Retrospective

Handout data set (samples of $n=5$)

Need to make (provisional) estimates of μ, σ.

(Use 1st 20 samples)

$$\bar{\bar{x}} = 5.16 \quad \leftarrow \text{use to estimate } \mu$$

$$\bar{R} = 3.6 \qquad \mu_R = d_2(n)\cdot\sigma$$

$$\hat{\sigma} = \frac{\bar{R}}{d_2(5)} = \frac{3.6}{2.326} = 1.548 \quad \leftarrow \text{use to estimate } \sigma.$$

$$UCL_{\bar{x}} = 5.16 + 3\cdot\frac{1.548}{\sqrt{5}} = 7.24$$

$$LCL_{\bar{x}} = 5.16 - 3\cdot\frac{1.548}{\sqrt{5}} = 3.08$$

$$CL_R = d_2(5)\cdot\hat{\sigma} = d_2(5)\cdot\frac{\bar{R}}{d_2(5)} = \bar{R} = 3.6$$

$$\underline{UCL_R} = D_2(5)\cdot\hat{\sigma} = D_2(5)\cdot\frac{\bar{R}}{d_2(5)}$$

$$= \underline{D_4(5)\cdot\bar{R}} \qquad \left(D_4(n) = \frac{D_2(n)}{d_2(n)}\right)$$

$$= (2.115)\cdot(3.6)$$

$$= 7.614.$$

By the way, the retrospective limits for $\bar{x}$ are of the form

$$\bar{x} \pm \underbrace{\left(\frac{3}{d_2(n)\sqrt{n}}\right)}_{A_2(n)} \cdot \bar{R}$$

Overview of §4.4 of V&J — variation on $\bar{x}/R$ chart analysis where natural sample size is $n=1$.

Shewhart Control Charting? One (unfortunately) common device is to plot both individuals and "moving ranges" — X/MR charting.

If consecutive samples of $n=1$ yield

$x_1, x_2, \cdots$

at time t plot

1) x_t on an "individuals" chart

($\bar{x}$ chart with $n=1$)

and 2) for $t \geq 2$ also plot

$MR_t = |x_t - x_{t-1}|$ on a

"moving range" chart

Control Limits?

Standard practice: X: $UCL_x = \mu + 3\sigma$

$LCL_x = \mu - 3\sigma$

MR: $UCL_{MR} = \underbrace{D_2}_{3.686}(2) \cdot \sigma$

"Better":
(with $\mathcal{M}$ and $\mathcal{R}$ suitably chosen)

X: $UCL_x = \mu + \mathcal{M} \cdot \sigma$

$LCL_x = \mu - \mathcal{M} \cdot \sigma$

MR: $UCL_{MR} = \mathcal{R} \cdot \sigma$

See Table 4.16 on page 172 of V&J.

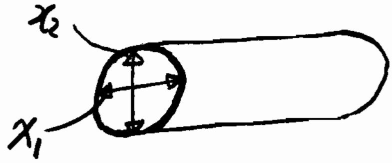
A related issue discussed in §4.4 of V&J is estimation of σ in $n=1$ situations —

$x_1 \sim N(\mu_1, \sigma^2)$, $x_2 \sim N(\mu_2, \sigma^2)$, $x_3 \sim N(\mu_3, \sigma^2)$, ...

There is really no completely honest way to estimate σ — about the best you can do is $\dfrac{\overline{MR}}{d_2(2)}$ ← this will tend to be conservative (but is less affected than other estimators by nonconstant mean).

§4.3 of V&J: Shewhart Control Charting for Multivariate Data

Example 2 distinguishable (and perpendicular) diameters on a "cylindrical" part



both x_1 and x_2 being slightly large (or slightly small) makes the part slightly big (or small) but of correct shape.

On the other hand one a little large and the other a little small produces a bad shape — I care about the relationship between x_1 and x_2.

Multivariate charts intend to monitor p variables simultaneously in contexts where relationships between them matter.

