Hypothesis Testing I

Agenda

- Hypothesis Testing
  - Terminologies
  - Decision making method
  - Test based on Z statistic
  - Error in hypothesis testing
- Large Sample Test
  - One Sample
  - Two Sample

Statistics

It is used to only describe the sample or summarize information about the sample

It is used to make inferences and generalizations about the broader population from the sample

Inferential Statistics

Testing of Hypothesis

Question:

A manufacturer produces batteries. He claims that the life of his batteries is 25,500 hours. Can we verify his claim?

Solution:

A trivial solution is to obtain the life of all the batteries produced by the manufacturer by using them and thereafter verify this claim

However, by doing so all the manufactured products (batteries) will be exhausted and he will not have any batteries to sell.

To avoid such scenarios, we statistically test the hypothesis

What is a hypothesis?

A hypothesis is a claim about the population parameter

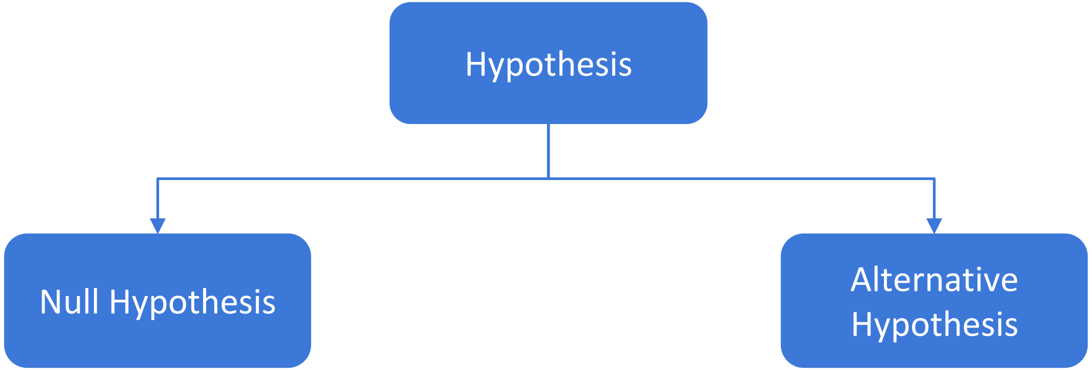
Examples:

- The average diameter of bolts manufactured is 200 cm
- The new drug is more effective

Hypothesis - example

A pharmaceutical has claimed to have produced a new drug is effective against cancer treatment.

To test the pharmaceuticals claim, first write the test hypotheses - Null Hypothesis and the Alternative Hypothesis



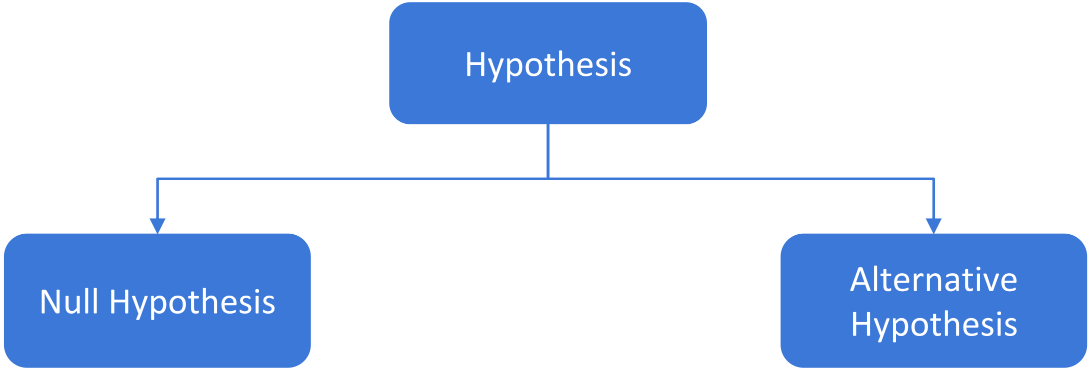
Null and alternative hypothesis

Hypothesis

Null Hypothesis	Alternative Hypothesis
• Hypothesis of no change	• It directly contradicts the null hypothesis
• Denoted by H_0	• Also known as research hypothesis
	• Denoted by H_a or H_1

Hypothesis - example

A pharmaceutical has claimed to have produced a new drug which treats the patients against cancer.



H_0: The new drug is not effective

H_1: The new drug is effective

Hypothesis - example

- Now consider the effectiveness of the new drug is determined by the number of days a patient requires to recover.

- The data for the number of days the patient take to recover are recorded

- Let the random variable X be the number of days required to recover

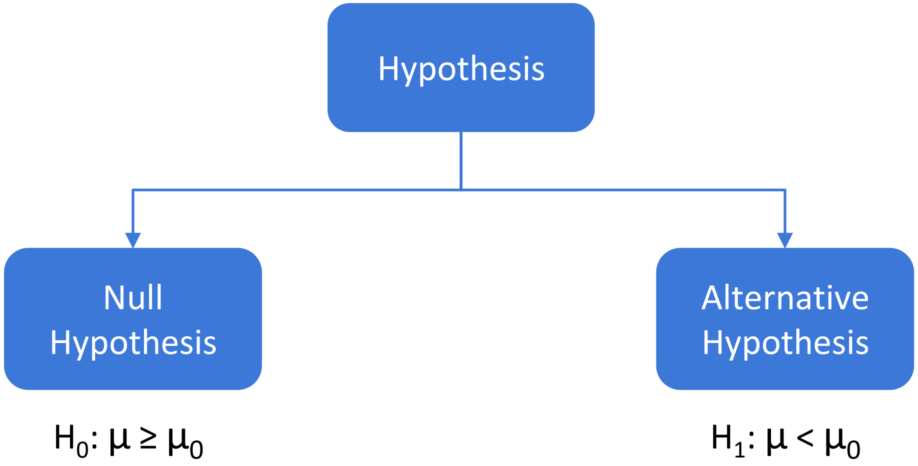
- Thus μ is the average number of days required to recover

State the hypothesis

It is claimed that the new drug reduces the recovery time

μ: the average number of days required to recover

Here, μ_0 is the general recovery time



Hypothesis testing

- The null and alternative hypothesis are contrasting
- Either the null or the alternative hypothesis can be true
- The decision of the test procedure is one of the following:
    - Reject H_0 implies the null hypothesis is false
    - Fail to reject H_0 implies the null hypothesis is true

“Fail to reject H_0”

- The null hypothesis is assumed to be true unless proven to be otherwise
- The null hypothesis gets the benefit of doubt
- The alternative hypothesis has the burden of proof, i.e. from the data it will have enough evidence to say the null hypothesis is false
- Hence, it is preferred to say ‘fail to reject H_0’ instead of accept H_0

Question:

Write the null and alternative hypothesis for the following scenarios:

1. A bolt manufacturing company claims that the average number of he manufactures in a day is 60
2. The new virus testing kit takes 20 hours less time than usual testing kit that takes 34 hours for the results
3. An analyst wants to test whether the Apple Inc. has outperformed in 2019 by 13.5%

1. A bolt manufacturing company claims that the average number of bolts he manufactures in a day is 60

Solution:

The null and alternative hypothesis for the following scenarios:

Let X: the number of bolts manufactured in a day and

μ: the average number of bolts manufactured in a day

$H_0: \mu = 60$ Against $H_1: \mu \neq 60$

Two tailed test

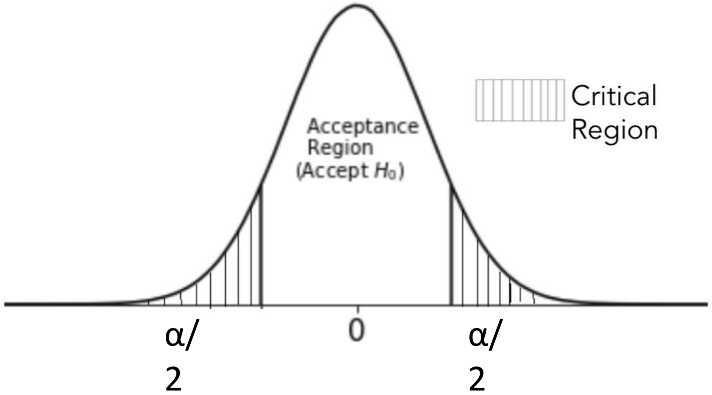
Two-tailed test is a method in which the critical region of a distribution is split in its two tails. It tests whether a sample is greater than or less than the population parameter.
A critical region is a region in which the null hypothesis is rejected. It is also known as 'Rejection Region'.

Example

Let X: the number of bolts manufactured in a day and

μ: the average number of bolts manufactured in a day

$H_0: \mu = 60$ Against $H_1: \mu \neq 60$



2. The new virus testing kit takes less time than usual testing kit that takes 34 hours for the results

Solution:

The null and alternative hypothesis for the following scenarios:

Let X: the time required for the results by a testing kit

μ: the average time required for the results by a testing kit

$H_0: \mu \geq 34$ Against $H_1: \mu < 34$

Left tailed test

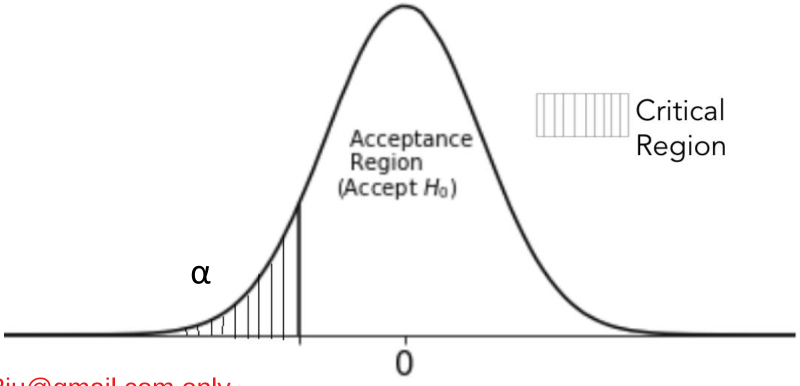
A left-tailed test is a statistical hypothesis test set up to show that the sample parameter would be lower than the population parameter.

Example

Let X: the time required for the results by a testing kit

μ: the average time required for the results by a testing kit

$H_0: \mu \geq 34$ Against $H_1: \mu < 34$



3. An analyst wants to test whether the Apple Inc. has outperformed in 2019 by 13.5%

Solution:

The null and alternative hypothesis for the following scenarios:

Let X: the performance of Apple Inc. stock

μ: the average performance of Apple Inc. stock

$H_0: \mu \leq 13.5$ Against $H_1: \mu > 13.5$

Right tailed test

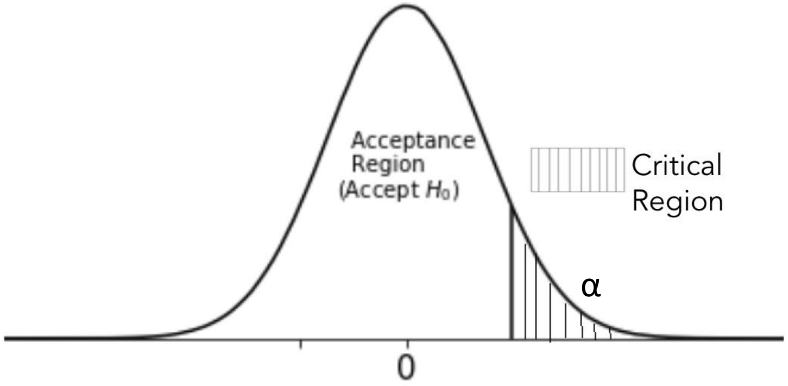
A right-tailed test is a statistical hypothesis test set up to show that the sample parameter would be higher than the population parameter.

Example

Let X: the performance of Apple Inc. stock

μ: the average performance of Apple Inc. stock

$H_0: \mu \leq 13.5$ Against $H_1: \mu > 13.5$



Write the null and alternative hypothesis and identify the type of the test.

- To test the hypothesis that the mean diastolic blood pressure for a group of 85 adults is less than 90mm
- The tensile strength of an alloy is more than 500 Pa
- The average heart rate of a healthy human is 85 beats per minute
- The yield BT cotton is more than 450 kg lint per hectare

Summary

Example	Hypothesis	Test Type
The company manager claims that the average weight of cookies in a gift box is 235 gm	$H_0: \mu = 235$ $H_1: \mu \neq 235$	Two Tailed Test
The station officer claims that the bullet train takes less than 2 hours from London to Brussels	$H_0: \mu \geq 2$ $H_1: \mu < 2$	Left Tailed Test
To test whether the average number of boat rides are more than 10 per day	$H_0: \mu \leq 10$ $H_1: \mu > 10$	Right Tailed Test

Decision Making Methods

Decision making methods

- Test statistic, p-value and confidence interval are the three different criterias to test the acceptance/ rejection of null hypothesis

- Test statistic is a sample statistic that is used to conduct a hypothesis test

- The p-value criteria returns the probability of getting the value greater than or equal to the test statistic

- The null hypothesis is accepted if the confidence interval contains the specified population parameter

Test statistic

- The decision whether to accept or reject the null hypothesis is based on the test statistic value
- Depending upon the problem in hand the test statistic changes
- The test statistic is used to quantify the sample data that distinguishes the null and alternative hypothesis
- The sampling distribution of a test statistic under the null hypothesis is used to calculate the p-value

Procedure - using test statistic

State the H_0 and H_1

↓

Compute the test statistic

↓

Compare the test statistic with the appropriate table value

↓

Based on the decision rule take a decision

↓

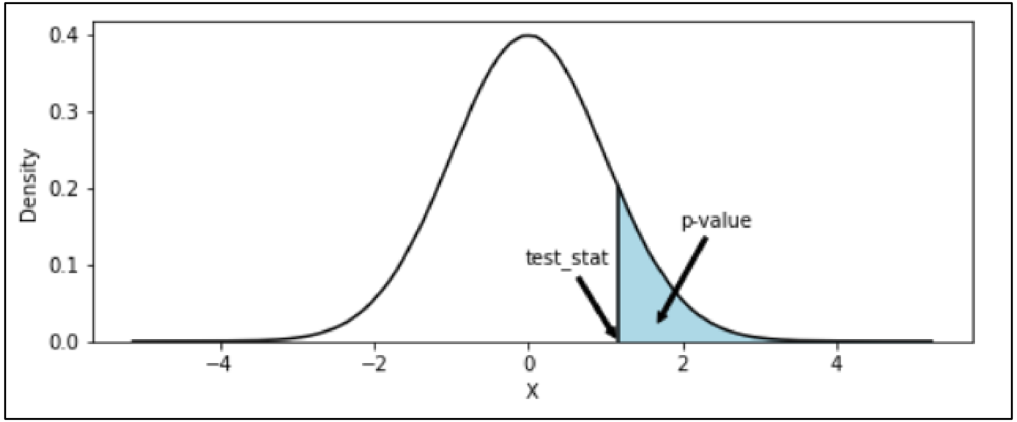
Draw the conclusion

The p-value

- The p-value is the mostly used criteria while performing the hypothesis test
- It is the probability of getting a value greater than or equal to the test statistic
- If T is the test statistic and T_0 is calculated value then observed level of significance or p-value is $P(T>T_0 | H_0 \text{ is true})$
- The test is significant if the p-value is less than or equal to the level of significance (α)

The p-value

- The smaller p-value supports the alternative hypothesis, as it exhibits the difference between the observed data and the null hypothesis



Procedure - using p-value

State the H_0 and H_1

↓

Compute the test statistic

↓

Compute the p-value

↓

Reject H_0, if p-value is less than or equal to α

↓

Draw the conclusion

Procedure - using confidence interval

State the H_0 and H_1

Compute the confidence interval

Reject H_0, if specified value of the population parameter does not lie in confidence interval

Draw the conclusion

Tests based on Z statistic

Tests based on Z statistic

Assumptions:

- The data is continuous in nature
- The sample is assumed to be drawn from a population following the normal distribution
- The sample is a simple random sample (each observation is equally likely to be drawn)

PLEASE!! NOTE

Test for normality

- The hypothesis to test whether the population is normally distributed

H_0: The data is normal against H_1: The data is not normal

- Failing to reject H_0, implies that the data is normally distributed

- The Shapiro-Wilk test is used

Note: The python code for Shapiro test is scipy.stats.shapiro(Sample)

Tests based on Z statistic

- The test statistic Z is given by

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

Sample mean

Specified mean

Population standard deviation

Sample size

- Under H_0, the test statistic follows normal distribution
- It is used to test whether the population mean is equal to a specific mean (μ)

PLEASE!! NOTE

The python code to conduct a Z test for population mean is

statsmodels.stats.weightstats.ztest(sample, value, alternative)

Decision rule

	H_1	Based on critical region	Based on p-value	Based on confidence interval
For two tailed test	$\mu \neq \mu_0$	Reject H_0 if $\|Z\| > Z_{\alpha/2}$	Reject H_0 if p-value is less than or equal to the level of significance	Reject H_0 if μ_0 does not lie in the confidence interval
For left tailed test	$\mu < \mu_0$	Reject H_0 if $Z < -Z_\alpha$		
For right tailed test	$\mu > \mu_0$	Reject H_0 if $Z > Z_\alpha$		

Two tailed test

Question:

A car manufacturing company claims that the mileage of their new car is 25 kmph with a standard deviation of 2.5 kmph. A random sample of 45 cars was drawn and recorded their mileage as per the standard procedure. From the sample, the mean mileage was seen to be 24 kmph. Is this evidence to claim that the mean mileage is different from 25kmph? (assume normality of data)

Test the claim using critical region technique, p-value technique and confidence interval technique. [Use $\alpha = 0.01$].

Two tailed test

Solution:

A car manufacturing company claims that the mileage of a car is 25 kmph with a standard deviation of 2.5 kmph.
Thus $\mu = 25$ kmph and $\sigma = 2.5$ kmph

A random sample of 45 cars was drawn and recorded their mileage as per the standard procedure. From the sample, the mean mileage was seen to be 24 kmph.
Thus $\bar{X} = 24$ kmph and $n = 45$

To find: Evidence to claim that the mean mileage is different from 25kmph

Two tailed test

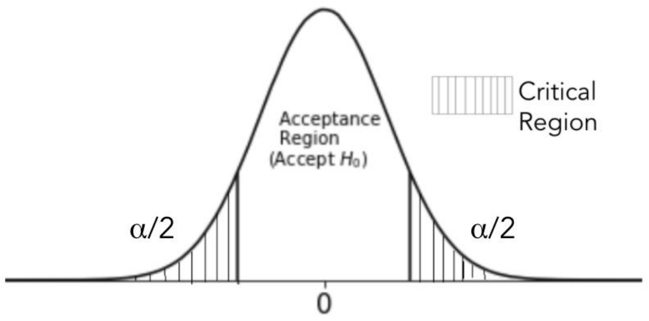
Solution:

Here X: the mileage of the new car

$X \sim N(\mu, \sigma^2)$

To test, $H_0 : \mu = 25$ against $H_1 : \mu \neq 25$

Obtain the solution using critical region technique.



The test statistics:

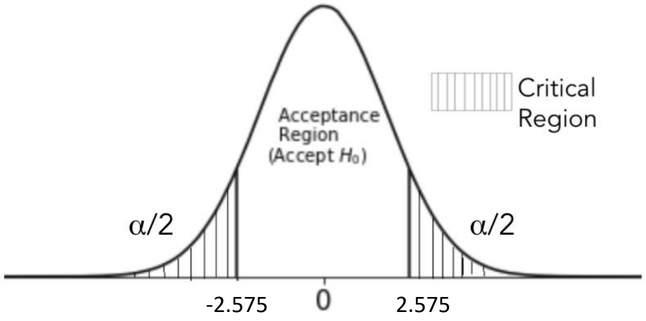
$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Two tailed test

Solution:

The test statistics:

$$Z_{calc} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{24 - 25}{2.5/\sqrt{45}} = -2.68$$



Here $\alpha = 0.01$, so $z_{0.005} = 2.575$

Decision: Reject H_0, if $|Z_{calc}| > Z_{\alpha/2}$. Here $|-2.68| > 2.575$.

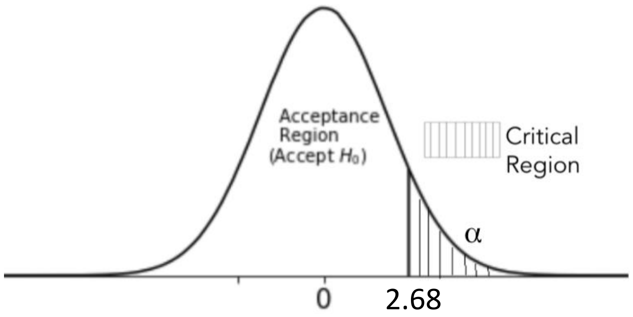
Thus reject H_0.
Thus there is enough evidence to conclude that the mean mileage is different from 25 kmph.

Two tailed test

Solution:

Let us obtain the decision using the p-value method.

$$Z_{calc} = -2.68$$



The p-value = 2 .P(|Z| > |Z_{calc}|, under H_0)

= 2 .P(|Z| > 2.68, μ = 25) … (due to symmetricity)

= 0.0074

Since p-value < 0.01, we reject H_0.

Thus there is enough evidence to conclude that the mean mileage is different from 25 kmph.

Two tailed test

Solution:

Solution using confidence interval method.

The confidence interval is given by $\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

The CI = ($\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$) = ($24 - 2.575\frac{2.5}{\sqrt{45}}, 24 + 2.575\frac{2.5}{\sqrt{45}}$)

= $(23.05, 24.95)$

Since $\mu_0 = 25$, does not lie in (23.05, 24.95), we reject H_0.

Thus there is enough evidence to conclude that the mean mileage is different from 25 kmph.

Two tailed test

Python solution: Calculate Z_{calc}

```
# define a function to calculate the Z-test statistic
# pass the population mean, population standard deviation, sample size and sample mean
def z_test(pop_mean, pop_std, n, samp_mean):

    # calculate the test statistic
    z_score = (samp_mean - pop_mean) / (pop_std / np.sqrt(n))

    # return the z-test value
    return z_score

# calculate the test statistic using the function 'z_test'
z_score = z_test(pop_mean = 25, pop_std = 2.5, n = 45, samp_mean = 24)
print("Z-score:", z_score)

Z-score: -2.6832815729997477
```

Reject H_0, as $|Z_{calc}| > Z_{\alpha/2}$. Here $|-2.68| > 2.575$

Two tailed test

Python solution: Calculate p-value

```
# calculate the corresponding p-value for the test statistic
# use 'cdf()' to calculate P(Z <= z_score)
p_value = stats.norm.cdf(z_score)

# for a two-tailed test multiply the p-value by 2
req_p = p_value*2
print('p-value:', req_p)
```

```
p-value: 0.007290358091535638
```

Since p-value < 0.01, we reject H_0.

Two tailed test

Python solution: Calculate 99% confidence interval

```
# calculate the 99% confidence interval for the population mean
# pass the sample mean to the parameter, 'loc'
# pass the scaling factor (pop_std / n^(1/2)) to the parameter, 'scale'
print('Confidence interval:', stats.norm.interval(0.99, loc = samp_mean, scale = pop_std / np.sqrt(n)))
```

```
Confidence interval: (23.040045096471452, 24.959954903528548)
```

Since $\mu_0 = 25$, does not lie in the confidence interval, we reject H_0.

Left tailed test

Question:

A sample of 900 PVC pipes is found to have an average thickness of 12.5 mm. Can we assume that the sample is coming from a normal population with mean 13mm against that it is less than 13 mm. The population standard deviation is 1 mm. Test the hypothesis using the p-value method at 5% level of significance.

Left tailed test

Solution:

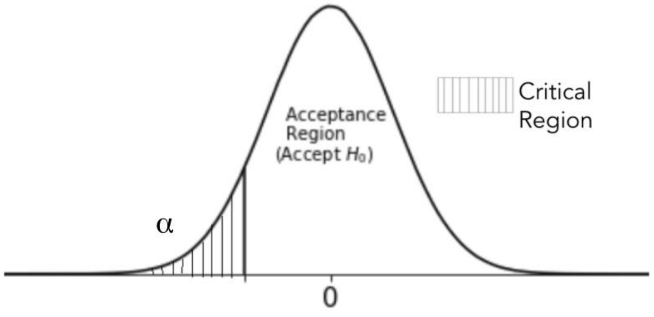
Here X: the thickness of PVC pipes

$X \sim N(\mu, \sigma^2)$

Here, $\mu = 13$mm , $\sigma = 1$ mm, $\bar{X} = 12.5$, $n = 900$

To test, $H_0 : \mu \geq 13$ against $H_1 : \mu < 13$

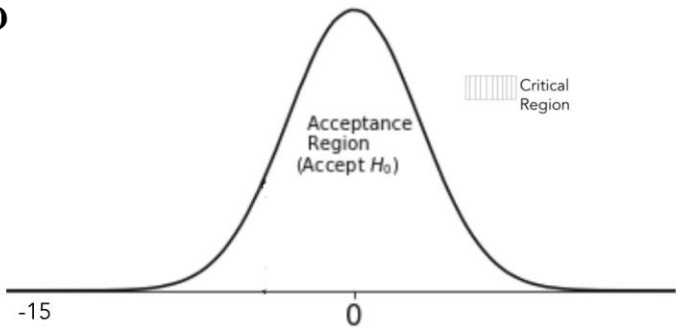
Obtain the solution using p-value technique.



Left tailed test

Solution:

The test statistics: $$Z_{calc} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{12.5-13}{1/\sqrt{900}} = -15$$



The p-value = P(Z < Z_{calc}, under H_0) = P(Z < -15, μ = 13) = 0.00

Since p-value < 0.05, we reject H_0.

Thus there is enough evidence to conclude that the sample may not be regarded as coming from population with mean 13

Left tailed test

Python solution: Calculate p-value

```
# calculate the corresponding p-value for the test statistic
# use 'cdf()' to calculate P(Z <= -15)
p_value = stats.norm.cdf(-15)

print('p-value:', p_value)

p-value: 3.6709661993126986e-51
```

Since p-value < 0.05, we reject H_0.

Right tailed test

Question:

An e-commerce company claims that the mean delivery time of food items on their website in NYC is 60 minutes with a standard deviation of 30 minutes. A random sample of 45 customers ordered from the website, and the mean time for delivery was found to be 75 minutes. Is this enough evidence to claim that the mean time to get items delivered is more than 60 minutes. (assume normality of data)

Test the claim using p-value technique. [Use $\alpha = 0.05$].

Right tailed test

Solution:

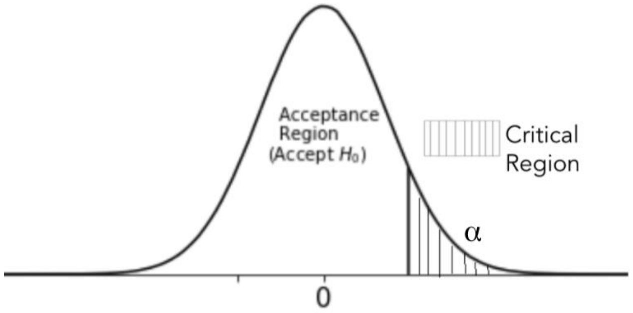
Here X: Time (in minutes) to receive the delivery

$X \sim N(\mu, \sigma^2)$

Here, $\mu = 60$, $\sigma = 30$, $\bar{X} = 75$, $n = 45$

To test, $H_0 : \mu \leq 60$ against $H_1 : \mu > 60$

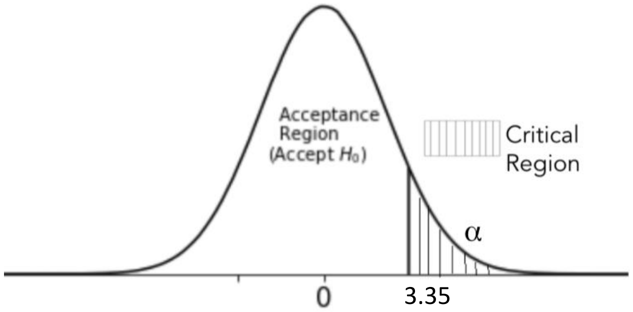
Obtain the solution using p-value technique.



Right tailed test

Solution:

The test statistics: $$Z_{calc} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{75-60}{30/\sqrt{45}} = \mathbf{3.35}$$



The p-value = $P(Z > Z_{calc}$, under $H_0) = P(Z > 3.35, \mu = 60) = 0.0004$

Since p-value < 0.05, we reject H_0.

Thus there is enough evidence to conclude that the mean delivery time is more than 60 min

Right tailed test

Python solution: Calculate p-value

```
# calculate the corresponding p-value for the test statistic
# use 'sf()' to calculate P(Z > 3.35)
p_value = stats.norm.sf(3.35)

print('p-value:', p_value)
```

```
p-value: 0.0004040578018640207
```

Since p-value < 0.05, we reject H_0.

Summary

- The Z test can be used if the sample data is drawn from a normally distributed population
- The normality of the dataset can be checked using Shapiro-Wilk test
- Two tailed as well as one tailed (left/ right) Z test can be performed
- The test statistic for the Z test follows normal distribution under the null hypothesis
- The test statistic, p-value and confidence interval can be used to analyze the significance of the test

Errors in Hypothesis Testing

Probability of committing an error

- P(Type I error) is the probability of wrongly rejecting a true null hypothesis
- It is the level of significance (α) set for the hypothesis
- An α of 0.05 suggests that one is willing to accept a 5% chance that you reject a true null hypothesis
- In order to lower this risk, set a lower α. However, it will be more difficult to reject the false null hypothesis

Error in hypothesis testing

		Test Decision	
		Reject H_0	Accept H_0
Actual Situation	H_0 is True	Type I error Wrongly reject true H_0	Correct Conclusion Correctly accept true H_0
	H_0 is False	Correct Conclusion Correctly reject false H_0	Type II error Wrongly accept false H_0

Probability of committing an error

- P(Type II error) is the probability of wrongly accepting a false null hypothesis
- Denoted by β
- 1 - β is the power of the test, i.e. correctly rejecting a false null hypothesis
- In order to lower the risk of type II error, ensure the power of the test is high
- In python, we calculate the power of a Z test using the 'power.zt_ind_solve_power()' from statsmodels library

PLEASE!! NOTE

Controlling the errors

- Both errors cannot be controlled simultaneously
- Fix one error and minimize the other
- Generally α is fixed and β is controlled

Previously

- The unknown population parameters are estimated by from sample
- The unknown population parameters are replaced by the sample estimates
- Test the hypothesis for population mean when the population variance is known

Now…

- Testing of Hypothesis
  - Large sample tests
  - Small sample tests
  - Tests for proportion

Large Sample Tests

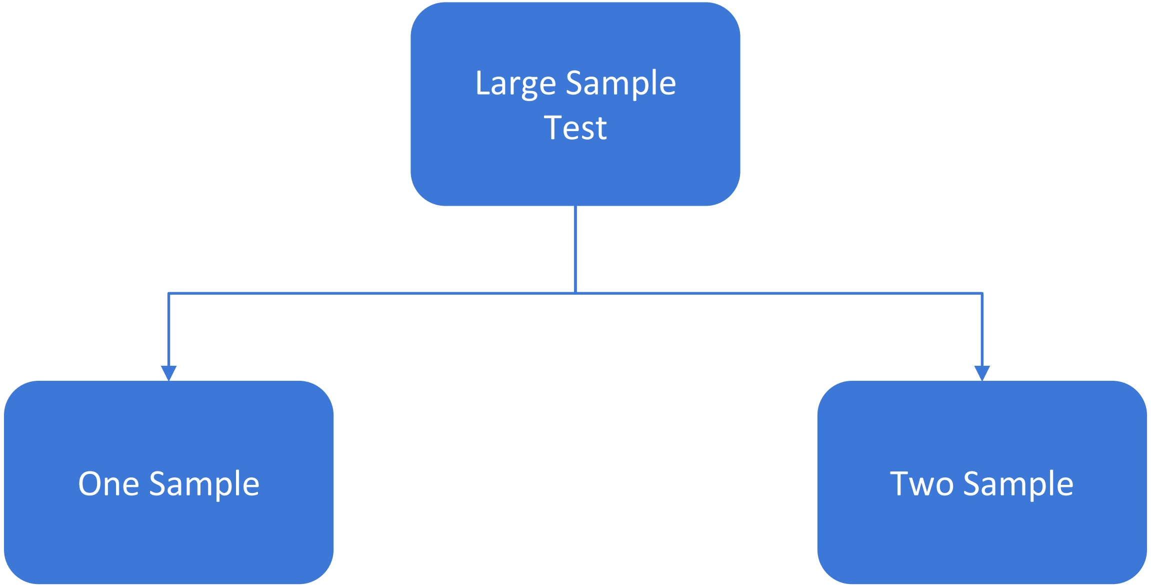
Large sample tests

- For large n, according to CLT, the sampling distribution of mean follows normal distribution

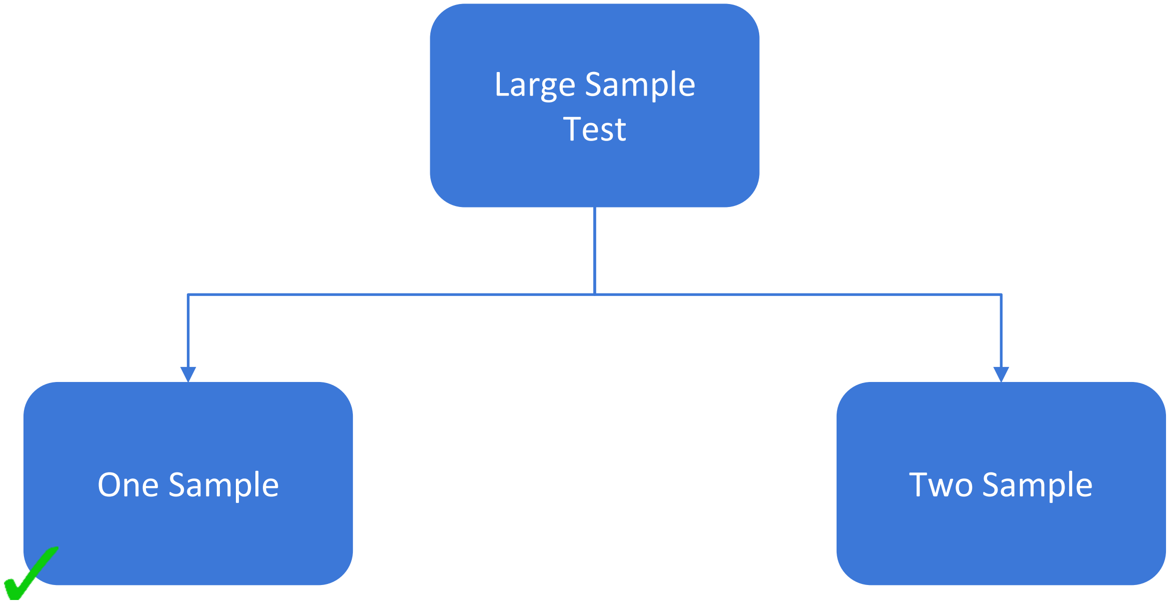
- For tests $n \geq 30$ are considered to be large sample tests

- If σ is unknown, its point estimator - sample standard deviation - is used

Large sample tests



Large sample tests



Large sample tests

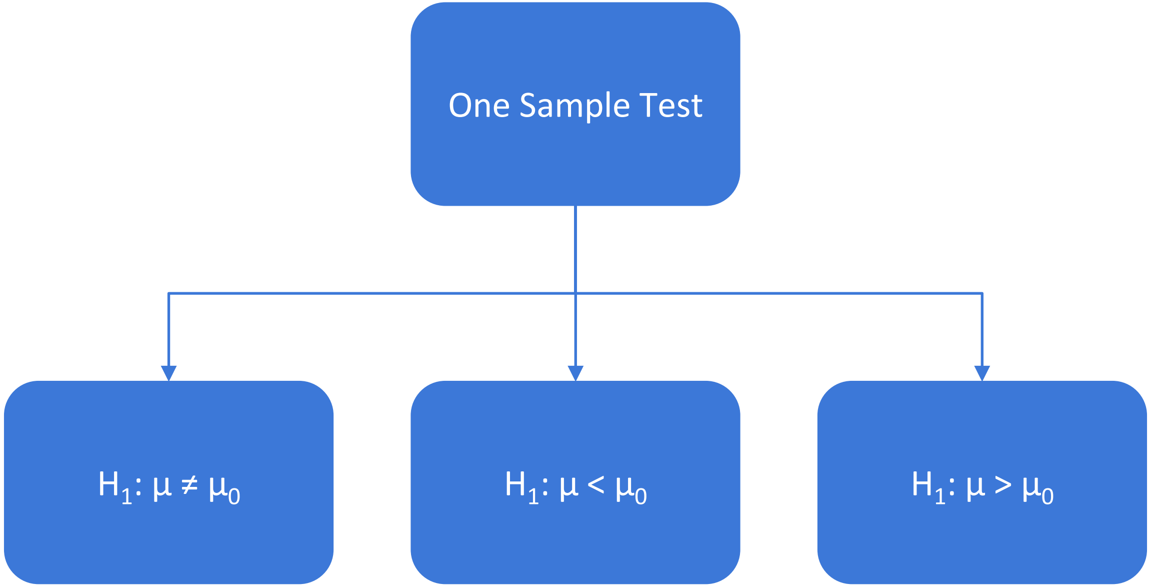
- One sample

- Two sample

Number of Defective Items in a Box
5
8
2
5
3
6

Can we say that each box has less than 3 defective items?

One sample test



One sample tests

- In the last session, we considered the case where σ^2 is known
- The test statistic changes slightly for σ^2 is unknown
- σ^2 replaced by sample variance (s^2)

Tests based on Z statistic

- The test statistic Z given by

Sample mean

Specified mean

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

Sample standard deviation

Sample size

where $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

- Under H_0, the test statistic follows normal distribution

PLEASE!! NOTE

The python code to conduct a Z test for population mean is

```
statsmodels.stats.weightstats.ztest(sample,  value, alternative)
```

Why n-1?

- The unbiased estimator of population variance is $s^2 = \frac{n(\text{sample variance})}{n-1}$

- Also $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ and $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ have the same numerator, thus dividing it by n-1 gives a smaller value which is unbiased

- If a sample of size 1 is drawn, say the observation is 475, the mean would be is 475 and there is no variation with just one observation. Thus, the denominator is n-1 is more appropriate

Test for population mean (σ unknown)

Question:

The manager of a packaging process at a protein powder manufacturing plant wants to determine if the protein powder packing process is in control. The correct amount of protein powder per box is 350 grams on an average. A sample of 80 boxes was drawn which gave a mean of 354.5 grams with a standard deviation of 15. At 5% level of significance, is there evidence to suggest that the weight is different from 350 grams.

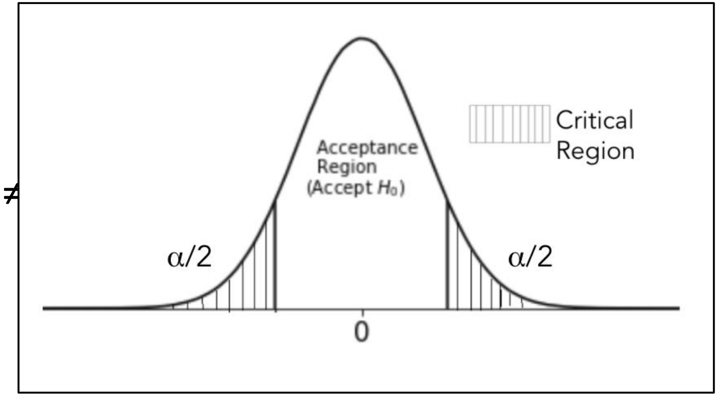
Test for population mean (σ unknown)

Solution:

X: Amount of protein powder in a box
$X \sim N(\mu, \sigma^2)$

Here n = 80, s = 15, μ =350, $\bar{X}$ = 354.5

To test H_0: μ = 350 against H_1: $\mu \neq$

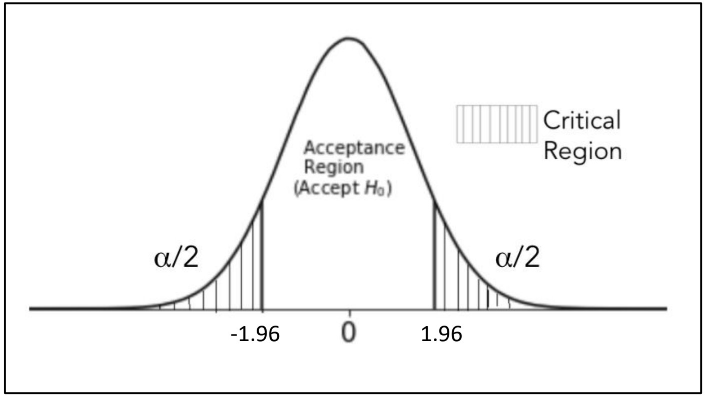


Test for population mean (σ unknown)

Solution:

The test statistic is

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{354.5 - 350}{15/\sqrt{80}} = 2.683$$



Decision Rule: Reject $|Z_{calc}| \geq Z_{\alpha/2}$

Here $Z_{\alpha/2} = 1.96$

Since 2.683 > 1.96, reject H_0.

We may conclude that the average weight of protein powder in the box is not 350 grams

Test for population mean (σ unknown)

Python solution: Calculate critical z-value

```
# calculate the z-value for 95% of confidence level
# use 'stats.norm.isf()' to find the z-value corresponding to the upper tail probability 'q'
# pass the value of 'alpha/2' for a two-tailed to the parameter 'q', here alpha = 0.05
# use 'round()' to round-off the value to 2 digits
z_val = np.abs(round(stats.norm.isf(q = 0.05/2), 2))

print('Critical value for two-tailed Z-test:', z_val)
```

```
Critical value for two-tailed Z-test: 1.96
```

i.e. If test statistic is less than -1.96 or greater than 1.96 then we reject H_0.

Test for population mean (σ unknown)

Python solution: Calculate test statistic

```python
# define a function to calculate the Z-test statistic
# here the population mean is unknown, thus use the sample standard deviation
# pass the population mean, sample standard deviation, sample size and sample mean as the function input
def z_test(pop_mean, samp_std, n, samp_mean):

    # calculate the test statistic
    z_score = (samp_mean - pop_mean) / (samp_std / np.sqrt(n))

    # return the z-test value
    return z_score

# calculate the test statistic using the function 'z_test'
z_score = z_test(pop_mean = 350, samp_std = 15, n = 80, samp_mean = 345.5)
print("Z-score:", z_score)
```

```
Z-score: 2.6832815729997477
```

As test statistic (=2.68) > critical value (=1.96), we reject H_0.

Test for population mean (σ unknown)

Python solution: Calculate p-value

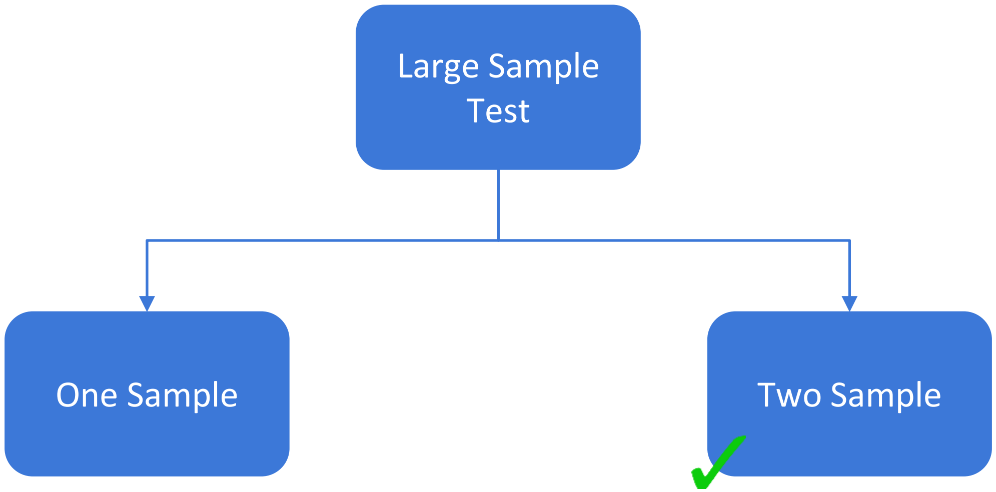
```
# calculate the corresponding p-value for the test statistic
# use 'sf()' to calculate P(Z > z_score)
p_value = stats.norm.sf(z_score)

# for a two-tailed test multiply the p-value by 2
req_p = p_value*2
print('p-value:', req_p)
```

```
p-value: 0.007290358091535638
```

As p-value < 0.05, we reject H_0.

Large sample tests



Large sample tests

- One sample
- Two sample

Number of Defective Items from Production Line A	Number of Defective Items from Production Line B
1	5
3	8
5	2
6	5
2	3
4	6

Can we say that Production Line B produces less defectives than Production Line A?

Two sample tests

- The two sample tests are used to compare the equality of means of two populations

- Used to compare:
  - Performance of two machineries
  - Performance of two portfolios

Two sample tests for population mean

- Let there be two different populations such that they follow normal distribution
- The two samples are independent of each other
- The samples must have equal variance (it can be tested statistically)

Test for equality of variances

- The hypothesis to test whether two populations have equal variances

H_0: The variances are equal against H_1: The variances are not equal

- Failing to reject H_0, implies that the two populations have equal variance

- The Levene's test used

Note: The python code for levene's test scipy.stats.levene()

Two sample test - hypothesis

- Let there be two samples of sizes n_1 and n_2 drawn from normal population $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively

- The hypothesis to test whether the population means are equal

$H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_1$

- It implies

H_0: The two population means are equal (i.e $\mu_1 = \mu_2$)

against H_1: The two population means are not equal μ_0 (i.e $\mu_1 \neq \mu_2$)

Two sample test - hypothesis

- Like the one sample tests, it is possible to test for

$$H_0 : \mu_1 \leq \mu_2 \quad \text{against} \quad H_1: \mu_1 > \mu_2$$

Or

$$H_0 : \mu_1 \geq \mu_2 \quad \text{against} \quad H_1: \mu_1 < \mu_2$$

- Failing to reject H_0 implies that the null hypothesis is true

Two sample tests - test statistic

- The test statistic is Z given by

Sample mean

Specified mean

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Under H_0, the test statistic follows normal distribution

Two sample tests - test statistic

- If σ_1^2 and σ_2^2 are not known then they are replaced the sample estimates s_1^2 and s_2^2 respectively

- The test statistic is Z given by

Sample mean

Specified mean

$$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

where $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

The python code to conduct a Z test for two population means is

```
statsmodels.stats.weightstats.ztest(Sample_1, Sample_2, value, alternative)
```

Two sample tests - decision rule

<table>
<tr><th></th><th>H_1</th><th>Based on critical region</th><th>Based on p-value</th><th>Based on confidence interval</th></tr>
<tr><td>For two tailed test</td><td>$\mu_1 \neq \mu_2$</td><td>Reject H_0 if $|Z| \geq Z_{\alpha/2}$</td><td rowspan="3">Reject H_0 if p-value is less than or equal to the level of significance</td><td rowspan="3">Reject H_0 if $\mu_1 - \mu_2$ does not lie in the confidence interval</td></tr>
<tr><td>For left tailed test</td><td>$\mu_1 < \mu_2$</td><td>Reject H_0 if $Z \leq Z_\alpha$</td></tr>
<tr><td>For right tailed test</td><td>$\mu_1 > \mu_2$</td><td>Reject H_0 if $Z \geq Z_\alpha$</td></tr>
</table>

Two sample test for population mean (σ unknown)

Question:

A study was carried out to understand amount of hemoglobin in blood for males and females. A random sample of 160 males and 180 females have means of 13 g/dl and 15 g/dl. The two samples have standard deviation of 4.1 g/dl for male donors and 3.5 g/dl for female donor . Can it be said the population means of hemoglobin are the same for men and women? Use $\alpha = 0.01$.

Two sample test for population mean (σ unknown)

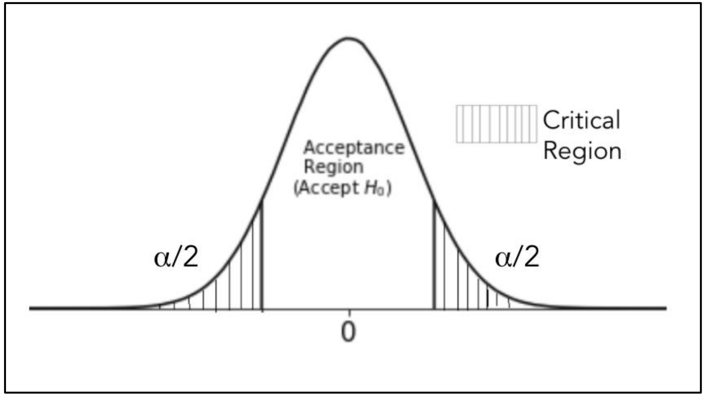
Solution:

X: Amount of hemoglobin in blood for males
Y: Amount of hemoglobin in blood for females

Here $n_1 = 160$, $s_1 = 4.1$, $\bar{X} = 13$

$n_2 = 180$, $s_2 = 3.5$, $\bar{Y} = 15$

To test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1$



Two sample test for population mean (σ unknown)

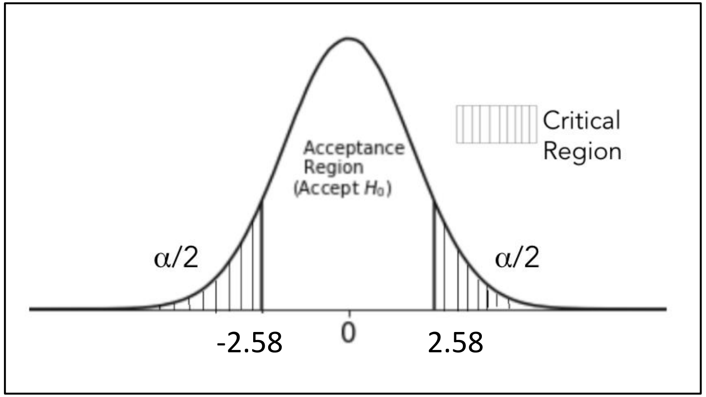
Solution:

The test statistic $Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{13 - 15}{\sqrt{\frac{4.1^2}{160} + \frac{3.5^2}{180}}} = -4.807$

Decision Rule: Reject $|Z_{calc}| \geq Z_{\alpha/2}$

Here $Z_{\alpha/2} = 2.58$

Since $4.807 > 2.58$, reject H_0.



We may conclude that both males and females have different hemoglobin averages

Two sample test for population mean (σ unknown)

Python solution: Calculate critical z-value

```
# calculate the z-value for 99% of confidence level
# use 'stats.norm.isf()' to find the z-value corresponding to the upper tail probability 'q'
# pass the value of 'alpha/2' for a two-tailed test to the parameter 'q', here alpha = 0.01
# use 'round()' to round-off the value to 2 digits
z_val = np.abs(round(stats.norm.isf(q = 0.01/2), 2))

print('Critical value for two-tailed Z-test:', z_val)
```

```
Critical value for two-tailed Z-test: 2.58
```

i.e. If test statistic is less than -2.58 or greater than 2.58 then we reject H_0.

Two sample test for population mean (σ unknown)

Python solution: Calculate test statistic

```python
# define a function to calculate the test statistic and corresponding p-value
# here the standard deviations for populations are unknown, thus use the sample standard deviations
# pass the sample mean, sample standard deviation and sample size for both the samples as the function input
# 'value' denotes the value in null hypothesis
def TwoSampZTest(samp_mean_1, samp_mean_2, samp_std_1, samp_std_2, value, n1, n2):

    # calculate the test statistic
    denominator = np.sqrt((samp_std_1**2 / n1) + (samp_std_2**2 / n2))
    zscore = ((samp_mean_1 - samp_mean_2) - (value)) / denominator

    # return the z-score
    return zscore

# pass the given data to the function 'TwoSampZTest'
zscore = TwoSampZTest(samp_mean_1 = 13, samp_mean_2 = 15, samp_std_1 = 4.1, samp_std_2 = 3.5, value = 0,
                      n1 = 160, n2 = 180)

print('z-score:', zscore)
```

```
z-score:  -4.806830552525058
```

As test statistic (=-4.8068) < critical value (=-2.58), we reject H_0.

Two sample test for population mean (σ unknown)

Python solution: Calculate p-value

```
# calculate the corresponding p-value for the test statistic
# use 'cdf()' to calculate P(Z <= z_score)
p_value = stats.norm.cdf(zscore)

# for a two-tailed test multiply the p-value by 2
req_p = p_value*2
print('p-value:', req_p)
```

```
p-value: 1.5334185117556497e-06
```

As p-value < 0.01, we reject H_0.

Thank You