Application of Student's *t*-test, Analysis of Variance, and Covariance

Abstract

Student's t test (*t* test), analysis of variance (ANOVA), and analysis of covariance (ANCOVA) are statistical methods used in the testing of hypothesis for comparison of means between the groups. The Student's *t* test is used to compare the means between two groups, whereas ANOVA is used to compare the means among three or more groups. In ANOVA, first gets a common *P* value. A significant *P* value of the ANOVA test indicates for at least one pair, between which the mean difference was statistically significant. To identify that significant pair(s), we use multiple comparisons. In ANOVA, when using one categorical independent variable, it is called one-way ANOVA, whereas for two categorical independent variables, it is called two-way ANOVA. When using at least one covariate to adjust with dependent variable, ANOVA becomes ANCOVA. When the size of the sample is small, mean is very much affected by the outliers, so it is necessary to keep sufficient sample size while using these methods.

Keywords: *Student's t test, analysis of variance, analysis of covariance, one-way, two-way*

Prabhaker Mishra, Uttam Singh, Chandra M Pandey, Priyadarshni Mishra[1], Gaurav Pandey[2]

Departments of Biostatistics and Health Informatics, [1]Ophthalmology and [2]Gastroenterology, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India

Introduction

Student's *t* test (*t* test), analysis of variance (ANOVA), and analysis of covariance (ANCOVA) are statistical methods used in the testing of hypothesis for comparison of means between the groups. For these methods, testing variable (dependent variable) should be in continuous scale and approximate normally distributed. Mean is the representative measure for normally distributed continuous variable and statistical methods used to compare between the means are called parametric methods. For non-normal continuous variable, median is representative measure, and in this situation, comparison between the groups is performed using non-parametric methods. Most parametric test has an alternative nonparametric test.[1-3]

There are many statistical tests within Student's *t* test (*t* test), ANOVA and ANCOVA, and each test has its own assumptions. Although not every method is popular, some of them can be managed from other available methods. The aim of the present article is to discuss the assumptions, application, and interpretation of the some popular T, ANOVA, and ANCOVA methods

i.e., one sample *t* test, independent samples *t* test, paired samples *t* test, one-way ANOVA, two-ways ANOVA, one-way repeated measures ANOVA, two-ways repeated measures ANOVA, one-way ANCOVA, and One-way repeated measures ANCOVA. To understand the above statistical methods, an example [Table 1] with a data set of 20 patients whose age groups, gender, body mass index (BMI), and diastolic blood pressure (DBP) measured at baseline (B/L), 30 min and 60 min are given below. Further, examples related to the above statistical methods are discussed from the given data.

T test, ANOVA, and ANCOVA

Basic concepts

The Student's *t* test (also called T test) is used to compare the means between two groups and there is no need of multiple comparisons as unique *P* value is observed, whereas ANOVA is used to compare the means among three or more groups.[4,5] In ANOVA, the first gets a common *P* value. A significant *P* value of ANOVA test indicates for at least one pair, between which the mean difference was statistically significant.[6] To identify that significant pair(s), post-hoc test (multiple comparisons) is used. In ANOVA test, when at least one covariate

Address for correspondence:
*Dr. Prabhaker Mishra,
Department of Biostatistics and Health Informatics,
Sanjay Gandhi Post Graduate Institute of Medical Sciences,
Lucknow, Uttar Pradesh, India.
E-mail: mishrapk79@gmail.com*

(continuous variable) is adjusted to remove the confounding effect from the result

called ANCOVA. ANOVA test (F test) is called "Analysis of Variance" rather than "Analysis of Means" because inferences about means are made by analyzing variance.[7-9]

Steps in hypothesis testing

Hypothesis building

Like other tests, there are two kinds of hypotheses; null hypothesis and alternative hypothesis. The alternative hypothesis assumes that there is a statistically significant difference exists between the means, whereas the null hypothesis assumes that there is no statistically significant difference exists between the means.

Computation of test statistics

In these test, first step is to calculate test statistics (called t value in student's *t* test and F value in ANOVA test) also called calculated value. It is calculated after putting inputs (from the samples) in statistical test formula. In student's *t* test, calculated t value is ratio of mean difference and standard error, whereas in the ANOVA test, calculated F value is ratio of the variability between groups with the variability of the observations within the groups.[1,4]

Tabulated value

At degree of freedom of the given observations and desired level of the confidence (usually at two-sided test, which is more powerful than one-sided test), corresponding tabulated value of the T test or F test is selected (from the statistical table).[1,4]

Comparison of calculated value with tabulated value and null hypothesis

If the calculated value is greater than the tabulated value, then reject the null hypothesis where null hypothesis states that means are statistically same between the groups.[1,4] As the sample size increases corresponding degree of freedom also increases. For a given level of confidence, higher degree of freedom has lower tabulated value. That's the reason, when the sample size increases, its significance level also improves (i.e., *P* value is decreasing).

