STAT 245 Course Notes

Stacy DeRuiter, Calvin University

2019-09-09

Contents

Chapter 1

Description

This is a set of course notes distributed in STAT 245 at Calvin University in Fall 2019. Contact sld33 at calvin.edu with comments, corrections or suggestions.

Chapter 2

Linear Regression

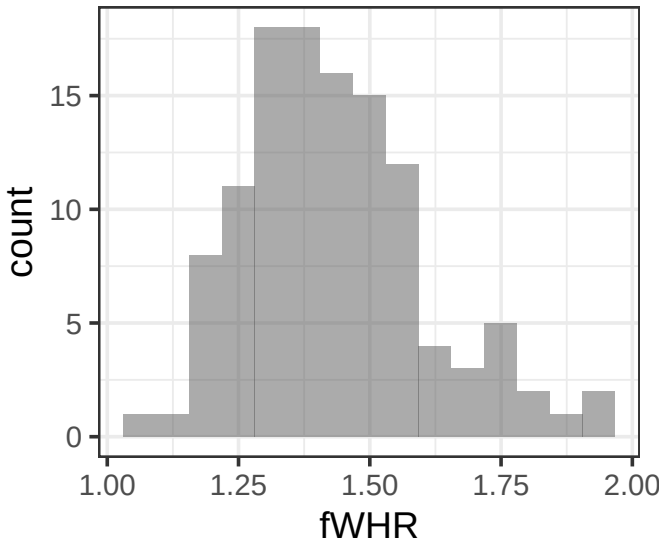
You probably learned something about linear regression in a previous course. Here, we briefly review the main concepts of simple linear regression and quickly expand our tool box to multiple regression (with both quantitative and categorical predictors).

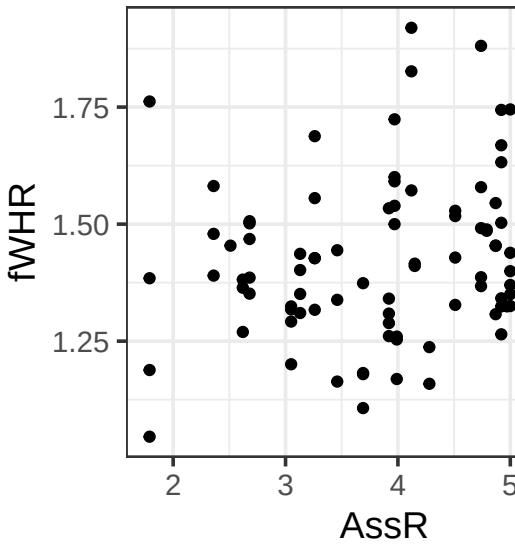
2.1 Data

We will consider a small dataset from an article by J.S. Martin and colleagues, titled *Facial width-to-height ratio is associated with agonistic and affiliative dominance in bonobos (**Pan paniscus**)*

Notes: variable `fWHR` is the facial width-height ratio and `AssR` is the Assertiveness score of affiliative dominance. `normDS` is another dominance score. A few figures of the data are below - we will do some more exploration together.

```
## Observations: 117
## Variables: 8
## $ Name   <fct> Zuani, Zuani, Zorba, Zorba, Zorba, Zomi, Zomi, Zamba, Z...
## $ Group  <fct> Apenheul, Apenheul, Wilhelma, Wilhelma, Wilhelma, Frank...
## $ Sex    <fct> Female, Female, Male, Male, Male, Female, Female, Male,...
## $ Age    <int> 22, 22, 34, 34, 34, 15, 15, 14, 14, 14, 18, 18, 18, 18,...
## $ fWHR   <dbl> 1.475052, 1.321814, 1.581446, 1.479237, 1.390086, 1.340...
## $ AssR   <dbl> 5.36, 5.36, 2.36, 2.36, 2.36, 3.92, 3.92, 4.74, 4.74, 4...
## $ normDS <dbl> 1.430, 1.430, 2.341, 2.341, 2.341, 3.087, 3.087, 3.035,...
## $ weight <dbl> 24.0, 24.0, NA, NA, NA, NA, NA, 41.6, 41.6, 41.6, 38.0,...
```





2.2 Simple linear regression, Residuals & Least squares

First, let's review and consider a simple (one-predictor) linear regression model. Fit the model

```
slr <- lm(fWHR ~ AssR, data=bonobos)
```

