Chapters 9-12

Stat 128: Elementary Statistics

Nathan Warner

Computer Science
Northern Illinois University
United States
July 27, 2023

Contents

Chapter 9

0.1   9.1 Estimating a Population Proportion

Learning Objectives For This Section:
1. **Obtain a Point Estimate for the Population Proportion**
2. **Construct and Interpret a Confidence Interval for the Population Proportion**
3. **Determine the Sample Size Necessary for Estimating a Population Proportion within a Specified Margin of Error**

Vocab:
- A **point estimate** is the value of a statistic that estimates the value of a parameter.
- **A Confidence Interval** for an unknown parameter consists of an interval of numbers based on a point estimate.
- **The level of confidence** represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained.
- **Critical Value** represents the number of standard deviations the sample statistic can be from the parameter and still result in an interval that includes the parameter.

Formulas/Notation:
- **Point Estimate = Sample Proportion**
$$\hat{p} = \frac{x}{n}.$$
- The **Margin of Error** is denoted:
$$E.$$
- **Confidence Intervals for a Proportion** are of the form:
$$\text{point estimate} \pm \text{margin of error}$$
  That is:
$$\hat{p} \pm E.$$
  We have 95% confidence that the sampling proportion will lie between:
$$\hat{p} \pm 1.96\sigma_{\hat{p}}.$$
- **The level of confidence** is denoted:
$$(1 - \alpha) \cdot 100\%.$$
- **Standard Error:**
$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$
- **Constructing any $(1-\alpha)\cdot 100\%$ Confidence Interval**
$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$
- **Critical Value**
$$z_{\frac{\alpha}{2}}.$$

- **Constructing a $(1-\alpha)\cdot$ 100% Confidence Interval for a Population Proportion:** Suppose that a simple random sample of size n is taken from a population or the data are the result of a randomized experiment. A $(1-\alpha)\cdot100\%$ confidence interval for p is given by the following quantities:

$$Lower\ bound:\ \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$Upper\ bound:\ \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

  Note: It must be the case that $n\hat{p}(1-\hat{p}) \geqslant 10$ and $n \leqslant 0.05N$ to construct this interval. Use $\hat{p}$ in place of p in the standard deviation. This is because p is unknown, and $\hat{p}$ is the best point estimate of p.

- The **margin of error**, E, in a $(1-\alpha)\cdot100\%$ confidence interval for a population proportion is given by

$$E = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$or:\ \frac{Upper\ Limit\ -\ Lower\ Limit}{2}.$$

- **Sample Size Needed for Estimating the Population Proportion**

$$n = \hat{p} \cdot (1 - \hat{p}) \cdot \left(\frac{z_{\alpha/2}}{E}\right)^2.$$

  (rounded up to the next integer) where $\hat{p}$ is a prior estimate of p.

  If a prior estimate of p is unavailable, the sample size required is

$$n = 0.25 \cdot \left(\frac{z_{\alpha/2}}{E}\right)^2.$$

  rounded up to the next integer. The margin of error should always be expressed as a decimal when using these formulas.

- **Width**

$$2(E).$$

Introduction

Two Types of Inferential Statistics:

1. Estimation

2. Hypothesis Testing

Obtain a Point Estimate for the Population Proportion

Suppose we want to estimate the proportion of adult Americans who believe that the amount they pay in federal income taxes is fair. It is unreasonable to expect that we could survey every adult American. Instead, we use a sample of adult Americans to arrive at an estimate of the proportion. We call this estimate a point estimate.

Example:

Problem:

The Gallup Organization conducted a poll in April 2017 in which a simple random sample of 1019 Americans aged 18 and older were asked, "Do you regard the income tax that you will have to pay this year as fair?" Of the 1019 adult Americans surveyed, 620 said yes. Obtain a point estimate for the proportion of Americans aged 18 and older who believe that the amount of income tax they pay is fair.

Approach: The point estimate of the population proportion is $\hat{p} = \frac{x}{n}$, where $x = 620$ and $n = 1019$.

Solution:

Substituting into the formula, we get $\hat{p} = \frac{x}{n} = \frac{620}{1019} = 0.608 = 60.8\%$.

We estimate that 60.8% of Americans aged 18 and older believe that the amount of income tax they pay is fair.

Construct and Interpret a Confidence Interval for the Population Proportion

What if we conducted a different random sample of 1019 Americans aged 18 years or older? Would we get the same sample proportion who believe the amount of income tax they pay is fair? Probably not. Why? Because statistics such as $\hat{p}$ vary from sample to sample. So a different random sample of adult Americans might result in a different point estimate of the population proportion, such as $\hat{p} = 0.593$.

If the method used to select the adult Americans was done appropriately, both point estimates would be good guesses of the population proportion. Due to variability in the sample proportion, we report a range (or interval) of values, including a measure of the likelihood that the interval includes the unknown population proportion.

Consider the following:

$$\hat{p} = 0.54 \quad E = 0.034.$$

Then:

$$0.54 - 0.034 = 0.506$$
$$0.54 + 0.034 = 0.574.$$

With this we can say "We are 95% confident that the proportion is between 0.506 and 0.574"

Figure:
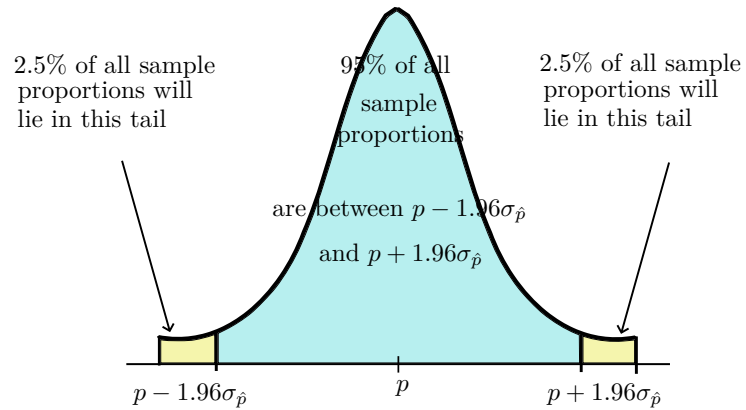
Where does the 1.96 comes from?

$$z_\alpha$$
$$z_{0.025} = 1.96.$$

This graphic also infers that 95% of all sample proportions will lie between:

$$p - 1.96\sigma_{\hat{p}} < \hat{p} < p + 1.96\sigma_{\hat{p}}.$$

By solving this inequality such that p is in the center:

$$\hat{p} - 1.96\sigma_{\hat{p}} < p < \hat{p} + 1.96\sigma_{\hat{p}}.$$

And we can write this in shorthand form:

$$\hat{p} \pm 1.96\sigma_{\hat{p}}$$
$$or: \ \hat{p} \pm E.$$

figure:

Summary:

- for a 95% confidence interval, any sample proportion that lies withing 1.96 standard errors of the population proportion will result in a confidence interval that includes p. This will happen in 95% of all possible samples.

- Any sample proportion that is more than 1.96 standard errors from the population proportion will result in a confidence interval that does not contain p. This will happen in 5% of all possible samples (those sample proportions in the tails of the distribution).
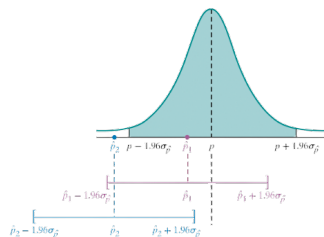
- A confidence interval for an unknown parameter consists of an interval of numbers based on a point estimate.

- The level of confidence represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained. The level of confidence is denoted $(1-\alpha) \cdot 100\%$.

- Whether a confidence interval contains the population parameter depends solely on the value of the sample statistic. Any sample statistic that is in the tails of the sampling distribution will result in a confidence interval that does not include the population parameter. See Figure 1. Notice that $\hat{p}_1$ results in a confidence interval that includes the population proportion, p. However, $\hat{p}_2$ results in a confidence interval that does not include the population proportion, p, because $\hat{p}_2$ is in the tails of the distribution.

Question 1

The horizontal axis in the sampling distribution of represents all possible sample proportions from a simple random sample of size n.

a.) What percent of sample proportions results in a 75% confidence interval that includes the population proportion?

b.) What percent of sample proportions results in a 75% confidence interval that does not include the population proportion?

Solution::   ☺

a.) 75%

b.) 25%

Caution:

95% confident does not mean 95% probability. Probability describes the likelihood of undetermined event.

It does not make sense to talk about the probability that the interval contains the parameter, because the parameter is an unknown, but fixed value. Thus, it is either in the interval, or not.

Consider a coin flip, if a coin has been flipped but its outcome has not been revealed, the flip resulting in a head is not $\frac{1}{2}$, because the outcome has already been determined. Instead the probability is either 0, or 1

Therefore, the probability that any confidence interval contains a parameter is either 0, or 1.

We do say "we are 95% confident the interval contains the parameter" because the method "works" in 95% of all samples.

Constructing any $(1 - \alpha) \cdot 100\%$ **Confidence interval**

We need a method for constructing any $(1 - \alpha) \cdot 100\%$ confidence interval. When $\alpha = 0.05$, we are constructing a 95% confidence interval.

We generalize

$$p \quad - \quad 1.96\sigma_{\hat{p}} \quad < \quad \hat{p} \quad < \quad p \quad + \quad 1.96\sigma_{\hat{p}}$$

$$\text{parameter} \ - \ 1.96 \, \text{standard error} \ < \ \text{point estimate} \ < \ \text{parameter} \ + \ 1.96 \, \text{standard error}$$

by first noting that $(1 - \alpha) \cdot 100\%$ of all sample proportions are in the interval as shown in Figure 2.

Figure 2
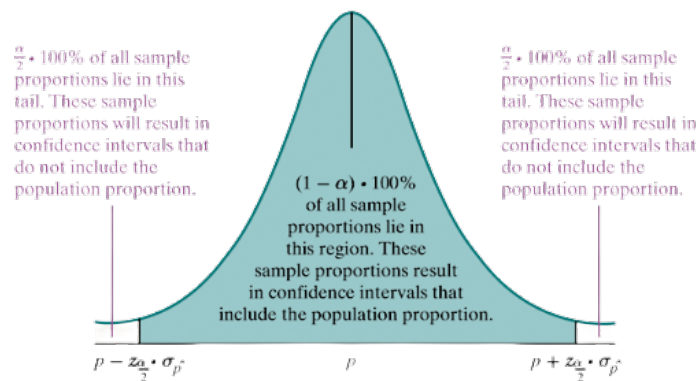
Rewrite this inequality with p in the middle and obtain

$$\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

So $(1 - \alpha)$ 100% of all sample proportions will result in confidence intervals that contain the population proportion. The sample proportions that are in the tails of the distribution in Figure 2 will not result in confidence intervals that contain the population proportion.

Table 1 shows some of the common critical values used in the construction of confidence intervals. Notice that higher levels of confidence correspond to higher critical values. After all, if your level of confidence that the interval includes the unknown parameter increases, the width of your interval (through the margin of error) should increase.

Level of confidence $(1 - \alpha) \cdot 100\%$	Area in each tail $\frac{\alpha}{2}$	Critical Value $z_{\frac{\alpha}{2}}$
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.575

Interpretation of a Confidence Interval

A $(1 - \alpha) \cdot 100\%$ confidence interval indicates that $(1 - \alpha) \cdot 100\%$ of all simple random samples of size n from the population whose parameter is unknown will result in an interval that contains the parameter.

For example, a 90% confidence interval for a parameter suggests that 90% of all possible samples will result in an interval that includes the unknown parameter and 10% of the samples will result in an interval that does not capture the parameter.

Example: Interpreting a Confidence Interval

Problem: The Gallup Organization conducted a poll in April 2017 in which a simple random sample of 1019 Americans aged 18 and older were asked, "Do you regard the income tax that you will have to pay this year as fair?" We learned from Example 1 that the proportion of those surveyed who responded yes was 0.608. Gallup reported its "survey methodology" as follows:

Approach: Confidence intervals for a proportion are of the form point estimate $\pm$ margin of error. So add and subtract the margin of error from the point estimate to obtain the confidence interval. Interpret the confidence interval, "We are 95% confident that the proportion of Americans aged 18 and older who believe that the income tax they will have to pay this year is fair is between lower bound and upper bound."

Solution:

The point estimate is 0.608, and the margin of error is 0.04. The confidence interval is 0.608±0.04. Therefore, the lower bound of the confidence interval is $0.608 - 0.04 = 0.568$ and the upper bound of the confidence interval is $0.608 + 0.04 = 0.648$. We are 95% confident that the proportion of Americans aged 18 and older who believe that the income tax they will have to pay this year is between 0.568 and 0.648.

We are now prepared to present a method for constructing a confidence interval about the population proportion, p

Constructing a $(1-\alpha)$ 100% Confidence Interval for a Population Proportion:

Suppose that a simple random sample of size n is taken from a population or the data are the result of a randomized experiment. A $(1-\alpha)$ 100% confidence interval for p is given by the following quantities:

$$Lower\ bound:\ \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$Upper\ bound:\ \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

Note:-

It must be the case that $n\hat{p}(1 - \hat{p}) \geqslant 10$ and $n \leqslant 0.05N$ to construct this interval. Use $\hat{p}$ in place of p in the standard deviation. This is because p is unknown, and $\hat{p}$ is the best point estimate of p.

Example: Constructing a Confidence Interval for a Population Proportion (Statcrunch)

Problem: In the Parent–Teen Cell Phone Survey conducted by Princeton Survey Research Associates International, 800 randomly sampled 16- to 17-year-olds living in the United States were asked whether they have ever used their cell phone to text while driving. Of the 800 teenagers surveyed, 272 indicated that they text while driving. Obtain a 95% confidence interval for the proportion of 16- to 17-year-olds who text while driving.

Approach: It is important to verify the requirements for constructing the confidence interval first. It must be the case that $n\hat{p}(1\hat{p}) \geqslant 10$ and the sample size is no more than 5% of the population size (n $\leqslant$ 0.05N). Then, construct the confidence interval either by hand or using technology.

Solution:

First, we obtain $\hat{p}$

$$\hat{p} = \frac{272}{800} = .34.$$

Verify Necessary Conditions:

$$n\hat{p}(1 - \hat{p}) \geqslant 10$$
$$800(.34)(.66) = 179.52 \geqslant 10.$$

Also, we know that $n \leqslant 0.05N$:

now in statcrunch:

1. Stat > Proportion Stats > One Sample > with summary

2. Input no. of successes

3. Input no. of observations

4. Input confidence level

5. Method: Standard-Wald

6. Compute

after doing the StatCrunch steps, we get a result of 0.307 and 0.373, therefore, we can conclude that we are 95% confident that the proportion of 16 to 17-year-olds who text while driving is between 0.307 and 0.373

Example: Constructing a Confidence Interval for a Population Proportion (By hand)

Problem: In the Parent–Teen Cell Phone Survey conducted by Princeton Survey Research Associates International, 800 randomly sampled 16- to 17-year-olds living in the United States were asked whether they have ever used their cell phone to text while driving. Of the 800 teenagers surveyed, 272 indicated that they text while driving. Obtain a 95% confidence interval for the proportion of 16- to 17-year-olds who text while driving.