T Test

It is one of the most popular statistical techniques used to test whether mean difference between two groups is statistically significant. Null hypothesis stated that both means are statistically equal, whereas alternative hypothesis stated that both means are not statistically equal i.e., they are statistically different to each other.[1,3,7] T test are three types i.e., one sample *t* test, independent samples *t* test, and paired samples *t* test.

One-sample t test

The one sample t test is a statistical procedure used to determine whether mean value of a sample is statistically same or different with mean value of its parent population from which sample was drawn. To apply this test, mean, standard deviation (SD), size of the sample (Test variable), and population mean or hypothetical mean value (Test value) are used. Sample should be continuous variable and normally distributed.[1,9-11] One-sample *t* test is used when sample size is <30. In case sample size is ≥30 used to prefer one sample z test over one sample *t* test although for one sample z test, population SD must be known. If population SD is not known, one sample *t* test can be used at any sample size. In one sample Z test, tabulated value is z value (instead of *t* value in one sample *t* test). To apply this test through popular statistical software i.e., statistical package for social sciences (SPSS), option can be found in the following menu [Analyze – compare means – one-sample *t* test].

Example: From Table 1, BMI (mean ± SD) was given 24.45 ± 2.19, whereas population mean was assumed to be 25.5. One sample *t* test indicated that mean difference between sample mean and population mean was statistically significantly different to each other ($P = 0.045$).

Independent samples t test

The independent *t* test, also called unpaired *t* test, is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated (independent) groups?

To apply this test, a continuous normally distributed variable (Test variable) and a categorical variable with two categories (Grouping variable) are used. Further mean, SD, and number of observations of the group 1 and group 2 would

Table 1: Data of the 20 patients

Sr. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Gender	M	M	M	M	M	M	M	M	M	M	F	F	F	F	F	F	F	F	F	F
Age groups	1	2	3	1	1	3	3	2	1	3	3	2	2	2	3	1	1	3	3	2
BMI	23	25	26	24	23	27	27	24	21	28	26	26	23	24	25	22	19	25	26	25
DBP (B/L)	68	80	85	75	77	82	80	82	80	74	75	86	85	87	83	79	73	77	82	81
DBP (at 30 min)	75	79	90	80	81	89	92	90	80	80	89	88	87	91	88	78	74	84	83	80
DBP (at 60 min)	70	78	85	75	72	72	75	85	80	75	85	80	82	82	85	74	87	86	85	72

Age groups: 1 (<30 years), 2 (30-50 years), 3 (>50 years). Gender: M=Male, F=Female, BMI=Body mass index, DBP=Diastolic blood pressure, B/L=Baseline, min=Minute

be used to compute significance level. In this procedure, first significance level of Levene's test is computed and when it is insignificant ($P > 0.05$), equal variances otherwise ($P < 0.05$), unequal variances are assumed between the groups and according P value is selected for independent samples *t* test.[1,10-12] In SPSS [Analyze – compare means – independent samples *t* test].

Example: From Table 1, mean BMI of the male (n = 10) and female (n = 10) were 24.80 ± 2.20 and 24.10 ± 2.23, respectively. Levene's test (*p* = 0.832) indicated that variances between the groups were statistically equal. At equal variances assumed, independent samples *t* test (*p* = 0.489) indicated that mean BMI of the male and female was statistically equal.

Paired samples t test

The paired samples *t* test, sometimes called the dependent samples *t*-test, is used to determine whether the change in means between two paired observations is statistically significant? In this test, same subjects are measured at two time points or observed by two different methods.[4] To apply this test, paired variables (pre-post observations of same subjects) are used where paired variables should be continuous and normally distributed. Further mean and SD of the paired differences and sample size (i.e., no. of pairs) would be used to calculate significance level.[1,11,13] In SPSS [Analyze – compare means – paired samples *t* test].

