Lab 9: ANOVA

Kyle Payne

October 21, 2015

# 1.

We wish to compare the mean values for a given variable (treatment) for three populations with five replicates per treatment. The sums of squares for treatment and error are given below.

Use this information to write the linear model.

# Exercise 2. Testing the assumptions of ANOVA

## a.

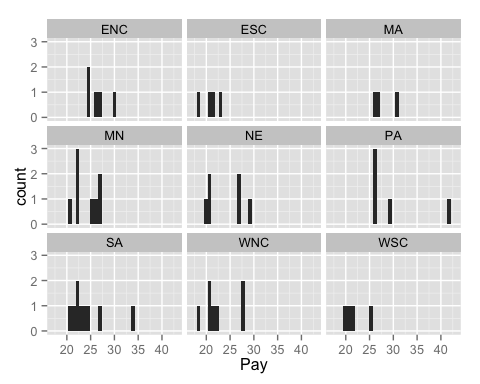
Abstract: Average salary paid to teachers and expenditures per pupil are two commonly used measures of the amount of money spent on education. Data on these two measures are provided by state, and states are classified by region of the country. The two measures have slightly skewed distributions, and Alaska is an outlier in both.

Looking at teachers pay as a dependent variable, then create scatter plots, histograms, and perform tests of normality for teachers pay for each region.

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.1.3

g <- ggplot(aes(x=Pay), data=educate) + geom\_histogram()   
g + facet\_wrap(~Region)



p <- ggplot(educate, aes(factor(Region), Pay)) + geom\_boxplot()

Conduct a test of normality for each of the subgroups using the shapiro wilks test at the level.

library(plyr)

## Warning: package 'plyr' was built under R version 3.1.3

ddply(.data=educate, .(Region), summarize, p\_val = shapiro.test(Pay)$p.value)

## Region p\_val  
## 1 ENC 0.460263456  
## 2 ESC 0.741186563  
## 3 MA 0.505939298  
## 4 MN 0.221714527  
## 5 NE 0.121683298  
## 6 PA 0.006917142  
## 7 SA 0.013304709  
## 8 WNC 0.160042976  
## 9 WSC 0.377781862

Now perform a test of the homogeneity of variance at the level.

bartlett.test(Pay~factor(Region), data=educate)

##   
## Bartlett test of homogeneity of variances  
##   
## data: Pay by factor(Region)  
## Bartlett's K-squared = 10.3129, df = 8, p-value = 0.2437

## b.

Now, perform an ANOVA of Region with Pay as the response and list the ANOVA table.

Based on these results, are any of the assumptions of ANOVA violated?

# 2.

Abstract: People who are concerned about their health may prefer hot dogs that are low in salt and calories. The "Hot dogs" datafile contains data on the sodium and calories contained in each of 54 major hot dog brands. The hot dogs are classified by type: beef, poultry, and meat (mostly pork and beef, but up to 15% poultry meat).

hot\_dogs <- read.delim("~/Desktop/hot\_dogs.txt")

## a.

Test for normality, homogeneity of variance, and produce boxplots and histograms

## b.

Perform an ANOVA with Sodium as the response and Type as the grouping variable. Output an ANOVA table of your results.