筑波大学
University of Tsukuba

Multiple factor Analysis of Variances (MANOVA)

Marco Antonio Florenzano Mollinetti[1]
[1]**University of Tsukuba, Systems Optimization Laboratory**
mollinetti@syou.cs.tsukuba.ac.jp

Before we Begin

- Go to the github repo:
  - https://github.com/Mollinetti/Statistics-R
- Download the script for this class! (in the ‘scripts’ folder, class_4.r!)
- Run the first lines to load/install the required libraries

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA (bonus round)

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA (bonus round)

Introduction

- MANOVA = Multiple Analysis of Variance
  - Multiple dependent variables (response)
  - Multiple independent variables (predictors)
- Generalization of the ANOVA
- Fitting multiple linear models
- Balanced/Unbalanced design
- Fixed/Random design

Introduction

- Sensitive to outliers
- Most of the analysis is done column wise
- MANOVA assumes that the dependent variables are sampled from a multivariate gaussian distribution

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA (bonus round)

Research Question of MANOVA

- Research questions for MANOVA are focused on differences
- Because MANOVA is a quantitative analysis, always identify your descriptors

Research Question of MANOVA

- Do differences exist in the percentages of male patients between or among this and that ethnic group that underwent treatment A or B, showing symptoms of disease 1 or 2?
- Do differences exist over different time periods of the shape of human skulls given measurements of width/length/depth?

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA (bonus round)

MANOVA

- Tests for the difference in two or more *vectors* of means
- Covariance is included
- Testing the multiple dependent variables is accomplished by creating new dependent variables that maximize group differences

MANOVA

- MANOVA is useful in experimental situations where at least some of the independent variables are manipulated
- Advantages:
  - Reduction of Type-I errors from multiple one-way ANOVA
  - Better chance of discovering significant factors

MANOVA

- However, one must be cautious:
  - The gain of power obtained from decreased SS error may be offset by the loss in these degrees of freedom
  - The dependent variables should be largely uncorrelated

MANOVA

- Special cases:
  - Unbalanced MANOVA: adjustment of the Sum squares must be done
  - Within-subject design: many dependent variables, there must be interaction terms

MANOVA

- We will use the egyptian skulls dataset
- Load the ’skulls.csv’ dataset
  - 1 independent variable (5 factors)
  - 4 dependent variables

