• CUSUM methods can also be used for attribute/discrete data-based detection. In fact, CUSUM methods are very popular in health-case monitoring and in public health surveillance.

• In principle, EWMA can be used for monitoring attribute data as well. But in actuality, especially in health care applications, EWMA is far less popular than CUSUM. So we will only discuss the application of CUSUM here.

• CUSUM charts for healthcare application are largely used in a one-sided fashion, in order to detect an improvement in a medical process or an increase in disease cases or a decrease in mortality.
Example 2.8: Suppose we mean to monitor the respiratory diseases in the Houston area, which is characterized by the presence of a large number of petro-chemical factories. The observed quantity is the monthly death rate due to respiratory diseases in male subjects during the period of 1980-89.

Some terminology:
- At-risk population: denote by \( n \)
- Baseline death rate: \( \lambda_a \)
- Expected number of deaths: \( m_a = n\lambda_a \)
- Observed death cases: \( x_i \), here it is reasonable to assume that \( x_i \) follows a Poisson distribution with parameter \( m_a \) in the absence of any change in case rate.

Given the observations of \( x_i \), people would like to know if there is any significant change (out-of-control) in the disease-induced death (if so, the public should be altered).
CUSUM for discrete data

- A standard CUSUM for discrete data is

\[ S_i = \max \{0, S_{i-1} + (x_i - K)\} \]

The sequence is considered out-of-control whenever \( S_i \geq a \) prescribed H.

- This formula is slightly different from the CUSUM statistic we introduced earlier because there is \( \mu_0 \) here. This should be fine because the value of H will take account of the absence of \( \mu_0 \) (note that \( \mu_0 \) is \( m_a \) in this example).

- How to determine K?

  If the in-control or historical expected number of deaths is \( m_a \), and the increased number at an out-of-control level is \( m_r = n\lambda_r \). Similar to the previous cases, \( K = \frac{(m_a + m_r)}{2} \).
CUSUM for discrete data

- How to determine $H$: using the Monte Carlo simulation

1. Using a given $m_a$ and Poisson distribution, simulate a sequence of data, and this is the in-control data;
2. Repeat (1) so that you have $N$ trials of the in-control data;
3. Choose an $H$ so that the $ARL_0$ is large enough (typically $ARL_0 = 500$).
4. Choose a $m_r > m_a$ (for example, $m_r = m_a + 1$), and simulate $N$ trials of the out-of-control data. Apply the CUSUM (with the chosen $H$) to evaluate the $ARL_1$ for this specific $m_r$.
5. Repeat (4) for other $m_r$’s and tabulate all ARL’s to see if they meet your requirement. If not, you need to adjust the value of $K$ and restart from (1).
CUSUM for discrete data

- How to determine $m_a$?

  Use the historical data, for example, the monthly death number in the same area for male subjects but in the decade before ('70-'79).

- It is typically difficult to estimate the baseline rate $m_a$ in the healthcare applications because many factors are changing over time. For example, the at-risk population could be changing; the age structure in a population may vary. So it may not be always valid to use the entire set of the historical data over a long time history to estimate the baseline. A few alternatives are:

  (a) Use the most recent week/month data to update the baseline rate constantly;

  (b) Let the monitoring process start anew whenever the CUSUM statistic becomes zero and re-estimate the baseline rate;

  (c) Perform a risk-adjustment (which will be presented in the sequel) that can more explicitly take care of the variability in the underlying risk.
CUSUM for discrete data

- Transformation used in the CUSUM method for attribute data.
- Because the attribute data are not normally distributed, the fundamental theory behind CUSUM method is not perfectly satisfied. So people suggested a number of transformations that will convert the original variable into a variable closer to being normally distributed.

- Assume $x \sim \text{Poisson}(m)$
  
  \[(a) \quad z_1 = \frac{x - m}{\sqrt{m}} \sim N(0, 1)\]
  because for $x \sim \text{Poisson}(m)$, $\mu_x = m$ and $\sigma_x = \sqrt{m}$

  \[(b) \quad z_2 = 2(\sqrt{x} - \sqrt{m}) \sim N(0, 1)\]
  - The square-root transformation of a Poisson variable is both normalizing and variance-stabilizing.
  - $E(\sqrt{x}) = \sqrt{m}$ and $\text{var}(\sqrt{x}) = \frac{1}{4}$ or $\text{stdev}(\sqrt{x}) = \frac{1}{2}$. 
CUSUM for discrete data

• Assume \( x \sim \text{Poisson}(m) \)

\[
 z_3 = \frac{x - 3m + 2\sqrt{xm}}{2\sqrt{m}}
\]

(c) Rossi (1999) suggested a half-sum of (a) and (b) as

The thought is that if the original variables \((z_1 \text{ and } z_2)\) are standard normal and independent, \(z_3\) should be standard normal as well. But because \(z_1\) and \(z_2\) are correlated, treating \(z_3\) as a standard normal variable includes further approximation.

