



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

PROYECTO FINAL

MÉTODOS ESTADÍSTICOS

Objetivo:

Encontrar el mejor modelo con
variables que sospechamos predictivas
para el PIB de Baja California

MET. Alejandra Cerda

FCFM , Lic. Actuarial

Nombres:

Ana Sofia Saucedo Garcia - 1946646

Natalia Jireh Valles Campusano - 1976629

Brayan Rangel Bazaldua - 1814466

Jazmin Alessandra Gonzalez Rodriguez - 1989943

Jorge Sebastian Aguirre Jimenez - 1947870

Índice

1	Introducción-----	3
1.1.	Descripción de datos-----	3
1.2.	Obtención de datos-----	5
2	Análisis de los datos-----	6
2.1.	Análisis exploratorio de los datos-----	6
2.2.	Prueba de bondad de ajuste por variable-----	7
2.3.	Multicolinealidad-----	8
2.4.	Tabla de Ajuste modelos lineales-----	9
2.5.	Tabla de Ajuste modelos No lineales-----	10
2.6.	Analisis de la existencia de datos atípicos, palanca y de influencia.---	11
3	Conclusiones-----	12
4	Referencias-----	13

1 Introducción

En este proyecto se busca encontrar el modelo que mediante variables que sospechamos predictivas para el PIB de Baja California , sea el que mejor predice el Producto Interno Bruto de Baja California.

1.1. Descripción de datos

A continuación tenemos las variables que sospechamos predictivas ,es decir, las actividades económicas que sospechamos que predicen el PIB de Baja California son las siguientes:

De las Actividades primarias tenemos:

Agricultura, ganadería, aprovechamiento forestal, pesca y caza. Este sector incluye unidades económicas dedicadas principalmente a la siembra, cultivo, cosecha y recolección de vegetales; a la explotación de animales en ambientes controlados; al aprovechamiento y recolección de recursos forestales; a la pesca, caza y captura de animales en su hábitat natural.

De las Actividades secundarias tenemos:

Construcción. Este sector comprende unidades económicas dedicadas principalmente a la edificación residencial, ya sea vivienda unifamiliar o multifamiliar; a la edificación no residencial, como naves y plantas industriales, inmuebles comerciales, institucionales y de servicios; a la construcción de obras de ingeniería civil, como puentes, carreteras, presas, vías férreas, centrales eléctricas y puertos; Puede tratarse de construcción nueva, ampliación, remodelación, mantenimiento o reparación integral de las construcciones.

Industrias manufactureras. Este sector comprende unidades económicas dedicadas principalmente a la transformación mecánica, física o química de materiales o sustancias con el fin de obtener productos nuevos. También se consideran manufacturas las actividades de maquila; el ensamble de partes y componentes o productos fabricados; la reconstrucción de maquinaria y equipo industrial, comercial, de oficina y otros; y el acabado de productos manufacturados mediante el teñido, tratamiento calorífico, enchapado y procesos similares.

Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final. Este sector comprende unidades económicas dedicadas principalmente a la generación, transmisión y suministro de energía eléctrica para su venta; a la captación, potabilización y suministro de agua, así como a la captación y tratamiento de aguas residuales; y al suministro de gas por ductos al consumidor final.

De las Actividades Terciarias tenemos:

Transportes, correos y almacenamiento. Este sector comprende unidades económicas dedicadas principalmente al transporte (de personas y de carga); a proporcionar servicios especializados relacionados directamente con el transporte; servicios de correo, y al

almacenamiento de bienes. En virtud de que las actividades de las oficinas postales y de los establecimientos de mensajería consisten en transportar bienes, fueron incluidas en este sector.

Siendo estas variables nuestras predictoras del PIB estatal (en este caso de Baja California) , donde :

Producto Interno Bruto por Entidad Federativa (PIBE): Es un cálculo anual que tiene como propósito, contribuir al conocimiento del desempeño económico de las entidades federativas; sus resultados, permiten conocer la estructura económica de cada entidad, su contribución al producto nacional, y posibilita la evaluación de la dinámica que presentan los 32 estados, así como la comparación entre los mismos.

En la *Tabla 1* observamos la relación de las variables , el tipo de variable y escala de medición de las variables con las que buscaremos encontrar el mejor modelo para predecir nuestro PIB estatal.

Variable	Tipo de relación	Tipo de Variable	Escala de medición
PIB estatal	Dependiente	Cuantitativa	De intervalo continua
Año	Independiente	Cuantitativa	De intervalo discreta
Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza	Independiente	Cuantitativa	De intervalo continua
Transportes, correos y almacenamiento	Independiente	Cuantitativa	De intervalo continua
Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final	Independiente	Cuantitativa	De intervalo continua
Construcción	Independiente	Cuantitativa	De intervalo continua
Industrias manufactureras	Independiente	Cuantitativa	De intervalo continua

Tabla 1. Tipo de variable, relación y escala de medición.

1.2. Obtención de datos

Los datos con los que se trabajaran para obtener el mejor modelo de nuestros datos es decir las variables que sospechamos predictoras del Producto Interno Bruto de Baja California , se obtuvieron del sitio web de INEGI de la pagina llamada *PIB por Entidad Federativa (PIBE)*. Base 2013. En donde en el apartado de *Tabulados* se encuentra una pestaña llamada *Serie detallada* en la cual como nos interesa saber el PIB estatal de Baja California y sus actividades economicas , entonces descargamos la base de datos en el apartado de *PIB de las entidades federativas por actividad económica* , de Baja California. Teniendo en excel la base de datos seleccionamos unicamente los datos de *Valores constantes en Millones de pesos* de los años del 2008 al 2020, de esta manera ya tendremos menos datos y con ayuda de un filtro seleccionamos las actividades economicas que observamos que más aportan al PIB estatal.

Obteniendo asi nuestros datos que se observan en la *Tabla 2* , siendo entonces Agricultura, ganadería, aprovechamiento forestal, pesca y caza, Transportes, correos y almacenamiento, Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final, Construcción e Industrias Manufactureras nuestras variables predictoras del PIB estatal de Baja California.

PIB estatal (millones de pesos)	Año	Actividades Primarias	Actividades secundarias			Actividades terciarias
		Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza	Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final	Construcción	Industrias Manufactureras	Transportes, correos y almacenamiento
457,557	2008	11,634	14,115	68,838	111,219	24,905
407,746	2009	13,218	14,502	52,431	91,268	21,252
428,163	2010	13,323	14,377	53,597	97,120	22,050
440,701	2011	13,184	14,908	54,618	94,819	22,815
456,024	2012	13,220	14,547	55,068	99,629	22,831
465,525	2013	14,182	15,155	54,082	100,696	22,879
478,122	2014	14,087	16,996	50,477	110,058	23,399
511,460	2015	14,861	16,876	55,325	123,651	24,344
535,553	2016	13,841	16,571	57,086	129,408	26,003
553,327	2017	15,073	15,392	60,685	132,550	26,883
566,046	2018	15,613	15,660	59,209	140,237	27,963
575,332	2019	15,460	16,180	57,732	146,764	27,969
554,008	2020	14,543	16,869	53,385	144,085	22,230

Tabla 2. PIB de Baja California (2008-2020) por actividad economica.

