Chapter 8 - Hypothesis Testing

In Chapters 2 and 3 we used "descriptive statistics" when we summarized data using tools such as graphs, and statistics such as the mean and standard deviation. Methods of inferential statistics use sample data to make an inference or conclusion about a population. The two main activities of inferential statistics are using sample data to (1) estimate a population parameter (such as estimating a population parameter with a confidence interval), and (2) test a hypothesis or claim about a population parameter.

In Chapter 7 we presented methods for estimating a population parameter with a confidence interval, and in this chapter, we present the method of hypothesis testing.

In statistics, a hypothesis is a claim or statement about a property of a population.

A hypothesis test (or test of significance) is a procedure for testing a claim about a property of a population.

Eg:

1. **Most people get their jobs through networks.**

   P > 0.50

2. **The average payload on trucks on highway 99 is 18000 lbs**

   $\mu = 18000$

Rare Event Rule for Inferential Statistics:

If under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

Eg:

1. Gender selection claim by a drug company: At least 80% Chance of Having a Girl if you use this Drug. A random sample of 100 Couples was studied.

Assumption 1 - The drug doesn't work. So 50% Girls & 50% Boys.

Case 1 - 52 Had Girls = Assumption 1 may be correct or it happened by chance.
Case 2 - 97 Had Girls = Assumption 1 is not correct.

Parts of a Hypothesis Tests:

1. Null Hypothesis: (H_o) - is a statement that the value of a population parameter (such as proportion, mean, or standard deviation) is equal to some claimed value. (The term null is used to indicate no change or no effect or no difference.)

Eg: $H_o : \mu = 5$, $H_o : p = 0.5$ **We test the null hypothesis directly in the sense that we assume (or pretend) it is true and reach a conclusion to either reject it or fail to reject it.**

NOTE: How to test a Hypothesis:

1. **Begin by assuming the H_o is a true statement.**
2. **Then use evidence to reach a conclusion**
   a. **Reject H_o - "I have enough evidence to prove H_o is wrong" (Note: Cannot accept H_o - We may reject the hypothesis but it doesn't necessarily mean we accept the hypothesis)**
   b. **Fail to Reject H_o: "I don't have enough evidence to prove H_o is wrong"**

2. Alternative Hypothesis: (H_1) - is the statement that the parameter (such as proportion, mean, or standard deviation) has a value that somehow differs from the null hypothesis. For the methods of this chapter, the symbolic form of the alternative hypothesis must use one of these symbols: $<$, $>$, $\neq$

Eg:

$H_1 : p < 0.53$, $H_1 : p > 0.53$, $H_1 : p \neq 0.53$

$H_1 : \mu < 12$, $H_1 : \mu > 12$, $H_1 : \mu \neq 12$

If you want to support a claim (to prove it right) you must state it as H_1 (Not H_0 because it can't determine it is right or not, it can only reject it or not reject it)

Eg: Suppose you want to prove that your fertility drug works:

Assuming drug doesn't work - $H_0 : p = 0.50$, $H_1 : p > 0.50$

How to determine H_0 & H_1:

1. **State your original claim symbolically, then state the opposite of the original claim as well. (Note: The original claim could be H_0 or H_1, depending where the equality symbol is)**

Eg:

1. "The mean of fluid is at least 12 ounces in a can"

Claim: $\mu \geq 12 \Rightarrow H_0 : \mu = 12$

Opposite: $\mu < 12 \Rightarrow H_1 : \mu < 12$

2. "The proportion of Male CEO's is Greater than 0.5" (Most meaning > 50%)

Claim: $p > 0.50 \Rightarrow H_1 : p > 0.50$

Opposite: $p \leq 0.50 \Rightarrow H_0 : p = 0.50$

3. "The mean weight of babies is at most 8.9 lbs"

Claim: $\mu \leq 8.9 \Rightarrow H_0 : \mu = 8.9$

Opposite: $\mu > 8.9 \Rightarrow H_1 : \mu > 8.9$

4. "The mean IQ score is 100"

Claim: $\mu = 100 \Rightarrow H_0 : \mu = 100$

Opposite: $\mu \neq 100 \Rightarrow H_1 : \mu \neq 100$

Test Statistic:

The test statistic is a value used in making a decision about the null hypothesis. It is found by converting the sample statistic (such as the sample proportion $\hat{p}$ the sample mean $\bar{x}$ or the sample standard deviation s) to a score (such as z, t, or χ^2) with the assumption that the null hypothesis is true.

For Proportion p:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p * q}{n}}}$$

For Mean μ:

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

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad If\ \sigma\ is\ unknown$$

Eg:

1. A sample of 706 companies found that 61% of CEO's were Male. Claim: Most CEOs are Male.

1. **Claim:** $p > 0.50 \Rightarrow H_1 : p > 0.50$ **(Most meaning more than)**
   Opposite: $p \leq 0.50 \Rightarrow H_0 : p = 0.50$

   0.50 because we are doing it for the Population.

