Academic Year 2022-23 **Semester :VI**

Class / Branch: TE IT A/B **Subject: DS using Python Lab**

Experiment No.3

Aim: To implement two-sample Z-test.

Prerequisites: python.

Objectives: - At the end of this experiment, you will be able to:

- solving real life problems based on Statistical analysis
- Use Z test on the given problem

Theory:

Two Sample Z Hypothesis Tests is a parametric test is to compare the means of two independent groups of samples drawn from a normal population. In other words, it is the test to know the means of two populations are the same or different. For example, the mean salaries of male workers are greater than the female workers for the same job?

When to use Two sample Z hypothesis tests

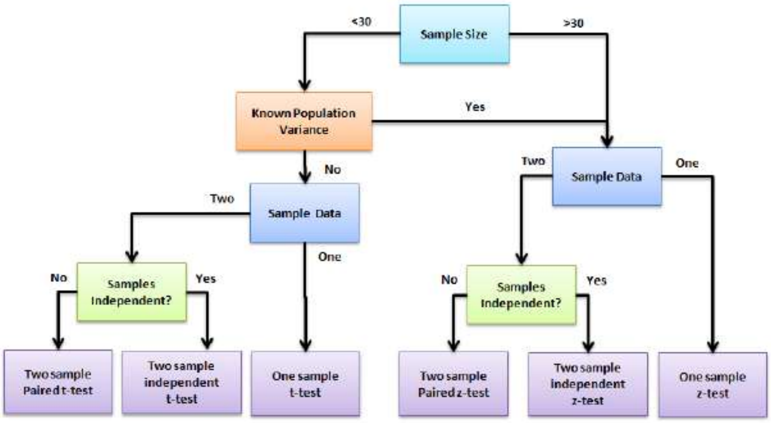
Two sample Z test compares the means of samples of independent groups

taken from a normal population. Also, in this test you will have only two groups with independent samples. To compare three groups, use One Way ANOVA instead of Z test.

Furthermore, Z test is similar to the student t-test. Z test is basically used for relatively large samples (say n>30) and the population standard deviation is known. Whereas, student t-test is for small sample size, and also t-test assumes the population standard deviation is unknown.

Academic Year 2022-23 **Semester :VI**

Class / Branch: TE IT A/B **Subject: DS using Python Lab**



Difference between One and Two sample Z hypothesis test

The two-sample z test is to tests the difference between means of two groups, whereas a one-sample z test is to tests the difference between a single group and the hypothesized population value.

Assumptions of Two sample Z hypothesis tests

- Population data is continuous
- Population follows a standard normal distribution
- Both sample ends must be higher than 30
- The population standard deviation is known
- Similar spread between the groups, in other words homogeneity of variance
- Both the samples should be randomly selected from the population

Two sample Z-test Formula

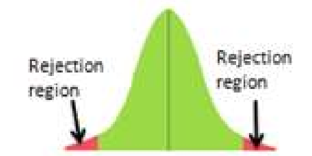
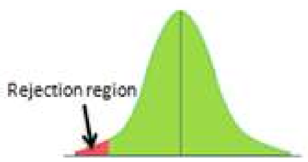
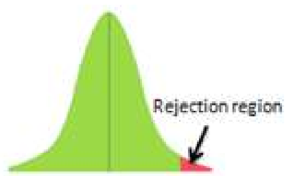
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- $\bar{x}1$ = sample mean of first sample
- $\bar{x}2$ = sample mean of second sample
- µ1= Mean of first population
- µ2= Mean of second population

- σ12= population variance in first population
- σ22= population variance in second population
- n1= sample size of first group
- n2= sample size of second group

Hypothesis Testing

A tailed hypothesis is an assumption about a population parameter. The assumption may or may not be true. A one-tailed hypothesis is a test of hypothesis where the area of rejection is only in one direction. Whereas two-tailed, the area of rejection is in two directions. The selection of one or two-tailed tests depends upon the problem.

