Test 2 - part 2

STAT 021

Swarthmore College

Short answer questions

Problem 5

The R-output below contains the results from a regression model fit to a data set concerning life expectancy in years, alcohol consumption in drinks per day, and smoking status (smoker - Yes vs. non-smoker - No). Answer the following questions pertaining to the R output below.

```
## Call:
## lm(formula = Lifespan ~ Alcohol + Smoker)

## Residuals:
## Min 1Q Median 3Q Max
## -30.796 -7.139 0.125 6.949 19.578

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 93.6818   2.1024      44.56   <2e-16 ***
## Alcohol     -3.2656   0.3147     -10.38   <2e-16 ***
## SmokerYes  -23.4392   1.9922     -11.77   <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 9.959 on 97 degrees of freedom
## Multiple R-squared: 0.7129, Adjusted R-squared: 0.7069
## F-statistic: 120.4 on 2 and 97 DF, p-value: < 2.2e-16
```

a) How does drinking alcohol affect average life expectancy? (2 points)

b) Suppose someone consumes about 2.5 alcoholic drinks per day and smokes regularly. Calculate (provide the formula for) the average life expectancy for this individual. (1 point)

c) What is the average difference in lifespan between smokers and non-smokers? (1 point)

d) Supposing we fix the amount of alcohol consumption; is there a statistically significant

relationship between life expectancy and smoking status? Justify your answer. (2 points)

For Problem 6 and 7 we are going to examine the impact of Vitamin C from two different sources and at two different dosages on the growth of teeth in a random sample of Guinea pigs. The variables in the data set are:

- *len* - a numeric variable representing the length of growth in the teeth (in cm)

- *supp.VC* - a binary categorical variable that is 1 if the Vitamin C supplement type is ascorbic acid and is 0 if the Vitamin C supplement type is orange juice.

- *dose.2.0* - is a binary categorical variable that is 1 if the amount of Vitamin C is 2.0 mg and is 0 if the amount of Vitamin C is 1.0 mg.

Problem 6

```
teeth_SLR_main<-lm(len ~ supp + dose)
summary(teeth_SLR_main)

## Call:
## lm(formula = len ~ supp + dose)

## Residuals:
##    Min     1Q Median     3Q    Max
## -6.697 -2.765 -1.005  2.178  9.262

## Coefficients:
##                 Estimate Std. Error t value   Pr(>|t|)
## (Intercept)       21.197      1.061  19.976    < 2e-16 ***
## supp.aa           -2.925      1.225  -2.387     0.0222 *
## dose.2.0           6.365      1.225   5.195   7.72e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 3.875 on 37 degrees of freedom
## Multiple R-squared:  0.469,  Adjusted R-squared:  0.4403
## F-statistic: 16.34 on 2 and 37 DF,  p-value: 8.202e-06
```

a) Based on the R output above, write the down the estimated main effects equations for predicting the average tooth growth for Guinea pics who (2 points):

   1) Have supplement type absorbic acid and dosage of 1.0 mg
   2) Have supplement type absorbic acid and dosage of 2.0 mg
   3) Have supplement type orange juice and dosage of 1.0 mg
   4) Have supplement type orange juice and dosage of 2.0 mg

b) explain the relationship between supplement type and tooth growth in language that can be understood by an average high school student. (2 points)

Problem 7

```
teeth_SLR_int<-lm(len ~ supp + dose + supp*dose)
summary(teeth_SLR_int)

## Call:
## lm(formula = len ~ supp * dose)

## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.200 -2.337 -0.005   2.147   7.760

## Coefficients:
##                     Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)           22.700      1.137    19.969    < 2e-16 ***
## supp.aa               -5.930      1.608    -3.689    0.00074 ***
## dose.2.0               3.360      1.608     2.090    0.04374 *
## supp.aa:dose.2.0       6.010      2.274     2.643    0.01208 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 3.595 on 36 degrees of freedom
## Multiple R-squared:  0.5553, Adjusted R-squared:  0.5183
## F-statistic: 14.99 on 3 and 36 DF,  p-value: 1.717e-06
```

a) Based on the R output above, write the down the estimated interaction effects MLR for predicting the average tooth growth for Guinea pics who (2 points):

1) Have supplement type absorbic acid and dosage of 1.0 mg
2) Have supplement type absorbic acid and dosage of 2.0 mg
3) Have supplement type orange juice and dosage of 1.0 mg
4) Have supplement type orange juice and dosage of 2.0 mg

b) explain the relationship between supplement type and tooth growth in language that can be understood by an average high school student. (2 points)

c) Based on the R output above and the residual plots on the next page, which model do you think is a better choice, the one **with** interactions or the one **without** interactions? Justify your answer. (2 points)

