

STAT 400 Homework 12

Spring 2018 / Dalpiaz / UIUC

Due: Wednesday, May 2, 2:00 PM

Exercise 1

Professor Professorson, a researcher at Greendale Community College, is interested in the effect of caffeine on the typing speed of students. Professorson obtains a random sample of 8 students who are given 400 mg of caffeine then given a typing test. They type an average of 51.4 words per minute (wpm), with a sample standard deviation of 12.3 wpm. He also obtains a random sample of 13 students who are given a placebo before the typing test. The placebo group types an average of 43.9 wpm, with a sample standard deviation of 15.1 wpm. Assume typing speeds follow a normal distribution in both groups.

(a) Construct a 99% confidence interval for $\mu_C - \mu_P$, the true difference in average typing speed between the caffeine and placebo groups. Assume that the two population variances are equal.

Solution:

We will use X for the caffeine (treatment) group.

- $\bar{x} = 51.4$
- $s_x = 12.3$
- $n_x = 8$

We will use Y for the placebo (control) group.

- $\bar{y} = 43.9$
- $s_y = 15.1$
- $n_y = 13$

The confidence interval is given by

$$\bar{x} - \bar{y} \pm t_{\alpha/2}(n_x + n_y - 2)s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

Here, the degrees of freedom for the t critical value is

$$n_x + n_y - 2 = 8 + 13 - 2 = 19.$$

With $\alpha = 0.01$, the critical value is then

$$t_{\alpha/2}(n_x + n_y - 2) = t_{0.005}(19) = 2.861.$$

The pooled standard deviation estimate is

$$s_p = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} = \sqrt{\frac{(8 - 1)(12.3)^2 + (13 - 1)(15.1)^2}{8 + 13 - 2}} = 14.13$$

Thus plugging everything in, we obtain

$$51.4 - 43.9 \pm 2.861 \cdot 14.13 \cdot \sqrt{\frac{1}{8} + \frac{1}{13}}$$

$$\boxed{7.5 \pm 18.17}$$

(b) Construct a 99% confidence interval for $\mu_C - \mu_P$, the true difference in average typing speed between the caffeine and placebo groups. *Do not assume that the two population variances are equal.* (Use Welch's T.)

Solution:

The confidence interval is given by

$$\bar{x} - \bar{y} \pm t_{\alpha/2}(df) \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

Here, the degrees of freedom for the t critical value is

$$df = \left\lfloor \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{1}{n_x-1} \left(\frac{s_x^2}{n_x}\right)^2 + \frac{1}{n_y-1} \left(\frac{s_y^2}{n_y}\right)^2} \right\rfloor = \left\lfloor \frac{\left(\frac{12.3^2}{8} + \frac{15.1^2}{13}\right)^2}{\frac{1}{8-1} \left(\frac{12.3^2}{8}\right)^2 + \frac{1}{13-1} \left(\frac{15.1^2}{13}\right)^2} \right\rfloor = \lfloor 17.31 \rfloor = 17$$

With $\alpha = 0.01$, the critical value is then

$$t_{\alpha/2}(df) = t_{0.005}(17) = 2.898.$$

Thus plugging everything in, we obtain

$$51.4 - 43.9 \pm 2.898 \cdot \sqrt{\frac{12.3^2}{8} + \frac{15.1^2}{13}}$$

$$\boxed{7.5 \pm 17.50}$$

(c) Calculate the the value of the test statistic for testing $H_0 : \mu_C = \mu_P$ versus $H_1 : \mu_C \neq \mu_P$. *Assume that the two population variances are equal.*

Solution:

The test statistic is given by

$$t = \frac{(\bar{x} - \bar{y}) - 0}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{(51.4 - 43.9) - 0}{14.13 \cdot \sqrt{\frac{1}{8} + \frac{1}{13}}} = \boxed{1.181}.$$

(d) State the critical value(s) for the test above test, and your statistical decision, using $\alpha = 0.05$. *Assume that the two population variances are equal.*

Solution:

Here, the degrees of freedom for the t critical value is

$$n_x + n_y - 2 = 8 + 13 - 2 = 19.$$

With $\alpha = 0.05$, the critical values are

$$t_{\alpha/2}(n_x + n_y - 2) = t_{0.025}(19) = \boxed{2.093}$$

$$-t_{\alpha/2}(n_x + n_y - 2) = -t_{0.025}(19) = \boxed{-2.093}$$

Since

$$1.181 < 2.093$$

we **Fail to Reject** the null hypothesis.

Exercise 2

Suppose we randomly sample 400 nucleotides from the human genome. (Consider only a single strand.) Does the occurrence of **adenine**, **thiamine**, **cytosine** and **guanine** (A, C, T and G) in this sample suggest that nucleotides in the human genome follow a uniform distribution?

