09

PSICOLOGÍA EXPERIMENTAL

Verónica Ventero Portelas
Psicóloga Especialista en Psicología Clínica.
FEA Psicología Clínica Complejo Hospitalario de Toledo.

Juan Antequera Iglesias
Psicólogo Especialista en Psicología Clínica.
FEA Psicología Clínica Hospital Virgen de la Misericordia de Toledo.

Laura Hernangómez Criado
Doctora en Psicología.
Psicóloga Especialista en Psicología Clínica.
Psicoterapeuta acreditada por ASEPCO.
FEA Psicología Clínica Hospital Complejo Hospitalario de Toledo. Hospital Virgen de la Salud. SESCAM.

Carla Tejero Berzosa
Psicóloga Especialista en Psicología Clínica.
FEA Psicología Clínica Hospital Universitario del Tajo.

Rosa Ruiz Girón
Residente de Psicología Clínica en el Hospital Universitario de La Paz. Madrid.

TODO EL MATERIAL,
EDITADO Y PUBLICADO
POR EL CENTRO DOCUMENTACIÓN
DE ESTUDIOS Y OPOSICIONES,
ES ÚNICO Y EXCLUSIVO
DE NUESTRO CENTRO.

ISBN obra completa: 978-84-18241-33-8
ISBN: 978-84-18241-36-9
Depósito Legal: M-3222-2022
EDITA Y DISTRIBUYE: CEDE

EDICIÓN: febrero 2022

ES PROPIEDAD DE:

© CEDE

CENTRO DOCUMENTACIÓN
DE ESTUDIOS Y OPOSICIONES

PRESENTACIÓN

La Psicología Experimental supone un compendio de las principales estrategias y herramientas que se utilizan para la realización de investigaciones y la interpretación de los resultados de las mismas.

Estos contenidos se estudian en la mayoría de carreras universitarias, realizándose especificaciones concretas relacionadas con cada área. En el grado de Psicología se dedican varias asignaturas (entre 3 y 4) al conocimiento de la metodología, el análisis de datos, la estadística y la psicometría; por lo que suponen una parte muy importante dentro de las competencias a adquirir por un psicólogo. Además, una adecuada adquisición de estos contenidos supone una ventaja a la hora de llevar a cabo el trabajo fin de carrera.

Por otro lado, durante la realización del ejercicio de la profesión de Psicólogo Clínico, este tipo de herramientas y de conocimientos acerca de la investigación científica son fundamentales para poder llevar a cabo cualquier tipo de estudio o publicación científica de cierto nivel. Aún más si pensamos en la realización de un máster y su trabajo fin de máster o en un doctorado.

Por todo ello, los conceptos relacionados con la Psicología Experimental han aparecido de manera constante en el examen PIR. El número de preguntas ha ido disminuyendo a medida que se ha reducido el número total de preguntas del examen PIR, pasando de una media de 20 preguntas cuando este constaba de 260 preguntas, a una media de 9 en los últimos años, en consonancia con los cambios en el número de preguntas del examen.

Podría parecer que el peso de la asignatura no es muy significativo en relación a otras áreas. Es cierto que el número de preguntas no es muy alto, pero también lo es, que las preguntas de esta asignatura pueden hacer que se consiga una mayor diferencia del resto de participantes en la convocatoria y, por lo tanto, se facilite la obtención de plaza.

La asignatura de Psicología Experimental está dividida en tres grandes bloques:

- *Fundamentos teóricos y diseños experimentales.*
- *Estadística.*
- *Psicometría.*

En cuanto al peso de cada uno de ellos dentro de la misma, nos encontramos con que la parte más relevante en cuanto a número de preguntas es el bloque dedicado a la Estadística; seguido de la parte de Metodología, y con un peso menor Psicometría.

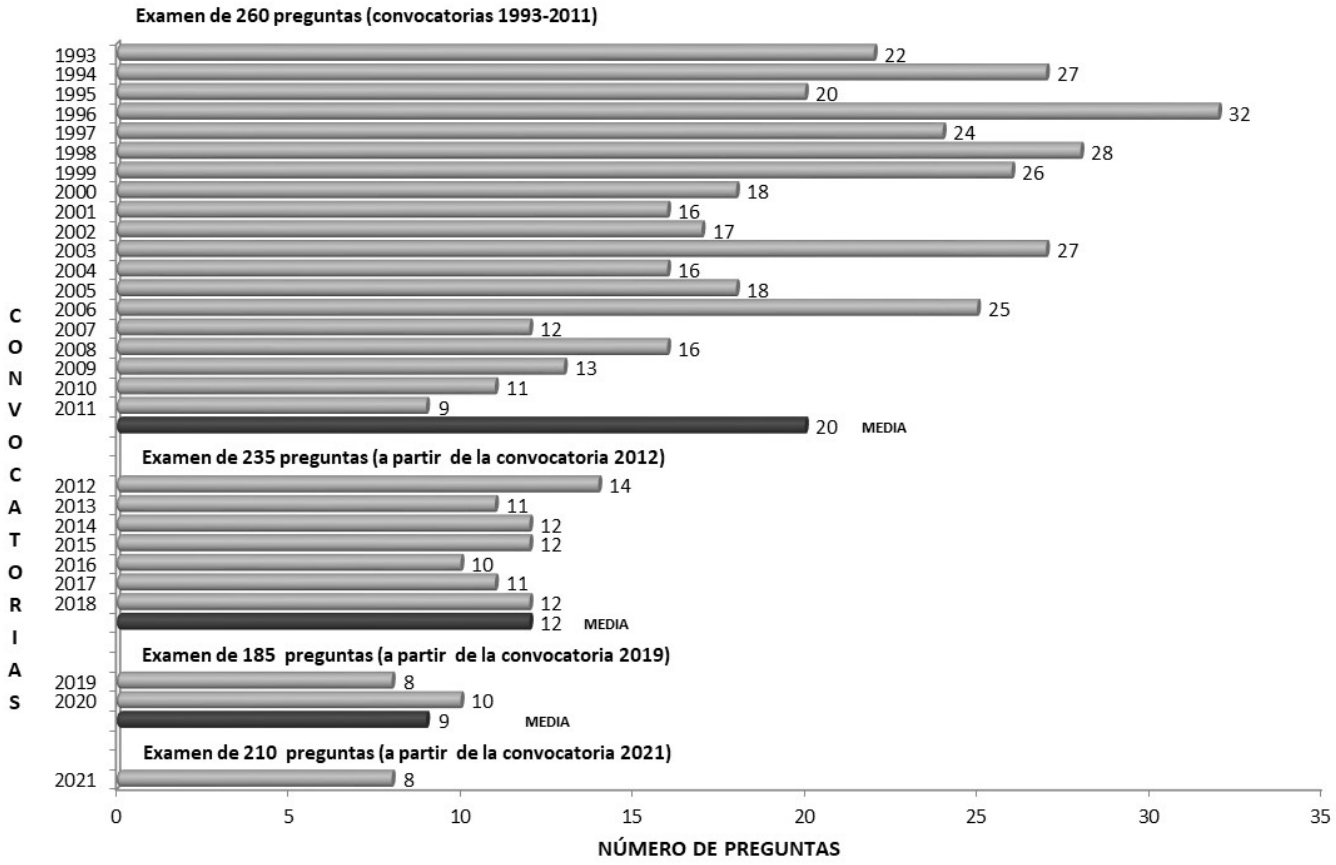
Las preguntas del área de Psicología Experimental suelen versar sobre conceptos principales y sus características generales; aunque en algunas convocatorias se ha pedido la interpretación de resultados mediante un ejemplo. Si bien es cierto que en los últimos años la tendencia ha sido a aumentar la difi-

cultad de las preguntas, una ventaja es que en esta asignatura el aumento de la dificultad no supone un mayor número de conocimientos y por tanto la consecuente ampliación del temario, sino simplemente una mayor comprensión de los conceptos básicos que manejamos.

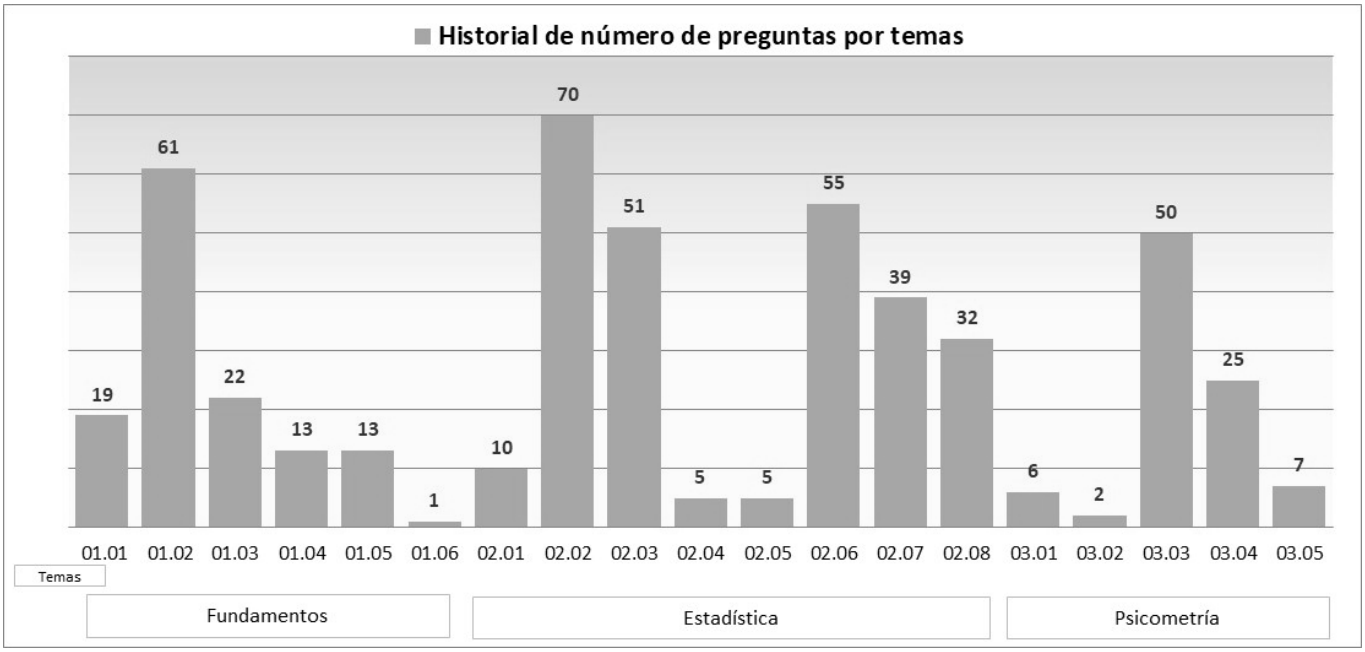
Dentro de la parte de Estadística, en el Manual de CEDE, los temas que aglutinan un mayor número de preguntas son los referentes a la estadística descriptiva –de una y de dos variables– y a la parte de estadística inferencial –fundamentos básicos y pruebas paramétricas, con una tendencia actual al alza en recientes convocatorias de este último apartado–. En el área dedicada a los Fundamentos teóricos y diseños experimentales, los temas a lo que se ha hecho referencia con más frecuencia son: características y clasificación del método científico, la intervención, manipulación y control de las variables, el diseño experimental y los diseños cuasi-experimentales y N = 1. Y por último, los temas referentes al cálculo del coeficiente de fiabilidad y a la validez, han aparecido de manera relevante dentro de la parte de Psicometría.