Example: From Table 1, DBP of the 20 patients (mean ± SD); at baseline, 30 min and paired differences (difference between baselines and 30 min) were 79.55 ± 4.87, 83.90 ± 5.58, and 4.35 ± 4.16. Paired samples *t* test indicated that mean difference of paired observations of DBP between baseline and 30 min was statistically significant ($P < 0.001$).

ANOVA test (F test)

A statistical technique used to compare the means between three or more groups is known as ANOVA or F test. It is important that ANOVA is an omnibus test statistic. Its significant *P* value indicates that there is at least one pair in which the mean difference is statistically significant. To determine the specific pair's, *post hoc* tests (multiple comparisons) are used. There are various ANOVAs test, and their objectives are varying from one test to another. There are two main types of ANOVA i.e., one-way ANOVA and one-way repeated measures ANOVA. First is used for independent observations and later for dependent observations. When used one categorical independent variable called one-way ANOVA, whereas for two categorical independent variables called two-way ANOVA. When used at least one covariate to adjust with dependent variable, ANOVA becomes ANCOVA.[1,11,14]

Post-hoc test (multiple comparisons): *Post hoc* tests (pair-wise multiple comparisons) used to determine the significant pair(s) after ANOVA was found significant.

Before applying post-hoc test (in between subjects factors), first need to test the homogeneity of the variances among the groups (Levene's test). If variances are homogeneous ($P \geq 0.05$), select any multiple comparison methods from least significant difference (LSD), Bonferroni, Tukey's, etc.[15,16] If variances are not homogeneous ($P < 0.05$), used to select any multiple comparison methods from Games-Howell, Tamhane's T2, etc.[15,16] Bonferroni is a good method for equal variances, whereas *Tamhane's T2* for unequal variances as both calculate significance level by controlling error rate. Similarly, for repeated measures ANOVA (RMA) (in within subjects factors), select any method from LSD, Boneferroni, Sidak although Bonferroni might be a better choice. The significance level of each of the multiple comparison method is varying from other methods as each used for a particular situation.

One-way ANOVA

The One-way ANOVA is extension of independent samples *t* test (In independent samples t test used to compare the means between two independent groups, whereas in one-way ANOVA, means are compared among three or more independent groups). A significant *P* value of this test refers to multiple comparisons test to identify the significant pair(s).[17] In this test, one continuous dependent variable and one categorical independent variable are used, where categorical variable has at least three categories. In SPSS [Analyze–compare means–one-way ANOVA].

Example: From Table 1, 20 patient's DBP (at 30 min) are given. One-way ANOVA test was used to compare the mean DBP in three age groups (independent variable), which was found statistically significant (*p* = 0.002). Levene test for homogeneity was insignificant (*p* = 0.231), as a result Bonferroni test was used for multiple comparisons, which showed that DBP was significantly different between two pairs i.e., age group of <30 to 30–50 and <30 to >50 ($P < 0.05$) but insignificant between one pair i.e., 30–50 to >50 ($P > 0.05$).

Two-way ANOVA

The two-way ANOVA is extension of one-way ANOVA [In one-way ANOVA, only one independent variable, whereas in two-way ANOVA, two independent variables are used]. The primary purpose of a two-way ANOVA is to understand whether there is any interrelationship between two independent variables on a dependent variable.[18] In this test, a continuous dependent variable (approximately normally distributed) and two categorical independent variables are used. In SPSS [Analyze –General Linear Model –Univariate].

Example: From Table 1, 20 patient's DBP (at 30 min) are given. Two-way ANOVA test was used to compare the mean DBP between age groups (independent variable_1) and gender (independent variable_2), which indicated that there was no significant interaction of DBP with age

groups and gender (tests of Between-Subjects effects in age groups*gender; $P = 0.626$) with effect size (Partial Eta Squared) of 0.065. The result also showed that there was significant difference in estimated marginal means (adjusted mean) of DBP between age groups ($P = 0.005$) but insignificant in gender ($P = 0.662$), where sex and age groups was adjusted.

