

SOLUCIONES GUIA 1 CONTROL 3
(preguntas 1 y 2)

P1)

- a) Para medir la relación entre dos variables cuantitativas conviene hacer un gráfico de dispersión. Este permite, en primer lugar, comparar las observaciones, ver en qué difieren y por otra parte, estudiar la relación entre las dos variables. Como el coeficiente de correlación lineal permite solamente de medir el grado de relación de tipo **lineal**, el gráfico de dispersión podrá orientar en buscar eventualmente otra relación si se detecta que no hay linealidad, pero otro tipo de relación. Si vale +1 se tiene una relación estrictamente lineal con pendiente positiva. Si vale -1 se tiene una relación estrictamente lineal con pendiente negativa. Si esta cercano a +1 hay una relación de tendencia lineal con pendiente positiva (salvo excepción con mezcla de poblaciones o observaciones atípicas). Si vale 0 no hay relación lineal, pero puede haber una relación funcional no lineal como, por ejemplo, una cuadrática.
- b) Para medir la relación entre una variable cuantitativa y una variable cualitativa, se usa un gráfico de cajas (Boxplot) para cada uno de los grupos definidos por la variable cualitativa. El gráfico permite de ver si las tendencias medias de los grupos son muy diferentes y la variabilidad al interior de cada grupo. La razón de correlación permite medir el grado de relación de tipo **funcional**. Se calcula como el cociente de la varianza Inter-grupo y de la varianza total. Si vale 1, existe una relación funcional estricta, ya que la varianza intra-grupo es nula. Si vale 0, no hay, ya que las medias de los grupos son todas iguales. Si esta cercano a 1, las medias por grupos varían más que la varianza promedio al interior de los grupos, se detectara entonces un cierto grado de relación.
- c) En el gráfico 1, se esperaría los puntos vecinos a la primera bisectriz y en el gráfico 2 que los puntos se distribuyen uniformemente, es decir que los residuos no dependen de la variable Y. Aquí observa una clara tendencia de los residuos con Y. Esto indica probablemente que la relación entre X e Y no es lineal. Es posible que sea logarítmica. Se podría probar de hacer la regresión de $\log(Y)$ con X.

P2)

a) (Manera corta y astuta)

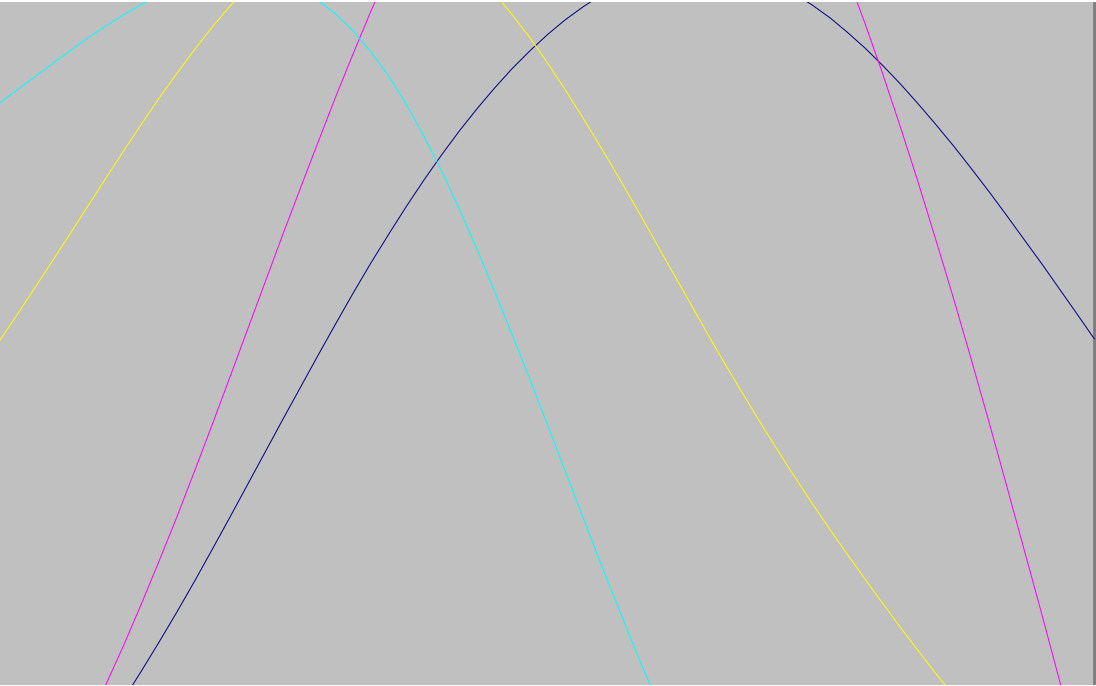
Los mínimos y máximos de cada grupo son:

grupo 1	3.62	4.79
grupo 2	3.6	4.42
grupo 3	3.25	4.41
grupo 4	2.97	4.01

Luego claramente se ve que el mínimo es 2.97 y el máximo 4.79.

