UNIVERSIDAD DE JAÉN

Material del curso "Recursos metodológicos y estadísticos para la docencia e investigación"

Manuel Miguel Ramos Álvarez

MATERIAL VI "INTRODUCCIÓN AL ANÁLISIS DE DATOS CATEGÓRICOS"

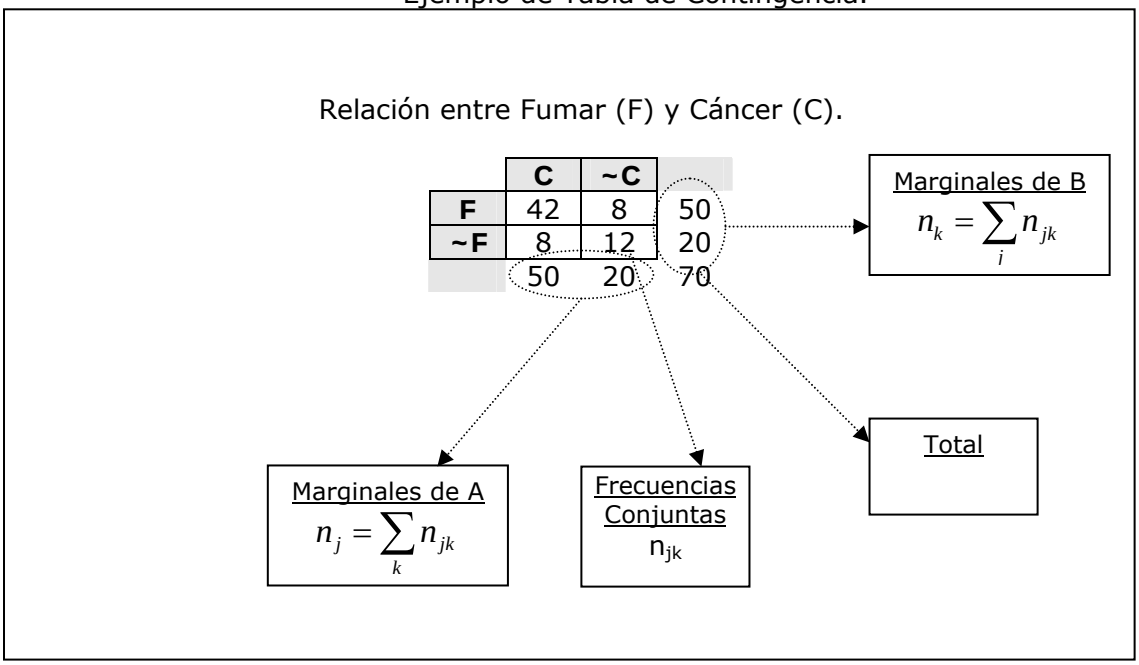
Índice

6. MARCO GENERAL DEL ANÁLISIS DE DISEÑOS CON VARIABLES CATEGÓRICAS

Bases:

- Cuando la investigación incluye variables categóricas, estrictamente tendremos frecuencias conjuntas.
- Ejemplo: La posible asociación entre fumar y desarrollar cáncer de pulmón.

Ejemplo de Tabla de Contingencia.



- Es decir, el estudio presenta dos variables, fumar y desarrollar cáncer de pulmón, que son categóricas y por lo tanto los datos se miden más bien a partir de la frecuencia de aparición de los valores de sendas variables (ó n_{jk}).

Objetivo:

- Hasta qué punto las dos variables se relacionan entre sí: los casos que confirman la idea de asociación son de la diagonal principal y los casos que no la confirman serían los de la diagonal secundaria. En principio, hay más casos favorables y por ende nos inclinaríamos a pensar que dichos datos apoyan la idea de que las dos variables del estudio están relacionadas.

Podemos trazar una analogía con el planteamiento básico de variables cuantitativas, donde:

- La frecuencia total n sería comparable a la gran media o media total $\overline{Y}$.
- Las frecuencias conjuntas n_{jk} serían comparables a las puntuaciones individuales Y_{jk}.
- Y las frecuencias marginales n_j y n_k serían comparables a las medias marginales $\overline{Y}_j$ e $\overline{Y}_k$ respectivamente.

Lógica estadística:

- Conceptos probabilísticos básicos. Dos sucesos son independientes cuando la **probabilidad de la conjunción equivale al producto de sus probabilidades individuales**:

$$p_{jk} = \frac{n_j}{n} \bullet \frac{n_k}{n} = p_j \bullet p_k$$

- Si nuestro modelo estadístico especificara la independencia entre las dos variables categóricas, para cada casilla esperaríamos obtener como frecuencia:

$$n\pi_{jk} \equiv m_{jk} = n\pi_j\pi_k$$

- Es decir, desde probabilidades ($P \; vs \; \Pi$) a frecuencias ($n_{jk} \; vs \; m_{jk}$), multiplicamos por n. Las frecuencias esperadas con la letra eme y Phi para las probabilidades o frecuencias relativas en la población.
- La ecuación básica que subyace al modelo es **multiplicativa** en lugar de ser aditiva pero si rescribimos las ecuaciones en logaritmos entonces volvemos a las bases estadísticas de tipo lineal-aditivo.

