Chapter 290

Control Charts for Means (Simulation)

Introduction

This procedure allows you to study the run length distribution of Shewhart (Xbar), Cusum, FIR Cusum, and EWMA process control charts for means using simulation. This procedure can also be used to study charts with a single observation at each sample. The in-control mean and standard deviation can be input directly or a specified number of in-control preliminary samples can be simulated based on a user-determined in-control distribution. The out-of-control distribution is flexible in terms of distribution type and distribution parameters. The Shewhart, Cusum, and EWMA parameters can also be flexibly input. This procedure can also be used to determine the necessary sample size to obtain a given run length.

Simulation Details

If the in-control mean and in-control standard deviation are assumed to be known, the steps to the simulation process are as follows (assume a sample consists of *n* observations).

1. An out-of-control sample of size *n* is generated according to the specified distribution parameters of the out-of-control distribution.

2. The average of the sample is produced and, if necessary for the particular type of control chart, the standard deviation.

3. Based on the control chart criteria, it is determined whether this sample results in an out-of-control signal.

4. If the sample results in an out-of-control signal, the sample number is recorded as the run length for that simulation. If the sample does not result in an out-of-control signal, return to Step 1.

5. Steps 1 through 4 are repeated until the number of simulations (*Nsim*) is reached. The result is *Nsim* run lengths.

6. The average or median or specified percentile of the run length distribution is reported.

If the in-control mean and in-control standard deviation are to be simulated based on in-control preliminary samples (*NPrelim*), the steps to the simulation process are as follows (assume a sample consists of *n* observations).

1. *NPrelim* in-control samples of size *n* are generated according to the specified distribution parameters of the in-control distribution.

2. The in-control average and standard deviation are calculated based on the *NPrelim* simulated in-control samples.

3. An out-of-control sample of size *n* is generated according to the specified distribution parameters of the out-of-control distribution.

4. The average of the sample is produced and, if necessary for the particular type of control chart, the standard deviation.

5. Based on the control chart criteria, it is determined whether this sample results in an out-of-control signal.

6. If the sample results in an out-of-control signal, the sample number is recorded as the run length for that simulation. If the sample does not result in an out-of-control signal, return to Step 3.

7. Steps 1 through 6 are repeated until the number of simulations (*Nsim*) is reached. The result is *Nsim* run lengths.

8. The average or median or specified percentile of the run length distribution is reported.

Data Distributions

A wide variety of distributions may be studied. These distributions can vary in skewness, elongation, or other features such as bimodality. A detailed discussion of the distributions that may be used in the simulation is provided in the chapter "Data Simulator".

Formulas for Constructing Control Charts

Suppose we have *k* subgroups, each of size *n*. Let x_{ij} represent the measurement in the jth sample of the ith subgroup. Three statistics that are routinely computed (depending on the type of control chart) for each subgroup are:

The subgroup mean

$$\bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n}$$

the subgroup range

$$R_i = x_{(n)} - x_{(1)}$$

and/or the subgroup standard deviation

$$s_i = \sqrt{\frac{\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2}{n-1}}$$

Estimating Sigma

Control limits vary according to the type of control chart used. These require an estimate of the process mean, μ_x (mu), and the process variability, σ_x (sigma). A known estimate of μ_x may be supplied by the user, or it may be estimated by the average of the averages of a number of in-control preliminary samples, 'x double bar' (also known as the grand mean):

$$\bar{\bar{x}}_i = \frac{\sum_{i=1}^{k} \bar{x}_{ij}}{k}$$

When sigma is not input directly, there are two methods available for estimating σ_x.

Method 1: Estimating Sigma from the Ranges

$$\hat{\sigma}_x = \frac{\overline{R}}{d_2}$$

where

$$\overline{R} = \frac{\sum_{i=1}^{k} R_i}{k}$$

$$d_2 = \frac{E(R)}{\sigma_x} = \frac{\mu_R}{\sigma_x}$$

Making the assumption that the x_{ij}'s follow the normal distribution with constant mean and variance, we can derive values for d_2 through the use of numerical integration.

