Statistics

MATH2089

Semester 1, 2018 – Lecture 9

This lecture

7. Inferences concerning a mean

Additional reading:
Sections 8.1, 8.2 (pp. 355-359) and 8.5 (pp. 389-391;393) (2nd edition)
Sections 8.1, 8.2 (pp. 363-367) and 8.5 (pp. 400-403;404-405) (3rd edition)

Hypotheses testing: Introduction

In the previous lectures we showed how a parameter of a population can be estimated from sample data, using either a point estimate or an interval of "plausible" values called a confidence interval.

However, there are many situations in which we must decide whether we believe a statement concerning a parameter is true or false, that is, we must **test a hypothesis about a parameter**.

For instance, suppose that a customer protection agency wants to test a paint manufacturer's claim that the average drying time of his new 'fast-drying' paint is 20 minutes.

It instructs a member of its staff to paint each of 36 boards using a different can of the paint: the observed average drying time for this sample is 20.75 minutes

$\rightarrow$ does that really contradict the manufacturer's claim ?

This type of question can be answered using a statistical inference technique called **hypothesis testing**.

Statistical hypotheses

- Many problems in engineering require that we decide which of two competing statements about some parameters are true

→ The statements are called hypotheses

- In the previous example, we might express the hypothesis as

$$H_0 : \mu = 20$$

  where μ is the 'true' mean drying time for this type of paint
- This statement is called the **null hypothesis**, and is usually denoted H_0
- H_0 is the default hypothesis we will assume is true <u>unless we have enough evidence</u> to compel us to change our minds
- If so, we will favour the **alternative hypothesis**, which is usually denoted H_a (or sometimes H_1 in some references)
- The choice of H_a depends mostly on the problem and what we hope to show
- In our example, it could be either $H_a : \mu \neq 20$ or $H_a : \mu > 20$

Null hypothesis

The value μ_0 of the population parameter specified in the null hypothesis

$$H_0 : \mu = \mu_0$$

is usually determined in one of three ways:

- it may result from past experience or knowledge of the process, or from previous tests or experiments

$\rightarrow$ determine whether the parameter value has changed

- it may be determined from some theory or some model

$\rightarrow$ check whether the theory or the model is valid

- it may result from external considerations, such as engineering specifications, or from contractual obligations

$\rightarrow$ conformance testing

Note: in some instances, a null hypothesis of the form $H_0 : \mu \geq \mu_0$ or $H_0 : \mu \leq \mu_0$ may seem appropriate. However, the test procedure for such an H_0 is the same as $H_0 : \mu = \mu_0$

$\rightarrow$ we always state a null hypothesis as an equality

Alternative hypothesis

The alternative hypothesis can essentially be of two types.

A **two-sided alternative** is when H_a is of the form

$$H_a : \mu \neq \mu_0$$

This is the exact negation of the null hypothesis $H_0 : \mu = \mu_0$
$\rightarrow \mu_0$ is the only value of some interest in the problem

However, in many situations, we may wish to favour a given direction for the alternative:

$$H_a : \mu < \mu_0 \qquad \text{or} \qquad H_a : \mu > \mu_0$$

These are called **one-sided alternatives**.

Continuing with our example, the customer protection agency may only wish to highlight that the average drying time of the paint is actually longer than the advertised 20 minutes (no criticisms if this time is even shorter).

The considered alternative might change the conclusion of a hypothesis test, and should be carefully formulated!

Hypothesis testing

A procedure leading to a decision about a particular hypothesis H_0 is called a test of hypothesis.

Such procedures rely on using the information contained in a random sample from the population of interest

$\rightarrow$ if this information is consistent with the hypothesis H_0, we will not reject H_0; however, if this information is inconsistent with H_0, we will reject it

Fact

The truth or the falsity of a particular hypothesis can never be known with certainty, unless we can examine the entire population.

The decision we make depends on a random sample, so is a kind of 'random object'

$\rightarrow$ a hypothesis test should be developed with the probability of reaching a wrong conclusion in mind

Hypothesis testing

To illustrate the general concepts, consider again the mean drying time problem. Imagine that we wish to test

$$H_0 : \mu = 20 \qquad \text{against} \qquad H_a : \mu \neq 20$$

We have a sample of $n = 36$ specimens and the sample mean $\bar{x}$ is observed

As the sample mean is a 'good' estimate of μ, we expect $\bar{x}$ to be reasonably close to μ

$\rightarrow$ if $\bar{x}$ falls 'close' to 20 min, no clear contradiction with H_0, we do not reject it

