DISCOVERING STATISTICS USING R

Comparing Several Means: One-Way ANOVA

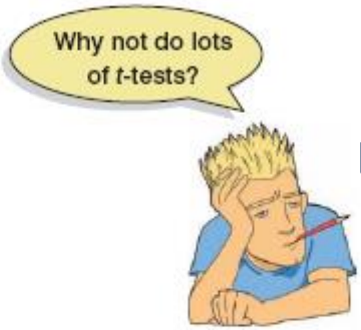
ANDY FIELD

Aims

- Understand the basic principles of ANOVA
  - Why it is done?
  - What it tells us?
- Theory of one-way independent ANOVA
- Following up an ANOVA:
  - *Post hoc* tests

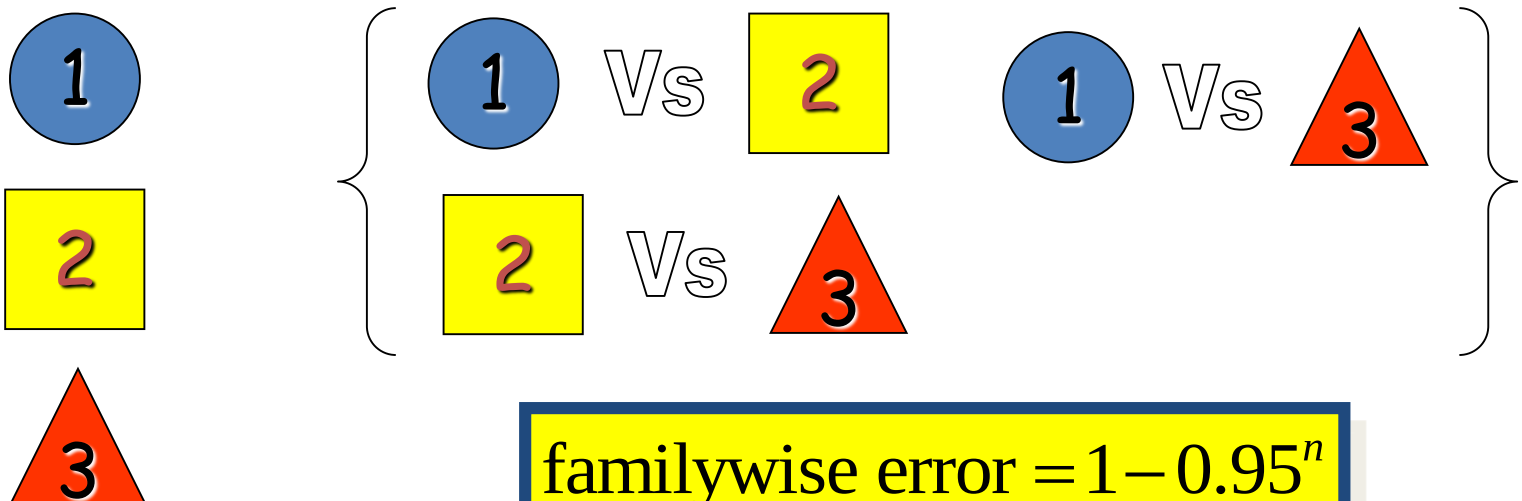
When and Why

- When we want to compare means we can use a *T*-test. This test has limitations:
  - You can compare only 2 means: often we would like to compare means from 3 or more groups.
  - It can be used only with one predictor/independent variable.
- ANOVA
  - Compares several means.
  - Can be used when you have manipulated more than one independent variable.
  - It is an extension of regression (the general linear model).



Why Not Use Lots of *T*-Tests?

- If we want to compare several means why don't we compare pairs of means with *T*-tests?
  - Can't look at several independent variables.
  - Inflates the Type I error rate.



$$\boxed{\text{familywise error} = 1 - 0.95^{n}}$$

Multiple Comparisons

Comparisonwise Error Rate (α=0.05)	Number of Comparisons	Experimentwise Error Rate (α=0.05)
.05	1	.05
.05	3	.14
.05	6	.26
.05	10	.40

What Does ANOVA Tell Us?

- Null hyothesis:
  - Like a *T*-test, ANOVA tests the null hypothesis that the means are the same.
- Experimental hypothesis:
  - The means differ.
- ANOVA is an omnibus test
  - It test for an overall difference between groups.
  - It tells us that the group means are different.
  - It doesn’t tell us exactly which means differ.

Assumptions ANOVA

1. The measure is interval-level continuous data.
2. The measure is normally distributed (within each group), *i.e.*, Y|X is normally distributed.
3. The variance of the measure is the same in each group.
4. The observations are independent