Approach: It is important to verify the requirements for constructing the confidence interval first. It must be the case that $n\hat{p}(1\hat{p}) \geqslant 10$ and the sample size is no more than 5% of the population size (n $\leqslant$0.05N).Then, construct the confidence interval either by hand or using technology.

Solution:

First, we obtain $\hat{p}$

$$\hat{p} = \frac{272}{800} = .34.$$

Verify Necessary Conditions:

$$n\hat{p}(1 - \hat{p}) \geqslant 10$$
$$800(.34)(.66) = 179.52 \geqslant 10.$$

Also, we know that $n \leqslant 0.05N$:

Now, to use the formula:

$$\hat{p} \pm E$$
$$Where\ E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

We will need to find alpha, given that the formula for confidence level is:

$$(a - \alpha) \cdot 100\%$$
$$1 - \alpha = 0.95$$
$$\alpha = 0.05.$$

So our critical value is:

$$z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}}$$
$$= z_{0.025}.$$

Using the normal distribution table, we get:

$$\alpha = 1.96.$$

Now we can compute our bounds:

$$LB := \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 0.34 - 19.6 \cdot \sqrt{\frac{0.34(0.66)}{800}}$$

$$= 0.307$$

$$UB = \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 0.34 + 1.96 \cdot \sqrt{\frac{0.34(0.66)}{800}}$$

$$= 0.373.$$

Therefore:

$$Upper\ bound:\ 0.307 \quad Upper\ Bound:\ 0.373$$
$$Or:(0.307, 0.373).$$

The Effect of Level of Confidence on the Margin of Error

Example: The Role of the Level of Confidence in the Margin of Error

Problem: In the Parent–Teen Cell Phone Survey conducted by Princeton Survey Research Associates International, 800 randomly sampled 16- to 17-year-olds living in the United States were asked whether they have ever used their cell phone to text while driving. Of the 800 teenagers surveyed, 272 indicated that they text while driving. From the last example, we concluded that we are 95% confident that the proportion of 16- to 17-year-olds who text while driving is between 0.307 and 0.373. Determine the effect on the margin of error by increasing the level of confidence from 95% to 99%.

Approach: We would expect the margin of error to increase with a larger level of confidence. Construct the confidence interval either by hand or using technology.

Solution:

The margin of error for the 95% confidence interval found in Example 3 is 0.033, and the margin of error for the 99% confidence interval is 0.043. So increasing the level of confidence increases the margin of error, resulting in a wider confidence interval.

The Effect of Sample Size on the Margin of Error

We know that larger sample sizes produce more precise estimates (the Law of Large Numbers). Given that the margin of error is $E = z_{\alpha/2} \cdot \hat{p} \cdot \sqrt{\frac{(1-\hat{p})}{n}}$, we can see that increasing the sample size n decreases the standard error, so the margin of error decreases. Therefore, larger sample sizes will result in narrower confidence intervals.

What If We Do Not Satisfy the Normality Condition?

When the normality condition is not satisfied, the proportion of intervals that capture the parameter is below the level of confidence.

Determine the Sample Size Necessary for Estimating a Population Proportion within a Specified Margin of Error

---

Example: Determining Sample Size

Problem: An economist wants to know if the proportion of the U.S. population who commutes to work via car-pooling is on the rise. What size sample should be obtained if the economist wants an estimate within 2 percentage points of the true proportion with 90% confidence?

Assume that the economist uses the estimate of 10% obtained from the American Community Survey.

Approach: Since $1 - \alpha = 0.9$, we know that $\alpha = 0.10$. Use $E = 0.02$, $z_{\alpha/2} = z_{0.12} = z_{0.05} = 1.645$, and $\hat{p} = 0.10$ (the prior estimate).

Solution: Using the formula assuming that a prior estimate is available, $n = \hat{p}(1 - \hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$, we obtain

$$n = \hat{p}(1 - \hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2 = 0.10 \times (1 - 0.10) \times \left(\frac{1.645}{0.02}\right)^2 = 608.9.$$

Round this value up to 609. So the economist must survey 609 randomly selected residents of the United States.

statcrunch steps:

First:

$$E = 0.02 \ (\text{The 2 percentage points})$$
$$W = 2(E) = 0.04$$
$$C.Level = 0.9$$
$$Target \ Proportion = 0.1.$$

1. Stats > Proportion Stats > One Sample > Width/Sample Size

2. Input confidence level

3. Input Target Proportion

4. Compute

And we find that the minimum sample size would be 609

Now consider that the economist does not use any prior estimates (target proportion), in this case, our target proportion will be 0.5

With this target proportion we get 1691

0.2   9.2: Estimating a Population Mean

Learning Objectives For This Section:

1. **Obtain a Point Estimate for the Population Mean**

2. **State Properties of Student's t-Distribution**

3. **Determine t-Values**

4. **Construct and Interpret a Confidence Interval for a Population Mean**

5. **Determine the Sample Size Necessary for Estimating a Population Mean within a Given Margin of Error**

Vocab:

- **The point estimate of the population mean**, μ, is the sample mean, $\overline{x}$.

- **T-Interval:** confidence interval that uses the t-distribution

Notation/Formulas:

- **Student's t-distribution**

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}.$$

- **Constructing a $(1-\alpha)\cdot 100\%$ Confidence Interval for μ** Provided:
    - sample data come from a simple random sample or randomized experiment
    - sample size is small relative to the population size ($n \leqslant 0.05N$)
    - the data come from a population that is normally distributed with no outliers, or the sample size is large

    A $(1-\alpha)\cdot 100\%$ confidence interval for μ is given by

$$LB := \overline{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$
$$UB := \overline{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$
$$.$$

    where $t_{\frac{\alpha}{2}}$ is the critical value with $n-1$ degrees of freedom.

- **Margin of Error:**

$$E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}.$$

- **Determine the Sample Size Necessary for Estimating a Population Mean within a Given Margin of Error**

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot s}{E} \right)^2.$$

Obtain a Point Estimate for the Population Mean

To find the point estimate of the population mean, compute the sample mean, $\bar{x}$

State Properties of Student's t-Distribution

a different random sample of 16 cars would likely result in a different point estimate of μ. For this reason, we want to construct a confidence interval for the population mean, just as we did for the population proportion.

A confidence interval for the population mean is of the form

$$\text{Point estimate} \pm \text{Margin of Error}$$

(just like the confidence interval for a population proportion). To determine the margin of error, we need to know the sampling distribution of the sample mean.

Recall that the distribution of $\bar{x}$ is approximately normal if the population from which the sample is drawn is normal or the sample size is sufficiently large. In addition, the distribution of $\bar{x}$ has the same mean as the parent population, $\mu_{\bar{x}} = \mu$, and a standard deviation equal to the parent population's standard deviation divided by the square root of the sample size, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Following the same logic used in constructing a confidence interval about a population proportion, our confidence interval would be

$$point\ estimate \pm Margin\ of\ Error$$
$$= \bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}.$$

This presents a problem because we need to know the population standard deviation to construct this interval. It does not seem likely that we would know the population standard deviation but not know the population mean. So what can we do? A logical option is to use the sample standard deviation, s, as an estimate of σ. Then the standard deviation of the sampling distribution of $\bar{x}$ would be estimated by $\frac{s}{\sqrt{n}}$ and our confidence interval would be

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}.$$

Unfortunately, there is a problem with this approach. The sample standard deviation, s, is a statistic and therefore will vary from sample to sample. Using the normal model to determine the critical value, $z_{\frac{\alpha}{2}}$, in the margin of error does not take into account the additional variability introduced by using s in place of σ. This is not much of a problem for large samples because the variability in the sample standard deviation decreases as the sample size increases (Law of Large Numbers), but for small samples, we have a real problem.

Put another way, the z-score of $\bar{x}$, $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$, is normally distributed with mean 0 and standard deviation 1 (provided $\bar{x}$ is normally distributed). However, $\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ is not normally distributed with mean 0 and standard deviation 1. So a new model must be used to determine the margin of error in a confidence interval that accounts for this additional variability. This leads to the story of William Gosset.

In the early 1900s, William Gosset worked for the Guinness brewery. Gosset was in charge of conducting experiments at the brewery to identify the best barley variety. When working with beer, Gosset was limited to small data sets. At the time, the model used for constructing confidence intervals about a mean was the normal model, regardless of whether the population standard deviation was known. Gosset did not know the population standard deviation, so he substituted the sample standard deviation for the population standard deviation and used the formula $\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$. While doing this, he was finding that his confidence intervals did not include the population mean at the rate expected. This led Gosset to develop a model that accounts for the additional variability introduced by using s in place of σ when determining the margin of error. Guinness would not allow Gosset to publish his results under his real name (Guinness was very secretive about its brewing practices), but did allow the results to be published under a pseudonym. Gosset chose "Student."

Definition: Student's T-Distribution

We know that we can create a standard normal random variable, with the formula:

$$z = \frac{x - \mu}{s}.$$

Where x is the value of the statistic, μ is the population mean, and σ is the population standard deviation, Thus, if we wanted to created a standard normal random variable with a sample mean statistic, we would have the equation:

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

However this poses a problem, as states previously, we may not always know the value of σ, in this case, we can use students $t - distribution$, which states:

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}.$$

So with the $t - distribution$, we are assuming we do not have access to σ.

It is important to note that when using students method, there will be additional variability in the $t - distribution$ compared to the $z - distribution$

Note:-

Student's t-distribution is asymptotically normal, this means as the sample size n increases, student's t-distribution looks more and more like the standard normal distribution

Properties of the $t - distribution$

1. The t-distribution is different for different degrees of freedom

2. the t-distribution is centered at 0 and is symmetric about 0

3. The area under the curve is 1. The area around each curve to the right of 0 equals the area under the curve to the left of 0, which equals $1/2$

4. As t increases or decreases without bound, the graph approaches, but never equals, zero.

$$\lim_{t \to -\infty} f(x) = 0$$
$$\lim_{t \to \infty} f(x) = 0.$$

5. The area in the tails of the t-distribution is a little greater than the area in the tails of the standard normal distribution, because we are using s as an estimate of σ, thereby introducing further variability into the t-statistic

6. As the sample size n increases, the density curve of t get closer to the standard normal density curve. This result occurs because, as the sample size increases, the values of s get closer to the value of σ, by the Law of Large Numbers.

Construct and Interpret a Confidence Interval for a Population Mean

We are now ready to construct a confidence interval for a population mean.

Constructing a $(1-\alpha)\cdot 100\%$ Confidence Interval for μ

Provided

- sample data come from a simple random sample or randomized experiment

- sample size is small relative to the population size ($n \leqslant 0.05N$)

- the data come from a population that is normally distributed with no outliers, or the sample size is large

A $(1-\alpha)\cdot 100\%$ confidence interval for μ is given by:

$$Lower\ Bound:\ \overline{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$
$$Upper\ Bonud:\ \overline{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

.

where $t_{\frac{\alpha}{2}}$ is the critical value with $n-1$ degrees of freedom.

Because this confidence interval uses the t-distribution, it is often referred to as the t-interval.

Note:-

The margin of error is $E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

A robust procedure

Notice that a confidence interval about μ can be computed for non-normal populations even though Student's t-distribution requires a normal population. This is because the procedure for constructing the confidence interval is robust—it is accurate despite minor departures from normality.

If a data set has outliers, the confidence interval is not accurate because neither the sample mean nor the sample standard deviation is resistant to outliers. Sample data should always be inspected for serious departures from normality and for outliers. This is easily done with normal probability plots and boxplots.

Example: Constructing a confidence interval about a population mean (By Hand)

Problem: The website fueleconomy.gov allows drivers to report the miles per gallon of their vehicle. The data in Table 3 show the reported miles per gallon of 2011 Ford Focus automobiles for 16 different owners. Treat the sample as a simple random sample of all 2011 Ford Focus automobiles. Construct a 95% confidence interval for the mean miles per gallon of a 2011 Ford Focus. Interpret the interval.

Before we begin, let's keep the following equation in mind:

$$\overline{x} \pm t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}.$$

Approach: Before we begin, let's insure that the model requirements for constructing a confidence interval about a mean are satisfied by drawing a normal probability plot and boxplot.

We can confidently conclude that the data was obtained randomly, and the sample size 16 is much less than 5% of the population size

Because the sample size is small, we first need to verify that the data came from a population that is *normally distributed*, and also that the sample has not outliers. If we take a lot at the probability plot, we can see that the data is roughly linear, which insinuates that the data did indeed come from a population that is *normally distributed*, furthermore, if we examine the boxplot, it becomes obvious that there are no outliers in the data.

Using technology, we can compute the sample mean $\overline{x}$ and the sample standard deviation s

$$\overline{x} = 36.8$$
$$s = 2.92.$$

Now let's find the critical value for the $t - distribution$

$$(1 - \alpha) = 0.95$$
$$\alpha = 0.05$$
$$\frac{\alpha}{2} = 0.025$$

.

Thus: $z_{0.025} = 2.131$ By the t-distribution area in right tail table. From here, we can compute our bounds:

$$LB : \ \overline{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$
$$= 36.8 - 2.131 \cdot \frac{2.92}{\sqrt{15}}$$
$$= 35.24.$$

$$UB : \ \overline{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$
$$= 36.8 + 2.131 \cdot \frac{2.92}{\sqrt{15}}$$
$$= 38.36.$$

We are 95% confident that the mean miles per gallon of all 2011 Ford Focus cars is between 35.24 and 38.36 mpg.

Example: Constructing a confidence interval about a population mean (Using Statcrunch)

First, verify that the population is normally distributed by looking at the probability plot, and also that there are no outliers by looking at the boxplot

Statcrunch Steps

1. Stat > T Stats > One Sample > with data

2. Select column containig data

3. Input confidence level

4. Compute

Note:-

In the above example (by hand), we found the critical value to be $t_{0.025} = 2.131$ for 15 degrees of freedom, whereas $z_{0.025} = 1.96$. The t-distribution gives a larger critical value, so the width of the interval is wider. This larger critical value using Student's t-distribution is necessary to account for the increased variability due to using s as an estimate of σ.

Remember, 95% confidence refers to our confidence in the method. If we obtained 100 samples of size $n = 16$ from the population of 2011 Ford Focuses, we would expect about 95 of the samples to result in confidence intervals that include μ. We do not know whether the interval in Example 3 includes μ or does not include μ.

