STAT 515 hw 9

Two-sample testing for differences in means and proportions

1. Two methods of measuring out one cup of flour from a large container of flour were compared. Under the *scoop* method no instruction was given on how to measure the flour, but a 1 cup measuring cup was provided. Under the *spoon* method, instructions were given to fluff up the flour in the large container with a spoon, then to scoop the fluffed-up flour into a 1 cup measuring cup, and then to scrap off the excess flour with a provided flat implement. The measurements were taken by students in a statistics class at different times in the semester. The weights in grams of the measured out cups of flour under the two methods are tabulated below:

Scoop			Spoon		
163	157	163	128	151	125
160	170	169	127	128	117
150	150	145	130	129	127
166	175	110	154	162	130
168	182	162	154	131	140
153	141	115			
152	122	146			

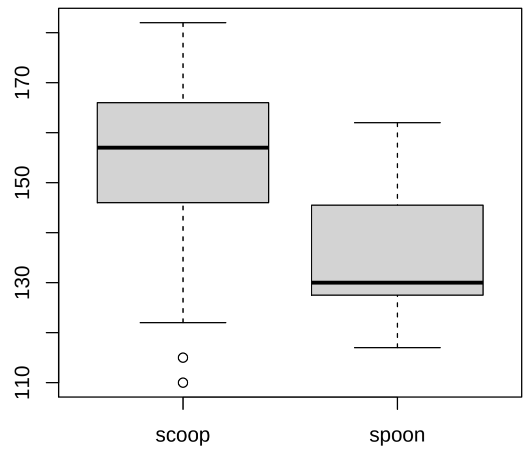
Serious bakers encourage measuring out flour by weighing it instead of by scooping it in a measuring cup, since flour is compacted by scooping. In the absence of a kitchen scale, some bakers recommend the *spoon* method described above. You will use the data to decide whether the *spoon* method results in smaller measurements of flour than the *scoop* method (assume throughout that the data follow Normal distributions):

(a) State the null and alternate hypotheses of interest (let μ_1 be the mean weight in grams of cups of flour measured out by the *scoop* method).

> We are interested in whether we can conclude $\mu_2 < \mu_1$, so the hypotheses of interest are
>
> $$H_0\colon\ \mu_1 - \mu_2 \leq 0 \text{ versus } H_1\colon\ \mu_1 - \mu_2 > 0.$$

(b) Make side-by-side boxplots of the weights of the measured out cups of flour. Make sure it is labeled so that someone can tell which plot belongs to which method. Access the documentation for the `boxplot` function by executing `?boxplot`.

> The code
>
> ```
> grams_spoon <- c(128,151,125,127,128,117,130,129,127,154,162,130,154,131,140)
> grams_scoop <- c(163,157,163,160,170,169,150,
>                  150,145,166,175,110,168,182,162,153,141,115,152,122,146)
> boxplot(grams_scoop,grams_spoon,names=c("scoop","spoon"))
> ```
>
> produces this figure:



(c) Based on your boxplots in part (b), do you recommend using the equal-variances or the unequal-variances version of the two-sample t-test?

Disregarding a couple of outliers in the *scoop* measurements, it appears that the spreads of the two boxplots are similar. We could probably safely use the equal-variances two-sample t-test.

(d) Give the value of the test statistic for the equal-variances two-sample t-test of the hypotheses in part (a).

We have

$$\bar{X}_1 = 153.2857, \quad S_1^2 = 356.9143, \quad n_1 = 21$$
$$\bar{X}_2 = 135.5333, \quad S_2^2 = 177.1238, \quad n_2 = 15,$$

based on which we have

$$S^2_{\text{pooled}} = \frac{(21-1)356.9143 + (15-1)177.1238}{21+15-2} = 282.8829.$$

The test statistic for the equal-variances two-sample t-test is equal to

$$T_{\text{test}} = \frac{153.2857 - 135.5333}{\sqrt{282.8829(1/21 + 1/15)}} = 3.122175.$$

(e) Give the p-value based on the data for testing the hypotheses in part (a).

The p-value is given by the area under the pdf of the $t_{21+15-2}$ distribution to the right of the value 3.122175. We can retrieve this with

$$\texttt{1 - pt(3.122175,21+15-2)} = 0.001826796.$$

(f) What is the smallest significance level α at which you would reject the null hypothesis in part (a) based on these data?

This is exactly the p-value. For all values of α greater than 0.001826796, we would reject H_0.

