

Modelo de Regresión Estadísticos para el Estudio de Crudos Pesados

Vega¹, Judith M.; Martinez², Julieta y Erdmann³, Eleonora.

(1) Instituto de Investigaciones para la Industria Química - INIQUI-CONICET, Consejo de Investigaciones- CIUNSa, Facultad de Ingeniería, Universidad Nacional de Salta. maju.iq@gmail.com

(2) Facultad de Ingeniería, Universidad Nacional de Salta, Consejo de Investigaciones- CIUNSa.

(3) Instituto de Investigaciones para la Industria Química – INIQUI (UNSa-CONICET), Instituto Tecnológico de Buenos Aires- ITBA Ciudad de Buenos Aires, Argentina.

Resumen. Con la gran demanda y los altos precios del petróleo y estando en declinación la producción de la mayoría de los yacimientos de petróleo convencionales, la atención de la industria en muchos lugares del mundo se está desplazando hacia la explotación de petróleo pesado y reutilización de los residuos de cola de Topping.

En este trabajo se hará uso de diferentes programas para obtener un modelo de regresión que permitirá estimar la variación de la concentración en función de la temperatura y presión en un diagrama de fases para diferentes crudos pesados. La importancia de esto radica en el hecho que estos datos de laboratorio, datos de destilación TBP (True Boiling Point) de un crudo, son datos incompletos para todo el rango de porcentajes de destilación, generalmente solo llegan hasta un 70-80% en volumen o en peso de destilado. En la destilación al vacío de petróleo de crudos pesados, es necesario contar con datos de destilación atmosférica a altas temperaturas para poder utilizar estos últimos valores, valores que son necesarios para el análisis una vez que se han transformado las curvas de TBP a EFV (Equilibrium Flash Vaporization) a diferentes presiones. Una vez obtenidas las curvas EFV a diferentes presiones, se realiza un Diagrama de Fases, el cual es de mucha importancia para el estudio de la factibilidad, aprovechamiento y rendimiento de un crudo para las empresas, ya que de él

podemos obtener ya sea la presión y/ o temperatura de los productos obtenidos de la destilación, destinados a la venta.

Palabras clave: Regresión, Modelos Estadísticos, Crudos Pesados.

1. Introducción

Una simulación es un proceso en el cual se diseña un modelo de un sistema real llevándose a cabo experiencias con ella, en si el objetivo de una simulación es comprender básicamente el comportamiento de un sistema frente a diversas situaciones, y de ellas evaluar nuevas estrategias siempre dentro de los límites que se imponen por un criterio o conjunto de ellos. La simulación nace en el instante en el cual las ecuaciones no son capaces de adaptarse a un modelo real, ya sea por ausencia de las mismas o por la no disponibilidad de los datos para aplicarlas, por esto es que los resultados obtenidos no son precisos sino que dan un rango de posibles resultados a ser interpretados de distintas maneras, llenando de esta forma el vacío que existiría si no se contara con esta herramienta.

Los métodos estadísticos de regresión diseñan un modelo –ecuación– de un sistema, con el cual podemos llevar a cabo diferentes experiencias con él creando distintos escenarios. La razón teórica o práctica para creer que una determinada variable es causalmente dependiente de una o más variables distintas. Si hay suficientes observaciones empíricas sobre estas variables, el análisis de regresión es un método apropiado para describir la estructura y sentido exacto de esta asociación. El objetivo de la técnica de regresión es establecer la relación estadística que existe entre la variable dependiente (Y) y una o más variables independientes (x_1, x_2, \dots, x_n). Para poder realizar esto, se postula una relación funcional entre las variables. Debido a su simplicidad analítica, la forma que más se utiliza en la práctica es la relación lineal, Ec. (1):

$$y = \beta_0 + \beta_1 * x_1 + \dots + \beta_n * x_n \quad (1)$$

Donde los coeficientes β_0 y $\beta_1 \dots \beta_n$ son los parámetros que definen la variación promedio de y , para cada valor de x . Estimada esta función teórica a partir de una muestra de observaciones sobre las variables y y x , cabe preguntarse qué tan bien se ajusta a la distribución real. En el análisis de regresión, estas estimaciones se obtienen por medio del método de mínimos cuadrados, logradas estas se evalúa la bondad de ajuste y significancia estadística de los resultados.

