Stat 431 Assignment 1 Spring 2021

Due by 4:00pm EDT on Friday May 28, 2021 via Crowdmark

Question 1 [10 marks] Adapted from Problems 6.9-6.14 of Kutner et al. (2005)

A large, national grocery retailer tracks productivity and costs of its facilities closely. Data were collected from a single distribution center for a one-year period. Each observation represents one week of activity. The variables included are the number of cases shipped (X_1), the indirect costs of the total labour hours as a percentage (X_2), a qualitative predictor called holiday that is coded 1 if the week has a holiday and 0 otherwise (X_3), and the total labour hours (Y).

The data are available as the `GroceryRetailer` data set in the `ALSM` R library. The code below will load the library and the data and display the internal structure of the data object.

```
> # install.packages('ALSM') # Should only need to run once, comment out
> # afterwards
> library(ALSM)
Warning: package 'ALSM' was built under R version 4.0.5
Warning: package 'leaps' was built under R version 4.0.5
Warning: package 'SuppDists' was built under R version 4.0.5
Warning: package 'car' was built under R version 4.0.5
> data(GroceryRetailer)
```

```
> str(GroceryRetailer)
'data.frame':	52 obs. of  4 variables:
 $ y : int  4264 4496 4317 4292 4945 4325 4110 4111 4161 4560 ...
 $ x1: int  305657 328476 317164 366745 265518 301995 269334 267631 296350 277223 ...
 $ x2: num  7.17 6.2 4.61 7.02 8.61 6.88 7.23 6.27 6.49 6.37 ...
 $ x3: int  0 0 0 0 1 0 0 0 0 0 ...
```

(a) **[2 marks]** Fit the main effects linear regression model $E[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. State the estimated regression equation. Give precise interpretations for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$.

```
> model <- lm(y ~ x1 + x2 + x3, data = GroceryRetailer)
> summary(model)

Call:
lm(formula = y ~ x1 + x2 + x3, data = GroceryRetailer)

Residuals:
    Min      1Q  Median      3Q     Max 
-264.05 -110.73  -22.52   79.29  295.75 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  4.150e+03  1.956e+02  21.220  < 2e-16 ***
x1           7.871e-04  3.646e-04   2.159   0.0359 *
x2          -1.317e+01  2.309e+01  -0.570   0.5712
x3           6.236e+02  6.264e+01   9.954 2.94e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 143.3 on 48 degrees of freedom
Multiple R-squared:  0.6883,    Adjusted R-squared:  0.6689
F-statistic: 35.34 on 3 and 48 DF,  p-value: 3.316e-12
```

The estimated regression Equation $E[\hat{y}] = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$ **is** $E[\hat{y}] = 4150 + 0.0007871x_1 - 13.17x_2 + 623.6x_3$

Precise interpretation for β**s:**

β_1=Expected change in total labour hour associated with a one unit increase in the number of cases shipped (x_1) with fixed indirect costs of the total labour hours percentage and holiday predictor.

