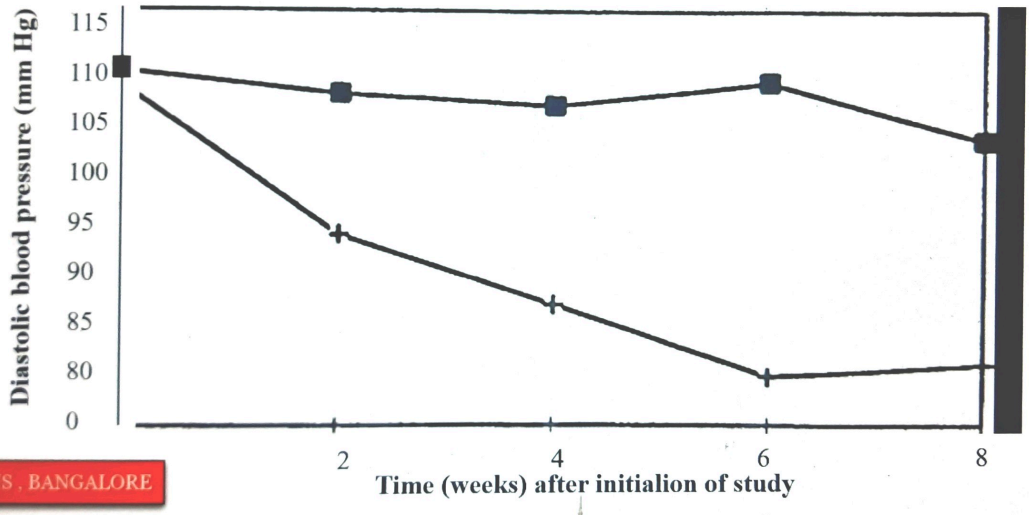
Proper construction and labeling of the typical rectilinear graph should include the following considerations:

1. A title should be given. The title should be brief and to the point, enabling the reader to understand the purpose of the graph without having to resort to reading the text. The title can be placed below or above the graph.

2. The axes should be clearly delineated and labeled. In general, the zero (0) points of both axes should be clearly indicated. The ordinate (the Y axis) is usually labeled with the description parallel to the Y axis. Both the ordinate and abscissa (X axis) should be each appropriately labeled and subdivided in units of equal width (of course, the X and Y axes almost always have different subdivisions).



Blood pressure as a function of time in a clinical study comparing drug and placebo with a regimen of one tablet per day. ■ placebo (average of 45 patients); +, drug (average of 50 patients).