$\rightarrow$ if $\bar{x}$ is considerably 'distant' from 20 min, evidence in support of H_a, we reject H_0

The numerical value which is computed from the sample and used to decide between H_0 and H_a, here the (possibly standardised) sample mean, is called the

test statistic

Hypothesis testing

Suppose that we decide to reject H_0 if $\bar{x}$ is smaller than 19.33 or larger than 20.67 (arbitrary criterion for illustrative purposes only)

$\rightarrow$ if $\bar{x} \in [19.33, 20.67]$, we do not reject $H_0 : \mu = 20$

The values for which we reject H_0, that is, values less than 19.33 and greater than 20.67, are called the rejection regions for the test, the limiting values (here 19.33 and 20.67) being the **critical values**.

This provides a clear-cut criterion for the decision; however, it is not infallible:

a) even if the true mean $\mu = 20$, there is a possibility that the sample mean $\bar{x}$ may be outside $[19.33, 20.67]$, due to bad luck
b) even if the true mean $\mu \neq 20$, say $\mu = 21$, there is a possibility that the sample mean may be in $[19.33, 20.67]$

Errors

So there are essentially two possible wrong conclusions:

a) rejecting H_0 when it is true: this is defined as a **type I error**

b) failing to reject H_0 when it is false: this is defined as **type II error**

Because the decision is based on a random sample, **probabilities** can be associated with these errors.

The probability of type I error is usually denoted α

$$\mathbb{P}(\text{type I error}) = \mathbb{P}(\text{reject } H_0 | \, H_0 \text{ is true}) = \alpha$$

The probability of type II error is usually denoted β

$$\mathbb{P}(\text{type II error}) = \mathbb{P}(\text{fail to reject } H_0 | H_0 \text{ is false}) = \beta$$

$1 - \beta = \mathbb{P}(\text{reject } H_0 \text{ when it is false})$ is also called the **power** of the test

Note that β actually depends on the true (unknown) value of μ.

Errors

		In Reality	
		H_0 True	H_0 False
Decision	Reject H_0	Type I Error	Correct Decision
	Fail to Reject H_0	Correct Decision	Type II Error

Quantifying Errors

Assume in our running example that it is known from past experience that the drying time is normally distributed with known standard deviation $\sigma = 2$ min.

Then we know that $Z = \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \sim \mathcal{N}(0,1)$, so:

$$\begin{aligned}
\mathbb{P}(\text{type I error}) &= \mathbb{P}((\bar{X} < 19.33) \cup (\bar{X} > 20.67) \text{ when } \mu = 20) \\
&= \mathbb{P}\left(Z < \sqrt{36}\,\frac{19.33-20}{2}\right) + \mathbb{P}\left(Z > \sqrt{36}\,\frac{20.67-20}{2}\right) \\
&= \mathbb{P}(Z < -2.01) + \mathbb{P}(Z > 2.01) = 0.044 = \alpha \qquad \text{(Matlab)}
\end{aligned}$$

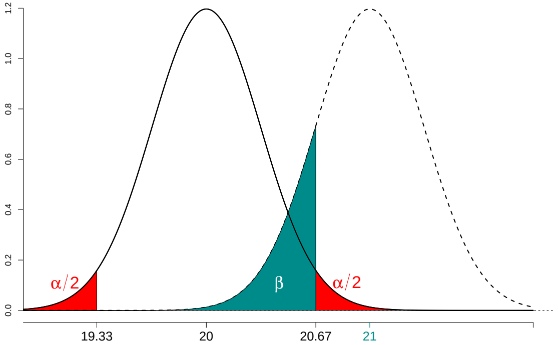
If H_0 is true, we have a 4.4% chance of rejecting it (with our rule)

Suppose now that $\mu = 21$ (so H_0 is not true!). We have:

$$\begin{aligned}
\mathbb{P}(\text{type II error}) &= \mathbb{P}(\bar{X} \in [19.33, 20.67] \text{ when } \mu = 21) \\
&= \mathbb{P}\left(\sqrt{36}\,\frac{19.33-21}{2} \le Z \le \sqrt{36}\,\frac{20.67-21}{2}\right) \\
&= \mathbb{P}(-5.01 \le Z \le -0.99) = 0.16 = \beta \qquad \text{(Matlab)}
\end{aligned}$$