• When use these transformations, the CUSUM should be

\[
 S_i = \max\{0, S_{i-1} + (z_i - K)\}
\]

where \(z_i\) is the value of either \(z_1\), \(z_2\), or \(z_3\) corresponding to \(x_i\). And \(K\) and \(H\) should be decided the same way as before.

It was claimed that using these transformation can improve the performance of the CUSUM detection method.
• (Example 2.8) The baseline monthly death rate is estimated from the data of 1970-1979 = 5.4. After a slight adjustment on the rate, people used $m_a = 5.33$ for 1980-89. The change of rate to be detected is about $m_r = 8.5$.

• When using a CUSUM without using the transformation, $H = 9$ and $K = 7$, simulations indicate that $ARL_0 = 500$ and $ARL_1 = 7$. When using the transformation $z_3$, $H = 3.80$ and $K = 0.60$ to achieve the same ARL performances.

- It turns out that the alarms on the left chart are false.
**Example 2.9:** Both pharmaceutical industry and regulatory agencies are aware of the importance of post-marketing surveillance of therapeutic drugs. So spontaneous reports of adverse drug reactions (ADR) are systematically collected by pharmaceutical companies and reported to regulatory authorities. One tries to detect any changes in the ADR rates in the course of time as soon as possible.

![Graph showing rates of injection site and serious ADRs to the DT vaccine, West Germany, 1986–1989](image-url)
CUSUM for discrete data: Example 2.9

• The figure shows the local injection sites and the serious ADR rates for a combined diphtheria/tetanus vaccine frequently used for children from 1986 to 1989 in West Germany.

• One liter of vaccine corresponds to 2,000 single doses, and the number of doses delivered is obtained from the factory sales record of the vaccine. The ADR rate is calculated as

\[
\text{rate} = \frac{\text{the number of ADRs}}{1 \text{ liter}} = \frac{\text{the number of ADRs}}{2,000 \text{ doses}}
\]

• The previous figure showed that the injection site reactions, not the serious ADRs, have a high percentage rate, especially in the second and fourth quarter of 1987. The question is how much increases can be assessed?
CUSUM for discrete data: Example 2.9

- Apply the CUSUM method to see what we detect.
- A few considerations:
  - In this example, we have a varying at-risk population (this is can be measured by the vaccine sales volume). So we will use the sales volume to do some adjustment.
  - The baseline incidence level, labeled as $k_0$ of the reported ADRs per unit of sales volume (called $\mu_0$ or $m_a$ in previous examples), and its standard deviation, are estimated from 1986's data.
  - The CUSUM score is

$$S_i = \max\{0, S_{i-1} + \frac{x_i}{c_i} - k_r\} \text{ for } i > 0 \text{ and } S_0 = 0$$

where
- $x_i = \text{number of ADRs of period } i$
- $c_i = \text{sales volume of period } i$
- $k_r = \text{offset threshold}$
CUSUM for discrete data: Example 2.9

- If $S_i$ exceeds a decision boundary $H$, an increased level in ADR is suspected and an alert is sent out.

- Choice of $k_r$.

  Let $k_1 > k_0$ be an increased incidence level which is considered to be the alert level. Again $k_r$ is typically take as $k_r = \frac{k_0 + k_1}{2}$.

- In this example:

  we choose $k_1 = 1.5k_0$ so $k_r = 1.25k_0$. $H$ is chosen so that $ARL_0 = 500$. Since the sampling interval is a quarter, $ARL_0 = 500$ implies, on average, one false alarm every $500/4 = 125$ years.
### The result

