Smt. Indira Gandhi College of Engineering
Computer Engineering Department
Ghansoli – Navi Mumbai
Academic Year 2023-24 (Even Sem)

Student Name: Khyati Garude **Roll No.:** 13 **Class:** BE **Sem:** VIII

Course Name: Applied Data Science Lab

Course Code: CSL8023

Experiment No. 06

Experiment Title: Implementing Inferential Statistic Computation On Selected Dataset

Date of Performance	Date of Submission	Marks (10)					Sign / Remark
		A	B	C	D	E	
		2	3	2	2	1	
8/2/24	15/2/24	2	3	2	2	0	12/4/24.
		Total Marks					
		09					

A: Prerequisite Knowledge
B: Implementation
C: Oral
D: Content
E: Punctuality & Discipline

Khyati Garude
BE COMPS
ADS

Date:

DATE	EXPERIMENT-6	SIGN OF FACULTY
15/2/24	Implementing Inferential statistics computation on the selected dataset.	12/2/24

AIM: Implementing Inferential Statistic on the selected dataset.

THEORY:

→ **Z-test:**
It is a statistical method used to determine if there is a significant difference between the means of two groups or if a single group mean is significantly different from a known value. Steps for Z-test:

(i) Formulate Hypothesis:

→ Null Hypothesis (H_0): This assumes that there is no significant difference between groups or population.

Date:

→ Alternate Hypothesis(H_1): This suggests that there is significant difference.

(ii) Select Significance level (α):
Choose a significance level, commonly denoted as α, typically set at 0.05. This represents the probability of rejecting the hypothesis.

(iii) Identify the type of z-test:
Decide whether we are conducting a on sample z-test, two sample z-test.

(iv) Calculate the z statistic:
For one sample z-test, use the formula:

$$z = \frac{\bar{x} - \mu}{\sigma - \sqrt{n}}$$

(v) Determine critical region:
Based on significance level (α) and the nature of hypothesis, find the critical-z value from the table.

(vi) Make a decision:
If the calculated value falls under the critical region, reject the null hypothesis.

Date:

(vii) Calculate P-value:
Calculate P-value associated with Z-score. If P-value is less than α, reject the null hypothesis.

→ ALGORITHM:

(i) Input:
obtain the population mean (μ), sample size (n) and significance level (α).

(ii) Null hypothesis:
Formulate the null hypothesis ($H_0 : \mu$ = population mean)

(iii) Alternative Hypothesis:
Formulate the alternative hypothesis ($H_1 : \mu \neq$ population mean).

(iv) Calculate Sample mean:
compute the sample mean ($\bar{x}$) from the given sample data.

(v) Calculate Population Standard deviation:
compute the population standard deviation (σ) using the sample data.

Date:

(vi) Calculate z-score:
compute the z-score using the formula
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

(vii) Determine the tail type:
Identify whether the test is left-tailed, right-tailed, or two tailed based on the sign of z-score.

(viii) Calculate p-value:
Based on tail-type, calculate p-value using standard normal distribution.

(ix) Comparison:
compare p-value with significance level α.

(X) Conclusion:
if p-value is less than or equal to α, reject the null hypothesis, otherwise, fail to reject the null hypothesis.

→ WORKING:

Q. Let's say we need to determine if girls on average score higher than 600 in exam. We have the information that the standard deviation for girls' scores is 100. So, we collect the data of 20 girls by using random samples and record α value to be 0.05.

Date:

Scores = 650, 730, 510, 670, 480, 800, 690, 530, 590, 620, 710, 670, 640, 780, 650, 490, 800, 600, 510, 700

Soln: $n = 20$

$\sigma = 100$

$\alpha = 0.05$

$\bar{x} = 641$

Step 1: $H_0: \mu = 600$

$H_a: \mu > 600$

{Right one sided test}

Step 2: $Z_{score} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \dfrac{641 - 600}{100/\sqrt{20}}$