En petroquímica, específicamente en el área de destilación del petróleo, son de mucha importancia los estudios previos, minuciosos y detallados del crudo. Ya que de ello depende la factibilidad de una empresa.

Para poder correlacionar los distintos puntos de ebullición con los porcentajes de vaporizado para cada cambio de presión de los distintos productos, se construye un diagrama de fases con las temperaturas EFV760 - Equilibrium Flash Vaporization a una presión de 760mmHg y EFV10 - Equilibrium Flash Vaporization a una presión de 10mmHg. El diagrama de fases para hidrocarburos está representado por la variación de las temperaturas con la presión para un porcentaje de destilado constante la que tiene una variación lineal cuando se representa logaritmo de las presiones versus la inversa de la temperatura, lo que también se puede correlacionar por medio de una longitud que está en función de las temperaturas EFV760 y EFV10 (2).

El software ASPEN HYSYS V8.3, utilizado para simular en estado estacionario la destilación del crudo, a través de cálculos internos resuelve automáticamente el diagrama de fases, en comparación con la dificultad que representan los cálculos manuales del mismo, tal como fueron explicitados posteriormente, pero sucede que para poder hacer uso de los resultados que nos arroja este software, necesitamos comparar con los resultados obtenidos de forma manual.

2. Desarrollo

Se trabajó con tres crudos cuyos datos y características son, Tabla 1:

Tabla 1. Crudos en estudio

Procedencia	Ojeda	Labuan	Grane
	Venezuela	Malasia	Noruega
Densidad API°	19	29,92	19
Densidad 60/60	0,94	0,8761	0,9403
Peso Molecular (PM)	460	347	465
Kuop	11,67	11,6	11,5

Por medio de la aplicación del método de Maxwell y Edmister y programando estos, se obtuvo las curvas EFV760 a partir de la curva TBP a 760mmHg, para cada crudo en estudio, Tabla 2:

Tabla 2. Datos de curvas TBP y EFV.

Crudo Ojeda			Crudo Labuan			Crudo Grane		
%D	TBP 760	EFV 760	%D	TBP 760	EFV 760	%D	TBP 760	EFV76 0
0	550	732	0,93	130	344,05	0,68	59	378,26
10	710	790	2,9	158	351,9	1,20	149	404,21
20	768	825	4,84	194	362,12	1,69	194	417,81
30	810	855	10	248	376,7	3,7	302	455,05
40	840	882	14,04	284	386,54	5,36	356	477,99
50	895	917	16,98	311	393,82	12	464	536,09
70	1130	1030	20	338	400,88	28,5	608	630,78
			22,33	365	408,31	40,2	707	697,63
			27,58	428	425,72	48,7	788	749,14
			30	442	428,9	69,7	977	872,66
			30,25	446	429,98	100	1022	973,59
			40	482	435,47			
			44,81	500	438,26			
			60	590	458,33			
			68,13	626	464,94			
			70	635	466,7			
			74,07	662	473,06			
			79,66	716	487,55			
			80	721	489,15			
			90	809	512,2			
			92,99	842	521,14			
			95,83	930	549,2			

Y transformando las curvas EFV a 760 mmHg en EFV a 10 mmHg por medio del Método de Maxwell, tenemos por ejemplo para el crudo Ojeda, Tabla 3:

Tabla 3. Curva EFV760 y EFV10 para el crudo Ojeda.

%D	EFV760	EFV10
0	732	403
10	790	461
20	825,56	496,56
30	855,6	526,6
40	882,8	553,8
50	917,7	588,7
70	1030	701

2.1. Construcción del Diagrama de Fases

Una sustancia a medida que cambia la presión cambia su punto de ebullición. Podemos caracterizar a las mezclas a través del punto de ebullición medio, pero para que una mezcla quede definida totalmente hay que especificar su composición. Con las curvas EFV760 y EFV10 construimos el diagrama de fase para el crudo reducido, la construcción del mismo se efectúa por medio de cada temperatura y presión de trabajo de las curvas EFV correspondientes con su porcentaje de destilado constante, para lo cual se hace necesario representar el $\log p$ vs. $1/T$, pero como la escala de temperatura no es lineal se deben calcular las mismas en función de la longitud definida por la Ec. (2) donde T_1 es la temperatura inicial de la EFV10 y T_2 es la temperatura final de la EFV760.

