# 13

# Cluster Analysis

**What is Cluster Analysis?**

Cluster analysis is used to classify observations into a finite and small number of groups based upon two or more variables (Finch, 2005). The term cluster analysis was first used in 1939 by Tryon (Tryon,1939). 'Numerical taxonomy' is another term used for cluster analysis in some areas of biology (Romesburg, 2004). There is no *a priori* hypothesis in cluster analysis, unlike other statistical analysis. In cluster analysis the variables are arranged in a natural system of groups (Kirkwood, 1989). The heterogeneous data collected are sorted into series of sets. Data in a cluster are considered to be 'similar' or highly correlated to each other. Clusters can be exclusive (a particular variable is included in only one cluster) and overlapping (a particular variable is included in more than one cluster). Cluster analysis method is used in a variety of research problems (Hartigan, 1975; Scoltock, 1982; Moore *et al.*, 2010). It is applied extensively in the fields of toxicogenomics (Hamadeh *et al*., 2002), genetics (Shannon *et al*., 2003; Makretsov *et al.*, 2004) and molecular biology (Furlan *et al.*, 2011). Cluster analysis only discovers structures in data, but does not explain why such structures exist.

Cluster analysis can be carried out using several methods. Three commonly used methods are described below:

*Hierarchical cluster analysis*

As the name indicates, hierarchical cluster analysis produces a hierarchy of clusters. The clusters thus produced are graphically presented. This graphical output is known as a dendrogram (from Greek *dendron* 'tree', *gramma* 'drawing'). The dendrogram can be used to examine how clusters

are formed in hierarchical cluster analysis (Schonlau, 2002). Hierarchical clustering can be of two types. One type is agglomerative clustering, where grouping of clusters is done small clusters to large ones. The other type is divisive clustering, where grouping of clusters is done large clusters to small ones. For illustrative purpose a dendrogram is given in Figure 13.1.
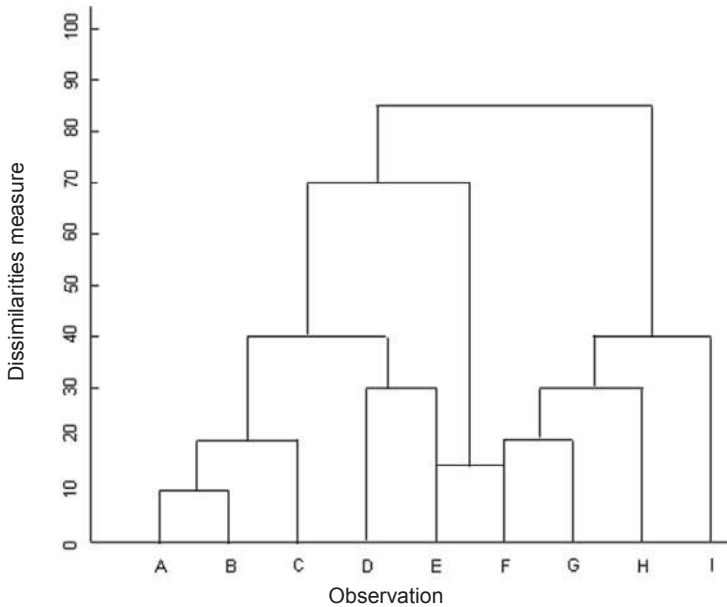


**Figure 13.1.** Dendrogram

The individual observations (A–I) are arranged evenly along the X axis of the dendrogram. They are called as leaf nodes. The vertical axis indicates a distance or dissimilarity measure. The height of a leaf node represents the distance of the two clusters that the node joins. In this dendrogram, the similarity of samples A and B is better than the other samples, and the first cluster is formed by these two samples.

***Ward's method of cluster analysis*** (Ward, 1963; Ward and Hook, 1963)

This method is more efficient than hierarchical cluster analysis. Ward's method uses the squared distances between-clusters and within-clusters (Rencher, 2002). Hence, Ward's method is also called as the 'incremental sum of squares' method.

### k-means cluster analysis

This method of clustering is used when *a priori* hypothesis concerning the number of clusters in variables are available. *k* is the number of clusters that we desire.

Data collected in repeated dose administration toxicity studies is enormous and are either qualitative or quantitative in nature. No observed adverse effect level (NOAEL) of the test substance is judged based on these data. Sometimes the toxicity effects manifested are not dose-dependent, which makes judging an NOAEL difficult. In such situations, cluster analysis is extremely useful for judging an NOAEL. Now the question is whether to consider only those data which show a significant difference compared to control for the cluster analysis or all data collected in the study, irrespective of their difference from the control is significant or not.

We shall try to understand cluster analysis with the help of an example. Groups (10/sex/dose) of seven-week-old Crj: CD rats were administered the test substance at low, mid, high and top doses by gastric intubation daily for 28 days. A concurrent control group was also maintained. Rats were daily examined for general behavior. During the dosing period, body weight, food and water consumption of the animals were measured. Animals were sacrificed on day 29 after overnight starvation for assessment of hematology, blood biochemistry, serum protein electrophoresis, urinalysis, myelogram and ophthalmologic and pathological (organ weight measurement and gross and histopathology) examinations (Kobayashi, 2004).

Salivation in both sexes in the high dose group, staggering gait in the top dose group, slight suppression of the body weight gain in males in the top dose group, slight anemic trend in both sexes in the top dose group, higher values in alkaline phosphatase in both sexes in the high dose and top dose groups, lower values in albumin in males in the top dose group and in females in the high dose and top dose groups, bone fractures, mobilization of the sinusoidal cell and extramedullary hematopoiesis in the liver in both sexes in the top dose group and squamous hyperplasia, and erosion of the fore-stomach in both sexes in the high and top dose groups were observed as the main changes attributable to the repeated oral administration of the test substance. Based on above observations and determinations, the NOAEL was considered to be the mid dose for both males and females.

