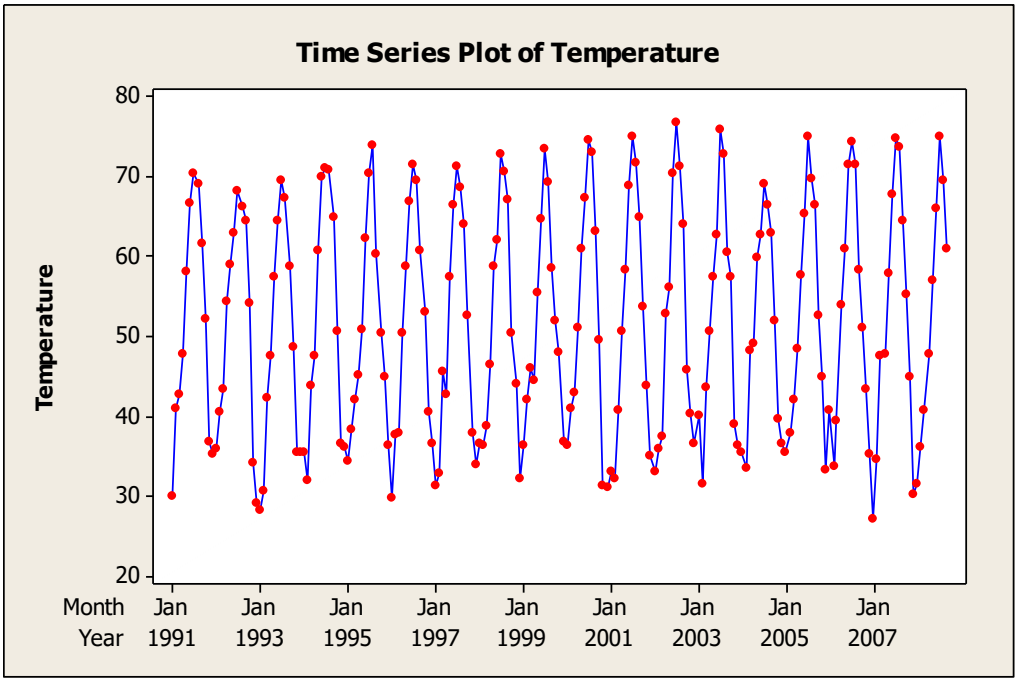
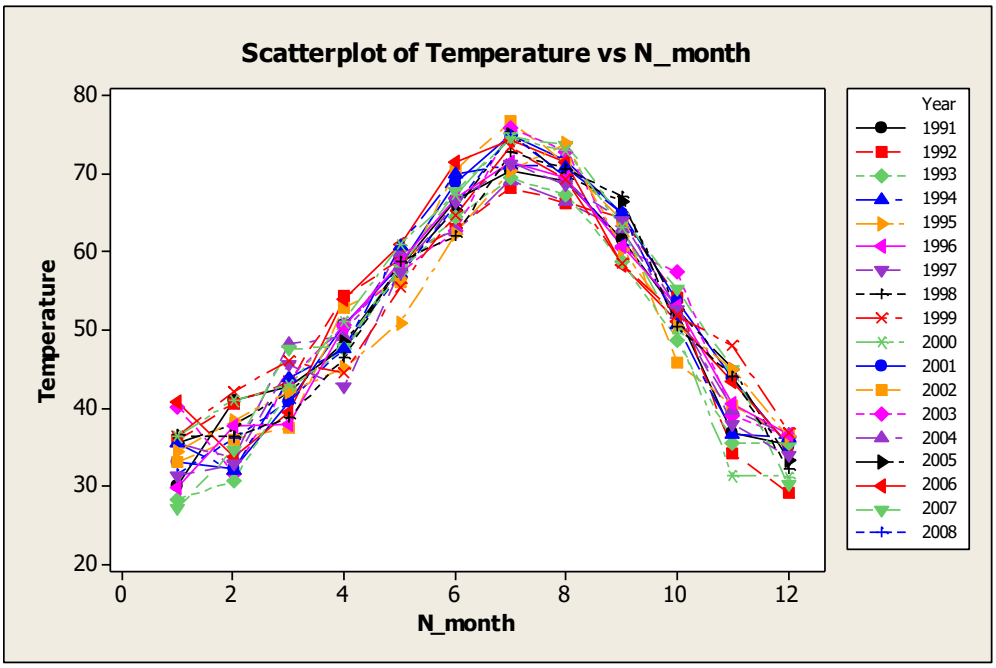
Chapter 2 ● Basic Tools for Forecasting
Outline Solutions

