Week 9 9.2 Note

Lecture 28: Hypothesis Testing

---

1. Hypothesis Testing

> **Hypothesis Testing**: A scientific method for **weighing up (assess) the evidence** given in the data against a given hypothesis (model).

- If the gap between the observed value (**OV from the data**) and the expected value (**EV from the hypothesis**) is too big (**over than 2 or 3 SEs**), we say that **the data is not consistent with the hypothesis**.

2. Framework of Hypothesis Testing (HATPC)

- Steps:
  1. Set up the research *question*.
    - **H:** Hypothesis H0 and H1
  2. Weigh up the *evidence*.
    - **A:** Assumptions
    - **T:** Test Statistic
    - **P:** P-Value
  3. Explain the *conclusion*.
    - **C:** Conclusion
- **HATPC** (Hypothesis, Assumption, Test Statistics, P-Value, Conclusion):
  - H: Hypothesis（假设）
    - **Null Hypothesis H0（零假设）**: Assume that the **difference** between the OV (data) and EV is **due to the chance alone**.
    - **Alternative Hypothesis H1（备选假设）**: Assume that the **difference** between the OV (data) and EV is **NOT due to the chance alone**.
    - 2 *box models* to represent H0 and H1.
    - Example (2 sided alternative):
      - Research Question: Does the probiotic treatment work for 80% of patients?
      - H0: 80% respond to the treatment. (H0: P = 0.8)
      - H1: More or less than 80% respond to the treatment. (H1: P =/ 0.8)
  - A: Assumption（推测）
    - A conclusion is **not transparent** if the assumptions are **not stated**.
    - A conclusion is potentially **invalid** if the assumptions are **not justified**.
    - Example:
      - Assume each child in the trial was *independent of each other* (not related or no similar health profile) - **state an assumption**
      - Assume each child had the *same chance of showing improvement* with allergy by using the probiotic. - **state another assumption**
      - Check these 2 assumptions by *looking at the records* of the medical trial. - **justify the assumptions**

- T: Test Statistic（检验统计量）
  - A test statistic measures the **difference** between what is **observed in the data** and what is **expected from the null hypothesis**.
  - Formula:

$$\text{test statistic} = \frac{\text{observed value (OV) - expected value (EV)}}{\text{standard error (SE)}}$$

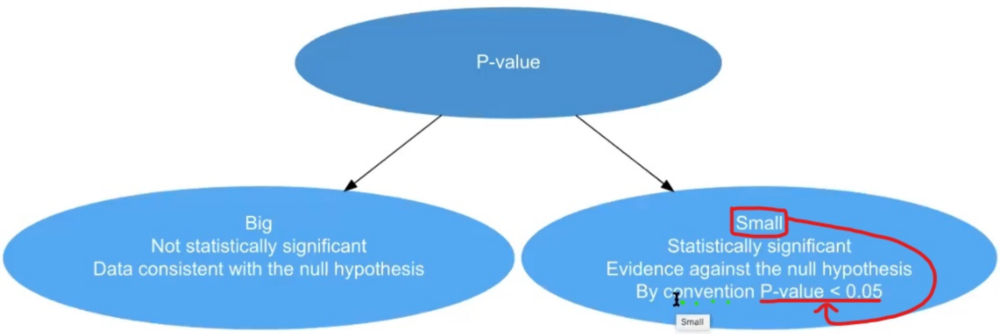
  - Note: **if the null hypothesis is true, then the test statistic is the *standard unit* corresponding to the *observed value*.** （很像standard unit的算法：Gap/SD）
  - Example:
    - Let X = the number of people in the medical trial who showed improvement in their peanut allergy, which for a particular sample is x.
    - If $H_0 : p = 0.8$ is true, we expect $\text{EV} = np$ improvements with $\text{SE} = \sqrt{np(1-p)}$.
    - So the test statistic is

$$\text{test statistic} = \frac{X - np}{\sqrt{np(1-p)}}$$

    - For the observed value of the test statistic, subsitute x for X.

Note (Extension): X has a Binomial distribution.

- P: P-value (observed significance level)（P值）
  - P-value is a way of **weighing up** whether the **sample** is consistent with H0.
  - P-value is the chance of observing the test statistic (or something more extreme) if H0 is true.
  - *Size* of P-value:



  - *Common Mistakes* with the P-value:
    - The P-value is *not the chance* that the *null hypothesis is true*.
    - A *large P-value* does *not* mean that *H0 is definitely true*.

Size of p-value	What not to say	What to say
Small	Ho is not true	There is evidence against Ho
	Ho is false	We reject Ho
Large	We accept Ho	Data is consistent with Ho
		We retain Ho

    - The significance level of 0.05 is a *convention*. Some people use 0.01 and say that the result is *highly significant*.

Lecture 29: Proportion Test

1. Proportion Test

- **Initial Trial Results**

Group	Participants	Numbers showing desensitisation
Treatment	29	26
Placebo	28	2

- **H: Hypothesis**

  Suppose the research team wants to claim that the new oral immunotherapy has a desensitisation rate of higher than 80%.

