

UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

INFORME
TAREA V
“DIAGNOSTICO DE
REGRESION,
SELECCIÓN DE
MODELOS Y
REGRESION CON
VARIABLES BINARIAS”

ECONOMETRIA

PROFESOR: RODRIGO ORTEGA
AYUDANTE: DIEGO BASCUÑAN

INTEGRANTE:
BARBARA LIZAMA

PROBLEMA

Se adjunta un set de datos de hamburguesas, de azúcar y datos de entrenamiento para generar diversas pruebas en relación a la población de estos data set.

OBJETIVO

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

I. DATA SET HAMBURGUESAS

En un negocio de venta de hamburguesas, se desea explicar las ventas en función del precio medio de las hamburguesas y los gastos en publicidad.

1. Corra dos modelos

Regression Analysis

OVERALL FIT

Multiple R	0,6695	AIC	240,8983
R Square	0,4483	AICc	241,4697
Adjusted R Square	0,4329	SBC	247,8507
Standard Error	4,8861		
Observations	75		

ANOVA				Alpha	0,05	
	df	SS	MS	F	p-value	sig
Regression	2	1396,5389	698,2695	29,2479	5,04086E-10	yes
Residual	72	1718,9429	23,8742			
Total	74	3115,4819				

	coeff	std err	t stat	p-value	lower	upper	vif
Intercept	118,9136	6,3516	18,7217	2,21429E-29	106,25185	131,57537	
price	-7,9079	1,0960	-7,2152	4,424E-10	-10,09268	-5,72303	1,00070
advert	1,8626	0,6832	2,7263	0,00804	0,50066	3,22451	1,00070

Ilustración 1 Con todas las variables en nivel

Regression Analysis

OVERALL FIT

Multiple R	0,6849	AIC	-413,2533
R Square	0,4691	AICc	-412,6819
Adjusted R Square	0,4544	SBC	-406,3008
Standard Error	0,0624		
Observations	75		

ANOVA				Alpha	0,05	
	df	SS	MS	F	p-value	sig
Regression	2	0,2475	0,1238	31,8101	1,25979E-10	yes
Residual	72	0,2801	0,0039			
Total	74	0,5276				

	coeff	std err	t stat	p-value	lower	upper	vif
Intercept	5,3199	0,1380	38,5640	7,96988E-50	5,0449	5,5949	
In price	-0,5749	0,0793	-7,2459	3,88144E-10	-0,7331	-0,4168	1,0000
In advert	0,0454	0,0137	3,3269	0,0014	0,0182	0,0727	1,0000

Ilustración 2 Con todas las variables en logaritmo natural

Determine la significancia de cada modelo y los coeficientes, e interprete los coeficientes.

$$H_0: \beta_1 = 0, \beta_2 = 0$$

$$H_a: H_0 \text{ no es verdadera}$$

Ilustración 3 Test de hipótesis Significancia Global

Modelo Nivel - Nivel				Alpha	0,05	
	df	SS	MS	F	p-value	sig
Regression	2	1396,5389	698,2695	29,2479	5,04086E-10	yes
Residual	72	1718,9429	23,8742			
Total	74	3115,4819				

Modelo Log - Log				Alpha	0,05	
	df	SS	MS	F	p-value	sig
Regression	2	0,2475	0,1238	31,8101	1,25979E-10	yes
Residual	72	0,2801	0,0039			
Total	74	0,5276				

Se puede concluir que se rechaza la hipótesis nula, dado que el P-value es menor al alpha, por lo cual si es significativo globalmente el modelo tanto a nivel como a logaritmo.

$$H_0: \beta_1 = 0 \quad H_0: \beta_2 = 0$$

$$H_a: \beta_1 \neq 0 \quad H_a: \beta_2 \neq 0$$

Ilustración 4 Test de hipótesis Significancia Local

	coeff	std err	t stat	p-value
Intercept	118,9136	6,3516	18,7217	2,21429E-29
price	-7,9079	1,0960	-7,2152	4,424E-10
advert	1,8626	0,6832	2,7263	0,00804