2. **Test Stat:** $\hat{p} = 0.61,\ p = 0.50,\ q = 0.50,\ n = 706$

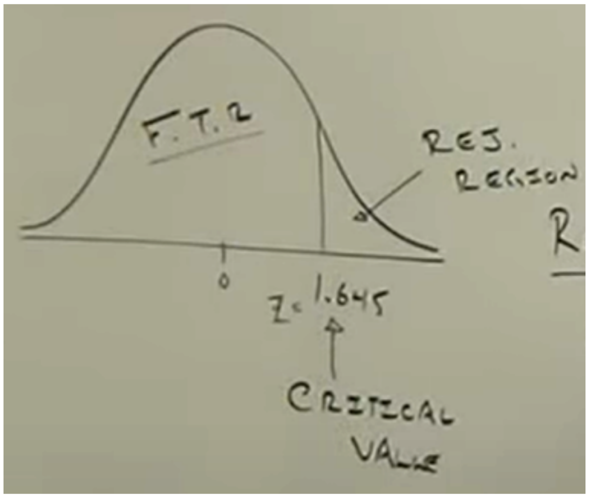
   $$z = \frac{\hat{p} - p}{\sqrt{\frac{p * q}{n}}} = \frac{\widehat{0.61} - 0.50}{\sqrt{\frac{0.50 * 0.50}{706}}} = \frac{0.11}{0.018817} = 5.845$$

   z = 5.845

3. **How to make a decision: by Significance level –** $\alpha\text{: } 0.10,\ 0.05,\ 0.01$
   Critical values – Separates rejection region from the fail to reject region.

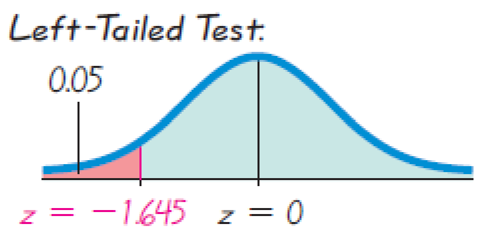
   $\alpha = 0.05 \rightarrow z = 1.645$

Rejection Region: If our test statistic falls into this region Reject H_0

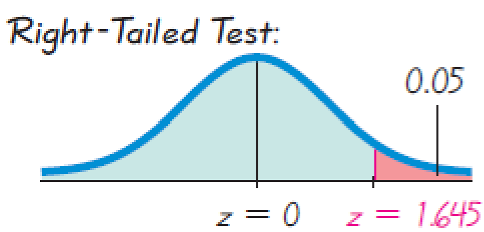


2. $\alpha = 0.05$

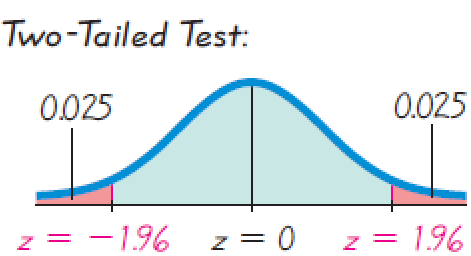
a. $H_1 : p < 0.05$



b. $H_1 : p > 0.05$

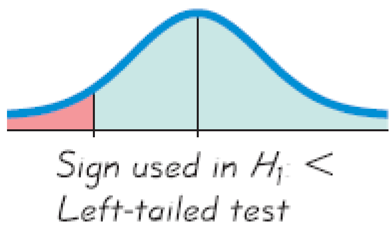


c. $H_1 : p \neq 0.05$

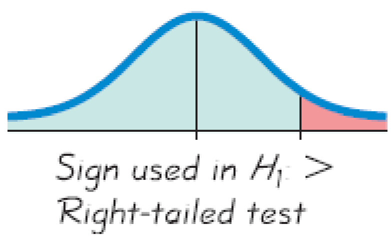


3 Types of tests: Determined by H_1

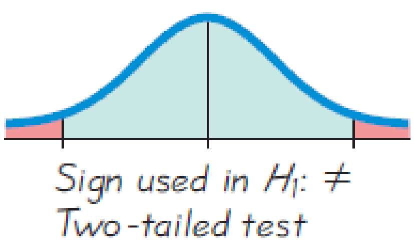
Left - Tail: If H_1 has <, Rejection Region is in left tail



Right - Tail: If H_1 has > Rejection region is in right tail



Two - Tail: If H_1 has ≠ , Rejection region is in both tails.



P-Value: **The P-value (or p-value or probability value) is the probability of getting a value of the test statistic that is at least as extreme as the one representing the sample data, assuming that the null hypothesis is true. P-values can be found after finding the area beyond the test statistic.**

Only two decisions are possible:

1. $Reject\ H_0 \rightarrow Accept\ H_1$
2. $Fail\ to\ Reject\ H_0 \rightarrow You\ know\ nothing$

Decisions and Conclusions:

The standard procedure of hypothesis testing requires that we directly test the null hypothesis, so our initial conclusion will always be one of the following:

1. Reject the null hypothesis.
2. Fail to reject the null hypothesis.

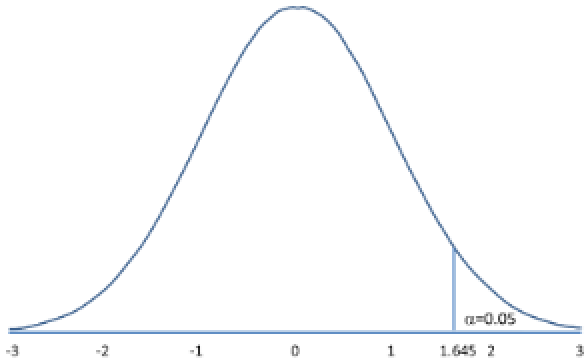
Decision Criterion - The decision to reject or fail to reject the null hypothesis is usually made using either the P-value method of testing hypotheses or the traditional method (or classical method). Sometimes, however, the decision is based on confidence intervals. In recent years, the use of the P-value method has been increasing along with the inclusion of P-values in results from software packages.