Para un buen estudio de la asignatura, es recomendable realizar una lectura comprensiva del Manual, poniendo especial énfasis en los conceptos principales que se apuntan dentro de cada tema. No es necesario un conocimiento profundo de cada uno de los aspectos que aparecen en el Manual, ni tampoco la memorización de fórmulas. Una vez que se tiene un conocimiento general de la asignatura, es momento de pasar a una fase en la que el objetivo será la fijación de los contenidos principales y de la manera en que han sido preguntados en las convocatorias PIR; para ello, será muy útil relacionar unos con otros y utilizar cualquier tipo de regla mnemotécnica. La realización de las preguntas de convocatoria y de los simulacros y exámenes de área es muy útil para aumentar el entrenamiento en la respuesta a preguntas tipo, y para reconocer lo más relevante a la hora de repasar. Por último, en las clases presenciales y en la web nos centraremos en el aumento de la comprensión sobre algunos procedimientos e interpretaciones, mediante explicaciones y ejercicios de ampliación; ofreciendo así una nueva forma de estudio del temario de Psicología Experimental.

EVOLUCIÓN DEL NÚMERO DE PREGUNTAS POR CONVOCATORIA



HISTORIAL DEL NÚMERO DE PREGUNTAS POR TEMA



09 PSICOLOGÍA EXPERIMENTAL

Índice general de temas

09.01. FUNDAMENTOS TEÓRICOS Y DISEÑOS EXPERIMENTALES

Página 102

09 01 05 DISEÑOS N = 1 Y METODOLOGÍA DE ENCUESTAS

1. Diseños N = 1
   1.1. Tipos de diseños N = 1
      1.1.1. Análisis de datos
2. Metodología de encuestas
   2.1. Introducción
   2.2. Tipos de encuestas según dimensión temporal
      2.2.1. Transversales
      2.2.2. Longitudinales
      2.2.3. Longitudinal-secuenciales
      2.2.4. Encuestas longitudinales retrospectivas
   2.3. La calidad de la encuesta

Página 113

09 01 06 LA OBSERVACIÓN Y LA INVESTIGACIÓN CUALITATIVA

1. La observación
2. Investigación cualitativa
   2.1. Definición, características y metodología
   2.2. Comparación con metodología cuantitativa
      2.2.1. Diferencias entre investigación cuantitativa y cualitativa
   2.3. Las técnicas cualitativas
   2.4. Los datos cualitativos

09.02. ESTADÍSTICA

Página 125

09 02 01 INTRODUCCIÓN A LA ESTADÍSTICA EN PSICOLOGÍA

1. Presentación
2. Concepto de medida en psicología
   2.1. La escala de medida
3. Vocabulario básico en estadística
   3.1. Población
   3.2. Muestra
   3.3. Parámetro
   3.4. Estadístico
4. Concepto de estadística
   4.1. Dos clases de estadística

Página 132

09 02 02 ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE UNA SOLA VARIABLE

1. Introducción
2. Organización de los datos
   2.1. Concepto y tipos de variable
   2.2. Modalidades y clases
   2.3. Distribución de frecuencias
   2.4. Diagrama de tallo y hojas
   2.5. Representación gráfica de la variabilidad: Diagrama de caja y bigotes
3. Estadísticos de tendencia central
   3.1. Media aritmética
   3.2. Mediana
   3.3. Moda
   3.4. Media, mediana, moda y asimetría
   3.5. Apuntamiento o curtosis
4. Estadísticos de posición: los cuantiles
   4.1. Cuartiles
   4.2. Deciles
   4.3. Percentiles
5. Estadísticos de variabilidad y dispersión
   5.1. Desviación media
   5.2. La varianza
   5.3. Amplitud total
   5.4. Amplitud semi-intercuartil
   5.5. Coeficiente de variación
6. Puntuaciones directas, diferenciales y típicas
   6.1. Puntuación directa
   6.2. Puntuación diferencial
   6.3. Puntuación típica
   6.4. Otras transformaciones de las puntuaciones
   6.5. Interpretación de puntuaciones directas, diferenciales y típicas
   6.6. La curva normal

Página 154

09 02 03 ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE DOS VARIABLES

1. Introducción
2. Distribución conjunta de frecuencias
   2.1. Covarianza
3. Relación lineal entre dos variables
   3.1. Covarianza
   3.2. Coeficiente de correlación de Pearson
   3.3. La ecuación de regresión
   3.4. El coeficiente de determinación y la recta de regresión
4. Relación curvilínea entre dos variables
   4.1. Propiedades de la razón de correlación
5. Relación entre variables ordinales
   5.1. Coeficiente de correlación de Spearman
   5.2. Coeficiente de correlación de Kendall
   5.3. Coeficiente de correlación de Goodman y Kruskal
6. Relación entre variables nominales
   6.1. Coeficiente Q de Yule
   6.2. Coeficiente χ^2
   6.3. Coeficiente C de contingencia
7. Otros coeficientes de correlación

Página 170

09 02 04
ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE TRES VARIABLES

1. Introducción
2. Correlación parcial
3. Coeficiente de correlación múltiple
   - 3.1. Propiedades del coeficiente de correlación múltiple
4. Regresión múltiple y su coeficiente de determinación
   - 4.1. Interpretaciones de $R^2_{1.23}$
     - 4.1.1. $R^2_{1.23}$ como índice de reducción de error en los pronósticos
     - 4.1.2. $R^2_{1.23}$ como aproximación de los puntos al plano de regresión
     - 4.1.3. $R^2_{1.23}$ como proporción de la varianza de X_1 asociada a la variación de X_2 y de X_3

Página 176

09 02 05
PROBABILIDAD

1. Introducción: conceptos básicos
2. Probabilidad y espacio muestral discreto
   - 2.1. Enfoque interpretativo
   - 2.2. Enfoque formal
   - 2.3. Probabilidad condicional
   - 2.4. Sucesos independientes
   - 2.5. Teorema de Bayes
3. Funciones de probabilidad y de distribución de probabilidad en variables aleatorias discretas
   - 3.1. Variable aleatoria
   - 3.2. Función de probabilidad
   - 3.3. Función de distribución
   - 3.4. Función de probabilidad y distribución con dos variables aleatorias discretas
   - 3.5. Algunas funciones de probabilidad y distribución en variables aleatorias discretas
4. Esperanza, covarianza, Pearson y varianza en variables aleatorias discretas
   - 4.1. Valor esperado o esperanza matemática
   - 4.2. Covarianza, Pearson y varianza
   - 4.3. Esperanza matemática y varianza en algunas distribuciones de probabilidad
5. Probabilidad y espacio muestral continuo
   - 5.1. Función de densidad de probabilidad uniforme o rectangular
   - 5.2. Función de densidad de probabilidad normal
   - 5.3. Función de densidad de probabilidad (χ^2)
   - 5.4. Función de densidad de probabilidad de Student (t)
   - 5.5. Función de densidad de probabilidad de Fisher (F)
   - 5.6. Función de densidad de probabilidad exponencial
6. Esperanza matemática y varianza en variables aleatorias continuas

Página 190

09 02 06
FUNDAMENTOS BÁSICOS DE LA ESTADÍSTICA INFERENCIAL

1. Introducción a la inferencia estadística
   - 1.1. Conceptos básicos
   - 1.2. Técnicas de muestreo
   - 1.3. Valor esperado y varianza de la media
2. Estimación puntual de parámetros
   - 2.1. Propiedades deseables de un estimador
3. Comprobación de hipótesis estadísticas e intervalos confidenciales
   - 3.1. Formulación de la hipótesis nula y alternativa
   - 3.2. Determinación del nivel de significación o α
   - 3.3. Estudiar las características de la población
   - 3.4. Especificar el tipo de muestreo realizado y el tamaño de la muestra o de las muestras
   - 3.5. Seleccionar el estadístico de contraste adecuado al caso
   - 3.6. Atender a la distribución muestral del estadístico de contraste
   - 3.7. Determinar la región crítica
   - 3.8. Rechazar o aceptar la hipótesis
   - 3.9. Determinación del intervalo confidencial del parámetro
   - 3.10. Ejemplo método de estimación por intervalos

Página 206

09 02 07
TÉCNICAS NO PARAMÉTRICAS

1. Introducción
2. Características de las técnicas no paramétricas
   - 2.1. Ventajas
   - 2.2. Desventajas
3. Principales pruebas no paramétricas
   - 3.1. Pruebas de bondad de ajuste
   - 3.2. Pruebas de independencia
   - 3.3. Prueba de Mann-Whitney
   - 3.4. Prueba de Wilcoxon
   - 3.5. Prueba de Kruskal-Wallis
   - 3.6. Prueba de Friedman
   - 3.7. Prueba de signos

Página 214

09 02 08
PRUEBAS PARAMÉTRICAS

1. Introducción
2. Supuestos pruebas paramétricas
   - 2.1. Normalidad
   - 2.2. Homocedasticidad
   - 2.3. Independencia
   - 2.4. Esfericidad
   - 2.5. Aditividad
3. Análisis de la regresión y correlación
   - 3.1. Análisis de la regresión
   - 3.2. Análisis de la correlación
4. Contraste de hipótesis sobre una sola media

5. Contraste de hipótesis sobre dos medias
   5.1. Contraste de hipótesis sobre dos medias independientes
   5.2. Contraste de hipótesis sobre dos medias relacionadas
6. Análisis de varianza: un solo criterio de clasificación (ANOVA I)
   6.1. Conceptos básicos en el análisis de varianza
   6.2. Esquema del análisis de varianza
   6.3. Un caso particular: el análisis de varianza con medidas repetidas
7. Análisis de varianza: doble criterio de clasificación (ANOVA II)
   7.1. Conceptos básicos en el ANOVA de doble criterio
   7.2. Esquema del ANOVA ll
   7.3. Contrastes o comparaciones múltiples
   7.4. Efectos factoriales
8. Análisis de covarianza (ANCOVA): conceptos básicos

09.03. PSICOMETRÍA

Página 237 **09 03 01**

INTRODUCCIÓN A LA PSICOMETRÍA

1. Desarrollo histórico de la Psicometría
2. Etapas en la construcción de los tests
3. La puntuación
   3.1. Tipos de puntuación
   3.2. Formas de distribución de las puntuaciones
4. La aptitud

Página 246 **09 03 02**

TEORÍA CLÁSICA DE LOS TESTS

1. Introducción
2. Supuestos básicos
3. Conclusiones de los supuestos básicos
4. Las medidas paralelas
5. Medidas equivalentes o tau-equivalentes
6. Consecuencias prácticas

Página 255 **09 03 03**

CÁLCULO DEL COEFICIENTE DE FIABILIDAD

1. Introducción
2. Concepto de fiabilidad
3. Cálculo del coeficiente de fiabilidad
   3.1. Métodos basados en dos aplicaciones
      3.1.1. Método de las formas paralelas o alternativas
      3.1.2. Procedimiento test-retest
      3.1.3. Test-retest con formas alternativas (formas alternativas en aplicación diferida)
   3.2. Procedimientos basados en una única aplicación del test: métodos basados en la consistencia interna
      3.2.1. Métodos basados en la división del test en dos mitades
      3.2.2. Métodos basados en las covarianzas entre ítems
   3.3. Fiabilidad entre evaluadores o calificadores
4. Relaciones entre la fiabilidad y otras variables
   4.1. Fiabilidad y homogeneidad de la muestra
   4.2. Fiabilidad y longitud del test
   4.3. Fiabilidad, longitud y varianza
   4.4. Razón señal-ruido
5. Estimación de la puntuación verdadera
   5.1. Errores de medida, estimación y predicción
6. Análisis convencional de un ítem
7. Valoración de la TCT
8. Teoría de la generalizabilidad
   8.1. Conceptos básicos
   8.2. Estudios G y D
   8.3. Optimización de un diseño
   8.4. Ejemplo de aplicación de la TG

Página 274 **09 03 04**

VALIDEZ

1. Introducción
2. Validez de contenido
3. Validez de constructo
   3.1. Validez multimétodo-multirasgo
   3.2. Análisis factorial
4. Validez relativa al criterio
   4.1. Coeficiente de validez
   4.2. Relaciones entre validez y fiabilidad: fórmulas de atenuación
   4.3. Validez y longitud del test
   4.4. Validez y homogeneidad de las muestras
   4.5. Modalidades del coeficiente de validez
   4.6. Estimación del criterio
   4.7. Validez de criterio: pronósticos mediante baterías de predictores
5. Otras aproximaciones de estudio de la validez en evaluación psicológica