Determine the Sample Size Necessary for Estimating a Population Mean within a Given Margin of Error

The margin of error in constructing a confidence interval about the population mean is $E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$.

Solving this for n, we obtain $n = \left(\frac{t_{\frac{\alpha}{2}} \cdot s}{E} \right)^2$.

The problem with this formula is that the critical value $t_{\frac{\alpha}{2}}$ requires that we know the sample size to determine the degrees of freedom, $n - 1$. Obviously, if we do not know n, we cannot know the degrees of freedom.

The solution to this problem lies in the fact that the t-distribution approaches the standard normal z-distribution as the sample size increases. To convince yourself of this, look at the last few rows of Table VII and compare them to the corresponding z-scores for 95

Determining the Sample Size n

The sample size required to estimate the population mean, μ, with a level of confidence $(1 - \alpha) \times 100\%$ within a specified margin of error, E, is given by

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot s}{E} \right)^2$$

Where n is rounded up to the nearest whole number.

Example: Determining Sample Size (By Hand)

Problem: We again consider the problem of estimating the miles per gallon of a 2011 Ford Focus. How large a sample is required to estimate the mean miles per gallon within 0.5 miles per gallon with 95% confidence? Note: The sample standard deviation is s=2.92 mpg.

Approach: Use $n = \left(\frac{z_{\alpha/2} \cdot s}{E}\right)^2$ with $z_{\alpha/2} = z_{0.025} = 1.96$, $s = 2.92$, and $E = 0.5$ to find the required sample size.

Solution:

$$n = \left(\frac{z_{\alpha/2} \cdot s}{E}\right)^2 = \left(\frac{1.96 \cdot 2.92}{0.5}\right)^2 = 131.02.$$

Round 131.02 up to 132. A sample size of $n = 132$ results in an interval estimate of the population mean miles per gallon of a 2011 Ford Focus with a margin of error of 0.5 mile per gallon with 95% confidence.

Example: Determining Sample Size (With Statcrunch)

1. Find Standard Deviation

2. Stat > Z Stats > One Sample > Width/Sample Size

3. Enter Confidence Lever

4. Enter Standard Deviation

5. Enter Sample Size

6. Compute

0.3   9.3: Putting It Together: Which Procedure Do I Use?

Learning Objectives For This Section:

1. **Determine the Appropriate Confidence Interval to Construct**

Determine the Appropriate Confidence Interval to Construct

Questions to ask

1. What is the variable of interest?

2. Are the conditions for constructing the interval satisfied?

For question one. *What is the variable of interest*, there are two possible situations for our purposes.

1. Qualitative variable with two outcomes: Analyze with proportions p

2. Quantitative variable: μ

Now that we know which model we are going to construct, we need to verify that the conditions for constructing an interval are satisfied.

1. Population Proportion:

   - $n\hat{p}(1-\hat{p}) \geqslant 10$, and $n \leqslant 0.05N$

2. Mean:

   - $n \geqslant 30$
   - if $n \leqslant 30$, we need to verify that the data comes from a population that is normally distributed. Thus, we need to make a normal probability plot to verify this and make a boxplot to check for outliers.

Chapter 10

10.1: The Language of Hypothesis Testing

Learning Objectives For This Section:

1. **Determine the Null and Alternative Hypotheses**

2. **Explain Type I and Type II Errors**

3. **State Conclusions to Hypothesis Tests**

Vocab:

- A **Hypothesis** is a statement regarding a characteristic of one or more populations

- **Hypothesis Testing** is used to test statements regarding a characteristic of one or more populations **Note:** The "heart" of hypothesis testing is making an assumption about reality, and collecting sample evidence to determine whether it contradicts the assumption.

- **Null Hypothesis** is a statement to be tested. The null hypothesis is a statement of no change, no effect or no difference and is assumed true until evidence indicates otherwise

- **alternative Hypothesis** is a statement that we are trying to find evidence to support

- Left- and right-tailed tests are referred to as **one-tailed tests**.

- Reject the null hypothesis when the null hypothesis is true. This decision would be incorrect. This type of error is called a **Type I error.**

- Do not reject the null hypothesis when the alternative hypothesis is true. This decision would be incorrect. This type of error is called a **Type II error.**

- **level of significance.** is the probability of making a type-1 error

Formulas/Notation:

- **Null Hypothesis**

$$H_0.$$

- **Alternative hypothesis**

$$H_1.$$

- **Level of Significance (Probability of making a type-1 error)**

$$\alpha = P(\text{Type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$$

.

- **Probability of making a type-2 error**

$$\beta = P(\text{Type II error}) = P(\text{not rejecting } H_0 \text{ when } H_1 \text{ is true}).$$

Determine the Null and Alternative Hypotheses

Steps in hypothesis testing:

1. Make a statement regarding the nature of the population

2. Collect evidence (sample data) to test the statement

3. Analyze the data to assess the plausibility of the statement

In this chapter, there are three ways to set up the null and alternative hypotheses

1. Equal versus not equal hypothesis (two-tailed test)

$$H_0 : \text{ parameter} = \text{some value}$$
$$H_1 : \text{ parameter} \neq \text{some value.}$$

2. Equal versus less than (left-tailed test)

$$H_0 : \text{ parameter} = \text{some value}$$
$$H_1 : \text{ parameter} < \text{some value.}$$

3. Equal versus greater than (right-tailed test)

$$H_0 : \text{ parameter} = \text{some value}$$
$$H_1 : \text{ parameter} > \text{some value.}$$

Left- and right-tailed tests are referred to as one-tailed tests. Notice that in the left-tailed test, the direction of the inequality sign in the alternative hypothesis points to the left ($<$), whereas in the right-tailed test, the direction of the inequality sign in the alternative hypothesis points to the right ($>$). In all three tests, the null hypothesis contains a statement of equality.

The statement we are trying to gather evidence for, which is dictated by the researcher before any data are collected, determines the structure of the alternative hypothesis (two-tailed, left-tailed, or right-tailed). For example, the label on a can of soda states that the can contains 12 ounces of liquid. A consumer advocate would be concerned if the mean contents were less than 12 ounces, so the alternative hypothesis is $H_1 : \mu < 12$ ounces. However, a quality control engineer for the soda manufacturer would be concerned if there was too little or too much soda in the can, so the alternative hypothesis would be $H_1 : \mu \neq 12$ ounces. In both cases, the null hypothesis is a statement of no difference between the manufacturer's assertion on the label and the actual mean contents of the can, so the null hypothesis is $H_0 : \mu = 12$ ounces.

Example: Forming Hypotheses

Problem: The Medco pharmaceutical company has just developed a new antibiotic for children. Two percent of children taking competing antibiotics experience headaches as a side effect. A researcher for the Food and Drug Administration wants to know if the percentage of children taking the new antibiotic and experiencing headaches as a side effect is more than 2%.

Approach: We must determine the parameter to be tested, the statement of no change or no difference (status quo), and the statement for which we are attempting to gather evidence.

Solution:

The hypothesis deals with a population proportion, p. If the new drug is no different from competing drugs, then the proportion of individuals taking it who experience a headache will be 0.02, so the null hypothesis is $H_0 : p = 0.02$. We want to determine whether the proportion of individuals who experience a headache is more than 0.02, so the alternative hypothesis is $H_1 : p > 0.02$. This is a right-tailed test because the alternative hypothesis contains a $>$ symbol.

Example: Forming Hypotheses

Problem: The Blue Book value of a used three-year-old Chevy Corvette Z06 is \$56,130. Grant wonders if the mean price of a used three-year-old Chevy Corvette Z06 in the Miami metropolitan area is different from \$56,130.

Approach: We must determine the parameter to be tested, the statement of no change or no difference (status quo), and the statement for which we are attempting to gather evidence.

Solution:

The hypothesis deals with a population mean, μ. If the mean price of a three-year-old Corvette Z06 in Miami is no different from the Blue Book price, then the population mean in Miami will be \$56,130, so the null hypothesis is $H_0 : \mu = \$56,130$. Grant wants to know if the mean price is different from \$56,130, so the alternative hypothesis is $H_1 : \mu \neq \$56,130$. This is a two-tailed test because the alternative hypothesis contains a $\neq$ symbol.

Explain Type I and Type II Errors

Sample data is used to decide whether to reject the statement in the null hypothesis. Because this decision is based on incomplete (sample) information, there is always the possibility of making an incorrect decision. In fact, there are four possible outcomes from hypothesis testing.

FOUR OUTCOMES FROM HYPOTHESIS TESTING

1. Reject the null hypothesis when the alternative hypothesis is true. This decision would be correct.

2. Do not reject the null hypothesis when the null hypothesis is true. This decision would be correct.

3. Reject the null hypothesis when the null hypothesis is true. This decision would be incorrect. This type of error is called a **Type I error.**

4. Do not reject the null hypothesis when the alternative hypothesis is true. This decision would be incorrect. This type of error is called a **Type II error.**

Note:-

The null hypothesis is never declared "true", either reject the null hypothesis if there is strong sample evidence against the statement in the null hypothesis, or do not reject the null hypothesis if there is not strong sample evidence against the statement in the null hypothesis

The Probability of Making a Type I or Type II Error

When we studied how to construct confidence intervals, we learned that we never know whether a confidence interval contains the unknown parameter. We only know the likelihood that a confidence interval captures the parameter. Similarly, we never know whether the conclusion of a hypothesis test is correct. However, just as we place a level of confidence in the construction of a confidence interval, we can assign probabilities to making Type I or Type II errors when testing hypotheses. The following notation is commonplace:

$$\alpha = P(\text{Type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$$
$$\beta = P(\text{Type II error}) = P(\text{not rejecting } H_0 \text{ when } H_1 \text{ is true}).$$

The symbol β is the Greek letter beta (pronounced "BAY tah"). The probability of making a Type I error, α, is chosen by the researcher before the sample data are collected. This probability is called the *level of significance.*

The choice of the level of significance depends on the consequences of making a Type I error. If the consequences are severe, the level of significance should be small (say, α=0.01). However, if the consequences are not severe, a higher level of significance can be chosen (say, α=0.05 or α=0.10).

Why is the level of significance not always set at $\alpha = 0.01$? Reducing the probability of making a Type I error increases the probability of making a Type II error, β. Using our court analogy from the video explaining Figure 1, a jury is instructed that the prosecution must provide proof of guilt "beyond all reasonable doubt." This implies that we are choosing to make α small so that the probability of convicting an innocent person is very small. The consequence of the small α, however, is a large β, which means many guilty defendants will go free. For now, we are content to recognize the inverse relation between α and β. (As one goes up, the other goes down.)

State Conclusions to Hypothesis Tests

Once the decision is made whether to reject the null hypothesis, the researcher must state his or her conclusion. It is important to recognize that we never accept the null hypothesis.

Sample evidence can never prove the null hypothesis to be true. By not rejecting the null hypothesis, we are saying that the evidence indicates that the null hypothesis could be true or that the sample evidence is consistent with the statement in the null hypothesis.

10.2A: Hypothesis Tests on a Population Proportion with Simulation

Learning Objectives For This Section:

1. **Explain the Logic of the Simulation Method**

2. **Test Hypotheses about a Population Proportion Using the Simulation Method**

Vocab:

- When observed results are unlikely under the assumption that the null hypothesis is true, the result is **statistically significant** and we reject the statement in the null hypothesis.

- A model that generates data randomly under the assumption the statement in the null hypothesis is true. Call this model the **null model**.

- A **P-value** is the probability of observing a sample statistic as extreme or more extreme than the one observed under the assumption that the statement in the null hypothesis is true. Put another way, the P-value is the likelihood or probability that a sample statistic will result in a statistic such as the one obtained if the statement in the null hypothesis is true.

- When observed results are unlikely under the assumption that the null hypothesis is true, we say the result is **statistically significant** and reject the statement in the null hypothesis

Explain the Logic of the Simulation Method

Suppose a friend claims to be able to predict the color of a playing card from a standard deck of playing cards. Because a standard deck of playing cards has 26 red and 26 black cards, your friend would be expected to guess the correct card color 50

To test your friend's ability, you shuffle a deck of cards, randomly draw a card from the deck, and ask your friend to determine the color. You repeat this for a total of 40 predictions. Suppose your friend is able to correctly report the color of the card in 24 of 40 repetitions of the experiment ($\hat{p} = \frac{24}{40} = 0.6$). Do these results suggest that your friend is able to correctly determine the card color more than 50

Hypothesis Testing Using the P-value Approach

- If the probability of getting a sample statistic as extreme or more extreme than the one obtained is small under the assumption the statement in the null hypothesis is true, reject the null hypothesis.

Of course, this leads to the question, "How low does the P-value need to be in order to reject the statement in the null hypothesis?" The answer is that it depends. The researcher must decide what represents statistically significant evidence against the statement in the null hypothesis. That said, typically we use the following rule.

A Rule Of Thumb

- If the P-value < 0.05 (the level of significance), we reject the statement in the null hypothesis.

The bottom line, however, is this. The smaller the P-value, the stronger the evidence against the statement in the null hypothesis. So, if P-value < 0.01, we say "strong evidence against the statement in the null hypothesis." If $0.01 <$ P-value < 0.05, we say "evidence against the statement in the null hypothesis." For $0.05 <$ P-value < 0.1, some may say "moderate evidence against the statement in the null hypothesis."

Test Hypotheses about a Population Proportion Using the Simulation Method

The following steps may be used to test hypotheses about a population proportion using the simulation approach.

Testing Hypotheses Regarding a Population Proportion Using Simulation

1. Verify that the variable of interest in the study is qualitative with two possible outcomes.

2. Determine the null and alternative hypotheses.

3. Build a null model that generates data randomly under the assumption the statement in the null hypothesis is true.

4. Estimate the P-value from the model in Step 3.

5. State the conclusion.

Example: Testing Hypotheses about a Population Proportion: Left-Tailed Test

Problem: The two major college entrance exams that a majority of colleges accept for admission are the SAT and ACT. ACT looked at historical records and established 22 as the minimum ACT math score for a student to be considered prepared for college mathematics. [Note: "Being prepared" means there is a 75% probability of successfully completing a course such as College Algebra in college.] An official with the Illinois State Department of Education wonders whether less than half of the students in her state who take the ACT are prepared for college mathematics. She obtains a simple random sample of 500 records of students who have taken the ACT and finds that 219 are prepared for college mathematics (that is, scored at least 22 on the ACT math test). Does this represent significant evidence that less than half of Illinois students who have taken the ACT are prepared for college mathematics upon graduation from a high school?

Approach: Follow Steps 1 through 5 for testing a hypothesis about a population proportion using simulation. There are two options for conducting the simulation. Option 1 is to use the coin-flipping applet to simulate outcomes and option 2 is to use the urn applet to simulate outcomes. We present both options in the video solution below.

