Week 5 Examples

Week 5: Example 1: A 2011 study compared 268 children with **Autism Spectrum Disorder** to 1507 randomly selected **Control** children without the disorder. Of the children with autism, 20 of the mothers had used antidepressant drugs during the year before pregnancy or the first trimester of pregnancy. Of the control children, 50 of the mothers had used the drug.

a. Test the claim that prenatal exposure to antidepressant medicine is associated with a higher risk of autism. Give the *p*-value.
b. Can we conclude that prenatal exposure to antidepressant drugs causes autism? Why or why not?

Type of Question: Hypothesis test on a difference of independent proportions

What claim(s)? $p_A > p_C$ where

$p_A =$ population proportion of children with autism from mothers who took the antidepressant medication

$p_C =$ population proportion of children without autism from mothers who took the antidepressant medication

Method: Hypothesis test of a difference of proportions

StatKey: Randomization Test for difference of proportions

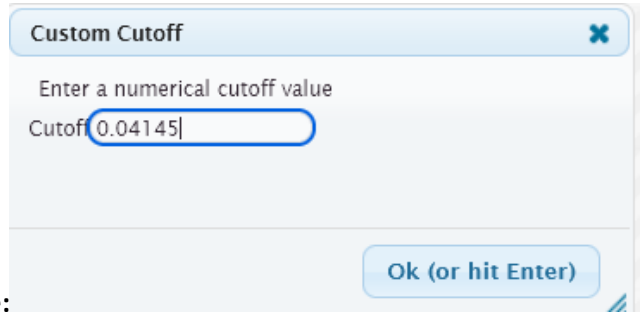
Write the hypotheses in symbols: $H_0 : p_A = p_C$
$H_A : p_A > p_C$

Identify the sample statistic(s) by symbols and with their numerical value(s). $\hat{p}_A = \frac{20}{268}$ $\hat{p}_C = \frac{50}{1507}$ $\hat{p}_A - \hat{p}_C = 0.04145$

Use simulations to find the p-value five times: 0.0010, 0.0020, 0.0010, 0.0030, 0.0020

Decide on the one you will use for the p-value: ___ Median = 0.0020 _____

Optional: Keep a screenshot of the last simulation**.** Here's where I put in the sample statistic value:



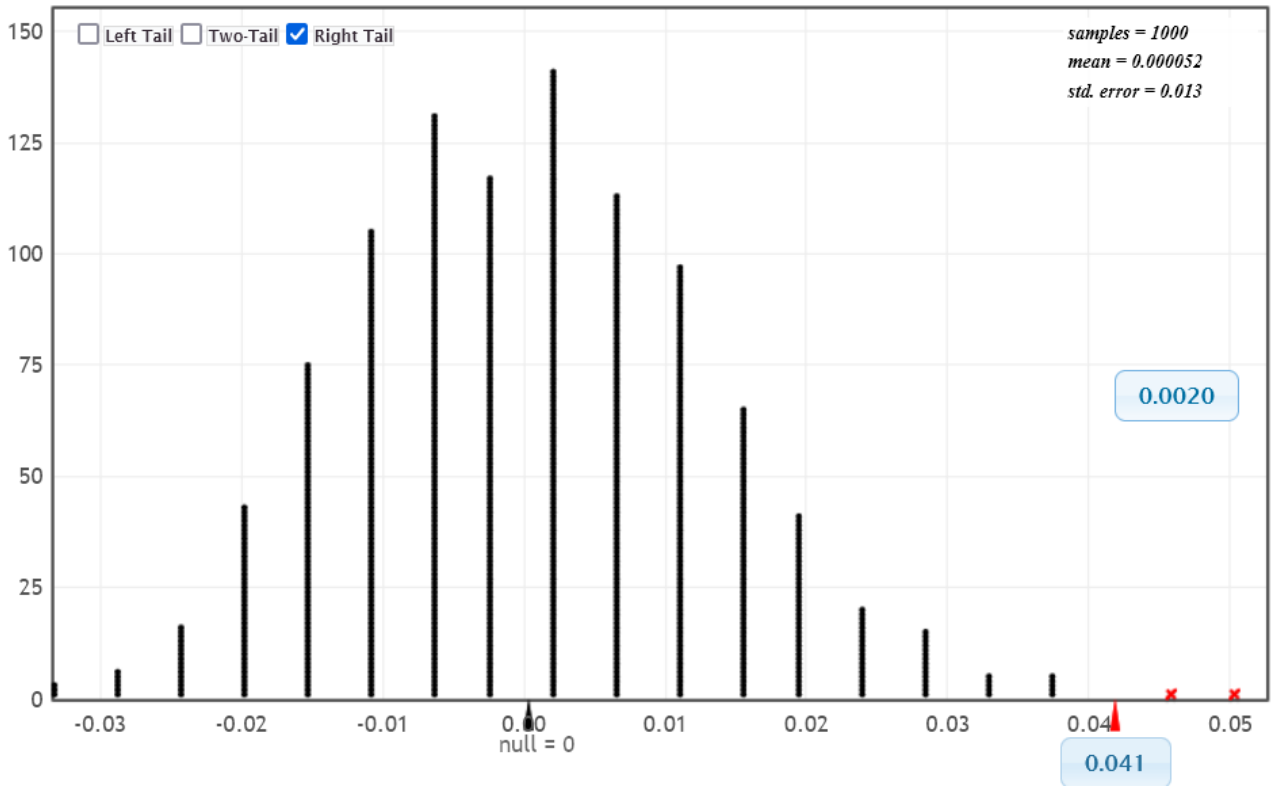
StatKey Randomization Test for a Difference in Proportions

