

ANÁLISIS CUANTITATIVO DE DATOS EN CIENCIAS SOCIALES CON EL SPSS (I)

*Tablas de contingencia y pruebas de
asociación*



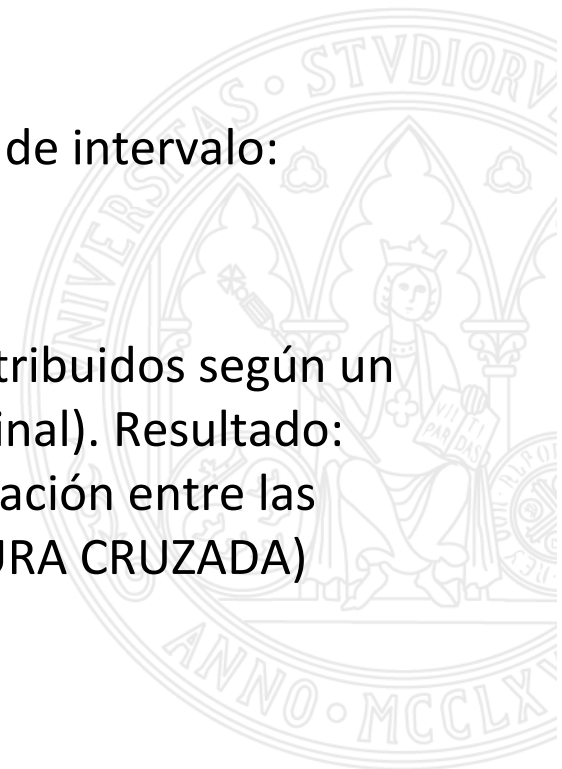
UNIVERSIDAD DE
MURCIA

Francisca José Serrano Pastor
Pedro A. Sánchez Rodríguez

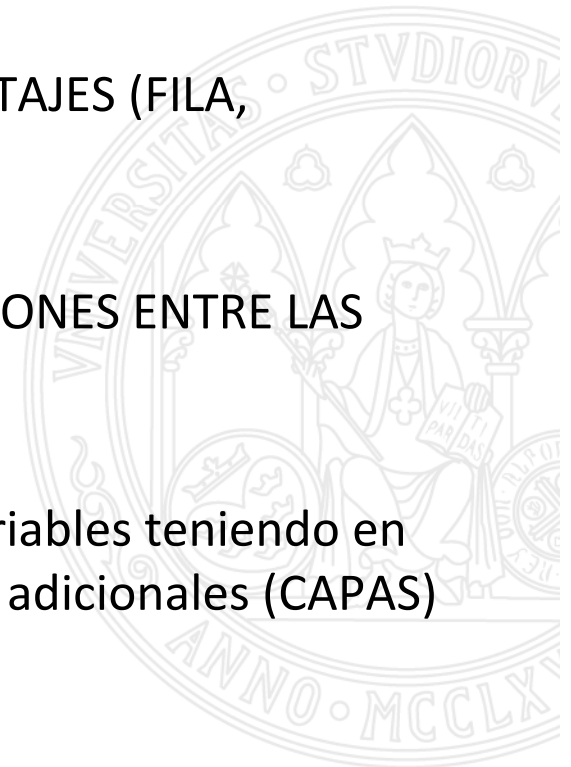


TABLAS DE CONTINGENCIA

- Implica siempre a variables cualitativas, categóricas o nominales, u ordinales con pocos valores: **nominal*nominal**, ordinal*ordinal; nominal*ordinal
- También puede implicar a una variable nominal y otra de intervalo: nominal*intervalo
- Los datos se organizan en tablas de doble entrada, distribuidos según un criterio de clasificación (variable nominal/variable ordinal). Resultado: frecuencias y porcentajes → Observar asociación o relación entre las categorías o valores de las variables implicadas (LECTURA CRUZADA)



