

G Chart

Summary	1
Data Input.....	3
G Chart.....	4
G Chart Report.....	7
Analysis Summary	8
Analysis Options.....	9
Runs Tests.....	10
Save Results	11
References.....	11

Summary

The **G Chart** procedure creates a control chart based on the intervals of time between the occurrence of rare events. It uses the geometric distribution, which assumes that every time an opportunity for an event arises, there is a probability p that the event will occur. This type of chart is used frequently in the health sciences to monitor the occurrence of events such as post-surgical infections.

The chart may be constructed in either *Initial Study* (Phase 1) mode, where the current data determine the control limits, or in *Control to Standard* (Phase 2) mode, where the limits come from either a known standard or from prior data.

The data for this analysis may be in either of 2 formats:

1. *Time of occurrence*. In this format, the data describe the time at which each event occurred:

$$t_i = \text{time of occurrence of event } i.$$

2. *Number of opportunities*. In this format, the data represent the number of opportunities between events.

$$x_j = \text{number of opportunities for an event to have occurred between event } j-1 \text{ and event } j.$$

Note that the domain of x_j is 0, 1, 2, ..., since it counts the number of opportunities for an event to have occurred between each event that did occur, which may equal 0. If the times of occurrence are given, the interevent times are calculated according to:

$$x_j = t_j - t_{j-1} \tag{1}$$

The number of interevent times will be represented by m .

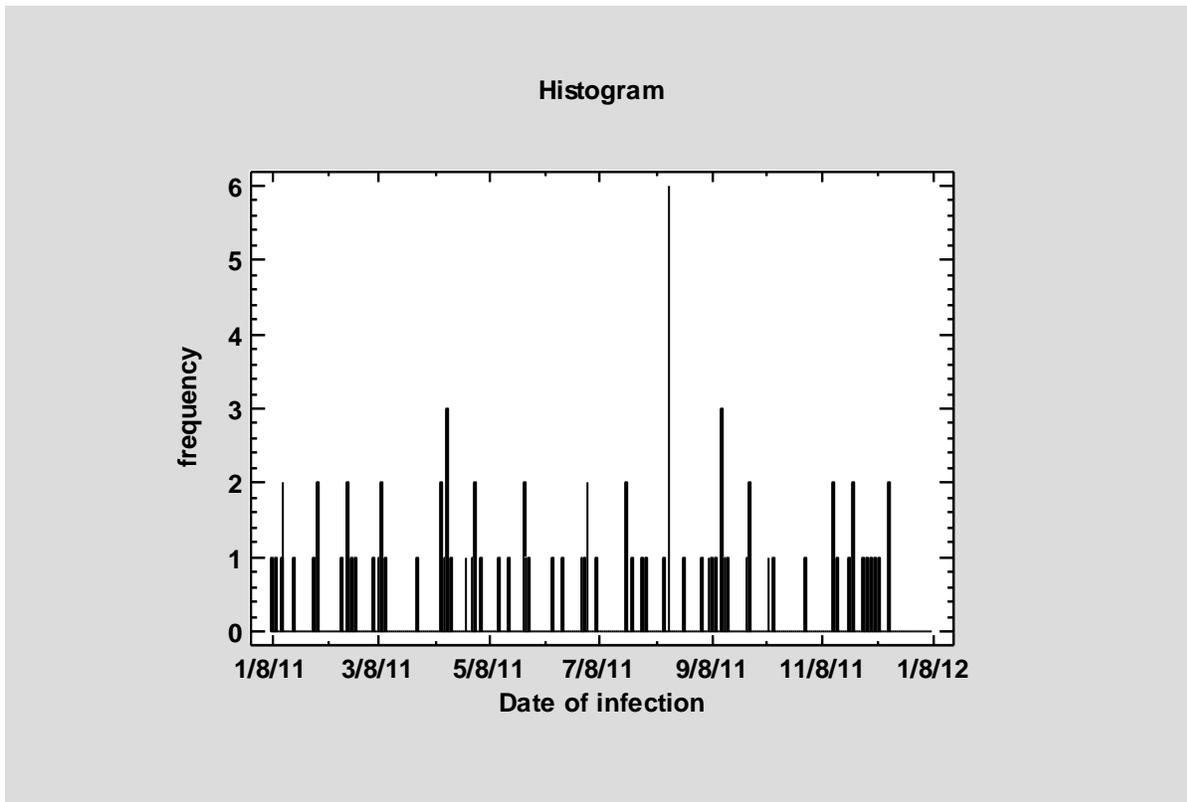
Sample StatFolio: *gchart.sgp*

Sample Data

The file *infections.sgd* contains the dates at which infections occurred at a hospital during 2011. Since there were 89 infections, this results in a total of $m = 88$ measurements of the time between infections. The first several rows of that file are shown below:

Date of infection
1/4/11
1/5/11
1/7/11
1/8/11
1/10/11
1/13/11

The chart below shows this data plotted as a histogram, with one class for each day of the year:



The number of infections on a given day ranged between 0 and 6.