Note that 2,3 and 4 is the same as saying

$$\epsilon_i \sim i.i.d. N(0, \sigma^2)$$

Theory of ANOVA

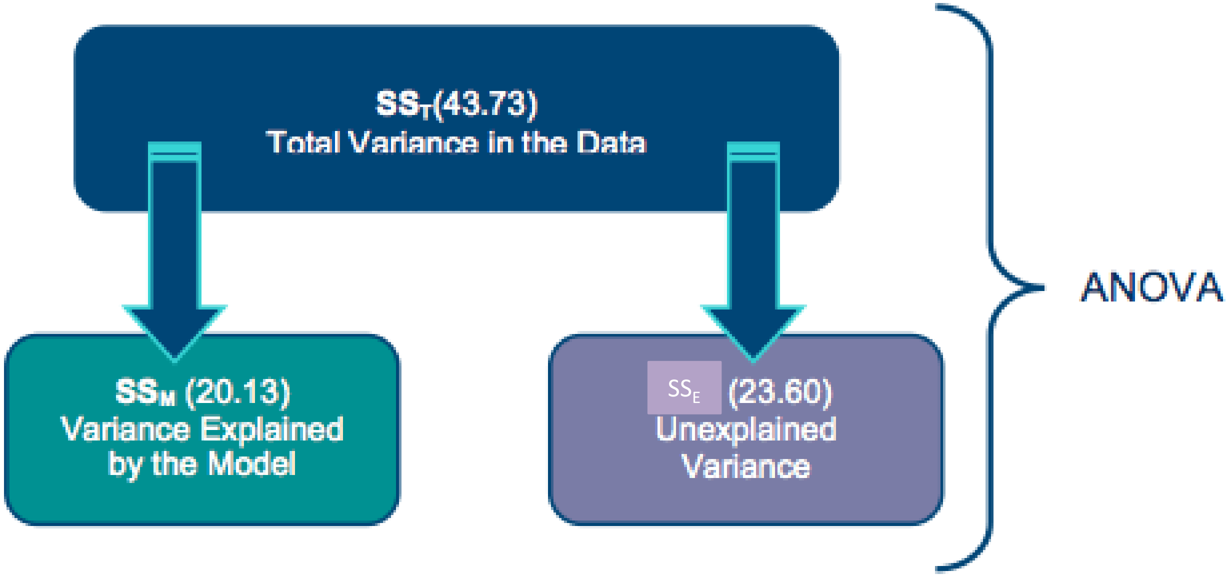


Figure 10.4: Partitioning variance for ANOVA

- If the experiment is successful, then the model will explain more variance than it can't
  - SS_M will be greater than SS_E

ANOVA by Hand

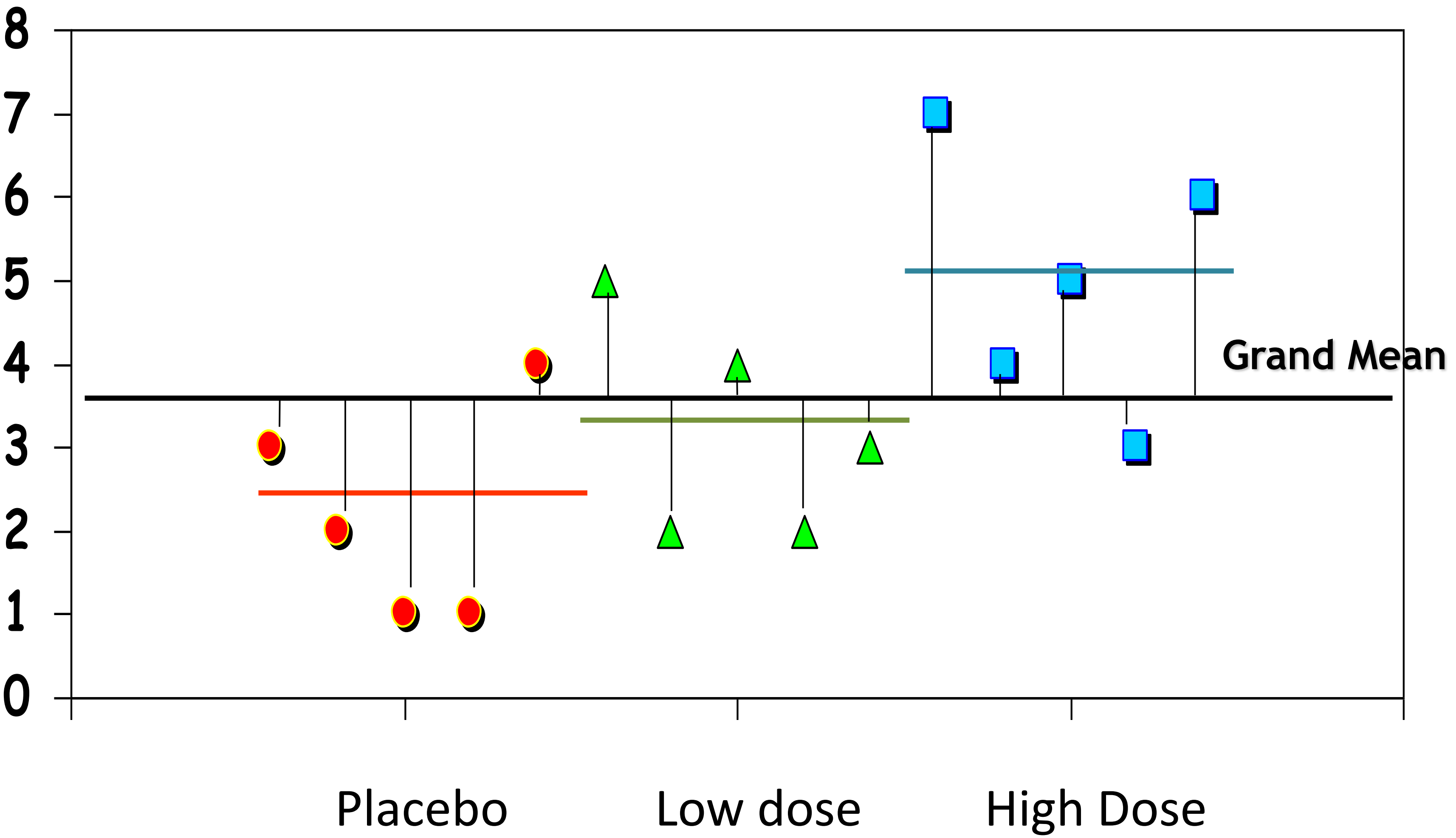
- Testing the effects of Viagra on libido using three groups:
  - Placebo (sugar pill)
  - Low dose viagra
  - High dose viagra
- The outcome/dependent variable (DV) was an objective measure of libido.