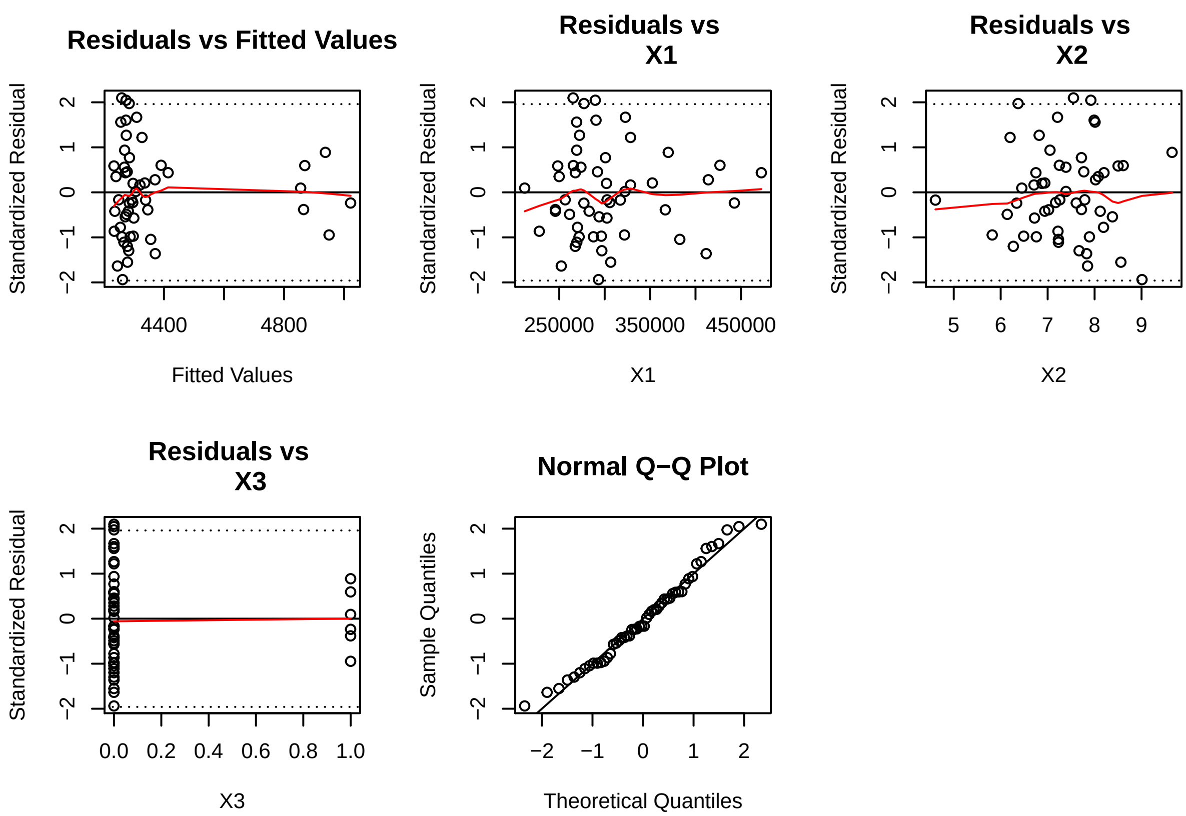
β_2=Expected change in total labour hour associated with a one percentage increase in the indirect costs of the total labour hours as percentage (x_2) with a fixed number of cases shipped and holiday predictor.

β_3=Expected change in total labour hour for holidays ($x_{i3} = 1$) versus non-holidays ($x_{i3} = 0$) with fixed number of cases shipped and percentage of indirect costs of the total labour hours.

(b) [**2 marks**] Plot the (standardized) residuals against $\hat{Y}$, x_1, x_2, and x_3 on four separate graphs. Also prepare a normal quantile-quantile plot. Summarize your findings from these plots with respect to whether model assumptions appear to be satisfied or not. Be specific.

```
> par(mfrow = c(2, 3))
> plot(model$fitted.values, rstandard(model), main = "Residuals vs Fitted Values",
+     ylab = "Standardized Residual", xlab = "Fitted Values")
> abline(h = 0)
> abline(h = 1.96, lty = 3)
> abline(h = -1.96, lty = 3)
> lines(lowess(model$fitted.values, rstandard(model)), col = "red")
>
> plot(GroceryRetailer$x1, rstandard(model), main = "Residuals vs
+      X1", ylab = "Standardized Residual",
+     xlab = "X1")
> abline(h = 0)
> abline(h = 1.96, lty = 3)
> abline(h = -1.96, lty = 3)
> lines(lowess(GroceryRetailer$x1, rstandard(model)), col = "red")
>
> plot(GroceryRetailer$x2, rstandard(model), main = "Residuals vs
+      X2", ylab = "Standardized Residual",
+     xlab = "X2")
> abline(h = 0)
> abline(h = 1.96, lty = 3)
> abline(h = -1.96, lty = 3)
> lines(lowess(GroceryRetailer$x2, rstandard(model)), col = "red")
>
> plot(GroceryRetailer$x3, rstandard(model), main = "Residuals vs
+      X3", ylab = "Standardized Residual",
```

