Chapter 6 – Tests of Differences

**Exercise 1**

The data below shows students’ scores before and after an intervention in a school class.

1. Explain what type of analysis is appropriate for this data set to determine whether or not there is a significant difference between the students’ scores before and after the intervention class.

A related sample design such as paired samples t-test is necessary to test the hypothesis since the grades to be compared were taken from the same students.

1. State the null and alternative hypotheses.

H0: There is no significant difference between the mean grades of students before and after the intervention class.

H1: There is a significant difference between the mean grades of students before and after the intervention class.

1. Find the data below by opening Student Grades.csv.

|  |  |  |
| --- | --- | --- |
| Student | Before Intervention | After Intervention |
| 1 | 18 | 22 |
| 2 | 21 | 25 |
| 3 | 16 | 17 |
| 4 | 22 | 24 |
| 5 | 19 | 16 |
| 6 | 24 | 29 |
| 7 | 17 | 20 |
| 8 | 21 | 23 |
| 9 | 23 | 19 |
| 10 | 18 | 20 |
| 11 | 14 | 15 |
| 12 | 16 | 15 |
| 13 | 16 | 18 |
| 14 | 19 | 26 |
| 15 | 18 | 18 |
| 16 | 20 | 24 |
| 17 | 12 | 18 |
| 18 | 22 | 25 |
| 19 | 15 | 19 |
| 20 | 17 | 16 |

> file = read.csv("Student Grades.csv")

> file

1. Perform the appropriate analysis and interpret the results according to the formed hypothesis.

> before = file$Grade\_before

> after = file$Grade\_after

> shapiro.test(before) # non-significant result

> shapiro.test(after) # non-significant result

> summary(file)



Let us assume that we expected the 'after' result to be greater than before.

> t.test(before, after, alternative="less", paired=TRUE) # A < B



There is a significant difference between the students’ grades before and after the intervention class (p < .05) with a mean difference of 2.

barplot(c(mean(before), mean(after)),

names.arg=c("before","after"),

xlab="Before", ylab="After")



**Exercise 2**

 A study assessed the effectiveness of a new drug designed to reduce repetitive behaviors in children affected with autism. A total of 8 children with autism enroll in the study and the amount of time that each child is engaged in repetitive behavior during three-hour observation periods are measured both before treatment and then again after taking the new medication for a period of 1 week. The data are shown below (find this data in Drug Treatment.csv file).

|  |  |  |
| --- | --- | --- |
| **Child** | **Before Treatment** | **After 1 Week of Treatment** |
| 1 | 85 | 75 |
| 2 | 70 | 50 |
| 3 | 40 | 50 |
| 4 | 65 | 40 |
| 5 | 80 | 20 |
| 6 | 75 | 65 |
| 7 | 55 | 40 |
| 8 | 20 | 25 |

1. Given this small dataset, explain what is the appropriate test to see if there is a significant difference between the median of two groups.

Wilcoxon signed rank test, which is a non-parametric counterpart of paired samples t-test is appropriate to use in this study.

1. State the null and alternative hypotheses.

H0: The average repetitive behavior scores before and after treatment are the same.

H1: The average repetitive behavior scores before and after treatment are not the same.

1. Use the data in Drug Treatment.csv.

> file = read.csv("Drug Treatment.csv")

> before = file$Score\_before

> after = file$Score\_after

1. Perform the appropriate analysis and interpret the results according to the formed hypothesis. As previously, look at descriptive statistics as well as the relevant statistical test.

> summary(file)



The median is smaller before the intervention than afterwards.

barplot(c(median(before), median(after)),

names.arg=c("before","after"),

xlab="Before", ylab="After")

Note the use of the median in the code for the bar plot.



> wilcox.test(before, after, paired=TRUE)



The plot shows that the average score before treatment is higher than the score after treatment; there is a median difference of 22.5. If we applied a 0.05 critical value, the Wilcoxon signed rank test results would not allow us to reject the null hypothesis that the scores are similar before and after the treatment. However, given the small number of cases, some would consider applying a critical value of 0.10, so *p* < .10; however, most would suggest using larger samples!

**Exercise 3**

In an attempt to improve students’ grades within one class, two sets of intervention classes were performed. The students’ grades were recorded as below.

|  |  |  |  |
| --- | --- | --- | --- |
| Student | Before Intervention | After First Intervention Class | After Second Intervention Class |
| 1 | 18 | 22 | 20 |
| 2 | 21 | 25 | 22 |
| 3 | 16 | 17 | 20 |
| 4 | 22 | 24 | 26 |
| 5 | 19 | 16 | 15 |
| 6 | 24 | 29 | 30 |
| 7 | 17 | 20 | 25 |
| 8 | 21 | 23 | 26 |
| 9 | 23 | 19 | 25 |
| 10 | 18 | 20 | 23 |
| 11 | 14 | 15 | 20 |
| 12 | 16 | 15 | 14 |
| 13 | 16 | 18 | 22 |
| 14 | 19 | 26 | 24 |
| 15 | 18 | 18 | 19 |
| 16 | 20 | 24 | 23 |
| 17 | 12 | 18 | 25 |
| 18 | 22 | 25 | 27 |
| 19 | 15 | 19 | 22 |
| 20 | 17 | 16 | 20 |

1. Explain why administering multiple t-tests would not be beneficial for this scenario. Which test would be appropriate?

