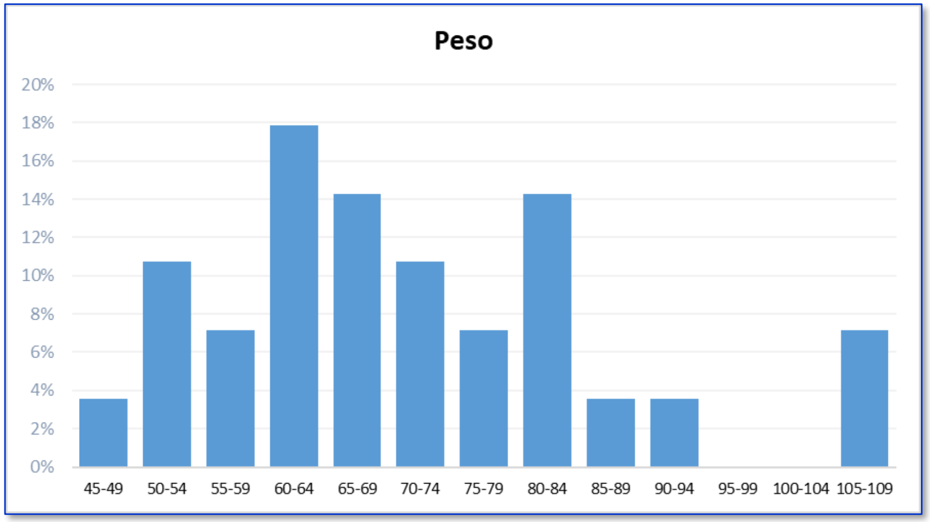
I.6. Ejemplos

Ejemplo 2. En un curso de Estadística Aplicada II, se solicitó a 28 alumnos que registraran su estatura (en metros y centímetros) y peso (en kilos hasta décimas). Los resultados aparecen en la siguiente Tabla.

Caso	Estatura	Peso	Caso	Estatura	Peso
1	1.74	85.0	*15*	1.72	76.0
2	1.58	54.0	*16*	1.58	47.0
3	1.80	70.0	*17*	1.63	60.0
4	1.77	70.0	*18*	1.67	70.0
5	1.64	56.0	*19*	1.65	60.0
6	1.68	62.0	*20*	1.83	69.0
7	1.63	68.0	*21*	1.60	60.0
8	1.79	64.0	*22*	1.80	83.0
9	1.82	80.0	*23*	1.91	107.3
10	1.72	65.0	*24*	1.90	105.0
11	1.66	50.0	*25*	1.75	80.0
12	1.82	90.0	*26*	1.72	67.5
13	1.67	58.0	*27*	1.64	54.0
14	1.76	77.0	*28*	1.78	80.0

I.6. Ejemplos

Imagine que, para empezar, es de interés describir la variable Peso en este banco de datos. Con tal propósito, se pueden producir algunos resúmenes, tanto gráficos como numéricos.



Mínimo = 47.0
Media = 70.3
Máximo = 107.3

Figura 4. *Distribución de frecuencias del Peso.*

I.6. Ejemplos

De la gráfica se observa que, aproximadamente 93% de los casos presentan un peso entre los 45 y los 94 kilos. Las excepciones, dos personas, se localizan entre 105 y 109 kilos.

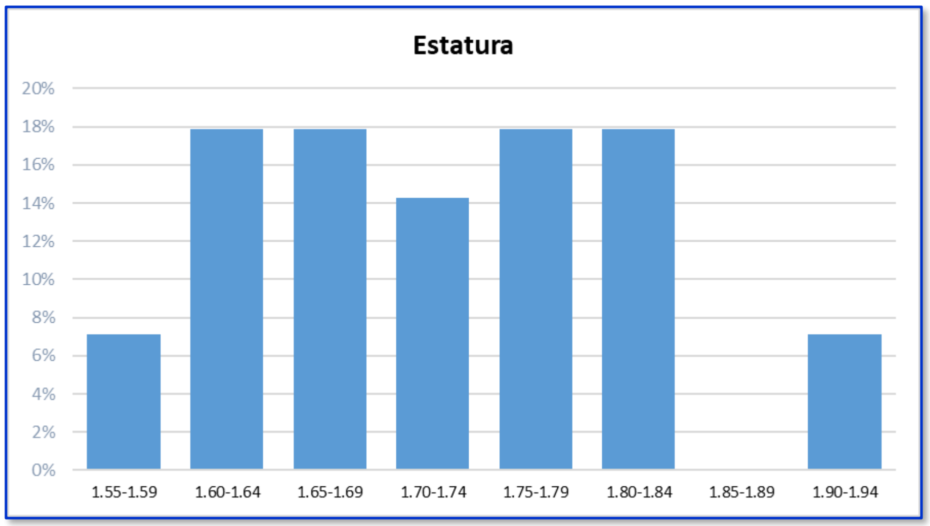
Una pregunta interesante es la siguiente: los dos casos con valores de la variable peso superiores al resto, ¿presentan sobrepeso u obesidad? Y de ser así, ¿serán los únicos candidatos para estas clasificaciones?

