Math 583 Quiz 10 fall 2021 Name_____

1) In numerical linear algebra, the least squares solution to "$\boldsymbol{Ax} = \boldsymbol{b}$" is of interest where the problem is actually the multiple linear regression model $\boldsymbol{b} = \boldsymbol{Ax} + \boldsymbol{\epsilon}$ where $\boldsymbol{A}$ has full rank p, and we will assume that $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$, and $Cov(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$.

a) What is the (formula for the) projection matrix $\boldsymbol{P}$ onto the column space of $\boldsymbol{A}$?

b) What is the OLS estimator $\hat{\boldsymbol{x}}$?

c) What is the vector of fitted values $\hat{\boldsymbol{b}}$?

d) What is the residual vector $\boldsymbol{e}$?

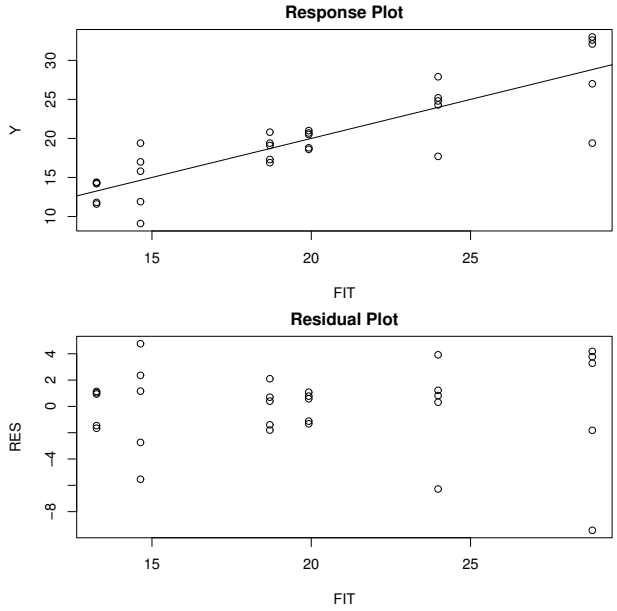

Figure 1: Plots for Clover Data

2) Following *SAS Institute* (1985, pp. 126-129), the mean nitrogen content of clover depends on the strain of clover (1=3dok1, 2=3dok5, 3=3dok4, 4=3dok7, 5=3dok13, 6=compos). So the Y_{ij} are the nitrogen contents of $n_i = 5$ clover plants for $i = 1, ..., 6$, and there are $p = 6$ groups. The least squares estimators of μ_i for the one way Anova model were $\overline{Y}_{10} = 28.2$, $\overline{Y}_{20} = 23.98$, $\overline{Y}_{30} = 14.64$, $\overline{Y}_{40} = 19.92$, $\overline{Y}_{50} = 13.26$, and $\overline{Y}_{60} = 18.7$. From the response and residual plots, the constant variance assumption is violated and $\max(R_1, ..., R_6) > 2\min(R_1, ..., R_6)$. In the response plot, the sample means of strains 3 and 5 are completely below the dot plots of the other 4 strains while the sample means of strains 1 and 2 are completely above the dot plots of the other 4 strains. Thus the response plot suggests that $\mu_3, \mu_5 < \mu_2, \mu_4 < \mu_1, \mu_2$. It is not clear from the response plot whether i) μ_3 and μ_5 differ, ii) μ_2 and μ_4 differ, or iii) μ_1 and μ_2 differ.

a) For the response plot shown above, draw in horizontal lines $Y = 23.98$ and $Y = 14.64$ to illustrate the above discussion.

b) In the residual plot, circle the two dot plots where the ranges are less than 4.

c) Do the plots suggest that linearity is reasonable?