GOI ESKOLA POLITEKNIKOA
ESCUELA POLITÉCNICA SUPERIOR
MONDRAGON UNIBERTSITATEA

ANÁLISIS DE DATOS

2. GRADO DE INGENIERÍA INFORMÁTICA - 2. SEMESTRE

Autores:

Aitor Landa
Ane Sajeras
Josu Garralda
Ibai Rodríguez
Oihane Lameirinhas

30 de mayo de 2019

MONDRAGON UNIBERTSITATEA

RESUMEN

En este trabajo se verifica si existe una relación lineal entre las variables *x* e *y*, dónde *x* recibirá el nombre de variable predictora o independiente mientras que la variable *y* será la variable dependiente o de respuesta. Para ello el objetivo es lograr a través de dicho modelo estimar el valor medio de la variable de respuesta para cada valor de la variable predictora o bien predecir el valor de dicha variable de respuesta para algún valor de la variable predictora. Para realizar dicho análisis se seguirán una serie de pasos recogidos a continuación.

ÍNDICE

MONDRAGON UNIBERTSITATEA

1. INTRODUCCIÓN

El análisis de regresión lineal es una técnica estadística utilizada para estudiar la relación entre variables, que se adapta a una amplia variedad de situaciones. En general interesa:

- Investigar si existe una asociación entre las dos variables testeando la hipótesis de independencia estadística.
- Estudiar la fuerza de la asociación, a través de una medida de asociación denominada coeficiente de correlación.
- Estudiar la forma de la relación. Usando los datos propondremos un modelo para la relación y a partir de ella será posible predecir el valor de una variable a partir de la otra.

Para ello propondremos un modelo que relaciona una variable dependiente (Y) con una variable independiente (X) donde debemos determinar cuál es el modelo teórico de dicha relación. En este caso, para poder comprobar que nuestra hipótesis cumple una regresión lineal hemos pensado en verificar que la cantidad de inmigrantes registrada en la ONG Afro corresponde al índice de pobreza de su país de origen. Para realizar el análisis de dichas variables seguiremos una serie de pasos que se recogen a continuación.

2. DEFINICIÓN Y RANGO POR ANALIZAR DE LAS VARIABLES X E Y

La función más simple para la relación entre dos variables es la función lineal **Y= a+bX** donde las variables utilizadas para la comprobación de dicha regresión lineal han sido las siguientes:

X = {Población bajo el nivel de pobreza (%) de un país de origen} [0-100]
Y = {N.º de personas de alta en la ONG por país de origen} [0-∞]

3. DIAGRAMA DE DISPERSIÓN XY

El diagrama de dispersión ofrece una idea bastante aproximada sobre el tipo de relación existente entre dos variables, utilizando las coordenadas cartesianas para mostrar los valores de éstas para un conjunto de datos. Además, representa la relación entre dos variables de forma gráfica y es muy útil para visualizar e interpretar los datos.

Mediante el diagrama de dispersión se puede estudiar la relación entre:

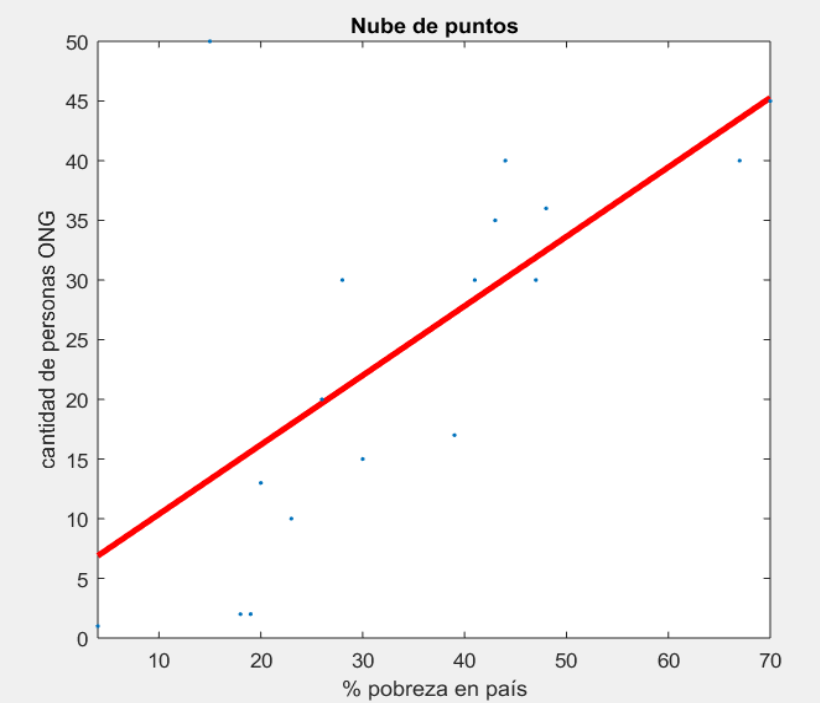
- Dos factores o causas relacionadas con la calidad
- Dos problemas de calidad
- Un problema de calidad y su posible causa

