

4.4 TEPAS: taller de estadística para administradores sanitarios



Se trata de analizar una base de datos ficticia correspondiente a tres unidades clínicas para ensayar habilidades de análisis de los datos en el programa EPI-INFO, y reforzar conceptos estadísticos y métodos de comparación.

Autores: Juan de Mata Donado Campos

Investigador y Docente del Centro Nacional de Epidemiología del ISCIII

/ José R. Repullo

Profesor de la Escuela Nacional de Sanidad¹

¹ Este tema se ha construido sobre la base de otra versión anterior en la que también participaron Joaquín Pereira (+), Ana González-Estéfani, Juana Casas y Beltrán Rubio

Se recomienda imprimir 2 páginas por hoja

Citación recomendada:

Donado Campos J. de M. Repullo Labrador J.R. TEPAS: taller de estadística para administradores sanitarios [Internet]. Madrid: Escuela Nacional de Sanidad; 2012 [consultado día mes año]. Tema 4.4. Disponible en: dirección url del pdf.



TEXTOS DE ADMINISTRACIÓN SANITARIA Y GESTIÓN CLÍNICA
by UNED Y ESCUELA NACIONAL DE SANIDAD
is licensed under a Creative Commons
Reconocimiento- No comercial-Sin obra Derivada
3.0 Unported License.



Introducción

Bienvenido a este Taller de Estadística para Administradores Sanitarios (TEPAS). Si usted tiene una formación amplia en manejo de bases de datos y paquetes de análisis estadístico, este tema será posiblemente sencillo para usted, y lo estudiará rápidamente. Si no es así, enhorabuena: el taller está particularmente diseñado para usted.

No se asuste; a continuación le ofrecemos una vía poco dolorosa (no nos atrevemos a decir que divertida), para repasar concep-

tos estadísticos, y entrenar sus habilidades en la utilización o aplicación de dichos conceptos al análisis de datos.

El objetivo del taller es analizar una base de datos. Tenemos datos de 500 pacientes, que con la misma enfermedad, han sido sometidos a dos tipos de intervención quirúrgica (una convencional y otra laparoscópica) por tres servicios de tres hospitales diferentes. Los datos los tenemos una base de datos de Epi-Info para Windows llamada TEPAS_2012.MDB que se encuentra en la misma carpeta desde donde usted ha descargado esta Unidad Didáctica. Cree una carpeta de trabajo en el escritorio (por ejemplo TEPAS) y copie en ella ese fichero. Es conveniente que lo haga así, porque los resultados de todas las operaciones que haga se irán guardando en este fichero del escritorio, en el mismo donde está la base de datos TEPAS_2012.

Introducción

1- Conceptos generales y variables

2- Visualización de la matriz de datos

3- Descripción de las variables cualitativas: porcentajes

4- Descripción de las variables cuantitativas

5- Comenzamos el análisis... reflexionando

6- Estudiamos la relación de algunas variables

7- El análisis de estancias por paciente de los tres servicios

8- La evaluación de tecnologías: ¿qué cirugía gasta menos estancias?

9- Correlación costes y estancias.

10- Las complicaciones

Epílogo

Anexo: soluciones

Siga las instrucciones del taller, y vaya resolviendo en cada etapa los problemas y aspectos que se le plantean; hemos intentado que la guía sea comprensible (hasta el límite donde divulgación deviene en vulgarización), y completa (hasta el límite donde la prolijidad deviene en neurosis); sabemos por experiencia que

nuestro consejo de ir paso a paso leyendo cuidadosamente las instrucciones será escasamente atendido (especialmente cuando el alumno se sienta ante la pantalla del ordenador), pero aun así queremos hacerlo constar.

RECORDATORIO DE LA INSTALACIÓN DE EPI INFO PARA WINDOWS

La Unidad Didáctica 4.3 versa sobre el manejo del programa EPI-info. Puede usted revisarla si no lo ha hecho antes. En todo caso, se precisa bajar el programa y activarlo para poder desarrollar este Taller. La dirección para descargarse gratuitamente el programa es <http://wwwn.cdc.gov/epiinfo/html/downloads.htm>.

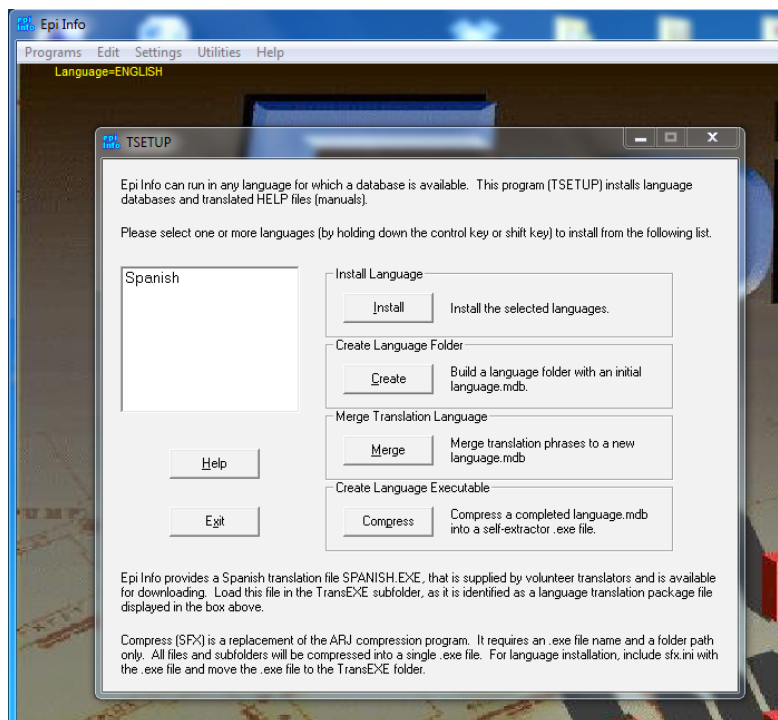
Usaremos en este taller la versión 3.5.4 (aunque hay una versión 7). Descárguese el archivo DOWNLOAD EPI INFO, y luego active el fichero de SET UP que ha descargado (EpiInfoSetup3_5.exe) ; se activará el programa de instalación de Epi-Info. Siga las instrucciones que aparecen en la pantalla.

Aparecerá un acceso directo de Epi-Info en el escritorio



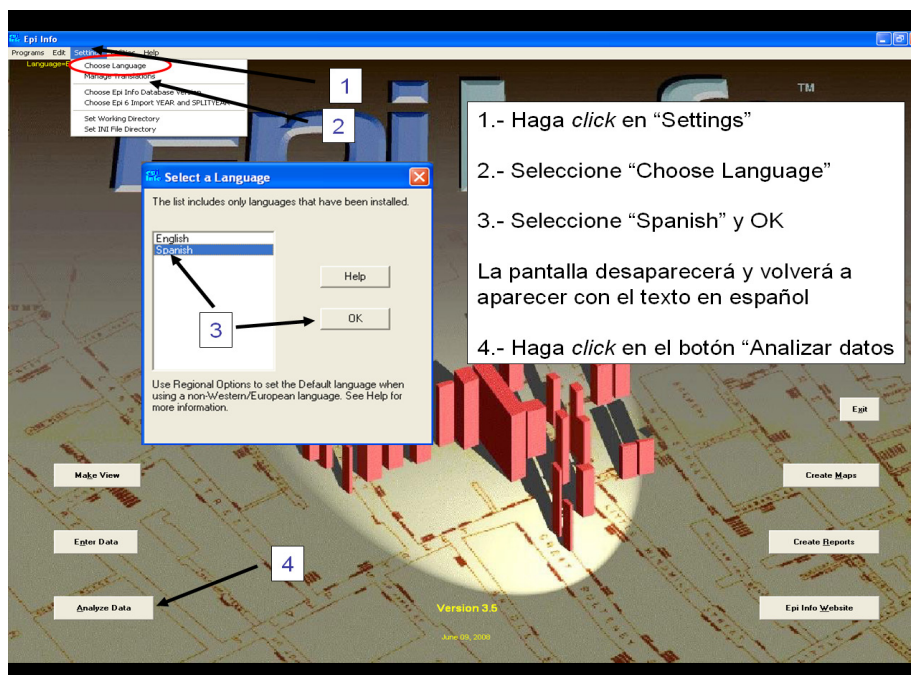
Haga doble click y aparecerá la pantalla principal de Epi-Info. Por defecto estará en inglés. Para cambiar el idioma, hay que hacer lo siguiente:

- 1- Bajarse la traducción al español: ftp://ftp.cdc.gov/pub/software/epi_info/Spanish/spanish.exe
- 2- Este fichero se ubicará en el directorio de DESCARGAS del usuario; hay que buscarlo y copiarlo, porque habrá que pegarlo dentro de EPI INFO:
- 3- Los ficheros de EPI INFO se habrán ubicado por defecto en el disco duro C: (o D:) ; en concreto nos interesa el directorio siguiente C:\Epi_Info\TransEXE que es donde debemos pegar el SPANISH.EXE que nos hemos descargado y copiado.
- 4- hacer click y ejecutar el programa SPANISH.EXE
- 5- Activar el programa EPI INFO, ir a SETTINGS -> MANAGE TRANSLATIONS: vea la imagen siguiente.



6- Si lo ha hecho bien, aparecerá la opción de activar el SPANISH; hágalo...

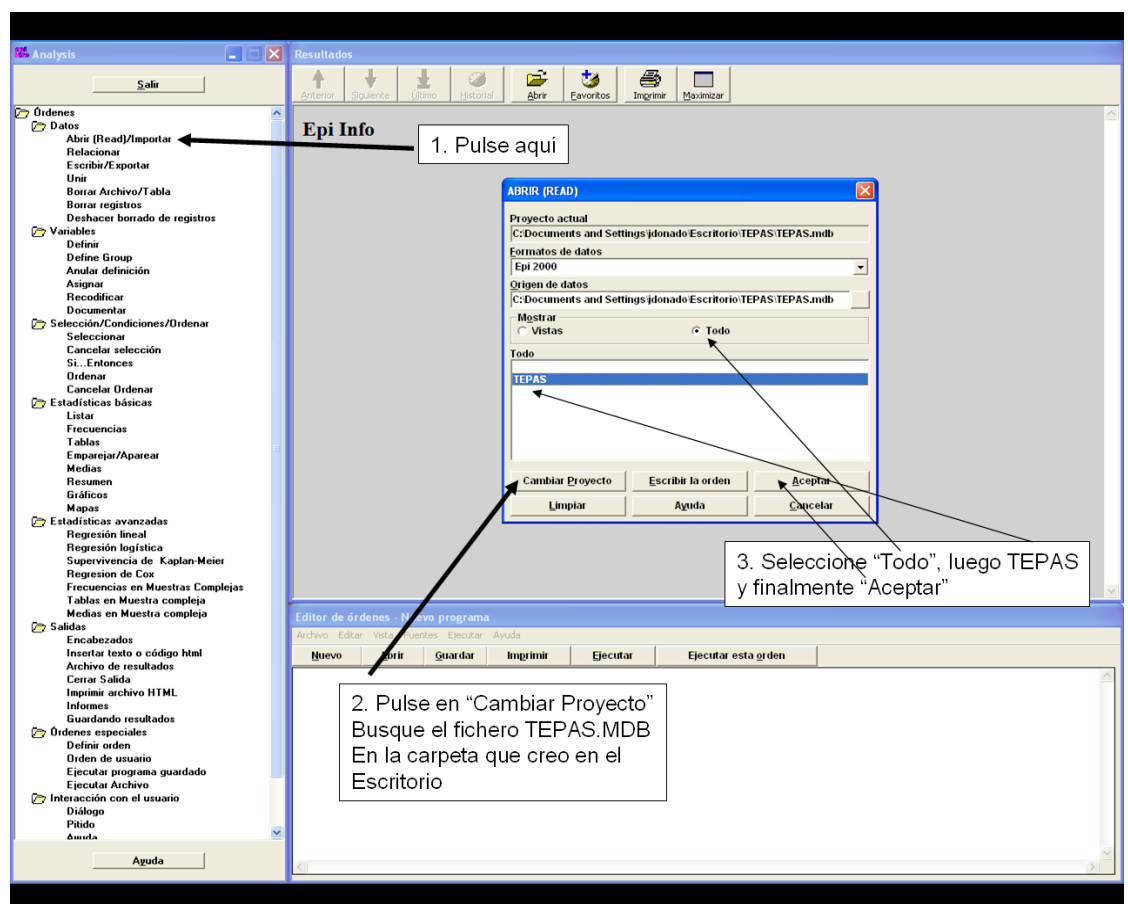
7- Ahora vaya a SETTINGS -> CHOOSE LANGUAGE... y elija el español ... vea esta imagen



