Lecture 7
One Sample Hypothesis Tests for a population mean

z-test for a population mean
t-test for a population mean

In the last lecture …

- We learnt that a 95% confidence interval was an interval which we believed, with 95% confidence, will contain the value of the population parameter of interest.
- The general form for a confidence interval is:

  sample estimate $\pm$ critical value $\times$ standard error
- We found confidence intervals for μ and π.
- In order to decrease the width of a confidence interval but at the same time retain our level of confidence (say 95%), then we need to increase the sample size.

Review Quiz 1

A 95% confidence interval for a mean is an interval containing:

a. the sample mean with 95% certainty
b. the population mean with 95% certainty
c. 95% of the data in the sample
d. 95% of the data in the population
e. the population mean in 95% of repeated samples

Review Quiz 2

A survey of 30 adults found that the mean age of a person's primary vehicle is between 5.3 and 5.9 years with 95% confidence.

a. What was the mean age of the vehicles in the sample used to calculate this interval?
b. How could you produce a narrower 95% confidence interval for the population mean?
c. If you were asked to provide a 99% confidence interval for the mean age of a person's primary vehicle, how would you change the following formula:

$$\bar{y} \pm t_{crit} \times \frac{s}{\sqrt{n}}$$

Review Exercise 1

Consider the output below which was obtained from the mammals data set we met in Lecture 1. The variable of interest is the number of hours sleep per day:

Descriptive Statistics: Sleep

Variable	Diet	Total Count	Mean	StDev
Sleep	carnivore	10	11.95	4.85
	herbivore	20	9.390	4.041

Provide a 95% confidence interval for the mean number of hours sleep per day among herbivores, assuming sleep times are drawn from a normal distribution.

Source: *Mammals database* (Lecture 1)

Review Exercise 2

A dean of a business school wanted to know whether the graduates of her school used a statistical inference technique during their first year of employment after graduation. She surveyed 314 graduates and asked about the use of statistical techniques. After tallying the responses, she found that 204 used statistical inference within one year of graduation.

Estimate with 95% confidence the proportion of all business school graduates who use statistical education within a year of graduation.

Source: Keller, G. (2012), ***Statistics for management and economics (ninth edition)***, South-Western (adapted)

Review Answers

Quiz 1:

Quiz 2: a. b.

Exercise 1:

c.

Exercise 2:

7.7A

Research Question

- Recall that we often want to answer a research question pertaining to some target population of interest.
- We design a study carefully, and obtain a representative sample, random if possible, from the target population.
- We use this sample to make inferences about the population and, thus, to answer the research question.

7.8

Answering a Research Question

Research Question:
Is the mean IQ of children who attend country schools the same as that of the general population?

We can perform a statistical test to answer the above research question, so long as our data are independent.

Let Y = variable of interest.

Here Y = IQ score of country students.

Recall that the sample consisted of 36 students, selected randomly, with a mean IQ of 103. (In the general population, IQ scores have mean μ = 100 and standard deviation σ = 15).

7.9

Null Hypothesis

- A null hypothesis is a statement (or claim) that a population parameter has a specified value μ_0 ie.

  H_0: $\mu = \mu_0$
- Here, we could claim that the mean IQ score of country school students is 100. This claim is based on the fact that the population mean is known to be 100, and we want to determine if this is the same for country school students. (In fact the word 'null' means 'no effect' or 'no change'.)
- So we express the null hypothesis as:

  $H_0 : \mu = 100$

7.10

Alternative Hypothesis

- We can also specify a second hypothesis H_1 which must be true if the null hypothesis is not.
- H_1 is often called the research hypothesis since it is often the hypothesis that is of greatest interest to the researcher.
- H_0 and H_1 are complementary. They do not overlap and should exhaust all the possible values of the parameter being tested.
- When testing a population mean, H_1 can take one of the following forms:

  H_1: $\mu \neq \mu_0$ Two-sided test (i.e. direction is not specified)

  H_1: $\mu > \mu_0$ One-sided test (i.e. direction is specified)

  H_1: $\mu < \mu_0$ One-sided test (i.e. direction is specified)

7.11

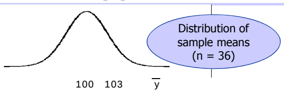
Alternative Hypothesis cont.