- Procedimiento de cálculo con SPSS:
- ANALIZAR → ESTADÍSTICOS DESCRIPTIVO → TABLAS DE CONTINGENCIA...
- TABLAS CON FRECUENCIAS (POR DEFECTO) Y PORCENTAJES (FILA, COLUMNA, TOTAL) EN **CASILLAS**...
- MEJOR PORCENTAJES (PERMITEN HACEN COMPARACIONES ENTRE LAS DIFERENTES CELDAS)
- TABLAS DE CONTINGENCIA SEGMENTADAS: Cruzar variables teniendo en cuenta los niveles o categorías de una o más variables adicionales (CAPAS)
- EJERCICIO PRÁCTICO (GATOS.SAV)



CONTRASTE DE HIPÓTESIS Y PRUEBAS DE ASOCIACIÓN

- No basta con observar las diferencias entre porcentajes
- Se debe utilizar alguna medida de asociación acompañada de su correspondiente prueba de significación para determinar si dos variables se encuentran relacionadas (asociadas) de forma estadísticamente significativa a un nivel crítico α , y el grado de relación entre ambas variables.
- Mayor interés: obtener un índice que describa la fuerza, la dirección y la naturaleza de la asociación (no interpretable en términos de relación causa-efecto) entre las variables. Éste es el objetivo de las MEDIDAS DE ASOCIACIÓN.

ANTES, debemos recordar algunos conceptos básicos:

Hipótesis Estadísticas: En Estadística Inferencial es habitual la aplicación de pruebas de significación estadística, que permiten determinar la existencia de diferencias entre grupos, la dependencia entre variables, etc. Tienen como punto de partida el establecimiento de las hipótesis estadísticas. Son dos:

* Hipótesis nula (H_0): La diferencia entre los grupos es estadísticamente nula; es debida al azar. Predice que la VI no tiene efecto sobre la VD para la población. Por ejemplo:

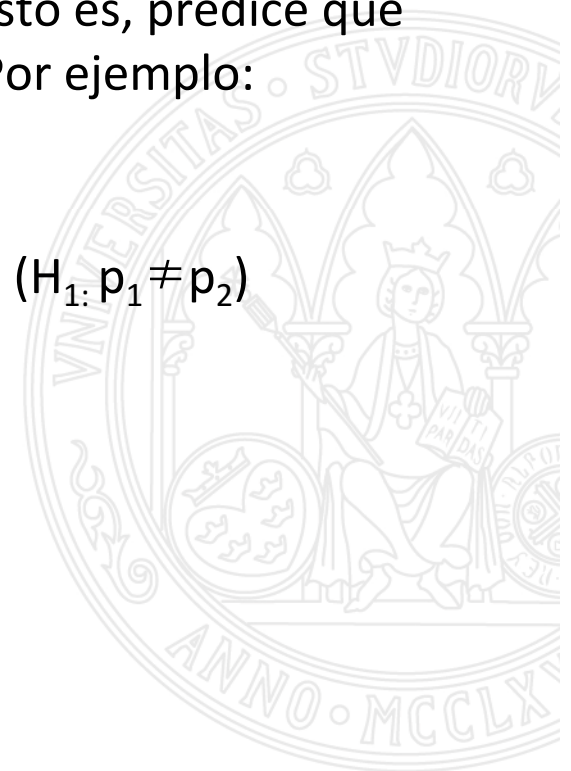
$$H_0: p_1 = p_2$$

* Hipótesis alterna (H_1) Las diferencias observadas no se deben al azar; son estadísticamente significativas entre los grupos; esto es, predice que la VI tendrá un efecto sobre la VD para la población. Por ejemplo:

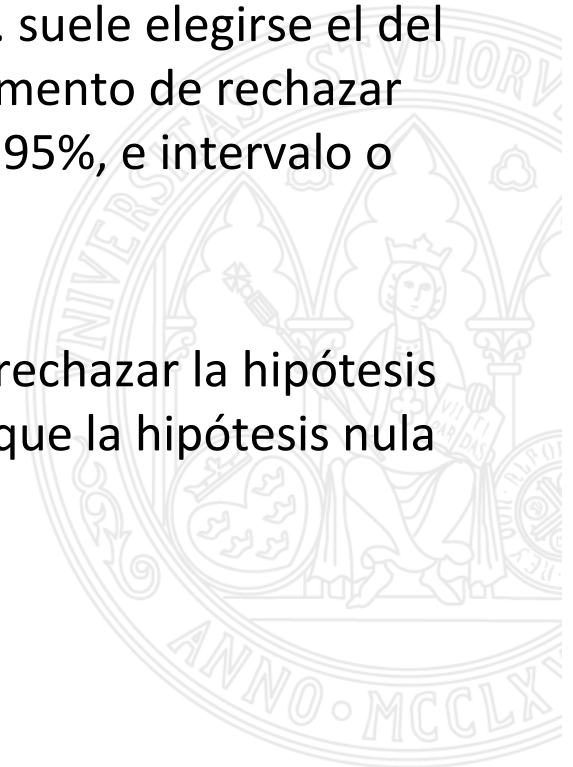
$$H_1: p_1 \neq p_2$$

- Hipótesis alterna no direccional o bidireccional ($H_1: p_1 \neq p_2$)
- Hipótesis direccional ($H_1: p_1 > p_2$)

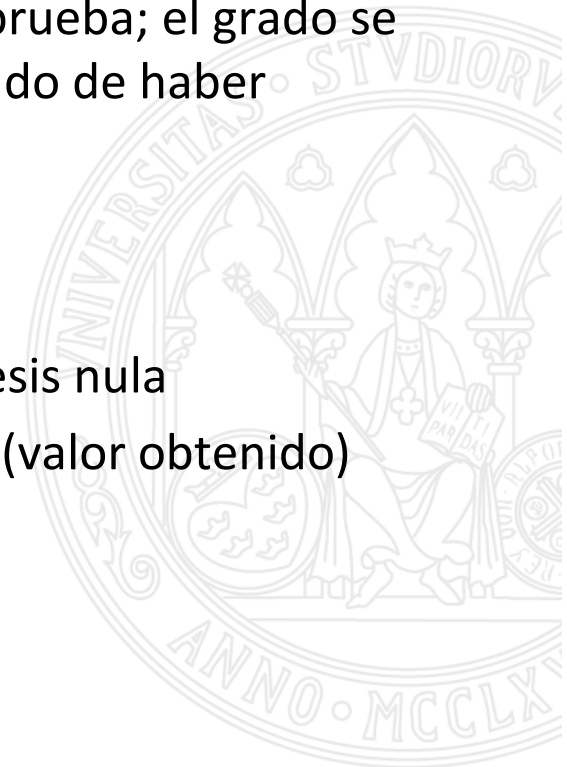
Se somete a comprobación siempre la Hipótesis Nula



- ❖ Nivel de significación (α): Riesgo de error que se está dispuesto a asumir en caso de rechazar la hipótesis nula. En CC.SS. suele elegirse el del 0.05; es decir, con un 5% de errores posibles en el momento de rechazar la Hipótesis nula (intervalo o margen de confianza del 95%, e intervalo o margen de error del 5%).
- ❖ Grado de significación (p): Probabilidad de error al rechazar la hipótesis nula. Cuanto más pequeña es “p”, más probable será que la hipótesis nula sea falsa. En SPSS aparece como “Sig.”



- ❖ Nivel y Grado de significación están relacionados.
- ❖ El nivel se establece *a priori*, antes de aplicar la prueba; el grado se calcula a posteriori, o sea, cuando se conoce el resultado de haber aplicado una prueba de significación estadística.
- ❖ Relación entre ambos conceptos:
 - Si $p > \alpha$ Nada se opone en aceptar la hipótesis nula
 - Si $p < \alpha$ Se rechaza la hipótesis nula con $p =$ (valor obtenido)