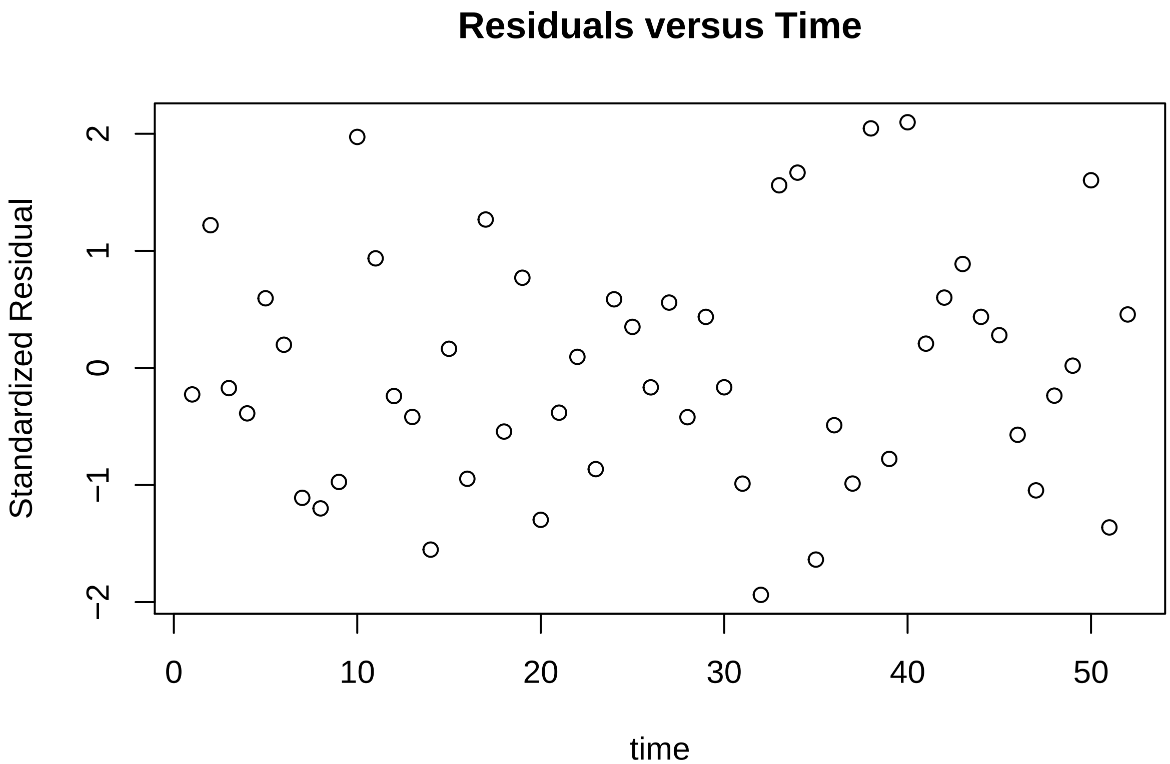
```
+     xlab = "X3")
> abline(h = 0)
> abline(h = 1.96, lty = 3)
> abline(h = -1.96, lty = 3)
> lines(lowess(GroceryRetailer$x3, rstandard(model)), col = "red")
>
> qqnorm(rstandard(model))
> abline(0, 1)
```



Comments: Since there are total 52 obs in the dataset, we expect maximum 2 points outside the interval (-1.96, 1.96). For all four points, we can see that all plots have exactly 2 points outside the range despite that a few points are at the boundary. For the fitted values, most of the data are located at the left. For x1, we see that data are poisitioned slightedly at left. For x2, there is a large cluster of points in the middle of the range. For x3, because of the nature of the data, all points are at 0 or 1. For all four plots, the lowess lines are going around 0, despite that lines of x1 and x2 are slightly off on the left side. From firt four graphs, we will say that the model assumption can be satisfied, but we need a more direct visualization of residuals. Therefore, we check the qq-plot. In this plot, data density is high in the middle around 0 and low at tails. This really saitsfies the shape of normal distribution. Therefore, we conclude that the model assumptions appear to satisfy.

(c) [**2 marks**] Plot the (standardized) residuals versus time. The order of observations in the data represents sequential weeks. Is there any indication that the error terms are correlated? Discuss.

```
> plot(rstandard(model), main = "Residuals versus Time", ylab = "Standardized Residual",
+     xlab = "time")
```



Comments: From the plot, we see that the residuals are uniformly distributed in the range. We don't see a clear pattern in the graph. Therefore, there is no indication that the error terms are correlated.

Regardless of your conclusions above, assume that the regression model is appropriate.

(d) **[2 marks]** Compute a 95% confidence interval for the regression parameter for holiday week. You can use the estimate and its standard error obtained from the R `summary()` object but you must show the details of the calculation (i.e. formula, values, result).

