ITMR paper B solution

Time: 2 hours
Total Marks: 60 marks

Note:

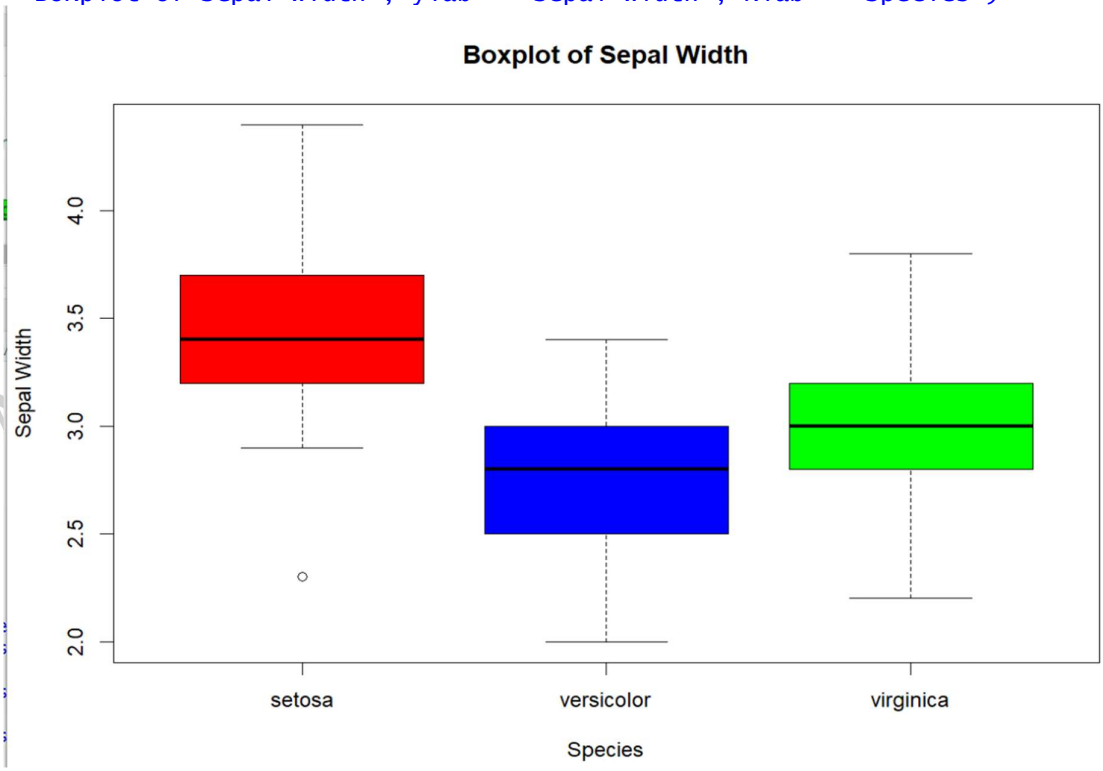
1. **The candidate has option to either attempt question 4A or question 4B. Rest all questions are mandatory.**
2. **Numbers to the right indicate full marks.**
3. **The candidates will be provided with the formula sheet and graph papers (if required) for the examination.**
4. **Use of approved scientific calculator is allowed.**

Q1 A **5 Marks**

```
> #Q1
> #A
> #(a)
> mu = 10000
> sigma = sqrt(90000)
> pnorm(10600,mu,sigma,lower.tail = FALSE)
[1] 0.02275013
> #(b)
> qnorm(0.99,mu,sigma)
[1] 10697.9
> #(c)
> var(rnorm(10,mu,sigma))
[1] 79647.27
>
> #The variance of the sample is quite lower compared to the variance of 9
0000 provided to us.
```

Q1 B **5 Marks**

```
> #(B)
> boxplot(Sepal.Width~Species, data = iris, col = c("red","blue","green"),
main = "Boxplot of Sepal Width", ylab = "Sepal Width", xlab = "Species")
```



Comment: Setosa has the highest Sepal Width and versicolor has the lowest. The

spread virginica is the lowest. Both Setosa and virginica seem to have a symmetric distribution but versicolor is slightly negatively skewed.