Solution:

1. For each student in the study, the researcher is interested in determining whether he or she is prepared for college mathematics, or not. This is a qualitative variable with two possible outcomes.

2. The goal of the research is to determine if less than half of the students in Illinois who have taken the ACT are prepared for college mathematics. Symbolically, we represent this as $p < 0.5$. The burden of proof lies in showing $p < 0.5$. We assume there is no difference between the proportion of students ready for college mathematics and the proportion not ready for college mathematics. Therefore, the statement in the null hypothesis is that $p = 0.5$, so:

$$H_0: \ p = 0.5 \quad versus \quad H_1: \ p < 0.5.$$

3. A null model that may be used to generate data randomly under the assumption the statement in the null hypothesis is true is a fair coin where $P(heads) = P(tails) = 0.5$

Statcrunch Urn:

1. Applets > Simulation > urn sampling

2. Fill urn with balls:

   - Green (Ready for college mathematics): half of the population (150,000) = 75,000
   - Red (Not ready for college mathematics): half of the population (150,000) = 75,000
   - Sampling: Number of balls to draw = 500 (sample size)
   - tally type 1 balls in sample: Number $<= 219$ (left tail test)
   - Compute: Conduct at least 1000 times, we will do it 5000 times

After viewing the results of the simulation, the count is 12, and we get the p-value 0.0024, that is, in only 12 of 5000 simulations, we observed 219 or fewer students who are ready for college mathematics assuming half are ready.

Because the P-value is low, we reject the statement in the null hypothesis

10.2B: Hypothesis Tests on a Population Proportion Using the Normal Model

Learning Objectives For This Section:

1. **Explain the Logic of Hypothesis Testing Using the Normal Model**

2. **Test Hypotheses about a Population Proportion Using the Normal Model**

3. **Test Hypotheses about a Population Proportion Using the Binomial Probability Distribution**

Formulas/Notation:

- $\mu_{\hat{p}}$

$$p_0.$$

- $\sigma_{\hat{p}}$

$$\sqrt{\frac{p_0(1 - p_0)}{n}}.$$

- **Conditions**

$$\text{Sample is simple random}$$
$$np_0(1 - p_0) \geqslant 10$$
$$and: \ n \leqslant 0.05N.$$

- **Computation of test statistic**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}.$$

Explain the Logic of Hypothesis Testing Using the Normal Model

One of the examples presented in the last section dealt with congressional districts. Recall the scenario. Prior to redistricting, the proportion of registered Republicans in a congressional district was 0.42. After redistricting, a random sample of 60 voters resulted in 22 being Republican. The goal of the research was to determine if the proportion of voters in the district registered as Republican decreased after redistricting. The hypotheses to be tested were $H_0 : p = 0.42$ versus $H_1 : p < 0.42$.

To test the hypothesis, the StatCrunch urn applet was used to simulate surveying 60 randomly selected voters from among 10,000,000 registered voters 5003 different times where the proportion of voters who were Republican is 0.42. Because hypothesis testing assumes the statement in the null hypothesis is true, $0.42 \times 10,000,000 = 4,200,000$ green balls were used to represent Republicans, and 5,800,000 red balls were used to represent non-Republicans.

The results of the simulation were used to determine if it is unusual to obtain 22 or fewer Republicans out of 60 ($\hat{p} = \frac{22}{60} = 0.367$) where the population proportion is 0.42. The following video continues that discussion.

From our study of sampling distributions, we know that the sampling distribution of the sample proportion, $\hat{p}$, is approximately normal with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, provided the following requirements are satisfied:

1. The sample is a simple random sample, or the result of a randomized experiment.

2. $np(1-p) \geqslant 10$

3. The sample values are independent of each other ($n \leqslant 0.05N$).

So, rather than using simulation to approximate a P-value, we can use a normal model to determine the P-value for a hypothesis test about a population proportion. Figure 6 shows the sampling distribution of the sample proportion from the redistricting problem (along with the shaded region that represents the P-value).

Figure
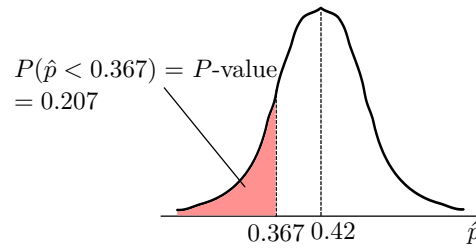
Remember, the P-value is the probability of observing a sample statistic as extreme or more extreme than the one observed under the assumption that the statement in the null hypothesis is true. Put another way, the P-value is the likelihood or probability that a sample will result in a statistic such as the one obtained if the statement in the null hypothesis is true.

If the probability of getting a sample statistic as extreme or more extreme than the one obtained is small under the assumption that the statement in the null hypothesis is true, reject the null hypothesis.

Testing Hypotheses Regarding a Population Proportion Using the Normal Model

We now formalize the procedure for testing hypotheses regarding a population proportion using the normal model.

Use steps 1-5 shown below provided that:

- The sample is a simple random sample, or the result of a randomized experiment.

- $np(1-p) \geqslant 10$

- The sample values are independent of each other ($n \leqslant 0.05N$).

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

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

3. Compute the **test statistic**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Notice in the computation of the test statistic that we are using p_0 (the proportion stated in the null hypothesis) in computing the standard error rather than $\hat{p}$, as we did in constructing confidence intervals about p. This is because H_0 is assumed to be true when we are performing a hypothesis test. So the assumed mean of the distribution of $\hat{p}$ is $\mu_{\hat{p}} = p_0$, and the assumed standard error is $\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$.

4. If P-value $< \alpha$, reject the null hypothesis.

5. State the conclusion

Example: Testing a Hypothesis about a Population Proportion: Left-Tailed Test

Problem: The two major college entrance exams that a majority of colleges accept for student admission are the SAT and ACT. ACT looked at historical records and established 22 as the minimum ACT math score for a student to be considered prepared for college mathematics. (Note: "Being prepared" means that there is a 75% probability of successfully completing College Algebra in college.) An official with the Illinois State Department of Education wonders whether less than half of the students in her state are prepared for College Algebra. She obtains a simple random sample of 500 records of students who have taken the ACT and finds that 219 are prepared for college mathematics (that is, scored at least 22 on the ACT math test). Does this represent significant evidence that less than half of Illinois students are prepared for college mathematics upon graduation from a high school? Use the $\alpha = 0.05$ level of significance. Data from ACT High School Profile Report.

Approach: The problem deals with a hypothesis test of a proportion. We want to determine whether the sample evidence suggests that less than half of the students are prepared for college mathematics. Symbolically, we represent this as $p < \frac{1}{2}$ or $p < 0.5$. Verify the three requirements to perform the hypothesis test:

1. The sample must be a simple random sample or the result of a randomized experiment.

2. $np_0(1 - p_0) \geqslant 10$

3. The sample size cannot be more than 5% of the population size (for independence).

Then follow Steps 1-5

Solution (by hand):

First, let's verify the conditions

$$p_0 = 0.5$$
$$n = 500$$
$$np_0(1 - p_0)$$
$$= 500(0.5)(1 - 0.5) = 125 \geqslant 10.$$

And we know:

$$H_0 : \ p = 0.5 \quad H_1 : \ p < 0.5 \text{ Left-Tail Test}$$
$$and : \ \alpha = 0.05.$$

Next:

$$\hat{p} = \frac{219}{500}$$
$$= 0.438.$$

Test statistic:

$$If : \ z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$
$$Then : \ z_0 = \frac{0.438 - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{500}}}$$
$$= -2.77.$$

$$P - value = P(z < -2.77)$$
$$= 0.0028.$$

Conclusion: Since the P-value is less than the level of significance, we reject the null hypothesis

Statcrunch Steps:

1. Verify Conditions

2. Determine null and alternative hypothesis, acknowledge α

3. Compute $\hat{p}$

4. Stat > Proportion Statistics > One Sample > With Summary

5. Input number of successes

6. Input number of observations

7. Input Hypothesis test for p (null and alternative)

Example: Testing a Hypothesis about a Population Proportion: Two-Tailed Test

Problem: When asked the following question, "Which do you think is more important—protecting the right of Americans to own guns or controlling gun ownership?", 46% of Americans said that protecting the right to own guns is more important. The Pew Research Center surveyed 1267 randomly selected Americans with at least a bachelor's degree and found that 559 believed that protecting the right to own guns is more important. Does this result suggest that the proportion of Americans with at least a bachelor's degree feel differently than the general American population when it comes to gun control? Use the $\alpha = 0.1$ level of significance

Approach: The problem deals with a hypothesis test of a proportion. Verify the three requirements to perform the hypothesis test. Then follow Steps 1 through 5.

Solution (by hand):

First, let's verify the conditions, to start, we know that the sample is a simple random sample.

$$If: \ p_0 = 0.46 \quad n = 1267$$
$$Then: \ np_0(1 - p_0) = 314.7 \geqslant 10.$$

We also know that

$$n \leqslant 0.05N.$$

now that we know the conditions are statisfied, let's compute $\hat{p}$ and, $\mu_{\hat{p}}$, $\sigma_{\hat{p}}$, and z_0

$$\mu_{\hat{p}} = p_0 = 0.46$$
$$\sigma_{\hat{p}} = \sqrt{\frac{0.46(0.54)}{1267}} = 0.014$$
$$\hat{p} = \frac{559}{1267} = 0.441$$
$$z_0 = \frac{.441 - .46}{\sqrt{\frac{.46(1-.46)}{1267}}} = -1.36.$$

because we are working with a two-tailed curve, we want to find the areas to the left of -1.36, and to the right of 1.36

$$P(p < -1.36) = 0.0867$$
$$2(0.0867) = 0.1738.$$

So are p-value is 0.1738, because the p-value is greater than the level of significance, we do not reject the null hypothesis

Test a Hypothesis Using a Confidence Interval

Two-Tailed Hypothesis Testing Using Confidence Intervals

- When testing $H_0 : p = p_0$ versus $H_1 : p \neq p_0$, if a $(1 - \alpha) \cdot 100\%$ confidence interval contains p_0, we do not reject the null hypothesis. However, if the confidence interval does not contain p_0, we conclude that $p \neq p_0$ at the level of significance, α.

Example: Testing a Hypothesis Using a Confidence Interval

Problem: A 2009 study by Princeton Survey Research Associates International found that 34% of teenagers text while driving. A recent study conducted by ATT found that 515 of 1200 randomly selected teens had texted while driving. Do the results of this study suggest that the proportion of teens who text while driving has changed since 2009? Use a 95% confidence interval to answer the question.

Approach: Find the 95% confidence interval. If the interval does not include 0.34, reject the null hypothesis $H_0 : p = 0.34$ in favor of $H_1 : p \neq 0.34$.

Solution: (by hand)

10.3 Hypothesis Tests for a Population Mean

Learning Objectives For This Section:

1. **Test Hypotheses about a Mean**

2. **Explain the Difference between Statistical Significance and Practical Significance**

Vocab:

- **Practical significance** refers to the idea that although small differences between the statistic and parameter stated in the null hypothesis are statistically significant, the difference may not be large enough to cause concern or be considered important.

Formula/Notations:

- **Test statistic** for hypothesis test about the mean

$$t_0 = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}.$$

Test Hypotheses about a Mean

Testing hypotheses about a mean follows the same logic as testing a hypothesis about a population proportion. The only difference is that we use Student's t-distribution rather than the normal distribution to estimate the P-value. The video explains the procedure.

Testing Hypotheses Regarding a Mean

Use Steps 1–5 shown below provided that

- The sample is obtained by simple random sampling or the data results from a randomized experiment;

- the sample comes from a population that is normally distributed with no outliers, or the sample size, n, is large $(n \geqslant 30)$; and

- the sampled values are independent of each other. This means that the sample size is no more than 5% of the population size $(n \leqslant 0.05N)$.

Steps:

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

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

3. Compute test-statistic

$$t_0 = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}.$$

- Two-Tailed:

$$p - value = 2P(t > |t_0|).$$

- Left-Tailed

$$p - value = P(t < t_0).$$

- Right-Tailed

$$p - value = P(t > t_0).$$

4. If P-value $< \alpha$, reject the null hypothesis.

5. State the conclusion.

Notice that the procedure just presented requires that the population from which the sample was drawn be normal or that the sample size be large (n $\geqslant$30). The procedure is robust, so minor departures from normality will not adversely affect the results of the test. However, if the data include outliers, the procedure should not be used.

We verify these assumptions by constructing normal probability plots (to assess normality) and boxplots (to discover whether there are outliers). If the normal probability plot indicates that the data do not come from a normal population or if the boxplot reveals outliers, nonparametric tests (not covered in this course) or the bootstrap, should be performed.

Before we look at a couple of examples, it is important to understand that we cannot find exact P-values using the t-distribution table (Table VII) because the table provides t-values only for certain areas. However, we can use the table to calculate lower and upper bounds on the P-value. To find exact P-values, we use statistical software or a graphing calculator with advanced statistical features.

Example: Testing a Hypothesis about a Population Mean: Large Sample

Problem: The mean height of American males is 69.5 inches. The heights of the 44 male U.S. presidents (Washington through Trump) have a mean of 70.84 inches and a standard deviation of 2.73 inches. Treating the 44 presidents as a simple random sample, determine whether there is evidence to suggest that U.S. presidents are taller than the average American male. Use the $\alpha = 0.05$ level of significance. (Note: Grover Cleveland was elected to two nonconsecutive terms, so technically there have been 45 presidents of the United States.)

Approach: After verifying the requirements, follow Steps 1 through 5 for testing a hypothesis about a mean.

Solution (by hand):

After verifying the requirements, we can compute our values:

$$\overline{x} = 70.84 \quad \mu_0 = 69.5$$

$$\sigma_{\overline{x}} = \frac{2.73}{\sqrt{44}} = 0.4116$$

$$t_0 = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{70.84 - 69.5}{0.4116} = 3.256.$$

Now:

$$p - value = P(t > 3.256) \quad \text{with n-1 } (44 - 1) \text{ degrees of freedom.}$$

And we find that the p-value is between 0.001 and 0.0025, that is:

$$0.001 < P - value < 0.0025.$$

Because the P-value is less than than the level of significance, we reject the null hypothesis. Thus, there is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that $U.S$ presidents are taller than the typical American male.

Explain the Difference between Statistical Significance and Practical Significance

When a large sample size is used in a hypothesis test, the results may be statistically significant even though the difference between the sample statistic and mean stated in the null hypothesis may have no practical significance.

Example: Statistical versus Practical Significance

Problem: According to the American Community Survey, the mean travel time to work in Collin County, Texas, is 27.6 minutes. The Department of Transportation reprogrammed all the traffic lights in Collin County in an attempt to reduce travel time. To determine whether there is evidence that travel time has decreased as a result of the reprogramming, the Department of Transportation obtains a random sample of 2500 commuters, records their travel time to work, and finds a sample mean of 27.3 minutes with a standard deviation of 8.5 minutes. Does this result suggest that travel time has decreased at the $\alpha = 0.05$ level of significance?

