CHAPTER

11

DATA DIAGNOSTICS AND REMEDIAL MEASURES

Eldho Varghese

Fishery Resources Assessment Division
ICAR-Central Marine Fisheries Research Institute

The raw data consist of measurements of some attribute on a collection of individuals. The measurement would have been made in one of the following scales *viz*., nominal, ordinal, interval or ratio scale.

Levels of Measurement

- **Nominal scale** refers to measurement at its weakest level when number or other symbols are used simply to classify an object, person or characteristic, *e.g.*, state of health (healthy, diseased).
- **Ordinal scale** is one wherein given a group of equivalence classes, the relation greater than holds for all pairs of classes so that a complete rank ordering of classes is possible, *e.g*., socio-economic status.
- When a scale has all the characteristics of an ordinal scale, and when in addition, the distances between any two numbers on the scale are of known size, **interval scale** is achieved, e.*g*., temperature scales like centigrade or Fahrenheit.
- An interval scale with a true zero point as its origin forms a ratio scale. In a **ratio scale**, the ratio of any two scale points is independent of the unit of measurement, e.g., height of trees.

The data can be classified as qualitative/quantitative depending on the levels based on which the observations are collected. There are several statistical procedures available in literature for the analysis of data which are broadly classified in to two categories *viz*., parametric tests and non-parametric tests. A parametric test specifies certain conditions about the distribution of responses in the population from which the research sample was drawn. The meaningfulness of the results of a parametric test depends on the validity of these assumptions. A nonparametric test is based on a model that specifies very general conditions and none regarding the specific form of the distribution from which the sample was drawn. Hence nonparametric tests are also known as distribution free tests. Certain assumptions are associated with most nonparametric statistical tests, but these are fewer and weaker than those of parametric tests.

Nonparametric test statistics utilize some simple aspects of sample data such as the signs of measurements, order relationships or category frequencies. Therefore, stretching

or compressing the scale does not alter them. As a consequence, the null distribution of the nonparametric test statistic can be determined without regard to the shape of the parent population distribution. These tests have the obvious advantage of not requiring the assumption of normality or the assumption of homogeneity of variance. They compare medians rather than means and, as a result, if the data have one or two outliers, their influence is negated.

Besides, the interpretation of data based on analysis of variance (ANOVA)/Regression is valid only when the following assumptions are satisfied:

1. The regression function is linear
2. The error terms do have constant variance
3. The error terms are independent
4. No outlying observations
5. The error terms are normally distributed
6. Predictor variables are uncorrelated.

Also the statistical tests t, F, z, *etc*. are valid under the assumption of independence of errors and normality of errors. The departures from these assumptions make the interpretation based on these statistical techniques invalid. Therefore, it is necessary to detect the deviations and apply the appropriate remedial measures.

The assumption of independence of errors, *i.e*., error of an observation is not related to or depends upon that of another. This assumption is usually assured with the use of proper randomization procedure. However, if there is any systematic pattern in the arrangement of treatments from one replication to another, errors may be non-independent. This may be handled by using nearest neighbour methods in the analysis of experimental data.

Normality of Errors

The assumptions of homogeneity of variances and normality are generally violated together. To test the validity of normality of errors for the character under study, one can take help of Normal Probability Plot, Anderson-Darling Test, D'Augstino's Test, Shapiro - Wilk's Test, Ryan-Joiner test, Kolmogrov-Smirnov test, *etc*. In general moderate departures from normality are of little concern in the fixed effects ANOVA as F - test is slightly affected but in case of random effects, it is more severely impacted by non-normality. The significant deviations of errors from normality, makes the inferences invalid. So before analyzing the data, it is necessary to convert the data to a scale that it follows a normal distribution. In the data from designed field experiments, we do not directly use the original data for testing of

normality or homogeneity of observations because this is embedded with the treatment effects and some of other effects like block, row, column, *etc*. So there is need to eliminate these effects from the data before testing the assumptions of normality and homogeneity of variances. For eliminating the treatment effects and other effects we fit the model corresponding to the design adopted and estimate the residuals. These residuals are then used for testing the normality of the observations. In other words, we want to test the null hypothesis H_0: errors are normally distributed against alternative hypothesis H_1: errors are not normally distributed. For details on these tests one may refer to D'Agostino and Stephens (1986). Most of the standard statistical packages available in the market are capable of testing the normality of the data. In SAS and SPSS commonly used tests are Shapiro-Wilk test and Kolmogrov-Smirnov test. MINITAB uses three tests *viz*. Anderson-Darling, Ryan-Joiner, Kolmogrov-Smirnov for testing the normality of data.

