Chapter 7 – DATA ANALYSIS

An important principle in the analysis of camouflage data is that the type of analysis to be carried out determines the way the data must be handled [11]. This principle will be explained and illustrated with practical examples. There are two different types of analysis that will be discussed in this report. In the first type of analysis, the data values are treated as sample points from a distribution and the analysis is concerned with statistical estimates of distribution parameters (Sections 7.2 and 7.3). In this case the order of the data values is not important and there is no other information associated with any given data point. In the second type of analysis, the distinguishing feature is that each data point has one or more covariates (factors) associated with it (Section 7.5). For example, if the analysis is concerned with the possible effect of observer age on detection, then age is a covariate and each detection range datum has an observer age associated with it. The time of each observation and even the order of observations are examples of possible covariates of interest.

Although detection, recognition and identification ranges are always reported as positive values, it is mathematically convenient in the analysis to treat ranges as negative values. In the subsequent discussion and examples of trial results, all ranges will be listed as negative values. Also, the subsequent discussion and examples will focus on detection results only, for brevity, but the procedures discussed apply equally to recognition and identification data.

7.1 DATA ADJUSTMENT

The recommended trial procedures described previously yield observer detection data that are interval-censored. In the experiments observers have the opportunity to make detections at pre-defined intervals, dictated by the associated ranges of the slides, digital images or viewing posts. In this circumstance, observers do not have an opportunity to make detections between the interval ranges. Thus when an observer detects a target, the detection range is biased towards a lower value. Given that the observer did not detect the target at the previous longer range, one would expect that detection would have taken place somewhere in the interval between the preceding detection opportunity and the one where detection occurred.

Similarly, if the observer did not detect the target after viewing all the images or observing from all the observing posts, one can speculate that detection would have taken place somewhere between the range of the last image/observing post and the target location (range = 0 m). Using a detection range value of zero in this circumstance biases the datum towards a lower value.

The bias associated with the raw detection results can be minimized by adjusting the detection range values. Two procedures are recommended:

- The uniform-spacing method, when there is no covariate information associated with the detection range data.
- The single-value spacing method when covariate data will be used in the analysis.

These data adjustments are particularly important when comparing detection data that has been collected using different range intervals. Adjusting the raw detection data reduces the potential for biased statistical analysis results due to imagery range-selection effects.

Table 2 presents hypothetical raw detection results for two camouflage treatments, which will be used to demonstrate the application of these two data-adjustment techniques. For demonstration purposes, the detection results are deemed to be based on detection opportunities at the ranges listed in Table 1.

Table 2: Hypothetical Raw Detection Results

Camouflage Net A			Camouflage Net B		
Observer Session	Age	Detection Range (m)	Observer Session	Age	Detection Range (m)
1	38	-1000	16	38	-900
2	40	-1100	17	40	-900
3	32	-1600	18	32	-1100
4	36	-900	19	36	-800
5	28	-900	20	28	-900
6	36	-900	21	36	-1000
7	37	-800	22	37	-1600
8	38	-1000	23	38	-700
9	38	-1000	24	38	-600
10	38	-1600	25	38	-500
11	33	-1100	26	33	-1000
12	37	-700	27	37	-900
13	25	-800	28	25	-900
14	32	-900	29	32	-500
15	37	-800	30	37	-1000

If one was solely interested in comparing the general detectability of the two camouflage nets, the detection data should be adjusted by the uniform-spacing method. To apply this method, one identifies the number of data samples applicable to each range value. Then these detections are equally spaced between the range where detection took place and the next longer range value. Consider the following example. Let R_a be the range where n detections took place and R_b is the range related to the next longer-range image. With the uniform spacing, the n observations are given range values that are equally distributed between R_a and R_b, with $(R_a – R_b) / (n + 1)$ as the spacing. In Table 2, there are three detections of camouflage net A with a range value of -1000 m. From Table 1, the next longer-range image has an associated range of -1100 m. The uniform spacing procedure would convert the three ranges to -1025, -1050 and -1075 m. These adjusted ranges cannot be associated with any particular observer. Applying the uniform-spacing procedure to the range data of Table 2 yields the results shown in Table 3. As the adjusted detection data using the uniform-spacing method is no longer associated with observers, it can be ordered as shown in Table 3, which facilitates subsequent analysis.