2.1

Time series plot for temperature

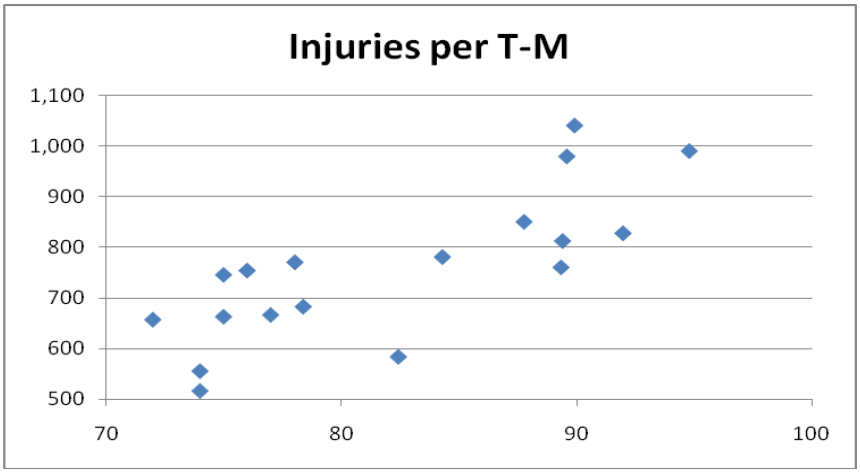


Seasonal plot: clear and stable seasonal pattern



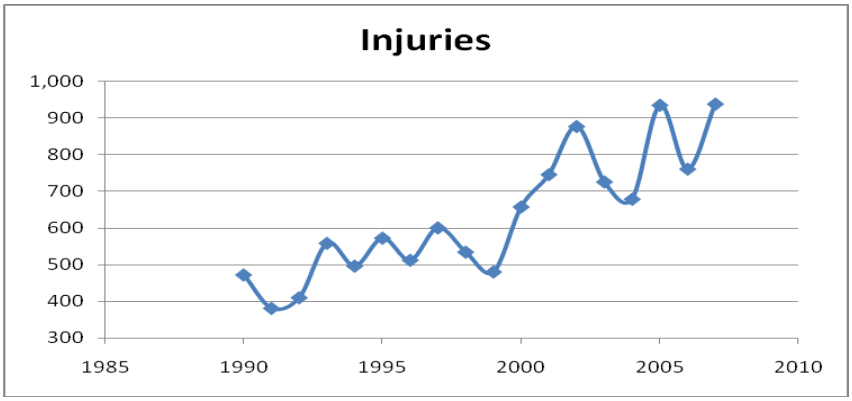
2.2

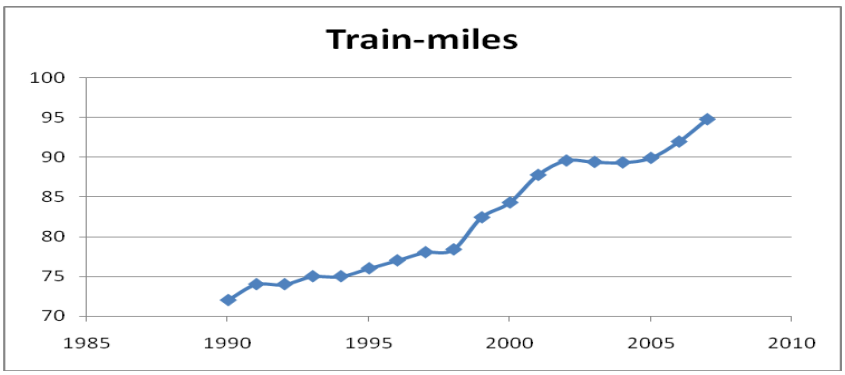
(a) Scatter plot for injuries versus train miles



