Module 12a – Inference for a Population Proportion (Confidence Intervals)

Review Against All Odds: Unit 28 (Inference for Proportions)

Stat Procedure Diagram – Where we are

		Descriptive Statistics (Describing Pops or Samples)		**Inferential Statistics (from Samples)**
	Variable Types	Display	Describe	Estimation
Univariate	categorical (nominal or ordinal)*	Bar Graph/Pie Chart	Counts/Percentages	*When binary/dichotomous:* Confidence interval for proportions
	quantitative/continuous	Histogram/Stem & Leaf Box Plot	Mean/St Dev (normal) Median/Min, Q1, Q3, Max (skewed)	Confidence interval for means
		Display	Describe	Significance Tests/Hypothesis Tests
Bivariate	2 categorical	Tables or Bar Graphs	Two-way tables/Crosstabulation	Chi-square test (for goodness of fit)
	1 categorical, 1 quant.	Bar Graphs	Comparison of means/averages	T-test (one sample/group, two samples/groups) ANOVA (two or more samples/groups)
	2 quant.	Scatterplot	Correlation Coef. (Coef. of determ)/ Regression Line	T-test for correlation
		Display	Describe	Significance Tests/Hypothesis Tests
Multivariate	Response Variable is Quant.	-	Ordinary Least Squares Regression (OLS)	F-test for overall model T-tests for each explanatory variable
	Response Variable is categorical (dichotomous)	-	Logistic Regression	Chi-square tests of significance

NOTE: Items highlighted in yellow are covered in this course.

**When a categorical variable has two categories, it is called dichotomous.*

Shifting focus to …***categorical (dichotomous) variable inference***

- ***Confidence intervals for a proportion (yes/no variable)***
  - Most common example, by far, is election polling
  - E.g. Based on a sample, what is the estimated percentage of the population who said "yes" to voting for candidate "X"?

Confidence interval for a proportion

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}$: proportion of "successes" in sample

n: number of cases in the sample

z: the appropriate z-score for the desired confidence level

***No degrees of freedom necessary for proportions!**

Confidence interval for a proportion (common confidence levels)

90% C.I. $\hat{p} \pm 1.645 * \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

95% C.I. $\hat{p} \pm 1.96 * \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

99% C.I. $\hat{p} \pm 2.576 * \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

DEGREES OF FREEDOM	CONFIDENCE LEVEL C											
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291

Confidence interval for a proportion

Suppose we conduct a random sample of university students at a large university. In asking about alcohol consumption habits we ask respondents if they drink alcohol on most days (yes/no).

We can use this information to create a confidence interval for the proportion who said "yes" to the question.

If 170 out of the 2,673 respondents said "yes" to the question, we can calculate a 99% confidence interval to make an estimate about the overall population.

Confidence interval for a proportion

$$99\%\ \text{C.I.} \qquad \hat{p} \pm 2.576 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}$: proportion of "successes" in sample = 170/2,673

n: number of cases in the sample = 2,673

z: 2.576

$0.0636 +/- 2.576 * \sqrt{[0.0636 * 0.9364] / 2{,}673}$

$0.0636 +/- 2.576 * \sqrt{0.0000223}$

0.0636 +/- 0.0122 (or 6.36% +/- 1.22%)

Based on the sample results, we can be 99% certain that the true proportion of college students at the university who would report drinking alcohol daily is 6.36% +/- 1.22% (or 5.14% to 7.58%)

**No degrees of freedom necessary for proportions!*