Chapter 8: Hypothesis Testing

8.1 Review & Preview

Inferential statistics involve using sample data to…

1.
2.

8.2 Basics of Hypothesis Testing

Part I: Basic Concepts of Hypothesis Testing

Hypothesis:

Hypothesis test (or test of significance):

Examples of typical hypotheses:

- the mean body temperature of humans is less than 98.6°F
  - using mathematical notation:
- the XSORT method of gender selection increases the probability that a baby born will be a girl, so the probability of a girl is greater than 0.5
  - using mathematical notation:
- the population of college students has IQ scores with a standard deviation equal to 15
  - using mathematical notation:

Rare Event Rule for Inferential Statistics

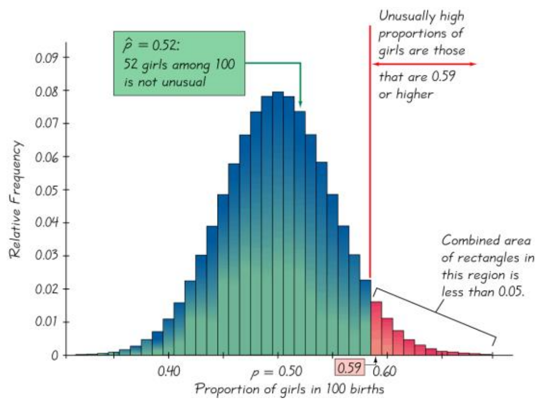
→ *To analyze sample data to test a claim we choose between the following 2 explanations:*

1. The sample results COULD easily occur by chance

Ex: In testing the XSORT gender selection method that is supposed to make babies more likely to be girls, the result of 52 girls in 100 births is greater than 50% but 52 girls could easily occur by chance, so there is not sufficient evidence to conclude that the XSORT method is effective.

2. The sample results are NOT LIKELY to occur by chance

Ex: In testing the XSORT gender-selection method, the result of 95 girls in 100 births is greater than 50% and 95 girls is so extreme that it could not easily occur by chance so there is sufficient evidence to conclude that the XSORT method is effective.



Ex: Assume that 100 babies are born to 100 couples treated with XSORT method of gender selection that is claimed to make girls more likely. If 58 of the 100 babies born are girls, test the claim that "with the XSORT method, the proportion of girls is greater than the proportion of 0.5 that occurs without any treatment."

- Using p to denote the proportion of girls born with XSORT method, test the claim that $p > 0.5$.

- Given the probability of getting 58 or more girls = 0.0548

Components of a Formal Hypothesis Test

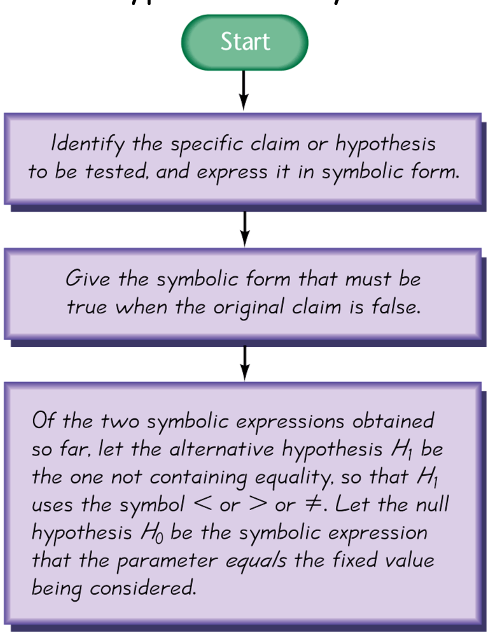
- *Null Hypothesis:*

  → denoted by H_0
  → We test the null hypothesis directly in the sense that we assume it is true and reach a conclusion to either reject H_0 or fail to reject H_0.
  → Ex: $H_0: p = 0.5$ $H_0: \mu = 98.6$ $H_0: \sigma = 15$

- *Alternative Hypothesis:*

  → denoted by H_1, H_a, or H_A
  → must use one of these symbols: $<$, $>$, or $\neq$
  → Ex: $H_1: p > 0.5$ $H_1: \mu < 98.6$ $H_1: \sigma \neq 15$

♦ Ex: Use the given claims to express the corresponding null and alternative hypothesis in symbolic form.



a. The proportion of workers who get jobs through networking is greater than 0.5.

