

Выборочное наблюдение



Понятие о выборочном наблюдении.

Выборочное наблюдение относится к разновидностям несплошного наблюдения. Оно охватывает отобранную часть единиц генеральной совокупности. Цель выборочного наблюдения - по отобранной части единиц дать характеристику всей совокупности единиц. Чтобы отобранная часть была репрезентативна (т.е. представляла всю совокупность единиц), выборочное наблюдение должно быть специально организовано. Следовательно, в отличие от генеральной совокупности, представляющей всю совокупность исследуемых единиц, выборочная совокупность представляет ту часть единиц генеральной совокупности, которая является объектом непосредственного наблюдения.



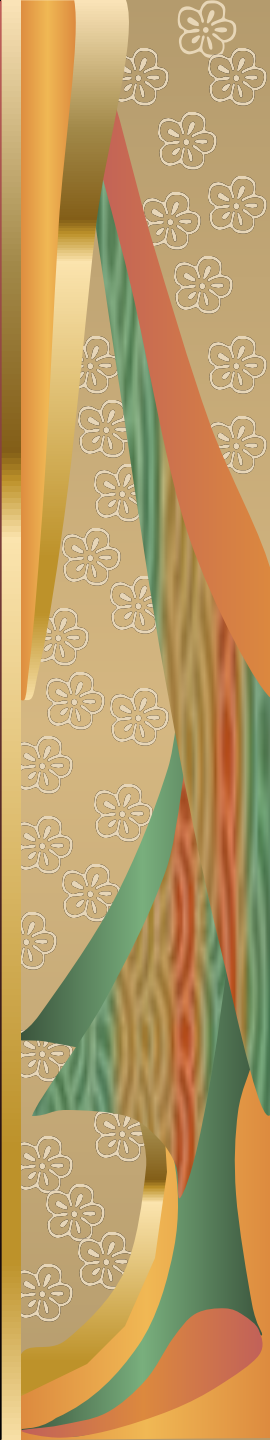
В статистике приняты следующие условные обозначения:

- *N - объем генеральной совокупности;*
- *n - объем выборочной совокупности;*
- *\bar{X} - средняя в генеральной совокупности;*
- *\bar{x} - средняя в выборочной совокупности;*
- *p - доля единиц в генеральной совокупности;*
- *w - доля единиц в выборочной совокупности;*
- *σ^2 - генеральная дисперсия;*
- *S^2 - выборочная дисперсия;*
- *σ - среднее квадратическое отклонение признака в генеральной совокупности;*
- *S - среднее квадратическое отклонение признака в выборочной совокупности*

Виды выборки, способы отбора и ошибки выборочного наблюдения

По способу отбора (способу формирования) выборки единиц из генеральной совокупности распространены следующие виды выборочного наблюдения:

- простая случайная выборка (собственно-случайная);**
- типическая (стратифицированная);**
- серийная (гнездовая);**
- механическая;**
- комбинированная;**
- ступенчатая.**



Простая случайная выборка (собственно-случайная) есть отбор единиц из генеральной совокупности путем случайного отбора, но при условии вероятности выбора любой единицы из генеральной совокупности. Отбор проводится методом жеребьевки или по таблице случайных чисел.

Типическая (стратифицированная) выборка предполагает разделение неоднородной генеральной совокупности на типологические или районированные группы по какому-либо существенному признаку, после чего из каждой группы производится случайный отбор единиц.

Для серийной (гнездовой) выборки характерно то, что генеральная совокупность первоначально разбивается на определенные равновеликие или неравновеликие серии (единицы внутри серий связаны по определенному признаку), из которых путем случайного отбора отбираются серии и затем внутри отобранных серий проводится сплошное наблюдение.

Механическая выборка представляет собой отбор единиц через равные промежутки (по алфавиту, через временные промежутки, по пространственному способу и т.д.). При проведении механического отбора генеральная совокупность разбивается на равные по численности группы, из которых затем отбирается по одной единице.

Комбинированная выборка основана на сочетании нескольких способов выборки.

Многоступенчатая выборка есть образование внутри генеральной совокупности вначале крупных групп единиц, из которых образуются группы, меньшие по объему, и так до тех пор, пока не будут отобраны те группы или отдельные единицы, которые необходимо исследовать.

Выборочный отбор может быть повторным и бесповторным. При повторном отборе вероятность выбора любой единицы не ограничена. При бесповторном отборе выбранная единица в исходную совокупность не возвращается.

Для отобранных единиц рассчитываются обобщенные показатели (средние или относительные) и в дальнейшем результаты выборочного исследования распространяются на всю генеральную совокупность.