CARGANDO EL FICHERO DE TEPAS PARA INICIAR EL TALLER

Active ANALIZAR DATOS; en la pantalla de ANALISIS elija ABRIR/ IMPORTAR -> CAMBIAR PROYECTO -> ESCRITORIO-> TEPAS -> TEPAS_2012 (cargar con doble click)

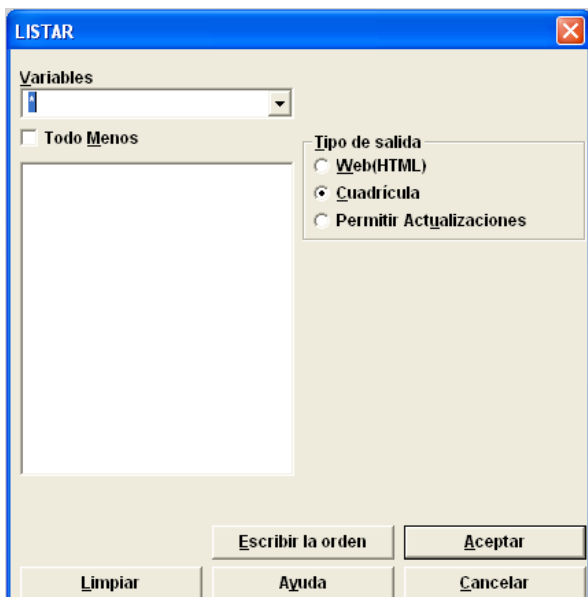
Seleccione en la pantallita, TODO + tepas + ACEPTAR... vea la siguiente imagen:



Y ya podemos empezar el Taller propiamente dicho...

1- Conceptos generales y variables

Empecemos asomándonos a la base de datos; miremos los 500 pacientes y las variables que tenemos. Para hacer un listado busque el comando "Listar" en la columna de órdenes



Resultados							
N_ID	TECNICA	EDAD	COMORBIL	SERVICIO	ESTANCIA	COSTE_€	COMPLIC
1	T1	E1	CO1	S3	8	3634,62	COMP1
2	T2	E3	CO2	S1	8	5019,95	COMP2
3	T2	E1	CO1	S1	3	2208,09	COMP1
4	T1	E2	CO1	S1	7	3072,9	COMP1
5	T1	E2	CO1	S3	10	4777,29	COMP1
6	T1	E1	CO1	S2	7	3366,59	COMP1
7	T1	E1	CO1	S3	8	3670,76	COMP1
8	T1	E2	CO1	S3	10	4698,69	COMP2
9	T1	E2	CO1	S1	6	3060,76	COMP1
10	T1	E3	CO1	S2	10	4253,8	COMP1
11	T1	E1	CO1	S3	8	3430,61	COMP1
12	T1	E1	CO2	S3	9	4762,48	COMP2
13	T2	E3	CO2	S1	8	4584,51	COMP1
14	T2	E3	CO2	S3	12	6366,57	COMP2
15	T1	E2	CO1	S2	9	4063,68	COMP2
16	T1	E2	CO1	S3	10	4107,24	COMP1
17	T1	E2	CO2	S1	8	4285,79	COMP2
...

Se observan los datos organizados en filas (individuos del 1 al 500), y columnas (variables), formando una matriz de datos ordenada de forma aleatoria.

Repasemos las variables que conocemos de cada uno de los 500 enfermos:

- 1- **N_ID**; es un número arbitrario de ordenación de los pacientes (tomarlo como ordenación al azar)
- 2- **TECNICA**; la **T1** se refiere al procedimiento quirúrgico convencional, mientras que la **T2** se refiere al nuevo procedimiento por vía laparoscópica.
- 3- **EDAD**; el valor **E1** se refiere a menores de 60 años; el **E2** a personas entre 60 y 69; el **E3** se refiere a personas de 70 años y más.
- 4- **COMORBILIDAD**; el valor **CO1** significa no existencia de otras patologías asociadas importantes en el episodio de cuidados; el valor **CO2** significa existencia de comorbilidad.
- 5- **SERVICIO**; **S1, S2 y S3** representan los tres servicios de los hospitales donde han sido atendidos los enfermos.
- 6- **ESTANCIAS**; número de días de estancias causados al alta de cada uno de los procesos atendidos.
- 7- **COSTE €**; coste del ingreso calculado en euros por proceso (coste al alta del paciente).
- 8- **COMPLICACIONES**; aparición tras el alta de complicaciones que han requerido atención sanitaria o reingreso para tratamiento de la misma enfermedad en los 3 meses siguientes a la intervención. **COMP1** significa la no existencia de complicaciones y **COMP2** la existencia de complicaciones.

Como se observa, las variables presentan diferencias en su naturaleza; a continuación tiene usted un cuadro con los principales tipos de variables; clasifique las de nuestra base de datos en las categorías que correspondan:

Variables: Aspectos observables objeto de estudio, cuyos valores son susceptibles de variación.

Cualitativas (características susceptibles más de catalogar

y cuantificar a través de un número que de medir)

-Nominales o cardinales: indican pertenencia a una clase o categoría, pueden tener dos valores (dicotómicas) o más. Ej.: Sexo (varón, mujer); Estado civil (soltero, casado, viudo, divorciado).

-Ordinales: cierta ordenación aunque sin proporcionalidad. Ej.: Satisfacción con un servicio (Poco Satisfecho/Regular/ Muy satisfecho).

Cuantitativas (características susceptibles de medir, cuantificar y expresar mediante un número)

-Discretas: hay saltos o interrupciones entre los valores de las variables, pero las distancias entre los intervalos se suponen iguales. Ej.: Número de hijos.

-Continuas: cualquier valor numérico en la medición. Ej.: Peso, Talla.

(Ojo...una variable inicialmente cuantitativa, por comodidad puede convertirse de hecho en cualitativa...ej: glucosa + / ++ / +++; no obstante cada vez que hacemos esto, aunque ganemos en comodidad, perdemos información...)

Técnica	Edad	Comorbilidad.	Servicio	Estancias	Coste	Complicac.

SOLUCIONES ANEXO 1

2- Visualización de la matriz de datos

Los datos no suelen mostrar sus secretos a simple vista; para ello debe usted comenzar un asedio sistemático; caben dos opciones; la primera es comenzar rápidamente una estadística descriptiva, y la segunda consiste en dedicar un rato a pensar sobre el significado de las variables de su base de datos. Recomendamos este segundo curso de acción, porque luego nos facilitará entrar en el análisis (construir modelos explicativos o predictivos para la relación de variables entre sí).

En una matriz de datos, ninguna variable es en principio más importante que otra...hasta que nos ponemos a pensar:

- Por una parte, existirán unas variables que son el objeto principal de nuestro interés, y que suelen asociarse a ciertos resultados de un fenómeno: suelen ser candidatas claras a **variables dependientes** (dependen de los valores de otras).
- Además, otras variables que suelen encontrarse "aguas arriba" de la relación entre variables (aunque no entremos a determinar aquí si dicha presunta causalidad es correcta), son candidatas a ser **variables independientes**, o también llamadas explicativas.

En notación matemática, $Y = f(X_1, X_2, X_3, \dots, X_n)$, la Y sería la variable dependiente de la(s) X.

Las variables independientes, en los diseños de estudio, a veces se refieren a características "**exógenas**" (como edad, sexo, residencia) que no son modificables en la intervención, y configuran el entorno (aunque hay que tenerlas en cuenta pues pueden actuar como modificadores importantes); nuestra mayor atención suele concentrarse en las características "**endógenas**", que representan intervenciones o comportamientos de agentes que trabajan en el sistema; éstas son las que a priori tienen más interés, y al menos presentan una cierta presunción de relación o asociación en los diseños (voluntariamente no queremos hablar de causalidad...).

Sobre la base de las ideas anteriores, clasifique las variables del

taller en los tres campos dibujados:

Variables

Exógenas (independientes)	Endógenas (independientes)
De resultados (dependientes)	

SOLUCIONES ANEXO 1

3- Descripción de las variables cualitativas: porcentajes

Comencemos por una variable cualitativa de la base de datos; **TECNICA**, que tiene sólo dos cualidades o caracteres, **T1** (convencional), y **T2** (laparoscópica). Un primer método sería el recuento de la frecuencia de cada técnica en la base de datos...

ANALISIS -> FRECUENCIAS (en el menú de órdenes) -> FRECUENCIA DE... desplegar y elegir TECNICA -> ACEPTAR... y nos sale la distribución de frecuencias entre T1 y T2

FREQ TECNICA

[Siguiente procedimiento](#)

[Adelante](#)

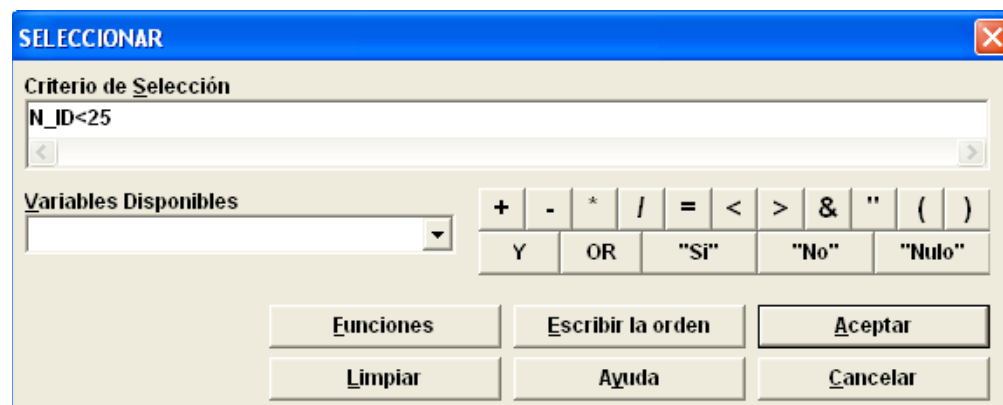
TECNICA	Frecuencia	Porcentaje	Porcentaje acumulado	
T1	300	60,0%	60,0%	<div style="width: 60%; background-color: yellow;"></div>
T2	200	40,0%	100,0%	<div style="width: 40%; background-color: yellow;"></div>
Total	500	100,0%	100,0%	<div style="width: 100%; background-color: orange;"></div>

Int. Conf. 95 %
 T1 55,5% 64,3%
 T2 35,7% 44,5%

Pero como esto es demasiado fácil, y queremos ir introduciendo la complejidad de trabajar con muestras, vamos a suponer que estamos en una **fase previa de nuestra investigación**.

