MAT 2379, Introduction to biostatistics

Solution to Assignment 5

Problem 10.6

[5] We compute Welch's number of degrees of freedom:

$$\begin{aligned}\nu &= \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)} \\ &= \frac{[(6.82)^2/9 + (16.75)^2/9]^2}{[(6.82)^2/9]^2/8 + [(16.75)^2/9]^2/8} \\ &= \frac{8[(6.82)^2 + (16.75)^2]^2}{(6.82)^4 + (16.75)^4} = 10.58.\end{aligned}$$

We round down this number. We obtain $\nu = 10$. From Table 18.4, we see that $t_{0.025,10} = 2.228$. The confidence interval for $\mu_1 - \mu_2$ is:

$$\bar{x}_1 - \bar{x}_2 \pm 2.228\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 99.6 - 85.3 \pm 2.228\sqrt{\frac{(6.82)^2}{9} + \frac{(16.75)^2}{9}} = [0.87; 27.73].$$

Because the interval contains only positive values, we are confident that $\mu_1 > \mu_2$. We have confident that babies whose mothers did not receive prenatal consultations have a smaller weight at birth on average.

Problem 10.10

[5] We will use a large sample test to compare the means, that is to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. The observed value of the test statistic is

$$\frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{3.5 - 2.4}{\sqrt{5.2^2/65 + 2.3^2/55}} = 1.54.$$

The p-value is $2\,P(Z > |1.54|) = 2\,P(Z > 1.54) = 2\,(1 - 0.9382) = 0.1236$. Since the p-value is larger than 5%, then we should not reject H_0. At a level of significance of $\alpha = 5\%$, we do not have evidence that the means are different.

Problem 10.16

(a) Both QQ-plots have a linear tendency, hence it is reasonable to assume that both populations are normal. Furthermore, the lines in the plots are approximately parallel, so it is reasonable to assume equality of variance.
(b) The t-test statistic is

$$-5.174 = \frac{\overline{x}_1 - \overline{x}_2}{s_p\,\sqrt{1/n_1 + 1/n_2}}.$$

Thus,

$$s_p = \frac{\overline{x}_1 - \overline{x}_2}{-5.174\,\sqrt{1/n_1 + 1/n_2}} = \frac{16.86667 - 19.09444}{-5.174\,\sqrt{1/15 + 1/18}} = 1.2316.$$

(c) A 95% confidence interval for difference between the mean recovery time on medication 1 and the mean recovery time on medication 2 is $[-3.10; -1.34]$.
(d) Since all the values in the interval are negative, we are 95% confident that the mean recovery time on medication 1 is smaller than the mean recovery time on medication 2. We are confident that medication 1 is the best on average.

Problem 11.4

[5] (a) Define the difference d as the speed on the filter without mantis excrement minus the speed on the filter with mantis excrement. We compute the $n = 15$ differences:

$$1.1,\ -0.3,\ 0.6,\ 1.1,\ 0.3,\ 0.1,\ 0.4,\ 0.6,\ -0.2,\ 0.2,\ 0.2,\ 0,\ 0.3,\ 0.8,\ 1.$$

The mean and the standard deviation of the differences are $\overline{d} = 0.4133$, and $s_d = 0.4454$. We test $H_0 : \mu_D = 0$ against $H_1 : \mu_D \neq 0$. The observed value of the test statistic is:

$$t_0 = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{0.4133 - 0}{0.4454/\sqrt{15}} = 3.59.$$

The p-value is $2\,P(T > 3.59)$, where T has a $T(14)$ distribution. Since $P(T > 3.59) < 0.005$, then p-value< 0.01. (With R, we compute p-value=0.003417.) Therefore, the speed of the spider on filters without mantis excrement is significantly different than the speed of the spider on filters with mantis excrement, on average.

[2] (b) A 95% confidence interval for μ_D is $\overline{d} \pm t\, s_d/\sqrt{n} = 0.4133 \pm (2.145)\,(0.4454)/\sqrt{15} = [0.17; 0.66]$. On average, the mantis excrement reduces the walking speed of the spider by 0.17 to 0.66 cm/s.