Página 285 **09 03 05**

TEORÍA DEL RASGO LATENTE

1. Introducción
2. Conceptos básicos
   2.1. Dimensionalidad
   2.2. Independencia local
   2.3. Curva característica del ítem
   2.4. Escala de aptitud
   2.5. Función de información del test

BIBLIOGRAFÍA COMENTADA
WEBGRAFÍA COMENTADA
PREGUNTAS PIR DE CONVOCATORIAS ANTERIORES

REFERENCIAS BIBLIOGRÁFICAS

FUNDAMENTOS TEÓRICOS Y DISEÑOS EXPERIMENTALES

- ÁLVAREZ CÁCERES, R. (1996): **El método científico en las ciencias de la salud. Las bases de la investigación biomédica**. Madrid: Díaz de Santos.
- ALVIRA MARTÍN, F. (2002): **Perspectiva cualitativa/ perspectiva cuantitativa en la metodología sociológica**. México DF.: McGraw Hill.
- AMÓN, J. (2006): **Estadística para psicólogos I: Estadística descriptiva**. Madrid: Pirámide.
- AMÓN, J. (2006): **Estadística para psicólogos II: Probabilidad, Estadística inferencial**. Madrid: Pirámide.
- ANGUERA, M.T., ARNAU, J.; ATO, M.; MARTÍNEZ, R.; PASCUAL, J. y VALLEJO, G. (1995): **Métodos de investigación en psicología**. Madrid: Síntesis.
- ARNAU, J. (1978): **Métodos de investigación en las Ciencias Humanas**. Barcelona: Omega.
- ARNAU, J.; ANGUERA, M.T. y GÓMEZ, J. (1990): **Metodología de la investigación en ciencias del comportamiento**. Murcia: Universidad de Murcia.
- BELLACK, A.S. y HERSEN, M. (1989): **Métodos de investigación en Psicología Clínica**. Bilbao: DDB.
- CABRERO GARCÍA, L.; RICHART MARTÍNEZ, M. (1996): **El debate investigación cualitativa frente a investigación cuantitativa**. México DF.: Enfermería clínica.
- CALERO, J.L. (2000): **Investigación cualitativa y cuantitativa. Problemas no resueltos en los debates actuales**. Rev. Cubana Endocrinol 2000; 11 (3): 192-8.
- CAMPBELL, D.; STANLEY, J. (2002). **Diseños experimentales y cuasi experimentales en la investigación social**. Buenos Aires (Argentina): Ammorrortu Editores.
- CASTRO, J.A. (2001): **Metodología de la investigación. Fundamentos**. Salamanca: Amarú.
- GARCÍA JIMÉNEZ, M.V. (1992): **El método experimental en la investigación psicológica**. Barcelona: PPU.
- MARTÍNEZ ARIAS, R. (1981): **Los métodos experimentales en psicología clínica**. En J.F. Morales: "Metodología y teoría de la psicología". Madrid: UNED.
- MARTÍNEZ HERNÁNDEZ, M. (1994): **Métodos y Diseños de Investigación en Psicología y Educación**. Madrid: Editorial Complutense.
- PELEGRINA, M. y SALVADOR, F. (1999): **La investigación experimental en Psicología**. Archidona (Málaga): Aljibe.

— PÉREZ SERRANO, G. (2002): **Investigación cualitativa. Retos e interrogantes. II Técnicas y análisis de datos**. Madrid: 1ª Muralla, S.A.
— RODRÍGUEZ GÓMEZ, G. y otros (1996): **Metodología de la investigación cualiativa**. Málaga: Ediciones Aljibe, S.L.
— RUBIO JERÓNIMO, A. (1989): **Bases teóricas de la investigación en Psicología**. Madrid: U.C.M.
— RUBIO JERÓNIMO, A. (1989): **Diseños experimentales: Teoría y comentarios de experimentos psicológicos**. Madrid: U.C.M.
— RUIZ OLABUÉNAGA, J.I. (1996): **Metodología de la investigación cualitativa**. Bilbao: Universidad de Deusto.
— ZINSER, O. (1987): **Psicología Experimental**. Bogotá: McGraw-Hill.

ESTADÍSTICA

— ABELSON, R.P. (1998). **La estadística razonada: Reglas y principios**. Barcelona: Paidós.
— AMÓN, J. (2006): **Estadística para psicólogos I: Estadística descriptiva**. Madrid: Pirámide.
— AMÓN, J. (2006): **Estadística para psicólogos II: Probabilidad, Estadística inferencial**. Madrid: Pirámide.
— BOTELLA, J.; LEÓN, O.F. y SAN MARTÍN, R. (1993). **Análisis de Datos en Psicología I: teoría y ejercicios**. Madrid: Pirámide.
— FERGUSON, G.A. (1989): **Análisis estadístico en educación y psicología**. Madrid: Anaya.
— GUARDIA, J. (2008). **Análisis de datos en Psicología**. Madrid: Delta.
— MARTÍNEZ ARIAS, R. (1995). **Psicometría: Teoría de los tests psicológicos y educativos**. Madrid: Síntesis.
— PARDO, A. y SAN MARTÍN, R. (1994). **Análisis de datos en psicología II**. Madrid: Pirámide (2ª edición: 1998).
— SAN MARTÍN, R.; ESPINOSA, L. y FERNÁNDEZ PEDREIRA, L. (1986). **Psicoestadística. Descriptiva**. Madrid: Pirámide.
— SAN MARTÍN, R.; ESPINOSA, L. y FERNÁNDEZ PEDREIRA, L. (1986). **Psicoestadística. Estimación y contraste**. Madrid: Pirámide.
— SAN MARTÍN, R. y PARDO, A. (1989). **Psicoestadística. Contrastes paramétricos y no paramétricos**. Madrid: Pirámide.

PSICOMETRÍA

— AMÓN, J. (1987): **Estadística para psicólogos 1: Probabilidad. Estadística inferencial**. Madrid: Pirámide.
— AMÓN, J. (1987): **Estadística para psicólogos 2: Estadística descriptiva**. Madrid: Pirámide.
— FERGUSON, G.A. (1989): **Análisis estadístico en educación y psicología**. Madrid: Anaya.
— MARTÍNEZ ARIAS, R. (1995): **Psicometría: Teoría de los tests psicológicos y educativos**. Madrid: Síntesis.
— MARTÍNEZ ARIAS, R. (2006): **Psicometría**. Madrid: Anaya.
— PONSODA, V. y OLEA, J. (1992): **Teoría de los tests**. Madrid: U.A.M.
— SANTISTEBAN REQUENA, C. (1990): **Psicometría: Teoría y práctica en la construcción de tests**. Madrid: Ediciones Norma.

BIBLIOGRAFÍA CLÁSICA

— AMÓN, J. (2000). **Estadística para psicólogos I. estadística descriptiva**. Madrid: Síntesis.
— AMÓN, J. (2000). **Estadística para psicólogos II. Probabilidad. Estadística inferencial**. Madrid: Síntesis.
— ANGUERA, M.T.; ARNAU, J.; ATO, M. et al. (1998). **Métodos de investigación en psicología**. Madrid: Síntesis.
— MARTÍNEZ, M. (1994). **Métodos y diseños de investigación en psicología**. Madrid: Complutense.
— MARTÍNEZ, R. (2005). **Psicometría: teoría de los test psicológicos y educativos**. Madrid: Síntesis.

BIBLIOGRAFÍA ACTUAL

— ALVARADO, J.M.; SANTISTEBAN, C. (2012). **La validez en la medición psicológica**. Madrid: UNED.
— BALLUERCA, N.; VERGARA, A.I. (2002). **Diseños de investigación en psicología experimental: modelos y análisis de datos mediante SPSS 10.0.** Pearson educación.
— BARBERO, M.I. (2010). **Psicometría (teoría, formulario y problemas resueltos)**. Madrid: Sanz y Torres.
— BOTELLA, J.; SUERÓ, M.; XIMÉNEZ, C. (2012). **Análisis de datos en psicología I**. Madrid: Pirámide.
— FONTES DE GRACIA, S.; GARCÍA, C.; QUINTANILLA, L. et al. (2010). **Fundamentos de investigación en psicología**. Madrid: UNED.
— LUBIN, P.; MACIÀ, A.; RUBIO DE LERMA, P. (2005). **Psicología matemática I y II**. Madrid: UNED.
— MARTÍNEZ, M.A.; HERNÁNDEZ, M.J.; HERNÁNDEZ, M.V. (2014). **Psicometría**. Madrid: Alianza.
— PARDO, A.; SAN MARTÍN, R. (2006). **Análisis de datos en psicología II**. Madrid: Pirámide.
— SANTISTEBAN, C. (2009). **Principios de psicometría**. Madrid: Síntesis.

BIBLIOGRAFÍA ESPECÍFICA

— ABAD, F.J. (2011). **Medición en ciencias sociales y de la salud**. Madrid: síntesis.
— BALLUERKA, N.; VERGARA, A.J. (2002). **Diseños de investigación en psicología experimental: Modelos y análisis de datos mediante SPSS 10.0**. Pearson Educación.
— BARBERO, M.I. (2011). **Introducción básica al análisis factorial**. Madrid: UNED.
— BOTELLA, J. (2015). **Meta-análisis en ciencias sociales y de la salud**. Madrid: Síntesis.
— CRONBACH, L.J. (1998). **Fundamentos de los test psicológicos**. Madrid: biblioteca nueva.
— KAZDIN, A.E. (2002). **Métodos de investigación en psicología clínica**. Méjico: Prentice Hall.
— MUÑIZ, J. (1997). **Introducción a la teoría de respuesta a los ítems**. Madrid: Pirámide.
— MUÑIZ, J. (2002). **Teoría clásica de los test**. Madrid: Pirámide.
— PANTOJA, A. (2009). **Manual básico para la realización de tesinas, tesis y trabajos de investigación**. Madrid: EOS.

ÚLTIMOS MANUALES PUBLICADOS

— MARTÍNEZ, R.; CASTELLANOS, M.A.; CHACÓN, J.C. (2015). **Análisis de datos en psicología y ciencias de la salud II. Inferencia estadística**. Madrid: EOS.
— MARTÍNEZ, R.; CHACÓN, J.C.; CASTELLANOS, M.A. (2015). **Análisis de datos en psicología y ciencias de la salud I. Exploración de datos y fundamentos probabilísticos**. Madrid: EOS.
— MENESES, J. (Coord.) (2013). **Psicometría**. UOC.
— MERINO, J.J.; MORENO, E.; PADILLA, M. et al. (2013). **Análisis de datos en psicología I**. Madrid: UNED.

CEDE - C/ Cartagena, 129 - 28002 Madrid
Tel.: 91 564 42 94

09.02.03 ESTADÍSTICA DESCRIPTIVA APLICADA AL ESTUDIO DE DOS VARIABLES

1. INTRODUCCIÓN

En este tema vamos a estudiar la estadística descriptiva centrada en dos variables.

En primer lugar abordaremos la organización de datos correspondiente a dos variables. Posteriormente comenzaremos tratando las relaciones entre variables, para a continuación abordar las relaciones curvilíneas.

Por último, dedicaremos algunos apartados a la relación entre dos variables que se encuentren a nivel de medida ordinal y nominal.

2. DISTRIBUCIÓN CONJUNTA DE FRECUENCIAS

En el tema anterior tratábamos la estadística descriptiva centrándonos en el estudio de una sola variable. Iniciamos ahora el estudio de dos variables; una persona puede evaluarse o medirse de acuerdo a una característica (peso, altura, inteligencia, nacionalidad, nivel cultural, etc.), pero también puede describirse de acuerdo a dos variables (inteligencia y nivel cultural, por ejemplo), de modo que a la persona le corresponden dos modalidades, una por variable, y consiguientemente, dos números como resultado de atribuir números a las modalidades de las dos variables. Las dos variables pueden encontrarse a igual nivel de medida (nominal-nominal, ordinal-ordinal, intervalos-intervalos, etc.) o bien pueden tener distinto nivel de medida; así podemos estudiar una serie de sujetos con respecto a su provincia de origen (nominal) y el grado académico (ordinal).

