Appendix B

Statistical Tests

Student's t-test for comparison of samples

B.1

Student's t-test for comparison of samples (eg. sets of mean speed measurements)

To determine whether the mean speed of one set of speed measurements is significantly different from another (ie. between a "before" and "after" study), it is appropriate to use Student's two-tailed t-test, making the reasonable assumption that the variances of the two sets of measurements are drawn from the same population. The null hypothesis is thus that there is no difference in the means (ie. that drivers' speed has not been affected by the scheme). It is first necessary to determine the standard deviation of the difference in means.

Let $b_1, b_2, \ldots\ldots. b_{nb}$ be the *Before* speed readings
and $a_1, a_2, \ldots\ldots. a_{na}$ be the *After* speed readings

We then calculate the equations below:

$$\text{Means:} \quad \bar{b} = \frac{\Sigma(b_i)}{n_b}, \quad \bar{a} = \frac{\Sigma(a_i)}{n_a}$$

$$\text{Standard deviation: } \sigma^2 = \sqrt{\frac{\Sigma(a_i^2) - \frac{(\Sigma(a_i))^2}{n_a} + \Sigma(b_i^2) - \frac{(\Sigma(b_i))^2}{n_b}}{(n_a + n_b - 2)}}$$

$$t = \frac{\bar{a} - \bar{b}}{\sigma} \sqrt{\frac{n_a \times n_b}{n_a + n_b}}$$

Having found the value of ***t*** we need to look at a table of Student's t values (see page B-4), with $(n_a + n_b - 2)$ degrees of freedom. If the value of ***t*** exceeds that for the 5% level (the ***t*** = 0.05 column) we can be 95% confident that the true mean speed has changed.

Example

Assume that number of speed readings

before a scheme,	$n_b = 210$
and the mean,	$\bar{b} = 37 \text{ mile/h}$
sum of readings	$\sum b_i = 7770$
sum of squares	$\sum (b_i)^2 = 291{,}142$
Similarly, after a scheme,	$n_a = 220$
	$\bar{a} = 33 \ mile/h$
	$\sum a_i = 7260$
	$\sum (a_i)^2 = 243{,}760$

From the above equations

$$\text{standard deviation, } \sigma = \sqrt{\frac{243760 - (7260)^2/220 + 291142 - (7770)^2/210}{220 + 210 - 2}}$$
$$= 18.299$$

$$t = \frac{33-37}{18.299} \cdot \sqrt{\frac{220 \times 210}{220+210}}$$
$$= 2.265$$

$$\text{for degrees of freedom, } \upsilon = 220 + 210 - 2$$
$$= 428$$

As the ***t*** value is greater than 1.96 (for the large number of degrees of freedom), then we can say that the mean difference in mean speeds (a 4 mile/h reduction) is significant at the 5% level.

Table of *t*-distribution

Degrees of Freedom, υ	*t*				
	0.10	0.05	0.02	0.01	0.001
1	6.314	12.706	31.821	63.657	636.619
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.941
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.859
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.405
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.053	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.310	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

Kolmogorov-Smirnov test

B.2

The ‘two-tailed’ Kolmogorov-Smirnov test determines whether two independent samples have been drawn from the same population (or from populations with the same distribution). If the two samples have in fact been drawn from the same population (the null hypothesis), then the cumulative distributions of both samples may be expected to be fairly close to each other, ie. they should show only random deviation from the population distributions. If the two sample cumulative distributions are too far apart at any point this suggests that they come from different populations. Thus a large enough deviation between the two sample cumulative distributions is evidence for rejecting the null hypothesis.

Let $S_{Na}(x)$ be the observed cumulative step function of the first speed sample

ie. $S_{Na}(x) = K/N_a$ where K is the number of vehicles equal to or less than x km/h and N_a is the total number of the sample. Let $S_{Nb}(x)$ be the cumulative step function of the second sample. Now the Kolmogorov-Smirnov two-tail test focuses on the maximum deviation, D.

$$D = \text{maximum} \mid S_{Na}(x) - S_{Nb}(x) \mid \quad \text{..........................(1)}$$

For large samples (N>40) Kolmogorov-Smirnov tables show that the value of D must equal or exceed the value of:

$$1.36.\sqrt{\frac{N_a + N_b}{N_a N_b}}$$

to reject the null hypothesis at the 5 per cent level, that is, that they are not from the same population.

