

# R for Psych Handbook

*David John Baker*

*2018-03-22*



# Contents

<b>1</b>	<b>Preface</b>	<b>5</b>
<b>2</b>	<b>Introduction to R</b>	<b>7</b>
<b>3</b>	<b>Data Manipulation in R</b>	<b>9</b>
<b>4</b>	<b>Episotomology of Statistics</b>	<b>11</b>
<b>5</b>	<b>Descriptive Statistics, z Scores, Central Limit</b>	<b>13</b>
5.1	Example one . . . . .	13
5.2	Example two . . . . .	13
<b>6</b>	<b>Data Transformations and Confidence Intervals</b>	<b>15</b>
<b>7</b>	<b>Power and Effect Size Measures</b>	<b>17</b>
<b>8</b>	<b>Correlation and Regression</b>	<b>19</b>
<b>9</b>	<b>Matched T Test</b>	<b>21</b>
<b>10</b>	<b>One Way ANOVA</b>	<b>23</b>
<b>11</b>	<b>Multiple Comparisons</b>	<b>25</b>
<b>12</b>	<b>Simple Effects, Interactions, Mixed Designs</b>	<b>27</b>
<b>13</b>	<b>Repeated Measures ANOVA:</b>	<b>29</b>
<b>14</b>	<b>Mixed Two Way ANOVA</b>	<b>31</b>
<b>15</b>	<b>Multiple Regression</b>	<b>33</b>
<b>16</b>	<b>Chi-Square</b>	<b>35</b>
<b>17</b>	<b>Non-Parametric Data</b>	<b>37</b>
<b>18</b>	<b>Advanced Data Cleaning</b>	<b>39</b>
<b>19</b>	<b>Advanced Multiple Regression</b>	<b>41</b>
<b>20</b>	<b>Logistic Regression</b>	<b>43</b>
<b>21</b>	<b>Mediation and Moderation</b>	<b>45</b>

<b>22 ANCOVA</b>	<b>47</b>
22.1 Theory . . . . .	47
22.2 Practice . . . . .	48
<b>23 MANOVA</b>	<b>53</b>
<b>24 Repeated Measures ANOVA</b>	<b>55</b>
<b>25 Factor Analysis</b>	<b>57</b>
<b>26 Mixed Effects Models</b>	<b>59</b>
<b>27 Confirmatory Factor Analysis</b>	<b>61</b>
<b>28 This is a template file</b>	<b>63</b>

# Chapter 1

## Preface

This book serves as a collection of resources used in the LSU psychological statistics courses. It contains some lecture notes, as well as R code and data to run each of the examples in R.

The book is still under construction.



## Chapter 2

# Introduction to R

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2017) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