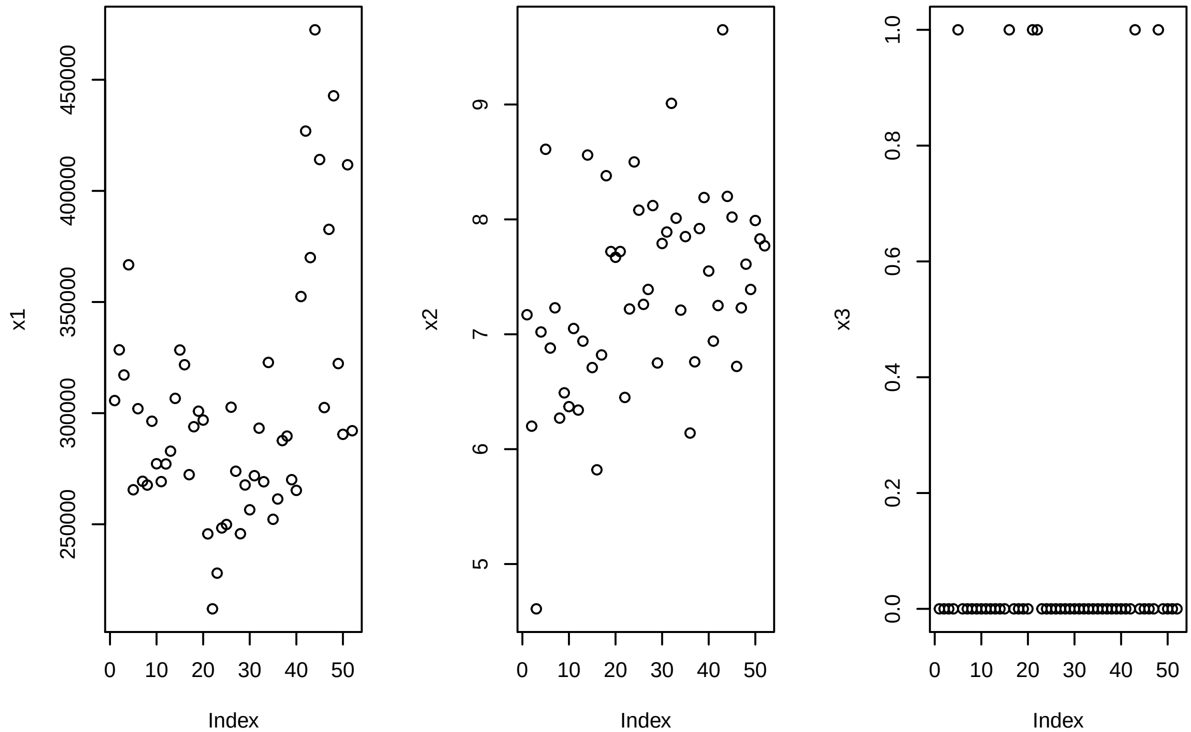
```
> summary(model)$coeff
                 Estimate   Std. Error    t value     Pr(>|t|)
(Intercept)  4.149887e+03 1.955654e+02 21.2199453 4.902653e-26
x1           7.870804e-04 3.645540e-04  2.1590228 3.587650e-02
x2          -1.316602e+01 2.309173e+01 -0.5701616 5.712274e-01
x3           6.235545e+02 6.264095e+01  9.9544230 2.940869e-13
> qt(0.975, 48)
[1] 2.010635
> 623.55 + c(-1, 1) * qt(0.975, 48) * 62.64
[1] 497.6038 749.4962
```

Steps:

$\hat{\beta}_3 \pm t_{52-3-1} se(\hat{\beta}_3) = 623.55 \pm 2.011(62.64) = (497.6038, 749.4962)$

(e) **[2 marks]** Based on the given data, would you consider a shipment of 400,000 cases with an indirect percentage of 7.2 on a nonholiday week to be within the scope of the model? What about a shipment of 400,000 cases with an indirect percentage of 9.9 on a nonholiday week? Support your answers by preparing a relevant plot. If appropriate, use the fitted regression model to predict the total labour hours for these weeks.

```
> par(mfrow = c(1, 3))
> plot(GroceryRetailer$x1, ylab = "x1")
> plot(GroceryRetailer$x2, ylab = "x2")
> plot(GroceryRetailer$x3, ylab = "x3")
```



Comment: From three graphs, we see that the maximum of x1 is over 450000, therefore 400000 falls in a reasonable range. The maximum of x2 is above 9 and the mean is around 7.4. I will say that 7.2 falls in a reasonable range but 9.9 is out of range. We do not need to care about holidays as both 0 or 1 can happen. So, the first case is in the scope of the model but the second is not.