(g) Give the p-value of the equal-variances t-test of

$$H_0\colon\ \mu_1 - \mu_2 = 0 \text{ versus } H_1\colon\ \mu_1 - \mu_2 \neq 0$$

based on these data.

This is just two times the p-value for the one-sided hypothesis test. This is equal to

$$2(0.001826796) = 0.003653592.$$

(h) Construct a 95% confidence interval for $\mu_1 - \mu_2$, assuming equal variances.

A 95% confidence interval for $\mu_1 - \mu_2$ is

$$153.2857 - 135.5333 \pm \underbrace{t_{21-15-2,0.05/2}}_{2.032245}\sqrt{282.8829(1/21 + 1/15)} = (6.197239, 29.30752).$$

(i) Give an interpretation of the confidence interval in part (h).

With 95% confidence, the expected weight of a cup of flour measured with the *scoop* method is between 6.2 and 29.3 grams greater than if measured with the *spoon* method.

(j) Use R to obtain the p-value of the unequal-variances two-sample t-test of the one-sided hypotheses from part (a). Comment on whether there is a large difference.

The coommand

```
t.test(grams_scoop,grams_spoon,alternative = "greater", var.equal = FALSE)
```

gives the output

```
        Welch Two Sample t-test

data:  grams_scoop and grams_spoon
t = 3.3077, df = 34, p-value = 0.001115
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 8.677268      Inf
sample estimates:
mean of x mean of y 
 153.2857  135.5333 
```

from which we see that the p-value is 0.001115. The p-value is very close to that of the equal-variances version of the test.

(k) You want to post a cookie recipe on your blog, and you want the directions to be as simple as possible, so you are reluctant to include instructions to fluff and spoon the flour. On the other hand, your recipe is very sensitive to poor flour measurements. You decide you will only include instructions to fluff and spoon the flour if you can conclude that the *spoon* method reduces the mean weight of a measured cup of flour by at least 10 grams.

i. Give the null and alternate hypothesis of interest.

The hypotheses of interest are

$$H_0\colon \mu_1 - \mu_2 \leq 10 \text{ versus } H_1 : \mu_1 - \mu_2 > 10.$$

ii. Obtain the p-value using the equal-variances two-sample t-test.

The test statistic is

$$T_{\text{test}} = T_{\text{test}} = \frac{(153.2857 - 135.5333) - 10}{\sqrt{282.8829(1/21 + 1/15)}} = 1.363439.$$

The p-value is

`1 - pt(1.363439,21+15-2)` $= 0.09085238.$

iii. What do you decide? Will you include instructions to fluff and spoon the flour in your recipe? Consider the significance levels $\alpha = 0.10$ and $\alpha = 0.05$.

At $\alpha = 0.10$, we would reject the null hypothesis and choose to include the instructions to fluff and spoon the flour in the recipe. At $\alpha = 0.05$, however, we would fail to reject the null hypothesis and decide not to include the instructions to fluff and spoon the flour.

2. Members of the 3D book club, whose names are of course aliases, are discussing whether to read

Bleak House by Charles Dickens:

1. Daragh will vote to read it if more than 10% of the sentences have at least 40 words.
2. Deaglan will vote to read it if fewer than 10% of the sentences have at least 40 words.
3. Deidre will vote to read it as long as the percentage of sentences with at least 40 words is about 10%.

They randomly sample 200 sentences from Bleak House and count 21 with at least 40 words.

(a) Write down the hypotheses each book club member wishes to test.

1. Daragh wishes to test H_0: $p \leq 0.10$ versus H_1: $p > 0.10$.
2. Deaglan wishes to test H_0: $p \geq 0.10$ versus H_1: $p < 0.10$.
3. Deidre wishes to test H_0: $p = 0.10$ versus H_1:$p \neq 0.10$.

(b) Give the p-value based the data for testing the hypotheses of each book club member.

The test statistic for testing the book club members' hypotheses about p is

$$Z_{\text{test}} = \frac{21/200 - 0.10}{\sqrt{0.10(1 - 0.10)/200}} = 0.236.$$

1. The p-value for Daragh's hypotheses, H_0: $p \leq 0.10$ versus H_1: $p > 0.10$, is

$$P(Z > 0.236) = \texttt{1 - pnorm(0.236)} = 0.4068319.$$

2. The p-value for Deaglan's hypotheses, H_0: $p \leq 0.10$ versus H_1: $p > 0.10$, is

$$P(Z < 0.236) = \texttt{pnorm(0.236)} = 0.5931681.$$

3. The p-value for Deirdre's hypotheses, H_0: $p \leq 0.10$ versus H_1: $p > 0.10$, is

$$2 \cdot P(Z > |0.236|) = \texttt{2*(1 - pnorm(0.236))} = 0.8134327.$$

(c) Who, if any, of the book club members will vote to read the book?