**Table III. Cusum scores for injection site and serious ADRs DT vaccine, 1986–1989**

<table>
<thead>
<tr>
<th>Baseline periods* (I/86–IV/86)</th>
<th>Injection site ADRs</th>
<th>Serious ADRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0 \pm SD$ (ADRs/litre)$†$</td>
<td>0.0574 ± 0.0163</td>
<td>0.0200 ± 0.0092</td>
</tr>
<tr>
<td>$k_r$ (ADRs/litre)</td>
<td>0.0716</td>
<td>0.0250</td>
</tr>
<tr>
<td>$k_1$ (=1.5$k_0$)</td>
<td>0.0861</td>
<td>0.0300</td>
</tr>
<tr>
<td>ARL$_0$</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>ARL$_1$</td>
<td>3.8</td>
<td>8.1</td>
</tr>
<tr>
<td>$h$ (detection boundary)</td>
<td>0.0435</td>
<td>0.0382</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observation periods*</th>
<th>Injection site ADRs</th>
<th>Serious ADRs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x_i/c_i$</td>
<td>$S_i$</td>
</tr>
<tr>
<td>I/87</td>
<td>0.0058</td>
<td>0</td>
</tr>
<tr>
<td>II/87</td>
<td>0.0954</td>
<td>0.0238</td>
</tr>
<tr>
<td>III/87</td>
<td>0.0731</td>
<td>0.0252</td>
</tr>
<tr>
<td>IV/87</td>
<td>0.1667</td>
<td>0.1203‡</td>
</tr>
<tr>
<td>I/88</td>
<td>0.0916</td>
<td>0.0200</td>
</tr>
<tr>
<td>II/88</td>
<td>0.0995</td>
<td>0.0479‡</td>
</tr>
<tr>
<td>III/88</td>
<td>0.1133</td>
<td>0.0417</td>
</tr>
<tr>
<td>IV/88</td>
<td>0.0329</td>
<td>0.0030</td>
</tr>
<tr>
<td>I/89</td>
<td>0.0418</td>
<td>0</td>
</tr>
<tr>
<td>II/89</td>
<td>0.0781</td>
<td>0.0065</td>
</tr>
<tr>
<td>III/89</td>
<td>0.1047</td>
<td>0.0395</td>
</tr>
<tr>
<td>IV/89</td>
<td>0.0809</td>
<td>0.0489‡</td>
</tr>
</tbody>
</table>

* Periods given in quarters (I–IV) and years.
† 1 litre of DT vaccine corresponds to 2000 vaccinations.
‡ Detection boundary exceeded, score reset to 0.
CUSUM for discrete data: Example 2.9

- The timeliness of reaction

<table>
<thead>
<tr>
<th></th>
<th>Days of delay between Occurrence and ADR report (N = 209)*</th>
<th>Exposure and ADR report (N = 238)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lower quartile</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Median</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Upper quartile</td>
<td>66</td>
<td>86</td>
</tr>
<tr>
<td>Maximum</td>
<td>2766</td>
<td>8285</td>
</tr>
</tbody>
</table>

* N = number of ADR reports with the respective information.

- On average, a ADR delay is about 25 to 30 days. Many times, people need to look back to the previous quarter(s) for the vaccinations causing the ADRs.
• When the characteristic under consideration is driven by complicated input or varying parameters, one will need to adjust the response before use a control chart for monitoring and detection. This adjustment procedure is called "risk adjustment."

• This is particularly true in healthcare applications, where health-related or patient-dependent variables are involved in the detection process. It may not be reasonable, for example, to assume that cardiac-surgery patients of widely differing ages and health conditions have the same probability of short-term survival.

• A good example is in the trial of Dr. Harold Shipman of Britain, who was eventually convicted in 2000 of murdering 15 of his patients and implicated in the killing of between 200-300 patients.
Risk adjustment for detection

• The raw data of Dr. Shipman's patient mortality clearly have an upward trend.

• But, at the trail, Dr. Shipman claimed that the increase in the mortality is due to that the recently treated patients are in the high-risk group (age, pre-conditions, etc.).
Risk adjustment for detection

• Risk adjusted mortality trend of Dr. Shipman's patients.

Statisticians did a risk adjustment to the raw data and found that there is a significant increase in the female case that cannot be explained by patient's own risk.
Risk adjustment for detection

• Typically, the way to conduct a risk adjustment is to establish a model connecting the raw observation \( x \) to the adjusted response \( y \).

• There are two (popular) models used:
  (1) regression model, for continuous data
  (2) logistic regression model, for attribute/discrete data.
• Assume that you are familiar with linear regression.
• Suppose that we have the following relationship:

\[
\begin{align*}
\text{x} & \quad \text{process} \quad \text{y} \\
\text{incoming characteristic} & \quad \text{outgoing characteristic}
\end{align*}
\]

When y is observed to have out-of-control observations, it may be caused by

1) the process
2) an out-of-control x
3) both

• If our intention is to detect change in either the process or in x, monitoring y does not appear to be an efficient way.
Regression-based risk adjustment

- Regression model:

\[
    y_k = \beta_0 + \beta_1 x_k + \epsilon_k,
\]

where \( k \) is a time variable; \( \epsilon_k \) is the error introduced by the process and is assumed to \( \sim N(0, \sigma^2) \); \( \beta_0, \beta_1 \) are unknown parameters, called regression coefficients.

