DEEPAN SHANMUGAM
RA2412044015083

# Life Long Learning Assessment

## 21CSC529T - INFERENTIAL STATISTICS

**Answer any 10 questions**

1. Calculate mean, median, mode, range, quartile deviation, standard deviation, variance and the coefficient of quartile deviation, standard deviation, coefficient of variation, Karl Pearson's coefficient of skewness, Bowley's coefficient of skewness, measures of skewness using $\gamma 1$ and measures of kurtosis using $\gamma 2$ for the following data:

| | | | | | | |
|---|---|---|---|---|---|---|
| Sales | 15 | 18 | 25 | 27 | 30 | 35 |
| Expenditu re | 50 | 65 | 82 | 95 | 110 | 120 |

Also find the Pearson's and Spearman's Correlation Co-efficient

2. A random variable X has the following distribution

| | | | | | | |
|---|---|---|---|---|---|---|
| X | -2 | -1 | 0 | 1 | 2 | 3 |
| P[X=x] | 0.1 | k | 0.2 | 0.2k | 0.3 | 0.3k |

Find (i) the value of k, (ii) the Distribution Function (CDF) (iii) P (0 < X < 3/X < 2) and

(iv) the smallest value of α for which P (X ≤ α) >½

3. Box 1 contains 1000 bulbs of which 10% are defective. Box 2 contains 2000 bulbs of which 5% are defective. Two bulbs are drawn (without replacement) from a randomly selected box. (i) Find the probability that both balls are defective and (ii) assuming that both are defective, find the probability that they came from box 1.

4. The chance of a doctor D will diagnose a disease Z correctly is 60%. The chance that the patient will survive by his treatment after a wrong diagnosis is 40% and the chance of surviving for a correct diagnosis is 70%. Find the probability of the patient surviving. A patient of doctor D, who has disease Z, survives. What is the chance that his disease was diagnosed wrongly?

5. Out of 500 companies with 4 secretaries each, how many companies would be expected to have
(a) 2 Men and 2 Women executives

(b) More than one men executive
(c) At least 2 women executives
(d) Executives of both genders
(Assume both men and women have equal probabilities)

6. The number of defective pins in a box of 100 follows the binomial probability law. The question is to find the probability that **more than 4 pins are defective** in a box. What is the probability that a box will fail to meet the guaranteed quality.

7. The life time of a certain brand of tube light may be considered as a random variable with mean 1200 hours and S.D. 250 hours. Find the probability using CLT that the average life time of 60 lights exceeds 1250 hours.

8. A salesman in a departmental store claims that at most 60% of the shoppers entering the store leaves without making a purchase. A random sample of 50 shoppers showed 35 of them without making a purchase. Are these sample results consistent with the salesman's claim?

9. The marks of 100 students in an exam are found to be normally distributed with mean 70 and S.D. 5. Estimate the number of students whose marks will be
(a) between 60 and 75.
(b) more than 75 marks.
(c) less than 65 marks.

10. Determine whether the average weight of a sample of 20 mangoes is significantly different from the population's average weight of 70 grams. The sample mean weight is 70.55 grams, and the sample standard deviation is 2.82 grams.

11. Determine if there is a significant difference in the average scores between the two teams. The following data is given:
Team A: Score: 65, 68, 70, 63, 67
Team B: Score: 62, 66, 69, 64, 68

12. You need to assess the effectiveness of a new teaching scheme by comparing the test scores of the same group of students before and after the implementation of the scheme. The following data is given:
Before (new teaching scheme scores): 76, 88, 65, 56, 76
After (new teaching scheme scores): 85, 95, 75, 60, 81
Determine if there is a significant difference in the average test scores before and after the implementation of the scheme.