$$\ln(m_{jk}) = \ln(n) + \ln(\pi_j) + \ln(\pi_k)$$

- Podríamos mantenernos, pues, dentro de la perspectiva lineal general pero incluir una **función de enlace** que nos permitiese ir desde las predicciones del modelo lineal hacia la variable criterio:
  - $\hat{Y} = g(\mu)$, donde μ expresa los valores esperados en la variable criterio, la esperanza a partir de las frecuencias observadas ($E(y)$).
  - En el modelo lineal clásico dicha función es del tipo identidad
  - En los diseños categóricos la función de enlace podría ser del tipo logarítmico: $\hat{Y} = \log(\mu)$.
- **Este tipo de modelos se denomina logarítmico-lineal y el planteamiento analítico es la perspectiva lineal generalizada (GLM).**
- En el contexto explicativo, la regresión de var. criterio de tipo categórico se puede realizar también dentro del contexto del Modelo Lineal Generalizado, especificando una función de enlace del tipo Logit: $\hat{Y} = \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{\mu}{N-\mu}\right)$

  - Se hacen predicciones probabilísticas a partir de la ecuación de regresión, es decir toda una gama posible de valores entre 0 y 1. Así, pues, en lugar de hacer predicciones sobre π_{jk} podríamos hacerlas más bien sobre el cociente entre dicha probabilidad y su complementaria (una **razón de probabilidades, ó en inglés *odds***).

Los **estimadores** se pueden deducir fácilmente a partir de la teoría de sucesos:

- En el caso de una tabla de contingencia bidimensional, la independencia vendrá dada por:

$$\log(m_{jk}) = \lambda + \lambda_A + \lambda_B$$

- Es decir, para cada casilla esperaríamos obtener como frecuencia:

$$n\pi_{jk} \equiv m_{jk} = n\pi_j\pi_k$$

- Sustituimos las probabilidades por sus estimadores y nos quedaría:

$$\hat{m}_{jk} = \cancel{n} \bullet \frac{n_j}{\cancel{n}} \bullet \frac{n_k}{n} = \frac{n_j n_k}{n}$$

6.1. Ejemplos de aplicación de la perspectiva de análisis categórico para diferentes tipos de diseños de investigación.

VARIABLES	MODELO	DISEÑO DE APLICACIÓN
Categóricas todas. Sin diferenciar estatus variables.	Log-lineal	Descriptivo
Categóricas todas. Unas son var.ind. y otra es var.dep. (mm. independientes)	Logit Probit	Explicativo(i.e. ≈Experimental)
Categóricas todas. Unas son var.ind. y otra es var.dep. (mm. relacionadas)	Logit-GSK	Explicativo(i.e. ≈Experimental medidas repetidas)
No categóricos los Predictores y Categórico el criterio	Regresión Logística	Explicativa(i.e. Correlacional)
Cadena causal de variables categóricas	Logit-causal	Explicativo(i.e. ≈Cuasiexperi-mental)
Categóricas a través del tiempo (t1, t2,...)	Logit-Markov	Descriptivo(i.e. longitudinal, observacional)
Categóricas en diversas var.criterio	Logit-Latente	Descriptivo (i.e. ≈Análisis Factorial en Tests)

6.2. La codificación e interpretación analítica de las variables categóricas

- Hay varios sistemas de codificación, los dos más destacados son los siguientes:
  - En el sistema de efectos se plantea la comparación de niveles dos a dos. Los parámetros indican la diferencia entre los valores de la variable a la que afecta.
  - Los parámetros recogen el efecto neto de cada variable **respecto al total**.

  - En el sistema de codificación ficticio (*dummy*) se toma una casilla como **punto de origen** situado en el cero y todas las demás se definen de manera comparativa con respecto a la misma (i.e. a1b1). La interpretación de los parámetros es el cambio con respecto al punto cero.
  - Para interpretar los coeficientes según este sistema de codificación, se asigna un parámetro a cada casilla de la matriz de contingencia que supone el cambio respecto a un vértice de la misma:

0,000	-0,154
-0,693	-0,310

Codificación mediante diferentes sistemas de los datos.

SISTEMA DE EFECTOS

INFORMAC.	ϕCONST.	ϕA	ϕB	ϕA*B
a_1b_1	1	1	1	1
a_1b_2	1	1	-1	-1
a_2b_1	1	-1	1	-1
a_2b_2	1	-1	-1	1

SISTEMA FICTICIO (*DUMMY*)

INFORMAC.	ϕCONST.	ϕA	ϕB	ϕA*B
a_1b_1	1	0	0	0
a_1b_2	1	0	1	0
a_2b_1	1	1	0	0
a_2b_2	1	1	1	1

6.3. La codificación de información en diseños categóricos mediante programas informáticos.

¿Cómo se obtienen las tablas de contingencia a partir de la matriz de información original?

- Los módulos de análisis estadístico relacionados con diseños categóricos son muy diversos en el programa SPSS y además funcionan según una estructura diferente. Esto hace por ejemplo que a veces el programa nos pida las variables de análisis en bruto, es decir las combinaciones de las variables categóricas para cada uno de los sujetos medidos; mientras que en otras ocasiones se nos pide la tabla de contingencia subyacente (el cómputo de las frecuencias para cada combinación). A continuación se expone cómo pasar de un tipo de codificación a la otra, Para facilitar la exposición proponemos un ejemplo, como el que se resumen en la siguiente figura:

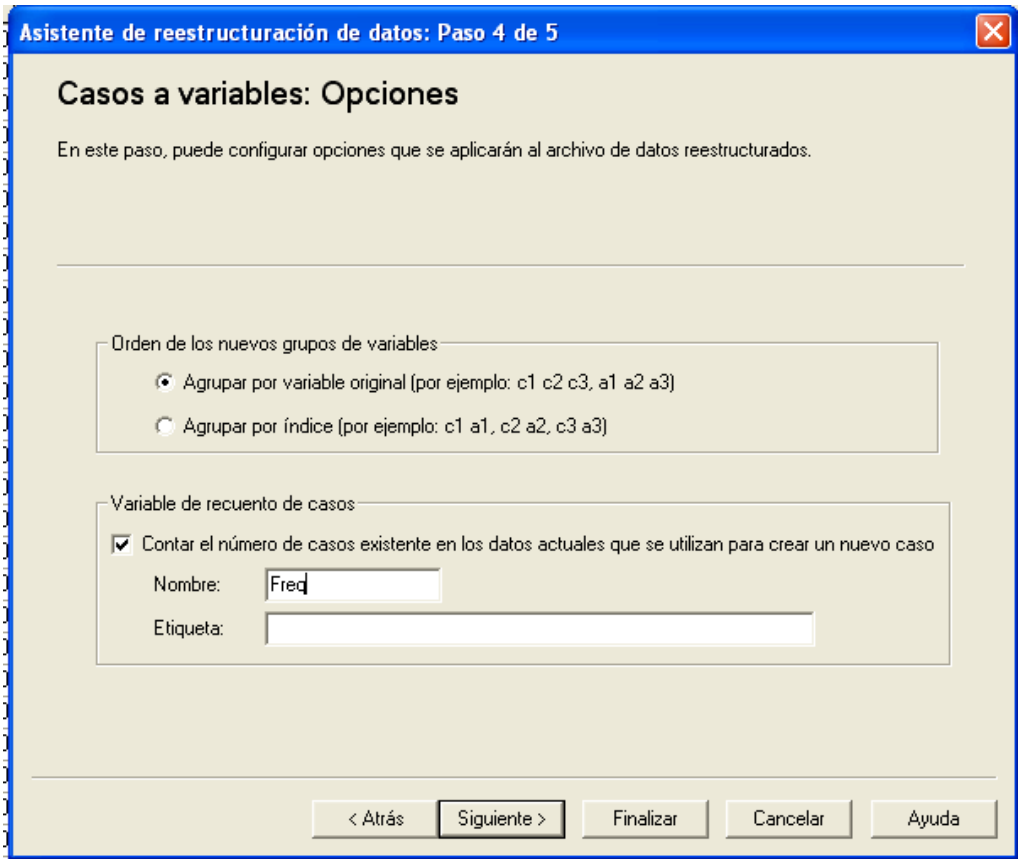
	entona	signifi
1	1,00	1,00
2	1,00	1,00
3	1,00	1,00
4	1,00	1,00
5	1,00	1,00
6	1,00	1,00
7	1,00	1,00
8	1,00	1,00
9	1,00	1,00
10	1,00	1,00
11	1,00	1,00
12	1,00	1,00
13	1,00	1,00
14	1,00	1,00
15	1,00	1,00
16	1,00	1,00
17	1,00	1,00
18	1,00	1,00
19	1,00	1,00
20	1,00	1,00
21	1,00	1,00

Fichero Datos Bruto
2 variables y 187 casos:
entona: Valores 1 y 2.
signifi: Valores 1 y 2.

	entona	signifi	freq
1	1,00	1,00	70
2	1,00	2,00	60
3	2,00	1,00	35
4	2,00	2,00	22

Fichero Frecuencias
3 variables y 4 casos:
entona: Valores 1 y 2.
signifi: Valores 1 y 2.
Freq: Los valores de las frecuencias

- Para obtener el Fichero de Frecuencias a partir del fichero de Datos en bruto, abrimos el correspondiente fichero de datos en bruto y entonces realizamos una reestructuración de datos según el comando **[Datos|Reestructurar]**, lo que nos despliega un asistente de 5 pasos. En el primer paso elegimos la opción Reestructurar los casos seleccionados en variables. En el segundo paso, definimos las dos variables como variables de identificación. En el 3º paso, elegimos la opción superior (que viene por defecto) de ordenación. En el paso 4º hay que definir la nueva variable que contendrá la frecuencia conjunta de aparición ("Freq"), es decir:



- En el sexto y último paso elegimos reestructurar los casos ahora y pulsamos el botón **[Finalizar]**. El programa cierra entonces el fichero original y nos crea uno nuevo con las variables deseadas. No olvidemos grabar el nuevo fichero.

➢ En el programa **Statistica**, todo es más simple a través del módulo especializado de análisis log-lineal. En concreto para obtener el Fichero de Frecuencias a partir del fichero de Datos en bruto, abrimos el correspondiente fichero de datos en bruto y entonces Statistics→ Advanced Linear/NonLinear Models → Log-Linear Analysis of Frequency Tables → Variables: Var1-Var2 → OK→ Input File: Raw Data → OK→ Pestaña Revew/Save → Save the table

6.4. Evaluar diferentes tipos de Hipótesis estadísticas en el contexto categórico

- En el marco del análisis categórico y en general, dentro del esquema de investigaciones descriptivas, se pueden utilizar multitud de pruebas estadísticas que van dirigidas al contraste de diferentes tipos de Hipótesis.
- El núcleo conceptual son las pruebas de **bondad de ajuste**.

A) Índice de Bondad de Ajuste

Información:

- Sobre el ajuste entre frecuencias observadas y esperadas a partir de algún Modelo estadístico o Hipótesis. Por ejemplo, el grado de ajuste a la Normal. Las frecuencias esperadas se obtienen a partir de lo especificado en tal Hipótesis o Modelo.

Matriz de Frecuencias observadas (n_{jk})

	b_1	b_2	n_k
a_1	75	125	200
a_2	30	70	100
n_j	105	195	300

$$m_{ij} = \frac{n_j \bullet n_k}{N}$$

Matriz de Frecuencias Esperadas (m_{jk})

	b_1	b_2	m_k
	70,00	130,00	200
	35,00	65,00	100
m_j	105	195	300

Hipótesis:

H_0: Ajuste.

H_1: No Ajuste.

$$H_0 : \pi_{jk} = \pi_k ; \forall_k$$

$$H_1 : \pi_{jk} \neq \pi_k ; \forall_k$$

Pruebas:

- Chi-Cuadrado Pearson (1900)
- Razón Verosimilitud Wilks (1935) ó G^2
- Mínimos Cuadrados Ponderados Neyman (1949).

Nomenclatura para identificar las frecuencias de las tablas de contingencia

VARIABLE A		VARIABLE B: 1	VARIABLE B: 0
	1	a	b
	0	c	d

Analizar → Estadísticos Descriptivos → Tablas de Contingencia → Filas: VarA; Columnas: VarB → Estadísticos *(seleccionar los asociados a Chi-Cuadrado)* → Casillas → Frecuencias Esperadas → Aceptar (análisis).

Statistics → Basicd Statistics → Tables and Banners → OK → Specify Tables → List1: VAR1; List2: VAR2 → OK → OK → *Pestaña* Options: Pearson & M-L Chi-square; Expected frequencies → Pestaña Advanced → Detailed two-way tables.

Interpretación Ejemplo:

A.1.) Pruebas de Contraste de Hipótesis para Proporciones

- **Una** muestra.
  - Pequeña (n<25): P = X/n; B (n, Π_0)
  - Intermedia (n ≈ 25): $Z = \dfrac{(P \pm 0{,}5/n) - \Pi_0}{\sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}}; N(0,1) \left\langle \begin{matrix} +: P > \Pi \\ -: P < \Pi \end{matrix} \right\rangle$
  - Grande (npq>3 ó n>25): $Z = \dfrac{P - \Pi_0}{\sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}}; N(0,1)$
- **Dos** Muestras.
  - **Independientes**
    - Sobre diferencia **nula** :

$$Z = \frac{P_1 - P_2}{\sqrt{\dfrac{n_1P_1 + n_2P_2}{n_1 + n_2}\left(1 - \dfrac{n_1P_1 + n_2P_2}{n_1 + n_2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}; N(0,1)$$

    - Sobre diferencia **no nula** (k):

$$Z = \frac{(P_1 - P_2) - k}{\sqrt{\dfrac{P_1(1-P_1)}{n_1} + \dfrac{P_2(1-P_2)}{n_2}}}; N(0,1)$$

  - **Relacionadas**: $Z = \dfrac{P_A - P_D}{\sqrt{\dfrac{A+D}{n^2}}}; N(0,1)$; Acuerdos y Desacuerdos.

Ver ejemplo de Excel: ContrProporMMRA.xls
En Statistica: Statistics→ Difference tests: r, %, means→OK→ Difference between two proportions

Interpretación Ejemplo:

B) Índices de Homogeneidad

Información:

- **Se fijan** las frecuencias marginales de una de las variables, por ejemplo A, y se clasifican las observaciones dentro de cada categoría de A en función de la otra variable considerada, B.
- Si la distribución de frecuencias de la variable que no estaba fijada es homogénea ó no a través de los niveles o categorías de la variable fijada.
- Ejemplo. la eficacia terapéutica de un nuevo método de tratamiento de la ansiedad en función del sexo. Si en el estudio fijamos un determinado número de varones, i.e. 200, y de mujeres, i.e. 100, entonces las Hipótesis estarán condicionadas por la variable que se ha fijado.

Matriz Probabilidades observadas $p(k/j)=n_{jk}/n_j$

	b_1	b_2
a_1	0,375	0,625
a_2	0,300	0,700
p_j	0,375	0,625

Matriz Probabilidades Esperadas

	b_1	b_2	P_k
	0,350	0,650	
	0,350	0,650	
P_j	0,350	0,650	0,075

Hipótesis:

H_0: Todas las (sub)poblaciones tienen la misma Distribución. $H_0 : \pi_{jk} = \pi_k ; \forall_k$

H_1: No todas las (sub)poblaciones tienen la misma Distribución. $H_1 : \pi_{jk} \neq \pi_k ; \forall_k$

Estadísticos:

- Generales $\chi^2; G^2; W^2$
- X^2 en 2x2
- Razón de Productos Cruzados (RPC) ó Razón Probabilidades (odds) y Logaritmo RPC (Log odd)
- Diferencia Probabilidades (Ver Pruebas Contraste Proporciones).
- Prueba McNemar (1955) para 2 muestras relacionadas (con frecuencias muy bajas).
- Prueba Cochran para k muestras relacionadas.

Analizar → Estadísticos Descriptivos → Tablas de Contingencia → Filas: VarA; Columnas: VarB → Estadísticos *(seleccionar los asociados a McNemar y Cochran)*→ Casillas → Frecuencias Esperadas → Aceptar (análisis).

En Statistica todo como en cuadro previo pero con los estadísticos: Fisher exact, Yates, McNemar (2x2).
Aparte, la prueba de Cochran se encuentra en Statistics→Nonparametrics.

Interpretación Ejemplo:

C) Índices de Independencia

Información:

- Se fija el tamaño de la muestra y se clasifica a los participantes simultáneamente en función de las variables de interés.
- Cuando los variables varían aleatoriamente, es decir no están prefijadas, la Hipótesis fundamentalmente suele versar sobre la posible Independencia o Dependencia de las variables del estudio.
- Comparable al concepto de interacción: ¿Los cambios de la frecuencia provocados por una de las variables son alterados o modulados por otras variables del estudio?

Matriz de Frecuencias observadas (n_{jk})

	b_1	b_2	n_k
a_1	75	125	200
a_2	30	70	100
n_j	105	195	300

$p(jk)=n_{jk}/N$ Conjuntas

Matriz Probabilidades Esperadas (p_{jk})

	b_1	b_2	P_k
	0,250	0,417	0,667
	0,100	0,233	0,333
P_j	0,350	0,650	1,000

Hipótesis:

H_0: Las variables son independientes. $H_0 : \pi_{jk} = \pi_j \bullet \pi_k ; \forall_{j,k}$

H_1: Las variables son dependientes. $H_1 : \pi_{jk} \neq \pi_j \bullet \pi_k ; \forall_{j,k}$

Pruebas:

- Generales $\chi^2; G^2; W^2$
- Coeficiente de Correlación Rho ó PHI (preferibles a las anteriores).
- Corrección de Yates, Cochran y Upton y la Prueba exacta de Fischer; que introducen la corrección por continuidad (si alguna de las fr. esperadas es menor que 1 y menos del 20% de las mismas es mayor que 5).

Aclaraciones:

- Homogeneidad es como Regresión ya que se toman muestras de varias poblaciones y el objetivo es demostrar si la respuesta es similar en dichas poblaciones.
- Independencia es como Correlación ya que una población se clasifica en dos categorías o atributos y el objetivo es evaluar si la respuesta a uno de los atributos es o no independiente de la respuesta al otro.
- El hecho de que las variables sean independientes estadísticamente es equivalente a afirmar que su asociación es nula y a la inversa.

D) Índices de Concordancia

- Análogos a las medidas de asociación, pero aplicables cuando las variables se computan en función de acuerdos-desacuerdos o concordancias-discrepancias. Son de utilidad para estimar la fiabilidad interjueces.

Estadísticos:

- Índice Concordancia (ó Porcentaje Acuerdos).
- Coef. Kappa de Cohen.

Analizar → Estadísticos Descriptivos → Tablas de Contingencia → Filas: VarA; Columnas: VarB → Estadísticos *(seleccionar los asociados a Chi-Cuadrado, Correlaciones y Nominal)*→ Casillas → Frecuencias Esperadas → Aceptar (análisis).

