213L3: Treatment of Data
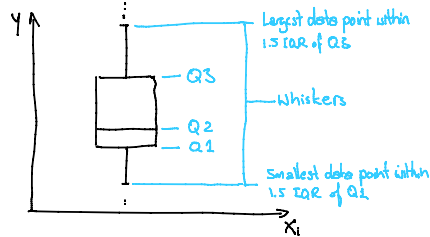
June 12, 2020   11:42 AM

Outline:
1. Data Summary & Display
2. Random Sampling
3. Stem-and-Leaf Diagram
4. Frequency Distributions & Histograms
5. Box Plots
6. Time Sequence Plots
7. Probability Plots

① ②

© Statistic
function of observations in a random sample

© Random Sampling
a) independent random variables X_i
b) same probability distribution $\forall X_i$

"Subset" of observations

© Sample Mean
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Reasonable Estimate

totality of observations

© Population Mean
$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

© Sample Range $r = \max(x_i) - \min(x_i)$

© Sample Variance
$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

© Population Variance
$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

© Sample Standard Deviation
$\sqrt{S^2}$

⑤

© Box Plots

Largest data point within 1.5 IQR of Q3

— Q3

— Whiskers

— Q2
— Q1

Smallest data point within 1.5 IQR of Q1

③ ⑥

Leading digits   trailing digits

© Stem & Leaf   frequency

1	x	1
2	x x x	3
3	x x x x x	5
4	x x x	3
5	x	1

© Median:
Measure of central tendency that divides data into 2 equal parts

Quartiles:
1: $(n+1)/4$
2: Median
3: $3(n+1)/4$

© Interquartile Range
Difference between upper/lower quartiles

© Range:
Measure of variability (Max - Min)

© Time Series/Sequence
Data set recorded in order of occurance

Leaf	Stem
	x
	x
	x
	x
	x

time
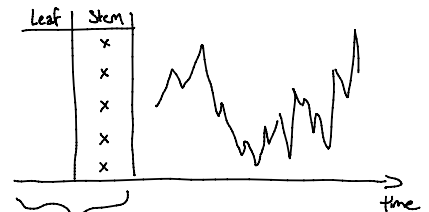
④ ⑦

© Frequency Distribution
- Compact data summary
- Divided into class intervals

frequency   range   x

© Probability Plots

Typically uses special paper

Visual examination of whether data conforms to hypothesized distribution

y_1 y_2
x_i

Negative/Left-Skew   Symmetric   Positive/Right-Skew

Light-tailed distribution   Heavy-tailed distribution   Right-skew distribution