

En esta sección haremos el análisis de las variables elegidas para formular el modelo predictivo. El tipo de variable, escala de medición, como se obtuvieron y una breve explicación de lo que nos dan de información las variables.

Para este ejercicio tomamos datos de 10 años atrás los cuales abarcan del 2009 al 2018 y las variables elegidas son las siguientes:

**Inflación:** es una medida económica que indica el crecimiento generalizado de los precios de bienes, servicios y factores productivos dentro de una economía en un periodo determinado. Se calcula a través del IPC (índice de precios al consumo, valor numérico que refleja las variaciones que experimentan los precios en un período determinado.). Un país altas tasas de inflación tendrá mayor capacidad de pago ya que altas tasas de inflación tienen como consecuencia un precio menos competitivo y problemas en la balanza de pagos.

Cabe recalcar que en cuanto mayor es la inflación, mayores serán los costes que sufra la economía, partiendo de la pérdida del poder adquisitivo del dinero. Altas tasas desincentivan el ahorro y fomentan la fuga de capitales.

Periodo	Inflación
2009	4.83
2010	6.77
2011	3.36
2012	2.13
2013	4.08
2014	3.97
2015	3.57
2016	3.82
2017	4.40
2018	3.57

**Tipo de variable:** Continua.

**Escala de medición:** Porcentual anual.

**De donde se obtuvieron:** <https://es.inflation.eu/tasas-de-inflacion/mexico/inflacionhistorica/ipc-inflacion-mexico.aspx>

**Tasa de desempleo:** también conocida como tasa de paro, mide el nivel de desocupación en relación con la población activa. En otras palabras, es la parte de la población que estando en edad, condiciones y disposición de trabajar (población activa), no tiene puesto de trabajo.

Esta se calcula tomando a la población de 15 años y más que no está trabajando y busca trabajo dividido por la población económicamente activa de 15 años y más (ocupados más desocupados).

Periodo	Tasa desempleo
2009	6.5
2010	5.6
2011	6.0
2012	6.2
2013	5.3
2014	5.6
2015	4.1
2016	2.6
2017	2.9
2018	2.5

**Tipo de variable:** Continua.

**Escala de medición:** Porcentual anual.

**De donde se obtuvieron:** <https://www.inegi.org.mx/app/estatal/?ag=07000002#grafica>

**Ventas internacionales:** Las Ventas Internacionales se refieren al total de las ventas hacia el extranjero de las personas morales cuyo domicilio fiscal se encuentra registrado en el estado de Baja California. Así, las Ventas Internacionales por Entidad Federativa representan un concepto diferente al de Exportaciones por Entidad Federativa, cuya publicación está a cargo del Instituto Nacional de Estadística y Geografía (INEGI).

Periodo	Ventas internacionales
2009	6398.65
2010	6603.74
2011	20471.5
2012	12915.5
2013	26376.2
2014	27552.6
2015	28132.9

2016	25875.2
2017	27229
2018	29784.7

**Tipo de variable:** Continua.

**Escala de medición:** Millones de pesos.

**De donde se obtuvieron:** <https://datamexico.org/es/profile/geo/baja-california-bc#economiaventas-internacionales>

**Compras internacionales:** Las Compras Internacionales se refieren al total de las compras a países extranjeros de las personas morales cuyo domicilio fiscal se encuentra registrado en el estado de Baja California. Así, las Compras Internacionales por Entidad Federativa representan un concepto diferente al de Importaciones, cuya publicación está a cargo del Instituto Nacional de Estadística y Geografía (INEGI).

Periodo	Compras internacionales
2009	12798.1
2010	23305.2
2011	22812.8
2012	24499.1
2013	22995.1
2014	27941.2
2015	26430
2016	23231
2017	27218.4
2018	28912.8

**Tipo de variable:** Continua.

**Escala de medición:** En millones de pesos.

**De donde se obtuvieron:** <https://datamexico.org/es/profile/geo/baja-california-bc#economiacompras-internacionales>

**IED:** La Inversión Extranjera Directa (IED) es aquella que tiene como propósito crear un vínculo duradero con fines económicos y empresariales de largo plazo, por parte de un inversionista extranjero en el país receptor.

Además tiene la ventaja de que suele ser más estable que otro tipo de flujos los cuales pueden ser más especulativos, esto se debe a que las inversiones extranjeras directas son dirigidas a proyectos a largo plazo, lo que implicará que haya un mayor nivel de crecimiento económico, más fuentes de empleo, más exportaciones y más divisas.

Periodo	IED
2009	829.8438
2010	1414.823
2011	766.09
2012	1017.843
2013	1290.554
2014	1215.892
2015	1166.431
2016	1565.163
2017	1677.488
2018	1633.636

**Tipo de variable:** Continua.

**Escala de medición:** En millones de dolares.

**De donde se obtuvieron:** <https://datamexico.org/es/profile/geo/baja-california-bc#economiaventas-internacionales> (En la sección de Inversión Extranjera Directa).

**PIB:** El Producto interno bruto es la suma de todos los bienes y servicios finales que produce un país o una economía, elaborados dentro del territorio nacional tanto por empresas nacionales como extranjeras, y que se registran en un periodo determinado (generalmente un año).

