STAT416 Assignment 1

Ashish Jain

Question 1:

i). The treatment factors considered in the experiment is Nitrogen.

ii). The treatment level considered in the experiment for Nitrogen is high and low.

iii). Yes this is a full factorial experimental design.

iv). The experimental units in this experiment is pots. There are 12 pots.

v). The observational unit is the seedling.

vi). Yes this experimental involves blocking. The block is the genotype of the seedling.

vii). It is a randomized complete block design. This is due to the fact that first the blocks are created on the basis of genotype and after that each block is treated with both the levels of nitrogen.

Question 2:

There are different sources of variabilty in this experiment.

i). Biological Variabilty

Genotype Variabilty, and Natural variation among biological replicates.

ii). Technical Variabilty

Sequencing depth, lane effect, adapter effect, and library prepration effect.

iii). Treatment Effect

The variability in the amount of nitrogen used for the treatment.

iv). Block effect

We have two different genotypes which will cause the block effect.

Yes, the seedlings with same genotype and same treatment are the biological replicates. There is no technical replicates involved in this experiment.

Question 3:

$$H_0 : \ p_1 = p_2 \ vs \ H_1 : p_1 \neq p_2$$

where,p_1 is the probability of drawing reads for gene g in the first lane and p_2 is the probability of drawing reads for gene g in the second lane.

```
library("tidyverse")
table <- data.frame(Lane = c("Gene g count", "Total read Count"),
    `1` = c("50", "10M"), `2` = c("25", "8M"))
table %>% knitr::kable(caption = "Number of Reads")
```

Table 1: Number of Reads

Lane	X1	X2
Gene g count	50	25
Total read Count	10M	8M

```
FisherTestMatrix <- matrix(c(50, 1e+07, 25, 8e+06), nrow = 2,
    dimnames = list(gene = c("yes", "no"), lane = c("1", "2")))
fisher.test(FisherTestMatrix)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  FisherTestMatrix
## p-value = 0.06245
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.9712752 2.6992872
## sample estimates:
## odds ratio
##   1.599997
```

As, seen from the Fisher's exact test the P-Value is 0.06245 which is greater than 0.05, so we fail to reject the null hypothesis with 95% confidence and hence there is no evidence for lane effect in terms of mapping rate.

Question 5:

i). The Hyopthesis test is:

$$H_0 : T.S. \ \sim \ \chi^2_{J-1} \ VS \ H_1 : T.S. \ \nsim \ \chi^2_{J-1}$$

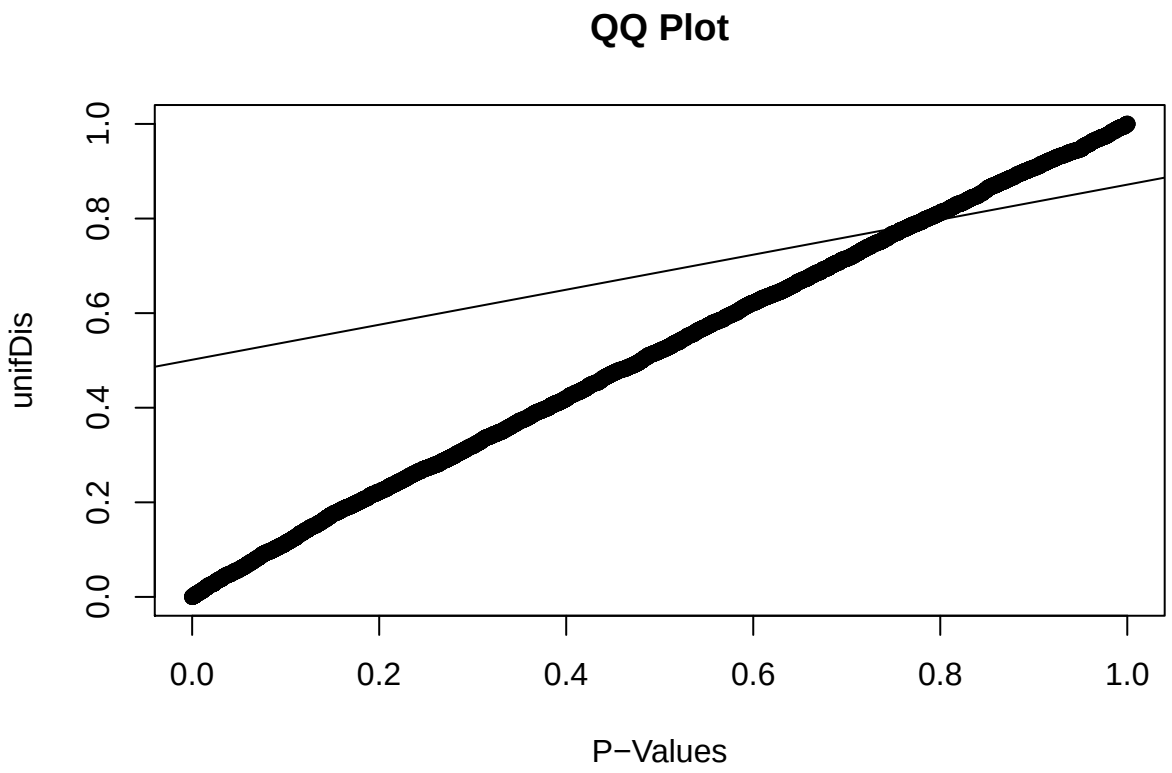
where,

$$T.S. = \sum \frac{Y_{gj} - C_j \hat{\mu_g}^2}{C_j \hat{\mu_g}}$$

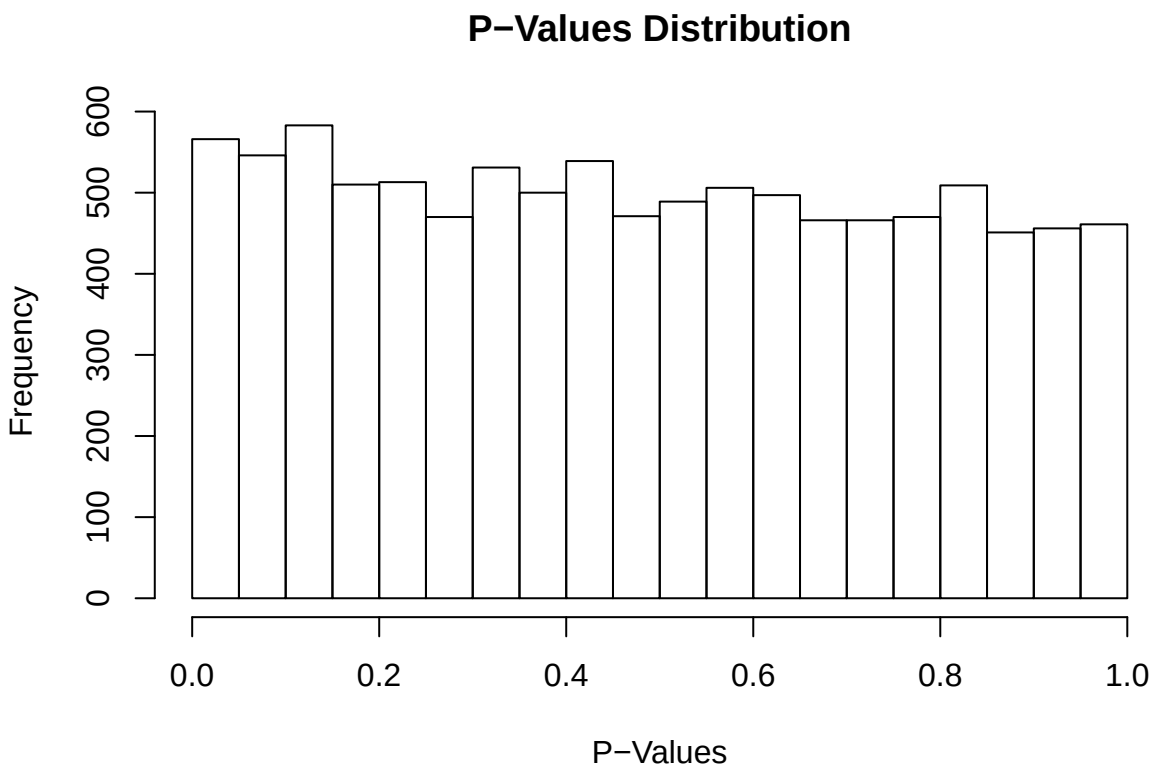
$J-1$ is the number of independent observations – the number of estimated parameters, j is the lane number, $\hat{\mu_g}$ is the estimated mean.

```r
library("MASS")
path <- "https://raw.githubusercontent.com/ashishjain1988/STAT416/master/HW1/"
countData <- read.table(paste0(path, "hwk1_4.csv"), header = TRUE,
    sep = ",")
totalCount <- sum(as.matrix(countData))
countLanes <- apply(as.matrix(countData), 2, function(x) {
    return(sum(x))
})

pValues = c()
pValues <- apply(countData, 1, function(x) {
    mu <- sum(x)/totalCount
    t <- sum(((x - mu * countLanes)^2)/(mu * countLanes))
    return(1 - pchisq(t, 3))
})
unifDis <- runif(10000)
qqplot(pValues, unifDis, plot.it = TRUE, xlab = "P-Values", main = "QQ Plot")
qqline(unifDis)
```




```r
# data.frame(pval=pValues) %>% ggplot(aes(sample = pval)) +
# stat_qq()
hist(pValues, xlab = "P-Values", main = "P-Values Distribution")
```




```
# data.frame(pval=pValues) %>% ggplot(aes(x=pval))
# +geom_histogram()
ks.test(pValues, "punif", 0, 1)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  pValues
## D = 0.02654, p-value = 1.524e-06
## alternative hypothesis: two-sided
```

