Engineering Mathematics 2B

Module 19: Statistics - Hypothesis Testing

Nick Polydorides

School of Engineering

THE UNIVERSITY *of* EDINBURGH

Module 19 contents

Motivation: The big questions (and their small answers)

"Is the mean IQ of the population increasing now that we become computer literate from a young age?"

"Is the average temperature of the planet going up?"

"Is the mean standard (or cost) of living decreasing in the UK?"

"Are we facing an extinction thread by the advancement of Artificial Intelligence?"

Hypothesis Testing

A **statistical hypothesis** is a conjecture concerning the whole population, affirming what we already know or saying something different, e.g.

"Mean human IQ is 5 points higher than 50 years ago", *"Mean temperature is 0.4 C higher than 2000"*, *"Halving the syllabus of EM2B will boost passing rates by 20%"*, *"The average 18 year old spends 2.3 h on social media every day"*, etc

Null & Alternative Hypotheses

We pose the alternative hypothesis (new knowledge) H_1 against the null hypothesis H_0 (current state of knowledge).

We then **attempt to show that**, in light of our collected data, **the null hypothesis is false**, i.e. the data tell us something new!

This involves calculating the probability of having the data we collected "by accident", if the null hypothesis was indeed true.

If this probability is very small it suggests that the null H_0 could be false, and we may have to reject it (and adopt H_1).

If this probability is not small then there is **not enough evidence to reject the null** H_0.

One-sided & Two-sided HT

Suppose we want to test the hypothesis that the mean precipitation in Edinburgh in September is **not** the nominal $\mu_0 = 2.2$ inches

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0.$$

We call this a **two-sided HT** since H_1 holds for $\mu < \mu_0$ and $\mu > \mu_0$.

If the alternative hypothesis is either of: the mean precipitation is **less than** μ_0 or **more than** μ_0, then we have two different **one-sided HT**

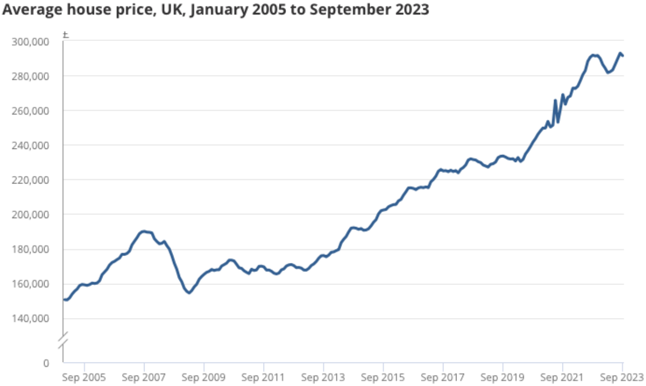
$$H_0 : \mu = \mu_0, \qquad H_1 : \mu < \mu_0$$
$$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0$$

One-sided and two-sided tests are treated differently.

Hypothesis Testing Procedure

1. Formulate the **null** and **alternative** hypotheses.
   $\downarrow$
2. Choose the **level of significance** or the right critical value for the HT.
   $\downarrow$
3. Design an experiment and collect (**data**).
   $\downarrow$
4. Calculate the relevant **test statistic** from the data.
   $\downarrow$
5. Compute the relevant **non-rejection region** assuming the null hypothesis holds.
   $\downarrow$
6. **Decision**: **Reject the null hypothesis** if the statistic falls outside the non-rejection region. Otherwise, retain it.

A matter that matters to everyone in the UK



H_0: The average house price is currently (late 2023) £291,385.

Source: Office for National Statistics. Annual house price data based on a sub-sample of the Regulated Mortgage Survey

Two sided (aka two-tail) test

The mean value £291385 from the ONS came up using data from mortgage surveys. We are also told that $\sigma = 55000$. Set $\mu_0 = 291385$.

Suppose that we table the alternative hypothesis that: "the average value of a house in the UK is not £291385".

To make the case we collect $n = 160$ independent sale prices from randomly chosen agencies across the UK, giving $\bar{x}_n = £298255$.

$$H_0 : \mu = \mu_0, \quad \text{and} \quad H_1 : \mu \neq \mu_0$$

As n is large, we may assume that $\bar{X}_n$ is normally distributed with a true mean μ. But is it equal to μ_0 or isn't?

