Chapter S7

Hypothesis Testing

In this chapter, students will learn

- concepts of null hypothesis (H_0) and alternative hypothesis (H_1), test statistic, critical region, critical value, level of significance and *p*-value,
- the formulation of hypotheses and tests for a population mean based on:
  - a sample from a normal population of known variance,
  - a large sample from any population,
- 1-tail and 2-tail tests,
- interpretation of the results of a hypothesis test in the context of the problem.

7.1 Introduction (Self reading)

Suppose a school superintendent reads an article which states that the overall mean rank points (RP) for the 2019 GCE 'A' Level Exams is 79. Furthermore, suppose that the average RP of a sample of students from the superintendent's school district is 83. Can the superintendent conclude that the students in the school district scored significantly higher than average? At first glance, you might be inclined to say yes, since 83 is higher than 79. But recall that the means of samples vary about the population mean when samples are selected from a specific population. So the question arises. Is there a real difference in the means, or is the difference simply due to chance i.e. sampling error? In this chapter, you will learn how to answer many questions of this type by using statistics that are explained in the theory of hypothesis testing.

Researchers are interested in answering many types of questions. For example, a doctor might want to know whether a new medication will lower a person's blood pressure. Automobile manufacturers are interested in determining whether seat belts will reduce the severity of injuries caused by accidents. These types of questions can be addressed through statistical **hypothesis testing**, which is a decision-making process for evaluating claims about a population.

In hypothesis testing, the researcher must define the population under study, state the particular hypotheses that will be investigated, give the significance level, select a sample from the population, collect the data, perform the calculations required for the statistical test, and reach a conclusion.

In other words, we will have a *hypothesis* about the population parameter, and we will use our sample values to conclude whether our assumptions are valid with a certain level of confidence. This is known as *hypothesis testing*. In this chapter, we will be focusing on **testing the population mean** and we will be learning a procedure to test the validity of statements (or claims) made.

A hypothesis test is a procedure whereby information from a random sample is used to determine whether there is evidence for rejecting the claim related to the population parameter.

7.1.1 Null and Alternative Hypothesis

When making a statistical enquiry, we often put forward a hypothesis concerning a population parameter.

For example:

The mean weight of JC2 male students is 70 kg.

This hypothesis is called the **null hypothesis**, denoted by H_0.

The **null hypothesis** (H_0) is a statement about the value of a population parameter which is **assumed to be true** unless sample results imply that it is not.

Contrasting to the null hypothesis is the **alternative hypothesis**, denoted by H_1. For example, we could set up the following alternative hypothesis:

The mean weight of JC2 male students is more than 70 kg.

Usually, we are hoping to reject the null hypothesis and establish the alternative hypothesis.

In order to test the validity of H_0, we consider observations made from a random sample taken from the population and perform a **statistical test**. We write the above statements concisely as

$$H_0: \mu = 70 \quad \text{vs} \quad H_1: \mu > 70$$

7.1.2 1-Tail and 2-Tail Tests

There are two types of test that can be performed, depending on the alternative hypothesis being made. These are **1-tail** or **2-tail** tests.

(i) A **1-tail test** is performed when the alternative hypothesis considers an increase or a decrease in the population parameter;

(ii) A **2-tail test** is performed when the alternative hypothesis considers any change in the parameter (either an increase or a decrease).

For testing of population mean, we formulate the null and alternative hypothesis as follows:

Null Hypothesis: $H_0: \mu = \mu_0$ (note: μ_0 is the hypothesized population mean)

Alternative Hypothesis: $H_1: \mu \neq \mu_0$ (any change) } 2-tail test

$H_1: \mu > \mu_0$ (definite increase)
$H_1: \mu < \mu_0$ (definite decrease) } 1-tail test

EXAMPLE 7.1:

For each of the following situations in **(a)** to **(c)**, write down the null and alternative hypothesis, and state whether they are one or two-tailed tests.

(a) A machine packs flour into bags, with average mass, which are supposed to be 1000 g. After adjustments to the machine, one is interested to find out whether the bags have mass that are more than 1000 g.

(b) A cake shop sells an average of 200 cakes per day. The manager claims that the sales per day should not be 200.

(c) Travelling time along a certain stretch of the CTE has a mean of 5 minutes; after an ERP gantry is set up, the relevant agency wish to check that the travelling time has improved.

SOLUTION:
(b) $H_0: \mu = 200$ $H_1: \mu \neq 200$ 2-tail test
(c) $H_0: \mu = 5$ $H_1: \mu < 5$ 1-tail test

SELF-REVIEW 7.1 [2016/H1/10]
A scientist claims that the mean top speed of cheetahs, in km/h, is 95. The top speed of each cheetah in a random sample of 40 cheetahs is recorded and the mean is found to be 96.3. It is known that the top speed of cheetahs are normally distributed with standard deviation 4.1. How would you test the scientist claim?

SOLUTION:

7.1.3 Level of Significance

The level of significance is the **conditional** probability of wrongly rejecting H_0 **when** H_0 **is true.** It is usually denoted by α. That is $\alpha = \text{P(rejecting } H_0 \mid H_0 \text{ is true)}$.

We usually say that the level of significance is 100α %.

In **EXAMPLE 7.1(a)**, if the level of significance is 5%, then P(wrongly reject $H_0 \mid H_0$ is true) = 0.05
Since α is the probability of rejecting H_0 wrongly when H_0 is true, we usually try to minimise this probability so that we are less likely to commit that error of making a wrong conclusion.

7.1.4 Test Statistic, Critical Region (Rejection region) and Critical Values

A **test statistic** is a random variable whose value is calculated from a given sample and it determines whether the null hypothesis is rejected. It provides the evidence whether to reject H_0 or not.

The **Critical Region (Rejection region)** is a set of values of the test statistic for which the null hypothesis will be rejected. It depends on the *type* (1-tail or 2-tail test) and *level* of the test chosen. The boundaries of the critical region are called the **critical values.**

7.2 Testing of Population Mean[1]

We will be using two methods to test the population mean, namely, the classical method and the *p*-value approach. Both methods follow through from the previous chapters on Normal distribution and Sampling Theory.

We will consider two main cases:

(i) Sample from a Normal population with known variance;
(ii) Large sample from any population, and this is where Central Limit Theorem will be applied.

7.2.1 Sample from a Normal Population with Known Variance

When performing a significance test it is useful to follow a set of procedures:

(a) **Null Hypothesis:** $H_0: \mu = \mu_0$

Alternative Hypothesis: $H_1: \mu \neq \mu_0$ (any change) } 2-tail test
$H_1: \mu > \mu_0$ (definite increase)
$H_1: \mu < \mu_0$ (definite decrease) } 1-tail test

A **significance level** will be given, say α (α usually takes value of 0.1, 0.05, 0.01).

(b) If the population variance σ^2 is known, then **under the assumption that H_0 is true**, we have $X \sim N(\mu, \sigma^2)$. Thus, $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Now, we want to investigate whether there is a significant difference between the sample mean and the population mean given by the null hypothesis.

After standardization, the test statistic is

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Using the observed sample values, we compute the value of the test statistic

$$z_{calc} = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}.$$

(c) Critical region (Rejection region)
The critical region depends on the alternative hypothesis H_1. For example, for $H_1: \mu > \mu_0$, we use the notation

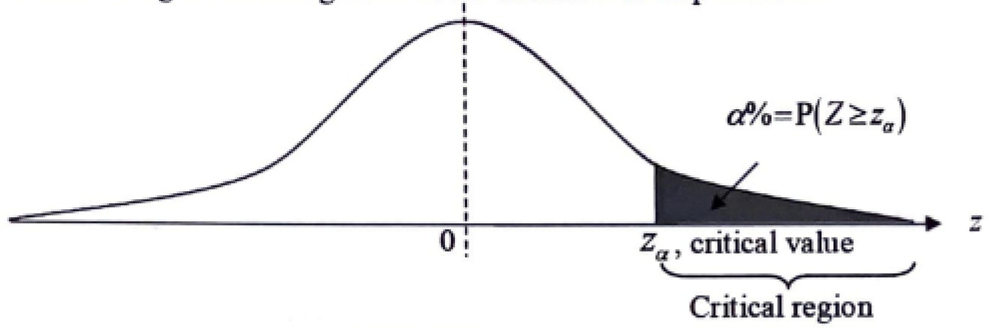
$$P(Z \geq z_\alpha) = \alpha$$

where z_α is called the <u>critical value</u>. With this notation, we can now state the critical region.

[1] Note: Our syllabus only covers hypothesis testing on population mean.