Claim:

H_1:
H_0:

b. The mean weight of airline passengers with carry on baggage is at most 195 lb.
Claim:

H_1:
H_0:

c. The standard deviation of IQ scores of actors is equal to 15.
Claim:

H_1: **H_0:**

- *Forming Your Own Claims (hypotheses):*
  If you are conducting a study and want to use a hypothesis test to support your claim, the claim must be worded so that it becomes the alternative hypothesis (and can be expressed using only the symbols <, >, or ≠).
  *You can never support a claim that some parameter is EQUAL to a specified value.

- *Significance level (denoted by α):*

  → If the test statistic falls in the critical region, we reject the null hypothesis, so α is the probability of making the mistake of rejecting the null hypothesis when it is true.
  → Most common choices for α are ________________
  (They correspond with confidence levels of 90%, 95%, and 99%)

- *Test Statistic:*

Parameter	**Sampling Distribution**	**Requirements**	**Test Statistic**
Proportion p		np ≥ 5 and nq ≥ 5	$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$
Mean μ		σ is unknown and normally distributed population OR σ is unknown and n > 30	$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$
Mean μ		σ is known and normally distributed population OR σ is known and n > 30	$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
Standard deviation σ or variance σ^2		normally distributed population (strict requirement)	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

- *Critical region (or rejection region):*

→ Ex: A survey of n = 703 randomly selected workers showed that 61% ($\hat{p}$ = 0.61) of those respondents found their job through networking. Find the value of the test statistic for the claim that most workers (more than 50%) get their jobs through networking.

→ Ex #2: A high school sampled 200 students and found the mean number of days absent per year was 36 and the standard deviation was 4 days. Find the value of the test statistic for the district's claim that the mean number of absences is less than 20 days per year.

→ Ex #3: Another high school attendance office sampled 150 students and found the mean number of days absent per year was 41 and the standard deviation was 3 days. Find the value of the test statistic for this district's claim that the standard deviation of days absent is greater than 4 days per year.

Completing a Hypothesis Test using the Critical Value Method:

- *Critical value (or traditional) method*:

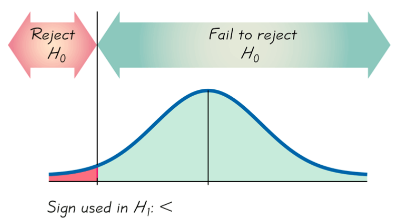
- The test statistic alone usually does not give us enough information to make a decision about the claim being tested

- We need to determine whether our hypothesis test is two-tailed, left-tailed, or right-tailed:

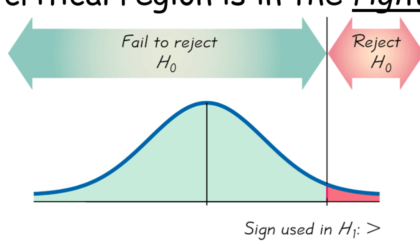
  → Depends on the nature of the null hypothesis, the sampling distribution that applies, and the significance level of α

  → Ex: Using significance level of α = 0.05, find the critical values for each of the following alternative hypothesis (assuming that the normal distribution can be used to approximate the binomial distribution).