No es necesario ser Médico para coincidir en que el sobrepeso y la obesidad dependen del peso pero no solamente del peso. En particular, no es lo mismo pesar, digamos, 100 kilos con una estatura de 1.6 metros que tener ese peso con una estatura de, por ejemplo, 2.1 metros.

Por esta razón es conveniente considerar la información de la variable Estatura en este grupo de personas.

I.6. Ejemplos

El patrón de la Estatura tiene parecido con el de los pesos. La mayoría de casos entre 1.55 y 1.84 metros y dos casos en el intervalo 1.90-1.94 metros. Sugiere una asociación entre las dos variables.



Mínimo = 1.58
Media = 1.91
Máximo = 1.72

Figura 5. *Distribución de frecuencias de Estatura.*

I.6. Ejemplos

El siguiente diagrama de puntos, también conocido como *scatter plot*, es una representación gráfica de la distribución de frecuencias conjuntas de las variables Peso y Estatura. Lo que puede observarse es una *tendencia* monótona creciente en este conjunto de datos.

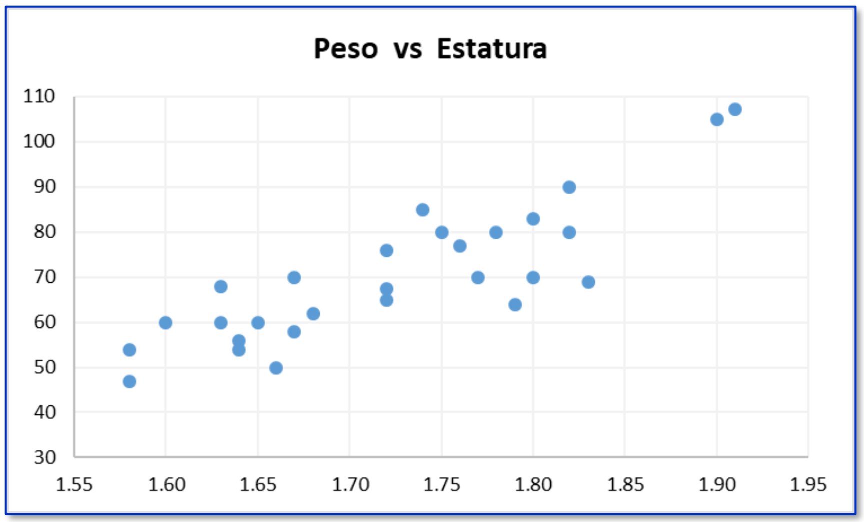


Figura 6. *Diagrama de puntos (scatterplot) Peso contra Estatura.*

I.6. Ejemplos

La tendencia monótona creciente significa que un incremento en la Estatura *tiende* a presentarse asociado a un incremento en el Peso. Una tendencia es una relación que se manifiesta con excepciones. La recta en la siguiente gráfica representa una *relación* lineal entre Peso y Estatura.