Z- Test	Null Hypothesis (H_0)	Alternative Hypothesis (H_1)	Statistical conclusion
Two-tailed	$\mu_1=\mu_2$	$\mu_1\neq\mu_2$	
Left-tailed	$\mu_1\geq \mu_2$	$\mu_1<\mu_2$	
Right-tailed	$\mu_1\leq \mu_2$	$\mu_1>\mu_2$	

Steps to Calculate Two Sample Z hypothesis test

- Select appropriate statistic- one-tailed or two-tailed?
- Determine the null hypothesis and alternative hypothesis
- Determine the level of significance

Academic Year 2022-23 **Semester :VI**

Class / Branch: TE IT A/B **Subject: DS using Python Lab**

- Find the critical value
- Calculate the standardized test statistics
- Then make a decision to reject or fail to reject the null hypothesis. Reject the null hypothesis, If the test statistic falls in the critical region.
- Finally, interpret the decision in the context of the original claim.

Example of Right-tailed test

Example: Princeton school science teacher claims that students in his section will score higher

marks than those in his colleague's section. The mean science score for 60 students in his section is 22.1, and the standard deviation is 4.8. The mean science score for 40 of the colleagues' sections is 18.8, and the standard deviation is 8.1. At α = 0.05, can the teacher's claim be supported?

- State the null and alternative hypothesis
  - Null hypothesis H0: μ1 ≤ μ2
  - Alternative hypothesis H1 : μ1 > μ2
- Select appropriate statistic- Since the claim is student scores higher marks than the colleagues section; it is a right-tailed test.
- Level of significance: α= 0.05
- Find the critical value: 1-α= 1-0.05=0.95
- Look at the 0.95 in z table= 1.645

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

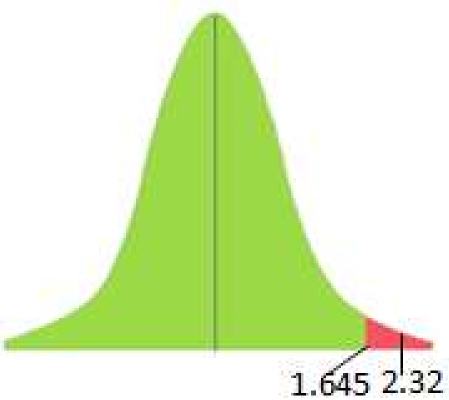
Academic Year 2022-23 **Semester :VI**

Class / Branch: TE IT A/B **Subject: DS using Python Lab**

- Calculate the test statistics
- $\bar{x}1 = 22.1$
- $\bar{x}2 = 18.8$
- $\sigma1 = 4.8$
- $\sigma2 = 8.1$
- n1= 60
- n2= 40

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(22.1 - 18.8) - (0)}{\sqrt{\frac{4.8^2}{60} + \frac{8.1^2}{40}}} = 2.32$$



Interpret the results: Compare Z calc to Z critical . In hypothesis testing, a critical value is a point on the test distribution compares to the test statistic to determine whether to reject the null

hypothesis. zcal value is in the rejection region. Hence, we have enough evidence to reject the null hypothesis. So, at 5%, we have enough evidence to support the teacher's claim that his students score higher than his colleague students.

Academic Year 2022-23 **Semester :VI**

Class / Branch: TE IT A/B **Subject: DS using Python Lab**

Example of left-tailed test

Example: A dietician claims that participants in the X program yield less

weight reduction than the Y program. The mean weight reduction for 100 participants in the X program is 19.7lb, and the standard deviation is 6.2. Similarly, for program Y, for 120 participants, it is 21.2lb, and the standard deviation is 7.6. At 1% level of significance, can you support the dietician claim?

