

Linear Models PEP Part 2: Fever

Due: 11:55 pm Friday 27 October 2023

In continuing with the study by Zhang et al. (2015), focusing on the occurrence of fever as an outcome, the analysis aimed to predict fever using six potential predictors: group (control or treatment), gender, age, BMI, duration of surgery, and maximum temperature on the first day post-surgery. The dataset was read into R from the file PEP1.csv. Given the binary nature of the outcome variable (fever: yes or no), a logistic regression model was considered appropriate for prediction. Graphical and numerical summaries for each variable were obtained and displayed compactly using the gridExtra library.

For each categorical predictor, reference levels were chosen based on logical or prevalent categories to facilitate interpretation. A scatterplot matrix of numeric variables was generated to assess correlations and inform whether all numeric predictors should be included in the logistic regression model. The association between categorical predictors was examined using chi-squared or Fisher's exact test, depending on assumptions, with results tabulated as p-values.

Relationships between numeric and categorical predictors were investigated using statistical graphics and formally tested with the median test, with p-values tabulated and a Bonferroni correction applied to determine significance. This step helped identify which predictor pairs should not be included together in the logistic regression model.

Model building using the AIC criterion identified the best additive model, with all fitted models and their AIC values listed in a table. The goodness of fit for the selected model was assessed, and the final model was written down, with each term explained. The final model was interpreted to understand the relationship between the predictors and the likelihood of fever, providing insights into which factors significantly influence postoperative fever in patients.

```
#read the dataset
pep1 <- read.csv("PEP1.csv", header = TRUE)
```

Linear regression is not appropriate because the response variable of interest, Fever, is binary with only values of 1,2. Therefore, it will be logical to consider the use of logistic regression which is best used for categorical variables such as Fever.

```
pep1$Fever<-as.factor(pep1$Fever)
pep1$Group<-as.factor(pep1$Group)
```

```
## Warning in Ops.factor(pep1$Group, as.factor(pep1$Group)): '-' not meaningful
## for factors
```

```
##      [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##     [26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##     [51] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##     [76] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [101] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [126] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [151] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [176] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##   [201] NA NA NA
```

```
p1<-ggplot(pep1, aes(x = Group, fill = Fever)) +
  geom_bar()
```

```
pep1$Gender<-as.factor(pep1$Gender)
```

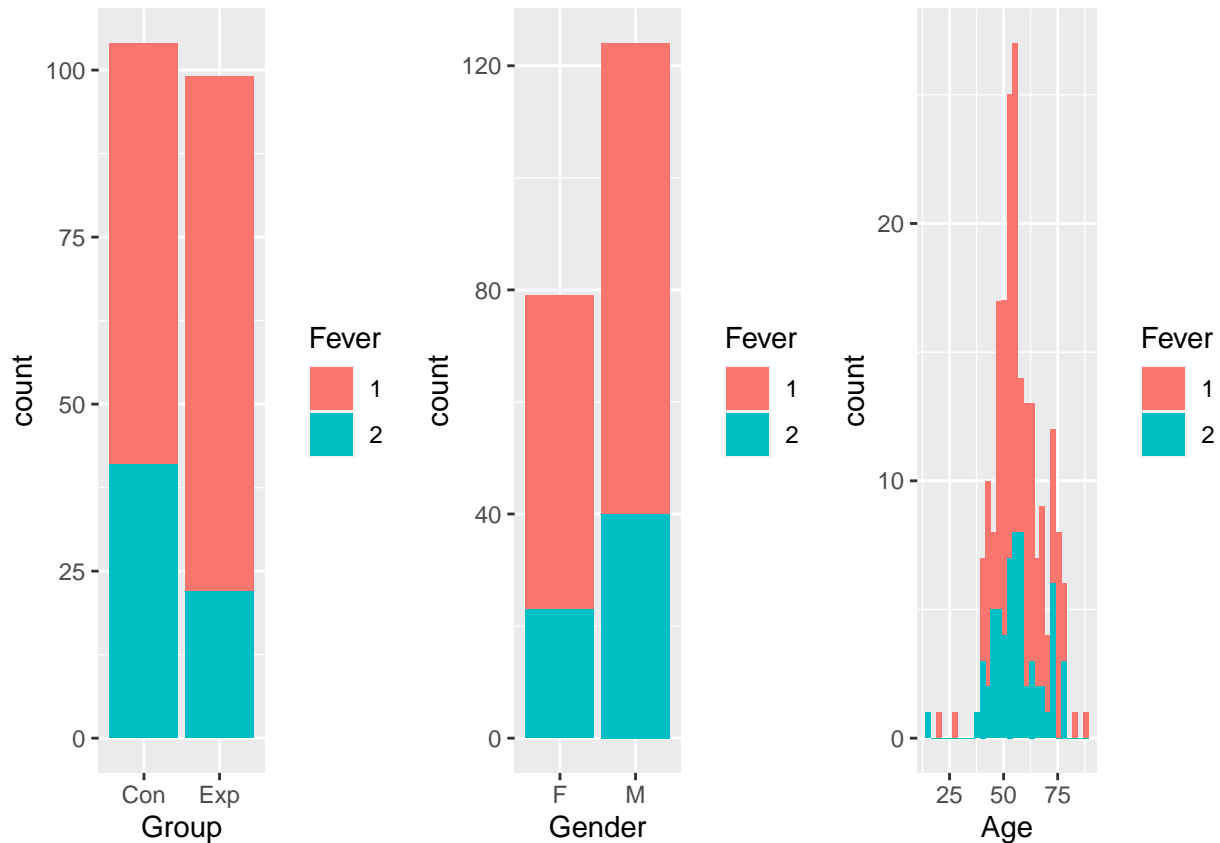
```
## Warning in Ops.factor(pep1$Gender, as.factor(pep1$Gender)): '-' not meaningful
## for factors
```

```
##      [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##     [26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##     [51] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##     [76] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [101] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [126] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [151] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [176] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
##    [201] NA NA NA
```

```
p2<-ggplot(pep1, aes(x = Gender, fill = Fever)) +
  geom_bar()
```

```
p3<-ggplot(pep1, aes(x = Age, fill = Fever)) +
  geom_histogram()
grid.arrange(p1,p2,p3, ncol = 3)
```

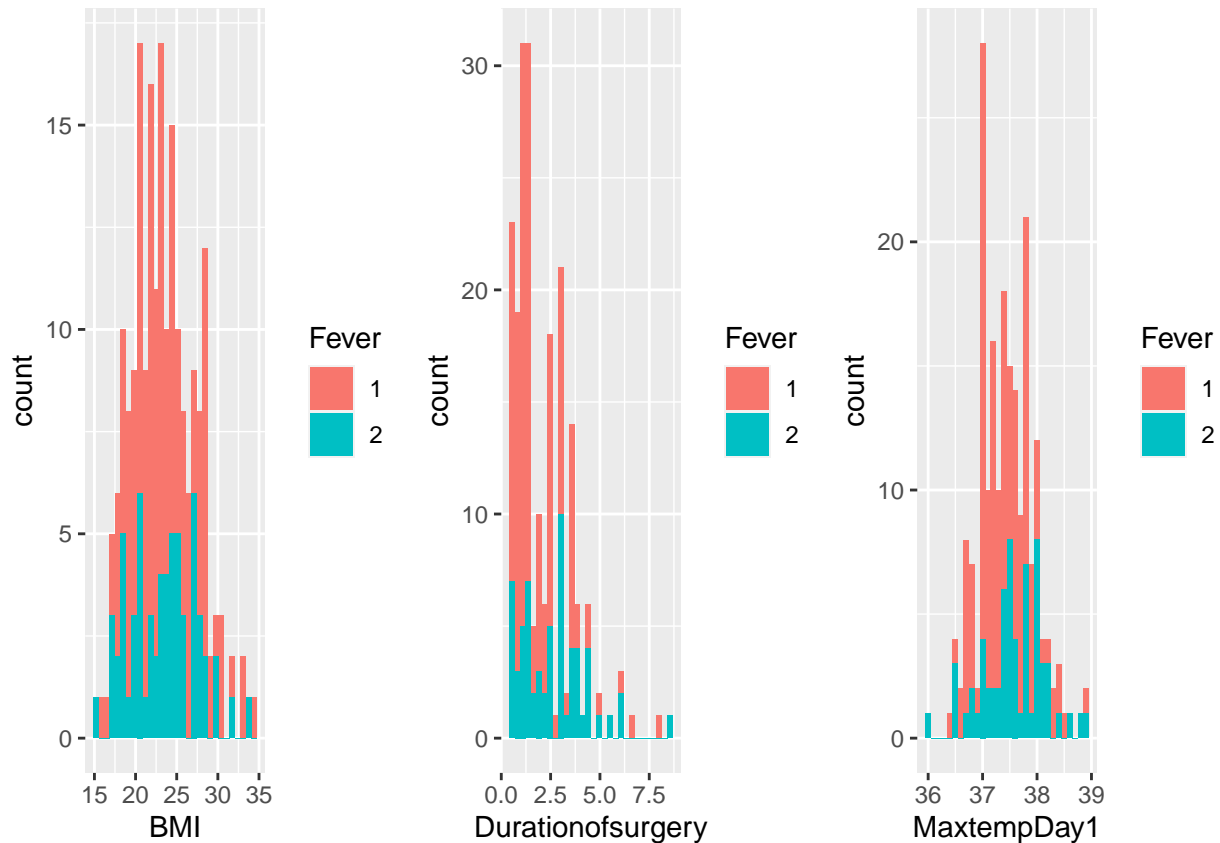
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Distribution of Group is relatively close for both groups and can be considered balanced. For Gender, however, it seems that there are more Males than Females. Distribution of Age is concentrated at the Age of 37-73 and looks approximately normally distributed. For these 3 variables, there are no clear associations with Fever yet since both Fever values of 0 and 1 appear to be proportional for every category and for age.

```
p4<-ggplot(pep1, aes(x = BMI, fill = Fever)) +
  geom_histogram()
p5<-ggplot(pep1, aes(x = Durationofsurgery, fill = Fever)) +
  geom_histogram()
p6<-ggplot(pep1, aes(x = MaxtempDay1, fill = Fever)) +
  geom_histogram()
grid.arrange(p4,p5,p6, ncol=3)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



For BMI, Durationof-surgery and MaxtempDay1 variables, there are no clear associations with Fever yet since both Fever values of 0 and 1 appear to be spread out across all numerical values of these variables. Moreover, for both BMI and MaxtempDay1, the distribution appears to be approximately normal. For Durationof-surgery on the other hand, distribution is right-skewed, and in an ideal setting, it is best to have this variable transformed by using log or square root transformations.

```
table(pep1$Group)
```

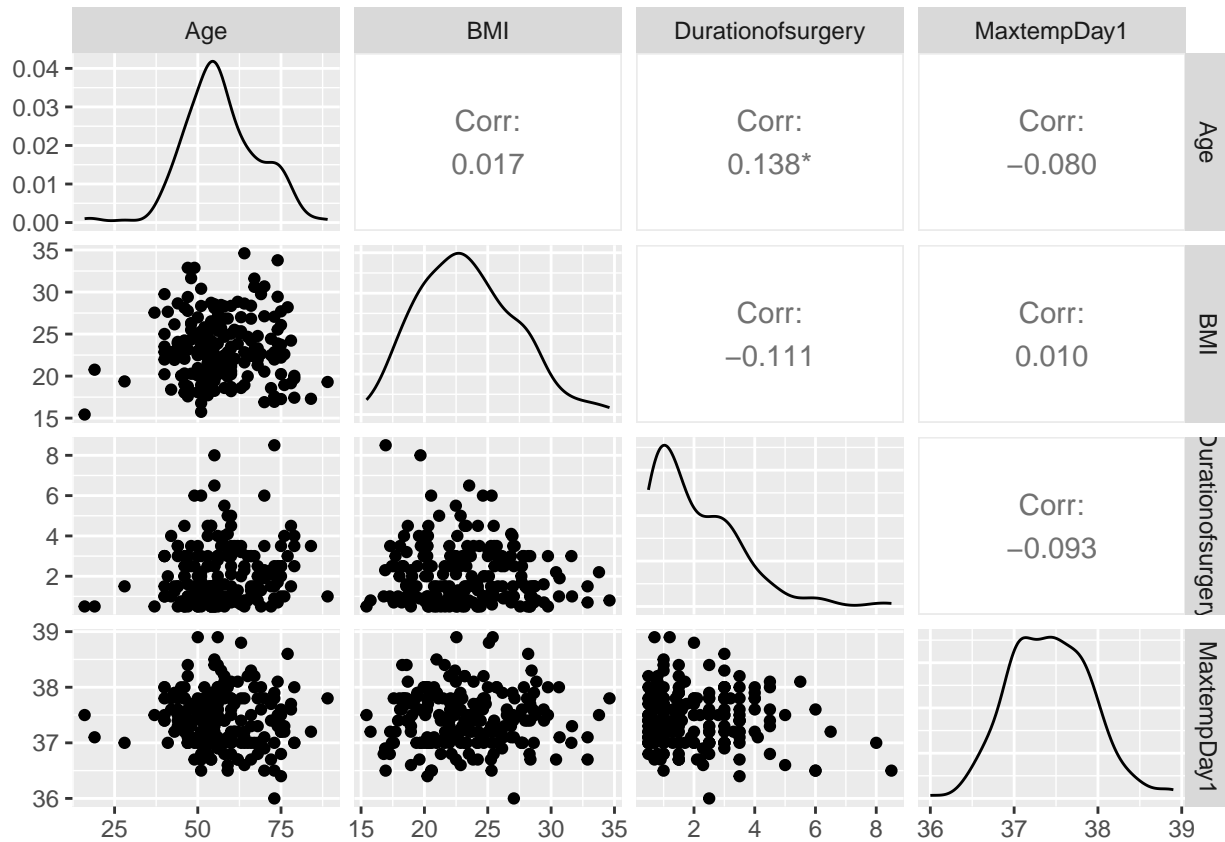
```
##
##  Con Exp
## 104  99
```

```
table(pep1$Gender)
```

```
##
##   F   M
##  79 124
```

Looking at the tables above, we can confirm that Group is relatively balanced but for Gender there are more Males than Females. For Group, it is ideal to use Con as the reference category since we have more records with Con category. For Gender on the other hand, the reference category should M for Male because it is significantly more than the Females from our dataset.