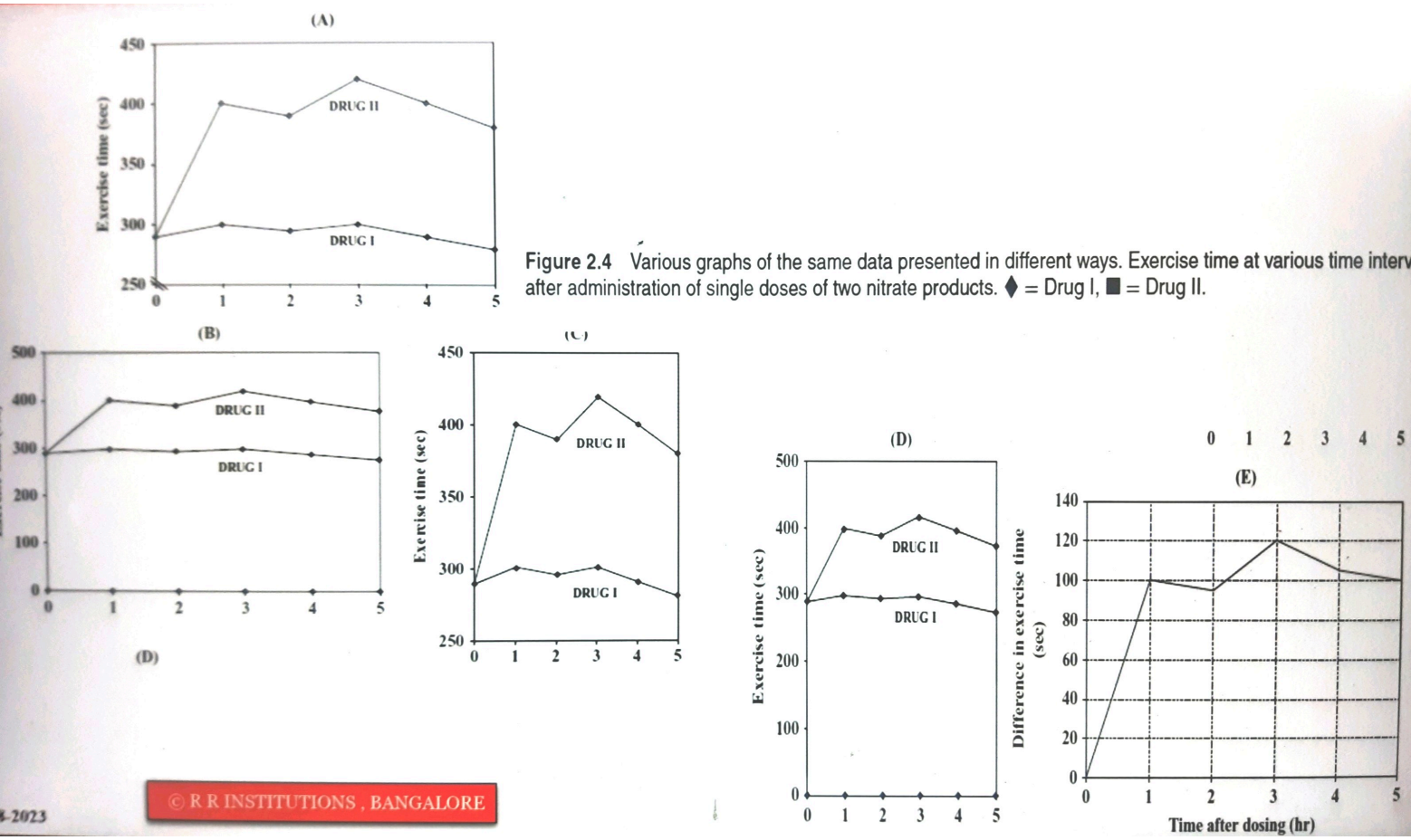


Figure 2.4 Various graphs of the same data presented in different ways. Exercise time at various time intervals after administration of single doses of two nitrate products. ♦ = Drug I, ■ = Drug II.

In the example ,note the units of mmHg and weeks for the ordinate and abscissa, respectively. Grid lines (E) may be added but, if used, should be kept to a minimum, not be prominent and should not interfere with the interpretation of the figure.

3. The numerical values assigned to the axes should be appropriately spaced so as to nicely cover the extent of the graph. This can easily be accomplished by trial and error and a little manipulation. The scales and proportions should be constructed to present a fair picture of the results and should not be exaggerated so to prejudice the interpretation. Sometimes, it may be necessary to skip or omit some of the data to achieve this objective. In these cases, the use of a "broken line" is recommended to clearly indicate the range of data not included in the graph (previous slide).

4. If appropriate, a key explaining the symbols used in the graph should be used. For example, at the bottom of Figure in (slide number 8) the key defines ■ as the symbol for placebo and + for drug. In many cases, labeling the curves directly on the graph (Fig) results in more clarity.

5. In situations where the graph is derived from laboratory data, inclusion of the source of the data (name, laboratory notebook number, and page number, for example) is recommended.

6. If more than one curve appears on the same graph, a convenient way to differentiate the curves is to use different symbols for the experimental points (e.g., ◦, ×, , , +) and, if necessary, connecting the points in different ways (e.g., —.—.—., ..., – .–.–.–). A key or label is used, which is helpful in distinguishing the various curves, as shown in Figures. Other ways of differentiating curves include different kinds of crosshatching and use of different colors.

7. One should take care not to place too many curves on the same graph, as this can result in confusion. There are no specific rules in this regard. The decision depends on the nature of the data, and how the data look when they are plotted. The curves graphed in below diagram are cluttered and confusing. The curves should be presented differently or separated into two or more graphs. Figure in next slide is a clearer depiction of the dissolution results of the five formulations shown in below diagram.