One-way repeated measures ANOVA

Repeated Measures ANOVA (RMA) is the extension of the paired *t* test. RMA is also referred to as within-subjects ANOVA or ANOVA for paired samples. Repeated measures design is a research design that involves multiple measures of the same variable taken on the same or matched subjects either under different conditions or more than two time periods. (In paired samples *t* test, compared the means between two dependent groups, whereas in RMA, compared the means between three or more dependent groups). Before calculating the significance level, Mauchly's test is used to assess the homogeneity of the variance (also called sphericity) within all possible pairs. When *P* value of Mauchly's test is insignificant ($P \geq 0.05$), equal variances are assumed and *P* value for RMA would be taken from sphericity assumed test (Tests of Within-Subjects effects). In case variances are not homogeneous (Mauchly's test: $P < 0.05$), epsilon (ε) value (which shows the departure of the sphericity, 1 shows perfect sphericity) decides the statistical method to calculate *P* value for RMA. When $\varepsilon \geq 0.75$ Huynh-Feldt while for $\varepsilon < 0.75$, Greenhouse-Geisser method (univariate method) or Wilks' lambda (multivariate method) is used to calculate *P* value for the RMA.[19] When the RMA is significant, pair-wise comparison contains multiple paired t tests with a Bonferroni correction is used.[20] In SPSS [Analyze – General Linear Model – Repeated Measures ANOVA].

Example: From Table 1, 20 patient's DBP were at baseline (79.55 ± 4.87), at 30 min (83.90 ± 5.58), and at 60 min (79.25 ± 5.68). The Mauchly's test of sphericity indicated that variances were equal ($P = 0.099$) between the pairs. RMA tests (i.e., Within-Subjects effects) was assessed using sphericity assumed test (*P* value = 0.001), which indicated that change in DBP over the time was statistically significant. Bonferroni multiple comparisons indicated that mean difference was statistically significant between DBP_B/l to DBP_30 min and DBP_30 min to DBP_60 min ($P < 0.05$) but insignificant between DBP_B/l to DBP_60 min ($P > 0.05$).

Two-way repeated measures ANOVA

Two-way Repeated Measures ANOVA is combination of between-subject and within-subject factors. A two-way RMA (also known as a two-factor RMA or a two-way "Mixed ANOVA") is extension of one-way RMA [In one-way RMA, use one dependent variable under repeated observations (normally distributed continuous variable) and one categorical independent variable (i.e., time points), whereas in two-way RMA; one additional categorical independent variable is used]. The primary purpose of two-way RMA is to understand if there is an interaction between these two categorical independent variables on the dependent variable (continuous variable). The distribution of the dependent variable in each combination of the related groups should be approximately normally distributed.[21] In SPSS [Analyze–General Linear Model – Repeated Measures], where second independent variable will be included as between subjects factor.

Example: From Table 1, 20 patient's DBP were at baseline (79.55 ± 4.87), at 30 min (83.90 ± 5.58), and at 60 min (79.25 ± 5.68). The Mauchly's test of sphericity ($P = 0.138$) indicated that variances were equal between the pairs. Two-way RMA tests for interaction (i.e., Within-Subjects effects) were assessed using sphericity assumed test (DBP*gender: *P* value = 0.214), which indicated that there was no interaction of gender with time and associated change in DBP over the time was statistically insignificant.

One-way ANCOVA

One-way ANCOVA is extension of one-way ANOVA [In one-way ANOVA, do not adjust the covariate, whereas in the one-way ANCOVA; adjust at least one covariate]. Thus, the one-way ANCOVA tests find out whether the independent variable still influences the dependent variable after the influence of the covariate(s) has been removed (i.e., adjusted). In this test, one continuous dependent variable, one categorical independent variable, and at least one continuous covariate for removing its effect/adjustment are used.[8,22] In SPSS [Analyze - General Linear Model – Univariate].

Example: From Table 1, 20 patient's DBP at 30 min are given. One-way ANCOVA test was used to compare the mean DBP in three age groups (independent variable) after adjusting the effect of baseline DBP, which was found to be statistically significant ($P = 0.021$). As Levene test for homogeneity was insignificant ($P = 0.601$), resultant Bonferroni test was used for multiple comparisons, which showed that DBP was significantly different between one pair i.e., age group of <30 to >50 ($P = 0.031$) and insignificant between rest two pairs i.e., <30 to 30–50 and 30–50 to >50 ($P > 0.05$).

One-way repeated measures ANOCOVA

One-way repeated measures ANCOVA is the extension of the One-way RMA. [In one-way RMA, we do not adjust the covariate, whereas in the one-way repeated measures ANCOVA, we adjust at least one covariate]. Thus, the One-way repeated Measures ANCOVA is used to test whether means are still statistically equal or different after adjusting the effect of the

covariate(s).[23,24] In SPSS [Analyze –General Linear Model – Repeated Measures ANOVA].

