Memo for Advanced Transportation Planning

1) Our class will basically be operated by “Real-time”. The application is Webex.
2) If you don’t have appropriate tele-communication method, please ask me.
3) This class requires basic statistics knowledge. If you don’t know the following keywords in the next page, please study by yourself. Google or Wikipedia may offer the useful information.

Mean, Average: $\overline{x} = \frac{\sum_{n=1}^{N} x_n}{N}$

Dispersion, Variance: $\sigma^2 = \frac{\sum_{n=1}^{N} (x_n - \overline{x})^2}{N}$

Standard Deviation, S.D.: $\sigma = \sqrt{\sigma^2}$

Coefficient of Variation: $C.V. = \frac{\sigma}{\overline{x}}$

HENSACHI（偏差値）： $50 + \frac{(x - \overline{x}) \times 10}{\sigma}$

Co-variance: $\sigma_{xy}^2 = \frac{\sum_{n=1}^{N} (x_n - \overline{x})(y_n - \overline{y})}{N}$

Correlation Coefficient: $\rho = \frac{\sigma_{xy}^2}{\sqrt{\sigma_x^2 \sigma_y^2}}$

Root Mean Square: $RMS = \sqrt{\frac{\sum_{n=1}^{N} x_n^2}{N}}$

Probabilistic Density function: $f(x)$

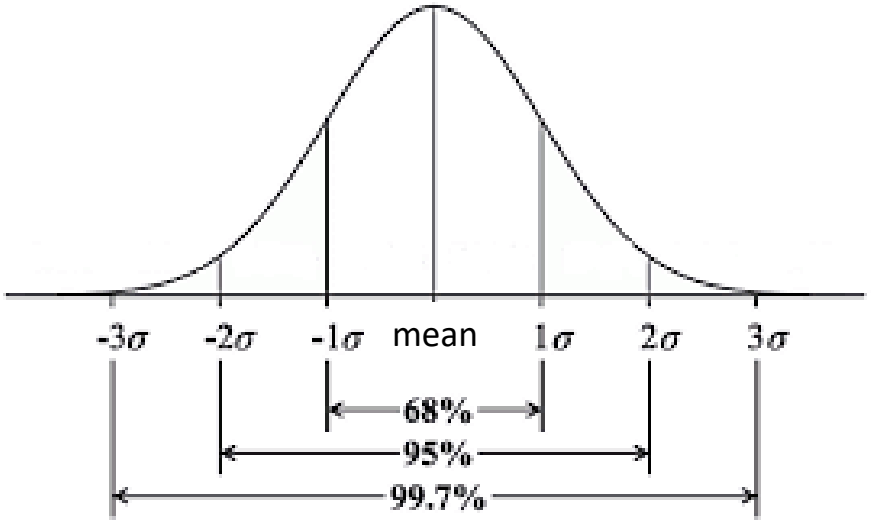
Probabilistic Distribution function: $F(x) = \int_{-\infty}^{x} f(z)dz$

Normal Distribution with mean μ & dispersion σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rightarrow N(\mu, \sigma^2)$$

another name is "Gaussian Distribution"

$$\int_{-\sigma}^{\sigma} f(x)dx = 0.68, \int_{-1.96\sigma}^{1.96\sigma} f(x)dx = 0.95$$



Please understand the differences between

Mode, Mean, Median

This class requires basic skill of “R” software.

You should install R package on your PC.

https://cran.r-project.org/bin/windows/base/

The latest version is 4.0.0 just released last month

If you want the basic information for R, the following YouTube may be helpful.

https://www.youtube.com/watch?v=_V8eKsto3Ug&list=RDCMUC8butISFwT-Wl7EV0hUK0BQ&start_radio=1#t=1001

Or you can find many PDFs for R beginners, you don’t need English. Japanese or Chinese are also available.

At least you should know the following commands in R.

test1.csv

```
x_1, x_2
15,200
18,210
20,220
14,190
12,150
30,230
28,250
16,190
13,120
22,230
```

setwd(), getwd()

1)Reading CSV file:

```
f <- read.csv(“test1.csv",header=TRUE)
f2 <- read.csv("filename", row.names=1)
```

Let’s try:

```
str(f),head(f),tail(f),summary(f)
```

f$x_1 and f[,1] is same meaning

2) Drawing histogram:

hist(f$x_1)

boxplot(f$x_1) shows basic statistics.

Mean, Median, Quartile

3) Relationship between two variables

plot(fx_1,fx_2) → scatterplot

4) Statistical indices between two variables

cor(fx_1,fx_2)→ correlation coefficient

var(f$x_1), mean(f$x_2), sd(f$x_1)

var(fx_1,fx_2) co-variance

5) Single regression model

$y_n = ax_n + b + \varepsilon_n$

plot(fx_1,fx_2)→ check the scatterplot

res <- lm(f$x_2 ~ f$x_1)

→ Estimating f$x_2=a* f$x_1+b, the results are stored in “res”

summary(res) → check the results

6) Generating array

```
fm <- matrix(values,# of rows,# of columns)
   Ex)fm <- matrix(1,10,5)
fm <- read.csv("Japan.csv",header=TRUE,
      row.names=1)
pairs(fm)
```

7) Making subgroups

By using "test1.csv", pick up data with first column is equal or over 20

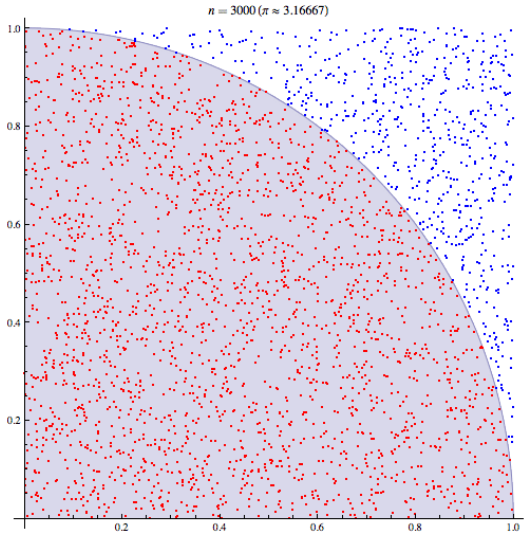
```
f[,1]>=20
```

What are generated ? By using this result, sub array can be generated by…

```
f2 <- f[f[,1]>20,]
```

8) Monte-Carlo simulation

By generating uniform random value 1000*1000, estimate PI




```
num <- 1000
x <- runif(num); y <- runif(num)
z <- sum(x^2+y^2<1)
z/num*4
```