total grade: 4.9/6

Date
Page

Q.1

a) Contingency Table :-

[0.8/1.0]

	Rescued	Not Rescued	Total
First Class	135	202	337
Second Class	160	125	285
Third Class	541	180	721
Staff	674	211	885
Total	1510	718	2228

✓

b) Conditional Relative Frequency distribution

	Rescued	Not Rescued
First Class	$\frac{135}{337} = 0.4006$	$\frac{202}{337} = 0.5994$
~~Seco~~		
Second Class	$\frac{160}{285} = 0.5614$	$\frac{125}{285} = 0.4385$
~~Thi~~		
Third Class	$\frac{541}{721} = 0.7503$	$\frac{180}{721} = 0.2496$
Staff	$\frac{674}{885} = 0.7615$	$\frac{211}{885} = 0.2384$

✓

There seems to be a higher percentage of rescues in lower-class passengers.

c) Formula ~~for~~ :- E_{ij} = (Row Total × Colum Total) / Grand Total.

	Rescued	Not Rescued
First Class	228.398	108.6023
Second Class	193.1553	91.8447
Third Class	488.649	232.3510
Staff	599.798	285.2020

✓

5.6911

Cramer's V

$$V = \sqrt{\frac{\chi^2}{n \times \min(r-1),(c-1)}}$$

$$\chi^2 = \frac{(135-228.398)^2}{228.398} + \frac{(202-108.6023)^2}{108.6023} + \dots + \frac{(718-285.2020)^2}{285.2020}$$

$$= 38.1929 + 80.3218 + 5.6911 + 11.9688 + 5.6086 + 11.7952 + 9.1797 + 19.3054$$

$$\chi^2 = 201.3689$$

minor computational error

$$V = \sqrt{\frac{201.3689}{2228 \times \min(3,1)}} = \sqrt{\frac{201.3689}{2228 \times 1}} = 0.3006$$

Which means association between variables X = "Travel class" & Y = "Rescue status") is moderate (or not strong.)

d) According to (a) & (~~c~~) it seems that being in a certain class may have influenced the likelihood of being rescued.

[1.0/1.2]

Date ______
Page ______

Q.3

a] The population mean (first moment) is $\frac{1}{p}$

first sample mean of the geometric distribution is, $\frac{1}{n}\sum_{i=1}^{n} x_i$

~~Set up the equation~~ we equate sample moment to their corresponding population moment.

$$\therefore \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{p}$$

$$\hat{p}_{MoM} = \frac{1}{\bar{x}}$$ ✓

b] Jensen inequality :-

$$f(E[x]) \leq E[f(x)]$$

The Jensen's inequality for $f(x) = \frac{1}{x}$ is ... strict convex function

$$\frac{1}{E[x]} \leq E\left[\frac{1}{x}\right]$$

$$\frac{1}{1/p} \leq E\left[\frac{1}{x}\right]$$

$$p \leq E\left[\frac{1}{x}\right]$$

$$\hat{p}_{MoM} = \frac{1}{\bar{x}}$$

i.e $\hat{p}_{MoM}$ is an biased estimator for p due to strict inequality.

✓

c] $P(X=k) = (1-p)^{k-1} \cdot p$ ... PMF of Geometric distribuⁿ

$L(p) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} (1-p)^{x_i - 1} \cdot p$

Log-likelihood function is :-

$l(p) = \sum_{i=1}^{n} [(x_i - 1) \cdot \log(1-p) + \log(p)]$

let's find derivative

$$\frac{d}{dp}\left(\log l(p)\right) = \frac{d}{dp}\left(n\log(p) + \sum_{i=1}^{n}(x_i-1)\log(1-p)\right)$$

$$= \frac{n}{p} - \sum_{i=1}^{n}\frac{x_i-1}{\cancel{p-p}\,1-p}$$

setting derivative zero

$$\frac{n}{p} - \sum_{i=1}^{n}\frac{x_i-1}{1-p} = 0$$

$$\frac{n}{p} = \sum_{i=1}^{n}\frac{x_i-1}{1-p}$$

Date ______
Page ______

$$\frac{1}{p} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i - 1}{1-p}$$

$$\frac{1}{p} = \frac{\sum_{i=1}^{n} x_i - 1}{n(1-p)}$$

$$n(1-p) = p \sum_{i=1}^{n} (x_i - 1)$$

$$n - np = p \sum_{i=1}^{n} x_i - p \sum_{i=1}^{n} 1$$

$$n = p \sum_{i=1}^{n} x_i - pn + np$$

$$n = p \sum_{i=1}^{n} x_i$$

$$\therefore p = \frac{n}{\sum_{i=1}^{n} x_i}$$

$\hat{P}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}$ ... we can see it's asymptotically unbiased where $\hat{P}_{MOM}$ was biased.