Set-up:

single observation $\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$

an average of n observations $\underset{\sim}{\bar{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$

"Stable Process" Model:

$$E(\underset{\sim}{x}) = E\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \underset{\sim}{\mu}$$

$$\operatorname{Cov} \underset{\sim}{x} = \underset{\sim}{V}$$

$$= \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \cdots\cdots \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{pmatrix}$$

$$E\underset{\sim}{\bar{x}} = \underset{\sim}{\mu}$$

$$\operatorname{Cov} \underset{\sim}{\bar{x}} = \frac{1}{n}\underset{\sim}{V}$$

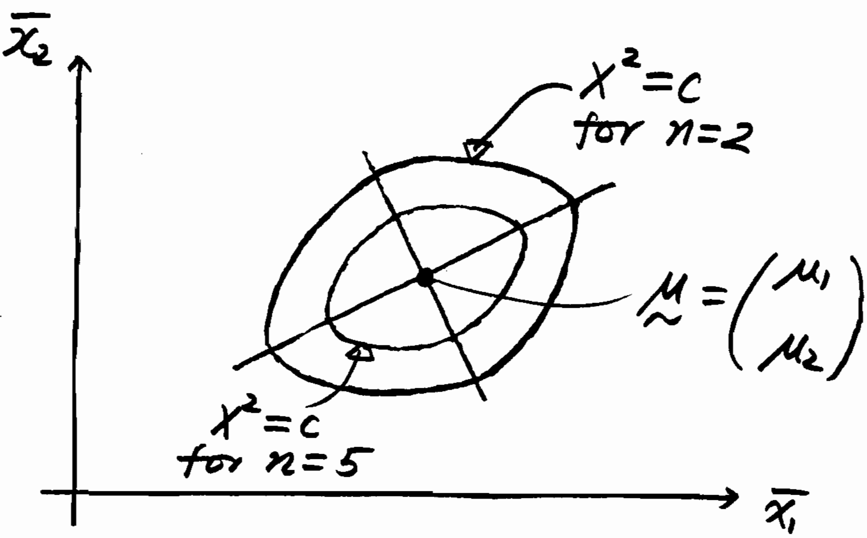
A (univariate) Summary Statistic:

$$Q = \chi^2 = \underset{1\times p}{(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'} \underset{p\times p}{\left(\frac{1}{n}\underset{\sim}{V}\right)^{-1}} \underset{p\times 1}{(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})}$$

$$= n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' \cdot \underset{\sim}{V}^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})$$

"Constant χ^2" surfaces in p-space are multivariate footballs centered at $\underset{\sim}{\mu}$ with size and orientation determined by $\underset{\sim}{V}$ (and n).

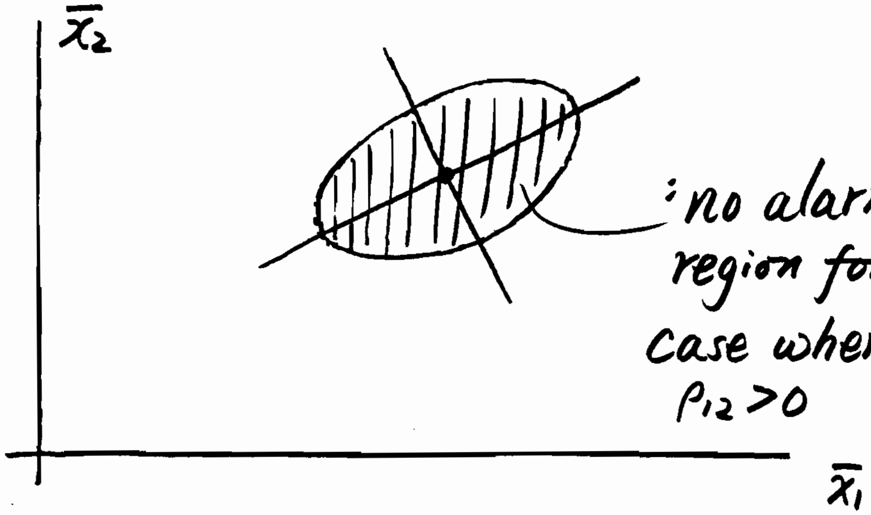
A $p=2$ picture (for $\rho_{12} > 0$)



- The "bigger" $\underset{\sim}{V}$ the larger the footballs.