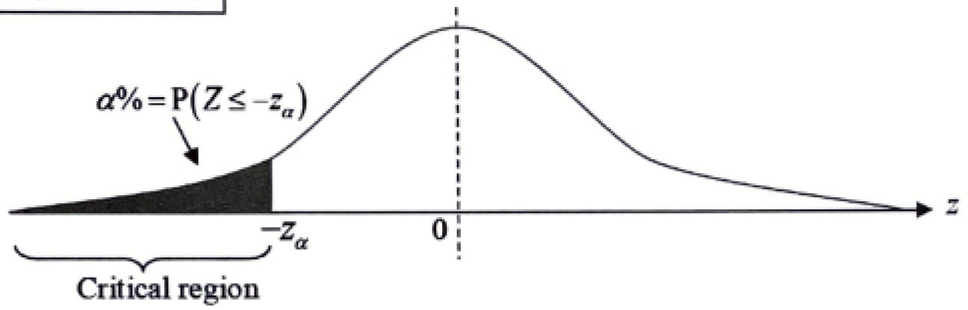
(i) $H_1: \mu > \mu_0$, (1-tail test)

The critical region is the right tail of the normal bell shaped curve.



The rejection region will be $z_{calc} \geq z_\alpha$.

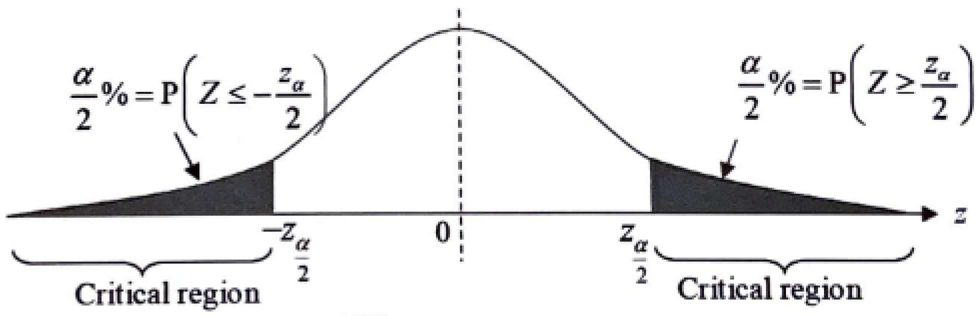
(ii) $H_1: \mu < \mu_0$, (1-tail test)



The rejection region will be $z_{calc} \leq -z_\alpha$ (by symmetry).

(iii) $H_1: \mu \neq \mu_0$, (2-tail test)

As the name of the test implies, the critical region is the two tail-ends of the normal bell shaped curve.



The rejection region will be $|z_{calc}| \geq z_{\frac{\alpha}{2}}$ (by symmetry).

(d) Conclusion

After we have computed the value of the test statistic using the sample data, we have to make conclusion of our finding.

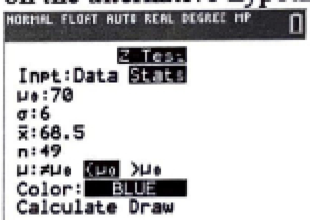
- If the value of the test statistics **lies** in the critical region, we **reject** H_0. There is sufficient evidence to conclude that ...... (write down the statement of H_1 **in context**).
- If the value of the test statistics **does not lie** in the critical region, we **do not reject** H_0. There is insufficient evidence to say that ...... (write down the statement of H_1 **in context**).

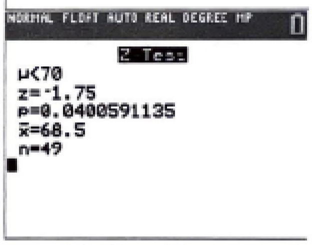
In hypothesis testing, there are **5 steps (T^2C^3 Approach)** that must be followed closely:

- **Step 1:** defining the null and alternative hypothesis and the level of significance;
- **Step 2:** choosing an appropriate test statistic;
- **Step 3:** obtaining the critical region of the test;
- **Step 4:** perform calculation of the test statistic based on random sample values;
- **Step 5:** make conclusion to reject or not to reject H_0.

EXAMPLE 7.2:

Experience has shown that the scores obtained in a particular math test are normally distributed with mean score 70 and variance 36. When the math test is taken by a random sample of 49 students, the mean score is 68.5. Is there sufficient evidence, at 3% level, to conclude that these students have not performed as well as expected?

<table>
<tr><th>SOLUTION:</th><th>THINKZONE:</th></tr>
<tr><td>
Let X denote the scores obtained by a randomly chosen student.
Given $X \sim N(70,36)$, $n = 49$
• To test $H_0: \mu = 70$
$H_1: \mu < 70$
at 3% level of significance.
• Test Statistic: Under H_0, $\bar{X} \sim N\left(70, \frac{36}{49}\right)$ and
test statistic, $Z = \frac{\bar{X} - 70}{6/\sqrt{49}} \sim N(0,1)$
• Critical Region:
• Calculation:
• Conclusion: Since $z_{calc} > -1.881$, it does not lie in the critical region, we do not reject H_0. There is insufficient evidence, at 3% level of significance, to say that these students have not performed as well as expected.
</td><td>
T^2C^3 Approach – these are 5 essential steps.
Note also the keywords "are normally distributed".
$\mu < 70$ means that the students have not performed as well as expected.
Recall: The critical region depends on the alternative hypothesis H_1.




Would the alternative hypothesis be different if the question is changed to "Is there sufficient evidence, at 3% level, to conclude that these students have under-performed?"
</td></tr>
</table>

EXAMPLE 7.3:

The amount of nicotine in a certain brand of cigarettes has mean μ milligrams and standard deviation 0.3 milligrams. The manufacturer claims that μ does not exceed 2.5. A random sample of 100 cigarettes is taken and the sample mean is found to be 2.6 milligrams. Test whether this provides significant evidence, at 7% significance level, that the manufacturer has understated the value of μ. Assume that the amount of nicotine in the brand of cigarettes is normally distributed.

SOLUTION:	**THINKZONE:**
Let X denote the amount of nicotine in the brand of cigarette. Given $X \sim N(\mu, 0.3^2)$, $n = 100$ • To test at 7% significance level • Test Statistic: Under H_0, $\bar{X} \sim N\left(2.5, \frac{0.3^2}{100}\right)$ and test statistic, $Z = \frac{\bar{X} - 2.5}{0.3/\sqrt{100}} \sim N(0,1)$ • Critical Region: Reject H_0 if $z_{calc} \geq z_{0.07} = 1.4758$. • Calculation: $z_{calc} = \frac{2.6 - 2.5}{0.3/\sqrt{100}} = 3.3333$ • Conclusion: Since $z_{calc} > 1.4758$, we reject H_0. There is sufficient evidence, at the 7% level of significance, to conclude that the sample data indicates the average amount of nicotine in the brand of cigarette is more than 2.5 milligram.	Note that the null hypothesis is always $\mu = \mu_0$. Can you identify μ_0? Also, the claim is that the mean mass **does not exceed 2.5 milligrams.**

EXAMPLE 7.4:

A machine produces elastic bands with breaking tension normally distributed with mean 45N and standard deviation 4.36N. On a certain day a sample of 50 was tested and it was found to have a mean breaking tension of 43.46N. Test at the 5% level of significance whether this indicates a change in the mean.

SOLUTION:	**THINKZONE:**
Let X denote the breaking tension of an elastic band. Given $X \sim N(45, 4.36^2)$, $n = 50$ • To test $H_0 : \mu = 45$ $H_1 : \mu \neq 45$ at 5% significance level • Test Statistic: Under H_0, $\bar{X} \sim N\left(45, \frac{4.36^2}{50}\right)$ and test statistic, $Z = \frac{\bar{X} - 45}{4.36/\sqrt{50}} \sim N(0,1)$	Remember the **T²C³ Approach.** Spot the keywords "*breaking tension normally distributed*".
• Critical Region: Reject H_0 if $\lvert z_{calc} \rvert \geq z_{0.025} = 1.96$. • Calculation:	Do you know why it is $z_{0.025}$? How are the values found?

• Conclusion: Since $\lvert z_{calc} \rvert > 1.96$, we reject H_0. There is sufficient evidence, at the 5% level of significance, to conclude that the sample data indicates a change in the mean breaking tension.	Without doing any calculation, would the conclusion be the same if the level of significance is changed to 8%?

REMARKS: The T^2C^3 acronym ensures the 5 essential steps to the testing procedure is adhered to.

SELF-REVIEW 7.2: (J82/II/10 MODIFIED)
The length of string in the balls of string made by a particular manufacturer has mean μ m and variance $27.4\,\text{m}^2$. The manufacturer claims that $\mu = 300$. A random sample of 100 balls of string is taken and the sample mean is found to be 299.2 m. Test whether this provides significant evidence, at the 3% level, that the manufacturer's claim overstates the value of μ. Assume that the length of string in each ball is normally distributed.

SOLUTION:

EXAMPLE 7.5:
It is claimed that the components produced at a particular workshop has mass that is normally distributed with mean 6g and a standard deviation 0.8g. If this claim is accepted, at the 5% level, on the basis of the mean mass obtained from a random sample of 50 components, what is the possible range of values for the mean mass of the sample?

