APPENDIX-2
STATISTICAL ANALYSIS

The petroleum engineer whether he/she is engaged in drilling, production, reservoir engineering or economics, each day is confronted with data. These data are the only facts the engineer has about the environment he is trying to describe, understand and control, consequently it is essential for him to evaluate his data before he proceeds to draw conclusions and act upon them.

Since the petroleum engineer is faced with the problem of drawing conclusions, under uncertainty, he needs every tool he can get to increase his confidence, or at least to measure his uncertainty. Important decisions often involve so many complicated factors that a complete analysis is not practical or even possible. Statistical methods can help to make the best "educated guess". All engineers are engaged in statistical analysis, though they don't think of themselves or statisticians. But each time they take the average set of number or draw a curve through plotted points, they are determining the central tendency of data. The average itself is a **statistic**, as statistic is, by definition ant value computed entirely from a sample. Statistics is the study of how to collect, organize, analyze and interpret the numerical information (data). The principal function of statistics is to describe certain characteristics of mass phenomena and repetitive events. Generally speaking, statistics can be thought of as a combination of two sub-fields called *descriptive statistics* and *statistical inference*. Descriptive statistics deals with the collection, tabulation and presentation of data and the calculation of measures, which describe the data in various ways. If the data are simply presented and described invariants informative ways, then the descriptive statistics is being utilized. If on the other hand, the data and the descriptions are used as a basis for making predictions or decisions relating to unobserved events (such as future events), then statistical inference is involved.

Data Significance and Error Analysis

Anyone who has done experimental work knows that when he makes N independent measurements of a variable, he can not expect to get N identical values. The values will vary from one measurement to the next randomly and most of the time they will be within certain limits depending upon the precision of the measuring device and the skill of its operation. Engineering applications require that principal features of a distribution be described numerically. Mainly we are interested in:

-measure of location (central tendency)

-measure of spread (standard deviation)

Measures of Central Tendency

Mean: It is the arithmetic average of the given set of observations.

$$\mathbf{X = \sum_{i=1}^{N} X_i / N = (sum\ of\ all\ observations) / (total\ number\ of\ observations)}$$

Median: Median is the valueoccurring midway in the data when the data are arranged in order of value (ascending or descending)

Mode: It is the value of the variable which occurs most frequently.

Variation: Variation is a characteristics of all data. Experience tells us, that, degree of variation will determine the measure of confidence we will have in any averages.

There are several measures of dispersion as range and standard deviation.

Range: It is the simplest of all measures of dispersion. The difference between the highest and lowest values of the variable in a sample. Range indicates the variation between the smallest and largest entries, but it does not tell us how much other values vary from one another.

Standard Deviation: It is a measurement that will give us a better idea of how the data entries

$$S = SQRT \left[\sum_{i=1}^{N} (X_i - X)^2 \right] / (N-1)$$

Example:

Bit No	X, footage drilled	X_i-X	$(X_i-X)^2$
1	69	-24.3	592.1
2	72	-21.33	454.9
3	89	-4.33	18.74
4	135	41.67	1736.38
5	72	-21.33	454.9
6	123	29.67	880.3

Mean = [(69+72+89+135+72+123) / 6] = **93.33**

Mode = **72**

Median = **89**

Range = 135 – 69 = **66**

Stan. Deviation = SQRT [4137.3 / (6-1)] = **28.7**

Frequency Distribution:

Sometimes it is enlightening to apply a descriptive device to data so that we can see how they vary. This device is called frequency distribution.

Normal Distribution: A special kind of frequency distribution that is very important in statistics is the normal and Gaussian distribution. It forms the basis for most statistical measures probability. Normal distribution is a bell shaped curve called Gaussian curve. The graph of the normal distribution is important because the portion of the area under the curve above a given interval represents the probability that measurements will lie in that interval.