# Chapter 5 Activity

217

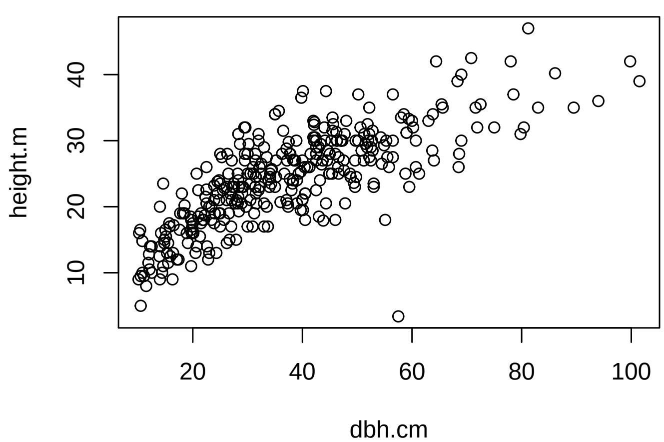
In this activity, we will be looking at two different datasets. The first is from a study at Upper Flat Creek study area in the University of Idaho Experimental Forest. A random sample of 336 trees were selected from the forest, with measurements recorded on Douglas Fir, Grand Fir, Western Red Cedar, and Western Larch trees. We will explore the relationship between tree height and tree diameter.

The second is from mark/recapture data on the Kootenai river in Northwestern Montana. Each year scientists use mark/recapture methods to track the population of various trout species downstream of Libby Dam. We will explore the relationship between annual outflow from the dam (measured in hundreds of thousands of cubic feet) and rainbow trout population.

### The Tree Data

```
require(spuRs)
data(ufc)
```

```
plot(height.m ~ dbh.cm, data = ufc)
```




1. The relationship in the plot above is positive, moderately strong with some curvature, and increasing variability as the diameter increases. There do not appear to be groups in the data set, but since this contains four different types of trees, we would want to look at a scatterplot split by group. Of particular interest is the observation with a diameter around 58 cm and a eight of less than 5 m. This tree may be an outlier. We can look at the correlation with and without it.

```
cor(ufc$dbh.cm, ufc$height.m)

## [1] 0.77

cor(ufc$dbh.cm[-168], ufc$height.m[-168])

## [1] 0.791
```

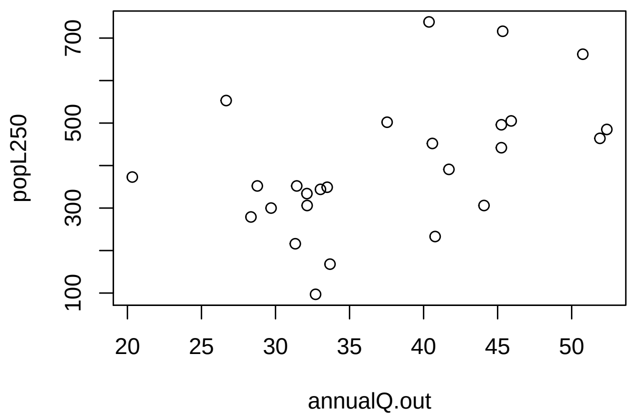
The correlation increases slightly when we exclude this outlier, but the change is relatively minor.

### The Trout Data

```
fish <- subset(recruitAbundLength.fwp, section == "FP")
fish2 <- fish[, c(6, 37)]
fish2$annualQ.out <- fish2$annualQ.out/100
cor(fish2)

##              popL250 annualQ.out
## popL250        1.000       0.488
## annualQ.out    0.488       1.000

plot(popL250 ~ annualQ.out, data = fish2)
```




Describe the relationship you seen the scatterplot of the fish data above.

2. **Describing the Relationship with a Regression Model:**

### The Tree Model

```
tree.lm <- lm(height.m ~ dbh.cm, data = ufc)
summary(tree.lm)

##
## Call:
## lm(formula = height.m ~ dbh.cm, data = ufc)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -27.331  -3.082   0.139   2.825  12.404
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.1077     0.6053    20.0   <2e-16
## dbh.cm        0.3239     0.0147    22.1   <2e-16
##
## Residual standard error: 4.65 on 334 degrees of freedom
## Multiple R-squared:  0.593,Adjusted R-squared:  0.592
## F-statistic:  486 on 1 and 334 DF,  p-value: <2e-16
```

Our model estimates that a one centimeter increase in tree diameter is associated with **an average increase** of 0.324 meters. Our model also estimates that a tree with diameter equal to zero has **an average** height of 12 meters. However, the interpretation of the y-intercept is somewhat nonsensical, as there are no trees with diameter equal to zero.

