

ORGANIZACIÓN PANAMERICANA DE LA SALUD



Análisis Epidemiológico de Datos Tabulados

VERSIÓN 2.0 PARA WINDOWS

Junio 1997

XUNTA DE GALICIA

EPIDAT 2.0
ANÁLISIS EPIDEMIOLÓGICO
DE DATOS TABULADOS

Programa desarrollado por el
Servicio de Información sobre Saúde Pública
de la Dirección Xeral de Saúde Pública
de la Consellería de Sanidade e Servicios Sociais (Xunta de Galicia),

como parte integrante del convenio de colaboración técnica con la
Unidade de Bioestatística de la Facultade de Medicina
(Universidad de Santiago de Compostela)

y de la carta de entendimiento con la

Organización Panamericana de la Salud (OPS-OMS)



**Consellería de Sanidade
e Servicios Sociais**

Avda. do Camiño Francés 10, baixo
15771 Santiago de Compostela
A Coruña. España



Organización Panamericana de la Salud

525 23rd Street, N.W.
Washington D.C. 20037
E.U.A.

Grupo de trabajo

Análisis y programación:

Lorenzo Rodríguez Currais

Área de epidemiología:

Carlos Castillo Salgado

Enma Fernández Nogueira

Xurxo Hervada Vidal

Área de estadística:

José M^º Alonso Meijide

Tomás Cotos Yáñez

Rubén Fernández Casal

M^º José Lombardía Cortiña

César Sánchez Sello

Coordinación:

Enrique Vázquez Fernández

Las opiniones expresadas por los autores son de su exclusiva responsabilidad y no reflejan necesariamente los criterios ni la política de la Organización Panamericana de la Salud ni de los Gobiernos miembros.

El grupo de trabajo agradece la colaboración prestada por Fernando Rodríguez Artalejo en el planteamiento y diseño de este programa. Y a Enrique Loyola Elizondo y Óscar Mújica Lengua por su colaboración en la revisión del manual.

© Organización Panamericana de la Salud, 1997

© Xunta de Galicia, 1997

1^ª Edición

Las publicaciones de la Organización Panamericana de la Salud están acogidas a la protección prevista por las disposiciones sobre reproducción de originales del Protocolo 2 de la Convención Universal sobre derechos de autor. Reservados todos los derechos.

La Organización Panamericana de la Salud y la Consellería de Sanidade e Servizos Sociais darán consideración muy favorable a las solicitudes de autorización para reproducir o traducir, íntegramente o en parte, esta publicación. Las solicitudes y las peticiones de información deberán dirigirse a la Consellería de Sanidade e Servizos Sociais.

La mención de determinadas sociedades mercantiles o del nombre comercial de ciertos productos no implica que la Organización Panamericana de la Salud los apruebe o recomiende con preferencia a otros similares no mencionados. Salvo en casos de error u omisión, las marcas registradas se destacan con letra mayúscula inicial

"Este libro está especialmente destinado a los estudiantes de ciencias de la salud de América Latina y se publica dentro de los programas de educación de la Organización Panamericana de la Salud, organismo internacional constituido por los países de las Américas, para la promoción de la salud de sus habitantes. Se deja constancia que este programa está siendo ejecutado con la cooperación financiera del Banco Interamericano de Desarrollo".

La Organización Panamericana de la Salud y la Consellería de Sanidade e Servizos Sociais de la Xunta de Galicia han firmado cartas de entendimiento con el objetivo de colaborar en tareas de investigación y capacitación en el área de la salud. Los resultados obtenidos de esta colaboración han sido muy satisfactorios para ambas instituciones. Las diversas actividades conjuntas realizadas y en particular los numerosos cursos, talleres y asesoramientos efectuados han sido de gran utilidad para las instituciones de salud de los países de las Américas y Galicia. Una importante línea de trabajo ha sido el desarrollo, la producción y la difusión de dos importantes programas informáticos: EPIDAT 1.0 y RSIDA, los cuales han sido incorporados en múltiples instituciones de 23 países como herramientas de ayuda al análisis epidemiológico y de la situación y tendencias del sida.

Avanzando en esta línea de apoyo al trabajo analítico de los profesionales de la salud, nos complace presentar aquí la segunda versión del programa EPIDAT. En esta versión se pueden destacar dos características básicas: la sencillez de su uso y la profundización en las técnicas modernas de análisis epidemiológico, usadas ampliamente en los servicios de salud. El programa incluye temas cotidianos para los epidemiólogos, como el ajuste de tasas, y otros inéditos en este tipo de programas, como un completo módulo de análisis de concordancia y validez. Asimismo, se han introducido procedimientos que permiten responder a una diversa gama de posibles usuarios, tales como epidemiólogos, gestores y docentes. Esta edición está dirigida especialmente a los profesionales de habla hispana, entre los que ya tuvo una excelente acogida la primera versión.

Estamos seguros de que el programa EPIDAT 2.0 podrá complementar a otras herramientas y programas de uso rutinario en las instituciones de salud y servirá de apoyo a los análisis de situación de salud y sus tendencias. Con su publicación y distribución, ofrecemos a los profesionales de la salud un instrumento útil que contribuirá modestamente a la consecución de las metas de salud que nos hemos propuesto.

George A.O. Alleyne
Director de la Oficina
Sanitaria Panamericana

José M^a. Hernández Cochón
Conselleiro de Sanidade
e Servizos Sociais

Novedades de la versión 2.0	11
Instalación	13
Entorno general de la aplicación	15
Ajustes	21
Método directo	23
Método indirecto	28
Muestras	35
Situación con una muestra. Proporción	37
<i>Cálculo con una proporción</i>	37
<i>Prueba de hipótesis</i>	39
Situación con una muestra. Media	40
<i>Cálculo de una media</i>	40
Situación con dos muestras. Proporción	41
<i>Cálculo de una diferencia</i>	41
<i>Prueba de hipótesis</i>	42
Situación con dos muestras. Media	44
<i>Muestras independientes.</i> <i>Cálculo de una diferencia</i>	44
<i>Muestras dependientes. Prueba de hipótesis</i>	45
Estudios casos–controles	46
<i>Pareado</i>	46
<i>No pareado</i>	47
Estudios de cohortes	49
<i>Estimación del riesgo relativo</i>	49
Verificación de la calidad de lotes	50
<i>Aceptación de una prevalencia</i>	50
<i>Rechazo de un lote</i>	51
Generación de números aleatorios	54
Concordancia	55
Dos observadores	57
<i>Dos categorías</i>	57
<i>Múltiples categorías</i>	59

Múltiples observadores	61
<i>Dos categorías</i>	61
<i>Múltiples categorías</i>	62
Hipótesis de igualdad de kappas	65
Tablas	67
Análisis de tablas	69
Tablas 2x2	69
Tablas MxN	79
Vigilancia Epidemiológica	83
<i>Captura-Recaptura</i>	83
<i>Detección de clusters</i>	85
Pruebas diagnósticas	87
Pruebas diagnósticas simples	89
Combinación de pruebas diagnósticas	92
<i>Datos tabulados</i>	92
<i>Valores diagnósticos</i>	94
Curvas ROC	96
<i>Tablas 2x2</i>	96
<i>Tablas 2xN</i>	98
Inferencias	101
Inferencia para la media. Una muestra	103
Inferencia para la proporción. Una muestra	105
Inferencia para la diferencia de medias. Dos muestras	106
<i>Independientes</i>	106
<i>Dependientes</i>	107
Inferencia para la diferencia de proporciones. Dos muestras	109
<i>Independientes</i>	109
<i>Dependientes</i>	111
Priorización	113
Determinación de necesidades en salud	115
Anexo I	119
Anexo II	123



EPIDAT 2.0

EPIDAT 2.0 está desarrollado para Windows® y las principales diferencias con la versión 1.0 son las siguientes:

- Ajuste de tasas** Permite introducir los datos manualmente o importarlos en formato .DBF. Pueden ajustarse simultáneamente tantas tasas como desee el usuario.
- Concordancia** Incorpora el análisis de concordancia entre múltiples observadores, tanto para dos categorías como para múltiples categorías de clasificación.
- Tablas** Analiza tablas 2x2 que provengan de estudios transversales, caso-control o cohortes, en tablas simples o estratificadas. En los estudios caso-control pareados admite hasta 4 controles por caso. Añade el método Bootstrap para el cálculo de intervalos de confianza. En tablas MxN realiza cálculos específicos si las categorías son ordenadas. Incorpora el método de captura-recaptura de Chandra-Sekar-Deming y el de detección de clusters de Knox.
- Pruebas diagnósticas** Calcula la sensibilidad, especificidad y valores predictivos de pruebas diagnósticas simples, en serie y en paralelo. En las pruebas combinadas realiza los cálculos tanto si dispone de los valores de las celdas como de valores de sensibilidad y especificidad de cada prueba. Permite dibujar curvas ROC calculando el área bajo la curva y el punto óptimo de corte si se dispone de la prevalencia y de la relación de costes.
- Priorización** Propone un método sencillo (CENDES/OPS) de determinación de prioridades en salud a partir de indicadores y áreas definidos por el usuario.

EPIDAT es un programa de libre distribución desarrollado por instituciones públicas. Por ello, no sólo se permite, sino que se agradece su difusión y cualquier tipo de crítica o comentario que ayude a mejorar futuras versiones.

Las instituciones responsables desean mostrar su agradecimiento a la Organización Mundial de la Salud por el permiso otorgado para incorporar, en el módulo de muestreo de EPIDAT, fragmentos de la publicación de Lwanga S.K. y Lemeshow S.. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico*. Ginebra: O.M.S.; 1991.

Para solicitar información adicional, copias o remitir comentarios, dirigirse a:

En España: Servicio de Información sobre Saúde Pública

DXSP

Avda. do Camiño Francés nº 10 baixo

15771 Santiago de Compostela

A Coruña

Tf: 981-542916

Fax: 981-542970

e-mail: dxsp3@jet.es

En América: Dr. Carlos Castillo–Salgado

Organización Panamericana de la Salud

525 23rd Street, N.W.

Washington D.C. 20037-2895

U.S.A.

Tf: 202-9743327

Fax: 202-9743674

e-mail: castillc@paho.org

EPIDAT ha sido desarrollado por el Servicio de Información sobre Saúde Pública de la Dirección Xeral de Saúde Pública de la Consellería de Sanidade e Servizos Sociais (Xunta de Galicia), como parte del Convenio de Colaboración Técnica con la Unidade de Bioestatística de la Facultade de Medicina (Universidade de Santiago de Compostela) y de la Carta de Entendimiento con la Organización Panamericana de la Salud (OPS–OMS) (Programa de Análisis de la Situación de Salud, HDA). Esta última institución se encarga de la supervisión y distribución para las Américas.

AVISO: EPIDAT es un programa de libre distribución. Los miembros del grupo de trabajo y las organizaciones que los sustentan declinan toda responsabilidad derivada del uso inadecuado del programa.

Requerimientos mínimos

- Ordenador IBM-PC o compatible, con disco duro y disquetera de 3,5"
- Procesador 386 o superior
- Tarjeta VGA
- Windows 3.1® o superior
- 5 Mb de espacio disponible en disco

Proceso de instalación

- Ejecutar: A:\instalar desde Windows®
- Una vez iniciado el proceso seguir las instrucciones que aparecen en pantalla
- EPIDAT crea una carpeta y un icono

El entorno general de la aplicación, manejo del ratón, teclado, configuración de impresoras, etc., es el de Windows®, ante cualquier duda consulte el manual de la versión de este sistema operativo que tenga instalada en su computadora.

Problemas tras la instalación

Si tras la instalación tiene algún problema (visualización incorrecta, ventanas de error en los módulos,...) puede ser debido a la configuración que se tenga instalada en Windows®.

Para más información consulte el HELP de EPIDAT 2.0. Este fichero puede leerse desde Windows® utilizando el explorador. Para ello, debe localizar el archivo EPIDAT.HLP en el directorio donde haya instalado EPIDAT 2.0 y abrirlo haciendo doble “click” sobre él.

Error de aplicación

- Imagine que se ha producido un Error que la Aplicación no puede resolver. Si tiene una impresora conectada, pulse la tecla “IMPR PANT” (PRINT SCREEN). Anote la fecha y hora del error y, si lo desea, remítala a una de las direcciones indicadas en la presentación de este manual.
- Pulse una tecla para Salir de la Aplicación.

Error recuperable

Suponga que se ha producido un error en un proceso debido al motivo que le ha sido especificado.

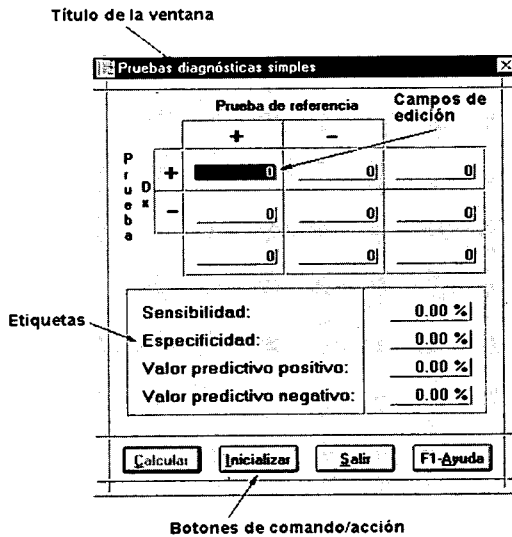
Se dispone de dos alternativas:

- Reintentar el proceso. Si es preciso, asegúrese de que la causa del error ha desaparecido o se ha resuelto. De no ser así se presentará de nuevo el error hasta que seleccione la otra alternativa.
- Ignorar el error. Se continua el proceso como si el error no se hubiese producido. Esta opción puede tener diversas implicaciones dependiendo del motivo y el lugar en que se produjo el error. Si es posible, es preferible utilizar siempre la otra alternativa.

Ver también: Error de la Aplicación.

En esta sección se detallan ciertos parámetros concernientes al manejo general de la aplicación, para ello se describen con detalles algunas pantallas que generalizan la operaciones básicas del entorno.

Ventanas comunes:



Título de la pantalla es la referencia a la selección de menú, indica el diálogo a ejecutar, la ventana esta definida como "NOCLOSE", lo que permite una salida indebida o errónea de la operación.

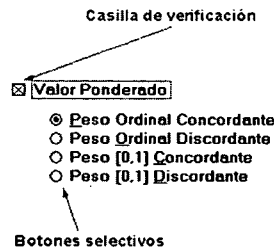
Campos de edición son los campos definidos como de introducción de datos, los que permiten al usuario proporcionar los datos para el proceso. Los campos contienen máscaras y validaciones según las definiciones estadísticas propias a cada ventana. Algunos de ellos se muestran de forma protegida, ya que son calculados de forma automática. También suelen aparecer con asteriscos, definición que se da a aquellos que no pueden ser calculados. Los campos que impliquen procesos de este tipo, pueden aparecer en la pantalla de entrada de datos o de resultados indistintamente.

Etiquetas definen en general referencias al campo de edición.

Botones de comando/acción por medio de estos se generan las acciones propuestas a cada proceso y no requieren ningún tipo de confirmación. El botón de acción que permite el cálculo asociado permanece desactivado, hasta que se complete la carga de datos mínimos para dicha ejecución.

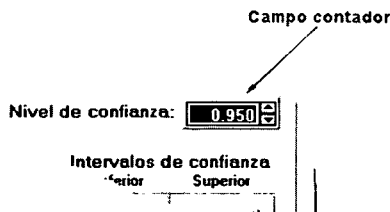
Objetos:

Se definen como objetos campos de edición, introducción, línea de datos, botones, listas, botones selectivos, y otros que se detallan a continuación y por separado para su mejor aprovechamiento:



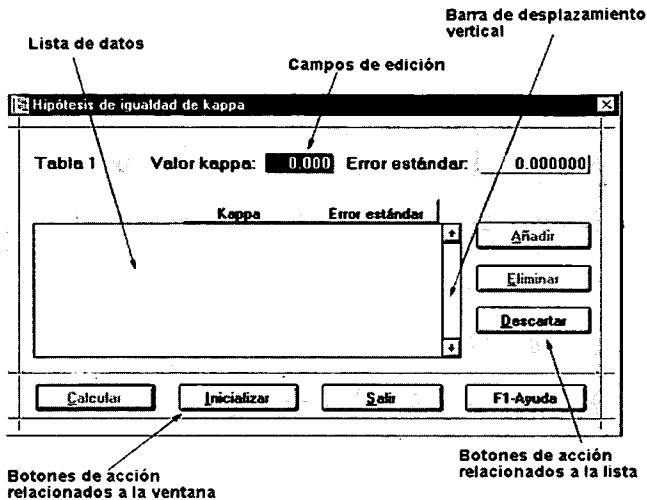
Casilla de verificación puede tomar dos estados, activado o desactivado, en el ejemplo se muestra con el estado de activado, lo que condiciona la acción, en este caso la habilitación de los botones selectivos.

Botones selectivos la elección de uno de ellos, invalida los otros, se presentan una serie de opciones que permiten al usuario la selección de un método o proceso sin temor a equivocarse ya que, por definición, al menos una debe ser elegida.



Campo contador tienen la propiedad de ser campos de edición con la característica de incrementar o disminuir en una cantidad constante el valor reflejado, en este caso permite un mínimo de 0,500 hasta un máximo de 0,995 en incrementos de 0,005.





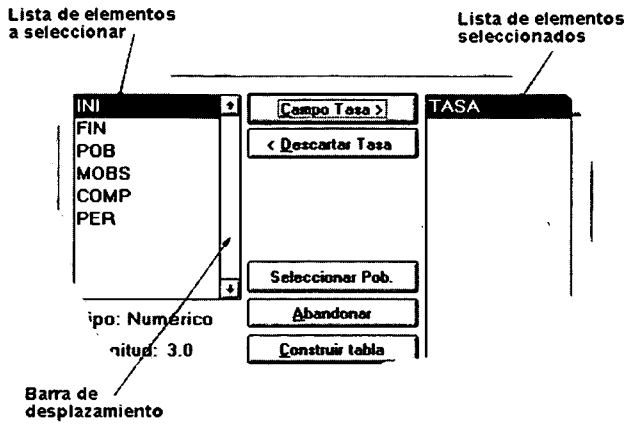
Listas de datos estos objetos permiten carga masiva de datos. EPIDAT, maneja gran cantidad de diálogos con estas características, con la finalidad de proporcionar al usuario flexibilidad a la hora de la introducción. El límite de carga viene sujeto a la capacidad de memoria y almacenamiento del ordenador donde se ejecute.

Existen varios objetos dentro de este tipo de ventanas que permiten un manejo adecuado de los datos.

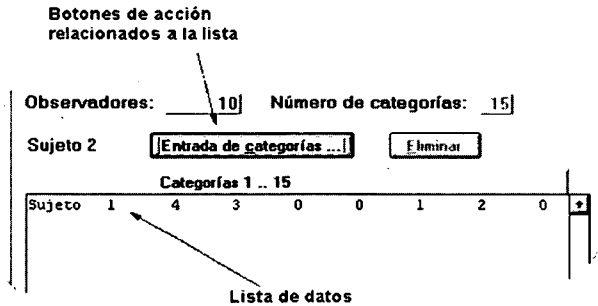
Botones de acción de la lista estos proporcionan una acción directa sobre la carga de datos de la lista. En el ejemplo propuesto, se añadirán tablas o filas compuestas por dos campos o columnas, una vez aplicadas las reglas de validación, se activa el botón “AÑADIR”, lo que permite agregar filas a la lista de datos. El objeto lista tiene la propiedad de poder ser seleccionada, lo cual permite una modificación sobre los datos. En el momento de la selección de la fila, los datos son cargados en los campos de edición, y el botón “AÑADIR”, cambia su contenido por el de “MODIFICAR”, y el botón “ELIMINAR” se activa, con lo cual el usuario elige la acción a tomar. El botón “DESCARTAR”, siempre activo, cumple una doble función ya que inicializa la acciones correspondientes a los botones de acción de la lista, permite blanquear dichos campos y anular la opción de carga para “MODIFICAR” o “ELIMINAR”, si es el caso.

La barra de desplazamiento vertical se activa cuando hay mas filas de las que se pueden visualizar en el cuadro de la lista.

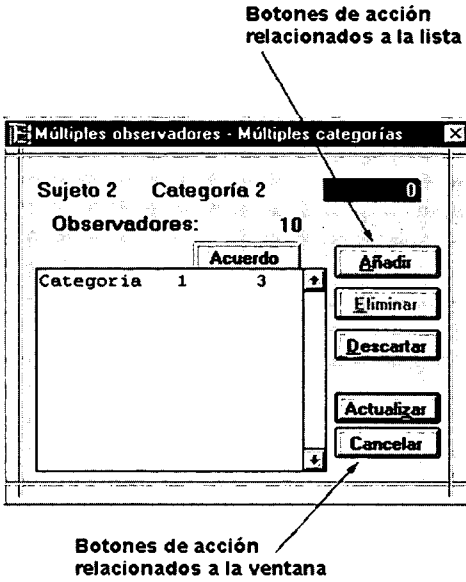
Botones de acción de la ventana tiene el mismo cometido que los descritos en el apartado de botones de comando/acción de ventanas comunes y son inherentes a todo el diálogo.



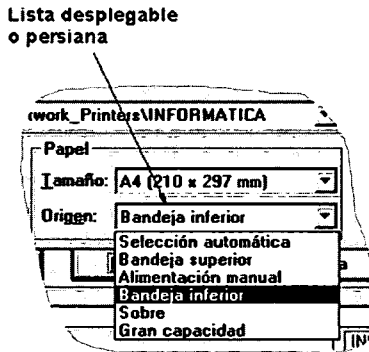
Lista de selección de elementos proporcionan al usuario varios elementos que se encuentran en una lista de elementos disponibles, todos aquellos elementos seleccionados se incluyen en otra lista. Para la inclusión se deben usar los botones asociados que se presentan en medio de ambas listas, su acción es directa sobre las listas y permite en este caso, seleccionar tasas para ajustar y aplica ciertas reglas de validación antes de la carga en la lista de seleccionados.



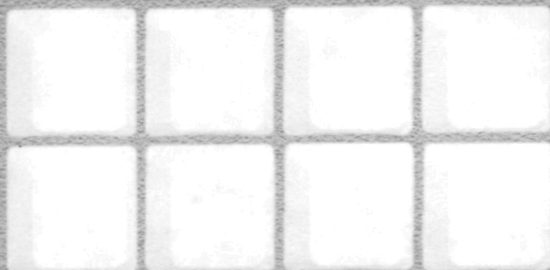
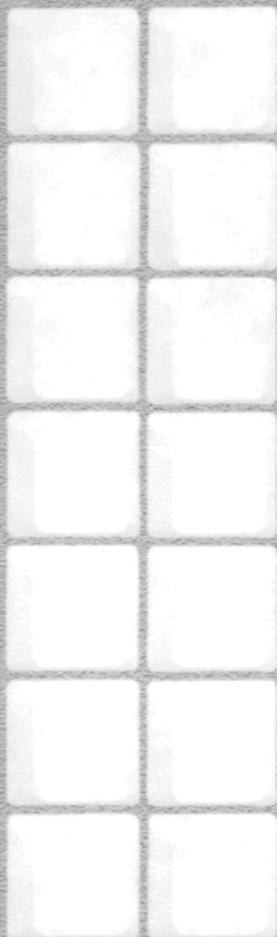
Lista de datos dependientes algunas de las listas de datos dependen de la introducción desde otro diálogo. En el caso del ejemplo, la entrada de datos se hace sobre una lista desplegada por la acción del botón “Entrada de categorías”. Presenta la peculiaridad que los datos en las columnas no se muestran en su totalidad, por lo cual presionando sobre la lista muestra nuevamente la ventana de entrada de datos, pero esta vez propone la modificación.



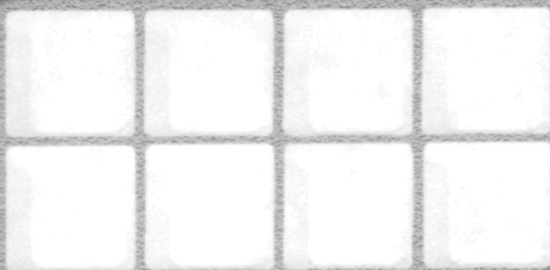
Esta ventana permite la entrada de datos del ejemplo anterior y funciona básicamente igual que las listas de datos comunes. Los datos son validados para su entrada y modificación y son agregados a la lista de control anterior, que es la que permite el proceso de cálculo.



Lista desplegable presenta un gran número de opciones y son excluyentes entre sí. En el EPIDAT sólo se muestra este tipo de objeto en los menús y ventanas de diálogos de impresión.

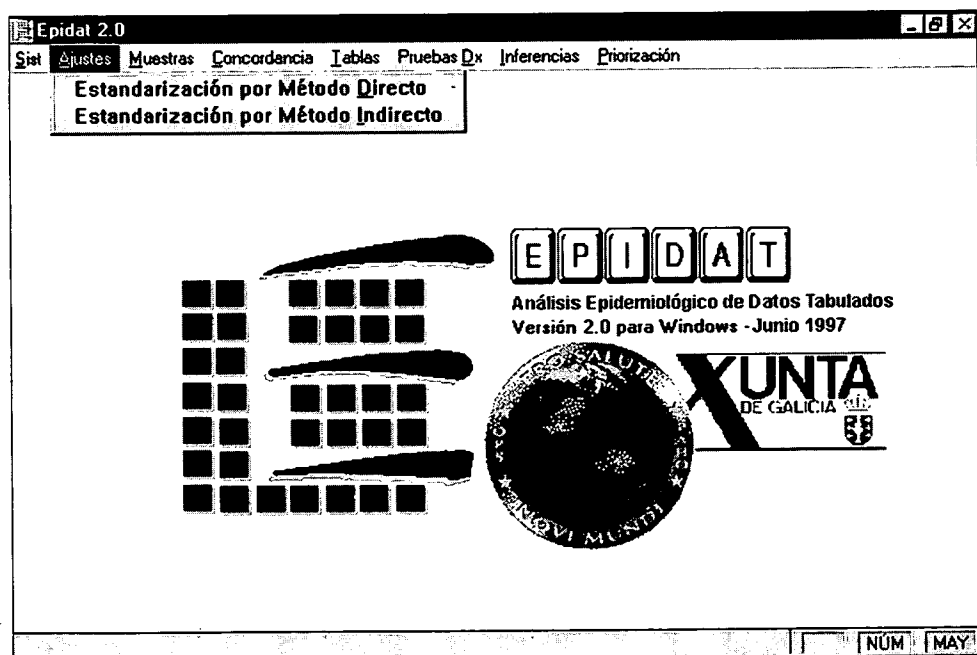


AJUSTES



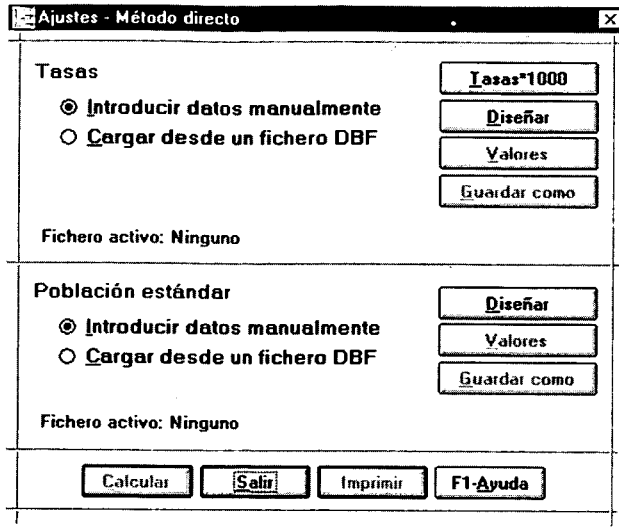
Este módulo permite ajustar tasas por una variable, tanto por el método directo como por el indirecto. Los datos pueden introducirse desde el teclado o importarse en formato .DBF, pudiéndose ajustar simultáneamente un número indefinido de tasas. EPIDAT 2.0 se distribuye con las poblaciones estándar europeas y mundial (Waterhouse, J. *et al.* eds.: *Cancer incidence in five continents*. Lyon: IARC; 1976; 3; 456).

Nota: Cuando se ajustan tasas, tanto por el método directo como por el indirecto, es necesario que los intervalos de las tasas y poblaciones sean idénticos. P. e., si se ajustan tasas por edad es preciso que los grupos de edad (intervalos) sean idénticos en todas las tasas y poblaciones.



Método directo

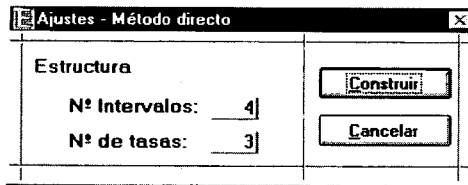
Esta opción permite ajustar tasas por el método directo. Para ello se necesita disponer de las tasas brutas que se quieren comparar por cada categoría de comparación (p. e. grupos de edad) y de la población que se desea utilizar como estándar. El número de categorías ha de ser el mismo. Si esto no sucediese, el programa utilizará solamente las primeras categorías de la población estándar que tienen correspondencia con las categorías de las tasas a comparar. Por ejemplo, si disponemos de las tasas en 5 grupos de edad y le aplicamos una población estándar de 8, EPIDAT sólo utilizará los 5 primeros de esta última.



Hay dos opciones para cargar los datos a ajustar:

Introducir los datos manualmente. —Deben seguirse los siguientes pasos:

- Pulsando el botón “Tasas*1000” repetidamente, seleccionar la opción en que estén expresadas las tasas a ajustar.
- Marcar “Diseñar” e introducir el número de tasas a ajustar y el de intervalos.



- Oprimiendo “Construir” se vuelve a la pantalla anterior en la que ya está activado el botón “Valores”. Pulsándolo aparecerá una tabla con tantas filas como intervalos y dos columnas por cada tasa, T1, P1,... Tn, Pn. En las columnas “T” deben introducirse las tasas de cada intervalo. Si se quieren obtener intervalos de confianza, en las columnas P

Interv	T1	P1	T2	P2	T3	P3
1	12.2500	14500.0000	7.2000	7800.0000	15.3200	9580.00
2	35.1400	18000.0000	39.4700	19650.0000	27.6000	14000.00
3	54.5800	25485.0000	65.2100	26840.0000	53.2600	26650.00
4	78.4600	3600.0000	58.3000	42000.0000	73.5400	24150.00

deberán introducirse los valores de las poblaciones (P1,... Pn) con que se obtuvieron las tasas. Si no interesan los intervalos de confianza, o no se dispone de las poblaciones, EPIDAT calculará las tasas ajustadas ocupando con valores "0" las celdas correspondientes a los IC.

- Introducidos los valores de las tasas y/o poblaciones, se abandonará la pantalla pulsando el icono de EPIDAT situado en la esquina superior izquierda de la ventana. De las opciones desplegadas seleccionar "Cerrar" o "Siguiete ventana".
- El proceso para introducir manualmente la población estándar es muy similar al descrito anteriormente.

Interv	Población
1	50000
2	200000
3	250000
4	120000

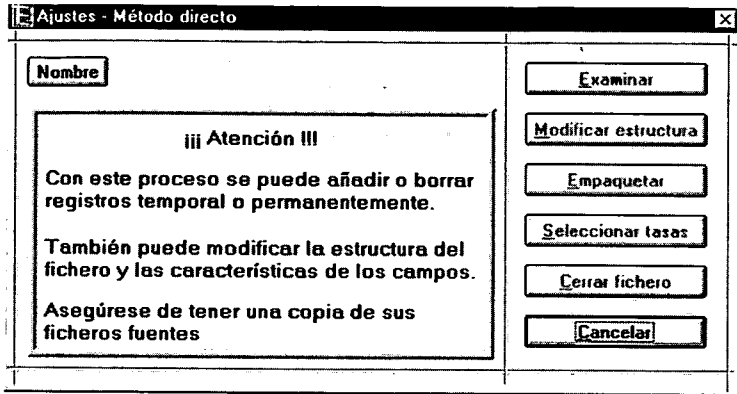
- Una vez finalizada la introducción manual de datos se activa el botón "Calcular". Pulsándolo aparece la tabla de resultados con los valores de las tasas ajustadas y los intervalos de confianza si en su momento se introdujeron las poblaciones correspondientes a cada tasa.

