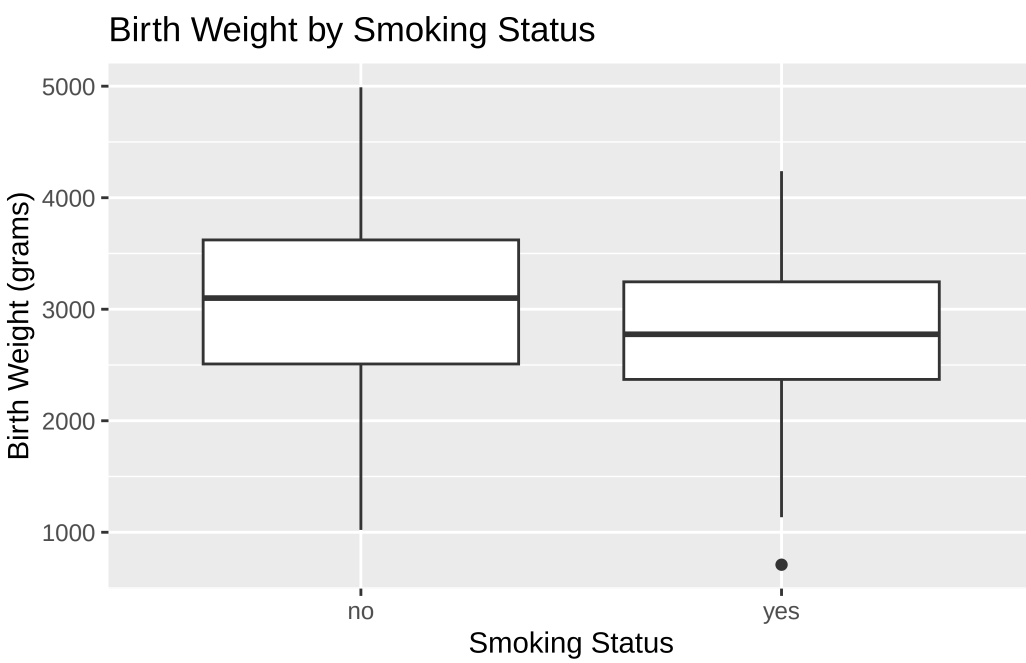
Medical Statistics – Answers lab 3

Part 1: Independent Samples t-test and Mann-Whitney U Test

Exploratory data analysis



Question 1

Based on the boxplot, do you expect the smoking status to have an effect on birth weight?

Answer question 1

Based on the boxplot, it appears that mothers who smoked have lower birth weights on average compared to those who did not smoke. This suggests that smoking status may indeed have an effect on birth weight.

Independent samples t-test

```
    Two Sample t-test

data:  bwt by smoke
t = 2.6336, df = 187, p-value = 0.009156
alternative hypothesis: true difference in means between group no and group yes is not equal
95 percent confidence interval:
  70.69274 492.73382
sample estimates:
 mean in group no mean in group yes
         3054.957          2773.243
```