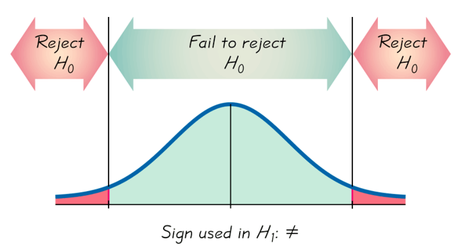
  - *Left-tailed test:* $p < 0.5$ (so the critical region is in the *left tail* of the normal distribution)



  - *Right-tailed test:* $p > 0.5$ (so the critical region is in the *right tail* of the normal distribution)



  - *Two-tailed test:* $p \neq 0.5$ (so the critical region is in *both tails* of the normal distribution)



→ Hint: The direction of the inequality symbol in the alternative hypothesis points to the tail of the critical region:

$>$ ______________ $<$ ________________

- *Conclusions in Hypothesis Testing:* We always test the null hypothesis. The initial conclusion will always be one of the following:

1.

2.

Ex: With a test statistic z = 1.60 and α = 0.05 where the critical region is everything to the right of the critical value z = 1.645, the test statistic ________fall within the critical region & we ______________________

Ex: With a test statistic z = 2.45 and α = 0.05 so the critical region is everything to the outside of the critical values of $-z_{\alpha/2}$ = _______ and $z_{\alpha/2}$ = _______ so the test statistic ________ fall within the critical region & we ___________

→ *Restate the Conclusion Using Simple & Non-technical Terms:*
--We always test the null hypothesis
--The written conclusion refers back to the original claim

Condition	Conclusion
Original claim does not include equality and you reject H_0	"There is sufficient evidence to support the claim that…(original claim)"
Original claim does not include equality and you fail to reject H_0	"There is not sufficient evidence to support the claim that…(original claim)"
Original claim includes equality and you reject H_0	"There is sufficient evidence to warrant rejection of the claim that…(original claim)"
Original claim includes equality and you fail to reject H_0	"There is not sufficient evidence to warrant rejection of the claim that …(original claim)"

--Recognize that we are not proving the null hypothesis. The term "accept" is somewhat misleading because it seems to imply that the null hypothesis has been proved. "Fail to reject" says more correctly that the available evidence isn't strong enough to warrant rejection of the null hypothesis.

--When stating the final conclusion, it's possible to get correct statements with up to 3 negative terms. Such conclusions are confusing & should be restated in a way that makes them more understandable (but does not change the meaning).

--Ex #1: Claim: $p = 0.75$ H_0: $p = 0.75$ H_1: $p \neq 0.75$
If we reject the null hypothesis, state the final conclusion:

--Ex #2: Claim: $\mu > 25$ H_0: $\mu = 25$ H_1: $\mu > 25$
If we fail to reject the null hypothesis, state the final conclusion:

Examples: First determine whether the given conditions result in a right-tailed test, a left-tailed test, or a two-tailed test, then find the P-value and the critical value(s). Use the critical value method to test the null hypothesis and state the conclusion.

a. A significance level of $\alpha = 0.05$ is used in testing the claim that $p > 0.25$, and the sample data result in a test statistic of $z = 1.18$.

   H_1:

   H_0:

   Critical value method:

   Conclusion:

b. A significance level of $\alpha = 0.05$ is used in testing the claim that $p \neq 0.25$, and the sample data result in a test statistic of $z = 2.34$

   H_1:

   H_0:

   Critical value method:

   Conclusion:

Part II: Errors in Hypothesis Tests:

→ Conclusions are sometimes correct & sometimes wrong (even if we do everything correctly)

--*Type I error (α):*

α = P(rejecting H_0 | H_0 is true)

--*Type II error (β):*

β = P(fail to reject H_0 | H_1 is true)

--Mneumonic device: use the consonants from "RouTiNe FoR FuN"
Type I error: "Reject True Null" (RTN)
Type II error: "Failure to Reject a False Null" (FRFN)

Table 8-1 Type I and Type II Errors

		True State of Nature	
		The null hypothesis is true	The null hypothesis is false
Decision	We decide to reject the null hypothesis	**Type I error** (rejecting a true null hypothesis) α	Correct decision
	We fail to reject the null hypothesis	Correct decision	**Type II error** (failing to reject a false null hypothesis) β