It seems Deirdre will vote to read the book, since there is little evidence to conclude that the true percentage of sentences which have at least 40 words differs from 10%. Daragh and Deaglan did not find grounds in the data to reject their null hypotheses, so they will not vote for the book.

(d) You conduct a text analysis of Bleak House, counting 20,488 sentences in the book, the lengths of which you summarize in the stem-and-leaf plot below, in which the decimal place is one digit to the right of the vertical line:

```
 0 | 11111111111111111111111111111111111111+8415
 1 | 00000000000000000000000000000000000000+5244
 2 | 00000000000000000000000000000000000000+3064
 3 | 00000000000000000000000000000000000000+1694
 4 | 00000000000000000000000000000000000000+911
 5 | 00000000000000000000000000000000000000+465
 6 | 00000000000000000000000111111111111111+176
 7 | 00000000000000011111112222222222223333+87
 8 | 00000000000111111111112222333333334445+18
 9 | 000011122233333555566667777899999
10 | 0112333444679
11 | 001134466777889
12 | 5
13 | 2335
14 | 8
15 | 24
16 |
17 | 6
18 |
19 | 7
20 |
21 | 0
```

Based on the stem-and-leaf plot, what is the true proportion of sentences in Bleak House which have at least 40 words? *To make sure you are reading the stem-and-leaf plot correctly: It shows that there are two sentences which have* 133 *words, four sentences which have* 90 *words, one sentence which has* 210 *words, and* 8,453 *sentences which have* 9 *or fewer words, for example.*

By carefully adding up counts from the stem-and-leaf plot, we see that there are

$$38 * 5 + 911 + 465 + 176 + 87 + 18 + 33 + 13 + 15 + 1 + 4 + 1 + 2 + 1 + 1 + 1 = 1919$$

sentences in "Bleak House" with at least 40 words. So the true proportion is

$$p = \frac{1919}{20488} = 0.09366458.$$

(e) Which, if any, of the book club members

i. committed a Type I error?

Since none of the book club members rejected his or her null hypothesis, none of them committed a Type I error.

ii. committed a Type II error?

For Deaglan and Deidre, the null hypothesis is false; yet each of them fail to reject the null. Therefore, they each committed a Type II error. In the case of Deirdre, however, the Type II error does not seem to be a grave error, since she will be content with a proportion of "long sentences" around 0.10, and the true proportion is very close to this.

(f) Suppose the book club members had, in their sample of 200 sentences, counted 16 with at least 40 words. Compared to the p-value previously computed, for which book club members, if any, would the p-value

i. increase?

For Daragh the p-value would increase, as these data would contain weaker evidence against the null hypothesis (actually stronger evidence *in favor* of the null hypothesis).

ii. decrease?

For Deaglan and Deirdre the p-value would decrease, as these data would contain stronger evidence against the null.

3. The book club members from Question 2 decide that they must read either Bleak House by Charles Dickens or Emma by Jane Austen in order to consider themselves a proper literary triumvirate.

1. Daragh will vote for Emma if it has a greater proportion of sentences with at least 40 words than Bleak House; otherwise he will vote for Bleak House.
2. Deaglan will vote for Emma if it has a smaller proportion of sentences with at least 40 words than Bleak House; otherwise he will vote for Bleak House.
3. Deirdre will vote for Bleak House if the proportion of its sentences with at least 40 words is no different from that of the sentences in Emma; otherwise she will vote for Emma.

They randomly sample 150 sentences from Emma; 22 of these have at least 40 words.

(a) State the null and alternate hypotheses of interest to each member of the 3D book club (formulate them in terms of $p_1 - p_2$, where Bleak House is 1 and Emma is 2).

1. Daragh is interested in the hypotheses

$$H_0 : p_1 - p_2 \geq 0 \text{ versus } H_1\colon p_1 - p_2 < 0.$$

2. Deaglan is interested in the hypotheses

$$H_0 : p_1 - p_2 \leq 0 \text{ versus } H_1\colon p_1 - p_2 > 0.$$

3. Deirdre is interested in the hypotheses

$$H_0 : p_1 - p_2 = 0 \text{ versus } H_1\colon\ p_1 - p_2 \neq 0.$$

(b) Using the data from Question 2, give the value of the test statistic for testing the hypotheses of interest to the book club members.