Every time you conduct a t-test, there is a chance of rejecting a true null hypothesis (a 'type 1 error'). Thus, conducting multiple paired sample t-tests in this case will increase the error rate.

To avoid this, with the same participants, the one-way Repeated Measures ANOVA is appropriate.

1. State the null and alternative hypotheses.

H0: There is no significant difference between the average grades among the three groups.

H1: There is a significant difference between the average grades among the three groups.

1. Load and view Student Grades Extended.csv.

> file = read.csv("Student Grades Extended.csv")

> file



1. Perform the appropriate analysis and interpret the results related to the hypothesis.

Perform the procedure for setting up Repeated Measures ANOVA in R as shown in the book, substituting the file name and variable names as appropriate, using Grade\_before, After\_1 and After\_2).

The results should end up looking something like this:



Since the *p* value is less than 0.05, the level of significance that was pre-determined, reject the null hypothesis that there is no significant difference between the average grades among the three groups.

Further adaptation of the code used in the Repeated Measures one-way ANOVA section will also allow you to find effect sizes (e.g. omega squared = 0.427) and multiple comparison tests. All pairings produce *p* values that are lower than 0.05. This is replicated in all the tests cited in this section of the book; the Bonferroni and 'BY' tests produce rather higher *p* values than the other tests for the pairing of the two 'after' results.

**Exercise 4**

Suppose that you want to compare the grades of students from two *different* classes, where the first class followed a traditional teaching method, whereas the second was an experimental class where a different teaching style was applied. Do note that each separate group needs to be represented by a number (here, within the 'Class' variable); this would be the same with an ANOVA (independent samples / unrelated), only with more than 2 such numbers.

This data set is available in Teaching Methods.csv

|  |  |  |
| --- | --- | --- |
| Student | Grade | Class |
| 1 | 18 | 1 |
| 2 | 21 | 1 |
| 3 | 16 | 1 |
| 4 | 22 | 1 |
| 5 | 19 | 1 |
| 6 | 24 | 1 |
| 7 | 17 | 1 |
| 8 | 21 | 1 |
| 9 | 23 | 1 |
| 10 | 18 | 1 |
| 11 | 14 | 1 |
| 12 | 16 | 1 |
| 13 | 16 | 1 |
| 14 | 19 | 1 |
| 15 | 18 | 1 |
| 16 | 20 | 1 |
| 17 | 12 | 1 |
| 18 | 22 | 1 |
| 19 | 15 | 1 |
| 20 | 17 | 1 |
| 21 | 22 | 2 |
| 22 | 25 | 2 |
| 23 | 17 | 2 |
| 24 | 24 | 2 |
| 25 | 16 | 2 |
| 26 | 29 | 2 |
| 27 | 20 | 2 |
| 28 | 23 | 2 |
| 29 | 19 | 2 |
| 30 | 20 | 2 |
| 31 | 15 | 2 |
| 32 | 15 | 2 |
| 33 | 18 | 2 |
| 34 | 26 | 2 |
| 35 | 18 | 2 |
| 36 | 24 | 2 |
| 37 | 18 | 2 |
| 38 | 25 | 2 |
| 39 | 19 | 2 |
| 40 | 16 | 2 |

1. State the null and alternative hypotheses.

H0: There is no significant difference between the average grades of students in the two classes.

H1: There is a significant difference between the average grades of students in the two classes.

1. Which test is appropriate here?

The Independent samples *t* test. There are only two groups. The grades being compared are from different students (students did not take both classes). An examination of the dependent variable (Grade) will show that the data can be considered normal; otherwise, a Mann-Whitney test would be preferred.

1. Input the data in R, perform the analysis and interpret the results.

Use the Independent samples *t* test procedure, as shown in the book. Assuming that we had no strong rationale for expecting a difference in one direction, the two-tailed hypothesis should be tested. A visualisation would also be nice.



> file = read.csv("Teaching methods.csv")

> experiment = with(file, data.frame(Grade, Class))

> experiment = na.omit(experiment)

> grade = experiment$Grade

> class = experiment$Class

> boxplot(grade ~ class)



Since the *p* value is greater than 0.05, we do not evidence supporting the rejection of the null hypothesis. There is no significant difference between the average grades of students in the two classes.