

Estadística (Q) - Clase 19

ANOVA (Práctica 7)

1. Se ha realizado un experimento para determinar si la resistencia de cierta tela se ve afectada por el porcentaje de poliéster con el que es fabricada. Para ello, se midió la resistencia de 18 de estas telas producidas con distintos porcentajes de poliéster y se obtuvieron los siguientes resultados:

Porcentaje de poliéster	Resistencia				
Bajo	21.274	20.822	20.452	21.291	19.897
Medio	22.309	23.999	23.304	21.532	
Alto	24.010	23.132	23.848	22.300	23.153
Muy alto	25.445	24.660	24.229	22.895	

Se quiere aplicar el modelo de análisis de la varianza (ANOVA) para decidir si existen diferencias en la resistencia media de los distintos grupos.

- Definir claramente las variables aleatorias, los parámetros involucrados, establecer los supuestos necesarios para aplicar dicho modelo y escribir la hipótesis a testear en este caso.
- Verificar si se cumplen los supuestos del ANOVA.
- Construya la tabla de análisis de la varianza. ¿Existen diferencias significativas entre los diferentes laboratorios a nivel 0.05?
- En el caso de rechazar H_0 , detecte para qué porcentajes de poliéster las resistencias difieren significativamente mediante intervalos de confianza con nivel simultáneo 95 %.

Resultados de R:

```
> resist <- scan()
> porc <- as.factor(c(rep(1,5),rep(2,4),rep(3,5),rep(4,4)))

> levene.test(resist,porc)

modified robust Brown-Forsythe Levene-type test based on the absolute
deviations from the median

data: resist
Test Statistic = 0.8505, p-value = 0.4892

> anovaresist <- aov(resist~porc)
> shapiro.test(anovaresist$res)

Shapiro-Wilk normality test

data: anovaresist$res
W = 0.9582, p-value = 0.5679
```

```

> summary(anovaresist)
          Df Sum Sq Mean Sq F value    Pr(>F)
porc       3 31.162 10.3873  14.255 0.0001544 ***
Residuals 14 10.202  0.7287
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Para el ítem *d*):

Cantidad de diferencias:

```

> m <- choose(4,2)
[1] 6

```

Construimos los IC de Bonferroni:

```

> alfa <- 0.05
> k <- 4
> n1 <- 5
> n2 <- 4
> n3 <- 5
> n4 <- 4
> nivel <- alfa/m
[1] 0.008333333
> sp2 <- 0.7287
> medias <- tapply(resist,porc,mean)
> gl <- 18 - 4
> tt <- qt(1-nivel/2,gl)

> i1i = medias[1] - medias[2] - tt*sqrt(sp2)*sqrt(1/n1 + 1/n2)
> i1s = medias[1] - medias[2] + tt*sqrt(sp2)*sqrt(1/n1 + 1/n2)
> i2i = medias[1] - medias[3] - tt*sqrt(sp2)*sqrt(1/n1 + 1/n3)
> i2s = medias[1] - medias[3] + tt*sqrt(sp2)*sqrt(1/n1 + 1/n3)
> i3i = medias[1] - medias[4] - tt*sqrt(sp2)*sqrt(1/n1 + 1/n4)
> i3s = medias[1] - medias[4] + tt*sqrt(sp2)*sqrt(1/n1 + 1/n4)
> i4i = medias[2] - medias[3] - tt*sqrt(sp2)*sqrt(1/n2 + 1/n3)
> i4s = medias[2] - medias[3] + tt*sqrt(sp2)*sqrt(1/n2 + 1/n3)
> i5i = medias[2] - medias[4] - tt*sqrt(sp2)*sqrt(1/n2 + 1/n4)
> i5s = medias[2] - medias[4] + tt*sqrt(sp2)*sqrt(1/n2 + 1/n4)
> i6i = medias[3] - medias[4] - tt*sqrt(sp2)*sqrt(1/n3 + 1/n4)
> i6s = medias[3] - medias[4] + tt*sqrt(sp2)*sqrt(1/n3 + 1/n4)

```

