

Lab#06

Rezaur Rashid

2022-04-16

Problem (A): One-way ANOVA Model

```
### LAB-06 ----- part-A -----

myT<-read.table("data/longitudinalRNASeqData/nc101_scaff_dataCounts.txt",sep="\t",header=TRUE,row.names=)

# remove rare genes
myT <- myT[ apply( myT,1, median)> 5,]

myTNorm <- myT
for ( i in 1:ncol(myT))
{
  colSum = sum(myT[,i])
  myTNorm[,i] =myTNorm[,i]/colSum
}

p_values_A = vector()

for (i in 1:nrow(myTNorm)) {

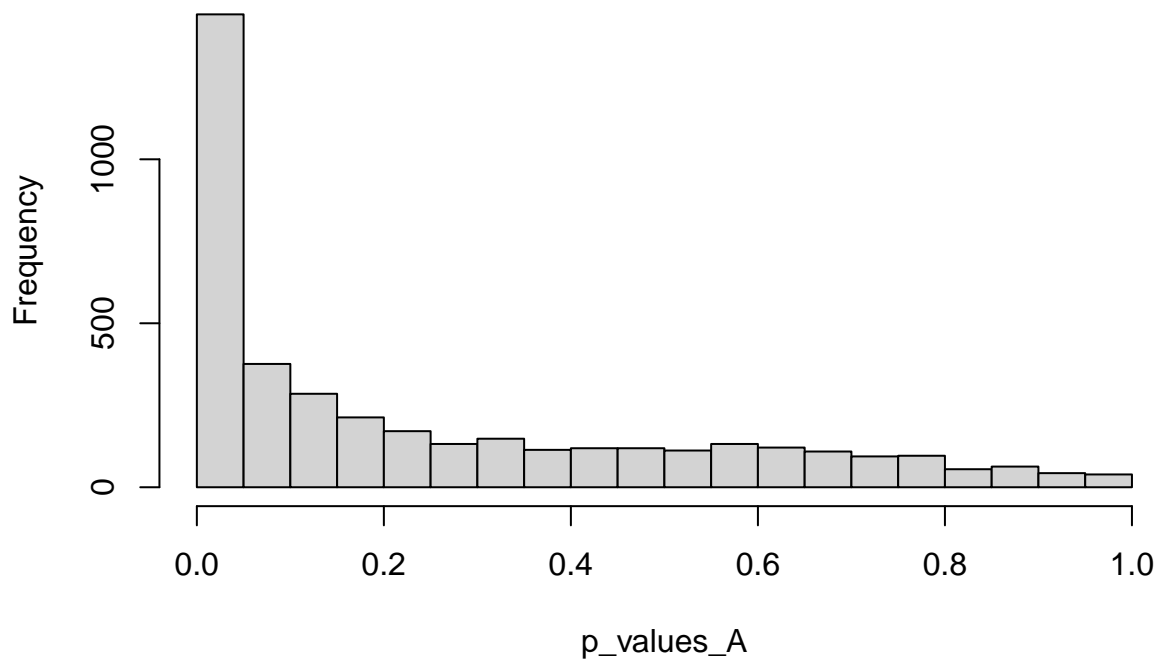
  Day_2 = as.numeric (myTNorm [i, 1:3])
  Week_12 = as.numeric(myTNorm[i, 4:6])
  Week_18 = as.numeric(myTNorm[i, 7:11])

  myData = c(Day_2, Week_12, Week_18)

  genotypes = c(rep('Day_2', length(Day_2)), rep('Week_12', length(Week_12)),
                rep('Week_18', length(Week_18)))
  genotypes = factor(genotypes)
  myLm = lm(myData ~ genotypes, x=TRUE)
  pVal = anova(myLm)$"Pr(>F)"[1]
  p_values_A[i] = pVal
}

hist(p_values_A, breaks = 25, main = 'Regression as category')
```

Regression as category



```
pValue_adj_A = p.adjust(p_values_A, method = 'BH')
paste('Significant #genes: ', sum(pValue_adj_A <= 0.05))
```

```
## [1] "Significant #genes: 612"
```

