Quiz 3

Name ______________________________

Problem 1.

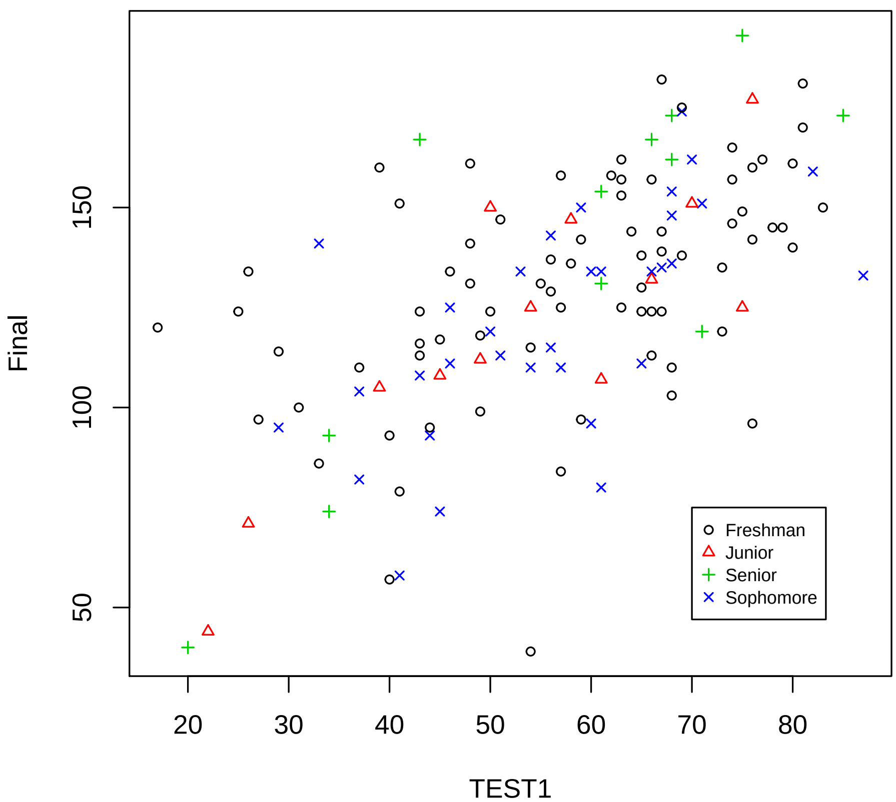

Figure 1: Final scores versus Test1.

A plot of final exam score versus the first test (out of four) is shown in Figure 1. There were 132 students in this class. Class-level is indicated for each student.

First, a simple linear regression of Final against Test1 yields a residual sum-of-squares of 77396.53. Second, R is used to fit a linear model using the function `lm` as follows:

```
> fit2 = lm(Final~TEST1 + Level + TEST1:Level, data=grades)
```

Here, `Level` is the categorical variable which indicates class level. The residual sum-of-squares for this model is 69355.57.

(a) Find the F-statistic used to test the hypothesis that the true model is described by the smaller of the two models fit to the data.

(b) The correct F distribution for this statistic has 0.90, 0.95, 0.975 and 0.99 quantiles

$$1.822190, \quad 2.172501, \quad 2.511870, \quad 2.950754,$$

respectively. Use this information to estimate the p-value for the associated hypothesis test.

(c) What do you conclude about the data?

(d) In the second model fit, a linear model of the form

$$\mathbb{E}[\boldsymbol{Y} \mid \boldsymbol{X}] = \boldsymbol{X}\boldsymbol{\beta}$$

is fit (using OLS). Specify the columns of the matrix $\boldsymbol{X}$ used in the fitted model.

Solution. The F statistic is

$$F = \frac{(\mathrm{RSS}_0 - \mathrm{RSS}_1)/\Delta\mathrm{df}}{\mathrm{RSS}_1/\mathrm{df}} = \frac{(77396.53 - 69355.57)/6}{69355.57/(132-8)} = 2.39$$

Since $2.17 < 2.39 < 2.51$, the p-value is between 0.05 and 0.975.

We reject the hypothesis that the smaller model is sufficient, i.e. the hypothesis that one line should be fit for all class levels.

Let $\boldsymbol{\delta}_k$ be the dummy variable indicating membership in class $k = 2, 3, 4$. (Here 2 is sophomore, 3, junior, 4 senior.) Let $\boldsymbol{t}$ be the TEST 1 scores. Then

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{1} & \boldsymbol{t} & \boldsymbol{\delta}_2 & \boldsymbol{\delta}_3 & \boldsymbol{\delta}_4 & \boldsymbol{\delta}_2\boldsymbol{t} & \boldsymbol{\delta}_3\boldsymbol{t} & \boldsymbol{\delta}_4\boldsymbol{t} \end{bmatrix}$$

□