HW 4

Mohammad Imtiaz Nur

4/16/2021

ID: 1878074

Task 1

Here we will perform Perform ANOVA analysis of covariance using FWT as the response, weaning time as the factor, and WWT as the covariate.

```
?Anova()
weaning_time = c("EARLY", "EARLY", "EARLY", "EARLY", "EARLY", "EARLY", "EARLY", "MEDIUM", "MEDIUM", "ME

WWT = c(9, 9, 12, 11, 15, 15, 14, 16, 16, 15, 14, 14, 12, 10, 18, 17, 16, 15, 14, 14, 13)

FWT = c(37, 28, 40, 45, 44, 50, 45, 48, 45, 47, 46, 40, 36, 33, 45, 38, 35, 38, 34, 37, 37)


df1 <- data.frame(FWT, weaning_time, WWT)

fit <- aov(FWT ~ weaning_time + WWT,df1)

Anova(fit, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: FWT
##              Sum Sq Df F value    Pr(>F)
## (Intercept)  138.13  1  12.003 0.0029634 **
## weaning_time 289.82  2  12.592 0.0004416 ***
## WWT          394.08  1  34.245 1.923e-05 ***
## Residuals    195.63 17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the anova test, the p value of weaning time (0.0004416) is much less than the significance level (0.05). Hence, we can say that there is significant evidence that weaning time is related to the response FWT.

```
TukeyHSD(fit, "weaning_time")
```

```
## Warning in replications(paste("~", xx), data = mf): non-factors ignored: WWT
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = FWT ~ weaning_time + WWT, data = df1)
##
## $weaning_time
##                     diff        lwr      upr     p adj
## LATE-EARLY   -3.5714286 -8.2231229 1.080266 0.1501964
## MEDIUM-EARLY  0.8571429 -3.7945514 5.508837 0.8849297
## MEDIUM-LATE   4.4285714 -0.2231229 9.080266 0.0634498
```

Using HSD for paired test we can see that the p-adjusted values for all pairs are greater than significance level (0.05). Hence, the groups related to different time period don't differ significantly in terms of means.

Task 2

(a)

```r
pave <- c("ASPHALT1", "ASPHALT1", "ASPHALT1", "ASPHALT1", "ASPHALT1", "ASPHALT1",
           "ASPHALT1", "CONCRETE", "CONCRETE", "CONCRETE", "CONCRETE", "CONCRETE",
           "CONCRETE", "CONCRETE", "CONCRETE", "CONCRETE", "CONCRETE", "ASPHALT2",
           "ASPHALT2", "ASPHALT2", "ASPHALT2", "ASPHALT2", "ASPHALT2", "ASPHALT2",
           "ASPHALT2", "ASPHALT2")

tread <- c(1, 1, 2, 2, 2, 6, 6, 1, 1, 1, 1, 2, 2, 2, 6, 6, 6, 1, 1, 1, 2, 2, 2, 6, 6, 6)

speed <- c(36.5, 34.9, 40.2, 38.2, 38.2, 43.7, 43.0, 40.2, 41.6, 42.6, 41.6,
           40.9, 42.3, 45.0, 47.1, 51.2, 51.2, 33.4, 38.2, 34.9, 36.8, 35.4,
           35.4, 40.2, 40.9, 43.0)

df2 <- data.frame(pave, tread, speed)

# using Dummy
df2$tread <- factor(tread)
model1 <- glm(speed ~ pave * tread, data = df2)
Anova(model1,type = 3)
```

```
## Analysis of Deviance Table (Type III tests)
##
## Response: speed
##            LR Chisq Df Pr(>Chisq)
## pave        29.6916  2  3.569e-07 ***
## tread       22.5807  2  1.249e-05 ***
## pave:tread   3.9013  4     0.4195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We used the glm model for the dummy variable analysis of variance and measured the significance using chi-squared test from Anova function considering type-3 error. As the p-value is less than 0.05 for both factors (pave and tread), hence each factor has significant effect on speed . But the interaction between these two variables is not significant as the p-value is more than 0.05.

Standard Approach

```r
model2 = aov(speed ~ pave * tread, data = df2)
summary(model2)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## pave          2 237.19  118.59  45.185 1.57e-07 ***
## tread         2 244.36  122.18  46.551 1.27e-07 ***
## pave:tread    4  10.24    2.56   0.975    0.447    
## Residuals    17  44.62    2.62                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here we used anova model taking both pave and tread as factor. As the p-value is less than 0.05 for both factors (pave and tread), hence each factor has significant effect on speed . But the interaction between these two variables is not significant as the p-value is more than 0.05.

(b)

To consider the tread as a measured variable we performed analysis of covariance considering tread as covariate.