The data obtained in the study was analyzed statistically. Continuous data was subjected to Bartlett's test for examining homogeneity of variance and was analysed (two-sided analysis) using the statistical techniques as

given in the decision tree proposed by Kobayashi *et al*. (2000) (Figure 13.2). Gross and histopathological findings were analyzed by the Fisher's exact test (Gad and Weil, 1986). The level of significance for the above mentioned statistical analysis was set at P<0.05.
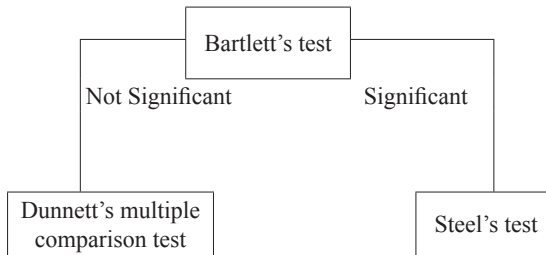


**Figure 13.2.** Analytical methods by a decision tree

We shall analyse the data of the study described above using Ward's method of cluster analysis (Milligan 1980). The software used for the analysis was JMP (version 5) of the SAS (SAS Institute, Japan).

## *Cluster-1*

The items in the dosed groups that showed a significant difference compared to the control group were—body weight gain, food efficiency, hematocrit, hemoglobin, red blood cell count, platelet count, neutrophil (%), lymphocytes (%), blood urea nitrogen, total protein, alanine aminotranferase, alkaline phosphatase, glucose, prothrombin time, albumin, albumin/globulin ratio, inorganic phosphorus in urine, lung weight, relative weights of the lung, liver, kidneys and testes, gross pathology findings, and microscopic findings. These items were grouped in Cluster 1.

Each dosed group was divided into Group 1 and Group 2. Group 1 was further divided into Subgroup 1 and Subgroup 2 (Table 13.1).

**Table 13.1.** Results of cluster analysis: Cluster 1—Items showing a significant difference (P<0.05) compared to control

| Dose Group | Number of animals | | |
|---|---|---|---|
| | Group 1 | | Group 2 |
| | Subgroup-1 | Subgroup-2 | |
| Control | 10 | 0 | 0 |
| Low | 10 | 0 | 0 |
| Mid | 10 | 0 | 0 |
| High | 2 | 8 | 0 |
| Top | 0 | 4 | 6 |

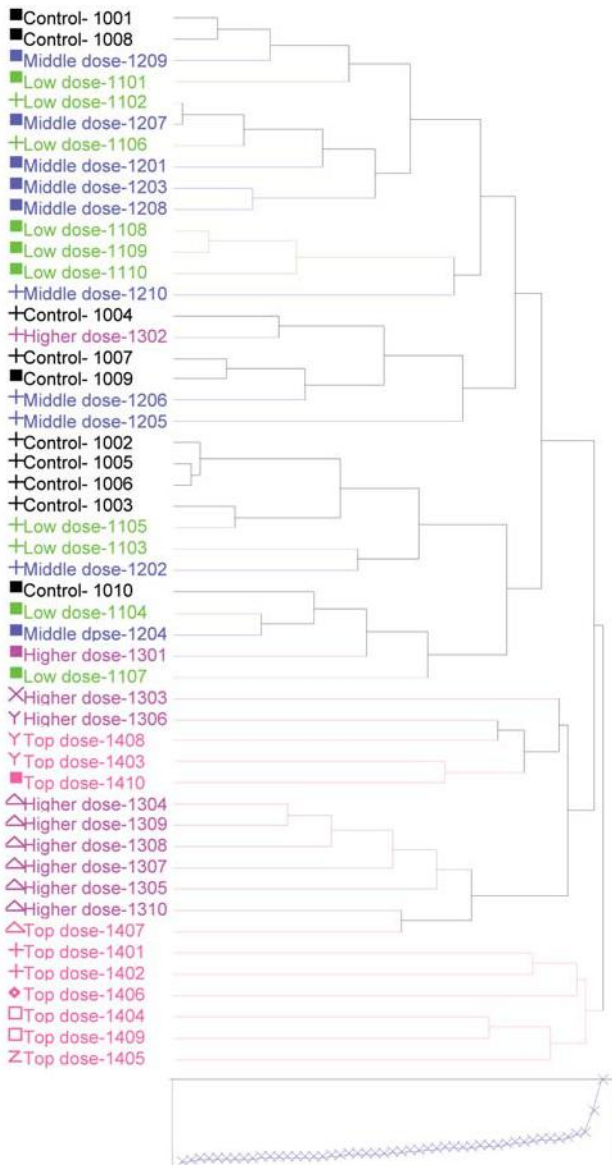The dendrogram obtained from the above data is given in Figure 13.3.



**Figure 13.3.** Dendrogram of items that are significantly different from control (Ward's method)

Note: Animal identification mark, dose group and animal number are given on the left side of the dendrogram.

## Cluster 2

The items which did not show a significant difference compared to control were—food and water consumption, leucocyte count, lymphocyte count, reticulocyte count, activated partial thromboplastin time, total cholesterol, free cholesterol, triglyceride, phospholipid, non esterified fatty acid, creatinine, total bilirubin, sodium, potassium, chloride, calcium, inorganic phosphorus, alanine aminotransferase, lactate dehydrogenase, alpha-1 (%), gamma (%), urine volume, urine specific gravity, and sodium, potassium, chloride, calcium and inorganic phosphorus in urine, and weights of the brain, heart, liver, kidneys, spleen, adrenals, testes, thyroid and thymus, and relative weights of the brain, heart, spleen, adrenals, thyroid and thymus. These items were grouped in Cluster 2.

Each dosed group was divided into Group 1 and Group 2. Groups 1 and 2 were further divided into two Subgroups each (Table 13.2).

**Table 13.2.** Results of cluster analysis: Cluster 2—Items showing no significant difference (P>0.05) compared to control

| Dose group | Number of animal | | | |
|---|---|---|---|---|
| | Group 1 | | Group 2 | |
| | Subgroup-1 | Subgroup-2 | Subgroup-1 | Subgroup-2 |
| Control | 8 | 2 | 0 | 0 |
| Low | 6 | 4 | 0 | 0 |
| Mid | 7 | 3 | 0 | 0 |
| High | 5 | 0 | 5 | 0 |
| Top | 0 | 0 | 5 | 5 |

The dendrogram obtained from the above data is given in Figure 13.4.