Task: Perform a HT at $\alpha = 0.05$.

Two-sided test rules

Rule: If the chosen data statistic falls within the $1-\alpha$ non-rejection region then we cannot reject the null H_0. Otherwise we do, and adopt H_1.

The appropriate data statistic to choose depends on whether the variance of the population is known or not.

1. If σ^2 is known then we perform a Z test using the Z statistic
$$\bar{Z}_n = \frac{\bar{X}_n - \mu_0}{\mathrm{SE}} = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$
2. If σ^2 is unknown and n is small (< 100) then we opt for a T test using the T statistic
$$\bar{T}_n = \frac{\bar{X}_n - \mu_0}{\widehat{\mathrm{SE}}} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$$
3. If σ^2 is unknown and n is large (> 100) then Z or T tests will give almost identical results.

Two-sided test rules

Upon choosing the appropriate statistic, the two-sided $1-\alpha$ non-rejection region for that statistic will be either

1. $\left[-z_{\alpha/2},\, z_{\alpha/2}\right]$ for $\bar{z}_n$ a sample of $Z \sim \mathcal{N}(0,1)$, or
2. $\left[-t_{\alpha/2},\, t_{\alpha/2}\right]$ for $\bar{t}_n$ a sample of $T \sim t(\nu = n-1)$

It is possible to map these regions to those referring to the original (non-standardised) variable.

On $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ the respective non-rejection region for $\bar{x}_n$ will be

$$\left[x^{(l)}_{\alpha/2},\, x^{(r)}_{\alpha/2}\right]$$

where

$$\mathbb{P}\left(x^{(l)}_{\alpha/2} \leq \bar{X}_n \leq x^{(r)}_{\alpha/2}\right) = 1-\alpha$$

Two-sided test example

Returning to our example, if H_0 holds true, then $\mu = \mu_0 = 291385$.

Since we know $\sigma = 55000$ then we know the distribution from which our data $\bar{x}_n = 298255$, $n = 160$ is sampled from

$$\bar{X}_n \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$$

So the non-rejection region centred at μ_0[1] has bounds $[\mu_0 - \bar{x}_{\alpha/2}, \mu_0 + \bar{x}_{\alpha/2}]$ where

$$\mathbb{P}\left(\mu_0 - \bar{x}_{\alpha/2} \leq \bar{X}_n \leq \mu_0 + \bar{x}_{\alpha/2}\right) = 1 - \alpha$$

There are two ways to go from here: Computer or Tables.

[1]Notice this is NOT centred at the data sample mean.

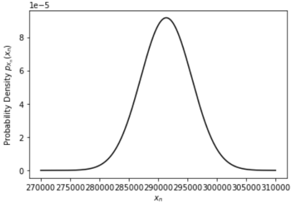
Two-sided tests on a computer

```
import matplotlib.pyplot as plt
import numpy as np
import scipy.stats as stats
import math

mu0 = 291385
variance = pow(55000,2)/160
sigma = math.sqrt(variance)
x = np.arange(270000,310000,0.1)
alpha=0.05

plt.plot(x, stats.norm.pdf(x, mu0, sigma), color = 'black')
plt.xlabel('$x_n$')
plt.ylabel('Probability Density $p_{X_n}(x_n)$')
plt.show()

val_left = stats.norm.ppf(alpha/2, loc=mu0, scale=sigma)
val_right = stats.norm.ppf(1- alpha/2, loc=mu0, scale=sigma)
print('The non-rejection region has bounds', val_left, val_right)
```




```
The non-rejection region has bounds 282862.81830581225 299907.18169418775
```

As $\bar{x}_n = 298255$ is inside this non-rejection region, we cannot reject H_0 at this α.

Two-sided tests with a table

As there are no tables for cumulative integrals of the Gaussian other than $\mathcal{N}(0,1)$ we will standardise the data and express the non-rejection region on the standard normal.

Rule: If the normalised data sample, assumed a realisation of the **the Z statistic**

$$\bar{Z}_n = \frac{\bar{X}_n - \mu_0}{\mathrm{SE}} \Rightarrow \bar{z}_n = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

falls within the $1-\alpha$ non-rejection region $[-z_{\alpha/2},\, z_{\alpha/2}]$ then we cannot reject the null H_0. Otherwise we do.