Residual plot for main effects mode

Residual plot for interaction effects
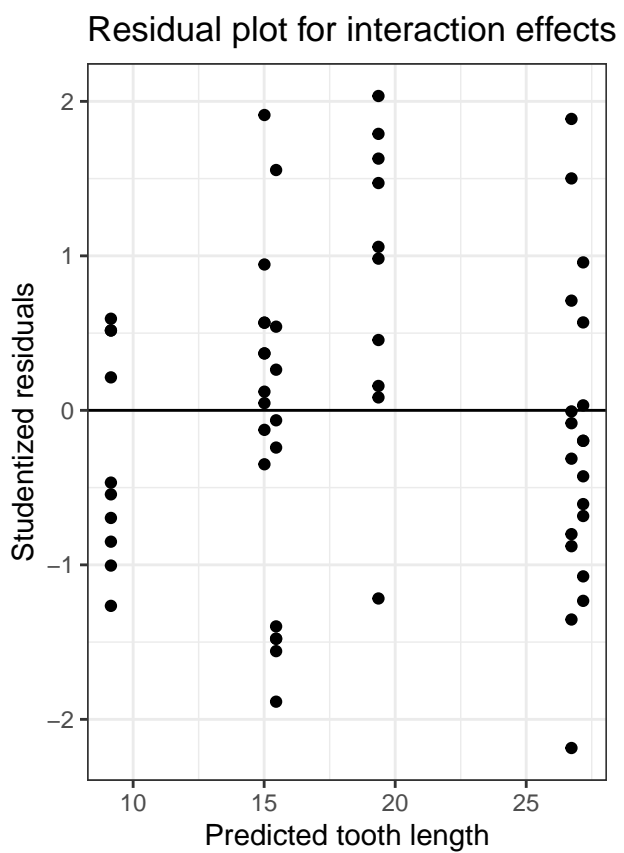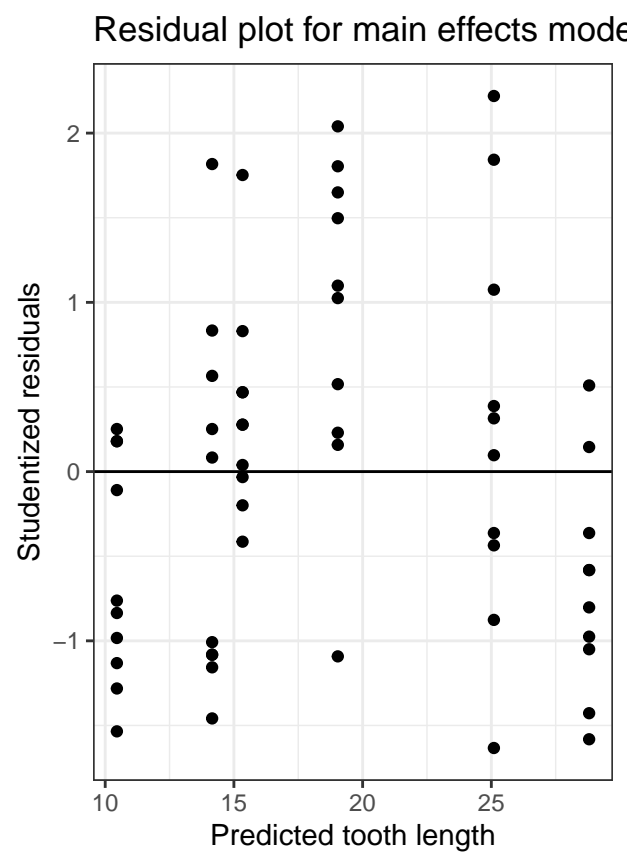
For Problem 8 consider a MLR model for vehicle fuel consumption, measured in miles per gallon.

Other attributes of cars like the weight, horsepower, number cylinders, and acceleration time were also recorded for each car.

