Independencia

Es de Cola derecha, siempre se mete con $>$ en la app, para χ^2_{obs} y valor P

El H_0 y H_1 siempre son los mismos, solo cambia el contexto

H_0: X, Y son independientes

H_1: X, Y NO son independientes

Con la tabla que nos dan (observados), Calcular los Totales de X (T_X) y los Totales de Y (T_Y) con:

$$e_{ij} = \frac{T_{X_i} \cdot T_{Y_j}}{n}$$

, T_X = Total de la Fila
T_Y = Total de la columna
n = Todos los T_X + Todos los T_Y

Hay que armar la tabla de esperados e_{ij} que sera nuestro "e_{ij}" y la tabla de observados sera nuestro "o_{ij}" en la formula

Para probar independencia cuando cada $e_{ij} \geq 5$

Estadístico de prueba: $$\chi^2 = \sum_{i=1}^{f} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \longrightarrow \chi^2_{obs}$$
con $\nu = (f-1)(c-1)$, si $\nu \geq 2$.

V = Filas -1 · Columnas -1

Cola derecha

Si se cumple
$V_{obs} < V_C$, NO se rechaza H_0

Si se cumple
$V_{obs} > V_C$, Se rechaza H_0

TODAS son de Cola derecha

Si se cumple
$P > \alpha$, NO se rechaza

$P < \alpha$, se rechaza

Ejemplo

Ejemplo

En una universidad se realiza un estudio para verificar si el tipo de trabajo (administrativo y docente) se relaciona con el grado de estrés (I, II y III) de los trabajadores. Para lo cual se elige una muestra aleatoria de 300 trabajadores y se clasifican en la tabla siguiente. **(5 puntos)**

	I	II	III
Administrativos	42	24	30
Docentes	54	78	72

Pruebe la hipótesis de que el tipo de trabajo afecta el grado de estrés del trabajador.

H_0: El trabajo y estres son independientes
H_1: El trabajo y estres no son independientes

Tabla inicial   A = Admistrativos   D = Docentes

	I	II	III	Total x
A	42	24	30	96
D	54	78	72	204
Total y	96	102	102	300 → n

Ojo!!! Tanto la tabla de observados como la de esperados en los Tx y Ty deben de dar igual

Calcular esperados $\frac{T_{x_i} \cdot T_{y_j}}{n}$, $n = 300$

$e_{11} = \frac{96 \cdot 96}{300} = 30{,}72$ $e_{21} = \frac{204 \cdot 96}{300} = 65{,}28$

$e_{12} = \frac{96 \cdot 102}{300} = 32{,}64$ $e_{22} = \frac{204 \cdot 102}{300} = 69{,}36$

$e_{13} = \frac{96 \cdot 102}{300} = 32{,}64$ $e_{23} = \frac{204 \cdot 102}{300} = 69{,}36$

Tabla de observados

	I	II	III	Total x
A	42	24	30	96
D	54	78	72	204
Total y	96	102	102	300

Tabla de esperados

	I	II	III	Total x
A	30,72	32,64	32,64	96
D	65,28	69,36	69,36	204
Total y	96	102	102	300

$$\chi^2_{obs} = \frac{(42-30{,}72)^2}{30{,}72} + \frac{(24-32{,}64)^2}{32{,}64} + \frac{(30-32{,}64)^2}{32{,}64} +$$

$$\frac{(54-65{,}28)^2}{65{,}28} + \frac{(78-69{,}36)^2}{69{,}36} + \frac{(72-69{,}36)^2}{69{,}36}$$

$\chi^2_{obs} = 9{,}7683$

$v = (2-1)\cdot(3-1) = 2$

$\alpha = 0{,}05$

$\chi^2_c = \chi^2_{0{,}05,\,2} = 5{,}99146$

$\boxed{\text{R/ Como } \chi^2_{obs} = 9{,}7683 > \chi^2_c = 5{,}99146 \text{ se rechaza Ho, NO son independientes}}$

Con valor P

$P(\chi^2 > 9{,}7683) = 0{,}007566 \qquad \alpha = 0{,}05$

$\boxed{\text{Como } P = 0{,}007566 < 0{,}05 \text{ se rechaza Ho, NO son independientes}}$

ANOVA

Es de Cola derecha, siempre se mete con > en la app, para χ^2_{obs} y valor P

El Ho y H1 siempre son los mismos, solo cambia el contexto

$H_0: u_1 = u_2 = u_3$ Todas iguales
$H_1: \exists\, i, j \mid u_i \neq u_j$ Al menos 2 diferentes

n_i = Cantidad de filas en la columna i, $i = 1, 2, \ldots K$

T_i = Suma de la columna i, $i = 1, 2, \ldots K$

$$\frac{T_i^2}{n_i} = \frac{(\text{Cantidad de filas en la columna i})^2}{\text{Suma de la columna i}}$$

y^2_{ij} = Suma de todos los elementos de todas las columnas, CADA UNO de ellos elevado a la 2

N = Suma de los n_i

T = Suma de los T_i

K = Cantidad de categorias

Para probar $H_0\colon \mu_1 = \cdots = \mu_k$

Fofs

Estadístico de prueba: $F = \dfrac{S_1^2}{S^2}$ con $(k-1, N-k)$ g.l.

donde

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2 = \underline{\sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}}$$

La suma aparte

$$\to \left(\sum_{i=1}^{k}\frac{T_i^2}{n_i}\right) - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = SST - SSA$$

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

Ejemplo

Ejemplo 5:

Se quiere determinar si la dosis de determinado tratamiento (Baja, Media, Alta) influye en el tiempo de sueño (en minutos) de los que la consumen.

Se plantea el problema de determinar si la tiempos de sueño varían o no, según la dosis del medicamento. ¿Puede afirmarse que los tiempos de sueño no varían según la dosis de medicamento?, use un nivel de significancia de 5%.

