

## Métodos Estadísticos en Economía. Grado en Administración y Dirección de Empresas

### PRÁCTICA 6. Análisis de la varianza (ANOVA)

Alumnos:

#### Ejercicio 1

¿Pueden ayudarnos ciertos aromas a hacer negocio? Sabemos que la música de fondo puede influir en el comportamiento de los clientes. ¿Se ven también influidos por los aromas? Para tratar de dar respuesta, un estudiante de ADE tomó datos en una pizzería de Valencia durante las tardes de tres sábados de noviembre.

Una tarde colocó ambientadores con olor a lavanda, otra tarde ambientadores con olor a limón, y otra tarde no puso ningún ambientador para tener datos de referencia (grupo de control).

Los resultados de las tres tardes son comparables en cuanto a las condiciones climáticas, número de clientes que accedió al local, etc ; por lo tanto, podemos considerar que los datos son muestras aleatorias simples (m.a.s.) representativas de la población de clientes que acuden a esta pizzería las tardes de los sábados de otoño.

Los datos siguientes recogen el tiempo (en minutos) que los clientes estuvieron en la pizzería (en general, cuando más tiempo pasa un cliente en un restaurante, mayor es su consumo).

Lavanda									
92	126	114	106	89	137	93	76	98	108
124	105	129	103	107	109	94	105	102	108
95	121	109	104	116	88	109	97	101	106

Limón									
78	104	74	75	112	88	105	97	101	89
88	73	94	63	83	108	91	88	83	106
108	60	96	94	56	90	113	97		

Sin aroma									
103	68	79	106	72	121	92	84	72	92
85	69	73	87	109	115	91	84	76	96
107	98	92	107	93	118	87	101	75	86

(a) Identifique la variable respuesta y el factor considerados en este estudio:

Variable respuesta:

Factor:

Identifique también las muestras y las poblaciones de este estudio:

Muestras:

Poblaciones:

Introduzca los datos en R. Preste atención al código a introducir:

```
# Una única columna con todos los datos de los tres grupos:
tiempo = c(92 , 126 , 114 , ... , 75 , 86)
# Una columna que indica a que aroma pertenece cada tiempo:
aroma = rep(c("lavanda" , "limón" , "sin aroma") , c(30 , 28 ,
30))
# Convertimos la columna `aroma` en factor:
aroma = factor(aroma)
```

(b) Obtener un diagrama de caja múltiple para el tiempo de permanencia de los clientes en función del aroma del restaurante. ¿Alguna distribución presenta valores aislados? ¿Alguna distribución presenta una asimetría muy marcada?

NOTA: La función para realizar un diagrama de caja múltiple en R es `boxplot(tiempo~aroma, horizontal = TRUE)`

(c) Obtenga un gráfico en papel probabilístico normal para cada grupo. ¿Podemos asumir que las distribuciones son normales?

NOTA: Feu servir el codi següent:

```
qqnorm(temps[aroma == "lavanda"] , datax = TRUE)
qqnorm(temps[aroma == "llima"] , datax = TRUE)
qqnorm(temps[aroma == "sense aroma"] , datax = TRUE)
```

(d) Realice un test F (ANOVA) en R para comparar las medias de las tres poblaciones.

Hipòtesis del test:

$H_0$ :

$H_1$ :

NOTA: Puede utilizar el código siguiente para ejecutar el análisis y mostrar la tabla resumen:

```
fm = aov(lm(temps~aroma))
summary(fm)
```

Valor de el estadístico  $F$ :

Grados de libertad del factor:

Grado de libertad residuales:

Valor  $p$ :

Explique sus conclusiones en el contexto del problema:

(e) Obtenga un gráfico de los intervalos HSD de Tukey del análisis anterior.

NOTA: Use este código:

```
intervals = TukeyHSD(fm)
plot(intervals)
```

¿Qué conclusión se puede extraer de estos intervalos?

(f) Utilice los residuos (R los almacena en la variable fm \$ residuales) para validar el modelo (es decir, para comprobar el cumplimiento de las tres condiciones asumidas en realizar el ANOVA):

1. *Independencia* (entre y dentro de las muestras). Esta condición debe garantizar la forma en que se seleccionan los individuos (utilizando un muestreo probabilístico o un diseño aleatorizado).

En R, podemos utilizar el gráfico plot (fm\$residuales) para validar esta hipótesis.

2. *Normalidad*. Compruebe si es aceptable pensar que los residuos proceden de una población normal de media igual a cero.

NOTA: Utilice el siguiente código de R:

```
summary(fm$residuals)
boxplot(fm$residuals)
hist(fm$residuals)
qqnorm(fm$residuals , datax = TRUE)
```

3. *Homoscedasticidad* (Igualdad de varianzas). Compruebe si la dispersión de los residuos es similar en los tres grupos. Puede utilizar un diagrama de caja múltiple y también el contraste de hipótesis con los residuos al cuadrado.

```
boxplot(fm$residuals~aroma)
res2=aov(lm(fm$residuals^2~aroma))
summary(res2)
```

Después de estos análisis, ¿cree que el ANOVA es un modelo válido para estos datos?

## Ejercicio 2

El conjunto de datos 'female.inc' que puede descargar desde PoliformaT contiene los ingresos ( 'income') y el grupo étnico ( 'race') de una muestra de mujeres estadounidenses de más de 15 años de edad en el 2001.

Importa este conjunto de datos a R y utilice el comando attach para tener acceso a las variables de este archivo.

NOTA: Al importar el archivo de datos, tenga en cuenta que, en esta ocasión, el separador de campos es la coma (`sep=","`) y el separador decimal es el punto (`dec="."`).

(a)

¿Cuántas mujeres hay en la muestra?

¿Cuántos grupos étnicos hay?

¿Cuántas mujeres hay en cada grupo?

NOTA: Puede utilizar la función de R table para obtener una `table` de frecuencias de la variable 'race'.

(b) Realice un análisis descriptivo de los ingresos ( 'income'). ¿Crees que esta variable sigue un modelo de distribución normal? ¿Por qué?

(c) A veces, una transformación de la variable respuesta permite aplicar técnicas paramétricas. Si utilizamos el logaritmo de los ingresos (`log(income)`) como variable respuesta, ¿podemos utilizar un ANOVA con estos datos?

Obtener la tabla resumen del ANOVA de log (income) con el factor race. ¿Existen diferencias significativas entre los ingresos medios de los diferentes grupos étnicos?

Valor  $p$ :

Conclusión:

(d) Obtenga un gráfico para los intervalos HSD de Tukey. Explique sus conclusiones:

(e) Compruebe si se cumplen las condiciones para utilizar la técnica del ANOVA mediante un análisis de los residuos:

1. *Independencia*. Asumimos que los datos se obtuvieron mediante un muestreo probabilístico (no hay que hacer nada)
2. *Normalidad*. Realice los análisis necesarios para comprobarla. ¿Podemos asumir que se cumple? ¿por qué?

3. *Homoscedasticidad*. Realice los análisis necesarios para comprobarla. ¿Podemos asumir que se cumple? ¿Por qué?

(f) Si no se cumplen las condiciones, utilice una prueba no paramétrica para poder obtener conclusiones válidas (`kruskal.test(income~race)`).

Valor  $p$ :

Conclusión:

¿Se obtiene la misma conclusión que con la técnica del ANOVA?