

Addis College

M.Sc : Construction Technology & Management

Course: Research Methods and Statistical Application

Lecture 3: Measurement Techniques

By: Dagnachew A. (PhD)

Contents of the presentation

- Lecture notes
- Individual assignment

Common statistical measures

- Statistics may be classified in several ways.
- One important, basic classification is ‘descriptive statistics’ & ‘analytic statistics’.
- Descriptive statistics are used to express the important features of a population (population *parameters*), *intended sample* or sample obtained (sample *statistics*) *numerically*; they include *percentages* (e.g. response rates) and other numerical descriptors of the distribution (e.g. of the sample of responses) under examination—measures of centrality (means, mode, median) and spread/variability (variance, standard deviation, coefficient of variation, interquartile range, range).
- Shapes of distributions should be described, including measures of skew and kurtosis (‘pointedness’) as appropriate. Use of descriptive statistics does NOT constitute analysis!

- A very common measurement is the *average*; in statistics, there are two important measurements of the average or *mean*:
- the *arithmetic mean*; which most people understand as the average the geometric mean.
- The *arithmetic mean* is expressed mathematically as:

$$(\bar{x}) = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{or} \quad \bar{x} = \frac{\sum x}{n}$$

Where;

n = no of items

x_i = an item of value i where the value of i ranges from 1 to n

x = arithmetic mean

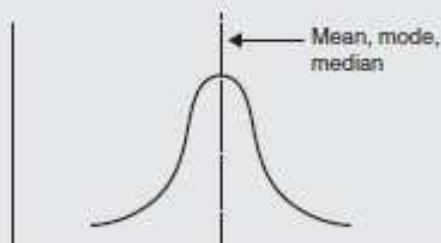
- The geometric mean is calculated as:

$$GM = \sqrt[n]{x_1 x_2 \dots x_n}$$

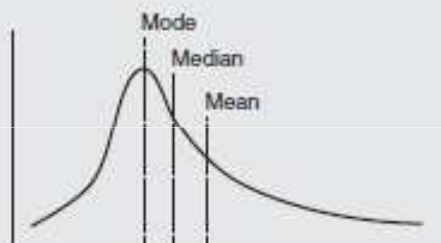
- The *median is the midpoint of a distribution. As extreme or 'outlier' values may distort* the mean, the median may be a preferable indicator of such a distribution's central tendency.
- The *mode is the value which occurs most frequently in a distribution.*
- *A distribution* may have more than one mode.

Example

- (1) Symmetrical distribution (e.g. normal distribution)
Mean, mode and median coincide.

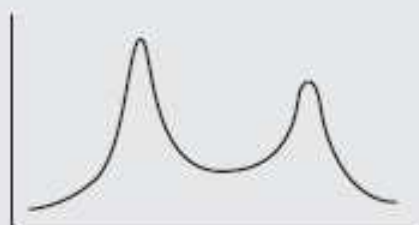


- (2) Skewed distribution
Mean and, to a lesser extent, median move away from the mode in the direction of the skew (long tail).



Note: 'Outlier' points of the distribution affect the location of the mean more than the median and the median more than the mode. Hence, if outliers cannot be disregarded, their effects must be acknowledged.

- (3) Bimodal distribution



Note: The modes need not be of equal size or value.

- The arithmetic mean is the value which, numerically, is the most representative of the series or distribution.
- The geometric mean is used where measurements of relative changes are being considered.
- The geometric mean is used in retail price indices; for such variables, it is usual to be interested in the percentage change between successive periods.

- Samples, and the populations from which samples are drawn and which the samples should represent, each have statistical distributions—commonly assumed to be normal; tests for normality should be applied.
- Parametric statistics depend upon the nature of the distribution from which the data have been obtained.
- However, many populations are structured in a particular way which is important to the research.
- In order to obtain a sample which is representative of the population, the sample should reflect the structure of the population closely.
- The structure of the population yields the sampling frame—the structure used to determine the form of the sample from which the data will be gathered.

- Generally, sample surveys (nomothetic studies) have the objective of either:
- *estimating a parameter in the population*—the estimate made should be accompanied by a note of its precision testing a statistical hypothesis concerning the population—the results of testing will lend support to
- or *lead to rejection of the hypothesis*; such tests require a criterion against which deviation of the result from the sample can be judged against the value hypothesized.

- **Normal distribution**

Much of statistics uses a particular distribution which has been found to occur very commonly—the *normal distribution*. *The normal distribution is a probability density function which has certain, particular properties.*

The main features of the normal distribution are:

- measurements of the random variable, X , are on an interval scale or a ratio scale;
- values of X are symmetrical about the mean, μ ;
- the distribution has a central tendency; that is, the closer values of X are to μ , the more frequently they occur;
- the tails of the distribution spread to infinity.

Thus, the distribution is uniform and ‘bell-shaped’.

- Another important measurement for a normal distribution concerns the variability or spread of the distribution. That measurement is the variance; the square root of the variance is the standard deviation, σ .
- As, in practice, samples are used, the standard deviation of the population must be estimated from that of the sample:

$$\text{Standard deviation of samples} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

Measurements of the spread of a distribution which are simpler than the standard deviation are:

- **Range**—*susceptible to distortion by extreme values (outliers).*
- **Interquartile range**—*the range between the first and third quartiles (i.e. 50% range of the distribution).*
- **Decile range**—*the range of the distribution excluding the lowest and highest 10% of data points (i.e. 80% range of the distribution).*
- The interquartile & decile ranges use the median as the centre of the distribution.
- For comparisons over a number of distributions, it is likely to be convenient to use the *coefficient of variation as the measure of spread.*
- *The coefficient of variation, cv , is calculated as $100(\sigma/\mu)$ using population parameters; or $100(s/x)$ using sample statistics.*

To calculate the level of confidence, the *standard normal variable, z, is employed*:

$$z = \frac{\bar{x} - \mu}{\sigma}$$

The usual levels of confidence are:

- 95% confidence is described as significant (5% significance)
 - 99% confidence is described as highly significant (1% significance)
 - 99.9% confidence is described as very highly significant (0.1% significance).
- Confidence refers to the level of probability that the test results are valid (such as accepting a null hypothesis correctly), that the interval estimate does include the population parameter; while the level of significance (1—the level of confidence) denotes the probability that the null hypothesis is correct but has been rejected incorrectly (a type 1 error).

Null hypotheses

- Another form of hypothesis which is encountered in statistical testing is the *null hypothesis*.
- Commonly, null hypotheses are employed in supplementing the overall hypothesis and any sub-hypotheses and they lend rigour to statistically testing particular, possible relationships between variables.
- A null hypothesis is tested in comparison with its opposite, the alternate hypothesis.
- The usual form is that the null hypothesis speculates that there is no difference between, say, the cost in \$ per kilometer of travel by coach and by train.

The alternate hypothesis speculates that there is a difference. Thus:

$$H_O : \mu_1 = \mu_2$$

$$H_A : \mu_1 \neq \mu_2$$

where:

μ_1 = population mean₁

μ_2 = population mean₂

H_O = null hypothesis

H_A = alternative hypothesis.

- To consider $H_O: \mu_1 = \mu_2$ against $H_A: \mu_1 \neq \mu_2$, again the appropriate level of confidence/significance must be decided (commonly 5%).
- Essentially, the test concerns whether the means of the samples x_1 and x_2 are drawn from populations with the same mean. The means of the populations are unknown. The method of testing requires the sampling distribution of the statistic $(x_1 - x_2)$ to be constructed; this is a normal distribution (Yeomans 1968), the mean of which is zero when $\mu_1 = \mu_2$.

- The method requires all possible means of each sample distribution to be calculated. All possible pairs of the sample means are considered, one mean being subtracted from the other (i.e. $(x_1 - x_2)$). *Then, if the mean of the resultant sampling distribution $(x_1 - x_2)$ is zero, H_0 is accepted.*
- A particular feature of null hypotheses is whether a 'one-tailed' or a 'two- tailed' test is required.
- If the null hypothesis is of the form $H_0: \mu_1 = \mu_2$, *a two-tailed test* is necessary where the alternative hypothesis considers $\neq$, in which both greater than and less than values are considered. However, if the alternate hypothesis is of the form $H_A: \mu_1 < \mu_2$, *a one-tailed (left-tailed, lower-tailed) test is required (if $H_A: \mu_1 > \mu_2$, a one-tailed, right-tailed, upper-tailed test is required).*
- For a two-tailed test, the region for rejection of the null hypothesis is in both tails of the distribution whilst in the one-tailed test, the region for rejection lies in the appropriate of either the lower or upper tails.
- Sometimes, testing a hypothesis yields an incorrect result—such error may occur

Conclusion

- This hypothesis, distinguished from a proposition, and proceeded to discuss the use and testing of any hypothesis and examining how appropriate it is to formulate hypotheses for different types of studies – usually, desirable for quantitative studies (as a focus for the research and its content) but more debatable for qualitative studies (often, the purpose of an exploratory study is to determine hypotheses for subsequent research).
- The essentials of a valid hypothesis were explained. Normally, a hypothesis must be *testable, positive and of constant meaning*.

- Types of research where it is desirable, if not essential, to have a hypothesis were discussed to aid both objectivity and delineation of the parameters of the research. That research rigorously tests the hypothesis, not seeks to prove it, was explained.
- The role of sampling was discussed, and basic statistical measures for distributions (centrality and variability/spread) were explained.
- Several common types of statistical distributions were noted, with particular attention to the properties and common measures regarding the normal distribution.

- A variety of statistical measures were outlined to inform the research design with a view to proper collection and analysis of data.
- The use of null hypotheses was examined to aid formal testing of hypotheses and how such hypotheses assist in determination of the level of confidence in research findings.

Individual Assignment 4 (5%)

1. *What is the difference between mean, median, variance, standard deviation, and Coefficient of variance (2.5%). Describe the uses of each of these statistical measures? (2.5%)*

Note: *All the assignments should be submitted through the class representative all in one **zipped file** through dagnachew2@gmail.com*

Submission: *August 10, 2020.*

Thank you

Your ? Through email