  H_0: 80% of people respond to the treatment. [Or $H_0 : p = 0.8$, where p = Proportion of patient who respond to the treatment (desensitise to peanut allergy).]

  H_1: More than 80% respond to the treatment. [Or $H_1 : p > 0.8$]

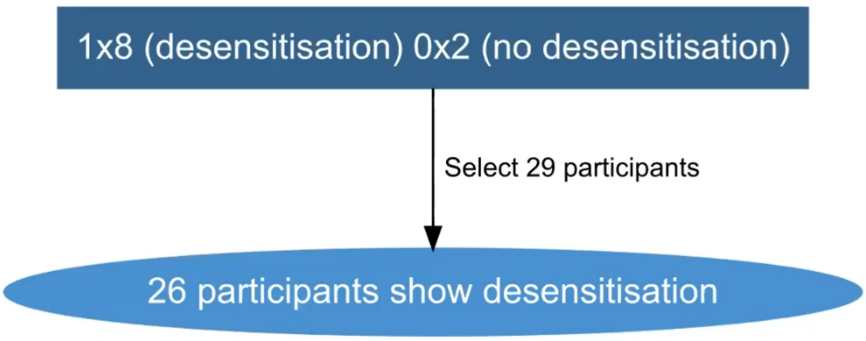
- **A: Assumptions**

  The participants in the treatment group are independent of each other.

  The chance of becoming desensitised is the same for all participants.

- **T: Test Statistic**

  If H_0 is true, then we can model the Treatment participants by a simple box model where a 1 ("shows desensitisation") and 0 ("doesn't show desensitisation").



For the box

- The mean is $\frac{1\times8+0\times2}{10} = 0.8$
- The SD is $(1-0)\sqrt{0.8\times0.2} = 0.4$

For the box model (modelling the Sum of the Sample)

- EV = $29 \times 0.8 = 23.2$ EV (Sum) = People * Mean
- SE = $\sqrt{29} \times 0.4 \approx 2.2$ SE (Sum) = sqrt(People) * SD

$$\text{test statistic} = \frac{\text{OV - EV}}{\text{SE}}$$

The observed value of the test statistic is:

standardized

$$t_{obs} = \frac{26 - 23.2}{2.2} \approx 1.3$$

- **P: P-value**

  The **p-value** is the chance of observing the test statistic (or something more extreme) if H_0 really is true.

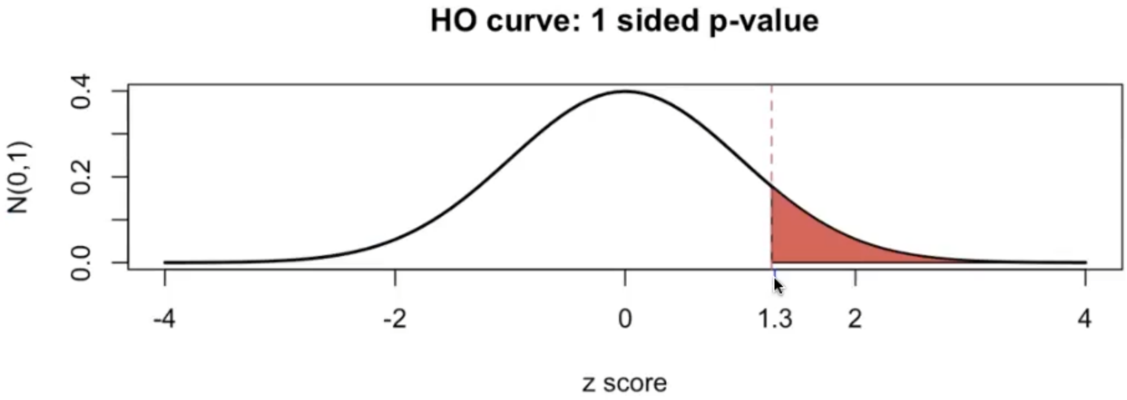
  $$P(\text{test statistic} \geq 1.3)$$

  Modelling by a Normal (assuming CLT), we find that

  $$P(\text{test statistic} \geq 1.3) \approx 0.097$$ vs. 0.05

```
1-pnorm(1.3)
```

```
## [1] 0.09680048
```



  - Note: Due to the *small sample size*, the *normality assumption may not be suitable*. We need to use simulation for the sample sum.

