Chapter 5

5.1:

$$\text{y} \sim \text{Bin}(n, \pi)$$
$$E(y) = n\pi$$
$$\text{Var}(y) = n\pi(1-\pi)$$

(a)

$$f(y;\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y} \text{ (for one observation)}$$
$$l(\pi) = y\ln(\pi) + (n-y)\ln(1-\pi) + \ln\left[\binom{n}{y}\right]$$
$$U = U(\pi) = \frac{\partial l}{\partial \pi} = \frac{y}{\pi} - \frac{n-y}{1-\pi} = \frac{y(1-\pi)-(n-y)\pi}{\pi(1-\pi)} = \frac{y - y\pi - n\pi + y\pi}{\pi(1-\pi)} = \frac{y-n\pi}{\pi(1-\pi)}$$

- The MLE of π is $\hat{\pi} = \frac{y}{n}$

Note that: $E(U) = E\left[\frac{y-n\pi}{\pi(1-\pi)}\right] = \frac{E(y)-n\pi}{\pi(1-\pi)} = \frac{n\pi-n\pi}{\pi(1-\pi)} = 0$

- Information:

$$J = \text{Var}(U) = \text{Var}\left[\frac{4-n\pi}{\pi(1-\pi)}\right] = \frac{\text{Var}(4)}{\pi^2(1-\pi)^2} = \frac{n\pi(1-\pi)}{\pi^2(1-\pi)^2} = \frac{n}{\pi(1-\pi)}$$

- Now:

$$(\hat{\pi}-\pi)^t J(\hat{\pi}-\pi) = \left(\frac{y}{n}-\pi\right)^2\left[\frac{n}{\pi(1-\pi)}\right] = \left(\frac{y-n\pi}{n}\right)^2\left(\frac{n}{\pi(1-\pi)}\right) = \frac{(y-n\pi)^2}{n\pi(1-\pi)} = \left[\frac{y-n\pi}{\sqrt{n\pi(1-\pi)}}\right]^2$$

Since $\text{y} \sim \text{Bin}(n, \pi)$, we have:

$$Z = \frac{y-E(y)}{\sqrt{\text{Var}(y)}} = \frac{y-n\pi}{\sqrt{n\pi(1-\pi)}} \approx N(0,1) \text{ (approximately)}$$

(Normal approximation to binomial)

$$Z^2 = \left[\frac{y-n\pi}{\sqrt{n\pi(1-\pi)}}\right]^2 \approx X^2_{(1)} \text{ (approximately)}$$

(b) Wald statistic is: $z^2 = \left[\frac{y-n\pi}{\sqrt{n\pi(1-\pi)}}\right]^2 = \frac{(y-n\pi)^2}{n\pi(1-\pi)} \to (*)$

and

$$U^t J^{-1} U = \left[\frac{y-n\pi}{\pi(1-\pi)}\right]^2 \frac{1}{\frac{n}{\pi(1-\pi)}} = \frac{(y-n\pi)^2}{\pi^2(1-\pi)^2}\frac{\pi(1-\pi)}{n} = \frac{(y-n\pi)^2}{n\pi(1-\pi)} \to (**)$$

$$(*) = (**)$$
$$\therefore \text{ Wald statistic } = U^t J^{-1} u$$

(c)

- $l(\pi) = y\ln(\pi) + (n-y)\ln(1-\pi) + \ln\left[\binom{n}{y}\right]$
- For the saturated model, we have:
  - MLE of π is $\hat{\pi} = \frac{y}{n}$
  - The fitted value $\hat{y} = n\hat{\pi} = n\left(\frac{y}{n}\right) = y$
  - The maximum value of $l(\pi)$ for the Saturated model is:

$$l(b_{\max}; y) = l(\hat{\pi}) = y\ln(\hat{\pi}) + (n-y)\ln(1-\hat{\pi}) + \ln\left[\binom{n}{y}\right] = y\ln\left(\frac{y}{n}\right) + (n-y)\ln\left(1-\frac{y}{n}\right) + \ln\left[\binom{n}{y}\right]$$

- For the model of interest:
  - Let $\tilde{\pi}$ be the MLE of π, and $\tilde{y} = n\tilde{\pi}$ be the fitted value.
  - The maximum value of $l(\pi)$ for the model of interest is:

$$l(b; y) = l(\tilde{\pi}) = y\ln(\tilde{\pi}) + (n-y)\ln(1-\tilde{\pi}) + \ln\left[\binom{n}{y}\right]$$
$$= y\ln\left(\frac{\tilde{y}}{n}\right) + (n-y)\ln\left(1-\frac{\tilde{y}}{n}\right) + \ln\left[\binom{n}{y}\right]$$