they are the same here!

d) $n = 15$

$\sum_{i=1} x_i = 2 + 12 + \ldots + 4 + 1 + 1$

$= 47$

$\hat{P}_{ML} = \frac{15}{47}$ ✓

```r
#FOS_HW9_Q3_e
# Given dataset
data <- c(2, 12, 2, 2, 2, 2, 1, 2, 9, 1, 2, 4, 4, 1, 1)

# Log-likelihood function for geometric distribution
log_likelihood <- function(p) {
  sum(log(p) + (data - 1) * log(1 - p))
}

# Numerical optimization to find maximum likelihood estimator
initial_guess <- 0.3191
result <- optim(initial_guess, log_likelihood, method = "Brent", lower = 0, upper = 1)

# Maximum likelihood estimator
p_ML <- result$par
print(p_ML)
```

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

R Notebook

2 - (a) [1/1]

```
library(ggplot2)
library(MASS)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

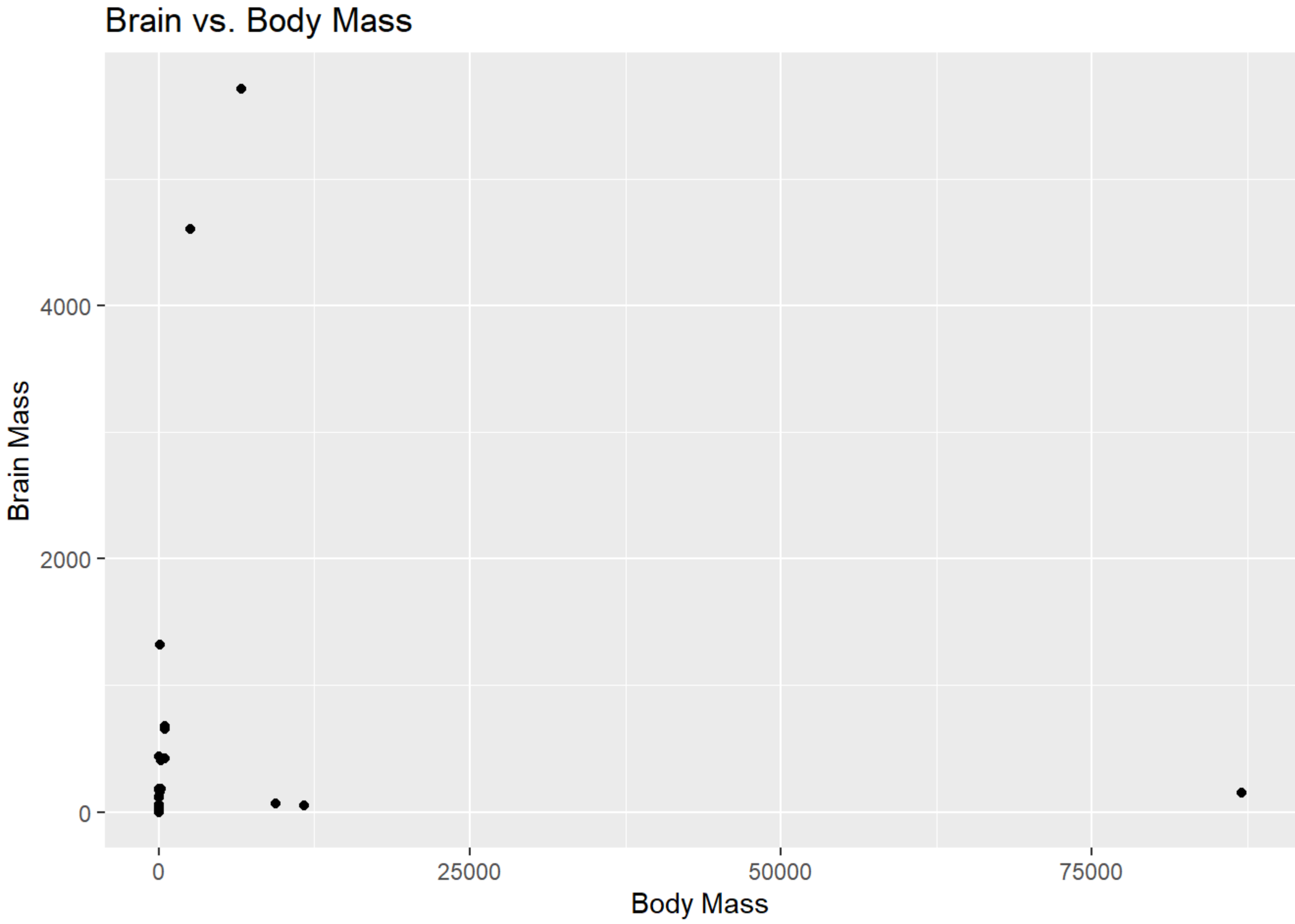
```
library(tibble)
library(ggrepel)
```