Errors

With the decision rule: reject H_0 if $\bar{x} \notin [19.33, 20.67]$



$\rightarrow \alpha = 0.044,$
$\beta = 0.16$ if $\mu = 21$

See that β would rapidly increase as μ approached the hypothesised value μ_0

Errors

Suppose that you want to reduce the type I error probability α

$\rightarrow$ widen the acceptance region, for instance say

$$\text{reject } H_0 \text{ if } \bar{x} \notin [19.2, 20.8]$$

(again for illustrative purpose only). Then (as on Slide 12),

$$\alpha = \ldots = 0.016 \quad (< 0.044)$$

but

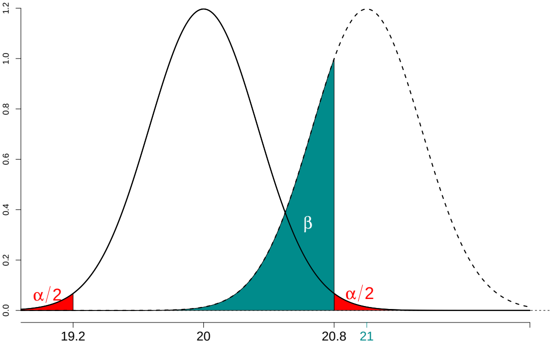
$$\beta = \ldots = 0.27 \quad (> 0.16) \qquad \text{when } \mu = 21$$

$\rightarrow$ if α decreases, β must increase and vice-versa !

$\rightarrow$ impossible to make both types of error as small as possible simultaneously

Errors

With the decision rule: reject H_0 if $\bar{x} \notin [19.2, 20.8]$



$\rightarrow \alpha = 0.016,$
$\beta = 0.27$ if $\mu = 21$

Errors

Usually, one decides to set α to a small predetermined level (and accept the resulting value of β).

This is because hypothesis testing was actually originally inspired by jury trials.

In a trial, defendants are initially **assumed innocent** (H_0). Then,

- if strong evidence is found to the contrary, then they are declared to be guilty (reject H_0)
- if there is insufficient evidence, they are declared not guilty (fail to reject H_0) $\rightarrow$ not the same as proving the defendant is innocent!

If the jury is wrong, either an innocent person is convicted (type I error) or a culprit is let free (type II error)

$\rightarrow$ The prevailing thought is that convicting an innocent person is more a serious problem than the contrary

Errors: analogy to criminal trials

		In Reality	
		Innocent	Guilty
Decision	Convict	Type I Error	Correct Decision
	Acquit	Correct Decision	Type II Error

Jury must be "convinced beyond a reasonable doubt" to convict

$\rightarrow$ usually, the type I error probability α is set to 0.10, 0.05 or 0.01, and the decision rule is fixed accordingly

In hypothesis testing, the value of α is called the **significance level** of the test.

Significance level and decision rule

Assume for the moment that **the population is normal with known standard deviation σ** $\quad \to \bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

<u>At significance level α</u>, we are after two constants ℓ and u such that

$$\alpha = \mathbb{P}(\bar{X} \notin [\ell, u] \text{ when } \mu = \mu_0) = \mathbb{P}\left(Z \notin \left[\sqrt{n}\,\frac{\ell - \mu_0}{\sigma}, \sqrt{n}\,\frac{u - \mu_0}{\sigma}\right]\right)$$

$$\to \sqrt{n}\,\frac{\ell - \mu_0}{\sigma} = z_{\alpha/2} = -z_{1-\alpha/2} \qquad \text{and} \qquad \sqrt{n}\,\frac{u - \mu_0}{\sigma} = z_{1-\alpha/2}$$

$$\to \boxed{\ell = \mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}} \qquad \text{and} \qquad \boxed{u = \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

$\to$ the decision rule is:

$$\text{reject } H_0 \text{ if } \bar{x} \notin \left[\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Significance level and decision rule: example

In our example (with $\sigma = 2$), suppose we want to test $H_0 : \mu = 20$ against $H_a : \mu \neq 20$ at the 5% significance level ($\alpha = 0.05$).

Then, $\ell = 20 - 1.96 \times \frac{2}{\sqrt{36}} = 19.35$ and $u = 20 + 1.96 \times \frac{2}{\sqrt{36}} = 20.65$

$\rightarrow$ At significance level 5%, the decision rule is

reject H_0 if $\bar{x} \notin [19.35, 20.65]$

Now, over our 36 samples, we have observed an average drying time of $\bar{x}$ =20.75 min $\rightarrow$ we reject H_0

$\rightarrow$ we contradict the manufacturer

Now, at significance level 1% ($\alpha = 0.01$),

$$\ell = 20 - 2.575 \times \frac{2}{\sqrt{36}} = 19.14 \text{ and } u = 20 + 2.575 \times \frac{2}{\sqrt{36}} = 20.86$$

and the decision rule becomes reject H_0 if $\bar{x} \notin [19.14, 20.86]$

$\rightarrow$ now we cannot reject H_0 from the observed sample with $\bar{x} = 20.75$

$\rightarrow$ you do not dare contradict the manufacturer

Reject / no reject of H_0: remark

The previous situation makes it clear why we do not say “we accept H_0”:

we reject H_0 at 5% level but we don't at 1% level. However H_0 is either true or not, regardless of the situation!

