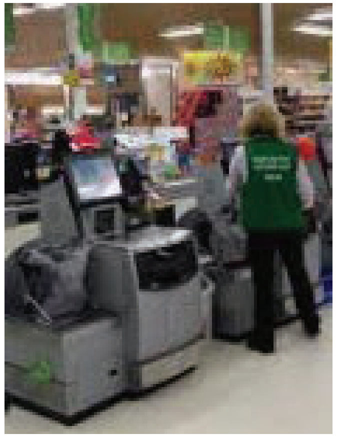
MODULE TWO: MEASURING UNCERTAINTY; AND DRAWING CONCLUSIONS ABOUT POPULATIONS BASED ON SAMPLE DATA

TOPIC 7: HYPOTHESIS TESTING



EFMD EPAS ACCREDITED

AACSB ACCREDITED

DEAKIN BUSINESS SCHOOL

+ Learning Objectives

At the completion of this topic, you should be able to:

- identify the basic principles of hypothesis testing
- use hypothesis testing to test a mean or proportion
- explain the assumptions of each hypothesis-testing procedure, how to evaluate them and the consequences if they are seriously violated
- recognise the pitfalls involved in hypothesis testing
- identify the ethical issues involved in hypothesis testing

+ Hypothesis Tests

We have two ways of using the sampling distribution to manage sampling error. They are:

- Confidence intervals (previous Topic)
  - Used if we have **no idea** about the value of the population parameter being investigated
- Hypothesis tests (this Topic)
  - Used when we **do have some idea** of the value of the population parameter being investigated, or a hypothesised value against which we can compare our sample results

These statistical techniques are widely used in decision making and research

+When are hypothesis tests used?

Generally in situations where we have:

- Prior knowledge
- Prior experience
- A standard
- A claim

In these situations we do generally have:

- Some idea of the value of the population parameter being investigated, or
- We have some hypothesised value against which we can compare our sample results.

Example: All jars of coffee should contain 500 grams of coffee and we wish to test whether a random sample meets the requirements of that standard.

+Hypothesis Tests

Hypothesis testing is about:

- Putting forward an assumption or claim about a population parameter
- Collecting data to determine the equivalent sample statistic
- Testing whether the sample statistic is consistent or inconsistent with the assumed population parameter
  - i.e. is it plausible that this data could be collected if the assumption was true?
- This approach underlies the 'scientific' method of discovery

+Hypothesis Tests

If the sample statistic is:

- **Consistent** with our assumption, then we have no reason to reject our assumption
  - i.e. It is plausible that the data could be collected if the assumption was true
- **Inconsistent** with our assumption, then we conclude that the population parameter is incorrect and we reject it in favour of another value
  - i.e. It is NOT plausible that the data could be collected if the assumption was true. In other words, <u>we have statistical evidence to suggest that our assumed population parameter is not correct</u>

+9.1 Hypothesis-Testing Methodology

The Null Hypothesis (H_0)

- States the belief or assumption in the current situation (status quo)
- Begin with the assumption that the null hypothesis is true (similar to the notion of innocent until proven guilty)
- Always contains '=', '≤' or '≥' sign
- May or may not be rejected
- Is always about a population parameter; e.g. μ, not about a sample statistic $\bar{X}$

Example H_0 : $\mu = 100\%$

+9.1 Hypothesis-Testing Methodology

The Alternative Hypothesis (H_1 OR H_A)

- Is the opposite of the null hypothesis e.g. The average number of TV sets in Australia homes is not equal to 3 (H1: $\mu \neq 3$)
- Challenges the status quo
- Can only can contain either the ‘<‘, ‘>’ or ‘≠’ sign
- May or may not be proven
- Is generally the claim or hypothesis that the researcher is trying to prove

+Hypothesis Tests

An analogy: a court case:

- An assumption is made at the beginning of a court case that the accused person is innocent (i.e. not guilty). We will call this the Null Hypothesis (H_0)
- The alternative to this assumption is that the accused person is guilty. We will call this the Alternative Hypothesis (H_1)
- The prosecution presents evidence that is inconsistent with the presumption of innocence
  - The standard used for this inconsistency is 'beyond reasonable doubt'

+Hypothesis Tests: The Critical Value of the Test Statistic Method

A critical value signifies the boundary between the regions of Rejecting the Null Hypothesis (H_0) and Not Rejecting the Null Hypothesis (H_0)