13. Customers are surveyed by a company to determine whether their age group (under 20, 20-40, over 40) and their preferred product category (food, apparel, or electronics) are related. The information gathered is:
   - Under 20: Electronic - 50, Clothing - 30, Food - 20
   - 20-40: Electronic - 60, Clothing - 70, Food - 50
   - Over 40: Electronic - 30, Clothing - 40, Food - 80

14. A researcher wants to determine whether three different teaching methods (A, B, and C) have different effects on students' test scores. She randomly assigns students to one of the three teaching methods and records their test scores. The data is as follows:
   - Method A: 85,90,78,92,88
   - Method B: 72,75,80,68,74
   - Method C: 88,85,84,90,86

15. In an experiment to see whether the amount of coverage of light-blue interior latex paint depends either on the brand of paint or on the brand of roller used, one gallon of each of four brands of paint was applied using each of three brands of roller, resulting in the following data (number of square feet covered). (a) Test whether the Roller Brands differ with respect to treatment. (b) Test whether the Paint Brands differ with respect to treatment. (c) Test whether the roller brands are the same for the different paint brands.

| **Coating** | **Soil Type** | | |
|---|---|---|---|
| | **A** | **B** | **C** |
| I | 454 | 446 | 451 |
| II | 446 | 444 | 447 |
| III | 439 | 442 | 444 |
| IV | 444 | 437 | 443 |

DEEPAN SHANMUGAM
RA2412044015083
INFERENTIAL STATISTICS

(i) Mean

For sales (15, 18, 25, 27, 30, 35).

$$\text{Mean} = \frac{15+18+25+27+30+35}{6}$$

$$\bar{x} = 25$$

(ii) Median

Arranging in order - 15, 18, 25, 27, 30, 35.

$n = 6$.

$$\text{Median} = \frac{25+27}{2} = 26.$$

(iii) Mode:

- No value repeats, so no mode.
- Multimodal.

(iv) Range:

Highest - Lowest = 35 − 15 = 20.

(v). Quartiles:

→ $Q_1$ : $(15+18)/2 = 16.5$

→ $Q_2$ (Median) = $\frac{25+27}{2} = 26.$

DEEPAN SHANMUGAM
RA2412044015083

$Q_3 : (30+27)/2 = 28.5$

Quartile deviation $= (Q_3 - Q_1)/2 = (28.5 - 16.5)/2$
$= 6$

(vi) Standard deviation & Variance.

| $X$ | $(X-\bar{X})$ $(\bar{x}=25)$ | $(X-\bar{X})^2$ | $(X-\bar{X})^3$ | $(X-\bar{X})^4$ |
|---|---|---|---|---|
| 15 | -10 | 100 | -1000 | 10000 |
| 18 | -7 | 49 | -343 | 2401 |
| 25 | 0 | 0 | 0 | 0 |
| 27 | 2 | 4 | 8 | 16 |
| 30 | 5 | 25 | 125 | 625 |
| 35 | 10 | 100 | 1000 | 10000 |
| | 0 | 278 | -210 | 23042 |

$$\text{Variance} = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{278}{5} = 55.6$$

$$\text{Standard deviation} = \sqrt{55.6} = 7.45$$

(vii) Coefficient of variation:

$$- \ CV = \left(\frac{SD}{Mean}\right) \times 100$$

$$= \frac{7.45}{25} \times 100$$

$$= 29.8\%$$

DEEPAN SHANMUGAM
RA2412044015083

Co-efficient of quartile deviation:

$$\Rightarrow \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)} = \frac{28.5 - 16.5}{28.5 + 16.5} = \frac{12}{45} = 0.267.$$

(ix) Karl Pearson's Co-efficient of Skewness:

$$\Rightarrow \frac{3\ (\text{Mean} - \text{Median})}{\text{Standard deviation.}}$$

$$= \frac{3(25 - 26)}{7.45}$$

$$= -0.402.$$

(x). Bowley's Co-efficient of Skewness.

$$\Rightarrow \frac{(Q_3 + Q_1 - 2Q_2)}{(Q_3 - Q_1)}$$

$$= \frac{(28.5 + 16.5 - 2(26))}{(28.5 - 16.5)}$$

$$= (45 - 52)/12$$

$$= -0.583$$

(xi) Measure of Skewness: calculated in step 6.

$$\Rightarrow \frac{\Sigma(x - \bar{x})^3}{n} = \frac{-210}{6} = -35$$

DEEPAN SHANMUGAM
RA2412044015083

$$\gamma_1 = \frac{m_3}{(SD)^3} = \frac{-35}{(7.45)^3} = -0.084.$$

(xii) $\gamma_2$ (Measure of kurtosis):

$$m_4 = \frac{\Sigma(x-\bar{x})^4}{n}$$

$$= \frac{23042}{6} = 3840.33.$$

$$\gamma_2 = (m_4 / (SD)^4) - 3$$

$$= \cancel{3840.33} - 3$$

$$= \frac{3840.33}{(7.45)^4} - 3$$

$$= 2.02$$

For expenditure data

(50, 65, 82, 95, 110, 120)