Para responder al problema planteado, se cuenta con los siguientes datos tomados de varios voluntarios expuestos a la aplicación de 3 dosis de medicamento (Alta, Media, Baja), registrando el número de minutos que duerme, una vez injerido la dosis:

DOSIS		
Baja	Media	Alta
67	96	74
69	98	24
72	130	15
79	65	33
		17

Sea u_1, u_2, u_3 los tiempos promedio de sueño en cada dosis (Baja, media, Alta)

H_0: $u_1 = u_2 = u_3$ Todas iguales

H_1: $\exists\, i,j \mid u_i \neq u_j$ Al menos 2 diferentes

Dosis

Baja	Media	Alta
67	96	79
69	98	29
72	130	15
79	65	33
		17

$T_3 = 79+29+15+33+17 = 163$

$n_3 = 5$

$n_1 = 4$ $\qquad$ $n_2 = 4$

$T_2 = 96+98+130+65 = 389$

$T_1 = 67+69+72+79 = 287$

n_i	4	4	5
T_i	287	389	163
$\frac{T_i^2}{n_i}$	$\frac{287^2}{4} = 20592{,}25$	$\frac{389^2}{4} = 37830{,}25$	$\frac{163^2}{5} = 5313{,}8$

$N = 4+4+5 = 13$

$T = 287+389+163 = 839$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$$

$$SSA = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - \frac{T^2}{N}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 = SST - SSA$$

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$$y_{ij}^2 = 67^2 + 69^2 + 72^2 + 79^2 + 96^2 + 98^2 + 130^2 + 65^2 + 79^2 + 29^2 + 15^2 + 33^2 + 17^2 = 68275$$

$$SST = y_{ij}^2 - \frac{T^2}{N} = 68275 - \frac{839^2}{13} = 19127{,}2308$$

$$SSA = \sum \frac{T_i^2}{n_i} - \frac{T^2}{N} = (20592{,}25 + 37830{,}25 + 5313{,}8) - \frac{839^2}{13}$$

$$SSA = 9588{,}53077$$

$$SSE = SST - SSA = 19127{,}2308 - 9588{,}53077$$

$$SSE = 9538{,}7$$

Estadístico de prueba: $F = \frac{S_1^2}{S^2}$ con $(k-1, N-k)$ g.l.

$$s_1^2 = \frac{SSA}{k-1} \text{ y } s^2 = \frac{SSE}{N-k}$$

$SSA = 9588{,}53077$
$SSE = 4538{,}7$
$N = 13$
$K = 3$

$S_1^2 = \frac{9588{,}53077}{3-1}$ $\quad S^2 = \frac{4538{,}7}{13-3}$

$$F_{obs} = \frac{S_1^2}{S^2} = \frac{\frac{9588{,}53077}{3-1}}{\frac{4538{,}7}{13-3}} = 10{,}5631$$

$V = K-1, N-K$
$V = 3-1, 13-3$
$V = 2, 10$

$F_c = f_{0.05,2,10} = 4{,}10282$

$$\boxed{\text{R/ Como } f_{obs} = 10{,}5631 > f_c = 4{,}10282 \text{ Se rechaza } H_0}$$

Valor P

$P(F > 10{,}5631) = 0{,}00342$ $\quad \alpha = 0{,}05$

$$\boxed{\text{R/ Como } P = 0{,}00342 < \alpha = 0{,}05 \text{ Se rechaza } H_0}$$

Regresion lineal Simple

Se observan las edades y las estaturas de un grupo de niños como se muestra en la siguiente tabla:

x	Edad	3,3	4,3	5,2	5,6	6,3	7,0	8,1	8,6	9,3	10,1
y	Estatura (cm)	94	112	115	123	117	117	125	139	134	140

$n=10$

Proponga un modelo de regresión lineal simple para los datos de la tabla anterior y conteste:

En la calcu
menu → 6 → 2 → meter datos → AC →
optn → Summation

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$a = \frac{\sum y - b\sum x}{n} = \bar{y} - b\bar{x}$$

Y ahi calculamos las sumas

$\sum xy = 8502{,}8 \quad \sum x = 67{,}8 \quad \sum y = 1216 \quad \sum x^2 = 509{,}54$

$n=10$

$$b = \frac{10 \cdot 8502{,}8 - 67{,}8 \cdot 1216}{10 \cdot 509{,}54 - 67{,}8^2} = 5{,}76$$

$$a = \frac{1216 - 5{,}76 \cdot 67{,}8}{10} = 82{,}55$$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$a = \frac{\sum y - b\sum x}{n} = \bar{y} - b\bar{x}$$

$\hat{y} = 82{,}55 + 5{,}76x$ → modelo de RLS

$x=5$

a. ¿Cuál es la estatura esperada para un niño de 5 años?

$\hat{y} = 82{,}55 + 5{,}76 \cdot 5 \approx \boxed{111{,}35 \text{ cm}}$

$\hat{y} = 125$

b. ¿Cuál es la edad estimada de un niño que mide 125cm de estatura?

$125 = 82{,}55 + 5{,}76x \rightarrow x = \dfrac{125 - 82{,}55}{5{,}76} \approx \boxed{7{,}37}$

$x = 25$

c. ¿Cuál sería la estatura esperada para un niño a los 25 años?