For our example

$$\bar{z}_n = \frac{298255 - 291385}{55000/\sqrt{160}} = 1.58, \quad \alpha = 0.05 \Rightarrow \text{NRR} : [-1.96, 1.96]$$

hence we cannot reject the null H_0.

Two-sided tests with a table

To confirm the computerised NRR, we return to the $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$ to find the NRR directly for $\bar{X}_n$. Let

$$x^{(l)}_{\alpha/2} \text{ s.t. } \mathbb{P}(\bar{X}_n \leq x^{(l)}_{\alpha/2}) = \frac{\alpha}{2}, \quad x^{(r)}_{\alpha/2} \text{ s.t } \mathbb{P}(\bar{X}_n \geq x^{(r)}_{\alpha/2}) = \frac{\alpha}{2}$$

Since we reject H_0 if $-1.96 \geq \bar{z}_n = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$ then we effectively reject H_0 if

$$\bar{x}_n \leq -1.96\frac{\sigma}{\sqrt{n}} + \mu_0 \Rightarrow \bar{x}_n \leq 282862.8 = x^{(l)}_{\alpha/2}.$$

Similarly, since we reject H_0 if $1.96 \leq \bar{z}_n = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$ then we reject H_0 if

$$\bar{x}_n \geq 1.96\frac{\sigma}{\sqrt{n}} + \mu_0 \Rightarrow \bar{x}_n \geq 299907.2 = x^{(r)}_{\alpha/2}$$

Errors in the decision

When can things go wrong: "a wrong decision is made". There are four possible HT outcomes:

	H_0 holds true	H_1 holds true
Rejecting H_0	**Type I error**	Right decision
Not rejecting H_0	Right decision	**Type II error**

Type I: False rejection of H_0, implying that we are too confident on the data.

Type II: False non-rejection of H_0, implying that we are too doubtful of the data.

Type I error

Rejecting the null H_0 while it is true is termed as a type I error. **The probability of committing a type I error is equal to the level of significance α** and it is found from $p_X(x|H_0)$.

The lower the significance level α, the less likely we are to commit a type I error. An $\alpha \leq 0.05$ allows a reasonable error tolerance. For two-sided tests where $\bar{X}_n \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$

$$\begin{aligned}
\mathbb{P}(\text{Type I error}) &= \mathbb{P}(x^{(l)}_{\alpha/2} \geq \bar{X}_n \geq x^{(r)}_{\alpha/2}) \\
&= \mathbb{P}(\bar{X}_n \leq x^{(l)}_{\alpha/2}) + \mathbb{P}(\bar{X}_n \geq x^{(r)}_{\alpha/2}) \\
&= \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

where

$$x^{(l)}_{\alpha/2} := \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad x^{(r)}_{\alpha/2} := \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

the left tail quantile and right critical points on the null PDF respectively.

Type II error

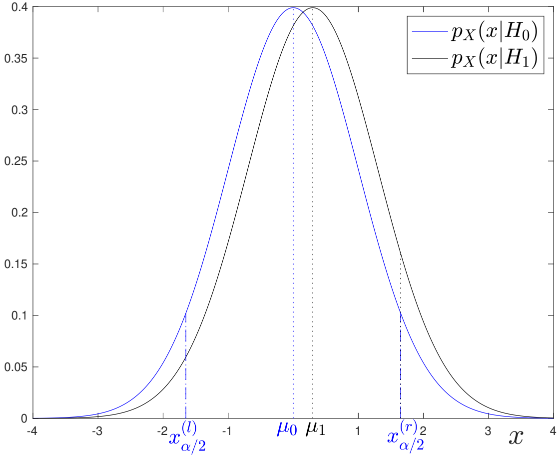
Failure to reject H_0 when H_1 is true is called a Type II error, and has a probability denoted by β. To compute β we need a **single valued** H_1 and a PDF for when H_1 holds $p_X(x \mid H_1)$.