(i) Mean = $\frac{50+65+82+95+110+120}{6}$

$= 87$

(ii) Median ⇒ 50, 65, $\underline{82}$, $\underline{95}$, 110, 120

$= \frac{82+95}{2} = 88.5$

Mode:
N

DEEPAN SHANMUGAM
RA2412044015083

Mode:

No mode (multimodal)

(iv) Range:

$120-50=70$

(v) Quartiles:

→ $Q_1 = (50+65)/2 = 57.5$

→ $Q_2 = 88.5$

→ $Q_3 = \frac{(110+95)}{2} = 102.5$

→ Quartile deviation $= \frac{(102.5-57.5)}{2}$

$= 22.5$

(vi) Standard deviation and variance:

$\bar{x} = 87$

| $x$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(x-\bar{x})^3$ | $(x-\bar{x})^4$ |
|---|---|---|---|---|
| 50 | -37 | 1369 | | |
| 65 | -22 | 484 | | |
| 82 | -5 | 25 | | |
| 95 | 8 | 64 | | |
| 110 | 23 | 529 | | |
| 120 | 33 | 1089 | | |
| | | 3560 | | |

DEEPAN SHANMUGA
RA2412044015083

$$\text{Variance} = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{3560}{5} = 712.$$

$$\text{Standard deviation} = \sqrt{712} = 26.68.$$

(vii) Co-efficient of variation:

$$\Rightarrow CV = \left(\frac{\text{Standard deviation}}{\text{Mean}}\right) \times 100$$

$$= \left(\frac{26.68}{87}\right) \times 100 = 30.67\%$$

(viii) Co-efficient of quartile deviation:

$$\Rightarrow \frac{(Q_3-Q_1)}{(Q_3+Q_1)} = \frac{102.5-57.5}{102.5+57.5} = \frac{45}{160} = 0.281$$

(ix) Karl Pearson's co-efficient of skewness:

$$\Rightarrow \frac{3(\text{Mean}-\text{Median})}{\text{Standard Deviation}} = \frac{3(87-88.5)}{26.68}$$

$$= -0.169$$

x) Bowley's Co-efficient of skewness:

$$\Rightarrow (Q_3+Q_1-2Q_2)/Q_3-Q_1$$

$$= \frac{(102.5+57.5-2(88.5))}{(102.5-57.5)} = -0.378$$

DEEPAN SHANMUGAM
RA2412044015083

Pearson's Correlation

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\left(\sqrt{\Sigma (x-\bar{x})^2}\right)\cdot\left(\sqrt{\Sigma (y-\bar{y})^2}\right)}$$

| x | $x-\bar{x}$ | y | $y-\bar{y}$ | $(y-\bar{y})(x-\bar{x})$ |
|---|---|---|---|---|
| 15 | ~~15-25~~ -10 | 50 | -37 | 370 |
| 18 | -7 | 65 | -22 | 154 |
| 25 | 0 | 82 | -5 | 0 |
| 27 | 2 | 95 | 8 | 16 |
| 30 | 5 | 110 | 23 | 115 |
| 35 | 10 | 120 | 33 | 330 |
| | | | | 985 |

We have already calculated $\Sigma(x-\bar{x})^2$ and $\Sigma(y-\bar{y})^2$

$\Sigma (x-\bar{x})^2 = 278$

$\Sigma (y-\bar{y})^2 = 3560$

$$r = \frac{985}{\sqrt{278 \times 3560}} = 0.99.$$

DEEPAN SHANMUGAM
RA2412044015083

## Spearman's correlation :

| Sales | Sales Rank | Expenditure | Expenditure Rank | d (SR - ER) | $d^2$ |
|---|---|---|---|---|---|
| 15 | 1 | 50 | 1 | 0 | 0 |
| 18 | 2 | 65 | 2 | 0 | 0 |
| 25 | 3 | 82 | 3 | 0 | 0 |
| 27 | 4 | 95 | 4 | 0 | 0 |
| 30 | 5 | 110 | 5 | 0 | 0 |
| 35 | 6 | 120 | 6 | 0 | 0 |

$$r_s = 1 - \frac{(6\sum d^2)}{n(n^2-1)} = 1 - \frac{6\times 0}{6(36-1)}$$

$$= 1 - \frac{6\times 0}{210}$$

$$r_s = 1$$

Pearson's r indicates near perfect linear correlation of $\underline{0.99}$.