```r
df2$tread <- as.numeric(tread)
model3 = aov(speed ~ pave + tread, data = df2)
summary(model3)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## pave          2 237.19  118.59   47.54 1.03e-08 ***
## tread         1 244.33  244.33   97.94 1.46e-09 ***
## Residuals    22  54.88    2.49                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The MSE of the model we get is 2.49 which is even better than the MSE from the ANOVA model of normal approach (MSE 2.62).

(c)

In this section we consider all pavement types characterized by their co-efficient of friction at 40 mph. So, the pave is also considered a measured variable instead of factor here.

```r
pave <- c(0.35, 0.35, 0.35, 0.35, 0.35, 0.35, 0.35, 0.48, 0.48, 0.48, 0.48, 0.48,
          0.48, 0.48, 0.48, 0.48, 0.48, 0.24, 0.24, 0.24, 0.24, 0.24, 0.24, 0.24,
          0.24, 0.24)

tread <- c(1, 1, 2, 2, 2, 6, 6, 1, 1, 1, 1, 2, 2, 2, 6, 6, 6, 1, 1, 1, 2, 2, 2,
           6, 6, 6)

speed <- c(36.5, 34.9, 40.2, 38.2, 38.2, 43.7, 43.0, 40.2, 41.6, 42.6, 41.6, 40.9,
```

```r
           42.3, 45.0, 47.1, 51.2, 51.2, 33.4, 38.2, 34.9, 36.8, 35.4, 35.4, 40.2,
           40.9, 43.0)

df3 <- data.frame(pave, tread, speed)


model4 <- aov(speed ~ pave * tread,data = df3)
summary(model4)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## pave          1 226.52  226.52  82.864 6.49e-09 ***
## tread         1 245.40  245.40  89.771 3.19e-09 ***
## pave:tread    1   4.35    4.35   1.592     0.22    
## Residuals    22  60.14    2.73                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparison of models

```r
mod1 <- mean(model1$residuals^2)
mod2 <- mean(model2$residuals^2)
mod3 <- mean(model3$residuals^2)
mod4 <- mean(model4$residuals^2)

print("Model1:")
```

```
## [1] "Model1:"
```

```r
mod1
```

```
## [1] 1.71609
```

```r
print("Model2:")
```

```
## [1] "Model2:"
```

```r
mod2
```

```
## [1] 1.71609
```

```r
print("Model3:")
```

```
## [1] "Model3:"
```

```r
mod3
```

```
## [1] 2.110934
```

```r
print("Model4:")
```

```
## [1] "Model4:"
```

```r
mod4
```

```
## [1] 2.313027
```

From the result of the comparison of the MSEs of all 4 models we can see that Model-1 and Model-2 where both pave and tread considered as factor gives the best result.

Task 3

(a)

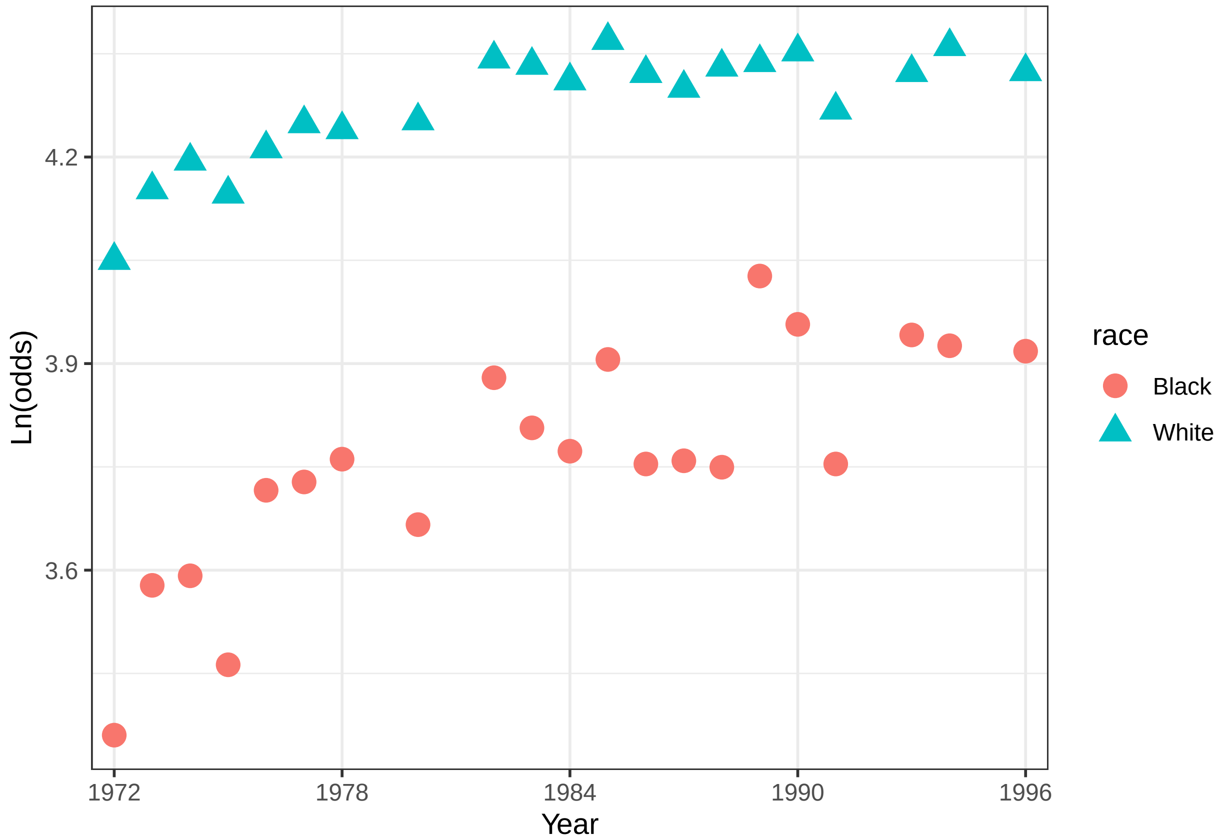
```r
year <- c(1972, 1973, 1974, 1975, 1976, 1977, 1978, 1980, 1982, 1983, 1984, 1985,
          1986, 1987, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1972, 1973, 1974,
          1975, 1976, 1977, 1978, 1980, 1982, 1983, 1984, 1985, 1986, 1987, 1988,
          1989, 1990, 1991, 1993, 1994, 1996)