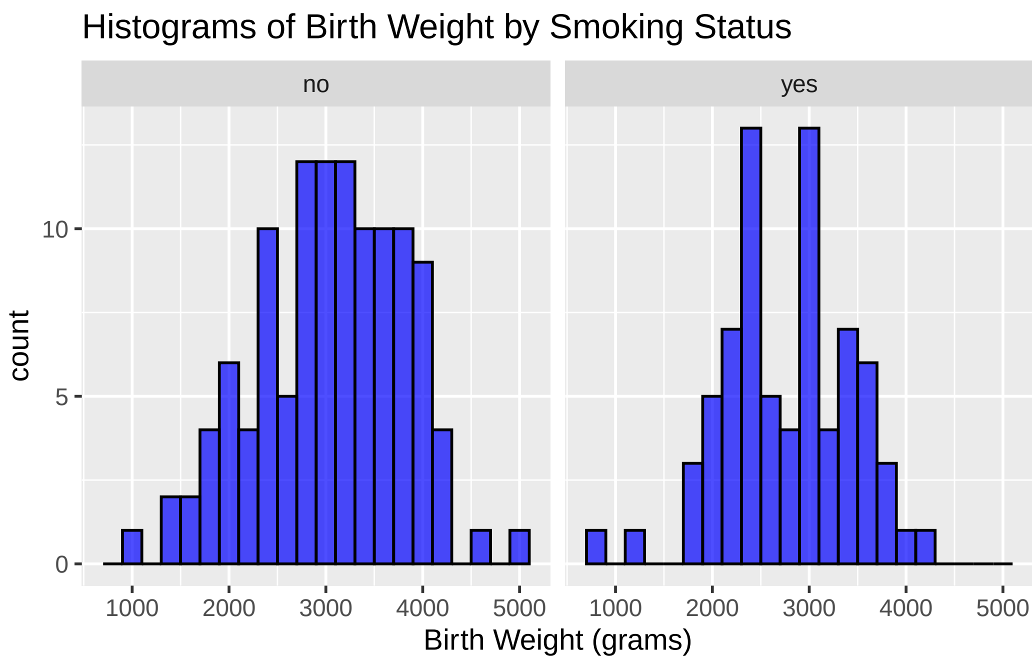
Question 2

Based on the results of the independent samples t-test, is there a statistically significant difference in birth weight between mothers who smoked and those who did not?

Answer question 2

The p-value of the independent samples t-test (p=0.009) is less than 0.05, which indicates that there is a statistically significant difference in the mean birth weight between mothers who smoked and those who did not.

Checking of assumptions



Question 3

Do the histograms indicate that the birth weight data are approximately normally distributed for both groups?

Answer question 3

Based on the histograms, the distributions in both groups appear reasonably normal for the purposes of conducting an independent samples t-test.

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  1.3901 0.2399
      187
```

Question 4

Based on Levene's test, does the assumption of equal variances hold?

Answer question 4

The p-value of Levene's test (p = 0.2399) is greater than 0.05, indicating insufficient evidence to reject the null hypothesis of equal variances. Therefore, we can assume that the assumption of equal variances holds.

95% Confidence Interval for the mean difference

```
# A tibble: 2 x 4
  smoke  mean    sd     n
  <fct> <dbl> <dbl> <int>
1 no    3055.  752.   115
2 yes   2773.  660.    74
```

Exercise

Based on these summary statistics, calculate the pooled standard deviation and the standard error of the mean difference. Then compute the 95% confidence interval for the mean difference in birth weight between mothers who smoked and those who did not.

Answer

Calculations for Birth Weight Difference

1. **Pooled Standard Deviation**
   The pooled standard deviation is calculated using the formula:

$$s_p = \sqrt{\frac{(n_{no} - 1) \cdot s_{no}^2 + (n_{yes} - 1) \cdot s_{yes}^2}{n_{no} + n_{yes} - 2}}$$

   Substituting the values:

$$s_p = \sqrt{\frac{(115 - 1) \cdot 752^2 + (74 - 1) \cdot 660^2}{115 + 74 - 2}} = 717.49$$

2. **Standard Error of the Mean Difference**
   The standard error (SE) of the mean difference is calculated as:

$$SE = s_p \cdot \sqrt{\frac{1}{n_{no}} + \frac{1}{n_{yes}}}$$

   Substituting the values:

$$SE = 717.49 \cdot \sqrt{\frac{1}{115} + \frac{1}{74}} = 106.93$$

3. **Mean Difference**
   The difference in means between the groups is:

$$\text{Mean Difference} = \bar{X}_{no} - \bar{X}_{yes} = 3055 - 2773 = 282$$

4. **95% Confidence Interval**
   To construct the exact 95% CI, we need the 97.5th percentile from the t-distribution with $115 + 74 - 2 = 187$ degrees of freedom. Since the table in Appendix B does not go beyond 150 df, we use 1.96 (the value from the standard normal distribution) as an approximation. The confidence interval is then calculated as:

$$\text{CI} = \text{Mean Difference} \pm 1.96 \cdot SE$$

   Substituting the values:

$$\text{CI} = 282 \pm 1.96 \cdot 106.93 = (72.43, 491.57)$$

Results

- **Pooled Standard Deviation:** 717.49
- **Standard Error of the Mean Difference:** 106.93
- **95% Confidence Interval for the Mean Difference:** (72.43, 491.57)

Question 5

Does your manually calculated 95% confidence interval for the mean difference in birth weight between the two groups agree with the one provided in the output of the `t.test()` function?

Answer question 5

The manually calculated 95% confidence interval for the mean difference in birth weight between mothers who smoked and those who did not is (72.43, 491.57), is approximately equal to the one provided in the output of the `t.test()` function, wich is (70.69, 492.73). The latter is marginally wider due to the use of the 97.5 the percentile of the t-distribution with 187 degrees of freedom (1.97) rather than the standard normal distribution (1.96). Additionally, small differences arise from rounding in the manual calculation.

Mann-Whitney U Test

