

Análisis de la relación entre la emisión de gases vehiculares y el kilometraje.

Estudio estadístico de regresión lineal sobre el aumento en la emisión de monóxido de carbono con respecto al kilometraje de un automóvil Fiat Palio 1.4cc

Autores: ALVAREZ PAREDES, Dario Ezequiel (1); MARAZ, Santiago Federico (2); REYNOSO AGÜERO, Franco Martín (3); VALERIANO JURADO, Brian Luciano (4).

Institución: UNIVERSIDAD NACIONAL DE JUJUY - FACULTAD DE INGENIERIA

Dato de contacto: alvarezdarioeze@gmail.com (+543884701294) (1).

aguero64franco@gmail.com (+543885803673) (3)

RESUMEN

En el presente estudio se buscará comprobar si existe una la relación entre la emisión del Monóxido de Carbono (CO) y el kilometraje de un automóvil Fiat Palio 1.4cc mediante los datos obtenidos por el RTO (Revisión Técnica Obligatoria), dicha comprobación se realizará mediante un modelo de regresión lineal simple. Con esto buscaremos encontrar si el desgaste del motor tiene relación con la emisión de los gases de efecto invernadero.

PALABRAS CLAVES

Modelo de Regresión Lineal - Monóxido de Carbono – Fiat Palio - Kilometraje

INTRODUCCIÓN

Lo que observamos diariamente en las calles de nuestra ciudad es un parque automotor muy variado en modelos de distintos años, en la cual los automóviles más longevos siguen circulando. Los cuales se observa que emiten mayor cantidad de gases, distintos colores y olores.

Lo que queremos comprobar con este estudio es ver si el desgaste del automóvil en kilómetros provoca una mayor emisión de gases de efecto invernadero, en este caso el (CO) monóxido de carbono. Para esto se eligió un automóvil que es muy común en las calles de nuestras ciudades, el Fiat Palio 1.4cc con año de fabricación anterior al 2015, ya que estos automóviles tuvieron que ser inspeccionados en la RTO por tener una antigüedad mayor a 3 años.

La recopilación de los datos necesarios se obtuvo mediante la visita al Taller RTO "TECNICAR" seguridad vial de Jujuy, con respecto a la base de datos de la máquina de medición de control de polución de gases cuya función es realizar las mediciones de los gases emitidos por los automóviles que asisten para realizar dicho control.

METODOLOGIA

El método estadístico utilizado para la realización de este informe es el Modelo de Regresión Lineal Simple, el cual nos permite encontrar la relación o desvinculación entre dos variables observadas.

El Método de Mínimos Cuadrados es la herramienta más eficiente para encontrar los estimadores B_0 y B_1 de la regresión lineal simple, con estos valores estimamos la recta de regresión lineal con la cual podremos calcular todo tipo de relación que pueda existir entre dichas variables.

Sujeto a este tipo de comprobación, no está de más la utilización de dichos valores estimativos con programas dedicados específicamente a calcular valores estadísticos como Minitab que nos permiten tener una solución profesional y rápida a nuestras investigaciones.

Introducción teórica

Monóxido de Carbono (CO)

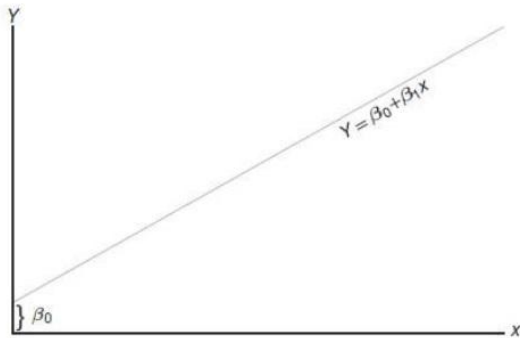
El **monóxido de carbono** es un gas incoloro y altamente tóxico. Puede causar la muerte cuando se respira en niveles elevados. Se produce por la combustión Deficiente de sustancias como gas, gasolina, queroseno, carbón, petróleo, tabaco o madera.

Una persona expuesta, aunque sea a pequeñas cantidades de este gas, pero por periodos continuos, puede morir sin darse cuenta.

De ahí la importancia de aprender a conocerlos, y, sobre todo, del cuidado que se debe tener con el sistema de escape para evitar que una fuga vaya a dar a la cabina.