Table 3: Detection Ranges Adjusted for Uniform Spacing

	Camouflage Net A	Camouflage Net B
Data Value	Detection Ranges (m)	Detection Ranges (m)
1	-1733	-1700
2	-1667	-1150
3	-1167	-1075
4	-1133	-1050
5	-1075	-1025
6	-1050	-983
7	-1025	-967
8	-980	-950
9	-960	-933
10	-940	-917
11	-920	-850
12	-875	-750
13	-850	-650
14	-825	-567
15	-750	-533

The single-value-spacing adjustment is appropriate when one wishes to relate some information to the detection ranges. For example, one may be interested in investigating if there is a relationship between detection range and observer age. In this case, each detection range must be associated with an observer age. Here, the only valid adjustment method to minimize bias is to replace each detection range with the mid-point value between the raw detection range and the next longer range, i.e. $(R_a + R_b)/2$. Applying this procedure to the range data in Table 2, produces the results shown in Table 4.

Table 4: Detection Results Adjusted for Single-Value Spacing

Camouflage Net A			Camouflage Net B		
Observer Session	Age	Adjusted Detection Range (m)	Observer Session	Age	Adjusted Detection Range (m)
1	38	-1050	16	38	-950
2	40	-1150	17	40	-950
3	32	-1700	18	32	-1150
4	36	-950	19	36	-850
5	28	-950	20	28	-950
6	36	-950	21	36	-1050
7	37	-850	22	37	-1700
8	38	-1050	23	38	-750
9	38	-1050	24	38	-650
10	38	-1700	25	38	-550
11	33	-1150	26	33	-1050
12	37	-750	27	37	-950
13	25	-850	28	25	-950
14	32	-950	29	32	-550
15	37	-850	30	37	-1050

7.2 CUMULATIVE FREQUENCY DISTRIBUTION

It is often useful and informative to create a cumulative frequency distribution curve for camouflage data. This curve shows the percentage of observations that occurred at or before each range. To create a cumulative frequency distribution curve, it is appropriate to use the uniform-spacing method to adjust the raw observations as no additional (covariate) information is desired. The data from Table 3 were used to create the curves shown in Figure 6.

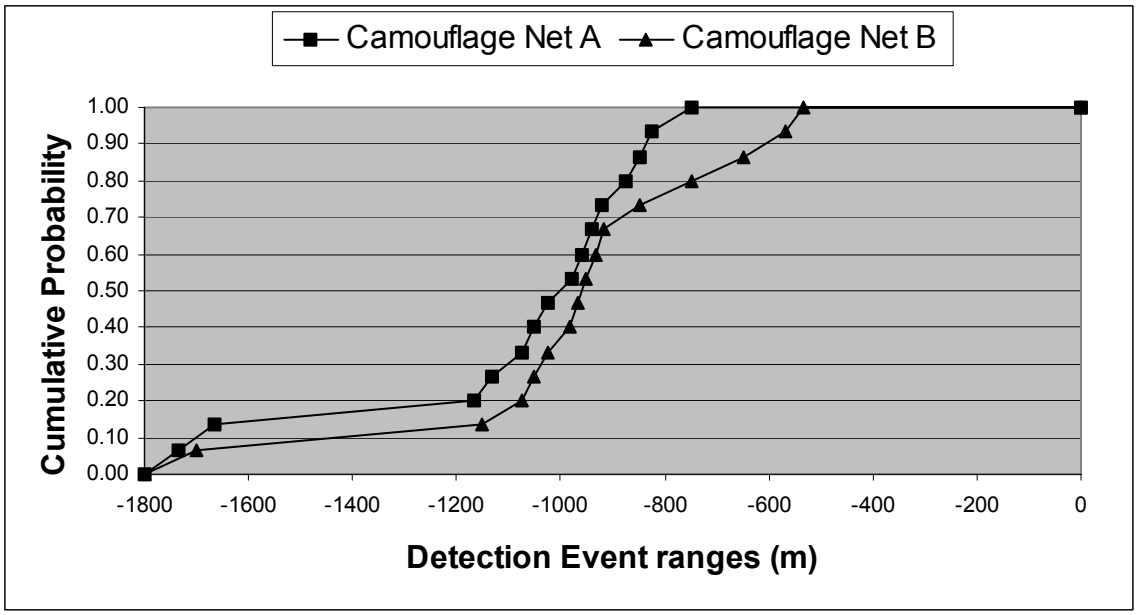

Figure 6: Cumulative Initial Detection Probability Curves.

