



Técnicas de Inferencia Estadística II

Segunda prueba parcial (A)

Ejercicio 1. El fichero de datos `DatosTumores.csv` corresponde a un estudio sobre dimensiones de tumores respecto a su tipo y forma. Las variables son:

Nombre	Descripción
TIPO	Tipo del tumor (1,2)
FORMA	Forma del tumor (1,2,3)
DIMENSION	Dimensión estimada del tumor

La siguiente salida de R muestra la lectura y las cinco primeras observaciones:

```
Datos <- read.table("DatosTumores.csv",header = TRUE,sep = ";",dec = ",")
Datos
  TIPO FORMA DIMENSION
1     1     1     598.0
2     1     1     762.5
3     1     1     841.5
4     1     1    1085.0
5     1     1    1400.0
```

- a) Realice un contraste de igualdad de medias de las dimensiones del tumor para los tumores de tipo 1, `Datos$DIMENSION[Datos$TIPO == 1]`, respecto a sus tres posibles formas. Plantee las hipótesis nula y alternativa, así como los supuestos necesarios para la validez del contraste. Utilice $\alpha=0.05$. Calcule e interprete el p-valor. (2 puntos)

El contraste pedido es $H_0: \mu_1 = \mu_2 = \mu_3$ donde μ_i es la media de las dimensiones de los tumores de tipo 1 y forma i , $H_1: \text{algún } \mu_i \text{ es diferente}$.

Los supuestos para este contraste son: las muestras se consideran independientes, y las dimensiones tienen distribución normal e iguales varianzas en las tres formas.

El contraste en R sería:

```
> Dtipol <- Datos$DIMENSION[Datos$TIPO == 1]
> Ftipol <- Datos$FORMA[Datos$TIPO == 1]
> Ftipol <- as.factor(Ftipol)
>
> aov.fit <- aov(Dtipol ~ Ftipol)
> summary(aov.fit)
              Df Sum Sq Mean Sq F value Pr(>F)
Ftipol         2 1171484   585742   7.244 0.00227 **
Residuals    36 2911077    80863
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-valor = 0.0027 < $\alpha=0.05$, por lo tanto, rechazamos H_0 , es decir, hay diferencias en las medias de las dimensiones según la forma del tumor.

- b) ¿Entre qué formas hay diferencias significativas en las dimensiones medias de los tumores de tipo 1? Plantee las hipótesis nula y alternativa, así como los supuestos necesarios para la validez del contraste. Utilice $\alpha=0.05$. Calcule e interprete el p-valor. (1 punto)

Se piden los contrastes $H_0: \mu_i = \mu_j$, para cada par (i,j) y donde μ_i es la media de las dimensiones de los tumores de tipo 1 y forma i. Estos contrastes se resuelven mediante el procedimiento de Tukey.

Los supuestos son los mismos que en el apartado anterior.

El contraste en R sería:

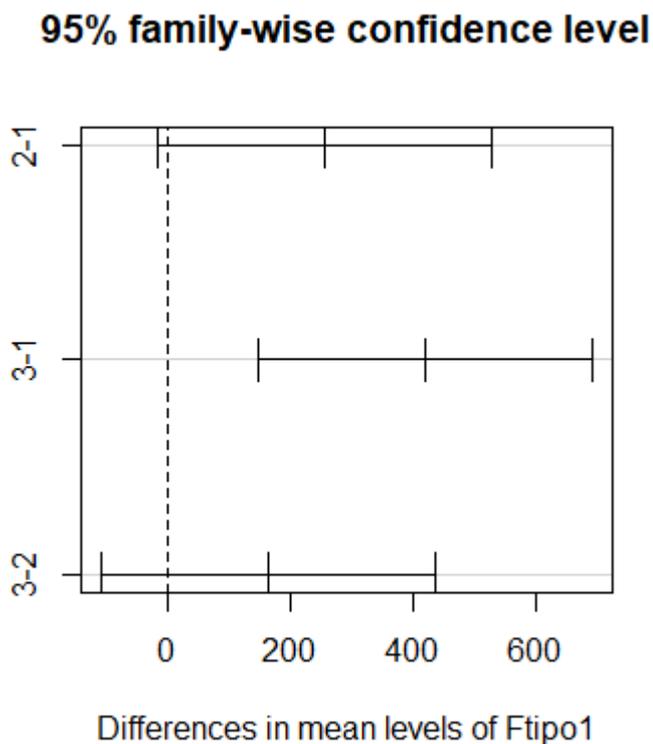
```
> TukeyHSD(aov.fit)
  Tukey multiple comparisons of means
  95% family-wise confidence level

Fit: aov(formula = Dtipol ~ Ftipol)

$Ftipol
      diff      lwr      upr    p adj
2-1 257.1154 -15.51418 529.7449 0.0678599
3-1 421.1154  148.48582 693.7449 0.0016347
3-2 164.0000 -108.62956 436.6296 0.3168789
```

Los p-valor ajustados son aproximadamente 0.068, 0.002 y 0.317 para los contrastes 2 versus 1, 3 versus 1 y 3 versus 2, respectivamente. Por tanto, usando una $\alpha=0.05$, solo rechazaríamos la igualdad de medias entre las formas 3 y 1.

- c) Represente gráficamente los resultados del contraste anterior. (1 punto)



- d) ¿Es necesario suponer normalidad en el contraste del apartado a)? Si es así, realice los contrastes de normalidad que considere necesarios. Plantee las hipótesis nula y alternativa, así como los supuestos necesarios para la validez de los contrastes. Utilice $\alpha=0.05$. Calcule e interprete los p-valores. (2 puntos)

En el apartado a) necesitamos que las dimensiones para cada una de las formas de los tumores de tipo 1 se distribuyan normal, es decir, tendríamos que realizar los siguientes contrastes de normalidad:

$H_0: X_i \sim N$
 $H_1: X_i \not\sim N$ donde X_i son las dimensiones en la forma i .