To construct the test statistic, we first need the value

$$\hat{p}_0 = \frac{200(21/200) + 150(22/150)}{200 + 150} = \frac{21 + 22}{350} = 0.1228571.$$

Then we obtain test statistic value

$$Z_{\text{test}} = \frac{21/200 - 22/150}{\sqrt{0.123(1 - 0.123)(1/200 + 1/150)}} = -1.175.$$

(c) For each book club member, give the p-value based on the data for testing his or her hypotheses.

1. For Daragh's hypotheses the p-value is

$$P(Z < -1.175) = \texttt{pnorm(-1.175)} = 0.120092.$$

2. For Deaglan's hypotheses the p-value is

$$P(Z > -1.175) = \texttt{1 - pnorm(-1.175)} = 0.879908.$$

3. for Deirdre's hypotheses the p-value is

$$2 \cdot P(Z > |1.175|) = \texttt{2*(1 - pnorm(abs(-1.175)))} = 2(0.120092) = 0.240184.$$

(d) State which book you think each member will vote for and explain why. In your explanation, give the significance level α you are assuming.

If we choose any significance level smaller than $\alpha \geq 0.120092$, which we usually do, then no book club members will reject their null hypotheses. So:

1. Since Daragh cannot conclude that Emma has a greater proportion of "long sentences" than Bleak House, he will vote for Bleak House.

2. Since Daragh cannot conclude that Emma has a lesser proportion of "long sentences" than Bleak House, he will vote for Bleak House.

3. It seems that Deirdre will also vote for Bleak House, since she cannot conclude that the proportion of "long sentences" in Emma is any different from that in Bleak House.

(e) Construct a 95% confidence interval for the difference between the proportions of sentences with at least 40 words between Bleak House and Emma (Bleak House minus Emma).

A 95% confidence interval for $p_1 - p_2$ is given by

$$21/200 - 22/150 - 1.96\sqrt{\frac{21/200(1-21/200)}{200} + \frac{22/150(1-22/150)}{150}} = (-0.112, 0.029).$$

(f) You conduct a text analysis of Emma and build the following stem-and-leaf plot of the lengths of all 7,319 sentences in Emma (the decimal point lies one digit to the right of the vertical line):

```
 0 | 11111111111111111111111111111111111111+2225
 1 | 00000000000000000000000000000000000000+2146
 2 | 00000000000000000000000000000000000000+1117
 3 | 00000000000000000000000000000000000000+630
 4 | 00000000000000000000000000000000000000+366
 5 | 00000000000000000000000000000011111111+188
 6 | 00000000000000000000000000111111111111+130
 7 | 00000000000111111111112222222222222233+42
 8 | 00000011111111222233444444555566666666+10
 9 | 00000122222223333444445555555668899
10 | 0111122333334466779999
11 | 002334455556778999
12 | 001123455567999
13 | 0048
14 | 0003344
15 | 599
16 | 014448
17 | 089
18 | 14667
19 | 9
20 |
21 | 2
22 | 0
23 | 19
```

Based on the stem-and-leaf plot, what is the true proportion of sentences in Emma which have at least 40 words?

By carefully adding up counts from the stem-and-leaf plot, we see that there are

$$38*5+366+188+130+42+10+35+22+18+15+4+7+3+6+3+5+1+1+1+2 = 1049$$

sentences in Emma with at least 40 words. So the true proportion is

$$p_2 = \frac{1049}{7319} = 0.1433256.$$

(g) Which, if any, of the book club members

i. committed a Type I error?

None of the book club members was able to reject his or her null hypothesis, so none of them could have committed a Type I error.

ii. committed a Type II error?

The alternate hypotheses of both Daragh and Deirdre were true, since the true difference in the proportions of "long sentences" was

$$p_1 - p_2 = 0.09366458 - 0.1433256 = -0.04966101.$$

Therefore both Daragh and Deirdre committed Type II errors. Deaglan's null hypothesis was true, and he failed to reject it, so he has made a correct statistical inference.

(h) Did the 95% confidence interval in part (e) contain the true difference in proportions?

Yes, the 95% confidence interval contained the value $p_1 - p_2 = -0.04966101$.

Optional (do not turn in) problems for additional study from McClave, J.T. and Sincich T. (2017) *Statistics*, 13th Edition: 8.76, 8.80, 8.129, 8.133, 8.144, 9.12, 9.18, 9.24, 9.26, 9.58, 9.68