$$L \text{ (mm origen)} = \text{long. total} \frac{\left(\frac{1}{T_1} - \frac{1}{T}\right)}{\left(\frac{1}{T_1} - \frac{1}{T_2}\right)} \quad (2)$$

Desarrollando la Ec. (2) para invertir la recta no lineal de $1/T$, sacamos logaritmos de las presiones entre temperatura y temperatura, obteniéndose la Ec. (3) que nos facilita el cálculo de cada recta, donde p_1 y p_2 son las presiones 760 y 10 mmHg y p la presión que uno fija su intervalo para calcular las temperaturas para cada porcentaje de destilado. Esta ecuación es la utilizada para obtener cada punto de cada recta del diagrama de fases en el software utilitario Excel.

$$\left(\frac{T_2-T_1}{\log_{10} p_2 - \log_{10} p_1}\right) * \log_{10} p - \left(\frac{T_2-T_1}{\log_{10} p_2 - \log_{10} p_1}\right) * \log_{10} p_1 + T_1 \quad (3)$$

Calculo del diagrama de fases, Tabla 4:

Tabla 4. Diagrama de fases para crudo Ojeda

p	T0%	T10%	T30%	T50%	T70%
10	445	493	545	600	690
60	563	615	674	732	831
110	603	657	717	777	879
160	628	683	744	804	909
210	646	701	764	824	930
260	660	716	779	840	947
310	672	728	792	853	961
360	682	738	802	864	973
410	690	747	812	874	983
460	698	755	820	882	992
510	705	762	827	890	1000
560	711	768	834	897	1008
610	716	774	840	903	1015
660	722	779	846	909	1021
710	727	784	851	914	1027
760	731	789	856	919	1032

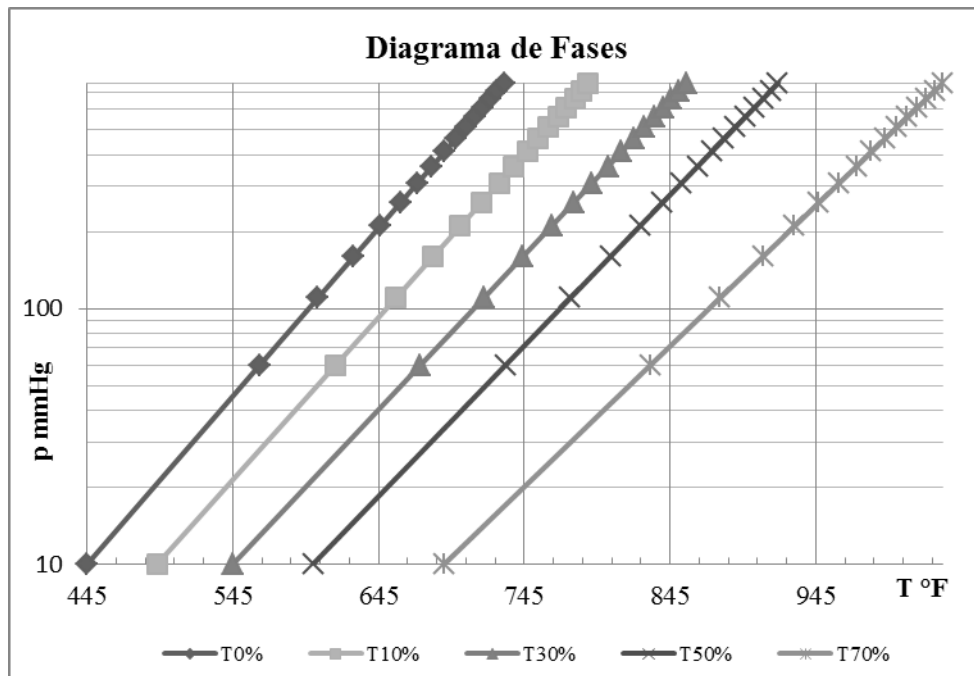


Fig. 1. Diagrama de Fases para crudo Ojeda.

Como se puede apreciar en la Fig. 1, un diagrama de fases es único para cada crudo, y de él podemos obtener la variación de la temperatura de ebullición en función de la presión de vaporizado constante o la presión de operación en función de la temperatura y la concentración.

