STAT 511A Homework 6

Kathleen Wendt

10/18/2019

Load packages

```
library(tidyverse)
library(broom)
library(car)
library(emmeans)
```

Question 1

Rat data

In an investigation of the possible influence of dietary chromium on diabetic symptoms, 14 rats were fed a low-chromium diet and 10 were fed a control diet. One response variable was activity of the liver enzyme GITH. The data is available as "RatLiver.csv".

```
rat_data <- readr::read_csv("RatLiver.csv") %>%
    mutate(Trt = as.factor(Trt))
```

```
## Parsed with column specification:
## cols(
##   Trt = col_character(),
##   Enzyme = col_double()
## )
```

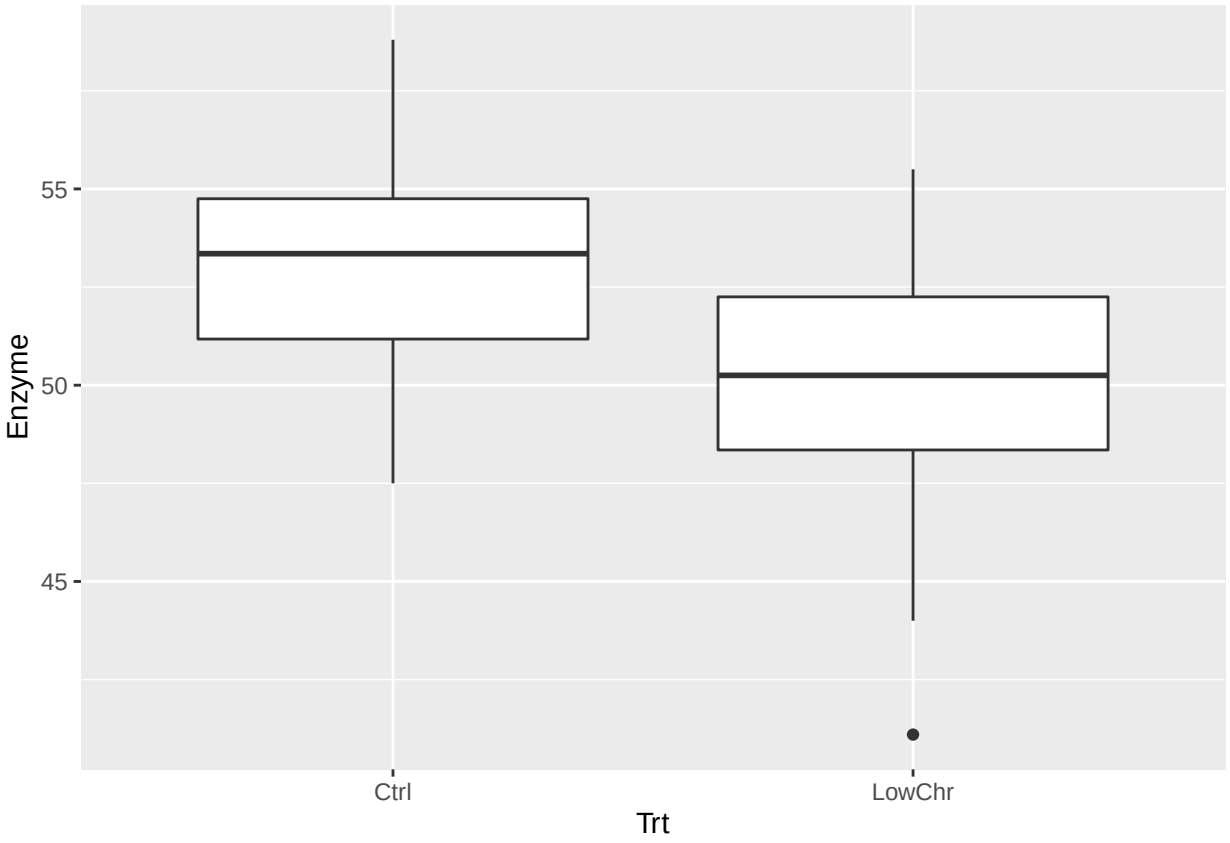
```
tibble::glimpse(rat_data)
```

```
## Observations: 24
## Variables: 2
## $ Trt    <fct> LowChr, LowChr, LowChr, LowChr, LowChr, LowChr, LowChr,...
## $ Enzyme <dbl> 44.0, 48.5, 50.7, 45.0, 53.0, 52.7, 51.8, 49.8, 48.3, 5...
```

Part 1A

Construct side-by-side boxplots of the data.

```
rat_data %>%
  ggplot(aes(y = Enzyme, x = Trt)) +
  geom_boxplot()
```



Part 1B

Use the F-test to test for equality of variances. Give the null hypothesis, test statistic, p-value and conclusion. (4 pts)

Null hypothesis

Equal variances; true ratio of variances is equal to 1.