Texto	Tasaest	Icinf	Icsup
Tasa estandarizada 1	49.52	49.42	49.61
Tasa estandarizada 2	50.89	50.82	50.97
Tasa estandarizada 3	45.85	45.78	45.92

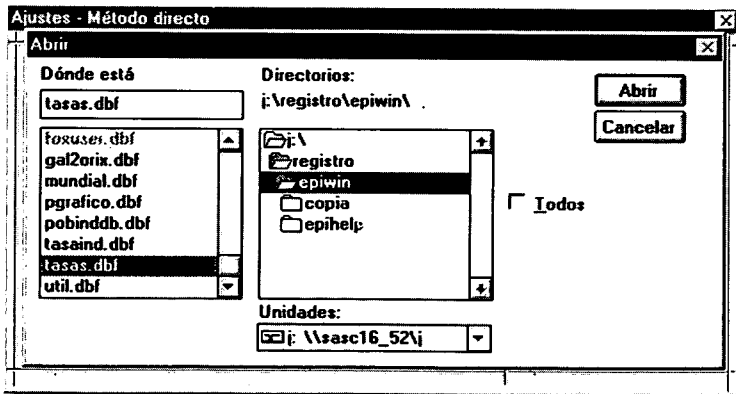
Cargar desde un fichero DBF. —Esta opción permite importar datos de cualquier archivo DBF. Si se desea utilizar datos de otro tipo de archivos (.xls, .wk1...) deberán guardarse previamente en formato .DBF.

Precaución: En la hoja de cálculo Excel las columnas deben etiquetarse previamente a la incorporación con N1, N2, N3, ...sino la primera fila con valores numéricos será eliminada. Si se utiliza otro programa consulte su manual.

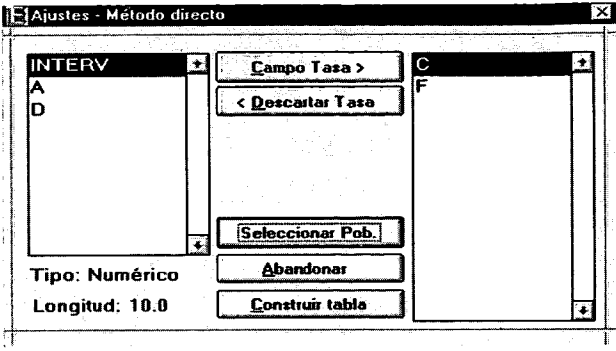
- Seleccionada la opción “Cargar desde un fichero DBF”, presionar “Diseñar”. Aparecerá una ventana que advierte sobre las posibilidades que ofrece esta alternativa y las precauciones aconsejables.



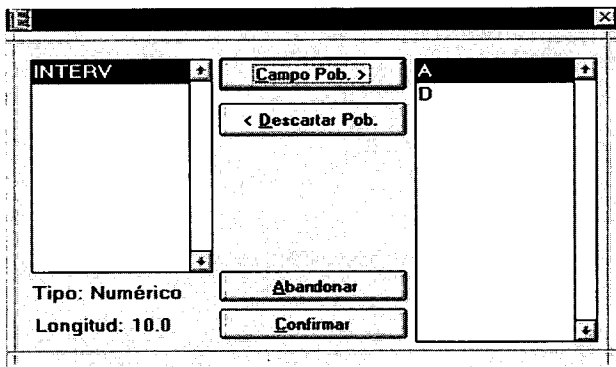
- Pulsando “Nombre” se pasa a una pantalla de búsqueda para facilitar la localización y selección del archivo del que se van a seleccionar las tasas para ajustar.



- Una vez seleccionado un archivo se vuelve a la anterior pantalla donde se ofrecen las posibilidades de “Examinar”, para ver el contenido del archivo, “Modificar estructura”, para, p. e., convertir campos carácter en numéricos y poder utilizarlos en el ajuste, “empaquetar” (eliminar físicamente los registros marcados previamente para borrar), o “Seleccionar tasas”.



- Este último botón abre una ventana en la que aparece a la izquierda el listado de variables del archivo importado. Deben marcarse las que correspondan a las tasas que se quieren ajustar, una a una, pasándolas al espacio de la derecha pulsando “Campo tasa”. Si se dispone de las poblaciones por intervalos e interesa obtener los intervalos de confianza, pulsando el botón “Seleccionar Pob.” aparece otra pantalla donde se marcan los campos que correspondan a las poblaciones **en el mismo orden** que las tasas, pasándolos, uno a uno, al espacio de la derecha, pulsando “Campo Pob.”.



- Una vez concluida la selección de campos de tasas y poblaciones, oprimir “Construir tabla” y se volverá a la pantalla inicial.

Texto	Tasaest	Icinf	Icsup
Tasa estandarizada 1	72.98	72.97	72.99
Tasa estandarizada 2	50.02	50.02	50.02

- Mediante un proceso similar se importará y seleccionará la población estándar. Una vez construida la tabla, EPIDAT volverá a la pantalla inicial y se activará el botón “Calcular”. Pulsándolo, presentará la pantalla de resultados igual a la descrita en la opción de introducir datos manualmente.

BIBLIOGRAFÍA

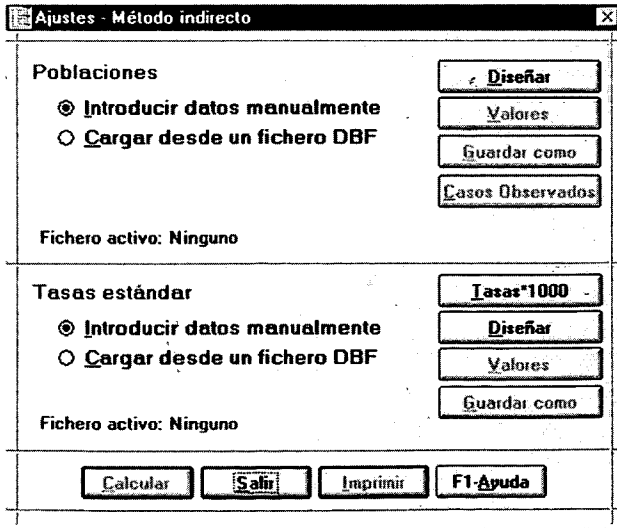
Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Kahn H.A, Sempos C.T. *Statistical methods in Epidemiology*. New York: Oxford University Press; 1989.

Rué M., Borrell C. Los métodos de estandarización de tasas. *Revisiones en Salud Pública*, 3: 263-295.

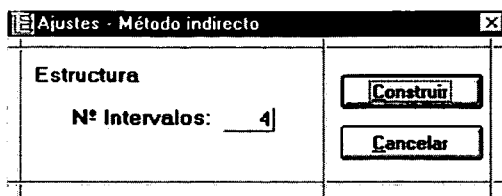
Método indirecto

Esta opción permite ajustar tasas por el método indirecto. Para ello se necesita disponer de las poblaciones por categoría de comparación (p. e. edad) de las que se obtuvieron las tasas a comparar y de la tasa que se utilizará como estándar. El número de categorías ha de ser el mismo. Si esto no sucediese, el programa utilizará solamente las primeras categorías de la tasa estándar que tienen correspondencia con las categorías de las poblaciones de las tasas a comparar. Por ejemplo, si disponemos de las poblaciones de las tasas en 5 grupos de edad y le aplicamos una tasa estándar de 8, EPIDAT sólo utilizará los 5 primeros de ésta última.



Hay dos opciones para cargar los datos a ajustar:

- Pulsando el botón “Tasas*1000” repetidamente, seleccionar la opción en que esté expresada la tasa estándar.
- El proceso para introducir manualmente la tasa estándar es muy parecido al descrito anteriormente.



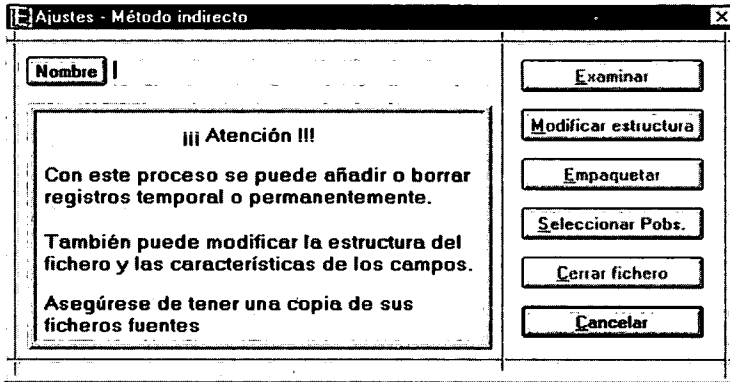
Interv	Tasastan
1	12.100
2	45.300
3	65.200
4	123.900

- Una vez finalizada la introducción manual de los datos se activa el botón “Calcular”. Pulsándolo aparece la tabla de resultados con los valores de las razones de mortalidad estandarizadas (RME o SMR) y sus intervalos de confianza.

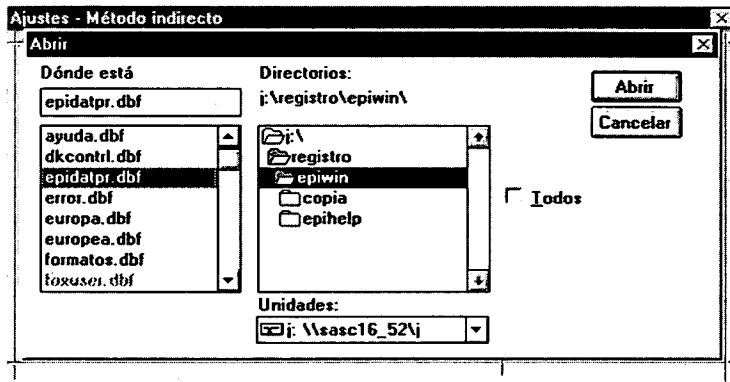
Texto	Rme	Icinf	Icsup
Población 1	123.38297125	93.39	153.38
Población 2	77.35483591	55.91	98.80

Cargar desde un fichero DBF. —Esta opción permite cargar datos desde cualquier archivo DBF. Si se desea utilizar datos de otro tipo de archivos (.xls, .wk1...) deberán guardarse previamente en formato .DBF. **Precaución:** ver aviso de Método-directo (ver pág. 26).

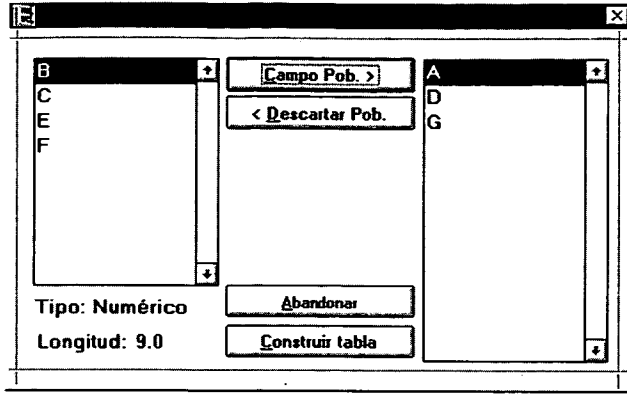
- Seleccionada la opción “Cargar desde un fichero DBF”, presionar “Diseñar”. Aparecerá una ventana que advierte sobre las posibilidades que ofrece esta alternativa y las precauciones aconsejables.



- Pulsando “Nombre” se pasa a una pantalla de búsqueda para facilitar la localización y selección del archivo del que se van a seleccionar las poblaciones de las tasas a ajustar. Una vez seleccionado un archivo se vuelve a la anterior pantalla donde se ofrecen las posibilidades de “Examinar”, para ver el contenido del archivo, “Modificar estructura”, para, p. e., convertir campos carácter en numéricos y poder utilizarlos en el ajuste, “Empaquetar” (eliminar físicamente los registros marcados previamente para borrar), o “Seleccionar Pob.”.



- Este último botón abre una ventana en la que aparece a la izquierda el listado de variables del archivo importado. Deben marcarse las que correspondan a las poblaciones de las tasas que se quieren ajustar, una a una, pasándolas al espacio de la derecha pulsando “Campo Pob”. Una vez concluida la selección de campos de poblaciones, pulsar “Construir tabla” y se volverá a la pantalla inicial.



- A continuación, presionando el botón “Casos Observados” aparecerá una pantalla con una tabla de 1 fila y tantas columnas como tasas a comparar. En cada campo deberá introducirse el total de casos observados en cada población.

Cobs1	Cobs2	Cobs3
72	350	42

- Pulsando el botón “Tasas*1000” repetidamente, seleccionar la opción en que estén expresadas las tasas a ajustar.
- Mediante un proceso similar se importará y seleccionará la tasa estándar. Una vez construida la tabla, Epidat volverá a la pantalla inicial y se activará el botón “Calcular”. Pulsándolo presentará la pantalla de resultados, igual a la descrita en la opción de introducir datos manualmente.

Ejercicio: 1.1. *Un investigador, en busca de factores de riesgo para el cáncer de cuello uterino, desea comparar las tasas de incidencia entre las mujeres de Cali (Colombia) y São Paulo (Brasil). Para ello cuenta con la siguiente información:*

Tabla 1 Población, número de casos y tasas anuales de incidencia de cáncer de cuello uterino en Cali, Colombia 1977–1981 y São Paulo, Brasil 1978.

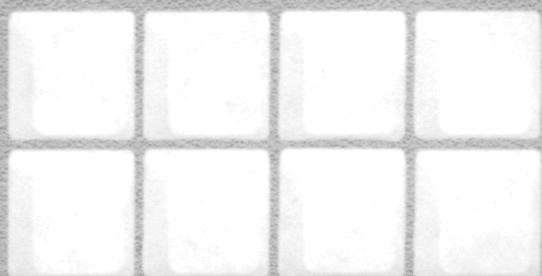
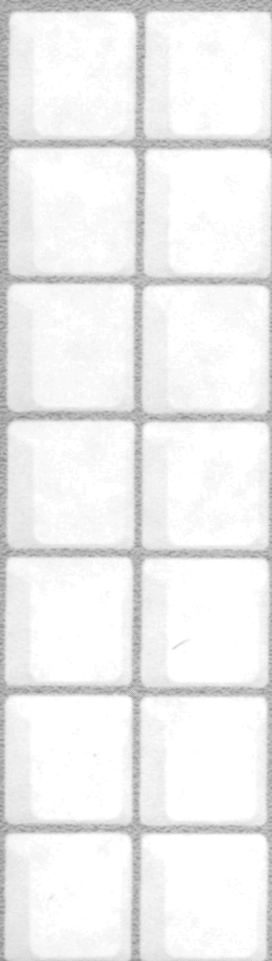
edad	Cali			São Paulo		
	población	número de casos	tasa x 10 ⁵	población	número de casos	tasa x 10 ⁵
0–14	217.645	0	0,0	992.534	0	0,0
15–24	145.409	2	1,4	746.750	14	1,9
25–34	86.644	16	18,5	639.214	76	11,9
35–44	63.454	34	53,6	423.847	195	46,0
45–54	41.180	44	106,8	328.074	266	81,1
55–64	24.551	36	146,6	208.108	228	109,6
65 y más	19.042	37	194,3	173.968	186	106,9
total	597.925	169		3.512.495	965	

Tabla 2 Población estándar: Celade. América Latina, proyecciones de población años calendarios 1950–2000. *Boletín Demográfico*. 1991, (48):31.

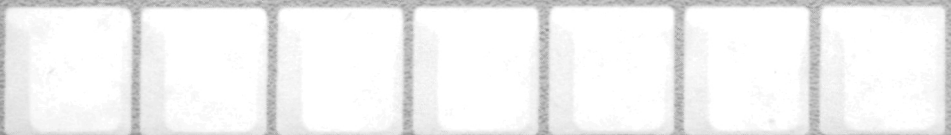
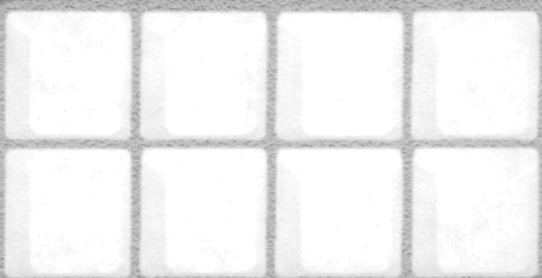
edad	población estándar
0–14	77.500
15–24	43.291
25–34	34.589
35–44	24.275
45–54	16.355
55–64	11.693
65 y más	11.220

- Preguntas:**
1. *Calcule las tasas brutas de incidencia acumulada de cáncer de cuello uterino en Cali y São Paulo.*
 2. *Calcule las tasas ajustadas por edad con sus intervalos de confianza, por el método directo, utilizando como población estándar la propuesta por el Celade (tabla 2).*
 3. *Calcule las tasas ajustadas por edad con sus intervalos de confianza, por el método indirecto, utilizando como tasas estándar las de São Paulo.*

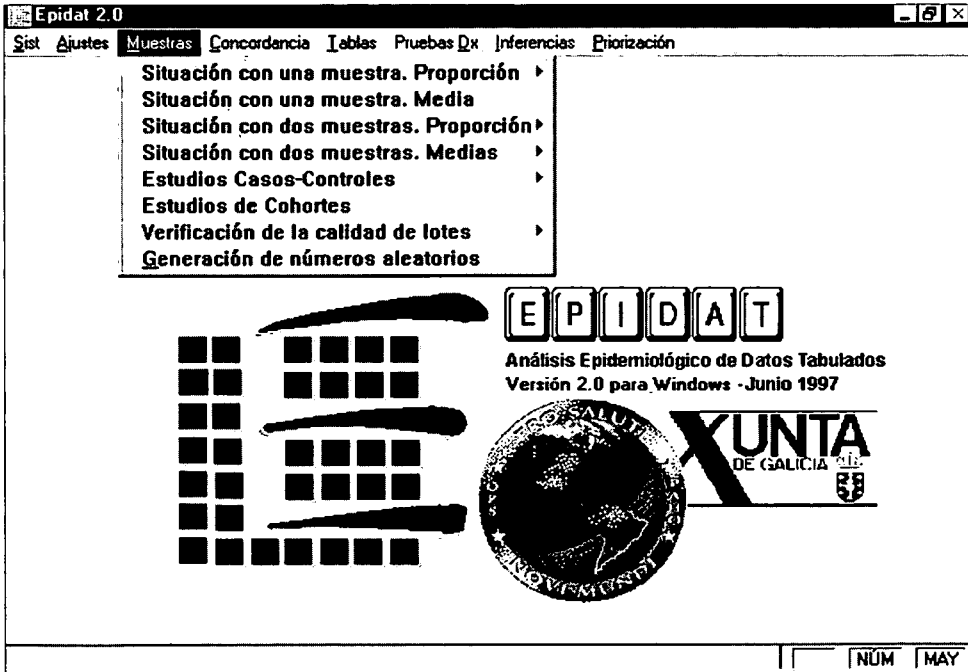
(Adaptado de: Londoño J.L., *Metodología de la investigación epidemiológica*. Ed. Universidad de Antioquia; 1995)



MUESTRAS



Este módulo aborda las situaciones más habituales y simples a las que se enfrentan los profesionales de la salud a la hora de determinar el tamaño de muestras probabilísticas en estudios sanitarios. En las opciones sobre proporciones, cohortes y verificación de la calidad de lotes se basa en el texto de Lwanga S.K. y Lemeshow S., *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico*. Ginebra: O.M.S.; 1991.



Situación con una muestra. Proporción

Cálculo con una proporción

Esta opción calcula el tamaño muestral mínimo necesario para tener una confianza del $100 \times (1 - \alpha) \%$ de obtener una proporción estimada que no difiera del valor verdadero más que la precisión absoluta (d). Para ello, introducir el tamaño de la población de la que se quiere extraer la muestra. Si es desconocida o sobrepasa el rango (999999) dejar el campo con 0. A continuación, introducir el valor aproximado de P , expresada en valores entre 0 y 1 (proporción de población prevista que presenta la característica que se está estudiando, p. e.,

hipertensión, diabetes, etc.). Si se desconoce, deberá asignársele el valor 0,5, opción más segura, dado que el caso más desfavorable (mayor tamaño muestral) se da cuando $P = 0,5$. Por último, el usuario debe definir la precisión absoluta o relativa, **expresada en porcentaje**, con que desea dar sus resultados (error de muestreo). P.e. si la proporción obtenida en un estudio fue del 0,2 (20%) y la precisión absoluta que se utilizó en la determinación del tamaño muestral fue de 0,01 (1%), quiere decir que la estimación va estar entre 0,19 (19%) y 0,21 (21%) y la relativa será del 5% ya que 0,01 (1%), la precisión absoluta, es del 5% de 0,2 (20%). Al introducir una de ellas, EPIDAT calcula la otra. Pulsando “Inicializar” se borran todos los datos introducidos hasta el momento.

Cálculo de una proporción	
Tamaño poblacional (N):	78452
Proporción en la población (P):	0.230
Precisión absoluta (d):	3.000 %
Precisión relativa :	13.043 %

- Una vez suministrados los datos al programa, pulsar “Calcular”. En la pantalla de resultados se muestran por defecto, para los niveles de confianza bilaterales del 99%, 95%, y 90%, los tamaños muestrales mínimos necesarios. El cursor se sitúa en la última posición por si se desea un nivel de confianza diferente de los anteriores.

Nivel de confianza bilateral	Tamaño muestral mínimo
1-alfa: 0.99	1284
1-alfa: 0.95	750
1-alfa: 0.90	530
N. de confianza: 0.950	750

Ejercicio: 2.1. *Un departamento local de salud quiere calcular la prevalencia de reacciones positivas a la tuberculina entre los menores de 5 años de su circunscripción. ¿Cuántos niños han de figurar en la muestra para que pueda calcularse la prevalencia con una precisión de 5 puntos porcentuales del valor real y un 95% de confianza, si se sabe que es poco probable que la verdadera tasa exceda del 20%?*

BIBLIOGRAFÍA

Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991.

Levy P.S., Lemeshow S. *Sampling of populations: Methods and Applications.* New York: John Wiley & Sons; 1991.

Prueba de hipótesis

Esta opción calcula el tamaño muestral mínimo necesario para someter a una prueba de hipótesis el supuesto de que la proporción de sujetos de una población con cierta característica sea igual a un valor determinado. Para ello, introducir la proporción en caso de hipótesis nula (P_0) y el valor esperado de la proporción de la población (P_a).

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT pasará a otra pantalla en la que por defecto se presentan los tamaños muestrales mínimos necesarios, para los niveles de confianza bilaterales del 99%, 95% y 90% y potencias del 80% y 90%. El cursor se sitúa en la última posición por si se desea un nivel de confianza o potencia diferentes de los anteriores.

Nivel de Confianza bilateral	Potencia de la Prueba	
	1-beta: 0.80	1-beta: 0.90
1-alfa: 0.99	1627	2102
1-alfa: 0.95	1103	1500
1-alfa: 0.90	875	1231

N.de confianza: 0.950

Potencia: 0.900

1500



Ejercicio: 2.2. Según las publicaciones científicas, la tasa de curación de determinado tipo de cáncer en 5 años (es decir, la proporción de pacientes que no presentan cáncer al cabo de 5 años de tratamiento) es del 50%. Un investigador desea someter a prueba la hipótesis de que esa tasa de curación es aplicable en cierto distrito sanitario. ¿Qué tamaño mínimo tendrá que tener la muestra si el investigador tiene interés en rechazar la hipótesis nula sólo si el valor real es inferior al 50% y si desea tener una seguridad del 90% de que detectará una tasa real del 40% al nivel de significación del 5%?

BIBLIOGRAFÍA

Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991.

Situación con una muestra. Media

Cálculo de una media

Cálculo de una media	
Tamaño poblacional (N):	23654
Desv. Estándar de la población (sd):	35.60
Precisión absoluta(d):	3.000
<input type="button" value="Calcular"/> <input type="button" value="Inicializar"/> <input type="button" value="Salir"/> <input type="button" value="F1-Ayuda"/>	

Esta opción calcula el tamaño muestral mínimo necesario para obtener, con una confianza del $100 \times (1 - \alpha)\%$, una media estimada que no difiera del valor verdadero más que la precisión absoluta (d). Para ello, introducir el tamaño de la población de la que se quiere extraer la muestra. Si es desconocida o sobrepasa el valor 999999, dejar el campo con 0. A continuación introducir el valor de la desviación estándar de la población. Por último, el usuario debe definir la diferencia con el valor verdadero (error de muestreo).

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT presentará otra pantalla en la que por defecto se muestran, para los niveles de confianza bilateral del 99%, 95% y 90%, los tamaños muestrales mínimos necesarios. El cursor se sitúa en la última posición por si se desea un nivel de confianza diferente de los anteriores.

Cálculo de una media - Resultados	
Nivel de confianza bilateral	Tamaño muestral mínimo
1-alfa: 0.99	929
1-alfa: 0.95	539
1-alfa: 0.90	379
N.de confianza: 0.950	539

Ejercicio: 2.3. Suponga que en su departamento la media de edad de los pacientes ingresados por herida de arma de fuego es de 45 años y la desviación estándar 34,8. Usted desea realizar un estudio similar en el hospital de referencia asumiendo que la desviación estándar de la media de edad es similar a la departamental. Quiere dar su estimación de la media de edad con una confianza del 95%. Calcule el tamaño muestral mínimo para que la longitud del intervalo de confianza sea 4 años.

BIBLIOGRAFÍA

Levy P.S., Lemeshow S. *Sampling of populations: Methods and Applications*. New York: John Wiley & Sons; 1991.

Situación con dos muestras. Proporción

Cálculo de una diferencia

Esta opción calcula el tamaño muestral mínimo necesario de cada población para poder asumir, con una confianza del $100 \times (1 - \alpha)\%$, que, la diferencia entre las proporciones muestrales no es distinta de la diferencia entre las proporciones verdaderas (poblacionales), en una cantidad mayor que la precisión absoluta establecida (d). Para ello, introducir los valores esperados de las proporciones en ambas poblaciones. Si se desconocen, deberá asignar el valor 0,5, la opción más segura, dado que

Cálculo de la diferencia entre dos proporciones	
Proporción esperada en la Población P1:	0.540
Proporción esperada en la Población P2:	0.513
Precisión absoluta (d):	2.500 %

el caso más desfavorable (mayor tamaño muestral) se da cuando ambas proporciones son iguales a 0,5. Por último, el usuario debe definir la precisión absoluta (error de muestreo) **expresada en porcentaje**, con que desea obtener los resultados.

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT pasará a otra pantalla en la que por defecto se presentan los tamaños muestrales mínimos necesarios para cada población, para niveles de confianza bilaterales del 99%, 95% y 90%. El cursor se sitúa en la última posición por si se desea un nivel de confianza diferente de los anteriores.

Nivel de confianza bilateral	Tamaño muestral mínimo
1-alfa: 0.99	5288
1-alfa: 0.95	3064
1-alfa: 0.90	2158
N.de confianza: 0.950	3064

Salir Pg-Ant Imprimir F1-Ayuda

Ejercicio: 2.4. *Un estudio piloto de 50 trabajadores agrícolas de un proyecto de regadío reveló que el 40% padecía esquistosomiasis activa. Un estudio piloto similar de otros 50 trabajadores distintos reveló que sólo la padecía el 32% de ellos. Si un epidemiólogo desea ampliar el estudio para calcular la diferencia entre el riesgo de esquistosomiasis con una precisión de 5 puntos porcentuales del valor real y con un 95% de confianza, ¿cuántas personas deberá estudiar de cada grupo?*

BIBLIOGRAFÍA

Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991.

Prueba de hipótesis

Esta opción calcula el tamaño muestral mínimo necesario de dos muestras para someter a prueba de hipótesis el supuesto de que la diferencia entre las proporciones de dos poblaciones que poseen cierta característica es igual a 0. Para ello, introducir las proporciones esperadas de cada población (P1 y P2).



Prueba de hipótesis. Proporción. Dos muestras

Proporción esperada en la Población P1:

Proporción esperada en la Población P2:

- Una vez suministrados los datos al programa, pulsar "Calcular". EPIDAT pasará a otra pantalla en la que por defecto se presentan los tamaños muestrales mínimos necesarios, para niveles de confianza bilaterales del 99%, 95% y 90% y potencias del 80% y 90%. El cursor se sitúa en las últimas posiciones por si se desea un nivel de confianza o potencia diferente de las anteriores.

Prueba de hipótesis. Proporción. Dos muestras - Resultados

Tamaño Muestral Mínimo

Nivel de Confianza bilateral	Potencia de la Prueba	
	1-beta: 0.80	1-beta: 0.90
1-alfa: 0.99	<input type="text" value="2014"/>	<input type="text" value="2566"/>
1-alfa: 0.95	<input type="text" value="1354"/>	<input type="text" value="1813"/>
1-alfa: 0.90	<input type="text" value="1067"/>	<input type="text" value="1477"/>

N. de confianza:

Potencia:

Ejercicio: 2.5. Se estima que la proporción de casos con complicaciones a raíz de cierta intervención quirúrgica es del 5% mientras que la proporción de complicaciones a raíz de otra intervención distinta es del 15%. ¿Cuál tendrá que ser el tamaño de la muestra de cada grupo de pacientes si un investigador desea comprobar, con una potencia del 90%, si el segundo procedimiento tiene una tasa de complicaciones significativamente mayor que el primero, a un nivel de significación del 5%?

BIBLIOGRAFÍA

Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991.

Lemeshow S., Hosmer D.W., Klar J., Lwanga, S.K. *Adequacy of Sample Size in Health Studies.* Chichester: John Wiley & Sons; 1990.

Situación con dos muestras. Media

Muestras independientes. Cálculo de una diferencia

Esta opción calcula el tamaño muestral mínimo necesario para cada población para poder suponer con una confianza del $100 \times (1 - \alpha) \%$ que la diferencia de las medias muestrales que obtendremos no excede de la diferencia de medias verdaderas (poblacionales) una cantidad mayor que la precisión absoluta establecida (d). Para ello, introducir los valores esperados de las desviaciones estándar de ambas poblaciones y la precisión absoluta.

Cálculo de la diferencia de dos medias	
Des. Estándar esperada en Pob. A (sd1):	12.45
Des. Estándar esperada en Pob. B (sd2):	19.35
Precisión absoluta(d):	3.000
<input type="button" value="Calcular"/> <input type="button" value="Inicializar"/> <input type="button" value="Salir"/> <input type="button" value="F1-Ayuda"/>	

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT pasará a otra pantalla en la que por defecto se presentarán los tamaños muestrales mínimos necesarios para cada población, para los niveles de confianza bilaterales del 99%, 95% y 90%. El cursor se sitúa en la última posición por si se desea un nivel de confianza diferente de los anteriores.

Cálculo de diferencia de dos medias - Resultados	
Nivel de confianza bilateral	Tamaño muestral mínimo
1-alfa: 0.99	391
1-alfa: 0.95	227
1-alfa: 0.90	160
N.de confianza: 0.950	227
<input type="button" value="Salir"/> <input type="button" value="Pg-Ant"/> <input type="button" value="Imprimir"/> <input type="button" value="F1-Ayuda"/>	

Ejercicio: 2.6. Suponga que en dos provincias de un país la media de consumo de cigarrillos/día se estima en 12,5 y 15, con desviaciones estándar de 11,7 y 12,2 respectivamente. Si usted desea realizar un estudio para comparar el consumo en ambas provincias, ¿qué tamaño de muestra mínimo necesitaría si quiere dar sus resultados con una confianza del 95% y una precisión absoluta de 2?

BIBLIOGRAFÍA

Sokal R.F., Rohlf F.J. *Biometry*. New York: W. H. Freeman; 1981.

Muestras dependientes. Prueba de hipótesis

Esta opción calcula el tamaño muestral mínimo necesario (n pares de sujetos) para someter a una prueba de hipótesis el supuesto de que la diferencia entre las medias (dif) de las dos poblaciones es igual a un valor determinado. Para ello, introducir, además de la diferencia entre medias, la desviación estándar de la diferencia esperada o las desviaciones estándar de las dos poblaciones y el coeficiente de correlación entre ambas.

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT pasará a otra pantalla en la que por defecto se presentan los tamaños muestrales mínimos necesarios, para niveles de confianza

Nivel de Confianza bilateral	Potencia de la Prueba	
	1-beta: 0.80	1-beta: 0.90
1-alfa: 0.99	399	508
1-alfa: 0.95	268	359
1-alfa: 0.90	211	293

N. de confianza: 0.950

Potencia: 0.900 → 359



bilaterales del 99%, 95% y 90% y potencias del 80% y 90%. El cursor se sitúa en las últimas posiciones por si se desea un nivel de confianza o potencia diferente de las anteriores.

