MANOVA, nested and other designs

(STATS546.2: Applied analysis of variance (work in progress))

Anonymous

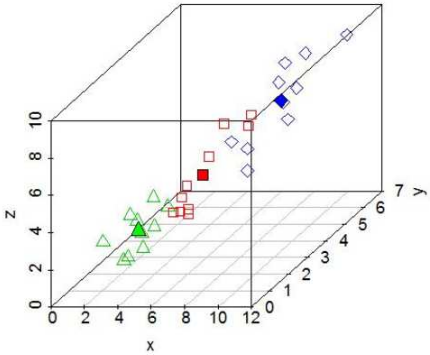
August 20, 2012

MANOVA

When we have many measurements on the same subject, e.g.

1. Petal length, sepal length etc. on plants
2. Heart rate, blood pressure etc on humans
3. Length, dorsal fin length, tail fin width etc on fish

we can use a multivariate ANOVA (MANOVA).
In ANOVA we are testing if means differ between groups but in MANOVA we test if there is a difference between group centroids.



Advantages and disadvantages

- Advantages
  - Controls the type I error.
  - More likely to observe difference between groups.
  - Take account of correlation between response variables.
- Disadvantages
  - can be difficult to see what variables are important.

One-way MANOVA

The simplest form of MANOVA is when we have one factor with two or more levels and p response variables.

The MANOVA is based on linear combination of the p variables.

$$: \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \\ \mu_{1p} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \\ \mu_{2p} \end{pmatrix} = \ldots = \begin{pmatrix} \mu_{g1} \\ \mu_{g2} \\ \\ \mu_{gp} \end{pmatrix}$$

$$H_1 \quad : \text{ not all } \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \\ \mu_{ip} \end{pmatrix} \text{are equal}$$

Test statistics

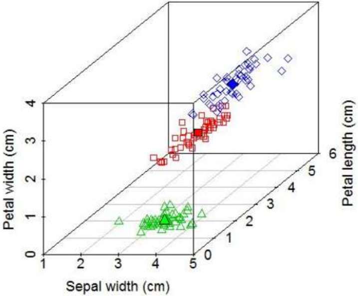
- Wilks
- Pillai
- Hotelling-Lawley
- Roy

Example - Iris

- Response variables
  - Petal length
  - Petal width
  - Sepal length
  - Sepal width
- Factor with 3 levels
  - Setosa
  - Versicolor
  - Virginica

Example - Iris

We want to know if there is a significant difference in the species group centroids.




```
Df Pillai approx F num Df den Df    Pr(>F)
Species     2 1.1899   71.485      6    292 < 2.2e-16 ***
Residuals 147
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

What variables are important?

It may be of interest to investigate what variables are the most important if the null hypothesis is rejected.

That can be investigated by doing

1. Univariate ANOVA
2. Step down analysis

Univariate ANOVA

Investigate the importance of each variables by doing univariate ANOVAs on each varibles separately.

Note that the univariate ANOVAs do not take correlation between response variables into account.

Step-down analysis

Step 1: Variables are ordered prior to the experiment of their importance to difference between groups.

Step 2: ANOVA carried out with the most important variable.

Step 3: ANCOVA carried out with the second most important variable as a response and the most important one as a covariate.

Step 4: ANCOVA carried out with the third most important variable as a response and the most and second most important ones as a covariates. This is the done sequentially for all the variables

Example - Skulls

Response variables

- maximum breaths of the skul l(mb)
- basibregmatic heights of the skull (bh)
- basialiveolar length of the skull (bl)
- nasal heights of the skull (nh)

Factor epoch with 5 levels

- 4000 BC
- 3300 BC
- 1850 BC
- 200 BC
- 150 AC

Example - Skulls

The centroids for the five epochs

$$\begin{pmatrix} \hat{\mu}_{mb} \\ \hat{\mu}_{bh} \\ \hat{\mu}_{bl} \\ \hat{\mu}_{nh} \end{pmatrix} \quad \begin{pmatrix} 131 \\ 134 \\ 99 \\ 51 \end{pmatrix} \quad \begin{pmatrix} 132 \\ 133 \\ 99 \\ 50 \end{pmatrix} \quad \begin{pmatrix} 134 \\ 134 \\ 96 \\ 51 \end{pmatrix} \quad \begin{pmatrix} 136 \\ 132 \\ 95 \\ 52 \end{pmatrix} \quad \begin{pmatrix} 136 \\ 130 \\ 94 \\ 51 \end{pmatrix}$$

```
           Df   Wilks approx F num Df den Df    Pr(>F)
epoch       4 0.66359   3.9009     16 434.45 7.01e-07 ***
Residuals 145
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Example - Skulls

There was found to be change over time in the Egyptian skulls but to see what variables were most important an individual analysis can be carried out.

variable	F-value	P-value
mb	5.95	0.0002
bh	2.45	0.0490
bl	8.31	$<$0.0001
nh	1.51	0.2032

Contrasts in MANOVA

It is possible to compare centroids between factor levels.

