UNIVERSITAS GADJAH MADA
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
BACHELOR IN CIVIL ENGINEERING

Statistics and Probability

Statistical Measures

Statistical measures

- Common statistical measures
  - Measure of central tendency
    - Mean
    - Mode
    - Median
  - Measure of variability
    - Range
    - Variance
    - Standard deviation
  - Measure of an individual in a population
    - z score
    - Percentile rank

Measure of central tendency

- Average
  - Mean
  - Mode
    - the highest frequency score
  - Median
    - score at the middle of a sorted data

Measure of central tendency

Example

Number of rainy days in the last eleven months:
21, 21, 21, 20, 18, 16, 12, 12, 6, 2, 1

mean	14	=AVERAGE(...)
mode	21	=MODE(...)
median	16	=MEDIAN(...)

Microsoft Excel

Which one is best to represent the number of rainy days in the last eleven months?

Example

- Give examples on statistical measures in civil engineering
  - number of bridges in a city
  - climatological data
  - river discharges
  - etc.
- Discuss on their statistical measures
  - mean
  - mode
  - median

Measure of central tendency

- Symbol and formula
  - Mean

$$\mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

mean of population

n = number of elements in the **population**

statistical parameter: based on population

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

sample mean

n = number of elements in the **sample**

statistical measure: based on sample

⇨ estimate on its population mean

Measure of central tendency

- Some mean characteristics

$$C\bar{X} = \frac{1}{n}\sum CX \qquad C + \bar{X} = \frac{1}{n}\sum (C + X) \qquad C = \text{constant}$$

- Weighted mean

$$\bar{X} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i}^{n} w_i} \Rightarrow$$

  - mean of class intervals
  - for example, in a frequency table of continuous variable
  - X_i is the median of the class interval

Measure of central tendency

- Mean
  - Arithmetic mean $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$ =AVERAGE(...)
  - Geometric mean $$\bar{X} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$ =GEOMEAN(...)
  - Harmonic mean $$\bar{X} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$ =HARMEAN(...)

Measure of central tendency

- Mode

$$\text{Mode} = M_o = L_{mo} + \frac{a}{a+b}c$$

mode interval is the class interval whose frequency is the highest

L_{mo}	lower bound of the mode interval (the class interval containing the mode)
a	frequency difference of the mode interval and the one that is less than it
b	frequency difference of the mode interval and the one that is greater than it
f_{md}	frequency of the median interval
c	class interval (class size)

Measure of central tendency

- Median

$$\text{Median} = M_d = L_{md} + \frac{n/2 - F}{f_{md}} c$$

interval median is the class interval containing the median, that is the $n/2$-th class of a sorted class according to the score

L_{md}	lower bound of the median interval (the class interval containing the median)
n	number of data
F	number of frequencies of class intervals less than the median interval
f_{md}	frequency of the median interval
c	class interval (class size)

Measure of variability

- Variability
  - Variability, scatter, spread
    - showing whether scores in a distribution are close each other or far separated
  - Range
    - difference between the highest and the lowest scores in a distribution
  - Standard deviation
    - commonly used in “technical” matters

Measure of Variability

- Symbol and formula
  - Standard deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

standard deviation of population =STDEV.P(...)

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}}$$

standard deviation of sample =STDEV.S(...)

estimate on its population standard deviation

Measure of variability

- Why do we put $n - 1$ as the denominator in calculating sample's standard deviation?
  - Results in higher standard deviation than if it is divided by n
    - this is to compensate the tendency of being under estimate of the true standard deviation (population standard deviation)
  - From practical point of view
    - there is no variability of sample having one element

Measure of variability

- Symbol and formula
  - Variance

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

variance of population

=VAR.P(...)

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$$

sample variance
estimate on its population variance

=VAR.S(...)

Measure of variability

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$$

$$s^2 = \frac{\sum_{i=1}^{n} {x_i}^2 - n\bar{X}^2}{n-1}$$

Try to derive this

Measure of variability

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n}({x_i}^2 - 2x_i\bar{X} + \bar{X}^2)}{n-1}$$

$$= \frac{\sum_{i=1}^{n}{x_i}^2 - 2\bar{X}\sum_{i=1}^{n}x_i + n\bar{X}^2}{n-1} = \frac{\sum_{i=1}^{n}{x_i}^2 - 2\frac{\sum_{i=1}^{n}x_i}{n}\sum_{i=1}^{n}x_i + n\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n}{x_i}^2 - \frac{(\sum_{i=1}^{n}x_i)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n}{x_i}^2 - n\bar{X}^2}{n-1}$$

Measure of variability

- Symbol and formula
  - coefficient of variance

$$c_v = \frac{\sigma}{\mu}$$

$$c_v = \frac{s_X}{\bar{X}}$$

Some measures of an individual in a population

- z score

$$z_X = \frac{x - \mu}{\sigma}$$

- Percentile rank

$$PR_X = \frac{B + \frac{1}{2}E}{n} \times 100$$

B number of scores whose values are less than that of X
E number of scores whose values are equal to that of X
n number of scores, $n \gg$

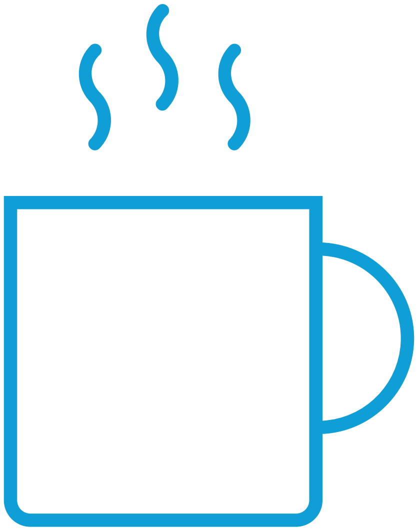
Some measures of an individual in a population

- Some functions available in Microsoft Excel
  - =RANK(...), =RANK.EQ(...), RANK.AVR(...)
    - score location in a sorted data
  - =PERCENTILE(...), =PERCENTILE.EXC(...), =PERCENTILE.INC(...)
    - percentile value of a series of scores
  - =PERCENTRANK(...), =PERCENTRANK.EXC(...), =PERCENTRANK.INC(...)
    - score location in a sorted data, in percent

$$= \frac{B}{B + A} \times 100$$

B number of scores whose values are less than that of X

A number of scores whose values are higher than that of X



Statistics and Probability

Statistical Measures