Custom Data | Edit Data

Randomization method: Reallocation

Generate 1 Sample | Generate 10 Samples | Generate 100 Samples | Generate 1000 Samples | Reset Plot

Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$ Null Hypothesis: $p_1 = p_2$



Original Sample

Group	Count	Sample Size	Proportion
Group 1	20	268	0.075
Group 2	50	1507	0.033
Group 1-Group 2	-30	n/a	0.041

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	11	268	0.041
Group 2	59	1507	0.039
Group 1-Group 2	-48	n/a	0.0019

Answers:

a. The p-value is 0.002**.** This is very strong evidence that the proportion of children with autism is higher for mothers who took anti-depressant drugs during pregnancy.

b. The design of the study was not adequate to conclude causation because this was an observational study – not an experiment with randomly assigned treatments.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Week 5: Example 2: In the dataset **ICU Admissions for the 2nd edition of Lock text**, the variable **Status** indicates whether the person lived or died after being admitted to tie ICU (0 for lived, 1 for died) and the variable **Infection** indicates whether the person had an infection when admitted (0 for No, 1 for Yes.)
Find a 92% confidence interval for the difference in the proportions of those who die between those with an infection and those without an infection.

Type of question: Confidence interval for the difference of independent proportions

Method: Bootstrap Confidence Interval for the difference of proportions

Identify, in words, the parameter to estimate:

$p_1 =$ proportion of people in the population who die among those who were admitted to ICU with an infection

$p_2 =$ proportion of people in the population who die among those who were admitted to ICU without an infection

Identify the sample statistic(s):

(1) Find the dataset.
(2) In StatKey use Descriptive Statistics and Graphs to make this table.
(3) Pay careful attention to how to read the labels on tables: First word is for vertical labels, second word is for horizontal labels.

(This is the most confusing kind of table because it has the same labels for the two different variables.
It was chosen to call your attention to being careful when pulling data from full datasets.)

Status \ Infection	1	0	Total
0	60	100	160
1	24	16	40
Total	84	116	200

$$\hat{p}_1 = \frac{\text{died and with infection}}{\text{with infection}} = \frac{24}{84}$$

$$\hat{p}_2 = \frac{\text{died and without infection}}{\text{without infection}} = \frac{16}{116}$$

Five simulations for the 92% confidence interval: 0.047 to 0.254, 0.040 to 0.254, 0.045 to 0.246, 0.052 to 0.253, 0.049 to 0.258

Compute as needed to answer the question in the assignment: Lengths: 0.207, 0.214, 0.201, 0.201, 0.209 Median of these is 0.207

Warning: Notice If the left-hand endpoint is below zero, be sure to subtract that negative number correctly when you compute the length of the interval.

