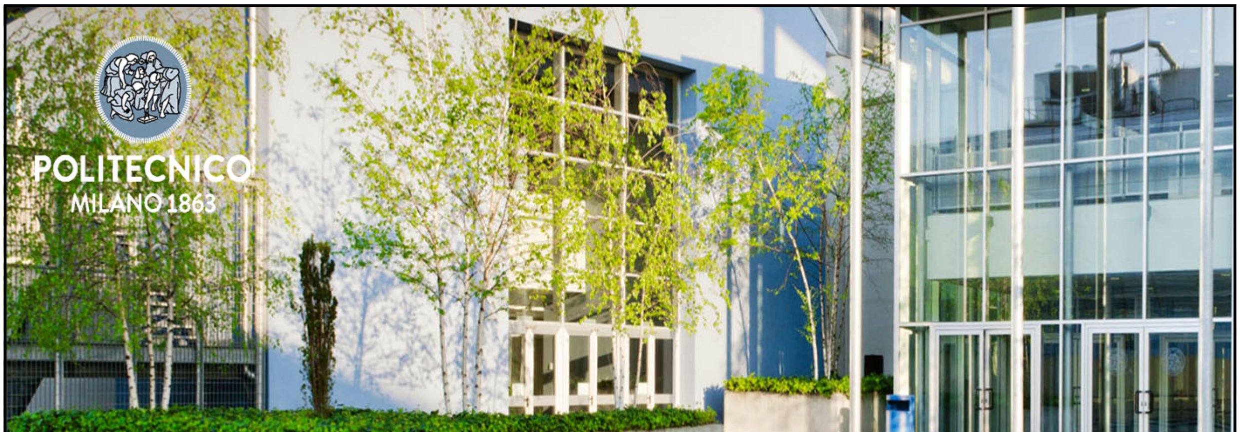


Quality Data Analysis

Control charts for variables - part 2

Bianca Maria Colosimo – biancamaria.colosimo@polimi.it

Reference:
Montgomery – Introduction to Statistical Quality Control

1

Control charts for variables and assumptions

Normality:

- *Xbar Control Chart*:
  - Known parameters: central limit theorem (sample mean is approx. Normal even though single observations are non-normal)
  - Unknwon parameters: we need an estimate of s based on R or s (R is better)
- R chart is more robust than S and S^2 charts with respect to small departures from normality

Solution: Box-Cox transformation on the original data

2

Control charts for variables and the assumptions: Non-random patterns

Pbm: A large variety of possible violations to random patterns do exist (linear and nonlinear trends):

- Trends
- Seasonal patterns
- Autocorrelation
- Other systematic patterns

and many possible combinations of the aforementioned features

- G.E.P. Box in chemical processes
- Montgomery and Friedman (1989) in manufacturing of integrated circuits
- Alwan and Bissel (1988) in clinical analyses

Alwan and Roberts (1995) in an empirical study on quality of products and services: systematic patterns represent **80%** of the real patterns observed in reality

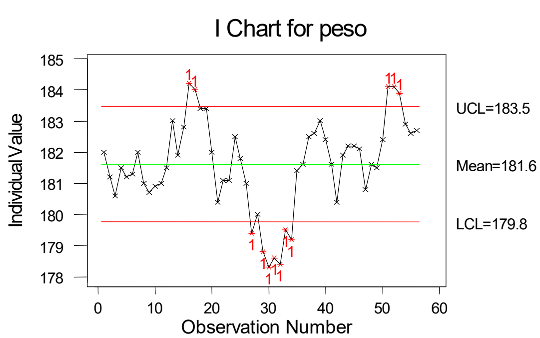
How does traditional SPC work in these cases?

Montgomery (1996): Among the [required] assumptions, the most important one is the independence assumption, … *Traditional control charts [Shewhart, CUSUM, EWMA] give unrealiable results when data arte correlated*

Example 1:

1. Daily weight – autocorreleted process

(weight.dat)



Three out of controls?

Looking for assignable causes - none

Advanced data monitoring vs Shewhart control chart

Main difference between advanced data monitoring and Shewhart approach

(Hoerl and Palm -1992)

Statistical modeling: ***"Fit the model to the process"***

Shewhart control charts: ***"Fit the process to the model"****, where*

$$model = NID$$

Advanced data monitoring

- Fit the model to the initial data and compute the residuals
- Design a control chart for the residuals. Residuals are differences between the forecast values (deterministic component of the model) and measured ones.
- If the model is correct, residuals are independent and identically (normally) distributed (when process is in-control).

- Two different 'charts' can be used:

  1. Fitted-Values Chart: it has no control limits; it shows just the fit versus the actual data to take a look to the non-random process pattern
  2. Special-Cause Chart: I-MR control chart on model residuals: it is used to monitor the random component of the process

Advanced data monitoring

Phase 1 vs Phase 2

Phase 1: consists of estimating the data model and parameters (indentifying and fitting the right model and estimate all the parameters)

Phase 2: model and parameters are assumed to be KNOWN and are just used to compute residuals and check if they are in control or not

Examples

Ex.1: Trend

Ex. 2: Non-linear trend + non-normality

Ex. 3: Autocorrelation (meandering process)

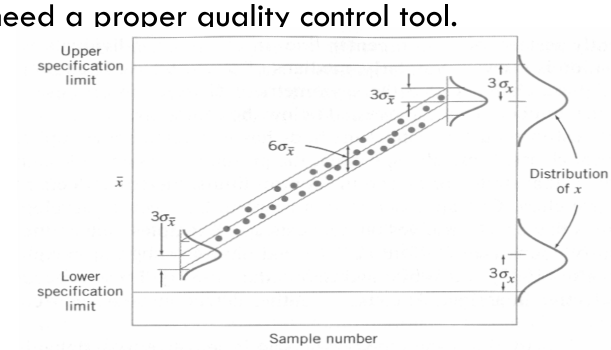
Ex. 4: Trend + autocorrelation

Ex. 1: Trend (Trend control chart)

Several processes, in practical applications, exhibit a systematic (and natural) change over time of the quality characteristic

- Examples: tool wear, continuous improvement, etc.

In these cases, we need a proper quality control tool.



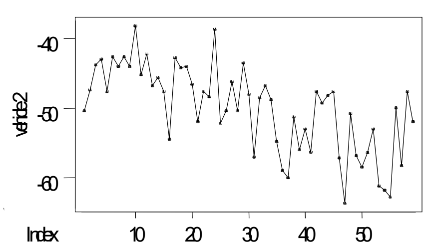
Example:

Camber measurements (in minutes) on one vehicle randomly selected in every production period

Table 5.4 Camber measurements

−50.4000	−47.4000	−43.8000	−43.0000	−47.6000	−42.6000	−44.0000
−42.6000	−44.0000	−38.2000	−45.2000	−42.2000	−46.8000	−45.6000
−47.6000	−54.5000	−42.8000	−44.2000	−44.0000	−46.6000	−52.0000
−47.6000	−48.4000	−38.6700	−52.2000	−50.4000	−46.2000	−50.4000
−43.5000	−48.0000	−57.0000	−48.5000	−46.8000	−48.8000	−54.8000
−59.0000	−60.0000	−51.2500	−56.0000	−53.0000	−56.3300	−47.6000
−49.2500	−48.2000	−47.6000	−57.2000	−63.6700	−50.8000	−56.8000
−58.5000	−56.4000	−53.0000	−61.2000	−61.8000	−62.8000	−50.0000
−58.2500	−47.6000	−52.0000				

(vehicle2.dat)



$$b_0 + b_1 t \pm 3\frac{\overline{MR}}{d_2(2)}$$

Model for process mean:

$b_0 + b_1 t$,

Contro limits for process mean:

$$b_0 + b_1 t \pm 3\frac{\overline{MR}}{d_2}$$

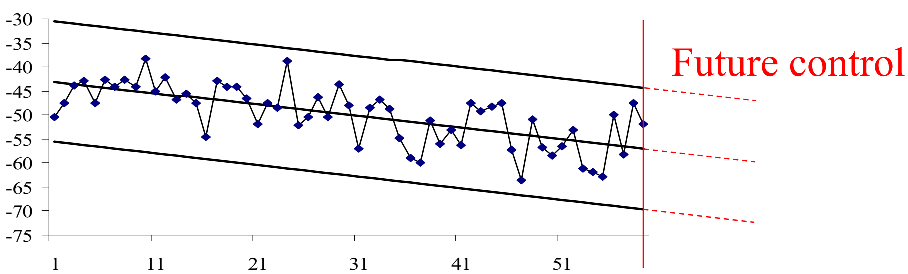
```
The regression equation is
vehicle 2 = - 42.8 - 0.241 t

Predictor        Coef     SE Coef        T       P
Constant      -42.844       1.204   -35.59   0.000
t            -0.24114     0.03490    -6.91   0.000

S = 4.565       R-Sq = 45.6%     R-Sq(adj) = 44.6%

Analysis of Variance

Source           DF          SS          MS        F        P
Regression        1      994.92      994.92    47.75    0.000
Residual Error   57     1187.73       20.84
Total            58     2182.65
```


If we try to detrend the process: control limits for future periods may indicate the success of the detrending operation

$$UCL = -42.8 - 0.241\,t + 3\left(\frac{4.748}{1.128}\right) = \text{-}30.2\text{-}0.241\,t$$

$$CL = -42.8 - 0.241\,t$$

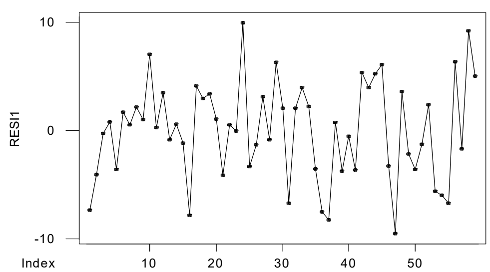
$$LCL = -42.8 - 0.241\,t - 3\left(\frac{4.748}{1.128}\right) = \text{-}55.4\text{-}0.241\,t$$



Fitted-Values Chart & Special-Cause Chart

Consider the previous example: let's analyse the residual time series.

