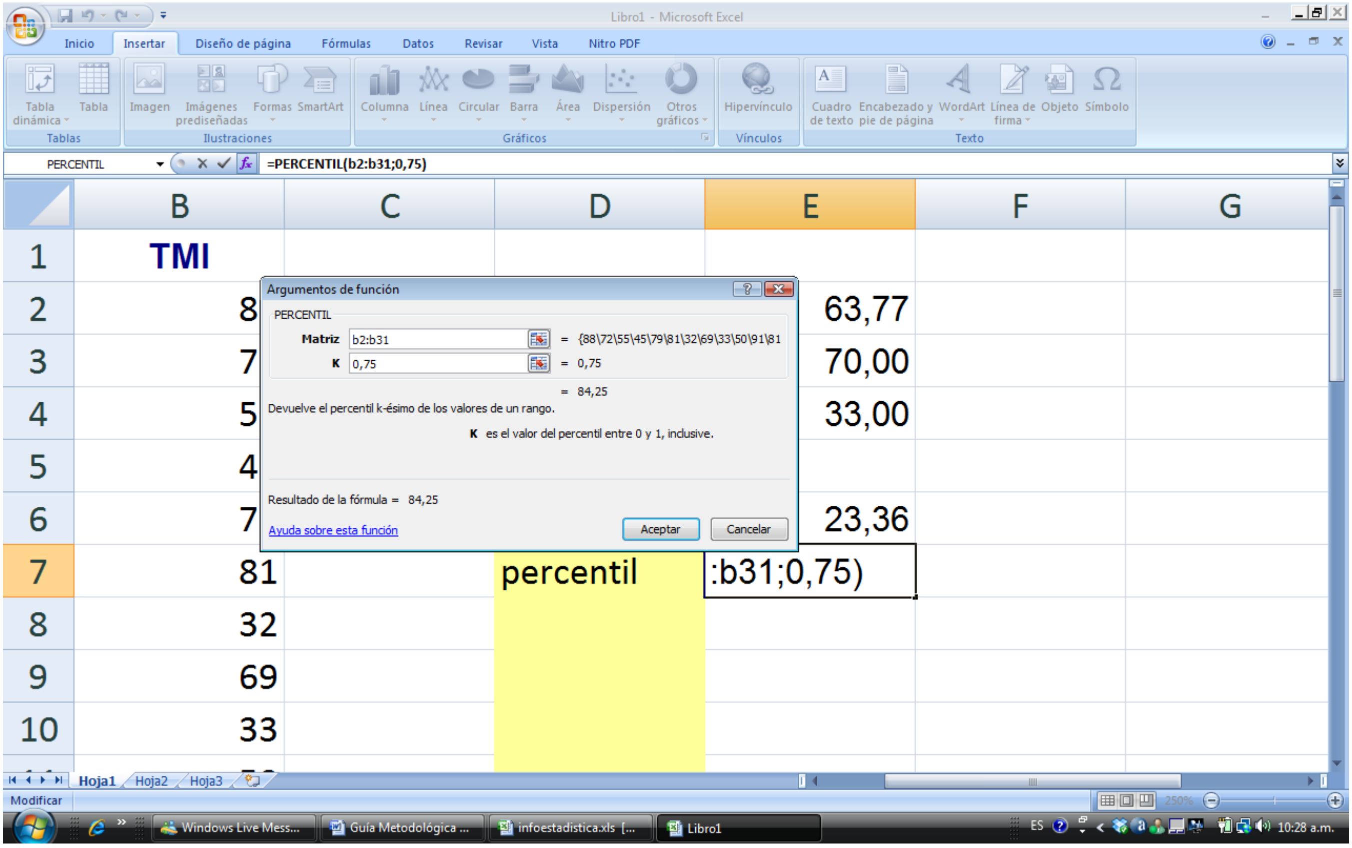
Para calcular los percentiles, la única diferencia con respecto a los pasos realizados, es que en la casilla "K" se debe incluir el percentil que se desea calcular, teniendo el cuidado de hacerlo en la escala del 0 a 1. Así que por ejemplo al percentil$_{75}$ corresponde 0,75.

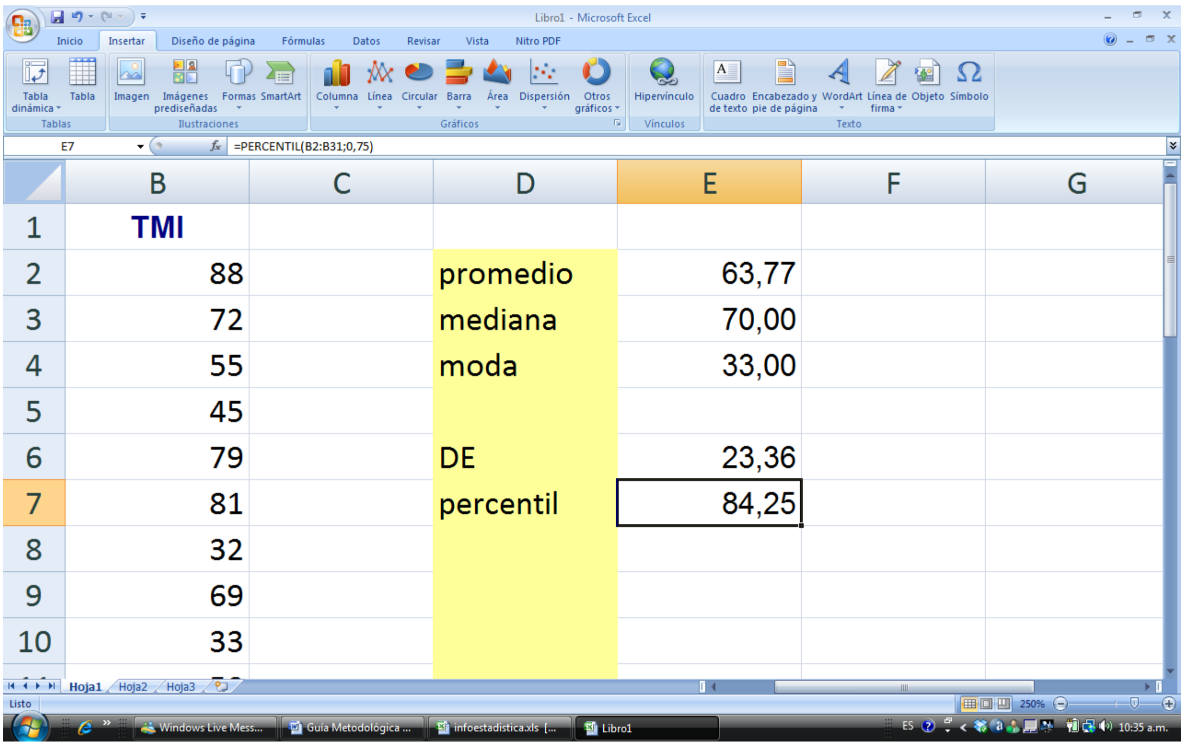
p_{75} = 84,25. Es decir que el 75 % de las observaciones son ≤ a 84.

---

[10] Recordemos que la TMI se expresa en defunciones en menores de 1 año por mil nacidos vivos. Con fines prácticos del ejercicio, solo se menciona el valor numérico.

[11] El coeficiente de variabilidad (V) = $\frac{DE}{X} \cdot 100$


Libro1 - Microsoft Excel
=PERCENTIL(b2:b31;0,75)
TMI
Argumentos de función
PERCENTIL
Matriz b2:b31 = {88\72\55\45\79\81\32\69\33\50\91\81
K 0,75 = 0,75
= 84,25
Devuelve el percentil k-ésimo de los valores de un rango.
K es el valor del percentil entre 0 y 1, inclusive.
Resultado de la fórmula = 84,25
Ayuda sobre esta función
Aceptar
Cancelar
63,77
70,00
33,00
23,36
81
percentil
:b31;0,75)
32
69
33


Libro1 - Microsoft Excel
E7
=PERCENTIL(B2:B31;0,75)
TMI
88
promedio
63,77
72
mediana
70,00
55
moda
33,00
45
79
DE
23,36
81
percentil
84,25
32
69
33

5.2. Introducción al análisis inferencial

La estimación del intervalo de confianza nos permite acceder a valiosas informaciones:

1- Estimar una valoración de una densidad porcentual de la serie analizada.

2- Estimar los límites de la serie estudiada, según la certeza deseada.

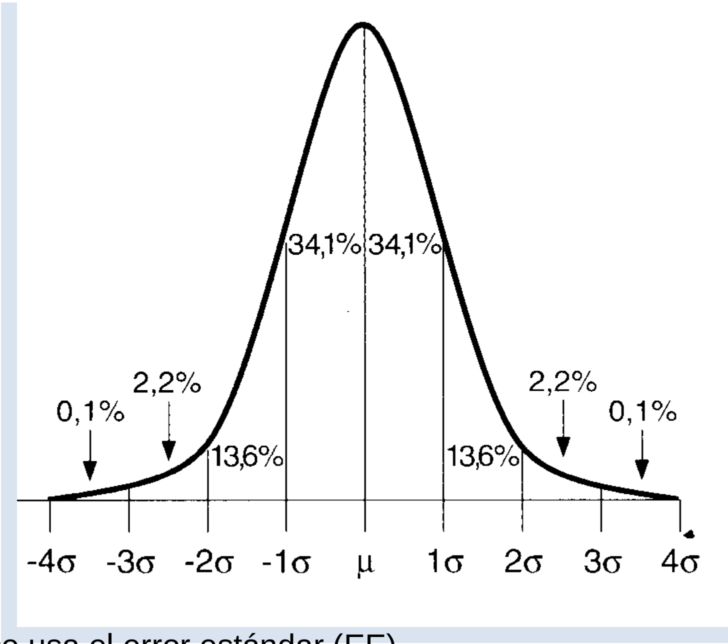
3- Estimar los límites del universo según la certeza deseada; y con ello, las áreas de significancia y no significancia estadística del universo a que pertenece los datos estudiados en el carácter de muestra.

Para las dos primeras situaciones, se usan el DE.

