



GUIA PARA USAR INFOSTAT EN ESTADÍSTICA DESCRIPTIVA

Autores:
Mgter. María Inés Stímolo
Cra. Mariana Gonzalez
Cra. Olga Padró

GUIA PARA CLASE DE GABINETE- INFOSTAT .

TABLAS Y GRÁFICOS

Esta clase tiene como objetivo introducirlo en el uso de un paquete estadístico, para lo cual se ha seleccionado software Infostat¹, El diseño de Infostat es similar a otros softwares estadísticos, esto permite poder adaptarse al uso de los mismos fácilmente. Es importante que tenga en claro que solo la práctica permite un mejor conocimiento del software, por lo que esta clase no es suficiente para lograr el conocimiento completo del programa.

Infostat cuenta con un manual de ayuda, para acceder al mismo se debe elegir en el menú AYUDA la opción MANUAL.

Análisis de una base de datos

Para trabajar en esta clase se utilizarán los resultados de las encuestas de la cátedra, los que se encuentran tabulados en el archivo alumnos.xls. (Infostat permite trabajar con bases creadas con otros programas como Excel). Para abrir el archivo existente seleccione menú ARCHIVO → ABRIR y en el recuadro que se despliega busque el archivo alumnos.xls y marque la opción ABRIR. Se abrirá una ventana con el título SELECCIONE LA CELDA INICIAL, en la cual debe elegir HOJAS EN EL LIBRO→BASE PRACTICOS

ACTIVIDAD 1

A partir del archivo alumnos:

- a) Describa la población de interés y la unidad estadística.
- b) Identifique las variables estudiadas. Clasifíquelas e indique la escala de medición

Análisis de una variable cualitativa o categórica

Antes de construir la tabla, debemos observar cómo está expresada en la base la variable. En general cuando se arma una base de datos, se realiza una codificación, esto significa que las categorías de las variables cualitativas son reemplazadas por números para agilizar la tarea. INFOSTAT requiere que le especifiquemos el código que identifica cada categoría (si no lo hacemos interpreta que se trata de una variable numérica), esto se realiza pintando la columna que contiene la variable SEXO y clickeando con el botón derecho de mouse sobre la misma columna, para que aparezca el menú emergente. En dicho menú debe seleccionar

CATEGORIZAR →ASIGNAR CATEGORIAS SEGÚN CODIGO DE VALORES

Designamos código 1 VARON código 2 MUJER

Luego aparece una nueva columna CatSEXO con los nombres de las categorías, esta columna es la que usamos para realizar las tablas y gráficos.

A continuación se explican los pasos para la construcción de una tabla de frecuencias para la variable sexo:

- 1) Se debe marcar ESTADISTICAS → TABLA DE FRECUENCIA

En la solapa VARIABLES seleccionar la variable CatSEXO (para elegir una variable marcar la variable y con la flecha pasarla al cuadro de la derecha).

¹ La versión estudiantil es libre y está disponible en la página : <http://www.infostat.com.ar/>, solicitar una clave gratuita ingresando por “Clave gratuita para versión estudiantil”.

En la ventana TABLA DE FRECUENCIAS marcar FA y FR para obtener una tabla con las frecuencias absolutas y relativas, como la siguiente:

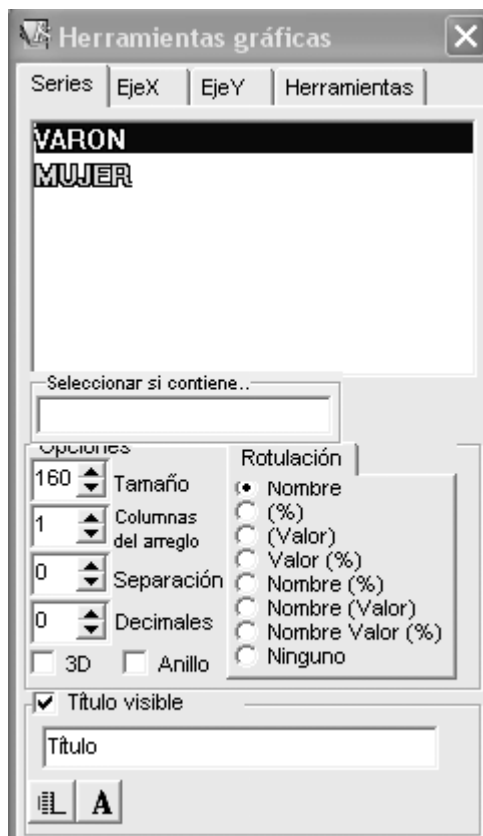
Tablas de frecuencias

Variable	Clase	FA	FR
Cat_SEXO	1	41	0,38
Cat_SEXO	2	68	0,62

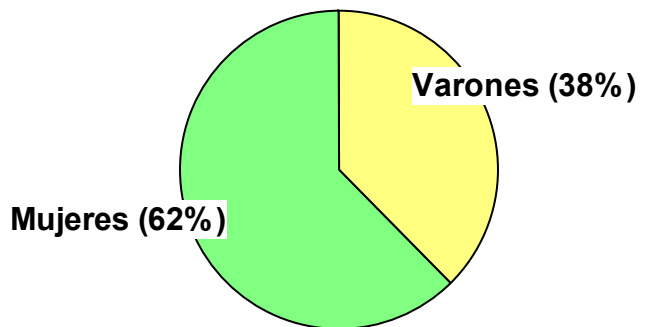
ACTIVIDAD 2

Organice adecuadamente los datos en una distribución de frecuencias para la variable LUGAR DE PROCEDENCIA. Recuerde codificar previamente la variable (1- Córdoba Capital 2- Provincia de Cba. 3- Otras provincias.

Para graficar estas variables se debe marcar GRAFICOS seleccionando el gráfico que desee. Para realizar un gráfico de sectores, por ejemplo, indicar la variable que va a graficar y especificar que las categorías están en filas. Luego con el gráfico aparece una ventana HERRAMIENTAS GRAFICAS que permite modificar el color, título, Etc.



Distribución de frecuencias por sexo



ACTIVIDAD 3

Realice un gráfico adecuado para la variable LUGAR DE PROCEDENCIA.

Análisis de una variable numérica discreta

En esta sección trabajaremos con la variable NOTA DE MATEMÁTICA I (N_MATEI). Como se trata de una variable discreta, que tiene pocos valores, no es necesario definir intervalos. Antes de construir la tabla, debemos asegurarnos de que la variable esté definida como discreta. Para ello pinte la columna que contiene la variable NOTA DE

MATEMÁTICA I y con el botón derecho del mouse haga click sobre la misma columna, para que aparezca el menú emergente. En dicho menú debe seleccionar

TIPO DE DATO → DISCRETO

Para la construcción de la tabla de distribución de frecuencias de esta variable siga estos pasos:

- 1) Se debe marcar ESTADÍSTICAS → TABLA DE FRECUENCIA

En la solapa VARIABLES seleccionar la variable N_MATEI (para elegir una variable marcar la variable y con la flecha pasarla al cuadro de la derecha).

- 2) En la ventana TABLA DE FRECUENCIAS marcar MC (marca de clase), FA, FR, FAA (frecuencias absolutas acumuladas) y FRA (frecuencias relativas acumuladas), para obtener una tabla como la siguiente:

Tablas de frecuencias

Variable	Clase	MC	FA	FR	FAA	FRA
N_MATEI	1	4	9	0,08	9	0,08
N_MATEI	2	5	11	0,10	20	0,18
N_MATEI	3	6	15	0,14	35	0,32
N_MATEI	4	7	24	0,22	59	0,54
N_MATEI	5	8	18	0,17	77	0,71
N_MATEI	6	9	13	0,12	90	0,83
N_MATEI	7	10	19	0,17	109	1,00

ACTIVIDAD 4

Organice adecuadamente los datos en una distribución de frecuencias para la variable CANTIDAD DE APLAZOS.

Análisis de una variable numérica continua

Para realizar la tabla de frecuencias para la variable estatura siga los siguientes pasos:

En INFOSTAT las tablas de frecuencias pueden, de acuerdo a los campos activados por el usuario, contener la siguiente información: límites inferiores (LI) y superiores (LS) de los intervalos de clase, marca de clase (MC), frecuencias absolutas (FA), frecuencias relativas (FR), frecuencias absolutas acumuladas (FAA) y frecuencias relativas acumuladas (FRA). El número de clases, puede ser obtenido en forma automática o definido por el usuario (PERSONALIZADO). Para la forma automática InfoStat obtiene el número de clases tomando el $\log_2(n+1)$. Para el caso personalizado, InfoStat permite especificar el mínimo, máximo y número de intervalos. Los intervalos que construye son cerrados a la derecha.

Para realizar la tabla de frecuencias para la variable estatura siga los siguientes pasos:

- 3) Se debe marcar ESTADÍSTICAS → TABLA DE FRECUENCIA
- 4) En la solapa VARIABLES elegir ESTATURA
- 5) En la ventana TABLA DE FRECUENCIAS marcar LI, LS, MC, FA, FR, FAA y FRA. Además indicar NUMERO DE CLASES PERSONALIZADO y poner 5 para obtener una tabla con las frecuencias absolutas y relativas, como la siguiente:

Tablas de frecuencias

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
ESTATURA	1	150,00	158,00	154,00	10	0,09	10	0,09
ESTATURA	2	158,00	166,00	162,00	27	0,25	37	0,34

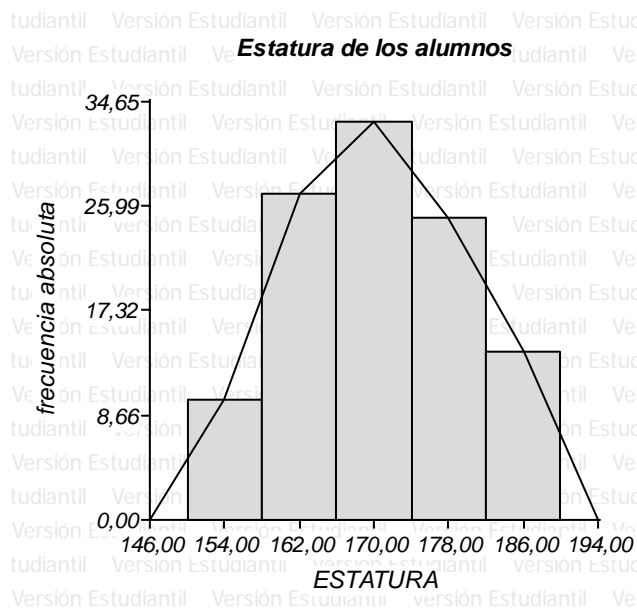
ESTATURA	3	166,00	174,00	170,00	33	0,30	70	0,64
ESTATURA	4	174,00	182,00	178,00	25	0,23	95	0,87
ESTATURA	5	182,00	190,00	186,00	14	0,13	109	1,00

ACTIVIDAD 5

Organice adecuadamente los datos en una distribución de frecuencias para la variable PESO

Para graficar estas variables se debe marcar GRAFICOS seleccionando HISTOGRAMA. Al construir un histograma, la ventana HERRAMIENTAS GRÁFICAS muestra un diálogo que permite modificar los atributos del histograma obtenido. En la solapa SERIES de esta ventana, hay un menú de opciones de histograma que permite cambiar el número de clases (Clases) que por defecto son calculadas como $\text{Log}_2(n+1)$. Además la ventana permite elegir la frecuencia representada en el histograma (Frec.). Las frecuencias a graficar pueden ser: frecuencia relativa (Frec. rel) que es la opción por defecto, frecuencia absoluta (Frec. abs.), frecuencia absoluta acumulada (Frec.abs. acum.) y frecuencia relativa acumulada (Frec. rel.acum.). El campo Bordes permite eliminar los contornos de las barras que forman el histograma. Se puede construir el polígono de frecuencia activando el campo Polígono. El campo Cuerpo permite eliminar el histograma de fondo a partir del cual se dibujó el polígono. Los campos LIPC y LSUC permiten ingresar los límites inferior y superior para la primera y última clase respectivamente. Para lograr que los "ticks" se correspondan con las marcas de clase de cada intervalo activar M. clases en la solapa Series.

En el siguiente gráfico se eligieron frecuencias absolutas, 5 clases y polígono:



ACTIVIDAD 6

Construya un histograma para la variable PESO. Realice un análisis del mismo.

Análisis de la variable estatura.

Para calcular las medidas descriptivas de la variable estatura, la secuencia a seguir es:

ESTADÍSTICAS → MEDIDAS RESUMENES

Luego de ingresar la variable y presionar aceptar, aparece la ventana ESTADÍSTICA DESCRIPTIVA que permite seleccionar las medidas a calcular, en nuestro caso elegimos las que se pueden observar en la siguiente salida:

Estadística descriptiva
Resumen ESTATURA

n	109,00
Media	170,65
D.E.	9,03
Var(n-1)	81,49
E.E.	0,86
CV	5,29
Mín	150,00
Máx	190,00
Mediana	170,00
Q1	165,00
Q3	177,00
Suma	18601,00
Asimetría	0,10
<u>Kurtosis</u>	<u>-0,51</u>

ACTIVIDAD 7 : Interprete cada medida que resulta de la tabla e indique su fórmula de cálculo.

ESTATURA			
Media		Curtosis	
Error típico		Coficiente de asimetría	
Mediana		Rango	
Moda		Mínimo	
Varianza de la muestra		Máximo	
Desviación estándar		Suma	

ACTIVIDAD 8:

A partir de las medidas descriptivas anteriores analice la simetría de la distribución. (Utilice todas las alternativas posibles)

ACTIVIDAD 9: Encuentre los valores del primer y tercer cuartil para la variable estatura. Calcule el recorrido intercuartílico

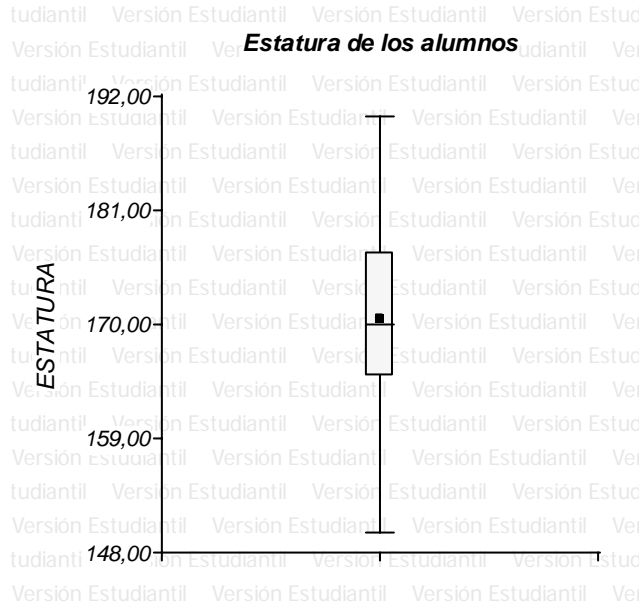
Q₁:..... Q₃:..... IQ:
.....

Otras funciones relacionadas con estas medidas de posición no centrales son :

- K.ESIMO MAYOR, devuelve el k-ésimo mayor valor de un conjunto de datos. Permite encontrar valores con una posición relativa específica dentro de un conjunto de datos.
- K.ESIMO MENOR, devuelve el k-ésimo menor valor de un conjunto de datos.

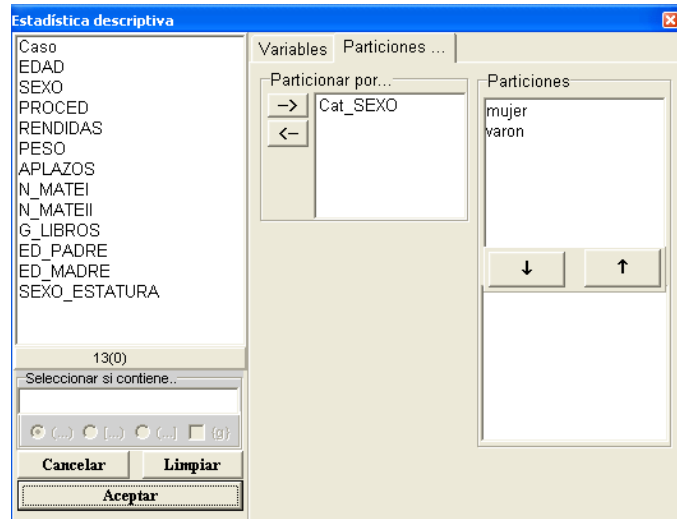
- PERCENTIL, devuelve el k-ésimo percentil de los valores de un rango

Para construir el diagrama de Caja y Brazos ingrese a GRÁFICOS → GRÁFICO DE CAJAS (BOX-PLOT)



Análisis de la variable estatura por sexo.

Se analizará la variable ESTATURA para el grupo de varones y el de mujeres. Para ello se debe ingresar en ESTADÍSTICAS→MEDIDAS RESUMENES y seleccionar la variable ESTATURA en la solapa VARIABLES y en la solapa PARTICIONES la variable CAT_SEXO como se observa a continuación:

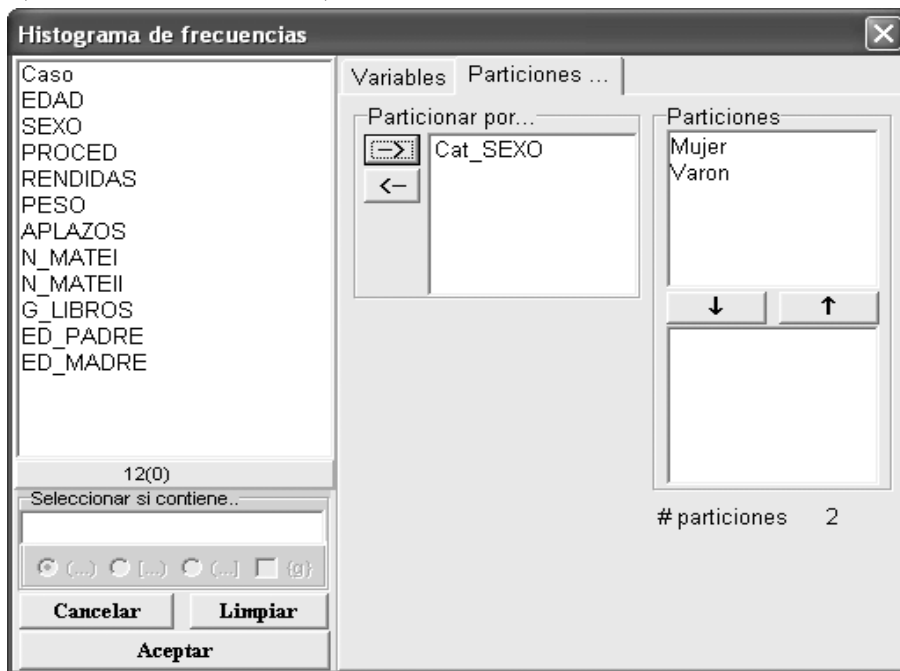


Las medidas para cada grupo son:

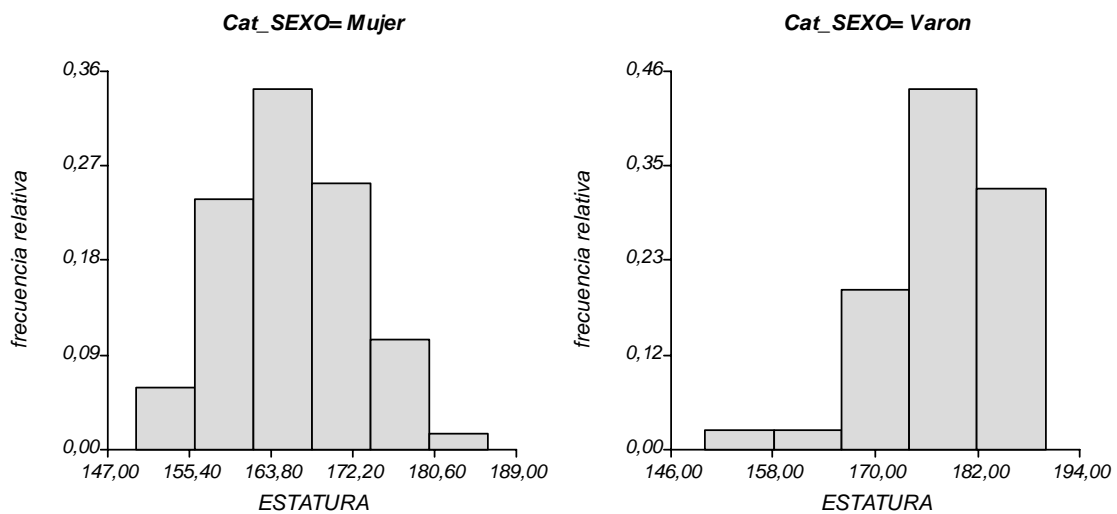
Estadística descriptiva		
Cat_SEXO	Resumen	ESTATURA
mujer	n	68,00
mujer	Media	166,40
mujer	D.E.	6,73
mujer	Var(n-1)	45,32
mujer	E.E.	0,82
mujer	CV	4,05
mujer	Mín	150,00
mujer	Máx	186,00
mujer	Mediana	166,00
mujer	Q1	161,00
mujer	Q3	170,00
mujer	Suma	11315,00
mujer	Asimetría	0,20
mujer	Kurtosis	0,22
varon	n	41,00

varon	Media	177,71
varon	D.E.	7,89
varon	Var(n-1)	62,31
varon	E.E.	1,23
varon	CV	4,44
varon	Mín	150,00
varon	Máx	190,00
varon	Mediana	178,00
varon	Q1	175,00
varon	Q3	184,00
varon	Suma	7286,00
varon	Asimetría	-1,05
varon	Kurtosis	2,35

Para obtener los histograma de la variable ESTATURA para cada grupo, dentro de la ventana HISTOGRAMA, seleccione la solapa PARTICIONES e ingrese la variable de interés, en este caso CatSEXO, tal como se muestra a continuación:



De esta manera el programa construirá un histograma para varones y otro para mujeres, como los que se muestran a continuación:



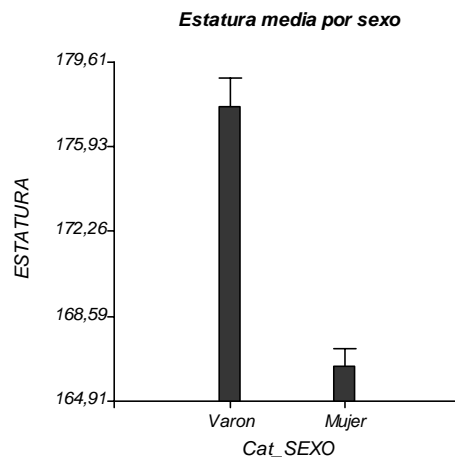
ACTIVIDAD 10: Encuentre y analice las medidas descriptivas para la variable PESO, según SEXO.

ACTIVIDAD 11: A partir de la media de cada grupo (varones y mujeres) reconstruya la media general de la variable PESO para todos los alumnos. ¿Qué propiedad aplicó?

Para obtener los gráficos de caja y brazos de la variable ESTATURA para cada grupo, dentro de la ventana GRÁFICO DE CAJAS (BOX-PLOT), seleccione la solapa PARTICIONES e ingrese la variable de interés, en este caso CatSEXO.

ACTIVIDAD 12: Construya el diagrama de caja y brazo de la variable ESTATURA para cada grupo. Interprete.

También es posible realizar un gráfico de barras que muestre la estatura promedio por sexo. Para ello, ingrese a GRÁFICOS → GRÁFICO DE BARRAS. Como Variable seleccione ESTATURA y como Criterio de Clasificación, seleccione CatSEXO. Al presionar Aceptar aparece una nueva pantalla donde debe seleccionar Media. El gráfico obtenido se muestra a continuación:



Análisis de la variable Gastos en libros.

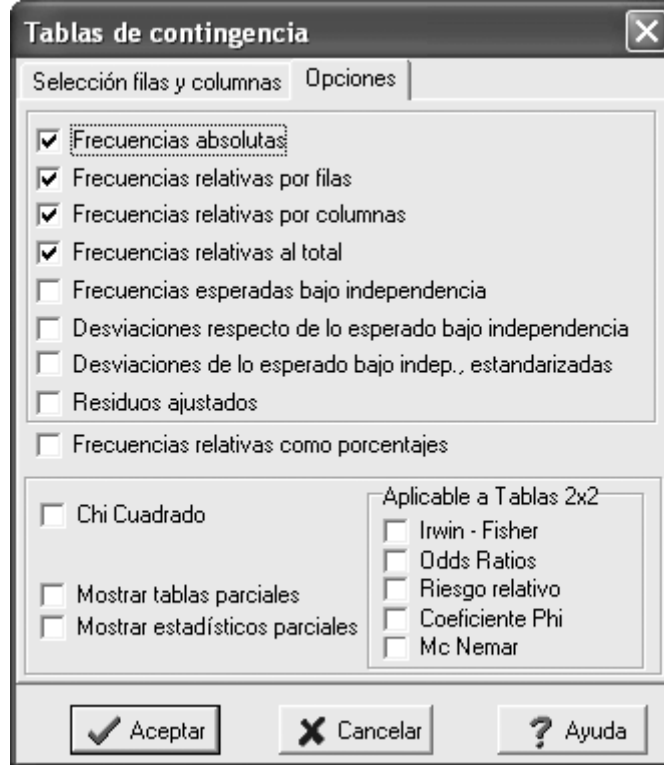
ACTIVIDAD 13: Calcule las medidas descriptivas de la variable GASTO EN LIBROS. Analice los resultados obtenidos.

ACTIVIDAD 14: Considerando que el gasto en los libros aumentó un 30% ¿cual será el nuevo gasto promedio en libros? Analice cómo afecta este aumento la varianza y el coeficiente de variación.

Análisis conjunto de la variable sexo y procedencia.

ESTADÍSTICAS → DATOS CATEGORIZADOS → TABLAS DE CONTINGENCIAS

Aparece una ventana en la que se deben desplazar las variables Cat-PROCED y Cat-SEXO a la ventana CRITERIOS DE CLASIFICACION de la solapa VARIABLES. Luego al aceptar aparece una ventana TABLA DE CONTINGENCIA con la solapa SELECCIÓN DE FILAS Y COLUMNAS para indicar cuál de las variables debe constituir las filas de la tabla y cuál las columnas. En la solapa OPCIONES se puede elegir las frecuencias como se observa a continuación:



Desactive las opciones Chi-cuadrado, Irwin-Fisher, Odds Ratios y Coeficiente Phi, que están activadas por defecto.