Homogeneity of Error Variances

A crude method for detecting the heterogeneity of variances is based on scatter plots of means and variance or range of observations or errors, residual vs fitted values, *etc*. To be clearer, let Y_{ij} be the observation pertaining to i^{th} treatment (i=1(1)v) in the j^{th} replication (j = 1(1)r_i). Compute the mean and variance for each treatment across the replications (the range can be used in place of variance) as

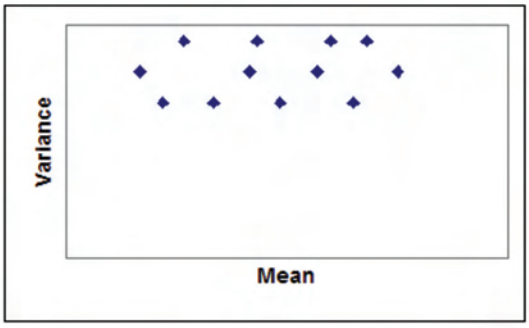
Mean = $\bar{Y}_{i.} = \frac{1}{r_i}\sum_{j=1}^{r_i} Y_{ij}$; Variance = $S_i^2 = \frac{1}{r_i - 1}\sum_{j=1}^{r_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$

Draw the scatter plot of mean vs variance (or range). If $S_{i.}^2$'s $\left(i = 1(1)v\right)$ are equal (constant) or nearly equal, then the variances are homogeneous. Based on these scatter plots, the heterogeneity of variances can be classified into two types:

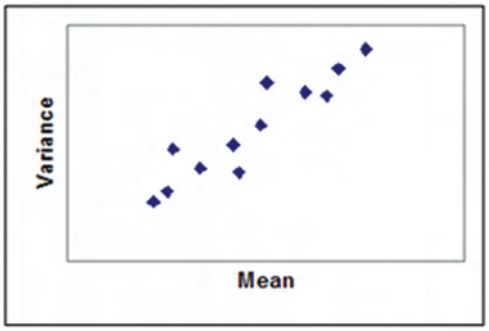
1. Where the variance is functionally related to mean.
2. Where there is no functional relationship between the variance and the mean.

For illustration some scatter - diagrams of mean and variances (or range) are given in below:

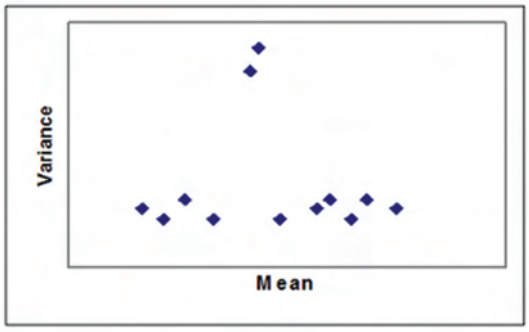
The first kind of variance heterogeneity (figure b) is usually associated with the data whose distribution is non-normal *viz.,* negative binomial, Poisson, binomial, *etc*. The second kind of variance heterogeneity usually occurs in experiments, where, due to the nature of treatments tested, some treatments have errors that are substantially higher (lower) than others. For example, in varietal trials, where various types of breeding material are being compared, the size of variance between plots of a particular variety will depend on the degree of genetic homogeneity of material being tested. The variance of F_2 generation, for example, can be expected to be higher than that of F_1 generation because genetic variability in F_2 is much higher than that in F_1. The variances of varieties that are highly tolerant of or



(a) Homogeneous variance



(b) Heterogeneous variance where variance is proportional to mean



(c) Heterogeneous variance without any functional relationship between variance and mean

highly susceptible to, the stress being tested are expected to be smaller than those of having moderate degree of tolerance. Also in testing yield response to a chemical treatment, such as, fertilizer, insecticide or herbicide, the non-uniform application of chemical treatments may result in a higher variability in the treated plots than that in the untreated plots.