X ± 1 D.E. = 68 % del área (68,27 %);

X ± 2 D.E. = 95 % del área (95,45%);

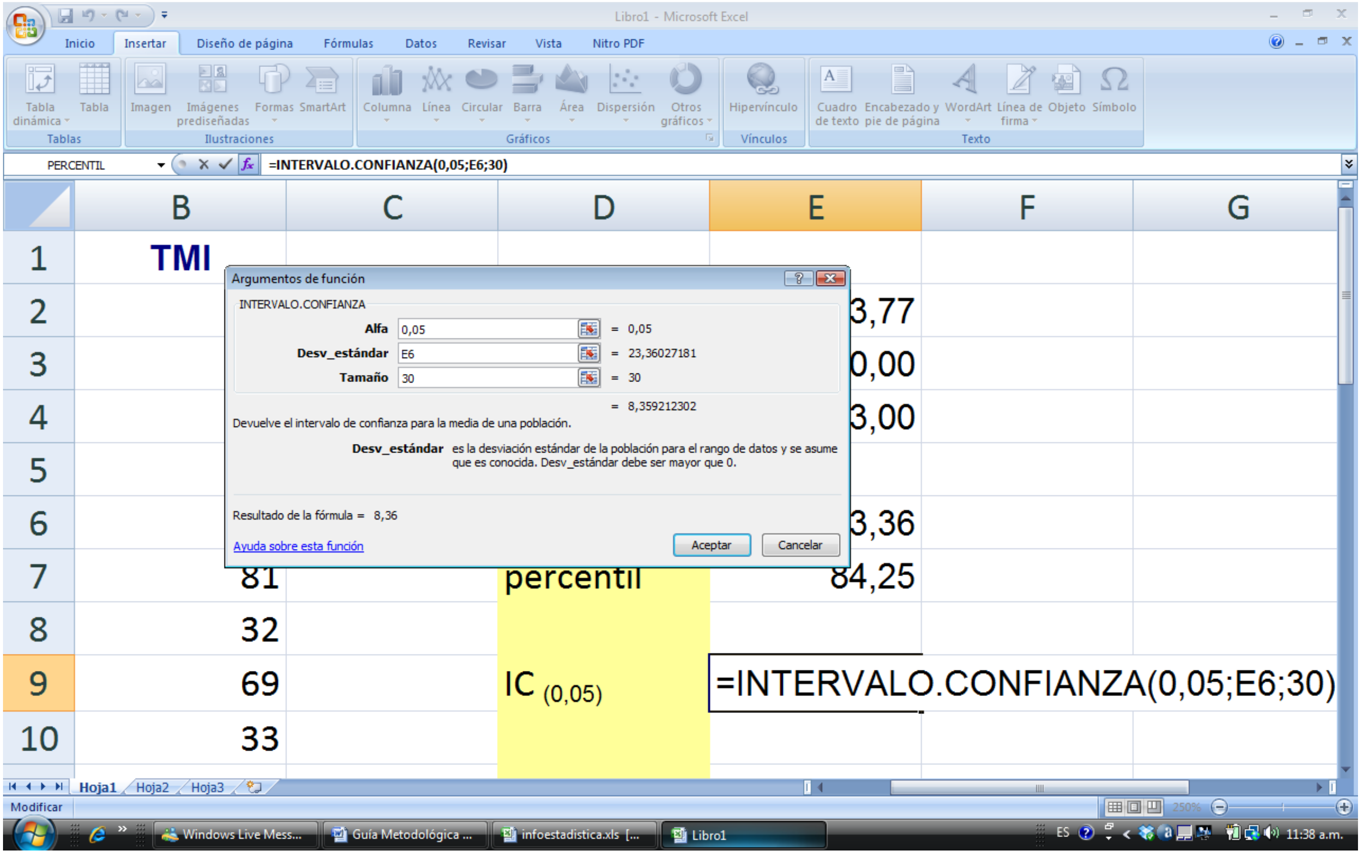
X ± 3 D.E. = 100 % del área (99,73%).



Para la tercera, se usa el error estándar (EE).

Excell nos da la posibilidad de analizar en el contexto de esta última situación (ver la descripción de la función).

El intervalo de confianza (IC) se estima usando la función “intervalo.confianza”.



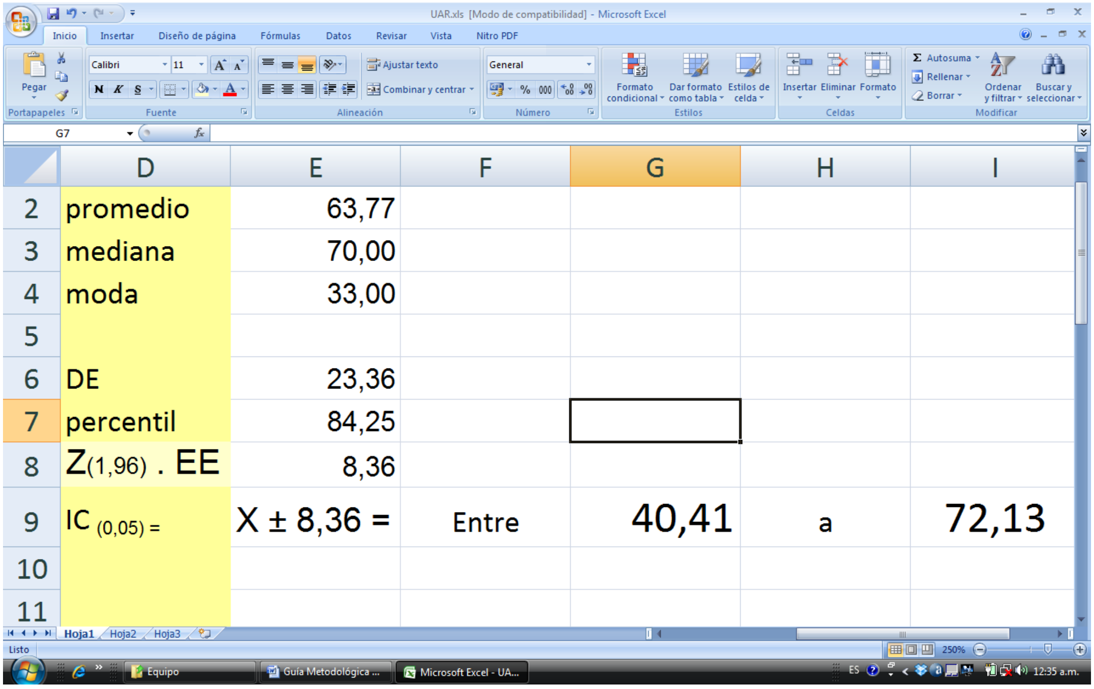
La casilla “Alfa” corresponde al nivel de significancia o error que se está dispuesto a correr. Se debe escribir según la escala 0 a 1. Alfa es el complemento del nivel de certeza. Así por ejemplo, si se desea un 95 % de certeza, corresponde un error tipo alfa del 5 % (se debe escribir 0,05).

Se completan las casillas para el desvío estándar y el tamaño de la muestra, es decir la cantidad de valores considerados (en nuestro ejemplo = 30).

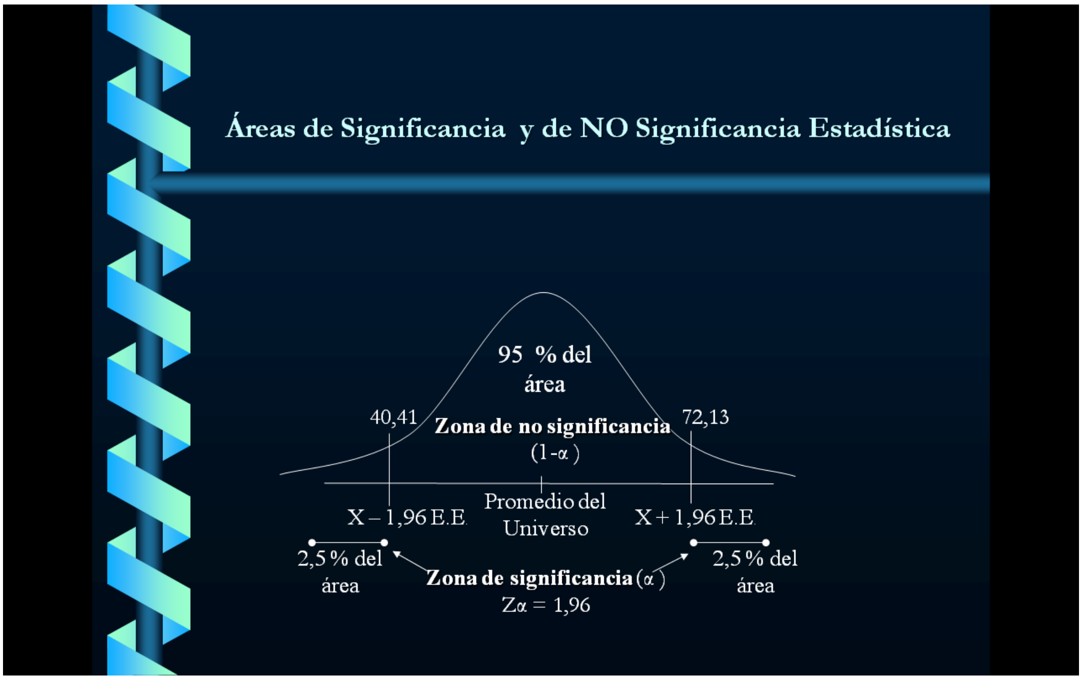
Si se conoce el DE poblacional, se la utiliza. Caso contrario se puede optar por el DE muestral[12] como estimador.

$IC_{0,05} = X \pm Z \,.\, EE$ ⟹ $IC_{0,05} = 63,77 \pm 8,36 =$ Entre **40,41** a **72,13**

[12] Se puede usar la función “desvetp” que estima el desvío estándar poblacional, cuando “n” es pequeña (n<30); ya que disminuirá el error estándar y con ello, los límites del IC, favoreciendo el hallazgo de diferencias estadísticamente significativas.



	D	E	F	G	H	I
2	promedio	63,77				
3	mediana	70,00				
4	moda	33,00				
5						
6	DE	23,36				
7	percentil	84,25				
8	Z(1,96) . EE	8,36				
9	IC (0,05) =	X ± 8,36 =	Entre	40,41	a	72,13
10						
11						



Interpretación:

Por tanto, los valores que caen por fuera de dichos límites (40, 41 a 72,13) del IC$_{0,05}$ tienen significancia estadística, por caer en el área de significancia; con una certeza del 95 %.

