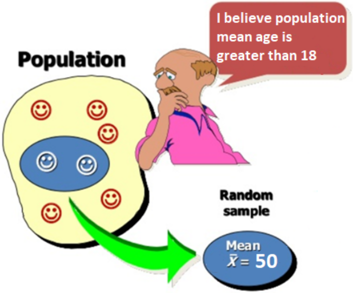
Hypothesis tests

Paula Moraga

Hypothesis tests

Hypothesis testing is a procedure to determine whether a claim about a population parameter is reasonable.

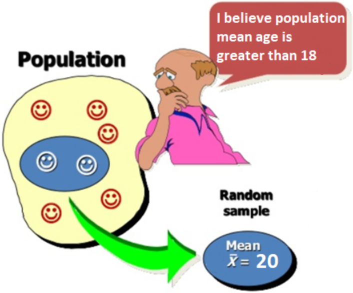
Currently accepted value of population mean age is 18.



Sample mean is 50, far from the currently accepted value 18.

So we reject the current belief that mean age is 18. The data provide evidence to say mean age is greater than 18.

Currently accepted value of population mean age is 18.



Sample mean is 20, close to the currently accepted value 18.

So we fail to reject the current belief that mean age is 18. The data do not provide enough evidence to say mean age is greater than 18.

Hypothesis tests

Hypothesis testing is a procedure to determine whether a claim about a population parameter is reasonable.

- Research question: Does a new drug reduce cholesterol?

The **alternative or research hypothesis** is the alternative to the currently accepted value for the population parameter

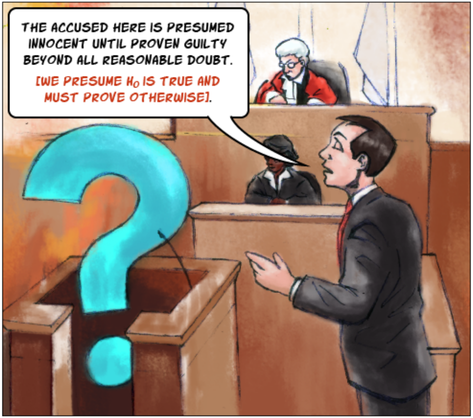
- H_1: The new drug reduces cholesterol

The **null hypothesis** represents "no difference" (currently accepted value of the population parameter)

- H_0: The new drug has no effect

To test a hypothesis, we select a sample from the population and use measurements from the sample and probability theory to determine how well the data and the null hypothesis agree.

We want to show the alternative hypothesis is reasonable (by showing there is evidence to reject the null hypothesis).



Example

- Research question we are trying to answer
  Does a new drug reduce cholesterol?
- Alternative hypothesis (investigator's belief)
  H_1: The new drug reduces cholesterol
- Null hypothesis (no difference, hypothesis to be rejected)
  H_0: The new drug has no effect

Let μ_1 and μ_2 be the cholesterol levels obtained with and without the drug, respectively.

$H_0 : \mu_1 = \mu_2$ (Null hypothesis)

$H_1 : \mu_1 < \mu_2$ (Alternative hypothesis)

Null and alternative hypothesis

Null Hypothesis H_0	Alternative Hypothesis H_1
Assumed true until evidence indicates otherwise	We try to find evidence for the alternative hypothesis
"Nothing is going on"	Opposite of the null hypothesis
There is no relationship between variables being studied	There is relationship between variables being studied
An intervention does not make a difference/has no effect	An intervention makes a difference/has an effect

Hypothesis are written in terms of population parameters.

- single mean (μ)
- single proportion (p)
- difference between two independent means ($\mu_1 - \mu_2$)
- difference between proportions ($p_1 - p_2$)

The null hypothesis H_0 contains the equal sign ($=$)

Write the null and alternative hypothesis

Is the average monthly rent of a one-bedroom apartment in Bath less than 800 pounds?

H_0: $\mu \geq 800$
H_1: $\mu < 800$

Do the majority of all university students own a dog?

$H_0 : p \leq 0.5$
$H_1 : p > 0.5$

In preschool, are the weights of boys and girls different?

H_0:
H_1:

In preschool, are the weights of boys and girls different?

$H_0 : \mu_b = \mu_g$
$H_1 : \mu_b \neq \mu_g$

Is the average intelligence quotient (IQ) score of all university students higher than 100?

H_0:
H_1:

Is the proportion of men who smoke cigarettes different from the proportion of women who smoke cigarettes in England?

H_0:
H_1:

Is the average intelligence quotient (IQ) score of all university students higher than 100?

H_0: $\mu \leq 100$
H_1: $\mu > 100$

Is the proportion of men who smoke cigarettes different from the proportion of women who smoke cigarettes in England?