- A one-sided test is appropriate when the researcher has specific ideas about the direction in which the population parameter lies.
- Researchers need to decide whether to use a one-sided or a two-sided test ***before*** they look at the data.
- For the country school students IQ data, the researchers were simply interested in determining whether the mean IQ score of country school students is 100 or not (i.e. no direction was suggested). Therefore a two-sided hypothesis test is appropriate and the alternative hypothesis for this problem takes the form:

  H_1: $\mu \neq 100$.

7.12

Testing a Null Hypothesis



- An important part of statistics involves testing hypotheses, which we do by comparing μ_0 with the corresponding sample value.
- For example, in the case of the IQ scores μ_0 = 100, and the sample mean of the country school students is 103.
- Testing this null hypothesis is the same as asking the question *'if the null hypothesis is true (ie. country students' have IQ scores the same as the general population), how likely is a sample mean to be 3 points (or more) away from the population mean, in <u>either</u> direction?"*

7.13

Test Statistic

- A test statistic is a scaled measure of the discrepancy between the observed sample statistic and the value specified by H_0.
- Dividing the discrepancy by the standard error gives a result that is independent of the measurement unit.
- For a one sample test of a mean where Y is the variable of interest, the test statistic takes the form:

$$\boxed{\text{test statistic} = \frac{\bar{y} - \mu_0}{\text{st. error of } \bar{y}}}$$

Note: The test statistic tells us how many standard errors the sample value is away from the hypothesised value.

7.14

Test Statistic: continued

So if σ is known, the test statistic is:

$$\boxed{z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}}$$

Note that we can only calculate this test statistic if we know that the sample mean arises from a normal population

For this study involving IQ scores, the sample mean is $\bar{y}$ = 103, and the null value is μ_0 = 100.

The sample size is n = 36, the population standard deviation σ = 15.

The test statistic is thus:

$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{103 - 100}{15/\sqrt{36}} = \frac{3}{2.5} = 1.2$$

A mean IQ of 103 is 1.2 standard errors from μ_0.

7.15

Size of Test Statistic

- If the absolute value of the test statistic is *large*, it means that there is a sizable discrepancy between the null value and the sample value.
- But if the test statistic is *small*, the null value is consistent with the data in the sample. (That is, the discrepancy could be attributable to chance ie. sampling variation)
- A more precise measure of the discrepancy is given by a *p-value.*

7.16

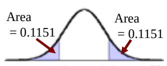
p-value for a two-sided test

- A p-value is the likelihood of getting such a large test statistic (or even larger), assuming that the null hypothesis is true.
- Note that we are interested in the measure of discrepancy ie. how likely are we to select a sample with a mean at least 1.2 standard errors away from the null value, if H_0 is true?
- The p-value is thus computed as the tail area of the distribution of the test statistic, assuming H_0 is true. For this z test of a population mean, the p-value is the area in *both* tails of the z distribution, because our measure of discrepancy could have been either side of the mean. This is known as a two-sided (or two-tailed) test.

7.17

Country students: p-value

For the study involving the country school students' IQ scores, the test statistic is z = 1.2. Using the standard normal distribution, the area in the two tails to the right of z = 1.2 and to the left of z = –1.2 is 0.2302.



This is the p-value. It is the probability of getting a z-value with magnitude greater than 1.2 if H_0 is true.

We use the standard normal distribution because the mean IQ score comes from a normal population.

7.18

Significance level

- A *small p-value* leads to a decision to *reject* the null hypothesis.
- By convention, a p-value is considered 'small' if it is less than 0.05. (This value of 0.05 is referred to as the significance level or *α-level* of the test.)
- The result is then said to be *'statistically significant'*.
- Conversely, if a *p-value is not small*, the decision is to *not reject* the null hypothesis. However, this does not mean that H_0 is accepted!
- So, for the country students' study, the correct decision is to *not reject H_0*, since the p-value is 0.2302.

7.19

Hypothesis Testing

Study Information

Specify the *target population* and outline the *sample*. Indicate why the sample is representative of the target population and how the measurements are independent of each other.