```
ggpairs(pep1, columns = c(3, 4, 5, 6))
```



All numerical variables have no apparent relationship since they all have weak relationships with one another. Therefore, all numerical variables can appear in the logistic regression model to be fitted since there is no evidence of multicollinearity.

(**HINT:** The 'xtabs()' function allows you to create crosstabulations using formula style input. Use

For the first part, we will be investigating whether the two categorical predictors, Group and Gender, are associated using a statistical test with the null and alternative hypotheses as stated below:

Ho: there is no significant association between the two categorical variables Group and Gender

Ha: there is a significant association between the two categorical variables Group and Gender

```
#Created crosstabulations on the categorical predictors
cat_table <- xtabs(~ Group + Gender, data = pep1)
cat_table
```

```
##      Gender
## Group  F  M
##   Con 41 63
##   Exp 38 61
```

```
#performed chi-test and got expected values
chisq<-chisq.test(cat_table)
chisq$expected
```

```
##      Gender
## Group      F      M
##   Con 40.47291 63.52709
##   Exp 38.52709 60.47291
```

Since expected values are all greater than 5, which is the Rule of Thumb for chi-squared test, then the assumptions for the chi-square test are met. No need to do Fisher test.

```
#displayed the percentages for each category against the other categories
prop.table(cat_table,1)
```

```
##      Gender
## Group      F      M
##   Con 0.3942308 0.6057692
##   Exp 0.3838384 0.6161616
```

```
#displayed the result of the chi-test
chisq
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  cat_table
## X-squared = 6.0884e-05, df = 1, p-value = 0.9938
```

As we can see, the p-value is a very large number, this means that we do not reject the null hypothesis hinting that there is no association between the two categorical predictors Group and Gender. Therefore, both Group and Gender can be used in the same model together. For the succeeding parts, we will now also be conducting the chi-squared test using the predictors against our response variable, Fever to have more information on the association of the categorical predictors against the response variable Fever.

First up, is the chi-squared test on Group

Ho: there is no significant association between the two categorical variables Group and Fever

Ha: there is a significant association between the two categorical variables Group and Fever

```
#Created crosstabulations on Group
group_table <- xtabs(~ Fever + Group, data = pep1)
group_table
```

```
##      Group
## Fever Con Exp
##    1  63  77
##    2  41  22
```

From the table above, it seems that there are more people who got fever from the con group.

```
#performed chi-test and got expected values
chisq_group<-chisq.test(group_table)
chisq_group$expected
```

```
##      Group
## Fever      Con      Exp
##      1 71.72414 68.27586
##      2 32.27586 30.72414
```

Since expected values are all greater than 5, which is the Rule of Thumb for chi-squared test, then the assumptions for the chi-square test are met. No need to do Fisher test.

```
#displayed the percentages for each category against the other categories
prop.table(group_table,1)
```

```
##      Group
## Fever      Con      Exp
##      1 0.4500000 0.5500000
##      2 0.6507937 0.3492063
```

Based on the table above, it can clearly be seen that there really are more people who got fever coming from the Con group.

```
#displayed the result of the chi-test
chisq_group
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  group_table
## X-squared = 6.2306, df = 1, p-value = 0.01256
```

From the result, we can see that the p-value is less than 0.05 which means that we can reject the null hypothesis and that we cannot ignore group because it may have an association with our response variable, Fever. Hence, it should be part of our fitted model.

Next up, we are going to test Gender on Fever as well.

Ho: there is no significant association between the two categorical variables Gender and Fever

Ha: there is a significant association between the two categorical variables Gender and Fever

```
#Created crosstabulations on Gender
gender_table <- xtabs(~ Fever + Gender, data = pep1)
gender_table
```

```
##      Gender
## Fever  F  M
##      1 56 84
##      2 23 40
```

From the table above, it seems that there are more Males in general for both Fever categories.

```
#performed chi-test and got expected values
chisq_gender<-chisq.test(gender_table)
chisq_gender$expected
```

```
##      Gender
## Fever      F      M
##      1 54.48276 85.51724
##      2 24.51724 38.48276
```

Since expected values are all greater than 5, which is the Rule of Thumb for chi-squared test, then the assumptions for the chi-square test are met. No need to do Fisher test.

```
#displayed the percentages for each category against the other categories
prop.table(gender_table,1)
```

```
##      Gender
## Fever      F      M
##      1 0.4000000 0.6000000
##      2 0.3650794 0.6349206
```

Based on the table above, it can clearly be seen that there really are more Males with about 60% of the dataset.

```
#displayed the result of the chi-test
chisq_gender
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  gender_table
## X-squared = 0.10019, df = 1, p-value = 0.7516
```

From the result, we can see that the p-value is a very high number 0.7516 which means that we do not reject the null hypothesis hinting that there is no association with our response variable, Fever; and that the model can be fitted without Gender.

```
#Produced a table containing p-values for the two categorical predictors.
p_value_table <- data.frame(Variable = c("Group", "Gender"),P_Value =c(0.01256, 0.7516))
p_value_table
```

```
##  Variable P_Value
## 1    Group 0.01256
## 2   Gender 0.75160
```

First off, we will be creating a series of boxplots for the pairs of numeric predictors against Group.


```

#created boxplots for the pairs of numeric and Group categorical predictor
b1 <- ggplot(data=pep1,aes(x=Group,y=Age)) + geom_boxplot() +theme_bw() + labs(caption="Con, Exp",x="Con, Exp")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

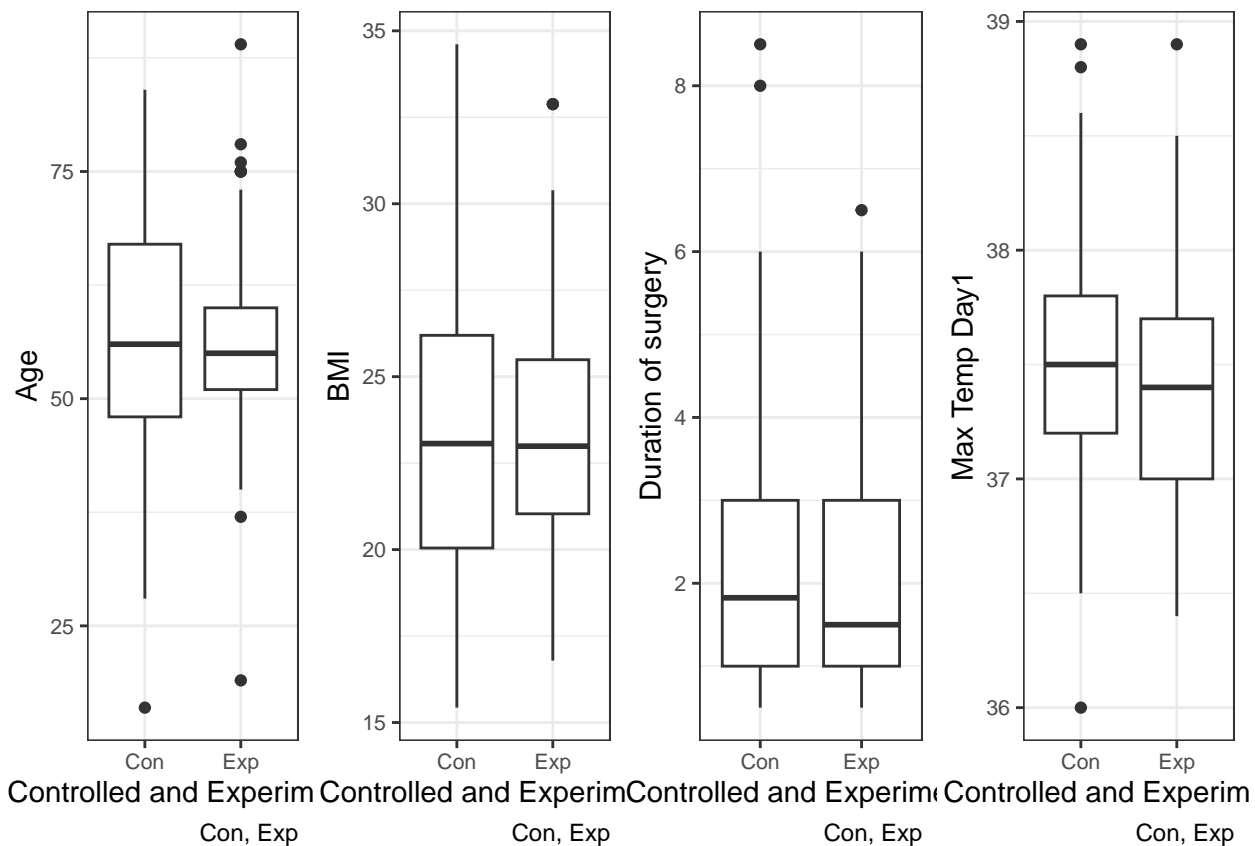
b2 <- ggplot(data=pep1,aes(x=Group,y=BMI)) + geom_boxplot() +theme_bw() + labs(caption="Con, Exp",x="Con, Exp")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

b3 <- ggplot(data=pep1,aes(x=Group,y=Durationofsurgery)) + geom_boxplot() +theme_bw() + labs(caption="Con, Exp",x="Con, Exp")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

b4 <- ggplot(data=pep1,aes(x=Group,y=MaxtempDay1)) + geom_boxplot() +theme_bw() + labs(caption="Con, Exp",x="Con, Exp")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

grid.arrange(b1,b2,b3,b4, ncol = 4)

```



From the boxplots of Group against Age, Duration of Surgery, and Max Temp Day1, it seems that the median age of subjects and median values of duration of surgery and Max Temp Day1 are slightly higher for the Con Group. This hints at a possibility of a relationship between Group and these predictors. As of now, it seems beneficial to include Age, Duration of Surgery and Max Temp Day1 for Group.

On the other hand, from the boxplot of Group and BMI, it seems that the median BMI level for both groups are the same. This hints at a possibility that there is no relationship between Group and BMI. As of now, it seems not beneficial to include BMI for Group.

Next, we will be creating a series of boxplots for the pairs of numeric predictors against Gender.

```

#created boxplots for the pairs of numeric and Gender categorical predictor
b5 <- ggplot(data=pep1,aes(x=Gender,y=Age)) + geom_boxplot() +theme_bw() + labs(caption="F, M",x="Female and Male")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

