R notes for Lecture 8

For all of these examples we use the tree data.

```
> tree.table <- read.table("tree.txt", header=T, sep="\t")
> names(tree.table)
[1] "Diameter" "Height"   "Volume"
> attach(tree.table)
```

1. Fit a regression

```
> plot(Height, Volume)
> tree.ht <- lm(Volume ~ Height)
> abline(tree.ht)
```

2. Evaluate residuals and find Cook's distances

```
> resid.ht <- residuals(tree.ht)  #saves residuals
> hist(resid.ht)    # crude check for normality, outliers
> qqnorm(resid.ht)  # better check for normality
> qqline(resid.ht)
> plot(Height, resid.ht) # check for deviations from
linearity, non-constant variance
> plot(tree.ht)
Hit <Return> to see next plot:
Hit <Return> to see next plot:
Hit <Return> to see next plot:
Hit <Return> to see next plot:
```

#2 out of 4 of these plots will show you residual diagnostic plots. Plot #4 gives Cook's distances.

3. Check the coefficients of the model

```
> coefficients(tree.ht)
(Intercept)      Height
 -87.123614    1.543350
```

4. T-tests that the slope is 0 or intercept is 0, find r-squared, find standard errors of estimated intercept and slope

```
> summary(tree.ht)

Call:
lm(formula = Volume ~ Height)

Residuals:
    Min      1Q  Median      3Q     Max
-21.274  -9.894  -2.894  12.067  29.852

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -87.1236    29.2731  -2.976 0.005835 **
Height        1.5433     0.3839   4.021 0.000378 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 `
' 1

Residual standard error: 13.4 on 29 degrees of freedom
Multiple R-Squared: 0.3579,   Adjusted R-squared: 0.3358
F-statistic: 16.16 on 1 and 29 DF,  p-value: 0.0003784
```

5. Estimate mean volume for all trees with height 72, 78, 79.

```
> new <- data.frame(Height=c(72,78,79)) # You must first
create a data frame with the new values in it.
> predict.lm(tree.ht, new, interval="confidence")
       fit      lwr      upr
1 23.99757 18.15976 29.83538
2 33.25767 28.09207 38.42327
3 34.80102 29.34525 40.25678
>
```

First column are the predictions: So the mean volume of trees 72 feet tall is 23.998. The next two columns give the lower and upper bounds on a 95% confidence interval. So the confidence interval for trees 72 feet tall is (18.16, 29.8).

6. Predict the volume of a tree that is 72 feet tall. Also one that is 78, and one that is 79.

```
> predict.lm(tree.ht, new, interval="prediction")
       fit       lwr      upr
1 23.99757 -4.017335 52.01247
```

```
2 33.25767  5.375087 61.14025
3 34.80102  6.863226 62.73881
```

The only change in the command was that interval is set to "prediction" instead of "confidence". Note that the first column is the same, but the confidence limits have changed. And the first set of limits show some problems with our model (we are predicting trees with negative volume).