Aunque el desarrollo tecnológico de los automóviles ha hecho que los sistemas de escape sean más amables con el medio ambiente, gracias a la implementación de los catalizadores, esto no significa que no sean nocivos para la salud humana (si no se toman las medidas de precaución pertinentes).

El concepto de análisis de regresión se refiere a encontrar la mejor relación entre X e Y, cuantificando la fuerza de esa relación, y empleando métodos que permitan predecir los valores de la respuesta dados los valores del regresor x. En muchas aplicaciones habrá más de un regresor, es decir, más de una variable independiente que ayude a explicar a Y.



$Y = \beta_0 + \beta_1 x$,
 β_0 es la **intersección** y β_1 es la **pendiente**

Al no poder calcular el valor real de los parámetros, se estima unos estadísticos con los valores de la muestra. Esos estadísticos serán b_0 y b_1 . Entonces nos queda la Recta de Regresión ajustada.

$$\hat{y} = b_0 + b_1 x \quad (1)$$

Donde Y es el valor estimado, pronosticado o ajustado.

Modelo de los mínimos Cuadrados

Encontraremos b_0 y b_1 , de modo que la suma de los cuadrados de los residuos sea mínima. A menudo la suma de los cuadrados de los residuos se llama suma de cuadrados de los errores alrededor de la línea de regresión y se denota con SCE. Este procedimiento de minimización para estimar los parámetros se llama método de mínimos cuadrados. Por ello, encontraremos b_0 y b_1 de modo que se minimice.

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (2)$$

$$b_0 = \frac{n \sum_{i=1}^n Y_i - b_1 \sum_{i=1}^n X_i}{n} = \bar{Y} - b_1 \cdot \bar{X} \quad (3)$$

El parámetro σ^2 , refleja una variación aleatoria o una variación del error experimental alrededor de la recta de regresión. Si para simplificar ponemos nombres breves a las ecuaciones

$$S_{xy} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (4)$$

$$S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2 \quad (5)$$

$$S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (6)$$

La SCE se puede escribir como sigue:

$$SCE = S_{yy} - b_1 \cdot S_{xy} \quad (7)$$

Un estimador insesgado de σ^2 es:

$$S^2 = \frac{SCE}{n-2} = \frac{S_{yy} - b_1 \cdot S_{xy}}{n-2} \quad (8)$$

Coeficiente de correlación muestral:

El coeficiente de correlación lineal puede ser positivo o negativo. Si r es positivo y tiende a aumentar la pendiente de la recta de mínimos cuadrados (es positiva), en tanto que si r es negativo tiende a disminuir y la pendiente (es negativa).

$$r = b_1 \sqrt{\frac{S_{xx}}{S_{yy}}} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} \quad (9)$$

Prueba de Hipótesis sobre la pendiente:

Para probar la hipótesis nula H_0 de que $\beta_1 = \beta_{10}$ en comparación con una alternativa posible, utilizamos la distribución t con n-2 grados de libertad con el fin de establecer una región crítica y después basar nuestra decisión en el valor de t.

$$T_{cal} = \frac{b_1 - \beta_{10}}{\frac{s}{\sqrt{S_{xx}}}} \quad (10)$$

Una prueba t importante sobre la pendiente es la prueba de la hipótesis:

$$H_0: b_1 = 0 \text{ en comparación con } H_1: b_1 \neq 0$$

Cuando no se rechaza la hipótesis nula la conclusión es que no hay relación lineal significativa entre X e Y.

El rechazo anterior de H_0 implica que hay una relación lineal significativa.

DESARROLLO

Se sospecha que la emisión de monóxido de carbono de los caños de escape de los autos se incrementa con el kilometraje de un automóvil. A fin de verificar dicha hipótesis se registraron las emisiones de monóxido de carbono (%) de 8 autos Fiat Palio 1.4cc, pero de distinto kilometraje. Los resultados fueron:

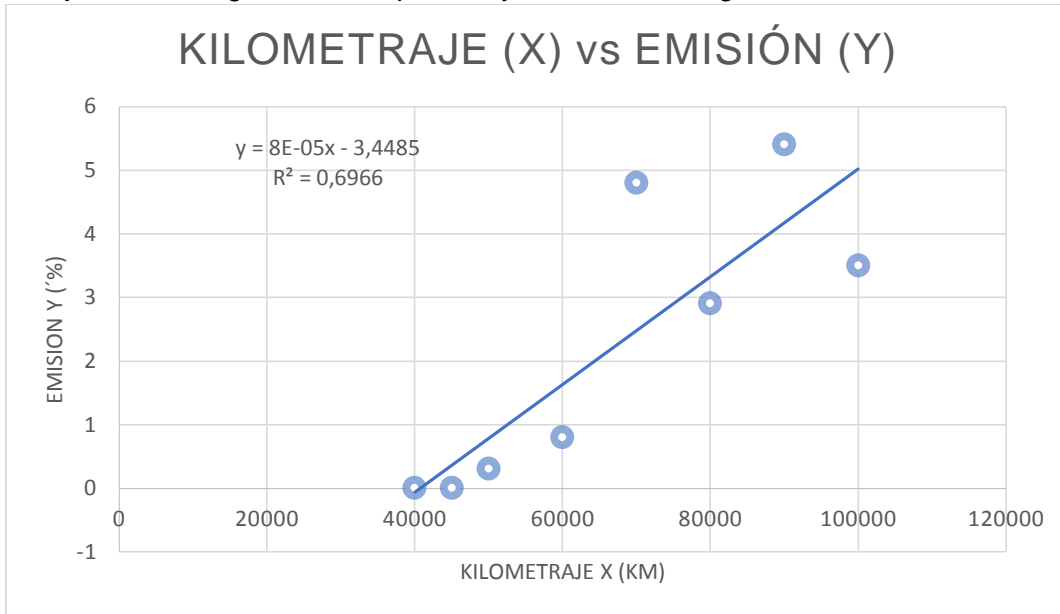
KILOMETRAJE (Km)	40000	45000	50000	60000	70000	80000	90000	100000
EMISIÓN (%)	0	0	0,3	0,8	4,8	2,9	5,4	3,5

- Variable Dependiente natural o de respuesta (Y): es el valor observado, Para este caso serán las emisiones de monóxido de carbono.
- Variable Independiente, predictora o explicativa (X): es la variable cuyo valor lo fijamos, para este caso será el kilometraje del automóvil.

Para encontrar el valor de las constantes primero debemos agrupar los datos en una tabla de frecuencias y calcular unos datos necesarios para poder utilizar las fórmulas.

No se pueden encontrar los valores reales, entonces procederemos a calcular un estimado de las constantes en base a los datos obtenidos en la muestra que es igual a 8 automóviles Fiat Palio 1.4cc.

Dibujamos el Diagrama de dispersión y la Recta de Regresión Lineal.



Obtenemos \bar{X} e \bar{Y} .

$$\bar{X} = \frac{\sum X_i}{n} = \frac{535000}{8} = 66875 \text{ km} \quad \text{e} \quad \bar{Y} = \frac{\sum Y_i}{n} = \frac{17,7}{8} = 2,2125$$

Cálculo empleando fórmulas para el método de mínimos cuadrados.

- Fórmula 4. Suma de Productos Cruzados XY

$$S_{xy} = \sum_{i=1}^n (X_i - \bar{X}) * (Y_i - \bar{Y})$$

$$S_{xy} = 283312,5$$

- Fórmula 5. Suma de los Cuadrados de X:

$$S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S_{xx} = 3346875000$$

- Fórmula 6. Suma de los Cuadrados de Y

$$S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$S_{yy} = 34,42875$$

Cálculo de los valores de los Estimadores muestrales por mínimos cuadrados:

- Fórmula 2. Sumatoria del cuadrado los valores individuales de la variable aleatoria (X) de la muestra.

$$b_1 = \frac{S_{xy}}{S_{xx}}$$

$b_1 = 8,465 \times 10^{-5}$

Representa la variación de la Variable explicada Y (Emisión de Monóxido de Carbono), cuando la variable explicativa X (Kilometraje de un automóvil), varía en una unidad de aumento de kilómetros. La pendiente de la recta es positiva (+) como se puede observar en el gráfico, por lo tanto, a medida que aumenta los kilómetros recorridos, aumenta la emisión de Monóxido de Carbono producidos por el automóvil.

- Fórmula 3. Fórmula de cálculo de la pendiente de la recta de regresión.

$$b_0 = \bar{Y} - b_1 * \bar{X}$$

$b_0 = -3,4485$

El intercepto representa el valor de la variable explicada Y (Emisión de Monóxido de Carbono), cuando la Variable explicativa X (Kilometraje de un automóvil) es igual a cero. Representa el valor de emisión inicial.

