STAT 635 F2022 Assignment 1 - Due on October 10, Monday, 2022, 11:59pm

Instructor: Xuewen Lu (This PDF is made from an R Markdown File)

2022-10-10

Contents

.

General Policies

- Your assignment solutions should be created in `"RStudio"` to include the R codes and should be saved as an R Markdown file and a PDF file generated from the R Markdown file. You should name your files as `"Lastname-Firstname-Stat635-A1.Rmd"` and `"Lastname-Firstname-Stat635-A1.pdf"`. To make an R Markdown file for your assignment, you can use the Rmd file posted in D2L for this assignment as a template to start.
- For numerical problems, use R codes for computation. Test your R codes before submission and make sure it can be executed successfully in `"RStudio"`. The R codes should be included in R chunks and R chunks should be inserted into your R Markdown file and included in the PDF.
- For each assignment, submit one PDF file and the associated Rmd file only to D2L for checking the reproducibility, upload one identical PDF to **gradescope.ca** for grading. Only the PDF file is graded. The Rmd file may be used to test your R programs when needed.
- If Monte-Carlo methods are used, you must fix the random seed in your R code by using **set.seed(2022)**.
- For all numerical problems, summarize the computer generated results in tables or figures and interpret them using your own words, then draw conclusions from them. **Show all your work and spell out the details**.
- Late submission is not acceptable.
- For guidelines on how to write a good assignment, go to D2L to read two sample assignments, **Bad-assignment-example.pdf** and **Good-assignment-example.pdf**, one is bad and the other is good, you are expected to do a good one.

Note: Materials are based on Dobson and Barnett (D&B or DB), 3rd Edition, 2008.

Problem 1.

Let $Y_1, \ldots, Y_n$ be independent random variables each with the distribution $N(\mu, \sigma^2)$. Let

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \text{ and } S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

(a) Show what is the distribution of $\bar{Y}$?

We compute the expectation value of $\bar{Y}$ first and then we compute the variance of $\bar{Y}$

$$E[\bar{Y}] = E[\frac{1}{n} \sum_{i=1}^{n} Y_i] = \frac{1}{n} \sum_{i=1}^{n} E[Y_i] = \frac{1}{n} n E(Y) = \mu$$

$$\begin{aligned}
Var[\bar{Y}] &= Var(\frac{1}{n} \sum_{i=1}^{n} Y_i) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} Var(Y_i) + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} Cov(Y_i, Y_j) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} Var(Y_i) \qquad\qquad\qquad \text{by independence} \\
&= \frac{n}{n^2} Var(Y_i) \\
&= \frac{1}{n} Var(Y_i) \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

Since each Y_i is normally distributed and independent, their linear combination, such as $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$, should also be normally distributed

Therefore, $\frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$

(b) Show that $S^2 = \frac{1}{n-1}[\sum_{i=1}^n (Y_i - \mu)^2 - n(\bar{Y} - \mu)^2]$.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu + \mu - \bar{Y})^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n [(Y_i - \mu) - (\bar{Y} - \mu)]^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n [(Y_i - \mu)^2 - 2(Y_i - \mu)(\bar{Y} - \mu) + (\bar{Y} - \mu)^2]$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - 2\sum_{i=1}^n (Y_i - \mu)(\bar{Y} - \mu) + \sum_{i=1}^n (\bar{Y} - \mu)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - 2\sum_{i=1}^n (Y_i\bar{Y} - Y_i\mu - \bar{Y}\mu + \mu^2) + \sum_{i=1}^n (\bar{Y} - \mu)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - (2\bar{Y}\sum_{i=1}^n Y_i - 2\mu n\bar{Y} - 2n\bar{Y}\mu + 2n\mu^2) + n(\bar{Y} - \mu)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - (2n\bar{Y}^2 - 2\mu n\bar{Y} - 2n\bar{Y}\mu + 2n\mu^2) + n(\bar{Y} - \mu)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - (2n\bar{Y}^2 - 2\mu n\bar{Y} - 2n\bar{Y}\mu + 2n\mu^2) + n\bar{Y}^2 - 2n\mu\bar{Y} + n\mu^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - 2n\bar{Y}^2 + 2\mu n\bar{Y} + 2n\bar{Y}\mu - 2n\mu^2 + n\bar{Y}^2 - 2n\mu\bar{Y} + n\mu^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - n\bar{Y}^2 + 2n\mu\bar{Y} - n\mu^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2 - n(\bar{Y}^2 - 2\mu\bar{Y} + \mu^2)$$

$$= \frac{1}{n-1} [\sum_{i=1}^n (Y_i - \mu)^2 - n(\bar{Y} - \mu)^2] \qquad \text{as required}$$

(c) From (b) it follows that the $\sum (Y_i - \mu)^2/\sigma^2 = (n-1)S^2/\sigma^2 + [(\bar{Y} - \mu)^2 n/\sigma^2]$. How does this allow you to deduce that $\bar{Y}$ and S^2 are independent? [Hint: use Cochran's theorem.]

Denote $\sum (Y_i - \mu)^2/\sigma^2 = \Phi$ and denote $(n-1)S^2/\sigma^2 = \Psi$ and $[(\bar{Y} - \mu)^2 n/\sigma^2] = \Omega$

Recall Cochran's Theorem: Suppose $U_1, ..., U_n$ are independent standard normally distributed random variables, and an identity of the form $\sum_{i=1}^n U_i^2 = Q_1 + ... + Q_K$ can be written where each Q_i is a sum of squares of linear combinations of the U_s. Further suppose that $r_1 + ... + r_k = n$ where r_i is the rank of Q_i. Cochran's theorem states that the Q_i are independent, and Q_i has a χ^2 distribution with r_i degrees of freedom.

Recall the relationships of Chi-Squared Distributions and its MGF:

Property 1. If Z is a standard normal random variable, i.e., $Z \sim N(0,1)$, then the distribution of Z^2 is chi-squared with $k = 1$ degree of freedom.

Property 2. If $X_1, ..., X_n$ is a collection of independent, chi-squared random variables each with 1 degree of freedom, i.e., $X_i \sim \chi_1^2$, for each $i = 1, ..., n$, then $\sum X_1 + ... + X_n$ is also chi-squared with $k = n$ degrees of freedom.

Property 3. If $X \sim \chi_{k_1}^2$ and $Y \sim \chi_{k2}^2$ are independent random variables, then $X + Y \sim \chi_{(k1+k2)}^2$

Property 4. A χ_n^2 random variable has the moment generating function $(1 - 2t)^{\frac{-n}{2}}$

So the proof is...

From (b) it follows that

$$\Phi = \sum(Y_i - \mu)^2/\sigma^2 = \Psi + \Omega = (n-1)S^2/\sigma^2 + [(\bar{Y} - \mu)^2 n/\sigma^2]$$

We notice that Φ is a sum of n independent $X_{(1)}^2$ random variable. This is because we have assumed that $Y_1, ..., Y_n$ are from a random sample of size n from $N(\mu, \sigma^2)$. Thus, $\frac{X_i - \mu}{\sigma}$ follows a standard normal distribution. Then by the aforementioned property (2): $\Phi = \sum(Y_i - \mu)^2/\sigma^2$ is a X^2 random variable with degree of freedom of n. By the aforementioned property (4), Φ should have a moment generating function (MGF) $M_\Phi(t) = (1 - 2t)^{\frac{-n}{2}}$

We notice that Ω is a $\chi_{(1)}^2$ random variable because the sample mean is normally distributed with mean of μ and variance of $\frac{\sigma^2}{n}$. Therefore, $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and thus $\Omega = Z^2 \sim X_{(1)}^2$ by the aforementioned property (1), Ω should have a moment generating function (MGF) $M_\Omega(t) = (1 - 2t)^{\frac{-1}{2}}$

Now we wish to compute the MGF of Ψ and we know $\Phi = \sum(Y_i - \mu)^2/\sigma^2 = \Psi + \Omega = (n-1)S^2/\sigma^2 + [(\bar{Y} - \mu)^2 n/\sigma^2]$ and the MGF of Φ and Ω, therefore, it is easy to compute the MGF of Ψ. Indeed, its MGF is $M_\Psi(t) = (1 - 2t)^{\frac{n-1}{-2}}$.

By Uniqueness Property of MGF and the aforementioned Property (4), we determine that Ψ is a χ^2 random variable with degree of freedom of $(n-1)$. That is $\Psi \sim \chi_{(n-1)}^2$

Also if we factor out all the constant terms, then we notice that Φ contains one random variable Yi, so that $\sum_{i=1}^n Y_i^2$ can be written in matrix format $Y^T I Y$ where I is the identity matrix. Similarly, Ψ contains random variable S^2, so that $\sum_{i=1}^n (Y_i - \bar{Y})^2$ can be written as $Y^T(I - \frac{1}{n}J)Y$ where J is matrix of ones. Similarly, Ω contains random variable $\bar{Y}^2$, and $\left(\frac{\sum_{i=1}^n Y_i}{n}\right)^2$ can be written as $Y^T(\frac{1}{n^2}J)Y$. As $Y_1, ..., Y_n$ are independent random variables and each with the normal distribution and also we have determined the quadratic form and the degree of freedom. We are now safe to say that by Cochran's Theorem, as we have shown $\Phi \sim \chi_{(n)}^2$, $\Psi \sim \chi_{(n-1)}^2$ and $\Omega \sim \chi_{(1)}^2$, Ψ and Ω are independent random variables. If we factor out all the constant terms, then clearly S^2 and $\bar{Y}$ are independent.