En la Tabla 4 podemos observar el cálculo realizado para obtener el diagrama de fases de la Fig. 1. De la misma manera se construye el diagrama de fases para cada sección de la torre obteniendo las temperaturas y presiones internas de la misma.

Por lo antes desarrollado observamos que al trabajar con distintos productos a distintas presiones, se necesita saber a qué temperatura se puede obtener el mismo porcentaje de destilado de algunos de los productos a esas distintas presiones. Por ejemplo: si queremos saber cuál es la temperatura que necesito para vaporizar GOLV a 760 mmHg nos fijamos en la curva entrando con el porcentaje y sacando la temperatura; pero si queremos saber esa temperatura pero a otras presiones, necesitamos conocer una relación entre la presión de trabajo y la temperatura de vaporización a distintos porcentajes de vaporizado, eso se obtiene del diagrama de fases.

3. Métodos Estadísticos

Para desarrollar el trabajo estadístico, además de Microsoft Excel se utilizaron los programas “R_x64 for Windows” y “IBM SPSS Statistics 20”, ambos softwares son muy utilizados para el análisis estadístico de datos. En el caso del software R (Lenguaje R) el mayor inconveniente que podría presentar es el hecho de funcionar mediante comandos, lo que para algunos usuarios puede resultar engorroso, pero para solventar esta dificultad existe un paquete llamado R Commander que permite utilizar R sin tener que escribir los comandos. Con respecto a IBM SPSS es un programa estadístico muy usado, más conocido por su capacidad para trabajar con grandes bases de datos y una interface sencilla para la mayoría de los análisis.

En principio se buscó por medio de la regresión lineal simple una relación entre el porcentaje de destilado (variable dependiente llamada C en SPSS e Y en R) y las temperaturas EFV en un rango de 50°F desde 0 a 1000°F (variable independiente T) a las presiones desde 10 a 760 mmHg en un rango de 150 mmHg para cada crudo

(variable independiente P) haciendo diferencia entre los tres crudos utilizados (variable dependiente COjeda ($CO = Y$) donde están todos los porcentajes de destilado de cada crudo ordenados de manera descendente en relación a la Temperatura y Presión de destilación, CGrane (CG) primera variable indicadora que toma el valor 1 cuando le corresponden el destilado y 0 cuando no, CLabuan (CL) segunda variable indicadora, similar a la primera, en las Tabla 5, 6 y 7 podemos ver cómo sería la tabla de estos factores:

Tabla 5. EFV a diferentes presiones, crudo Ojeda

EFV	Ojeda (0)					
	p0 10	p1 160	p2 310	p3 460	p4 610	p5 760
450	3,39					
500	22,34					
550	52,96					
600	73,35					
650	86,95	10,87				
700	95,35	41,18	8,46			
750		67,74	39,08	15,87	1,66	
800		85,00	67,08	48,79	30,99	15,96
850		95,45	85,35	74,18	62,36	50,02
900			96,21	90,29	83,48	76,01
950				99,28	96,03	92,22

Tabla 6. EFV a diferentes presiones, crudo Grane

EFV	Grane(1)					
	p0 10	p1 160	p2 310	p3 460	p4 610	p5 760
0	0,91					
50	1,40	0,24				
100	2,26	0,56	0,20			
150	3,86	0,96	0,54	0,27	0,07	
200	6,74	1,53	0,95	0,64	0,41	0,23
250	11,54	2,49	1,56	1,12	0,83	0,61
300	18,80	4,18	2,58	1,86	1,43	1,13
350	27,80	7,11	4,34	3,12	2,42	1,94
400	36,77	11,84	7,35	5,30	4,10	3,31
450	45,07	18,70	12,17	8,92	6,96	5,65
500	53,65	27,05	19,01	14,48	11,54	9,49
550	66,01	35,76	27,28	21,90	18,09	15,26

600	96,00	44,48	36,01	30,39	26,15	22,79
650		54,27	45,00	39,27	34,92	31,36
700		68,46	55,50	48,89	44,23	40,51
750		96,83	70,97	61,18	55,22	50,85
800				81,11	71,09	64,64
850						87,48