Q1 C **5 Marks**

```
> #(C)
> #(a)
> x = c(4,5,6,7,8,9)
> obs = c(3,9,18,29,25,16)
> p = 1/(sum(x*obs)/sum(obs))
> p
[1] 0.1404494
> exp = sum(obs)*c(pgeom(4,p),dgeom(5:8,p),pgeom(8,p,lower = FALSE))
> chisq_cal = sum((obs-exp)^2/exp)
> df = length(x) - 1 - 1#-1 fore estimating p
> pchisq(chisq_cal,df,lower = FALSE)
[1] 4.578713e-64
> #Since the p-value is very small we can conclude that Geometric is not a
good fit.
```

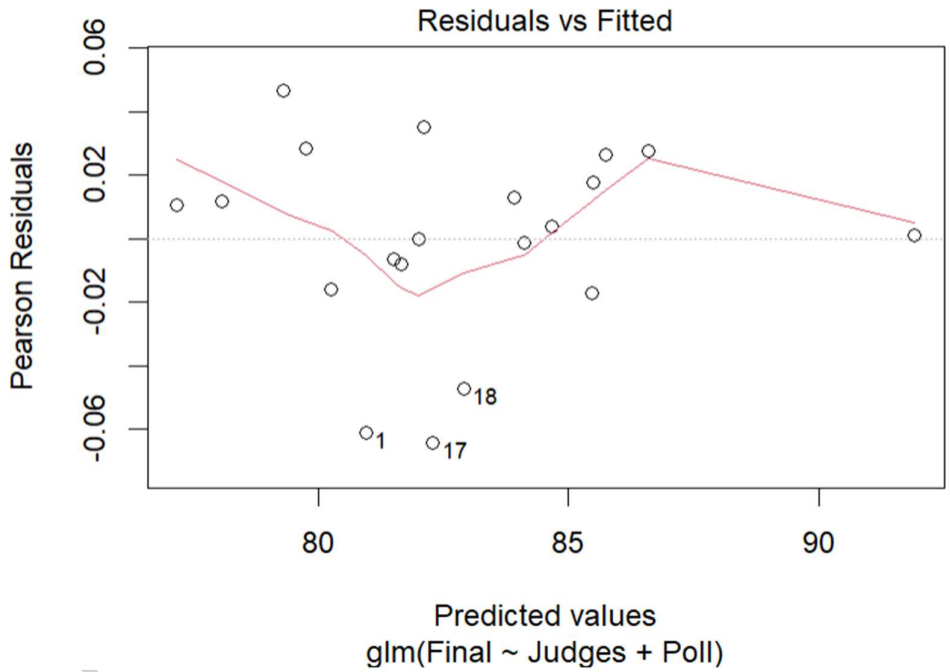
Q2 A **5 Marks**

```
> #Q2
> #A
> #(a)
> dance = read.csv("dance.csv")
> fitA = glm(Final~Judges+Poll, family = Gamma("identity"), data = dance)
> fitA

Call:  glm(formula = Final ~ Judges + Poll, family = Gamma("identity"),
    data = dance)

Coefficients:
(Intercept)       Judges         Poll
    11.2205       0.7058       0.1518

Degrees of Freedom: 19 Total (i.e. Null);  17 Residual
Null Deviance:      0.04907
Residual Deviance: 0.01746     AIC: 100.5
> #(b)
> plot(fitA,1)
```



There is some pattern between the Pearson residuals and the fitted value which needs further investigation. The outliers (1,17,18) observations might also need some evaluation.