Para apreciar más claramente cómo es una distribución conjunta de frecuencias partiremos del estudio de dos variables cuantitativas y con un mismo nivel de medida, en este caso como mínimo a nivel de intervalos.

Supongamos una muestra de 10 sujetos y sus correspondientes calificaciones en un examen (X); a continuación les aplicamos un test de inteligencia (Y). Los resultados en cada variable, sin agrupar en intervalos, son los siguientes:

X	Y
34	100
20	93
10	130
48	125
15	97
60	125
34	110
25	105
22	98
50	115

A continuación agrupamos esos datos en intervalos, pues lo más frecuente es que una distribución conjunta se muestre con sus datos agrupados. El tipo de intervalos no tiene por qué ser igual en las dos variables, de modo que para la variable X elegimos los intervalos (1-20 / 21-40 / 41-60); para la variable Y elegimos estos otros (91-100 / 101-110 / 111-120 / 121-130). La distribución conjunta de frecuencias con estos intervalos tendrá el siguiente aspecto:

		X			
		1-20	21-40	41-60	
Y	121-130	1	0	2	3
	111-120	0	0	1	1
	101-110	0	2	0	2
	91-100	2	2	0	4
		3	4	3	10

El conjunto de casillas grises constituye la **distribución de frecuencias conjunta**, pues cada una de las casillas recoge el número (o porcentaje o proporción en otros casos) de sujetos u observaciones comprendidos entre los pares

de intervalos de cada variable. Así, en la casilla superior izquierda aparece la frecuencia conjunta de 1, que indica que en nuestra muestra hay un único sujeto cuya puntuación en el examen se encuentra entre 10-20 y cuya inteligencia comprende un valor entre 121 y 130.

Además de la distribución conjunta de X e Y tenemos dos **distribuciones marginales**. Estas distribuciones se representan en los márgenes de la tabla.

La distribución marginal de X es la distribución en X de todas las observaciones, independientemente de sus puntuaciones en Y:

X	nj	Xj
41-60	3	50,5
21-40	4	30,5
10-20	3	10,5
	10	

X representa los intervalos definidos en la variable; nj la frecuencia absoluta en cada intervalo y Xj el punto medio.

La otra distribución marginal es la distribución en Y de todas las observaciones, independientemente de sus puntuaciones en X:

Y	nj	Yj
121-130	3	125,5
111-120	1	115,5
101-110	2	105,5
90-100	4	95,5
	10	

Cada una de las distribuciones marginales tendrá su correspondiente media, y varianza, que en este caso pasan a denominarse media y varianza marginales.

3. RELACIÓN LINEAL ENTRE DOS VARIABLES

Existe **relación** o **correlación** entre dos variables, cuando las modalidades de una de las dos variables están ligadas a las modalidades de la otra. La correlación entre dos variables implica una **variación conjunta**. Esta correlación puede darse en distintos niveles de medida, así, a un nivel nominal, la correlación entre dos variables implica que pertenecer a una clase o modalidad de una variable se asocia con pertenecer a una clase o modalidad determinada de otra variable.

En un primer momento vamos a abordar la correlación entre variables cuantitativas, para pasar a considerar, más adelante, las variables cuasicuantitativas (nivel de medida ordinal) y cualitativas (nivel de medida nominal). Así mismo, comenzaremos por estudiar la relación lineal.

3.1. COVARIANZA (PIR 03,50)

La covarianza entre dos variables, por ejemplo X e Y, es la media aritmética de los productos entre la diferencia $(X_i - \bar{X})$ y la diferencia $(Y_i - \bar{Y})$ correspondientes a cada uno de los n elementos que componen un grupo:

$$\text{cov}\,(X, Y) = S_{xy} = \Sigma\,(X_i - \bar{X})\,(Y_i - \bar{Y}) \Big/ n = (\Sigma X_i Y_i \Big/ n) - \bar{X}\bar{Y}$$

En el caso de que los datos estén agrupados en intervalos, como en nuestra distribución conjunta del inicio del capítulo, la fórmula será:

$$\text{cov}\,(X, Y) = S_{xy} = \Sigma\Sigma n_{ij}\,(X_i - \bar{X})\,(Y_i - \bar{Y})_{/} n =$$

$$(\Sigma\Sigma n_{ij} X_i Y_{i\,/} n) - \bar{X}\bar{Y}$$

donde n_{ij} representa las frecuencias conjuntas, esto es, la frecuencia con que cada par de puntos medios de los respectivos intervalos (X_i, Y_i) aparece en la distribución conjunta.

3.1.1. Propiedades de la covarianza

• Si a nuestras dos variables X e Y se les aplica una transformación lineal del tipo (V = bX + a; W = cY + d), resulta que la covarianza de las dos nuevas variables es:

$$S_{vw} = (bc)\, S_{xy}$$

• Cuando existen varios grupos de personas u observaciones con puntuaciones en las variables X e Y, la covarianza del grupo total es igual a la media de las covarianzas de cada grupo más la covarianza de las medias del grupo.

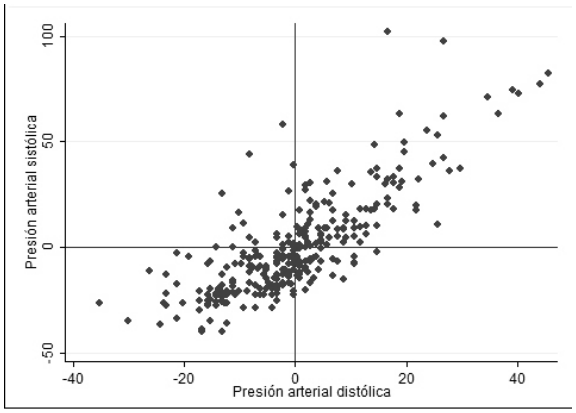
• La covarianza no indica la intensidad de la relación y está ligada a las unidades de medida de los datos.

3.2. COEFICIENTE DE CORRELACIÓN DE PEARSON

El **coeficiente de correlación de Pearson** es un índice que mide la correlación o variación conjunta entre dos variables, X e Y, siempre y cuando ambas sean cuantitativas y además se relacionen linealmente. Esto es, indica la co-variación lineal (PIR 13, 79). Por relación lineal se entiende que al representar en un diagrama cartesiano los pares de puntuaciones (X, Y), el resultado es una nube de

puntos (gráfico de dispersión) que se aproximan a la línea recta **(PIR 13, 81)**.

RELACIÓN LINEAL ENTRE DOS VARIABLES



Decimos que la correlación entre X e Y es positiva cuando a una puntuación en X que se encuentra por encima de la media, le corresponde una puntuación en Y también por encima de la media, y a una puntuación X por debajo de la media le corresponde una puntuación Y por debajo de la media. La correlación será negativa cuando la covariación se produzca en el sentido opuesto. Finalmente, la correlación será nula cuando a una puntuación en X por encima de la media le corresponde una puntuación en Y que puede estar por encima o por debajo de la media, y a una puntuación X por debajo de la media, le corresponde una puntuación Y que puede estar por debajo o por encima de la media. En este último caso se dice que no covarían. Dicho de otra manera, hablaremos de correlación positiva cuando al aumentar la variable X aumenta también la variable Y, o al disminuir la X disminuye también la variable Y. En cambio, cuando nos encontramos ante una correlación negativa, al aumentar la variable X disminuye Y y viceversa. Por último, cuando la correlación lineal es nula ambas variables varían, aunque sin seguir una regla concreta; diremos por lo tanto, que son linealmente independientes.

Según lo anterior se entiende fácilmente que la covarianza (Sxy) puede ser un índice apropiado para medir la correlación, pues si la relación entre las dos variables es positiva, las dos diferencias $(X - \bar{X})$ e $(Y - \bar{Y})$ serán del mismo signo y por tanto el resultado del producto llevará el signo positivo, esto es, Sxy será positivo. Si la relación es negativa, cada diferencia tendrá un signo distinto, y el resultado del sumatorio del producto de estas diferencias llevará el signo negativo. En el caso de la relación nula, la mitad de los productos serán positivos y la otra mitad negativos, con lo cual el sumatorio dará un valor muy próximo a cero.

No obstante no podemos usar la covarianza como un índice del todo apropiado, pues está muy ligado a la unidad de medida de las variables. Para solucionar este problema necesitamos un número abstracto, y éste se consigue si dividimos las diferencias $(X - \bar{X})$ e $(Y - \bar{Y})$ por sus desviaciones típicas correspondientes (recordemos que en el caso del Cociente de Variación hacíamos lo mismo para comparar variabilidades entre dos variables). El resultado es el **coeficiente de correlación de Pearson**:

$$\frac{\Sigma(X - \bar{X}/S_x)(Y - \bar{Y}/S_y)}{n} = \frac{\Sigma[(X - \bar{X})(Y - \bar{Y})]/n}{S_x S_y} = \frac{S_{xy}}{S_x S_y}$$

El coeficiente de correlación de Pearson viene designado habitualmente por $\mathbf{r}_{xy}$ y adopta varias versiones:

Con puntuaciones típicas:

$$r_{xy} = \Sigma z_x z_y / n$$

Con puntuaciones diferenciales:

$$r_{xy} = \Sigma x y / n S_x S_y$$

Con puntuaciones directas:

$$r_{xy} = (n\Sigma XY - \Sigma X \Sigma Y) / \sqrt{n\Sigma X^2 - (\Sigma X)^2}\ \sqrt{n\Sigma Y^2 - (\Sigma Y)^2}$$

3.2.1. Propiedades del Coeficiente de correlación de Pearson o Coeficiente de correlación simple

• El coeficiente no puede valer más de 1 ni menos de –1. Así, podremos conocer tanto la intensidad como la dirección de la correlación interpretando el signo de pearson, como veremos más adelante **(PIR 02, 156)**.

• Si aplicamos una transformación lineal a las variables X e Y (V = bX + a; W = cY + d), resulta que el coeficiente de correlación entre las nuevas variables V y W es igual en valor absoluto al coeficiente de correlación entre X e Y:

$$|r_{xy}| = |r_{vw}|$$

Si las pendientes b y c son del mismo signo, entonces los coeficientes de correlación serán también de idéntico signo. Si las pendientes b y c son de diferente signo, en-

tonces los coeficientes de correlación también lo serán **(PIR 01, 116; PIR 05, 96)**.

• El coeficiente de correlación entre dos variables X e Y aumenta cuando aumenta la variabilidad de una o de las dos variables, y disminuye cuando desciende la variabilidad de una o de las dos variables, se verá reducido si aumenta la homogeneidad de la muestra **(PIR 17, 219)**. Por lo tanto, el efecto de la "restricción del rango" disminuirá el valor de la correlación observada entre dos variables **(PIR 18, 52)**.

• El coeficiente de correlación entre dos variables X e Y puede ser elevado por el influjo de una tercera variable; una forma de contrarrestar este efecto es identificar grupos en función de los niveles de la tercera variable, y calcular el coeficiente de correlación entre X e Y para cada grupo. En definitiva se trata de controlar la tercera variable como si fuera una variable extraña o contaminante.

• Si entre dos variables no existe ninguna relación, el coeficiente de correlación de Pearson será cero necesariamente. Ahora bien, dado un coeficiente de correlación de Pearson igual a cero, no puede concluirse necesariamente que no exista ninguna relación, pues es posible que exista una relación no lineal, y por tanto no va a ser detectada por dicho coeficiente.

3.2.2. Interpretación del Coeficiente de correlación de Pearson

Cuando el coeficiente de correlación de Pearson vale (± 1) estamos ante un caso de correlación lineal perfecta, esto es, los pares de puntuaciones (X, Y) se distribuyen en la representación cartesiana coincidiendo con una línea recta, sin que ninguno de ellos se desvíe de la misma.

