Stat 371-003, Solutions to Homework #6

1. **7.31 (pg 245)**

   The estimated standard error of the difference between the sample means is

   $$\sqrt{\frac{(8.73)^2}{5} + \frac{(7.19)^2}{5}} \approx \mathbf{5.06}$$

   (a) Our test statistic is $t = (\bar{y}_1 - \bar{y}_2)/\hat{\text{SE}}(\bar{y}_1 - \bar{y}_2) \approx (31.72 - 29.22)/5.06 \approx$ **0.494**.

   For a test at significance level $\alpha = 0.10$, we need the 95th percentile of a t distribution with 7.7 degrees of freedom. From the table at the back of the book, we get **1.86**. In R, we get `qt(0.95, 7.7)` $\approx$ 1.87.

   Since 0.494 $<$ 1.87, we **fail to reject** the null hypothesis.

   Note that we could calculate the P-value in R as `2*(1-pt(0.494, 7.7))` $\approx$ 0.64. (Using the table, we would just see that the P-value is $>$ 0.4.)

   (b) The result should not be surprising. Such data could reasonably ascribed to chance variation. If the underlying averages were the same, we would get data like this most of the time.

2. **7.36 (pg 246)**

   (a) True (we **reject H$_0$**), since the P-value = 0.07 $< \alpha = 0.10$.

   (b) True (we **fail to reject H$_0$**), since the P-value = 0.07 $> \alpha = 0.05$.

   (c) **False!** The P-value is the chance of getting data this extreme (or more so), if $\mu_1 = \mu_2$. First, it doesn't concern a statement about $\bar{y}_1$ and $\bar{y}_2$. Second, it concerns the chance of data *given* $\mu_1 = \mu_2$, not the other way around.

3. **7.44 (pg 255)**

   Yes, we would **reject H$_0$**. The 95% confidence interval is the set of plausible values for $\mu_1 - \mu_2$, and so is the set of values, δ, for which a test of $H_0 : \mu_1 - \mu_2 = \delta$ would *not* be rejected. Since 0 is not in the confidence interval, we would reject the hypothesis that $\mu_1 = \mu_2$.

4. **7.51 (pg 264–265)**

   Our estimate of the standard error of $\bar{y}_1 - \bar{y}_2$ is

   $$\sqrt{\frac{(0.621)^2}{8} + \frac{(0.652)^2}{8}} \approx \mathbf{0.318}$$

   Our test statistic is $t = (\bar{y}_1 - \bar{y}_2)/\hat{\text{SE}}(\bar{y}_1 - \bar{y}_2) \approx (6.169 - 5.291)/0.318 \approx$ **2.758**.

(a) For the two-sided test at $\alpha = 0.05$, we need the 97.5 percentile of a t distribution with 14 degrees of freedom. From the table, we find this is 2.145. (In R, we would use `qt(0.975, 14)`.)

Since 2.76 $>$ 2.145, we would **reject H**$_0$.

From the table, we see that our P-value is between 0.02 and 0.01. In R, we could calculate the p-value via `2*( 1-pt(2.758, 14) )` $\approx$ 1.5%.

(b) For the one-sided test at $\alpha = 0.05$, we need the 95th percentile of a t distribution with 14 degrees of freedom. From the table, we find this is 1.761. (In R, we would use `qt(0.95, 14)`.)

Since 2.76 $>$ 1.761, we again **reject H**$_0$.

The p-value is half of what was calculated in part (a). From the table, we see that our P-value is between 0.01 and 0.005. In R, we could calculate the p-value via `1-pt(2.758, 14)` $\approx$ 0.8%.

(c) The two-sided test (against the “nondirectional alternative”) is more appropriate, since we couldn’t anticipate the direction of the possible effect in advance. In appears that the choice of direction was made after looking at the data.

5. **9.4 (pg 356–357)**

We first load the data:

```
dat <- read.csv("http://www.biostat.wisc.edu/%7Ekbroman/teaching/stat371/data_9-4.csv")
```

We pull out the responses under treatment and control and assign them to `x` and `y`, respectively.

```
x <- dat$treated
y <- dat$control
```

(a) To get a 95% confidence interval for the true mean difference, we use `t.test()` with the differences, `x - y`.

```
t.text(x - y)
```

The 95% confidence interval is **(–0.50, 0.74)**.

(b) To get the *incorrect* 95% confidence interval for the true mean difference, ignoring the paired nature of the data, we type:

```
t.text(x, y)
```

This gives a 95% confidence interval of **(–0.83, 1.06)**.

Note that this interval is **wider** than that calculated in part (a).