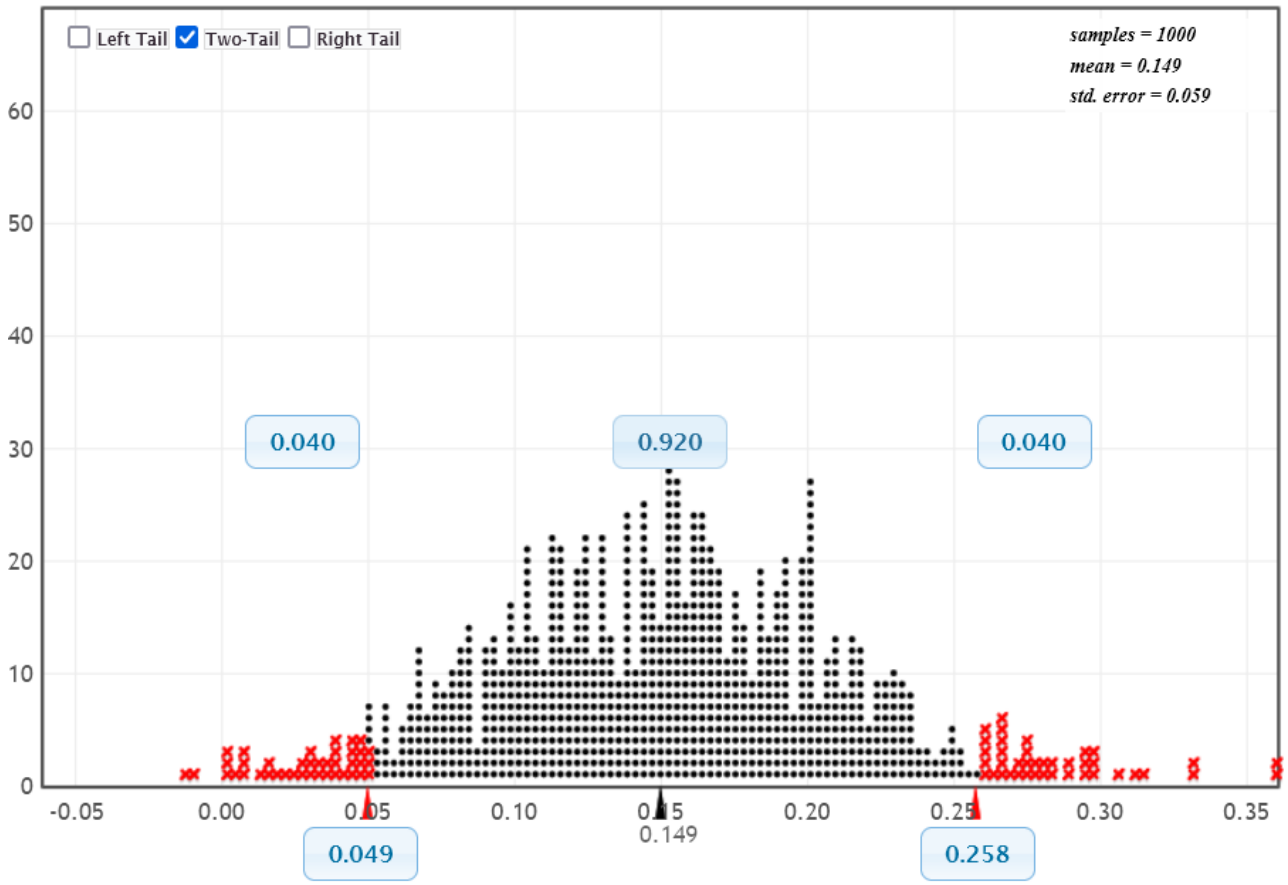
Optional: Keep a screenshot of the last simulation.

StatKey Confidence Interval for a Difference in Proportions

Custom Data | Edit Data

Generate 1 Sample | Generate 10 Samples | Generate 100 Samples | Generate 1000 Samples | Reset Plot

Bootstrap Dotplot of $\hat{p}_1 - \hat{p}_2$



Original Sample

Group	Count	Sample Size	Proportion
Group 1	24	84	0.286
Group 2	16	116	0.138
Group 1-Group 2	8	n/a	0.148

Bootstrap Sample

Group	Count	Sample Size	Proportion
Group 1	21	84	0.250
Group 2	16	116	0.138
Group 1-Group 2	5	n/a	0.112

xxxxxxxxxxxxxxxxxxxxxxxxxxxx

Week 5: Example 3. Consider the sampling distribution of the sample proportion when the population proportion is 0.03 and the sample size is 450. What is the standard error of the sampling distribution? Does this distribution appear to be symmetric?

Type of question: Sampling distribution of a proportion.

For a proportion, merely by the population proportion. $p = 0.03$

For a mean, you need the data for the whole population_____

Size of the sample: 450

What is the question? What is the standard error?

Does this distribution appear to be symmetric?

