One and Two Sample Tests

Evan Misshula

2017-10-20

Outline

1 Learning objectives

A test by many names

- two functions we introduce
  - `t.test`
  - `wilcox.text`

A test by many names

- two functions we introduce
    - `t.test`
    - `wilcox.text`
- can be used with either

A test by many names

- two functions we introduce
  - `t.test`
  - `wilcox.text`
- can be used with either
- Note the "two sample Wilcox test" is also called the "Mann Whitney test"

one sample t-test

- t-test is based on the idea that the data $x_1, ..., x_n$
  - are independant
  - identically distributed

one sample t-test

- t-test is based on the idea that the data $x_1, ..., x_n$
  - are independant
  - identically distributed
- denoted *iid*

- t-test is based on the idea that the data $x_1, \ldots, x_n$
  - are independant
  - identically distributed
- denoted *iid*
- from a distribution that is Gaussian with mean
  - mean μ
  - standard deviation σ

one sample t-test

- t-test is based on the idea that the data $x_1, ..., x_n$
  - are independant
  - identically distributed
- denoted *iid*
- from a distribution that is Gaussian with mean
  - mean μ
  - standard deviation σ
- $N(\mu, \sigma)$

A tale of two hypotheses

- The *null* hypotheses H_o is the one we think is not true
- The *research* hypothesis H_1 is the one we think may be true

A tale of two hypotheses

- The *null* hypotheses H_o is the one we think is not true
- The *research* hypothesis H_1 is the one we think may be true
- We can only reject the *null*

In the one sample t-test

- we are testing that $\mu = \mu_0$

Key Conept is the standard error of the mean

- The *standard error of the mean* is the variation of the mean of the average of n numbers with mean μ and standard deviation σ

$$SEM = \frac{\sigma}{\sqrt{n}}$$

- That variation goes to 0 as n goes to infinity

But we don't know the real parameters of the distribution

- How would we find the mean, μ?

But we don't know the real parameters of the distribution

- How would we find the mean, μ?

take the sample mean,

$\bar{x} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} x_i$

These two facts allows us to create a test statistic

These two facts allows us to create a test statistic

that should be normally distributed:

$$t = \frac{\bar{x} - \mu}{SEM} \sim N(0, 1)$$

These two facts allows us to create a test statistic

that should be normally distributed:

$$t = \frac{\bar{x} - \mu}{SEM} \sim N(0,1)$$

So if I have guessed right:

t should be between -2 and 2

$\mu \pm 2\sigma < 2$ about 95% of the time

Two sample T-test

The data are in two groups

- $x_{1\ 1}, \ldots, x_{1\ n_1} \sim N(\mu_{1,\sigma_1^2})$
- $x_{2\ 1}, \ldots, x_{2\ n_2} \sim N(\mu_{2,\sigma_2^2})$

So the null is hypthesis is

$0 = \mu_1 - \mu_2$

and the test statistic

$$t = \frac{\bar{x_1} - \bar{x_2}}{SEDM}$$

SEDM is the

standard difference of the means

$$SEDM = \sqrt{SEM_1^2 + SEM_2^2}$$