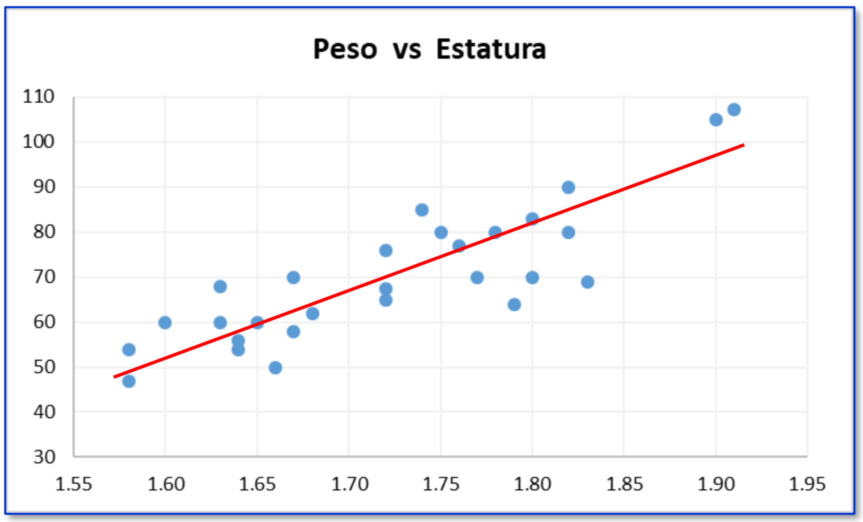


Figura 7. *Tendencia y Relación.*

I.6. Ejemplos

Mientras que en una relación, a un incremento en la Estatura *siempre* corresponde un incremento en el Peso, en el caso de una tendencia, el incremento ocurre en general pero, *con excepciones*.

Para precisar estas ideas, sea Y la variable Peso y X la Estatura. Entonces, una *relación lineal* entre Y y X se representa como:

$$Y = \beta_0 + \beta_1 X.$$

Por otra parte, una *tendencia lineal* entre estas variables se puede describir de la siguiente manera:

$$Y \approx \beta_0 + \beta_1 X$$

o, con una notación mas precisa,

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

donde ε es un elemento que describe las diferencias entre la *relación* y la *tendencia*. En términos del proceso de modelado, si X es la *variable explicativa* con la que se busca describir a Y, entonces ε representa el efecto (aditivo) de todos los factores (diferentes a la Estatura) que pueden estar asociados con Y.

Una situación en la cual <u>no</u> podríamos esperar que el modelo lineal

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \ldots \quad (ML)$$

nos fuera útil para describir una relación entre Y y X, puede darse si por ejemplo el diagrama de dispersión no muestra una forma de asociación

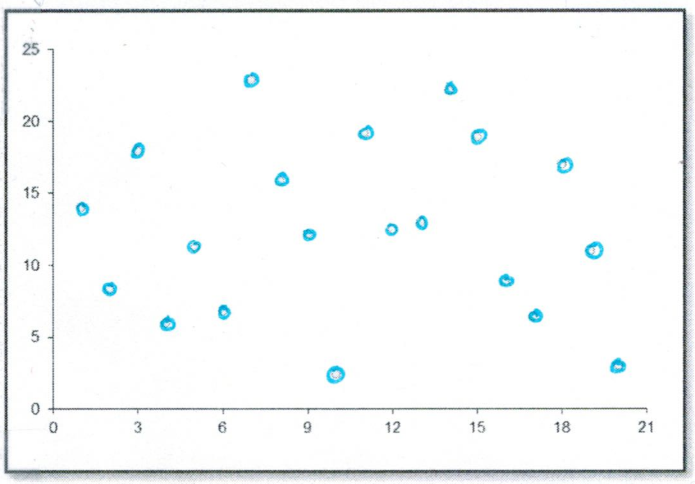


Figura 9. *Sin evidencia de asociación.*

En este caso, de hecho no podemos proponer una relación funcional entre X y Y. Otro caso en el que el modelo (ML) tampoco sería util, es cuando de existir alguna regla de asociación entre X y Y, esta no es lineal, pero si está dada a través de una relación más compleja, quizás un polinomio de orden mayor.

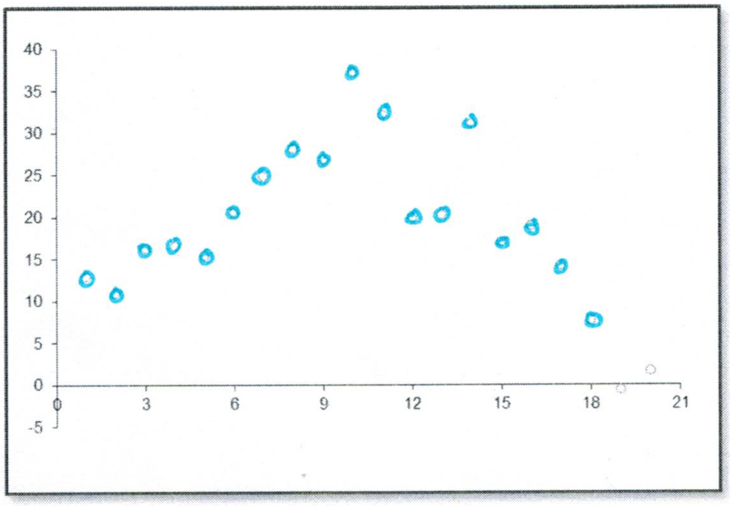


Figura 10. *Tendencia Compleja* (NO LINEAL)