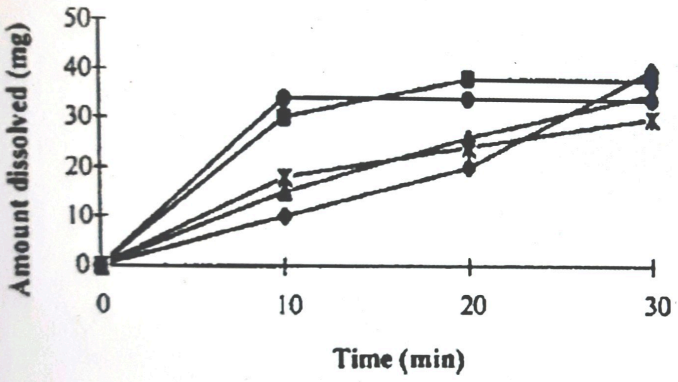


Figure 2.7 Plot of dissolution time of five different commercial formulations of the same drug. • = product A, ■ = product B, × = product C, ▲ = product D, ♦ = product E.

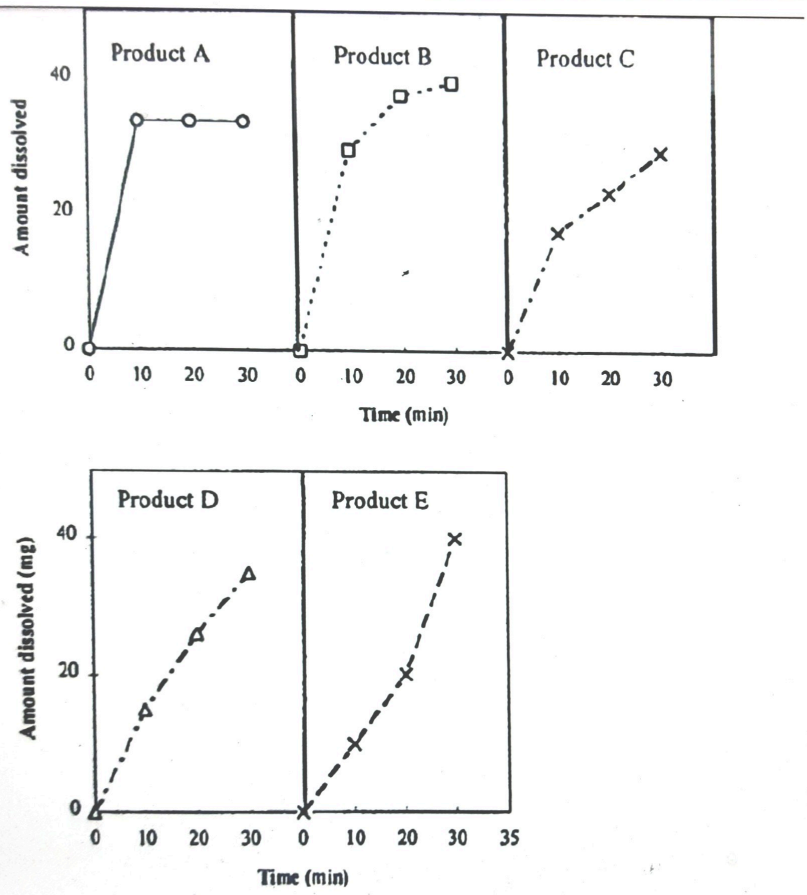


Figure 2.8 Individual plots of dissolution of the five formulations shown in Fig. 2.7.

8. The standard deviation may be indicated on graphs as shown in below figure . However, when the standard deviation is indicated on a graph (or in a table, for that matter), it should be made clear whether the variation described in the graph is an indication of the ***standard deviation (S) or the standard deviation of the mean ($S^{-}x$).*** The standard deviation of the mean, if appropriate, is often preferable to the standard deviation not only because the values on the graph are mean values, but also because $S^{-}x$ is smaller than the s.d., and therefore less cluttering. Overlapping standard deviations, as shown in Figure (next slide), should be avoided, as this representation of the experimental results is usually more confusing than clarifying.