→ Descriptions of type I and II errors refer to the null hypothesis being true or false, but when wording a statement representing a type I or type II error, be sure that the conclusion addresses the original claim (which may or may not be the null hypothesis)

→ Ex: Assume that we are conducting a hypothesis test of the claim that $p < 0.5$. Here are the null and alternative hypotheses:

H_0: $p = 0.5$ H_1: $p < 0.5$

a. Type I error—the mistake of rejecting a true null hypothesis so this is a type I error:

b. Type II error—the mistake of failing to reject hypothesis when it is false so this is a type II error:

→ Controlling type I & II errors: Since α, β, and the sample size n are all related, when you choose or determine any two of them, the 3rd is automatically determined. The usual practice is to select the values of α and n so the value of β is determined.

1. For any fixed α, an increase in the sample size n will cause a decrease in β. (A larger sample will lessen the chance that you make the error of not rejecting the null hypothesis when it's actually false).
2. For any fixed sample size n, a decrease in α will cause an increase in β. (Conversely, an increase in α will cause a decrease in β.)
3. To decrease both α and β, increase the sample size.

-Ex: In testing the claim that μ = 0.8535 g. for M&M's, we might choose α = 0.05 and n = 100. In testing the claim that μ = 325 mg. for Bufferin brand aspirin tablets, we might choose α = 0.01 and n = 500 because of the more serious consequences associated with testing a commercial drug.

8.3 Testing a Claim about a Proportion

Claims about a population proportion are usually tested by using a normal distribution as an approximation to the binomial distribution. Instead of using the same exact methods of section 6.7, we use a different but equivalent form of the test statistic and don't include the correction for continuity (because its effect tends to be very small with large samples).

Testing Claims about a Population Proportion p:

→ *Requirements:*

1. The sample observations are a simple random sample
2. The conditions for a binomial distribution are satisfied (fixed # of trials having consistent probabilities, and each trial has 2 outcome categories—success and failure)
3. The conditions $np \geq 5$ and $nq \geq 5$ are both satisfied, so the binomial distribution of sample proportions can be approximated by a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$

    --Note that p is the ***assumed*** proportion used in the claim, not the sample proportion

→ *Notation:*

n =

$\hat{p} = \frac{x}{n}$

p =

q =

→ *Test Statistic for Testing a Claim about a Proportion:* $z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$

→ *Critical values:* use the standard normal distribution (table A-2) and find the z scores that correspond to the given significance level

Example #1: Among 703 randomly selected workers, 61% got their jobs through networking. Use the following sample data with 0.05 significance level to test the claim that most workers (more than 50%) get their jobs through networking.

Claim: Most workers get their jobs through networking ($p > 0.5$)
Sample data: n = 703 and $\hat{p}$ = 0.61

--Are the requirements satisfied?

--Critical value method of hypothesis testing:

Step 1: The original claim in symbolic form:
Step 2: The opposite of the original claim:
Step 3: H_0:
H_1:
Step 4: Use a significance level of α = ________
Step 5: The normal distribution is used for this test (because we are testing a claim about a population proportion p & the sampling distribution of sample proportions $\hat{p}$ is approximated by a normal distribution)
Step 6: Calculate the test statistic $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$ =

Step 7: This is a right-tailed test, so the area of the critical region is an area of α = ___ in the right tail. From table A-2, we find z = ________________
Step 8: Because the test statistic (z = 5.83) falls within the critical region, we ___________________________
Step 9: We conclude that

Example #2: When Gregor Mendel conducted his famous hybridization experiments with peas, one such experiment resulted in offspring consisting of 428 peas with green pods and 152 peas with yellow pods. According to Mendel's Theory, $\frac{1}{4}$ of the offspring peas should have yellow pods. Use a 0.05 significance level to test the claim that the proportion of peas with yellow pods is equal to $\frac{1}{4}$.