Расчет средней ошибки повторной простой случайной выборки производится следующим образом:

$$\mu_x = \sqrt{\frac{S^2}{n}}$$

средняя ошибка для средней

$$\mu_p = \sqrt{\frac{w(1-w)}{n}}$$

средняя ошибка для доли



$$\mu_x = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$$

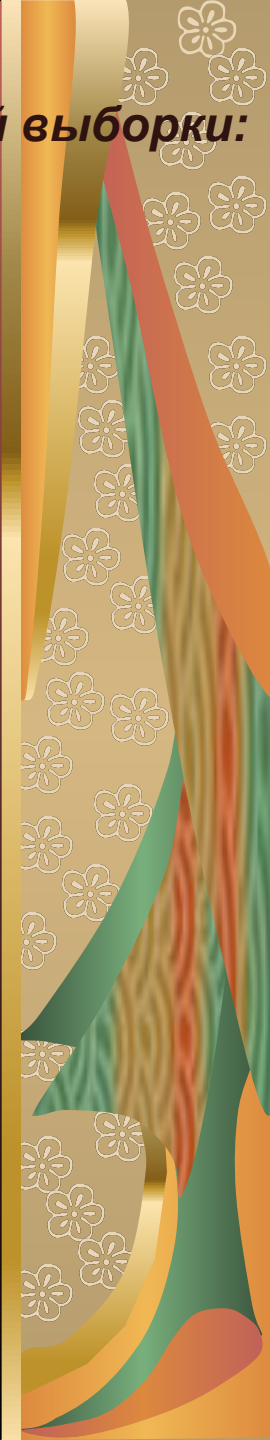
**Расчет средней ошибки бесповторной случайной выборки:
средняя ошибка для средней**

$$\mu_p = \sqrt{\frac{W(1-W)}{n} \left(1 - \frac{n}{N}\right)}$$

средняя ошибка для доли

$$\Delta_x = t \sqrt{\frac{S^2}{n}}$$

предельная ошибка для средней



$$\Delta_p = t \sqrt{\frac{w(1-w)}{n}}$$

предельная ошибка для доли

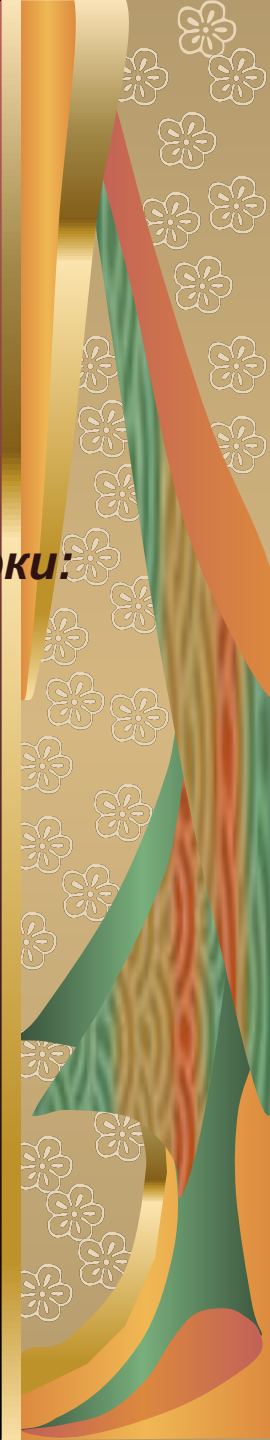
Расчет предельной ошибки бесповторной случайной выборки:

$$\Delta_{\bar{x}} = t \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$$

предельная ошибка для средней

$$\Delta_p = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$$

предельная ошибка для доли



Чтобы рассчитать численность n повторной и бесповторной простой случайной выборки, можно использовать следующие формулы:

$$n = \frac{t^2 S^2}{\Delta_x^2}$$

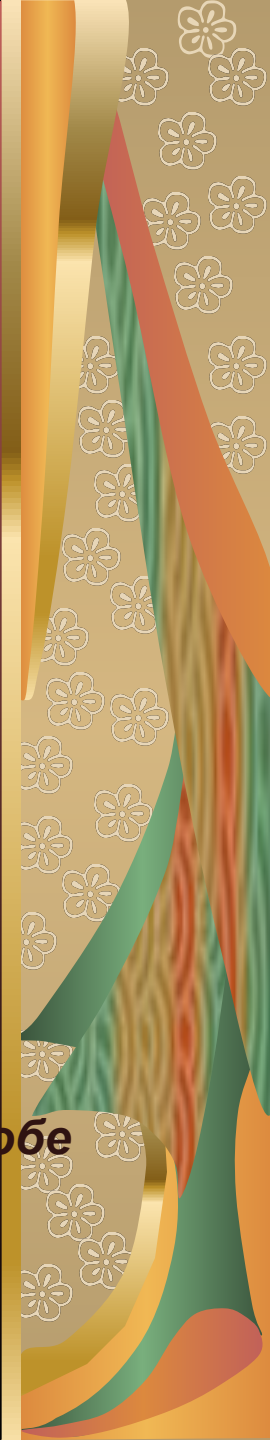
для средней при повторном способе



$$n = \frac{t^2 NS^2}{\Delta_x^2 N + t^2 S^2} \quad \text{для средней при бесповторном способе}$$

$$n = \frac{t^2 w(1-w)}{\Delta_p^2} \quad \text{для доли при повторном способе}$$

$$n = \frac{t^2 Nw(1-w)}{\Delta_p^2 N + t^2 w(1-w)} \quad \text{для доли при бесповторном способе}$$