As you would have observed from the dendrograms, when the number of observations are more, it is very difficult to distinguish each observation. Dendrograms are only suitable for hierarchical cluster analysis. Schonlau (2002) proposed a clustergram, which is suitable for non-hierarchical cluster analysis. For hierarchical cluster analysis, a radial clustergram was proposed by Agrafiotis *et al*. (2007). In radial clustergram, clusters are arranged into a series of layers, each representing a different level of the tree. However, for small set of data, a dendrogram is still preferable to a clustergram.
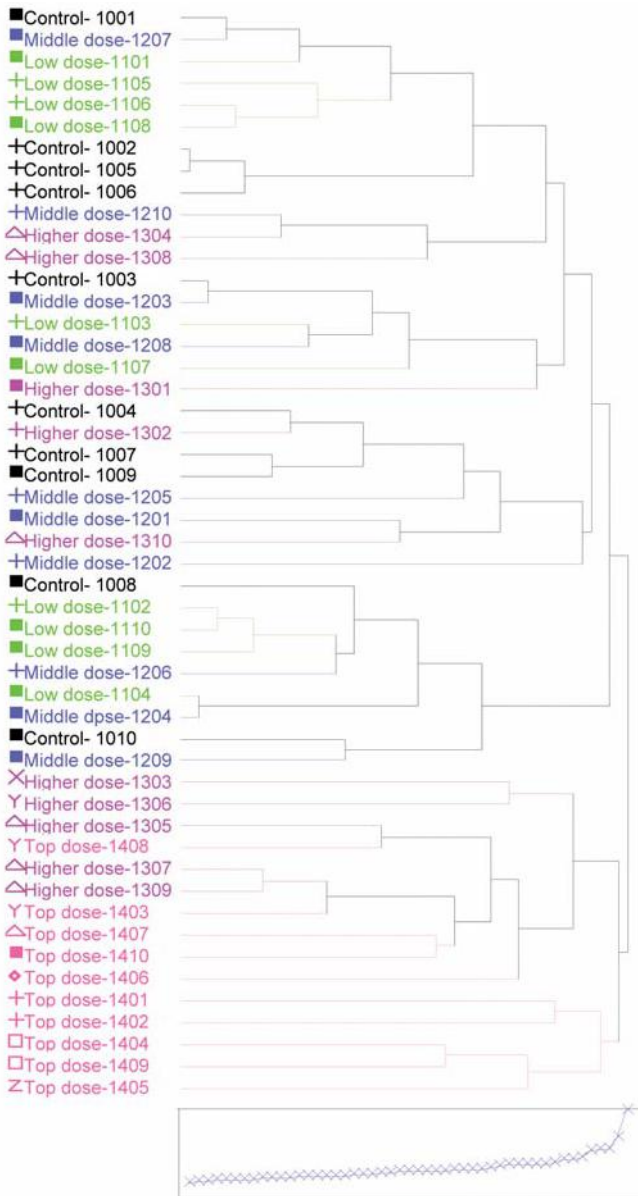
**Figure 13.4.** Dendrogram of items that are not significantly different from control (Ward's method)

Note: Animal identification mark, dose group and animal number are given on the left side of the dendrogram.

# References

Agrafiotis, D.K., Bandyopadhyay, D. and Farnum, M. (2007): Radial clustergrams: visualizing the aggregate properties of hierarchical clusters. J. Chem. Inf. Model, 47, 69–75.

Finch, H. (2005): Comparison of distance measures in cluster analysis with dichotomous data. J. Data Sci., 3, 85–100.

Furlan, D., Carnevali, I.W., Bernasconi, B., Sahnane, N., Milani, K., Cerutti, R., Bertolini, V., Chiaravalli, A.M., Bertoni, F., Kwee, I., Pastorino, R. and Carlo, C. (2011): Hierarchical clustering analysis of pathologic and molecular data identifies prognostically and biologically distinct groups of colorectal carcinomas. Modern Path., 24, 126–137.

Gad, S. and Weil, C.S. (1986): Statistics and Experimental Design for Toxicologists, The Telford Press, New Jersey, USA.

Hamadeh, H.K., Bushel, P.R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, C., Paules, R.S., Blanchard, K. and Afshari, C.A. (2002): Prediction of compound signature using high density gene expression profiling. Toxicol. Sci., 67, 232–240.

Hartigan, J.A. (1975): Clustering Algorithms. John Wiley & Sons, Inc., New York, USA.

Kikwood, B. (1989): Medical Statistics, Blackwell Scientific Publications, London, UK.

Kobayashi, K. (2004): Evaluation of toxicity dose levels by cluster analysis. J. Toxicol. Sci., 29(2), 125–129.

Kobayashi, K., Kanamori, M., Ohori, K. and Takeuchi, H. (2000): A new decision tree method for statistical analysis of quantitative data obtained in toxicity studies on rodent. San Ei Shi, 42, 125–129.

Makretsov, N.A., Huntsman, D.G., Nielsen, T.O., Yorida, E., Peacock, M., Cheang, M.C.U., Dunn, S.E., Hayes, M., van de Rijn, M., Bajdik, C. and Gilks, C.B. (2004): Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma. Clin. Cancer Res., 10, 6143–6151.

Milligan, G.W. (1980): An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 45, 325–342.

Moore, C.W., Meyers, D.A., Wenzel, S.E., Teague, G.W., Li, H., Li, X., D'Agostino, Jr., R., Castro, M., Curran-Everett, D., Fitzpatrick, A.M., Gaston, B., Jarjour, N.N., Sorkness, R., Calhoun, W.J., Chung, K.F., Comhair, S.A.A., Dweik, R.A., Israel, E., Peters, S.P., Busse, W.W., Erzurum, S.C. and Bleecker, E.R. (2010): Identification of asthma phenotypes using cluster analysis in the severe asthma research program. Am. J. Resp. Crit. Care Med., 181, 315–323.

Rencher, A.C. (2002): Methods of Multivariate Analysis. 2nd Edition, Wiley-Interscience, New York, USA.

Romesburg, H. (2004): Cluster Analysis for Researchers. Lulu Press, North Carolina, USA.

Schonlau, M. (2002): The Clustergram: A graph for visualizing hierarchical and non-hierarchical cluster analyses. The Stata J., 3, 316–327.