Periodo	PIB
2009	407,746
2010	428,163
2011	440,701
2012	456,024
2013	465,525
2014	478,122
2015	511,460
2016	535,553

<b>2017</b>	553,327
<b>2018</b>	565,895

**Tipo de variable:** Continua.

**Escala de medición:** En millones de pesos.

**De** **donde** **se** **obtuvieron:**  
<https://www.inegi.org.mx/app/tabulados/default.aspx?pr=17&vr=7&in=2&tp=20&wr=1&cno=2>

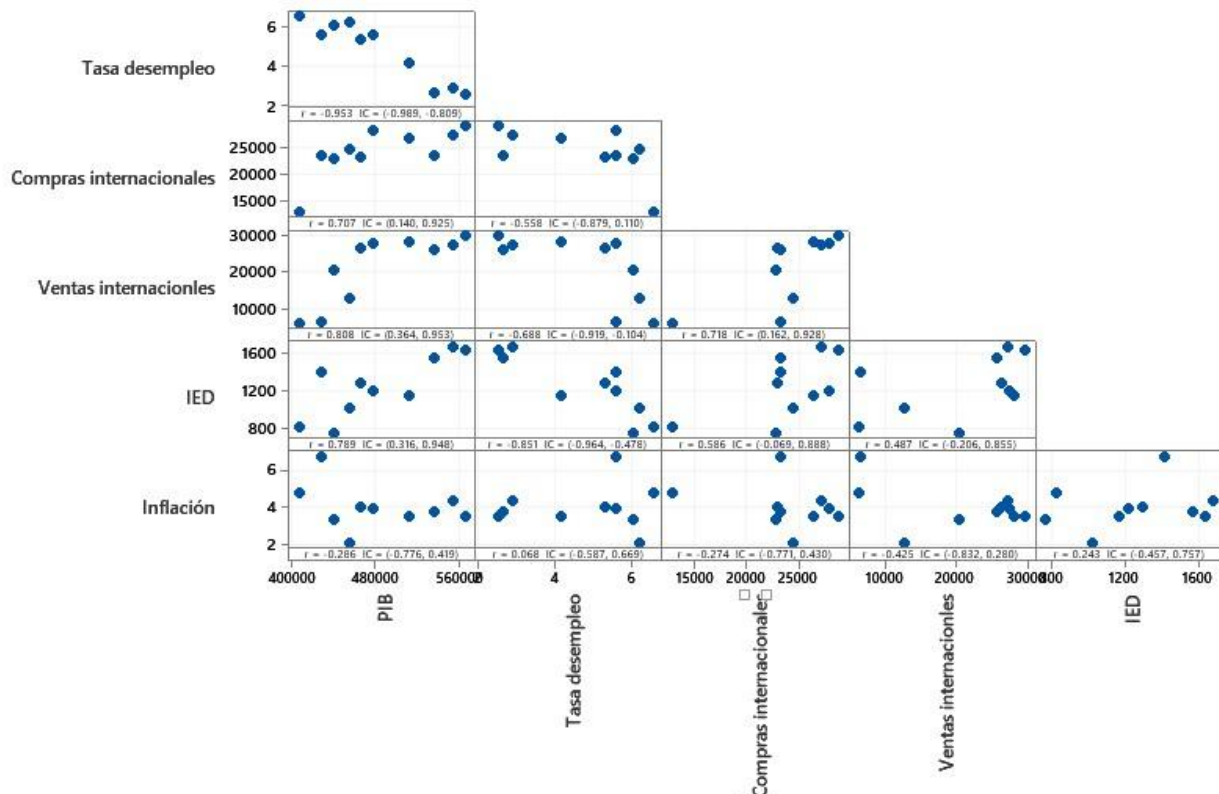
## **Desarrollo:**

### **1. Análisis exploratorio de los datos:**

Primero observamos la correlación de nuestras variables y su dispersion por medio de gráficas y tablas que ayudan al análisis preliminar del modelo de regresion.

Utilizamos la correlación de Pearson con un intervalo de confianza del 95%.

Gráfica de matriz de PIB, Tasa desempleo, Compras internacionales, Ventas internacionales, IED, Inflación  
IC de 95% para la correlación de Pearson



## Correlaciones

	PIB	Tasa desempleo	Compras internacionales	Ventas internacionales	IED
Tasa desempleo	-0.953				
Compras internacionales	0.707	-0.558			
Ventas internacionales	0.808	-0.688	0.718		
IED	0.789	-0.851	0.586	0.487	
Inflación	-0.286	0.068	-0.274	-0.425	0.243

Las correlaciones más significativas son las siguientes:

Existe una relación lineal positiva entre:

- PIB y Compras internacionales de 0.707.
- PIB y Ventas internacionales de 0.808.
- PIB y IED de 0.789.
- IED y compras internacionales de 0.586.
- IED y ventas internacionales de 0.487.
- Compras internacionales y ventas internacionales de 0.718.

Estos valores indican que existe una relación positiva moderada entre las variables. Esto quiere decir que a mayor PIB, hay mayor compras internacionales, ventas internacionales, mayor inversión extranjera, etc.

**Existe una relación lineal negativa para los siguientes pares, con coeficientes de correlación de Pearson negativos:**

- PIB y tasa de desempleo de -0.953.
- Tasa de desempleo y ventas internacionales de -0.688.
- Tasa de desempleo y compras internacionales de -0.558.
- Tasa de desempleo y IED de -0.851.
- Inflación y ventas internacionales de -0.425.

La relación entre estas variables es negativa, lo cual indica que a mayor PIB, hay menor tasa de desempleo, a menor tasa de desempleo, hay menos compra y venta internacional, IED, etc.