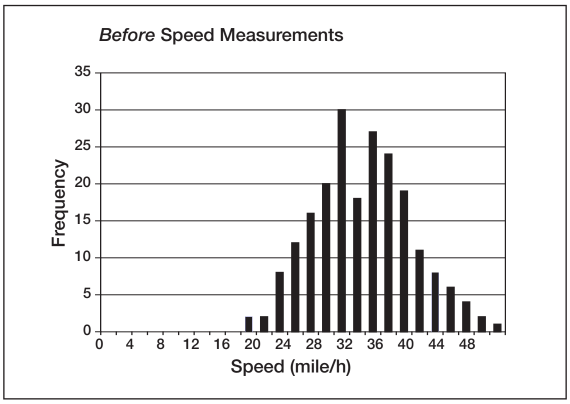
The ‘one-tailed’ Kolmogorov-Smirnov test determines whether the two samples have been drawn from the same population or whether the values of one sample are stochastically larger than the values of the population from which the other sample was drawn. The maximum deviation is again calculated using equation (1) and the significance of the observed value of D can be computed by reference to the chi-squared distribution. It has been shown that for large samples:

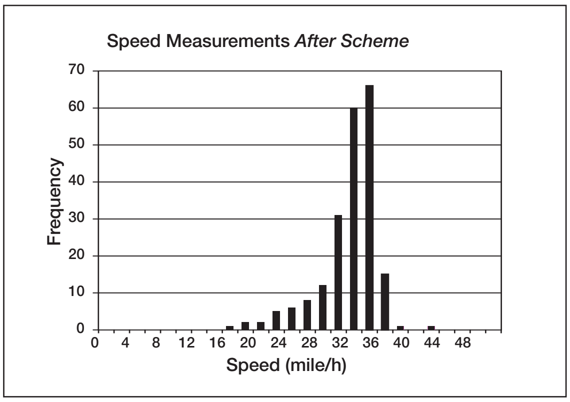
$$\chi^2 = 4D^2 \frac{N_a N_b}{N_a + N_b}$$

has a sampling distribution which is approximated to the chi-square distribution with two degrees of freedom. A chi-squared table for reference is given on page B-7.

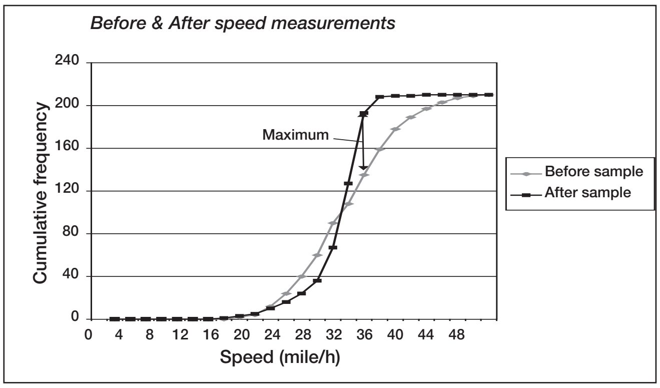
Example

Let us assume that *Before* and *After* speed measurements have given the following two distributions:–





If we plot these as cumulative speed distributions:



The observed cumulative step function of the *After* speed sample,

$$S_{Na}(x) = K/N_a = 193/210$$
$$= 0.919$$

For the *Before* sample,

$$S_{Nb}(x) = K/N_b = 135/210$$
$$= 0.643$$

The maximum deviation, $D = 0.919 - 0.643$
$$= 0.276$$

The Kolmogorov-Smirnov value at the 5% level

$$= 1.36\ (210+210/(210 \times 210))^{-}$$
$$= 0.133$$

which is less than the maximum deviation, and thus we can reject the null hypothesis at the 5% level. That is, in this case there is a significant difference between the two speed samples.