Scoltock, J. (1982): A survey of the literature of cluster analysis. Computer J., 25(1), 130–134.

Shannon, W., Culverhouse, R. and Duncan, J. (2003): Analyzing microarray data using cluster analysis. Pharmacogenomics, 4(1), 41–52.

Tryon, R.C. (1939): Cluster Analysis. Edward Brothers, Ann. Arbor., MI, USA.

Ward, J.H., Jr. (1963): Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc., 58(301), 235–244.

Ward, J.H., Jr. and Hook, M.E. (1963): Application of an hierarchical grouping procedure to a problem of grouping profiles. Edu. Psych. Measurement, 23, 69–81.

# Trend Tests

## Introduction

In pharmacology and toxicology experiments three or more than three treatment groups are usually used. One of the objectives for carrying out the experiment with three or more than three groups is to assess the dose-dependency of the test substance. Dose-dependency is an important concept for evaluating toxicological data (Hamada *et al*., 1997). In order to examine whether the change in a parameter observed in a study is dose-dependent, a trend test is used. A trend test examines whether the results in all dose groups together increase as the dose increases (EPA, 2005). Trend tests have been recommended as a customary method for analyzing data from subchronic and chronic animal studies (Selwyn, 1995). For examining quantitative data, Jonckheere's trend test (Jonckheere, 1954) is generally used. The frequency data are examined by Cochran-Armitage trend test (Cochran, 1954; Armitage, 1955).

## *Jonckheere's trend test*

Jonckheere's test is a frequently used nonparametric trend test for the evaluation of preclinical studies and clinical dose-finding trials (Neuhäuser *et al*., 1999). Predicted trend can be evaluated using this test (Cohen and Holliday, 2001). Since it does not require specification of a covariate, it has generated a continued interest (Jones, 2001). Jonckheere's test is based on the idea of taking a score in a particular condition and counting how many scores in subsequent conditions are smaller than that score (Field, 2004). In order to use the Jonckheere's test, the number of groups should be 3 or more than 3 and each group should have equal number of observations.

Water consumption of B6C3F1 mice fed on a diet containing a test substance at week eight is given in Table 14.1. There are three dose groups and one control group. Let us examine whether there is a trend in the water consumption across the groups.

**Table 14.1.** Water consumption (g/week) of B6C3F1 mice fed on a diet containing a test substance at week eight

| Group | Group 1 (Control) | Group 2 (Low dose) | Group 3 (Mid dose) | Group 4 (High dose) |
|---|---|---|---|---|
| | 40.6 | 31.9 | 32.7 | 30.6 |
| | 38.0 | 36.8 | 31.3 | 35.9 |
| | 41.1 | 32.4 | 32.9 | 29.6 |
| | 52.7 | 34.8 | 31.9 | 29.2 |
| | 48.8 | 43.1 | 28.5 | 28.5 |
| | 41.1 | 39.0 | 31.2 | 30.8 |
| | 39.9 | 33.6 | 33.1 | 30.5 |
| | 43.1 | 34.3 | 34.1 | 29.4 |
| | 32.7 | 34.0 | 31.2 | 30.8 |
| | 30.1 | 33.8 | 31.7 | 32.0 |
| Mean | 40.8 | 35.4 | 31.9 | 30.7 |
| SD | 6.7 | 3.4 | 1.5 | 2.1 |
| N | 10 | 10 | 10 | 10 |

Formula:

$$J = \frac{\left[\sum ijTij + \dfrac{\sum ijSij}{2} - \dfrac{N^2 \sum iN_1^{\,2}}{4}\right] - 0.5}{\sqrt{V}}$$

$$V = \frac{N(N-1)(2N+5) - \sum iN_1(N_i-1)(2Ni+5)}{72}$$

$$+ \frac{\left\{\sum iNi(ni-1)(Ni-2)\right\}\left\{\sum i\tau(\tau i-1)(\tau i-2)\right\}}{36N(N-1)(N-2)}$$

$$+ \frac{\left\{\sum iNi(Ni-1)\right\}\left\{\sum i\tau i(\tau i-1)\right\}}{8N(N-1)}$$

If the computed $J$ value is greater than the $Z$ value given in the standard normal distribution Table, it is considered to be significantly different.

Calculation of $T$ values:

We need this information for the calculation of $J$. Arrange the data in each group in the order of prediction. Let us calculate $T_{12}$ (Control Group *vs* Low Dose Group). For each control value the number of values that are lesser than it in the low dose group are counted, and their total is calculated:

$$T_{12} = 9+ 8+ 9+10+10+9+ 9+ 9+ 2+ 0 = 75$$

The first value of the control group is 40.6 and there are 9 values of the low dose group, which are lesser than 40.6. The second value of the control group is 38.0 and there are 8 values of the low dose group, which are lesser than 38.0, and so on.

Similarly, values are counted for other trends.

$$T_{13} = 10+10+10+10+10+10+10+10+ 6+ 1 \quad = 87$$
$$T_{14} = 10+10+10+10+10+10+10+10+ 9+ 4 \quad = 93$$
$$T_{23} = 5+10+ 6+10+10+10+ 9+10+ 9+ 9 \quad = 88$$
$$T_{24} = 8+10+ 9+ 9+10+10+ 9+ 9+ 9+ 9 \quad = 92$$
$$T_{34} = 9+ 8+ 9+ 8+ 0+ 8+ 9+ 9+ 8+ 8 \quad = 76$$

where, $T_{13}$ is Control Group *vs* Mid Dose Group, $T_{14}$ is Control Group *vs* High Dose Group, $T_{23}$ is Low Dose Group *vs* Mid Dose Group, $T_{24}$ is Low Dose Group *vs* High Dose Group and $T_{34}$ is Mid Dose Group *vs* High Dose Group.

