

X-Bar/R Control Charts

Control charts are used to analyze variation within processes. There are many different flavors of control charts, categorized depending upon whether you are tracking **variables** directly (e.g. height, weight, cost, temperature, density) or **attributes** of the entire process (e.g. number of defective parts produced, proportion of defectives). The **X-Bar/R control chart** is one of these flavors. It's used for **variable data** when the data is **readily available**. This is one of the most commonly encountered control chart variants, and leverages two different views:

- The X-Bar chart shows how much variation exists in the process *over time*.
- The Range (R) chart shows the variation *within each variable (called "subgroups")*.

A process that is in statistical control is *predictable*, and characterized by points that fall between the lower and upper *control limits*. When an X-Bar/R chart is in statistical control, the average value for each subgroup is consistent over time, and the variation within a subgroup is also consistent. Control limits are not the same as specification limits, but both are important when we are performing process analysis:

- **Control limits** are characteristics of the process. They reflect the *actual* amount of variation that is observed. We assume that the variation can be described by the normal distribution, which means that 99.73% of all of our observations will fall somewhere between three standard deviations below the mean (-3σ) and three standard deviations above the mean ($+3\sigma$). We use this principle to set our control limits.
- **Specification limits** give us a sense of the voice of the customer. They reflect the variation that the customer is willing to accept from the process. We use the *target* value for our variable, and the customer-specified *tolerance* around that variable to determine the specification limits. (There is no connection between the tolerance that the customer specifies, and the observed standard deviation of the process.)

| Control Limits for X-Bar Chart | Control Limits for R Chart | Specification Limits |
|---------------------------------|---------------------------------|---------------------------|
| UCL = Process Mean + A_2R | UCL = Average Range x D_4 | USL = Target + Tolerance |
| Center Line = Average of X-Bars | Center Line = Average of Ranges | CL = Center Line = Target |
| LCL = Process Mean - A_2R | LCL = Average Range x D_3 | LSL = Target - Tolerance |

Calculate Control Limits

To determine the control limits, we apply the following formulas, where \bar{X} with a double bar indicates the average of all the X-Bars, and the constants A2, D3, and D4 are selected from the table below based on the number of items in each of your subgroups:

$$UCL_{\bar{x}} = \bar{\bar{X}} + A_2 \bar{R}$$

$$LCL_{\bar{x}} = \bar{\bar{X}} - A_2 \bar{R}$$

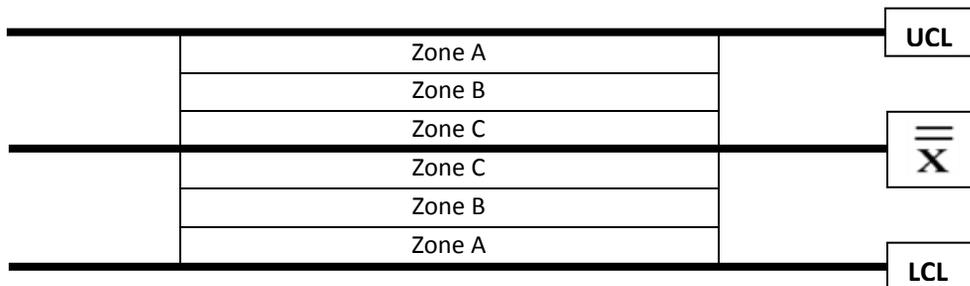
$$UCL_x = D_4 \bar{R}$$

$$LCL_x = D_3 \bar{R}$$

| n | A2 | A3 | d2 | D3 | D4 | B3 | B4 |
|----|-------|-------|-------|-------|-------|-------|-------|
| 2 | 1.88 | 2.659 | 1.128 | 0 | 3.267 | 0 | 3.267 |
| 3 | 1.023 | 1.954 | 1.693 | 0 | 2.574 | 0 | 2.568 |
| 4 | 0.729 | 1.628 | 2.059 | 0 | 2.282 | 0 | 2.266 |
| 5 | 0.577 | 1.427 | 2.326 | 0 | 2.114 | 0 | 2.089 |
| 6 | 0.483 | 1.287 | 2.534 | 0 | 2.004 | 0.03 | 1.97 |
| 7 | 0.419 | 1.182 | 2.704 | 0.076 | 1.924 | 0.118 | 1.882 |
| 8 | 0.373 | 1.099 | 2.847 | 0.136 | 1.864 | 0.185 | 1.815 |
| 9 | 0.337 | 1.032 | 2.97 | 0.184 | 1.816 | 0.239 | 1.761 |
| 10 | 0.308 | 0.975 | 3.078 | 0.223 | 1.777 | 0.284 | 1.716 |
| 11 | 0.285 | 0.927 | 3.173 | 0.256 | 1.744 | 0.321 | 1.679 |
| 12 | 0.266 | 0.886 | 3.258 | 0.283 | 1.717 | 0.354 | 1.646 |
| 13 | 0.249 | 0.85 | 3.336 | 0.307 | 1.693 | 0.382 | 1.618 |
| 14 | 0.235 | 0.817 | 3.407 | 0.328 | 1.672 | 0.406 | 1.594 |
| 15 | 0.223 | 0.789 | 3.472 | 0.347 | 1.653 | 0.428 | 1.572 |
| 16 | 0.212 | 0.763 | 3.532 | 0.363 | 1.637 | 0.448 | 1.552 |
| 17 | 0.203 | 0.739 | 3.588 | 0.378 | 1.622 | 0.466 | 1.534 |
| 18 | 0.194 | 0.718 | 3.64 | 0.391 | 1.608 | 0.482 | 1.518 |
| 19 | 0.187 | 0.698 | 3.689 | 0.403 | 1.597 | 0.497 | 1.503 |
| 20 | 0.18 | 0.68 | 3.735 | 0.415 | 1.585 | 0.51 | 1.49 |
| 21 | 0.173 | 0.663 | 3.778 | 0.425 | 1.575 | 0.523 | 1.477 |
| 22 | 0.167 | 0.647 | 3.819 | 0.434 | 1.566 | 0.534 | 1.466 |
| 23 | 0.162 | 0.633 | 3.858 | 0.443 | 1.557 | 0.545 | 1.455 |
| 24 | 0.157 | 0.619 | 3.895 | 0.451 | 1.548 | 0.555 | 1.445 |
| 25 | 0.153 | 0.606 | 3.931 | 0.459 | 1.541 | 0.565 | 1.435 |

Is the Process in Control? Using Run Rules

To determine whether the process is in control, we apply *run rules*, FIRST to the R chart and then to the X-Bar chart. Some of the run rules are based on the concept of *zones*:



The upper zones are obtained by splitting the distance between the center line and the upper limit by 3. The lower zones are obtained by splitting the distance between the center line and the lower limit by 3. You test the run rules by asking yourself the following questions, first for the R chart, and then for the X-Bar chart:

1. Are there any points below the LCL or above the UCL?
2. Are 2 out of 3 subsequent points in Zone A or beyond?
3. Are 4 out of 5 subsequent points in Zone B or beyond?
4. Are 8 or more points on one side of the center line?
5. Are there 6 or more subsequent points steadily increasing? Are there 6 or more subsequent points steadily decreasing?

If you answer YES to any of the above questions, then the R chart is out of control and the process limits on the X-Bar chart are *meaningless*. If the R Chart appears to be in control, check the run rules against the X-Bar chart. If you answer YES to any of these 5 questions, then the process is out of control. This indicates the presence of *special cause variation*, which means that action should be taken to bring the process back into control.

Is the Process Capable?

Process capability provides a way to assess whether the control limits and the specification limits are sufficiently lined up. Because who cares if the process is in control if it's not generating the output we want? To make this decision, we use the process capability index (C_{pk}) like this:

1. First, find the difference between X-Bar and the LSL, and divide by 3σ .
2. Next, find the difference between X-Bar and the USL, and divide by 3σ .
3. Compare those two values, and pick the smaller one. That will be your C_{pk} .
 - If $C_{pk} \geq 1.33$, then the process is capable.
 - If $C_{pk} < 1.33$, then the process is *not* capable.