Problem (B): Regression Model

```
### LAB-06 ----- part-B -----

p_values_B = vector()

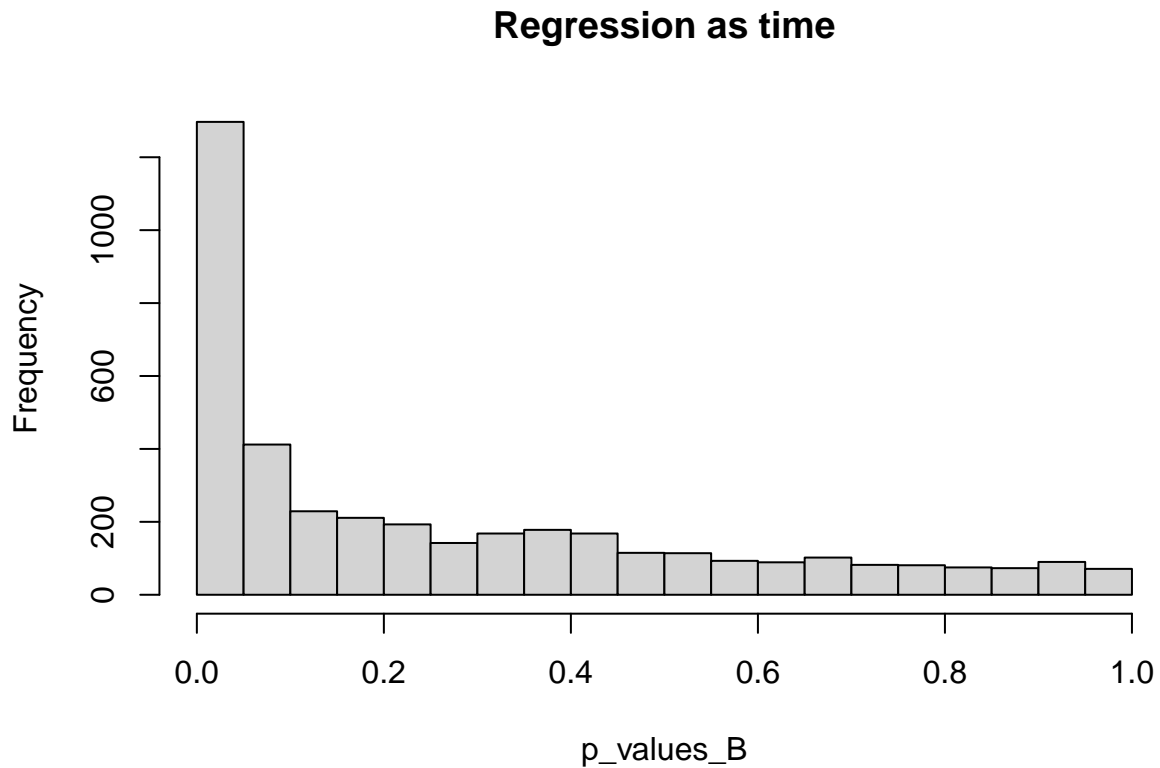
for (i in 1:nrow(myTNorm)) {

  Day_2 = as.numeric(myTNorm[i, 1:3])
  Week_12 = as.numeric(myTNorm[i, 4:6])
  Week_18 = as.numeric(myTNorm[i, 7:11])

  myData = c(Day_2, Week_12, Week_18)

  genotypes = c(rep(2, length(Day_2)), rep(86, length(Week_12)),
                rep(128, length(Week_18)))
  myLm = lm(myData ~ genotypes, x=TRUE)
  pVal = anova(myLm)$"Pr(>F)"[1]
  p_values_B[i] = pVal
}
```

```
hist(p_values_B, breaks = 25, main = 'Regression as time')
```



```
pValue_adj_B = p.adjust(p_values_B, method = 'BH')
paste('Significant #genes: ', sum(pValue_adj_B <= 0.05))
```

```
## [1] "Significant #genes: 448"
```

Problem (C): Model Difference

```
### LAB-06 ----- part-C -----

p_values_C = vector()

for (i in 1:nrow(myTNorm)) {

  Day_2 = as.numeric (myTNorm [i, 1:3])
  Week_12 = as.numeric(myTNorm[i, 4:6])
  Week_18 = as.numeric(myTNorm[i, 7:11])

  myData = c(Day_2, Week_12, Week_18)

  genotypes = c(rep('Day_2', length(Day_2)), rep('Week_12', length(Week_12)),
                rep('Week_18', length(Week_18)))
  genotypes = factor(genotypes)
  myLm = lm(myData ~ genotypes, x=TRUE)
  full_residuals = sum(residuals(myLm)^2)
```