- **C: Conclusion**

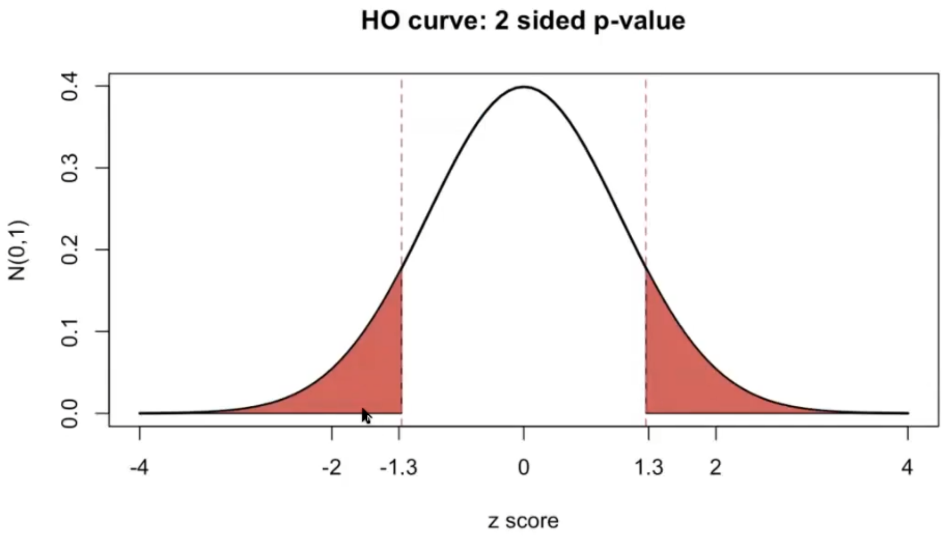
  We conclude that the data is consistent with the null hypothesis. H_0

  ie the new treatment does not seem to have an effectiveness rate higher than 0.8.

  Note we have not proved that the effectiveness rate is 0.8. Rather, we have just failed to find sufficient evidence to claim an effectiveness rate of higher than 0.8.

- Effect on H1:
  - 2 types of alternative hypotheses:
    - 1 sided: **specifies the direction** of the H1.
      - Ex. H1 : p > 0.8.
    - 2 sided: does not specify the direction of the H1.
      - ex. H1 : p =/ 0.8.
      - In this case, we usually **double the p-value**.

      -



2. Simulating P-value

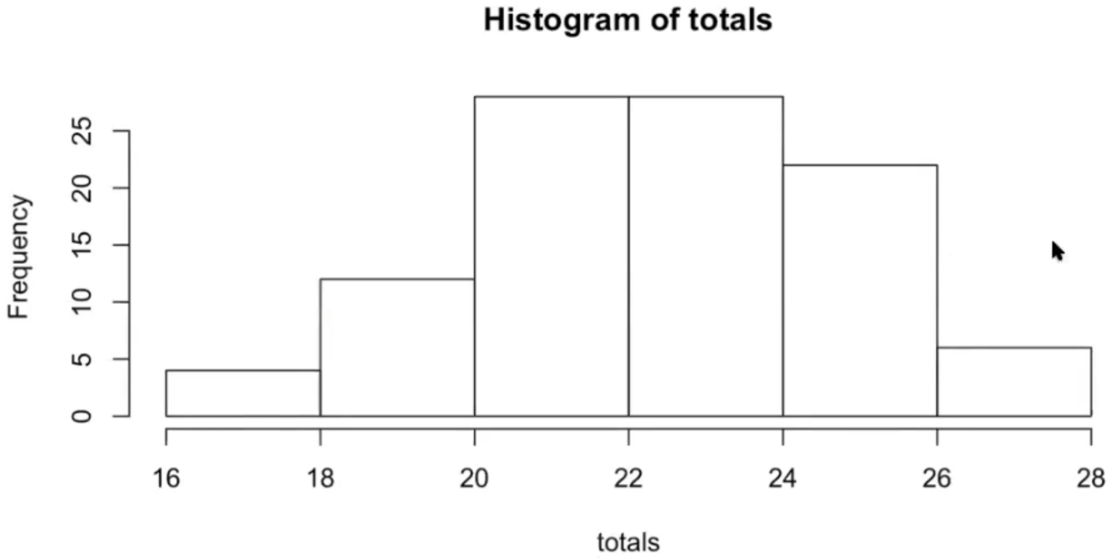
- It's possible that the 26 participants showing desensitisation was a rare result.
- Simulating 100 times:

```
set.seed(1)
box=c(1,1,1,1,0)  # using proportions of 80% and 20% in box
totals = replicate(100, sum(sample(box, 29, rep = T)))
table(totals)
```

```
## totals
## 16 17 18 19 20 21 22 23 24 25 26 27 28
##  1  2  1  7  5 12 16 15 13 16  6  4  2
```

```
hist(totals)
abline(v=28,col="green")
```

```
## totals
## 16 17 18 19 20 21 22 23 24 25 26 27 28
##  1  2  1  7  5 12 16 15 13 16  6  4  2
```



The estimated p-value here would be (6+4+2)/100.

- The p-value = (88242 + 39131 + 11517 + 1512)/1000000 = 0.14 (2dp).
- This is the approximation/estimate of the chance of getting 26 or more participants who show desensitisation.
- Compare this with 0.09, using the CLT.
- Why the difference? The much higher proportion of 1s vs. 0s (4-fold ratio) combined with a relatively small sample size (29) implies that the distribution of the number of 1s in the sample is left skewed.
- Extension: The exact p-value is based on the Binomial model.

```
x=c(26,27,28,29)
sum(dbinom(x,29,0.8))
```

```
## [1] 0.1403805
```