Method 2: Estimating Sigma from the Standard Deviations

$$\hat{\sigma}_x = \frac{\overline{s}}{c_4}$$

where

$$\overline{s} = \frac{\sum_{i=1}^{k} s_i}{k}$$

$$c_4 = \frac{E(s)}{\sigma_x} = \frac{\mu_s}{\sigma_x}$$

. Making the assumption that the x_{ij}'s follow the normal distribution with constant mean and variance, we can derive values for c_4 from the following formula.

$$c_4 = \sqrt{\frac{2}{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

Estimating Sigma when *n* = 1

When *n* is one, we cannot calculate R_i or s_i since these require at least two measurements. In this case, we use the standard deviation of all *k* measurements. Unfortunately, this method does not approximate the within-subgroup variation. Rather, it combines the within and the between subgroup variation.

Xbar Chart Limits

The lower and upper control limits for the Xbar chart are calculated using the formula

$$LCL = \overline{\overline{x}} - z\left(\frac{\hat{\sigma}_x}{\sqrt{n}}\right)$$

$$UCL = \overline{\overline{x}} + z\left(\frac{\hat{\sigma}_x}{\sqrt{n}}\right)$$

where z is a multiplier (often set to three) chosen to reduce the possibility of false alarms (signaling an out-of-control situation when the process is in control).

Cusum and FIR Cusum Charts

The Cusum chart has been shown to detect small shifts in the process average much quicker than the Xbar chart.

In **PASS** we use the Cusum procedure presented by Ryan (1989). This procedure may be summarized as follows:

1. Calculate all statistics as if you were going to generate an Xbar chart.

2. Calculate the z_i using the formula

$$z_i = \frac{\bar{x}_i - \bar{\bar{x}}}{\hat{\sigma}_{\bar{x}}}$$

3. Calculate the lower and upper cumulative sums as follows

$$S_{Li} = -\max\left[0, \left(-z_i - K\right) + S_{Li-1}\right]$$

$$S_{Hi} = \max\left[0, \left(z_i - K\right) + S_{Hi-1}\right]$$

4. The control limits are chosen as plus or minus h. Often, K is set to 0.5 (for detecting one-sigma shifts in the mean) and h is set to 5.

5. Usually, the starting value for S_{Li} and S_{Hi} is zero. Occasionally, however, a "fast initial response" (FIR) value of $h/2$ is used.

EWMA Chart Limits

The lower and upper control limits for the exponentially weighted moving-average (EWMA) chart are calculated using the formula

$$LCL_i = \bar{\bar{x}} - L\left(\frac{\hat{\sigma}_x}{\sqrt{n}}\right)\sqrt{\frac{R}{2-R}\left[1 - \left(1-R\right)^{2i}\right]}$$

$$UCL_i = \bar{\bar{x}} + L\left(\frac{\hat{\sigma}_x}{\sqrt{n}}\right)\sqrt{\frac{R}{2-R}\left[1 - \left(1-R\right)^{2i}\right]}$$

where L is a multiplier (usually set to three) and R is smoothing constant.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

Solve For

Solve For

Solve for either the *run length distribution* or the *sample size*. This is the parameter displayed on the vertical axis of the plot.

Note: The search for the sample size may take several minutes to complete. You may find it quicker and more informative to solve for run length distribution for a range of sample sizes.

Simulations

Simulations

Specify the number of Monte Carlo iterations. Each simulation consists of the full process required to obtain a run length. That is, the number of simulations here is the number of run lengths that are used to form the run length distribution.

A 2007 article in Communication in Statistics-Simulation and Computation suggests that 5000 to 10000 simulations is usually sufficient. For estimating percentiles near 0 or 1, more simulations may be needed.

For searches for N, you may wish to use fewer simulations to determine the ballpark for N and then larger numbers of simulations for fine-tuning.

Maximum Run Length

For each simulation, if the run length reaches the maximum run length, the maximum run length is used for that simulation. This keeps simulations from going on indefinitely when run lengths are extremely long.

