Now let's look at a hypothesis test $H_0: \ p_1 = p_2$.

We have $\widehat{p}_1$ and $\widehat{p}_2$ as before, but also a *pooled estimate* computed under the assumption that H_0 is correct: $\widehat{p} = \frac{5+6}{78+31} = .1009$.

The pooled standard error is defined to be

$$
\begin{aligned}
SE &= \sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \\
&= \sqrt{.1009 \times .8991 \times \left(\frac{1}{78}+\frac{1}{31}\right)} \ = \ 0.0639.
\end{aligned}
$$

Note that this is similar to, but not the same as, the SE used in the confidence interval calculation.

Calculate the z score:

$$
\begin{aligned}
z &= \frac{\widehat{p}_2 - \widehat{p}_1}{SE} \\
&= \frac{.1935 - .0641}{.0639} = 2.03.
\end{aligned}
$$

The one-sided probability associated with this (Table A) is .0212.

Therefore, the two-sided P-value is .0424.

Just significant at the .05 level.

A caveat: This again uses the normal distribution in a situation where it is not strictly justified. It is possible to make the calculation without using a normal approximation (the method is called *Fisher's exact test*), but we shall not do that.

General method

Sample proportions $\widehat{p}_1,\ \widehat{p}_2$ based on sample sizes $n_1,\ n_2$.

Standard error for a confidence interval is

$$SE \;=\; \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1}+\frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}.$$

This formula assumes INDEPENDENCE of the two samples.

Confidence interval for $p_2 - p_1$ is $\widehat{p}_2 - \widehat{p}_1 \pm z \times SE$ where z is the z statistic appropriate to the confidence level, e.g. $z = 1.96$ for a 95% CI or $z = 2.58$ for a 99% CI.

For a hypothesis test, compute pooled proportion $\widehat{p} = \frac{n_1 \widehat{p}_1 + n_2 \widehat{p}_2}{n_1 + n_2}$.

Standard error and z score

$$
\begin{aligned}
SE &= \sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \\
z &= \frac{\widehat{p}_2 - \widehat{p}_1}{SE}.
\end{aligned}
$$

Compute one-sided or two-sided P-value associated with z.

We usually interpret this as rejecting the null hypothesis if P$<$ .05, otherwise do not reject H_0.

Example: Problem 10.3, page 480.

In two surveys of drinking habits, students were asked whether about binge drinking. In 1993, 31.2% of 159 students reported binge drinking; in 2005, 38.2% of 485 students did so.

(a) Estimate the difference between the proportions in 1993 and 2005, and interpret.

(b) Find the standard error of the difference, and interpret.

(c) Construct and interpret a 95% confidence interval for the true change, explaining how your interpretation reflects whether the interval includes 0.

(d) State and check the assumptions for the interval in (c) to be valid.

(a) $0.382 - 0.312 = 0.070$; proportion has increased

(b) $SE = \sqrt{\frac{.312 \times .688}{159} + \frac{.382 \times .618}{485}} = .0429$.

(c) $.070 \pm 1.96 \times .0429 = (-0.014, 0.154)$. Does include 0; therefore change is not statistically significant (though it's close).

(d) The main assumptions are independence of the two samples, and random sampling; these would appear to be satisfied. Also, the conditions $np > 15, n(1-p) > 15$ should be satisfied for each sample; obviously true here (e.g. with $n = 159$, $p = 0.312$, we have $np = 49.6$).

Other Examples of Testing for Equal Proportions

1. Sexual abstinence example (page 4, class 8): among 95 students in an “abstinence only” class, 31 (32.6%) reported having ever had sexual intercourse. Among 88 students in the control group, 41 (46.6%) reported sex.

2. Marathon runners example (class 24): 24/210 marathon runners had a potentially dangerous skin lesion, against 14/210 in the control group

3. 6/31 ”heavy training” marathon runners had a potentially dangerous skin lesion, against 14/210 in the control group

Solutions

1. $\hat{p}_1 = \frac{31}{95} = .3263$; $\hat{p}_2 = \frac{41}{88} = .4659$; $\hat{p} = \frac{72}{183} = .3934$. SE$=\sqrt{.3934 \times .6066 \times \left(\frac{1}{95} + \frac{1}{88}\right)} = .0723$. $z = \frac{.4659 - .3263}{.0723} =$ 1.93. One-sided P-value is .0268; two-sided is .0536.