```
	Wilcoxon rank sum test with continuity correction

data:  bwt by smoke
W = 5243.5, p-value = 0.007109
alternative hypothesis: true location shift is not equal to 0
```

Question 6

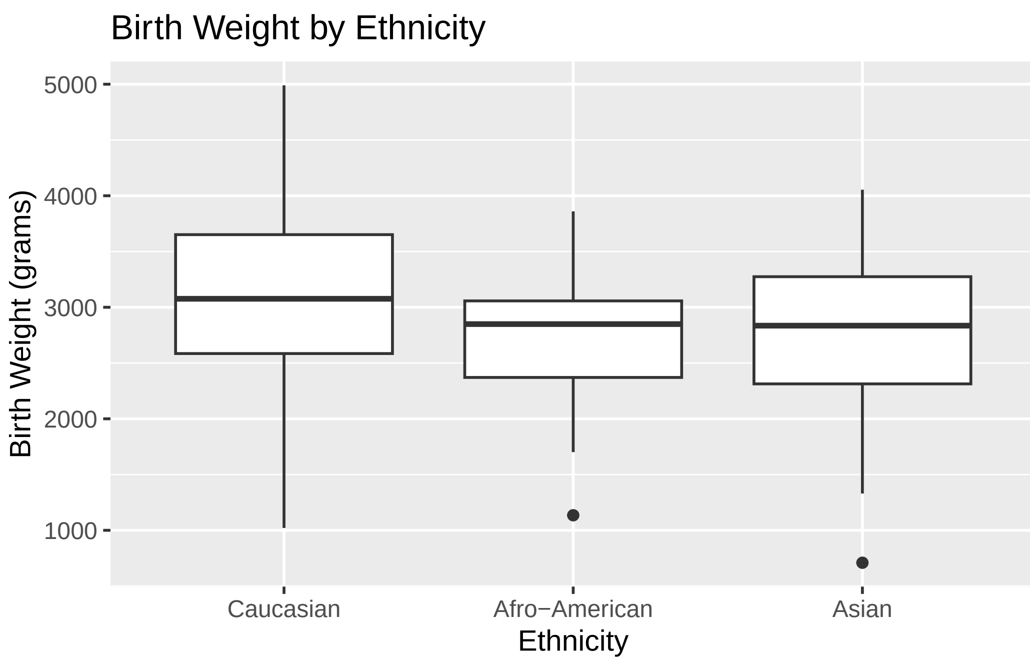
What are the null and alternative hypotheses for the Mann-Whitney U test, and what does the p-value indicate about the difference in birth weight between mothers who smoked and those who did not?

Answer question 6

The null hypothesis for the Mann-Whitney U test is that there is no difference in the median birth weight between mothers who smoked and those who did not. The alternative hypothesis is that there is a difference in median birth weight between the two groups. The p-value of the test is 0.0071, which is less than 0.05. Therefore, we reject the null hypothesis and conclude that there is a statistically significant difference in median birth weight between mothers who smoked and those who did not.

Part 2: One-Way ANOVA and Kruskal-Wallis Test

Explaratory data analysis



Question 7

What does the boxplot suggest about the distribution of birth weights across different ethnic groups?

Answer question 7

The boxplot suggests that there may be differences in birth weight across the three ehtnic groups. More specifically, The mean birth weight appears to be higher for the Causasian group compared to the Afro-American and Asian groups.

One-way ANOVA