Table of χ^2

Degrees of Freedom, υ	0.99	0.98	0.95	0.90	0.50	0.10	0.05	0.02	0.01	0.001
1	0.000	0.001	0.004	0.015	0.455	2.710	3.840	5.410	6.640	10.830
2	0.020	0.040	0.103	0.211	1.386	4.610	5.990	7.820	9.210	13.820
3	0.115	0.185	0.352	0.584	2.366	6.250	7.820	9.840	11.340	16.270
4	0.297	0.429	0.711	1.064	3.357	7.780	9.490	11.670	13.280	18.470
5	0.554	0.752	1.145	1.610	4.351	9.240	11.070	13.390	15.090	20.520
6	0.872	1.134	1.635	2.204	5.350	10.650	12.590	15.030	16.810	22.460
7	1.239	1.564	2.167	2.833	6.350	12.020	14.070	16.620	18.480	24.320
8	1.646	2.032	2.733	3.490	7.340	13.360	15.510	18.170	20.090	26.130
9	2.088	2.532	3.325	4.168	8.340	14.680	16.920	19.680	21.670	27.880
10	2.558	3.059	3.940	4.865	9.340	15.990	18.310	21.160	23.210	29.590
11	3.050	3.610	4.570	5.580	10.340	17.280	19.680	22.620	24.730	31.260
12	3.570	4.180	5.230	6.300	11.340	18.550	21.030	24.050	26.220	32.910
13	4.110	4.760	5.890	7.040	12.340	19.810	22.360	25.470	27.690	34.120
14	4.660	5.370	6.570	7.790	13.340	21.060	23.690	26.870	29.140	36.120
15	5.230	5.990	7.260	8.550	14.340	22.310	25.000	28.260	30.580	37.700
16	5.810	6.610	7.960	9.310	15.340	23.540	26.300	39.360	32.000	39.250
17	6.410	7.260	8.670	10.090	16.340	24.770	27.590	31.000	33.410	40.790
18	7.020	7.910	9.390	10.870	17.340	25.990	28.870	32.350	34.810	42.310
19	7.630	8.570	10.120	11.650	18.340	27.200	30.140	33.690	36.190	43.820
20	8.260	9.240	10.850	12.440	19.340	28.410	31.410	35.020	37.570	45.320
21	8.900	9.910	11.590	13.340	20.340	29.610	32.670	36.340	38.930	46.800
22	9.540	10.600	12.340	14.040	21.340	30.810	33.920	37.660	40.290	48.270
23	10.200	11.290	13.090	14.850	22.340	32.010	35.170	38.970	41.640	49.730
24	10.860	11.990	13.850	15.660	23.340	33.200	36.420	40.270	42.980	51.180
25	11.520	12.700	14.610	16.470	24.340	34.380	37.650	41.570	44.310	52.620
26	12.200	13.410	15.380	17.290	25.340	35.560	38.890	42.860	45.640	64.050
27	12.880	14.120	16.150	18.110	26.340	36.740	40.110	44.140	46.960	55.480
28	13.560	14.850	16.930	18.940	27.340	37.920	41.340	45.420	48.280	56.890
29	14.260	15.570	17.710	19.770	28.340	39.090	42.560	46.690	49.590	58.300
30	14.950	16.310	18.490	20.600	29.340	40.260	43.770	47.960	50.890	59.700
40	22.164	23.838	26.509	29.051	39.335	51.805	55.759	60.436	63.691	73.402
50	29.707	31.664	37.689	37.689	49.335	63.167	67.505	72.613	76.154	86.661
60	37.485	39.699	43.188	46.459	59.335	74.397	79.082	84.580	88.379	99.607
70	45.442	47.839	51.739	55.329	69.334	85.527	90.531	96.388	100.425	112.317
80	53.539	56.213	60.391	64.278	79.334	96.578	101.880	108.069	112.329	124.839
90	61.754	64.634	69.126	73.291	89.334	107.565	113.145	119.646	124.116	137.208
100	70.065	73.142	77.929	82.358	99.334	118.498	124.342	131.142	135.807	149.449

The Tanner k test

B.3

The Tanner *k* test can be used to show how the accident numbers at a site change relative to control data.

For a given site or group of similarly treated sites, let:

a = before accidents at site
b = after accidents at site
c = before accidents at control
d = after accidents at control

then:

$$k = \frac{b/a}{d/c}$$

or, if any of the frequencies are zero then ½ should be added to each, ie:

$$k = \frac{(b + ½).(c + ½)}{(a + ½).(c + ½)}$$

If $k < 1$ then there has been a *decrease* in accidents relative to the control;
if $k = 1$ then there has been *no change* relative to the control; and
if $k > 1$ then there has been an *increase* relative to the control.

The percentage change at the site is given by:–

$$(k\text{-}1) \times 100\%$$

Example

Let us assume that the table below gives the annual injury accident totals for a priority T-junction in a semi-urban area which had Stop signs on the minor road originally, but where a roundabout was installed three years ago. The control data used are accidents on all other priority junctions in the Authority over exactly the same 3-year before and 3-year after periods.