The evidence shown by the sample indicates that we would be wrong with a chance between 1% and 5% if we rejected H_0

Testing at the 5% level essentially means that we tolerate being wrong in at most 5% of the cases $\rightarrow$ we may reject H_0

Testing at the 1% level means that we tolerate being wrong in at most 1% of the cases

$\rightarrow$ it is then too risky to reject H_0 if we need to be 99% confident in our decision

$\rightarrow$ we do not reject H_0, the sample has not shown *enough* evidence against it. We don't know whether H_0 is true but we cannot exclude the possibility it is (which doesn't mean it is for sure!)

p-value

In the previous situation, we said that "we would be wrong with a chance between 1% and 5% if we rejected H_0"

$\rightarrow$ it would be interesting to know more about how confident we can be in our decision

That is, what is the ***p*-value**.

Definition

The ***p*-value** is the smallest level of significance that would lead to rejection of H_0 with the observed sample

Concretely, the p-value is the probability that the test statistic will take on a value that is at least as extreme as the observed value when H_0 is true ('extreme' to be understood in the direction of the alternative).

It might be *roughly* interpreted as the chance of being wrong if we reject H_0

p-value

When testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$, the p-value will be the probability of finding the random variable $\bar{X}$ more different to μ_0 than the observed $\bar{x}$, that is,

$$\begin{aligned} p &= \mathbb{P}\left(\bar{X} \notin [\mu_0 \pm |\bar{x} - \mu_0|] \text{ when } \mu = \mu_0\right) \\ &= 1 - \mathbb{P}\left(\bar{X} \in [\mu_0 \pm |\bar{x} - \mu_0|] \text{ when } \mu = \mu_0\right) \end{aligned}$$

Define z_0 as the z-score of $\bar{x}$ if $\mu = \mu_0$, i.e.

$$z_0 \doteq \sqrt{n}\,\frac{\bar{x} - \mu_0}{\sigma} \qquad \rightarrow \text{"observed value of the test statistic"}$$

As we know that $Z = \sqrt{n}\,\frac{\bar{X}-\mu_0}{\sigma} \sim \mathcal{N}(0,1)$, we have

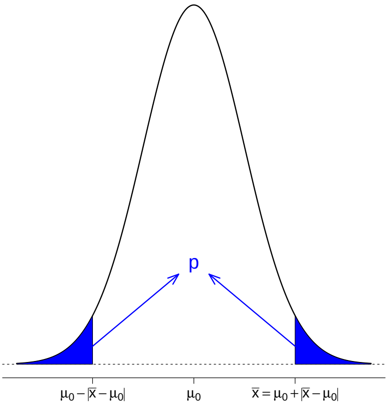
$$\begin{aligned} p &= 1 - \mathbb{P}\left(\sqrt{n}\,\frac{\bar{X} - \mu_0}{\sigma} \in \left[\sqrt{n}\,\frac{\mu_0 \pm |\bar{x} - \mu_0| - \mu_0}{\sigma}\right]\right) \\ &= 1 - \mathbb{P}(Z \in [-|z_0|, |z_0|]) = 2 \times (1 - \Phi(|z_0|)) \end{aligned}$$

(by symmetry of the $\mathcal{N}(0,1)$ distribution)

p-value

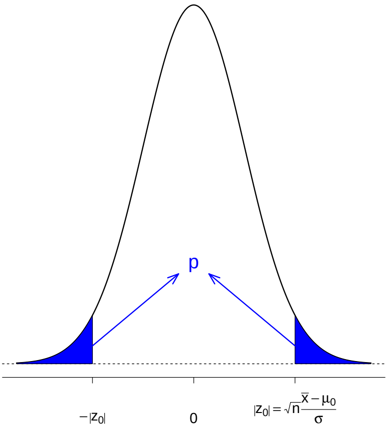
Distribution of $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

if $H_0 : \mu = \mu_0$ is true



Standardised distribution of

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



$$\rightarrow p = 2 \times (1 - \Phi(|z_0|))$$

p-value: example

In our example, we have observed a sample mean of $\bar{x} = 20.75$ min, so that the p-value is given by

$$p = 1 - \mathbb{P}(\bar{X} \in [19.25, 20.75] \text{ when } \mu = 20)$$

We have here that

$$z_0 = \sqrt{36}\,\frac{20.75 - 20}{2} = 2.25,$$

so that the p-value can easily be computed:

$$p = 1 - \mathbb{P}(-2.25 \leq Z \leq 2.25) = 2 \times (1 - \Phi(2.25)) \overset{\text{Matlab}}{=} 0.024$$

