Inference about Two Population Proportions

Definition

A sampling method is independent when the individuals selected for one sample do not dictate which individuals are to be in a second sample. A sampling method is dependent when the individuals selected to be in one sample are used to determine the individuals in the second sample. Dependent samples are often referred to as matched-pairs samples. It is possible for an individual to be matched against him- or herself.

Sampling Distribution of the Difference Between Two Proportions

Suppose a simple random sample of size n_1 is taken from a population where x_1 of the individuals have a specified characteristic, and a simple random sample of size n_2 is independently taken from a different population where x_2 of the individuals have a specified characteristic.

Sampling Distribution of the Difference Between Two Proportions (cont.)

The sampling distribution of $\hat{p}_1 - \hat{p}_2$, where $\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$, is approximately normal, with mean $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ and standard deviation $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$, provided that $n_1\hat{p}_1(1-\hat{p}_1) \geq 10$ and $n_2\hat{p}_2(1-\hat{p}_2) \geq 10$ and each sample size is no more than 5% of the population size

Sampling Distribution of the Difference Between Two Proportions (cont.)

The standardized version of $\hat{p}_1 - \hat{p}_2$ is then written as

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Which has an approximate standard normal distribution

Hypothesis Test Regarding the Difference Between Two Population Proportions

Requirements:

1. Samples are independently obtained using simple random sampling or through a randomized experiment
2. $n_1\hat{p}_1(1-\hat{p}_1) \geq 10$ and $n_2\hat{p}_2(1-\hat{p}_2) \geq 10$
3. Each sample size is no more than 5% of the population size

Classical Approach (TI-83/84) – Independent Sample

1. Write down a shortened version of claim
2. Come up with null and alternate hypothesis (H_o always has the equals part on it)
3. See if claim matches H_o or H_1
4. Draw the picture and split α into tail(s)

   H_1: $p_1 \neq p_2$ Two Tail

   H_1: $p_1 < p_2$ Left Tail

   H_1: $p_1 > p_2$ Right Tail
5. Find critical values (INVNORM)
6. Find test statistic (2-PROPZTEST)
7. If test statistic falls in tail, Reject H_o. If test statistic falls in main body, Accept H_o. Determine the claim based on step 3

Classical Approach (By Hand) – Independent Sample

1. Write down a shortened version of claim
2. Come up with null and alternate hypothesis (H_o always has the equals part on it)
3. See if claim matches H_o or H_1
4. Draw the picture and split α into tails

   H_1: $p_1 \neq p_2$ Two Tail

   H_1: $p_1 < p_2$ Left Tail

   H_1: $p_1 > p_2$ Right Tail

Classical Approach (By Hand) – Independent Sample (cont.)

5. Find critical values: Use Standard Normal Distribution table
6. Find test statistic: $z_o = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

   where $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$
7. If test statistic falls in tail, Reject H_o. If test statistic falls in main body, Accept H_o. Determine the claim based on step 3

P-Value Approach (TI-83/84) – Independent Sample

1. Write down a shortened version of claim
2. Come up with null and alternate hypothesis (H_o always has the equals part on it)
3. See if claim matches H_o or H_1
4. Find p-value (2-PROPZTEST)
5. If p-value is less than α, Reject H_o. If p-value is greater than α, Accept H_o. Determine the claim based on step 3

P-Value Approach (By Hand) – Independent Sample

1. Write down a shortened version of claim
2. Come up with null and alternate hypothesis (H_o always has the equals part on it)
3. See if claim matches H_o or H_1
4. Find test statistic: $z_o = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

   where $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$

P-Value Approach (By Hand) (cont.)

5. Lookup the z-score from step 4 in the Standard Normal Distribution table and find the value from table (VFT)
   - Left Tail: p-value is the VFT
   - Right Tail: p-value is 1-VFT
   - Two Tail:

     If VFT $\leq$ 0.5: p-value is 2 times VFT

     If VFT > 0.5: p-value is 2 times (1 – VFT)
6. If p-value is less than α, Reject H_o. If p-value is greater than α, Accept H_o. Determine the claim based on step 3

1. Claim

In testing a new drug (Zizar) there was a concern of migraines occurring during usage. 300 out of 1000 people in the Zizar group experienced migraines while 250 out of 1000 people in the placebo group experienced migraines. Is the proportion of people getting migraines greater for those taking Zizar than those taking the placebo at the $\alpha = 0.05$ level of significance?