Para este caso se puede intentar "el modelo lineal"(1)

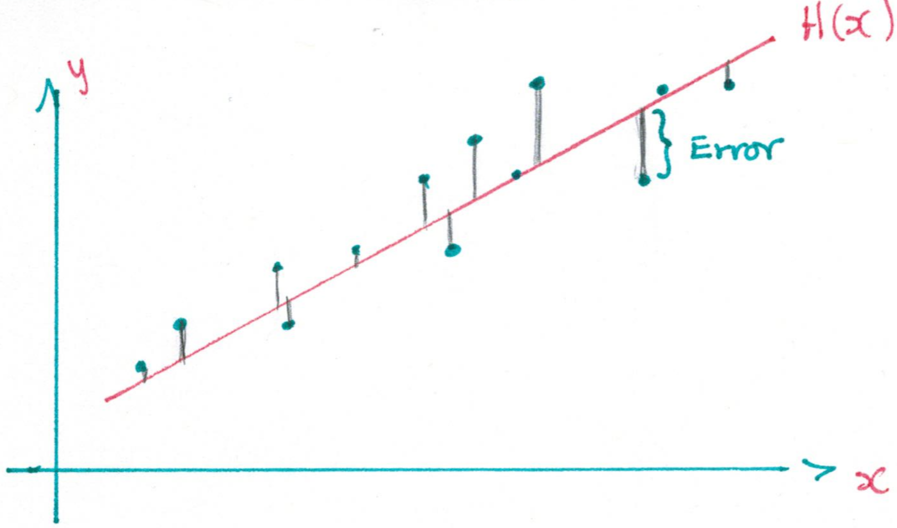
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3.$$

Una vez que se ha encontrado evidencia de una posible relación de asociación entre Y y X, el problema es producir una función que describa esta asociación en forma aproximada. Generalmente un conjunto de datos no se puede describir de manera perfecta con una función, ya que en muchos conjuntos de datos se presentan pares (x_1, y_1) y (x_2, y_2) en donde $x_1 = x_2$ pero $y_1 \neq y_2$, un ejemplo de lo anterior se da en los datos de estaturas y pesos

(1) Más adelante veremos es posible definir $X_1 \equiv X$; $X_2 \equiv X^2$ $X_3 \equiv X^3$ y escribir $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ una función lineal de las variables explicativas X_1, X_2, X_3

dos personas pueden tener la misma estatura pero su peso es diferente. De cualquier manera podemos pensar en encontrar "la mejor aproximación" a la relación entre X y Y usando una curva. Esta curva puede seleccionarse como "la más cercana" a los pares $(x_1, y_1), \ldots, (x_N, y_N)$, como nuestro objetivo es describir a Y para cada valor fijo de X, entonces la distancia apropiada a considerar es, para cada valor fijo x de X, la distancia que existe entre el valor y que le corresponde en los datos y el valor de $H(x)$. Así, si se cuenta con un conjunto de N datos $(x_1, y_1), \ldots, (x_N, y_N)$ y se busca describir que tan cerca se encuentra la función $H(x)$ a la nube de puntos, es posible evaluar la distancia de cada punto en el conjunto de datos a la curva descrita por la función

X	Y	H(x)	Error
x_1	y_1	$H(x_1)$	$y_1 - H(x_1)$
x_2	y_2	$H(x_2)$	$y_2 - H(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_N	y_N	$H(x_N)$	$y_N - H(x_N)$



El problema es hallar $H(x)$ tal que $\|\mathbf{e}\|$ sea lo más pequeña posible

$$\mathbf{e} = \begin{pmatrix} y_1 - H(x_1) \\ y_2 - H(x_2) \\ \vdots \\ y_N - H(x_N) \end{pmatrix}$$

la norma $\|\cdot\|$ puede seleccionarse de entre varias normas posibles, aquí usaremos $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{N} x_i^2}$ ya que este problema ha sido estudiado con esta elección y se han encontrado propiedades matemáticas que permiten un análisis más inmediato[1].