Tabla de datos:

PAÍS	ÍNDICE DE POBREZA (%)	PERSONAS DE ALTA EN LA ONG
Guinea-Bissau	67	40
Guinea Ecuatorial	44	40
Marruecos	15	50
Perú	23	10
Tanzania	23	10
Venezuela	20	13
Colombia	28	30

Angola	41	30
Argelia	23	10
Gambia	48	36
Kenia	43	35
Nigeria	70	45
Senegal	47	30
Ecuador	26	20
Pakistán	30	15
Camerún	30	15
Portugal	19	2
Bolivia	39	17
Polonia	18	2
Brasil	4	1

A continuación, se representa gráficamente la distribución mediante una nube de puntos o diagrama de dispersión. El gráfico de dispersión es un diagrama que se utiliza las coordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos.



Sample size: 17
Mean x (x): 34.235294117647
Mean y (y): 24.470588235294
Intercept (a): 4.5594929807823
Slope (b): 0.58159556585344
Regression line equation: y=4.5594929807823+0.58159556585344x

Mirando el resultado de la gráfica podemos observar que no todos los datos son lineales, ya que hay países donde el índice de pobreza es menor como por ejemplo Marruecos, pero por la situación geográfica, la cercanía, ley de memoria histórica etc. existen más inmigrantes de ese país en la ONG.

4. CÁLCULOS INTERMEDIOS

En la tabla inferior se recogen todos los datos utilizados para realizar las operaciones necesarias para el análisis de la regresión lineal.

X	Y	X^2	Y^2	X*Y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})*(y-\bar{y})$	y calculada	e(error)	e^2
4	1	16	1	4	-30,23529412	-23,47058824	914,1730104	550,8685121	709,6401384	6,885875244	-5,885875244	34,64352739
15	50	225	2500	1250	-19,23529412	25,52941176	369,9965398	651,7508651	-491,0657439	13,28342647	36,71657353	1348,106772
18	2	324	4	36	-16,23529412	-22,47058824	263,5847751	504,9273356	364,816609	15,02821317	-13,02821317	169,7343383
19	2	361	4	38	-15,23529412	-22,47058824	232,1141869	504,9273356	342,3460208	15,60980873	-13,60980873	185,2268937
20	13	400	169	260	-14,23529412	-11,47058824	202,6435986	131,5743945	163,2871972	16,1914043	-3,191404298	10,18506139
23	10	529	100	230	-11,23529412	-14,47058824	126,2318339	209,3979239	162,5813149	17,936191	-7,936190995	62,98312752
26	20	676	400	520	-8,235294118	-4,470588235	67,8200692	19,98615917	36,816609	19,68097769	0,319022307	0,1017752324
28	30	784	900	840	-6,235294118	5,529411765	38,87889273	30,57439446	-34,47750865	20,84416882	9,155831175	83,82924451
30	15	900	225	450	-4,235294118	-9,470588235	17,93771626	89,69204152	40,11072664	22,00735996	-7,007359956	49,10309356
39	17	1521	289	663	4,764705882	-7,470588235	22,70242215	55,80968858	-35,59515571	27,24172005	-10,24172005	104,8928296
41	30	1681	900	1230	6,764705882	5,529411765	45,76124567	30,57439446	37,40484429	28,40491118	1,595088819	2,544308341
43	35	1849	1225	1505	8,764705882	10,52941176	76,8200692	110,8685121	92,28719723	29,56810231	5,431897688	29,50551249
44	40	1936	1600	1760	9,764705882	15,52941176	95,34948097	241,1626298	151,6401384	30,14969788	9,850302122	97,02845189
47	30	2209	900	1410	12,76470588	5,529411765	162,9377163	30,57439446	70,58131488	31,89448458	-1,894484576	3,589071808
48	36	2304	1296	1728	13,76470588	11,52941176	189,467128	132,9273356	158,6989619	32,47608014	3,523919858	12,41801117
67	40	4489	1600	2680	32,76470588	15,52941176	1073,525952	241,1626298	508,816609	43,52639589	-3,526395893	12,43546799
70	45	4900	2025	3150	35,76470588	20,52941176	1279,114187	421,4567474	734,2283737	45,27118259	-0,271182590	0,0735399974
582	**416**	**25104**	**14138**	**17554**	**0**	**0**	**5179,058824**	**3958,235294**	**3012,117647**	**416**	**0**	**2206,401027**

