Applied statistics in vascular surgery Part III: Understanding Analysis of Variance (ANOVA) and its subtypes

Constantine N. Antonopoulos[1], Efthymios D. Avgerinos[2], Larry Fish[3], John Kakisis[1]

[1]Department of Vascular Surgery, Athens University Medical School, "Attikon" Hospital, Athens, Greece
[2]Department of Vascular Surgery, University of Pittsburgh, UPMC Presbyterian Hospital, Pittsburgh, PA
[3]Department of Surgery, University of Pittsburgh, UPMC Presbyterian Hospital, Pittsburgh, PA

Abstract:

Statistical testing is rather simple in the case of a continuous dependent variable and one independent two-group variable; the t-test will determine whether the two means differ significantly from each another. However, when the independent variable has more than two independent levels, then further t-test calculations potentially introduce a type I error. The one-way analysis of variance (ANOVA) will test for differences among three or more groups while controlling experiment-wise type I error. A subsequent multiple comparison (post-hoc) test will identify pairs of groups that are significantly different. The ANOVA can be used in several research designs, such as repeated-measures (or within-subjects), mixed-model (or within-between), factorial (or two-way, three-way ANOVA etc), analysis of covariance (ANCOVA) and multivariate analysis of variance (MANOVA), after taking into consideration the levels of the independent variable and its covariates.

INTRODUCTION

A student was asked to compare maximum aortic diameter in patients with abdominal aortic aneurysm (AAA) at the end of the first month after open surgical repair (OSR; group A) and endovascular aortic repair (EVAR; group B). The null hypothesis is that the mean maximum aortic diameter after OSR is the same compared to EVAR. The statistical testing looks rather straightforward; there is one continuous dependent variable (maximum aortic diameter) and one independent variable (type of AAA repair) with two independent groups (OSR and EVAR). The student will compute mean maximum aortic diameter for OSR and EVAR and then the *independent samples t-test* will examine whether the two means greatly differ from each other[1].

When, however, we want to test for maximum aortic diameter differences among patients treated with seven different endografts at the end of the first month after EVAR, analysis might be more complicated. In that case, there would be 21 different comparisons and thus we would need to perform 21 different t-tests simultaneously (Supplemental Material). However, for each individual t-test, the probability of type I error would be 5% and "experiment-wise" the probability of type I error, across all pair-wise comparisons, would be higher. To avoid inflating the experiment-wise probability of type I error when we compare more than two means, we should use *Analysis of Variance (ANOVA),* which provides only one overall test; a test of whether all the population means are equal[2]. The t-test is a type of ANOVA dealing with one continuous dependent variable and one independent variable with two groups. In our second scenario, ANOVA is the method of choice, as we have one continuous dependent variable (maximum aortic diameter) and one independent variable (type of endograft used) with seven groups (seven different types of endograft).

Author for correspondence:

Constantine N. Antonopoulos

Department of Vascular Surgery, Athens University Medical School, Attikon University Hospital, Athens, Greece
E-mail: kostas.antonopoulos@gmail.com
ISSN 1106-7237/ 2019 Hellenic Society of Vascular and Endovascular Surgery Published by Rotonda Publications

BASIC CONCEPTS OF ANOVA

The one-way ANOVA is used to test for differences among the levels of an independent variable, with respect to the mean value of the independent variable, also called a factor. ANOVA is used as a test for comparison between means, but it uses variance as a tool for evaluating the sizes of those mean differences. As a result, intragroup and intergroup variability is what is actually being analysed in ANOVA. A main result of the ANOVA model is the F statistic (also called the F-ratio), The formula, which conveys the general concept behind the F-ratio, is: *F-ratio = Between-groups variance / Within-groups variance*. The formula for F varies among models, and even within models. If no differences are found among the tested groups - the null hypothesis -the F-ratio will be close to 1[3].

BASIC ASSUMPTIONS OF ANOVA

ANOVA controls for multiple type I errors and thus maintains the experiment-wise alpha level at the desired value (usually 0.05). To be valid, the ANOVA requires that two assumptions be satisfied: a) The dependent variable (eg. maximum aortic diameter) must be normally distributed in each of the populations being compared (eg. in each group of the 7 different endografts used), b) The variance of the dependent variable is constant among the populations being compared (eg. the 7 groups of different endografts used)[3]. If the first assumption is violated, the researcher should perform data transformation or perform nonparametric *Kruskal-Wallis H Test*, while if the second assumption is violated, other tests like *Welch* of *Brown and Forsythe* can be used[3]. Certain ANOVA models, to be discussed below, may require additional assumptions.

OTHER ANOVA MODELS

As reported in the previous section, when dealing with one continuous dependent and one independent variable with multiple groups we should use one-way ANOVA (Table 1). If, however, the dependent variable was measured on multiple occasions (or under different conditions), but on the same subjects, *repeated-measures (or within-subjects) ANOVA* is used. In this model the levels of the independent variable will represent the occasions, or the circumstances, in which each subject was observed. When examining differences in the continuous dependent variable both by groups and time, then a *mixed-model (or within-between) ANOVA* should be used. In the case of one continuous dependent and two or more independent variables, the appropriate test is *factorial ANOVA (or two-way, three-way ANOVA etc)*[2,4].