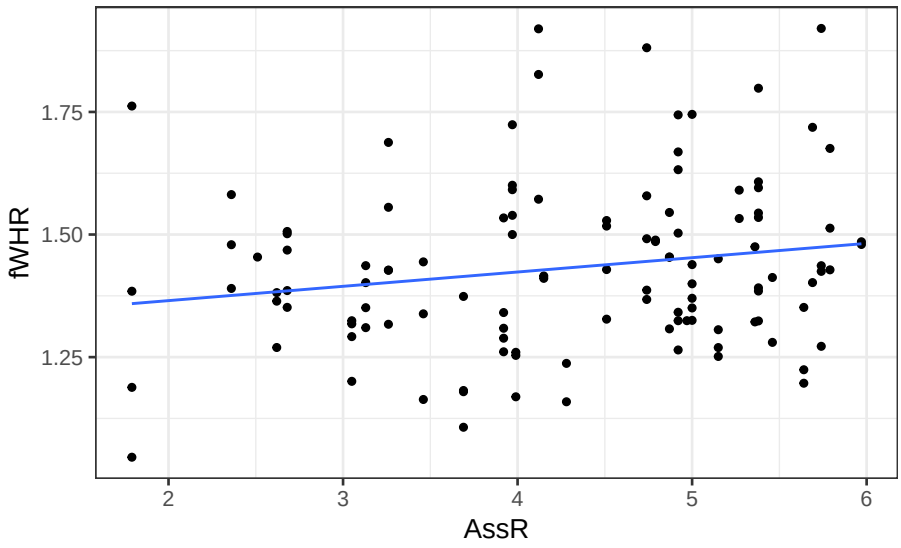
Extract the slope and intercept values:

```
coef(slr)
```

```
## (Intercept)        AssR
##  1.30685287  0.02918242
```

Add the regression line to the plot:

```
gf_point(fWHR ~ AssR, data=bonobos) %>%
  gf_lm()
```




```
summary(slr)
```

```
##
## Call:
## lm(formula = fWHR ~ AssR, data = bonobos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31320 -0.11369 -0.01242  0.09008  0.49241
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.30685    0.06283  20.801   <2e-16 ***
## AssR         0.02918    0.01420   2.055   0.0421 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1689 on 115 degrees of freedom
## Multiple R-squared:  0.03542,    Adjusted R-squared:  0.02704
## F-statistic: 4.223 on 1 and 115 DF,  p-value: 0.04213
```

2.2.1 Using `lm()` to fit a linear regression in R

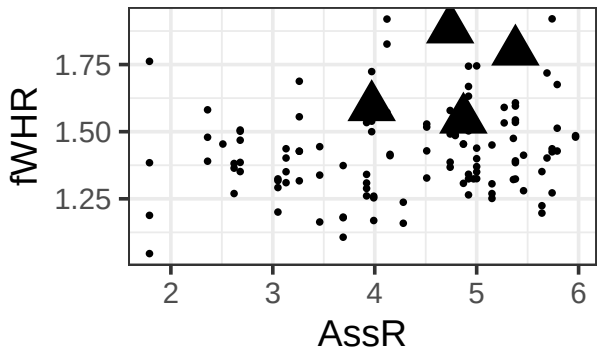
2.2.2 Equation of the fitted regression line

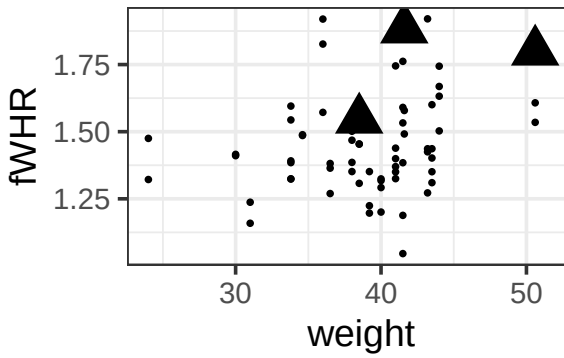
2.3 Multiple regression

Rarely does our response variable **really** depend on only one predictor. Can we improve the model by adding more predictors?

```
mlr <- lm(fWHR ~ AssR + weight, data=bonobos)
coef(mlr)
```

```
## (Intercept)        AssR      weight
## 0.944790930 0.039888045 0.008644299
```





2.3.1 Is it really better?

How do we know if the model with more predictors is "better"? (For a more detailed answer, wait about a week…) But before we can define a "beter" model: how did R find the "best" intercept and slopes?