Residuals:



Verify the goodness of the trend model. Check assumptions on residuals:

- Independence
- Normality


```
residual
K =     0.0000

The observed number of runs  =  30
The expected number of runs  =  30.4237
31 Observations above K 28 below
    The test is significant at 0.9112
    Cannot reject at alpha   =  0.05
```

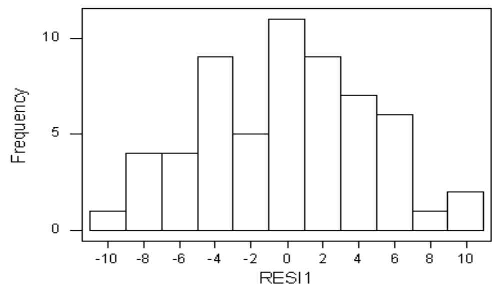
(a) **Runs test**

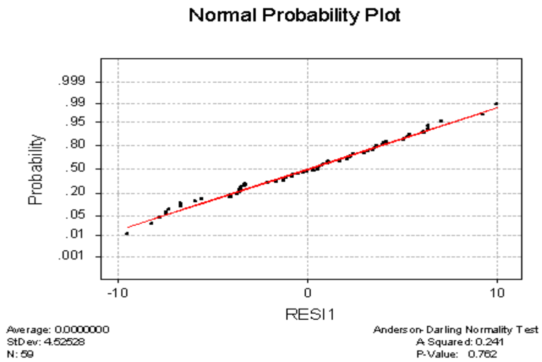
Check the absence of systematic patterns in the residual time series (IID assumption)

```
ACF of residual

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
 1   0.115                               XXXX
 2  -0.001                               X
 3  -0.096                             XXX
 4  -0.058                              XX
 5  -0.042                              XX
 6  -0.032                              XX
 7  -0.155                            XXXXX
 8  -0.163                            XXXXX
 9  -0.148                            XXXXX
10   0.141                               XXXXX
11  -0.073                             XXX
12  -0.093                             XXX
13  -0.083                             XXX
14   0.201                               XXXXXX
15   0.194                               XXXXXX
```

(b) **ACF**

The ACF shows that no significant correlation exists (first 15 lags)



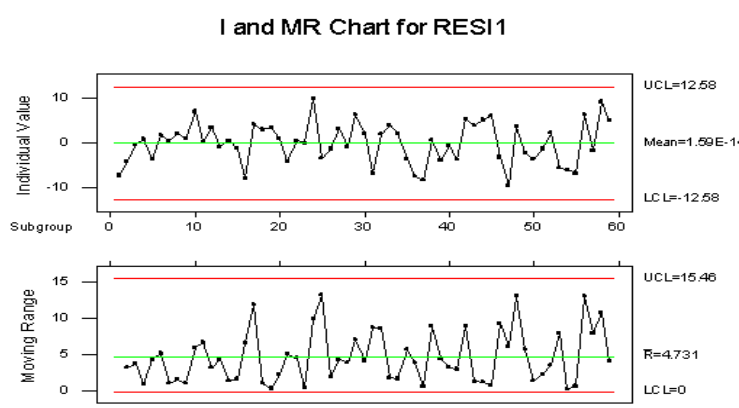


Verify that residuals are normally distributed

Anderson-Darling Test:
p-value=0.762

Fitted-Values Chart & Special-Cause Chart

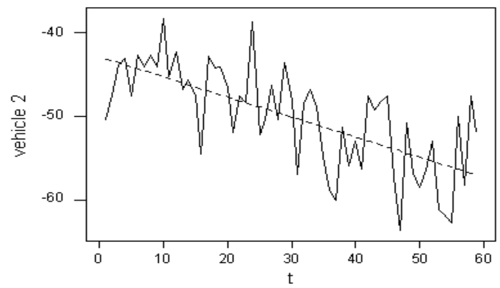
Since residuals are compliant with the Shewhart's control chart assumptions, we can design an I-MR chart on residuals (SCC):



Note: the MR chart helps one to check if 'homoscedasticity' assumption (constant variance) is met on the random component of the model

Fitted-Values Chart & Special-Cause Chart

We can design the **fitted-values chart (FVC)** on the same data:



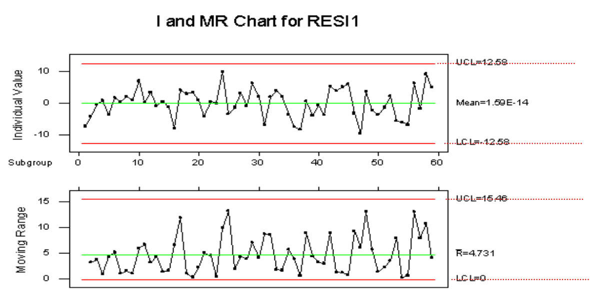
Both the control charts work by applying a retrospective analysis to the process.

Indeed, they rely on values (fitted values or residuals) computed *after* process observation.

The FVC provides a tool to determine the natural process behaviour (deterministic component) that may be helpful to improve the process itself.
The SCC provides a tool for process improvement, because it highlights any possible special cause that is not related with the natural process behaviour.

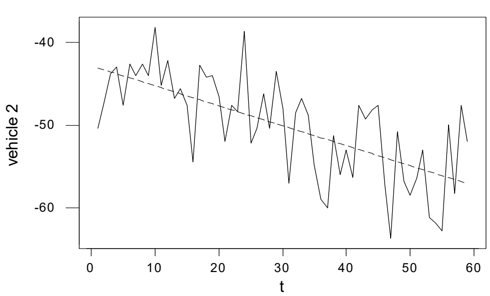
In conclusion: FVC+SCC:

SCC



Future control limits on residuals

FVC



Benefits:

✓ Additional out-of-control detection criteria can be used (e.g., run-rules)
✓ Two clearly distinct effects
✓ Suitable approach for any kind of systematic pattern

Ex 2. *I* control chart in the presence of non-linear trend and non-normality

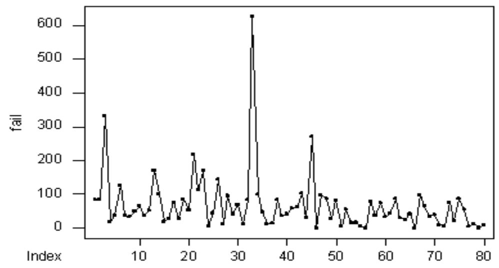
Example: time between computer failures

Time (in hours) between crashes of an information system of a Mid-West bank (USA)

(failure.dat)

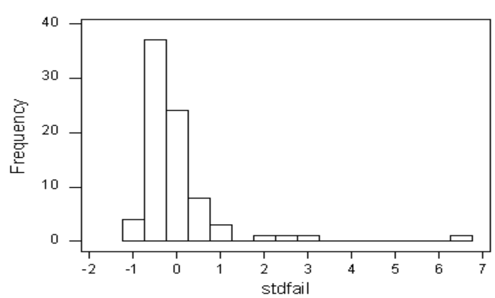
Table 5.5 Time between computer failures

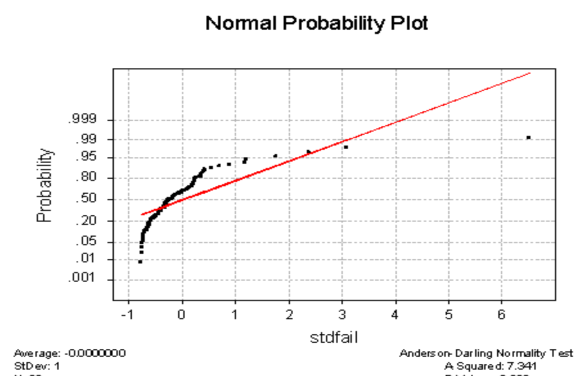
83.483	86.267	331.750	17.783	37.967	126.417	38.917
32.533	50.467	64.534	38.700	51.267	170.390	100.640
19.683	28.683	74.817	27.667	85.500	54.083	217.583
113.550	168.200	5.867	44.400	142.600	12.567	95.917
40.883	68.933	13.500	84.000	624.819	99.150	49.083
13.083	14.450	83.883	36.550	40.950	58.750	61.917
103.050	30.283	270.000	1.233	97.183	86.883	28.717
81.817	3.800	55.483	15.633	15.417	4.833	1.000
78.400	37.683	73.467	32.617	43.833	86.650	29.350
24.000	42.000	1.500	97.500	65.750	34.083	39.167
8.750	5.250	75.917	22.483	88.100	54.500	4.667
12.233	1.183	9.667				



Positive asymmetry seems to be present: Normality assumption is verified?

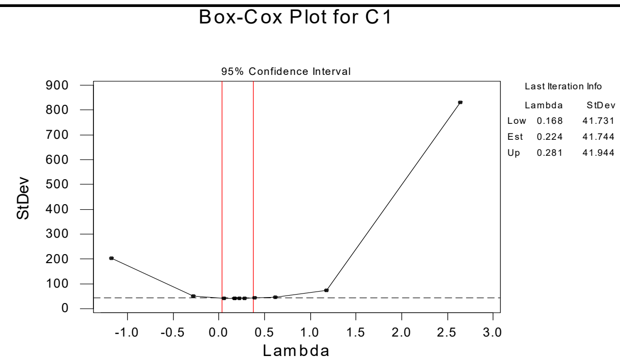
Standardized data: $\dfrac{x - \overline{x}}{s_x}$



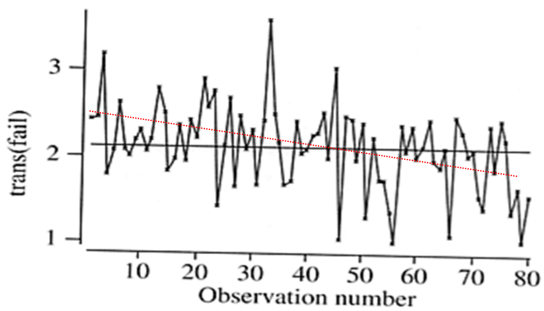


p-Value: 0.000

Remind: if time between two occurrences of a given event follows a Poisson distrib. –memoryless processes: the time between two occurrences has an exponential distrib. → look for transformation



$$trans(fail) = (failure_t)^{0.2} = \sqrt[5]{failure_t}$$



Trend?
It seems linear.
Anyway, try stepwise regression with t, t^2 and 1/t

Quality Data Analysis

26

26

Step	1
Constant	2.327
t**2	-0.00010
T-Value	-3.62
P-Value	0.001
S	0.455
R-Sq	14.35
R-Sq(adj)	13.26
C-p	0.3

With linear trend

Regression Analysis: trans(fail) versus t2**

The regression equation is

trans(fail) = 2.33 -0.000095 t**2

Predictor	Coef	SE Coef	T	P
Constant	2.32724	0.07656	30.40	0.000
t**2	-0.00009523	0.00002634	-3.62	0.001

S = 0.4547 R-Sq = 14.4% R-Sq(adj) = 13.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2.7030	2.7030	13.07	0.001
Residual Error	78	16.1284	0.2068		
Total	79	18.8313			

