Outline

1. Introduction
2. Data Summary

Overview

- Life in e-world: SWOT
  - Numericalized/Digital world: Determinism and Stochasticism
  - Decisions and choices: what are they based upon?
- Examples
  - How many studying hours does it take for a math student to survive?
  -

See Better, Look Better

Guess is human nature;
Statistics is human nurture

Everybody has some ability to predict and estimate. Statistics enhances and sharpens this ability with stat/comp powers.

Why statistics? Alternatives?

Questions to be answered, the way to be answered, the way of formulating the problem.

Conce.165 176.103 Td[(Why)-334(statistics)-1(?)-444(Alterna)1(t

Intro to DS

$x_1, \cdots, x_n$ vs $X_1, \cdots, X_n$

- Why bother?
  - Complete data is hard to understand and usually noninformative.
  - Data compression: Small and Useful. Few and informative
  - Example: MP3

Intro to DS

$x_1, \cdots, x_n$ vs $X_1, \cdots, X_n$

- Why bother?
  - Complete data is hard to understand and usually noninformative.
  - Data compression: Small and Useful. Few and informative
  - Example: MP3
- What are we summarizing for?
  - Trend or randomness
  - "Distribution": central tendency, variation, skewness, extreme values, etc.
  - Example: Monthly pocket money of a NDHU undergrad
- How? Numerical summary (Descriptive statistics) and Graphical summary (Stat graphs)

Numerical Summary

- Central Tendency: Mean (average) vs. Median ("The middle one")
- Variation: (sample) standard deviation, IQR=Q3-Q1, Range=Max-Min
- Easily calculable from R

Remark: Q1: middle of lower half; Q3: middle of upper half

Numerical Summary

- Central Tendency: Mean (average) vs. Median (“The middle one”)
- Variation: (sample) standard deviation (SD), Interquartile Range (IQR)=Q3-Q1, Range=Max-Min
- Relative frequencies
- Easily calculable from R

Remark: Q1: middle of lower half; Q3: middle of upper half

- HW1: Write down “possible” definitions of Median and Q3. Explain briefly why you define them so.
- HW2: Give examples to illustrate that
  - Mean is more sensitive to outliers than median
  - SD is more sensitive to outliers than IQR
  - Definitions: Q3, Q2, Q1, Median, IQR, $SD = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$

Graphical Displays

- Stem-and-leaf plot
- Box plot
- Histogram
- Time plot (for observations over time)
- Easily constructable from `R`
- `http://en.wikipedia.org/wiki/Category:`

Next Step

-