Comparison of Two Independent Samples Method Based on the Normal Distribution

Osman Osmanaj*, Lazim Kamberi

University of Tetovo
Faculty of Natural Science and Mathematics
Department of Mathematic and Informatics
1200 Tetovo, Macedonia

**Corresponding author's email:osmanaj_osos [AT] yahoo.com*

ABSTRACT– ***In order to choose the right statistical test, when analyzing the data from an experiment, we must have at least a decent understanding of some basic statistical terms and concepts;***
some knowledge about few aspects related to the data we collected during the research/experiment (e.g. what types of data we have - samples, normal distribution with parameters, power of tests, how the data are organized, how many study groups (usually experimental and control at least) we have, are the samples paired or unpaired, and are the sample(s) extracted from a normally distributed/Gaussian population);a good understanding of the goal of our statistical analysis; we have to parse the entire statistical protocol in an well structured - decision tree /algorithmic manner, in order to avoid some mistakes.

Keywords— Samples, Paired samples, parameters, statistics

1. INTRODUCTION

Suppose we have a sample $X_1, \dots\dots..X_n$ that depends on a normal distribution with parameters
μ_X And σ^2 and other independent sample depends on from another parameters of normal distribution
μ_Y Andσ^2. Now compare them taken by the change of parameters
μ_x- μ_y. In fact the word μ_x- μ_y is $\bar{X}$ - $\bar{Y}$.now we have:

$$\bar{X} - \bar{Y} \sim N\left[\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right]$$

If $\boldsymbol{\sigma}^2$ is known, the confidence interval for μ_x- μ_y is based on:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\boldsymbol{\sigma}\sqrt{\frac{1}{\mathrm{n}} + \frac{1}{\mathrm{m}}}}$$

It is having normal distribution. And confidence interval will be in the form:

$$(\bar{X} - \bar{Y}) \pm z(\alpha/2)\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}$$

Usually, $\boldsymbol{\sigma}^2$ it is not known and shall be calculated from the data of the joint calculation of a sample variance,

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{m + n - 2}$$

Where: $s_x^2 = (n-1)\sum_{i=1}^{n}(X_i - \bar{X})^2$ and similar tos_y^2.

Theorem.*1*: Assume that $X_1, \dots \dots X_2$ are independent random variables with normal distribution with parameters μ_X And σ^2 and $Y_1, \dots \dots Y_m$ are independent random variables of normal distribution on the parameters μ_Y And σ^2 and X_i They are independent of Y_i. And statistics:

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

It has distribution t with $m + n - 2$ degree of freedom.
Proof: We use the definition of distribution t and a theorem:

Definition: If $Z \sim N(0,1)$ any$U \sim \chi^2_n$, Z and U They are independent, then expression

$Z/\sqrt{U/n}$ It has distribution t with n degree of freedom.

Theorem*:* Expression $(n-1)S^2/\sigma^2$ has hi-square distribution with $n - 1$ Degree of freedom.

Now, based on the above definition we must show that statistics be expressed as the quotient of a random variable having a normal distribution of regulated and square root of a random variable with the distribution of hi-square dashed for $m + n - 2$ degree of freedom. Now we remember the above theorem that $(n-1)s_X^2/\sigma^2$ And $(m-1)s_Y^2/\sigma^2$ are random variables having distribution Hi-square with $n - 1$ and $m - 1$ degrees of freedom and are independent respectively after X_i and Y_i They are such. So their sum is more hi-square with m + n-2 degrees of freedom. Now, we express statistics t as report U/V , where

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$V = \sqrt{\left[\frac{(n-1)s_X^2}{\sigma^2} + \frac{(m-1)s_Y^2}{\sigma^2}\right]\frac{1}{m+n-2}}$$

The standard deviation for $\bar{X} - \bar{Y}$ found through:

$$s_{\bar{X}-\bar{Y}} = s_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

Corollary: Regarding assumptions Theorem 1, the confidence interval $100(1-\alpha)\%$
For $\mu_X - \mu_Y$ is

$$(\bar{X} - \bar{Y}) \pm t_{m+n-2}(\alpha/2)s_{\bar{X}-\bar{Y}}$$

Example 1: Two methods A and B are used for determining the enthalpy of fusion of ice. Now we compare these two methods and see the change. Below table provides the values of enthalpy of ice from water -0.72 ° C to 0 ° C in calories per gram of the measure:

Method A	Method B
79.98	80.02
80.04	79.94
80.02	79.98
80.04	79.97
80.03	79.97
80.03	
80.04	
79.97	

So obviously change of two methods. If assumptions Theorem 1, can form a 95% confidence interval estimate From table:

$$\overline{X_A} = 80.02 \qquad S_a = 0.024$$
$$\overline{X_B} = 79.98 \qquad S_b = 0.031$$
$$s_p^2 = \frac{8 \times S_a^2 + 5 \times S_b^2}{13} = 0.01513885$$
$$s_p = 0.123$$

Change two methods is $\overline{X_A} - \overline{X_B} = 0.04$ and the standard deviation is:

$$s_{\overline{X_A}-\overline{X_B}} = s_p\sqrt{\frac{1}{8} + \frac{1}{5}} = 0.07012079$$

From the distribution table t the distribution quintile 0.975 degrees of freedom is,
t with 19 degrees of freedom is,

2.093, therefore $t_{19}(0.025) = 2.093$ confidence interval is

95% is $(\overline{X_A} - \overline{X_B}) \pm t_{19}(0.25)s_{\overline{X_A}-\overline{X_B}}$ or$(0.015, 0.65)$.