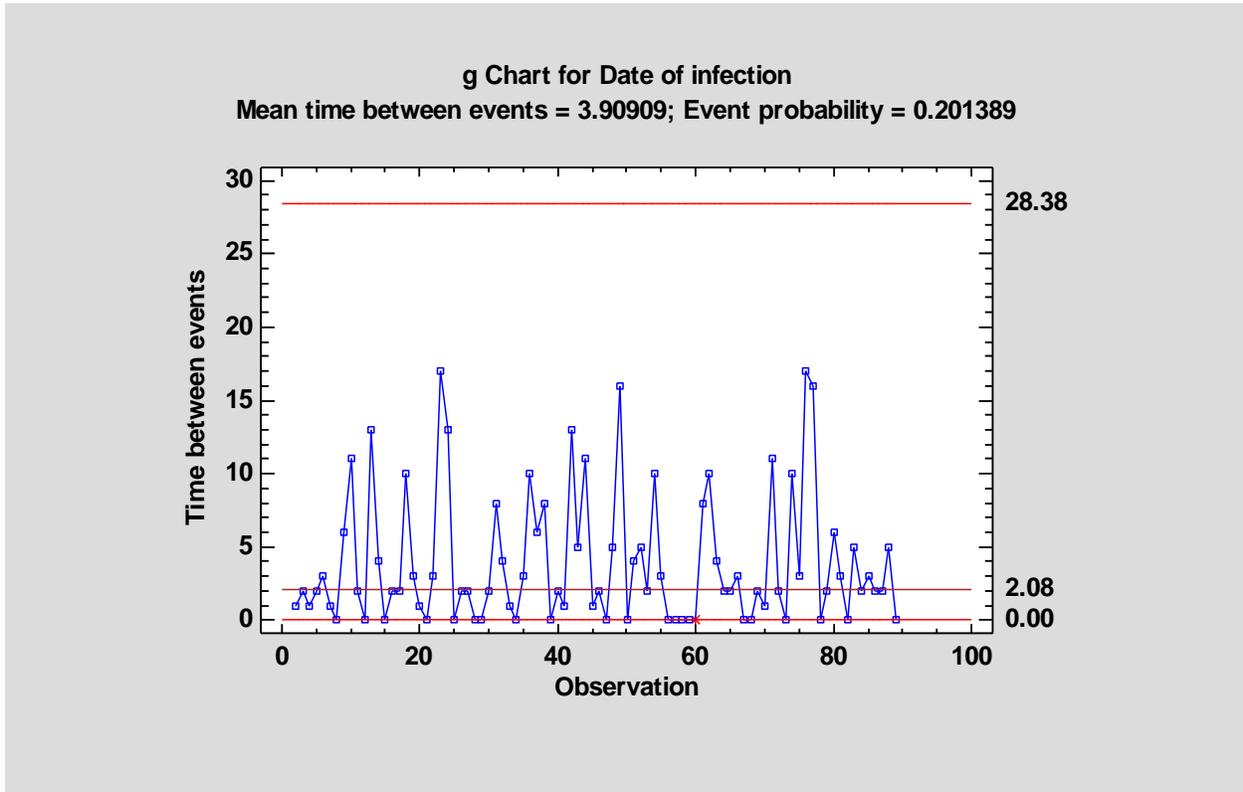
Data Input

The data input dialog box for the g chart is shown below:

- **Data:** the data to be analyzed in the format specified by the *Type of data* radio buttons.
- **Type of data:** an indication of the format in which the data have been entered. If *Time of occurrence* is selected, then the data specify the time or trial number at which each event occurred. If *Time between occurrences* is selected, then the data represent the number of opportunities between each event (not counting the event itself). Note that the g chart assumes that events occur at discrete points in time. Events occurring over a continuous interval should be plotted using a *t chart*.
- **Labels:** optional labels for each data value.
- **LSL, Nominal, USL:** optional lower specification limit, nominal (target) value, and upper specification limit for the time between events.
- **Select:** subset selection.