En Statistica todo como en cuadro previo pero con los estadísticos: Fisher exact, Yates, McNemar (2x2) y Phi (2x2 tables) & Cramér's

Interpretación Ejemplo:

Analizar → Estadísticos Descriptivos → Tablas de Contingencia → Filas: VarA; Columnas: VarB → Estadísticos *(seleccionar los asociados a Kappa)*→ Casillas → Frecuencias Esperadas → Aceptar (análisis).

➢ *Ver también capítulo de Diseños categóricos en libro texto recomendado.*

E) Índices de Asociación

Información:

- Para cuantificar el grado de asociación cuando se piensa que las variables están relacionadas –que no son independientes-
- Se fija el tamaño de la muestra y se clasifica a los participantes simultáneamente en función de las variables de interés.
- Es comparable a RPE en Modelización o estimación del Efecto de Tratamiento.

Hipótesis:

H_0: La asociación entre variables es nula. $H_0 : \rho = 0$

H_1: La asociación entre variables es significativa. $H_1 : \rho \neq 0$

Estadísticos:

E.1.- Variables Nominales:

- Coeficiente PHI
- Coeficiente C de contingencia (y Ajustado)
- V de Cramer
- Coeficiente Rho -Equivale a Pearson y a PHI-
- Prueba de Mantel-Haenszel
- Lambda de Goodman y Kruskal (Basada en la medida de Concentración).
- Coeficiente de Incertidumbre (Basada en la medida de Entropía).
- Razón de Productos Cruzados (RPC) para cada sub-tabla 2x2 (como en Homogeneidad).
- Q. de Yule
- Coeficiente de Coaligación Y de Yule

E.2.- Variables ordinales

- Rangos de Spearman -Equivale a Pearson-
- Gamma de Goodman y Kruskal (1979)
- D de Somers (1962)
- Tau- B de Kendall (1979)
- Tau-C de Stuart (1953)
- Asociac. Parcial Tau- B de Kendall (Comparable a Correlac. Parcial)

E.3.- Casos Mixtos.

- Continua vs Dicotómica: Biserial-puntual.
- Dicotómica vs Dicotómica: PHI.
- Continua vs Dicotomizada –Normal-: Biserial.
- Dicotomizada –Normal- vs Dicotomizada –Normal-: Tetracórica.

Analizar → Estadísticos Descriptivos → Tablas de Contingencia → Filas: VarA; Columnas: VarB → Estadísticos *(seleccionar los asociados a Correlaciones, Nominal, Ordinal y Nominal por intervalo)*→ Casillas → Frecuencias Esperadas → Aceptar (análisis).

En Statistica todo como en cuadro previo pero con los estadísticos: Coefficient Phi (2x2 tables), Cramer's V and C, Kendall Tau, Gamma, Spearman R (rank order correlation), Sommer's d,

Interpretación Ejemplo:

6.5. Anexo. Tablas de contingencia: Estadísticos en los programas de análisis

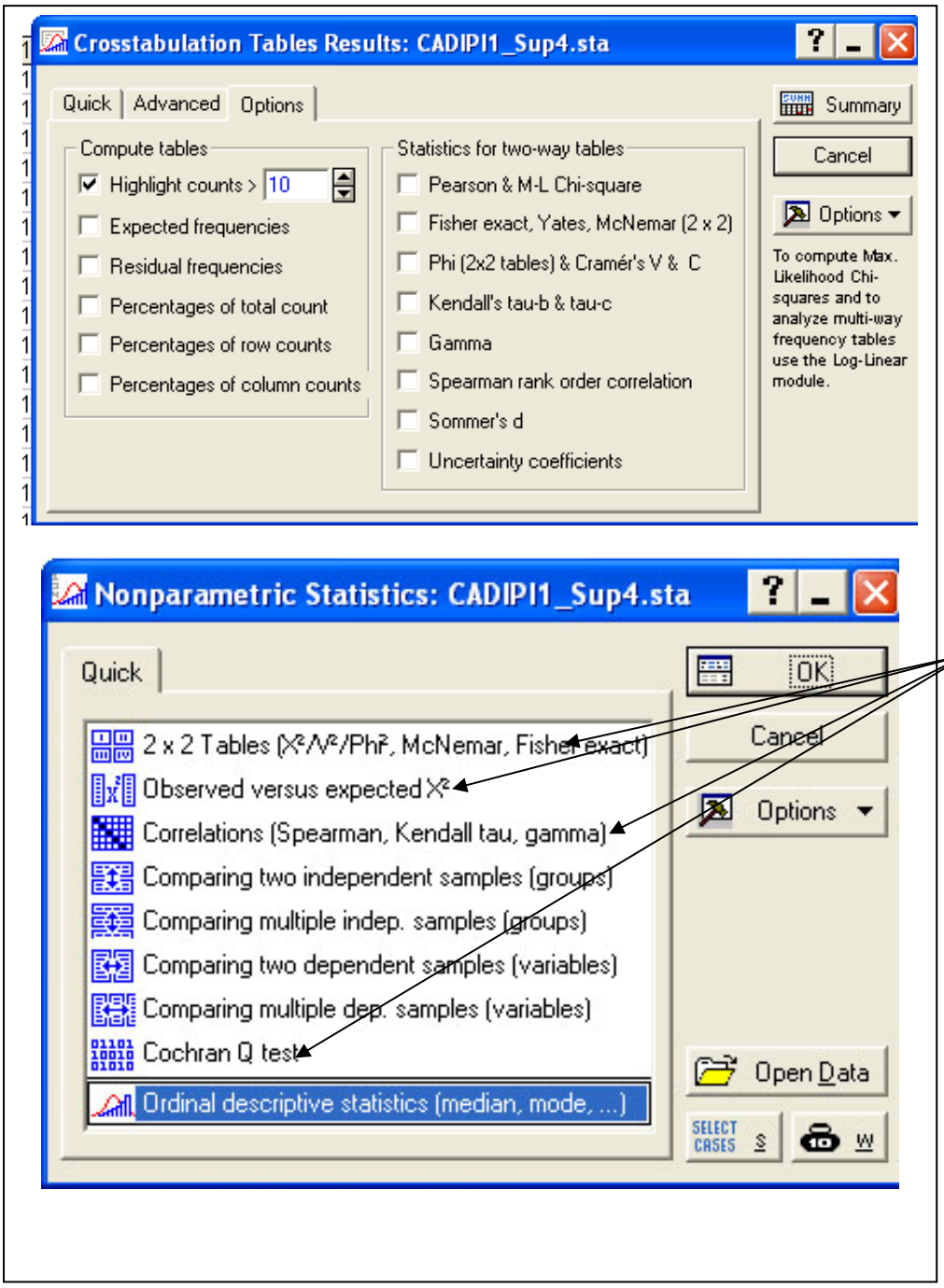
- **Chi-cuadrado**. Para las tablas con dos filas y dos columnas, seleccione Chi-cuadrado para calcular el chi-cuadrado de Pearson, el chi-cuadrado de la razón de verosimilitud, la prueba exacta de Fisher y el chi-cuadrado corregido de Yates (corrección por continuidad). Para las tablas 2 × 2, se calcula la prueba exacta de Fisher cuando una tabla (que no resulte de perder columnas o filas en una tabla mayor) presente una casilla con una frecuencia esperada menor que 5. Para las restantes tablas 2 × 2 se calcula el chi-cuadrado corregido de Yates. Para las tablas con cualquier número de filas y columnas, seleccione Chi-cuadrado para calcular el chi-cuadrado de Pearson y el chi-cuadrado de la razón de verosimilitud. Cuando ambas variables de tabla son cuantitativas, Chi-cuadrado da como resultado la prueba de asociación lineal por lineal.
- **Correlaciones**. Para las tablas en las que tanto las columnas como las filas contienen valores **ordenados**, Correlaciones da como resultado rho, el coeficiente de correlación de Spearman (sólo datos numéricos). La rho de Spearman es una medida de asociación entre órdenes de rangos. Cuando ambas variables de tabla (factores) son cuantitativas, Correlaciones da como resultado r, el coeficiente de correlación de Pearson, una medida de asociación lineal entre las variables.
- **Nominal**. Para los datos nominales (sin orden intrínseco, como católico, protestante o judío), puede seleccionar el coeficiente Phi y V de Cramér, el Coeficiente de contingencia, Lambda (lambdas simétricas y asimétricas y tau de Kruskal y Goodman) y el Coeficiente de incertidumbre.
    - **Coeficiente de contingencia**. Medida de asociación basada en chi-cuadrado. El valor siempre está comprendido entre 0 y 1. El valor 0 indica que no hay asociación entre la fila y la columna. Los valores cercanos a 1 indican que hay gran relación entre las variables. El valor máximo posible depende del número de filas y columnas de la tabla.
    - **Phi y V de Cramer**. Phi es una medida de asociación basada en chi-cuadrado que conlleva dividir el estadístico chi-cuadrado por el tamaño muestral y calcular la raíz cuadrada del resultado. V de Cramer es una medida de asociación basada en chi-cuadrado.
    - **Lambda**. Medida de asociación que refleja la reducción proporcional en el error cuando se utilizan los valores de la variable independiente para pronosticar los valores de la variable dependiente. Un valor igual a 1 significa que la variable independiente pronostica perfectamente la variable dependiente. Un valor igual a 0 significa que la variable independiente no ayuda en absoluto a pronosticar la variable dependiente.
    - **Coeficiente de incertidumbre**. Medida de asociación que indica la reducción proporcional del error cuando los valores de una variable se emplean para pronosticar los valores de la otra variable. Por ejemplo, un valor de 0,83 indica que el conocimiento de una variable reduce en un 83% el error al pronosticar los valores de la otra variable. SPSS calcula tanto la versión simétrica como la asimétrica del coeficiente de incertidumbre.
- **Ordinal**. Para las tablas en las que tanto las filas como las columnas contienen valores ordenados, seleccione Gamma (orden cero para tablas de doble clasificación y condicional para tablas cuyo factor de clasificación va de 3 a 10), Tau-b de Kendall y Tau-c de Kendall. Para pronosticar las categorías de columna de las categorías de fila, seleccione d de Somers.
    - **Gamma**. Medida de asociación simétrica entre dos variables ordinales cuyo valor siempre está comprendido entre menos 1 y 1. Los valores próximos a 1, en valor absoluto, indican una fuerte relación entre las dos variables. Los valores próximos a cero indican que hay poca o ninguna relación entre las dos variables. Para las tablas de doble clasificación, se muestran las gammas de