Injury accident totals in 3-year periods at treated site and controls

	Site	*Control*	*Total*
Before	20 (a)	418 (c)	438 (g)
After	6 (b)	388 (d)	394 (h)
Total	26 (e)	806 (f)	832 (n)

Using the notation and formula above,

$$k = \frac{6/20}{388/418} = 0.323$$

Therefore, as $k < 1$ there has been a decrease in accidents relative to the controls of:

$$(k - 1) \times 100\% = 67.7\%$$

The Chi-Squared test

B4

This test can be used to determine whether the change in accidents was produced by the treatment or whether this occurred by chance. This test thus determines whether the change is statistically significant. The test is based on a table showing both the observed values of a set of data (O) and the corresponding expected values (E). The chi-squared statistic is then given by:

$$\chi^2 = \sum_{i=1,k=1}^{n,m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} is the observed value in column j, row i of the table
E_{ij} is the expected value in column j a, row i of the table
m is the number of columns
n is the number of rows

A chi-squared table (as on page B.2-3) is then used to look up this value which shows the probability that the 'expected' value and the 'observed' values are drawn from the same population. The number of degrees of freedom is also required and this is given by:

Degrees of freedom, υ = (n-1)(m-1).

For a site accident evaluation, where its accidents are compared in similar periods before and after treatment with a set of control sites for the same periods, we have a 2 by 2 contingency table (2 columns and 2 rows with degrees of freedom =1). For the test to be valid the value of any cell of the table should not ideally be less than 5. However, when testing an individual site for accidents then this situation can, of course, be quite common and so a slight modification (known as Yates' correction) is normally applied.

Example
Consider the same example as given in Appendix B3:

Injury accident totals in 3-year periods at treated site and controls

	Site	*Control*	*Total*
Before	20 (a)	418 (c)	438 (g)
After	6 (b)	388 (d)	394 (h)
Total	26 (e)	806 (f)	832 (n)

For such a 2x2 table, a special simplified formula can be used for chi-squared which, using the notation from the above table, is:

$$\chi^2 = \frac{\left(|ad - bc| - \frac{n}{2}\right)^2 .n}{efgh}$$

Its value is then compared with values in the Chi-squared table (page B-7) with degrees of freedom, $\upsilon = 1$, and if it is just greater than a particular value it is said to be statistically significant at at least that percentage level.

$$\chi^2 = \frac{\left(|\,20 \times 388 - 6 \times 418\,| - \frac{832}{2}\right)^2 \times 83}{26 \times 806 \times 438 \times 394}$$

$$\chi^2 = 5.38$$

Now looking at the chi-squared distribution table (page B-7) and the first line (one degree of freedom, $\upsilon = 1$), the value for chi-square of 5.38 lies between 3.84 and 5.41. This corresponds to a value of significance level (on the column header line) between 0.05 and 0.02, which is normally quoted as greater than the lower level, ie. better than the 5% level of significance.

This means that there is only a 5% likelihood (or 1 in 20 chance) that the change in accidents is due to random fluctuation. Another way of stating this is that there is a 95% (100%-5%) confidence that a real change in accidents has occurred at the junction.

The 5% level or better is widely accepted as the level in which the remedial action has certainly worked, though the 10% level can be regarded as an indication of an effect.

For groups of sites that have been given the same treatment, these can be grouped together and analysed using the chi-squared test as for a single site. This will enable the overall benefit to be evaluated, and any specific sites can be analysed separately.

Test for statistical significance between two proportions

B.5

This test is used to determine whether proportions (of accident types, or of any other characteristic) in a study area are significantly different from the proportion in a control area. The null-hypothesis tested is that the proportion from the sample is the same as the proportion from the control, and the test tells us if we can reject this hypothesis.

There are two situations to consider, firstly where the study area is not contained within the control area and secondly where it is within the control area.

Suppose that we are interested in the proportion of all accidents area that involve serious injury within a study as compared to a control area. We test the hypothesis that the proportions are the same. If the number of all accidents in the study area is n_s and in the control area is n_c, and we observe m_s serious accidents in the study area and m_c in the control area, then:

1. Study area not within control area

The proportion in the *Study* area is given by: $p_s = m_s / n_s$,
and the proportion in the *Control* area by: $p_c = m_c / n_c$
and the overall proportion in the *Total* area (both study and control areas) by:

$$p = (m_s + m_c) / (n_s + n_c)$$

The test statistic '**t**' is calculated by: $\mathbf{t} = (p_s - p_c) / (p(1\text{-}p) (1/n_s + 1/n_c))^{1/2}$

with $(n_s + n_c - 2)$ degrees of freedom. If the degrees of freedom are greater than 120, and **t** is greater than 1.96 then we can be 95% sure that the two proportions are from different populations.

2. Study area within control area

Suppose the study area is a local authority area and national data are being used as a control. Then, for the purposes of this test, the *Study* accidents need to be excluded from the *Control* and the numbers of accidents in the *Control* area is calculated as 'the *Total* (national) accidents – *Study* accidents'.