race <- c("White", "White", "White", "White", "White", "White", "White", "White", "White",
          "White", "White", "White", "White", "White", "White", "White", "White", "White",
          "White", "White", "White", "Black", "Black", "Black", "Black", "Black", "Black",
          "Black", "Black", "Black", "Black", "Black", "Black", "Black", "Black", "Black",
          "Black", "Black", "Black", "Black", "Black", "Black")
opinion <- c(57.4, 63.6, 66.3, 63.2, 67.5, 70.0, 69.4, 70.3, 76.9, 76.2, 74.5, 79.0,
             75.3, 73.7, 76.0, 76.5, 77.7, 71.4, 75.4, 78.3, 75.5, 28.8, 35.8, 36.3,
             31.9, 41.1, 41.6, 43.0, 39.1, 48.4, 45.0, 43.5, 49.7, 42.7, 42.9, 42.5,
             56.1, 52.3, 42.7, 51.5, 50.7, 50.3)

df4 <- data.frame(year, race, opinion)

ggplot(df4, aes(x=year, y = log(opinion))) +
  geom_point(aes(shape=race,color=race),size=4) +
  theme_bw()+
  scale_x_discrete(limits= c(1972, 1978, 1984, 1990, 1996))+
  labs(y="Ln(odds)", x = "Year")
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean ‘limits = factor(...)‘ or ‘scale_*_continuous()‘?
```



The plot shows that the ln(odds) of favoring the death penalty are much higher among whites. In both races, the ln(odds) increases quickly in the 1970s but the rate of increase slowed by the mid 1980s. The size of the gap between whites and blacks seems stable across the years.

(b)

To create the regression model first the odds are converted into factor opinions where 1 for support and 0 for non-support. To convert it, it is assumed each year has 1100 voters sample and 400 black voters. We fit the model that replaced YEAR with YEARS = YEAR - 1972, to avoid very large numbers in the quadratic term. We coded RACE = 0 for white and 1 for blacks.

```
years <- c()
races <- c()
opinions <- c()
i = 1
while (i<= length(df4$year )){
  if (df4$race[i]=="White"){
    x <- rbinom(1100,1,df4$opinion[i]/100)
      j = 1
    while(j <= 1100){
      races <- c(races, 0)
      years <- c(years, df4$year[i]-1972)
      j = j + 1
    }
  }
```

```
  if (df4$race[i]=="Black"){
    x <- rbinom(400,1,df4$opinion[i]/100)

    j = 1
    while(j <= 400){
      races <- c(races, 1)
      years <- c(years, df4$year[i]-1972)
      j = j + 1
    }
  }

  opinions <- c(opinions,x)
  i = i + 1
}

years_square <- years**2
df5 <- data.frame(years, years_square,races, opinions)


model_1 <- glm(opinions ~ races + years, data = df5, family = "binomial")
model_1$coefficients
```

```
## (Intercept)       races       years
##  0.62367296 -1.25237721  0.03002836
```

```
model_2 <- glm(opinions ~ races+ years + years_square, data = df5, family = "binomial")
model_2$coefficients
```