$\hat{y} = 82{,}55 + 5{,}76 \cdot 25 \approx \boxed{226{,}55 \text{ cm}}$

Ejemplo 88 *Un estudiante de computación, que tiene una pequeña empresa de elaboración de software, llevó a cabo un estudio para determinar la relación entre el tiempo que tarda elaborando un software en días* (X) *y el ingreso obtenido por su venta en dólares* (Y). *Se recolectaron los valores de estas variables para 14 software, obteniendo los siguientes datos:*

X:	2	3	5	9	10	16	19	20	24	27	32	41	55	60
Y:	100	200	250	400	500	850	930	900	1300	1360	1500	2050	2800	2900

Interpretación de los coeficientes

a : valor promedio esperado cuando X es cero.

b : razón de cambio del valor promedio de Y por cada unidad adicional en X.

$$\hat{Y} = 8{,}52 + 49{,}29X$$

2. *¿Aproximadamente, al elaborar un software, cuánto es el valor promedio del ingreso fijo (ingreso que no depende del tiempo de elaboración)?*

$\boxed{8{,}52}$

3. *¿Aproximadamente, al elaborar un software, cuánto es el aumento promedio del ingreso por día de elabolación?*

$\boxed{49{,}29}$

Metodo de minimos cuadrados

Cuando hay que estimar β

1) Se iguala todo a 0, ejemplo

$$Y = a + \beta x \rightarrow Y - a - \beta x = 0$$

$$Y = \beta + \beta x \rightarrow Y - \beta - \beta x = 0$$

2) Se pone todo a la 2 dentro de una suma

$$\sum_{i=1}^{n} (Y - a - \beta x)^2 = 0$$

, n = Cantidad de filas de la tabla que nos den

3) Se deriva con $[(f(x))^n]' = n[f(x)]^{n-1} f'(x)$

$$\sum_{i=1}^{n} 2(y-a-\beta x)\cdot(\cancel{y}-\cancel{a}-\beta x)' = 0$$

A dividir se cancela

$\hookrightarrow \beta x' = x$

$$\sum_{i=1}^{n} (y-a-\beta x)\cdot -x = 0$$

$$\sum_{i=1}^{n} (xy - ax - \beta x^2) = 0$$

4) Se reparte la suma

$$\sum_{i=1}^{n} xy - a\sum_{i=1}^{n} x - \beta \sum_{i=1}^{n} x^2 = 0$$

4) Se despeja β

$$-\beta \sum x^2 = -\sum xy + a \sum x$$

$$\beta = \frac{\sum xy - a \sum x}{\sum x^2}$$

En la calcu
menu -> 6 -> 2 -> meter datos
-> AC -> optn -> Summation
Y ahi calculamos las
sumas

5) Dar la respuesta

$$y = a + \beta x$$

$\hookrightarrow$ El a original

Ejemplo

Ejemplo 89 *Considere los datos de la siguiente tabla:*

X:	2	3	5	9	13
Y:	8	10	18	30	40

$n = 5$

A partir de estos datos, estime el coeficiente β de la ecuación de regresión $\mu_{Y|x} = 2 + \beta x$ utilizando el método de mínimos cuadrados.

$$y = 2 + \beta x \longrightarrow y - 2 - \beta x = 0$$ Igualando a 0

($2 \to a$, $\beta \to \beta$)

$$\sum_{i=1}^{5} (y - 2 - \beta x)^2$$ Poniendo en suma y elevado a la 2

$$\sum_{i=1}^{5} 2(y - 2 - \beta x) \cdot -x = 0$$ Derivando

$$\sum_{i=1}^{5} (y - 2 - \beta x) \cdot x$$ Eliminando el -2 pasandolo a dividir

$$\sum_{i=1}^{5} (xy - 2x - \beta x^2)$$ Repartiendo el x

$$-\sum_{i=1}^{5} xy + 2\sum_{i=1}^{5} x + \beta \sum_{i=1}^{5} x^2 = 0$$ Repartiendo la suma

$$\beta = \frac{\sum xy - 2\sum x}{\sum x^2}$$ Despejando β

$$\beta = \frac{926 - 2 \cdot 32}{288} = 2{,}99$$

$\sum xy = 926$
$\sum x = 32$
$\sum x^2 = 288$

Calcu

$$\boxed{y = 2 + 2{,}99x}$$ Ecuacion final

Ejemplo extra

Ejercicio 44 *Considere los datos de la siguiente tabla:*

$x:$	1	5	10	20
$y:$	4	12	21	43

A partir de estos datos, estime el coeficiente β de la ecuación de regresión $\mu_{Yx} = \beta + \beta x$ utilizando el método de mínimos cuadrados. $R/\quad b = \frac{567}{281}$

$$\sum (y - \beta - \beta x)^2$$

$$\sum 2 (y - \beta - \beta x)(-1 - x) = 0$$

$$\sum (y - \beta - \beta x)(-1 - x) = 0$$

$$\sum (-y + \beta + \beta x - xy + \beta x + \beta x^2) = 0$$

$$\sum (-y + \beta + 2\beta x + \beta x^2 - xy) = 0$$

$$\sum (-y + \beta + 2\beta x + \beta x^2 - xy) = 0$$

$$-\sum y + \beta \sum 1 + 2\beta \sum x + \beta \sum x^2 - \sum xy = 0$$

$$\beta \sum 1 + 2\beta \sum x + \beta \sum x^2 = \sum xy + \sum y$$

$$\beta \left(\sum 1 + 2 \sum x + \sum x^2\right) = \sum xy + \sum y$$

$$\beta = \frac{\sum xy + \sum y}{\sum 1 + 2\sum x + \sum x^2}$$

$\sum xy = 1134$

$\sum y = 80$

$\sum 1 = 1 \cdot 4 = 4$

$\sum x = 36$

$\sum x^2 = 526$

$$\beta = \frac{1134 + 80}{4 + 2 \cdot 36 + 526}$$

$$\beta = 2{,}0166$$

$$\boxed{y = 2{,}0166 + 2{,}0166x}$$

IC Para RLS

Ejemplo 90 *Un estudiante de computación, que tiene una pequeña empresa de elaboración de software, llevó a cabo un estudio para determinar la relación entre el tiempo que tarda elaborando un software en días (X) y el ingreso obtenido por su venta en dólares (Y). Se recolectaron los valores de estas variables para 14 software, obteniendo los siguientes datos:*

$$\sum x = 323 \qquad \sum y = 16040 \qquad \sum x^2 = 11871$$
$$\sum y^2 = 29162000 \qquad \sum xy = 587890$$

