



From the SelectedWorks of D. K. Chaturvedi
Dr.

February 2010

SELECTION CRITERIA TO STATISTICAL MODELS

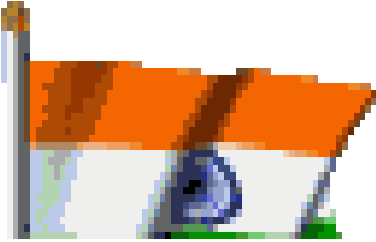
Contact
Author

Start Your Own
SelectedWorks

Notify Me
of New Work

Available at: http://works.bepress.com/dk_chaturvedi/39

SELECTION CRITERIA TO STATISTICAL MODELS AND THEIR COMPARISON



D.K. Chaturvedi

Professor

Dayalbagh Educational Institute

(Deemed University)

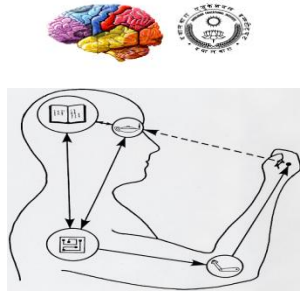
Dayalbagh, Agra

Email: dkc.foe@gmail.com

Web Page: www.works.bepress.com/dk_chaturvedi



Dayalbagh Educational Institute Soft Computing Edge Cutting Tech. Lab (DEISEL)



Soft Computing Group

Professor Incharge

- 4 Teaching Staff members
- 5 Ph.D. Students

- 4 Non-Teaching Staff Members
- 6 M. Tech. Students

Aim

To exploit the tolerance for imprecision uncertainty, approximate reasoning and partial truth to achieve **tractability, robustness, low solution cost, and close resemblance** *with human like decision making* to find an approximate solution to an imprecisely/precisely formulated problem.

Challenge

The challenge is to exploit the tolerance for imprecision by devising methods of computation which lead to an optimal solution at low cost.

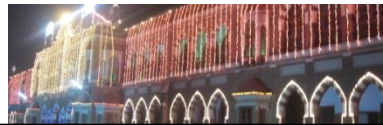
Dayalbagh Educational Institute Soft Computing Edge Cutting Tech. Lab (DEISEL)



Ph.D.



Supporting staff



Colloborations



UNIVERSITY OF
CALGARY



Christian-
Albrechts-
Universität zu
Kiel



Imperial College
London



ADRDE Lab
Agra



IIT, Kanpur



IIT, Rajstan



JMI, Delhi

Work Done in DEISEL

Feed-forward back-propagation ANN

- Developed Product / Sum Type Neuron based ANN
- Study of Aggregation functions on ANN performance

Generalized Neuron and its variations

Benchmark Testing

GN Applications in Machine & Power System

Development of Fuzzy Systems and Evolutionary Algorithms

Integration of ANN/GNN, FS and EA

Quantum Neuron

Outcome from Lab

M.Tech. Dissertations

– 42

Ph.D.

- 8

Books Published

– 4

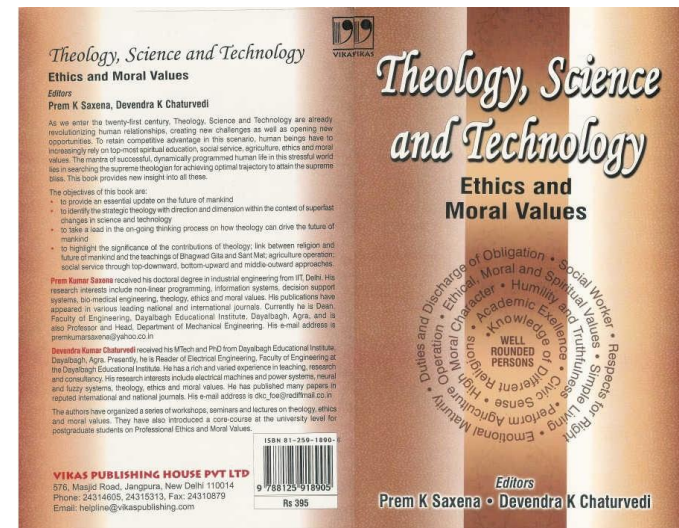
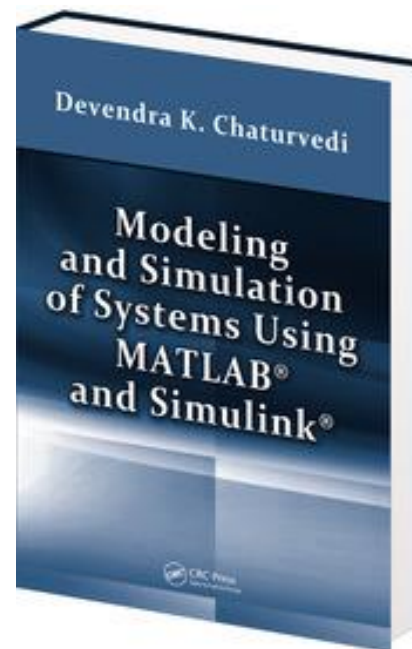
Int. and national research Papers – 140

Reviewed Int. J. Papers

– 45

Funded projects Completed

- 5



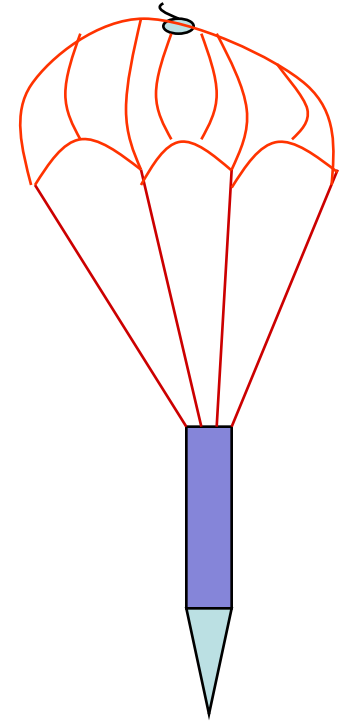
Modeling and Simulation of Parachute.

M.Tech. Dissertations guided

- B.R. Gupta (ADRDE), *An innovative approach for modeling and simulation of parachute inflation and its performance*, 1993.
- Bahadur, K. (ADRDE), *Modelling and Simulation Performance Characteristics of Flexible Aerodynamic Deceleration Device*, Dec., 1993.