Here for simplicity, we assume that there is only one incoming variable \( x \). The model could be expanded to have multiple input variables such as

\[
    y_k = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} \ldots + \beta_p x_{pk} + \epsilon_k,
\]

The data pair \( \{x_k, y_k\} \) are supposed to be available.
Regression-based risk adjustment

- Regression-based detection:

  Instead of monitoring \( y \), we will monitor \( z \), the residual from the regression model,

  \[
  z_k = \hat{\epsilon}_k = y_k - \beta_0 - \beta_1 x_k
  \]

  Under \( H_0 \), i.e., no change in the process,
  
  \[
  z_k = \epsilon_k \sim N(0, \sigma^2).
  \]

  Otherwise, \( z_k \) will depart from this normal distribution.
Regression-based risk adjustment

- Estimate $\beta_0$ and $\beta_1$ from historical data

Suppose that we have the historical dataset $\{x_i, y_i\}_{i=1}^{m}$. Then,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \text{ for } i = 1, \ldots, m$$

$$\Rightarrow \begin{cases} y_1 = \beta_0 + \beta_1 x_1 + \epsilon_1 \\ \vdots \\ y_m = \beta_0 + \beta_1 x_m + \epsilon_m \end{cases}$$

$$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix}$$

Expressed in matrix/vector form $\Rightarrow \mathbf{Y} = \mathbf{X} \cdot \mathbf{\beta} + \mathbf{\epsilon}$. 
Regression-based risk adjustment

- Least-squares estimation (LSE) of $\beta$

$$\min_{\hat{\beta}} \{(Y - X\hat{\beta})^T (Y - X\hat{\beta})\} \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

Once $\hat{\beta}$ is estimated, calculate $z_k = y_k - \hat{y}_k = y_k - \hat{\beta}_0 - \hat{\beta}_1 x_k$, $k = 1, \ldots, m$

- Control limits for $z$

Suppose it is an individual chart. Then use the moving range (MR) statistics to estimate $\hat{\sigma}$, i.e.,

$$\bar{MR} = \frac{1}{m-1} \sum_{k=2}^m |MR_k| = \frac{1}{m-1} \sum_{k=2}^m |z_k - z_{k-1}|,$$

and then

$$\bar{z} = \frac{1}{m} \sum_{k=1}^m z_k, \quad \hat{\sigma} = \frac{\bar{MR}}{d_2}, \text{ where } d_2 = 1.128 \text{ is corresponding to } n = 2.$$

If choose $L = 3$, then

$$UCL/LCL = \bar{z} \pm 3 \cdot \hat{\sigma} = \bar{z} \pm \frac{3}{1.128} \bar{MR} = \bar{z} \pm 2.66\bar{MR}$$

$$CL = \bar{z}$$
Regression-based risk adjustment

- Decision rule: use $x$ and $z$ charts

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$x$-chart</th>
<th>$z$-chart</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>signal</td>
<td>signal</td>
<td>both $x$ and the process are out of control</td>
</tr>
<tr>
<td>2</td>
<td>signal</td>
<td>no signal</td>
<td>$x$ (input) out of control</td>
</tr>
<tr>
<td>3</td>
<td>no signal</td>
<td>signal</td>
<td>process out of control</td>
</tr>
<tr>
<td>4</td>
<td>no signal</td>
<td>no signal</td>
<td>both in control</td>
</tr>
</tbody>
</table>

It is more transparent to use $x$ and $z$ charts for detection. The $y$-chart does not contribute much to the decision making process in addition to the $x$- and $z$- charts.
Example 2.10: consider a component part for the braking system for an automobile. The variables measured are $x = \text{ROLLWT}$ as the input variable and $y = \text{BAKEWT}$ as the response variable.

Regression model (using a least-squares estimation):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$$

Data: A total of 70 data pairs ($x$ and $y$). The first 45 of those were used to establish the relationship between the two variables and to calculate the control limits.
Regression-based risk adjustment: Example 2.10

