Estadística (Q): Resolución del ejercicio 5, de la lista de ejercicios de la clase práctica 21: 30/10/2012

Ejercicio 5. Una compañía pretende decidir si basado solamente en el aspecto del envase será posible modificar el precio de un perfume. Para ello selecciona 15 clientes al azar les hace "probar" el perfume presentado en un envase de aspecto tradicional y les solicita que indiquen cual es el máximo valor que estarían dispuestos a pagar por el mismo. Selecciona otros 15 clientes al azar y repite la prueba pero usando el envase moderno. El precio máximo reportado por cada uno de los 30 clientes se muestra en la tabla siguiente.

Envase	Máximo precio que pagaría														
MODERNO	20	25	40	44	43	13	32	34	35	11	12	46	13	17	47
TRADICION	5	7	10	11	12	17	21	28	33	35	40	40	41	44	45

a) Realice gráficos descriptivos de ambos conjuntos de datos: histogramas, boxplot, qqplots.

b) Decida qué test realizar para definir si el precio máximo que están dispuestos a pagar los clientes depende del envase. Indique en base a qué gráficos hace su elección.

c) Realice el test de hipótesis correspondiente a nivel 0.05 e indique claramente la conclusión.

a) Antes de hacer los gráficos, definimos las variables aleatorias:

X_i = máximo precio que estaría dispuesto a pagar el iésimo cliente que prueba el perfume en el envase moderno

Y_i = ídem envase tradicional

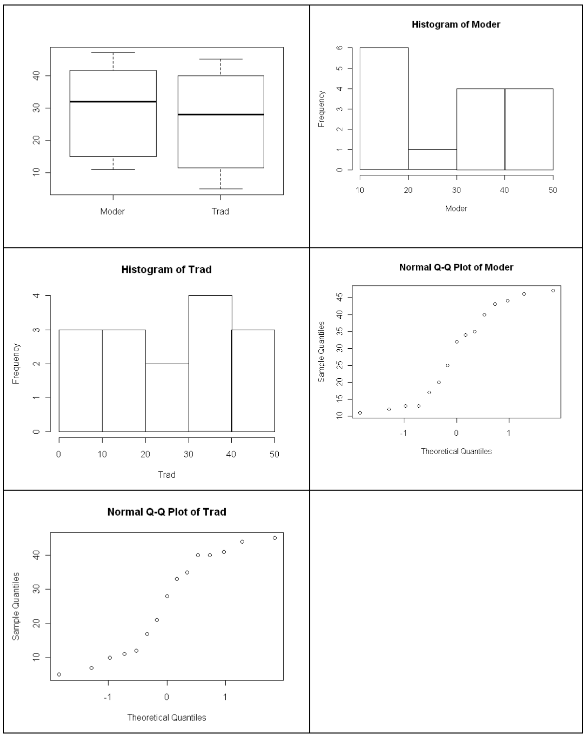
Entonces, resulta que las X_i son v.a.i.i.d. ($1 \leq i \leq 15$), las Y_i son v.a.i.i.d. ($1 \leq i \leq 15$) y las dos muestras (la de las X's y la de las Y's) son independientes entre sí (puesto que se trata de clientes distintos). Luego, para muestras independientes tenemos, esencialmente, dos posibilidades:

→ Hacer tests bajo el supuesto de que la distribución tando de las X's como la de las Y's son normales

→ Hacer un test no paramétrico, si llamamos F a la distribución de las X's y G a la distribución de las Y's, tenemos dos modelos posibles donde trabajar para realizar el test de sumas de rangos de Wilcoxon para dos muestras independientes, o test de Mann-Whitney-Wilcoxon para dos muestras independientes.

Analicemos los gráficos de los datos para decidir cuál camino tomar. Las instrucciones y los gráficos figuran a continuación.

```
> perfumes <- read.table("C:/datos/perfumes.txt", header=T, quote="\"")
> summary(perfumes)
     Moder            Trad      
 Min.   :11.0   Min.   : 5.00  
 1st Qu.:15.0   1st Qu.:11.50  
 Median :32.0   Median :28.00  
 Mean   :28.8   Mean   :25.93  
 3rd Qu.:41.5   3rd Qu.:40.00  
 Max.   :47.0   Max.   :45.00  
> attach(perfumes)
> boxplot(perfumes)
> hist(Moder)
> hist(Trad)
> qqnorm(Moder,main="Normal Q-Q Plot of Moder")
> qqnorm(Trad,main="Normal Q-Q Plot of Trad")
> shapiro.test(Moder)

Shapiro-Wilk normality test

data:  Moder
W = 0.8885, p-value = 0.06362

> shapiro.test(Trad)

Shapiro-Wilk normality test

data:  Trad
W = 0.8921, p-value = 0.07203
```



b) Los boxplots muestran distribución no normal (variabilidad muy grande). Lo mismo sucede en los q-q plots: los puntos no parecen seguir una línea recta, indicando falta de ajuste a la distribución normal. También puede verse esto en los histogramas (pareciera haber distribución uniforme en ambos grupos de observaciones) y los p-valores de Shapiro Wilks (que son 0,0636 y 0,072, respectivamente) no son mayores que 0,20 (indicando rechazar la distribución normal tanto para una muestra como para la otra), resulta que no es posible utilizar los tests bajo el supuesto de normalidad de las observaciones. Por otro lado, como tanto los boxplots como los histogramas son muy similares entre sí en forma, se puede asumir que vale que

$$F(x) = G(x - \Delta) \text{ para algún } \Delta \in \mathbb{R}. \tag{1}$$

Bajo este modelo para las distribuciones de las observaciones, el test de M-W-W para dos muestras independientes se puede expresar como un test para Δ:

$$H_0 : \Delta = 0$$
$$H_1 : \Delta \neq 0$$

Si no puediéramos sustentar el modelo (1) de que la distribución de las X's es un corrimiento de la de las Y's (es decir, ambas distribuciones tienen la misma forma, difieren a lo sumo en un desplazamiento), de todos modos podríamos realizar el test de M-W-W pero en tal caso las hipótesis que testearíamos serían

$$H_0 : F = G$$
$$H_1 : F \neq G$$

Observemos que ambas hipótesis nulas coinciden (si $\Delta = 0$ en el modelo (1), resulta que $F = G$).

c) Corramos el test M-W-W en R.

```
> wilcox.test(Moder, Trad, paires=FALSE,alternative = "two.sided",exact=FALSE)

  Wilcoxon rank sum test with continuity correction

data:  Moder and Trad
W = 130.5, p-value = 0.4674
alternative hypothesis: true location shift is not equal to 0
```

Si miramos la salida del test de M-W-W vemos que el p-valor es 0,4674 > 0,05. Luego, no tenemos razones suficientes para rechazar H_0, y resulta que el envase no influye en el precio máximo que un cliente está dispuesto a pagar por el perfume.