For example, if the average run length is close to the maximum run length, the maximum run length should be increased.

Run Length

Run Length Summary 1 and 2

Select the first and second run length distribution summary value to be reported.

If a value between 0 and 1 (not inclusive) is entered, the corresponding distribution percentile is reported.

For example, if 0.40 is entered, the run length such that 40% of simulated run lengths are shorter is reported.

Percentile or Average

When solving for N, the value entered for "Percentile or Average" indicates whether the search is based on the average run length or some percentile, such as the median.

If a value between 0 and 1 (not inclusive) is entered, the corresponding distribution percentile is used in the search for N.

For example, if ARL (Average Run Length) is entered, and "Run Length" is 10, the program will search for the smallest N for which the average run length is 10.if 0.40 is entered, and "Run Length" is 10, the program will search for the smallest N for which 40% of simulated run lengths are shorter than 10.

Run Length

When solving for N, this is run length that is searched for based on the specified "Percentile or Average".

For example, if ARL (Average Run Length) is entered, and "Run Length" is 10, the program will search for the smallest N for which the average run length is 10.if 0.40 is entered, and "Run Length" is 10, the program will search for the smallest N for which 40% of simulated run lengths are shorter than 10.

Sample Size

N (Sample Size)

Enter a value for the sample size, N. This is the number of observations in each sample. This sample size is also used when preliminary samples are used.

When a sample size of one is used, the preliminary sample standard deviation is estimated with all the preliminary data jointly.

You may enter a range such as 2 to 10 by 2 or a list of values separated by commas or blanks.

You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

Distributions Tab

In-Control and Out-of-Control Distributions

In-Control Distributions Specified By

Specify whether a known mean and standard deviation will be used for the centerline and standard deviation for boundaries, or if the centerline and standard deviation for boundaries will be based on simulated preliminary samples.

In-Control and Out-of-Control Distributions - Known

In-Control Mean (Center Line)

This is the centerline mean from which boundaries and the in-control standard deviation are based.

In-Control Standard Deviation

This is an assumed known standard deviation from which the boundaries are calculated.

In-Control and Out-of-Control Distributions – Preliminary Samples

In-Control Distribution

This is the distribution from which in-control preliminary samples are simulated. The mean and standard deviation of the in-control samples are used to define the centerline and the control boundaries for that simulation.

The parameters of the distribution can be specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value "M0" is reserved for the value of the mean under the null hypothesis.

M0 (In-Control Mean)

This is the mean of the in-control distribution. Preliminary samples are simulated based on a distribution with this mean.

These values are substituted for the "M0" in the distribution specification given above.

You can enter a list of values using the syntax *0 2 3* or *0 to 3 by 1*.

Note that whether "M0" is the mean of the simulated distribution depends on the formula you have entered. For example, "N(M0 S)" does have M0 as its mean,

but "N(M0 S)-N(M0 S)" has a mean of zero.

Number of Preliminary Samples

For each run length simulation, this is the number of samples that are simulated to determine the centerline and standard deviation for the control chart.

For example, if a control chart uses 50 samples with N = 5 to establish the in control centerline and standard deviation, 50 should be entered here. The program will generate 50 samples of size 5 each, obtain means and standard deviations (or ranges) for each sample and use those to produce a centerline and standard deviation estimate as a basis for the simulated control chart.

Commonly recommended numbers of in-control preliminary samples are 30, 50, and 60.

Varying this number allows one to see the effect on the distribution of run lengths.

Standard Deviation Estimation Method

Specify whether the overall control chart standard deviation for each simulation is to be estimated from the ranges of the preliminary samples or the standard deviations of the preliminary samples.

If N is one, then this selection is ignored and the standard deviation is calculated based on all the preliminary samples jointly.

In-Control and Out-of-Control Distributions – Out-of-Control

Out-of-Control Distribution

This is the distribution from which out-of-control samples are simulated.

The parameters of the distribution can be specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value "M1" is reserved for the value of the mean under the alternative hypothesis.

M1 (Out-of-Control Mean)

This is the mean of the out-of-control distribution. Out-of-control samples are simulated based on a distribution with this mean.