P-value method:

Using the significance level: If $p\ value \ \leq\ \alpha,\ Reject\ H_0$**, If** $p\ value\ >\ \alpha,\ Fail\ to\ Reject\ H_0$

Traditional Method:

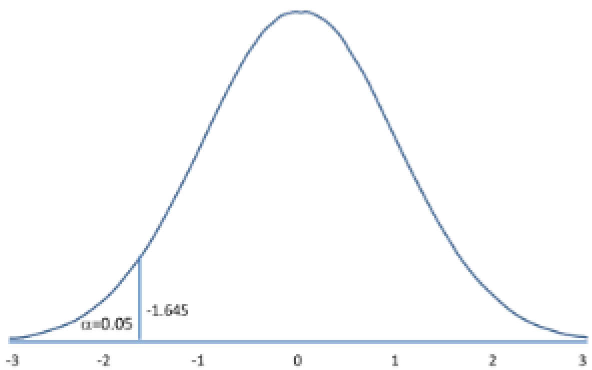
If the test statistic falls within the critical region, reject H_0 **. If the test statistic does not fall within the critical region, fail to reject** H_0**.**



Rejection Region for Upper-Tailed Z Test (H_1: $\mu > \mu_0$) with α=0.05

The decision rule is: Reject H_0 if $Z \geq 1.645$.

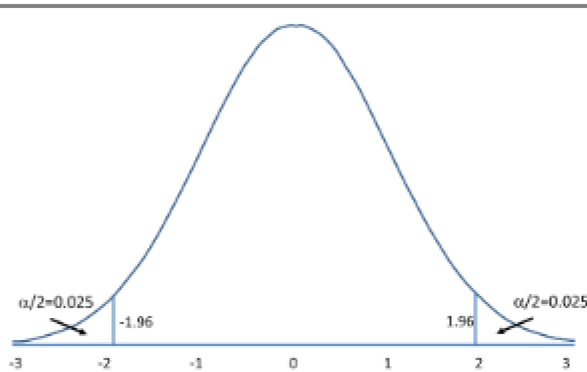
Upper-Tailed Test	
α	Z
0.10	1.282
0.05	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090
0.0001	3.719



Rejection Region for Lower-Tailed Z Test (H_1: $\mu < \mu_0$) with α =0.05

The decision rule is: Reject H_0 if $Z \leq 1.645$.

Lower-Tailed Test	
a	Z
0.10	-1.282
0.05	-1.645
0.025	-1.960
0.010	-2.326
0.005	-2.576
0.001	-3.090
0.0001	-3.719



Rejection Region for Two-Tailed Z Test (H_1: $\mu \neq \mu_0$) with α =0.05

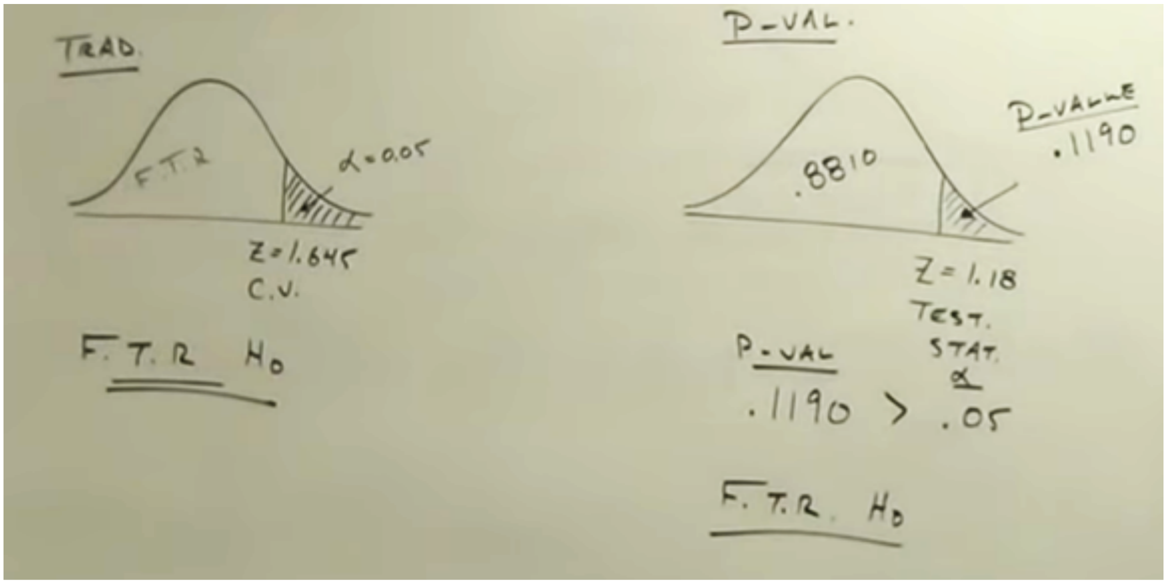
The decision rule is: Reject H_0 if $Z \leq -1.960$ or if $Z \geq 1.960$.