The critical value is derived from the Level of Significance, α, and is a reflection of the level of risk you are prepared to take

A **Test Statistic** is then calculated (using sample data) and then compared to the **Critical Value** to determine if the Null Hypothesis is Rejected or Not Rejected

+Regions of Rejection and Non-Rejection

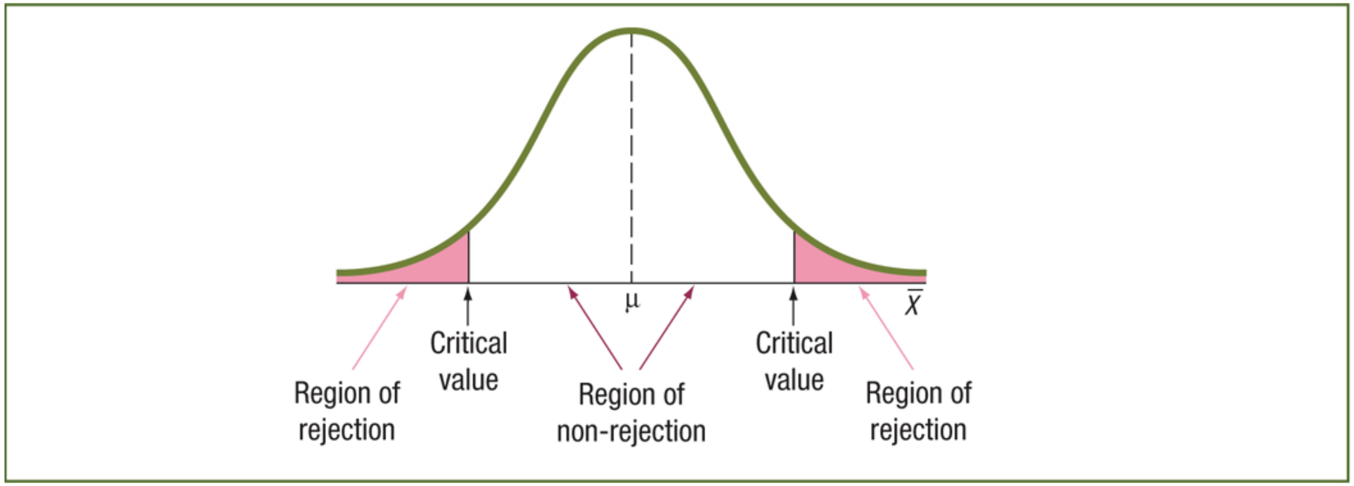
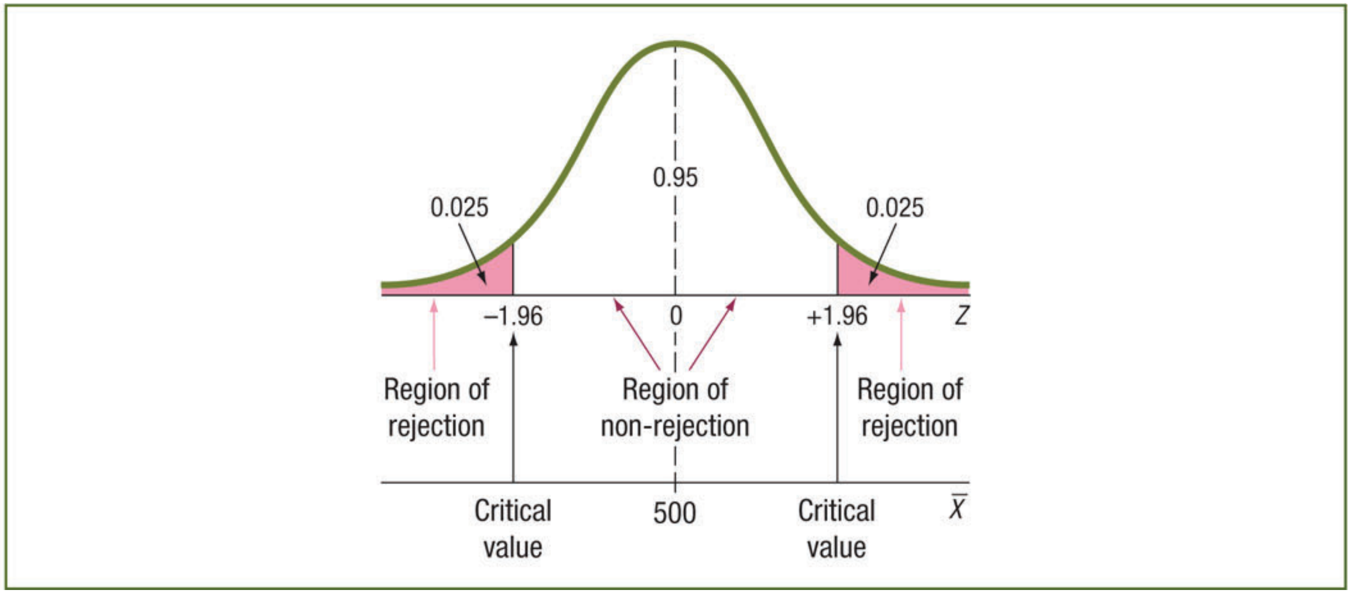


