Proportions

Inference for categorical data

Inferences for a single proportion

What is proportion?

- sample proportion can be described as a sample mean
- if success = 1 and fail = 0
- sample proportion ($\hat{p}$) is mean of numerical outcomes

Sampling distribution

- sample distribution is nearly normal when
  - observations are independent
  - At least 10 success and 10 failure
  - if conditions are met
    - ↳ $SE = \sqrt{\frac{p(1-p)}{n}}$

CI = Point estimate $\pm z^* SE$ → $z^* SE$ = margin of error

A recent estimate of congress approval rating was 19%. What sample size does this estimate suggest we should use for a margin of error of 0.04 with 95% confidence?

$n = ?$ $\quad 0.04 = 1.96 * \sqrt{\frac{p(1-p)}{n}} \Rightarrow .04 = 1.96 * \sqrt{\frac{.19(1-.19)}{n}}$

$n =$

When expected proportion is unknown...
50/50 is a good guess

- Do a majority of Americans support nuclear arms reduction? Set up a one-sided hypothesis test to evaluate this question.
- A simple random sample of 1028 US adults in March 2013 found that 56% support nuclear arms reduction. Does this provide convincing evidence that a majority of Americans supported nuclear arms reduction at the 5% significance level?

$n = 1028 \quad p = .56$

$H_0: \mu \leq .5 \qquad H_a: \mu > .5$

$z = \frac{.56 - .5}{\sqrt{.50(1-.50)/1028}} = 3.847$

Differences of Two Proportions

$$SE = \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

Independence within groups vs independence between groups

Construct a 95% confidence interval for the difference between the proportions of Duke students and Americans who would be bothered a great deal by the melting of the northern ice cap ($p_{Duke} - p_{US}$).

Data	Duke	US
A great deal	69	454
Not a great deal	36	226
Total	105	680
$\hat{p}$	0.657	0.668

$$(p_{Duke} - p_{US}) \pm z^* \cdot \sqrt{p_{Duke}(1-p_{Duke})/n_{Duke} + p_{US}(1-p_{US})/n_{US}}$$

$$= (.657 - .668) \pm 1.96 \cdot \sqrt{\frac{.657 \cdot .343}{105} + \frac{.668 \cdot .332}{680}}$$

$$= (-.108, .086)$$

Pooled proportion

$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$ ← only for 2 tailed test

No pooled proportion for the duke example...

$$\frac{69 + 454}{105 + 680} = .666$$

$$Z = \frac{p_{Duke} - p_{US}}{\sqrt{\frac{p(1-p)}{n_{Duke}} + \frac{p(1-p)}{n_{US}}}} = \frac{.657 - .668}{\sqrt{\frac{.666 \cdot .334}{105} + \frac{.666 \cdot .334}{680}}} = -.22$$

$P\text{value} = 2 \times P(z < -.22) = 2 \cdot .41 = .82$

$p > .05 \rightarrow$ fail to reject null hypothesis