7.3 DESCRIPTIVE STATISTICS

7.3.1 Median

A standard way to summarize a set of data using numerical methods is to calculate a central value for the data such as the mean or median. The median is preferred over the mean for camouflage data because extreme data points, such as unusually high or low initial detection ranges, could skew the mean and make the central value appear higher or lower. The median is unaffected by such outliers.

To calculate the median, first adjust the raw data observations, if necessary, using the uniform-spacing method. The median is the middle value of an ordered set of observations (i.e., the ordered set of the adjusted values). When the number of observations is odd, the median is the middle value. When the number of observations is even, the median is the average of the two middle values.

For example, the median of the 15 ordered data values in Table 3 is the eighth value, -980 m for camouflage net A and -950 m for camouflage net B.

7.3.2 Confidence Interval

The calculated median for the data sample is an estimate of the true value for the total population, which is unknown. It is understood that the true population median will likely be a value somewhat higher or lower than the estimate. The range of values about the estimated median, within which the true median is expected to lie, can be determined. This is the uncertainty or **confidence interval** associated with the estimate of the median. The confidence interval is an estimate of upper and lower limits, with an associated confidence level, within which the true statistic is likely to lie. The confidence level (a percentage, usually 95%) indicates how certain the analyst wants to be that the interval includes the true median. Higher confidence levels increase the likelihood that the true median will be in the interval, but widen the range of the interval. Smaller sample sizes also tend to increase the width of the confidence interval.

Methods to calculate the confidence interval associated with an estimate of the median can be found in standard statistical textbooks, such as References 12 and 13. Also, standard statistical analysis software, such as Analyse-it® for Microsoft Excel® can compute confidence intervals with the estimate for the median. For example, for the data in Table III, the 95% confidence interval for the median for camouflage net A is -1133 to -873 m while for camouflage net B the 95% confidence interval is -1050 to -750 m.

7.3.3 First Percentile

In some cases one may be less interested in understanding average performance and more concerned about exceptional performance. There are several possible statistics that can be used to describe the longest range at which initial detection might take place.

It is recommended that the estimated first percentile of the distribution of observations be used as a standard descriptive statistic. The first percentile is the distance at which an observer has a one percent probability of initially detecting the target. The number of observers used in typical tests is not large enough to actually encounter this threshold, but it can be derived from the test distribution.

In order to calculate the first percentile, first adjust the N raw data observations, if necessary, using the uniform spacing method. Then sort the data in decreasing order of magnitude and denote the data value with the largest magnitude as $X_{(1)}$. Next calculate the median as explained above and denote it by $X_{.50}$. The estimated first percentile, denoted by $X_{.01}$ is given by the following formula:

$$X_{.01} = X_{.50} - 2.326\left(\frac{X_{(1)} - X_{.50}}{\alpha(N)}\right)$$

The value $\alpha(N)$ may be obtained from Annex A.8. A detailed derivation of the first percentile formula is given in [11]. To illustrate the calculation of the first percentile, consider the data in Table 3. In this case of camouflage net A, $N = 15$, $X_{.50} = -980$, $X_{(1)} = -1733$ and $\alpha(15) = -1.736$. Therefore the estimated range of the first percentile for camouflage net A is -1989 m.

$$X_{.01} = -980 - 2.326\left(\frac{-1733 - (-980)}{-1.736}\right) = -1989$$

Applying the same formula to the data for camouflage net B yields a first-percentile range of -1955 m.

This formula assumes that the probability density function of the distribution of detection ranges has a peak near the median and decreases steadily towards longer ranges. This is to say that the greatest number of detections occurs near the median and the numbers of detections decrease as the magnitude of the range value get larger and larger. Practical experience with photosimulation and field observation trials has demonstrated this assumption to be reasonable and valid.

While, in general, it is recommended that assumptions about the nature of the distribution of detection ranges be avoided in the analysis of results, such an assumption is unavoidable in creating a simple, practical formula for the calculation of the first-percentile descriptive statistic.

7.4 TESTING FOR SIGNIFICANT DIFFERENCES

The analysis of camouflage effectiveness is often concerned with the evaluation of observer performance for differing camouflage treatments. Several statistical tests are used depending on the experimental design of the camouflage trial. The two main characteristics that determine which statistical test is appropriate are the number of camouflage treatments being evaluated and whether the observations are paired or independent. If each observer views more than one of the camouflage treatments being evaluated, then the data are paired. Conversely, if a different group of observers views each camouflage treatment, then the data are independent. Table 5 lists the recommended statistical procedure for the four situations that commonly occur. The mathematical derivation of these tests is explained in most statistical handbooks, such as References 12 and 13. The practical application of these tests will be discussed below.

