3 One-sample and two-sample problems

This chapter will discuss the following two problems.

- One–sample problems: we have a ***single random sample*** from a ***Normally distributed population*** with mean μ and variance σ^2. We wish to make inferences about these parameters. There are two basic approaches to such statistical inference:
  (i) ***Estimation*** (including confidence intervals), and
  (ii) ***Hypothesis testing***.
- Two–sample problems: we wish to compare two populations. Both will be assumed Normal. We may wish to test for a common ***mean***.

3.1 Estimation and confidence intervals

Definition: A ***point estimate*** of a parameter is a sample statistic (i.e. a value calculated from a sample) which is chosen to be as close to the (unknown) value of the parameter as possible. For example,

- $\bar{x}$ estimates μ, and
- s estimates σ.

However, both $\bar{x}$ and s vary from sample to sample! We need to know how reliable our estimates are! To solve this problem, we can either

- give the ***standard error*** of the estimator, or
- construct an ***interval estimate*** or ***confidence interval***.

3.2 The Central Limit Theorem (CLT)

The Central Limit Theorem (CLT) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (IID) random variables with common mean μ and common variance σ^2 (both assumed to exist). Then

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n} \overset{\text{approx.}}{\sim} N\left(\mu, \sigma^2/n\right),$$

when the sample size n is sufficiently large for ***whatever*** the distribution of X.

The main value of the result is as an approximation for finite n. The rate of convergence depends mainly on the symmetry or asymmetry of the distribution of X.

3.3 Confidence interval for the population mean μ

Definition: A ***confidence interval*** is a range of plausible values for a parameter. There is an associated ***confidence level*** which indicates how likely it is that the interval will include the true value of the parameter *in repeated sampling.*

For example, if we can find μ_1 and μ_2 such that

$$P(\mu_1 < \mu < \mu_2) = 95\%,$$

then (μ_1, μ_2) is the *confidence interval* of μ with 95% *confidence level.*

Case 1: σ known

Suppose $X_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$, where the population standard deviation σ is *known.* Then we have (from the Central Limit Theorem, $\bar{X}$ is approximately normal when n is sufficiently large even if X is not normally distributed)

$$\bar{X} \quad \sim \quad N\left(\mu, \sigma^2/n\right)$$

and hence

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

Using statistical tables, we can find that

$$\mathsf{Pr}\left(-1.96 < Z < 1.96\right) \quad = \quad 0.95$$

Hence, we can be 95% sure that

$$-1.96 < \quad \tfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad < 1.96$$

leading to

$$-1.96 \times \sigma/\sqrt{n} < \quad \bar{X} - \mu \quad < 1.96 \times \sigma/\sqrt{n}$$

i.e.

$$\bar{X} - 1.96 \times \sigma/\sqrt{n} < \quad \mu \quad < \bar{X} + 1.96 \times \sigma/\sqrt{n}.$$

This gives a ***95% confidence interval for the population mean μ***. The value 1.96 can be replaced by other values leading to other levels of confidence. Again, from statistical tables, we get:

Confidence coefficient	90%	95%	99%	99.9%
z_α	***1.645***	***1.960***	***2.576***	***3.291***

Remarks.

- As the level of confidence increases, the ***critical value*** z_α increases;
- This will in turn lead to a wider confidence interval;
- If the aim is to "capture" the population mean μ with our confidence interval, why do we not construct a 100% confidence interval and so be *certain* of capturing μ?

Case 2: σ unknown

In practice, σ is rarely known. In this case, we replace σ by its estimator, s, and so Z must be replaced by

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}.$$

Remarks.

- Z and T are identical but for σ being replaced with s;
- Z follows a ***standard normal*** distribution;
- T does not follow a standard normal distribution, but ***Student's t distribution***

Some notes on *Student's t distribution*:

- It's similar in shape to the normal distribution;
- It has a larger ***spread*** (or "heavier tails");

- The exact shape depends on the parameter ν (degrees of freedom), which itself depends on the sample size n ($\nu = n - 1$);
- As $n \to \infty$, $T \to Z$, i.e. the standard normal distribution;
- The heavier tails account for ***uncertainty*** in σ.

From this result, we get a corresponding formula for a confidence interval for μ when the population standard deviation is *unknown*:

$$\bar{X} \quad \pm \quad t_{n-1,\alpha} \times s/\sqrt{n},$$

where the value $t_{n-1,\alpha}$ depends on the level of confidence α and the sample size, and can be obtained from tables.

Examples (will be discussed in lecture)

1. In an air pollution study, the following amounts of suspended benzene soluble organic matter (μg/m^3) were obtained for 7 random m^3 of air:

2.2 1.8 3.1 2.0 2.4 2.0 1.2

(a) Construct a 95% confidence interval for the population mean.

(b) Construct a 99% confidence interval for the population mean.

2. A company packs sacks of flour. The variance of the filling process is known to be 100g. A sample of 50 bags is taken and weighed and the resulting sample mean is 750g. Compute a 90% and a 95% confidence interval for the mean weight of a bag of flour.