Para el caso particular en que $H(x) = \beta_0 + \beta_1 x$,

(1) Entre otras cosas debido a que se puede utilizar cálculo diferencial para minimizar $\|\mathbf{e}\|_2$

encontrar $H(x)$ que minimize $\|\mathbf{e}\|_2$ corresponde con encontrar β_0 y β_1 tales que

$$\|\mathbf{e}\|_2 = \sqrt{\sum_{i=1}^{N} (y_i - H(x_i))^2}$$

$$= \sqrt{\sum_{i=1}^{N} (y_i - (\beta_0 + \beta_1 x_i))^2} \equiv \sqrt{\Delta(\beta_0, \beta_1)}$$

sea lo más pequeña posible. En otras palabras la tarea es minimizar $\Delta(\beta_0, \beta_1)$ con respecto a β_0 y β_1 [(1)], ya que la raíz cuadrada es una función monótona en los reales no negativos, de donde encontrar β_0 y β_1 que minimizan $\|\mathbf{e}\|_2$ equivale a encontrar β_0 y β_1 que minimizan $\Delta(\beta_0, \beta_1)$.

Para encontrar el mínimo de $\Delta(\beta_0, \beta_1)$ y los valores β_0 y β_1 donde este mínimo se alcanza, notemos que

(1) Conocido como el método de Mínimos Cuadrados, esta idea fue generada por Gauss y Legendre a principios del siglo 19

como función de β_0 y β_1, Δ es un polinomio de grado dos y en particular tiene derivadas en todo su dominio (en todo $\mathbb{R}^2$). Entonces podemos usar cálculo diferencial y calcular derivadas parciales con respecto a β_0 y β_1 para encontrar los puntos críticos de $\Delta(\beta_0, \beta_1)$:

$$\frac{\partial \Delta(\beta_0,\beta_1)}{\partial \beta_0} = \sum_{i=1}^{N} 2(y_i - \beta_0 - \beta_1 x_i)(-1)$$

$$= -2\left\{\sum_{i=1}^{n} y_i - \beta_0 N - \beta_1 \sum_{i=1}^{N} x_i\right\}$$

$$\frac{\partial \Delta(\beta_0,\beta_1)}{\partial \beta_1} = \sum_{i=1}^{N} 2(y_i - \beta_0 - \beta_1 x_i)(-x_i)$$

$$= -2\left\{\sum_{i=1}^{N} y_i x_i - \beta_0 \sum_{i=1}^{N} x_i - \beta_1 \sum_{i=1}^{N} x_i^2\right\}$$

A continuación debemos resolver el sistema de ecuaciones simultáneas

$$\frac{\partial \Delta(\beta_0,\beta_1)}{\partial \beta_0} = 0$$

$$\frac{\partial \Delta(\beta_0,\beta_1)}{\partial \beta_1} = 0$$

Este sistema de ecuaciones es

$$\sum_{i=1}^{N} y_i - \beta_0 N - \beta_1 \sum_{i=1}^{N} x_i = 0$$

$$\sum_{i=1}^{N} y_i x_i - \beta_0 \sum_{i=1}^{N} x_i - \beta_1 \sum_{i=1}^{N} x_i^2 = 0$$

ó equivalentemente

$$\begin{aligned} \beta_0 N + \beta_1 \sum_{i=1}^{N} x_i &= \sum_{i=1}^{N} y_i \\ \beta_0 \sum_{i=1}^{N} x_i + \beta_1 \sum_{i=1}^{N} x_i^2 &= \sum_{i=1}^{N} x_i y_i \end{aligned} \quad \ldots\ldots \quad (EN)$$

El sistema de ecuaciones es conocido como Sistema Normal de ecuaciones (ó ecuaciones Normales) y es un sistema de ecuaciones lineales en β_0 y β_1. Este sistema tiene solución siempre que el discriminante

$$D = \begin{vmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{vmatrix} = N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2$$

$$= N\left(\sum_{i=1}^{N} x_i^2 - N\bar{x}^2\right)$$

sea distinto de 0. Notemos que D es una función de los datos $x_1, \ldots, x_N$, pero NO depende de los valores $y_1, \ldots, y_N$

¿Qué restricción sobre $x_1, \ldots, x_N$ hay que hacer, para que exista la solución (los valores de β_0 y β_1 donde Δ puede tener un mínimo)?

