# Homework - Week 9

*Solution*

Questions marked with "(OS3: X.X)" are from the textbook with "X.X" as the exercise number. The answers to the odd questions (odd by book numbering that is) will be in the back of the book.

1. (OS3: 5.3) An independent random sample is selected from an approximately normal population with an unknown standard deviation. Find the p-value for the given set of hypotheses and $T$ test statistic. Also determine if the null hypothesis would be rejected at $\alpha = 0.05$.

   a. $H_A : \mu > \mu_0$, $n = 11$, $T = 1.91$

   ```
   p.a <- pt(1.91, df=10, lower.tail=F)
   p.a
   ```

   ```
   ## [1] 0.04260244
   ```

   **p $<$ 0.05. Reject null hypothesis.**

   b. $H_A : \mu < \mu_0$, $n = 17$, $T = -3.45$

   ```
   p.b <- pt(-3.45, df=16)
   p.b
   ```

   ```
   ## [1] 0.001646786
   ```

   **p $<$ 0.05. Reject null hypothesis.**

   c. $H_A : \mu \neq \mu_0$, $n = 7$, $T = 0.83$

   ```
   p.c <- pt(0.83, df=6, lower.tail=F) * 2
   p.c
   ```

   ```
   ## [1] 0.4383084
   ```

   **p $>$ 0.05. Do not reject null hypothesis.**

   d. $H_A : \mu > \mu_0$, $n = 28$, $T = 2.13$

   ```
   p.d <- pt(2.13, df=27, lower.tail=F)
   p.d
   ```

   ```
   ## [1] 0.02121769
   ```

   **p $<$ 0.05. Reject null hypothesis.**

2. (OS3: 5.7) New York is known as "the city that never sleeps". A random sample of 25 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are shown below. Do these data provide strong evidence that New Yorkers sleep less than 8 hours a night on average?

| n | $\bar{x}$ | s | min | max |
|---|---|---|---|---|
| 25 | 7.73 | 0.77 | 6.17 | 9.78 |

   a. Write the hypotheses in symbols and in words.

   $H_0 : \mu = 8$. **The true mean hours New Yorkers sleep is 8 hrs.**

   $H_1 : \mu < 8$. **The true mean hours New Yorkers sleep is less than 8 hrs.**

   b. Check conditions, then calculate the test statistic, $T$, and the associated degrees of freedom.

```r
# t = (x bar - 8) / (sd / sqrt(n))
t <- (7.73 - 8) / (0.77 / sqrt(25))
t
```

```
## [1] -1.753247
```

```r
df <- 25 -1
df
```

```
## [1] 24
```

c. Find and interpret the p-value in this context. Drawing a picture may be helpful.

```r
p.sleep <- pt(t, df=df)
p.sleep
```

```
## [1] 0.04616261
```

d. What is the conclusion of the hypothesis test?

**p < 0.05. Reject null hypothesis. There is evidence that New Yorkers sleep less than 8 hours per night on average.**

e. If you were to construct a 90% confidence interval that corresponded to this hypothesis test, would you expect 8 hours to be in the interval?

**The one-sided p-value is less than 0.05, so we would not expect 8 hours to be in a 90% confidence interval.**

3. (OS3: 5.19) Is there strong evidence of global warming? Let's consider a small scale example, comparing how temperatures have changed in the US from 1968 to 2008. The daily high temperature reading on January 1 was collected in 1968 and 2008 for 51 randomly selected locations in the continental US. Then the difference between the two readings (temperature in 2008 - temperature in 1968) was calculated for each of the 51 different locations. The average of these 51 values was 1.1 degrees with a standard deviation of 4.9 degrees. We are interested in determining whether these data provide strong evidence of temperature warming in the continental US.

a. Is there a relationship between the observations collected in 1968 and 2008? Or are the observations in the two groups independent? Explain.

**Because the observations are for the same locations at two different points in time, the observations are related. They are not independent.**

b. Write hypotheses for this research in symbols and in words.

$H_0 : \mu_d = 0$. **There is no change in temperatures. Or, the mean difference in temperatures is zero.**

$H_a : \mu_d > 0$. **Temperatures have increased. Or, the mean difference in temperatures is greater than zero.**

c. Check the conditions required to complete this test.

d. Calculate the test statistic and find the p-value.

```r
t.temp <- (1.1) / (4.9/sqrt(51))
t.temp
```

```
## [1] 1.603178
```

```r
p.temp <- pt(t.temp, df=50, lower.tail=F)
p.temp
```

```
## [1] 0.05759731
```

e. What do you conclude? Interpret your conclusion in context.

   **p = 0.0576 > 0.05. Fail to reject null hypothesis. There is not evidence that temperatures have increased.**

f. What type of error might we have made? Explain in context what the error means.

   **Since we failed to reject the null hypothesis, it is possible we have a false negative. That is, it is possible that temperatures have increased but we did not detect it. This is a type II error.**

g. Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the temperature measurements from 1968 and 2008 to include 0? Explain your reasoning.

