Inferential Statistics

Examples are taken from edx course on Statistics 2.3x

Introduction to Data Science

M. Sakthi Balan

Associate Professor
LNM Institute of Information Technology
Jaipur

Contents

Introduction

Important terms

Population: A collection of same units that we are interested in
Parameter: A number we are interested in about the population
Sample: A subset of population
Estimate: A guess for parameter calculated from the sample

Estimate will be **good** if the sample is **good**
Sample should avoid **Bias**

Introduction

Important terms

Population: A collection of same units that we are interested in
Parameter: A number we are interested in about the population
Sample: A subset of population
Estimate: A guess for parameter calculated from the sample

Estimate will be **good** if the sample is **good**
Sample should avoid **Bias**

Introduction

Does the sample size matter? – large sample size versus small sample size

US Election 1936

- Roosevelt versus Landon
- Literary Digest predicted Landon (Republican) win and predicted Democratic will get only 43%
- Literary Digest used a sample size of 10 million (2.4 million responded)
- Gallup used a good scientific method and predicted that Literary Digest will predict 44%!! They used a sample size of just 3000!!
- Gallup also predicted that Roosevelt will win with 56% votes
- Gallup used a sample size of 50000 only!!
- Roosevelt won the election with 61% votes

Introduction

- Selection Bias – Literary Digest used phone book!
- Non-response Bias – only 2.4 million responded out of 10 million!
- Sample should not be a sample of convenience
- Should draw uniformly at random with or without replacement from the population

Simple Random Sample – Random sample with all units in population equally likely in the sample before selection
Random Sample – all units need not be equally likely

Introduction

- Selection Bias – Literary Digest used phone book!
- Non-response Bias – only 2.4 million responded out of 10 million!
- Sample should not be a sample of convenience
- Should draw uniformly at random with or without replacement from the population

Simple Random Sample – Random sample with all units in population equally likely in the sample before selection
Random Sample – all units need not be equally likely

Introduction

Problem

The population consists of only four people: A, B, C and D. A is chosen in the sample and other three are chosen only if we get heads when we toss a fair coin. Is this a good sample?

Estimation and Standard Error

Problem

In a large population, average age = 37 and standard deviation = 15. A sample size of 200 is taken from the population. Average age will be what?

Expected age = 37 and standard error = $15/\sqrt{200} = 1.06$

Estimation and Standard Error

Problem

In a large population, average age = 37 and standard deviation = 15. A sample size of 200 is taken from the population. Average age will be what?

Expected age = 37 and standard error = $15/\sqrt{200}$ = 1.06

Estimation and Standard Error

Problem

For a random sample size of 100, average age = 38 and standard deviation = 14. What will be the average age of the population?

Estimated age = 38 and standard error = $14/\sqrt{100} = 1.4$

Bootstrap

A considerable large sample size is like a population, so SD of the sample is an approximation to the SD of the population.

Estimation and Standard Error

Problem

For a random sample size of 100, average age = 38 and standard deviation = 14. What will be the average age of the population?

Estimated age = 38 and standard error = $14/\sqrt{100} = 1.4$

Bootstrap

A considerable large sample size is like a population, so SD of the sample is an approximation to the SD of the population.

Estimation and Standard Error

Problem

For a random sample size of 100, average age = 38 and standard deviation = 14. What will be the average age of the population?

Estimated age = 38 and standard error = $14/\sqrt{100} = 1.4$

Bootstrap

A considerable large sample size is like a population, so SD of the sample is an approximation to the SD of the population.

Confidence Intervals

Problem

In a simple random sample of 605 households in a city the average income is \$63000 with SD as \$40000. Find an approximate 95% confidence interval for the average income of huseholds in the city.

$63000 \pm 2 * \frac{40000}{\sqrt{625}}$
$= 63000 \pm 2 * 1600$
$\implies$ 59800 to 66200 (Confidence Interval)

Given distribution is not normal but how did we use approximation for a normal distribution?

Confidence Intervals

Problem

In a simple random sample of 605 households in a city the average income is $63000 with SD as $40000. Find an approximate 95% confidence interval for the average income of husholds in the city.

$63000 \pm 2 * \frac{40000}{\sqrt{625}}$
$= 63000 \pm 2 * 1600$
$\implies$ 59800 to 66200 (Confidence Interval)

Given distribution is not normal but how did we use approximation for a normal distribution?