- Fórmula 1. Entonces la Recta Muestral Estimada de Regresión es:

$$\hat{Y} = b_0 + b_1 * X_i$$

$\hat{Y} = -3,4485 + 8 \times 10^{-5} * X_i$

- Fórmula 7: Suma de Cuadrados explicada por la Regresión

$$SCE = S_{yy} - b_1 * S_{xy}$$

$SCE = 11,76375$

- Fórmula 9. Fórmula de cálculo del coeficiente de correlación

$$R = \frac{S_{xy}}{\sqrt{S_{xx}} * \sqrt{S_{yy}}}$$

$$R = 0,834661$$

Significa que el 83.46% de la variabilidad de la emisión de monóxido de carbono en gramos está explicada por la variabilidad de los kilómetros de un automóvil. Puede decirse que el ajuste es “suficientemente bueno”.

- Fórmula 8. Fórmula de cálculo de la varianza residual.

$$s_e^2 = \frac{\sum(\hat{Y}_i - Y_i)^2}{n - 2}$$

$$s_e^2 = 1,960625$$

- Desviación estándar muestral residual.

$$Se = 1,40$$

Datos y Cálculos con Excel ®

	kilometraje (X)	Emisión (Y)	X_i^2	Y_i^2	$X_i * Y_i$
	40000	0	1600000000	0	0
	45000	0	2025000000	0	0
	50000	0,3	2500000000	0,09	15000
	60000	0,8	3600000000	0,64	48000
	70000	4,8	4900000000	23,04	336000
	80000	2,9	6400000000	8,41	232000
	90000	5,4	8100000000	29,16	486000
	100000	3,5	10000000000	12,25	350000
\sum	535000	17,7	39125000000	73,59	1467000

Estos valores son necesarios para la estimación de los valores de la regresión lineal mediante el método de mínimos cuadrados por el cual calculamos los valores totales de las mediciones del kilometraje y las emisiones, además el total del cuadrado de ambas variables y el producto de ambas.

\bar{X}	66875
\bar{Y}	2,2125
B ₀	-3,4485
B ₁	0,00008
S _{xx}	3346875000
S _{yy}	34,42875
S _{xy}	283312,5
B ₁ estimado	0,0001
B ₀ estimado	-3,4485
R	0,8346
S _{ce}	10,4464
S ₂	1,7411
S	1,3195
$T_{0,025,6}$	2,447

Tabla 1. Resumen de cálculo de coeficientes del MRLS para la emisión de monóxido de carbono con respecto al kilometraje de un automóvil con planilla Excel®

Comprobación de los cálculos del MRLS empleando software Minitab 18®.

Análisis de regresión: Emisión (%) vs. kilometraje (km)

La ecuación de regresión es

$$\text{Emisión (\%)} = - 3,448 + 0,000085 \text{ kilometraje (km)}$$

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)
1,31949	69,66%	64,60%

Análisis de Varianza

Fuente	GL	SC	MC	F	P
Regresión	1	23,9824	23,9824	13,77	0,010
Error	6	10,4464	1,7411		
Total	7	34,4288			

Resumen de cálculo de coeficientes del MRLS para la emisión de monóxido de carbono con respecto al kilometraje de un automóvil Minitab 18®.