Asuma las hipótesis de regresión. En un ejemplo anterior se determinó que la ecuación de regresión lineal para el ingreso como función del tiempo de elaboración de un software es

$$\widehat{y} = 49.2935x + 8.44282$$

1. *Determine un intervalo de confianza del 90% para el valor promedio del ingreso fijo (ingreso que no depende del tiempo de elaboración)* $\rightarrow a$

$a = 8{,}44282 \quad b = 49{,}2935 \quad n = 14$

$$a \pm t_{\delta/2,\nu}\, s\sqrt{\frac{\sum x^2}{nS_{xx}}} \text{ con } \nu = n-2$$

$$S_{xx} = \sum(x_i-\bar{x})^2 = (n-1)s_x^2 = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$S_{yy} = \sum(y_i-\bar{y})^2 = (n-1)s_y^2 = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$S_{xy} = \sum(x_i-\bar{x})(y_i-\bar{y}) = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^n (y_i-\hat{y}_i)^2}{n-2}} = \sqrt{\frac{S_{yy}-bS_{xy}}{n-2}}$$

$$S_{xx} = 11871 - \frac{323^2}{14} = \frac{61865}{14}$$

$$S_{yy} = 29162000 - \frac{16040^2}{14} = 10784742{,}86$$

$$S_{xy} = 587890 - \frac{323 \cdot 16040}{14} = \frac{1524770}{7}$$

$$s = \sqrt{\frac{10784742{,}86 - 49{,}2935 \cdot \frac{1524770}{7}}{14-2}} = 62{,}86323259$$

$$\alpha = 0{,}10 \rightarrow \frac{\alpha}{2} = 0{,}05 \rightarrow t_{0{,}05,12} = \pm 1{,}78229 \quad V = 12$$

$$\text{Lim inf} = 8{,}44282 - 1{,}78229 \cdot 62{,}86323259\sqrt{\frac{11871}{14 \cdot \frac{61865}{14}}}$$
$$= -40{,}63623863$$

$$\text{Lim sup} = 8{,}44282 + 1{,}78229 \cdot 62{,}86323259\sqrt{\frac{11871}{14 \cdot \frac{61865}{14}}}$$
$$= 57{,}52187863$$

$$\boxed{\text{IC }]-40.63623863,\ 57.52187863[}$$

2. *Determine un intervalo de confianza del 90% para el $\underbrace{\text{aumento promedio del ingreso por día de}}_{b}$ elabolación*

$b = 49{,}2935$

IC para β: $b \pm t_{\delta/2,\nu} s \sqrt{\frac{1}{S_{xx}}}$ con $\nu = n - 2$

$t_{0.05,12} = \pm 1{,}78229$ $\qquad S_{xx} = \frac{61865}{14}$

$s = 62{,}86323259$

$$49{,}2935 - 1{,}78229 \cdot 62{,}86323259 \cdot \sqrt{\frac{1}{\frac{61865}{14}}} = 47{,}60804$$

$$49{,}2935 + 1{,}78229 \cdot 62{,}86323259 \cdot \sqrt{\frac{1}{\frac{61865}{14}}} = 50{,}97895$$

$\boxed{]47.60804, 50.97895[}$

Determine el IC de 90% para el ingreso esperado por venta de software con un tiempo de elaboración de 30 días

$y = 49{,}2935 \cdot 30 + 8{,}44282$ $\qquad n = 14$

$y = 1487{,}24782$ $\qquad \nu = 12$

$\hat{y} \pm t_{\delta/2,\nu} s \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$

con $\nu = n - 2$ y $\hat{y} = a + bx$

$t_{0.05,12} = \pm 1{,}78229$ $\qquad s = 62{,}86323259$

$S_{xx} = \frac{61865}{14}$ $\qquad x = 30$

$\bar{x} = \frac{\Sigma x}{n} = \frac{323}{14}$

Lim inf:

$$1487{,}24782 - 1{,}78229 \cdot \sqrt{\frac{1}{14} + \frac{\left(30 - \frac{323}{14}\right)^2}{\frac{61865}{14}}}$$

$= 1486{,}736542$

Lim Sup:

$$1487,24782-1,78229\cdot 62,86323259\cdot\sqrt{\frac{1}{14}+\frac{\left(30-\frac{323}{14}\right)^2}{\frac{61865}{14}}}$$

$= 1519,38843$

$$\boxed{\left]1486.736542,\ 1519.38843\right[}$$

Determine un intervalo de predicción de 90% para el ingreso por venta de software con un tiempo de elaboración de 30 días

$y = 49,2935\cdot 30+8,44282$ $\quad n=14$

$y = 1487,24782$ $\quad v=12$

Intervalo de predicción para $Y|x$ cuando se cumplen las hipótesis de regresión

Extremos del intervalo: $\hat{y}\pm t_{\delta/2,\nu}s\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}$

con $\nu = n-2$ y $\hat{y}=a+bx$

$t_{0.05,12}=\pm 1,78229$ $\quad s = 62,86323259$

$S_{xx}=\frac{61865}{14}$ $\quad x=30$

$\bar{x}=\frac{\Sigma x}{n}=\frac{323}{14}$

Lim inf:

$$1487,24782-1,78229\cdot 62,86323259\cdot\sqrt{1+\frac{1}{14}+\frac{\left(30-\frac{323}{14}\right)^2}{\frac{61865}{14}}}$$

$= 1370,688415$

$$1487,24782+1,78229\cdot 62,86323259\cdot\sqrt{1+\frac{1}{14}+\frac{\left(30-\frac{323}{14}\right)^2}{\frac{61865}{14}}}$$