Use a χ^2 test with a significance level of $\alpha = 0.01$ to test the following hypotheses:

- $H_0 : p_A = p_C = p_T = p_G = 0.25$. (p_A is the proportion of adenine in the human genome.)
- H_A : At least one of p_A, p_C, p_T, p_G is not 0.25.

The data obtained was:

	A	C	T	G
Count	90	110	85	115

(a) Calculate the value of the appropriate test statistic.

Solution:

We first calculate the table of expected counts.

	A	C	T	G
Expected	100	100	100	100

We then calculate the test statistic.

$$X^2 = \frac{(90 - 100)^2}{100} + \frac{(110 - 100)^2}{100} + \frac{(85 - 100)^2}{100} + \frac{(115 - 100)^2}{100} = \boxed{6.5}$$

```
observed = c(90, 110, 85, 115)
expected = rep(100, times = 4)
sum(((observed - expected) ^ 2) / expected )
```

```
## [1] 6.5
```

(b) State the critical value for this test, and your statistical decision.

Solution:

Under the null hypothesis, the test statistic follows a chi-square distribution with $k - 1$ degrees of freedom. In this case,

$$X^2 \sim \chi^2(3)$$

With $\alpha = 0.01$, we have

$$\chi_{0.01}^2(3) = \boxed{11.34}$$

Since

$$6.5 < 11.34$$

we **Fail to Reject** the null hypothesis. (Also note, this is a valid result, as each of the expected counts is at least 5.)

```
qchisq(0.01, df = 3, lower.tail = FALSE)
```

```
## [1] 11.34487
```

Note: Even if A, C, T and G follow a uniform distribution in the human genome, that does **not** mean that nucleotides occur randomly! Even with a uniform distribution, certain patterns can (and do) occur.

Exercise 3

At Anytown College, the administration would like the students' grade distribution to be 25% A's, 30% B's, 25% C's, and 20% D's. The school's president thinks that the instructors may not be following this guideline, so he takes a random sample of grades to check his suspicion. A random sample of 30 grades yields 9 A's, 15 B's, 3 C's, and 3 D's. The president wishes to test

$$H_0 : p_A = 0.25, p_B = 0.30, p_C = 0.25, p_D = 0.20$$

(a) Calculate the value of the appropriate test statistic.

	A	B	C	D
Observed	9	15	3	3
Expected	$30 \cdot 0.25 = 7.5$	$30 \cdot 0.30 = 9$	$30 \cdot 0.25 = 7.5$	$30 \cdot 0.20 = 6$
$\frac{(O-E)^2}{E}$	$\frac{(9-7.5)^2}{7.5} = 0.3$	$\frac{(15-9)^2}{9} = 4.0$	$\frac{(3-7.5)^2}{7.5} = 2.7$	$\frac{(3-6)^2}{6} = 1.5$

$$X^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E} = 0.3 + 4.0 + 2.7 + 1.5 = \boxed{8.5}$$

(b) State the critical value for this test, and your statistical decision using a 5% significance level.

$$df = k - 1 = 4 - 1 = 3$$

$$\chi_{0.05}^2(3) = \boxed{7.815}$$

$$X^2 = 8.5 > \chi_{0.05}^2(3) = 7.815$$

Reject H_0

Exercise 4

In a random sample of 1,000 voters in Neverland, each individual was asked to name the issue that is most important to them in the upcoming presidential election. The individuals were also classified by party affiliation. The results were as follows:

Party Affiliation	Most Important Issue				Total
	Economy	Terrorism	Wealth Distribution	Climate Change	
Democrat	193	135	112	60	500
Republican	176	129	64	31	400
Independent	31	36	24	9	100
Total	400	300	200	100	1000

We wish to test whether political party affiliation and the most important issue are independent.

(a) Calculate the value of the appropriate test statistic.

Solution:

Party Affiliation	Most Important Issue				Total
	Economy	Terrorism	Wealth Distribution	Climate Change	
Democrat	193 200 <i>0.245</i>	135 150 <i>1.5</i>	112 100 <i>1.44</i>	60 50 <i>2</i>	500
Republican	176 160 <i>1.6</i>	129 120 <i>0.675</i>	64 80 <i>3.2</i>	31 40 <i>2.025</i>	400
Independent	31 40 <i>2.025</i>	36 30 <i>1.2</i>	24 20 <i>0.8</i>	9 10 <i>0.1</i>	100
Total	400	300	200	100	1000

$$\begin{array}{c} O \\ E \\ \frac{(O - E)^2}{E} \end{array}$$

$$X^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E} = \boxed{16.81}$$

(b) Calculate the p-value (approximately) of the appropriate test statistic, and your statistical decision using a 5% significance level.

Solution:

$$df = (r - 1)(c - 1) = (3 - 1)(4 - 1) = 6$$

$$\text{p-value} = P(\chi^2(6) > 16.81) = \boxed{0.01}$$

$\boxed{\text{Reject } H_0}$