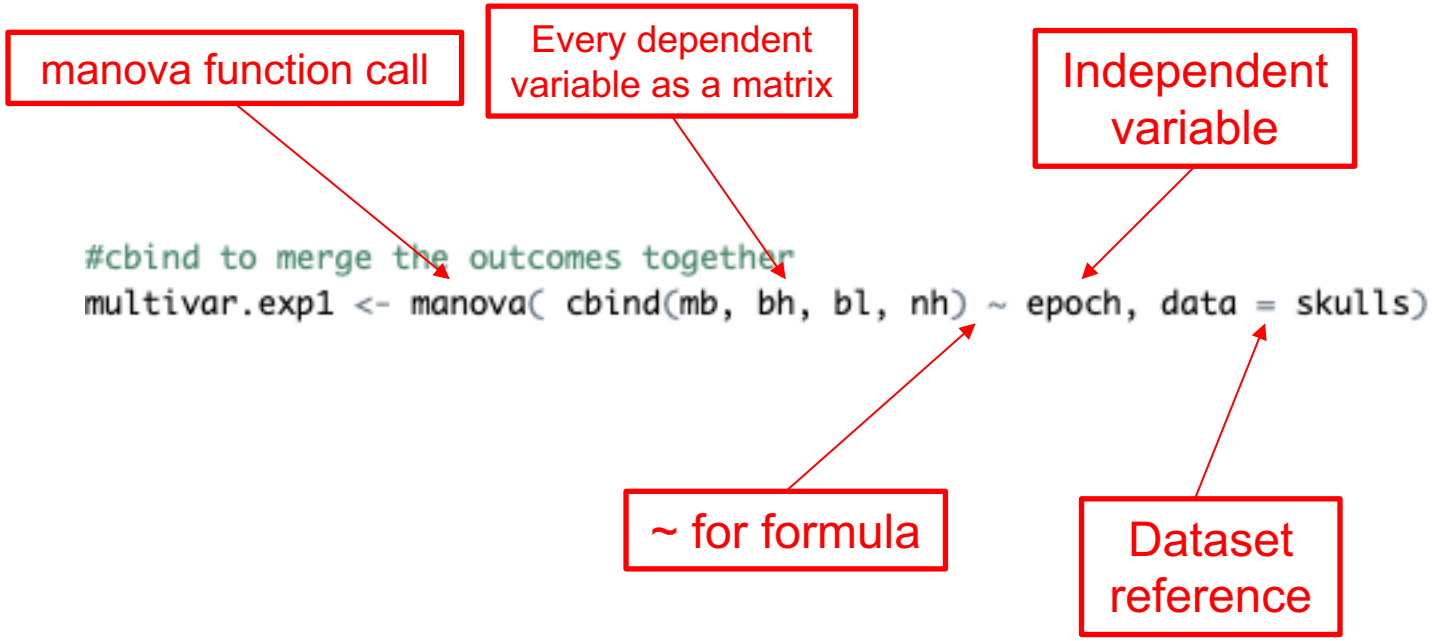
MANOVA

- R provides the manova function

```
#cbind to merge the outcomes together
multivar.exp1 <- manova( cbind(mb, bh, bl, nh) ~ epoch, data = skulls)
```

MANOVA

- R provides the manova function



Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA (bonus round)

Evaluating Extremeness

- The difference in means is called between-group sum of squares and cross-products matrix (S_B)
- Scatter within samples is called within-group sum of squares and cross-products matrix (S_w)
- MANOVA A-statistic: $A = SS_W^{-1} SS_B$

Evaluating Extremeness

- Because of the multivariate nature of the data, there is no unique measurement of "extremeness" A

Pillai's trace	$\sum \frac{1}{1+\lambda_k}$
Wilk's Λ	$\prod \frac{1}{1+\lambda_k}$
Hotelling-Lawley trace	$\sum \lambda_k$
Roy's largest root	$\frac{\lambda_{max}}{1+\lambda_{max}}$

λ are eigenvalues of matrix A

Evaluating Extremeness

```
> summary(multivar.exp1, test = "Pillai")
          Df  Pillai approx F num Df den Df    Pr(>F)
epoch      4 0.35331    3.512     16    580 4.675e-06 ***
Residuals 145
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(multivar.exp1, test = "Wilks")
          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
> summary(multivar.exp1, test = "Hotelling-Lawley")
          Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
epoch      4          0.48182    4.231     16    562 8.278e-08 ***
Residuals 145
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(multivar.exp1, test = "Roy")
          Df    Roy approx F num Df den Df    Pr(>F)
epoch      4 0.4251    15.41      4    145 1.588e-10 ***
Residuals 145
```

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA(bonus round)

Assumptions

1. Linearity of the dependent variables
2. Multivariate normality
3. Multivariate homogeneity of variance between groups
4. Multivariate homogeneity of covariance between groups

Assumptions

1. Linearity of the dependent variable Pearson r-test
2. Multivariate normality Multivariate Shapiro-wilk test
3. Multivariate homogeneity of variance between groups Levene's test
4. Multivariate homogeneity of covariance between groups Box-M test

Assumptions

Linearity of the dependent variable

- Pairwise analysis by Pearson-r test (H_0 =no correlation)
- Dependent variables can be correlated

```
> results.pearson
  var         mb          bh          bl          nh
1  mb 0.00000000 0.451725451 0.055067477 0.02536095
2  bh 0.45172545 0.000000000 0.001079603 0.07313252
3  bl 0.05506748 0.001079603 0.000000000 0.93823655
4  nh 0.02536095 0.073132521 0.938236552 0.00000000
```

Assumptions

Multivariate normality

- Done by multivariate Shapiro-Wilk test

```
> mshapiro.test(t(data.matrix(skulls[,2:5])))

        Shapiro-Wilk normality test

data:  Z
W = 0.98687, p-value = 0.1685
```