- Data table and the fitted line

<table>
<thead>
<tr>
<th>ID</th>
<th>RWT</th>
<th>BWT</th>
<th>ID</th>
<th>RWT</th>
<th>BWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>210</td>
<td>200</td>
<td>36</td>
<td>209</td>
<td>199</td>
</tr>
<tr>
<td>2</td>
<td>211</td>
<td>200</td>
<td>37</td>
<td>208</td>
<td>201</td>
</tr>
<tr>
<td>3</td>
<td>208</td>
<td>199</td>
<td>38</td>
<td>210</td>
<td>202</td>
</tr>
<tr>
<td>4</td>
<td>208</td>
<td>200</td>
<td>39</td>
<td>210</td>
<td>200</td>
</tr>
<tr>
<td>5</td>
<td>209</td>
<td>203</td>
<td>40</td>
<td>212</td>
<td>200</td>
</tr>
<tr>
<td>6</td>
<td>210</td>
<td>203</td>
<td>41</td>
<td>209</td>
<td>201</td>
</tr>
<tr>
<td>7</td>
<td>211</td>
<td>202</td>
<td>42</td>
<td>212</td>
<td>203</td>
</tr>
<tr>
<td>8</td>
<td>211</td>
<td>201</td>
<td>43</td>
<td>211</td>
<td>201</td>
</tr>
<tr>
<td>9</td>
<td>210</td>
<td>201</td>
<td>44</td>
<td>212</td>
<td>204</td>
</tr>
<tr>
<td>10</td>
<td>213</td>
<td>203</td>
<td>45</td>
<td>209</td>
<td>200</td>
</tr>
<tr>
<td>11</td>
<td>210</td>
<td>200</td>
<td>46</td>
<td>209</td>
<td>201</td>
</tr>
<tr>
<td>12</td>
<td>211</td>
<td>203</td>
<td>47</td>
<td>206</td>
<td>200</td>
</tr>
<tr>
<td>13</td>
<td>210</td>
<td>201</td>
<td>48</td>
<td>210</td>
<td>200</td>
</tr>
<tr>
<td>14</td>
<td>210</td>
<td>201</td>
<td>49</td>
<td>208</td>
<td>199</td>
</tr>
<tr>
<td>15</td>
<td>209</td>
<td>202</td>
<td>50</td>
<td>208</td>
<td>198</td>
</tr>
<tr>
<td>16</td>
<td>211</td>
<td>202</td>
<td>51</td>
<td>208</td>
<td>200</td>
</tr>
<tr>
<td>17</td>
<td>211</td>
<td>201</td>
<td>52</td>
<td>206</td>
<td>197</td>
</tr>
<tr>
<td>18</td>
<td>211</td>
<td>202</td>
<td>53</td>
<td>208</td>
<td>199</td>
</tr>
<tr>
<td>19</td>
<td>212</td>
<td>202</td>
<td>54</td>
<td>211</td>
<td>201</td>
</tr>
<tr>
<td>20</td>
<td>208</td>
<td>200</td>
<td>55</td>
<td>214</td>
<td>204</td>
</tr>
<tr>
<td>21</td>
<td>212</td>
<td>202</td>
<td>56</td>
<td>212</td>
<td>203</td>
</tr>
<tr>
<td>22</td>
<td>209</td>
<td>201</td>
<td>57</td>
<td>209</td>
<td>200</td>
</tr>
<tr>
<td>23</td>
<td>210</td>
<td>202</td>
<td>58</td>
<td>209</td>
<td>204</td>
</tr>
<tr>
<td>24</td>
<td>210</td>
<td>201</td>
<td>59</td>
<td>206</td>
<td>201</td>
</tr>
<tr>
<td>25</td>
<td>211</td>
<td>201</td>
<td>60</td>
<td>214</td>
<td>202</td>
</tr>
<tr>
<td>26</td>
<td>210</td>
<td>200</td>
<td>61</td>
<td>211</td>
<td>202</td>
</tr>
<tr>
<td>27</td>
<td>210</td>
<td>200</td>
<td>62</td>
<td>212</td>
<td>202</td>
</tr>
<tr>
<td>28</td>
<td>210</td>
<td>200</td>
<td>63</td>
<td>210</td>
<td>203</td>
</tr>
<tr>
<td>29</td>
<td>211</td>
<td>201</td>
<td>64</td>
<td>214</td>
<td>201</td>
</tr>
<tr>
<td>30</td>
<td>211</td>
<td>202</td>
<td>65</td>
<td>209</td>
<td>200</td>
</tr>
<tr>
<td>31</td>
<td>211</td>
<td>201</td>
<td>66</td>
<td>210</td>
<td>201</td>
</tr>
<tr>
<td>32</td>
<td>210</td>
<td>201</td>
<td>67</td>
<td>212</td>
<td>201</td>
</tr>
<tr>
<td>33</td>
<td>209</td>
<td>201</td>
<td>68</td>
<td>214</td>
<td>206</td>
</tr>
<tr>
<td>34</td>
<td>212</td>
<td>203</td>
<td>69</td>
<td>214</td>
<td>204</td>
</tr>
<tr>
<td>35</td>
<td>209</td>
<td>200</td>
<td>70</td>
<td>212</td>
<td>203</td>
</tr>
</tbody>
</table>

\[ \hat{y} = 104.325 + 0.4604 \cdot x \]
Regression-based risk adjustment: Example 2.10

- Calculate $z_k$ for the last 25 samples

<table>
<thead>
<tr>
<th>ID</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>0.440</td>
</tr>
<tr>
<td>47</td>
<td>0.821</td>
</tr>
<tr>
<td>48</td>
<td>-1.021</td>
</tr>
<tr>
<td>49</td>
<td>-1.100</td>
</tr>
<tr>
<td>50</td>
<td>-2.100</td>
</tr>
<tr>
<td>51</td>
<td>-0.100</td>
</tr>
<tr>
<td>52</td>
<td>-2.179</td>
</tr>
<tr>
<td>53</td>
<td>-1.100</td>
</tr>
<tr>
<td>54</td>
<td>-0.481</td>
</tr>
<tr>
<td>55</td>
<td>1.137</td>
</tr>
<tr>
<td>56</td>
<td>1.058</td>
</tr>
<tr>
<td>57</td>
<td>-0.560</td>
</tr>
<tr>
<td>58</td>
<td>3.440</td>
</tr>
<tr>
<td>59</td>
<td>1.821</td>
</tr>
<tr>
<td>60</td>
<td>-0.863</td>
</tr>
<tr>
<td>61</td>
<td>0.519</td>
</tr>
<tr>
<td>62</td>
<td>0.058</td>
</tr>
<tr>
<td>63</td>
<td>1.979</td>
</tr>
<tr>
<td>64</td>
<td>-1.863</td>
</tr>
<tr>
<td>65</td>
<td>-0.560</td>
</tr>
<tr>
<td>66</td>
<td>-0.021</td>
</tr>
<tr>
<td>67</td>
<td>-0.942</td>
</tr>
<tr>
<td>68</td>
<td>3.137</td>
</tr>
<tr>
<td>69</td>
<td>1.137</td>
</tr>
<tr>
<td>70</td>
<td>1.058</td>
</tr>
</tbody>
</table>