Spearman's $r_s$ indicates perfect rank correlation of $\underline{1.00}$

DEEPAN SHANMUGAM
RA2412044015083

Box 1 ⇒ 1000 bulbs.
10% defective

Therefore, defective bulbs $= \frac{10}{100} \times 1000$
$= 100$

Good bulbs $= 1000 - 100$
$= 900.$

Box 2 ⇒ 2000 bulbs
5% defective

Therefor, defective bulbs $= \frac{5}{100} \times 2000$
$= 100$

Good bulbs $= 2000 - 100$
$= 1900.$

In From box 1:

$P(\text{first defective}) = \frac{100}{1000} = 0.1$

without replacement ⇒ $P(\text{second defective}) = \frac{99}{999} = 0.099$

From box 2:

$P(\text{first defective}) = \frac{100}{2000} = 0.05$

$P(\text{second defective}) = \frac{99}{1999} = 0.0495$

DEEPAN SHANMUHAM
RA2412044015083

P( corre
P(

Probability of selecting each box :

$P(Box1) = 1/2$

$P(Box2) = 1/2$.

(i) Total probability = P(Box 1) × P(first defective in box 1) × P(second defective in box 1)

+ P(Box 2) × P(first defective in box 2) × P(second defective in box 2)

$= \frac{1}{2}(0.1)(0.099) + \frac{1}{2}(0.05)(0.0495)$

$= 0.00495 + 0.001238$

$= 0.006188$

$= 0.62\%$.

(ii) P(Box 1 | both are defective)

Baye's Theorem = $\frac{P(\text{both are defective} \mid \text{Box 1}) \cdot P(\text{Box 1})}{P(\text{both are defective})}$

$= \frac{(0.1 \times 0.099) \times 0.5}{0.006188}$

$= \frac{0.00495}{0.006188} = 0.80\ (80\%)$.

DEEPAN SHANMUGAM
RA2412044015083

P( correct diagnosis) = 0.60

P( wrong diagnosis) = 0.40.

P( survive | wrong diagnosis) = 0.40

P( survive | correct diagnosis) = 0.70.

(i) P( survive) – using law of total probability.

P( survieve ) = P( survive | wrong diagnosis) × P( ~~correct~~ wrong)

+ P( survive | correct diagnosis) × P( correct)

= 0.40 × 0.40 + 0.70 × 0.60

= 0.16 + 0.42

= 0.58 (58%)

(ii). P ( wrong diagnosis | survive)

using Bayes' theorem:

$$P(\text{wrong} \mid \text{survive}) = \frac{P(\text{survive} \mid \text{wrong}) \cdot P(\text{wrong})}{P(\text{survive})}$$

$$= \frac{\cancel{0.5}\ 0.40 \times 0.40}{0.58}$$

= 0.276 (27.6%)

DEEPAN SHANMUGAM
RA2412044015083

5) (a). P(man) = P(woman) = 0.5

Binomial probability for n = 4 (Two men + Two wom

$P(x=r) = nC_r \times P^r \times q^{(n-r)}$

n = 4
r = number of successes (2 men in this case)
p = probability of success (0.5 for men in this case)
q = probability of failure (0.5 for woman in this case)

$= 4C_2 \times (0.5)^2 \times (0.5)^{(4-2)}$

$= \frac{4 \times 3}{2 \times 1} \times 0.25 \times 0.25$

$= 0.375$

Expected companies to have 2 men and 2 women executives = 0.375 × 500

= 187.5

≈ 188 companies

(b). More than one men executive:

$\Rightarrow P(x \geq 2) = 1 - [P(x=0) + P(x=1)]$

$P(x=0) = 4C_0 \times (0.5)^4 = 1 \times 0.0625$

$= 0.0625$

DEEPAN SHANMUGAM
RA2412044015083

$P(x=1) = 4C_1 \times (0.5)^1 \times (0.5)^3$

$= 4 \times 0.5 \times 0.125$

$= 0.25$

So, $P(x \geq 2) = 1-(0.0625+0.25)$

$= 0.6875$

Expected number $= 500 \times 0.6875$

$= 343.75$

$\approx 344$ companies

c). At least 2 women executives:

Since $P(\text{man}) = P(\text{woman}) = 0.5$,

the answer will be the same as for

atleast 2 men,

$P(x \geq 2) = 1-(0.0625+0.25)$

$E(x \geq 2) = 343.75$

$\approx 344$ Companies.

d). Executives of both genders.