Regression Analysis: trans(fail) versus t

The regression equation is

trans(fail) = 2.43 - 0.00770 t

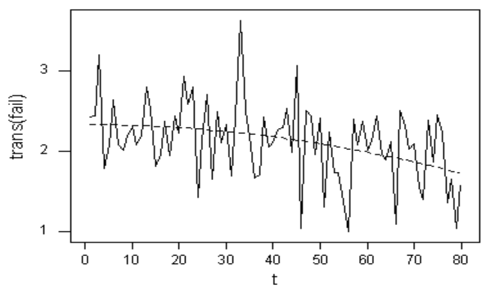
Predictor	Coef	SE Coef	T	P
Constant	2.4321	0.1032	23.57	0.000
t	-0.007699	0.002213	-3.48	0.001

S = 0.4572 R-Sq = 13.4% R-Sq(adj) = 12.3%

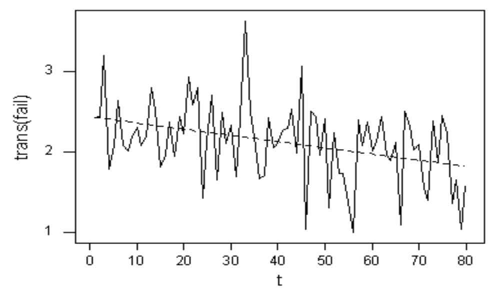
Quality Data Analysis- BM Colo

27

FVC (on transformed data)



quadratic



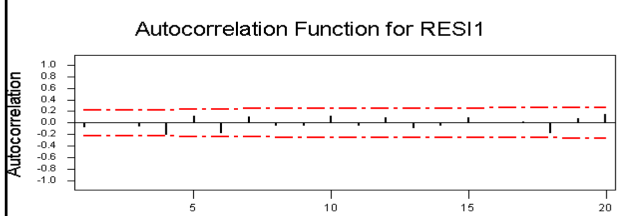
linear

28

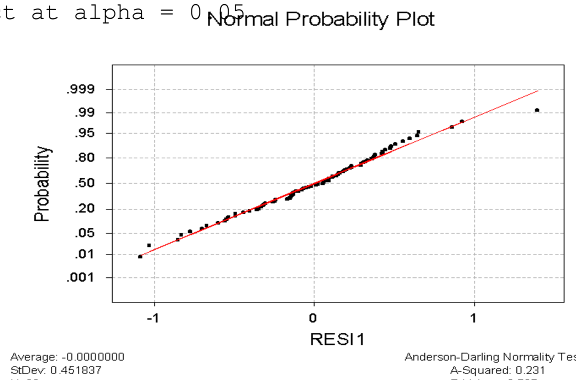
```
trans(fail) = 2.33 -0.000095 t**2
```

Runs Test: RESI1

```
RESI1
K =     -0.0000
The observed number of runs =  40
The expected number of runs =  40.9000
42 Observations above K   38 below
          The test is significant at  0.8391
          Cannot reject at alpha = 0.05
```

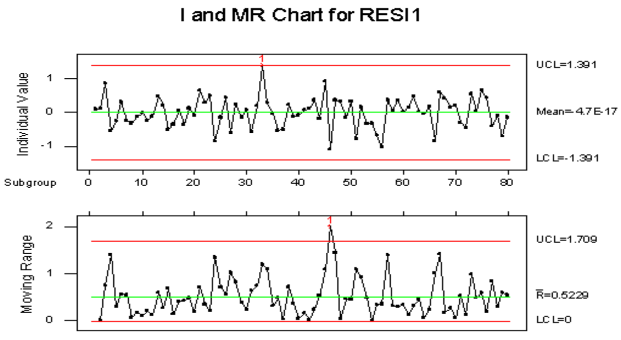


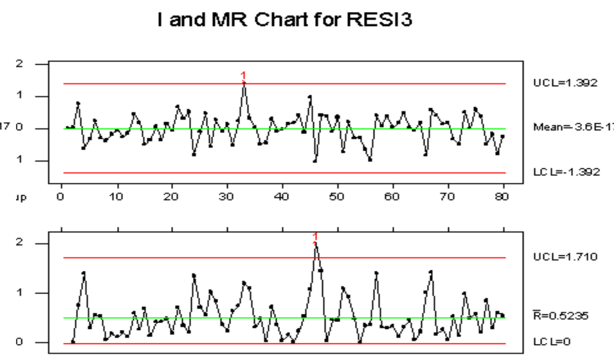
Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.09	-0.78	0.64	8	-0.06	-0.52	10.63	15	0.08	0.64	15.30
2	-0.01	-0.08	0.65	9	-0.06	-0.51	11.00	16	-0.02	-0.14	15.33
3	-0.07	-0.66	1.12	10	0.11	0.88	12.13	17	0.01	0.11	15.35
4	-0.21	-1.87	4.99	11	-0.06	-0.46	12.45	18	-0.19	-1.49	19.27
5	0.12	1.00	6.22	12	0.09	0.72	13.25	19	0.08	0.57	19.89
6	-0.19	-1.59	9.43	13	-0.10	-0.80	14.26	20	0.15	1.11	22.29
7	0.10	0.78	10.26	14	-0.06	-0.47	14.62				



29

With linear trend





Nothing changes

Assignable cause found for out-of-control in *I chart*: define independent variable: special cause whose value is 1 at time 33 and 0 otherwise:

Box & Tiao (1975): how to model a shock and other anomalies in time series (by means of a tranfer function). Shock=impulse

Stepwise Regression: trans(fail) versus t, 1/t, t2, special**

```
 Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05

Response is trans(fa on  4 predictors, with N =   80

   Step          1        2
Constant     2.327    2.298

t**2      -0.00010 -0.00009
T-Value      -3.62    -3.62
P-Value      0.001    0.001

special               1.42
T-Value               3.30
P-Value              0.001

S            0.455    0.428
R-Sq         14.35    24.94
R-Sq(adj)    13.26    22.99
C-p           10.1      1.4
```

Regression Analysis: trans(fail) versus t2, special**

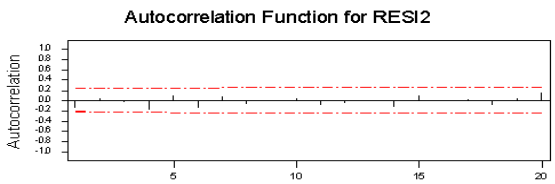
The regression equation is

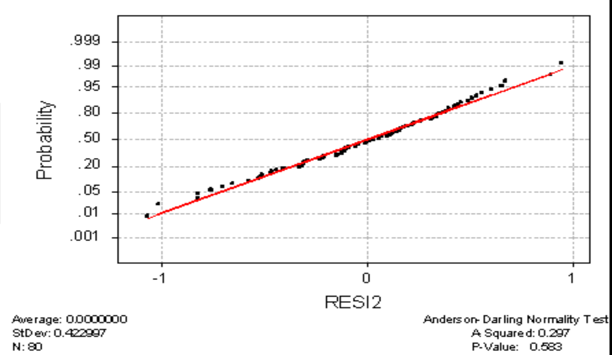
trans(fail) = 2.30 -0.000090 t2 + 1.42 special**

```
Predictor         Coef      SE Coef       T       P
Constant       2.29819      0.07268   31.62   0.000
t**2       -0.00009005   0.00002487   -3.62   0.001
special         1.4236       0.4320    3.30   0.001

S = 0.4285      R-Sq = 24.9%     R-Sq(adj) = 23.0%

Analysis of Variance

Source          DF        SS        MS       F       P
Regression       2    4.6961    2.3481   12.79   0.000
Residual Error  77   14.1352    0.1836
Total           79   18.8313
```

Runs Test: RESI2

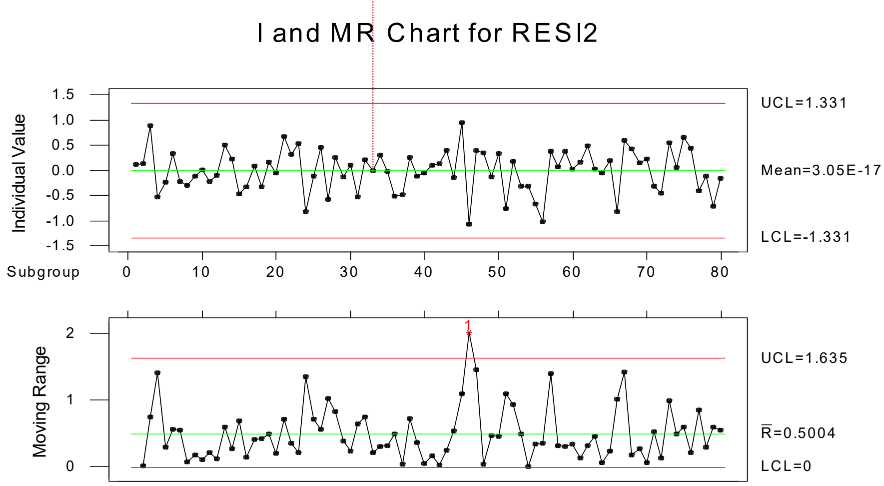
```
K =      0.0000
The observed number of runs =  44
The expected number of runs =  40.9000
42 Observations above K   38 below
             The test is significant at  0.4843
             Cannot reject at alpha = 0.05
```

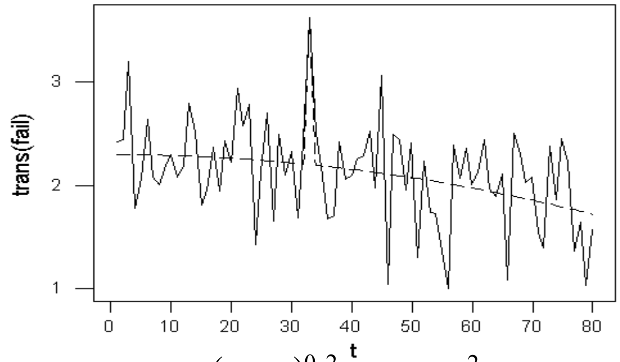




Special Cause Chart (after removing the datum with assignable cause):

Res $_{33}$=0: special cause

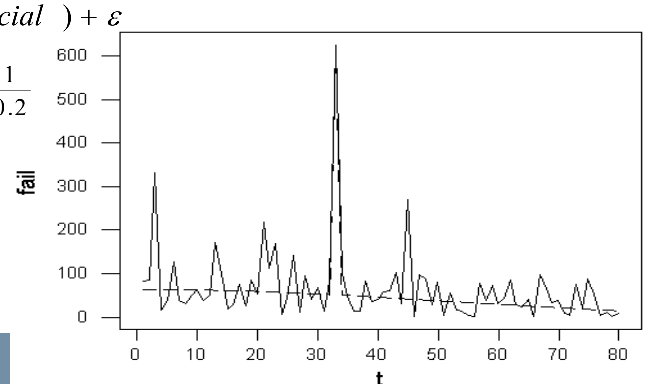




Fitted Value Chart:
For a better interpretation, let’s look back to original data (before transformation)

$$trans\,(fail)_t = (fail_t)^{0.2} = fit2(t^2, special) + \varepsilon$$

$$\left[(fail_t)^{0.2}\right]^{\frac{1}{0.2}} = \left[fit2(t^2, special) + \varepsilon\right]^{\frac{1}{0.2}}$$



Attention: time between crashes has been decreasing!