Returning to our example, we may modify $H_1 : \mu \neq \mu_0$ by picking for example $H_1 : \mu_1 = \bar{x}_n$. In effect, assuming $\bar{X}_n \sim \mathcal{N}(\mu_1, \frac{\sigma^2}{n})$, we have

$$\begin{aligned}
\mathbb{P}(\text{Type II error}) &= \mathbb{P}\big(x^{(l)}_{\alpha/2} \leq \bar{X}_n \leq x^{(r)}_{\alpha/2} \mid \bar{X}_n \sim \mathcal{N}(\mu_1, \frac{\sigma^2}{n})\big) \\
&= \mathbb{P}(\bar{X}_n \leq x^{(r)}_{\alpha/2} \mid \mu = \mu_1) - \mathbb{P}(\bar{X}_n \leq x^{(l)}_{\alpha/2} \mid \mu = \mu_1) \\
&:= \beta
\end{aligned}$$

To evaluate β in this example we can standardise $\bar{X}_n$ and use $\Phi(z)$.

Probabilities of error and power as areas



Shade the areas corresponding to the probabilities of type I, type II errors and power of the 2-sided test.

Power of the hypothesis test

The power of a test is the probability of rejecting H_0 given that the true parameter is not in H_0 but rather a value in H_1, i.e. "to trust what the data are telling us against H_0"

$$\text{Power} = 1 - \beta.$$

In summary we have:

	H_0 holds true	H_1 holds true
Rejecting H_0	$\mathbb{P}$(**Type I error**) $= \alpha$	$\mathbb{P}$(Right decision) $=$ Power $= 1 - \beta$
Not rejecting H_0	$\mathbb{P}$(Right decision) $= 1 - \alpha$	$\mathbb{P}$(**Type II error**) $= \beta$

One-sided (aka single tail) tests

We now turn our attention to hypothesis tests where

$$H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0 \qquad \text{left tail test}$$

or

$$H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0 \qquad \text{right tail test}$$

As with the two-tail case, to do these tests we require a level of significance α and some data that yield $\bar{x}_n$ a sample of $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n^2})$

The process is similar to the two-sided test, but without splitting α in two and the respective confidence intervals have one of their bounds equal to $\pm\infty$.

Example: Left-sided test

We have data

$$\{12.1, 6.0, 12.5, 9.1, 7.9, 9.1, 7.7, 7.8, 10.2\}$$

from a normal distribution whose mean is said to be $\mu_0 = 10$.

Task: Perform a hypothesis test at $\alpha = 0.01$ to verify if the true mean is less than μ_0.

Notice:

- $n = 9$, sample is small,
- We are not given σ^2,
- We are exclusively testing for the true mean being less than μ_0 (as opposed to different than).

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu < \mu_0$$

Example: Left-sided test

Since we have a small sample (i.e. $n = 9$) we will use **the T statistic**

$$\bar{t}_n = \frac{\bar{x}_n - \mu_0}{\widehat{\mathrm{SE}}} = \frac{\bar{x}_n - \mu_0}{S_n/\sqrt{n}}$$

since we expect $\bar{t}_n$ to be a realisation of

$$T \sim t(\nu), \quad \nu = n - 1 = 8.$$

where from the data we get

$$\bar{x}_n = 9.15, \quad \widehat{\mathrm{SE}} = \frac{\widehat{\sigma}}{\sqrt{n}} = \frac{S_n}{\sqrt{n}} = \frac{2.13}{\sqrt{9}} = 0.71$$

and the respective NRR is

$$[t_\alpha, +\infty) \quad \text{where} \quad \mathbb{P}(T \leq t_\alpha) = \alpha = 0.01,$$

so t_α is a quantile on the $t(8)$ distribution.

Example: Left-sided test

Rule: If the realisation of the T or Z statistic (whichever is applicable) via

$$\bar{t}_n = \frac{\bar{x}_n - \mu_0}{\widehat{\text{SE}}} = \frac{\bar{x}_n - \mu_0}{\hat{\sigma}/\sqrt{n}}, \quad \text{or} \quad \bar{z}_n = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

falls within the $1 - \alpha$ NRR $[t_\alpha, +\infty)$ then we cannot reject the null H_0. Otherwise we do.

Returning to our example, three things remain:

1. Evaluate the T statistic,
2. Find the quantile t_α from the t-distribution tables, and
3. Check if the t statistic is inside the NRR. If it is we cannot reject H_0 otherwise we do, and go with H_1.