- Control limits
  
  $\text{UCL/LCL} = \pm 2.42$ and $\text{CL} = 0$ because the sample average of $z$ is (nearly) zero.
Regression-based risk adjustment: Example 2.10

- Three control charts

- Comparison of out-of-control among y, x, and z charts
  - y-chart $\rightarrow \{50, 52, 68\}$
  - z-chart $\rightarrow \{58, 68\}$
  - x-chart $\rightarrow \{47, 52, 55, 59, 60, 64, 68, 69\}$
Risk adjustment for attribute data: logistic regression

- Logistic regression is the method often used for handling discrete data.
- Suppose Y can only take two distinct values 0 or 1, a binary variable, then

\[ E(Y|X=x) = \sum_{k=0}^{1} k \cdot \Pr[Y = k|X = x] = \Pr[Y = 1|X = x] \]

so the interpretation of \( E(Y|X=x) \) is that it is the conditional probability of that \( Y=1 \).

- Denote by \( E(Y|X=x) = \Pr[Y=1 \mid X=x] \) by \( \pi(x) \). In the meanwhile,

\[ \text{Var}(Y|X=x) = \pi(x)[1-\pi(x)], \quad \text{(why this?)} \]
Risk adjustment for attribute data: logistic regression

- Suppose that we assume a regular linear regression relation between $Y$ and $X$, so
  \[ Y = \beta_0 + \beta_1 X + \epsilon \]
  Then, we have
  \[ E(Y|X = x) = \beta_0 + \beta_1 x \Rightarrow \pi(x) = \beta_0 + \beta_1 x \]
  since $E(Y|X = x) = \pi(x)$.

- Here is a problem:
  $\pi(x)$ is a probability, taking values only from $[0, 1]$, while $\beta_0 + \beta_1 x$ could return any value.

  $\Rightarrow$ regular linear regression does not model a discrete $Y$ well (a binary $Y$ is a special form of discrete responses).
Risk adjustment for attribute data: logistic regression

- To address the previous modeling difficulty, people suggest using a logarithm transformation

\[ 0 < \pi(x) < 1 \Rightarrow 0 < \frac{\pi(x)}{1-\pi(x)} < \infty \Rightarrow -\infty < \log \frac{\pi(x)}{1-\pi(x)} < \infty \]

\[ \Rightarrow \log \frac{\pi(x)}{1-\pi(x)} \text{ can take any value from } -\infty \text{ to } +\infty \]

so it is reasonable to let \( \log \frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x \).

This model is called "logistic regression" model, where \( \log \frac{\pi(x)}{1-\pi(x)} \) is the so-called "logit" function of \( \pi(x) \), namely

\[ \text{logit}(\pi(x)) = \log \frac{\pi(x)}{1-\pi(x)}. \]
Remarks on logistic regression:

(a) In the above derivation, \( x \) is a scalar. But \( x \) can be a vector, representing a number of explanatory factors.

(b) Logistic regression can be used to model a discrete response \( Y \) taking more than two values. For a binary \( Y \), this is a two-class problem, while, for a \( Y \) taking more than two values, it becomes a multi-class problem. For a multi-class problem, the logistic formulation needs some revision but we will focus on the two-class problem here.

(c) Linear regression and logistic regression are typical predictive models and thus are the subjects covered with more details by ISEN 619, Analysis and Prediction.
Risk adjustment for attribute data: logistic regression

- How to estimate the coefficient in a logistic regression: use a maximum likelihood estimation (MLE) method.

\[ \mathbf{x}_i = (1 \ x_{i1} \ ... \ x_{ip})^T, \] a $p \times 1$ input vector with $p$ explanatory variables; $\mathbf{x}_i$ is the $i^{th}$ observation of the input, $i = 1, ..., N$;

\[ y_i = 1 \text{ or } 0 \text{ is the } i^{th} \text{ observation of the binary response}; \]

\[ \mathbf{\beta} = (\beta_0 \ \beta_1 \ ... \ \beta_p)^T \] is the coefficient to be estimated.