**En base al análisis exploratorio podemos decir que las variables que sospechamos tendrán mayor impacto en el modelo de regresión, son las variables de compra internacional, IED, inflación y tasa de desempleo.**

El modelo a plantear inicialmente es el siguiente:

#### **Ecuación de regresión**

$$\text{PIB} = 567146 - 25073 \text{ Tasa desempleo} + 1,55 \text{ Compras internacionales} \\ + 0,550 \text{ Ventas internacionales} + 17,6 \text{ IED} - 8626 \text{ Inflación}$$

Pero primero revisaremos la distribución de los datos.

## **2. Pruebas de bondad de ajuste usando un IC de 95%.**

**\*Para todas las pruebas siguientes se rechaza H0 si p-valor < a alfa.**

**Prueba para PIB.**

### Conteos observados y esperados

<u>Categoría</u>	<u>Observado</u>	<u>Proporción</u> <u>de prueba</u>	<u>Esperado</u>	<u>Contribución a</u> <u>chi-cuadrada</u>
1	3	0.249417	2.49417	0.102586
2	3	0.269326	2.69326	0.034936
3	1	0.260703	2.60703	0.990608
4	3	0.220555	2.20555	0.286168

4 (100.00%) de los conteos esperados son menores que 5.

### Prueba de chi-cuadrada

#### N GL Chi-cuad. Valor p

10 3 1.41430 0.702

H0: Los datos se distribuyen normal ( 484,252, 54,660.5).

H1: los datos no se distribuyen normal.

El estadístico de chi-cuadrada general es 1.141 y tiene un valor p de 0.702 . Puesto que el valor p es mayor que el nivel de significancia de 0.05, No se rechaza H0, podemos decir con un 95% de confianza, que los datos siguen una distribución normal con media 484,252 y sigma 54,660.5.

Prueba para tasa de desempleo.

### Conteos observados y esperados

<u>Categoría</u>	<u>Observado</u>	<u>Proporción</u> <u>de prueba</u>	<u>Esperado</u>	<u>Contribución a</u> <u>chi-cuadrada</u>
1	3	0.213732	2.13732	0.34820
2	1	0.226002	2.26002	0.70249
3	1	0.248287	2.48287	0.88563
4	5	0.311979	3.11979	1.13316

### Prueba de chi-cuadrada

#### N GL Chi-cuad. Valor p

10 3 3.06948 0.381

H0: Los datos se distribuyen normal (5, 1.564). H1:

los datos no se distribuyen normal.



El estadístico de chi-cuadrada general es 3.069 y tiene un valor p de 0.381 Puesto  
que el valor p es mayor que el nivel de significancia de 0.05, No se rechaza H0, y podemos decir con un  
95% de confianza, que los datos siguen una distribución normal con media 5 y sigma 1.564

Prueba para compras internacionales.

Conteos observados y esperados

<u>Categoría</u>	<u>Observado</u>	<u>Proporción</u>		<u>Contribución a</u>
		<u>de prueba</u>	<u>Esperado</u>	
1	1	0.056521	0.56521	0.33446
2	0	0.186551	1.86551	1.86551
3	5	0.332961	3.32961	0.83799
4	4	0.423966	4.23966	0.01355

4 (100.00%) de los conteos esperados son menores que 5.

Prueba de chi-cuadrada

<u>N GL Chi-cuad. Valor p</u>			
10	3	3.05151	0.384

H0: Los datos se distribuyen normal ( 24,014, 4535.71). H1:  
los datos no se distribuyen normal.

El estadístico de chi-cuadrada general es 3.051 y tiene un valor p de 0.384 Puesto  
que el valor p es mayor que el nivel de significancia de 0.05, No se rechaza H0, y podemos decir con un  
95% de confianza, que los datos siguen una distribución normal con media 24,014 y sigma 4535.718

Prueba para Ventas internacionales.

Conteos observados y esperados

<u>Categoría</u>	<u>Observado</u>	<u>Proporción</u>		<u>Contribución a</u>
		<u>de prueba</u>	<u>Esperado</u>	
1	2	0.164834	1.64834	0.07503

2	1	0.204495	2.04495	0.53396
3	1	0.251447	2.51447	0.91217
4	6	0.379224	3.79224	1.28530

4 (100.00%) de los conteos esperados son menores que 5.

## Prueba de chi-cuadrada

### N GL Chi-cuad. Valor p

10 3 2.80646 0.422

H0: Los datos se distribuyen normal ( 21,134, 9118.772). H1: los datos no se distribuyen normal.

El estadístico de chi-cuadrada general es 2.8 y tiene un valor p de 0.422  
p es mayor que el nivel de significancia de 0.05, No se rechaza H0,  
de confianza, que los datos siguen una distribución normal con media 21,134 y sigma 9118.772

Puesto que el valor  
y podemos decir con un 95%

## Prueba para IED.

### Conteos observados y esperados

Categoría	Observado	Proporción		Contribución a chi-cuadrada
		de prueba	Esperado	
1	2	0.205917	2.05917	0.001700
2	3	0.249519	2.49519	0.102132
3	2	0.269241	2.69241	0.178067
4	3	0.275323	2.75323	0.022117

4 (100.00%) de los conteos esperados son menores que 5.

## Prueba de chi-cuadrada

### N GL Chi-cuad. Valor p

10 3 0.304016 0.959

H0: Los datos se distribuyen normal ( 1,258, 321.48).

H1: los datos no se distribuyen normal.