An Example with Simulated Data

We want to examine the **within-piece variability** for some PVC pipe that our company manufactures. With a digital caliper, we measure the outside diameter of the pipe on one end, and then repeat that measurement with the caliper oriented *perpendicular* to its original position. That way, we capture some of the variability in diameter just on that end of the pipe. Next, we measure the outside diameter of the same piece of pipe, but at the other end. We repeat the measurement with the caliper oriented perpendicular to its original position to capture the variability on that end.

Together, these four measurements are called a *subgroup*, and become our first *sample*. Our sample size (4) is the number of items in each subgroup. We can track as many subgroups as we want, **but we need to make sure that each subsequent subgroup represents measurements that are made later in time.**

| Subgroup # | Measurements | | | | Sample Average (X-Bar) | Range (R) |
|------------|--------------|----|----|----|------------------------|-----------|
| | #1 | #2 | #3 | #4 | | |
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | | | | | | |

We want to examine the **within-piece variability** in the R statistical software. First, install the `qcc` library from CRAN and then call it into active memory using the `library` command:

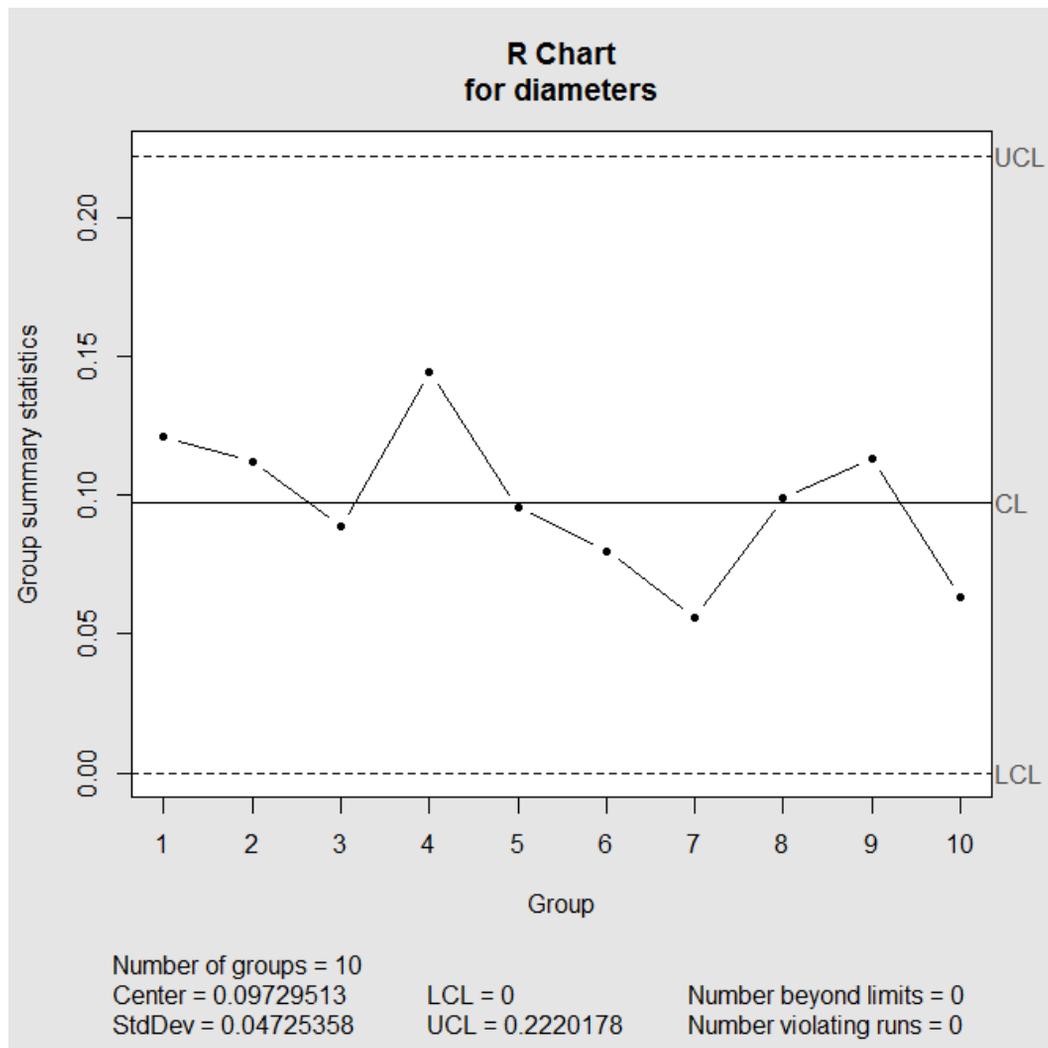
```
install.packages("qcc")  
library(qcc)
```

