



---

# Hypothesis Testing

*Sarangan Ravichandran*

# Hypothesis Testing



## Overview

- Setup a hypothesis that can be falsifiable
- Test the hypothesis

- Example

- If we come across a sample of 49 total cholesterol levels with an average of 244 mg/100 ml. Can we say whether this sample could have come from the population?

$$\mu_X = 237 \text{ mg/100ml}$$
$$\sigma_X = 44.7 \text{ mg/100ml}$$

- What about 250 mg/100 ml?
  - Can we formally test these values?



## Let us formalize the procedure

**Procedure:** Compares one or more population parameters (proportions). We do this by selecting a sample from the population.

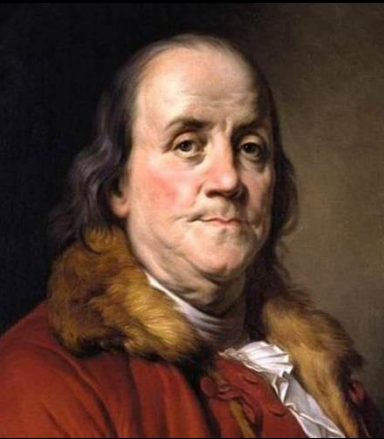
# Hypothesis Testing



- Choose a test (one sample Z test; t-test)
- Null Hypothesis
  - $H_0: \mu_0 = 237$  (cholesterol level)
- Alternate Hypothesis
  - $H_A: \mu_0 \neq 237$  (two-sided; decided up-front)
- Gather/Analyze Data (**random sample**)
- Decide on a cut-off (alpha) = 95% (arbitrary; decided up-front)
- Note the experiment is carried out assuming that the Null Hypothesis is true. Draw a sample and test to see how extreme the value is and use it to **reject the null hypothesis** or **fail to reject the null hypothesis**

$$\mu_x = 237 \text{ mg/100ml}$$
$$\sigma_x = 44.7 \text{ mg/100ml}$$

# Hypothesis Testing & US Judicial System



"that it is better [one hundred] guilty Persons should escape than that one innocent Person should suffer."

Jury	Person	
	Innocent	Guilty
Not Guilty	✓	✗
Guilty	✗	✓

# Hypothesis Testing & US Judicial System



US	Population	
	$\mu = \mu_0$	$\mu \neq \mu_0$
Fail to Reject	✓	✗
Reject	✗	✓

# Hypothesis Testing & US Judicial System



US	Population	
	$\mu = \mu_0$	$\mu \neq \mu_0$
Fail to Reject	✓	Type-II
Reject	Type-I	✓

# Hypothesis Testing & US Judicial System



US	Population	
	$\mu = \mu_0$	$\mu \neq \mu_0$
Fail to Reject	✓	Type-II = $\beta$
Reject	Type-I = $\alpha$	✓

$1 - \beta = \text{Power}$

Goal is to make  $\alpha$  and  $\beta$  smaller



## Error type summary

- $P(\text{rejecting } H_0 \mid H_0 \text{ is correct}) = \alpha$
- $P(\text{not rejecting } H_0 \mid H_A \text{ is correct}) = \beta$



# Hypothesis Testing by Example



A large population, study identified the mean Intraocular pressure (IOP) to be 15.5 mm Hg ([www.aaio.org](http://www.aaio.org)) with SD ( $\sigma = 2.6$  mm Hg).

If we come across a sample mean of **16.5 mm Hg** ( $n = 49$ ), can we say whether this could have come from the above population?

Let us set up the problem?