**The Trout Model**

```
fish.lm <- lm(popL250 ~ annualQ.out, data = fish2)
summary(fish.lm)
```

```
##
## Call:
## lm(formula = popL250 ~ annualQ.out, data = fish2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -259.8  -49.7  -16.7   30.2  311.6
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    59.95     127.34    0.47    0.642
## annualQ.out     9.08       3.31    2.74    0.011
##
## Residual standard error: 140 on 24 degrees of freedom
## Multiple R-squared:  0.238,Adjusted R-squared:  0.206
## F-statistic:  7.5 on 1 and 24 DF,  p-value: 0.0114
```

Give an interpretation of the coefficients in the trout model.

3. **Prediction:**
**The Tree Model:** It is predicted that **on average** a tree with a diameter of 24 cm has height of 23.8 m.

**The Trout Model:** Give the predicted rainbow trout population for a year in which 3,000,000 cubic feet of water were release from the dam (remember how the outflow is measured in the data).

4. **Finding and Interpreting $R^2$:**
**Tree Model**
In the R output below, we can see that the $R^2$ for our tree model is .5928. That means that 59.28% of the variability in tree heights is explained by tree diameters in this population of Idaho State University experimental forest trees.

```
tree.lm <- lm(height.m ~ dbh.cm, data = ufc)
summary(tree.lm)

##
## Call:
## lm(formula = height.m ~ dbh.cm, data = ufc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -27.331  -3.082   0.139   2.825  12.404
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.1077     0.6053    20.0   <2e-16
## dbh.cm        0.3239     0.0147    22.1   <2e-16
##
## Residual standard error: 4.65 on 334 degrees of freedom
## Multiple R-squared:  0.593,Adjusted R-squared:  0.592
## F-statistic:  486 on 1 and 334 DF,  p-value: <2e-16
```

**Trout Model**
Give the $R^2$ for the trout model and interpret it in the context of the problem.

```
fish.lm <- lm(popL250 ~ annualQ.out, data = fish2)
summary(fish.lm)

##
## Call:
## lm(formula = popL250 ~ annualQ.out, data = fish2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -259.8  -49.7  -16.7   30.2  311.6
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    59.95     127.34    0.47    0.642
## annualQ.out     9.08       3.31    2.74    0.011
##
## Residual standard error: 140 on 24 degrees of freedom
## Multiple R-squared:  0.238,Adjusted R-squared:  0.206
## F-statistic:  7.5 on 1 and 24 DF,  p-value: 0.0114
```

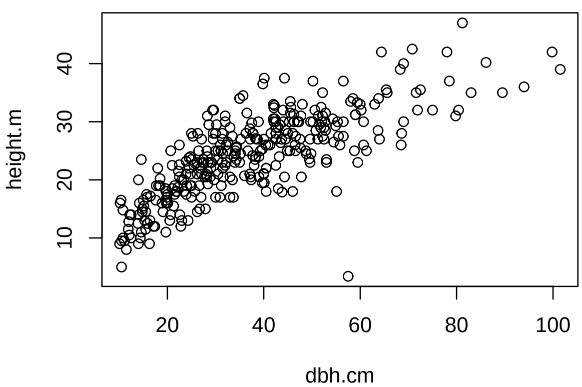
5. **Assessing Assumptions:**

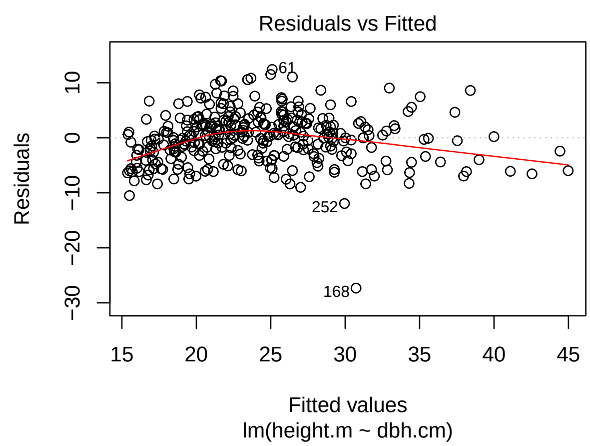
**Tree Model**

(a) Quantitative variables: Both tree height and tree diameter are quantitative.