The predicted total labour hours for the first case is:

```
> 4150 + 0.0007871 * 4e+05 - 13.17 * 7.2 + 623.6 * 0
[1] 4370.016
```

Question 2 [10 marks] Adapted from Problem 8.6 of Casella and Berger

Suppose we have two independent samples: $X_1, \dots, X_n$ are iid Exponential random variables with mean θ_1 and $Y_1, \dots, Y_n$ are iid Exponential random variables with mean θ_2.

(a) **[2 marks]** Find the likelihood and log likelihood functions for the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$.

Likelyhood Function:

$$\begin{aligned}
L(\theta_1, \theta_2) &= \prod_{i=1}^{n} \frac{1}{\theta_1} exp(-\frac{x_i}{\theta_1}) \prod_{j=1}^{n} \frac{1}{\theta_2} exp(-\frac{y_j}{\theta_2}) \\
&= \frac{1}{\theta_1^n} exp(-\frac{\sum_{i=1}^n x_i}{\theta_1}) \frac{1}{\theta_2^n} exp(-\frac{\sum_{j=1}^n y_j}{\theta_2}) \\
&= \frac{1}{(\theta_1 \theta_2)^n} exp(-\frac{\sum_{i=1}^n x_i}{\theta_1} - \frac{\sum_{j=1}^n y_j}{\theta_2})
\end{aligned}$$

Log Likelyhood Function:

$$\begin{aligned}
l(\theta_1, \theta_2) &= log(L(\theta_1, \theta_2)) \\
&= log(\frac{1}{(\theta_1 \theta_2)^n} exp(-\frac{\sum_{i=1}^n x_i}{\theta_1} - \frac{\sum_{j=1}^n y_j}{\theta_2})) \\
&= -nlog(\theta_1) - nlog(\theta_2) - \frac{\sum_{i=1}^n x_i}{\theta_1} - \frac{\sum_{j=1}^n y_j}{\theta_2}
\end{aligned}$$

(b) **[2 marks]** Obtain the score vector and information matrix and find the maximum likelihood estimator of $\boldsymbol{\theta}$.

Score Vector:

$$\begin{aligned}
S(\theta_1, \theta_2) &= \begin{bmatrix} \frac{\partial l}{\partial \theta_1} \\ \\ \frac{\partial l}{\partial \theta_2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\theta_1} + \frac{\sum_{i=1}^n x_i}{\theta_1^2} \\ \\ -\frac{n}{\theta_2} + \frac{\sum_{j=1}^n y_i}{\theta_2^2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\theta_1} + \frac{n\bar{x}}{\theta_1^2} \\ \\ -\frac{n}{\theta_2} + \frac{n\bar{y}}{\theta_2^2} \end{bmatrix}
\end{aligned}$$

Information Matrix:

$$I(\theta_1, \theta_2) = \begin{bmatrix} -\frac{\partial^2 l}{\partial \theta_1^2} & -\frac{\partial^2 l}{\partial \theta_1 \theta_2} \\ \\ -\frac{\partial^2 l}{\partial \theta_2 \theta_1} & -\frac{\partial^2 l}{\partial \theta_2^2} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{n}{\theta_1^2} + \frac{2\sum_{i=1}^n x_i}{\theta_1^3} & 0 \\ \\ 0 & -\frac{n}{\theta_2^2} + \frac{2\sum_{j=1}^n y_i}{\theta_2^3} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{n}{\theta_1^2} + \frac{2n\bar{x}}{\theta_1^3} & 0 \\ \\ 0 & -\frac{n}{\theta_2^2} + \frac{2n\bar{y}}{\theta_2^3} \end{bmatrix}$$

