Computer exercise 3

Simon Sigurdhsson, 900322–0291

This is a report on exercise 11.50 in Rice (2007), which concerns a data set of body temperature readings and heart rates in males and females. The exercise consists of 3 parts labeled (a) through (c). The exercise has been solved using the R statistical software and the code used is included throughout the report. The preamble of this code simply loads a couple of packages along with the data set, converting the temperature to degrees Celsius and dividing the dataset by gender (where 1 denotes male samples and 2 female samples):

```
library(ggplot2)
library(tikzDevice)
data <- read.csv("bodytemp.txt", header=TRUE)
data$temperature <- (data$temperature-32)*(5/9)
men   <- data[data$gender == 1,]
women <- data[data$gender == 2,]
```

Here, the `ggplot2` library is used to generate figures and the `tikzDevice` library is used to export them as Ti*k*Z code.

a. *Confidence interval of body temperature difference.*

```
temp.mean.m <- mean(men$temperature)
temp.mean.w <- mean(women$temperature)
temp.var.m <- var(men$temperature)
temp.var.w <- var(women$temperature)
s.p.2 <- ((length(men[,1])-1)*temp.var.m +
          (length(women[,1])-1)*temp.var.w) /
          (length(men[,1]) + length(women[,1]) - 2)
s.xy <- sqrt(s.p.2)*sqrt(1/length(men[,1]) + 1/length(women[,1]))
z.975 <- 1.980 # z_{\alpha/2} \approx 1.980 for n+m = 130
temp.ci.low  <- temp.mean.m - temp.mean.w - z.975*s.xy
temp.ci.high <- temp.mean.m - temp.mean.w + z.975*s.xy
tikz("tikz/box-temp.tex", width=2.5, height=2.5)
qplot(factor(gender), temperature, data=data, geom='boxplot')
dev.off()
```

As seen in figure 1 on page 3, the normal approximation seems reasonable for the body temperature data set, although the data may be slightly skewed for both genders. Using normal theory, we have $\mu_1 = \overline{X} = 36.72$ and $\mu_2 = \overline{Y} = 36.89$. We can form a confidence interval of $\mu = \mu_1 - \mu_2$ using the point estimate $\hat{\mu} = \overline{X} - \overline{Y}$ with

$$s_{\overline{X}-\overline{Y}} = \sqrt{\frac{(n-1)\operatorname{Var}(X) + (n-1)\operatorname{Var}(Y)}{2n-2}}\sqrt{\frac{2}{n}}, \quad (n = 130)$$

yielding a 95% confidence interval $\hat{\mu} \pm 1.980 s_{\overline{X}-\overline{Y}}$, which in this case is $[-0.30, -0.02]$. It seems that the difference in this case is significant, but this will be explored further in part (c).

b. *Confidence interval of heart rate difference.*

```
rate.mean.m <- mean(men$rate)
rate.mean.w <- mean(women$rate)
rate.var.m <- var(men$rate)
rate.var.w <- var(women$rate)
s.p.2 <- ((length(men[,1])-1)*rate.var.m +
          (length(women[,1])-1)*rate.var.w) /
          (length(men[,1]) + length(women[,1]) - 2)
s.xy <- sqrt(s.p.2)*sqrt(1/length(men[,1]) + 1/length(women[,1]))
z.975 <- 1.980 # z_{\alpha/2} \approx 1.980 for n+m = 130
rate.ci.low  <- rate.mean.m - rate.mean.w - z.975*s.xy
rate.ci.high <- rate.mean.m - rate.mean.w + z.975*s.xy
tikz("tikz/box-rate.tex", width=2.5, height=2.5)
qplot(factor(gender), rate, data=data, geom='boxplot')
dev.off()
```

As seen in figure 2 on the next page, the normal approximation is unreasonable for the heart rate data set. The male heart rates are skewed upwards, and the female heart rates are skewed downwards. As such, any results obtained directly from the data set are invalid. Despite this, we proceed as in part (a), yielding a 95% confidence interval of $[-3.24, 1.67]$.

c. *Comparing the body temperatures and heart rates.*

```
temp.t.reject <- t.test(temperature ~ gender, data=data)$p.value < 0.025
rate.t.reject <- t.test(rate ~ gender, data=data)$p.value < 0.025
```

Performing the (parametric) t test on both data sets tells us that the null hypothesis $H_0 : \mu_1 = \mu_2$ should be rejected for the temperature data set while it shouldn't for the heart rates (with 95% confidence).

```
temp.w.reject <- wilcox.test(temperature ~ gender, data=data)$p.value < 0.025
rate.w.reject <- wilcox.test(rate ~ gender, data=data)$p.value < 0.025
```

Instead performing the nonparametric Mann-Whitney (Wilcoxon rank) test on both data sets tells us that the null hypothesis shouldn't be rejected for any of the data sets.

Since the nonparametric Mann-Whitney test doesn't assume that the data has any specific distribution, it is likely more reliable here than the parametric t test which incorrectly assumes that the data is normally distributed. As such, one should conclude that with 95% confidence there is no actual difference in heart rates or temperatures between males and females.

References

Rice, John (2007). *Mathematical statistics and data analysis*. Australia Belmont, CA: Thompson/Brooks/Cole. ISBN: 9780495118688.

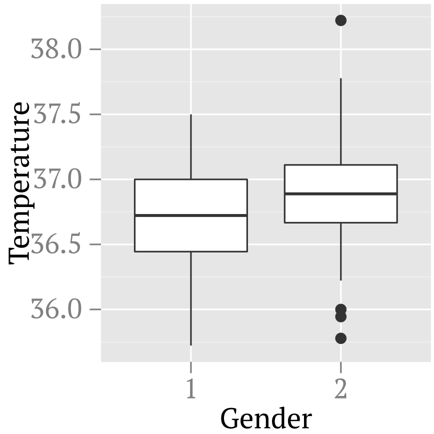

Figure 1: *Boxplot of the temperatures divided by gender.*

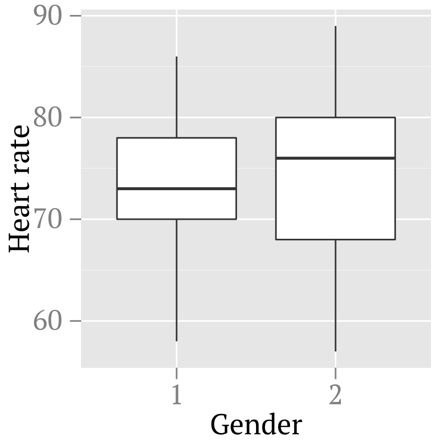

Figure 2: *Boxplot of the heart rates divided by gender.*