Tabla 7. EFV a diferentes presiones, crudo Labuan

EFV	Labuan(2)					
	p0 10	p1 160	p2 310	p3 460	p4 610	p5 760
50	2,11					
100	9,69					
150	18,79					
200	29,21	4,70				
250	45,11	12,98	5,79	1,66		
300	62,25	22,58	14,34	9,29	5,66	2,88
350	75,91	35,26	24,52	18,59	14,37	11,05
400	86,56	51,75	38,13	30,27	24,98	21,00
450	94,03	67,72	54,81	45,72	38,99	33,83
500	98,10	80,67	70,41	62,40	55,72	50,04
550		90,65	83,10	76,84	71,41	66,52
600		96,68	92,45	88,28	84,27	80,51
650			97,98	95,72	93,46	91,14
700					98,97	97,68

Lo primero que debemos hacer es plantear nuestras hipótesis de investigación, las cuales son:

- H1: La Temperatura (EFV), la presión (p) y las variables indicadoras porcentaje de destilado del Crudo Grane (CG) y Crudo Labuan (CL), predicen el porcentaje de destilado de las muestras Y (Porcentaje de destilado del Crudo Ojeda, que recordemos contiene todos los porcentajes de los tres crudos en estudio).
- H0: La T, p, CG y CL NO predicen la cantidad de porcentaje destilado Y (CO).

En segunda medida vamos a realizar una tabla de Correlaciones para ver qué tan vinculadas/ asociadas están las variables entre sí, Tabla 8:

Tabla 8. Correlaciones

		%CO (Y)	T (X1)	p (X2)	%CG (X3)	%CL (X4)
% de dest Crudo Ojeda (%CO)	Corr. de Pearson	1	,707**	-,041	-,462**	,265**
	Sig. (bilateral)		,000	,598	,000	,001
	N	169	169	169	169	169
Temperatura Flash (T)	Corr. de Pearson	,707**	1	,307**	-,254**	-,203**
	Sig. (bilateral)	,000		,000	,001	,008
	N	169	169	169	169	169
Presiones de trabajo (p)	Corr. de Pearson	-,041	,307**	1	,029	-,003
	Sig. (bilateral)	,598	,000		,705	,971
	N	169	169	169	169	169
% de dest Crudo Grane (%CG)	Corr. de Pearson	-,462**	-,254**	,029	1	-,690**
	Sig. (bilateral)	,000	,001	,705		,000
	N	169	169	169	169	169
% de dest Crudo Labuan (%CL)	Corr. de Pearson	,265**	-,203**	-,003	-,690**	1
	Sig. (bilateral)	,001	,008	,971	,000	
	N	169	169	169	169	169

** . La correlación es significativa al nivel 0,01 (bilateral).

En la Tabla 8, tabla de correlaciones podemos ver la correlación que existe entre cada variable, antes de hacer un análisis de esta tabla tenemos que tener en cuenta que tanto la temperatura como la presión de trabajo son variables termodinámicamente asociadas y que el hecho de haber definido las variables indicadoras %CG y %CL en relación con %CO (Y) fue para facilitar el modelo, por lo que comprensiblemente estarán vinculadas entre sí.

Con lo antes aclarado es que solo vamos a analizar lo que está en negrita en la Tabla 8, que denota la fuerte asociación entre la T y la variable dependiente Y, no así con la variable presión de trabajo p.

La primera metodología de estudio es la de regresión jerárquica para poder encontrar el modelo final, primer modelo (M1) que consta de las cuatro variables independientes (T+P+CG+CL), como segundo modelo (M2) que omite la variable presión de trabajo p (T+CG+CL) y como tercer modelo (M3) el que omite la

variable temperatura T (P+CG+CL), al hacer este análisis se podrá ver que tanto aporta cada una de ellas al modelo final.

