Control Charts for Attributes

Terminology

“Attributes” refers to qualitative variables like conforming/nonconforming, as opposed to “Variables”, which refers to quantitative variables like physical dimensions.

Variables that are small counts are quantitative, but are often handled better as attribute data.

Control charts for attributes and for variables are constructed differently, but with the same general goal of monitoring performance.
Examples

- A glass container for a liquid product is classified as conforming or nonconforming, based on the absence or presence of flaws.
- The number of defects on a silicon wafer is counted.

Three different charts are used for data like these: the $p$ chart, the $c$ chart, and the $u$ chart.
Control Chart for Fraction Nonconforming

Statistical framework
Suppose that items are produced, or services are provided, in a sequence, and that each is classified as conforming or nonconforming.

Suppose that when the process is in control, successive items independently have probability $p$ of being nonconforming.

Then if a sample of $n$ items has $D$ nonconforming items, $D$ has the binomial distribution with parameters $n$ and $p$:

$$D \sim \text{Bin}(n, p).$$
We estimate $p$ by

$$\hat{p} = \frac{D}{n},$$

and from the properties of $\text{Bin}(n, p)$ we know that

$$E(\hat{p}) = p$$

and

$$\text{Var}(\hat{p}) = \frac{p(1 - p)}{n}.$$ 

We use these properties to construct a control chart called the $p$ chart.
Development of the $p$ chart

If $p$ is known, plot each $\hat{p}$ on a chart with a center line at $p$ and three-sigma control limits:

$$UCL = p + 3 \sqrt{\frac{p(1-p)}{n}}$$

Center line $= p$

$$LCL = p - 3 \sqrt{\frac{p(1-p)}{n}}.$$

In this chart and later, a negative LCL is replaced by 0, and in the $p$ chart, a UCL greater than 1 is replaced by 1.
If $p$ is unknown, take $m$ preliminary samples of size $n$.

Let

$$\hat{p}_i = \frac{D_i}{n}, \quad i = 1, 2, \ldots, m,$$

and

$$\bar{p} = \frac{1}{m} \sum_{i=1}^{m} \hat{p}_i = \frac{1}{mn} \sum_{i=1}^{m} D_i.$$
Plot each $\hat{p}$ on a chart with a center line at $\bar{p}$ and three-sigma control limits:

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

Center line = $\bar{p}$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}.$$
Example: Orange juice cans

Cans are inspected for defective seals on side seam or bottom joint.

Development data: \( m = 30 \) samples of size \( n = 50 \), selected at 30 minute intervals.

In R:

cans <- read.csv("Data/Table-07-01-02.csv")
cansDev <- with(cans, Nonconforming[Phase == "I"])
summary(qcc(cansDev, type = "p", sizes = 50))
Data points 15 and 23 are above the UCL; inspection reveals assignable causes to both (new batch of cardboard, inexperienced operator).

Recompute limits, excluding those points:

```
summary(qcc(cansDev[-c(15, 23)], type = "p", sizes = 50))
```

One more data point is above the new UCL, but no assignable cause was found, and the process was declared in-control.

Note: Montgomery suggests including points 15 and 23 in the chart, just leaving them out of the calculations; not currently possible in `qcc()`.
The percentage of nonconforming cans (21%) was not acceptable, so the equipment was adjusted.

Using the chart for subsequent operations; 24 additional samples:

```r
cansUse <- with(cans, Nonconforming[Phase == "II"])
summary(qcc(cansDev[-c(15, 23)], type = "p", sizes = 50,
             newdata = cansUse, newsizes = 50))
```

The lengthy run below the center line and the one point below the LCL indicate that the fraction nonconforming has been reduced.

The control limits were recomputed based on the later samples.
The *np* chart

In some areas the convention is to chart the *number* nonconforming, instead of the *fraction* nonconforming.

When the sample sizes are all equal, this is a trivial change:

```r
summary(qcc(cansDev[-c(15, 23)], type = "np", sizes = 50,
    newdata = cansUse, newsizes = 50))
```
Unequal sample sizes

If the development data have sample sizes \( n_i, i = 1, 2, \ldots, m \), we estimate \( p \) by

\[
\bar{p} = \frac{\sum_{i=1}^{m} D_i}{\sum_{i=1}^{m} n_i}
\]

Three-sigma control limits vary by sample:

\[
\text{UCL} = \bar{p} + 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_i}} \\
\text{Center line} = \bar{p} \\
\text{LCL} = \bar{p} - 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_i}}.
\]
Example: Erroneous purchase orders

Purchase orders issued by an aerospace company may not be correct. Samples are based on weeks of operation, and different numbers of POs are issued each week.

```r
orders <- read.csv("Data/Table-07-04.csv")
summary(qcc(orders$Nonconforming, sizes = orders$Size,
    type = "p"))
```

Other solutions are to use the average sample size (not recommended) or the **standardized** chart, based on

\[ Z_i = \frac{\hat{p}_i - p}{\sqrt{\frac{p(1-p)}{n_i}}} \]
Operating characteristics

The operating characteristic (OC) curve is a graph of the $\beta$-risk against the actual fraction nonconforming.