SOLUTION:	THINKZONE:
Let X denote the mass of a component produced by the factory. Given $X \sim N(6, 0.8^2)$, $n = 50$. • To test $H_0: \mu = 6$ $H_1: \mu \neq 6$ at 5% level of significance	

• Test Statistics: Under H_0, $\bar{X} \sim N\left(6, \frac{0.8^2}{50}\right)$ and test statistic, $Z = \frac{\bar{X}-6}{0.8/\sqrt{50}} \sim N(0,1)$. • Critical Region: Reject H_0 if $\lvert z_{calc} \rvert \geq 1.960$. Since H_0 is 'accepted', we have Thus the sample mean mass must lie in the interval [5.78, 6.22].	

NOTE:

- **Not rejecting H_0 is not equivalent to accepting H_0.**
  It just means that we do not have sufficient evidence to reject H_0.
- Smaller significance levels result in smaller area of the critical regions.
- Rejecting H_0 at $\alpha\%$ (say 1%) significance level also means rejecting H_0 at $\beta\%$ (say 5%) level, for all $\beta > \alpha$
- The smaller the level of significance, the stronger the evidence against H_0 if the test statistic falls in the critical region.

7.3 The *p*-value Approach

We introduce the term ***p*-value**, which provides an alternative way to carry out our hypothesis test.

> **Definition:** Corresponding to an observed value of the test statistic, ***p*-value** is the **smallest level of significance** at which the null hypothesis can be rejected.

> ***p*-value** is also defined as the probability that the test statistic would take a value as extreme or more extreme than that actual observed, assuming that H_0 is true.

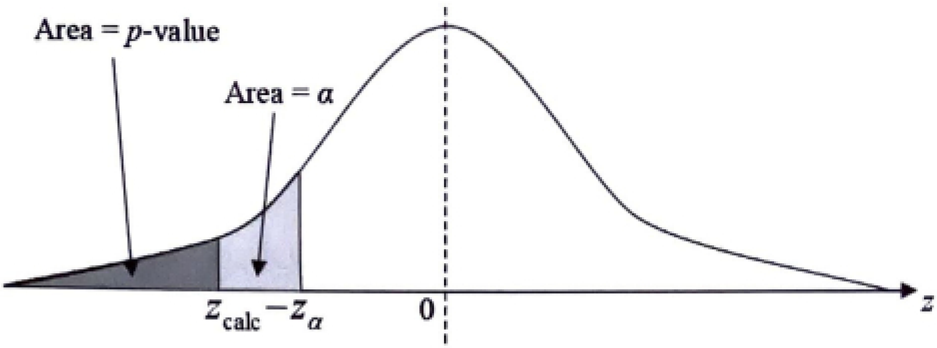
The smaller the *p*-value, the stronger the evidence against H_0 provided by the data.

i.e. **We reject H_0 if *p*-value is less than or equal to the significance level α.**

Suppose Z is the test statistic and z_{calc} is the observed value of the test statistic.

Then

- If we are testing against the alternative hypothesis $H_1: \mu < \mu_0$, then the *p*-value is $P(Z \leq z_{calc})$.
  For example,



Since p-value $= P(Z \le z_{calc}) \le \alpha$, we reject H_0.

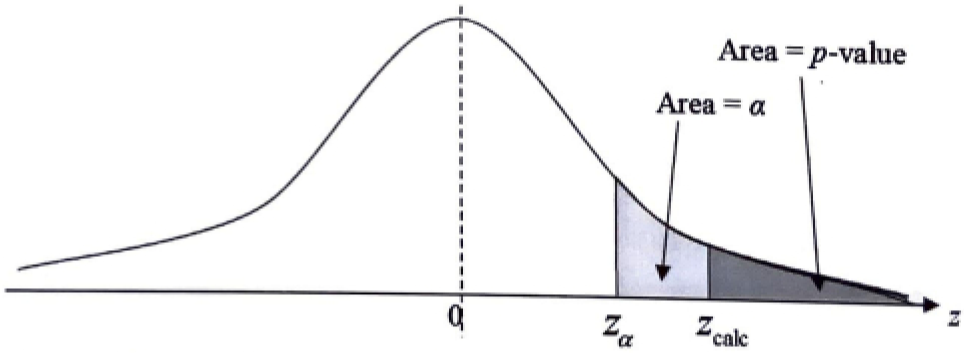
In **EXAMPLE 7.2**, $H_1 : \mu < 70$ and we decided to test at the 3% level of significance, we can compute the p-value $P(\overline{X} \le 68.5) = P(Z \le -1.75)$ and compare it against $\frac{\alpha}{100} = 0.03$

If p-value ≤ 0.03, we reject H_0
If p-value > 0.03, we do not reject H_0

REMARKS:
p-value $= P(\overline{X} \le 68.5) = P(Z \le -1.75) = 0.0401$ (see Annex - Example 7.2) when $H_0 : \mu = 70$ is true.

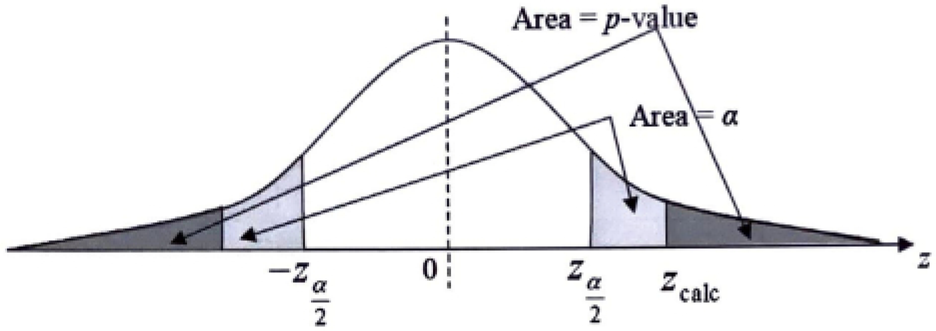
> The **meaning of this p-value** in the context of this question is
> "If $\mu = 70$, out of 100 tests conducted, we expect using the same method to have a test statistics that is than this set of data."

- If we are testing against the alternative hypothesis $H_1 : \mu > \mu_0$, then the p-value is $P(Z \ge z_{calc})$. For example,



Since p-value $= P(Z \ge z_{calc}) \le \alpha$, we reject H_0

- If we are testing against the alternative hypothesis $H_1: \mu \neq \mu_0$, then the p-value is $2\,P\left(Z \geq |z_{cal}|\right)$. For example,



Since p-value $= 2\,P\left(Z \geq |z_{calc}|\right) \leq \alpha$, we reject H_0

In **Example 7.4**, $H_1: \mu \neq 45$, then

p-value $= 2P(\overline{X} \leq 43.46) = 2P(Z \leq -2.4976) = 2P(Z \geq 2.4976) = 0.0125 < 0.05$.

Therefore, we reject H_0.

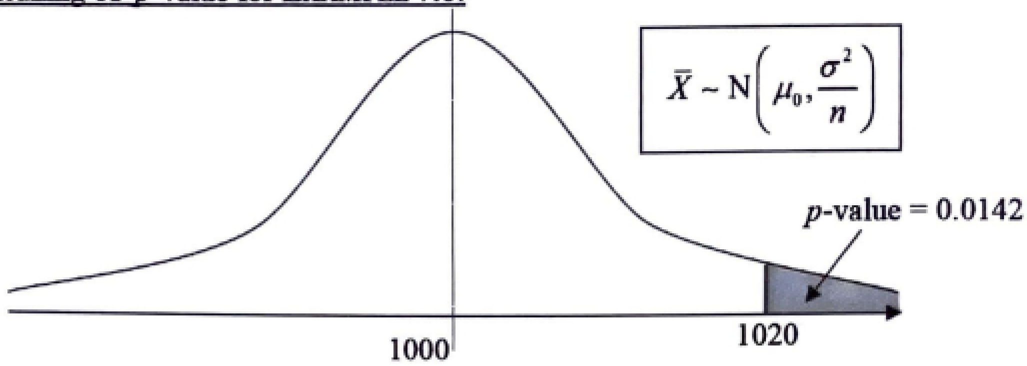
Example 7.6:

A machine packs flour into bags, with average mass, which are supposed to be 1000g, and with variance 50^2 g^2. After adjustments to the machine, it is claimed that the bags have mass that are more than 1000g. By assuming the mass of flour is normally distributed, investigate this claim by finding the p-value for a random sample of 30 bags that has a mean mass of 1020 g. Find the range of level of significance level that will lead to rejection of H_0.

Solution:	**ThinkZone:**
Let X be the mass of a bag of flour. Given $X \sim N(1000, 50^2)$, n =30. To test $H_0: \mu = 1000$ $H_1: \mu > 1000$ at α% level of significance Under H_0, $\overline{X} \sim N\left(1000, \frac{50^2}{30}\right)$. Thus $Z = \frac{\overline{X} - 1000}{50/\sqrt{30}} \sim N(0,1)$ Using the above definition, the p-value for a random sample of 30 bags that has a mean mass of 1020 g is To reject H_0 if	Note that the claim is that after adjustment, the mean mass is **more than** 1000

In this **Example 7.6**, the p-value is 0.0142. It represents the probability of getting a sample mean of 1020 or greater is 0.0142.