Table 5: Statistical Tests for Different Experimental Designs

	Independent Observations	Paired Observations
2 Camouflage Treatments	Wilcoxon Two-Sample Test (Mann-Whitney Test)	Wilcoxon Two-Sample Paired Test
3 or more Treatments	Kruskal-Wallis Test	Friedman Test

Note that each of the recommended statistical tests is a non-parametric test. Nonparametric tests make no assumptions about the nature of the distribution of observations. As such, the statistical tests are robust. However, the tests may be less sensitive to small differences between medians of different data sets than statistical tests that make assumptions about the nature of the data distributions. As it is felt that camouflage assessment trials are conducted to identify significant differences between camouflage treatments, the non-parametric statistical tests achieve this goal. Parametric tests require the user to confirm that the nature of the distribution of observations fits the assumptions of the test. This requires additional tests on the data. For simplicity and robustness, non-parametric statistical test are recommended for camouflage analysis.

7.4.1 Wilcoxon Two-Sample Test

The Wilcoxon two-sample test, also referred to as the Mann-Whitney test, provides a methodology to determine if there is a statistically significant difference between the medians of two sets of independent data. The Wilcoxon test is most often used to assess differences between detection performance against two

differently camouflaged targets for two groups of observers.[1] Note that the Wilcoxon two-sample test requires that the two samples of observers be independent. This means that each observer views only one of the targets under investigation. If the same observer views both targets, the observations are said to be paired and this Wilcoxon test is not appropriate. The analysis of paired observations will be discussed below.

The output from a Wilcoxon test is a "p-value" indicating the strength of evidence that the difference between the median detection ranges for the two camouflage treatments is a random effect and the detectability of the targets is the same. For camouflage effectiveness analysis, a p-value less than or equal to 0.05 indicates a statistically significant difference. A p-value between 0.05 and 0.10 indicates some evidence of a difference.

In order to apply the Wilcoxon two-sample test, first adjust the values of the raw observations, if necessary, using the uniform spacing method. Then, a statistical software package such as the Statistical Analysis System (SAS) or Analyse-it® for Microsoft Excel® can be used to apply the test on the adjusted data values.

7.4.1.1 Example of the Application of the Wilcoxon Two-Sample Test

An example of the use of the Wilcoxon two-sample test is the comparison of observer detection performance for the two camouflage treatments shown in Table 2. To test simply for a difference in detection performance between the camouflage types, the data are first adjusted using the uniform spacing method with results shown in Table 3. Recall that the median detection range for camouflage A is -980 m, while the median detection range for camouflage B is -950 m. The Wilcoxon two-sample test is then applied to the data of Table 3. The results of this Wilcoxon test using the Analyse-it software are contained in file 'Wilcoxon _Example.xls' on this CD-Rom, on the RTO Website and in Annex A.9. The results of the Wilcoxon test produce a p-value of 0.3245, indicating that there is no statistically significant difference in the median detection ranges for camouflage types A and B. In other words, there is no evidence to suggest that the detectability of the two camouflage types is different.

7.4.2 Wilcoxon Two-Sample Paired Test

The Wilcoxon two-sample paired test is used to assess differences between detection performance against two differently camouflaged targets when the observations are paired.[2] Paired observations occur when each observer views both of the targets under investigation. As each observer views both targets, the observations are not independent and the earlier Wilcoxon two-sample test is not applicable or appropriate.

In order to apply the Wilcoxon two-sample paired test, first adjust the values of the raw observations, if necessary, using the single-value spacing method. The single-value spacing method is applied because the link between the detection ranges for each camouflage treatment and the observer the data relate to must be retained. Once the data are adjusted, a statistical software package can be used to apply the test.

7.4.2.1 Example of the Wilcoxon Two-Sample Paired Test

The data in Table 2 will be used in this example. For the purposes of this demonstration, it will be assumed that observers in Observing Sessions 1 and 16 are the same person, as is the observer in Observing Sessions 2

[1] This test can also be used to evaluate the detection performance of two independent groups of observers for the same target on two different occasions.