--Are the requirements satisfied?

--Critical value method of hypothesis testing:

8.4 Testing a Claim about a Mean

Part I: Testing Claims about a Population Mean (with σ not known):
More practical and realistic because we don't usually know the value of σ

→ *Requirements:*
1.

2.

3.

→ *Test Statistic for Testing a Claim about a Mean (with σ not known):*

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

→ *P-values & Critical Values:* use table A-3 and use df = n – 1 for the number of degrees of freedom.

→ Ex #1: Data set 20 in Appendix B includes weights of 13 red M&M candies randomly selected from a bag containing 465 M&M's. The sample weights are listed below and they have a mean of $\bar{x}$ = 0.8635 and a standard deviation of s = 0.0576. The bag states that the net weight of the contents is 396.9 g, so the M&M's must have a mean weight that is at least 396.9/465 = 0.8535 g in order to provide the amount claimed. Use the sample data with a 0.05 significance level to test the claim of a production manager that the M&M's have a mean that is actually greater than 0.8535 g, so consumers are being given more than the amount indicated on the label.

0.751	0.841	0.856	0.799	0.966	0.859
0.942	0.873	0.809	0.890	0.878	0.905
0.857					

--Check the requirements: simple random sample, we are not using a known value of σ, the sample size n = 13 is not greater than 30 but the normal quantile plot suggests the weights are normally distributed.

--*Critical value method of hypothesis testing*:

Step 1: The original claim:

Step 2: The opposite of the original claim:

Step 3: H_0:
H_1:

Step 4: Use a significance level of α = ________

Step 5: Because the claim is made about the population mean μ and because the requirements for using the t test statistic are satisfied, we use the t distribution.

Step 6: Calculate the test statistic: $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ =

The correct df = ___________
The test is right-tailed with α = ______
Use table A-3 to find the critical value of t = ___________

Step 7: Because the test statistic of t = 0.626 does not fall in the critical region, we ______________________________

Step 8: We conclude that

--Ex #2: Listed below are the measure radiation emissions (in W/kg) corresponding to a sample of cell phones. Use a 0.05 significance level to test the claim that cell phones have a mean radiation level that is less than 1.00 W/kg. Assume the sample is a simple random sample and that the data is from a population that is normally distributed.

0.38 0.55 1.54 1.55 0.50 0.60 0.92 0.96 1.00 0.86 1.46

-Critical value method:

<u>Part II: Testing Claims about a Population Mean (with σ known):</u>
Uses the normal distribution with the same components of hypothesis testing as section 8-3.

→ *<u>Requirements:</u>*
1.

2.

3.

→ *<u>Test Statistic for Testing a Claim about a Mean (with σ known):</u>* $z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$

→ *<u>Critical values:</u>* use the standard normal distribution (table A-2) and find the z scores

→ Ex #1: Data set 13 in Appendix B includes weights of 13 red M&M candies randomly selected from a bag containing 465 M&M's. The standard deviation of the weights of all M&M's in the bag is σ = 0.0565 g. The sample weights are listed below and they have a mean of $\bar{x}$ = 0.8635. The bag states that the net weight of the contents is 396.9 g, so the M&M's must have a mean weight that is at least 396.9/465 = 0.8535 g in order to provide the amount claimed. Use the sample data with a 0.05 significance level to test the claim of a production manager that the M&M's have a mean that is actually greater than 0.8535 g, so consumers are being given more than the amount indicated on the label.

0.751	0.841	0.856	0.799	0.966	0.859
0.942	0.873	0.809	0.890	0.878	0.905
0.857					

--Check the requirements: simple random sample, value of σ = 0.0565 g, the sample size n = 13 is not greater than 30 but the normal quantile plot suggests the weights are normally distributed.