If $H_0 = 20$ is true, the probability of obtaining another random sample whose mean is at least as far from 20 as our 20.75 is 0.024.

$\rightarrow$ if we rejected H_0, we would be wrong with a 2.4% chance

p-value and significance level

This means that $H_0 : \mu = 20$ would be rejected in favour of $H_a : \mu \neq 20$ at any level of significance greater than or equal to 0.024.

Operationally, once a p-value is computed, we typically compare it to a predefined significance level α to make a decision:

$$\text{if } \boxed{p < \alpha}, \text{ reject } H_0, \qquad \text{if } \boxed{p \geq \alpha}, \text{ do not reject } H_0$$

In presenting results and conclusions, it is standard practice to report the observed p-value along with the decision that is made regarding H_0.

This gives potential other decision makers the possibility to draw a conclusion at any specified level, not only the one you impose to them.

Here, the conclusion would be:

$$p = 0.024 < \alpha = 0.05 \quad \rightarrow \text{reject } H_0 \text{ (at significance level 5\%)}$$

or

$$p = 0.024 > \alpha = 0.01 \quad \rightarrow \text{not reject } H_0 \text{ (at significance level 1\%)}$$

Procedures in hypothesis testing

1. State the null and alternative hypotheses: H_0 and H_a
2. Determine the rejection criterion
3. Compute the appropriate test statistic and determine its distribution
4. Calculate the p-value using the test statistics computed
5. Conclusion: reject/do not reject H_0, relate back to the research question

One-sided alternatives

The whole development, yet very similar, must be slightly adapted when **one-sided alternatives** are concerned.

First, *it might occasionally be difficult to choose the appropriate formulation of the alternative*.

In our running example, suppose now that we would like to highlight that the average drying time is actually longer than the advertised 20 min. Would we test for

- $H_0 : \mu = 20$ against $H_a : \mu > 20$, hoping to reject H_0, or
- $H_0 : \mu = 20$ against $H_a : \mu < 20$, hoping not to reject H_0 ?

Recall that rejecting H_0 is a strong conclusion (we have enough evidence to do it),
unlike not rejecting (we do not have enough evidence to conclude, the decision is dictated by risk aversion, not by facts $\rightarrow$ weak conclusion)

$\rightarrow$ always put what we want to prove in the alternative hypothesis

Here, we should test $H_0 : \mu = 20$ against $H_a : \mu > 20$

One-sided alternatives

In a two-sided test, i.e. with alternative $H_a : \mu \neq \mu_0$, an observed value $\bar{x}$ of $\bar{X}$ much smaller than μ_0 **or** much larger than μ_0 is evidence in direction of H_a.

However, if the alternative is $H_a : \mu > \mu_0$, a small value of $\bar{x}$ is not evidence against $H_0 : \mu = \mu_0$ in favour of H_a (H_0 is more likely than H_a if $\bar{X}$ takes a small value even very different to μ_0!)

$\rightarrow$ we must only seek evidence against H_0 in the direction of H_a!

1. Thus, in testing
$$H_0 : \mu = \mu_0 \qquad \text{against } H_a : \mu > \mu_0$$
we should only reject H_0 if $\bar{X}$ is much greater than μ_0
2. Similarly, in testing
$$H_0 : \mu = \mu_0 \qquad \text{against } H_a : \mu < \mu_0$$
we should only reject H_0 if $\bar{X}$ is much smaller than μ_0

The critical region is again determined by the significance level α.

