

## ENDGAMES

## STATISTICAL QUESTION

## Parametric statistical tests for independent groups: numerical data

Philip Sedgwick *reader in medical statistics and medical education*

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

Researchers investigated whether a school based educational programme aimed at reducing consumption of carbonated drinks prevented excessive weight gain in children. A cluster randomised controlled trial study design was used. The intervention, which was delivered over one school year, included focused education promoting a healthy diet together with discouragement of carbonated drink consumption. The control group received no intervention. Children were followed for three years from baseline.<sup>1</sup>

The main outcome measures included body mass index (BMI) converted to age and sex specific z scores. A total of 644 children aged between 7 and 11 years from six schools were recruited. Measurements were obtained from 434 children three years after baseline. Distributional assumptions of normality in the BMI z scores were verified. At follow-up the age and sex specific BMI z scores had increased in the control group by a mean of 0.1 (SD 0.53) but decreased in the intervention group by 0.01 (SD 0.58). The mean difference between treatment groups was not significant (0.1 (95% confidence interval 0 to 0.21); P=0.06).

Which one of the following statistical tests was most likely used to compare treatment groups in the mean difference in the mean change in BMI z score over three years from baseline?

- Paired *t* test
- Student's *t* test
- Wilcoxon rank sum test
- Wilcoxon signed ranks test

### Answers

Student's *t* test (answer *b*) would most likely have been used to compare the treatment groups in the mean difference in mean change in BMI z score over three years from baseline.

The trial investigated whether an educational programme aimed at reducing consumption of carbonated drinks prevented excessive weight gain in children. The outcome measures included BMI, recorded at baseline and after three years. The distribution of BMI would have been different for children of

different ages, with each distribution described by a unique mean and standard deviation. Boys would have been expected to have on average a greater BMI than girls. Each child's BMI was therefore transformed to a z score specific for their age and sex. The standardisation of outcome measures by using z scores has been described in a previous question.<sup>2</sup> Each child's change in BMI z score was calculated—that is, his or her BMI z score at baseline was subtracted from that after three years. The mean change in BMI z score over three years in the intervention group was compared with that in the control group, providing an estimate of the true effect of the school based educational programme at preventing excessive weight gain in children.

Student's *t* test (answer *b*) would most likely have been used to compare treatment groups in the mean change in BMI z score over the three years of follow-up. Student's *t* test, also known as the independent samples *t* test and described in a previous question,<sup>3</sup> compares the means of a variable measured on a continuous scale in two independent groups. It is a parametric test that assumes that the distribution of change scores in the age and sex specific BMI z scores were approximately normally distributed in both groups and that the variances for the two groups were equal. The researchers reported that distributional assumptions of normality in the BMI z scores had been verified. Parametric tests have been described in a previous question.<sup>4</sup>

A further indication that Student's *t* test (answer *b*) would most likely have been used was that a 95% confidence interval for the mean difference between treatment groups in mean change in BMI z scores from baseline was presented. A 95% confidence interval for the mean difference should have been derived only if the assumption of normality required for a parametric test could be made. If the assumption of normality could not have been made, the Wilcoxon rank sum test (answer *c*)—a non-parametric test described below—would have been used. However, under such circumstances it would not have been sensible to derive the 95% confidence interval for the mean difference between the treatment groups.

In the statistical test to compare the intervention with the control treatment, the null hypothesis stated that in the population from

which the sample was taken there was no difference between treatment groups in the mean change in BMI z score over three years of follow-up. The alternative hypothesis was two sided: in the population the intervention, when compared with control treatment, resulted in a larger or a smaller mean change in BMI z score over three years of follow-up. Although at follow-up the BMI z scores had increased in the control group by a mean of 0.1 but decreased in the intervention group by 0.01, the mean difference between the groups was not significant ( $P=0.06$ ) at the 5% level of significance. Therefore, there was no evidence of a difference between treatments in the effect on the mean change in BMI z score over three years.

The paired  $t$  test (answer *a*) is used to compare two related measurements of a continuous variable. The age and sex specific BMI z score of each child were recorded at baseline and again three years later. The paired  $t$  test could have been used to test whether the mean change in BMI z scores over three years was significantly different from zero in each treatment group. The paired  $t$  test is a parametric test and would have assumed that the distribution of the change in BMI z scores in the population was normally distributed. Although analysis of the mean change in BMI z scores in each treatment group may be useful, it would be more informative to compare the mean change in the intervention group with that in the control group. Children in the control group did not receive any intervention, and thus this group provided an estimate of the natural change in BMI z scores over three years. However, investigating the mean change within groups can be misleading, as a previous question has described.<sup>5</sup>

The Wilcoxon rank sum test (answer *c*) and Wilcoxon signed ranks test (answer *d*) are non-parametric methods that have been

described in previous questions.<sup>6,7</sup> The Wilcoxon rank sum test is used to compare two independent groups in a variable measured on a continuous or ordinal scale. The Wilcoxon signed ranks test is used to compare two related samples in a variable that is continuous or ordinal. The Wilcoxon rank sum test and Wilcoxon signed ranks test are non-parametric methods and therefore make no assumption about the distribution of the variable in the population. The tests are used when the distribution of the variable does not satisfy the assumption of normality or when it is not achieved after a transformation of the data. The log transformation of data has been described in a previous question.<sup>8</sup> If the assumption of normality could not have been made, and non-parametric tests had been used, then it would not have been sensible to calculate a 95% confidence interval for the mean difference between treatment groups in the mean change in BMI z scores from baseline.

Competing interests: None declared.

- 1 James J, Thomas P, Kerr D. Preventing childhood obesity: two year follow-up results from the Christchurch obesity prevention programme in schools (CHOPPS). *BMJ* 2007;335:762-5.
- 2 Sedgwick P. Standardisation of outcome measures (z scores). *BMJ* 2012;345:e6178.
- 3 Sedgwick P. Independent samples  $t$  test. *BMJ* 2010;340:c2673.
- 4 Sedgwick P. Parametric v non-parametric statistical tests. *BMJ* 2012;344:e1753.
- 5 Sedgwick P. Analysis of outcome measures within treatment groups. *BMJ* 2012;345:e7201.
- 6 Sedgwick P. Non-parametric statistical tests for independent groups: numerical data. *BMJ* 2012;344:e3354.
- 7 Sedgwick P. Non-parametric statistical tests for two related groups: numerical data. *BMJ* 2012;344:e2537.
- 8 Sedgwick P. Log transformation of data. *BMJ* 2012;345:e6727.

Cite this as: *BMJ* 2012;345:e8145

© BMJ Publishing Group Ltd 2012