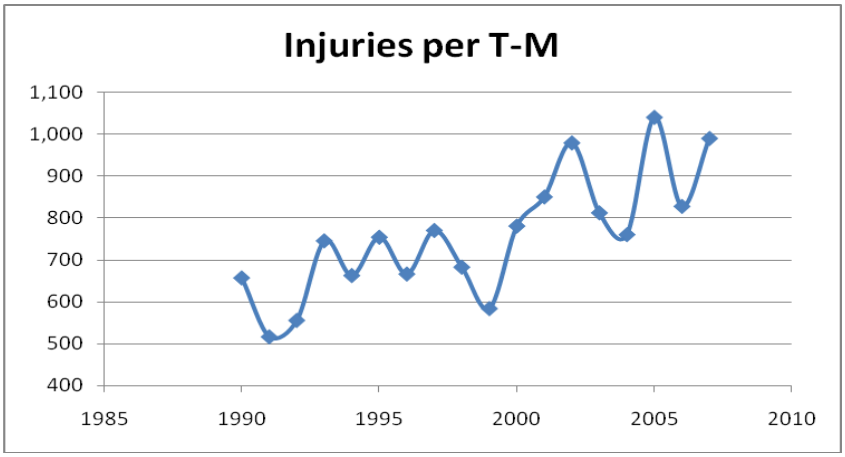
© Cengage Learning 2013

(b) Time plots for Injuries, Train Miles and Injuries per 100 million train miles





© Cengage Learning 2013



(c) The level of injuries fluctuates but shows an upward trend over time.

2.3

The standard Excel output shows that the median is unchanged but the mean is much smaller and the standard deviation much larger. A few outliers can greatly distort the analysis.

Returns		*Returns with outlier*	
Mean	7.48	Mean	-3.08
Standard Error	3.32	Standard Error	12.46
Median	6.20	Median	6.20
Mode	#N/A	Mode	#N/A
Standard Deviation	9.96	Standard Deviation	37.39
Sample Variance	99.19	Sample Variance	1398.31
Kurtosis	-0.20	Kurtosis	7.65
Skewness	0.58	Skewness	-2.68
Range	30.40	Range	125.40
Minimum	-5.00	Minimum	-100.00
Maximum	25.40	Maximum	25.40
Sum	67.30	Sum	-27.70
Count	9.00	Count	9.00

The MAD is 7.6 without the outlier and 21.7 with the outlier.

2.4 Means and medians are close suggesting a symmetric pattern. November and the winter months are somewhat more variable.

Month	N	Mean	Std Dev	Median	MAD
JAN	18	34.0	3.8	35.0	3.06
FEB	18	36.0	3.6	36.1	2.99
MAR	18	42.5	3.1	42.6	2.35
APR	18	48.8	3.1	48.2	2.49
MAY	18	58.0	2.3	58.1	1.58
JUN	18	66.1	2.9	66.3	2.36
JUL	18	72.8	2.6	73.2	2.24
AUG	18	70.3	2.3	70.2	1.88
SEP	18	62.6	2.7	63.0	2.30
OCT	17	51.9	2.6	51.9	1.91
NOV	17	40.4	4.6	40.3	3.75
DEC	17	34.5	2.4	35.3	1.99

2.5

The standard summary statistics are given in the table. To compute the MAD in each case, calculate the absolute values of the deviations about the mean and compute their average.

Measure	*Train-miles*	*Injuries*	*Injuries per T-M*
Mean	82.2	630.3	757.4
Standard Error	1.8	40.3	34.4
Median	80.4	587.0	757.0
Mode	74.0	#N/A	#N/A
Standard Deviation	7.5	171.1	146.1
Sample Variance	55.5	29280.2	21331.3
Kurtosis	-1.6	-0.7	-0.3
Skewness	0.2	0.5	0.4
Range	22.8	556.0	523.9
Minimum	72.0	382.0	516.2
Maximum	94.8	938.0	1040.1
Sum	1478.8	11345.0	13632.7
Count	18.0	18.0	18.0

Variable	Train-miles	Injuries	Injuries per T-M
MAD	6.7	142.0	110.4

Since the variables display a rising trend, the calculation of these averages is not particularly informative.

2.6

The correlation between the 2002 and 2003 figures is 0.955. The associated P-value is 0.000 to 3 decimal places.

2.7

The three sets of results are given below. The variation among the first 50 values is much greater than the second 50, so that the results for the first 50 and for all 100 are very similar. The high negative correlation between rank and return on capital (ROC) simply reflects the fact that the rank ordering placed heavy weight on ROC.

