INTRODUCTION TO PROBABILITY MODELS

Lecture 33

Qi Wang, Department of Statistics

Nov 9, 2018

MEASURES OF SPREAD

- Range
- Variance
- Standard deviation
- p_{th} percentile
- Interquartiles Range(IQR)

RANGE

- Range = max - min

VARIANCE

Variance: based on the difference between each observation and the mean

- Population variance:
  $$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$
- Sample variance:
  $$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

STANDARD DEVIATION

Standard deviation: most commonly used for measuring how far observation are from the mean

- Population version:
  $\sigma = \sqrt{\sigma^2}$
- Sample version:
  $s = \sqrt{s^2}$

p_{th} PERCENTILE

p_{th} percentile: value such that p% of the observation fall at or below it

- Median: $M = 50_{th}$ percentile
- First quartile: $Q_1 = 25_{th}$ percentile
- Third quartile: $Q_3 = 75_{th}$ percentile

HOW TO FIND A PERCENTILE FOR DATA

1. Order the data in increasing order
2. Calculate $i = \frac{np}{100}$, where n is the sample size, p is the percentile
3. - If i is not an integer, round i up to the next integer. Then take the i_{th} value
   - If i is an integer, take an average of the i_{th} and $(i+1)_{th}$ values

Example: -20, 1, 23, 25, 32.5, 33, 67

INTERQUARTILES RANGE(IQR)

- IQR = $Q_3 - Q_1$
- Outliers: an observation is said to be a suspected outlier if it is
  $> Q_3 + 1.5 * IQR$
  OR
  $< Q_1 - 1.5 * IQR$