In multiple comparisons the significance level needs to be adjusted by using Bonferroni's method. This is however a conservative method.

Assumptions

Normality

Constant variance

Difficult to test in the multivariate case

Check these assumption for each variable separately.

multi-factor MANOVA

Just like in ANOVA we can have more than one factor in MANOVA.

MANCOVA

It is also possible to correct for some continuous variable, e.g. age, temperature etc.

MANOVA and repeated measures

MANOVA can also be used when we have repeated measures.

The easurement at each time is then considered separate variable and the number of the response variables will then be equal to the number of times measurements are made.

Example - Soils

Nine measurement of both physical and chemical characteristics of soil were made in microtopographic areas, categorized as Top, Slope and Depression were sampled at four different depth layers (0–10, 10–30, 30–60, 69–90 cm).

The question is, is there a difference in soil characteristics between areas and depths?

```
          Df Pillai approx F num Df den Df    Pr(>F)
Contour    2 1.1183   4.9323     18     70 5.929e-07 ***
Depth      3 1.6444   4.8524     27    108 1.761e-09 ***
Residuals 42
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Example - Soils

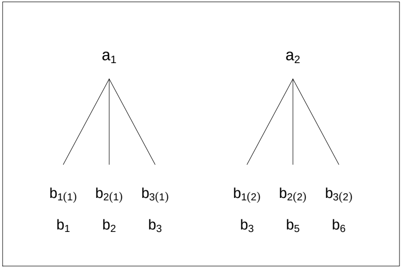
What variables are most important when it comes to difference in soil characteristics between depths and contour?

Results from step down analysis
Contour: K>Na>Ca>P>Conduc>Mg>N>Dens>pH
Depths: Na>Conduc>Ca>pH>P>N>Dens>K>Mg

Nested design

In nested designs we have factors nested within other factors.

We have two levels of factor A and in each level of factor A we have three levels of factor B, but these levels are of factor B are not the same across all levels of factor A as in a regular two-way ANOVA.



The nested model

The nested model is on the form

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

or

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$$

with the restrictions $\sum \alpha_i = 0$ and $\sum \beta_{j(i)} = 0$.

μ is the overall mean, α_i is the effect of level i in factor A and $\beta_j(i)$ is the effect of level j of factor B within level i of factor A and ϵ_{ijk} is the error as usual.

$$= y_{ijk} - \hat{y}_{ijk}$$

$$:\alpha_1 = \alpha_2 = ... = \alpha_i = 0$$

$$\mathrm{H}_1 \quad :\alpha_i \neq 0 \text{for all } i$$

Is there an effect of factor B within factor A

$$H_0 : \beta_{1(1)} = \beta_{2(1)} = ... = \beta_{j(1)} = ... = \beta_{1(2)} = \beta_{2(2)} = ... = \beta_{j(2)} ... = \beta j(i) =$$

$$H_1 : \beta_{j(i)} \neq 0 \text{for all } j \text{within } i$$

Nested model with random effects

The factors in nested designs can be random and it is common that the nested factor, factor B is random.

$$y_{ijk} = \mu + \alpha_i + b_{j(i)} + \epsilon_{ijk}$$

$b_{j(i)}$ follos a normal distribution with mean zero and variance σ_b^2.

ANOVA table for nested designs

Source	df	SS	MS
Factor A	$a-1$	$SSA = na\sum(\bar{y}_{i}.. - \bar{y}...)^2$	$MSA = \frac{SSA}{a-1}$
FactorB(A)	$a(b-1)$	$SSB(A) = n\sum\sum(\bar{y}_{j(i)}. - \bar{y}_{i}..)^2$	$MSB(A) = \frac{SS}{a(}$
Error	$ab(n-1)$	$SSE = \sum\sum\sum(y_{ijk} - \bar{y}_{j(i)}.)^2$	$MSE = \frac{SSE}{ab(n-}$
Total	$abn-1$	$SSTOT = \sum\sum\sum(y_{ijk} - \bar{y}...)^2$	

Test for effects of the factors

To test if there are factor effects a F-test is carried out.

$F_A = \frac{MSA}{MSE}$ follows a F-distribution with $a-1$ and $ab(n-1)$ degrees of freedom and the null hypothesis H_0 : all $\alpha_i = 0$ is rejected if $F_A > F_{1-\alpha,a-1,ab(n-1)}$.

$F_B = \frac{MSB(A)}{MSE}$ follows a F-distribution with $a(b-1)$ and $ab(n-1)$ degrees of freedom and the null hypothesis H_0 : all $\beta_{j(i)} = 0$ is rejected if $F_B > F_{1-\alpha,a(b-1),ab(n-1)}$.

Example