Donde para el trabajo de la obtencion del mejor modelo de nuestros datos , algunas de nuestras variables predictoras seran abreviadas como:

PIBE = PIB estatal de Baja California (en millones de pesos)

AGFPC = Agricultura, ganadería, aprovechamiento forestal, pesca y caza

EAG = Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final

IM= Industrias Manufactureras

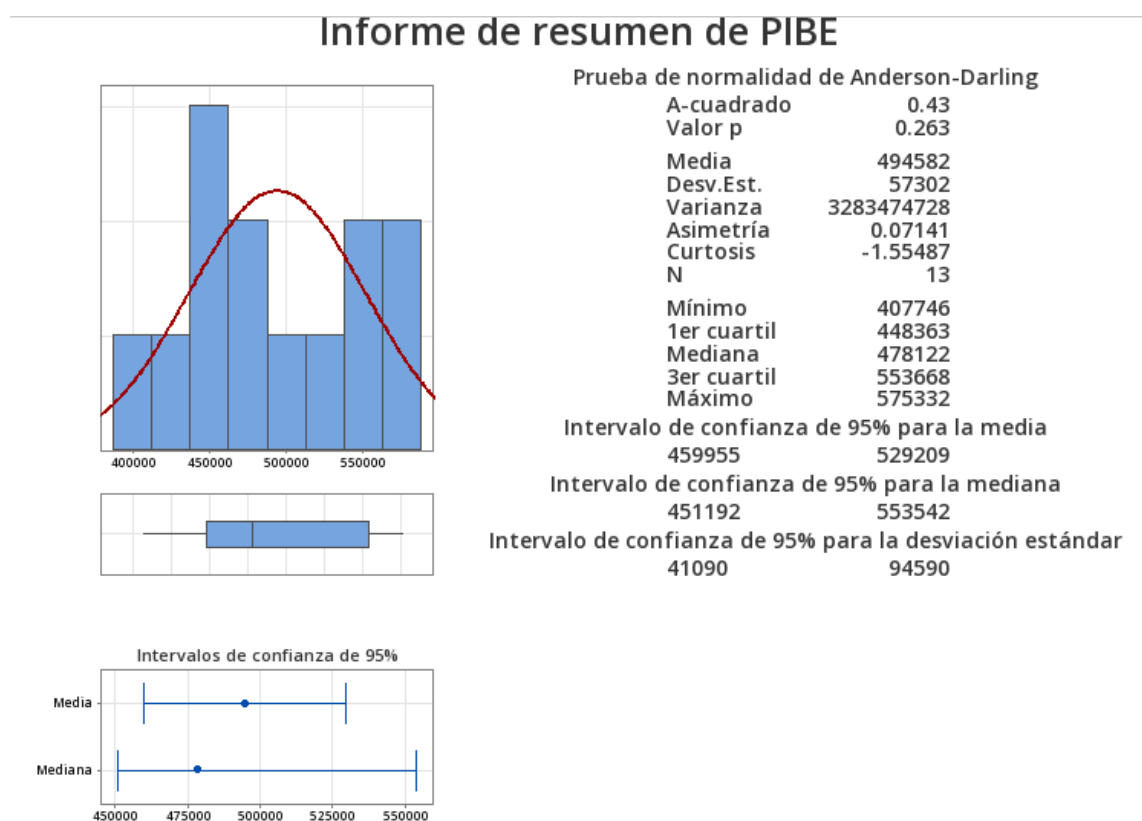
TCA = Transportes, correos y almacenamiento

2 Análisis de los datos

2.1. Análisis exploratorio de los datos

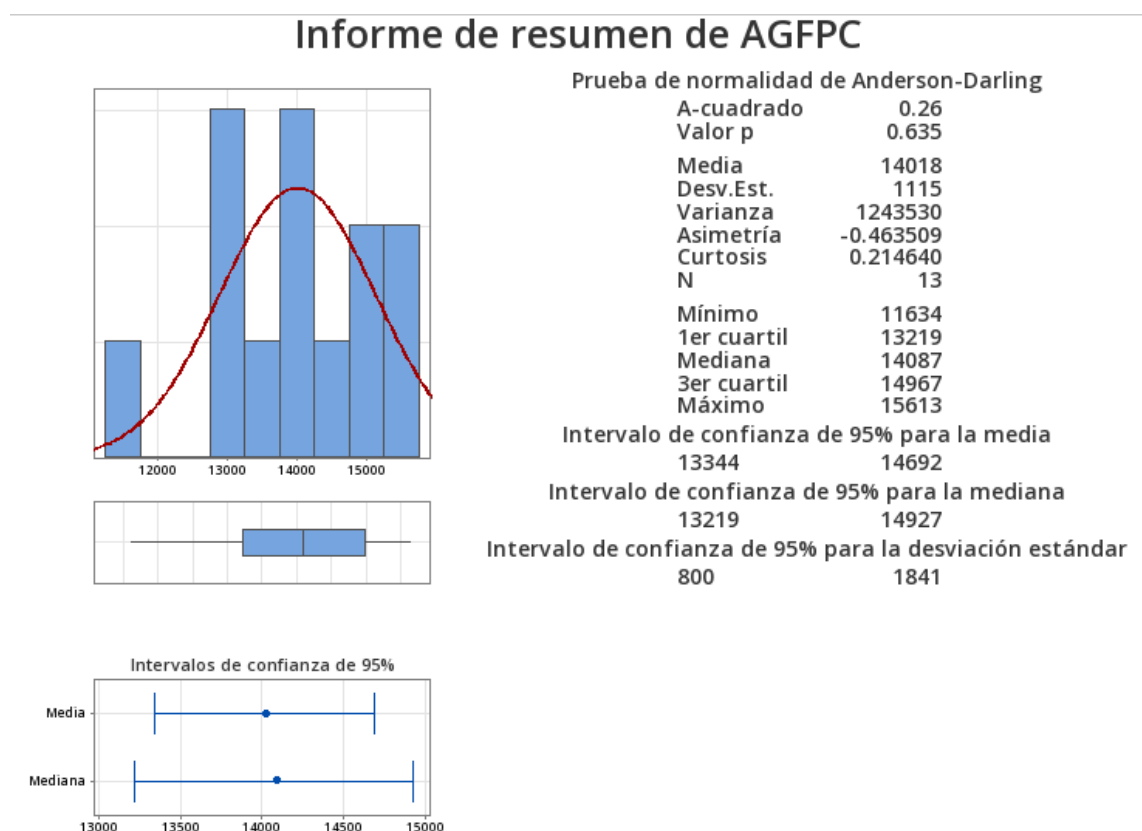
El análisis de nuestros datos conlleva el analizar las estadísticas de nuestros datos esto para entender más sobre como se comportan nuestras variables, es decir para describir y representar los datos de manera mas sencilla.

En la *Grafica 1* observamos las estadísticas del PIB de Baja California donde observamos que tiene una media de 494,582, teniendo una desviación estándar de 57,302. El histograma nos muestra que nuestros valores con mayor frecuencia se encontraron en 450,000. Siendo este estado de los que mas aportan al PIB nacional en los ultimos años según lo encontrado en las paginas del inegi, llegando a aportar hasta 575,332 (millones de pesos) observando que es el maximo valor o PIB de Baja California que se encontro en los años del 2008-2020.



Grafica 1. Resumen estadístico descriptivo y gráfico del PIB de Baja California (2008-2020).

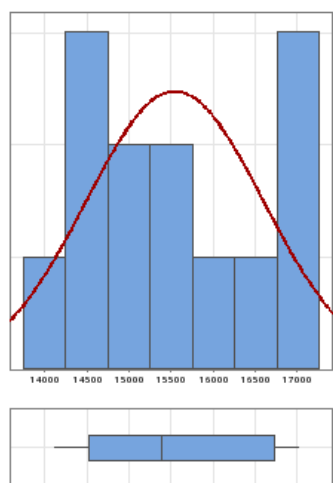
En la *Grafica 2* observamos las estadísticas de la aportación al PIB en millones de pesos de la actividad económica “Agricultura, ganadería, aprovechamiento forestal, pesca y caza de Baja California” (AGFPC), en donde muestra una media de 14,018 ; teniendo una desviación estándar de 1,115 ; El histograma nos muestra que la aportacion de esta actividad con mayor frecuencia se encuentra en 13,000 y 14,000. Siendo la agricultura una de las actividades económicas que caracterizan al estado de Baja California por la producción de uva para vino, puesto que es uno de los estados con mayor producción de uvas a nivel nacional.