- Two-sided Hypothesis (interested in both direction)
- One sample Z-test (CLT)
- Alpha of 0.05

# Hypothesis testing by example



–  $H_0: \mu_0 = 15.5 \text{ mm Hg}$

–  $H_A: \mu_0 \neq 15.5 \text{ mm Hg}$

**Sample**

$\bar{X} = 16.5 \text{ mm Hg}$

$n = 49$

**Population**

$\mu_0 = 15.5 \text{ mm Hg}$

$\sigma = 2.6 \text{ mm Hg}$

– Establish what value(s) we will accept to be different from the NULL distribution ( $\alpha \text{ level} = 0.05$ )

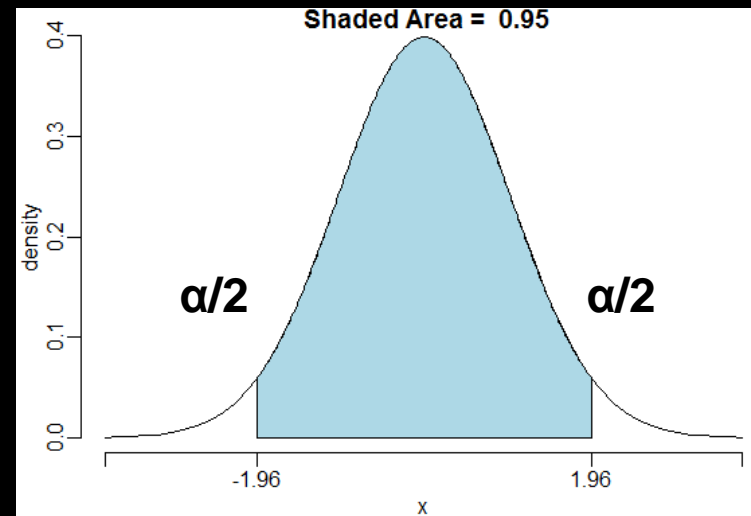
– Assuming CLT, we can perform an one sample Z-test

# Hypothesis testing by example



**Alpha values lead us  
to Critical Values**

**Values that represent Z  
values that correspond  
to  $\alpha = \pm 2.5\%$  are called  
critical values ( $\pm 1.96$ )**

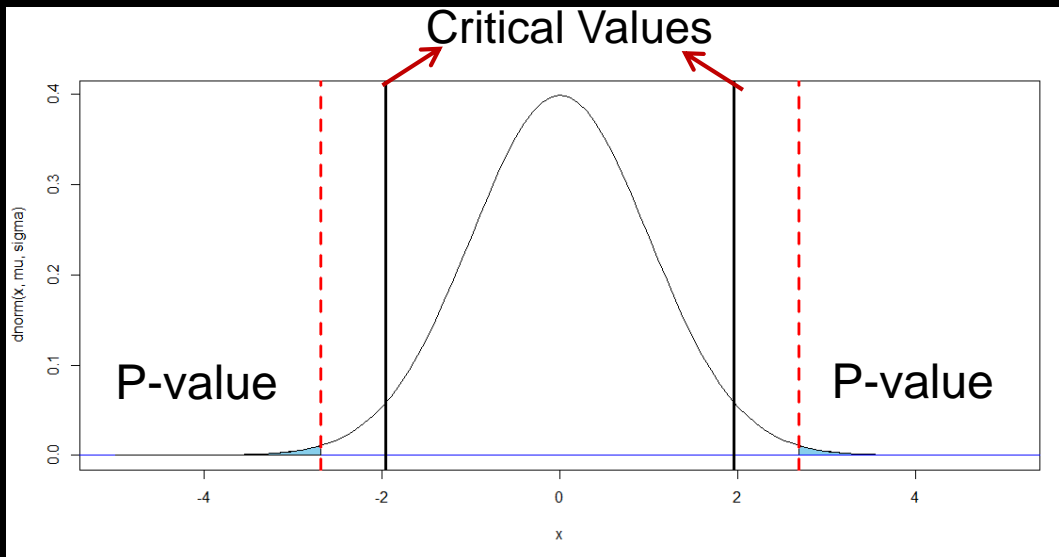


# Hypothesis testing by example



Z of 2.69 or GREATER) = 0.36%  
Z of 2.69 or LOWER) = 0.36%

$$Z = \frac{16.5 - 15.5}{\frac{2.6}{\sqrt{49}}} = 2.69$$



P-value of the 2-tailed  
Z-test: is 0.0072

What is P-value?