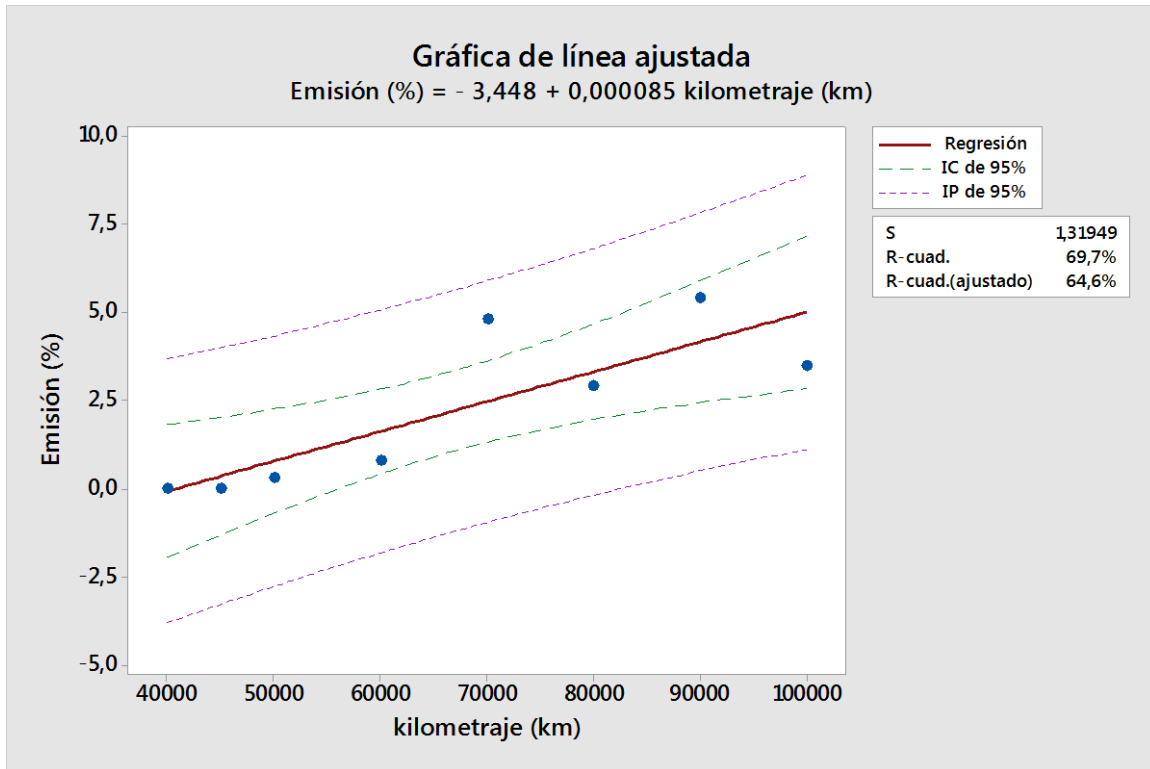


Gráfico.
Recta del MRL
S, IC e IP de 95% utilizando Minitab 18®

Ahora probaremos la hipótesis nula

de que la pendiente de la línea es cero contra la alternativa de que la pendiente no es cero al nivel de significancia del 0.05 y comprobaremos si existe o no una relación entre las 2 variables.

Calculamos también intervalo de confianza del 95% para el parámetro β_1 en la recta de regresión:

$$b_1 - t_{\frac{\alpha}{2}} * \frac{Se}{\sqrt{S_{xx}}} < \beta_1 < b_1 + t_{\frac{\alpha}{2}} * \frac{Se}{\sqrt{S_{xx}}}$$

$$8 \times 10^{-5} - 2,447 * \frac{1,40}{\sqrt{3346875000}} < \beta_1 < 8 \times 10^{-5} + 2,447 * \frac{1,40}{\sqrt{3346875000}}$$

$$2,078 \times 10^{-5} < \beta_1 < 1,3922 \times 10^{-4}$$

Según el IC del 95% podemos decir que el verdadero valor de la pendiente estará "comprendido" entre los valores de $2,078 \times 10^{-5}$ y $1,3922 \times 10^{-4}$.

Empezamos por plantear las hipótesis.

$$H_0: b_1 = 0$$

$$H_1: b_1 \neq 0$$

Para probar esta hipótesis con un nivel de significancia de 0,05 utilizamos la distribución t con n-2 grados de libertad con el fin de establecer una región crítica.

Al ser H_1 "≠" tendremos 2 regiones críticas.

$$\alpha = 0,05 \quad \frac{\alpha}{2} = 0,025 \quad n=8 \quad t_{0,025;6} = 2,447$$

$$T_{cal} < -t_{\frac{\alpha}{2};n-2} \quad \text{ó} \quad T_{cal} > t_{\frac{\alpha}{2};n-2}$$

$$T_{cal} < -t_{0,025;6} \quad \text{ó} \quad T_{cal} > t_{0,025;6}$$

$$\boxed{T_{cal} < -2,447} \quad \text{ó} \quad \boxed{T_{cal} > 2,447}$$

Antes de poder calcular T_{cal} vamos a utilizar la desviación estándar muestral (Se) calculada en la Formula 8.