Figure 9.1
Regions of rejection and non-rejection in hypothesis testing

+Test of Hypothesis

The Critical Value Approach to Hypothesis Testing

Figure 9.2
Testing a hypothesis about the mean (σ known) at the 0.05 level of significance



H_0: $\mu = 500$
H_0: $\mu \neq 500$

level of significance = 0.05

+Risks in Decision Making Using Hypothesis Testing

Errors in hypothesis tests

Correct outcomes:

- H_0 is true and we do not reject H_0
- H_0 is false and we reject H_0

Incorrect outcomes: Errors

- Type I Error:
  We reject H_0 when in fact H_0 is true
- Type II Error:
  We do not reject H_0 when in fact H_0 is false

+Risks in Decision Making Using Hypothesis Testing

	Actual situation	
Statistical decision	**H_0 true**	**H_0 false**
Do not reject H_0	Correct decision Confidence = $1 - \alpha$	Type II error P(Type II error) = β
Reject H_0	Type I error P(Type I error) = α	Correct decision Power = $1 - \beta$

Table 9.1
Hypothesis testing and decision making

+ Risks in Decision Making Using Hypothesis Testing

Minimise the chance of a serious error

- It would be a serious error to conclude someone was guilty when in fact they weren't
- As a general rule, we set up our two hypotheses so that we can control, and minimise, the chance of making the more serious error
- Type I Error is normally the more serious of the two, so we try to set up H_0 and H_1 accordingly – to control the Type I error
- This is also a conservative approach to decision making. We only conclude the alternative is true if there is sufficient evidence
- Thus to set up H_0 and H_1 we need to think about:
  - Which error is the more serious (Type I)
  - The risk we are willing to run of making this error

