enerdì 26 febbraio 2010

Quando le assunzioni dell'ANOVA sono violate: tests e post-hoc tests

Il capitolo dei confronti multipli è uno dei più vasti nella letteratura statistica. Cominciamo con un riepilogo. Le due principali assunzioni della ANOVA sono:
1. tutti i gruppi seguono una distribuzione normale
2. le varianze sono omogenee (omoschedasticità)

Queste due assunzioni vengono verificate rispettivamente con i test di normalità, e con i test di omogeneità della varianza.
Se queste due assunzioni sono confermate, si può procedere con una **ANOVA parametrica omoschedastica**, già discussa in questo post. Se essa risulta significativa, si procede con i test post-hoc per i confronti multipli, tra cui il test HSD di Tukey, il test LSD di Fisher, il test di Bonferroni e il test di Scheffé, e il test di Student-Newman-Keuls (SNK test).

Supponiamo che sia violata la assunzione 1., ma che sia valida la assunzione 2. In questo caso si procede con una **ANOVA omoschedastica non parametrica**, più nota come test di Kruskal-Wallis, di cui si è parlato in questo post. Se questo test risulta significativo (ossia se le medie risultano non uguali), occorre procedere con test post-hoc non parametrici, tra cui:
- **Nemenyi-Damico-Wolfe-Dunn test**
- **Behrens-Fisher test**
- **Steel test (Nonparametric Multiple Comparisons)**

Ricordo che per poter procedere con il test di Kruskal-Wallis, occorre fare l'assunzione di omogeneità delle varianze dei k gruppi in analisi.

Supponiamo ora che i dati siano normali, ma che le varianze non siano omogenee. In questo caso si procede con una **ANOVA eteroschedastica**: gli approcci sono:
- **Welch one-way ANOVA** (disponibile in R)
- **Brown-Forsythe procedure** (non disponibile in R, ma presente in SPSS)
- **James first-and-second-order procedure** (non disponibile in R)

Tra breve parlerò del primo metodo.

Se il test di Welch (siamo quindi sotto l'assunzione di non omogeneità delle varianze) risulta significativo, si deve procedere con test post-hoc per i confronti multipli non omoschedastici. Tra i test noti vi sono:
- **test di Tamhane**
- **test di Games-Howell**
- **test di Duncan**

i quali però non sono disponibili in R (se non ricordo male sono però implementati in SPSS).

Infine può accadere che i dati non siano normali, e che le varianze non siano omogenee. In tal caso si procede con il **test di Friedman**, e con i relativi post-hoc test: **Schaich and Hamerle test**, e **Conover test**.

Prima di procedere con gli esempi, ecco una tabella riassuntiva sulla situazione. Mi preme però fare 3 precisazioni:

1) esistono altri test oltre a quelli qui citati e discussi;
2) ciascun test offre a considerare proprie caratteristiche che qui non vengono considerate. Ad esempio il test di Friedman è maggiormente raccomandato per una "ANOVA con misure ripetute", mentre qui si parla di violazioni delle assunzioni della ANOVA classica. Non è un errore, ma è sempre bene fare riferimento alla letteratura specifica di ogni test;
3) ogni tanto può capitare qualche errore. Per cui consiglio sempre di controllare quanto è scritto qui (prima di incorrere in "rimproveri" da parte dei referer statistici!). Se trovate qualche errore, sarei grato se me lo segnalate nei commenti, così che possa correggere il tutto.

Ho compilato questa tabella per riassumere i test che possono essere utilizzati nei vari casi:

Normalità	Omoschedasticità	Test raccomandato	Post-hoc tests
Verificata	Verificata	ANOVA parametrica omoschedastica	• Test HSD di Tukey • Test LSD di Fisher • Test di Bonferroni • Test di Scheffè • Test di Student-Newman-Keuls
Non verificata	Verificata	Test di Kruskal-Wallis	• Nemenyi-Damico-Wolfe-Dunn test • Behrens-Fisher test • Steel test (Nonparametric Multiple Comparisons)
Verificata	Non verificata	• Welch one-way ANOVA • Brown-Forsythe procedure • James first-and-second-order procedure	• Test di Tamhane • Test di Games-Howell • Test di Waller-Duncan
Non verificata	Non verificata	Friedman test	• Schaich and Hamerle test • Conover test

Tenendo ben presente questi concetti, vediamo qui di seguito 3 esempi:
- come fare i confronti multipli dopo un test di Kruskal-Wallis significativo
- come procedere con una ANOVA eteroschedastica
- come effettuare una ANOVA nonparametrica eteroschedastica, e il relativo post-hoc test