```
## Warning: package 'ggrepel' was built under R version 4.3.2
```

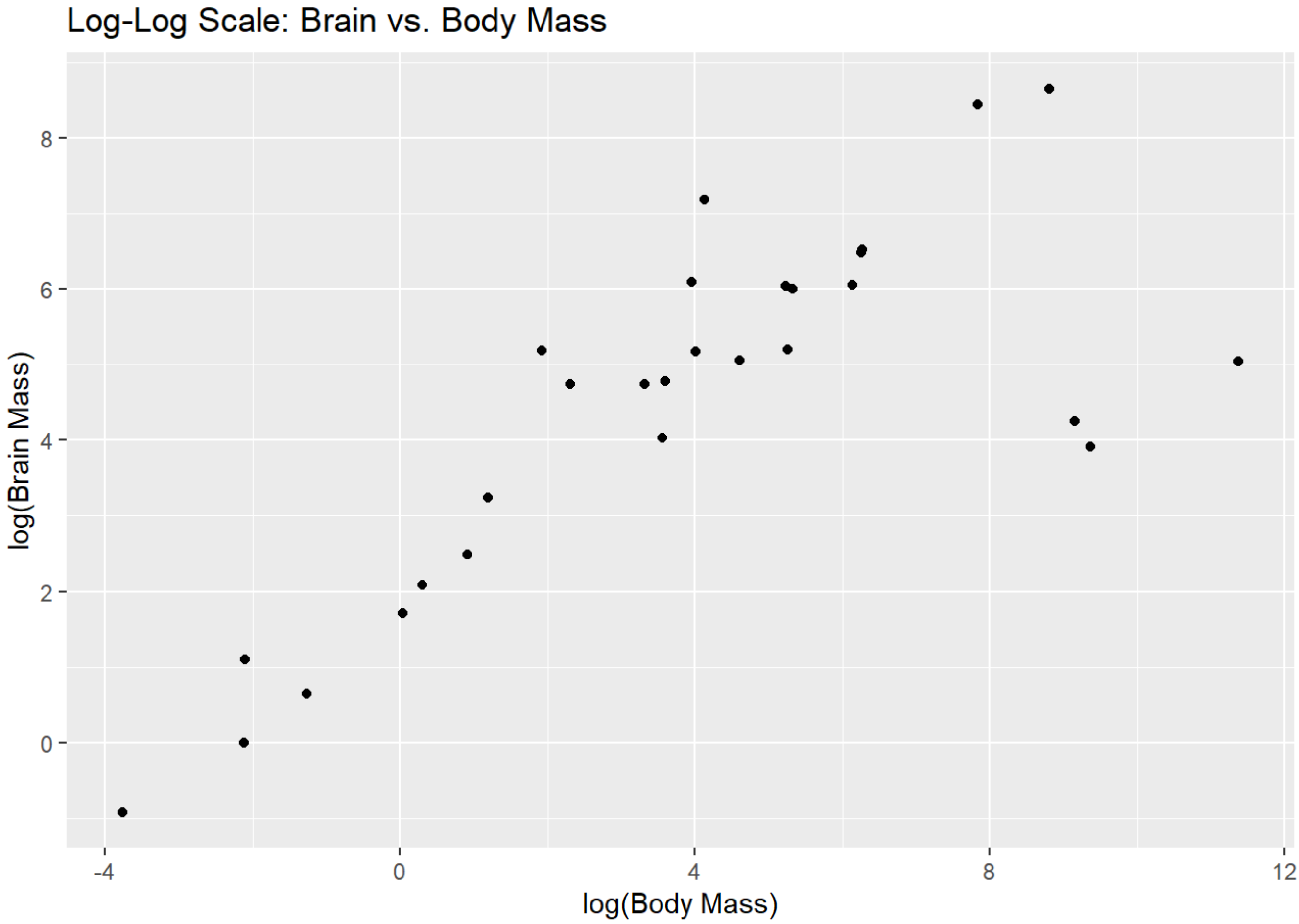
```
data <- Animals
data <- data %>%
  rownames_to_column(var = "species")