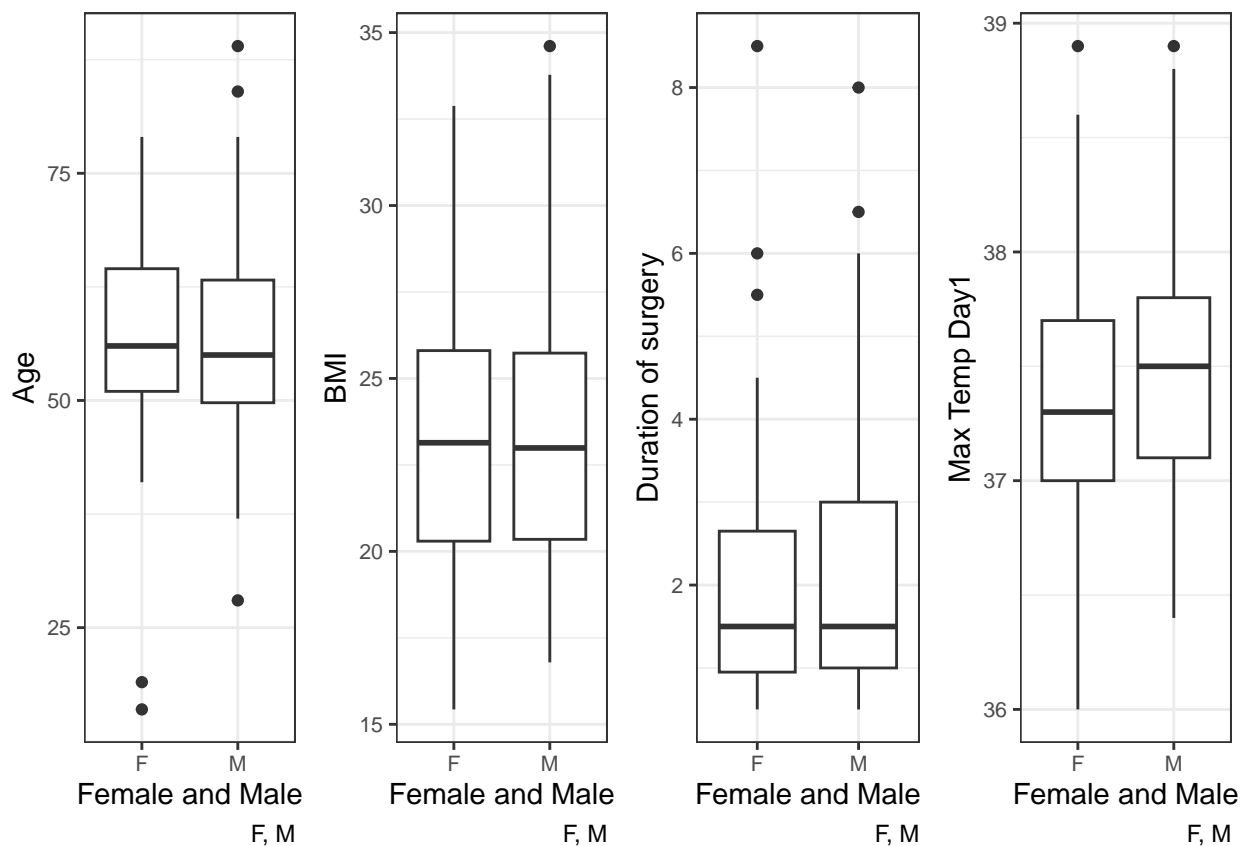
b6 <- ggplot(data=pep1,aes(x=Gender,y=BMI)) + geom_boxplot() +theme_bw() + labs(caption="F, M",x="Female and Male")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

b7 <- ggplot(data=pep1,aes(x=Gender,y=Durationofsurgery)) + geom_boxplot() +theme_bw() + labs(caption="F, M",x="Female and Male")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

b8 <- ggplot(data=pep1,aes(x=Gender,y=MaxtempDay1)) + geom_boxplot() +theme_bw() + labs(caption="F, M",x="Female and Male")
theme(axis.text = element_text(size = 8)) + theme(axis.text.x = element_text(size = 8))

grid.arrange(b5,b6,b7,b8, ncol = 4)

```



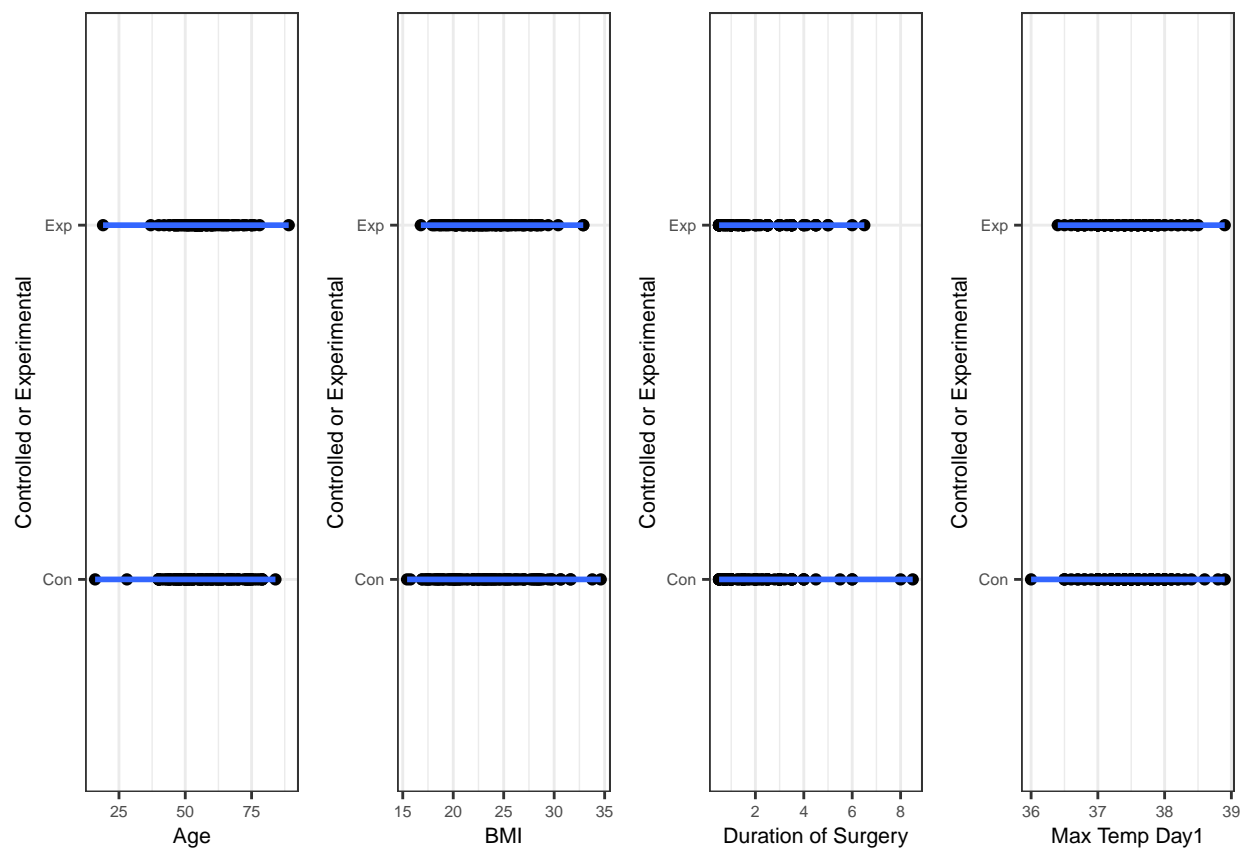
From the boxplot of Gender against Age, it seems that the median age of subjects is slightly higher for the Female Gender. This hints at a possibility that there is a relationship between Gender and Age .As of now, it seems beneficial to include Age for Gender.

From the boxplot of Gender against Max Temp Day1, it is evident that the Max Temp Day1 values are higher for the Male Gender. This hints at a possibility that there is a relationship between Gender and Max Temp Day1.As of now, it seems beneficial to include Max Temp Day1 for Gender.

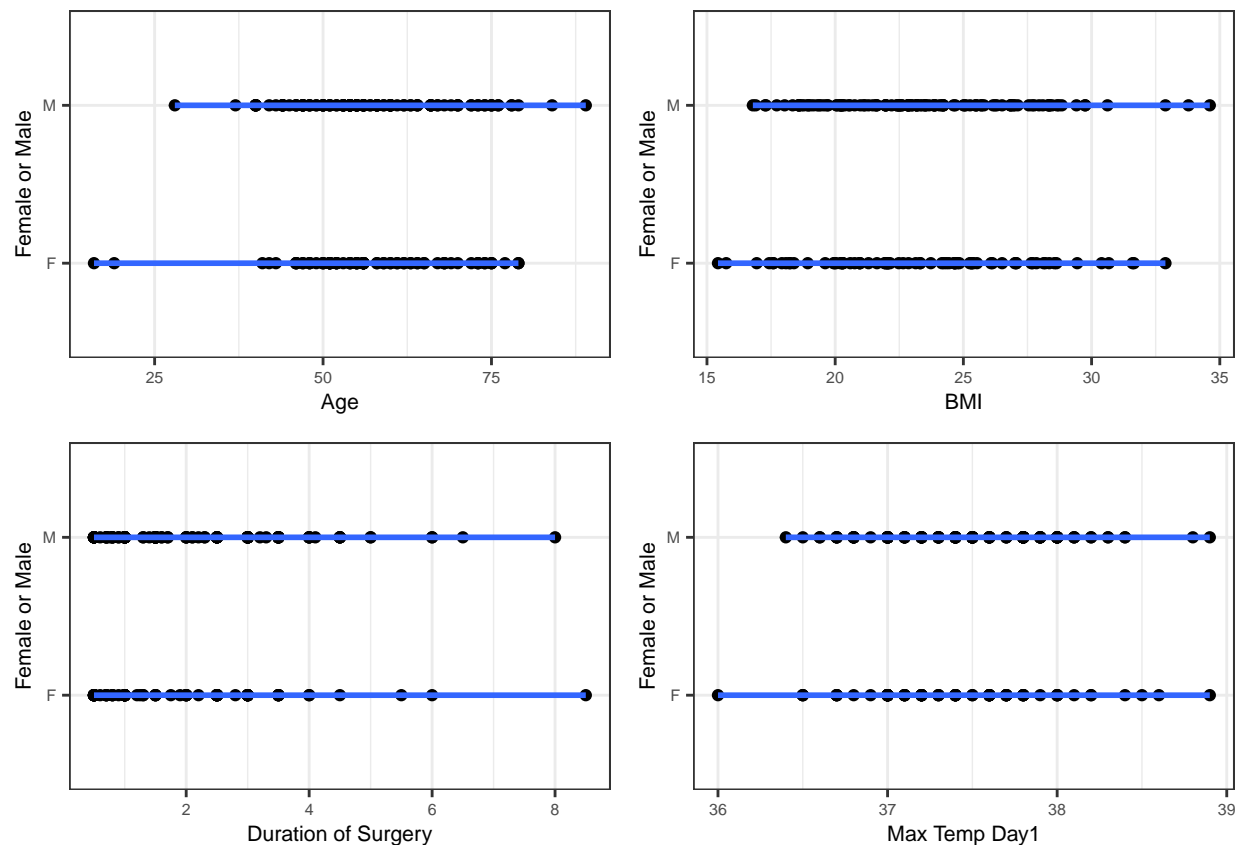
On the other hand, from the boxplots of Gender against BMI and Duration of surgery, it seems that the median BMI level and median Duration of surgery for both genders are the same. This hints at a possibility that there is no relationship between Gender and the two numerical predictors, BMI and Duration of surgery. As of now, it seems not beneficial to include BMI and Duration of Surgery for Gender.

For the next part, we will also be creating scatter plots.

```
#created scatterplots with splines for the pairs of numeric and Group categorical predictor
sc1 <- ggplot(data=pep1,aes(y=Group,x=Age)) + geom_point()+geom_smooth(method = lm, formula = y ~ spline
sc2 <- ggplot(data=pep1,aes(y=Group,x=BMI)) + geom_point()+geom_smooth(method = lm, formula = y ~ spline
sc3 <- ggplot(data=pep1,aes(y=Group,x=Durationofsurgery)) + geom_point()+geom_smooth(method = lm, formul
sc4 <- ggplot(data=pep1,aes(y=Group,x=MaxtempDay1)) + geom_point()+geom_smooth(method = lm, formula = y
grid.arrange(sc1,sc2,sc3,sc4, ncol = 4)
```



```
#created scatterplots with splines for the pairs of numeric and Gender categorical predictor
sc5 <- ggplot(data=pep1,aes(y=Gender,x=Age)) + geom_point()+geom_smooth(method = lm, formula = y ~ spline
sc6 <- ggplot(data=pep1,aes(y=Gender,x=BMI)) + geom_point()+geom_smooth(method = lm, formula = y ~ spline
sc7 <- ggplot(data=pep1,aes(y=Gender,x=Durationofsurgery)) + geom_point()+geom_smooth(method = lm, formul
sc8 <- ggplot(data=pep1,aes(y=Gender,x=MaxtempDay1)) + geom_point()+geom_smooth(method = lm, formula = y
grid.arrange(sc5,sc6,sc7,sc8)
```



Looking at the scatter plots above, we cannot clearly see the relationships between the numerical and categorical pairs. To further help us make sense of these and make sense of our investigation, we will be using a series of overlapping histograms.