   **Because we failed to reject the null, we would expect the appropriate confidence interval to contain zero.**

4. (OS3: 5.28) Prices of diamonds are determined by what is known as the 4 Cs: cut, clarity, color, and carat weight. The prices of diamonds go up as the carat weight increases, but the increase is not smooth. For example, the difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but the price of a 1 carat diamond tends to be much higher than the price of a 0.99 diamond. In this question we use two random samples of diamonds, 0.99 carats and 1 carat, each sample of size 23, and compare the average prices of the diamonds. In order to be able to compare equivalent units, we first divide the price for each diamond by 100 times its weight in carats. That is, for a 0.99 carat diamond, we divide the price by 99. For a 1 carat diamond, we divide the price by 100. The distributions and some sample statistics are shown below.

   Conduct a hypothesis test to evaluate if there is a difference between the average standardized prices of 0.99 and 1 carat diamonds. Make sure to state your hypotheses clearly, check relevant conditions, and interpret your results in context of the data.

| | 0.99 carats | 1 carat |
|---|---|---|
| Mean | $ 44.51 | $ 56.81 |
| SD | $ 13.32 | $ 16.13 |
| n | 23 | 23 |

   $H_0 : \mu_{0.99} = \mu_1$

   $H_a : \mu_{0.99} \neq \mu_1$

```
t.diam <- (44.51 - 56.58) / (sqrt(13.32^2/23 + 16.13^2/23))
t.diam
```

```
## [1] -2.767151
```

```
p.diam <- pt(t.diam, df=22) * 2
p.diam
```

```
## [1] 0.01124216
```

   **p < 0.05. Reject null hypothesis. There is evidence that the price per carat is different between 0.99 carat diamonds and 1 carat diamonds.**

5. Dataset `smoking.csv` on D2L contains the results of a clinical trial testing a drug to reduce smoking. The trial was a double-blinded randomized controlled trial with a control group which received a placebo (a fake version of the drug) and two control groups ("A" and "B"). The data represent the cigarettes per day smoked at the end of the trial for each participant.

   For each scenario below, state the null and alternative hypothesis, conduct the test, report a decision on the null hypothesis and state your conclusion in the context of the research question.

```r
smoke <- read.csv("../../Data/smoking.csv")
str(smoke)
```

```
## 'data.frame':    48 obs. of  3 variables:
##  $ control: int  17 10 12 8 12 17 12 19 12 17 ...
##  $ group.A: int  11 10 16 12 11 16 11 6 7 13 ...
##  $ group.B: int  10 7 18 9 10 17 10 6 4 15 ...
```

a. The mean cigarettes per day smoked by participants before the trial started was 14.6. Use the control group to test for a placebo effect. That is, test whether the control group has a mean cigarettes per day different from 14.6.

   $H_0 : \mu = 14.6$

   $H_a : \mu \neq 14.6$

   ```r
   t.test(smoke$control, mu=14.6)
   ```

   ```
   ##
   ##  One Sample t-test
   ##
   ## data:  smoke$control
   ## t = -2.2433, df = 47, p-value = 0.02963
   ## alternative hypothesis: true mean is not equal to 14.6
   ## 95 percent confidence interval:
   ##  12.39501 14.47999
   ## sample estimates:
   ## mean of x
   ##   13.4375
   ```

   **p $<$ 0.05. Reject null hypothesis. There is evidence that the control group has a mean cigarettes per day different from 14.6.**

b. Use the control group and group A to test whether the drug was effective. That is, does group A smoke less cigarettes per day than the control group?

   $H_0 : \mu_a = \mu_c$

   $H_a : \mu_a < \mu_c$

   ```r
   t.test(smoke$group.A, smoke$control, alternative="less")
   ```

   ```
   ##
   ##  Welch Two Sample t-test
   ##
   ## data:  smoke$group.A and smoke$control
   ## t = -3.0748, df = 93.955, p-value = 0.001379
   ## alternative hypothesis: true difference in means is less than 0
   ## 95 percent confidence interval:
   ##       -Inf -1.024813
   ## sample estimates:
   ## mean of x mean of y
   ##  11.20833  13.43750
   ```

   **p $<$ 0.05. Reject null hypothesis. There is evidence that group A smoke less cigarettes per day than the control group.**

c. Assume groups A and B are two separate groups of subjects. Test whether there a difference between the goups.

$H_0 : \mu_a = \mu_b$

$H_a : \mu_a \neq \mu_b$

```
t.test(smoke$group.A, smoke$group.B)
```

```
##
##  Welch Two Sample t-test
##
## data:  smoke$group.A and smoke$group.B
## t = 1.5603, df = 90.422, p-value = 0.1222
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3414348  2.8414348
## sample estimates:
## mean of x mean of y
## 11.208333  9.958333
```

**p $>$ 0.05. Do not reject null hypothesis. There is not evidence that group A smoke a diffenent number cigarettes per day than group B.**

d. Assume groups A and B are the same group of subjects measured at two different points in time. Test whether there is a difference in smoking over time.

$H_0 : \mu_d = 0$

$H_a : \mu_d \neq 0$

```
t.test(smoke$group.A, smoke$group.B, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  smoke$group.A and smoke$group.B
## t = 3.5933, df = 47, p-value = 0.0007792
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5501842 1.9498158
## sample estimates:
## mean of the differences
##                    1.25
```

**p $<$ 0.05. Reject null hypothesis. There is evidence that the number cigarettes per day smoked changed over time.**