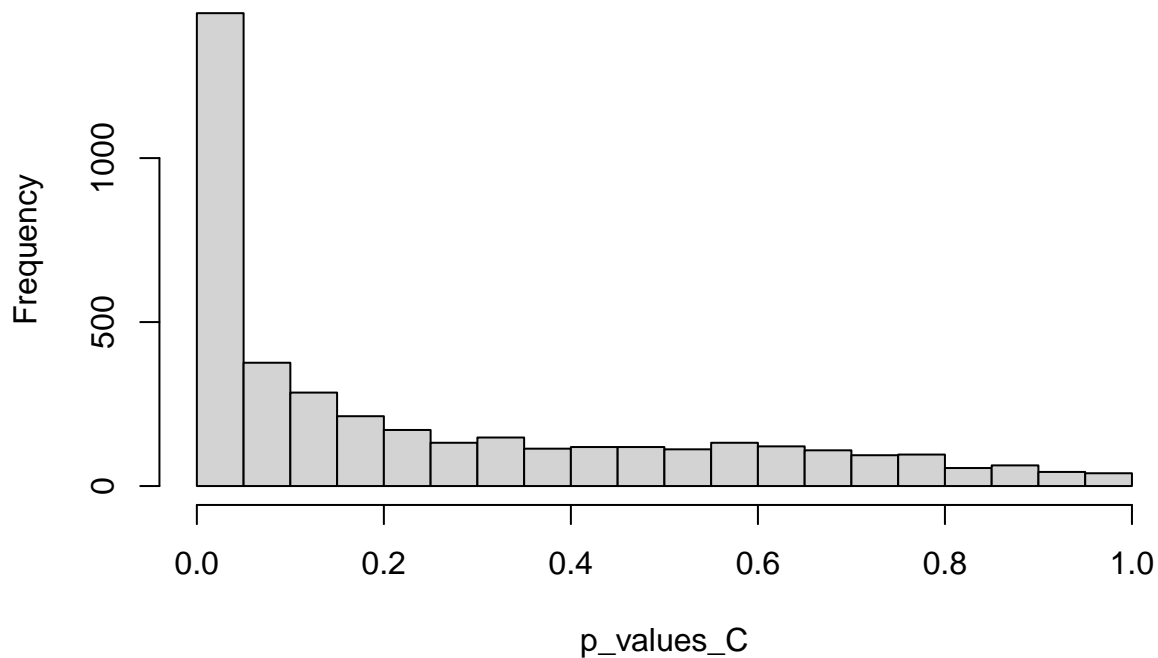
```

reduced_residuals = sum((myData - mean(myData))^2)
myF = ((reduced_residuals - full_residuals)/(10-8)/(full_residuals/8))
pVal = pf(myF,2,8, lower.tail = FALSE)
p_values_C[i] = pVal
}

hist(p_values_C, breaks = 25, main = 'Regression as time')

```

Regression as time



```

pValue_adj_C = p.adjust(p_values_C, method = 'BH')
paste('Significant #genes: ', sum(pValue_adj_C <= 0.05))

```

```
## [1] "Significant #genes: 612"
```

Problem (D): Graph: Box Plot

```
### LAB-06 ----- part-D -----
```

```

index = 1:length(p_values_A)
length(index)

