



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

INFORME

TAREA III

“REGRESION SIMPLE”

ECONOMETRIA

**PROFESOR: RODRIGO
ORTEGA**

INTEGRANTE:
BARBARA LIZAMA

27 DE ABRIL DE 2021

PROBLEMA

Se adjunta un set de datos TV con la muestra de número de horas gastadas frente al televisor y la edad de la persona, generando diversas pruebas en relación a la población.

OBJETIVO

Realizar diversas pruebas de hipótesis con diversos análisis como modelos de regresión, interpretando sus resultados viendo los valores observados y estimados, para establecer diversas conclusiones en relación a estas hipótesis planteadas.

1) Estime un modelo de regresión para explicar el tiempo gastado mirando televisión en función de la edad de la persona. Estime el modelo usando las tres formas aprendidas en clase:

- Vía fórmula

Formula Regresión Simple

Se tomara como variable dependiente $Y = \text{TV (horas/semana)}$

Se tomara como variable independiente $X = \text{Edad (años)}$

$$B1 = \frac{\text{Covarianza (X, Y)}}{\text{Varianza (X)}} = \frac{283,4067}{283,9574} = 0,9981 \text{ hora semana/año}$$

$$\text{Promedio X} = 41,5340$$

$$\text{Promedio Y} = 78,7330$$

$$Y \text{ promedio} = \beta_0 + \beta_1 * X \text{ promedio}$$

$$B0 = Y \text{ promedio} - B0 * X \text{ promedio} = 37,2796$$

$$Y = B0 + B1 * X + \mu = 37,2796 + 0,9981 * X$$

Ilustración 1 Formulas para sacar los betas

TV (horas/semana) - Y	Edad (años) - X	Y - estimado	μ - estimado
48	21	58,24	-10,24
47	21	58,24	-11,24
73	18	55,24	17,76
65	23	60,23	4,77

Ilustración 2 Muestra del cálculo del Y estimado con su Residual aplicando formulas

- **Utilizando el método de minimización de la suma de cuadrados de residuales (SSR) con Solver**

Lo que se busca aquí es minimizar la suma de residuales al cuadrado, por lo cual se debe empezar haciendo un supuesto para β_0 y β_1 , con lo cual se obtiene el Y estimado, el μ estimado (residual) y los residuales al cuadrado.

TV (horas/semana) - Y	Edad (años) - X	Y - estimado	μ - estimado (residual)	SSR (suma de residuales $\wedge 2$)
48	21	58,24	-10,24	104,83
47	21	58,24	-11,24	126,31
73	18	55,24	17,76	315,25
65	23	60,23	4,77	22,71

Ilustración 3 Muestra de datos con el cálculo de los supuestos de betas

Luego se aplica Solver de la siguiente manera, nuestra función objetivo es SSR, variables a cambiar los betas y con la restricción de que SR sea igual a 0, así se obtiene la minimización de la SSR.

β_0	37,2796
β_1	0,9981
r	0,8300
Suma de residuales (SR)	0
Suma de cuadrados residuales (SSR)	26.176,6209

Ilustración 4 Datos obtenidos con Solver

Ilustración 5 Generación de Optimización

○ Utilizando la función regresión en Excel o Real Statistics

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0,8300
Coefficiente de determinación R ²	0,6890
R ² ajustado	0,6874
Error típico	11,3277
Observaciones	206

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	57.985,6947	57.985,6947	451,8949	1,23517E-53
Residuos	204	26.176,6209	128,3168		
Total	205	84.162,3155			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	37,2796	2,1037	17,7210	0,0000	33,1318	41,4273	33,1318	41,4273
Edad (años) - X	0,9981	0,0470	21,2578	0,0000	0,9055	1,0906	0,9055	1,0906

Ilustración 6 Función Regresión por Análisis de datos de Excel

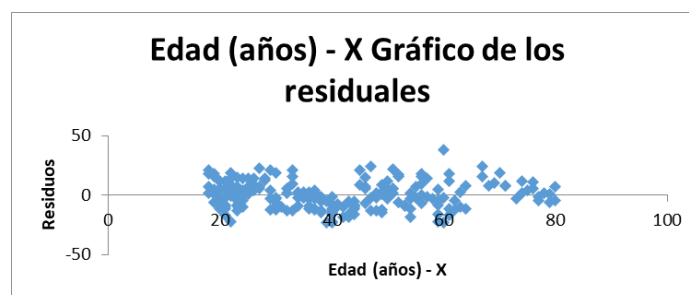


Ilustración 7 Grafico de Residuales

2) Estableciendo las pruebas de hipótesis correspondientes, interprete los resultados en términos de la significancia del modelo y de los parámetros (betas) estimados. Determine e interprete la bondad de ajuste. Interprete la pendiente del modelo.