SUPONGAMOS...que un colega nos dice: *"la técnica laparoscópica se usa muy poco...como mucho un 15% de todas las intervenciones"*; como **ya hemos revisado 24 enfermos** de nuestra base de datos, la idea es ver si dichos datos confirman o desmienten la anterior afirmación (supuesta la representatividad de nuestra muestra).

Vamos a seleccionar los 24 primeros pacientes (recordamos que el número N-ID es aleatorio, y por lo tanto lo podemos usar como muestra insesgada de la población de 500 pacientes); para dicha selección, se precisan las siguientes instrucciones: **Busque el comando "Seleccionar" en la lista de órdenes**



ACEPTAMOS... y ya quedan seleccionados sólo los 24 primeros pacientes...

Pedimos a continuación que nos calcule la Frecuencia de las Técnicas empleadas con estos 24 enfermos (como hicimos al principio de este apartado):

Aparecerán los siguientes resultados

[Librería de resultados](#)
 Vista actual: d:\users\jrepullo\Desktop\Tepas_2012.mdb:tepas
 Selección: N_ID < 25
 Número de registros: 24 Fecha: 23/08/2012 12:37:11

FREQ TECNICA

[Siguiete procedimiento](#)

[Adelante](#)

TECNICA	Frecuencia	Porcentaje	Porcentaje acumulado	
T1	17	70,8%	70,8%	
T2	7	29,2%	100,0%	
Total	24	100,0%	100,0%	

Int. Conf. 95 %
 T1 48,9% 87,4%
 T2 12,6% 51,1%

La cirugía laparoscópica (T2) representa aproximadamente un 30% en la muestra; para saber si podemos desmentir la afirmación de nuestro colega (refutar el 15% que él estimaba como máximo), debemos buscar el intervalo de confianza en torno al cual puede ser cierto ese 30% de utilización de la laparoscopia (con un error de equivocarnos en dicho intervalo de tan sólo el 5%...la conocida $p < 0,05$).

El problema es que 24 enfermos son muy pocos, y el intervalo de confianza saldrá muy ancho; nos obstante lo intentamos... recordamos la fórmula del **"Error estándar del porcentaje"**

$$Sp = \sqrt{\frac{p \cdot q}{n}}$$

Siendo "p" la probabilidad del evento que nos interesa (técnica laparoscópica) "q" la probabilidad del otro evento (1-p), y "n" el número total de sujetos de la muestra.

Tras la operación, resulta que $Sp = 9,3$

Para hallar el intervalo de confianza respecto al valor de p que nos salió en la muestra (29,2%), deberemos sumar y restar el Sp (error estándar del porcentaje) multiplicado por 1,96 (para conseguir un 95% de seguridad):

$$IC = p \pm z_{\alpha} * Sp$$

$$IC = 29,2 \pm 1,96 * 9,3$$

$$\qquad \qquad \qquad \downarrow$$

$$\qquad \qquad \qquad 18,2$$

El resultado es 29,2 +/- 18,2

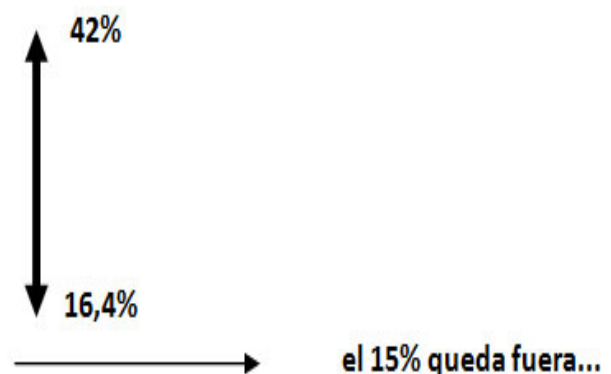


Luego, como podemos ver, la opinión de nuestro colega no puede darse como falsa...(aunque tiene mucha pinta de serlo).

Repitamos ahora el experimento, pero esta vez doblaremos el número de pacientes, hasta 48...

Primero debemos anular la selección anterior. Busque el comando "Cancelar selección" y pulse la opción "Aceptar".

El resultado debería ser







A continuación vamos a seleccionar los 48 primeros registros (N_ID < 49). Siga los pasos que vimos anteriormente.

y por lo tanto podemos darnos la (siempre sana) satisfacción de decirle a un colega que se ha equivocado.

Ahora, **utilizando todos los registros**, para ello active el comando "Cancelar selección", tocaría hacer una estadística descriptiva de dos variables cualitativas: Técnica (T1 y T2) y Servicios (S1, S2 y S3) para ver la proporción de las mismas en la población de 500 pacientes atendidos.

Ya hicimos la FRECUENCIA de TECNICA al principio de este apartado; ahora haremos igualmente la FRECUENCIA de SERVICIO

[Adelante](#)

SERVICIO	Frecuencia	Porcentaje	Porcentaje acumulado	
S1	125	25,0%	25,0%	
S2	150	30,0%	55,0%	
S3	225	45,0%	100,0%	
Total	500	100,0%	100,0%	

Veámos al principio, que en los datos globales la técnica convencional **T1** era más frecuente (300 pacientes y un 60%) que la laparoscópica **T2** (200 pacientes y 40%); también vemos que el Servicio 1 atiende 125 pacientes (25%), que el Servicio 2 atiende algo más (150 – 30%), y que el Servicio 3 asume la mayor carga (225 – 45%)...pero... ¿hasta qué punto cada uno de dichos 3 servicios desarrolla una u otra técnica?...

Pero, para dar más emoción... hagamos primero una reflexión metodológica ...

Si los tres servicios hicieran ambas técnicas proporcionalmente, ¿qué valores tendría que darnos en la tabla (casillas centrales vacías)?

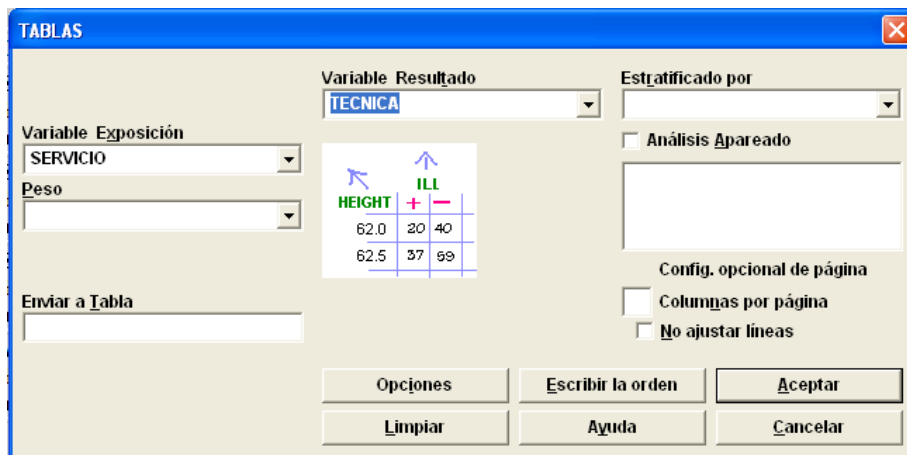
	T1	T2	Total
S1			125
S2			150
S3			225
Total	300	200	500

(para calcular la celdilla S1-T1 multiplicar 300 por 125 y dividirlo por 500, y así sucesivamente)

El resultado debería ser...

	T1	T2	Total
S1	75	50	125
S2	90	60	150
S3	135	90	225
Total	300	200	500

Buscamos ahora los datos las frecuencias reales encontradas... Localice el comando TABLAS en la lista de órdenes



[Adelante](#)

TECNICA

SERVICIO	T1	T2	TOTAL
S1	50	75	125
% Fila	40,0	60,0	100,0
% Columna	16,7	37,5	25,0
S2	75	75	150
% Fila	50,0	50,0	100,0
% Columna	25,0	37,5	30,0
S3	175	50	225
% Fila	77,8	22,2	100,0
% Columna	58,3	25,0	45,0
TOTAL	300	200	500
% Fila	60,0	40,0	100,0
% Columna	100,0	100,0	100,0

Análisis de tabla simple

Chi cuadrado gl Probabilidad
56,7130 2 0,0000

Llevados a la presentación de tabla anterior, los resultados serían

	T1	T2	Total
S1	50	75	125
S2	75	75	150
S3	175	50	225
Total	300	200	500

Por lo tanto vemos que hay diferencias...que el **Servicio 1** es muy activo con la nueva técnica (superando a la convencional), que el **Servicio 2** hace igual de ambas, y que el **Servicio 3** confía mucho más en la convencional que en la laparoscópica...pero **¿existe asociación significativa entre servicio y técnica o podrían las diferencias ser explicadas por el azar?**

Para saberlo, precisamos usar el estadístico que compara proporciones, y que es el famosísimo **Chí cuadrado** - χ^2 - (que es precisamente lo que nos ha salido al pie de la tabla en Epi6). Recordamos que el χ^2 de Pearson, se calcula haciendo un sumatorio de todas las diferencias entre valores encontrados (f_e) y teóricos (f_t) (elevada dicha diferencia al cuadrado y dividida por el valor teórico); el número que nos sale (mayor cuanto mayor es la diferencia entre valores encontrados y teóricos), se lleva a una tabla (ver ANEXO 2) donde vemos si puede ser explicada dicha diferencia por el azar (para una $p > 0,05$ o para una $p > 0,01$); igualmente recordamos que para usar esa tabla, hay que señalar los "grados de libertad" con los que entramos en la tabla (que sería el resultado de multiplicar el número de filas de datos menos 1, por el número de columnas menos 1...en este caso $(3-1) \cdot (2-1) = 2$ grados de libertad).

La fórmula es la siguiente:

$$\chi^2 = \sum \frac{(f_e - f_t)^2}{f_t}$$

Afortunadamente, EpiInfo tiene la amabilidad de darnos el

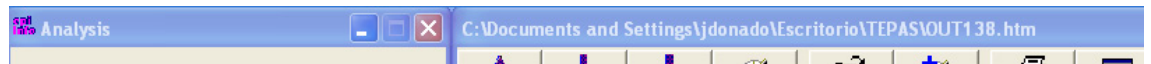
resultado (y los grados de libertad)...comprobamos que los servicios son MUY DIFERENTES (existe asociación significativa entre servicio y técnica utilizada) en su uso de las técnicas convencionales y laparoscópicas, y que al decir esto, nuestra posibilidad de equivocarnos es menos que $p= 0,000000001$).