We will study the relationship between mpg and weight (in lbs).

```
##
## Call:
## lm(formula = mpg ~ weight + transmission_type + weight:transmission_type,
##     data = car_dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4534 -1.8453  0.3717  1.4173  4.9229
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              29.4530696  3.2887177   8.956 1.04e-09 ***
## weight                   -0.0030367  0.0008114  -3.743 0.000834 ***
## transmission_typeM       28.6553504  6.2299643   4.600 8.28e-05 ***
## weight:transmission_typeM -0.0094807  0.0022902  -4.140 0.000289 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.689 on 28 degrees of freedom
## Multiple R-squared:  0.8365, Adjusted R-squared:  0.8189
## F-statistic: 47.73 on 3 and 28 DF,  p-value: 3.908e-11

##
## Call:
## lm(formula = log_mpg ~ weight + transmission_type + weight:transmission_type,
##     data = car_dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37955 -0.09260  0.02747  0.08665  0.27744
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               3.548e+00  1.710e-01  20.748  < 2e-16 ***
## weight                   -1.786e-04  4.219e-05  -4.233 0.000224 ***
## transmission_typeM        8.895e-01  3.240e-01   2.746 0.010429 *
## weight:transmission_typeM -2.908e-04  1.191e-04  -2.442 0.021194 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
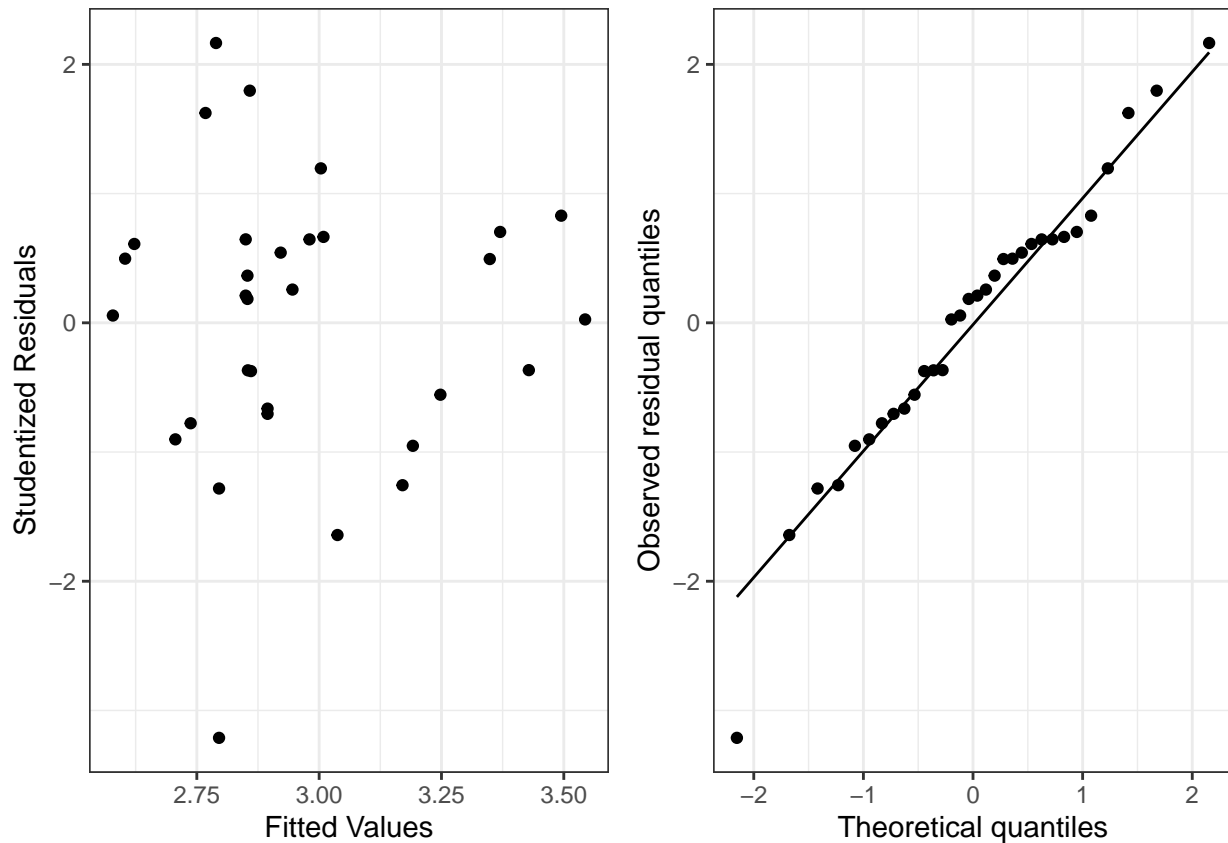