- For a two-class problem, the log-likelihood function is (skip the details).

\[ \ell(\mathbf{\beta}) = \sum_{i=1}^{N} \{ y_i \cdot \mathbf{\beta}^T \mathbf{x}_i - \log[1 + \exp(\mathbf{\beta}^T \mathbf{x}_i)] \}. \]

- The MLE method is to find the $\mathbf{\beta}$ that maximizes the above log-likelihood function, given the training dataset \{\mathbf{x}_i, y_i\}, $i = 1, ..., N$. 


Risk adjustment for attribute data: logistic regression

- How to maximize the log-likelihood function: need helps from some numerical solver.

- Can use the \texttt{fminunc} function in MATLAB.

Syntax: \texttt{BETA = fminunc('log-likelihood function', BETA0)}

Remark:
- \texttt{fminunc} solves a minimization problem. You need to multiply a (-1) to the returning value of your log-likelihood function so that the \texttt{fminunc} can solve a maximization problem.

- You can declare \texttt{X} and \texttt{Y} (the data matrix/vector) as global variables to transfer values between functions.

- \texttt{BETA0} is the initially value. You can use \([0 \ 0 \ 0]^T\) if you have no better choices.
**Example 2.11**: an example of monitoring the mortality rate in a cardiac surgery.

The measurement data includes a patient's risk characteristic $x$ and the surgical outcome $y$.

\[ y_t = 1 \text{ if patient } t \text{ dies} \]
\[ y_t = 0 \text{ if patient } t \text{ survives} \]

Risk of a surgery consists of two factors: (1) that depends on a patient's own risk; (2) that is due to the surgical procedure (including surgeon's skill). Here we assumed these two factors are independent.

Usually the objective of a hospital or a medical team is to detect any significant change in the risk associated with the surgical procedure, and if it happens, take action to lower it.

But if we directly monitor $y_t$, it would be difficult to tell which factor has caused it when a change in $y_t$ is observed.
Risk adjustment for attribute data: Example 2.11

• Surgery's risk (odds ratio)

Denote by $R$ the odds ratio related to a surgical procedure. Odds ratio is a probability ratio of one outcome (say, a patient dies) versus the alternative (the patient survives).

$$R = \frac{\Pr[Y_{t} = 1 \mid \text{a given surgical procedure}]}{\Pr[Y_{t} = 0 \mid \text{a given surgical procedure}]}$$

• Patient's risk

Patient's risk may be characterized by $x_{t}$, a vector of factors affecting the chance of a patient's survival after a surgery.

Denote by $p_{t}$ the patient's risk, namely the probability of a patient dying after a surgery, given his/her risk vector $x_{t}$,

$$p_{t} = \Pr[Y_{t} = 1 \mid x_{t}]$$
A logistic regression model for patient's risk

- Available mortality data are from 1992-1998. Use the first two years data to establish a logistic regression model. There are a total of 2,218 surgeries and 143 deaths observed (the overall mortality rate is about 6.5%)

- In this study, the patient's risk vector degenerates to a scalar $x_t$, the Parsonnet score. A Parsonnet score is a medical risk parameter based on a combination of explanatory variates thought to be important in cardiac surgery.

- Use the historical data, the logistic model is obtained as

$$\text{logit}(p_t) = -3.68 + 0.77x_t$$

From this model, a low risk patient (say, $x_t = 0$) has a prior risk of death of just 2.5% following surgery, while a patient with the highest risk ($x_t = 71$) has a prior risk of 86% mortality.
Assuming that the two risks are independent, the combined odds for a patient to die after an operation is

\[
\frac{\Pr[Y_t=1|\text{procedure, } x_t]}{\Pr[Y_t=0|\text{procedure, } x_t]} = \frac{\Pr[Y_t=1|\text{procedure}]}{\Pr[Y_t=0|\text{procedure}]} \cdot \frac{\Pr[Y_t=1|x_t]}{\Pr[Y_t=0|x_t]} = \frac{R \cdot p}{1-p}
\]

Recall that people intends to detect whether there is an increase in the risk of surgical procedures. So

\[H_0: \text{odds ratio} = R_0\]
\[H_1: \text{odds ratio} = R_1\]

and \(R_1 > R_0\) and oftentimes, \(R_0\) is set as 1.

Researcher [Steiner et al. 2000; Moustakides 1986] showed that an optimal detection scheme is to apply the CUSUM method to a log-likelihood ratio of \(Y_t = 1\) versus \(Y_t = 0\).
Risk adjustment for attribute data: Example 2.11

- The log-likelihood ratio is defined as follows.