Para completar esta fase descriptiva, **haga las frecuencias de todas las variables cualitativas**, incluyendo la EDAD (que hemos convertido en grupos para explotación cualitativa):

TECNICA SERVICIO EDAD COMORBIL COMPLIC

NOTA SOBRE LOS RESULTADOS

- Todos los resultados que se obtienen en EPI-INFO se guardan en un documento con formato HTML (el formato de las páginas Web)
- El nombre del documento donde se guardan aparece en la zona superior de la ventana de resultados. El nombre se compone de la palabra OUT y un número y están guardados en la carpeta donde está la base de datos TEPAS.MDB



- También puede ver un listado de todos los resultados que ha hecho hasta ahora pulsando en el enlace de HISTORIAL que aparece en la ventana de resultados, o moviéndose con las flechas hacia operaciones anteriores o siguientes.

Historial de Sesión		
jueves, 23 de agosto de 2012		
Hora	Orden	Archivo
12:28	FREQ TECNICA	OUT1.htm
12:31	FREQ TECNICA	OUT1.htm
12:37	FREQ TECNICA	OUT2.htm
12:45	FREQ TECNICA STRATAVAR = SERVICIO	OUT3.htm
12:49	TABLES SERVICIO TECNICA	OUT3.htm
12:56	FREQ SERVICIO	OUT4.htm
01:02	TABLES SERVICIO TECNICA	OUT4.htm

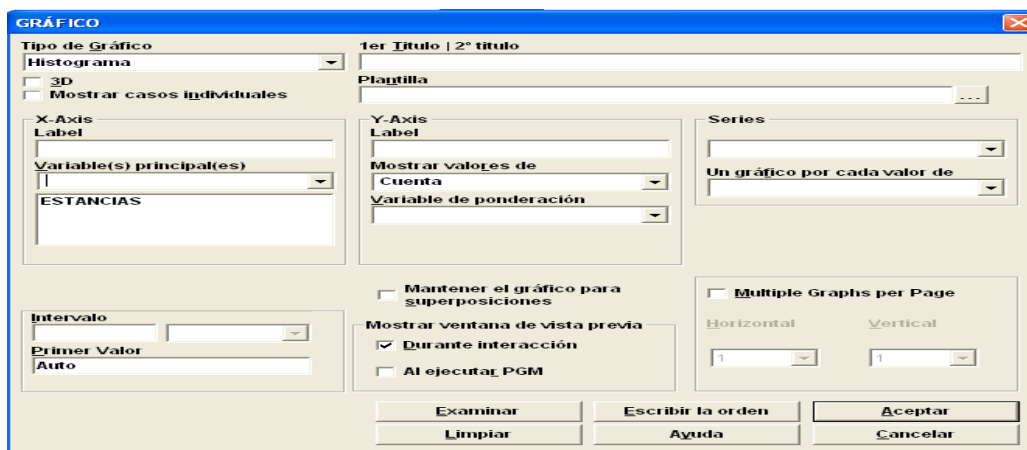
- Para abrir un fichero OUT desde Word primero salga de Epi-Info. Luego abra Word y seleccione Archivo à Abrir... o bien, al seleccionar el fichero con el botón derecho del ratón elegir ABRIR CON ... WORD
- Localice la carpeta donde tenga guardados los ficheros OUT y en ventana de "Tipos de Archivo" selecciones "Todas las páginas Web"
- Seleccione el archivo que le interese y edítelo como quiera. Lugo lo puede guardar como un documento Word.

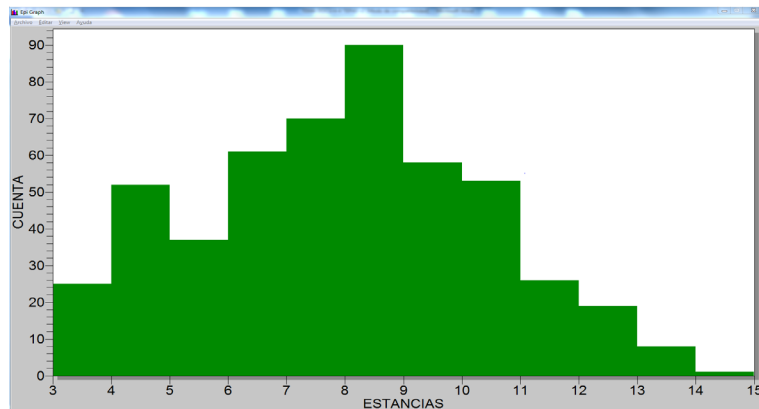
4- Descripción de las variables cuantitativas

Las variables cuantitativas (sean discretas como ESTANCIAS, o continuas como COSTES) requieren otros estadísticos bien conocidos como la **MEDIA, MEDIANA, MODA, DESVIACIÓN TÍPICA y VARIANZA.**

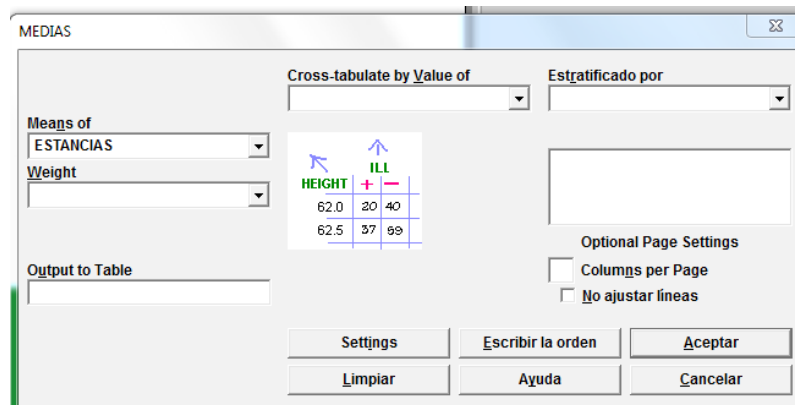
Aquí nos basaremos tan sólo en la **MEDIA** como valor central de la distribución (no obstante deberían revisarse los conceptos de MEDIANA –valor que divide a la distribución en dos partes de igual número de individuos- y MODA –valor más frecuente en la distribución)

Tomemos por ejemplo la variable **ESTANCIA** y pidamos para empezar una distribución de frecuencias y su dibujo en forma de HISTOGRAMA ...Para hacer un HISTOGRAMA seleccione en la lista de órdenes la opción GRÁFICOS





La media de las estancias se calcula fácilmente con las órdenes MEDIAS -> ver imagen



ESTANCIAS	Frecuencia	Porcentaje	Porcentaje acumulado
3	25	5,0%	5,0%
4	52	10,4%	15,4%
5	37	7,4%	22,8%
6	61	12,2%	35,0%
7	70	14,0%	49,0%
8	90	18,0%	67,0%
9	58	11,6%	78,6%
10	53	10,6%	89,2%
11	26	5,2%	94,4%
12	19	3,8%	98,2%
13	8	1,6%	99,8%
14	1	0,2%	100,0%
Total	500	100,0%	100,0%

Observaciones	Total	Media	Varianza	Desviación típica
500	3728,0000	7,4560	6,0121	2,4520
Mínimo	25%	Mediana	75%	Máximo
3,0000	6,0000	8,0000	9,0000	14,0000
				8,0000

Vemos que los 500 valores quedan representados por una **Media de 7,45 días**, y que la dispersión de los datos alrededor de esta media se representa por la **Varianza (6,012)** y por la **Desviación Estándar o Típica (2,452)**. También nos calcula la mediana y la moda (8 días ambas).

Si queremos salir de nuestra muestra y plantearnos la validez de esta media que hemos encontrado, la pregunta correcta sería... ¿Cuál sería el **Sm (Error Estándar)** de la media de 7,45 días de estancia de nuestro experimento?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

O mejor, considerando los grados de libertad ... poniendo n-1 en el denominador

Si el alumno es laborioso, ya estará dándole al Excel o a la calculadora...

a	varianza	6,01	
b	desv típica	2,45	raiz de a
c	n-1	459,00	
d		21,42	raiz de c
e	Sm	0,11	c/d

¿Cómo lo podemos interpretar?; Es sencillo: si al valor de la media le añadimos y le restamos 1,96 veces dicho valor, encontraremos el famosísimo intervalo de confianza:

$$IC = \mu \pm z_{\alpha} * \sigma / \sqrt{(n-1)}$$

\downarrow \downarrow \downarrow
 IC= 7,45 ± 1,96 * 0,11

$$7,45 \left\{ \begin{array}{l} + (1,96 * 0,11) = 7,66 \text{ días} \\ - (1,96 * 0,11) = 7,23 \text{ días} \end{array} \right.$$

la notación sería 7,45 (+/- 0,2156) p=0,05

o bien señalando el intervalo de confianza 7,45 (7,23-7,66) p=0,05

La interpretación sería: si repitiéramos muchísimas veces una medida similar de 500 pacientes de las mismas características, elegidos aleatoriamente, el valor de la media de estancias causadas estaría entre 7,23 días y 7,66 días, en 95 de cada 100 veces que hiciéramos este experimento o medición (este es el significado de la probabilidad de error $p=0,05$).

Cuando las muestras son pequeñas, el multiplicador 1,96 veces que corresponde a la $p=0,05$ debería aumentarse; de ahí el famoso estadístico de la t de student, que animamos a repasar por su cuenta a los alumnos (se usa una tabla que nos da el valor a aplicar según el número de individuos de la muestra (menos 1 para convertirlos en "grados de libertad"); luego se haya el intervalo de confianza... media $\pm (t * S_m)$. La tabla y su explicación puede encontrarse en muchos lugares; por ejemplo en este enlace: http://es.wikibooks.org/wiki/Tablas_estad%C3%ADsticas/Distribuci%C3%B3n_t_de_Student

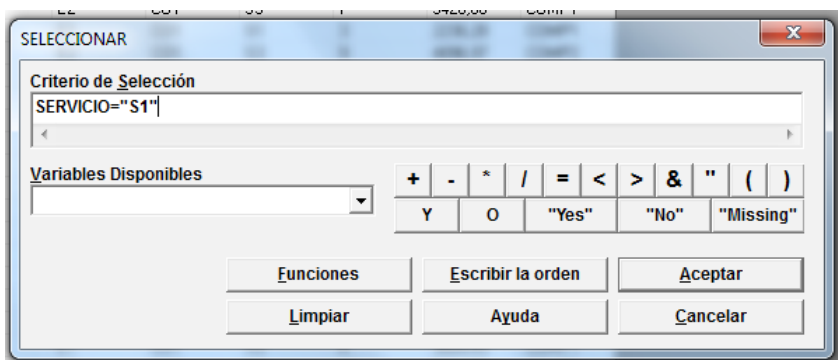
¿Para qué puede sernos útil todas estas molestias de calcular los intervalos de confianza?