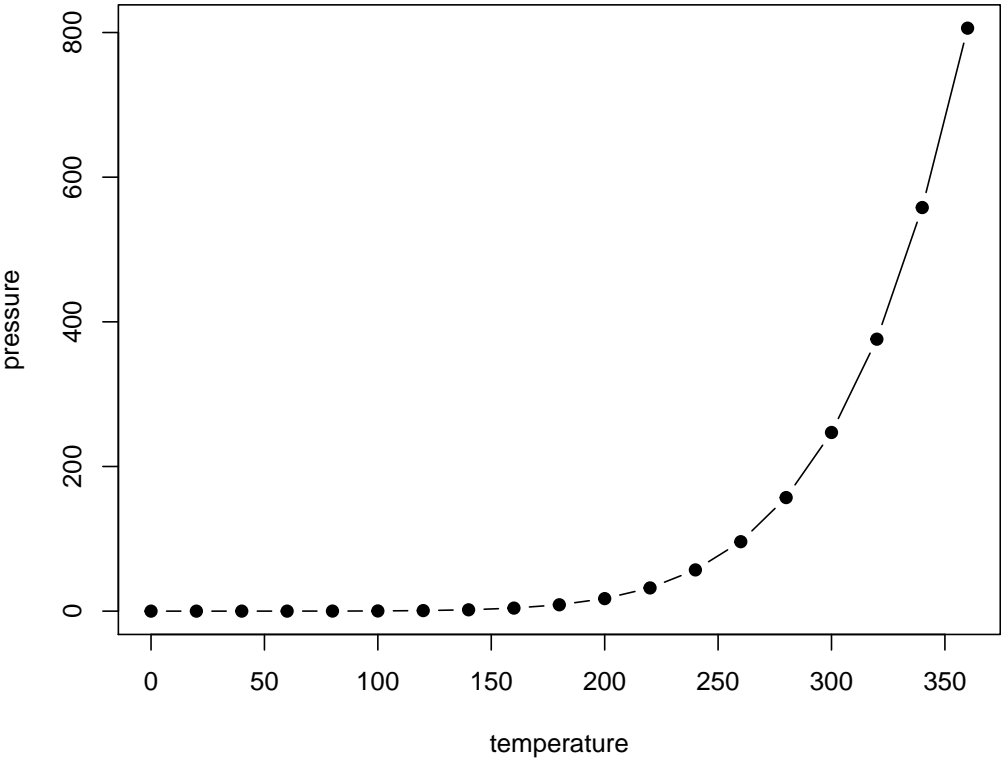


Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa



## Chapter 3

# Data Manipulation in R

Here is a review of existing methods.



## Chapter 4

# Episotomology of Statistics

We describe our methods in this chapter.



## Chapter 5

# Descriptive Statistics, z Scores, Central Limit

Some *significant* applications are demonstrated in this chapter.

### 5.1 Example one

### 5.2 Example two



## Chapter 6

# Data Transformations and Confidence Intervals

We have finished a nice book.





## Chapter 7

# Power and Effect Size Measures

attempt here



## Chapter 8

# Correlation and Regression

attempt here



## Chapter 9

# Matched T Test

Template file



## Chapter 10

# One Way ANOVA

Template file





## Chapter 11

# Multiple Comparisons

Template file



## Chapter 12

# Simple Effects, Interactions, Mixed Designs

Template file



## Chapter 13

# Repeated Measures ANOVA:

Template file



## Chapter 14

# Mixed Two Way ANOVA

Template file





## Chapter 15

# Multiple Regression

Template file



## Chapter 16

# Chi-Square

Template file



## Chapter 17

# Non-Parametric Data

Template file



## Chapter 18

# Advanced Data Cleaning

Template file





## Chapter 19

# Advanced Multiple Regression

Template file



## Chapter 20

# Logistic Regression

Template file



## Chapter 21

# Mediation and Moderation

Template file



# Chapter 22

## ANCOVA

### 22.1 Theory

#### 22.1.1 Reducing Noise

Just like the ANOVA (Analysis of Variance), the ANCOVA (Analysis of Covariance) is used to analyze experiments by calculating an F ratio for more than 2 groups. As with any F calculation, differences in the dependent variable are caused by two things:

1. The independent variable (signal, systematic variation)
2. Error (noise, unsystematic variation)

The F statistic captures this

$$F = \frac{\text{Variation Due to IV}}{\text{Variation Due to Error}}$$

When there is small variation within groups and large variation between groups, we get a large F.

CHART HERE

When there is large variation within groups and small variation between groups, we get a small F.

The idea with the ANCOVA is that you can reduce your error term (the denominator from above) by choosing a covariate (CV) that is related to the DV and will soak up some random variation in your F test to give you a clearer picture of what is going on. Usually this means controlling for some sort of variable. The classic example is if you wanted to give some kids a test of some mental ability in a between subjects design but you don't want their age to skew your results. You might have had a control condition, an intervention, and some sort of alternate intervention. Typically you would run an ANOVA on the three groups, but since you know age has been accounted for and you want to remove that effect from the model, you enter age as a covariate in this calculation. Outside of this there are **three** major applications for ANCOVA.

1. Increase test sensitivity by using the CV(s) to account for more of the error variance
2. Adjust DV scores to what they would be if everyone scored the same on the CV(s)
3. Adjustment of a DV for other DVs taken as CVs

Note that use of a CV can adjust DV scores and show a larger effect or the CV can even eliminate the effect. Looking at this another way:

1. Reduces random error by increasing the size of F
2. Reduces systematic error by adjusting for differences in means
3. May increase differences by soaking up error

It's a **good** idea to use ANCOVA when you are removing variance in the DV related to covariate, but not related to the grouping variable. This decreases the error term and increases power.

It's a **bad** idea to use ANCOVA when groups differ on their mean level of the covariate. Usually here the covariate and the grouping variable are not independent. An example of this might be when "controlling" for anxiety when studying people with and without depression. Clearly people with depression will have higher levels of anxiety than their controls by nature of having depression!