Two-Tailed Test	
α	Z
0.20	1.282
0.10	1.645
0.05	1.960
0.010	2.576
0.001	3.291
0.0001	3.819

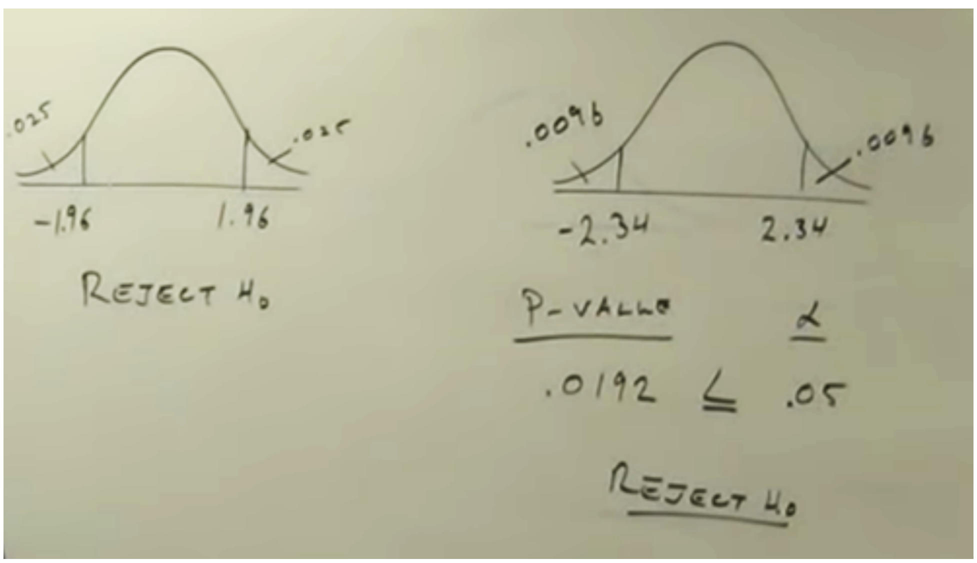
Eg:

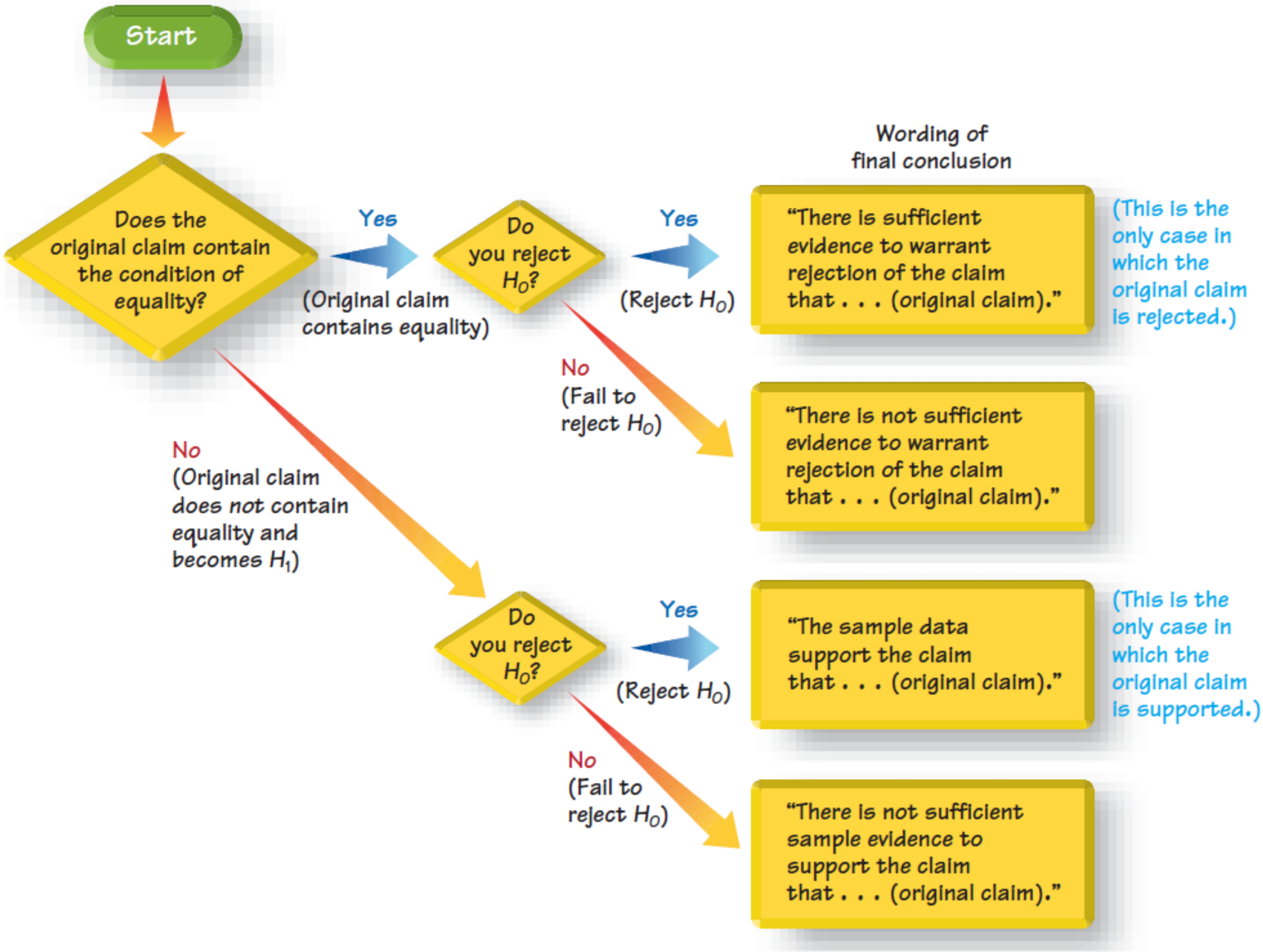
Find P - Values:

1. $\alpha = 0.05$, $H_1: p > 0.25$. **You get a test statistic of z = 1.18**



2. $\alpha = 0.05$, $H_1: p \neq 0.25$, $z = 2.34$




Start
Does the original claim contain the condition of equality?
Yes
(Original claim contains equality)
Do you reject H_0?
Yes
(Reject H_0)
Wording of final conclusion
"There is sufficient evidence to warrant rejection of the claim that . . . (original claim)."
(This is the only case in which the original claim is rejected.)
No
(Fail to reject H_0)
"There is not sufficient evidence to warrant rejection of the claim that . . . (original claim)."
No
(Original claim does not contain equality and becomes H_1)
Do you reject H_0?
Yes
(Reject H_0)
"The sample data support the claim that . . . (original claim)."
(This is the only case in which the original claim is supported.)
No
(Fail to reject H_0)
"There is not sufficient sample evidence to support the claim that . . . (original claim)."

Errors in Hypothesis Tests:

When testing a null hypothesis, we arrive at a conclusion of rejecting it or failing to reject it. Such conclusions are sometimes correct and sometimes wrong (even if we do everything correctly). The table summarises the two different types of errors that can be made, along with the two different types of correct decisions. We distinguish between the two types of errors by calling them type I and type II errors.

Type I error - The mistake of rejecting the null hypothesis when it is actually true. The symbol α (alpha) is used to represent the probability of a type I error.

Type II error - The mistake of failing to reject the null hypothesis when it is actually false. The symbol β (beta) is used to represent the probability of a type II error.

(RouTiNe FoR FuN), we can easily remember that a type I error is RTN: Reject True Null (hypothesis), whereas a type II error is FRFN: Fail to Reject a False Null (hypothesis).

α (alpha) probability of a type I error (the probability of rejecting the null hypothesis when it is true)

β (beta) probability of a type II error (the probability of failing to reject a null hypothesis when it is false)

		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) $P(\text{type I error}) = \alpha$	Correct decision
	We fail to reject the null hypothesis	Correct decision	**Type II error** (failing to reject a false null hypothesis) $P(\text{type II error}) = \beta$

P-Value Method

Start

1. Identify the specific claim or hypothesis to be tested, and put it in symbolic form.
2. Give the symbolic form that must be true when the original claim is false.
3. Of the two symbolic expressions obtained so far, let the alternative hypothesis H_1 be the one not containing equality, so that H_1 uses the symbol $>$ or $<$ or $\neq$. Let the null hypothesis H_0 be the symbolic expression that the parameter *equals* the fixed value being considered.
4. Select the significance level α based on the seriousness of a type 1 error. Make α small if the consequences of rejecting a true H_0 are severe. The values of 0.05 and 0.01 are very common.
5. Identify the statistic that is relevant to this test and determine its sampling distribution (such as normal, *t*, chi-square).
6. Find the test statistic and find the *P*-value (see Figure 8-5). Draw a graph and show the test statistic and *P*-value.
7. Reject H_0 if the *P*-value is less than or equal to the significance level α. Fail to reject H_0 if the *P*-value is greater than α.
8. Restate this previous decision in simple, nontechnical terms, and address the original claim.

Stop

Traditional Method

Start

1. Identify the specific claim or hypothesis to be tested, and put it in symbolic form.
2. Give the symbolic form that must be true when the original claim is false.
3. Of the two symbolic expressions obtained so far, let the alternative hypothesis H_1 be the one not containing equality, so that H_1 uses the symbol $>$ or $<$ or $\neq$. Let the null hypothesis H_0 be the symbolic expression that the parameter equals the fixed value being considered.
4. Select the significance level α based on the seriousness of a type 1 error. Make α small if the consequences of rejecting a true H_0 are severe. The values of 0.05 and 0.01 are very common.
5. Identify the statistic that is relevant to this test and determine its sampling distribution (such as normal, *t*, chi-square).
6. Find the test statistic, the critical values, and the critical region. Draw a graph and include the test statistic, critical value(s), and critical region.
7. Reject H_0 if the test statistic is in the critical region. Fail to reject H_0 if the test statistic is not in the critical region.
8. Restate this previous decision in simple, nontechnical terms, and address the original claim.