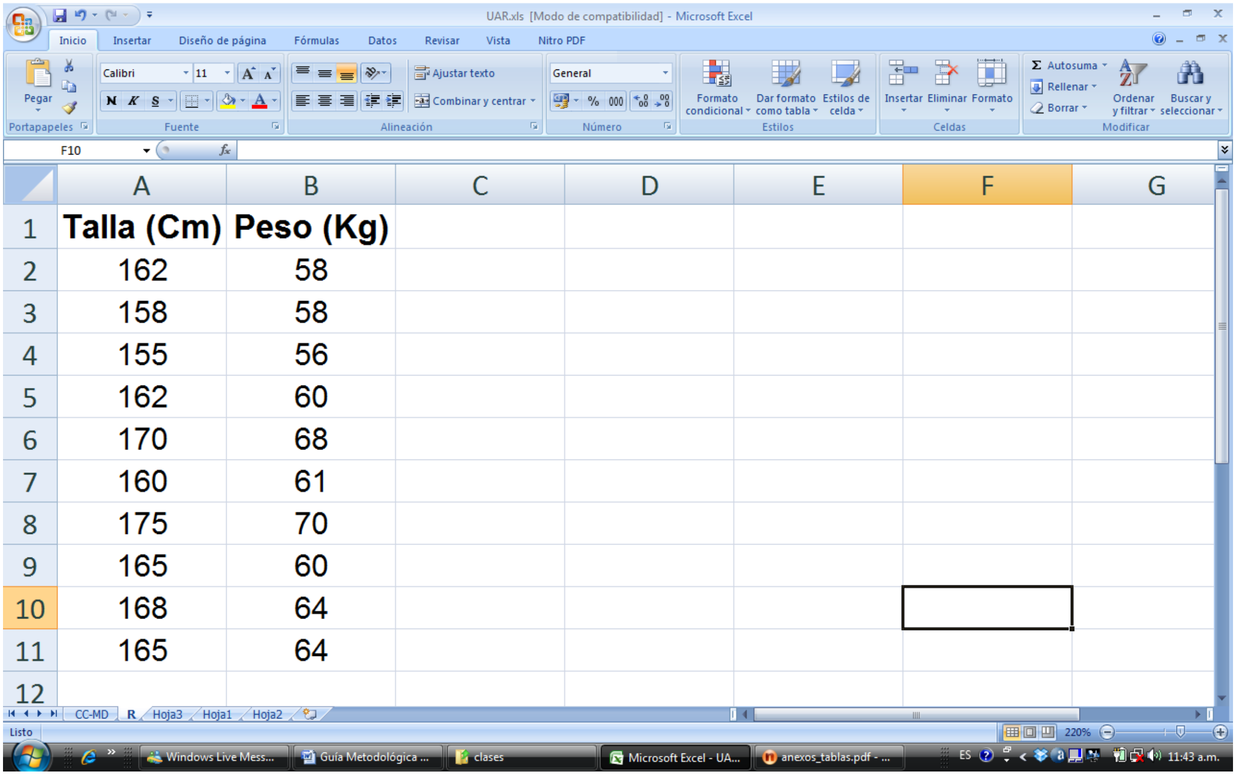
5.3. Análisis de dos variables o de Datos de Asociación

Métodos de Correlación y Regresión

En ocasiones enfrentamos el reto de determinar si dos variables se hallan asociadas siguiendo un modelo lineal (problema de correlación lineal). Esta situación puede plantearse para investigar una situación de consistencia de datos. Una vez demostrado esto, se procede a establecer la significancia estadística.

Establecida la asociación significativa de las variables, se plantea el predecir o estimar un valor desconocido de la variable respuesta o dependiente (problema de regresión lineal); como podría ser el estimar una tendencia de una serie cronológica. En este momento se debe evaluar que tan bueno es el modelo para explicar las variaciones de la variable dependiente.

Veamos el "paso a paso" del análisis con un ejemplo. Supongamos que queremos determinar la asociación o no de dos variables: peso (kg) y talla (cm).

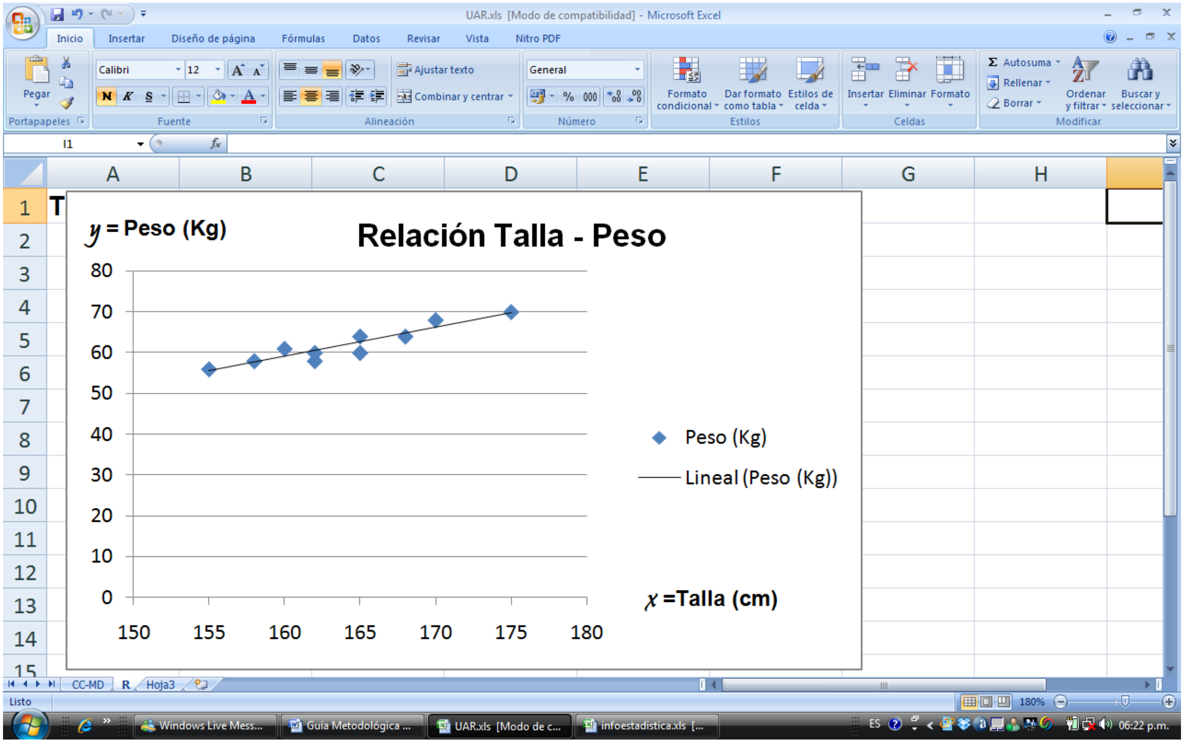


	A	B
1	Talla (Cm)	Peso (Kg)
2	162	58
3	158	58
4	155	56
5	162	60
6	170	68
7	160	61
8	175	70
9	165	60
10	168	64
11	165	64

1. Determinar la variable dependiente (y) e independiente (x). Peso (y); talla (x).

2. Hacer un gráfico de correlación para establecer visualmente una relación lineal.

   Repasar que se cumplen con todos los supuestos o requisitos de este método paramétrico de análisis.



El gráfico nos presenta una clara relación lineal (fijarse como todos los puntos están muy próximos a línea de regresión), por tanto seguimos al siguiente paso:

3. Calcular el coeficiente de correlación de Pearson (“r”). Recordando que cuanto más cercano a “1”, más perfecta es la asociación; y que el signo determina el sentido de la relación (directa o inversa).

Con la función “COEF.DE.CORREL” se introducen en las casillas matriz 1 y 2, el rango de valores correspondientes a las dos variables consideradas (sin necesidad de especificar la dependiente o la independiente). En nuestro ejemplo, r = 0,935.

Un “r” tan próximo a la unidad, casi asegura una relación significativa, no obstante, debemos confirmarla al comparar con el valor “r” crítico de la tabla (ver anexo[13]).

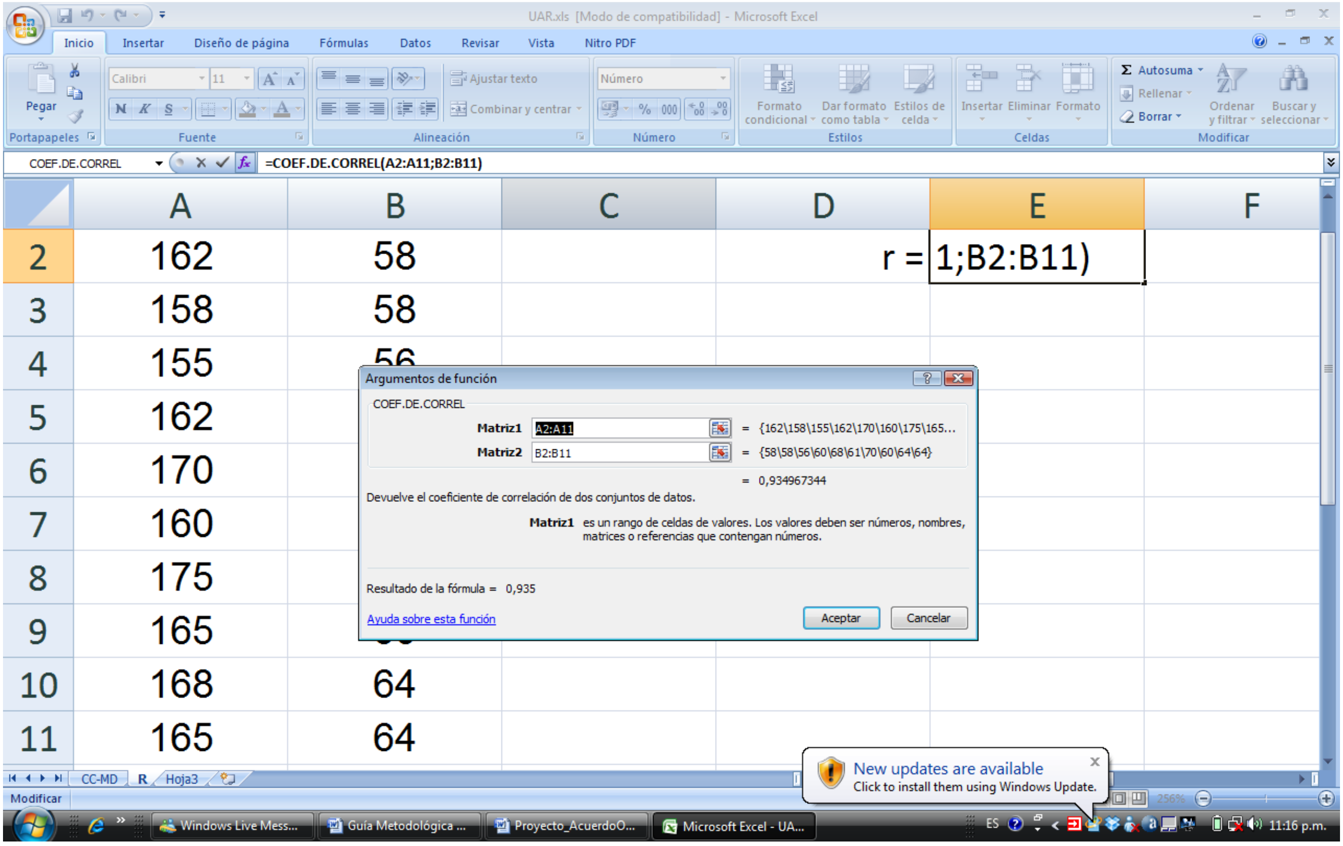
$$r_{tabulado}\ (n\text{-}2 = 8;\ \alpha = 0{,}05) = 0{,}632$$

La regla de decisión es: $r_{observado} \geq r_{tabulado}$ (según un grado de libertad n-2; y un nivel de significancia $\leq$ 0,05 %) = relación significativa.

0,935 ($r_{observado}$) > $r_{tabulado}$ (n-2 = 8; α = 0,05) 0,632 > 0,765 ($r_{tabulado}$ n-2 = 8; α = 0,01)

13 Para buscar el valor crítico de la “r” en la tabla de la sección “Anexos”; ver en la columna α para un test de dos colas = 0,05. La primera columna “gl” equivale a n-2. En nuestro ejemplo = n – 2 = gl. 8.