F-test for equality of variances

```
rat_ftest <- tidy(var.test(Enzyme ~ Trt, data = rat_data))
```

```
## Multiple parameters; naming those columns num.df, denom.df
```

```
rat_ftest
```

```
## # A tibble: 1 x 9
##   estimate `num df` `denom df` statistic p.value conf.low conf.high method
##      <dbl>    <int>      <int>     <dbl>   <dbl>    <dbl>     <dbl> <chr>
## 1    0.790        9         13     0.790   0.737    0.238      3.03 F tes~
## # ... with 1 more variable: alternative <chr>
```

Test statistic

The test statistic (F) for the F-test for equality of variances is 0.7897775.

P-value

The p-value for the F-test for equality of variances is 0.7373033.

Conclusion

Fail to reject null hypothesis. Because the above p-value is above $\alpha = 0.05$, there is no evidence to suggest that the true ratio of variances is not equal to 1.

Part 1C

Use Levene's test (with center="median") to test for equality of variances. Give the p-value and conclusion.

Levene's Test

```
rat_levene <- tidy(leveneTest(Enzyme ~ Trt, data = rat_data,
                center = "median"))
rat_levene
```

```
## # A tibble: 2 x 4
##   term     df statistic p.value
##   <chr> <int>     <dbl>   <dbl>
## 1 group     1     0.176   0.679
## 2 ""       22    NA      NA
```

P-value

The p-value for the Levene Test for equality of variances with median as center is 0.6788657.

Conclusion

Fail to reject null hypothesis. Because the above p-value is above $\alpha = 0.05$, there is no evidence to suggest that the true ratio of variances is not equal to 1.

Part 1D

Based on your conclusions from the two previous questions, would the pooled variance t-test or Welch-Satterthwaite t-test be preferred?

Pooled t-test

Part 1E

Regardless of your answer to the previous question, run a two-sample t-test assuming equal variances. Give the null hypothesis, test statistic, p-value and conclusion. (4 pts)

Null hypothesis

$\mu 1 = \mu 2$. No difference between sample means.

Pooled two-sample t-test

```
rat_ttest <- tidy(t.test(Enzyme ~ Trt,
                         data = rat_data,
                         var.equal = TRUE))
rat_ttest
```

```
## # A tibble: 1 x 9
##   estimate1 estimate2 statistic p.value parameter conf.low conf.high method
##       <dbl>     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
## 1      52.9      49.5      2.17  0.0410        22    0.151      6.59 " Two~
## # ... with 1 more variable: alternative <chr>
```

Test statistic

The test statistic (t) for the pooled two-sample t-test of means is 2.1708862.

P-value

The p-value for the pooled two-sample t-test of means is 0.041005.

Conclusion

Reject null hypothesis: $0.041005 < \alpha = 0.05$ level. There is evidence to suggest that there is a difference between sample means, such that the mean level of enzyme in Control rats is higher than that of the Treatment group.

Part 1F

Rerun the analysis as a one-way ANOVA. Give the ANOVA table in your assignment. Compare your results to the previous question and notice that the p-value is the same and F = t2.

```r
rat_lm <- lm(Enzyme ~ Trt, data = rat_data)
anova(rat_lm)
```

```
## Analysis of Variance Table
##
## Response: Enzyme
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Trt        1  66.249  66.249  4.7127  0.041 *
## Residuals 22 309.261  14.057
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 2

Read Problem 8.32 which concerns corn yield. The data is available as "CornYield.csv".

Corn data

```r
corn_data <- readr::read_csv("CornYield.csv") %>%
  mutate(Variety = as.factor(Variety)) %>%
  group_by(Variety) %>%
  mutate(Mean = mean(Yield),
         SE = sd(Yield)/sqrt(length(Yield))) %>%
  ungroup()
```

```
## Parsed with column specification:
## cols(
##   Variety = col_character(),
##   Yield = col_double()
## )
```

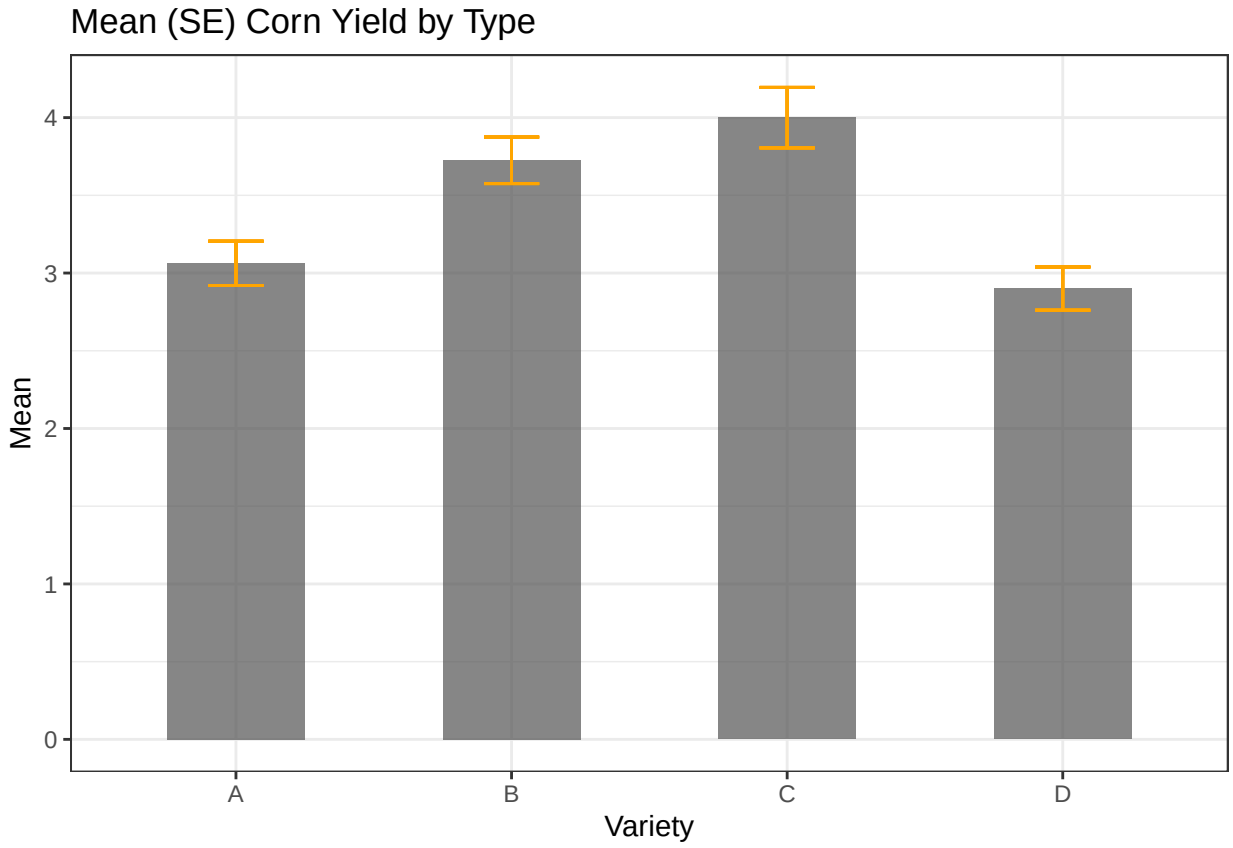
```r
tibble::glimpse(corn_data)
```

```
## Observations: 32
## Variables: 4
## $ Variety <fct> A, A, A, A, A, A, A, A, B, B, B, B, B, B, B, B, C, C, ...
## $ Yield   <dbl> 2.5, 3.6, 2.8, 2.7, 3.1, 3.4, 2.9, 3.5, 3.6, 3.9, 4.1,...
## $ Mean    <dbl> 3.0625, 3.0625, 3.0625, 3.0625, 3.0625, 3.0625, 3.0625...
## $ SE      <dbl> 0.1426002, 0.1426002, 0.1426002, 0.1426002, 0.1426002,...
```

Part 2A

Construct a bar plot showing means and SEs for each variety. (4 pts)

```
corn_data %>%
  dplyr::select(-Yield) %>%
  ggplot(aes(x = Variety, y = Mean)) +
  geom_col(alpha = 0.15, position = "dodge", width = 0.5) +
  geom_errorbar(aes(ymin = Mean - SE, ymax = Mean + SE),
                width = 0.2,
                color = "orange",
                position = position_dodge()) +
  ggtitle("Mean (SE) Corn Yield by Type") +
  theme_bw()
```



Part 2B

Carry out a one-way ANOVA analysis to determine whether there is evidence of differences (using alpha 0.05) in the mean yield for the different varieties. State the null hypothesis, give the test statistic, p-value and conclusion. (4 pts)

Null hypothesis

$\mu 1 = \mu 2 = \mu 3 = \mu 4$. No difference between sample means.

One-way ANOVA

```
corn_lm <- lm(Yield ~ Variety, data = corn_data)
corn_lm
```

```
##
## Call:
## lm(formula = Yield ~ Variety, data = corn_data)
##
## Coefficients:
## (Intercept)     VarietyB     VarietyC     VarietyD
##      3.0625       0.6625       0.9375      -0.1625
```

```
summary(corn_lm)
```

```
##
## Call:
## lm(formula = Yield ~ Variety, data = corn_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82500 -0.28750  0.01875  0.34688  0.70000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0625     0.1580  19.380  < 2e-16 ***
## VarietyB      0.6625     0.2235   2.964 0.006133 **
## VarietyC      0.9375     0.2235   4.195 0.000249 ***
## VarietyD     -0.1625     0.2235  -0.727 0.473184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.447 on 28 degrees of freedom
## Multiple R-squared:  0.542,  Adjusted R-squared:  0.493
## F-statistic: 11.05 on 3 and 28 DF,  p-value: 5.85e-05
```

```
anova(corn_lm)
```

```
## Analysis of Variance Table
##
## Response: Yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Variety    3 6.6209 2.20698  11.047 5.85e-05 ***
## Residuals 28 5.5938 0.19978
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
corn_anova_tidy <- tidy(anova(corn_lm))
corn_anova_tidy
```

```
## # A tibble: 2 x 6
##   term         df sumsq meansq statistic    p.value
##   <chr>     <int> <dbl>  <dbl>     <dbl>      <dbl>
## 1 Variety       3  6.62  2.21       11.0  0.0000585
## 2 Residuals    28  5.59  0.200      NA   NA
```

Test statistic

The test statistic (F) for the one-way ANOVA on corn yield by variety is 11.0472253.

P-value

The p-value for the one-way ANOVA on corn yield by variety is 0.0000585 `corn_anova_tidy$p.value[1]`.

Conclusion

Reject null hypothesis at $\alpha = 0.05$. There is evidence to suggest a difference in mean yields across the four varieties of corn.

Part 2C

Run (unadjusted) pairwise comparisons of means. Give the estimated difference and p-value for each comparison. (4 pts)

Pairwise comparisons (`emmeans`)

```
corn_em <- emmeans(corn_lm, pairwise ~ Variety, adjust = "none")
corn_em
```

```
## $emmeans
##  Variety emmean    SE df lower.CL upper.CL
##  A         3.06 0.158 28     2.74     3.39
##  B         3.73 0.158 28     3.40     4.05
##  C         4.00 0.158 28     3.68     4.32
##  D         2.90 0.158 28     2.58     3.22
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate    SE df t.ratio p.value
##  A - B      -0.662 0.223 28 -2.964  0.0061
##  A - C      -0.938 0.223 28 -4.195  0.0002
##  A - D       0.163 0.223 28  0.727  0.4732
##  B - C      -0.275 0.223 28 -1.231  0.2287
##  B - D       0.825 0.223 28  3.692  0.0010
##  C - D       1.100 0.223 28  4.922  <.0001
```

Tidy pairwise comparisons

```
corn_em_info <- emmeans(corn_lm, ~ Variety, adjust = "none")
corn_em_tidy <- tidy(corn_em_info)
corn_em_tidy
```

```
## # A tibble: 4 x 6
##   Variety estimate std.error    df conf.low conf.high
##   <fct>      <dbl>     <dbl> <dbl>    <dbl>     <dbl>
```

```
## 1 A           3.06     0.158    28     2.74      3.39
## 2 B           3.72     0.158    28     3.40      4.05
## 3 C           4.00     0.158    28     3.68      4.32
## 4 D           2.90     0.158    28     2.58      3.22
```

```
corn_em_tidy_pairs <- tidy(pairs(corn_em_info), adjust = "none")
corn_em_tidy_pairs
```

```
## # A tibble: 6 x 7
##   level1 level2 estimate std.error    df statistic   p.value
##   <chr>  <chr>     <dbl>     <dbl> <dbl>     <dbl>     <dbl>
## 1 A      B        -0.662     0.223    28    -2.96  0.00613
## 2 A      C        -0.938     0.223    28    -4.19  0.000249
## 3 A      D         0.162     0.223    28     0.727 0.473
## 4 B      C        -0.275     0.223    28    -1.23  0.229
## 5 B      D         0.825     0.223    28     3.69  0.000955
## 6 C      D         1.10      0.223    28     4.92  0.0000343
```

Part 2D

Calculate the LSD(0.05) value. Recall that this is the 95% ME for pairwise comparisons of means.

ME = LSD = t(alpha/2)*sw(sqrt(s/n))

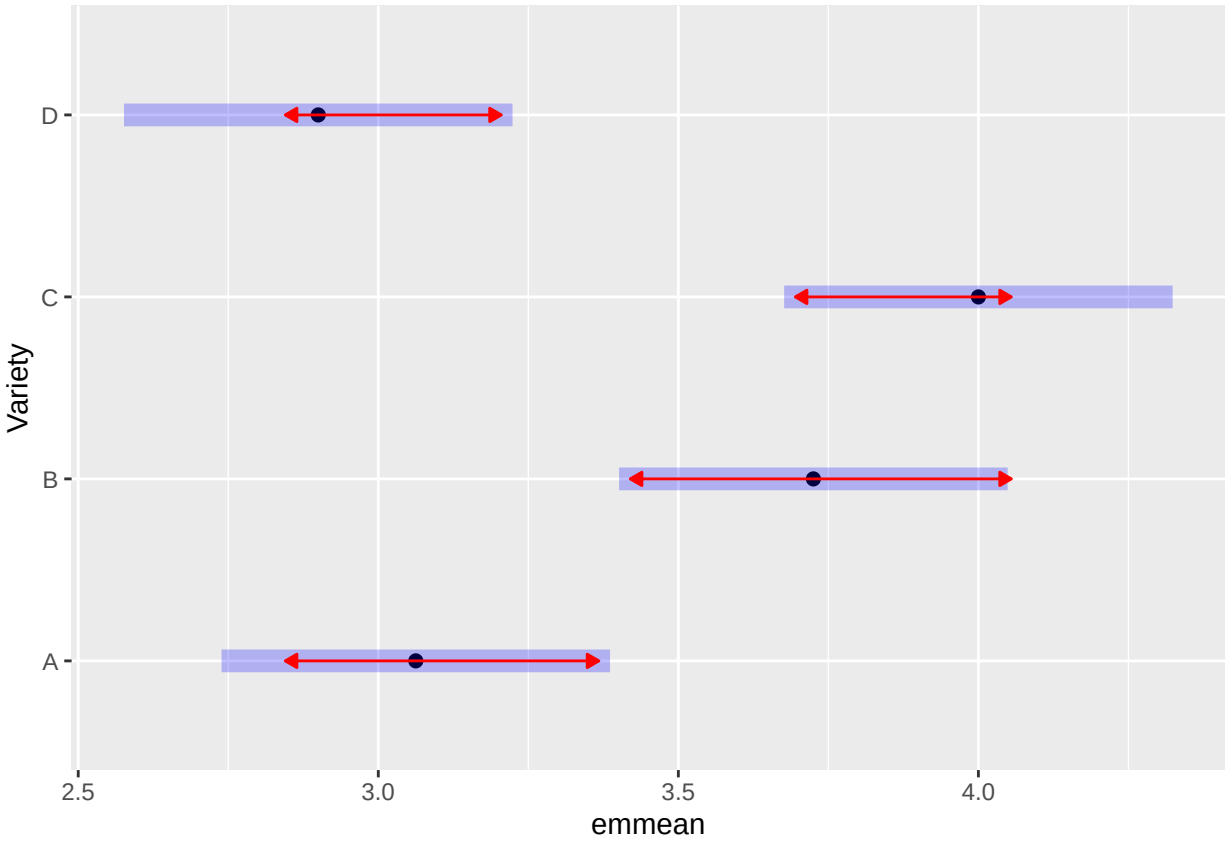
Part 2E

Construct an (unadjusted) "cld" display including the mean for each variety and assigning number groups (or underlining) varieties that are not "significantly" different. (4 pts)

```
CLD(corn_em$emmeans, adjust = "none")
```

```
##  Variety emmean    SE df lower.CL upper.CL .group
##  D         2.90 0.158 28     2.58     3.22  1
##  A         3.06 0.158 28     2.74     3.39  1
##  B         3.73 0.158 28     3.40     4.05   2
##  C         4.00 0.158 28     3.68     4.32   2
##
## Confidence level used: 0.95
## significance level used: alpha = 0.05
```