The Data

Table 10.1: Data in **Viagra.dat**

	Placebo	Low Dose	High Dose
x_{ij}	3	5	7
	2	2	4
	1	4	5
	1	2	3
	4	3	6
$\bar{x}_{.j}$	2.20	3.20	5.00
s_j	1.30	1.30	1.58
s_j^2	1.70	1.70	2.50
Grand Mean = 3.467 Grand SD = 1.767 $\bar{x}_{..}$ Grand Variance = 3.124 s^2			

Total Sum of Squares (SS_T):



Step 1: Calculate SS_T

$$SS_T = \sum_{i=1}^{N} (x_i - \bar{x}_{..})^2$$

$$s^2 = \frac{SS}{N-1}$$

$$SS = s^2(N-1)$$

$$SS_T = s^2(N-1)$$

$$SS_T = 3.124(15-1)$$

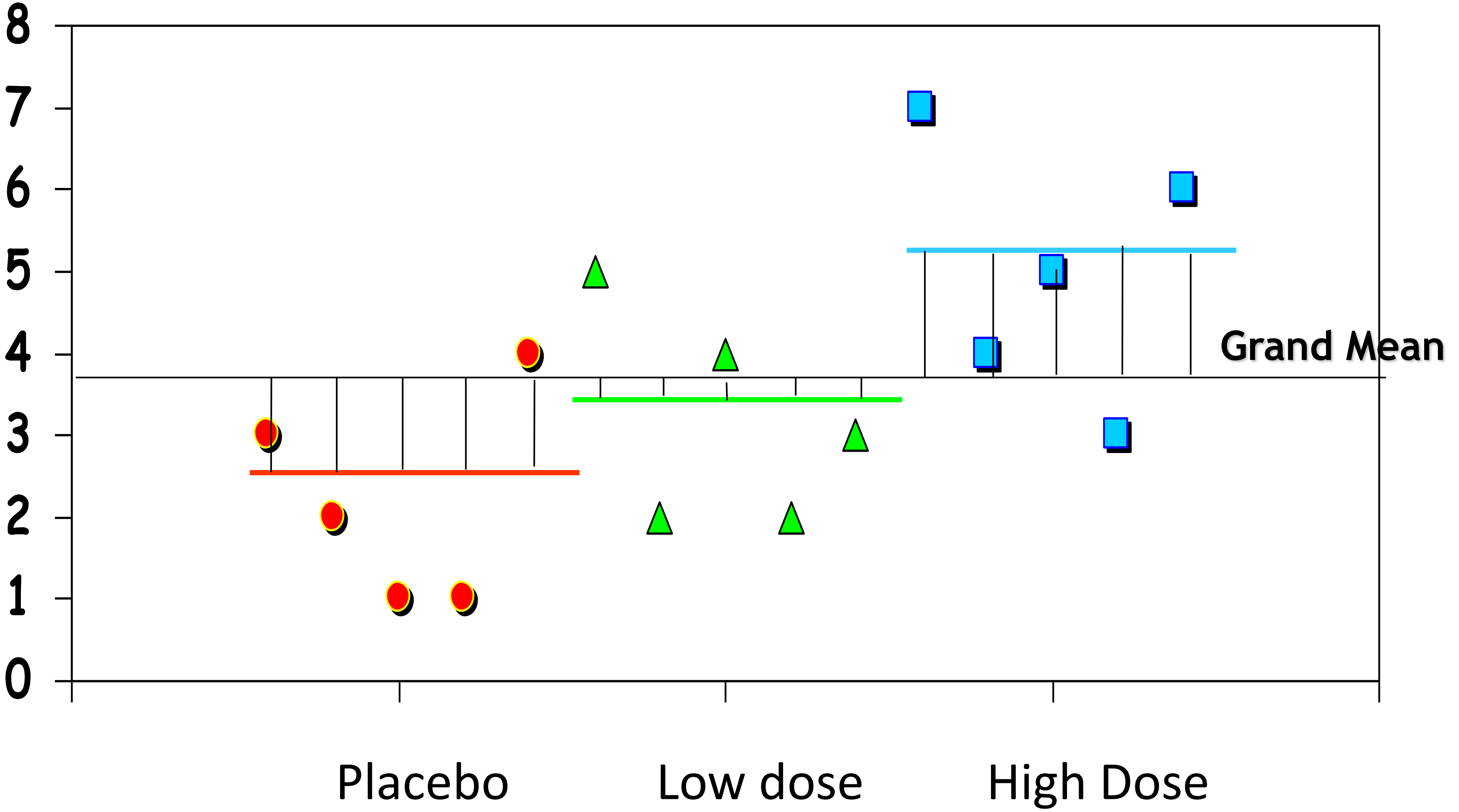
$$SS_T = 43.74$$

Degrees of Freedom

- Degrees of freedom (*df*) are the number of values that are free to vary.
  - Think about rugby teams!
- In general, the *df* are one less than the number of values used to calculate the SS.

$$df_T = N - 1 = 15 - 1 = 14$$

Model Sum of Squares (SS_M):



Step 2: Calculate SS_M

$$SS_M = \sum_{j=1}^{k} n_j(\bar{x}_{.j} - \bar{x}_{..})^2$$

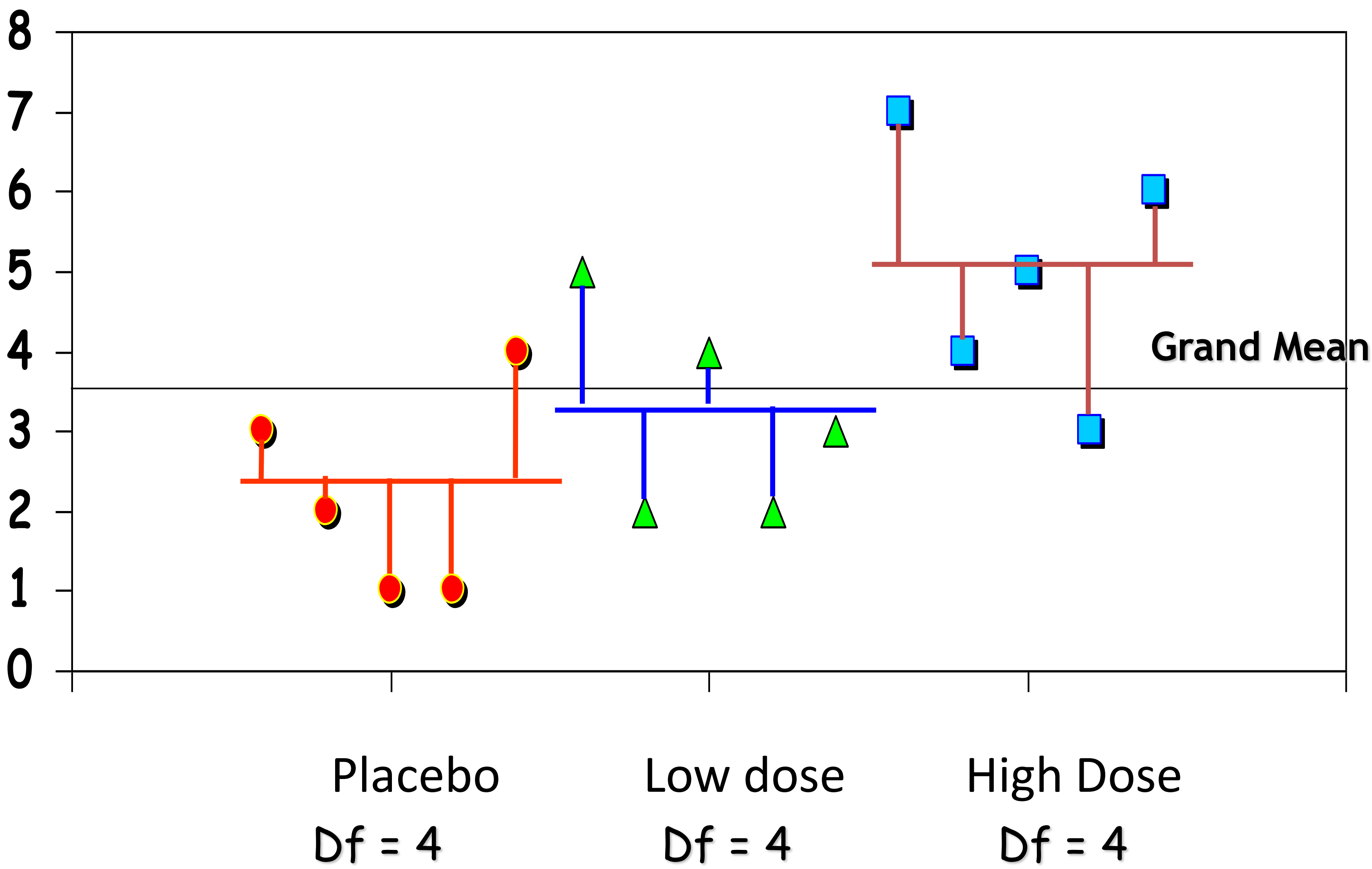
$$SS_M = 5(2.2 - 3.467)^2 + 5(3.2 - 3.467)^2 + 5(5.0 - 3.467)^2$$
$$SS_M = 20.135$$