If the answer to the question is numerical,

then give the five values of that number for five simulations. std error: 0.0080, 0.0083, 0.0083, 0.0080, 0.0082

Decide on a number for the answer to the numerical question: std error. The median of these five values: 0.0082

Additional Question: Does the distribution appear to be symmetric?

Answer: No. It looks definitely skewed to the right.

Comment: When using "theoretical dist'n" formulas to do statistical calculations in a typical statistics class, it would be considered acceptable to approximate this distribution by a normal distribution, which is, of course, symmetric. Many students in such courses have not paid attention to such discrepancies. Those formulas were convenient to use in the early and mid-20th century before our computer capabilities allowed us to do more precise work.

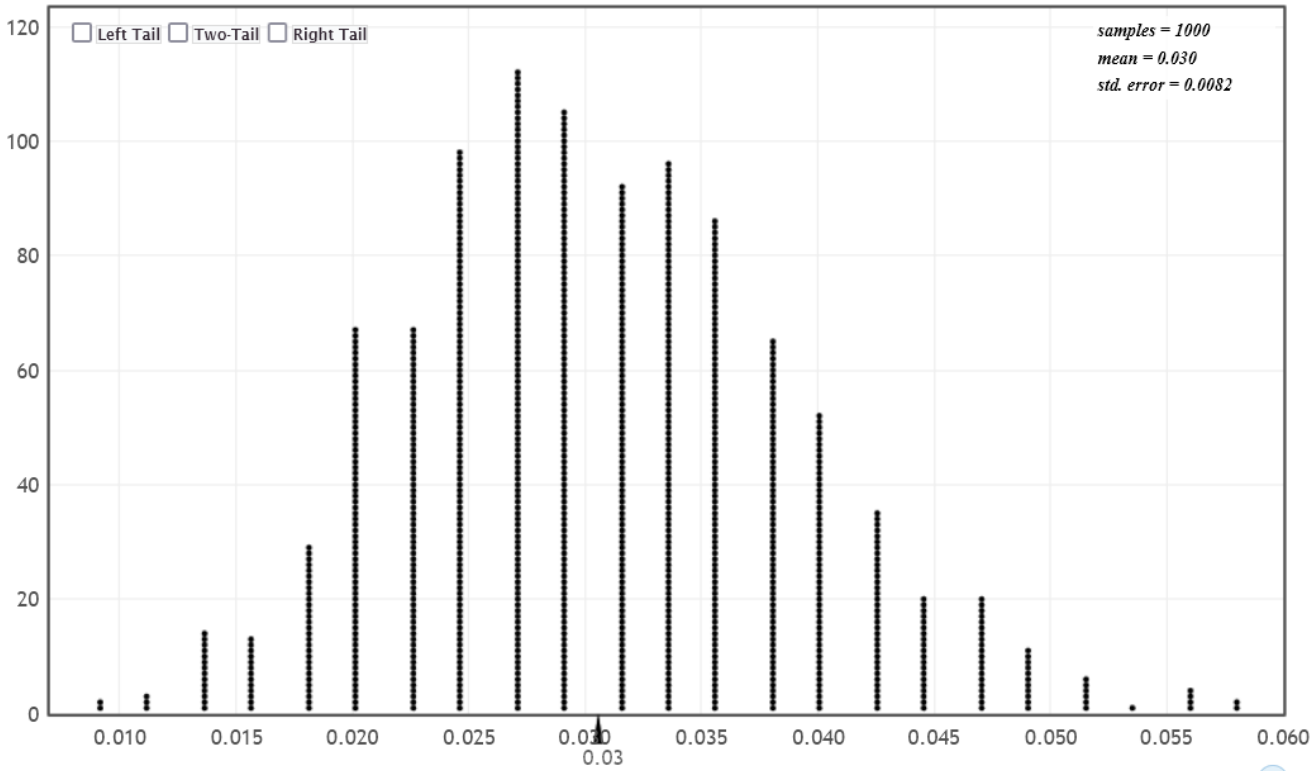
Optional: Keep a screenshot of the last simulation.

StatKey Sampling Distribution for a Proportion

Custom Data | Edit Proportion | Edit Data | Choose samples of size n = 450

Generate 1 Sample | Generate 10 Samples | Generate 100 Samples | Generate 1000 Samples | Reset Plot

Sampling Dotplot of Proportion



Data Tables | Confidence Intervals

Original Population

Proportion
0.03

Sample Show Data Table

Count	Sample Size	Proportion
20	450	0.044

XXXXXXXXXXXXXXXXXXXXXXXXXXXX