#Bravais-Pearson correlation coefficient
coeff <- cor(data$body, data$brain)
cat("Bravais-Pearson Correlation Coefficient for the given data is ", coeff, "\n")
```

```
## Bravais-Pearson Correlation Coefficient for the given data is  -0.005341163
```

```
#Plotting the data in normal scale
ggplot(data, aes(x = body, y = brain)) +
  geom_point() +
  labs(title = "Brain vs. Body Mass", x = "Body Mass", y = "Brain Mass")
```



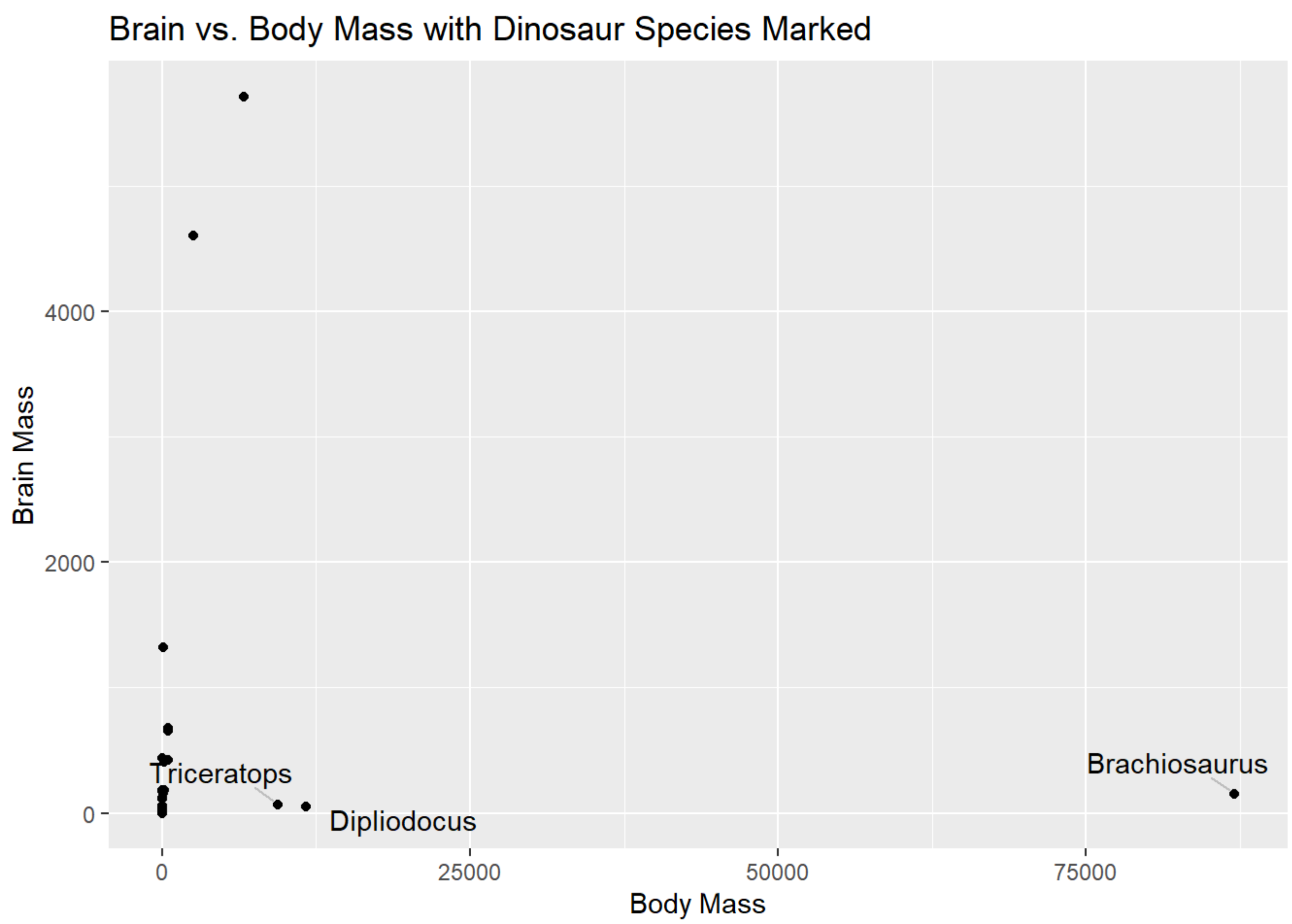

