

ENDGAMES

STATISTICAL QUESTION

One way analysis of variance

Philip Sedgwick *senior lecturer in medical statistics*

Centre for Medical and Healthcare Education, St George's, University of London, Tooting, London, UK

Researchers investigated current trends in the use of bariatric surgery in England. In particular, they looked at the surgical techniques used and factors that influenced postoperative outcomes. A population cohort study was performed. All NHS adult patients with a primary diagnosis of obesity who had undergone a primary elective bariatric procedure (gastric bypass, gastric banding, or sleeve gastrectomy) in England between April 2000 and March 2008 were studied. The main outcome measures included mortality at 30 days and one year after surgery, unplanned readmission to hospital within 28 days, and duration of stay in hospital.¹

In total, 3649 gastric band procedures, 3191 gastric bypasses, and 113 sleeve gastrectomies were performed. Patients' characteristics at time of surgery were compared between the procedures to establish differences that might have influenced postoperative outcome. The patient characteristics compared included age. No significant difference was reported between surgical procedures in mean age (gastric bypass 42.25 years (standard deviation 9.50), gastric banding 42.44 (9.79), sleeve gastrectomy 44.18 (8.96); one way analysis of variance $P=0.364$).

The researchers reported that the number of bariatric surgical procedures had increased in England in recent years. Gastric banding and gastric bypass were the most prevalent procedures, with sleeve gastrectomy first recorded in 2006. Patients selected for gastric banding had lower postoperative mortality and readmission rates plus a shorter length of stay than those selected for gastric bypass.

Which of the following statements, if any, are true for one way analysis of variance?

- It is a non-parametric statistical test
- It was assumed that the three surgical groups were sampled from populations with a common variance for age
- The null hypothesis involved pairwise comparisons between surgical groups in mean age

Answers

Statement *b* is true, whereas *a* and *c* are false.

One way analysis of variance, sometimes abbreviated to analysis of variance (ANOVA), is a statistical test used to compare the mean of a variable measured on a continuous scale in three or more independent groups. The bariatric surgical procedures of gastric bypass, gastric banding, and sleeve gastrectomy formed independent groups—that is, a patient could have undergone only one of the procedures as primary elective surgery. Mean age was compared between the three groups to establish whether the groups were equivalent; otherwise age may have confounded the comparison of postoperative outcomes after bariatric surgery.

Analysis of variance is a parametric statistical test (*a* is false). The two types of statistical tests—parametric and non-parametric methods—have been described in a previous question.² Parametric methods make the assumption that the variable being compared between groups has a particular distribution, typically Normal, in the population from which each group was sampled. The Normal distribution, described in a previous question,³ is a theoretical distribution described by its mean and standard deviation. Non-parametric methods make no such assumption about the distribution of the variable in the population. When using analysis of variance to compare the three surgical procedures in mean age it was assumed that the distribution of age was Normal in the population from which each patient group was sampled. Each population consisted of all NHS adult patients with a primary diagnosis of obesity who had undergone one of the surgical procedures. A further assumption was also made when undertaking analysis of variance—that the samples came from populations with a common variance for age (*b* is true). Equality of variances for age between groups could have been verified by a statistical test, such as Levene's test, which is provided routinely by statistical software.

The null hypothesis in analysis of variance is a global comparison between groups. In the example above, the null hypothesis was that the mean age of the surgical groups (gastric bypass, gastric banding, and sleeve gastrectomy) was equal in the populations from which the samples were obtained. The null hypothesis did not involve pairwise comparisons of mean age between the groups (*c* is false). The alternative hypothesis stated that the mean ages of patients in the three surgical groups were not equal in the respective populations. It was not specified

whether a particular surgical group had a younger mean age when compared with the other two groups—the alternative hypothesis was in effect a two sided alternative. The reported P value for this statistical test was 0.364. Therefore, there was no evidence to reject the null hypothesis in favour of the alternative; it was concluded that the three groups had a common mean age in the populations from which the samples were taken. Analysis of variance is called such because the total variation between sample members for the variable analysed can be partitioned into two sources—the variation between groups and the variation within groups. In the example above, the between group variation was the variation in age that existed between patients belonging to the different surgical groups. The within group variation was the variability that existed between patients in age within each surgical group. The within group variation is sometimes called the residual or unexplained variability. The ratio of these two variances forms the basis for the statistical test of one way analysis of variance. Under the null hypothesis of no difference between the three surgical groups in mean age, the between group and within group variances would be expected to be equal. If the three groups differed in mean age then the variation between groups would be larger than the variation within groups and the null hypothesis would be rejected in favour of the alternative.

When only two groups are compared using one way analysis of variance, the statistical test is equivalent to the *t* test for two independent groups; the two statistical tests will give the same P value and result in the same conclusion. The *t* test for two independent groups has been described in a previous question.⁴ Although the *t* test for two independent groups could have been used to undertake pairwise comparisons between surgical

groups, this would not have been sensible because multiple significance tests would have increased the probability of a type I error. A type I error, described in a previous question,⁵ would have occurred if the null hypothesis had been rejected in favour of the alternative when in fact the mean age between surgical groups in the populations did not differ. For that reason, analysis of variance was used to make a global comparison between groups because such an approach does not affect the probability of making a type I error. If analysis of variance had shown a significant difference between surgical groups in mean age, it would then have been suitable to undertake pairwise comparisons between groups to establish which groups differed in mean age. Many methods have been proposed for making these post hoc comparisons (analyses not specified in advance of data collection), including Scheffe's test, Dunnett's test, and Newman Keuls method. The application of these tests will depend on a variety of circumstances, including whether groups have equal numbers of patients. The tests are routinely provided by statistical software and have all been designed to ensure the probability of making a type I error is kept to a minimum.

Competing interests: None declared.

- 1 Burns EM, Naseem H, Bottle A, Lazzarino AI, Aylin P, Darzi A, et al. Introduction of laparoscopic bariatric surgery in England: observational population cohort study. *BMJ* 2010;341:c4296.
- 2 Sedgwick P. Parametric v non-parametric statistical tests. *BMJ* 2012;344:e1753.
- 3 Sedgwick P. The Normal distribution. *BMJ* 2010;341:c6085.
- 4 Sedgwick P. Independent samples t test. *BMJ* 2010;340:c2673.
- 5 Sedgwick P. Multiple significance tests. *BMJ* 2010;340:c2963.

Cite this as: [BMJ 2012;344:e2427](https://doi.org/10.1136/bmj.e2427)

© BMJ Publishing Group Ltd 2012