```
pep1$Group<-as.factor(pep1$Group)
pep1$Group=as.factor(pep1$Group)
```

```
## Warning in Ops.factor(pep1$Group, as.factor(pep1$Group)): '-' not meaningful
## for factors
```

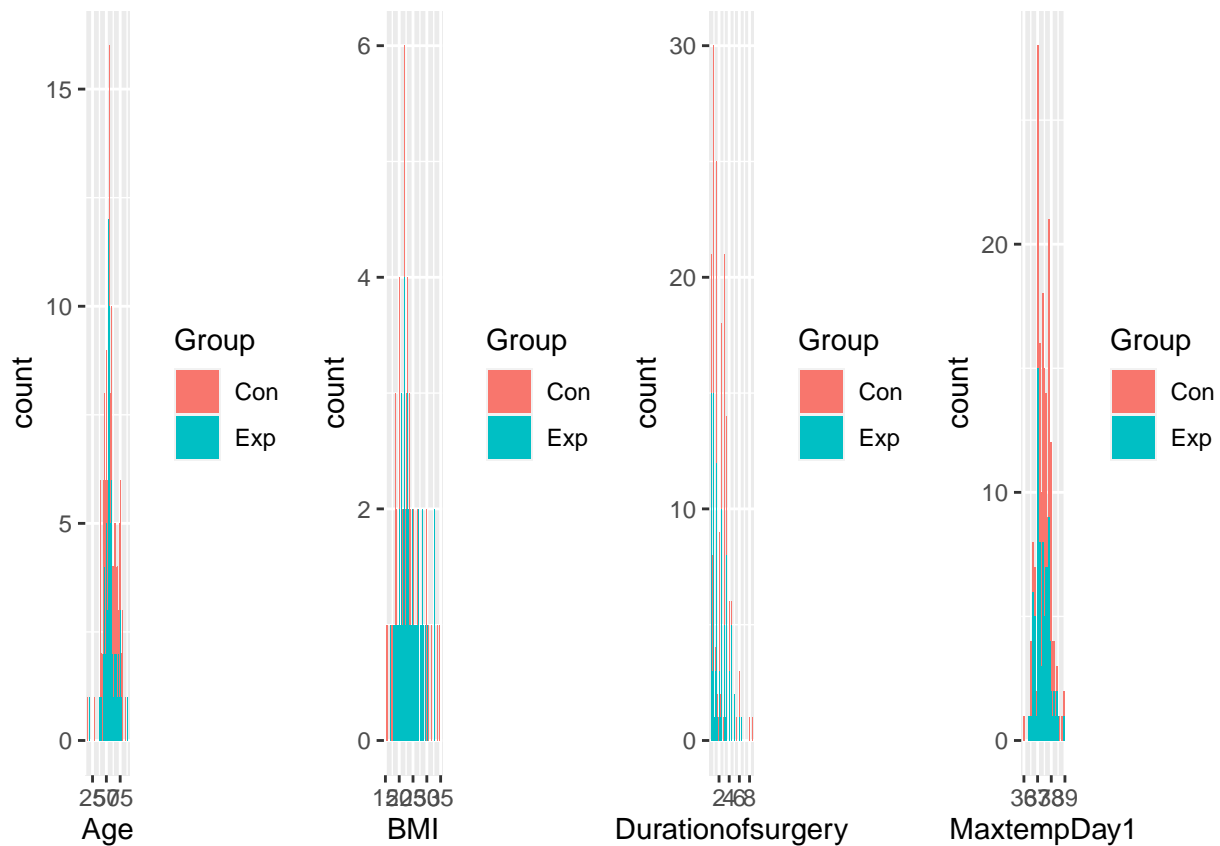
```
## [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [51] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [76] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [101] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [126] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [151] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [176] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [201] NA NA NA
```

```
h1<-ggplot(pep1, aes(x = Age, fill = Group)) +
geom_bar()
```

```
h2<-ggplot(pep1, aes(x = BMI, fill = Group)) +
geom_bar()
```

```
h3<-ggplot(peg1, aes(x = Durationofsurgery, fill = Group)) +
geom_bar()

h4<-ggplot(peg1, aes(x = MaxtempDay1, fill = Group)) +
geom_bar()
grid.arrange(h1,h2,h3,h4, ncol = 4)
```



```
peg1$Gender<-as.factor(peg1$Gender)
peg1$Gender=as.factor(peg1$Gender)
```

```
## Warning in Ops.factor(peg1$Gender, as.factor(peg1$Gender)): '-' not meaningful
## for factors
```

```
## [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [51] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [76] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [101] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [126] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [151] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [176] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [201] NA NA NA
```

```

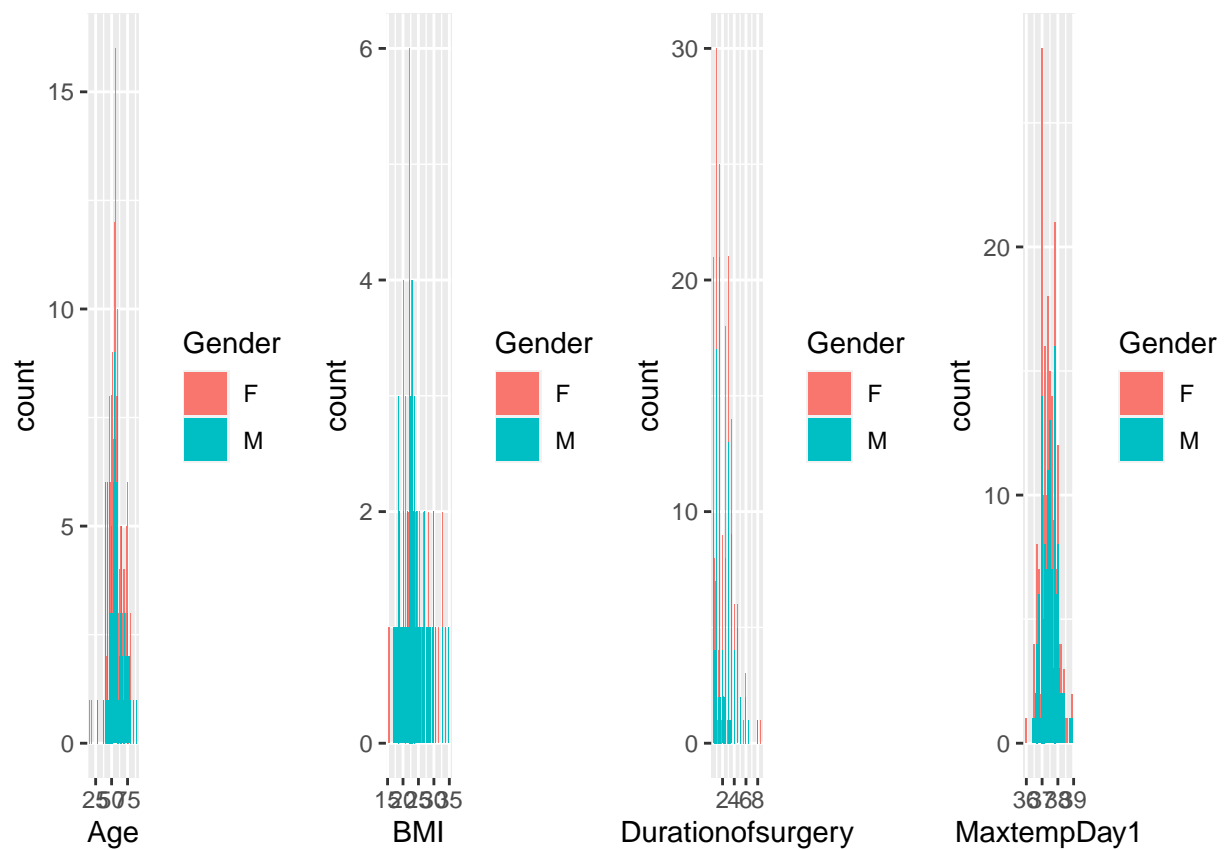
h5<-ggplot(pep1, aes(x = Age, fill = Gender)) +
geom_bar()

h6<-ggplot(pep1, aes(x = BMI, fill = Gender)) +
geom_bar()

h7<-ggplot(pep1, aes(x = Durationofsurgery, fill = Gender)) +
geom_bar()