τ Values: We also need to know how many times a value repeated within a group and across the groups. 43.1 is repeated twice-one each in Groups 1 and 2 ($\tau_1$), 41.1is repeated twice within the Group 1 ($\tau_2$), 32.7 is repeated twice-one each in Groups 1 and 3 ($\tau_3$), 31.9 is repeated twice-one each in Groups 2 and 3 ($\tau_4$), 30.8 is repeated twice within Group 4 ($\tau_5$), 31.2 is repeated twice within Group 3 ($\tau_6$)and 28.5 is repeated twice-one each in Groups 3 and 4 ($\tau_7$).

$$V = \frac{40(40-1)(2 \times 40 + 5) - 10(10-1)(20+5) \times 4}{72}$$

$$+ \frac{4 \times 10(10-1)(10-2) \times \left\{ 2(2-1)(2-2) + 2(2-1)(2-2) + 2(2-1)(2-2) + 2(2-1)(2-2) + 2(2-1)(2-2) + 2(2-1)(2-2) + 2(2-1)(2-2) \right\}}{36 \times (40-1)(40-2)}$$

$$+ \frac{10(10-1) \times 4 \left\{ 2(2-1) + 2(2-1) + 2(2-1) + 2(2-1) + 2(2-1) + 2(2-1) + 2(2-1) \right\}}{8 \times 40(40-1)} = 1717.003$$

$$J = \cfrac{\left\{75+87+93+88+92+76+\cfrac{1+1+0+1+0+1}{2}-\cfrac{40^2-10^2\times 4}{4}\right\}-0.5}{\sqrt{1717.003}} = \cfrac{212.5}{41.4} = 5.13$$

Note: 40 is total number of observations, 10 is number of observations in each group, 4 is total number of groups and the denominators 2 and 4, and 0.5 are the constants.

$1+1+0+1+0+1=$ $S_{12}+S_{13}+S_{14}+S_{23}+S_{24}+S_{34}$; Number of values repeated across the groups (not within the groups)—the value 43.1 repeated in Groups 1 and 2 ($S_{12}=1$), 32.7 is repeated in Groups 1 and 3 ($S_{13}=1$), no value is repeated in Groups 1 and 4 ($S_{14}=0$), 31.9 is repeated in Groups 2 and 3 ($S_{23}=1$), no value is repeated in Groups 2 and 4 ($S_{24}=0$), and 28.5 is repeated in Groups 3 and 4 ($S_{34}=1$).

Computed value for $J=5.13$ is greater than the point and ($\alpha$) = 3.290 (Table 14.2). Therefore, it could be stated that there is a dose-related trend in the decrease of water consumption of B6C3F1 mice fed on diet containing the test substance at week eight.

**Table 14.2.** Standard normal distribution Table (Yoshimura, 1987)

| Two tailed $P$ | Upper $P$ | % point |
|---|---|---|
| $2\alpha$ | $\alpha$ | U($\alpha$) |
| 0.00100 | 0.000500 | **3.290527** |
| 0.00200 | 0.0010000 | 3.090232 |

### *The Cochran-Armitage test*

The Cochran-Armitage trend test is commonly used to examine whether a dose-response relationship exists in toxicological risk assessment, carcinogenicity studies and several other biomedical experiments (Mehta *et al*., 1998) including mutagenicity studies (Kim *et al*., 2000). It is also widely used in genetics and epidemiology to test linear trend (Buonaccorsi *et al*., 2011). The Cochran-Armitage test for trend is used in categorical data analysis. It can be used to test for linear correlation between a binomial response and an ordinal group variable (Walker and Shostak, 2010). In 1985, the US Federal Register recommended that the analysis of tumour incidence data is carried out with a Cochran-Armitage's trend test (Gad, 2009).

The presence of the antibody to the house dust was investigated in individuals of different age groups (see Table 14.3). Let us examine whether there is a tendency to increase the antibodies to the house dust as the age of the individuals increases.

**Table 14.3.** Individuals of different age groups expressing antibodies to house dust

| Age | Conversion value | Independent variable (log transformed) | Number of investigations | Number of antibody positives |
|---|---|---|---|---|
| One's sixties | 2.5 | 0.398 | 10 | 2 |
| One's fifties | 5 | 0.699 | 10 | 4 |
| One's forties | 10 | 1.000 | 10 | 6 |
| One's thirties | 20 | 1.301 | 10 | 8 |

A value of 10 is assigned to the age forties. Half of the value of the age forties (10/2=5) is assigned to the age fifties and half of the value of age fifties (5/2=2.5) is assigned to the age sixties. The value assigned for the age thirties is 20 (10x2).

Number of group = 4, Sum of number of sample = 40, rate of positive in total = (2+4+6+8)/40= 20/40= 0.5

$$\text{Mean} = \frac{(10 \times 0.398 + 10 \times 0.699 + 10 \times 1.000 + 10 \times 1.301}{40} = 0.8495$$

$$X^2 = \frac{\left\{(2 \times 0.398 + 4 \times 0.699 + 6 \times 1.000 + 8 \times 1.301) - 40 \times 0.5 \times 0.8495\right\}^2}{0.5 \times (1 - 0.5) \times 10 \times \left\{(0.398 - 0.8495)^2 + (0.699 - 0.8495)^2 + (1.000 - 0.8495)^2 + (1.301 - 0.8495)^2\right\}}$$

$$= \frac{9.0601}{1.1325} = 8.000$$

From the chi-square Table (Table 14.4), for one degree of freedom, we find that the calculated value (8.000) is greater than the chi-square Table value (6.635) at 0.01 probability level. Hence, we conclude that there is a tendency to increase the antibodies to the house dust as the age of the person increases.

**Table 14.4.** Chi-square (Yoshimura, 1987)

| DF | α | | | |
|---|---|---|---|---|
| | 0.100 | 0.050 | 0.010 | 0.001 |
| 1 | 2.705 | **3.841** | **6.635** | 10.82 |
| 2 | 4.605 | 5.991 | 9.210 | 13.81 |
| 3 | 6.251 | 7.814 | 11.34 | 16.26 |
| 4 | 7.779 | 9.487 | 13.27 | 18.46 |
| 5 | 9.236 | 11.07 | 15.08 | 20.51 |
| 6 | 10.64 | 12.59 | 16.81 | 22.45 |
| 7 | 12.01 | 14.06 | 18.47 | 24.32 |
| 8 | 13.36 | 15.50 | 20.09 | 26.12 |
| 9 | 14.68 | 16.91 | 21.66 | 27.87 |
| 10 | 15.98 | 18.30 | 23.20 | 29.58 |

Armitage (1955) recommended the Cochran-Armitage test in case there is no *a priori* knowledge of the type of the trend. The Cochran-Armitage test is asymptotically efficient for all monotone alternatives (Tarone and Gart, 1980). But, this test should not be used for the data showing an extra-Poisson variability (Astuti and Yanagawa, 2002), where estimated variance exceeds estimated means. Antonello *et al*. (1993) stated that Tukey trend test is more powerful for monotonic dose-response toxicologic effects than the pair-wise comparison tests. But dichotomous endpoints are frequently observed in several toxicologic effects. For analysing dichotomous endpoints, Neuhauser and Hothorn (1997) proposed a trend test analogous to the nonparametric Jonckheere's trend test.

