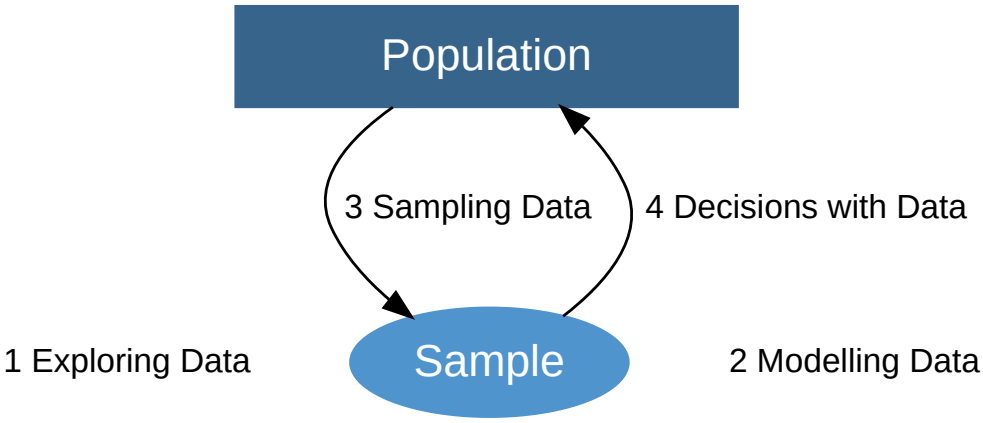
# Hypothesis Testing

## Decisions with Data | Test for a Proportion

# Unit Overview



# Module4 Decisions with Data

## Test for a Proportion

How can we make evidence based decisions? Is an observed result due to chance or something else? How can we test whether a population has a certain proportion?

## Tests for a Mean

How can we test whether a population has a certain mean? Or whether 2 populations have the same mean?

## Tests for a Relationship [DATA1001/MATH1115]

How can we test whether 2 variables are linearly related? How can we test whether a categorical variable is in certain proportions?

# Framework for Hypothesis Testing

# Data Story

Does a probiotic cure peanut allergy?

# Does a probiotic treatment cure peanut allergy?

- Food allergies have risen dramatically in the past 20 years.
- Whereas allergies to egg, milk, wheat, and soy generally resolve during childhood, nut and seafood allergies often persist throughout life.
- Only around 20 percent of children grow out of their peanut allergy.
- The constant vigilance required to avoid allergens substantially impairs quality of life of both food-allergic children and their families.

# ABC release in 2017

ABC

# The PPOIT study in Melbourne

- A team from Murdoch has been studying long-term clinical and immunological effects of probiotic and peanut oral immunotherapy after treatment cessation.
- The trial consisted of a 4-year follow-up of a randomised, double-blind, placebo-controlled trial.

## Statistical Thinking

With the person next to you:

- If you had a peanut allergy, would you take part in the probiotic trial?
- If you had a small child with a peanut allergy, would you let them take part in the probiotic trial?
- What would convince you that it worked?

# Why hypothesis testing?

# Why hypothesis testing?

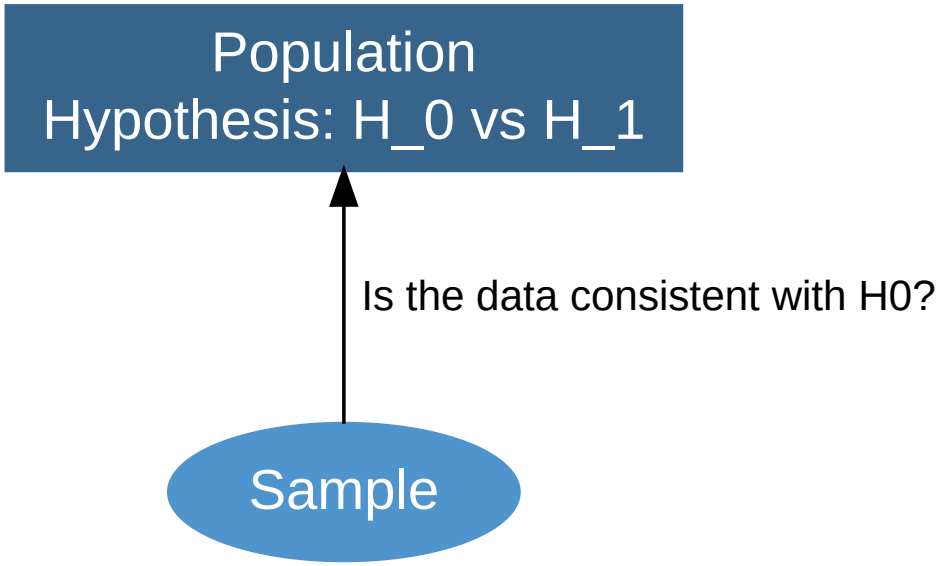
- To make evidence based decisions, we need to **weigh up** evidence.
- Hypothesis Testing is a scientific method for weighing up the evidence given in the data against a given hypothesis (model).

### Hypothesis testing

We say that the data is not consistent with the hypothesis if the gap between the observed value (data) and the expected value (assuming the hypothesis) is too big (eg more than 2 or 3 standard errors).