Data Summaries

Identify the variable/s of interest.

Look at the plots and numerical summaries.
These may be given to you or you may need to obtain them from the data.

7.20

Steps in Hypothesis Testing

Hypothesis Test

H *Hypothesis:*
State the *null* and the *alternative* hypothesis in terms of the parameter of interest.

A *Assumptions:*
Check the underlying assumptions of the test.

T *Test Statistic:*
Calculate the test statistic.

P *p-value:*
Obtain the p-value for the test and make a decision about the believability of H_0.

C *Conclusion:*
Write a conclusion framed in terms of the original research question and include a confidence interval for the population parameter (for two sided tests only).

7.21

Example: Stopping Distance

Research Question:
Is the average stopping distance of school buses travelling at 75 km per hour equal to 80 metres as claimed by public transport authorities?

Studies by public transport authorities have indicated that the average stopping distance of a bus travelling at 75 km per hours is 80 metres. The standard deviation of stopping distances at that speed is known to be 1.2 metres.

Automotive engineers tested a randomly selected sample of 20 school buses and found that the average stopping distance for this sample, when travelling at 75 km per hour, was 80.74 metres.

7.22

Example: continued

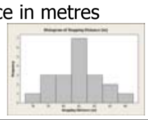
Study Information

The target population is school buses. We can assume our sample is representative of all school buses and that the 20 school buses sampled are independent since they were selected randomly.

Stopping Distance (m)
78.4
78.5
78.7
79.4
79.8
80
80.4
80.5
80.5
80.5
80.6
80.7
81
81
81.5
81.9
82
82.5
83.3
83.6

Data Summaries

Variable: Y = stopping distance in metres
$n = 20, \quad \bar{y} = 80.74, \quad \sigma = 1.2$



7.23

Example: continued

Hypothesis Test

H H_0: $\mu = 80$
H_1: $\mu \neq 80$ (i.e. no direction is specified so this is a two- sided test)

A From the histogram, the normality assumption appears reasonable, since the sample seems to be drawn from a population which is normally distributed.

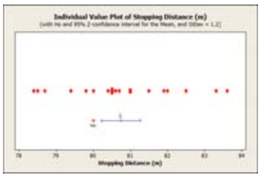
T $$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{80.74 - 80}{1.2/\sqrt{20}} = 2.76$$

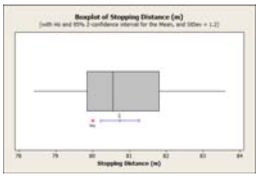
P p-value = 0.0058. Since p-value < 0.05, reject H_0.

C The average stopping distance for school buses is <u>more</u> than 80 metres. We are 95% confident that the average stopping distance is between 80.21 and 81.27 metres.

7.24

Computer Output






```
One-Sample Z: Stopping Distance (m)
Test of mu = 80 vs not = 80
The assumed standard deviation = 1.2

Variable              N    Mean  StDev  SE Mean       95% CI
Stopping Distance (m)  20  80.740  1.445    0.268  (80.214, 81.266)

Variable                Z      P
Stopping Distance (m)  2.76  0.006
```

7.25

Exercise 2

Research Question:
Is the average weight of a plate produced by a machine equal to 25kg, as claimed?

A company builds large harvesters. For a harvester to be properly balanced when operating, a 25 kg plate is installed on its side. The machine producing these plates is set to yield plates averaging 25kg with a standard deviation of 1kg. The supervisor is concerned that the machine is out of adjustment and is producing plates that do not average 25kg. To test this concern, she randomly selects 20 plates produced the day before and weighs them.