The scatter-diagram of means and variances of observations for each treatment across the replications gives only a preliminary idea about homogeneity of error variances. Statistically the homogeneity of error variances is tested using Bartlett's test for normally distributed errors and Levene test for non-normal errors. These tests are described in the sequel.

Bartlett's Test for Homogeneity of Variances

Let there are v- independent samples drawn from same population and i^{th} sample is of size r_i and $(r_1 + r_2 + \ldots + r_v) = N$. In the present case, the independent samples are the residuals of the observations pertaining to v treatments and i^{th} sample size is the number of

replications of the treatment *i*. One wants to test the null hypothesis H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_v^2$ against the alternative hypothesis H_1 : at least two of the σ_1^2's are not equal, where σ_1^2 is the error variance for treatment *i*.

Let e_{ij} denotes the residual pertaining to the observation of treatment *i* from replication *j*, then it can easily be shown that the sum of residuals pertaining to a given treatment is zero. In this test $S_i^2 = \frac{1}{r_i - 1}\sum_{j=1}^{r_i}(e_{ij} - \bar{e}_i)^2 = \frac{1}{r_i - 1}\sum_{j=1}^{r_i} e_{ij}^2$ is taken as unbiased estimate of σ_i^2. The procedure involves computing a statistic whose sampling distribution is closely approximated by the χ^2 distribution with *v - 1* degrees of freedom. The test statistic is

$$\chi_0^2 = 2.3026\frac{q}{c}$$

and null hypothesis is rejected when $\chi_0^2 > \chi^2_{\alpha,v-1}$, where $\chi^2_{\alpha,v-1}$ is the upper α percentage point of χ^2 distribution with *v - 1* degrees of freedom.

To compute χ_0^2, follow the steps:

Step 1: Compute mean and variance of all *v*-samples.

Step 2: Obtain pooled variance $S_p^2 = \frac{\sum_{i=1}^{v}(r_i - 1)S_i^2}{N - v}$

Step 3: Compute $(N - v)\log_{10} S_p^2 - \sum_{i=1}^{v}(r_i - 1)\log_{10} S_i^2$

Step 4: Compute $1 + \frac{1}{3(v-1)}\left(\sum_{i=1}^{v}(r_i - 1)^{-1} - (N - v)^{-1}\right)$

Step 5: Compute χ_0^2.

Bartlett's χ^2 test for homogeneity of variances is a modification of the normal-theory likelihood ratio test. While Bartlett's test has accurate Type I error rates and optimal power when the underlying distribution of the data is normal, it can be very inaccurate if that distribution is even slightly non-normal (Box 1953). Therefore, Bartlett's test is not recommended for routine use.

An approach that leads to tests that are much more robust to the underlying distribution is to transform the original values of the dependent variable to derive a *dispersion variable* and then to perform analysis of variance on this variable. The significance level for the test of homogeneity of variance is the *p*-value for the ANOVA *F*-test on the dispersion variable. Commonly used test for testing the homogeneity of variance using a dispersion variable is Levene Test given by Levene (1960). The procedure is described in the sequel.