Visual meaning of *p*-value for **Example 7.6**:



Example 7.7:
Solve **Example 7.4** using the *p*-value approach.

Solution: | **ThinkZone:**

Let X denote the breaking tension of an elastic band.
Given $X \sim \mathrm{N}(45, 4.36^2)$, $n = 50$

- To test $\mathrm{H}_0 : \mu = 45$
  $\mathrm{H}_1 : \mu \neq 45$
  At 5% significance level.
- Test Statistic: Under H_0, $\bar{X} \sim \mathrm{N}\left(45, \frac{4.36^2}{50}\right)$ and test statistic,
  $$Z = \frac{\bar{X} - 45}{4.36/\sqrt{50}} \sim \mathrm{N}(0,1)$$
- Critical Region: Reject H_0 if *p*-value ≤ 0.05
- Calculation: $z_{calc} = \frac{43.46 - 45}{4.36/\sqrt{50}} = -2.4976$
- Conclusion: Since *p*-value < 0.05, we reject H_0. There is sufficient evidence, at 5% level of significance, to conclude that the sample data indicates a change in the mean breaking tension.

ThinkZone: See Annex on obtaining *p*-value

Example 7.8:
A sample of size 16 is taken from a distribution of $X \sim \mathrm{N}(\mu, 9)$ and a hypothesis test is carried out at the α% level of significance. Given that $\bar{x} = 101.23$, and that the null hypothesis $\mu = 100$ is rejected in favour of the alternative hypothesis $\mu > 100$, find the *p*-value for this test. Hence, find the range of values that α satisfies.

Solution: | **ThinkZone:**

Given $X \sim \mathrm{N}(\mu, 9)$, $n = 16$

- To test $\mathrm{H}_0 : \mu = 100$
  $\mathrm{H}_1 : \mu > 100$
  at α% level of significance
- Test Statistic: Under H_0, $\bar{X} \sim \mathrm{N}\left(\mu, \frac{9}{16}\right)$ and test statistic,
  $$Z = \frac{\bar{X} - 100}{\sqrt{9/16}} \sim \mathrm{N}(0,1)$$

- Critical Region: Reject H_0 if p-value $\le \frac{\alpha}{100}$
- Calculation: $z_{calc} = \frac{101.23-100}{3/\sqrt{16}} = 1.64$

  p-value = 0.050503.

  Since H_0 is rejected, p-value $\le \frac{\alpha}{100}$

7.3.1 <u>Large</u> Sample from <u>any</u> Population

Recall from Chapter 6 that if the population variance σ^2 is unknown, we will estimate it by the unbiased estimator s^2 where

$$s^2 = \frac{n}{n-1} \times \text{sample variance}$$

$$= \frac{n}{n-1}\left(\frac{\sum(x-\bar{x})^2}{n}\right) = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right] \quad \text{(in List MF26)}$$

$$= \frac{1}{n-1}\left(\sum(x-a)^2 - \frac{\left(\sum(x-a)\right)^2}{n}\right)$$

When the sample size is large ($n \ge 30$) and distribution of X is **<u>not normally distributed</u>**, the distribution of $\bar{X}$ is approximately normal by Central Limit Theorem.

Hence (i) $\bar{X} \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$ approximately, if population variance is known, or

(ii) $\bar{X} \sim \text{N}\left(\mu, \frac{s^2}{n}\right)$ approximately, if population variance is unknown.

The test statistic is $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \text{N}(0,1)$ or $Z = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim \text{N}(0,1)$ approximately.

EXAMPLE 7.9:

A manufacturer claims that the average life span of his electric light bulbs is 2000 hours. A random sample of 64 bulbs is tested and the life span x in hours recorded. The results obtained are as follows:

$$\Sigma x = 127808, \quad \Sigma(x-\bar{x})^2 = 9694.6$$

Is there sufficient evidence, at 2% level of significance, that the manufacturer is over-estimating the lifetime of his light bulbs?

SOLUTION:	**THINKZONE:**
Unbiased estimates for population mean and population variance, $\bar{x}=\frac{127808}{64}=1997$, $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=153.8825$ Let μ denote the mean lifetime of an electric light bulb. • To test $H_0:\mu=2000$ $H_1:\mu<2000$ at 2% level of significance • Test Statistic: Under H_0, since $n=64$ is large by Central Limit Theorem, $\bar{X}\sim N\left(2000,\frac{153.8825}{64}\right)$ approximately, and test statistic, $Z=\frac{\bar{X}-2000}{\sqrt{153.8825/64}}\sim N(0,1)$ approx. • Critical Region: Reject H_0 if p-value ≤ 0.02 • Calculations : • Conclusion: Since p-value > 0.02, we do not reject H_0. There is insufficient evidence, at 2% level of significance, to say that the manufacturer is over-estimating the lifetime of his light bulbs.	If the manufacturer is over-estimating the lifetime, the population mean is

SELF-REVIEW 7.3 (N91/II/10)

A supermarket manager investigated the lengths of time that customers spent shopping in the store. The time, x minutes, spent by each of a random sample of 150 customers was measured, and it was found that $\Sigma x = 2871$, $\Sigma x^2 = 60\ 029$. Test, at the 5% level of significance, the hypothesis that the mean time spent shopping by customers is 20 minutes, against the alternative that it is less than this.

EXAMPLE 7.10:

Suppose that 100 tires made by a certain manufacturer lasted on average 21819 miles with a standard deviation of 1295 miles. Test the null hypothesis $\mu = 22000$ miles against the alternative hypothesis $\mu \neq 22000$ miles at the 5% level of significance.

SOLUTION:	**THINKZONE:**
Unbiased estimates for population mean and population variance, $\bar{x} = 21819$, $s^2 = \frac{100}{99} \times \text{sample variance} = \frac{100}{99}(1295)^2 = 1301.5^2$ • To test $H_0 : \mu = 22000$ $H_1 : \mu \neq 22000$ at 5% level of significance • Test Statistics: Under H_0, since n = 100 is large, by Central Limit Theorem, $\bar{X} \sim N\left(22000, \frac{1301.5^2}{100}\right)$ approximately, and test statistic, $Z = \frac{\bar{X} - 22000}{\sqrt{1301.5^2/100}} \sim N(0,1)$ approx. • Critical Region: Reject H_0 if p-value ≤ 0.05 • Calculations : $z_{cal} = \frac{21819 - 22000}{1301.5/\sqrt{100}} = -1.3907$ p-value $= 2P(Z \leq -1.3907) = 0.16432$ • Conclusion: Since p-value > 0.05, we do not reject H_0. There is insufficient evidence, at 5% level of significance, to conclude that the average miles lasted by the tires is not 22000 i.e. $\mu \neq 22000$.	The standard deviation given is obtained from the sample. This is not the population variance nor it is the unbiased estimate of the population variance.

Example 7.11:

A manufacturer of sports equipment has developed a new synthetic fishing line that he claims to have a mean breaking strength of 8 kg. A random sample of 150 fishing lines is tested and the breaking strength x kg recorded. The results are as follows: $\Sigma(x-8) = 74.535$, $\Sigma(x-\bar{x})^2 = 1974.035$. Is there sufficient evidence, at 4% level of significance, that the manufacturer has under-estimated the breaking strength of the fishing line? Is it necessary for the manufacturer to know anything about the distribution of the breaking strength of the fishing line?

SOLUTION:	THINKZONE:
Unbiased estimates for population mean and population variance, $\bar{x}=\frac{1}{n}\sum(x-a)+a=\frac{74.535}{150}+8=8.4969$ $s^2=\frac{1}{n-1}\sum\left(x-\bar{x}\right)^2=13.2486$ • To test $H_0:\mu=8$ $H_1:\mu>8$ at 4% level of significance • Test Statistics: Under H_0, since n = 150 is large, by Central Limit Theorem, $\bar{X}\sim N\left(8,\frac{13.2486}{150}\right)$ approximately, and test statistic, $Z=\frac{\bar{X}-8}{\sqrt{13.2486/150}}\sim N(0,1)$ approximately. • Critical Region: Reject H_0 if p-value ≤ 0.04 • Calculations : $z_{cal}=\frac{8.4969-8}{\sqrt{13.2486/150}}=1.67198$ p-value $=P(Z\geq 1.67198)=0.047264$ • Conclusion: Since p-value > 0.04, we do not reject H_0. There is insufficient evidence, at 4% level of significance, to conclude that the manufacturer has under-estimated the average breaking strength of the fishing line. Since the sample size is large, by Central Limit Theorem, the **mean** breaking strength of the fishing line follows the standard normal distribution. As such, he does not need to know the distribution of the breaking strength of the fishing line.	Note that we are testing if "*the manufacturer's claim is under-stated*".