Work Published

- Chaturvedi, D.K., and Gupta, B.R., *Simulation of Temperature Variation in Parachute Inflation*, J. of The Institution of Engineers (India), AS, Vol. 76, Sept. 1995, pp. 29-31.
- Chaturvedi, D.K., and Gupta, B.R., *Heat Generation Modelling During Parachute Packing and Deployment*, National Conference on System Design and Simulation (SYDSIM), Agra, April 30 – May 2, 1992.
- Chaturvedi, D.K., Bahadur, K., and Gupta, B.R., *An Innovative Approach for Predicting the Selected Performance Characteristics of Aerodynamics deceleration Device*, Proc. Of National Systems Conference (NSC – 92), pp. 123 – 126, Madras, 1992.
- Chaturvedi, D.K., Gupta, B.R., and Bahadur Kunwer, *Performance Evaluation of Parachute During Inflation*, Proc. Of National Systems Conference (NSC – 94), D.E.I., Dayalbagh, pp. 232, Agra, Jan. 14 – 16, 1995.



Aircraft Landing Control System

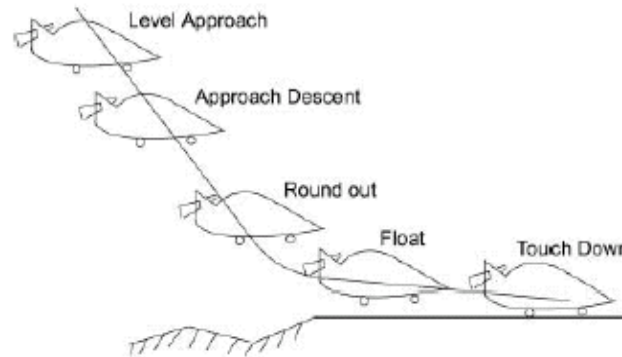


Fig. 1. Final approach extends from line upto touch down

M.Tech. Dissertations Guided

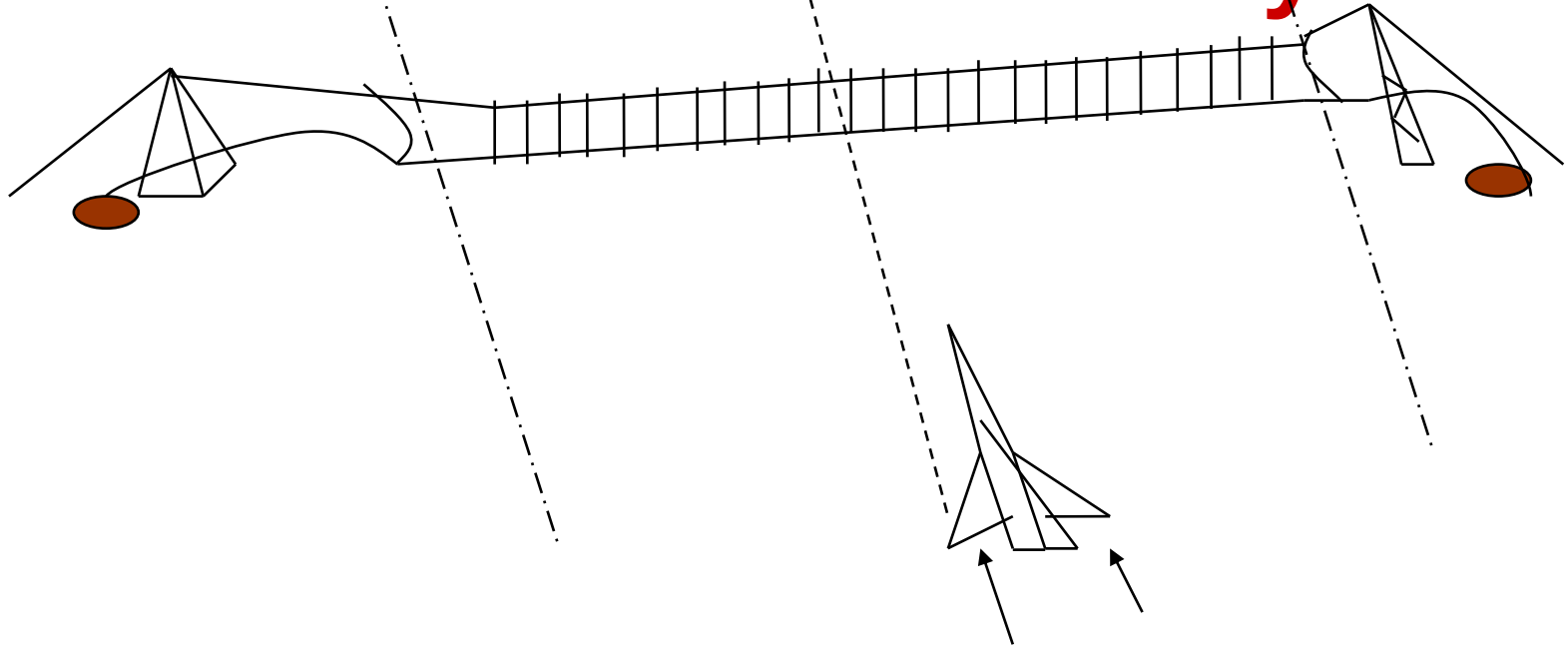
Mahadev B. Alloli, *Intelligent Control System for Aircraft Landing*, April 98.

Ramawadh Chauhan, *Automatic Aircraft Landing Control System Using Neural Networks*, Dec. 1998,

Published

Chaturvedi, D.K., Ramawadh Chauhan & Kalra, P.K, *Applications of Generalised Neural Network for Aircraft Landing Control System*, Int. J. on Soft Computing, Springer- Verlag, Vol. 6, No. 6, Sept. 2002, pp. 441-448.

Aircraft Arrestor Barrier System



M.Tech. Dissertation Guided

R.K. sharma (ADRDE), **Performance characteristics of Aircraft Arrestor Barrier System**", 1994

Published

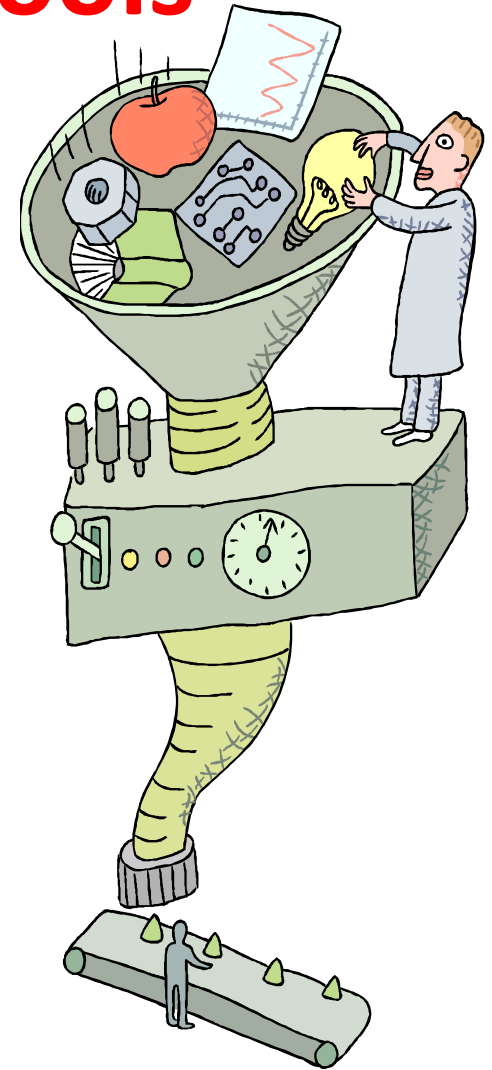
- Chaturvedi, D.K., and Sharma, R.K., **An Experimental Study of Initial Tension of Suspension Strap of Aircraft Arrestor Barrier System**, National Systems Conference, PSG College of Technology, Coimbatore, pp. 474 – 481, Coimbatore, Dec. 14 – 16, 1995.
- Chaturvedi, D.K., and Sharma, R.K., **Modelling and Simulation of Force Generated in Stanchion System of Aircraft Arrestor Barrier System**, Int. J. of Modelling, Measurements, and Control, France, B, Vol. 64, No. 2, 1996, pp.33-51.