Assumptions

Multivariate homogeneity of variance between groups

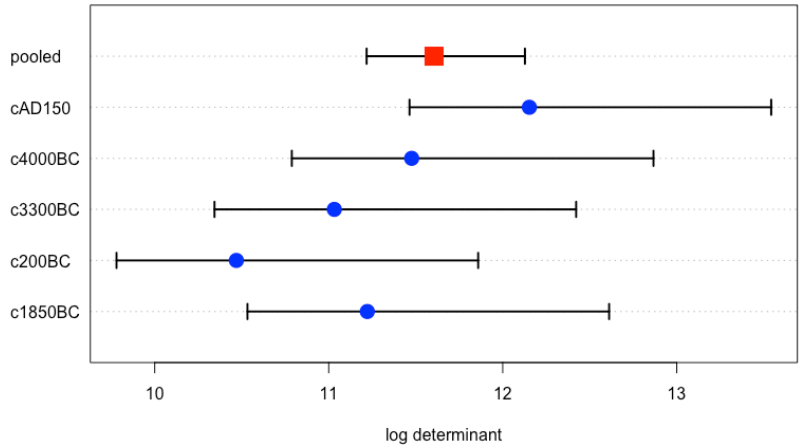
- Analysis by column-wise Levene test (Fligner-Kileen or Bartlett's test also ok)

```
> results.levene
  var         p
1  mb 0.3904702
2  bh 0.5815558
3  bl 0.6071019
4  nh 0.3878457
```

Assumptions

Multivariate homogeneity of covariance between groups

- Column-wise analysis by Box-M test




```
        Box's M-test for Homogeneity of Covariance Matrices

data:  skulls[, 2:5]
Chi-Sq (approx.) = 45.667, df = 40, p-value = 0.2483
```

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA(bonus round)

Post-hoc Analysis

- As a follow-up to the MANOVA some post-hoc tests can be done
- It is possible to verify which dependent variable influences the difference between groups

Post-hoc Analysis

- Possible tests:
  - One-way ANOVA for each dependent variable
  - LDA for each variable
  - Density plots
  - Tukey HSD for each variable

Post-hoc Analysis

```
> summary.aov(multivar.exp1)
 Response mb :
             Df  Sum Sq Mean Sq F value    Pr(>F)
epoch         4  502.83 125.707  5.9546 0.0001826 ***
Residuals   145 3061.07  21.111
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response bh :
             Df Sum Sq Mean Sq F value  Pr(>F)
epoch         4  229.9  57.477  2.4474 0.04897 *
Residuals   145 3405.3  23.485
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response bl :
             Df Sum Sq Mean Sq F value    Pr(>F)
epoch         4  803.3 200.823  8.3057 4.636e-06 ***
Residuals   145 3506.0  24.179
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response nh :
             Df Sum Sq Mean Sq F value Pr(>F)
epoch         4   61.2  15.300   1.507 0.2032
Residuals   145 1472.1  10.153
```

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA (bonus round)

Power Analysis

- For any post-hoc power analysis of MANOVA, it is recommended repeated experiments techniques such as:
  - Bootstrapping
  - Monte-Carlo simulations

Agenda

- Introduction
- Research Question of MANOVA
- MANOVA
- Evaluating Extremeness
- Assumptions
- Post-hoc Analysis
- Power Analysis
- Within-Subjects MANOVA (bonus round)

Within-Subjects MANOVA

- We will use the wolves skulls dataset
- Load the ’Wolves.csv’ dataset
  - 3 independent variables
  - 9 dependent variables

