筑波大学
University of Tsukuba

One-way Analysis of Covariance (ANCOVA)

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_5.r!)
- Run the first lines to load/install the required libraries

Agenda

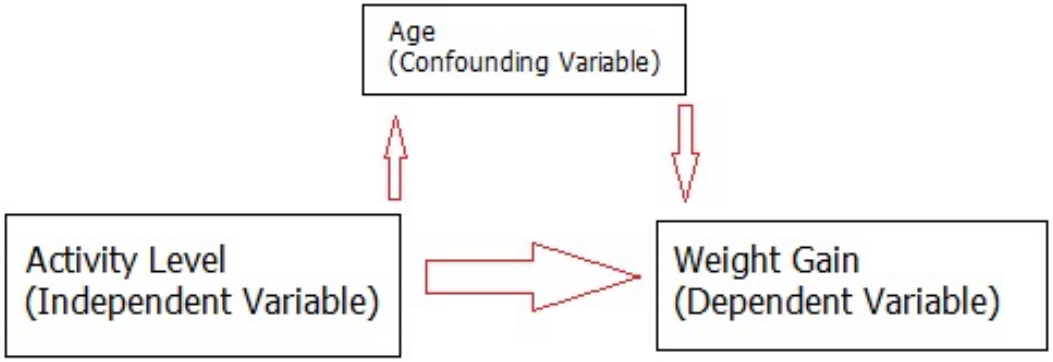
- Introduction
- ANCOVA
- ANOVA review

ANCOVA

- In clinical data, many confounding factors can compromise the test interpretation
- Example:
  - Fatigue
  - Motivation
  - Age
  - Education

ANCOVA

- Confounding factor: variable not accounted for
- Can suggest there is correlation when none
- Can introduce bias
- Absolute care must be taken to not include confounding variables



ANCOVA

- If we are interested in comparison between groups, we must take the following into account:
  - Decide which variables are significant
  - Ensure that the chosen variables are able to generalize the model (i.e. confounding variables do not bias the experiment)

ANCOVA

- Quality control of the experiments can be done in two ways:
  - Methodological
  - Statistically

ANCOVA

- Methodological quality control
  - Sample choice
  - Randomization
  - Validity
  - Stratification
  - Cutoffs
- Problem: large sample size, hard to generalize, budget constraints

ANCOVA

- Statistical quality control: if we can measure the covariant, it can be included in the regression model
- Error is now accounted
- Takes into account the inevitability of sample characteristics that cannot be factored out by design:
  - Age
  - Education
  - Intelligence

ANCOVA

- Covariate: continuous uncontrollable variable directly related to dependent variable
- The relationship between variables may be clouded by the covariate
- If we can measure the covariates, ANCOVA is preferred over ANOVA

ANCOVA

- Purpose of ANCOVA:
  - Error reduction, especially if group sizes are small
  - Control for factors which cannot be randomized but which can be measured on an interval scale.
  - Remove the effects of variables which modify the relationship of the independent to the dependent variable

ANCOVA

- Assumptions:
  - Linearity of regression
  - Homogeneity of error variances
  - Normality of residuals
  - Equality of groups on the covariate
  - equality of slopes–interaction

ANCOVA

- Assumptions:
  - Linearity of regression: plot of slopes
  - Homogeneity of error variances: levene test
  - Normality of residuals: shapiro-wilk test
  - Equality of groups on the covariate: anova or t-test between covariate and independent variable
  - equality of slopes–interaction: t-test or anova between covariate and the variables

ANCOVA

- Remember: just like ANOVA, ANCOVA can be:
  - Balanced/Unbalanced
  - One-way / Multiple (MANCOVA)
  - Fixed/ Random/ Mixed effects
- Usually, ANCOVA is fixed and balanced

ANCOVA

- We will use the ‘leprosy.dat’ dataset
- Variables
  - Dependent: Drug
  - Independent: X
  - Covariate: Y
- Neither Tukey nor mmc are suitable for post-hoc tests
- Instead the general linear hypothesis will be used for analysis

ANCOVA

- Before conducting the ANCOVA, some assumptions must be tested:

```
#equality of slopes interaction
fitted.lepr <- aov(After ~ Drug * Before, data = lepr)
```