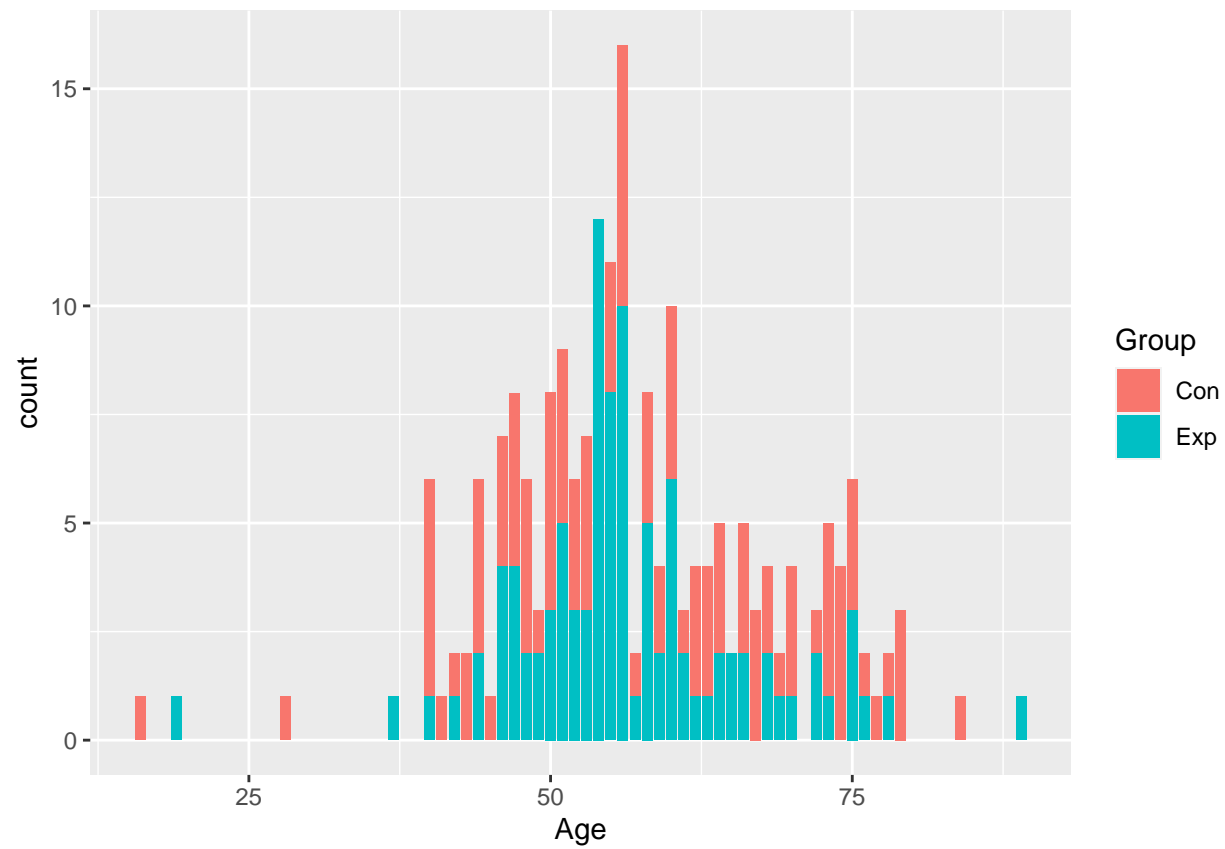
h8<-ggplot(pep1, aes(x = MaxtempDay1, fill = Gender)) +
geom_bar()
grid.arrange(h5,h6,h7,h8, ncol = 4)

```

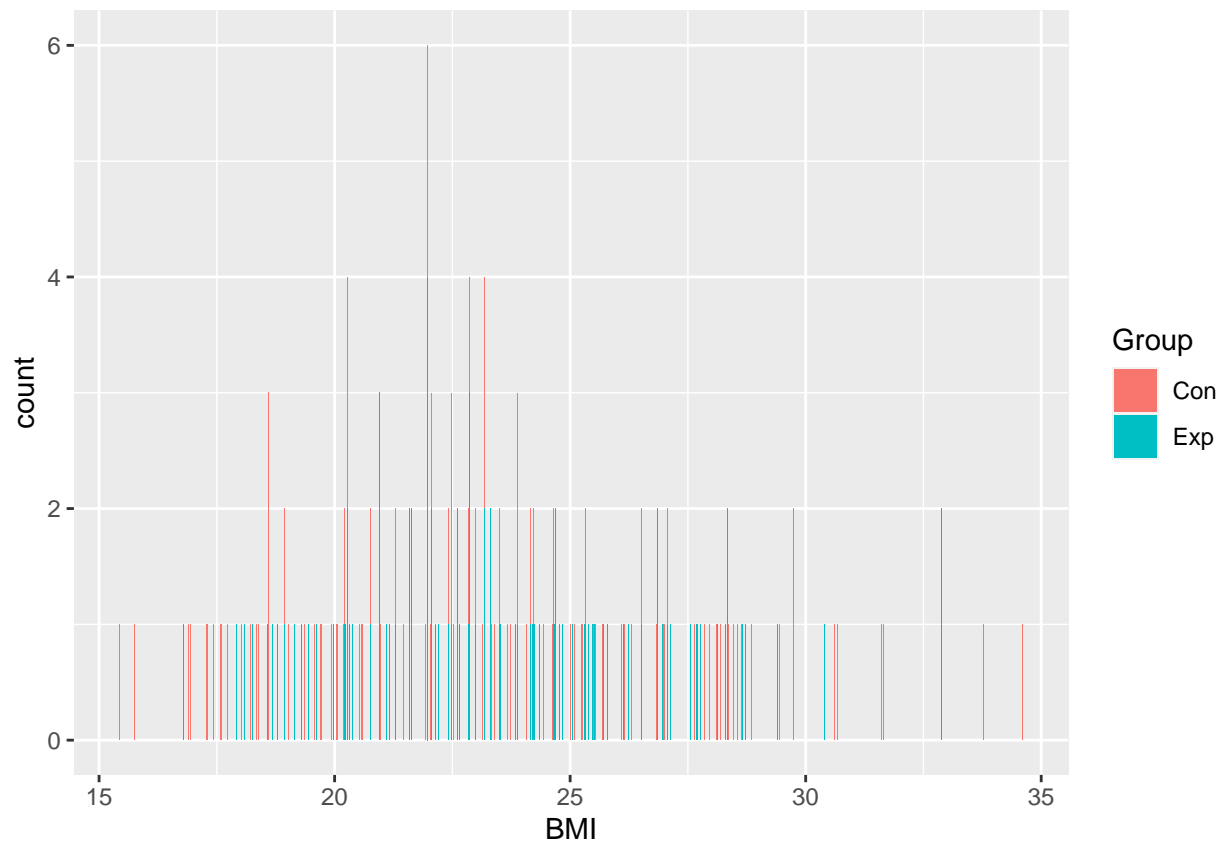


To see each graph clearly we will zoom in.

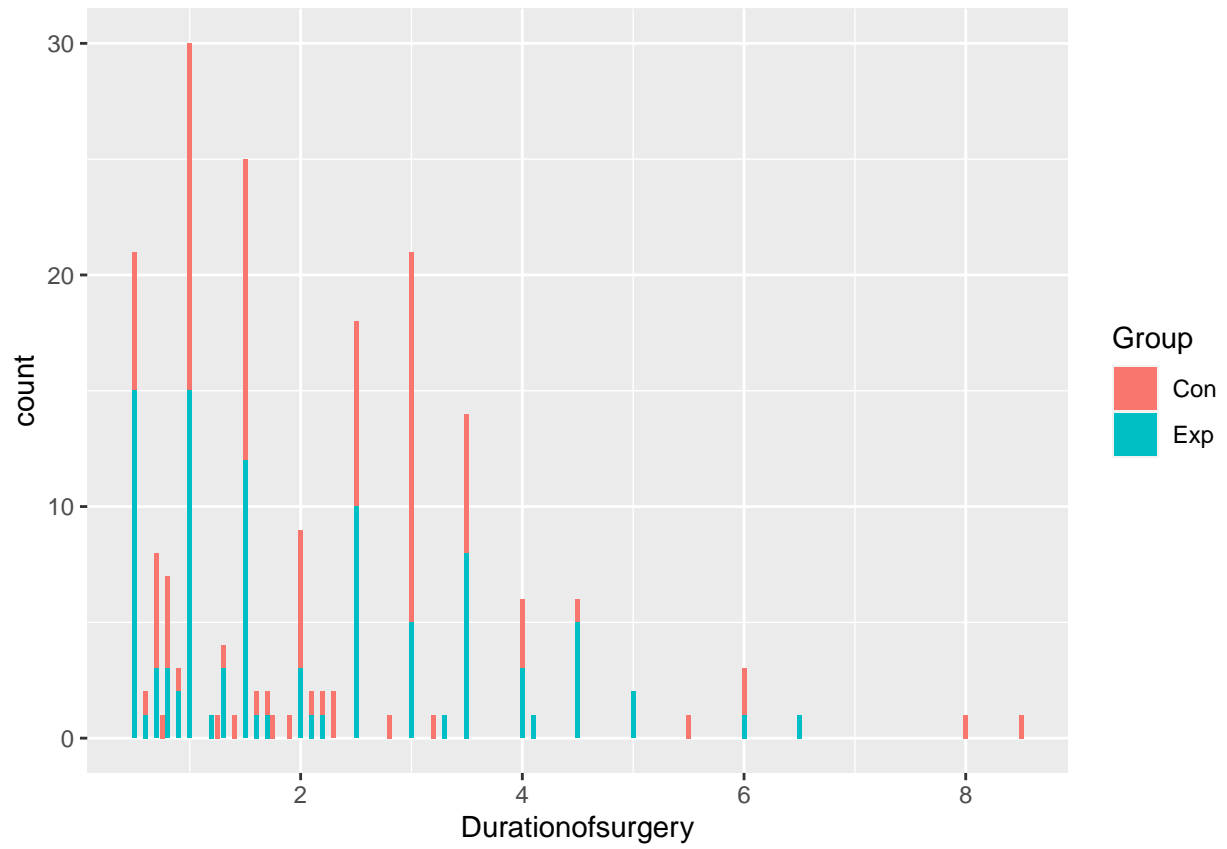
```
h1
```



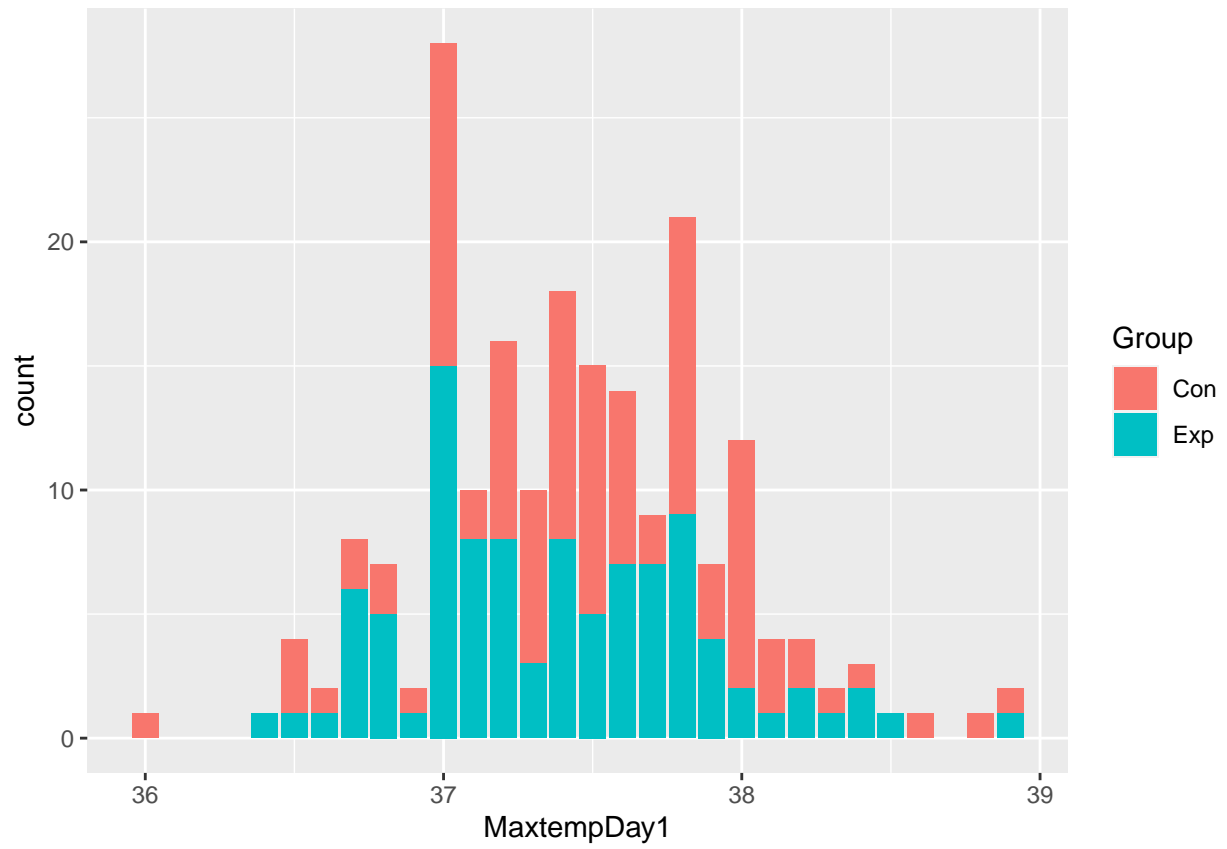
h2



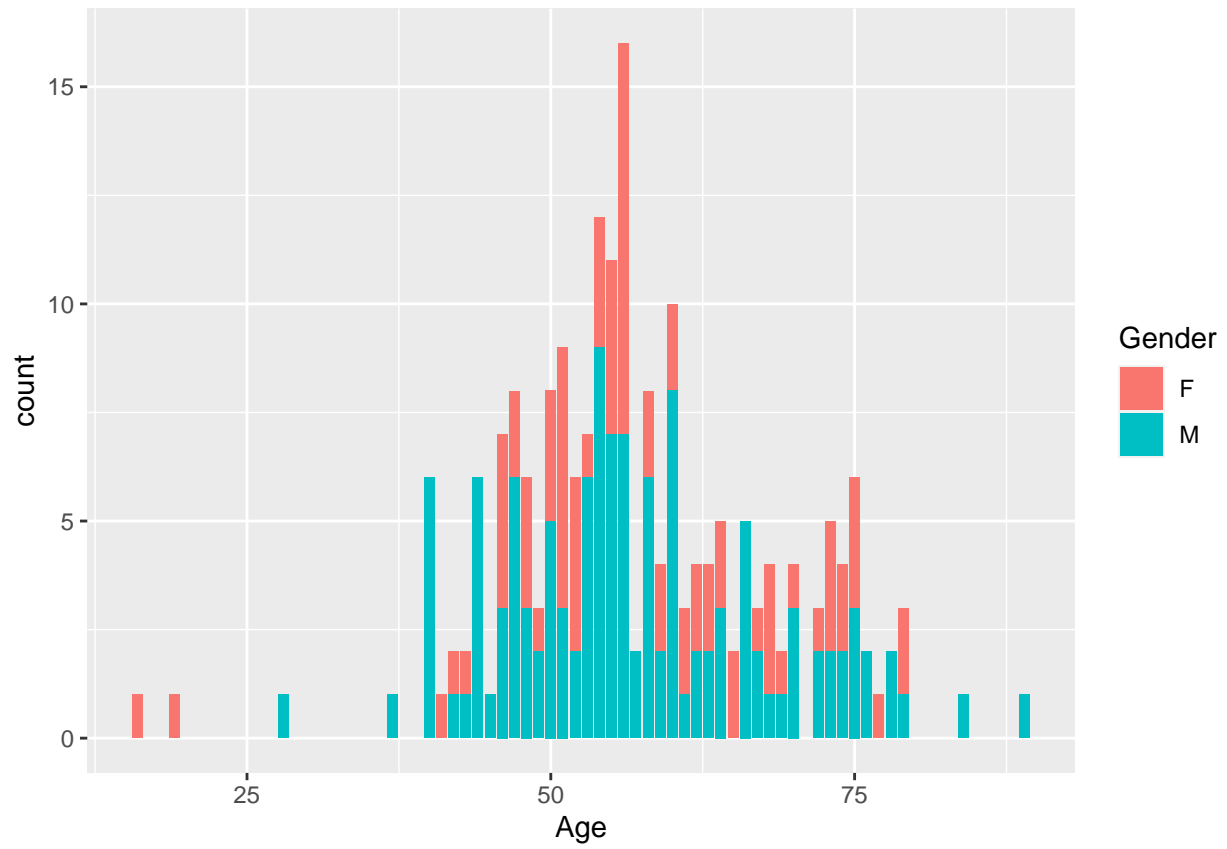
h3



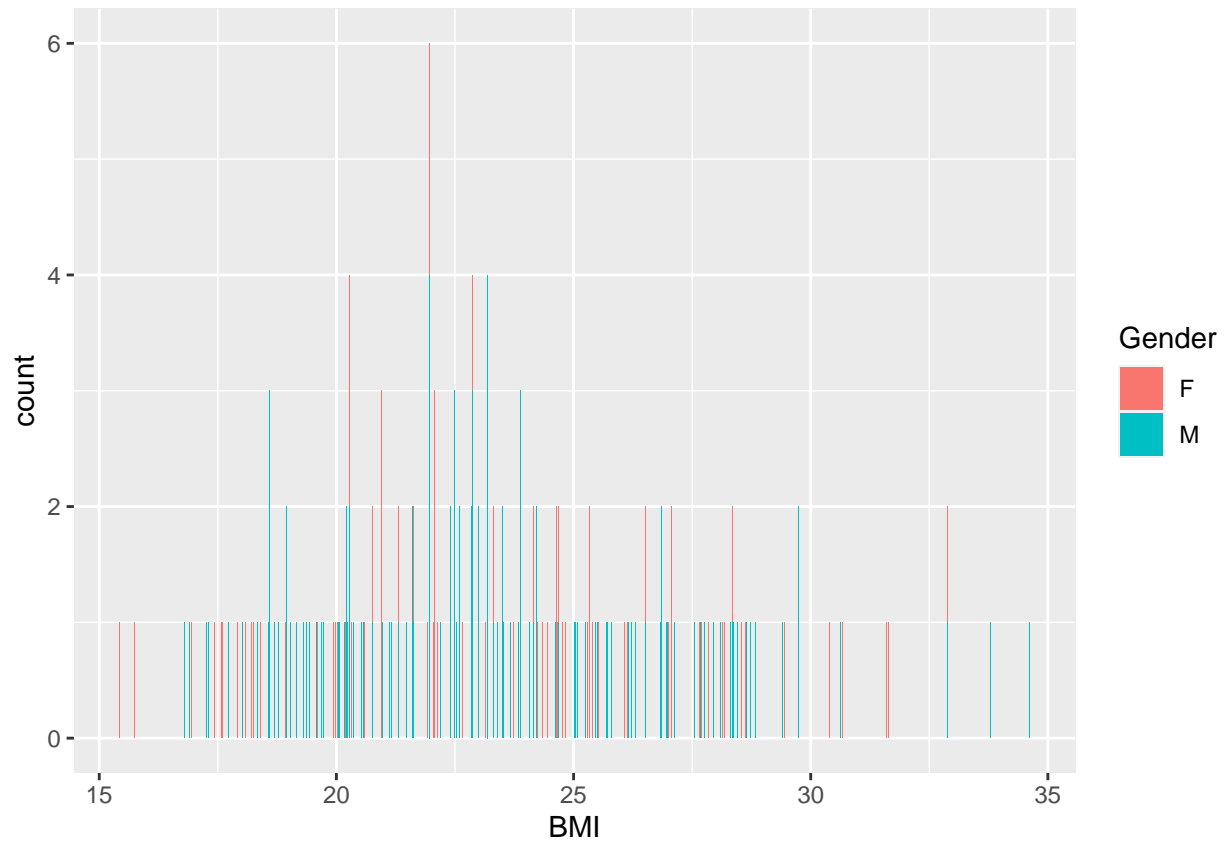
h4



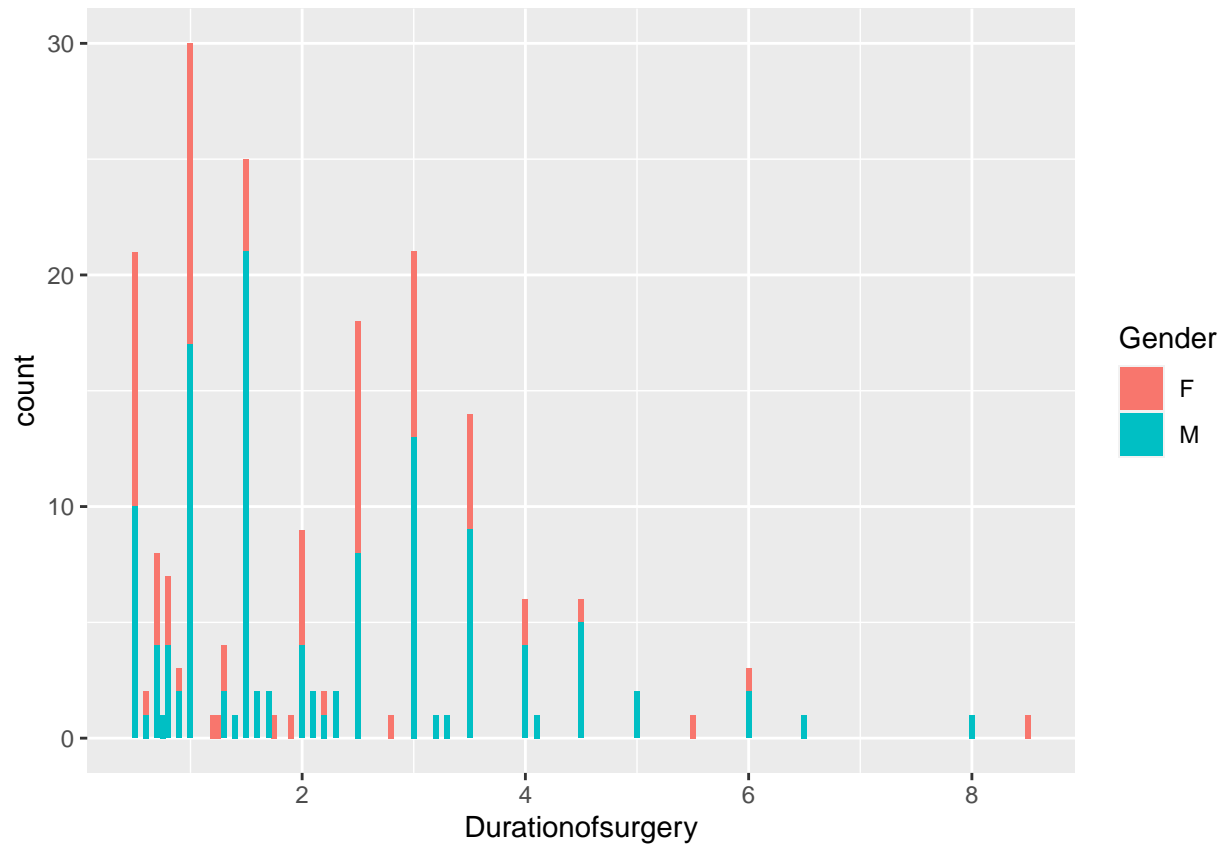
h5



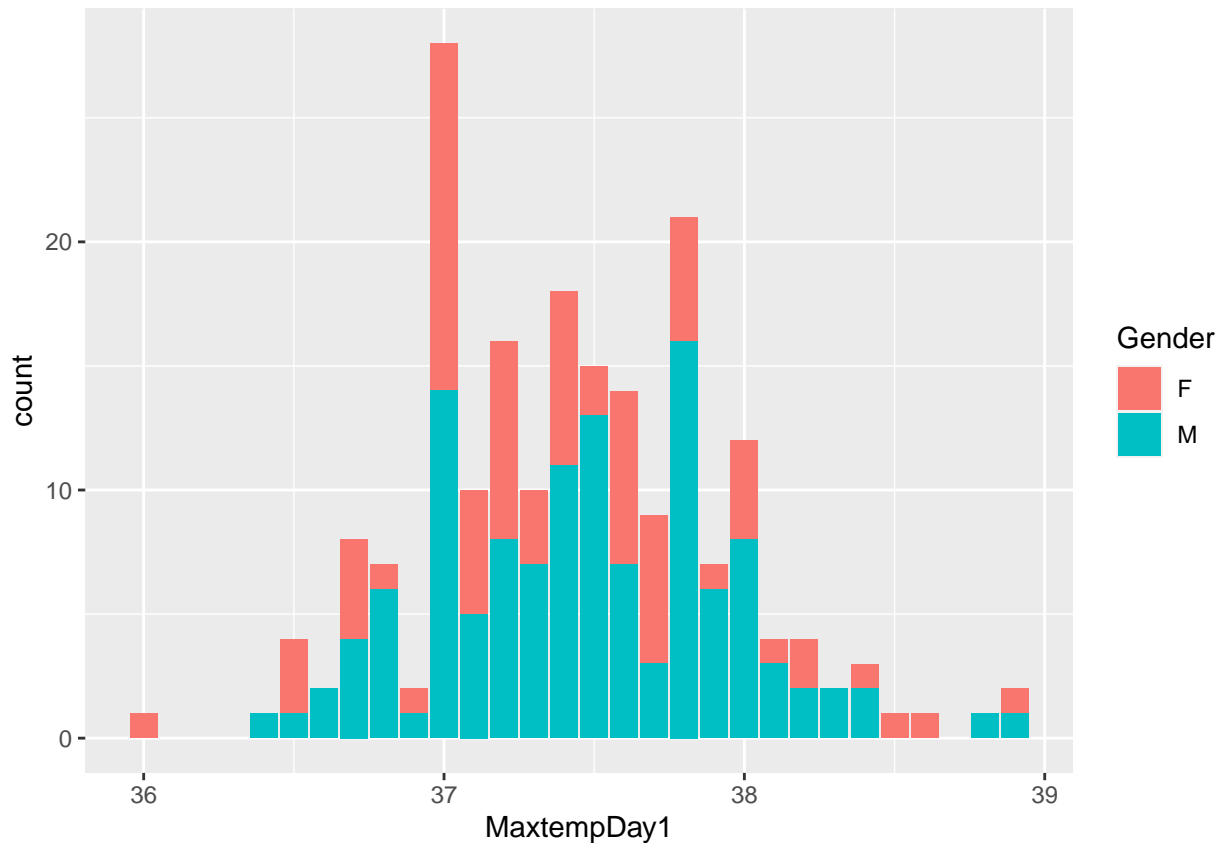
h6



h7



h8



After close inspection of each of the histograms above, it seems that there are no numerical predictors that have significant relationship with either Group or Gender since the values are relatively spread out and proportional. There was no clear concentration of values present.

Now, to formally test our observations from the graphs of the numerical predictors and the categorical predictors, we will be performing a series of median tests. For each test to be conducted, the hypothesis test is as follows.

Ho: The medians of the the numerical and categorical variable pair are equal Ha: The medians of the the numerical and categorical variable pair are different