Remark:

If after control chart design on residuals there is an out of control with assignable cause: the observation shall be removed

– Re-estimation of linear regression coefficients

– Control chart design on new residuals

Ex. 3. Autocorrelation (meandering process)

Steel production plant: 90 measures of phosphorus percentage in consecutive lots

(steel1.dat)

Table 5.6 Percentage of phosphorus series

0.045123	0.039692	0.042499	0.039905	0.039580	0.044515
0.042653	0.050871	0.051175	0.046192	0.045998	0.045287
0.043689	0.045428	0.037653	0.037126	0.038197	0.041416
0.039248	0.038642	0.042793	0.041774	0.041212	0.042322
0.043960	0.043397	0.049041	0.041046	0.040260	0.044297
0.036117	0.039916	0.044106	0.046558	0.041433	0.046357
0.039078	0.042592	0.043579	0.045991	0.045503	0.045529
0.043794	0.043867	0.040168	0.038482	0.041719	0.045132
0.039334	0.039720	0.038269	0.045140	0.044500	0.044594
0.046301	0.042862	0.043058	0.041517	0.042217	0.035953
0.038042	0.039301	0.043625	0.045408	0.047051	0.045239
0.041177	0.042900	0.048665	0.048620	0.050647	0.050802
0.046519	0.045999	0.046880	0.042557	0.039337	0.037700
0.038583	0.038924	0.041805	0.043915	0.038175	0.041224
0.041572	0.042356	0.037447	0.039595	0.046171	0.045421

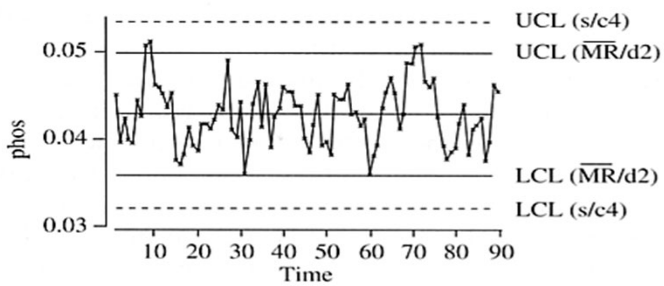


Figure 5.28 X chart for the phosphorus series with standard control limits.

Is there a unique explanation for the systematic pattern?

```
phos
K = 0.0428
The observed number of runs  =  27
The expected number of runs  =  45.9778
44 Observations above K  46 below
   The test is significant at 0.0001
```

(a) Runs test

It looks like an autocorrelated process, but a *stationary* one (stable mean)

PACF

```
ACF of phos
        -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
          +----+----+----+----+----+----+----+----+----+----+
 1   0.518                           XXXXXXXXXXXXXX
 2   0.294                           XXXXXXXX
 3   0.128                           XXXX
 4  -0.013                           X
 5  -0.101                        XXXX
 6  -0.125                        XXXX
 7  -0.201                      XXXXXX
 8  -0.222                     XXXXXXX
 9  -0.203                      XXXXXX
10  -0.224                     XXXXXXX
11  -0.185                      XXXXXX
12  -0.172                       XXXXX
13  -0.090                         XXX
14  -0.035                          XX
15   0.010                           X
16   0.082                           XXX
17   0.078                           XXX
18   0.078                           XXX
19  -0.006                           X
20  -0.071                         XXX
21  -0.069                         XXX
22  -0.132                        XXXX
```

(b) ACF

```
PACF of phos
        -1.0 -0.8 -0.6 -0.4 -0.2 0.0  0.2  0.4  0.6  0.8  1.0
         +----+----+----+----+----+----+----+----+----+----+
 1  0.518                          XXXXXXXXXXXXXX
 2  0.035                          XX
 3 -0.050                         XX
 4 -0.095                        XXX
 5 -0.074                        XXX
 6 -0.028                         XX
 7 -0.131                       XXXX
 8 -0.077                        XXX
 9 -0.045                         XX
10 -0.113                       XXXX
11 -0.043                         XX
12 -0.092                        XXX
13  0.016                          X
14 -0.026                         XX
15 -0.028                         XX
16  0.042                          XX
17 -0.056                         XX
18 -0.016                          X
19 -0.139                       XXXX
20 -0.106                       XXXX
21 -0.018                          X
22 -0.155                      XXXXX
```

AR(1)

```
The regression equation is
phos = 0.0205 + 0.521 phos-1

89 cases used 1 cases contain missing values

Predictor        Coef      StDev        T       P
Constant     0.020486   0.003931     5.21   0.000
phos-1        0.52124    0.09158     5.69   0.000

S = 0.003039    R-Sq = 27.1%    R-Sq(adj) = 26.3%

Analysis of Variance

Source            DF          SS          MS       F       P
Regression         1  0.00029917  0.00029917   32.39   0.000
Residual Error    87  0.00080345  0.00000924
Total             88  0.00110261
```

Adjustment? Assume target=0.045 and:

- departure from target determines computable loss (e.g., quadratic?)
- adjustment cost (to get back to target value) is known

It is possible to compute values for which we have expected loss = adjustment cost - ACTION LIMITS

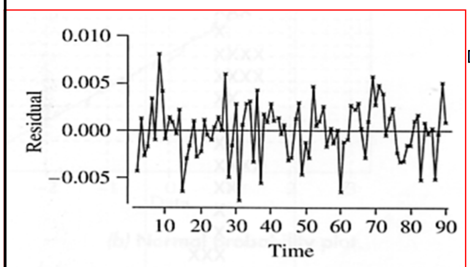
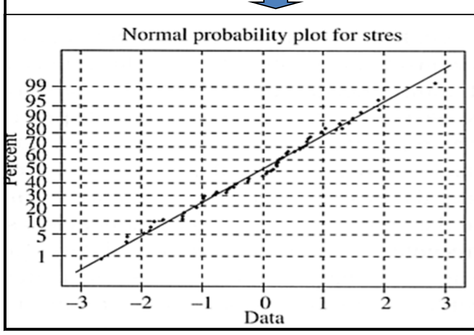


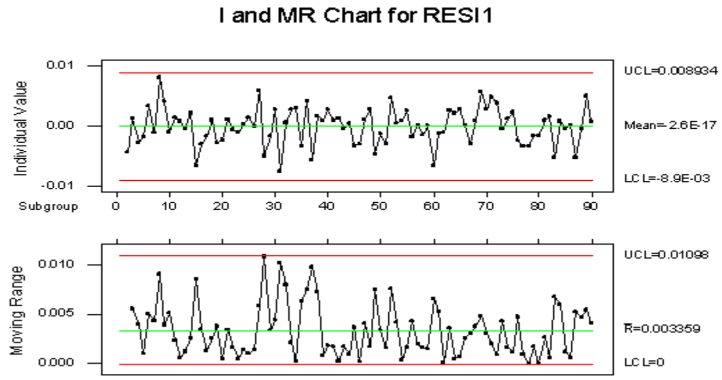
Figure 5.32 Time-series plot of the residuals from the fitted lag 1 model for the phosphorus series.



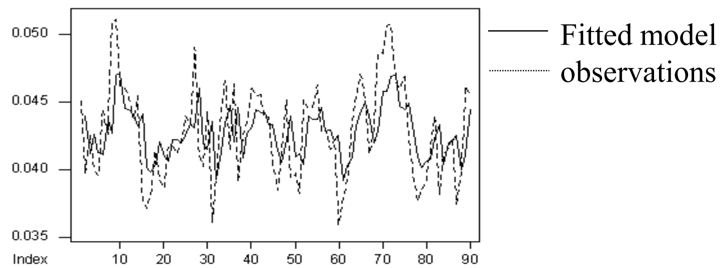

```
residual

K =     -0.0000

The observed number of runs   =  50
The expected number of runs   =  45.2247
48 Observations above K   41 below
   The test is significant at 0.3056
   Cannot reject at alpha = 0.05
```

```
ACF of residual
        -1.0 -0.8 -0.6 -0.4 -0.2 0.0  0.2  0.4  0.6  0.8  1.0
         +----+----+----+----+----+----+----+----+----+----+
 1 -0.007                          X
 2  0.055                          XX
 3  0.030                          XX
 4 -0.046                         XX
 5 -0.087                        XXX
 6  0.000                          X
 7 -0.125                       XXXX
 8 -0.119                       XXXX
 9 -0.030                         XX
10 -0.116                       XXXX
11 -0.046                         XX
12 -0.105                       XXXX
13 -0.021                         XX
14  0.009                          X
15 -0.012                          X
16  0.090                          XXX
17  0.021                          XX
18  0.092                          XXX
19 -0.002                          X
20 -0.065                        XXX
21  0.030                          XX
22 -0.056                         XX
```

Residuals (SCC)



FVC



Note: Also in this case we could apply a trend control chart – like approach:

$$0.0205 + 0.521 \text{ phos}_{t-1} \pm 3\frac{\overline{MR}_{res}}{d_2} \rightarrow \text{Note: residuals}$$

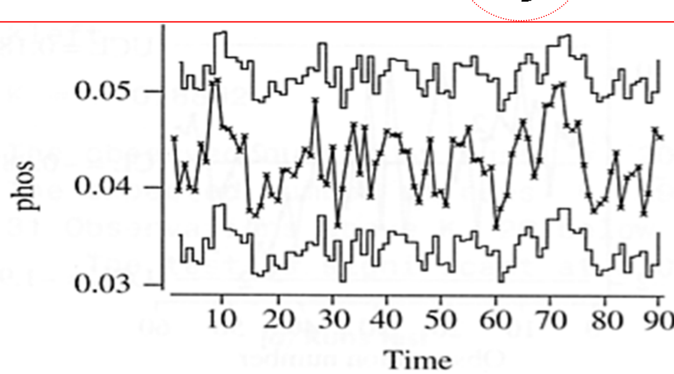
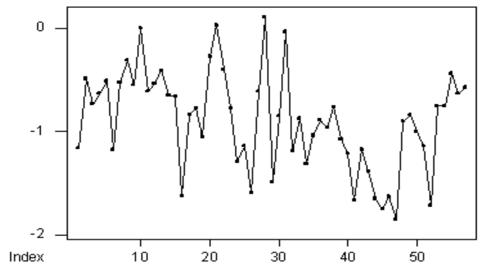


Figure 5.37 Phosphorus series with time-varying control limits.