GOALS

- Determine the components of a time series
- Calculate moving average
- Determine a linear trend equation
- Work out a trend equation for a nonlinear trend



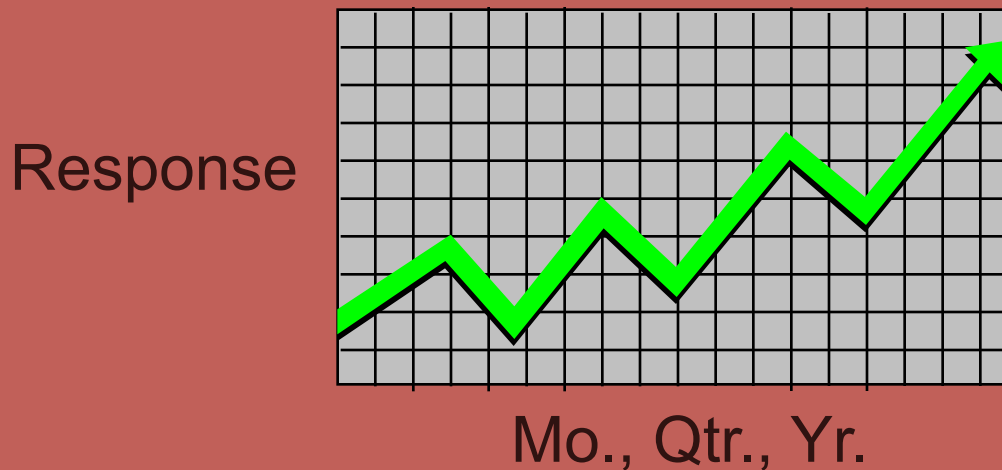
Components of a Time Series

- Secular Trend – the smooth long term direction of a time series
- Cyclical Variation – the rise and fall of a time series over periods longer than one year
- Seasonal Variation – Patterns of change in a time series within a year which tends to repeat each year
- Irregular Variation – classified into:
 - Episodic – unpredictable but identifiable
 - Residual – also called chance fluctuation and unidentifiable



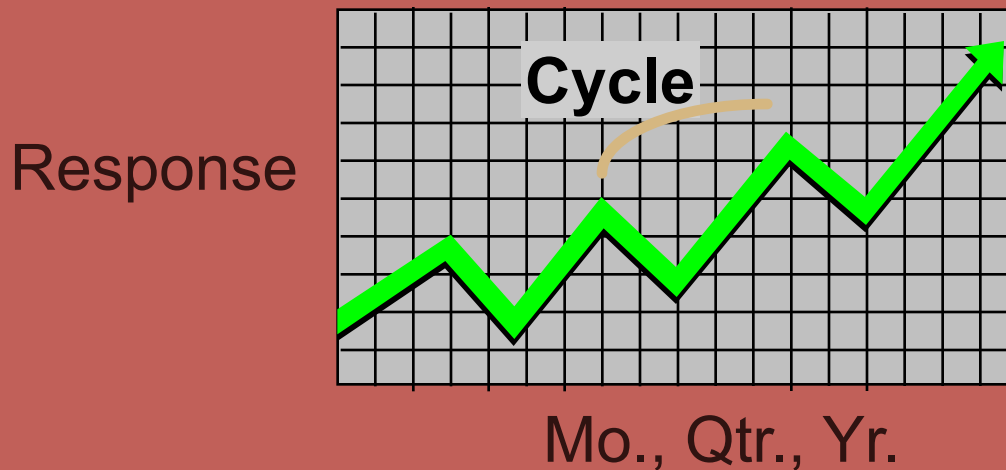
Trend Component

- 1. Persistent, Overall Upward or Downward Pattern
- 2. Due to Population, Technology etc.
- 3. Several Years Duration
-



Cyclical Component

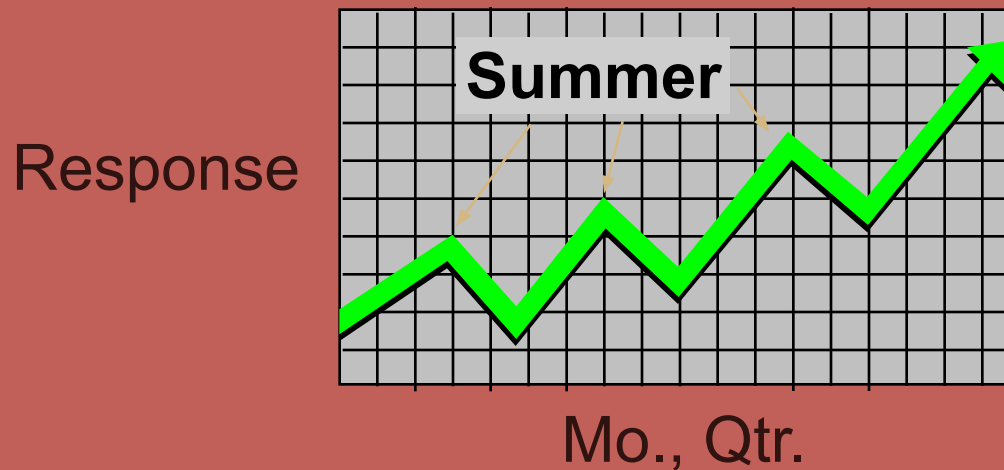
- 1. Repeating Up & Down Movements
- 2. Due to Interactions of Factors Influencing Economy
- 3. Usually 2-10 Years Duration