Ilustración 5 Significancia local Modelo Nivel

	coeff	std err	t stat	p-value
Intercept	5,3199	0,1380	38,5640	7,96988E-50
ln price	-0,5749	0,0793	-7,2459	3,88144E-10
ln advert	0,0454	0,0137	3,3269	0,0014

Ilustración 6 Significancia Local Modelo Log

Se puede concluir que se rechaza la hipótesis nula, dado que el P-value es menor al alpha, por lo cual si es significativo localmente los coeficientes en el modelo tanto a nivel como a logaritmo.

	coeff
Intercept	118,9136
price	-7,9079
advert	1,8626

Ilustración 7 Coeficientes Modelo Nivel

β_1 (Price): Por cada dólar adicional en el precio de la hamburguesa, las ventas disminuirán en 7.908 dólares

β_2 (Advert): Por cada dólar adicional gastado en publicidad, las ventas aumentaran en 1.863 dólares

	<i>coeff</i>
Intercept	5,3199
ln price	-0,5749
ln advert	0,0454

Ilustración 8 Coeficientes Modelo Log

β_1 (Ln Price): Por cada 1% adicional de aumento en el precio de la hamburguesa, las ventas disminuirán en 0,5749%

β_2 (Ln Advert): Por cada 1% adicional gastado en publicidad, las ventas aumentaran en 0,0454%

2. Realice un diagnostico completo de cada regresión incluyendo:

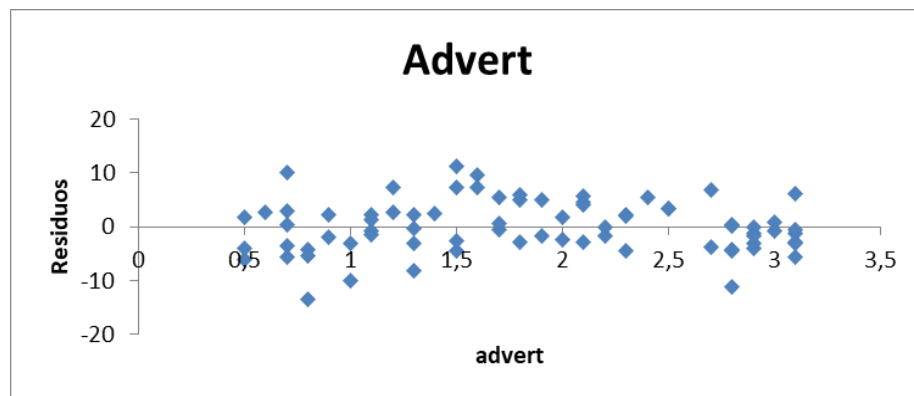
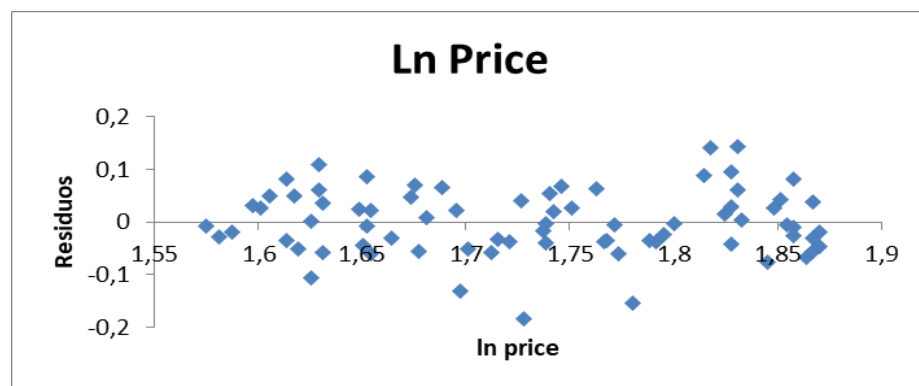


Ilustración 9 Gráficos Residuales Modelo Nivel

A la vista se podría indicar que no hay problemas de heterocedasticidad en los gráficos de residuales del modelo nivel.



