BMJ

BMJ 2012;344:e3354 doi: 10.1136/bmj.e3354 (Published 16 May 2012)

ENDGAMES

STATISTICAL QUESTION

Non-parametric statistical tests for independent groups: numerical data

Philip Sedgwick *senior lecturer in medical statistics*

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

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

In total, 3191 gastric bypass, 3649 gastric band, and 113 sleeve gastrectomy procedures were performed. A significant difference was reported between surgical procedures in the median length of stay (gastric bypass: 5 days (interquartile range 3-7 days); gastric banding: 2 (1-3); sleeve gastrectomy 4 (3-7); P<0.001). A non-parametric test was used to compare groups in the length of hospital stay, as distributional assumptions could not be made.

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

Which one of the following statistical tests was most likely used to compare length of hospital stay in the three groups?

a) Analysis of variance

b) Friedman two way analysis of variance by ranks

c) Kruskal-Wallis test

d) Wilcoxon rank sum test

Answers

The Kruskal-Wallis test (answer *c*) would most likely have been used to compare length of hospital stay in the three groups.

The Kruskal-Wallis test compares three or more groups in a variable that is continuous or ordinal. The groups must be independent—that is, study participants can belong to only one group. In the example above, cohort members could have had only one type of surgery as their primary elective bariatric procedure. Because participants in the groups are unrelated to each other, the comparison is made between groups or between participants. The number of participants does not have to be the same in each group. The Kruskal-Wallis test is a non-parametric test that makes no assumption about the distribution of the variable in the population; non-parametric tests have been described in a previous question.[2] Two types of statistical test exist: parametric and non-parametric. Parametric methods assume that the variable being compared across the groups has a particular distribution, typically normal, in the population from which each group was sampled. Analysis of variance (answer *a*), described in a previous question,[3] is the parametric equivalent of the Kruskal-Wallis test. The researchers reported that distributional assumptions could not be made about the length of hospital stay, and therefore analysis of variance was not appropriate.

The null hypothesis for the Kruskal-Wallis test is a global comparison between groups. In the above example, the null hypothesis was that the distributions of length of hospital stay for the three surgical groups (gastric bypass, gastric banding, and sleeve gastrectomy) were the same in the population from which the samples were taken. The alternative hypothesis stated that the distributions of length of hospital stay for the three surgical groups were not the same in the population from which the samples were taken. No direction was specified as to whether a particular surgical group had a shorter or longer length of stay than the other two groups—the alternative hypothesis was in effect a two sided alternative.

The Kruskal-Wallis test is based on the sum of ranks of the values of length of hospital stay in each of the three groups. Under the null hypothesis, if the distribution of length of stay were the same for each group in the population, then the average rank for values of length of stay would be expected to be equal for the three groups in the sample. The reported P value was <0.001. Therefore, there was evidence to reject the null

p.sedgwick@sgul.ac.uk

hypothesis in favour of the alternative: it was concluded that the distribution of length of hospital stay differed between the three groups in the population from where the samples were taken. Those patients who underwent gastric banding had the shortest length of stay, while patients undergoing gastric bypass had the longest.

The Wilcoxon rank sum test (answer *d*) is used to compare two independent groups in a variable measured on a continuous or ordinal scale. The Wilcoxon rank sum test is equivalent to the Mann-Whitney U test in that both give the same P value, resulting in the same conclusion with respect to statistical hypothesis testing. These tests are non-parametric methods and therefore make no assumption about the distribution of the variable in the population. The Kruskal-Wallis test is thought of as an extension of the Mann-Whitney U and Wilcoxon rank sum tests, as it compares the distribution of a variable between three or more independent groups.

Friedman two way analysis of variance by ranks (answer *b*) compares three or more related groups in a variable that is continuous or ordinal. It is a non-parametric test that makes no assumption about the distribution of the variable in the population. An example where the test may be used is the comparison of effectiveness of three non-steroidal anti-inflammatory drugs in reducing knee pain and disability in patients with osteoarthritis of the knee. Each patient would receive each drug in a random sequence, and knee pain and disability would be measured on an ordinal scale by a questionnaire. The number of participants in each group would be equal, with each drug representing a different group. The comparison would be between treatments within patients. Friedman two way analysis of variance by ranks is thought of as an extension to the Wilcoxon signed ranks test. Described in a previous question,[4] the Wilcoxon signed ranks test is a non-parametric test that compares two related samples in a variable that is continuous or ordinal.

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. One way analysis of variance. *BMJ* 2012;344:e2427.
4 Sedgwick P. Non-parametric statistics tests for two related groups: numerical data. *BMJ* 2012;344:e2537.

Cite this as: *BMJ* 2012;344:e3354