Example: From Table 1, 20 patient's DBP were at baseline (79.55 ± 4.87), at 30 min (83.90 ± 5.58), and at 60 min (79.25 ± 5.68). The Mauchly's test of sphericity indicated that variances were equal ($P = 0.093$) between the pairs. RMA tests (i.e., Within-Subjects effects) were assessed using sphericity assumed test (DBP*BMI: P value = 0.011), which indicated that change in DBP over the time was statistically significant after adjusting BMI. Bonferroni multiple comparisons indicated that mean difference was statistically significant between DBP_B/l to DBP_30 min and DBP_30 min to DBP_60 min but insignificant between DBP_B/l to DBP_60 min after adjusting BMI.

Conclusions

Student's *t* test, ANOVA, and ANCOVA are the statistical methods frequently used to analyze the data. Two common things among these methods are dependent variable must be in continuous scale and normally distributed, and comparisons are made between the means. All above methods are parametric method.[2] When the size of the sample is small, mean is very much affected by the outliers, so it is necessary to keep sufficient sample size while using these methods.

Conflicts of interest

There are no conflicts of interest.

References

1. Sundaram KR, Dwivedi SN, Sreenivas V. Medical Statistics: Principles and Methods. 2nd ed. New Delhi: Wolters Kluwer India; 2014.
2. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth 2019;22:67072.
3. Jaykaran. How to select appropriate statistical test?. J Pharm Negative Results 2010;1:61-3.
4. Altman DG. Practical Statistics for Medical Research. Boca Raton, Florida: CRC Press; 1990.
5. McDonaldJH. Handbook of Biolological Statistics. Third Edition. University of Delaware. Baltimore, Maryland, U.S.A.: Sparky House Publishing; 2014.
6. Kao LS, Green CE. Analysis of variance: Is there a difference in means and what does it mean? J Surg Res 2007;144:158-70.
7. Whitley E, Ball J. Statistics review 5: Comparison of means. Crit Care 2002;6:424-8.
8. Kim HY. Statistical notes for clinical researchers: Analysis of covariance (ANCOVA). Restor Dent Endod 2018;43:e43.
9. One-Sample T-Test using SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/one-sample-t-test-using-spss-statistics.php. [Last accessed on 2019 Jun 09].
10. Barton B, Peat J. Medical Statistics: A Guide to SPSS, Data Analysis and Clinical Appraisal. Second edition. Wiley Blackwell, BMJ Books; 2014.
11. Peat J, Barton B. Medical Statistics: A Guide to Data Analysis and Critical Appraisal. Hoboken, New Jersey: John Wiley and Sons; 2008.
12. Independent *t*-test using SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/independent-t-test-using-spss-statistics.php. [Last accessed on 2019 Jun 09].
13. Dependent T-Test using SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/dependent-t-test-using-spss-statistics.php. [Last accessed on 2019 Jun 09].
14. Kim HY. Analysis of variance (ANOVA) comparing means of more than two groups. Restor Dent Endod 2014;39:74-7.
15. One-way ANOVA in SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php. [Last accessed on 2019 Jun 09].
16. Lee S, Lee DK. What is the proper way to apply the multiple comparison test?. Korean J Anesthesiol 2018;71:353-60.
17. One-way ANOVA *Post hoc* tests. Available from: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/base/idh_onew_post.html. [Last accessed on 2019 Jun 10].
18. Two-way ANOVA in SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php. [Last accessed on 2019 Jun 09].
19. ANOVA with Repeated Measures using SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/one-way-anova-repeated-measures-using-spss-statistics.php. [Last accessed on 2019 Jun 09].
20. Mauchly'ssphericity test. Available from: https://en.wikipedia.org/wiki/Mauchly%27s_sphericity_test. [Last accessed on 2019 Jun 10].
21. Two-way repeated measures ANOVA in SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/two-way-repeated-measures-anova-using-spss-statistics.php.[Last accessed on 2019 Jun 09].
22. One-way ANCOVA in SPSS Statistics. Available from: https://statistics.laerd.com/spss-tutorials/ancova-using-spss-statistics.php. [Last accessed on 2019 Jun 09].
23. Conduct and Interpret Repeated MeasuresANCOVA. Available from: https://www.statisticssolutions.com/repeated-measures-ancova/. [Last accessed on 2019 Jun 09].
24. How Can I Do Repeated Measures ANOVA With Covariates In SPSS?. Available from: https://stats.idre.ucla.edu/spss/faq/how-can-i-do-repeated-measures-anova-with-covariates-in-spss/. [Last accessed on 2019 Jun 09].