Monitoring Quality with Principal Components

- Conventional multivariate control-charting procedures are reasonably effective as long as p (the number of process variables) is not very large.
- Today, it is not uncommon for data to be collected on tens or even hundreds of process variables in electronic, chemical, and manufacturing processes.
- As p increases, the ARL performance to detect a specified shift in the mean of these variables for multivariate control charts also increases, because the shift is "diluted" in the p-dimension space of process variables.
- In the situations where it is suspected that most of the variability in the process is in a relatively small subset of process variables, sensitivity to detect special causes of variation can be enhanced based on principal components.

Phase I: Two-Part Procedure

- First part: construct an ellipse format chart (chi-square chart for two variables) for the scores of the first two principal components for each observation.
- Second part: create a chart from the remaining principal components.
- Example 8.10 and 8.11 of J&W

Phase I Charts for Ex 8.10&11

First part:

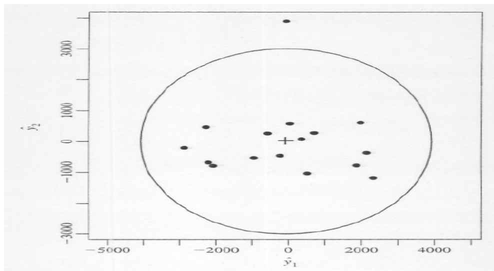
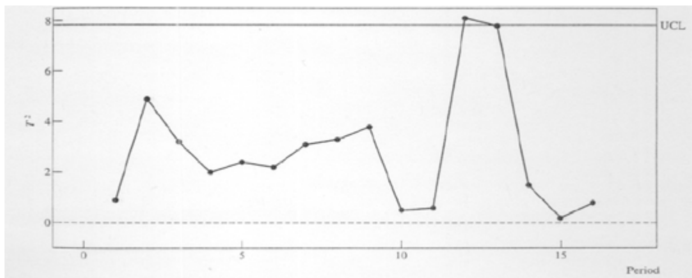


Figure 8.7 The 95% control ellipse based on the first two principal components of overtime hours.

Second part:



79

Factor Analysis

80

Introduction

- Factor analysis (FA) is to describe the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called **factors**.
- The goal of factor analysis is to reduce the redundancy among the variables by using a smaller number of factors.
- Factor analysis can be considered an extension of principal component analysis. The major differences between FA and PCA are:
  - Principal components are defined as linear combinations of the original variables. In FA, the original variables are expressed as linear combinations of the factors
  - In PCA, we explain a large part of the total variance of the variables. In FA, we seek to account for the covariances or correlations among the variables.

81

Factor Analysis Model

- The observed random vector **X**, with p components has mean μ and covariance matrix Σ. The factor model assumes that **X** is linearly dependent on a few unobservable random variables $F_1, F_2, \ldots, F_m$, called **common factors**. Let $\mathbf{F}=[F_1, \ldots, F_m]^T$, in matrix notation,

  $\mathbf{X}-\mu=\mathbf{L}\mathbf{F}+\varepsilon$
- The p-by-m matrix **L** is the matrix of **factor loadings**.
- $F_1, F_2, \ldots, F_m, \varepsilon_1, \ldots, \varepsilon_p$ are all *unobservable random variables*.

82

Orthogonal Factor Model with *m* Common Factors

- Orthogonal factor model: $\mathbf{X}=\boldsymbol{\mu}+\mathbf{LF}+\boldsymbol{\varepsilon}$ with $E[\mathbf{F}]=\mathbf{0}$, $\text{cov}[\mathbf{F}]=\mathbf{I}$, $E[\boldsymbol{\varepsilon}]=\mathbf{0}$, $\text{cov}[\boldsymbol{\varepsilon}]=\boldsymbol{\Psi}$, a diagonal matrix, and $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent.
- Under the orthogonal factor model,
  - $\text{cov}[\mathbf{X}]=\mathbf{LL}^T+\boldsymbol{\Psi}$
  - $\text{cov}[\mathbf{X}, \mathbf{F}]=\mathbf{L}$
- The variance of X_i, $i=1, \ldots, p$ due to the m common factors is called the i^{th} **communality**. The portion of variance of X_i due to ε_i is called **specific variance**.
- Example 9.1 of J&W

83

Determination of Factor Loadings

- Factor analysis is most useful when m is small relative to p to provide "simple" explanation of the covariation in $\mathbf{X}$.
- Unfortunately, most covariance matrices cannot be factored when m is much less than p (Example 9.2).
- When $m>1$, there is always some inherent ambiguity associated with the factor model. Factor loadings $\mathbf{L}$ are determined only up to an orthogonal matrix $\mathbf{T}$. Thus the loadings $\mathbf{L}^*=\mathbf{LT}$ and $\mathbf{L}$ both give the same representation. The communalities are also unaffected by the choice of $\mathbf{T}$.
- Factor analysis strategy:
  - Uniquely estimate $\mathbf{L}$ and $\boldsymbol{\Psi}$ by imposing additional conditions.
  - Rotate loading matrix based on some criteria.

84