--Critical value method of hypothesis testing:

Step 1: The original claim:

Step 2: The opposite of the original claim:

Step 3: H_0:
H_1:

Step 4: Use a significance level of __________

Step 5: Because σ is known and the population appears to be normally distributed, the central limit theorem indicates that the distribution of sample means can be approximated by a normal distribution.

Step 6: Calculate the test statistic: $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ =

Step 7: This is a right-tailed test, so the area of the critical region is an area of α = ___ to the right so 0.95 is the area from the left. From table A-2, we find z = _________

Step 8: Because the test statistic (z = 0.64) does not fall within the critical region, ______________________________

Step 9:

8.5 Testing a Claim about a Standard Deviation or Variance

Testing Claims about a Population Standard Deviation σ or a Population Variance σ²
Uses the chi-squared distribution from section 7-4

→ *Requirements:*
1.
2.

→ *Test Statistic for Testing a Claim about σ or σ²:* $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

where n =
s =
σ =
s^2 =
σ^2 =

→ *P-values and Critical Values:* Use table A-4 with df = n – 1 for the number of degrees of freedom

*Remember that table A-4 is based on cumulative areas from the ________

→ *Properties of the Chi-Square Distribution:*
1. All values of χ^2 are nonnegative and the distribution is not symmetric
2. There is a different χ^2 distribution for each number of degrees of freedom
3. The critical values are found in table A-4 (based on cumulative areas from the right)
   --locate the row corresponding to the appropriate number of degrees of freedom (df = n – 1)
   --the significance level α is used to determine the correct column
   --Right-tailed test: Because the area to the right of the critical value is 0.05, locate 0.05 at the top of table A-4
   --Left-tailed test: With a left-tailed area of 0.05, the area to the right of the critical value is 0.95 so locate 0.95 at the top of table A-4
   --Two-tailed test: Divide the significance level of 0.05 between the left and right tails, so the areas to the right of the two critical values are 0.975 and 0.025. Locate 0.975 and 0.025 at the top of table A-4

→ Ex #1: The industrial world shares this common goal: Improve quality by reducing variation. Quality control engineers want to ensure that a product has an acceptable mean, but they also want to produce items of consistent quality so that there will be few defects. The Newport Bottling Company had been manufacturing cans of cola with amounts having a standard deviation of 0.051 oz. A new bottling machine is tested, and a simple random sample of 24 cans results in the amounts (in ounces) listed below. (Those 24 amounts have a standard deviation of s = 0.039 oz). Use a 0.05 significance level to test the claim that cans of cola from the new machine have amounts with a standard deviation that is less than 0.051 oz.

11.98	11.98	11.99	11.98	11.90	12.02	11.99	11.93
12.02	12.02	12.02	11.98	12.01	12.00	11.99	11.95
11.95	11.96	11.96	12.02	11.99	12.07	11.93	12.05

--*Check requirements:* simple random sample, histogram & normal quantile plot show that the sample appears to come from a population with a normal distribution

--*Critical value method of testing hypotheses:*

Step 1: The original claim: ________

Step 2: If the original claim is false, then __________

Step 3: H_0:

H_1:

Step 4: The significance level is __________

Step 5: Because the claim is made about σ, we use the chi-square distribution and we have a ___________ test

Step 6: The test statistic is $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$ =

Using table A-4, df = __ and the column corresponding to ___, we find the critical value = __________

Step 7: Because the test statistic is not in the critical region, we

____

Step 8:

→ Ex #2: The Skytek Avionics company uses a new production method to manufacture aircraft altimeters. A simple random sample of new altimeters resulted in the errors listed below. Use a 0.01 significance level to test the claim that the new production method has errors with a standard deviation greater than 32.2 feet, which was the standard deviation for the old production method. If it appears that the standard deviation is greater, does the new production method appear to be better or worse than the old method? Should the company take any action?

-42 78 -22 -72 -45 15 17 51 -5 -53 -9 -109

--*<u>Critical value method of testing hypotheses:</u>*