Si el coeficiente es cero, la correlación lineal es nula **(PIR 05, 87)**, aunque como ya hemos dicho, los pares de puntuaciones pudieran formar en el plano otro tipo de relación que no fuera lineal.

Cuando el coeficiente de correlación tiene signo positivo, concluiremos que existe correlación positiva. Si su signo es negativo, entonces la correlación será negativa o inversa.

Para valores de Pearson entre (0) y (± 1) no existe una interpretación unívoca, pues para afirmar si un valor es alto o bajo debemos comparar nuestro dato con otras investigaciones con las mismas variables y en circunstancias parecidas.

Una alta correlación no implica necesariamente causalidad. Es cierto que si entre dos variables existe una relación causal y ésta es lineal, el coeficiente de correlación de Pearson será elevado, pero esta conclusión no puede deducirse de modo inverso. La relación entre dos variables X e Y pudiera estar basada en el influjo de una tercera variable, la cual sería la responsable de esta correlación. La correlación indica, en principio, una mera covariación entre dos variables y nada más.

3.2.3. Matriz de varianzas-covarianzas y de correlaciones

Para poder conocer las relaciones de diferentes variables que combinan linealmente podemos utilizar la matriz de varianzas-covarianzas o la matriz de correlaciones.

En la matriz de varianzas-covarianzas obtendremos los valores de las covarianzas de dichas variables, y en la diagonal de la matriz obtendremos la varianza de cada variable **(PIR 03, 45)**, ya que la varianza se puede definir como la covarianza de una variable consigo misma.

La matriz de correlaciones la calcularemos cuando queremos organizar las correlaciones de diferentes variables. En este caso, en la diagonal de dicha matriz tendremos unos, ya que la correlación de una variable consigo misma es perfecta **(PIR 01, 118)**.

MATRIZ DE VARIANZAS-COVARIANZAS

	1	2	3
1	S_1^2	S_{12}	S_{13}
2		S_2^2	S_{23}
3			S_3^2

MATRIZ DE CORRELACIONES

	1	2	3
1	1	r_{12}	r_{13}
2		1	r_{23}
3			1

3.3. LA ECUACIÓN DE REGRESIÓN

La regresión significa **predicción** o **pronóstico**; en el caso de dos variables que mantienen una relación lineal, hablamos de regresión simple. Para poder predecir los resultados de un individuo en la variable Y (criterio o variable dependiente) **(PIR 01, 125)** a partir de su puntuación en la variable X (causa o variable independiente) **(PIR 01, 122)**, necesitamos una ecuación que relacione ambas variables

(PIR 06, 24; PIR 08, 232). Esta ecuación no es más que la ecuación de una recta, y que por tanto adquiere esta forma:

$$Y = bX + a$$

Debemos recordar que b es llamada también beta y a, alfa.

b y a son dos constantes propias de cada tipo de recta: (a) es la ordenada en el origen, esto es, representa el valor de Y (ordenada) cuando X = 0. La otra constante, (b) es la pendiente de la recta **(PIR 01, 124)**, y representa la inclinación de la misma. Indica el cambio que se produce en la variante respuesta (Y) por una unidad de cambio en el predictor (X). (Tener en cuenta que existen diversas formas de nombrar a la recta, por ejemplo: Y = a + bX).

A nivel gráfico, X se representará en el eje de abscisas **(PIR 01, 123)**, mientras que Y se representa en el eje de ordenadas.

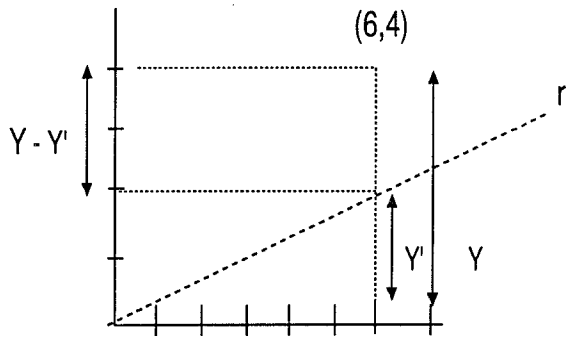
Aunque para algunos autores las dos constantes de la ecuación de regresión son también denominadas b0 en el caso de alfa, y b1 en el caso de beta.; la mayoría coinciden en llamar alfa a la constante sumativa que se representa en la ordenada en el origen y beta a la constante multiplicativa que tiene que ver con la pendiente de la recta (b y a, tal y como las hemos denominado a lo largo del manual).

β	α
Beta Constante multiplicativa Pendiente de la recta B1	Alfa Constante Aditiva Ordenada en el origen B0

El programa SPSS, entre otros, incluye estos valores (β y α) en una columna denominada coeficientes no estandarizados, e indica que dicha columna contiene los coeficientes de regresión parcial que definen la ecuación de regresión en puntuaciones directas **(PIR 21, 29)**. Por otro lado, los coeficientes de regresión parcial estandarizados son los que definen a la ecuación de regresión cuando ésta se obtiene tras estandarizar las variables originales, es decir, tras convertir las puntuaciones directas en típicas.

La construcción de una ecuación de regresión (o recta de regresión) requiere seleccionar de entre todas las posibles aquélla con la que cometamos los mínimos errores.

En el apartado anterior decíamos que en el caso de una relación lineal perfecta ($r = \pm 1$) los pares de puntuaciones (X, Y) coincidían todos ellos con una línea recta, siendo esta línea (la única que corta todos los puntos X, Y) la ecuación de regresión. Pero en el resto de los casos, cuando la relación lineal no es perfecta, no todos los puntos (X, Y) coinciden con una línea recta (Ver gráfico: **Relación lineal entre dos variables**), aunque la tendencia sea distribuirse linealmente. Este hecho obliga a seleccionar una recta de entre varias posibles, con la que, aun cometiendo error, haga que éste sea mínimo.



Según el gráfico un sujeto puntúa 6 y 4 en las variables X e Y respectivamente. Si elegimos como recta de regresión la que figura en el gráfico, vemos que para una puntuación en X de 6, la recta asigna al sujeto una puntuación en Y de 2. Llamamos puntuación Y a la puntuación real del sujeto en la variable Y; por otra parte, la puntuación que la recta de regresión asigna a este sujeto en la variable Y recibe el nombre de puntuación pronosticada Y'. Como vemos se produce un error al realizar el pronóstico:

$$E = Y - Y' = (4 - 2) = 2$$

Si calculamos las diferencias Y - Y' para todas las puntuaciones de n sujetos, tendremos un conjunto de errores. Si se elevan al cuadrado estos errores y se suman tendremos una suma de errores cuadráticos. Ahora bien, de todas las rectas de regresión posibles elegiremos aquélla que haga mínima dicha suma; en esto consiste el método de construcción de las rectas de regresión de Y sobre X por el método de mínimos cuadrados.

En resumen, para construir la recta de regresión de Y sobre X empleamos un conjunto de individuos con puntuaciones en ambas variables para determinar una recta que haga mínima la suma de los errores cuadráticos. A continuación, utilizamos dicha recta para predecir las puntuaciones en la variable Y en un conjunto de sujetos, similares a los anteriores, acerca de los cuales sólo conocemos sus puntuaciones en la variable X.

Por ejemplo, podríamos tener un grupo en el que hallamos su puntuación en un test de extroversión (X) y por otro lado contamos el número de contactos sociales que tienen en una fiesta (Y). Podríamos encontrar una ecuación lineal

que relacionará las dos variables, como por ejemplo Y = X/2+1. Así, si un sujeto tiene una puntuación de 50 en el test, podríamos saber que el número de contactos sociales es de 50/2 + 1 = 26 **(PIR 03, 59; PIR 05, 95)**. Tras hallar la ecuación de regresión, ya no haría falta que en otros sujetos midiéramos las dos variables, sino que pasando el cuestionario podríamos predecir el número de contactos sociales que van a realizar (Y').

Cuando hablamos de regresión lineal y de la recta de regresión solemos hacerlo en puntuaciones directas, si bien dicha recta puede transformase a puntuaciones diferenciales y a puntuaciones típicas. Los valores de beta y alfa tienen relación con los de la desviación típica y la media de sus respectivas puntuaciones. Por ejemplo, la recta de regresión en puntuaciones diferencias (cuya media es 0 y desviación típica la misma que la de las directas) tiene 0 como ordenada en el origen y la pendiente igual a la de puntuaciones directas. Al igual que la recta de regresión expresada en diferenciales, la recta en típicas tampoco tiene coeficiente alfa (su valor es cero), dado que el eje de coordenadas estaría ahora sobre el punto (0, 0) que es la media de las puntuaciones típicas de X e Y, y el valor de la pendiente coincide con el del coeficiente de correlación de Pearson (en vez de con el de la desviación típica que sería 1), en ocasiones a esa pendiente en la recta se le denomina coeficiente de regresión.

3.3.1. La ecuación de regresión en puntuaciones directas

Siendo la ecuación de regresión una recta su formulación matemática en puntuaciones directas será la siguiente:

$$Y' = BX + A$$

Donde B equivale a:

$$B = \Sigma X Y - n \Sigma \bar{X} \bar{Y} \Big/ n \Sigma X^2 - (\Sigma X)^2$$

Y A viene expresada como:

$$A = \bar{Y} - B\bar{X}$$

3.3.2. La ecuación de regresión en puntuaciones diferenciales

La formulación de la recta en puntuaciones diferenciales se ajusta al siguiente formato:

$$y' = bx + a$$

donde (b) y (a) valen:

$$a = 0$$

Como (a = 0) la recta de regresión, expresada en puntuaciones diferenciales, pasará por el origen de coordenadas (0, 0).

$$b = \Sigma x y \Big/ \Sigma x^2 = r_{xy} (Sy \Big/ Sx)$$

Transformando en la expresión matemática de (b) las puntuaciones diferenciales en directas se observa que:

$$b = \Sigma x y \Big/ \Sigma x^2 = \Sigma (X - X)(Y - Y) \Big/ \Sigma (X - X)^2 =$$

$$= (n \Sigma X Y - \Sigma X \Sigma Y) \Big/ n \Sigma X^2 - (\Sigma X^2) = B$$

De lo cual se deduce que al tener idéntica pendiente, la recta de regresión en puntuaciones diferenciales será siempre paralela a la recta de regresión en puntuaciones directas **(PIR 03, 57)**.

3.3.3. La ecuación de regresión en puntuaciones típicas

A partir de la expresión de la recta en puntuaciones típicas:

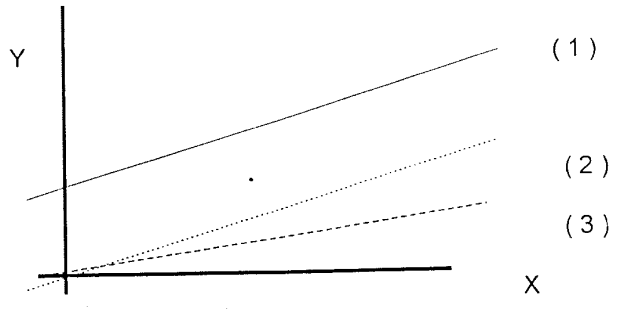
$$z'_y = bz_x + a$$

Presentamos las expresiones equivalentes en estas puntuaciones para (b) y (a):

$$a = 0; b = r_{xy}$$

Como en el caso de las puntuaciones diferenciales, al ser (a = 0) la recta pasa por el origen de coordenadas, si bien en este caso la recta no va a ser paralela a la expresada en puntuaciones directas pues la pendiente no coincide, ya que tiene que ver con el coeficiente de correlación entre VI y VD.

Gráficamente podemos resumir cómo se sitúan en el plano las tres rectas; las tres indicarían una relación positiva. Esto se puede deducir del signo de la pendiente: si ésta es positiva, la recta asciende de izquierda a derecha (como en el gráfico). Si la recta en cambio desciende de izquierda a derecha, la pendiente será negativa y la relación entre las dos variables también.