Model Degrees of Freedom

- How many values did we use to calculate SS_M?
  - We used the 3 means.

$$df_M = k - 1 = 3 - 1 = 2$$

Residual Sum of Squares (SS_E):



Step 3: Calculate SS_E

$$SS_E = \sum_{j=1}^{k} \sum_{i=1}^{N} (x_{ij} - \bar{x}_{.j})^2$$

$$s^2 = \frac{SS}{N-1}$$

$$SS = s^2(N-1)$$

$$SS_E = \sum_{j=1}^{k} s_j^2(n_j - 1)$$

$$SS_E = s_1^2(n_1 - 1) + s_2^2(n_2 - 1) + s_3^2(n_3 - 1)$$

Step 3: Calculate SS_E

$$
\begin{aligned}
SS_E &= s_1^2(n_1 - 1) + s_2^2(n_1 - 1) + s_3^2(n_1 - 1) \\
&= 1.70(5 - 1) + 1.70(5 - 1) + 2.50(5 - 1) \\
&= 23.60
\end{aligned}
$$

Residual Degrees of Freedom

- How many values did we use to calculate SS_E?
  - We used the 5 scores for each of the SS for each group.

$$df_E = df_1 + df_2 + df_3$$
$$df_E = (n_1 - 1) + (n_2 - 1) + (n_3 - 1)$$
$$df_E = N - k = 12$$

Double Check

$$SS_T = SS_M + SS_E$$
$$43.74 = 20.14 + 23.60$$
$$43.74 = 43.74$$

$$df_T = df_M + df_E$$
$$14 = 2 + 12$$

Step 4: Calculate the Mean Squared Error

$$MS_M = \frac{SS_M}{df_M} = \frac{20.135}{2} = 10.067$$
$$MS_E = \frac{SS_E}{df_E} = \frac{23.60}{12} = 1.967$$

