Development Statistics

SO1 Population & Sample

Fujikawa, Kiyoshi
Nagoya University, GSID

Objective of this class = statistical inference

1. Collect data (or samples)
   - First source
   - Second source
2. Process the data (or samples)
   - Descriptive statistics
   - Make tables or graphs
3. Analyze the data (or samples)
   - Statistical Estimation (推定)
   - Statistical Test （検定)

Mathematical Statistics

- Based on the information of a set of data (or sample) ある標本の情報を基礎にして
  1. To estimate population parameters
     母数を推定すること
  2. To test hypotheses
     仮説を検定をすること

Statistical Estimation

- You are a sales manager in Nagoya
  - You have to make a sales plan in Nagoya.
  - First of all, you want to know the size of the market (demands) in Nagoya.
  - in other words, you want to predict the relation between “the price and quantity” or “the income and quantity” by region, age, time, and so on.

Statistical Test

- You are a general sales manager
  - You want to know the difference of the consumer's preferences between Osaka and Nagoya.
  - In other words, you want to do a statistical test on the hypothesis that Osaka is same as Nagoya

Population & Sample
母集団と標本

- Population parameter (母数)
  - Fixed number
  - Only the God knows them
  - The objectives of estimation
- Sample statistic (標本統計量)
  - Sample = a part of the population
  - Sample statistic is different in each sampling case
  - Sample statistic = a random variable

Population Mean 平均 (of a Dice)

- Population mean of a dice

$$\mu = \frac{1}{6}\sum_k X_k = 3.5$$

- Population mean

$$\mu = \frac{1}{N}\sum_k X_k$$

Sample Mean 平均

- Sample mean (=average)
- =AVERAGE( )

$$\bar{x} = \frac{1}{n}\sum_k x_k$$

Population Variance 分散 of a Dice

- Population variance of a dice

$$\sigma^2 = \frac{1}{6}\sum_k (X_k - \mu)^2 = 2.9167$$

- Population variance
- =VARP()

$$\sigma^2 = \frac{1}{N}\sum_k (X_k - \mu)^2$$

Two kinds of sample variance

- Max. likelihood variance(最尤分散)
- =VARP()

$$s^2 = \frac{1}{n}\sum_k (x_k - \bar{x})^2$$

- Unbiased variance(不偏分散)
- =VAR()

$$s^2 = \frac{1}{n-1}\sum_k (x_k - \bar{x})^2$$

Standard deviation
標準偏差 of “Dice”

- Population SD of a dice

$$\sigma = \sqrt{\frac{1}{6}\sum_k (X_k - \mu)^2} = 1.7078$$

- Population SD
- =STDEVP()

$$\sigma = \frac{1}{N}\sqrt{\sum_k (X_k - \mu)^2}$$

Population Skewness 歪度

- Population skewness

$$Skew = \frac{1}{N}\sum_k \left(\frac{x_k - \bar{x}}{\sigma}\right)^3$$

- Meaning of skewness
  - Zeoro : symmetry
  - Positive: right-tail skewed （右に尾が長い）
  - Negative: left-tail skewed （左に尾が長い）
- No function in EXCEL

Sample Skewness 歪度

- Sample skewness
  (Estimator of population skewness)

$$skew = \frac{n}{(n-1)(n-2)} \sum_k \left(\frac{x_k - \bar{x}}{s} \right)^3$$

- =SKEW() in EXCEL

Population kurtosis 尖度

- Population kurtosis

$$Kurt = \frac{1}{N} \sum_k \left(\frac{x_k - \bar{x}}{\sigma} \right)^4 - 3$$

- Meaning of relative kurtosis
  - Zero : same as normal distribution
  - Positive: lepto-kurtic (急尖的)
  - Negative: platy-kurtic （緩尖的)
- No function in EXCEL

Sample kurtosis 尖度

- Sample kurtosis
  (Estimator of population kurtosis)

$$kurt = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_k \left(\frac{x_k - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

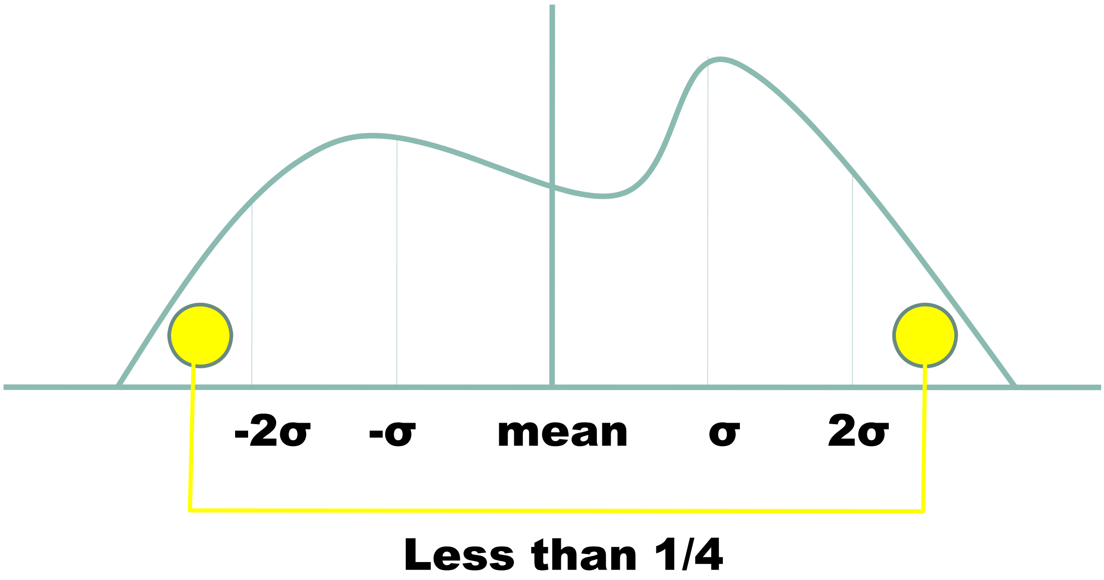
  - =KURT() in EXCEL

Chebyshev's theory

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- k=2
  - 3 / 4 of the random values are in the rage of "mean ±2σ"
- k=3
  - 8 / 9 of the random values are in the rage of "mean ±3σ"

Chebyshev's theory



Appendix: Desirable character of sample statistic

- Unbiasedness ← unbiased variance
  - If you take samples many times and each time you calculate a sample parameter
  - You can make a distribution of the parameter
  - The mean of this distribution = true parameter
- Consistency ← maximum likelihood variance
  - The larger the size is, the sample parameter becomes closer to the true parameter