```
plot(corn_em$emmeans, comparisons = TRUE)
```



Part 2F

Summarize your findings from parts C and E.

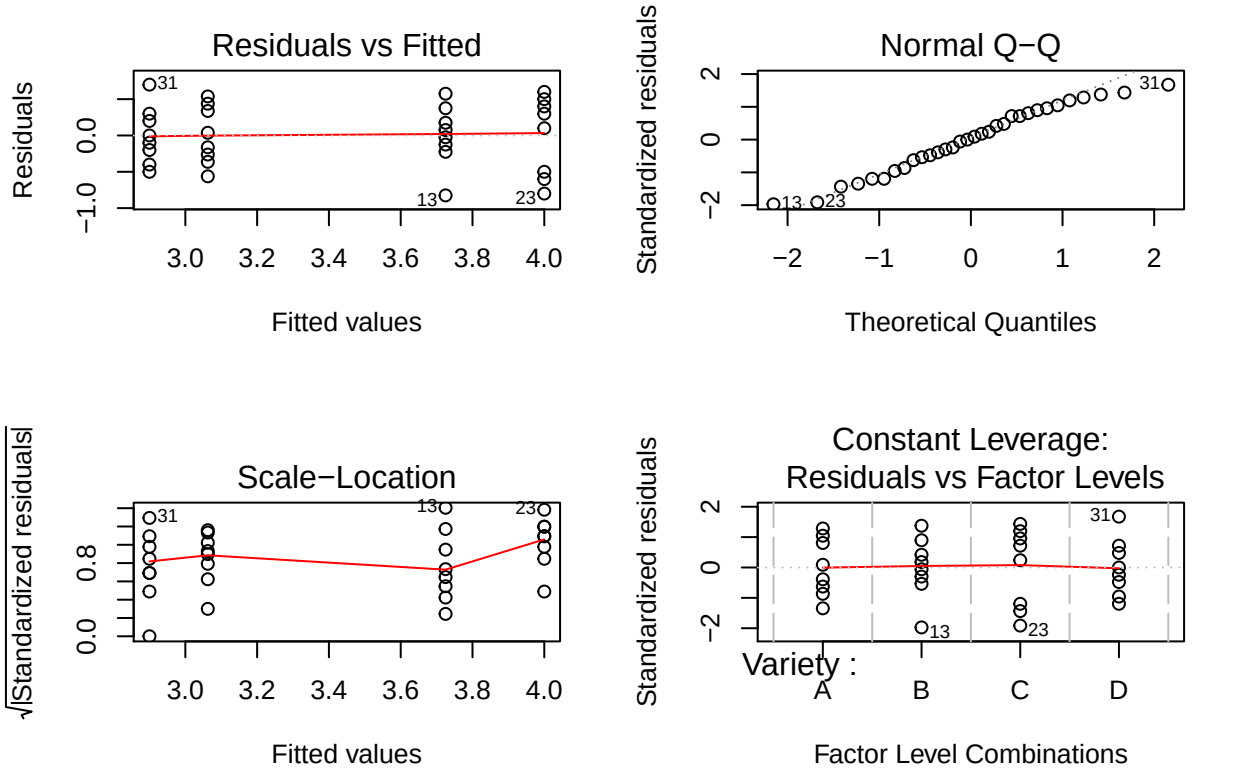
Corn D and Corn A are in Group 1. Corn B and Corn C are are in Group 2. There are between-group differences in means (Group 1 vs. Group 2) but not within-group differences in means (Corn D/Corn A; Corn B/Corn C).

Part 2G

Use the plot() function to generate the diagnostic plots from the model from part B. You do not have to include the graphs in your assignment, but discuss the plots of (1) Residuals vs Fitted values and (2) qqplot of residuals and whether assumptions appear to be satisfied based on each plot. (4 pts)

Check ANOVA assumptions visually

```
par(mfrow=c(2,2))
plot(corn_lm)
```



(1) Residuals vs Fitted values: This plot checks for violations of constant variance and linearity. Because there is neither a mega-phone or bow pattern of the residuals, there is no evidence for the violation of the assumption of constant variance or linearity, respectively.

(2) Q-Q plot of residuals: This plot checks for normality of residuals. There are slight deviation from the Q-Q line on the upper-end of the plot. This indicates there might be evidence of heavy tail, and, therefore, a potential violation of the assumption of normally distributed residuals.