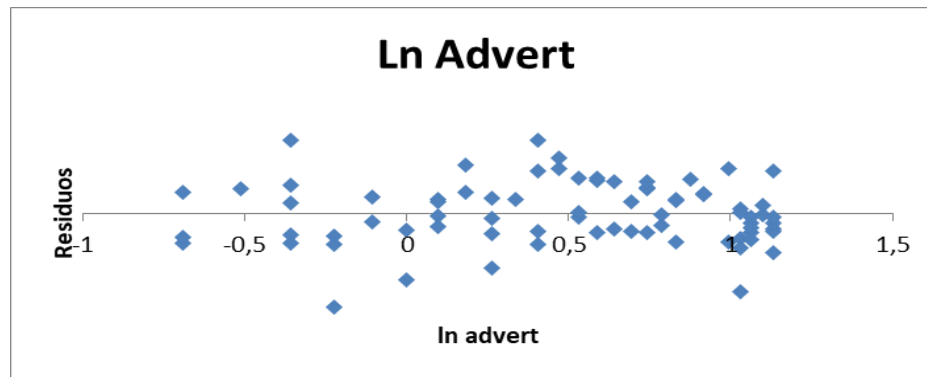


Ilustración 10 Gráficos Residuales Modelo Log

A la vista se podría indicar que no hay problemas de heterocedasticidad en los gráficos de residuales del modelo logarítmico.

Normalidad de los residuales

En ambos modelos se puede ver que sus residuales son normales.

Shapiro-Wilk Test

	<i>Residual</i>
W-stat	0,9879654
p-value	0,7028158
alpha	0,05
normal	yes

d'Agostino-Pearson

DA-stat	0,5351967
p-value	0,7652151
alpha	0,05
normal	yes

Ilustración 11 Prueba de Shapiro y D'Agostino Modelo Nivel

Shapiro-Wilk Test

	<i>Residuos</i>
W-stat	0,9802118
p-value	0,290078
alpha	0,05
normal	yes

d'Agostino-Pearson

DA-stat	1,8342917
p-value	0,3996581
alpha	0,05
normal	yes

Ilustración 12 Prueba de Shapiro y D'Agostino Modelo Log

Factor de Inflación de la Varianza

VIF	
price	1,00070
advert	1,00070

Ilustración 13 VIF Modelo Nivel

VIF	
ln price	1,0000010
ln advert	1,0000010

Ilustración 14 VIF Modelo Log

Al revisar los VIF de cada Beta en cada modelo, se puede concluir que ninguno de ellos tiene problema de multicolinealidad ya que el estadístico es menor a 10 en todas las variables, debido a que todas tienen un valor ínfimo.

Observaciones influyentes e Interprete los resultados

	Valor corte
Leverage	> 0,080
RStudent	> 3
T-Test	< 0,050
Cook's D	> 0,053
DFFITS	> 0,400
Cov	< 0,880 ; > 1,120

Ilustración 15 Valores de Corte

Obs	price	advert	sales	Pred Y	Residual	Leverage	RStudent	T-Test	Cook's D	DFFITS	Cov
3	5,63	0,8	62,4	75,882	-13,482	0,035	-2,956	0,004	0,095	-0,560	0,763
9	6,24	0,7	81	70,872	10,128	0,055	2,188	0,032	0,089	0,530	0,908
38	6,16	1,5	84,3	72,995	11,305	0,027	2,424	0,018	0,051	0,405	0,845
39	5,93	2,8	66	77,235	-11,235	0,034	-2,417	0,018	0,064	-0,452	0,851

Ilustración 16 Observaciones Influyentes Modelo Nivel

Luego de revisar los criterios utilizados en el modelo Nivel se deberían eliminar cuatro observaciones influyentes que se repitieron en más de 2 criterios.

Obs	price	advert	sales	Pred Y	Residual	Leverage	RStudent	T-Test	Cook's D	DFFITS	Cov
3	1,728	-0,223	4,134	4,316	-0,183	0,038	-3,168	0,002	0,116	-0,627	0,729
9	1,831	-0,357	4,394	4,251	0,143	0,063	2,456	0,016	0,126	0,636	0,871
20	1,852	-0,693	4,265	4,224	0,042	0,103	0,701	0,485	0,019	0,237	1,138
38	1,818	0,4055	4,434	4,293	0,141	0,025	2,366	0,021	0,045	0,379	0,852
39	1,78	1,0296	4,19	4,343	-0,154	0,031	-2,601	0,011	0,066	-0,464	0,819
63	1,87	-0,693	4,167	4,213	-0,047	0,110	-0,788	0,433	0,026	-0,277	1,142

Ilustración 17 Observaciones Influyentes Modelo Log

Luego de revisar los criterios utilizados en el modelo Nivel se deberían eliminar seis observaciones influyentes que se repitieron en más de 2 criterios.