For the orange juice cans:

```r
oc.curves(qcc(cansDev, type = "p", sizes = 50))
```

The flat top means that small changes in $p$ are hard to detect in a single sample, but the $ARL_1 = 1/(1 - \beta)$ may be acceptable.
Control Chart for Nonconformities

In some contexts, a product or a service as a whole may be classified as nonconforming if it contains too many individual nonconformities.

For example: flaws in the paint of a car; one or two may be acceptable, but more than two is not.
Statistical framework

The simplest distribution for counted data like these is the Poisson distribution:

\[ P(X = x) = \frac{e^{-c}c^x}{x!}, \]

where \( X \) is the number of nonconformities and \( c \) is the parameter of the distribution.

Since

\[ E(X) = \text{Var}(X) = c, \]

three-sigma limits for known \( c \) are:

\[ \text{UCL} = c + 3\sqrt{c} \]
\[ \text{Center line} = c \]
\[ \text{LCL} = c - 3\sqrt{c}. \]
When \( c \) is unknown, \( m \) preliminary counts \( x_1, x_2, \ldots, x_m \) are made, and \( c \) is estimated by

\[
\bar{c} = \bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i.
\]

Three-sigma limits for unknown \( c \) are then:

- **UCL** = \( \bar{c} + 3\sqrt{\bar{c}} \)
- **Center line** = \( \bar{c} \)
- **LCL** = \( \bar{c} - 3\sqrt{\bar{c}} \).

This is the \( c \) control chart.
Example: printed circuit boards

The inspection unit is a sample of 100 boards, and \( X \) is the number of nonconformities in the sample. Preliminary counts are for \( m = 26 \) samples.

In R:

```r
b我不懂 coaches <- read.csv("Data/Table-07-07-08.csv")
boardsDev <- with(boards, Nonconformities[Phase == "I"])
summary(qcc(boardsDev, type = "c"))
```
Two data points are outside the control limits, and investigation showed that both have assignable causes (inexperienced inspector, problem with wave soldering), so they are omitted:

summary(qcc(boardsDev[-c(6, 20)], type = "c"))

Remaining points are all within limits, so these limits are used for subsequent operations:

summary(qcc(boardsDev[-c(6, 20)], type = "c", newdata = boardsUse))
Further analysis

The number of nonconformities is high, so a Pareto chart was used to show the types of nonconformity:

```r
boardsPareto <- read.csv("Data/Figure-07-14.csv")
Frequency <- boardsPareto$Frequency
names(Frequency) <- boardsPareto$Defect.Code
pareto.chart(Frequency)
```

The chart shows that most of the nonconformities were of just two types, both related to soldering.
Unequal sample sizes

In some situations, nonconformities are counted for different numbers of sampling units in each sample.

If \( X_i \) is the number of defects for \( n_i \) sampling units, and the average number of defects per sampling unit is \( u \), then \( X_i \sim \text{Poisson}(n_iu) \), and

\[
E(X_i) = \text{Var}(U_i) = n_iu.
\]

If these counts \( x_i \) are plotted on a \( c \) chart, the center line and the control lines all vary from sample to sample; \( \text{qcc}() \) refuses to make such a chart.
Instead, let

\[ U_i = \frac{X_i}{n_i}, \]

so that

\[ \mathbb{E}(U_i) = u, \quad \text{Var}(U_i) = \frac{u}{n_i}. \]

These may be plotted on a \( u \) chart with three-sigma limits, which for known \( u \) are:

\[
\begin{align*}
\text{UCL} &= u + 3 \sqrt{\frac{u}{n_i}} \\
\text{Center line} &= u \\
\text{LCL} &= u - 3 \sqrt{\frac{u}{n_i}}.
\end{align*}
\]
For unknown \( u \), estimate \( u \) from \( m \) preliminary samples:

\[
\bar{u} = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m} n_i u_i}{\sum_{i=1}^{m} n_i}.
\]

The three-sigma limits become

\[
UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n_i}}
\]

Center line = \( \bar{u} \)

\[
LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n_i}}.
\]
Example: Dyed cloth

Dyed cloth is inspected for defects per 50 square meters. Ten rolls are inspected, with varying areas.

cloth <- read.csv("Data/Table-07-12.csv")
summary(qcc(cloth$Nonconformities, type = "u",
sizes = cloth$Square.Meters / 50))

The $u$ chart may also be used with equal sample sizes, when it is desired to report results on a per-unit basis instead of a per-sample basis.