Data from vehicle component (measurement related to stability)

(vehicle3.dat)



→ Trend?

```
kleft

K =  -0.8882

The observed number of runs  =  20
The expected number of runs  =  29.2807
31 Observations above K  26 below
   The test is significant at 0.0124
```

ACF+PACF ⇨ AR(1)

42

Let's try: kleft $_{t-1}$, t , t^2 and 1/t:

Stepwise Regression: kleft versus t, t2, 1/t, kleft_1

```
  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

    Step            1        2
Constant      -0.5381  -0.3860

t             -0.0117  -0.0084
T-Value         -3.19    -2.17
P-Value         0.002    0.035

kleft_1                   0.28
T-Value                   2.11
P-Value                  0.040

S               0.444    0.431
R-Sq            15.83    22.33
R-Sq(adj)       14.27    19.39
```

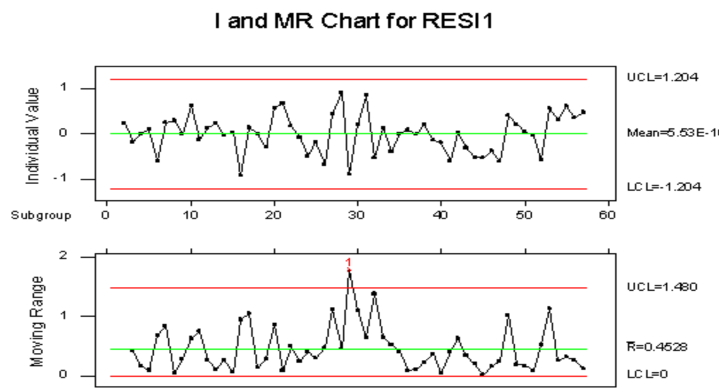
```
The regression equation is
kleft = -0.386 + 0.279 kleft-1 - 0.00842 t

56 cases used 1 cases contain missing values

Predictor          Coef       StDev        T        P
Constant        -0.3860      0.1399    -2.76    0.008
kleft-1          0.2785      0.1323     2.11    0.040
t             -0.008419    0.003888    -2.17    0.035

S = 0.4308      R-Sq = 22.3%      R-Sq(adj) = 19.4%

Analysis of Variance

Source            DF          SS         MS        F        P
Regression         2      2.8267     1.4133     7.62    0.001
Residual Error    53      9.8346     0.1856
Total             55     12.6613
```

Checking assumptions on residuals (they result to be NID)

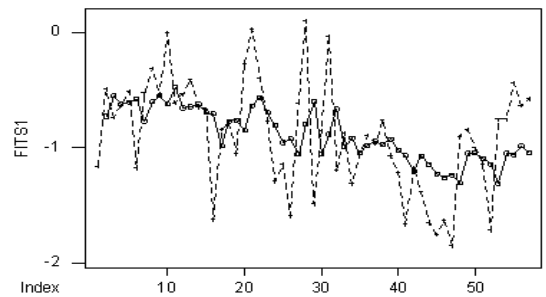
43

SCC

(no assignable cause found for out-of-control in MR chart)



FVC



Control charts for variables and assumptions

Non-random pattern (process is not IID)?

- For I-MR chart (between observations)
- For chart with n>1: non-random pattern **within** the sample

Example 1: control chart for process mean

Quality control manual, Ishikawa (1986):

Humidity content of a textile product measured at (5) regular intervals:

6:00 10:00 14:00 18:00 22:00 on 25 consecutive days

(ishikawa.dat)

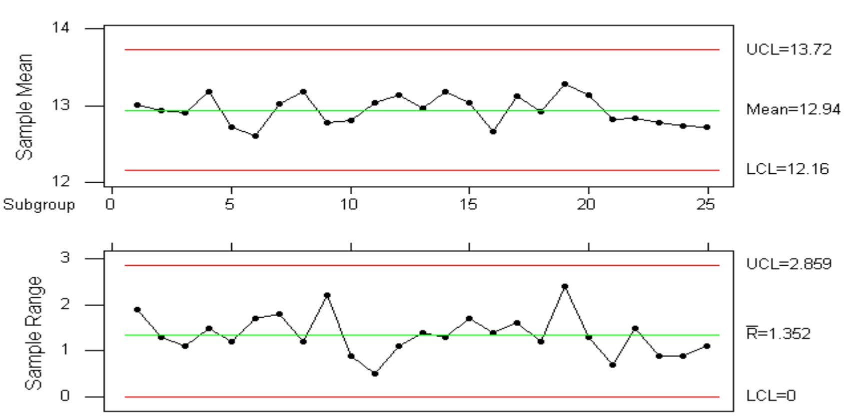
subgroup	6:00	10:00	14:00	18:00	22:00	mean	range
1	14.0	12.6	13.2	13.1	12.1	13.00	1.9
2	13.2	13.3	12.7	13.4	12.1	12.94	1.3
3	13.5	12.8	13.0	12.8	12.4	12.90	1.1
4	13.9	12.4	13.3	13.1	13.2	13.18	1.5
5	13.0	13.0	12.1	12.2	13.3	12.72	1.2
6	13.7	12.0	12.5	12.4	12.4	12.60	1.7
…							
20	13.9	13.0	13.0	13.2	12.6	13.14	1.3
21	13.3	12.7	12.6	12.8	12.7	12.82	0.7
22	13.9	12.4	12.7	12.4	12.8	12.84	1.5
23	13.2	12.3	12.6	13.1	12.7	12.78	0.9
24	13.2	12.8	12.8	12.3	12.6	12.74	0.9
25	13.3	12.8	13.0	12.3	12.2	12.72	1.1

$\bar{\bar{x}} = 12.94$ $\bar{R} = 1.352$ $\hat{\sigma} = \dfrac{\bar{R}}{d_2(5)} = \dfrac{1.352}{2.326} = 0.5812$

$$\text{UCL} = \bar{\bar{x}} + A_2(n)\bar{R} = 12.94 + 0.577(1.352) = 13.72$$
$$\text{CL} = \bar{\bar{x}} = 12.94$$
$$\text{LCL} = \bar{\bar{x}} - A_2(n)\bar{R} = 12.94 - 0.577(1.352) = 12.16$$

$$\text{UCL} = D_4(n)\bar{R} = 2.114(1.352) = 2.858$$
$$\text{CL} = \bar{R} = 1.352$$
$$\text{LCL} = D_3(n)\bar{R} = 0$$

Xbar/R Chart for ishikawa



→ "hugging" or stratification

↓

Common cause: Systematic pattern within the sample

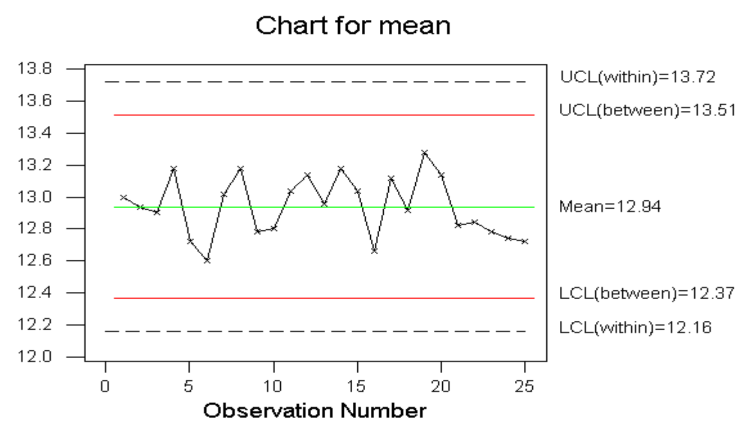
$$\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{\overline{R}/d_2(5)}{\sqrt{n}} = 0.2599$$

Process means regarded to as individuals $: Y = \overline{X}$

$$\hat{\sigma}_Y = \frac{s_Y}{c_4(25)} = \frac{0.1892}{0.9897} = 0.1912$$

$$UCL = \overline{\overline{x}} + 3\hat{\sigma}_Y = 13.51$$

$$LCL = \overline{\overline{x}} - 3\hat{\sigma}_Y = 12.37$$



Quality Data Analysis

48

48

```
ACF of ishikawa
          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
 1  -0.083                        XXX
 2  -0.145                      XXXXX
 3   0.005                          X
 4  -0.104                       XXXX
 5   0.371                          XXXXXXXXXX
 6   0.001                          X
 7  -0.087                        XXX
 8  -0.155                      XXXXX
 9  -0.078                        XXX
10   0.310                          XXXXXXXXX
11  -0.156                      XXXXX
12  -0.074                        XXX
13  -0.154                      XXXXX
14  -0.154                      XXXXX
15   0.454                          XXXXXXXXXXXX
…
20   0.284                          XXXXXXXX
 …
25   0.369                          XXXXXXXXXX
…
29  -0.147                      XXXXX
30   0.359                          XXXXXXXXXX
31  -0.051                         XX
```

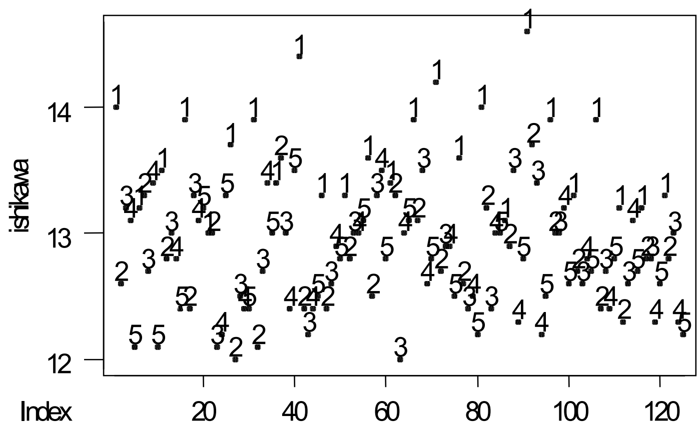
$$\pm 2/\sqrt{25(5)} = 0.1789$$

Period: 5

Quality Data Analysis

49

49



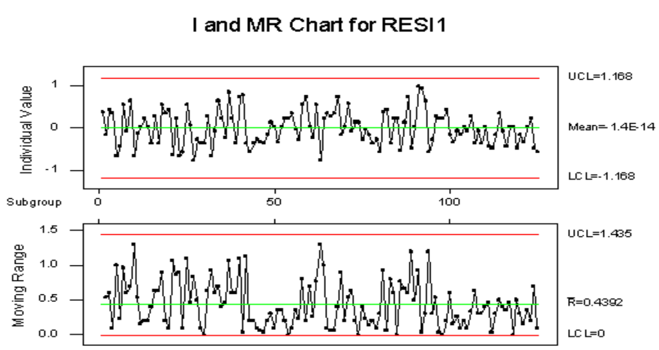
First obs. (1) is the largest one in the sample for 21 days in 25 (expected value (1/5)25=5 gg)

