Section 4.2 (con't.)

01/22/2019

1 The Behrens-Fisher problem (when $\sigma_1^2 \neq \sigma_2^2$)

Suppose that the following assumptions hold

1. $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$
2. $\sigma_1^2 \neq \sigma_2^2$
3. The X sample and Y sample are independent

Welch (1947) showed that

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

has an approximate t-distribution with degrees of freedom close to

$$v^* = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

Then using the pivot method again,

$$.95 = P\left[-t_{.025,v^*} \leq \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \leq t_{.025,v^*}\right]$$

$$\vdots$$

$$= P\left[(\overline{X} - \overline{Y}) - t_{.025,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{X} - \overline{Y}) + t_{.025,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right]$$

so that a 95% confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y}) \pm t_{.025,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

This is called the Welch-Satterthwaite (or simply Welch) confidence interval for the difference between two means.

This is the default method given by the *t.test()* function in R.

(Example here)

In general, the Welch-Satterthwaite $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is

$$(\overline{X} - \overline{Y}) \pm t_{\alpha/2,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Comments:

1. Welch-t requires normality, but relaxes the equal variance assumption required by pooled-t
2. If $\sigma_1^2 = \sigma_2^2$, Welch-t and pooled-t have approximately the same performance. If $\sigma_1^2 \neq \sigma_2^2$, Welch-t is better because pooled-t confidence interval may not have the desired coverage probability
3. Simulations suggest that conducting the Welch-t all the time is better than the two-stage test
   - Conduct a test for equal variance
   - Conduct pooled-t or Welch-t depending on outcome of test for equal variance

   This is because the test for equal variance is not sensitive enough, and the pooled-t ends up getting used even when variances are not equal
4. When normality is violated and sample sizes are small, then use nonparametric methods like the rank sum test (from Stat 5660)

2 Comparison of independent proportions

Suppose we want to compare two population proportions p_1 and p_2. The difference between two proportions $p_1 - p_2$ is estimated by $\hat{p}_1 - \hat{p}_2$. Since the two sample proportions are independent,

$$\begin{aligned}\text{Var}(\hat{p}_1 - \hat{p}_2) &= \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2)\\ &= \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\end{aligned}$$

so that

$$\text{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Furthermore, a $(1-\alpha)100\%$ confidence interval for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

<u>Example</u> Maternal Flu Vaccine
In the study, respiratory illness with fever occurred in 110 infants out of 153 in the vaccine group, and 153 infants out of 172 in the control group. Calculate a 95% confidence interval for the difference in proportions that develop respiratory illness with fever.

Soln:

$$\hat{p}_1 - \hat{p}_2 = 153/172 - 110/168 = .8895 - .6548 = .2348$$

The standard error is

$$\text{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{.8895(1-.8895)}{172} + \frac{.6547(1-.6548)}{168}} = .0438$$

A 95% confidence interval is

$$\begin{gathered}\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}(\text{SE})\\ .2348 \pm 1.96(.0438)\\ (.1490, .3206)\end{gathered}$$

3 Equivalence of confidence intervals and test of hypothesis

Consider the two-sample problem comparing two means. Recall that a test for

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2$$

with level of significance $\alpha = .05$ will reject the null hypothesis H_0 if

$$\left| \frac{\overline{X} - \overline{Y}}{\text{SE}} \right| > t_{.025,df}$$

The degrees of freedom for the critical value t depends on whether the Welch-t or pooled-t is used to calculate the SE. In either case, the test rejects H_0 if and only if

$$\left| \overline{X} - \overline{Y} \right| > t_{.025,df} \text{ (SE)}$$

or, equivalently, if and only if

$$\overline{X} - \overline{Y} \pm t_{.025,df} \text{ (SE)}$$

does not contain 0.

Similarly for proportions, to test for

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

with level of significance $\alpha = .05$, we will reject the null hypothesis H_0 if

$$\left| \frac{\hat{p}_1 - \hat{p}_2}{\text{SE}} \right| > z_{.025}$$

The test rejects H_0 if and only if

$$|\hat{p}_1 - \hat{p}_2| > z_{.025} \text{ (SE)}$$

or, equivalently, if and only if

$$\hat{p}_1 - \hat{p}_2 \pm t_{.025,df} \text{ (SE)}$$

does not contain 0.

4 The p-value

While a test concludes whether to reject the null hypothesis or not, the p-value provides a measure of evidence towards rejection.

Definition: p-value
The p-value of a test statistic is the probability of *observing a value as extreme as, or more extreme than, the one observed.*

Example Maternal Flu Vaccine
Respiratory illness with fever occurred in 110 infants out of 153 in the vaccine group, and 153 infants out of 172 in the control group. Calculate a p-value for testing equality of proportions that develop respiratory illness with fever.

Soln: The Z-test statistic is

$$\left| \frac{\hat{p}_1 - \hat{p}_2}{\text{SE}} \right| = \left| \frac{8895 - .6547}{.0438} \right| = 5.36$$

A standard normal variable Z exceeds 5.36 in absolute value with probability smaller than .0001, so p-value$< .05$. We reject the null and conclude significant difference between the two proportions.

Example Suppose we toss a coin 100 times, and observe 65 heads and 35 tails. Does this allow us to reject $H_0 : p = .5$? Under $H_0 : p = .5$, how often does one observe an imbalance as extreme as 65 percent? Here the p-value is the probability of getting 65 percent or more heads, plus the the probability of getting 35 or fewer heads, i.e.

$$\begin{aligned} \text{p-value} &= P_{p=.50}(65 \text{ or more heads}) + P_{p=.50}(35 \text{ or fewer heads}) \\ &= .0015 + .0015 \\ &= .003 \end{aligned}$$

Comment:

We reject H_0 if p-value $\leq .05$ so 'acceptance' consists of the 95% region of expected behavior, while 'rejection' takes the extreme 5% region outside that. In the example, 65 heads is in the extreme 5% (and extreme 1%) rather than the middle 95%. This definition of the rejection region provides a conditional test for comparison of two proportions.

Example Fisher Exact Test

In a smaller vaccine study, suppose that 5 out of 5 subjects in the control group develop illness, and 1 out of 5 in the treatment group. Is this enough evidence to conclude significant difference?

Soln: The data may be arranged in a contingency table as follows

	Yes	No	
Ctl	5	0	5
Trt	1	4	5
	6	4	10

Conditional on the marginals, there are 5 possible arrangements of the cells

$$\begin{array}{|cc|}\hline 5 & 0 \\ 1 & 4 \\ \hline\end{array} \quad \begin{array}{|cc|}\hline 4 & 1 \\ 2 & 3 \\ \hline\end{array} \quad \begin{array}{|cc|}\hline 3 & 2 \\ 3 & 2 \\ \hline\end{array} \quad \begin{array}{|cc|}\hline 2 & 3 \\ 4 & 1 \\ \hline\end{array} \quad \begin{array}{|cc|}\hline 1 & 4 \\ 5 & 0 \\ \hline\end{array}$$

which occur under the hypergeometric distribution with respective probabilities .0238, .2381, .4762, .2381, .0238. None of the other configurations are as extreme as or more extreme than the observed confiduration, so p-value=.0238.