[2] This test can also be used to evaluate the detection performance of paired observations for the same target on two different occasions.

and 17, similarly for Observing Sessions 3 and 18, etc. In this case, the detection data for the two camouflage treatments are not independent, so the Wilcoxon two-sample paired test is the appropriate test to apply. The single-value-spacing-adjusted data of Table 4 is used in the test. The input data and results are contained in the file named 'Wilcoxon_Paired_Example.xls' on this CD-Rom, on the RTO Website and in Annex A.9.

The results using the Analyse-it software show that the p-value is 0.2412, indicating that there is no statistically significant detection difference between camouflage treatment A and B. When an experiment uses the same observers to obtain detection data for different targets, one expects to see less variability in the results than for different observers. So, any variability observed in data obtained in these circumstances is a strong indicator of differences. This is why the p-value is much lower using this paired test compared to the standard two-sample test, even though the same input data was used in both tests. This indicates the importance of understanding the nature of the data and applying the appropriate statistical test.

7.4.3 Kruskal-Wallis Test

The Kruskal-Wallis test is a non-parametric test that is used to determine if there is a statistically significant difference between three or more **independent** groups of sampled data, regardless of their distribution. When there are only two groups in the comparison, the test becomes a Wilcoxon two-sample test. To apply the Kruskal-Wallis test, first adjust the values of the raw observations, if necessary, using the uniform-spacing method. Then, a statistical package such as Analyse-it can be used. If the test returns a p-value that is less than or equal to 0.05, then two or more groups of the sampled data have medians that are significantly different. In this case, further analysis is required in order to determine how to partition the datasets into groups with equal medians. A "maximum likelihood" method, which will be described subsequently, is recommended to accomplish this further analysis.

7.4.3.1 Example of the Kruskal-Wallis Test

For this example we will assume that another observer trial with a third camouflage treatment (Camouflage Net C) was conducted. The observer data and detection results are shown in Table 6. Table 6 also contains adjusted detection results according to the single-value-spacing method, to preserve the association with the observer and the observer age. The detection results for camouflage treatment C were also adjusted using the uniform-spacing method to allow a comparison of these results with those for camouflage treatments A and B to be made. The detection results adjusted by the uniform-spacing method are shown in Table 7, sorted by detection range.

Table 6: Detection Results for Camouflage Treatment C

Camouflage Net C			
Observer Session	Age	Raw Detection Range (m)	Adjusted Single-Spacing Detection Range (m)
31	38	-1000	-1050
32	40	-700	-750
33	32	-900	-950
34	36	-1000	-1050
35	28	-1200	-1300
36	36	-900	-950
37	37	-900	-950
38	38	-800	-850
39	38	-1200	-1300
40	38	-1800	-1900
41	33	-800	-850
42	37	-800	-850
43	25	-900	-950
44	32	-700	-750
45	37	-800	-850

Table 7: Detection Results Adjusted by Uniform Spacing

Camouflage Net C	
Data Value	Uniform-Spacing Detection Range (m)
1	-1900
2	-1267
3	-1233
4	-1067
5	-1033
6	-980
7	-960
8	-940
9	-920
10	-880
11	-860
12	-840
13	-820
14	-767
15	-733

From Table 7, it can be seen that the median detection range is -940 m. Applying the procedure to calculate the first-percentile range returns a value of -2226 m for camouflage C.

If one wished to test for differences in detectability between the three camouflage treatments (A, B, and C), the Kruskal-Wallis test should be used as these are independent sets of data and involve more than two cases. Again, a statistical software package can be used to conduct the test. The file named 'Kruskal-Wallis_

Example.xls' on this CD-Rom, on the RTO Website and in Annex A.9 contains the results of applying this test to the detection data (adjusted by the uniform-spacing method) for camouflage treatments A, B, and C. The Kruskal-Wallis test returns a p-value of 0.5588, indicating that there is no statistically significant difference between the median detection ranges of the three camouflage treatments.

7.4.4 Friedman Test

The Friedman Test is a non-parametric test that is used to determine if there is a statistically significant difference between three or more groups of paired data, regardless of their distribution. When there are only two groups in the comparison, the Wilcoxon two-sample paired test should be used. To apply the Friedman Test, first adjust the values of the raw observations, if necessary, using the single-value spacing method. Then, the test can be applied using a statistical package such as SAS or Analyse-it. If the test returns a p-value that is less than or equal to 0.05, then two or more groups of the sampled data have medians that are significantly different. In this case, further analysis is required in order to determine how to partition the datasets into groups with equal medians. If there had been a difference, the different dataset groupings could have been identified by using the maximum likelihood method.