```
#Ploting the data in log-log scale
ggplot(data, aes(x = log(body), y = log(brain))) +
  geom_point() +
  labs(title = "Log-Log Scale: Brain vs. Body Mass", x = "log(Body Mass)", y = "log(Brain Mas
s)")
```



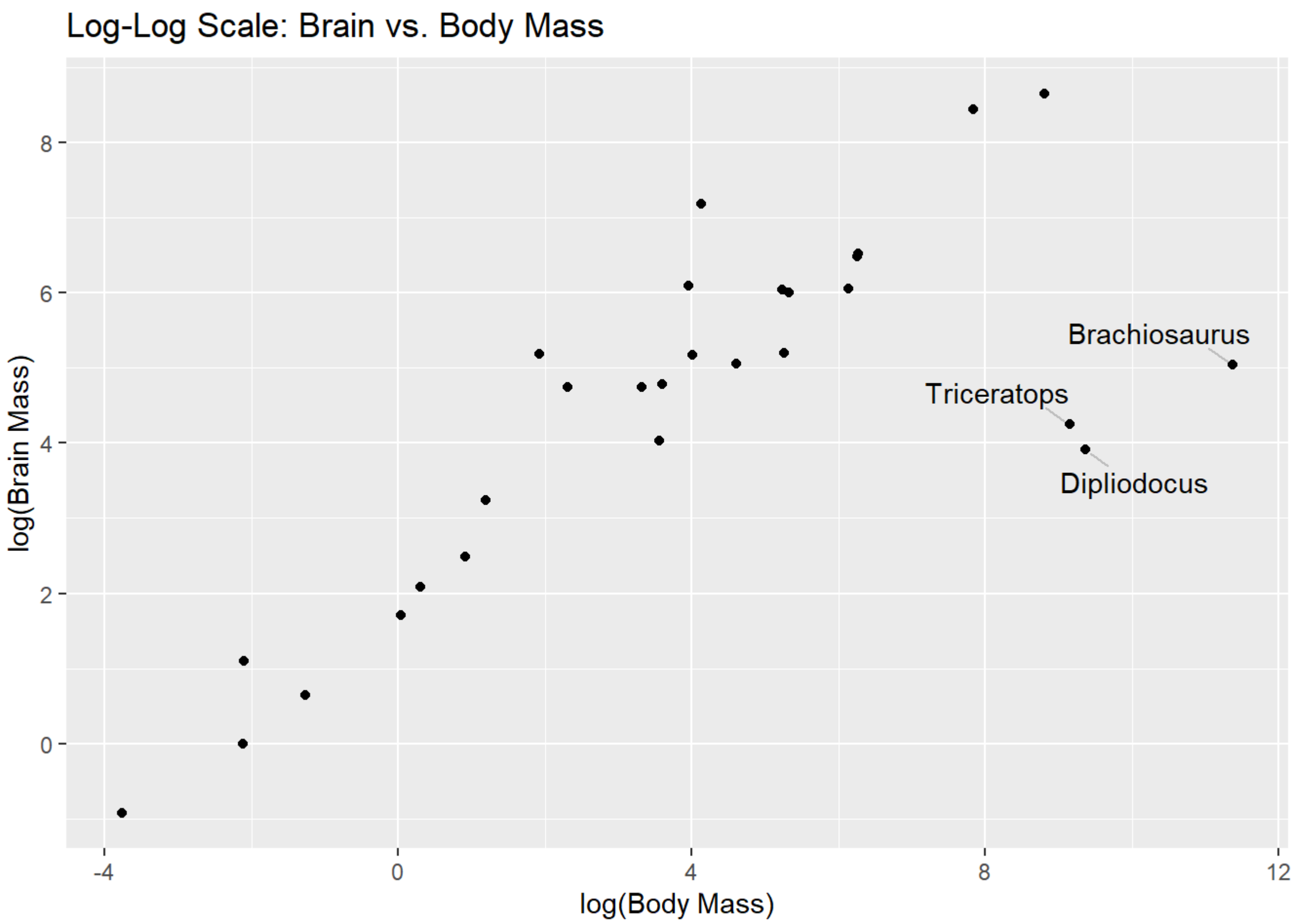
The Bravais-Pearson Correlation Coefficient between body mass and brain size is -0.005341163. This value is close to zero, indicating a very weak and practically insignificant linear relationship between the two variables. Contrary to the hypothesis, there is no evidence of a positive correlation; instead, the weak correlation observed is negative, suggesting a slight tendency for brain size to decrease as body mass increases, although the relationship is not practically significant.

2 - (b)

