

```
# IRDR04 Research Tools Module
# Week 4 Tutorial
# 23/10/18 9-11am @ Birkbeck, Malet Street 414/415 Public Cluster
# Tutor: Dr Alison Fairbrass, UCL Centre for Biodiversity & Environment Research, alison.fairbrass@gmail.com
```

```
## Exercises ##
```

```
# Exercise 1
```

```
# Ten people have their height and weight measured
# Create a dataframe with their data
```

```
# Person Height (cm) Weight (kg)
# 1 140 50
# 2 171 73
# 3 162 52
# 4 159 50
# 5 169 57
# 6 175 80
# 7 168 64
# 8 145 47
# 9 175 86
# 10 180 86
```

```
# Solution 1
```

```
person <- c(1:10)
height <- c(140, 171, 162, 159, 169, 175, 168, 145, 175, 180)
weight <- c(50, 73, 52, 50, 57, 80, 64, 47, 86, 86)
df <- data.frame(person, height, weight)
```

```
# Exercise 2
```

```
# Calculate the total, mean, median, standard deviation and variance for your height and weight data
# Create a new dataframe with these summary statistics
# Export you dataframe with only the height and weight descriptive statistics to a csv file called "summaryStatistics.csv" in your working directory
```

```
# Solution 2
```

```
df_total <- colSums(df)
df_mean <- apply(df, 2, mean)
df_median <- apply(df, 2, median)
df_var <- apply(df, 2, var)
df_sd <- apply(df, 2, sd)
df2<-data.frame(df_total, df_mean, df_median, df_var, df_sd)
df2 <- data.frame(t(df2))
df2 <- df2[c('height', 'weight')]
write.csv(df2, "summaryStatistics.csv")
```

```
# Exercise 3
```

```
# Create a new dataset of height and weight data for 30 study participants.
# The data must be normally distributed.
# The height data must have a mean of 170 and standard deviation of 15.
# The weight data must have a mean of 60 and a standard deviation of 10.
# Make sure you can make the dataset again, so your experiment is repeatable.
```

```
# Solution 3
```

```
set.seed(10)
height <- rnorm(n=30, mean = 170, sd = 15)
set.seed(10)
weight <- rnorm(n=30, mean = 60, sd = 10)
df <- data.frame(height, weight)
```

```
# Exercise 4
```

```
# Generate histograms and density plots of your study participants' characteristics
# Do some googling and play around with changing the appearance of these plots e.g. title, axis labels, colour, histogram breaks
hist(df$height, breaks = 10)
hist(df$weight)
plot(density(df$height))
plot(density(df$weight))
```

```
# Exercise 5
```

```
# Create a csv file of your study participants' data and save it in your working directory.
```

```
# Solution 5
```

```
write.csv(df, "studyParticipants.csv", row.names = F)
```

```
# Exercise 6
```

```
# Assume that a waiting time of patients in a hospital waiting room is normally distributed
# with a mean of 50 minutes and a standard deviation of 15 minutes
```

```
# a) Generate a density plot of the distribution of the waiting time data.
# Again play arounds with making the plot look a bit nicer
```

```
# Solution 6a)
```

```
plot(density(rnorm(10000, mean=50, sd=15)))
```

```
# b) Compute the probability that a patient (who is going to be randomly selected)
# will need to wait less than 45 minutes
```

```
# Solution 6b)
```

```
pnorm(45, mean = 50, sd = 15)
```

```
# c) Compute the probability that a patient (who is going to be randomly selected)
# will need to wait for between 30 and 60 minutes
```

```
# Solution 6c)
```

```
pnorm(60, mean = 50, sd = 15) - pnorm(30, mean = 50, sd = 15)
```