Ejercicio: 2.7. Para evaluar una intervención con flúor en un colegio se plantea estudiar un grupo de niños antes y después del uso de colutorios. ¿Cuántos niños deberá estudiar para comprobar que la diferencia entre los índices CAO (CPO) posterior y anterior a la intervención es igual a 1,5? Suponga que espera una desviación estándar de la diferencia de índices CAO (CPO) de 3,4 y desea dar sus resultados con una confianza del 95% y una potencia del 90%.

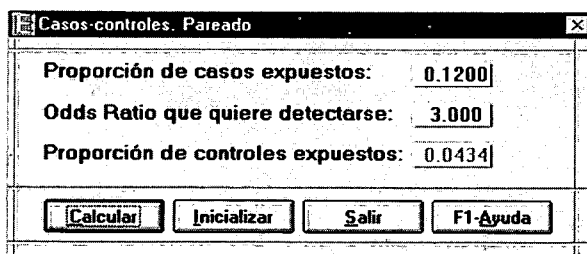
BIBLIOGRAFÍA

Sokal R.F., Rohlf F.J. *Biometry*. New York: W.H. Freeman; 1981.

Estudios casos–controles

Pareado

Esta opción calcula el número de pares mínimo necesario para detectar una determinada odds ratio (razón de ventaja, razón de productos cruzados, razón de oportunidad). Para ello, debemos introducir dos de los siguientes tres parámetros: proporción estimada de casos expuestos, odds ratio que queremos detectar y la proporción estimada de controles expuestos.



- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT pasará a otra pantalla en la que por defecto se presentará el mínimo número de pares necesarios para los niveles de confianza bilaterales del 99%, 95% y 90% y las potencias del 80% y 90%. El cursor se sitúa en la última posición por si se desea un nivel de confianza o potencia diferente de los anteriores.

Casos-contrroles. Pareado - Resultados

Tamaño Muestral Mínimo

Nivel de Confianza bilateral	Potencia de la Prueba	
	1-beta: 0.80	1-beta: 0.90
1-alfa: 0.99	288	360
1-alfa: 0.95	190	249
1-alfa: 0.90	151	203

N.de confianza:

Potencia:

249

Ejercicio: 2.8. Usted va a diseñar un estudio de casos y controles pareado sobre tabaquismo y cáncer de laringe utilizando controles poblacionales. Suponiendo que la prevalencia del hábito tabáquico en la población es del 45%, ¿cuántos pares de casos y controles necesitará para detectar una OR de 3? Desea tener un nivel de confianza del 95% y un poder estadístico del 80%.

BIBLIOGRAFÍA

Schlesselman J.J. *Case-Control Studies*. New York: Oxford University Press; 1982.
 Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

No pareado

Esta opción calcula el tamaño muestral mínimo necesario en cada grupo, para detectar una determinada odds ratio. Para ello, debemos introdu-

Casos-contrroles. No pareado

Proporción de expuestos en casos (P1):

Proporción de expuestos en controles (P2):

ODDS-RATIO a detectar:

Número de controles por caso:

dos de los siguientes tres parámetros: odds ratio que se pretende detectar, proporción estimada de casos expuestos y la proporción estimada de controles expuestos.

- Una vez suministrados los datos al programa, pulsar “Calcular”. Epidat pasará a otra pantalla en la que por defecto se presentarán los tamaños muestrales mínimos necesarios para los niveles de confianza bilaterales del 99%, 95% y 90% y las potencias del 80% y 90%. El cursor se sitúa en la última posición por si se desea un nivel de confianza o potencia diferente de los anteriores.

Tamaño Muestral Mínimo				
Nivel de Confianza bilateral	Potencia de la Prueba			
	1-beta: 0.80		1-beta: 0.90	
	Casos	Controles	Casos	Controles
1-alfa: 0.99	69	138	90	180
1-alfa: 0.95	47	94	65	130
1-alfa: 0.90	37	74	53	106

N. de confianza:	<input type="text" value="0.950"/>	}	<input type="text" value="65"/>
Potencia:	<input type="text" value="0.900"/>		<input type="text" value="130"/>

Salir	Pg-Ant	Imprimir	F1-Ayuda
-------	--------	----------	----------

Ejercicio: 2.9. Queremos realizar un estudio de casos y controles para analizar la influencia del consumo de alcohol en la aparición de cáncer de esófago. ¿Cuántos casos debemos estudiar si se quiere detectar una OR de 2,5 y sabemos que la exposición entre los controles es del 60%? Calcule el número de casos necesarios para obtener un nivel de confianza del 95% y un poder del 80%.

BIBLIOGRAFÍA

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Lwanga S.K, Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico*. Ginebra: O.M.S.; 1991.

Estudios de cohortes

Estimación del riesgo relativo

Esta opción calcula el tamaño muestral mínimo necesario para detectar un determinado riesgo relativo. Para ello, debemos introducir dos de los siguientes tres parámetros: el riesgo relativo previsto, la proporción de expuestos con la enfermedad y la proporción de no expuestos con la enfermedad. También debe introducirse, en porcentaje, la precisión relativa (ϵ).

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT presentará otra pantalla en la que por defecto se muestra el tamaño muestral mínimo para cada cohorte para los niveles de confianza bilaterales del 99%, 95% y 90%. El cursor se sitúa en la última posición por si se desea un nivel de confianza diferente de los anteriores.

Nivel de confianza bilateral	Tamaño muestral mínimo
1-alfa: 0.99	7888
1-alfa: 0.95	4571
1-alfa: 0.90	3219
N.de confianza: 0.950	4571

Ejercicio: 2.10. Un epidemiólogo proyecta un estudio sobre la posibilidad de que cierta enfermedad pulmonar esté relacionada con la exposición a un contaminante atmosférico recién detectado. ¿Qué tamaño tendrá que tener la muestra de cada grupo (expuesto y no expuesto) si el epidemiólogo desea calcular el riesgo relativo con una precisión del 10% del valor real (probablemente de 2) con un 95% de confianza? La enfermedad se manifiesta en el 20% de las personas no expuestas al contaminante atmosférico.

BIBLIOGRAFÍA

Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991.

Verificación de la calidad de lotes

Aceptación de una prevalencia

(Texto entre comillas tomado de: Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991).

“Esta opción se refiere al modo de determinar el tamaño de la muestra necesaria de una población para que, si cierta característica sólo se encuentra en un número determinado de sujetos de la muestra, pueda admitirse que la prevalencia de esa característica en la población no excede de cierto valor (valor umbral)”.

Aceptación de que una proporción no supera un valor específico

Tamaño de la Población (N): 14526

Proporción esperada en la población (P): 0.240

Máximo nº admisible de casos en la muestra (Valor Umbral): 3

Calcular Inicializar Salir F1-Ayuda

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT pasará a otra pantalla en la que por defecto se presentan los tamaños muestrales mínimos necesarios para los niveles de confianza unilaterales del 99%, 95% y 90%. El cursor se sitúa en la última posición por si se desea un nivel de confianza diferente de los anteriores.

Aceptación de que la proporción en una población no supere u...

Nivel de confianza unilateral	Tamaño muestral
1-alfa: 0.99	39
1-alfa: 0.95	30
1-alfa: 0.90	26

N.de confianza: 0.950

Salir Pg-Ant Imprimir F1-Ayuda



Ejemplo: En una escuela con 2.500 niños, ¿a cuántos habrá que examinar para que, si sólo se detecta parasitemia palúdica en dos, pueda concluirse, con un nivel de confianza del 95%, que la prevalencia del paludismo en la escuela no excede del 10%?

Solución:

- prevalencia prevista en la población:..... 10%
- tamaño de la población:..... 2500
- número máximo de paludismo en la muestra:..... 2
- nivel de confianza:..... 95%

MUESTRA NECESARIA: 61 NIÑOS

Ejercicio: 2.11. Si su municipio tiene 5.840 habitantes, calcule cuántos deberá examinar para que, si sólo encuentra 3 hipertensos, pueda concluir, con un nivel de confianza del 90%, que la prevalencia de hipertensión en su municipio no excede del 15%.

BIBLIOGRAFÍA

Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991.

Rechazo de un lote

(Texto entre comillas tomado de: Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991).

“Esta opción se refiere a los estudios diseñados para probar si un «lote» (una población objeto de muestreo) cumple un requisito específico. La hipótesis nula es aquella en la que la proporción de sujetos de la población que presentan cierta característica es igual a un valor dado; la prueba unilateral permite aceptar que el lote cumple el requisito específico sólo si se puede rechazar la hipótesis nula. Para ello se establece un «valor umbral» de sujetos con la característica (d) como base para la decisión; si el número de sujetos de la muestra que poseen la característica no excede el umbral, se rechaza la hipótesis nula (y se acepta el lote), mientras que si excede del umbral se rechaza el lote”.

Máxima proporción aceptable de población con el problema en estudio		<input type="text" value="0.450"/>
Proporción esperada (Meta) de población con el problema en estudio		<input type="text" value="0.350"/>
<input type="button" value="Calcular"/>		<input type="button" value="Inicializar"/>
<input type="button" value="Salir"/>		<input type="button" value="F1-Ayuda"/>

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT pasará a otra pantalla en la que por defecto se presentan los tamaños muestrales y los valores umbral necesarios, para los niveles de confianza unilaterales del 99%, 95% y 90% y las potencias del 80% y 90%. El cursor se sitúa en las últimas posiciones por si se desean otras opciones diferentes de las anteriores.

Tamaño muestral				
Nivel de Confianza unilateral	Potencia de la Prueba			
	1-beta: 0.80		1-beta: 0.90	
	T. Muestral	V. Umbral	T. Muestral	V. Umbral
1-alfa: 0.99	<input type="text" value="243"/>	<input type="text" value="92"/>	<input type="text" value="313"/>	<input type="text" value="121"/>
1-alfa: 0.95	<input type="text" value="149"/>	<input type="text" value="58"/>	<input type="text" value="205"/>	<input type="text" value="81"/>
1-alfa: 0.90	<input type="text" value="108"/>	<input type="text" value="42"/>	<input type="text" value="156"/>	<input type="text" value="63"/>
N.de confianza:	<input type="text" value="0.950"/>		<input type="text" value="205"/> (Tamaño muestral)	
Potencia:	<input type="text" value="0.900"/>		<input type="text" value="81"/> (Valor umbral)	
<input type="button" value="Salir"/>		<input type="button" value="Pg-Ant"/>		
<input type="button" value="Imprimir"/>		<input type="button" value="F1-Ayuda"/>		

Ejemplo: En una gran ciudad, las autoridades sanitarias desean obtener una cobertura de vacunación en la población infantil del 90%. En vista de los inquietantes brotes de enfermedades infantiles observados en ciertas zonas de la ciudad, un equipo de investigadores del servicio de sanidad proyecta hacer una encuesta que detecte los barrios donde la cobertura es del 50% o menor, para que puedan adoptarse las medidas apropiadas. ¿Cuántos niños habrá que estudiar en cada zona y qué valor umbral habrá de utilizarse para probar la hipótesis de que la proporción de niños no vacunados es del 50%? Los investigadores desean tener una confianza del 90% de que reconocerán las zonas donde se ha alcanzado la cobertura de vacunación preestablecida (es decir, en las que el 10% de los niños no están inmunizados).

Solución:

- valor de la proporción en la población sometida a prueba 50%
- valor previsto de la proporción de la población: 10%
- nivel de significación: 5%
- potencia de la prueba: 90%

Como el error más importante es aceptar que ciertos grupos de niños están suficientemente vacunados cuando en realidad su cobertura es del 50% o menor, $P_o = 0,5$ y $P_a = 0,1$, tenemos:

- TAMAÑO DE LA MUESTRA = 10
- VALOR UMBRAL = 2 (ver nota)

Por tanto, deberá tomarse en cada una de las 10 zonas estudiadas una muestra de 10 niños. Si se descubre que en una muestra hay más de dos niños deficientemente vacunados, se “rechaza” el lote (la población origen de la muestra) y las autoridades sanitarias podrán tomar medidas para mejorar la cobertura en la zona de que se trate. En cambio, si se descubre que sólo están deficientemente vacunados 2 niños (o menos), deberá rechazarse la hipótesis nula y se podrá estimar que el grupo de niños no es una prioridad inmediata en una campaña intensiva de vacunación”.

NOTA: EPIDAT redondea el valor umbral hacia “arriba”, a diferencia de lo propuesto en las tablas de Lwanga & Lemeshow. Así pues, en el ejemplo anterior, el valor umbral obtenido con EPIDAT sería 3 y su interpretación la siguiente: “Si se descubre que en una muestra hay TRES O MÁS niños deficientemente vacunados...”

Ejercicio: 2.12. *En su departamento, la cobertura prevista para un programa de tamizaje de cáncer de mama es del 70%. Detectadas ciertas deficiencias en el programa usted desea encuestar a mujeres del grupo de edad seleccionado por el programa para que le permita detectar barrios y zonas donde la cobertura es del 45% o menor. ¿Cuántas mujeres habrá que encuestar en cada barrio o zona y qué valor umbral habrá de utilizarse para probar la hipótesis de que la proporción de mujeres perdidas por el programa es del 55%? Usted desea tener una confianza del 90% de que reconocerá zonas donde se ha alcanzado la cobertura de tamizaje preestablecida.*

BIBLIOGRAFÍA

Lwanga S.K., Lemeshow S. *Determinación del tamaño muestral en los estudios sanitarios. Manual práctico.* Ginebra: O.M.S.; 1991.

Generación de números aleatorios

Esta opción genera una lista de números pseudo aleatorios entre un mínimo y un máximo. Para ello es preciso introducir tres valores: mínimo, máximo y cantidad de números a extraer entre ambos.

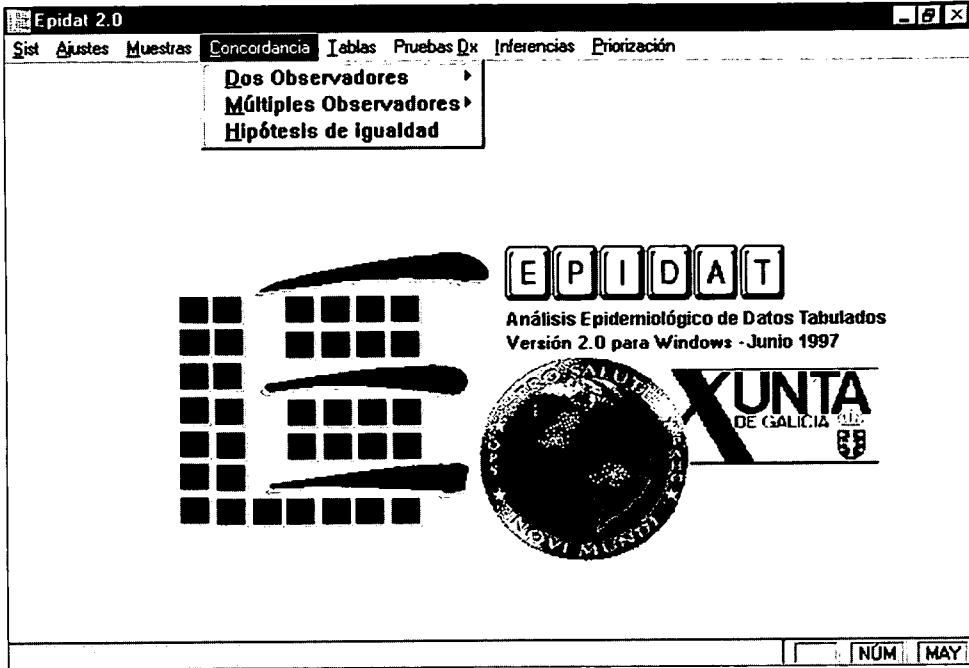
- Pulsando “Calcular” EPIDAT presenta en otra pantalla el listado de números ordenados de menor a mayor.

A	B	C	D	E	F	G	H	I	J
8559	8675	8892	9004	9144	9168	9475	9591	9709	9727
9844	9994	10008	10086	10247	10311	10637	10689	10710	10815
10820	10874	10878	11002	11008	11012	11035	11040	11041	11068
11086	11091	11128	11163	11217	11288	11332	11367	11439	11440
11449	11508	11599	11660	11697	11850	11857	11922	12052	12094
12145	12339	12583	12610	12849	12880	12985	12987	13078	13247
13270	13307	13340	13475	13564	13573	13783	13948	13966	13985
14050	14085	14304	14359	14372	14535	0	0	0	0

CONCORDANCIA

--	--	--	--	--	--	--

Análisis de concordancia entre dos o más observadores y dos o más categorías de observación e hipótesis de igualdad de kappas.



Dos observadores

Dos categorías

Esta opción calcula los índices de concordancia bruta, J de Youden y kappa de Cohen para el caso particular de dos observadores y dos categorías de clasificación de las observaciones. Para la J de Youden y kappa de Cohen se calculan los intervalos de confianza. Introducir el número de acuerdos y desacuerdos entre los observadores A y B y pulsar "Calcular".

		A		Marginales
		+	-	
B	+	<input type="text" value="36"/>	<input type="text" value="12"/>	<input type="text" value="48"/>
	-	<input type="text" value="14"/>	<input type="text" value="142"/>	<input type="text" value="156"/>
		<input type="text" value="50"/>	<input type="text" value="154"/>	<input type="text" value="204"/>

- En la pantalla de resultados se presentan los valores de la concordancia bruta, J de Youden y kappa. Para esta última se obtienen, además, el error estándar (necesario para contrastar la igualdad de kappas) y el valor p para contrastar la hipótesis nula de $k=0$.

Concordancia. 2 observadores - 2 categorías			
Concordancia bruta	0.873		
J de Youden	0.642		
I.C.	0.510	0.773	
Kappa	0.651		
I.C.	0.527	0.773	
Err.Estánd. kappa	0.06272		
Ho: k=0 . P-valor	0.00000		
Nivel de confianza:	0.950		
<input type="button" value="Pág.Ant."/> <input type="button" value="Salir"/> <input type="button" value="Imprimir"/> <input type="button" value="F1-Ayuda"/>			

Ejercicio: 3.1. Suponga que radiografías de tórax de 170 sujetos que están controlados en una unidad de neumología se remiten a dos radiólogos del programa de tuberculosis con el objetivo de estudiar la concordancia entre ellos. Los radiólogos A y B tienen que clasificar las radiografías como "positivas" (sospechosas de lesión tuberculosa) o "negativas" (no sospechosas de lesión tuberculosa). Los resultados se muestran a continuación:

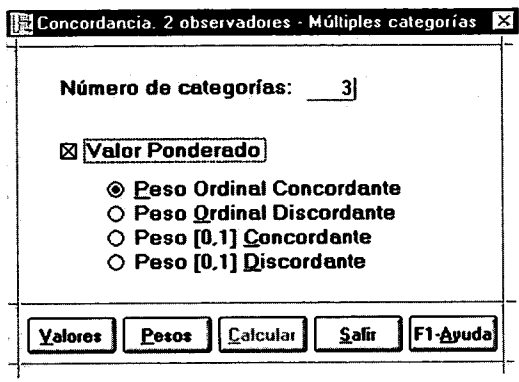
	Radiólogo A	
	+	-
Radiólogo B		
+	58	39
-	12	61

Calcule la concordancia bruta entre los radiólogos, la J de Youden y kappa, con sus intervalos de confianza. Coméntelos.

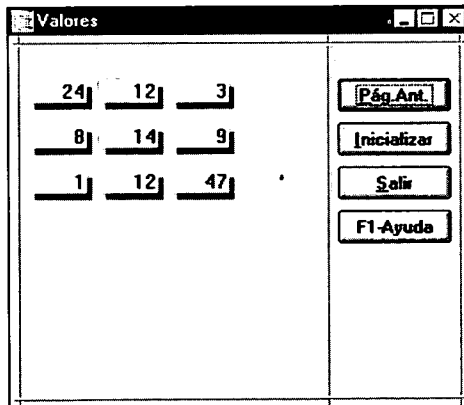
BIBLIOGRAFÍA

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

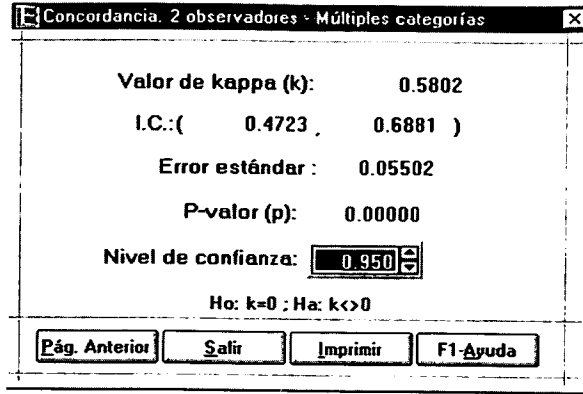
Múltiples categorías



Esta opción calcula el índice de concordancia kappa de Cohen para el caso particular de dos observadores y más de dos categorías de clasificación de las observaciones. En la pantalla inicial, una vez definido el tamaño de la matriz, se puede elegir k ponderada o no. Si se desea ponderar, se pueden seleccionar las opciones de ordinal de concordancia y discordancia al igual que los pesos entre 0 y 1. En las dos primeras opciones los pesos son fijos dependiendo del tamaño de la matriz y en las dos últimas la tabla de pesos se puede modificar por el usuario, siempre que se mantenga la simetría de la matriz.



- Una vez cubierta la tabla de valores EPIDAT calcula el valor de k , sus intervalos de confianza, el error estándar (necesario para contrastar la igualdad de kappas) y el valor p para contrastar la hipótesis nula de $k=0$.



Ejercicio: 3.2. *Se desea analizar la concordancia entre dos pruebas diagnósticas de diabetes, el Diabur-Test 5000® y el Clinitest® en 1.677 muestras de orina. Los valores obtenidos se muestran en la tabla siguiente:*

		Diabur-Test 5000®					
		negativo	trazas	1	2	3	5
Clinitest®	negativo	452	5				
	trazas	133	270	28	1	2	
	1	4	36	107	5	2	2
	2		5	53	76	28	4
	3			12	28	81	35
5				2	11	44	251

Modificado de: Banauch D., Koller P.U., Bablok W. *Evaluation of Diabur-Test 5000: a cooperative study carried out at 12 diabetes centers.* Diabetes Care: 1983; May-Jun. 6(3) 213-218.

Calcule el valor de kappa sin ponderar y ponderado. Para la ponderación utilice la opción [0,1] concordante considerando que los resultados que "caen" en la misma celda de la tabla de correlación, a efectos prácticos, tienen completa concordancia y, por tanto, les daremos un peso 1. Un desacuerdo en solamente una celda (p. e. trazas vs 1) lo pesaremos con 0,8, en dos celdas con 0,6 y así sucesivamente, el desacuerdo en 5 celdas tendría un peso 0.

BIBLIOGRAFÍA

Fleiss J.L. *Statistical methods for rates and proportions.* New York: John Wiley and Sons; 1981.

Kramer M.S., Feinstein A.R. *Clinical Biostatistics. The Biostatistics of Concordance.* Clin. Pharmacol. Ther. 1981; 29: 111-123.



Múltiples observadores

Dos categorías

Concordancia. Múltiples observadores - 2 categorías

Sujeto 8 Observadores: Clasificación +:

	Observadores	Clasificación +	
Sujeto 1	5	3	+
Sujeto 2	4	4	
Sujeto 3	3	2	
Sujeto 4	4	2	
Sujeto 5	4	4	
Sujeto 6	3	3	
Sujeto 7	5	3	-

Esta opción calcula el índice de concordancia kappa de Cohen para el caso particular de múltiples observadores y dos categorías de clasificación de las observaciones. En la pantalla inicial se introduce el número de observadores para el sujeto 1, a continuación el número de acuerdos positivos. Una vez aceptado se repite la operación con el sujeto 2 y así sucesivamente. El número de observadores no tiene que ser necesariamente el mismo para todos los sujetos.

Concordancia. Múltiples observadores - 2 categorías

Valor de kappa (k): -0.0328

Ho: $k=0$: Ha: $k > 0$

P-valor (p): 0.83129

- En los resultados se presentan la kappa estimada y el error estándar para probar la hipótesis de $k=0$. EPIDAT no calcula los intervalos de confianza de kappa debido a la complejidad del proceso de cálculo.

Ejercicio: 3.3. Suponga ahora que una selección de 29 radiografías correspondientes a otros tantos pacientes del servicio de neumología del problema 3.1., se entregan a un grupo de radiólogos para que, de forma independiente, estudien su concordancia. Por diversas razones no todos los radiólogos del equipo pudieron emitir juicio sobre todas las radiografías. Los resultados se presentan en la siguiente tabla:



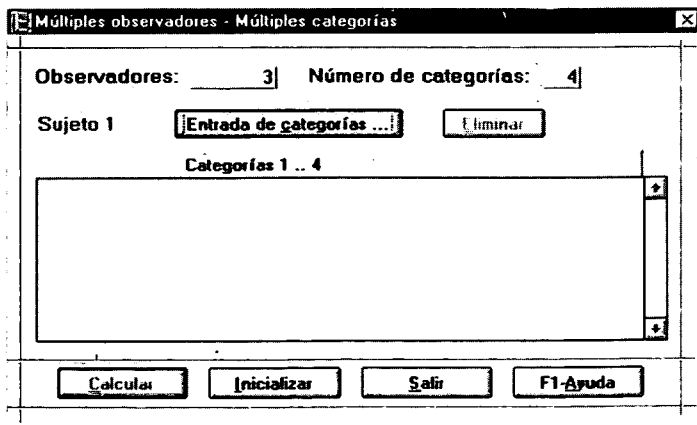
Paciente	Número de radiólogos	Número de positivos	Paciente	Número de radiólogos	Número de positivos
1	4	3	16	5	5
2	3	2	17	5	0
3	4	2	18	3	2
4	5	4	19	4	3
5	3	3	20	4	2
6	4	2	21	3	1
7	4	3	22	2	0
8	5	3	23	5	0
9	5	4	24	4	4
10	5	5	25	4	3
11	3	0	26	3	2
12	2	0	27	3	2
13	4	2	28	2	1
14	4	0	29	5	4
15	3	2	Total	110	64

Calcule el valor de k y el p -valor correspondiente al contraste.

BIBLIOGRAFÍA

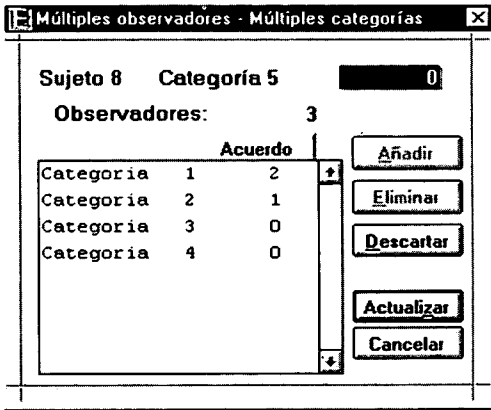
Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Múltiples categorías

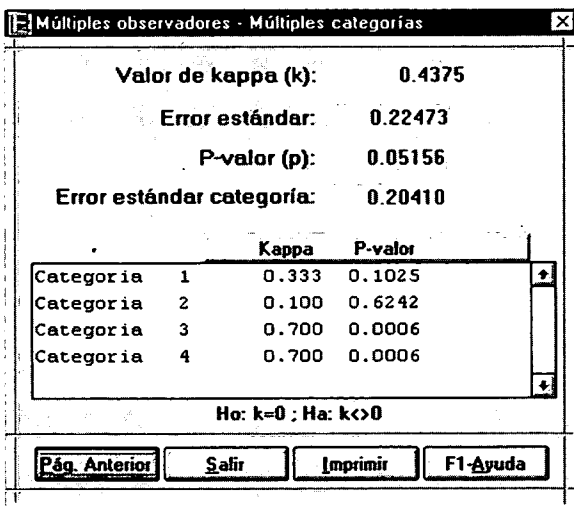


Esta opción calcula el índice de concordancia kappa de Cohen para el caso particular de múltiples observadores y más de dos categorías de cla-

sificación de las observaciones. En la pantalla inicial se fijan el número de observadores y categorías de clasificación. En ambos casos deben ser más de dos. Seleccionando el botón “Entrada de categorías” se accede a otra pantalla donde se introducen los acuerdos de cada categoría de clasificación para el sujeto 1.



- Al aceptar los valores introducidos EPIDAT regresa a la anterior pantalla desde la que se pueden añadir, siguiendo el mismo proceso, los datos correspondientes al resto de los sujetos.
- En los resultados se presentan la kappa estimada y el error estándar para probar la hipótesis de $k=0$. EPIDAT no calcula los intervalos de confianza de kappa debido a la complejidad del proceso de cálculo.



Ejercicio: 3.4. A usted le interesa afinar más su análisis de concordancia entre los radiólogos del ejemplo anterior y selecciona un subgrupo de 15 pacientes cuyas placas entrega a los 5 radiólogos del equipo para que las clasifiquen de forma independiente como "muy sospechosas de lesión tuberculosa" (categoría 1), "sospecha ligera de lesión tuberculosa" (categoría 2) y "sin sospecha alguna de lesión tuberculosa" (categoría 3). Los resultados se muestran en la siguiente tabla:

Número de radiólogos que clasificaron en cada categoría			
Paciente	categoría 1	categoría 2	categoría 3
1	2	2	1
2	5	0	0
3	0	1	4
4	1	1	3
5	4	1	0
6	1	2	2
7	0	0	5
8	0	1	4
9	3	1	1
10	4	0	1
11	1	0	4
12	0	1	4
13	1	3	1
14	1	4	0
15	2	3	0
Total	25	20	30

Calcule ahora el valor de kappa global y por cada categoría de clasificación.

BIBLIOGRAFÍA

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Landis J.R., Koch G.G. A one-way components of variance model for categorical data. *Biometrics*, 33, 671-679.

Hipótesis de igualdad de kappas

Esta opción está destinada a contrastar la homogeneidad entre diversas kappas. Se calcula la k global y su intervalo de confianza. Se permite variar el nivel de confianza que por defecto es del 95%.

Hipótesis de igualdad de kappa

Tabla 4 Valor kappa: 0.000 Error estándar: 0.000000

		Kappa	Error estándar	
Tabla	1	0.680	0.01500	+
Tabla	2	0.740	0.00340	
Tabla	3	0.590	0.40000	-

Añadir
Eliminar
Descartar

Calcular Inicializer Salir F1-Ayuda

Hipótesis de igualdad de kappa.

Valor de kappa global: 0.727

I.C. (0.625 , 0.830)

Valor de la Chi-Cuadrado: 0.2435

P-valor (p): 0.88534

H₀: No hay diferencias significativas entre los valores de Kappa
H_a: Las diferencias entre los valores de Kappa son significativas

Pg-Ant Salir Imprimir F1-Ayuda

Ejercicio: 3.5. De usted dependen dos unidades de diagnóstico de tuberculosis (A y B). Cada seis meses remite a cada una de ellas un porcentaje de las tinciones realizadas en el semestre anterior con el objeto de comprobar la concordancia entre ellas. A las dos unidades le envía las mismas muestras, que tienen que clasificar como "positivas" o "negativas". Los valores obtenidos en los dos últimos años se reflejan a continuación:

Semestre 1:

		A	
		+	-
B	+	350	120
	-	70	550

Semestre 2:

		A	
		+	-
B	+	280	80
	-	60	550

Semestre 3:

		A	
		+	-
B	+	320	30
	-	120	29

Semestre 4:

		A	
		+	-
B	+	890	210
	-	290	700

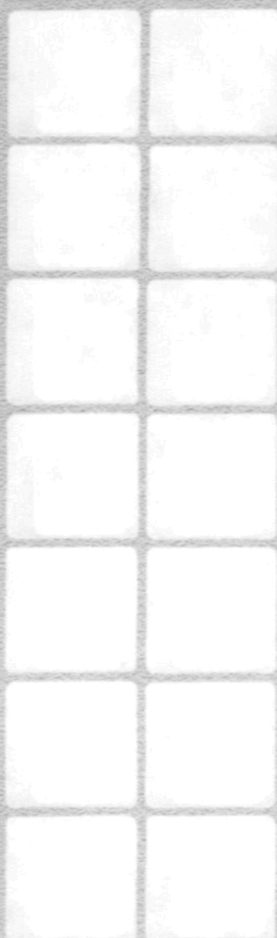
Calcule el k global. ¿Son diferentes los kappas obtenidos en cada semestre? Sustituya los valores del 4º semestre por:

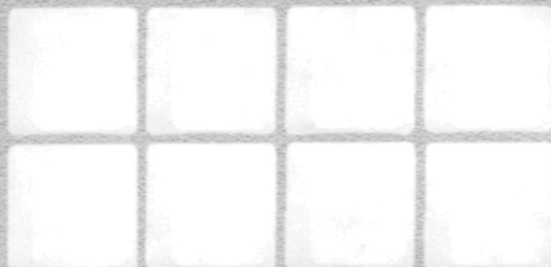
Semestre 4:

		A	
		+	-
B	+	350	310
	-	195	240

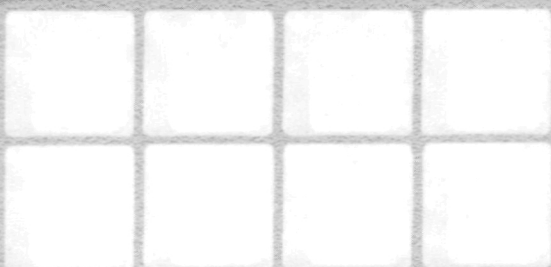
¿Qué ocurre ahora?