3. Si existieran observaciones influyentes para cada modelo, elimínelas y ajuste los modelos nuevamente.

Regression Analysis

OVERALL FIT

Multiple R	0,75177	AIC	205,12269
R Square	0,56515	AICc	205,72875
Adjusted R Square	0,55236	SBC	211,91073
Standard Error	4,15311		
Observations	71		

ANOVA						Alpha	0,05
	df	SS	MS	F	p-value	sig	
Regression	2	1524,35094	762,17547	44,18829	5,05203E-13	yes	
Residual	68	1172,88849	17,24836				
Total	70	2697,23944					

	coeff	std err	t stat	p-value	lower	upper	vif
Intercept	121,19656	5,45756	22,20710	7,0431E-33	110,30617	132,08694	
price	-8,39011	0,94689	-8,86075	6,0238E-13	-10,27959	-6,50063	1,00205
advert	2,12932	0,60183	3,53809	0,00073159	0,92839	3,33025	1,00205

Ilustración 18 Regresión sin Observaciones Influyentes

	Modelo Original	Modelo Sin Observaciones
R ²	0,44826	0,56515
Valor F	29,24785	44,18829
β ₁	-7,90785	-8,39011
β ₂	1,86258	2,12932

Ilustración 19 Comparación entre Modelos Nivel

Realizando la comparación se puede ver que el R² aumento en un 11,689%, por lo cual el modelo sin las observaciones influyentes explica mejor el modelo. También se puede ver que el valor de F aumento de 29 a 44, que los betas en el caso del β₁ la disminución de las ventas será mayor y en el caso de β₂ las ventas aumentarán más sin las observaciones influyentes. Por lo tanto, las observaciones eliminadas si tenían una gran influencia en el modelo.

Regression Analysis

OVERALL FIT

Multiple R	0,75053	AIC	-405,40620
R Square	0,56329	AICc	-404,78120
Adjusted R Square	0,55005	SBC	-398,70388
Standard Error	0,05187		
Observations	69		

ANOVA						Alpha	0,05
	df	SS	MS	F	p-value	sig	
Regression	2	0,22906	0,11453	42,56467	1,34E-12	yes	
Residual	66	0,17759	0,00269				
Total	68	0,40664					

	coeff	std err	t stat	p-value	lower	upper	vif
Intercept	5,37927	0,11976	44,91653	3,32739E-51	5,14016	5,61838	
In price	-0,61054	0,06963	-8,76813	1,12373E-12	-0,74956	-0,47151	1,01589
In advert	0,05100	0,01292	3,94605	0,000195205	0,02519	0,07680	1,01589

Ilustración 20 Regresión sin Observaciones Influyentes

	Modelo Original	Modelo Sin Observaciones
R ²	0,4691	0,56329
Valor F	31,8101	42,5647
β_1	-0,5749	-0,6105
β_2	0,0454	0,051

Ilustración 21 Comparación entre Modelos Log

Realizando la comparación se puede ver que el R² aumento en un 9,419%, por lo cual el modelo sin las observaciones influyentes explica mejor el modelo. También se puede ver que el valor de F aumento de 31 a 42, que los betas en el caso del β_1 la disminución de las ventas en porcentaje será mayor y en el caso de β_2 las ventas aumentarán en un mayor porcentaje sin las observaciones influyentes.

Por lo tanto, las observaciones eliminadas si tenían una gran influencia en el modelo.