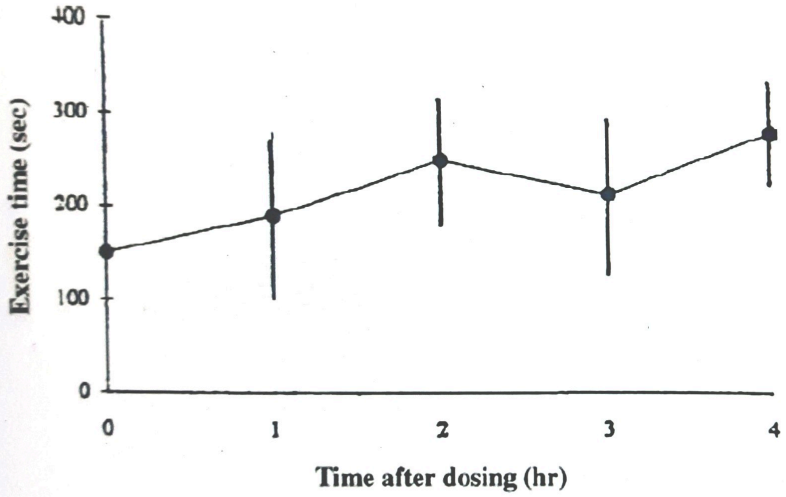


Figure 2.9 Plot of exercise time as a function of time for an antianginal drug showing mean values and standard error of the mean.

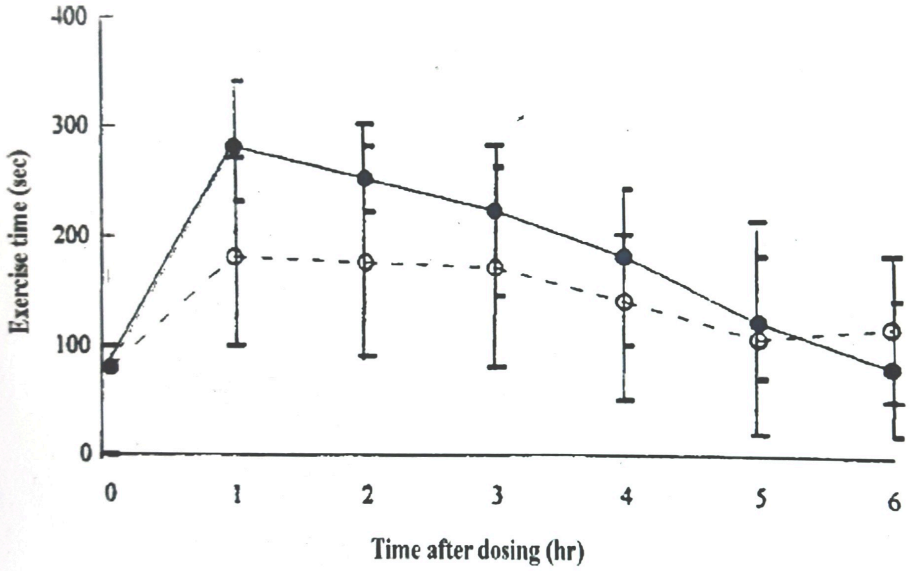


Figure 2.10 Graph comparing two antianginal drugs that is confusing and cluttered because of the overlapping standard deviations. •, drug A; o, drug B.

9. The manner in which the points on a graph should be connected is not always obvious. Should the individual points be connected by straight lines, or should a smooth curve that approximates the points be drawn through the data. . (See this slide) If the graphs represent functional relationships, the data should probably be connected by a smooth curve. ***For example, the blood level versus time data shown in Figure*** are described most accurately by a smooth curve. Although, theoretically, the points should not be connected by straight lines as shown in Figure (A) , such graphs are often depicted this way. Connecting the individual points with straight lines may be considered acceptable if one recognizes that this representation is meant to clarify the graphical presentation, or is done for some other appropriate reason. ***In the blood-level example, the area under the curve is proportional to the amount of drug absorbed***. The area is often computed by the trapezoidal rule and depiction of the data as shown in Figure (A) makes it easier to visualize and perform such calculations.