Next, let's generate some simulated data. We will use the `rnorm` function to pull diameters for 10 different subgroups, each of which is composed of 4 measurements:

```
> diameters <- as.data.frame(replicate(4, rnorm(10,mean=1.31,sd=0.05)))
> diameters
      V1      V2      V3      V4
1 1.397742 1.399917 1.278918 1.279828
2 1.283877 1.307215 1.341566 1.396107
3 1.313634 1.278839 1.242498 1.331201
4 1.245943 1.303432 1.390168 1.298949
5 1.188624 1.226905 1.217450 1.284127
6 1.287875 1.283185 1.234186 1.314055
7 1.276711 1.284376 1.305309 1.249184
8 1.312219 1.297509 1.272367 1.371223
9 1.378350 1.312981 1.381944 1.268875
10 1.332196 1.268824 1.299608 1.329053
```

First, we use the `qcc` function to generate an R chart:

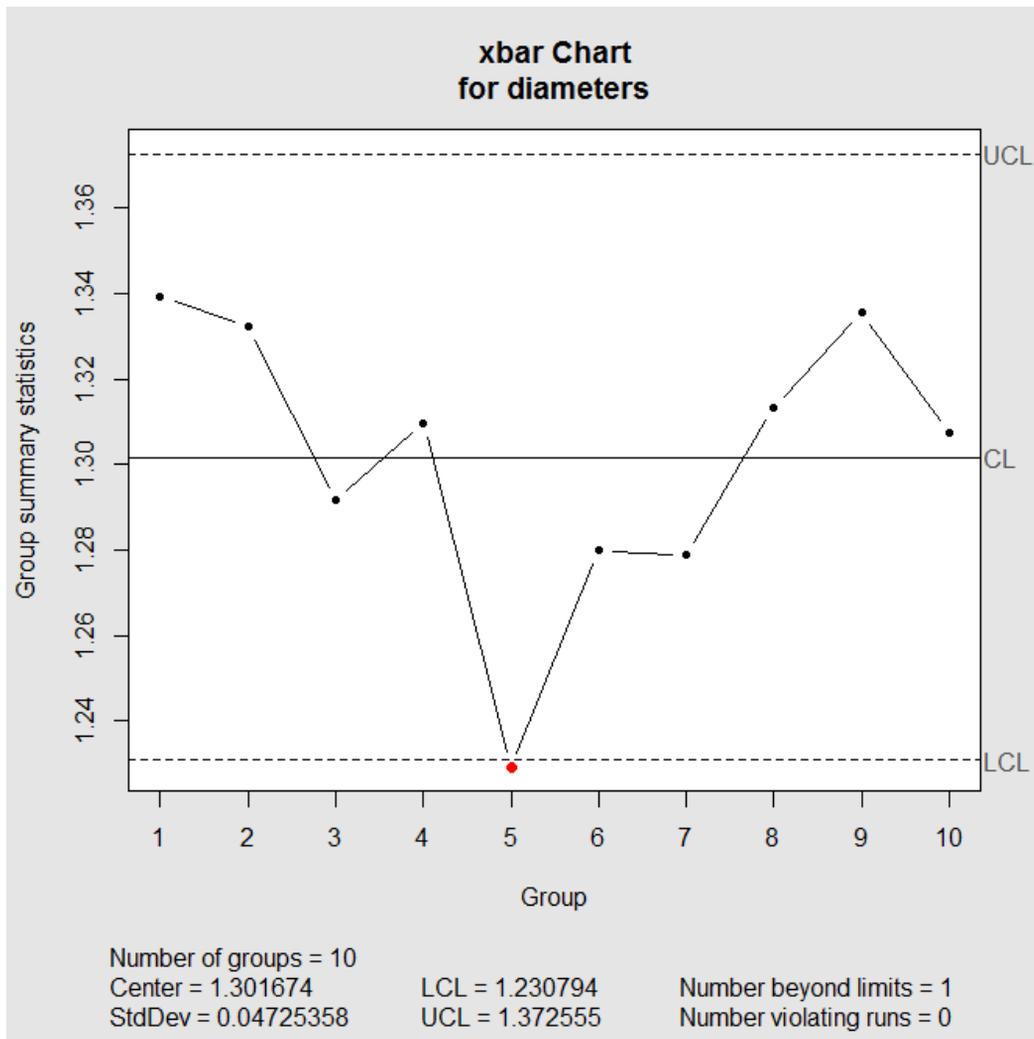
```
> q <- qcc(diameters, type="R", nsigmas=3)
```



