QUESTIONS

10.3 The following table shows the serum cholesterol concentrations of 20 smokers and 20 non-smokers, matched for age and sex, recruited for a study of glucose response. Statistical analysis was by two-tailed, unpaired Student's t test (Facchini *et al.* 1992).

Cholesterol concentration, mean (SEM) in mmol/l

	Smokers	Non-smokers	t test
Total plasma	4.49 (0.19)	4.48 (0.17)	NS
Very low density lipoprotein (VLDL)	0.45 (0.06)	0.23 (0.04)	P< 0.005
Intermediate density lipoprotein (IDL)	0.18 (0.03)	0.22 (0.04)	NS
Low density lipoprotein (LDL)	2.72 (0.17)	2.53 (0.13)	NS
High density lipoprotein (HDL)	1.16 (0.05)	1.51 (0.08)	P< 0.005

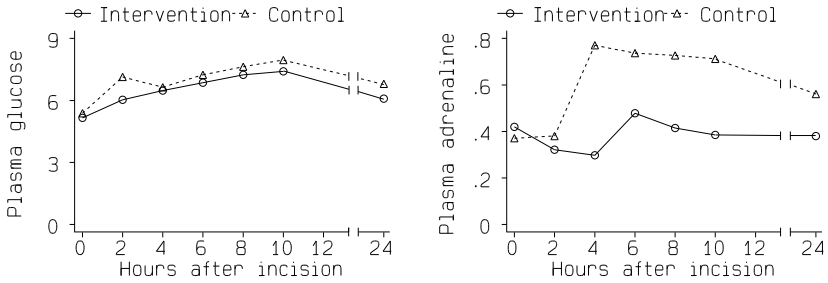
10.3.1 What is meant by a 'two-tailed, unpaired Student's t test'?

10.3.2 What conditions must the data satisfy for these t tests to be valid? Are these likely to satisfied here?

10.3.3 What extra information could be given in the table?

! 10.3.4 What aspect of the data has been ignored in the analysis?

10.4 In a randomized controlled trial carried out to evaluate the influence of preoperative abstinence on postoperative outcome in alcohol misusers, 42 alcoholic patients admitted for elective colorectal surgery were allocated either to withdrawal from alcohol consumption for one month before operation (disulfiram controlled) or to carry on with their usual drinking (Tonnesen *et al.* 1999). Among others, glucose and adrenalin were measured at several times. The means at each time were as follows:



The authors analysed these as the area under the curve (AUC). The treatment difference was significant for adrenalin, not for glucose.

+ 10.4.1 What is meant by 'area under the curve' and how can this be calculated?

+ 10.4.2 What are the advantages of this approach compared to analysing each time point separately?

ANSWERS

10.3.1 The two-tailed, unpaired Student's t test is used to compare means from two independent samples. It tests the null hypothesis that the means are the same in the populations from which the samples are drawn against the alternative hypothesis of a difference in either direction. If the null hypothesis is true then the difference between means divided by the standard error of the difference follows the t distribution with 38 $(= 20 - 1 + 20 - 1)$ degrees of freedom (*Intro* §9.5, 10.3).

10.3.2 The assumptions are that the cholesterol data are from Normal distributions with the same variance. In general serum concentrations are often positively skew with the variance increasing with the mean. Here, the standard errors are consistently bigger for the larger of the two means for each type of cholesterol. Since the sample size is equal for the smokers and non-smokers, the standard deviations must also increase with the means. Hence, the assumptions of the tests may not be met. However, with equal numbers in 2 groups the test is very robust, though some power may be lost (*Intro* §10.3, 10.5).

10.3.3 It would be useful to show the difference between the means and a 95% confidence interval for that difference. In addition, the actual p-value is more informative than 'NS'.

10.3.4 The analysis has ignored the matching for age and sex and has treated the groups as independent. A matched analysis, such as using a paired t test would take the structure of the data into account. If cholesterol is actually related to the matching variables age and sex, a paired test would remove some of the variation which is included in the standard error in the unpaired test. The paired test would be more powerful. If cholesterol were unrelated to the matching variables then a paired test would not be necessary. When the sample is very small, the loss of degrees of freedom may even make a paired test less powerful and so be counter-productive in these circumstances.

10.4.1 The area under the curve is a method for combining a series of measurements made at different times into a single observation. It takes into account the actual time intervals. These may vary, as in this example, where the last time interval (14 hours) is much longer than the others (all 2 hours). It is usually calculated by the trapezium method. For each time interval, the average of the measurements at the beginning and end is multiplied by the length of the interval. These are then added to give the AUC (*Intro* §10.7).

10.4.2 As the authors of this paper say, it reduces the number of tests and so reduces the risk of a spurious significant difference when the null hypothesis is true. It also uses all the information in a single test and so increases the power of the test, the chance of getting a significant difference when the null hypothesis is false (*Intro* §9.10, 10.7).