Dada una recta de regresión expresada en puntuaciones directas (1) y con ($A \neq 0$), vemos cómo esta misma expresada en puntuaciones diferenciales (2) es paralela, porque ambas tienen la misma pendiente ($B = b$), y atraviesa el origen de coordenadas pues ($a = 0$). Por último, la expresión de la recta de regresión en puntuaciones típicas (3) se caracteriza porque la recta atraviesa también el origen de coordenadas, pues también aquí ($a = 0$), pero no es paralela a las otras dos pues la pendiente no es la misma ($b \neq b = B$).

Por lo tanto, en el modelo de regresión lineal simple, cuando trabajamos en puntuaciones típicas, el coeficiente de regresión (beta) de la VI es igual a la correlación de Pearson entre la VD y la VI **(PIR 18, 37)**.

3.3.4. Propiedades de las puntuaciones pronosticadas Y'

• La media de las puntuaciones pronosticadas es igual a la media de las puntuaciones criterio:

$$\bar{Y}' = \bar{Y}$$

• La varianza de las puntuaciones pronosticadas es igual al producto del cuadrado del coeficiente de correlación de Pearson por la varianza de las puntuaciones reales.

$$S^2_{y'} = (r_{xy})^2 S^2_Y$$

• Obviamente la desviación típica de las puntuaciones pronosticadas será:

$$S_{y'} = |r_{xy}| S_Y$$

• La media de las diferencias entre puntuaciones reales y pronosticadas, es decir, los errores, es igual a cero:

$$\bar{E} = \Sigma(Y - Y') / n = 0$$

• La varianza de los errores, o como veremos posteriormente varianza de Y no asociada a X, equivale:

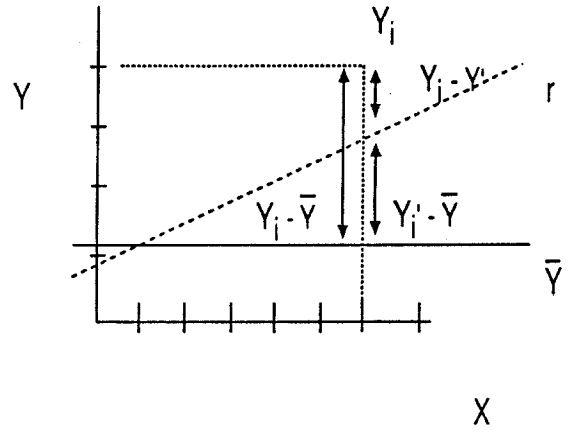
$$S^2_E = (1 - (r_{xy})^2)\, S^2_Y$$

3.4. EL COEFICIENTE DE DETERMINACIÓN Y LA RECTA DE REGRESIÓN

El coeficiente de correlación de Pearson, como hemos visto, está muy ligado a la ecuación de la recta de regresión. En relación con ésta, adquiere otras interpretaciones y sentidos aparte de los que ya han sido citados. Hay que tener en cuenta que estas interpretaciones se harán acerca de su valor elevado al cuadrado, también llamado **coeficiente de determinación**. El coeficiente de determinación da lugar a tres interpretaciones.

3.4.1. Índice de reducción de los errores

Supongamos que en vez de utilizar la recta de regresión para realizar la predicción sobre un sujeto con la puntuación Y_i, mediante la cual sabemos que cometemos un error ($Y - Y'$), utilizamos la media de las puntuaciones reales del grupo al que pertenece el sujeto. Mediante la media cometeremos también un error $(Y - \bar{Y})$.



Si sumamos todos los errores que cometeríamos al usar la media para realizar predicciones sobre las puntuaciones en Y y elevamos al cuadrado estos errores, además de dividir este sumatorio por el número de puntuaciones, obtenemos el error cuadrático medio que cometemos al atribuir a cada persona, como puntuación, la media. Y si desarrollamos matemáticamente esta expresión encontramos que equivale:

$$\Sigma(Y - \bar{Y})^2 / n = [\Sigma(Y' - \bar{Y})^2 / n] + [\Sigma(Y - Y')^2 / n]$$

Despejando:

$$\Sigma(Y' - \bar{Y})^2 / n = [\Sigma(Y - \bar{Y})^2 / n] - [\Sigma(Y - Y')^2 / n]$$

La anterior expresión refleja la diferencia entre el error cuadrático medio cometido al aplicar a cada sujeto la media (S^2_Y) y el cometido al aplicar a cada sujeto la puntuación pronosticada ($S^2_{Y.X}$). En síntesis el resultado de esta diferencia es el error cuadrático medio que dejamos de cometer al emplear la puntuación pronosticada en vez de la media a la hora de hacer pronósticos ($S^2_{Y'}$).

Por lo tanto el cociente,

$$(S^2_{Y'}) \Big/ (S^2_Y) = S^2_{Y'} \Big/ (S^2_{Y'} + S^2_{Y.X}) = r^2_{XY}$$

puede ser interpretado de manera que en el numerador figura la parte de error cuadrático medio eliminado (al utilizar el pronóstico) y en el denominador la suma del error cuadrático medio eliminado y de aquella parte de error no eliminado. En resumen, el cociente viene a expresar **la proporción de error que se ha dejado de cometer usando el pronóstico en lugar de la media, y esta proporción es igual al coeficiente de correlación de Pearson al cuadrado, también llamado coeficiente de determinación**.

De modo que si el coeficiente de correlación al cuadrado entre dos variables (X, Y) es de 0,70 significa que si usamos la recta de regresión para hacer pronósticos en Y a partir de las puntuaciones en X, en lugar de la media, reducimos el error en un setenta por ciento, o, lo que es lo mismo, sólo cometemos el 30% del error que se cometería usando la media.

Pearson simple (correlación) y Pearson al cuadrado (determinación) sólo coincidirán en el caso de que la correlación sea perfecta, puesto que el cuadrado de 1 es 1 **(PIR 05, 90)**.

3.4.2. Índice de aproximación de los puntos a la recta de regresión

En el apartado anterior hacíamos equivalente el coeficiente de determinación a la siguiente expresión:

$$r^2_{XY} = S^2_{Y'} \Big/ S^2_Y$$

Si desarrollamos esta expresión llegamos a la siguiente formulación:

$$S^2_{Y'} \Big/ S^2_Y = (S^2_Y - S^2_{Y.X}) \Big/ S^2_Y = 1 - (S^2_{Y.X} \Big/ S^2_Y) = r^2_{XY}$$

De acuerdo con esta expresión, si todas las puntuaciones Y se encuentran encima de la recta de regresión, entonces ($S^2_{Y.X}$), o error cometido al emplear las puntuaciones pronosticadas, será cero, por lo tanto el coeficiente de determinación (Pearson al cuadrado) valdrá 1.

Vemos entonces como Pearson al cuadrado es también **un índice de aproximación de los puntos a la recta de regresión**, aproximación que es máxima para un valor de 1, pues los puntos coinciden plenamente con la recta de regresión, esto es, se cumple que (Y = Y') para todas las puntuaciones **(PIR 03, 58; PIR 04, 94)**.

3.4.3. Proporción de la varianza de Y asociada a la variación de X

Sabemos que (Y' = bX + a), o lo que es lo mismo, que Y' depende, es función, está asociada a X, de manera que las variaciones en X se van a ver reflejadas en variaciones en la puntuación pronosticada.

Por otra parte, el error que cometemos al usar la puntuación pronosticada es (Y - Y'), y este error no depende, no está asociado a X. Las variaciones en X no tienen porqué corresponderse con variaciones en los errores.

Por otra parte también sabemos que,

$$\Sigma(Y - \bar{Y})^2 \Big/ n = [\Sigma(Y' - \bar{Y})^2 \Big/ n] + [\Sigma(Y - Y')^2 \Big/ n]$$

expresión que equivale a esta otra:

$$S^2_Y = S^2_{Y'} + S^2_{Y.X}$$

En definitiva, significa que la varianza total de Y se descompone en dos partes o sumandos, una, ($S^2_{Y'}$), asociada a, dependiente de, X, pues ya hemos dicho que Y' dependía o estaba asociada a X, y otra, ($S^2_{Y.X}$), no asociada a, no dependiente de, X, pues es la varianza de los errores (Y – Y'), y éstos no dependen de la variable X.

Por otra parte hemos demostrado antes que el coeficiente de correlación de Pearson al cuadrado era equivalente a:

$$r^2_{XY} = S^2_{Y'} \Big/ S^2_Y = S^2_{Y'} \Big/ (S^2_{Y'} + S^2_{Y.X})$$

Y por lo tanto, puede ser interpretado como la proporción de varianza asociada (cociente entre la varianza asociada y la varianza total de Y). Si tenemos un coeficiente de correlación de Pearson de 0,60 podríamos hallar la proporción de varianza asociada elevando 0,60 al cuadrado. Podremos concluir entonces que 0,36 es la proporción de varianza asociada o varianza común y que 0,64 es la proporción de varianza no asociada. Explicado en palabras

más claras, con esto queremos decir que el 36% de las diferencias individuales en Y está asociado, depende de o es explicado por las variaciones o diferencias individuales en X. Por otro lado, diremos que el 64% de dichas diferencias en Y no están asociadas o no dependen de las variaciones en X. De aquí se deduce que cuanto mayor sea un coeficiente de correlación de Pearson, mayor será la cantidad de varianza que se asocie a las variaciones en X y menor la varianza no asociada.

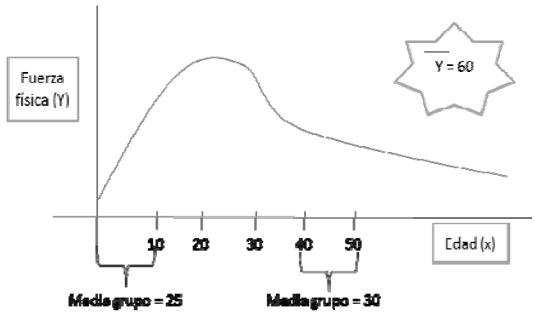
El análisis de regresión lineal es una técnica estadística utilizada para estudiar la relación (en sentido amplio) entre variables, y se adapta a una amplia variedad de situaciones. Tanto en el caso de dos variables (regresión simple) como en el de más de dos (regresión múltiple), el análisis de regresión lineal puede utilizarse para explorar y cuantificar la relación entre una variable llamada dependiente o criterio (Y) y una o más variables llamadas independientes o predictoras (X), así como para desarrollar una ecuación lineal con fines predictivos. Aunque el modelo de regresión lineal parece indicado cuando la naturaleza de ambas variables X e Y sean cuantitativas, no obstante, es demostrable que no es problema operar con variables independientes cualitativas **(PIR 19, 159)**. En el caso de una variable X dicotómica, la regresión simple equivale a un contraste de medias. En el caso de que la variable independiente presente más categorías, será asimilable al hecho de realizar un análisis de la varianza, habiendo una total equivalencia de la regresión con ambas pruebas, con la ventaja de que la regresión ofrece un enfoque más parsimonioso y permite además conocer la proporción de variabilidad explicada por la variable independiente (Rxy2).

La codificación de un predictor categórico o factor utiliza en esencia un conjunto de variables pseudo-numéricas, cada una de las cuales representa un contraste particular entre las medias de los niveles del factor. En general, un predictor categórico o factor con a categorías o niveles requiere un total de **a – 1 variables pseudo-numéricas** **(PIR 21, 30)**. Por ejemplo, para codificar la variable nivel socioeconómico, con tres niveles o categorías (bajo, medio y alto), se requieren 3 – 1 = 2 variables pseudo-numéricas.

Aunque existen muchos tipos de codificación, en la investigación aplicada suelen emplearse tres tipos básicos. Para factores nominales hay dos sistemas muy extendidos, la **codificación tipo regresión o codificación ficticia** (*dummy coding*) y la **codificación tipo anova o codificación de efectos** (*effect coding*). Para factores ordinales el sistema de codificación más común es la **codificación ortogonal**.