ANCOVA

- Before conducting the ANCOVA, some assumptions must be tested:

```
#equality of slopes interaction
fitted.lepr <- aov(After ~ Drug * Before, data = lepr)
```

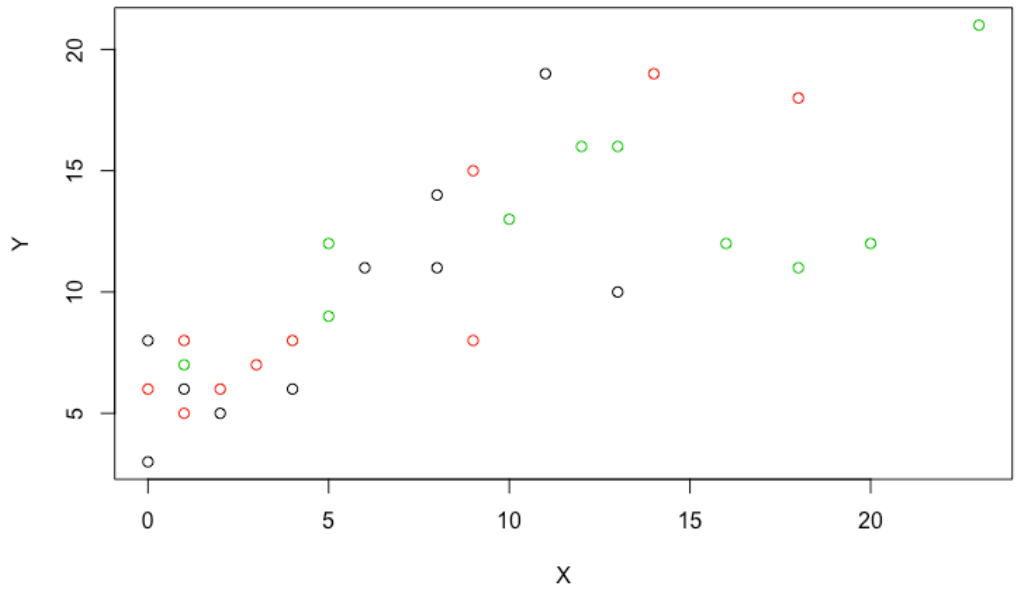
```
> summary(fitted.lepr)
             Df Sum Sq Mean Sq F value   Pr(>F)
Drug          2  293.6   146.8   8.862  0.00131 **
Before        1  577.9   577.9  34.887 4.29e-06 ***
Drug:Before   2   19.6     9.8   0.593  0.56058
Residuals    24  397.6    16.6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

H_0: equality of slopes interaction is not significant

ANCOVA

- Before conducting the ANCOVA, some assumptions must be tested:

```
#linearity of slopes (Eyeballing, just like Rick's neutrino bomb)
plot(After, Before, col = Drug, xlab = "Before", ylab = "After")
```



Pay attention to the pattern of each variable!

ANCOVA

- Before conducting the ANCOVA, some assumptions must be tested:

```
#Equality of the groups on the covariate
#Anova with the groups
fitted.eq<- aov(Before ~ Drug, data = lepr)
summary(fitted.eq)
```

ANCOVA

- Before conducting the ANCOVA, some assumptions must be tested:

```
#Equality of the groups on the covariate
#Anova with the groups
fitted.eq<- aov(Before ~ Drug, data = lepr)
summary(fitted.eq)
```