G Chart

The *G Chart* plots the time between events x :



It is assumed that the time between events follows a geometric distribution where the probability that the time between events equals x is given by:

$$p(x) = p(1 - p)^x \tag{2}$$

The parameter p represents the probability that an event occurs whenever there is an opportunity for an event. In the example, p represents the probability that there are $x = 0$ days until the next event. For a geometric distribution, the mean is

$$\mu = \frac{1 - p}{p} \tag{3}$$

and the standard deviation is

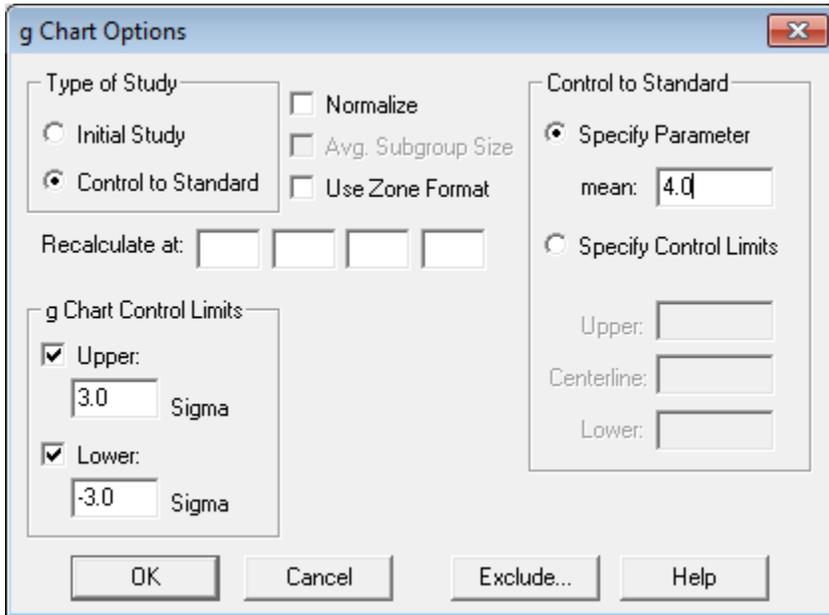
$$\sigma = \frac{\sqrt{1 - p}}{p} \tag{4}$$

In Phase 1 (*Initial Studies*) mode, the centerline and control limits are determined from the data. To calculate the limits, the maximum likelihood estimator for p is first calculated from

$$\hat{p} = \frac{\left(\frac{m-1}{m}\right)}{\bar{x} + 1} \tag{5}$$

where \bar{x} is the average of the m interevent times. Using the estimated value of p , the centerline is placed at the 50th percentile of the geometric distribution. The lower and upper control lines are placed at the 0.135th percentile and the 99.865th percentile (assuming that 3-sigma limits are requested).

In Phase 2 (*Control to Standard*) mode, the user specifies the assumed mean time between events on the *Analysis Options* dialog box, as shown below:



Any points beyond the control limits will be flagged using a special point symbol. Any point excluded from the analysis, usually by clicking on a point on the chart and pressing the *Exclude/Include* button, will be indicated by an X. If so indicated on the *Pane Options* dialog box, unusual sequences of points may also be flagged. In the current chart, all samples are within the control limits, although one point is flagged because it is the fifth consecutive interval that is equal to 0.

Pane Options

- **Outer Warning Limits:** check this box to add warning limits at the specified multiple of sigma, usually at 2 sigma.
- **Inner Warning Limits:** check this box to add warning limits at the specified multiple of sigma, usually at 1 sigma.
- **Moving Average:** check this box to add a moving average smoother to the chart. In addition to the sample data, the average of the most recent q points will also be displayed, where q is the *order* of the moving average.
- **Exponentially Weighted Moving Average:** check this box to add an EWMA smoother to the chart. In addition to the sample data, an exponentially weighted moving average of the sample data will also be displayed, where λ is the smoothing parameter of the EWMA.
- **Decimal Places for Limits:** the number of decimal places used to display the control limits.
- **Mark Runs Rules Violations:** flags with a special point symbol any unusual sequences or runs. The runs rules applied by default are specified on the *Runs Tests* tab of the *Preferences* dialog box.
- **Color Zones:** check this box to display green, yellow and red zones.
- **Display Specification Limits:** whether to add horizontal lines to the chart displaying the location of the specification limits (if any).

