R Workshop Module 4: Some Statistical Analyses

Katherine Thompson (katherine.thompson@uky.edu)
Department of Statistics, University of Kentucky
October 14, 2016

Reading in Data

Start by reading the dataset 'practicedata.csv' into R. (Having trouble? See the instructions in Module 1.)

Two-Sample *t*-tests

```
  ## Find group means for response variable
  aggregate((practicedata$respvar)~practicedata$groupvar*practicedata$groupvar2,FUN=mean)
```

```
##   practicedata$groupvar practicedata$groupvar2 (practicedata$respvar)
## 1               Control                      A               39.92996
## 2             Treatment                      A               46.32736
## 3               Control                      B               37.24586
## 4             Treatment                      B               45.49279
```

```
  ## Find group standard deviations for response variable
  aggregate((practicedata$respvar)~practicedata$groupvar*practicedata$groupvar2,FUN=sd)
```

```
##   practicedata$groupvar practicedata$groupvar2 (practicedata$respvar)
## 1               Control                      A               8.523437
## 2             Treatment                      A               7.430197
## 3               Control                      B               7.160007
## 4             Treatment                      B               8.608150
```

```
  ## Perform a t-test
  t.test(respvar~groupvar,data=practicedata,
         var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  respvar by groupvar
## t = -4.6117, df = 98, p-value = 1.207e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.472948  -4.171384
## sample estimates:
##   mean in group Control mean in group Treatment
##                38.58791                45.91007
```

```
  t.test(respvar~groupvar2,data=practicedata,
         var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  respvar by groupvar2
## t = 1.0097, df = 98, p-value = 0.3151
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.698629  5.217300
## sample estimates:
## mean in group A mean in group B
##        43.12866        41.36932
```

Analysis of Variance (ANOVA)

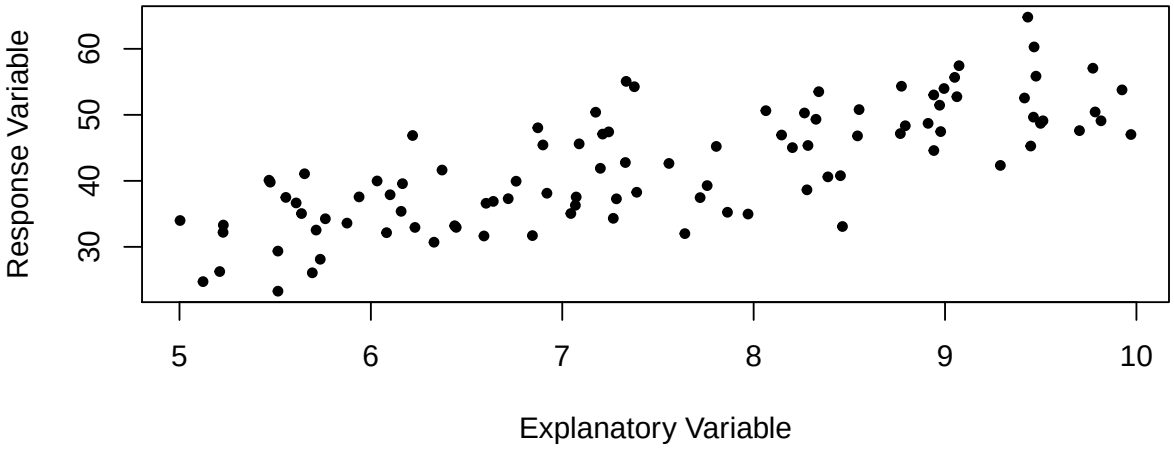
```
  ## Perform an ANOVA
  anova.results=lm(respvar~groupvar+groupvar2+groupvar*groupvar2,data=practicedata)

  anova.results

##
## Call:
## lm(formula = respvar ~ groupvar + groupvar2 + groupvar * groupvar2,
##     data = practicedata)
##
## Coefficients:
##                   (Intercept)             groupvarTreatment
##                        39.930                         6.397
##                    groupvar2B  groupvarTreatment:groupvar2B
##                        -2.684                         1.850

  anova(anova.results)

## Analysis of Variance Table
##
## Response: respvar
##                     Df Sum Sq Mean Sq F value    Pr(>F)
## groupvar             1 1340.4 1340.35 21.1727 1.283e-05 ***
## groupvar2            1   77.4   77.38  1.2223    0.2717
## groupvar:groupvar2   1   21.4   21.38  0.3377    0.5625
## Residuals           96 6077.3   63.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  summary(anova.results)
```

```
##
## Call:
## lm(formula = respvar ~ groupvar + groupvar2 + groupvar * groupvar2,
##     data = practicedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.6440  -5.7905   0.0287   5.7551  19.3054
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    39.930      1.591  25.093  < 2e-16 ***
## groupvarTreatment               6.397      2.250   2.843  0.00546 **
## groupvar2B                     -2.684      2.250  -1.193  0.23593
## groupvarTreatment:groupvar2B    1.850      3.183   0.581  0.56251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.956 on 96 degrees of freedom
## Multiple R-squared:  0.1915,Adjusted R-squared:  0.1662
## F-statistic: 7.578 on 3 and 96 DF,  p-value: 0.0001333
```