Las tablas obtenidas son:

Tablas de contingencia

Frecuencias absolutas

En columnas:Cat_SEXO

<u>Cat_PROCED</u>	<u>Mujer</u>	<u>Varon</u>	<u>Total</u>
Cba. Capital	33	22	55
Otras pcias.	10	8	18
Pcia. de Cba.	25	11	36
Total	68	41	109

Frecuencias relativas por filas

En columnas:Cat_SEXO

<u>Cat_PROCED</u>	<u>Mujer</u>	<u>Varon</u>	<u>Total</u>
Cba. Capital	0,60	0,40	1,00
Otras pcias.	0,56	0,44	1,00
Pcia. de Cba.	0,69	0,31	1,00
Total	0,62	0,38	1,00

Frecuencias relativas por columnas

En columnas:Cat_SEXO

<u>Cat_PROCED</u>	<u>Mujer</u>	<u>Varon</u>	<u>Total</u>
Cba. Capital	0,49	0,54	0,50
Otras pcias.	0,15	0,20	0,17
Pcia. de Cba.	0,37	0,27	0,33
Total	1,00	1,00	1,00

Frecuencias relativas al total

En columnas:Cat_SEXO

<u>Cat_PROCED</u>	<u>Mujer</u>	<u>Varon</u>	<u>Total</u>
-------------------	--------------	--------------	--------------

Cba. Capital	0,30	0,20	0,50
Otras pcias.	0,09	0,07	0,17
Pcia. de Cba.	0,23	0,10	0,33
Total	0,62	0,38	1,00

ACTIVIDAD 15:

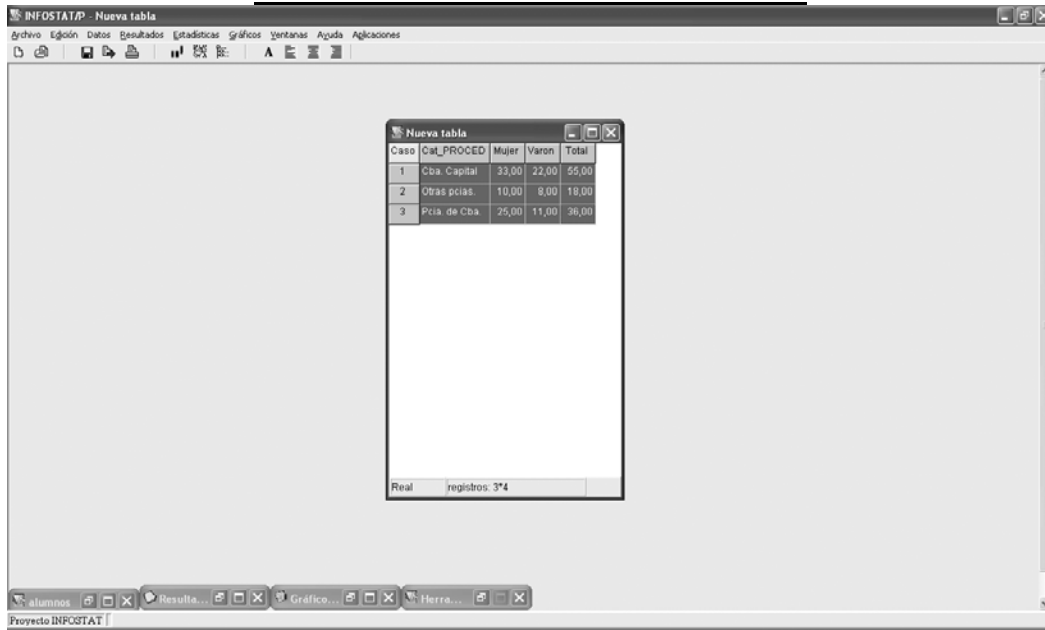
Interprete las frecuencias señaladas en negrita.