```
##  (Intercept)        races        years years_square
##  0.473557926 -1.254691416  0.074387645 -0.001956685
```

The equation for the model for calculating the probability that a person will support the death penalty, as a function of race and year is:

$$Ln(odds) = 0.6237 + (-1.2524)Race + 0.03Years$$

Adding a quadratic term $Years^2$, makes the model equation as follows:

$$Ln(odds) = 0.4736 + (-1.2547)Race + 0.0744Years + (-0.002)Years^2$$

```
anova(model_1, model_2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: opinions ~ races + years
## Model 2: opinions ~ races + years + years_square
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1     31497      38450
## 2     31496      38396  1   54.103 1.902e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the significance test using Chi-square results, it is evident that the quadratic term improves the model (model 2) as the p value is much less than 0.05.

(c)

```
model_3 <- glm(opinions ~ races+ years + years_square + races * years + races* years_square,
               data = df5, family = "binomial")
model_3$coefficients
```

```
##        (Intercept)              races              years       years_square
##        0.433180428       -1.109114030        0.084797021       -0.002382520
##        races:years races:years_square
##       -0.035246206        0.001416135
```

Adding an interaction of RACE with YEARS and RACE with $YEARS^2$ gave the equation:

$$Ln(odds) = 0.4332+(-1.1091)Race+0.0848Years+(-0.0024)Years^2+(-0.0352)Race*Year+0.0014Race*Year^2$$

```
anova(model_3, model_2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: opinions ~ races + years + years_square + races * years + races *
##     years_square
## Model 2: opinions ~ races + years + years_square
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1     31494      38389
## 2     31496      38396 -2  -6.5952  0.03697 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The significance test using ANova shows that there is no significant evidence of an interaction.

(d)

It is reasonable to refer to the gap as enduring. The first model showed extremely strong evidence for a gap. Since the interactions were not significant, the gap in the ln(odds) does not appear to be changing much with time. However, the plot of the probabilities might show a slight change in the size of the gap, but it would be small.

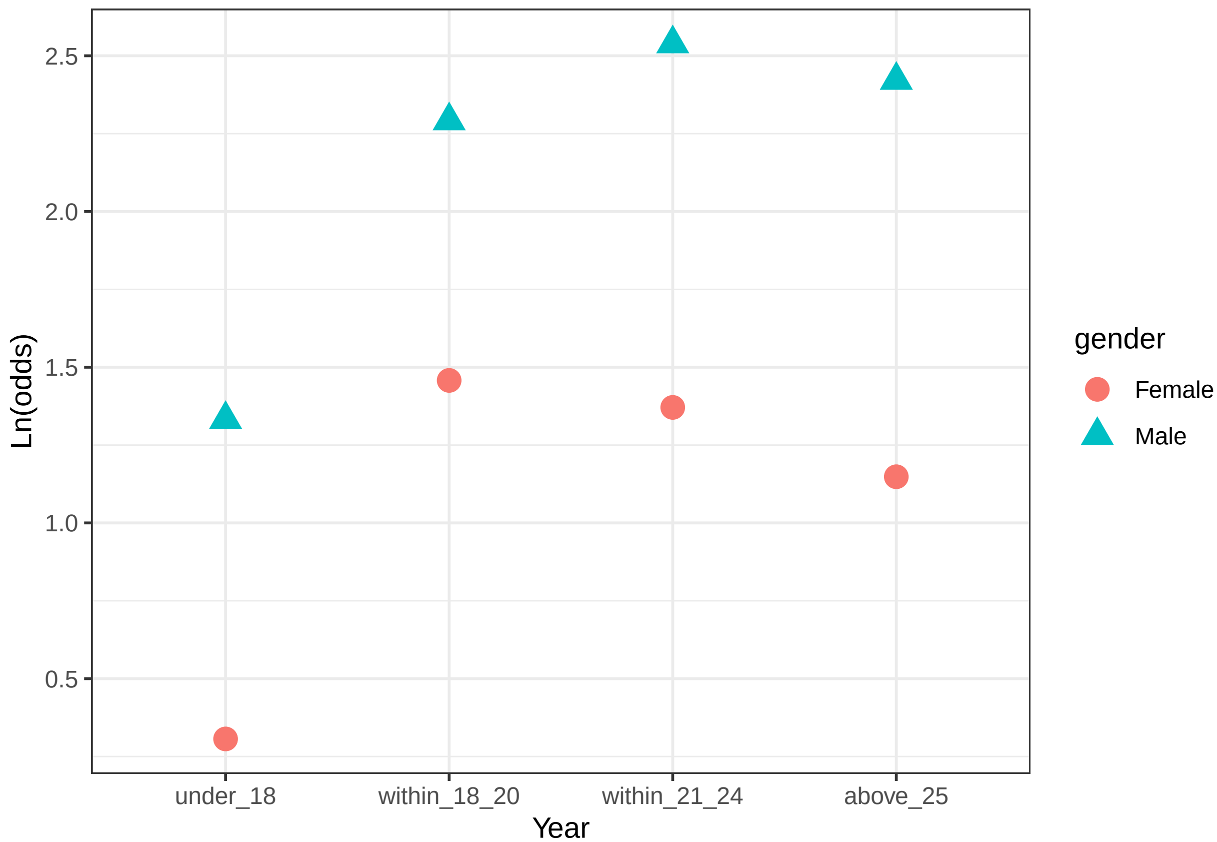
Task 4

(a)

