NOTE: These questions are representative of what will be on the exam.

Questions 1-7. Forty-six mountains in the Adirondacks of upstate New York are known as the High Peaks with elevations near or above 4000 feet. Below is some R output from a linear regression model of $Y = Time$ (expected trip time to hike the peak, in hours) on $X = Ascent$ (in feet).

```
> summary(time.lm)
Call:
lm(formula = Time ~ Ascent, data = HighPeaks)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.2100541  1.8661683   2.256  0.02909 *
Ascent      0.0020805  0.0005909   3.521  0.00101 **

Residual standard error: 2.496 on 44 degrees of freedom
Multiple R-squared: 0.2198,     Adjusted R-squared: 0.2021
F-statistic: 12.4 on 1 and 44 DF, p-value: 0.001014
```

1. Interpret the slope with appropriate units.

   A: The slope is 0.00208 hours per foot, which means that for every foot of ascent, the expected trip time increases by 0.00208 hours. Put another way, for every 100 feet of ascent, the expected trip time increases by 0.208 hours, or about 12 minutes.

2. The 95% confidence interval for β_1, the coefficient of *Ascent*, is (0.00089, 0.00327). Interpret this interval using an increment of 1000 feet of Ascent.

   A: We are 95% confident that the expected trip time increases between 0.89 and 3.27 hours per 1000 feet of ascent.

3. Interpret the residual standard error.

   A: The residual standard error is 2.496 hours. This value measures the variation in the residuals of the fitted model. It can also be interpreted as an approximate standard deviation of a predicted mean response (expected trip time).

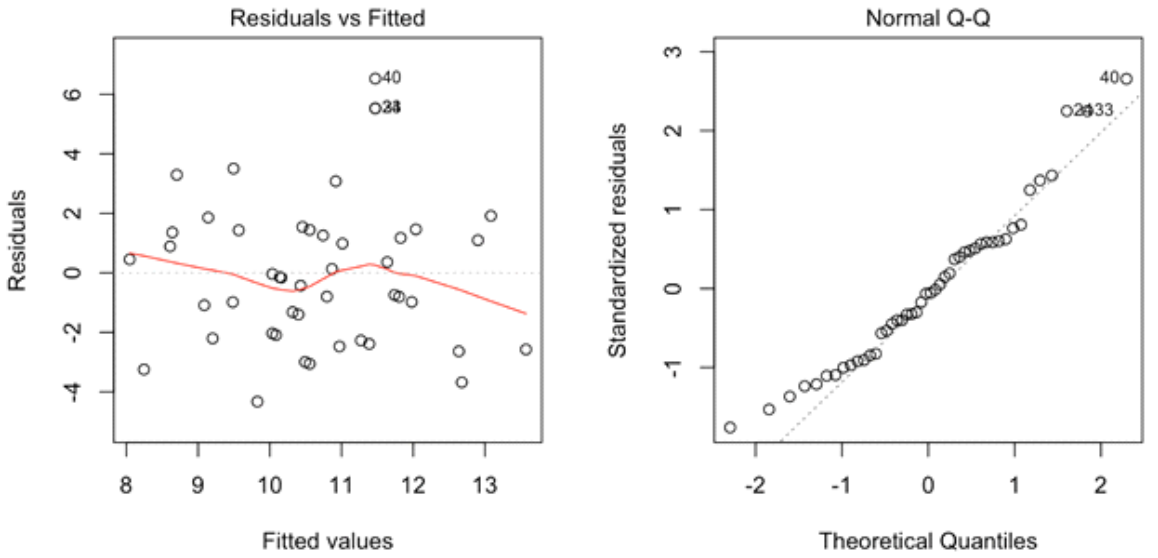
4. Interpret the coefficient of determination.

   A: The coefficient of determination is 0.2198; thus, ascent explains nearly 22% of the variation in expected trip time.

5. Using this model, predict the hiking time for a mountain with an ascent of 4000 feet.

   A: The predicted hike time for an ascent of 4000 feet is 12.5 hours.

6. Interpret the residual plots from this fit:



A: The first residual plot does not indicate any obvious concerns with non-linearity or non-constant residual variance. The are two possible outliers corresponding to rows 40 and 33. In fact, there appear to be two points plotted where the label 33 is. The second residual plot does not indicate any obvious differences from a normal distribution.