```
             Df   Sum Sq Mean Sq F value  Pr(>F)
ethnicity     2  5070608 2535304   4.972 0.00788 **
Residuals   186 94846445  509927
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 8

What conclusions can be drawn from the results of the one-way ANOVA?

Answer question 8

The p-value from the one-way ANOVA (p=0.0079) is less than 0.05, indicating that there is at least one ethic group with a significantly different birth weight.

Post-hoc tests

```
    Pairwise comparisons using t tests with pooled SD

data:  lowbwt$bwt and lowbwt$ethnicity

              Caucasian Afro-American
Afro-American 0.048     -
Asian         0.027     1.000

P value adjustment method: bonferroni
```

Question 9

What conclusions can be drawn from the post-hoc comparisons?

Answer question 9

It follows from the post-hoc comparisons with the Bonferroni-adjusted p-values that the mean birth weight of Caucasian infants is significantly different from that of Afro-American infants (p=0.048) and Asian infants (p=0.027). There is no significant difference in mean birth weight between Afro-American and Asian infants (p=1).

Kruskal-Wallis Test

```
    Kruskal-Wallis rank sum test

data:  bwt by ethnicity
Kruskal-Wallis chi-squared = 8.5909, df = 2, p-value = 0.01363
```

Question 10

Are the results of the Kruskal-Wallis test consistent with the one-way ANOVA results?

Answer question 10

The p-value from the Kruskal-Wallis test (p=0.014) is consistent with the one-way ANOVA results, indicating that there is at least one ethnic group with a significantly different median birth weight.

```
  Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 8.5909, df = 2, p-value = 0.01

                           Comparison of x by group
                                 (Bonferroni)
Col Mean-|
Row Mean |   Afro-Ame      Asian
---------+----------------------
   Asian |  -0.667476
         |     1.0000
         |
Caucasia |  -2.380803  -2.337641
         |     0.0518     0.0582

alpha = 0.05
Reject Ho if p <= alpha

$chi2
[1] 8.590907

$Z
[1] -0.6674762 -2.3808031 -2.3376414

$altP
[1] 0.50446798 0.01727494 0.01940586

$altP.adjusted
[1] 1.00000000 0.05182482 0.05821758

$comparisons
```

```
[1] "Afro-American - Asian"     "Afro-American - Caucasian"
[3] "Asian - Caucasian"
```

Question 11

What conclusions can be drawn from Dunn's test? Are these consistent with the post-hoc comparisons from the one-way ANOVA?

Answer question 11

Unlike the post-hoc comparisons from the one-way ANOVA, Dunn's test with Bonferroni correction does not show statistically significant differences between any pair of groups at the 0.05 level. Although the adjusted p-values for Caucasian vs. Afro-American (p=0.052) and Caucasian vs. Asian (p=0.058) are close to 0.05, they do not reach statistical significance. This illustrates that non-parametric post-hoc tests can be more conservative than their parametric counterparts.

