Booklet of Code and Output
for
STAD29/STA 1007 Midterm Exam

List of Figures in this document by page:

List of Figures

```
library(MASS)
library(tidyverse)

## -- Attaching packages ---------------------------------------
tidyverse 1.3.1 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()

library(broom)
library(survival)
library(survminer)

## Loading required package:  ggpubr
##
## Attaching package:  'survminer'
## The following object is masked from 'package:survival':
##
##     myeloma
```

Figure 1: Packages

```
minutes  machines
  97.0    7.0
  86.0    6.0
  78.0    5.0
  10.0    1.0
  75.0    5.0
  62.0    4.0
 101.0    7.0
  39.0    3.0
  53.0    4.0
  33.0    2.0
 118.0    8.0
  65.0    5.0
  25.0    2.0
  71.0    5.0
 105.0    7.0
  17.0    1.0
  49.0    4.0
  68.0    5.0
```

Figure 2: Copiers data

```
ggplot(copiers, aes(x=machines, y=minutes))+geom_point()
```

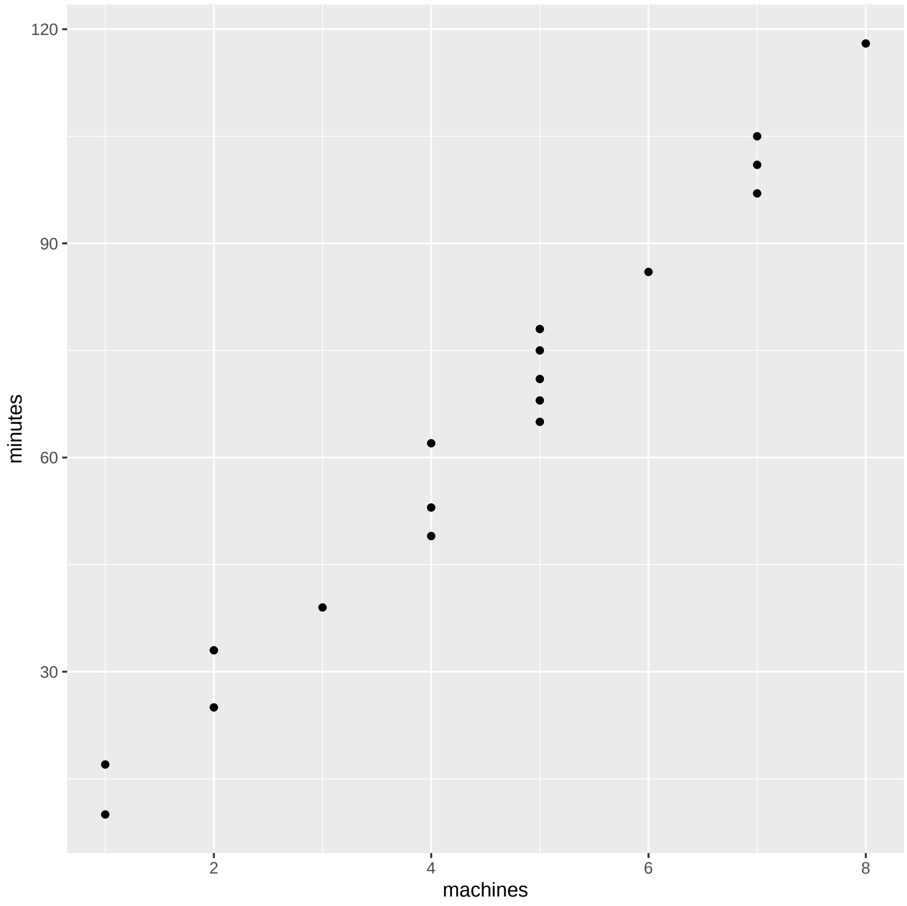


Figure 3: Copiers scatterplot

```
copiers.1=lm(minutes~machines, data=copiers)
summary(copiers.1)

##
## Call:
## lm(formula = minutes ~ machines, data = copiers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6309 -3.2500 -0.2383  4.0235  6.6309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.3221     2.5644  -0.906    0.379
## machines     14.7383     0.5193  28.383  4.1e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.482 on 16 degrees of freedom
## Multiple R-squared:  0.9805,Adjusted R-squared:  0.9793
## F-statistic: 805.6 on 1 and 16 DF,  p-value: 4.097e-15
```