Peak
Contraction
Trough
Expansion

Seasonal Component

- 1. Regular Pattern of Up & Down Fluctuations
- 2. Due to Weather, Customs etc.
- 3. Occurs Within 1 Year
-



The Moving Average Method

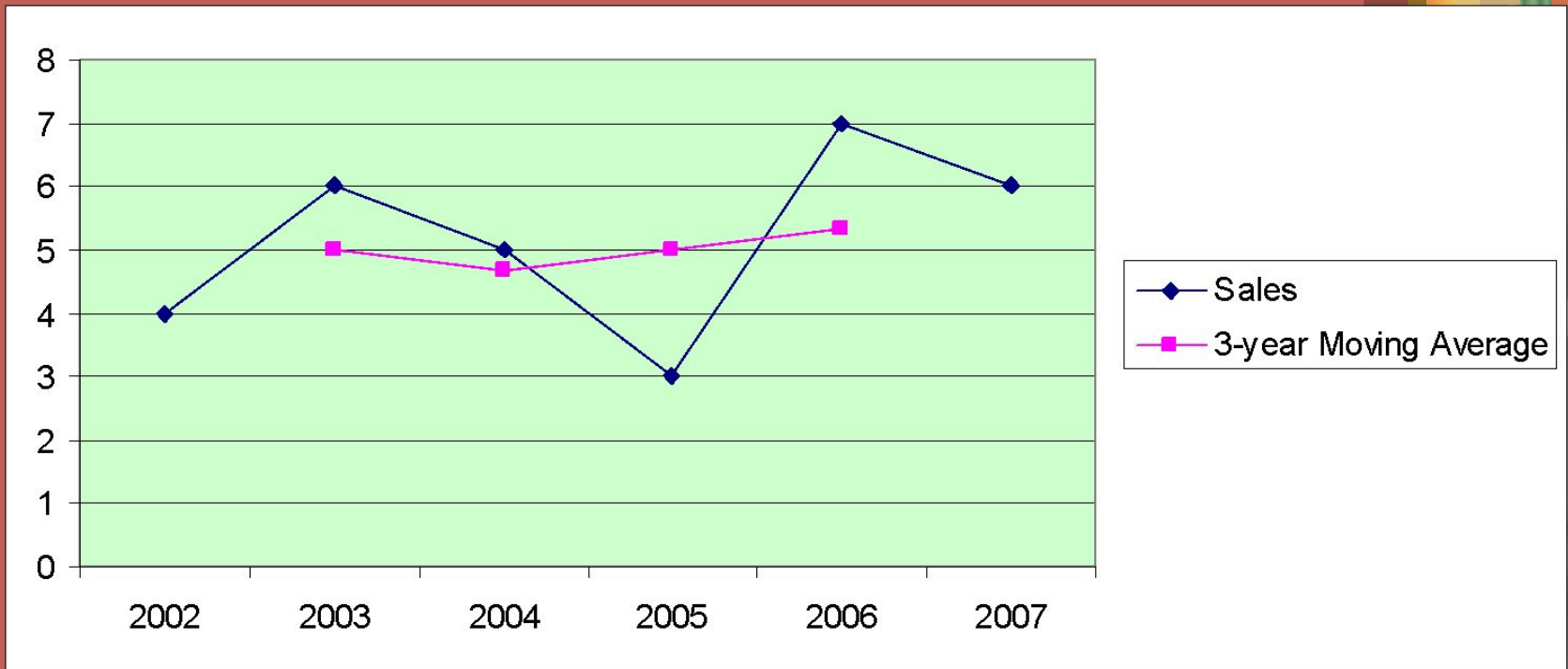
- Smoothing time series to see its trend
- Fundamental method used in measuring seasonal fluctuation
- Time series follows fairly linear trend that have definite rhythmic pattern is most applicable



Moving Average Method - Example

Time	Response	Moving Total	Moving
2002	4	NA	NA
2003	6	$4 + 6 + 5 = 15$	$15/3 = 5.0$
2004	5	$6 + 5 + 3 = 14$	$14/3 = 4.7$
2005	3	$5 + 3 + 7 = 15$	$15/3 = 5.0$
2006	7	$3 + 7 + 6 = 16$	$16/3 = 5.3$
2007	6	NA	NA