22.6	22.2
23.2	27.4
24.5	27.0
26.6	28.1
26.9	24.9
26.2	25.3
23.1	24.2
26.1	25.8
30.4	28.6
23.5	23.6

μ = ?
σ = 1

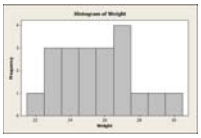
Source: Black et al (2010) *Australian Business Statistics, Wiley (adapted)*

7.26Q

Answer to Exercise 2

Study Information

Data Summaries



7.27A

Answer to Exercise 2

Hypothesis Test

H

A

T

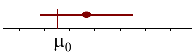
P

C

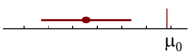
7.28A

p-value and confidence interval

In the case of a ***two-sided hypothesis test***, a 95% confidence interval can be used to test the null hypothesis.
If the confidence interval does not include the null value, the decision is 'reject H_0'. If the confidence interval includes the null value, the decision is 'do not reject H_0'.
A p-value is a refinement of this rule. If the p-value is less than 0.05, H_0 is rejected, whereas if the p-value is greater than 0.05, H_0 is not rejected.



p-value > 0.05 *do not reject* H_0



p-value < 0.05 *reject* H_0

A two-sided 95% confidence interval can ***not*** be used to test a null hypothesis associated with a ***one-sided hypothesis test***. 7.29

One-sided hypothesis test

As mentioned earlier, a one-sided test is used when there is a specific idea about the ***direction*** in which the population parameter lies. We need to decide whether to use a one sided or a two sided test ***before*** we look at the data. That is, we can not let the data suggest a hypothesis, then 'test' whether the data support the hypothesis.

Furthermore, since H_0 and H_1 must be complementary (i.e. not overlapping) and since they should exhaust all the possible values of the parameter being tested, the null/alternative hypothesis set for one-sided tests can take one of the following forms:

$H_0: \mu \le \mu_0$ OR $H_0: \mu \ge \mu_0$

$H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$

where μ_0 is some specified value.

7.30

One-sided hypothesis test cont.

- Each of the sets of hypotheses on previous slide account for all possible values of μ.
- However, only the parameter value in H_0 that is closest to H_1 influences the form of the test.
- **We will therefore write H_0 in a simpler form where the parameter takes on the specific value of μ_0.** For one-sided hypothesis tests we will therefore write the null/alternative hypothesis set as either

$H_{0:}\ \mu = \mu_0$, $H_{1:}\ \mu > \mu_0$ OR $H_{0:}\ \mu = \mu_0$, $H_{1:}\ \mu < \mu_0$

- All the steps outlined earlier for hypothesis testing still apply when doing a one-sided hypothesis test. Only the determination of the p-value is different.

7.31

p-values

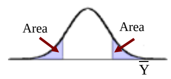
Two-sided tests

Research hypothesis H_1 *does not* specify direction.

H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$

where μ_0 is some value

p–value



p–value = Area in both tails

One-sided tests

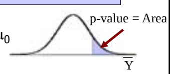
Research hypothesis H_1 *does* specify direction.

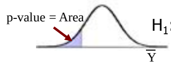
H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$ OR H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$

where μ_0 is some value

p–value

H_1: $\mu > \mu_0$





H_1: $\mu < \mu_0$

7.32

Exercise 3

Research Question:
Is the average Degree of Reading Power (DRP) score of third graders in a suburban school district higher than the national average DRP score, which is 32.

The DRP scores for 44 third-grade students are given on the right. These students can be considered to be a random sample of the third-graders in a suburban school district. DRP scores are approximately normally distributed. Suppose that the standard deviation of scores in this school district is known to be σ = 11. The researcher believes that the average score μ of third graders in this district is higher than the national average, which is 32.

40	47	52	47
26	19	25	35
39	26	35	48
14	35	35	22
42	34	33	33
18	15	29	41
25	44	34	51
43	40	41	27
46	38	49	14
27	31	28	54
19	46	52	45

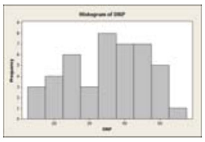
Source: Moore and McCabe (1998) *Introduction to the practice of statistics (third edition), Freeman (adapted)*

7.33Q

Answer to Exercise 3

Study Information

Data Summaries



7.34A

Answer to Exercise 3

Hypothesis Test

H

A

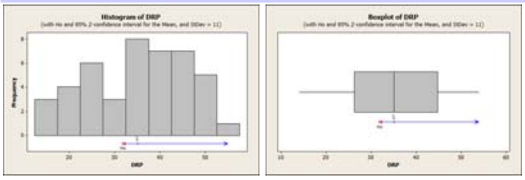
T

P

C

7.35A

Computer Output:



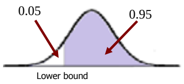
```
One-Sample Z: DRP
Test of mu = 32 vs > 32
The assumed standard deviation = 11

                                          95% Lower
Variable   N    Mean  StDev  SE Mean      Bound      Z       P
DRP       44   35.09  11.19     1.66      32.36   1.86   0.031
```

When doing a one-sided hypothesis test, only a one-sided confidence bound is given in the output.