El estadístico de chi-cuadrada general es 0.30 y tiene un valor p de 0.959

Puesto que el valor p es mayor que el nivel de significancia de 0.05, No se rechaza H0, y podemos decir con un 95% de confianza, que los datos siguen una distribución normal con media 1,258 y sigma 321.48

## Prueba para inflación.

### Conteos observados y esperados

<u>Categoría</u>	<u>Observado</u>	<u>Proporción de prueba</u>	<u>Esperado</u>	<u>Contribución a chi-cuadrada</u>
1	1	0.262075	2.62075	1.00232
2	7	0.369204	3.69204	2.96384
3	1	0.273190	2.73190	1.09795
4	1	0.095531	0.95531	0.00209

### Prueba de chi-cuadrada

#### N GL Chi-cuad. Valor p

10 3 5.06620 0.167

H0: Los datos se distribuyen normal 4,1.1931).

H1: los datos no se distribuyen normal.

El estadístico de chi-cuadrada general es 5.066 y tiene un valor p de 0.167 . Puesto que el valor p es mayor que el nivel de significancia de 0.05, No se rechaza H0, y podemos decir con un 95% de confianza, que los datos siguen una distribución normal con media 4 y sigma 1.1931.

## 3. Prueba de multicolinealidad.

### Coeficientes

<u>Término</u>	<u>Coef</u>	<u>EE del coef.</u>	<u>IC de 95%</u>	<u>Valor T</u>	<u>Valor p</u>	<u>FIV</u>
Constante	567146	54497	(415840, 718453)	10.41	0.000	
Tasa desempleo	-25073	5925	(-41523, -8622)	-4.23	0.013	7.46
Compras internacionales	1.55	1.33	(-2.15, 5.25)	1.17	0.308	3.17
Ventas internacionales	0.550	0.722	(-1.456, 2.556)	0.76	0.489	3.77
IED	17.6	30.2	(-66.3, 101.5)	0.58	0.592	8.20
Inflación	-8626	4213	(-20322, 3070)	-2.05	0.110	2.19

En la tabla de coeficientes Podemos ver que no existe ningún FIV > 10, por lo que no hay problemas de multicolinealidad con nuestras variables.

#### 4. Ajuste modelos lineales en tabla.

##### Comparación de modelos cuadrático y cúbico.

MODELO	ecuación estimada	Modelo lineal asociado
Cuadrático	$y = 607884 - 15626 x - 2002 x^2$	No aplica
Cúbico	$y = 614368 - 20390x - 917 x^2 - 78x^3$	No aplica

##### Prueba de significancia

MODELO				R2
	hipótesis	p valor	conclusión	
Cuadrático	H0:regresión no significativa, H1: Regresión significativa	0	Como P valor(0)<alfa 0.05, la regresión es significativa	91.13%
Cúbico	H0:regresión no significativa, H1: Regresión significativa	0.001	Como P valor(0)<alfa 0.05, la regresión es significativa	91.13%

Ambos modelos tienen la misma significancia del 91.13%, sin embargo, el modelo cuadrático tiene mejor R2 ajustada (88.59%).