4. RELACIÓN CURVILÍNEA ENTRE DOS VARIABLES

En el punto 3 hemos estudiado la relación lineal entre dos variables, su ecuación y su coeficiente de correlación. Ahora bien, existen muchas variables que no se ajustan a la linealidad, como por ejemplo las características físicas y la edad, e incluso algunas no físicas, como la inteligencia, ya que, en estudios transversales, se observa cómo crece hasta los 20 años aproximadamente y desciende suavemente hasta los 60 años, edad en la que se observa un rápido declinar. Si llamamos X a la edad e Y a la fuerza física, al representar gráficamente la relación se obtiene una similar a la siguiente:



Cuando la relación entre dos variables no es lineal, no es adecuado emplear el coeficiente de correlación de Pearson, pues puede darnos un valor muy próximo a cero aun en los casos de una fuerte relación entre las dos variables. En estos casos hay que utilizar la **razón de correlación** de Y sobre X (η_{yx}).

Siguiendo con el ejemplo de la variable Y (inteligencia, fuerza física, habilidad, etc.) dependiente de la edad (X); imaginemos que tenemos un grupo amplio de unas 100 personas. En este grupo hay distintas edades y, elaborando intervalos de edad, agrupamos a los sujetos en subgrupos o categorías (0-10 años / 11-20 / 21-... etc.).

GRUPO 1	GRUPO 2	.....	GRUPO C
Y_{11}	Y_{12}	.....	Y_{1C}
Y_{21}	Y_{22}	.....	Y_{2C}
.....	.....	.....	.....
Y_{n1}	Y_{n2}	.....	Y_{nc}
Y_1	Y_2	.....	Y_C

Cada grupo de edad tiene su propia media, que en principio, y puesto que Y varía con la edad, será diferente para cada grupo. También existe una media total, una media de Y para las 100 personas (en este caso 60).

Ahora bien, si a cada persona le atribuimos como puntuación en Y la media del grupo total estaremos cometiendo

una determinada cantidad de error. Si en lugar de utilizar la media del grupo total utilizamos la media de cada grupo, de modo que a cada individuo le atribuimos la media de su grupo de edad, también estaremos cometiendo un error, pero, probablemente, de menor cuantía.

La razón de correlación (**al cuadrado**) no es más que lo que acabamos de expresar, esto es, la proporción de error que dejamos de cometer cuando a cada individuo le atribuimos (como pronóstico) la media de su grupo de edad en lugar de la media del grupo total.

$$\eta^2_{YX} = \Sigma\, n_c (\bar{Y}_c - \bar{Y})^2 \Big/ \Sigma\Sigma (Y_{ic} - \bar{Y})^2$$

Obviamente, calculando la raíz cuadrada de la anterior expresión, obtendremos: (η_{YX}).

4.1. PROPIEDADES DE LA RAZÓN DE CORRELACIÓN

• η^2_{YX} es igual o mayor que cero e igual o menor que 1.

• Cuando ($\eta^2_{YX} = 1$) significa que al emplear la media de cada grupo como pronóstico estamos reduciendo a cero el error que habríamos cometido en el caso de emplear como pronóstico la media total.

• Para unos mismos datos siempre ocurre que:

$$r^2_{YX} \leq \eta^2_{YX}$$

Esto ocurre porque Pearson, como hemos dicho anteriormente, sólo mide relación lineal, mientras que la razón de correlación mide cualquier tipo de relación (lineal y no lineal).

• La diferencia ($\eta^2_{YX} - r^2_{YX}$) mide el grado de alejamiento mayor o menor de unos datos de la linealidad. Cuanto mayor sea el valor de esta diferencia, menos lineales serán los datos, y viceversa.

• (η^2_{YX}) es función del número de categorías o subgrupos establecidos en la variable X. Cuando aumenta su número, la razón de correlación al cuadrado también lo hace, y a la inversa.

4.1.1. Interpretación de la razón de correlación

La razón de correlación es un índice de correlación solamente adecuado para el caso de **variables cuantitativas** y que mantengan una relación curvilínea entre las mismas. Sólo respetando estas condiciones puede interpretarse este índice con sentido. La razón de correlación simple tiene una interpretación complicada, ya que la relación puede no ser exclusivamente lineal, por lo que se suele interpretar la razón de correlación al cuadrado, que arroja datos más fácilmente interpretables.

Determinar si (η^2_{YX}) es alta o baja para un conjunto de datos es una tarea similar a la que recomendábamos con el coeficiente de correlación de Pearson. La única manera de determinar si la correlación entre dos variables es alta, media o baja es comparar nuestro estudio con otras investigaciones que hayan estudiado las mismas variables y en condiciones similares.

5. RELACIÓN ENTRE VARIABLES ORDINALES

En el tema introductorio a la estadística decíamos que una variable se encuentra a nivel de medida ordinal cuando entre los números atribuidos a sus modalidades sólo podían establecerse relaciones de igualdad-desigualdad y de orden. Esto es, podemos decir que una modalidad o el número que la representa es distinto a otra, y que es superior o inferior en un orden determinado. Por ejemplo, la puntuación del examen PIR es una variable cuantitativa (podemos decir cuántas veces la diferencia entre dos puntuaciones contiene a otra), pero, como sucede realmente, si utilizamos las distintas puntuaciones asignándoles un orden (el número 1 al sujeto con mayor puntuación en el examen; el número 2 al sujeto con la puntuación siguiente; el número 3 a... etc.) hemos convertido la variable cuantitativa en una variable ordinal. Del número 1 sólo puede decirse que es diferente al 2 y mayor que él, nada más.

En este apartado vamos a estudiar algunos de los principales coeficientes de correlación empleados con variables ordinales.

5.1. COEFICIENTE DE CORRELACIÓN DE SPEARMAN (PIR 05, 88, 93; PIR 12, 15; PIR 14, 28; PIR 15, 8)

Supongamos que tenemos una muestra de cinco personas, con sus correspondientes puntuaciones en el examen PIR (Variable X). Ordenamos esas puntuaciones de mayor a menor, tal como hemos explicado en el apartado anterior. Por otro lado, consideramos también su expediente académico (Variable Y), que, igualmente, transformamos en una variable a nivel ordinal, ordenando las calificaciones de mayor a menor. Si presentamos los datos en una tabla:

	X	Y
SUJETO A	4	2
SUJETO B	2	1
SUJETO C	3	4
SUJETO D	1	3
SUJETO E	5	5

Esta tabla se interpreta del siguiente modo: el sujeto D ha obtenido en el examen (X) la puntuación más alta, con lo cual se le asigna el valor ordinal (1). En cuanto a sus calificaciones académicas, comparándole con sus compañeros, le corresponde el valor ordinal (3), esto es, su calificación académica le coloca en tercer lugar.

No siempre las cosas resultan tan sencillas, pues podría suceder que dos individuos, por ejemplo el sujeto B y C, hubieran obtenido la misma puntuación en el examen (X), con lo cual a ambos les correspondería el 2° puesto o valor ordinal (2). Para evitar esto se realiza la media aritmética de los valores ordinales que les hubieran correspondido de no haber existido empate, en este caso, (2) y (3). A continuación, se le asignaría a ambos sujetos en la variable X el resultado de dicha media (2,5).

Teniendo en cuenta estas observaciones el **coeficiente de correlación de Spearman** viene expresado de este modo:

$$r_S = 1 - (6 \Sigma d_i^2) \Big/ n(n^2 - 1)$$

(d^2_i) es el cuadrado de la diferencia entre el valor ordinal en X y el valor ordinal en Y del sujeto i.

5.1.1. Propiedades del coeficiente de correlación de Spearman

• El coeficiente de correlación de Spearman está comprendido entre (-1) y (+1).

• Una correlación de (+1) significa que cada sujeto ocupa el mismo lugar ordinal en ambas variables, esto es, que al sujeto con valor ordinal (1) en X, le corresponde también el valor (1) en Y; que el sujeto con valor ordinal (2) en X le corresponden el valor (2) en Y, etc.

• Una correlación de (-1) significa que al sujeto con valor ordinal (1) en X le corresponde el último valor ordinal en Y; que al sujeto con valor (2) en X le corresponden el penúltimo valor en Y, etc.

5.2. COEFICIENTE DE CORRELACIÓN DE KENDALL

Supongamos n personas y dos variables X e Y. Si elegimos dos personas, A y B, puede suceder que A sea superior a B en X e inferior a B en Y o inferior a B en X y superior a B en Y. En estos casos se dice que existe una inversión.

Por el contrario, si sucede que A es superior a B en X y también en Y o inferior a B en X e inferior a B en Y, diremos que estamos ante un caso de no-inversión. Para los casos de empates, hay una fórmula especial pero en la práctica se suele aplicar el coeficiente de correlación de Kendall sólo si no aparecen empates. En caso de que aparezcan empates, se emplea el coeficiente de Goodman y Kruskal.

Para calcular el coeficiente de correlación de Kendall realizamos todas las comparaciones entre los pares posibles, esto es, con los n (n - 1) / 2 pares resultantes que se obtienen con n elementos con tal de que cada par difiera de los restantes en uno, al menos, de sus elementos. Si llamamos P al número de no-inversiones y Q al de inversiones el coeficiente de correlación de Kendall vendrá definido por la siguiente expresión:

$$\tau = (P - Q) \Big/ (P + Q)$$

5.2.1. Propiedades del coeficiente de correlación de Kendall

• El coeficiente de correlación de Kendall se encuentra comprendido entre (-1) y (+1).

• Cuando no existe ninguna inversión (Q = 0) entonces τ es (+1), cuando en todos los pares hay inversión (P = 0) entonces τ es (-1).

5.3. COEFICIENTE DE CORRELACIÓN DE GOODMAN Y KRUSKAL

Este coeficiente, de cálculo más complejo, es el apropiado para aquellos casos en que hay muchas observaciones o sujetos y son muy pocos los valores ordinales, produciéndose por tanto numerosos **empates**.

Supongamos que tenemos dos variables X e Y en las que sólo son posibles definir tres valores ordinales para cada una. Si tenemos cincuenta sujetos, es fácil ver que muchos compartirán la misma posición ordinal, produciéndose muchos empates.

Siguiendo con el ejemplo, con estas cincuenta personas podemos formar (50) 49/2 = 1225 pares que difieran al

menos en un elemento. Analizando estos pares son posibles tres condiciones:

Par semejante: cuando la primera persona del par es superior a la segunda tanto en X como en Y o si es inferior a la segunda tanto en X como en Y.

Par desemejante o inverso: si la primera persona es superior a la segunda en X e inferior a ella en Y o si es inferior a la segunda en X y superior a ella en Y.

Par empatado: si la primera persona es igual que la segunda bien sólo en X, bien sólo en Y, bien simultáneamente en X y en Y.

$$\gamma = (n_s - n_d) \Big/ (n_s + n_d)$$

n_s = número de pares semejantes o no-inversos.
n_d = número de pares desemejantes o inversos.

5.3.1. Propiedades del coeficiente de correlación de Goodman y Kruskal

• El coeficiente de correlación de Goodman y Kruskal está comprendido entre (–1) y (+1).

• Cuando todos los pares no empatados son semejantes, γ vale (+1).

• Cuando todos los pares no empatados son desemejantes, γ vale (–1).

6. RELACIÓN ENTRE VARIABLES NOMINALES

Al estudiar la relación entre variables nominales tendremos que tener en cuenta que ninguno de los índices nos dará la dirección de la correlación. Para ello tendremos que ir a la tabla de distribución conjunta y comparar la frecuencia teórica y empírica. La frecuencia teórica, también llamada: esperada, esperada por azar, esperada supuesta verdadera la hipótesis nula, se define como el número de sujetos que van a aparecer en cada una de las casillas de la distribución en el caso de que ambas variables no estén relacionadas. Dicho de otra manera, sería la cantidad de sujetos que caerían en cada casilla si sólo estuviera influyendo el azar.

La frecuencia empírica u observada se define como el número de sujetos que aparecen en cada casilla en nuestra distribución. Dicho de otra manera, serían los sujetos que realmente caen en cada casilla.

Como decíamos al comienzo de este punto, para poder saber si la relación es positiva o negativa se compararían las frecuencias téoricas de cada casilla con sus respectivas casillas empíricas.