```
age <- c("under_18","under_18", "within_18_20","within_18_20", "within_21_24", "within_21_24",
         "above_25","above_25")
gender <- c("Male", "Female", "Male", "Female", "Male", "Female", "Male", "Female")
total <- c(14589, 8612, 21708, 10941, 25664, 13709,  41304, 25183)
A_R <- c(   553,    117, 2147,    470,    3250,   540,    4652,   794)

df6 <- data.frame(age, gender, total, A_R)

ggplot(df6, aes(x=age, y = log(A_R/total*100))) +
  geom_point(aes(shape=gender,color=gender),size=4) +
  theme_bw() +
  scale_x_discrete(limits= c("under_18", "within_18_20", "within_21_24", "above_25"))+
  labs(y="Ln(odds)", x = "Year")
```

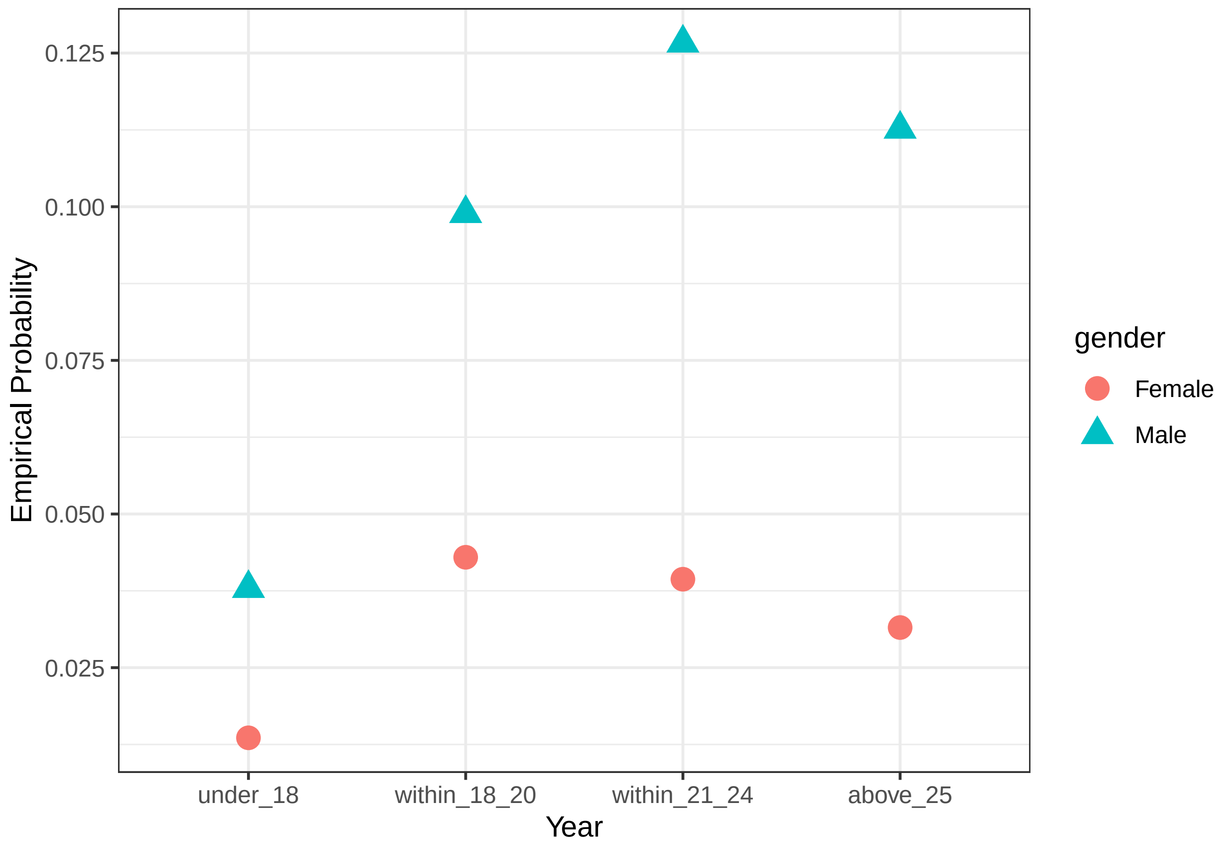


From the plot it is apparent that male drivers are more likely to be in a accident than female drivers. The gap between male and female drivers increases= with the age category. So there is a strong indication of interaction.

(b)

