STAT 513 fa 2020 hw 7

Do NOT use fancy R functions like `lm()` or `t.test()` to do any part of this homework.

1. Bring into the workspace the built-in R data set called `Puromycin` using the command `data(Puromycin)`. You may type `?Puromycin` to read more about the data. It contains the variables `rate`, `conc`, and `state`. Let $Y_1, \dots, Y_n$ denote the values in the column `rate`, $x_{11}, \dots, x_{n1}$ the values in the column `conc`, and $x_{12}, \dots, x_{n2}$ the values defined by

$$x_{i2} = \begin{cases} 1 & \text{if } \texttt{state}_i = \texttt{treated} \\ 0 & \text{if } \texttt{state}_i = \texttt{untreated} \end{cases} \qquad \text{for } i = 1, \dots, n,$$

where $\texttt{state}_i$ is the ith value of the column `state`. Suppose we wish to fit the model

$$Y_i = \beta_0 + \beta_1 \log(x_{i1}) + \beta_2 x_{i2} + \beta_3 x_{i2} \log(x_{i1}) + \varepsilon_i, \quad i = 1, \dots, n,$$

where we believe that $\varepsilon_1, \dots, \varepsilon_n$ are independent $\text{Normal}(0, \sigma^2)$ random variables.

(a) Define new covariate values u_{i1}, u_{i2}, and u_{i3}, $i = 1, \dots, n$, such that the above model can be expressed as

$$Y_i = \beta_0 + \beta_1 u_{i1} + \beta_2 u_{i2} + \beta_3 u_{i3} + \varepsilon_i, \quad i = 1, \dots, n.$$

Your answer should be like $u_{i1} =$, $u_{i2} =$, and $u_{i3} =$.

(b) State the regression model for the treated and the untreated cases; that is, give an expression for Y_i when $x_{i2} = 1$ and when $x_{i2} = 0$.

(c) Use R to construct the design matrix

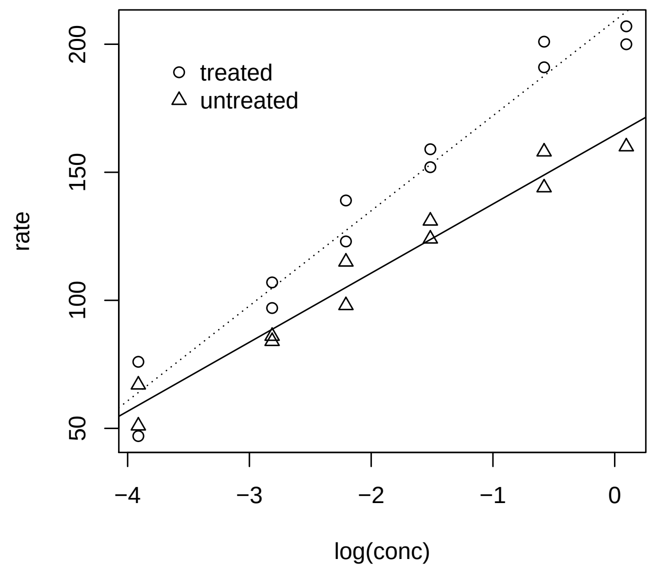
$$\mathbf{X} = \begin{bmatrix} 1 & u_{11} & u_{12} & u_{13} \\ \vdots & \vdots & \vdots & \\ 1 & u_{n1} & u_{n2} & u_{n3} \end{bmatrix}$$

and then compute the least-squares estimators of β_0, β_1, β_2, and β_3.

(d) The following R code produces a scatterplot of $Y_1, \dots, Y_n$ against the values $\log(x_{11}), \dots, \log(x_{n1})$, where circles are used for the treatment cases and triangles for the untreated cases:

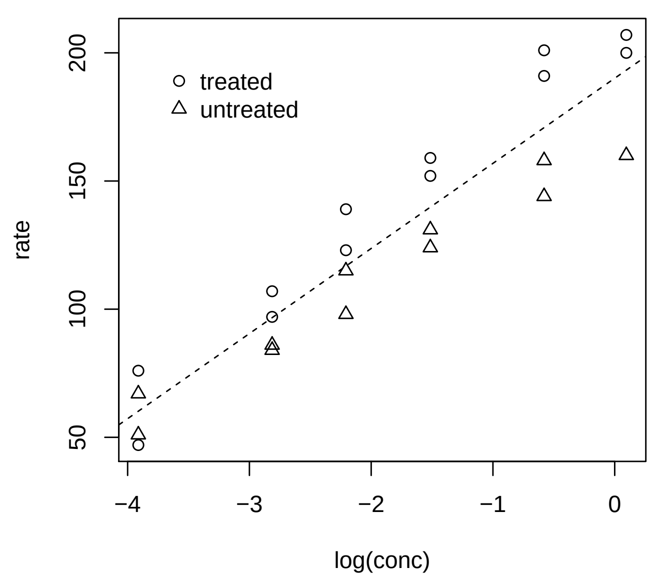
```
plot(Y~u1,pch=ifelse(u2==1,1,2),xlab="log(conc)",ylab="rate")
legend(x = -3.75, y = 200,legend=c("treated","untreated"),pch=c(1,2),bty="n")
```

Add some commands to this R code in order to produce a plot like the one below, with least-squares lines fitted to the treated and untreated cases. *Hint: Think about how to use the least-squares estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ to produce these lines with the* `abline()` *function.*



(e) Suppose we wish to investigate whether the treatment has any effect; more precisely, suppose we wish to test whether the regression functions for treated and untreated cases are the same. Give a set of hypotheses we could test in order to decide whether the two regression functions are equal.

(f) Consider the plot below, in which a least-squares line has been fitted to the points without regard for the treatment variable. Give the intercept and slope of this line.



(g) Compute the sum of the squared residuals for the model in part (f).

(h) Compute the sum of the squared residuals for the model in part (a).

(i) Compute the F-statistic for testing the hypotheses in part (e) with the full-reduced model F-test.

(j) Give the critical value for the full-reduced model F-test at the $\alpha = 0.01$ significance level.

(k) State your conclusion about the hypotheses in part (e) at the $\alpha = 0.01$ significance level.

(l) Give the p-value of these data for testing the hypotheses in part (e).

(m) Using the full model in part (a), construct a 95% confidence interval for the height of the true regression function for treated cases when the natural log of the concentration is equal to -1.

(n) Using the full model in part (a), construct a 95% confidence interval for the difference in the heights of the true regression functions for treated and untreated cases when the natural log of the concentration is equal to -1.

(o) Give the ANOVA table resulting from fitting the full model in part (a).

2. Show that in the $p = 1$ case (simple linear regression), the matrix formula $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ gives

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1\bar{x}_n \quad \text{and} \quad \hat{\beta}_1 = S_{xY}/S_{xx}.$$

Hint: for any 2×2 invertible matrix we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$