

Estadística (Q) Clase del 5/3/2014 – Ejemplo de ANOVA~

Un investigador registró la transpiración de plantas de tabaco de 5 variedades distintas. Para cada una de las variedades registró la transpiración de 12 plantas elegidas al azar. Se decide aplicar un modelo de análisis de la varianza (ANOVA) para analizar si existen diferencias significativas entre las variedades. Los datos observados son

Tabaco1	Tabaco2	Tabaco3	Tabaco4	Tabaco5
9.080729	8.798094	8.619636	9.003297	9.948916
8.898094	8.980729	9.403858	8.919114	9.509691
8.70729	8.577438	8.819636	8.785475	9.462271
8.577438	9.027343	9.103858	9.339848	9.56095
9.127343	8.607059	9.02778	9.226791	9.116161
8.607059	9.234845	9.443001	8.559244	9.727411
8.934845	8.459148	8.919614	8.425281	9.757861
8.759148	8.668539	9.263839	9.208764	9.087074
8.768539	8.828109	9.092951	8.680994	9.392484
8.428109	8.993441	8.869939	8.779978	9.53709
9.23441	8.798094	9.17763	8.978667	9.284733
8.98094	8.980729	8.817533	9.003297	9.432191

- Plantee un modelo para este problema definiendo claramente las variables aleatorias y los parámetros involucrados en este problema. Indique cuáles son los supuestos necesarios para aplicar el anova. Escriba las hipótesis que se desea testear.
- Analizar la validez de los supuestos.
- Testee las hipótesis planteadas en a) con nivel de significación 0.05. ¿Qué test/salida emplea para tomar esta decisión? Plantee el p-valor. Escriba la conclusión del test.