Three-year Moving Averages



Weighted Moving Average

- A simple moving average assigns the same weight to each observation in averaging
- Weighted moving average assigns different weights to each observation
- Most recent observation receives the most weight, and the weight decreases for older data values
- In either case, the sum of the weights = 1



Weighted Moving Average - Example

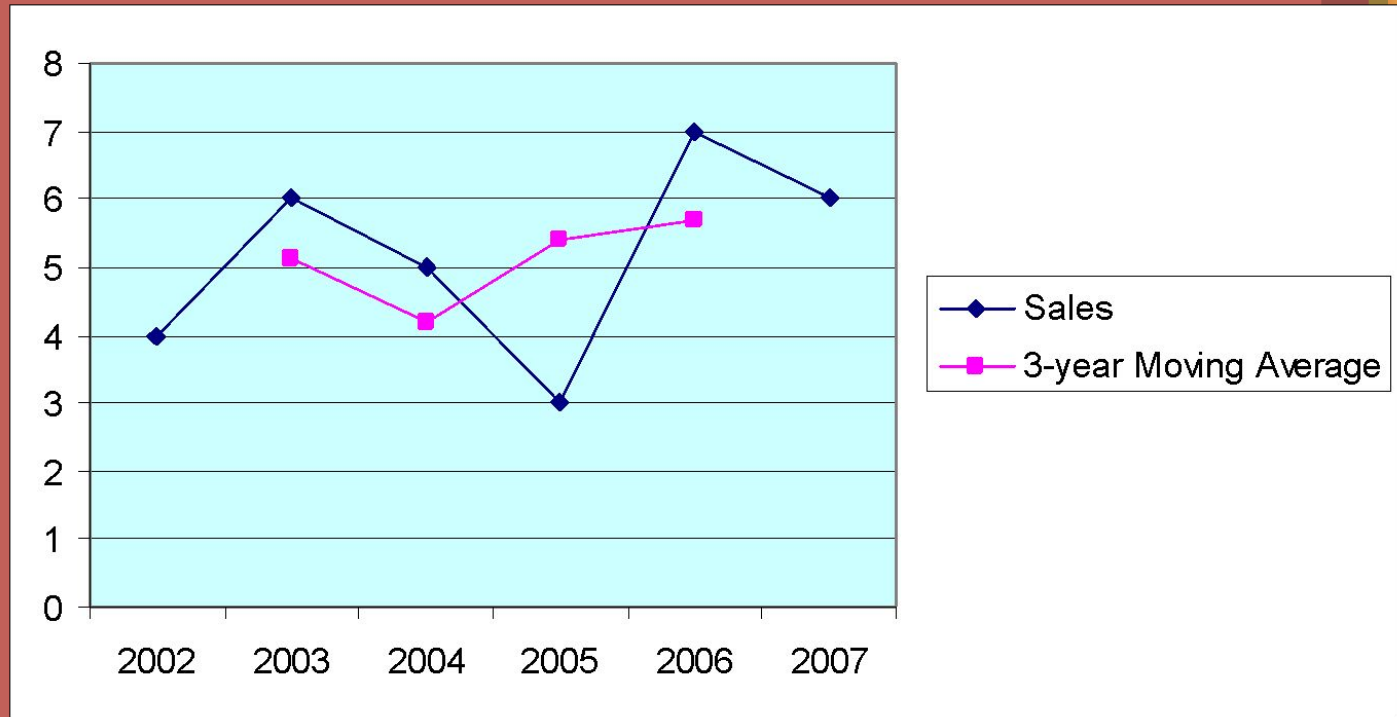
Compute a three-year moving average and a three-year weighted moving average with weights of **0.2**, **0.3**, and **0.5** for successive years.



Weighted Moving Average - Example

Year	Sales	3-year Moving Average	3-year Moving Average
2002	4		
2003	6	$0.2*4+0.3*6+0.5*5$	5.1
2004	5	$0.2*6+0.3*5+0.5*3$	4.2
2005	3	$0.2*5+0.3*3+0.5*7$	5.4
2006	7	$0.2*3+0.3*7+0.5*6$	5.7
2007	6		

Weighed Moving Average – An Example



Linear Trend

- The long term trend of many business series often approximates a straight line

Linear Trend Equation: $\hat{Y} = a + bt$

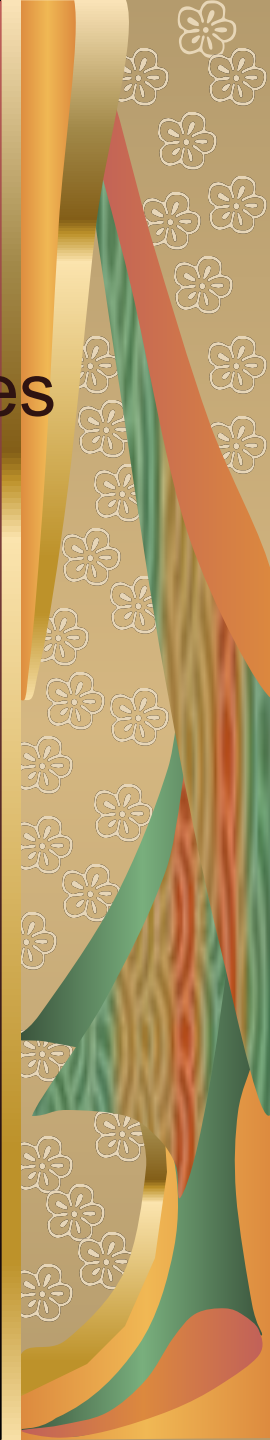
where:

\hat{Y} – read "Y hat", is the projected value of the variable of interest (response variable)

a – the Y - intercept
(estimated value of Y when $t = 0$)

b – the slope of the line
(average change in Y for each unit change in t)

t – any value of time (coded) that is selected



Linear Trend Plot

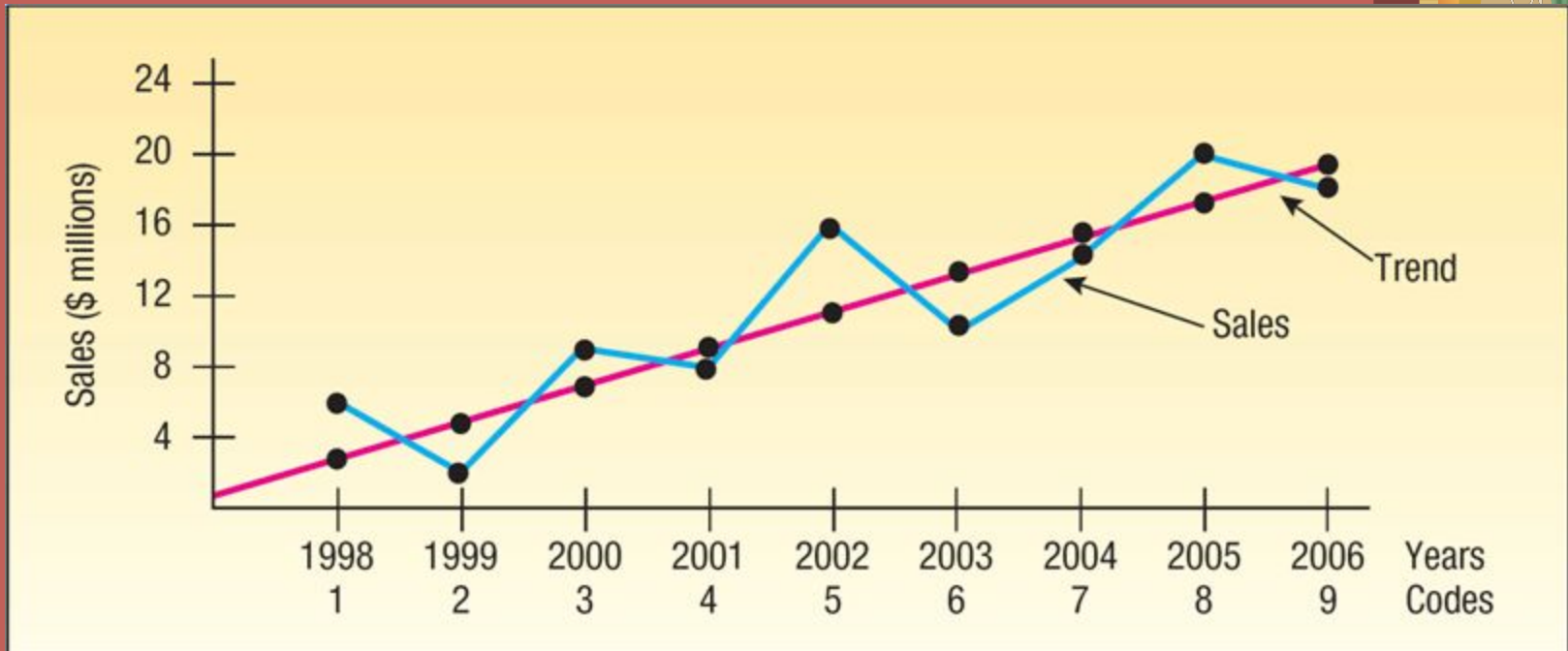


CHART 16-5 A Straight Line Fitted to Sales Data

Linear Trend – Using the Least Squares Method

- Use the least squares method in Simple Linear Regression to find the best linear relationship between 2 variables
- Code time (t) and use it as the independent variable
- E.g. let t be 1 for the first year, 2 for the second, and so on (if data are annual)