EXAMPLE 7.12:

The security department of a factory wants to know whether or not the true average time required by the night guard to walk his round is 30 minutes. In a random sample of 100 rounds, the night guard averaged 30.8 minutes with a standard deviation of 3.8 minutes. Determine the minimum value of the level of significance to claim that the true average time is not 30 minutes.

SOLUTION:	THINKZONE:
Unbiased estimates for population mean and population variance, $\bar{x}=30.8$ $s^2=\frac{100}{99}\times\text{sample variance}$ $=\frac{100}{99}\times 3.8^2$ $=14.5859=3.8191^2$ Let X be the time required by the night guard to walk his round.	The standard deviation given is obtained from the sample. This is **not** the population parameter (variance).

• To test $H_0 : \mu = 30$ $H_1 : \mu \neq 30$ at α % level of significance • Test Statistics: Under H_0, since n = 100 is large by and test statistic, • Critical Region: Reject H_0 if $p-\text{value} \leq \frac{\alpha}{100}$ • Calculations : $z_{cal} = \frac{30.8-30}{3.81914/\sqrt{100}} = 2.0947$ p-value = 2 P($Z \geq 2.0947$) = 0.036198 Given the true average time is not 30 minutes i.e. to reject H_0. The minimum value of level of significance is	What would the minimum value of level of significance be if the answer is to be given to 2 s.f. of accuracy?

Example 7.13: (H2 Mathematics Specimen/P2/Q10)
The average time required for the manufacture of a certain type of electronic control panel is 17 hours. An alternative manufacturing process is trialled, and the time taken, t hours, for the manufacture of each of 50 randomly chosen control panel using the alternative process is recorded. The results are summarized as follows.

$$n = 50, \quad \Sigma t = 835.7, \quad \Sigma t^2 = 14067.17.$$

The Production Manager wishes to test whether the average time taken for the manufacture of a control panel is different using the alternative process, by carrying out a hypothesis test.

(i) Explain whether the Production Manager should use a 1-tail test or 2-tail test.

(ii) Explain why the Production Manager is able to carry out a hypothesis test without knowing anything about the distribution of the times taken to manufacture the control panels.

(iii) Find unbiased estimates of the population mean and variance and carry out the test at the 10% level of significance for the Production Manager.

(iv) Suggest a reason why the Production Manager might be prepared to use an alternative process that takes a longer average time than the original process.

The Finance Manger wishes to test whether the average time taken for the manufacture of a control panel is **shorter** using the alternative process. The Finance Manager finds that the average time taken for the manufacture of each of 40 randomly chosen control panels, using the alternative process, is 16.7 hours. He carries out a hypothesis test at the 10% level of significance.

(v) Explain, with justification, how the population variance of the times will affect the conclusion made by the Finance Manager.

SOLUTION:		**THINKZONE:**
(i)	The Production Manager should use a 2-tail test since he is interested on find out whether the average time taken has changed.	*"average time taken… is different"*
(ii)	Since the sample size is large, by Central Limit Theorem, the test statistics will follow the standard normal distribution approximately. Thus he does not need to have the knowledge of the distribution.	
(iii)	The unbiased estimates of the population mean and variance are: $\bar{t}=\frac{\Sigma t}{n}=16.714,$ $s^2=\frac{1}{n-1}\left[\Sigma t^2-\frac{(\Sigma t)^2}{n}\right]$ $=\frac{1}{49}\left[14067.17-\frac{835.7^2}{50}\right]=2.0261$ • To test $H_0:\mu=17$ $H_1:\mu\neq17$ at 10% level of significance • Test Statistics: Under H_0, since n = 50 is large, by Central Limit Theorem, $\bar{T}\sim N\left(17,\frac{2.0261}{50}\right)$ approximately and test statistic, $Z=\frac{\bar{T}-17}{\sqrt{2.0261/50}}\sim N(0,1)$ approximately. • Critical Region: Reject H_0 if p-value ≤ 0.1 • Calculations : $z_{cal}=\frac{16.714-17}{\sqrt{2.0261/50}}=-1.4208$ $p\text{-value}=2P(Z\leq-1.4208)=0.15539$ • Conclusion: Since p-value > 0.1, we do not reject H_0. There is insufficient evidence, at 10% level of significance, to conclude that the average time taken for the manufacture of a control panel is different using the alternative process.	
(iv)	The control panel manufactured by the alternative process might be of a higher quality.	
(v)	• To test $H_0:\mu=17$ $H_1:\mu<17$ at 10% level of significance	The key phrase is *"how the population variance of the times will affect…"*, we need to find a condition on the population variance to decide the outcome of test.

	• Test Statistics: Under H_0, since $n = 40$ is large by Central Limit Theorem, $\bar{T} \sim N\left(17, \frac{\sigma^2}{40}\right)$ approximately and test statistic, $Z = \frac{\bar{T}-17}{\sigma/\sqrt{40}} \sim N(0,1)$ approximately • Critical Region: Reject H_0 if $z_{cal} \le -1.282$ • Calculations : $z_{cal} = \frac{16.7-17}{\sigma/\sqrt{40}}$ To reject H_0, $z_{cal} = \frac{16.7-17}{\sigma/\sqrt{40}} \le -1.282 \Rightarrow \sigma^2 \le 2.19$ Thus, if the population variance is less than 2.19, the Finance Manager will conclude that the alternative process is shorter.	

Summary

1) **Hypothesis Test on the population mean**

Note that a sample random sample of size n is assumed in each case.

Variance	Population	Sample size	Test statistic	Remark
Known	Normal	Any	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	
Known	Any	Large	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ approx	Use CLT if non-normal
Unknown	Any	Large	$Z = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim N(0,1)$ approx	Use CLT if non-normal

2) **To Test : Null Hypothesis:** $H_0: \mu = \mu_0$ vs

Alternative Hypothesis	**Classical Method**	***p*-value Method**
$H_1: \mu \ne \mu_0$ (any change)	**Reject H₀ if** $\lvert z_{calc} \rvert \ge z_{\frac{\alpha}{2}}$	**Reject H₀ if** p-value $\le \frac{\alpha}{100}$
$H_1: \mu > \mu_0$ (definite increase)	**Reject H₀ if** $z_{calc} \ge z_\alpha$	**Reject H₀ if** p-value $\le \frac{\alpha}{100}$
$H_1: \mu < \mu_0$ (definite decrease)	**Reject H₀ if** $z_{calc} \le -z_\alpha$	**Reject H₀ if** p-value $\le \frac{\alpha}{100}$

3) **Extra Reading on *p*-value:**

Example (based on information in **Example 7.6**):

Take X to be mass of a bag of flour where $X \sim N(1000, 50^2)$, and suppose five different samples of 30 bags of flour give sample means $\bar{x}$ of

i) 990 g ii) 1010 g iii) 1013 g. iv) 1020 g v) 1030 g

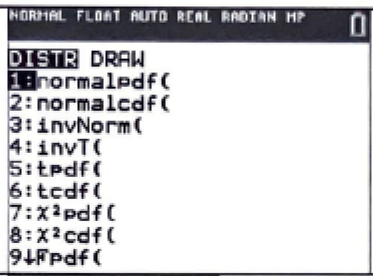
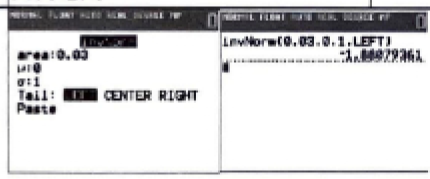
Investigate the claim that the mean mass of a bag of flour is 1000 g, versus the alternative claim that it is more than 1000 g by finding the p-values for the 5 samples.

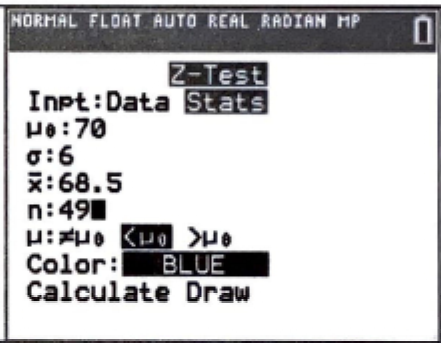
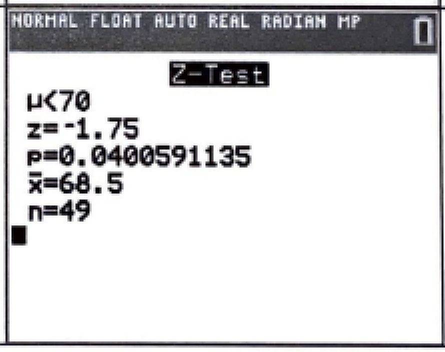
Which sample provides the strongest evidence to reject H_0? Which sample provides the least evidence to reject H_0? What do you notice about the p-value for these samples?