The R chart appears to be in control (none of the run rules have been violated). Next, we use the `qcc` function to generate the X-Bar control chart:

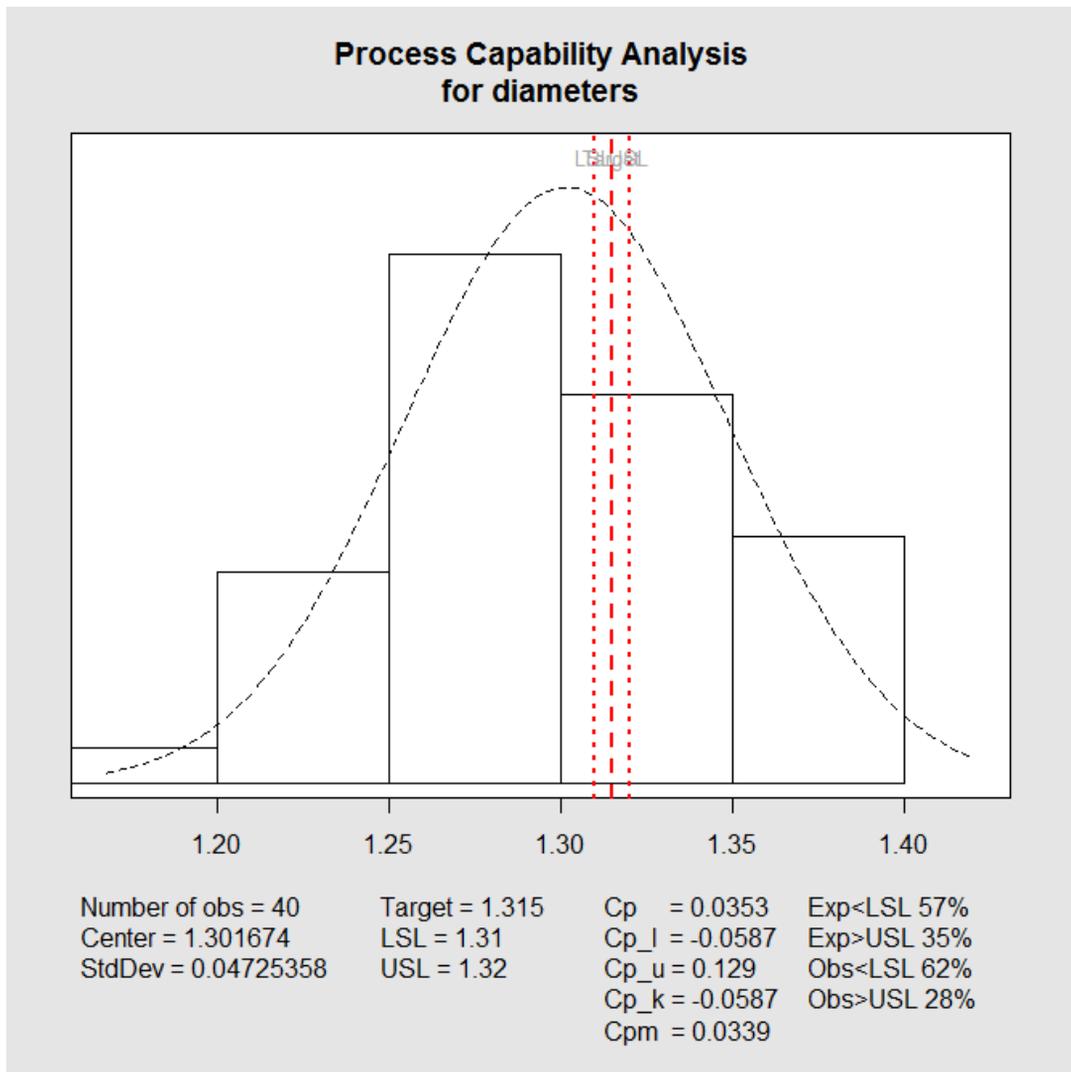
```
> q <- qcc(diameters, type="xbar", nsigmas=3)
```