```
#there are 3 dinosour species in dataset
dinosaur_species <- c('Dipliodocus', 'Brachiosaurus', 'Triceratops')

#Plotting the data with dinosaur species marked
ggplot(data, aes(x = body, y = brain)) +
  geom_point() +
  geom_text_repel(data = subset(data, species %in% dinosaur_species),
                  aes(label = species),
                  box.padding = 0.5,
                  point.padding = 0.2,
                  force = 5,
                  segment.color = "grey",
                  direction = "both") +
  labs(title = "Brain vs. Body Mass with Dinosaur Species Marked",
       x = "Body Mass",
       y = "Brain Mass")
```



```r
#In log-log scale
ggplot(data, aes(x = log(body), y = log(brain))) +
  geom_point() +
  geom_text_repel(data = subset(data, species %in% dinosaur_species),
                  aes(label = species),
                  box.padding = 0.5,
                  point.padding = 0.2,
                  force = 5,
                  segment.color = "grey",
                  direction = "both") +
  labs(title = "Log-Log Scale: Brain vs. Body Mass", x = "log(Body Mass)", y = "log(Brain Mas
s)")
```




```r
#Removing the dinosaur species from the dataset
data_no_dinosaurs <- subset(data, !(species %in% dinosaur_species))

#Bravais-Pearson correlation coefficient without dinosaur species outliers
coeff_no_dinosaurs <- cor(data_no_dinosaurs$body, data_no_dinosaurs$brain)
cat("Bravais-Pearson Correlation Coefficient without Dinosaur Species Outliers is", coeff_no_
dinosaurs, "\n")
```

```
## Bravais-Pearson Correlation Coefficient without Dinosaur Species Outliers is 0.9318502
```

2 - (c)

```
library(coin)
```

```
## Warning: package 'coin' was built under R version 4.3.2
```

```
## Loading required package: survival
```

```
#Spearman's rank correlation coefficient with outliers
spearman_with_outliers <- cor(data$body, data$brain, method = "spearman")
cat("Spearman's Rank Correlation Coefficient with Outliers:", spearman_with_outliers, "\n")
```

```
## Spearman's Rank Correlation Coefficient with Outliers: 0.7162994
```

```
#Spearman's rank correlation coefficient without outliers
spearman_no_outliers <- cor(data_no_dinosaurs$body, data_no_dinosaurs$brain, method = "spearm
an")
cat("Spearman's Rank Correlation Coefficient without Outliers:", spearman_no_outliers, "\n")
```