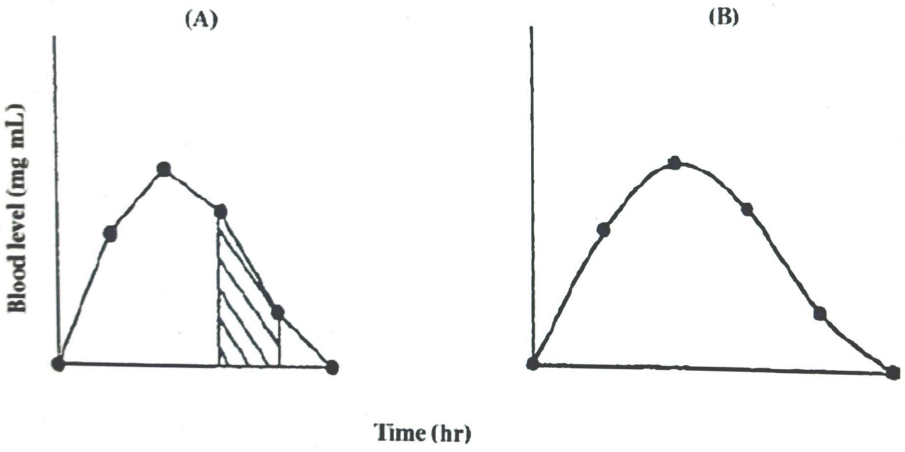
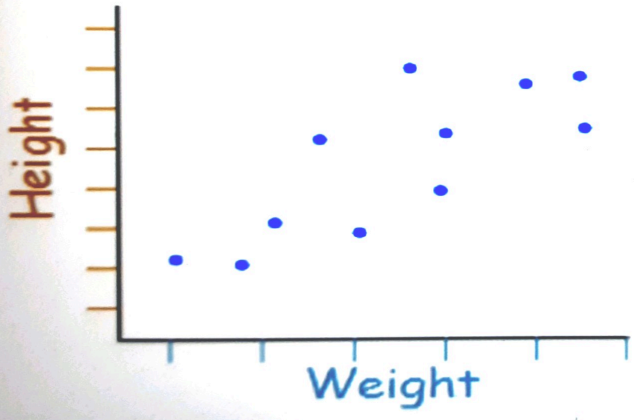


Figure 2.11 Plot of blood level versus time data illustrating two ways of drawing the curves.

SCATTER PLOTS (CORRELATION DIAGRAMS) A graph of plotted points that show the relationship between two sets of data.

Scatter plots (also called correlation diagrams or scatter diagrams) at this time. This type of plot or diagram is commonly used when presenting results of experiments. ***A scatter diagram is a graphic representation of the relationship between two variables. Scatter diagrams help teams identify and understand cause-effect relationships***.

A typical scatter plot is illustrated in Figure .Data are collected in pairs (X and Y) with the objective of demonstrating a trend or relationship (or lack of relationship) between the X and Y variables. Usually, we are interested in showing a linear relationship between the variables (i.e., a straight line). For example, one may be interested in demonstrating a relationship (or correlation) between time to 80% dissolution of various tablet formulations of a particular drug.



In this example, each dot represents one person's weight versus their height.

and the fraction of the dose absorbed when human subjects take the various tablets. The data plotted in Figure show pictorially that as dissolution increases (i.e., the time to 80% dissolution decreases) in vivo absorption increases. Scatter plots involve data pairs, X and Y, both of which are variable. In this example, dissolution time and fraction absorbed are both random variables.