TABLAS

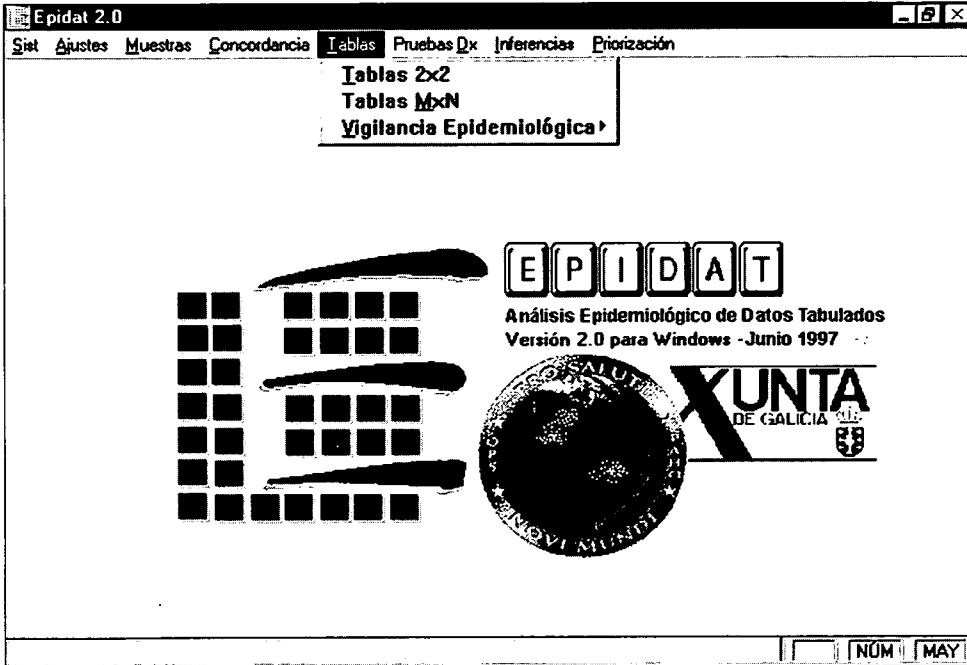




--	--	--	--	--	--	--

Análisis de tablas

Este módulo permite realizar las pruebas de homogeneidad y fuerza de asociación más comunes en los estudios epidemiológicos que analizan tablas 2x2 simples o estratificadas. Calcula los índices de homogeneidad, basados en el estadístico chi-cuadrado, en tablas MxN e incorpora algunos tópicos de interés en vigilancia epidemiológica.



Tablas 2x2

Tablas simples. Estudios transversales/cohortes. —Análisis de datos transversales y de seguimiento sin estratificar.

Análisis de tablas 2x2		
Título del Estudio: <input type="text" value="Cáncer de pulmón y tabaquismo"/>		
Tipo de Tabla	Tipo de Estudio	Forma
<input checked="" type="radio"/> Simple <input type="radio"/> Estratificada	<input checked="" type="radio"/> Transversal / Cohortes <input type="radio"/> Caso-control	<input type="radio"/> No Pareado <input type="radio"/> Pareado Controles: <input type="text" value="0"/>
<input type="button" value="Valores"/>	<input type="button" value="Iniciatizar"/>	<input type="button" value="Salir"/> <input type="button" value="F1-Ayuda"/>

Análisis de tablas 2x2				
Simple - Transversal / Cohortes				
		Enfermos	Sanos	
Fr	+	65	12	77
	-	15	145	160
		80	157	237

Calcular

Reg. Ant.

Inicializar

Salir

F1-Ayuda

Esta opción analiza los datos tabulados a partir de un estudio transversal (descriptivo) o de seguimiento (cohortes). EPIDAT calcula el chi-cuadrado con y sin la corrección de Yates, así como el chi-cuadrado de Mantel-Haenszel con el valor p correspondiente a una cola de cada uno de ellos. Cuando $n \leq 20$ ó $20 < n < 40$ y alguna frecuencia esperada es menor de 5, se calculan los valores p de la prueba de Fisher, tanto para una cola como para dos. Asimismo, se calcula la odds ratio con sus intervalos de confianza, utilizando si es posible el método de Cornfield y en su defecto, el de Wolf. Se permite variar el nivel de confianza que por defecto es del 95%. Pulsando el botón “Intervalo de Confianza (Bootstrap)” aparece un aviso en el que se indica que el tiempo de cálculo puede prolongarse durante unos segundos, dependiendo del número de iteraciones seleccionadas, los valores de las celdas y las características de la computadora. EPIDAT permite realizar entre 400 y 2.000 iteraciones.

Intervalo de confianza (Bootstrap)

!!! AVISO !!!

Este procedimiento tiene un tiempo de ejecución dependiente del número de interacciones seleccionadas, de los valores introducidos y de las características de su computadora. Los intervalos de confianza Bootstrap se calculan con un nivel de confianza 95 %.

Número de iteraciones: 2000

Aceptar
Cancelar

- La odds ratio se puede utilizar en diseños transversales situando en las filas la variable “antecedente” y en las columnas el evento que hipotéticamente le sucede, midiendo el grado de asociación entre la primera y el segundo.

Análisis de tablas 2x2 - Resultados																					
<table border="1"> <tr> <th>Enfermos</th> <th colspan="2">Sanos</th> </tr> <tr> <td>Fr. +</td> <td>65</td> <td>12</td> <td>77</td> </tr> <tr> <td>Fr. -</td> <td>15</td> <td>145</td> <td>160</td> </tr> <tr> <td></td> <td>80</td> <td>157</td> <td>237</td> </tr> </table>		Enfermos	Sanos		Fr. +	65	12	77	Fr. -	15	145	160		80	157	237	<p>Cáncer de pulmón y tabaquismo</p> <p>Chi-cuadrado P-valor (p)</p> <p>No corregido: 130.90 0.0000</p> <p>Corrección de Yates: 127.57 0.0000</p> <p>Mantel-Haenszel: 130.35 0.0000</p> <p>Prob. exacta de Fisher:</p> <table border="1"> <tr> <td>Unilateral</td> <td>0.0000</td> </tr> <tr> <td>Bilateral</td> <td>0.0000</td> </tr> </table>	Unilateral	0.0000	Bilateral	0.0000
Enfermos	Sanos																				
Fr. +	65	12	77																		
Fr. -	15	145	160																		
	80	157	237																		
Unilateral	0.0000																				
Bilateral	0.0000																				
Simple - Transversal / Cohortes																					
Odds ratio (OR): 52.36	I.C. li	I.C. ls																			
Intervalo de confianza (Cornfield)	21.71	129.87																			
Intervalo de confianza (Bootstrap):	22.53	107.35																			
Nivel de confianza: 0.950																					
<input type="button" value="Más ..."/> <input type="button" value="Página Anterior"/> <input type="button" value="Salir"/> <input type="button" value="Imprimir"/> <input type="button" value="F1-Ayuda"/>																					

- Al pulsar “Más” en la pantalla de resultados, se presentan los cálculos del riesgo relativo con los intervalos de confianza de Katz y del riesgo atribuible poblacional con los intervalos de confianza de Walter.

Los riesgos deben utilizarse en estudios de seguimiento.

Análisis de tablas 2x2 - Resultados		
Cáncer de pulmón y tabaquismo		
Simple - Transversal / Cohortes		
Riesgo Relativo (RR): 9.00	I.C. li	I.C. ls
Intervalo de confianza (Katz):	5.50	14.71
Intervalo de confianza (Bootstrap):	0.00	0.00
Riesgo atribuible poblacional a partir de un estudio de cohortes.		
Proporción de expuestos:	0.32	
Riesgo atribuible poblacional:	0.72	
Intervalo de confianza (Walter):	0.60	0.83
Nivel de confianza: 0.950		
<input type="button" value="Página Anterior"/> <input type="button" value="Salir"/> <input type="button" value="Imprimir"/> <input type="button" value="F1-Ayuda"/>		

Ejercicio: 4.1. En un estudio de seguimiento de malaria decide estudiar el factor de riesgo “trabajador agrícola (TA)” en relación al resto de la población, encontrando que, a lo largo del año, 202 TA enfermaron de malaria y 2.014 no lo hicieron. En el mismo periodo 81 individuos de otras profesiones padecieron la misma enfermedad y 2.811 no. Calcule las tasas de incidencia en expuestos y en no expuestos al trabajo agrícola; calcule el riesgo relativo del factor de riesgo TA, con sus intervalos de confianza. ¿Qué porcentaje de riesgo poblacional se puede atribuir a este factor?

BIBLIOGRAFÍA

Kahn H.A., Sempos C.T. *Statistical Methods in Epidemiology*. New York: Oxford University Press; 1989.

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Nieto F.J., Peruga A. Riesgo atribuible: sus formas, usos e interpretación. *Gaceta Sanitaria* 1990; 18, vol 4: 112-117.

Efron B., Tibshirani R.J. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.

Tablas simples. Estudios caso-control no pareado. —Análisis de tablas, caso-control no pareados y no estratificados.

Análisis de tablas 2x2

Título del Estudio: Cáncer de pulmón y tabaquismo

Tipo de Tabla	Tipo de Estudio	Forma
<input checked="" type="radio"/> Simple <input type="radio"/> Estratificada	<input type="radio"/> Transversal / Cohortes <input checked="" type="radio"/> <u>Caso-control</u>	<input checked="" type="radio"/> No Pareado <input type="radio"/> Pareado Controles: <input type="text" value="0"/>

Análisis de tablas 2x2

Simple - Caso Control - No pareado

		Enfermos	Sanos	
Fr	+	57	14	71
	-	6	45	51
		63	59	122

Esta opción analiza tablas provenientes de estudios caso-control no pareados y sin estratificar. EPIDAT calcula la prueba chi-cuadrado con o sin la corrección de Yates, así como el chi-cuadrado de Mantel-Haenszel con el valor p correspondiente a una cola de cada uno de ellos. Cuando $n \leq 20$ ó $20 < n < 40$ y alguna frecuencia esperada es menor de 5, se calculan los valores de p de la prueba de Fisher, tanto para una cola como para dos. Asimismo se calcula la odds ratio con los intervalos de confianza, utilizando si es posible el método de Cornfield y en su defecto, el de Wolf. Pulsando el botón "Intervalo de Confianza (Bootstrap)" aparece un aviso en el que se indica que el tiempo de cálculo puede alargarse unos segundos, dependiendo del número de iteraciones, los valores de las celdas y las características de la computadora. EPIDAT permite realizar entre 400 y 2.000 iteraciones.

También se estima el riesgo atribuible poblacional con los intervalos de confianza de Walter.

Excepto para el bootstrap se permite variar el nivel de confianza que por defecto es del 95%.

Intervalo de confianza (Bootstrap)

!!! AVISO !!!
 Este procedimiento tiene un tiempo de ejecución dependiente del número de interacciones seleccionadas, de los valores introducidos y de las características de su computadora. Los intervalos de confianza Bootstrap se calculan con un nivel de confianza 95 %.

Número de iteraciones:

Análisis de tablas 2x2 - Resultados

	Enfermos	Sanos		Cáncer de pulmón y tabaquismo	
Ft. +	57	14	71	Chi-cuadrado	P-valor 95%
Ft. -	6	45	51	No corregido:	55.79 0.0000
	63	59	122	Corrección de Yates:	53.08 0.0000
				Mantel-Haenszel:	55.33 0.0000
				Prob. exacta de Fisher:	0.0000 Unilateral
					0.0000 Bilateral

Simple - Caso Control - No pareado

Odds Ratio (OR): I.C. li I.C. ls

Intervalo de confianza (Cornfield)

Intervalo de confianza (Bootstrap):

Nivel de confianza:

Aproximación al riesgo atribuible poblacional a partir de un estudio caso-control.

Riesgo atribuible poblacional:

Intervalo de confianza (Walter):

Ejercicio: 4.2. En un estudio de casos y controles en el que se estudió la diabetes como factor de riesgo para la enfermedad carotídea obstructiva, realizado en Medellín (Colombia), se obtuvieron los siguientes resultados:

		enfermedad carotídea		
		si	no	total
diabetes	si	81	26	107
	no	419	474	893
Total		500	500	1000

Fuente: Cadavid G., Londoño J.L., Zapata A.M. *Algunos factores de riesgo para la enfermedad carotídea obstructiva: estudio de casos y controles en ejecución.* Centro de investigaciones médicas de Antioquia, CIMA, 1993. Adaptado de Londoño Fernández J.L. *Metodología de la investigación epidemiológica.* Editorial Universidad de Antioquia, Medellín, 1995.

Calcule los chi-cuadrados crudo, corregido, sus valores p y la odds ratio con sus intervalos de confianza. Estime el riesgo atribuible poblacional con sus intervalos de confianza. Comente los resultados.

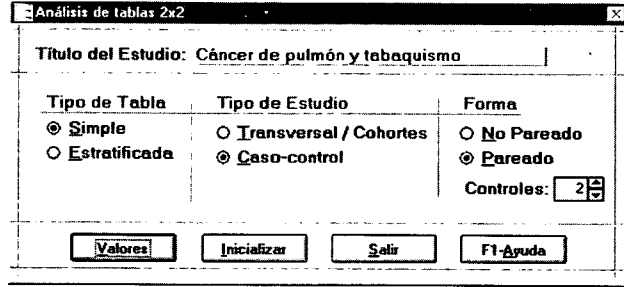
BIBLIOGRAFÍA

Kahn H.A., Sempos C.T. *Statistical Methods in Epidemiology.* New York: Oxford University Press; 1989.

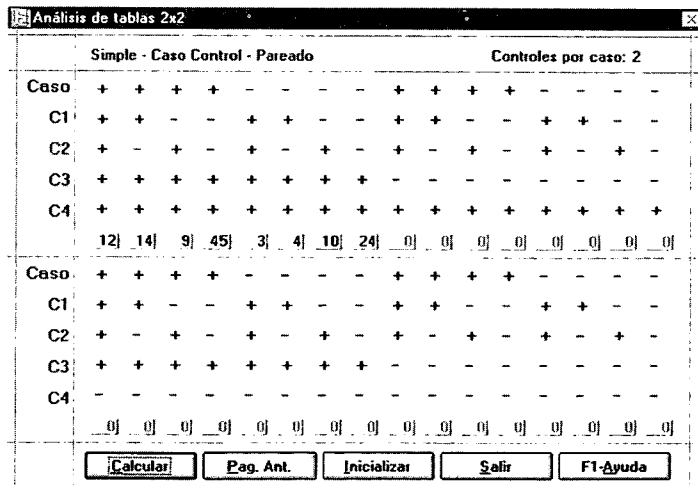
Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Efron B., Tibshirani R.J. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.

Tablas simples. Estudios caso-control pareado. —Análisis de estudios caso-control pareados.



Esta opción permite analizar tablas provenientes de estudios caso-control pareados, con 1, 2, 3 ó 4 controles por caso. EPIDAT presenta una pantalla en la que, según el número de controles por caso, se activan las posibles combinaciones de caso-contróles.



- En los resultados se presentan la odds ratio estimada con los intervalos de confianza y el chi-cuadrado corregido. Se permite variar el nivel de confianza que por defecto es del 95%.

En el análisis pareado con dos controles o más por caso, se calculan las odds ratio de Mantel-Haenszel propuestas por Miettinen y los chi-cuadrados de Mantel-Haenszel propuestos por Pike & Morrow.

Análisis de tablas 2x2 - Resultados

Simple - Caso Control - Pareado

Cáncer de pulmón y tabaquismo Controles por caso: 2

	I.C.li	I.C.la
Odds Ratio (OR):	5.65	4.27 7.47
Chi Cuadrado:	49.24	Nivel de Confianza: 0.950
P_valor (p):	0.0000	

Ejercicio: 4.3. Suponga que un estudio de casos y controles tiene como objetivo analizar la posible asociación entre tabaquismo de la madre gestante y bajo peso al nacer. Al diseñarlo decide equiparar individualmente cada caso con un control de similar edad. Seleccionados 50 casos y sus controles correspondientes obtiene la siguiente distribución:

		control	
		+	-
caso	+	8	26
	-	11	5

Estime el OR existente entre el tabaquismo de la madre y el bajo peso al nacer, eliminada la confusión que pueda ocasionar la edad de la madre. Calcule los intervalos de confianza de su estimación.

(Adaptado de: Londoño Fernández J.L. Metodología de la investigación epidemiológica. Editorial Universidad de Antioquia. Medellín 1995).

BIBLIOGRAFÍA

Schlesselman J.J. Case-Control Studies. Design, Conduct, Analysis. Oxford University Press. New York. 1982. 207-220.

Tablas estratificadas. Estudios transversales y cohortes. —Análisis de datos transversales y de seguimiento estratificados.

Análisis de tablas 2x2

Título del Estudio: Cáncer de pulmón y tabaquismo

Tipo de Tabla	Tipo de Estudio	Forma
<input type="radio"/> Simple	<input checked="" type="radio"/> Transversal / Cohortes	<input type="radio"/> No Pareado
<input checked="" type="radio"/> Estratificada	<input type="radio"/> Caso-control	<input type="radio"/> Pareado
		Controles: 0

Análisis de tablas 2x2

Estratificada - Transversal / Cohortes

Estrato: >74 a

	Enfermos	Sanos	
Fr. +	45	12	57
Fr. -	13	24	37
	58	36	94

Añadir
Eliminar
Descartar

	a	b	c	d	abcd
Tabla 1	24	12	11	53	100
Tabla 2	29	14	9	32	84

Calcular Pag. Ant. Inicializar Salir F1-Ayuda

Esta opción analiza los datos tabulados en estratos a partir de un estudio transversal (descriptivo) o de seguimiento (cohortes). EPIDAT calcula las odds ratios crudas de cada estrato con sus intervalos de confianza, la OR cruda y ajustada por los métodos de Mantel-Haenszel (M-H) y Wolf. También realiza la prueba de asociación de M-H con su valor p .

Análisis de tablas 2x2. Resultados

Estratificada - Transversal / Cohortes

Cáncer de pulmón y tabaquismo

Estrato	a	b	c	d	OR	I.C.li	I.C.la
<55 a	24	12	11	53	9.64	3.73	24.91
55-74 a	29	14	9	32	7.37	2.77	19.56
>74 a	45	12	13	24	6.92	2.74	17.51

Nivel de confianza: 0.950

O.R. crudo: 8.51 I.C.li 4.96 I.C.la 14.62

O.R. Mantel-Haenszel: 7.87 I.C.li 4.54 I.C.la 13.62

O.R. Wolf: 7.47 I.C.li 4.36 I.C.la 12.82

Prueba de asociación (M-H): 57.600 | P-valor: 0.0000

Más ...
Página Anterior
Salir
Imprimir
F1-Ayuda

- Al pulsar "Más" en la pantalla de resultados, se presentan los cálculos de los riesgos relativos de cada estrato con sus intervalos de confianza, el RR crudo y ajustado por los métodos de Mantel-Haenszel y Katz y la prueba de asociación de M-H.

Análisis de tablas 2x2 - Resultados								
Estratificada - Transversal / Cohortes								
Cáncer de pulmón y tabaquismo								
Estrato	a	b	c	d	RR	I.C.li	I.C.ls	
<55 a	24	12	11	53	3.88	2.16	6.97	+
55-74 a	29	14	9	32	3.07	1.66	5.67	
>74 a	45	12	13	24	2.25	1.42	3.55	+

Nivel de confianza:	<input type="text" value="0.950"/>	I.C.li	I.C.ls	
R.R. crudo:	<input type="text" value="3.10"/>	<input type="text" value="2.25"/>	<input type="text" value="4.25"/>	<input type="button" value="Página Anterior"/>
R.R. Mantel-Haenszel:	<input type="text" value="2.87"/>	<input type="text" value="2.11"/>	<input type="text" value="3.88"/>	<input type="button" value="Salir"/>
R.R. Katz:	<input type="text" value="2.76"/>	<input type="text" value="2.04"/>	<input type="text" value="3.74"/>	<input type="button" value="Imprimir"/>
Prueba de asociación (M-H):	<input type="text" value="57.600"/>	P-valor:	<input type="text" value="0.0000"/>	<input type="button" value="F1-Ayuda"/>

Ejercicio: 4.4. Continuando con el ejercicio 4.1., suponga que decide estratificar a los individuos de su estudio en tres grupos, según la frecuencia de utilización de mosquitero durante las noches, obteniendo la siguiente distribución:

		enfermo		sano
utilización	nunca	TA	112	126
		otro	42	277
de	ocasionalmente	TA	75	765
		otro	21	985
mosquitero	siempre	TA	29	1.123
		otro	32	1.549

Calcule los riesgos relativos de cada estrato con sus intervalos de confianza y el riesgo relativo global. Comente los resultados.

BIBLIOGRAFÍA

Kahn H.A., Sempos C.T. *Statistical Methods in Epidemiology*. New York: Oxford University Press; 1989.

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Tablas estratificadas. Estudios caso-control. —Análisis de datos de estudios caso-control estratificados.

Análisis de tablas 2x2

Título del Estudio: Cáncer de pulmón y tabaquismo

Tipo de Tabla	Tipo de Estudio	Forma
<input type="radio"/> Simple	<input type="radio"/> Transversal / Cohortes	<input type="radio"/> No Pareado
<input checked="" type="radio"/> Estratificada	<input checked="" type="radio"/> Caso-control	<input type="radio"/> Pareado
		Controles: <input type="text" value="0"/> <input type="text" value="0"/>

Análisis de tablas 2x2

Estratificada - Caso Control

Estrato: >74

		Enfermos	Sanos	
Fr.	+	<input type="text" value="31"/>	<input type="text" value="28"/>	<input type="text" value="59"/>
	-	<input type="text" value="36"/>	<input type="text" value="24"/>	<input type="text" value="60"/>
		<input type="text" value="67"/>	<input type="text" value="52"/>	<input type="text" value="119"/>

	a	b	c	d	abcd
Tabla 1	25	12	32	69	138
Tabla 2	24	12	21	58	115

Esta opción analiza los datos tabulados en estratos a partir de un estudio de casos y controles. EPIDAT calcula las odds ratios crudas de cada estrato con sus

Análisis de tablas 2x2. Resultados

Estratificada - Caso Control

Cáncer de pulmón y tabaquismo

Estrato	a	b	c	d	OR	I.C.li	I.C.la
<55 a	25	12	32	69	4.49	2.01	10.06
55-74 a	24	12	21	58	5.52	2.35	12.98
>74 a	31	28	36	24	0.74	0.36	1.53

Nivel de confianza:

	I.C.li	I.C.la
O.R. crudo:	<input type="text" value="2.61"/>	<input type="text" value="4.03"/>
O.R. Mantel-Haenszel:	<input type="text" value="2.29"/>	<input type="text" value="3.54"/>
O.R. Wolf:	<input type="text" value="2.30"/>	<input type="text" value="3.62"/>

Prueba de asociación (M-H): P-valor:

intervalos de confianza, la OR cruda y ajustada por los métodos de Mantel–Haenszel y Wolf. También realiza la prueba de asociación de M–H con su valor p .

Ejercicio: 4.5. Siguiendo con el estudio del problema 4.2., los investigadores, ante la posible asociación entre la obesidad y la enfermedad vascular obstructiva y de la primera con la diabetes, decidieron ajustar el valor de la OR por obesidad para controlar el posible efecto confusor de esta variable. Para ello se conformaron los dos estratos que se muestran en las siguientes tablas:

Obesidad	si	enfermedad carotídea			no	enfermedad carotídea		
		si	no	total		si	no	total
diabetes	si	37	8	45	si	44	18	62
	no	288	298	586	no	131	176	307
	Total	325	306	631	Total	175	194	369

Calcule los OR parciales, crudo y ajustado con sus intervalos de confianza y las pruebas de interacción y homogeneidad. Comente e interprete los resultados.

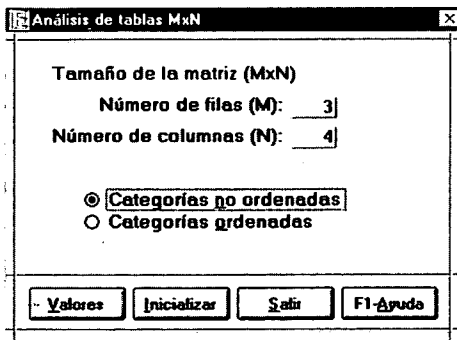
BIBLIOGRAFÍA

Kahn H.A., Sempos C.T. *Statistical Methods in Epidemiology*. New York: Oxford University Press; 1989.

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Tablas MxN

Esta opción permite realizar pruebas y medidas de asociación entre 2 variables de hasta un máximo de 10 categorías cada una. Una vez definido el tamaño de la tabla es necesario señalar si las categorías están ordenadas o no. Pulsando la tecla “Valores” se pasa a la pantalla de introducción de datos. Pulsando “Calcular” aparecerá la pantalla de resultados.



32	21	14	6	Calcular
65	45	8	24	Pag.Ant.
14	12	35	24	Salir
				F1-Ayuda

Categorías no ordenadas

Se calculan las siguientes pruebas de asociación/independencia:

chi-cuadrado de Pearson, con $(m-1)(n-1)$ grados de libertad. Con esta prueba pueden examinarse dos hipótesis:

1. Las variables A y B son independientes.
2. Si A representa a una variable categórica de interés y B tiene N niveles que representan a N poblaciones, la prueba contrasta si la distribución de frecuencias de A es la misma en cada población.

G cuadrado. —Es el estadístico chi-cuadrado de la razón de verosimilitud. Constituye una alternativa a la prueba chi-cuadrado de Pearson. Está basado en la teoría de la máxima verosimilitud y se utiliza sobre todo en tablas de contingencia multivariante.

También se calculan las medidas de asociación, que son independientes del tamaño muestral (lo que no ocurre con las pruebas de asociación anteriormente descritas):

Coefficiente Phi. —Corrige la dependencia del estadístico chi-cuadrado respecto al tamaño muestral global n . Si la tabla es 2×2 puede tomar valores entre 0 y 1. En analogía al coeficiente de correlación, el valor 0 indica que no existe asociación y el valor 1 asociación completa. En el caso de que las filas o columnas sean >2 , puede tomar valores por fuera de ese rango.

Coefficiente de contingencia C. —Se interpreta como el coeficiente Phi, sin embargo, en el caso de asociación entre A y B, ésta depende del número de filas y columnas de la tabla.

Coefficiente V de Cramer. —Con la misma interpretación que los anteriores, corrige el efecto del número de filas y columnas del coeficiente de contingencia C.

	P_valor (p)
Chi cuadrado de Pearson: 64.434	0.0000
G2: 67.388	0.0000
Phi: 0.463	0.0000
Coefficiente de contingencia C: 0.420	0.0000
V de Cramer: 0.327	0.0000
Tau de Goodman y Kruskal	
Filas/columnas:	0.11264
Columnas/filas:	0.06903
<input type="button" value="Pág Ant"/> <input type="button" value="Salir"/> <input type="button" value="Imprimir"/> <input type="button" value="F1 Ayuda"/> <input type="button" value="Más"/>	

Tau de Goodman–Kruskal. —Mide qué tal explica una variable A la otra variable B. Se calcula comparando la probabilidad de error en dos situaciones: A en función de B y B en función de A. Su valor se encuentra entre 0 y 1. Cuanto más se acerque a 0 menor es la relación entre la variable considerada como independiente y la dependiente. Cuanto más se acerque a 1, más significativamente disminuye el error de clasificación la variable independiente.

Ejercicio: 4.6. El número de consultas según el sexo de los/as pacientes por servicios en el hospital de su área de salud se refleja en la siguiente tabla:

	neumología	traumatología	neurología	digestivo
hombres	1.254	658	478	986
mujeres	987	427	652	684

Pruebe la hipótesis de que la distribución por sexos es independiente del tipo de consulta. Obtenga los coeficientes de asociación apropiados y coméntelos.

BIBLIOGRAFÍA

Everit B.S. *The Analysis of Contingency Tables*. London: Chapman & Hall; 1977.

Categorías ordenadas

Calcula los mismos coeficientes que la opción *categorías no ordenadas*. Al pulsar la tecla “Más” aparecerá otra pantalla con los resultados de los coeficientes propios de categorías ordenadas:

Coefficiente de correlación de Spearman. —Se calcula aplicando el coeficiente de correlación de Pearson a dos variables ordinales cuyas puntua-

ciones han sido sustituidas por sus rangos. Toma valores entre -1 y 1. Un valor 0 indica que no existe ninguna relación lineal entre las variables.

Tau-b de Kendall. —Para el cálculo de este coeficiente se contabilizan las concordancias, discordancias y empates entre pares de casos, a partir de los que se genera el índice de asociación. Puede tomar valores entre -1 (máxima discordancia) y 1 (máxima concordancia). El valor 0 indica que no existe relación entre ambas variables.

Tau-c de Stuart. —Similar al anterior utiliza para el cálculo el menor número de filas y columnas en lugar del número de empates. Cuanto más probable sea que aparezcan empates, es decir que el número de filas o columnas sea menor, el índice disminuye.

Correlación producto-momento. —Calcula el coeficiente de correlación utilizando las entradas de las celdas como frecuencias y los índices de las celdas como valores observados.

D de Sommers. —Permite realizar un análisis de relación entre dos variables considerando una de ellas como dependiente, por lo que se obtendrán dos índices, uno considerando la variable situada en las filas como independiente y otro haciendo lo mismo con la variable situada en las columnas. Para el cálculo utiliza los empates que se dan en la variable dependiente.

Comentario. —El índice más utilizado es el de Spearman, por lo que su uso garantiza mejor comunicación de los resultados. En caso de muchos empates se recomienda utilizar las tau o la D de Sommers. Si interesa contar con una medida asimétrica para cada variable, utilice la D de Sommers.

Tablas NxN

1. **kappa.** —ver opciones *Concordancia: 2 observadores. 2 categorías y 2 observadores. Múltiples categorías.* En el módulo de tablas, EPIDAT sólo da el resultado de kappa sin ponderar con sus intervalos de confianza.
2. **homogeneidad NxN.** —Prueba la hipótesis de que las probabilidades marginales de las filas son iguales a las probabilidades marginales de las columnas correspondientes.

Tablas 2xN

1. **Prueba de tendencia lineal (prueba de regresión lineal).** —Realiza una regresión lineal de cada elemento de la tabla sobre el total de su columna y calcula la prueba chi-cuadrado con 1 grado de libertad.



Análisis de tablas MxN - Resultados			
Correlación de Spearman:	0.29909		
Tau-b de Kendall:	0.25618	I.C.li	I.C.ls
Tau-c de Stuart:	0.26140	0.15	0.37
Correlación producto-momento:	0.29998	0.19	0.40
D de Sommers	Filas/columnas:	0.15	0.32
	Columnas/filas:	0.17	0.37
Kappa:		I.C.li	I.C.ls
Homeogeneidad NxN:			P-valor (p):
Prueba de tendencia lineal:			P-valor (p):
Nivel de confianza:	0.950		
<input type="button" value="Pág Ant"/> <input type="button" value="Salir"/> <input type="button" value="Imprimir"/> <input type="button" value="F1 Ayuda"/>			

Ejercicio: 4.7. Suponga que en su municipio pretenden analizar la posible relación entre las afecciones víricas comunes de la infancia y la edad de los niños, como paso previo para estudiar la influencia de la escolarización como factor de riesgo en la transmisión de estas enfermedades. Para ello se estudia la situación inmunológica frente al virus del sarampión, rubéola y parotiditis, con los siguientes resultados:

	edad		
	< 6 años (%)	6-7 años (%)	> 7 años (%)
niveles de anticuerpos que sugieren haber padecido			
sarampión	35	9	1
rubéola	60	20	18
parotiditis	70	28	24

Estudie la posible asociación u homogeneidad entre las variables. Comente los resultados.