$\bar{x}$= 582/17 = 34,23529412 $\bar{y}$ = 416/17 = 24,47058824

MONDRAGON UNIBERTSITATEA

5. ESTIMACIÓN DE ß0 Y ß1

Este método pretende encontrar el valor de los parámetros de la curva de regresión y=f(x) que minimicen la diferencia entre el valor observado para la variable explicada y correspondiente a un valor observado de x con el valor teórico que obtendremos utilizando la curva de regresión. Para lograr la ecuación de regresión han sido necesarias las siguientes fórmulas:

$$SS_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 14138 - \frac{416^2}{17} = 3958.23$$

$$SS_{xy} = \sum_{i=1}^{n} [(x_i - \bar{x}) \times (y_i - \bar{y})] = 3012.12$$

$$SS_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = 25104 - \frac{582^2}{17} = 5179.06$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{3012.12}{5179.06} = 0.58$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 24.47 - 0.58 \times 34.23 = 4.56$$

Después de realizar las siguientes operaciones podemos deducir que nuestra ecuación es la siguiente: **y = 4.56+0.58x**, donde la pendiente de la recta de regresión es 0.58 y el término independiente 4.56. Además, un salto en el índice de pobreza genera un 0.58 de cambio en el número de inmigrantes.

6. ESTIMACIÓN DEL COEFICIENTE DE CORRELACIÓN

El coeficiente de correlación de Pearson es una medida lineal entre dos variables aleatorias cuantitativas, que tiene un índice que puede utilizarse para medir el grado de relación de dos variables siempre y cuando ambas sean cuantitativas y continuas.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \times SS_{yy}}} = \frac{3012.12}{\sqrt{5179.06 \times 3958.23}} = 0.66$$

Para verificar el coeficiente de correlación utilizaremos la siguiente expresión: -1≤ r ≤+1 con independencia de la escala en que se midan x e y. En nuestro caso como r tiene valores próximos a 1 indica que hay una fuerte relación lineal entre ambas variables. Por otra parte, debido a que la estimación del coeficiente de correlación es 0.66, un número bastante bajo, podemos decir que la correlación es floja.

7. TEST DE HIPÓTESIS SOBRE EL COEFICIENTE DE CORRELACIÓN

Para verificar la prueba de hipótesis estos han sido los pasos seguidos:

7.1. Hipótesis

La prueba de hipótesis que definimos para $ß_1$ es equivalente a la que podemos definir para el coeficiente de correlación de la siguiente forma:

$$\begin{cases} H_0 : r = 0 \\ H_A : r \neq 0 \end{cases}$$

7.2. Significación

El nivel de significación para la prueba estadística es de $\boldsymbol{\alpha}$=0.05.