- **En primer lugar**, cuando reflejemos un valor muestral en un trabajo científico, es casi obligatorio añadir el intervalo de confianza o el error estándar con su p . Es lógico si pensamos que a nadie le interesa nuestra muestra, sino las inferencias poblacionales de los datos (que afectan a sus posibles casos o problemas). Además así quedamos bien, y no sufrimos el riesgo de que un revisor neurótico nos rechace el artículo con insultos de incultura estadística (habitualmente merecidos).
- **En segundo lugar**, nos sirve para comparar nuestro resultado con un valor teórico: En la variable que nos ocupa... si a nivel nacional nos encontramos que la Estancia media para este tipo de enfermos es de 7,8 días...¿podría nuestro excelente resultado de 7,45 días ser atribuido a nuestros méritos, o ser explicado por el azar?
- **En tercer lugar**, puede servirnos para comparar nuestro experimento con otro experimento y saber si la diferencia entre los resultados puede ser explicada por el azar.

La pregunta ahora sería la siguiente: si comparamos los resultados del Servicio 1 y el Servicio 3, ¿pueden ser la diferencia en la media de estancias debida al azar?

Para ello, en la base de datos seleccionamos primero los datos del Servicio 1

- Recuerde que debe utilizar la orden SELECCIONAR, pero en este caso ponga como criterio = "S1" (entre comillas, ya que no se trata de un número sino de un texto)



- extraemos el valor de la media y el error estándar (**orden MEDIAS -> ESTANCIAS**)

MEANS ESTANCIAS

[Siguiente procedimiento](#)

[Adelante](#)

ESTANCIAS	Frecuencia	Porcentaje	Porcentaje acumulado	
3	25	20,0%	20,0%	■
4	22	17,6%	37,6%	■
5	15	12,0%	49,6%	■
6	23	18,4%	68,0%	■
7	16	12,8%	80,8%	■
8	15	12,0%	92,8%	■
9	6	4,8%	97,6%	■
10	3	2,4%	100,0%	■
Total	125	100,0%	100,0%	

Observaciones Total Media Varianza Desviación típica
 125 692,0000 5,5360 3,8636 1,9656

Mínimo 25% Mediana 75% Máximo Moda
 3,0000 4,0000 6,0000 7,0000 10,0000 3,0000

- quitamos la selección previa con el comando **Cancelar Selección**
- y repetimos lo anterior para el Servicio S2 y para el Servicio S3...

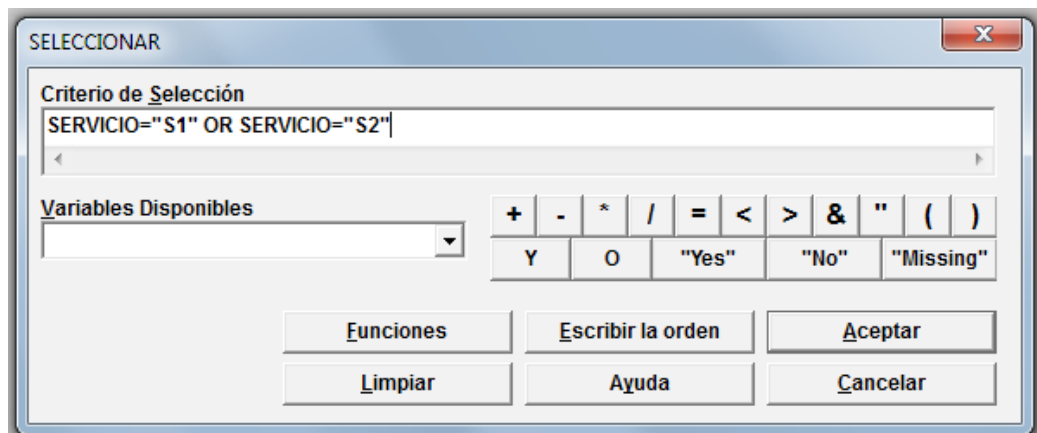
Estancia media por servicio	S1	S2	S3
Media	5,54	6,84	8,93
Desviación Típica	1,97	2,13	1,93
Número de pacientes	125	150	225
Error estandar de la media	0,02	0,01	0,01
Límite superior intervalo confianza	5,57	6,87	8,95
Límite inferior intervalo confianza	5,50	6,81	8,92

Vemos que las estancias medias son bastantes diferentes, y que sus intervalos de confianza no llegan a tocarse...

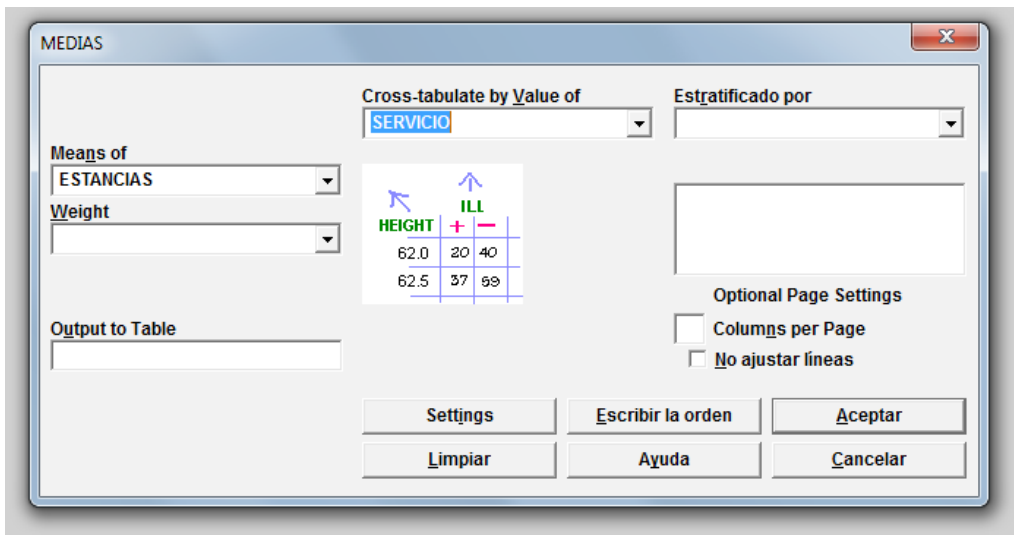
Otro método más fino sería la comparación de medias, a través del **error estándar de la diferencia de medias**, (que no vamos a utilizar aquí), o bien a través del análisis de la comparación de las dispersiones de las diferentes distribuciones, que lleva el nombre de **ANÁLISIS DE LA VARIANZA**, y que es el utilizado por EPI INFO.

Comencemos comparando el Servicio S1 y el Servicio S2 ...

SELECCIONAR -> SERVICIO="S1" OR SERVICIO="S2"



Hacemos la comparación de medias. **Para ello también utilizamos la orden MEDIAS**



Los resultados que nos da EPI INFO aportan dos sistemas de comparación; el ANOVA (análisis de la varianza) que como test paramétrico podemos utilizar en el caso de que la distribución de los datos sea normal (en forma de campana de Gauss); y el KRUSKAL WALLIS, que es una alternativa a utilizar si no cabe usar el ANOVA; para decidir el uso se recurre al test de Bartlett (Un valor de p menor de 0,05 sugiere que las varianzas no son homogéneas y que el test ANOVA no es apropiado).

Estadísticas descriptivas de la variable de cruce

	Obs	Total	Media	Varianza	Desviación típica
S1	125	692,0000	5,5360	3,8636	1,9656
S2	150	1026,0000	6,8400	4,5380	2,1303

	Mínimo	25%	Median	75%	Maximum	Mode
S1	3,0000	4,0000	6,0000	7,0000	10,0000	3,0000
S2	4,0000	5,0000	7,0000	8,0000	12,0000	7,0000

ANOVA, test paramétrico para comparación de medias

(Para datos normalmente distribuidos)

Variación	SC	gl	MS	Estadístico F
Entre grupos	115,9375	1	115,9375	27,3975
Intra grupo	1155,2480	273	4,2317	
Total	1271,1855	274		

Estadístico T = 5,2343

Valor p= 0,0000

Bartlett's Test for Inequality of Population Variances

Bartlett's chi square= 0,8674 df=1 valor p=0,3517

Un valor de p pequeños (menor de 0,05) sugiere que las varianzas no son homogéneas y que el test ANOVA no es apropiado.

Test de dos muestras de Mann-Whitney/Wilconxon (Test de Kruskal-Wallis para dos grupos)

H de Kruskal-Wallis (equivalente a Chi cuadrado)= 23,4813

Grados de libertad= 1

Valor p= 0,0000

Como vemos, podemos usar ANOVA, y cabe afirmar que S1 Y S2 son servicios con estancias medias diferentes con altísima significación estadística ($p > 0,0000...$)

5- Comenzamos el análisis... reflexionando

En realidad el análisis implica fundamentalmente comparación entre variables; algo de comparación ya hemos hecho en las fases anteriores (pero era sólo a efectos de adquirir habilidades y fijar conceptos); obviamente todas las variables pueden compararse con todas, pero el trabajo es complejo y tedioso, y a veces se pierde el sentido común y se nos escapa el significado de las relaciones que van apareciendo.

Por ello es importante establecer una guía de explotación con la lógica (a-priori) de relación entre variables.

Como habíamos comentado en la Fase 1, existen tres variables que son claras candidatas a variables dependientes (a explicar): Estancia, Costes y Complicaciones.

En esta fase comenzaremos por las **Estancias**, que será la variable principal en el taller (las otras dos serán útiles posteriormente para complementar algunos conceptos y aplicaciones).

¿qué puede repercutir y cómo en la estancia?

Le invitamos a que rellene este cuadro de relaciones...

VARIABLE EXPLICATIVA: ESTANCIAS

Si claramente /dudoso/No parece

TECNICA	
EDAD	
COMORBILIDAD	
SERVICIO	

Acometeremos pues un análisis entre dos variables cada vez (bivariante).

Puede ser útil anticipar en forma de cuadro, los métodos estadísticos y las órdenes de EPI INFO que usaremos en la comparación de variables según sean la naturaleza de ambas (reseñamos sólo las paramétricas, que se utilizan cuando la distribución es normal).

Variable 1	Variable 2	Método estadístico	Orden EPI INFO
Cualitativa	Cualitativa	χ^2 (Chí o Jí cuadrado)	TABLAS
Cualitativa	Cuantitativa	Comparación de medias	MEDIAS
Cualitativa (más de dos categorías)	Cuantitativas	ANOVA (análisis de la varianza)	MEDIAS
Cuantitativa	Cuantitativa	Correlación / Regresión	DISPERSION XY

TECNICA Y ESTANCIAS

Supongamos que tenemos dos poblaciones; la de los pacientes operados con la técnica **convencional**, y los operados vía **laparoscópica**... ¿hay diferencias significativas entre las estancias causadas en cada caso?

Pedimos a Epi-Info que nos compare las medias (**orden MEDIAS**) de la ESTANCIA dependiendo de la TÉCNICA empleada

Vemos que ...

Estadísticas descriptivas de la variable de cruce					
	Obs	Total	Media	Varianza	Desviación típica
T1	300	2616,0000	8,7200	3,1120	1,7641
T2	200	1112,0000	5,5600	4,3783	2,0924

Test de dos muestras de Mann-Whitney/Wilcoxon (Test de Kruskal-Wallis para dos grupos)

H de Kruskal-Wallis (equivalente a Chi cuadrado)= 206,9818
 Grados de libertad= 1
 Valor p= 0,0000

La **diferencia de las medias** (8,72 y 5,56) es de 3,16 días... Y en este caso, el test no paramétrico recomendado por EPI INFO nos dice que es altísimamente significativo...