Recordemos que

$$\sum_{i=1}^{N} x_i^2 - N\bar{x}^2 = \sum_{i=1}^{N} x_i^2 - 2N\bar{x}^2 + N\bar{x}^2$$

$$= \sum_{i=1}^{N} x_i^2 - 2N\left(\frac{\sum_{1}^{N} x_i}{N}\right)\left(\frac{\sum_{1}^{N} x_i}{N}\right) + N\bar{x}^2$$

$$= \sum_{i=1}^{N} x_i^2 - 2\left(\sum_{i=1}^{N} x_i\right)\bar{x} + \sum_{i=1}^{N} \bar{x}^2$$

$$= \sum_{i=1}^{N} \left\{ x_i^2 - 2x_i\bar{x} + \bar{x}^2 \right\}$$

$$= \sum_{i=1}^{N} (x_i - \bar{x})^2 = N\, S_{xx}$$

donde $S_{xx} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2$ es la varianza muestral correspondiente a $x_1, \ldots, x_N$. Claramente si $D=0$ entonces $S_{xx}=0$, lo cual implicaría que $x_1 = x_2 = \cdots = x_N$. Por otra parte, si $S_{xx}=0$ (de nuevo sucede si y sólo si $x_1 = x_2 = \cdots = x_N$) entonces $D=0$.

$\therefore$ La solución de (EN) no existe solo cuando los datos $(x_1, y_1), \ldots, (x_N, y_N)$ tienen <u>todos</u> el mismo valor de la variable explicativa: $x_1 = x_2 = \cdots = x_N$, pero lo anterior es un caso que carece de todo sentido práctico, ¡a nosotros nos interesa modelar como cambia Y cuando X cambia!

En conclusión, basta con que existan x_i y x_j con $i \neq j$ en $\{x_1, \ldots, x_N\}$ tales que $x_i \neq x_j$ para que exista la solución al sistema (EN).

Procedemos a determinar la solución, usando la regla de Cramer

$$\hat{\beta}_1 = \frac{\begin{vmatrix} N & \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i y_i \end{vmatrix}}{N^2 S_{xx}}$$

$$= \frac{N \sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N^2 S_{xx}}$$

$$= \frac{N\left\{\sum_{i=1}^{N} x_i y_i - N \bar{x}\bar{y}\right\}}{N^2 S_{xx}}$$

$$= \frac{\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y} + N\bar{x}\bar{y} - N\bar{x}\bar{y}}{N S_{xx}}$$

$$= \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{N S_{xx}}$$

NOTACIÓN: $N S_{xy} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$

$$\hat{\beta}_1 = \frac{N\,S_{xy}}{N\,S_{xx}} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})}$$

NOTA: En algunas fuentes se usa la notación

$$S_{xy} = \sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})$$

$$S_{xx} = \sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})$$

en cuyo caso $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$

Usaremos la primera ecuación del sistema para encontrar $\hat{\beta}_0$

$$\hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i$$

$$\Rightarrow \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x}$$

Definición: Los coeficientes del ajuste lineal via mínimos cuadrados son

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

La correspondiente Recta de mínimos cuadrados es

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

<u>ejemplo:</u> Para los datos de Estatura y Peso, a partir de los 28 datos en la primera página del archivo "Regre5.pdf", obtenemos

$$\bar{x} = 1.72 \qquad \bar{y} = 70.3$$

$$\hat{\beta}_0 = -169.94 \qquad \hat{\beta}_1 = 139.37$$

Recta de mínimos cuadrados

$$\hat{Y} = -169.94 + 139.37 X$$

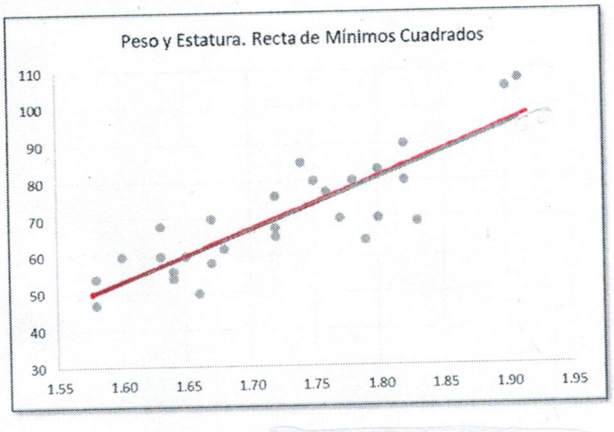


Figura 14.

Una pregunta interesante: De acuerdo a este ajuste, a una persona que mida cero metros, le corresponde un peso de -170 Kilos, es claro que este valor negativo resulta contra intuitivo ¿Cuál es el problema ó cómo podemos interpretar o entender este?