```
## Spearman's Rank Correlation Coefficient without Outliers: 0.9328717
```

The coefficient obtained without outliers (0.9328717) is higher than the coefficient obtained with outliers (0.7162994). In general, a higher correlation coefficient indicates a stronger monotonic relationship between the variables.

Therefore, in our case Spearman's rank correlation coefficient without outliers is more robust to the presence of outliers in the dataset. The removal of outliers has resulted in a higher correlation coefficient, suggesting a stronger monotonic relationship between the variables when the influence of outliers is reduced.

4-a

[1.0/1.2]

here $X_1, \ldots X_n$ are i.i.d from $U(a,b)$

pdf is given by;

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Likelihood function;

$$L(a,b) = \prod_{i=1}^{n} f(x_i)$$

converting into log-likelihood for easier calculation;

$$l(a,b) = \sum_{i=1}^{n} \log f(x_i)$$

since $\log f(x_i)$ is maximized when $f(x)$ is maximized we focus on maximizing that.

(i) MLE for a :

$$L_a(a) = \prod_{i=1}^{n} f(x_i) = \left(\frac{1}{b-a}\right)^n$$

$$\therefore l_a(a) = n \log \left(\frac{1}{b-a}\right)$$

here to maximize $l_a(a)$ we need to minimize $(b-a)$ that happens where a is minimize among the sample points;

$$\boxed{\therefore \hat{a}_{MLE} = \min\{x_1, \ldots x_n\}}$$

(ii) MLE for b;

$$L_b(b) = \prod_{i=1}^{n} f(x_i) = \left(\frac{1}{b-a}\right)^n$$

$$\therefore l_b(b) = n \log\left(\frac{1}{b-a}\right)$$

TO maximize $l_b(b)$ we minimize $(b-a)$ and this occurs when b is largest, meaning maximum value among the sample points.

$$\boxed{\therefore \hat{b}_{MLE} = \max\{x_1, \ldots x_n\}}$$

4-b.

from H.W.-7 2-C

$$F_{\hat{a}_n}(x) = \begin{cases} 0 & x \le a \\ 1 - \left(\frac{\cancel{x}-a}{b-a}\right)^n & a < x < b \\ 1 & x \geq b \end{cases}$$

$\therefore$

$$\lim_{n \to \infty} F_{\hat{a}_n}(x) = \begin{cases} 0 & x \le a \\ 1 & x > a \end{cases}$$

$$\therefore \quad \underset{n \to \infty}{E}[\hat{a}_n] = \int_0^\infty (1 - F(x))\, dx$$

✓

$$= \int_0^a (1-0)\, dx + \int_a^\infty (1-1)\, dx$$

It seems that here, from convergence in distribution, derived in HW 2(c), you obtain convergence of the first moment.
Note that this is in general not true.
But here is possible because the random variables are bounded.

$$= [x]_0^a + 0$$

$$\therefore \quad \underset{n \to \infty}{E}[\hat{a}_n] = a$$

means ; as $n \to \infty$ $E[\hat{a}_n] \to a$

and ; $F_{\hat{b}_n}(x) = \begin{cases} 0 & x \le a \\ \left(\frac{x-a}{b-a}\right)^n & a < x < b \\ 1 & x \ge b \end{cases}$

$\therefore \lim_{n \to \infty} F_{\hat{b}_n}(x) = \begin{cases} 0 & x < b \\ 1 & x \ge b \end{cases}$

$$\therefore \underset{n \to \infty}{E[\hat{b}_n]} = \int_0^\infty (1 - F(x))\,dx$$

$$= \int_0^b (1-0)\,dx + \int_b^\infty (1-1)\,dx$$

$$= [x]_0^b + 0$$

$$\therefore \lim_{n \to \infty} E[\hat{b}_n] = b$$

means ; as $n \to \infty$ $E[\hat{b}_n] \to b$

So $\hat{a}_n$ and $\hat{b}_n$ are asymptotically unbiased.

4 (c)?

[0.5/0.6]

Exercise 05

(a) The exponential distribution with rate parameter λ has probability density function (PDF):

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Now, Let's find the Expected value of $\bar{x}$:

$$E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

Since x_i's are independent & identically distributed, we can use the linearity of expectations:

$$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$$

The expected value of single x_i is given by;

$$E(x_i) = \frac{1}{\lambda}$$

Therefore

$$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda}$$

$$E(\bar{x}) = \frac{1}{n}\cdot n\cdot\frac{1}{\lambda}$$

$$E(\bar{x}) = \frac{1}{\lambda}$$ ✓

$$\mu = \frac{1}{\lambda}$$

Since the expected value of $\bar{x}$ is equal to μ, the sample mean $\bar{x}$ is an unbiased estimator for the population mean μ in the case of an exponential distribution with parameter λ.

(b) The PDF of M_n can be derived as follows:

$$P(M_n \geq x) = P(X_1 \geq x, X_2 \geq x, \ldots, X_n \geq x)$$

Since the x_i's are independent,

$$P(M_n \geq x) = P(X_1 \geq x) \cdot P(X_2 \geq x) \cdots P(X_n \geq x)$$

For an exponential distribution with rate parameter λ,

Then CDF is given by $F(x) = 1 - e^{-\lambda x}$.