$$Se = 1,40$$

Calculamos T_{cal} con (formula nº 10)

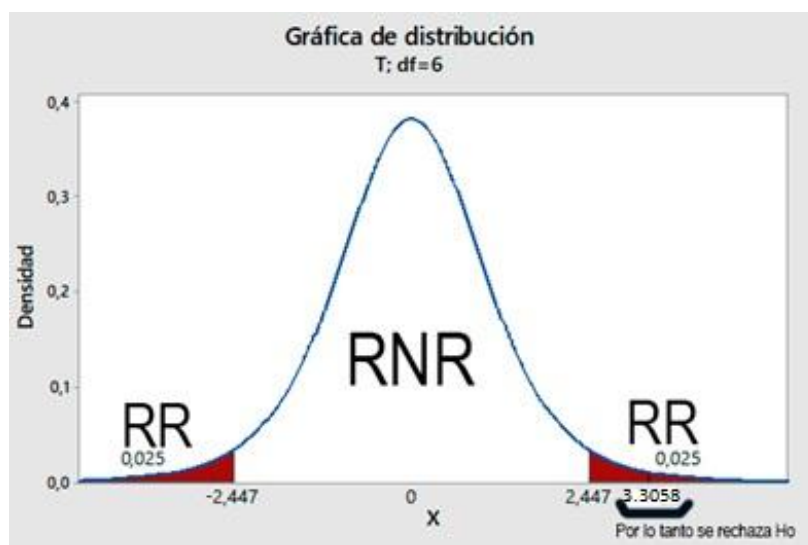
$$T_{cal} = \frac{b_1 - \beta_{10}}{\frac{se}{\sqrt{S_{xx}}}} = \frac{8 \times 10^{-5} - 0}{\frac{1,40}{\sqrt{3346875000}}} = \boxed{3,3058}$$

$$\boxed{3,3058 < -2,447 \text{ FALSO}}$$

ó

$$\boxed{3,3058 > 2,447 \text{ VERDADERO}}$$

Gráfico.
utilizando



Distribución t
Minitab18®

Cae en la región crítica por lo tanto se rechaza la hipótesis nula para un intervalo de significancia de 0,05.

El hecho de que se rechace la Hipótesis nula implica que hay una relación lineal entre la emisión de monóxido de carbono y el kilometraje de un automóvil.

CONCLUSIONES

Una vez terminada la realización de los cálculos y las interpretaciones de los gráficos pertinentes podemos afirmar que dicha hipótesis es cierta debido a que existen relación lineal entre ambas variables comprobadas con los valores obtenidos en el coeficiente de correlación y el rechazo de la Hipótesis nula (h_0).

Esto significa que existe una relación en donde el nivel de emisiones va aumentando con el kilometraje del automóvil. Puede deberse al desgaste de las piezas internas del motor y la falta (o mal) mantenimiento del vehículo, ya que en nuestra experiencia en el taller pudimos comprobar como automóviles de otras marcas, con mucho kilometraje, pero en un estado de conservación excelente poseían niveles de emisión similares a la de automóviles con mucho menor kilometraje.

De igual modo hemos observado como autos considerados “nuevos” contaminan, entonces también concluimos que todos los coches contaminan, uno más que otros, pero todos contaminan al fin. Debemos buscar alternativas tecnológicas, como los coches eléctricos, o métodos que nos permitan reducir las emisiones de dicho gas toxico para así poder aumentar la calidad de vida las personas.

BIBLIOGRAFIA

“*Probabilidad y Estadística para Ingeniería y Ciencias*” Walpole, Myers, Myers y Ye. Octava Edición. Ed. Pearson/ Prentice Hall. 2007.

“*Teoría Regresión Lineal*” Catedra Probabilidad y Estadística 2018.

“*Regresión Lineal Modelo envases PET*” Mautino, Gisella Carla.

“*Probabilidad y Estadística TEORIA y 760 problemas resueltos*” Murray, Spiegel.

<https://www.eltiempo.com/archivo/documento/MAM-3223819>

https://es.wikipedia.org/wiki/Mon%C3%B3xido_de_carbono

Base de datos TALLER RTO “*Tecnica*” Seguridad Vial Jujuy – Ing. Sarmantano, Pablo.

SOFTWARE UTILIZADO:

Microsoft Office 2019®

Minitab 18®