Seguimos con las otras relaciones bivariantes...

COMORBILIDAD Y ESTANCIAS

Comparación de medias de ESTANCIAS según COMORBIL

	Obs	Total	Media	Varianza	Desviación típica
CO1	350	2349,0000	6,7114	4,5784	2,1397
CO2	150	1379,0000	9,1933	5,0698	2,2516

Se nota que la comorbilidad añade dos días y medio de estancia, y que esta diferencia resulta significativa ...

SERVICIOS Y ESTANCIAS

En este caso, nos encontramos que la variable cualitativa SERVICIOS tiene **tres categorías** posibles (1, 2 y 3). Para comparar globalmente las dos variables, se usa también el **Análisis de la Varianza**.

Como ya comentamos, en el ANALISIS DE LA VARIANZA (ANOVA), se exige que las distribuciones sean normales y homogéneas; profundizando un poco más en el método, se compara la varianza imputable al factor que estudiamos (SERVICIO), con la varianza que no puede atribuirse a dicho factor (la interna que se produce dentro de los pacientes del 1 del 2 y del 3); el resultado de dicha división se lleva a la tabla de la F de Snédecor (por ejemplo <http://www.itl.nist.gov/div898/handbook/eda/section3/eda3673.htm>), y nos da los valores para las probabilidades de error requeridas.

Estadísticas descriptivas de la variable de cruce

	Obs	Total	Media	Varianza	Desviación típica
S1	125	692,0000	5,5360	3,8636	1,9656
S2	150	1026,0000	6,8400	4,5380	2,1303
S3	225	2010,0000	8,9333	3,7321	1,9319
	Mínimo	25%	Median	75%	Maximum
S1	3,0000	4,0000	6,0000	7,0000	10,0000
S2	4,0000	5,0000	7,0000	8,0000	12,0000
S3	5,0000	8,0000	9,0000	10,0000	14,0000

ANOVA, test paramétrico para comparación de medias

(Para datos normalmente distribuidos)

Variación	SC	gl	MS	Estadístico F
Entre grupos	1008,7840	2	504,3920	125,8923
Intra grupo	1991,2480	497	4,0065	
Total	3000,0320	499		

Valor p= 0,0000

Vemos que S1 tiene 5,5 días de estancia, menos que S2 (6,8) y que S3 (8,9); la varianza total se descompone con la interna de cada servicio (intra-grupo – de todos los pacientes de cada Servicio con respecto a su propia media) y la factorial –del “factor explicativo servicios”- (inter-grupos - de la media de los grupos comparados con la media general); de esta comparación surge la estimación de significación estadística; en este caso el Estadístico F muestra significación.

EDAD Y ESTANCIAS

Realizamos el análisis de ambas variables como en el caso anterior: **Comparación de medias de ESTANCIAS según EDAD.**

Estadísticas descriptivas de la variable de cruce					
	Obs	Total	Media	Varianza	Desviación típica
E1	250	1590,0000	6,3600	3,7173	1,9280
E2	150	1169,0000	7,7933	5,1852	2,2771
E3	100	969,0000	9,6900	4,9029	2,2143

Comprobamos que la edad más avanzada añade más de tres días de estancia respecto de la más baja; En este caso las varianzas no son homogéneas, pero el test no paramétrico nos indica que las diferencias son también estadísticamente significativas.

En resumen...

Vemos que en todos los casos se demuestran diferencias significativas.

Por lo tanto tenemos una buena pista a seguir...parecen existir diferencias, lo que significa que las variables están enlazadas de alguna forma (busquemos asociación, relación, y seguimos sin querer hablar de causalidad, aunque nos gustaría...).

Pero ... sería importante llegar a identificar no sólo si son significativas, sino **EN QUE MEDIDA ESTAN ASOCIADAS LAS VARIABLES, O CON QUÉ FUERZA SE RELACIONAN.** Pasemos a la siguiente fase.

6- Estudiamos la relación de algunas variables

Hemos visto que las variables TECNICA, EDAD, COMORBILIDAD Y SERVICIO, influían en las ESTANCIAS de forma estadísticamente significativa.

Para intentar cuantificar la intensidad de dicha relación, por ejemplo en el caso de TECNICAS, podríamos ver hasta qué punto la técnica convencional provoca más estancias que la técnica laparoscópica. Una forma es hacer un cociente (o razón) con la primera en el numerador y la segunda en el denominador (**PROPORCIÓN RELATIVA** que guarda semejanza con el Riesgo Relativo que se usa en estudios epidemiológicos).

Tomemos los datos de ESTANCIAS y TÉCNICA, a través de la ya conocida orden de MEDIAS

	Obs	Total	Media	Varianza	Desviación típica
T1	300	2616,0000	8,7200	3,1120	1,7641
T2	200	1112,0000	5,5600	4,3783	2,0924

El cociente o proporción de la Estancia Media de T1 sobre al de T2 es $8,72 / 5,56 = 1,56$

El cociente o razón 1,56 significa que la técnica 1 gasta el 56% más de días de estancia que la técnica 2, y mide la intensidad de influencia de dicha variable en la estancia.

El **riesgo relativo** es una medida de "efecto"; también podemos representar el mismo fenómeno con la **proporción atribuible**, que es una medida de "impacto" cuya interpretación puede ser incluso más directa y clara (igualmente dicho concepto guarda relación con el riesgo atribuible de estudios epidemiológicos):

El cálculo es sencillo:

$$\begin{aligned} & \text{Riesgo relativo - 1} \\ \text{Proporción atribuible} &= \frac{\text{Riesgo relativo - 1}}{\text{Riesgo Relativo}} \\ \text{En el caso anterior...} & \\ & \frac{(1,56 - 1)}{1,56} = 0,36 = 36 \% \end{aligned}$$

La interpretación sería la siguiente; en la variable dependiente (las estancias), un 36% de la misma es atribuible a la Técnica 1; es decir, si pasáramos a realizar solamente la técnica 2, nos ahorraríamos un 36% del total de estancias; por ello se dice que es una medida de impacto (del problema o de la solución), y también nos permite entender la importancia que tiene para la salud pública.

A continuación hacemos lo mismo con las siguientes variables

EDAD (cada uno de los tres tramos de edad) Y ESTANCIAS

Los resultados deberían ser los siguientes:

Estancia Media para pacientes de edad E1= 6,36 días

Estancia Media para pacientes de edad E2= 7,79 días

Estancia Media para pacientes de edad E3= 9,69 días

Asumiendo la edad más anciana E3 como numerador, hallar el cociente o proporción de E3 sobre E1, y realizar la misma operación con E2 sobre E1

Los resultados deberían ser los siguientes:

Cociente E3/E1= 1,52 (52% más de estancias)

Cociente E2/E1= 1,22 (22% más de estancias)

Calcule ahora la proporción atribuible en ambos casos

.....

El siguiente paso es hacer lo mismo con la COMORBILIDAD... (renunciamos a ayudar a los alumnos que ya habrán adquirido notables habilidades prácticas tras los cálculos anteriores).

Los resultados deberían ser

Estancia Media para pacientes sin comorbilidad (CO1) = 6,71 d.

Estancia Media para pacientes con comorbil. (CO2) = 9,19 d.

Cociente CO2/CO1 = 1,37 (37% más de estancias)

Finalmente hacemos otro tanto con los tres servicios...

Estancia Media en S1= 5,54 d.

Estancia Media en S2= 6,84 d.

Estancia Media en S3= 8,93 d.

Cociente S3/S2 = 1,31 (31% más de días)

Cociente S3/S1 = 1,61 (61% más de días)

(tomando como referencia el S3)

Algún gestor incauto, podría llamar a capítulo y regañar al Servicio 3...obviamente éste se defendería diciendo que él usa la técnica convencional porque es la que mejor sabe hacer, y que sus pacientes son ancianos y con comorbilidad...en todo caso, y dado que dichas variables están relacionadas, necesitaríamos buscar una forma de comparación entre servicios que dejara a un lado (estandarizara) la influencia de otras variables exógenas. A ello dedicaremos la próxima fase del estudio.

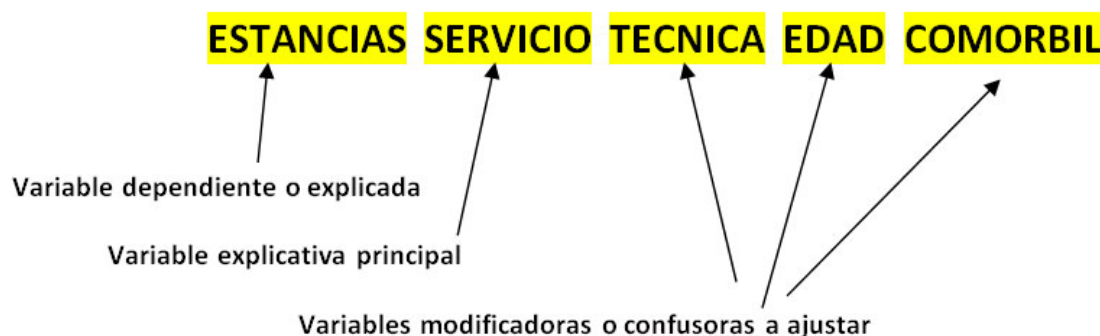
7- El análisis de estancias por paciente de los tres servicios

El tema es un poquito laborioso, pero en su desarrollo adquiriremos nuevas habilidades. Se trata de **estratificar las variables** para comparar los servicios quitando el **efecto confusor** de otras variables...a veces el número de pacientes no es suficiente para ello, pues cuando vamos estratificando posibilidades, se nos puede quedar un número excesivamente reducido...no obstante lo intentaremos.

La función que emplearemos, será **MEDIA (MEANS)**, que permite un estudio estratificado de variables; en esencia, cuando damos la orden **MEANS ESTANCIAS V1 V2 V3...Vn**, el estudio va ordenando la comparación de medias (y el análisis de varianza) entre todos los grupos posibles de las variables cualitativas anteriores; de dichas comparaciones extraemos el dato de la Estancia Media para cada combinación de variables, y si existe significación estadística a través del **Análisis de Varianza** (o en su defecto del test no paramétrico de **Kruskal Wallis** que tiene similar sentido, y que se aconseja cuando el ordenador detecta que hay problemas, y no se dan las condiciones para usar el estadístico anterior¹).