$= 1603,807225$

$$\boxed{\left]1370.688415,\ 1603.807225\right[}$$

PH Para RLS

Ejemplo 90 *Un estudiante de computación, que tiene una pequeña empresa de elaboración de software, llevó a cabo un estudio para determinar la relación entre el tiempo que tarda elaborando un software en días (X) y el ingreso obtenido por su venta en dólares (Y). Se recolectaron los valores de estas variables para 14 software, obteniendo los siguientes datos:*

$$\sum x = 323 \quad \sum y^2 = 29162000 \quad \sum y = 16040 \quad \sum xy = 587890 \quad \sum x^2 = 11871$$

Asuma las hipótesis de regresión. En un ejemplo anterior se determinó que la ecuación de regresión lineal para el ingreso como función del tiempo de elaboración de un software es

$$\hat{y} = 49.2935x + 8.44282$$

Con un nivel de signicancia del 10%, ¿Hay evidencia de que el aumento promedio del ingreso por día de elabolación sea de 50 dólares por día?

$H_0: \beta = 50$ $\quad b = 49{,}2935$ $\quad S_{xx} = \frac{61865}{14}$ $\quad n = 14$

$H_1: \beta \neq 50$ $\quad \beta_0 = 50$ $\quad v = 12$

$S = 62{,}86323259$ $\quad T = (B - \beta_0)\frac{\sqrt{S_{xx}}}{S}$ con $\nu = n - 2$

$$t_{0,05} = \frac{(49{,}2935 - 50) \cdot \sqrt{\frac{61865}{14}}}{62{,}86323259} = -0{,}74709$$

$$\alpha = 0{,}1 \rightarrow \frac{\alpha}{2} = 0{,}05 \rightarrow t_{0.05,12} = \pm 1{,}78229$$

$\boxed{\text{R/ } -1{,}78229 < -0{,}74709 < 1{,}78229, \text{ No se rechaza } H_0}$

Con valor P

$$2P(t > -0{,}74709) = 0{,}46940 \qquad \alpha = 0{,}10$$

$\boxed{\text{R/ Como } P = 0{,}46940 > \alpha = 0{,}10, \text{ No se rechaza } H_0}$

- Una fabrica recolectó la siguiente información de 8 de sus trabajadores sobre el número de minutos que llegaron tarde al trabajo (X) y el número de piezas defectuosas que fabricaron ese día (Y):

X	2	5	10	15	20	25	30	40
Y	2	4	9	12	15	20	24	30

Suponiendo que se cumplen las hipótesis de regresión.

- A un nivel de significancia del 5%, ¿existe evidencia de que el aumento promedio en el número de piezas defectuosas que realiza un empleado por cada minuto que llega tarde al trabajo es mayor a 0.7?

$H_0: \beta = 0,7 \ (\leq)$

$H_1: \beta > 0,7$

$n = 8$

$b = 0,749230606$

$\Sigma x^2 = 3879$

$\Sigma x = 147$

$\Sigma xy = 3014$

$\Sigma y^2 = 2346$

$\Sigma y = 116$

$\alpha = 0,05$

$v = 6$

$\beta_0 = 0,7$

Como es 1 cola se usa directa sin dividir

$$T = (B - \beta_0)\frac{\sqrt{S_{xx}}}{S} \text{ con } \nu = n - 2$$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

$$S_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

$$s = \sqrt{\frac{S_{yy} - bS_{xy}}{n-2}}$$

$$S_{xx} = 3879 - \frac{147^2}{8} = 1177,875$$

$$S_{yy} = 2346 - \frac{116^2}{8} = 664$$

$$S_{xy} = 3014 - \frac{147 \cdot 116}{8} = 882,5$$

$$S = \sqrt{\frac{664 - 0,749230606 \cdot 882,5}{8-2}} = 0,68361663$$

$$t_{obs} = (0,749230606 - 0,7) \cdot \frac{\sqrt{1177,875}}{0,68361663} = 2,47157$$

$$t_c = t_{0.05,6} = 1,94318$$

R/ Como $t_{obs} = 2,47157 > t_c = 1,94318$ se rechaza H_0

Con valor P

$P(t > 2,47757) = 0,02438$ $\quad \alpha = 0,05$

$\boxed{\text{R/ Como } P = 0,02438 < 0,05 \text{ se rechaza } H_0}$

Correlacion y determinacion

Interpretación de los coeficientes

X **a : valor promedio esperado cuando X es cero.**

Y **b : razón de cambio del valor promedio de Y por cada unidad adicional en X.**

Algunas fórmulas útiles

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}.$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}.$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

r^2 = coeficiente de determinacion

r^2 = proporcion de variacion que depende de x

$1-r^2$ = proporcion de variacion que NO depende de x

$$r^2 = b\frac{s_x}{s_y} = \left(b\sqrt{\frac{S_{xx}}{S_{yy}}}\right)^2$$

r = coeficiente de correlacion

$$r = b\frac{s_x}{s_y} = b\sqrt{\frac{S_{xx}}{S_{yy}}}$$

Ejemplo

Una cooperativa de ahorro ha sufrido un cierre técnico debido a una deficiente administración. Esta entidad financiera estima que el tiempo (en días) para la devolución de dinero depende linealmente de la cantidad de dinero (millones de colones) que el cliente tiene en su cuenta. Los siguientes datos se refieren a información de 10 clientes sobre el dinero en su cuenta (x) y el tiempo que han esperado para su devolución (y).

$$\sum_{i=1}^{10} x_i^2 = 433 \qquad \sum_{i=1}^{10} x_i = 51 \qquad \sum_{i=1}^{10} y_i^2 = 1545 \qquad \sum_{i=1}^{10} y_i = 95 \qquad \sum_{i=1}^{10} x_i y_i = 808$$