6.1. COEFICIENTE Q DE YULE

Este coeficiente sólo es válido para cuando tratamos con dos variables nominales, ambas con sólo dos modalidades.

Supongamos dos variables X e Y, a un nivel de medida nominal, y que constan cada una de ellas de dos modalidades. Su representación en la tabla de frecuencias conjuntas tendría el siguiente aspecto:

		X		
		X_1	X_2	
Y	Y_2	X_1Y_2	X_2Y_2	Y_2
	Y_1	X_1Y_1	X_2Y_1	Y_1
		X_1	X_2	n

Las casillas de tono gris representan las frecuencias conjuntas, esto es, el número de observaciones o sujetos que tienen un valor en X e Y determinado. Como sólo son dos modalidades por dos variables, resultan cuatro tipos de frecuencias conjuntas.

Para el cálculo de la relación Yule propuso el coeficiente Q:

$$Q = ((X_1Y_1)(X_2Y_2) - (X_1Y_2)(X_2Y_1)) \Big/ ((X_1Y_1)(X_2Y_2) + (X_1Y_2)(X_2Y_1))$$

Si la relación es nula, el numerador valdrá cero y Q será cero. Si la relación es perfecta $(X_1Y_2) = (X_2Y_1) = 0$, o $(X_1Y_1) = (X_2Y_2) = 0$. En el primer caso Q valdrá (1) y en el segundo (-1). Sin embargo, puede ocurrir que aunque Q = 1 o Q = –1, la relación no sea necesariamente perfecta. Tampoco su signo habla de la relación entre variables, ya que el signo viene determinado por la organización del cuadro de frecuencias. No olvidemos que para interpretar Q es necesario consultar la tabla de frecuencias, pues su significado dependerá de la colocación de las variables.

Q no puede ser mayor de (+1) ni menor de (-1).

6.2. COEFICIENTE χ^2

Este estadístico se denomina así porque su distribución de probabilidad se aproxima a la distribución de probabilidad llamada χ^2, a medida que aumenta más y más el tamaño de la muestra.

χ^2 es función del tamaño de la muestra, de modo que, si mantenemos constantes las proporciones de la tabla de frecuencias conjuntas, pero aumentamos n, sucede que aquél también aumenta, aunque esto no implica que la relación entre las variables aumente. Este coeficiente se calcula con el objeto de identificar el coeficiente de contingencia, que veremos más adelante.

Tiene la ventaja frente a Q de que puede ser aplicado sin la restricción de dos modalidades por variable. Supongamos las dos variables anteriores, pero ahora añadimos una tercera modalidad a la variable X. La distribución de frecuencias tendrá el siguiente aspecto:

		X			
		X_1	X_2	X_3	
Y	Y_2	X_1Y_2	X_2Y_2	X_3Y_2	Y_2
	Y_1	X_1Y_1	X_2Y_1	X_3Y_1	Y_1
		X_1	X_2		n

Llamemos **frecuencias empíricas u observadas** a las seis frecuencias conjuntas de la tabla de frecuencias conjuntas anterior. Por otra parte, definamos la **frecuencia teórica** (o frecuencia esperada por azar o frecuencias esperada supuesta verdadera la hipótesis nula) como el producto de las dos frecuencias marginales dividido por el número total de personas **(PIR 15, 3)**. A continuación, calculamos la frecuencia teórica para cada casilla (en la tabla se muestra cómo se calcula la frecuencia teórica de cada casilla). Finalmente, si las frecuencias empíricas, las obtenidas de la muestra de sujetos, coinciden con las frecuencias teóricas calculadas diremos que no existe relación entre esas dos variables, o, lo que es lo mismo, que son independientes para el grupo de personas que estamos estudiando. Por el contrario, si no coinciden, concluiremos que no son independientes para ese grupo, es decir, que existe cierta relación entre X e Y. Cuanta más diferencia entre las frecuencias empíricas y las teóricas mayor grado de relación. Esta diferencia vendrá dada por la fórmula:

$$\chi^2 = \sum (f_e - f_t)^2 \Big/ f_t$$

Con tablas de dos filas por dos columnas, y cuando una de las frecuencias teóricas es muy pequeña (menos de 10), se aplica la corrección de Yates, que consiste en restarle al numerador (0,5).

6.3. COEFICIENTE DE CONTINGENCIA, C

Para contrarrestar el efecto del aumento del estadístico χ^2 en función del tamaño de la muestra, se utiliza el coeficiente de contingencia o C:

$$C = \sqrt{\chi^2 \Big/ (\chi^2 + n)}$$

El coeficiente de contingencia será siempre menor que (1) o mayor o igual que (0) **(PIR 04, 91)**, y siempre va a adquirir valores positivos. Nunca podría llegar a 1, ya que el denominador siempre va a ser mayor que el numerador. C nos va a indicar la fuerza de la relación entre las dos variables, si bien el sentido de la correlación nos lo indicará la tabla de frecuencias conjuntas.

C es función del número de filas y columnas, de manera que sólo son comparables dos coeficientes de contingencia si se han calculado a partir de tablas con el mismo número de columnas y filas.

Si C se calcula a partir de una tabla con igual número (k) de filas y columnas, podremos hallar el valor máximo de C (aquél que alcanzaría C para esa tabla en el caso de que la correlación fuera perfecta). Este valor máximo nos servirá para apreciar la magnitud de la correlación entre las dos variables. El valor máximo de C viene dado por la siguiente expresión:

$$C_{máx} = \sqrt{(k-1)/k}$$

7. OTROS COEFICIENTES DE CORRELACIÓN

Para terminar este tema queremos hacer mención a una serie de coeficientes de correlación que se emplean en variables con unas condiciones muy especiales, como es el caso de las variables dicotómicas o dicotomizadas.

Variable dicotómica:

Es aquella variable que sólo puede adoptar dos modalidades: sexo, estado vital (vivo/muerto).

Variable dicotomizada:

Es aquella variable que pudiendo adoptar varias modalidades, se la obliga artificialmente a adoptar sólo dos: calificación académica (aprobado/suspenso); altura (alto/bajo).

En estos casos se suelen emplear los siguientes coeficientes de correlación:

Coeficiente de correlación biserial puntual: r_{bp}

Es una aplicación del coeficiente de correlación de Pearson a dos variables, siendo una de ellas continua y la otra dicotómica **(PIR 15, 10)**. Su valor está comprendido entre (–1) y (+1).

Coeficiente de correlación: φ (PHI)

Es una aplicación de Pearson a dos variables, ambas dicotómicas **(PIR 05, 100)**. Los límites son idénticos a los del anterior coeficiente.

Coeficiente de correlación biserial: r_b

Supongamos dos variables, ambas continuas. Una de ellas aparece como continua, la otra aparece dicotomizada artificialmente **(PIR 02, 169)**. El coeficiente de correlación biserial es una estimación de Pearson si la variable dicotomizada artificialmente se hubiera mantenido continua y se cumplen las condiciones de que la distribución de Y considerada como continua es normal y la relación entre X e Y (consideradas ambas como continuas) es lineal. Para unos mismos datos, el coeficiente de correlación biserial es siempre mayor o igual que el coeficiente de correlación biserial puntual. Puede valer más que uno y menos que menos uno.

Coeficiente de correlación tetracórica: r_t

Supongamos dos variables continuas. Ambas aparecen dicotomizadas artificialmente. El coeficiente de correlación tetracórica es una estimación de Pearson si ambas variables se hubieran mantenido continuas, tienen una distribución normal y su relación es lineal. Para unos mismos datos, el coeficiente de correlación tetracórica es siempre mayor o igual al coeficiente de correlación φ. Su valor, igual que el de Pearson, se encuentra entre –1 y +1. Tanto desde el punto de vista cuantitativo como cualitativo, su interpretación es igual que el coeficiente de Pearson.

En cualquier procedimiento científico de observación o diagnóstico es muy importante determinar la fiabilidad del procedimiento utilizado y/o de los observadores implicados. Este tipo de fiabilidad se denomina **fiabilidad inter-jueces o inter-evaluadores**, y para su cálculo se recurre a coeficientes de correlación que manejan variables a escala nominal.

Supongamos que tenemos dos observadores que han de registrar la presencia o ausencia de una conducta a lo largo de una serie de intervalos temporales. En la mayoría de los intervalos coincidirán en apreciar la presencia o ausencia de la conducta (acuerdos), pero en algunos se producirán discrepancias en el juicio emitido (desacuerdos). Un procedimiento utilizado para evaluar la fiabilidad de los observadores es el **Porcentaje de acuerdo**:

$$P = \frac{\text{núm. acuerdos}}{\text{núm. acuerdos} + \text{núm. desacuerdos}} \times 100$$

El problema de este índice es que sobrevalora el grado de acuerdo, puesto que no tienen en cuenta que algunos de los acuerdos pueden producirse por azar. Para superar esta dificultad, Cohen propuso el coeficiente **Kappa** **(PIR 07, 55)**:

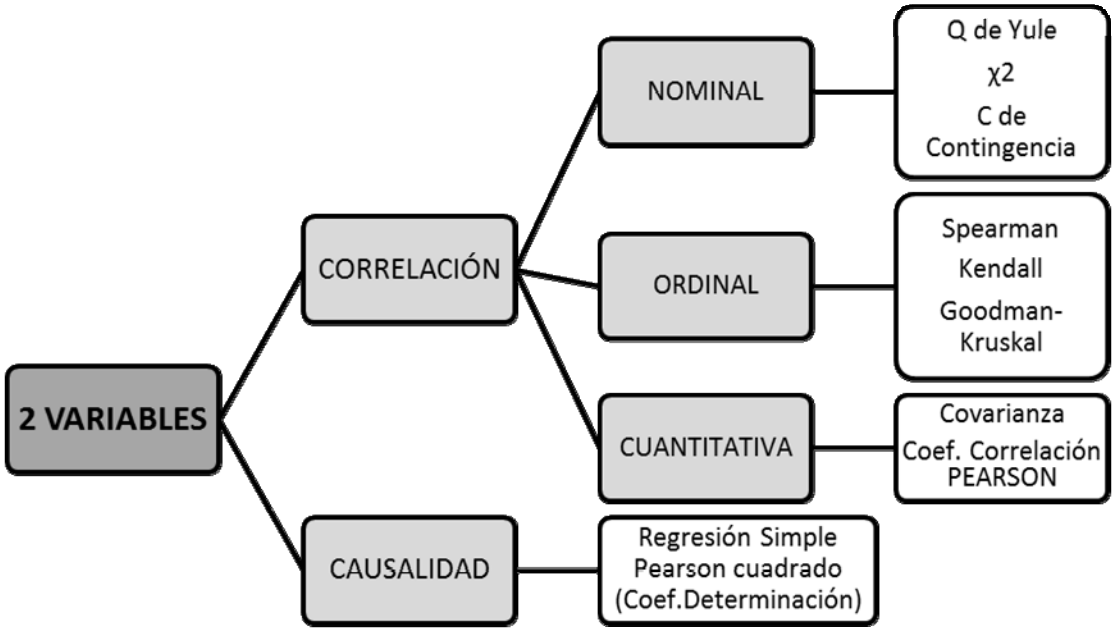
$$K = \frac{P_0 - P_E}{1 - P_E} \times 100$$

Donde P_0 es la proporción de acuerdos obtenidos y Pe es la proporción de acuerdos esperados por azar, que se calcula del siguiente modo:

$$P_E = p_1 \times p_2$$

Siendo P_1 la proporción de ocurrencia de la conducta para el observador 1 y P_2 la proporción de ocurrencia para el observador 2.

ESQUEMA DE CONTENIDOS



	1 variable continua y otra...	Las dos variables	
Dicotómicas	BISERIAL PUNTUAL : r_{bp}	PHI *	APLICACIONES DE PEARSON
Dicotomizadas	BISERIAL: r_b *	TETRACÓRICA: r_t	ESTIMACIONES DE PEARSON