Module 12b – Comparing Two Proportions (intro)

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

Stat Procedure Diagram – Where we are

We can use a z-test for proportions
if the DV is dichotomous

		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***

- ***One sample z-test for a proportion (yes/no variable)***
  - A hypothesis test for a proportion (generally comparing to 50% *standard*)
  - E.g. Based on a sample, do we have good evidence that the population proportion is greater than/less than/not equal to 50%?

One sample z-test (for a proportion)

Formula:

Standard Error

$$z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1 - p_o)}{n}}}$$

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

p_o: null hypoth. proportion (usually .50)

n : number of cases in the sample

Null Hyp. H_0: The true population proportion = .50 (50%)

Alt Hyp. H_a: The true population proportion > .50 (50%) (1-sided hypothesis)

OR

Alt Hyp. H_a: The true population proportion < .50 (50%) (1-sided hypothesis)

OR

Alt Hyp. H_a: The true population proportion is greater or less than .50 (50%) (2-sided hypothesis)

One sample z-test (for a proportion)

Are newborn babies more likely to be boys than girls (to compensate for higher male mortality rates)?

A random sample of 25,468 births records 13,173 are boys and the remainder are girls. We want to test the null hypothesis that 50% of births are boys.

- **Null Hyp. H_0**: The true proportion of male births = .50 (50%)

- **Alt Hyp. H_a:** The true proportion or male births > .50 (50%) (1-sided hypothesis)

One sample z-test (for a proportion)

$$z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1 - p_o)}{n}}}$$

$\hat{p}$: proportion of "successes" in sample = [13,173 / 25,468] = 0.5172

p_o : null hypoth. proportion (usually .50) = 0.50 (50%)

n : number of cases in the sample = 25,468

$z = [0.5172 - 0.50] / \sqrt{[0.50 * 0.50] / 25{,}468}$

$z = 0.0172 / \sqrt{0.00000982}$

$z = 0.0172 / .003133$

$z = 5.49$

One sample z-test (for a proportion)

- z = 5.49
- The sample average is 5.49 standard error terms above the hypothesized proportion of 50% (.50)

- We reject the null hypothesis if
  - …the absolute value of ***"z" (our test statistic)*** is greater than the ***critical value.***
  - …because if "z" is greater than the critical value, that means the ***p-value*** is less than the ***alpha level.***

Two sample t-test (matched pairs practice)

- The critical value for a one-tailed z-test at alpha=.05 is 1.645

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
One-sided P	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
Two-sided P	.50	.40	.30	.20	.10	.05	.04	.02	.01	.005	.002	.001

- Our z-statistic of 5.49 is greater than the critical value of 1.645.
- Therefore, we know that our p-value is below the alpha level of .05.
- Our result is statistically significant and we can reject the null hypothesis. We have good evidence that boys are born at slightly higher rates than girls.

Significance test results

for one-sided z-test if alpha=.05