Es decir, la “r” estimada es significativa con una certeza mayor al 99 %

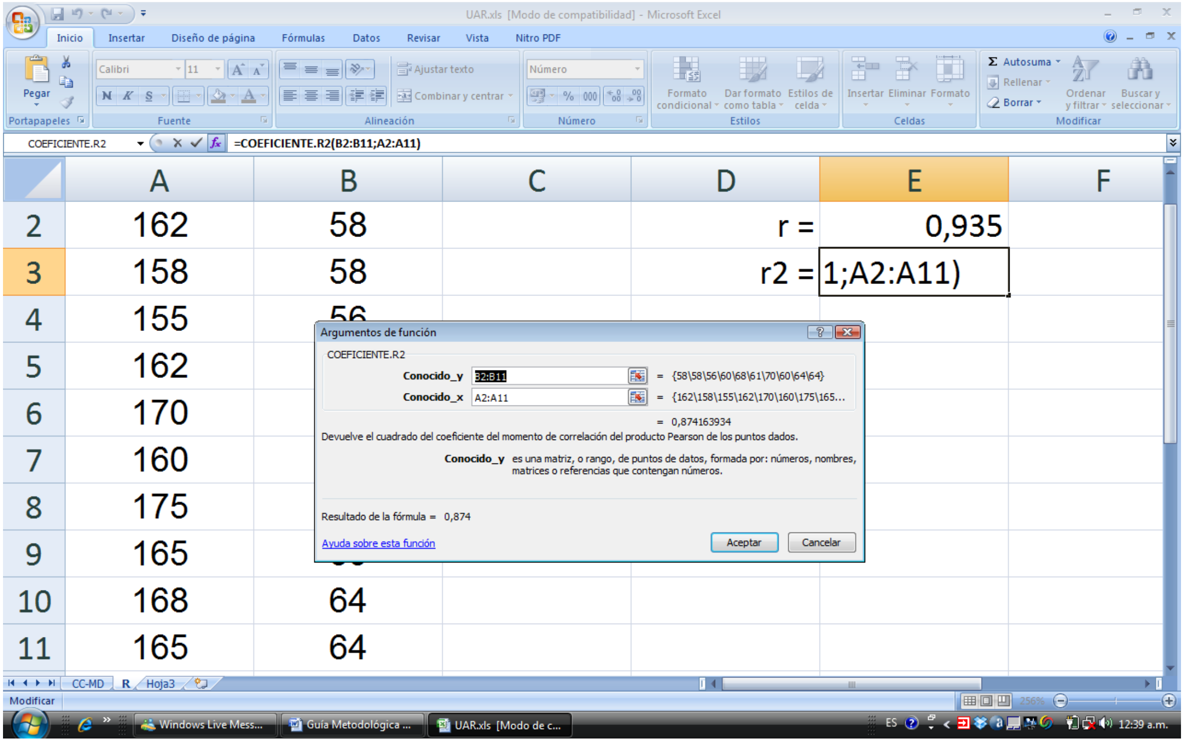


4. Establecer la bondad del ajuste del modelo de regresión, es decir que parte de la variación de *y* está explicada por *x* (variación explicada). Para esto debemos calcular el coeficiente de determinación (r^2).

5. La función “COEFICIENTE.R2”, eleva al cuadrado “r”. En nuestro ejemplo r^2 = 0,874. Si lo expresamos en porcentaje nos da el porcentaje de la variación explicada. El 87 % de la variación del peso está explicada en función de la talla.

La función “COEFICIENTE.R2”, eleva al cuadrado “r”. En nuestro ejemplo r^2 = 0,874. Si lo expresamos en porcentaje nos da el porcentaje de la variación explicada. El 87 % de la variación del peso está explicada en función de la talla.

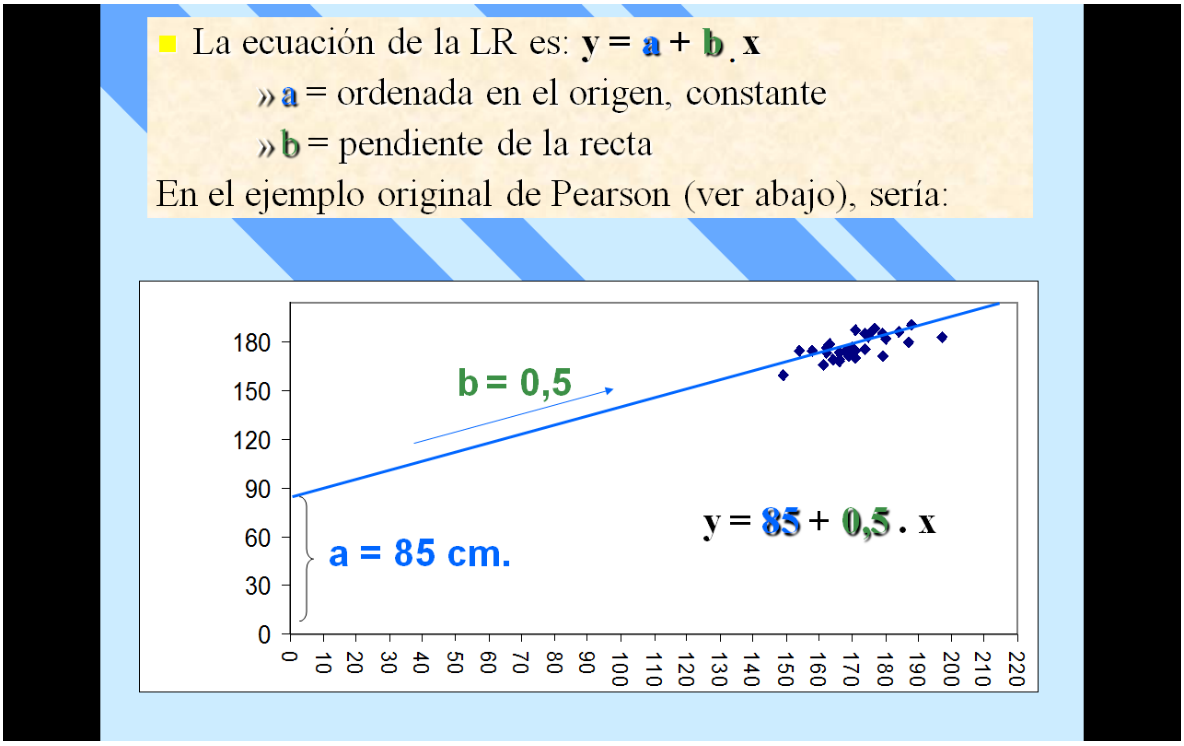
Con la información analizada podemos pasar a estimar valores de la dependiente según valores conocidos de la variable independiente con el aval de que la estimación tendrá el rigor técnico suficiente. Es decir, usar el método de regresión lineal.



Si no hubiéramos establecido una correlación significativa entre las variables esta estimación no tendría sentido.

6. Análisis de la Regresión

Estimación de los parámetros de la ecuación de la recta de regresión

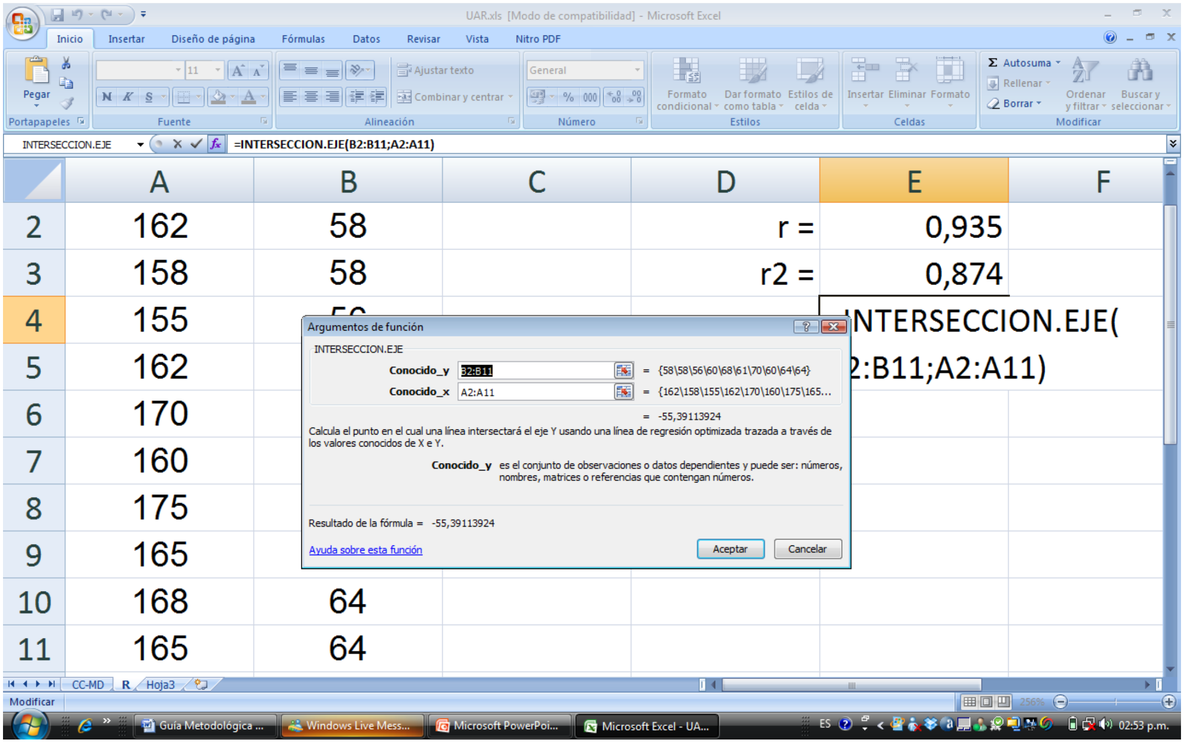


7. Para estimar "a" usamos la función "INTERSECCION.EJE". En la casilla "Conocido_y" van el rango de valores de la dependiente. En "Conocido_x" van el rango de valores de la independiente.

