Monitoring and data filtering
I. Classical Methods

Advanced Herd Management
Cécile Cornou, IPH

Outline
Framework and Introduction
Shewart Control chart
• Basic principles
• Examples: milk yield and daily gain
• Alarms
Moving Average Control Chart
EWMA Control Chart
→→ Break & exercises
Monitoring autocorrelation
• Model for autocorrelation
• Use EWMA
Concluding remarks

Introduction (1/3)
So far
Control: compare key figures (k) with expected results
\[ k = \theta + e_{st} + e_{ot} \]
Deviation: look if significant from a statistical point of view
If deviation: adjustment plan or/and implementation
Problem: we assume that results can be evaluated without considering results from the previous period

Introduction (2/3)
Key figures regarded as a time series of observations, treated as a whole
How to model the results?
\[ k_t = \theta + e_{st} + e_{ot} = \theta + v_t \]
\[ k_t \]: observed value of the key figure
\[ \theta \]: true underlying value
\[ e_{st} \]: sample error (biological variation)
\[ e_{ot} \]: observation error (observation method)

Results from 2 herds

Is the conclusion the same in both herds?
**The Shewart Control Chart: basic principles (1/2)**

- **Upper Control Limit (UCL)**
- **Center Line**
- **Lower Control Limit (LCL)**

Here: all the points fall inside the CL. Process in control.

**Sample quality characteristic**

- Sample number, or time

**Upper Control Limit (UCL)**

\[ \text{Center Line} = \text{target value} \]

\[ \text{UCL} = \theta' + a \sigma_t \]

\[ \text{LCL} = \theta' - a \sigma_t \]

Usually distance parameter \( a = 2 \) or 3

- If \( a = 2 \) : “2-sigma” control limit

We test the hypothesis \( H_0: \theta' = \theta \)

\( a = 2 \) corresponds to approx. 5% precision level.

**Example 1: milk yield**

Target value:

\[ \text{CL} = 25.60 \text{ kg for first lactation} \]

Control limits:

\[ \text{UCL} = \theta' + a \sigma_t \]

\[ \text{LCL} = \theta' - a \sigma_t \]

Standard deviation calculated according to number of cows behind the average.

**Control and warning limits (1/3)**

UCL and LCL determined by \( a \) (\( a = 2 \leftrightarrow p = 0.05 \))

Possible that a change in \( \theta \) is not detected (type II error)

- Lower \( a \) reduces type II but increases type I

Possible that alarm is given even though no change (type I error)

Choice of significance level / distance parameter: tradeoff between number of type I and II errors.

- High \( a \) reduces type I but increases type II (and vice versa)
Control and warning limits (2/3)

**Sampling Frequency**

The more frequent $\kappa$ is calculated, the higher $a$

Average Run Length $\text{ARL} = 1/q$

$\text{ARL}$: expected number of obs between 2 out-of-control alarms

$q$ is the probability of an arbitrary point exceeding the control limits

Average Time to Signal $\text{ATS} = \text{ARL}/\nu$

Sampling frequency defined as $\nu$ observations per time unit

<table>
<thead>
<tr>
<th>Example: Process in control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarterly obs $\text{ATS}=\text{ARL}/\nu=20/4=5$</td>
</tr>
<tr>
<td>Two obs per second $\text{ATS}=\text{ARL}/\nu=20/2=10$</td>
</tr>
</tbody>
</table>

Control and warning limits (3/3)

What is the cost of a type II error? (change not detected)
- Not detected decrease in average milk
- Not detected illness, oestrus

What is the cost of type I error? (false alarm)
- Time spent checking a false alarm
- Risk decrease the reactivity of the farmer to an alarm

Alternative: use of warning limits (fig. 1.5 vs. 2)

Pattern detection

What do we detect?
- Level change, outliers, increase in variation (control limits)
- Trend (increase, decrease), cyclic pattern, autocorrelation

Rules of thumb:
1. One point outside the control limits
2. Two out of three consecutive points outside the warning limits
3. Four out of five consecutive points at a distance of more than $\sigma$ from the expected level
4. Eight consecutive points on the same side of the expected level

Illustration pattern detection

From Example 1

Example 2: daily gain

Precision estimates ($\nu$)
Random sampling: 20.2 g

Target value:
$\theta' = \text{CL} = 775$ g

Control limits:
$\text{UCL} = 775 + \sigma \cdot a = 2$
$\text{LCL} = 775 - \sigma \cdot a = 2$