Toda la descripción del método de mínimos cuadrados dada, no asume supuestos sobre una distribución de probabilidades para Y ó X o $Y|X=x$,[1] de manera que el método de mínimos cuadrados en su formulación original NO es un procedimiento estadístico, es una técnica determinista para el ajuste de curvas.

A continuación evaluaremos que tan pequeño es el mínimo encontrado

(1) En todo caso, para la existencia de soluciones sólo se asume $\mathbb{P}(X=c) \neq 1$.

Una pregunta interesante: De acuerdo a este ajuste, a una persona que mida cero metros, le corresponde un peso de -170 Kilos, es claro que este valor negativo resulta contra intuitivo ¿Cuál es el problema ó cómo podemos interpretar o entender este?

Toda la descripción del método de mínimos cuadrados dada, <u>no</u> asume supuestos sobre una distribución de probabilidades para Y ó X o $Y|X=x$ (1), de manera que el método de mínimos cuadrados en su formulación original <u>NO</u> es un procedimiento estadístico, es una técnica determinista para el ajuste de curvas.

A continuación evaluaremos qué tan pequeño es el mínimo encontrado, con este fin evaluamos Λ en el punto $(\hat{\beta}_0, \hat{\beta}_1)$

$$\Lambda(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

(1) En todo caso, para la existencia de soluciones sólo se asume $\mathbb{P}(X=c) \neq 1$.

$$\Delta(\hat\beta_0, \hat\beta_1) = \sum_{i=1}^{N}\left\{y_i - \left[(\bar y - \hat\beta_1 \bar x) + \hat\beta_1 x_i\right]\right\}^2$$

$$= \sum_{i=1}^{N}\left\{(y_i - \bar y) - \hat\beta_1 (x_i - \bar x)\right\}^2$$

$$= \sum_{i=1}^{N}\left[(y_i-\bar y)^2 - 2\hat\beta_1 (y_i-\bar y)(x_i-\bar x) + \hat\beta_1^2 (x_i-\bar x)^2\right]$$

$$= \sum_{i=1}^{N}(y_i-\bar y)^2 - 2\hat\beta_1 \sum_{i=1}^{N}(y_i-\bar y)(x_i-\bar x) + \hat\beta_1^2 \sum_{i=1}^{N}(x_i-\bar x)^2$$

Pero $\sum_{i=1}^{N}(x_i-\bar x)(y_i-\bar y) = \hat\beta_1 \sum_{i=1}^{N}(x_i-\bar x)(x_i-\bar x)$, luego

$$\Delta(\hat\beta_0, \hat\beta_1) = \sum_{i=1}^{N}(y_i-\bar y)^2 - 2\hat\beta_1^2 \sum_{i=1}^{N}(x_i-\bar x)(x_i-\bar x) + \hat\beta_1^2 \sum_{i=1}^{N}(x_i-\bar x)^2$$

$$= \sum_{i=1}^{N}(y_i-\bar y)^2 - \hat\beta_1^2 \sum_{i=1}^{N}(x_i-\bar x)^2. \quad \cdots\cdots (\gamma')$$

Notando que $0 \le \Delta(\hat\beta_0, \hat\beta_1)$ tenemos

$$0 \le \hat\beta_1^2 \sum_{i=1}^{N}(x_i-\bar x)^2 \le \sum_{i=1}^{N}(y_i-\bar y)^2,$$

de donde

$$0 \le \Delta(\hat\beta_0, \hat\beta_1) \le \sum_{i=1}^{N}(y_i-\bar y)^2.$$

Esta desigualdad nos sugiere un método para re-escalar o estandarizar $\Delta(\hat\beta_0, \hat\beta_1)$, de manera que el resultado sea una cantidad que no dependa de las unidades de medición en la variable de respuesta Y y cuya magnitud se puede evaluar con mayor facilidad, la estandarización está dada por

$$D(\hat{\beta}_0,\hat{\beta}_1) = \frac{\Delta(\hat{\beta}_0,\hat{\beta}_1)}{\sum_{i=1}^{N}(y_i-\bar{y})^2},$$

lo cual supone que $\sum_{i=1}^{N}(y_i-\bar{y})^2 > 0$, pero está no resulta una restricción de importancia, como ya discutimos antes para la variable explicativa X, $\sum_{i=1}^{N}(y_i-\bar{y})^2=0$ sucedería sólo cuando $y_1 = y_2 = \cdots = y_N$; que es una situación que <u>núnca</u> nos interesaría modelar.