Para representar gráficamente la información conjunta de SEXO y LUGAR DE PROCEDENCIA, trabajaremos de la siguiente manera:

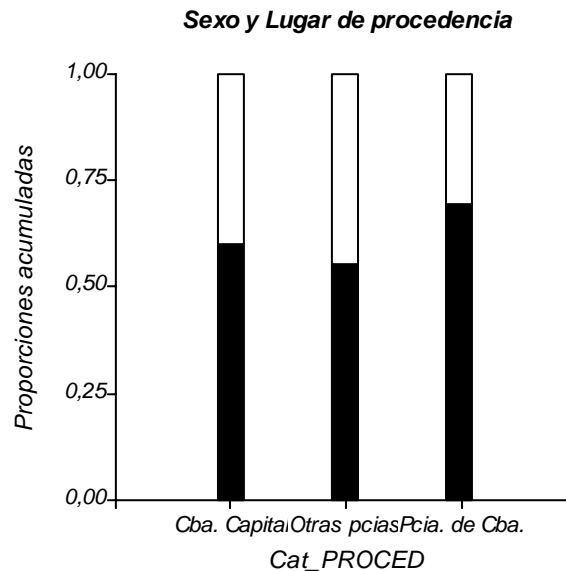
- 1) En la salida, seleccionar las filas que se muestran remarcadas, abrir una nueva tabla y pegar con nombre de columna, tal como se muestra en la pantalla.

Frecuencias absolutas
En columnas: Cat_SEXO

Cat_PROCED	Mujer	Varon	Total
Cba. Capital	33	22	55
Otras pcias.	10	8	18
Pcia. de Cba.	25	11	36
Total	68	41	109

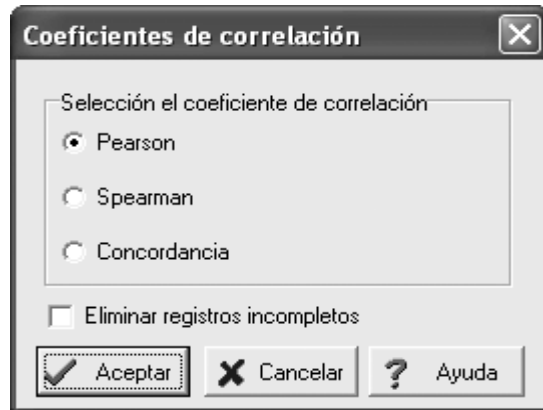


- 2) Ingrese a GRÁFICOS → GRÁFICO DE BARRAS APILADAS. Como Variables seleccione Varón y Mujer y como Criterio de Clasificación, seleccione Cat_PROCED.. El gráfico obtenido se muestra a continuación:



Análisis conjunto de la variable peso y estatura

Para calcular el coeficiente de correlación entre PESO y ESTATURA ingrese a ESTADÍSTICAS → ANÁLISIS DE CORRELACIÓN → COEFICIENTES DE CORRELACIÓN. Al ingresar las variables presionar Aceptar aparece una ventana como la siguiente:



Deje seleccionado Pearson y obtendrá la siguiente salida:

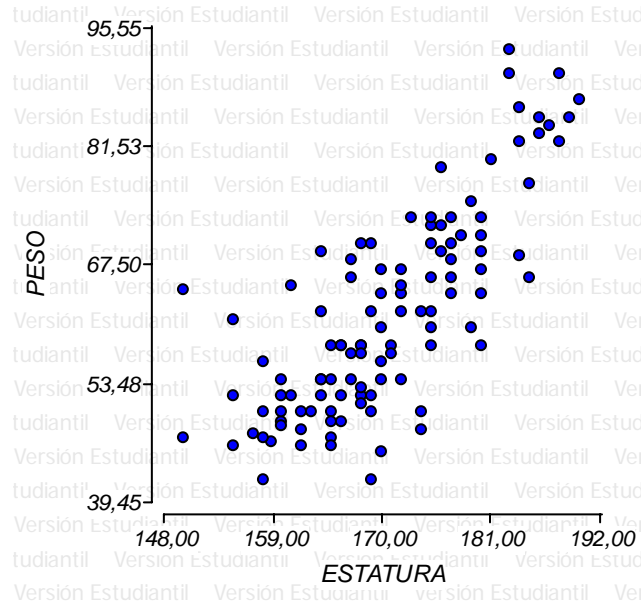
Coeficientes de correlación

Correlacion de Pearson: coeficientes\probabilidades

	<u>PESO</u>	<u>ESTATURA</u>
<u>PESO</u>	1,00	0,00
<u>ESTATURA</u>	0,78	1,00

La salida es una matriz que se denomina de correlación. En la diagonal principal aparecen unos. Fuera de la diagonal principal aparece el coeficiente de correlación entre las dos variables estudiadas, que en el ejemplo es 0,78.

Se puede realizar un gráfico de dispersión como el siguiente desde el menú de GRAFICOS →DIAGRAMA DE DISPERSION



ACTIVIDAD 16:

Analice la correlación existente entre las variables PESO y ESTATURA en base al coeficiente calculado y al diagrama de dispersión.