$\to r^2$

b) ¿Qué porcentaje de la variación en los días de espera para la devolución del dinero depende de factores diferentes a la cantidad de dinero en la cuenta del cliente? **(3 puntos)**

No dependiendo de x, usar $1-r^2$

$$S_{xx} = 433 - \frac{51^2}{10} = 172,9$$

Algunas fórmulas útiles

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}.$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}.$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$S_{yy} = 1545 - \frac{95^2}{10} = 642,5$$

$$r^2 = b\frac{s_x}{s_y} = \left(b\sqrt{\frac{S_{xx}}{S_{yy}}}\right)^2$$

$$1-r^2 = 1 - \left(1,87 \cdot \sqrt{\frac{172,9}{642,5}}\right)^2 = 0,059 \approx 0,6$$

$\boxed{\text{Al rededor del 6\% de variacion}}$

Regresion NO lineal simple

Hay 4 modelos, lo relevante es conocer el X_1, Y_1, α y β

Modelo exponencial

$$\ln(x)^y = y\ln(x)$$
$$\ln(x \cdot y) = \ln(x) + \ln(y)$$

$$Y = \alpha \beta^X$$
$$\ln(Y) = \ln(\alpha) + X\ln(\beta)$$ Se debe Aplicar ln

Entonces $Y_1 = \ln(Y)$ $X_1 = X$ $\alpha = \ln(\alpha)$ $\beta = \ln(\beta)$, con esto se debe armar una tabla nueva a partir de la vieja

X	Y	X_1	Y_1
w	z	$\ln(w)$	$\ln(z)$

y asi con todos los valores

w, z son numeros

Si bien la calculadora da $Y = \alpha + \beta X$, se debe transformar

Al $\alpha \neq \alpha$ $(\alpha = \ln(\alpha))$ y $\beta \neq \beta$ $(\beta = \ln(\beta))$, se debe hacer un ajuste, despejando el $\ln(\alpha)$ y $\ln(\beta)$, usando e

$$\ln(\alpha) = \alpha \qquad \ln(\beta) = \beta$$
$$e^{\ln(\alpha)} = e^{\alpha} \qquad e^{\ln(\beta)} = e^{\beta}$$
$$\alpha = e^{\alpha} \qquad \beta = e^{\beta}$$

Y al final se da la respuesta con el modelo original

$$Y = \alpha \beta^X$$

Suelen decir cual modelo usar

Ejemplo

Un sector del Estadio Nacional de Costa Rica dispone de 980 entradas para los aficionados. Para el partido de despedida de la Selección Nacional se midió el ritmo de ventas de estas entradas en el sitio web. Con la infomación se construyó la tabla siguiente en la que se tienen las horas transcurridas desde que se inició la venta de entradas y la cantidad de entradas disponibles hasta ese momento.

Horas transcurridas	0.5	1	2	3	4	4.5	5	6	7	8
Entradas Disponibles	975	850	760	550	385	250	230	185	150	130

Determine la ecuación de regresión exponencial ($y = \alpha\beta^x$) para estimar la cantidad de entradas disponibles a partir del tiempo de ventas transcurrido. **(4 puntos)**

$$y = \alpha \cdot \beta^x$$

$$\ln(y) = \ln(\alpha) + x\ln(\beta)$$

$y_1 = \ln(y)$ $\quad x_1 = x$ $\quad \alpha = \ln(\alpha)$ $\quad \beta = \ln(\alpha)$

x_1 (x)	y_1 $(\ln(y))$
0,5	$\ln(975)$
1	$\ln(850)$
2	$\ln(760)$
3	$\ln(550)$
4	$\ln(385)$
4,5	$\ln(250)$
5	$\ln(230)$
6	$\ln(185)$
7	$\ln(150)$
8	$\ln(130)$

Calcu

$$y = \underbrace{7{,}06228244}_{\alpha} - \underbrace{0{,}294635677}_{\beta} x$$

$\ln(\alpha) = \alpha$ $\qquad\qquad \ln(\beta) = \beta$

$e^{\ln(\alpha)} = e^{\alpha}$ $\qquad\qquad e^{\ln(\beta)} = e^{\beta}$

$\alpha = e^{7{,}06228244}$ $\qquad\qquad \beta = e^{-0{,}294635677}$

$\alpha = 1171{,}720246$ $\qquad\qquad \beta = 0{,}7448028866$

$$\boxed{y = 1171{,}720246 \cdot 0{,}7448028866^{x}}$$

Modelo potencial

$$y = \alpha \cdot x^{\beta}$$

$$\ln(y) = \ln(\alpha) + \beta \ln(x)$$ Se debe Aplicar ln

Entonces $y_I = \ln(y)$ $x_I = \ln(y)$ $\alpha = \ln(\alpha)$ $\beta = \beta$, con esto se debe armar una tabla nueva a partir de la vieja

Si bien la calculadora da $y = \alpha + \beta x$, se debe transformar

Al $\alpha \neq \alpha$ ($\alpha = \ln(\alpha)$) se debe hacer un ajuste, despejando el $\ln(\alpha)$

$$\ln(\alpha) = \alpha$$
$$e^{\ln(\alpha)} = e^{\alpha}$$
$$\alpha = e^{\alpha}$$

$\beta = \beta$, entonces se deja igual

Y al final se da la respuesta con el modelo original

$$y = \alpha \cdot x^{\beta}$$

Ejemplo

Ejemplo 98 *Seguidamente se muestra la distancia de frenado D (en pies) de un automóvil que viaja a la velocidad V (en millas por hora) a partir del instante en que se percibe el peligro.*

X	*Velocidad V (mi / h)*	20	30	40	50	60	70
Y	*Distancia de frenado D (pies)*	54	90	138	206	292	396