STATISTICS

- The systematic and scientific treatment of quantitative measurement is precisely known as statistics.
- Statistics is concerned with collection, classification (or organization) and relationship in data.

Need of Statistical Tools

- Modern problems and needs are forcing statistical methods and statistical ideas more and more to the fore.
- There are many things we wish to know which cannot be discovered by a single observation or be single measurement.



Various statistical Tools

- The statistical tools are commonly used are
 - Average or mean,
 - Error,
 - correlation coefficient,
 - variance and
 - to develop forecasting models etc.



Variables

- **Qualitative**

Color of ball (e.g., red, green, blue) or breed of a dog (e.g., collie, shepherd, terrier)



or

- **Quantitative**

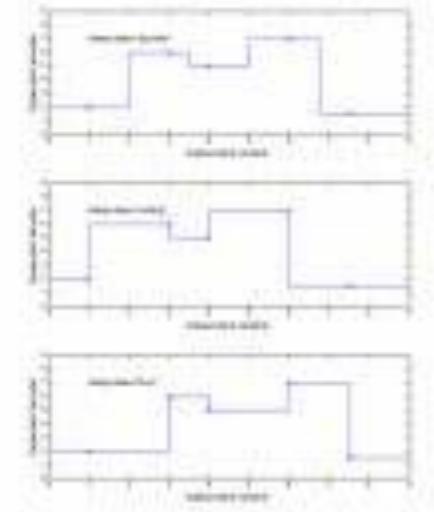
number of people in the city



Variable

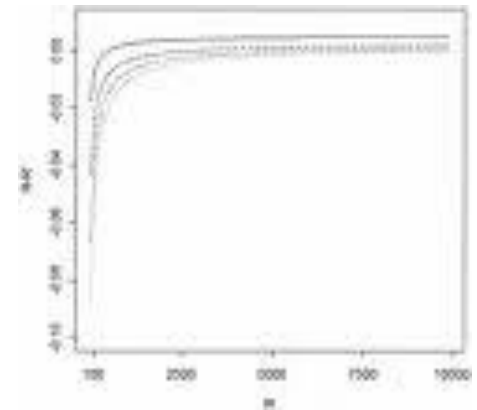
Discrete

flip a coin and count the number of heads



Continuous

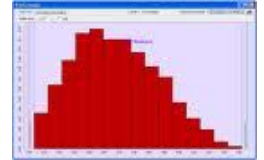
Velocity of vehicle.



Variable

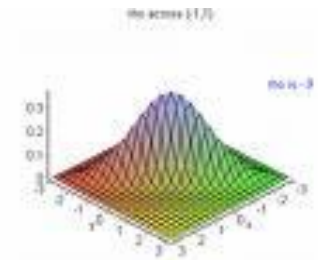
- **Univariate data.**

conducted a survey to estimate the average weight of high school students.



- **Bivariate data.**

examines the relationship between two variables eg. relationship between the height and weight of high school students.



Descriptive Statistics

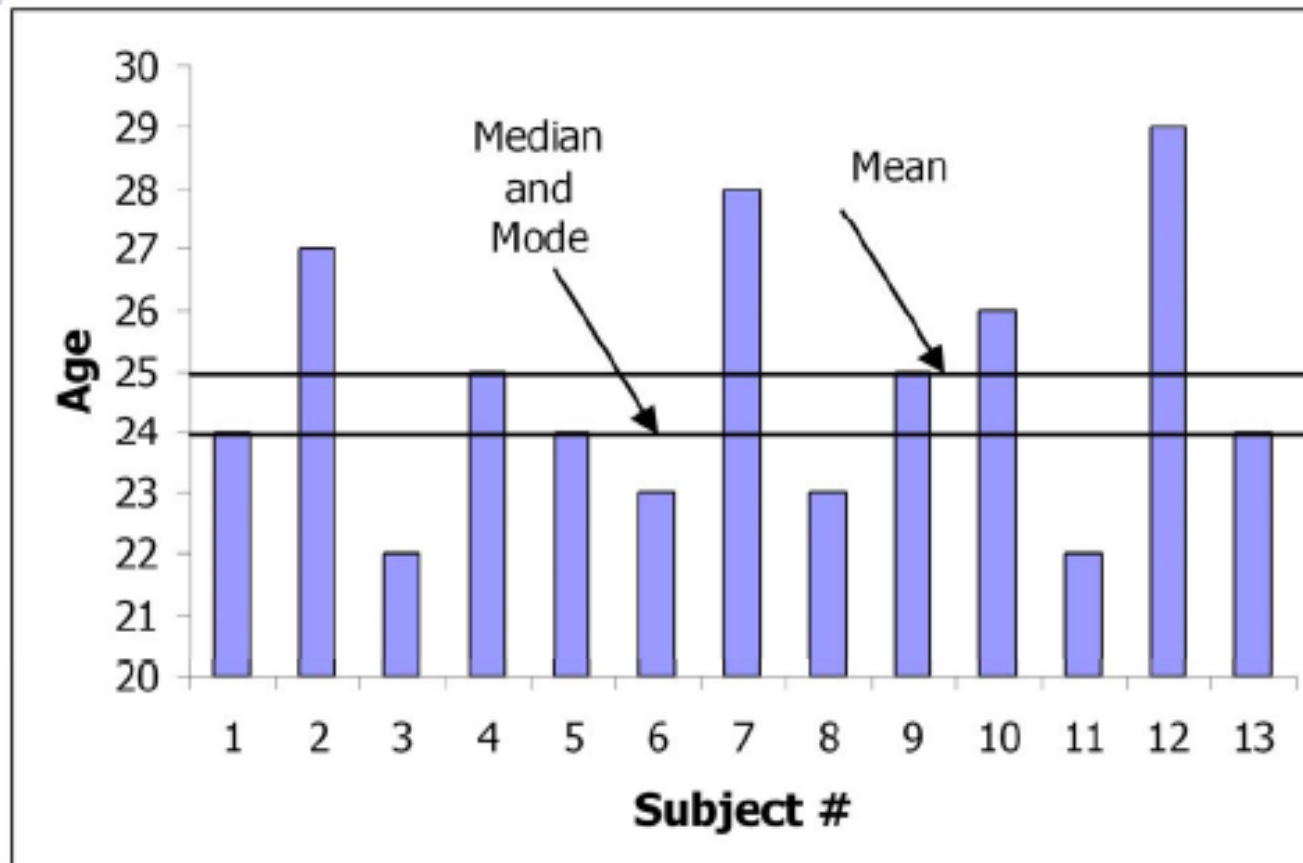
- Measures of Central Tendency
 - Mean (average)
 - Median (middle)
 - Mode (most frequent)
- Measures of Dispersion
 - variance
 - standard deviation
 - standard error
- Measures of Association
 - correlation