- The deviance is:

$$D = 2[l(b_{max}) - l(b)]$$
$$= 2\left\{\left[y\ln\left(\frac{y}{n}\right) + (n-y)\ln\left(1-\frac{y}{n}\right) + \ln\left[\binom{n}{y}\right]\right] - \left[y\ln\left(\frac{\tilde{y}}{n}\right) + (n-y)\ln\left(1-\frac{\tilde{y}}{n}\right) + \ln\left[\binom{n}{y}\right]\right]\right\}$$
$$= 2\left\{y\ln\left(\frac{y}{\tilde{y}}\right) + (n-y)\ln\left(\frac{1-\frac{y}{n}}{1-\frac{\tilde{y}}{n}}\right)\right] = 2\left[y\ln\left(\frac{y}{\tilde{y}}\right) + (n-y)\ln\left(\frac{n-y}{n-\tilde{y}}\right)\right]$$

(d)

Wald statistic $= (\hat{\pi}-\pi)t^2J(\pi-\pi) : \dfrac{(y-n\pi)^2}{n\pi(1-\pi)} \approx \chi^2_{(1)}$

Score statistic: $U^tJ^{-1}U \approx \chi^2_{(1)}$

Deviance: $D = 2\left[y\ln\left(\frac{y}{\tilde{y}}\right) + (n-y)\ln\left(\frac{n-y}{n-\tilde{y}}\right)\right]$

- We have shown that:

$$\text{Wald Statistic} = \text{Score Statistic} = \frac{(y-n\pi)^2}{n\pi(1-\pi)}$$
$$n = 10, \quad y = 3, \quad \chi^2_{0.05(1)} = 3.84$$

(i) For the model with $\pi = \pi_0 = 0.1$: $\begin{Bmatrix} H_0: \pi = 0.1 \\ H_1: \pi \neq 0.1 \end{Bmatrix}$

$\tilde{y} = n\pi_0 = (10)(0.1) = 1$

- Wald statistic $= \frac{(y-n\pi_0)^2}{n\pi_0(1-\pi_0)} = \frac{(3-1)^2}{(10)(0.1)(0.9)} = 4.444$
  Since $4.444 > X^2_{0.05}(1)$, we reject $H_0: T = 0.1$
- wald statistic = Score statistic = 4.444 we will have the same calculations and the same conclusion. Therefore, the model is not adequate bused on There two statistics.
- Deviance or log-likelihood statistic

$$D = 2\left[y\ln\left(\frac{y}{n\pi_0}\right) + (n-y)\ln\left(\frac{n-y}{n-n\pi_0}\right)\right] = 2\left[(3)\ln\left(\frac{3}{1}\right) + (10-3)\ln\left(\frac{10-3}{10-1}\right)\right] = 3.07327$$

Since $3.07327 < X^2_{0.05(1)}$, we donot reject to based on the deviance statistic, and we conclude that the model is a adequate.

We notice that their statistics do not lead to the same conclusions.

(ii) Both statistics equal zero and would not suggest rejecting $\pi = 0.3$.

(iii) Wald/score statistic $= 1.60$, log-likelihood statistic $= 1.65$, so neither would suggest rejecting $\pi = 0.5$.

5.2:

$y_1, y_2, \dots, y_N$ are independent.
$y_i \sim \text{Exp}(\theta_i)$ for all $i = 1,2,\dots,N$

$$f(y_i,\theta_i) = \theta_i e^{-\theta_i y_i}, \qquad y_i > 0 \text{ and } \theta_i > 0$$

- For maximal model, the θ_i' 's ane different.
- For the model of interest, the θ_i' s are the same

$$\theta_1 = \theta_2 = \cdots = \theta_N = \theta.$$

(1) For the maximal model:
The log-likelihood function is:

$$l = l(\theta_1,\dots,\theta_N) = \sum_{i=1}^{N} \ln\left(\theta_i e^{-\theta_i y_i}\right) = \sum_{i=1}^{N} \ln(\theta_i) - \sum_{i=1}^{N} \theta_i y_i$$

$$\frac{\partial l}{\partial \theta_i} = \frac{\partial l(\theta_1,\dots,\theta_N)}{\partial \theta_i} = \frac{1}{\theta_i} - y_i$$

$$\frac{\partial l}{\partial \theta_i} \overset{\text{set}}{=} 0 \Leftrightarrow \frac{1}{\theta_i} - y_i = 0 \Leftrightarrow \hat{\theta}_i = \frac{1}{y_i}$$