$P(\text{both genders}) = 1-(P(\text{all men})+P(\text{all women}))$

$P(\text{all men}) = 4C_4 \times (0.5)^4 \times 1 = 0.0625$

$P(\text{all women}) = 4C_4 \times (1) \times (0.5)^4 = 0.0625$

DEEPAN SHANMUGA
RA2412044015083

$P(\text{both}) = 1 - 0.125 = 0.875$.

Expected number $= 500 \times 0.875$

$= 437.5$.

$\approx 438$ companies.

6). $n = 100$ (pins).

Assume $p = 0.05$.

$q = 0.95$.

$r = 0, 1, 2, 3, 4$.

Need to calculate $P(X > 4)$.

$$P(X>4) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)]$$

$$= 1 - [100C_0 \times (0.05)^0 \times (0.95)^{100-0} + 100C_1 \times (0.05)^1 \times (0.95)^{100-1} + 100C_2 \times (0.05)^2 \times (0.95)^{100-2} + 100C_3 \times (0.05)^3 \times (0.95)^{100-3} + 100C_4 \times (0.05)^4 \times (0.95)^{100-4}]$$

$= 0.0485$

$= 4.85\%$

Probability that it will fail to meet the guaranteed quality of 95% is 4.85%

DEEPAN SHANMUGAM
RA2412044015083

Population mean ($\mu$) = 1200 hours.

Population SD ($\sigma$) = 250 hours

Sample size ($n$) = 60

Since sample > 30.

$P(\bar{x} > 1250) = ?$

$$\text{Standard error (SE)} = \frac{\sigma}{\sqrt{n}} = \frac{250}{\sqrt{60}} = 32.27$$

$$Z = \frac{\text{Expected Mean} - \text{Population Mean}}{\text{Standard Error}}$$

$$= \frac{1250 - 1200}{32.27}$$

$$= 1.55$$

using Z-table.

$$P(Z > 1.55) = 0.5 - 0.4394$$

$$= 0.0606$$

$$= 6.06\,\%$$

(8). $H_0 : P_0 \leq 0.60$ (proportion of shoppers leaving without purchasing)

$H_1 : P > 0.60$ (we are trying to reject $H_0$).

~~Standard error~~

DEEPAN SHANMUGAM
RA2412044015083

Num

(a)

Sample proportion $(\hat{p}) = 35/50 = 0.70$.

Standard Error:

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$= \sqrt{\frac{0.6 \times 0.4}{50}}$$

$$= \sqrt{\frac{0.24}{50}} = 0.0693$$

Z-Score:

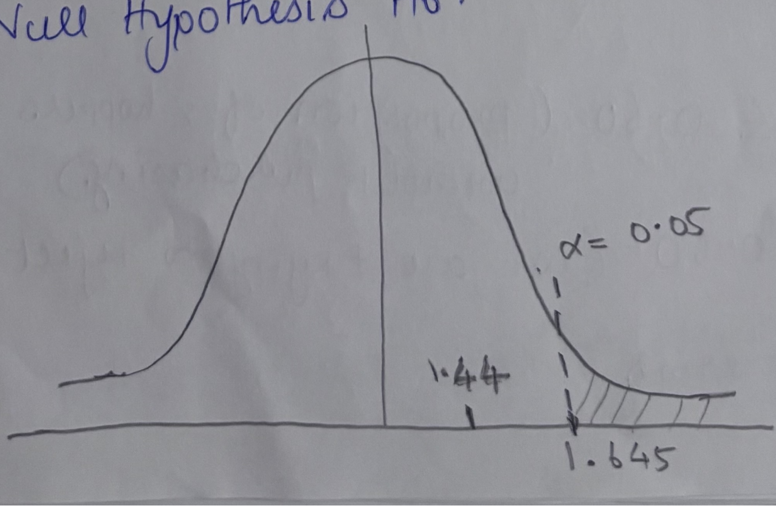
$$Z = (\hat{p} - p_0)/SE$$

$$= \frac{(0.70 - 0.60)}{0.0693}$$

$$= 1.44$$

At $\alpha = 0.05$, critical value $= 1.645$.