7.4.4.1 Example of the Friedman Test

This example will use the data presented in Tables 4 and 6. It is assumed that the same observers viewed each of the three camouflage treatments (A, B and C). It is assumed that Observer 1 participated in Observing Sessions 1, 16 and 31; Observer 2 took part in Observing Sessions 2, 17 and 32, and so on. Then, the detection data generated from the 45 observing sessions are related to only 15 observers. As the detection range data are not independent, it is important to retain the linkage between the data and the observer that generated it. Hence, the data adjusted by the single-value spacing method will be used for the statistical analysis.

Applying the Friedman test on the adjusted data of Tables 4 and 6 produces a p-value of 0.8040, indicating that there is no statistically significant difference between the detection results of the three camouflage treatments. The input data and results of the Friedman test using the Analyse-it software are contained in the file 'Friedman_Example.xls', on this CD-Rom, on the RTO Website and in Annex A.9. If there had been a difference, the different groups could have been identified by doing pair-wise comparisons using the Wilcoxon two-sample paired test.

7.4.5 Maximum Likelihood Method

The maximum likelihood method is applied when the Kruskal-Wallis or Friedman tests identify that there is a difference among the median values associated with three or more datasets. In such a circumstance, further analysis is required to determine which datasets have statistically similar medians and which are different. The maximum likelihood method utilizes the previously described statistical tests to determine the "best" way to combine the datasets into groups of statistically similar median detection ranges. The maximum likelihood method has the desirable attribute that it will always separate the datasets into at least two groups, consistent with the results of the Kruskal-Wallis or Friedman tests.

To apply the method, one begins by ordering the datasets according to the median detection range values. Potential separation points are identified between adjacent data sets, as shown in Figure 7. Each separation point divides the datasets into two data groups formed by combining all the detection range data from all the datasets on the same side of the separation point. Note that the original detection range results must be used to produce the combined dataset and the detection ranges must then be re-adjusted using the uniform spacing

method. The two data groups are compared using the Wilcoxon two-sample test for differences in their median detection ranges. The p-value from the Wilcoxon test indicates the strength of evidence that the datasets should be divided at the separation point being examined. The p-value for each potential separation point is calculated and compared. The lowest p-value indicates the strongest evidence that the associated separation point should be chosen to split the datasets. After splitting the datasets into two groups based on the most likely separation point, there may be more than one dataset in either group. When this occurs the datasets in the group(s) are tested for differences using the appropriate statistical test. If the test indicates that a difference exists, then the datasets within the group must be separated using either the Wilcoxon two-sample test when there are two datasets in the group or the maximum likelihood method if there are more than two datasets in the group. This process is repeated until there are no differences indicated within any of the defined data groups. It is possible that the process could separate every dataset into a different group, in which case every dataset is different from every other dataset.

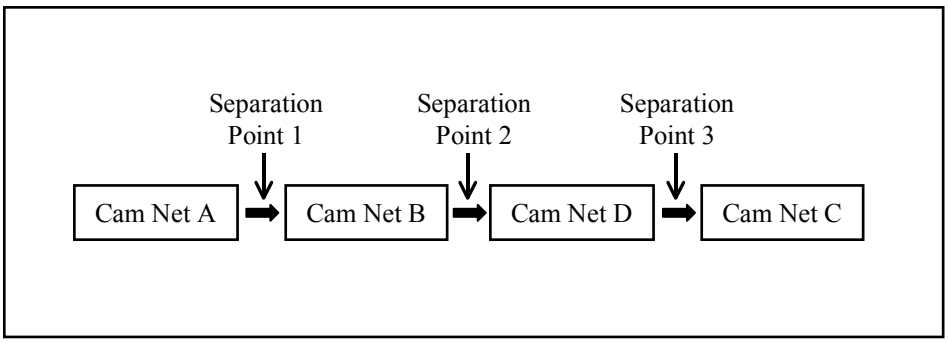

Figure 7: Possible Separation Points.

7.4.5.1 Example of the Maximum Likelihood Method

The maximum likelihood method will be demonstrated using the detection range data shown in Table 8. The detection ranges in Table 8 are fictitious and associated with four hypothetical camouflage nets for vehicles. The ranges are raw detection ranges derived from imagery with the associated detection ranges shown in Table 1. It is also assumed that different observers were used to collect the data for each net, so the data are independent. Results of the applied statistical tests will be reported, but the calculations will be left to the reader to perform.