The MLE of θ_i is $\hat{\theta}_i = \frac{1}{y_i}$ $\left(\text{Note:} \frac{\partial^2 l}{\partial \theta_i^2} = -\frac{1}{\theta_i^2} < 0\right)$.
The maximum value of $l(\theta_1,\dots,\theta_N)$ is:

$$l(\underset{\sim}{b}max) = l\left(\hat{\theta}_1,\dots,\hat{\theta}_N\right) = \sum_{i=1}^{N} \ln\left(\frac{1}{y_i}\right) - \sum_{i=1}^{N} \frac{1}{y_i}(y_i) = \sum_{i=1}^{N} \ln\left(\frac{1}{y_i}\right) - N$$

(2) For the model of internet $(\theta_1 = \theta_2 = \cdots = \theta_N = \theta)$:
The log-likelihood function is:

$$l = l(\theta) = \sum_{i=1}^{N} \ln\left(\theta e^{-\theta y_i}\right) = N\ln(\theta) - \theta \sum_{i=1}^{N} y_i$$

$$\frac{\partial l}{\partial \theta} = \frac{N}{\theta} - \sum_{i=1}^{N} y_i$$

$$\frac{\partial l}{\partial \theta} \overset{\text{set}}{=} 0 \Leftrightarrow \frac{N}{\theta} - \sum_{i=1}^{N} y_i = 0 \Leftrightarrow \hat{\theta} = \frac{N}{\sum_{i=1}^{N} y_i} = \frac{1}{\bar{y}}$$

The MLE of θ is $\hat{\theta} = \frac{1}{\bar{y}}$ $\left(note\ that\ \frac{\partial l^2}{\partial \theta^2} = -\frac{N}{\theta^2} < 0\right)$
The maximum value of $l(\theta)$ is

$$l(\underset{\sim}{b}\quad) = l\left(\hat{\theta}\right) = N\ln\left(\hat{\theta}\right) - \hat{\theta}\sum y_i = N\ln\left(\frac{1}{\bar{y}}\right) - \frac{1}{\bar{y}}\sum_{i=1}^{N} y_i = N\ln\left(\frac{1}{\bar{y}}\right) - N, \left(\sum y_i = N\bar{Y}\right)$$

For (1) and (2):
The deviance is

$$D = 2[l(\underset{\sim}{b}max) - l(\underset{\sim}{b}\quad)] = 2\left\{\left[\sum_{i=1}^{N} \ln\left(\frac{1}{y_i}\right) - N\right] - \left[N\ln\left(\frac{1}{\bar{y}}\right) - N\right]\right\} = 2\left[\sum_{i=1}^{N} \ln\left(\frac{1}{y_i}\right) - N\ln\left(\frac{1}{\bar{y}}\right)\right]$$

$$= 2\left[\sum_{i=1}^{N} \ln\left(\frac{1}{y_i}\right) - \sum_{i=1}^{N} \ln\left(\frac{1}{\bar{y}}\right)\right] = 2\sum_{i=1}^{N}\left[\ln\left(\frac{1}{y_i}\right) - \ln\left(\frac{1}{\bar{y}}\right)\right] = 2\sum_{i=1}^{N} \ln\left[\frac{1/y_i}{1/\bar{y}}\right] = 2\sum_{i=1}^{N} \ln\left(\frac{\bar{y}}{y_i}\right)$$

5.3:

$y_1, y_2, \dots, y_N$ are independent.
$y_i \sim \text{Pareto}(\theta)$ for all $i = 1,2,\dots,N$

$$f(y_i;\theta) = \frac{\theta}{y_i^{\theta+1}} = \theta y_i^{-\theta-1}\,; y_i > 1 \text{ and } \theta > 0$$

(a) The log-likelihood function is:

$$l = l(\theta) = \sum_{i=1}^{N} \ln\left(\theta y_i^{-\theta-1}\right) = N\ln(\theta) - (\theta+1)\sum_{i=1}^{N} \ln(y_i)$$

$$\frac{\partial l}{\partial \theta} = \frac{N}{\theta} - \sum \ln(y_i)$$

$$\frac{\partial l}{\partial \theta} \overset{\text{set}}{=} 0 \Leftrightarrow \frac{N}{\theta} - \sum \ln(y_i) = 0 \Leftrightarrow \hat{\theta} = \frac{N}{\sum \ln(y_i)}$$

$$\text{The MLE of } \theta \text{ is } \hat{\theta} = \frac{N}{\sum \ln(y_i)}\,, \left(\text{note that } \frac{\partial^2 l}{\partial \theta^2} = -\frac{N}{\theta^2} < 0\right)$$

(b) Recall From Exercise 3.11:

- Score statistic is: $U = \frac{N}{\theta} - \sum_{i=1}^{N} \ln(y_i)$
- The information is: $J = \text{Var}(U) = \frac{N}{\theta^2}$

The variance of $\hat{\theta} = \frac{N}{\sum \ln(y_i)}$ is:

$$\text{Var}(\hat{\theta}) = \mathcal{J}^{-1} = \frac{\theta^2}{N}$$

$$\hat{\theta} \approx N(\theta, \mathcal{T}^{-1}) = N\left(\theta, \frac{\theta^2}{N}\right) \text{ (asymptotically)}$$