Confidence Intervals

Problem

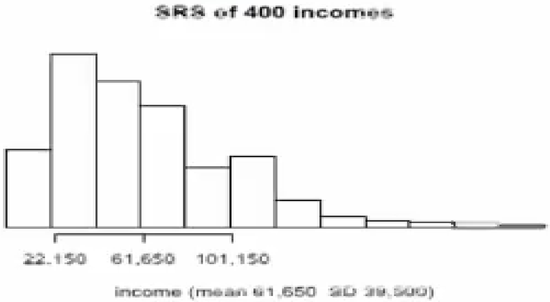
In a simple random sample of 605 households in a city the average income is $63000 with SD as $40000. Find an approximate 95% confidence interval for the average income of huseholds in the city.

$63000 \pm 2 * \frac{40000}{\sqrt{625}}$
$= 63000 \pm 2 * 1600$
$\implies$ 59800 to 66200 (Confidence Interval)
Given distribution is not normal but how did we use approximation for a normal distribution?

LNMIIT

Confidence Intervals

A sample distribution:

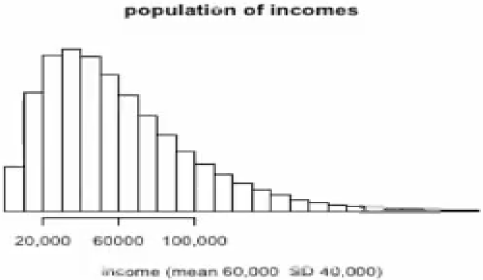


- sample mean $61,650
- sample SD $39,500

from edx course on Statistics

Confidence Intervals

Original distribution that we may not have:

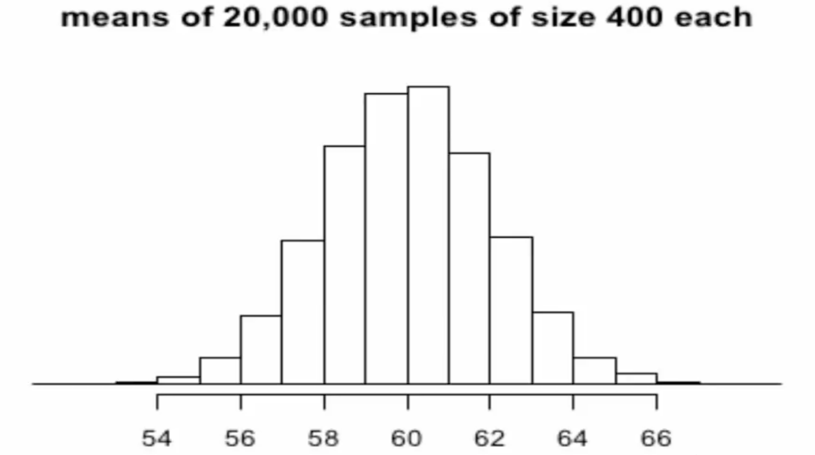


- population mean $60,000
- population SD $40,000

from edx course on Statistics

Confidence Intervals

After many trials:



from edx course on Statistics

Testing Hypothesis

- Suppose you toss a coin 15 times and observe that you get 10 times head. Is it a fair coin or not?
- Coin is **fair**: H_0 – called as **Null hypothesis**
- Coin is **not fair** and it is **biased towards heads**: H_A – called as **Alternate hypothesis**
- Assume that H_0 is true and calculate the chance of getting H_A, i.e., the above observation of 10 times head out of 15 – called as **P-value**.

Method

If $P - value < 5\%$ (small) then coose H_A
We call the above situation as **Statistically Significant**
Otherwise ($P - value \geq 5\%$) no change in our assumption that H_0 is true

Testing Hypothesis

- Suppose you toss a coin 15 times and observe that you get 10 times head. Is it a fair coin or not?
- Coin is **fair**: H_0 – called as **Null hypothesis**
- Coin is **not fair** and it is **biased towards heads**: H_A – called as **Alternate hypothesis**
- Assume that H_0 is true and calculate the chance of getting H_A, i.e., the above observation of 10 times head out of 15 – called as **P-value**.

Method

If $P - value < 5\%$ (small) then coose H_A
We call the above situation as **Statistically Significant**
Otherwise ($P - value \geq 5\%$) no change in our assumption that H_0 is true

Testing Hypothesis

- Suppose you toss a coin 15 times and observe that you get 10 times head. Is it a fair coin or not?
- Coin is **fair**: H_0 – called as **Null hypothesis**
- Coin is **not fair** and it is **biased towards heads**: H_A – called as **Alternate hypothesis**
- Assume that H_0 is true and calculate the chance of getting H_A, i.e., the above observation of 10 times head out of 15 – called as **P-value**.