+Flow Chart for Selecting a One-Sample Test

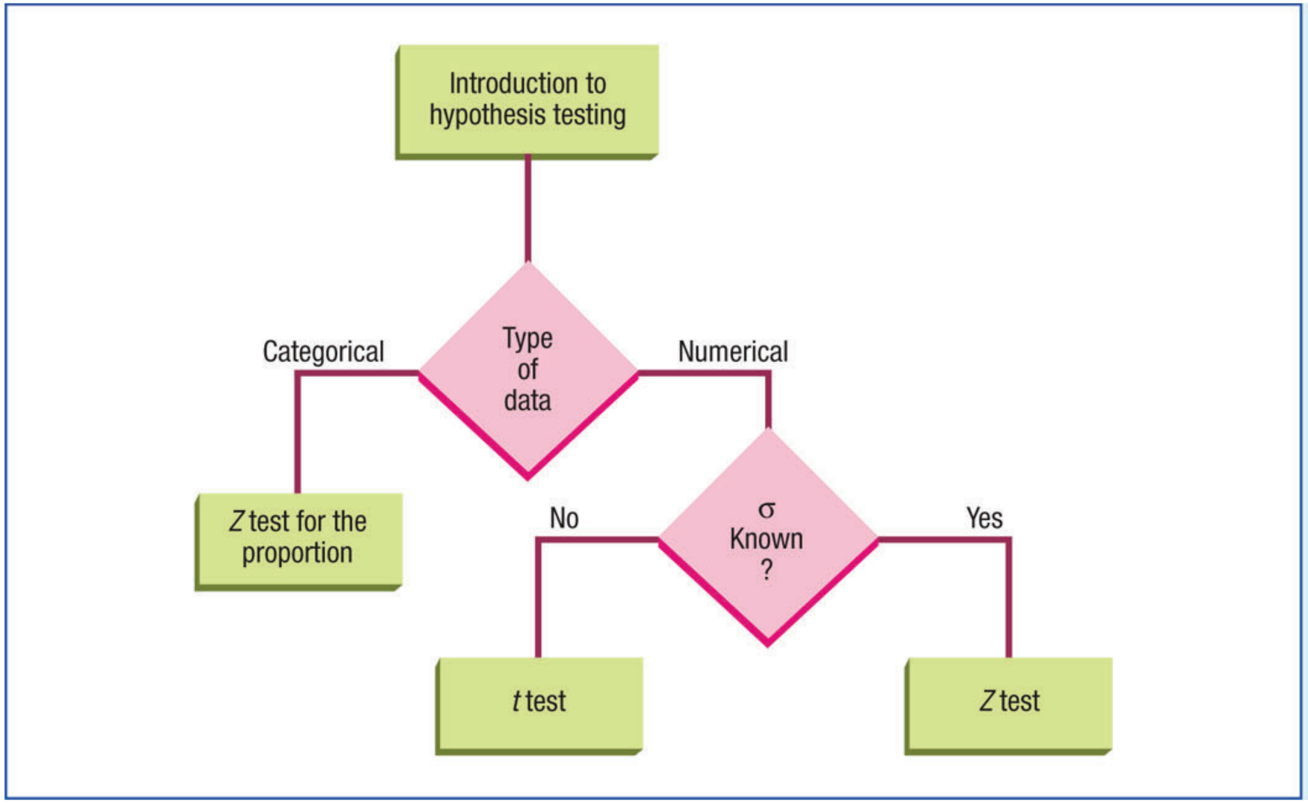


Figure 9.22
Flow chart for selecting a one-sample test

+Test of Hypothesis (cont)

Six Steps in Hypothesis Testing:

1. State the null hypothesis, H_0 and the alternative hypothesis, H_1 in symbols and words
2. Choose the level of significance, α, and the sample size, *n*
3. Determine the appropriate test statistic and sampling distribution (i.e. Z if σ known, or for a proportion; and *t* if σ is unknown)
4. Determine the critical value(s) that divide the rejection and non-rejection regions
5. Collect the data and calculate the value of the test statistic
6. Make the statistical decision (i.e. Reject or Do Not Reject the Null Hypothesis) and state the managerial conclusion (written in the context of the real world problem)

+Step 1 - Setting up H_0 and H_1 Prior Knowledge

Assume prior knowledge is still current, therefore set as H_0

Example: A detailed survey in 2007 showed the average kilometres travelled yearly per car was 14,500k. Has there been any change in 2017? We take a random sample of cars from 2017

We are interested in whether the population mean, μ, is different (has it increased or decreased)

- H_0: $\mu = 14{,}500$
  No change: Average is still 14,500k in 2017
- H_1: $\mu \neq 14{,}500$
  Change: Average is not 14,500k in 2017

+Step 1 - Setting up H_0 and H_1 Prior Experience

Assume prior experience is still reliable, therefore set as H_0

Example: Past experience has shown that no more than 3% of components from a particular supplier are defective. A new batch has just arrived: we test a random sample of components from this batch

We are interested in whether the population proportion, π, is greater than 3%

- H_0: $\pi \leq 3\%$
  The new batch has no more than 3% defective components
- H_1: $\pi > 3\%$
  The new batch has more than 3% defective components

+Step 1 - Setting up H_0 and H_1 Claim

Sometimes we treat a claim as H_0 (i.e. we accept the claim is true until we have evidence to the contrary). Which way depends on who is making the claim, the seriousness of the claim, etc.