```
#loaded the coin library
```

```
library(survival)
```

```
## Warning: package 'survival' was built under R version 4.2.3
```

```
library(coin)
```

```
## Warning: package 'coin' was built under R version 4.2.3
```

```
#Performed the series of median tests for Group and the other numerical variables
```

```
pep1$Group <- as.factor(pep1$Group)
```

```
median_test(Age~Group, data = pep1)
```

```
##
```

```
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: Age by Group (Con, Exp)
## Z = 1.6722, p-value = 0.09448
## alternative hypothesis: true mu is not equal to 0
```

```
median_test(BMI~Group, data = pep1)
```

```
##
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: BMI by Group (Con, Exp)
## Z = 0.07176, p-value = 0.9428
## alternative hypothesis: true mu is not equal to 0
```

```
median_test(Durationofsurgery~Group, data = pep1)
```

```
##
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: Durationofsurgery by Group (Con, Exp)
## Z = 1.1996, p-value = 0.2303
## alternative hypothesis: true mu is not equal to 0
```

```
median_test(MaxtempDay1~Group, data = pep1)
```

```
##
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: MaxtempDay1 by Group (Con, Exp)
## Z = 1.3516, p-value = 0.1765
## alternative hypothesis: true mu is not equal to 0
```

```
#Performed the series of median tests for Gender and the other numerical variables
```

```
pep1$Gender <- as.factor(pep1$Gender)
median_test(Age~Gender, data = pep1)
```

```
##
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: Age by Gender (F, M)
## Z = -0.071373, p-value = 0.9431
## alternative hypothesis: true mu is not equal to 0
```

```
median_test(BMI~Gender, data = pep1)
```

```
##
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: BMI by Gender (F, M)
## Z = 0.19948, p-value = 0.8419
## alternative hypothesis: true mu is not equal to 0
```

```
median_test(Durationofsurgery~Gender, data = pep1)
```

```
##
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: Durationofsurgery by Gender (F, M)
## Z = -0.15143, p-value = 0.8796
## alternative hypothesis: true mu is not equal to 0
```

```
median_test(MaxtempDay1~Gender, data = pep1)
```

```
##
## Asymptotic Two-Sample Brown-Mood Median Test
##
## data: MaxtempDay1 by Gender (F, M)
## Z = -1.5415, p-value = 0.1232
## alternative hypothesis: true mu is not equal to 0
```

```
#Computed for the Bonferonni correction
```

$$0.05/8 = 0.00625$$

The new value is 0.00625 after the correction.

```
#Created a table with the p-value results from the series of median tests performed above
median_table <- data.frame(Variable = c("Age", "BMI", "Duration of Surgery", "Max Temp Day1"),Group = c
median_table
```

```
##           Variable  Group Gender
## 1              Age 0.09448 0.9431
## 2              BMI 0.94280 0.8419
## 3 Duration of Surgery 0.23030 0.8796
## 4      Max Temp Day1 0.17650 0.1232
```

Based on the results of the median test, the p-values of all the tests performed are greater than the bonferonni corrected value 0.00625. This means that we do not reject the null hypothesis which implies that we do not have sufficient evidence to say that there is a statistically significant difference in the medians of the numerical and categorical predictor pairs above. It is also important to note that even if we don't adjust the significance value using the Bonferonni approach, we will still not be rejecting the null hypothesis for all pairs, since all p-values will still be above the 0.05 mark. Therefore, we can conclude that none of the pairs have different medians hinting that the numerical predictors do not significantly change as categories or levels of the categorical predictors (Con,Exp,M,F) change which could mean that the categorical predictors and the numerical predictors do not have any significant relationship.

