SOS POLITICAL SCIENCE & PUBLIC ADMINISTRATION, JIWAJI UNIVERSITY GWALIOR

MBA HRD II SEM

PAPER CODE 206

SUBJECT NAME: QUANTITATIVE TECHNIQUES FOR MANAGERS

TOPIC NAME: T TESTS

The t tests

Previously we have considered how to test the null hypothesis that there is no difference between the mean of a sample and the population mean, and no difference between the means of two samples. We obtained the difference between the means by subtraction, and then divided this difference by the standard error of the difference. If the difference is 196 times its standard error, or more, it is likely to occur by chance with a frequency of only 1 in 20, or less.

- With small samples, where more chance variation must be allowed for, these ratios are not entirely accurate because the uncertainty in estimating the standard error has been ignored. Some modification of the procedure of dividing the difference by its standard error is needed, and the technique to use is the *t* test. Its foundations were laid by WS Gosset, writing under the pseudonym "Student" so that it is sometimes known as Student's *t* test. The procedure does not differ greatly from the one used for large samples, but is preferable when the number of observations is less than 60, and certainly when they amount to 30 or less.
- The application of the *t* distribution to the following four types of problem will now be considered.
- The calculation of a confidence interval for a sample mean.
- The mean and standard deviation of a sample are calculated and a value is postulated for the mean of the population. How significantly does the sample mean differ from the postulated population mean?
- The means and standard deviations of two samples are calculated. Could both samples have been taken from the same population?
- Paired observations are made on two samples (or in succession on one sample). What is the significance of the difference between the means of the two sets of observations?

- In each case the problem is essentially the same - namely, to establish multiples of standard errors to which probabilities can be attached. These multiples are the number of times a difference can be divided by its standard error. We have seen that with large samples 1.96 times the standard error has a probability of 5% or less, and 2.576 times the standard error a probability of 1% or less (Appendix table A). With small samples these multiples are larger, and the smaller the sample the larger they become.

Confidence interval for the mean from a small sample

- A rare congenital disease, Everley's syndrome, generally causes a reduction in concentration of blood sodium. This is thought to provide a useful diagnostic sign as well as a clue to the efficacy of treatment. Little is known about the subject, but the director of a dermatological department in a London teaching hospital is known to be interested in the disease and has seen more cases than anyone else. Even so, he has seen only 18. The patients were all aged between 20 and 44.

- The mean blood sodium concentration of these 18 cases was 115 mmol/l, with standard deviation of 12 mmol/l. Assuming that blood sodium concentration is Normally distributed what is the 95% confidence interval within which the mean of the total population of such cases may be expected to lie?
- The data are set out as follows:

Number of observations	18
Mean blood sodium concentration	115 mmol/l
Standard deviation	12 mmo1/l
Standard error of mean	$SD/\sqrt{n} = 12/\sqrt{18} = 2.83$ mmol/l

- To find the 95% confidence interval above and below the mean we now have to find a multiple of the standard error. In large samples we have seen that the multiple is 1.96 (Chapter 4). For small samples we use the table of *t* given in Appendix Table B.pdf. As the sample becomes smaller t becomes larger for any particular level of probability. Conversely, as the sample becomes larger t becomes smaller and approaches the values given in table A, reaching them for infinitely large samples.

- Since the size of the sample influences the value of t, the size of the sample is taken into account in relating the value of *t* to probabilities in the table. Some useful parts of the full t table appear in . The left hand column is headed d.f. for "degrees of freedom". The use of these was noted in the calculation of the standard deviation (Chapter 2). In practice the degrees of freedom amount in these circumstances to one less than the number of observations in the sample. With these data we have 18 - 1 = 17 d.f. This is because only 17 observations plus the total number of observations are needed to specify the sample, the 18th being determined by subtraction.

- To find the number by which we must multiply the standard error to give the 95% confidence interval we enter table B at 17 in the left hand column and read across to the column headed 0.05 to discover the number 2.110. The 95% confidence intervals of the mean are now set as follows:
- Mean + 2.110 SE to Mean - 2.110 SE
- which gives us:
- 115 - (2.110 x 283) to 115 + 2.110 x 2.83 or 109.03 to 120.97 mmol/l.

- We may then say, with a 95% chance of being correct, that the range 109.03 to 120.97 mmol/l includes the population mean.
- Likewise from Table the 99% confidence interval of the mean is as follows:
- Mean + 2.898 SE to Mean - 2.898 SE
- which gives:
- 115 - (2.898 x 2.83) to 115 + (2.898 x 2.83) or 106.80 to 123.20 mmol/l.
- Difference of sample mean from population mean (one sample *t* test)

- Estimations of plasma calcium concentration in the 18 patients with Everley's syndrome gave a mean of 3.2 mmol/l, with standard deviation 1.1. Previous experience from a number of investigations and published reports had shown that the mean was commonly close to 2.5 mmol/l in healthy people aged 20-44, the age range of the patients. Is the mean in these patients abnormally high?