Set $S(\theta_1, \theta_2) = 0$, then:

$$-\frac{n}{\theta_1} + \frac{n\bar{x}}{\theta_1^2} = 0 \text{ and } -\frac{n}{\theta_2} + \frac{n\bar{y}}{\theta_2^2} = 0$$

$$\hat{\theta}_1 = \bar{x} \text{ and } \hat{\theta}_2 = \bar{y}$$

$$\text{Therefore, } \hat{\theta} = \begin{bmatrix} \bar{x} \\ \\ \bar{y} \end{bmatrix}$$

$$I(\hat{\theta}_1, \hat{\theta}_2) = \begin{bmatrix} -\frac{n}{\hat{\theta}_1^2} + \frac{2n\bar{x}}{\hat{\theta}_1^3} & 0 \\ \\ 0 & -\frac{n}{\hat{\theta}_2^2} + \frac{2n\bar{y}}{\hat{\theta}_2^3} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{n}{\bar{x}^2} & 0 \\ \\ 0 & \frac{n}{\bar{y}^2} \end{bmatrix} > 0 \text{ since } x_i, y_i \geq 0 \text{ and } \theta_1, \theta_2 > 0 \text{ and } n > 0$$

Therefore, $\hat{\theta}$ is the MLE of θ.

(c) [**3 marks**] Consider a likelihood ratio test of $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 \neq \theta_2$. Show that the test statistic can be written as a function of $T = \bar{X}/(\bar{X} + \bar{Y})$. Note: $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Perform the test with observed data $n = 100$, $\bar{x} = 3$, $\bar{y} = 4$. Report the p-value. Do you reject the null hypothesis?

$$
\begin{aligned}
r(\theta) &= l(\theta) - l(\hat{\theta}_1, \hat{\theta}_2) \\
&= -nlog(\theta) - nlog(\theta) - \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{\sum_{j=1}^{n} y_j}{\theta} + nlog(\hat{\theta}_1) + nlog(\hat{\theta}_2) + \frac{\sum_{i=1}^{n} x_i}{\hat{\theta}_1} + \frac{\sum_{j=1}^{n} y_j}{\hat{\theta}_2} \\
&= -nlog(\theta) - nlog(\theta) - \frac{n\bar{X}}{\theta} - \frac{n\bar{Y}}{\theta} + nlog(\bar{X}) + nlog(\bar{Y}) + \frac{n\bar{X}}{\bar{X}} + \frac{n\bar{Y}}{\bar{Y}} \\
&= -2nlog(\theta) - \frac{n(\bar{X}+\bar{Y})}{\theta} + nlog(\bar{X}\bar{Y}) + 2n \\
&= -nlog((\frac{\bar{X}+\bar{Y}}{2})^2) - \frac{n(\bar{X}+\bar{Y})}{\frac{\bar{X}+\bar{Y}}{2}} + nlog(\bar{X}\bar{Y}) + 2n \text{ Since } \hat{\theta} = \frac{\bar{X}+\bar{Y}}{2} \\
&= -nlog(\frac{(\bar{X}+\bar{Y})^2}{4}) + nlog(\bar{X}\bar{Y}) \\
&= nlog(\frac{4\bar{X}\bar{Y}}{(\bar{X}+\bar{Y})^2}) \\
&= nlog(4\frac{\bar{X}}{\bar{X}+\bar{Y}}\frac{\bar{Y}}{\bar{X}+\bar{Y}}) \\
&= nlog(4\frac{\bar{X}}{\bar{X}+\bar{Y}}\frac{1-\bar{X}}{\bar{X}+\bar{Y}}) \\
&= nlog(4T(1-T)) \text{ where } T = \frac{\bar{X}}{\bar{X}+\bar{Y}}
\end{aligned}
$$

Therefore, the test statistic $-2r(\theta_1, \theta_2) = -2nlog(4T(1-T))$

Since $n = 100$**,** $\bar{x} = 3$**,** $\bar{y} = 4$**, we will have** $-2r(\theta) = -200log(4\frac{3}{7}\frac{4}{7}) \approx 4.12386$

$$
\begin{aligned}
p &= P(\chi^2_{(1)} > -2r(\theta)) \\
&= P(\chi^2_{(1)} > 4.12386) \\
&= 0.04228256
\end{aligned}
$$

```
> 1 - pchisq(-200 * log(4 * 3/7 * 4/7), 1)
[1] 0.04228256
```

Since 0.04228256 < 0.05, there is an evidence against the null hypothesis. We reject H_0**:** $\theta_1 = \theta_2$**. The distributions of** X **and** Y **are likely to be not the same.**

(d) **[3 marks]** Regardless of your conclusion in part (c) assume that the X's and Y's were generated from a single Exponential distribution with mean θ. Find a Wald based 95% confidence interval for θ given the observed data in part (c).