All 100	*Rank*	*Return on Capital*	*P-E Ratio*
Rank	1.00		
Return on Capital	-0.65	1.00	
P-E Ratio	-0.27	0.31	1.00

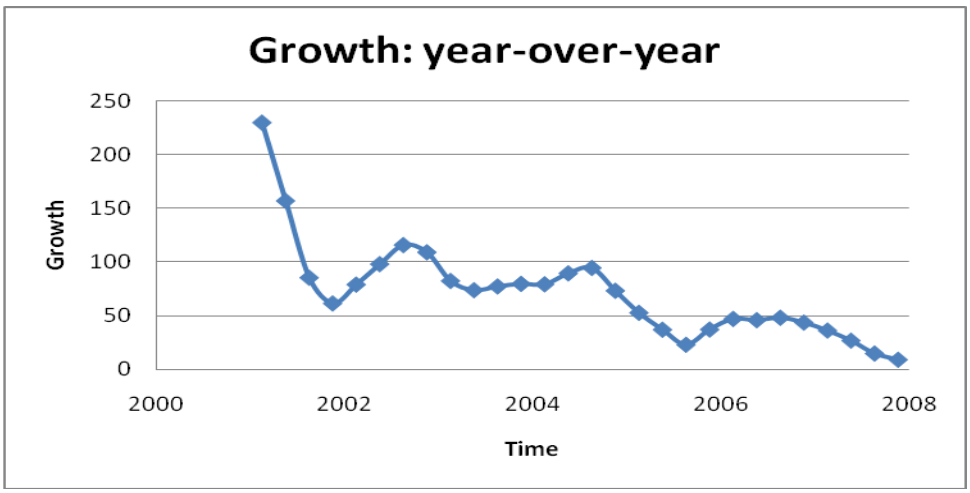
First 50	*Rank*	*Return on Capital*	*P-E Ratio*
Rank	1.00		
Return on Capital	-0.66	1.00	
P-E Ratio	-0.24	0.27	1.00

Second 50	*Rank*	*Return on Capital*	*P-E Ratio*
Column 1	1.00		
Column 2	-0.19	1.00	
Column 3	0.12	-0.14	1.00

2.8

a. The mean and median are not useful, given the strong trend in the data.

b. There is a clear slowing of percentage growth over time



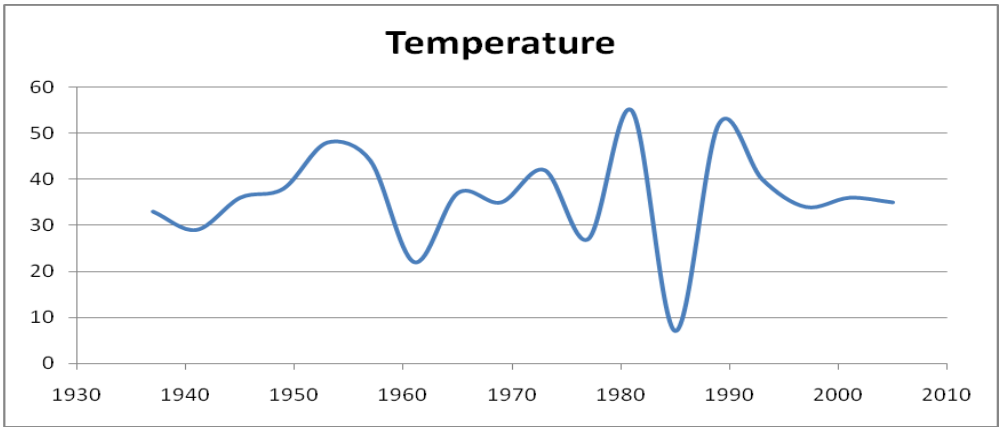
2.9

The use of the means produces apparently better results, but it should be recognized that the means include the observations actually being forecast. Only prior information should be used. As expected, one-month-ahead forecasts are not useful in this case.

	12 month errors	*Means errors*	*1-step errors*
ME	0.02	0.03	0.13
MAE	3.45	2.45	6.93
RMSE	4.45	3.03	8.23
MASE	0.50	0.35	1.00

2.10

The mean is around 36 degrees and the standard deviation is 11 degrees. With 18 observations, the 95% t-value from tables is 2.11 so the prediction interval is $36.11 \pm 2.11 * 11.02 = 36.11 \pm 23.25 = [12.9, 59.4]$. In other words, be prepared for almost any kind of weather. Clearly much better forecasts are possible closer to the event!!



Temperature	
Mean	36.11
Standard Error	2.60
Median	36.00
Mode	36.00
Standard Deviation	11.02
Sample Variance	121.40
Kurtosis	1.98
Skewness	-0.76
Range	48.00
Minimum	7.00
Maximum	55.00
Sum	650.00
Count	18.00

2.11

The prediction limits are ±1.96*4.445 or ±8.71. Out of 201 monthly values 4 are below the lower limit and 2 are above the upper limit, making just under 3 percent overall.