```
ggplot(df6, aes(x=age, y = A_R/total)) +
  geom_point(aes(shape=gender,color=gender),size=4) +
  theme_bw() +
```

```
scale_x_discrete(limits= c("under_18", "within_18_20", "within_21_24", "above_25"))+
labs(y="Empirical Probability", x = "Year")
```



From the plot using empirical probability, it is also apparent that male drivers are more likely to be in a accident than female drivers. The gap between male and female drivers increase more strongly with the age category. So there is a strong indication of interaction.

(c)

Here, I have created dummy variables using fastDummies. This function has created four dummy variable for age group, age_under_18, age_within_18_20, age_within_21_24, age_above_25 and two variables for gender group gender_Male, gender_Female. The presence of values for the variable treated as 1 and absence as 0. Later to construct the binary factored response, total number of accidents are converted into total number of rows.

```
df7 <- fastDummies::dummy_cols(df6, select_columns = c("age","gender"))

age_under_18 <- c()
age_within_18_20 <- c()
age_within_21_24 <- c()
age_above_25 <- c()
gender_Male <- c()
gender_Female <- c()
alocohole_Related <- c()
```

```r
i = 1
while (i<= length(df7$A_R )){

  j = 1

  while(j <= df7$total[i]){

      age_under_18 <- c(age_under_18, df7$age_under_18[i])
      age_within_18_20 <- c(age_within_18_20, df7$age_within_18_20[i])
      age_within_21_24 <- c(age_within_21_24, df7$age_within_21_24[i])
      age_above_25 <- c(age_above_25, df7$age_above_25[i])
      gender_Male <- c(gender_Male, df7$gender_Male[i])
      gender_Female <- c(gender_Female, df7$gender_Female[i])

    if (j <= df7$total[i] - df7$A_R[i]){
      alocohole_Related <- c(alocohole_Related, 0)
    }
    else{
      alocohole_Related <- c(alocohole_Related, 1)
    }

    j = j + 1

  }
  i = i + 1
}
df8 <- data.frame(alocohole_Related, age_under_18,age_within_18_20, age_within_21_24, age_above_25, gend

model_1 <- glm(alocohole_Related ~ age_under_18 + age_within_18_20 + age_within_21_24 + age_above_25 + g
model_1$coefficients
```

```
##      (Intercept)     age_under_18 age_within_18_20 age_within_21_24
##      -3.30285118      -1.11826716      -0.06724028       0.14977620
##     age_above_25      gender_Male    gender_Female
##               NA       1.21439080               NA
```

```r
summary(model_1)
```

```
##
## Call:
## glm(formula = alocohole_Related ~ age_under_18 + age_within_18_20 +
##     age_within_21_24 + age_above_25 + gender_Male + gender_Female,
##     family = "binomial", data = df8)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5185  -0.4833  -0.2893  -0.2688   2.9776
##
## Coefficients: (2 not defined because of singularities)
##                  Estimate Std. Error  z value Pr(>|z|)
## (Intercept)      -3.30285    0.02536 -130.258  < 2e-16 ***
## age_under_18     -1.11827    0.04183  -26.731  < 2e-16 ***
## age_within_18_20 -0.06724    0.02502   -2.688  0.00719 **
```

```
## age_within_21_24  0.14978    0.02239    6.690 2.23e-11 ***
## age_above_25           NA         NA       NA       NA    
## gender_Male       1.21439    0.02542   47.773  < 2e-16 ***
## gender_Female          NA         NA       NA       NA    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 88124  on 161709  degrees of freedom
## Residual deviance: 84064  on 161705  degrees of freedom
## AIC: 84074
##
## Number of Fisher Scoring iterations: 8
```

The model equation can be written as:

Ln(odds) = -3.30 -1.12 * age_under_18 - 0.07 * age_within_18_20 + 0.15 * age_within_21_24 + 1.21 * gender

From the summary of the regression model we can see the age groups age_under_18, age_within_18_20 and age_within_21_24 are significant. So by default age_above_25 is significant. Similarly gender_Male is significant, so by default gender_Female is significant. From figure a, we can interpret it as, Females have significantly lower odds of alcohol related crash and older age groups has higher odds than younger age groups.

(d)