$$
\begin{aligned}
l(\theta) &= -2nlog(\theta) - \frac{n(\bar{X}+\bar{Y})}{\theta} \\
I(\theta) &= \frac{\partial^2 l}{\partial \theta^2} \\
&= -\frac{2n}{\theta^2} + \frac{2n(\bar{X}+\bar{Y})}{\theta^3} \\
&= \frac{2n(\bar{X}+\bar{Y}-\theta)}{\theta^3} \\
I(\hat{\theta}) &= \frac{2n(\bar{X}+\bar{Y}-\frac{\bar{X}+\bar{Y}}{2})}{(\frac{\bar{X}+\bar{Y}}{2})^3} \\
&= \frac{n(\bar{X}+\bar{Y})}{\frac{(\bar{X}+\bar{Y})^3}{8}} \\
&= \frac{8n}{(\bar{X}+\bar{Y})^2}
\end{aligned}
$$

Therefore, we will have Wald based 95% Wald Confidence Interval:

$$
\begin{aligned}
\hat{\theta} \pm 1.96[I(\hat{\theta})]^{-\frac{1}{2}} &= \frac{\bar{X}+\bar{Y}}{2} \pm 1.96[\frac{8n}{(\bar{X}+\bar{Y})^2}]^{-\frac{1}{2}} \\
&= \frac{3+4}{2} \pm 1.96[\frac{800}{(3+4)^2}]^{-\frac{1}{2}} \\
&= 3.5 \pm 1.96[\frac{800}{49}]^{-\frac{1}{2}} \\
&= (3.0149, 3.9851)
\end{aligned}
$$

Question 3 [10 marks]

A random variable Y has a negative binomial distribution with parameters (μ, k), if its probability mass function can be written as

$$f(y;\mu,k) = \binom{k+y-1}{k-1}\left(\frac{k}{\mu+k}\right)^k\left(\frac{\mu}{\mu+k}\right)^y \qquad y = 0,1,2,\ldots$$

where $\mu > 0$ and $k > 0$.

(a) **[3 marks]** Show that this distribution belongs to the exponential family if k is known by identifying the canonical parameter θ, the dispersion parameter ϕ, and the functions $a(\phi)$, $b(\theta)$, $c(y;\phi)$.

$$\begin{aligned}
f(y;\mu;k) &= \binom{k+y-1}{k-1}(\frac{k}{\mu+k})^k(\frac{\mu}{\mu+k})^y \\
&= exp\{log(\binom{k+y-1}{k-1}(\frac{k}{\mu+k})^k(\frac{\mu}{\mu+k})^y)\} \\
&= exp\{log[\binom{k+y-1}{k-1}] + k\cdot log(\frac{k}{\mu+k}) + y\cdot log(\frac{\mu}{\mu+k})\} \\
&= exp\{y\cdot log(\frac{\mu}{\mu+k}) + k\cdot log(\frac{k}{\mu+k}) + log[\binom{k+y-1}{k-1}]\}
\end{aligned}$$

Therefore, $\phi = 1, \theta = log(\frac{\mu}{\mu+k})$,

$$a(\phi) = 1, b(\theta) = -k\cdot log(\frac{k}{\mu+k}) = -k\cdot log(1-exp(\theta))$$

$$c(y;\phi) = log[\binom{k+y-1}{k-1}]$$

(b) **[3 marks]** Obtain an expressions for the mean and variance of Y and the canonical link function.

$$
\begin{aligned}
E(Y) &= b'(\theta) \\
&= -k \cdot \frac{1}{1 - exp(\theta)} \cdot -exp(\theta) \\
&= -k \cdot \frac{-exp(\theta)}{1 - exp(\theta)} \\
&= -k \cdot \frac{-\frac{\mu}{\mu+k}}{1 - \frac{\mu}{\mu+k}} \\
&= -k \cdot \frac{-\frac{\mu}{\mu+k}}{\frac{k}{\mu+k}} \\
&= \mu
\end{aligned}
$$

$$
\begin{aligned}
Var(Y) &= b''(\theta) \cdot a(\phi) \\
&= \frac{k \cdot exp(\theta)}{(1 - exp(\theta))^2} \cdot 1 \\
&= \frac{k \cdot \frac{\mu}{\mu+k}}{(1 - \frac{\mu}{\mu+k})^2} \\
&= \frac{\frac{k\mu}{\mu+k}}{(\frac{k}{\mu+k})^2} \\
&= \frac{\mu(\mu+k)}{k}
\end{aligned}
$$

$$
\begin{aligned}
\text{Set } g(\mu) &= \theta = \eta = x^T b \\
&= log(\frac{\mu}{\mu+k}) \\
\text{Therefore, the canonical link is } g(\mu) &= log(\frac{\mu}{\mu+k})
\end{aligned}
$$