Al hacer la prueba de regresión jerárquica, en relación a la prueba de supuestos, hacemos el diagnostico de colinealidad para saber si hay o no colinealidad en los datos y la prueba de Durbin-Watson que nos permite saber si hay independencia en los errores, también se realizó el diagnostico por casos, ya que en algunas situaciones tenemos una distribución de puntos que maso menos configuran una línea pero puede ocurrir que existan puntos por fuera del resto de la regresión y esos puntos pueden modificar los resultados, es por esto que es de utilidad identificar esos casos atípicos y se harán con un análisis cualitativos, y si se dieran estos se estudiará si se pueden retirar estos valores y ver si mejora el modelo:

Tabla 9. Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1 (T+p+CG+CL)	,884	,782	,777	15,83086	,878
2 (T+CG+CL)	,822	,677	,671	19,23602	,813
3 (p+CG+CL)	,468	,219	,205	29,88429	1,068

En la Tabla 9 tenemos los tres modelos en estudio y puede verse que cuando se omite la variable p en el segundo modelo la varianza (el valor de R) disminuye y más aún cuando se coloca como variable independiente a la presión y las variable indicadoras, por lo que según este análisis el modelo que mejor ajustaría sería el que toma en cuenta a todas las variables independientes ya que con ellas estamos explicando el 78.2% de la varianza, cuando sacamos la p del modelo podemos ver que disminuye más del 10% y cuando saco la variable T disminuye mucho más, al 21.9%.

Con respecto a la prueba de supuestos, el resultado de Durbin- Watson nos dice que el análisis o puntaje de esta prueba indica que hay independencia de errores para el modelo tres (1.058) sin embargo el valor de esta prueba para el modelo 1 y 2 nos da un valor que está por debajo de 1 (M1: 0.878 y M2: 0.813) por lo que denota que hay una dependencia de errores con estos modelos.

Para el modelo de regresión probado con las 4 variables independientes se explica el 78.2% de la varianza de la variable dependiente (R^2 : 0.782).

Tabla 10. ANOVA

	Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	147631,563	4	36907,891	147,269	,000
	Residual	41101,044	164	250,616		
	Total	188732,607	168			
2	Regresión	127678,564	3	42559,521	115,018	,000
	Residual	61054,043	165	370,025		
	Total	188732,607	168			
3	Regresión	127678,564	3	42559,521	115,018	,000
	Residual	61054,043	165	370,025		
	Total	188732,607	168			

Para los tres modelos podemos ver en la Tabla 10, que la significancia está por debajo de 0.05 y por lo tanto el ANOVA de los modelos indica que mejoran significativamente la predicción de la variable dependiente, pero el Modelo 1 (que incluye a todas las variables) es el que tiene el mayor valor de F por lo que él es el que mejor ajusta para explicar la variable dependiente.

Tabla 11. Coeficientes

Modelo	Coef. no estandarizados		Coef. tipificados	t	Sig.	Estadísticos de colinealidad	
	B	Error típ.	Beta			Tolerancia	FIV
	-						
(Constante) β_0	40,269	5,922		-6,8	,000		
1							
Temperatura Flash β_1	,150	,007	1,008	20,59	,000	,554	1,804
presiones de trabajo β_2	-,046	,005	-,355	-8,92	,000	,837	1,195
% de destilado Crudo Grane β_3	16,425	4,209	,246	3,90	,000	,335	2,986
% de destilado Crudo Labuan β_4	45,508	4,415	,638	10,38	,000	,347	2,885
2							
(Constante) β_0	-36,777	7,180		-5,12	,000		
Temperatura Flash β_1	,124	,008	,832	15,28	,000	,661	1,512
% de destilado Crudo Grane β_3	6,333	4,926	,095	1,28	,200	,361	2,771
% de destilado Crudo Labuan β_4	35,602	5,192	,499	6,85	,000	,370	2,703
3							
(Constante) β_0	60,133	6,343		9,47	,000		
presiones de trabajo β_2	-,003	,009	-,026	-,37	,711	,999	1,001
% de destilado Crudo Grane β_3	-35,508	6,361	-,531	-5,58	,000	,522	1,914
% de destilado Crudo Labuan β_4	-7,277	6,785	-,102	-1,07	,285	,523	1,912

En la Tabla 11 podemos ver los valores de los coeficientes de los tres modelos en estudio, los cuales nos darían las siguientes ecuaciones, Ec. 4, 5, 6 y 7 respectivamente:

$$Y(\text{CO}) = \beta_0 + \beta_1 * T + \beta_2 * p + \beta_3 * \text{CG} + \beta_4 * \text{CL} \quad (4)$$

$$Y(\text{CO}) = -40.259 + 0.15 * T - 0.046 * p + 16.425 * \text{CG} + 45.506 * \text{CL} \quad \text{M1} \quad (5)$$

$$Y(\text{CO}) = -36.777 + 0.124 * T + 6.333 * \text{CG} + 35.602 * \text{CL} \quad \text{M2} \quad (6)$$

$$Y(\text{CO}) = 60.133 - 0.003 * p - 6.351 * \text{CG} - 7.277 * \text{CL} \quad \text{M3} \quad (7)$$