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

Decisions in hypothesis testing

We make one of these two decisions:

- **Reject the null**
  (if data provide evidence to disprove it)
- **Fail to reject the null**
  (if data do not provide enough evidence to disprove it)
- Failing to reject the null does not mean we accept the null. It is possible the data we have do not provide enough evidence against the null. But maybe we can collect additional data that provide evidence against the null.

A man goes to trial where he is being tried for a murder.

H_0: Man is not guilty
H_1: Man is guilty

It is possible the data we have do not provide enough evidence to reject the null and conclude the man is guilty. But we do not say he is innocent! We may collect additional data that provide evidence and then we can reject the null and say the man is guilty.

Errors in hypothesis testing

	H_0 is True	H_0 is False
Reject null	Type I error	Correct
Fail to reject null	Correct	Type II error

Type I error: Reject the null hypothesis when it is true

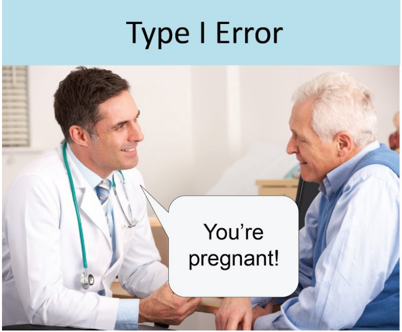
Type II error: Fail to reject the null hypothesis when it is false

α = Probability of Type I error = P(rejecting H_0 | H_0 true)

β = Probability of Type II error = P(failing to reject H_0 | H_0 false)

Example

H_0: patient is not pregnant
H_1: patient is pregnant



- Type I error: Reject the null hypothesis when it is true. Patient is not pregnant but we conclude he is pregnant.
- Type II error: Fail to reject the null hypothesis when it is false. Patient is pregnant but we conclude she is not pregnant.

Example

A man goes to trial where he is being tried for a murder.

H_0: Man is not guilty
H_1: Man is guilty

- Type I error: Reject the null hypothesis when it is true.
- Type II error: Fail to reject the null hypothesis when it is false.

Example

A man goes to trial where he is being tried for a murder.

H_0: Man is not guilty
H_1: Man is guilty

- Type I error: Reject the null hypothesis when it is true.
  The man did not kill the person but was found guilty and was punished for a crime he did not really commit.
- Type II error: Fail to reject the null hypothesis when it is false.
  The man killed the person but was found not guilty and was not punished.

Significance level

- Before testing the hypothesis, we select the significance level of the test α.
- The **significance level** is the maximum probability of incorrectly rejecting the null when it is true we are willing to tolerate. Typical α values are 0.05 and 0.01.
  $\alpha = \text{P(Type I Error)} = \text{P(reject } H_0 \mid H_0 \text{ is true)}$
- If we choose $\alpha = 0.05$, we have a maximum of 5% chance of incorrectly rejecting the null when it is true.

Steps in hypothesis testing

1. State the null and alternative hypotheses
2. Choose the significance level. Usually $\alpha = 0.05$
3. Calculate the test statistic using the sample observations

$$\text{test statistic} = \frac{\text{sample statistic - null parameter}}{\text{standard error}}$$

4. Find the p-value for the observed data
   (p-value is the probability of obtaining a test statistic as extreme as or more extreme than the one observed in the direction of the alternative, assuming the null is true)
5. Make a decision and state a conclusion
   If p-value $< \alpha$, we reject the null
   If p-value $\geq \alpha$, we fail to reject the null

Example

A company manufacturing computer chips claims the defective rate is 5%. Let p denote the true defective probability. We use a sample of 1000 chips from the production to determine their claim is reasonable. The proportion of defectives in the sample is 8%.

1. State the null and alternative hypotheses

$H_0 : p = p_0$ (proportion of defective chips is 5%)
$H_1 : p > p_0$ (proportion of defective chips is greater than 5%)
where $p_0 = 0.05$

2. Choose the significance level

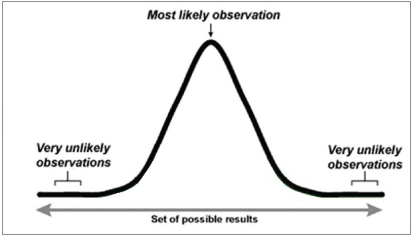
We use a significance level $\alpha = 0.05$
So the maximum probability of incorrectly rejecting the null when it is true we are willing to tolerate is 5%.

3. Calculate the test statistic

$$\text{test statistic} = \frac{\text{sample statistic - null parameter}}{\text{standard error}}$$

The test statistic is a value calculated from a sample that summarizes the characteristics of the sample and is used to determine whether to reject or fail to reject the null hypothesis.