Asi entonces,

$$0 \leq D(\hat{\beta}_0,\hat{\beta}_1) \leq 1,$$

si $D=0$, entonces $\Delta(\hat{\beta}_0,\hat{\beta}_1)=0$ que se interpretaría como que la recta se ajusta perfectamente al conjunto de datos, de hecho sólo sucedería cuando todas las parejas (x_i,y_i) yacen sobre la línea recta determinada por mínimos cuadrados, es decir, cuando $y_i=\hat{\beta}_0+\hat{\beta}_1 x_i$ $\forall\, i=1,2,\ldots,N$.

En el extremo opuesto, $D=1$ corresponde al caso en que $\Delta(\hat{\beta}_0,\hat{\beta}_1) = \sum_{i=1}^{n}(y_i-\bar{y})^2$ y este debe ser el peor escenario, el valor $\Delta(\hat{\beta}_0,\hat{\beta}_1)$ es un mínimo[1], pero es el mínimo más grande que pudimos obtener. De la relación (γ^*) en la página 42 se sigue que si $\Delta(\hat{\beta}_0,\hat{\beta}_1) = \sum_{i=1}^{n}(y_i-\bar{y})^2$ entonces

$$\hat{\beta}_1^2 \cdot \sum_{i=1}^{N}(x_i-\bar{x})^2 = 0$$

y como la situación en que $x_1=x_2=\cdots=x_N$ está descartada, esto implicaría que $\hat{\beta}_1=0$ de forma que la recta de mínimos cuadrados es una constante lo cual sugiere que no existe una relación de asociación entre X y Y.

En resumen tenemos que $0 \leq D(\hat{\beta}_0,\hat{\beta}_1) \leq 1$ y en la peor situación

$D(\hat{\beta}_0,\hat{\beta}_1) = 1$ ($\hat{\beta}_1=0$ y no hay evidencia de asociación lineal)

(1) Se puede probar que $\Delta(\beta_0,\beta_1)$ tiene un mínimo en $\beta_0=\hat{\beta}_0$ y $\beta_1=\hat{\beta}_1$, APENDICE de Calculo Diferencial al final →

En la mejor situación

$$D(\hat{\beta}_0, \hat{\beta}_1) = 0 \quad \text{(ajuste perfecto)}$$

Podemos entonces usar a D como un índice que nos dice qué tan malo es el modelo (la línea recta) para describir la relación entre X y Y.

Alternativamente se puede definir un índice (de bondad de ajuste) dado por $B(\hat{\beta}_0, \hat{\beta}_1) = 1 - D(\hat{\beta}_0, \hat{\beta}_1)$

para el cual

$$0 \leq B(\hat{\beta}_0, \hat{\beta}_1) \leq 1.$$

En este caso $B(\hat{\beta}_0, \hat{\beta}_1) = 0 \Rightarrow$ no hay evidencia de asociación lineal entre X y Y ($\hat{\beta}_1 = 0$).

y $B(\hat{\beta}_0, \hat{\beta}_1) = 1 \Rightarrow$ ajuste perfecto.

La siguiente relación respecto a $B(\hat{\beta}_0, \hat{\beta}_1)$ es interesante:

$$B(\hat{\beta}_0, \hat{\beta}_1) = 1 - D(\hat{\beta}_0, \hat{\beta}_1)$$
$$= 1 - \frac{\Delta(\hat{\beta}_0, \hat{\beta}_1)}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

$$B(\hat{\beta}_0,\hat{\beta}_1) = 1 - \frac{\sum_{i=1}^{N}(y_i-\bar{y})^2 - \hat{\beta}_1^2\sum_{i=1}^{N}(x_i-\bar{x})^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2}$$

$$= \hat{\beta}_1^2 \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2},$$

pero $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$, así que

$$B(\hat{\beta}_0,\hat{\beta}_1) = \frac{\left(\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})\right)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2 \cdot \sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$= \left(\frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}}\right)^2$$

$$= \left(\frac{\frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{N}}{\sqrt{\frac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{N}}\sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}}}\right)^2$$

$$= r_{XY}^2,$$

donde $r_{XY} = \dfrac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$ es la correlación muestral entre $x_1,\ldots,x_N$ y $y_1,\ldots,y_N$.