Since $1.44 < 1.645$, we fail to reject the salesman's claim meaning & we fail to reject the Null Hypothesis $H_0$.



DEEPAN SHANMUGAM
RA2412044015083

(a) Number of students = 100

Mean $\mu = 70$

Standard deviation $\sigma = 5$.

Between 60 and 75,

Z-score for 60 = $\frac{x-\mu}{\sigma} = \frac{60-70}{5} = -2$

Z-score for 75 = $\frac{x-\mu}{\sigma} = \frac{75-70}{5} = 1$.

$P(-2 < Z < 1) = [0.8413 + 0.4772] = 0.8415$.

Number of students = $100 \times 0.8415$

$= 84.15$

$\approx 84$ students.

(b). More than 75:

Z-score for 75 = $\frac{75-70}{5} = 1$.

$P(z>1) = 0.5 - 0.3413$ → z table value.

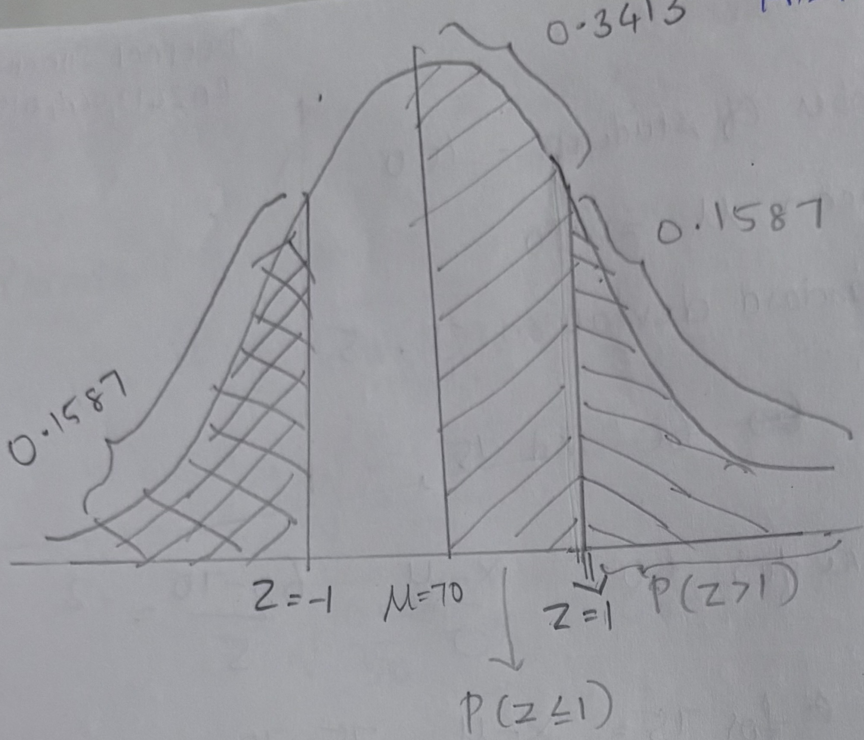
$= 0.1587$.

Number of students = $100 \times 0.1587$

$= 15.87$

$\approx 16$ students

DEEPAN SHANMU
RA2412044015083




c) Less than 65 marks.

Z-score for 65 : $\frac{65-70}{5} = -1$.

$P(Z < -1) = 0.1587$

Number of students $= 100 \times 0.1587$

$= 15.87$

$\simeq 16.$ students.

10). Population mean $(\mu_0) = 70$ gms

Sample size $(n) = 20$

Sample mean $(\bar{x}) = 70.55$ gms

DEEPAN SHANMUGAM
RA2412044018083

Standard deviation ($s$) = 2.82 g.

Since $n < 30$, t-test.

$H_0 : \mu_0 = 70$.