The test statistic differs from sample to sample. The sampling distribution of the test statistic under the null must be known so we can compare the results observed to the results expected from the null (e.g., test statistics with normal or t distributions).

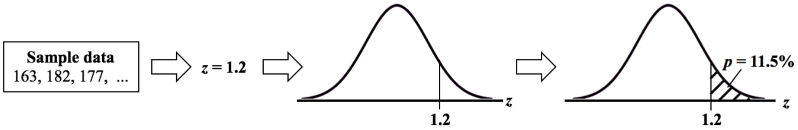


4. Find the p-value

We reject the null if the observed value of the test statistic is so extreme that it is unlikely to occur if the null is true (unlikely to occur is given by the significance level α). Otherwise we fail to reject the null.

The p-value is the probability of obtaining a test statistic as extreme as or more extreme than the one observed in the direction of the alternative hypothesis, assuming the null hypothesis is true.

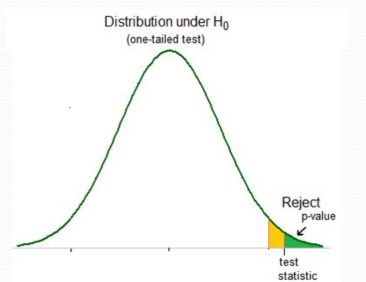
To calculate the p-value we compute the test statistic value using the sample data, and calculate the area under the test statistic distribution curve beyond the test statistic value.



5. Compare p-value and α and make a decision

We compare the p-value to the significance level α and decide whether to reject or fail to reject the null.

- If p-value is small (p-value $< \alpha$) we had little chance of getting our data if the null hypothesis is true. So we reject the null. We conclude there is evidence against the null hypothesis.
- If p-value is big (p-value $\geq \alpha$) there is reasonable chance of getting our data if the null hypothesis is true. So we fail to reject the null. We conclude there is not sufficient evidence against the null hypothesis.



1. State the null and alternative hypotheses

$H_0 : p = p_0$ (proportion of defective chips is 5%)
$H_1 : p > p_0$ (proportion of defective chips is greater than 5%)
where $p_0 = 0.05$

2. Choose the significance level

We use a significance level $\alpha = 0.05$
So the maximum probability of incorrectly rejecting the null when it is true we are willing to tolerate is 5%.

3. Calculate the test statistic

Sample statistic is proportion of defective chips in the sample ($\hat{P}$).
Test statistic

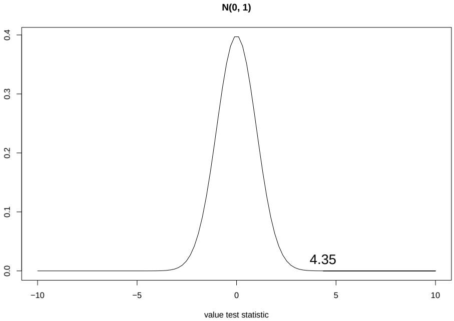
$$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$$

The test statistic Z follows a N(0, 1) distribution under the null hypothesis (this holds if we have independent observations and the number of successes and failures are both greater than 10).
$z_{obs} = \frac{0.08-0.05}{\sqrt{0.05(1-0.05)/1000}} = 4.35.$

4. Find the p-value

p-value is the area under the N(0,1) curve beyond the test statistic observed in the direction of the alternative hypothesis

p-value $= P(Z > 4.35) \approx 0$



5. Make a decision and state a conclusion

p-value is approximately 0 and $\alpha = 0.05$

p-value $< \alpha$ so we reject the null, data provide evidence to say the population proportion of defective chips is greater than 5%.

One-sided and two-sided tests

One-sided (or one-tailed) test

The alternative hypothesis specifies that the population parameter lies entirely above or below the value specified in H_0

$H_0 : \mu = 100$
$H_1 : \mu > 100$

This is tested in the same way as

$H_0 : \mu \leq 100$
$H_1 : \mu > 100$

(if we conclude $\mu > 100$, we also conclude $\mu > 90$, $\mu > 80$, etc)

Two-sided (or two-tailed) test

The alternative hypothesis specifies that the parameter can lie on either side of the value specified by H_0

$H_0 : \mu = 100$
$H_1 : \mu \neq 100$

p-values in one- and two- tailed tests

- Left-tailed tests: p-value is the area under the test statistic distribution curve to the left of the test statistic
- Right-tailed tests: p-value is the area under the test statistic distribution curve to the right of the test statistic
- Two-tailed tests: we use the symmetry of the test distribution curve. Calculate the p-value for a one-sided test and double it