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Subject #	Age		mean - add up all ages and divide by the total												
2	1	24		Excel command is =average(b2:b25)												
3	2	27		(24+27+22+25+24+23+28+23+25+26+22+29+24) / 13												
4	3	22		25												
5	4	25														
6	5	24		median - halfway point, equal number of variables on both sides												
7	6	23		Excel command is =median(b2:b23)												
8	7	28		22,22,23,23,24,24,24,25,25,26,27,28,29												
9	8	23		24												
10	9	25														
11	10	26		mode - most frequent												
12	11	22		Excel command is =mode(b2:b23)												
13	12	29		22,22,23,23,24,24,24,25,25,26,27,28,29												
14	13	24		24												

Descriptive Stats

Central Tendency

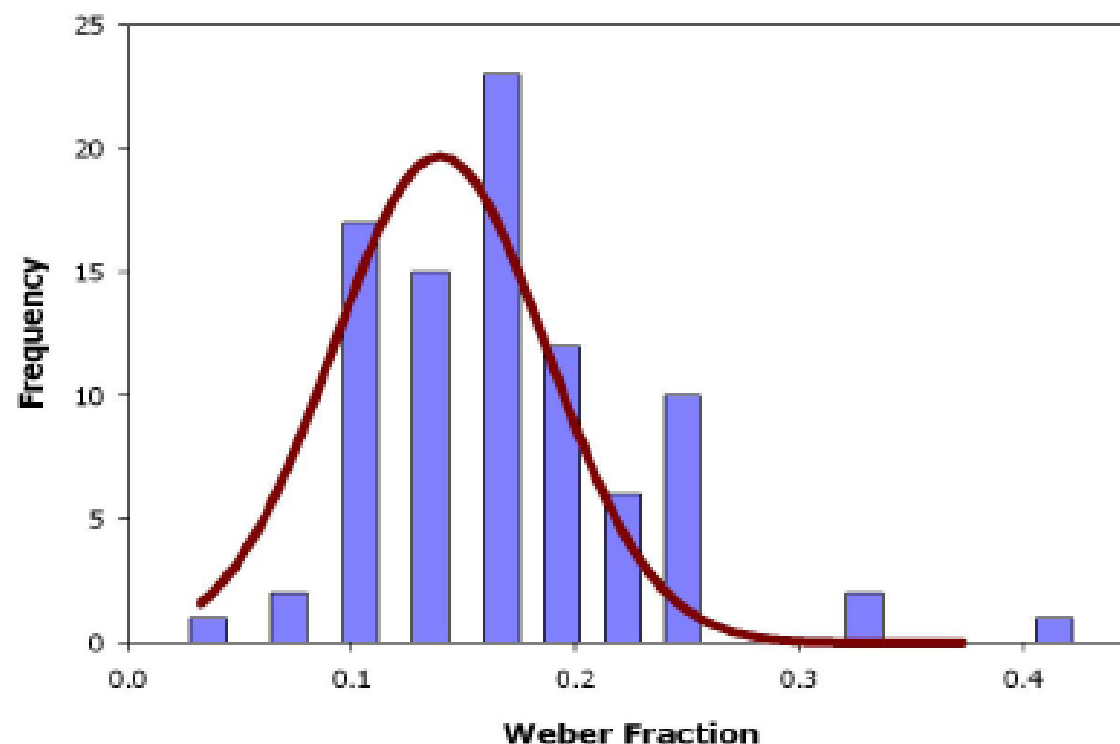


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Subject #	Age		mean - add up all ages and divide by the total												
2	1	24		Excel Command is =average(b2:b14)												
3	2	27		(24+27+22+25+24+23+28+23+25+26+22+29+24)/13 =												
4	3	22		25												
5	4	25														
6	5	24		Standard Deviation - square root of the sum of the squared												
7	6	23		individual differences with the mean divided by												
8	7	28		the total number of data points minus 1.												
9	8	23		S.D. = $\sqrt{[\sum(y_i - y_{\text{mean}})^2 / (N - 1)]}$												
10	9	25		Excel command is = stdev(b2:b14)												
11	10	26		2.2												
12	11	22														
13	12	29		Standard Error - Represents the spread in means if many												
14	13	24		samples of the same size are taken from the population.												
15				S.E. = S.D. / $\sqrt{N}$												