```

```
## [1] 3983
```

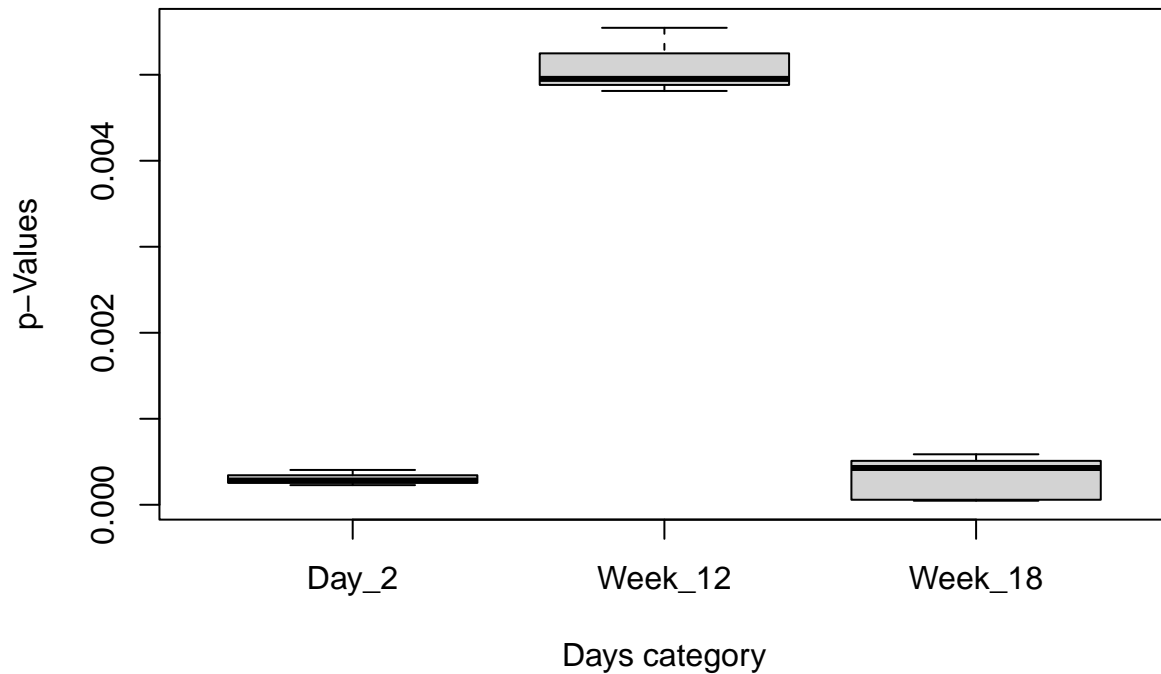
```

genotypes_A_C = c(rep('Day_2', 3), rep('Week_12', 3),
                  rep('Week_18', 5))
cats_A_C = factor(genotypes_A_C)
cats_B = c(rep(2, 3), rep(86, 3),
           rep(128, 5))

```

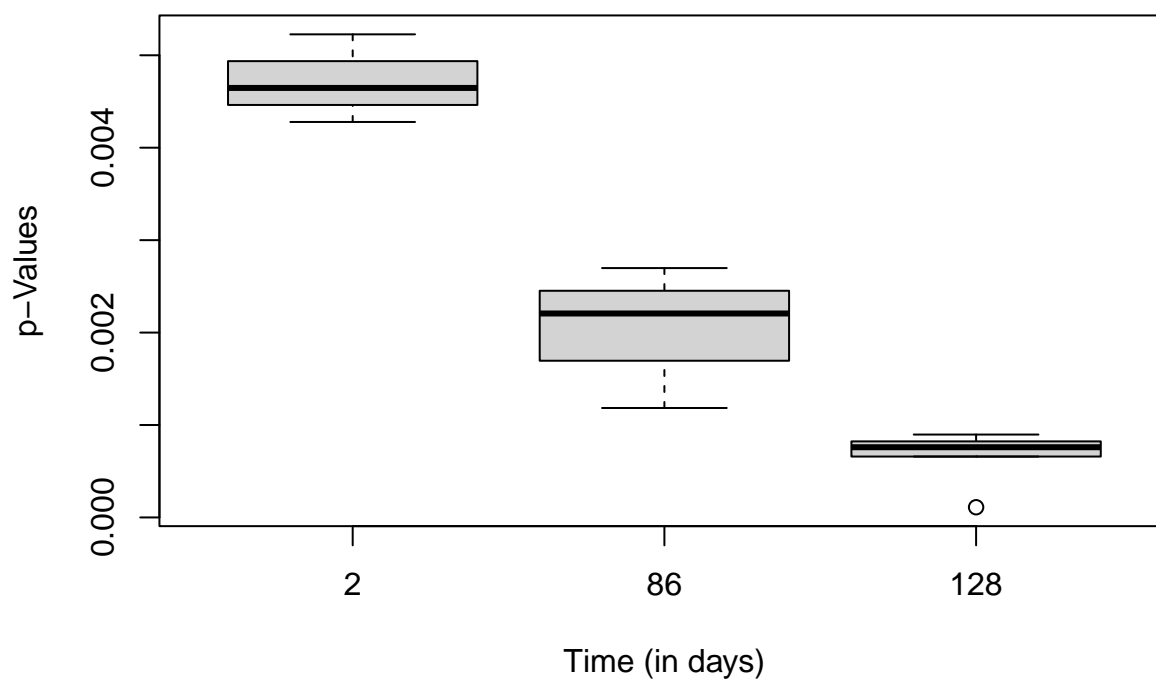
```
# box plotting graph for (A)
myFrame <- data.frame( index, p_values_A,p_values_B,p_values_C)
myFrame <- myFrame[ order(myFrame$p_values_A), ]
boxplot( as.numeric( myTNorm[ myFrame$index[1],]) ~ cats_A_C, main = 'Graph: Oneway ANOVA model',
        xlab = 'Days category', ylab = 'p-Values')
```