Insert seasonality index:

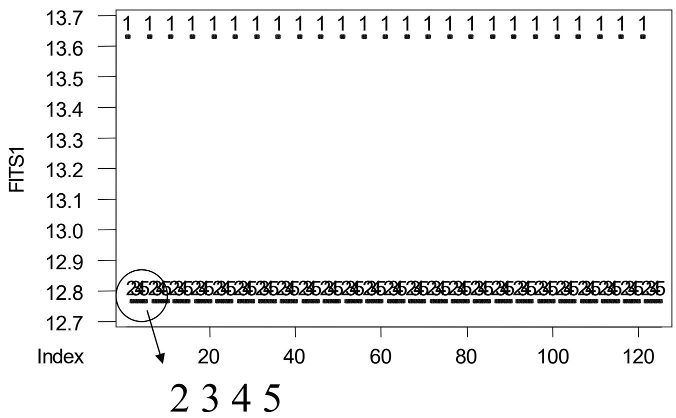
Independent variable *6hours* (=1 if the observation is the first one in the sample, 0 otherwise) and estimate the deterministic component of the model via regression

Regression Analysis: ishikawa versus ore6

```
The regression equation is
ishikawa = 12.8 + 0.865 ore6
Predictor        Coef     SE Coef          T        P
Constant      12.7670      0.0408     312.61    0.000
ore6          0.86500     0.09132       9.47    0.000    p-value<0.0005

S = 0.4084      R-Sq = 42.2%     R-Sq(adj) = 41.7%
```



I-MR on residuals

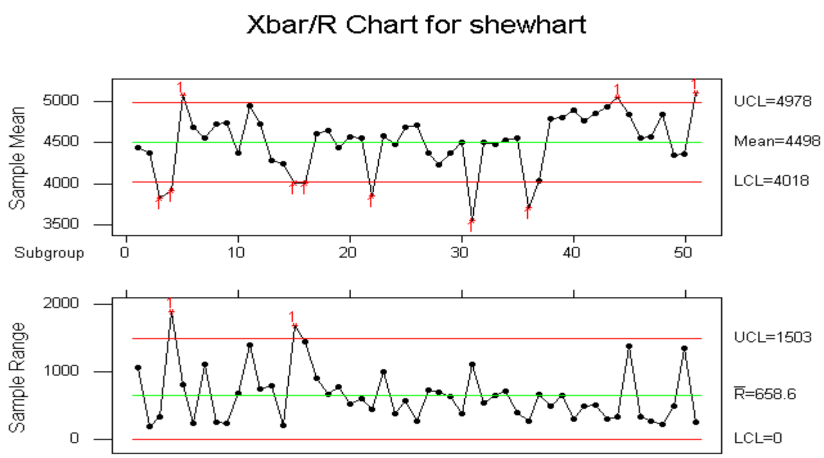


Fitted line plot: by removing the start-up effect, a variability reduction of about 42.2% would be achieved (R^2)

Conclusion 1: important information may be lost (masked) by studying the sample statistics only
Conclusion 2: keep track of how data have been collected within the sample!!

W.A. Shewhart (1931) *Economic Control of Quality Manufactured product:*

- 204 electrical resistance consecutive measurements (MΩ) (shewhart.dat)
- n=4 (arbitrary choice – as stated by Shewhart)



✓10 mean values (19.6% of all values) are out-of-control
✓with run rules: 10 more out-of-control (globally 17 samples - 35% of total- seem to be out-of-control)

Analogously to the previous example: I control chart directly applied on process means

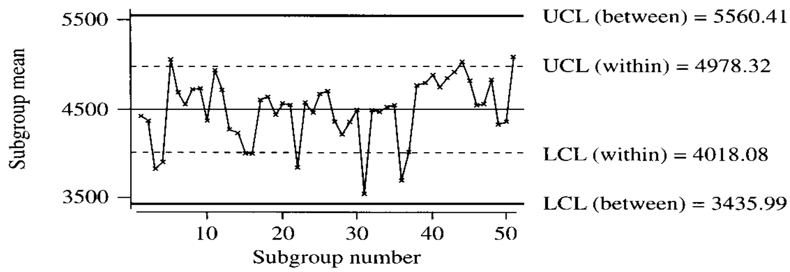


Figure 6.18 Subgroup means for megohm data with different control limits.

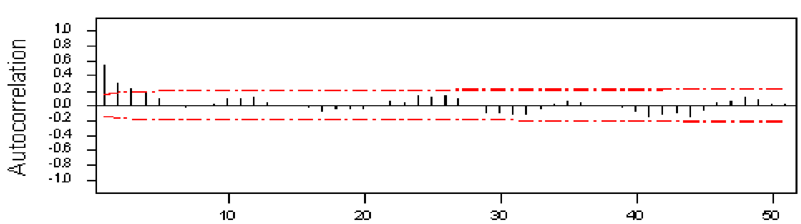
Contrary to the previous example: between-sample standard deviation

Let's analyse the measurement sequence

Runs Test: shewhart

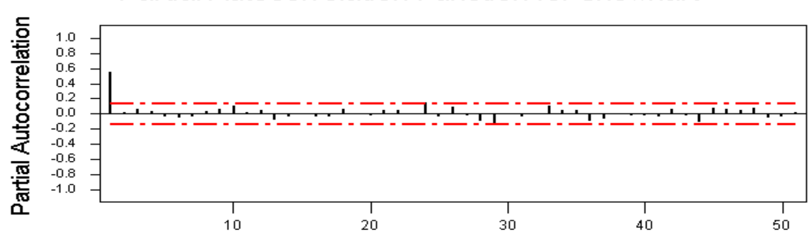
```
    K =  4498.1765
  The observed number of runs =   49
  The expected number of runs = 102.0196
112 Observations above K   92 below
            The test is significant at  0.0000
```

Autocorrelation Function for shewhart



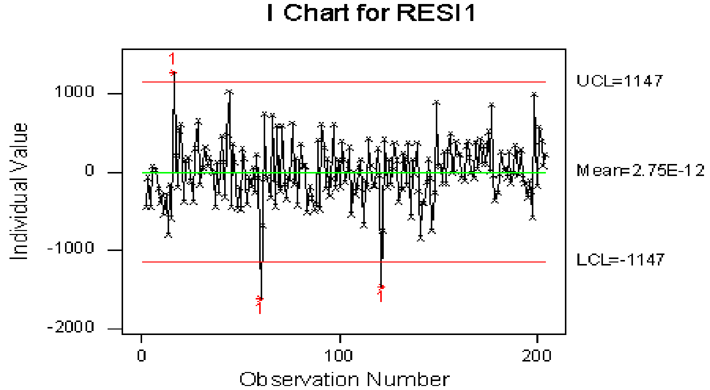
Partial Autocorrelation Function for shewhart

AR(1)

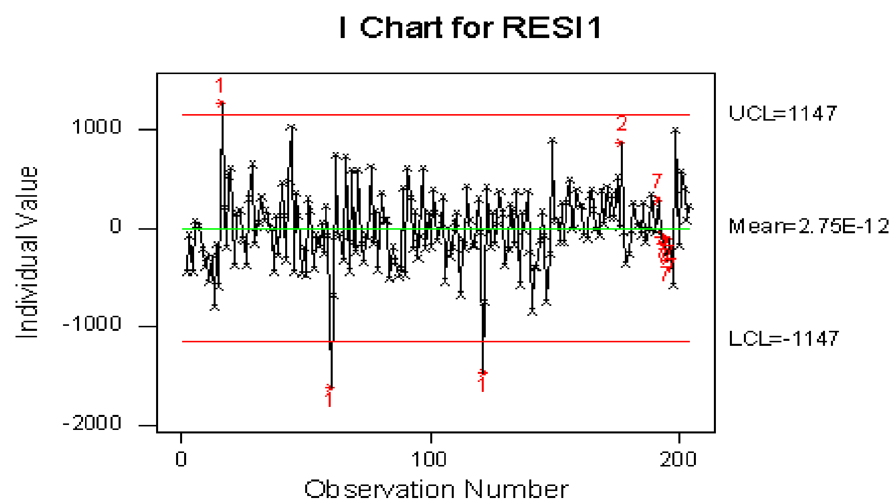


Regression Analysis: shewhart versus shewhart t_1

```
The regression equation is
shewhart = 2029 + 0.549 shewhart t_1
203 cases used 1 cases contain missing values
Predictor         Coef      SE Coef          T         P
Constant        2028.8        266.3       7.62     0.000
shewhart       0.54867      0.05892       9.31     0.000
S = 390.4        R-Sq = 30.1%      R-Sq(adj) = 29.8%
```



I (estimate based on MR)

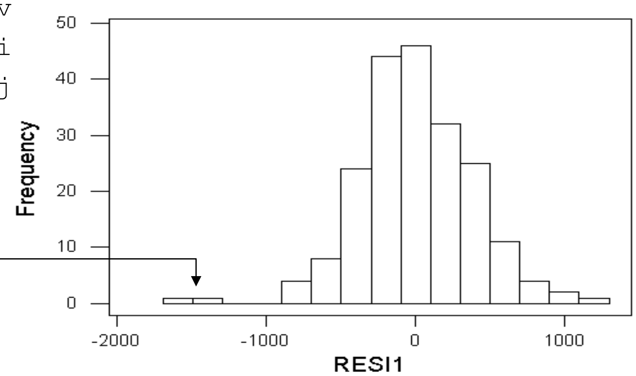


+run rules

```
TEST 1. One point more than 3.00 sigmas from center line.
Test Failed at points: 16 60 121
TEST 2. 9 points in a row on same side of center line.
Test Failed at points: 177
TEST 7. 15 points within 1 sigma of center line (above and below CL).
Test Failed at points: 192 193 194 195 196 197
```

Runs Test: RESI1

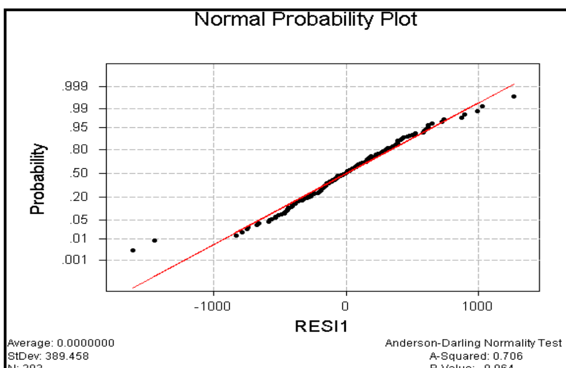
```
K =      0.0000
The observed number of runs = 102
The expected number of runs = 102.4778
100 Observations abov
            The test i
            Cannot rej
```