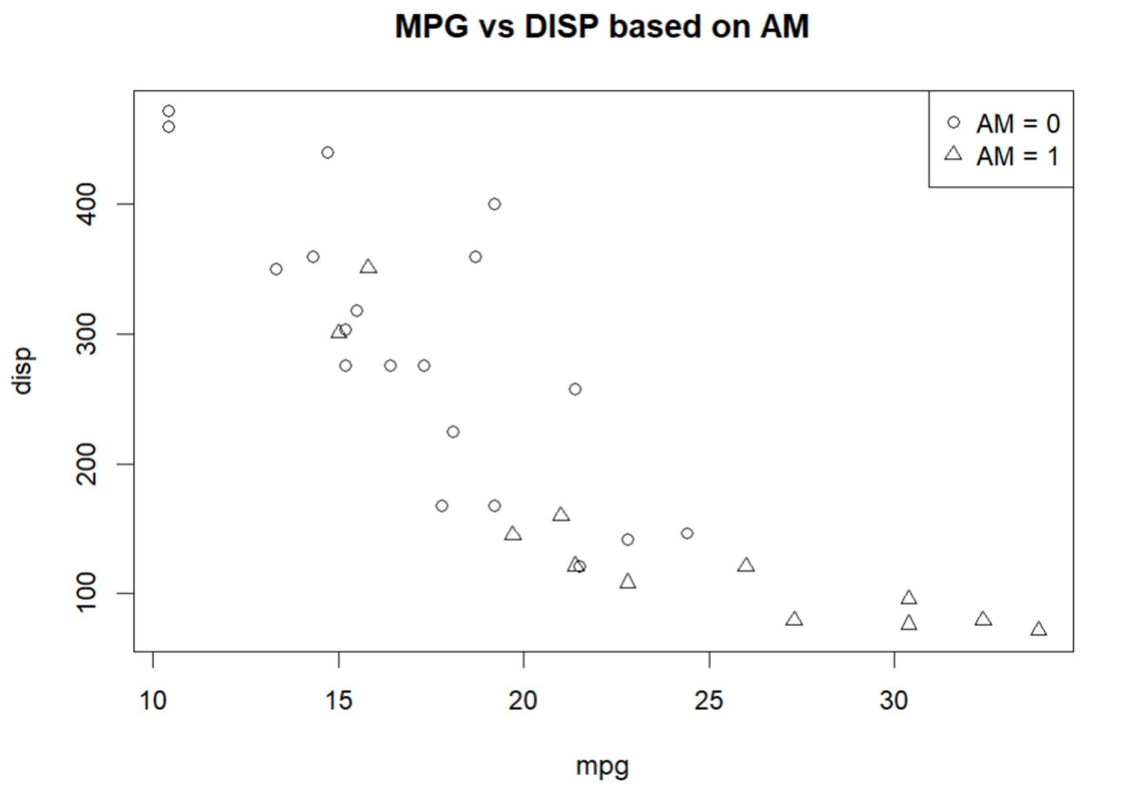
Q2 B **5 Marks**

```
> #(B)
> data = read.csv("data_Q2B_PaperB.csv")
> weight = data$Weight
> n = length(weight)
> S = sd(weight)
> sigma_0 = sqrt(121)
> chisq_cal = (n-1)*S^2/sigma_0^2
> chisq_cal
[1] 87.44421
> pchisq(chisq_cal,n-1,lower.tail = FALSE)
```

```
[1] 0.790362
```

Q2 C **5 Marks**

```
> #(C)
> #(a)
> aggregate(mpg~am+vs, data = mtcars, FUN = mean)
  am vs      mpg
1  0  0 15.05000
2  1  0 19.75000
3  0  1 20.74286
4  1  1 28.37143
> #(b)
> plot(disp~mpg, data = mtcars, pch = am+1, main = "MPG vs DISP based on A
M")
> legend("topright", legend = c("AM = 0","AM = 1"), pch = 1:2)
```