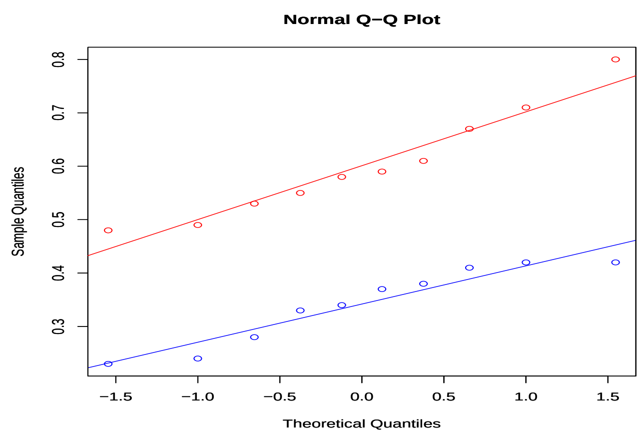
Problem 1

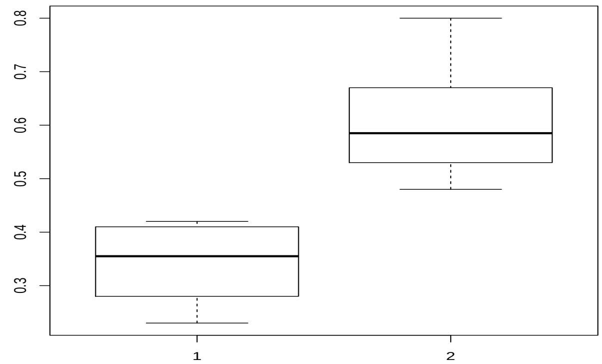
(a) By the `R` code given below, it is reasonable to assume that both populations are normal since both QQ-plots have a linear tendency. Moreover, the best fitted lines in the plots are approximately parallel, so it is reasonable to assume equality of variance.

```
x1=c(0.42, 0.24, 0.41, 0.34, 0.23, 0.28, 0.37, 0.42, 0.33, 0.38)
x2=c(0.55, 0.80, 0.61, 0.58, 0.48, 0.53, 0.67, 0.71, 0.49, 0.59)

limits=range(x1,x2) # vertical limits
qqnorm(x1,ylim=limits,col="blue") # produces the QQ-plot for x in blue
abline(mean(x1),sd(x1),col="blue") # produces the line of best-fit for x
par(new=T) # begin overlay
qqnorm(x2,ylim=limits,col="red") # produces the QQ-plot for y in red
abline(mean(x2),sd(x2),col="red") # produces the line of best fit for y
par(new=F) # end overlay

boxplot(x1,x2)
```





(b) Let μ_1 be the mean stem weight without nitrogen and let μ_2 be the mean stem weight with nitrogen. We want to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$. Using the following command in R we get

```
> t.test(x1,x2,alternative="less",var.equal=TRUE) # for equal variances

        Two Sample t-test

data:  x1 and x2
t = -6.6275, df = 18, p-value = 1.6e-06
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
       -Inf -0.1912333
sample estimates:
mean of x mean of y
    0.342     0.601
```

In fact, since the p-value=1.6×10^{-6} is less than the level of significance of $\alpha = 0.05$, then we can reject H_0. We can conclude that $\mu_2 > \mu_1$.

(b) To find a 95% confidence interval for the difference $\mu_1 - \mu_2$, we use the following command

```
> t.test(x1,x2,var.equal=TRUE)$conf.int # for equal variances
[1] -0.3411034 -0.1768966
attr(,"conf.level")
[1] 0.95
```

So that, a 95% confidence interval for the difference $\mu_1 - \mu_2$ is $[-0.341; -0.177]$. Now, since this interval contains only the negative values we conclude that the mean stem weight without nitrogen is less than the mean stem weight with nitrogen ($\mu_1 < \mu_2$).

[**/17**]