Method

If $P - value < 5\%$ (small) then coose H_A
We call the above situation as **Statistically Significant**
Otherwise ($P - value \geq 5\%$) no change in our assumption that H_0 is true

Testing Hypothesis

- Suppose you toss a coin 15 times and observe that you get 10 times head. Is it a fair coin or not?
- Coin is **fair**: H_0 – called as **Null hypothesis**
- Coin is **not fair** and it is **biased towards heads**: H_A – called as **Alternate hypothesis**
- Assume that H_0 is true and calculate the chance of getting H_A, i.e., the above observation of 10 times head out of 15 – called as **P-value**.

Method

If $P - value < 5\%$ (small) then coose H_A
We call the above situation as **Statistically Significant**
Otherwise ($P - value \geq 5\%$) no change in our assumption that H_0 is true

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Is theory good or bad?

Solution

- $H_0 : p = 0.25$ and $H_A : p < 0.25$
- P-value: binomial with $p = 0.25$ and $k \leq 88 \implies P = 9.08\%$
- $P > 5\%$ hence the theory looks fine!

This is called as **exact binomial test**

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Is theory good or bad?

Solution

- $H_0 : p = 0.25$ and $H_A : p < 0.25$
- P-value: binomial with $p = 0.25$ and $k \leq 88 \implies P = 9.08\%$
- $P > 5\%$ hence the theory looks fine!

This is called as **exact binomial test**

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Is theory good or bad?

Solution

- $H_0 : p = 0.25$ and $H_A : p < 0.25$
- P-value: binomial with $p = 0.25$ and $k \leq 88 \implies P = 9.08\%$
- $P > 5\%$ hence the theory looks fine!

This is called as **exact binomial test**

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Solution

- The binomial distribution is approximately a normal distribution according to **central limit theorem**
- Calculate μ and σ
- $\mu = 400 * 0.25 = 100$
- $\sigma = \sqrt{400 * 0.25 * 0.75} = 8.66$
- Calculate $z - value$

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Solution

- The binomial distribution is approximately a normal distribution according to central limit theorem
- Calculate μ and σ
- $\mu = 400 * 0.25 = 100$
- $\sigma = \sqrt{400 * 0.25 * 0.75} = 8.66$
- Calculate $z - value$

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Solution

- The binomial distribution is approximately a normal distribution according to **central limit theorem**
- Calculate μ and σ
- $\mu = 400 * 0.25 = 100$
- $\sigma = \sqrt{400 * 0.25 * 0.75} = 8.66$
- Calculate $z - value$

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Solution

- The binomial distribution is approximately a normal distribution according to **central limit theorem**
- Calculate μ and σ
- $\mu = 400 * 0.25 = 100$
- $\sigma = \sqrt{400 * 0.25 * 0.75} = 8.66$
- Calculate $z - value$

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Solution

- $z = \frac{88.5-100}{8.66} = -1.328$
- Get the area under the normal curve
- $P = 9.21\%$ (approx) $> 5\% \implies$ theory looks good!
- This is called as **one-sample z test**

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Solution

- $z = \frac{88.5 - 100}{8.66} = -1.328$
- Get the area under the normal curve
- $P = 9.21\%$ (approx) $> 5\% \implies$ theory looks good!
- This is called as **one-sample z test**

Testing Hypothesis – Example 1

Problem

- Genetic theory says that each plant has 25% chance of being red-flowering
- Data: 400 plants and 88 of them are red-flowering

Solution

- $z = \frac{88.5 - 100}{8.66} = -1.328$
- Get the area under the normal curve
- $P = 9.21\%$ (approx) $> 5\% \implies$ theory looks good!
- This is called as **one-sample z test**

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Selection of 100 men is like a trial conducted 100 times with selecting black as 1 and selecting whites as 0
- This distribution is approximately a normal distribution with μ as mean and σ as SD. Calculate them!
- $\mu = 100 * 0.26 = 26$
- $\sigma = \sqrt{100 * 0.26 * 0.74} = 4.39$
- Calculate the z-value

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Selection of 100 men is like a trial conducted 100 times with selecting black as 1 and selecting whites as 0
- This distribution is approximately a normal distribution with μ as mean and σ as SD. Calculate them!
- $\mu = 100 * 0.26 = 26$
- $\sigma = \sqrt{100 * 0.26 * 0.74} = 4.39$
- Calculate the z-value