Example: A newspaper claims that at least 40% of readers will see a particular type of ad in its paper. In this example, we will adopt a conservative approach and assume the claim is true

- H_0: $\pi \geq 40\%$
  At least 40% of readers will see the ad
- H_1: $\pi < 40\%$
  Less than 40% of readers will see the ad

+Step 1 - Setting up H_0 and H_1 Standard

We normally set H_0 equal to the standard

Example: We wish to test whether employees overall at a given company are producing 100+ widgets per day. We take a random sample of 48 staff

- H_0: $\mu \geq 100$
  Employees on average are meeting, or exceeding, the 100 widgets per day standard
- H_1: $\mu < 100$
  Employees on average are failing to meet the 100 widgets per day standard

+Some Common Mistakes in Setting up H_0 and H_1

The **equality** part of the hypotheses always appears in the **null hypothesis**

$$H_0: \pi > 100$$
$$H_1: \pi \leq 100$$

Hypotheses are statements about **population parameters**, not **sample statistics**

$$H_0: \bar{x} = 100$$
$$H_1: \bar{x} \neq 100$$

+Step 1 - Different Types of Tests

In general, a hypothesis test about the value of a population mean, μ, or proportion, π, must take one of the following three forms:

Numerical Data	$H_0: \mu \geq A$ $H_1: \mu < A$	$H_0: \mu \leq A$ $H_1: \mu > A$	$H_0: \mu = A$ $H_1: \mu \neq A$
Categorical Data	$H_0: \pi \geq A$ $H_1: \pi < A$	$H_0: \pi \leq A$ $H_1: \pi > A$	$H_0: \pi = A$ $H_1: \pi \neq A$

+Step 1 - One-Tail Tests

In some cases, the alternative hypothesis focuses on a particular direction

This is a **lower**-tail test since the alternative hypothesis is focused on the lower tail below the mean of 3

H_0: $\mu \geq 3$

H_1: $\mu < 3$

This is an **upper**-tail test since the alternative hypothesis is focused on the upper tail above the proportion of 8%

H_0: $\pi \leq 8\%$

H_1: $\pi > 8\%$

+Step 1 - Two-Tail Tests

In other cases, the alternative hypothesis focuses on both directions

This is a **two**-tail test since the alternative hypothesis is focused on either values below the mean of 25 or values above 25

H_0: $\mu = 25$

H_1: $\mu \neq 25$

H_1 : $\mu < A$ -- > Lower Tail Test

H_1 : $\mu > A$ -- > Upper Tail Test

H_1 : $\mu \neq A$ -- > Two Tail Test

+Step 2 - The Level of Significance, α, and the Sample Size, *n*

We specify the (maximum) risk we are willing to run of making Type I Error

This is normally set at 1%, 5% or 10%

In Hypothesis Testing, this risk is called the Level of Significance, α

In summary, the level of significance, α, is the maximum risk we are willing to run of making a Type I Error (rejecting H_0 when we should not)

The Level of Significance will dictate the sample size, *n*

Confidence level = $1 - \alpha$

+Step 3 - Determine the Test Statistic and Sampling Distribution

Determine the appropriate test statistic and sampling distribution

Numerical Data (i.e. Hypothesis Tests for a Population Mean, μ)

- Z test, if σ known, and Normal Distribution
- *t* test, if σ is unknown, and Student *t* Distribution

Categorical Data (i.e. Hypothesis Tests for a Population Proportion, π)

- Z test, and Normal Distribution

+Step 4 - Determine the Critical Values

Once we:

- decide on the direction of the test (Step 1); and
- specify the Level of Significance, α (Step 2)

we can next determine the critical (rejection) region which is the area in the sampling distribution where we will reject H_0

The **critical value**(s) are the border(s) of the critical region

These values are known as either:

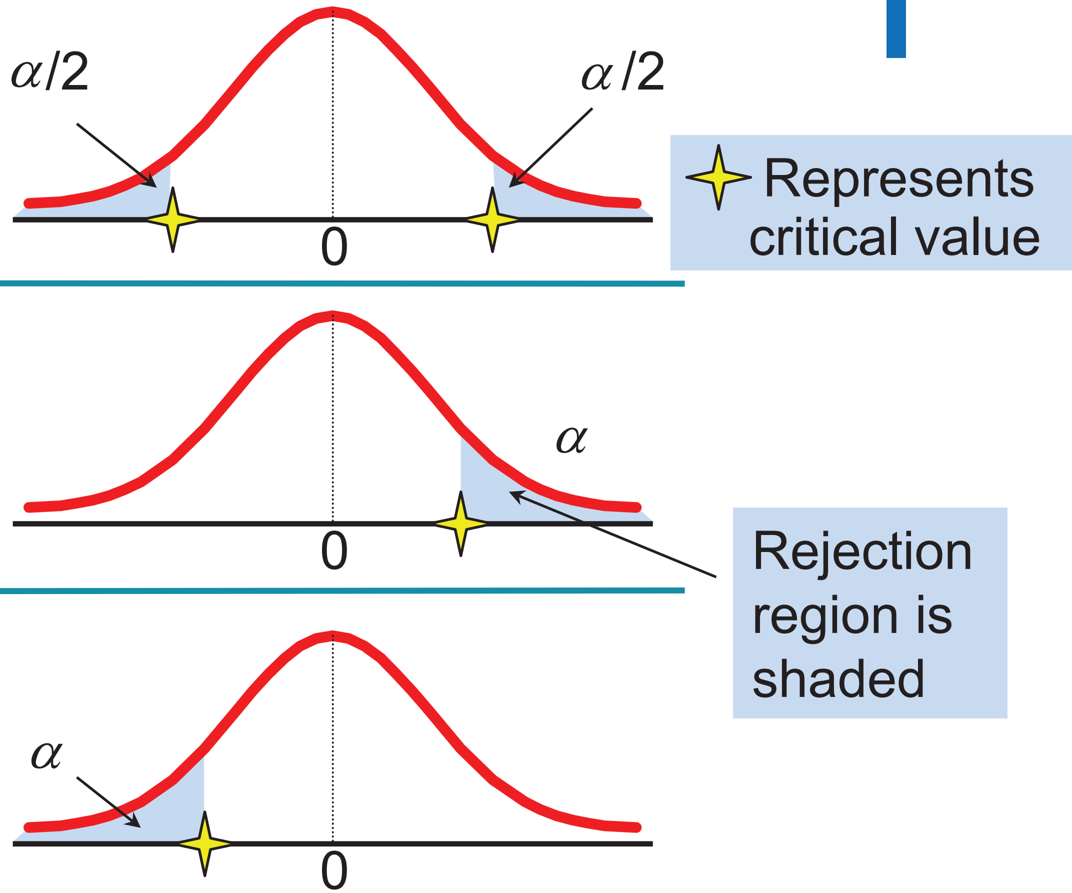
- The Critical Value of Z; or
- The Critical Value of *t*

+Step 4 - Determine the Critical Values

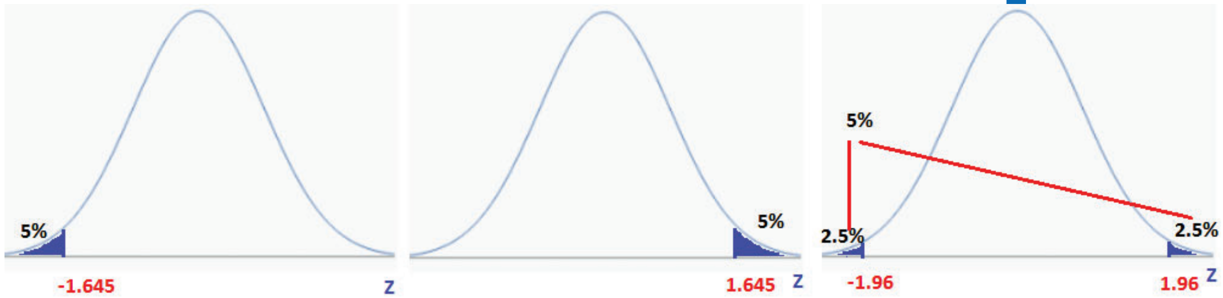
H_0: $\mu = 0$
H_1: $\mu \neq 0$ Two-tail test

H_0: $\mu \leq 0$
H_1: $\mu > 0$ Upper-tail test

H_0: $\mu \geq 0$
H_1: $\mu < 0$ Lower-tail test

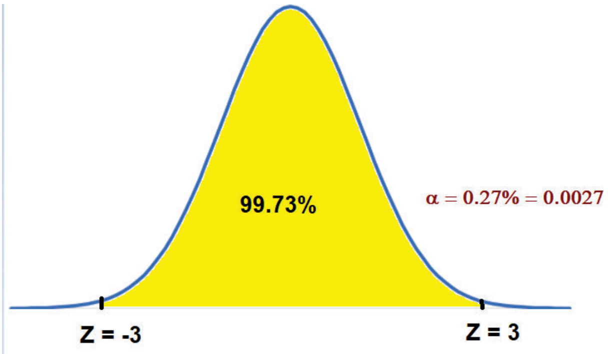


Step 4 - Determine the Critical Values (cont)