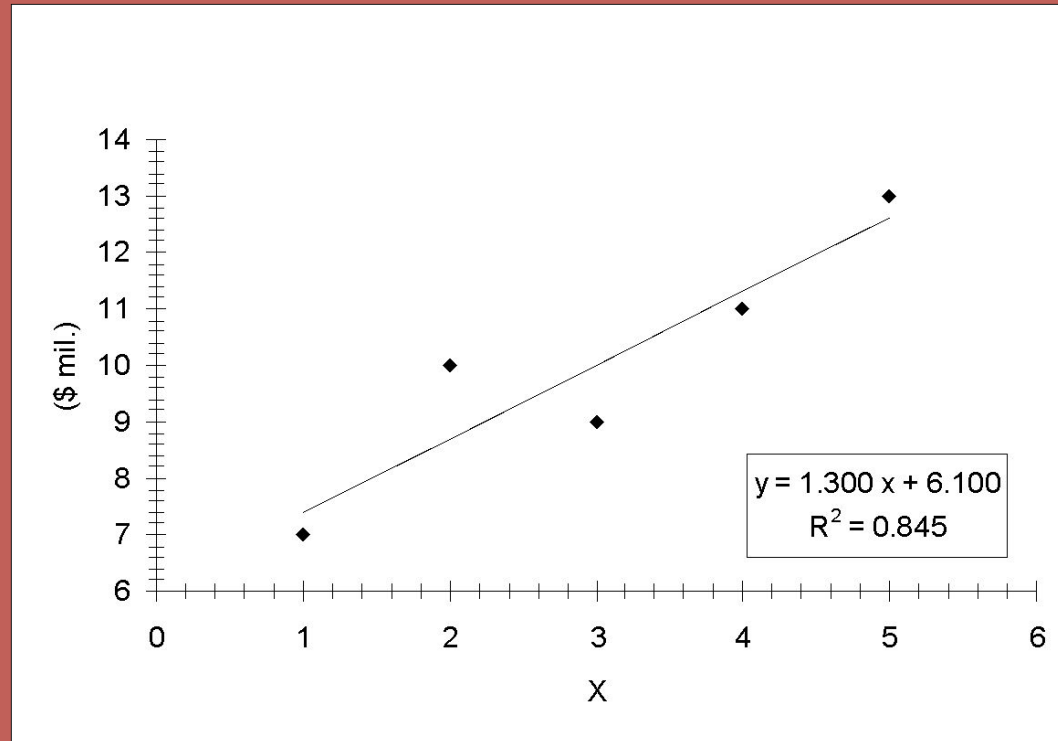
Linear Trend – Using the Least Squares Method: An Example

Below shows the sales of a NTUC outlet in Ang Mo Kio,

Year	Sales (\$ mil.)
2002	7
2003	10
2004	9
2005	11
2006	13

Year	t	Sales (\$ mil.)
2002	1	7
2003	2	10
2004	3	9
2005	4	11
2006	5	13

Linear Trend – Using the Least Squares Method: An Example Using Excel



Linear Trend – Using the Least Squares to forecast

What is the forecasted sales for the year 2007?

Linear Trend Equation : $\hat{Y} = 6.1 + 1.3t$

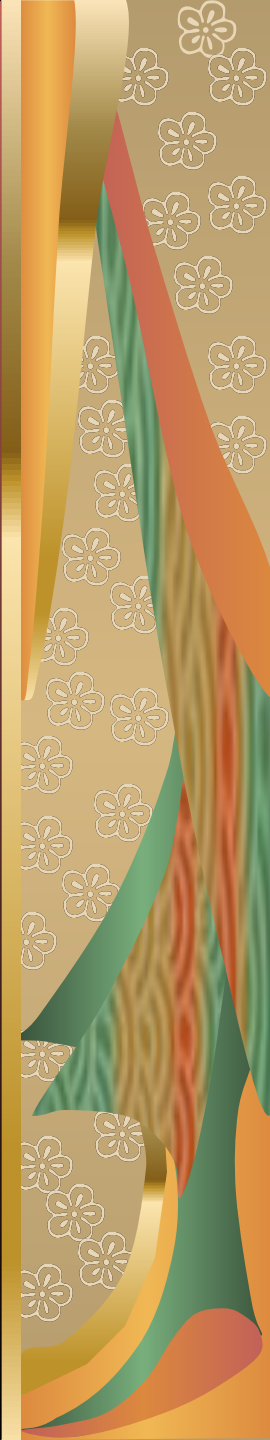
where :

$$\hat{Y} = 6.1 + 1.3(6) = 13.9$$



Nonlinear Trends

- Linear trend equation is used when the data are increasing (or decreasing) by equal amounts
- Nonlinear trend equation is used when the data are increasing (or decreasing) by increasing amounts over time
- When data increase (or decrease) by equal *percents or proportions* plot will show curvilinear pattern

