Into to Inferential Statistics

Lesson 1

Klout score example - Online score (1 -99) credit score for online reputation . Klout score follows bimodal distribution . Having 40 is good Klout score.

```
library(tidyverse)
```

Reading the data

```
klout_data=read_csv('datasets/klout_score.csv')
```

```
## Parsed with column specification:
## cols(
##   score = col_double()
## )
```

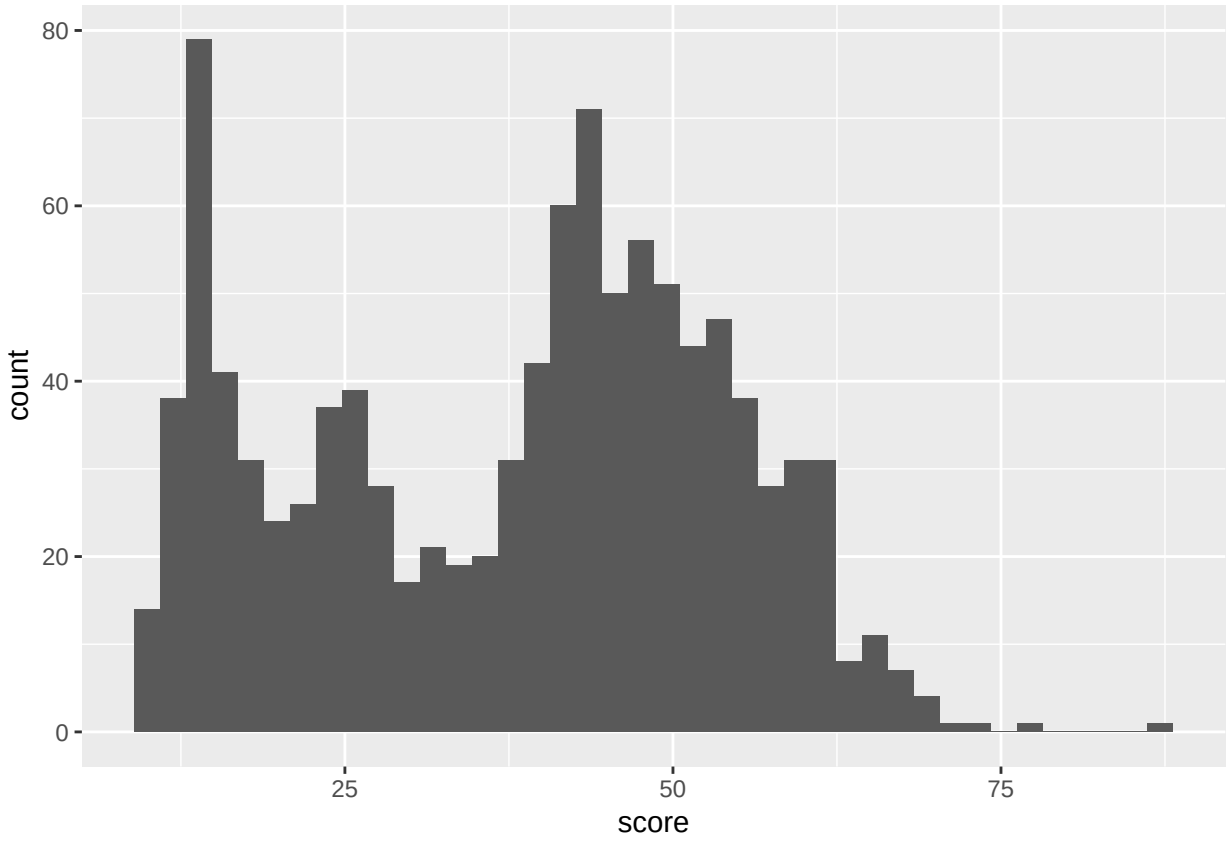
summary statistics

```
summary(klout_data)
```

```
##      score
##  Min.   :10.00
##  1st Qu.:23.17
##  Median :41.09
##  Mean   :37.72
##  3rd Qu.:50.28
##  Max.   :87.25
```