For **Normal Distribution** (from Table E.2)

Level of Significance α	*Critical Value of Z (Lower Tail)*	*Critical Value of Z (Upper Tail)*	*Critical Value of Z (Two Tail)*
10%	-1.28	1.28	±1.645
5%	-1.645	1.645	± 1.96
1%	-2.33	2.33	± 2.575



For **Student t Distribution** (from Table E.3), first determine:

- degrees of freedom (n-1), then
- the value of α within a right tail

Note: Table E.3 only returns the value for an Upper-Tail area

+Step 5 – Collect Data and Calculate Value of Test Statistic

For **Numerical**, σ Known (i.e. Z Test)

Example:

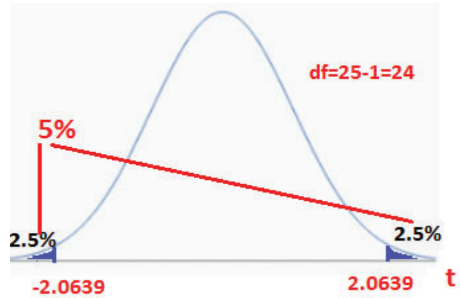
Suppose I wish to test the assumption, or status quo, that the true mean of TV sets in Australian homes is equal to 3. I take a sample with the following results:

- n = 100, $\bar{X}$= 2.84, σ = 0.8 (for this example, assume that σ is known)

Therefore, the test statistic would be calculated as:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = \frac{-.16}{.08} = -2.0$$

+ Step 5 – Collect Data and Calculate Value of Test Statistic



For **Numerical**, σ **Un**known (i.e. *t* Test)

Example:

Suppose the average cost of a hotel room in Sydney is said to be \$168 per night.

My hypothesis would be H_0: μ = 168 and H_1: μ ≠ 168 – Two tail Test

Set α= 5% then Critical values: ±2.0639 (df = 25-1 = 24 and right tail area = 0.025; TableE.3)

A random sample of 25 hotels resulted in an average of \$172.50 and standard deviation of \$15.40. Therefore, the test statistic is calculated as:

$$t_{n-1} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{172.50 - 168}{\frac{15.40}{\sqrt{25}}} = 1.46$$

+Step 5 – Collect Data and Calculate Value of Test Statistic (cont)

For **Categorical** (i.e. **Z Test**)

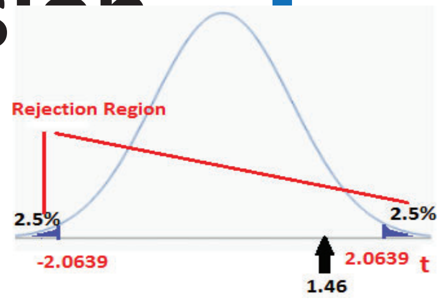
Example:

A marketing company claims that it receives 8% responses from its mailing. To test this claim, a random sample of 500 were surveyed with 25 responses which is an actual response rate of 5%

Therefore, the test statistic is calculated as:

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.05 - 0.08}{\sqrt{\frac{0.08(1-0.08)}{500}}} = -2.47$$

Step 6 – Make Statistical Decision and Draw Conclusion



Statistical Decision

Compare the Test Statistic (calculated in Step 5) with the Critical value(s) (determined in Step 4)

If the Test Statistic falls in the Reject Region, we Reject the Null Hypothesis H_0. Otherwise we Do Not Reject the Null Hypothesis H_0

Draw Conclusion

For the Sydney hotel example, we would not have rejected H_0 so our conclusion would have been: there is insufficient evidence to disprove the claim that the average Sydney hotel room to be $168 per night.

