Universidad De Atacama

Facultad De Ciencias Jurídicas / Carrera De Trabajo Social

TECNOLOGÍA INFORMÁTICA I (SPSS)

ESTADÍSTICA DESCRIPTIVA CON MÁS DE UNA VARIABLE

Profesor: Hugo S. Salinas. Primer Semestre 2008

1. Relación Lineal

1.1. Tipos de relación lineal

1. DIRECTA o POSITIVA

   Por ejemplo las variables Inteligencia y Rendimiento.

Tabla 1: Inteligencia y Rendimiento.

X	3	1	2	3	1
Y	6	3	5	7	4

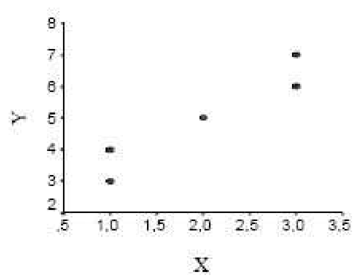

Figura 1: Inteligencia y Rendimiento.

2. INVERSA O NEGATIVA

   Por ejemplo las variables Ansiedad y Aciertos.

Tabla 2: Ansiedad y Aciertos.

S	2	4	6	2	1
T	4	2	1	3	5

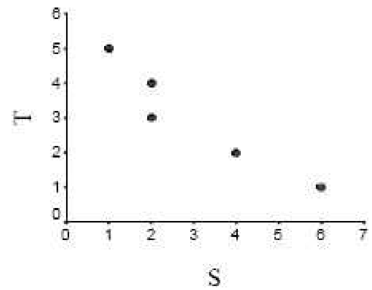

Figura 2: Ansiedad y Aciertos.

3. RELACIÓN LINEAL NULA

   Por ejemplo las variables Extroversión y Absentismo.

Tabla 3: Extroversión y Absentismo.

U	2	4	4	2	2
W	5	4	7	4	8

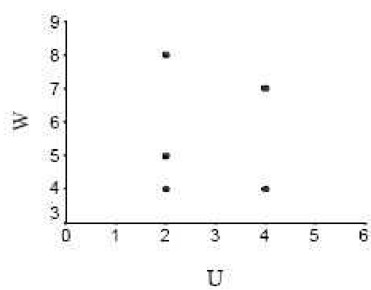

Figura 3: Extroversión y Absentismo.

1.2. Representación gráfica

1.2.1. Diagrama de Dispersión

Se dispone de n pares de puntuaciones (en el ejemplo para 10 sujetos) en dos variables cuantitativas (Tienen que referirse a los mismos sujetos).

Tabla 4: Ejemplo para 10 sujetos.

Sujeto	1	2	3	4	5	6	7	8	9	10
X	4	4	5	5	7	9	9	9	10	10
Y	6	7	8	10	12	10	13	15	13	15

Aquí $\overline{X} = 7{,}2$ y $\overline{Y} = 10{,}9$. Entonces el gráfico de dispersion es:

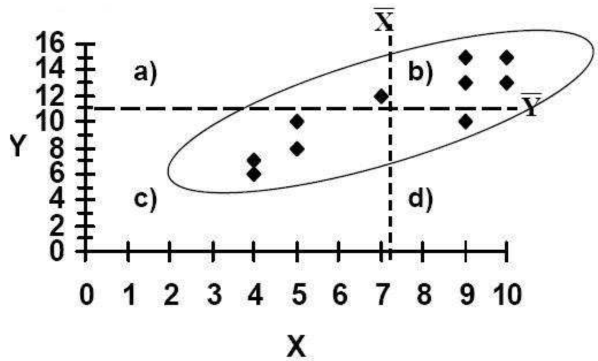

Figura 4: Gráfico de dispersion.

2. Cuantificación de la relación lineal

2.1. Covarianza

$$S_{XY} = \overline{xy} - \overline{x}\,\overline{y}, \qquad -\infty \leq S_{XY} \leq \infty, \tag{1}$$

donde $\overline{x} = \frac{1}{n}\sum x_i$ y $\overline{xy} = \frac{1}{n}\sum x_i y_i$

2.2. Correlación de Pearson

$$r_{XY} = \frac{S_{XY}}{S_X\, S_Y}, \qquad -1 \leq r_{XY} \leq 1, \tag{2}$$

NOTA: En transformaciones lineales la covarianza se altera pero la correlación de Pearson no. En efecto, si: $U = a\,X + c$ y $V = c\,Y + d$, entonces $S_{UV} = a\,b\,S_{XY}$ y $r_{UV} = r_{XY}$ (Si a y b tienen signos contrarios, entonces $r_{UV} = -r_{XY}$).