```
##
## Residual standard error: 0.1398 on 28 degrees of freedom
## Multiple R-squared:  0.7906, Adjusted R-squared:  0.7681
## F-statistic: 35.23 on 3 and 28 DF,  p-value: 1.215e-09
```
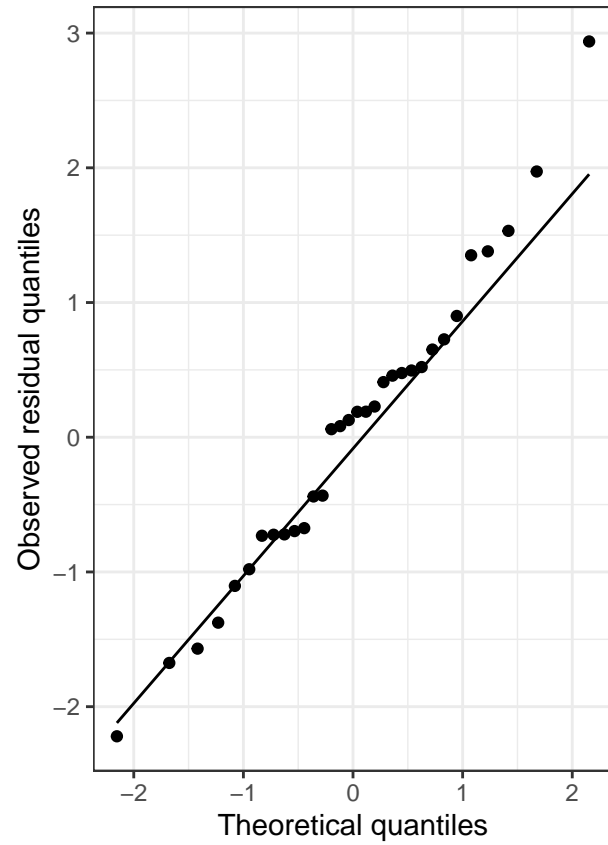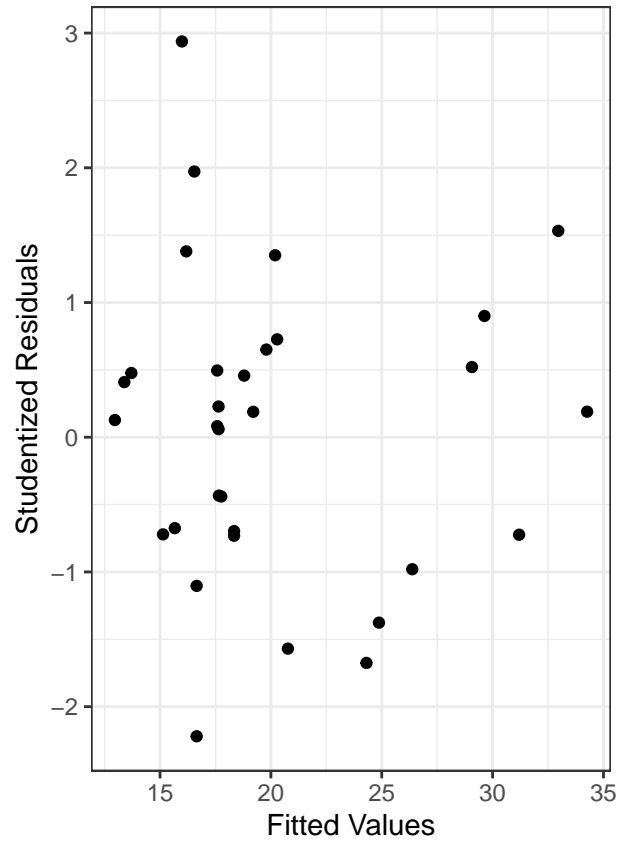
Problem 8

(6pts) Someone argues that a linear regression model would be a good fit. Make a residual plot of ei versus the fitted values.

What does the plot suggest about the linearity assumption of the regression model?

(b) (8pts) Apply the log transformation on weight, refit the linear regression model, and produce a new residual plot. What does the plot suggest about the linearity assumption now? What else does it suggest? What assumptions are violated?

(c) (6pts) Prepare a qq-plot of the residuals in (b). Do the residuals appear to be Normally distributed?

(5pts) How do you interpret the estimated intercept and slope?

(b) (10pts) Test whether or not there is a linear association between mpg (after the transformation you selected, if any) and weight (after the transformation you selected, if any), using alpha= 0.05. State the alternative hypothesis, decision rule, and conclusion. What is the p-value of the test? (c) (5pts) Find a 90% confidence interval for beta1. How do you interpret it?