##### Comprobación de supuestos del modelo.

###### La ecuación de regresión es:

$$y = 607884 - 15626 x - 2002 x^2$$

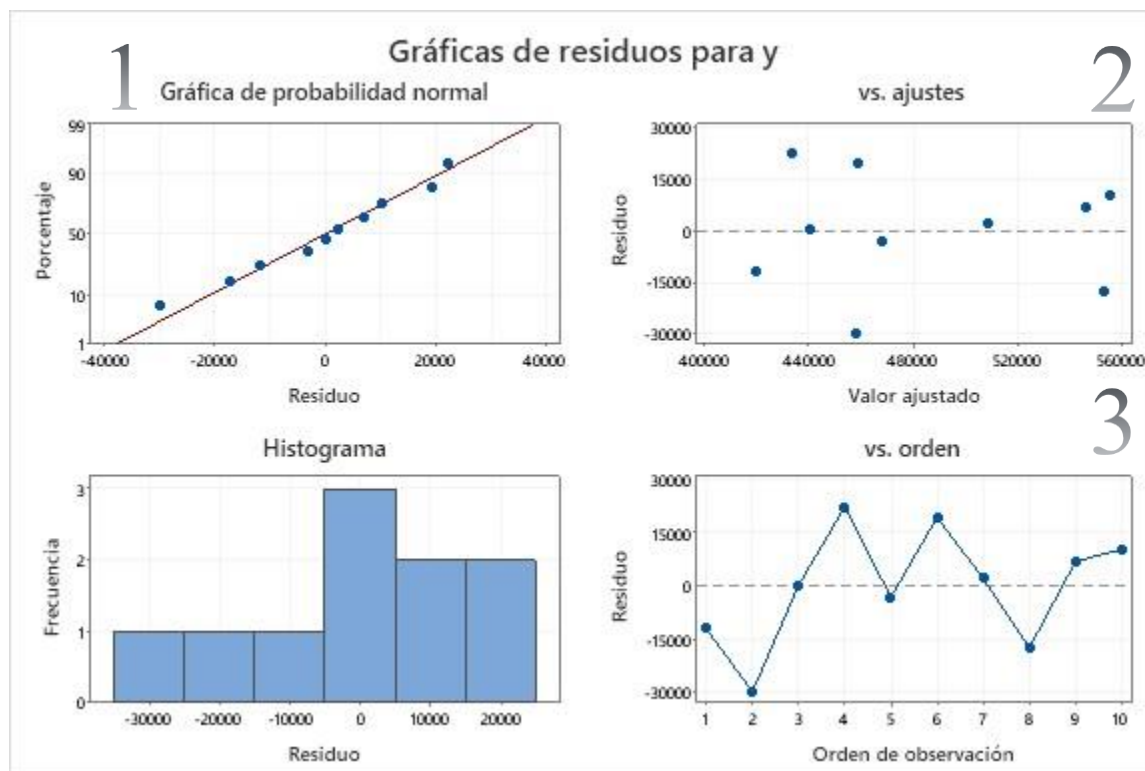
###### Resumen del modelo

	R-cuad.
S R-cuad. (ajustado)	88.59%
18461.6	91.13%

###### Análisis de Varianza

Fuente	GL	SC	MC	F	P
Regresión	2	2.45041E+10	1.22521E+10	35.95	0.000
Error	7	2.38582E+09	3.40832E+08		
Total	9	2.68899E+10			

En esta tabla, vemos que el p-valor de la regresión es de 0 y por ser menor a alfa de 0.05, se dice que la regresión es significativa.



1.- Normalidad: Podemos ver que la dispersion se acerca mucho a la linea, por lo que Podemos decir que el supuesto se cumple.

2.- Varianza constante: Ya que los valores ajustados no tienen un patrón visible, el supuesto se cumple.

3.- Aleatoriedad: Ya que no existe un patrón visible en el orden de los residuos, el supuesto se cumple.

## 5. Ajuste de modelos NO lineales.

MODELO	ecuación estimada	Modelo lineal asociado	Prueba de significancia hipótesis
lineal	$y = 642330 - 33314 x$	No aplica	H0: regresión no significativa, H1: Regresión signific
potencia	$y = 731,169.075 X^{(-0.2788)}$	$y^* = 13.5024 - 0.2788 x^*$	H0: regresión no significativa, H1: Regresión signific
exponenci	$y = 664,838,877e^{(-0.06799(x))}$	$y^* = 13.4073 - 0.06799 x$	H0: regresión no significativa, H1: Regresión signific
logaritmo	$y = 689595 - 137,020x^*$	$y = 689595 - 137020 \ln(x^*)$	H0: regresión no significativa, H1: Regresión signific
recíproco	$y = x / (0.000003x - 0.000002)$	$1/y = 0.000003 - 0.000002(1/x)$	H0: regresión no significativa, H1: Regresión signific

MODELO	Prueba de significancia			R2
	p valor	conclusión		
lineal	0	Como P valor(0) < alfa 0.05, la regresión es significativa		90.87%
potencia	0	Como P valor(0) < alfa 0.05, la regresión es significativa		87.13%
exponenci	0	Como P valor(0) < alfa 0.05, la regresión es significativa		89.46%
logaritmo	0	Como P valor(0) < alfa 0.05, la regresión es significativa		89.07%
recíproco	0	Como P valor(0) < alfa 0.05, la regresión es significativa		80.88%

Ajustando los modelos lineales, encontramos que el mejor modelo para predecir nuestra variable PIB es un modelo lineal, ya que cuenta con una R2 del 90.87%. A continuación probaremos los supuestos de nuestro modelo lineal.

### Comprobación de supuestos del modelo lineal.

#### Ecuación de regresión

$$y = 642330 - 33314x$$

#### Coeficientes

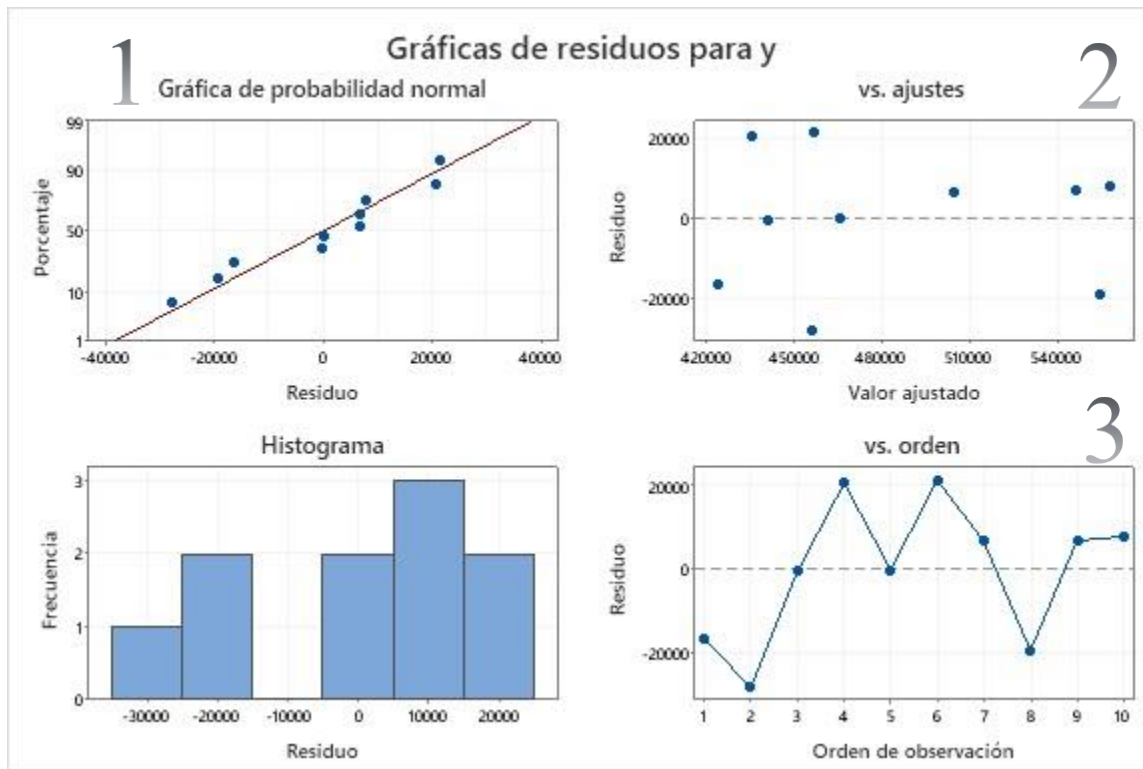
Término	Coef	EE del coef.	IC de 95%	Valor T	Valor p	FIV
Constante	642330	18562	(599525, 685135)	34.60	0.000	
x	-33314	3734	(-41924, -24704)	-8.92	0.000	1.00

