Exercise 2 - The Six Sigma Method

Alexander J Ohrt - Statistics for Business Management

06.10.2021

Exercise done in place of lecture

A process is assumed optimal if the production time (in minutes) per unit is distributed as a Normal (30, 2), i.e. a normal distribution with mean equal 30 and variance equal 2. A set of production times are observed for 5, 000 units. The file "prod.csv" contains the data

```
# Import data.
Y<-read.csv("prod.csv",sep = ",",col.names = 1)
Y<-as.matrix(Y)
head(Y)
```

```
#>            X1
#> [1,] 27.15271
#> [2,] 29.92395
#> [3,] 29.41166
#> [4,] 30.98629
#> [5,] 29.31611
#> [6,] 24.30063
```

```
summary(Y)
```

```
#>        X1
#>  Min.   :17.42
#>  1st Qu.:27.99
#>  Median :30.05
#>  Mean   :30.04
#>  3rd Qu.:32.06
#>  Max.   :40.28
```

```
# Parameters.
mu=30
sigma=sqrt(2)
```

1. Calculate the LSL and USL values in a Six Sigma environment.

```
(LSL <- (-6) * sigma + mu)
```

```
#> [1] 21.51472
```

```
(USL <- 6 * sigma + mu)
```

```
#> [1] 38.48528
```

As is apparent from the code above, LSL ≈ 21.51 and USL ≈ 38.49.

2. Analyse how the actual production process (data set information) is not adapted to Six Sigma environment

First of all, DPMO $\approx$ 3.4, as is seen from the output below also.

```
(DPMO <- 1000000*(1-pnorm(USL,mu+1.5*sigma,sigma)))
```

```
#> [1] 3.397673
```

This means that we will have approximately 3.4 defective parts per million opportunities in the long run. Hence, this is a Six Sigma process (This is the theoretical processs).

Furthermore, the amount of observations that are outside the limits of LSL and USL are

```
(out <- length(Y[Y>USL | Y < LSL]))
```

```
#> [1] 20
```

which means that

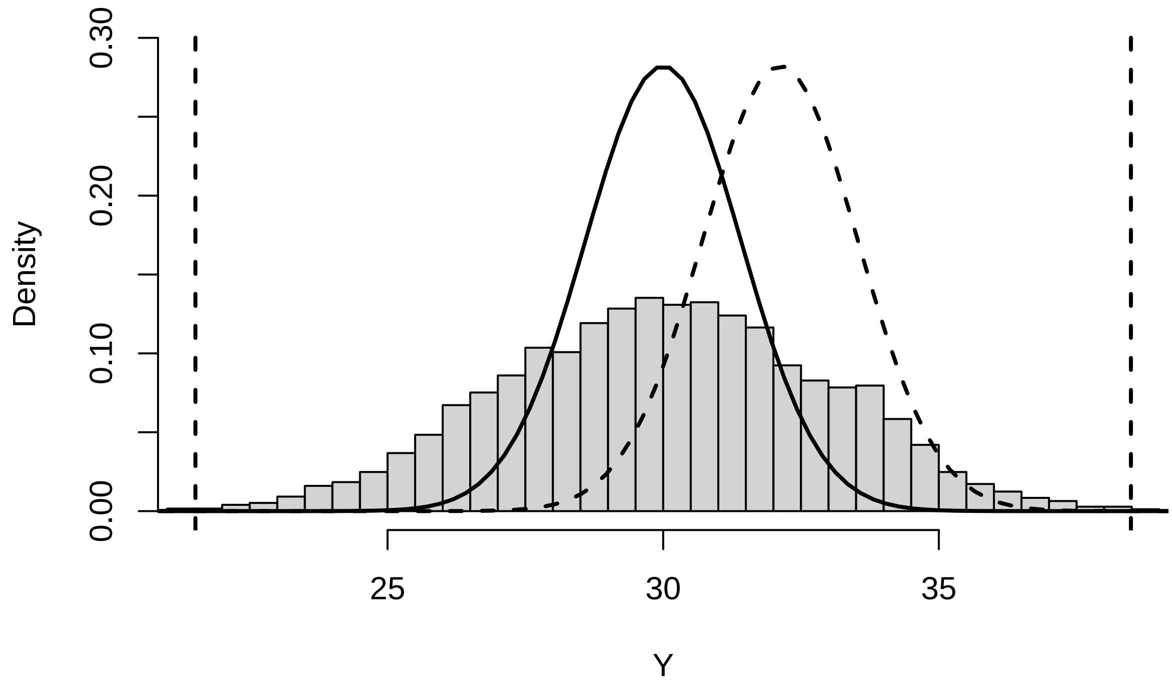
```
out/dim(Y)[1]*10^6
```

```
#> [1] 4000
```

and hence, from the Sigma Scale table, this is a 4 sigma process. **This is a contradiction, what is correct here?** –> This is the empirical process, while the above shows that the theoretical process is optimal, i.e. Six Sigma.

Moreover, the code block below shows a plot of LSL, USL, the standard normal probability distribution with mean 0 and mean 1.5 and a histogram of the data.

```
x <- seq(min(Y), max(Y), length = 100)
f1 <- dnorm(x, mean = mu+1.5*sigma, sd = sigma)
f2 <- dnorm(x, mean = mu, sd = sigma)
hist(Y,xlim = c(LSL,USL),ylim=c(0,0.3),probability = T,main = "", breaks = 50)
lines(x,f1,type="l",lwd=2,lty=2)
lines(x,f2,type="l",lwd=2,lty=1)
abline(v=LSL,lwd=2,lty=2)
abline(v=USL,lwd=2,lty=2)
```



We can see that there are no defects above the USL, which again shows that this is a Six Sigma process.

3. Calculate the number of faults per million following the observed production process.

```
(faults<-sum((Y>USL)+(Y<LSL)))
```

```
#> [1] 20
```

As seen from the calculation above, the number of fault per million is approximately 20 (I already found this above also).

4. Obtain the $k\sigma$ process that is adapted to the observed data

```
prob<-faults/nrow(Y)
prob
```

```
#> [1] 0.004
```

```
k<-qnorm(1-prob,1.5,1)
k
```

```
#> [1] 4.15207
```

```
sqrt(var(Y))
```

```
#>          X1
#> X1 2.946705
```

What does this mean? –> Hence, one needs to reduce the variance from 2.95 (which is the std in the data, as seen above) to 2 (which is the given std for the theoretical optimal process) in order to reah the optimal process.

5. In Atenea: Calculated C_p also.