**Tarea 2 STATA**

El gusto de queso madurado está relacionado con la concentración de varios productos químicos en el producto final.  Esta [base de datos](http://pedinta.cl/mod/resource/view.php?id=6468) contiene la una muestra de 30 quesos maduros Cheddar.  Las variables son:

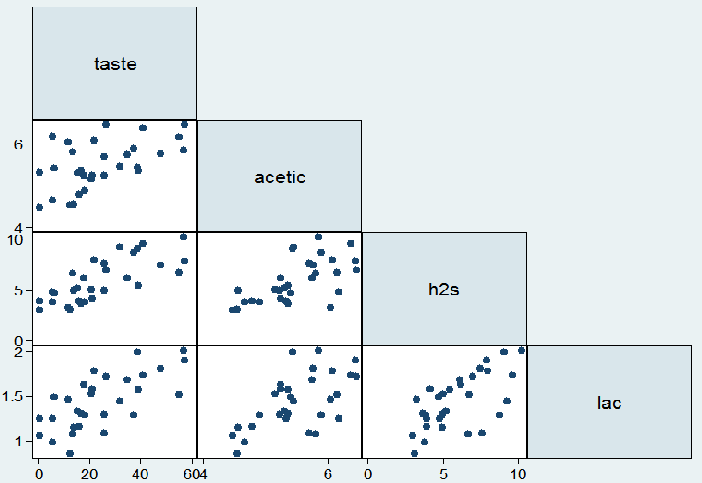
1. Case: Numero de la muestra
2. Taste: Puntaje subjetivo del gusto, obtenido como el promedio de varios catadores.
3. Acetic: logaritmo natural de concentración ácido acético
4. H2S: logaritmo natural de la concentración de Natural de sulfuro de hidrogeno
5. Lac: Concentración  de ácido láctico

Verificar si existe asociación entre el puntaje del gusto con las concentraciones de los productos químicos mencionadas arriba.

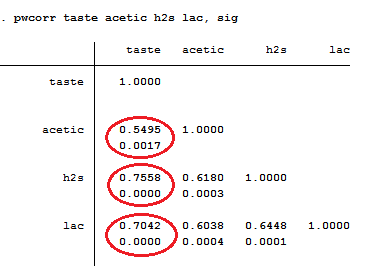
P.D.: Realizar la regresión y verificar todos los supuestos vistos en clases con sus interpretaciones.

1. **Diagrama de dispersión**. Observar si hay relación lineal entre la variable dependiente con cada una de las independientes. Se observa una tendencia lineal entre taste y las variables independientes.



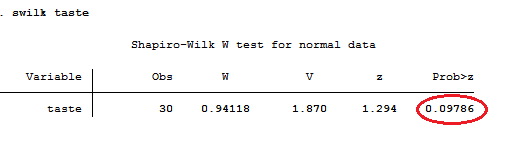


1. **Supuesto de Correlación.**



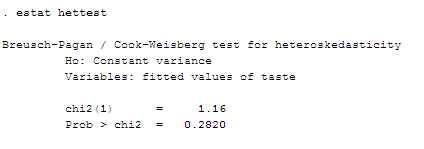
Con una confianza del 95 %, existe una correlación lineal entre la variable dependiente con cada una de las variables independientes. Estas tres correlaciones son positivas.

1. **Comprobar supuesto de Normalidad**.



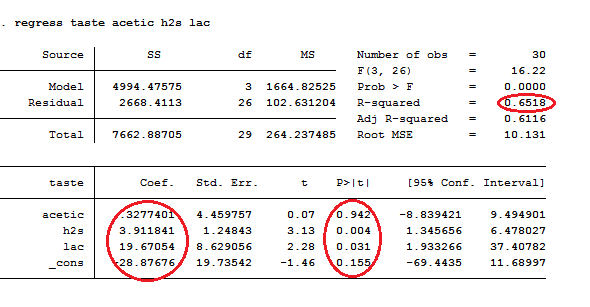
Test S- Wilk no rechaza H0, por lo que no se rechaza que los residuos tengan una distribución normal, y se continua con el procedimiento.