Stop

Confidence Interval Method

Construct a confidence interval with a confidence level selected as in Table 8-2.

Because a confidence interval estimate of a population parameter contains the likely values of that parameter, reject a claim that the population parameter has a value that is not included in the confidence interval.

Table 8-2 Confidence Level for Confidence Interval

		Two-Tailed Test	One-Tailed Test
Significance Level for Hypothesis Test	0.01	99%	98%
	0.05	95%	90%
	0.10	90%	80%

Testing a Claim About a Proportion:

Testing a Claim About a Proportion, P for claims involving percentages and proportions. (35%, 0.20, Most)

Requirements:

1. **Random Sample size**
2. $np \geq 5, nq \leq 5$
3. **n = sample size (Number of trials)**
   $\hat{p} = \frac{x}{n}$ **p = Population proportion**
   $q = 1 - p$

   $z = \frac{\hat{p} - p}{\sqrt{\frac{p*q}{n}}}$

Eg: In a sample of 300 corporations, 183 CEO's were male. Test the claim that most CEO's are male. Use a 0.05 significance level.

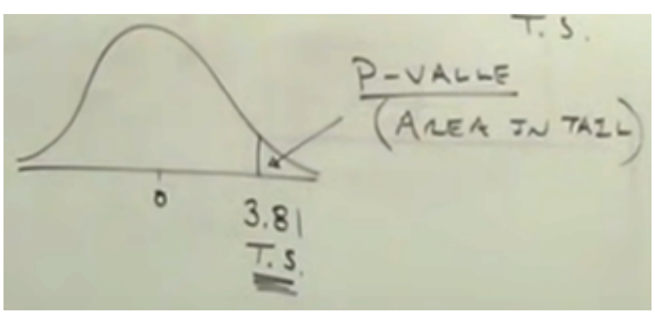
P value Method:

Claim: $p > 0.50 \Rightarrow H_1 : p > 0.50$
Opposite: $p \leq 0.50 \Rightarrow H_0 : p = 0.50$

$\alpha = 0.05$

$\hat{p} = \frac{183}{300} = 0.61,\ p = 0.50,\ q = 1 - 0.50 = 0.50$

Test statistic - $z = \frac{\hat{p} - p}{\sqrt{\frac{p*q}{n}}} = \frac{0.61 - 0.50}{\sqrt{\frac{0.50*0.50}{300}}} = \frac{0.61 - 0.50}{0.028867} = 3.810$



From table: 1 - 0.999 = 0.0001

Decision:

p-value - 0.0001
$\alpha = 0.05$

$p\ value \leq \alpha \Rightarrow Reject\ H_0$

$p\ value > \alpha \Rightarrow Failed\ to\ Reject\ H_0$

Since, $0.0001 \leq 0.05 \Rightarrow Reject\ H_0,\ H_1\ is\ true$

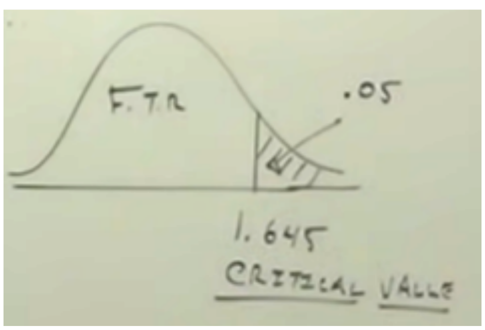
Interpretation:

If we $Reject\ H_0$ **- There is enough evidence to support the claim that most CEO's are men.**

If we $Failed\ to\ Reject\ H_0$ **- There is not enough evidence to support the claim that most CEO's are men.**

Traditional Method:

Till finding Z same steps.



Compare test stat to the critical value: If it falls in the rejection region reject H_0 **or if its in failed to reject region, failed to reject** H_0

3.81 falls is the rejection region.

Eg: Peas: 428 Green Pods, 152 yellow pods, Use a 0.05 significance level to test the claim that the proportion of yellow pods is equal to ¼.