2.12

There are two values below the lower limit and none above, possibly suggesting either milder weather or a cost-saving effort by the home-owner.

Period	Actual	Forecast	Error	Lower PI	Upper PI	Below Lower Limit	Above Upper Limit
Jan-03	790	820	-30	473	1167	0	0
Feb-03	810	790	20	443	1137	0	0
Mar-03	680	720	-40	373	1067	0	0
Apr-03	500	640	-140	293	987	0	0
May-03	520	780	-260	433	1127	0	0
Jun-03	810	980	-170	633	1327	0	0
Jul-03	1120	1550	-430	1203	1897	1	0
Aug-03	1840	1850	-10	1503	2197	0	0
Sep-03	1600	1880	-280	1533	2227	0	0
Oct-03	1250	1600	-350	1253	1947	1	0
Nov-03	740	890	-150	543	1237	0	0
Dec-03	610	690	-80	343	1037	0	0
					Counts	2	0

2.13

Let S = sales and I = inventory. Then the cost will be:

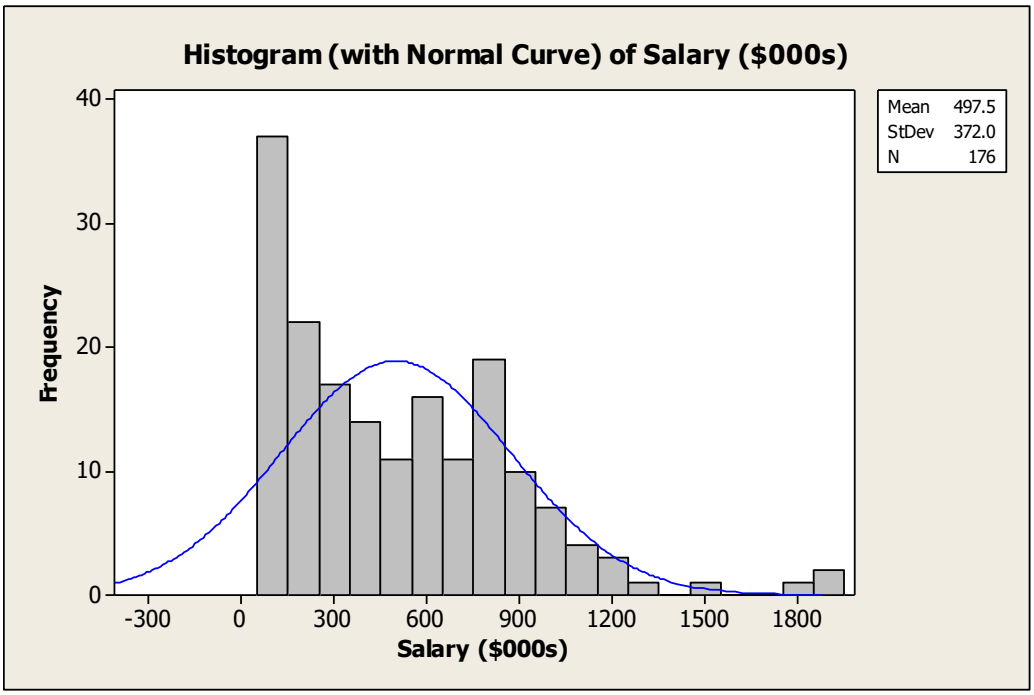
$$\begin{aligned} \text{Cost} &= C*(S-I), \text{if } S > I \\ &= (I-S), \text{ if } I > S \\ &= 0, \text{ if } I = S \end{aligned}$$

Given the probabilities, we can compute the expected cost for given values of C and I. The solution may be determined numerically by various means. As expected, when C=1, the optimal level of inventory to hold is I=100 and the total expected cost is 5.24. When C=3, the optimal level of inventory to hold is I=105 and the total expected cost is 7.86.

Minicases

The aim is to be creative – in all the minicases. These summary results are designed to provide basic information and not a complete solution.