Grafica 2. Resumen estadístico descriptivo y gráfico de la aportacion al PIB de la actividad economica Agricultura, ganadería, aprovechamiento forestal, pesca y caza (AGFPC) de Baja California. (2008-2020).

En la *Grafica 3* observamos las estadísticas de la aportación al PIB en millones de pesos de la actividad económica “Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final.” (EAG), en donde muestra una media de 15,550 ; teniendo una desviación estándar de 1,048 ; El histograma nos muestra que la aportacion de esta actividad con mayor frecuencia se encuentra en 14,500 y 17,000.

Informe de resumen de EAG



Prueba de normalidad de Anderson-Darling

A-cuadrado 0.44
Valor p 0.252

Media 15550
Desv.Est. 1048
Varianza 1097748
Asimetría 0.16378
Curtosis -1.60649
N 13

Mínimo 14115
1er cuartil 14525
Mediana 15392
3er cuartil 16720
Máximo 16996

Intervalo de confianza de 95% para la media

14917 16183

Intervalo de confianza de 95% para la mediana

14533 16665

Intervalo de confianza de 95% para la desviación estándar

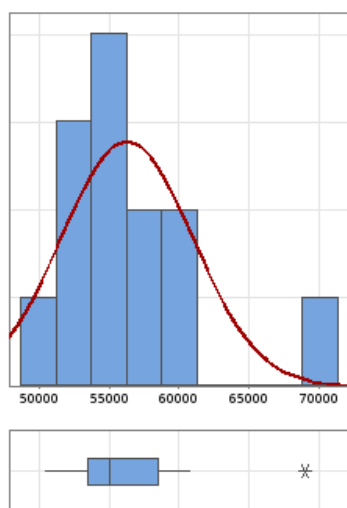
751 1730



Grafica 3. Resumen estadístico descriptivo y gráfico de la aportacion al PIB de la actividad economica Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final (EAG) de Baja California. (2008-2020).

En la *Grafica 4* observamos las estadísticas de la aportación al PIB de Baja California en millones de pesos de la actividad económica “Construcción”, en donde muestra una media de 56,349 ; teniendo una desviación estándar de 4,676; El histograma nos muestra que la aportacion de esta actividad con mayor frecuencia se encuentra en 55,000. Siendo esta actividad economica de las que mas aportan al PIB de Baja california llegando a aportar hasta 68,830 (millones de pesos) observando que es el maximo valor o aportacion de la actividad economica “Construcción” que se encontro en los años del 2008-2020.

Informe de resumen de Construcción



Prueba de normalidad de Anderson-Darling

A-cuadrado	0.65
Valor p	0.070
Media	56349
Desv.Est.	4676
Varianza	21869002
Asimetría	1.66287
Curtosis	3.67053
N	13

Mínimo	50477
1er cuartil	53491
Mediana	55068
3er cuartil	58470
Máximo	68838

Intervalo de confianza de 95% para la media

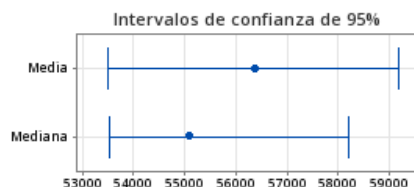
53523	59175
-------	-------

Intervalo de confianza de 95% para la mediana

53530	58198
-------	-------

Intervalo de confianza de 95% para la desviación estándar

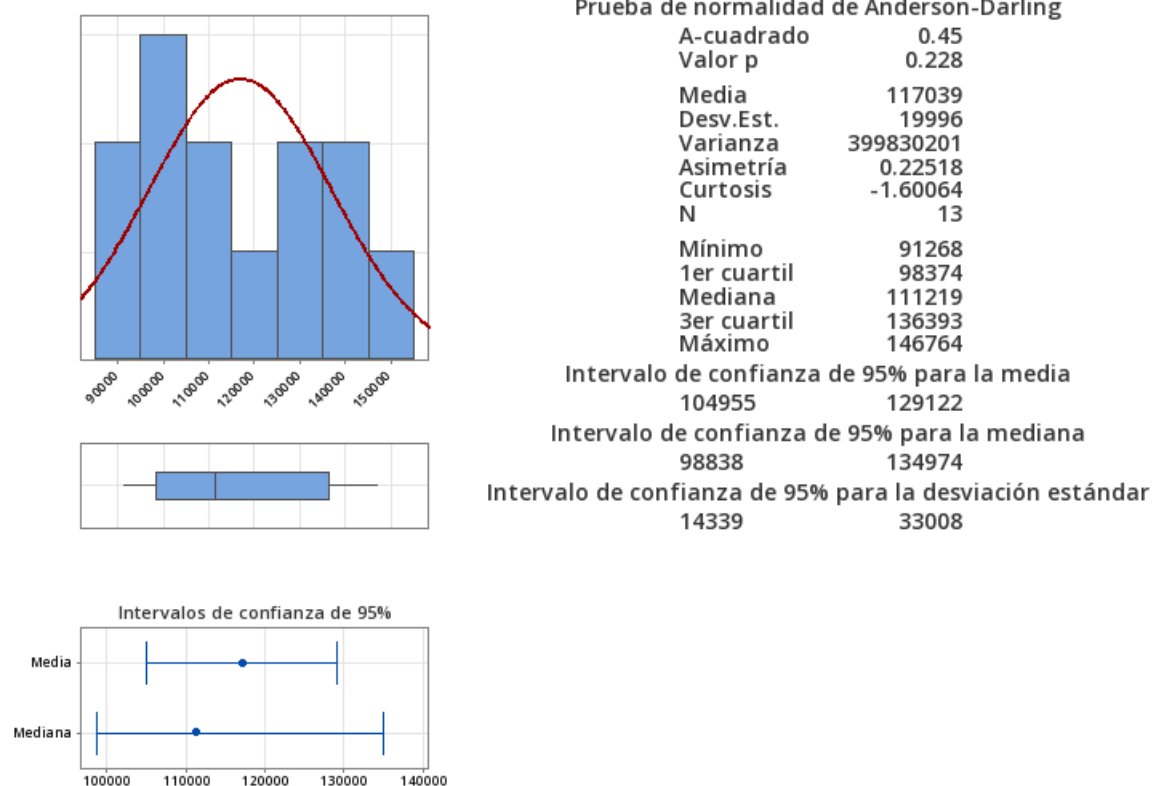
3353	7720
------	------



Grafica 4. Resumen estadístico descriptivo y gráfico de la aportacion al PIB de la actividad economica Construcción de Baja California. (2008-2020).

En la *Grafica 5* observamos las estadísticas de la aportación al PIB de Baja California en millones de pesos de la actividad económica “Industrias Manufactureras”, en donde muestra una media de 117,039 ; teniendo una desviación estándar de 19,996; El histograma nos muestra que la aportacion de esta actividad con mayor frecuencia se encuentra en 100,000. Siendo esta actividad economica de las que mas aportan al PIB de Baja california llegando a aportar hasta 146,764 (millones de pesos) observando que es el maximo valor o aportacion de la actividad economica “Industrias Manufactureras” que se encontro en los años del 2008-2020.

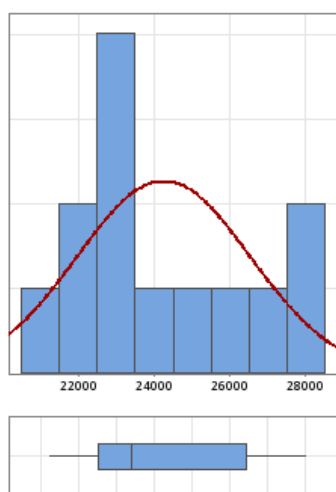
Informe de resumen de IM



Grafica 5. Resumen estadístico descriptivo y gráfico de la aportacion al PIB de la actividad economica Industrias Manufactureras (IM) de Baja California. (2008-2020).

En la *Grafica 6* observamos las estadísticas de la aportación al PIB de Baja California en millones de pesos de la actividad económica “Transportes, Correos y Almacenamiento”(TCA) , en donde muestra una media de 24,271 ; teniendo una desviación estándar de 2,285; El histograma nos muestra que la aportacion de esta actividad con mayor frecuencia se encuentra en 23,000 (millones de pesos).Y llegando a aportar un máximo de 27,969 (millones de pesos).

Informe de resumen de TCA



Prueba de normalidad de Anderson-Darling

A-cuadrado	0.49
Valor p	0.178
Media	24271
Desv.Est.	2285
Varianza	5219731
Asimetría	0.55954
Curtosis	-1.06147
N	13

Mínimo	21252
1er cuartil	22522
Mediana	23399
3er cuartil	26443
Máximo	27969

Intervalo de confianza de 95% para la media

22890	25652
-------	-------

Intervalo de confianza de 95% para la mediana

22630	26281
-------	-------

Intervalo de confianza de 95% para la desviación estándar

1638	3771
------	------

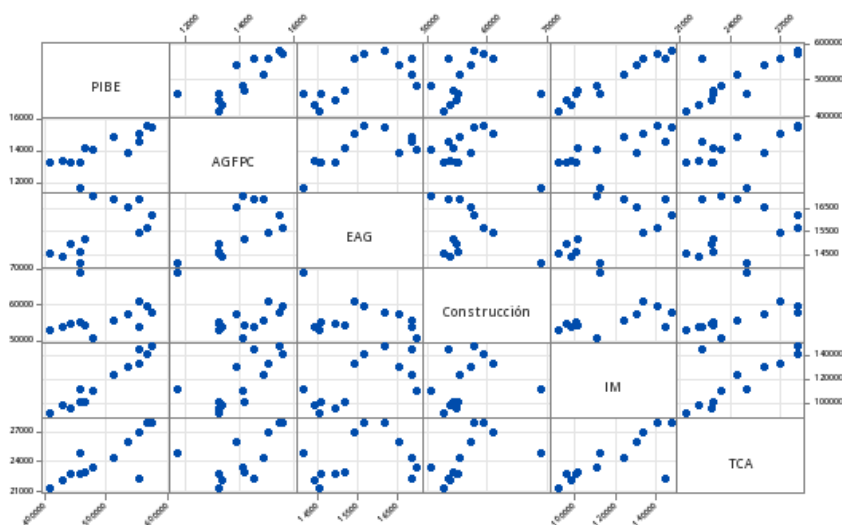


Grafica 6. Resumen estadístico descriptivo y gráfico de la aportacion al PIB de la actividad economica Transportes, Correos y Almacenamiento (TCA) de Baja California. (2008-2020).

Continuaremos con el analisis de tablas y gráficas para encontrar relaciones y patrones en nuestros datos.

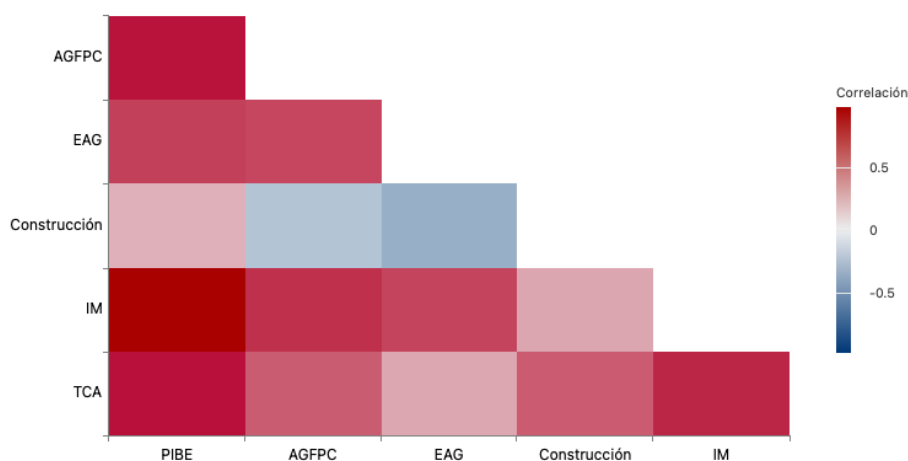
En las *Grafica 7 y 8* observamos la dispersion de los datos que nos ayuda para encontrar alguna relacion o patron con la cual el PIB de Baja california tiene mayor relación donde observando nuestras gráficas podemos ver que se muestra un patrón con la actividad economica Industrias Manufactureras (IM) , con Agricultura, ganadería, aprovechamiento forestal, pesca y caza (AGFPC) y con Transportes, correos y almacenamiento (TCA) mostrando como un modelo exponencial.

Gráfica de matriz de PIBE, AGFPC, EAG, Construcción, IM, TCA



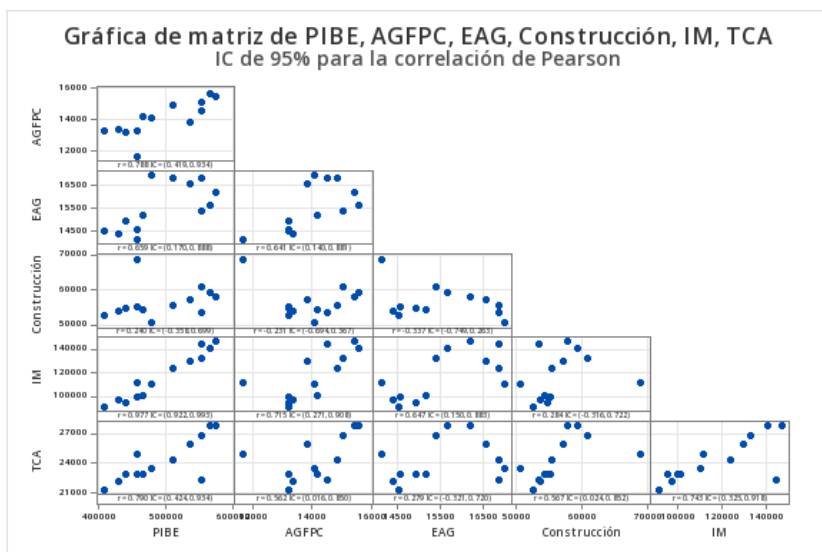
Grafica 7. Matriz de Dispersion de nuestras variables.

Correlograma de PIBE,AGFPC,EAG,Construcción,IM,TCA



Grafica 8. Correlograma de nuestras variables.

La *Grafica 9* y la *Tabla 3* tambien nos ayuda para observar de manera mas visible la relacion de las variables mostrando tambien lo mencionado anteriormente que la actividades economica que mayor relacion tienen con la respuesta es Industrias Manufactureras (IM) con un coeficiente de correlación de 0.977 que representa una relacion positiva entre las variables. Es decir a medida que el PIB de Baja California aumenta , Industrias Manufactureras tambien aumenta.



Grafica 9. Grafica de dispersion , con IC de 95% para la correlación.

Correlaciones

	PIBE	AGFPC	EAG	Construcción	IM
AGFPC	0.788				
EAG	0.659	0.641			
Construcción	0.240	-0.231	-0.337		
IM	0.977	0.715	0.647	0.284	
TCA	0.790	0.562	0.279	0.567	0.743

Tabla 3. Correlaciones de nuestras variables.

Por lo que con lo analizado anteriormente podemos sospechar que el mejor modelo para nuestros datos seria un modelo que incluya las variables o actividades economicas : Industrias Manufactureras (IM) , Agricultura, ganadería, aprovechamiento forestal, pesca y caza (AGFPC) y Transportes, correos y almacenamiento (TCA).