2.3. Matriz de varianzas y covarianzas

$$S = \begin{pmatrix} S_X^2 & S_{XY} & \cdots & S_{XZ} \\ & S_Y^2 & \cdots & S_{YZ} \\ & & \cdots & \vdots \\ & & & S_Z^2 \end{pmatrix},$$

2.4. Matriz de correlaciones

$$R = \begin{pmatrix} 1 & r_{XY} & \cdots & r_{XZ} \\ & 1 & \cdots & r_{YZ} \\ & & \cdots & \vdots \\ & & & 1 \end{pmatrix},$$

donde $r_{XY} = \frac{S_{XY}}{S_X\, S_Y}, \qquad -1 \leq r_{XY} \leq 1,$

2.4.1. Interpretación de r_{XY}

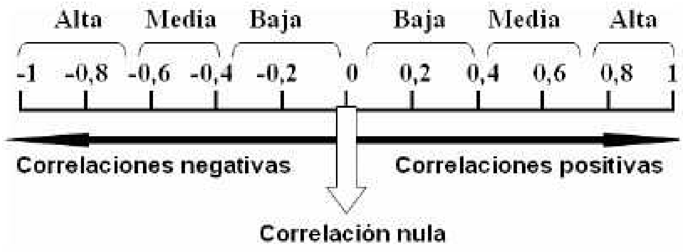

Figura 5: Interpretation de r_{XY}

1. Examinar su magnitud absoluta:
   a) Si $|r_{XY}| = 0$, entonces hay relación lineal nula.
   b) Si $|r_{XY}| \to 1$, entonces hay relación lineal
2. Examinar su signo:
   a) Si $r_{XY} > 0$, entonces hay relación lineal directa
   b) Si $r_{XY} < 0$, entonces hay Relación lineal inversa

2.4.2. Propiedades de r_{XY}

1. $r_{XY}^2 \times 100\,\%$ de variabilidad común entre X e Y. Por ejemplo si $r_{XY} = 0{,}5$, entonces existe un 25 % de variabilidad común entre las variables X e Y.
2. $r_{XY} = 0$ no implica que no haya relación entre X e Y (la relación puede ser de otro tipo).
3. Factores que afectan a r_{XY}:
   a) Variabilidad de X, Y y XY,
   b) Terceras variables (efectos moderadores),
   c) Que X e Y estén bien medidas (fiabilidad) y
   d) La muestra en que se evalúen X e Y.
4. Para decidir si X e Y están linealmente relacionadas, se realiza una **prueba de significación estadística** (contraste de hipótesis sobre ρ_{XY})
5. La correlación **NUNCA IMPLICA CAUSALIDAD**, sólo grado de relación lineal.

3. Regresión Lineal

3.1. Introducción

OBJETIVO: Hacer predicciones o pronósticos en una variable (Y) a partir de otra (X): Regresión de Y sobre X.

- Variable predictora: X (o variable independiente). La que se utiliza para hacer pronósticos.

- Variable criterio: Y (o variable dependiente). Sobre la que se pronostica

Para esto, se crea un MODELO (la recta de regresión $Y = \beta_0 + \beta_1 X$, donde β_0 es el origen y β_1 la pendiente) que se aproxime o ajuste lo más posible a los datos observados en X y en Y.

Ejemplos:
X: Minutos en llamada; Y: Gasto en el teléfono móvil (en miles pesos/mes)

X: Ansiedad ; Y: número de aciertos en la PSU.

3.2. Identificación del modelo

3.2.1. Criterio de mínimos cuadrados

Se trata de encontrar aquella recta de regresión $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X$ en que la distancia al cuadrado entre ella y las puntuaciones observadas sea mínima. Es decir:

$$\frac{1}{n}\sum (Y - \widehat{Y})^2 \to 0, \tag{3}$$

De modo que se tiene lo siguiente: $\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$, donde $\widehat{\beta}_1 = \frac{S_{XY}}{S_X^2} = r_{XY}\,\frac{S_Y}{S_X}$.

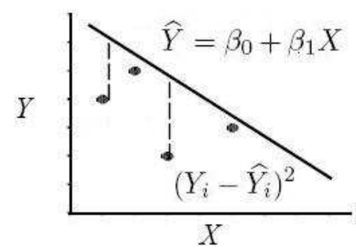

Figura 6: Estimación de Mínimos Cuadrados

3.3. Valoración del modelo

3.3.1. Coeficiente de determinación

r_{XY}^2 : Proporción de varianza de Y que queda explicada por X.
Descomposición de la varianza del criterio (Y):

$$Y_i = \widehat{Y}_i + (Y_i - \widehat{Y}_i), \tag{4}$$

donde:

- Y_i son las puntuaciones observadas (empíricas).
- $\widehat{Y}_i$ son las puntuaciones pronosticadas (estimadas).
- $Y_i - \widehat{Y}_i$ son los errores en los pronósticos.

Por lo tanto, la varianza de $Y_i = \widehat{Y}_i + (Y_i - \widehat{Y}_i)$ está dada por:

$$S_Y^2 = S_{\widehat{Y}}^2 + S_{Y-\widehat{Y}}^2 + 2\,S_{\widehat{Y}(Y-\widehat{Y}_i)} \tag{5}$$

donde:

- $S^2_{\widehat{Y}} = \frac{1}{n}\sum(\widehat{Y}_i - \overline{\widehat{Y}})^2$,
- $S^2_{Y-\widehat{Y}} = \frac{1}{n}\sum(Y_i - \widehat{Y})^2$,
- $S_{\widehat{Y}(Y-\widehat{Y})} = 0$,

Luego, reemplazando:

$$S^2_Y = S^2_{\widehat{Y}} + S^2_{Y-\widehat{Y}}, \tag{6}$$

donde la varianza de la variable dependiente (criterio) S^2_Y se descompone en:

- $S^2_{\widehat{Y}}$: varianza de las estimaciones (pronósticos). **Varianza explicada**.
- $S^2_{Y-\widehat{Y}}$: error cuadrático medio. **Varianza No explicada**.