Therefore, the probability of $P(X_i \geq x) = 1 - F(x)$

$$P(X_i \geq x) = 1 - (1 - e^{-\lambda x})$$

$$= e^{-\lambda x}$$

~~Now~~, Therefore,

$$P(M_n \geq x) = (e^{-\lambda x})^n$$

$$= e^{-n\lambda x}$$

Differentiate both sides with respect to x to obtain the probability density function of M_n

$$\cancel{f(M} \quad f_{M_n}(x) = \frac{d}{dx} P(M_n \geq x)$$

$$f_{M_n}(x) = (-n\lambda) e^{-n\lambda x}$$

The PDF of M_n is given by:

$$f_{M_n}(x) = n\lambda e^{-n\lambda x}$$

This is the PDF of an exponential distribution with rate parameter $n\lambda$.

$\therefore$ M_n follows an exp. distribution with rate parameter $n\lambda$.

CamScanner

The expected value of M_n.

$$E(\bar{M}_n) = E(nM_n)$$

$$E(\bar{M}_n) = n\,E(M_n)$$

Since M_n follows an exp. distribution with rate parameter $n\lambda$

$$E(M_n) = \frac{1}{n\lambda}$$

Therefore

$$E(\bar{M}_n) = n\frac{1}{n\lambda}$$

$$E(\bar{M}_n) = \frac{1}{\lambda}$$ ✓

$\cancel{\mu = \mu.}$

$E(\bar{M}_n)$ is the same as the population mean μ, which means that $\bar{\mu}_n = nM_n$ is an unbiased estimator for μ as well.

c) Variance

CamScanner

(e) Variance of $\bar{X}_n$

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

Since X_i follows an exponential distribution with rate λ, the variance of X_i is $\frac{1}{\lambda^2}$.

$$Var(\bar{X}_n) = \frac{1}{n^2} \cdot n \cdot \frac{1}{\lambda^2} = \frac{1}{n\lambda^2}$$ ✓

$$Var(\hat{\mu}_n) = \frac{1}{n\lambda^2}$$ ✓

Variance of M_n:

$$Var(\bar{M}_n) = Var(nM_n) = n^2 Var(M_n)$$

Since M_n follows an exponential distribution with rate $n\lambda$:

$$Var(\bar{M}_n) = n^2 \frac{1}{(n\lambda)^2}$$

$$= \frac{1}{\lambda^2}$$ ✓

opposite

$Var(\bar{X}_n) > Var(\bar{M}_n)$ for $n > 1$ ✗

($\tilde{\mu}_n$)

Therefore, nM_n is a preferred estimator for μ as it has a smaller variance.

Exercise 6 :

(a) For a Gamma distribution $\Gamma(\alpha,\beta)$,

[0.6/1] the population mean and variance are given by,

$$E(x) = \frac{\alpha}{\beta} \qquad \text{(from slide 24 chapter 1.6)}$$

✓

$$Var(x) = \frac{\alpha}{\beta^2}$$

According to that,

we can get sample moments equal to population moments,

× $\therefore$ $E(\hat{x}) = \frac{\hat{\alpha}}{\hat{\beta}}$ and $Var(\hat{x}) = \frac{\hat{\alpha}}{\hat{\beta}^2}$

what is \hat{x}

Let's assume that, ~~sample moments equal~~

× $E(\hat{x}) = \bar{x}$ and $Var(\hat{x}) = s^2$

$\therefore$ $\bar{x} = \frac{\hat{\alpha}}{\hat{\beta}} \longrightarrow \hat{\alpha} = \bar{x}\hat{\beta}$ — (1)

From $Var(\hat{x}) \longrightarrow s^2\hat{\beta}^2 = \hat{\beta}\bar{x}$

$$\hat{\beta} = \frac{\bar{x}}{s^2} \quad \text{— (2)}$$

Now substitude $\hat{\beta}$ back into the (1) eq.

from (1), $\hat{\alpha} = \bar{x}\left(\frac{\bar{x}}{s^2}\right) = \frac{\bar{x}^2}{s^2}$ ✓ $\quad \hat{\beta} = \frac{\bar{x}}{s^2}$ ✓

These are the method of moments estimators for the parameters α and β of the gamma distribution based on the sample mean $\bar{x}$ and variance s^2.

Scanned with CamScanner

(b) Log likelihood function. $l(\alpha,\beta)$ for a gamma distribution is given by,

$$l(\alpha,\beta) = \sum_{i=1}^{n} \left[(\alpha-1)\log(X_i) - \beta X_i - \log(\Gamma(\alpha)) \right]$$

some terms missing

Then we need to take the partial derivatives of l with respect to α and β.

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n} \left[\log(X_i) - \psi(\alpha) \right]$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{\alpha}{\beta} - X_i \right]$$

And, we need to set the partial derivatives equal to zero to find the maximum.

$$\sum_{i=1}^{n} \left[\log(X_i) - \psi(\alpha) \right] = 0$$

$$\sum_{i=1}^{n} \left(\alpha/\beta - X_i \right) = 0$$

Solving this system of equations explicitly may not be straightforward due to the presence of the digamma function, denoted as $\psi(\alpha)$, is the logarithmic derivative of the gamma distribution.

Scanned with CamScanner

Explicit solutions may be obtained for specific cases of the gamma distribution,

In general, solving $\hat{\alpha}$, $\hat{\beta}$ may require numerical method.

For a specific dataset, we can use numerical optimization techniques like gradient descent, Newton's method or other optimization algorithms.

Scanned with CamScanner