2.2. Prueba de bondad de ajuste por variable

Prueba de bondad de ajuste de la variable PIBE

HIPÓTESIS
H0: Los valores de la variable provienen de una distribución normal
H1: Los valores de la variable provienen de una distribución normal

PIBE
457556.57
407745.95
428162.55
440700.66
456024.47
465524.70
478121.94
511459.53
535552.65
553327.21
566045.91
575332.46
554008.04

Máx	575332.46
Mín	407745.95
Rango	167586.52
n	13
Clases	3.60555128
Ancho de clase	41896.6288
Media est	494581.74
Varianza est	3030899749

4
41896.6288

Prueba de bondad de ajuste						
Clase	LI	LS	Frec Obs	Probabilidad	Frec. Esp	Cociente
1	407745.95	449642.58	3	0.207169977	2.69320969	0.03494726
2	449642.58	491539.21	4	0.270793708955136	3.52031822	0.06536188
3	491539.21	533435.83	1	0.281864555176162	3.66423922	1.93714716
4	533435.834	575332.463	5	0.240171759	3.12223287	1.129323
			13	1	13	3.1667793

Región de Rechazo
Rechazar H0 si $\chi^2 > \chi^2_{\alpha}$

Estadístico de prueba	
χ^2	3.1668
Valor crítico	
α	0.05
χ^2_{α}	3.84145882

Conclusión

No se rechaza H0, con 95% de confianza se afirma que los datos provienen de una distribución normal.

Prueba de bondad de ajuste de la variable TCA

HIPÓTESIS
H0: Los valores de la variable provienen de una distribución normal
H1: Los valores de la variable provienen de una distribución normal

TCA
24904.87
21252.32
22050.09
22814.86
22830.64
22878.77
23399.41
24344.05
26003.34
26882.74
27962.59
27969.03
22229.84

Máx	27969.03
Mín	21252.32
Rango	6716.71
n	13
Clases	3.60555128
Ancho de clase	1679.1775
Media est	24270.97
Varianza est	4818212.83

4
1679.1875

Prueba de bondad de ajuste						
Clase	LI	LS	Frec Obs	Probabilidad	Frec. Esp	Cociente
1	21252.32	22931.51	6	0.270857909	3.521152821	1.74507715
2	22931.51	24610.69	2	0.290640920781751	3.77833197	0.8370002
3	24610.69	26289.88	2	0.259652379746422	3.375480937	0.56049726
4	26289.8805	27969.068	3	0.17884879	2.325034272	0.19594495
			13	1	13	3.33851956

Región de Rechazo
Rechazar H0 si $X^2 > X^2_{\alpha}$

Estadístico de prueba	
X^2	3.3385
Valor crítico	
α	0.05
X^2_{α}	3.84145882

Conclusión

No se rechaza H0, con 95% de confianza se afirma que los datos provienen de una distribución normal.

Prueba de bondad de ajuste de la variable EAG

HIPÓTESIS	
H0:	Los valores de la variable provienen de una distribución normal
H1:	Los valores de la variable provienen de una distribución normal

EAG
14114.78
14502.10
14377.47
14907.91
14547.45
15154.99
16996.04
16876.33
16571.15
15392.24
15659.80
16180.20
16869.01

Máx	16996.04
Mín	14114.78
Rango	2881.26
n	13
Clases	3.60555128
Ancho de clase	720.31375
Media est	15549.96
Varianza est	1013306.21

4
720.32375

Prueba de bondad de ajuste						
Clase	LI	LS	Frec Obs	Probabilidad	Frec. Esp	Cociente
1	14114.78	14835.10	4	0.23880672	3.104487365	0.25831733
2	14835.10	15555.43	3	0.263360483101239	3.42368628	0.05243181
3	15555.43	16275.75	2	0.262381442320087	3.41095875	0.58364957
4	16275.7513	16996.075	4	0.235451354	3.060867605	0.28814368
			13	1	13	1.18254238

Región de Rechazo
Rechazar H0 si $\chi^2 > \chi^2_{\alpha}$

Estadístico de prueba	
χ^2	1.1825
Valor crítico	
alfa	0.05
χ^2_{α}	3.84145882

Conclusión

No se rechaza H0, con 95% de confianza se afirma que los datos provienen de una distribución normal.

Prueba de bondad de ajuste de la variable Construcción

HIPÓTESIS
H0: Los valores de la variable provienen de una distribución normal
H1: Los valores de la variable provienen de una distribución normal

CONSTRUCCION
68838.23
52431.46
53597.43
54617.68
55067.52
54081.59
50477.08
55325.46
57086.24
60685.05
59208.88
57732.09
53385.19

Máx	68838.23
Mín	50477.08
Rango	18361.16
n	13
Clases	3.60555128
Ancho de clase	4590.289
Media est	56348.76
Varianza est	20186770.8

4
4590.299

Prueba de bondad de ajuste						
Clase	LI	LS	Frec Obs	Probabilidad	Frec. Esp	Cociente
1	50477.08	55067.38	6	0.387746319	5.040702151	0.18256432
2	55067.38	59657.68	5	0.381529791521249	4.95988729	0.00032441
3	59657.68	64247.97	1	0.191360997118959	2.487692963	0.88967183
4	64247.974	68838.273	1	0.039362892	0.511717597	0.46592047
			13	1	13	1.53848103

Región de Rechazo
Rechazar H0 si $\chi^2 > \chi^2_{\alpha}$

Estadístico de prueba	
χ^2	1.5385
Valor crítico	
alfa	0.05
χ^2_{α}	3.84145882

Conclusión

No se rechaza H0, con 95% de confianza se afirma que los datos provienen de una distribución normal.

Prueba de bondad de ajuste de la variable IM

HIPÓTESIS
H0: Los valores de la variable provienen de una distribución normal
H1: Los valores de la variable provienen de una distribución normal

IM
111219.06
91268.25
97119.67
94818.58
99629.21
100695.55
110058.12
123650.72
129407.68
132549.64
140237.05
146763.88
144085.13

Máx	146763.88
Mín	91268.25
Rango	55495.63
n	13
Clases	3.60555128
Ancho de clase	13873.9075
Media est	117038.66
Varianza est	369074031

4
13873.9175

Prueba de bondad de ajuste						
Clase	LI	LS	Frec Obs	Probabilidad	Frec. Esp	Cociente
1	91268.25	105142.17	5	0.267877723	3.482410401	0.661346
2	105142.17	119016.09	2	0.273113242118021	3.550472148	0.67708287
3	119016.09	132890.00	3	0.254353166965671	3.306591171	0.02842751
4	132890.005	146763.922	3	0.204655868	2.660526281	0.04331564
			13	1	13	1.41017202

Región de Rechazo
Rechazar H0 si $X^2 > X^2_{\alpha}$

Estadístico de prueba	
X^2	1.4102
Valor crítico	
α	0.05
X^2_{α}	3.84145882

Conclusión

No se rechaza H0, con 95% de confianza se afirma que los datos provienen de una distribución normal.

Prueba de bondad de ajuste de la variable AGFPC

HIPÓTESIS
H0: Los valores de la variable provienen de una distribución normal
H1: Los valores de la variable provienen de una distribución normal

AGFPC
11634.06
13217.61
13323.31
13184.16
13219.84
14181.79
14087.13
14860.53
13841.08
15072.72
15613.14
15459.81
14542.51

Máx	15613.14
Mín	11634.06
Rango	3979.09
n	13
Clases	3.60555128
Ancho de clase	994.7715
Media est	14018.28
Varianza est	1147873.61

4
994.7815

Prueba de bondad de ajuste						
Clase	LI	LS	Frec Obs	Probabilidad	Frec. Esp	Cociente
1	11634.06	12628.84	1	0.097339019	1.265407242	0.05566667
2	12628.84	13623.62	4	0.258960964587716	3.36649254	0.1192136
3	13623.62	14618.40	4	0.356004915527658	4.628063902	0.08523311
4	14618.4015	15613.183	4	0.287695101	3.740036316	0.01806964
			13	1	13	0.27818302

Región de Rechazo
Rechazar H0 si $X^2 > X^2_{\alpha}$

Estadístico de prueba	
X^2	0.2782
Valor crítico	
alfa	0.05
X^2_{α}	3.84145882

Conclusión

No se rechaza H0, con 95% de confianza se afirma que los datos provienen de una distribución normal.

