Section 4.2 Supplement

Day 6 (1/24)

1 Dependent Means and Proportions

1.1 Paired data

Let $(X_1, Y_1), \ldots (X_n, Y_n)$ be a random sample of paired observations, where $(X_1, \ldots, X_n)$ is a random sample from a distribution $f_1(\cdot)$ with mean μ_1 and $(Y_1, \ldots, Y_n)$ is a random sample from a distribution $f_2(\cdot)$ with mean μ_2. Then the differences $(D_1, \ldots, D_n) = Y_1 - X_1, \ldots, Y_n - X_n$ constitute a random sample from a distribution $g(\cdot)$ with mean $\mu_2 - \mu_1$ (call this μ_d).

To test $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_i \neq \mu_2$, we instead test

$$H_0 : \mu_d = 0 \text{ versus } H_1 : \mu_d \neq 0$$

which is a one-sample problem on the differences $D_1, \ldots, D_n$. Consequently, we can construct a 95% confidence interval for μ_d as follows:

$$\overline{D} \pm t_{.025,n-1} \frac{S_d}{\sqrt{n}}$$

A test with level of significance $\alpha = .05$ will reject $H_0 : \mu_d = 0$ if

$$\left| \frac{\overline{D}}{S_d/\sqrt{n}} \right| \geq t_{.025,n-1}$$

1.2 Paired binary data

Example: BMR data from handout

		Week 2	
		N	Ab
Week 0	N	5	4
	Ab	1	0

In generic, suppose we express the data as

		Post	
		N	Ab
Pre	N	a	b
	Ab	c	d

<u>McNemar's test</u>
To test

$$H_0 : p_1 = p + 2 \text{ versus } H_1 : p_1 \neq p_2$$

where p_1 and p_2 are the marginal probabilities (of "Normal", say), reject H_0 if

$$M = \frac{(b-c)^2}{b+c} > \chi^2_{.05,1}$$

BMR example:

Since $M = \frac{(4-1)^2}{4+1} = \frac{9}{5} = 1.8$ which is not greater than $\chi^2_{.05,1} = 3.84$, then we do not reject the null hypothesis. The percentage of normal (or abnormal) is not significantly different between Week 0 and Week 2.

1.3 Dependent proportions

When two proportions $\hat{p}_1$ and $\hat{p}_2$ are two categories of a multinomial, then the two proportions are not independent. In fact, the sum $\hat{p}_1 + \hat{p}_2$ cannot exceed 1.0, so when one exceeds .5, then the other cannot.

In this case, the estimator of $p_1 - p_2 = \hat{p}_1 - \hat{p}_2$ is the same as before, but the variance formula is different.

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} + \frac{2\hat{p}_1\hat{p}_2}{n}$$

Example:

Suppose that $n = 50$ people were asked whether they were optimistic about the economy. The data is shown below. Are there significant more people who said 'yes' than who said 'no'?

Yes	No	Not sure
22	15	13

Solution:

$$\hat{p}_1 - \hat{p}_2 = \frac{22}{50} - \frac{15}{50} = .44 - .30 = .14$$

$$\text{Var}(\hat{p}_1 - \hat{p}_2) == \frac{(.44)(.56)}{50} + \frac{(.30)(.70)}{50} + \frac{(.44)(.30)}{50} = .0118$$

The standard error of $\hat{p}_1 - \hat{p}_2 = .14$ is

$$\text{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{.0118} = .11$$

The 95% confidence interval of the difference is $.14 \pm 1.96(.11)$ or

$$(-.08, .25)$$