$= 1.8336$

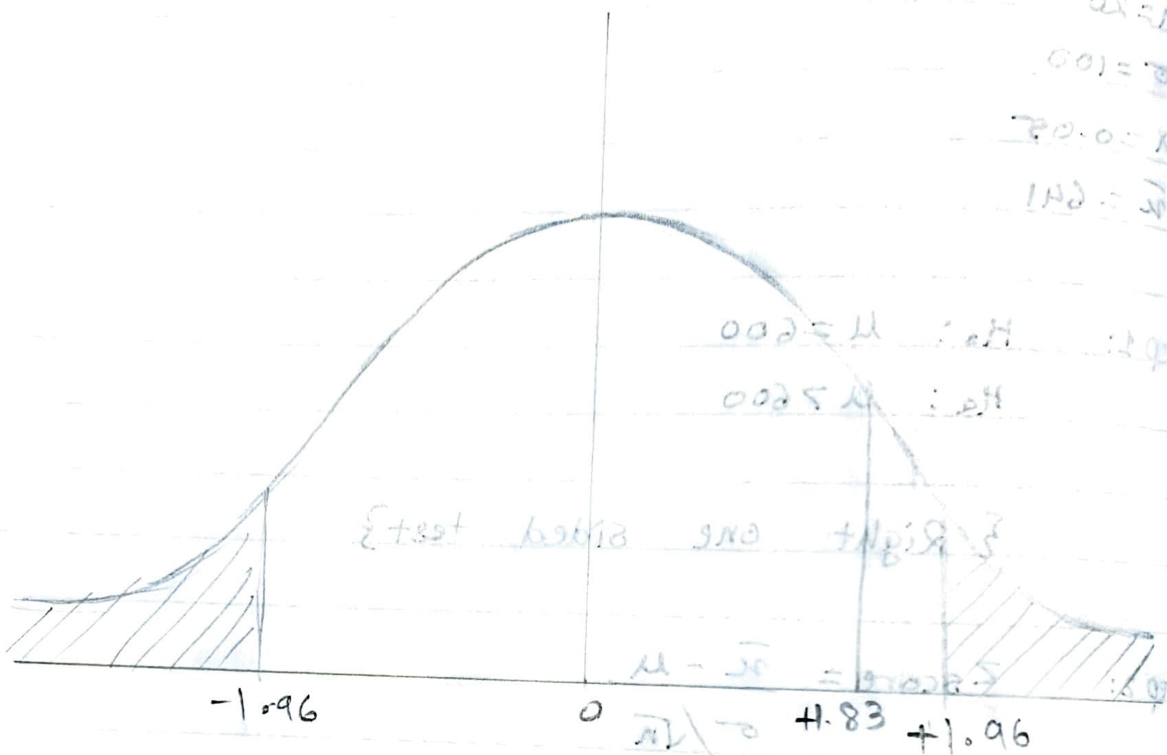
lookup table → 0.9664

Step 3: $P_{value} = 1 - Z_{score}$

$= 1 - 1.8336$

$= 0.033357$

~~0.8~~ $0.0336 < 0.05$ (α)



☐ Region of acceptance

▨ Region of rejection

Date:

H_0 is rejected

H_a : $\mu > 600$

CONCLUSION:

z-score is a metric for interential statistic using a z-test. It quantifies the standard deviation by which the sample mean differs from the population mean. The Calculated z-score is compared with critical value signifies informed decision about rejecting or accepting null hypothesis.

