Question 4.4

a. The results of `summary()` is shown in Table 1.

Table 1. Summary of simple linear regression results of mpg on horsepower.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 46.92659    1.92184   24.42  < 2e-16
hp          -0.21762    0.01829  -11.90 1.03e-14
---
Residual standard error: 3.096 on 40 degrees of freedom
Multiple R-squared: 0.7796,Adjusted R-squared: 0.7741
F-statistic: 141.5 on 1 and 40 DF,  p-value: 1.027e-14
```

- The "Intercept" "Estimate" (46.93) is the estimated mean mpg for a car with no horsepower.
- The "hp" "Estimate" (i.e., the slope) says that for every increase in one horsepower the mpg will decrease by -0.22, on average.
- The "Intercept" p-value ($p < 0.00005$) shows that the mean mpg for a car with no horsepower is different than zero (a nonsensical significance test).
- The "hp" p-value ($p < 0.00005$) shows that the slope is significantly different from zero indicating that there is a significant relationship between the horsepower and mpg of a car.
- The "residual standard error" (3.096) is a measure of the standard deviation **about the line**, or an indication of how far an "average" individual deviates from the best-fit line.
- The "residual degrees of freedom" (40) is two less then the number of observations.
- The "multipe r-squared" (0.78) shows the proportion of the total variability in mpg (ignoring horsepower) that is explained away by knowing a horespower value.
- The F (141.5) and corresponding p-value ($p < 0.00005$) show that the full model including the slope is significantly "better" than the simple model with no slope. Thus, a slope "is needed" and it can be concluded that there is a significant relationship between a car's horespower and its gas mileage.

b. The results of `confint()` is shown in Table 2.

Table 2. Coefficient confidence intervals for simple linear regression results of mpg on horsepower.

```
                2.5 %    97.5 %
(Intercept) 43.0424051 50.810780
hp          -0.2545932 -0.180651
```

- The "intercept" CI says that the mean mpg for a car with no horsepower is between 43.0 and 50.8.
- The "hp" CI says that the mpg will *decrease* between 0.18 and 0.25 for a one unit increase in horespower.

c. The results of `anova()` is shown in Table 3.

Table 3. ANOVA table for simple linear regression results of mpg on horsepower.

```
           Df  Sum Sq Mean Sq F value    Pr(>F)
hp          1 1356.83 1356.83  141.53 1.027e-14
Residuals 40  383.48    9.59
```

- The "Residuals", or "error", df is two less than the number of observations.
- The "hp", or "regression", df is one less than the number of parameters estimated (two – intercept and slope).
- The "Residual", or "error", MS is the total variance in mpg after considering horsepower or the variability of individuals around the best-fit line (i.e., the full model).

- The "hp", or "regression", MS is the variance in mpg that can be explained by knowing the value of horespower.
- The F test statistic is the ratio of variability in mpg explained by knowing the value of horespower to the variability unexplained even after knowing the value of the horsespower.

d. There is a significant relationship as indicated by the very small slope and F-test p-values ($p < 0.00005$).

Question 4.5

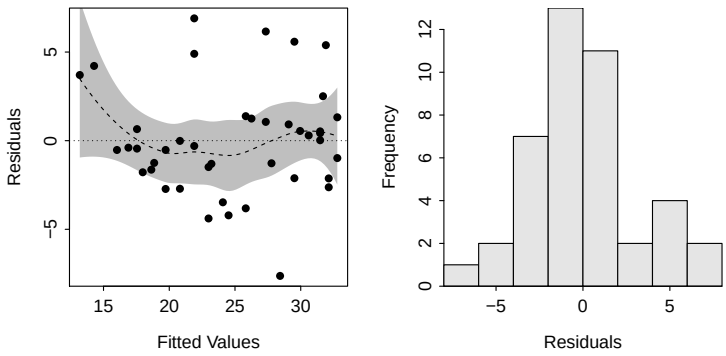

Figure 1. Residual plot (Left) and histogram of residuals (Right) for the simple linear regression of car mpg on car horespower.

a. The cars appear to be independent as no car was used twice and they were all of different makes and models.
b. No non-linearity was detected as the residual plot does not exhibit a distinct curvature (Figure 1-Left).
c. The residual plot does show a slight heteroscedasticity (Figure 1-Left).
d. The Anderson-Darling test shows weak but insignificant evidence for non-normality ($p = 0.0784$) and the histogram of residuals is approximately symmetric (Figure 1-Right).
e. There are no significant outliers according to the outlier test ($p = 0.4334$)

R Commands

```
> car <- read.csv("https://raw.githubusercontent.com/droglenc/NCData/master/CarMPG.csv")
> lm1 <- lm(mpg~hp,data=car)
> summary(lm1)
> confint(lm1)
> anova(lm1)
```

```
> residPlot(lm1,main="")
> adTest(lm1$residuals)
> hist(lm1$residuals,xlab="Residuals",main="")
> outlierTest(lm1)
```

Notes from the Professor

-