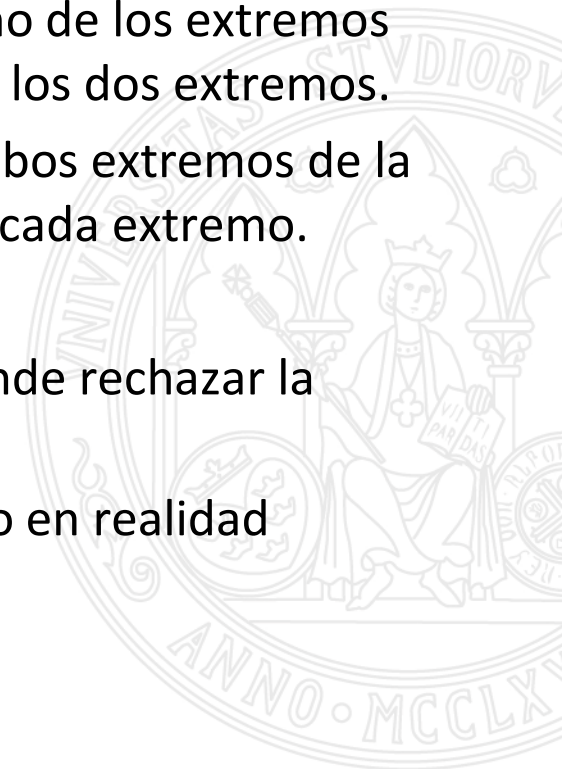


Dos tipos de contraste de hipótesis:

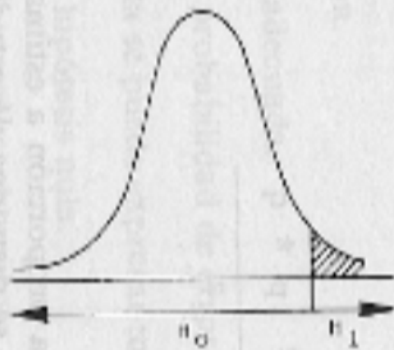
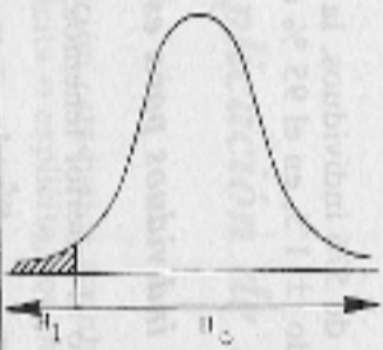
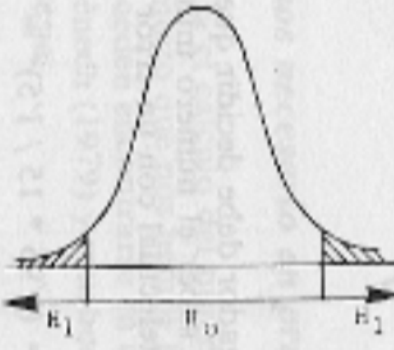
- ❖ Contraste unilateral: Región crítica situada en uno de los extremos de la distribución. El riesgo α se encuentra en uno de los dos extremos.
- ❖ Contraste bilateral: Región crítica ubicada en ambos extremos de la distribución. El riesgo α se divide en dos, la mitad en cada extremo.

Región crítica: Zona de la distribución en la que corresponde rechazar la hipótesis nula.

Riesgo α : Error tipo I. Se rechaza la hipótesis nula, siendo en realidad verdadera.