(d) Show what is the distribution of $(n-1)S^2/\sigma^2$?

Indeed we have determined its distribution is $\chi_{(n-1)}^2$ using MFG on part c).

$$M_\Psi(t) = \frac{M_\Phi(t)}{M_\Omega(t)} = \frac{(1 - 2t)^{\frac{-n}{2}}}{(1 - 2t)^{\frac{-1}{2}}} = (1 - 2t)^{\frac{n-1}{-2}}$$

It is more efficient to find it by Cochran's Theorem now.

(e) Show what is the distribution of $\frac{\bar{Y} - \mu}{S/\sqrt{n}}$?

5

By definition of t-distribution, it is defined as the ratio of two independent random variables. The numerator has the standard normal distribution and the denominator is the square root of a central chi-squared random variable divided by its degree of freedom; that is,

$$T = \frac{Z}{\sqrt{(X^2/n)}}$$

where $Z \sim N(0,1), X^2 \sim \chi^2(n)$ and Z and X^2 are independent. This is denoted by $T \sim t(n)$

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{S/\sigma}$$

$$= \frac{\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}}$$

$$= \frac{\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}}$$

Now the numerator part, let's say $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} = A$ and within the denominator part, let's denote $(n-1)S^2/\sigma^2 = B$. We can see that $A \sim N(0,1)$ and $B \sim \chi^2_{(n-1)}$, which is the result from previous question. In addition to that, A and B are independent because of the fact that we proved on part c).

By definition of t-distribution stated above, $\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$

Problem 2.

For each of the two different forms of simple linear regression, namely

$$E[y_i] = \alpha + \gamma x_i \text{ and } E[y_i] = \mu + \gamma(x_i - \bar{x})$$

for $i = 1, 2, \ldots, n$:

(a) For each model write $E[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\beta}$, specifying $\boldsymbol{X}$ and $\boldsymbol{\beta}$. Note: $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{\beta}$ are vectors or matrices.

For the first form of simple linear regression, we have $E[y_i] = \alpha + \gamma x_i$ for i = 1,2,...,n.

$$E[y_1] = \alpha + \gamma x_1$$
$$E[y_2] = \alpha + \gamma x_2$$
$$E[y_3] = \alpha + \gamma x_3$$
$$\ldots$$
$$E[y_n] = \alpha + \gamma x_n$$

We can also write this in matrix form: $E[Y] = X\beta$ where

$$E[Y]_{n \times 1} = \begin{pmatrix} E[Y_1] \\ E[Y_2] \\ \vdots \\ E[Y_n] \end{pmatrix}$$

6

where

$$X_{n \times 2} = \begin{pmatrix} 1 & x_1 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

where

$$\beta_{2 \times 1} = \begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$$

For the second form of simple linear regression, we have $E[y_i] = \mu + \gamma(x_i - \bar{x})$ for i = 1,2,...,n.

$$E[y_1] = \mu + \gamma(x_1 - \bar{x})$$
$$E[y_2] = \mu + \gamma(x_2 - \bar{x})$$
$$E[y_3] = \mu + \gamma(x_3 - \bar{x})$$
$$...$$
$$E[y_n] = \mu + \gamma(x_n - \bar{x})$$

We can also write this in matrix form: $E[Y] = X\beta$ where

$$E[Y]_{n \times 1} = \begin{pmatrix} E[Y_1] \\ E[Y_2] \\ \vdots \\ E[Y_n] \end{pmatrix}$$

where

$$X_{n \times 2} = \begin{pmatrix} 1 & (x_1 - \bar{x}) \\ 1 & (x_2 - \bar{x}) \\ \vdots & \vdots \\ 1 & (x_n - \bar{x}) \end{pmatrix}$$

where

$$\beta_{2 \times 1} = \begin{pmatrix} \mu \\ \gamma \end{pmatrix}$$

(b) For each model obtain $\boldsymbol{X}^\top \boldsymbol{X}$ and $\hat{\boldsymbol{\beta}}$.

For the first form, we could obtain:

$$X^T X = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \times \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

Using Least Squares Method and the idea of minimizing the sum of squares:

$$\sum_i [Y_i - (\alpha + \gamma x_i)]^2 = (Y - X\beta)^T (Y - X\beta)$$

Take derivative of above expression with respect to parameter β

$$\frac{\partial}{\partial \beta}[(Y - X\beta)^T (Y - X\beta)] = 2(X^T X\beta - X^T Y)$$

So that using Least Squares Method, one can compute $\hat{\beta}$ in matrix format:

$$\hat{\beta} = (X^TX)^{-1}X^TY = \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n x_iY_i \end{pmatrix}$$

$$= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 \sum_{i=1}^n Y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_iY_i \\ n\sum_{i=1}^n x_iY_i - \sum_{i=1}^n x_i \sum_{i=1}^n Y_i \end{pmatrix}$$

For the second form, we could obtain:

$$X^TX = \begin{pmatrix} 1 & \cdots & 1 \\ (x_1 - \bar{x}) & \cdots & (x_n - \bar{x}) \end{pmatrix} \times \begin{pmatrix} 1 & (x_1 - \bar{x}) \\ 1 & (x_2 - \bar{x}) \\ \vdots & \vdots \\ 1 & (x_n - \bar{x}) \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n(x_i - \bar{x}) \\ \sum_{i=1}^n(x_i - \bar{x}) & \sum_{i=1}^n(x_i - \bar{x})^2 \end{pmatrix} = \begin{pmatrix} n & 0 \\ 0 & \sum_{i=1}^n(x_i - \bar{x})^2 \end{pmatrix}$$

Similarly, one can compute $\hat{\beta}$ in matrix format using Least Squares Method:

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

$$= \frac{1}{n\sum_{i=1}^n(x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n(x_i - \bar{x})^2 & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ (x_1 - \bar{x}) & \cdots & (x_n - \bar{x}) \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$= \frac{1}{n\sum_{i=1}^n(x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n(x_i - \bar{x})^2 & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n(x_i - \bar{x})Y_i \end{pmatrix}$$

$$= \frac{1}{n\sum_{i=1}^n(x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n(x_i - \bar{x})^2 \sum_{i=1}^n Y_i \\ n\sum_{i=1}^n(x_i - \bar{x})Y_i \end{pmatrix}$$

(c) For each model obtain $var(\hat{\beta})$. Do you see any advantage of one model equation over the other?

Note: we assume that X is of full rank and X^TX is a symmetric matrix. Also, we assume that $Y \sim MVN(X\beta, \sigma^2 I)$ in linear regression

$$var(\hat{\beta}) = (X^TX)^{-1}X^T var(Y)X(X^TX)^{-1}$$
$$= (X^TX)^{-1}X^T\sigma^2 IX(X^TX)^{-1}$$
$$= \sigma^2(X^TX)^{-1}$$

The second form of the model has advantage over the the first form because the computation of the $(X^TX)^{-1}$ of the second one is more computational friendly than the first form since the off-diagonal terms are zero in the second form

(d) For each model obtain the variance of the estimated value of the mean of y_i at a new value of x, denoted as x^*.

Note in simple linear regression $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ and we have the following properties:

$\hat{\beta}_0 \sim N\left(\beta_0, (\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n(x_i - \bar{x})^2})\sigma^2\right)$ and $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n(x_i - \bar{x})^2}\right)$

For the first form of simple linear regression, we have $E[y_i|x^*] = \mu_{x^*} = \alpha + \gamma x^*$. So that we will estimate it with $\hat{\mu_{x^*}} = \hat{\alpha} + \hat{\gamma}x^*$

8

Compute the variance of the estimated value of the mean of y_i at a new value of x^*

$$\begin{aligned} Var(\hat{\mu_{x^*}}) &= Var(\hat{\alpha} + \hat{\gamma}\hat{x^*}) \\ &= Var(\bar{y} - \hat{\gamma}\bar{x} + \hat{\gamma}x^*) \\ &= Var(\bar{y} + \gamma(x^* - \bar{x})) \\ &= Var(\bar{y}) + (x^* - \bar{x})^2 Var(\hat{\gamma}) \\ &= \frac{\sigma^2}{n} + \frac{(x^* - \bar{x}^2)\sigma^2}{\sum_{i=1}^{n}(x^* - \bar{x})^2} \\ &= [\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x^* - \bar{x})^2}]\sigma^2 \end{aligned}$$

note: $Cov(\bar{y}, \hat{\gamma}) = 0$

For the second form of simple linear regression, we have $E[y_i|x^*] = \xi_{x^*} = \mu + \gamma(x^* - \bar{x})$. So that we will estimate it with $\hat{\xi_{x^*}} = \hat{\mu} + \hat{\gamma}(x^* - \bar{x})$

Compute the variance of the estimated value of the mean of y_i at a new value of x^*

$$\begin{aligned} Var(\hat{\xi_{x^*}}) &= Var(\hat{\mu} + \hat{\gamma}(x^* - \bar{x})) \\ &= Var(\bar{y} + \gamma(x^* - \bar{x})) \\ &= Var(\bar{y}) + (x^* - \bar{x})^2 Var(\hat{\gamma}) \\ &= \frac{\sigma^2}{n} + \frac{(x^* - \bar{x}^2)\sigma^2}{\sum_{i=1}^{n}(x^* - \bar{x})^2} \\ &= [\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x^* - \bar{x})^2}]\sigma^2 \end{aligned}$$

note: $Cov(\bar{y}, \hat{\gamma}) = 0$ and $\hat{\mu} = \bar{y}$. This is because the second form of linear regression is just redefining the regressor variable as the deviation from its own average.