Use el modelo potencial para estimar la distancia de frenado dependiendo de la velocidad

$$y = \alpha \cdot x^{\beta}$$
$$\ln(y) = \ln(\alpha) + \beta \ln(x)$$

$y1 = \ln(y)$ $\quad x1 = x$ $\quad \alpha = \ln(\alpha)$ $\quad \beta = \beta$

$x1\,(x)$	$y1\,(\ln(y))$
$\ln(20)$	$\ln(54)$
$\ln(30)$	$\ln(90)$
$\ln(40)$	$\ln(138)$
$\ln(50)$	$\ln(206)$
$\ln(60)$	$\ln(292)$
$\ln(70)$	$\ln(396)$

Calcu

$$Y = -0{,}87022 8393 + 1{,}59455 6411\, X$$

$$\ln(a) = a$$
$$\ln(a) = -0{,}870228393$$
$$e^{\ln(a)} = e^{a}$$
$$a = e^{a}$$
$$a = e^{-0{,}870228393} = 0{,}41886$$

En este modelo hay que hacer este despeje para el a

El modelo sería $y = a \cdot x^{b}$, potencial

$$\boxed{Y = 0{,}41886 \cdot X^{1{,}594556411}}$$

El IC para cualquier modelo es en base a X_1 y Y_1 y son las mismas formulas

3. *Encuentre un intervalo de confianza de 90% para α y β*

Para α

X_1	Y_1
ln(20)	ln(54)
ln(30)	ln(90)
ln(40)	ln(138)
ln(50)	ln(206)
ln(60)	ln(292)
ln(70)	ln(396)

$$a \pm t_{\delta/2,\nu} s \sqrt{\frac{\sum x^2}{n S_{xx}}} \text{ con } \nu = n-2$$

$$S_{xx} = \sum x_1^2 - \frac{(\sum x_1)^2}{n} \quad S_{yy_1} = \sum y_1^2 - \frac{(\sum y_1)^2}{n}$$

$$S_{xy} = \sum xy_1 - \frac{(\sum x_1)(\sum y_1)}{n} \quad s = \sqrt{\frac{S_{yy_1} - bS_{xy_1}}{n-2}}$$

$Y = -0,870228393 + 1,594556411X$

$S_{xx} = 1,083410023$ $\quad n = 6 \quad V = 4$

$S_{yy} = 2,779363722$ $\quad a = -0,870228393$

$S_{xy} = 1,727558398$ $\quad \Sigma x^2 = 84,26768032$

$S = 0,0785372573$ $\quad \alpha = 0,10$

$\alpha = 0,05$

$$a \pm t_{\delta/2,\nu} s \sqrt{\frac{\sum x^2}{n S_{xx}}} \text{ con } \nu = n-2$$

$t_{0.05,4} = \pm 2,13185$

$$-0,870228393 \pm 2,13185 \cdot 0,0785372573 \cdot \sqrt{\frac{84,26768032}{6 \cdot 1,083410023}}$$

$]-1,473052773, -0,2674039134[$

Para el caso de alfa, al no ser este parámetro equivalente a Bo, se debe realizar la siguiente transformación para calcular su IC de 90% como sigue:

$a = e^{a}$, entonces

$]e^{-1,473052773}, e^{-0,2674039134}[$

$\boxed{\alpha \;]0.229225, 0.765363[}$

$$b \pm t_{\delta/2,\nu} s \sqrt{\frac{1}{S_{xx}}} \text{ con } \nu = n - 2 \qquad b = 1{,}594556411$$

$$1{,}594556411 \pm 2{,}13185 \cdot 0{,}07853725723 \cdot \sqrt{\frac{1}{1{,}083410023}}$$

$$\boxed{R \,|\,]1.43370091,\ 1.755411932[}$$

Modelo Reciproco

$Y = \alpha + \dfrac{\beta}{X}$, Aquí no se usa ln, solo se escribe así

$Y = \alpha + \beta \cdot \dfrac{1}{X}$

Entonces $Y1 = Y$ $\quad X1 = \frac{1}{X}$ $\quad \alpha = \alpha$ $\quad \beta = \beta$,
con esto se debe armar una tabla nueva
a partir de la vieja

$\alpha = \alpha \wedge \beta = \beta$,
NO ocupan
transformación

Y al final se da la
respuesta con el modelo
original

$Y = \alpha + \dfrac{\beta}{X}$

Sacado de chatgpt, lo importante es el procedimiento

Un laboratorio de ingeniería analiza el **tiempo promedio de respuesta** y (en segundos) de un servidor cuando se incrementa la **capacidad asignada de procesamiento** x (en unidades de cómputo).
Se observa que, a medida que la capacidad aumenta, el tiempo de respuesta disminuye de forma no lineal y tiende a estabilizarse, por lo que se propone un **modelo recíproco**.

x	2	3	5	8	12	25
y	27.3	20.1	14.6	11.4	9.2	6.9

$$y = \alpha + \frac{\beta}{x} \longrightarrow y = \alpha + \beta \cdot \frac{1}{x}$$

$$y_1 = y \qquad x_1 = \frac{1}{x} \qquad \alpha = \alpha \qquad \beta = \beta$$

$x_1\left(\frac{1}{x}\right)$	$y_1(y)$
$\frac{1}{2}$	27,3
$\frac{1}{3}$	20,1
$\frac{1}{5}$	14,6
$\frac{1}{8}$	11,4
$\frac{1}{12}$	9,2
$\frac{1}{25}$	6,9

$$y = \underbrace{5{,}583523359}_{\alpha} + \underbrace{43{,}69221837}_{\beta}\,x$$

$$y = \alpha + \frac{\beta}{x}$$

$$\boxed{y = 5{,}583523359 + \frac{43{,}69221837}{x}}$$

Modelo Hiperbolico

$$y = \frac{x}{\alpha x + \beta}$$, Aquí no se usa ln, solo se escribe así

$$\frac{1}{y} = \alpha + \beta \cdot \frac{1}{x}$$

Entonces $y1 = \frac{1}{y}$ $x1 = \frac{1}{x}$ $\alpha = \alpha$ $\beta = \beta$,
con esto se debe armar una tabla nueva a partir de la vieja