```r
model_2 <- glm(alocohole_Related ~ age_under_18 + age_within_18_20 + age_within_21_24 +
                 age_above_25 + gender_Male + gender_Female + age_under_18 * gender_Male +
                 age_under_18 * gender_Female + age_within_18_20 * gender_Male +
                 age_within_18_20 * gender_Female + age_within_21_24 * gender_Male +
                 age_within_21_24 * gender_Female + age_above_25 * gender_Male +
                 age_above_25 * gender_Female, data = df8, family = "binomial")

model_2$coefficients
```

```
##                    (Intercept)                   age_under_18
##                    -3.42480403                    -0.86025507
##               age_within_18_20               age_within_21_24
##                     0.32117191                     0.23075231
##                   age_above_25                    gender_Male
##                             NA                     1.36063327
##                  gender_Female       age_under_18:gender_Male
##                             NA                    -0.30959690
##     age_under_18:gender_Female   age_within_18_20:gender_Male
##                             NA                    -0.46646742
## age_within_18_20:gender_Female   age_within_21_24:gender_Male
##                             NA                    -0.09761231
## age_within_21_24:gender_Female       age_above_25:gender_Male
##                             NA                             NA
##     age_above_25:gender_Female
##                             NA
```

```r
summary(model_2)
```

```
##
## Call:
## glm(formula = alocohole_Related ~ age_under_18 + age_within_18_20 +
##     age_within_21_24 + age_above_25 + gender_Male + gender_Female +
##     age_under_18 * gender_Male + age_under_18 * gender_Female +
##     age_within_18_20 * gender_Male + age_within_18_20 * gender_Female +
##     age_within_21_24 * gender_Male + age_within_21_24 * gender_Female +
##     age_above_25 * gender_Male + age_above_25 * gender_Female,
##     family = "binomial", data = df8)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5204  -0.4889  -0.2963  -0.2531   2.9322
##
## Coefficients: (7 not defined because of singularities)
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -3.42480    0.03606 -94.971  < 2e-16 ***
## age_under_18                    -0.86026    0.09982  -8.618  < 2e-16 ***
## age_within_18_20                 0.32117    0.05936   5.411 6.28e-08 ***
## age_within_21_24                 0.23075    0.05682   4.061 4.88e-05 ***
## age_above_25                          NA         NA      NA       NA
## gender_Male                      1.36063    0.03928  34.642  < 2e-16 ***
## gender_Female                         NA         NA      NA       NA
## age_under_18:gender_Male        -0.30960    0.10994  -2.816  0.00486 **
## age_under_18:gender_Female            NA         NA      NA       NA
## age_within_18_20:gender_Male    -0.46647    0.06544  -7.128 1.02e-12 ***
## age_within_18_20:gender_Female        NA         NA      NA       NA
## age_within_21_24:gender_Male    -0.09761    0.06183  -1.579  0.11439
## age_within_21_24:gender_Female        NA         NA      NA       NA
## age_above_25:gender_Male              NA         NA      NA       NA
## age_above_25:gender_Female            NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 88124  on 161709  degrees of freedom
## Residual deviance: 84011  on 161702  degrees of freedom
## AIC: 84027
##
## Number of Fisher Scoring iterations: 21
```

The interaction model equation stands as :

Ln(odds) = -3.42 - 0.86 * age_under_18 + 0.32 * age_within_18_20 + 0.23 * age_within_21_24 + 1.36 * gender_Male - 0.31 * age_under_18:gender_Male - 0.47 * age_within_18_20:gender_Male - 0.098 * age_within_21_24:gender_Male

It can be written as:

Ln(odds) = -3.34 - 0.86 * age_under_18 + 0.32 * age_within_18_20 + 0.23 * age_within_21_24 + 1.36 * gender - 0.31 * age_under_18:gender - 0.47 * age_within_18_20:gender - 0.098 * age_within_21_24:gender

(e)

```r
anova(model_1, model_2, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: alocohole_Related ~ age_under_18 + age_within_18_20 + age_within_21_24 +
##     age_above_25 + gender_Male + gender_Female
## Model 2: alocohole_Related ~ age_under_18 + age_within_18_20 + age_within_21_24 +
##     age_above_25 + gender_Male + gender_Female + age_under_18 *
##     gender_Male + age_under_18 * gender_Female + age_within_18_20 *
##     gender_Male + age_within_18_20 * gender_Female + age_within_21_24 *
##     gender_Male + age_within_21_24 * gender_Female + age_above_25 *
##     gender_Male + age_above_25 * gender_Female
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1    161705      84064
## 2    161702      84011  3   52.952 1.877e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result of the anova “Likelihood Ratio Test” (LRT) for the null hypothesis is that none of the interactions are significant. But the p_value (1.88*10-11) for model-2 is very much less than significance level 0.05. So we can reject the null hypothesis and say at least one interaction is significant and expected difference in ln(odds) different in at least one age group.