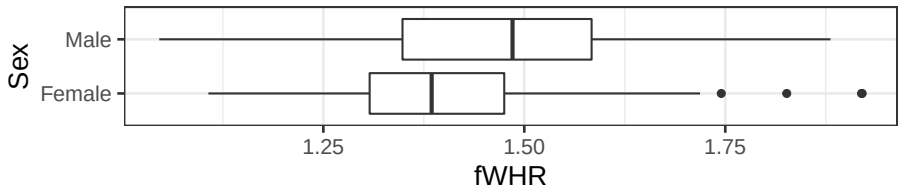
2.3.2 Regression residuals = "errors"

2.3.3 Computing Predictions

Use the regression equation to compute **predicted values** for the three data points below:

```
##        fWHR AssR weight
## 8  1.880866 4.74   41.6
## 25 1.798387 5.38   50.6
## 41 1.591440 3.97     NA
## 65 1.545019 4.87   38.5
```

2.4 Predictors with two categories



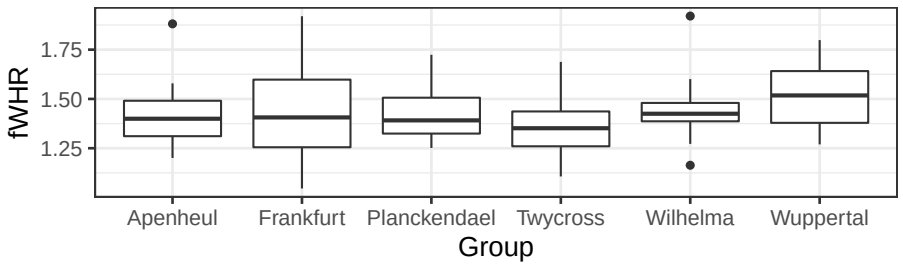

```
mlr2 <- lm(fWHR ~ AssR + weight + Sex, data = bonobos)
coef(mlr2)
```

```
## (Intercept)        AssR      weight     SexMale
## 1.065420976 0.058435841 0.002257142 0.128484275
```

How does the model incorporate this covariate mathematically?

2.4.1 Predictors with more categories

```
gf_boxplot(fWHR ~ Group, data = bonobos)
```



```
mlr3 <- lm(fWHR ~ AssR + weight + Sex + Group, data = bonobos)
coef(mlr3)
```

```
##       (Intercept)              AssR            weight           SexMale
##       1.007734691       0.064361973       0.003458979       0.124854271
##    GroupFrankfurt GroupPlanckendael     GroupTwycross     GroupWilhelma
##       0.037426358      -0.008464572      -0.112907589       0.011186724
##    GroupWuppertal
##      -0.004364826
```

How does the model incorporate **this** covariate mathematically?

Chapter 3

Literature

Here is a review of existing methods.

Chapter 4

Methods

We describe our methods in this chapter.

Chapter 5

Applications

Some *significant* applications are demonstrated in this chapter.

5.1 Example one

5.2 Example two

Chapter 6

Final Words

We have finished a nice book.

Chapter 7

Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 7. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter 4.

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 7.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 7.1.

```
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, 2019) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

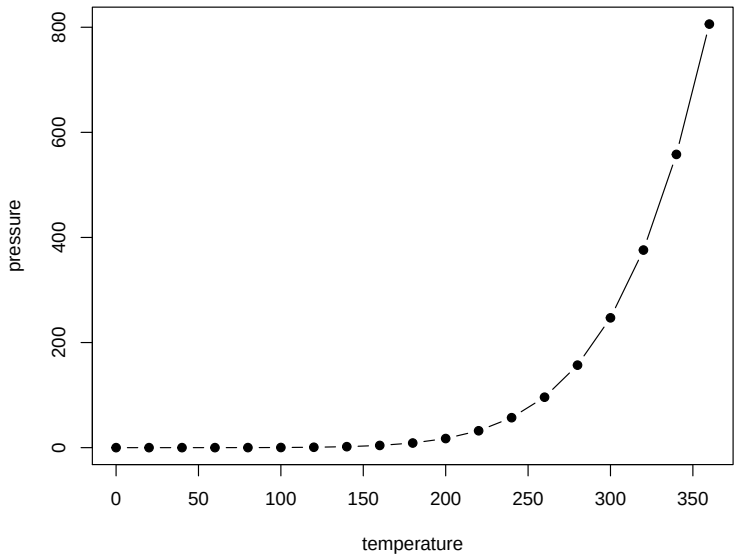


Figure 7.1: Here is a nice figure!

Table 7.1: 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

Bibliography

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

Xie, Y. (2019). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.13.