

Pregunta 1

Correcta

Se puntúa 0,00 sobre 1,00

Considere estimar por el Método de Mínimos Cuadrados Ordinarios (MCO) el modelo de regresión lineal $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

¿Cuál de las siguientes afirmaciones es correcta?

- I. Cuanto más alta la correlación entre las variables explicativas x_{1i} y x_{2i} más serio será el problema de multicolinealidad.
 - II. Cuanto más alta la correlación entre las variables explicativas x_{1i} y x_{2i} mayores serán los errores estándares de los beta sombrero de x_{1i} y x_{2i} que se reportan en la segunda columna de la regresión.
 - III. Cuanto más alta la correlación entre las variables explicativas x_{1i} y x_{2i} mayores serán los *p-values* para testear la hipótesis nula que los beta sombrero de x_{1i} y de x_{2i} son iguales a cero (individualmente), que se reportan en la cuarta columna de la regresión (contra la hipótesis alternativa de que son diferentes a cero).
- ☐ a. Solo I es correcta
 - ☐ b. Solo II es correcta
 - ☐ c. Solo I y II son correctas
 - ☐ d. Solo I y III son correctas
 - ☒ e. I, II y III son correctas ✓

Como $Var(\hat{\beta}_1) = \frac{\sigma^2 S_{22}}{S_{11}S_{22} - S_{12}^2} = \frac{\sigma^2}{S_{11}(1 - r_{12}^2)}$, cuanto más alta sea la correlación entre las variables explicativas (o sea, r_{12}^2 más cercano a 1) menor es el denominador en la fórmula de la varianza, por lo tanto mayor será la varianza, y consecuentemente el error estándar. Análogamente para $Var(\hat{\beta}_2)$. Como consecuencia, el estadístico t que se reporta en las regresiones será más bajo, y el p-value asociado, más alto. Precisamente en esto consiste el problema de la multicolinealidad.

La respuesta correcta es:

I, II y III son correctas

Correcta

Puntos para este envío: 1,00/1,00. Contando con los intentos anteriores, daría **0,00/1,00**.

Pregunta 2

Correcta

Se puntúa 0,67 sobre 1,00

Suponga que usted corre una regresión simple con constante por el Método de Mínimos Cuadrados Ordinarios y obtiene la recta estimada


$$\hat{y}_i = 0.6 + 2 x_i$$

con un $R^2 = 0.80$.

Suponga que usted decide ahora regresar x_i en y_i , usando las mismas observaciones, obteniendo la siguiente ecuación estimada:

$$\hat{x}_i = \hat{\alpha} + \hat{\beta} y_i$$

¿Cuál de las siguientes es la opción correcta respecto de esta última ecuación estimada? Hint: multiplicando ambas pendientes (las fórmulas), ¿qué obtiene?

- ☒ a. $\hat{\beta} = 0.4 \quad R^2 = 0.80$ 
- ☐ b. $\hat{\beta} = 0.5 \quad R^2 = 0.40$
- ☐ c. $\hat{\beta} = 0.5 \quad R^2 = 0.80$
- ☐ d. $\hat{\beta} = 1.6 \quad R^2 = 0.80$
- ☐ e. $\hat{\beta} = 2.5 \quad R^2 = 0.80$

Claramente ambas regresiones deben tener el mismo $R^2 = 0.8$, ya que el R^2 es la correlación al cuadrado entre estas dos variables en una regresión simple. En ambas regresiones son las mismas dos variables.

Sea $\hat{\delta}$ el estimador de la primera regresión. Se nos informa que $\hat{\delta} = 2$.

Sabemos que $\hat{\delta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)}$

Por otro lado,

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(y)}$$

Multiplicando las pendientes,

$$\hat{\delta}\hat{\beta} = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)} \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(y)} = \widehat{Corr}(X, Y)^2 = R^2 = 0.8$$

Consecuentemente, $\hat{\beta} = \frac{0.8}{2} = 0.4$

La respuesta correcta es:

$$\hat{\beta} = 0.4 \quad R^2 = 0.80$$

Correcta

Puntos para este envío: 1,00/1,00. Contando con los intentos anteriores, daría 0,67/1,00.