$\alpha = \alpha \wedge \beta = \beta$, NO ocupan transformacion

Y al final se da la respuesta con el modelo original

$$y = \frac{x}{\alpha x + \beta}$$

Ejemplo

Ejemplo 97 *Considere los datos en la siguiente tabla:*

X varia poco

Y varia mucho

x	18	19	19.5	19.7	19.9
Y	19	40	79	130	397

2. *Encuentre los coeficientes de la ecuación de regresión para el modelo* hiperbolico

$$Y = \frac{X}{\alpha x + \beta} \longrightarrow \frac{1}{Y} = \alpha + \beta \cdot \frac{1}{X}$$

$$Y1 = \frac{1}{Y} \quad X1 = \frac{1}{X} \quad \alpha = \alpha \quad \beta = \beta$$

$X_1\left(\frac{1}{X}\right)$	$Y_1\left(\frac{1}{Y}\right)$
$\frac{1}{18}$	$\frac{1}{19}$
$\frac{1}{19}$	$\frac{1}{40}$
$\frac{1}{19,5}$	$\frac{1}{79}$
$\frac{1}{19,7}$	$\frac{1}{130}$
$\frac{1}{19,9}$	$\frac{1}{397}$

Estos a la calcu para encontrar $y = a + bx$

$$Y = \underbrace{-0,46977455}_{\alpha} + \underbrace{9,403241079}_{\beta} X$$

El modelo seria

$$\boxed{Y = \frac{X}{\alpha x + \beta} = \frac{X}{-0,46977455X + 9,403241079}}$$

Regresion lineal multiple

La forma SIEMPRE es

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, entonces con esto se debe construir una matriz

β_0 = una columna de 1s, el resto son mas columnas,

Forma aumentada del sistema:

$$\underbrace{\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}}_{\text{Matriz asociada al sistema}} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Matriz asociada al sistema: X — Vector de incógnitas: Y — Vector solución: Y

Y se debe realizar la operacion

$$b = (X^T \cdot X)^{-1} \cdot (X^T \cdot Y)$$

Ejemplo 100 *Se tienen las siguientes observaciones*

x_1	0	1	1	0
x_2	1	0	−1	0
Y	1	4	6	3

1. *Estime la ecuación de regresión de Y como función lineal de x_1,x_2.*

X_1	X_2	Y
0	1	1
1	0	4
1	-1	6
0	0	3

$$\hat{Y} = b_0 + b_1 x_1 + b_2 \cdot x_2$$

Se debe agregar una columna de 1s al principio y luego los demas e igualar a y

$$\hat{Y} = b_0 + b_1 x_1 + b_2 \cdot x_2$$

$$X = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 4 \\ 6 \\ 3 \end{pmatrix} \qquad b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$$

$$b = (X^T \cdot X)^{-1} \cdot (X^T \cdot y) = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$$

$\boxed{\text{R/ El modelo de RLM es } \hat{Y} = 3 + 1 \cdot x_1 - 2 \cdot x_2}$

La calcu solo permite 4x4 max.

Regresion NO lineal multiple

La forma varia, pues hay que linealizar el modelo normalmente con ln y haciendo un cambio de variable

B_0 = una columna de 1s, el resto son mas columnas,

Forma aumentada del sistema:

$$\underbrace{\begin{pmatrix} \boxed{\begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix}} & \begin{matrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{matrix} \end{pmatrix}}_{} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Matriz asociada al sistema — X

Vector de incógnitas — Y

Vector solución — Y

Y se debe realizar la operacion

$$b = (X^T \cdot X)^{-1} \cdot (X^T \cdot Y)$$

Ejemplo 2: Dadas las siguientes observaciones estime los coeficientes de la ecuación

$$z = \beta_0 x^{\beta_1} y^{\beta_2}$$

- Determine z cuando $x = 2$ y $y = 7$

X	Y	Z
1	0,5	0,06
1,5	1	6,7
2	2	512
1	2	64

Linealizando el modelo

$$z = \beta_0 \cdot x^{\beta_1} \cdot y^{\beta_2}$$

$$\ln(z) = \ln(\beta_0 x^{\beta_1} \cdot y^{\beta_2})$$ Aplicando ln

$$\ln(z) = \ln(\beta_0 \cdot x^{\beta_1}) + \ln(y)^{\beta_2}$$

$$\ln(z) = \ln(\beta_0) + \beta_1 \ln(x) + \beta_2 \ln(y)$$

$z_1 = \ln(z)$ $\quad x_1 = \ln(x)$ Haciendo cambio

$\beta_0 = \ln(\beta_0)$ $\quad \beta_1 = \beta_1$ $\quad \beta_2 = \beta_2$ de variable

Modelo de estimacion para $z = \beta_0 \cdot x^{\beta_1} y^{\beta_2}$

$z_1 = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2$ Cambio de variable Aplicado

$$Z1 = b_0 + b_1 \cdot x1 + b_2 \cdot x2$$

Construyendo nueva tabla

$x1 = \ln(x)$	$x2 = \ln(y)$	$Z1 = \ln(z)$
$\ln(1)$	$\ln(0{,}5)$	$\ln(0{,}06)$
$\ln(1{,}5)$	$\ln(1)$	$\ln(6{,}7)$
$\ln(2)$	$\ln(2)$	$\ln(512)$
$\ln(1)$	$\ln(2)$	$\ln(64)$

X	Y	Z
1	0,5	0,06
1,5	1	6,7
2	2	512
1	2	64