These values are substituted for the "M1" in the distribution specification given above.

You can enter a list of values using the syntax *0 2 3* or *0 to 3 by 1*.

Note that whether "M1" is the mean of the simulated distribution depends on the formula you have entered. For example, "N(M1 S)" does have M1 as its mean, but "N(M1 S)-N(M1 S)" has a mean of zero.

In-Control and Out-of-Control Distributions – S and Other Parameters

Parameter Values (S, A, B, C)

Enter the numeric value(s) of parameter listed above. These values are substituted for the corresponding letter in the distribution specifications.

You can enter a list of values using syntax such as *0 1 2 3* or *0 to 3 by 1*.

You can also change the letter that is used as the name of this parameter.

Tests Tab

Include Tests

Specify whether to include this test in the Reports and Plots. The specific options for each test are specified below.

Shewhart (Xbar) Options

Use Z-Multiplier or Probability

Specify whether the Z-multiplier will be input directly or calculated based on the two-sided probability.

The Z-multiplier is the value of Z in the Shewhart Xbar chart formula for the limits:

meanline +/- Z * SDmean

Z-Multiplier

The Z-multiplier is the value of Z in the Shewhart Xbar chart formula for the limits:

meanline +/- Z * SDmean

Two-Sided Probability

This probability is used to calculate the Z-multiplier to create the Shewhart limits.

The Z-multiplier is the value of Z in the Shewhart Xbar chart formula for the limits:

meanline +/- Z * SDmean

Cusum and FIR Cusum Options

K

K is the value that is subtracted from the z-score in the CUSUM procedure formula.

H

H is the out-of-control threshold for the CUSUM procedure.

Typical choices are 4 and 5.

FIR

FIR or Fast Initial Response is a value that may be used to shorten the run lengths of the CUSUM procedure.

FIR is often set to H/2.

EWMA Options

R

R is the weight assigned to the terms of the moving average chart.

It is sometimes known as lambda or w.

Values in the range of 0.10 to 0.30 are common.

L

L defines the distance to the boundary. It is essentially a Z-multiplier used to create L-sigma limits.

It is sometimes known as q.

A typical value is 3.

Example 1 – Run Length Distribution

A researcher wishes to examine the run length distribution for a process monitored by a Shewhart (Xbar) chart. Six observations are to make up the sample examined at each hour. The in-control mean and standard deviation are known to be 5.2 and 3.1, respectively. The researcher would like see the run length distribution if the out-of-control mean and standard deviation are 6.2 and 3.1, respectively. A Z-Multiplier of 3.0 is to be used in the control chart for the boundaries.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Control Charts for Means (Simulation)** procedure window by clicking on **Quality Control**, and then clicking on **Control Charts for Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	**Value**
Design Tab	
Solve For	**Run Length Distribution**
Simulations	**5000**
Maximum Run Length..............................	**5000**
Run Length Summary 1	**ARL (Average Run Length)**
Run Length Summary 2	**MRL (Median Run Length)**
N (Sample Size)....................................	**6**
Distributions Tab	
In-Control Distributions Specified By	**Mean, Standard Deviation Directly**
In-Control Mean	**5.2**
In-Control Standard Deviation	**3.1**
Out-of-Control Distribution	**N(M1 S)**
M1 (Out-of-Control Mean)	**6.2**
S..	**3.1**
Tests Tab	
Shewhart (Xbar).....................................	**Checked**
All Other Tests	**Unchecked**
Use Z-Multiplier or Probability	**Z-Multiplier**
Z-Multiplier	**3.0**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Control Charts for a Process Mean
Shewhart control limits were determined by specifying a centerline mean and standard deviation directly.
Out-of-Control Distribution (determines size of shift): Normal(M1 S)

Chart Type	Average Run Length	Median Run Length	N	In-Cntrl Mean	Z	In-Cntrl SD	LCL	UCL	Out-of-Cntrl Mean M1	S
Shewhart	72.3	50.0	6	5.20	3.000	3.10	1.40	9.00	6.20	3.10

Notes
Simulations: 5000. Run Time: 4.0 seconds.