Pregunta 3

Correcta

Se puntúa 0,67 sobre 1,00

Se corre una regresión lineal, obteniéndose la siguiente salida:

```
. regress saleprice lotsize bedroom bath stories driveway recroom basement gas aircond
> garage desireloc // Regresion por MCO
```

Source	SS	df	MS	Number of obs	=	546
Model	2.6158e+11	11	2.3780e+10	F(11, 534)	=	99.97
Residual	1.2703e+11	534	237874666	Prob > F	=	0.0000
				R-squared	=	0.6731
				Adj R-squared	=	0.6664
Total	3.8860e+11	545	713032635	Root MSE	=	15423

saleprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lotsize	3.546303	.3503	10.12	0.000	2.858168	4.234438
bedroom	1832.003	1047	1.75	0.081	-224.7409	3888.748
bath	14335.56	1489.921	9.62	0.000	11408.73	17262.38
stories	6556.946	925.2899	7.09	0.000	4739.291	8374.6
driveway	6687.779	2045.246	3.27	0.001	2670.065	10705.49
recroom	4511.284	1899.958	2.37	0.018	778.9759	8243.592
basement	5452.386	1588.024	3.43	0.001	2332.845	8571.926
gas	12831.41	3217.597	3.99	0.000	6510.706	19152.11
aircond	12632.89	1555.021	8.12	0.000	9578.182	15687.6
garage	4244.829	840.5442	5.05	0.000	2593.65	5896.008
desireloc	9369.513	1669.091	5.61	0.000	6090.724	12648.3
_cons	-4038.35	3409.471	-1.18	0.237	-10735.97	2659.271

El valor 0.0000 correspondiente a **Prob > F** = **0.0000** refleja

- ☐ a. Que rechazo la hipótesis nula $H_0: R^2 = 0$
- ☐ b. Que rechazo la hipótesis nula $H_0: R^2 = 1$
- ☐ c. Que no rechazo la hipótesis nula $H_0: R^2 = 0$
- ☐ d. Que no rechazo la hipótesis nula $H_0: R^2 = 1$
- ☒ e. Que rechazo la hipótesis nula de que todas las pendientes (todos los betas), excepto el intercepto, son simultáneamente iguales a cero ✔

Efectivamente este p-value es el p-value asociado al estadístico F que debe emplearse para testear la hipótesis nula de que todas las pendientes (todos los betas), excepto el intercepto, son simultáneamente iguales a cero, contra la hipótesis alternativa que al menos uno de los betas es diferente de cero.

La respuesta correcta es:

Que rechazo la hipótesis nula de que todas las pendientes (todos los betas), excepto el intercepto, son simultáneamente iguales a cero

Correcta

Puntos para este envío: 1,00/1,00. Contando con los intentos anteriores, daría 0,67/1,00.

Pregunta 4

Correcta

Se puntúa 1,00 sobre 1,00

Al ejecutar `print(modelo1.summary())` se obtiene la siguiente salida usando StatsModels,

```

=====
                        OLS Regression Results
=====
Dep. Variable:      Compensacion_CEO      R-squared:                0.434
Model:              OLS                  Adj. R-squared:           0.426
Method:             Least Squares        F-statistic:              52.24
Date:               Fri, 25 Jul 2025      Prob (F-statistic):       5.50e-10
Time:               10:30:04              Log-Likelihood:          -73.655
No. Observations:   70                  AIC:                     151.3
Df Residuals:       68                  BIC:                     155.8
Df Model:           1
Covariance Type:    nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          0.6000      0.112       5.342      0.000      0.376      0.824
Ganancias      0.0008      0.000       7.228      0.000      0.001      0.001
=====
Omnibus:                 21.745    Durbin-Watson:           2.045
Prob(Omnibus):            0.000    Jarque-Bera (JB):         36.165
Skew:                     1.136    Prob(JB):                 1.40e-08
Kurtosis:                  5.691    Cond. No.                  1.29e+03
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.29e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

```

¿Qué valor obtiene si ejecuta `modelo1.params[0]` ?

- ☐ a. 0.0001
- ☐ b. 0.112
- ☐ c. 0.426
- ☐ d. 0.434
- ☒ e. 0.6000 ✓

StatsModels guarda los estimadores bajo ".params" y en este caso es el primero, correspondiente a la constante (el alfa sombrero).