Wald Statistic is:

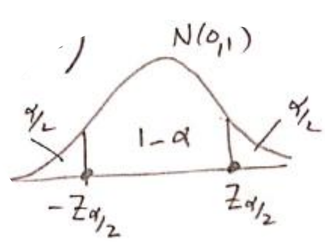
$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{\text{Var}(\hat{\theta})}} = \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta^2}{N}}} \doteq N(0,1) \text{ (approximately)}$$

$$Z^2 = \frac{(\hat{\theta} - \theta)^2}{\theta^2/N} = \frac{N(\hat{\theta} - \theta)^2}{\theta^2} = N\left(\frac{\hat{\theta} - \theta}{\theta}\right)^2 \approx \chi^2_{(1)}$$

(c)

<u>1st Method:</u>

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta^2}{N}}} = \frac{\sqrt{N}(\hat{\theta} - \theta)}{\theta} = \sqrt{N}\left(\frac{\hat{\theta}}{\theta} - 1\right)$$



$$1-\alpha = P\left[-Z_{\alpha/2} < Z < Z_{\alpha/2}\right] = P\left[-Z_{\alpha/2} < \sqrt{N}\left(\frac{\hat{\theta}}{\theta}-1\right) < Z_{\alpha/2}\right] = P\left[-Z_{\alpha/2}\frac{1}{\sqrt{N}} < \frac{\hat{\theta}}{\theta}-1 < Z_{\alpha/2}\frac{1}{\sqrt{N}}\right]$$

$$= P\left[1-Z_{\alpha/2}\frac{1}{\sqrt{N}} < \frac{\hat{\theta}}{\theta} < 1+Z_{\alpha/2}\frac{1}{\sqrt{N}}\right] = P\left[\frac{1}{1+Z_{\alpha/2}\frac{1}{\sqrt{N}}} < \frac{\theta}{\hat{\theta}} < \frac{1}{1-Z_{\alpha/2}\frac{1}{\sqrt{N}}}\right]$$

$$= P\left[\frac{\hat{\theta}}{1+\frac{Z_{\alpha/2}}{\sqrt{N}}} < \theta < \frac{\hat{\theta}}{1-\frac{Z_{\alpha/2}}{\sqrt{N}}}\right]$$

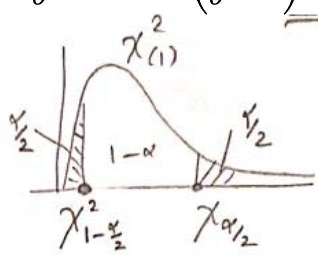
$\therefore$ $(1-\alpha)100\%$ C.I. for θ is

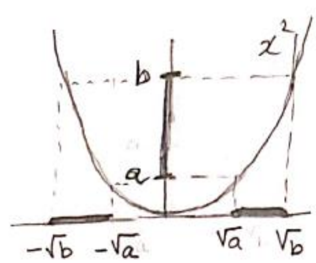
$$\frac{\hat{\theta}}{1+\frac{Z_{\frac{\alpha}{2}}}{\sqrt{N}}} < \theta < \frac{\hat{\theta}}{1-\frac{Z_{\frac{\alpha}{2}}}{\sqrt{N}}}$$

The 95% C.I. for θ is $\left(\frac{\hat{\theta}}{1+\frac{1.96}{\sqrt{N}}}, \frac{\hat{\theta}}{1-\frac{1.96}{\sqrt{N}}}\right)$.

<u>2nd Method:</u>

$$\frac{N(\hat{\theta}-\theta)^2}{\theta^2} = N\left(\frac{\hat{\theta}}{\theta}-1\right)^2 \approx \chi^2_{(1)}$$





$a < |x| < b \quad \Leftrightarrow \quad (a < x < b) \quad \text{OR} \quad (-b < x < -a).$

Visual Representation

```
<-----|-----|-----|-----|-----|-----|-----|----->
     -b    -a     0     a     b
```