(t Test statistic =1.46 and Critical values = ±2.0639 (TableE.3))

+Test of Hypothesis (cont)

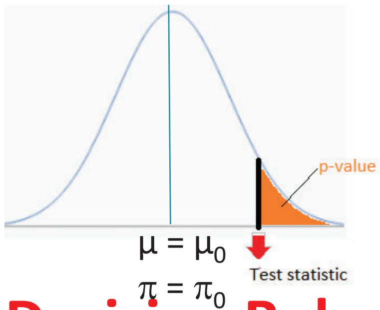
Six Steps in Hypothesis Testing:

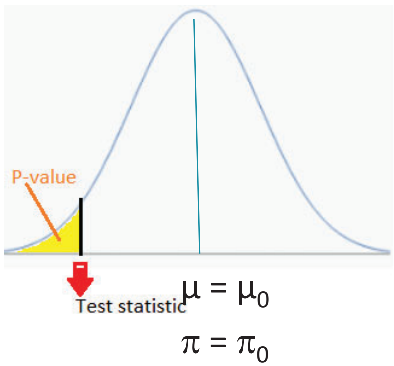
1. State the null hypothesis, H_0 and the alternative hypothesis, H_1 in symbols and words
2. Choose the level of significance, α, and the sample size, *n*
3. Determine the appropriate test statistic and sampling distribution (i.e. Z if σ known, or for a proportion; and *t* if σ is unknown)
4. Determine the critical value(s) that divide the rejection and non-rejection regions
5. Collect the data and calculate the value of the test statistic
6. Make the statistical decision (i.e. Reject or Do Not Reject the Null Hypothesis) and state the managerial conclusion (written in the context of the real world problem)

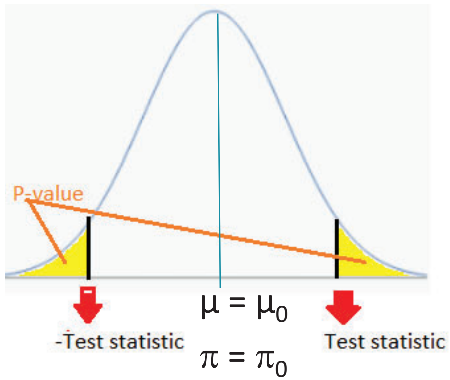
+Hypothesis Tests: The *p*-Value Approach

An alternative to the **Critical Value Approach to Hypothesis Testing** there is the ***p*-Value** method

The *p*-value is the probability of getting a test statistic more extreme than the sample result, given that the Null Hypothesis, H_0, is true







Decision Rule:

- If the p-value is **greater than** or **equal to** α, you **do not reject** the Null Hypothesis
- If the p-value is **less than** α, you **reject** the Null Hypothesis

+Hypothesis Tests: The p-Value Approach

In the example below, the p-value is **greater than** α, the Level of Significance, so we **do not reject** the **Null Hypothesis**

	A	B	
1	**Muesli packaging process hypothesis test**		
2			
3	**Data**		
4	**Null hypothesis μ =**	**500**	
5	**Level of significance**	**0.05**	
6	**Population standard deviation**	**15**	
7	**Sample size**	**25**	
8	**Sample mean**	**504.5**	
9			
10	Intermediate calculations		
11	Standard error of the mean	3	**=B6/SQRT(B7)**
12	***Z* test statistic**	**1.5**	**=(B8 – B4)/B11**
13			
14	**Two-tail test**		
15	**Lower critical value**	**–1.9600**	**=NORM.S.INV(B5/2)**
16	**Upper critical value**	**1.9600**	**=NORM.S.INV(1 – B5/2)**
17	***p*-value**	**0.1336**	**=2*(1 – NORM.S.DIST(ABS(B12),1))**
18	**Do not reject the null hypothesis**		

Figure 9.5
Microsoft Excel 2013 *Z* test worksheet for the muesli-packaging example

+A Connection Between Confidence Interval Estimation and Hypothesis Testing

Confidence Intervals are used to estimate parameters

Hypothesis testing is used when making decisions about specified values of a population parameter

- i.e. when trying to prove that a parameter is less than, more than or not equal to a specified value

Proper interpretation of a Confidence Interval, however, can also indicate whether a parameter is less than, more than or not equal to a specified value

+9.7 Potential Hypothesis-Testing Pitfalls and Ethical Issues

Sample data must be collected **randomly**

- this reduces selection biases or coverage errors
- if, however, non-random data has been collected, our results will be called into question

Human subjects cannot be surveyed without informed consent

Choose the level of significance, α, and the type of test (one-tail or two-tail) **before** data collection

Data Snooping is never permissible. It is where tests are performed **prior** to choosing between one-tail and two-tail test, or the determination of the level of significance