---

<u>ANOVA non parametrica in R e Kruskal-Wallis post-hoc test</u>

Consideriamo i seguenti dati:

```
a <- c(4, 3, 2, 3, 14, 4, 3, 6, 5, 54, 2, 4, 4, 2, 3, 4)
b <- c(7, 56, 4, 6, 7, 2, 9, 35, 5, 9, 3, 8, 6, 4, 7, 6)
c <- c(12, 6, 14, 12, 10, 9, 12, 17, 7, 56, 12, 11, 6, 13, 10, 14)
d <- c(17, 15, 7, 20, 13, 11, 16, 25, 11, 24, 18, 21, 16, 19, 59, 23)
```

Verifichiamo la normalità dei 4 gruppi con il test di Shapiro-Wilk:

```
> shapiro.test(a)
        Shapiro-Wilk normality test
data:  a
```

```
W = 0.4172, p-value = 4.706e-07
> shapiro.test(b)
        Shapiro-Wilk normality test
data:  b
W = 0.5481, p-value = 5.458e-06
> shapiro.test(c)
        Shapiro-Wilk normality test
data:  c
W = 0.5142, p-value = 2.786e-06
> shapiro.test(d)
        Shapiro-Wilk normality test
data:  d
W = 0.6986, p-value = 0.000163
```

Tutt'e 4 i gruppi risultano non normali. Ora verifichiamo la omoschedasticità. Essendo gruppi che non seguono una distribuzione normale, dobbiamo usare il test di Levene non parametrico:

```
dati <- c(a, b, c, d)
gruppi <- factor(rep(letters[1:4], times=c(length(a),length(b), length(c),
length(d))))
> levene.test(dati, gruppi, location="mean", kruskal.test=T)
        rank-based (Kruskal-Wallis) classical Levene's test based on the absolute
deviations from
        the mean ( none not applied because the location is not set to median )
data:  dati
Test Statistic = 5.484, p-value = 0.1396
```

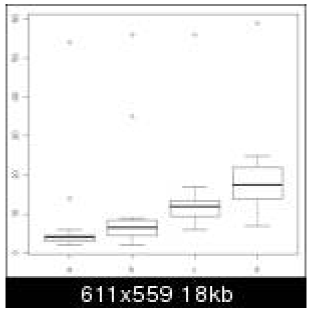
Essendo p-value > 0.05, accettiamo l'ipotesi di omogeneità delle varianze.

Alla luce di questi due dati (distribuzione non-normale, varianze omogenee), possiamo procedere con il test di Kruskal-Wallis:

```
kruskal.test(dati, gruppi)
        Kruskal-Wallis rank sum test
data:  dati and gruppi
Kruskal-Wallis chi-squared = 31.4139, df = 3, p-value = 6.955e-07
```

Essendo p-value < 0.05, rifiutiamo l'ipotesi nulla: le medie sono significativamente differenti.
Dobbiamo quindi indagare tra quali gruppi vi è una forte differenza.
Prima di procedere, proviamo a mettere in grafico i dati, in questo modo:

```
plot(gruppi, dati)
```



Se osserviamo il grafico, risulta abbastanza evidente che:
1. le medie sono differenti (come afferma il test di Kruskal-Wallis che è significativo)
2. le varianze sono omogenee (questo non è molto evidente, essendo presente degli outliers).

I test post-hoc non parametrici a nostra disposizione sono:
- **Nemenyi-Damico-Wolfe-Dunn test**
- **Behrens-Fisher test (Nonparametric Multiple Comparisons)**
- **Steel test**

Il codice per effettuare il **Nemenyi-Damico-Wolfe-Dunn test** è disponibile nel package `coin`, nell'help relativo alla funzione `oneway_test`. Occorre installare anche la library `multcomp`.
Innanzitutto creiamo un dataframe con i dati, e i gruppi:

```
dft <- data.frame(dati, gruppi)
```

Ora utilizziamo il seguente codice, e analizziamo l'output:

```
library(coin)
if (require("multcomp")) {
    NDWD <- oneway_test(dati ~ gruppi, data = dft,
        ytrafo = function(data) trafo(data, numeric_trafo = rank),
        xtrafo = function(data) trafo(data, factor_trafo = function(x)
            model.matrix(~x - 1) %*% t(contrMat(table(x), "Tukey"))),
        teststat = "max", distribution = approximate(B = 90000))

    print(pvalue(NDWD))

    print(pvalue(NDWD, method = "single-step"))
  }
#output
[1] 0
99 percent confidence interval:
 0.000000e+00 5.886846e-05

b - a 0.3819666667
c - a 0.0015555556
d - a 0.0000000000
c - b 0.2104555556
d - b 0.0009666667
d - c 0.3237666667
```

Le coppie con medie significativamente differenti sono quelle con p-value < 0.05. Quindi risultano significativamente differenti le coppie a-c, a-d, b-d (come possiamo controllare dal plot).