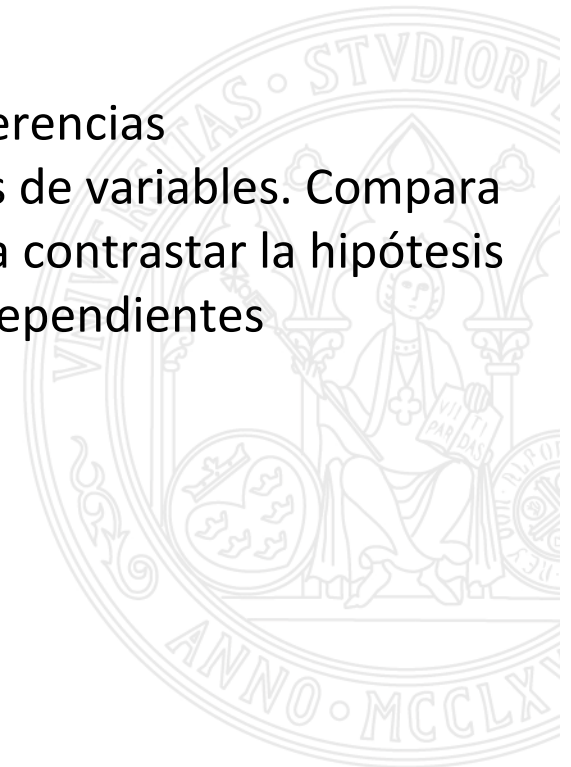
Pruebas Bilaterales y Unilaterales

H_1	$>$	$<$	\neq
Región crítica	Derecha	Izquierda	Ambos lados
Prueba	Unilateral	Unilateral	Bilateral
Representación gráfica			
Ejemplo (al rechazar H_0)	La VL de los sujetos que han seguido un PEL es superior a μ .	El índice de morbilidad de los no fumadores es inferior a los fumadores.	El CI de los alumnos de carreras de Ciencias es diferente del de Letras.
Hipótesis	$H_0 : \bar{X} = \mu$ $H_1 : \bar{X} > \mu$	$H_0 : p_1 = p_2$ $H_1 : p_1 < p_2$	$H_0 : \bar{X}_1 = \bar{X}_2$ $H_1 : \bar{X}_1 \neq \bar{X}_2$



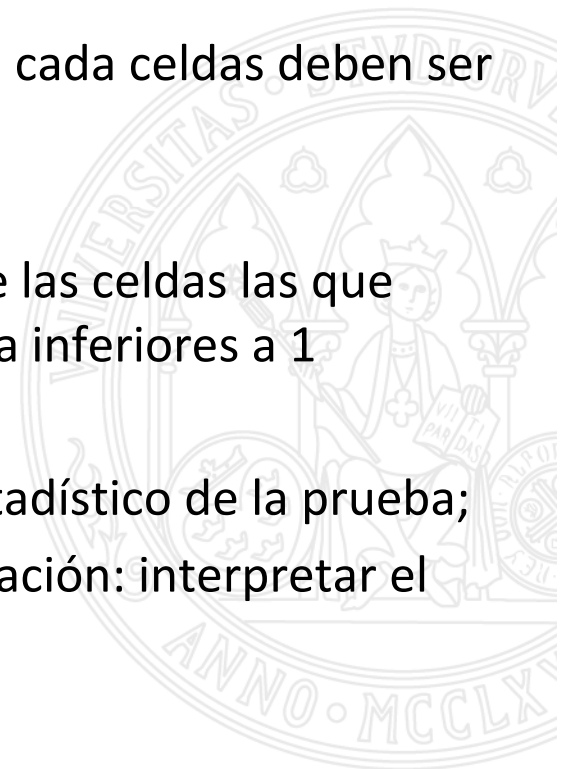
JI CUADRADO DE PEARSON

- La prueba de contraste de hipótesis utilizada que suele realizarse
- Finalidad prueba Ji Cuadrado: comprobar si existen diferencias estadísticamente significativas entre dos distribuciones de variables. Compara frecuencias observadas con frecuencias esperadas para contrastar la hipótesis de que las dos variables categóricas implicadas son independientes