2.3. Multicolinealidad

Como se puede observar en la *Tabla 4*, las variables no muestran problemas graves de multicolinealidad, puesto que ninguno de los Factores de Inflación de Varianza (FIV) supera las 10 unidades, sino que se mantienen bajos. Dichos resultados indican que es posible trabajar con las variables que sospechamos predictoras del Producto Interno Bruto de Baja California y por lo tanto no es necesario excluir alguna, ya que no hay dependencia lineal entre las variables de regresión.

Coefficientes

Término	Coef	EE del coef.	IC de 95%	Valor T	Valor p	FIV
Constante	-10533	169246	(-410737, 389671)	-0.06	0.952	
AGFPC	8.35	7.45	(-9.27, 25.97)	1.12	0.299	7.16
EAG	3.29	5.85	(-10.55, 17.12)	0.56	0.592	3.89
Construcción	0.15	1.87	(-4.27, 4.57)	0.08	0.938	7.93
IM	2.024	0.427	(1.015, 3.033)	4.75	0.002	7.54
TCA	3.77	3.38	(-4.22, 11.76)	1.12	0.301	6.18

Tabla 4. Revisión de Multicolinealidad con los FIV.

2.4. Tabla de Ajuste modelos lineales

Observando en la *Tabla 5* tenemos que el mejor modelo es de 3 variables siendo estas variables Industrias Manufactureras (IM) , Agricultura, ganadería, aprovechamiento forestal, pesca y caza (AGFPC) y Transportes, correos y almacenamiento (TCA) con un desempeño del 97.1% .

la respuesta es PIBE

Vars	R-cuadrado	R-cuadrado (ajust)	R-cuadrado (pred.)	Cp de Mallows	S	AGFPC	EAG	Construcción	IM	TCA
1	95.4	95.0	93.6	6.7	12835				X	
1	62.5	59.1	49.5	118.7	36662					X
2	97.0	96.4	95.4	3.1	10806	X			X	
2	96.3	95.6	94.7	5.5	12021				X	X
3	97.8	97.1	90.9	2.4	9775.8	X			X	X
3	97.6	96.7	85.5	3.3	10333			X	X	X
4	97.9	96.9	80.9	4.0	10069	X	X		X	X
4	97.9	96.8	85.6	4.3	10288	X		X	X	X
5	97.9	96.5	71.7	6.0	10759	X	X	X	X	X

Tabla 5. Mejores subconjuntos

Por tanto una vez teniendo que el mejor modelo es el mencionado anteriormente haremos un analisis del modelo seleccionado,

Haremos una prueba de significancia

H_0 : Regresion no significativa, es decir, $\beta_1 = \beta_2 = \beta_3 = 0$

H_a : Regresion significativa, si Al menos un $\beta_i \neq 0$

Rechazamos H_0 si $p - \text{valor} < \alpha$, por tanto

Rechazamos H_0 con un 95% de confianza sabemos que al menos un coeficiente de β es diferente de 0, por lo que la regresión es significativa. Siendo nuestra α del 0.05 y con nuestra *Tabla 6* de análisis de varianza sabemos que nuestro p-valor es 0.000.

Fuente	GL	SC Sec	Contribución	SC Ajust.	MC Ajust.	F-Valor	P-Valor
Regression	3	38541606192	97.82%	38541606192	12847202064	134.43	0.000
TCA	1	24616796966	62.48%	307609638	307609638	3.22	0.106
IM	1	13339787625	33.86%	7118505810	7118505810	74.49	0.000
AGFPC	1	585021601	1.48%	585021601	585021601	6.12	0.035
Error	9	860090540	2.18%	860090540	95565616		
Total	12	39401696733	100.00%				

Tabla 6. Analisis de Varianza

Ahora en nuestra ecuación de regresión podemos ver que la variable que mas afecta mayormente la respuesta de forma positiva es AGFPC.

Ecuación de regresión

$$\text{PIBE} = 35470 + 8.98 \text{ AGFPC} + 2.159 \text{ IM} + 3.32 \text{ TCA}$$

Resumen del modelo

S	R-cuadrado	R-cuadrado(ajustado)	PRESS	R-cuadrado (pred)	AICc	BIC
9775.77	97.82%	97.09%	3600269210	90.86%	289.56	283.82

Tabla 7. Resumen del modelo que elegimos como el "mejor".

Por tanto ahora tenemos que comprobar el cumplimiento de los supuestos con ayuda de los residuales para este modelo.

Con ayuda de la *Grafica 10* y *11* podemos revisar el cumplimiento de los supuestos.

En la *Grafica 10* la *Grafica de vs. Ajustes* Considerando que los ejes no tienen la misma escala, no se observa un patrón fuerte, por lo tanto concluimos que se cumple el supuesto de varianza constante.

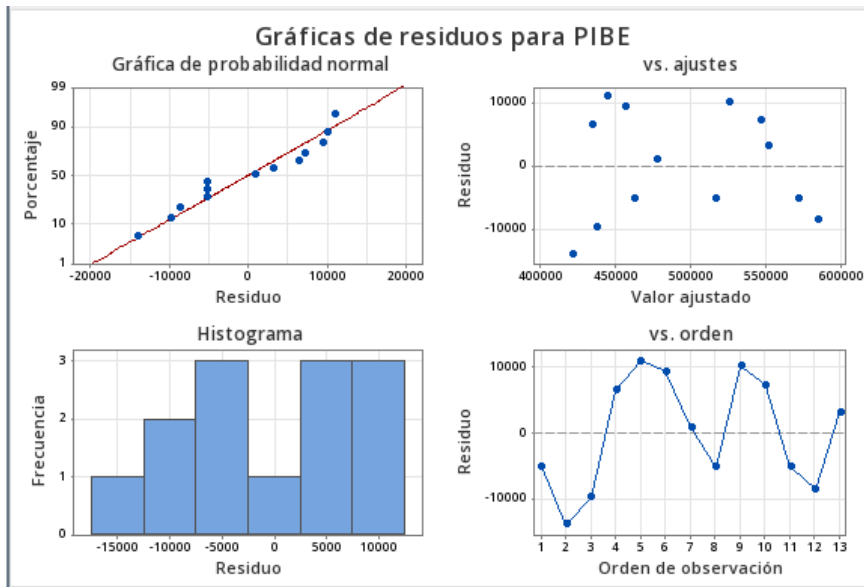
En la grafica vs. Orden no se observa un patrón aparente; por lo tanto decimos que se cumple el supuesto de incorrelación.