Levene Test for homogeneity of Variances

The test is based on the variability of the residuals. The larger the error variance, the larger the variability of the residuals will tend to be. To conduct the Levene test, we divide the data into different groups based on the number of treatments if the error variance is either increasing or decreasing with the treatments, the residuals in the one treatment will tend to be more variable than those in others treatments. The Levene test than consists simply F – statistic based on one way ANOVA used to determine whether the mean of absolute/ Square root deviation from mean are significantly different or not. The residuals are obtained from the usual analysis of variance. The test statistic is given as

$$F = \frac{\left\{\sum_{i=1}^{v}(r_i - 1)\right\}\left\{\sum_{i=1}^{v} r_i(\bar{d}_{i.} - \bar{d}_{..})^2\right\}}{v-1} \Big/ \frac{}{\sum_{i=1}^{v}\sum_{j=1}^{r_i}(d_{ij} - \bar{d}_{i.})^2} \sim F\left((v-1), \sum_{i=1}^{v}(r_i - 1)\right)$$

where $d_{ij} = |e_{ij} - \bar{e}_{i.}|;\ \bar{d}_{i.} = \frac{\sum_{j=1}^{r_i} d_{ij}}{r_i};\ \bar{d}_{..} = \frac{\sum_{i=1}^{v}\sum_{j=1}^{r_i} d_{ij}}{\sum_{j=1}^{r_i} r_i}$ and d_{ij} is the j^{th} residual for the i^{th} plot, e_i is the mean of the residuals of the i^{th} treatment.

This test was modified by Brown and Forsythe (1974). In the modified test, the absolute deviation is taken from the median instead of mean in order to make the test more robust.

In the present investigation, the Bartlett's χ^2 – test has been used for testing the homogeneity of error variances when the distribution of errors is normal and Levene test for non-normal errors.

Remark 1: In a block design**,** it can easily be shown that the sum of residuals within a given block is zero. Therefore, the residuals in a block of size 2 will be same with their sign reverse in order. As a consequence, *q* in Bartlett's test and numerator in Levene test statistic becomes zero for the data generated from experiments conducted to compare only two treatments in a RCB design. Hence, the tests for homogeneity of error variances cannot be used for the experiments conducted to compare only two treatments in a RCB design. Inferences from such experiments may be drawn using Fisher-Behren t-test. Further, Bartlett's test cannot be used for the experimental situations where some of the treatments are singly replicated.

Remark 2: In a RCB design, it can easily be shown that the sum of residuals from a particular treatment is zero. As a consequence, the denominator of Levene test statistic is

zero for the data generated from RCB designs with two replications. Therefore, Levene test cannot be used for testing the homogeneity of error variances for the data generated from RCB designs with two replications.

Presence of Outliers

Outliers are extreme observations. Residual outliers can be identified from residual plots against *X or Y*. Outliers can create great difficulty. When we encounter one, our first suspicion is that the observation resulted from a mistake or other extraneous effect. On the other hand, outliers may convey significant information, as when an outlier occurs because of an interaction with another predictor omitted from the model. A safe rule frequently suggested is to discard an outlier only if there is direct evidence that it represents in error in recording, a miscalculation, a malfunctioning of equipment, or a similar type of circumstances.

Omission of Important Predictor Variables

Residuals should also be plotted against variables omitted from the model that might have important effects on the response. The purpose of this additional analysis is to determine whether there are any key variables that could provide important additional descriptive and predictive power to the model. The residuals are plotted against the additional predictor variable to see whether or not the residuals tend to vary systematically with the level of the additional predictor variable.

Overview of Tests

Graphical analysis of residuals is inherently subjective. Nevertheless, subjective analysis of a variety of interrelated residuals plots will frequently reveal difficulties with the model more clearly than particular formal tests.

Tests for Randomness

A run test is frequently used to test for lack of randomness in the residuals arranged in time order. Another test, specially designed for lack of randomness in least squares residuals.

Durbin-Watson test:

The Durbin-Watson test assumes the first order autoregressive error models. The test consists of determining whether or not the autocorrelation coefficient (p, say) is zero. The usual test alternatives considered are:

$H_0: p = 0$

$H_0: p > 0$

The Durbin-Watson test statistic D is obtained by using ordinary least squares to fit the

regression function, calculating the ordinary residuals: $e_t = Y_t - Y_{t'}$ and then calculating the statistic:

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{n\sum_{t=1} e_t^2}$$

Exact critical values are difficult to obtain, but Durbin-Watson have obtained lower and upper bound and such that a value of D outside these bounds leads to a definite decision. The decision rule for testing between the alternatives is:

if $D > d_U$, conclude H_0

if $D < d_L$, conclude H_1

if $dL < D < d_U$, test is inconclusive.

