Applied Technology

www.e-AT-USA.com

A Bell Shaped Distribution Does NOT Imply Only Common Cause Variation

by John J. Flaig, Ph.D.

The notion that if data from a process has a "bell shaped" histogram, then the system is experiencing only common cause variation is incorrect. This is serious error because it reflects a fundamentally incorrect understanding of the sources of variation. However, even knowledgeable people sometime make this mistake. For example, here is a quote from a popular Six Sigma textbook, "When most values fall in the middle and tail off in either direction we have statistical evidence of common cause variation". This is not correct and the misunderstanding is probably caused by the belief that the Central Limit Theorem applies to all systems, which is also an incorrect assumption. Hence, a histogram is not a robust tool for determining the type of variation a process is experiencing. However, it can give us a clue to the answer, but a control chart is the best way to arrive at the correct conclusion.

A mathematical proof is given below for the argument that a Normal or "bell shaped" histogram does not imply that the system is experiencing only random variation and inversely a system experiencing only random variation will not necessarily produce a Normal distribution.

Theorem: Normal ←/→ Random (i.e., Normal does not imply Random, and Random does not imply Normal)

Proof:

Part 1. (not ←) The proof that this implication is false is obvious as you can generate random distributions that are uniform, triangular, Weibull, Poisson, binomial, etc., and yes even Normal (see JMP or Minitab for examples).

Part 2. (not →) The proof that this implication is false is by counter example provided by Dr. Fred Khorasani and is given below. In this example the histogram is "bell shaped" but the system is experiencing both special cause (systematic) variation and common cause (random) variation. In the graph the slope of the trend line characterizes special cause (systematic) variation and common (random) variation is characterized by the spread of the point about the trend line.

Sales data in thousands of dollars:

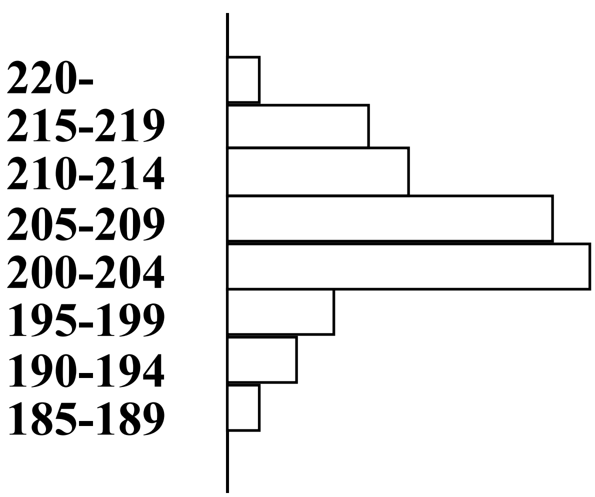
188, 191, 194, 198, 199, 196, 204, 202, 203

204, 200, 200, 201, 204, 203, 202, 206, 206

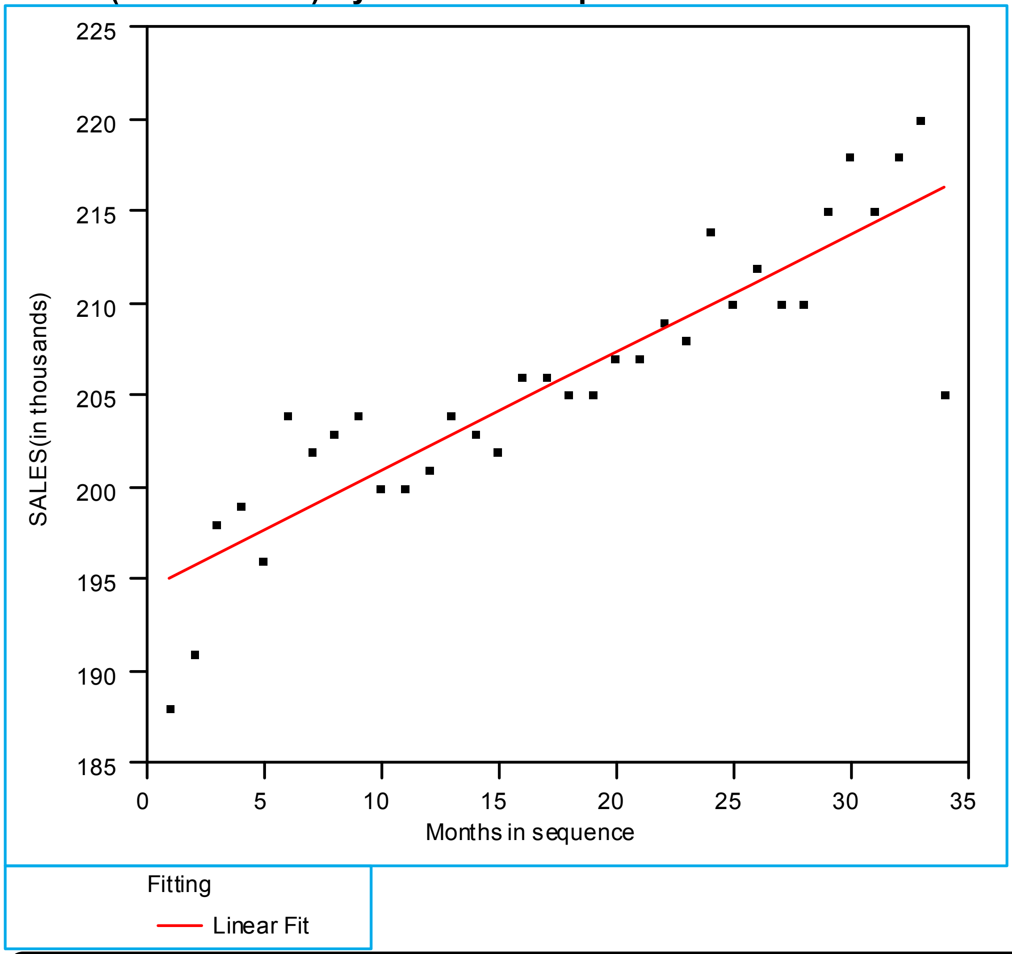
Ź205, 205, 207, 207, 209, 208, 214, 210, 212,

Ź210, 210, 215, 218, 215, 218, 220, 205

Histogram of the sales data:



SALES(in thousands) By Months in sequence



Linear Fit

Summary of Fit

Rsquare	0.78425
Root Mean Square Error	3.415869
Mean of Response	205.7353
Observations (or Sum Wgts)	34

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1357.2364	1357.24	116.3196
Error	32	373.3812	11.67	**Prob>F**
C Total	33	1730.6176		0.0000

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	194.46524	1.19797	162.33	0.0000
Months in sequence	0.6440031	0.05971	10.79	0.0000

Applied Technology

www.e-AT-USA.com

John J. Flaig, Ph.D.
Managing Director
Applied Technology
Tel: 408-266-5174
E-mail: JohnFlaig@yahoo.com
Website: www.e-AT-USA.com

© 2004 Applied Technology. All rights reserved.