(c) **[4 marks]** Given the data $y_1, y_2, \ldots, y_n$ and the linear predictor $\eta_i = \sum_{j=0}^{p-1} x_{ij}\beta_j$ where $x_{i0} = 1$, find the specific form of the score and information function under the canonical link. Briefly describe how you would obtain maximum likelihood estimates of $\beta_0, \beta_1, \ldots, \beta_{p-1}$ for a specified value of the parameter k.

$$\begin{aligned}
\text{Since } g(\mu_i) &= x_i^T\beta, \\
log(\frac{\mu_i}{\mu_i + k}) &= x_i^T\beta \\
\mu_i &= exp(x_i^T\beta) \\
\mu_i &= (\mu_i + k)exp(x_i^T\beta) \\
\mu_i &= \frac{k \cdot exp(x_i^T\beta)}{1 - exp(x_i^T\beta)}
\end{aligned}$$

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{k}{\mu_i(\mu_i + k)} = \frac{(1 - exp(x_i^T\beta))^2}{k \cdot exp(x_i^T\beta)}$$

$$Var(y_i) = \frac{mu_i(mu_i + k)}{k} = \frac{k \cdot exp(x_i^T\beta)}{(1 - exp(x_i^T\beta))^2}$$

$$\begin{aligned}
W_i^{-1} &= Var(y_i)(\frac{\partial \eta_i}{\partial \mu_i})^2 \\
&= \frac{k \cdot exp(x_i^T\beta)}{(1 - exp(x_i^T\beta))^2}(\frac{(1 - exp(x_i^T\beta))^2}{k \cdot exp(x_i^T\beta)})^2 \\
&= \frac{(1 - exp(x_i^T\beta))^2}{k \cdot exp(x_i^T\beta)} \\
\text{Then, } W_i &= \frac{k \cdot exp(x_i^T\beta)}{(1 - exp(x_i^T\beta))^2}
\end{aligned}$$

Therefore, the score vector is:

$$\begin{aligned}
S_j(\beta) &= \sum_{i=1}^{n}(y_i - \mu_i) \cdot W_i \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij} \\
&= \sum_{i=1}^{n}(y_i - \frac{k \cdot exp(x_i^T\beta)}{1 - exp(x_i^T\beta)}) \cdot \frac{k \cdot exp(x_i^T\beta)}{(1 - exp(x_i^T\beta))^2} \cdot \frac{(1 - exp(x_i^T\beta))^2}{k \cdot exp(x_i^T\beta)} \cdot x_{ij} \\
&= \sum_{i=1}^{n}(y_i - \frac{k \cdot exp(x_i^T\beta)}{1 - exp(x_i^T\beta)}) \cdot x_{ij}
\end{aligned}$$

Since the information function is under canonical link, we will have the expected information matrix equals the observed information matrix.
Therefore, the Information matrix is:

$$\begin{aligned}
I(\beta)_{jk} &= E[I(\beta)_{jk}] \\
&= \sum_{i=1}^{n} x_{ij} \cdot W_i \cdot x_{jk} \\
&= \sum_{i=1}^{n} x_{ij} \cdot \frac{k \cdot exp(x_i^T\beta)}{(1 - exp(x_i^T\beta))^2} \cdot x_{jk}
\end{aligned}$$

To find maximum likelihood estimates of $\beta_0, \beta_1, \ldots, \beta_{p-1}$ for a specified value of the parameter k, we can use Newton Raphson Method to solve them iteratively.

1. We start with an initial guess $\beta^{(0)}$

2. evaluate score vector $S(\hat{\beta}^{(r)})$ the information matrix $I(\hat{\beta}^{(r)})$ for each iteration. We can use the expected information matrix instead of observed one.
3. update coefficients by $\hat{\beta}^{(r+1)} = \hat{\beta}^{(r)} + I^{-1}(\hat{\beta}^{(r)})S(\hat{\beta}^{(r)})$
4. Repeat step 2 and step 3 until the new one is sufficiently close to the old one.
5. We get the MLE of $\beta_0, \beta_1, \ldots, \beta_{p-1}$.