En esta sección, Podemos ver que no hay evidencia de una ecuación al origen, ya que el IC de nuestra constante no pasa por 0.

#### Análisis de Varianza

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	1	24434623508	90.87%	24434623508	24434623508	79.61	0.000
x	1	24434623508	90.87%	24434623508	24434623508	79.61	0.000
Error	8	2455310400	9.13%	2455310400	306913800		
Total	9	26889933908	100.00%				

En esta sección Podemos verificar que nuestro p-valor es menor a nuestro alfa de 0.05, lo cual quiere decir que la regresión es significativa.



1.- Normalidad: Podemos ver que la dispersion se acerca mucho a la linea, por lo que Podemos decir que el supuesto se cumple.

2.- Varianza constante: Ya que los valores ajustados no tienen un patrón visible, el supuesto se cumple.

3.- Aleatoriedad: Ya que no existe un patrón visible en el orden de los residuos, el supuesto se cumple.

**Hasta el momento**, nuestro modelo que mejor se ajusta a los datos es el modelo de **función cuadrática con R2 de 91.13% y R2 ajustada de 88.59%**. A continuación evaluaremos la tabla de mejores subconjuntos para contrastar la mejor regresión contra la actual.

**la respuesta es PIB**

C o  
m V p e r n  
a t s a      s i  
T n i a t n  
s e t

<div> <div>are</div> <div>nr</div> <div>danl</div> <div>ecansicf</div> <div>eoilmnoap</div> <div>anc</div> <div>l l l l i</div> <div>e e e E ó</div> <div>S o s s D n</div> </div>									
Vars	R-cuad.	R-cuad. (ajust)	R-cuad. (pred.)	Cp de Mallows					
1	90.9	89.7	85.8	17.7	17519	X			
1	65.4	61.0	51.2	83.9	34124		X		
2	95.8	94.6	92.6	6.9	12705	X			X
2	95.3	94.0	86.6	8.1	13397	X	X		
3	98.2	97.3	86.3	2.7	9035.1	X X	X 3	97.0	95.5
4	98.2	96.8	84.3	4.6	9742.9	X	X		X X
5	98.5	96.5	69.8	6.0	10180	X	X	X X	X

Podemos observar que el “mejor” modelo es de que incluye las 5 variables por su R2 y R2 ajustada, sin embargo, su R2 pronosticado es de 69.8%. Lo que indica que probablemente hay un sobreajuste, esto sucede cuando hay variables que no tienen tanta importancia para nuestros datos. El modelo se adapta a los datos existentes, pero no cumple con lo necesario para predecir el comportamiento de nuevos datos.

Por lo que descartamos esa opción y elegimos la que tiene las 3 variables con mayor R2, R2 ajustada y R2 pronóstico , que es la que incluye las variables: Tasa de desempleo, Compras Internacionales e Inflación.

**Comprobando supuestos del nuevo mejor modelo de 3 variables. (Tasa de desempleo, Compras Internacionales e Inflación)**



## Ecuación de regresión

PIB = 598994 - 29111 Tasa desempleo + 2.338 Compras internacionales - 8088 Inflación

## Coeficientes

Término	Coef	EE del coef.	IC de 95%	Valor T	Valor p	FIV
Constante	598994	32316	(519919, 678069)	18.54	0.000	
Tasa desempleo	-29111	2334	(-34822, -23399)	-12.47	0.000	1.47
Compras internacionales	2.338	0.835	(0.295, 4.381)	2.80	0.031	1.58
Inflación	-8088	2639	(-14546, -1629)	-3.06	0.022	1.09

## Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	PRESS	R-cuad. (pred)	AICc	BIC
9035.05	98.18%	97.27%	3694974430	86.26%	230.45	216.96

En la tabla de coeficientes podemos ver que el IC de la constante no pasa por el 0, lo cual indica que no existe evidencia de una ecuación al origen.