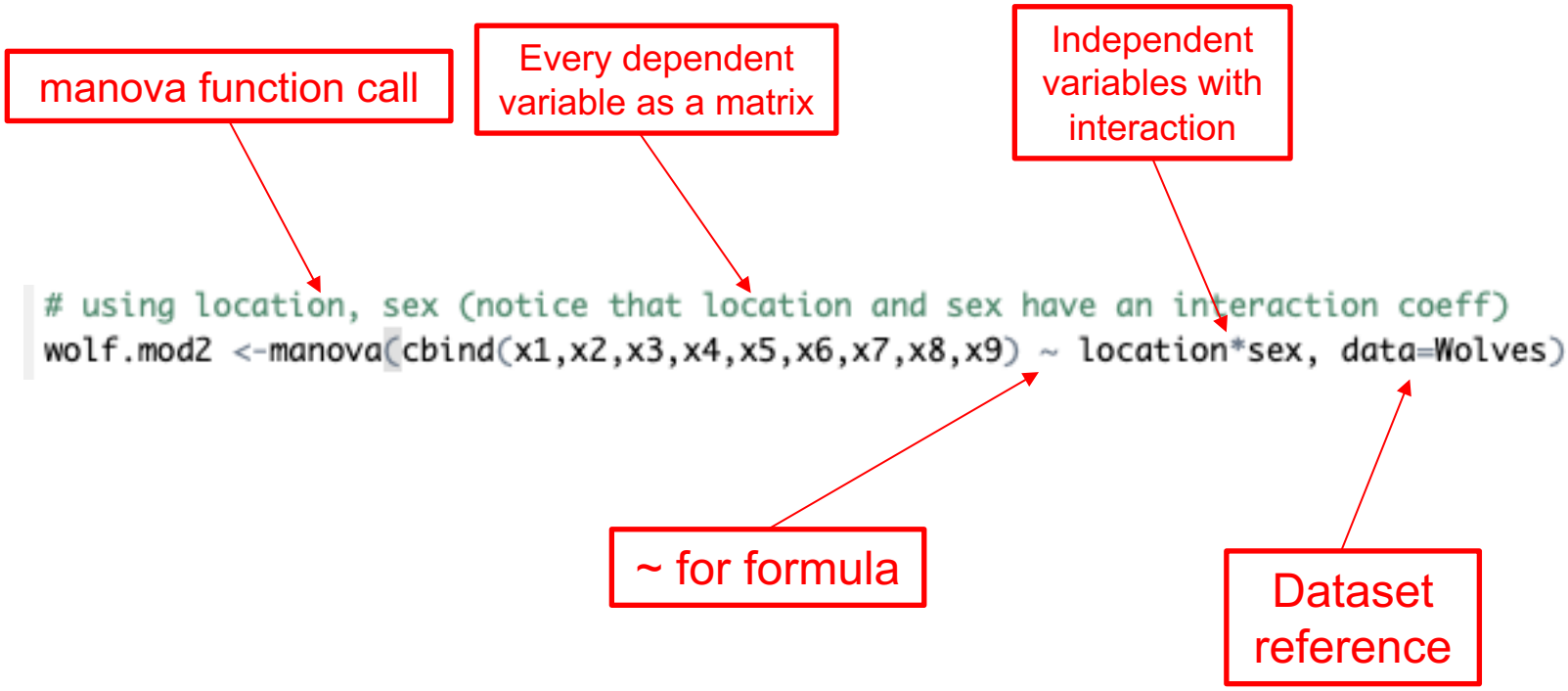
Within-Subjects MANOVA

- All assumptions and post-hoc tests are the same from the regular MANOVA
- However, interaction terms must be included for each independent variable

Within-Subjects MANOVA

```
# using location, sex (notice that location and sex have an interaction coeff)
wolf.mod2 <-manova(cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9) ~ location*sex, data=Wolves)
```

Within-Subjects MANOVA



Within-Subjects MANOVA

```
> summary(wolf.model)
          Df Pillai approx F num Df den Df    Pr(>F)
group      3 2.2454   4.9592     27     45 1.191e-06 ***
Residuals 21
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary.aov(wolf.model)
 Response x1 :
            Df Sum Sq Mean Sq F value    Pr(>F)
group        3 781.16  260.39  54.355 4.522e-10 ***
Residuals   21 100.60    4.79
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response x2 :
            Df Sum Sq Mean Sq F value    Pr(>F)
group        3 445.51 148.502  18.794 3.705e-06 ***
Residuals   21 165.93   7.902
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Next Episode

- Next episode, we will wrap-up with a review of everything we have seen so far
- Bring out your data if you have any questions!

筑波大学
University of Tsukuba

Multiple factor Analysis of Variances (MANOVA)

Marco Antonio Florenzano Mollinetti[1]
[1]**University of Tsukuba, Systems Optimization Laboratory**
mollinetti@syou.cs.tsukuba.ac.jp