Q3 A **30 Marks**

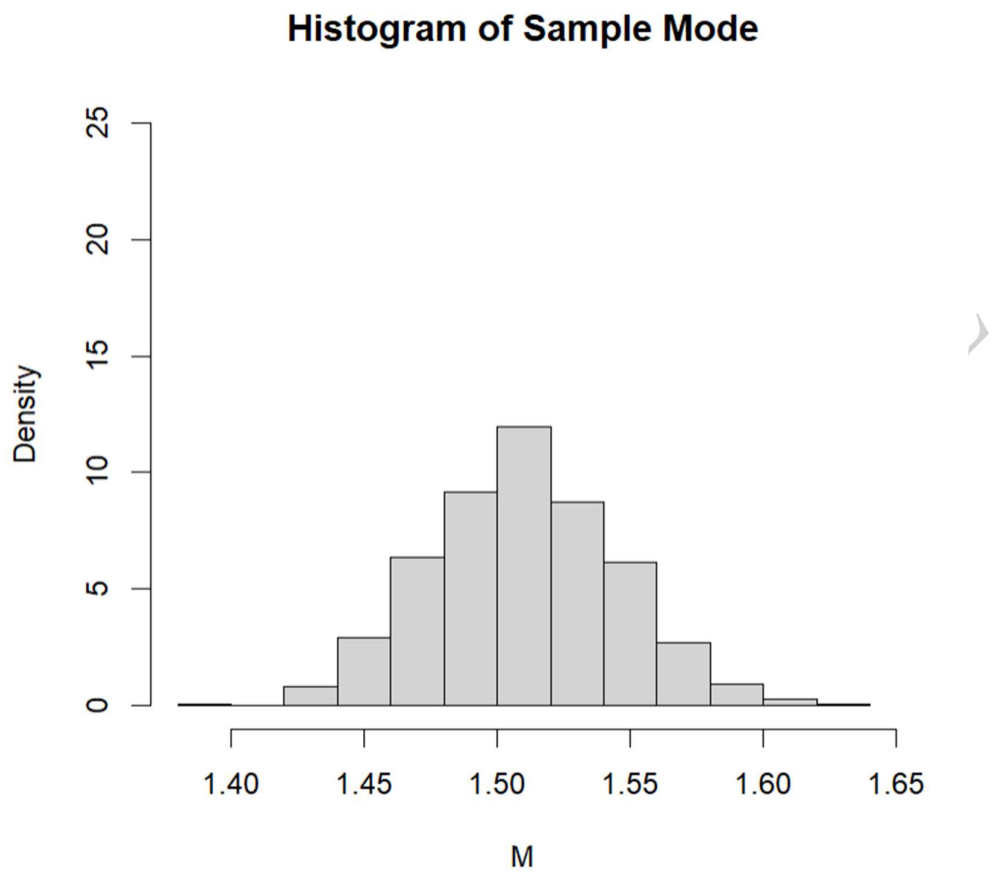
```
> #Q3
> #A
> mu = 1.512
> sigma = 0.0741
>
> #(a)
> mu
[1] 1.512
> sigma/sqrt(15)
[1] 0.01913254
```

The distribution would be N(1.512,0.019^2)