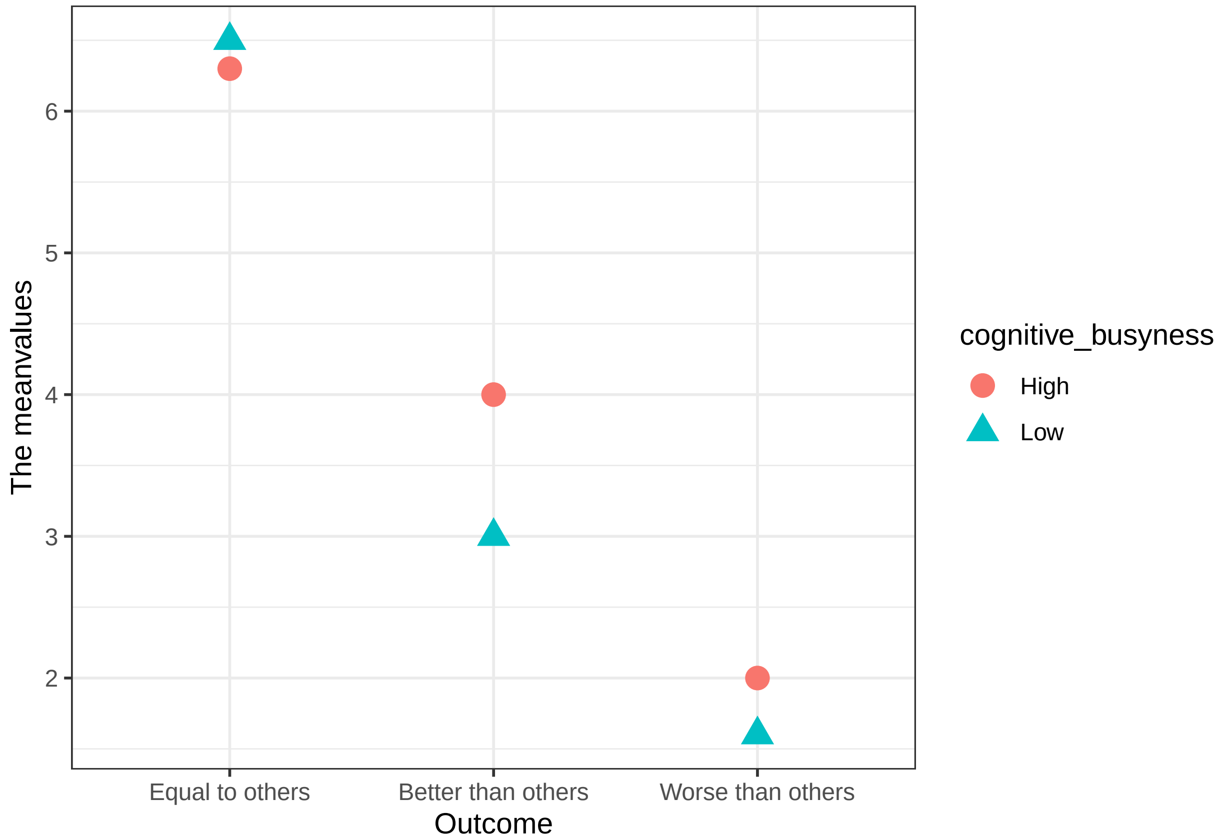
Task 5

(a)

```r
cognitive_busyness <- c("Low", "Low", "Low", "High", "High", "High")
outcome  <- c("Equal to others", "Better than others", "Worse than others",
              "Equal to others", "Better than others", "Worse than others")
values <- c(6.5, 3.0, 1.6, 6.3, 4.0, 2.0)

df9 <- data.frame(cognitive_busyness,outcome,values)

ggplot(df9, aes(x=outcome, y = values)) +
  geom_point(aes(shape=cognitive_busyness,color=cognitive_busyness),size=4) +
  theme_bw() +
  scale_x_discrete(limits= c("Equal to others", "Better than others", "Worse than others"))+
  labs(y="The meanvalues", x = "Outcome")
```



(b)

The plot and test statistics are consistent in showing a very strong main effect of the outcome, with participants showing the highest satisfaction when outcome is perceived as equal. There is a weak main effect for Cognitive Busyness, with a tendency for those with the Low Busyness to have less satisfaction. However, this tendency seems strongest in the better outcome category, leading to a significant interaction.

(c)

Total cell 6. Total category 5 in two groups. Number of independent t-test = $(6 * 5)/2 = 15$

(d)

Yes, as noted in (b), the difference between the High and Low Busyness categories is most pronounced in the Better outcome category. There is less of a difference in the Equal and Worse categories.