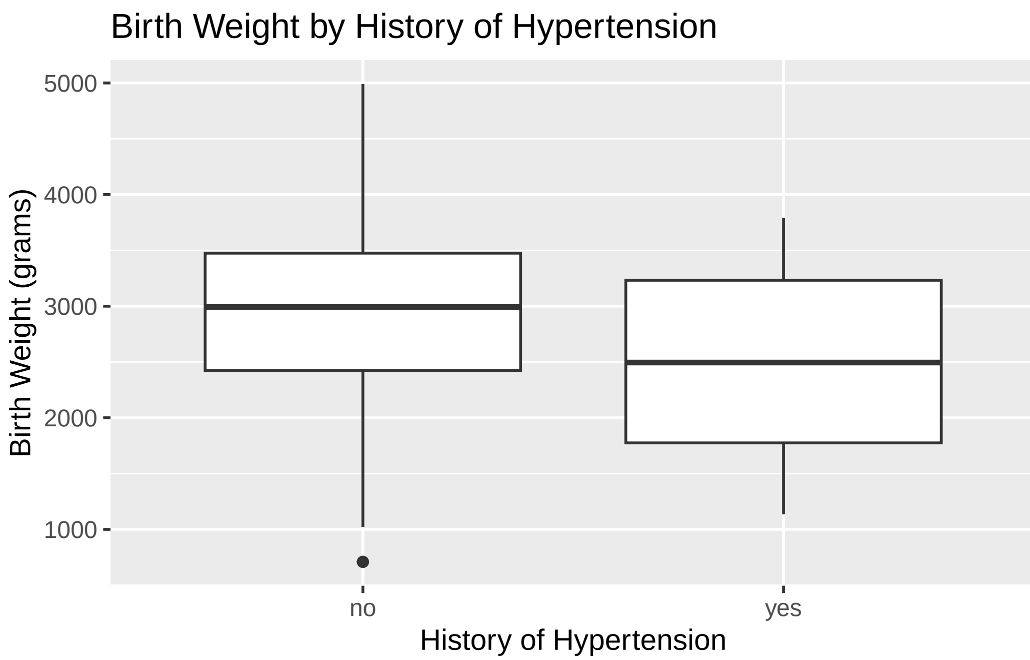
Part 3: Unguided exercises

Effect of hypertension on birth weight

Examine the effect of history of hypertension on birth weight by performing the following steps:

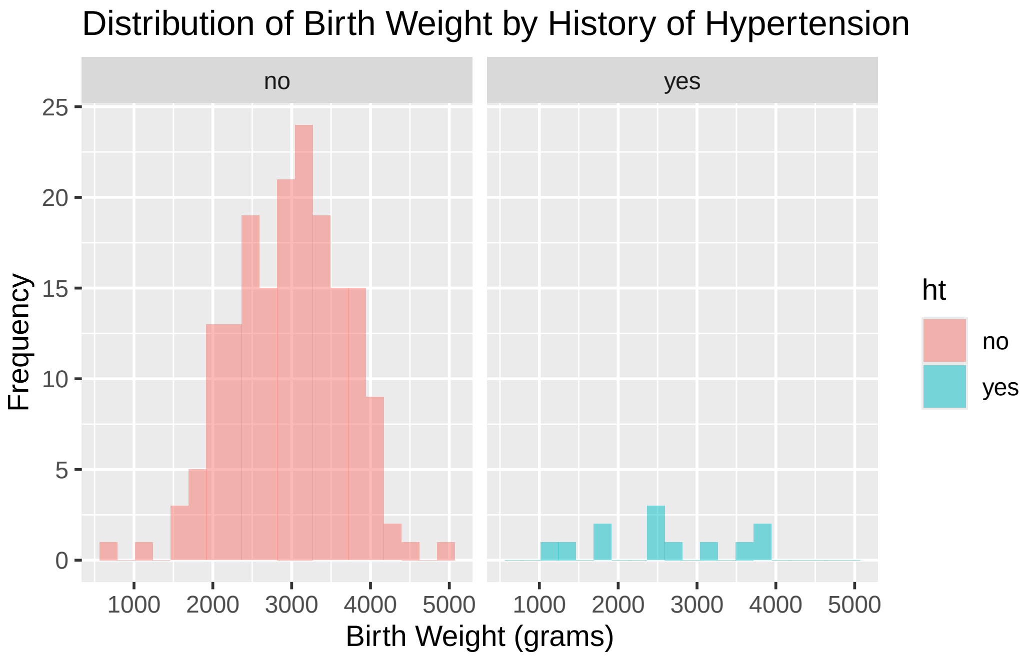
- Create a boxplot to visualize the distribution of birth weights by history of hypertension
- Perform an independent samples t-test to compare the mean birth weights between mothers with and without a history of hypertension
- Check the assumptions of the t-test, including normality and homogeneity of variances
- If the assumptions of the t-test are violated, perform a Mann-Whitney U test as a non-parametric alternative