The statistic evaluates to

$$\bar{t}_n = \frac{\bar{x}_n - \mu_0}{\widehat{\text{SE}}} = \frac{9.15 - 10}{0.71} = -1.1972$$

Example: Left-sided test

One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	
	50%	60%	70%	80%	90%	95%		99%	99.5%	
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.321	3
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.089	2
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	1
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5
	0.711	0.896	1.119	1.415	1.895	2.365		3.499	4.029	4
8	0.706	0.889	1.108	1.397	1.860	2.30	2.896	355	3.833	4
	0.703	0.883	1.100	1.383	1.833	2.262		3.250	3.690	4

Looking at the $t(\nu = 8)$ table, one-sided test, we notice that there is no column for $\alpha = 1\%$ as needed in $t_{0.01}$.

Recall that $t(\nu)$ are symmetric wrt $t = 0$, and thus

$$t_{0.01} = -t_{1-0.01} = -t_{0.99} = -2.896$$

$$\bar{t}_n \in [-2.896, +\infty) \quad \Longrightarrow \quad \text{we cannot reject } H_0 \text{ at } \alpha = 1\%.$$

Example: Right-sided test

Suppose now we have data

$$\{12.9, 5.7, 13.9, 8.5, 7.3, 8.8, 5.3, 9.7, 11.0, 8.1\}$$

from a normal distribution whose mean is said to be $\mu_0 = 7$.

Task: Perform a hypothesis test at $\alpha = 0.05$ to verify if the true mean is more than μ_0.

As with the left-sided test example, we have a small sample of $n = 10$, we are not given σ^2, and we are only interested to see if the true mean is more than μ_0 (as opposed to different than).

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0$$

Example: Right-sided test

The combination of small sample and unknown variance, impose the use of HT based on the T statistic, hence from the data

$$\bar{t}_n = \frac{\bar{x}_n - \mu_0}{S_n/\sqrt{n}} = \frac{9.12 - 7}{2.83/\sqrt{10}} = 2.3689,$$

and the right tail NRR is

$$(-\infty, t_\alpha], \quad \text{where} \quad \mathbb{P}(T \geq t_\alpha) = \alpha = 0.05,$$

so t_α is a critical value on the $t(9)$ distribution.

Notice: The sought t_α is the point on the horizontal axis on $t(\nu)$ so that the area underneath it and to right equals α. This is the same point as the $1 - \alpha$ quantile.

Example: Right-sided test

One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%
Two-sided	50%	60%	70%	80%	[illegible]	95%	98%	99%
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	[illegible]	[illegible]306	2.896	3.355
9	0.703	0.883	1.100	1.3[illegible]	1.833	[illegible]62	2.821	3.250
10	0.700	0.879	1.093	1.372	[illegible]	[illegible].228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106

In effect,

$$t_\alpha \quad \text{such that} \quad \mathbb{P}(T \geq t_\alpha) = 0.05$$

is the same as

$$t_\alpha \quad \text{such that} \quad \mathbb{P}(T \leq t_{1-\alpha}) = 1 - 0.05 = 0.95$$

Example: Right-sided test

From the table we can fix our non-rejection region as

$$(-\infty, 1.833],$$

and since out T statistic $\bar{t}_n = 2.3689$ is outside this NRR, we have to reject H_0 and adopt H_1.

We can be 95% confident that the true mean is larger than μ_0.

Summary

	H_0 is true	H_1 is true
Reject H_0	Type I error	Correct decision
Don't reject H_0	Correct decision	Type II error

$$\begin{aligned}
\alpha = \text{significance level} &= \mathbb{P}(\text{type I error}) \quad (\text{want close to } 0) \\
&= \mathbb{P}(\text{incorrectly reject } H_0) \\
&= \mathbb{P}(\text{test statistic outside NRR } \mid H_0) \\
&= \mathbb{P}(\text{false positive})
\end{aligned}$$

$$\begin{aligned}
\text{Power} &= 1 - \mathbb{P}(\text{type II error}) \quad (\text{want close to } 1) \\
&= \mathbb{P}(\text{correctly reject } H_0) \\
&= \mathbb{P}(\text{test statistic outside NRR } \mid H_1) \\
&= \mathbb{P}(\text{true positive})
\end{aligned}$$

Main outcomes of module 19

You MUST know:

1. What is the purpose of hypothesis testing at a significance level α.
2. To perform two- and one- sided hypothesis tests with Z and T statistics (using the tables).
3. The type I and II errors in HT and the power of a test.