Note: the warning limits are placed relative to the control limits. By default, they are located at one-third and two-thirds of the distance to the “3-sigma” limits.

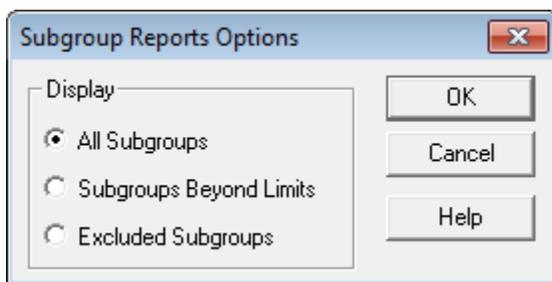
G Chart Report

This pane tabulates the values plotted on the control chart:

G Chart Report		
All Observations		
X = Excluded * = Beyond Limits		
<i>Observation</i>	<i>Date of infection</i>	<i>Time between events</i>
1	1/4/11	
2	1/5/11	1.0
3	1/7/11	2.0
4	1/8/11	1.0
5	1/10/11	2.0
6	1/13/11	3.0
7	1/14/11	1.0
8	1/14/11	0.0
9	1/20/11	6.0
10	1/31/11	11.0
11	2/2/11	2.0
12	2/2/11	0.0
13	2/15/11	13.0
14	2/19/11	4.0
15	2/19/11	0.0
16	2/21/11	2.0
17	2/23/11	2.0
18	3/5/11	10.0
19	3/8/11	3.0
20	3/9/11	1.0

Out-of-control points are indicated by an asterisk. Points excluded from the calculations are indicated by an X.

Pane Options



- **Display:** specify the observations to display in the report.

Analysis Summary

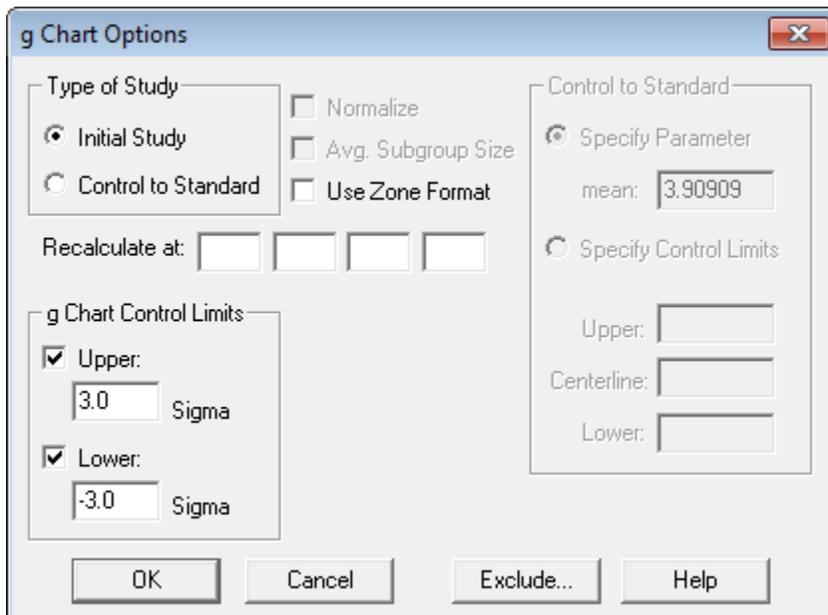
The *Analysis Summary* summarizes the data and the control chart:

<u>G Chart - Date of infection</u>	
Number of observations = 88	
0 observations excluded	
g Chart	
<i>Period</i>	#1-89
UCL: +3.0 sigma	28.3829
Centerline	2.08228
LCL: -3.0 sigma	0.0
0 beyond limits	
Estimates	
<i>Period</i>	#2-89
Mean time between events	3.90909
Event probability (p)	0.201389

Included in the table are:

- **Sample Information:** the number of interevent times m and the number of observations (if any) that have been excluded from the calculations.
- **g Chart:** a summary of the centerline and control limits for the g chart. Using *Analysis Options*, separate limits may be calculated for different periods (sets of samples).
- **Estimates:** estimates of the process mean time between events $\hat{\mu}$ and the event probability \hat{p} .