References
Ryan, T.P. 1989. Statistical Methods for Quality Improvement. Wiley. New York.
Montgomery, D.C. 1991. Introduction to Statistical Quality Control. Wiley. New York.

Report Definitions
Average Run Length is the mean of the run lengths across all simulations.
Median Run Length is the median of the run lengths across all simulations.
N is the number of units measured in each sample.
In-control Mean is the assumed known value of the center line of the control chart.
Z is the Z-multiplier which corresponds to the two-sided probability of a single sample mean outside the
    control limits.
In-control SD is the assumed known standard deviation that is used in the calculation of limits.
LCL and UCL are the lower and upper control chart limits, respectively.
Out-of-Control Mean (M1) is the mean of the distribution from which out-of-control samples are drawn. It's
    difference from the Mean Line is the shift to detect.
Other parameters, often S (standard deviation), define the distribution from which out-of-control samples are
    drawn.

Summary Statements
For a Xbar control chart with a mean line at 5.20 and lower and upper Shewhart control limits
of 1.40 and 9.00, respectively, samples of size 6 from the distribution Normal(M1 S) with mean
6.20 have an average run length of 72.3 and a median run length of 50.0. These results are
based on 5000 Monte Carlo samples.

Individual Summaries
Out-of-Control Distribution (determines size of shift): Normal(M1 S)

Chart Type	Average Run Length	Median Run Length	N	In-Cntrl Mean	Z	In-Cntrl SD	LCL	UCL	Out-of-Cntrl Mean M1	S
Shewhart	72.3	50.0	6	5.20	3.000	3.10	1.40	9.00	6.20	3.10

Average Run Length 95% CI: (70.3, 74.3)
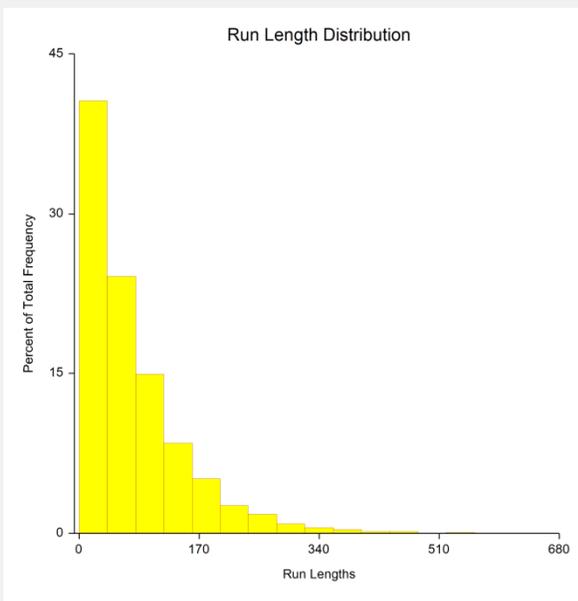Median Run Length 95% CI: (48.0, 52.0)

Average Run Length and Percentiles

Avg	1%	5%	10%	25%	50%	75%	90%	95%	99%
72.3	1.0	4.0	8.0	21.0	50.0	97.0	170.0	221.0	331.0

The results will vary slightly because they are based on simulation.

Plots Section

This plot shows the distribution of run lengths of 5000 simulated runs.