Linear Regression

For this example, we will investigate the relationship between the variables `respvar` and `expvar` from the data set, `practicedata`. Remember than by using a '$', we can refer to the variable as `practicedata$respvar` in the following code. To fit a linear model, we use the function `lm()` as follows.

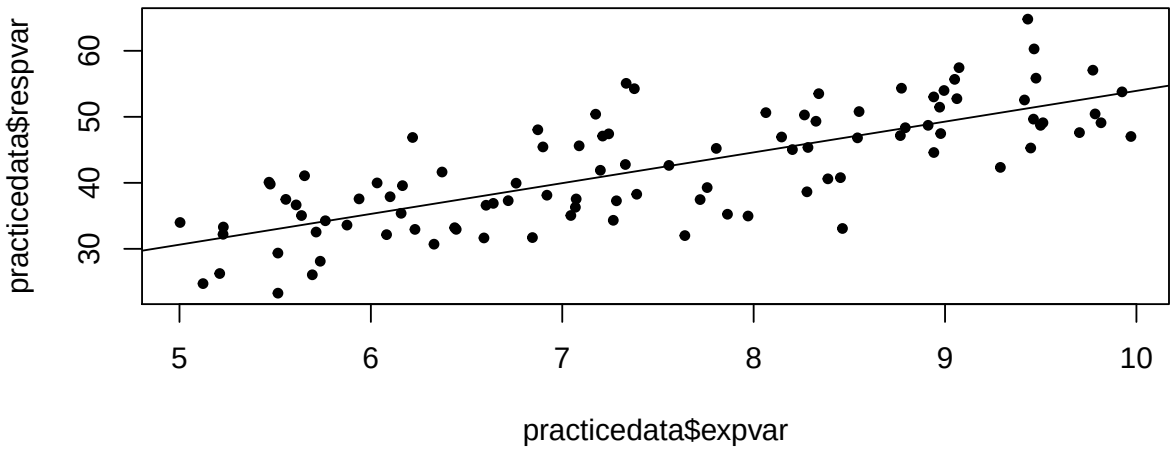
```
## Scatter plot of the data
plot(practicedata$expvar,practicedata$respvar,xlab='Explanatory Variable',
      ylab='Response Variable',pch=20)
```