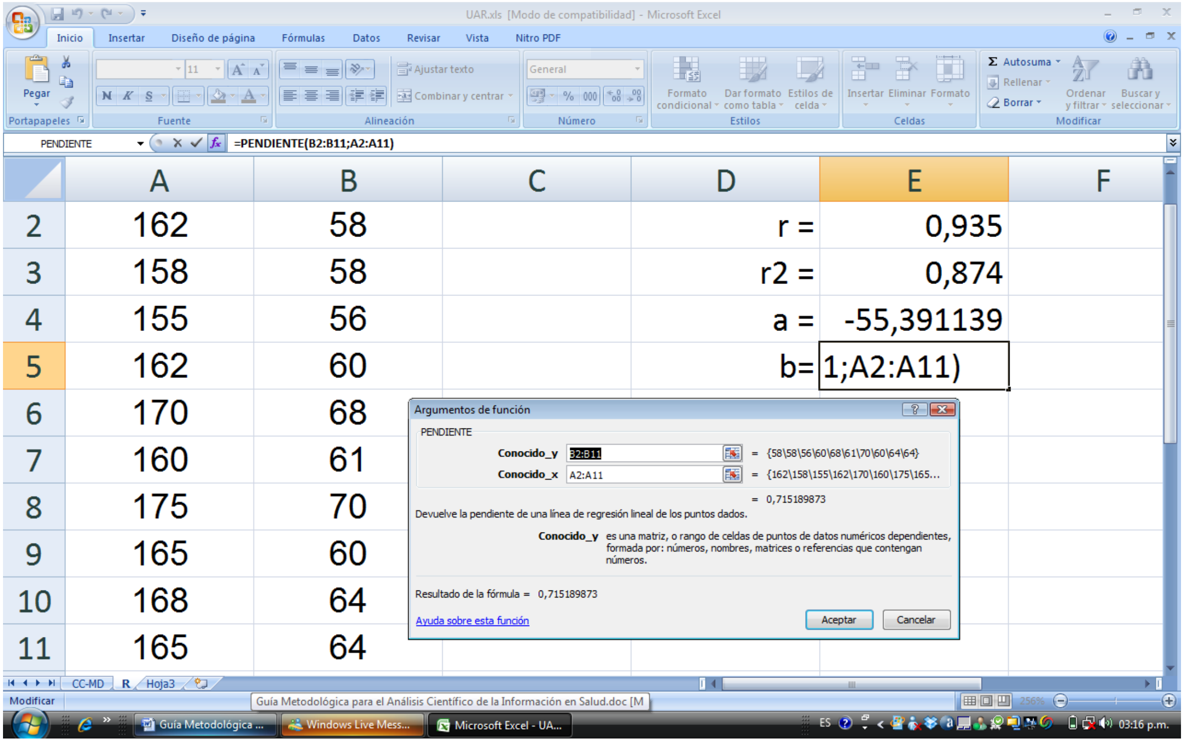
8. Para estimar "b" usamos la función "PENDIENTE". Las casillas son similares a la anterior función.

La ecuación de la LR: $y = -55{,}391 + 0{,}715 . x$

<u>La estimación de "a" – Función "INTERSECCION.EJE"</u>

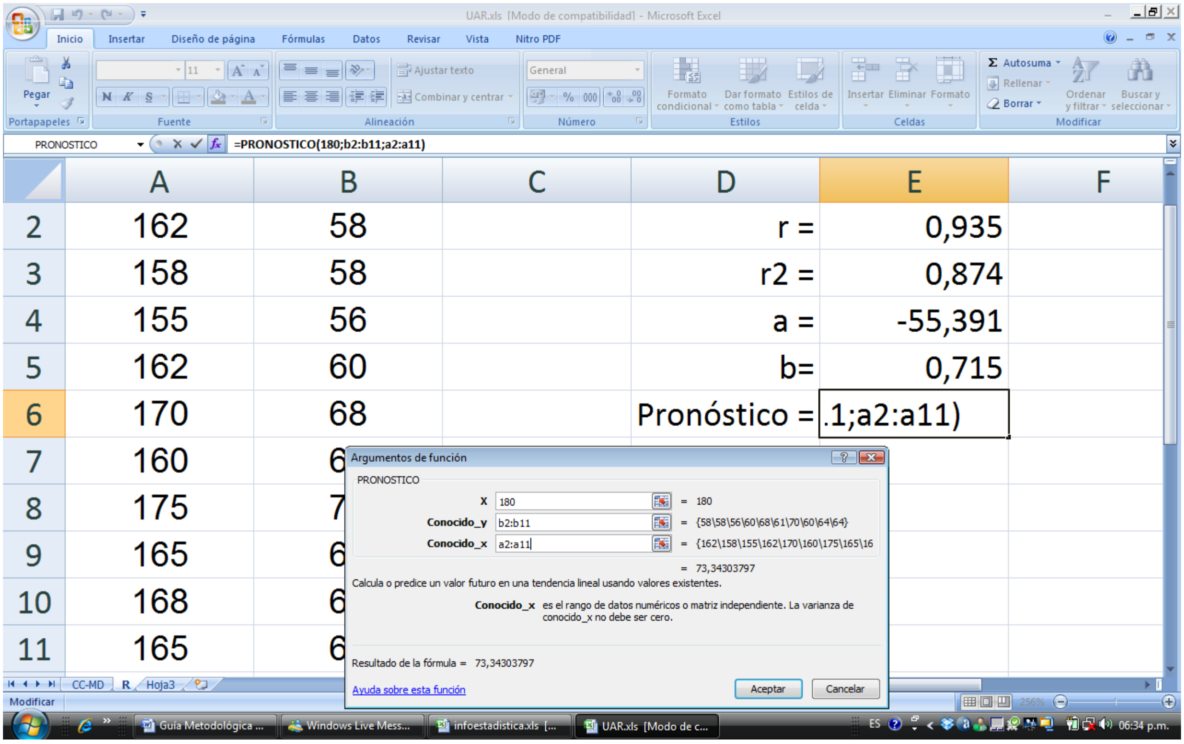


<u>La estimación de "b"- Función "PENDIENTE"</u>



9. La estimación más importante con el método de regresión es la predicción o tendencia de un valor y, dado un valor conocido de x. Supongamos que queremos estimar que peso tendría una sujeto con 180 cm de talla

Para tal efecto, usamos la función "PRONOSTICO".



En la casilla "x" se pone el valor conocido de la variable independiente. En nuestro ejemplo, 180 cm. Las casillas "conocido_y"; "conocido_x" son los mismos que en las funciones ya usadas. El y_{180} = 73,343 Kg.

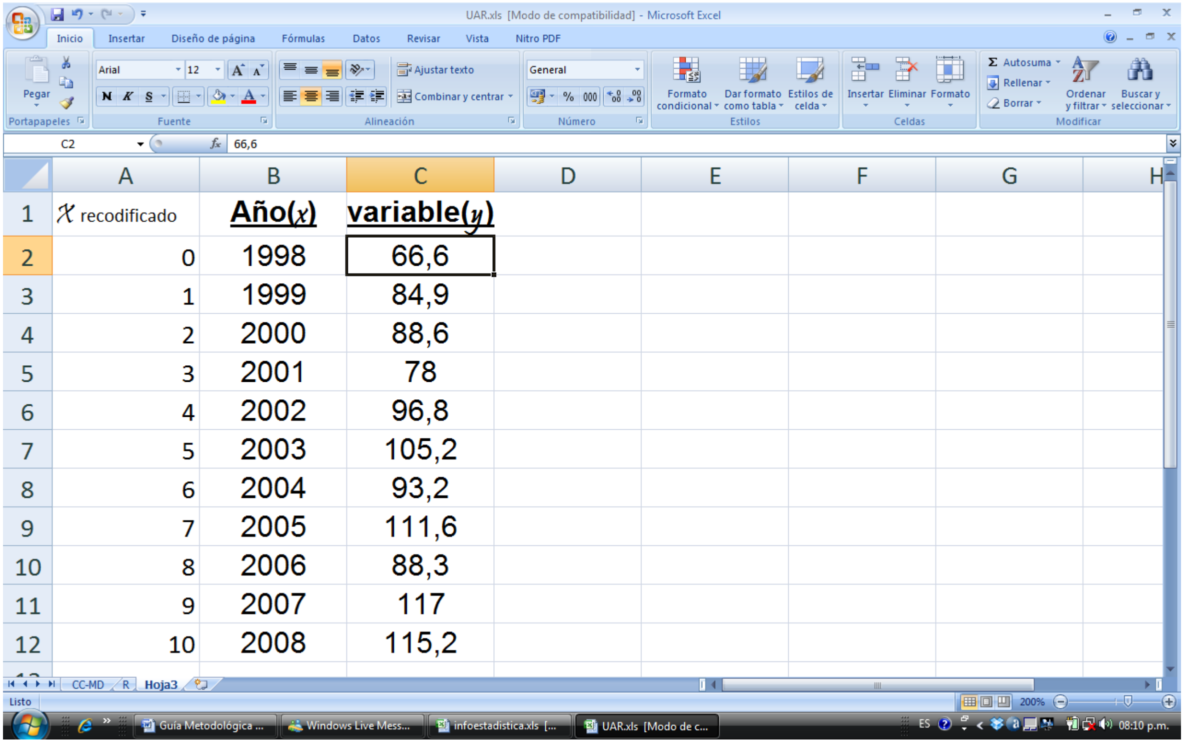
Para ver la diferencia entre un valor estimado según el método de regresión y un valor real de la serie de datos, estimemos y_{160}. La tendencia predice que un sujeto de 160 cm tendrá un peso de 60,470 Kg. Si comparamos con el valor real = a 60 kg. Vemos que hay muy poca diferencia, esto era esperable dado la bondad de ajuste del método en este caso.

En este momento es bueno recordar un axioma de correlación: *"La correlación no es sinónimo de causalidad".*

5.4. Análisis de tendencias de series cronológicas con la regresión lineal

Supongamos que tenemos una serie cronológica de una década que muestra el comportamiento de una variable "y" queremos predecir el valor de la dependiente en el año inmediato siguiente al último año de la serie.

Como no se puede usar los años como valor numérico, sino en el sentido nominal; se tiene que recodificar un valor para los mismos.



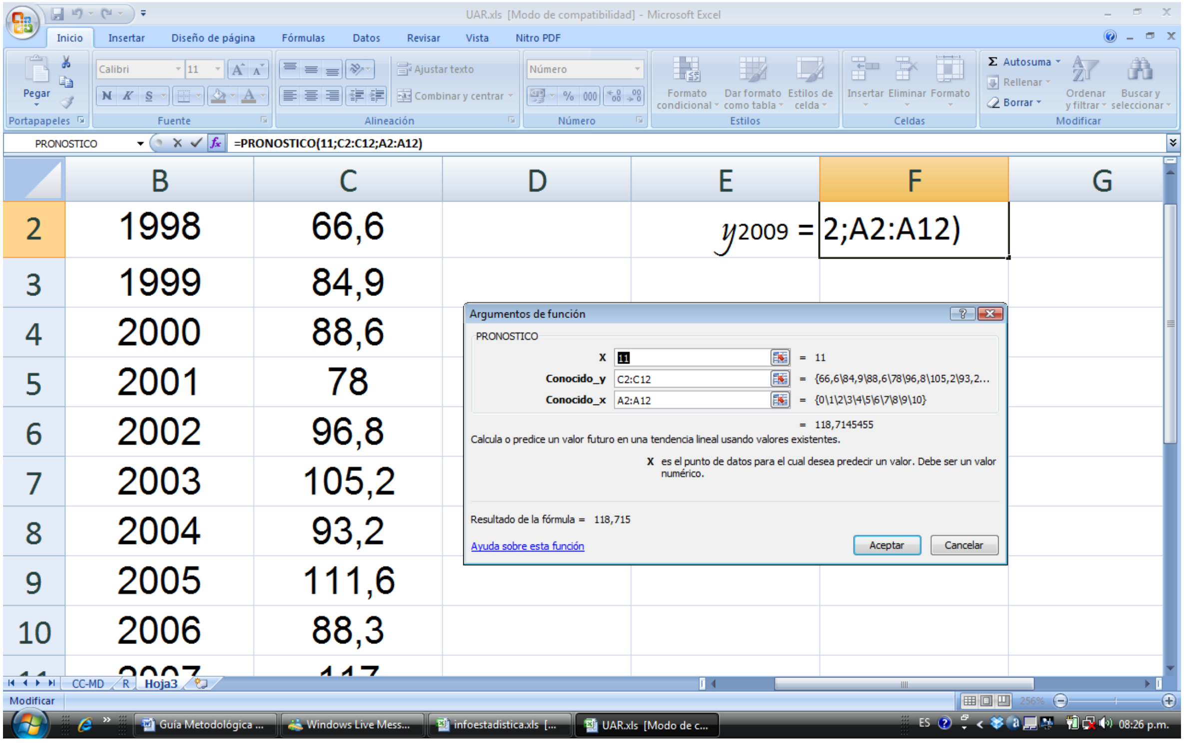
x recodificado	Año(x)	variable(y)
0	1998	66,6
1	1999	84,9
2	2000	88,6
3	2001	78
4	2002	96,8
5	2003	105,2
6	2004	93,2
7	2005	111,6
8	2006	88,3
9	2007	117
10	2008	115,2