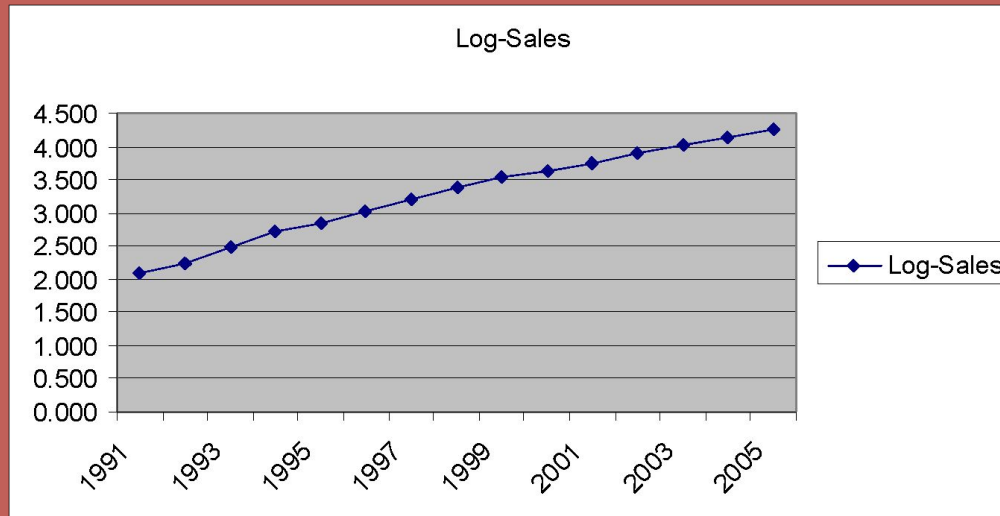
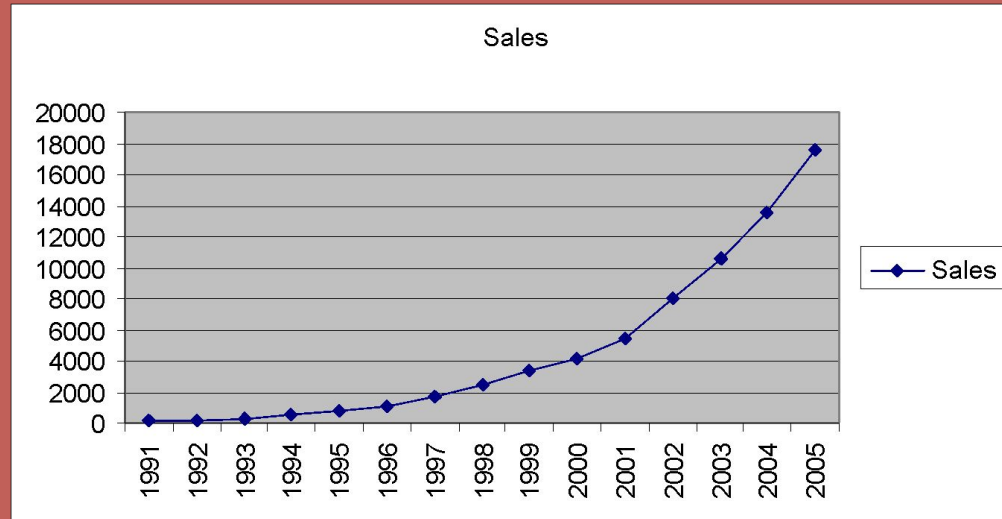


Log Trend Equation – Gulf Shores Importers Example

- Top graph is plot of the original data
- Bottom graph is the log base 10 of the original data which now is linear
(Excel function: $=\log(x)$ or $\log(x,10)$)
- Using Data Analysis in Excel, generate the linear equation
- Regression output shown in next slide

Year	Sales	Log-Sales
1991	124.2	2.094
1992	175.6	2.245
1993	306.9	2.487
1994	524.3	2.720
1995	714	2.854
1996	1052	3.022
1997	1638.3	3.214
1998	2463.2	3.391
1999	3358.2	3.526
2000	4181.3	3.621
2001	5388.5	3.731
2002	8027.4	3.905
2003	10587.2	4.025
2004	13537.4	4.132
2005	17515.6	4.243

Log Trend Equation – Gulf Shores Importers Example



Log Trend Equation – Gulf Shores Importers Example

The Linear Equation is :

$$\hat{y} = 2.053805 + 0.153357t$$

Regression Analysis

r^2 0.988 n 15
 r 0.994 k 1
 Std. Error 0.079 Dep. Var. **Log-Sales**

ANOVA table

Source	SS	df	MS	F	p-value
Regression	6.5851	1	6.5851	1065.10	7.37E-14
Residual	0.0804	13	0.0062		
Total	6.6654	14			

Regression output

variables	coefficients	std. error	t (df=13)	p-value	confidence interval	
					95% lower	95% upper
Intercept	2.0538	0.0427	48.072	4.99E-16	1.9615	2.1461
t	0.1534	0.0047	32.636	7.37E-14	0.1432	0.1635

Log Trend Equation – Gulf Shores Importers Example

Estimate the Import for the year 2009 using the linear trend

$$\hat{y} = 2.053807 + 0.153357t$$

Substitute into the linear equation above the code (19) for 2009

$$\hat{y} = 2.053805 + 0.153357(19)$$

$$\hat{y} = 4.967588$$

Then find the antilog of $\hat{y} = 10^{\hat{y}}$

$$= 10^{4.967588}$$

$$= 92,808$$