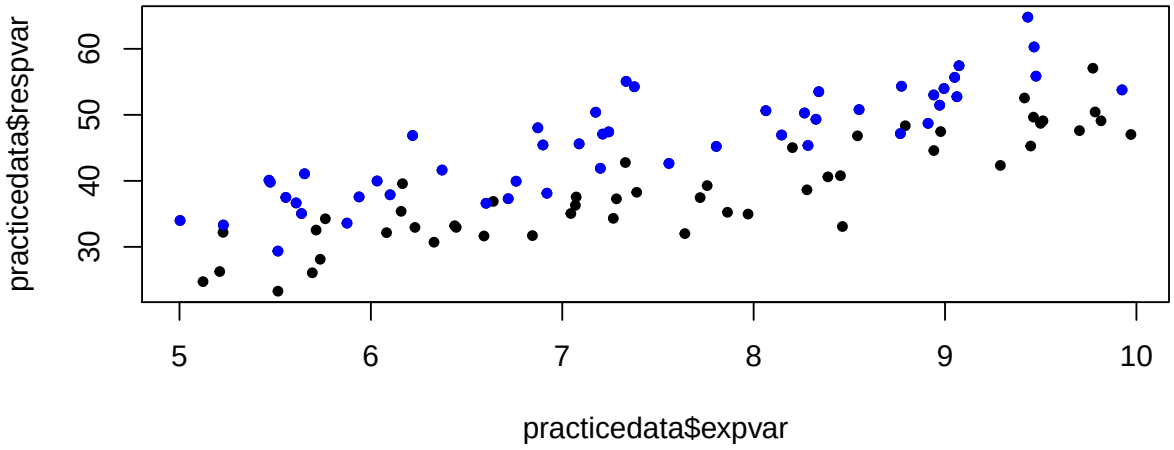

```
## Fit a linear model to the Data
fitted.model = lm(respvar~expvar,
                  data=practicedata) # Fit a linear model with
                  # y-variable respvar and x-variable expvar
summary(fitted.model) #Summarize the linear model

##
## Call:
## lm(formula = respvar ~ expvar, data = practicedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.6918  -4.1475  -0.2382   4.2102  13.5615
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.3299     3.0495   2.404   0.0181 *
## expvar        4.6604     0.3999  11.654   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.67 on 98 degrees of freedom
## Multiple R-squared:  0.5809,Adjusted R-squared:  0.5766
## F-statistic: 135.8 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
## Check to make sure that the model looks appropriate
plot(fitted.model)
```

```r
#Plot the data with the fitted regression line
plot(practicedata$expvar,practicedata$respvar,pch=20)
abline(fitted.model)
```




```r
##Plotting the data and the fitted model
plot(practicedata$expvar,practicedata$respvar,pch=20)
points(practicedata$expvar[practicedata$groupvar=='Treatment'],
       practicedata$respvar[practicedata$groupvar=='Treatment'],
       col='blue',pch=20)
```




```
##You can also fit a linear model with more than one variable
fitted.group.model = lm(respvar~expvar + groupvar,
                        data=practicedata) # Fit a linear model with
                        # y-variable respvar and x-variables expvar and groupvar
summary(fitted.group.model) #Summarize the linear model
```

```
##
## Call:
## lm(formula = respvar ~ expvar + groupvar, data = practicedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.7226 -2.4108 -0.1681  2.1640  9.4101
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.4628     2.1175   0.691    0.491
## expvar              4.8846     0.2694  18.129   <2e-16 ***
## groupvarTreatment   8.3739     0.7640  10.960   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.809 on 97 degrees of freedom
## Multiple R-squared:  0.8128,Adjusted R-squared:  0.8089
## F-statistic: 210.5 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
##You can also fit a linear model with both variables and their interaction
fitted.int.model = lm(respvar~expvar + groupvar + expvar*groupvar,
```

```
                      data=practicedata) # Fit a linear model with
                      # y-variable respvar and
                      # x-variables expvar, groupvar and their interaction
summary(fitted.int.model) #Summarize the linear model

##
## Call:
## lm(formula = respvar ~ expvar + groupvar + expvar * groupvar,
##     data = practicedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6342 -2.4548 -0.0682  2.2381  9.4161
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.2408     2.8744   0.780    0.438
## expvar                     4.7822     0.3714  12.875   <2e-16 ***
## groupvarTreatment          6.7406     4.1313   1.632    0.106
## expvar:groupvarTreatment   0.2182     0.5423   0.402    0.688
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.826 on 96 degrees of freedom
## Multiple R-squared:  0.8131,Adjusted R-squared:  0.8072
## F-statistic: 139.2 on 3 and 96 DF,  p-value: < 2.2e-16
```