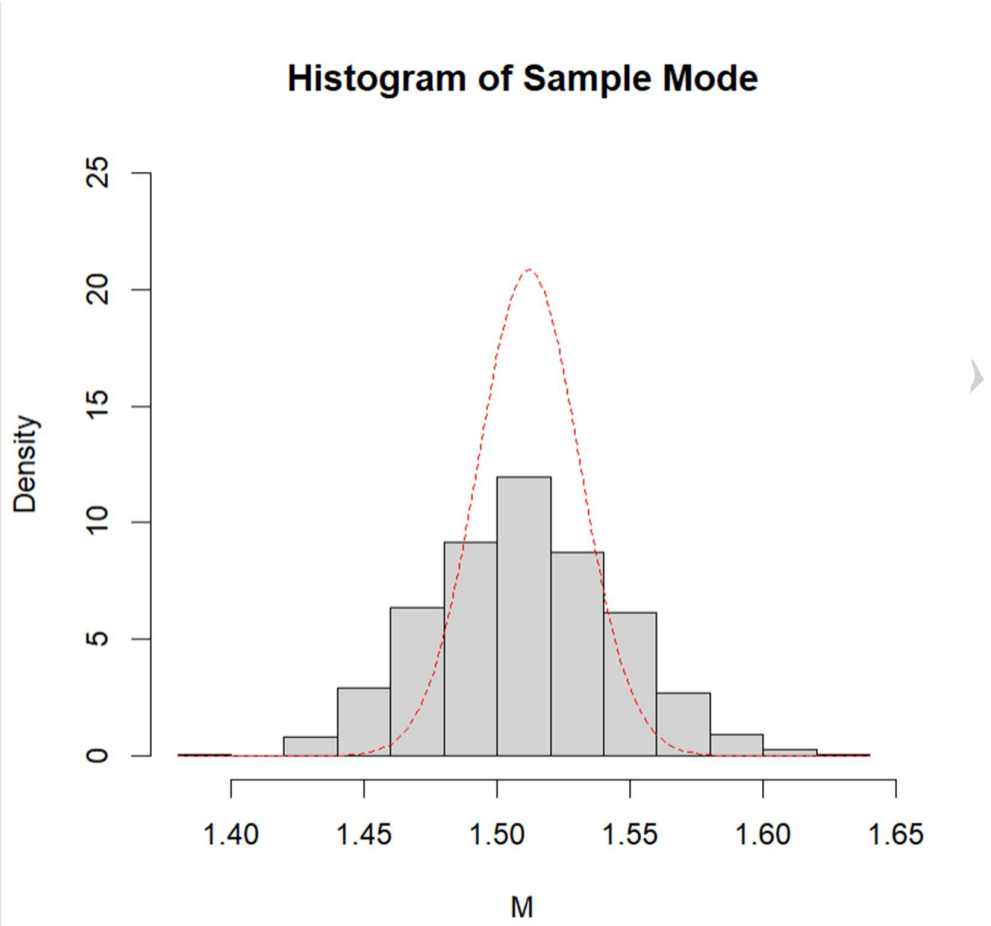
```
> #(b)
> n = 15
> set.seed(2024)
> y = rnorm(n,mu,sigma)

> #(c)
> dens = density(y)
> M = dens$x[which.max(dens$y)]
> M
[1] 1.549025
> #(d)
> M = numeric(1000)
> set.seed(2024)
> for(i in 1:1000){
+   y = rnorm(n,mu,sigma)
+   dens = density(y)
+   M[i] = dens$x[which.max(dens$y)]
+ }

> #(e)
> hist(M, ylim = c(0,25), main = "Histogram of Sample Mode", freq = FALSE)
```

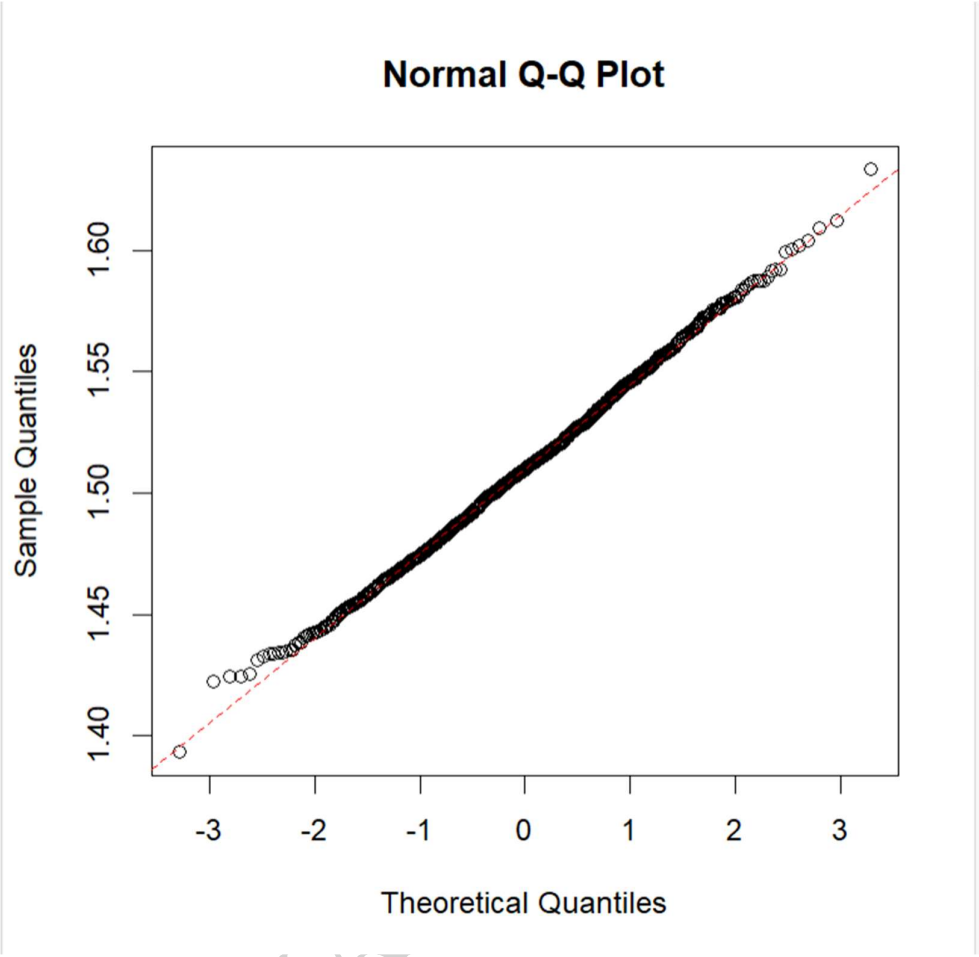



```
> #(f)
> curve(dnorm(x,mu,sigma/sqrt(n)),add = TRUE, col = "red",lty =2)
```



