B.TECH. (TT/TE) V SEMESTER

SUBJECT – PRACTICAL APPLICATION OF STATISTICS

MODEL QUESTION PAPER WITH ANSWER KEY

MID TERM I

1. For the given set of data, apply t test and find whether the difference between the means is significant or not.

Table 1: Scores for First Graders

Experimental	Comparison
35	2
40	27
12	38
15	31
21	1
14	19
46	1
10	34
28	3
48	1
16	2
30	3
32	2
48	1
31	2
22	1
12	3
39	29
19	37
25	2
Mean = 27.15	11.95
SD = 12.51	14.62

Formula for T-test for indepentdent groups

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{var_1}{n} + \frac{var_2}{n}}}$$ Substituting our values: $$t = \frac{27.15 - 11.95}{\sqrt{\frac{12.5^2}{20} + \frac{14.6^2}{20}}}$$

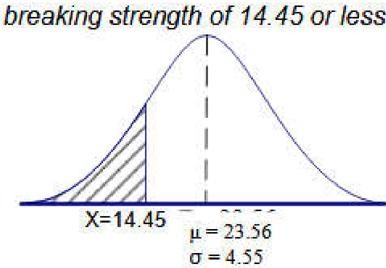
$$t = \frac{27.15 - 11.95}{\sqrt{\frac{12.5^2}{20} + \frac{14.6^2}{20}}} = \frac{15.2}{\sqrt{\frac{156.25}{20} + \frac{213.45}{20}}} = \frac{15.2}{\sqrt{7.81 + 10.67}} = \frac{15.2}{\sqrt{18.48}} = \frac{15.2}{4.298} = 3.54$$

Our obtained, or calculated t value is 3.54. Our degrees of freedom equals the total group size (40) minus 2, or 38. Entering a t table with 38 degrees of freedom, we see that for alpha = .05 the tabled value is 2.03 and for alpha = .01, the tabled value is 2.72.

Our calculated value is larger than the tabled value at alpha = .01, so we reject the null hypothesis and accept the alternative hypothesis, namely, that the difference in gain scores is likely the result of the experimental treatment and not the result of chance variation.

Wool fibre breaking strengths are normally distributed with mean μ = 23.56 Newtons and standard deviation, σ = 4.55.
What proportion of fibres would have a breaking strength of 14.45 or less?

- **Draw a diagram, label and shade area required:**



- **Convert** raw score (X) to standard score (Z) : $Z = \frac{14.45 - 23.56}{4.55} = -2.0$

  That is, the raw score of 14.45 is equivalent to a standard score of -2.0.
  It is negative because it is on the left hand side of the curve.

- **Use tables** to find probability and adjust this result to required probability:

$$p(X < 14.45) = p(Z < -2.0) = 0.5 - p(0 < Z < 2)$$
$$= 0.5 - 0.4772$$
$$= 0.0228$$

Suppose the National Transportation Safety Board (NTSB) wants to examine the safety of compact cars, midsize cars, and full-size cars. It collects a sample of three for each of the treatments (cars types). Using the hypothetical data provided below, test whether the mean pressure applied to the driver's head during a crash test is equal for each types of car. Use $\alpha =$ 5%.

Table ANOVA.1

	Compact cars	Midsize cars	Full-size cars
	643	469	484
	655	427	456
	702	525	402
$\overline{X}$	666.67	473.67	447.33
S	31.18	49.17	41.68

(1.) State the null and alternative hypotheses
The null hypothesis for an ANOVA always assumes the population means are equal. Hence, we may write the null hypothesis as:

H_0: $\mu_1 = \mu_2 = \mu_3$ - The mean head pressure is statistically equal across the three types of cars.

Since the null hypothesis assumes all the means are equal, we could reject the null hypothesis if only mean is not equal. Thus, the alternative hypothesis is:
H_a: At least one mean pressure is not statistically equal.

(2.) Calculate the appropriate test statistic
The test statistic in ANOVA is the ratio of the *between* and *within* variation in the data. It follows an F distribution.

Total Sum of Squares – the total variation in the data. It is the sum of the between and within variation.

Total Sum of Squares (SST) = $\sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij} - \overline{\overline{X}})^2$, where r is the number of rows in the table, c is the number of columns, $\overline{\overline{X}}$ is the grand mean, and X_{ij} is the i th observation in the j th column.

