Appendix F Statistical Methods For Data Analysis And Derivation

In the main text of Deliverable 2, reference is made to the need for a statistical approach towards data analysis. This Appendix contains brief details of some of the possible methods that could be used to check the statistical significance of the experimental data. The techniques are split according to the nature of the data to be analysed. The data types are defined as follows:

Quantitative, or 'hard', parameters: parameters that are expressed in terms of counts, measurements, or other physical units. The applied measurement techniques are usually human observation and automatic measurement.

Qualitative, or 'soft', parameters: parameters that are expressed in terms of people's attitudes, perceptions, or observation. Measurements are usually conducted by interviews and questionnaires, e.g. "Stated preference".

F.1 TOOLS FOR HARD PARAMETERS

Hard parameters refer to variables which are directly measured such as speed or journey time, i.e. they are usually obtained by automatic measurement. Such parameters are often presented on an interval scale with a recognisable distribution function. Therefore, appropriate techniques for comparing mean values are:

- 't' test,
- Analysis of variance - 'one-way',
- Analysis of variance - 'many-way'

F.1.1 The 't'-Test

The 't' test is appropriate for comparing the mean values from two samples. There are different formulae for computing the 't' value, depending on whether the standard deviations from the two samples are assumed to be the same or different. In the most likely case the population mean and standard deviation are unknown and estimated from the sample. If it can be established, before the 't' statistic is calculated, that the variability in one sample is different from the other then appropriate formula can be used.

If the samples are matched, then the differences between the matched responses are computed and these become the sample data. The null-hypothesis is that the mean difference is zero, i.e. that there is no difference between the two sets of results. This is a more powerful version of the test because the differences between individuals have been controlled for.

Once the 't' statistic has been calculated then it must be checked to see if it exceeds the critical value. This depends upon the Type I error, probably 5%, and the degrees of freedom, i.e. on the sample size. Tables of critical 't' values are readily available, and most computer packages will compute the probability of the 't' value occurring for the appropriate degrees of freedom. The user should have defined the Type I error that he is going to consider significant and if the test is one-sided or two-sided before the 't' value is calculated.

F.1.2 Analysis of Variance - 'one-way'

The technique of analysis of variance assumes that the total variation in the data can be partitioned into component variances attributable to groups of interest, plus a residual within-group variation. In 'one-way' analysis of variance only one factor (which may have many groups), is considered. For example, when regarding journey times, the sample variation from the whole data set can be split into the variation due to different route guiding system applications and the residual journey time variation within each of the applications, i.e. for each group member.

The variance between the guiding system applications divided by the residual within each group variance gives an 'F' value. The probability of this 'F' value occurring can be found from published tables or will be computed by the computer package. If the 'F' value is sufficiently large so that one can be 95% certain that it was not by chance, then the null hypothesis - that there are no differences between the route guiding systems applications - can be rejected. If we are considering just one guiding system application versus the pre-implementation situation i.e. 2 groups, then the analysis is complete because we know the mean values are significantly different from each other. (The 't' test could also have been used and would have given the same result). However, if there are several applications then we know that there is a significant difference, but we do not know between which applications.

The significantly different application(s) may or may not be obvious from looking at the mean values of each application, and so formal tests for significant differences should be computed. There are several techniques for comparing group means, i.e. application means. It is certainly feasible to construct a number of 't' tests between each pair but it is important to dynamically adjust the Type I error for each 't' test in order to maintain the overall Type I error of 5% say. This is necessary because the overall Type I error will be greater than 5% if each possible pair is compared.

Most statistical packages have a variety of techniques for controlling the overall Type I error levels when multiple comparisons are requested. It is important to be aware of the potential problem and take advice from the statistical package manual and / or from a statistician.

F.1.3 Analysis of Variance - 'many-way'

There may be more than one factor of interest, for example there may be alternative route guiding system applications and it may be of importance to consider the time of day that the journey is taken. In this case, time of day is another factor, with perhaps morning / mid-day / evening as three times of interest. It is desirable when more than one factor is involved to have the same number of data values per combination of factors, i.e. the same sample for pre-implementation morning as post-implementation guiding system application in the evening etc. The components of variance can only be uniquely estimated if this condition is met.

There are methods of handling so called unbalanced designs, but it is best if they can be avoided. With any experiment where more than one factor is being evaluated it is important to have an appropriate design. The statistical analysis should reflect the experimental design. There are many experimental designs that can be used, but if complex experiments are intended it is best to discuss the appropriate design with an expert.

F.2 TOOLS FOR SOFT PARAMETERS

Soft parameters will be on nominal or at best ordinal scales, like on a Lickert scale or perhaps yes / no responses on a questionnaire. Sometimes responses based on a Lickert scale are assumed to be sufficiently as an interval scale because they can be summarised by a mean and standard deviation and analysed using 't'-Test etc. Generally, soft parameters are analysed by comparing category counts and use chi-squared or log-linear modelling.