  Given the risk of a patient, let \( f(y_t|H_0, x_t) \) and \( f(y_t|H_1, x_t) \) be the pdf’s of possible outcomes under \( H_0 \) and \( H_1 \), respectively. The log-likelihood ratio is

  \[
  W_t = \log \frac{f(y_t|H_1, x_t)}{f(y_t|H_0, x_t)}
  \]

- Use this log-likelihood ratio score in a CUSUM procedure:

  \[
  Z_t = \max\{0, Z_{t-1} + W_t\}, \quad t = 1, 2, \ldots, \text{ and } Z_0 = 0,
  \]

  When \( Z_t \geq h \), then it indicates a significant increase in the procedure’s risk.

  To detect a decrease, one can use \( Z_t = \min\{0, Z_{t-1} + W_t\} \), \( Z_0 = 0 \)
Risk adjustment for attribute data: Example 2.11

• How to calculate $W_t$ after observing the outcome of a given patient after surgery?

Given that $y_t$ is a binary variable, the sequence of $y_t$ follows a Binomial distribution, namely

$$
f(y_t|H_0, x_t) = (\Pr[y_t = 1|H_0, x_t])^{y_t}(1 - \Pr[y_t = 1|H_0, x_t])^{1-y_t}
$$

$$
f(y_t|H_1, x_t) = (\Pr[y_t = 1|H_1, x_t])^{y_t}(1 - \Pr[y_t = 1|H_1, x_t])^{1-y_t}
$$

Recall that $\frac{\Pr[Y_t=1|\text{procedure}, x_t]}{\Pr[Y_t=0|\text{procedure}, x_t]} = \frac{R \cdot p_t}{1-p_t}$, and

$\Pr[Y_t = 1|\text{procedure}, x_t] = 1 - \Pr[Y_t = 0|\text{procedure}, x_t]$, so we have $\Pr[Y_t = 1|\text{procedure}, x_t] = \frac{R \rho_t}{1-\rho_t + R \rho_t}$.

Under $H_0$: $\Pr[y_t = 1|H_0, x_t] = \frac{R_0 \rho_t}{1-\rho_t + R_0 \rho_t}$ and

$\Pr[y_t = 0|H_0, x_t] = \frac{1-\rho_t}{1-\rho_t + R_0 \rho_t}$

Under $H_1$: $\Pr[y_t = 1|H_1, x_t] = \frac{R_1 \rho_t}{1-\rho_t + R_1 \rho_t}$ and

$\Pr[y_t = 0|H_1, x_t] = \frac{1-\rho_t}{1-\rho_t + R_1 \rho_t}$
Two possible log-likelihood ratio score for patient $t$ are:

$$W_t = \begin{cases} \log\left[\frac{(1-p_t+R_0p_t)R_1}{(1-p_t+R_1p_t)R_0}\right] & \text{if } y_t = 1 \\ \log\left[\frac{1-p_t+R_0p_t}{1-p_t+R_1p_t}\right] & \text{if } y_t = 0 \end{cases}$$

So the procedure to compute the $W_t$ is:
- Given a patient, knowing his/her $x_t$, calculate $p_t$
- Set $R_0$ and $R_1$. Typically set $R_0=1$. $R_1$ is the level of risk increase (or decrease) one attempts to detect.
- Upon observing $y_t$ after a surgery, compute $W_t$ using the corresponding formula.
Risk adjustment for attribute data: Example 2.11

- Control limit $h$, to be determined via simulation
  
  For instance, we try to detect a doubling of the odds of death. then choose $R_0 = 1$ and $R_1 = 2$.

  1. Use the historical dataset to calculate the prior risk factor $p_t$ from $x_t$ of patient $t$;
  2. Use $R_0 = 1$ and $p_t$ to calculate $\Pr[y_t = 1|H_0, x_t]$ and $\Pr[y_t = 0|H_0, x_t]$;
  3. Use the probabilities in (2) to simulate a sequence of Bernoulli trial $\{y_t\}_{t=1}^m$ for a large enough $m$;
  4. Apply the CUSUM method to the simulated $\{y_t\}_{t=1}^m$ to get a run length;
  5. Repeat (1) - (4) $N$ times to get $ARL_0$. Adjust $h$ so that $ARL_0$ is large enough (for example, $\geq 500$). In practice, the desirable level of $ARL_0$ can be decided based on the yearly number of operations and the tolerable false alarms every a number of years.
  6. Change $R_0$ to $R_1$, and repeat (1)-(4) to get $ARL_1$. 

ISEN 614 Advanced Quality Control (Anomaly and Change Detection)  Dr. Yu Ding
• In this example, $\text{ARL}_0$ set as 500, $h = 4.5$ for detecting an increase and $h = -4$ for detecting a decrease.

Trainee surgeons CUSUM; unadjusted CUSUMs on top, risk-adjusted CUSUMs on the bottom.
Risk adjustment for attribute data: Example 2.11

Experienced surgeons have a high mortality rate in the unadjusted chart primarily because of the patient's own risks.

Experienced surgeon CUSUM; unadjusted CUSUMs on the top, risk-adjusted CUSUMs on the bottom.