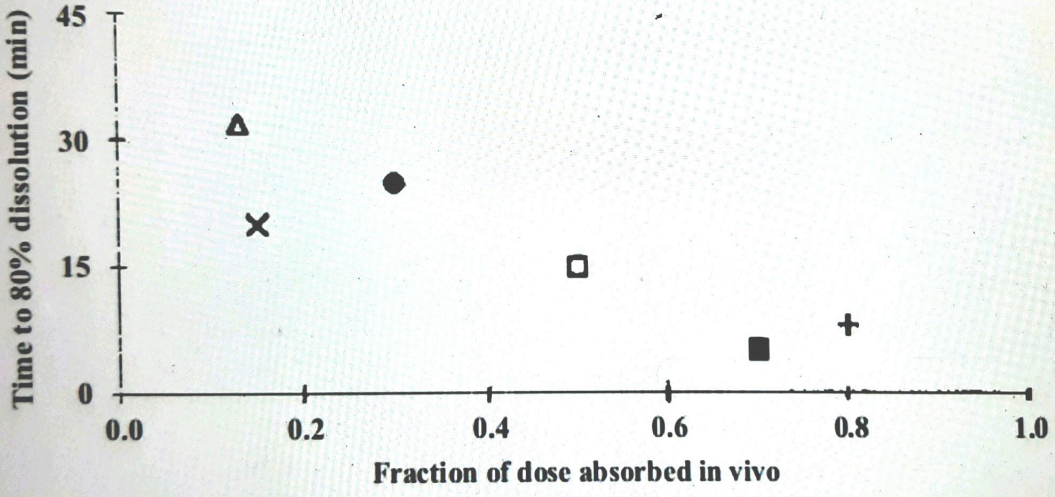


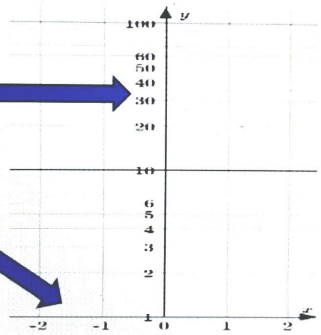
Figure 2.13 Scatter plot showing the correlation of dissolution time and in vivo absorption of six tablet formulations. △, formulation A; ×, formulation B; •, formulation C; □, formulation D; ■, formulation E; +, formulation F.

SEMI LOGARATHIMIC PLOT

A semi logarithmic plot is a graphical representation of a time series. It is defined by an arithmetic scale of time t plotted on the abscissa axis and a logarithmic scale plotted on the ordinate axis, meaning that the logarithm of the observed value Yt will be transcribed in ordinate on an arithmetic scale. In general, printed semi logarithmic sheets of paper are used. In a semi logarithmic graph, one axis has a logarithmic scale and the other axis has a linear scale.

Semi-Logarithmic Graphs

- In the following set of axes, the vertical scale is **logarithmic** (equal scale between powers of 10) and the horizontal scale is **linear** (even spaces between numbers).

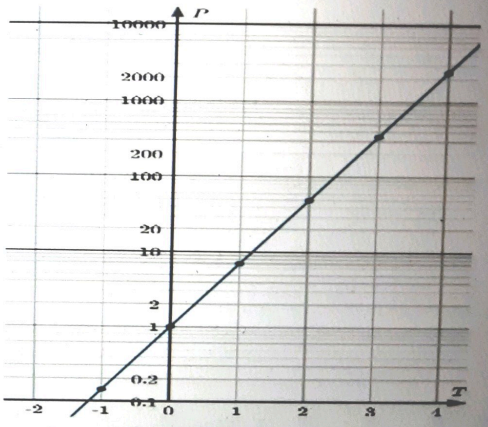


Semilogarithmic axes.

There are no negative numbers on the y-axis, since we can only find the logarithm of positive numbers.

37

Semi-logarithmic axes



Graph of $P = 7^T$ on semilogarithmic axes.

43

ANOVA TEST

- An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.

- Basically, you're testing groups to see if there's a difference between them. Examples of when you might want to test different groups:

- A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.

- A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.

- What Does "One-Way" or "Two-Way Mean?
- One-way or two-way refers to the number of independent variables (IVs) in your Analysis of Variance test.

- One-way has one independent variable (with 2 levels). For example: brand of cereal,
- Two-way has two independent variables (it can have multiple levels). For example: brand of cereal, calorie

Types of Tests.

There are two main types: one-way and two-way. Two-way tests can be with or without replication.

One-way ANOVA between groups: used when you want to test two groups to see if there's a difference between them. Two way ANOVA without replication: used when you have one group and you're double-testing that same group. For example, you're testing one set of individuals before and after they take a medication to see if it works or not. Two way ANOVA with replication: Two groups, and the members of those groups are doing more than one thing. For example, two groups of patients from different hospitals trying two different therapies.

One Way ANOVA

A one way ANOVA is used to compare two means from two independent (unrelated) groups using the F-distribution. The null hypothesis for the test is that the two means are equal. Therefore, a significant result means that the two means are unequal.

Examples of when to use a one way ANOVA

Situation 1: You have a group of individuals randomly split into smaller groups and completing different tasks. For example, you might be studying the effects of tea on weight loss and form three groups: green tea, black tea, and no tea.