- $\rho_{12} < 0$ orients footballs in the other direction.

Compute and plot X^2 using only an upper control limit



What control limit to use?

Basic probability fact used is

$$\underset{\sim}{\bar{x}} \sim MVN \implies X^2 \sim \chi_p^2$$

(Multivariate normal)

Example (of this argument) $p = 2$, $\rho_{12} = 0$

$\bar{x}_1$ and $\bar{x}_2$ are independent

$$\underset{\sim}{V} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

$$\underset{\sim}{V}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix}$$

$$Q = X^2 = n\left[\left(\frac{\bar{x}_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{\bar{x}_2-\mu_2}{\sigma_2}\right)^2\right]$$

$$= \left(\frac{\bar{x}_1-\mu_1}{\underbrace{\sigma_1/\sqrt{n}}}\right)^2 + \left(\frac{\bar{x}_2-\mu_2}{\underbrace{\sigma_2/\sqrt{n}}}\right)^2$$

$$\sim \chi^2_2$$

Recall that $W \sim \chi^2_p \Rightarrow EW = p$

$\text{Var}\, W = 2p$

So standards given 3-sigma control limits for X^2 are

$$CL_{X^2} = p$$

$$UCL_{X^2} = p + 3\sqrt{2p}$$

Retrospective analysis?

I need ways to take several p-dimensional samples and get estimates of (supposedly constants) $\underset{\sim}{\mu}$ and $\underset{\sim}{V}$.

$\bar{\underset{\sim}{x}}_i$ = sample mean vector for sample i

and if $n_i > 1$:

$$\hat{\underset{\sim}{V}}_i = \text{sample variance-covariance matrix for period } i$$

$$\underset{\sim}{\bar{x}}_{pooled} = \sum_{i=1}^{r} n_i \underset{\sim}{\bar{x}}_i \Big/ \sum_{i=1}^{r} n_i$$

can be used to estimate $\underset{\sim}{\mu}$

and $$\hat{\underset{\sim}{V}}_{pooled} = \left(\frac{1}{\sum_{i=1}^{r} n_i - r} \right) \sum_{i=1}^{r} (n_i - 1) \cdot \hat{\underset{\sim}{V}}_i$$

can be used to estimate $\underset{\sim}{V}$.

Retrospective analysis with all $n_i = 1$ requires some means of estimating $\underset{\sim}{V}$?!?!

See V&J problem 4.35 for a possible method (like the univariate

$$\hat{\sigma} = \frac{\overline{MR}}{d_2}$$

method).

For very large p, reduction of data from a given period to a single 1-dimensional Q may not be so helpful —

Example.

50 3-d positions of identifiable points (landmarks) on an auto body.

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ x_2 \\ y_2 \\ z_2 \\ \vdots \\ x_{50} \\ y_{50} \\ z_{50} \end{pmatrix}$$

I might compute an overall X^2...
practical interest will center on interpretable summaries of this 150-dimensional data vector...
They will need to be monitored as well.

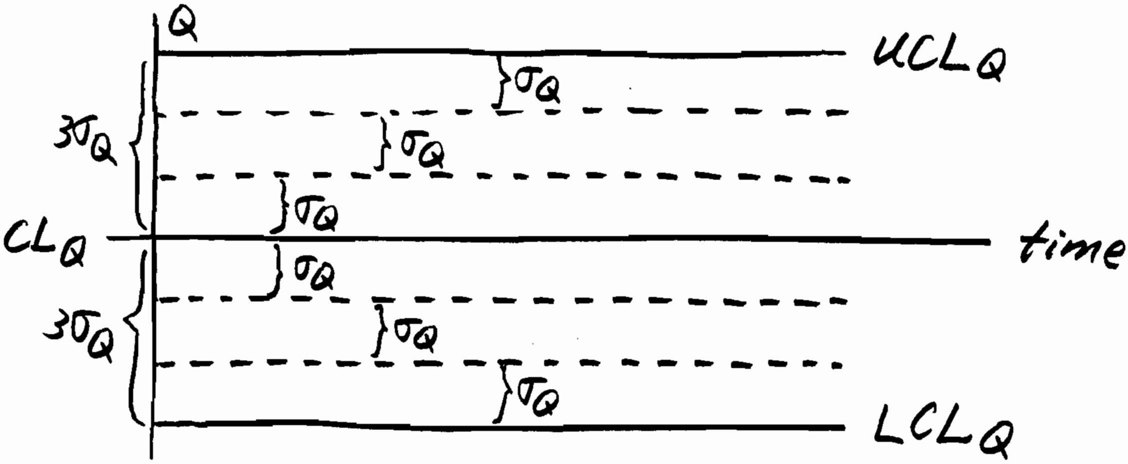
Current best practice seems to be also (in addition to overall X^2 statistic) keeping track of summaries related to principal components (derived from $\underset{\sim}{X}$)

---

§3.4, 3.5 of V&J.

There is more to be seen in a plot of Q's than only whether they are inside control limits — people examine plots looking for trends and interpretable patterns.

Over the years various systems have been developed for identifying "nonrandom" patterns — The most famous is the set of "WE Alarm Rules" (page 90 of V&J)



(might) Alarm if:

1) one point outside 3 sigma limits
2) two out of any three consecutive points outside of 2 sigma limits on one side of center
3) four out of any five consecutive points outside 1 sigma limits on one side of center
4) eight straight points fall on a single side of center

 Pandora's box is open!!

How to judge or choose??

Basic notion used to quantify the properties of process monitoring schemes is that of "run length" — i.e., if $Q_1, Q_2, \cdots$ are being fed into an algorithm used to produce "out of control" signals, let

T = the period/index at which the monitoring scheme 1st issues an "out of control" signal

run length

probability model for evolution of $Q_1, Q_2, \cdots$ + form of monitoring scheme $\Rightarrow$ dsn for T

run length dsn

$ET = ARL$ = average run length

$\sqrt{\operatorname{Var} T} = SDRL$ = standard deviation of run length

If all is simple, then "elementary" methods yield ARLs, etc. (see §3.5 of V&J and Ch.2 of the Notes)

Example (§3.5 of V&J)

Shewhart chart (with single alarm rule)
$Q_1, Q_2, \cdots$ iid with some (fixed marginal) dsn (could be one used to set up control limits, but it need not be)

Let $q = P[Q_1$ falls outside control limits$]$.

Then "clearly" $T \sim$ geometric (q), i.e.

$$P[T=t] = \begin{cases} q(1-q)^{t-1} & t=1,2,3,\cdots \\ 0 & \text{otherwise} \end{cases}$$

So $ARL = ET = \frac{1}{q}$.

So if, for example, $Q = \bar{x}$ and a 3-sigma chart is used:

<u>For μ, σ at standard values</u>:

$q = P[\bar{x}_1$ plots outside limits$]$

$= P[|Z|>3] \approx .0027$

$ARL = \frac{1}{.0027} = 370$

For σ at its standard value, $\mu = \text{centerline} + \sigma$ and $n=5$:

$$q = P[\bar{x}_1 \text{ plots outside limits}]$$

$$= P[Z > .764] + P[Z < -5.236]$$

$$\approx .222$$

So $ARL = \dfrac{1}{.222} = 4.5$.

Stat 531: Data for Retrospective Variables Control Charts

Time	Sample Measurements					Sum	Average, $\overline{X}$	Range, R
1	5	4	7	4	4	24	4.8	3
2	6	4	5	3	7	25	5.0	4
3	3	4	4	5	6	22	4.4	3
4	5	5	6	7	8	31	6.2	3
5	3	6	7	5	5	26	5.2	4
6	4	4	5	4	7	24	4.8	3
7	6	5	7	2	5	25	5.0	5
8	4	5	6	4	9	28	5.6	5
9	7	4	6	4	5	26	5.2	3
10	4	5	2	5	6	22	4.4	4
11	6	6	4	4	5	25	5.0	2
12	2	7	7	7	6	29	5.8	5
13	6	8	6	6	6	32	6.4	2
14	3	6	6	5	4	24	4.8	3
15	8	5	3	7	6	29	5.8	5
16	2	5	6	7	3	23	4.6	5
17	6	4	6	5	5	26	5.2	2
18	6	3	2	5	5	21	4.2	4
19	7	7	7	3	4	28	5.6	4
20	6	6	3	5	6	26	5.2	3
21	11	12	14	6	0	43	8.6	14
22	12	13	12	6	8	51	10.2	7
23	9	11	4	11	12	47	9.4	8
24	11	13	16	10	8	58	11.6	8
25	9	12	6	9	8	44	8.8	6

Note: Physical change was made at the end of time 20.

Stat 531/ Bivariate Data from Ghare and Torgeson

$p = 2$, $n = 3$
$r = 10$ $(n_i = 3)$

x_1 = time delay between detonation and explosion (10^{-2} sec)
x_2 = explosive force generated (ft lbs)

Sample	Item	x_1	x_2	Sample Mean Vector	Sample Covariance Matix
1	1	332	253	$\begin{pmatrix} 318.33 \\ 240.33 \end{pmatrix}$	$\begin{pmatrix} 152.33 & 143.83 \\ 143.83 & 136.33 \end{pmatrix}$
	2	308	230		
	3	315	238		
2	1	335	242	$\begin{pmatrix} 330.67 \\ 245.67 \end{pmatrix}$	$\begin{pmatrix} 520.33 & 561.83 \\ 561.83 & 660.33 \end{pmatrix}$
	2	351	273		
	3	306	222		
3	1	355	267	$\begin{pmatrix} 354.00 \\ 273.00 \end{pmatrix}$	$\begin{pmatrix} 421.00 & 364.50 \\ 364.50 & 351.00 \end{pmatrix}$
	2	374	294		
	3	333	258		
4	1	323	238	$\begin{pmatrix} 336.00 \\ 253.33 \end{pmatrix}$	$\begin{pmatrix} 133.00 & 127.00 \\ 127.00 & 257.33 \end{pmatrix}$
	2	340	270		
	3	345	252		
5	1	350	260	$\begin{pmatrix} 336.67 \\ 249.33 \end{pmatrix}$	$\begin{pmatrix} 358.33 & 346.67 \\ 346.67 & 341.33 \end{pmatrix}$
	2	315	228		
	3	345	260		
6	1	349	268	$\begin{pmatrix} 343.67 \\ 259.00 \end{pmatrix}$	$\begin{pmatrix} 22.33 & 44.50 \\ 44.50 & 133.00 \end{pmatrix}$
	2	340	246		
	3	342	263		
7	1	330	246	$\begin{pmatrix} 351.00 \\ 265.33 \end{pmatrix}$	$\begin{pmatrix} 330.00 & 319.50 \\ 319.50 & 380.33 \end{pmatrix}$
	2	363	285		
	3	360	265		
8	1	340	255	$\begin{pmatrix} 342.67 \\ 255.67 \end{pmatrix}$	$\begin{pmatrix} 366.33 & 419.33 \\ 419.33 & 484.33 \end{pmatrix}$
	2	325	234		
	3	363	278		
9	1	355	282	$\begin{pmatrix} 358.33 \\ 269.33 \end{pmatrix}$	$\begin{pmatrix} 57.33 & 38.33 \\ 38.33 & 220.33 \end{pmatrix}$
	2	367	273		
	3	353	253		
10	1	367	283	$\begin{pmatrix} 342.33 \\ 261.33 \end{pmatrix}$	$\begin{pmatrix} 505.33 & 439.33 \\ 439.33 & 382.33 \end{pmatrix}$
	2	337	256		
	3	323	245		

(Weighted) Average sample mean vector = $\begin{pmatrix} 341.367 \\ 257.233 \end{pmatrix}$

(Weighted) Average sample variance-covariance matrix = $\begin{pmatrix} 286.933 & 280.433 \\ 280.433 & 334.667 \end{pmatrix}$