Il test non parametrico per i confronti multipli di Behrens-Fisher e quello di Steel sono disponibili nella library `npmc`. Anche in questo caso dobbiamo creare un dataframe, in cui necessariamente le due colonne relative ai dati e ai gruppi devono avere i nomi rispettivamente `var` e `class`:

```
var <- dati
class <- gruppi
dft <- data.frame(var, class)
> summary(npmc(dft))

 /////////////////////////////////////////////////////////////
 / npmc executed                                             /
 /                                                           /
 / NOTE:                                                     /
 / -Used Satterthwaite t-approximation (df=15.4762442214526) /
 / -Calculated simultaneous (1-0.05) confidence intervals    /
 / -The one-sided tests 'a-b' reject if group 'a' tends to   /
 /  smaller values than group 'b'                            /
 /////////////////////////////////////////////////////////////


$`Data-structure`
  group.index class.level nobs
a           1           a   16
b           2           b   16
c           3           c   16
d           4           d   16
$`Results of the multiple Behrens-Fisher-Test`
  cmp    effect  lower.cl upper.cl   p.value.1s   p.value.2s
1 1-2 0.7558594 0.4902196 1.021499 3.184958e-02 0.0613214501
2 1-3 0.8867188 0.6698053 1.103632 2.877271e-04 0.0005863507
3 1-4 0.9257812 0.7521589 1.099404 7.622567e-06 0.0000321084
4 2-3 0.8007812 0.5482831 1.053279 8.410051e-03 0.0170268058
5 2-4 0.8652344 0.6317030 1.098766 9.485910e-04 0.0020723217
6 3-4 0.7988281 0.5544711 1.043185 7.261303e-03 0.0142616548
$`Results of the multiple Steel-Test`
  cmp    effect  lower.cl upper.cl   p.value.1s   p.value.2s
1 1-2 0.7558594 0.4917138 1.020005 0.0329254942 0.0616483846
2 1-3 0.8867188 0.6213672 1.152070 0.0005619949 0.0010558237
3 1-4 0.9257812 0.6600126 1.191550 0.0002169494 0.0001998778 *
4 2-3 0.8007812 0.5353070 1.066256 0.0096495950 0.0186865845
5 2-4 0.8652344 0.5990248 1.131444 0.0012742103 0.0025523271
6 3-4 0.7988281 0.5326920 1.064964 0.0102897944 0.0206920249
```

Nell'output vengono forniti i risultati dei due test separatamente.

ANOVA parametrica eteroschedastica in R

Utilizziamo il dataset presente in questa pagina. Ho caricato i dati nel dataframe di nome `data`.
Per prima cosa effettuiamo un test di Levene non parametrico, poichè 4 dei 5 gruppi risultano non-normali secondo il test di Shapiro-Wilk, per la verifica dell'omoschedasticità:

```
data[1:5,]
  group smell
1     1 1.381
2     1 1.322
3     1 1.162
4     1 1.275
5     1 1.381
...
group5 <- data[139:180, 2]
group4 <- data[96:138, 2]
group3 <- data[75:95, 2]
group2 <- data[39:74, 2]
group1 <- data[1:38, 2]
> shapiro.test(group1)
        Shapiro-Wilk normality test
data:  group1
W = 0.9201, p-value = 0.009898
> shapiro.test(group2)
        Shapiro-Wilk normality test
data:  group2
W = 0.9167, p-value = 0.01008
> shapiro.test(group3)
        Shapiro-Wilk normality test
data:  group3
W = 0.9323, p-value = 0.1529
> shapiro.test(group4)
        Shapiro-Wilk normality test
data:  group4
W = 0.8499, p-value = 5.181e-05
> shapiro.test(group5)
        Shapiro-Wilk normality test
data:  group5
W = 0.9034, p-value = 0.001820
dati <- c(group1, group2, group3, group4, group5)
gruppi <- factor(rep(letters[1:5],
times=c(length(group1),length(group2),length(group3),length(group4),length(group5))
))
levene.test(dati, gruppi, location="mean", kruskal.test=T)
        rank-based (Kruskal-Wallis) classical Levene's test based on the absolute
        deviations from the mean ( none not applied because the location is not set
        to median )
data:  dati
Test Statistic = 35.486, p-value = 3.691e-07
```