Table 8: Detection Range Datasets

Detection Range (m)	Cam Net A	Cam Net B	Cam Net C	Cam Net D
Observer 1	-1400	-1400	-800	-1200
Observer 2	-1200	-1400	-800	-1200
Observer 3	-1100	-1200	-800	-900
Observer 4	-1100	-1000	-800	-800
Observer 5	-1100	-1000	-700	-800
Observer 6	-1000	-1000	-600	-800
Observer 7	-1000	-1000	-500	-700
Observer 8	-1000	-1000	-400	-700
Observer 9	-1000	-900	-400	-700
Observer 10	-1000	-900	-400	-700
Observer 11	-900	-900	-400	-700
Observer 12	-900	-900	-300	-600
Observer 13	-900	-900	-300	-500
Observer 14	-900	-900	-300	-400
Observer 15	-900	-900	-300	-400
Median Detection Range	-1050	-1017	-480	-767

In Table 8, the median detection range for each dataset is shown at the bottom and is based on the detection range data adjusted by the uniform-spacing method. The first step in the process is to apply the Kruskal-Wallis test to determine if there is a difference among any of the datasets in the group. Using the adjusted detection range data, this test produces a p-value less than 0.0001, indicating that there is at least one difference within the group, i.e. at least one dataset (or group of datasets) that is different from the others. To locate the difference(s), the maximum likelihood method is next applied.

With the maximum likelihood method, one begins by ordering the datasets according to their associated median detection range. Going from longest median detection range to the shortest range, the sample data yields an order of net A, followed by net B, followed by net D and lastly by net C. As shown in Figure 7, there are three possible separation points: between the datasets for net A and net B, between the datasets for net B and net D and between the datasets for net D and net C. Each separation point divides the datasets into two groups. Within each group, all the detection range data of the datasets are combined into one large dataset and adjusted according to the uniform-spacing method. For separation point one, the dataset for net A is left in one group, while the data for nets B, D and C are combined. For separation point two, the data for nets A and B are combines in one group and the data for nets D and C are combined in the other group. Lastly, for separation point three, the data for nets A, B and D are combined in one group, while the net C data forms the second group.

The Wilcoxon two-sample test is used to calculate the p-value for each of the pairs of data groups formed by the separation points. The following results are obtained:

Separation Point 1 : p-value = 0.0027

Separation Point 2 : p-value = <0.0001

Separation Point 3 : p-value = <0.0001

Separation points two and three have the lowest p-values, which are equal in value. This indicates that both these separation points are equally good choices as the most likely separation points to use. In this case, the choice between separation point two and three is arbitrary. Separation point two is chosen.

The datasets have now been divided into two groups (net A + net B and net D + net C) that are known to have different median detection ranges. The next step is to test each of these data groups to determine if there are any differences between the datasets within the groups. As there are only two datasets in each group, the Wilcoxon two-sample test is applied with the following results:

Cam Net A versus Cam Net B : p-value = 0.5668
Cam Net D versus Cam Net C : p-value = 0.0264

The p-values obtained indicate that there is no statistically significant difference between cam nets in the first group (A and B), but there is a difference between cam nets in the second group (D and C). With this result, one would now apply the maximum likelihood method to determine the best separation point for the datasets in the second group. However, as there are only two datasets in the group, net D and net C, there is only one possible separation point, so the maximum likelihood method is not required. The median detection ranges for camouflage nets D and C have been shown to be different.

The overall results for the analysis of the detection range data for the four camouflage nets indicates that there is no statistically significant difference in the median detection ranges of net A and net B, net C and Net D are different from nets A and B and from each other.

7.5 INFLUENCE OF PARAMETERS ON DETECTABILITY

In some circumstances, the factors that may affect detectability of a camouflage treatment are of interest. For example, having collected information on the observers participating in the trial, one can investigate the relationship between observer age and experience on initial detection range of the target. One begins by establishing categories or ratings for the factors of interest and grouping the detection data by the associated factors' categories. To maintain the proper association between the covariate factors and the detection data, the detection data should be adjusted only by the single-spacing method. Having established the factor categories and having adjusted the associated detection data, one can apply the afore-described statistical tests to determine if the detection results are affected by the factors.