```
#fitted models first using single predictors
modell1<-glm(Fever ~ Group, data=pep1,family="binomial")
summary(modell1)
```

```
##
## Call:
## glm(formula = Fever ~ Group, family = "binomial", data = pep1)
```



```
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.001  -1.001  -0.709   1.364   1.734
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.4296     0.2007  -2.141  0.03229 *
## GroupExp     -0.8232     0.3142  -2.620  0.00879 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 244.37  on 201  degrees of freedom
## AIC: 248.37
##
## Number of Fisher Scoring iterations: 4
```

```
model2<-glm(Fever ~ Gender, data=pep1,family="binomial")
summary(model2)
```

```
##
## Call:
## glm(formula = Fever ~ Gender, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8826  -0.8826  -0.8296   1.5043   1.5710
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.8899     0.2477  -3.593 0.000327 ***
## GenderM       0.1479     0.3134   0.472 0.636973
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 251.24  on 201  degrees of freedom
## AIC: 255.24
##
## Number of Fisher Scoring iterations: 4
```

```
model3<-glm(Fever ~ Age, data=pep1,family="binomial")
summary(model3)
```

```
##
## Call:
## glm(formula = Fever ~ Age, family = "binomial", data = pep1)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0230  -0.8732  -0.8331   1.4779   1.6424
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.16130    0.79095  -0.204   0.838
## Age         -0.01123    0.01374  -0.817   0.414
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 250.80  on 201  degrees of freedom
## AIC: 254.8
##
## Number of Fisher Scoring iterations: 4
```

```
model4<-glm(Fever ~ BMI, data=pep1,family="binomial")
summary(model4)
```

```
##
## Call:
## glm(formula = Fever ~ BMI, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8796  -0.8652  -0.8572   1.5181   1.5596
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.650251    0.951518  -0.683   0.494
## BMI         -0.006346    0.040240  -0.158   0.875
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 251.44  on 201  degrees of freedom
## AIC: 255.44
##
## Number of Fisher Scoring iterations: 4
```

```
model5<-glm(Fever ~ Durationofsurgery, data=pep1,family="binomial")
summary(model5)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery, family = "binomial",
##      data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7417  -0.8388  -0.7196   1.2429   1.7988
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.5745     0.2904  -5.422 5.88e-08 ***
## Durationofsurgery  0.3554     0.1089   3.265  0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 239.95  on 201  degrees of freedom
## AIC: 243.95
##
## Number of Fisher Scoring iterations: 4
```

```
model6<-glm(Fever ~ MaxtempDay1, data=pep1,family="binomial")
summary(model6)
```

```
##
## Call:
## glm(formula = Fever ~ MaxtempDay1, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3183  -0.8712  -0.7367   1.3429   2.0412
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -30.1982     11.8991  -2.538  0.0112 *
## MaxtempDay1  0.7847      0.3173   2.473  0.0134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 245.11  on 201  degrees of freedom
## AIC: 249.11
##
## Number of Fisher Scoring iterations: 4
```

#Created a table with the AIC values of the models fitted with the single predictors

```
AIC_table1 <- data.frame(Model_Number = c(1, 2, 3, 4, 5, 6), Predictors = c("Group", "Gender", "Age", "BMI", "Duration of Surgery", "Max Temp Day1"))
AIC_table1
```

```
##   Model_Number      Predictors    AIC
## 1           1             Group 248.37
## 2           2             Gender 255.24
## 3           3              Age 254.80
## 4           4              BMI 255.44
## 5           5 Duration of Surgery 243.95
## 6           6      Max Temp Day1 249.11
```

From the table, we can see that the best model is model number 5(Duration of Surgery) with the lowest AIC of 243.95. As a consequence, we will be pairing Duration of Surgery with other predictors.

#fitted models using Duration of surgery with the other predictors

```
model7<-glm(Fever ~ Durationofsurgery+Group, data=pep1,family="binomial")
summary(model7)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + Group, family = "binomial",
##      data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9225  -0.8414  -0.6336   1.2342   2.0047
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.2182     0.3206  -3.799 0.000145 ***
## Durationofsurgery  0.3619     0.1130   3.201 0.001369 **
## GroupExp        -0.8283     0.3241  -2.556 0.010596 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 233.19  on 200  degrees of freedom
## AIC: 239.19
##
## Number of Fisher Scoring iterations: 4
```

```
model8<-glm(Fever ~ Durationofsurgery+Gender, data=pep1,family="binomial")
summary(model8)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + Gender, family = "binomial",
##      data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7496  -0.8226  -0.7070   1.2320   1.8181
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.61686     0.34412  -4.699 2.62e-06 ***
## Durationofsurgery  0.35354     0.10907   3.241 0.00119 **
## GenderM          0.07513     0.32367   0.232 0.81644
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 251.47 on 202 degrees of freedom
## Residual deviance: 239.90 on 200 degrees of freedom
## AIC: 245.9
##
## Number of Fisher Scoring iterations: 4

model9<-glm(Fever ~ Durationofsurgery+Age, data=pep1,family="binomial")
summary(model9)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + Age, family = "binomial",
## data = pep1)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.8129 -0.8426 -0.7100 1.2239 1.9180
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.54232 0.82218 -0.660 0.509503
## Durationofsurgery 0.37714 0.11081 3.403 0.000666 ***
## Age -0.01902 0.01438 -1.323 0.185959
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 251.47 on 202 degrees of freedom
## Residual deviance: 238.19 on 200 degrees of freedom
## AIC: 244.19
##
## Number of Fisher Scoring iterations: 4
```

```
model10<-glm(Fever ~ Durationofsurgery+BMI, data=pep1,family="binomial")
summary(model10)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + BMI, family = "binomial",
## data = pep1)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.7323 -0.8435 -0.7137 1.2552 1.8379
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.82408 1.05284 -1.733 0.08318 .
## Durationofsurgery 0.35847 0.10954 3.273 0.00107 **
## BMI 0.01038 0.04199 0.247 0.80474
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 239.89  on 200  degrees of freedom
## AIC: 245.89
##
## Number of Fisher Scoring iterations: 4
```

```
model11<-glm(Fever ~ Durationofsurgery+MaxtempDay1,data=pep1,family="binomial")
summary(model11)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1, family = "binomial",
##      data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6517  -0.8223  -0.6464   1.0991   2.1423
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -36.9997    12.3490  -2.996 0.002734 **
## Durationofsurgery  0.3980     0.1107   3.597 0.000322 ***
## MaxtempDay1      0.9428     0.3280   2.875 0.004044 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 231.30  on 200  degrees of freedom
## AIC: 237.3
##
## Number of Fisher Scoring iterations: 4
```

```
#Updated the AIC table with the AIC values of the new models fitted using duration of surgery in combin
AIC_table2 <- data.frame(Model_Number = c(1, 2, 3, 4, 5, 6,7,8,9,10,11), Predictors = c("Group","Gender",
AIC_table2
```

##	Model_Number	Predictors	AIC
## 1	1	Group	248.37
## 2	2	Gender	255.24
## 3	3	Age	254.80
## 4	4	BMI	255.44
## 5	5	Duration of Surgery	243.95
## 6	6	Max Temp Day1	249.11
## 7	7	Duration of Surgery+Group	239.19
## 8	8	Duration of Surgery+Gender	245.90
## 9	9	Duration of Surgery+Age	244.19
## 10	10	Duration of Surgery+BMI	245.89
## 11	11	Duration of Surgery+Max Temp Day1	237.30

From the updated table above, we can see that the best model is model number 11(Duration of Surgery+Max Temp Day1) with the lowest AIC of 237.30. As a consequence, we will be pairing Duration of Surgery and Max Temp Day 1 with the other predictors.

```
#fitted models using Duration of surgery and Max Temp Day1 with the other predictors
model12<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Group, data=pep1,family="binomial")
summary(model12)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group,
##      family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8274  -0.8334  -0.6181   1.0940   2.1992
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -34.6476    12.5407  -2.763  0.005730 **
## Durationofsurgery  0.4026     0.1146   3.513  0.000443 ***
## MaxtempDay1      0.8889     0.3329   2.670  0.007584 **
## GroupExp        -0.7667     0.3307  -2.318  0.020435 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 225.78  on 199  degrees of freedom
## AIC: 233.78
##
## Number of Fisher Scoring iterations: 4
```

```
model13<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Gender, data=pep1,family="binomial")
summary(model13)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Gender,
##      family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6516  -0.8222  -0.6465   1.0990   2.1422
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.700e+01  1.238e+01  -2.990  0.002793 **
## Durationofsurgery  3.980e-01  1.110e-01   3.587  0.000335 ***
## MaxtempDay1      9.429e-01  3.291e-01   2.865  0.004168 **
## GenderM         -6.101e-04  3.330e-01  -0.002  0.998538
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 231.30  on 199  degrees of freedom
## AIC: 239.3
##
## Number of Fisher Scoring iterations: 4
```

```
model14<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Age, data=pep1,family="binomial")
summary(model14)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Age,
##      family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7146  -0.8161  -0.6513   1.0798   2.1939
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -35.50513    12.46278  -2.849 0.004387 **
## Durationofsurgery  0.41574     0.11235   3.700 0.000215 ***
## MaxtempDay1      0.92767     0.32988   2.812 0.004921 **
## Age            -0.01702     0.01472  -1.156 0.247490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 229.96  on 199  degrees of freedom
## AIC: 237.96
##
## Number of Fisher Scoring iterations: 4
```

```
model15<-glm(Fever ~ Durationofsurgery+MaxtempDay1+BMI, data=pep1,family="binomial")
summary(model15)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + BMI,
##      family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6415  -0.8293  -0.6434   1.0986   2.1598
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```



```
## (Intercept)      -37.35669    12.42320   -3.007  0.002638 **
## Durationofsurgery  0.40221     0.11165    3.602  0.000315 ***
## MaxtempDay1       0.94434     0.32818    2.877  0.004009 **
## BMI               0.01243     0.04323    0.287  0.773804
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 231.22  on 199  degrees of freedom
## AIC: 239.22
##
## Number of Fisher Scoring iterations: 4
```

#Updated the AIC table with the AIC values of the new models fitted using duration of surgery and Max T
AIC_table3 <- data.frame(Model_Number = c(1, 2, 3, 4, 5, 6,7,8,9,10,11,12,13,14,15), Predictors = c("Gr

AIC_table3

##	Model_Number	Predictors	AIC
## 1	1	Group	248.37
## 2	2	Gender	255.24
## 3	3	Age	254.80
## 4	4	BMI	255.44
## 5	5	Duration of Surgery	243.95
## 6	6	Max Temp Day1	249.11
## 7	7	Duration of Surgery+Group	239.19
## 8	8	Duration of Surgery+Gender	245.90
## 9	9	Duration of Surgery+Age	244.19
## 10	10	Duration of Surgery+BMI	245.89
## 11	11	Duration of Surgery+Max Temp Day1	237.30
## 12	12	Duration of Surgery+Max Temp Day1+Group	233.78
## 13	13	Duration of Surgery+Max Temp Day1+Gender	239.30
## 14	14	Duration of Surgery+Max Temp Day1+Age	237.96
## 15	15	Duration of Surgery+Max Temp Day1+BMI	239.22

From the updated table above, we can see that the best model is now model number 12(Duration of Surgery+Max Temp Day1+Group) with the lowest AIC of 233.78. As a consequence, we will be pairing Duration of Surgery, Max Temp Day 1 and Group with the other predictors.

#fitted models using Duration of surgery, Max Temp Day1 and Group with the other predictors
model16<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Group+Gender, data=pep1,family="binomial")
summary(model16)

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group +
##      Gender, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8267  -0.8326  -0.6190   1.0947   2.2003
```

```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -34.667141  12.595177  -2.752 0.005916 **
## Durationofsurgery  0.402732  0.115035   3.501 0.000464 ***
## MaxtempDay1      0.889472  0.334876   2.656 0.007904 **
## GroupExp       -0.766736  0.330731  -2.318 0.020433 *
## GenderM        -0.005652  0.338143  -0.017 0.986664
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 225.78  on 198  degrees of freedom
## AIC: 235.78
##
## Number of Fisher Scoring iterations: 4

modell17<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Group+Age, data=pep1,family="binomial")
summary(modell17)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group +
##      Age, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8957  -0.8374  -0.6016   1.0598   2.1964
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -32.72238  12.68258  -2.580 0.009877 **
## Durationofsurgery  0.42020  0.11592   3.625 0.000289 ***
## MaxtempDay1      0.86414  0.33510   2.579 0.009916 **
## GroupExp       -0.78224  0.33230  -2.354 0.018570 *
## Age            -0.01812  0.01475  -1.229 0.219161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 224.26  on 198  degrees of freedom
## AIC: 234.26
##
## Number of Fisher Scoring iterations: 4
```

```
modell18<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Group+BMI, data=pep1,family="binomial")
summary(modell18)
```

```
##
```

```
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group +
##     BMI, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8167  -0.8407  -0.6054   1.1065   2.2197
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -35.00570    12.60658  -2.777  0.005490 **
## Durationofsurgery  0.40705     0.11544   3.526  0.000422 ***
## MaxtempDay1      0.88976     0.33313   2.671  0.007565 **
## GroupExp       -0.76799     0.33082  -2.322  0.020259 *
## BMI             0.01348     0.04298   0.314  0.753852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 225.68  on 198  degrees of freedom
## AIC: 235.68
##
## Number of Fisher Scoring iterations: 4
```

```
#Updated the AIC table with the AIC values of the new models fitted using duration of surgery, Max Temp
AIC_table4 <- data.frame(Model_Number = c(1, 2, 3, 4, 5, 6,7,8,9,10,11,12,13,14,15,16,17,18), Predictors=, AIC=)
AIC_table4
```

##	Model_Number	Predictors	AIC
## 1	1	Group	248.37
## 2	2	Gender	255.24
## 3	3	Age	254.80
## 4	4	BMI	255.44
## 5	5	Duration of Surgery	243.95
## 6	6	Max Temp Day1	249.11
## 7	7	Duration of Surgery+Group	239.19
## 8	8	Duration of Surgery+Gender	245.90
## 9	9	Duration of Surgery+Age	244.19
## 10	10	Duration of Surgery+BMI	245.89
## 11	11	Duration of Surgery+Max Temp Day1	237.30
## 12	12	Duration of Surgery+Max Temp Day1+Group	233.78
## 13	13	Duration of Surgery+Max Temp Day1+Gender	239.30
## 14	14	Duration of Surgery+Max Temp Day1+Age	237.96
## 15	15	Duration of Surgery+Max Temp Day1+BMI	239.22
## 16	16	Duration of Surgery+Max Temp Day1+Group+Gender	235.78
## 17	17	Duration of Surgery+Max Temp Day1+Group+Age	234.26
## 18	18	Duration of Surgery+Max Temp Day1+Group+BMI	235.68

From the updated table above, we can see that the best model is still model number 12(Duration of Surgery+Max Temp Day1+Group) with the lowest AIC of 233.78. Therefore, model number 12 with Duration of Surgery, Max Temp Day1 and Group as predictors is still the best fitted model based on AIC. However, we also have to consider the interaction terms among these predictors.

#fitted models using Duration of surgery, Max Temp Day1 and Group with interaction terms