2. Claim

In 2008, a study was done asking 500 people if they read at least 2 books a year, 20 responded yes. In 2012, a study was done asking the same question of 800 people and 45 responded yes. Are the proportions different at $\alpha = 0.10$?

Constructing a Confidence Interval for the Difference between Two Population Proportions (Independent Samples)

Confidence interval for $p_1 - p_2$ is given by:

Lower bound:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Upper bound:

$$(\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

TI-83/84 Instructions

1. STAT Button
2. Right Arrow to TESTS
3. Down arrow and choose 2-PropZInt
4. Enter x_1, n_1, x_2, n_2, and C-Level and then Calculate

This will give you the interval.

$$Point\ Estimate = \frac{UB - LB}{2}$$

$$Margin\ of\ Error\ (E) = UB - Point\ Estimate$$

3. Confidence Intervals

In a poll of 1000 men, 200 said they believed in UFO's. In a poll of 780 women, 100 said they believed in UFO's. Construct a 95% confidence interval for the difference between the two population proportions, $p_{men} - p_{women}$

Testing a Hypothesis Regarding the Difference of Two Proportions: Dependent Samples (McNemar's Test)

To test hypothesis regarding two population proportions p_1 and p_2, where the samples are dependent, arrange the data in a contingency table as follows:

		Treatment A	
		Success	Failure
Treatment B	Success	f_{11}	f_{12}
	Failure	f_{21}	f_{22}

Requirements

1. Samples are dependent and are obtained randomly
2. Total number of observations where the outcomes differ must be greater than or equal to 10. That is $f_{12} + f_{21} \geq 10$

Classical Approach (TI-83/84) – Dependent Sample

1. Write down a shortened version of claim (always difference)
2. Come up with null and alternate hypothesis (H_o: $p_1 = p_2$ and H1: $p_1 \neq p_2$)
3. Draw the picture (always two tails) and split α into tail(s)
4. Find critical values (INVNORM)
5. Find test statistic: $z_o = \frac{|f_{12}-f_{21}|-1}{\sqrt{f_{12}+f_{21}}}$
6. If test statistic falls in tail, Reject H_o. If test statistic falls in main body, Accept H_o. Determine the claim based which one the claim matches

P-Value Approach (By Hand) – Dependent Sample

1. Write down a shortened version of claim (always difference)
2. Come up with null and alternate hypothesis (H_o: $p_1 = p_2$ and H1: $p_1 \neq p_2$)
3. Find test statistic: $z_o = \frac{|f_{12}-f_{21}|-1}{\sqrt{f_{12}+f_{21}}}$

P-Value Approach (By Hand) (cont.)

4. Lookup the z-score from step 3 in the Standard Normal Distribution table and find the value from table (VFT)
   - Left Tail: p-value is the VFT
   - Right Tail: p-value is 1-VFT
   - Two Tail:
     If VFT $\leq$ 0.5: p-value is 2 times VFT
     If VFT > 0.5: p-value is 2 times (1 – VFT)
5. If p-value is less than α, Reject H_o. If p-value is greater than α, Accept H_o. Determine the claim (remember always matches H_1 in these claims)

4. Claim

Given a survey of 1000 adults, they were asked whether they believe in UFO's and whether they believe in ESP. Given the table below, is there a significant difference in the proportion of adults who believe in UFO's and believe in ESP at α = 0.05.

	Believe in UFO's (success)	Don't Believe in UFO's (failure)
Believe in ESP (success)	100	400
Don't Believe in ESP (failure)	200	300

Sample Size for Estimating $p_1 - p_2$

The sample size necessary to obtain a confidence interval is given by:

$$n = n_1 = n_2 = [\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)]\left(\frac{z_{\alpha/2}}{E}\right)^2$$

assuming prior estimate is known

$$n = n_1 = n_2 = 0.5\left(\frac{z_{\alpha/2}}{E}\right)^2$$

assuming no prior estimate is known

Note: Always round up and E should be in decimal form

5. Sample Size

We want to determine the difference in the proportion of men versus the proportion of women who believe in UFOs. What sample size should be obtained with 95% confidence within 5 percentage points:

a) Assuming a prior estimate of 20% for men and 17% for women
b) Assuming no prior estimate