Now we discuss the initial testing. Hypothesis $H_0{:}\mu_X = \mu_Y$

And alternative hypotheses are:

$$H_1{:}\mu_X \neq \mu_Y$$
$$H_2{:}\mu_X > \mu_Y$$
$$H_3{:}\mu_X < \mu_Y$$

Test statistics used for rejecting the initial hypothesis is:

$$t = \frac{\bar{X} - \bar{Y}}{s_{\bar{X}-\bar{Y}}}$$

And regions of rejection for three alternatives are:

$$H_1, |t| > t_{m+n-2}(\alpha/2)$$

$$H_2, t > t_{m+n-2}(\alpha)$$
$$H_3, t < -t_{m+n-2}(\alpha)$$

1.2. Power of Test

Power calculations are part of important experiments that defines what should be the size of four factors of four samples. Depends:

1. Difference,$\Delta = |\mu_X - \mu_Y|$. The greater the difference, the greater the power.
2. Significance level α in which the test is concluded. So greater, it is the power.
3. Smaller standard deviation greater power.
4. The volume of sample n and m. The largest volume of samples greater power.

Required sample size can be determined from the importance of the test, standard deviation, and the power required alternative hypothesis,

$$H_1: \mu_X - \mu_Y = \Delta$$

Calculation of the power of the volume of sample testing is done by aligning large normal distribution.

We assume that given σ, α and Δ and the volume of the two samples are n. Then

$$Var(\bar{X} - \bar{Y}) = \sigma^2\left(\frac{1}{n} + \frac{1}{n}\right) = \frac{2\sigma^2}{n}$$

Testing level α of $H_0: \mu_X = \mu_Y$ against the alternative
$H_1: \mu_X \neq \mu_Y$ Based on test-statistics:

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{2/n}}$$

and the interval of refusal for this test is:$|Z| > z(\alpha/2)$,or

$$|\bar{X} - \bar{Y}| > z(\alpha/2)\sigma\sqrt{\frac{2}{n}}$$

The power of the test if $\mu_X - \mu_Y = \Delta$ is likely to test-statistics refused range of rejection, or

$$P\left[|\bar{X} - \bar{Y}| > z(\alpha/2)\sigma\sqrt{\frac{2}{n}}\right] = P\left[\bar{X} - \bar{Y} > z(\alpha/2)\sigma\sqrt{\frac{2}{n}}\right] + P\left[\bar{X} - \bar{Y} < -z(\alpha/2)\sigma\sqrt{\frac{2}{n}}\right]$$

Now we have:

$$P\left[\bar{X} - \bar{Y} > z(\alpha/2)\sigma\sqrt{\frac{2}{n}}\right] = 1 - \Phi\left[z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right]$$

And:

$$P\left[\bar{X} - \bar{Y} < -z(\alpha/2)\sigma\sqrt{\frac{2}{n}}\right] = \Phi\left[-z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right]$$

Therefore, likely to test-statistics refused rejection interval is:

$$1 - \Phi\left[z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right] + \Phi\left[-z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{\frac{n}{2}}\right].$$

2. COMPARING PAIRS OF SAMPLES

Often samples may be pairs. Pairs can be an effective technique and will demonstrate comparing couple model and non-model pairs. See consider first couple of samples as pairs. We write
(X_i, Y_i) Where $i = 1, \dots \dots n,$ and assume that
X-is and Y-is they have between μ_X and μ_Y and variance σ_X^2 andσ_Y^2. Assume that pairs of samples have independent distribution and$Cov(X_i, Y_i) = \sigma_{XY}$. We will work with change $D_i = \bar{X} - \bar{Y}$ and they are independent of the:

$$E(D_i) = \mu_X - \mu_Y$$

$$Var(D_i) = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_{XY} = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$$

Where ρ is the correlation coefficient. A calculation of the ordinary
Is $\mu_X - \mu_Y$ and$\bar{D} = \bar{X} - \bar{Y}$, the average change. From D_i derived from that:

$$E(\bar{D}) = \mu_X - \mu_Y$$

$$Var(\bar{D}) = \frac{1}{n}(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)$$

Now on the other hand we assume that the experiment is made by taking two independent samples with
n X-a and n Y-a. Therefore, $\mu_X - \mu_Y$ It can be calculated more $\bar{X} - \bar{Y}$ and

$$E(\bar{X} - \bar{Y}) = \mu_X - \mu_Y$$

$$Var(\bar{X} - \bar{Y}) = \frac{1}{n}(\sigma_X^2 + \sigma_Y^2)$$

By comparing the two calculations variances see¯variance D is small if the correlation is positive. In these circumstances it is more effective coupling. In a simple case in which $\sigma_X = \sigma_Y = \sigma$ two variances can be expressed simply more with:

$$Var(\bar{D}) = \frac{2\sigma^2(1-\rho)}{n}$$

In pair case:

$$Var(\bar{X} - \bar{Y}) = \frac{2\sigma^2}{n}$$

And in the case of non-couple and relative efficiency is:

$$\frac{Var(\bar{D})}{Var(\bar{X} - \bar{Y})} = 1 - \rho$$

3. REFERENCES

[1] Arellano-Valle RB, Gomez HW, Quintana FA. (2004). A new class of skew normal distributions. Communications in Statistics: Theory and Methods 33(7): 14651480
[2] Azzalini A. (1985). A class of distributions which includes the normal ones. Scandinavian journal of statistics 12:171 - 178.
[3] Azzalini A. (1986). Further result on a class of distribution which includes the normal ones. Statistica, 46, 199-208.
[4] Cheng, R. C. H., & Amin, N. A. K. (1983). Estimating parameters in continuous univariante distributions with a shifted origin. Journal of the Royal Statistical Society. Series B (Methodological), 394-403