Example solution




```
    Two Sample t-test

data:  bwt by ht
t = 2.0192, df = 187, p-value = 0.04489
alternative hypothesis: true difference in means between group no and group yes is not equal
95 percent confidence interval:
  10.02413 861.09734
sample estimates:
 mean in group no mean in group yes
         2972.311          2536.750
```




```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  1.2851 0.2584
      187
```

Evaluation of the results: Although the p-value from the independent samples t-test suggests that there is no statistically significant difference in birth weight between mothers with and without a history of hypertension, the group sizes are highly imbalanced (177 vs. 12). This imbalance reduces the reliability of the t-test results and may limit its power to detect true differences. Additionally, the small sample size in the hypertension group makes it harder to assess assumptions such as normality. To check the robustness of the findings, we also conduct the Mann-Whitney U test as a non-parametric alternative.

```
    Wilcoxon rank sum test with continuity correction

data:  bwt by ht
W = 1350, p-value = 0.1169
alternative hypothesis: true location shift is not equal to 0
```

Overall conclusion: Both the independent samples t-test and the Mann-Whitney U test suggest that there is no statistically significant difference in birth weight between mothers with and without a history of hypertension. This consistency across methods supports the robustness of the findings, despite the imbalance in group sizes.

Comparing red cell folate levels across ventilation strategies in cardiac bypass patients

Twenty-two patients undergoing cardiac bypass surgery were randomized to one of three ventilation groups:

- **Group I**: Received a 50% nitrous oxide and 50% oxygen mixture continuously for 24 hours
- **Group II**: Received a 50% nitrous oxide and 50% oxygen mixture only during the operation
- **Group III**: Received no nitrous oxide and a 35-50% oxygen mixture continuously for 24 hours

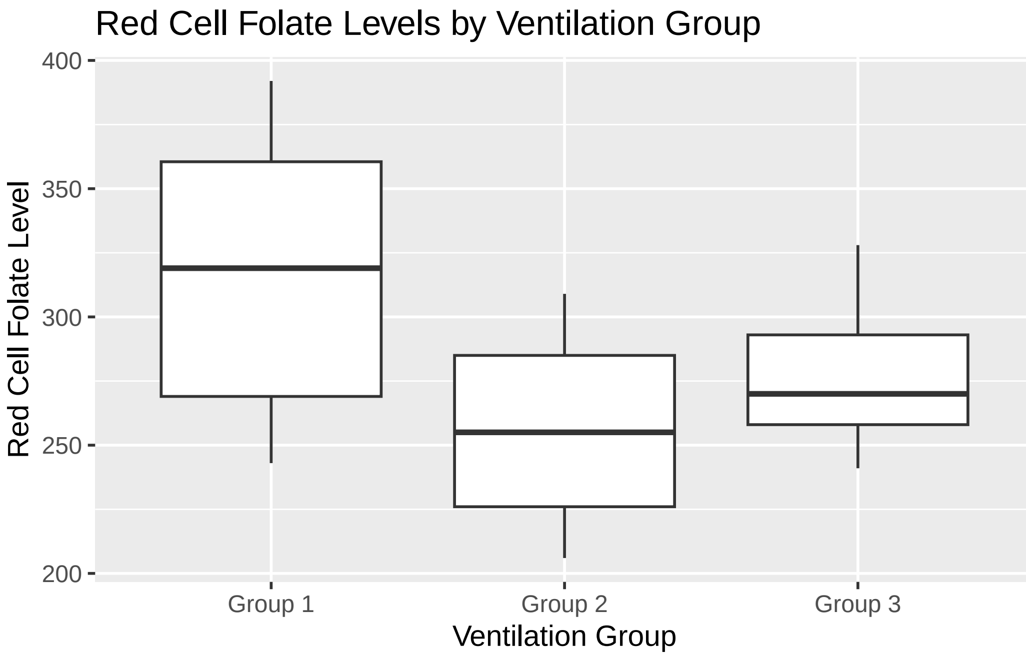
The data file `ex5_6.sav` (available on Brightspace) contains the red cell folate levels for the three groups after 24 hours of ventilation. The aim of this study is to compare the three groups and test whether they have the same red cell folate levels.