```

> rbind(c(i1i,i1s), c(i2i,i2s), c(i3i,i3s),c(i4i,i4s), c(i5i,i5s), c(i6i,i6s))
      1      1
[1,] -3.796102 -0.2814985
[2,] -4.198200 -0.8846002
[3,] -5.317352 -1.8027485
[4,] -2.259902  1.2547015
[5,] -3.373608  0.3311085
[6,] -2.775952  0.7386515

```

Intervalos de confianza y Tests simultáneos de Tukey:

```

USAR PAQUETE stats
> TukeyHSD(anovaresist)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = resist ~ porc)
$porc
      diff      lwr      upr    p adj
2-1 2.03880  0.3744021 3.703198 0.0146909
3-1 2.54140  0.9721906 4.110609 0.0016907
4-1 3.56005  1.8956521 5.224448 0.0001178
3-2 0.50260 -1.1617979 2.166998 0.8162604
4-2 1.52125 -0.2331794 3.275679 0.0998211
4-3 1.01865 -0.6457479 2.683048 0.3231746

```

Comparemos las longitudes de los IC de los 2 métodos:

```

> tt
[1] 3.068779
> qq <- qtuchy(1-alfa,4,18-4)/sqrt(2)
[1] 2.906567

```

2. Un gobierno provincial realiza una evaluación de opción múltiple a estudiantes de quinto año en 5 localidades, calificándolos entre 0 y 100. Desea saber si hay diferencias significativas entre los desempeños de las distintas localidades, con el objeto de reforzar el presupuesto educativo en aquellas con peor desempeño. A pesar de su buena voluntad a futuro, no tiene presupuesto asignado para pagar las correcciones de estos exámenes, así que el ministro elige apenas 20 exámenes de cada localidad al azar entre las respectivas pilas y se pone a corregir.

a) (4p) Describir las variables aleatorias involucradas, los parámetros y los supuestos para realizar un test F de Análisis de la varianza. ¿Cuáles son las hipótesis del test?

.....

A continuación se detallan las salidas obtenidas en R:

```
> anovanotas<-aov(notas~localidad)
> shapiro.test(anovanotas$res)

      Shapiro-Wilk normality test

data:  anovanotas$res
W = 0.9837, p-value = 0.2559

> levene.test(notas,localidad)

      modified robust Brown-Forsythe Levene-type test based on the absolute
      deviations from the median

data:  notas
Test Statistic = 0.4128, p-value = 0.799

> summary(anovanotas)
      Df Sum Sq Mean Sq F value    Pr(>F)
localidad    4   4796   1199.1    5.298 0.000678 ***
Residuals   95  21501    226.3
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

Fit: aov(formula = notas ~ localidad)

```
$localidad
      diff      lwr      upr      p adj
2-1 -6.75 -19.9796199  6.47962 0.6171542
3-1 10.00  -3.2296199 23.22962 0.2278144
4-1 12.50  -0.7296199 25.72962 0.0734032
5-1  4.10  -9.1296199 17.32962 0.9100341
3-2 16.75   3.5203801 29.97962 0.0058434
4-2 19.25   6.0203801 32.47962 0.0009858
5-2 10.85  -2.3796199 24.07962 0.1602573
4-3  2.50 -10.7296199 15.72962 0.9845509
5-3 -5.90 -19.1296199  7.32962 0.7279397
5-4 -8.40 -21.6296199  4.82962 0.3995901
```

A continuación se calculan las medias muestrales de notas por localidad

```
> tapply(notas,localidad,mean)
 1      2      3      4      5
63.95 57.20 73.95 76.45 68.05
```

b) (4p) ¿Se cumplen los supuestos del modelo para aplicar el test Anova? Fundamente brevemente:

.....

.....

.....

c) (4p) ¿Qué conclusión puede sacar el ministro?

.....

.....

.....

d) (4p) ¿Existen pares de localidades para las cuales se puede decir que la media de sus notas es significativamente diferente con nivel simultáneo 5%? Fundamente brevemente

.....

.....

.....
e) (4p) Si se decide subir el presupuesto en el mayor grupo de localidades de peores resultados en el que no se encuentren diferencias significativas a nivel simultáneo 5%. ¿En qué localidades deberían subir el presupuesto? Fundamente brevemente

.....
.....
.....