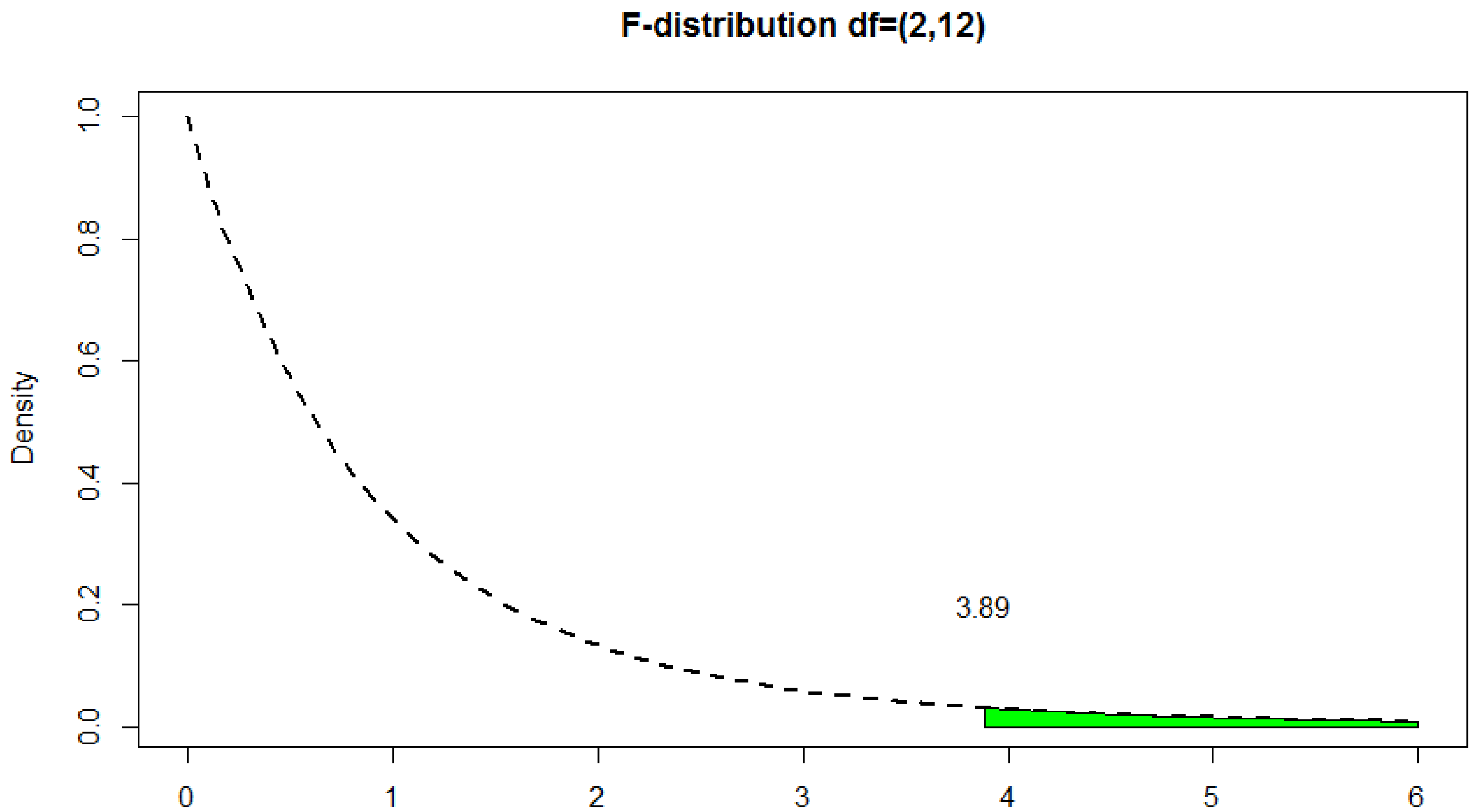
Step 5: Calculate the *F*-Ratio

$$F = \frac{MS_M}{MS_E} = \frac{10.067}{1.967} = 5.12$$

Step 6: Construct a Summary Table

Source	SS	*df*	MS	*F*
Model	20.14	2	10.067	5.12*
Residual	23.60	12	1.967	
Total	43.74	14		

F Statistic and Critical Values at α=0.05



F(Model df, Error df)=MS_M / MS_E

One-Way ANOVA using **R** When the Test Assumptions Are Met

- Using *lm()*:

```
viagraModel <- lm(libido~dose, data = viagraData)
Anova(viagraModel)
```

- In practice we verify the assumptions

Output

```
             Df Sum Sq Mean Sq F value  Pr(>F)
dose          2 20.133 10.0667  5.1186 0.02469 *
Residuals    12 23.600  1.9667
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Output 10.5

Verifying the assumptions

> plot(viagraModel)

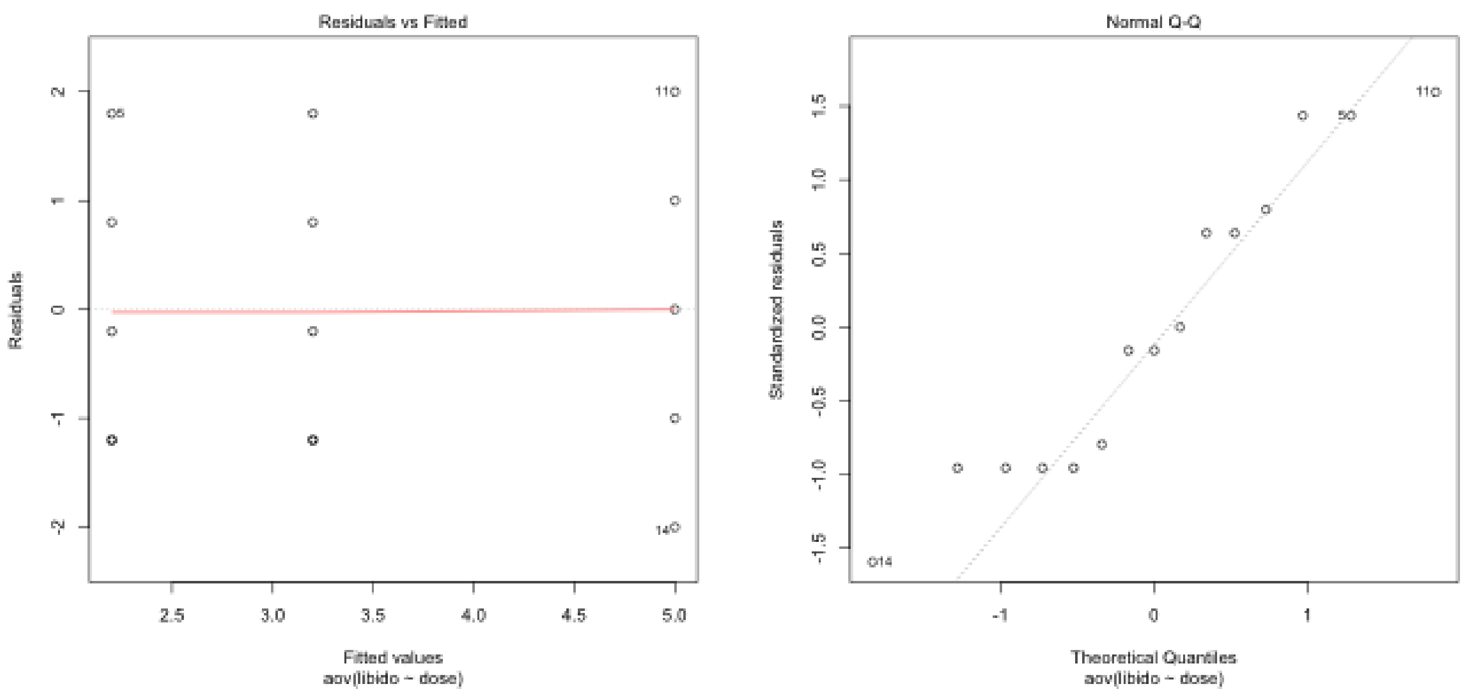


Figure 10.13: Plots of an ANOVA model

Why Use Follow-Up Tests?

- The *F*-ratio tells us only that the experiment was successful
  - i.e. group means were different
- It does not tell us specifically which group means differ from which.
- We need additional tests to find out where the group differences lie.

Multiple Comparison Methods

Control Comparisonwise Error Rate	**Pairwise t-tests**
Control Experimentwise Error Rate	**Compare All Pairs** **Tukey** **Compare to Control** **Dunnett**

Post Hoc Tests

- Compare each mean against all others.
- In general terms they use a stricter criterion to accept an effect as significant.
  - Hence, control the familywise error rate.
  - Simplest example is the Bonferroni method:

$$\boxed{Bonferroni\ \alpha = \frac{\alpha}{\text{number of tests}}}$$

Tukey's Multiple Comparison Method

- This method is appropriate when you consider pairwise comparisons only.
- The experimentwise error rate is
  - equal to alpha when ***all*** pairwise comparisons are considered
  - less than alpha when ***fewer*** than all pairwise comparisons are considered.

Special Case of Comparing to a Control

- Comparing to a control is appropriate when there is a natural reference group, such as a placebo group in a drug trial.
  - Control comparison computes and tests k-1 groupwise differences, where k is the number of levels of the classification variable.
  - An example is the *Dunnett* method.

Tukey {multcomp}

- For the Viagra data, we can obtain Tukey *post hoc* tests by executing:

{emmeans}