7.36

One-sided confidence bounds

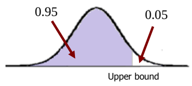
A 95% one-sided lower confidence bound indicates that 95% of the population is greater than the value given by the lower bound.



The calculation of the 95% lower bound in Example 3 is as follows:

$$\bar{y} - z^* \frac{\sigma}{\sqrt{n}} = 35.09 - 1.645 * \frac{11}{\sqrt{44}} = 32.36$$

A 95% one-sided upper confidence bound indicates that 95% of the population is less than the value given by the upper bound.



The calculation of lower and upper confidence bounds is NOT examinable in STAT170.

7.37

The t-test for a population mean

- The z-test assumes that the population standard deviation (σ) is known, but usually σ is unknown. In which case its value can be estimated using the sample standard deviation, s.
- The z-test statistic is then replaced by with ν = n – 1 degrees of freedom (df)

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

This is called *Student's t statistic*.

- A p-value is obtained from the t-distribution (instead of the standard normal distribution).

7.38

Determining the p-value

p two-tail	*0.0005*	*0.001*	*0.002*	*0.005*	*0.01*	*0.02*	*0.05*	*0.1*	*0.2*	*0.5*
p one-tail	*0.00025*	*0.0005*	*0.001*	*0.0025*	*0.005*	*0.01*	*0.025*	*0.05*	*0.1*	*0.25*
ν = 10	5.049	4.587	4.144	3.581	3.169	2.764	2.228	1.812	1.372	0.700

Using the t-table, you cannot usually determine the p-value exactly. However, you can get a *range* of values for it. The first and the second lines of the table give the two-tailed probability (used for two-sided tests) and the one-tailed (used for one-sided tests) probability respectively.

For example, suppose | t| is 2.31 and ν is 10 and suppose that you are doing a two-sided hypothesis test.
Now p-value = 0.02 when | t| is 2.764 and p-value = 0.05 when | t| is 2.228.

So 0.02 < p-value < 0.05

Most statistical packages calculate the exact p-value.

7.39

Quiz 3

Use the t-tables to determine a range of p-values for the following test statistics when performing

1) one-sided hypothesis test:

a. t = 3.326 with a sample size of 15

b. t = 3.855 with a sample size of 10

c. t = 1.276 with a sample size of 20

2) two-sided hypothesis test:

a. t = 0.544 with a sample size of 30

b. t = 2.925 with a sample size of 58

c. t = 1.650 with a sample size of 200

7.40Q

Answers to Quiz 3

1) one-sided hypothesis test:

	t-value	n	df	p-value
a.	3.326	15		
b.	3.855	10		
c.	1.276	31		

2) two-sided hypothesis test:

	t-value	n	df	p-value
a.	0.544	30		
b.	2.925	58		
c.	1.650	200		

7.41A

Example: Trout fishing

Source: *Snowy Mountains Region, New South Wales* (2005)

Research Question: Is the length of trout caught in the Snowy Mountains region equal to 25cm, on average?

In April, Eric went trout fishing at 12 of his favourite locations in the Snowy Mountains region. He caught one trout at each location then moved on to the next.

location	length	location	length
Eucumbene Dam	32	Tumut River	28
Three-mile Dam	21	Thredbo River	20
Yarrangobilly River	25	Jindabyne Dam	18
Goobarragandra	23	Blowering Dam	33
Talbingo Dam	15	Snowy River	19
Wee Jasper	26	Batlow River	22

Eric claims that the average length of a trout in the Snowy Mountains must be equal to 25cm.

7.42

Example: continued

Study Information

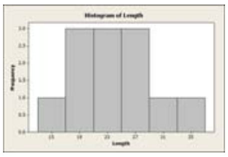
Target population: <u>all</u> trout from the Snowy Mountains region.