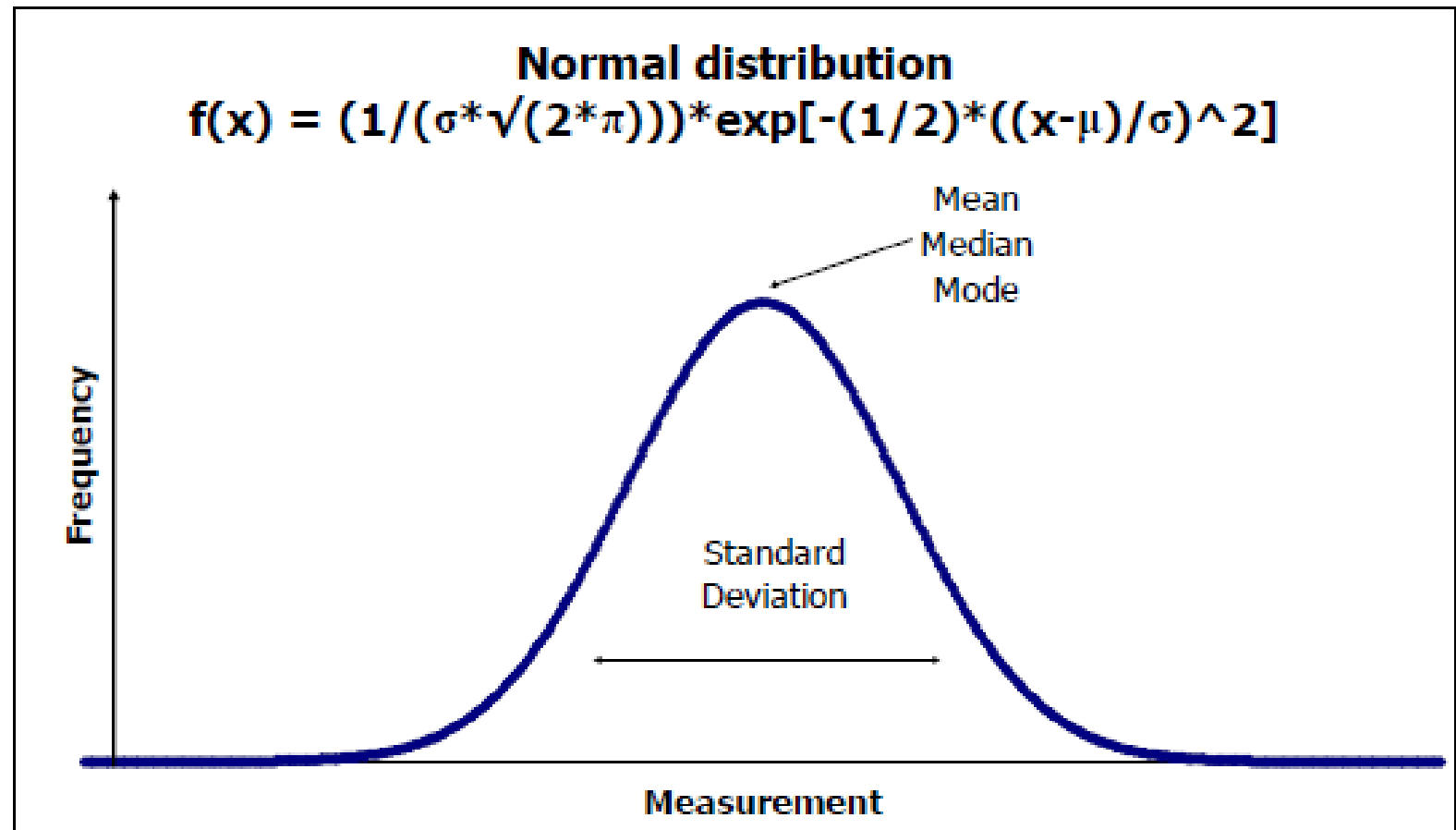
Descriptive Statistics

- Data can usually be characterized by a normal distribution.
- Central tendency is represented by the peak of the distribution.
- Dispersion is represented by the width of the distribution.