Result:

The P-value is 0.0389 (by hand: $0.025 < $ P-value $ < 0.05$). Because the P-value is less than the level of significance, $\alpha = 0.05$, we reject the null hypothesis that $\mu = 27.6$ minutes. There is sufficient evidence to conclude that the mean travel time to work has decreased. While the difference between 27.3 and 27.6 minutes is statistically significant, it has no practical meaning. After all, is 0.3 minute (18 seconds) really going to make anyone feel better about his or her commute to work?

The reason that the results from Example 3 were statistically significant had to do with the large sample size. The moral of the story is this:

Large sample sizes can lead to results that are statistically significant, whereas the difference between the statistic and parameter in the null hypothesis is not enough to be considered practically significant.

10.4: Putting It Together: Which Procedure Do I Use?

Learning Objectives For This Section:

1. **Determine the Appropriate Hypothesis Test to Perform**

Determine the Appropriate Hypothesis Test to Perform

Questions to ask yourself:

1. What is the variable of interest?

2. Is the purpose of the study to estimate the value of a parameter or to test a statement regarding a parameter?

   - Estimate: Confidence Interval
   - Test a statment: Hypothesis Test

3. Are the conditions for performing the inference satisfied?

Figure:
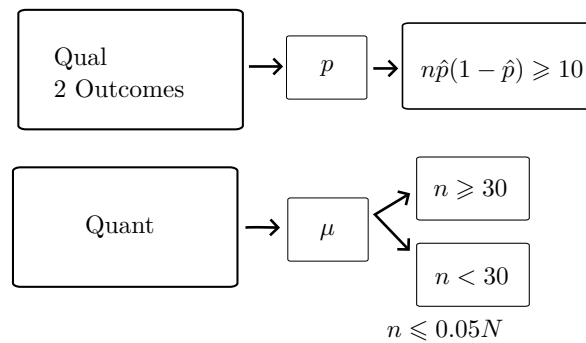
Chapter 11

11.1A: Using Randomization Techniques to Compare Two Proportions

Learning Objectives For This Section:

1. **Use Randomization to Compare Two Population Proportions**

Introduction

Earlier in the course, we focused on determining whether a sample may come from a population with a specified characteristic (such as a specified proportion of the population with a given characteristic, or a specified population mean).

Now, we shift gears to comparing the characteristics of two groups. Comparison of two groups may occur through either observational studies (such as comparing the proportion of Republicans to the proportion of Democrats in favor of some policy) or through designed experiments. Recall in a completely randomized design, a group of individuals is randomly assigned to two or more treatment groups, the treatment is imposed on the individuals, and a response variable is measured. The methods of this section apply to completely randomized designs where there are two levels of the treatment (or two distinct groups) and the response variable is qualitative with two possible outcomes. Because the response variable is qualitative with two possible outcomes, we analyze the data using proportions.

Use Randomization to Compare Two Population Proportions

To help understand the logic behind the approach that may be used to compare two groups with a qualitative response variable, we lay out a scenario. Last semester, Joliet Junior College implemented a redesign of its Intermediate Algebra course. The Math Redesign Program (MRP) is a mastery-based learning model in which students must work at a minimally prescribed pace and each module must be mastered prior to moving to the next module. The goal of the research was to determine if the MRP course results in a higher pass rate than the traditional lecture course. During the Fall, 2015 semester, Professor Egner recruited 48 volunteer students who were going to enroll in Intermediate Algebra in the Spring, 2016 semester. She randomly assigned the students to either the MRP course or traditional course. Professor Egner taught both sections. At the end of the semester, she determined whether a student passed, or failed, the course. Table 1 summarizes the results of the study in a contingency table. Does this evidence suggest the MRP course is superior to the traditional course as measured by pass rates?

Figure:

	MRP Course	Traditional Course	Total
Pass	19	14	33
Fail	5	10	15
Total	24	24	48

Notice that the response variable in the study is whether the student passed, or not. This is a qualitative variable with two possible outcomes. There are two levels of the treatment – the MRP course or the traditional course. Because the students were randomly assigned to the course, this is a completely randomized designed experiment where the experimental units are the 48 students. The assumption is that the group of students in the two courses (after random assignment, but before the treatment) are similar in terms of academic ability, attendance patterns, and other variables that may affect whether a student passes, or not. This, of course, is the beauty of randomly assigning experimental units to treatment groups – any differences in the experimental units after the treatment is imposed can be attributed to the effect of the treatment. One variable that potentially could confound the results is the time the class meets. Professor Egner, obviously, cannot teach each course at the same time. Even if one course is offered MWF at 8 am and the other at 9 am, class time could be a variable that affects the response variable. Unfortunately, there is nothing we can do about this (other than acknowledge the potential shortcoming of the study).

In this study, the proportion of students in the MRP course who passed is $\hat{p}^{\text{MRP}} = \frac{19}{24} = 0.792$ and the proportion of students in the traditional course who passed is $\hat{p}^{\text{Trad}} = \frac{14}{24} = 0.583$. The difference in sample proportions is $0.792 - 0.583 = 0.209$.

Here are the questions we have:

1. is the difference in sample proportions statistically significant, suggesting the MRP course has a higher pass rate than the traditional course?

2. Or, is it possible that the difference is due to random chance and there is no difference in the pass rates of the two courses? In other words, is it possible the 33 students who passed would have passed regardless of which course they were enrolled in, and a higher proportion of these students just happened to get placed into the MRP course?

The answers to these questions suggest two possibilities:

1. The MRP course is not effective and the higher pass rate in the MRP course was due to random chance. That is, the proportion of students who pass the MRP course equals the proportion who pass the traditional course.

2. The MRP course is effective and this explains the difference in pass rates. That is, the proportion of students who pass the MRP course is greater than the proportion of students who pass the traditional course.

We can state these two possibilities using the notation of hypothesis tests.

$$H_0 : \ p_{MRP} = p_{Trad}$$
$$H_1 : \ p_{MRP} > p_{Trad}.$$

Write each of the hypotheses as a difference in proportions as shown below.

$$H_0 : p_{\text{MRP}} - p_{\text{Trad}} = 0$$
$$H_1 : p_{\text{MRP}} - p_{\text{Trad}} > 0.$$

To answer the questions posed, assume that Statement (1), the null hypothesis, is true because this is the statement of "no effect" or "no difference". Use this statement to build the null model. In building the null model, it is important to understand that the statement in the null hypothesis is saying that the pass rate for each course is the same (or, there is no difference in the pass rate). Therefore, whether a student passes the class or not is not dependent on the course enrolled in. Put another way, whether the student passes the course or not has nothing to do with whether the student is in the traditional course or MRP course. Therefore, we build the null model by randomly assigning each student to a course type. We could do this with a physical simulation by using 33 green index cards (for the students who pass) and 15 purple index cards (for the students who do not pass). So, each index card represents a student. Shuffle the cards and then deal 24 to the traditional course and 24 to the MRP course. In each course, determine the proportion who passed and use this to determine the difference in pass rates.

So, the observed results had 19 passing in the MRP course and 14 passing in the Traditional course. After randomly assigning the students to a treatment, we had 15 passing in the MRP course and 18 passing in the Traditional course. Remember, the null model assumes the students would pass regardless of which class they were enrolled in. So, random assignment to a treatment represents the situation in which treatment given (MRP or Traditional) does not play a role in whether the student passes. The difference in pass rate after randomly assigning students to a treatment is $\hat{p}^{\text{MRP}} - \hat{p}^{\text{Trad}} = \frac{15}{24} - \frac{18}{24} = -0.125$. This random assignment suggests students did better in the traditional course.

Rather than using index cards to do a random assignment, we could use the Urn applet in StatCrunch. Let 33 green balls represent the 33 students who passed the course and let 15 red balls represent the 15 students who did not pass. So, each ball represents a student. Mix the 48 balls in the urn and randomly choose 24 balls. These 24 balls will represent the 24 students who were assigned to the MRP course.

In, StatCrunch after clicking "1 run". We notice that 17 students passed (green balls) and 7 failed (red balls). This represents the outcome of the MRP course. From the MRP course results, we can easily figure out the corresponding outcome of the traditional course. There are a total of 33 passing students, and 17 were in the MRP course, so $33 - 17 = 16$ passing students are in the traditional course. The difference in pass rates is $\frac{17}{24} - \frac{16}{24} = 0.042$.

In Run 2, we have 16 students passed (green balls) and 8 failed (red balls) in the MRP course. Therefore, $33 - 16 = 17$ students passed in the traditional course. The difference in pass rates is $\frac{16}{24} - \frac{17}{24} = -0.042$.

Each of these results represent the proportion who pass the MRP course if there is no effect due to course delivery method (that is, no difference in the pass rate of MRP and traditional courses). Neither of these random assignments resulted in a difference of pass rates as extreme, or more extreme, than the observed result of 0.209

Of course, to get a sense of how likely a difference in pass rates of 0.209 or greater is (assuming the two courses have the same pass rate), we need to repeat this random assignment many, many times. Figure 3 shows the results of 2000 more random assignments using the Urn applet. Notice the proportion of times we observed 19 or more students passing in the MRP course was 0.1114. This, of course, is the approximate P-value for the hypothesis test. It says that if the statement in the null hypothesis were true (i.e., there is no difference in pass rates in MRP versus Traditional), we would observe 19 or more passing in the MRP course in about 11 of every 100 repetitions of this study. The observed results are not unusual under the assumption the statement in the null hypothesis is true, so we do not reject the null hypothesis.

Rather than using the Urn applet to do the random assignment, we can use the "Randomization test for two proportions" applet in StatCrunch. To use this applet, you may either have raw data from a data table or summary statistics. Figure 4(a) shows a partial representation of raw data from the MRP study entered into the StatCrunch spreadsheet. In StatCrunch, select Applets > Resampling > Randomization test for two proportions. Figure 4(b) shows how to enter the results into the dialogue box. Now click Compute!. In the applet shown in Figure 4(c), click "1000 times" at least two times. The applet uses the difference of sample proportions, 0.208, as the test statistic (The difference in sample proportions of 0.208 versus 0.209 used earlier is due to rounding error). In the 2000 runs (2000 different random assignments of students to the classes), 198 resulted in a difference in sample proportions of 0.208 or higher for a P-value of 0.099. This means that if there is no difference in the proportion of students passing each class and we conducted this study 100 times, we would expect to observe a difference in sample proportions of 0.208 or higher in about 10 of the studies. The observed results are not unusual under the assumption that the statement in the null hypothesis is true. So, there is not sufficient evidence to reject the statement in the null hypothesis. Note: These results are similar to those obtained using the Urn applet.

Suppose we wanted to determine if there was a difference in the pass rates of the MRP course versus the Traditional course. Then, we would be testing the hypotheses

$$H_0: \ P_{MRP} = P_{Trad}$$
$$H_1: \ P_{MRP} \neq P_{Trad}$$

.

or, equivalently,

$$H_0: \ P_{MRP} - P_{Trad} = 0$$
$$H_1: \ P_{MRP} - P_{Trad} \neq 0$$

.

Testing Hypotheses Regarding Two Proportions Using Random Assignment

1. Verify that the explanatory variable (or treatment) has two levels and the response variable in the study is qualitative with two possible outcomes.

2. Determine the null and alternative hypotheses.

3. Build a null model that randomly assigns the individuals in the study under the assumption the statement in the null hypothesis is true.

4. Estimate the P-value from the model in Step 3.

5. State the conclusion.

Example: Testing Hypotheses Regarding Two Population Proportions

Problem: Zoloft and Trintellix are drugs meant to alleviate symptoms associated with major depressive disorder. In clinical trials of 3066 patients taking Zoloft, 790 reported nausea as a side effect. In clinical trials of 1013 patients taking Trintellix, 216 reported nausea as a side effect. Is there significant evidence to conclude a difference in the proportion of patients who report nausea as a side effect for these two drugs?

Approach: There are two groups being compared – the patients who took Zoloft and the patients who took Trintellix. The response variable is whether or not the patient reports nausea as a side effect. This is a qualitative response variable with two possible outcomes. The individuals in the study are the 3066+1013=4079 patients taking one of the two drugs.

We want to determine if the sample evidence suggests that the proportion of patients who report nausea as side effect differs between the two drugs. We will call the group that took Zoloft sample 1 and the group that took Trintellix sample 2

Solution: (Statcrunch)

$$H_0 : \ p_1 = p_2$$
$$H_1 : \ p_1 \neq p_2$$
$$\hat{p}_1 = \frac{790}{3066} = 0.258$$
$$\hat{p}_2 = \frac{216}{1013} = 0.213$$
$$\hat{p}_1 - \hat{p}_2 = 0.258 - 0.213 = 0.045.$$

Question: Could we obtain results as extreme, or more extreme than those observed, if the statement in the null hypothesis is true?

1. Applets > Resampling > Randomization test for two proportions

2. Enter from summary: Sample 1

3. Enter from summary: Sample 2

4. Compute

5. Run a couple thousand times

6. Because this is a two tailed test, observe the p-value in the *total* row

Because 0.0047 is unusual, there is sufficient evidence to conclude the proportion of Zoloft users who experience nausea as a side effect is different from the proportion of Trintellix users who experience nausea as a side effect.

11.1: Inference about Two Population Proportions: Independent Samples

Learning Objectives For This Section:

1. **Distinguish between Independent and Dependent Sampling**

2. **Test Hypotheses Regarding Two Population Proportions from Independent Samples**

3. **Construct and Interpret Confidence Intervals for the Difference between Two Population Proportions**

4. **Determine the Sample Size Necessary for Estimating the Difference between Two Population Proportions**

Vocab:

- A sampling method is **independent** when an individual selected for one sample does not dictate which individual is to be in the 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. (matched pairs samples)

Formula/Notations:

- **Sampling Distribution of the Difference between Two Proportions (Independent Sample)**

  The **sampling distribution** of $\hat{p}_1 - \hat{p}_2$, where:

  $$\hat{p}_1 = \frac{x_1}{n_2}.$$

  Is approximately normal, with mean:

  $$\mu_{\hat{p}_1 - \hat{p}_2} = \hat{p}_1 - \hat{p}_2.$$

  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:

  $$n_1 \hat{p}_1(1 - \hat{p}_1) \geqslant 10$$
  $$And : n_2 \hat{p}_2(1 - \hat{p}_2) \geqslant 10$$

  .

  And each sample is no more than 5% of the population

  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.