Problem 3.

An experiment analyzes imperfection rates for two processes used to fabricate silicon wafers for computer chips. For treatment A applied to 10 wafers, the numbers of imperfections are 8, 7, 6, 6, 3, 4, 7, 2, 3, 4. Treatment B applied to 10 other wafers has 9, 9, 8, 14, 8, 13, 11, 5, 7, 6 imperfections. Treat the counts as independent Poisson variates having means μ_A and μ_B. Consider the model $\log \mu = \alpha + \beta x$, where $x = 1$ for treatment B and $x = 0$ for treatment A.

(a) Show that $\beta = \log \mu_B - \log \mu_A = \log(\mu_B/\mu_A)$ and $e^\beta = \mu_B/\mu_A$.

$\log \mu_A = \alpha$ and $\log \mu_B = \alpha + \beta$.

Therefore, $\log \mu_B - \log \mu_A = \log(\mu_B/\mu_A) = \alpha + \beta - \alpha = \beta$

If we do exponentiation on both sides, then $e^\beta = \mu_B/\mu_A$

(b) Fit the model. Report the prediction equation and interpret β.

```
x = c(rep(c(0,1), each=10))
imperfection = c(8, 7, 6, 6, 3, 4, 7, 2, 3, 4, 9, 9, 8, 14, 8, 13, 11, 5, 7, 6)
dat = data.frame(cbind(x,imperfection))
model = glm(imperfection~x, family = poisson(link=log), data = dat)
summary(model)
```

```
##
## Call:
## glm(formula = imperfection ~ x, family = poisson(link = log),
##     data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5280  -0.7622  -0.1699   0.6938   1.5399
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.6094     0.1414  11.380  < 2e-16 ***
## x             0.5878     0.1764   3.332 0.000861 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 27.857  on 19  degrees of freedom
## Residual deviance: 16.268  on 18  degrees of freedom
## AIC: 94.349
##
## Number of Fisher Scoring iterations: 4
```

```
## the inverse of the information matrix
summary(model)$cov.unscaled
```

```
##             (Intercept)           x
## (Intercept)  0.01999992 -0.01999992
## x           -0.01999992  0.03111102
```

```
## p value using the Deviance Test
1-pchisq(model$deviance, 18)
```

```
## [1] 0.5738748
```

So the prediction equation is $log(\mu) = 1.6094 + 0.5878x$. In this model, an exponentiated coefficient $exp(0.5878) = 1.800024$ represents increasing x1 by one unit (i.e., get treatment B instead of treatment A), μ is going to increase 80%.

(c) Test $H_0: \quad \mu_A = \mu_B$ by conducting the Wald and likelihood-ratio test of $H_0: \quad \beta = 0$. Interpret the result.

Wald Test:

Step 1: $H_0 : \beta = 0$ and $H_a : \beta \neq 0$

Step 2: Wald Statistic: $W = \frac{(\hat{\beta}-0)^2}{I(\hat{\beta})^{-1}} = \frac{0.5878^2}{0.03111102} = 11.10567 \sim \chi^2_{(df=1)}$ This is equivalent to $z = \frac{\hat{\beta}}{se(\hat{\beta})} = \frac{0.5878}{\sqrt{0.03111102}} = 3.332518 \sim N(0,1)$

Step 3: from the summary table above, we see $p-value = 2 \times P(Z > |z|) = 0.000861 < 0.05$ where $Z \sim N(0,1)$

Step 4: we reject the null hypothesis, which means that applying treatment A or B do make a difference, and this variable is statistically significant.

Likelihood-ratio Test:

Step 1: $H_0 : \beta = 0$ and $H_a : \beta \neq 0$

Step 2: Likelihood Ratio (LR)/Deviance Test Statistic: $D =$ Residual Deviance $\sim \chi^2_{(df=n-p)}$

Step 3: from the summary table above, we see the residual deviance $= 16.268$ and degree of freedom is 18 and also we could compute $p-value = P[\chi^2_{(df=18)} > D] = 0.5738748 > 0.05$

Step 4: we do not reject the null hypothesis, which means that applying treatment A or B does not make a difference,and this variable is not statistically significant.

(d) Construct a 95% confidence interval for μ_B/μ_A. [Hint: Construct one for $\beta = \log(\mu_B/\mu_A)$ and then exponentiate.]

We compute a 95% confidence interval for $\beta = log(\mu_B/\mu_A)$ first:

$$\hat{\beta} \pm 1.96se(\hat{\beta}) = 0.5878 \pm 1.96 \times 0.1764 = (0.242056, 0.933544)$$

Then we exponentiate this interval to the 95% CI for μ_B/μ_A and we get: $(1.273866, 2.543507)$

Note: For the following Problems [4] and [5], use R to implement algorithm for numerical results. Based on the R code you develop, please write down the key steps in the algorithm and explain those steps in plain language, and save the R code in Rmd (with the datasets) for further examination.

Problem 4.

The Puromycin Data. Bates and Watts (1988) present data on the velocity of an enzymatic reaction, where the substrate has been treated with puromycin at several concentrations. The velocity and concentration data are shown in Table 1. Bates and Watts propose fitting the Michaelis-Menten model for chemical kinetics

$$E(y_i) = f(x_i, \beta) = \frac{\beta_1 x_i}{\beta_2 + x_i}$$

to the data. Note that the expectation function of the Michaelis-Menten model can easily be linearized because

$$\frac{1}{f(x_i, \beta)} = \frac{\beta_2 + x_i}{\beta_1 x_i} = \alpha_0 + \alpha_1 z_i,$$

where $\alpha_0 = 1/\beta_1$, $\alpha_1 = \beta_2/\beta_1$, $z_i = 1/x_i$. Therefore we are tempted to fit the linear regression model

$$y_i^* = \alpha_0 + \alpha_1 z_i + \epsilon_i,$$

where $y_i^* = 1/y_i$ is the reciprocal of the observed velocity.

Table 1: Reaction Velocity and Substrate Concentration for Puromycin Experiment

Substrate Concentration (ppm)	Velocity (counts/min2)	
0.02	47	76
0.06	97	107
0.11	123	139
0.22	152	159
0.56	191	201
1.10	200	207

```
concentration = c(rep(c(0.02,0.06,0.11,0.22,0.56,1.10), each=2))
velocity = c(47,76,97,107,123,139,152,159,191,201,200,207)
dat = data.frame(cbind(concentration,velocity))
head(dat)
```

```
##    concentration velocity
## 1           0.02       47
## 2           0.02       76
## 3           0.06       97
## 4           0.06      107
## 5           0.11      123
## 6           0.11      139
```

```
zi = 1/concentration
yi_star = 1/velocity
model = lm(yi_star~zi)
summary(model)
```

```
##
## Call:
## lm(formula = yi_star ~ zi)
##
## Residuals:
##         Min         1Q     Median         3Q        Max
## -0.0043103 -0.0003742 -0.0000510  0.0004549  0.0038084
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0051072  0.0007040   7.255 2.74e-05 ***
## zi          0.0002472  0.0000321   7.700 1.64e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001892 on 10 degrees of freedom
## Multiple R-squared:  0.8557, Adjusted R-squared:  0.8413
## F-statistic:  59.3 on 1 and 10 DF,  p-value: 1.642e-05
```

Do the following:

(a) Find the least squares fit for y_i^*.