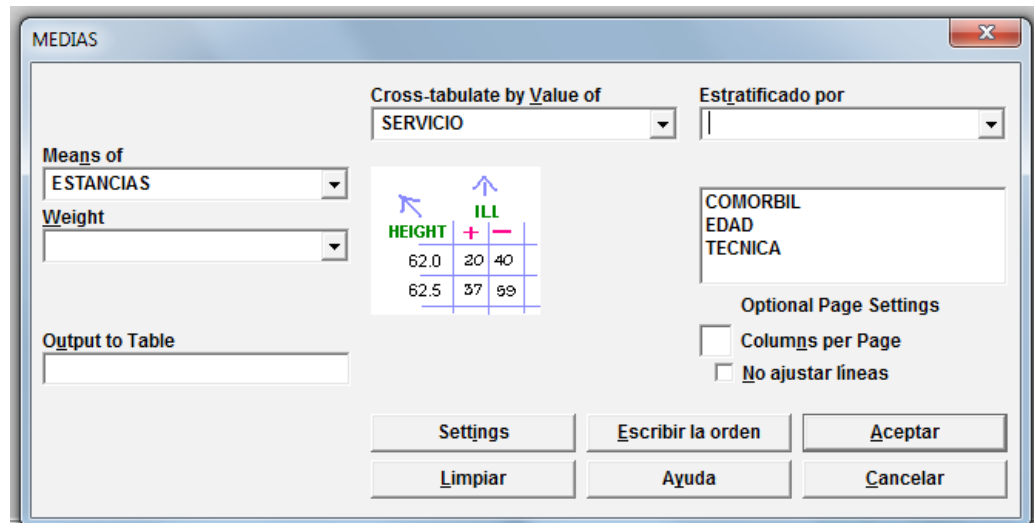
Dado que el interés básico al estratificar consiste en comparar los servicios S1, S2 y S3, la fórmula debería ser:

Seleccionamos todos los registros y utilizamos la orden MEDIAS teniendo en cuenta que



¹ Recordamos que l

La aplicación del análisis de la varianza (ANOVA) exige el cumplimiento de una serie de condiciones, entre las que se encuentra la existencia de al menos una variable cuantitativa, que la variable cuantitativa tenga una distribución normal (gaussiana), que la muestra no sea demasiado pequeña, y que las varianzas sean homogéneas; EPI-INFO nos ayuda con el tema de la homogeneidad de las varianzas (test de Barlet), y nos aconseja usar el ANOVA si son homogéneas. Si no lo son, debemos usar el test no paramétrico de Kruskal Wallis.



E I

resultado se puede llevar al siguiente cuadro para hacer comparaciones estratificadas:

S1	T1	E1	CO1	5,94	S1	T2	E1	CO1	3,14
			CO2	7,4				CO2	5
		E2	CO1	6,85			E2	CO1	4
		CO2	8,38			CO2	5,57		
	E3	CO1	7,5	E3		CO1	5,29		
		CO2	9,29			CO2	7,5		
S2	T1	E1	CO1	6,73	S2	T2	E1	CO1	4,03
			CO2	8				CO2	6,09
		E2	CO1	8,18			E2	CO1	5,38
		CO2	9,83			CO2	7,80		
	E3	CO1	9,37	E3		CO1	7,13		
		CO2	11,22			CO2	10,25		
S3	T1	E1	CO1	7,97	S3	T2	E1	CO1	5,44
			CO2	9,47				CO2	7,8
		E2	CO1	9,37			E2	CO1	6,54
		CO2	11			CO2	10		
	E3	CO1	10,37	E3		CO1	9		
		CO2	12,5			CO2	12		

Vemos que todas las relaciones son estadísticamente significativas a través del Análisis de Varianza (o Kruskal-Wallis).

¿Cómo podemos usar estos datos para responder a la pregunta de cuál Servicio trabaja con menos días de estancia, sin tener en cuenta los factores confusores?

Si la variable dependiente fuera cualitativa, existiría un estadístico específico; aquí de forma sencilla, podemos comparar S1, S2 y S3 a través de la Estancia Media de cada combinación de variables cualitativas. Una forma cómoda de disponer el cuadro para dichas comparaciones sería:

Técnica	Edad	Comorbilidad	S1	S2	S3
T1	E1	CO1	5,94	6,73	7,97
		CO2	7,4	8	9,47
	E2	CO1	6,85	8,18	9,37
		CO2	8,38	9,83	11
	E3	CO1	7,5	9,37	10,37
		CO2	9,29	11,22	12,5
T2	E1	CO1	3,14	4,03	5,44
		CO2	5	6,09	7,8
	E2	CO1	4	5,38	6,54
		CO2	5,57	7,80	10
	E3	CO1	5,29	7,13	9
		CO2	7,5	10,25	12

Por lo tanto, de la comparación horizontal (ajustada por las variables que podrían confundir), vemos claramente que **el Servicio 1 merece una felicitación justificada**. Puede ser útil el hacer un riesgo relativo de cada fila, para saber hasta qué punto hay que felicitarle al Servicio 1, o advertir seriamente al Servicio 3; para aquellos alumnos entregados y laboriosos, les facilitamos una rejilla para el cálculo...podrán ellos mismos comprobar que parte importante de las mayores estancias del Servicio 3 sobre el 1 (Riesgo Relativo de 1,61) se han disuelto al hacer la comparación estratificada...Para aquellos alumnos menos laboriosos, les damos hecha la tabla...

Tabla para cálculos (mirar)

S3	S1	S3/S1
7,97	5,94	
9,47	7,4	
9,37	6,85	
11	8,38	
10,37	7,5	
12,5	9,29	
5,44	3,14	
7,8	5	
6,54	4	
10	5,57	
9	5,29	
12	7,5	

Vemos que para la Técnica 2 (parte blanca del cuadro), el Servicio 3 resulta especialmente ineficiente en el uso de estancias.

Tabla con soluciones (no mirar)

S3	S1	S3/S1
7,97	5,94	1,34
9,47	7,4	1,27
9,37	6,85	1,36
11	8,38	1,31
10,37	7,5	1,38
12,5	9,29	1,34
5,44	3,14	1,73
7,8	5	1,56
6,54	4	1,63
10	5,57	1,79
9	5,29	1,70
12	7,5	1,6

8- La evaluación de tecnologías: ¿qué cirugía gasta menos estancias?

Como en este punto del Taller los alumnos habrán adquirido una importante pericia, les invitamos a que reflexionen y respondan solos a la pregunta anterior..

Recordamos que puede ser interesante un primer análisis general (que en éste caso ya hicimos en el apartado 6), pero luego es inevitable hacer un estudio estratificado por las variables COMORBILIDAD y EDAD.

Finalmente, y para los alumnos aventajados, se trataría de plantearles la siguiente pregunta complementaria: **¿cuál de los servicios es más eficiente –estancias- en el uso de la tecnología convencional y cuál es el más eficiente en la laparoscópica?** (ah...diferencias significativas, por supuesto).

Los resultados de dicho estudio se reseñan a continuación.

La estancia media de T1 es de 8,72 días, mientras que la de T2 es de 5,56 (significativo, y cociente T1 / T2 de 1,57) (como se analizó en el apartado 6).

El análisis estratificado daría los siguientes resultados, teniendo todas la comparaciones entre T1 y T2 de los distintos grupos significación estadística (se incluye el cálculo del Cociente de T1 / T2)

(Se invita al alumno a calcular proporción atribuible)

Comparación de medias de ESTANCIAS por TÉCNICA estratificado por EDAD y COMORBIL

Cociente (Prop.Atrib)

T1	E1	CO1	7,35	1,82 ()	4,03	CO1	E1	T2
		CO2	8,68	1,40 ()	6,18	CO2		
	E2	CO1	8,79	1,89 ()	5,17	CO1	E2	
		CO2	10,07	1,27 ()	7,9	CO2		
	E3	CO1	9,61	1,45 ()	6,59	CO1	E3	
		CO2	11,5	1,24 ()	9,3	CO2		

¿qué llama la atención de estos datos?

...PISTAS... ¿acaso no parece que la técnica laparoscópica donde es realmente mejor es cuando se aplica en pacientes menos ancianos y sin comorbilidad?...reflexionar mirando a la tabla.

Para aquellos alumnos aventajados que hayan investigado la eficiencia relativa de los 3 servicios con las dos tecnologías, habrán visto que el Servicio 1 es mucho más eficiente en el consumo de estancias que el Servicio 3, especialmente en el caso de la nueva tecnología en la cual el sobreconsumo de estancias en el Servicio 3 es de más de un 60%.

9- Correlación costes y estancias.

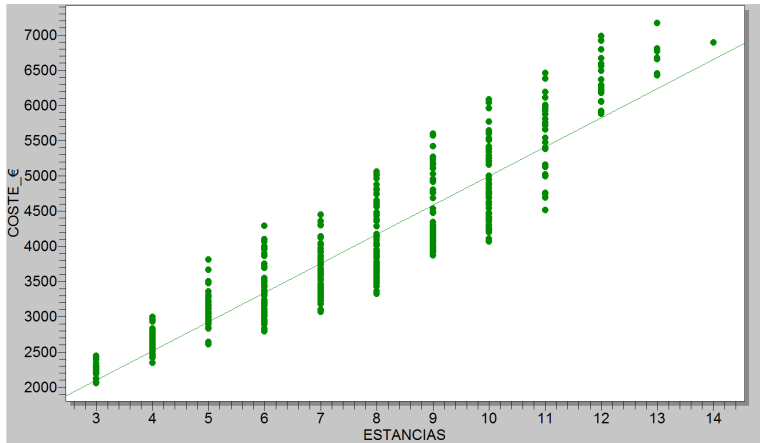
Nos dirigimos ahora a una variable como los COSTES que se encuentra "aguas abajo" de las estancias; además son dos variables de resultados que tienen carácter cuantitativo (una discreta y otra continua), y que nos vienen muy bien para introducirnos en la estadística de correlación.

El método que usaremos en principio será el de la regresión lineal; el término parece difícil pero el concepto no es demasiado complejo, y además el ordenador hace los cálculos por nosotros.

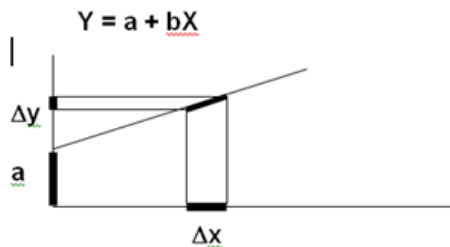
Si tomamos las ESTANCIAS como variable independiente (explicativa) y el COSTE como variable dependiente (explicada por las estancias), y lo llevamos a un eje de coordenadas donde la X sean las estancias y la Y el coste, nos saldrá una nube de puntos (cada punto es una intersección de valores de X y su correspondiente de Y). No hablemos más...en Epi-Info **localice la orden Gráficos ...**

The screenshot shows the 'GRÁFICO' window in Epi-Info. The 'Tipo de Gráfico' is set to 'Dispersión XY'. The '1er Título | 2º título' field is empty. The 'Plantilla' field is also empty. The 'Variables' section has 'Etiqueta' empty and 'Variables X, Y (y Z)' set to '|'. The 'Eje Y' section has 'Etiqueta' empty. The 'Series' section has 'Estilo por cada valor de' set to a dropdown menu and 'Un gráfico por cada valor de' set to a dropdown menu. The variable list at the bottom includes 'ESTANCIAS' and 'COSTE_€'.

Vemos un bello gráfico en el cual se aprecia una clara tendencia creciente de los costes según aumentan los días de estancia. La nube de puntos presenta una cierta regularidad ascendente; la regresión lineal busca una recta que produzca el mejor ajuste posible en esa nube de puntos (minimizando la varianza entre los valores teóricos de Y y los valores reales de Y).