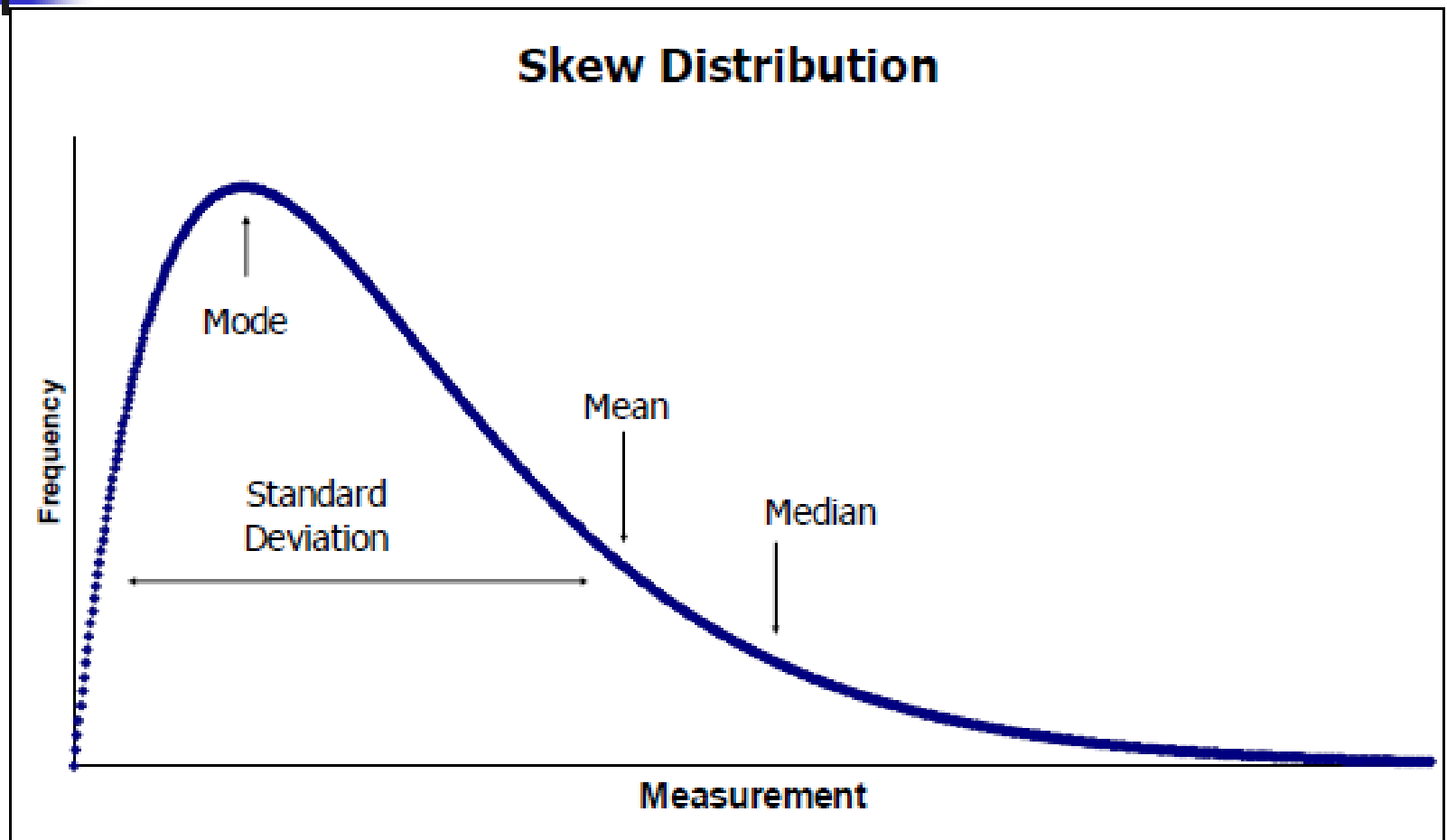
**Mastery Levels
Visit 4**



Descriptive Statistics

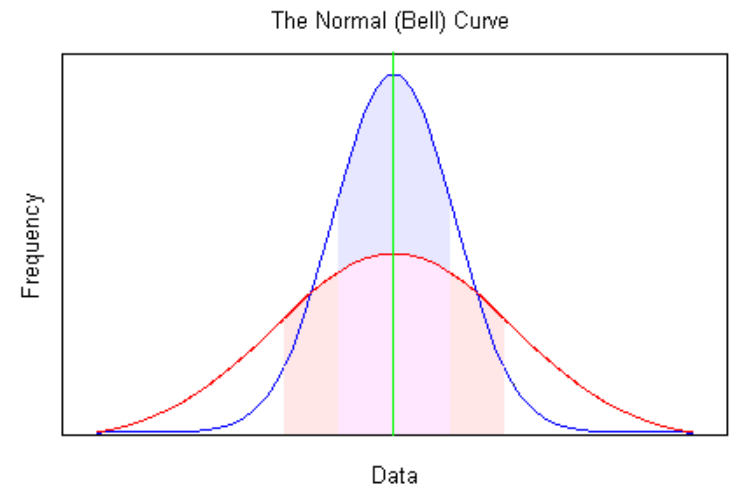


Descriptive Statistics



Standard Deviation

- A larger standard deviation (shown in light pink) indicates more scatter -- less precision -- in the results. A smaller standard deviation (shown in light blue) indicates less scatter. Both sets of results have the same mean (the green line).



$$\sigma^2 = \Sigma (X_i - \mu)^2 / N$$

Measures of Variability

- **Range** is the difference between the largest and smallest values in a set of values.
- **Variance** is the average squared deviation from the population mean, as defined by the following formula:

$$\sigma^2 = \sum (X_i - \mu)^2 / N$$

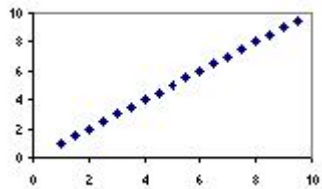
- **Standard Deviation** is the square root of the variance. Thus, the standard deviation of a population is:

$$\sigma = \text{sqrt} [\sigma^2]$$

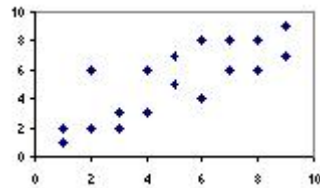
Correlation



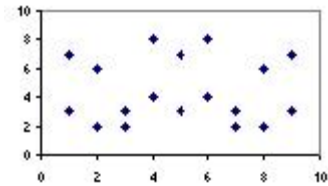
- **Correlation coefficients** measure the strength of association between two variables



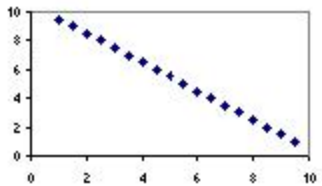
Maximum positive correlation



Strong positive correlation

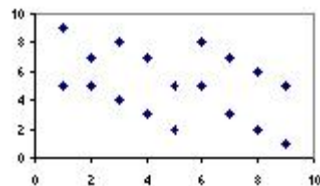


Zero correlation



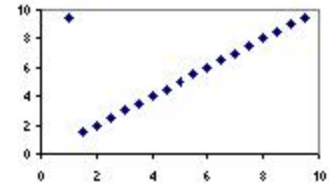
Minimum negative correlation

($r = -1.0$)



Moderate negative correlation

($r = -0.43$)



Strong correlation & outlier

($r = 0.71$)

The correlation r between two variables is: $r = \left[1 / (n - 1) \right] * \sum \left\{ \left[(x_i - \bar{x}) / s_x \right] * \left[(y_i - \bar{y}) / s_y \right] \right\}$

Regression

- Regression was first used by British biometrician Sir Francis Galton (1822-1911).
- According to his law of universal regression “each peculiarity in a man is shared by his kinsman but on the average in a less degree”
- Tall fathers do tend to have tall sons yet the average height of tall fathers is more than the average height of tall sons.
- Short fathers do tend to have short sons yet the average height of short fathers is less than the average height of short sons

© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com



search ID: gwbn119

***"This week we're going
to try some regression
therapy."***

Regression Analysis

- Regression Analysis: Dictionary meaning of regression is “stepping back”
- Regression Analysis: Some sort of functional relation ship between two or more related variable
- Regression Analysis: in regression any of the variable may be consider as a independent or dependent

Regression Analysis

- Easy to use and applies to many situations

Simple Regression: single explanatory variable

Multiple Regression: includes any number of explanatory variables.

Regression Analysis

- Dependant variable: the single variable being explained/ predicted by the regression model (response variable)
- Independent variable: The explanatory variable(s) used to predict the dependant variable. (predictor variable)

Regression Analysis

- Linear Regression: straight-line relationship

Form: $y = mx + b$

- Non-linear: implies curved relationships, for example logarithmic relationships

LEAST SQUARE ERROR TECHNIQUE

- Suppose $y=mx+c$ represents a straight line .
- Given a set of observed data $(a_1,b_1).....,(a_n,b_n)$.
- According to least square method the total error of the estimated points on a straight line with respect to observed points is minimal.

t-test

compare respiratory rates
between group of horses

group 1	group 2
11	11
12	12
13	13
14	13
15	14
	25

Horses in group 1 are
independent
from horses in group 2

H_0 : mean 1 = mean 2

H_a : mean 1 $\neq$ mean 2

What is the probability that the
means will differ?

t-test calculations

group 1	group 2
---------	---------

11	11
12	12
13	13
14	13
15	14
	25

mean 13

s1= 1.58

14

s2 = 5.16

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\text{var} \sqrt{(1/n_1) + (1/n_2)}}$$

$$\text{var} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

Statistical Forecasting

Forecasting means

“Looking into the future”