$\hat{y}_i^* = 0.00511 + 0.00025\hat{z}^*$

(b) Presents a scatter diagram of the transformed data with this straight-line fit in (a) superimposed.

```
# Add scatterplot
plot(zi,yi_star, main="Scatterplot of the Transformed Data",
    xlab="transformed concentration ", ylab="transformed velocity", pch=19)
# Add fit lines
abline(lm(yi_star~zi), col="red") # regression line
```


Scatterplot of the Transformed Data
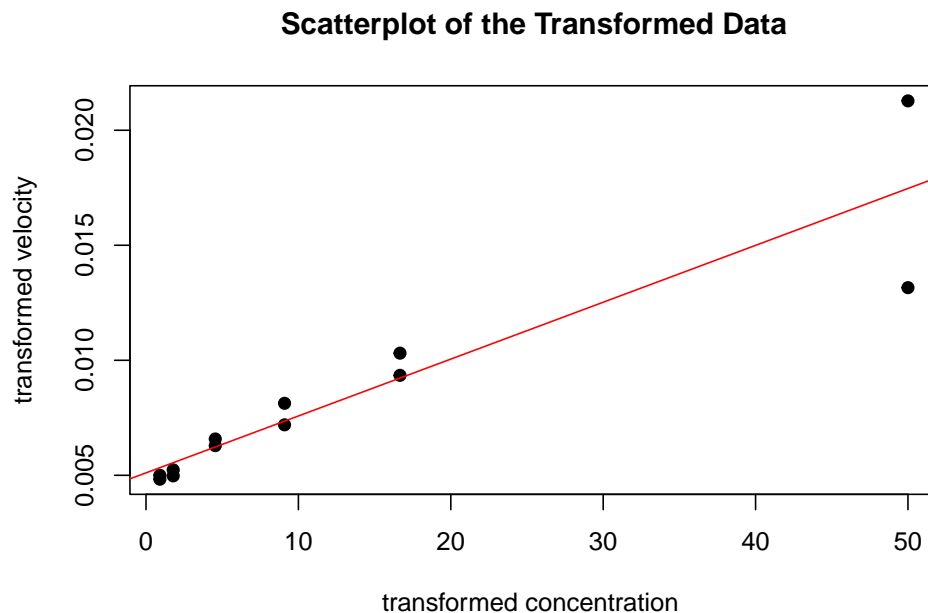
(c) Find estimates β_1 and β_2 in the original nonlinear model.

From part b), we know $\hat{\alpha}_0 = 1/\beta_1 = 0.00511$ and $\hat{\alpha}_1 = \beta_2/\beta_1 = 0.00025$. Therefore, $\hat{\beta}_1 = \frac{1}{0.00511} = 195.6947$ and $\hat{\beta}_2 = 0.00025 \times 195.6947 = 0.04892$

(d) Presents a scatter diagram to show the fitted nonlinear model in the original scale superimposed on the scatter plot of the data.

```
# Add scatterplot
plot(concentration,velocity, main="Scatterplot of the Original Data",
    xlab="concentration", ylab="velocity ", pch=19)
# Add this nonlinear model fitted line
lines(concentration,195.6947*concentration/(0.04892+concentration),col="red",lty = 1)
```


Scatterplot of the Original Data
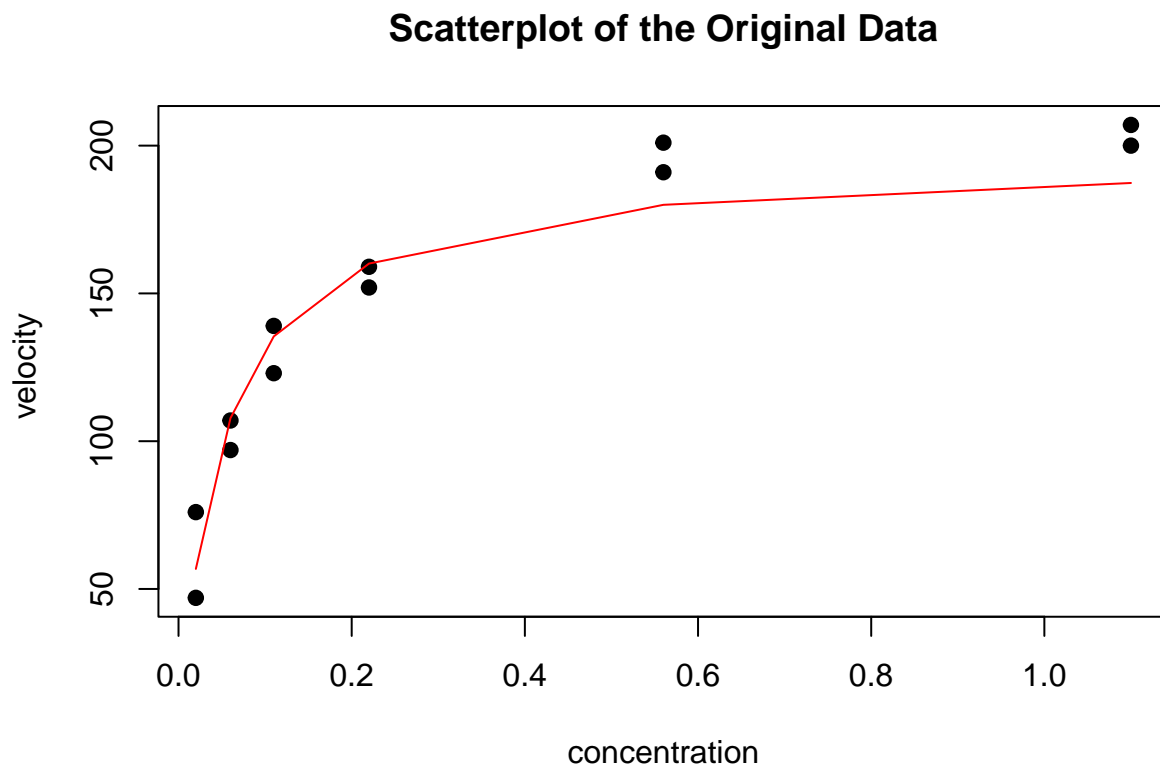
(e) Comment on how the variance of the replicated response observations influences the least squares fit.

Randomness plays an important role in statistics because it ensures that a study is conducted without bias. Replication means that samples can no longer be considered random, which is a big problem. In fact, variation may be reduced if sampling is considered as not random. If there is some correlation between the replicates, the variance is less than would be expected if the numbers were randomly selected.

Problem 5.

In this problem, we use a different approach to the Puromycin Data. We reconsider the puromycin data and use Newton-Raphson algorithm to fit the Michaelis-Menten model. We assume a nonlinear model as follows:

$$y_i = f(x_i, \beta) + \epsilon_i,$$

where $f(x_i, \beta) = \frac{\beta_1 x_i}{\beta_2 + x_i}$, $\epsilon_i \sim N(0, \sigma^2)$ and are iid.

(a) Write down the log-likelihood function for the unknown parameters based on the observed data.

The normality of random errors imply the normality of response. So we know $y_i \sim N(\frac{\beta_1 x_i}{\beta_2 + x_i}, \sigma^2)$ and keep in mind that $y_1, y_2, ...y_n$ are independent.

Likelihood function ($\theta = [\beta_1, \beta_2, \sigma^2]^T$):

$$L(\theta) = f(y_1, y_2, ..., y_n)$$
$$= \prod_{i=1}^{n} f(y_i; \beta_1, \beta_2, \sigma^2)$$
$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(y_i - \frac{\beta_1 x_i}{\beta_2 + x_i})^2)$$

Log-likelihood function ($\theta = [\beta_1, \beta_2, \sigma^2]^T$):

$$l(\theta) = log[L(\theta)]$$
$$= \sum_{i=1}^{n}[-\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_i - \frac{\beta_1 x_i}{\beta_2 + x_i})^2]$$
$$= -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \frac{\beta_1 x_i}{\beta_2 + x_i})^2$$

(b) Develop a Newton-Raphson algorithm to estimate β_1 and β_2 iteratively, using as the starting values for the parameters $b_{10} = 205$ and $b_{20} = 0.08$.

first we compute $\frac{\partial}{\partial \beta_1}l(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \frac{\beta_1 x_i}{\beta_2 + x_i})(\frac{x_i}{\beta_2 + x_i})$ and set it to zero

then we compute $\frac{\partial}{\partial \beta_2}l(\theta) = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \frac{\beta_1 x_i}{\beta_2 + x_i})(\frac{\beta_1 x_i}{(\beta_2 + x_i)^2})$ and set it to zero

so that the score function is

$$U(\theta) = \left[\frac{\partial}{\partial \beta_1}l(\theta), \frac{\partial}{\partial \beta_2}l(\theta)\right]^T = \left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \frac{\beta_1 x_i}{\beta_2 + x_i})(\frac{x_i}{\beta_2 + x_i}), -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \frac{\beta_1 x_i}{\beta_2 + x_i})(\frac{\beta_1 x_i}{(\beta_2 + x_i)^2})\right]^T$$

we now determine the information matrix $I_n(\theta)$

$$I_n(\theta) = -\begin{bmatrix} \frac{\partial^2}{\partial \beta_1^2}l(\theta) & \frac{\partial^2}{\partial \beta_1 \partial \beta_2}l(\theta) \\ \frac{\partial^2}{\partial \beta_1 \partial \beta_2}l(\theta) & \frac{\partial^2}{\partial \beta_2^2}l(\theta) \end{bmatrix} = -\begin{bmatrix} \sum_{i=1}^{n} -\frac{x_i^2}{(\beta_2 + x_i)^2} & \sum_{i=1}^{n} 3\frac{\beta_1 x_i^2}{(\beta_2 + x_i)^3} - \frac{x_i y_i}{(\beta_2 + x_i)^2} \\ \sum_{i=1}^{n} 3\frac{\beta_1 x_i^2}{(\beta_2 + x_i)^3} - \frac{x_i y_i}{(\beta_2 + x_i)^2} & \sum_{i=1}^{n} 3\frac{(\beta_1 x_i)^2}{(\beta_2 + x_i)^4} - 2\frac{\beta_1 x_i y_i}{(\beta_2 + x_i)^3} \end{bmatrix}$$