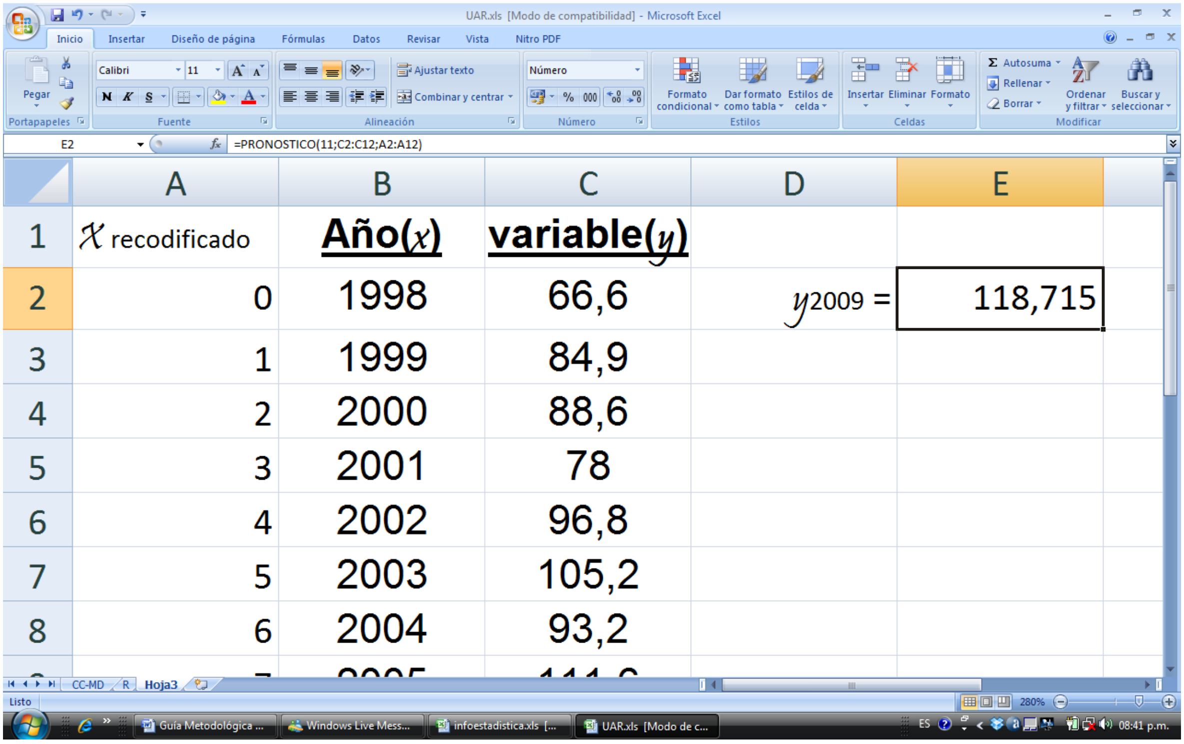
Luego procedemos como un análisis de regresión que ya estudiamos:

Así para estimar que valor tendrá "y" en el 2009 (valor $x_{recodificado}$ = 11); se usa la función "PRONOSTICO" ya conocida.

En las casillas "x"; "conocido_x" se usan los valores correspondientes de la variable $x_{recodificado}$ (11; A2:A12 respectivamente).

El y_{2009} = 118,715





6. Estructura de un Informe Estadístico

El Objetivo de este capítulo es estandarizar el formato de salida que debe tener el análisis estadístico realizado, conforme a la metodología sugerida en las secciones precedentes. Pues, así como la metodología analítica debe ser comparable, también lo debe ser la redacción de una producción científica.

Debemos pensar que el lector final, en la mayoría de los casos, no formará parte del equipo técnico encargado del análisis de la información; y sobre todo en el caso que tengan un nivel gerencial, para tomar las decisiones basadas en dicho informe técnico.

Por tanto, es muy importante mantener una estructura básica y uniforme de tal modo de conseguir una familiaridad con el modelo de reporte.

También es bueno recordar, que si bien tenemos que apegarnos a una correcta construcción gramatical, lo importante es el fondo. Se sugiere un lenguaje sobrio que transmita la información a un lector que estará más interesado en la interpretación y las conclusiones más que en el método de análisis usado.

Estamos escribiendo una redacción científica más que una obra literaria, donde buscamos a la vez traducir un lenguaje técnico, pero sin perder el rigor científico.

A continuación esquematizamos la estructura básica de un informe estadístico, junto con sugerencias para su correcto llenado.

Estructura Básica de un Informe[14]

Página 1	Página 2	Página 3 en adelante
• Título y Autor	• Resumen	• Introducción / justificación • Métodos • Resultados • Análisis • Discusión / Conclusiones • Bibliografía[15]

[14] Para información complementaria se recomienda leer" REQUISITOS DE UNIFORMIDAD PARA MANUSCRITOS PRESENTADOS A REVISTAS BIOMÉDICAS - Normas de Vancouver". www.upch.edu.pe/vrinve/doc/nvanco.htm.

[15] Si fuera necesario.

Título

Debe dar una idea clara del tema estudiado. En la medida de lo posible, cautivando el interés del lector, pero sin sacrificar por ello la sobriedad y el perfil técnico-científico que debe tener.

El titulo debe ser breve generalmente con un máximo de 40 caracteres (incluyendo letras y espacios); en caso de no poder hacerse así, conviene dividirlo en título y subtítulo, separados por dos puntos.

Debe ser lo más corto posible, preferiblemente sin interrogaciones ni exclamaciones y con carácter afirmativo.

Los autores deben incluir en el titulo toda la información que maximice la indexación; y la sensibilidad y especificidad en una revisión sistemática de la literatura médica.

Colocar en todas las hojas, junto con el número de página.

Autoría

Además de los nombres de los autores debe incluir sus referencias de contacto. Ejemplo: dirección, teléfono, correo electrónico, etc.

Para concederle a alguien el crédito de autor, hay que basarse únicamente en su contribución esencial en lo que se refiere a:

1) La concepción y el diseño del estudio, o recogida de los datos, o el análisis y la interpretación de los mismos.
2) Redacción del primer borrador o una revisión crítica importante de conceptos e interpretaciones en un borrador posterior.
3) Capacidad para defender y argumentar intelectualmente las conclusiones de la publicación, incluyendo la defensa de las evidencias.

Resumen

El propósito es dar al lector no especialista, los hechos más relevantes y las conclusiones del análisis, sin entrar en los detalles estadísticos. Tener presente que los lectores pueden pertenecer a distintas áreas, por lo cual quizá estén poco interesados en los aspectos técnicos del informe. Sin embargo, debe garantizársela inclusión de toda la información relevante.

Si bien el resumen precede al informe propiamente dicho, se lo escribe luego de haber finalizado el análisis, pues el contenido del resumen surge del informe dado que contiene sintetizados las cinco secciones del cuerpo del informe.
En general se escribe con los verbos en tiempo pasado.

Consejos útiles:

- El resumen no excederá de 250 palabras. Debe contener:
- Objetivos del estudio.
- Procedimientos básicos (selección de los sujetos del análisis, variables de estudio, métodos de observación y análisis).
- Resultados más destacados (mediante la presentación de datos concretos y, de ser posible, de su significación estadística).
- Principales conclusiones.
- Se hará hincapié en aquellos aspectos del estudio o de las observaciones que resulten más novedosos o de mayor importancia.
- Debe incluirse el resumen en inglés (abstract)[16].
- Evite el uso de referencias bibliográficas.

Cuerpo del Informe

Introducción

Explica la pertinencia (**porqué**) del estudio o análisis, y la necesidad de una acción estratégica. Describe la naturaleza, alcance y objetivos del problema.

Da a conocer los rasgos generales del estudio y motiva a los lectores para abordar el resto del trabajo.

Incluye:

- Justificación y alcance del estudio.
- Antecedentes
- Objetivos general y específico.

Métodos

Describe cómo se llegan a los resultados.

Debe permitir que otra persona pueda repetir la metodología de análisis.

Sintetiza el origen o la unidad de análisis o investigación.

Conceptualiza la definición de las variables.

Define las herramientas, procedimientos y métodos estadísticos usados para el análisis.

Incluye, cuando se trate de trabajos de revisión, una sección en la que se describen los métodos utilizados para localizar, seleccionar, recoger y sintetizar los datos. Estos métodos se describen también en el resumen del trabajo.

[16] Aunque esto es más propio de las redacciones de investigaciones originales.

Los métodos estadísticos deben describirse en detalle para que sean susceptibles de verificación por expertos.

Si se está trabajando con muestras, se menciona las características y tipo
Indicarse si se siguieron las normas de ética y confidencialidad.
Se redacta en tiempo pasado.
Por su importancia se recomienda leer esta sección en las Normas de Vancouver[17]

Resultados

Establece **que** se ha encontrado. Este es el paso en que se realiza el análisis estadístico propiamente dicho: descriptivo y/o inferencial (que se ha descripto en la sección "métodos").

Se aconseja presentar los cálculos y estimaciones con el rigor científico necesario para validar la prueba o método estadístico utilizado, pero no incluir información que no sea necesaria.

No interpretar los resultados hasta la sección siguiente.

Resumir y resaltar los resultados más relevantes para las conclusiones y sugerencias del análisis.

Presente los resultados en el texto, tablas y gráficos siguiendo una secuencia lógica, y sin repeticiones ni redundancias.

Las tablas y figuras deben limitarse solo a las estrictamente necesarias para explicar el argumento del análisis.

No deben presentarse solo frecuencias relativas (por ejemplo, porcentajes), sino también los valores absolutos de donde provienen.

Se redacta en tiempo pasado.

Análisis

Se interpreta integralmente los resultados en función a las características intrínsecas de los métodos y herramientas de análisis empleados.

Es una sección muy técnica, pero que incluso el lector poco familiarizado, pero interesado, lo leerá con el mayor detenimiento.