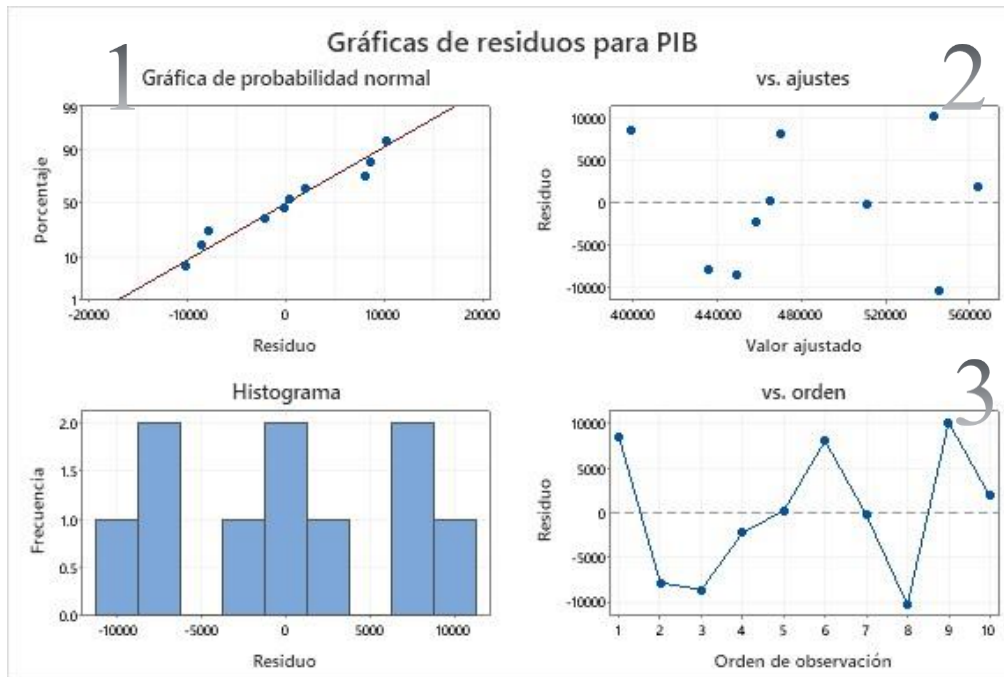
También Podemos ver que los valores FIV son todos menores a 10, por lo tanto, no hay problemas de multicolinealidad.

## Análisis de Varianza

Fuente	GL	SC Sec.	Contribución	SC Ajust.	MC Ajust.	Valor F
Regresión	3	26400140894	98.18%	26400140894	8800046965	107.80
Tasa desempleo	1	24434623508	90.87%	12697322112	12697322112	155.54
Compras internacionales	1	1199035241	4.46%	640123542	640123542	7.84
Inflación	1	766482145	2.85%	766482145	766482145	9.39
Error	6	489793014	1.82%	489793014	81632169	
Total	9	26889933908	100.00%			

Fuente	Valor p
Regresión	0.000
Tasa desempleo	0.000
Compras internacionales	0.031
Inflación	0.022
Error	
Total	

En esta tabla podemos ver que el P-valor es 0, al ser menor que alfa de 0.05, podemos decir que la regresión es significativa.



1.- Normalidad: Podemos ver que la dispersion se acerca mucho a la linea, por lo que Podemos decir que el supuesto se cumple.

2.- Varianza constante: Ya que los valores ajustados no tienen un patrón visible, el supuesto se cumple.

3.- Aleatoriedad: Ya que no existe un patrón visible en el orden de los residuos, el supuesto se cumple.

**6. Revisión la existencia de datos atípicos, apalancamiento y de influencia del modelo escogido.**

Periodo	RESIDUALES	RESIDUALES ESTANDARIZADOS
2009	8539.354925	2.053219387
2010	-7798.242706	-1.818333745
2011	-8559.537239	-1.073901682
2012	-2186.381216	-0.349163577
2013	292.1398304	0.034361014
2014	8089.782253	1.118336731
2015	-178.6523853	-0.021343272

2016	-10274.27928	-1.552025712
2017	10159.71481	1.321864652
2018	1916.101013	0.261253624

Podemos  
observar  
un dato

**DATO ATÍPICO**

atípico en el año 2009, ya que el residual estandarizado es mayor a 2.

### Apalancamiento.

Para revisar si existe apalancamiento en nuestros datos, checamos cualquier valor  $H_i$  el cual se obtiene al medir la distancia del valor  $X$  con el promedio de los valores de  $X$  de las demás observaciones, si  $H_i$  es mayor a  $3p/n$ , donde  $p$  = Número de términos del modelo incluyendo la constante y  $n$  = Número de observaciones, habrá necesidad de investigar esa observación.

$$3p/n = 1.2$$

Periodo	RESID	RESIDEST	HI	COOK
2009	8539.354925	2.053219387	0.788106282	3.919922
2010	-7798.242706	-1.81833375	0.774687668	2.842031
2011	-8559.537239	-1.07390168	0.22176602	0.082159
2012	-2186.381216	-0.34916358	0.519677897	0.032976
2013	292.1398304	0.034361014	0.114500609	3.82E-05
2014	8089.782253	1.118336731	0.358986539	0.175104
2015	-178.6523853	-0.02134327	0.141709477	1.88E-05
2016	-10274.27928	-1.55202571	0.4631616	0.519549
2017	10159.71481	1.321864652	0.276351412	0.16682
2018	1916.101013	0.261253624	0.341052494	0.008832

Ya que no hay valores mayores a 1.2 ( $ep/n$ ), no hay evidencia de observaciones con apalancamiento significativo que pueda tener algún efecto en el modelo.

### Influencia.

Para medir si existen observaciones que tengan predictores poco comunes o influencia, se utilizan la distancia de cook, la cual se obtiene a partir de medir la distancia entre los valores ajustados calculados con y sin la  $i$ ésima observación.