If we want to examine differences in a continuous dependent variable between groups of an independent variable (ANOVA), but we need to control for potential confounders, the analysis is called *analysis of covariance (ANCOVA)*. ANCOVA can be applied to a between-subjects design, a within-subjects design, or a mixed-model design. Finally, when the model incorporates multiple continuous dependent variables, then a multivariate analysis of variance (MANOVA) is used to investigate differences among the levels of the independent variables, with respect to their effects on a composite of the dependent variables. *Multivariate analysis of covariance* (MANCOVA) is the extension of MANOVA to incorporate control for potential confounders[2,4].

<table>
<tr><th rowspan="2" colspan="2"></th><th colspan="2">Dependent Variable(s)</th></tr>
<tr><th>1
(eg. maximum aortic diameter)</th><th>≥2
(eg. maximum aortic diameter and
aortic neck length)</th></tr>
<tr><td rowspan="7">Independent Variable(s)</td><td>1
with 1 group
(eg. all patients with AAA)</td><td>One sample t-test</td><td>-</td></tr>
<tr><td>1
with 2 groups
(eg. patients after EVAR and patients after OSR)</td><td>Independent samples t-test</td><td>Hotelling's t-test</td></tr>
<tr><td>1
with ≥3 groups
(eg. patients treated with 7 types of endograft)</td><td>one-way ANOVA</td><td>one-way MANOVA</td></tr>
<tr><td rowspan="2">1
measured multiple times over time or under different conditions but on the same subjects
(eg. same patients with AAA, with post-treatment follow-up measures for each patient post EVAR)</td><td>2
measurements per subject
Dependent samples t-test</td><td rowspan="2">repeated-measures MANOVA</td></tr>
<tr><td>≥3
measurements per subject
repeated-measures ANOVA</td></tr>
<tr><td>1
with control for potential confounders
(eg. type of operation, controlled for age)</td><td>ANCOVA</td><td>MANCOVA</td></tr>
<tr><td>≥2
(eg. type of operation and age-group)</td><td>factorial ANOVA
(two-way, three-way ANOVA etc).</td><td>factorial MANOVA
(two-way, three-way MANOVA etc).</td></tr>
</table>

Abbreviations: AAA: abdominal aortic aneurysm, EVAR: endovascular aortic repair, OSR; open surgical repair, ANOVA: analysis of variance, ANCOVA: analysis of covariance (ANCOVA), MANOVA: multiple analysis of variance MANCOVA: multiple analysis of covariance

Table 1. Types of suggested analysis according to the number of dependent, number and type of independent and groups of independent variable(s)

POST-HOC TESTING

Although ANOVA tests for differences among groups of an independent variable, it cannot tell us which particular group means differed from each other. If we want to determine which particular group means differed from each other, a *post hoc test* is usually required. There are several post-hoc tests available and the choice depends on the form of the post-hoc hypothesis being tested, or on how badly the homogeneity of variance assumption has been violated. Among them, when equal variances are assumed, available multiple post-hoc tests are Bonferroni, Tukey's honestly significant difference test, Sidak, Gabriel, Hochberg, Dunnett, Scheffé, and LSD (least significant difference). When equal variances are not assumed, the appropriate post-hoc tests, which may be more robust, include Tamhane's T2, Dunnett's T3, Games-Howell, and Dunnett's C[5].

A special assumption, called sphericity, must be satisfied in repeated-measures ANOVA; the assumption is that the variances of the differences between all possible group pairs are equal. When Mauchly's Test of Sphericity indicates that data violates this assumption, corrections like the lower-bound estimate, Greenhouse-Geisser and Huynh-Feldt adjustments are applied to the degrees of freedom in the F-ratio in order to avoid increases in the type I error rate[3].

CONCLUSION

One-way ANOVA is a statistical model, which tests the hypothesis that the means of a dependent measure differ among two or more populations. When a significant F-ratio is found, a multiple comparison (post-hoc) test may be used in order to identify which of the three or more groups are different. This, and other models of ANOVA, such as repeated-measures (or within-subjects), mixed-model (or within-between), factorial (or two-way, three-way ANOVA etc), ANCOVA and MANOVA can be used appropriately after taking into consideration the levels of the independent variable and its covariates.

REFERENCES

1 Antonopoulos CN, Kakisis JD. Applied statistics in vascular surgery Part 1: Choosing between parametric and non-parametric tests. *HelJVES*. 2019;1(1):36-7.

2 Moye L. Statistical Methods for Cardiovascular Researchers. *Circ Res*. 2016;118(3):439-53.

3 Gaddis GM, Gaddis ML. Introduction to biostatistics: Part 4, statistical inference techniques in hypothesis testing. *Ann Emerg Med*. 1990;19(7):820-5.

4 Tabachnick BG, Fidell LS. Using Multivariate Statistics. Boston: Allyn and Bacon; 2013.

5 McHugh ML. Multiple comparison analysis testing in ANOVA. *Biochem Med (Zagreb)*. 2011;21(3):203-9.

Supplemental Material

In case of (**n**) groups, the number of possible comparisons (**C**) among them is given by the formula: **C= [n* (n-1)]/2**. For n=7 groups (7 different types of endograft) we get: C=[7*(7-1)]/2 = 21.

https://www.dcode.fr/combinations