```
> summary(fitted.eq)
            Df Sum Sq Mean Sq F value Pr(>F)
Drug         2   72.9   36.43   1.659  0.209
Residuals   27  593.0   21.96
```

H_0: No difference between variances

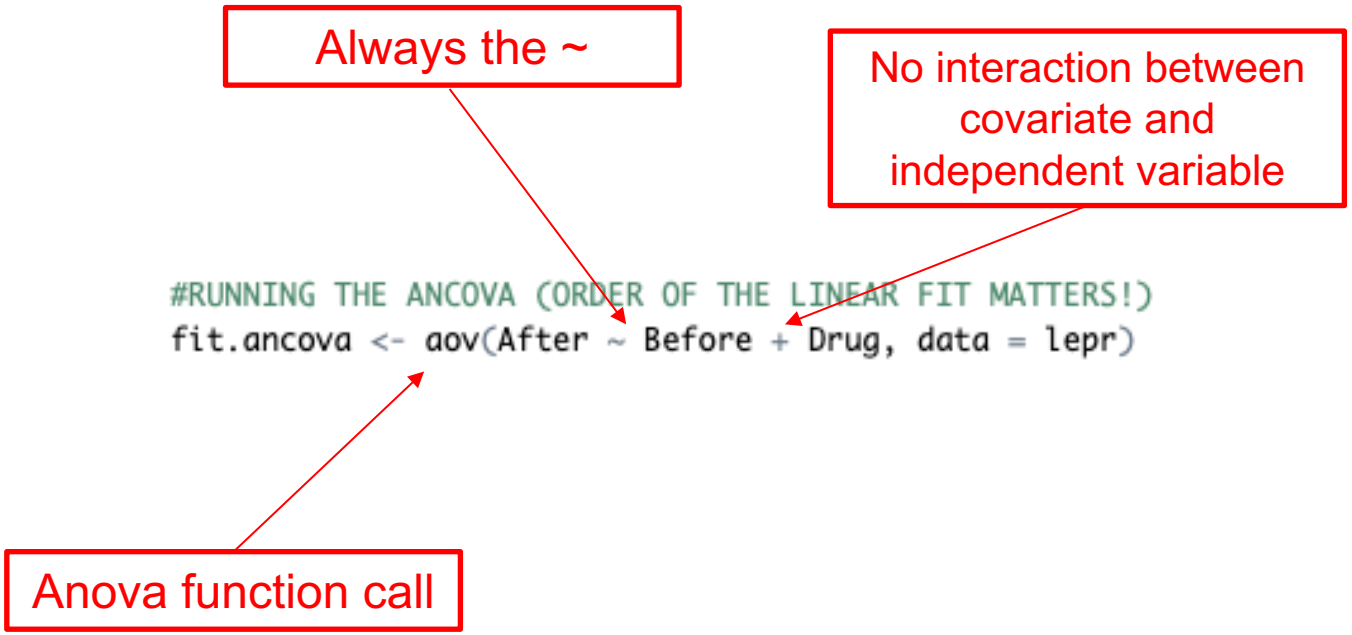
ANCOVA

- Clearing these assumptions, we call the ANCOVA

```
#RUNNING THE ANCOVA (ORDER OF THE LINEAR FIT MATTERS!)
fit.ancova <- aov(After ~ Before + Drug, data = lepr)
```

ANCOVA

- Clearing these assumptions, we call the ANCOVA



ANCOVA

- Posthoc is conducted by the glht function

```
# to get the 'real' p value and difference,
# uses the general linear hypothesis function
posthoc <- glht(fit.ancova, linfct = mcp(Drug = "Tukey"))

> summary(posthoc)

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = After ~ Before + Drug, data = lepr)

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
B - A == 0    0.109      1.795   0.061    0.998
C - A == 0    3.446      1.887   1.826    0.181
C - B == 0    3.337      1.854   1.800    0.189
(Adjusted p values reported -- single-step method)
```

ANCOVA

- Posthoc analysis is conducted by the glht function

```
# to get the 'real' p value and difference,
# uses the general linear hypothesis function
posthoc <- glht(fit.ancova, linfct = mcp(Drug = "Tukey"))

> summary(posthoc)

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = After ~ Before + Drug, data = lepr)

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
B - A == 0    0.109      1.795   0.061    0.998
C - A == 0    3.446      1.887   1.826    0.181
C - B == 0    3.337      1.854   1.800    0.189
(Adjusted p values reported -- single-step method)
```

H_0: No significant differences of means

Agenda

- Introduction
- ANCOVA
- ANOVA review

ANOVA review

- We will use the Insect sprays dataset that comes with the 'car' library
- Script will be written now!
- Answers will be in the 'Class_4_5_ans.r' file

Next Episode

- We will now proceed to fitting linear models with Linear regression

筑波大学
University of Tsukuba

One-way Analysis of Covariance (ANCOVA)

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