$H_0: \beta_1 = 0$ No hay Regresión

$H_a: \beta \neq 0$ Hay Regresión

Ilustración 8 Test de Hipótesis

ANOVA				Alpha	0,05	
	df	SS	MS	F	p-value	sig
Regression	1	57985,69467	57985,69467	451,894909	1,2352E-53	yes
Residual	204	26176,62086	128,3167689			
Total	205	84162,31553				

Ilustración 9 Calculo con Real Statistics

Se puede concluir que se rechaza la hipótesis nula, dado que el p value es menor al alpha, por lo cual si existe regresión lineal y si es significativo globalmente el modelo.

	coeff	std err	t stat	p-value	lower	upper
Intercept	37,2795789	2,103692477	17,72102116	3,7291E-43	33,1318107	41,4273471
Edad (años) - X	0,99806063	0,046950281	21,25781996	1,2352E-53	0,9054906	1,09063066

Ilustración 10 Calculo con Real Statistics

Para revisar si hay significancia a nivel local, es decir si algún beta es significativo o no para el modelo, se revisó los p values de cada beta, lo cuales indican que ambos betas si son significativos para el modelo.

OVERALL FIT

Multiple R	0,8300
R Square	0,6890
Adjusted R Square	0,6874
Standard Error	11,3277
Observations	206

Ilustración 11 Bondad de Ajuste

La bondad de ajuste es igual al R^2 que permite ver cuánto está explicando las variables independientes a las variables dependientes, en este caso se puede deducir que el 68,9% de la variación esta explicada por el modelo y el restante se puede atribuir a otras variables que no están en nuestro modelo.

	coeff
Intercept	37,2795789
Edad (años) - X	0,99806063

Ilustración 12 Pendiente del Modelo

Como conclusión de la pendiente se puede deducir del β_1 que por cada año extra de edad la persona vera 0,998 horas más de televisión por semana.

3) Grafique los valores observados y estimados por el modelo. Para estos últimos utilice una línea.

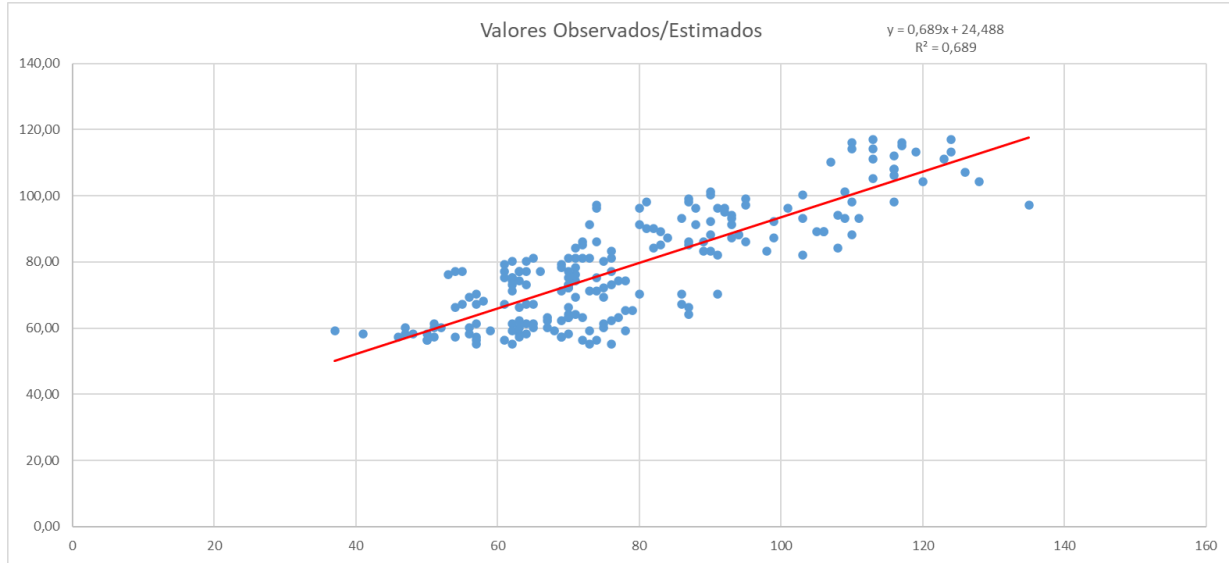


Ilustración 13 Grafico Dispersión Valores Observados y Estimados

4) Transforme ambas variables a logaritmo y corra nuevamente el modelo, esta vez con la función regresión de Excel o Real Statistics. Comente sobre la significancia del modelo y su bondad de ajuste. Interprete la pendiente del modelo.