By using Newton-Raphson Method demonstrated below, we have determined that $\hat{\beta}_1 = 212.6837$ and $\hat{\beta}_2 = 0.0641$

```r
## import the data
x = c(rep(c(0.02,0.06,0.11,0.22,0.56,1.10), each=2))
y = c(47,76,97,107,123,139,152,159,191,201,200,207)
```

```r
library(matlib)
library(numDeriv)

loglike = function(beta)
  {-sum(y^2)+2*sum(y*((beta[1]*x)/(beta[2]+x)))-sum(((beta[1]*x)^2/(beta[2]+x)^2))}

newton.raphson = function(beta,
                          maxIteration, # max number of iterations
                          tolerance# parameters for the test
){
  ## initialize
  i = 0
  error = 1
  ## loop
  while(error > tolerance & i <= maxIteration){
    betaOld = beta
    score = grad(func=loglike,beta)
    information = -hessian(func=loglike,beta)
    beta = beta + inv(information)%*%score
    ## update iteration
    error = abs(sum(betaOld-beta))
    i = i+1
  }
  ## return last value and whether converged or not
  list(beta = beta,
       iteration = i
       )
}
```

```r
beta = c(205,0.08)
#estimate of beta
newton.raphson(beta,100,1E-5)
```

```
## $beta
##             [,1]
## [1,] 212.68374316
## [2,]   0.06412128
##
## $iteration
## [1] 7
```

(c) State the fitted model and plot the fitted model with the original data superimposed on it.

Fitted model:

$$\hat{y}_i = \frac{212.6837 x_i}{0.0641 + x_i}$$

```r
# Add scatterplot
plot(concentration,velocity, main="Scatterplot of the Original Data",
```
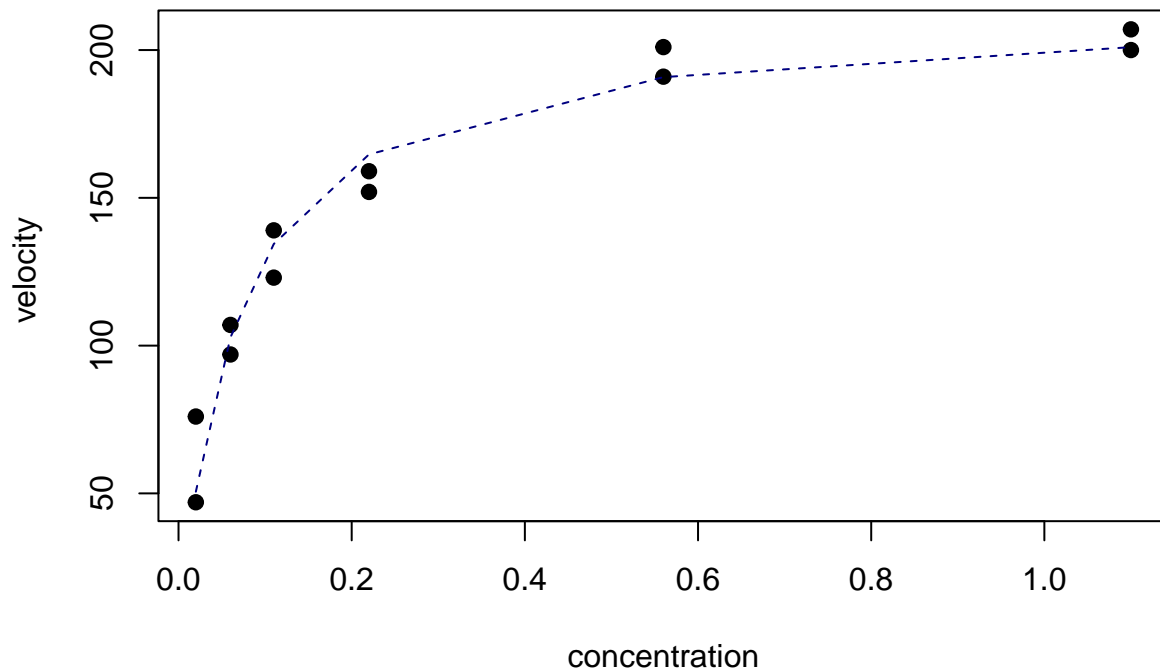
```
xlab="concentration", ylab="velocity ", pch=19)
# Add this nonlinear model fitted line
lines(concentration,212.6837*concentration/(0.0641+concentration),col="navy",lty = 2)
```

Scatterplot of the Original Data

(d) Compare this nonlinear model with the transformation followed by linear regression did in Problem 4.

The estimate of β is very close and I would say that the transformation did a good job. However, one can clearly see that the nonlinear model has a better fit, especially when the velocity and the concentration is high (top right part of the figure). Since it is nonlinear, more errors will be generated when the data points are far from the majority of the data points.

(e) When the estimation procedure converges to a final vector of parameter estimates, obtain an estimate of the error variance σ^2 from the residual, adjusted from the number of regression parameters.

The estimate of the error variance σ^2 from the residual is 119.5451:

$$\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y_i})^2$$

```
x = c(rep(c(0.02,0.06,0.11,0.22,0.56,1.10), each=2))
y = c(47,76,97,107,123,139,152,159,191,201,200,207)
n = length(x)
## fitted model
y_hat = 212.6837*x/(0.0641+x)
error_variance = (1/(n-2))*sum((y_hat-y)^2)
error_variance
```

```
## [1] 119.5451
```

17

Problem 6.

The gamma distribution is denoted by $G(\mu, \nu)$, its probability density function is given by

$$f(y) = \frac{y^{-1}}{\Gamma(\nu)} \left(\frac{y\nu}{\mu}\right)^{\nu} e^{-y\nu/\mu}, \quad y > 0.$$

(a) Write down the likelihood equations (log-likelihood and score functions) for the estimation of μ and ν of the gamma distribution.

$$f(y) = \frac{y^{-1}}{\Gamma(\nu)} \left(\frac{y\nu}{\mu}\right)^{\nu} e^{-y\nu/\mu} = \frac{y^{\nu-1}}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} e^{-y\nu/\mu}$$

Likelihood function:

$$L(\mu, \nu) = \prod_{i=1}^{n} f(y_i) = \frac{\prod_{i=1}^{n} y_i^{\nu-1}}{\Gamma(\nu)^n} \left(\frac{\nu}{\mu}\right)^{n\nu} e^{-\sum_{i=1}^{n} y\nu/\mu}$$

Log-likelihood function:

$$l(\mu, \nu) = log(L(\mu, \nu)) = (\nu - 1) \sum_{i=1}^{n} log(y_i) + n\nu log(\frac{\nu}{\mu}) - \frac{\nu}{\mu} \sum_{i=1}^{n} y_i - n log(\Gamma(\nu))$$

Score Functions:

$$\frac{\partial l(\nu, \mu)}{\partial \mu} = -n\frac{\nu}{\mu} + \frac{\nu}{\mu^2} \sum_{i=1}^{n} y_i$$

$$\frac{\partial l(\nu, \mu)}{\partial \nu} = \sum_{i=1}^{n} log(y_i) + n log(\nu/\mu) + n - \frac{1}{\mu} \sum_{i=1}^{n} y_i - n\frac{\Gamma'(\nu)}{\Gamma(\nu)}$$

(b) Show that the MLE of μ is $\hat{\mu} = \bar{y}$.

Set $\frac{\partial l(\nu, \mu)}{\partial \mu} = -n\frac{\nu}{\mu} + \frac{\nu}{\mu^2} \sum_{i=1}^{n} y_i = 0$. We have $n\frac{\nu}{\mu} = \frac{\nu}{\mu^2} \sum_{i=1}^{n} y_i = \frac{\nu}{\mu^2} n\bar{y}$. This means that $\hat{\mu} = \bar{y}$

Problem 7.

Vehicle insurance data. This data set is based on one-year vehicle insurance policies taken out in 2004 or 2005. There are 67856 policies of which 4624 (6.8%) had at least one claim. To get the data, you can install the R package "`insuranceData`", then type

```
library(insuranceData)
data(dataCar)
head(dataCar)
```

to see the first 6 observations. Do the following and save the R code and datasets for further examination.

```
library(insuranceData)
data(dataCar)
head(dataCar)
```

18

```
##   veh_value  exposure clm numclaims claimcst0 veh_body veh_age gender area
## 1      1.06 0.3039014   0         0         0    HBACK       3      F    C
## 2      1.03 0.6488706   0         0         0    HBACK       2      F    A
## 3      3.26 0.5694730   0         0         0      UTE       2      F    E
## 4      4.14 0.3175907   0         0         0    STNWG       2      F    D
## 5      0.72 0.6488706   0         0         0    HBACK       4      F    C
## 6      2.01 0.8542094   0         0         0    HDTOP       3      M    C
##   agecat         X_OBSTAT_
## 1      2 01101   0   0   0
## 2      4 01101   0   0   0
## 3      2 01101   0   0   0
## 4      2 01101   0   0   0
## 5      2 01101   0   0   0
## 6      4 01101   0   0   0
```

(a) For claim size in the vehicle insurance data [claim amount, the fifth column in the data set], write a program to estimate μ and ν. Since there are many zero responses in claim size, let's take only non-zero responses in data analysis, we can use R command `subset(y, (y != 0))` to get a subset of responses with only non-zero values. In programming, you may need to evaluate functions such as $\psi(\nu) = \Gamma'(\nu)/\Gamma(\nu)$ and $\psi'(\nu)$, these are called `digamma` and `trigamma` functions in R, they can be evaluated by $psigamma(\nu, deriv = 0)$ and $psigamma(\nu, deriv = 1)$, respectively.