Figure 4: Regression for copiers data

```
new=tibble(machines=6)
p1=predict(copiers.1, new, interval="c")
cbind(new,p1)

##   machines      fit      lwr      upr
## 1        6 86.10738 83.32504 88.88973

p2=predict(copiers.1, new, interval="p")
cbind(new,p2)

##   machines      fit      lwr      upr
## 1        6 86.10738 76.20721 96.00756
```

Figure 5: Predictions for copiers data

```
bottles

## # A tibble: 6 x 3
##   deposit  sold returned
##     <dbl> <dbl>    <dbl>
## 1       2   500       72
## 2       5   500      103
## 3      10   500      170
## 4      20   500      296
## 5      25   500      406
## 6      30   500      449
```

Figure 6: Soft drink bottle return data

```
bottles.1=glm(y~deposit, data=bottles, family=binomial)
summary(bottles.1)

##
## Call:
## glm(formula = y ~ deposit, family = binomial, data = bottles)
##
## Deviance Residuals:
##       1        2        3        4        5        6
##  0.1754   0.4330   0.5784  -2.9193   1.2710   1.2209
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.076565   0.084839  -24.48   <2e-16 ***
## deposit      0.135851   0.004772   28.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1108.171  on 5  degrees of freedom
## Residual deviance:   12.181  on 4  degrees of freedom
## AIC: 53.419
##
## Number of Fisher Scoring iterations: 3
```

Figure 7: Bottles logistic regression

```
probability=predict(bottles.1, bottles, type="response")
preds=cbind(bottles, probability)
preds

##   deposit sold returned probability
## 1       2  500       72   0.1412601
## 2       5  500      103   0.1982432
## 3      10  500      170   0.3278210
## 4      20  500      296   0.6548554
## 5      25  500      406   0.7891326
## 6      30  500      449   0.8806877
```

Figure 8: Predictions for bottles data

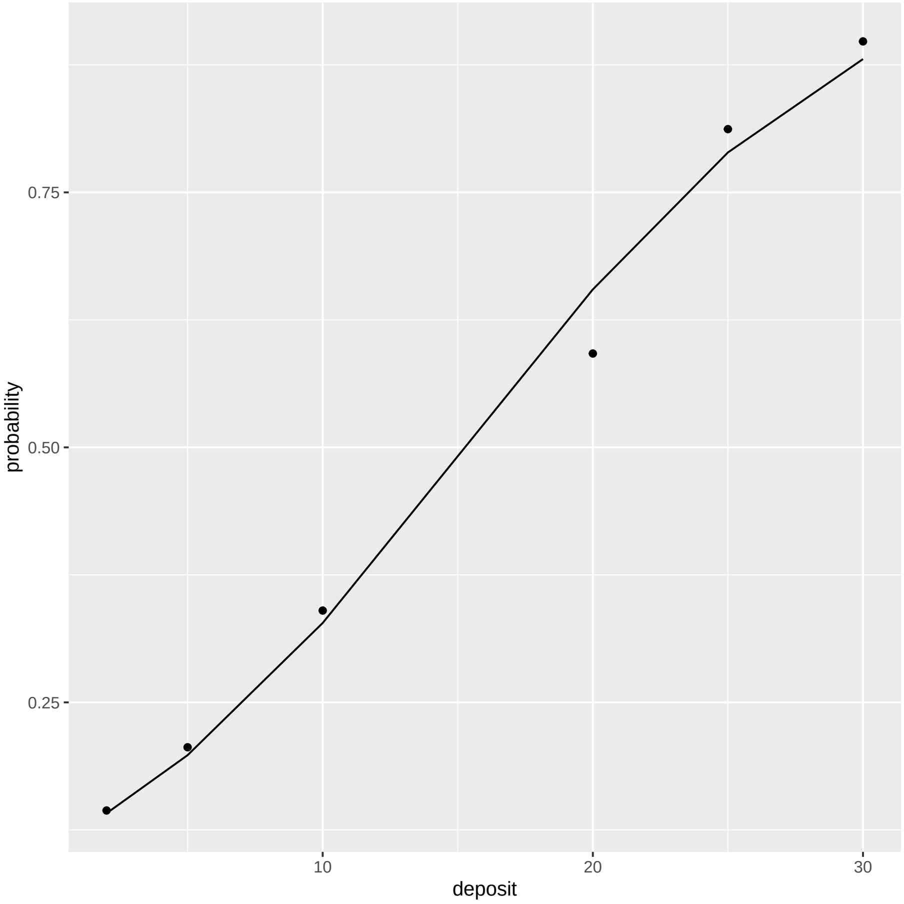


Figure 9: Plot of predictions and proportions

```
arthritis=read_table("arthritis.txt")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   id = col_double(),
##   treatment = col_character(),
##   sex = col_character(),
##   age = col_double(),
##   impr = col_double()
## )
## Warning:  43 parsing failures.
## row col  expected    actual            file
##   1  -- 5 columns 6 columns 'arthritis.txt'
##   2  -- 5 columns 6 columns 'arthritis.txt'
##   3  -- 5 columns 6 columns 'arthritis.txt'
##   4  -- 5 columns 6 columns 'arthritis.txt'
##   5  -- 5 columns 6 columns 'arthritis.txt'
## ... ... ......... ......... ...............
## See problems(...)  for more details.
```

```
arthritis
```

```
## # A tibble: 85 x 5
##       id treatment sex     age  impr
##    <dbl> <chr>     <chr> <dbl> <dbl>
##  1    57 Treated   Male     27     1
##  2    46 Treated   Male     29     0
##  3    77 Treated   Male     30     0
##  4    17 Treated   Male     32     2
##  5    36 Treated   Male     46     2
##  6    23 Treated   Male     58     2
##  7    75 Treated   Male     59     0
##  8    39 Treated   Male     59     2
##  9    33 Treated   Male     63     0
## 10    55 Treated   Male     63     0
## # ... with 75 more rows
```

Figure 10: Arthritis data (some)

```
arthritis.1=polr(factor(impr)~treatment+sex+age, data=arthritis)
drop1(arthritis.1, test="Chisq")

## Single term deletions
##
## Model:
## factor(impr) ~ treatment + sex + age
##           Df    AIC     LRT  Pr(>Chi)
## <none>       155.46
## treatment  1 168.17 14.7095 0.0001254 ***
## sex        1 159.15  5.6880 0.0170812 *
## age        1 158.03  4.5715 0.0325081 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 11: Arthritis model

```
##   treatment    sex age         0          1          2
## 1   Placebo Female  46 0.6849211 0.15738536 0.15769352
## 2   Placebo Female  63 0.5318884 0.20439438 0.26371724
## 3   Placebo   Male  46 0.8837222 0.06545116 0.05082667
## 4   Placebo   Male  63 0.7988957 0.10817798 0.09292635
## 5   Treated Female  46 0.2751180 0.20743935 0.51744266
## 6   Treated Female  63 0.1655413 0.16217047 0.67228824
## 7   Treated   Male  46 0.5702499 0.19503589 0.23471420
## 8   Treated   Male  63 0.4095362 0.22067590 0.36978787
```

Figure 12: Arthritis predictions

```
## Rows:  15 Columns:  5
## -- Column specification -------------------------------------------------------
## Delimiter:  "\t"
## chr (1):  status
## dbl (4):  id, survtime, treatment, age
##
## i Use `spec()` to retrieve the full column specification for
this data.
## i Specify the column types or set `show_col_types = FALSE` to
quiet this message.
```

```
## # A tibble: 15 x 5
##       id survtime status   treatment   age
##    <dbl>    <dbl> <chr>    <chr>     <dbl>
##  1     1        1 Died     B            75
##  2     2        1 Died     B            79
##  3     3        4 Died     B            85
##  4     4        5 Died     B            76
##  5     5        6 Unknown  B            66
##  6     6        8 Died     A            75
##  7     7        9 Survived B            72
##  8     8        9 Died     B            70
##  9     9       12 Died     A            71
## 10    10       15 Unknown  A            73
## 11    11       22 Died     B            66
## 12    12       25 Survived A            73
## 13    13       37 Died     A            68
## 14    14       55 Died     A            59
## 15    15       72 Survived A            61
```

Figure 13: Patient survival data

```
patients.1=coxph(y~treatment+age, data=patients)
summary(patients.1)

## Call:
## coxph(formula = y ~ treatment + age, data = patients)
##
##   n= 15, number of events= 10
##
##               coef exp(coef) se(coef)     z Pr(>|z|)
## treatmentB 1.88484   6.58531  0.96833 1.946  0.05160 .
## age        0.21739   1.24283  0.08429 2.579  0.00991 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##            exp(coef) exp(-coef) lower .95 upper .95
## treatmentB     6.585     0.1519     0.987    43.936
## age            1.243     0.8046     1.054     1.466
##
## Concordance= 0.873  (se = 0.034 )
## Likelihood ratio test= 14.41  on 2 df,   p=7e-04
## Wald test            = 9.03  on 2 df,   p=0.01
## Score (logrank) test = 12.61  on 2 df,   p=0.002
```

Figure 14: Patient survival Cox model

Treatment	Male	Female
A	22	21
	25	19
	26	18
	27	24
	24	25
B	14	21
	17	20
	19	23
	20	27
	17	25
C	15	37
	17	34
	19	36
	14	26
	12	29

Figure 15: Headache pain relief times, original layout of data

```
## Rows:  15 Columns:  3
## -- Column specification --------------------------------------------------------
##
## chr (1):  Treatment
## dbl (2):  Male, Female
##
## i Use ‘spec()‘ to retrieve the full column specification for
this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to
quiet this message.
```

```
## # A tibble: 30 x 3
##    Treatment Gender  Time
##    <chr>     <chr>  <dbl>
##  1 A         Male      22
##  2 A         Female    21
##  3 A         Male      25
##  4 A         Female    19
##  5 A         Male      26
##  6 A         Female    18
##  7 A         Male      27
##  8 A         Female    24
##  9 A         Male      24
## 10 A         Female    25
## # ... with 20 more rows
```

```
painrelief.1=aov(Time~Treatment*Gender, data=painrelief)
summary(painrelief.1)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment         2   71.5   35.73   3.822   0.0362 *
## Gender            1  313.6  313.63  33.544 5.70e-06 ***
## Treatment:Gender  2  521.9  260.93  27.907 5.46e-07 ***
## Residuals        24  224.4    9.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 16: ANOVA with interaction for headache pain relief times

```
painrelief.2=update(painrelief.1, .~.-Treatment:Gender)
summary(painrelief.2)

##             Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment    2   71.5   35.73   1.245 0.30456
## Gender       1  313.6  313.63  10.927 0.00277 **
## Residuals   26  746.3   28.70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(painrelief.2)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Time ~ Treatment + Gender, data = painrelief)
##
## $Treatment
##     diff       lwr      upr     p adj
## B-A -2.8 -8.753648 3.153648 0.4819497
## C-A  0.8 -5.153648 6.753648 0.9405350
## C-B  3.6 -2.353648 9.553648 0.3062075
##
## $Gender
##                  diff       lwr       upr     p adj
## Male-Female -6.466667 -10.48785 -2.445488 0.0027695
```

Figure 17: Analysis of main effects for headache pain relief times

```
painrelief %>% filter(Gender=="Male") -> painrelief_male
painrelief.3a=aov(Time~Treatment, data=painrelief_male)
summary(painrelief.3a)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment    2  245.2  122.60   22.56 8.59e-05 ***
## Residuals   12   65.2    5.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(painrelief.3a)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Time ~ Treatment, data = painrelief_male)
##
## $Treatment
##     diff        lwr       upr     p adj
## B-A -7.4 -11.333026 -3.466974 0.0008076
## C-A -9.4 -13.333026 -5.466974 0.0000968
## C-B -2.0  -5.933026  1.933026 0.3928918
```

Figure 18: Simple effects of treatment for males

```
painrelief %>% filter(Gender=="Female") -> painrelief_female
painrelief.3b=aov(Time~Treatment, data=painrelief_female)
summary(painrelief.3b)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment    2  348.1  174.07   13.12 0.000955 ***
## Residuals   12  159.2   13.27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(painrelief.3b)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Time ~ Treatment, data = painrelief_female)
##
## $Treatment
##     diff       lwr       upr     p adj
## B-A  1.8 -4.345745  7.945745 0.7210392
## C-A 11.0  4.854255 17.145745 0.0012146
## C-B  9.2  3.054255 15.345745 0.0046858
```

Figure 19: Simple effects of treatment for females

```
## Rows:  30 Columns:  3
## -- Column specification --------------------------------------------------------
## Delimiter:  ","
## dbl (3):  Eth, Phil, Other
##
## i Use ‘spec()‘ to retrieve the full column specification for
this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to
quiet this message.

## # A tibble: 20 x 2
##    discipline score
##    <fct>      <dbl>
##  1 Eth            3
##  2 Other          5
##  3 Other          5
##  4 Phil           1
##  5 Other          4
##  6 Other          5
##  7 Phil           5
##  8 Phil           2
##  9 Phil           5
## 10 Eth            4
## 11 Other          5
## 12 Eth            5
## 13 Eth            5
## 14 Phil           5
## 15 Phil           4
## 16 Eth            2
## 17 Other          4
## 18 Eth            3
## 19 Other          5
## 20 Eth            5
```

Figure 20: Morality of eating meat data (some randomly chosen rows)

```
meat %>% group_by(discipline) %>%
    summarize(mean_score=mean(score))

## # A tibble: 3 x 2
##   discipline mean_score
##   <fct>           <dbl>
## 1 Eth              4.17
## 2 Other            4.63
## 3 Phil             3.73
```

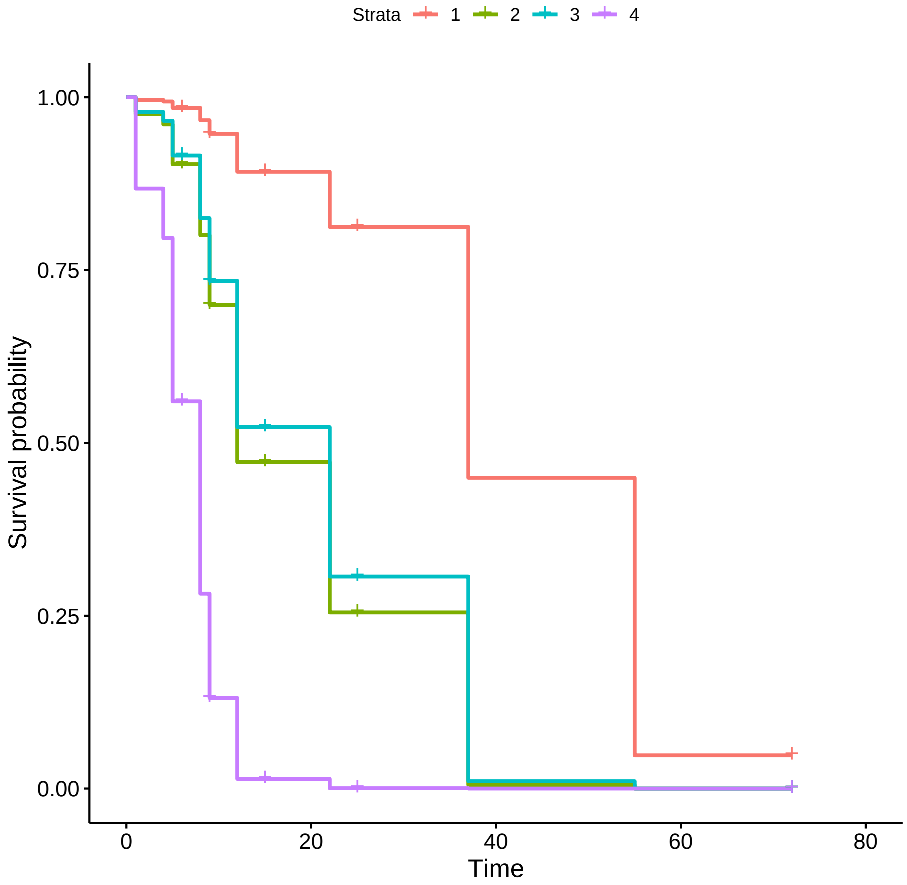
Figure 21: Mean scores by group for meat data

```
meat.1=lm(score~discipline, data=meat)
summary(meat.1)

##
## Call:
## lm(formula = score ~ discipline, data = meat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6333 -0.7333  0.3667  0.8333  1.2667
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.1778     0.1180  35.409  < 2e-16 ***
## disciplinec1    0.2167     0.1445   1.499  0.13739
## disciplinec2    0.4556     0.1669   2.730  0.00766 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.119 on 87 degrees of freedom
## Multiple R-squared:  0.1003,Adjusted R-squared:  0.07965
## F-statistic: 4.851 on 2 and 87 DF,  p-value: 0.01006
```

Figure 22: Testing of contrasts for meat data

```
## Warning:  `gather_()` was deprecated in tidyr 1.2.0.
## Please use `gather()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this
warning was generated.
```



```
## # A tibble: 4 x 2
##     age treatment
##   <dbl> <chr>
## 1    67 A
## 2    67 B
## 3    75 A
## 4    75 B
```

Figure 23: Predicted survival curves for ages and treatments given below plot

```
## `summarise()` has grouped output by 'Treatment'.  You can
override using the
## `.groups` argument.
```

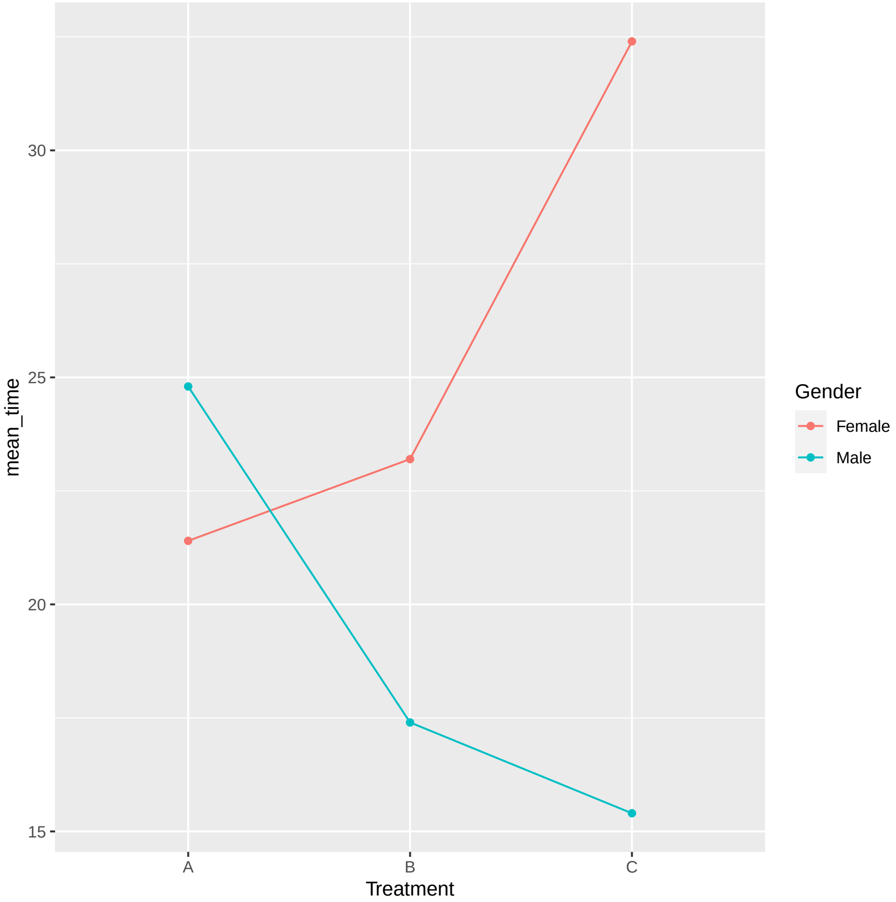


Figure 24: Headache pain relief times, interaction plot