Minicase 2.1: Baseball Salaries



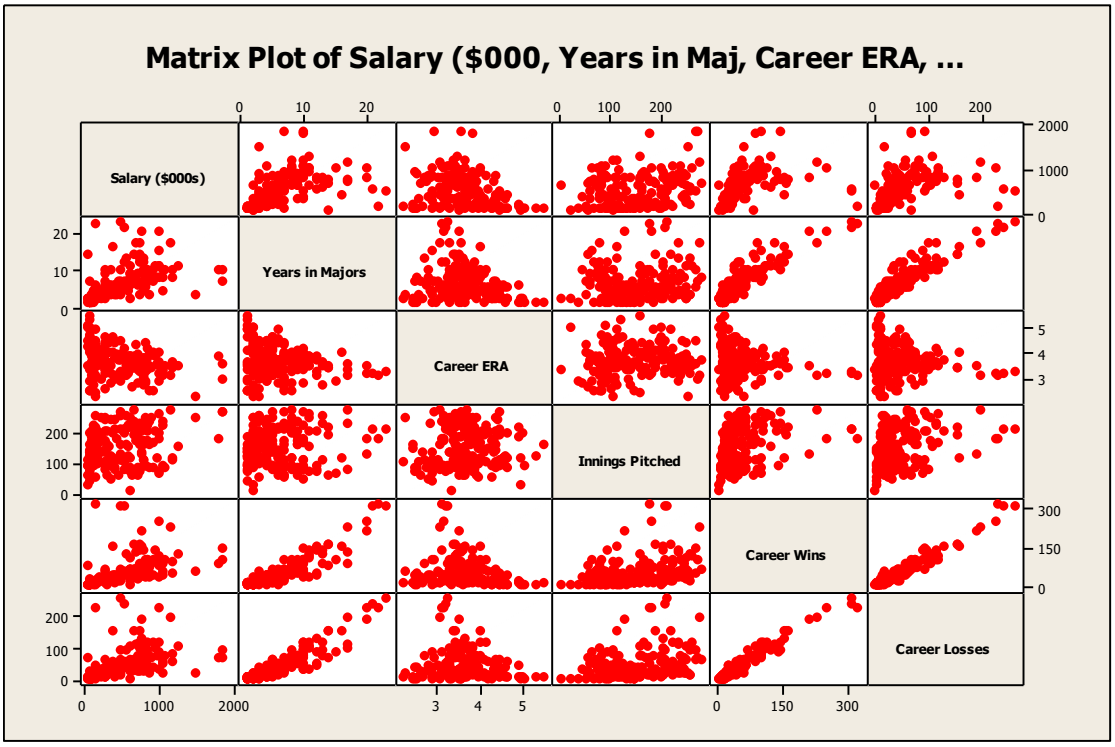
Descriptive Statistics: Salary ($000s)

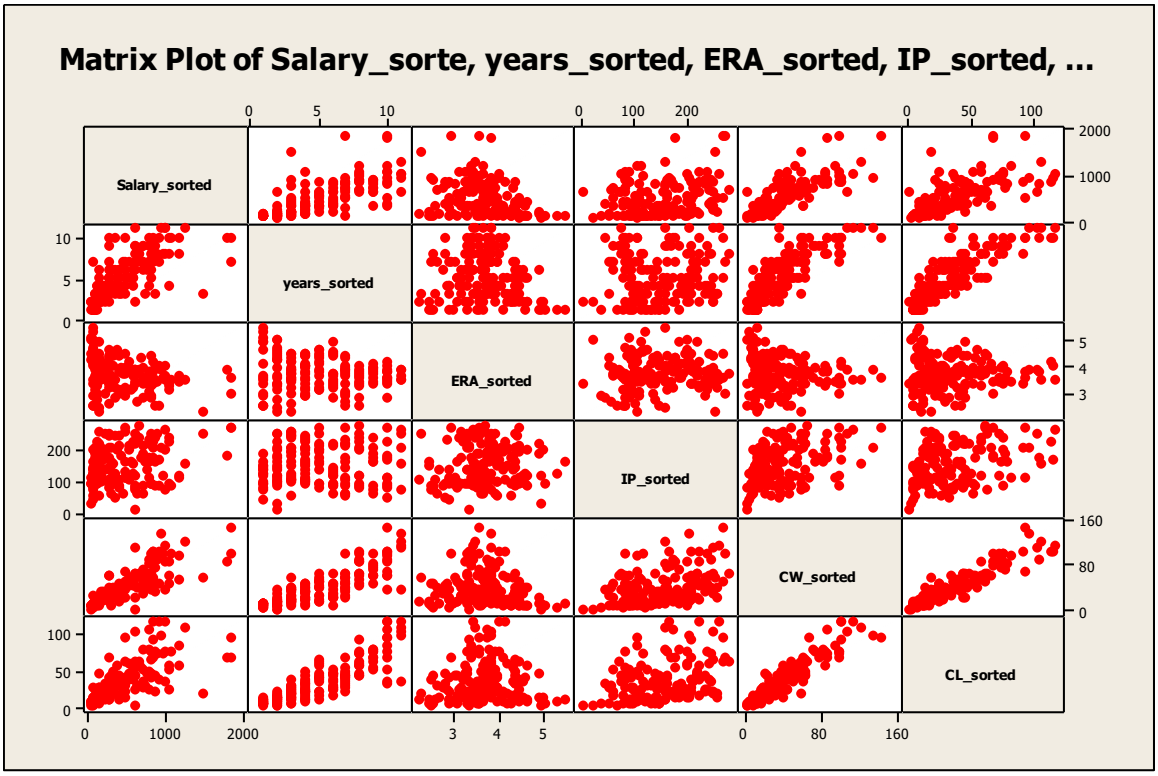
Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Salary ($000s)	176	497.5	372.0	62.5	156.3	417.5	768.8	1850.0