Load Forecasts

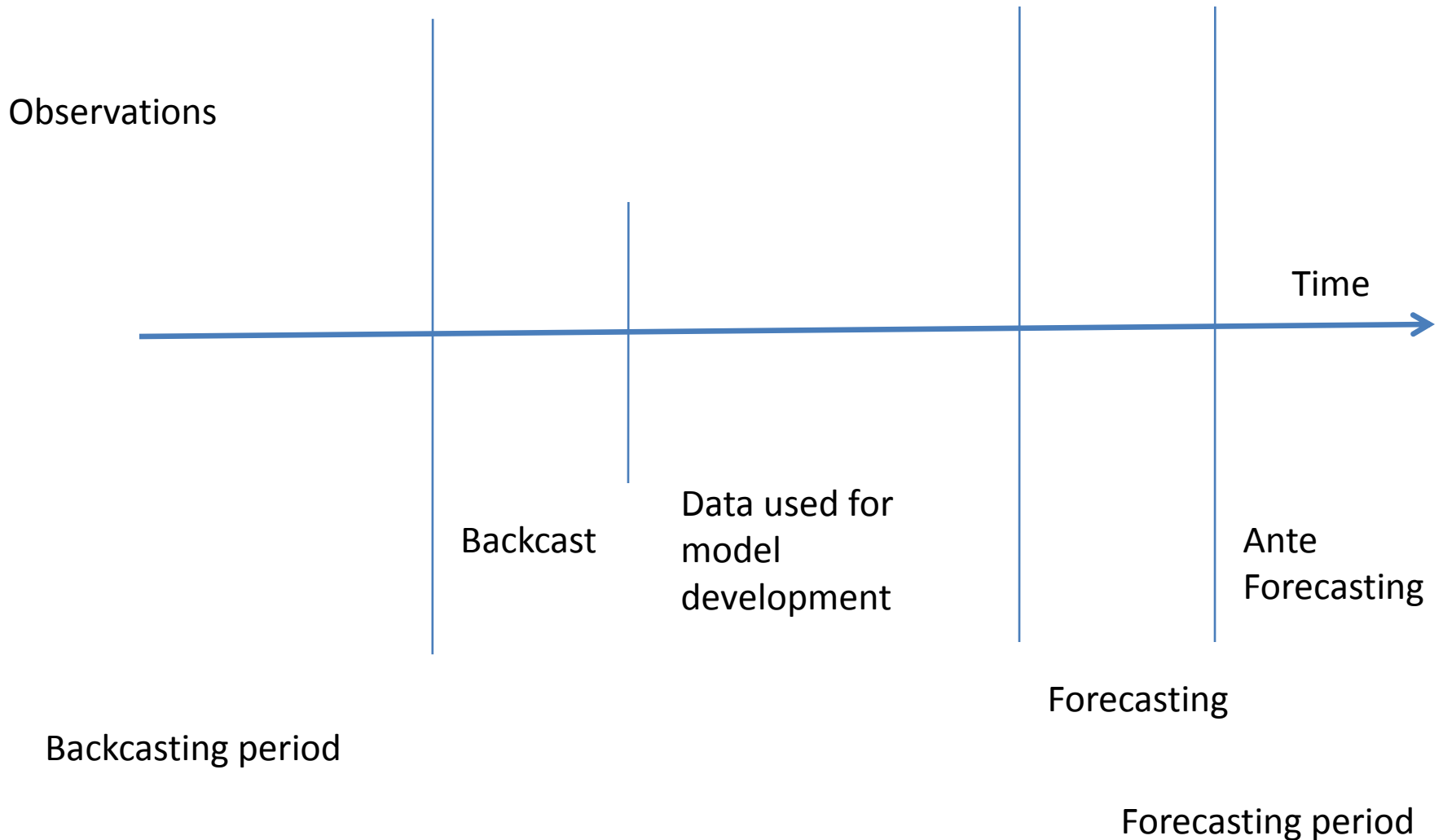
```
graph TD; A[Load Forecasts] --> B[Short term forecasts<br/>(one hour to a week)]; A --> C[Medium forecasts<br/>(a month up to a year)]; A --> D[Long term forecasts<br/>(over one year)];
```

Short term forecasts
(one hour to a week)

Medium forecasts
(a month up to a year)

Long term forecasts
(over one year)

Time Line in Forecasting

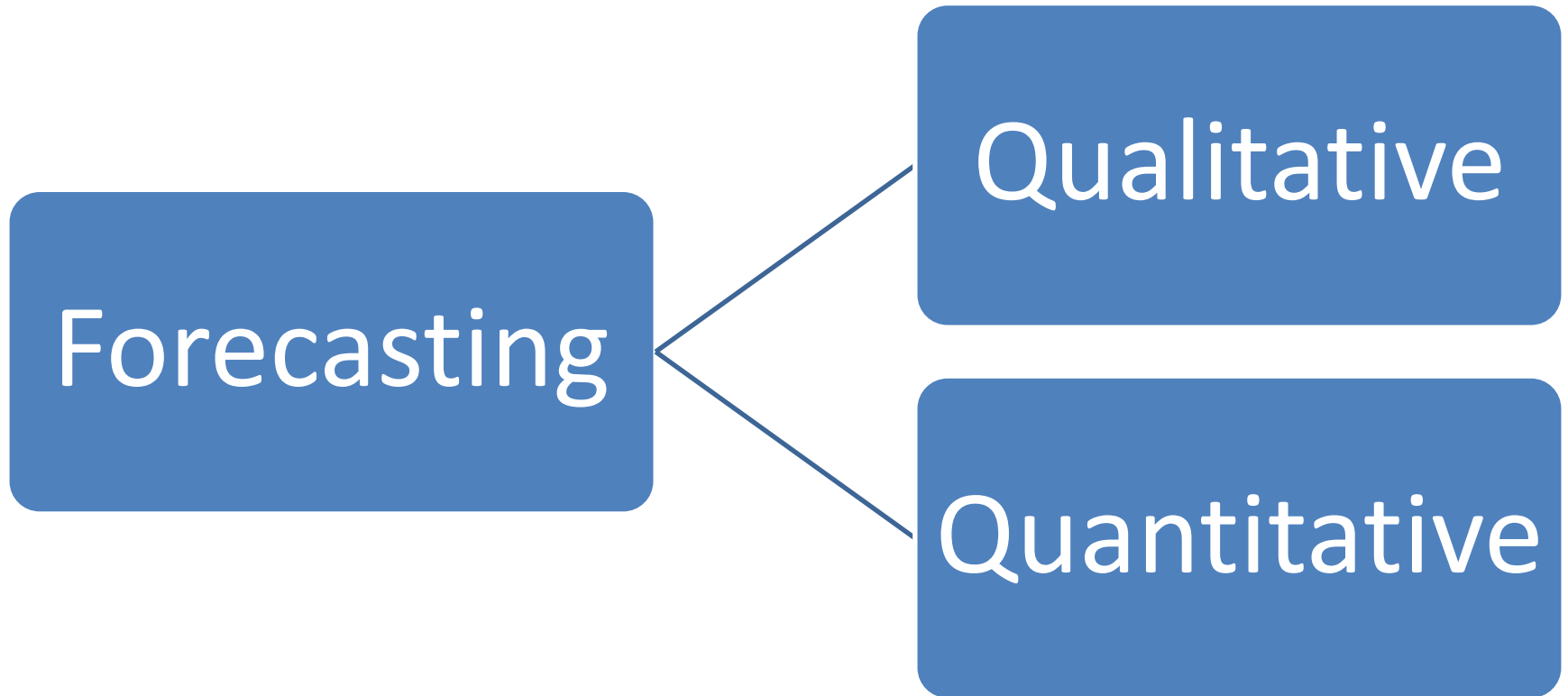


Statistical Forecasting Models

The selection criteria for statistical forecasting models have been well established for:

- Satisfaction of specific conditions (assumptions) under laying the statistical formulation of the model;
- Statistical significance of the model;
- Evaluation of forecast summary statistics over both the estimation and the ex post time periods.

