Section 11.1 – Inference about Two Population Proportions

Independent/Dependent

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 (Independent Sample)

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.

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.

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.

Section 11.1 – Inference about Two Population Proportions

Hypothesis Test Regarding the Difference between Two Population Proportions

To test hypotheses regarding two population proportions, p_1 and p_2, use the steps that follow, provided that

- the samples are independently obtained using simple random sampling or through a randomized experiment with two levels of treatment.
- $n_1\hat{p}_1(1-\hat{p}_1)\geq 10$ and $n_2\hat{p}_2(1-\hat{p}_2)\geq 10$, and
- $n_1 \leq 0.05N_1$ and $n_2 \leq 0.05N_2$ (the sample size is no more than 5% of the population size); this requirement ensures the independence necessary for a binomial experiment.

Step 1 – Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two- Tailed	Left-Tailed	Right-Tailed
$H_0:\ p_1 = p_2$	$H_0:\ p_1 = p_2$	$H_0:\ p_1 = p_2$
$H_1:\ p_1 \neq p_2$	$H_1:\ p_1 < p_2$	0o9

Note: p_1 is the population proportion for population 1, and p_2 is the population proportion for population 2.

Step 2 – Select a level of significance, α, depending on the seriousness of making a Type I error.

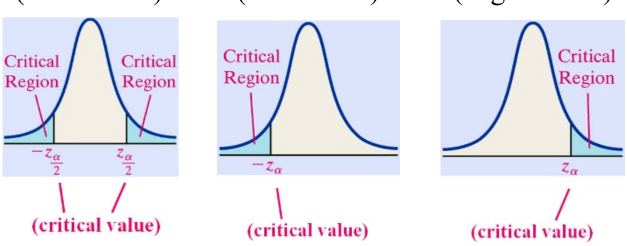
Step 3 – Compute the **test statistic** $z_0 = \dfrac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\cdot\sqrt{\left(\dfrac{1}{n_1}\right)+\left(\dfrac{1}{n_2}\right)}}$, where $\hat{p} = \dfrac{x_1+x_2}{n_1+n_2}$.

CLASSICAL APPROACH OR ***P*-VALUE APPROACH**

CLASSICAL APPROACH

Use Table V to determine the critical value.

$\pm z_{\alpha/2}$ (Two-tailed) or $-z_\alpha$ (Left-tailed) or $+z_\alpha$ (Right-tailed)

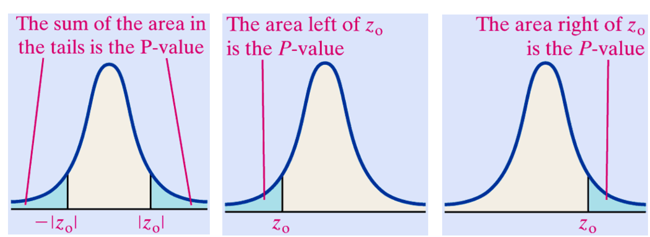


Step 4 – Compare the critical value with the test statistic.

Two- Tailed	Left-Tailed	Right-Tailed
If $z_0 < -z_{\frac{\alpha}{2}}$ or $z_0 > z_{\frac{\alpha}{2}}$ reject the null hypothesis	If $z_0 < -z_\alpha$ reject the null hypothesis	If $z_0 > z_\alpha$ reject the null hypothesis

***P*-VALUE APPROACH**

By Hand:
Use Table V to determine the *P*-value.



By Technology: Use a statistical spreadsheet or calculator with statistical capabilities to obtain the *P*-value.

Step 4 – If *P*-value $\leq \alpha$, reject the null hypothesis.

Step 5 – State the conclusion.

Section 11.1 – Inference about Two Population Proportions

Constructing a $(1-\alpha)\cdot 100\%$ Confidence Interval for the Difference between Two Population Proportions (Independent Samples)

To construct a $(1-\alpha)\cdot 100\%$ confidence interval for the difference between two population proportions from independent samples, the following requirements must be satisfied:

- the samples are obtained independently, using simple random sampling or from a randomized experiment.
- $n_1\hat{p}_1(1-\hat{p}_1)\geq 10$ and $n_2\hat{p}_2(1-\hat{p}_2)\geq 10$, and
- $n_1\leq 0.05N_1$ and $n_2\leq 0.05N_2$ (the sample size is no more than 5% of the population size); this ensures the independence necessary for a binomial experiment.

Provided that these requirements are met, $(1-\alpha)\cdot 100\%$ confidence interval for p_1-p_2 is given by

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

Sample Size for Estimating p_1-p_2

The sample size required to obtain a $(1-\alpha)\cdot 100\%$ confidence interval with a margin of error, *E*, is given by

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

rounded up to the next integer, if prior estimates of p_1 and p_2, $\hat{p}_1$ and $\hat{p}_2$, are available.

If prior estimates of p_1 and p_2 are unavailable, the sample size is

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

rounded up to the next integer.

Section 11.1 – Inference about Two Population Proportions

☺ **Exercises:**

1) **Vasectomies and Prostate Cancer.** Approximately 450,000 vasectomies are performed each year in the United States. In this surgical procedure for contraception, the tube carrying sperm from the testicles is cut and tied. Several studies have been conducted to analyze the relationship between vasectomies and prostate cancer. The results of one such study by E. Giovannucci et al. appeared in the paper "A Retrospective Cohort Study of Vasectomy and Prostate Cancer in U.S. Men" (*Journal of the American Medical Association*, Vol. 269(7), pp. 878-882). Of 21,300 men who did not have a vasectomy, 69 were found to have prostate cancer; of 22,000 men who had a vasectomy, 113 were found to have prostate cancer. At the 1% significance level, do the data provide sufficient evidence to conclude that men who have had a vasectomy are at greater risk of having prostate cancer? (Use either the classical or *P*-value approach and assume all the assumptions that are required for this test have been met).

Section 11.1 – Inference about Two Population Proportions

☺ **Exercises:**

2) **Raising Taxes.** Time magazine reported the result of a telephone poll of 800 adult Americans. The question posed of the Americans who were surveyed was: "Should the federal tax on cigarettes be raised to pay for health care reform?" The results of the survey were:

Non-Smokers	**Smokers**
$n_1 = 605$	$n_2 = 195$
$\overline{x}_1 = 351$ (said 'yes')	$\overline{x}_2 = 41$ (said 'yes')
$\hat{p}_1 = \frac{351}{605} = 0.58$	$\hat{p}_2 = \frac{41}{195} = 0.21$

Is there sufficient evidence at the $\alpha = 0.01$ level, say, to conclude that the two populations, smokers and non-smokers, differ significantly with respect to their opinions?

☺ **Exercises:**

3) **Body Mass Index.** The body mass index (BMI) of an individual is one measure that is used to judge whether an individual is overweight or not. A BMI between 20 and 25 indicates that one is at a normal weight. In a survey of 750 men and 750 women, the Gallup organization found that 203 men and 270 women were normal weight. Construct a 90% confidence interval to gauge whether there is a difference in the proportion of men and women who are normal weight. Interpret the interval. (Use either the classical or *P*-value approach).

4) An educator wants to determine the difference between the proportion of males and females who have completed 4 or more years of college. What sample size should be obtained if she wishes the estimate to be within 2 percentage points with 90% confidence, assuming that

a) she uses the 1999 estimates of 27.5% male and 23.1% female from the U.S. Census Bureau?

b) she does not use any prior estimates?