Example 2: daily gain

Shewart control chart, 2-sigma CL
Example 2: daily gain

Process out of control 7 obs out of 16

Seasonal variation is to be expected in slaughter pig production

If there is an expected pattern: use of other monitoring techniques to take it into account
  e.g. other classical techniques (presented next) or state space models (chapter 8)

If no expected pattern: further analysis / intervention

Moving Average Control Charts (1/2)

The moving average is the average of the most recent $n$ observations

$$M_t(n) = \frac{k_t^n + k_{t-1}^{n-1} + \cdots + k_1}{n}, \quad t \geq n$$

with variance

$$\sigma^2 = \frac{1}{n}$$

The moving average control chart is built the same way as the Shewart control chart

Moving Average Control Charts (2/2)

What can we conclude?

Exponentially Weighted Moving Average control charts (1/3)

The EWMA is a weighted average of all observations until now

$$\hat{z}_t = \lambda K_t + (1 - \lambda) \hat{z}_{t-1}$$

with variance, for large $t$,

$$\sigma_t^2 = \sigma^2 \left( \frac{1}{2 - \lambda} \right)$$

The most recent observations are given highest weights

The EWMA control chart is built the same way as the Shewart control chart

Exponentially Weighted Moving Average control charts (2/3)

First lactation, $a=2, \lambda=0.68$

Choice of lambda:

Small values favor detection of small shifts of $\theta$

Can take time to detect: small lambda = low weight to new obs

Shewart control chart is suggested for detecting large shifts

Combination of EWMA + Shewart for both small and large shifts

Exponentially Weighted Moving Average control charts (3/3)
Monitoring autocorrelated data

Our time series is modeled as

\[ \kappa_t = \theta + \nu_t \]

Assumption: error terms independent

Sample error: often autocorrelated due to repeated measurements on same animal, environmental effects...

Observation error: often independent but depends of measurement method

Daily gain example

Daily gain, slaughter pigs

Dias 25

Correlation

Present versus previous observation
Not obvious

Previous Quarter

Present versus same quarter last year
Seems clear

Milk Yield example - Check for autocorrelation

Milk Yield example - Check for autocorrelation

Dias 26

A model for autocorrelation

First order autoregressive model

\[ u_t = \beta u_{t-1} + \epsilon_t \]

where \( \beta \) is the autoregressive coefficient and \( \epsilon_t \) is an independent random term

When a model is defined, it is used for prediction of the next observation, given the information available at time \( t \):

\[ \hat{\kappa}_{t+1} = \theta + \beta \left( \kappa_t - \theta \right) \]

At time \( t+1 \), the true value is observed and the forecast error calculated,

\[ e_{t+1} = \kappa_{t+1} - \hat{\kappa}_{t+1} \]

Control chart – correlated data

Construct a model describing the correlation

Use the model to predict next observation

Calculate the forecast error

= difference between the observed and predicted value

Calculate the standard deviation of the forecast error

Create a usual control chart for the prediction error
EWMA for autocorrelated data

Use EWMA as one-step-ahead predictor for autocorrelated data

\[ \hat{\kappa}_{t+1} \approx z_t \]

Choose \( \lambda \) by minimizing the sum of the squares of the forecast errors

\[ \epsilon_t = \hat{\kappa}_t - \kappa_{t-1} \]

\[ \sum_{t=1}^{T} \epsilon_t^2 \]

The variance of the forecast errors is calculated as

\[ \sigma^2 = \frac{\sum_{t=1}^{T} \epsilon_t^2}{T} \]

- Raw data (Shewart control chart)
- Averaged data (Moving / Exponentially Moving Average)

We observed seasonality

In the next lecture we will see how to model it using cyclic components

Concluding remarks

We have shifted focus from observing a key figure \( \kappa \) at time \( t \) to an entire time series \( \kappa_1, \kappa_2, \ldots, \kappa_T \).

We tried to detect changes in process (alarms)
- Raw data (Shewart control chart)
- Averaged data (Moving / Exponentially Moving Average)

We observed autocorrelation: model, EWMA

We observed seasonality

In the next lecture we will see how to model it using cyclic components