T TEST OF LARGE, SMALL &CORRELATED DATA

KHAGENDRA KUMAR

T TEST

- IN STATISTICS, THE TERM “T-TEST” REFERS TO THE HYPOTHESIS TEST IN WHICH THE TEST STATISTIC FOLLOWS A GROUP’S T-DISTRIBUTION. IT IS USED TO CHECK WHETHER TWO DATA SETS ARE SIGNIFICANTLY DIFFERENT FROM EACH OTHER OR NOT.
-

One Sample t test
one-sample t-test which is used to determine if the sample is significantly different from the population.

$$t = (\bar{x} - \mu) / (s / \sqrt{n})$$

$\bar{x}$ = Observed Mean of the Sample
μ = Theoretical Mean of the Population
s = Standard Deviation of the Sample
n = Sample Size

2 SAMPLES T TEST

- IN CASE STATISTICS OF TWO SAMPLES ARE TO BE COMPARED, THEN A TWO-SAMPLE T-TEST IS TO BE USED AND ITS FORMULA IS EXPRESSED USING RESPECTIVE SAMPLE MEANS, SAMPLE STANDARD DEVIATIONS, AND SAMPLE SIZES. MATHEMATICALLY, IT IS REPRESENTED AS,

$$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{[(s^2_1 / n_1) + (s^2_2 / n_2)]}$$

Where,

- **$\bar{x}_1$** = Observed Mean of 1st Sample
- **$\bar{x}_2$** = Observed Mean of 2nd Sample
- **s_1** = Standard Deviation of 1st Sample
- **s_2** = Standard Deviation of 2nd Sample
- **n_1** = Size of 1st Sample
- **n_2** = Size of 2nd Sample

Formula for small samples

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_1^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

For large sample (earlier discussed)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

PURPOSE (CORRELATED T TEST)

- THE CORRELATED T-TEST ALLOWS TO CONSIDER DIFFERENCES BETWEEN TWO GROUPS OR SETS OF SCORES THAT ARE RELATED.

LIKELY CONDITIONS FOR CORRELATED OR DEPENDENT GROUPS/SAMPLES

CONDITION 1 BEFORE-AFTER STUDIES

- THESE TYPES OF DATA OCCUR MOST OFTEN WITH PRETEST-TREATMENT-POSTTEST EXPERIMENTAL DESIGNS. THE PRETEST AND POSTTEST SCORES ARE RELATED, OR CORRELATED, BECAUSE THE SCORES ARE TAKEN FROM THE SAME INDIVIDUALS, I.E., EACH PERSON IS MEASURED TWICE.

EXAMPLES:

- A STUDENT TAKES THE TET, ENROLLS IN A TET ENHANCEMENT CLASS, AND THEN RETAKES THE TET. TWO SCORES FROM THE SAME STUDENT EXIST.
- A TEACHER MEASURED THE READING PERFORMANCE OF A THIRD-GRADER, PRESENTED SOME TREATMENT DESIGNED TO INCREASE READING PERFORMANCE, THEN RE-MEASURED THE STUDENT'S READING PERFORMANCE AGAIN (TWO SCORES FROM SAME INDIVIDUAL.

CONDITION 2 NATURAL PAIRS

- NATURALLY OCCURRING PAIRS, NATURAL PAIRS, SUCH AS HUSBANDS AND WIVES, TWINS, BROTHERS, SISTERS, BROTHERS AND SISTERS, PARENTS AND THEIR CHILDREN, ETC. WITH NATURALLY OCCURRING PAIRS, ONE WOULD EXPECT THE PAIRS TO HOLD SIMILAR FEELINGS, BELIEFS, ATTITUDES, ETC., SO THEIR SCORES WILL GENERALLY BE RELATED.

EXAMPLES

- DETERMINING WHETHER HUSBANDS' ATTITUDES TOWARD POLITICS ARE SIMILAR TO THEIR WIVES' ATTITUDES. SINCE PEOPLE TEND TO MARRY OTHERS LIKE THEMSELVES, ONE WOULD EXPECT THAT MOST HUSBANDS AND WIVES TO HOLD SIMILAR POLITICAL VIEWS.
- DETERMINING WHETHER BOYS' IQ DIFFERS FROM GIRLS' IQ. SINCE BROTHERS AND SISTERS ARE SIMILAR GENETICALLY, ONE MIGHT ANTICIPATE THE TWO TO HAVE SIMILAR IQS, THAT IS, THEIR IQS ARE LIKELY TO BE RELATED; THEREFORE, BROTHERS AND SISTERS SHOULD BE MATCHED.