```
> lsm = emmeans(viagraModel, ~ dose)
> contrast(lsm, method="pairwise", adjust="tukey")
```

Output

```
 Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Linear Hypotheses:
                             Estimate Std. Error t value Pr(>|t|)
Low Dose - Placebo == 0        1.0000     0.8869   1.127   0.5162
High Dose - Placebo == 0       2.8000     0.8869   3.157   0.0208 *
High Dose - Low Dose == 0      1.8000     0.8869   2.029   0.1474
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dunnett {multcomp}

- For the Viagra data, we can obtain Dunnett *post hoc* tests by executing:

  {emmeans}

```
> lsm = emmeans(viagraModel, ~ dose)
> contrast(lsm, method="trt.vs.ctrl", ref=1, CIs=TRUE)
```

When variances are not equal across groups

- If Levene's test is significant then it is reasonable to assume that population variances are different across groups.
- We can get the output for Welch's *F* (variance weighted) for the current data by executing:

```
oneway.test(libido ~ dose, data = viagraData)
```

Output

```
	One-way analysis of means (not assuming equal variances)

data:  libido and dose
F = 4.3205, num df = 2.000, denom df = 7.943, p-value = 0.05374
```

Homogeneity of Variance

- Unless the group variances are extremely different or the number of groups is large, the usual ANOVA test is relatively robust when the groups are all about the same size.
- As Box (1953) notes,

"To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!"

quiz

- In the viagra dataset we compare 3 groups. How many dummy variables do we have in the regression model?
  a) 1
  b) 2
  c) 3
  d) 4

quiz

- In the viagra dataset we compare 3 groups. How many dummy variables doe we have in the regression model
  a) 1
  b) 2
  c) 3
  d) 4

workflow

- Summary statistics (cfr indep T-test)
- Graphs (cfr indep T-test)
- Statistical model
- Test the assumptions of a linear model (cfr regression)
- Post-hoc tests

ANOVA as Regression

$$libido_i = b_0 + b_1 X_1 + b_2 X_2 + e_i$$

X_1 = 1 if observation belongs to the Low Dose group and 0 otw

X_2 = 1 if observation belongs to the High Dose group and 0 otw

Dummy coding

observation	Intercept	Low Dose	High dose
Placebo	1	0	0
Placebo	1	0	0
Placebo	1	0	0
Placebo	1	0	0
Placebo	1	0	0
Low Dose	1	1	0
Low Dose	1	1	0
Low Dose	1	1	0
Low Dose	1	1	0
Low Dose	1	1	0
High Dose	1	0	1
High Dose	1	0	1
High Dose	1	0	1
High Dose	1	0	1
High Dose	1	0	1

Placebo Group

- $X_1 = X_2 = 0$

$$E(libido|Placebo) = b_0 = \bar{x}_{placebo}$$

Low Dose Group

- $X_1=1$, $X_2=0$

$$E(libido|Low) = b_0 + b_1 = \bar{x}_{Low}$$

$$b_1 = \bar{x}_{Low} - \bar{x}_{placebo}$$

High Dose Group

- X_1=0, X_2=1

$$E(libido|High) = b_0 + b_2 = \bar{x}_{high}$$

$$b_2 = \bar{x}_{high} - \bar{x}_{placebo}$$

Output from Regression

```
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.2000     0.6272   3.508  0.00432 **
## doseLow Dose      1.0000     0.8869   1.127  0.28158
## doseHigh Dose     2.8000     0.8869   3.157  0.00827 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.402 on 12 degrees of freedom
## Multiple R-squared:  0.4604, Adjusted R-squared:  0.3704
## F-statistic: 5.119 on 2 and 12 DF,  p-value: 0.02469
```

Question: What is the average value for the Low Dose group?