FORECASTING TECHNIQUES



QUALITATIVE TECHNIQUES

Naïve Extrapolation

Sales Force Composite

Jury of Experts Opinion

Scenario Methods

Delphi Technique

Historical Analogy



Quantitative methods



Statistical methods

- Market Testing
- Market Survey

Time Series methods

- ARX method
- ARMAX

Association or causal method

- Correlation models
- Econometric model

Qualitative Approaches to Forecasting



■ Delphi Approach

- A panel of experts, each of whom is physically separated from the others and is anonymous, is asked to respond to a sequential series of questionnaires.
- After each questionnaire, the responses are tabulated and the information and opinions of the entire group are made known to each of the other panel members so that they may revise their previous forecast response.
- The process continues until some degree of consensus is achieved.

Qualitative Approaches (continued)

■ Scenario Writing

- Scenario writing consists of developing a conceptual scenario of the future based on a well defined set of assumptions.
- After several different scenarios have been developed, the decision maker determines which is most likely to occur in the future and makes decisions accordingly.

Qualitative Approaches (continued)

■ Subjective or Interactive Approaches

- These techniques are often used by committees or panels seeking to develop new ideas or solve complex problems.
- They often involve "brainstorming sessions".
- It is important in such sessions that any ideas or opinions be permitted to be presented without regard to its relevancy and without fear of criticism.

Quantitative Approaches to Forecasting

- Quantitative methods are based on an analysis of historical data concerning one or more time series.
- A time series is a set of observations measured at successive points in time or over successive periods of time.
- If the historical data used are restricted to past values of the series that we are trying to forecast, the procedure is called a time series method.
- If the historical data used involve other time series that are believed to be related to the time series that we are trying to forecast, the procedure is called a causal method.
- Quantitative approaches are generally preferred.

Time Series Data

- Time Series Data is usually plotted on a graph to determine the various characteristics or components of the time series data.
- There are 4 Major Components: Trend, Cyclical, Seasonal, and Irregular Components.

Components of a Time Series

- The trend component accounts for the gradual shifting of the time series over a long period of time.
- Any regular pattern of sequences of values above and below the trend line is attributable to the cyclical component of the series.
- The seasonal component of the series accounts for regular patterns of variability within certain time periods, such as over a year.
- The irregular component of the series is caused by short-term, unanticipated and non-recurring factors that affect the values of the time series. One cannot attempt to predict its impact on the time series in advance.

Time Series Data

- Forecasting Approaches:
Smoothing
Trend Projections

Smoothing Methods

- In cases in which the time series is fairly stable and has no significant trend, seasonal, or cyclical effects, one can use smoothing methods to average out the irregular components of the time series.
- Three common smoothing methods are:
 - Moving average
 - Weighted moving average
 - Exponential smoothing

Smoothing Methods: Moving Average

■ Moving Average Method

The moving average method consists of computing an average of the most recent n data values for the series and using this average for forecasting the value of the time series for the next period.

Smoothing Methods: Weighted Moving Average

- **Weighted Moving Average Method**

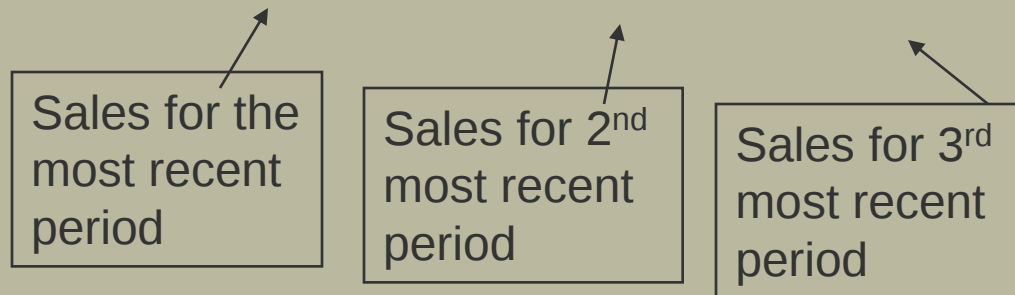
The weighted moving average method consists of computing a weighted average of the most recent n data values for the series and using this weighted average for forecasting the value of the time series for the next period. The more recent observations are typically given more weight than older observations. For convenience, the weights usually sum to 1.

- The regular moving average gives equal weight to past data values when computing a forecast for the next period. The weighted moving average allows different weights to be allocated to past data values.

Smoothing Methods: Weighted Moving Average

- Use a 3 period weighted moving average to forecast the sales for week 11 giving a weight of 0.6 to the most recent period, 0.3 to the second most recent period, and 0.1 to the third most recent period.

$$F_{11} = (0.6)*130 + (0.3)*110 + (0.1)*115 = 122.5$$



Thus we would forecast the sales for week 11 to be 122.5.

Smoothing Methods: Exponential Smoothing

- **Exponential Smoothing**

- Using exponential smoothing, the forecast for the next period is equal to the forecast for the current period plus a proportion (α) of the forecast error in the current period.
- Using exponential smoothing, the forecast is calculated by:

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t$$

where:

← This is the same as
 $F_{t+1} = F_t + \alpha (Y_t - F_t)$

α is the smoothing constant (a number between 0 and 1)

F_t is the forecast for period t

F_{t+1} is the forecast for period $t+1$

Y_t is the actual data value for period t

Trend Projection

- If a time series exhibits a linear trend, the method of least squares may be used to determine a trend line (projection) for future forecasts.
- Least squares, also used in regression analysis, determines the unique trend line forecast which minimizes the mean square error between the trend line forecasts and the actual observed values for the time series.
- The independent variable is the time period and the dependent variable is the actual observed value in the time series.

Trend Projection

- Using the method of least squares, the formula for the trend projection is:

$$Y_t = b_0 + b_1 t$$

where:

Y_t = trend forecast for time period t

b_1 = slope of the trend line

b_0 = trend line projection for time 0

$$b_1 = \frac{n \sum t Y_t - \sum t \sum Y_t}{n \sum t^2 - (\sum t)^2}$$

$$b_0 = \bar{Y} - b_1 \bar{t}$$

where: Y_t = observed value of the time series at time period t

$\bar{Y}$ = average of the observed values for Y_t

$\bar{t}$ = average time period for the n observations

Causal Method: Regression Analysis

- Regression Analysis is similar to trend analysis, except the independent variable is not restricted to time. Refer to Robert's Drug example. Instead of letting time represent our independent variable, we could forecast sales based upon the price of the product. Since products often go on sale, we could collect data over several months collecting the weekly price and number of items sold for the week. For this model, we would find the regression equation in the same manner in which we found the trend line except we would call the independent variable x , instead of t .

Conclusions

Various stats tools

Mean, Mode, Median

Variance, SD

Regression

Discussed forecast methods

Qualitative and Quantitative

Thanks

