ACTL2131 2.4 - Hypothesis Testing

Tadhg Xu-Glassop

2025T1

MyExperience

Please do it!

Motivating Hypothesis Testing

- When we collect data, we may want to make claims from our results.

    E.g. the heights of two different teams is different, the population mean of the returns of a stock is non-zero...

- But our observations are random, so in theory we could observe any value! This makes things tricky.
- We need a formal way of testing claims (hypothesis) about our statistics/data - a way of determining when something is "too unlikely" to be true.

Setting up a test

Obviously, we need two things coming into the test:

- A statement / hypothesis about our data;

Setting up a test

Obviously, we need two things coming into the test:

- A statement / hypothesis about our data;

  A null hypothesis $\mathcal{H}_0$ that we want to reject, and an alternate hypothesis $\mathcal{H}_1$ we wish to accept instead.

Setting up a test

Obviously, we need two things coming into the test:

- A statement / hypothesis about our data;

  A null hypothesis $\mathcal{H}_0$ that we want to reject, and an alternate hypothesis $\mathcal{H}_1$ we wish to accept instead.
- What constitutes as something "too unlikely"

Setting up a test

Obviously, we need two things coming into the test:

- A statement / hypothesis about our data;

  A null hypothesis $\mathcal{H}_0$ that we want to reject, and an alternate hypothesis $\mathcal{H}_1$ we wish to accept instead.

- What constitutes as something "too unlikely"

  A level of significance, α. If the probability of the null being true is less than this, then we reject our null.

Example

Suppose the height of UNSW students has unknown population mean μ. We may wish to test

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu \neq 2,$$

or

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu < 2,$$

under the significance level of $\alpha = 5\%$.

How do I find out how “rare” my data is?

Remember, we can compute statistics from our sample such as $\bar{X}$ and s^2.

If only there was a way to **pivot** from these sample statistics to a known distribution...

How do I find out how “rare” my data is?

Remember, we can compute statistics from our sample such as $\bar{X}$ and s^2.

If only there was a way to **pivot** from these sample statistics to a known distribution...

- Pivots (from confidence intervals) let us do exactly this!

Then, we assume (almost like a contradiction) that the null is true, and see how likely the things we observe are. If it is less than α and more likely to be $\mathcal{H}_1$, we reject $\mathcal{H}_0$ in favour of $\mathcal{H}_1$.

Example (cont.)

Suppose the height of UNSW students is normally distributed with unkown mean μ and known variance σ^2. We collect n heights and compute $\bar{X}$. We want to test

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu \neq 2,$$

with $\alpha = 5\%$.

Example (cont.)

Suppose the height of UNSW students is normally distributed with unkown mean μ and known variance σ^2. We collect n heights and compute $\bar{X}$. We want to test

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu \neq 2,$$

with $\alpha = 5\%$.

A suitable pivot here that incorporates everything I want is

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

Testing our Null Hypothesis

As mentioned before, we assume our null is true and aim to test how plausible this assumption is. If it's unlikely to be a good assumption (under α), we reject.

There are two ways of achieving this:

- If $\mathcal{H}_0$ is true, I can use my pivot to make a $1-\alpha$ confidence interval on the value of my statistics! If what I observe falls outside, then my assumption is bad.
- If $\mathcal{H}_0$ is true, I can find the probability **my observation and anything else more extreme** is! If it's very low (less than α), then my assumption is bad.

These are called **critical region** and *p*-**values** respectively. You only need to do one, and they are equivalent.

Example (cont.)

Again, suppose the height of UNSW students X is normally distributed with unkown mean μ and known variance 0.025. We collect 40 heights and compute $\bar{X} = 1.94$. We want to test

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu \neq 2,$$

with $\alpha = 5\%$.

Example (cont.)

Again, suppose the height of UNSW students X is normally distributed with unkown mean μ and known variance 0.025. We collect 40 heights and compute $\bar{X} = 1.94$. We want to test

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu \neq 2,$$

with $\alpha = 5\%$.

A suitable pivot here that incorporates everything I want is

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

Who thinks we're going to reject?