We propose Jonckheere's trend test for the analysis of quantitative data, such as body weight, erythrocyte count, alkaline phosphatase and organ weights. For qualitative data, such as a macroscopic-, microscopic-pathological findings and urinalysis (color, pH, protein, glucose, ketone, bilirubin and urobilinogen) we propose Cochran-Armitage test.

## References

Antonello, J.M., Clark, R.L. and Heyse, J.F. (1993): Application of the Tukey trend test procedure to assess developmental and reproductive toxicity I. Measurement data. Tox. Sci., 21(1), 52–58.

Armitage, P. (1955): Tests for linear trends in proportions and frequencies. Biometrics, 11 (3), 375–386.

Astuti, E.T. and Yanagawa, T. (2002): Testing trend for count data with extra-Poisson variability. Biometrics, 58(2), 398–402.

Buonaccorsi, J.P., Laake, P. and Veierød, M.B. (2011): On the power of the Cochran-Armitage test for trend in the presence of misclassification. Stat. Methods Med. Res., August 2011; doi:10.1177/0962280211406424.

Cochran, W.G. (1954): Some methods for strengthening the common chi-square tests. Biometrics, 10(4), 417–451.

Cohen, L. and Holliday, M. (2001): Practical Statistics for Students. SAGE Publications Inc., California, USA.

EPA (2005): United States Environmental Protection Agency. Guidelines for Carcinogen Risk Assessment. U.S. Environmental Protection Agency, EPA/630/P-03/001F. USEPA, Washington D.C., USA.

Field, A.P. (2004): Discovering Statistics Using SPSS, 2nd Edition, SAGE, London, UK.

Gad, S.C. (2009): Drug Safety Evaluation, 2nd Edition. John Wiley & Sons, Inc., New Jersey.

Hamada, C., Yoshino, K., Matsumoto, K., Ikumi Abe, I., Yoshimura, I. and Nomura, M. (1997): A study on the consistency between statistical evaluation and toxicological judgment. Drug Inf. J., 31, 413–421.

Jonckheere A.R. (1954): A distribution-free *k*-sample test against ordered alternatives. Biometrika, 41, 133–145.

Jones, M.P. (2001): Unmasking the trend sought by Jonckheere-type tests for right censored data. Scand. J. Stat., 28(3), 527–535.

Kim, B.S., Zhao, B., Kim, H.J. and Cho, M.H. (2000): The statistical analysis of the *in vitro* chromosome aberration assay using Chinese hamster ovary cells. Mut. Res., 469( 2), 243–252.

Mehta, C.R., Patel, N.R. and Senchaudhuri, P. (1988): Exact power and sample size computations for the Cochran-Armitage trend test. Biometrics, 54, 1615–1621.

Neuhauser, M. and Hothorn, L.A. (1997): Trend tests for dichotomous end points with application to carcinogenicity studies. Drug Inf. J., 31, 463–469.

Neuhäuser, M., Liu, P.Y. and Hothorn, L.A. (1999): Nonparametric tests for trend: Jonckheere's test, a modification and a maximum test. Biometrical J., 40(8), 899–909.

Selwyn, M.R. (1995): The use of trend tests to determine a no-observable-effect level in animal safety studies. Int. J. Toxicol., 14(2), 158–168.

Tarone, R.E. and Gart, J.J. (1980): On the robustness of combined tests for trends in proportions. J. Am. Stat. Assoc., 75, 110–116.

Walker, G.A. and Shostak, J. (2010): Common Statistical Methods for Clinical Research with SAS Examples, 3rd Edition. SAS Inst. Inc., North Carolina.

Yoshimura, I. (1987): Statistical Analysis of Toxicological Data, Scientist Press, Tokyo, Japan.

# 15
# Survival Analysis

## Introduction

Survival analysis is one of the oldest fields of statistics, going back to the 17th century. The first life-table was presented by John Graunt in 1662 (Kreager, 1988). Life-tables are used extensively in analysing the mortality data obtained from toxicology studies, especially carcinogenicity and long-term repeated dose administration studies (Portier, 1988; FDA, 2007) and ecotoxicology studies (Gentile *et al*., 1982; Van Leeuwen *et al*., 1985; Bechmann, 1994). A major advancement in the survival analysis took place in 1958, when Kaplan and Meier proposed their 'estimator of the survival curve' (Kaplan and Meier, 1958). Since then, the field of survival analysis progressed significantly with the contributions from several statisticians (Mantel and Haenszel, 1959; Cox, 1972; Aalen, 1976; Aalen, 1980; Diggle, *et al*., 2007; Aalen *et al*., 2008). The term "survival" is a bit misleading. Originally the analysis was concerned with time from treatment until death, hence the name, "survival analysis". Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs (Kleinbaum and Klein, 2005). According to Akritas (2004), survival analysis is a method for the analysis of data on an event observed over time and the study of factors associated with the occurrence rates of this event. The event could be the time until a generator's bearing seizes, the time until a patient dies or the time until a person finds employment (Cleves *et al*., 2008). Survival analysis can be used in many fields, such as medicine, biology, public health and epidemiology (Kul, 2010). In pharmacology and toxicology survival analysis is used in analyzing the events like time to death, time to signs occurrence**,** disappearance and reoccurrence, time to recovery etc. of the experimental animals.

Another terminology that we need to understand in survival analysis is 'censored observation'. When animals do not have an event during the observation time, they are described as censored. Censored animals may or may not have an event after the end of observation time.