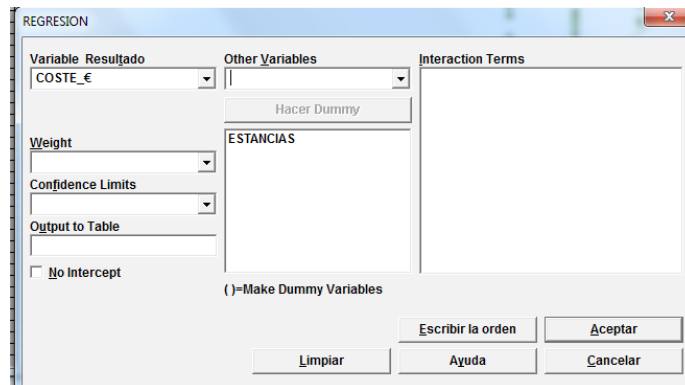


Recordamos la fórmula general de la recta...



En la fórmula, "**a**" representa el punto de corte entre la recta y el eje Y de ordenadas (interceptor), y "**b**" representa la pendiente: lo que crece el valor de y (Δy) por cada unidad de aumento en el valor de x (Δx); por lo tanto la pendiente $b = \Delta y / \Delta x$.

También veremos la recta de regresión ... Si pedimos a Epi-Info que nos calcule la ecuación de esta recta (regresión de COSTE sobre ESTANCIAS) localice la orden **Regresión lineal**



...nos devuelve la información que precisamos para trazar la recta de regresión: la intersección y (valor a de la ecuación anterior, con un valor de 858,62 €), y de la pendiente b , para la cual nos da el valor de 413 €.

REGRESS COSTE_€ = ESTANCIAS

[Siguiete procedimiento](#)

Regresión lineal

Variable	Coefficiente	Error Estandar	Valor F	Valor p
ESTANCIAS	413,785	7,844	2782,4549	0,000000
CONSTANT	858,621	61,563	194,5167	0,000000

Coefficiente de correlación: $r^2 = 0,85$

Fuente	gl	Suma de Cuadrados	Media de cuadrados	Estadístico F
Regresión	1	513660423,489	513660423,489	2782,455
Residuales	498	91934245,987	184606,920	
Total	499	605594669,476		

El estadístico que nos permite medir el grado de correlación entre las dos variables es el **coeficiente de correlación de Pearson**, y su notación es una r ; la r puede oscilar entre 1 y -1 ; cuando es 1 se habla de una correlación positiva perfecta (todo incremento de x explica todo el incremento de y); cuando es -1 se habla de una correlación negativa perfecta (todo incremento de x explica todo el decremento de y); cuando es 0, implica que

ninguna variación de x explica o predice ninguna variación de y (sería una nube de puntos sin ningún patrón o regularidad, o bien significaría que los valores se agrupan en torno a una recta paralela al eje de las X, y por tanto la Y siempre tiene el mismo valor con independencia del valor que tome la X).

El valor que nos dio Epi-Info de la r es de 0,92 (se calcula haciendo la raíz cuadrada de $r^2=0,85$), lo cual implica un importantísimo nivel de correlación positiva.

Esta $r^2 = 0,85$ indica qué parte de la varianza de los costes viene explicada por las estancias; en este caso es de 0,85 (un 85%)

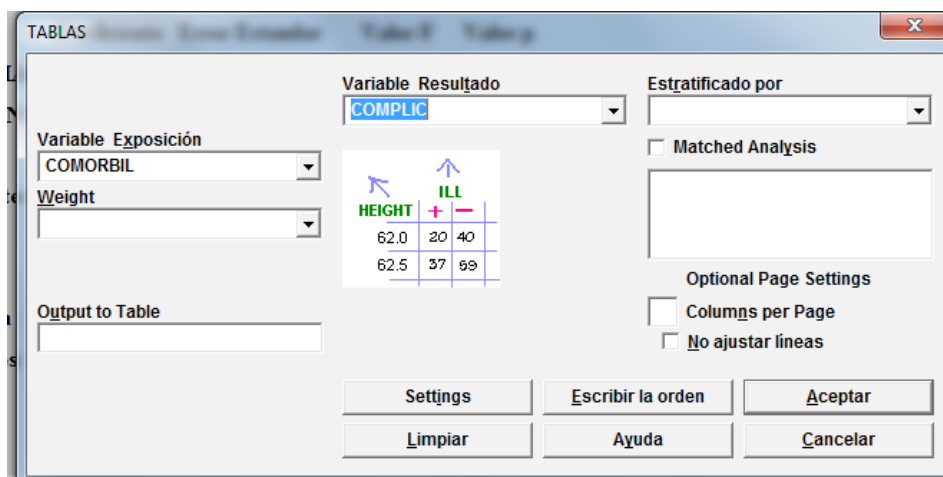
10- Las complicaciones

Vamos finalmente a analizar la tercera variable de resultados, que son las COMPLICACIONES; es una variable cualitativa dicotómica (no hay / si hay), y ello nos permite utilizar otra herramienta estadística.

En primer lugar, cabría hacer un análisis bivalente con aquellas variables que consideremos más importantes a efectos de la aparición de complicaciones: tanto las exógenas EDAD y COMORBILIDAD, como aquellas endógenas SERVICIO y TECNICA.

Como en este caso todas son cualitativas, el estadístico de comparación es el χ^2 ya explicado, y la función del Epi-Info es **Tablas**

Al aplicar la función **Tablas** con dos variables cualitativas dicotómicas, por ejemplo COMORBILIDAD COMPLICACIONES,



Epi-Info nos ofrece los resultados en tablas 2x2 como las que habitualmente se estudian en epidemiología. En las tablas donde una de las variables tiene tres categorías (por ejemplo **COMPLICACIONES SERVICIOS**), Epi-Info nos ofrece el χ^2 del conjunto de la tabla, que revela la significación estadística de la relación de ambas variables tomando en cuenta todos los valores de las casillas.

Realice la tabla COMPLICACIONES EDAD

Realice la tabla COMPLICACIONES SERVICIO

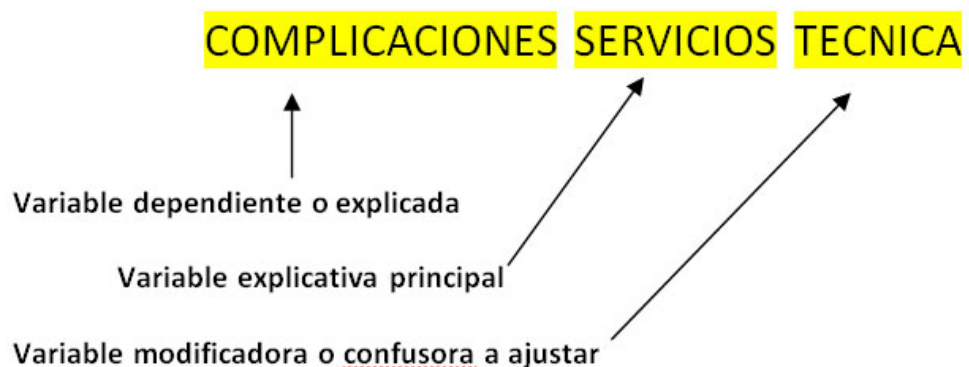
Realice la tabla COMPLICACIONES TECNICA

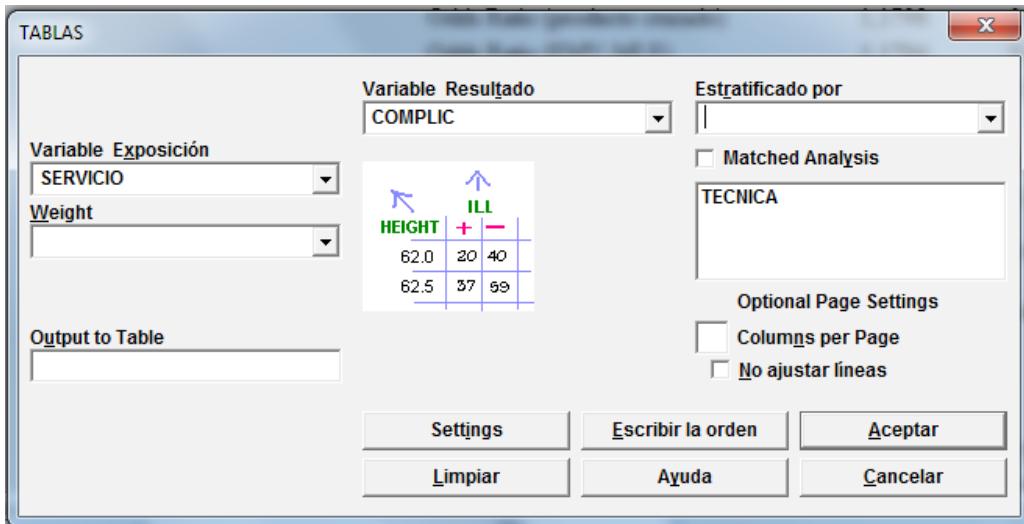
Como veremos en el análisis bivariante, la edad y la comorbilidad influyen de forma estadísticamente significativa en la aparición de complicaciones; los Servicios no tienen influencia significativa, y tampoco la técnica.

A partir de aquí caben hacer análisis estratificados con más de una variable explicativa y de ajuste de las complicaciones.

A modo de sugerencia... **¿si analizamos las complicaciones de los tres servicios, ajustadas por la técnica que utilizan... no parece que el S3 tiene más complicaciones de las esperadas cuando realiza la T2?**

Para hacer esto, y de forma similar a lo que ensayamos en la comparación estratificada de medias de estancias, podemos usar la función **Tablas**:





Vemos con sorpresa, que aunque en la relación bivalente anterior entre COMPLICACIONES Y SERVICIOS no había diferencias significativas, ahora aparecen para la Técnica 2 diferencias que sí lo son entre los servicios...si miramos atentamente al Servicio 3 y sus complicaciones al usar la Técnica 2 nos damos cuenta de donde está el problema... (y las recomendaciones para solucionarlo).

Epílogo

Si ha llegado usted hasta aquí, nuestras más sinceras felicitaciones; si ha seguido fielmente los pasos establecidos, resuelto apropiadamente los ejercicios, y adquirido los conocimientos y habilidades previstas, por favor comuníquenoslo enseguida (sería un caso realmente sorprendente y meritorio). No obstante, si ha conseguido mejorar su comprensión de algunos conceptos fundamentales, y refrescado sus conocimientos, nos sentimos ya profundamente satisfechos.

A través de este taller pensamos que puede haber nacido en usted un renovado interés en el uso de conceptos y técnicas estadísticas para el manejo e interpretación de datos cuantitativos. Y también deseamos que le haya generado el interés necesario para desempolvar (¿comprar si no lo tiene?) un manual de estadística de los muchos que existen.

Anexo:

CLASIFICACIÓN DE LAS VARIABLES

TÉCNICA: Cualitativa nominal o cardinal

EDAD: Cuantitativa discreta

COMORBILIDAD: Cualitativa nominal-dicotómica

SERVICIO: Cualitativa cardinal

ESTANCIAS: Cuantitativa discreta

COSTE: Cuantitativa continua

COMPLICACIONES: Cualitativa nominal-dicotómica

CARACTERÍSTICA DE LAS VARIABLES

Exógenas (independientes) EDAD COMORBILIDAD	Endógenas (independientes) TÉCNICA SERVICIO
De resultados (dependientes) ESTANCIAS COSTES COMPLICACIONES	