```
model19<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Group+Durationofsurgery*MaxtempDay1, data=pep1,family=
summary(model19)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group +
##     Durationofsurgery * MaxtempDay1, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7566  -0.8264  -0.6163   1.0697   2.1725
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -25.6274     21.5892  -1.187   0.235
## Durationofsurgery    -3.4707      7.5531  -0.460   0.646
## MaxtempDay1         0.6476      0.5762   1.124   0.261
## GroupExp          -0.7788      0.3322  -2.345   0.019 *
## Durationofsurgery:MaxtempDay1  0.1038      0.2025   0.513   0.608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 225.52  on 198  degrees of freedom
## AIC: 235.52
##
## Number of Fisher Scoring iterations: 4
```

```
model20<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Group+Durationofsurgery*Group, data=pep1,family="bin
summary(model20)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group +
##     Durationofsurgery * Group, family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6343  -0.8521  -0.5938   1.1047   2.2757
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -34.0453     12.5557  -2.712  0.00670 **
## Durationofsurgery    0.3274      0.1463   2.238  0.02520 *
## MaxtempDay1         0.8772      0.3330   2.634  0.00843 **
## GroupExp          -1.1736      0.6272  -1.871  0.06132 .
## Durationofsurgery:GroupExp  0.1746      0.2254   0.775  0.43845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 251.47 on 202 degrees of freedom
## Residual deviance: 225.18 on 198 degrees of freedom
## AIC: 235.18
##
## Number of Fisher Scoring iterations: 4
```

```
model21<-glm(Fever ~ Durationofsurgery+MaxtempDay1+Group+MaxtempDay1*Group, data=pep1,family="binomial")
summary(model21)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group +
## MaxtempDay1 * Group, family = "binomial", data = pep1)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.8245 -0.8360 -0.6157 1.0922 2.1955
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -35.2261 16.5730 -2.126 0.033543 *
## Durationofsurgery 0.4029 0.1147 3.514 0.000442 ***
## MaxtempDay1 0.9043 0.4404 2.053 0.040063 *
## GroupExp 0.5644 24.8733 0.023 0.981897
## MaxtempDay1:GroupExp -0.0355 0.6633 -0.054 0.957318
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 251.47 on 202 degrees of freedom
## Residual deviance: 225.78 on 198 degrees of freedom
## AIC: 235.78
##
## Number of Fisher Scoring iterations: 4
```

```
#Updated the AIC table with the AIC values of the new models fitted using duration of surgery, Max Temp
AIC_table5 <- data.frame(Model_Number = c(1, 2, 3, 4, 5, 6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21), I
AIC_table5
```

```
## Model_Number
## 1 1
## 2 2
## 3 3
## 4 4
## 5 5
## 6 6
## 7 7
## 8 8
## 9 9
## 10 10
```

```
## 11      11
## 12      12
## 13      13
## 14      14
## 15      15
## 16      16
## 17      17
## 18      18
## 19      19
## 20      20
## 21      21
##
##                                Predictors    AIC
## 1                                Group 248.37
## 2                                Gender 255.24
## 3                                Age 254.80
## 4                                BMI 255.44
## 5                    Duration of Surgery 243.95
## 6                    Max Temp Day1 249.11
## 7            Duration of Surgery+Group 239.19
## 8            Duration of Surgery+Gender 245.90
## 9                    Duration of Surgery+Age 244.19
## 10                   Duration of Surgery+BMI 245.89
## 11            Duration of Surgery+Max Temp Day1 237.30
## 12            Duration of Surgery+Max Temp Day1+Group 233.78
## 13            Duration of Surgery+Max Temp Day1+Gender 239.30
## 14            Duration of Surgery+Max Temp Day1+Age 237.96
## 15            Duration of Surgery+Max Temp Day1+BMI 239.22
## 16            Duration of Surgery+Max Temp Day1+Group+Gender 235.78
## 17            Duration of Surgery+Max Temp Day1+Group+Age 234.26
## 18            Duration of Surgery+Max Temp Day1+Group+BMI 235.68
## 19 Durationofsurgery+MaxtempDay1+Group+Durationofsurgery*MaxtempDay1 235.52
## 20            Durationofsurgery+MaxtempDay1+Group+Durationofsurgery*Group 235.18
## 21            Durationofsurgery+MaxtempDay1+Group+MaxtempDay1*Group 235.78
```

Even after checking the interaction terms, from the updated table above, we can now confirm that the best model really is model number 12(Duration of Surgery+Max Temp Day1+Group) with the lowest AIC of 233.78. It is also important to note that all the interaction terms are insignificant with very high p-values whereas model number 12 have all 3 predictors as significant with p-values less than 0.05.This phenomenon confirms our finding from question 6 earlier that the categorical predictors and the numerical predictors do not have any significant relationship. With all these said, based on AIC and significance of the predictors, model number 12 with Duration of Surgery, Max Temp Day1 and Group as predictors without any interaction terms is the best fitted model.

```
#got summary of the model
summary(model12)
```

```
##
## Call:
## glm(formula = Fever ~ Durationofsurgery + MaxtempDay1 + Group,
##      family = "binomial", data = pep1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.8274 -0.8334 -0.6181 1.0940 2.1992
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -34.6476    12.5407  -2.763 0.005730 **
## Durationofsurgery  0.4026     0.1146   3.513 0.000443 ***
## MaxtempDay1      0.8889     0.3329   2.670 0.007584 **
## GroupExp        -0.7667     0.3307  -2.318 0.020435 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.47  on 202  degrees of freedom
## Residual deviance: 225.78  on 199  degrees of freedom
## AIC: 233.78
##
## Number of Fisher Scoring iterations: 4
```

From the summary above, we are able to identify the following:

```
residual_deviance <- 225.78
degrees_of_freedom <- 199

dev_df_ratio <- residual_deviance/degrees_of_freedom
dev_df_ratio
```

```
## [1] 1.134573
```

The deviance-to-df ratio is very high with a value of 1.134573. This means that the model is good. However, this may be an indicator that the model is too good and may be overfitting.

To further investigate, we will also be computing the Sensitivity and Specificity of the model. Note: Prior probability of having fever is

$$63/203 = 0.3103448$$

```
#computed sensitivity and specificity

#created new data for the prediction
newdata<-data.frame(Durationofsurgery=pep1$Durationofsurgery,MaxtempDay1=pep1$MaxtempDay1, Group = pep1$Group)

#created table of the predicted and observed values
probabilities<- model12 %>% predict(newdata,type="response")
predicted<-ifelse(probabilities>0.3103448,"have fever","no fever")
fever2<-pep1$Fever
fever2<-as.numeric(as.character(fever2))
observed<-ifelse(feve2 > 1,"have fever","no fever")
xtabs(~observed+predicted,data=newdata)
```

```
##              predicted
## observed   have fever no fever
##   have fever      46      17
##   no fever       48      92
```

Based on the table above, we can compute the following values: Sensitivity(True Positives):

$$46/63 = 0.7301587$$

Specificity(True Negatives):

$$92/140 = 0.6571429$$

Based on Sensitivity, the model is good because it can predict subjects with fever with 73% accuracy. However, for specificity, it's relatively lower in accuracy but still moderately good.

To further assess and confirm our findings, we will also be plotting an ROC curve.

```
#plotted ROC curve
```

```
#imported pROC library  
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 4.2.3
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
```

```
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:mosaic':
```

```
##
```

```
##      cov, var
```

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

```
##
```

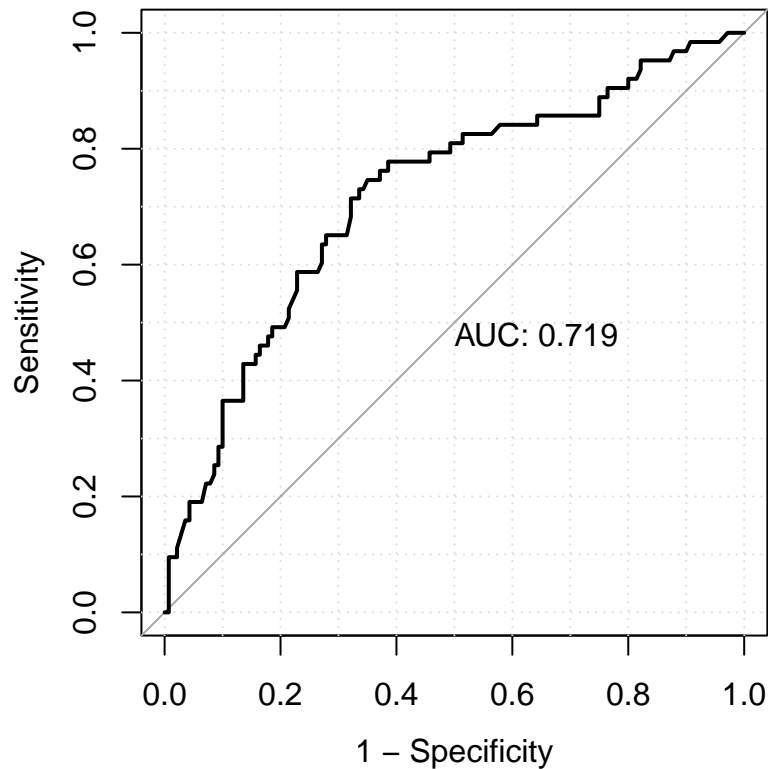
```
##      cov, smooth, var
```

```
test_prob<-predict(model12,newdata,type="response")  
test_roc<-roc(newdata$Fever ~ test_prob)
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```

```
par(pty="s")  
plot(test_roc, grid=TRUE,legacy.axes=TRUE, plot=TRUE,print.auc=TRUE)
```

Based on the result, AUC value is greater than 0.7 (0.719) so we can say that the model is good for prediction or a good fit.

```
#got summary of coefficients of the model
summary(model12)$coefficients
```

	Estimate	Std. Error	z value	Pr(> z)
## (Intercept)	-34.6476059	12.5406829	-2.762817	0.0057304971
## Durationof surgery	0.4025637	0.1145842	3.513257	0.0004426492
## MaxtempDay1	0.8888667	0.3329043	2.670037	0.0075842967
## GroupExp	-0.7667027	0.3307233	-2.318260	0.0204351794

$$\text{logit}\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) = -34.6476059 + 0.4025637 \text{Durationof surgery}_i + 0.8888667 \text{MaxtempDay1}_i - 0.7667027 \text{Group}_i$$

where:

$$\hat{\pi}_i$$

is equal to the P(i'th observed person that has a fever) and fever has a Bernoulli

$$(\pi)$$

distribution.

Durationof surgery_i is the duration of surgery for the i'th subject.

MaxtempDay1i is the maximum temperature for the ith subject.
Groupi is Con for Controlled and Exp for Experimental.

Durationofsurgery is significant with p-value of 0.000443.

MaxtempDay1 is significant with p-value of 0.007584

Group is significant with p-value of 0.020435

```
#got summaries of the significant predictors
summary(pep1$Durationofsurgery)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.50   1.00   1.50   2.08   3.00   8.50
```

```
summary(pep1$MaxtempDay1)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     36.00   37.00   37.40   37.43   37.80   38.90
```

```
summary(pep1$Group)
```

```
## Con Exp
## 104  99
```

Since Durationofsurgery goes from 0.50 to 8.50 and the min value is 0.50 and beginning from 1st Quantile until the Mean the value is increased by 0.50, it is reasonable to look at an increase of 0.50 in the case of Durationofsurgery. The expected result of Durationofsurgery is

$$e^{0.4025637*0.5} = 1.223$$

Therefore, a 0.5 increase in Durationofsurgery results in a 22.3% increase in the odds of getting a fever (conditioned on MaxtempDay1 and Group)

Since MaxtempDay1 only goes from 36.0 to 38.90, it is reasonable to look at increments of an increase of 0.1 in MaxtempDay1. The expected result of MaxtempDay1 is

$$e^{0.8888667*0.1} = 1.093$$

Therefore, a 0.1 increase in MaxtempDay1 results in a 9.3% increase in the odds of getting a fever (conditioned on Durationofsurgery and Group)

Here, we see that Con has more subjects than Exp which makes sense that we are using it as our reference category.

For Con, the expected fever result is

$$e^{-34.6476059+0.4025637Durationofsurgeryi+0.8888667MaxtempDay1i-0.7667027}$$

OR

$$e^{-34.6476059+0.4025637Durationofsurgeryi+0.8888667MaxtempDay1i} X e^{-0.7667027}$$

OR

$$e^{-34.6476059+0.4025637Durationofsurgeryi+0.8888667MaxtempDay1i} X 0.4645$$

which is 0.4645 times the expected fever result for the Exp group as the Exp group's expected fever result is

$$e^{-34.6476059+0.4025637Durationofsurgeryi+0.8888667MaxtempDay1i}$$

A subject belonging to the Con group results in a 53.55% decrease in the expected odds of having a fever, conditioned on Duration of surgery and Max temp Day1. This is the multiplicative effect of belonging to the Con group on the odds of having a fever. In other words, subjects belonging to the Exp group increases odds of having a fever by 53.55%, conditioned on Duration of surgery and Max temp Day1.