## P-value Definition



- Given a  $H_0$ ,  $H_A$  and a Test Statistic  $T$ , the p-value can be defined as

*“the probability, computed assuming that  $H_0$  is true, that the test statistic would take a value as extreme or more extreme than that actually observed”*

*Moore, D.S. (2007) The Basic Practice of Statistics*

## P-value Definition (using Math)



- Given a  $H_0$ ,  $H_A$  and a Test Statistic  $T$ , the p-value can be defined as

$$p = P(T \geq t_{obs} | H_0)$$

Duncan Murdoch, Univ, of Western Ontario

- What is extreme?
- Larger values of  $T$

# Consequence of P-values & beyond



- P-values
  - are random variables; probabilities (0 to 1)
  - just show how compatible/incompatible your data with the assumed statistical model
  - does not measure size effect
  - Give no indication to the clinical important of an obs. Effect
  - Using P-values to group significant/non-significant is a poor idea
- Obsession of low p-values have led us into selective reporting/publishing
  - Inflation bias (p-hacking)



# P-Hacking



- If I pulled out a coin and flipped 10 coins in a row, will you be surprised?
- Let us set this up in R
- Now what! Can we write a paper!

```
set.seed(6107)
```

```
rbinom(n=10, size = 1, prob = 0.5)  
# [1] 1 1 1 1 1 1 1 1 1 1
```

```
# Probability  
0.5^10  
#[1] 0.000976562
```

# P-Hacking



- How could this have happened?
  - A fishing expedition
- R code uses 1000 different random seeds
- Flip the coin until we find what we like (10 H)
- Take the seed and run it to surprise someone

```
set.seed(9382)
rSeeds <- round(runif(1000) * 10000 )
for(s in rSeeds)
{
  set.seed(s)
  heads <- sum( rbinom( n=10,
                        size=1, prob=0.5) )
  if(heads == 10)
    break
}
s
## [1] 6107
```

# What other options do we have?



- 95% Confidence Interval

$$\Pr\left\{-1.96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq +1.96\right\} = 0.95$$
$$\Pr\left\{\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right\} = 0.95$$
$$\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

Assume  $n$  large enough. Variable collection in the middle is the one that varies.

$n$ ,  $\sigma$  positive. Let us cross-multiply, multiply by  $-1$ ; reverse inequalities; We have created two quantities that vary without the unknown  $\mu$

We have isolated something that varies from sample to sample on both sides

Let us use this in our previous example



$$\left[ \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right]$$

**Population**

$\mu_0 = 15.5$  mm Hg  
 $\sigma = 2.6$  mm Hg

$$\begin{aligned} &[14.5 - 0.728, 14.5 + 0.728] \\ &[13.772, 15.228] \end{aligned}$$