BIBLIOGRAFÍA

Everitt B.S. *The Analysis of Contingency Tables*. London: Chapman and Hall; 1977.
 Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.
 Manzano V. *Inferencia estadística. Aplicaciones con SPSS/PC+*. Madrid: Ra- ma; 1995.

Vigilancia epidemiológica. —Métodos de vigilancia epidemiológica.

Captura-Recaptura

Existen diversos métodos para evaluar y mejorar la integridad y perfección de sistemas de registro y vigilancia en salud pública. El más clásico es el

propuesto por Chandra Sekar–Deming para estimar el número de casos no notificados por dos sistemas de registro que traten de obtener información sobre un mismo tipo de eventos.

Este método asume que:

1. La población de donde se obtiene la información es cerrada.
2. El método de cruce de los dos sistemas de información identifica correctamente los casos notificados por ambos y los notificados por uno de ellos pero no por el otro.
3. Los casos identificados por ambos sistemas son casos reales.
4. Los dos sistemas son independientes.

		Sistema 1				
S i s t e m a		Incluido	No Incluido	Marginales		
		Incluido	36	12	48	<input type="button" value="Calcular"/> <input type="button" value="Inicializar"/> <input type="button" value="Salir"/> <input type="button" value="F1-Ayuda"/>
		No Incluido	8	2	10	
		Marginales	44	14	58	

- En la casilla **a** debe introducirse el número de casos notificados por los dos sistemas, en la **b** los que notifica el sistema 2 pero no el 1. En la casilla **c** los que notifica el sistema 1 pero no el 2. Al pulsar “Calcular” Epidat estima la casilla **d**, es decir, los que no notifica ninguno de los dos sistemas de vigilancia evaluados. También calcula en porcentaje la integridad de la notificación de los dos sistemas por separado y combi-

		Sistema 1				
S i s t e m a		Incluido	No Incluido	Marginales		
		Incluido	36	12	48	Integridad de notificación Sistema 1: 75.09% Sistema 2: 81.91% Los dos combinados: 95.57% Nivel de confianza: 0.950 I.C.Ii: 54 I.C.Is: 62 Número de casos estimados: 58
		No Incluido	8	2	10	
		Marginales	44	14	58	

dados. Se presenta el número total de casos (registrados y estimados) con los intervalos de confianza. Se permite variar el nivel de confianza que por defecto es del 95%.

Ejercicio: 4.8. En su región funciona un registro de mortalidad perinatal que recoge información de dos fuentes independientes: certificados estadísticos de defunción e informes de alta hospitalaria. Una vez cruzados los datos de ambos sistemas, observa que el primero recoge 25 casos que no figuran en el segundo, éste contiene 17 casos que no constan en el primero y 113 casos aparecen registrados en ambos sistemas. Estime el total de casos de mortalidad perinatal en su región con sus intervalos de confianza y valore la integridad de la notificación de cada sistema por separado y en conjunto.

BIBLIOGRAFÍA

Chandra Sekar C., Deming W.E. On a method of estimating birth and death rates and the extent of registration. *J Am Stat Assoc.* 1949; 44: 101-15.

Teutsch S.M., Churchill R.E. Principles and Practice of Public Health Surveillance. *Oxford University Press.* New York. 1994. 143-45.

Detección de clusters

Esta opción utiliza el método de Knox para detectar clusters o agrupamiento de casos cercanos en tiempo y espacio. Calcula la probabilidad de observar casos cercanos en tiempo y espacio, usando la distribución de Poisson, dado un número esperado de casos calculado en base a los valores marginales de la tabla.

El principal problema que presenta este método radica en la arbitrariedad del usuario o investigador al definir lo que es cercano en tiempo y espacio y lo que no lo es.

		Cercanos en tiempo		Marginales
		Si	No	
C e r c a n o s	Si	<input type="text" value="36"/>	<input type="text" value="24"/>	<input type="text" value="60"/>
	No	<input type="text" value="12"/>	<input type="text" value="74"/>	<input type="text" value="86"/>
Marginales		<input type="text" value="48"/>	<input type="text" value="98"/>	<input type="text" value="146"/>

Buttons: **Calcular**, **Inicializar**, **Salir**, **F1-Ayuda**

- Una vez introducidos los valores de las celdas y pulsado el botón “Calcular”, EPIDAT presenta la pantalla de resultados con los casos esperados y la probabilidad de que los valores de la celda a formen un cluster o agrupamiento.

Vigilancia epidemiológica - Detección de clusters - Resulta...				
C e r c a n o s	Cercanos en tiempo			
		Si	No	Marginales
	Si	36	24	60
No	12	74	86	
Marginales	48	98	146	
Valor esperado:		19.7260		
Probabilidad:		0.0001		
<input type="button" value="Pag. Anterior"/> <input type="button" value="Imprimir"/> <input type="button" value="Salir"/> <input type="button" value="F1-Ayuda"/>				

Ejercicio: 4.9. En su unidad de vigilancia epidemiológica vienen observando un incremento de casos de hepatitis A en los niños de su región. Deciden estudiar si existe un agrupamiento de los casos para lo que deciden clasificarlos, en base al período de incubación, como “cercanos en el tiempo” a los que se diagnosticaron con menos de 50 días de diferencia. Subjetivamente, y considerando las características de su zona, decide considerar como “cercanos en espacio” a aquellos casos en los que los domicilios o colegios distan menos de 500 metros. Tabulados sus datos comprueba que 18 pares de enfermos eran cercanos en tiempo pero no en espacio, que 35 lo eran en espacio pero no en tiempo, que 45 lo eran en tiempo y espacio, mientras que 112 no lo eran ni en tiempo ni en espacio. ¿Podría usted decir que en su región ha habido un agrupamiento de casos de hepatitis A mayor que el esperado? Razone la respuesta.

Bibliografía

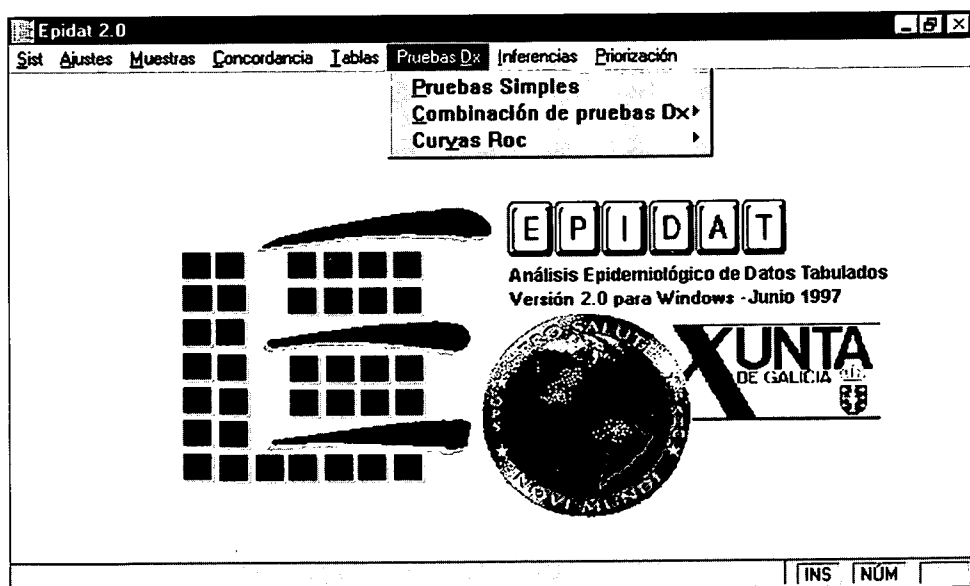
Knox E.G. The detection of space-time interaction. *Appl Statist.* 1964; 13: 25-9.
 Teutsch S.M., Churchill R.E. Principles and Practice of Public Health Surveillance. *Oxford University Press.* New York. 1994; 142-3.



PRUEBAS DIAGNÓSTICAS



Este módulo tiene tres submódulos: Pruebas simples, Combinación de pruebas diagnósticas y Curvas ROC (Receiver Operating Characteristic). En cualquiera de ellos se puede trabajar con tablas 2x2, además en el submódulo de Pruebas agrupadas se tiene la opción de trabajar directamente con los índices de las distintas pruebas que se quieren agrupar, y en el submódulo de Curvas ROC los datos pueden introducirse por medio de una tabla 2xN, donde N es el número de categorías en las que se clasifican los valores que toman los individuos en la prueba. En cualquiera de los submódulos los índices que se calculan intentan describir la relación entre los resultados de un prueba y los diagnósticos “verdaderos”.



Pruebas diagnósticas simples

Calcula, con los intervalos correspondientes, la sensibilidad, especificidad, valores predictivos positivo y negativo y las razones de verosimilitudes positiva y negativa de una prueba diagnóstica en relación a otra prueba de referencia o estándar.

		Enfermedad	
		+	-
Prueba diagnóstica	+	a	b
	-	c	d

sensibilidad: es la proporción de individuos con la enfermedad que tienen una prueba positiva.

$$a/(a+c)$$

especificidad: es la proporción de sanos que tienen una prueba negativa.

$$d/(b+d)$$

valor predictivo positivo: es la probabilidad de que un individuo con resultado positivo en la prueba tenga la enfermedad.

$$a/(a+b)$$

valor predictivo negativo: es la probabilidad de que un individuo con resultado negativo en la prueba no tenga la enfermedad.

$$d/(c+d)$$

válor global de la prueba: es la probabilidad de que un individuo sea clasificado correctamente por la prueba.

$$(a+d)/(a+b+c+d)$$

razón de verosimilitud positiva: es la probabilidad de una prueba positiva en presencia de la enfermedad, dividida por la probabilidad de una prueba positiva en ausencia de la enfermedad.

$$(a/a+c)/(b/b+d)$$

razón de verosimilitud negativa: es la probabilidad de una prueba negativa en presencia de la enfermedad, dividida por la probabilidad de una prueba negativa en ausencia de la enfermedad.

$$(c/a+c)/(d/b+d)$$

Hay dos alternativas para realizar los cálculos. La primera y más habitual es introducir los valores de a, b, c, d, y calcular los parámetros descritos

Pruebas diagnósticas simples				
Prueba de referencia				
		+	-	
P r u e b a	+	145	24	169
	-	36	425	461
		181	449	205
Sensibilidad:		80.11 %		
Especificidad:		94.65 %		
Valor predictivo positivo:		85.79 %		
Valor predictivo negativo:		92.19 %		
Calcular		Inicializer	Salir	F1-Ayuda

anteriormente. Otra sería introducir los valores de alguna de las casillas y fijar el valor de algún parámetro (sensibilidad, especificidad...). EPIDAT calculará el valor de las casillas vacías. Si alguno de estos cálculos no dan un valor entero exacto, se redondea al entero más cercano.

- Una vez que se pulse “Calcular” se recalculan, a partir de los datos de la tabla, todos los índices.
- En la pantalla de resultados se presentan los valores calculados con los intervalos de confianza correspondientes. Se permite variar el nivel de confianza que por defecto es del 95%.

		Prueba de referencia		Pruebas diagnósticas simples	
		+	-		
P r u e b a	+	145	24	Nivel de confianza: <input style="width: 50px;" type="text" value="0.950"/>	
	-	26	425		
		181	449	630	

				Intervalos de confianza	
				Inferior	Superior
Sensibilidad:	80.11%	73.39	85.51		
Especificidad:	94.65%	92.03	96.47		
Valor predictivo positivo:	85.79%	79.40	90.51		
Valor predictivo negativo:	92.19%	89.25	94.39		
Valor global de la prueba:	90.47%	87.84	92.59		
Razón de verosimilitud positiva:	14.98	10.08	22.26		
Razón de verosimilitud negativa:	0.21	0.15	0.28		

Página Anterior
Salir
Imprimir
F1-Ayuda

Ejercicio: 5.1. Suponga que se dispone a realizar una prueba de tamizaje de cáncer de mama mediante mamografía en una población de 45.000 mujeres con edades comprendidas entre los 50 y 64 años de edad. Por estudios previos, sabe que la sensibilidad de la prueba es del 80% y la especificidad del 97%. En el supuesto que la prevalencia de cáncer de mama en la población de estudio sea del 66,6x100.000, calcule los valores predictivos de la mamografía con sus intervalos de confianza.

BIBLIOGRAFÍA

Fletcher R.H., Fletcher S.W., Wagner E.H. *Clinical Epidemiology. The essentials.* Baltimore: Williams & Wilkins; 1988.

Fleiss J.L. *Statistical methods for rates and proportions.* New York: John Wiley & Sons; 1981.

Silva L.C. *Métodos estadísticos para la investigación epidemiológica.* Seminario Internacional de Estadística de Euskadi. 1987



Combinación de pruebas diagnósticas

Datos tabulados

Esta opción permite calcular la sensibilidad, especificidad, valores predictivos, valor global y razones de verosimilitud de pruebas diagnósticas múltiples, realizadas en serie y en paralelo, a partir de tablas 2x2.

Pruebas en serie. —Consisten en realizar otra prueba diagnóstica (o la misma) a los sujetos clasificados como positivos en una prueba anterior. Los sujetos con resultado negativo en una prueba ya se consideran “negativos” y no se someten a la siguiente prueba.

Por tanto, “a+b” de la primera tabla=“a+b+c+d” de la segunda, “a” de la primera=“a+c” de la segunda y “b” de la primera=“b+d” de la segunda. Las mismas condiciones tendrían que cumplir los valores de una tercera tabla respecto a los de la segunda y así sucesivamente.

Combinación de pruebas diagnósticas - Datos tabulados

Prueba de referencia

	+	-	
P r u e b a	+	50	10
	-	20	10
		70	20

En serie
 En paralelo

[Añadir] [Eliminar] [Descartar]

	a	b	c	d	ab	abcd
Tabla 1	100	50	40	60	150	250
Tabla 2	70	20	30	30	90	150

[Calcular] [Inicializar] [Salir] [F1-Ayuda]

Las pruebas en serie disminuyen la sensibilidad y aumentan la especificidad.

Pruebas diagnósticas

Prueba de referencia

	+	-	
P r u e b a	+	50	10
	-	90	100
		140	110

Combinación de pruebas diagnósticas - Datos tabulados

Nivel de confianza: 0.950

		Intervalos de confianza
		Inferior Superior
Sensibilidad:	35.71%	27.92 44.29
Especificidad:	90.90%	83.52 95.31
Valor predictivo positivo:	83.33%	71.02 91.29
Valor predictivo negativo:	52.63%	45.29 59.86
Valor global de la prueba:	60.00%	53.61 66.06
Razón de verosimilitud positiva:	3.92	2.08 7.38
Razón de verosimilitud negativa:	0.70	0.61 0.81

[Página Anterior] [Salir] [Imprimir] [F1-Ayuda]

Combinación de pruebas diagnósticas - Datos tabulados

		Prueba de referencia		
		+	-	
P r u e b a	+	8	10	18
	-	2	170	172
		10	180	190

En serie
 En paralelo

	a	b	c	d	cd	abcd	
Tabla 1	100	50	40	200	240	390	+
Tabla 2	30	20	10	180	190	240	

Pruebas en paralelo. —Consisten en realizar otra prueba diagnóstica (o la misma) a los sujetos clasificados como negativos en una prueba anterior. Los sujetos con resultado positivo en una prueba ya se consideran “positivos” y no se les someten a la siguiente prueba.

Por tanto, “c+d” de la primera tabla=“a+b+c+d” de la segunda, “c” de la primera=“a+c” de la segunda y “d” de la primera=“b+d” de la segunda. Las mismas condiciones tendrían que cumplir los valores de una tercera tabla respecto a los de la segunda y así sucesivamente.

Combinación de pruebas diagnósticas - Datos tabulados

		Prueba de referencia		
		+	-	
P r u e b a	+	8	10	18
	-	2	170	172
		10	180	190

En serie
 En paralelo

	a	b	c	d	cd	abcd	
Tabla 1	100	50	40	200	240	390	+
Tabla 2	30	20	10	180	190	240	

Las pruebas en paralelo aumentan la sensibilidad y disminuyen la especificidad.

Pruebas diagnósticas

		Prueba de referencia		Combinación de pruebas diagnósticas - Datos tabulados	
		+	-		
Prueba	+	138	80	218	Nivel de confianza: <input style="width: 50px;" type="text" value="0.950"/>
	-	2	170	172	
		140	250	390	

		Intervalos de confianza	
		Inferior	Superior
Sensibilidad:	98.57%	94.40	99.75
Especificidad:	68.00%	61.77	73.66
Valor predictivo positivo:	63.30%	56.49	69.63
Valor predictivo negativo:	98.83%	95.42	99.79
Valor global de la prueba:	78.97%	74.52	82.84
Razón de verosimilitud positiva:	3.08	2.56	3.69
Razón de verosimilitud negativa:	0.02	0.00	0.08

Una vez construidas las tablas, los cálculos se realizan como en la opción pruebas diagnósticas simples.

En la pantalla de resultados se presentan los valores calculados con los intervalos de confianza correspondientes.

Se permite variar el nivel de confianza que por defecto es del 95%.

Ejercicio: 5.2. *Un centro de referencia de VIH/SIDA utiliza como prueba de tamizaje la prueba ELISA. Sobre los resultados de un grupo de 1.000 sujetos que acudieron a efectuar la prueba, procedentes de un grupo con una prevalencia conocida del 5% de positivos, quiere estudiar los cambios en la sensibilidad, especificidad y valores predictivos al someter a una prueba de Western Blot a aquellos que tienen ELISA positivo. Resuélvale el problema al centro de referencia sabiendo que la sensibilidad y especificidad del ELISA son del 98% y 99,8% respectivamente y del Western blot, del 85% y 99,99% ¿Qué pasaría si en vez de realizar la prueba de confirmación a los positivos se la hacemos a los ELISA negativos?*

Valores diagnósticos

Esta opción permite calcular la sensibilidad, especificidad, valores predictivos, valor global y razones de verosimilitud de pruebas diagnósticas múltiples realizadas en serie y en paralelo, *a partir de la sensibilidad y especificidad de cada prueba.* Para el cálculo de los valores predictivos y de los *intervalos de confianza* es preciso introducir la prevalencia de la enfermedad y el tamaño de la población a la que se pretende realizar la prueba.



	Sensibilidad	Especificidad
1	98.00 %	92.00 %
2	94.00 %	87.00 %
3	86.00 %	91.00 %
4	80.00 %	94.00 %

Prevalencia de enfermedad: 0.24000

Población: 145800

La interpretación de las pruebas en serie y en paralelo es la misma que en “Combinación de pruebas diagnósticas. Datos tabulados”. Se permite variar el nivel de confianza que por defecto es del 95%.

	Sensibilidad	Especificidad	VP+	VP-
1	98.00 %	92.00 %	79.45 %	99.31 %
2	94.00 %	87.00 %	69.54 %	97.86 %
3	86.00 %	91.00 %	75.10 %	95.36 %
4	80.00 %	94.00 %	80.80 %	93.70 %
Serie	63.37 %	99.99 %	99.97 %	89.63 %
Paralelo	99.99 %	68.46 %	50.03 %	99.99 %

Prevalencia: 0.24000

Serie			Paralelo		
	I.C.6	I.C.1s		I.C.6	I.C.1s
SEN	63.37 %	62.87 63.88	SEN	99.99 %	99.98 99.99
ESP	99.99 %	99.98 99.99	ESP	68.46 %	68.19 68.73
VP+	99.97 %	99.93 99.98	VP+	50.03 %	49.66 50.40
VP-	89.63 %	89.46 89.80	VP-	99.99 %	99.99 99.99

Nivel de confianza: 0.950

Ejercicio: 5.3. La angiografía es la prueba de referencia para diagnosticar aneurisma cerebral. Sin embargo esta es una prueba costosa y no exenta de riesgo. Suponga que dispone de otras dos pruebas alternativas, "A" que tiene una sensibilidad del 80% y una especificidad del 60% y "B", con un 90% de sensibilidad y especificidad. En el grupo de pacientes seleccionados (100 individuos) la prevalencia de la enfermedad es del 20%.

Calcule la sensibilidad y especificidad de ambas pruebas realizadas en serie y en paralelo. Si su intención es clasificar como positivos al máximo de enfermos posibles, ¿cuál de las dos estrategias escogería? ¿Si su objetivo fuese utilizar una secuencia que incremente al máximo la proporción de individuos con el resultado de la prueba positivo y que tengan la enfermedad, por qué estrategia optaría? Razónelo.

Curvas ROC

Tablas 2x2

La curva ROC (Receiver Operating Characteristic) es la representación gráfica de la capacidad discriminativa de una prueba a lo largo de todos los posibles "puntos de corte". En el eje de ordenadas se sitúan los distintos valores de la sensibilidad y en el eje de abscisas los valores de 1-especificidad, es decir, la proporción de falsos positivos. EPIDAT utiliza el método no paramétrico para representar la curva.

El programa calcula el área bajo la curva, el error estándar, los intervalos de confianza y representa la curva ROC a partir de tablas 2x2. El total de enfermos y sanos deberá ser igual en todas las tablas introducidas, es decir a+c debe tener el mismo valor en todas las tablas, al igual que b+d.

Curvas Roc 2x2

Prueba de referencia

		+	-	
P r u e b a	+	80	2	82
	-	80	238	318
		160	240	

Tabla 1 a: 150 b: 126 c: 10 d: 114 ac: 160 bd: 240

Tabla 2 a: 102 b: 8 c: 58 d: 232 ac: 160 bd: 240

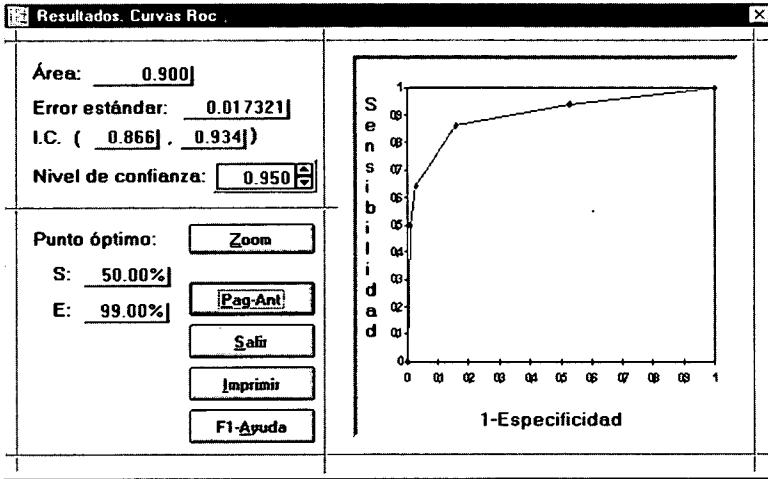
Tabla 3 a: 138 b: 38 c: 22 d: 202 ac: 160 bd: 240

Prevalencia: 0.04000 Razón de costes (FN/FP): 0.500

Buttons: Añadir, Eliminar, Descartar, Calcular, Inicializar, Salir, F1-Ayuda

Se permite variar el nivel de confianza que por defecto es del 95%.

EPIDAT incorpora un método simplificado para la estimación del umbral óptimo de sensibilidad–especificidad de las tablas introducidas. El método se basa en la razón de costes (coste de un falso positivo/coste de un falso negativo) y en la prevalencia de la enfermedad.



Ejercicio: 5.4. Suponga que quiere definir el mejor punto de corte, entre los cuatro que siguen, para una prueba diagnóstica de diabetes. Para ello determina cuantitativamente la glucosa en sangre a 80 diabéticos y 120 no diabéticos clasificándolos en las siguientes tablas para las que se consideran los puntos de corte de 90, 110, 130 y 150 mg/100 ml.

Punto de corte: 90 mg/100 ml

		Diabetes	
		+	-
Prueba dx	+	75	63
	-	5	57

Punto de corte: 110 mg/100 ml

		Diabetes	
		+	-
Prueba dx	+	69	19
	-	11	101

Punto de corte: 130 mg/100 ml

		Diabetes	
		+	-
Prueba dx	+	51	4
	-	29	116

Punto de corte: 150 mg/100 ml

		Diabetes	
		+	-
Prueba dx	+	40	1
	-	40	119

Dibuje la curva ROC para las distintas sensibilidades y especificidades, calcule el área por debajo de la curva y defina el punto de corte más eficiente suponiendo que la relación (ratio) entre el coste de un falso positivo y un falso negativo es de 1:5 y la prevalencia de diabetes en la comunidad objeto del programa de tamizaje es de aproximadamente 0,8%.



BIBLIOGRAFÍA

Sackett D.L., Haynes R.B., Tugwell P. *Epidemiología clínica. Una ciencia básica para la medicina clínica*. Madrid: Díaz de Santos S.A., 1989.

Hanley J.A., McNeil B.J. The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 1982; 143: 29-36.

Zweig M.H., Campbell G. Receiver-Operating Characteristic (ROC) Plots: A fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; 39: 561-577.

Silva L.C. *Métodos estadísticos para la investigación epidemiológica*. Seminario internacional de estadística de Euskadi. 1987.

Burgueño M.J., García-Bastos J.L., González-Buitrago J.M. Las curvas ROC en la evaluación de las pruebas diagnósticas. *Med Clin* 1995; 104: 661-670.

Tablas 2xN

Esta opción para representación de curvas ROC calcula el área bajo la curva, el error estándar, los intervalos de confianza y dibuja la curva, a partir de una tabla 2xN. En este tipo de presentación, en una columna se incluyen los sanos y en la otra los enfermos. En la primera fila se clasifican los individuos utilizando el punto de corte que garantice mayor sensibilidad, es decir, aquel con el que la probabilidad de que un enfermo tenga un resultado positivo sea mayor. En la segunda fila se incluyen sujetos utilizando un punto de corte de menor sensibilidad que el anterior pero mayor que el de la tercera fila y así sucesivamente hasta la última fila. Esta incluirá a aquellos que clasifique el punto de corte que nos garantice mayor especificidad, es decir, aquel con el que la probabilidad de que un sano tenga un resultado negativo sea mayor.

Curvas Roc

Prueba de referencia

+	-	
•		

Añadir
Eliminar
Descartar

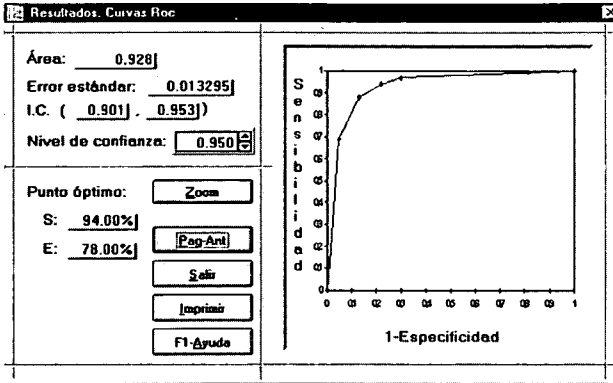
	a	b	
Tabla 1	6	165	↕
Tabla 2	6	21	
Tabla 3	12	21	
Tabla 4	36	18	
Tabla 5	135	12	↕

Prevalencia: 0.39000|

Razón de costes (FN/FP): 0.300|

Calcular Inicializar Salir F1-Ayuda

También permite identificar el umbral óptimo de sensibilidad–especificidad entre los puntos de corte introducidos (ver Curvas ROC. 2x2).



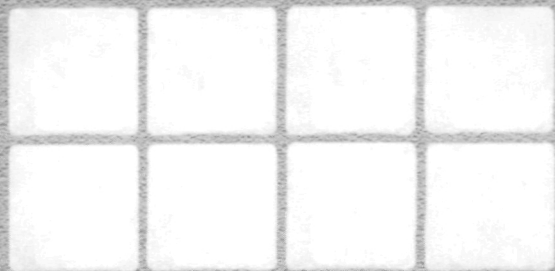
Ejercicio: 5.5. Un radiólogo de un programa de tamizaje de cáncer de mama desea analizar las variaciones de la sensibilidad y especificidad de la mamografía en un grupo seleccionado de mujeres con elevada prevalencia del problema. Para ello utilizó diferentes puntos de corte, basados en la definición de la imagen, a la hora de clasificar a las mujeres seleccionadas. Los valores obtenidos se muestran en la tabla siguiente:

		situación real			
		enfermas	sanas	total	
definición de imagen	claramente sana	(--)	2	53	55
	probablemente sana	(-)	2	7	9
	dudosa	(±)	4	7	11
	probablemente enferma	(+)	12	6	18
	claramente enferma	(++)	45	4	49
	total		65	77	142

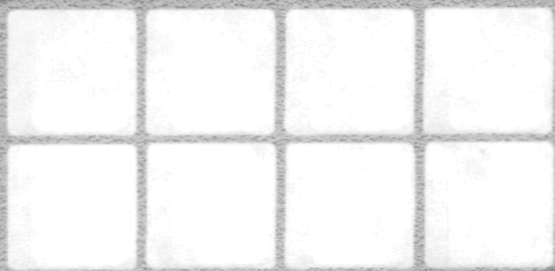
Dibuje la curva ROC y calcule el área bajo la curva. Defina el punto óptimo considerando que la relación (ratio) de costes entre un resultado falso positivo y un falso negativo, a efectos de este ejercicio, es 0,2. Interprete los resultados.

BIBLIOGRAFÍA

Véase bibliografía de Curvas ROC. 2x2.

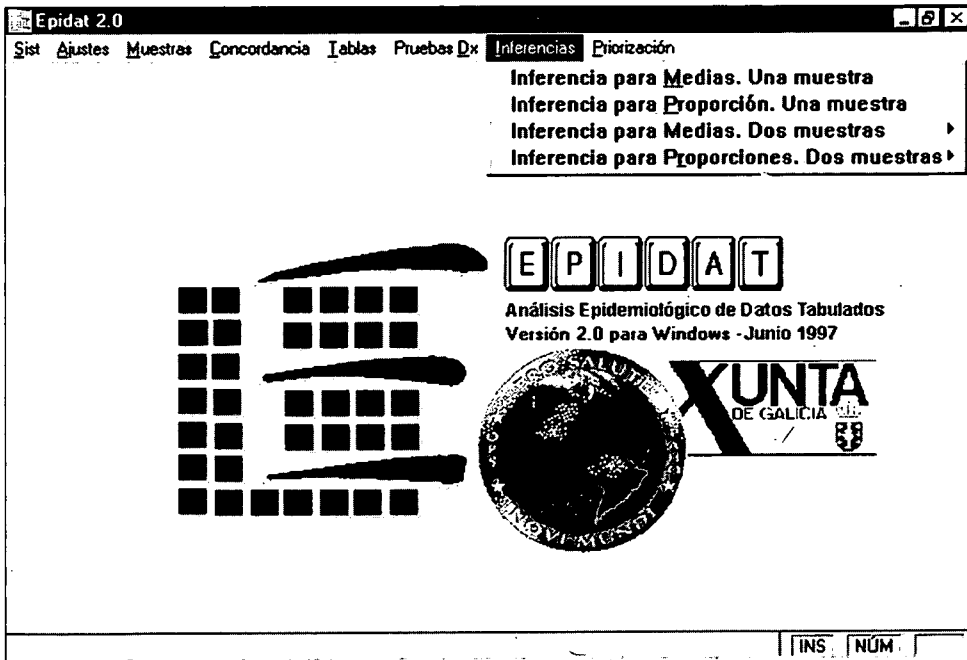


INFERENCIAS



Este módulo permite contrastar hipótesis de igualdad de medias y proporciones y calcular intervalos de confianza.

NOTA: Por defecto en el módulo de inferencia los contrastes de hipótesis son **bilaterales**. Si interesa realizar pruebas **unilaterales** hay que dividir el valor de p por dos, ya que se da por defecto el valor p del estadístico en valor absoluto.