Diremos que un punto es influyente a nivel alfa si:

$D(i) \geq F_{\alpha, k+1, n-k-1}$ , donde  $F_{\alpha, k+1, n-k-1}$  es el percentil  $(1-\alpha)100\%$  de la ley F de Fisher con  $k+1$  y  $n-k-1$  grados de libertad, donde  $k$  = variables regresoras incluyendo la constante y  $n$  = número de observaciones.

$$F(0.05, 5, 5) = 5.050329058$$

Periodo	RESID	RESIDEST	HI	COOK
2009	8539.354925	2.053219387	0.788106282	3.919922
2010	- 7798.242706	-1.81833375	0.774687668	2.842031
2011	- 8559.537239	-1.07390168	0.22176602	0.082159
2012	- 2186.381216	-0.34916358	0.519677897	0.032976
2013	292.1398304	0.034361014	0.114500609	3.82E-05
2014	8089.782253	1.118336731	0.358986539	0.175104
2015	- 178.6523853	-0.02134327	0.141709477	1.88E-05
2016	- 10274.27928	-1.55202571	0.4631616	0.519549
2017	10159.71481	1.321864652	0.276351412	0.16682
2018	1916.101013	0.261253624	0.341052494	0.008832

Al observar la tabla no hay observaciones con distancia de cook mayor a 5.050, por lo tanto, no hay evidencia de observaciones influyentes.

## Conclusión.

Después de un análisis exhaustivo se obtuvo que el mejor modelo para predecir el PIB de la entidad federativa de Baja California con los datos tomados de fuentes confiables como el inegi, páginas del gobierno de México y datamexico.org la cual es una plataforma que permite la integración, visualización y análisis de datos públicos de la economía Mexicana, es el **modelo de regresión lineal multiple construido por 3 variables** las cuales se obtuvieron por medio de una metodología analítica. Las 3 variables con más peso en la predicción del comportamiento del PIB son **Tasa de desempleo, Compras internacionales e Inflación.**

En la primer sección del análisis vemos que la correlación de las variables elegidas en el modelo son las siguientes:

- PIB y Compras internacionales con una correlación de 0.707.
- PIB y tasa de desempleo con una correlación de -0.953.
- PIB e inflación con una correlación de -0.286

Como se mencionaba con anterioridad, el PIB tiene una correlación negativa muy fuerte esto quiere decir que a mayor PIB, menor tasa de desempleo y viceversa, así como también con un mayor PIB hay más compras internacionales, etc.

Al hacer el primer análisis exploratorio de los datos, sospechamos que las variables que desempeñarían un mayor impacto en el modelo de regresión, eran las variables de compra internacional, IED, inflación y tasa de desempleo, esto, debido a que tenían mayor grado de correlación con la variable PIB, pero al hacer un análisis más profundo vimos que la inflación a pesar de no tener un coeficiente de correlación fuerte, en conjunto con la tasa de desempleo y compras internacionales, nos proporcionaron un mejor modelo para el comportamiento del PIB de Baja California.

Al final, hace mucho sentido que esas 3 variables definan el PIB del estado, ya que la tasa de inflación es un macro indicador de la economía en general, a mayor inflación, mayor incertidumbre, menos gastos no necesarios, etc.

En cuanto a las compras internacionales, tampoco sorprende, ya que estas son de productos que se venden con un margen de ganancia y eso se ve reflejado en la declaración de impuestos ante hacienda.

Y en cuanto a la tasa de desempleo, también era lógico ya que la población es la que con su trabajo genera pagos de impuestos, empleos, etc.

Sin embargo, gracias al análisis matemático, pudimos descartar variables que se pensaban eran lógicas para modelar en comportamiento del PIB, como la de ventas internacionales y las IED.

## **Bibliografía:**

### **Fuentes de datos:**

<https://datamexico.org/es/profile/geo/baja-california-bc#economia-ventas-internacionales>  
<https://www.inegi.org.mx/app/estatal/?ag=07000002#grafica>  
<https://www.inegi.org.mx/app/tabulados/default.aspx?pr=17&vr=7&in=2&tp=20&wr=1&cno=2>  
<https://es.inflation.eu/tasas-de-inflacion/mexico/inflacion-historica/ipc-inflacion-mexico.aspx>  
<https://datamexico.org/es/profile/geo/baja-california-bc#economia-inversion-extranjera>  
<https://datamexico.org/es/profile/geo/baja-california-bc#economia-ventas-internacionales>  
<https://www.inegi.org.mx/app/estatal/?ag=07000002#grafica>

### **Material de apoyo para la investigación:**

[http://www.est.uc3m.es/agrone/ficheros\\_docencia/MetodosRegresion/teoria/slides\\_tema5\\_reducido.pdf](http://www.est.uc3m.es/agrone/ficheros_docencia/MetodosRegresion/teoria/slides_tema5_reducido.pdf)  
<https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/regression/howto/best-subsets-regression/interpret-the-results/all-statistics/#r-sq>  
<https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/model-assumptions/multicollinearity-in-regression/#what-is-multicollinearity>  
<https://support.minitab.com/en-us/minitab/18/help-and-how-to/statistics/tables/how-to/chi-square/goodness-of-fit-test/before-you-start/example/>  
<https://repositorio.comillas.edu/jspui/bitstream/11531/6845/1/TFM000312.pdf>  
<https://www.youtube.com/watch?v=ikNk4dwh8Vk> Bondad de ajuste en minitab  
[https://www.youtube.com/watch?v=xc\\_X9GFVuVU](https://www.youtube.com/watch?v=xc_X9GFVuVU) Palanca e influencia