3.4 Hypothesis tests

Confidence intervals can be used to make inferences about population parameters. Sometimes, you may be asked to assess whether or not a parameter takes a specific value. For example, whether the population mean $\mu = 5$.

One way of re–expressing this question is to ask whether the parameter value is plausible in light of the data. A simple check to see whether the value is contained in a 95% confidence interval will provide an answer. An alternative method, called a ***hypothesis test***, is also available.

Illustrative example. The average score by 11 year old in a standard reading test is 5.7. Suppose that a group of 10 such children are given special coaching. They obtain an average mark of 6.2 with standard deviation 1.1. Does this show that the coaching has a real effect?

Remark: Without coaching, we would expect $\mu = 5.7$. We need to know whether the population mean after coaching is still 5.7. In other words, is the average mark of 6.2 a *real* effect, or could it be due to chance?

Idea of a hypothesis test: We first assume

$$H_0 : \mu = 5.7 \text{ and } H_1 : \mu > 5.7,$$

where H_0 is called the ***null hypothesis*** and H_1 is called the ***alternative hypothesis***. The idea of a hypothesis test is to make a decision on whether we should accept H_0 or reject H_0. Intuitively, a large value of $\bar{X}$ would be in favour of H_1. So, we can adopt the following *decision rule*:

$$\text{if } \bar{X} > a, \text{ we reject } H_0;$$

where $\bar{X} > a$ is called a *rejection region.* Now, the problem is how to find the value of a? We usually find such a value to control the so called *type I error*:

$$P(\bar{X} > a | \text{if } H_0 \text{ is true}) = \alpha,$$

where α is called *significant level*, taking a small value such 0.1 or 0.05.

Using the fact that

$$\frac{\bar{X} - 5.7}{1.1/\sqrt{10}} \sim t_9 \text{ if } H_0 \text{ is true},$$

and $P(t_9 > 1.383) = 0.1$, we have

$$0.1 = P(\frac{\bar{X} - 5.7}{1.1/\sqrt{10}} > 1.383 | H_0) = P(\bar{X} > 6.18 | H_0),$$

thus the rejection region with 10% level is $\bar{X} > 6.18$.

We have observed the average of 6.2, which is in the rejection region. We can therefore conclude that we should reject H_0 and accept H_1 with 10% level; i.e., the coaching seems have a real effect.

Alternatively, we can calculate the *P-value* – the probability of observing a test statistic as extreme as that obtained if the null hypothesis is true.

$$\begin{aligned} P - value &= \mathsf{Pr}\left(\bar{X} \geq 6.2\right) \\ &= \mathsf{Pr}\left(T_9 \geq \frac{6.2 - 5.7}{1.1/\sqrt{10}}\right) \\ &= \mathsf{Pr}\left(T_9 \geq 1.437\right) = 0.0923. \end{aligned}$$

Note that

$$\{\text{P-value} < \alpha\} \Longleftrightarrow \{\bar{X} \text{ is in the rejection region}\} \Longleftrightarrow \{\text{Reject } H_0\}.$$

For the above illustrative example, $P - value = 0.0923 < 0.1$, so we should reject H_0.

In practice, we usually use the following decision rule:

***p*–value**	**Interpretation**
$p > 0.1$	no evidence against the null hypothesis
p lies between 0.05 and 0.1	***slight*** evidence against H_0
p lies between 0.01 and 0.05	***moderate*** evidence against H_0
p is smaller than 0.01	***strong*** evidence against H_0

The following table gives a general framework in hypothesis testing.

General	**Coaching example**
1. State a ***null hypothesis***, H_0	$H_0 : \mu = 5.7$
2. Decide on a ***test***	Use a T-test, i.e. $\frac{\bar{X}-5.7}{s/\sqrt{n}} \sim t_9$ distribution
3. Calculate a ***test statistic***	$t_9 = 1.437$
4. Find the ***p*–*value***	0.0923
5. Form your ***Conclusions***	Slight evidence against H_0

Conclusion for Coaching example: we have only ***slight*** evidence against $H_0 : \mu = 5.7$, or we can conclude that the coaching has a slight effect.

Errors

Our final decision is subject to two types of error.

A ***Type I error*** occurs when we reject the null hypothesis when really it is true.

A ***type II error*** occurs when we fail to reject the null hypothesis when in fact it is false.

Decision	$\rightarrow$	Do not reject	Reject
Hypothesis is	True	Correct!	**Type I error**
(fact)	False	**Type II error**	Correct!

One-sample test

We now go back to assuming we have a sample from a Normal distribution.

One-sample Z–test. This is a test of $H_0 : \mu = \mu_0$, with σ assumed known. The test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

and we find our p-value from $N(0, 1)$ tables.

