Name:
Quiz 3, Stat 431 (Summer 2012)
July 23, 2012

You have 20-25 minutes to answer these questions.

1. Suppose we have (X_i, Y_i) for $i = 1, ..., n$ and we want to study the relationship between X_i and Y_i. We fit the regression and obtain $\hat{\beta}_0$ and $\hat{\beta}_1$

   (a) Suppose we decide to reformat X_i and Y_is and obtain $X_{i,new}$ and $Y_{i,new}$:

   $$X_{i,new} = X_i - \bar{X} \quad , \quad Y_{i,new} = Y_i - \bar{Y}$$

   What are the estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ under the reformatted data $(X_{i,new}, Y_{i,new})$? Please provide a short, *mathematical* justification.

   (b) How do the estimate in (a) compare to those obtained from the unformatted data(X_i, Y_i)?

2. Your instructor gets easily depressed, especially after not being able to solve a problem for six months. To remedy this depression, he has resorted to buying wine with lots and lots of alcohol. Today, you're going to help him find the wine with the highest alcohol concentration by looking at the wine's magnesium content.

   Your instructor wanted to know whethere there is any relationship between magnesium content (in grams) and alcohol by volume (in 100% percents). He wants to use this information to predict alcohol content based solely on magnesium content of the wine. He runs a regression in R and obtains the following output.

```
Call:
lm(formula = alcohol ~ magnesium)

Residuals:
     Min       1Q   Median       3Q      Max
-1.74371 -0.52989 -0.00989  0.59396  1.88394
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.465359   0.415551   ?      < 2e-16 ***
magnesium    ?          0.004124   3.732      ?   ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.7837 on ?   degrees of freedom
Multiple R-squared: 0.07333,    Adjusted R-squared: 0.06807
F-statistic: ?     on 1 and 176 DF,  p-value: 0.0002562
```

Sadly, he was slightly "tipsy" while he was doing this and "unintentionally" lost a couple of values from the R output.

(a) Fill out the following ANOVA table based on the above output.

SS	MSE	DF
SSE =	MSE =	DFE =
SSR =	MSR =	DFR =
SST =	MST =	DFT =

Table 1: ANOVA table for the regression in question 2

(b) Suppose I want to test whether there is a positive relationship between magnesium content and alcohol content. Set up the hypothesis, propose a test statistic, and *actually compute* the p-value of the test *from the table*. Do you reject the null at $\alpha = 0.01$?
Hint: Think about the tails of the sampling distribution and the difference between one-sided and two-sided tests when it comes to computing the tail areas.

(c) If a bottle of wine contains 109 grams of magnesium, what is its predicted alcohol content?

(d) Which one is the 95% prediction interval and which one is the95% confidence interval for (c)?

i. $[10.87344, 11.2342]$
ii. $[13.00486, 13.2814]$
iii. $[11.59028, 14.69598]$
iv. $[14.12344, 16.23039]$

(e) From the plot below, should I worry about homoscedasticity? How about linearity and normality of the errors? Are there any regression outliers, leverage points, or influential points. On the plot, identify, if any, regression outliers, leverage points, or influential points by circling the points.