$H_1 : \mu \neq 70$.

$$SE = \frac{s}{\sqrt{n}} = \frac{2.82}{\sqrt{20}} = 0.63.$$

$$t = \frac{\bar{x} - \mu_0}{SE} = \frac{70.55 - 70}{0.63}$$

$$= 0.873$$

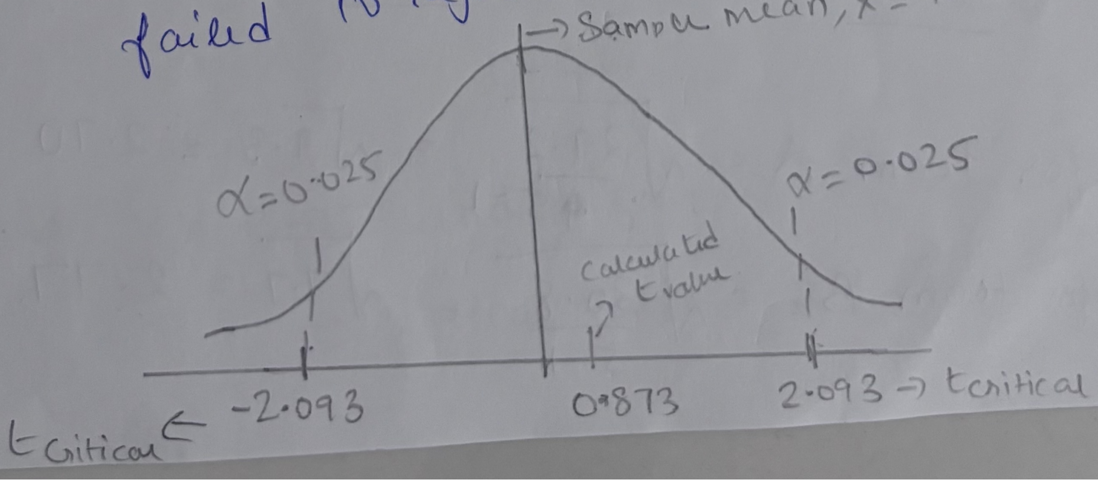
$\alpha = 0.05$ and $df = (n-1) = 20 - 1 = 19$

from t-table, critical value = $\pm 2.093$ (two tailed)

~~Since~~

Since $|t| < t_{critical}$, $|0.873| < 2.093$,

failed to reject $H_0$.



DEEPAN SHANMUGAM
RA2412044015083

11). Team A: 65, 68, 70, 63, 67.

$$\text{Mean } (\bar{x}_1) = \frac{65+68+70+63+67}{5}$$

$$= 66.6$$

Team B: 62, 66, 69, 64, 68.

$$\text{Mean } (\bar{x}_2) = \frac{62+66+69+64+68}{5}$$

$$= 65.8$$

| $x_1$ | $(x_1-\bar{x}_1)^2$ | $x_2$ | $(x_2-\bar{x}_2)^2$ |
|---|---|---|---|
| 65 | 2.56 | 62 | 14.44 |
| 68 | 1.96 | 66 | 0.04 |
| 70 | 11.56 | 69 | 10.24 |
| 63 | 12.96 | 64 | 3.24 |
| 67 | 0.16 | 68 | 4.84 |
| | 29.2 | | 30.8 |

$$S_1, \text{ SD for team A} = \sqrt{\frac{29.2}{\cancel{4}5-1}} = \sqrt{7.3} = 2.70$$

$$S_2, \text{ SD for team B} = \sqrt{\frac{30.8}{5-1}} = \sqrt{7.7} = 2.77.$$

DEEPAN SHANMUGAM
RA2412044015083

Pooled standard deviation,

$$= \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{(n_1+n_2-2)}}$$

$$= \sqrt{\frac{(2.70)^2(4) + (2.77)^2(4)}{(10-2)}}$$

$Sp = 2.74.$

$$\text{T-statistic} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{Sp^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{66.6 - 65.8}{\sqrt{2.74^2\left(\frac{1}{5}+\frac{1}{5}\right)}}$$

$$= \frac{0.8}{1.73}$$

$$= 0.46.$$

At $\alpha = 0.05$ and $df = (10-2) = 8$,

$\rightarrow$ Critical value $= \pm 2.306.$

DEEPAN SHANMUGAM
RA2412044015008-3

→ Since $|t_{statistic}| < t_{critical}$, $|0.46| < 2.306$,
failed to reject $H_0$.

Hence there is no significant difference in average scores between the two teams.