- The solution consists of two intervals:
  - $x \in (a, b)$ (right side of the number line)
  - $x \in (-b, -a)$ (left side of the number line)

$$1-\alpha = P\left[\chi^2_{1-\frac{\alpha}{2}} < N\left(\frac{\hat{\theta}}{\theta}-1\right)^2 < \chi^2_{\alpha/2}\right] = P\left[\frac{\chi^2_{1-\frac{\alpha}{2}}}{N} < \left(\frac{\hat{\theta}}{\theta}-1\right)^2 < \frac{\chi^2_{\alpha/2}}{N}\right]$$

$$= P\left[\left(\sqrt{\frac{\chi^2_{1-\alpha/2}}{N}} < \frac{\hat{\theta}}{\theta}-1 < \sqrt{\frac{\chi^2_{\alpha/2}}{N}}\right) \cup \left(-\sqrt{\frac{\chi^2_{\alpha/2}}{N}} < \frac{\hat{\theta}}{\theta}-1 < -\sqrt{\frac{\chi^2_{1-\alpha/2}}{N}}\right)\right]$$

$$= P\left\{\left[\frac{1}{\hat{\theta}}\left(1+\sqrt{\frac{\chi^2_{1-\alpha/2}}{N}}\right) < \frac{1}{\theta} < \frac{1}{\hat{\theta}}\left(1+\sqrt{\frac{\chi^2_{\alpha/2}}{N}}\right)\right] \cup \left[\frac{1}{\hat{\theta}}\left(1-\sqrt{\frac{\chi^2_{\alpha/2}}{N}}\right) < \frac{1}{\theta} < \frac{1}{\hat{\theta}}\left(1-\sqrt{\frac{\chi^2_{1-\alpha/2}}{N}}\right)\right]\right\}$$

$$= P\left\{\left(\frac{\hat{\theta}}{1+\sqrt{\frac{\chi^2\alpha/2}{N}}} < \theta < \frac{\hat{\theta}}{1+\sqrt{\frac{\chi^2_1-\alpha/2}{N}}}\right) \cup \left(\frac{\hat{\theta}}{1-\sqrt{\frac{\chi^2_{1-\alpha/2}}{N}}} < \theta < \frac{\hat{\theta}}{1-\sqrt{\frac{\chi^2_{\alpha/2}}{N}}}\right)\right\}$$

$\therefore$ $(1-\alpha)100\%$ C.I. for θ is (the region):

$$\therefore \left(\frac{\hat{\theta}}{1+\sqrt{\frac{\chi^2_{\alpha/2}}{N}}}, \frac{\hat{\theta}}{1+\sqrt{\frac{\chi^2_{1-\alpha/2}}{N}}}\right) \cup \left(\frac{\hat{\theta}}{1-\sqrt{\frac{\chi^2_{1-\alpha/2}}{N}}}, \frac{\hat{\theta}}{1-\sqrt{\frac{\chi^2_{\alpha/2}}{N}}}\right)$$

(d) $y \sim \text{pareto}(\theta)$; $f(y;\theta) = \theta y^{-\theta-1}$; $y > 1$ and $\theta > 0$

The cumulative dist. function is:

$$F(y) = \int_1^y \theta x^{-\theta-1} dx = \theta\left[\frac{x^{-\theta-1+1}}{-\theta-1+1}\right]_{x=1}^{y} = \theta\left[\frac{x^{-\theta}}{-\theta}\right]_{x=1}^{y} = \left[-x^{-\theta}\right]_{x=1}^{y} = 1 - y^{-\theta}$$

Suppose $U \sim \text{Uniform}(0,1)$.

To generate y, we first generate a random number U from $\text{Uniform}(0,1)$, and then equate, $U = F(y)$ and then solve for $y = F^{-1}(u)$

For Pareto,

$$u = 1 - y^{-\theta} \Leftrightarrow y^{-\theta} = 1 - u \Leftrightarrow \left(\frac{1}{y}\right)^{\theta} = 1 - u \Leftrightarrow \frac{1}{y} = (1-u)^{1/\theta} \Leftrightarrow y = (1-u)^{-1/\theta}$$

$$\therefore y = \left(\frac{1}{1-u}\right)^{\frac{1}{\theta}}$$

Note: $U \sim$ Unifrom (0,1) $\Leftrightarrow 1 - U \sim$ Unifrom(0,1)

Thus, we can use:

$$y = \left(\frac{1}{u}\right)^{1/\theta}$$

We will use computer to do that. (See Minitab output) For $\theta = 2$.

- For $\theta = 2$:

Generating Process: Using Minitab, we generate 100 random numbers from uniform (0,1);

$$y = \left(\frac{1}{u}\right)^{1/\theta} = \left(\frac{1}{u}\right)^{\frac{1}{2}} = \sqrt{\frac{1}{u}}.$$

u_i	y_i
0.1290	2.7842
0.4867	1.4335
⋮	⋮
0.7293	1.1710

$$\hat{\theta} = \frac{N}{\sum \ln(y_i)}, \quad N = 100, \quad \sum_{i=1}^{N} \ln(y_i) = 51.786$$

$$\Rightarrow \hat{\theta} = \frac{100}{51.786} = 1.931$$

95% C.I. for θ: $\dfrac{\hat{\theta}}{1+\frac{1.96}{\sqrt{N}}} < \theta < \dfrac{\hat{\theta}}{1-\frac{1.96}{\sqrt{N}}} \Rightarrow \dfrac{1.931}{1+\frac{1.96}{10}} < \theta < \dfrac{1.931}{1-\frac{1.96}{10}} \Rightarrow 1.7247 < \theta < 2.4017$