## Hazard Rate

'Hazard rate' is an important concept in survival analysis. It provides information on the risk of event happening as a function of time, condition on not having happened previously (Aalen *et al.*, 2009), whereas survival curve provides information on how many have survived upto a certain time. Hazard function can be estimated using the equation:

H (t) = Number of individuals experiencing an event in interval beginning at t/(number of individuals surviving at time *t*) x (interval width)

The hazard function describes the risk of an outcome of an event in an interval after time *t*, conditional on the individual having experienced the event to time *t*. The hazard function is useful in determining whether toxicity is constant over time, or it increases or decreases as the exposure continues (Wright and Welbourn, 2002).

## Kaplan-Meier Method

Survival analysis is normally carried out using Kaplan-Meier method or the log rank test. The log rank test is ideal for the analysis of two groups. The Kaplan–Meier estimator uses product-limit methods to estimate the survival ratio (Kaplan and Meier, 1958). This is a nonparametric maximum likelihood estimate of survival analysis and is used in animal experiments to measure the fraction of animals that lives after treatment.

Distribution of the survival time T from the start of the experiment (first dose administration) to the event of interest (for example mortality) is considered as a random variable. The survival rate, $S_t$, is defined as the probability that an animal survives longer than *t* units of time:

$S_t$=P (T> *t*); for example, if *t* is in years, $S_2$ is the two-year survival rate; if $S_2$=P (T> 2)=0.10, it indicates 10% is the probability the time from a treatment to death is greater than 2 years

### Kaplan-Meier product-limit estimator

$$S_t = \prod \frac{r_i - d_i}{r_i},$$

$r_i$ is the number of animals lived just before $t_i$; $d_i$ is the number of animals which died in $t_i$. $\Pi$ denotes the product (geometric sum) across all cases less than or equal to $t$. Kaplan-Meier product-limit estimator measures the fraction of animals living for a certain amount of time after treatment.

Let us review an example to understand Kaplan-Meier product-limit estimator. The survival rate of F344 rats in a 110-week chronic toxicity study is given in Table 15.1. The experimental group of rats (20 rats/group) was treated with 1000 ppm pesticide in diet. The control group of rats (20 rats/group) was given normal diet without the pesticide.

**Table 15.1.** Survival rate of F344 rats in a 110-week chronic toxicity (Funaki and Origasda, 2001)

| Control group (Normal diet) | | | | Treatment group (1000 ppm pesticide in diet) | | | |
|---|---|---|---|---|---|---|---|
| Animal ID-No. | Survival period (week) | Survival rate ($s_t$) | Size of effective sample (n') | Animal ID-No. | Survival period (week) | Survival rate ($s_t$) | Size of effective sample (n') |
| 1001 | 85 | 0.950 | 20 | 1101 | 66 | 0.900 | 20 |
| 1002 | 87 | 0.900 | 19 | 1102 | 66 | | |
| 1003 | 95 | 0.800 | 18 | 1103 | 62 | 0.850 | 18 |
| 1004 | 95 | | | 1104 | 63 | 0.800 | 17 |
| 1005 | 99 | 0.650 | 16 | 1105 | 68 | 0.750 | 16 |
| 1006 | 99 | | | 1106 | 70 | 0.650 | 15 |
| 1007 | 99 | | | 1107 | 70 | | |
| 1008 | 101 | 0.550 | 13 | 1108 | 72 | 0.550 | 13 |
| 1009 | 101 | | | 1109 | 72 | | |
| 1010 | 102 | 0.500 | 11 | 1110 | 75 | 0.400 | 11 |
| 1011 | 103 | 0.350 | 10 | 1111 | 75 | | |
| 1012 | 103 | | | 1112 | 75 | | |
| 1013 | 103 | | | 1113 | 77 | 0.300 | 8 |
| 1014 | 104 | 0.250 | 7 | 1114 | 77 | | |
| 1015 | 104 | | | 1115 | 78 | 0.57 | 7 |
| 1016 | 106 | 0.150 | 5 | 1116 | 79 | 0.154 | 5 |
| 1017 | 106 | | | 1117 | 79 | | |
| 1018 | 110 | 0.050 | 3 | 1118 | 80 | 0.051 | 3 |
| 1019 | 112 | 0.025 | 2 | 1119 | 80 | | |
| 1020 | 120 | - | 1 | 1120 | 88 | - | 1 |

The survival rate is calculated using the equation:

$$S_t = \prod \frac{r_i - d_i}{r_i}$$

Calculation procedure of $S_t$ is given below by working out few selected survival period of control group:

Week 85 ($S_{85}$):  $0.950 = \dfrac{20-1}{20}$

Week 87 ($S_{87}$):  $0.900 = 0.950 \times \dfrac{19-1}{19}$

Week 95 ($S_{95}$):  $0.800 = 0.900 \times \dfrac{18-2}{18}$

Week 112 ($S_{112}$):  $0.025 = 0.05 \times \dfrac{2-1}{2}$

Thus $S_t$ is calculated for all survival periods of control group and given in Table 15.1.

Calculation of standard error of $S_t$:

$$SE = \sqrt{\frac{S_t(1-S_t)}{n'}}$$

Let us calculate SE for ($S_{104}$):

Week 104 ($S_{104}$):  $0.25 = 0.35 \times \dfrac{7-2}{7}$

$$SE = \sqrt{\frac{0.25(1-0.25)}{7}} = 0.164$$

Survival rate at 95% confidence interval is:

$$0.25 - (1.96 \times 0.164) \sim 0.25 + (1.96 \times 0.164) = 0\text{–}0.57$$

The 95% confidence interval exploded in a wide range, because of the small sample size (N=7).

Similarly survival rate of the treatment group is computed and given in Table 15.1. Plot of survival curves is an important part of survival analysis (Freeman *et al.*, 2008). A plot of the survival curves of data (Table 15.1) is given in Figure 15.1.

