

```
# IRDR04 Research Tools Module
# Week 5 Tutorial
# 30/10/18 9-11am @ Birkbeck, Malet Street 414/415 Public Cluster
# Tutor: Dr Alison Fairbrass, UCL Centre for Biodiversity & Environment Research, alison.fairbrass@gmail.com
```

```
# # # Exercises and Solutions # # #
```

```
# Exercise 1
```

```
# You want to investigate whether the academic performance of a group of school children changed
# significantly after they experienced the 2011 Japanese earthquake and tsunami.
```

```
# There are 130 children in your study. Before the event they had a mean test score of 100 with a
# standard deviation of 10. After the event their mean score was 85 with a standard deviation of 11.
```

```
# A) Create a dataframe of your data by randomly selecting test scores for your 130 children from the
# appropriate normal distributions.
```

```
# Group 1: Before event
set.seed(10)
score <- rnorm(n=130, mean=100, sd=10)
group <- 1:length(score)
group <- rep("Group 1",length(group))
df1 <- data.frame(score, group)
```

```
# Group 2: After event
set.seed(10)
score <- rnorm(n=130, mean=85, sd=11)
group <- 1:length(score)
group <- rep("Group 2",length(group))
df2 <- data.frame(score, group)
```

```
# Combine into one dataframe
df <- rbind(df1, df2)
```

```
# B) Check the assumptions of parametric testing and justify your choice of a statistical test.
```

```
# Check distribution of data
hist(df$score[df$group=="Group 1"], main = "Group 1", xlab = "Score")
hist(df$score[df$group=="Group 2"], main = "Group 2", xlab = "Score")
```

```
# Check homogeneity of the groups variance
library(car)
leveneTest(score ~ group, data = df)
```

```
# Justification. A parametric test is suitable as the data is normally distributed and the variances
# of the two groups are equal. The sample size is large.
# A paired t-test should be used as the data of the groups are not independent.
```

```
# C) Use an appropriate statistical test to draw a conclusion about the childrens test scores
# before and after the event. Report the results of your test appropriately.
```

```
t.test(df$score[df$group=="Group 1"], df$score[df$group=="Group 2"], paired=T)
# Calculate standard deviation of the two groups
mean(df$score[df$group=="Group 1"])
sd(df$score[df$group=="Group 1"])
mean(df$score[df$group=="Group 2"])
sd(df$score[df$group=="Group 2"])
```

```
# Report: There was a significant difference in the academic scores of students before (M=100, SD=9.26)
# and after (M=84, SD=10.19), t=185.9, p<0.001, the event.
```

```
# Exercise 2
```

```
# You want to investigate whether the happiness of men changes significantly with age.
# You survey a group of 15 men at age 20 and then at age 35, and ask them on a scale of 1 to 5
# how happy they feel, from 1 = very unhappy to 5 = very happy. Here are the results:
# Men aged 20 happiness scores: 4,5,3,4,5,5,2,4,3,3,4,5,2,1,4
# Men aged 35 happiness scores: 3,4,5,5,1,1,3,2,2,4,2,5,3,3,2
```

```
# A) Create a dataframe for this data
```

```
men_20 <- c(4,5,3,4,5,5,2,4,3,3,4,5,2,1,4)
men_35 <- c(3,4,5,5,1,1,3,2,2,4,2,5,3,3,2)
# Create a data frame
df <- data.frame(
  group = rep(c("20", "35"), each = length(men_20)),
  happiness_score = c(men_20, men_35)
)
```

```
# B) Justify why a parametric test would not be appropriate for your investigation.
```

```
# Data is not normally distributed. Ordinal data is appropriate for non-parametric tests.
# Small sample size.
```

```
# C) Using a boxplot assess whether there is likely to be a significant difference between the
# happiness scores of the men at age 20 and 35.
```

```
boxplot(happiness_score~group, data = df)
```

```
# D) Using an appropriate statistical test draw a conclusion about the happiness of aging men.
# Report the results of your test appropriately.
wilcox.test(men_20, men_35, paired = T, exact=F)
```

```
median(men_20)
n <- length(men_20)
# Lower CI
rank <- (n/2) - ((1.96*(sqrt(n)))/2)
rank <- round(rank)
LowerCI <- sort(men_20)[rank]
# Upper CI
rank <- 1 + (n/2) + ((1.96*(sqrt(n)))/2)
rank <- round(rank)
UpperCI <- sort(men_20)[rank]
```

```

median(men_35)
n <- length(men_35)
# Lower CI
rank <- (n/2) - ((1.96*(sqrt(n)))/2)
rank <- round(rank)
LowerCI <- sort(men_35)[rank]
# Upper CI
rank <- 1 + (n/2) + ((1.96*(sqrt(n)))/2)
rank <- round(rank)
UpperCI <- sort(men_35)[rank]

# There was no significant difference between the happiness scores of men aged 20 (Mdn=4, CI[3,5])
# and men aged 35 (Mdn=3, CI[2,4]), U=69, P= >0.05.

# E) If we surveyed the same group of men again when they were 50, what would be the appropriate test
# to investigate differences in happiness at 20, 35 and this third age point? Why?

# Friedman test is appropriate for 3 related groups of ordinal data.

```