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Selection of 100 men is like a trial conducted 100 times with selecting black as 1 and selecting whites as 0
- This distribution is approximately a normal distribution with μ as mean and σ as SD. Calculate them!
- $\mu = 100 * 0.26 = 26$
- $\sigma = \sqrt{100 * 0.26 * 0.74} = 4.39$
- Calculate the z-value

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Selection of 100 men is like a trial conducted 100 times with selecting black as 1 and selecting whites as 0
- This distribution is approximately a normal distribution with μ as mean and σ as SD. Calculate them!
- $\mu = 100 * 0.26 = 26$
- $\sigma = \sqrt{100 * 0.26 * 0.74} = 4.39$
- Calculate the z-value

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Calculate the z-value
- $z = \frac{8.5-26}{4.39} = -3.99$ (approx)

z-value is almost 4 standard errors away from the mean! What do you think?!

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Calculate the z-value
- $z = \frac{8.5-26}{4.39} = -3.99$ (approx)

z-value is almost 4 standard errors away from the mean! What do you think?!

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Calculate the z-value
- $z = \frac{8.5 - 26}{4.39} = -3.99$ (approx)

z-value is almost 4 standard errors away from the mean! What do you think?!

Testing Hypothesis – Example 2

Problem

- There are 16000 jurors and 26% of them are blacks
- Data: 100 men on jury and 8 of them are blacks

Question: Is this fair?

Solution

- Calculate the z-value
- $z = \frac{8.5-26}{4.39} = -3.99$ (approx)

z-value is almost 4 standard errors away from the mean! What do you think?!

Testing Hypothesis – Example 3

Problem

- Coin is tossed 20 times – $H_0 : p = 0.5$ and $H_A : p = 0.8$
- If the number of head is more than 14 then select H_A, otherwise select H_0

Possible Scenario

Reality/Observation	$p = 0.5$	$p = 0.8$
$p = 0.5$	Correct	Type I error Probability: 5.8%
$p = 0.8$	Type II error	Correct Probability: 91.3%

LNMIIT

Testing Hypothesis – Example 3

Problem

- Coin is tossed 20 times – $H_0 : p = 0.5$ and $H_A : p = 0.8$
- If the number of head is more than 14 then select H_A, otherwise select H_0

- Calculate probability for Type I error
- Its binomial with $n = 20, p = 0.5$ and $k \geq 14!$ $probability = 5.8\%$ – called as **Significance level**.
- Calculate probability for Type II error
- Its binomial with $n = 20, p = 0.8$ and $k \geq 14!$ $probability = 91.3\%$ – called as **Power**.

Testing Hypothesis – Example 3

Problem

- Coin is tossed 20 times – $H_0 : p = 0.5$ and $H_A : p = 0.8$
- If the number of head is more than 14 then select H_A, otherwise select H_0

- Calculate probability for Type I error
- Its binomial with $n = 20, p = 0.5$ and $k \geq 14$! $probability = 5.8\%$ – called as **Significance level**.
- Calculate probability for Type II error
- Its binomial with $n = 20, p = 0.8$ and $k \geq 14$! $probability = 91.3\%$ – called as **Power**.

Testing Hypothesis – Example 3

Problem

- Coin is tossed 20 times – $H_0 : p = 0.5$ and $H_A : p = 0.8$
- If the number of head is more than 14 then select H_A, otherwise select H_0

- Calculate probability for Type I error
- Its binomial with $n = 20, p = 0.5$ and $k \geq 14!$ $probability = 5.8\%$ – called as **Significance level**.
- Calculate probability for Type II error
- Its binomial with $n = 20, p = 0.8$ and $k \geq 14!$ $probability = 91.3\%$ – called as **Power**.

Testing Hypothesis – Example 3

Problem

- Coin is tossed 20 times – $H_0 : p = 0.5$ and $H_A : p = 0.8$
- If the number of head is more than 14 then select H_A, otherwise select H_0

- Calculate probability for Type I error
- Its binomial with $n = 20, p = 0.5$ and $k \geq 14!$ $probability = 5.8\%$ – called as **Significance level**.
- Calculate probability for Type II error
- Its binomial with $n = 20, p = 0.8$ and $k \geq 14!$ $probability = 91.3\%$ – called as **Power**.

Testing Hypothesis – Example 3

Problem

- Coin is tossed 20 times – $H_0 : p = 0.5$ and $H_A : p = 0.8$
- If the number of head is more than 14 then select H_A, otherwise select H_0

- Calculate probability for Type I error
- Its binomial with $n = 20, p = 0.5$ and $k \geq 14!$ $probability = 5.8\%$ – called as **Significance level**.
- Calculate probability for Type II error
- Its binomial with $n = 20, p = 0.8$ and $k \geq 14!$ $probability = 91.3\%$ – called as **Power**.

Testing Hypothesis

- **Significance Level:** Error probability. Should be **small**
- **Power:** Probability that the test correctly predicts that the guess is false. This should be **large**.
- Significance level is actually the tolerance level that we want.

Neyman-Pearson Lemma

Design the test in such a way that we maximize the power given the significance level

Testing Hypothesis

Significance level – what is the connection with the previous slides on testing hypothesis?

- Toss the coin many times and observe the number of heads
- $H_0 : p = 0.5$ and $H_A : p > 0/5$
- Test: If number of heads is large then take H_A otherwise take H_0
- **Method:** If *P-value* is $< 5\%$ then H_A otherwise it is H_0
- If *P-value* is $< 5\%$ then $z > 1.645$
- Actually this 5% is our significant level! – this is what percentage we may conclude that the coin is not fair but it is actually fair – this is our tolerance level
- 5% is our **significance level** and observed percentage is our **observed significance level**

Testing Hypothesis

- **Observation:** Coin tossed 400 times and observed 225 heads
- H_0: coin is fair: $p = 0.5$
- Under H_0 expect 200 heads with standard error of 10 (??)
- How did you get SE as 10? Its binomial distribution! So $\sqrt{400 * 0.5 * 0.5}$
- **First alternative:** Coin is biased towards – $p > 0.5$
- **Second alternative:** Coin is not fair – $p \neq 0.5$
- **Third alternative:** Coin is biased towards tails $p < 0.5$

Find the P-value of all the above cases.

Testing Hypothesis

- **Observation:** Coin tossed 400 times and observed 225 heads
- H_0: coin is fair: $p = 0.5$
- Under H_0 expect 200 heads with standard error of 10 (??)
- How did you get SE as 10? Its binomial distribution! So $\sqrt{400 * 0.5 * 0.5}$
- **First alternative:** Coin is biased towards – $p > 0.5$
- **Second alternative:** Coin is not fair – $p \neq 0.5$
- **Third alternative:** Coin is biased towards tails $p < 0.5$

Find the P-value of all the above cases.

Testing Hypothesis – first alternative

- The chance of getting 225 heads or more assuming the coin is fair
- $z = \frac{225-200}{10} = 2.5, \; P = 0.71\%$ (approx)
- $P < 5\%$ so choose the first alternative
- Called as one-tailed test

LNMIIT

Testing Hypothesis – second alternative

- The chance of getting the number of heads as 25 more or less the expected value of 200 assuming that the coin is fair
- $z = \pm 2.5, \;\; P = 2 * 0.71\% = 1.42\%$ (approx)
- $P < 5\%$ so choose the second alternative
- Called as two-tailed test

Testing Hypothesis – third alternative

- The chance of getting the number of heads is 225 heads or fewer assuming that coin is fair.
- $z = \frac{226-200}{10} = 2.6, \; P = 99.4\%$ (approx)
- $P > 5\%$ so choose H_0
- This is another one-tailed test

LNMIIT

z-test

- Average height of Indian men = 175cm
- **Data:** Simple random sample of 100 students taken from the universities in India with ten of thousands of students
  - Average height of sample = 174cm
  - Standard deviation = 5cm

- **Question:** Is the average height of students shorter than Indian men?
- What's our population of interest here? students in the universitues in India
- What is the unknown parameter here? population mean

LNMIIT

z-test

- Average height of Indian men = 175cm
- **Data:** Simple random sample of 100 students taken from the universities in India with ten of thousands of students
  - Average height of sample = 174cm
  - Standard deviation = 5cm

- **Question:** Is the average height of students shorter than Indian men?
- What's our population of interest here? students in the universitues in India
- What is the unknown parameter here? population mean

z-test

- Average height of Indian men = 175cm
- **Data:** Simple random sample of 100 students taken from the universities in India with ten of thousands of students
  - Average height of sample = 174cm
  - Standard deviation = 5cm

- **Question:** Is the average height of students shorter than Indian men?
- What's our population of interest here? students in the universitues in India
- What is the unknown parameter here? population mean

LNMIIT

z-test

- Average height of Indian men = 175cm
- **Data:** Simple random sample of 100 students taken from the universities in India with ten of thousands of students
  - Average height of sample = 174cm
  - Standard deviation = 5cm

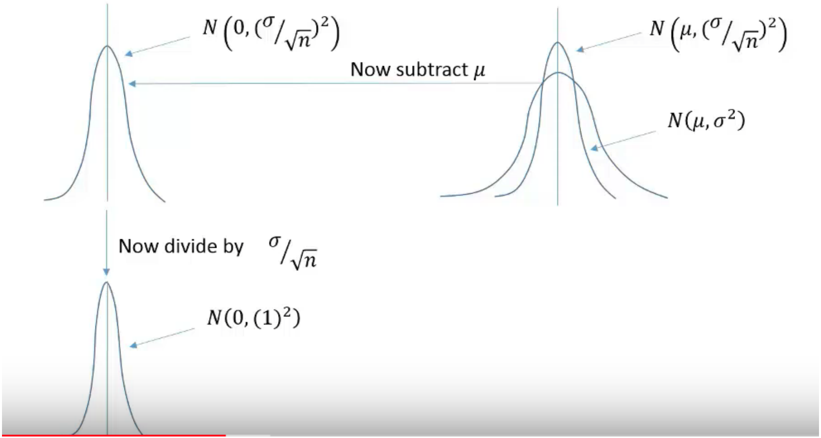
- $H_0 : \mu = 175cm$ and $H_A : \mu < 175cm$
- Calculate the P-value: Assuming H_0 what is the probability that the sample mean would be less than 175?
- If H_0 is true then the expected sample mean would be 175 with standard error given by $\sigma/\sqrt{100}$ where σ is approxiamtely 5cm
- Hence $SE = 5/\sqrt{100} \implies SE = 0.5$

z-test

- Average height of Indian men = 175cm
- **Data:** Simple random sample of 100 students taken from the universities in India with ten of thousands of students
  - Average height of sample = 174cm
  - Standard deviation = 5cm

- Now $z = (174 - 175)/0.5 \implies z = -2 \implies P = 2.5\%$ (approx)
- $P < 5\%$ so reject the null hypothesis *implies* average height of students is less than the average height of Indian men

z-test



$N\left(0, \left(\sigma/\sqrt{n}\right)^2\right)$
$N\left(\mu, \left(\sigma/\sqrt{n}\right)^2\right)$
Now subtract μ
$N(\mu, \sigma^2)$
Now divide by $\sigma/\sqrt{n}$
$N(0, (1)^2)$

t-test

- Population of weights: approximately normal. Mean (μ) and SD (σ) unknown (unknown parameters)
- **Data:** 31.8, 30.9, 34.2, 32.1, 28.8 (in gm)
- Sample mean = 31.56 gm
- Belief is that the population mean = 30 gm

- $H_0 : \mu = 30$ and $H_A : \mu > 30$
- Calculate the *P*-value: If the Null were true what is the probability of sample eman to be greater than 31.56?

t-test

- Population of weights: approximately normal. Mean (μ) and SD (σ) unknown (unknown parameters)
- **Data:** 31.8, 30.9, 34.2, 32.1, 28.8 (in gm)
- Sample mean = 31.56 gm
- Belief is that the population mean = 30 gm

- If the null were true expected value of the sample mean would be 30 and SE of sample mean = $\sigma/\sqrt{5}$
- How to find σ?
- Calculate it from the data that you know! (divide by $n - 1$)

t-test

- Population of weights: approximately normal. Mean (μ) and SD (σ) unknown (unknown parameters)
- **Data:** 31.8, 30.9, 34.2, 32.1, 28.8 (in gm)
- Sample mean = 31.56 gm
- Belief is that the population mean = 30 gm

- If the null were true expected value of the sample mean would be 30 and SE of sample mean = $\sigma/\sqrt{5}$
- How to find σ?
- Calculate it from the data that you know! (divide by $n - 1$)

t-test

- Population of weights: approximately normal. Mean (μ) and SD (σ) unknown (unknown parameters)
- **Data:** 31.8, 30.9, 34.2, 32.1, 28.8 (in gm)
- Sample mean = 31.56 gm
- Belief is that the population mean = 30 gm

- $\sigma/\sqrt{5} \implies 1.96/\sqrt{5} = 0.87$
- Now calculate a score called *t*-score: $(31.56 - 30)/0.87 = 1.79 \implies P = 7.2\%$
- $P > 5\%$ so keep the null hypothesis