Con respecto a las columnas de los valores t y Sig. de la tabla anterior, denotan cuanto podemos generalizar la información que tenemos y también el aporte de estos, nos dicen que tan diferente o importante o que tan en cuenta debo tenerlos, por lo que para los coeficientes del M1 las puntuaciones t indican que las variables tenidas en cuenta aportan significativamente al modelo, lo que indica que los valores obtenidos se pueden generalizar a la población, podemos ver que sucede lo mismo con el modelo 2 (M2) salvando el valor de $t:0.2$ para CG ya que se trata de una variable indicadora, y volvemos a notar la falta de ajuste que tiene el modelo 3 (M3) al ver valores significativos mayores a 0.

En los Estadísticos de Colinealidad, el FIV (Factor de Varianza Inflada) interpretamos que dado que estos valores son para los tres modelos menores que 10, indicaría que se cumple el supuesto de no multicolinealidad. En el caso del M1 y M2 los valores son mayores que 1, pero en el M1 son un poquito más grandes que en M2, por lo que puedo suponer que ajustan mejor los coeficientes del modelo 2 (M2).

Por el estudio previo que se hizo de las variables en SPSS, suponemos que tanto el modelo M1 como el modelo M2 pueden definir de manera acertada al conjunto de datos, aun con estos resultados, se realizó en los programas Excel, R y SPSS un análisis del modelo M1 y M2 teniendo en cuenta todo el conjunto de datos y comparando los mismo con un conjunto de datos acotados, esto es debido a que mirando los datos y teniendo en cuenta el estudio previo, la mayor dispersión se encuentra en los valores iniciales y finales de cada porcentaje de destilado, y como comentamos anteriormente, estos pueden ser valores que afecten significativamente.

Acomodando los datos y realizando las regresiones con Excel, SPSS y R para comparar las variables independientes con diferentes combinaciones de las mismas $D(T+P+CG+CL)$, $D(T+CG+CL)$ para todos los valores de la Tabla 5, 6 y 7 y los valores acotados en un rango de porcentaje de destilado mayores que 10% y menores que el 90%, obtenemos las Tablas 12 a 17:

Tabla 12.

	D(EFV+p+CG+CL)					
	Excel	Excel acot	SPSS	SPSS acot	R	R acot
R	0,88	0,88	0,88	0,88		
R ²	0,78	0,78	0,78	0,78	0,78	0,78
R ² correg.	0,78	0,77	0,78	0,77	0,78	0,77

Tabla 13.

	D(EFV+CG+CL)					
	Excel	Excel acot	SPSS	SPSS acot	R	R acot
R	0,82	0,69	0,82	0,69		
R ²	0,68	0,48	0,68	0,48	0,68	0,48
R ² correg.	0,67	0,46	0,67	0,46	0,67	0,46

Tabla 14.

		D(EFV+p+CG+CL)					
		Excel	Excel acot	SPSS	SPSS acot	R	R acot
Regresión	suma de cuadr.	147632,22	45823,48	147631,56	45823,39		
	gl	4,00	4,00	4,00	4,00	4,00	4,00
	med. cuadr.	36908,06	11455,87	36907,89	11455,85		
	F	147,27	82,70	147,27	82,69	147,3	82,69
	Sig	0,00	0,00	0,00	0,00	2.2e-16	
Residual	suma de cuadr.	41100,25	13159,94	41101,04	13160,80	41101,00	13160,8
	gl	164,00	95,00	164,00	95,00	164,00	95,00
	med. cuadr.	250,61	138,53	250,62	138,53	251,00	138,5
Total	suma de cuadr.	188732,47	58983,42	188732,61	58984,19		
	gl	168,00	99,00	168,00	99,00	168,00	99,00

Tabla 15.