(b) Independence: The documentation of these data state that the sample was random. Therefore, our observations must be independent.

(c) Linearity of relationship:

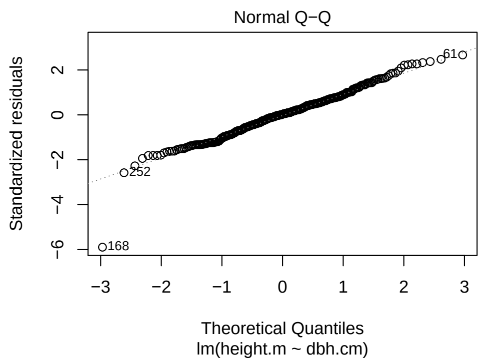


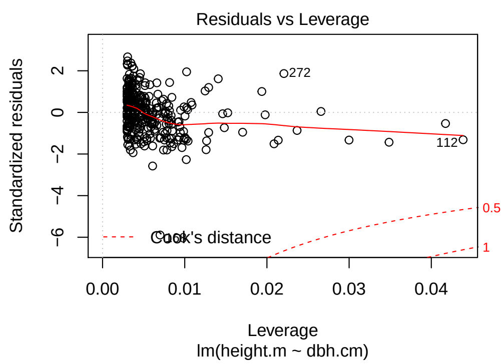

We can see from the scatterplot that the relationship between tree height and tree diameter appears reasonably linear. From the Residuals vs. Fitted plot we observe a slight curvature. If we were looking adding terms to our model to account for curvature, I would want to explore this. But right now you should just make note of it and move on.

(d) Constant variance: We can see from the Residuals vs. Fitted plot (above) that the spread is fairly constant the range of fitted values. It may appear as thought larger fitted values have somewhat less spread, but I think this is more due to there being fewer observations in that range.

(e) Normality of residuals: The Normal Q-Q plot looks just fine. I see nothing to be concerned about.

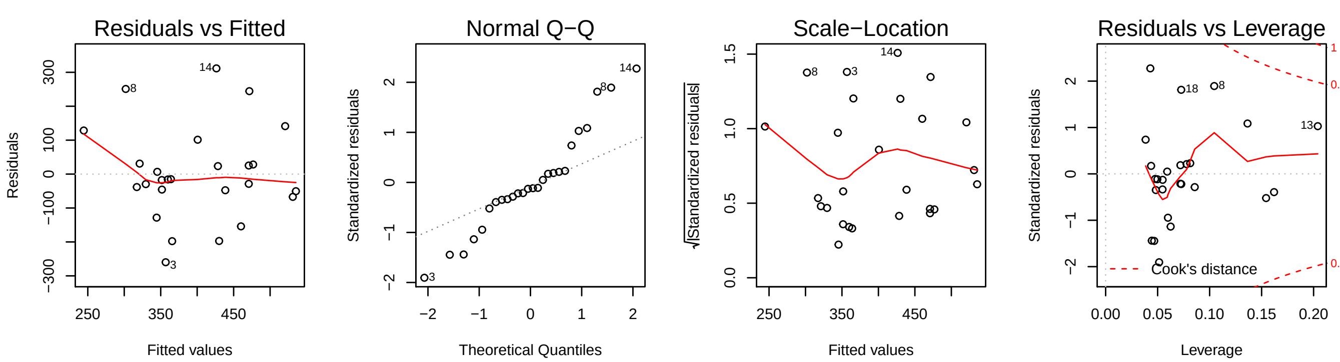



(f) No influential points: We can see from the Scale-Location plot that these data contain no influential points.




## Trout Model

```r
par(mfrow = c(1, 4))
plot(fish.lm)
```




(a) Quantitative variables condition:

(b) Independence of observations:

(c) Linearity of relationship:

(d) Equal (constant) variance:

(e) Normality of residuals:

(f) No influential points: