Multivariate Statistical Process Control Charts

Dr. Austin Brown

Kennesaw State University

Introduction

- In modern organizations, processes have become increasingly complex and intertwined with other processes.
- Think about our Dunkin example from early in the semester. We drew out a flow chart showing that customer satisfaction, our ultimate goal, can be measured by several different items, including speed of service and perceived quality.
- Or in a manufacturing example, suppose we need to monitor both the inner and outer diameter of a bearing to ensure the bearing's usefulness for whatever its intended purpose is.

Introduction

- We can (and certainly people do) use individual control charts for each of the quality characteristics we're monitoring.
- For example, in the bearing monitoring example, we could use a Shewhart $\bar{X}$ and s chart for both the inner and outer diameters. However, this may not be optimum.

Introduction

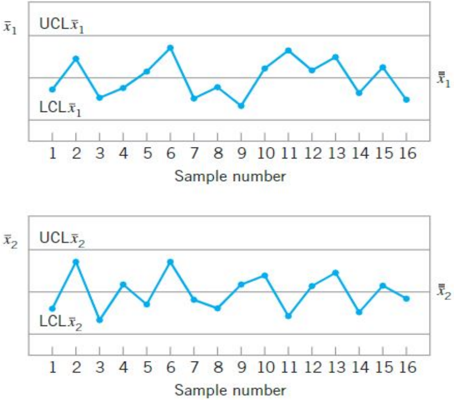

Figure 1: Figure 11.1 from Text

Introduction

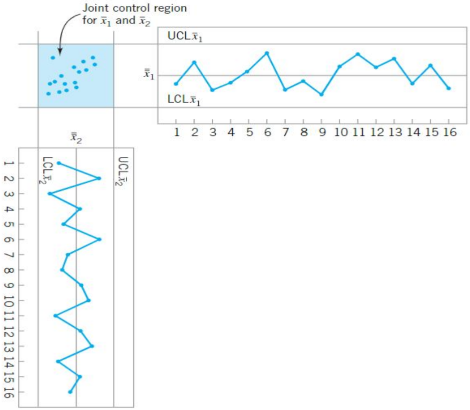

Figure 2: Figure 11.2 from Text

Introduction

- So it's clear that simultaneous monitoring of these two univariate processes is giving us insight that isn't so clear with separate monitoring.

- However, this procedure illustrated by Figure 11.2 isn't super useful beyond three processes (how do we visualize four dimensions??).

- Moreover, such a procedure becomes incredibly cumbersome when you're trying to set up and maintain multiple charts (especially Shewhart $\bar{X}$ and R or s charts as they're already two chart schemes!).

Introduction

- We also run into another problem. Suppose the inner and outer diameters of the bearings are independent of each other. Further suppose that we set up individual $\bar{X}$ and s charts for both processes such that the probability of a false alarm is α for both charts.

- What's the probability that both charts (focused just on $\bar{X}$ here) plot in between their respective control limits?

$$P\left[(LCL_1 < \bar{x}_1 < UCL_1) \cap (LCL_2 < \bar{x}_2 < UCL_2) \middle| \text{IC}\right] =$$

$$P\left[(LCL_1 < \bar{x}_1 < UCL_1) \middle| \text{IC}\right] \times P\left[(LCL_2 < \bar{x}_2 < UCL_2) \middle| \text{IC}\right] = (1-\alpha)^2$$

Introduction

- This then suggests that the true probability of a false alarm is:

$$\alpha^{'} = 1 - (1 - \alpha)^2$$

- So for example, if $\alpha = 0.0027$, then $\alpha^{'} = 0.0054$.

- This is obviously not ideal. So what do we do? We can take advantage of control charts specifically designed for monitoring multivariate processes.

A Review of the Multivariate Normal Distribution

- But first, we need to review the multivariate normal distribution. Recall the univariate distribution:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{1}{2}(\frac{x - \mu}{\sigma})^2 \right), \quad x \in \mathbb{R}$$

- We can think of the exponentiated term as the standardized square distance an observation, x, is from its mean, μ. We can rewrite this as:

$$(x - \mu)(\sigma^2)^{-1}(x - \mu)$$

A Review of the Multivariate Normal Distribution

- This general structure is referred to as the quadratic form and is foundational in multivariate statistical methods.
- The multivariate normal distribution is a generalization of the univarite normal distribution where we have p-jointly distributed normal variables with possibly differing means and variances, but also potentially having covariance, too.

$$f(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}((\mathbf{x}-\mu)^{\mathbf{T}}\Sigma^{-1}(\mathbf{x}-\mu))\right)$$

A Review of the Multivariate Normal Distribution

- You don't need to worry too much about some of the complicated aspects of the multivariate normal distribution as software will do the heavy lifting for us, but it is important to notice that we have a quadratic form in the exponentiated term, here, too.

A Review of the Multivariate Normal Distribution

- As was the case in univariate monitoring, we have a Phase I and Phase II component. As before, many times we don't have a good estimate of the mean or variance and thus have to estimate them using Phase I data. How do we do that with multivariate data?

- When we take a sample of size n with multivariate data, what we get is n-vectors each with p-variables

$$\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$$

where:

$$\mathbf{x_i} = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{p1} \end{bmatrix}$$

A Review of the Multivariate Normal Distribution

- So then the sample mean will also be a vector where each element is the univariate sample mean of the variable held in its position.

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i}$$

A Review of the Multivariate Normal Distribution

- What about the variance-covariance matrix? Let's first consider the structure of the matrix when it's known:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \sigma_{p3} & \dots & \sigma_p^2 \end{bmatrix}$$

A Review of the Multivariate Normal Distribution

- When Σ is unknown, we have to estimate it. All of the diagonal elements, being the variances, can be estimated in the typical way:

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

- and all of the off-diagonal elements, being the covariances, can be estimated by:

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

A Review of the Multivariate Normal Distribution

- We can generalize this into one step (typically done using software):

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T$$

- Let's see how we can do this using R.

The Hotelling's T^2 Control Chart

- Now that we know a little bit more about how to estimate the multivariate mean and variance-covariance matrix, how can this be used in a control charting scheme.

- One of the most well-known and widely used multivariate control charts is called, "Hotelling's T^2 Control Chart" or "Hotelling's T^2" for short. Hotelling developed this chart during WW2 as a way of monitoring bombsight data.

- This chart is the multivariate analogue to the Shewhart $\bar{X}$ chart.

The Hotelling's T^2 Control Chart

- As before, we are most likely going to have to estimate the mean vector and variance-covariance matrix using Phase I data.

- Like in the univariate case, the text recommends taking m samples of size n and estimating the mean vector and v-c matrix by averaging the sample statistics (means, variances, and covariances, in this case) over the m samples.

The Hotelling's T^2 Control Chart

- Once we have our estimates, we can construct the T^2 plotting statistic as:

$$T^2 = n(\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}})^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}})$$

- If the process is in-control, T^2 follows an F distribution. The text recommends using different control limits in Phase I and in Phase II.

The Hotelling's T^2 Control Chart

- For Phase I:

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$
$$LCL = 0$$

- For Phase II:

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$
$$LCL = 0$$

The Hotelling's T^2 Control Chart

- Let's look at an example using data which jointly monitors the tensile strength and diameter of textile fibers using data from this week's Excel file.

The Hotelling's T^2 Control Chart

- If we don't have points plotting out-of-control in Phase I, then this is evidence to us that we have statistical control and that our estimated mean vector and v-c matrix are adequate for Phase II monitoring.

- So somewhat unlike the Shewhart $\bar{X}$ Chart, we don't bring our control limit forward per se, but we do bring forward our mean vector and v-c matrix that we used in Phase I.