7.3. Test estadístico

Utilizaremos como estadístico de prueba la siguiente fórmula, que sigue una distribución t-Student con n-2 grados de libertad.

$$t = \frac{r - p}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.665 - 0}{\sqrt{\frac{1-(0.665)^2}{17-2}}} = \frac{0.665}{0.1928} = 3.45$$

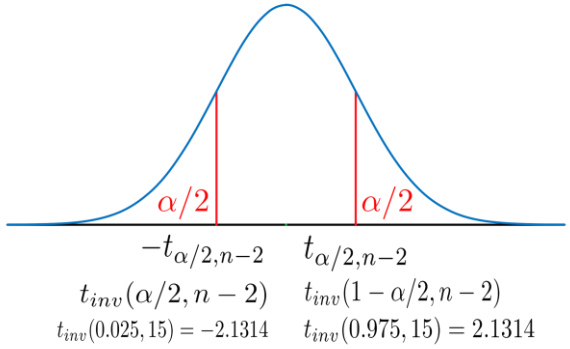
La fórmula anteriormente citada es una equivalente de la siguiente fórmula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.665\sqrt{17-2}}{\sqrt{1-0.655^2}} = 3.45$$

7.4. Valores críticos

La región crítica por su lado corresponderá con:

$$|T| > t_{\alpha/2,n-2}$$



7.5. Decisión

Teniendo en cuenta que |t|>|c|, es decir, (3.45>2.1314) y que rechazamos H_0 en favor de H_A siendo 5% el nivel de significación, podemos decir que el coeficiente de correlación de la población no es cero, por lo que rechazamos la hipótesis nula.

8. COEFICIENTE DE DETERMINACIÓN

El coeficiente de determinación denominado como R^2 y pronunciado como R cuadrado se utiliza para predecir futuros resultados o probar una hipótesis. Además, determina la calidad del modelo para replicar los resultados, y la proporción de variación de los resultados que puede explicarse por el modelo. Teniendo en cuenta que la recta de regresión lineal es la siguiente: **y = 4.56+0.58x**, para implementar el coeficiente de determinación se ha utilizado la siguiente fórmula:

$$R^2 = \frac{\sum(\hat{Y} - \bar{y})^2}{\sum(Y - \bar{y})^2} = 1 - \frac{\sum e^2}{\sum(Y - \bar{y})^2} = 1 - \frac{2206.40}{3958.23} = 0.44$$

La fórmula anteriormente citada es una equivalente de la siguiente fórmula:

$$R = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}} = 1 - \frac{2206.40}{3958.23} = 0.44$$

Dónde SS_{yy} y SSE se resolverán con las siguientes fórmulas:

$$SS_{yy} = \sum(y_i - \bar{y})^2 = 3958.23$$

$$SSE = \sum(y_i - \hat{y})^2 = 2206.4$$

En conclusión, en un modelo de regresión lineal se puede probar que $R = r^2$, por lo que $0 \leq R \leq 1$. Este coeficiente establece que 100*R representa el tanto por ciento de la suma total de las desviaciones de los valores y respecto a la media $\bar{y}$ que se puede explicar o atribuir a x en un modelo de línea recta.

MONDRAGON UNIBERTSITATEA

Por tanto, se puede deducir que:

$$R = r^2 = 0.665^2 = 0.44$$

En consecuencia, podemos ver que utilizando las diferentes fórmulas podemos obtener el mismo resultado.

9. PRUEBA DE HIPÓTESIS SOBRE SIGNIFICACIÓN DE LA REGRESIÓN LINEAL

En este apartado realizaremos la prueba de hipótesis sobre significación de regresión lineal. Para ello, un caso especial, importante, para esta prueba de hipótesis que venimos de analizar se corresponde al siguiente caso:

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_A : \beta_1 \neq 0 \end{cases}$$

Además de realizar la hipótesis para la ordenada también realizaremos la hipótesis para la pendiente para el siguiente caso, teniendo en cuenta que nuestra muestra no contempla el índice de pobreza 0:

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_A : \beta_0 \neq 0 \end{cases}$$

Para realizar las siguientes hipótesis tendremos en cuenta que $\hat{\beta}_1 \rightsquigarrow N\left(\beta_1, \frac{\sigma}{\sqrt{SS_{xx}}}\right)$, pero cómo **σ** es desconocida, usamos S con lo cual seguirá una T Student. Para ello, utilizaremos las siguientes fórmulas, para calcular los errores estándar:

$$S = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{2206.4}{15}} = 12.13$$

$$T_0 = \frac{\hat{\beta}_0 - \beta_{00}}{S \times \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} = \frac{4.56 - 0}{12.13 \times \sqrt{\frac{1}{17} + \frac{1172.05}{5179.06}}} = 0.7$$