- We repeat this process 20 times using Minitab Macro, and found that:
  (1) $\bar{\hat{\theta}} = 1.99$
  (2) Out of 20 95% C. I.'s for θ, 19 intervals contain $\theta = 2.0$ (See Minitab Output)

5.4:
Using data in Exercise 4.2:

y_i	x_i
65	3.36
156	2.88
$\vdots$	$\vdots$
65	5

y_i = time to death (in weeks)
$x_i = \log_{10}$ (initial while bland while call)

$$\mu_i = E(y_i) = e^{\beta_1+\beta_1 x_i} = e^{\eta(x_i)} = e^{\eta_i} \text{ (This ensure that } \mu_i > 0)$$

$$\Leftrightarrow \ln(\mu_i) = \beta_1 + \beta_1 x_i = \eta(x_i) = \eta_i$$

link function

$$g(\mu_i) = \ln(\mu_i)$$

We will use the following form for Exp(θ)

$$f(y_i, \theta) = \frac{1}{\theta} e^{-y/\theta}; y > 0, \theta > 0$$

$$E(y_i) = \theta_i$$

$$\text{Var}(Y_i) = \theta_i^2$$

(a) Wald Statistic:

$$(\underset{\sim}{b} - \underset{\sim}{\beta})^t \underline{J} (\underset{\sim}{b} - \underset{\sim}{\beta}) \sim \chi^2(p)$$

From Exercises 4.2, we have:

$$b_1 = 8.4775$$

$$b_2 = -1.1093$$

$$\underline{J}(\underset{\sim}{b}) = \begin{bmatrix} 17 & 69.63 \\ 69.63 & 291.457 \end{bmatrix} = \text{Var}(\underset{\sim}{U})$$

$$\underline{J}^{-1}(\underset{\sim}{b}) = \begin{bmatrix} 2.73839 & -0.654209 \\ -0.654209 & 0.159724 \end{bmatrix} = \text{Var}(\underset{\sim}{b})$$

$$\underset{\sim}{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 8.4775 \\ -1.1093 \end{pmatrix}$$

- In dividuals C.I. for β_1 and β_2 :
  95% C. I. for β_1 is:

$$b_1 \pm 1.96\ \mathrm{se}(b_1) = 8.4775 \pm 1.96\sqrt{2.73839}$$
$$\Rightarrow 5.23408 < \beta_1 < 11.72092$$

95% C.I, for β_2 is:

$$b_2 \pm 1.96\ \mathrm{se}(b_2) = -1.1093 \pm 1.96\sqrt{0.159724}$$
$$\Rightarrow\ -1.89262 < \beta_2 < -0.32598$$

- Simultancous C.I. for β_1 and β_2 :

A $(1-\alpha)100\%$ confidence Region for $\underset{\sim}{\beta}$ is

$$\left\{\underset{\sim}{\beta} : X^2_{1-\alpha/2(2)} < \left(\underset{\sim}{b} - \underset{\sim}{\beta}\right)^t \underline{J}\left(\underset{\sim}{b} - \underset{\sim}{\beta}\right) < X_{\alpha/2(2)}\right\}$$

where:

$$\left(\underset{\sim}{b} - \underset{\sim}{\beta}\right)^t \underline{J}\left(\underset{\sim}{b} - \underset{\sim}{\beta}\right) = (8.4775-\beta_1, -1.1093-\beta_2)\begin{bmatrix} 17 & 69.63 \\ 69.63 & 291.457 \end{bmatrix} * \begin{pmatrix} 8.9775-\beta_1 \\ -1.1093-\beta_2 \end{pmatrix}$$
$$= 17(8.4775-\beta_1)^2 + 291.457(-1.1093-\beta_2)^2 + 2(69.63)(8.4775-\beta_1)(-1.1093-\beta_2)$$

(b) $\begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases} \Leftrightarrow \begin{cases} H_0: \mu_i = e^{\beta_1} \\ H_1: \mu_i = e^{\beta_1+\beta_2 x_i} \end{cases}$