Inferencia para la media. Una muestra

Esta opción permite construir los intervalos de confianza para la media a partir de la media muestral y la desviación estándar, y realizar, si se desea, contrastes relacionados con la media, cuando se trabaja con una población.

- Para construir los intervalos de confianza se debe introducir la media, el tamaño de la muestra y la desviación estándar (indicando si es muestral o poblacional). Si se quiere contrastar que la media de una variable aleatoria, que sigue una distribución normal de parámetros desconocidos, es igual a una teórica, debe marcarse con una "X" la casilla correspondiente a "Contraste de hipótesis" e introducir la media teórica.

Inferencia para la media. 1 muestra	
Media muestral:	25.630
Tamaño muestral:	245
Desviación estándar:	11.320
<input checked="" type="radio"/> Muestral <input type="radio"/> Poblacional	
<input checked="" type="checkbox"/> Contraste de hipótesis	
Media teórica:	22.000
<input type="button" value="Calcular"/> <input type="button" value="Inicializar"/> <input type="button" value="Salir"/> <input type="button" value="F1-Ayuda"/>	

- Una vez suministrados los datos al programa, pulsar “Calcular”. EPIDAT presentará una pantalla de resultados en la que aparecerá la media muestral, el intervalo de confianza construido a partir de la media muestral y la desviación estándar indicada, y el nivel de confianza elegido, que por defecto es 95%. La opción de Contraste de hipótesis presenta el estadístico evaluado y el valor *p* o nivel crítico (si no se selecciona esta opción aparecerán asteriscos).

Inferencia para la media. 1 muestra - Resultados	
Media muestral:	25.630
Intervalo de confianza:	24.212 27.047
Nivel de confianza:	0.950
Contraste de hipótesis H ₀ : $\mu = \mu_0$ - H _a : $\mu \neq \mu_0$	
Estadístico evaluado:	5.019300
P-valor (p):	0.0000002
<input type="button" value="Salir"/> <input type="button" value="Pg-Ant"/> <input type="button" value="Imprimir"/> <input type="button" value="F1-Ayuda"/>	

Ejercicio: 6.1. En un estudio de salud bucodental realizado en los escolares de su municipio, entre otros indicadores, se calculó el índice CAO (CPO)-co que resultó ser igual a 2,47. Si el total de niños examinados fue 1.600, y la desviación estándar 0,87. Estimar el índice CAO (CPO)-co mediante un intervalo de confianza, utilizando distintos niveles de confianza.

BIBLIOGRAFÍA

Peña D. *Estadística. Modelos y Métodos. 1 Fundamentos*. Madrid: Alianza Universidad Textos; 1993.



Inferencia para la proporción. Una muestra

Esta opción permite construir los intervalos de confianza para la proporción a partir de la proporción muestral y realizar, si se desea, contrastes relacionados con las proporciones cuando se trabaja con una población.

- Para construir los intervalos de confianza se debe introducir la proporción y el tamaño muestral. Si se desea someter a prueba de hipótesis un valor predeterminado para la proporción poblacional, debe marcarse con una “X” la casilla “Contraste de hipótesis” e introducir la proporción teórica.
- Una vez suministrados los datos al programa, pulsar “Calcular” y se presentará una pantalla de resultados en la que aparecerá la proporción muestral, el intervalo de confianza construido a partir de esta proporción y el nivel de confianza elegido, que por defecto es 95%. La opción de “Contraste de hipótesis” presenta el estadístico evaluado y el valor p o nivel crítico (si no se elige esta opción aparecerán asteriscos).

Ejercicio: 6.2. Para la evaluación de una prueba diagnóstica, se examina 1.000 individuos, de los cuales 700 tienen la enfermedad relacionada con la prueba. Un investigador quiere hacer un contraste de hipótesis, suponiendo como hipótesis nula que la sensibilidad de la prueba es del 80%. ¿Se aceptará dicha hipótesis teniendo en cuenta los datos anteriores y sabiendo que tenemos 100 falsos negativos? Tomar como nivel de significación el 5%.

BIBLIOGRAFÍA

Peña D. *Estadística. Modelos y Métodos 1. Fundamentos*. Madrid: Alianza Universidad Textos; 1993.

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.

Inferencia para la diferencia de medias. Dos muestras

Independientes

Esta opción permite construir los intervalos de confianza para la diferencia de medias de dos poblaciones independientes y realizar, si se desea, contrastes relacionados con dichas diferencias.

- Para la construcción de los intervalos de confianza se dispone de dos poblaciones independientes de las que se introduce la media y el tamaño muestral de cada población, así como las desviaciones estándar de cada una de ellas (indicando previamente si es muestral o poblacional). Si se desea realizar el contraste de hipótesis es necesario marcar una "X" en la casilla correspondiente al contraste, indicando la diferencia entre las medias poblacionales (Población 2-Población 1), de no hacerse así, tomará el valor 0 por defecto.

Población 1		Población 2	
Media muestral:	<input type="text" value="41.500"/>	Media muestral:	<input type="text" value="38.700"/>
Tamaño muestral:	<input type="text" value="416"/>	Tamaño muestral:	<input type="text" value="327"/>
Desviación Estándar			
<input checked="" type="radio"/> Muestral		Población 1:	<input type="text" value="14.580"/>
<input type="radio"/> Poblacional		Población 2:	<input type="text" value="17.200"/>
<input checked="" type="checkbox"/> Contraste de hipótesis			
		Diferencia entre medias poblacionales: <input type="text" value="0.000"/>	
(Población 2 - Población 1)			
<input type="button" value="Calcular"/>		<input type="button" value="Inicializar"/>	
<input type="button" value="Salir"/>		<input type="button" value="F1-Ayuda"/>	

- En la pantalla de resultados se presentan, en la opción del intervalo de confianza, las medias muestrales de ambas poblaciones y el intervalo de confianza construido para la diferencia de medias a partir de un valor p calculado de un contraste de hipótesis de igualdad de varianzas (sólo en el caso de varianzas desconocidas), y el nivel de confianza introducido (por defecto se toma el 95%). La opción de “Contraste de hipótesis” presenta el estadístico evaluado, el valor p y los grados de libertad (si no se selecciona la opción de contraste se visualizan asteriscos).

Inferencia para la media. 2 muestras independiente - Resultados	
Intervalo de confianza	
Media 1:	41.500
Media 2:	38.700
I.C. para la diferencia:	-5.132 -0.467
P-valor (p):	0.0033 para Ho : igualdad de varianzas (Sólo se calcula en el caso de varianzas desconocidas)
Nivel de confianza:	0.950
Contraste de hipótesis	
Ho: $\mu_2 - \mu_1 =$	0.000
Ha: $\mu_2 - \mu_1 \neq$	0.000
Diferencia de medias observada:	-2.800
Estadístico evaluado:	0.718
P-valor (p):	0.0000 g.l.: 0 (Si corresponde)
<input type="button" value="Salir"/> <input type="button" value="Pg-Ant"/> <input type="button" value="Imprimir"/> <input type="button" value="F1-Ayuda"/>	

Ejercicio: 6.3. Volviendo a utilizar los datos del ejercicio 6.1., contrastar la igualdad del índice CAO (CPO)-co de su municipio con los del vecino, si los datos de un estudio realizado en este último son los siguientes:

Tamaño muestral 1.500 niños, índice CAO (CPO)-co 2,89 y desviación estándar 0,89. Nivel de significación 5%.

BIBLIOGRAFÍA

Peña D. Estadística. *Modelos y Métodos I. Fundamentos*. Madrid: Alianza Universidad Textos; 1993.

Dependientes

A diferencia del contraste de medias entre dos muestras independientes, en este caso tenemos n pares de datos tomados en condiciones muy seme-

jantes, con lo que la variabilidad de la diferencia entre sus medias será mucho menor. Esta opción permite construir los intervalos de confianza para la diferencia de medias de dos poblaciones dependientes y realizar, si se desea, contrastes relacionados con dichas diferencias.

- Para construir el intervalo de confianza se dispone de dos muestras dependientes de las que se deben introducir las medias muestrales, las desviaciones estándar y el número de pares muestrales. También es necesario introducir el coeficiente de correlación entre las dos muestras. Si se desea contrastar si la diferencia entre las medias corresponde a un valor predeterminado (Población 2-Población 1) debe marcarse con "X" la opción de "Contraste de hipótesis", introduciendo el valor de la diferencia a comparar entre medias poblacionales (tomando por defecto 0).

Población 1		Población 2	
Media muestral:	<input type="text" value="14.300"/>	Media muestral:	<input type="text" value="17.400"/>
Desv. Estándar:	<input type="text" value="11.400"/>	Desv. Estándar:	<input type="text" value="9.650"/>
Tamaño muestral:	<input type="text" value="217"/>		
Coeficiente de correlación entre las dos muestras: <input type="text" value="0.6800"/>			
<input checked="" type="checkbox"/> Contraste de hipótesis			
Diferencia a contrastar entre medias poblacionales: <input type="text" value="0.000"/> (Población 2 - Población 1)			
<input type="button" value="Calcular"/>		<input type="button" value="Inicializar"/>	
<input type="button" value="Salir"/>		<input type="button" value="F1-Ayuda"/>	

Esta opción se utiliza cuando las muestras dependientes A y B están formadas por los mismos individuos, observados en distintas situaciones, por lo que el tamaño muestral es único. Si se desconoce el coeficiente de

Intervalo de confianza	
Media 1:	<input type="text" value="14.300"/> Media 2: <input type="text" value="17.400"/>
I.C. para la diferencia:	<input type="text" value="-13.703"/> <input type="text" value="19.903"/>
Nivel de confianza:	<input type="text" value="0.950"/>
Contraste de hipótesis	
H ₀ : $\mu_2 - \mu_1 =$	<input type="text" value="0.000"/>
H _a : $\mu_2 - \mu_1 \neq$	<input type="text" value="0.000"/>
Diferencia de medias observada:	<input type="text" value="3.100"/>
Estadístico evaluado:	<input type="text" value="0.361"/>
P-valor (P):	<input type="text" value="0.3588"/> g.l.: <input type="text" value="216"/>
<input type="button" value="Salir"/>	
<input type="button" value="Pg-Ant"/>	
<input type="button" value="Imprimir"/>	
<input type="button" value="F1-Ayuda"/>	

correlación de Pearson no se puede usar este contraste de hipótesis. En ese caso no deberá utilizarse la opción “muestras independientes”, ya que si el coeficiente de correlación fuese negativo se obtendrían intervalos de confianza más pequeños que los reales y que, si fuera positivo, los intervalos de confianza obtenidos serían mayores que los verdaderos.

- En la pantalla de resultados, la opción de intervalos de confianza presenta las medias muestrales de ambas poblaciones y el intervalo de confianza para la diferencia una vez introducido el nivel de confianza correspondiente, que por defecto es 95%. La opción contraste de hipótesis muestra el contraste realizado y el estadístico evaluado, o bien asteriscos en los respectivos campos si no se seleccionó. También se muestra el valor p y los grados de libertad (0 por defecto).

Ejercicio: 6.4. *Para comparar la eficacia de dos medicamentos se administraron éstos a 12 parejas de enfermos, emparejados según la gravedad de su estado. El tiempo medio de recuperación usando el primer medicamento fue de 6 días, y con el segundo 5 días, con desviaciones estándar de 2 y 1,5 respectivamente. Si la covarianza entre las dos muestras es 0,8, ¿existen diferencias significativas entre los tiempos de recuperación de los dos medicamentos?*

BIBLIOGRAFÍA

Peña D. *Estadística. Modelos y Métodos 1. Fundamentos*. Madrid: Alianza Universidad Textos; 1993.

Inferencia para la diferencia de proporciones. Dos muestras Independientes

En el supuesto de dos proporciones hipotéticamente independientes se construyen los intervalos de confianza para la diferencia de proporciones y se presenta la opción de realizar contrastes de hipótesis.

- En el primer caso se necesitan las proporciones muestrales y los tamaños muestrales de ambas poblaciones.
- Si se desea contrastar si la diferencia entre las proporciones es significativamente diferente de un valor predeterminado (Población 2- Población 1), debe marcarse con “X” la opción de “Contraste de hipótesis”, introduciendo el valor estimado para esa diferencia (tomando por defecto 0).

Población 1		Población 2	
Proporción muestral:	<input type="text" value="0.74000"/>	Proporción muestral:	<input type="text" value="0.71000"/>
Tamaño muestral:	<input type="text" value="547"/>	Tamaño muestral:	<input type="text" value="362"/>
<input checked="" type="checkbox"/> Constraste de hipótesis			
Diferencia entre proporciones poblacionales:		<input type="text" value="0.000"/>	
(Población 2 - Población 1)			
<input type="button" value="Calcular"/>		<input type="button" value="Inicializar"/>	
<input type="button" value="Salir"/>		<input type="button" value="F1-Ayuda"/>	

- En la pantalla de resultados, la primera opción presenta las proporciones muestrales de ambas poblaciones y el intervalo de confianza para la diferencia, una vez introducido el nivel de confianza correspondiente (por defecto toma 95%). La segunda opción muestra el contraste realizado, el estadístico evaluado y el valor p . En caso de no haber seleccionado esta opción aparecerán 0 y asteriscos.

Intervalo de confianza			
Proporción muestral 1:	<input type="text" value="0.74000"/>		
Proporción muestral 2:	<input type="text" value="0.71000"/>		
I.C. para la diferencia:	<input type="text" value="-0.091"/>	<input type="text" value="0.031"/>	
(Población 2 - Población 1)			
Nivel de confianza:	<input type="text" value="0.950"/>		
Contraste de hipótesis			
H ₀ : P2-P1=	<input type="text" value="0.00000"/>		
H _a : P2-P1<>	<input type="text" value="0.00000"/>		
Diferencia de proporciones observadas: <input type="text" value="-0.03000"/>			
Estadístico evaluado: <input type="text" value="-1.064"/> P-valor (p): <input type="text" value="0.2871"/>			
<input type="button" value="Salir"/>		<input type="button" value="Pg-Ant"/>	
<input type="button" value="Imprimir"/>		<input type="button" value="F1-Ayuda"/>	

Ejercicio: 6.5. En el hospital A, 37 miembros del personal de un total de 135 son mujeres. En el hospital B, de un total de 234 miembros, 53 son mujeres. ¿Existen diferencias significativas entre el porcentaje de miembros mujeres de los dos hospitales? Utilizar como nivel de significación 5%.

BIBLIOGRAFÍA

Peña D. *Estadística. Modelos y Métodos 1. Fundamentos*. Madrid: Alianza Universidad Textos; 1993.

Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.



Dependientes

Opción destinada a construir intervalos de confianza para la diferencia de proporciones y/o contrastar la hipótesis según la cual la diferencia entre las proporciones poblacionales es significativamente diferente de un valor predeterminado. El estudio es **pareado** y se realiza en **distintos tiempos**; por tanto se trata de muestras dependientes.

- Se dispone de una muestra pareada (mismos individuos en distintos tiempos) de la que se deben introducir los siguientes datos: los casos positivo-positivo, positivo-negativo, negativo-positivo y negativo-negativo; si se desea realizar contraste de hipótesis debe marcarse con "X" esta opción e introducir las proporciones esperadas.

Inferencia para la proporción. 2 muestras dependientes			
	Tiempo 1	Tiempo 2	
Pares	+	+	45
	+	-	23
	-	+	7
	-	-	33
<input checked="" type="checkbox"/> Contraste de hipótesis			
Proporción esperada en el Tiempo 1:			0.25000
Proporción esperada en el Tiempo 2:			0.35000
<input type="button" value="Calcular"/> <input type="button" value="Inicializar"/> <input type="button" value="Salir"/> <input type="button" value="F1-Ayuda"/>			

- En la pantalla de resultados, la primera opción presenta las proporciones de pares positivos en el tiempo 1 y pares positivos en tiempo 2,

Inferencia para la proporción. 2 muestras dependientes. Resultados	
Intervalo de confianza	
Proporción de pares + en 1:	0.629
Proporción de pares + en 2:	0.481
I.C. para la diferencia:	-0.252 -0.043
(+ tiempo 2) - (+ tiempo en 1)	
Nivel de confianza:	0.950
Contraste de hipótesis	
H ₀ : P2-P1=	0.10000
H _a : P2-P1<>	0.10000
Diferencia de proporciones observadas:	-0.148
Estadístico evaluado:	-1.179
P-valor (p):	0.238
<input type="button" value="Salir"/> <input type="button" value="Pg-Ant"/> <input type="button" value="Imprimir"/> <input type="button" value="F1-Ayuda"/>	



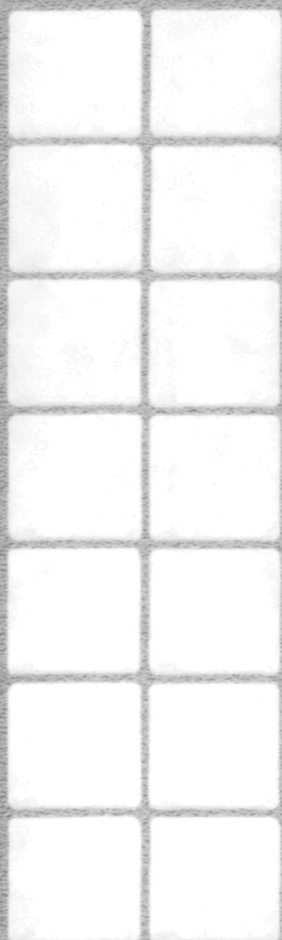
y el intervalo de confianza para la diferencia de pares positivos entre ambos tiempos, una vez introducido el nivel de confianza correspondiente (por defecto 95%). La segunda opción muestra el contraste realizado, la diferencia de proporciones observada, el estadístico evaluado y el valor p , presentando 0 y asteriscos en el caso de no seleccionar esta opción.

Ejercicio: 6.6. De un total de 300 enfermos se eligieron 100 para comprobar la eficacia de un tratamiento. Entre estos pacientes elegidos 65 estaban en un estado grave de la enfermedad antes de iniciarse el tratamiento, de los cuales 45 pasaron a un estado más leve una vez finalizado este, entre los 35 que estaban en estado leve, 5 pasaron a un estado más grave. Entre los 100 pacientes elegidos al azar, ¿existe diferencia significativa entre los porcentajes de enfermos graves y leves antes y después del tratamiento? Construir un intervalo de confianza para la diferencia de proporciones, con un nivel de confianza del 90%.

BIBLIOGRAFÍA

Peña D. Estadística. *Modelos y Métodos 1. Fundamentos*. Madrid: Alianza Universidad textos; 1993.

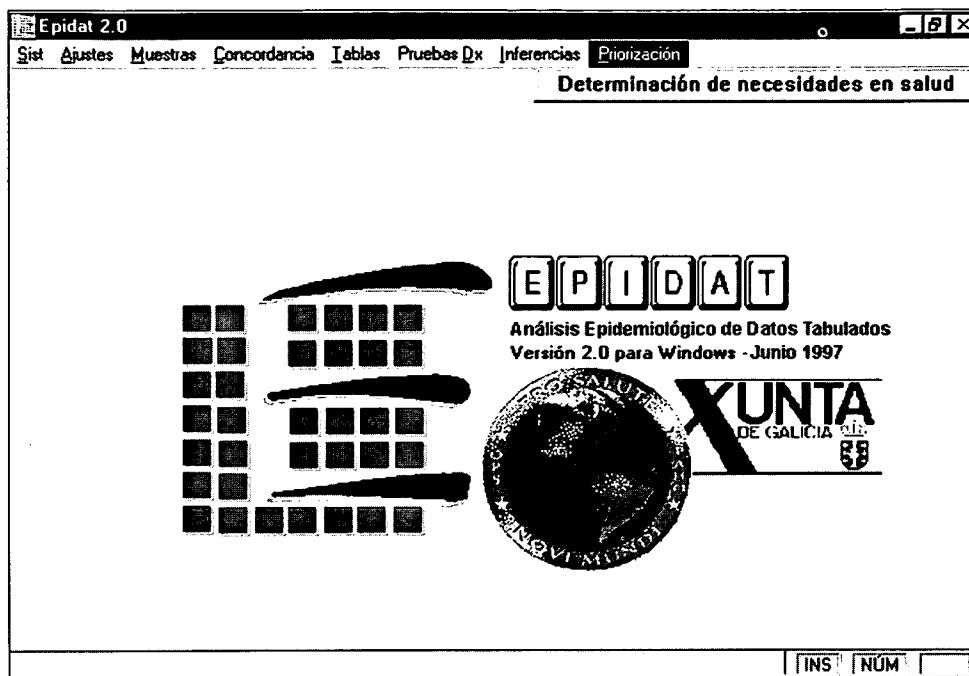
Fleiss J.L. *Statistical methods for rates and proportions*. New York: John Wiley & Sons; 1981.



PRIORIZACIÓN



Determinación de necesidades en salud



Se presenta un método simple de determinación de necesidades en salud. Para utilizarlo se necesita definir una serie de **unidades de análisis** (países, regiones, provincias, áreas...) que se desea ordenar en base a necesidades definidas por una serie de **indicadores** o variables de salud o enfermedad seleccionados por el usuario.

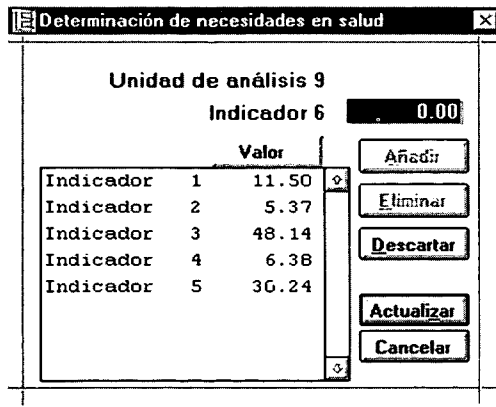
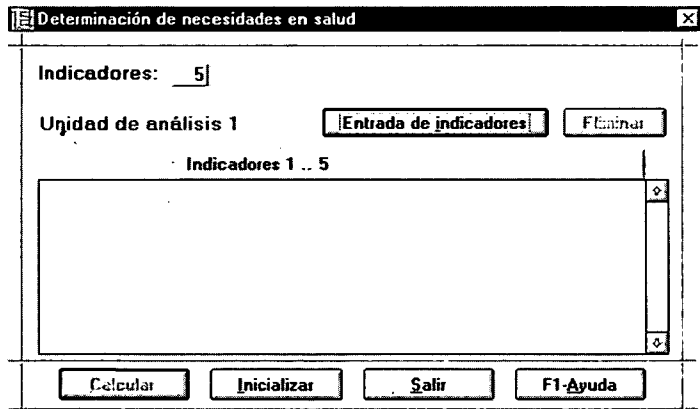
El método asigna puntuaciones positivas y negativas a cada unidad de análisis. En el caso de que los valores elevados de los indicadores seleccionados impliquen peor situación sanitaria (p. e. mortalidad perinatal), los valores positivos de los resultados del índice señalarían mayor necesidad y los negativos menor. En caso contrario (p. e. camas x 1.000 habitantes) los valores positivos señalarían comunidades con menor necesidad.

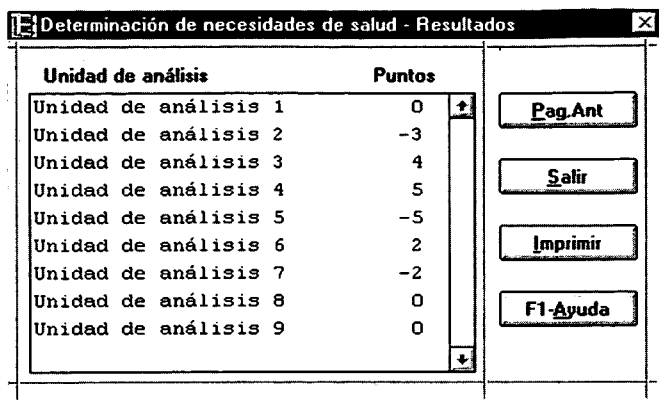
Advertencia: Es imprescindible tener en cuenta ciertos aspectos a la hora de utilizar este método:

- 1º La selección de indicadores es empírica y depende del criterio del investigador. Deben ser sólidos y reflejar condiciones de salud-enfermedad.

- 2º Los indicadores deben tener el mismo sentido, es decir los valores elevados han de tener una misma interpretación, o indicar buena salud o mala. No se pueden mezclar indicadores como mortalidad infantil en el que una tasa elevada denota un problema de salud, con el número de médicos por 1.000 habitantes, que cuanto mayor sea, presupone una mejor situación sanitaria.
- 3º Ante la presencia de "valores aberrantes" en algún indicador y unidad de análisis, dada su influencia en el resultado final, se propone sustituirlo por el valor máximo o mínimo, según el caso, considerado aceptable.

En primer lugar teclee el número de indicadores que va a utilizar. A continuación pulse "Entrada de indicadores" y pasará a la segunda pantalla en la que deberá teclear, para la unidad de análisis 1, los valores de cada indicador seleccionado. Una vez introducido el último, pulsar "Actualizar" y regresará a la pantalla anterior. Repitiendo la operación introduzca los indicadores de todas las unidades de análisis en la que desee priorizar. Al final del proceso pulse "Calcular". Si desea modificar algún dato, marque con el ratón la fila de la unidad de análisis que contenga el dato, haga doble "click" y el programa le presentará la pantalla donde podrá modificar los datos que precise.



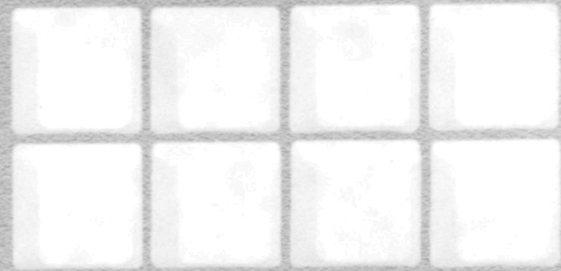
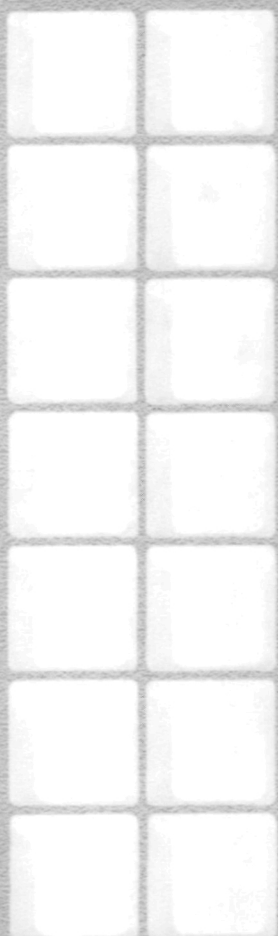


BIBLIOGRAFÍA

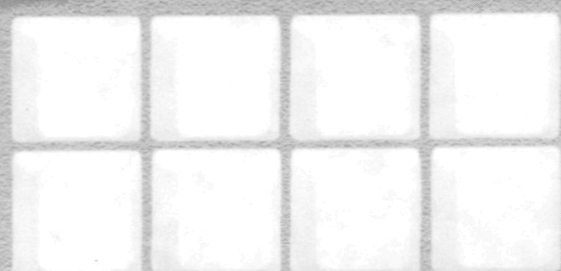
Castillo-Salgado C. *Los Servicios de Salud en las Américas: Análisis de Indicadores Básicos*. Cuaderno Técnico nº 14. Organización Panamericana de la Salud. Washington DC, 1988; 147-152, 221-230.

Ejercicio: 7.1. En una región de un país existen 9 áreas administrativas y usted desea establecer prioridades con el objeto de asignar recursos, para lo que dispone de una serie de indicadores, que se especifican a continuación. Seleccione 5 indicadores y, utilizando la opción de "Determinación de necesidades en salud" de EPIDAT, señale las áreas dónde considera prioritario invertir recursos en función de una mayor necesidad. Justifique los indicadores seleccionados.

Área	mortalidad infantil (x 1.000 nv)	esperanza de vida al nacer	años de vida perdidos (x 1.000) en < 55 años	personas con limitaciones permanentes (x 100)	TBC (x 10 ³)	odólogos (x 10.000)	% hogares sin servicios higiénicos en vivienda
1	15,18	74,46	45,41	11,22	16,89	1,45	11,16
2	12,78	76,54	39,44	6,90	27,79	1,95	4,36
3	17,,29	74,81	46,40	6,46	31,96	1,98	9,93
4	15,94	75,76	49,29	9,74	21,52	1,74	7,72
5	12,66	74,24	46,95	13,39	30,68	1,45	4,54
6	12,51	75,46	40,90	14,39	18,82	1,94	9,81
7	12,61	75,60	43,57	6,24	35,58	1,37	16,65
8	11,37	76,34	42,80	7,96	20,27	1,00	18,63
9	13,78	76,88	34,92	9,38	20,87	2,01	3,07



Anexo I



intervalos de confianza “Margen de valores de una variable de interés, por ejemplo, una tasa, dispuestos de tal modo que dicho margen tiene una determinada probabilidad (confianza) de incluir el verdadero valor de la variable. La probabilidad especificada se denomina nivel de confianza, y los puntos extremos del intervalo, límites de confianza”.

Last, J.M. *Diccionario de epidemiología*. Barcelona: Salvat Editores; 1989.

NOTA: En algunas ocasiones, dependiendo de los valores de las celdas, los intervalos de confianza pueden tomar valores imposibles como, por ejemplo, límites de confianza inferiores de odds ratios con valores menores de 0, o límites superiores de pruebas diagnósticas mayores de 100. Esto no constituye un error del programa. Se debe al proceso de cálculo y debe interpretarse como que el intervalo de confianza correspondiente toma el valor límite.

nivel de confianza Es la probabilidad de aceptar la hipótesis nula cuando es cierta.

EPIDAT utiliza por defecto un nivel de confianza, $1-\alpha$, de 0,950 (95%). Usted puede utilizar otro diferente y para ello debe pulsar repetidamente el botón izquierdo del ratón con el cursor situado en las flechas de la ventana “nivel de confianza”. Este aumenta de 0,005 en 0,005 al pulsar sobre la flecha superior y disminuye al hacerlo sobre la inferior. Una vez visualizado el nivel de confianza que se desea, debe confirmarse pulsando “enter” en el teclado. EPIDAT recalcula los valores que dependen del nivel de confianza elegido.

potencia

Capacidad de un estudio para demostrar significación estadística, cuando existe una diferencia o una asociación verdadera de una fuerza determinada en la población de la que se ha extraído la muestra (sinónimo: poder estadístico, poder de resolución).

Riegelman, R.K., Hirsch R.P. *Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica*. O.P.S. 1992; Publicación Científica Nº 531.

error de muestreo El error de muestreo o aleatorio “es aquella porción de variación en una medida que no tiene relación aparente con ninguna otra medida o variable, y que generalmente se considera debida al azar”.

Last, J.M. *Diccionario de epidemiología*. Barcelona: Salvat Editores; 1989.

Se denomina error de muestreo a aquel “que se comete debido al hecho de que se sacan conclusiones sobre cierta realidad, a partir de la observación de sólo una parte de ella”.

Silva, L.C. *Muestreo para la investigación en ciencias de la salud*. Madrid: Ediciones Díaz de Santos; 1993.

valor p

“Probabilidad de que las cifras estadísticas de una prueba tengan valores iguales a los observados, o más extremos que ellos, si la hipótesis

nula fuera verdadera. La letra p , seguida de la abreviatura n.s. (no significativa) o del símbolo $<$ (inferior a) y una cifra decimal, tal como 0,01 ó 0,05, es una constatación de la probabilidad de que la diferencia observada haya ocurrido puramente por azar, si los grupos son realmente semejantes, es decir, bajo la hipótesis nula.

Aunque los investigadores pueden fijar arbitrariamente sus propios niveles de significación, en la mayoría de los trabajos médicos y epidemiológicos se considera que es suficientemente improbable que el resultado de una investigación cuyos valores de probabilidad sean menores del 5% ($p < 0,05$) ó del 1% ($p < 0,01$) haya ocurrido por azar, por lo que se justifica la designación de “estadísticamente significativo”.

Last, J.M. *Diccionario de epidemiología*. Barcelona: Salvat Editores; 1989.

prueba bilateral “Prueba de significación estadística en la que se toman en cuenta las desviaciones de la hipótesis nula en cualquier dirección. El uso de una prueba bilateral implica que el investigador deseaba considerar las desviaciones en cualquier dirección antes de recoger los datos”.

