The first assignment was generally well done. Please note the following comments and go through the solutions carefully. Please take note of any comments you received and look carefully at the R code in the solutions, since marks will be deducted for poor presentation of graphs (title, units etc.), poor grammar, and incorrect, irrelevant or incomplete R code.

Common errors:

- Many students did not include the R code. Ensure that you only include relevant and correct R code. You should edit the code before placing in your assignment.

- Give labels and titles to your plots and tables. Refer to them by label in your text. For example:

  “From the numerical summary of ATP given in Table 1, we see that the average for all types is 1.1 nmol/mg with a std dev of 0.6 nmol/mg.”

- When presenting results in tables remember Ehrenberg’s principles. Round your results.

**Reminder:** most assignment questions are directly related to examples or exercises in the lecture notes. Work through those before attempting the assignment questions. Solutions to the exercises are posted on the stat100 homepage each Thursday - go to the Tutorial and Practical Exercises link.

**Question 1**

(b) From the numerical summary of ATP given in Table 1, we see that the average for all types is 1.1 nmol/mg with a std dev of 0.6 nmol/mg. When we examine each type (Table 2) we see that the lowest ATP occurs in the flooded European birch (mean = 0.3 nmol/mg, sd = 0.2 nmol/mg) and the greatest occurs for the control River birch (mean = 1.8 nmol/mg, sd = 0.2 nmol/mg).

```
mean  sd  0%  25%  50%  75%  100%  n
1.12 0.579 0.11 0.887 1.18 1.46 2.04 16
```

Table 1: Numerical summary of ATP (nmol/mg)
Table 2: Numerical summary of ATP (nmol/mg) by type

(c) & (d) From Figure 1,
- There is a lot of overlap in the distribution of ATP for the controlled European birch and the flooded river birch (compare $Q_1$ & $Q_3$).
- However, there is no overlap between the distribution of ATP for the controlled River birch and the distributions for the other three types. On average, the ATP for the control River birch is greater than the ATP for the other three types.
- Similarly, there is no overlap between the distribution of ATP for flooded European birch and the rest. It is likely that flooding European birch results in the lowest ATP concentration.
- Furthermore, the reduction in ATP concentration after flooding is greater for European birch than for River birch.

![Figure 1: ATP by tree type](image_url)
Question 2
Refer to Figure 2.

NB: The values given in the solutions were calculated precisely using R. Allowance was made for estimation by eye.

(a) The interval which contains the middle 60% of adult female heights is (156.7, 168.7) cm.

(b) The first quartile corresponds to the 25th percentile and is approximately 157.9cm. The third quartile corresponds to the 75th percentile and is approximately 167.5cm.

(c) Approximately 73% of adult women are shorter than 167cm.

![Figure 2: Distribution Function for Adult Female Heights](image)

Question 3

(a) & (b) From the boxplots (Fig. 3) and the numerical summaries (Table 3), we can see that there is some overlap between the two distributions.

- The mean and median study times are greater for females (165min & 175 min, respectively) than for males (117 mins & 120 mins,respectively), although there is a lot of variability in the distributions (sd = 57 min & 74 mins for females and males respectively).
- Furthermore, we see that while all females spent at least one hour studying on a typical night (minimum: 0% = 60), this was not the case for males. In fact, 25% of the males studied less than one hour on a typical night (25% =60).
• Interestingly there are two outliers. One female reported spending typically 6 hours per night studying and one male reported spending 5 hours per night. Are these errors? Have the students exaggerated? Certainly worth exploring.

**Remember** You can’t compare measures of centre (e.g., means) without including measures of variability!

Figure 3: Study time by gender

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>0%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>165</td>
<td>57</td>
<td>60</td>
<td>120</td>
<td>175</td>
<td>180</td>
<td>360</td>
<td>30</td>
</tr>
<tr>
<td>M</td>
<td>117</td>
<td>74</td>
<td>0</td>
<td>60</td>
<td>120</td>
<td>150</td>
<td>300</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 3: Summary of study time by gender
Appendix: R code

Question 1

# limit the number of significant figures
options(digits=3)

# summary statistics of ATP
numSummary(birch[,"ATP"], statistics=c("mean", "sd", "quantiles"))

# summary statistics of ATP by type
numSummary(birch[,"ATP"], groups=birch$type,
          statistics=c("mean", "sd", "quantiles"))

# boxplots of ATP for each of the four types.
boxplot(ATP~type, ylab="ATP", xlab="type", data=birch)

Question 3

# Numerical summary of study times by gender
numSummary(study[,"Minutes"], groups=study$Gender,
           statistics=c("mean", "sd", "quantiles"))

# Boxplots of study times by gender
boxplot(Minutes~Gender, ylab="Minutes", xlab="Gender", data=study)