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

  To construct a $(1 - \alpha)\cdot100\%$ confidence interval for the difference between two population proportions from independent samples, the following requirements must be satisfied:

  - 1. The samples are obtained independently, using simple random sampling or from a completely randomized experiment with two levels of treatment.
  - $n_1 \hat{p}_1(1 - \hat{p}_1) \geqslant 10 \quad and \quad n_2 \hat{p}_2(1 - \hat{p}_2) \geqslant 10$
  - The sampled values are independent of each other. This means that each sample size is no more than 5% of the population size ($n_1 \leqslant 0.05N_1 \ and \ n_2 \leqslant 0.05N_2$).

  Provided that these requirements are met, a $(1 - \alpha)\cdot100\%$ confidence interval for $p_1 - p_2$ is given by:

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

  Notice that we do not pool the sample proportions. This is because we are not making any assumptions regarding their equality, as we did in hypothesis testing.

Introduction

Chapter 11:

1. Comparing two population proportions: Independent or dependent?

2. Comparing two population means: Dependent sample or matched-pairs design

3. Comparing two population means: Independent sample or completely randomized design with two treatments

Distinguish between Independent and Dependent Sampling

Example: Distinguishing between Independent and Dependent Sampling

Problem: Joliet Junior College decided to implement a course redesign of its developmental math program. Students enrolled in either a traditional lecture format course or a lab-based format course in which lectures and homework were done using video and the course management system MyMathLab. There were 1100 students enrolled in the traditional lecture format and 500 enrolled in the lab-based format. Once the course ended, the researchers determined whether the student passed the course. The goal of the study was to determine whether the proportion of students who passed the lab-based format exceeded that of students who passed the lecture format.

Approach: Determine whether the individuals in one group were used to determine the individuals in the other group. If so, the sampling method is dependent. If not, the sampling method is independent. Finally, consider the response variable in the study. Is it qualitative with two outcomes? If so, inferential methods based on proportions are appropriate. Is it quantitative? If so, inferential methods based on means may be appropriate.

Solution:

The sampling method is independent because the individuals in the lecture format are not related to the individuals in the lab-based format. The response variable is whether the student passed the course or not, so it is qualitative. Because there are two outcomes, pass or do not pass, the researchers can compare the proportion of students passing the lecture course to those passing the lab-based course.

Test Hypotheses Regarding Two Population Proportions from Independent Samples

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

- The samples are independently obtained using simple random sampling or the data result from a completely randomized experiment with two levels of treatment.

- $n_1 \hat{p}_1 (1 - \hat{p}_1) \geqslant 10$ and $n_2 \hat{p}_2 (1 - \hat{p}_2) \geqslant 10$.

- The sampled values are independent of each other. This means that each sample size is no more than 5% of the population size ($n_1 \leqslant 0.05 N_1$ and $n_2 \leqslant 0.05 N_2$). This ensures the independence necessary for a binomial experiment.

1. Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways: **Note:** p_1 is the population proportion for population 1, and p_2 is the population proportion for population 2.

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

3. Compute the test statistic.

$$z_o = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}, \ where \ \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}, \ \text{the pooled statistic of p..}$$

Use table V to find the P-value

4. If P-value $< \alpha$, reject the null hypothesis.

5. State the conclusion.

Example: Testing a Hypothesis Regarding Two Population Proportions

Problem: In clinical trials of Nasonex®, 3774 adult and adolescent allergy patients (patients 12 years and older) were randomly divided into two groups. The patients in group 1 (experimental group) received 200 μg of Nasonex, while the patients in group 2 (control group) received a placebo. Of the 2103 patients in the experimental group, 547 reported headaches as a side effect. Of the 1671 patients in the control group, 368 reported headaches as a side effect. It is known that over 10 million Americans who are 12 years and older are allergy sufferers. Is there significant evidence to conclude that the proportion of Nasonex users who experienced headaches as a side effect is greater than the proportion in the control group at the $\alpha = 0.05$ level of significance?

Approach: Note that this is a completely randomized experiment with two levels of treatment: Nasonex or placebo. The response variable is whether the subject experienced a headache as a side effect, which is qualitative with two outcomes. Because of this, use a hypothesis test for two independent proportions.

Solution (Statcrunch):

$$\hat{p}_1 = \frac{x_1}{n_1} = 0.260$$
$$\hat{p}_2 = \frac{x_2}{n_2} = 0.220$$
$$n\hat{p}_1(1 - \hat{p}_1) = 404.62 \geqslant 10$$
$$n\hat{p}_2(1 - \hat{p}_2) = 286.74 \geqslant 10.$$

$$H_0 : \ p_1 = p_2$$
$$H_1 : \ p_1 > p_2.$$

the level of significance is $\alpha = 0.05$

Statcrunch Steps

1. Stat > Proportion Stats > Two Sample > With Summary

2. Enter Sample 1 data

3. Enter Sample 2 data

4. Enter hypothesis test for $p_1 - p_2$

5. Compute

Because the p-value is less than the level of significance, we reject the null hypothesis.

Construct and Interpret Confidence Intervals for the Difference between Two Population Proportions

The sampling distribution of the difference of two proportions, $\hat{p}_1 - \hat{p}_2$, from independent samples can also be used to construct confidence intervals for the difference of two proportions.

Example: Constructing a Confidence Interval for the Difference between Two Population Proportions

Problem: The Gallup organization surveyed 1100 adult Americans on May 6–9, 2002, and conducted an independent survey of 1100 adult Americans on May 3–7, 2017. In both surveys, Gallup asked the following: "Right now, do you think the state of moral values in the country as a whole is getting better or getting worse?" On May 3–7, 2017, 846 of the 1100 surveyed responded that the state of moral values is getting worse; on May 6–9, 2002, 737 of the 1100 surveyed responded that the state of moral values is getting worse. Construct and interpret a 90% confidence interval for the difference between the two population proportions, $p_{2017} - p_{2002}$.

Solution: (statcrunch)

$$\hat{p}_1 = \frac{x_1}{n_1} = 0.769$$
$$\hat{p}_2 = \frac{x_2}{n_2} = 0.67$$
$$n_1 \hat{p}_1 (1 - \hat{p}_1) = 195.4 \geqslant 10$$
$$n_2 \hat{p}_2 (1 - \hat{p}_2) = 243 \geqslant 10.$$

Statcrunch steps:

1. Stat > Proportion Stats > Two Sample > With Summary

2. Enter sample 1:

3. Enter sample 2:

4. Enter confidence interval:

5. Compute:

We are 90% confident that the difference between the proportion of adult Americans who believed that the state of moral values in the country as a whole was getting worse from 2002 to 2017 is between 0.068 and 0.130

11.2 Inference about Two Population Means: Dependent Samples

Learning Objectives For This Section:

1. **Test Hypotheses for a Population Mean from Matched-Pairs Data**

2. **Construct and Interpret Confidence Intervals about a Population Mean Difference of Matched-Pairs Data**

Formula/Notations:

- **Population Mean Difference**

$$\mu_d.$$

- **Sample Mean Difference**

$$\bar{d}.$$

- **Sample Standard Deviation Difference**

$$s_d.$$

- **Testing Hypotheses Regarding the Difference between Two Population Means: Dependent Samples**

  Test Statistic:

  $$t_0 = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}.$$

  Confidence Interval for Matched-Pairs Data:

  $$Lowerbound : \bar{d} - t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$$
  $$Upperbound : \bar{d} + t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}.$$

  The critical value $t_{\alpha/2}$ is determined using $n - 1$ degrees of freedom. The values of $\bar{d}$ and s_d are the mean and standard deviation of the differenced data.

  > **Note:-**
  >
  > The interval is exact when the population is normally distributed and approximately correct for nonnormal populations, provided that n is large.

Test Hypotheses for a Population Mean from Matched-Pairs Data

Testing Hypotheses Regarding the Difference between Two Population Means: Dependent Samples

To test hypotheses regarding the mean difference of data obtained from a dependent sample (matched-pairs data), use the following steps, provided that

- The sample is obtained by simple random sampling or the data result from a matched-pairs design experiment.

- The sample data are matched-pairs (dependent).

- The differences are normally distributed with no outliers or the sample size, n, is large ($n \geqslant 30$).

- The sampled values are independent of each other [that is, the sample size is no more than 5% of the population size ($n \leqslant 0.05N$)].

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

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

3. Compute the test statistic

$$t_0 = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}.$$

4. If P-value $< \alpha$, reject the null hypothesis.

5. State the conclusion.

The Normality Requirement

The procedures just presented are robust, which means that minor departures from normality will not adversely affect the results of the test. If the data have outliers, however, the procedure should not be used.

Verify the assumption that the differenced data come from a population that is normally distributed by constructing a normal probability plot. Use a boxplot to determine whether there are outliers. If the normal probability plot indicates that the differenced data are not normally distributed or the boxplot reveals outliers, then either bootstrapping or nonparametric tests should be performed. Nonparametric tests are not covered in this course.

Example: Testing a Hypothesis for Matched-Pairs Data

Problem: Professor Andy Neill measured the time (in seconds) required to catch a falling meter stick for 12 randomly selected students' dominant hand and non-dominant hand. Professor Neill wants to know if the reaction time in an individual's dominant hand is less than the reaction time in his or her non-dominant hand. A coin flip is used to determine whether reaction time is measured using the dominant or non-dominant hand first. Conduct the test at the $\alpha = 0.05$ level of significance. The data obtained are presented in Table 1.

Approach: This is a matched-pairs design because the response variable, reaction time, is measured on the same subject for both the dominant and non-dominant hand. Because the response variable is quantitative, test the hypothesis using the difference of two dependent samples.

Solution: (statcrunch)

First verify the requirements:

- The sample is obtained randomly

- the sample data are matched pairs

- (Normal Probability Plot), QQ Plot shows linear relation

- Boxplot does not show outliers (differences boxplot)

$$H_0 : \ \mu_d = 0$$
$$H_1 : \ \mu_d < 0$$
$$\alpha = 0.05.$$

P-value(with Statcrunch):

$$t_0 = P - value = 0.009.$$

Because the P-value is less than the level of significance, we reject the null hypothesis.

there is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that the reaction time in the dominant hand is less than the reaction time in the nondominant hand

Statcrunch Steps:

1. Stat > T-Stats > Paired

Construct and Interpret Confidence Intervals about a Population Mean Difference of Matched-Pairs Data

---

Example: Constructing a Confidence Interval for Matched-Pairs Data

Problem: Using the data from Table 1, construct a 95% confidence interval estimate of the mean difference, μ_d.

Solution: (statcrunch)

First, compute the differences

1. Data > Compute Expression > Build

2. Columns: Dominant - Nondominant

3. Name: Differences

4. Compute

Next, construct the normal probability plot with the differences, and create a boxplot with the differences

1. Stat > T-stats > Paired

2. Input data

3. Enter confidence interval

4. Compute

11.3A Using Randomization Techniques to Compare Two Independent Means

Learning Objectives For This Section:

1. **Use Randomization to Compare Two Population Means**

Use Randomization to Compare Two Population Means

Suppose Professor Sullivan suspects that the female students in his Elementary Algebra course spend more time on homework than the male students. To confirm his suspicion, he randomly selects 12 male and 12 female students and records the amount of time (in minutes) on task for a particular section based on MyLab Math reports. See Table 1.

The sample mean time spent on homework for the 12 males is 80.6 minutes, while the sample mean time spent on homework for the 12 females is 99.8 minutes. The dot plot and sample means both suggest that females are spending more time on homework, on average. The sample mean difference is 99.8 - 80.6 = 19.2 minutes. The value of 19.2 is the test statistic for this study.

Is the difference in sample means of 19.2 minutes statistically significant, suggesting that females spend more time on homework than males? Or, is it possible that the difference is due to random chance and there is no difference in the study times in the two groups? There are two possibilities:

1. There is no association between gender and study times, and the difference in study times shown in Table 1 is due to random chance. That is, study time is independent of gender.

2. There is an association between gender and study times. Namely, females, on average, spend more time studying, and this explains the difference in sample means. That is, study time is not independent of gender.

We can state these two possibilities using the language of hypothesis testing. We want to know if females spend more time on homework than males. If true, then the mean time females spend on homework would exceed the mean time that males spend on homework, or $\mu_1 > \mu_2$. Of course, the statement of "no change" or "no difference" would be that $\mu_1 = \mu_2$ and we have the following null and alternative hypotheses.

$$H_0: \ \mu_1 = \mu_2$$
$$H_1: \ \mu_1 > \mu_2.$$

Write each of the hypotheses as a difference in means as shown below.

$$H_0: \ \mu_1 - \mu_2 = 0$$
$$H_1: \ \mu_1 - \mu_2 > 0$$
.

As always, the statement in the null hypothesis is used to build the null model. The logic behind building the null model is the same as it was when using randomization to compare two proportions. Statement (1) suggests that the times observed in Table 1 are independent of gender. The null model is based on this statement. Under the assumption that Statement (1) is true, we can randomly assign a gender to each time. One way this can be accomplished physically is by writing the study times on 24 index cards. Shuffle the cards and deal 12 to Group 1 – the female group. The remaining 12 cards are part of Group 2 – the male group. For each group compute the sample mean and then compute the sample mean difference. It is important to recognize that the random assignment just described conforms to Statement (1) – study time is not associated with gender. Table 2 shows the results we obtained when doing this random assignment once.

Testing Hypotheses Regarding Two Independent Means Using Random Assignment

1. Verify that the explanatory variable (or treatment) has two levels and the response variable in the study is quantitative.

2. Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways.

3. Build a null model that randomly assigns the individuals in the study to the value of the response variable under the assumption the statement in the null hypothesis is true.

4. Estimate the P-value from the model in Step 3.

5. State the conclusion

Example: Testing Hypotheses Regarding Two Independent Means

Problem: In the Spacelab Life Sciences 2 payload, 14 male rats were sent to space. Upon their return, the red blood cell mass (in milliliters) of the rats was determined. A control group of 14 male rats was held under the same conditions (except for space flight) as the space rats, and their red blood cell mass was also determined when the space rats returned. The project resulted in the data listed in Table 4. Does the evidence suggest that the flight animals have a different red blood cell mass from the control animals?

Solution: (statcrunch)

$$H_0 : \ \mu_1 - \mu_2 = 0$$
$$H_1 : \ \mu_1 - \mu_2 \neq 0.$$

We know:

$$\overline{x}_1 = 7.881 \ ml$$
$$\overline{x}_2 = 8.430 \ ml.$$

So:

$$\overline{x}_1 - \overline{x}_2 = -0.549 \ ml.$$

Statcrunch Steps:

1. Applets > resampling > Rondamization test for two means

2. Input Sample 1 and Sample 2

After running the applet a few thousand times, we can conclude that the total proportion of runs is 0.1628, therefore, if red blood cell mass and flight condition are not associated, we would expect results as extreme or more extreme than those observed in about 16 of every 100 repetitions of this study.