Use results to discuss shape of distribution, etc.

The matrix plot shows that there is a lot of variation even after allowing for the number of years played. Note the decline in salaries in later years.

When the long serving pitchers are removed the positive relationship between years and salary becomes clearer (higher correlation). Correlations that change substantially are highlighted.





Correlations: Salary ($000, Years in Maj, Career ERA, Innings Pitc, ...

	Salary ($000s)	Years in Majors	Career ERA
Years in Majors	0.535		
	0.000		
Career ERA	-0.337	-0.224	
	0.000	0.003	
Innings Pitched	0.265	0.107	0.087
	0.000	0.159	0.250
Career Wins	0.508	0.894	-0.213
	0.000	0.000	0.004
Career Losses	0.493	0.914	-0.138
	0.000	0.000	0.068

	Innings Pitched	Career Wins
Career Wins	0.328	
	0.000	
Career Losses	0.300	0.972
	0.000	0.000

Cell Contents: Pearson correlation
P-Value

Correlations: Salary_sorte, years_sorted, ERA_sorted, IP_sorted, CW_sorted, ...

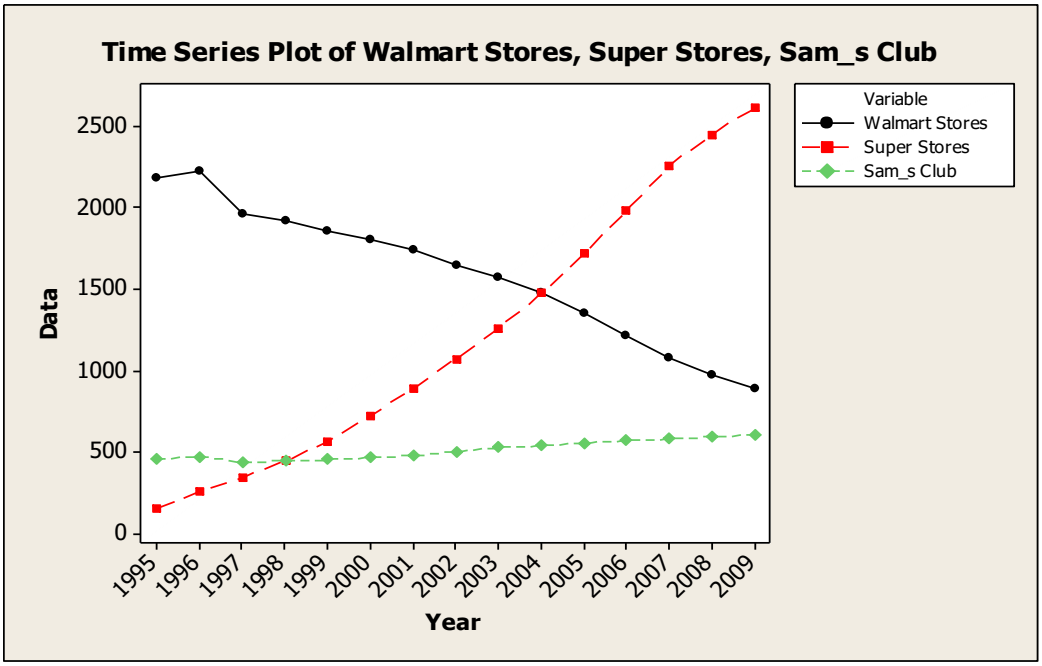
	Salary_sorted	years_sorted	ERA_sorted	IP_sorted
years_sorted	0.723			
	0.000			
ERA_sorted	-0.314	-0.132		
	0.000	0.100		
IP_sorted	0.272	0.086	0.101	
	0.001	0.287	0.212	
CW_sorted	0.824	0.827	-0.139	0.417
	0.000	0.000	0.083	0.000
CL_sorted	0.718	0.854	-0.001	0.347
	0.000	0.000	0.988	0.000