Es aquí donde se hace "hablar científicamente" al resultado extraído de la prueba o

[17] Normas de Vancouver www.upch.edu.pe/vrinve/doc/nvanco.htm.

método estadístico empleado. Por ejemplo (según la situación):

- El porqué de la interpretación del tipo de distribución y nivel simetría de la serie.
- Significado del valor numérico analizado con las medidas centrales, dispersión y posición; medidas de variabilidad del fenómeno; correlación, regresión.
- La significancia estadística según nivel de certeza, potencia del estudio (especialmente en los hallazgos no significativos), IC. El valor *p*
- Cumplimiento de supuestos o requisitos para métodos paramétricos.
- La interpretación y lectura crítica de cuadros y gráficos.
- Entre otros.

Discusión y Conclusiones

Esta sección probablemente, sea la más importante de la redacción, y es, con toda seguridad la más leída, juntamente con el resumen.

La importancia[18] de las conclusiones se observa al repetirlas (con diferentes niveles de complejidad) en el Resumen; Introducción y aquí, en Discusión.

Por todo esto, es la sección más difícil de escribir:

- Muchos documentos son rechazados a causa de una discusión deficiente
- Existen documentos en que la significación de los resultados no se expone o se expone insuficientemente.
- Muchos segmentos de la discusión resultan demasiado largos.

Esta sección tiene como propósito:

- Responder a la problematización manifestada en la introducción y el objetivo general de un trabajo de investigación.
- Explicar cómo los resultados apoyan la respuesta.
- Explicar como la respuesta encaja con el conocimiento existente.
- Comparar las observaciones realizadas con las de otros estudios pertinentes.

El contenido debe:

1. Constatar que la respuesta al problema planteado responda íntegramente la pregunta tal como se preguntó, y esté limitado a la población apropiada.
2. Apoyar la respuesta en los resultados propios, y si fuera apropiado (y necesario) en los resultados de otros. Esto debe estar relacionado con lo expuesto en la introducción.
    - Citar los gráficos y cuadros que apoyen la discusión

[18] Por su relevancia volvemos a referenciar la necesidad de lectura complementaria en las normativas de Vancouver.

- Citar referencias apropiadas de resultados ajenos.

3. Explicar por qué su respuesta es razonable y cómo encaja con lo ya publicado.
4. Defender su respuesta, incluyendo los argumentos a favor y en contra.
5. Explicar los resultados propios que no apoyan a su pregunta de investigación.
6. Establecer la novedad de su respuesta.
7. Explicar los hallazgos no esperados, y si caben las relaciones causales.
8. Explicar las limitaciones o debilidades del diseño o la metodología empleada.
9. Explicar la validez de las presunciones. Esto es particularmente importante en los experimentos.
10. Explicar la discrepancia en los resultados publicados.
11. Establecer la importancia, aplicación y alcance de las conclusiones.
12. Justificar, si fuera necesario, nuevos estudios (basados en diseños, refinamiento del método, otras poblaciones).

Formato y Estructuración de la discusión
Debe estar redactado en tiempo presente y ser objetivo.

- **Inicio**:

Hacerlo con algo interesante vinculado a la problematización central del estudio. “En el presente estudio hemos demostrado que… (conteste la pregunta y /o objetivo general del estudio)”.

- **Medio**:

Ir de los más importante a lo menos importante utilizando conectores gramaticales (encontramos que, específicamente).

- **Final**:

Dar el mensaje central: “En conclusión este estudio demuestra que...” “En resumen nuestros resultados indican que...”
Las conclusiones se escriben en el párrafo final del a Discusión, allí se hace una recapitulación puntualizada y breve de los hallazgos; una por cada objetivo planteado.

7. Bibliografía

1. Bancroft, Huldah. Introducción a la bioestadística/ Huldah Bancroft. Buenos Aires: Editorial Universitaria de Buenos Aires, 1971. 246 p.
2. Camel, Fayad. Estadística médica y de salud pública/ Fayad Camel. Venezuela-Mérida: Talleres gráficos universitarios, 1970. 528 p.
3. Centro Cochcrane Iberoamericano. Manual de la Colaboración Cochcrane / Centro Cochcrane Iberoamericano. Area de Bioestadística. España. 2004.
4. Espasa Calpe. Gran Enciclopedia Universal Espasa Calpe. Ed. Espasa Calpe SA. 2005.
5. G. Piédrola Gil et al. Medicina. Medicina Preventiva y Salud Pública. Barcelona. Ed. Masson – Salvat. 1991. 1476 pag.
6. Guzmán - Cólera - Salvador. Matemáticas / Miguel de Guzmán - José Cólera - Adela Salvador. Madrid - España: Anaya; 1987. 359 p.
7. Hill, Bradford. Principios de estadística médica/ Bradford Hill. 3a. ed. - Buenos Aires: El Ateneo, 1965. 365 p.
8. Javaloyes y González J. Manual de Estadística/ J. González y Javaloyes. Asunción. 1975, 199 p.
9. Jenicek M., Cleroux R. Epidemiología. Principios-Técnicas-Aplicaciones. Barcelona: Salvat; 1987.
10. Mallorga C. Tendencia de la Mortalidad y sus Determinantes. Revista Gerencia y Políticas de Salud. Colombia. 2004.
11. Moreno, A. Programa de Mejoramiento de la Calidad de la Información en Salud: Marco conceptual y metodológico. Asunción. USAID – CIRD.2010.
12. Moreno A., Rojas M., Galeano J. SIECS: Propuesta para el Sistema de Información en Salud. Asunción: USAID - CIRD. 2009.
13. Moreno, A. Apuntes de Bioestadística. En Prensa. 2007.
14. Norman y Streiner. Bioestadística / Geoffrey Norman - David Streiner. Madrid - España: Mosby / Doyma Libros SA; 1996. 260 p.
15. OMS. Informe mundial sobre el conocimiento orientado a mejorar la salud: Fortalecimiento de los sistemas sanitarios. OMS. Ginebra. 2004.
16. OMS. RMS Red de la Métrica de Salud. Hacia El Marco Y Estándares Para El

Desarrollo del Sistema De Información En Salud Del País. México. 2006.

17. Red de la Métrica de Salud. http://www.who.int/healthmetrics/

18. Rodríguez, Pedro Carvalho. Exercicios de estadística/ Pedro Carvalho Rodriguez. Río de Janeiro: Centro Pan-Americano de Febre Aftosa, 1976. 112 p.

19. Rovirosa, Mario. Enseñanza de la demografía en las facultades de medicina; Urbanización y salud/ Mario Robirosa. 3a. reim. Bogotá - Colombia: Federación Panamericana de Asociados de Facultades de Medicina, 1974. 149 p.

20. Ruiz - Morillo. Epidemiología Clínica: Investigación Clínica Aplicada. Bogotá. Ed. Médica Panamericana. 2004. 576 pag.

21. Sackett, D.L., Haynes, R.B., Guyatt, G.H., Tugwell, P. Epidemiología clínica. Ciencia básica para la medicina clínica. 2ª ed. Madrid: Editorial Médica Panamericana; 1994.

22. Salvat. Enciclopedia Salvat Diccionario. Salvat editores SA. Barcelona. 1972.

23. Tapia Conyer et al. El Manual de Salud Pública. México. Intersistemas. 2da. Ed. 2006.

24. Toranzos, Fausto. Iniciación en estadística aplicada/ Fausto Toranzos. Buenos Aires. Ed. Macchi, 1968. 225 p.

25. USAID, Management Sciences for Health, Quality Assurance Project, Health Systems 20/20. Health Systems Assessment Approach: A How-to Manual. Islam, M ed. USA. 2007.

26. Weintraub - Douglass - Gillings. Bioestadística en Salud Bucodental. 2ª. Ed. North Carolina - USA: CAVCO Publications; 1985. 316 p.

8. ANEXOS

Tabla de la "t" de Student

TABLA

Valores críticos del test de la t

	α para un test de una cola					
	0,10	0,05	0,025	0,01	0,005	0,0005
	α para un test de dos colas					
gl	0,20	0,10	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,820	63,656	636,615
2	1,886	2,920	4,303	6,965	9,925	31,599
3	1,638	2,353	3,182	4,541	5,841	12,942
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,408
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,141
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,849
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,768
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,310	1,697	2,042	2,457	2,750	3,646
35	1,306	1,690	2,030	2,438	2,724	3,591
40	1,303	1,684	2,021	2,423	2,704	3,551
45	1,301	1,679	2,014	2,412	2,690	3,520
50	1,299	1,676	2,009	2,403	2,678	3,496
55	1,297	1,673	2,004	2,396	2,668	3,476
60	1,296	1,671	2,000	2,390	2,660	3,460
70	1,294	1,667	1,994	2,381	2,648	3,435
80	1,292	1,664	1,990	2,374	2,639	3,416
90	1,291	1,662	1,987	2,368	2,632	3,402
100	1,290	1,660	1,984	2,364	2,626	3,390

Tabla del Chi Cuadrado

TABLA

Valores críticos para el test de chi-cuadrado

gl	α 0,10	0,15	0,25	0,01	0,005	0,001
1	2,706	3,842	5,024	6,635	7,879	10,828
2	4,605	5,992	7,378	9,210	10,597	13,816
3	6,251	7,815	9,348	11,345	12,838	16,266
4	7,779	9,489	11,143	13,277	14,860	18,467
5	9,236	11,071	12,833	15,086	16,750	20,515
6	10,645	12,592	14,449	16,812	18,548	22,457
7	12,017	14,067	16,013	18,475	20,278	24,321
8	13,362	15,507	17,535	20,090	21,955	26,124
9	14,684	16,919	19,023	21,666	23,589	27,877
10	15,987	18,307	20,483	23,209	25,188	29,588
11	17,275	19,675	21,920	24,725	26,757	31,264
12	18,549	21,026	23,336	26,217	28,299	32,909
13	19,812	22,362	24,736	27,688	29,819	34,528
14	21,964	23,685	26,120	29,141	31,319	36,123
15	22,307	24,996	27,488	30,578	32,801	37,697
16	23,542	26,296	28,845	32,000	34,267	39,252
17	24,769	27,587	30,191	33,409	35,718	40,790
18	25,989	28,869	31,526	34,805	37,156	42,312
19	27,204	30,144	32,852	36,191	38,582	43,820
20	28,412	31,410	34,170	37,566	39,997	45,314
21	29,615	32,671	35,479	38,932	41,401	46,797
22	30,813	33,924	36,781	40,289	42,796	48,268
23	32,007	35,172	38,076	41,638	44,181	49,728
24	33,196	36,415	39,365	42,980	45,558	51,178
25	34,382	37,652	40,647	44,314	46,928	52,620
26	35,563	38,885	41,924	45,642	48,290	54,052
27	36,741	40,113	43,195	46,963	49,645	55,476
28	37,916	41,337	44,461	48,278	50,993	56,892
29	39,087	42,557	45,723	49,588	52,336	58,301
30	40,256	43,773	46,980	50,892	53,672	59,703

Tabla de valores cíticos de la "r" de Pearson

TABLA

Valores críticos del coeficiente de correlación de Pearson (r)