We used the ks.test to check whether the p-values follows that U(0,1) or not. As you can see, the P-value is significant and from that we can reject the null hypothesis which means that the P-Values do not follow U(0,1). From this we can conclude that the data do not follow poisson distribution.

Question 4:

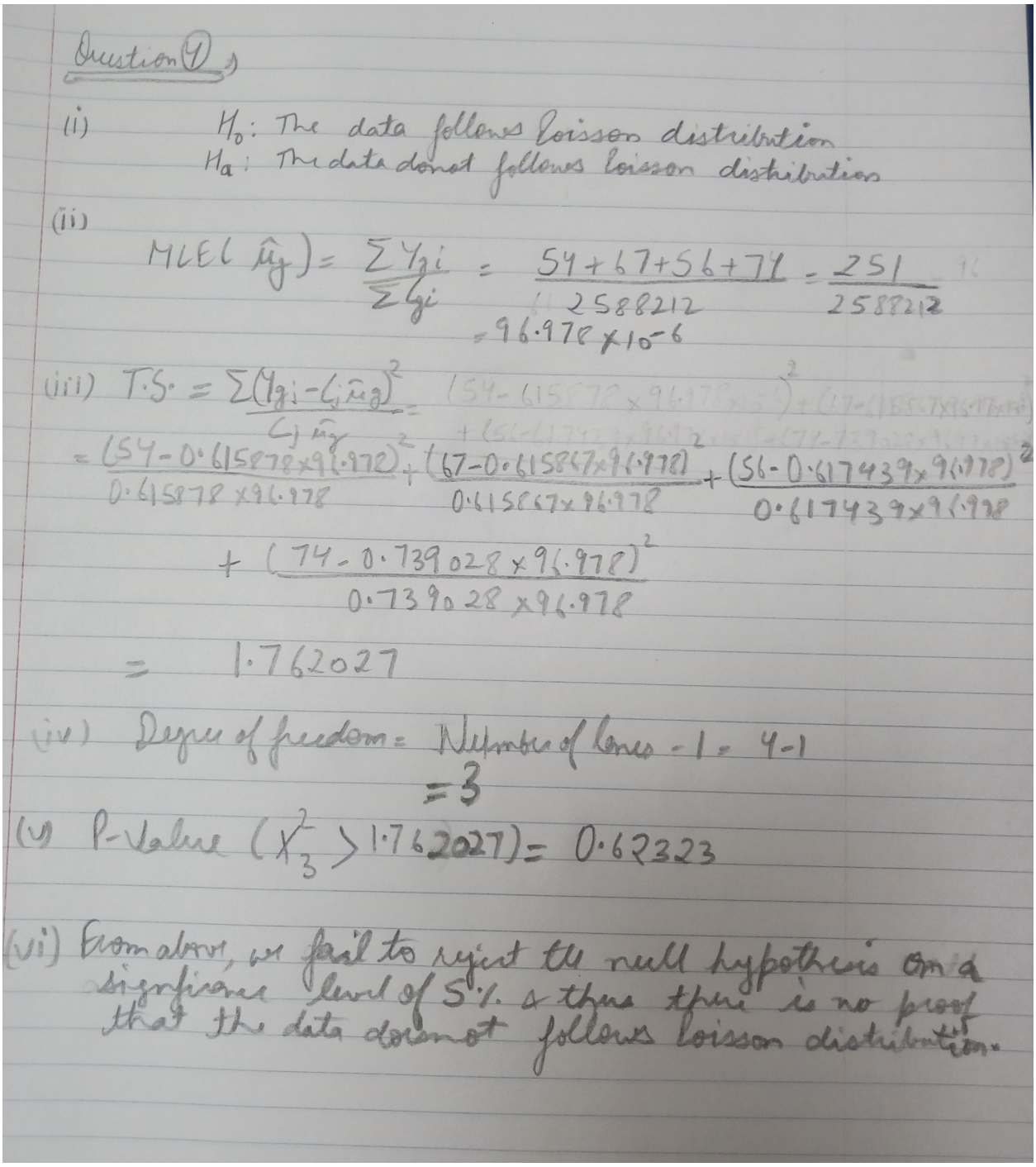


Figure 1: Question 4