	CW_sorted
CL_sorted	0.931
	0.000

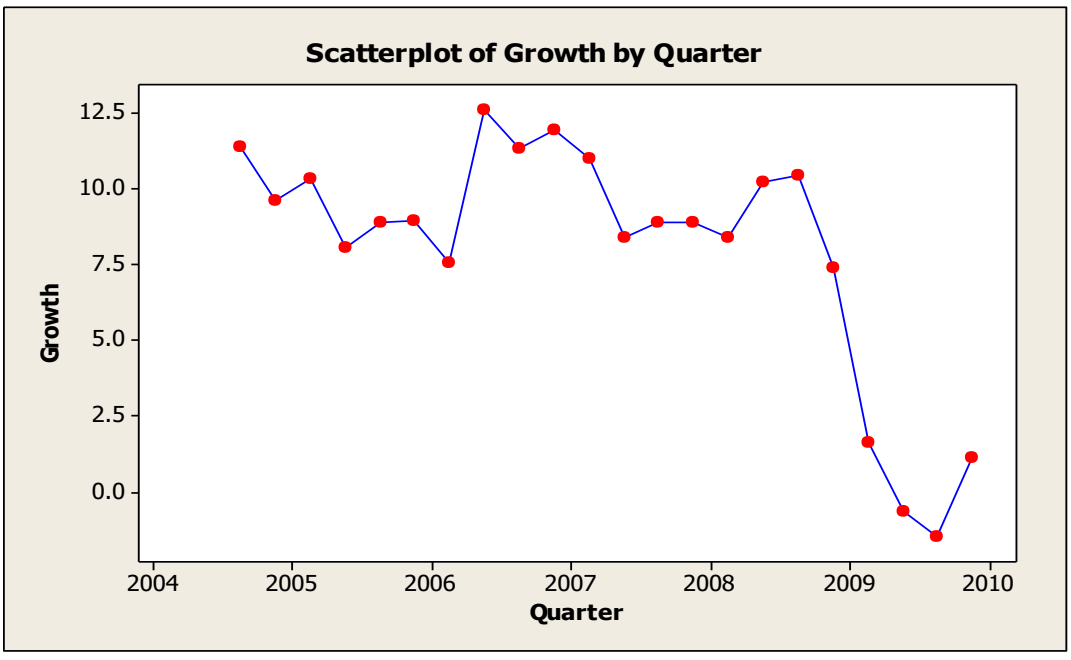
Cell Contents: Pearson correlation
P-Value

Minicase 2.2: Whither Walmart?

The plot shows clearly the movement towards Super Stores and away from the regular stores; some were doubtless conversions. The Sam's Club count is stable over the first half of the period but shows a steady growth in later years. These trends are likely to persist in later years.



As for many other companies, Walmart experienced a slowdown in growth as the result of the "Great Recession".



Minicase 2.3: Economic Recessions

Descriptive Statistics: Duration, Gap

```
Variable   N   Mean  SE Mean  StDev  Minimum  Median  Maximum
Duration  14  13.36     2.49   9.31     6.00   10.50    43.00
Gap       13  59.38     9.05  32.63    12.00   50.00   120.00
```

One-Sample T: Duration, Gap

```
Variable   N   Mean  StDev  SE Mean       95% CI
Duration  14  13.36   9.31     2.49  ( 7.98, 18.73)
Gap       13  59.38  32.63     9.05  (39.66, 79.10)
```

The length of the Great Depression from 1929 on greatly affects the analysis. Discuss the impact of omitting this observation. It is clear from the data that the "business cycle" can vary considerably in length, a point worth making in anticipation of the discussion of seasonality later.

There is essentially no evidence that the length of a growth period is related to the length of a recession either preceding or following it.

Correlations: Gap, Duration _after, Duration_before

```
                            Gap  Duration _after
Duration _after          -0.321
                          0.286

Duration_before           0.016           -0.040
                          0.960            0.896


Cell Contents: Pearson correlation
               P-Value
```