Lecture 24 Hypothesis Testing Terms

BIO210 Biostatistics

Xi Chen

Spring, 2024

School of Life Sciences
Southern University of Science and Technology

南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Hypothesis Testing

Logic:

Some info./questions/claims:

- parameters of interest
- compare various groups

→ Hypothesis →

Tests:
If the opposite were true, the probability of observing ... is ...

- 9% of people have blood type AB
- $> 25\%$ of MCQs with B as the correct answer
- normal body temperature is 37 °C
- Compare mean test scores of two groups of students

Design experiments and collect data, you come up with some hypotheses:
- The proportion of people with blood type AB is not 9%
- The proportion of MCQs whose correct answers are Bs is $> 25\%$
- The normal body temp. is not 37 °C
- The Mean test score in group A is higher than that in group B

Is the probability small or large?

The Null Hypothesis And The Alternative Hypothesis

- In inferential statistics, the null hypothesis is a general statement or default position that there is no relationship/difference/association between measured phenomena/groups.

- The alternative hypothesis is the opposite to the null hypothesis. They are collectively exhaustive and mutually exclusive.

The Null Hypothesis And The Alternative Hypothesis

H_0	H_1 or H_a
The proportion of blood type AB in the COVID-19 patients is 0.09. $(H_0 : \pi = 0.09)$	The proportion of blood type AB in the COVID-19 patients is not 0.09. $(H_0 : \pi \neq 0.09)$
The proportion of MCQs whose correct answers are Bs is equal to or less than 0.25. $(H_0 : \pi \leqslant 0.25)$	The proportion of MCQs whose correct answers are Bs is higher than 0.25. $(H_1 : \pi > 0.25)$
The mean body temperature of normal people is 37 °C. $(H_0 : \mu = 37)$	The mean body temperature of normal people is not 37 °C. $(H_1 : \mu \neq 37)$

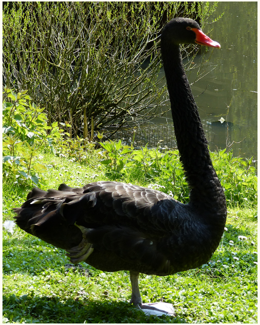
Null Hypothesis Significance Testing (NHST)

- $\boldsymbol{H_0}$ vs $\boldsymbol{H_1}$: H_1 is the negation of H_0, and vice versa.

- Why test the null hypothesis ?

  - Scientific methods: can be falsified/disproved.
  - Introduce less bias, such as confirmation bias.
  - Practically easier.

All swans are white.



Fisher vs Neyman-Pearson

https://en.wikipedia.org/wiki/Statistical_hypothesis_testing

The p-value And The Test Statistics

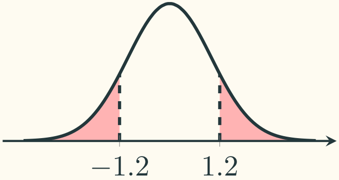
- Given that the null hypothesis is true, the probability of obtaining a measurement as extreme as or more extreme than the observed sample is called the *p*-value.

$$\mathbb{P}\left(\text{data or more extreme} \mid H_0 \text{ is true}\right)$$

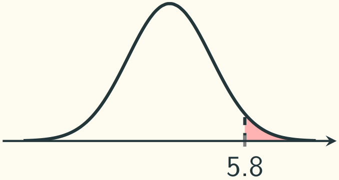
- How do we perform the calculation ?

✓ By Calculating **the test statistic** and use the properties of the sampling distributions.

