

Exploratory Data Analysis

Enqun Wang

April 23, 2016

Input the data

```
setwd("/Users/ewenwang/Dropbox/Data Science/DMAIC/Case Study/3-Analyze")

require(dplyr)

df = data.frame(read.csv("data.csv", header = T)[-c(1:4),-1])
df <- tbl_df(df)
glimpse(df)
```

```
## Observations: 78
## Variables: 16
## $ Trading.Volume      (dbl) 84797489, 93074580, 118554712, 91550434, 81...
## $ Increase.Rate.1Day  (dbl) 26.009443, 9.761009, 27.376037, -22.777903,...
## $ Increase.Rate.5Day  (dbl) 0.7633873, 0.7973166, 1.3512288, 1.1438346,...
## $ Rate.of.Return      (dbl) 9.338962, 9.927363, 9.368053, 9.168878, 9.3...
## $ Term                (dbl) 4.365299, 4.067725, 2.676194, 2.518341, 5.3...
## $ Balance             (dbl) 174346100, 248412317, 205746271, 194141712,...
## $ Capital.Inflow      (dbl) 6291296.0, -14027301.8, -22458699.0, 231910...
## $ TZD.Inflow          (int) 11817494, 9723379, 5491484, 4786453, 125599...
## $ TZD.Account         (int) 24661231, 13514789, 5067603, 4875912, 18356...
## $ LB.Received         (dbl) 990006.8, 483006.8, 489531.0, 825899.4, 317...
## $ LB.Used             (int) 670929, 492902, 727923, 621826, 686442, 162...
## $ R.007              (dbl) 6.000, 4.350, 3.600, 3.002, 2.200, 2.470, 3...
## $ Interbank.Rate      (dbl) 3.3802, 3.3824, 3.5145, 3.5345, 3.4598, 3.3...
## $ SHIBOR             (dbl) 3.3500, 3.3470, 3.4644, 3.4360, 3.4030, 3.3...
## $ SHA                (dbl) 3286.07, 3290.90, 3349.32, 3372.91, 3449.30...
## $ GEM                (dbl) 2045.32, 2035.83, 2018.13, 2069.07, 2142.67...
```

Defect, Baseline and Goal

```
summary(df$Increase.Rate.5Day)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.5714  0.9051  1.0170  1.0370  1.1300  1.6500
```

```
defect.value = quantile(df$Increase.Rate.5Day, 0.15)
```

```
baseline.IR = mean(df$Increase.Rate.5Day)
```

```
sd.IR = sd(df$Increase.Rate.5Day)
```

```
baseline.defect = sum(df$Increase.Rate.5Day < defect.value)/sum(df$Increase.Rate.5Day)
```

```
goal = 1.10
```

```
goal.defect = pnorm(defect.value, goal, sd.IR)
```

```
improvement = (baseline.defect - goal.defect)/goal.defect

hist(df$Increase.Rate.5Day, br = 30, freq = F)
x = seq(-0.65, 1.65, 0.01)
lines(x, dnorm(x, baseline.IR, sd.IR), col = "green")
abline(v = defect.value, col = "red")
lines(x, dnorm(x, goal, sd.IR), col = "blue")
```

Histogram of df\$Increase.Rate.5Day

