12.2 Comparing two proportions

ex. Is there statistical evidence that the proportion of m&ms that are yellow is lower than the proportion of skittles that are yellow?

Very important distinction for 2 sample setting:

For confidence intervals, use

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

For hypothesis tests, use

$$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

This is so that the two-sample test statistic remains consistent with the test methods in later chapters.

As a result, no exact correspondence between a two-sided test and a CI exists.

Confidence Intervals for Comparing Two Proportions
page 704

Conditions

2 random

independent

ten %

and large enough ns ← **this lets us use z scores and normal approximations**

Confidence Intervals for Comparing Two Proportions

page 704

- 2 random samples
- that are independent
- with less than ten % of each population (N>10n), but
- a large enough n in each that we see (CI) or expect (HT/TOS) at least 10 successes and at least 10 failures

$$n\hat{p}_1 \ge 5 \quad n(1-\hat{p}_1) \ge 5$$
$$n\hat{p}_2 \ge 5 \quad n(1-\hat{p}_2) \ge 5$$

$$np_1 \ge 5 \quad n(1-p_1) \ge 5$$
$$np_2 \ge 5 \quad n(1-p_2) \ge 5$$

In a CI, we don't propose p, so we need to see $\ge$ 5 successes and $\ge$ 5 failures

In H_0: $p=p_0$, we propose p, so we need to expect $\ge$ 5 successes and $\ge$ 5 failures

Significance test for two proportions pages 708-709

To test the hypothesis: $H_o: p_1=p_2$
first find the pooled proportion p-hat

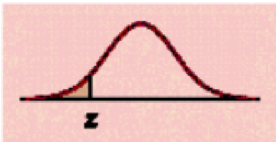
$$\hat{p} = \frac{\text{total successes in both samples}}{\text{total sample size of both samples}} = \frac{X_1 + X_2}{n_1 + n_2}$$

of sucesses in both samples combined and then compute the z statistic

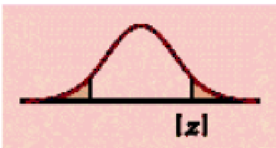
$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

In terms of using a variable Z having the standard normal distribution, the P-value for a test of H_o against
H_a: $p_1 > p_2$ is $P(Z \geq z)$

H_a: $p_1 < p_2$ is $P(Z \leq z)$



H_a: $p_1 \neq p_2$ is $2P(Z \geq |z|)$



Conditions for use: the populations should be 10 times larger than the samples and the values
n_1("p-hat$_1$") n_1(1-"p-hat$_1$") n_2("p-hat$_2$") n_2(1-"p-hat$_2$")
are all at least 5.

page 719 has a nice flowchart for selecting which inference procedures to use

zap ta$\overline{x}$

Pizza Time, Zeuσ!

proportions use z!

t is for means!

except that z is for when you know sigma!

page 703

Based on our work with random variables in ch. 7:

the mean of $\hat{p}_1 - \hat{p}_2 = \mu_{\hat{p}_1 - \hat{p}_2} = \mu_{\hat{p}_1} - \mu_{\hat{p}_2} = p_1 - p_2$

and the variance of $\hat{p}_1 - \hat{p}_2 = \sigma^2_{\hat{p}_1 - \hat{p}_2} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

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