Situation 2: Similar to situation 1, but in this case the individuals are split into groups based on an attribute they possess. For example, you might be studying leg strength of people according to weight. You could split participants into weight categories (obese, overweight and normal) and measure their leg strength on a weight machine

Limitations of the One Way ANOVA

A one way ANOVA will tell you that at least two groups were different from each other. But it won't tell you which groups were different. If your test returns a significant f-statistic, you may need to run an ad hoc test (like the Least Significant Difference test) to tell you exactly which groups had a difference in means.

Back to Top

Two Way ANOVA

A Two Way ANOVA is an extension of the One Way ANOVA. With a One Way, you have one independent variable affecting a dependent variable. With a Two Way ANOVA, there are two independents. Use a two way ANOVA when you have one measurement variable (i.e. a quantitative variable) and two nominal variables. In other words, if your experiment has a quantitative outcome and you have two categorical explanatory variables, a two way ANOVA is appropriate.

For example, you might want to find out if there is an interaction between income and gender for anxiety level at job interviews. The anxiety level is the outcome, or the variable that can be measured. Gender and Income are the two categorical variables. These categorical variables are also the independent variables, which are called factors in a Two Way ANOVA.

The factors can be split into levels. In the above example, income level could be split into three levels: low, middle and high income. Gender could be split into three levels: male, female, and transgender. Treatment groups are all possible combinations of the factors. In this example there would be 3 x 3 = 9 treatment groups.

14, Why ANOVA is preffered over student-t-test :-

1) ANOVA is an observable technique used to compare the means of more than two population groups.

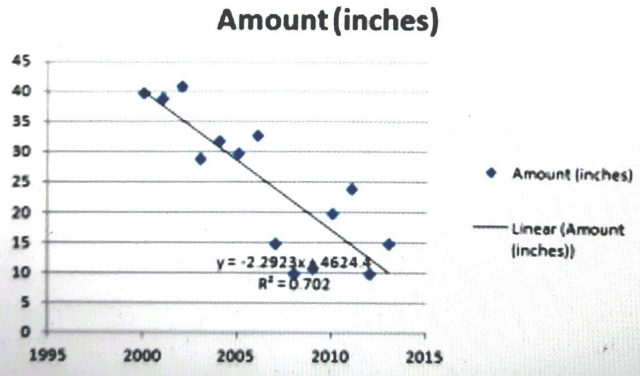
2) ANOVA equates three or more such groups.

3) ANOVA is one-sided test due to no negative variance.

4) ANOVA is used for huge population counts.

LINEAR REGRESSION AND CORRELATION

Regression analysis is used in stats to find trends in data. For example, you might guess that there's a connection between how much you eat and how much you weigh; regression analysis can help you quantify that. Regression analysis will provide you with an equation for a graph so that you can make predictions about your data. For example, if you've been putting on weight over the last few years, it can predict how much you'll weigh in ten years time if you continue to put on weight at the same rate. It will also give you a slew of statistics (including a p-value and a correlation coefficient) to tell you how accurate your model is. Most elementary stats courses cover very basic techniques, like making scatter plots and performing linear regression. However, you may come across more advanced techniques like multiple regression.



Definition

- *The correlation is the measure of the extent and the direction of the relationship between two variables in a bivariate distribution.*

Example:

(i) Height and weight of children.

(ii) An increase in the price of the commodity by a decrease in the quantity demanded.

Types of Correlation: The following are the types of correlation

(i) Positive and Negative Correlation

(ii) Simple, Partial and Multiple Correlation

(iii) Linear and Non-linear Correlation

Type of correlation coefficient

1. Perfect ***Positive correlation***
2. Perfect ***negative correlation***
3. ***Moderately*** Positive correlation
4. ***Moderate*** negative correlation
5. Absolute ***no correlation***

SPEARMANN'S CO RELATION

In statistics, Spearman's rank correlation coefficient or Spearman's ρ, named after Charles Spearman and often denoted by the Greek letter (rho) or as is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic function.

The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

Intuitively, the Spearman correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) rank (i.e. relative position label of the observations within the variable: 1st, 2nd, 3rd, etc.) between the two variables, and low when observations have a dissimilar (or fully opposed for a correlation of −1) rank between the two variables.