### 22.1.2 Assumptions of ANOVA

1. All ANOVA assumptions apply (errors are RIND)
  - Random
  - Independent
  - Normally distributed
2. The Covariate  $z$  is unrelated to  $x$ , and that  $z$  is related to, and in a sense, acts as a suppressor of  $y$
3. The covariate has a **linear** relationship with the dependent variable
4. Variance of groups are equal.
5. Correlation between  $y$  and  $z$  is equal for all levels of  $x$  (Homogeneity of Regression)

You test for homogeneity of regression slopes by including a covariate by group interaction. If it's significant, you violated an assumption and can't do ANCOVA!

CHART

CHART

### 22.1.3 Best Practice

Things you should control for: 1. Broad Descriptors (Age, Social Class) 2. General Ability (intelligence, memory, speed of movement) 3. Personality measures (extraversion, neuroticism) 4. Things you think are relevant 5. Pick CVs that are correlated with DV \* Height good for basketball \* Height poor for test of math skills 6. CV should not be influenced by treatment

## 22.2 Practice

In order to get our hands dirty with ANCOVA we're going to look at a dataset where a bunch of sheep were slaughtered. We have three types of sheep (ewe, wether, ram) and a measure of fatness and weight.

Let's first run an ANOVA to see if fatness differs between animal type.

```
library(data.table)
deadsheep <- fread("datasets/deadsheep.csv")
deadsheep[, Animal := as.factor(Animal)]
sheep.model.1 <- aov(fatness ~ Animal, data=deadsheep)
summary(sheep.model.1)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Animal      2  124.4    62.21   3.165 0.0566 .
## Residuals   30  589.6    19.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Looking at the output, we note that There was no significant effect of fatness on levels of animal,

$$F(2, 30) = 3.165, p > .05$$

.

Now since we know that the whole weight of the animal might be confounding our answer (it's related to the DV of fatness, but not necessarily related to the type of sheep), let's run an ANCOVA with carcass weight as a covariate and see if that changes our answer.

Before we run the ANCOVA, we need to first check that our IV is not significantly related to our CV. We do this by running an ANOVA predicting weight by animal.

```
#Test IND of IV and CV
ind.sheep <- aov(weight ~ Animal, data=deadsheep) ## Check CV for IV indep, want non-sig
summary(ind.sheep) # p=.104

##           Df Sum Sq Mean Sq F value Pr(>F)
## Animal      2  14.20   7.102   2.438  0.104
## Residuals   30  87.41   2.914
```

Here we get

$$F(2, 30) = 2.44, p > .05$$

and can move forward. Let's now run our ANCOVA. Note that R's default is Type I Sum of Squares, we need to load the `car` package to give us the correct sum of squares. For further reading on this, check out the Andy Field book on this.

```
sheepANCOVA <- aov(fatness ~ Animal + weight, data=deadsheep)
library(car)
Anova(sheepANCOVA, type="III")

## Anova Table (Type III tests)
##
## Response: fatness
##           Sum Sq Df F value    Pr(>F)
## (Intercept) 117.17  1  26.529 1.669e-05 ***
## Animal      332.31  2  37.621 8.772e-09 ***
## weight      461.56  1 104.506 3.994e-11 ***
## Residuals   128.08 29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can then report

The covariate, weight, was not significantly related to the the fatness of the animal,

$$F(1, 29) = , p = .0565$$

and that there was a significant effect of the animal after controlling for weight of the animal,

$$F(2, 28) = 37.62, p < .05.$$

.

Let's now check on our marginal means. These are our group means after we have controlled for our covariate. For this we need the `effects` package.

```
library(effects)
mean.sheep <- effect("Animal", sheepANCOVA, se=TRUE)
summary(mean.sheep)
```

```
##
## Animal effect
## Animal
##      Ewe      Ram    Wether
## 18.90219 10.53996 14.83058
##
## Lower 95 Percent Confidence Limits
## Animal
##      Ewe      Ram    Wether
## 17.565871  9.183322 13.532447
##
## Upper 95 Percent Confidence Limits
## Animal
##      Ewe      Ram    Wether
## 20.23851 11.89660 16.12870
```

We can then note the estimated marginal means here (18.9 for Ewe, 10.53 for Ram, and 14.83 for Wether) are mean values for fatness once the CV of weight has been controlled for.