Sample: 12 trout, selected from different locations, so it is reasonable to assume that this would constitute a representative sample of independent trout from this area.

Data Summaries

Variable: Y = length (cm)

n = 12, $\bar{y}$ = 23.5, s = 5.52



7.43

Example: continued

Hypothesis Test

H H_0: μ = 25
H_1: $\mu \neq$ 25 (note that no direction was suggested)

A The measurements, from the histogram, appear to be drawn from a normally distributed population.

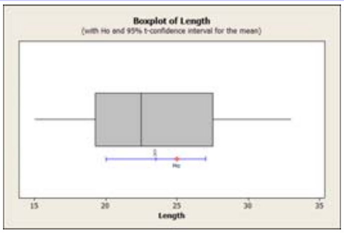
T $$t = \frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{23.5 - 25}{5.52/\sqrt{12}} = -0.94 \quad \text{with 11df}$$

P 0.2 < p-value < 0.5. Since p-value > 0.05 don't reject H_0.

C There is insufficient evidence to reject Eric's claim. The mean length of trout caught in the Snowy Mountains region could be 25cm. 95% confidence interval for μ is (19.99, 27.01).

7.44

Computer output

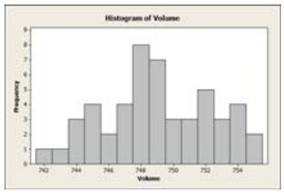



```
One-Sample T: Length
Test of mu = 25 vs not = 25

Variable   N   Mean  StDev  SE Mean      95% CI          T      P
Length    12  23.50   5.52     1.59  (19.99, 27.01)  -0.94  0.367
```

7.45

Barossa Valley wine

The computer output shown below was obtained from 50 randomly selected bottles of wine from a Barossa Valley winery.




```
One-Sample T: Volume
Test of mu = 750 vs not = 750

Variable   N     Mean  StDev  SE Mean        95% CI            T      P
Volume    50  749.040  3.338    0.472  (748.091, 749.989)  -2.03  0.047
```

7.46

Exercise 4

Research Question:
Is the average volume of wine dispensed into bottles by a machine at the Barossa Valley winery 750 ml as labelled?

The computer output was obtained from a random sample of 50 bottles from a Barossa Valley winery. The wine is dispensed into the bottles by a machine and labelled as containing 750ml.

Use the computer output on the previous slide to answer the question posed above and determine whether the machine is dispensing 750ml on average, as claimed.

7.47Q

Answer to Exercise 4

Hypothesis Test

H

A

T

P

C

7.48A

Lecture 7 summary (1)

A null hypothesis (H_0) is a statement that a population parameter has a particular value, known as the null value.

A test statistic is a scaled measure of the discrepancy between the null value and the corresponding value obtained from a sample.

A p-value is the probability of getting such a test statistic or even more extreme in a study, assuming H_0 is true.

7.49

Lecture 7 summary (2)

Hypothesis testing involves:

study information: outlining information about the target population and the sample.

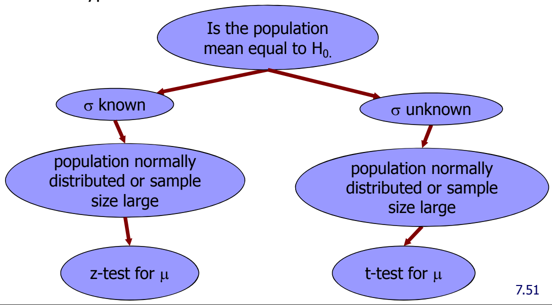
data summaries: stating the variable/s of interest and providing numerical summaries and visual display/s from the data.

answering the research question: H stating the null and the alternative hypothesis, A checking the assumptions of the test, T calculating the test statistic, P obtaining the p-value and making a decision about the believability of H_0 based on this and C writing a conclusion, which should include a confidence interval if a two-sided hypothesis test was carried out and the H_0 was rejected.

7.50

Lecture 7 concept map

The following concept map applies for both one-sided and two-sided hypothesis test for means.



7.51