1. **Comprobar supuesto de Homocedasticidad de la varianza.**

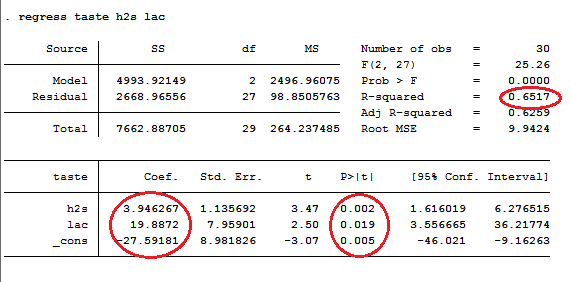


No se rechaza Ho, por lo que no se rechaza igualdad de varianzas. (Este test se calculó después del cálculo de regresión lineal, pero por cuestión de orden de los supuestos se colocó antes del cálculo de regresión).

1. **Regresión lineal múltiple**.

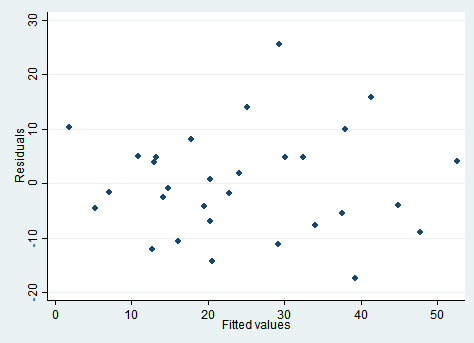


* **R- cuadrado** **indica que:** un 65 % de la variabilidad de la variable “taste” se explica con el modelo, el otro 35% es explicado por el azar.
* **Las significancias de los coef indican que:** Con un nivel de significación del 5% hay evidencia para rechazar H0, paras las variables independientes h2s y lac, esto significa que h2s y lac son variables significativas para explicar la variabilidad de taste. En cuanto a la variable acetic, no se rechaza H0, por lo que esta variable no es significativa para explicar la variabilidad de taste.
* **Los coeficientes indican que:** por cada unidad de aumento de la concentración del h2s, taste aumenta 3.92 en el puntaje. Por otro lado, por cada unidad de aumento en la concentración de lac, taste aumenta 19,67 en el puntaje.
* Al demostrar que la variable acetic no influye en el modelo como se mencionó anteriormente, se vuelve a calcular la regresión sin considerar esta variable.



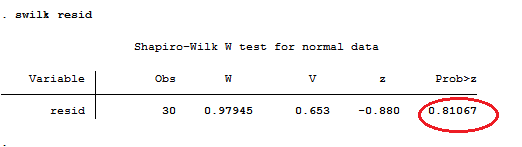
* No existe gran diferencia en R- cuadrado, ni en los coeficientes. En cuanto a los ajustes de las variables, estas mejoran al sacar la variable acetic.
* El modelo de regresión sería y= -27,59 + 3.95x1 + 19,89x2.

1. **Análisis de residuos.**



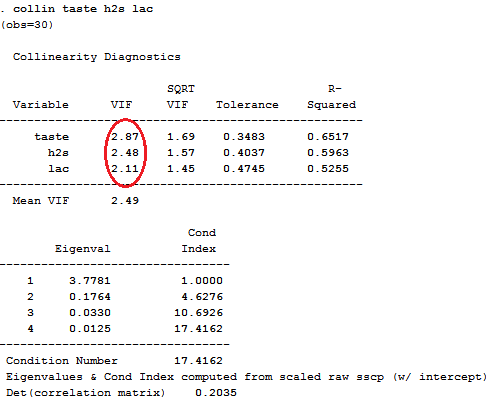
Los residuos se reparten aleatoriamente alrededor del cero y solo uno se aleja del cero en comparación de los demás. Esto demuestra que el modelo de regresión es lineal y se ajusta correctamente a la nube de puntos.

1. **Normalidad de residuos.**



No se rechaza H0, los residuos tienen una distribución normal.

1. **Colinealidad.**



Considerando solo las variables que influyen en el modelo, la colinealidad (vif) es menor que 10 por lo tanto no hay problemas de colinealidad.

Por lo tanto, existe asociación entre el puntaje del gusto con las concentraciones de sulfato de hidrógeno y con la concentración de ác. Láctico. Mientras que el ác acético no es significativo para el puntaje del gusto.