11.3 Inference about Two Population Means: Independent Samples

Learning Objectives For This Section:

1. **Test Hypotheses Regarding Two Population Means from Independent Samples**

2. **Construct and Interpret Confidence Intervals about the Difference of Two Independent Means**

Formula/Notations:

- **Welch's approximate t (Sampling Distribution of the Difference of Two Means: Independent Samples with Population Standard Deviations Unknown** Suppose that a simple random sample of size n_1 is taken from a population with unknown mean μ_1 and unknown standard deviation σ_1. In addition, a simple random sample of size n_2 is taken from a second population with unknown mean μ_2 and unknown standard deviation σ_2. If the two populations are normally distributed or the sample sizes are sufficiently large ($n_1 \geqslant 30$ and $n_2 \geqslant 30$), then

$$t_0 = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

  approximately follows Student's t-distribution with the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom, where $\overline{x}_1$ is the sample mean and s_1 is the sample standard deviation from population 1, and $\overline{x}_2$ is the sample mean and s_2 is the sample standard deviation from population 2.

- **degrees of freedom:** The degrees of freedom used to determine the P-value presented in the by-hand solution of Example 1 are conservative (this means we would require more evidence against the statement in the null hypothesis than if we use the formula for degrees freedom given below).

  Results that are more accurate can be obtained using the following formula for degrees of freedom:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}.$$

  When using this formula to compute degrees of freedom, round down to the nearest integer to use Table VII. For hand inference, it is recommended that you use the smaller of $n1-1$ or $n2-1$ as the degrees of freedom to ease computation. However, for increased precision in determining the P-value, computer software will use the formula above when computing the degrees of freedom.

- **Constructing a $(1-\alpha)\cdot100\%$ Confidence Interval for the Difference of Two Means**

$$Lower\ Bound: \ (\overline{x}_1 - \overline{x}_2) - t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$Upper\ Bound: \ (\overline{x}_1 + \overline{x}_2) - t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$.$$

  where $t_{\frac{\alpha}{2}}$ is computed using the smaller of n_1-1 or n_2-1 degrees of freedom or using the degrees of freedom formula.

Test Hypotheses Regarding Two Population Means from Independent Samples

Suppose we want to know whether a new experimental drug relieves symptoms attributable to the common cold. The response variable for the study is the time until the cold symptoms go away—a quantitative variable. If the drug is effective, the mean time until the cold symptoms go away should be less for individuals taking the drug than for those not taking the drug. If we let μ_1 represent the mean time until cold symptoms go away for the individuals taking the drug and μ_2 represent the mean time until cold symptoms go away for individuals taking a placebo, the null and alternative hypotheses will be:

$$H_0 : \ \mu_1 = \mu_2$$
$$H_1 : \ \mu_1 < \mu_2.$$

Or equivalently:

$$H_0 : \ \mu_1 - \mu_2 = 0$$
$$H_1 : \ \mu_1 - \mu_2 < 0.$$

To conduct this test, we might randomly divide 500 volunteers who have a common cold into two groups: an experimental group (group 1) and a control group (group 2). The control group will receive a placebo, and the experimental group will receive a predetermined amount of the experimental drug. Next, determine the time until the cold symptoms go away. Compute $\bar{x}_1$, the sample mean time until cold symptoms go away in the experimental group, and $\bar{x}_2$, the sample mean time until cold symptoms go away in the control group. Now we determine whether the difference in the sample means, $\bar{x}_1 - \bar{x}_2$, is significantly less than 0, the assumed difference stated in the null hypothesis. To do this, we need to know the sampling distribution of $\bar{x}_1 - \bar{x}_2$.

It is unreasonable to expect to know information regarding σ_1 and σ_2 without knowing information regarding the population means. Therefore, we must develop a sampling distribution for the difference of two means when the population standard deviations are unknown.

The comparison of two means with unequal (and unknown) population variances is called the Behrens–Fisher problem. While an exact method for performing inference on the equality of two means with unequal population standard deviations does not exist, an approximate solution is available. The approach that we use is known as Welch's approximate t, in honor of English statistician Bernard Lewis Welch (1911–1989).

Testing Hypotheses Regarding the Difference between Two Population Means: Independent Samples

To test hypotheses regarding two population means, μ_1 and μ_2, obtained from independent samples, use the following steps, provided that

- The two independent samples are obtained by simple random sampling or the data result from a completely randomized experiment with two treatments.

- The sample data come from a population that is normally distributed with no outliers or the sample sizes are large ($n_1 \geqslant 30$ and $n_2 \geqslant 30$).

- For each sample, the sample size is no more than 5% of the population size ($n \leqslant 0.05N$).

1. Determine the null and alternative hypotheses. The hypotheses are structured in one of three ways:

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

3. Compute the test statistic.

$$t_0 = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

   which approximately follows Student's t-distribution. Use Table VII to approximate the P-value using the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom

4. If P-value $< \alpha$, reject the null hypothesis.

5. State the conclusion.

Statcrunch Steps:

1. Stat > T Stat > Two Sample > With data

11.4: Putting It Together: Which Procedure Do I Use?

```
                                    ┌──────────┐      ┌──────────┐        ┌──────────┐
                                    │ Qual     │      │ Sampling │        │   One    │
                    ┌──────────┐    │ Two      │  →   │ Method   │        │          │
                    │ Response │ ↗  │ Outcomes │      │          │ ↗      │          │
                    │ Variable?│    └──────────┘      └──────────┘        │   Two    │
                    │          │ ↘                                        └──────────┘
                    └──────────┘    ┌──────────┐      ┌──────────┐        ┌──────────┐
                                    │          │      │ Sampling │   →    │   One    │
                                    │  Quant   │  →   │ Method   │        └──────────┘
                                    │          │      │          │   →    ┌──────────┐
                                    └──────────┘      └──────────┘        │   Two    │
                                                                          └──────────┘
                                                                           ↓        ↓
                                                                        ┌─────┐  ┌──────┐
                                                                        │ Dep │  │Indep │
                                                                        └─────┘  └──────┘
```
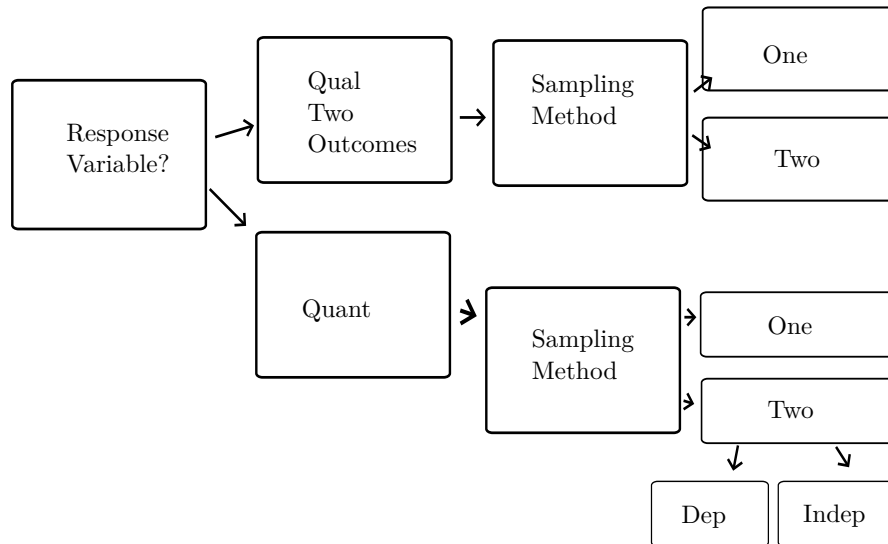
Chapter 12

12.1 Goodness-of-Fit Test

Learning Objectives For This Section:

1. **Perform a Goodness-of-Fit Test**

Vocab:

- A **goodness-of-fit** test is an inferential procedure used to determine whether a frequency distribution follows a specific distribution.

Formula/Notations:

- **Critical Value of Chi-Square distribution**

$$\chi^2_\alpha.$$

- **Expected Counts** Suppose there are n independent trials of an experiment with $k \geqslant 3$ mutually exclusive possible outcomes. Let p_1 represent the probability of observing the first outcome and E_1 represent the expected count of the first outcome; p_2 represent the probability of observing the second outcome and E_2 represent the expected count of the second outcome; and so on.

  The expected counts for each possible outcome are given by:

$$E_i = \mu_i = np_i \quad \text{for } i = 1, 2, \ldots, k.$$

- **Test Statistic for Goodness-of-Fit Tests** Let O_i represent the observed counts of category i, E_i represent the expected counts of category i, k represent the number of categories, and n represent the number of independent trials of an experiment. Then the formula

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \text{for } i = 1, 2, \ldots, k.$$

- **Conditions for test-statistic**, the test-statistic approximately follows the chi-square distribution with $k - 1$ degrees of freedom, provided that:

  - all expected frequencies are greater than or equal to 1 (all $E_{i \geqslant 1}$) and
  - no more than 20% of the expected frequencies are less than 5.

Characteristics of the Chi-Square Distribution

- It is not symmetric.

- The shape of the chi-square distribution depends on the degrees of freedom, just like Student's t-distribution.

- As the number of degrees of freedom increases, the chi-square distribution becomes more symmetric. See Figure 1.

- The values of χ^2 are nonnegative (greater than or equal to 0).

Figure:

Chi-square distributions

d.f. $= 2$

0.2

d.f. $= 5$

d.f. $= 10$

0.1

d.f. $= 15$

d.f. $= 30$

10   20   30   40   50
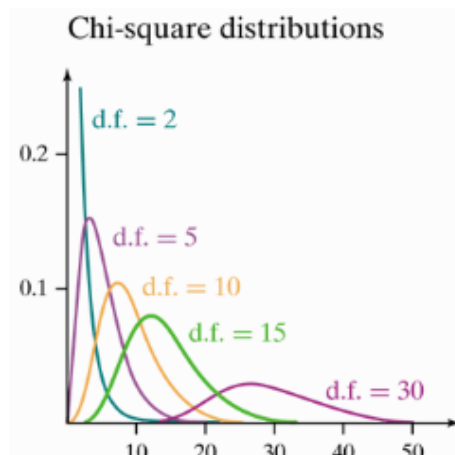
Example: Finding Critical Values for the Chi-Square Distribution

Problem: Find the critical values that separate the middle 90% of the chi-square distribution from the 5% area in each tail, assuming 15 degrees of freedom.

Approach:

1. Draw a chi-square distribution with the critical values and areas labeled.

2. Use Table VIII to find the critical values.

Solution (by hand):

First, we must determine the tails, if we want to separate the middle 90%, we now we will have 5% in each tail

So we need to find:

$$\chi^2_{0.95}, \quad \chi^2_{0.05}.$$

Using table VIII to find the areas, we find that:

$$\chi^2_{0.95} = 7.261$$
$$\chi^2_{0.05} = 24.996.$$

Solution: (By hand)

1. Stat > Calculators > Chi-Square

2. Enter degrees of freedom

3. Adjust sign and enter α

4. Compute

What If the Degrees of Freedom I Need Is Not In the Table?

In studying Table VIII, notice that the degrees of freedom are numbered 1 to 30, inclusive, then 40,50,60,…,100. If the number of degrees of freedom is not in the table, choose the degrees of freedom closest to that desired. If the degrees of freedom are exactly between two values, find the mean of the values.

For example, to find the critical value corresponding to 75 degrees of freedom, compute the mean of the critical values corresponding to 70 and 80 degrees of freedom.

Perform a Goodness-of-Fit Test

We might test whether a die is fair. This means the probability of each outcome is $\frac{1}{6}$ when a die is cast. Because we give the die the benefit of the doubt (that is, assume that the die is fair), the null hypothesis is:

$$H_0: \ p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}.$$

As another example, according to the U.S. Census Bureau in 2000, 19.0% of the U.S. population resided in the Northeast, 22.9% in the Midwest, 35.6% in the South, and 22.5% in the West. We might want to test whether the distribution of U.S. residents is the same today as it was in 2000. Since the null hypothesis is a statement of "no change," we have

$$H_0: \ \text{The distribution of residents in the U.S. is the same today as it was in 2000.}$$

The idea behind testing these types of hypotheses is to compare the actual (or observed) number of observations for each category of data with the number of observations we would expect if the null hypothesis were true. If a significant difference exists between the observed counts and expected counts, we have evidence against the statement in the null hypothesis.

The method for obtaining the expected counts is an extension of the expected value of a binomial random variable. Recall that the mean (or expected value) of a binomial random variable with n independent trials and probability of success, p, is given by $E = \mu = np$

Example: Finding Expected Counts

Problem: One growing concern regarding the U.S. economy is the inequality in the distribution of income. The data in Table 1 represent the distribution of household income for various levels of income in 2000. An economist would like to know if income distribution is changing, so she randomly selects 1500 households and obtains the household income. Find the expected number of households in each income level assuming that the distribution of income has not changed since 2000.

Income	Percent
Under $15,000	10.2
$15,000 to $24,999	9.8
$25,000 to $34,999	9.9
$35,000 to $49,999	13.3
$50,000 to $74,999	18.1
$75,000 to $99,999	13.1
$100,000 to $149,999	14.6
$150,000 to $199,999	5.8
At least $200,000	5.2

Approach: Any quantitative variable (such as income) can become qualitative by creating categories of the variable. For example, in this data, we have seven categories of income. There are two steps to follow to determine the expected counts for each category.

1. Determine the probabilities for each outcome.

2. There are $n = 1500$ trials (the 1500 households surveyed) in the experiment. We expect $np_{\text{Under \$15,000}}$ of the households surveyed to have an income under $15,000.

