ANOVA Examples in R

AGH Practicals 2

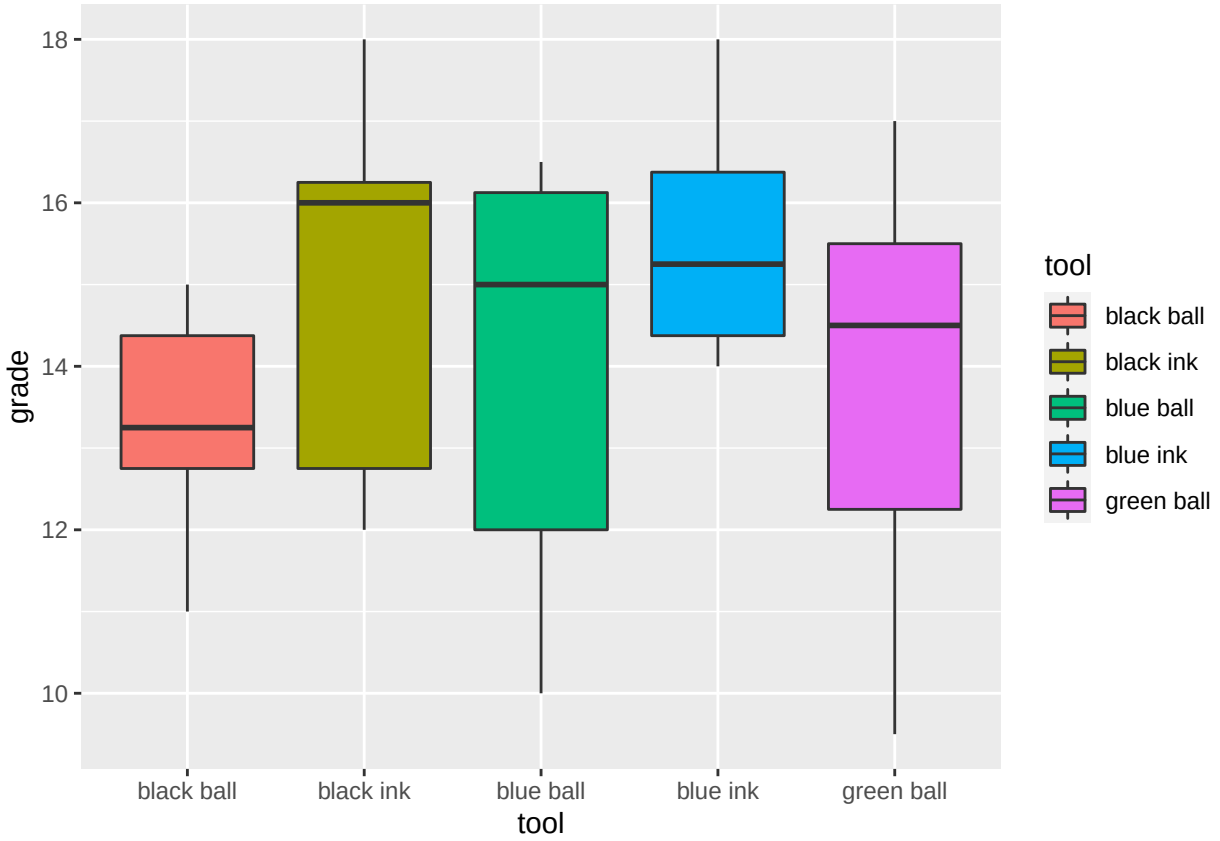
Problem 1:

```
library(ggplot2)
library(grid)
library(gridExtra)
library(dplyr, warn.conflicts = FALSE)
library(xtable)

# Read in the data
grades <- c(16.25, 12.5, 15, 16.5, 11.5, 10, 16, 14, 14.75, 13, 11, 12.5, 13.25,
15, 15, 12.75, 16, 9.5, 11.75, 14.5, 17, 18, 14.25, 17.5, 14, 15.25,
15.25, 14.5, 16, 18, 13.5, 16.5, 12, 16, 12)

# Define factor "pen" corresponding to pen-type for each grade
pen <- factor(c(rep("ball",7*3), rep("ink", 7*2)))

# Define factor "color" corresponding to pen color for each grade
color <- factor(c(rep(c("blue","black","green"), each = 7, times = 2)))
color <- color[1:(7*5)]

# Combine color and pen to get "writing tool"
tool <- paste(color, pen)

# Put data in a long-form (melted) data frame (one observation per row)
grades.df <- data.frame(pen = pen, color = color, tool = tool, grade = grades)
head(grades.df)
```

```
##    pen color      tool grade
## 1 ball  blue blue ball 16.25
## 2 ball  blue blue ball 12.50
## 3 ball  blue blue ball 15.00
## 4 ball  blue blue ball 16.50
## 5 ball  blue blue ball 11.50
## 6 ball  blue blue ball 10.00
```

Before conducting the ANOVA, it might be an idea to visualize the data to see if there is a difference between the groups using boxplots.

```
ggplot(grades.df) + geom_boxplot(aes(x = tool, y = grade, fill = tool))
```



Given these boxplots, it appears that there is a significant difference between grades received using black ball and blue ink. However, we also notice that the variability in the other three groups is extremely large, so this difference may not be enough to conclude that there is a difference. To test this statistically, the ANOVA table is given by

```
# One-way ANOVA for pen-type
anova.tool <- aov(grade ~ tool, grades.df)
print(summary(anova.tool))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## tool         4  21.72   5.431    1.15  0.352
## Residuals   30 141.63   4.721
```

```
# Solve p-value
pf(1.15, 4, 30, lower.tail = FALSE)
```

```
## [1] 0.3523392
```

So the p-value is given by

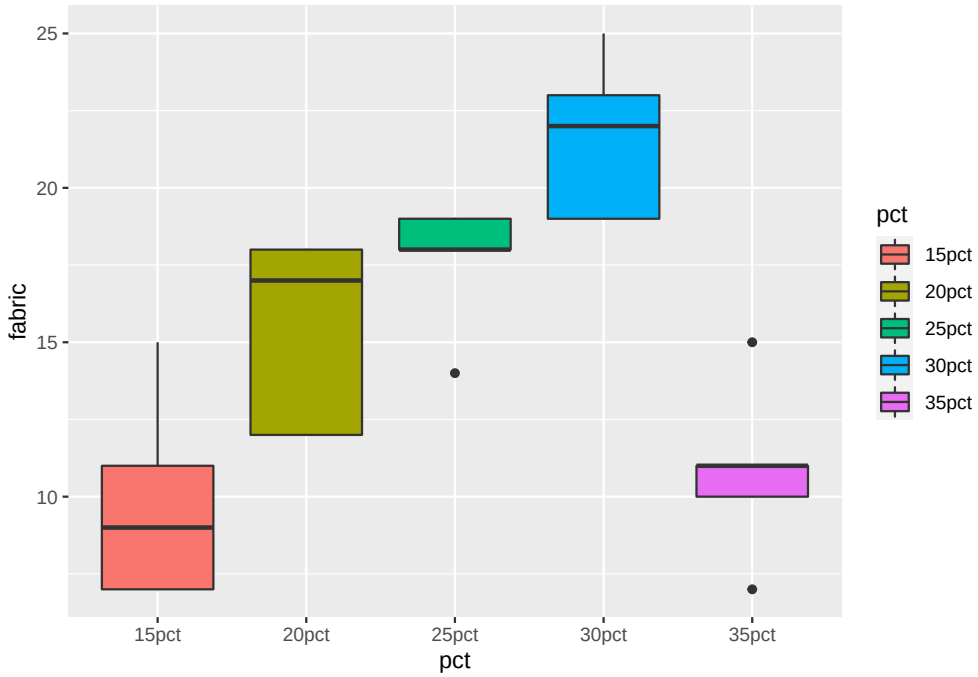
$$P(F_{4,30} > 1.15) = 0.3523392$$

which implies that we do not have enough evidence to reject H_0 at the 5% level. That is, we fail to reject the null hypothesis that there is no difference between writing tool in terms of grade received.

Problem 2:

Before conducting the ANOVA, it might be an idea to visualize the data to see if there is a difference between the groups using boxplots.

```
ggplot(cotton.df) + geom_boxplot(aes(x = pct, y = fabric, fill = pct))
```



Given these boxplots, it appears that there is a significant difference among the treatment means. However, we also notice that the variability in the other three groups is extremely large, so this difference may not be enough to conclude that there is a difference. To test this statistically, the ANOVA table is given by

Analysis of Variance

```
cotton.aov <- aov(fabric ~ pct, cotton.df)
print(summary(cotton.aov))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## pct          4  475.8  118.94   14.76 9.13e-06 ***
## Residuals   20  161.2    8.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This implies that we have enough evidence to reject H_0 at the 5% level. We conclude that there is a significant difference between the tensile strength among the different level of cotton percentages.

Pairwise Comparisons

```
library(DescTools)

PostHocTest(cotton.aov, method = "scheffe")
```

```
##
##   Posthoc multiple comparisons of means: Scheffe Test
##     95% family-wise confidence level
##
## $pct
##              diff      lwr.ci     upr.ci    pval
## 20pct-15pct   5.6  -0.4795547 11.6795547 0.08118 .
## 25pct-15pct   7.8   1.7204453 13.8795547 0.00761 **
## 30pct-15pct  11.8   5.7204453 17.8795547 7.9e-05 ***
## 35pct-15pct   1.0  -5.0795547  7.0795547 0.98830
## 25pct-20pct   2.2  -3.8795547  8.2795547 0.82349
## 30pct-20pct   6.2   0.1204453 12.2795547 0.04411 *
## 35pct-20pct  -4.6 -10.6795547  1.4795547 0.20320
## 30pct-25pct   4.0  -2.0795547 10.0795547 0.32566
## 35pct-25pct  -6.8 -12.8795547 -0.7204453 0.02323 *
## 35pct-30pct -10.8 -16.8795547 -4.7204453 0.00024 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
PostHocTest(cotton.aov, method = "bonferroni")
```

```
##
##   Posthoc multiple comparisons of means : Bonferroni
##     95% family-wise confidence level
##
## $pct
##              diff      lwr.ci    upr.ci    pval
## 20pct-15pct   5.6  -0.0620885 11.262089  0.0541 .
## 25pct-15pct   7.8   2.1379115 13.462089  0.0031 **
## 30pct-15pct  11.8   6.1379115 17.462089 2.1e-05 ***
## 35pct-15pct   1.0  -4.6620885  6.662089  1.0000
## 25pct-20pct   2.2  -3.4620885  7.862089  1.0000
## 30pct-20pct   6.2   0.5379115 11.862089  0.0251 *
## 35pct-20pct  -4.6 -10.2620885  1.062089  0.1859
## 30pct-25pct   4.0  -1.6620885  9.662089  0.3754
## 35pct-25pct  -6.8 -12.4620885 -1.137911  0.0116 *
## 35pct-30pct -10.8 -16.4620885 -5.137911 7.0e-05 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
PostHocTest(cotton.aov, method = "lsd")
```

```
##
##   Posthoc multiple comparisons of means : Fisher LSD
##     95% family-wise confidence level
##
## $pct
```

```
##              diff      lwr.ci     upr.ci    pval
## 20pct-15pct   5.6   1.8545482  9.3454518 0.00541 **
## 25pct-15pct   7.8   4.0545482 11.5454518 0.00031 ***
## 30pct-15pct  11.8   8.0545482 15.5454518 2.1e-06 ***
## 35pct-15pct   1.0  -2.7454518  4.7454518 0.58375
## 25pct-20pct   2.2  -1.5454518  5.9454518 0.23471
## 30pct-20pct   6.2   2.4545482  9.9454518 0.00251 **
## 35pct-20pct  -4.6  -8.3454518 -0.8545482 0.01859 *
## 30pct-25pct   4.0   0.2545482  7.7454518 0.03754 *
## 35pct-25pct  -6.8 -10.5454518 -3.0545482 0.00116 **
## 35pct-30pct -10.8 -14.5454518 -7.0545482 7.0e-06 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
PostHocTest(cotton.aov, method = "hsd")
```

```
##
##   Posthoc multiple comparisons of means : Tukey HSD
##     95% family-wise confidence level
##
## $pct
##              diff      lwr.ci     upr.ci    pval
## 20pct-15pct   5.6   0.2270417 10.9729583  0.0385 *
## 25pct-15pct   7.8   2.4270417 13.1729583  0.0026 **
## 30pct-15pct  11.8   6.4270417 17.1729583 1.9e-05 ***
## 35pct-15pct   1.0  -4.3729583  6.3729583  0.9798
## 25pct-20pct   2.2  -3.1729583  7.5729583  0.7372
## 30pct-20pct   6.2   0.8270417 11.5729583  0.0189 *
## 35pct-20pct  -4.6  -9.9729583  0.7729583  0.1163
## 30pct-25pct   4.0  -1.3729583  9.3729583  0.2101
## 35pct-25pct  -6.8 -12.1729583 -1.4270417  0.0091 **
## 35pct-30pct -10.8 -16.1729583 -5.4270417 6.2e-05 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
PostHocTest(cotton.aov, method = "duncan")
```

```
##
##   Posthoc multiple comparisons of means : Duncan's new multiple range test
##     95% family-wise confidence level
##
## $pct
##              diff      lwr.ci     upr.ci    pval
## 20pct-15pct   5.6   1.6685342  9.5314658 0.00718 **
## 25pct-15pct   7.8   3.7503181 11.8496819 0.00055 ***
## 30pct-15pct  11.8   7.6677507 15.9322493 4.8e-06 ***
## 35pct-15pct   1.0  -2.7454516  4.7454516 0.58375
## 25pct-20pct   2.2  -1.5454516  5.9454516 0.23471
## 30pct-20pct   6.2   2.2685342 10.1314658 0.00340 **
## 35pct-20pct  -4.6  -8.3454516 -0.8545484 0.01859 *
## 30pct-25pct   4.0   0.2545484  7.7454516 0.03754 *
## 35pct-25pct  -6.8 -10.7314658 -2.8685342 0.00159 **
## 35pct-30pct -10.8 -14.8496819 -6.7503181 1.3e-05 ***
```

```
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
PostHocTest(cotton.aov, method = "newmankeuls")
```

```
##
##   Posthoc multiple comparisons of means : Newman-Keuls
##     95% family-wise confidence level
##
## $pct
##               diff      lwr.ci     upr.ci    pval
## 20pct-15pct   5.6   1.0572910 10.1427090  0.0143 *
## 25pct-15pct   7.8   2.7743700 12.8256300  0.0017 **
## 30pct-15pct  11.8   6.4270417 17.1729583 1.9e-05 ***
## 35pct-15pct   1.0  -2.7454516  4.7454516  0.5838
## 25pct-20pct   2.2  -1.5454516  5.9454516  0.2347
## 30pct-20pct   6.2   1.6572910 10.7427090  0.0068 **
## 35pct-20pct  -4.6  -8.3454516 -0.8545484  0.0186 *
## 30pct-25pct   4.0   0.2545484  7.7454516  0.0375 *
## 35pct-25pct  -6.8 -11.3427090 -2.2572910  0.0032 **
## 35pct-30pct -10.8 -15.8256300 -5.7743700 3.9e-05 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
DunnettTest(fabric, pct, control = "15pct")
```

```
##
##   Dunnett's test for comparing several treatments with a control :
##     95% family-wise confidence level
##
## $`15pct`
##              diff     lwr.ci    upr.ci    pval
## 20pct-15pct   5.6  0.8373325 10.362668  0.0186 *
## 25pct-15pct   7.8  3.0373325 12.562668  0.0012 **
## 30pct-15pct  11.8  7.0373325 16.562668 2.8e-05 ***
## 35pct-15pct   1.0 -3.7626675  5.762668  0.9469
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 3:

```
library(ggplot2)
library(grid)
library(gridExtra)
library(dplyr, warn.conflicts = FALSE)
library(DescTools)

# Read in the data
trt1 <- c(35,37,49,46,63,39,46,56,63,65,56,65,70,63,65,70,77,81,86,70,70,77,77,81,77)
trt2 <- c(40,37,44,47,47,47,68,47,54,61,71,75,89,58,59,62,79,96,58,62,70,72,75,96,75)
```

```
trt3 <- c(46,42,65,46,58,42,48,58,50,80,63,65,70,70,72,97,46,56,70,70,72,76,90,76,92)
trt4 <- c(21,40,44,54,36,40,56,60,48,53,60,60,65,68,60,81,81,48,48,56,68,75,81,48,68)
trt5 <- c(16,19,19,32,33,33,30,42,42,33,26,30,40,54,34,34,47,47,42,47,54,54,56,60,44)

trt <- c(trt1, trt2, trt3, trt4, trt5)
desc <- rbind(summary(trt1), summary(trt2), summary(trt3),summary(trt4),summary(trt5))
desc2 <- cbind(desc, c(sd(trt1), sd(trt2), sd(trt3), sd(trt4), sd(trt5)))
rownames(desc2) <- c("None", "1 pregnant", "1 virgin", "8 pregnant", "8 virgin")
colnames(desc2) <- c("Min", "Q1","Median", "Mean", "Q3", "Max", "Std Dev")
print(desc2)
```

```
##            Min Q1 Median  Mean Q3 Max  Std Dev
## None        35 56     65 63.36 77  86 14.53983
## 1 pregnant  37 47     62 63.56 75  96 16.45215
## 1 virgin    42 50     65 64.80 72  97 15.65248
## 8 pregnant  21 48     56 56.76 68  81 14.92838
## 8 virgin    16 32     40 38.72 47  60 12.10207
```

```
# Define factor
group <- factor(c(rep("None",25), rep("1 pregnant", 25),
                  rep("1 virgin",25), rep("8 pregnant", 25),
                  rep("8 virgin",25)))

# Put data in a long-form (melted) data frame (one observation per row)
fly.df <- data.frame(trt = trt, group = group)
fly.aov <- aov(trt ~ group, fly.df)
print(summary(fly.aov))
```

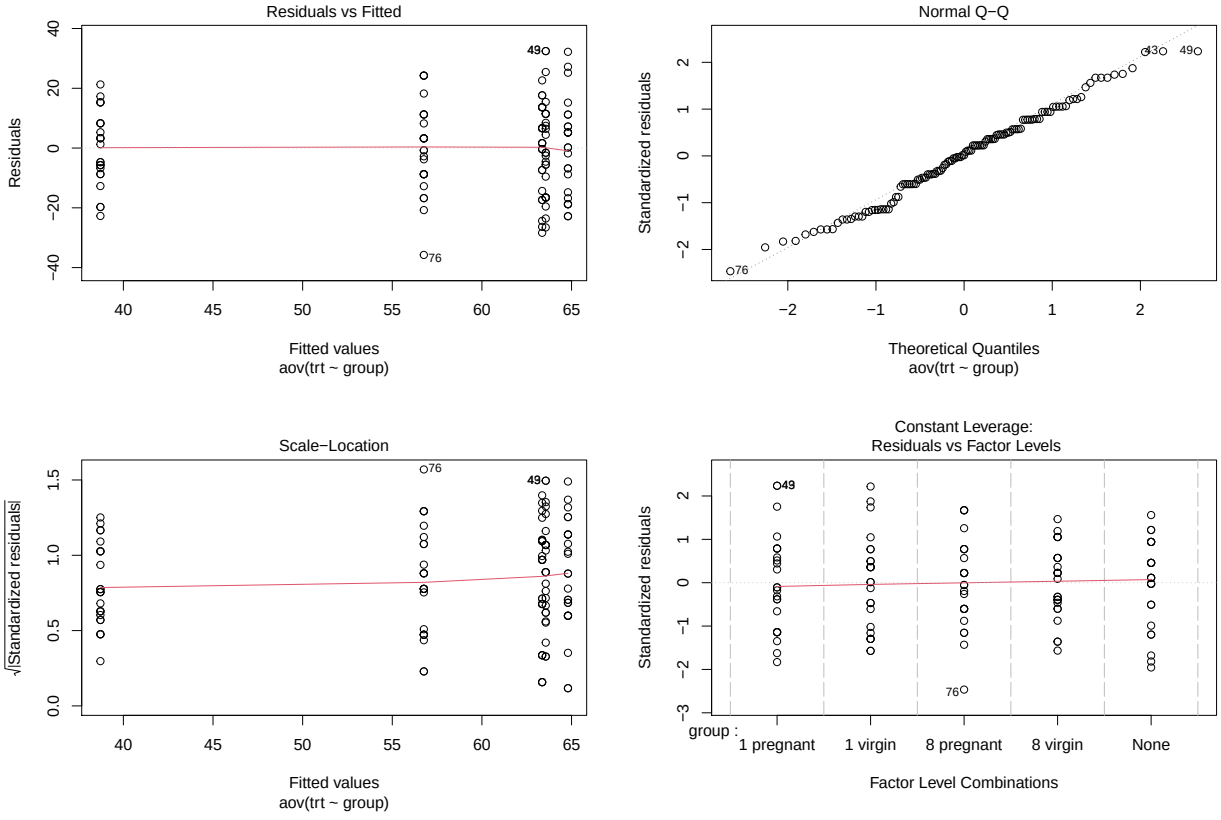
```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## group         4  11939  2984.8   13.61 3.52e-09 ***
## Residuals   120  26314   219.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##   Posthoc multiple comparisons of means: Scheffe Test
##     95% family-wise confidence level
##
## $group
##                          diff     lwr.ci     upr.ci    pval
## 1 virgin-1 pregnant     1.24 -11.864239  14.344239  0.9991
## 8 pregnant-1 pregnant  -6.80 -19.904239   6.304239  0.6217
## 8 virgin-1 pregnant   -24.84 -37.944239 -11.735761 2.9e-06 ***
## None-1 pregnant        -0.20 -13.304239  12.904239  1.0000
## 8 pregnant-1 virgin    -8.04 -21.144239   5.064239  0.4540
## 8 virgin-1 virgin     -26.08 -39.184239 -12.975761 7.9e-07 ***
## None-1 virgin          -1.44 -14.544239  11.664239  0.9983
## 8 virgin-8 pregnant   -18.04 -31.144239  -4.935761  0.0016 **
## None-8 pregnant         6.60  -6.504239  19.704239  0.6486
## None-8 virgin          24.64  11.535761  37.744239 3.6e-06 ***
```

```
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##   Posthoc multiple comparisons of means : Bonferroni
##     95% family-wise confidence level
##
## $group
##                          diff     lwr.ci     upr.ci    pval
## 1 virgin-1 pregnant     1.24 -10.738139  13.218139 1.00000
## 8 pregnant-1 pregnant  -6.80 -18.778139   5.178139 1.00000
## 8 virgin-1 pregnant   -24.84 -36.818139 -12.861861 3.0e-07 ***
## None-1 pregnant        -0.20 -12.178139  11.778139 1.00000
## 8 pregnant-1 virgin    -8.04 -20.018139   3.938139 0.57282
## 8 virgin-1 virgin     -26.08 -38.058139 -14.101861 7.3e-08 ***
## None-1 virgin          -1.44 -13.418139  10.538139 1.00000
## 8 virgin-8 pregnant   -18.04 -30.018139  -6.061861 0.00034 ***
## None-8 pregnant         6.60  -5.378139  18.578139 1.00000
## None-8 virgin          24.64  12.661861  36.618139 3.7e-07 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##   Posthoc multiple comparisons of means : Fisher LSD
##     95% family-wise confidence level
##
## $group
##                          diff     lwr.ci      upr.ci    pval
## 1 virgin-1 pregnant     1.24  -7.052658   9.5326583  0.7677
## 8 pregnant-1 pregnant  -6.80 -15.092658   1.4926583  0.1071
## 8 virgin-1 pregnant   -24.84 -33.132658 -16.5473417 3.0e-08 ***
## None-1 pregnant        -0.20  -8.492658   8.0926583  0.9620
## 8 pregnant-1 virgin    -8.04 -16.332658   0.2526583  0.0573 .
## 8 virgin-1 virgin     -26.08 -34.372658 -17.7873417 7.3e-09 ***
## None-1 virgin          -1.44  -9.732658   6.8526583  0.7316
## 8 virgin-8 pregnant   -18.04 -26.332658  -9.7473417 3.4e-05 ***
## None-8 pregnant         6.60  -1.692658  14.8926583  0.1177
## None-8 virgin          24.64  16.347342  32.9326583 3.7e-08 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##   Posthoc multiple comparisons of means : Tukey HSD
##     95% family-wise confidence level
##
## $group
##                          diff     lwr.ci     upr.ci    pval
## 1 virgin-1 pregnant     1.24 -10.360468  12.840468 0.99830
## 8 pregnant-1 pregnant  -6.80 -18.400468   4.800468 0.48542
## 8 virgin-1 pregnant   -24.84 -36.440468 -13.239532 3.0e-07 ***
## None-1 pregnant        -0.20 -11.800468  11.400468 1.00000
## 8 pregnant-1 virgin    -8.04 -19.640468   3.560468 0.31265
```

```
## 8 virgin-1 virgin     -26.08 -37.680468 -14.479532 7.2e-08 ***
## None-1 virgin          -1.44 -13.040468  10.160468 0.99696
## 8 virgin-8 pregnant   -18.04 -29.640468  -6.439532 0.00032 ***
## None-8 pregnant         6.60  -5.000468  18.200468 0.51577
## None-8 virgin          24.64  13.039532  36.240468 3.7e-07 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##   Dunnett's test for comparing several treatments with a control :
##     95% family-wise confidence level
##
## $None
##                    diff     lwr.ci     upr.ci    pval
## 1 pregnant-None    0.20 -10.163843  10.563843  1.0000
## 1 virgin-None      1.44  -8.923843  11.803843  0.9909
## 8 pregnant-None   -6.60 -16.963843   3.763843  0.3291
## 8 virgin-None    -24.64 -35.003843 -14.276157 8.8e-08 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```




```
##
##  Bartlett test of homogeneity of variances
##
## data:  trt by group
```

```
## Bartlett's K-squared = 2.4196, df = 4, p-value = 0.6591
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value Pr(>F)
## group   4  0.5405 0.7062
##       120
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   4  0.4916 0.7419
##       120
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```