MATH 124 Spring 2005

Lecture: 21
Date: May 10, 2005

Skills you should acquire from this lecture:
- Ability to find confidence interval for a single population proportion
- Ability to carry out hypothesis tests for a single population proportion

Related readings in the textbook:
- Section 8.1

We have been considering, to this point, confidence intervals for μ and hypothesis tests about μ also. Today we consider those same topics in relation to p the population proportion. Note that your text book uses slightly different formula.

Sample proportion

$$\hat{p} = \frac{X}{n} \quad \text{- estimate}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \quad \text{- standard deviation}$$

} discussed before in the context of sampling distributions.

CI for p

~~Recall that~~ when the sample size is large (ie n is big) then $\hat{p}$ is approximately normal with mean p and sd $\sqrt{\frac{p(1-p)}{n}}$. Using similar thinking to the derivation of the CI for μ we can show that an approximate level C confidence interval for p is given by

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ ← note the "hats in this term"

where $P(-z^* < z < z^*) = C$

Unfortunately, because we don't know p we use our sample estimate $\hat{p}$ in place of p in the formula for standard deviation

<u>Note</u> this estimated standard deviation of a parameter based on sample data is called the <u>standard error</u> (SE).

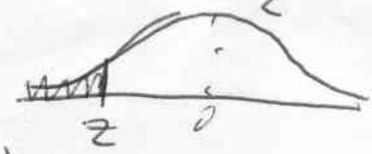
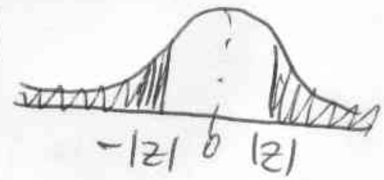
<u>Hypothesis testing for p</u>

$$\left.\begin{aligned} H_0&: p = p_0 \\ H_A&: p \neq p_0 \end{aligned}\right\}$$ typical null and alternative hypotheses.

test statistic $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ ← this test statistic has approximately the $N(0,1)$ distribution.

P-values come from the normal distribution.

Note that which region you use depends on H_A:

Alternative	Pvalue	
$H_A: p > p_0$	$P(Z \geq z)$	
$H_A: p < p_0$	$P(Z \leq z)$	
$H_A: p \neq p_0$	$2P(Z \geq \lvert z \rvert)$	

Again as with CI for μ I don't expect you to know how to compute sample sizes.

An Example (of how margins of error are computed for political polls)

Typical poll result

Candidate 1	47%
Candidate 2	45%
Candidate 3	3%
undecided	5%

In small print
margin of error is $\pm 3\%$
Survey of adults/likely voters aged 18+ conducted on ___

Where does the margin of error come from?

It is from a 95% confidence interval. But which one?

Rather than choosing only one candidate (or making confidence intervals for each candidate) the widest possible confidence interval for the given sample size is used. (this works out to choosing $\hat{p} = 0.5$)

<u>Example</u>

<u>Sample Size</u>	<u>Margin of error</u>
100	$1.96\sqrt{\frac{.5(1-.5)}{100}} = .098$ (+/- 9.8%)
250	$1.96\sqrt{\frac{.5(1-.5)}{250}} = .062$ (+/- 6.2%)
500	$1.96\sqrt{\frac{.5(1-.5)}{500}} = .044$ (+/- 4.4%)
1000	$1.96\sqrt{\frac{.5(1-.5)}{1000}} = .031$ (+/- 3.1%
2000	$1.96\sqrt{\frac{.5(1-.5)}{2000}} = .022$ (+/- 2.2%)

As sample size increases we see margin of error decreases.