The proportion in the *Study* area is given by: $p_s = m_s / n_s$

and the proportion in the *Control* area by: $p_c = (m_c - m_s) / (n_c - n_s)$

and the overall proportion in the *Total* area by: $p = m_c / n_c$

The test statistic '**t**' is calculated by: **t** = $(p_s - p_c) / (p(1-p) (1/n_s + 1/(n_c - n_s)))^{1/2}$

with $(n_c - 2)$ degrees of freedom. If the degrees of freedom are greater than 120, and **t** is greater than 1.96 then we can be 95% sure that the two proportions are from different populations. (If n_c is large compared to n_s, then we can ignore the fact that the study area is within the national area and use method 1).

Example

Suppose that we are interested in whether the proportion of accidents at rural junctions in the study area is different from the proportion nationally. Then consider the following (fictitious) data:

	Rural junction accidents	*All Rural accidents*	*Proportion at junctions*
Total accidents nationally	32,000	80,000	0.400
Study area	3200	7750	0.4129

Since rural junction accidents are included within all rural accidents, approach 2 is the appropriate test. The null-hypothesis is that the proportion of rural accidents that are at junctions in the study area is the same as the proportion of accidents elsewhere in the country that are at junctions.

The proportion in the *Study* area is given by: p_s = 3200/7750=0.4129
and the proportion in the *Control* area by: p_c = (32,000-3200)/(80,000-7750)=0.3986
and the overall proportion in the *Total* area by: p = 32,000/80,000=0.400

The test statistic '**t**' is calculated by:

$$\mathbf{t} = (0.4129-0.3986)/(0.4*(1-0.4)*(1/7750+1/(80,000-7750)))^{1/2}$$

$$= 2.44$$

with (80,000-2) i.e. 79,998 degrees of freedom

So since the number of degrees of freedom is greater than 120 and **t** is greater than 1.96, we can be at least 95% sure that the proportion of accidents at junctions in our rural study area is greater than the proportion at junctions on other rural roads. Therefore we would recommend that further investigations are carried out to try and explain this result (see Barker et al (1999) for a more detailed explanation of how to interpret the result).

Regression-to-the-mean correction

B.6

To correct for the regression-to-the-mean effect it is necessary to estimate the true underlying accident rate. Several statisticians have proposed ways of doing this, eg. Hauer (1992) extended the Empirical Bayes' model to estimate the true underlying accident rate and then based the evaluation on this rather than the raw data. However, an approach that is simpler to apply for a single site was described by Abbess et al (1981), in which they adjusted the data to correct for biases using assumptions about the distribution of accidents over a period of years.

Accident data must be gathered for similar sites to the treated site over the same time period: the control sites. Using this full dataset the mean number of accidents, *a*, and the variance of accidents *var* (*a*) are calculated. The regression-to-the-mean effect, R (in per cent) was shown to be given by the following formula:

$$R = \left(\frac{(A_t + A)\, n}{(n_t + n) A} - 1 \right) .100$$

where A = the number of accidents at the site over a period of n years

$$A_t = \frac{a^2}{(var\ (a) - a)}$$

$$n_t = \frac{a}{(var\ (a) - a)}$$

A_t and n_t are the estimates of the parameters of the statistical distribution showing the true underlying accident rates, ie. the probability distribution of the accident rate before any data are available. The main assumption is, therefore, that the study site with a particular accident history will behave in the same way as the set of all similar sites with the same accident history.

Example

Let us consider a junction, which has had an average of 15 accidents per year over the past 5 years. The site was widened, large new junction signing, splitter islands and STOP signs installed, after which the site has averaged 10 accidents per year over a similar period.

To correct for the regression-to-mean effect, we need to select similar uncontrolled junction sites with similar traffic flows. If all these sites have produced a mean, *a*, of 12.6 accidents per year with a variance, *var(a)*, of 2.91, then using the equation above, the input values are:

n = 5 (years)

A = 75 (accidents)

$A_t = 12.6^2 / (2.91 - 12.6) = -16.38$

$n_t = 12.6 / (2.91 - 12.6) = -1.3$

Thus the Regression effect:-

$$R = \left(\frac{(-16.38 + 75)5}{(-1.3 + 5)75} - 1 \right).100$$

$$= 5.2\%$$

That is, during the after period we would expect that if nothing were done to the site, the accidents would reduce by 5.2 per cent, or to 14.25 accidents per year. Thus it is the figure of 14.25 accidents per year that should be compared with the 10 accidents per year that actually occurred to determine whether the reduction in accident frequency due to the improvements is statistically significant.