ANEXO II. TRATAMIENTO DE DATOS.

Los valores de cada una de las características estudiadas para seguir la evolución de la calidad de las distintas muestras a estudio, con el tanto por ciento de agua y con el tiempo de almacenamiento, se recogen en las tablas presentadas en el anexo I. Corresponden al valor promedio obtenido a partir de los resultados correspondientes a tres alícuotas de cada muestra; entre paréntesis se muestra el valor correspondiente de la desviación estándar, calculada a partir de la fórmula:

$$SD = \sqrt{\frac{\sum\left(Xi - \overline{X}\right)^2}{n-1}}$$, en donde $\bar{x}$ representa el valor promedio.

Se ha estudiado, en el apartado III.3, la existencia de correlación lineal entre parejas de algunas de las variables estudiadas, pues consideramos que si disponemos de dos series de datos emparejadas, es útil comprobar si ambas están relacionadas, y en ese caso encontrar la expresión de dicha relación. Si la ecuación que las relaciona es una recta, decimos que entre ellas existe correlación lineal. En el caso de no encontrar correlación lineal entre la variabilidad de dos características no quiere decir que no exista entre ellas otro tipo de correlación.

La cuantificación de la fuerza de la relación lineal entre dos variables cuantitativas, se estudia por medio del cálculo del coeficiente de correlación de Pearson (r). Este coeficiente oscila entre -1 y 1. Un valor de 1 indica una relación lineal positiva perfecta. Un valor próximo a cero indica que no existe relación lineal entre las dos variables. Y en general se considera que valores absolutos de r mayores que 0,80 indican una correlación lineal importante, aunque esto depende del número de parejas de datos utilizados y del nivel de seguridad elegido.

El coeficiente de correlación mide sólo la relación con una línea recta. Si las variables tienen una relación curvilínea fuerte no se detecta a través de éste, por lo que el primer paso que haremos en el estudio de la correlación será la representación gráfica.

Indicar que correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación.

El coeficiente de Pearson r relaciona la covarianza con el producto de las desviaciones típicas de cada variable:

$$r = \frac{Sxy}{Sx.Sy}$$

El valor de r, como ya dijimos, puede variar de -1 a 1 con los significados siguientes:

- -1/1 indica correlación lineal negativa/positiva perfecta.
- -0.9/0,9 correlación negativa/positiva muy fuerte.
- -0.75 correlación negativa moderada.
- -0.5 correlación negativa media.
- -0.1 correlación negativa débil.
- 0,0 no existe correlación entre las variables.

Tras realizar el cálculo del coeficiente de correlación de Pearson (r) determinamos si dicho coeficiente es estadísticamente diferente de cero. Para dicho cálculo se aplica un test basado en la distribución de la t de Student**.**

⇨ Si el valor del r calculado es mayor que el producto del error estándar multiplicado por la t de Student con n-2 grados de libertad, decimos que el coeficiente de correlación es significativo (n es el número puntos).

El nivel de significación viene dado por la decisión que adoptemos al buscar el valor en la tabla de la t de Student. Trabajaremos con un nivel de seguridad del 95 % ($p<0,05$).

Cálculo del test de hipótesis de r:

Conocido r, se calcula el error estándar: Error = $\sqrt{\frac{1-r^2}{n-2}}$

Para n-2 grados de libertad y $p < 0,05$ se determina la t Student a partir de la tabla 47 Para nuestro caso, con $p< 0,05$ y 4 grados de libertad obtenemos, a través de la tabla, el valor 2,776.

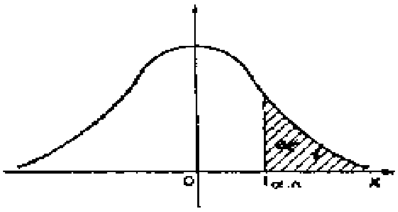
Significado de r y R^2:

El coeficiente de correlación lineal **r** indica el grado de relación existente entre las dos variables. Si elevamos al cuadrado el coeficiente de correlación obtendremos el coeficiente de determinación **R^2,** que indica el porcentaje de la variabilidad de los datos que se explica por la asociación entre las dos variables.

⇨ *Por ejemplo un R^2 = 0.95, significa que el valor de la variable dependiente podrá ser predecido en un 95% de las casos por el valor de la variable independiente.*

Por lo tanto el valor de **R^2** indica el porcentaje de variabilidad de los valores de Y que pueden ser explicados en función de la variabilidad de los valores de X.

Para cada una de las series estudiadas en este trabajo mostraremos la gráfica de puntos de las dos variables objeto de estudio, daremos el valor del r correspondiente junto con el valor de la p utilizada, y se indicará además el número de observaciones realizadas.



gl \ α/2	0,40	0,30	0,20	0,10	0,050	0,025	0,010	0,005	0,001	0,0005
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,192	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,255	0,259	0,851	1,303	1,648	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262	3,495
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,254	0,256	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Tabla 47.- Distribución t de Student

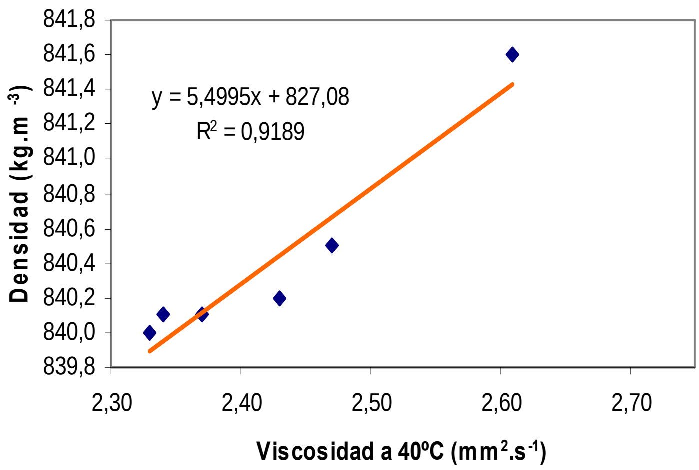
Los cálculos de promedios, desviaciones estándar, así como las gráficas de las distintas correlaciones se han realizado utilizando las herramientas disponibles en el programa de Microsoft Office, con Excel 2003.

Del tratamiento de datos con dicho programa se obtienen las gráficas de correlación correspondientes así como la recta de regresión y el valor del coeficiente de determinación R^2. A partir de éste se ha calculado el coeficiente de correlación r, y como se ha indicado anteriormente se han realizado los correspondientes test de hipótesis.

A modo de ejemplo se detalla el cálculo realizado para una de las series de muestras estudiadas.

⇨ ***Correlación densidad/viscosidad para muestras de gasóleo A con el tiempo de almacenamiento.***

μ ($mm^2\ s^{-1}$)	d ($kg\ m^{-3}$)	**0A**
2,33	840,0	**0A1**
2,34	840,1	**0A2**
2,37	840,1	**0A3**
2,43	840,2	**0A4**
2,47	840,5	**0A5**
2,61	841,6	**0A6**



⇨ A partir del valor de R^2, proporcionado por Excel, calculamos el coeficente de Pearson, r = 0,96.

$$\text{Error stándar} = \sqrt{\frac{1-r^2}{n-2}} = \mathbf{0{,}14}$$

- ⇨ Para n-2 = 4 grados de libertad y $p<0,05$; $\alpha/2 = 0,025$, se obtiene a través de la tabla de la t de student el valor 2,78.
- ⇨ Como 0,96 > 2,78 . 0,14, se puede asegurar que el coeficiente de correlación es significativo $p< 0,05$.