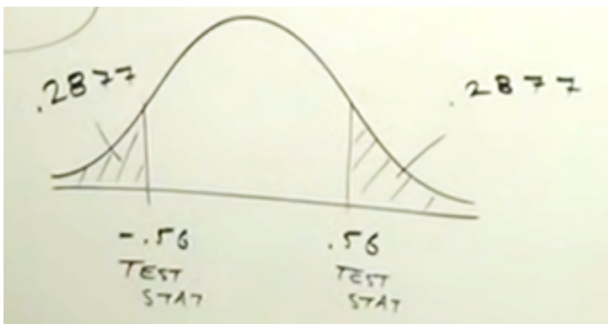
P value Method:

Claim: $p = 0.25 \Rightarrow H_0 : p = 0.25$
Opposite: $p \neq 0.25 \Rightarrow H_1 : p \neq 0.25$

$\alpha = 0.05$

$\hat{p} = \frac{152}{580} = 0.2620,\ p = 0.25,\ q = 1 - 0.25 = 0.75$

Test statistic - $z = \frac{\hat{p} - p}{\sqrt{\frac{p * q}{n}}} = \frac{0.26 - 0.25}{\sqrt{\frac{0.25 * 0.75}{580}}} = \frac{0.01}{0.01797} = 0.56$



P value = 0.5754

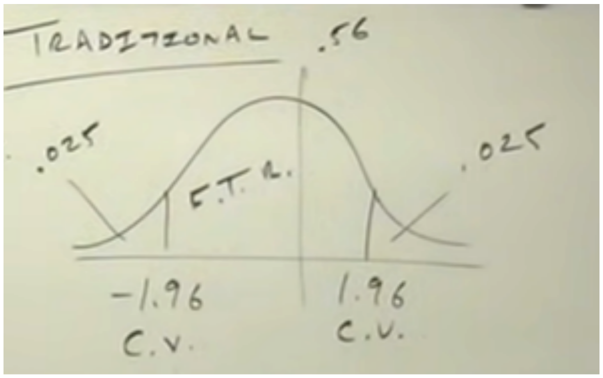
$p\ value \leq \alpha \Rightarrow Reject\ H_0$
$p\ value > \alpha \Rightarrow Failed\ to\ Reject\ H_0$

$Since,\ 0.5754 > 0.05 \Rightarrow We\ Failed\ to\ Reject\ H_0$

There is not enough evidence to support the claim that the proportion of yellow pods is equal to ¼.

Traditional Method:

Till finding Z same steps.



0.56 lies in Failed to reject region.

Testing a Claim About a Population Mean: σ Known

Requirements:

1. **Random sample size.**
2. **σ is given.**
3. $n > 30$ **or Population is normally distributed.**

Test Statistic:

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

n - sample size
$\bar{x}$ **- sample mean**
$\mu_{\bar{x}}$ **- population mean of all sample means from samples of size n (this value is based on the claim and is used in the null hypothesis)**
σ **- known value of the population standard deviation**

Eg: You sample 465 chocolates, the sample had a mean of 0.8635. The standard deviation for population is 0.0565g. Test the claim that the mean weight is greater than 0.8535g. Significance level is 0.01

Claim: $\mu > 0.8535 \Rightarrow H_1 : \mu > 0.8535$

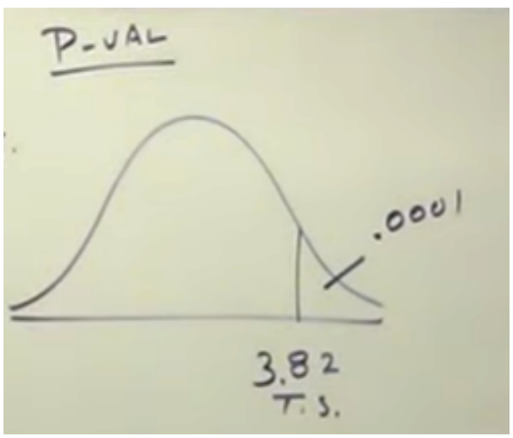
Opposite: $\mu \le 0.8535 \Rightarrow H_0 : \mu = 0.8535$

$\alpha = 0.01$

$\overline{x} = 0.8635,\ \mu = 0.8535,\ \sigma = 0.565, n = 465$

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.8635 - 0.8535}{\frac{0.565}{\sqrt{465}}} = \frac{0.8635 - 0.8535}{\frac{0.565}{21.564}} = \frac{0.01}{0.02620} = 3.82$$

P value Method:



P value = 0.0001

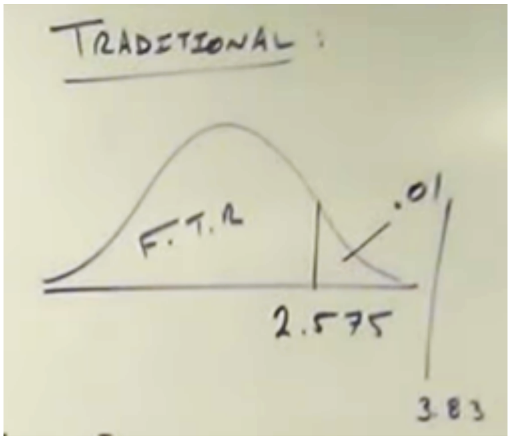
$p\ value \le \alpha \Rightarrow Reject\ H_0$

$p\ value > \alpha \Rightarrow Failed\ to\ Reject\ H_0$

$Since,\ 0.0001 \le 0.01 \Rightarrow We\ Reject\ H_0$

There is enough evidence to support the claim that the mean weight of chocolates are greater than 0.8535g.