# Framework for hypothesis testing

# Framework for hypothesis testing



# 3 main steps

## 1. Set up research question

H: Hypothesis $H_0$ vs $H_1$

## 2. Weigh up evidence

A: Assumptions
T: Test Statistic
P: P-value

## 3. Explain conclusion

C: Conclusion

# H: Hypotheses

## Hypotheses

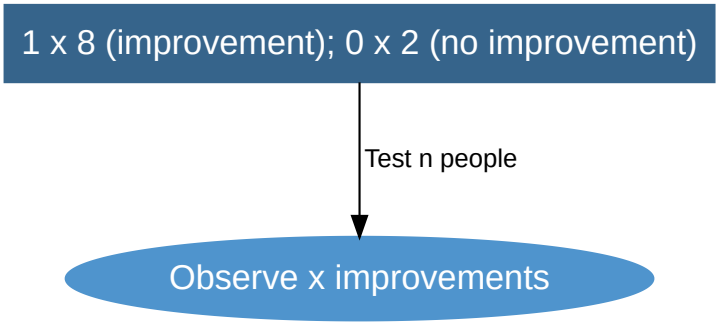
- The **null hypothesis** $H_0$ assumes that the difference between the observed value (data) and expected value (EV) is due to chance alone.
- The **alternative hypothesis** $H_1$ assumes that the difference between the observed value (data) and expected value (EV) is NOT due to chance alone.
- We set up 2 box models to represent $H_0$ and $H_1$.

### Example (2 sided alternative)

Research Question: Does the probiotic treatment work for 80% of patients?

- $H_0$: 80% of the patients respond to the treatment. [Or $H_0 : p = 0.8$, where $p$ is the proportion of patients who respond to treatment (desensitise to peanut allergy)].
- $H_1$: More or less than 80% of the patients respond to treatment ( desensitise to peanut allergy). [Or $H_1 : p \neq 0.8$]

## Null hypothesis $H_0$



Note:

- The Sum of the $n$ draws = the number of people who showed an improvement.
- For the box: $\text{mean} = 0.8$ and $\text{SD} = (1-0)\sqrt{0.8 \times 0.2} = 0.4$.
- For the Sum of the Sample (draws):
  - $\text{EV} = n \times 0.8$.
  - $\text{SE} = \sqrt{n} \times 0.4$

```
box = c(0, 0, 1, 1, 1, 1, 1, 1, 1, 1)
mean(box)
```

```
## [1] 0.8
```

```
library(multicon)
popsd(box)
```

```
## [1] 0.4
```

### Example (1 sided alternative)

Research Question: Does the probiotic treatment work for more than 80% of the patients?

- $H_0$ : The probiotic treatment works for 80% of the patients.
- $H_1$ : The probiotic treatment works for more than 80% of the patients.

# Assumptions

- A conclusion is not **transparent**, if the assumptions are not stated.
- A conclusion is potentially **invalid**, if the assumptions are not justified.

### Example

- We assume each child in the trial was independent of each other (eg not related; similar health profile).
- We assume each child had the same chance of showing improvement with their peanut allergy by using the probiotic treatment.
- We could check these assumptions by looking at the records of the medical trial.

# Test Statistic

## Test statistic

- A **test statistic** measures the difference between what is **observed** in the data and what is **expected** from the null hypothesis.
- It takes the form

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

Note: If the null hypothesis is true, then the test statistic is the standard unit corresponding to the observed value.

### 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 $\mathrm{EV} = np$ improvements with $\mathrm{SE} = \sqrt{np(1-p)}$.
- 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 $X$.

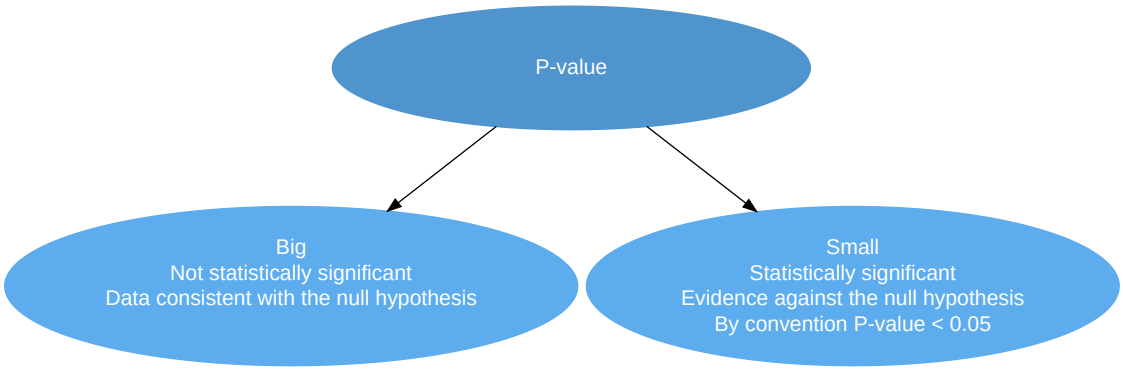
Note (Extension): $X$ has a Binomial distribution.

# P-value

## P-value (observed significance level)

- The **p-value** is a way of weighing up whether the sample is consistent with $H_0$.
- The **p-value** is the chance of observing the test statistic (or something more extreme) if $H_0$ really is true.

# Size of P-value



# Common mistakes with the P-value

**1.** The p-value is not the chance that the null hypothesis is true.

**2.** A large p-value does not mean that Ho is 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 |

**3.** The use of 0.05 is not mandatory. The significance level of 0.05 is a convention. Some people use 0.01 and say that the result is **highly significant**.

### Example

p-value = $P_{H_0}$(“X is at least as extreme as x”)

# Controversy about the P-value

Nature March 2016

# Summary

## Key Words

evidence-based decisions, hypothesis testing, null hypothesis, alternative hypothesis, test statistic, p-value, statistically significant, conclusion

# Further Thinking

## P-value

## P-value & tea