MONDRAGON UNIBERTSITATEA

$$T_0 > |t_{\alpha/2,n-2}| \Rightarrow 0.7 > 2.1314 \Rightarrow \text{No se rechaza}$$

$$T_1 = \frac{\hat{\beta}_1 - \beta_{10}}{\frac{S}{\sqrt{SS_{xx}}}} = \frac{0.58 - 0}{\frac{12.13}{\sqrt{5179.06}}} = 3.45$$

$$T_1 > |t_{\alpha/2,n-2}| \Rightarrow 3.45 > 2.1314 \Rightarrow \text{Se rechaza } H_0$$

estadístico t		valor crítico	Resolución	Conclusión
$\lvert t_{\beta_0} \rvert$	**vs**	$t_{\alpha/2,n-2}$	**0,70<2.1314**	El estadístico de prueba $t_{ß0}$, es menor que el valor crítico de tablas, por lo que **no podemos rechazar H_0**. En este caso, sabemos que la ordenada al origen (el valor de **$ß_0$**), no es significativa, es decir, que va a ser igual a 0. Entonces los datos no contienen evidencias de que **H_0** sea falsa.
$\lvert t_{\beta_1} \rvert$		$t_{\alpha/2,n-2}$	**3,451039361 > 2.1314**	En cuanto al estadístico de prueba $t_{ß1}$, es mayor que el valor crítico de tablas, por lo que **rechazamos H_0**. En este caso, **aceptamos H_1**, si es significativa la pendiente. **$ß_0$**, no es significativa, es decir, que va a ser igual a 0, pero la pendiente (**m**), si es significativa.

MONDRAGON UNIBERTSITATEA

10. PRUEBA DE HIPÓTESIS SOBRE LA PENDIENTE DE LA RECTA DE REGRESIÓN LINEAL

Si suponemos que se desea contrastar la hipótesis de que dicha pendiente es igual a un cierto valor **ß₁₀**, estableceremos como hipótesis adecuadas:

$$\begin{cases} H_0 : \beta_1 = 3 \\ H_A : \beta_1 \neq 3 \end{cases}$$

$$T_3 = \frac{\hat{\beta}_0 - \beta_{00}}{S - \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SS_{xx}}}} = \frac{4.55 - 3}{12.13 - \sqrt{\frac{1}{17} + \frac{1172.05}{5107.06}}} = 0.24$$

$$T_3 > |t_{\alpha/2,n-2}| \Rightarrow 0.24 > 2.1314 \Rightarrow \text{No se rechaza.}$$

Por tanto, podemos decir que los datos no contienen evidencias de que **H₀** sea falsa.

MONDRAGON UNIBERTSITATEA

11. PRUEBA DE HIPÓTESIS SOBRE LA ORDENADA EN EL ORIGEN DE LA RECTA DE REGRESIÓN LINEAL

Podemos utilizar un proceso semejante para establecer pruebas de hipótesis sobre la ordenada en el origen de la recta de regresión planteamos las hipótesis:

$$\begin{cases} H_0 : \beta_0 = 0.4 \\ H_A : \beta_0 \neq 0.4 \end{cases}$$

$$T_{0.4} = \frac{\hat{\beta}_1 - \beta_{10}}{\frac{S}{\sqrt{SS_{xx}}}} = \frac{0.58 - 0.4}{\frac{12.13}{\sqrt{5179.06}}} = 1.06$$

$$T_{0.4} > |t_{\alpha/2,n-2}| \Rightarrow 1.06 < 2.1314 \Rightarrow \text{No se rechaza.}$$

Por tanto, podemos decir que los datos no contienen evidencias de que H0 sea falta.