II. DATA SET AZUCAR

Use el set de datos denominado azúcar. Utilizando un $\alpha=0.05$, seleccione las variables que deberían incluirse en el modelo para explicar el rendimiento de azúcar_2009 (TM_2009), en función del rendimiento de azúcar_2008 (TM_2008) y el resto de las variables del set de datos, incluidas longitud (X) y latitud (Y). Utilice los procedimientos: forward, Backward y Stepwise. Compare los modelos obtenidos en cada caso.

```
. regress tm_2009 x y tm_2008 ph mo ntotal c_n p__disp ca__disp mg_disp k_disp na
> _disp cic
```

Source	SS	df	MS	Number of obs	=	116
Model	2853.44715	13	219.495935	F(13, 102)	=	13.38
Residual	1673.19394	102	16.4038622	Prob > F	=	0.0000
				R-squared	=	0.6304
				Adj R-squared	=	0.5833
Total	4526.64109	115	39.3620965	Root MSE	=	4.0502

tm_2009	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	.000154	.0002847	0.54	0.590	-.0004108	.0007187
y	.000321	.000162	1.98	0.050	-2.37e-07	.0006422
tm_2008	.8040304	.0761438	10.56	0.000	.6529996	.9550612
ph	1.780168	1.102421	1.61	0.109	-.4064782	3.966815
mo	3.519469	2.555962	1.38	0.172	-1.55027	8.589207
ntotal	-98.08027	49.49606	-1.98	0.050	-196.2555	.0949351
c_n	.2342664	.406922	0.58	0.566	-.5728616	1.041394
p__disp	.0690868	.0383456	1.80	0.075	-.0069715	.1451451
ca__disp	.7388254	.377166	1.96	0.053	-.0092816	1.486932
mg_disp	-.698306	1.083999	-0.64	0.521	-2.848413	1.451801
k_disp	1.684827	2.669662	0.63	0.529	-3.610435	6.980088
na_disp	5.480707	8.831116	0.62	0.536	-12.03577	22.99718
cic	-.2688905	.2207388	-1.22	0.226	-.7067248	.1689438
_cons	-631.6016	325.3358	-1.94	0.055	-1276.904	13.70042

Ilustración 22 Regresión Múltiple Stata

En el procedimiento de Backward nos da 8 variables a remover

```
. stepwise,pr(.05): regress tm_2009 x y tm_2008 ph mo ntotal c_n p__disp ca__disp
> mg_disp k_disp na_disp cic
begin with full model
p = 0.5898 >= 0.0500 removing x
p = 0.6057 >= 0.0500 removing c_n
p = 0.5271 >= 0.0500 removing na_disp
p = 0.5290 >= 0.0500 removing mg_disp
p = 0.5597 >= 0.0500 removing k_disp
p = 0.0737 >= 0.0500 removing cic
p = 0.1163 >= 0.0500 removing y
p = 0.0876 >= 0.0500 removing ph
```

Source	SS	df	MS	Number of obs	=	116
Model	2684.48555	5	536.89711	F(5, 110)	=	32.06
Residual	1842.15554	110	16.7468686	Prob > F	=	0.0000
				R-squared	=	0.5930
				Adj R-squared	=	0.5745
Total	4526.64109	115	39.3620965	Root MSE	=	4.0923

tm_2009	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
p__disp	.0983904	.0324978	3.03	0.003	.0339872	.1627935
ca__disp	.5437067	.2238836	2.43	0.017	.1000219	.9873915
tm_2008	.8540111	.0717129	11.91	0.000	.7118929	.9961293
ntotal	-132.3441	34.48641	-3.84	0.000	-200.688	-64.00012
mo	4.565516	1.441845	3.17	0.002	1.708118	7.422914
_cons	4.565192	2.622436	1.74	0.085	-.6318608	9.762244

Ilustración 23 Procedimiento Backward Stata

En el procedimiento de Forward nos da 4 variables a añadir

```
. stepwise,pe(.05): regress tm_2009 x y tm_2008 ph mo ntotal c_n p__disp ca__disp
> mg_disp k_disp na_disp cic
begin with empty model
p = 0.0000 < 0.0500 adding tm_2008
p = 0.0042 < 0.0500 adding c_n
p = 0.0492 < 0.0500 adding cic
p = 0.0225 < 0.0500 adding ca__disp
```