Spearman's coefficient is appropriate for both continuous and discrete ordinal variables. Both Spearman's rho and Kendall's tau can be formulated as special cases of a more general correlation coefficient.

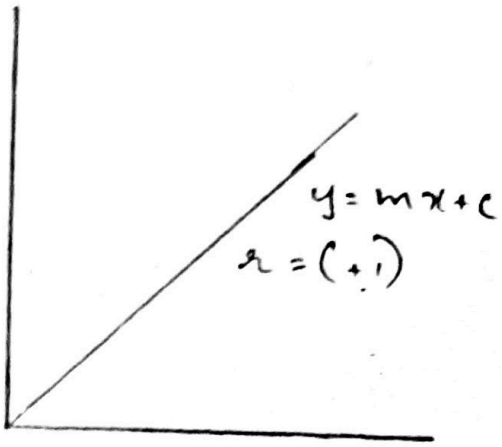
The Spearman rank correlation coefficient, rs, is the nonparametric version of the Pearson correlation coefficient. Your data must be ordinal, interval or ratio. Spearman's returns a value from -1 to 1, where:
+1 = a perfect positive correlation between ranks
-1 = a perfect negative correlation between ranks
0 = no correlation between ranks.

The formula for the Spearman rank correlation coefficient when there are no tied ranks is:

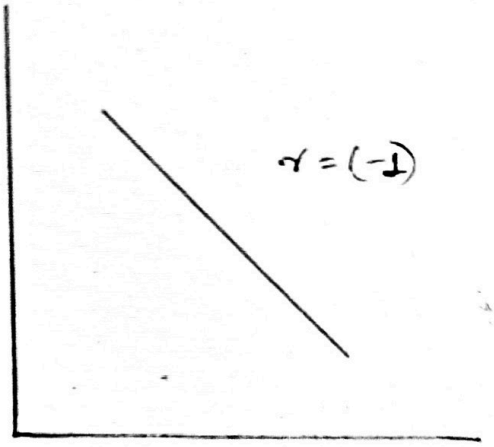
$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

R values of Regression

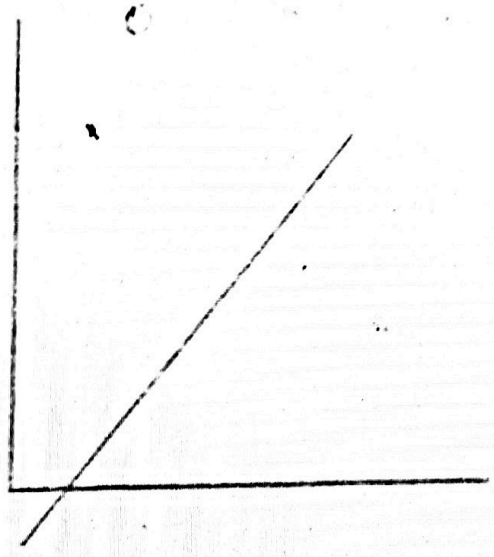
- Positive Co-relation



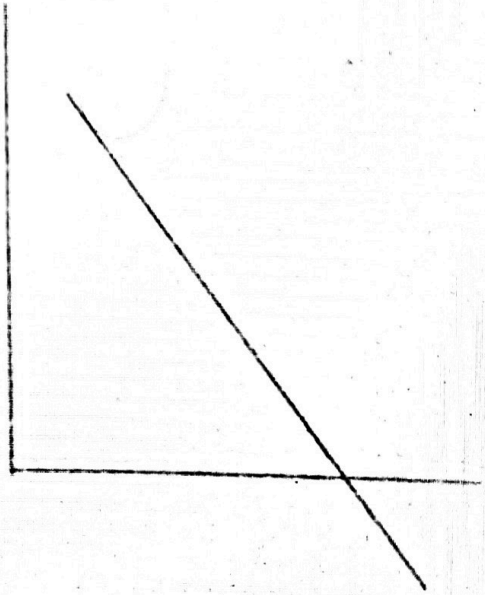
- Negative Correlatio



- Moderately Positive



- Moderately negative



- No Co-relation

$r = 0$