- ¿Qué pares de variedades difieren significativamente entre sí con un nivel de significación simultáneo del 5%? Explique en qué método/s y en qué salida/s basa su conclusión.
- Usando algún método para calcular todos los intervalos de confianza de nivel simultáneo 0.95 para la diferencia de medias entre las variedades, construya un intervalo para la diferencia entre la variedad 1 y la 5.

```
> tabac<-scan(file.choose()) #tabac.txt contiene al vector de datos
Read 60 items
> varie<-c(rep("var1",12), rep("var2",12), rep("var3",12),
rep("var4",12), rep("var5",12))
> varie.f<-factor(varie)

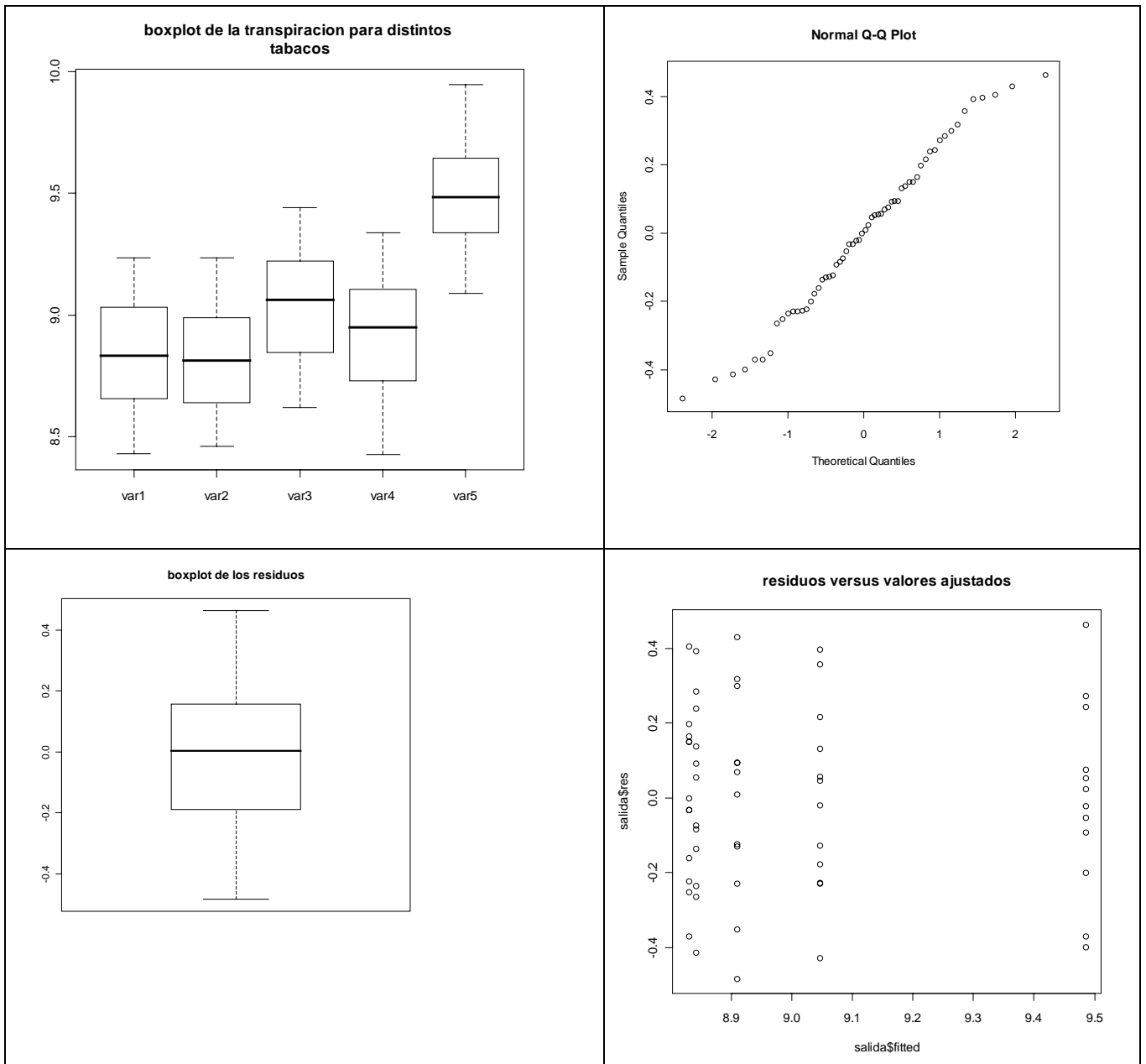
> salida<-aov(tabac~varie.f)
> summary(salida)
              Df Sum Sq Mean Sq F value    Pr(>F)
varie.f         4  3.5630  0.89076   14.257 4.782e-08 ***
Residuals      55  3.4363  0.06248
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

> names(salida)
[1] "coefficients" "residuals" "effects" "rank" "fitted.values"
[6] "assign" "qr" "df.residual" "contrasts" "xlevels"
[11] "call" "terms" "model"
>
> qf(0.95,df1=4,df2=55)
[1] 2.539689
> boxplot(tabac~varie.f,main="boxplot de la transpiracion para distintos
+ tabacos")
> boxplot(salida$res,main="boxplot de los residuos")

> qqnorm(salida$res)
> plot(salida$fitted,salida$res,main="residuos versus valores ajustados")
> plot(salida) #otra instrucción que proporciona gráficos