**n = 49**

**Sample1**

$\bar{X} = 14.5$  mm Hg

**Misses  $\mu_0$**

$$[14.872, 16.328]$$

**Sample2**

$\bar{X} = 15.6$  mm Hg

**Includes  $\mu_0$**

$$[17.272, 18.728]$$

**Sample3**

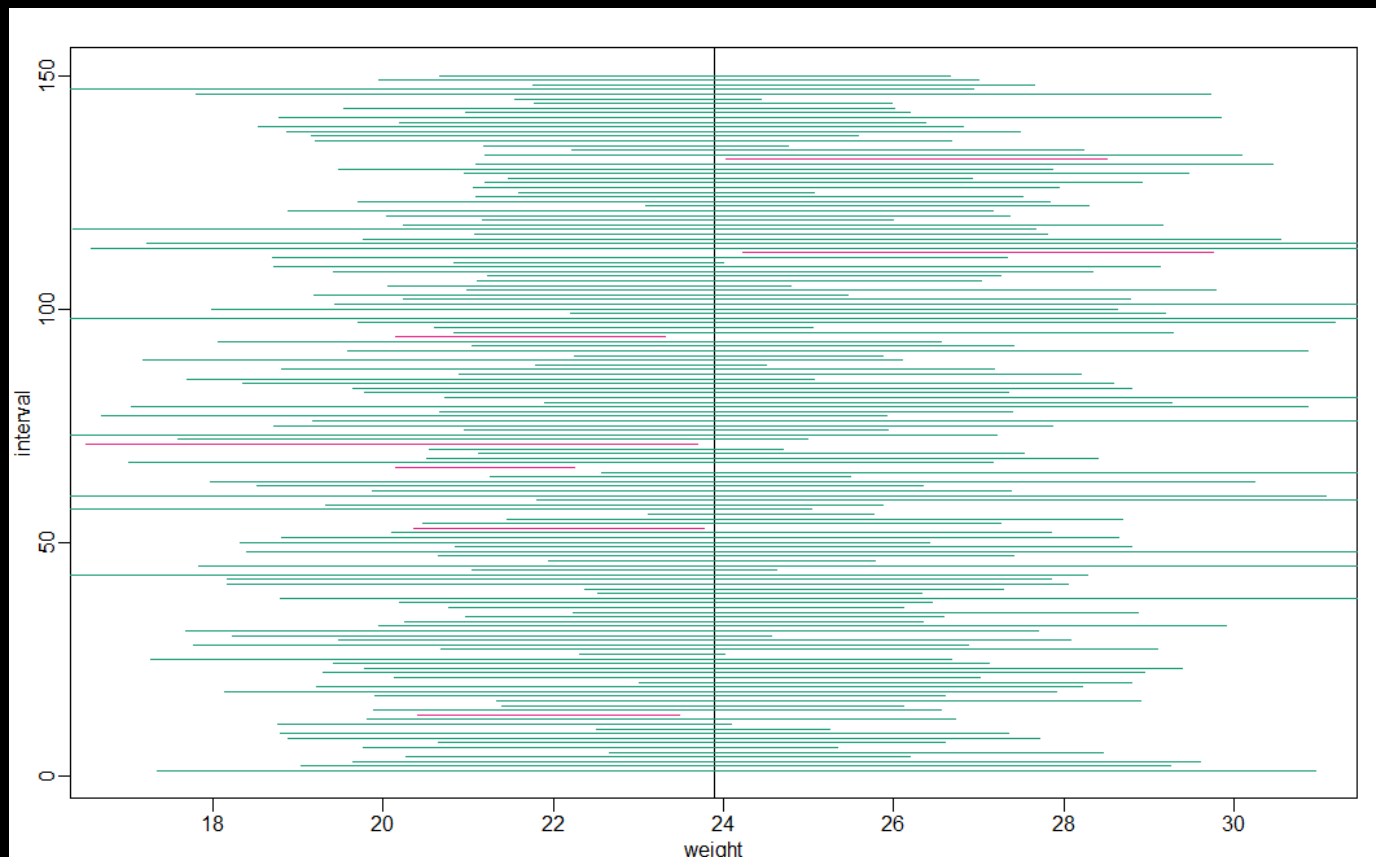
$\bar{X} = 18.0$  mm Hg

**Misses  $\mu_0$**

# Confidence Interval



$\mu = 23.89$  ;  $\sigma = 2.42$ ; Number of samples = 150; Sample size = 5  
95% CI: red that doesn't include  $\mu$



Code from Rafael. A. Irizarry and Michael I. Love

# Hypothesis Testing: From Z to t



- In reality, we use t-test and not the Z-test which has been shown to be well behaved for small sample sizes
- We have used population standard deviation, which we do not have access to in almost all cases
  - If we have reasonable sample size, we can replace sample standard deviation instead of population standard deviation

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

# What if sigma is unknown?



$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

$$t = \frac{\bar{X} - \mu}{sd / \sqrt{n}}$$

Just as the Mean of the sample means is the same as the population mean, so too is the mean of the sample variance is equal to the population variance

Unlike Z, t has a distribution that depends on the sample size  
Precisely, n-1 degrees of freedom

Assumption

Population is approximately normal