CONDITION 3: MATCHED-PAIRS

- TWO SETS OR GROUPS OF SCORES ARE INVOLVED IN THE STUDY, AND SETS ARE MATCHED ON SOME EXTRANEOUS VARIABLE (OR PERHAPS MULTIPLE VARIABLES) THAT IS RELATED TO THE DEPENDENT VARIABLE EXAMINED.

EXAMPLE

- A TEACHER IS INTERESTED IN DETERMINING WHETHER "HOOKED ON PHONICS" INCREASES THIRD-GRADE STUDENTS' READING PERFORMANCE. USING TWO GROUPS OF STUDENTS, GROUP A (THE EXPERIMENTAL GROUP) WILL USE "HOOK ON PHONICS" FOR ONE MONTH, AND GROUP B (THE CONTROL) WILL BE EXPOSED TO THE USUAL READING LESSONS DURING THE MONTH.
- THE TEACHER KNOWS THAT IQ INFLUENCES READING PERFORMANCE, SO TO CONTROL FOR THE EFFECTS OF IQ ON THE DEPENDENT VARIABLE (WHICH IS A POSTTEST ON READING PERFORMANCE), THE RESEARCHER MATCHES STUDENTS IN THE TWO GROUPS ON THEIR UNIQUE IQ LEVELS .

HYPOTHESIS

- THE HYPOTHESIS TESTED WITH THE CORRELATED T-TEST IS THE SAME AS IN THE INDEPENDENT-SAMPLES T-TEST.

EXAMPLE 1- IS THERE A DIFFERENCE IN MEAN SYSTOLIC BLOOD PRESSURE BEFORE AND AFTER TAKING 25MG OF METOLAR?

- NULL HYPOTHESIS- THE MEAN SYSTOLIC BLOOD PRESSURE IS THE SAME BEFORE AND AFTER TAKING LOSARTAN.
- ALTERNATIVE - THE MEAN SYSTOLIC BLOOD PRESSURE BEFORE AND AFTER TAKING METOLAR DIFFERS.

EXAMPLE 2 - DO IDENTICAL TWINS, SEPARATED AT BIRTH AND RAISED BY DIFFERENT FAMILIES, SHOW SIMILAR IQ SCORES?

- NULL HYPO- THERE IS NO DIFFERENCE IN MEAN IQ SCORES BETWEEN TWINS RAISED BY DIFFERENT FAMILIES.
- ALTERNATIVE- THERE IS A DIFFERENCE IN MEAN IQ SCORES BETWEEN TWINS RAISED BY DIFFERENT FAMILIES.

- LIKE THE TWO-SAMPLES T-TEST, THE PAIRED-SAMPLES T-TEST FORMS A RATIO OF MEAN DIFFERENCES DIVIDED BY THE STANDARD ERROR OF THAT DIFFERENCE:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2} - 2r\left(\frac{S_1}{\sqrt{N_1}}\right)\left(\frac{S_2}{\sqrt{N_2}}\right)}}$$

X	C
$N1 = 20$	$N2 = 20$
$S1^2 = 54.76$	$S2^2 = 42.25$
$X1 = 53.20$	$X2 = 49.80$
r = + .60	df = 19

f = $54.76/42.25$ = 1.30 (variance are homogenous)

$$t = \frac{53.20-49.80}{\sqrt{\frac{54.76}{20}+\frac{42.85}{20}-2\,(+.60)\left(\frac{7.40}{4.47}\right)\left(\frac{6.50}{4.47}\right)}}$$

$$= \frac{3.40}{\sqrt{2.74+2.14-1.20(1.66)(1.45)}} = \frac{3.40}{\sqrt{4.84-2.89}}$$

$$= \frac{3.40}{\sqrt{1.95}} = \frac{3.40}{1.40} = \mathbf{2.43}$$

Pretest	Test after TM
$N2 = 30$	$N1 = 30$
$S2^2 = 37.21$	$S1^2 = 36.10$
$X2 = 44.80$	$X1 = 49.10$
r = + .84	df = 29

F = $37.21/36.10$ = 1.03 (variance are homogenous)

$$t = \frac{49.10 - 44.80}{\sqrt{\frac{37.21}{30} + \frac{36.10}{30} - 2\,(+.84)\left(\frac{6.10}{5.48}\right)\left(\frac{6.01}{5.48}\right)}}$$

= 6.94