SOLUTION:	THINKZONE:
$H_0: \mu = 1000$ $H_1: \mu > 1000$ For sample size 30, assuming H_0, we have $\bar{X} \sim N\left(1000, \frac{50^2}{30}\right)$ and the respective p-values (i.e., probabilities) for the different samples are: i) $P(\bar{X} \geq 990) =$ ii) $P(\bar{X} \geq 1010) =$ iii) $P(\bar{X} \geq 1013) =$ iv) $P(\bar{X} \geq 1020) =$ v) $P(\bar{X} \geq 1030) =$ The sample with sample mean likely to provide the strongest evidence to reject H_0. The sample with sample mean is likely to provide the least evidence to reject H_0 as the sample mean is not that far. The sample with sample mean providing the strongest evidence to reject H_0 has the p-value. However the sample with sample mean providing the least evidence to reject H_0 has the p-value.	$H_0: \mu = 1000$ $H_1: \mu > 1000$ Important note: we calculate GREATER than sample mean value in this exercise because the alternative hypothesis is that the mean mass is more than 1000 g. Recall that the more extreme the value of the sample mean, that is the value being larger than 1000, the more evidence we have to support the alternative hypothesis. Note that in this case, the p-value is smallest for the sample that provides the strongest evidence to reject H_0.

Annex A – GC Keystrokes for the Examples

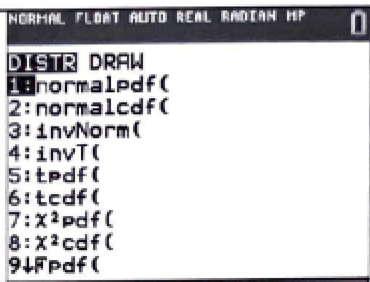
EXAMPLE 7.2:

<table>
<tr>
<td>Press 2nd VARS, highlight DISTR and choose 3: invNorm(</td>
<td>
</td>
</tr>
<tr>
<td>Then press ENTER and enter as shown.
Go to "Paste"
Then press ENTER ENTER</td>
<td>
</td>
</tr>
<tr>
<td>Choose 1: Z-Test.
Select and enter as shown.

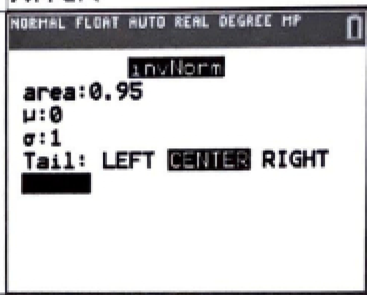
Select Calculate and press ENTER.</td>
<td>
</td>
</tr>
<tr>
<td>The value of z_{cal} is shown
$z_{calc} = -1.75$
p-value = 0.0401</td>
<td>
</td>
</tr>
</table>

EXAMPLE 7.4 AND EXAMPLE 7.7:

Press [2nd] [VARS], highlight DISTR and choose 3: invNorm(

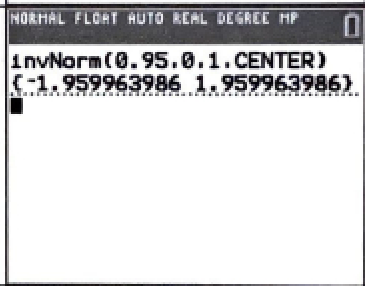


Then press [ENTER] and enter as shown. (Since level of significance is 0.05, the area of the center region should be 0.95)



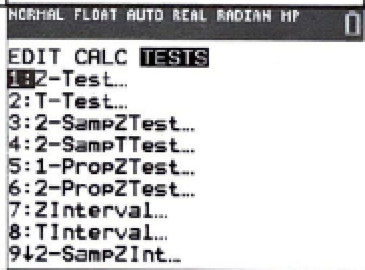
Go to "Paste"

Then press [ENTER] [ENTER]

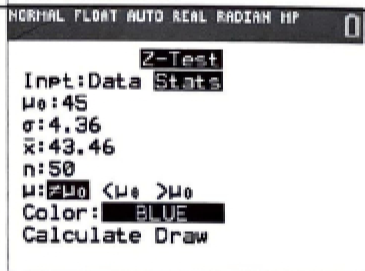


Press [STAT] and select TESTS.

Next choose 1: Z-Test and press [ENTER]

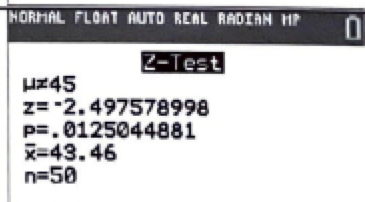


Select and enter as shown.
(Make sure Stats is highlighted)
Since we are testing $H_1 : \mu \neq 45$, we highlight $\neq \mu_0$ to conform to H_1



Select Calculate and press [ENTER].

The value of z_{cal} is shown
$z_{calc} = -2.4976$
p-value = 0.0125



NANYANG JUNIOR COLLEGE

DEPARTMENT OF MATHEMATICS

H2 Mathematics **Tutorial S7 – Hypothesis Testing** **Year 2/2020**

1. **[N2004/II/30 (OR)]**

The mean of a random variable X is denoted by μ. A sample of 50 random observations of X is taken and the results are summarised by $\sum x = 527.1$.

(i) It is given that the population variance is 15. Carry out a 2-tail test of the null hypothesis $\mu = 9.5$, at the 5% significance level. [do not reject H_0]

(ii) It is given instead that $\sum x^2 = 6172.31$. In a 1-tail test of the null hypothesis $\mu = 11.5$, the alternative hypothesis is accepted. State the alternative hypothesis, and find an inequality satisfied by the significance level of the test. [$H_1 : \mu < 11.5$, $\alpha \geq 2.80$]

2. A random sample of 90 batteries, used in a particular model of mobile phone, is tested and the 'standby-time', x hours, is measured. The results are summarised by $\Sigma x = 3040.8$ and $\Sigma x^2 = 115773.66$. Test, at the 1% significance level, whether the mean standby-time is less than 36.0 hours. [Do not reject H_0]

State, with a reason, whether in using the above test, it is necessary to assume that the 'standby-time' has a normal distribution.

In a test at the 5% significance level, it is found that there is significant evidence that the population mean talk-time is less than 5 hours. Using only this information, and giving a reason in each case, state whether each of the following statements is (i) necessarily true, (ii) necessarily false, or (iii) neither necessarily true nor false.

(i) There is significant evidence at the 10% significance level that the population mean talk-time is less than 5 hours.

(ii) There is significant evidence at the 5% significance level that the population mean talk-time is not 5 hours.

3. At an early stage in analysing the marks scored by the large number of candidates in an examination paper, the Examining Board took a random sample of 250 candidates and found that the mark, x, of these candidates gave $\sum x = 11872$, $\sum x^2 = 646193$. Using these figures obtained in this sample, the null hypothesis $\mu = 49.5$ is tested against the alternative hypothesis $\mu < 49.5$ at the α% significance level. Determine the set of values of α for which the null hypothesis is rejected in favour of the alternative hypothesis.

It is subsequently found that the population mean and standard deviation for the paper are 45.292 and 18.761 respectively. Find the probability of a random sample of size 250 giving a sample mean at least as high as one found in the sample above.

[$\alpha > 4.02$; 0.0321]

It is not the answers that enlightens but the questions... Eugene Ionesco

4. **[ACJC/2018/II/05]**
The management of a supermarket wishes to analyse the effectiveness of an advertising campaign. Before the campaign, the daily takings, $\$X$, is normally distributed with mean \$72 300 and standard deviation \$4410.
Immediately after the campaign, a sample of 30 shopping days is taken and the mean daily takings was found to be $\$\bar{x}$. A test is carried out, at the 5% significance level, to determine whether the campaign is effective, assuming that there is no change in the standard deviation of the daily takings after the campaign.
(i) State appropriate hypotheses for the test.
(ii) By stating a necessary assumption, find the set of values of $\bar{x}$, to the nearest dollar, for which the result of the test would be to reject the null hypothesis at 5% level of significance.
(iii) The 30 readings taken on X immediately after the campaign is summarised as follows.

$$\sum(x-70\ 000)=129\ 000$$

By finding $\bar{x}$, state the conclusion of the test, at the 5% significance level, whether the campaign is effective. [$\bar{x}>73624$; reject H_0]

5. **[2012/II/6]**
On a remote island a zoologist measures the tail lengths of a random sample of 20 squirrels. In a species of squirrel known to her, the tail lengths have mean 14.0 cm. She carries out a test, at the 5% significance level, of whether squirrels on the island have the same mean tail length as the species known to her. She assumes that the tail lengths of squirrels on the island are normally distributed with standard deviation 3.8 cm.
(i) State appropriate hypotheses for the test.
The sample mean tail length is denoted by $\bar{x}$ cm.
(ii) Use an algebraic method to calculate the set of values of $\bar{x}$ for which the null hypothesis would not be rejected. (Answers obtained by trial and improvement from a calculator will obtain no marks.) [$\{x\in\mathbb{R}:12.3<\bar{x}<15.7\}$]
(iii) State the conclusion of the test in the case where $\bar{x}=15.8$. [Rejected]