One-sided alternatives

1. With $H_a : \mu > \mu_0$, we are after a constant u such that

$$\mathbb{P}(\bar{X} > u \text{ when } \mu = \mu_0) = \alpha$$

As we know that $Z = \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim \mathcal{N}(0,1)$,

$$\boxed{u = \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}}$$

$\rightarrow$ the decision rule is reject H_0 if $\bar{x} > \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$

Again with the 'observed value of the test statistic' $z_0 = \sqrt{n}\frac{\bar{x}-\mu_0}{\sigma}$, the p-value is given by

$$p = \mathbb{P}(\bar{X} > \bar{x} \text{ when } \mu = \mu_0) = \mathbb{P}\left(Z > \sqrt{n}\,\frac{\bar{x}-\mu_0}{\sigma}\right) = 1 - \Phi(z_0)$$

In our example, with $\bar{x} = 20.75$ and $H_a : \mu > 20$, we have, at significance level 5%, $u = 20 + 1.645 \times \frac{2}{\sqrt{36}} = 20.548$

$\rightarrow$ we reject H_0. The p-value is

$$p = \mathbb{P}\left(Z > \sqrt{36}\,\frac{20.75-20}{2}\right) = \mathbb{P}(Z > 2.25) \overset{\text{Matlab}}{=} 0.012 \quad (< 0.05)$$

One-sided alternatives

2. With $H_a : \mu < \mu_0$, we are after a constant l such that

$$\mathbb{P}(\bar{X} < \ell \text{ when } \mu = \mu_0) = \alpha$$

As we know that $Z = \sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \sim \mathcal{N}(0,1)$, $\boxed{\ell = \mu_0 - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}}$

$\rightarrow$ the decision rule is reject H_0 if $\bar{x} < \mu_0 - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$

The p-value is given by

$$p = \mathbb{P}(\bar{X} < \bar{x} \text{ when } \mu = \mu_0) = \mathbb{P}\left(Z < \sqrt{n}\,\frac{\bar{x}-\mu_0}{\sigma}\right) = \Phi(z_0)$$

In our example, with $\bar{x} = 20.75$ and $H_a : \mu < 20$, we have, at significance level 5%, $\ell = 20 - 1.645 \times \frac{2}{\sqrt{36}} = 19.452$

$\rightarrow$ we do not reject H_0. The p-value is

$$p = \mathbb{P}\left(Z < \sqrt{36}\,\frac{20.75-20}{2}\right) = \mathbb{P}(Z < 2.25) \overset{\text{Matlab}}{=} 0.988 \quad (\gg 0.05)$$

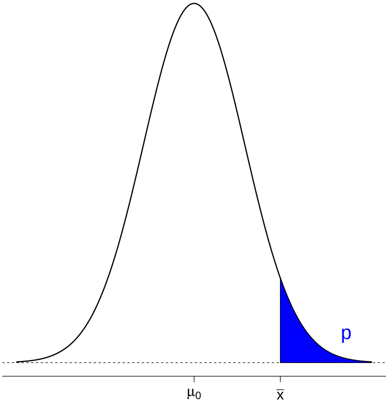
One-sided p-values

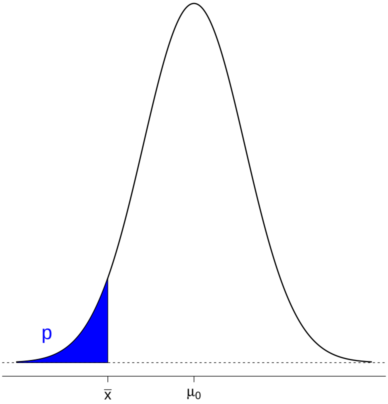
Alternative $H_a : \mu > \mu_0$

("Upper-tailed test")

Alternative $H_a : \mu < \mu_0$

("Lower-tailed test")





One-sided alternatives: remark

Remark 1: as announced earlier, the selected alternative does affect the reject/fail to reject of the same null hypothesis at the same significance level! It is therefore important to choose the alternative in a meaningful way

Remark 2: in most situations, the only meaningful way of writing H_a is in the direction which has been observed in the sample

For instance, it is obvious that an observed sample mean $\bar{x}$ greater than μ_0 can never bring evidence in favour of $H_a : \mu < \mu_0$!

In other words, we will never reject a null hypothesis which is not 'challenged' by some evidence in favour of the alternative

$\rightarrow$ if we want the test to be relevant, the alternative must be somewhat supported by the observed evidence

The question is just to check whether that evidence is sufficiently against H_0 to reject it or not.

Normal populations with unknown standard deviation

So far, we have assumed that we had a normal population with known standard deviation σ, which allowed us to directly use the procedure

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

Suppose now that σ is unknown (still in a normal population)

Like we did when deriving confidence intervals, we can replace the unknown σ by its estimator

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

and work with

$$T = \sqrt{n}\,\frac{\bar{X} - \mu}{S}$$

However, we know that the randomness of S affects the distribution of T, which is no longer normal but

$$\boxed{T \sim t_{n-1}}$$

Normal populations with unknown standard deviation

That apart, everything happens as with a known standard deviation.