2. $\hat{p}_1 = \frac{24}{210} = .1143$; $\hat{p}_2 = \frac{14}{210} = .0667$; $\hat{p} = \frac{38}{420} = .0905$. SE$=\sqrt{.0905 \times .9095 \times \left(\frac{1}{210} + \frac{1}{210}\right)} = .0280$. $z = \frac{.1143 - .0667}{.0280} =$ 1.7. One-sided P-value is .0446; two-sided is .0892.

3. $\hat{p}_1 = \frac{6}{31} = .1935$; $\hat{p}_2 = \frac{14}{210} = .0667$; $\hat{p} = \frac{20}{241} = .0830$. SE$=\sqrt{.083 \times .917 \times \left(\frac{1}{31} + \frac{1}{210}\right)} = .0531$. $z = \frac{.1935 - .0667}{.0531} =$ 2.388. One-sided P-value is .0085; two-sided is .017.

Comment

We still need the condition $n\hat{p} > 15$, $n(1-\hat{p}) > 15$ (for *both* samples) to justify the use of the normal approximation to the binomial distribution.

Especially in example 3, this is not satisfied ($n = 31$, $\hat{p} = .083$, $31 \times .083 = 2.6$).

There is another test, called *Fisher's exact test*, that doesn't require this condition. For the three examples considered here, the P-values according to Fisher's exact test are .069, .12 and .029. In each case, the P-value is a little larger than the one calculated from the normal distribution, but it doesn't change the overall conclusion.

Comparing Two Means

Example: 14 men in this class have a mean height of 72.57 inches and a standard deviation 3.131 inches.

69 women in this class have a mean height of 65.71 inches and a standard deviation 2.815 inches.

Is this a statistically significant difference?

Standard error for the women:

$$SE_1 = \frac{2.815}{\sqrt{69}} = .3389$$

Standard error for the men:

$$SE_2 = \frac{3.131}{\sqrt{14}} = .8368$$

Standard error for the difference:

$$SE = \sqrt{SE_1^2 + SE_2^2} = .9028$$

t statistic:

$$t = \frac{\bar{x}_2 - \bar{x}_1}{SE} = \frac{72.57 - 65.71}{0.9028} = 7.60.$$

Is this statistically significant?

Before we go on...

In this instance, the answer is clearly yes. For a t value as large as 7, we don't need a detailed computation of the P-value to conclude that the result is statistically significant.

However, in other cases it could be more critical, so we show how to obtain the P-value.

Degrees of Freedom

The Welch-Satterthwaite Formula:

$$df \;=\; \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$$

where s_1, s_2 are the individual standard deviations and n_1, n_2 are the sample sizes.

In this case $s_1 = 2.815$, $s_2 = 3.131$, $n_1 = 69$, $n_2 = 14$ and hence $df = 17.4$ (say 17 for looking up in table).

However there is also the simpler formula

$$df = \min(n_1, n_2) - 1.$$

In this case, that leads to $df = 13$.

In practice, it's good enough to use the simpler formula, certainly in this course.

Extract from the table of critical values of the t statistic with $df = 13$ or 17:

	Confidence level					
df	80%	90%	95%	98%	99%	99.8%
13	1.350	1.771	2.160	2.650	3.012	3.852
17	1.333	1.740	2.110	2.567	2.898	3.646

At P=.05, the critical value for significance is either 2.160 or 2.110 depending on which df you use.

Either way, $t = 7.60$ is clearly significant.

Suppose we want a 95% confidence interval for the difference in height between men and women.

Recall: the observed mean difference is $72.57 - 65.71 = 6.86$ and the standard error is 0.9028.

Based on $df = 13$, the critical value for a 95% confidence interval is 2.16.

Therefore, the desired 95% confidence interval is

$$6.86 \pm 2.16 \times 0.9028 \ = \ (4.91, 8.81).$$