Graph: Oneway ANOVA model



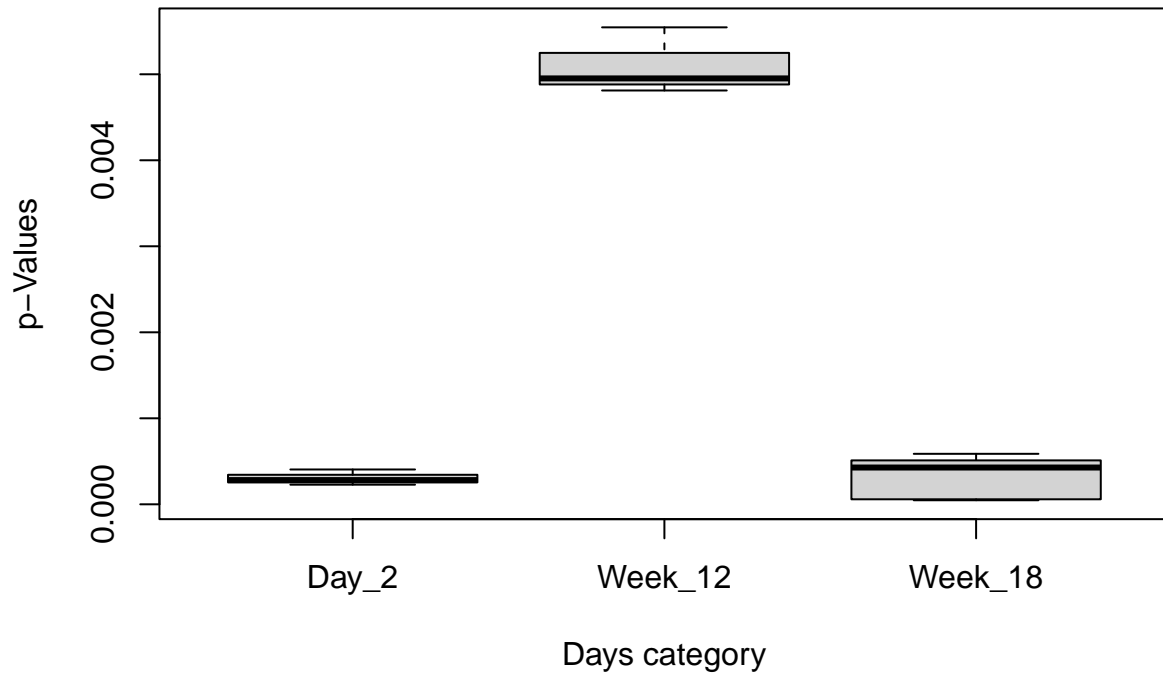
```
# box plotting graph for (B)
myFrame <- data.frame( index, p_values_A,p_values_B,p_values_C)
myFrame <- myFrame[ order(myFrame$p_values_B), ]
boxplot( as.numeric( myTNorm[ myFrame$index[1],]) ~ cats_B, main = 'Graph: Regression Model',
        xlab = 'Time (in days)', ylab = 'p-Values')
```

Graph: Regression Model



```
# box plotting graph for (C)
myFrame <- data.frame( index, p_values_A,p_values_B,p_values_C)
myFrame <- myFrame[ order(myFrame$p_values_C), ]
boxplot( as.numeric( myTNorm[ myFrame$index[1],]) ~ cats_A_C, main = 'Graph: Model Difference',
        xlab = 'Days category', ylab = 'p-Values')
```

Graph: Model Difference



Problem (E): Model Evaluation

Overall, we think the two-parameter reduced model is appropriate for these data. For these data, we think the interceptor is not that significant compared to the slope therefore we can assume that the interception term is zero and adding more degrees of freedom doesn't produce better results. Therefore, a reduced model is more appropriate.