Source	SS	df	MS	Number of obs	=	116
Model	2614.80554	4	653.701385	F(4, 111)	=	37.95
Residual	1911.83555	111	17.2237437	Prob > F	=	0.0000
				R-squared	=	0.5776
				Adj R-squared	=	0.5624
Total	4526.64109	115	39.3620965	Root MSE	=	4.1501

tm_2009	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tm_2008	.8147823	.072287	11.27	0.000	.6715407	.9580238
c_n	.7857547	.2298856	3.42	0.001	.330221	1.241288
cic	-.4655003	.171208	-2.72	0.008	-.8047603	-.1262403
ca__disp	.5432744	.2347573	2.31	0.022	.0780871	1.008462
_cons	-2.700738	3.386396	-0.80	0.427	-9.411108	4.009631

Ilustración 24 Procedimiento Forward Stata

En el procedimiento de Stepwise nos da 8 variables a remover

```
. stepwise,pr(.05)pe(.0499): regress tm_2009 x y tm_2008 ph mo ntotal c_n p_disp
> ca_disp mg_disp k_disp na_disp cic
      begin with full model
p = 0.5898 >= 0.0500 removing x
p = 0.6057 >= 0.0500 removing c_n
p = 0.5271 >= 0.0500 removing na_disp
p = 0.5290 >= 0.0500 removing mg_disp
p = 0.5597 >= 0.0500 removing k_disp
p = 0.0737 >= 0.0500 removing cic
p = 0.1163 >= 0.0500 removing y
p = 0.0876 >= 0.0500 removing ph
```

Source	SS	df	MS	Number of obs	=	116
Model	2684.48555	5	536.89711	F(5, 110)	=	32.06
Residual	1842.15554	110	16.7468686	Prob > F	=	0.0000
				R-squared	=	0.5930
				Adj R-squared	=	0.5745
Total	4526.64109	115	39.3620965	Root MSE	=	4.0923

tm_2009	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
p_disp	.0983904	.0324978	3.03	0.003	.0339872	.1627935
ca_disp	.5437067	.2238836	2.43	0.017	.1000219	.9873915
tm_2008	.8540111	.0717129	11.91	0.000	.7118929	.9961293
ntotal	-132.3441	34.48641	-3.84	0.000	-200.688	-64.00012
mo	4.565516	1.441845	3.17	0.002	1.708118	7.422914
_cons	4.565192	2.622436	1.74	0.085	-.6318608	9.762244

Ilustración 25 Procedimiento Stepwise Stata

Al comparar los diferentes procedimientos se puede deducir que Backward y Stepwise remueven 8 variables, por lo que se dejan solo las variables que son significativas localmente para el modelo.

III. DATA SET DE ENTRENAMIENTO

Utilice el set de datos denominado entrenamiento para probar que el subsidio a la capacitación no tiene efecto sobre un mayor entrenamiento en las empresas.

Para efectuar los análisis se tuvieron que eliminar algunas observaciones que tenían como dato un punto, para poder correr el modelo.

1. Establezca el modelo econométrico. Transforme las variables que considere necesarias.

$$\text{Ventas anuales} = 3.012.831,536 + 192,634 * \text{salario_prom} - 734.315,008 * \text{subsidio} + 26.790,904 * \text{total_emp_entr} + 88.458,476 * \text{n}^\circ \text{ empleados}$$

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0,806517422
Coefficiente de determinación R^2	0,650470351
R^2 ajustado	0,646309284
Error típico	4878776,503
Observaciones	341

ANÁLISIS DE VARIANZA

	Grados de libertad		Media de cuadrados		F	Valor crítico de F
Regresión	4	1,48835E+16	3,72087E+15	156,3229607	2,22745E-75	
Residuos	336	7,99763E+15	2,38025E+13			
Total	340	2,28811E+16				

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	-3012831,536	821033,9961	-3,669557595	0,000282461	-4627845,953	-1397817,12	-4627845,953	-1397817,12
salario_prom	192,6337305	38,46239816	5,008365045	8,88481E-07	116,9762939	268,2911671	116,9762939	268,2911671
Subsidio	-734315,0079	803162,7458	-0,914279219	0,361225768	-2314175,792	845545,7765	-2314175,792	845545,7765
total_emp_entr	26790,90396	8172,688161	3,278101822	0,00115436	10714,82274	42866,98517	10714,82274	42866,98517
n°empleados	88458,47612	6008,829045	14,72141668	3,34114E-38	76638,81265	100278,1396	76638,81265	100278,1396