Using the R-code below, one can determine that $\hat{\mu} = 2014.404$ and $\hat{\nu} = 0.7501495$

```
dataNew = subset(dataCar,claimcst0!=0)
n = length(dataNew$claimcst0)
n
```

```
## [1] 4624
```

```
head(dataNew)
```

```
##    veh_value  exposure clm numclaims claimcst0 veh_body veh_age gender area
## 15      1.66 0.4845996   1         1  669.5100    SEDAN       3      M    B
## 17      1.51 0.9938398   1         1  806.6100    SEDAN       3      F    F
## 18      0.76 0.5393566   1         1  401.8055    HBACK       3      M    C
## 41      1.89 0.6543463   1         2 1811.7100    STNWG       3      M    F
## 65      4.06 0.8514716   1         1 5434.4400    STNWG       2      M    F
## 66      1.39 0.3175907   1         1  865.7900    HBACK       3      F    A
##    agecat         X_OBSTAT_
## 15      6 01101   0   0   0
## 17      4 01101   0   0   0
## 18      4 01101   0   0   0
## 41      2 01101   0   0   0
## 65      3 01101   0   0   0
## 66      4 01101   0   0   0
```

```
y = dataNew$claimcst0
mu = mean(y)
mu
```

```
## [1] 2014.404
```

Consider the score function derived in Problem 6a), we have

$$\frac{\partial l(\nu,\mu)}{\partial \nu} = \sum_{i=1}^{n} log(y_i) + nlog(\nu/\mu) + n - \frac{1}{\mu}\sum_{i=1}^{n} y_i - n\frac{\Gamma'(\nu)}{\Gamma(\nu)}$$

$$= \sum_{i=1}^{n} log(y_i) + nlog(\nu/\mu) + n - \frac{1}{\mu}n\bar{y} - n\frac{\Gamma'(\nu)}{\Gamma(\nu)}$$

$$= \sum_{i=1}^{n} log(y_i) + nlog(\nu/\bar{y}) + n - \frac{1}{\bar{y}}n\bar{y} - n\frac{\Gamma'(\nu)}{\Gamma(\nu)} \qquad \text{subsititution}$$

$$= \sum_{i=1}^{n} log(y_i) + nlog(\nu/\bar{y}) - n\frac{\Gamma'(\nu)}{\Gamma(\nu)}$$

Consider its second derivative with respect to ν

$$\frac{\partial^2 l(\nu,\mu)}{\partial \nu^2} = n\frac{1}{\nu} - n\psi'(\nu)$$

```
psiFn = function(nu){sum(log(y))+n*log(nu/mu)-n*psigamma(nu,deriv = 0)}
psiPrimeFn = function(nu){return(y-y+n*(1/nu)-n*psigamma(nu,deriv = 1))}
curve(expr = psiFn,from = 0,to = 5)
abline(h=0, col = "navy")
```
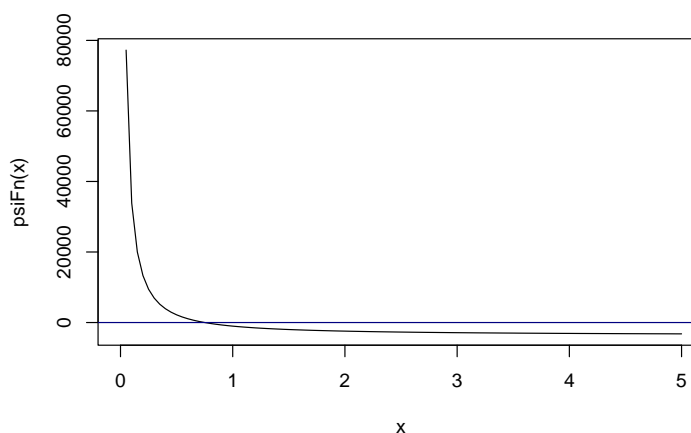


```
newton.raphson_2 = function(nu,
                            maxIteration, # max number of iterations
                            tolerance# parameters for the test
){
    ## initialize
    i = 0
    error = 1
    ## loop
    while(error > tolerance & i <= maxIteration){
    nuOld = nu

    nu = nu - psiFn(nu)/psiPrimeFn(nu)
```

```
    ## update iteration
    error = abs(sum(nuOld-nu))
    i = i+1
  }
## return last value and whether converged or not
list(nu = nu,iteration = i)
}
```

```
psiFn = function(nu){return(sum(log(y))+n*log(nu/mean(y))-n*psigamma(nu,deriv = 0))}
psiPrimeFn = function(nu){return(n*(1/nu)-n*psigamma(nu,deriv = 1))}
newton.raphson_2(0.05,20,1E-6)
```

```
## $nu
## [1] 0.7501495
##
## $iteration
## [1] 9
```

(b) Plot the fitted $G(\hat{\mu}, \hat{\nu})$ density, superimposed on a histogram of claim size.

```
#G(alpha, beta)
alpha = 0.7501495 # alpha = nu
beta = 0.75/2014 # beta = alpha/mu
n = length(dataNew$claimcst0)
series = rgamma(n,shape=alpha,rate=beta)
library(tidyverse)
y = dataNew$claimcst0
y = as.data.frame(y)
ggplot(data = y) +
  geom_histogram(aes(x=y, y=..density..)) +
  geom_line(aes(x=y, y=dgamma(y,shape = alpha, rate = beta), color="red")) +
  theme_classic()
```


21
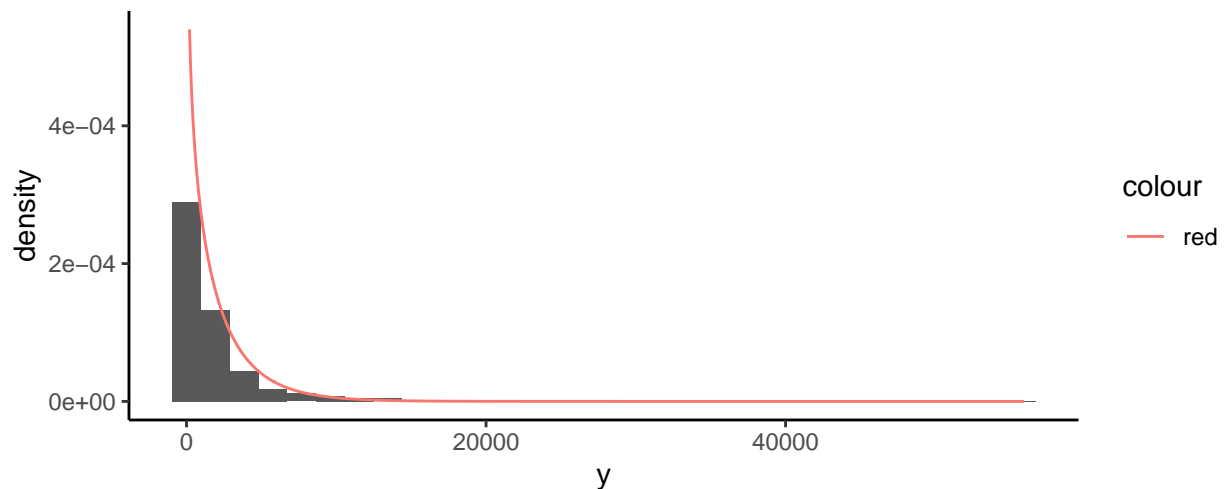
Problem 8.

In an experiment conducted to investigate the amount of drug retained in the liver of a rat, see Table 2, nineteen rats were randomly selected, weighed, and placed under light anesthetic and given an oral dose of the drug. Because large livers were thought to absorb more of a given dose than a small liver, the dose was approximately deter- mined as 40mg of the drug per kg of body weight. After a fixed length of time, each rat was sacrificed, the liver weighed, and the percentage dose in the liver y determined.