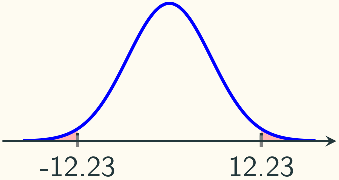
$$z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = 1.2$$



$$z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \geqslant 5.8$$



$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = -12.23$$



Significance Test

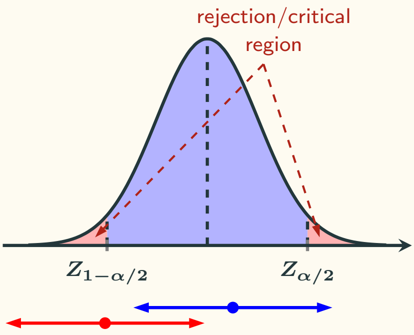
- If the p-value is **is small**, we reject the **null hypothesis**.
- How small ?

\- Ronald Fisher. **Statistical Methods for Research Workers (1925).**

> DISTRIBUTIONS 45
>
> only once in 370 trials, while Table II. shows that to exceed the standard deviation sixfold would need nearly a thousand million trials. The value for which P = ·05, or 1 in 20, is 1·96 or nearly 2 ; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant. Using this criterion, we should be led to follow up a false indication only once in 22 trials, even if the statistics were the only guide available. Small effects will still escape notice *if the data are insufficiently numerous to bring them* out, but no lowering of the standard of significance would meet this difficulty.

Significance Test And Confidence Interval

- Common p-value cutoffs: 0.01, 0.05, 0.10.
- In 1933, Jerzy Neyman and Egon Pearson called those cutoffs as significance levels, denoted by α. A significance level must be decided ahead of time. $Z_\alpha, Z_{1-\alpha}, Z_{\alpha/2}, Z_{1-\alpha/2}$, are called critical values.
- When the p-value is smaller than α, we reject the null hypothesis, and we say the result is statistically significant.



$p \geqslant \alpha \;\Leftrightarrow$ the test statistic falls into the middle zone $\Leftrightarrow\; (1-\alpha)\times 100\%$ covers π_0 or μ_0

$p < \alpha \;\Leftrightarrow$ the test statistic falls into the rejection/critical region $\Leftrightarrow (1-\alpha)\times 100\%$ does NOT cover π_0 or μ_0

Interpret The Result

- p-value $< \alpha$, reject H_0, **accept** H_1;
- p-value $\geqslant \alpha$, do not reject H_0 .

- Meanings and warnings:

  1. failing to reject H_0 does NOT mean that the null hypothesis is true. The same goes to H_1.
  2. The test is about the data, NOT your theory or hypothesis! Remember the p-value is $\mathbb{P}\,(\text{data or more extreme} \mid H_0 \text{ is true})$.
  3. **Accept**: act/behave as if ...

courtroom trial

Innocent H_0	Guilty H_1
Assumed to be true until convincing evidence suggests otherwise	Accepted if H_0 is rejected by convincing evidence

Rationale of Significance Test

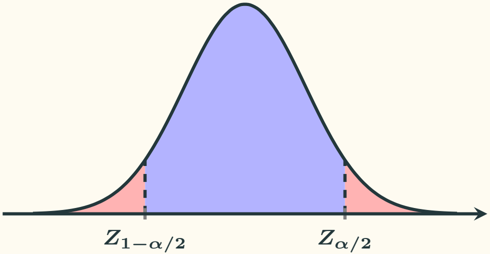
We assume that the POPULATION parameter of interest takes a certain value π_0 *or* μ_0

↓

A random sample is collected

↓

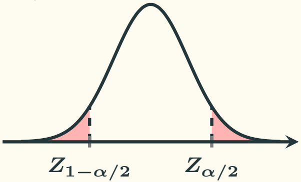
We use the properties of the sampling distribution of the sample proportion/mean to analyse the sample statistic: High chance; Low chance



Steps For Hypothesis Testing

1. Specify what you are comparing
2. Formulate hypotheses
3. Check assumptions
4. Determine significance level α
5. Compute the test statistic
6. Check significance
7. Make a decision about whether to reject H_0
8. Interpret findings

6a)



6b) Calculate the p-value

6c) Construct $(1-\alpha)\times 100\%$ confidence interval to see if it covers the H_0 value.