Suppose the data consists of counts, for example the responses to an interview or a questionnaire. As an example let us assume that an RDS-TMC system for the main routes into a city centre has been implemented and the impact of this on congestion is being assessed. Congestion is being measured on a 5 point scale, (none / a little / some / quite a lot / extensive), and a sample of drivers have been asked by a team of interviewers to give an opinion. A sample of drivers were asked before and after the RDS-TMC system went 'live', on the same days of the week, the same time of year and during the same time of day. Time had been left between samples for the new system to 'settle down'.

The null-hypothesis to be tested is that the distribution of responses for the before and after implementation samples is the same, i.e. there is no interaction between the perception of congestion and the implementation of the RDS-TMC application. We have a 2-way table of counts with a cell for each point on the 5-point scale both 'before' and 'after', i.e. 10 cells. The chi-squared statistic is an appropriate method to analyse this table and to test this hypothesis.

If we can assume that the 5-point scale is ordinal (one-way) then the Kruskall-Wallis test is more appropriate, which is a non-parametric test for differences. That means it will indicate a statistical difference between any of the means.

The group of drivers who thought that there was extensive congestion may be more of interest. The proportion of drivers in that group 'before' and 'after' can be estimated and compared under the null-hypothesis that there is no change from before to after. A special form of the 't' test is appropriate in this instance.

Suppose that the time of day and the day of week were considered to be influential factors in the response patterns of driver perceived congestion. The 2-way table discussed above becomes a 4-way table. If 3 time periods in the day and 7 days of the week are being considered the 4-way table now has a total of (5x2x3x7)=210 cells. Clearly, not only does the sample of drivers have to be sufficiently large to give an adequate number per cell, but a different analytical technique, i.e. log-linear modelling is also required.

Generally, soft parameters are analysed by the following tools:

1. chi-squared
2. comparing category counts/ proportions
3. log-linear modelling

F.2.1 Chi-squared

The well known chi-squared test provides an appropriate method of analysing count data. The test is usually applied to 2-way tables of counts. i.e. a table of responses to a survey about congestion before and after the implementation of a city RDS-TMC service, perhaps like the table below:

Table F1. Opinion about level of congestion.

		None	little	Some	quite a bit	extensive	Total
Before	Number	10	20	40	20	30	120
	%	8.3%	16.7%	33.3%	16.7%	25.0%	100%
After	Number	10	30	50	10	10	110
	%	9.1%	27.3%	45.5%	9.1%	9.1%	100%
Total	Number	20	50	90	30	40	230
	%	8.7%	21.7%	39.1%	13.0%	17.4%	

The null-hypothesis for the chi-squared test is that the rows and columns in the table to be analysed are independent of each other. If this is true then the counts in the table can be predicted just using the row totals and the column totals and a chi-squared value is computed. If the chi-squared value is large relative to what is expected then the null-hypothesis is rejected. The degrees of freedom for a chi-square are the same as its expected mean value and is computed by (rows-1)*(columns-1), so in this example would be (2-1)*(5-1)=4.

The probability for a value of chi-squared occurring, given the degrees of freedom, are readily available from published tables or any statistical package. The null-hypothesis may thus be accepted or rejected at the chosen Type I error level.

In the example above the chi-squared value is 16.04, based on 4 degrees of freedom, so the value greatly exceeds the expected mean value. The probability of this value of chi-squared occurring by chance is about 1 in 200, hence the null-hypothesis would be rejected. This is not surprising as the row percentages in the table show that there has been a clear shift in opinion about congestion. In other words there is a relationship between the rows and columns, and they cannot be regarded as independent of one another.

The chi-squared test makes no assumptions about any natural ordering of the rows or columns, only that they are mutually exclusive. If there is a natural ordering in the rows or the columns then other techniques, such as the Kruskall-Wallis test or the Friedman test may be more appropriate.

F.2.2 Proportions

Sometimes the impact measure of interest is a proportion of occurrences, for example the proportion of drivers in our example who thought that congestion was excessive. We can see from the table above that there were 25% of drivers before and 9.1% of drivers after the implementation of an RDS-TMC system who thought that congestion was excessive, and the sample sizes were 120 and 110 respectively. Are these proportions significantly different?

If we assume that the sample proportions have been drawn from a binomial distribution then we can compute a version of the 't' test statistic. The null-hypothesis is that both the sample proportions are binomially distributed with a mean proportion of p say, which for the example is 0.174. The standard deviation for a Binomial distribution is given by $\sqrt{p(1-p)}$, which in this example is 0.38. A 't' statistic can then be computed and a decision made whether to accept or reject the null-hypothesis.

F.2.3 Log-linear modelling

Often, when faced with cross-tabulated data then a chi-squared test of independence is computed for each sub-table (as discussed in section F2.1). This can be a useful first step in checking which categorical variables seem important, but it can estimate only the effects between just two variables. This can be problematic, especially if there are significant interactions between variables.

The technique of log-linear models is appropriate for the analysis of the inter-relationships between categorical variables. In many respects it can be regarded as an extension to the chi-squared test, but with interaction terms specifiable and with the possibility to include as many variables as required.