6. **[SAJC/2019/II/11]**
The time T seconds required for a computer to boot up, from the moment it is switched on, is a normally distributed random variable. The specifications for the computer state that the population mean time should not be more than 30 seconds. A Quality Control inspector checks the boot up time using a sample of 25 randomly chosen computers.
A particular sample yielded $\sum t=802.5$ and $\sum t^2=26360.25$.
(i) Calculate the unbiased estimates of the population mean and variance. [32.1, 25]
(ii) What do you understand by the term "unbiased estimate"?
(iii) Test, at the 5% level of significance level, whether the specification is being met. Explain in the context of the question, the meaning of "5% level of significance". [p-value = 0.0179 , reject H_0].
(iv) Find the range of values of $\bar{t}$ such that the specification will be met in the test carried out in part (iii). [$0<\bar{t}<31.6$]
(v) A new Quality Control policy is that when the specification is not met, all the computers will be sent back to the manufacturer for upgrading. The inspector tested a second random sample of 25 computers, and the boot up time, y seconds, of each computer is measured, with $\bar{y}=32.4$. Using a hypothesis test at the 5% level of significance, find the range of values of the population standard deviation such that the computers will not be sent back for upgrading. [$\sigma>7.30$]

It is not the answers that enlightens but the questions... Eugene Ionesco

7. **[DHS/2019/II/9]**
The time taken, T (in minutes), for a 17-year-old student to complete a 5-km run is a random variable with mean 30. After a new training programme is introduced for these students, a random sample of n students is taken. The mean time and standard deviation for the sample are found to be 28.9 minutes and 4.0 minutes respectively.

(a) Find the unbiased estimate of the population variance in terms of n. $[s^2 = \frac{16n}{n-1}]$

(b) Using $n = 40$,

(i) carry out a test at the 10% significance level to determine if the mean time taken has changed. State appropriate hypotheses for the test and define any symbols you use.
$[\,p\text{-value} = 2\mathrm{P}(\overline{T} \le 28.9) = 0.0859 \le 0.10, \text{Reject } \mathrm{H}_0\,]$

(ii) State what it means by the p-value in this context.

(iii) Give a reason why no assumptions about the population are needed in order for the test to be valid.

(c) The trainer claims instead that the new training programme is able to improve the mean of T, 30 minutes, by at least 5%. The school wants to test his claim.

(i) Write down the null and alternative hypothesis.

(ii) Using the existing sample, the school carried out a test at 1% significance level and found that there was sufficient evidence to reject the trainer's claim. Find the set of values that n can take, stating any necessary assumption(s) needed to carry out the test. $[\{n \in \mathbb{Z} : n \ge 543\}]$

8. **[EJC/2018/I/08]**
A teacher, Mr. Ku, suspects that the average time a student spends on his or her mobile phone per day is μ_0 minutes. He selected a random sample of 97 students in the school who own mobile phones and recorded the amount of time each student spent on his or her phone in a randomly selected day. The results are displayed in the table below.

Time spent per day (to nearest minute)	60	65	72	90	110	180
Number of people	11	20	32	18	10	6

(i) Calculate unbiased estimates of the population mean and variance of the time a student spends on his or her mobile phone per day. [83.1, 842]

The null hypothesis that the average time a student spends on his or her mobile phone per day is μ_0 minutes is tested, at 5% level of significance, against the alternative hypothesis that the average time a student spends on his or her mobile phone per day differs from μ_0 minutes.

(ii) Determine the range of values of μ_0 for which the null hypothesis is rejected.
$[\mu_0 \ge 88.9 \quad \text{or} \quad \mu_0 \le 77.3]$

(iii) Explain, in the context of this question, the meaning of 'at 5% level of significance'.

9. Experimental data concerning a variable X, which measures the reliability of a certain electronic component, is as follows : $\sum x = 1164.2$, $\sum x^2 = 13911.6$, $n = 100$. Calculate the unbiased estimates of the population mean and standard deviation from this figures. Explain whether, on the evidence of this sample, you would reject the hypothesis that the mean value of X is 12. Figures collected over a long period have established that the mean and standard deviation of X are 12 and 2 respectively. After a change in the manufacturing process, it is expected that the mean will have increased, but it may be assumed that the standard deviation remains equal to 2. A sample of n values of X is taken, with sample mean m; if m is greater than some critical value, it will be accepted that the mean has in fact increased, but if m is less than the critical value, the increase is not established. State carefully appropriate null and alternative hypotheses for this situation, and find, in terms of n, the critical value for a 1% significance level. [11.642, 1.911, do not reject at 5% level of sig : $12 + \frac{4.652}{\sqrt{n}}$]

10. **[H1 N2008/I/10]**
A consumer association is testing the lifetime of a particular type of battery that is claimed to have a lifetime of 150 hours. A random sample of 70 batteries of this type is tested and the lifetime, x hours, of each battery is measured. The results are summarised by

$$\Sigma x = 10\,317, \qquad \Sigma x^2 = 1\,540\,231.$$

The population mean lifetime is denoted by μ hours. The null hypothesis $\mu = 150$ is to be tested against the alternative hypothesis $\mu < 150$. Find the p-value of the test and state the meaning of this p-value in the context of the question. [5]
A second random sample of 50 batteries of this type is to be tested and the lifetime, y hours, of each battery is measured, with results summarized by

$$\Sigma y = 7331, \qquad \Sigma y^2 = 1\,100\,565.$$

Combining the two samples into a single sample, carry out a test, at the 10% significance level, of the same null and alternative hypotheses. [6]
[p-value = 0.0975 < 0.10, reject H_0]

It is not the answers that enlightens but the questions... Eugene Ionesco

Assignment:

1. **[2013/Prelim/RI/II/8]**

The manufacturer of a particular type of battery claims that the lifetime is distributed with mean 150 hours and standard deviation 16.877 hours. A random sample of n batteries, where n is large, is taken and the lifetime, t hours, of each battery is measured.

It is found that $\sum t = 10\ 317$.

A test is carried out to determine whether the manufacturer has overstated the mean lifetime of the batteries, and the p-value is found to be 0.0975.

(i) Write down appropriate hypotheses for the test, defining any symbols you use. State, in context, the conclusion of the test if the level of significance is 8%. [3]

(ii) Show that n satisfies the equation $\frac{10317}{n} + \frac{21.871}{\sqrt{n}} - 150 = 0$, and solve it numerically. [5]

(iii) State, with a reason, whether it is necessary to assume that the lifetime of the batteries is normally distributed in order to carry out the calculation in (ii). [1]

2. **[2017 Prelim ACJC /II/7 modified]**

It has been suggested that the optimal pH value for shampoo should be 5.5, to match the pH level of healthy scalp. Any pH value that is too low or too high may have undesirable effects on the user's hair and scalp. A shampoo manufacturer wants to investigate if the pH level of his shampoo is at the optimal value, by carrying out a hypothesis test at the 10% significance level. He measures the pH value, x, of n randomly chosen bottles of shampoo, where n is large.

(a) In the case where $n = 30$, it is found that $\sum x = 178.2$ and $\sum x^2 = 1238.622$.

(i) Find unbiased estimates of the population mean and variance, and carry out the test at the 10% significance level. [6]

(ii) Explain if it is necessary for the manufacturer to assume that the pH value of a bottle of shampoo follows a normal distribution. [1]

(iii) State what it means by the p-value in this context. [1]

(b) In the case where n is unknown, assume that the sample mean is the same as that found in **(a)**.

(i) State the critical region for the test. [1]

(ii) Given that n is large and that the population variance is found to be 6.5, find the greatest value of n that will result in a favourable outcome for the manufacturer at the 10% significance level. [3]

EXTRA PRACTICE QUESTIONS

1. **2014/DHS/II/11**

In Factory A, it is claimed that the mean mass of each bag of beans produced is 22 kg. To investigate this claim, the mass, x kg, of a random sample of 50 bags of beans are obtained and summarised by:

$$\sum(x-18)=165.3 \text{ and } \sum(x-18)^2=876.5.$$

(i) What do you understand by the term "unbiased estimate"? [1]
(ii) Find the unbiased estimates of the population mean and variance. [2]
(iii) Test at the 5% level of significance whether this claim is valid. [4]
(iv) State the meaning of the p-value obtained in part **(iii)**. [1]

[(ii) 21.3, 6.74, Do not reject H_0, p-value = 0.0586]

2. **2013/MI/II/9**

In XYZ School, the students' mean running time in the 2.4 km run is claimed to be 705 seconds. A programme to help the students perform better in the run is introduced. After the programme, a sample of 100 students was taken and their mean running time is 690 seconds. At a significance level of 7%, a z-test was done to test if the students are performing better. It was concluded that there is sufficient evidence that the students are performing better in the 2.4 km run.