(1) Model under H_0:

$$\mu_i = e^{\beta_1}, i = 1, \ldots, N \ \ (\text{Recall } \mu_i = \theta_i)$$

For this model

$$\mu_1 = \mu_2 = \cdots = \mu_N = \mu = \theta$$

The log-likelihood function is:

$$l_0 = l(\theta) = \sum_{i=1}^{N} \ln\left(\frac{1}{\theta}e^{-\frac{y_i}{\theta}}\right), \left(f(y;\theta) = \frac{1}{\theta}e^{-y/\theta}\right) = -N\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{N} y_i = -N\ln(\theta) - \frac{N}{\theta}\bar{y}$$
$$\frac{\partial l_0}{\partial \theta} = -\frac{N}{\theta} + \frac{N\bar{y}}{\theta^2}$$
$$\frac{\partial l_0}{\partial \theta} \overset{\text{set}}{=} 0 \Leftrightarrow \frac{N\bar{Y}}{\theta^2} = \frac{N}{\theta} \Leftrightarrow \frac{\bar{Y}}{\theta} = 1 \Leftrightarrow \hat{\theta} = \bar{Y}$$

The MLE of θ is $\hat{\theta} = \bar{Y}$ $\left(\frac{\partial^2 l_0}{\partial \theta^2}\Big|_{\theta=\hat{\theta}} = \cdots = -\frac{N}{(\bar{y})^2} < 0\right)$

The maximum value of the loy-litelikood function

$$\hat{l}_0 = l(\hat{\theta}) = -N\ln(\hat{\theta}) - \frac{N}{\hat{\theta}}\bar{y} = -N\ln(\bar{y}) - N\left(\frac{\bar{y}}{\bar{y}}\right) = -N[\ln(\bar{y}) + 1], (\bar{y} = 62.4706)$$
$$= -17[\ln(62.4706) + 1] = -87.289832$$

(2) Model under H_1 :

$$\mu_i = e^{\beta_1+\beta_2 x_i}, (\mu_i = \theta_i)$$

We have found that the MLE's of β_1 and β_2 are:

$$b_1 = 8.4775 \text{ and } b_2 = -1.1093$$

Therefore, the MLE of $\mu_i = \theta_i$ is

$$\hat{\theta}_i = e^{b_1+\beta_2 x_i} = e^{8.4775-1.1093x_i}$$

The log-likelihood function is:

$$l_1 = l(\theta_1, \ldots, \theta_N) = \sum_{i=1}^{N} \ln\left(\frac{1}{\theta_i}e^{-y_i/\theta_i}\right) = \sum_{i=1}^{N}\left[-\ln(\theta_i) - \frac{y_i}{\theta_i}\right] = -\sum_{i=1}^{N} \ln(\theta_i) - \sum_{i=1}^{N} \frac{y_i}{\theta_i}$$

The maximum value of the log-likelihood function is:

$$\hat{l}_1 = l(\hat{\theta}_1, \dots, \hat{\theta}_N) = -\sum_{i=1}^{N} \ln(\hat{\theta}_i) - \sum_{i=1}^{N} \frac{y_i}{\hat{\theta}_i} = -\sum_{i=1}^{N} \ln(e^{8.4775-1.1093x_i}) - \sum_{i=1}^{N} \frac{y_i}{e^{8.4775-1.1093x_i}}$$
$$= -\sum_{i=1}^{N} (8.4775 - 1.1093x_i) - \sum_{i=1}^{N} \frac{y_i}{e^{8.4775-1.1093x_i}} = -(66.8769) - (17.0001)$$
$$= -83.877 \quad \text{(calculatedfrom the data)}$$

The test statistic is:

$$\Delta D = 2[\hat{l}_1 - \hat{l}_0] = 2[-83.877 - (-87.289832)] = 6.825664$$

under H_0, $\Delta D \sim \chi^2_{(2-1)} = \chi^2_{(1)}$

$$\chi^2_{0.05,(1)} = 3.842$$

Since $\Delta D > \chi^2_{0.05,(1)}$, we reject $H_0: \beta_2 = 0$ and we conclude that x has a significant importance in predicting y.

Solution of parts a and b by using R:

```
# y:times to death .
y<-c(65,156,100,134,16,108,121,4,39,143,56,26,22,1,1,5,65)
y

##  [1]  65 156 100 134  16 108 121   4  39 143  56  26  22   1   1   5  65

#x:log10(inital white blood cell count).
x<-c(3.36,2.88,3.63,3.41,3.78,4.02,4,4.23,3.73,3.85,3.97,4.51,4.54,5,5,4.72,5)
x

##  [1] 3.36 2.88 3.63 3.41 3.78 4.02 4.00 4.23 3.73 3.85 3.97 4.51 4.54 5.00 5.00
## [16] 4.72 5.00
```