12/2/24

```
# One sample Z test
import scipy.stats as stats

# Sample data
scores = [650, 730, 510, 670, 480, 800, 690, 530, 590, 620, 710, 670, 640, 780

# User inputs
population_mean = float(input("Enter the population mean to formulate the null
sample_size = int(input("Enter the sample size (n): "))
alpha = 0.05  # Significance level

# Formulate null and alternative hypothesis statements
null_hypothesis = "H0: μ = {}".format(population_mean)
alternative_hypothesis = "H1: μ ≠ {}".format(population_mean)

print("Null Hypothesis:", null_hypothesis)
print("Alternative Hypothesis:", alternative_hypothesis)

# Calculate sample mean
sample_mean = sum(scores) / len(scores)
# Calculate sample standard deviation
mean_difference_squared = [(x - sample_mean) ** 2 for x in scores]
population_std = (sum(mean_difference_squared) / (len(scores) - 1)) ** 0.5
print("Sample Mean (x̄):", sample_mean)
print("Standard Deviation (σ):", population_std)

# Calculate z-score
z_score = (sample_mean - population_mean) / (population_std / (sample_size **

# Determine the type of tailed test
if z_score < 0:
    tail_type = "left-tailed"
elif z_score > 0:
    tail_type = "right-tailed"
else:
    tail_type = "two-tailed"

# Calculate p-value based on the type of tailed test
if tail_type == "left-tailed":
    p_value = stats.norm.cdf(z_score)
elif tail_type == "right-tailed":
    p_value = 1 - stats.norm.cdf(z_score)
else:
    p_value = 2 * (1 - stats.norm.cdf(abs(z_score)))

print("Z-score:", z_score)
print("P-value:", p_value)
```

```
print("Type of test:", tail_type)

# Compare p-value with significance level (alpha)
if p_value <= alpha:
    print("Reject null hypothesis")
else:
    print("Fail to reject null hypothesis")
```

```
Enter the population mean to formulate the null hypothesis (μ): 600
Enter the sample size (n): 20
Null Hypothesis: H0: μ = 600.0
Alternative Hypothesis: H1: μ ≠ 600.0
Sample Mean (x̄): 641.0
Standard Deviation (σ): 100.09995004993759
Z-score: 1.831744911595958
P-value: 0.03349471703839335
Type of test: right-tailed
Reject null hypothesis
```

```
import scipy.stats as stats

# Sample data
sample_data = [0.88, 0.66, 0.89, 0.51, 0.76, 0.9, 0.54, 0.9, 0.79, 0.55, 0.92,

# User inputs
population_mean = float(input("Enter the population mean to formulate the null
sample_size = int(input("Enter the sample size (n): "))
alpha = 0.05  # Significance level

# Formulate null and alternative hypothesis statements
null_hypothesis = "H0: μ = {}".format(population_mean)
alternative_hypothesis = "H1: μ ≠ {}".format(population_mean)

print("Null Hypothesis:", null_hypothesis)
print("Alternative Hypothesis:", alternative_hypothesis)

# Calculate sample mean
sample_mean = sum(sample_data) / len(sample_data)
# Calculate sample standard deviation
mean_difference_squared = [(x - sample_mean) ** 2 for x in sample_data]
population_std = (sum(mean_difference_squared) / (len(sample_data) - 1)) ** 0.!
print("Sample Mean (x̄):", sample_mean)
print("Standard Deviation (σ):", population_std)

# Calculate z-score
z_score = (sample_mean - population_mean) / (population_std / (sample_size **

# Determine the type of tailed test
if z_score < 0:
```

```
        tail_type = "left-tailed"
    elif z_score > 0:
        tail_type = "right-tailed"
    else:
        tail_type = "two-tailed"

    # Calculate p-value based on the type of tailed test
    if tail_type == "left-tailed":
        p_value = stats.norm.cdf(z_score)
    elif tail_type == "right-tailed":
        p_value = 1 - stats.norm.cdf(z_score)
    else:
        p_value = 2 * (1 - stats.norm.cdf(abs(z_score)))

    print("Type of test:", tail_type)
    print("Z-score:", z_score)
    print("P-value:", p_value)

    # Compare p-value with significance level (alpha)
    if p_value <= alpha:
        print("Reject null hypothesis")
    else:
        print("Fail to reject null hypothesis")
```

```
Enter the population mean to formulate the null hypothesis (μ): 0.78
Enter the sample size (n): 30
Null Hypothesis: H0: μ = 0.78
Alternative Hypothesis: H1: μ ≠ 0.78
Sample Mean (x̄): 0.7889999999999999
Standard Deviation (σ): 0.15979189051970707
Type of test: right-tailed
Z-score: 0.30849519343652076
P-value: 0.37885277855208555
Fail to reject null hypothesis
```

12/4/24