The sample mode does not follow the Normal distribution similar to that of the sample mean. The spread of the sample mode is higher than the mean and hence there is a high difference between the empirical distribution and the distribution given by the sample mean.

```
> #(h)
> qqnorm(M)
> qqline(M, lty = 2, col = "red")
```



The majority of the points seem to line up with the red line specifying a Normal fit. So the sample mode seems to be following a Normal distribution. There are small differences at the tails which should be studied further.

OR

Q3 B

(I)

```
> #Q3
> #B
> #I
> #(a)
> T1 = c(2.36,4.85,3.65,7.82,5.63,4.89,10.1,8.53,3.14,6.58)
> C1 = c(5.63,5.2,4.78,9.89,8.72,5.98,0.56,4.12,1.85,4.02)
> t.test(T1,C1, alt = "g",var.equal = TRUE)

        Two Sample t-test

data:  T1 and C1
t = 0.57376, df = 18, p-value = 0.2866
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.375136       Inf
sample estimates:
mean of x mean of y
```

```
    5.755     5.075

> #(b)
> var.test(T1,C1, alt = "t")

        F test to compare two variances

data:  T1 and C1
F = 0.79282, num df = 9, denom df = 9, p-value = 0.7351
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1969261 3.1919023
sample estimates:
ratio of variances
         0.7928234

> #(c)
> obs = mean(C1) - mean(T1)
> results = c(C1,T1)
> index = 1:length(results)
> p = combn(index,length(index)/2)
> dif = numeric(nrow(p))
> for(i in 1:nrow(p)){
+   dif[i] = mean(results[p[,i]]) - mean(results[-p[,i]])
+ }
> length(dif[dif<=obs])/length(dif)
[1] 0.4
```

(II)

```
> #II
> #(a)
> claims = 0:8
> obs = c(28,126,0,111,31,30,0,0,39)
> lambda = sum(claims*obs)/sum(obs)
> exp = sum(obs)*c(dpois(0:7,lambda),ppois(7,lambda,lower.tail = FALSE))
> chisq_cal = sum((obs-exp)^2/exp)
> pchisq(chisq_cal,length(obs)-1-1,lower  = FALSE)
[1] 2.615949e-122
>
> #(b)
> n = 8
> p = 0.358
> exp = sum(obs)*dbinom(0:8,n,p)
> chisq_Cal = sum((obs-exp)^2/exp)
> pchisq(chisq_Cal,length(obs)-1,lower = FALSE)
[1] 0
```