La respuesta correcta es:

0.6000

Correcta

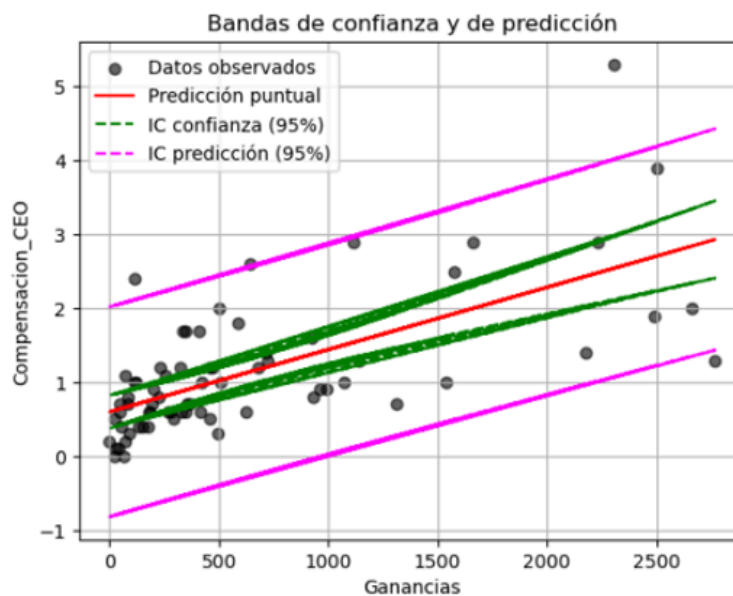
Puntos para este envío: 1,00/1,00.

Pregunta 5

Correcta

Se puntúa 0,67 sobre 1,00

A continuación, se presenta un gráfico con las bandas de predicción y de confianza al 95% para el premio o compensación a los CEO, que surge de una regresión simple con constante en las ganancias de las empresas de primera línea:



¿Cuál de las siguientes afirmaciones es correcta?

- ☐ a. Las bandas de confianza al 95% son utilizadas para construir intervalos de predicción de observaciones individuales futuras, de manera que tiene en cuenta la varianza del valor predicho de la variable dependiente como también la asociada al residuo.
- ☐ b. Las bandas de confianza al 95% son más amplias que las bandas de predicción, ya que incorporan más incertidumbre.
- ☒ c. Las bandas de predicción al 95% representan un intervalo dentro del cual se espera que caiga una observación individual de la compensación a los CEO, dado un nivel específico de ganancias de las empresas. ✓
- ☐ d. Las bandas de confianza al 95% aseguran que el 95% de las observaciones históricas están contenidas dentro del intervalo.
- ☐ e. Las bandas de predicción se construyeron para estimar la media poblacional de las ganancias de las empresas de primera línea dado un valor específico de compensación a los CEO para un determinado nivel de confianza.

Las bandas de predicción son un intervalo de confianza (IC) para capturar qué valor tomaría una observación (futura) individual de la variable dependiente y_i dado un cierto valor de la variable independiente x_i . Estadísticamente, el objetivo es estimar un IC para $y_i = \hat{y}_i + e_i$. Esto incluye la incertidumbre en la estimación del valor esperado \hat{y}_i y también la variabilidad inherente a la parte residual e_i . Por eso las bandas de predicción son más anchas que las bandas de confianza.

La opción A) es incorrecta porque las bandas de confianza no sirven para predecir observaciones individuales futuras sino para estimar un intervalo de confianza para el valor predicho de la variable dependiente, es decir, para \hat{y}_i , donde $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$. B) es incorrecta porque las bandas de predicción al 95% son más amplias que las bandas de confianza, ya que incorporan más incertidumbre (aquella asociada al residuo). D) es incorrecta porque las bandas de confianza no están diseñadas para contener observaciones individuales, ni pasadas ni futuras. E) confunde que la variable a explicar es la compensación a los CEO.

La respuesta correcta es:

Las bandas de predicción al 95% representan un intervalo dentro del cual se espera que caiga una observación individual de la compensación a los CEO, dado un nivel específico de ganancias de las empresas.

Correcta

Puntos para este envío: 1,00/1,00. Contando con los intentos anteriores, daría 0,67/1,00.