The calculation of p-value depends on the alternative hypothesis. Let $\bar{x}$ be the value of $\bar{X}$ for a set of sample, and let $b = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$, then for a two-tailed test (two-sided test) $H_1 : \mu \neq \mu_0$,

$$p - value = P(|Z| > |b|) = 2P(Z > |b|), \quad Z \sim N(0, 1).$$

For one-tailed (one-sided) test,

$$p - value = P(Z > b), \quad \text{if } H_1 : \mu > \mu_0;$$

and

$$p - value = P(Z < b), \quad \text{if } H_1 : \mu > \mu_0.$$

The test can be performed in `Minitab` using `Stat` – `Basic Statistics` – `1-Sample Z`.

One–sample t–test. This is the same as for the Z–test except that σ is unknown and must be estimated by s. Thus, our test statistic becomes

$$t_{n-1} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}};$$

we can find our p–value from tables of probabilities for the t distribution on $\nu = n - 1$ degrees of freedom.

The test can be performed in `Minitab` using `Stat` – `Basic Statistics` – `1-Sample t`.

One–tailed versus two–tailed tests. The standard approach is to use a two–tailed alternative hypothesis (i.e. $H_1 : \mu \neq \mu_0$) and hence a two–tailed test, unless there are compelling arguments for a one–tailed alternative (i.e. either $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$):

- It is inherent in the context that departures from H_0 can only conceivably be in one direction.
- The test is constructed to be one–sided, e.g. a goodness–of–fit test, where only high values of the statistic indicate poor fit.

Examples (will be discussed in lecture)

1. In an air pollution study, the following amounts of suspended benzene soluble organic matter (μg/m^3) were obtained for 7 random m^3 of air:

$$2.2 \quad 1.8 \quad 3.1 \quad 2.0 \quad 2.4 \quad 2.0 \quad 1.2$$

(a) Construct a 95% confidence interval for the population mean.

(b) Construct a 99% confidence interval for the population mean.

(c) Test the null hypothesis that the population mean is 2.8 μg/m^3.

2. A machine for filling cans of Coke has a process variance of 400ml. A sample of 100 cans is taken and it is found that the average contents are 240ml. Is this consistent with the cans containing the stated weight of 250ml?

3.5 Two–sample problems

Here, we wish to compare two populations. Both will be assumed Normal. We may wish to test for a common ***mean***.

A typical context is where we wish to compare a ***treatment*** with a ***control*** under similar conditions.

3.5.1 Comparison of sample means

Here, we wish to test ***independent*** samples from two populations (i.e. there is no natural pairing between the two).

We assume both populations are Normally distributed (and we can check this assumption by looking at plots of the sample data).

As with one–sample hypothesis tests for the population mean, there are two cases to consider:

- Both population standard deviations are ***known*** (rare), and
- Both population standard deviations are ***unknown***.

Two–sample Z–test. Let $\bar{X}_i$ be the sample mean from a sample with sample size n_i and population $N(\mu_i, \sigma_i^2)$ for $i = 1, 2$. Those two samples are assumed to be independent. We test $H_0 : \mu_1 = \mu_2$ when both population standard deviations (σ_1 and σ_2) – or indeed the variances – are ***known***.

The test statistic is

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

which follows a standard Normal distribution.

The `Minitab` commands are `Stat` – `Basic Statistics` – `2-Sample Z`.

Two–sample t–test. This is equivalent to the previous test, but is used when both population standard deviations are ***unknown***.

The test statistic is

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

which follows a t distribution on $\nu = n_1 + n_2 - 2$ degrees of freedom, and

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

is a "pooled standard" deviation.

The `Minitab` commands are `Stat`– `Basic Statistics`–`2-Sample T`.

The usual assumption of Normality applies, but in this test we also assume the population standard deviations are equal.

The validity of a two–sample t test depends in three assumptions:

(i) Normality,

(ii) independence, and

(iii) common variance (homogeneity)

It is, however, a **robust** test and will work well if these are only approximately valid.

As a guide for (iii), neither standard deviation should exceed twice the other. What about a two–sample test when σ_1 and σ_2 are unknown but *cannot be assumed equal*? When this cannot be assumed, there are other tests available for this two–sample test, but these are beyond the scope of this course.

Examples (will be discussed in lecture)

1. Fifteen containers of water were taken from each of two different stations A and B on a river. Determinations of the lead content of each sample were made and the results (ppm) are given below:

A	9.6	10.7	10.6	10.0	11.1	10.7	10.3	10.7
	12.0	11.3	11.6	10.5	10.8	11.0	11.1	
B	9.7	11.8	11.9	10.5	11.7	10.5	11.5	11.4
	12.4	12.1	9.8	11.7	11.2	11.2	11.1	

   Suppose the samples from stations A and B are independent. Perform an appropriate hypothesis test to see if there is a significant difference between the lead content of samples taken at the two stations. State any assumptions implicit in your test.

2. Before a training session for call centre employees, a sample of 50 calls to the centre had an average duration of 5 minutes, whereas after the training session a sample of 45 calls had an average duration of 4.5 minutes. The population variance is known to have been 1.5 minutes before the course and 2 minutes afterwards. Has the training course affected the average call duration?