Small value of D lead to the conclusion that $p>0$.

Multi-collinearity

The use and interpretation of a multiple regression model depends implicitly on the assumption that the explanatory variables are not strongly interrelated. In most regression applications the explanatory variables are not orthogonal. Usually the lack of orthogonality is not serious enough to affect the analysis. However, in some situations the explanatory variables are so strongly interrelated that the regression results are ambiguous. Typically, it is impossible to estimate the unique effects of individual variables in the regression equation. The estimated values of the coefficients are very sensitive to slight changes in the data and to the addition or deletion of variables in the equation. The regression coefficients have large sampling errors which affect both inference and forecasting that is based on the regression model. The condition of severe non-orthogonality is also referred to as the problem of multicollinearity.

Data transformation

In this section, we shall discuss the remedial measures for non-normal and/or heterogeneous data in greater details.

Data transformation is the most appropriate remedial measure, in the situation where the variances are heterogeneous and are some functions of means. With this technique, the original data are converted to a new scale resulting into a new data set that is expected to satisfy the homogeneity of variances. Because a common transformation scale is applied to all observations, the comparative values between treatments are not altered and comparison between them remains valid.

Error partitioning is the remedial measure of heterogeneity that usually occurs in experiments, where, due to the nature of treatments tested some treatments have errors that are substantially higher (lower) than others.

Here, we shall concentrate on those situations where character under study is non-normal and variances are heterogeneous. Depending upon the functional relationship between variances and means, suitable transformation is adopted. The transformed variate should satisfy the following:

1. The variances of the transformed variate should be unaffected by changes in the means. This is also called the variance stabilizing transformation.
2. It should be normally distributed.
3. It should be one for which effects are linear and additive.
4. The transformed scale should be such for which an arithmetic average from the sample is an efficient estimate of true mean.

The following are the three transformations, which are being used most commonly, in biological research.

a) Logarithmic Transformation

b) Square root Transformation

c) Arc Sine or Angular Transformation

a) Logarithmic Transformation

This transformation is suitable for the data where the variance is proportional to square of the mean or the coefficient of variation (S.D./mean) is constant or where effects are multiplicative. These conditions are generally found in the data that are whole numbers and cover a wide range of values. This is usually the case when analyzing growth measurements such as trunk girth, length of extension growth, weight of tree or number of insects per plot, number of eggmass per plant or per unit area *etc.*

For such situations, it is appropriate to analyze *log X* instead of actual data, *X*. When data set involves small values or zeros, *log (X+1), log (2X + 1)* or *log* $\left(X+\frac{3}{8}\right)$ should be used instead of *log X*. This transformation would make errors normal, when observations follow negative binomial distribution like in the case of insect counts.

b) Square-Root Transformation

This transformation is appropriate for the data sets where the variance is proportional to the mean. Here, the data consists of small whole numbers, for example, data obtained in

counting rare events, such as the number of infested plants in a plot, the number of insects caught in traps, number of weeds per plot, parthenocarpy in some varieties of mango. This data set generally follows the Poisson distribution and square root transformation approximates Poisson to normal distribution.

For these situations, it is better to analyze $\sqrt{X}$ than that of *X*, the actual data. If X is confirmed to small whole numbers then, $\sqrt{X+\frac{1}{2}}$ or $\sqrt{X+\frac{3}{8}}$ should be used instead of $\sqrt{X}$.

This transformation is also appropriate for the percentage data, where, the range is between 0 to 30% or between 70 and 100%.

c) Arc Sine Transformation

This transformation is appropriate for the data on proportions, *i.e.,* data obtained from a count and the data expressed as decimal fractions and percentages. The distribution of percentages is binomial and this transformation makes the distribution normal. Since the role of this transformation is not properly understood, there is a tendency to transform any percentage using arc sine transformation. But only that percentage data that are derived from count data, such as % barren tillers (which is derived from the ratio of the number of non-bearing tillers to the total number of tillers) should be transformed and not the percentage data such as % protein or % carbohydrates, %nitrogen, *etc.* which are not derived from count data. For these situations, it is better to analyze $\sin^{-1}(\sqrt{X})$ than that of *X*, the actual data. If the value of X is 0%, it should be substituted by $\left(\frac{1}{4n}\right)$ and the value of 100% by $\left(100-\frac{1}{4n}\right)$, where *n* is the number of units upon which the percentage data is based.