(a) Plot DoseInLiver against each explanatory variable, and identify important features to be modeled. From the pairwise scatter plot shown below, I would say that none of the scatter plot show a linear relation between the explanatory variable and the response variable.

```
BodyWt = c(176,176,190,176,200,167,188,195,176,165,158,148,149,163,170,186,146,181,149)
LiverWt = c(6.5,9.5,9.0,8.9,7.2,8.9,8.0,10.0,8.0,7.9,6.9,7.3,5.2,8.4,7.2,6.8,7.3,9.0,6.4)
Dose = c(0.88,0.88,1.00,0.88,1.00,0.83,0.94,0.98,0.88,0.84,0.80,0.74,0.75,0.81,0.85,0.94,
         0.73,0.90,0.75)
DoseInLiver = c(0.42,0.25,0.56,0.23,0.23,0.32,0.37,0.41,0.33,0.38,0.27,0.36,0.21,0.28,0.34,
               0.28,0.30,0.37,0.46)
par(mfrow=c(1,3))
plot(BodyWt, DoseInLiver,ylab = "DoseInLiver", xlab = "BodyWt")
plot(LiverWt, DoseInLiver,ylab = "DoseInLiver", xlab = "LiverWt")
plot(Dose, DoseInLiver,ylab = "DoseInLiver", xlab = "Dose")
```
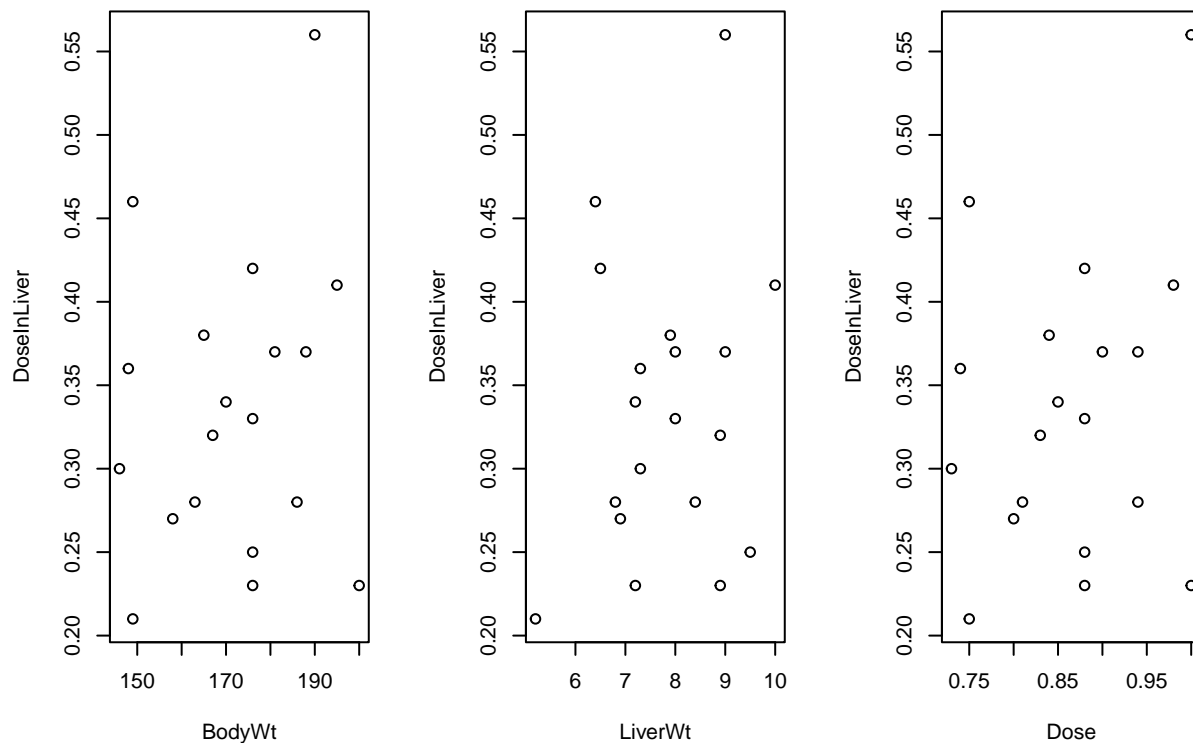

(b) Fit a linear regression model with systematic component DoseInLiver   BodyWt + LiverWt + Dose.

Fitted Model: $DoseInLiver_i = 0.265922 - 0.021246 BodyWt_i + 0.014298 LiverWt_i + 4.178111 Dose_i$

```
model = lm(DoseInLiver~BodyWt+LiverWt+Dose)
summary(model)
```

```
##
## Call:
## lm(formula = DoseInLiver ~ BodyWt + LiverWt + Dose)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.100557 -0.063233  0.007131  0.045971  0.134691
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.265922   0.194585   1.367   0.1919
## BodyWt      -0.021246   0.007974  -2.664   0.0177 *
## LiverWt      0.014298   0.017217   0.830   0.4193
## Dose         4.178111   1.522625   2.744   0.0151 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07729 on 15 degrees of freedom
## Multiple R-squared:  0.3639, Adjusted R-squared:  0.2367
## F-statistic:  2.86 on 3 and 15 DF,  p-value: 0.07197
```

(c) Using t-tests, show that `BodyWt` and `Dose` are significant for modelling DoseInLiver.

Consider the model: $DoseInLiver_i = \beta_0 + \beta_1 BodyWt_i + \beta_2 LiverWt_i + \beta_3 Dose_i + \epsilon_i$

$H_0 : \beta_1 = 0$ and $H_a : \beta_1 \neq 0$

$t = \frac{\hat{\beta_1}}{se(\hat{\beta_1})} = \frac{-0.021246}{0.007974} = -2.664 \sim t_{15}$

p-value$= P(|T| > |t|) = 2 \times P(T > |t|) = 0.0177$

Thus, we reject H_0 and we conclude that $BodyWt$ is statistically significant @ 95% significant level.

$H_0 : \beta_2 = 0$ and $H_a : \beta_2 \neq 0$

$t = \frac{\hat{\beta_2}}{se(\hat{\beta_2})} = \frac{0.014298}{0.017217} = 0.830 \sim t_{15}$

p-value$= P(|T| > |t|) = 2 \times P(T > |t|) = 0.4193$

Thus, we do not reject H_0 and we conclude that $LiverWt$ is NOT statistically significant @ 95% significant level.

$H_0 : \beta_3 = 0$ and $H_a : \beta_3 \neq 0$

$t = \frac{\hat{\beta_3}}{se(\hat{\beta_3})} = \frac{4.178111}{1.522625} = 2.744 \sim t_{15}$

p-value$= P(|T| > |t|) = 2 \times P(T > |t|) = 0.0151$

Thus, we reject H_0 and we conclude that $Dose$ is statistically significant @ 95% significant level.

We have determined that $BodyWt$ and $Dose$ are significant for modelling $DoesInLiver$

(d) In the study, the dose was determined as an approximate function of body weight, hence both variables `BodyWt` and `Dose` measure almost the same physical quantity. Why should both covariates be necessary in the model? By computing the appropriate statistics, show that Observation 3 has high leverage and is influential. [hint: one can use R function *influence.measures()* for this.]

Indeed we should not include BodyWt and Dose in the model together because they have a strong multicollinearity issue. Also, by computing the Cook's Distance (an overall measure of the influence)