MONDRAGON UNIBERTSITATEA

12. INTERVALO DE CONFIANZA PARA LOS PARÁMETROS DE LA RECTA DE REGRESIÓN

En este apartado, además de las estimaciones puntuales que hemos deducido para β_0 y para β_1, las distribuciones muestrales que hemos deducido para los estadísticos utilizados en las pruebas de hipótesis anteriores nos permiten, utilizando dichos estadísticos como expresiones pivótales, deducir intervalos de confianza del (100 − α) por cien para $ß_0$ y $ß_1$. Operando en la forma habitual llegaríamos a:
Intervalo de confianza del 100*(1-**α**) por cien $ß_1$:

$$\hat{\beta}_1 - t_{\alpha/2,n-2}\frac{S}{\sqrt{SS_{xx}}} < \beta_1 < \hat{\beta}_1 + t_{\alpha/2,n-2}\frac{S}{\sqrt{SS_{xx}}}$$

Intervalo de confianza del 100*(1-**α**) por cien para $ß_0$:

$$\hat{\beta}_0 - t_{\alpha/2,n-2}\sqrt{\frac{1}{n}+\frac{\bar{x}}{SS_{xx}}} < \beta_0 < \hat{\beta}_0 + t_{\alpha/2,n-2}S\sqrt{\frac{1}{n}+\frac{\bar{x}}{SS_{xx}}}$$

	left	right	Intervalos
β_1	$0.58 - 2.13 \times \frac{12.13}{\sqrt{5179.06}} = 0.22$	$0.58 + 2.13 \times \frac{12.13}{\sqrt{5179.06}} = 0.94$	**0.22<0.58<0.94**
β_0	$4.56 - 2.13 \times \sqrt{\frac{1}{17}+\frac{34.23}{5179.06}} = 4.01$	$4.56 + 2.13 \times \sqrt{\frac{1}{17}+\frac{34.23}{5179.06}} = 5.1$	**4.01<4.55<5.10**

MONDRAGON UNIBERTSITATEA

13. INTERVALO DE ESTIMACIÓN PARA LA RESPUESTA MEDIA

En este apartado podemos obtener un intervalo de confianza para la respuesta media en un valor específico de x por ejemplo x_p, en nuestro caso x_p=4.

$$\hat{\mu}_{Y|x_p} = \hat{\beta}_0 + \hat{\beta}_1 \times x_p = 4.56 + 0.58 \times 4 = 6.88$$

$$Var\left(\hat{\mu}_{Y|x_p}\right) = \sigma^2 \times \left[\frac{1}{n} + \frac{(x_p - \bar{x})^2)}{SS_{xx}}\right] = 12.12^2 \times \left(\frac{1}{17} + \frac{914.17}{5179.06}\right) = 34.61$$

Haciendo uso de s^2 como estimador de $\boldsymbol{\sigma}^2$, siguiendo un proceso ya conocido demostraremos que el estadístico:

$$T = \frac{\hat{\mu}_{Y|x_p} - \mu_{Y|x_p}}{s \times \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}} = \frac{6.88 - 6.89}{12.12 \times \sqrt{\frac{1}{17} + \frac{914.17}{5179.06}}} = 5.71$$

sigue una distribución t de Student con (n-2) grados de libertad.

Esto nos permite establecer el siguiente intervalo de estimación para la respuesta media del 100*(1-$\boldsymbol{\alpha}$) por ciento.

$$\hat{\mu}_{Y|x_p} - t_{\alpha/2,n-2} \times s \times \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} < \mu_{Y|x_p} < \hat{\mu}_{Y|x_p} + t_{\alpha/2,n-2} \times s\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

MONDRAGON UNIBERTSITATEA

$$I_{\mu_p/x_p} = \hat{\mu}_{p/x_p} \pm t_{\alpha/2,n-2} \times s \times \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

	left	right	Intervalos
$\mu_{Y\|x_p}$	$6.88 - 2.13 \times 12.12 \times \sqrt{\frac{1}{17} + \frac{914.17}{5179.06}} = -5.65$	$6.88 + 2.13 \times 12.12 \times \sqrt{\frac{1}{17} + \frac{914.17}{5179.06}} = 19.42$	$-5.56 < \mu_{Y\|x_p} < 19.42$

Podemos observar que el ancho del intervalo de confianza es función del valor x_p, siendo mínimo cuando $x_p = \bar{x}$ y aumentando cuando $|x_p = \bar{x}$ y| aumenta.

MONDRAGON UNIBERTSITATEA

14. INTERVALO DE PREDICCIÓN PARA LA OBSERVACIÓN FUTURA

El intervalo de confianza que hemos obtenido nos permite estimar el resultado medio de un elevado número de observaciones realizadas para un determinado valor de la variable de regresión x_p en nuestro caso x_p=4. Se ha utilizado el intervalo de estimación de la respuesta media más la variabilidad del error aleatorio en una observación particular, que como hemos establecido es constante y de valor σ^2. Así pues, la varianza de una observación futura vendrá dada por:

$$Var(y-\hat{y}) = \sigma^2\left[1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}\right] = 12.12^2 \times \left(1+\frac{1}{17}+\frac{914.17}{5179.06}\right) = 181.71$$

El razonamiento seguido en apartados anteriores nos permite establecer el siguiente intervalo de predicción del 100*(1-α) por ciento para una observación futura y cuando $x=x_p$:

$$\hat{y}_p - t_{\alpha/2,n-2} \times s \times \sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}} < y_p < \hat{y}_p + t_{\alpha/2,n-2} \times s \times \sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}}$$

$$I_{Y_p/x_p} = \hat{y}_{p/x_p} \pm t_{\alpha/2,n-2} \times s \times \sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{SS_{xx}}}$$

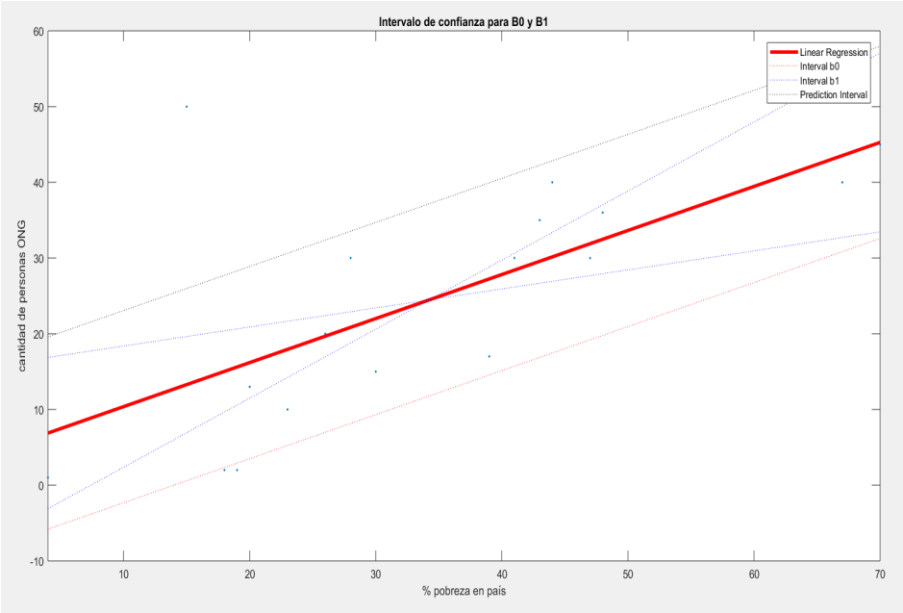
	left	right	Intervalos
$\mu_{Y\|x_p}$	$6.88 - 2.13 \times 12.12 \times \sqrt{1+\frac{1}{12}+\frac{914.17}{5179.06}} = -21.85$	$6.88 + 2.13 \times 12.12 \times \sqrt{1+\frac{1}{12}+\frac{914.17}{5179.06}} = 35.62$	$-21.85 < \mu_{Y\|x_p} < 35.62$

Al igual que en el intervalo de estimación de la respuesta media, este intervalo de predicción es más estrecho cuando $x_p = \bar{x}$ y aumenta conforme $|x_p = \bar{x}|$ se hace mayor.

MONDRAGON UNIBERTSITATEA

15. GRÁFICO PARA LOS INTERVALOS DE CONFIANZA ß0 Y ß1

En este apartado, se representa gráficamente el intervalo de confianza para ß0 Y ß1 calculados con anterioridad. Además, está representada la línea de regresión y también el intervalo de predicción.



MONDRAGON UNIBERTSITATEA