orden cero. Para las tablas de tres o más factores de clasificación, se muestran las gammas condicionales.

- **d de Somers**. Medida de asociación entre dos variables ordinales que toma un valor comprendido entre -1 y 1. Los valores próximos a 1, en valor absoluto, indican una fuerte relación entre las dos variables. Los valores próximos a cero indican que hay poca o ninguna relación entre las dos variables. La d de Somers es una extensión asimétrica de gamma que difiere sólo en la inclusión del número de pares no empatados en la variable independiente. También se calcula una versión simétrica de este estadístico.
- **Tau-b de Kendall**. Medida no paramétrica de la correlación para variables ordinales o de rangos que tiene en consideración los empates. El signo del coeficiente indica la dirección de la relación y su valor absoluto indica la magnitud de la misma, de tal modo que los mayores valores absolutos indican relaciones más fuertes. Los valores posibles van de -1 a 1, pero un valor de -1 o +1 sólo se puede obtener a partir de tablas cuadradas.
- Tau-c de Kendall. Medida no paramétrica de asociación para variables ordinales que ignora los empates. El signo del coeficiente indica la dirección de la relación y su valor absoluto indica la magnitud de la misma, de tal modo que los mayores valores absolutos indican relaciones más fuertes. Los valores posibles van de -1 a 1, pero un valor de -1 o +1 sólo se puede obtener a partir de tablas cuadradas.

- **Nominal por intervalo**. Cuando una variable es categórica y la otra es cuantitativa, seleccione Eta. La variable categórica debe codificarse numéricamente.
  - **Eta**. Medida de asociación cuyo valor siempre está comprendido entre 0 y 1. El valor 0 indica que no hay asociación entre las variables de fila y de columna. Los valores cercanos a 1 indican que hay gran relación entre las variables. Eta resulta apropiada para una variable dependiente medida en una escala de intervalo (por ejemplo, ingresos) y una variable independiente con un número limitado de categorías (por ejemplo, sexo). Se calculan dos valores de eta: uno trata la variable de las filas como una variable de intervalo; el otro trata la variable de las columnas como una variable de intervalo.
- **Kappa**. La kappa de Cohen mide el acuerdo entre las evaluaciones de dos jueces cuando ambos están valorando el mismo objeto. Un valor igual a 1 indica un acuerdo perfecto. Un valor igual a 0 indica que el acuerdo no es mejor que el que se obtendría por azar. Kappa sólo está disponible para las tablas cuadradas (tablas en las que ambas variables tienen el mismo número de categorías y utilizan los mismos valores de categoría).
- **Riesgo**. Para las tablas 2 x 2, medida del grado de asociación entre la presencia de un factor y la ocurrencia de un evento. Si el intervalo de confianza para el estadístico incluye un valor de 1, no se podrá asumir que el factor está asociado con el evento. Cuando la ocurrencia del factor es poco común, se puede utilizar la razón de ventajas como estimación del riesgo relativo.
- **McNemar**. Prueba no paramétrica para dos variables dicotómicas relacionadas. Contrasta los cambios en las respuestas utilizando la distribución de chi-cuadrado. Es útil para detectar cambios en las respuestas debidas a la intervención experimental en los diseños del tipo "antes-después". Para tablas cuadradas mayores, se utiliza la prueba de simetría de McNemar-Bowker.
- Estadísticos de **Cochran** y de Mantel-Haenszel. Estos estadísticos se pueden utilizar para contrastar la independencia entre una variable dicotómica de factor y una variable dicotómica de respuesta, condicionada por los patrones en las covariables, los cuales vienen definidos por la variable o variables de las capas (variables de control). Tenga en cuenta que mientras que otros estadísticos se calculan capa por capa, los estadísticos de Cochran y Mantel-Haenszel se calculan una sola vez para todas las capas.

En Statistica las opciones principales son las siguientes:



- Pearson Chi-square
- Maximum-Likelihood (M-L)Chi-square
- Fisher Exact Test
- Yates Correction
- McNemar Chi-square
- Coefficient Phi (2x2 tables)
- Cramer's V and C
- Kendall Tau
- Gamma
- Spearman R (rank order correlation)
- Sommer's d
- Uncertainty Coefficients