23

and plot the Residuals vs Leverage Plots, it is so clear to see that Observation 3 has high leverage and is influential.

```
influence.measures(model)
```

```
## Influence measures of
##   lm(formula = DoseInLiver ~ BodyWt + LiverWt + Dose) :
##
##        dfb.1_  dfb.BdyW  dfb.LvrW dfb.Dose    dffit  cov.r   cook.d     hat inf
## 1   -0.03835   0.31492 -0.704363 -0.24375   0.8920 0.631 1.69e-01 0.1780
## 2    0.14256  -0.09774 -0.481778  0.12561  -0.6088 1.016 8.85e-02 0.1793
## 3   -0.23100  -1.66770  0.304572  1.74720   1.9048 7.401 9.30e-01 0.8509    *
## 4    0.12503  -0.12686 -0.303651  0.14009  -0.4944 0.860 5.72e-02 0.1076
## 5    0.52161  -0.39627  0.550016  0.27474  -0.9095 1.524 2.03e-01 0.3915
## 6    0.00229   0.01360  0.029003 -0.01715   0.0427 1.567 4.87e-04 0.1612
## 7   -0.18376   0.15044 -0.083355 -0.11839   0.3096 1.289 2.46e-02 0.1369
## 8   -0.29725   0.05936  0.246500 -0.04042   0.4262 1.520 4.69e-02 0.2537
## 9   -0.00968   0.01791  0.000168 -0.01673   0.0427 1.402 4.88e-04 0.0670
## 10 -0.00566   0.00993 -0.003365 -0.00929  -0.0140 1.496 5.23e-05 0.1197
## 11 -0.29053   0.19381  0.100742 -0.17288  -0.4104 1.066 4.14e-02 0.1195
## 12  0.21742  -0.02526  0.051721 -0.00920   0.2689 1.444 1.89e-02 0.1724
## 13 -0.77232   0.14391  0.766461 -0.12005  -1.0989 0.972 2.73e-01 0.3162
## 14 -0.03482  -0.04618 -0.076534  0.05990  -0.1423 1.461 5.37e-03 0.1314
## 15  0.01868   0.04063 -0.054603 -0.03817   0.1188 1.359 3.73e-03 0.0762
## 16  0.12309  -0.00557  0.328720 -0.04910  -0.4475 1.375 5.10e-02 0.2166
## 17 -0.10359   0.01454 -0.027611  0.00215  -0.1262 1.607 4.25e-03 0.1952
## 18 -0.15423   0.19035  0.162330 -0.18929   0.3522 1.270 3.16e-02 0.1487
## 19  0.85580  -0.25010 -0.294617  0.17123   0.9952 0.517 2.00e-01 0.1780
```

```
# Residuals vs Leverage Plot
plot(model, 5)
```
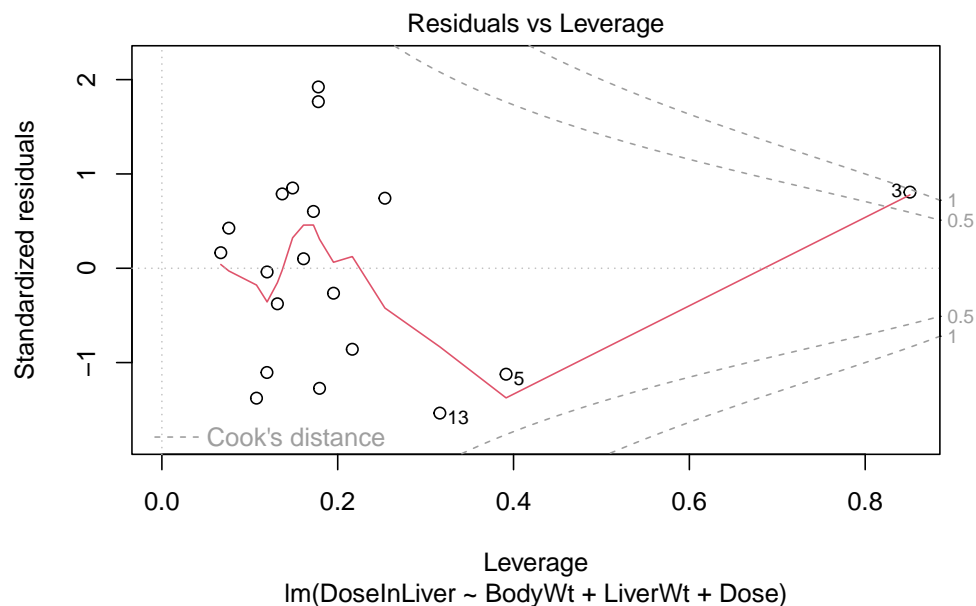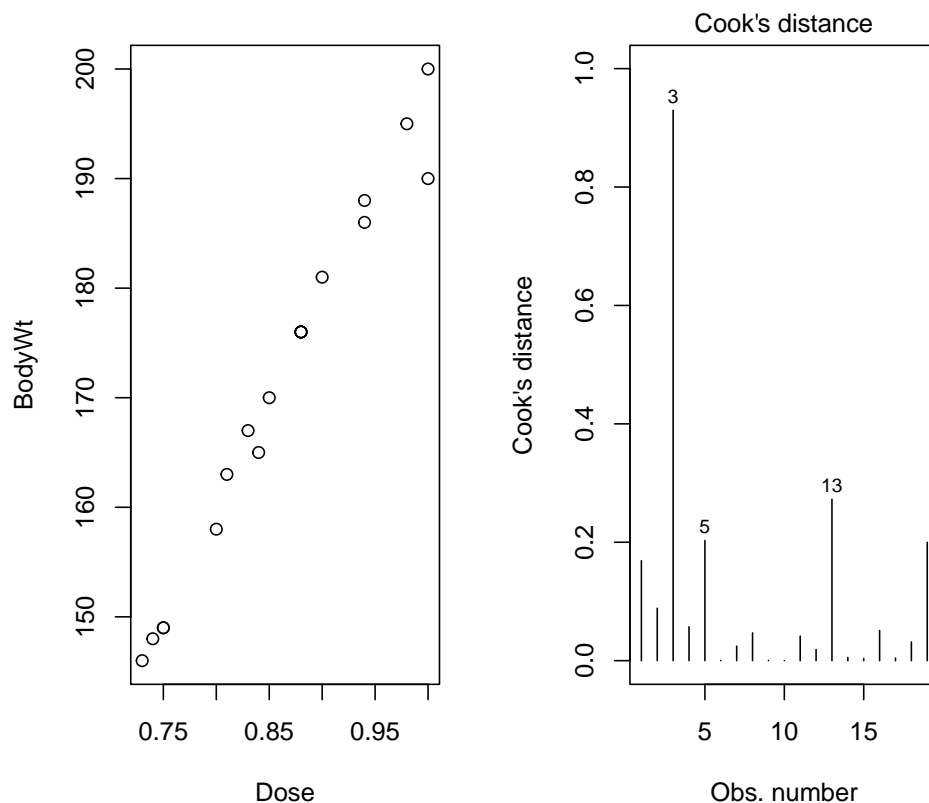
(e) Plot `BodyWt` against `Dose`, and identify Observation 3 to see the problem.

We can clearly see that Obs.3 has a high Cook's Distance Value from the plot below

```r
par(mfrow=c(1,2))
# Scatterplot
plot(Dose,BodyWt,xlab = "Dose", ylab = "BodyWt")
# Cook's distance
plot(model, 4)
```


(f) Fit the same linear regression model, after omitting Observation 3. Use t-tests to show that none of the covariates are now statistically significant.

Use t-test and R code shown below, it is clear to see that none of the covariates are now statistically significant.

```r
data =  as.data.frame(cbind(BodyWt,LiverWt,Dose,DoseInLiver))
dataNew = data[-3,]
model2 = lm(dataNew$DoseInLiver~dataNew$BodyWt+dataNew$LiverWt+dataNew$Dose)
summary(model2)
```

```
##
## Call:
## lm(formula = dataNew$DoseInLiver ~ dataNew$BodyWt + dataNew$LiverWt +
```

```
##      dataNew$Dose)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.102154 -0.056486  0.002838  0.046519  0.137059
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.311427   0.205094   1.518    0.151
## dataNew$BodyWt  -0.007783   0.018717  -0.416    0.684
## dataNew$LiverWt  0.008989   0.018659   0.482    0.637
## dataNew$Dose     1.484877   3.713064   0.400    0.695
##
## Residual standard error: 0.07825 on 14 degrees of freedom
## Multiple R-squared:  0.02106,    Adjusted R-squared:   -0.1887
## F-statistic: 0.1004 on 3 and 14 DF,  p-value: 0.9585
```

Table 2: Drug doses retained in the liver of rats. See the text for an explanation of the data. `BodyWt` is the body weight of each rat (in g); `LiverWt` is liver weight (in g); `Dose` is the dose relative to largest dose; `DoseInLiver` is the proportion of the dose in liver, as percentage of liver weight

	BodyWt	LiverWt	Dose	DoseInLiver
1	176	6.5	0.88	0.42
2	176	9.5	0.88	0.25
3	190	9.0	1.00	0.56
4	176	8.9	0.88	0.23
5	200	7.2	1.00	0.23
6	167	8.9	0.83	0.32
7	188	8.0	0.94	0.37
8	195	10.0	0.98	0.41
9	176	8.0	0.88	0.33
10	165	7.9	0.84	0.38
11	158	6.9	0.80	0.27
12	148	7.3	0.74	0.36
13	149	5.2	0.75	0.21
14	163	8.4	0.81	0.28
15	170	7.2	0.85	0.34
16	186	6.8	0.94	0.28
17	146	7.3	0.73	0.30
18	181	9.0	0.90	0.37
19	149	6.4	0.75	0.46

Problem 9.

By expanding $\sum_{i=1}^{n}(y_i - z_i^{\top}\beta)^2 = (Y - Z\beta)^{\top}(Y - Z\beta)$ and using vector differentiation and matrix calculus, derive the least squares estimate of β, where $Y = (y_1, \ldots, y_n)^{\top}$ and $Z^{\top} = (z_1, \ldots, z_n)$, $z_i = (z_{i1}, \ldots, z_{ip})^{\top}$, $\beta = (\beta_1, \ldots, \beta_p)^{\top}$.

Recall for a n by n matrix A, row vector $a = (a_1, ..., a_n)$ and column vector $b = (b_1, ..., b_n)^T$ and $\beta = (\beta_1, ..., \beta_n)^T$, we have the following vector differentiation and matrix calculus properties:

$$\frac{\partial}{\partial \beta}(a\beta) = a^T$$

$$\frac{\partial}{\partial \beta}(\beta^T b) = b$$

$$\frac{\partial}{\partial \beta}(\beta^T A \beta) = (A + A^T)\beta$$

Take derivative of the given expression with respect to the parameter vector β

$$
\begin{aligned}
\frac{\partial}{\partial \beta}[(Y - Z\beta)^T(Y - Z\beta)] &= \frac{\partial}{\partial \beta}[(Y^T - \beta^T Z^T)(Y - Z\beta)] \\
&= \frac{\partial}{\partial \beta}[Y^T Y - Y^T Z\beta - \beta^T Z^T Y + \beta^T Z^T Z\beta] \\
&= -(Y^T Z)^T - Z^T Y + [Z^T Z + (Z^T Z)^T]\beta \qquad \text{by the properties listed above} \\
&= 2(Z^T Z\beta - Z^T Y) \qquad\qquad\qquad \text{set it to equal to zero and solve it} \\
2(Z^T Z\beta - Z^T Y) &= 0 \\
Z^T Z\beta &= Z^T Y \\
\hat{\beta} &= (Z^T Z)^{-1} Z^T Y \\
&= \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}
\end{aligned}
$$