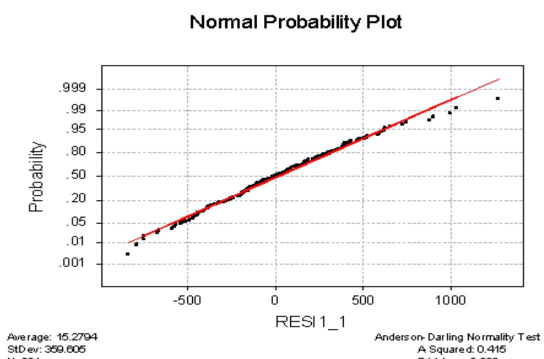


Observations 60, 121

Quality Data Analysis

58





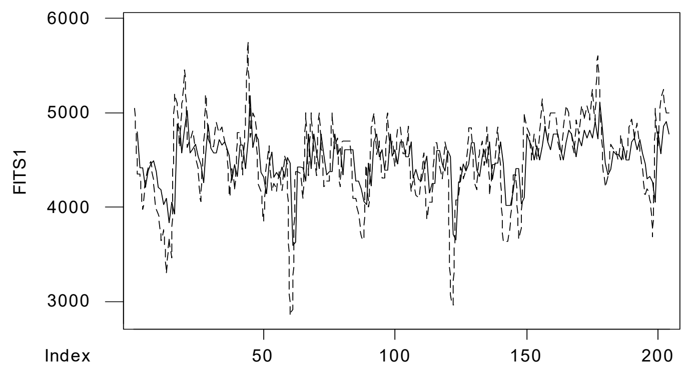
p-value 0.064

By excluding observations 60, 121

p-value 0.332

59

Fitted
line plot



- Conclusions:
  - –Detrimental effects due to non-random patterns within the subgroup. E.g., Shewhart with positive autocorr. → standard dev. is understimated → many unjustified out-of-control observations

Control chart for process mean: gapping-batching

Non-random patterns within the sample yield:

- Wrong estimation of process dispersion;
- Non-randomness that may characterize the sample mean sequence too
  1. Identify a model for non-random pattern directly on the sample mean sequence
  2. Gapping (sampling)-Batching

Strategy n° 2: What types of time series does it work for?

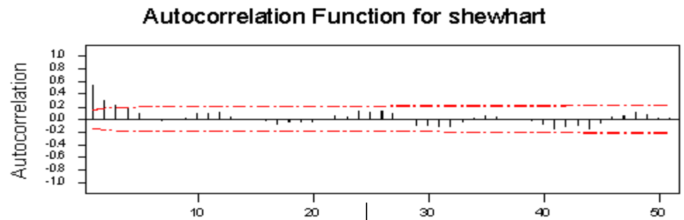
The process that generates the observed data must be **stationary**.

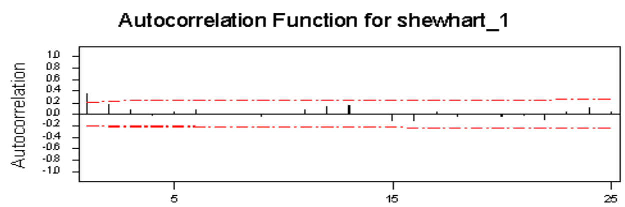
- Theoretically speaking, if the process is not stationary (e.g., trend, random walk, …) it is not possible to remove the relationship between observed means
- Practically speaking, it might look like one could remove the relationship between observations even though the process is not stationary

Example – Shewhart's time series (individual measurements):

1. One observation out of 2

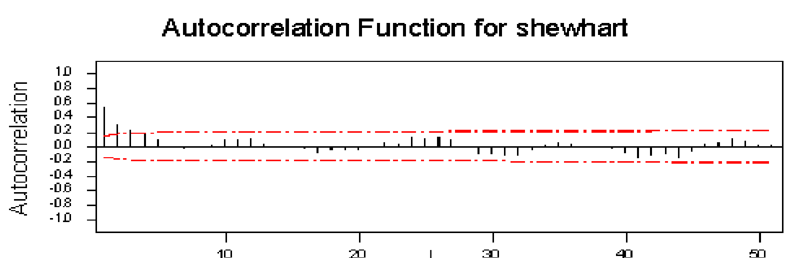
shewhart	shewhart_1
5045	5045
4350	
4350	4350
3975	
4290	4290
4430	
4485	4485
4285	…
…	

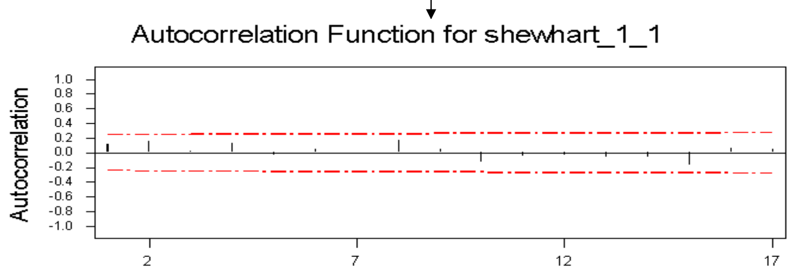




2. One observation out of 3

shewhart	shewhart_1_1
5045	5045
4350	
4350	
3975	3975
4290	
4430	
4485	4485
4285	…
…	

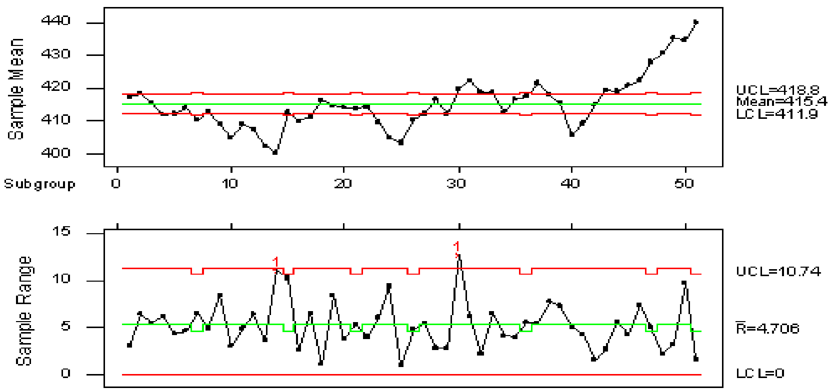




Data: daily values of Standard & Poors index (S&P) from January 6, 1992 to December 24, 1992 (sp500.dat)

- Remind: time series of economic/finantial indexes often follows a random walk
- BATCHING: consider subgroup=week (from Monday to Friday) :

Xbar/R Chart for S&P



✓means: non-random pattern
✓Control limits are so tight that more than 50% of observations are out-of-control
✓Missing data (days off)

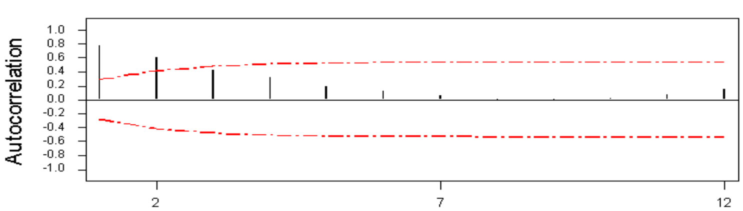
Quality Data Analysis

64

64

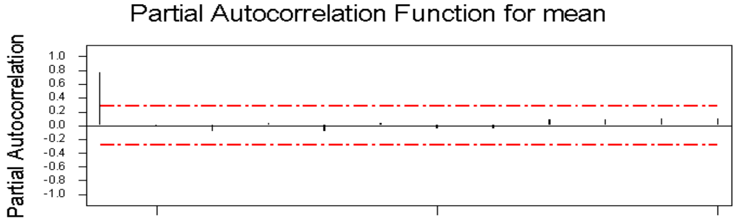
Process means exhibit AR(1) autocorrelation

Autocorrelation Function for mean



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.78	5.54	32.51	8	-0.01	-0.03	71.43
2	0.60	2.88	52.23	9	-0.01	-0.05	71.44

Partial Autocorrelation Function for mean



Random walk is a special case of AR(1) with ϕ=1

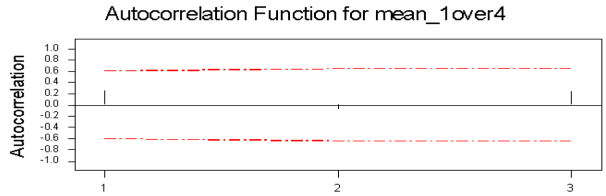
$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

Quality Data Analysis

65

65

After batching: gapping: consider one subgroup every three weeks: from 51 subgroups to 13



Autocorrelation seems to be filtered out (but only because the number of data is reduced): add means (with gapping) from dec. 92 to dec. 93

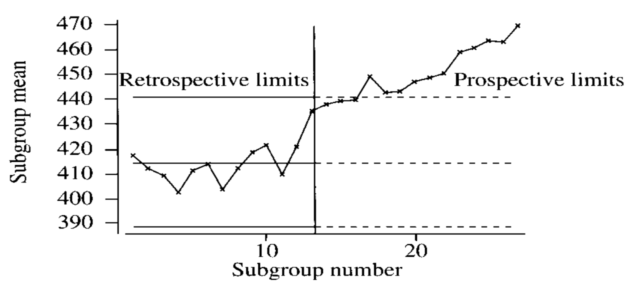


Figure 6.28 Subgroup mean chart for gapped S&P subgroups.