It is interesting to note here that not all percentage data need to be transformed and even if they do, arc sine transformation is not the only transformation possible. The following rules may be useful in choosing the proper transformation scale for percentage data derived from count data.

Rule 1: The percentage data lying within the range 30 to 70% is homogeneous and no transformation is needed.

Rule 2: For percentage data lying within the range of either 0 to 30% or 70 to 100%, but not both, the square root transformation should be used.

Rule 3: For percentage that do not follow the ranges specified in Rule 1 or Rule 2, the Arc Sine transformation should be used.

The other transformations used are reciprocal square root $\left[\frac{1}{\sqrt{X}}\right]$, when variance is proportional to cube of mean], reciprocal $\left[\frac{1}{X}\right]$, when variance is proportional to fourth power of mean] and tangent hyperbolic transformation.

The transformation discussed above are a particular case of the general family of transformations known as Box-Cox transformation.

d) Box-Cox Transformation

By now we know that if the relation between the variance of observations and the mean is known then this information can be utilized in selecting the form of the transformation. We now elaborate on this point and show how it is possible to estimate the form of the required transformation from the data. The transformation suggested by Box and Cox (1964) is a power transformation of the original data. Let y_{ut} be the observation pertaining to the u^{th} plot; then the power transformation implies that we use y_{ut}'s as

$$y_{ut}^{*} = y_{ut}^{\lambda}.$$

The transformation parameter λ in $y_{ut}^{*} = y_{ut}^{\lambda}$ may be estimated simultaneously with the other model parameters (overall mean and treatment effects) using the method of maximum likelihood. The procedure consists of performing, for the various values of λ, a standard analysis of variance on

$$y_{ut}^{(\lambda)} = \begin{cases} \dfrac{y_{ut}^{\lambda} - 1}{\lambda \dot{y}_{ut}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y}_{ut} \ln y_{ut} & \lambda = 0 \end{cases} \qquad \text{(A)}$$

where $\dot{y}_{ut} = \ln^{-1}\left[(1/n)\sum_{u=1}^{N}\sum_{t=1}^{n_u} \ln y_{ut}\right]$.

y_{ut}^{*} is the geometric mean of the observations. The maximum likelihood estimate of λ is the value for which the error sum of squares, say SSE (λ), is minimum. Notice that we cannot select the value of λ by directly comparing the error sum of squares from analysis of variance on y^{λ} because for each value of l the error sum of squares is measured on a different scale. Equation (A) rescales the responses so that the error sums of squares are directly comparable. This is a very general transformation and the commonly used transformations follow as particular cases. The particular cases for different values of are given below.

λ	Transformation
1	No Transformation
½	Square Root
0	Log
-1/2	Reciprocal Square Root
-1	Reciprocal

Remark 3: If any one of the observations is zero then the geometric mean is undefined. In the expression (A), geometric mean is in denominator so it is not possible to compute that expression. For solving this problem, we add a small quantity to each of the observations.

Note: It should be emphasized that transformation, if needed, must take place right at the beginning of the analysis, all fitting of missing plot values, all adjustments by covariance *etc*. being done with the transformed variate and not with the original data. At the end, when the conclusions have been reached, it is permissible to 're-transform' the results so as to present them in the original units of measurement, but this is done only to render them more intelligible.

As a result of this transformation followed by back transformation, the means will rather be different from those that would have been obtained from the original data. A simple example is that without transformation, the mean of the numbers 1, 4, 9, 16 and 25 is 11. Suppose a square root transformation is used to give 1, 2, 3, 4 and 5, the mean is now 3, which after back- transformation gives 9. Usually the difference will not be so great because data do not usually vary as much as those given, but logarithmic and square root transformation always lead to a reduction of the mean, just as angles of equal formation usually lead to its moving away from the central value of 50%.