Traditional Method:



Since 3.83 falls in the rejection region $We\ Reject\ H_0$

Testing a Claim About a Population Mean: σ UnKnown

Requirements:

1. **Random samples given.**
2. **σ is not known (S is known)**
3. $n > 30$ **or Population is normally distributed.**

Test Statistic:

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

n - sample size
$\bar{x}$ **- sample mean**
μ **- population mean of all sample means from samples of size n (this value is based on the claim and is used in the null hypothesis)**
s **- known value of the sample standard deviation**

Traditional Method Only:

Eg: Assume you are in soda business and sample 39 soda cans, mean volume of soda was 12.11 ounces with a sample std dev of 0.27 ounces. Test the claim that the mean volume of soda is greater than 12 ounces with a 0.01 significance level.

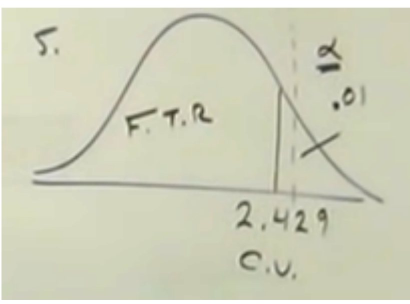
Claim: $\mu > 12 \Rightarrow H_1 : \mu > 12$

Opposite: $\mu \leq 12 \Rightarrow H_0 : \mu = 12$

$\alpha = 0.01$

$\overline{x} = 12.11, \ \mu = 12, \ s = 0.27, \ n = 465$

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{12.11 - 12}{\frac{0.27}{\sqrt{39}}} = \frac{0.11}{\frac{0.27}{6.245}} = \frac{0.11}{0.04323} = 2.54$$



2.544 > 2.429

2.544 falls in rejection region $\Rightarrow We\ Reject\ H_0$

There is enough evidence to support the claim that the mean level of soda is greater than 12 ounces.

Eg: Assume you are in soda business and sample 39 soda cans, mean volume of soda was 12.11 ounces with a sample std dev of 0.27 ounces. Test the claim that the mean volume of soda is not equal to 12 ounces with a 0.01 significance level.

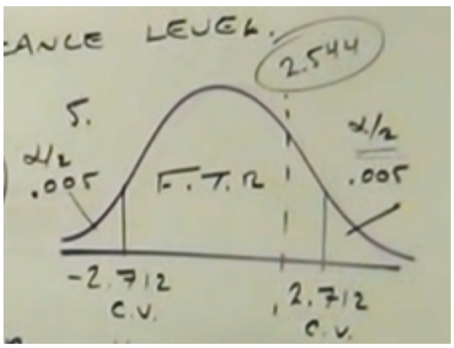
Claim: $\mu \neq 12 \Rightarrow H_1 : \mu \neq 12$

Opposite: $\mu = 12 \Rightarrow H_0 : \mu = 12$

$\alpha = 0.01$

$\bar{x} = 12.11, \ \mu = 12, \ s = 0.27, \ n = 465$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{12.11 - 12}{\frac{0.27}{\sqrt{39}}} = \frac{0.11}{\frac{0.27}{6.245}} = \frac{0.11}{0.04323} = 2.54$$



± 2.712

2.54 < 2.712

2.54 falls in Fail to reject region $\Rightarrow$ *We Fail to Reject* H_0

There is not enough evidence to support the claim that the mean level of soda is greater than 12 ounces.

Testing a Claim About a Standard Deviation or Variance:

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma^2}$$

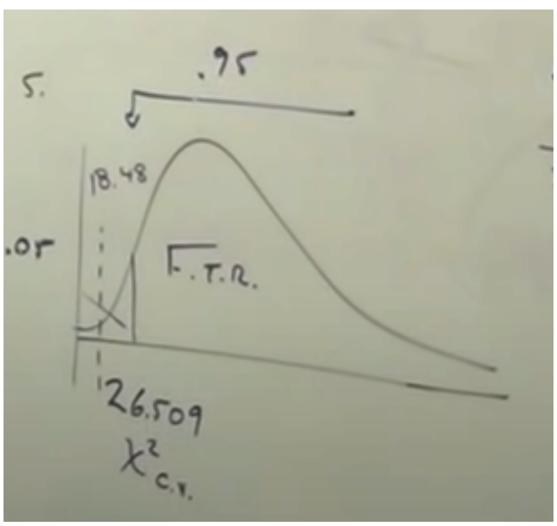
n = sample size
s = sample standard deviation
s^2 = sample variance
σ = claimed value of the population standard deviation
σ^2 = claimed value of the population variance

Eg: In a sample of 37 coins with a mean of 2.49910 gm and std dev of 0.01648 gm, test the claim that the population std dev is less than 0.023gm at a 0.05 significance level.

Claim: $\sigma < 0.023 \Rightarrow H_1 : \sigma < 0.023$
Opposite: $\sigma \geq 0.023 \Rightarrow H_0 : \sigma \geq 0.023$

$\alpha = 0.05,\ n = 37,\ s = 0.01648,\ \sigma = 0.023$

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma^2} = \frac{(37-1)\,0.01648^2}{0.023^2} = \frac{36 * 0.002715904}{0.000529} = 18.4825$$



18.4825 < 26.509 Hence $\Rightarrow We\ Reject\ H_0$
There is enough evidence to support the claim that the population std dev is less than 0.023gm