Ilustración 26 Regresión

Sub modelos				
	Bo	Bsalario prom	Btotal empl entren	Bn°empleados
Sin subsidio	-3.012.831,54	192,63	26.790,90	88.458,48
Con subsidio	-3.747.146,54	192,63	26.790,90	88.458,48

Ilustración 27 Submodelos

2. Establezca la prueba de hipótesis respectiva

$$H_0: \beta_1 = 0 \quad H_a: \beta_1 \neq 0$$

Ilustración 28 Test de hipótesis

	Probabilidad
Intercepción	0,000282461
salario_prom	8,88481E-07
Subsidio	0,361225768
total_emp_entr	0,00115436
n°empleados	3,34114E-38

Ilustración 29 Significancia Local

No hay antecedentes para rechazar la Hipótesis nula, ya que el p value es mayor al alpha, por lo cual es una variable no significativa localmente.

3. Ajuste el modelo

Resumen

Estadísticas de la regresión	
Coficiente de correlación múltiple	0,8060
Coficiente de determinación R ²	0,6496
R ² ajustado	0,6465
Error típico	4877588,5604
Observaciones	341

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	Valor crítico de F
Regresión	3	1,48636E+16	4,95453E+15	208,2534	2,15089E-76
Residuos	337	8,01752E+15	2,37909E+13		
Total	340	2,28811E+16			

	Coficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%
Intercepción	-3104558,7631	814682,7046	-3,8108	0,0002	-4707062,6861	-1502054,8401
salario_prom	192,1545	38,4495	4,9976	9,34688E-07	116,5233	267,7856
total_emp_entr	24799,8466	7875,2835	3,1491	0,0018	9308,9411	40290,7520
n°empleados	89358,0115	5926,2897	15,0782	1,26658E-39	77700,8321	101015,1909

Ilustración 30 Regresión con modelo ajustado

4. Realice un diagnóstico de la regresión (heterocedasticidad, colinealidad, observaciones influyentes). Seleccione las herramientas necesarias.

VIF
1,0011
1,9842
1,9849

Ilustración 31 Cálculo VIF

Al revisar los VIF de cada Beta en el modelo, se puede concluir que ninguno de ellos tiene problema de multicolinealidad ya que el estadístico es menor a 10 en todas las variables, debido a que todas tienen un valor ínfimo.

Criterio	Mayor a
Leverage	0,0235
Rstudent	> 3
Cook's D	0,0117
DFFITS	0,2166

Ilustración 32 Valor corte

Luego de revisar los criterios utilizados en el modelo se deberían eliminar 46 observaciones influyentes que se repitieron en más de 2 criterios.

Regression Analysis

OVERALL FIT

Multiple R	0,909521047	AIC	8459,683904
R Square	0,827228535	AICc	8459,891516
Adjusted R Squ	0,825447386	SBC	8474,431805
Standard Error	1675653,869		
Observations	295		

ANOVA				Alpha	0,05	
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p-value</i>	<i>sig</i>
Regression	3	3,91215E+15	1,30405E+15	464,4353038	1,4006E-110	yes
Residual	291	8,17074E+14	2,80782E+12			
Total	294	4,72922E+15				

	<i>coeff</i>	<i>std err</i>	<i>t stat</i>	<i>p-value</i>	<i>lower</i>	<i>upper</i>	<i>vif</i>
Intercept	-1001089,989	327087,4065	-3,060619177	0,0024	-1644846,922	-357333,0557	
salario_prom	57,12727363	16,07580626	3,553617946	0,0004	25,4876827	88,76686455	1,007478306
total_emp_eni	18821,78779	4794,749491	3,92549972	0,0001	9385,003712	28258,57186	1,618862048
n°empleados	90327,17872	3390,430142	26,641805	4,90251E-80	83654,30515	97000,0523	1,610184078

Ilustración 33 Modelo sin observaciones Influyentes

5. Establezca sus conclusiones.

El modelo econométrico permite identificar que la variable binaria subsidio no tiene gran relevancia para identificar si la empresa tiene mayor entrenamiento.

Al correr el modelo sin las observaciones influyentes se puede ver que el R^2 aumento a un 82,72%, por lo cual se puede explicar mejor el modelo.