Below histograms shows distribution is bimodal

```
ggplot(klout_data,aes(x=score))+geom_histogram(bins=40)
```



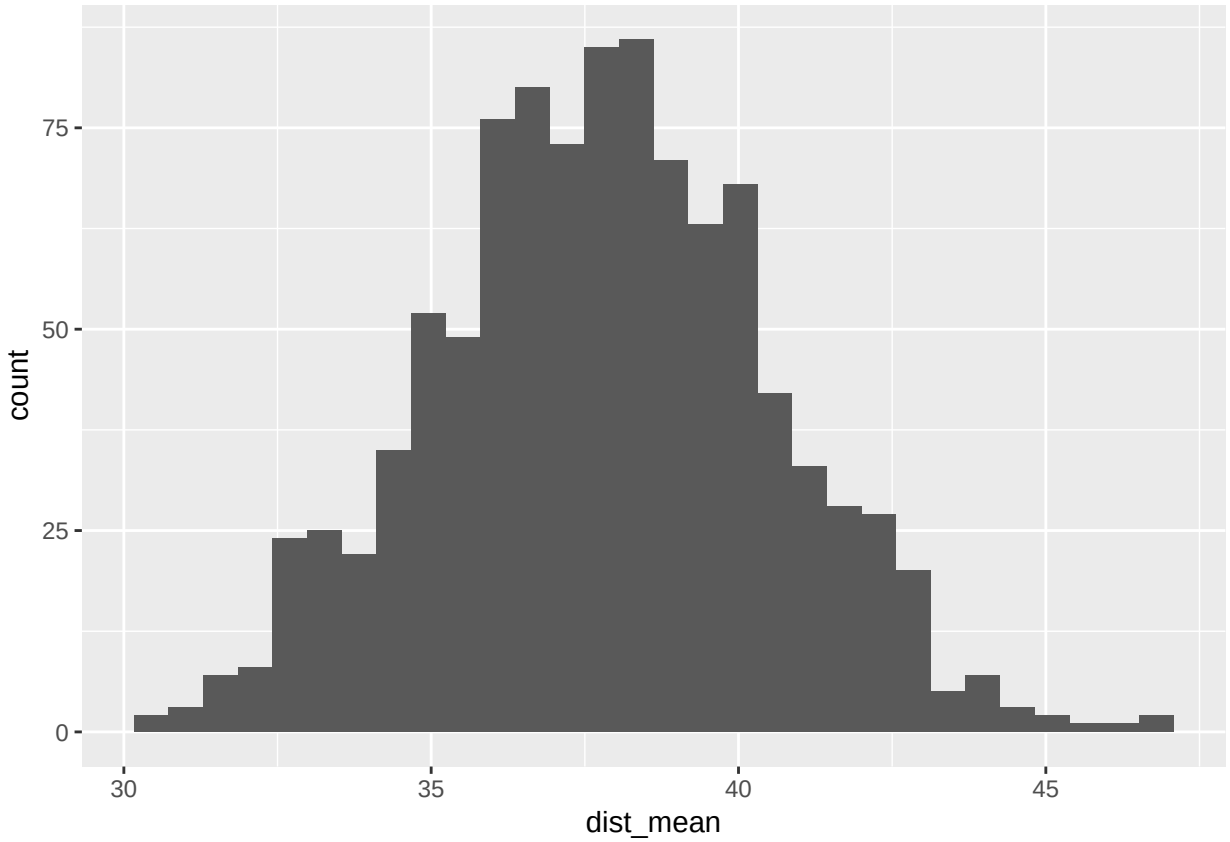
As per Central Limit therom , mean of random samples from the distribution is normally distributed.

```
sample_mean=c()

# taking the 35 sample 1000 times taking the average and saving it in the variable.
for (i in 1:1000){
  sample_mean=c(sample_mean,mean(sample(klout_data$score,35,replace=TRUE)))
}

ggplot(data.frame(dist_mean=sample_mean),aes(x=dist_mean))+geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As per central limit therom , sample distribution mean is same as population mean

$$\mu_{sampledistribution} \approx \mu_{population}$$