		D(EFV+CG+CL)					
		Excel	Excel acot	SPSS	SPSS acot	R	R acot
Regresión	suma de cuadr.	127679,82	28273,01	127678,56	28272,41		
	gl	3,00	3,00	3,00	3,00	3,00	3,00
	med. cuadr.	42559,94	9424,34	42559,52	9424,14		
	F	115,02	29,46	115,02	29,46	115,00	29,46
	Sig	0,00	0,00	,000 ^b	,000 ^b	2.2e-16	1.37e-13
Residual	suma de cuadr.	61052,65	30710,41	61054,04	30711,78	61054,00	30712,0
	gl	165,00	96,00	165,00	96,00	165,00	96,00
	med. cuadr.	370,02	319,90	370,02	319,91	370,00	319,9
Total	suma de cuadr.	188732,47	58983,42	188732,61	58984,19		
	gl	168,00	99,00	168,00	99,00	168,00	99,00

Tabla 16.

M1	Excel	Excel acot.	SPSS	SPSS acot.	R	R acot.	Error tip. SPSS	Error tip.SPSS acotado	Prob. Excel	Prob. Excel acot.
Cte	-40,27	-80,82	-40,27	-80,82	-40,27	-80,82	5,92	8,52	1,85E-10	2,11E-15
T (EFV)	0,15	0,21	0,15	0,21	0,15	0,21	0,01	0,01	2,47E-47	4,35E-31
p	-0,05	-0,07	-0,05	-0,07	-0,05	-0,07	0,01	0,01	8,69E-16	3,54E-19
CG	16,42	20,98	16,42	20,98	16,42	20,98	4,21	3,83	1,39E-04	3,47E-07
CL	45,51	67,09	45,51	67,10	45,51	67,10	4,41	5,33	1,60E-19	6,09E-22

Tabla 17.

M2	Excel	Excel acot.	SPSS	SPSS acot.	R	R acot.	Error tip. SPSS	Error tip.SPSS acotado	Prob. Excel	Prob. Excel acot.
Cte	-36,78	-38,13	-36,78	-38,13	-36,78	-38,13	7,18	11,60	8,34E-07	1,41E-03
T (EFV)	0,12	0,12	0,12	0,12	0,12	0,12	0,01	0,01	2,03E-33	1,27E-13
CG	6,33	4,35	6,33	4,35	6,33	4,35	4,93	5,37	2,00E-01	4,19E-01
CL	35,60	35,82	35,60	35,82	35,60	35,82	5,19	6,92	1,34E-10	1,23E-06

4. Resultados y Conclusiones

En las Tablas 12 a 17, podemos apreciar la gran similitud en los resultados de los tres programas utilizados por lo que un análisis general será suficiente para explicar el mejor resultado obtenido.

Cuando se realizó la primera aproximación a un posible resultado por medio del software SPSS sacábamos como posibles modelos el M1 y en menor medida el M2, sin embargo lo mismo se realizó otro estudio más para poder corroborar esto. Al acotar los valores, lo que hacía suponer que se obtendría un mejor resultado, es decir un mayor ajuste al modelo, se pudo apreciar que esta suposición fue errada ya que tanto los valores de R, R² y R² corregida son centésimas mejores que los datos sin acotar, lo mismo sucede con los resultados obtenidos para el ANOVA de cada suposición, acotado y sin acotar. Donde podemos observar una diferencia en cuanto a los resultados obtenidos por acotar o no los datos en las Tablas 16 y 17, en donde si observamos las probabilidades vemos que la que mejores resultados da es en el modelo M1 para valores acotados.

Luego de varios análisis realizados podemos concluir que el mejor modelo, que ajusta en un 88.14% es el modelo M1 para los datos acotados, Ec.8:

$$Y(CO) = -80.82 + 0.21 * T - 0.07 * p + 20.98 * CG + 67.09 * CL \quad M1acot \quad (8)$$

Referencias

Capilla M. E (2014), *Métodos Estadísticos para el Diseño y Análisis de Investigaciones*, Cartillas Curso de posgrado, Facultad de Ingeniería- UNSa. Argentina.

Ojeda, A.L. et al. (1978), *Destilación Aplicada a Sistemas de Hidrocarburos*. YPF-PGM, Tomos I y II.

SEQC (2007), *Análisis de la Varianza ANOVA*. Cartilla on-line.

Special Report the Hydrocarbon Processing (February 2010), *Consider new methods for bottom of barrel processing- Advanced methods use molecule management to upgrade heavy ends*, Pgs 35-38.