(i) In the test above, explain if any assumption about the students' 2.4 km running times is necessary. [1]
(ii) If a two-tailed test had been conducted, is it necessarily true, necessarily false or not necessarily true or false that the conclusion will remain the same? [2]
(iii) What can you do to increase the p-value in the test in part **(b)**? [1]

3. **2017/RVHS/II/11**

Physicists are conducting an experiment involving collisions between protons and anti-protons. The mean amount of energy, $\overline{x}$ MeV, released in n collisions is found to be 1864 MeV.
One model predicts the energy released would be 1860 MeV with standard deviation 40 MeV. This is tested at a 1% level of significance against a newer model that claims a higher value.

(i) Find the least value of n such that the hypothesis that the mean amount of energy released is 1860 MeV is rejected. [5]

Given instead that $n = 600$.

(ii) Calculate the p-value and state its meaning in context of the question. [3]
(iii) State, with a reason, whether it is necessary to assume the amount of energy released in collisions to be normally distributed for this test to be valid. [1]

Two-sigma is an indicative of how confident researchers feel their results are. For researchers to feel confident, they must be able to produce a "two-sigma" result – that is the experimental result must be at least two standard deviations away from the predicted mean under the null hypothesis.

(iv) Calculate the level of significance that corresponds to a "two-sigma" test. Hence, using your answer from part **(ii)** determine whether the experiment has met the "two-sigma" threshold. [3]

[(i) 542, (ii) p-value = 0.00715, (iv) 2.28%]

4. **2017/CJC/II/9**

A computer hard drive manufacturer claims that the mean usage hours before failure of their R series hard drives is 50 thousand hours. A technology columnist wishes to investigate this claim and collected the usage hours, t thousand hours for each of the 50 randomly chosen hard drives which were submitted to the local service centre for drive failures. The data is summarized as follows.

$$n = 50, \quad \Sigma t = 2384.5, \quad \Sigma t^2 = 115885.23$$

The technology columnist wants to use hypothesis testing to test whether the mean usage hours before failure of a hard drive is different from what the manufacturer has stated.

(i) Explain whether it is necessary for the columnist to know about the distribution of the usage hours before failure of the drives in order to carry out a hypothesis test. [1]

(ii) Find the unbiased estimates of the population mean and variance and carry out the test at 1% level of significance for the columnist. [6]

The columnist published the data and the results of the hypothesis testing in an online article.

(iii) Suggest a reason why the test result might not be useful to a reader of the article who is deciding whether to buy an R series hard drive from the manufacturer. [1]

(iv) State an alternative hypothesis that is more relevant to the decision making process and explain whether the result will differ from the earlier test carried out by the columnist at 1% level of significance. [2]

(v) State a necessary assumption that was made for all the tests carried out. [1]

[(ii) 47.69, 44.3, p-value = 0.0141, do not reject H_0]

5. **2017/SAJC/II/11**

The Kola Company receives a number of complaints that the volume of cola in their cans are less than the stated amount of 500 ml. A statistician decides to sample 50 cola cans to investigate the complaints. He measures the volume of cola, x ml, in each can and summarised the results as follows:

$$\sum x = 24730, \sum x^2 = 12242631.$$

(i) Find unbiased estimates of the population mean and variance correct to 2 decimal places and carry out the test at the 1% level of significance. [6]

(ii) One director in the company points out that the company should test whether the volume of cola in a can is 500 ml at the 1% significance level instead. Using the result of the test conducted in (i), explain how the p-value of this test can be obtained from p-value in part (i) and state the corresponding conclusion. [2]

The head statistician agrees the company should test that the volume of cola in a can is 500 ml at the 1% level of significance. He intends to make a simple rule of reference for the production managers so that they will not need to keep coming back to him to conduct hypothesis tests. On his instruction sheet, he lists the following:

1. Collect a random sample of 40 cola cans and measure their volume.
2. Calculate the mean of your sample, $\bar{x}$ and the variance of your sample, s_x^2.
3. Conclude that the volume of cola differs from 500 ml if the value of $\bar{x}$ lies…

(iii) Using the above information, complete the decision rule in step 3 in terms of s_x. [4]

A party organiser has n cans of cola and $2n$ packets of grape juice. Assume now that the volume of a can of cola has mean 500 ml and variance 144 ml^2, and the volume of a packet of grape juice has mean 250 ml and variance 25 ml^2. She mixes all the cola and grape juice into a mocktail, which she pours into a 120-litre barrel. Assume that n is sufficiently large and that the volumes of the cans of cola and packets of grape juice are independent.

(iv) Show that if the party organiser wants to be at least 95% sure that the barrel will not overflow, n must satisfy the inequality $1000n + 22.9\sqrt{n} - 120,000 \le 0$. [4]

[(i) 494.60, 228.02, p-value = 0.00572, Reject H_0, (iii) $500 - 0.412s_x \ge \bar{x}$ or $\bar{x} \ge 500 + 0.412s_x$]

6. **2017/TPJC/II/9**

The town council is investigating the mass of rubbish in domestic dustbins. In 2016, the mean mass of rubbish in domestic dustbins was 20.0 kg per household per week. The town council starts a recycling initiative and wishes to determine whether there has been a reduction in the mass of rubbish in domestic dustbins.

The mass of rubbish in a domestic dustbin is denoted by X kg. A random sample of 50 domestic dustbins is selected and the results are summarised as follows.

$$n = 50 \qquad \sum x = 924.5 \qquad \sum x^2 = 18249.2$$

(i) Explain what is meant in this context by the term 'a random sample'. [2]

(ii) Explain why the town council is able to carry out a hypothesis test without knowing anything about the distribution of the mass of rubbish in domestic dustbins. [2]

(iii) Find the unbiased estimates of the population mean and variance and carry out the test at the 1% level of significance for the town council. [6]

(iv) Use your results in part (iii) to find the range of values of n for which the result of the test would be that the null hypothesis is rejected at the 1% level of significance. [3]

[(iii) 18.5, 23.6, p-value = 0.013937 do not reject H_0, $n \ge 56$, $n \in \mathbb{Z}^+$]

7. **2019/HCI/II/9**

A company purchased a machine to pack shower gel into its bottles. The expected mean volume of shower gel in a bottle is 950 ml.

(a) The floor supervisor believes that the machine is packing less amount of shower gel than expected. A random sample of 80 bottles is taken and the data is as follows:

Volume of shower gel in a bottle (correct to nearest ml)	948	949	950	951	952	953	955
Number of bottles	9	22	36	6	4	1	2

(i) Find unbiased estimates of the population mean and variance, giving your answers correct to 2 decimal places. [2]

(ii) Write down the appropriate hypotheses to test the floor supervisor's belief. You should define any symbols used. [2]

(iii) Using the given data, find the p-value of the test. State what is meant by this p-value in the context of this question. [2]

(iv) It was concluded at α% level of significance that the machine is indeed packing less amount of shower gel than expected. State the set of values of α. [1]

(b) Due to a change in marketing policy, the machine is being recalibrated to pack smaller bottles of shower gel with mean volume of 250 ml. The volume of a recalibrated bottle of shower gel is denoted by Y ml. A random sample of 50 bottles of y ml each is taken and the data obtained is summarised by:

$$\sum(y-250) = -25, \qquad \sum(y-250)^2 = k.$$

average weight for the second sample is k grams.
Based on the combined sample of 100 mini breads, find the range of values of k such that the customer's claim is valid at the 4% level of significance. [4]

[44.6, 3.80, p-value = 0.0482 do not reject H_0, $k \leq 45.1$]

10. **2019/VJC/II/9**

Exposure to Volatile Organic Compounds (VOCs), which have been identified in indoor air, is suspected as a cause for headaches and respiratory symptoms. Indoor plants have not only a positive psychological effect on humans, but may also improve the air quality. Certain species of indoor plants were found to be effective removers of VOCs.
A commonly known VOC is Benzene. The following data gives the benzene levels, x (in ppm) in 40 test chambers containing the indoor plant *Epipremnum aureum*.

$$n = 40, \sum(x-26.0) = -30.1, \sum(x-26.0)^2 = 214.61.$$

The initial mean Benzene level (in ppm) without *Epipremnum aureum* was found to be 26.0.

(i) Test, at the 5% level of significance, the claim that the mean Benzene level, μ (in ppm), has decreased as a result of the indoor plant *Epipremnum aureum*. You should state your hypotheses clearly. [5]

(ii) State, giving a reason, whether there is a need to make any assumptions about the population distribution of the Benzene level in order for the test to be valid. [2]

The Benzene levels of another 50 test chambers containing the indoor plant *Epipremnum aureum* were recorded, The sample mean is $\bar{x}$ ppm and the sample variance is 8.33 ppm^2.

(iii) The acceptance region of a test of the null hypothesis $\mu = 26.0$ is $\bar{x} > 25.1$. State the alternative hypothesis and find the level of significance of the test. [4]

(iv) If the null hypothesis is $\mu = \mu_0$, where $\mu_0 > 26.0$, would the significance level of a test with the same acceptance region in part **(iii)** be larger or smaller than that found in part **(iii)**? Give a reason for your answer. [2]

[(i) p-value = 0.0160, Reject H_0, (iii) 1.45% (iv) smaller]