Analysis Options



- **Type of Study:** determines how the control limits are set. For an *Initial Study* (Phase 1) chart, the limits are estimated from the current data. For a *Control to Standard* (Phase 2) chart, the control limits are determined from the information in the *Control to Standard* section of the dialog box.
- **Use Zone Format:** if checked, the statistics will be plotted using a zone chart rather than the usual format.
- **Recalculate at:** the control limits of the chart may be recalculated at up to 4 locations along the X axis by specifying the sample numbers at which new estimates are to be initiated. Separate estimates of the mean time between events will be obtained using the data in the different sections. In such cases, the control limits will be adjusted at the start of each new period.
- **g Chart Control Limits:** specify the multiple k to use in determining the upper and lower control limits on the g chart. To suppress a limit completely, enter 0.
- **Control to Standard:** to perform a Phase 2 analysis, select *Control to Standard* for the *Type of Study* and then enter either of the following:

Specify Parameters: specify the established standard mean time between events.

Specify Control Limits: specify the location of the centerline and control limits exactly where you wish them to be placed.

- **Exclude:** use this button to exclude specific samples from the calculations.

Runs Tests

The *Runs Tests* pane displays the results of standard tests applied to the g chart to look for unusual sequences of points.

Runs Tests	
Rules	
(A) runs above or below centerline of length 8 or greater.	
(B) runs up or down of length 8 or greater.	
(H) 5 consecutive zeroes (alpha = 0.135%)	
Violations	
<i>Observation</i>	<i>g Chart</i>
60	H

Depending on the default settings on the *Runs Tests* tab of the *Preferences* dialog box, STATGRAPHICS will look for up to 8 different types of patterns:

- A: A group of 8 or more points, all above or all below the centerline.
- B: A group of 8 or more points, all increasing or all decreasing.
- C: A group of 5 points in which at least 4 are more than 1-sigma away from the centerline, on the same side of the centerline.
- D: A group of 3 points in which at least 2 are more than 2-sigma away from the centerline, on the same side of the centerline.
- E: A group of 15 or more points, all within 1-sigma.
- F: A group of 8 or more points, all beyond 2-sigma, but not necessarily on the same side of the centerline.
- G: A group of 8 or more points, all following an alternating up and down pattern.
- H: A group of k or more points, all of which equal 0, where k is the smallest integer that satisfies the inequality

$$k \geq \frac{\log(\alpha_{LCL})}{\log(p)} \tag{6}$$

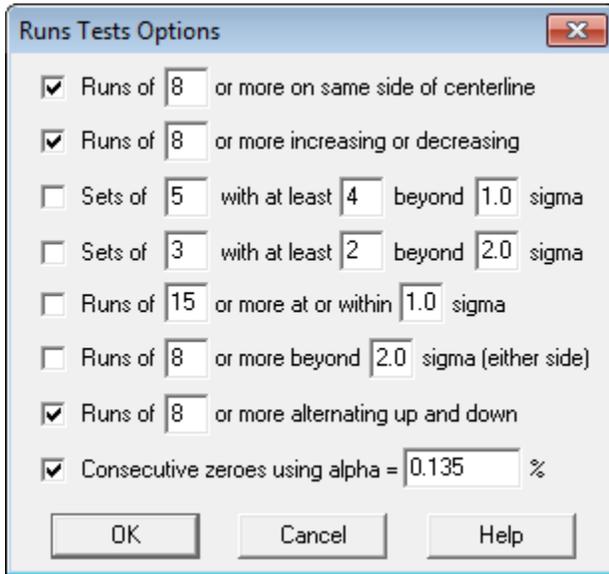
and α_{LCL} is the false alarm probability associated with the lower control limit.

Any such runs will be indicated in the above table and also on the control charts (unless suppressed).

Test H, due to Benneyan (1999), is often the best way to detect a sudden increase in the rate of events, since the lower control limit is often equal to 0 and therefore not able to signal a significant increase.

Note: For run tests C through F, the “1-sigma” and “2-sigma” limits are replaced by one-third and two-thirds of the distance to the control limits, respectively.

Pane Options



Select the runs tests to be applied and the parameters that define those tests.

Save Results

The following results can be saved to the datasheet:

1. *Times* – the interevent times.
2. *Labels* – the sample labels.
3. *Mean* – the estimated mean time between events.
4. *p* – the estimated probability of an event.

References

Benneyan, James C. (1999) *Geometric-Based g-Type Statistical Control Charts for Infrequent Adverse Events: New Quality Control Charts for Hospital Infections*. Institute of Industrial Engineers Society for Health Systems 1999 Conference Proceedings.

Benneyan, James C. (2001) *Number-Between g-Type Statistical Quality Control Charts for Monitoring Adverse Events*. Health Care Management Science 4, 305-318.