Example 2 – Comparing Tests

Continuing with the Example 1 parameters, the researchers would like to compare the various control chart tests available.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Control Charts for Means (Simulation)** procedure window by clicking on **Quality Control**, and then clicking on **Control Charts for Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	**Value**
Design Tab	
Solve For	**Run Length Distribution**
Simulations	**5000**
Maximum Run Length.............................	**5000**
Run Length Summary 1	**ARL (Average Run Length)**
Run Length Summary 2	**MRL (Median Run Length)**
N (Sample Size).....................................	**6**
Distributions Tab	
In-Control Distributions Specified By	**Mean, Standard Deviation Directly**
In-Control Mean	**5.2**
In-Control Standard Deviation	**3.1**
Out-of-Control Distribution	**N(M1 S)**
M1 (Out-of-Control Mean)	**6.2**
S..	**3.1**
Tests Tab	
All Tests	**Checked**
Use Z-Multiplier or Probability	**Z-Multiplier**
Z-Multiplier	**3.0**
K..	**0.5**
H	**5**
FIR	**2.5**
R	**0.25**
L..	**3**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Control Charts for a Process Mean
Shewhart control limits were determined by specifying a centerline mean and standard deviation directly.
Out-of-Control Distribution (determines size of shift): Normal(M1 S)
Cusum Parameters: K: 0.5, H: 5, FIR: 2.5
EWMA parameters: R: 0.25, L: 3

Chart Type	Average Run Length	Median Run Length	N	In-Cntrl Mean	Z	In-Cntrl SD	LCL	UCL	Out-of-Cntrl Mean M1	S
Shewhart	74.3	53.0	6	5.20	3.000	3.10	1.40	9.00	6.20	3.10
Cusum	15.6	13.0	6	5.20		3.10			6.20	3.10
Cus+Shew	14.5	12.0	6	5.20	3.000	3.10	1.40	9.00	6.20	3.10
FIR Cusum	10.0	7.0	6	5.20		3.10			6.20	3.10
FIR+Shew	10.0	7.0	6	5.20	3.000	3.10	1.40	9.00	6.20	3.10
EWMA	17.1	13.0	6	5.20		3.10			6.20	3.10
EWMA+Shew	16.6	13.0	6	5.20	3.000	3.10	1.40	9.00	6.20	3.10

The FIR tests show the process is out-of-control much sooner than the other tests.

Example 3 – Validation Using Montgomery

A table in Montgomery (1991), page 298, gives the average run lengths for a mean shift of one standartd deviation to be Cusum (10.4), Cusum with Shewhart (10.20), Cusum with FIR (6.35), Cusum with FIR and Shewhart (6.32).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Control Charts for Means (Simulation)** procedure window by clicking on **Quality Control**, and then clicking on **Control Charts for Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Run Length Distribution**
Simulations **100000**
Maximum Run Length............................. **5000**
Run Length Summary 1.......................... **ARL (Average Run Length)**
Run Length Summary 2.......................... **MRL (Median Run Length)**
N (Sample Size)...................................... **1**

Distributions Tab
In-Control Distributions Specified By ...... **Mean, Standard Deviation Directly**
In-Control Mean **0**
In-Control Standard Deviation **1**
Out-of-Control Distribution...................... **N(M1 S)**
M1 (Out-of-Control Mean) **1**
S... **1**

Control Charts for Means (Simulation)

Tests Tab

Cusum .... **Checked**

Cusum + Shewhart .... **Checked**

FIR Cusum.. **Checked**

FIR Cusum + Shewhart .... **Checked**

All Other Tests .... **Unchecked**

Use Z-Multiplier or Probability .... **Z-Multiplier**

Z-Multiplier **3.5**

K.. **0.5**

H **5**

FIR **2.5**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Control Charts for a Process Mean
Shewhart control limits were determined by specifying a centerline mean and standard deviation directly.
Out-of-Control Distribution (determines size of shift): Normal(M1 S)
Cusum Parameters: K: 0.5, H: 5, FIR: 5
Cusum Parameters: K: 0.5, H: 5, FIR: 2.5

Chart Type	Average Run Length	Median Run Length	N	In-Cntrl Mean	Z	In-Cntrl SD	LCL	UCL	Out-of-Cntrl Mean M1	S
Cusum	10.36	9.00	1	0.00		1.00			1.00	1.00
Cus+Shew	10.26	9.00	1	0.00	3.500	1.00	-3.50	3.50	1.00	1.00
FIR Cusum	6.35	5.00	1	0.00		1.00			1.00	1.00
FIR+Shew	6.33	5.00	1	0.00	3.500	1.00	-3.50	3.50	1.00	1.00

The average run lengths are very close to those presented in Montgomery (1991).