1)Hypothesis: H_0 vs H_1

$$H_0: \beta_2 = 0 \quad \gg \quad H_0: \mu_i = e^{\beta_1}$$

Vs

$$H_1: \beta_2 \neq 0 \quad \gg \quad H_1: \mu_i = e^{\beta_1 + \beta_2 * x_i}$$

a-Model under H_1 :

```
#To Test H0:Beta2 equal 0 vs H1:Beta2 not equal 0 :
# 1) find model under H0 and H1 :
#model under H1 (M1) :
m1<-glm(y~x ,family =Gamma(link = "log"))
summary(m1,dispersion =1)

##
## Call:
## glm(formula = y ~ x, family = Gamma(link = "log"))
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   8.4775     1.6548   5.123 3.01e-07 ***
## x            -1.1093     0.3997  -2.776  0.00551 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 1)
##
##     Null deviance: 26.282  on 16  degrees of freedom
## Residual deviance: 19.457  on 15  degrees of freedom
## AIC: 173.97
##
## Number of Fisher Scoring iterations: 8
```

b-Model under H_0 :

```r
#model under H0  (M0) :
m0<-glm(y~1 ,family =Gamma(link = "log"))
summary(m0,dispersion =1)
```

```
##
## Call:
## glm(formula = y ~ 1, family = Gamma(link = "log"))
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.1347     0.2425   17.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 1)
##
##     Null deviance: 26.282  on 16  degrees of freedom
## Residual deviance: 26.282  on 16  degrees of freedom
## AIC: 178.09
##
## Number of Fisher Scoring iterations: 6
```

2)Find test statistics: $D_0 \geq D_1$

$$\Delta D = D_0 - D_1 \sim \chi^2_{p-q}$$

D0=Deviance of model M0
D1= Deviance of model M1
p= number of parameters in model M_1 =2
q= number of parameters in model M_0 =1

```r
# 2) Find test statistics
# we can find test statistics by useing Deviance (delta(D)= D0-D1)
D0<- deviance(m0)
D0
```

```
## [1] 26.2821
```

```r
D1<- deviance(m1)
D1
```

```
## [1] 19.45653
```

```r
# or by code : anova(m0 ,m1 , test = "Chisq")
delta_D <- D0-D1
delta_D
```

```
## [1] 6.825567
```

$$\Delta D = D_0 - D_1 = 26.2821 - 19.4563 = 6.825567$$

3)To find table value:

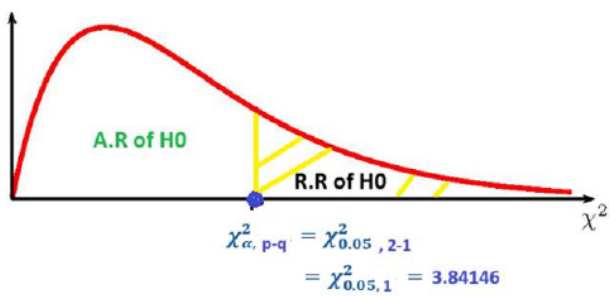
use the chi-square table and the degrees of freedom (df) are calculated as p-q.

```r
# critical value "from chi table " df=p-q =2-1=1
# p= number of parameter in model M1=2 ,, q= number of parameter in model M0=1
df_diff<- m0$df.residual - m1$df.residual
df_diff

## [1] 1

critical_value<- qchisq(1-0.05 , df_diff)
critical_value

## [1] 3.841459
```



4)Decision:

Since $\Delta D > \chi^2_{0.05,1}$, $We\ Reject\ H_0{:}\ \beta_2 = 0$., and we conclude that x has a significant importance in predicting

```r
#Decision:
if (delta_D >critical_value) {
  print("Reject H0")
}else{
  print("Do not Reject H0")
}

## [1] "Reject H0"
```

##

Hypothesis Testing Steps for Model Parameters:

1- **Hypothesis:**

$$\mathbf{H_0{:}\ \beta_i = 0 \quad Vs \quad H_1{:}\ \beta_i \neq 0}$$

2- **Test statistics:**

$$\Delta \mathbf{D} = \mathbf{D_0} - \mathbf{D_1} \sim \chi^2_{\mathbf{p-q}}$$

$$= \mathbf{2}\ [\ \ell(\mathbf{b_1}; \boldsymbol{y}) - \ell(\boldsymbol{b_0}; \boldsymbol{y})]$$

D_0 = Deviance of model M0 **and** D_1 = Deviance of model M1
($D_0 \geq D_1$)
b_0 is the MLE of β_0 for model M_0 **and** b_1 is the MLE of β_1 for model M_1
$\ell(\mathbf{b_1}; \boldsymbol{y}) \geq \ell(\mathbf{b_0}; \boldsymbol{y})$

3- **Critical value**

$$\chi^2_{\mathbf{p-q}}$$

p= number of parameters in model M_1 **and** q= number of parameters in model M_0

4- **Decision:**

If $\Delta D > \chi^2_{\alpha,p-q}$, $We\ Reject\ H_0$
If $\Delta D \leq \chi^2_{\alpha,p-q}$, $We\ do\ not\ \ Reject\ H_0$, and choose model M0