Specifically, for the two-sided test $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$, the decision rule is

$$\text{reject } H_0 \text{ if } \bar{x} \notin \left[\mu_0 - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}, \mu_0 + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}\right],$$

with the observed sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

and from the observed value of the test statistic

$$t_0 = \sqrt{n}\,\frac{\bar{x} - \mu_0}{s}$$

we can compute the p-value

$$p = 1 - \mathbb{P}(T \in [-|t_0|, |t_0|]) = 2 \times \mathbb{P}(T > |t_0|) \quad \text{where } T \sim t_{n-1}$$

Normal populations with unknown standard deviation

For the one-sided alternatives, the rejection criteria are

$$\text{reject } H_0 \text{ if } \bar{x} > \mu_0 + t_{n-1,1-\alpha}\frac{s}{\sqrt{n}} \quad \text{or} \quad \text{reject } H_0 \text{ if } \bar{x} < \mu_0 - t_{n-1,1-\alpha}\frac{s}{\sqrt{n}}$$

and the associated p-values are

$$p = \mathbb{P}(T > t_0) \quad \text{or} \quad p = \mathbb{P}(T < t_0)$$

It is no surprise that this test is often called the ***t*-test**, in contrast with the test based on the Normal distribution, called the ***z*-test**.

Note: as for confidence intervals, the t-distribution, with its heavier tails (compared to $\mathcal{N}$), reflects the extra variability introduced in the procedure by the estimation of σ

$\rightarrow$ we must be more careful when making the decision, and we reject H_0 'less easily'

t-test: example

Example

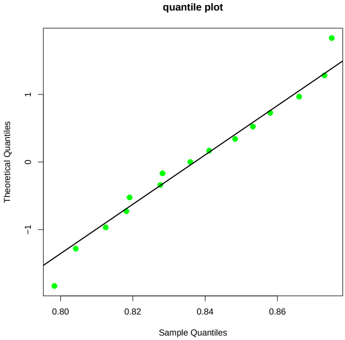
The quality of a golf club is, among other things, measured by its ‘coefficient of restitution’, the ratio of the outgoing velocity of the ball to the incoming velocity of the club. An experiment was performed in which 15 clubs produced by a particular club maker were selected at random and their coefficient of restitution measured:

```
0.8411 0.8191 0.8182 0.8125 0.8750 0.8580 0.8532 0.8483
0.8276 0.7983 0.8042 0.8730 0.8282 0.8359 0.8660
```

The maker claims that the mean coefficient of restitution of its clubs exceeds 0.82. From the observations we have, is there evidence (at level 0.05) to support the maker’s claim? (**Hint:** You can use the following Matlab output: $\texttt{tinv}(0.95, 14) = 1.76$, $\texttt{tcdf}(2.72, 14) = 0.992$)

t-test: example

The quantile plot of the data supports the assumption that the coefficient of restitution is normally distributed



t-test: example

Non-normal populations

Assuming that the population is normal, we use the results

$$Z = \sqrt{n}\,\frac{\bar{X} - \mu}{\sigma} \sim \mathcal{N}(0,1) \quad \text{or} \quad T = \sqrt{n}\,\frac{\bar{X} - \mu}{S} \sim t_{n-1}$$

What if the population is not normal ?

By the Central Limit Theorem (CLT)

$$\sqrt{n}\,\frac{\bar{X} - \mu}{\sigma} \overset{a}{\sim} \mathcal{N}(0,1),$$

we can carry over all our z-test procedures to the arbitrary population case, bearing in mind that the results require n 'large enough' and are only approximately right.

Further, we know that the estimation of σ by S does not dramatically affect that result:

$$\sqrt{n}\,\frac{\bar{X} - \mu}{S} \overset{a}{\sim} \mathcal{N}(0,1),$$

$\rightarrow$ the above observation holds true even if σ needs to be estimated

Non-normal populations: two-sided large sample test

Hence, the test for

$$H_0 : \mu = \mu_0 \qquad \text{against} \qquad H_a : \mu \neq \mu_0$$

(two-sided test) using the decision rule

$$\text{reject } H_0 \text{ if } \bar{x} \notin \left[\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

(if σ is known) or

$$\text{reject } H_0 \text{ if } \bar{x} \notin \left[\mu_0 - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \mu_0 + z_{1-\alpha/2}\frac{s}{\sqrt{n}}\right]$$

(σ unknown), will have an **approximate significance level** α, provided that n is large enough, regardless of the population distribution

As on Slide 22, the associated **approximate p-value** will be given by

$$p = 2 \times (1 - \Phi(|z_0|)),$$

with $z_0 = \sqrt{n}\,\frac{\bar{x}-\mu_0}{\sigma}$ or $z_0 = \sqrt{n}\,\frac{\bar{x}-\mu_0}{s}$, the observed value of the test statistic.