One concern to keep in mind when performing this type of analysis is the sample size. Each factor category under investigation can be considered a separate distribution or population. The data related to each factor category is a sample of the population. As was discussed earlier, ideally one should have at least 15 data values for each population being compared to provide reasonable confidence that true differences will be statistically discernable and good estimates for descriptive statistics can be obtained. If one collects detection data for a camouflage treatment using 15 observers, this will be adequate for comparing the detectability of

this camouflage treat to others. However, if one then wishes to investigate if the detectability of the camouflage treatment is a function of the age of the observer, the sample size will be lower than the recommended level. If the observers are divided into two age groups (i.e. young and mature) then there will be less that eight data samples in one of the groups. At least one group will have a sample size much lower than the recommended level. If it is anticipated that this type of analysis will be conducted, the number of observers in the initial data should be increased to ensure that adequate sample sizes will be available for all components of the analysis.

7.5.1 Example of Factor Analysis

To demonstrate this type of analysis the data shown in Table 8 will be used. For conciseness only 15 data samples will be used. This sample size will be sufficient to demonstrate the technique, but would be much too small to be used in practice.

During the hypothetical detection trial for camouflage X, each observer is required to complete a questionnaire indicating their age and previous experience at target detection. The results are shown in Table 9. Examination of the ages of the observers indicates that the median age of the group was 37. This value was used to divide the observers into two age groups, Young for observers under 37 years of age and Mature for observers 37 years of age and older. The data on experience level naturally divides into two categories. With these factor categories, one can now test to determine if detection range is related to observer age or experience. To statistically test for the effect of observer age on detection range, begin by separating the detection range data into a group associated with Young observers and another related to Mature observers. Analyzing the detection data based on age category indicates that the median detection range for Young observers is -800 m and for Mature observers, it is -1050 m. The Wilcoxon (Mann-Whitney) two-sample test can be used to determine if the perceived difference in median detection range between the age groups is statistically significant. The Wilcoxon test is the appropriate statistical test as all the data are independent. The resulting p-value from the Wilcoxon test is 0.6943, indicating there is no statistically significant difference in the detection performance between the observer age groups. The input data and results for this analysis (and the analyses that will be discussed below) using the Analyse-it software are contained in the file ‘Age_Experience_Tests.xls’ on this CD-Rom, on the RTO Website and in Annex A.9.

The same statistical procedure can be used to test for the effect of observer experience on detection performance. Again, the detection data are separated in two groups based on the categories of experience, Yes and No. The two sets of detection range data are then compared using the Wilcoxon test. Here, the median detection range for observers with no experience is -650 m, while for experienced observers the median detection range is -1200 m. The p-value from the Wilcoxon test is 0.0120, indicating that there is a statistically significant difference in the detection results for the two categories of observer experience.

Table 9: Detection Results for Camouflage Treatment X

Observer Session	Age	Age Category	Experienced	Adjusted Detection Range (m)
1	38	Mature	Yes	-1050
2	40	Mature	Yes	-1150
3	32	Young	No	-1700
4	36	Young	No	-950
5	28	Young	No	-650
6	36	Young	No	-450
7	37	Mature	No	-850
8	38	Mature	Yes	-1250
9	38	Young	Yes	-1350
10	38	Mature	Yes	-1700
11	33	Young	Yes	-1150
12	37	Mature	No	-750
13	25	Young	No	-550
14	32	Young	No	-650
15	37	Mature	No	-350

Lastly, one can also check if there is an effect on target detection for the interaction between age and experience. To do this a set of groupings must be created for the combinations of age and experience categories. Here there are two age categories, Young and Mature, and there are two categories of experience, Yes and No. This produces four combinations of age and experience: Young-No, Young-Yes, Mature-No and Mature-Yes. Having established the applicable interaction categories, the detection results are grouped under the appropriate categories. For the detection range data being analyzed here, the median detection ranges for the four Age-Experience categories are: -650 m for the Young-No category, -1250 m for Young-Yes, -750 m for Mature-No and -1200 m for the category of Mature-Yes. As there are more than two groups of independent data, the Kruskal-Wallis test is applied. The results of the Kruskal-Wallis test indicate a p-value of 0.0835, revealing that there is no statistically significant difference in the detection results for the combined age-experience factor. However, as the p-value is between 0.05 and 0.10, this suggests that there is some evidence for an age-experience effect.