- State the null and alternative hypothesis
    - Null hypothesis H0: μ1 ≥ μ2
    - Alternative hypothesis H1 : μ1 < μ2
- Select appropriate statistic- Since the claim is participants in X program yields less weight reduction than Y program; it is a left-tailed test.
- Level of significance: α= 0.01
- Find the critical value: 1-α= 1-0.01=0.99
- Look at the 0.99 in z table= 2.330

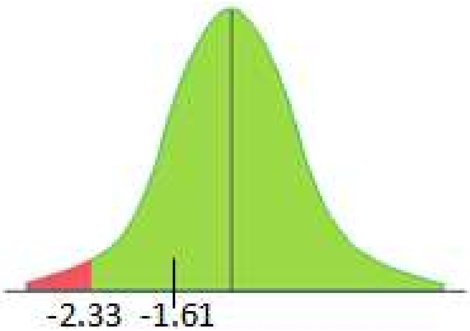
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964

- Calculate the test statistics
- $\overline{x}1 = 19.7$
- $\overline{x}2 = 21.2$

- σ1= 6.2
- σ2= 7.6
- n1= 100
- n2= 120

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(19.7 - 21.2) - (0)}{\sqrt{\frac{6.2^2}{100} + \frac{7.6^2}{120}}} = -1.612$$



Interpret the results: Compare Z calc to Z critical . In hypothesis testing, a critical value is a point on the test distribution compares to the test statistic to determine whether to reject the null hypothesis. zcal value is not in the rejection region. Hence, we failed to reject the null hypothesis. So, at 1% we don’t have enough evidence to support the dietician’s claim.

Example of Two-tailed test

Example: Glenbrook sports center is planning to compare the ages from a random sample of

---

male and female swimmers in their coaching center. The swimming coordinator collected 62 female swimmers' data, and the mean age is 23.1 with a standard deviation of 3.5. Similarly, collected 46 male swimmers data and the mean average is 19.2 with a standard deviation of 4.8. Assume the population follows a standard normal distribution. At 5% significance level, test whether there is a significant difference in age between the sexes?

- State the null and alternative hypothesis
  - Null hypothesis H0: $\mu 1 = \mu 2$
  - Alternative hypothesis H1 : $\mu 1 \neq \mu 2$
- Select appropriate statistic- Since the Glenbrook sports center planning to test whether there is a significant difference in age between the sexes ; it is a two-tailed test.
- Level of significance: $\alpha = 0.05$
- Since it is a two tailed test $\alpha/2 = 0.05/2 = 0.025$
- Find the critical value: $1-\alpha/2 = 1-0.025 = 0.975$
- Then, look at the 0.975 in z table= ± 1.96

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

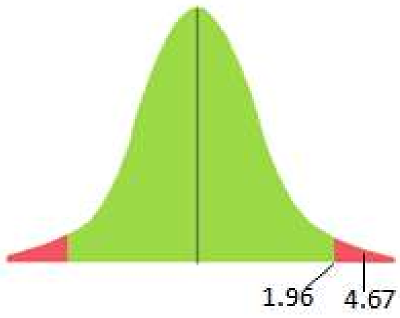
- Calculate the test statistics
- $\bar{x}1 = 23.1$
- $\bar{x}2 = 19.2$
- $\sigma 1 = 3.5$
- $\sigma 2 = 4.8$
- n1= 62
- n2= 46

Academic Year 2022-23 **Semester :VI**

Class / Branch: TE IT A/B **Subject: DS using Python Lab**

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(23.1 - 19.2) - (0)}{\sqrt{\frac{3.5^2}{62} + \frac{4.8^2}{46}}} = 4.66$$



Interpret the results: Compare Z calc to Z critical . In hypothesis testing, a critical value is a point on the test distribution compares to the test statistic to determine whether to reject the null

hypothesis. zcal value is in the rejection region. Hence, we have enough evidence to reject the null hypothesis. So, at 5% we have enough evidence to support the significant difference in age between the sexes.

Conclusion: - In this experiment, we have validated dataset by performing two-sample Z-test