Tasks

1. **Exploratory data analysis** Create a boxplot to visualize the distribution of red cell folate levels by ventilation group. Based on this plot:
   - What are your first conclusions regarding the means and variances of the different groups?
2. **Perform a one-way ANOVA**:
   - Interpret the results
   - Are the assumptions satisfied?
3. **Try a log transformation on the data**:
   - Perform another one-way ANOVA
   - Are the assumptions satisfied after the transformation?
4. **Determine which means differ**:
   - Which means do you think differ?
   - Explain your reasoning.
5. **Try a non-parametric approach**:
   - What are your conclusions from this method?

Example solution

Exploratory data analysis



The groups seem to be different with respect to their means, but also with respect to their variation (range has different length)

Perform a one way ANOVA, and interpret the results. Are the conditions satisfied?

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        2  15516    7758   3.711 0.0436 *
Residuals   19  39716    2090
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  3.6413 0.04585 *
      19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA reveals significant results P= 0.044; this means that the null hypothesis of equal means in the three groups is rejected. However, according to the test of homogeneity of variances, the assumption of equal variances is violated (P-value = 0.044).

Try a log transformation on the data, and perform again a one-way ANOVA. Are now the assumptions satisfied?

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        2 0.1784 0.08922   3.539 0.0494 *
Residuals   19 0.4790 0.02521
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2   1.889 0.1785
      19
```

The test of equal variances for the log-transformed data reveals that the assumption of equal variances can be made. Test results for the ANOVA: P-value = 0.049. This is on the boundary of significance. There is indication that at least one pair of groups have different means.

Which means do differ according to you? Why?

```
    Pairwise comparisons using t tests with pooled SD

data:  df$rcfl and df$group

        Group 1 Group 2
Group 2 0.042   -
Group 3 0.464   1.000

P value adjustment method: bonferroni


    Pairwise comparisons using t tests with pooled SD

data:  df$log_rcfl and df$group
```

```
        Group 1 Group 2
Group 2 0.047   -
Group 3 0.597   1.000

P value adjustment method: bonferroni
```

From the paired comparisons with Bonferroni-corrections it appears that group1 and group 2 differ significantly with respect to their means (but on the boundary, P-value = 0.047).

(e) Try a non-parametric approach on these data. What are now your conclusions?

```
    Kruskal-Wallis rank sum test

data:  rcfl by group
Kruskal-Wallis chi-squared = 4.1852, df = 2, p-value = 0.1234
```

The Kruskal-Wallis test gives P-value = 0.123, indicating that there are no significant differences between the three groups.

Caveat

In this example, the sample size is very limited, making it difficult — if not impossible — to verify whether the assumptions underlying the one-way ANOVA are met. The Kruskal-Wallis test is a good alternative in such cases.