Essendo p-value < 0.05, rifiutiamo l'ipotesi di omogeneità delle varianze. Per confrontare le medie, dobbiamo quindi procedere con una ANOVA eteroschedastica (questi sono metodi parametrici, anche se i dati sono non-normali. Non avendo trovato dataset maggiormente adatti a questo esempio, dobbiamo fingere che i dati siano normali), utilizzando una delle seguenti procedure:
- **Welch one-way ANOVA**
- **Brown-Forsythe procedure** (non disponibile in R)
- **James first-and-second-order procedure** (non disponibile in R)

Per procedere con il test di Welch, il codice è il seguente:

```
oneway.test(smell ~ group, data=data, var.equal=F)
        One-way analysis of means (not assuming equal variances)
data:  smell and group
F = 13.7208, num df = 4.000, denom df = 78.749, p-value = 1.555e-08
```

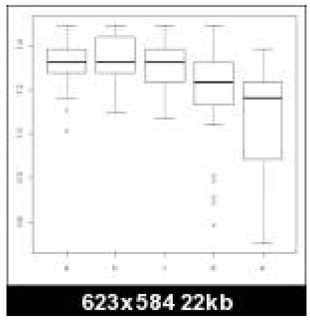
Essendo p-value < 0.05, rifiutiamo l'ipotesi nulla: le medie tra gruppi sono significativamente differenti.
A questo punto dovremmo procedere con i confronti multipli. Nel caso di eteroschedasticità, i test che dovremmo eseguire sono:
- **Games-Howell post-hoc test** (non disponibile in R)
- **Tamhane post-hoc test** (disponibile in SPSS)
- **Duncan-Waller post-hoc test**

Solo il terzo test è disponibile in R, e questo è il codice:

```
library(agricolae)
model <- lm(dati ~ gruppi)
df<-df.residual(model)
MSerror<-deviance(model)/df
Fc<-anova(model)[1,4]
comparison <- waller.test(dati, gruppi, df, MSerror, Fc, group=FALSE)
Study:
Waller-Duncan K-ratio t Test for dati
This test minimizes the Bayes risk under additive
loss and certain other assumptions.
                               ......
K ratio                  100.00000000
Error Degrees of Freedom 175.00000000
Error Mean Square          0.03211259
F value                   16.65064329
Critical Value of Waller   1.79200000
Treatment Means
  gruppi     dati    std.err replication
1      a 1.316895 0.01681486          38
2      b 1.345139 0.01762372          36
3      c 1.306143 0.02782045          21
4      d 1.201093 0.03349086          43
5      e 1.059619 0.03795035          42
Different MSD for each comparison
Comparison between treatments means
   tr.i tr.j       diff        MSD significant
1     1    2 0.02824415 0.07468760       FALSE
2     1    3 0.01075188 0.08731729       FALSE
3     1    4 0.11580171 0.07149772        TRUE
4     1    5 0.25727569 0.07189592        TRUE
5     2    3 0.03899603 0.08817637       FALSE
6     2    4 0.14404587 0.07254438        TRUE
7     2    5 0.28551984 0.07293687        TRUE
8     3    4 0.10504983 0.08549128        TRUE
9     3    5 0.24652381 0.08582458        TRUE
10    4    5 0.14147398 0.06966687        TRUE
```

Nell'ultima tabella, la colonna `significant` ci dice quali sono le coppie di gruppi con medie statisticamente differenti.
In pratica sono i gruppi 4 e 5 che presentano differenze tra loro e con tutti gli altri gruppi. Si può notare questo plottando i dati con il codice:

```
plot(gruppi, dati)
```



ANOVA non-parametrica eteroschedastica in R

Immaginiamo infine che entrambe le assunzioni (normalità ed omoschedasticità) siano violate. In questo caso si può procedere con il test di Friedman, e con i post-hoc test di Schaich-Hamerle, e Conover.
Il test di Friedman viene facilmente richiamato con la funzione `friedman.test()`, mentre il codice per il test post-hoc è disponibile qui.

Rimando a questo link per un esempio svolto: LINK.