	α para un test de una cola			
	0,05	0,025	0,01	0,005
	α para un test de dos colas			
gl	0,10	0,05	0,02	0,01
1	0,988	0,997	0,9995	0,9999
2	0,900	0,950	0,980	0,990
3	0,805	0,878	0,934	0,959
4	0,729	0,811	0,882	0,917
5	0,669	0,755	0,833	0,874
6	0,621	0,707	0,789	0,834
7	0,582	0,666	0,750	0,798
8	0,549	0,632	0,715	0,765
9	0,521	0,602	0,685	0,735
10	0,497	0,576	0,658	0,708
11	0,476	0,553	0,634	0,684
12	0,458	0,532	0,612	0,661
13	0,441	0,514	0,592	0,641
14	0,426	0,497	0,574	0,623
15	0,412	0,482	0,558	0,606
16	0,400	0,468	0,543	0,590
17	0,389	0,456	0,529	0,575
18	0,378	0,444	0,516	0,561
19	0,369	0,433	0,503	0,549
20	0,360	0,423	0,492	0,537
21	0,352	0,413	0,482	0,526
22	0,344	0,404	0,472	0,515
23	0,337	0,396	0,462	0,505
24	0,330	0,388	0,453	0,496
25	0,323	0,381	0,445	0,487
26	0,317	0,374	0,437	0,479
27	0,312	0,367	0,430	0,471
28	0,306	0,361	0,423	0,463
29	0,301	0,355	0,416	0,456
30	0,296	0,349	0,409	0,449
35	0,275	0,325	0,381	0,418
40	0,257	0,304	0,358	0,393
45	0,243	0,288	0,338	0,372
50	0,231	0,273	0,322	0,354
55	0,220	0,261	0,307	0,339
60	0,211	0,250	0,295	0,325
70	0,195	0,232	0,274	0,302
80	0,183	0,217	0,257	0,283
90	0,173	0,205	0,242	0,267
100	0,164	0,195	0,230	0,254
125	0,147	0,174	0,206	0,228
150	0,134	0,159	0,189	0,208
175	0,124	0,147	0,175	0,193
200	0,116	0,138	0,164	0,181
250	0,104	0,124	0,146	0,162

TABLA DE EQUIVALENCIA DE Z / AREA

Z	AREA	Z	AREA	Z	AREA	Z	AREA
0,00	0,0000	1,00	0,3413	2,00	0,4772	3,00	0,4987
0,01	0,0039	1,01	0,1437	2,01	0,4778	3,01	0,4987
0,02	0,0079	1,02	0,3461	2,02	0,4783	3,02	0,4987
0,03	0,0119	1,03	0,3485	2,03	0,4788	3,03	0,4988
0,04	0,0159	1,04	0,3508	2,04	0,4793	3,04	0,4988
0,05	0,0199	1,05	0,3531	2,05	0,4798	3,05	0,4989
0,06	0,0239	1,06	0,3554	2,06	0,4803	3,06	0,4989
0,07	0,0279	1,07	0,3577	2,07	0,4808	3,07	0,4989
0,08	0,0318	1,08	0,3599	2,08	0,4812	3,08	0,4990
0,09	0,0358	1,09	0,3621	2,09	0,4817	3,09	0,4990
0,10	0,0398	1,10	0,3643	2,10	0,4821	3,10	0,4990
0,11	0,0438	1,11	0,3665	2,11	0,4826	3,11	0,4991
0,12	0,0477	1,12	0,3686	2,12	0,4830	3,12	0,4991
0,13	0,0517	1,13	0,3708	2,13	0,4834	3,13	0,4991
0,14	0,0556	1,14	0,3729	2,14	0,4838	3,14	0,4992
0,15	0,0596	1,15	0,3749	2,15	0,4842	3,15	0,4992
0,16	0,0635	1,16	0,3770	2,16	0,4846	3,16	0,4992
0,17	0,0675	1,17	0,3790	2,17	0,4850	3,17	0,4992
0,18	0,0714	1,18	0,3810	2,18	0,4854	3,18	0,4993
0,19	0,0753	1,19	0,3830	2,19	0,4857	3,19	0,4993
0,20	0,0792	1,20	0,3849	2,20	0,4861	3,20	0,4993
0,21	0,0831	1,21	0,3869	2,21	0,4864	3,21	0,4993
0,22	0,0870	1,22	0,3888	2,22	0,4868	3,22	0,4994
0,23	0,0909	1,23	0,3906	2,23	0,4871	3,23	0,4994
0,24	0,0948	1,24	0,3925	2,24	0,4875	3,24	0,4994
0,25	0,0987	1,25	0,3943	2,25	0,4878	3,25	0,4994

0,26	0,1025	1,26	0,3962	2,26	0,4881	3,26	0,4994
0,27	0,1064	1,27	0,3980	2,27	0,4884	3,27	0,4995
0,28	0,1102	1,28	0,3997	2,28	0,4887	3,28	0,4995
0,29	0,1141	1,29	0,4015	2,29	0,4890	3,29	0,4995
0,30	0,1179	1,30	0,4032	2,30	0,4893	3,30	0,4995
0,31	0,1217	1,31	0,4049	2,31	0,4896	3,31	0,4995
0,32	0,1255	1,32	0,4066	2,32	0,4898	3,32	0,4995
0,33	0,1293	1,33	0,4082	2,33	0,4901	3,33	0,4996
0,34	0,1330	1,34	0,4099	2,34	0,4904	3,34	0,4994
0,35	0,1363	1,35	0,4115	2,35	0,4906	3,35	0,4996
0,36	0,1406	1,36	0,4131	2,36	0,4909	3,36	0,4996
0,37	0,1443	1,37	0,4147	2,37	0,4911	3,37	0,4996
0,38	0,1480	1,38	0,4162	2,38	0,4913	3,38	0,4996
0,39	0,1517	1,39	0,4177	2,39	0,4916	3,39	0,4997
0,40	0,1554	1,40	0,4192	2,40	0,4918	3,40	0,4997
0,41	0,1591	1,41	0,4207	2,41	0,4920	3,41	0,4997
0,42	0,1627	1,42	0,4222	2,42	0,4922	3,42	0,4997
0,43	0,1664	1,43	0,4236	2,43	0,4925	3,43	0,4997
0,44	0,1700	1,44	0,4251	2,44	0,4927	3,44	0,4997
0,45	0,1736	1,45	0,4265	2,45	0,4929	3,45	0,4997
0,46	0,1772	1,46	0,4279	2,46	0,4931	3,46	0,4997
0,47	0,1808	1,47	0,4292	2,47	0,4932	3,47	0,4997
0,48	0,1844	1,48	0,4306	2,48	0,4932	3,48	0,4997
0,49	0,1879	1,49	0,4319	2,49	0,4936	3,49	0,4998
0,50	0,1914	1,50	0,4332	2,50	0,4938	3,50	0,4998
0,51	0,1950	1,51	0,4345	2,51	0,4940	3,51	0,4998
0,52	0,1985	1,52	0,4357	2,52	0,4941	3,52	0,4998
0,53	0,2019	1,53	0,4370	2,53	0,4943	3,53	0,4998

0,54	0,2054	1,54	0,4382	2,54	0,4945	3,54	0,4998
0,55	0,2088	1,55	0,4394	2,55	0,4946	3,55	0,4998
0,56	0,2122	1,56	0,4406	2,56	0,4948	3,56	0,4998
0,57	0,2156	1,57	0,4418	2,57	0,4949	3,57	0,4998
0,58	0,2190	1,58	0,4429	2,58	0,4951	3,58	0,4998
0,59	0,2224	1,59	0,4441	2,59	0,4952	3,59	0,4998
0,60	0,2257	1,60	0,4452	2,60	0,4953	3,60	0,4998
0,6]	0,2291	1,61	0,4463	2,61	0,4955	3,61	0,4998
0,62	0,2324	1,62	0,4474	2,62	0,4956	3,62	0,4999
0,63	0,2356	1,63	0,4484	2,63	0,4957	3,63	0,4999
0,64	0,2389	1,64	0,4495	2,64	0,4959	3,64	0,4999
0,65	0,2421	1,65	0,4505	2,65	0,4960	3,65	0,4999
0,66	0,2454	1,66	0,4515	2,66	0,4961	3,66	0,4999
0,67	0,2486	1,67	0,4525	2,67	0,4962	3,67	0,4999
0,68	0,2517	1,68	0,4535	2,68	0,4963	3,68	0,4999
0,69	0,2549	1,69	0,4545	2,69	0,4964	3,69	0,4999
0,70	0,2580	1,70	0,4554	2,70	0,4965	3,70	0,4999
0,71	0,2611	1,71	0,4564	2,71	0,4966	3,71	0,4999
0,72	0,2642	1,72	0,4673	2,72	0,4967	3,72	0,4999
0,73	0,2673	1,73	0,4582	2,73	0,4968	3,73	0,4999
0,74	0,2703	1,74	0,4591	2,74	0,4969	3,74	0,4999
0,75	0,2734	1,75	0,4599	2,75	0,4970	3,75	0,4999
0,76	0,2764	1,76	0,4608	2,76	0,4971	3,76	0,4999
0,77	0,2793	1,77	0,4616	2,77	0,4972	3,77	0,4999
0,78	0,2823	1,78	0,4625	2,78	0,4973	3,78	0,4999
0,79	0,2852	1,79	0,4633	2,79	0,4975	3,79	0,4999
0,80	0,2881	1,80	0,4641	2,80	0,4974	3,80	0,4999
0,81	0,2910	1,81	0,4649	2,81	0,4975	3,81	0,4999

0,82	0,2932	1,82	0,4656	2,82	0,4976	3,82	0,4999
0,83	0,2967	1,83	0,4664	2,83	0,4977	3,83	0,4999
0,84	0,2995	1,84	0,4671	2,84	0,4977	3,84	0,4999
0,85	0,3023	1,85	0,4678	2,85	0,4978	3,85	0,4999
0,86	0,3051	1,86	0,4686	2,86	0,4979	3,86	0,4999
0,87	0,3078	1,87	0,4693	2,87	0,4979	3,87	0,4999
0,88	0,3106	1,88	0,4699	2,88	0,4980	3,88	0,4999
0,89	0,3133	1,89	0,4706	2,89	0,4981	3,89	0,4999
0,90	0,3159	1,90	0,4713	2,90	0,4981	3,90	0,5000
0,91	0,3186	1,91	0,4719	2,91	0,4982	3,91	0,5000
0,92	0,3212	1,92	0,4726	2,92	0,4982	3,92	0,5000
0,93	0,1238	1,93	0,4732	2,93	0,4983	3,93	0,5000
0,94	0,3264	1,94	0,4738	2,94	0,4984	3,94	0,5000
0,95	0,3289	1,95	0,4744	2,95	0,4984	3,95	0,5000
0,96	0,3315	1,96	0,4750	2,96	0,4985	3,96	0,5000
0,97	0,3340	1,97	0,4756	2,97	0,4985	3,97	0,5000
0,98	0,3365	1,98	0,4761	2,98	0,4985	3,98	0,5000
0,99	0,1389	1,99	0,4767	2,99	0,4986	3,99	0,5000
						4,00	0,5000