Y con ayuda de la *Grafica 11* podemos revisar el supuesto de normalidad con media cero, y si

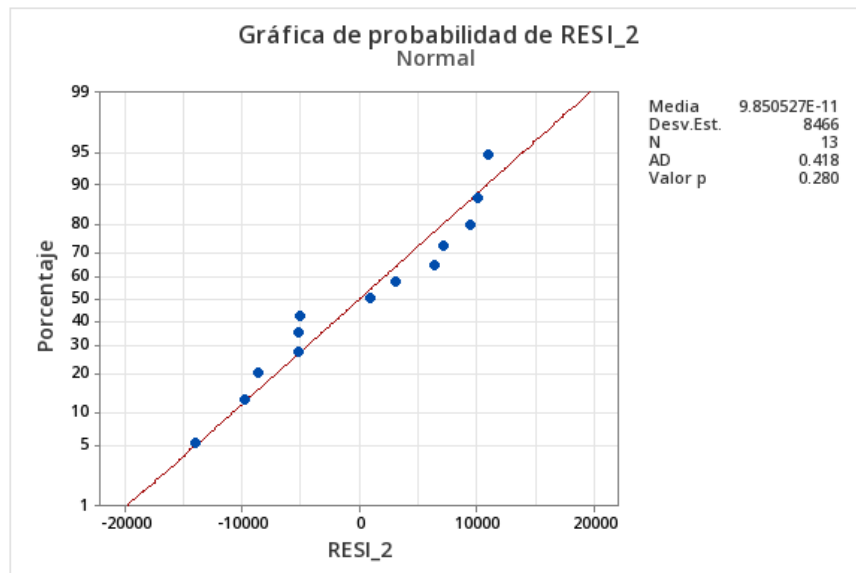
H_0 : los residuales provienen de distribución normal con media cero
 H_1 : los residuales NO provienen de distribución normal con media cero

Rechazamos H_0 si $p - \text{valor} < \alpha$, por tanto

Como $p - \text{valor}$ es de 0.280 entonces no es menor a nuestra α de 0.05 por lo que No rechazamos nuestra hipótesis nula y nuestros residuales provienen de distribución normal con media cero.



Gráfica 10. Verificación del cumplimiento de los supuestos.



Gráfica 11. Revisión de que los residuales provienen de distribución normal y media 0

Entonces como cumple los supuestos tenemos que el mejor modelo es de 3 variables siendo estas variables IM, AGFPC y TCA con un desempeño del 97.1% .

2.5. Tabla de Ajuste modelos No lineales

Para el ajuste modelos no lineales , tomaremos la variable Industrias Manufactureras (IM) siendo esta la que mejor desempeño tiene , por lo observado en la *Tabla 5*, vemos que este modelo de una variable tiene un 95% de desempeño.

Por tanto observando la *Tabla 7* tenemos que el “mejor” modelo no lineal es el logaritmo con un desempeño del 95.22%.

Modelo	Ecuacion estimada	Modelo lineal asociado	Prueba de hipotesis			R2
			Hipotesis (con un nivel de significancia del 0.05)	p valor (Rechazamos H_0 si p valor < alpha)	Conclusion	
Lineal	PIBE estimado = 166989 + 2.799 IM		$H_0: B_1 = 0$, no significativa vs $H_a: \beta \neq 0$, es significativa	0.000	Rechazo H_0 con 95% de confianza , el modelo es significativo	94.98%
Cuadratico	PIBE estimado = 24396 + 5.26 IM - 0.000010 IM ²		$H_0: B_1 = B_2 = 0$, no significativa vs H_a : Al menos una $\beta \neq 0$, es significativa	0.000	Rechazo H_0 con 95% de confianza , el modelo es significativo	94.76%
Cubico	PIBE estimado = 477773 - 6.4 IM + 0.000089 IM ² - 0.000000 IM ³		$H_0: B_1 = B_2 = B_3 = 0$, no significativa vs H_a : Al menos una $\beta \neq 0$, es significativa	0.000	Rechazo H_0 con 95% de confianza , el modelo es significativo	94.23%
Logaritmo	PIBE estimado = -3333638+ 328411ln(IM)	$y = -3333638 + 328411 \ln x$	$H_0: B_1 = 0$, no significativa vs $H_a: \beta \neq 0$, es significativa	0.000	Rechazo H_0 con 95% de confianza , el modelo es significativo	95.22%
Potencia	PIBE estimado =(210.819*(IM))^(0.0652)	$y^* = 5.351 + 0.6652 \ln x$	$H_0: B_1 = 0$, no significativa vs $H_a: \beta \neq 0$, es significativa	0.000	Rechazo H_0 con 95% de confianza , el modelo es significativo	94.78%
Exponencial	PIBE estimado =(253495.157)(exp(0.000006*IM))	$y^* = 12.4431 + 0.000006x$	$H_0: B_1 = 0$, no significativa vs $H_a: \beta \neq 0$, es significativa	0.000	Rechazo H_0 con 95% de confianza , el modelo es significativo	94.11%
Reciproco	PIBE estimado = IM/(0.000001+0.1567IM)	$1/y = 0.000001 + 0.1567 \ln x$	$H_0: B_1 = 0$, no significativa vs $H_a: \beta \neq 0$, es significativa	0.000	Rechazo H_0 con 95% de confianza , el modelo es significativo	94.39%

Tabla 8. Modelos no lineales

Por tanto ahora tenemos que comprobar el cumplimiento de los supuestos para este modelo.

Con ayuda de la *Grafica 12* podemos comprobar el cumplimiento de estos supuestos , teniendo que con la *Grafica de probabilidad normal* o qqplot podemos observar que la mayoría de los datos se encuentran en la linea por lo cual supondremos que provienen de una distribucion normal.

En nuestra *Grafica 12* tambien observamos que la *Grafica vs ajuste* no sigue ningun patron por lo que nuestra varianza es constante. Al igual que con la *Grafica vs orden* no muestra ningun patron por lo que los residuales estan incorrelacionados.

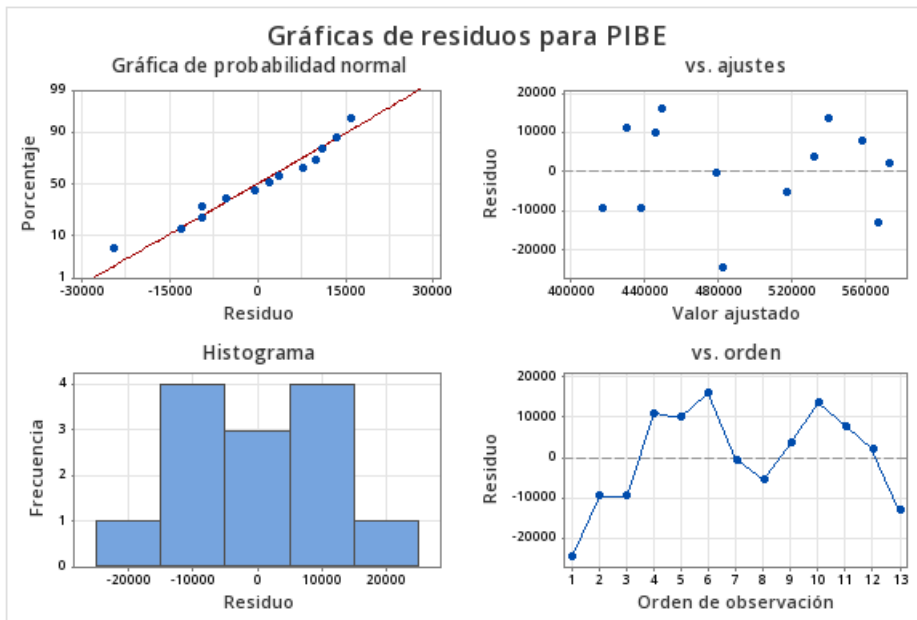
Y observamos con la ayuda de las *Graficas 12 y 13* que nuestro *Histograma* de los residuos muestran media 0 y realizando la prueba de significancia tenemos que