```



```

> shapiro.test(salida$res)

      Shapiro-Wilk normality test

data:  salida$res
W = 0.9811, p-value = 0.4761

> bartlett.test(salida$res)
Error en length(g) : g' esta perdido
> bartlett.test(tabac,varie.f)

      Bartlett test of homogeneity of variances

data:  tabac and varie.f
Bartlett's K-squared = 0.4754, df = 4, p-value = 0.9758

> install.packages("lawstat")
> levene.test(tabac,varie.f)

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

data:  tabac
Test Statistic = 0.1292, p-value = 0.9711

> medias<-tapply(tabac,varie.f,mean)
> medias
      var1      var2      var3      var4      var5
8.841995 8.829464 9.046606 8.909229 9.484736
> varianzas<-tapply(tabac,varie.f,var)
> varianzas
      var1      var2      var3      var4      var5
0.05912479 0.05037476 0.06259336 0.07631125 0.06398942
> sum(varianzas)*11/55      #reconstruimos el S_p al cuadrado, a partir de
las varianzas muestrales.
[1] 0.06247872

> tabaco.tuk<-TukeyHSD(salida,"varie.f",ordered=FALSE,conf.level=0.95)
> tabaco.tuk
      Tukey multiple comparisons of means
      95% family-wise confidence level

Fit: aov(formula = tabac ~ varie.f)

$varie.f
      diff      lwr      upr      p adj
var2-var1 -0.01253133 -0.30033079 0.2752681 0.9999470
var3-var1  0.20461092 -0.08318854 0.4924104 0.2772498
var4-var1  0.06723383 -0.22056562 0.3550333 0.9642318
var5-var1  0.64274075  0.35494129 0.9305402 0.0000005
var3-var2  0.21714225 -0.07065721 0.5049417 0.2231868
var4-var2  0.07976517 -0.20803429 0.3675646 0.9348568
var5-var2  0.65527208  0.36747263 0.9430715 0.0000003
var4-var3 -0.13737708 -0.42517654 0.1504224 0.6638932
var5-var3  0.43812983  0.15033038 0.7259293 0.0006650

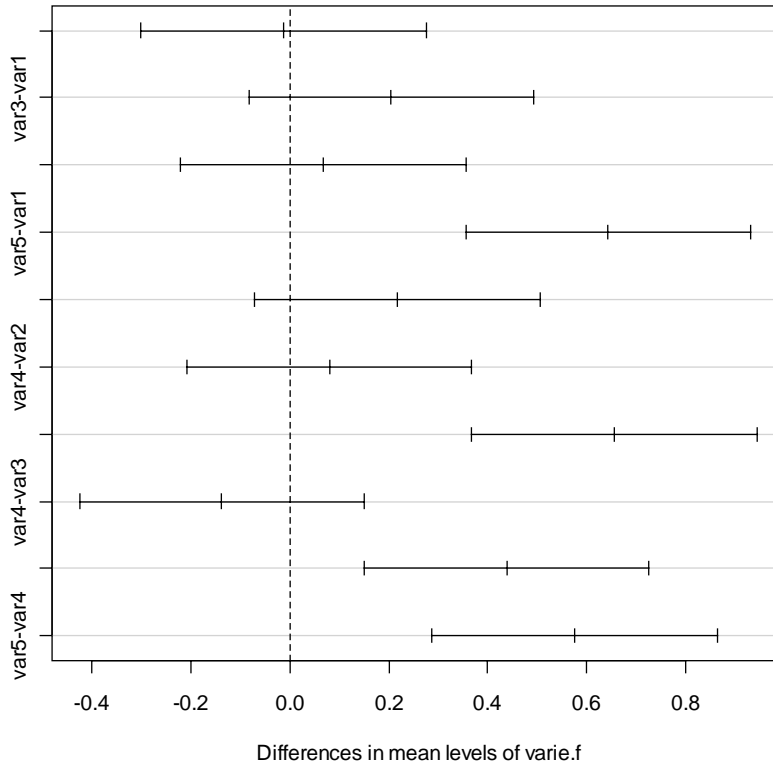
```

```

var5-var4  0.57550692  0.28770746  0.8633064  0.0000059
> qtukey(0.95,5,55)
[1] 3.988545
> plot(tabaco.tuk)

```

**95% family-wise confidence level**



```

> pairwise.t.test(tabac,varie.f,p.adj="bonf")

```

Pairwise comparisons using t tests with pooled SD

data: tabac and varie.f

	var1	var2	var3	var4
var2	1.00000	-	-	-
var3	0.49882	0.37841	-	-
var4	1.00000	1.00000	1.00000	-
var5	5.3e-07	3.3e-07	0.00072	6.1e-06

P value adjustment method: bonferroni

```

> 1-0.05/20
[1] 0.9975
> qt(0.9975,df=55)
[1] 2.924701

```