Solution:

The probabilities are the relative frequencies from the 2000 distribution:

$p_{\text{Under \$15,000}} = 0.102, p_{\text{\$15,000}-\text{\$24,999}} = 0.098, p_{\text{\$25,000}-\text{\$34,999}} = 0.099, p_{\text{\$35,000}-\text{\$49,999}} = 0.133,$

$p_{\text{\$50,000}-\text{\$74,999}} = 0.181, p_{\text{\$75,000}-\text{\$99,999}} = 0.131, p_{\text{\$100,000}-\text{\$149,999}} = 0.146, p_{\text{\$150,000}-\text{\$199,999}} = 0.058,$

$p_{\geqslant \text{\$200,000}} = 0.052.$

The expected counts for each income are as follows:

Expected count of Under \$15,000: $E_{\text{Under \$15,000}} = np_{\text{Under \$15,000}} = 1500 \times 0.102 = 153$
Expected count of \$15,000 to \$24,999: $E_{\text{\$15,000}-\text{\$24,999}} = np_{\text{\$15,000}-\text{\$24,999}} = 1500 \times 0.098 = 147$
Expected count of \$25,000 to \$34,999: $E_{\text{\$25,000}-\text{\$34,999}} = np_{\text{\$25,000}-\text{\$34,999}} = 1500 \times 0.099 = 148.5$
Expected count of \$35,000 to \$49,999: $E_{\text{\$35,000}-\text{\$49,999}} = np_{\text{\$35,000}-\text{\$49,999}} = 1500 \times 0.133 = 199.5$
Expected count of \$50,000 to \$74,999: $E_{\text{\$50,000}-\text{\$74,999}} = np_{\text{\$50,000}-\text{\$74,999}} = 1500 \times 0.181 = 271.5$
Expected count of \$75,000 to \$99,999: $E_{\text{\$75,000}-\text{\$99,999}} = np_{\text{\$75,000}-\text{\$99,999}} = 1500 \times 0.131 = 196.5$
Expected count of \$100,000 to \$149,999: $E_{\text{\$100,000}-\text{\$149,999}} = np_{\text{\$100,000}-\text{\$149,999}} = 1500 \times 0.146 = 219$
Expected count of \$150,000 to \$199,999: $E_{\text{\$150,000}-\text{\$199,999}} = np_{\text{\$150,000}-\text{\$199,999}} = 1500 \times 0.058 = 87$
Expected count of At least \$200,000: $E_{\geqslant \text{\$200,000}} = np_{\geqslant \text{\$200,000}} = 1500 \times 0.052 = 78$

Of the 1500 households surveyed, the economist expects to have 153 households that earned under \$15,000, 147 that earned \$15,000 to \$24,999, and so on, assuming the distribution of income is the same today as it was in 2000.

To conduct a goodness-of-fit hypothesis test, compare the observed counts with the expected counts. If the observed counts are significantly different from the expected counts, we have evidence against the statement in the null hypothesis. To perform such a test, we need a test statistic and sampling distribution.

The Goodness-of-Fit Test:

1. Determine the null and alternative hypotheses:

    H_0 : The random variable follows a certain distribution.
    H_1 : The random variable does not follow the distribution in the null hypothesis..

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

3. a.) Calculate the expected counts, E_i, for each of the k categories: $E_i = n \cdot p_i$ for $i = 1, 2, \ldots, k$, where n is the number of trials and p_i is the probability of the ith category, assuming that the null hypothesis is true.

   b.) Verify that the requirements for the goodness-of-fit test are satisfied.
   - All expected counts are greater than or equal to 1 (all $E_i \geqslant 1$).
   - No more than 20% of the expected counts are less than 5.

   c.) Compute the test statistic

   $$\chi_0^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

   Note: O_i is the observed count for the ith category.

   d.) Use Table VIII to approximate the P-value by determining the area under the chi-square distribution with k−1 degrees of freedom to the right of the test statistic.

4. If P-value $< \alpha$, reject the null hypothesis.

5. State the conclusion.

Example: Conducting a Goodness-of-Fit Test

Problem: One growing concern regarding the U.S. economy is the inequality in the distribution of income. The data in Table 2 represent the distribution of household income in 2000. An economist would like to know if income distribution is changing, so she randomly selected 1500 households and obtained the household income. Table 3 contains the results of this survey. Does the evidence suggest that the distribution of income has changed since 2000 at the α=0.05 level of significance?

$$H_0 : \text{The distribution of income is the same as it was in the year 2000}$$
$$H_1 : \text{The distribution of income is different than it was in the year 2000}$$
$$\alpha = 0.05.$$

Verify Conditions: After computing all of the expected counts, we can conclude that all $E_i \geqslant 1$.

Statcrunch Steps: Input the data into a statcrunch spreadsheet, then:

1. Stat > Goodness-of-Fit > Chi-Square test

2. Select columns

3. Compute

A Word about Conclusions of Hypothesis Tests

Whenever we are testing a hypothesis, the evidence can never prove the statement in the null hypothesis to be true. For example, in performing the goodness-of-fit test from this section, we tested

$$H_0 \text{ the distribution of household income is the same today as in 2000.}$$

Suppose we failed to reject the null hypothesis. We would not be able to say the distribution today is the same as it was in 2000. Instead, we say that we do not have enough evidence to suggest the distribution has changed significantly.

Unfortunately, goodness-of-fit tests cannot be used to test whether sample data follow a specific distribution. We can only say the data is consistent with a distribution stated in the null hypothesis.

12.2 Tests for Independence and the Homogeneity of Proportions

Learning Objectives For This Section:

1. **1Perform a Test for Independence**

2. **2Perform a Test for Homogeneity of Proportions**

Vocab:

- The **chi-square test for independence** is used to determine whether there is an association between a row variable and a column variable in a contingency table constructed from sample data.

  - The null hypothesis is that the variables are not associated, or independent.
  - The alternative hypothesis is that the variables are associated, or dependent.

- In a **chi-square test for homogeneity of proportions**, we test whether different populations have the same proportion of individuals with some characteristic.

Formula/Notations:

- **Expected Frequencies in a Chi-Square Test for Independence**

$$Expected\ Frequency = \frac{(Row\ Total)(Column\ Total)}{TableTotal}.$$

- **Test Statistic for the Chi-Square Test of Independence** Let O_i represent the observed number of counts in the ith cell and E_i represent the expected number of counts in the ith cell. Then

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

- **Chi-Square test for independence degrees of freedom**

$$(r-1)(c-1).$$

  Where r is the number of rows and c is the number of columns in the contingency table, provided that

  - All expected frequencies are greater than or equal to 1 and
  - No more than 20% of the expected frequencies are less than 5.

Introduction

In Section 4.4, we discussed the association between two variables by examining the conditional distribution in a contingency table. The data presented in a contingency table represent the measurement of two categorical variables on each individual in the study to describe the association between the two variables.

We now introduce an inferential method to determine whether an apparent association between two categorical variables is statistically significant.

Perform a Test for Independence

Is there a relationship between marital status and happiness? The data in Table 5 show the marital status and happiness of individuals who participated in the General Social Survey.

Tatble 5:

Sheet1

	Marital Status		
	Married	Widowed	Divorced/Separated
Very Happy	600	63	112
Pretty Happy	720	142	355
Not Too Happy	93	51	119

Happiness

For the data in this contingency table, we are looking at counts of a single group of individuals categorized based on the row variable "happiness" and column variable "marital status." Put another way, we are measuring two variables on each individual.

It seems reasonable to wonder whether one's marital status is associated with one's level of happiness.

To determine if there is a statistically significant association between two categorical variables, we perform a chi-square test of independence.

The chi-square test for independence is used to determine whether there is an association between a row variable and a column variable in a contingency table constructed from sample data.

- The null hypothesis is that the variables are not associated, or independent.

- The alternative hypothesis is that the variables are associated, or dependent.

The idea behind testing these types of hypotheses is to compare actual (or observed) counts to the counts we would expect assuming the variables are independent (that is, if the statement in the null hypothesis were true). If a significant difference between the observed counts and expected counts exists, we take this as evidence against the statement in the null hypothesis.

To obtain the expected counts, compute the number of observations expected within each cell under the assumption of independence (again, assuming the statement in the null hypothesis is true).

The Multiplication Rule for Independent Events can be used to obtain the expected proportion of observations within each cell under the assumption of independence. Multiply each proportion by n, the sample size, to obtain the expected count within each cell. Remember, the expected value of a binomial random variable for n independent trials of a binomial experiment with probability of success p is given by $E = \mu = np$

Example:

Problem: Is there a relationship between marital status and happiness? The data in Table 5 show the marital status and happiness of individuals who participated in the General Social Survey. Compute the expected counts within each cell, assuming that marital status and happiness are independent.

Approach:

1. Compute the row and column totals.

2. Compute the relative marginal frequencies for the row variable and column variable.

3. Use the Multiplication Rule for Independent Events to compute the expected proportion of observations within each cell assuming independence.

4. Multiply the proportions by the sample size to obtain the expected counts within each cell.

Solution:

First, we need to find the probabilities, to do this, find the totals with a frequency marginal distribution. Next, find the proportion of the totals ($\frac{Total}{n}$), and then use the Multiplication rule for independent events $P(E \text{ and } F) = P(E) \cdot P(F)$ to find the proportion for each row / column characteristic.

Then, we find the expected counts. Recall that:

$$E_i = \mu_i = np.$$

The technique just outlined in the Approach from Example 1 to find the expected counts might seem rather tedious. It certainly would be more pleasant if we could determine a shortcut formula to obtain the expected counts. Consider the expected count for "very happy" and "married." This expected count was obtained by multiplying the proportion of individuals who are "very happy," the proportion of individuals who are "married," and the number of individuals in the sample. That is,

$$
\begin{aligned}
\text{Expected count} &= (\text{ proportion "very happy" }) \cdot (\text{ proportion "married" }) \cdot (\text{ sample size}) \\
&= \frac{919}{2985} \cdot \frac{1413}{2985} \cdot 2985 \\
&= \frac{919 \cdot 1413}{2985} \\
&= \frac{(\text{row total for happiness}) \, (\text{column total for marital status})}{\text{table total}}.
\end{aligned}
$$

This leads to the following general result:

Expected Frequencies in a Chi-Square Test for Independence

To find the expected frequency in a cell when performing a chi-square independence test, multiply the cell's row total by its column total and divide this result by the table total. That is,

$$Expected\ Frequency = \frac{(row\ total)(column\ total)}{tabletotal}.$$

We need a test statistic and sampling distribution to see whether the expected and observed counts are significantly different.

Test Statistic for the Chi-Square Test of Independence

Let O_i represent the observed number of counts in the i th cell and E_i represent the expected number of counts in the i th cell. Then

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

approximately follows the chi-square distribution with $(r-1)(c-1)$ degrees of freedom, where r is the number of rows and c is the number of columns in the contingency table, provided that

- all expected frequencies are greater than or equal to 1 and

- no more than 20% of the expected frequencies are less than 5 .

Now that we know the distribution of goodness-of-fit tests, we can present a method for testing hypotheses regarding the association between two variables in a contingency table.

The Chi-Square Test for Independence

To test hypotheses regarding the association between (or independence of) two variables in a contingency table, use the following steps.

1. Determine the null and alternative hypotheses:

$$H_0 : \text{ The row variable and column variable are independent.}$$
$$H_1 : \text{ The row variable and column variable are dependent..}$$

2. Decide on a level of significance, depending on the seriousness of making a Type I error.

3. (a) Calculate the expected frequencies (counts) for each cell in the contingency table.

   (b) Verify that the requirements for the test for independence are satisfied.

      i. All expected frequencies are greater than or equal to 1 (all $E_1 \geqslant 1$).
      ii. No more than 20% of the expected frequencies are less than 5.

   (c) Compute the test statistic

   (d) Use Table VIII to approximate the P-value by determining the area under the chi-square distribution with degrees of freedom to the right of the test statistic, where is the number of rows and is the number of columns. See Figure 5.

4. If P-value $< \alpha$, then reject the null hypothesis.

5. State the conclusion.

Example: Performing a Chi-Square Test for Independence

Problem: Does one's happiness depend on one's marital status? The data in Table 5 represent the marital status and disclosed level of happiness for a random sample of adult Americans who participated in the General Social Survey. Use this data to answer the question at the α=0.05 level of significance.

Happiness	Married	Widowed	Divorced/Separated	Never Married
Very Happy	600	63	112	144
Pretty Happy	720	142	355	459
Not Too Happy	93	51	119	127

Solution: (statcrunch)

$$H_0 : \text{Happiness and marital status are independent}$$
$$H_1 : \text{Happiness and marital status are dependent}$$
$$\alpha = 0.05.$$

1. Stat > Table > Contingency > With Summary

2. Select Columns

3. Select Row Label

4. Display: Expected Count

After verifying the conditions, We can then observe the test statistic and p-value

Visual Support of Conclusions from Inference

To see the association between happiness and marital status, draw bar graphs of the conditional distributions of happiness by marital status. In this regard, we are treating marital status as the explanatory variable. It seems reasonable to treat marital status as the explanatory variable because we would like to see how a change in marital status is associated with a change in level of happiness. For example, we can calculate the relative frequency of "very happy," given that an individual is "married." We repeat this for each remaining category of marital status. This is illustrated in the next example.

Perform a Test for Homogeneity of Proportions

A second type of chi-square test exists that can be used to compare the population proportions from two or more independent samples. The test is an extension of the two-sample Z-test introduced in Section 11.1, where we compared two population proportions from independent samples.

For example, we might look at the proportion of individuals who experience headaches as a side effect for a placebo group (group 1) and for one experimental group that receives 50 milligrams (mg) per day of a medication (group 2) and another that receives 100 mg per day (group 3). We assume that the proportion of individuals who experience a headache as a side effect is the same in each population (because the null hypothesis is a statement of "no difference"). So our hypotheses are

$$H_0 : p_1 = p_2 = p_3$$

H_1 : At least one of the population proportions is different from the others..

The procedures for performing a test of homogeneity are identical to those for a test of independence.

While the procedures for the test for independence and the test of homogeneity of proportions are the same, the data differ. How?

In the test for independence, we are measuring two variables (such as marital status and level of happiness) on each individual. In other words, a single population is segmented based on the value of two variables.

In the test of homogeneity of proportions, we consider whether the proportion of individuals among different populations have the same value.

So, if you have a single population in which two variables are measured on each individual to assess whether one variable might be associated with another, conduct a test of independence. If you have two or more populations in which you want to determine equality of proportions among the choices, conduct a test of homogeneity of proportions.

Example: A Test for Homogeneity of Proportions

Problem: Zocor is a drug (Merck and Co.) that is meant to reduce the level of LDL (bad) cholesterol and increase the level of HDL (good) cholesterol. In clinical trials of the drug, patients were randomly divided into three groups. Group 1 received Zocor, group 2 received a placebo, and group 3 received cholestyramine, a cholesterol-lowering drug currently available. Table 7 contains the number of patients in each group who did and did not experience abdominal pain as a side effect. Is there evidence to indicate that the proportion of subjects in each group who experienced abdominal pain is different at the $\alpha = 0.01$ level of significance?

	Group 1 (Zocor)	Group 2 (placebo)	Group 3 (cholestyramine)
Number of people who experienced abdominal pain	51	5	16
Number of people who did not experience abdominal pain	1532	152	163

Solution: (statcrunch)

$$H_0: \ p_1 = p_2 = p_3$$
$$H_1: \ \text{At least one of the proportions is different than the others.}$$

1. Stat > Tables > Contingency > With Summary

2. Enter Columns

3. Enter Row Label

4. Display: Expected Count

5. Compute

So we can see that the conditions have been met, and the p-value is less than α. So we reject H_0

If Model Requirements Fail

The requirements for performing a chi-square test are that all expected frequencies are greater than or equal to 1 and that at most 20% of the expected frequencies can be less than 5.

If these requirements are not satisfied, the researcher has one of two options:

1. combine two or more columns (or rows) to increase the expected frequencies or

2. increase the sample size.