Work on 51 process mean values as they were individuals:

```
The regression equation is
mean = 18.3 + 0.957 mean t-1

50 cases used 1 cases contain missing values
```

Predictor	Coef	SE Coef	T	P
Constant	18.33	33.76	0.54	0.590
mean t-1	0.95691	0.08133	11.77	0.000

S = 4.154   R-Sq = 74.3%   R-Sq(adj) = 73.7%

It looks like a random walk even on process means

Indeed:

$$Y_{t+5} = \mu + Y_{t+4} + \varepsilon_{t+5} = 2\mu + Y_{t+3} + \varepsilon_{t+4} + \varepsilon_{t+5} = 3\mu + Y_{t+2} + \varepsilon_{t+3} + \varepsilon_{t+4} + \varepsilon_{t+5} = \dots = 5\mu + Y_t + \sum_{j=1}^{5} \varepsilon_{t+j}$$

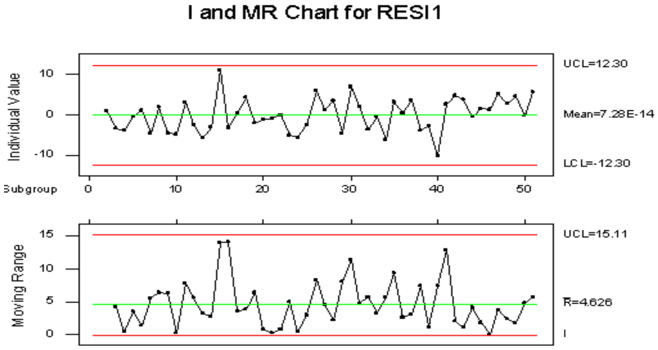
$$Y_{t+k+5} - Y_{t+k} = 5\mu + \sum_{j=1}^{5} \varepsilon_{t+k+j}$$

Sample mean

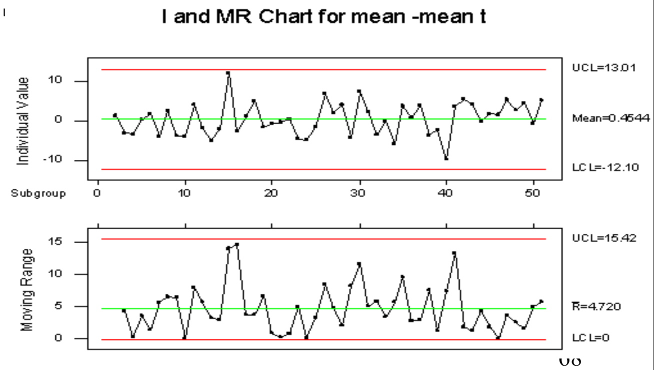
$$\frac{1}{5}\left(\sum_{k=6}^{10} Y_{t+k}\right) - \frac{1}{5}\left(\sum_{k=1}^{5} Y_{t+k}\right) = \frac{1}{5}\left(\sum_{k=1}^{5} (Y_{t+k+5} - Y_{t+k})\right) = \frac{1}{5}\left(\sum_{k=1}^{5}\left(5\mu + \sum_{j=1}^{5} \varepsilon_{t+k+j}\right)\right) = 5\mu + \underbrace{\frac{1}{5}\left(\sum_{k=1}^{5}\sum_{j=1}^{5} \varepsilon_{t+k+j}\right)}_{\text{comb. lineare: } \varepsilon_t'}$$

AR(1)

With regression model



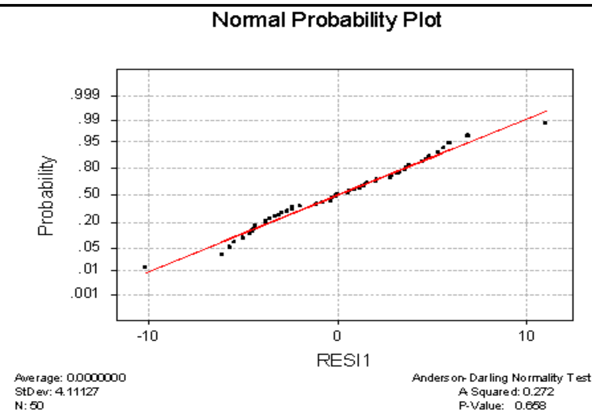
With differences between means

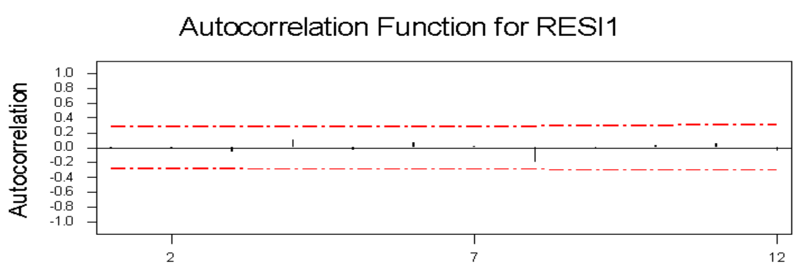


Quality Data Analysis

68

Residual analysis





Quality Data Analysis

69

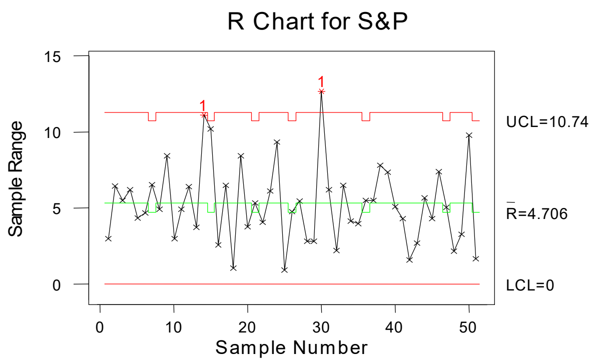
69

R chart with correlation within the group

Thus far, our attention has been on the chart for process mean

R chart: even with severe autocorrelation, the sample range exhibits a random pattern, but:

The distribution of R values becomes more and more asymmetric as the autocorrelation increases



Two out-of-control data: by the way, control limits at $\pm 3\sigma_R$ are appropriate?

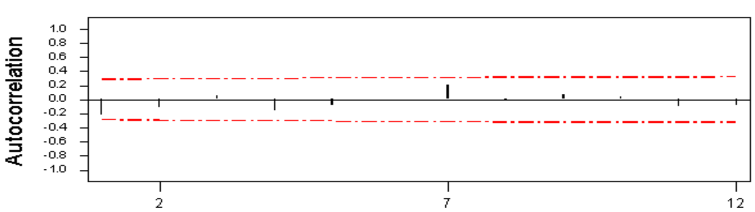
-Probability limits (if data follow NID distribution)

-Data transformation (Range)

Runs Test: range

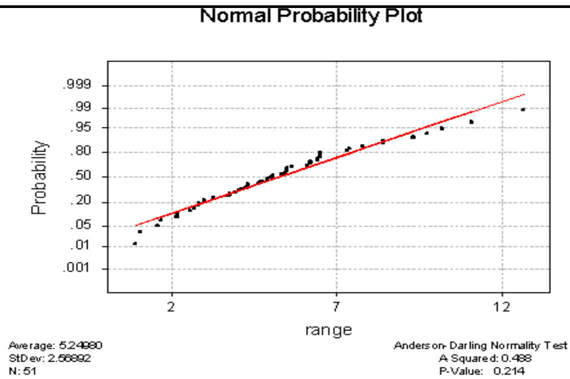
```
K =      5.2498
The observed number of runs =  33
The expected number of runs =  26.4118
24 Observations above K   27 below
        The test is significant at  0.0614
```

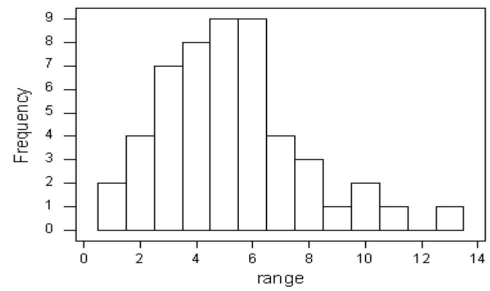
Autocorrelation Function for range^0.5

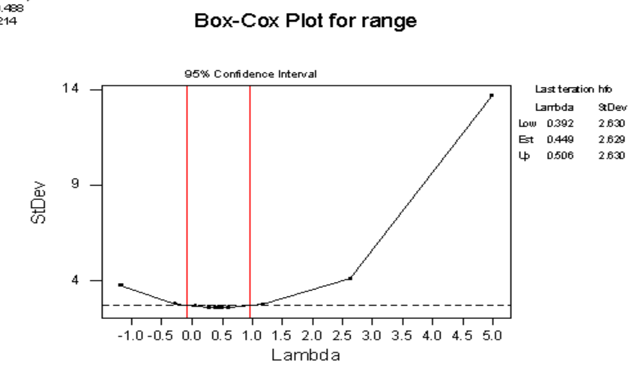


Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.22	-1.54	2.52	8	-0.02	-0.14	7.87
2	-0.11	-0.78	3.25	9	0.08	0.48	8.24
3	0.05	0.35	3.40	10	0.03	0.19	8.30
4	-0.15	-1.02	4.72	11	-0.09	-0.60	8.91
5	-0.08	-0.52	5.09	12	-0.08	-0.48	9.31
6	0.02	0.13	5.11				
7	0.21	1.38	7.84				

Despite non randomness of original data, the ranges exhibit a random pattern







Despite normality null hypothesis can not be rejected, the range distribution is strongly asymmetric

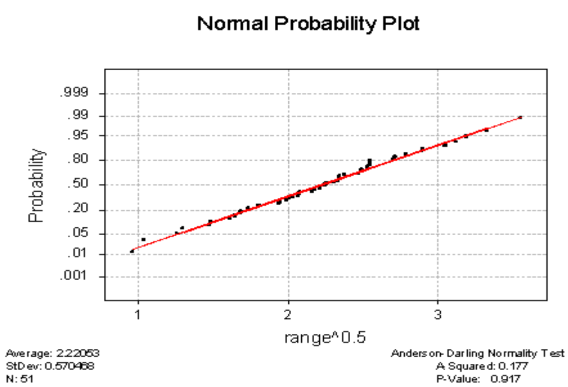
Use λ=0.5

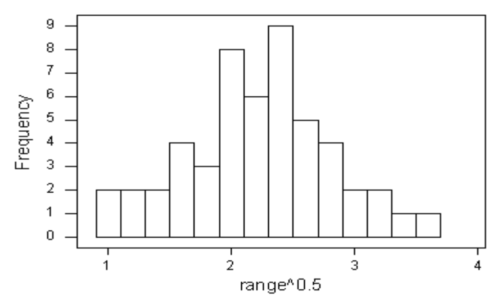
Quality Data Analysis

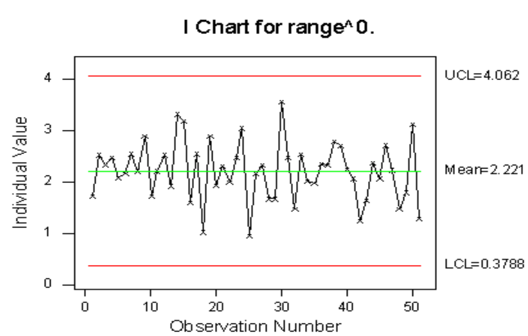
Quality Engineering

72

72







No more out-of-control data:

Quality Data Analysis

73