That command *immediately* plots the control chart, highlighting out of control points in red using the run rules described above. Notice that there is a standard deviation reported in the legend for the chart on the next page: this is the standard deviation *of the X-Bars*:



It is easy to see that for this particular sample, the average diameter of the measurements in subgroup 5 were below the Lower Control Limit so **the process is not in control**. Let's say we know that our customers have demanded that the diameters be within 0.005 in (the tolerance) of 1.315 in (the target). The LSL is $1.315 - 0.005 = 1.31$, and the USL is $1.315 + 0.005 = 1.32$. We can also use the `qcc` package to see these values on a full process capability analysis:

```
> process.capability(q, spec.limits=c(1.31,1.32))
```



This plot shows that the distribution of diameters that we measured is *way* beyond what we would like to see: the LSL, the target, and the USL (the vertical orange lines, from left to right) are embedded within the distribution of diameters. In addition to being out of control, this process is *not capable*.

Note: This example uses **SIMULATED** data. The results you get when **YOU** create your own simulated data will be *different* than the charts that appear above.

Analyzing Your Own Data

Here is an example of how to analyze three different subgroups, each of sample size 4. You can process as many subgroups of your own data using the R code below. Be sure to put **your** measured values in each line. Create a variable for EACH of your subgroups (so if you have more than 3 subgroups, you might also define `subgroup4`, `subgroup5`, etc. and also include them in the `rbind` command).

```
# Load your data FROM A CSV FILE with one subgroup per row
my.data <- read.csv("my-data.csv",header=FALSE)

# OR, Load your data for each subgroup MANUALLY:
sg1 <- c(1.397742,1.399917,1.278918,1.279828) # Fill in YOUR data!
sg2 <- c(1.397742,1.399917,1.278918,1.279828) # Fill in YOUR data!
sg3 <- c(1.397742,1.399917,1.278918,1.279828) # Fill in YOUR data!
# If you have more than one subgroup, include it in this list:
my.data <- rbind(sg1,sg2,sg3)

# Draw the R Chart and calculate relevant metrics
q1 <- qcc(my.data, type="R", nsigmas=3)

# Draw the X-Bar Chart and calculate relevant metrics
q2 <- qcc(my.data, type="xbar", nsigmas=3)

# Establish the LSL and USL as set by customer specs, then
# draw the process capability chart and calculate metrics:
lsl <- 1.31 # Fill in YOUR LSL here!
usl <- 1.32 # Fill in YOUR USL here!
process.capability(q2, spec.limits=c(lsl,usl))
```

Other Resources

- **Where do control chart constants come from?** The short answer is that they are based on measures of dispersion that are derived from the normal model. The longer (and fully derived) answer is here: <https://www.spcforexcel.com/knowledge/control-chart-basics/control-limits>
- **Control chart constants** can also be found here: <http://www.dfx.nl/userfiles/file/pdf/Control%20Chart%20Constants%20and%20Formulas.pdf>