```
> library(nortest)
> lillie.test(Dtipol[Ftipol==1])

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  Dtipol[Ftipol == 1]
D = 0.16287, p-value = 0.4529

> lillie.test(Dtipol[Ftipol==2])

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  Dtipol[Ftipol == 2]
D = 0.18942, p-value = 0.2281

> lillie.test(Dtipol[Ftipol==3])

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  Dtipol[Ftipol == 3]
D = 0.1859, p-value = 0.2525
```

En los tres contrastes no se puede rechazar la hipótesis de normalidad.

Ejercicio 2. En el archivo `NLYS.csv` tenemos una submuestra de la Encuesta Nacional Longitudinal de Jóvenes (*National Longitudinal Youth Survey*). Las variables son:

Nombre	Descripción
GENERO	Género (Hombres 1, Mujeres 2)
ALTURaft	Altura pies
ALTURain	Altura pulgadas
PESO	Peso en libras
AUTO	Auto descripción de su (percepción de su) peso 1 = muy bajo peso 2 = ligeramente bajo peso 3 = sobre el peso correcto 4 = ligeramente sobrepeso 5 = muy sobrepeso

```

> NLYS <- read.table("NLYS.csv",header = TRUE,sep = ";",dec = ",")
> NLYS[1:5,]
  GENERO EDAD ALTURaft ALTURain PESO AUTO
1      2   18      5      7  150    4
2      1   17      5      7  140    3
3      2   16      5      7  100    2
4      2   18      5      2  185    4
5      1   17      5      6  145    4

```

- a) En hombres, $\text{GENERO} = 1$, ¿la proporción de las cinco categorías de auto descripción es la misma? Plantee las hipótesis nula y alternativa, así como los supuestos necesarios para la validez del contraste. Utilice $\alpha=0.05$. Calcule e interprete el p-valor. (2 puntos).

Nos piden contrastar si las proporciones son iguales, por tanto, podemos realizar un contraste de bondad de ajuste a una distribución uniforme discreta. Esto es $H_0: p_1 = p_2 = p_3 = p_4 = p_5$ donde p_i es la proporción de hombres con valor i en la variable AUTO. $H_1: \text{algún } p_i \text{ es diferente}$

Realizaremos un contraste de bondad de ajuste χ^2 que supone que las observaciones son independientes.

```

> AUTO1 <- NLYS$AUTO[NLYS$GENERO==1]
> AUTO1 <- as.factor(AUTO1)
> summary(AUTO1)
  1     2     3     4     5 NA's
90  636 2523  788  102  460
> length(AUTO1)
[1] 4599
> length(AUTO1)-460
[1] 4139

```

Debemos tener en cuenta la presencia de NA, por tanto, el número efectivo de datos es 4139. La instrucción `summary()` nos aporta el número observado de cada categoría.

El contraste en R sería:

```
> chisq.test(c(90,636,2523,788,102),p=rep(1/5,5))
```

Chi-squared test for given probabilities

```
data:  c(90, 636, 2523, 788, 102)
X-squared = 4811.8, df = 4, p-value < 2.2e-16
```

p-valor < 2.2e-16, por tanto, rechazamos la hipótesis nula. Es decir, alguna proporción es diferente.

- b) El **índice de masa corporal (IMC)** se calcula dividiendo los kilogramos de peso por el cuadrado de la estatura en metros ($\text{IMC} = \text{peso [kg]} / \text{estatura [m}^2\text{]}$). El índice de masa corporal “normal” en hombres de esas edades está entre 19 y 25. Contraste si la proporción de hombres con un índice de masa corporal “normal” es superior al 50%. Plantee las hipótesis nula y alternativa, así como los supuestos necesarios para la validez del contraste. Utilice $\alpha=0.05$. Calcule e interprete el p-valor. (2 puntos)

En este ejercicio lo primero que debemos hacer es obtener las alturas y pesos en las unidades de medidas metros y kg, respectivamente.

```

> A1 <- NLYS$ALTURaft[NLYS$GENERO==1]
> A2 <- NLYS$ALTURain[NLYS$GENERO==1]
> Altura <- A1/3.2808 + (A2/0.39370)/100
>
> P1 <- NLYS$PESO[NLYS$GENERO==1]
> Peso <- P1/2.2046
>
> IMC <- Peso/Altura^2

```

Definimos el IMCnormal como:

```

> IMCnormal <- (IMC>=19) & (IMC<=25)
> summary(IMCnormal)
  Mode   FALSE   TRUE   NA's
logical  1644   2469   486

```

Realizamos el contraste $H_0: p_{\text{IMC}} = 0.5$ donde p_{IMC} es la proporción de hombres con IMC en el intervalo [19,25]. Este contraste supone que tenemos una m.a.s.

El contraste en R sería:

```

> binom.test(2469,length(IMC)-486,p=0.5,alternative="greater")

Exact binomial test

data: 2469 and length(IMC) - 486
number of successes = 2469, number of trials = 4113, p-value < 2.2e-16
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.5875571 1.0000000
sample estimates:
probability of success
 0.6002918

```

También podríamos usar el contraste asintótico:

```

> prop.test(2469,length(IMC)-486,p=0.5,alternative="greater")

1-sample proportions test with continuity correction

data: 2469 out of length(IMC) - 486, null probability 0.5
X-squared = 165.08, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.5875445 1.0000000
sample estimates:
      p
0.6002918

```

En ambos casos rechazaríamos la hipótesis nula, es decir, la proporción es superior al 50%.