Riegelman, R.K., Hirsch, R.P. *Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica*. O.P.S. 1992; Publicación Científica Nº 531.

prueba unilateral “Prueba de significación estadística en la cual sólo se toman en cuenta las desviaciones respecto a la hipótesis nula en una sola dirección. El empleo de una prueba unilateral implica que el investigador no considera posible una desviación verdadera en dirección opuesta”.

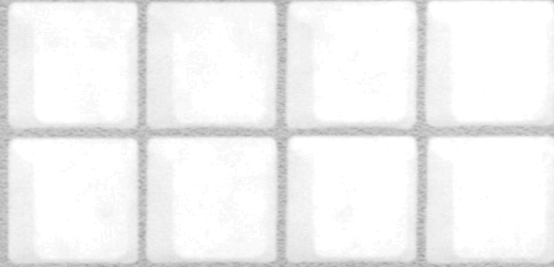
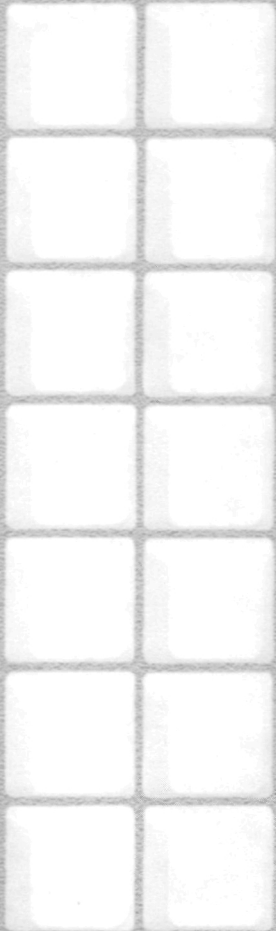
Riegelman, R.K., Hirsch, R.P. *Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica*. O.P.S. 1992; Publicación Científica Nº 531.

contraste de hipótesis Contraste para decidir si una determinada hipótesis, afirmación o conjetura, sobre la distribución poblacional estudiada, es confirmada o invalidada estadísticamente a partir de las observaciones contenidas en una muestra.

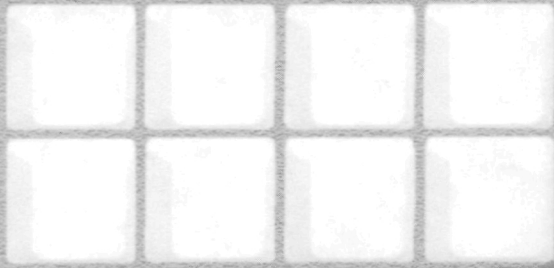
intervalo de confianza Bootstrap Técnica para calcular la varianza y el sesgo de un método de estimación, mediante la extracción repetida de muestras, al azar, con sustitución a partir de las observaciones que se poseen. Se aplica el método de estimación a cada muestra extraída, con lo que se obtiene una serie de estimaciones. La varianza observada de esta serie es la estimación “bootstrap” de la varianza. La diferencia entre el promedio de las estimaciones de la serie y la estimación original es la estimación “bootstrap” del sesgo.

Last, J.M. *Diccionario de epidemiología*. Barcelona: Salvat Editores; 1989.

EPIDAT permite realizar entre 400 y 2.000 iteraciones.



Anexo II



Módulo 1: ajuste de tasas

Método directo:

Tasa estandarizada:

$$test = \sum_{j=1}^m t_j \frac{p_{ej}}{p_e} \quad (1.1)$$

Varianza de la tasa estandarizada:

$$var(test) = \sum_{j=1}^m \frac{t_j}{p_j} \left(\frac{p_{ej}}{p_e} \right)^2 \quad (1.2)$$

Intervalo de confianza de la tasa estandarizada:

$$\left(test - z_{\alpha/2} \sqrt{var(test)}, \quad test + z_{\alpha/2} \sqrt{var(test)} \right) \quad (1.3)$$

Donde

m es el número de intervalos de edad,

p_j es la población en cada grupo de edad ($j=1, \dots, m$),

p_{ej} es la población estándar en cada grupo de edad ($j=1, \dots, m$),

p_e es la población estándar total,

$$p_e = \sum_{j=1}^m p_{ej}$$

t_j es la tasa específica en cada grupo de edad ($j=1, \dots, m$),

z_{α} es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1 - \alpha) \%$ es el nivel de confianza.

Método indirecto:

Índice de mortalidad estandarizada (IME, SMR):

$$IME = \frac{\sum_{j=1}^m t_j p_j}{\sum_{j=1}^m t_{ej} p_j} = \frac{CO}{CE} \quad (1.4)$$

Varianza del IME:

$$\text{var}(IME) = \sum_{j=1}^m \left(\frac{p_j}{t_j p_j} \right)^2 \text{var}(t_j) = \frac{CO}{\left(\sum_{j=1}^m t_{ej} p_j \right)^2} \quad (1.5)$$

Intervalo de confianza del IME:

$$\left(IME - z_{\alpha/2} \sqrt{\text{var}(IME)}, IME + z_{\alpha/2} \sqrt{\text{var}(IME)} \right) \quad (1.6)$$

Donde

m es el número de intervalos de edad,

p_j es la población en cada grupo de edad ($j=1, \dots, m$),

p_{ej} es la población estándar en cada grupo de edad ($j=1, \dots, m$),

$$p_e = \sum_{j=1}^m p_{ej}$$

p_e es la población estándar total,

t_j es la tasa específica en cada grupo de edad ($j=1, \dots, m$),

t_{ej} es la tasa estándar en cada grupo de edad ($j=1, \dots, m$),

CO son los casos observados,

z_{α} es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1-\alpha)\%$ es el nivel de confianza.

Módulo 2: muestras

Situación con una muestra. Proporción:

Cálculo de una proporción. Tamaño muestral mínimo si N es desconocido (N=0):

$$n = \frac{z_{1-\alpha/2}^2 P(1-P)}{d^2} \quad (2.1)$$

Tamaño muestral mínimo si N es conocido :

$$n = \frac{N z_{1-\alpha/2}^2 P(1-P)}{(N-1)d^2 + z_{1-\alpha/2}^2 P(1-P)} \quad (2.2)$$

Prueba de hipótesis. Tamaño muestral mínimo:

$$n = \frac{\left(z_{1-\alpha/2} \sqrt{P_0(1-P_0)} + z_{1-\beta} \sqrt{P(1-P)} \right)^2}{(P_0 - P)^2} \quad (2.3)$$

Donde

N es el tamaño poblacional,

P es la proporción esperada de la población,

P_0 es la proporción de la hipótesis nula,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1-\alpha)\%$ es el nivel de confianza,

$100 \times (1-\beta)\%$ es la potencia de la prueba,

d es la precisión absoluta requerida a ambos lados de la proporción.

Situación con una muestra. Media:

Cálculo de una media. Tamaño muestral mínimo si N es desconocido (N=0):

$$n = \frac{z_{1-\alpha/2}^2 s d^2}{d^2} \quad (2.4)$$

Tamaño muestral mínimo si N es conocido:

$$n = \frac{N z_{1-\alpha/2}^2 V_x^2}{(N-1)d^2 + z_{1-\alpha/2}^2 V_x^2} \quad (2.5)$$

Donde

N es el tamaño poblacional,

V_x es el coeficiente de variación,

$$V_x = \frac{sd}{\bar{X}}$$

sd es la desviación estándar esperada en la población,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,



$100 \times (1 - \alpha) \%$ es el nivel de confianza,

d es la precisión absoluta requerida a ambos lados de la proporción.

Situación con dos muestras. Proporción:

Cálculo de una diferencia. Tamaño muestral mínimo:

$$n = \frac{z_{1-\alpha/2}^2 (P_1(1-P_1) + P_2(1-P_2))}{d^2} \quad (2.6)$$

Prueba de hipótesis. Tamaño muestral mínimo si P_1 y P_2 toman valores entre 0,001 y 0,05:

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{2(\arcsin\sqrt{P_2} - \arcsin\sqrt{P_1})^2} \quad (2.7)$$

En cualquier otro caso

$$n = \frac{(z_{1-\alpha/2}\sqrt{2\bar{P}(1-\bar{P})} + z_{1-\beta}\sqrt{P_1(1-P_1) + P_2(1-P_2)})^2}{(P_1 - P_2)^2} \quad (2.8)$$

Donde

P_1 es la proporción esperada de la población 1,

P_2 es la proporción esperada de la población 2,

\bar{P} es la media de las proporciones anteriores,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1 - \alpha) \%$ es el nivel de confianza,

$100 \times (1 - \beta) \%$ es la potencia de la prueba,

d es la precisión absoluta

Situación con dos muestras. Media:

Muestras independientes. Cálculo de una diferencia. Tamaño muestral mínimo si las muestras son independientes:

$$n = \frac{z_{1-\alpha/2}^2 (sd_1^2 + sd_2^2)}{d^2} \quad (2.9)$$

Muestras dependientes. Prueba de hipótesis. Tamaño muestral mínimo si las muestras son dependientes:

$$n = \frac{sd^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{dif^2} \quad (2.10)$$

Donde

sd_1 es la desviación estándar esperada de la población 1,

sd_2 es la desviación estándar esperada de la población 2,

$$sd^2 = sd_1^2 + sd_2^2 - 2r_{12}sd_1sd_2$$

r_{12} es el coeficiente de correlación entre ambas poblaciones,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1 - \alpha) \%$ es el nivel de confianza,

100x(1-β)% es la potencia de la prueba,
d es la precisión absoluta,
dif es la diferencia entre las medias.

Estudios casos–controles:

Pareado. Mínimo número de pares discordantes requerido:

$$m = \frac{\left(\frac{z_{\alpha/2}}{2} + z_{\beta} \sqrt{P(1-P)} \right)^2}{\left(P - \frac{1}{2} \right)^2} \quad (2.11)$$

Número de pares requeridos para detectar *m* pares discordantes:

$$M = m / P_e \quad (2.12)$$

No pareado. Número de casos que se toman:

$$n = \frac{\left(z_{1-\alpha/2} \sqrt{(1+c) \bar{P} \bar{Q}} + z_{1-\beta} \sqrt{P_1 Q_1 + P_2 Q_2 c} \right)^2}{(P_1 - P_2)^2} \quad (2.13)$$

Número de controles: *c x n*

Donde

c es el número de controles por caso,

P₁ es la proporción de casos expuestos,

P₂ es la proporción de controles expuestos,

$$P_1 = P_2 OR / (1 + (OR - 1))$$

OR es la odds ratio,

$$P_c = (P_2 Q_1 + P_1 Q_2)$$

$$\bar{P} = 1 / (2(P_1 + P_2))$$

$$Q_1 = 1 - P_1$$

$$Q_2 = 1 - P_2$$

$$\bar{Q} = 1 - P$$

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

100x(1-α)% es el nivel de confianza,

100x(1-β)% es la potencia de la prueba.

Estudio de cohortes:

Estimación del riesgo relativo. Tamaño muestral mínimo para cada cohorte:

$$n = \frac{z_{1-\alpha/2}^2 \left[\frac{1 - P_1}{P_1} + \frac{1 - P_2}{P_2} \right]}{\log(1 - \epsilon)^2} \quad (2.14)$$

Donde

P₁ es la proporción de expuestos con la enfermedad,

P₂ es la proporción de no expuestos con la enfermedad,

ε es la precisión relativa,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100x(1-\alpha)\%$ es el nivel de confianza,

$100x(1-\beta)\%$ es la potencia de la prueba.

Verificación de la calidad de lotes

Aceptación de una prevalencia. En el caso de población finita, el tamaño muestral se obtiene resolviendo la desigualdad:

$$\sum_{x=0}^{d^*} M C_x^{(N-M)} C_{(n-x)}^N / C_n^N < \alpha \quad (2.15)$$

En el caso de población infinita:

$$\sum_{d=0}^{d^*} n C_d P^d (1-P)^{n-d} < \alpha \quad (2.16)$$

Rechazo de un lote. Tamaño muestral:

$$n = \left[z_{1-\alpha} \sqrt{P_0(1-P_0)} + z_{1-\beta} \sqrt{P(1-P)} \right]^2 / (P_0 - P)^2 \quad (2.17)$$

Donde

N es el tamaño de la población,

$$M = NxP$$

P es el valor previsto de la proporción en la población,

P_0 es el valor de la proporción de la población sometido a prueba en caso de hipótesis nula,

d^* es el número máximo de sujetos de la muestra que presentan la característica,

en el caso de RECHAZO DE UN LOTE:

$$d^* = \left[n P_0 - z_{1-\alpha} \sqrt{n P_0 (1 - P_0)} \right]$$

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100x(1-\alpha)\%$ es el nivel de confianza,

$100x(1-\beta)\%$ es la potencia de la prueba.

Módulo 3: concordancia

Dos observadores

Dos categorías

+	a	-	b
-	c		d

Denotando por $n = a+b+c+d$

CONCORDANCIA BRUTA:

$$\frac{a+d}{n} \quad (3.1)$$

J DE YODEN:

$$\frac{a}{a+c} + \frac{d}{b+c} - 1 \quad (3.2)$$

desviación estándar de J de Youden:

$$sd(J) = \sqrt{\frac{ac}{(a+c)^3} + \frac{bd}{(b+c)^3}} \quad (3.3)$$

Intervalo de confianza de J de Youden:

$$\left(J - z_{1-\alpha/2} sd(J), \quad J + z_{1-\alpha/2} sd(J) \right) \quad (3.4)$$

KAPPA (K):

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (3.5)$$

desviación estándar de K:

$$sd(K) = \frac{\sqrt{A+B-C}}{(1-P_e)\sqrt{n}} \quad (3.6)$$

Intervalo de confianza de K:

$$\left(K - z_{1-\alpha/2} sd(K), \quad K + z_{1-\alpha/2} sd(K) \right) \quad (3.7)$$

Donde

$$P_0 = \frac{a+d}{n}$$

$$P_e = \frac{(a+b)(a+c) + (c+d)(b+d)}{n^2}$$

$$A = \sum_{j=1}^2 P_{.j} [(1 - P_{.j} + P_{.j}) (1 - K)]^2$$

$$B = (1 - K)^2 \sum_{i,j} P_{ij} (P_{.i} + P_{.j})^2$$

$$C = [K - P_e (1 - K)]^2$$

$P_{.i}$ y $P_{.j}$ son los valores marginales de la tabla,

$$P_{ij} = \frac{x_{ij}}{n}$$

x_{ij} es el dato de la i-ésima fila y la j-ésima columna,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,
 $100x(1-\alpha)\%$ es el nivel de confianza.

Para contrastar que $K=0$ (la concordancia se debe al azar), se supone la normalidad asintótica y se calcula el valor p de

$$z = \frac{K}{sd_0(K)} \quad (3.8)$$

desviación estándar de K para contrastar $H_0: K=0$:

$$sd_0(K) = \frac{1}{(1-P_e)\sqrt{n}} \sqrt{P_e + P_e^2 - \sum_{i=1}^2 P_{i\bullet} P_{\bullet i} (P_{i\bullet} + P_{\bullet i})} \quad (3.9)$$

Múltiples categorías. KAPPA (K):

$$\frac{P_0 - P_e}{1 - P_e} \quad (3.10)$$

Desviación estándar de K :

$$sd(K) = \frac{1}{(1 - P_e)\sqrt{n}} \sqrt{\sum_{i=1}^n \sum_{j=1}^n P_{ij} [w_{ij} - (\bar{w}_{i\bullet} + \bar{w}_{\bullet j})(1 - K)]^2 - [K - P_e(1 - K)]^2} \quad (3.11)$$

Intervalo de confianza de K :

$$(K - z_{1-\alpha/2} sd(K), K + z_{1-\alpha/2} sd(K)) \quad (3.12)$$

Donde

$$P_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} P_{ij}$$

$$P_e = \sum_{i=1}^n \sum_{j=1}^n w_{ij} P_{i\bullet} P_{\bullet j}$$

$$P_{ij} = \frac{x_{ij}}{n}$$

n es la suma de todas las celdas

w_{ij} es la ponderación de la i -ésima fila y la j -ésima columna,

x_{ij} es el dato de la i -ésima fila y la j -ésima columna,

$P_{i\bullet}$ y $P_{\bullet i}$ son los valores marginales de la tabla,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100x(1-\alpha)\%$ es el nivel de confianza.

Para contrastar que $K=0$ (la concordancia se debe al azar), se supone la normalidad asintótica y se calcula el valor de p de

$$z = \frac{K}{sd_0(K)} \quad (3.13)$$

$$sd_0(K) = \frac{1}{(1 - P_e)\sqrt{n}} \sqrt{\sum_{i=1}^n \sum_{j=1}^n P_{i\bullet} P_{\bullet j} [w_{ij} - (\bar{w}_{i\bullet} + \bar{w}_{\bullet j})]^2 - P_e^2} \quad (3.14)$$

Múltiples observadores:

Dos categorías. (Kappa):

$$\hat{\kappa} = 1 - \frac{\sum_i x_i (m_i - x_i)}{n(\bar{m} - 1)\bar{p}\bar{q}} \quad (3.15)$$

desviación estándar para contrastar la hipótesis $H_0: K=0$:

$$sd(\hat{\kappa}) \cong \frac{\sqrt{2(\bar{m}_{11} - 1) + \frac{(\bar{m} - \bar{m}_{11})(1 - 4\bar{p}\bar{q})}{\bar{m}\bar{p}\bar{q}}}}{(\bar{m} - 1)\sqrt{n\bar{m}_{11}}} \quad (3.16)$$

Donde

n es el número de sujetos,

m_i es el número de clasificadores del sujeto i -ésimo,

x_i es el número de clasificaciones positivas (de definición arbitraria) del sujeto i -ésimo,

$$\begin{aligned} \bar{m} &= \sum_i m_i / n \\ \bar{p} &= \sum_i x_i / (n\bar{m}) \\ \bar{q} &= 1 - \bar{p} \\ \bar{m}_{11} &= \frac{n}{\sum_i 1/m_i} \end{aligned}$$

Múltiples categorías. Valor de kappa para la categoría j :

$$\hat{\kappa}_j = 1 - \frac{\sum_i x_{ij}(m - x_{ij})}{nm(m-1)\bar{p}_j\bar{q}_j} \quad (3.17)$$

Valor de kappa global:

$$\hat{\kappa} = 1 - \frac{nm^2 - \sum_i \sum_j x_{ij}^2}{nm(m-1)\sum_j \bar{p}_j\bar{q}_j}$$

Desviaciones estándar para contrastar la hipótesis $H_0: K=0$:

$$sd(\hat{\kappa}_j) = \sqrt{\frac{2}{nm(m-1)}} \quad (3.19)$$

$$sd(\hat{\kappa}) = \sqrt{\frac{2}{nm(m-1)} \frac{(\sum_j \bar{p}_j\bar{q}_j)^2 - \sum_j \bar{p}_j\bar{q}_j(\bar{q}_j - \bar{p}_j)}{\sum_j \bar{p}_j\bar{q}_j}} \quad (3.20)$$

Donde

n es el número de sujetos,

k es el número de categorías,

m es el número de observadores por sujeto (cte.),

x_{ij} es el número de clasificaciones del sujeto i en la categoría j ,

$$\sum_j x_{ij} = m \quad \forall i$$

\bar{P}_j es la proporción total de observadores en la categoría j ,

$$\bar{q}_j = 1 - \bar{p}_j$$

Hipótesis de igualdad de Kappas:

Test χ^2 para la prueba de homogeneidad:

$$\chi_{igual} = \sum_{i=1}^g \frac{(k_i - k_{global})^2}{\text{var}(k_i)} \quad (3.21)$$

Intervalo de confianza bilateral del 95%:

$$\left(k_{global} - z_{1-\alpha/2} \sqrt{\frac{1}{\sum_{i=1}^g \frac{1}{\text{var}(k_i)}}}, k_{global} + z_{1-\alpha/2} \sqrt{\frac{1}{\sum_{i=1}^g \frac{1}{\text{var}(k_i)}}} \right) \quad (3.22)$$

Donde

k_{global} es la kappa global,

z_{α} es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1-\alpha)\%$ es el nivel de confianza.

Módulo 4: Tablas

Tablas 2x2:

	Enfermos	Sanos
Fr +	a	b
Fr -	c	d

Denotando por $n = a+b+c+d$.

ESTIMADOR DE LA ODDS-RATIO (razón de ventajas o producto cruzado) :

$$OR = \frac{ad}{bc} \quad (4.1)$$

Estimador de la varianza de $\text{Ln}(OR)$:

$$\hat{Var}(\text{Ln}(OR)) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \quad (4.2)$$

Intervalo de confianza utilizando el estimador de $\text{Var}(\text{Ln}(OR))$ propuesto por Wolf:

$$\left(OR \text{Exp}\left(-z_{1-\alpha/2} \sqrt{\hat{Var}(\text{Ln}(OR))}\right), OR \text{Exp}\left(z_{1-\alpha/2} \sqrt{\hat{Var}(\text{Ln}(OR))}\right) \right) \quad (4.3)$$

Para el intervalo de confianza utilizando el método de Cornfield ver la aproximación descrita por Fleiss.

ESTIMADOR DEL RIESGO RELATIVO, se obtiene a partir de:

$$RR = \frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}} \quad (4.4)$$

Estimador de la varianza de $\text{Ln}(RR)$:

$$\hat{Var}(\text{Ln}(RR)) = \frac{b}{a(a+b)} + \frac{d}{c(c+d)} \quad (4.5)$$

Intervalo de confianza utilizando el estimador de $\text{Var}(\text{Ln}(RR))$ propuesto por Katz:

$$\left(RR \text{Exp}\left(-z_{1-\alpha/2} \sqrt{\hat{Var}(\text{Ln}(RR))}\right), RR \text{Exp}\left(z_{1-\alpha/2} \sqrt{\hat{Var}(\text{Ln}(RR))}\right) \right) \quad (4.6)$$

Donde

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

100x(1- α)% es el nivel de confianza.

MEDIDAS DE ASOCIACIÓN: las medidas de asociación utilizadas están basadas en la distribución χ^2 (1 g.l.). Para utilizar las pruebas Chi-Cuadrado, sin y con corrección de Yates, y la Chi-Cuadrado de Mantel-Haenszel, se utilizan las fórmulas:

$$\chi^2 = \frac{n(ad - cb)^2}{(a+b)(a+c)(c+d)(b+d)} \quad (4.7)$$

$$\chi^2_{\text{Yates}} = \frac{n\left(|ad - cb| - \frac{n}{2}\right)^2}{(a+b)(a+c)(c+d)(b+d)} \quad (4.8)$$

$$\chi^2_{M-ii} = \frac{(n-1)(ad-cb)^2}{(a+b)(a+c)(c+d)(b+d)} \quad (4.9)$$

Los valores p correspondientes (una c6la) se obtienen a partir de la distribuci6n χ^2 con 1 grado de libertad.

TEST EXACTO DE FISHER: la probabilidad de una tabla 2x2 bajo la suposici6n de independencia entre filas y columnas (fijando los totales marginales) se obtiene a partir de la distribuci6n hipergeom6trica:

$$P(\bar{a}) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} \quad (4.10)$$

- (Si se fijan los marginales la configuraci6n est1 caracterizada por el valor de a).

Para el c1culo de las probabilidades de otras configuraci6nes distintas se puede utilizar:

$$\bar{P}(a+1) = \frac{bc}{(a+1)(d+1)} P(a) \quad (4.11)$$

y

$$\bar{P}(a-1) = \frac{ad}{(c+1)(b+1)} P(a) \quad (4.12)$$

Ent6nces si:

$$P(+) = \sum_{P(a+k) \leq P(a)} P(a+k)$$

y

$$P(-) = \sum_{P(a-k) \leq P(a)} P(a-k)$$

$$P(\text{Min}) = \begin{cases} P(+) & \text{si } \frac{ad}{(b+1)(c+1)} > \frac{bc}{(a+1)(d+1)} \\ P(=) & \text{en caso contrario} \end{cases} \quad (4.13)$$

las probabilidades correspondientes a una y dos c6las vienen dadas por:

$$P1C6la = P(\text{Min}) + P(a) \quad (4.14)$$

$$P2C6las = P(-) + P(a) + P(+) \quad (4.15)$$

Caso-control. RIESGO ATRIBUIBLE POBLACIONAL:

$$\bar{P}AR = \frac{ad-bc}{d(a+c)} \quad (4.16)$$

Estimador de la varianza de $\bar{P}AR$:

$$PAR = \frac{P(RR=1)}{1+P(RR=1)} \quad (4.17)$$

Cohortes. RIESGO ATRIBUIBLE POBLACIONAL:

$$\bar{P}AR = \frac{P(RR=1)}{1+P(RR=1)} \quad (4.18)$$

Estimador de la varianza de $\bar{P}AR$:

$$\hat{Var}(PAR) = \frac{cn[ad(n-c) + bc^2]}{(a+c)^3(c+d)^3} \quad (4.19)$$

El intervalo de confianza de nivel α se obtiene a partir de:

$$\left(PAR - z_{1-\alpha/2} \sqrt{\hat{Var}(PAR)}, \quad PAR + z_{1-\alpha/2} \sqrt{\hat{Var}(PAR)} \right) \quad (4.20)$$

Donde

P es la proporción de expuestos que suele ser estimada por:

$$\hat{p} = \frac{a+b}{n}$$

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1-\alpha)\%$ es el nivel de confianza.

Intervalos de confianza Bootstrap. Se calculan para la odds-ratio y el riesgo relativo. La técnica utilizada es el *Bootstrap percentil t-simetrizado*: se obtienen n^* cuantiles ordenados en sentido ascendente,

$$cuantil_i^* = \frac{|\hat{\Theta}_i^* - \hat{\Theta}|}{\sqrt{\hat{var}(\hat{\Theta})^*}} \quad i=1, \dots, n^* \quad (4.21)$$

tomando el cuantil que ocupa la posición parte entera de $n^* \times \alpha$ se calcula el intervalo de confianza para Θ :

$$\left(\text{Exp}(\hat{\Theta} - cuantil \times \sqrt{\hat{var}(\hat{\Theta})}), \quad \text{Exp}(\hat{\Theta} + cuantil \times \sqrt{\hat{var}(\hat{\Theta})}) \right) \quad (4.22)$$

Donde

n^* número de réplicas bootstrap,

Θ es la odds-ratio o el riesgo relativo que se desea estimar,

$\hat{\Theta}$ es el estimador del Ln(OR) (ver Fleiss) o el estimador del Ln(RR) propuesto por Katz (ver Kahn),

$$\hat{\Theta} = \text{Ln} \left[\frac{(a+0.5)(d+0.5)}{(c+0.5)(b+0.5)} \right] \quad \hat{\Theta} = \text{Ln} \left[\frac{\frac{a}{a+b}}{\frac{c}{c+d}} \right]$$

$\hat{\Theta}^*$ es el estimador bootstrap del Ln(OR) o el estimador bootstrap del Ln(RR) propuesto por Katz, $\hat{var}(\hat{\Theta})$ es el estimador de la varianza del Ln(OR) (ecuación 4.2) o el estimador de la varianza del Ln(RR) propuesto por Katz (ecuación 4.5)

$\hat{var}(\hat{\Theta})^*$ es el estimador bootstrap de la varianza del Ln(OR) o el estimador bootstrap de la varianza del Ln(RR) propuesto por Katz.

Casos–controles pareados:

Un control por caso

		Control	
		+	-
Caso	+	a	b
	-	c	d

$$N = B + C$$

Estimador de máxima verosimilitud de la ODDS-RATIO, condicional en el número de pares discordantes b+c :

$$OR = \frac{B}{C} \quad (4.23)$$

TEST DE McNEMAR: Test bilateral para chequear la hipótesis nula de que la odds-ratio es la unidad, se obtiene calculando el estadístico:

$$\chi^2 = \frac{(|B - C| - 1)^2}{(B + C)} \quad (4.24)$$

El valor p correspondiente se obtiene a partir de la distribución χ^2 con 1 grado de libertad.

Para el cálculo de los intervalos de confianza se utiliza la aproximación descrita por Schlesselman.

Dos controles por caso

Caso	+	+	+	+	-	-	-	-
Control 1	+	+	-	-	+	+	-	-
Control 2	+	-	+	-	+	-	+	-
Nº	n_0	n_1	n_2	n_3	n_4	n_5	n_6	n_7

ESTIMADOR MANTEL-HAENSZEL de la odds-ratio:

$$OR = \frac{n_1 + n_2 + 2n_3}{2n_4 + n_5 + n_6} \quad (4.25)$$

El test para chequear la hipótesis $H_0: OR=1$ se obtiene calculando el estadístico:

$$\chi^2_{M-H} = \frac{(N_1 - N_2)^2}{N_2} \quad (4.26)$$

cuya distribución es aproximadamente una χ^2 con 1 grado de libertad bajo H_0 .

ESTIMADOR DE LA VARIANZA de $Ln(OR)$:

$$\hat{Var}(Ln(OR)) = \frac{1}{\chi^2_{M-H}} \quad (4.27)$$

Intervalo de confianza aproximado para la odds-ratio:

$$\left(OR \text{ Exp} \left(-z_{1-\alpha/2} \sqrt{\hat{Var}(Ln(OR))} \right), OR \text{ Exp} \left(z_{1-\alpha/2} \sqrt{\hat{Var}(Ln(OR))} \right) \right) \quad (4.28)$$

Donde

$$N_1 = [n_1 + n_2 + 2(n_3 - n_4) - (n_5 + n_6)]/3$$

$$N_2 = 2[n_1 + n_2 + n_3 + n_4 + n_5 + n_6]/9$$

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,
 $100 \times (1 - \alpha)\%$ es el nivel de confianza.

Tres o más controles por caso

ESTIMADOR MANTEL-HAENSZEL de la odds-ratio:

$$OR = \frac{\sum_{j=0}^c (c-j)n_j(+)}{\sum_{j=0}^c jn_j(-)} \quad (4.29)$$

Test de Mantel-Haenszel para contrastar la hipótesis $H_0: OR=1$:

$$\chi_{M-H}^2 = \frac{(|T_1 - T_2| - 1/2)^2}{T_3} \quad (4.30)$$

Estimador de la varianza de $Ln(OR)$:

$$\hat{Var}(Ln(OR)) = \frac{1}{\chi_{M-H}^2} \quad (4.31)$$

A partir del cual se puede obtener un intervalo de confianza aproximado para la odds-ratio.

Donde

$c = n^\circ$ de controles por caso,

$n_j(+)$ = n° de pares con caso (+) y j controles (+),

$n_j(-)$ = n° de pares con caso (-) y j controles (+),

$$T_1 = \sum_{j=0}^{c-1} n_j(+)$$

$$T_2 = \sum_{j=0}^c j m_j / (c+1)$$

$$T_3 = \sum_{j=0}^c j(c+1-j) m_j / (c+1)^2$$

$$m_j = n_{j-1}(+) + n_j(-)$$

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1-\alpha)\%$ es el nivel de confianza.

Tablas 2x2 estratificadas:

Dadas m tablas 2x2:

	Enfermos	Sanos
Fr +	a_i	b_i
Fr -	c_i	d_i

Denotando por $n_i = a_i + b_i + c_i + d_i$.

MEDIDAS DE ASOCIACIÓN BRUTAS: son las correspondientes a la tabla 2x2 dada por: sa, sb, sc y sd

$$sa = \sum_{i=1}^m a_i, \quad sb = \sum_{i=1}^m b_i, \quad (4.32)$$

$$sc = \sum_{i=1}^m c_i, \quad sd = \sum_{i=1}^m d_i$$

MÉTODO BASADO EN VARIANZAS (OR Wolf y RR Katz): medida de asociación global (media ponderada):

$$\bar{y} = \frac{\sum_{i=1}^m w_i y_i}{\sum_{i=1}^m w_i} \quad (4.33)$$

Varianza de la medida de asociación global:

$$\text{var}(\bar{y}) = \frac{1}{\sum_{i=1}^m w_i} \quad (4.34)$$

Bajo hipótesis de homogeneidad el estadístico

$$\chi^2 = \sum_{i=1}^m w_i (y_i - \bar{y})^2 \quad (4.35)$$

sigue una distribución Chi-Cuadrado con $m-1$ grados de libertad. Para más detalles acerca de este método ver Fleiss.