Using the data in Table ANOVA.1 we may find the grand mean:

$$\overline{\overline{X}} = \frac{\sum X_{ij}}{N} = \frac{(643+655+702+469+427+525+484+456+402)}{9} = 529.22$$

SST =

$$(643-529.22)^2 + (655-529.22)^2 + (702-529.22)^2 + (469-529.22)^2 + \ldots + (402-529.22)^2 = 96303.55$$

Between Sum of Squares (or Treatment Sum of Squares) – variation in the data between the different samples (or treatments).

Treatment Sum of Squares (SSTR) = $\sum r_j(\overline{X}_j - \overline{\overline{X}})^2$, where r_j is the number of rows in the j th treatment and $\overline{X}_j$ is the mean of the j th treatment.

Using the data in Table ANOVA.1,

SSTR = $[3*(666.67-529.22)^2] + [3*(473.67-529.22)^2] + [3*(447.33-529.22)^2] = 86049.55$

Error Sum of Squares (SSE) = $\sum\sum(X_{ij} - \bar{X}_j)^2$

From Table ANOVA.1,

SSE=$[(643-666.67)^2 + (655-666.67)^2 + (702-666.67)^2]+$
$[(469-473.67)^2 + (427-473.67)^2 + (525-473.67)^2]+$
$[(484-447.33)^2 + (456-447.33)^2 + (402-447.33)^2] = 10254.$

Note that SST = SSTR + SSE (96303.55 = 86049.55 + 10254).

Hence, you only need to compute any two of three sources of variation to conduct an ANOVA. Especially for the first few problems you work out, you should calculate all three for practice.

The next step in an ANOVA is to compute the "average" sources of variation in the data using SST, SSTR, and SSE.

Total Mean Squares (MST) $= \dfrac{SST}{N-1}$ → "average total variation in the data" (N is the total number of observations)

MST = $\dfrac{96303.55}{(9-1)} = 12037.94$

Mean Square Treatment (MSTR) = $\dfrac{SSTR}{c-1}$ → "average between variation" (c is the number of columns in the data table)

MSTR = $\dfrac{86049.55}{(3-1)} = 43024.78$

Mean Square Error (MSE) = $\dfrac{SSE}{N-c}$ → "average within variation"

MSE = $\dfrac{10254}{(9-3)} = 1709$

Note: MST $\neq$ MSTR + MSE

The test statistic may now be calculated. For a one-way ANOVA the test statistic is equal to the ratio of MSTR and MSE. This is the ratio of the "average between variation" to the "average within variation." In addition, this ratio is known to follow an F distribution. Hence,

$F = \frac{MSTR}{MSE} = \frac{43024.78}{1709} = 25.17$. The intuition here is relatively straightforward. If the average between variation rises relative to the average within variation, the F statistic will rise and so will our chance of rejecting the null hypothesis.

(3.) Obtain the Critical Value
To find the critical value from an F distribution you must know the numerator (MSTR) and denominator (MSE) degrees of freedom, along with the significance level.

F^{CV} has df1 and df2 degrees of freedom, where df1 is the numerator degrees of freedom equal to c-1 and df2 is the denominator degrees of freedom equal to N-c.

In our example, df1 = 3 - 1 = 2 and df2 = 9 - 3 = 6. Hence we need to find $F_{2,6}^{CV}$ corresponding to α = 5%. Using the F tables in your text we determine that $F_{2,6}^{CV} = 5.14$.

(4.) Decision Rule
You reject the null hypothesis if: F (observed value) > F^{CV} (critical value). In our example 25.17 > 5.14, so we reject the null hypothesis.

(5.) Interpretation
Since we rejected the null hypothesis, we are 95% confident (1-α) that the mean head pressure is not statistically equal for compact, midsize, and full size cars.

. Consider the following set of points: {(-2 , -1) , (1 , 1) , (3 , 2)}

a) Find the least square regression line for the given data points.

b) Plot the given points and the regression line in the same rectangualr system of axes.

a) Let us organize the data in a table.

x	y	x y	x^2
-2	-1	2	4
1	1	1	1
3	2	6	9
$\Sigma x = 2$	$\Sigma y = 2$	$\Sigma xy = 9$	$\Sigma x^2 = 14$

We now use the above formula to calculate a and b as follows

$a = (n\Sigma x\ y - \Sigma x\Sigma y) / (n\Sigma x^2 - (\Sigma x)^2) = (3*9 - 2*2) / (3*14 - 2^2) = 23/38$

$b = (1/n)(\Sigma y - a\ \Sigma x) = (1/3)(2 - (23/38)*2) = 5/19$

b) We now graph the regression line given by y = ax + b and the given points.