Regression Analysis

OVERALL FIT

Multiple R	0,7852	AIC	-760,760774
R Square	0,6166	AICc	-760,641962
Adjusted R Square	0,6147	SBC	-754,105022
Standard Error	0,1570		
Observations	206		

ANOVA				Alpha	0,05	
	df	SS	MS	F	p-value	sig
Regression	1	8,0888	8,0888	328,0478	2,43001E-44	yes
Residual	204	5,0301	0,0247			
Total	205	13,1190				

	coeff	std err	t stat	p-value	lower	upper
Intercept	2,6123	0,0957	27,3010	4,82475E-70	2,4237	2,8010
Ln Edad	0,4728	0,0261	18,1121	2,43001E-44	0,4213	0,5243

Ilustración 14 Calculo con Real Statistics

Se puede concluir que se rechaza la hipótesis nula, dado que el p value es menor al alpha, por lo cual si existe regresión lineal y si es significativo globalmente el modelo.

La bondad de ajuste en este caso se puede deducir que el 61,66% de la variación esta explicada por el modelo y el restante se puede atribuir a otras variables que no están en nuestro modelo.

Como conclusión de la pendiente se puede deducir del β_1 que cada un 1% de aumento en la edad, la persona vera un 0,4728% de horas más de televisión por semana.

5) Realice un gráfico de residuales en cada caso (modelo sobre variables originales y modelo con logaritmos). ¿Se observa algún problema de heterocedasticidad?

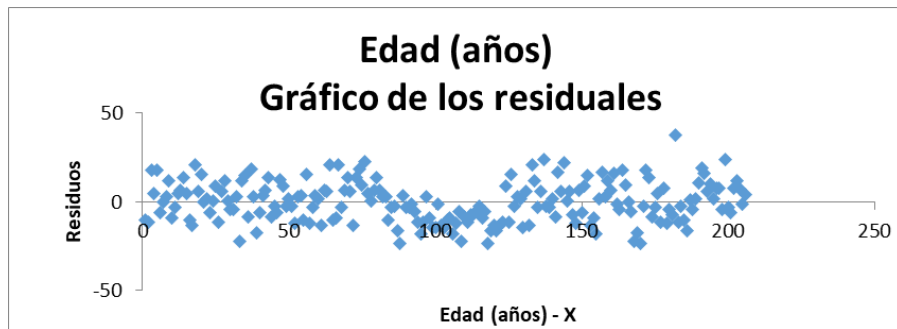


Ilustración 15 Grafico con Variables Originales

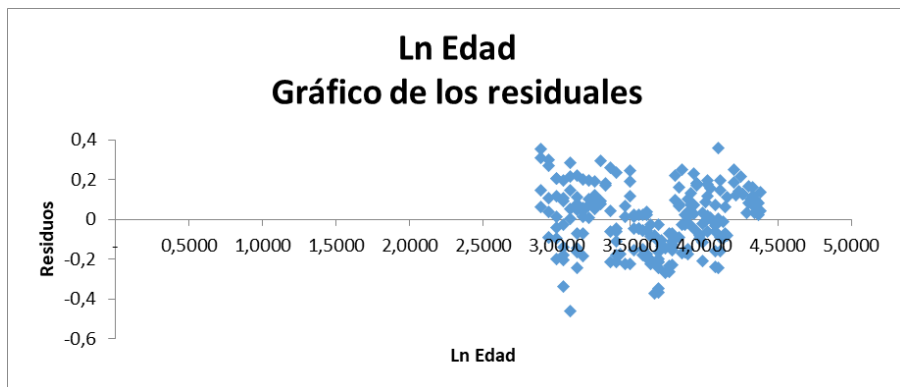


Ilustración 16 Grafico con Logaritmos

Al comparar ambos modelos, no se observa algún problema de Heterocedasticidad, ya que se podría indicar que la varianza de los errores es constante en todas las observaciones realizadas

CONCLUSION

Por medio de las diferentes pruebas que se realizó durante el informe se puede concluir que las variables TV y Edad son significativas tanto global como localmente en el modelo.

Cuando se revisó la bondad de ajuste en ambos casos fue más de un 60% de la variación se explica por el modelo y se puede indicar que el restante se puede atribuir a otras variables que no están presentes en nuestro modelo.