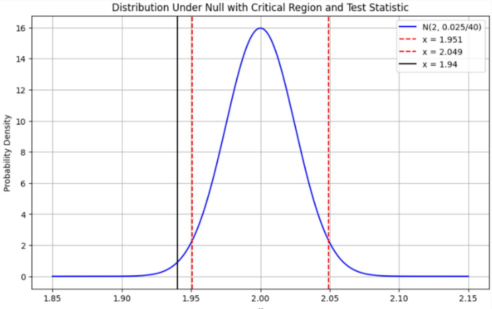
Example - Critical Region

Let's assume that $\mathcal{H}_0$ is true, and find a confidence interval for my statistic $\bar{X}$.

Then, $X|\mathcal{H}_0 \sim \mathcal{N}(2, 0.025)$, and so I find a $1-\alpha = 0.95$ confidence interval for $\bar{X}$ using my pivot.

$$\implies 2 \pm 1.96 \cdot \sqrt{0.025/40} \implies (1.951, 2.049).$$

Clearly, 1.9 falls out of this region. So, we reject $\mathcal{H}_0$ in favour of $\mathcal{H}_1$. That is, there is significant statistical evidence that $\mu \neq 2$.



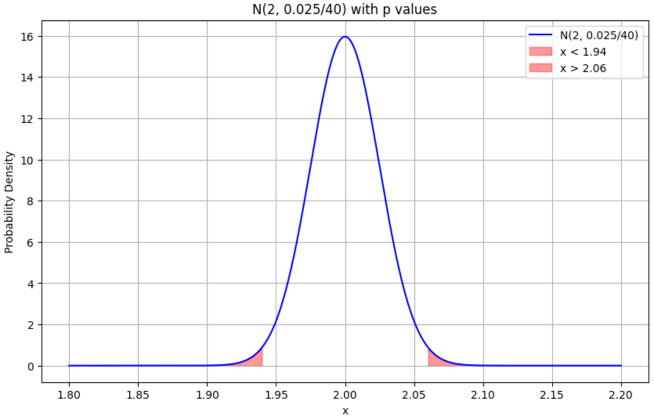
Example - p-value

Let's assume that $\mathcal{H}_0$ is true, and find the probability of observing my value of $\bar{X}$ or anything more extreme!

Example - p-value

Let's assume that $\mathcal{H}_0$ is true, and find the probability of observing my value of $\bar{X}$ or anything more extreme!

Drawing a graph here is very useful here - since our distribution is symmetrical and our test is two tailed*, this includes both tails.



Example - p-value (cont.)

Now, under the null,

$$T|\mathcal{H}_0 = \frac{1.94 - 2}{\sqrt{0.025/40}} = -2.4.$$

So, the p-value is

$$p = 2 \cdot \Phi(-2.4) = 2(1 - \Phi(2.4)) = 0.0164.$$

Since $0.0164 < \alpha$, we reject our null. That is, there is significant statistical evidence that $\mu \neq 2$.

Hypothesis Testing

Method

1. Establish $\mathcal{H}_0$ and $\mathcal{H}_1$ and some significance level α.
2. Find a pivot that connects what we're trying to test with what's available.
3. Either find the critical region or the *p*-value, and make a conclusion.

One-tailed vs Two-tailed tests

But what if we wanted to test

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu > 2$$

under $\alpha = 5\%$? If we used the same process, would you say that $\mu > 2$?

One-tailed vs Two-tailed tests

But what if we wanted to test

$$\mathcal{H}_0 : \mu = 2 \qquad \text{v.s.} \qquad \mathcal{H}_1 : \mu > 2$$

under $\alpha = 5\%$? If we used the same process, would you say that $\mu > 2$?

No! We found that $\bar{X} = 1.94$ - how is this evidence that $\mu > 2$?

To adapt,

- Our critical region would be from $(-\infty, L)$ (one tail).
- Our *p*-value would only be the area that the alternate is implying (one tail).

This contrasts one-tailed and two-tailed tests.

Advice

There is so much to consider, but you can still approach each test intuitively.

Think:

- When would I reject $\mathcal{H}_0$ for $\mathcal{H}_1$? Under big or small or positive or negative T ?
- Does your result make sense?

Extra Terminology

Type I Error - α

- Reject $\mathcal{H}_0$ when $\mathcal{H}_0$ is true

Type II Error - β

- Don't reject $\mathcal{H}_0$ when $\mathcal{H}_0$ is false.

Power - $\pi = 1 - \beta$.

Tutorial Questions

2.4.6 (1-3), 2.4.5, 2.5.4