$$H_0: \text{los residuales provienen de distribucion normal con media cero}$$

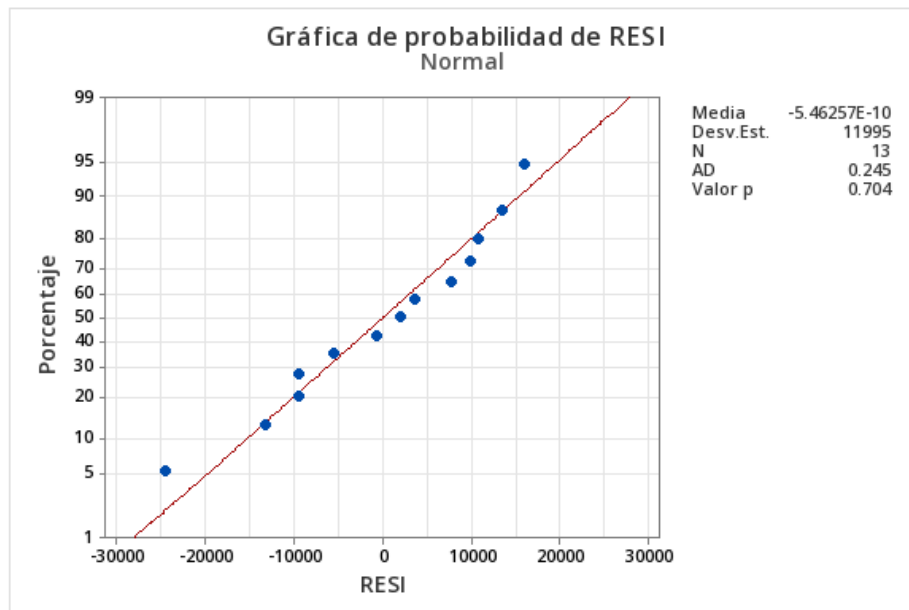
$$H_1: \text{los residuales NO provienen de distribucion normal con media cero}$$

Rechazamos H_0 si $p - \text{valor} < \alpha$, por tanto

Como p - valor es de 0.704 entonces no es menor a nuestra alfa de 0.05 por lo que No rechazamos nuestra hipotesis nula y nuestros residuales provienen de distribucion normal con media cero.



Gráfica 12. Revisión de cumplimiento de supuestos



Gráfica 13. Revisión de que los residuos provienen de distribución normal y media 0.

Por tanto como se cumplen los supuestos , entonces el modelo no lineal que elegimos como el “mejor” es el logaritmo con un desempeño del 95.22%.

2.6. Analisis de la existencia de datos atípicos, palanca y de influencia

Tomando el mejor modelo lineal de nuestros datos como el modelo de 3 variables, entonces para encontrar la existencia de datos atípicos tenemos que tomar los residuos estandarizados y observar que ningun dato sea mayor a 3 o menor 3 por lo tanto como se observa en la *Tabla 9* no se encontro ningun dato atipico. Ya que ninguno es mayor a 3 o menor a 3.

Residual estandarizado	dato atipico. (es mayor a 3 o -3)
-1.15116398	no atipico
-1.651095824	no atipico
-1.101628019	no atipico
0.727640488	no atipico
1.212322489	no atipico
1.089521578	no atipico
0.095918307	no atipico
-0.575007394	no atipico
1.149935441	no atipico
0.818537076	no atipico
-0.65725214	no atipico
-1.069698871	no atipico
1.145974791	no atipico

Tabla 9. Analisis de la existencia de datos atipicos.

Los valores de H_i están entre 0 y 1. Minitab identifica las observaciones con valores de apalancamiento superior a $3p/n$ o 0.99, el valor que sea menor, mediante una X en la tabla de ajustes y diagnósticos de observaciones poco usuales.

El H_i , también denominado apalancamiento, mide la distancia del valor x de una observación hasta el promedio de los valores x de todas las observaciones en un conjunto de datos.

HI	¿Apalancamiento?
0.79001882	FALSO
0.24888737	FALSO
0.17151063	FALSO
0.18663626	FALSO
0.14167053	FALSO
0.22425847	FALSO
0.1073724	FALSO
0.14327836	FALSO
0.19163518	FALSO
0.20808915	FALSO
0.34645819	FALSO
0.31671105	FALSO
0.92347359	VERDADERO

p (Numero de coeficientes en el modelo)	4
n (Numero de observaciones)	13
Regla ($3p/n$)	0.92307692

Tabla 10. Analisis de la existencia de datos con apalancamiento

Por tanto aunque se observe en la *Tabla 10* que hay un dato con apalancamiento esto porque estamos usando la regla de que sea superior a $3p/n$ como en Minitab , sin embargo todos los valores estan entre 0 y 1.

Por ultimo para conocer si una observación tiene influencia, usamos $D_i > 1$, donde D_i es la distancia "COOK" de la observación i . Donde las influencias de los factores indican cuánto explica un factor a una variable. Las influencias pueden ir de -1 a 1. Las influencias cercanas a -1 o 1 indican que el factor afecta considerablemente a la variable. Las influencias cercanas a 0 indican que el factor tiene poca influencia en la variable.

Lo cual como se observa si se encontraron 2 valores con influencia lo que indica que el factor afecta considerablemente a la variable.

COOK	¿Influencia?
1.24644023	VERDADERO
0.22583038	FALSO
0.06280787	FALSO
0.03037281	FALSO
0.06064595	FALSO
0.08579135	FALSO
0.00027667	FALSO
0.01382381	FALSO
0.07837089	FALSO
0.0440139	FALSO
0.05725079	FALSO
0.1325934	FALSO
3.96189774	VERDADERO

Tabla 11. Analisis de de la existencia de datos con influencia

3 Conclusiones

En conclusión , en este analisis como lo habiamos mencionado nuestro objetivo era encontrar el modelo que mediante variables que sospechamos predictivas para el PIB de Baja California , sea el que mejor predice nuestros datos .

Entonces pudimos encontrar que las variables que mas relación tenían con la respuesta era Industrias Manufactureras (IM), Agricultura, ganadería, aprovechamiento forestal, pesca y caza (AGFPC) y Transportes, correos y almacenamiento (TCA). Lo cual se pudo observar desde nuestro analisis exploratorio de los datos mostrando en la matriz de dispersion patrones y en el correlograma y la tabla de correlación eran las que mas relación mostraban .

Sin embargo con el coeficiente de determinacion ajustado que es el porcentaje de la variacion en la respuesta que es explicada por el modelo , obtenemos que el nivel del desempeño es muy bueno con un coeficiente de determinacion del 97.1% siendo este modelo el mismo del que sospechamos en el analisis exploratorio , es decir , el modelo lineal fue el de 3 variables siendo estas variables IM, AGFPC y TCA, las cuales cumplen con los supuestos , por lo cual podria ser propuesto como el mejor modelo de nuestros datos.

Tambien se pudo encontrar un modelo no lineal el cual cumpliera con los supuestos y mostraba un muy buen desempeño con un coeficiente de determinacion del 95.22% , siendo este modelo no lineal el modelo logaritmo ,para el cual tomamos la variable Industrias Manufactureras (IM) siendo esta la que mejor desempeño mostro en modelos de una variables , por lo observado en la *Tabla 5*, teniendo un 95% de desempeño.

Por lo cual las variables o actividades economicas que mayor relacion tienen o que mejor predicen el PIB de Baja California según lo analizado son Industrias Manufactureras (IM), Agricultura, ganadería, aprovechamiento forestal, pesca y caza (AGFPC) y Transportes, correos y almacenamiento (TCA). Por lo cual el modelo de estas tres variables es el mejor modelo.

Referencias

INEGI. (2013). *Sistema de Cuentas Nacionales de México*. Obtenido de inegi.org.mx:

https://www.inegi.org.mx/contenidos/programas/pibent/2013/doc/met_pibefa.pdf

INEGI. (2022). *PIB por Entidad Federativa (PIBE). Base 2013*. Obtenido de inegi.org.mx:

<https://www.inegi.org.mx/programas/pibent/2013/#Tabulados>

Minitab. (s.f.). Obtenido de app.minitab: <https://app.minitab.com>

Minitab. (2023). *Ejemplo de Mostrar estadísticos descriptivos*. Obtenido de Soporte de Minitab:

<https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistics/basic-statistics/how-to/display-descriptive-statistics/before-you-start/example/>

Presentaciones de Métodos Estadísticos . MET. Alejandra Cerda.