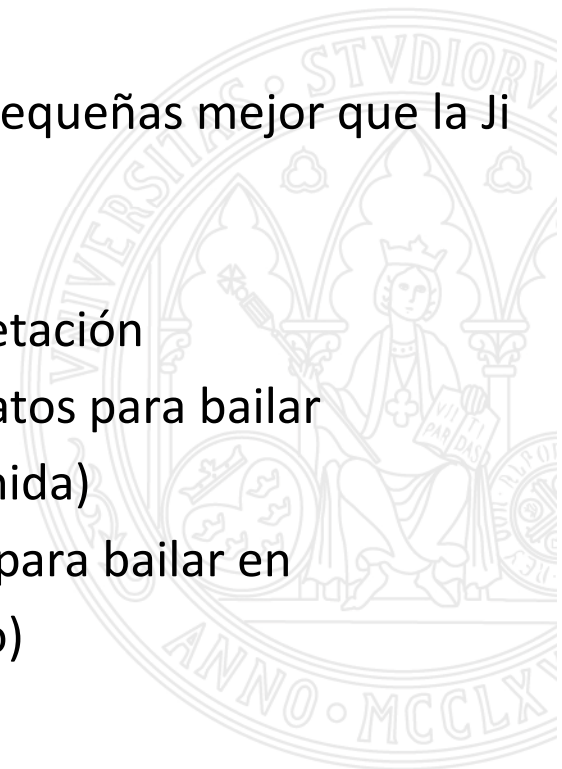
JI CUADRADO DE PEARSON

- Requisito de aplicación prueba Ji Cuadrado:
 - En tablas 2x2: las frecuencias teóricas o esperadas en cada celdas deben ser iguales o superiores a 5
 - En tablas más grandes: se acepta que sean un 20% de las celdas las que tengan frecuencias esperadas inferiores a 5, pero nunca inferiores a 1
 - Si no se cumple este presupuesto → Menos poder estadístico de la prueba; la prueba puede no detectar un efecto real. Recomendación: interpretar el **Estadístico exacto de Fisher**



JI CUADRADO DE PEARSON

- Razón de verosimilitud: igual que Ji Cuadrado de Pearson. Carácter aditivo para modelos log lineales.
- Corrección por continuidad de Yates: para muestras pequeñas mejor que la Ji Cuadrado (no unanimidad)
- ACTIVIDAD GATOS.SAV: Cálculo Ji Cuadrado e interpretación
 - H_0 : No existen diferencias en el aprendizaje de los gatos para bailar en función del tipo de recompensa (afecto o comida)
 - H_1 : Existen diferencia en el aprendizaje de los gatos para bailar en función del tipo de recompensa (comida o afecto)



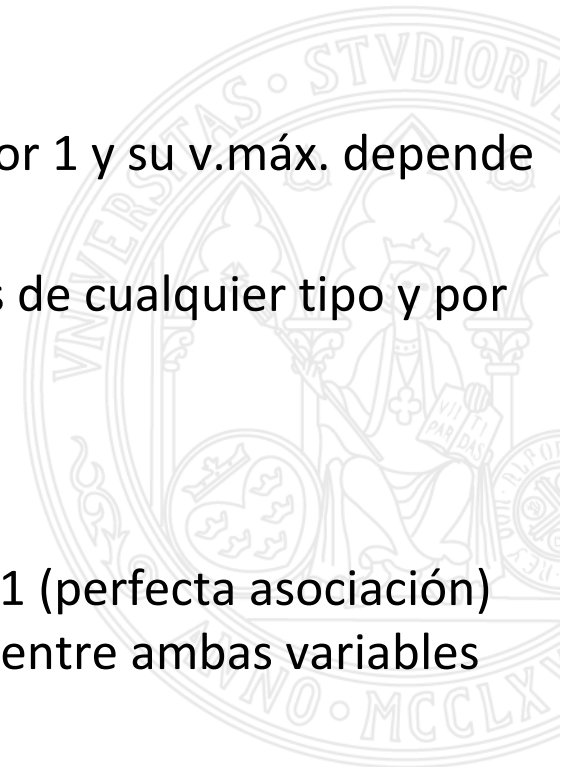
MEDIDAS DE ASOCIACIÓN QUE ACOMPAÑAN A LA JI CUADRADO DE PEARSON (nominales)

Permiten obtener un índice que describa la fuerza, la dirección y la naturaleza de la asociación (no interpretable en términos de relación causa-efecto) entre las variables.

- Dos tipos de medidas de asociación:
 - Simétricas
 - Asimétricas o direccionales.
- SIMÉTRICAS: Se interpretan como el índice del grado o magnitud de la asociación entre las dos variables, pero sin supeditar una a la otra.
- ASIMÉTRICAS: Proporcionan una medida del grado en que se reduce el error al pronosticar una variable (VD o criterio) a partir de la otra (VI o predictora). La persona investigadora es quien las define previamente. Objetivo que se persigue: determinar en qué medida la VD puede ser “explicada” por la VI.