Si dividimos el rango en 5 intervalos y contamos la cantidad de datos que caen en cada intervalo podemos hacer histogramas para cada grupo ¿no?

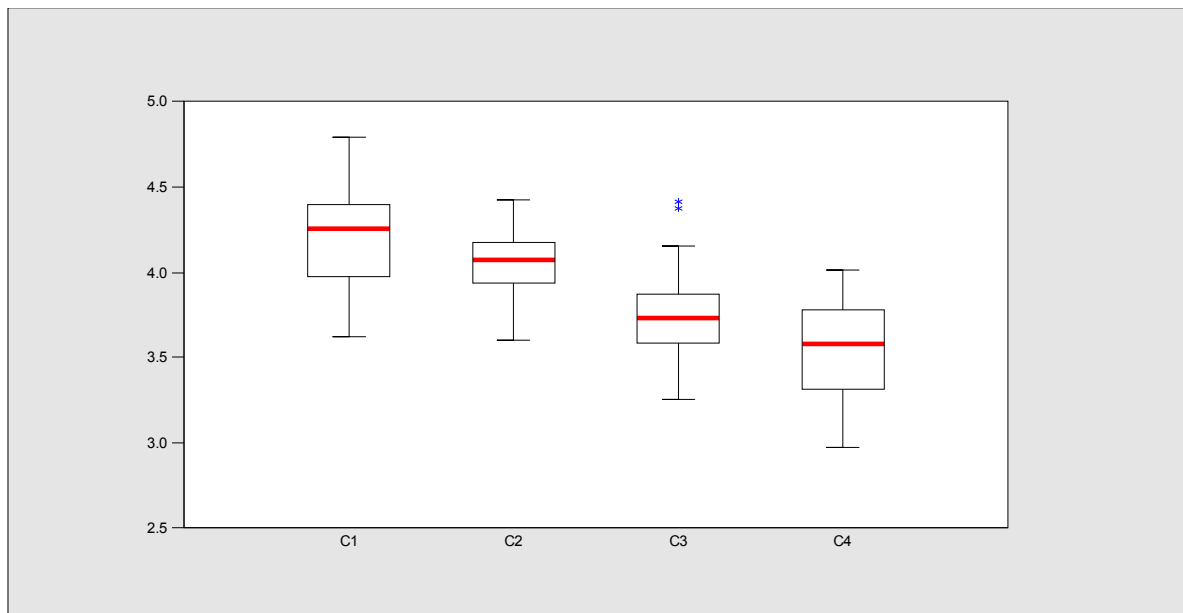
[2.97,3.334]	(3.334,3.698]	(3.698,4.062]	(4.062,4.426]	(4.426-4.79]	
0	1	12	17	10	1 prom
0	1	17	21	1	3 prom
2	11	18	9	0	5 prom
10	15	15	0	0	7 prom



El gráfico muestra que hay diferencias, lo cual debe ser confirmado por el test ANOVA.

(La manera larga)

Con gráficos de caja (boxplots):



Resumen de los datos:

n	Media	S	Grupo
40	4.224	0.2734	1 prom
40	4.062	0.1742	3 prom
40	3.759	0.252	5 prom
40	3.548	0.275	7 prom

Están las siguientes dos fórmulas para efectuar el cálculo: $S^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ y

$$S^2 = \frac{1}{n} \sum (x_i)^2 - \bar{x}^2 .$$

Si hicieron el cálculo con la primera fórmula entonces el tiempo de demora es infinito, si lo hicieron utilizando la segunda ecuación el tiempo era de aproximadamente 10 - 15 minutos con calculadora normal.

Tabla ANOVA:

ANOVA table

Fuente	gl	Suma de Cuadrados	Mean Square	F-Ratio	P-Value
Tratamientos (B)	3	10.988543	3.662848	59.903437	0.000
Error (W)	156	9.538755	0.061146		
Total (T)	159	20.527298			

b)

Al observar que los valores de S son parecidos por tanto es razonable pensar que el supuesto es válido.

c)

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1 : Alguno distinto

De la tabla ANOVA de la parte a) se tiene un valor $F = 59,9$, que es un valor alto, por lo que el P-Valor asociado tiende a 0 => Existen diferencias en los precios para los cuatro grupos.