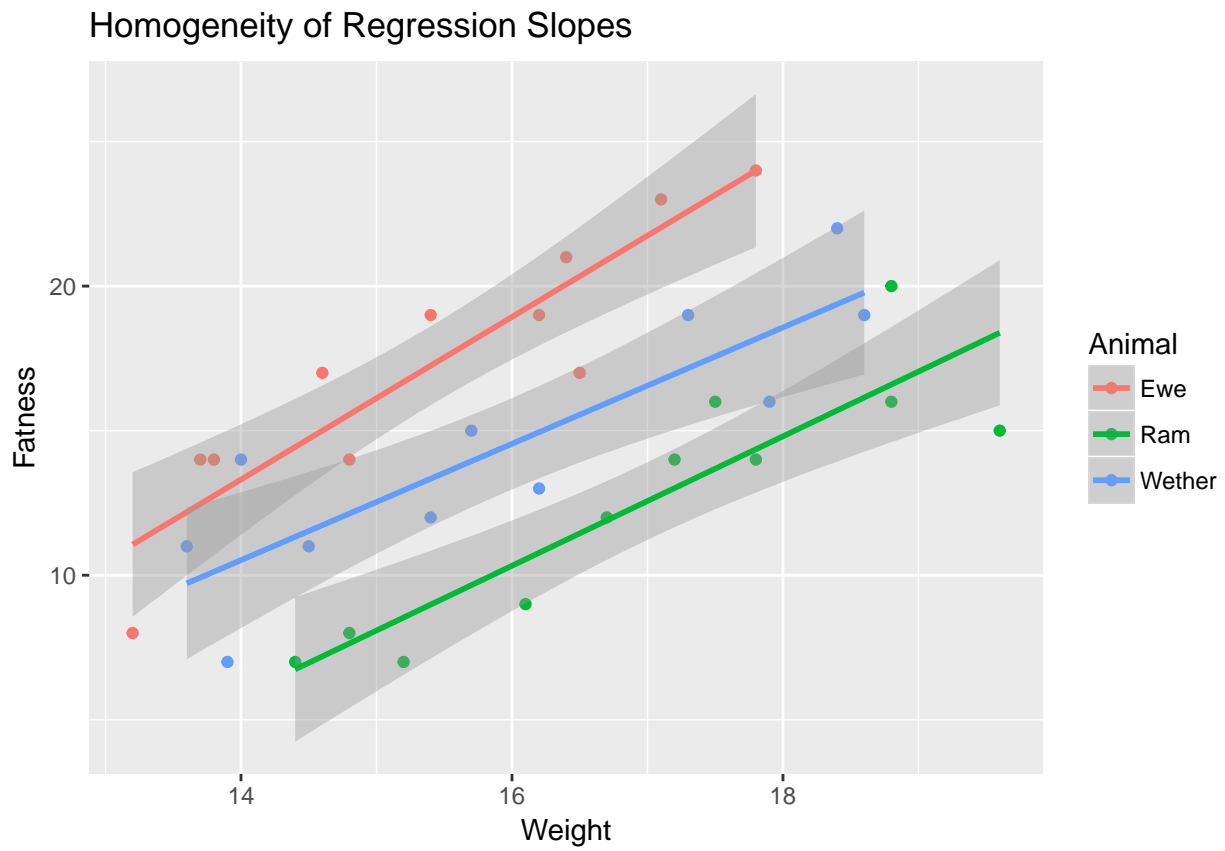
Lastly, let's test the homogeneity of the regression slopes assumption (though we should have done this first!). We do this by running the model with an interaction term included.

```
hoRS.sheep <- aov(fatness ~ weight + Animal + animal:weight, data=deadsheep)
Anova(hoRS.sheep, type="III")
```

```
## Anova Table (Type III tests)
##
## Response: fatness
##           Sum Sq Df F value    Pr(>F)
## (Intercept)  64.571  1 14.6355 0.0007007 ***
## weight      180.144  1 40.8308 7.624e-07 ***
## Animal        5.293  2  0.5998 0.5560672
## weight:animal  8.957  2  1.0151 0.3757923
## Residuals    119.123 27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction term is not significant, thus no violation. We can also plot this to see it. All regression lines should be pointing in the same direction when we plot the groups separately.

```
library(ggplot2)
ggplot(deadsheep, aes(x = weight, y = fatness, color = Animal)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Homogeneity of Regression Slopes", x = "Weight", y = "Fatness")
```



If one of these lines would be pointed downwards, we would have an interaction.



## Chapter 23

# MANOVA

Template file



## Chapter 24

# Repeated Measures ANOVA

Template file





## Chapter 25

# Factor Analysis

Template file



## Chapter 26

# Mixed Effects Models

Template file



## Chapter 27

# Confirmatory Factor Analysis

Template file



## Chapter 28

# This is a template file

Template file





# Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2017). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.5.