However, in practice, computing treatment means from original data is more frequently used because of its simplicity, but this may change the order of ranking of converted means for comparison. Although transformations make possible a valid analysis, they can be very awkward. For example, although a significant difference can be worked out in the usual way for means of the transformed data, none can be worked out for the treatment means after back transformation.

Non-parametric tests in the Analysis of Experimental Data

When the data remains non-normal and/or heterogeneous even after transformation, a recourse is made to non-parametric test procedures. A lot of attention is being paid to develop non-parametric tests for analysis of experimental data. Most of these non-parametric test procedures are based on rank statistic. The rank statistic has been used in development of these tests as the statistic based on ranks is

1. distribution free
2. easy to calculate and
3. simple to explain and understand.

Another reason for use of rank statistic is due to the well known result that the average rank approaches normality quickly as *n* (number of observations) increases, under the rather

general conditions, while the same might not be true for the original data {see e.g. Conover and Iman (1976, 1981)}. The non-parametric test procedures available in literature cover completely randomized designs, randomized complete block designs, balanced incomplete block designs, design for bioassays, split plot designs, cross-over designs and so on. For an excellent and elaborate discussions on non-parametric tests in the analysis of experimental data, one may refer to Siegel and Castellan Jr. (1988), Deshpande, Gore and Shanubhogue (1995), Sen (1996), and Hollander and Wolfe (1999).

Kruskal-Wallis Test can be used for the analysis of data from completely randomized designs. Skillings and Mack Test helps in analyzing the data from a general block design. Friedman Test and Durbin Test are particular cases of this test. Friedman Test is used for the analysis of data from randomized complete block designs and Durbin test for the analysis of data from balanced incomplete block designs.

References

Anderson, V. L. and McLean, R. A. (1974). *Design of Experiments: A realistic approach*. Marcel Dekker Inc., New York.

Bartlett, M. S. (1947). The use of transformation. *Biometrics*, 3, 39-52.

Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69, 364-367.

Box, G. E. P., (1953). Non-normality and tests on variances. *Biometrika*, 40 (1 & 2), 318-335.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformation. *J. Roy. Stastist. Soc.* B, 26, 211-252.

Conover, W. J. and Iman, R. L. (1976). In Some Alternative Procedure Using Ranks for the Analysis of Experimental Designs. *Comm. Statist.: Theory & Methods,* A5(14), 1349-1368.

Conover, W. J. and Iman, R. L. (1981). Rank transfomations as a bridge between parametric and non-parametric statistics (with discussion). *American Statistician*, 35, 124-133.

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-fit Techniques*. Marcel Dekkar, Inc., New York.

Deshpande, J. V., Gore, A. P. and Shanubhogue, A. (1995). *Statistical Analysis of Non-normal Data*. Wiley Eastern Limited, New Delhi.

Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments*. Springer, New York.

Dolby, J. L. (1963). A quick method for choosing a transformation. *Technometrics*, 5, 317-326.

Draper, N. R. and Hunter, W. G. (1969). Transformations: Some examples revisited. *Technometrics*, 11, 23-40.

Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods*, 2nd Ed, John Wiley, New York.

Levene, H., (1960). Robust Tests for Equality of Variances. I Olkin, ed., *Contributions to Probability and Statistics*, Palo Alto, Calif.: Stanford University Press, 278-292.

Royston, P. (1992). Approximating the Shapiro-Wilk W-Test for non-normality. *Statistics and Computing*, **2**, 117 -119.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variances test for normality (Complete Samples). *Biometrika*, 52, 591-611

Siegel, S. and Catellan, N. J. Jr. (1988*). Non parametric statistics for the Behavioural Sciences*, 2nd Ed., McGraw Hill, New York.

Tukey, J. W. (1949). One degree of freedom for non-additivity. *Biometrics,* 5, 232-242.