Non-normal populations: one-sided large sample test

For the one-sided test for $H_0 : \mu = \mu_0$ against

$$H_a : \mu > \mu_0 \qquad \text{or} \qquad H_a : \mu < \mu_0,$$

the decision rules

$$\text{reject } H_0 \text{ if } \bar{x} > \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}} \qquad \text{or} \qquad \text{reject } H_0 \text{ if } \bar{x} < \mu_0 - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$$

(σ known) and

$$\text{reject } H_0 \text{ if } \bar{x} > \mu_0 + z_{1-\alpha}\frac{s}{\sqrt{n}} \qquad \text{or} \qquad \text{reject } H_0 \text{ if } \bar{x} < \mu_0 - z_{1-\alpha}\frac{s}{\sqrt{n}}$$

(σ unknown) will have **approximate significance level** α, provided that n is large enough, regardless of the population distribution

As on Slides 29-30, the associated **approximate p-values** are

$$p = 1 - \Phi(z_0) \qquad \text{or} \qquad p = \Phi(z_0)$$

These tests are called **large sample tests**, as they require n large ($n > 40$, say) to be (approximately) valid.

Large sample test: example

Example

A manager in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contacts a week (he would like to increase this number). As a check on his claim, $n = 49$ salespeople are selected at random and the number of contacts made by each of them is recorded for a single randomly selected week. The mean and variance of the 49 measurements were $\bar{x} = 17$ and $s^2 = 9$, respectively. Do these figures contradict the manager's claim, at the significance level $\alpha = 0.05$? (**Hint:** You can use the Matlab outputs: $\texttt{norminv}(0.95) = 1.645$, $\texttt{normcdf}(4.67) = 1$)

Large sample test: example

Hypothesis tests and confidence intervals

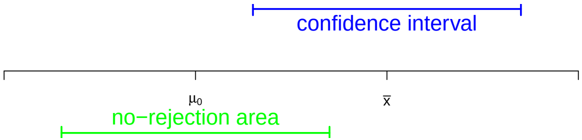
You might have noticed that the critical values for the (two-sided test) decision rule

$$\text{reject } H_0 \text{ if } \bar{x} \notin \left[\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

look like the limits of the (two-sided) confidence interval for μ

$$\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

As the interval widths are the same, the confidence interval (centred at $\bar{x}$) cannot contain μ_0 if the 'no-rejection area' (centred at μ_0) does not contain $\bar{x}$ (and vice-versa)



Hypothesis tests and confidence intervals

Generally speaking, there is always a **close relationship between the test of hypothesis about any parameter, say θ, and the confidence interval for θ**:

If $[\ell, u]$ is a $100 \times (1 - \alpha)\%$ confidence interval for a parameter θ, then the hypothesis test for

$$H_0 : \theta = \theta_0 \qquad \text{against} \qquad H_a : \theta \neq \theta_0$$

will reject H_0 at significance level α if and only if θ_0 is not in $[\ell, u]$

$\rightarrow$ hypothesis tests and CIs are more or less equivalent, however each provides somewhat different insights:

- CIs provide a range of likely values for θ
- tests easily display the risk levels, such as *p*-values, associated with a specific decision

Note: the same analogy exists between one-sided tests and one-sided confidence intervals

Objectives

Now you should be able to:

- structure engineering decision-making problems as hypothesis tests □
- understand the concepts of significance level, power, error of type I and of type II □
- test hypotheses on the mean of a normal distribution using either a z-test or a t-test □
- test hypotheses on the mean of an arbitrary distribution using the Central Limit Theorem □
- use the p-value approach for making decisions in hypothesis tests □
- explain and use the relationship between confidence intervals and hypothesis tests □

Recommended exercises:

$\rightarrow$ Q2, Q3, Q5, Q8, Q9, Q11 p.354, Q15 p.355, Q17, Q18, Q19 p.367, Q21, Q23 p.368, Q55 p.394, Q62 p.396 (2nd edition)

$\rightarrow$ Q2, Q3, Q5 p.361, Q8, Q9, Q11, Q15 p.362, Q18, Q19, Q20 p.374, Q22, Q24 p.375, Q65 p.406, Q73 p.408 (3rd edition)