ANOVA as Regression

- No intercept model
- In the absence of an intercept, the regression coefficients correspond to the mean of each group
- Formula: y ~ 0 + dose
- Formula: y ~ dose - 1

ANOVA as Regression

$$libido_i = b_1X_1 + b_2X_2 + b_3X_3 + e_i$$

X_1 = 1 if observation belongs to the Placebo group and 0 otw

X_2 = 1 if observation belongs to the Low Dose group and 0 otw

X_3 = 1 if observation belongs to the High Dose group and 0 otw

"*different parameterization*"

No intercept model

observation	Intercept	Low Dose	High dose
Placebo	1	0	0
Placebo	1	0	0
Placebo	1	0	0
Placebo	1	0	0
Placebo	1	0	0
Low Dose	0	1	0
Low Dose	0	1	0
Low Dose	0	1	0
Low Dose	0	1	0
Low Dose	0	1	0
High Dose	0	0	1
High Dose	0	0	1
High Dose	0	0	1
High Dose	0	0	1
High Dose	0	0	1

Conditional models

- Placebo group: X_1=1

$$E(libido|Placebo) = b_1 = \bar{x}_{placebo}$$

- Placebo group: X_2=1

$$E(libido|Low) = b_2 = \bar{x}_{low}$$

- Placebo group: X_3=1

$$E(libido|High\ ) = b_3 = \bar{x}_{high}$$

lsmeans

- Least-squares means are predictions from a linear model
  - Balanced data: lsmeans = arithmetic group means
    - All SE are equal (due to homogeneity of variance assumption)
  - Unbalanced data: lsmeans ≠ arithmetic group means, adjusted for imbalance

Linear hypotheses {multcomp}

- Intercept model
- Eg. Test all pairwise comparisons
  - Low vs Placebo

    $E(libido|Placebo) = b_0$

    $E(libido|Low) = b_0 + b_1$

    Thus testing Low-Placebo is equal to testing b_1=0
  - High vs Low

    $E(libido|High) = b_0 + b_2$

    $E(libido|Low) = b_0 + b_1$

    Thus testing High-Low is equal to testing b_2-b_1=0
- R code

```
> L2 = rbind(c(0,1,0),c(0,0,1),c(0,-1, 1))
> rownames(L2) = c("low-placebo","high-placebo", "high-low")
> summary(glht(viagraModel, linfct = L2))
```

Linear hypotheses {multcomp}

- No intercept model
- Eg. Test vs control
  - Low vs Placebo

    $E(libido|Placebo) = b_1$

    $E(libido|Low) = b_2$

    Thus testing Low-Placebo is equal to testing b_2-b_1=0
  - High vs Placebo

    $E(libido|Placebo) = b_1$

    $E(libido|High) = b_3$

    Thus testing High-Placebo is equal to testing b_3-b_1=0
- R code

```
> L1 = rbind(c(-1,1,0),c(-1,0,1))
> rownames(L1) = c("low-placebo","high-placebo")
> summary(glht(viagraModel, linfct = L1))
```

Model comparison

- Compare residual SS from the one-way anova model to a model where dose is considered as a linear covariate
  - > anova(reduced model, full model)

10.1 Multiple Choice Poll

- If you have 20 observations in your ANOVA and you calculate the residuals, to which of the following would they sum?

a. -20
b. 0
c. 20
d. 400
e. Unable to tell from the information given

10.1 Multiple Choice Poll – Correct Answer

- If you have 20 observations in your ANOVA and you calculate the residuals, to which of the following would they sum?

a. -20
b. 0
c. 20
d. 400
e. Unable to tell from the information given

10.2 Multiple Choice Poll

- If you have 20 observations in your ANOVA and you calculate the squared residuals, to which of the following would they sum?

a. -20
b. 0
c. 20
d. 400
e. Unable to tell from the information given

10.2 Multiple Choice Poll – Correct Answer

- If you have 20 observations in your ANOVA and you calculate the squared residuals, to which of the following would they sum?

a. -20
b. 0
c. 20
d. 400
e. Unable to tell from the information given

10.3 Multiple Choice Poll

- Which part of the ANOVA tables contains the variation due to nuisance factors?

a. Sum of Squares Model
b. Sum of Squares Error
c. Degrees of Freedom

10.3 Multiple Choice Poll – Correct Answer

- Which part of the ANOVA tables contains the variation due to nuisance factors?

a. Sum of Squares Model
b. Sum of Squares Error
c. Degrees of Freedom

10.4 Multiple Answer Poll

- A study is conducted to compare the average monthly credit card spending for males versus females. Which statistical method might be used?

  a. One-sample *t*-test
  b. Two-sample *t*-test
  c. One-way ANOVA
  d. Two-way ANOVA

10.4 Multiple Answer Poll – Correct Answers

- A study is conducted to compare the average monthly credit card spending for males versus females. Which statistical method might be used?

  a. One-sample *t*-test
  b. Two-sample *t*-test
  c. One-way ANOVA
  d. Two-way ANOVA

DEMO ONE-WAY ANOVA

Open the program Ch10_glm1.R

Exercises

- Task 1
  - The **Superhero.dat** dataset contains data relating to children's injuries while wearing superhero costumes. Children reporting to the emergency centre at hospitals had the severity of their injury (injury) assessed (on a scale from 0, no injury, to 100, death). In addition, a note was taken of which superhero costume they were wearing (hero): Spiderman, Superman, the Hulk or a teenage Mutant Ninja Turtle. Use one-way ANOVA and multiple comparisons to test the hypotheses that different costumes are associated with more severe injuries.