Though the survival curves provide a good information on the mortality rates in two groups, the comparison of the curves should be made using a statistical tool (Altman, 1991). Log-rank test is the common method used to compare survival curves (Cox, 1972). This test assigns equal weight to each event at whatever time it occurs (Tinazzi *et al.*, 2008). The null
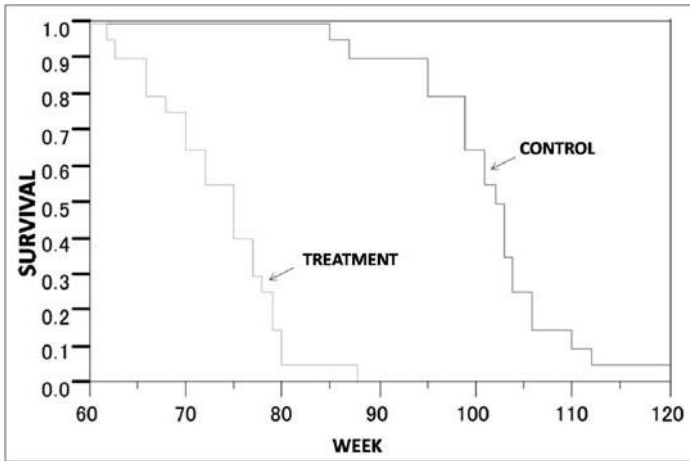
**Figure 15.1.** Survival rate of F344 rats in a 110-week chronic toxicity

hypothesis for the log-rank test is that there is no difference between the survivals of two or more populations that are being compared. The comparison is based on the difference between the observed number of events in each group and the expected number of events in case of non-difference between the two groups. The $\chi^2$ equation is:

$$\chi^2{}_{log-rank} = \sum_g \frac{\left(O_g - E_g\right)^2}{E_g}$$

where $O$ is the number of observed events in each group $g$, and $E$ is the total number of expected events in each group $g$. $O$ and $E$ are computed each time an event happens; if a survival time is censored, then the subject is considered to be at risk during the interval of censoring, but not anymore for the subsequent intervals. The test statistic is then compared with a $\chi^2$ with $g$-1 degrees of freedom. The limitation of log rank test and Cox's proportional hazards model is that they are based on the assumption that the hazard ratio is constant over time (Bewick *et al.*, 2004).

Both the life-table and the Kaplan-Meier methods have advantages and disadvantages. In small data sets in which the time of occurrence event is measured precisely the Kaplan-Meier method is best used, whereas the life-table methods works well with large data sets and when the time of occurrence of an event cannot be measured precisely. The Kaplan-Meier method handles censored data better than life-table method.

# References

Aalen, O.O. (1976): Nonparametric inference in connection with multiple decrement models. Scand. J. Stat., 3, 15–27.

Aalen, O.O. (1980): A model for non-parametric regression analysis of lifetimes. In: Mathematical Statistics and Probability Theory. Editors, Klonecki, W., Kozek, A. and Rosinski, J. Lecture Notes in Statistics, Vol. 2, Springer-Verlag, NewYork.

Aalen, O.O., Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (2009): History of applications of martingales in survival analysis. Electronic J. History Probability Stat., 5(1), 1–28.

Aalen, O.O., Borgan, Ø. and Gjessing, H.K. (2008): Survival and Event History Analysis: A Process Point of View. Springer-Verlag, NewYork.

Akritas, M.G. (2004): Nonparametric survival analysis. Stat. Sci., 19(4), 615–623.

Altman, D.G. (1991): Practical Statistics for Medical Research. Chapman & Hall/CRC, London.

Bechmann, R.K. (1994): Use of life tables and $lC_{50}$ tests to evaluate chronic and acute toxicity effects of copper on the marine copepod *Tisbe furcata* (baird), Environmental Toxicology and Chemistry. 13(9), 1381–1548.

Bewick, V., Cheek, L. and Ball, J. (2004): Statistics review 12: Survival analysis. Crit Care, 8(5), 389–394.

Cleves, M., Gutierrez, R., Gould, W. and Marchenko, Y. (2008): An Introduction to Survival Analysis Using Data. 2nd Edition, Stata Press, Texas, USA.

Cox, D.R. (1972): Regression models and life tables (with discussion). J. Royal Stat. Soc. Ser., B34, 187–220.

Diggle, P., Farewell, D.M. and Henderson, R. (2007): Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. J. Royal Stat. Soc. Ser. C (Applied Statistics), 56, 499–550.

FDA (2007): United States Food and Drug Administration. Redbook 2000: IV.B.4 Statistical Considerations in Toxicity Studies. USFDA, MD, USA.

Freeman, J.V., Walters, S.J. and Campbell, M.J. (2008): How to display data. Blackwell BMJ Books, Oxford, UK.

Funaki, K. and Origasda, H. (2001): Statistics with Confidence, Scientist Press, Tokyo, Japan.

Gentile, J.H., Gentile, S.M., Hairston, N.G. and Sullivan, B.K. (1982): The use of life-tables for evaluating the chronic toxicity of pollutants to *Mysidopsis bahia*. Hydrobiologia, 93(1-2), 179–187.

Kaplan, E. L. and Meier, P. (1958): Nonparametric estimation from incomplete observations. J. Am. Stat. Assn., 53, 457–481.

Kleinbaum, D.G and Klein, M. (2005): Survival Analysis-A Self Learning Text. 2nd Edition, Springer+Business Media, Inc., New York, USA.

Kreager, P. (1988): Newlighton Graunt. Population Studies, 42,129–140.

Kul, S. (2010): The use of survival analysis for clinical pathways. Intl. J. Care Path Ways, 14, 23–26.

Mantel, N. and Haenszel, W. (1959): Statistical aspects of the analysis of data from retrospective studies of disease. J. National Cancer Inst., 22, 719–748.

Portier, C.J. (1988): Life table analysis of carcinogenicity experiments. Int. J.Tox., 7(5), 575–582.

Tinazzi, A., Scott, M. and Compagnoni, A. (2008): A gentle introduction to survival analysis. SAS Conference Proceedings: PhUSE 2008, October 12–15, 2008, Manchester, UK.

Van Leeuwen, C.J., Moberts, F. and Niebeek, G. (1985): Aquatic toxicological aspects of dithiocarbamates and related compounds. II. Effects on survival, reproduction and growth of *Daphnia magna*. Aquatic Toxicol., 7(3), 165–175.

Wright, D.A. and Welbourn, P. (2002): Environmental Toxicity. Cambridge Environmental Chemistry Series 11. Cambridge University Press, Cambridge, UK.