MEDIDAS DE ASOCIACIÓN QUE ACOMPAÑAN A LA JI CUADRADO DE PEARSON (nominales)

- Medidas Simétricas:
 - Coeficiente de contingencia: No puede alcanzar el valor 1 y su v.máx. depende del tamaño de la muestra
 - V de Cramer: la recomendada en principio para tablas de cualquier tipo y por no presentar problemas su interpretación
 - Phi: 2x2; con más supera el valor 1
 - Interpretación: coeficiente entre 0 (independencia) y 1 (perfecta asociación) en términos de la magnitud o fuerza de la asociación entre ambas variables



MEDIDAS DE ASOCIACIÓN QUE ACOMPAÑAN A LA JI CUADRADO DE PEARSON (nominales)

Medidas Direccionales:

- Coeficiente de incertidumbre: %
- Lambda de Goodman y Kruskal
- Tau de Goodman y Kruskal
- Interpretación: “reducción proporcional del error” que se consigue al tener en cuenta una variable (VI o predictora) en la predicción de la otra (VD o criterio)
- Actividad con Gatos.sav: cálculo medidas de asociación e interpretación:

Pregunta 1: *¿Cuál es la magnitud o fuerza entre las variables “tipo de recompensa” y “aprender a bailar”?*

Pregunta 2: *¿Qué error se reduce cuando predecimos el aprendizaje del baile (VD) en función del tipo de recompensa?*

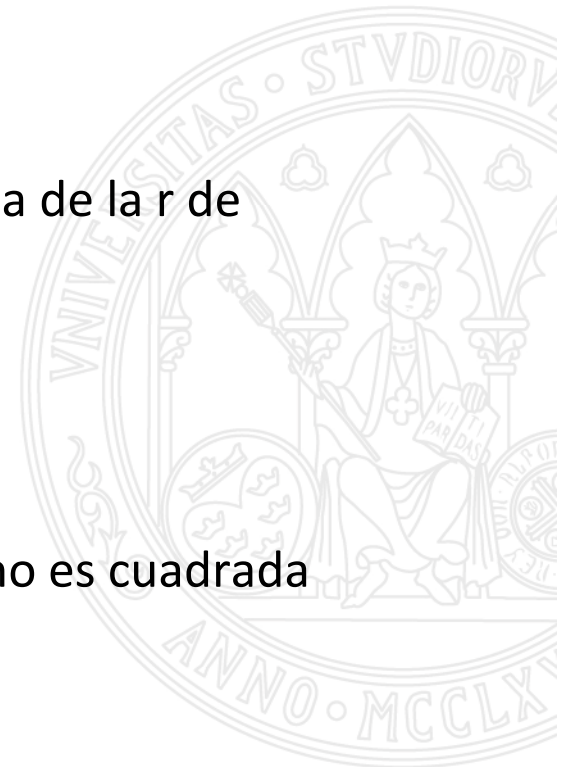
CALCULAR EL TAMAÑO DEL EFECTO JI CUADRADO DE PEARSON (nominales)

- Aunque la V de Cramer calcula un tamaño del efecto adecuado y es muy fácil de interpretar, también en tablas 2x2...
- Puede ser más útil medir el efecto con el cociente de probabilidad; esto es, la probabilidad de que un hecho ocurra dividido por la probabilidad de que este hecho no ocurra
- Probabilidad = $P(\text{ocurre}) / P(\text{no ocurre})$
- Actividad con Gatos.sav: cálculo tamaño efecto e interpretación
Pregunta: ¿Cuántas veces es superior el efecto de que los gatos cuando son entrenados con comida aprenden a bailar en mayor proporción que cuando lo son con afecto?

MEDIDAS DE ASOCIACIÓN QUE ACOMPAÑAN A LA CORRELACIÓN DE PEARSON (ordinales)

Son pruebas de asociación, pero añaden además dirección: +/-

- Pueden utilizarse con variables de intervalo
- **Medidas Simétricas:**
 - Correlación de Spearman (versión no paramétrica de la r de Pearson). OK para variables ordinales.
 - Tau-b de Kendall
 - Tau- c de Kendall: la más adecuada
 - Gamma de Goodman y Kruskal
 - Los otros no alcanzan el valor 1 cuando la tabla no es cuadrada
- **Medidas Direccionales:**
 - d de Somers



MEDIDAS DE ASOCIACIÓN (nominal por intervalo)

- Cuando se cruzan variables nominales u ordinales con variables de intervalo o de razón/proporción
- Medida Simétrica única que proporciona:
 - Coeficiente Eta: proporción de la varianza de la variable cuantitativa explicada por la variable cualitativa
- Para evitar errores: mejor colocar en “Filas” la variable de intervalo.