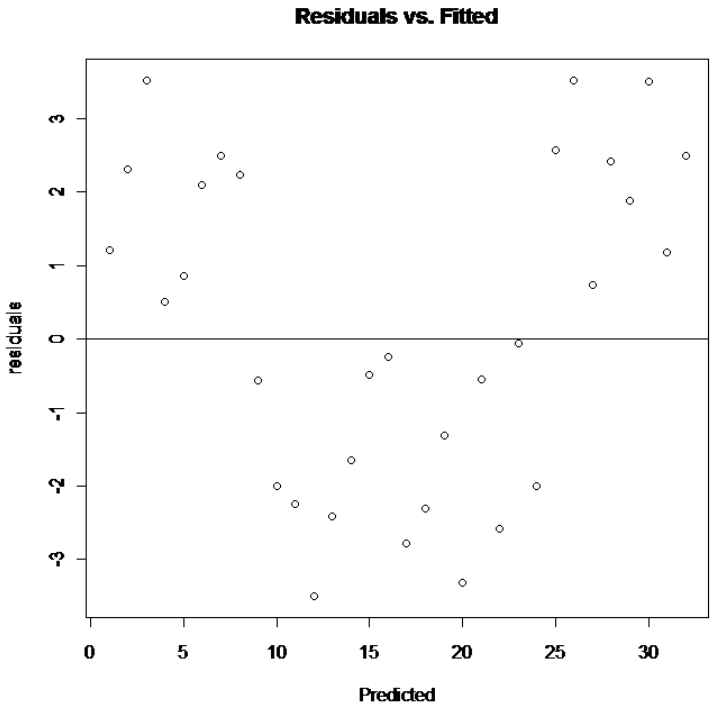
7. Using the R output below, report and interpret a 95% confidence interval for the mean trip time when *Ascent* is 3000 feet.

```
> predict.lm(time.lm, list(Ascent=3000),interval="confidence")
       fit      lwr      upr
1 10.45163 9.701043 11.20222
> predict.lm(time.lm, list(Ascent=3000),interval="prediction")
       fit      lwr      upr
1 10.45163 5.365099 15.53816
```

A: The expected mean trip time for all trips where the ascent is 3000 feet is 10.45 hours. We are 95% confident that the population mean expected time is between 9.70 and 11.20 hours.

Note: this is the end of the questions based on the Adirondack data.

8. A residuals vs. fitted value plot for a regression model is shown below. Based only on the information in this plot, do you feel the condition of linearity is *reasonable*, *problematic*, or you *can't judge* (from the plot shown)?



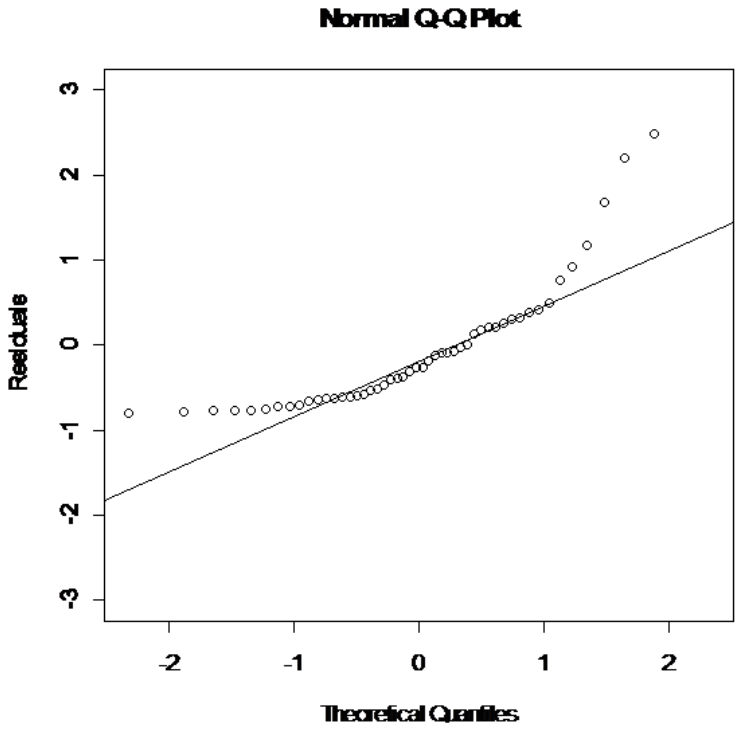
9. Using the same plot (above), do you feel the condition of equal variance is *reasonable*, *problematic*, or you *can't judge*?

A: Reasonable

10. Using the same plot (above), do you feel the condition of normality is *reasonable*, *problematic*, or you *can't judge*?

A: Can't judge

11. A normal quantile plot for a regression model is shown below. Based only on the information in this plot, do you feel the condition of linearity is *reasonable*, *problematic*, or you *can't judge*?



A: Can't judge (the normal quantile plot only checks for the residuals having a normal distribution).

12. Using the same plot (above), do you feel the condition of normality is *reasonable*, *problematic*, or you *can't judge*?

A: Problematic

Other topics to review:

1. Outliers, leverage, and influential points
2. ANOVA table calculations