MÉTODO DE MANTEL-HAENSZEL: odds-ratio ajustada:

$$OR_{M-H} = \frac{\sum_{i=1}^m (a_i d_i)}{\sum_{i=1}^m (c_i b_i)} \quad (4.36)$$

Intervalo de Confianza de la odds-ratio ajustada:

$$\left(OR_{M-H} \text{Exp}\left(-z_{1-\alpha/2} \sqrt{V_{OR}}\right), OR_{M-H} \text{Exp}\left(z_{1-\alpha/2} \sqrt{V_{OR}}\right) \right) \quad (4.37)$$

Riesgo relativo ajustado:

$$RR_{M-H} = \frac{\sum_{i=1}^m a_i (c_i + d_i)}{\sum_{i=1}^m c_i (a_i + b_i)} \quad (4.38)$$

Intervalo de confianza del riesgo relativo ajustado:

$$\left(RR_{M-H} \text{Exp}\left(-z_{1-\alpha/2} \sqrt{V_{RR}}\right), RR_{M-H} \text{Exp}\left(z_{1-\alpha/2} \sqrt{V_{RR}}\right) \right) \quad (4.39)$$

Donde

y_i es la medida de asociación elegida correspondiente a la tabla i -ésima, ($i=1, \dots, m$)

$$Y_i = \ln(OR_i) \text{ o } Y_i = \ln(RR_i) \quad (i=1, \dots, m)$$

$\text{var}(y_i)$ es la varianza de la medida de asociación, ($i=1, \dots, m$)

w_i son los pesos ($i=1, \dots, m$)

$$w_i = \frac{1}{\text{var}(y_i)}$$

$$V_{OK} = \frac{\sum_{i=1}^m a_i d_i (a_i + d_i) + \sum_{i=1}^m [a_i d_i (a_i + d_i) + c_i b_i (c_i + b_i)]}{2 \left(\sum_{i=1}^m a_i d_i \right)^2} + \frac{\sum_{i=1}^m c_i b_i (c_i + b_i)}{2 \left(\sum_{i=1}^m c_i b_i \right)^2}$$

$$V_{WR} = \frac{\sum_{i=1}^m [(a_i + b_i) + (c_i + d_i)(a_i + c_i) - a_i c_i n_i]}{\sum_{i=1}^m a_i (a_i + c_i) \sum_{i=1}^m c_i (a_i + b_i)}$$

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,
 $100x(1-\alpha)\%$ es el nivel de confianza.

Tablas MxN:

CHI-CUADRADO (χ^2):

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (4.40)$$

g.l. = (m-1)(n-1)

CHI-CUADRADO RAZÓN DE VEROSIMILITUDES (G^2):

$$G^2 = \sum_{i=1}^m \sum_{j=1}^n n_{ij} \ln \left(\frac{n_{ij}}{e_{ij}} \right) \quad (4.41)$$

g.l. = (m-1)(n-1)

PHI (Φ^2):

$$\phi^2 = \sqrt{\frac{\chi^2}{n_{..}}} \quad (4.42)$$

COEFICIENTE DE CONTINGENCIA (C):

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n_{..}}} \quad (4.43)$$

V DE CRAMER:

$$V = \sqrt{\frac{\chi^2}{n_{..}(\min - 1)}} \quad (4.44)$$

D'S DE SOMERS:

$$d_{m/n} = \frac{P - Q}{n_{..}^2 - \sum_{j=1}^n n_{*j}^2} \quad d_{n/m} = \frac{P - Q}{n_{..}^2 - \sum_{i=1}^m n_{i*}^2} \quad (4.45)$$

TAUS DE GOODMAN Y KRUSKALL (τ):

$$\tau_{m/n} = \frac{n_{..} \sum_{j=1}^n \sum_{i=1}^m \frac{n_{ij}^2}{n_{i*}} - \sum_{j=1}^n n_{*j}^2}{n_{..}^2 - \sum_{j=1}^n n_{*j}^2} \quad \tau_{n/m} = \frac{n_{..} \sum_{i=1}^m \sum_{j=1}^n \frac{n_{ij}^2}{n_{*j}} - \sum_{i=1}^m n_{i*}^2}{n_{..}^2 - \sum_{i=1}^m n_{i*}^2} \quad (4.46)$$

TAU-B DE KENDALL (τ_b):

$$\tau_b = \frac{(P-Q)}{\sqrt{(n_{..}^2 - \sum_{i=1}^m n_{i.}^2)(n_{..}^2 - \sum_{j=1}^n n_{.j}^2)}} \quad (4.47)$$

TAU-C DE STUART (τ_c):

$$\tau_c = \frac{(P-Q)min}{n_{..}^2(min-1)} \quad (4.48)$$

CORRELACIÓN PRODUCTO-MOMENTO (r):

$$r = \frac{\sum_{i=1}^m \sum_{j=1}^n n_{ij}(i-\bar{i})(j-\bar{j})}{\sqrt{\sum_{i=1}^m n_{i.}(i-\bar{i})^2 \sum_{j=1}^n n_{.j}(j-\bar{j})^2}} \quad (4.49)$$

COEFICIENTE DE CORRELACIÓN POR RANGOS DE SPEARMAN (r_s): Fórmula para el caso de empates (Siegel; 1956):

$$r_s = \frac{\frac{1}{2}(v+w) - 6 \sum_{i=1}^m \sum_{j=1}^n n_{ij} d_{ij}^2}{\sqrt{vw}} \quad (4.50)$$

Donde

min es el mínimo entre m y n ,

n_{ij} son las frecuencias observadas en la tabla de contingencia $M \times N$ ($i=1, \dots, m$; $j=1, \dots, n$),

$n_{i.}$, $n_{.j}$ y $n_{..}$ son los totales por fila, columna y total de la tabla de contingencia,

$$n_{i.} = \sum_{j=1}^n n_{ij}$$

$$n_{.j} = \sum_{i=1}^m n_{ij}$$

$$n_{..} = \sum_{i=1}^m \sum_{j=1}^n n_{ij}$$

$$e_{ij} = \frac{n_{i.} n_{.j}}{n_{..}}$$

$$C_{ij} = \sum_{k>i} \sum_{l>j} n_{ij} + \sum_{k<i} \sum_{l<j} n_{ij}$$

$$D_{ij} = \sum_{k>i} \sum_{l<j} n_{ij} + \sum_{k<i} \sum_{l>j} n_{ij}$$

$$P = \sum_{i=1}^m \sum_{j=1}^n n_{ij} C_{ij}$$

$$Q = \sum_{i=1}^m \sum_{j=1}^n n_{ij} D_{ij}$$

$$\bar{i} = \sum_{i=1}^m i n_{i.} / n_{..}$$

$$\bar{j} = \sum_{j=1}^n j n_{.j} / n_{..}$$

$$v = n_{..}^3 - \sum_{i=1}^m n_{i.}^3$$

$$w = n_{..}^3 - \sum_{j=1}^n n_{.j}^3$$

$$d_{ij} = \sum_{k < j} n_{.k} - \sum_{l < i} n_{l.} + (n_{.j} - n_{i.})/2.$$

Tablas MxM y 2xM:

TEST PARA LA HOMOGENEIDAD MARGINAL: Para contrastar la hipótesis $H_0: P_{i.} = P_{.j} (i=1, \dots, m)$ se utiliza el estadístico

$$\chi^2 = \mathbf{d}'\mathbf{V}^{-1}\mathbf{d} \quad (4.51)$$

Si H_0 es cierta el estadístico χ^2 sigue una distribución Chi-Cuadrado con $(m-1)$ grados de libertad. Ver Everit.

COEFICIENTE DE CONCORDANCIA DE COHEN: Es un caso particular de la concordancia para dos observadores y múltiples categorías tomando pesos unitarios en la diagonal ($w_{ii}=1$) y nulos en el resto ($w_{ij}=0$).

TEST DE TENDENCIA LINEAL: En una tabla 2xM se hace la regresión lineal de $n_{ij}/n_{.j}$ sobre j , y se calcula el test Chi-Cuadrado con un grado de libertad:

$$\chi^2 = \frac{\left[n_{..} \sum_j (n_{1j} - \frac{n_{.j}n_{1.}}{n_{..}})(j - \bar{j}) \right]^2}{n_{1.}n_{2.} \sum_j n_{.j}(j - \bar{j})} \quad (4.52)$$

Donde

$$p_{i.} = n_{i.}/n_{..}$$

$$p_{.j} = n_{.j}/n_{..}$$

\mathbf{d} es un vector de dimensión $(m-1)$ formado por cualquiera de las diferencias

$$d_{i.} = n_{i.} - n_{.i}$$

\mathbf{V} es la matriz de covarianzas de los d_i 's cuyos elementos son:

$$v_{ii} = n_{i.} + n_{.i} - 2n_{ii}$$

$$v_{ij} = -(n_{ij} + n_{ji})$$

$$\bar{j} = \sum_{j=1}^n j \frac{n_{.j}}{n_{..}}$$

Vigilancia epidemiológica

Captura-Recaptura

		Sistema 1		
		Detectados	No detectados	Total
Sistema 2	Detectados	<i>a</i>	<i>b</i>	<i>n₁</i>
	No detectados	<i>c</i>	<i>d</i>	
	Total	<i>m₁</i>		<i>n</i>

Casos estimados:

$$n = \frac{(n_1 + 1)(m_1 + 1)}{a + 1} \quad (4.53)$$

Estimación de la varianza de n :

$$\hat{Var}(n) = \frac{(n_1 + 1)(m_1 + 1)cb}{(a + 1)^2(a + 2)} \quad (4.54)$$

La integridad de los sistemas se calcula como sigue:

$$\begin{aligned} \text{Sistema 1} &= \frac{m_1}{n} & \text{Sistema 2} &= \frac{n_1}{n} \\ \text{Sistemas 1 y 2} &= \frac{(a + b + c)}{n} \end{aligned} \quad (4.55)$$

Detección de clusters

Método de Knox

		Tiempo		
		+	-	Total
Espacio	+	a	b	n_1
	-	c	d	n_2
Total		m_1	m_2	n

El estadístico de interés es $X = \text{“número de pares cercanos en espacio y tiempo”}$ (a es el valor observado). A partir de los marginales se puede estimar su valor esperado:

$$m = \frac{n_1 m_1}{n} \quad (4.56)$$

Suponiendo que X sigue una distribución de Poisson de parámetro m (media), la probabilidad de observar por lo menos a pares cercanos en espacio y tiempo es:

$$P(X \geq a) = \sum_{x=a}^{\infty} \frac{e^{-m} m^x}{x!} \quad (4.57)$$

Para más información acerca de este método ver Steven M.T. y Elliott C.

	Enfermos	Sanos
Fr +	a	b
Fr -	c	d

Sensibilidad: $\frac{a}{a+c}$ (5.1)

Especificidad: $\frac{d}{b+d}$ (5.2)

Valor predictivo positivo: $\frac{a}{a+b}$ (5.3)

Valor predictivo negativo: $\frac{d}{c+d}$ (5.4)

Valor global de la prueba: $\frac{a+d}{a+b+c+d}$ (5.5)

Razón de verosimilitud positiva: $\frac{a/(a+c)}{b/(b+d)}$ (5.6)

Razón de verosimilitud negativa: $\frac{c/(a+c)}{d/(b+d)}$ (5.7)

Los intervalos de confianza se calculan a partir de los datos para proporciones en el módulo de inferencias, ecuación (6.17).

Curva ROC

Tablas 2x2. Se transforman hasta conseguir tablas 2xN y se tratan como éstas.

Tablas 2xN. Punto óptimo:

$$m = \left(\frac{\text{coste de falso - positivo}}{\text{coste de falso - negativo}} \right) \times \frac{1-P}{P} \quad (5.8)$$

Estimación del área verdadera bajo la curva Roc, Θ :

$$\Theta = P\{x_A > x_N\}$$

$$\hat{\Theta} = \hat{P}\{x_A > x_N\} = \frac{1}{n_A n_N} \sum_1^{n_A} \sum_1^{n_N} S(x_A, x_N) \quad (5.9)$$

Desviación estándar del área bajo la curva Roc estimada:

$$sd(\hat{\Theta}) = \sqrt{\frac{\hat{\Theta}(1-\hat{\Theta}) + (n_A - 1)(Q_1 - \hat{\Theta}^2) + (n_N - 1)(Q_2 - \hat{\Theta}^2)}{n_A n_N}} \quad (5.10)$$

Intervalo de confianza:

$$\left(\hat{\Theta} - z_{1-\alpha/2} sd(\hat{\Theta}), \hat{\Theta} + z_{1-\alpha/2} sd(\hat{\Theta}) \right) \quad (5.11)$$

Donde

P es el predominio o probabilidad anterior de enfermedad,

x_A es la distribución de los valores de x para imágenes no normales seleccionadas aleatoriamente,

x_N es la distribución de los valores de x para imágenes normales seleccionadas aleatoriamente,

n_A es el tamaño muestral para la población A,

n_N es el tamaño muestral para la población N,

$$S(x_A, x_N) = \begin{cases} 1 & \text{si } x_A > x_N \\ \frac{1}{2} & \text{si } x_A = x_N \\ 0 & \text{si } x_A < x_N \end{cases} \quad (\text{sólo datos discretos})$$

Q_1 es la probabilidad de que dos imágenes aleatorias no normales puedan ser clasificadas con mayor sospecha que una imagen aleatoria normal,

Q_2 es la probabilidad de que una imagen aleatoria no normal pueda ser clasificada con mayor sospecha que dos imágenes normales elegidas aleatoriamente,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100(1-\alpha)\%$ es el nivel de confianza.

Módulo 6: inferencias

Inferencias para medias. Una muestra:

Desviación estándar desconocida. Estadístico evaluado:

$$d = \frac{\bar{x} - \mu_0}{sd/\sqrt{n}} \quad (6.1)$$

Valor p :

$$p = P\left\{ |t| > \left| \frac{\bar{x} - \mu_0}{sd/\sqrt{n}} \right| \right\} \quad (6.2)$$

Intervalo de confianza de la media poblacional:

$$\left(\bar{x} - t_{1-\alpha/2} \frac{sd}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2} \frac{sd}{\sqrt{n}} \right) \quad (6.3)$$

Donde

d es el estadístico evaluado,

\bar{x} es la media muestral,

μ_0 es la media poblacional,

sd es la desviación estándar muestral,

n es el tamaño muestral,

p es el p -valor,

t es una variable t de Student con $n-1$ grados de libertad,

t_α es el cuantil de la distribución t de Student que deja una cola α a la izquierda,

$100 \times (1-\alpha)\%$ es el nivel de confianza.

Desviación estándar conocida. Las fórmulas son las mismas que en el caso anterior, sólo varía en que en lugar de la variable t de Student aparece una variable normal, tanto en la fórmula del valor p como en el intervalo de confianza.

Inferencia para medias. Dos muestras:

Independientes. Varianzas conocidas. Estadístico evaluado:

$$d = \frac{(\bar{x}_1 - \bar{x}_2) - dif}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (6.4)$$

Valor p :

$$p = P\left\{ |z| > \left| \frac{(\bar{x}_1 - \bar{x}_2) - dif}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \right\} \quad (6.5)$$

Intervalo de confianza de la media poblacional

$$\left((\bar{x}_1 - \bar{x}_2) - z_{1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right) \quad (6.6)$$

Donde

d es el estadístico evaluado,

dif es la diferencia entre las medias que se van a contrastar,

\bar{x}_1 es la media muestral de la primera población,

\bar{x}_2 es la media muestral de la segunda población,

s_1 es la desviación estándar muestral de la primera población,

s_2 es la desviación estándar muestral de la segunda población,

n_1 es el tamaño muestral de la primera población,

n_2 es el tamaño muestral de la segunda población,

p es el p -valor,

z es una variable aleatoria normal,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100 \times (1 - \alpha)\%$ es el nivel de confianza.

Independientes. Varianzas desconocidas. Fórmula empleada para calcular el valor p del contraste de igualdad de varianzas:

$$p = 2 P \left\{ F_{n_1-1, n_2-1} \geq \frac{s_1^2}{s_2^2} \right\} \quad (6.7)$$

Si el valor p del contraste anterior es menor de 0,05 EPIDAT 2.0 adopta la opción de **VARIANZAS DESCONOCIDAS Y DESIGUALES**. Las fórmulas son iguales que en el caso de varianzas conocidas, exceptuando que, en lugar de una variable normal, aparece una variable t de Student con f grados de libertad, en la fórmula del valor p y en la del intervalo de confianza. El cálculo de f se realiza mediante la siguiente fórmula (aproximación de Welch):

$$f = \frac{\left(\frac{s_1^2}{m_1} + \frac{s_2^2}{m_2} \right)^2}{\left(\frac{s_1^2}{m_1} \right)^2 \frac{1}{m_1 + 1} + \left(\frac{s_2^2}{m_2} \right)^2 \frac{1}{m_2 + 1}} - 2 \quad (6.8)$$

En el caso de que en el contraste de varianzas, el valor p sea mayor que 0,05, EPIDAT 2.0 adopta la opción de **VARIANZAS DESCONOCIDAS E IGUALES**, y emplea las siguientes fórmulas:

Estadístico evaluado:

$$d = \frac{(\bar{x}_1 - \bar{x}_2) - dif}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \cdot \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.9)$$

Valor p :

$$p = P \left\{ |t| > \left| \frac{(\bar{x}_1 - \bar{x}_2) - dif}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \cdot \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \right\} \quad (6.10)$$

Intervalo de confianza de la media poblacional:

$$\left((\bar{x}_1 - \bar{x}_2) - t_{1-\alpha/2} \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\alpha/2} \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \quad (6.11)$$

Donde

d es el estadístico evaluado,

dif es la diferencia entre las medias que se van a contrastar,

\bar{x}_1 es la media muestral de la primera población,

\bar{x}_2 es la media muestral de la segunda población,

s_1 es la desviación estándar muestral de la primera población,

s_2 es la desviación estándar muestral de la segunda población,

n_1 es el tamaño muestral de la primera población,

n_2 es el tamaño muestral de la segunda población,

p es el p -valor,

F es la función de distribución F de *Snedecor*,

t es una variable t de *Student* con $n-1$ grados de libertad,

t_α es el cuantil de la distribución t de *Student* que deja una cola α a la izquierda,

$100x(1-\alpha)\%$ es el nivel de confianza.

Dependientes. Estadístico evaluado:

$$d = \frac{(\bar{x}_1 - \bar{x}_2) - dif}{\sqrt{s_1^2 + s_2^2 - 2 cov}} \quad (6.12)$$

Valor p :

$$p = P \left\{ |t| > \frac{(\bar{x}_1 - \bar{x}_2) - dif}{\sqrt{s_1^2 + s_2^2 - 2 cov}} \right\} \quad (6.13)$$

Intervalo de confianza:

$$\left((\bar{x}_1 - \bar{x}_2) - t_{1-\alpha/2} \sqrt{s_1^2 + s_2^2 - 2 cov}, (\bar{x}_1 - \bar{x}_2) + t_{1-\alpha/2} \sqrt{s_1^2 + s_2^2 - 2 cov} \right) \quad (6.14)$$

Donde

d es el estadístico evaluado,

dif es la diferencia entre las medias que se va a contrastar,

\bar{x}_1 es la media muestral de la primera población,

\bar{x}_2 es la media muestral de la segunda población,

s_1 es la desviación estándar muestral de la primera población,

s_2 es la desviación estándar muestral de la segunda población,

cov es la covarianza entre las dos variables,

n es el tamaño muestral (número de pares),

p es el p -valor,

F es la función de distribución F de *Snedecor*,

t es una variable t de *Student* con $n-1$ grados de libertad,

t_α es el cuantil de la distribución t de *Student* que deja una cola α a la izquierda,

$100x(1-\alpha)\%$ es el nivel de confianza.

Inferencia para proporciones. Una muestra

Estadístico evaluado:

$$z = \frac{|p - P_0| - \frac{1}{2n}}{\sqrt{\frac{pq}{n}}} \quad (6.15)$$

Valor p :

$$p = P \left\{ |z| \geq \frac{|p - P_0| - \frac{1}{2n}}{\sqrt{\frac{pq}{n}}} \right\} \quad (6.16)$$

Intervalo de confianza de la proporción poblacional:

$$\left(p - z_{1-\alpha/2} \sqrt{\frac{pq}{n}} - \frac{1}{2n}, \quad p + z_{1-\alpha/2} \sqrt{\frac{pq}{n}} + \frac{1}{2n} \right) \quad (6.17)$$

Donde

z es el estadístico evaluado,

p es la proporción muestral,

$$q = 1 - p$$

P_0 es la proporción bajo la hipótesis nula,

n es el tamaño muestral,

z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,

$100(1-\alpha)\%$ es el nivel de confianza.

Inferencia para proporciones: Dos muestras

Independiente. Estadístico evaluado:

$$z = \frac{(p_2 - p_1) - (P_2 - P_1) - \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \quad (6.18)$$

Valor p :

$$p = P \left\{ |z| \geq \frac{(p_2 - p_1) - (P_2 - P_1) - \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \right\} \quad (6.19)$$

Intervalo de confianza de la diferencia entre las proporciones poblacionales:

$$\left((p_2 - p_1) - z_{1-\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} - \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right), \quad (p_2 - p_1) + z_{1-\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} + \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right) \quad (6.20)$$

Donde

z es el estadístico evaluado,

p_1 es la proporción muestral en el tiempo 1,

p_2 es la proporción muestral en el tiempo 2,

$$q_1 = 1 - p_1$$

$$q_2 = 1 - p_2$$

P_1 es la proporción bajo la hipótesis nula de la primera población,
 P_2 es la proporción bajo la hipótesis nula de la segunda población,
 n_1 es el tamaño muestral de la primera población,
 n_2 es el tamaño muestral de la segunda población,
 p es el p -valor,
 z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,
 $100x(1-\alpha)\%$ es el nivel de confianza.

Dependientes (pareados). Estadístico evaluado:

$$z = \frac{(p_2 - p_1) - (P_2 - P_1) - \frac{1}{n}}{\frac{\sqrt{(a+d)(b+c) + 4bc}}{n\sqrt{n}}} \quad (6.21)$$

Valor p :

$$p = P \left\{ |z| \geq \left| \frac{(p_2 - p_1) - (P_2 - P_1) - \frac{1}{n}}{\frac{\sqrt{(a+d)(b+c) + 4bc}}{n\sqrt{n}}} \right| \right\} \quad (6.22)$$

Intervalo de confianza de la diferencia entre las proporciones poblacionales:

$$\left((p_2 - p_1) - z_{1-\alpha/2} \frac{\sqrt{(a+d)(b+c) + 4bc}}{n\sqrt{n}} - \frac{1}{n}, (p_2 - p_1) + z_{1-\alpha/2} \frac{\sqrt{(a+d)(b+c) + 4bc}}{n\sqrt{n}} + \frac{1}{n} \right) \quad (6.23)$$

Donde

z es el estadístico evaluado,
 p_1 es la proporción muestral de la primera población,
 p_2 es la proporción muestral de la segunda población,
 P_1 es la proporción bajo la hipótesis nula de la primera población,
 P_2 es la proporción bajo la hipótesis nula de la segunda población,
 n es el tamaño muestral en número de pares,
 a es el número de casos positivo-positivo,
 b es el número de casos positivo-negativo,
 c es el número de casos negativo-positivo,
 d es el número de casos negativo-negativo,
 p es el p -valor,
 z_α es el cuantil de la distribución normal que deja una cola α a la izquierda,
 $100x(1-\alpha)\%$ es el nivel de confianza.



Módulo 7: priorización

Determinación de necesidades en salud:

Media:

$$m = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (7.1)$$

Desviación estándar:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n - 1}} \quad (7.2)$$

Valores z de las x's:

$$z_i = \frac{(x_i - m)}{s} \quad (7.3)$$

$i = 1, \dots, n$

Valores z calculados:

z < -3,5	= -4
-3,5 ≤ z < -2,5	= -3
-2,5 ≤ z < -1,5	= -2
-1,5 ≤ z < -0,5	= -1
-0,5 ≤ z < 0,0	= 0
0,0 ≤ z < 0,5	= 0
0,5 ≤ z < 1,5	= 1
1,5 ≤ z < 2,5	= 2
2,5 ≤ z < 3,5	= 3
z > 3,5	= 4

entonces el puntaje asignado es:

El valor del puntaje final de cada unidad de análisis es la suma de los puntajes asignados a cada indicador.

Módulo 1: AJUSTES

Ejercicio 1.1. $T_{Cali} = \frac{16900000}{597925} = 28,26441$ $T_{São Paulo} = \frac{96500000}{3512495} = 27,447335$

	T. estándar Cali	T. estándar São Paulo
MD	Tasa = 34,91 IC = (34,91; 34,92)	Tasa = 24,75 IC = (24,75; 24,75)
MI	RME = 137,48 IC = (116,76; 158,21)	RME = 99,96 IC = (93,66; 106,27)

Módulo 2: MUESTRAS

- Ejercicio 2.1. n = 246
- Ejercicio 2.2. n = 259
- Ejercicio 2.3. n = 1164
- Ejercicio 2.4. n = 704
- Ejercicio 2.5. n = 188
- Ejercicio 2.6. n = 275
- Ejercicio 2.7. n = 54
- Ejercicio 2.8. n = 56
- Ejercicio 2.9. n = 92
- Ejercicio 2.10. n = 1905
- Ejercicio 2.11. n = 43
- Ejercicio 2.12. 1- α = 90%, 1- β = 80%: TM = 17, VU = 7
1- α = 90%, 1- β = 90%: TM = 25, VU = 11

Módulo 3: CONCORDANCIA

- Ejercicio 3.1. Concordancia bruta = 70%
J de Youden = 0,439 IC = (0,308; 0,568)
kappa = 0,415 IC = (0,286; 0,543)
- Ejercicio 3.2. kappa sin ponderar = 0,67 IC = (0,645; 0,696)
kappa ponderado = 0,859 IC = (0,846; 0,871)
- Ejercicio 3.3. kappa = 0,233
p-valor = 0,00329
- Ejercicio 3.4. kappa global = 0,280
kappa(1) = 0,310
kappa(2) = 0,114
kappa(3) = 0,389

- Ejercicio 3.5.** **Semestre 1:** kappa = 0,64
Error estándar = 0,02356
- Semestre 2:** kappa = 0,687
Error estándar = 0,02432
- Semestre 3:** kappa = 0,132
Error estándar = 0,0426
- Semestre 4:** kappa = 0,518
Error estándar = 0,01871
kappa global = 0,537 p-valor = 0,1547
- Semestre 4 (bis):** kappa = 0,078
Error estándar = 0,02945
kappa global = 0,433 p-valor = 0,0124

Módulo 4: TABLAS

- Ejercicio 4.1.** $T_{exp} = 0,0912$; $T_{no\ exp} = 0,028$
RR = 3,25 IC = (2,53; 4,18)
RAP = 49% IC = (0,40; 0,58)
- Ejercicio 4.2.** Chi-cuadrado crudo = 31,65 p-valor = 0,000
Chi-cuadrado corregido = 30,51 p-valor = 0,000
OR = 3,52 IC = (2,17; 5,74)
RAP = 11% IC = (0,07; 0,15)

- Ejercicio 4.3.** OR = 2,36 IC = (1,06; 5,49)

Ejercicio 4.4.

Estrato	RR	IC
1	3,57	2,62; 4,89
2	4,28	2,66; 6,88
3	1,24	0,76; 2,04
global	2,96	2,34; 3,74

Ejercicio 4.5.

	OR	IC
Estrato 1	4,79	2,19; 10,45
Estrato 2	3,28	1,81; 5,94
crudo	3,52	2,22; 5,58
Mantel-Hanenszel	3,83	2,39; 6,93

Prueba de asociación = 33,38 p-valor = 0,000

- Ejercicio 4.6.** Chi-cuadrado de Pearson = 99,467
 $G^2 = 99,179$
Phi = 0,127
Coeficiente de contingencia C = 0,126
V de Cramer = 0,127

Tau de Goodman y Kruskal

Filas/columnas	0,01624
Columnas/filas	0,00433

- Ejercicio 4.7.** Chi-cuadrado de Pearson = 9,053
 $G^2 = 12,057$
Phi = 0,184
Coeficiente de contingencia C = 0,181
V de Cramer = 0,130
Correlación de Spearman = 0,14118

Tau-b de Kendall = 0,12853
 Tau-c de Stuart = 0,11176
 Correlación producto-momento = 0,15824
 kappa = 0,052
 Homogeneidad NxN = 106,921

Tau de Goodman y Kruskal

Filas/columnas	0,01235
Columnas/filas	0,01751

D de Somers

Filas/columnas	0,13804
Columnas/filas	0,11969

Ejercicio 4.8. Casos estimados = 158 IC = (154; 163)

Integridad

Sistema 1	86,94%
Sistema 2	81,90%
Ambos	97,65%

Ejercicio 4.9. Se esperaban 24 y se registraron 45
 P = 0,0000

Módulo 5: PRUEBAS DIAGNÓSTICAS

Ejercicio 5.1. VPP = 1,74 IC = (1,14; 2,63)
 VPN = 99,98 IC = (99,96; 99,99)

Ejercicio 5.2.

	Sensibilidad	Especificidad	VP+	VP-
Serie (sobre +)	84,0%	100%	100%	99,16%
Paralelo (sobre -)	100%	99,78%	96,15%	100%

Ejercicio 5.3.

	Sensibilidad	Especificidad
Serie	72%	96%
Paralelo	98%	54%

Ejercicio 5.4. Área = 0,9
 Pto. óptimo: S = 64%, E = 97% (130 mg/100)

Ejercicio 5.5. Área = 0,926
 Pto. óptimo: S = 97%, E = 69%

Módulo 6: INFERENCIAS

Ejercicio 6.1.

Nivel de confianza (1- α)	IC
90%	(2,43; 2,50)
95%	(2,42; 2,51)
99%	(2,41; 2,52)

Ejercicio 6.2. Proporción muestral = 0,857 IC = (0,828; 0,881)
 p-valor = 0,00019
 Se rechaza P = 80%

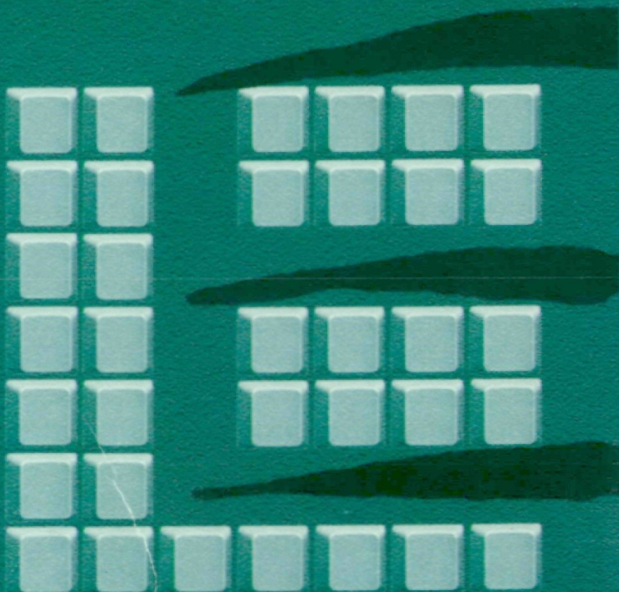
- Ejercicio 6.3.** Diferencia de medias = 0,42
 Se acepta la igualdad de varianzas: p-valor = 0,3268
 Se rechaza la igualdad de los índices: p-valor = 0,0
- Ejercicio 6.4.** Coeficiente de correlación muestral = $0,8/2 \times 1,5 = 0,266$
 Estadístico = -0,463
 No existe diferencia significativa: p-valor = 0,652
- Ejercicio 6.5.** Proporción muestral 1 = 0,274
 Proporción muestral 2 = 0,226
 Diferencia de proporciones observadas = 0,048
 Estadístico = -1,142
 No existen diferencias significativas: p-valor = 0,2569
- Ejercicio 6.6.** Diferencia de proporciones observada = -0,4 IC = (-0,524; -0,275)
 No existen diferencias significativas: p-valor = 0,0000

Módulo 7: PRIORIZACIÓN

Ejercicio 7.1.

Unidades de análisis

	1	2	3	4	5	6	7	8	9
Ptos.	1	-4	3	1	1	-1	1	-1	-4



CONSELLERÍA DE SANIDADE
E SERVIZOS SOCIAIS
Dirección Xeral de Saúde Pública



ORGANIZACIÓN PANAMERICANA
DE LA SALUD

525 23rd Street, N.W.
Washington D.C. 20037
E.U.A.