ILRST/STSCI 2100 Discussion 3: Ch. 4 Numerical Methods

Review summations

- Suppose c is a constant, then $\sum_{i=1}^{n} c = n * c$

  Ex. $\sum_{i=1}^{3} 2 = 2 + 2 + 2 = 3 * 2 = 6$

- $\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$

  Ex. Suppose $a = \{1, 2, 3, 4\}$ and $b = \{5, 6, 7, 8\}$
  $\sum_{i=1}^{4}(a_i + b_i) = \sum_{i=1}^{4} a_i + \sum_{i=1}^{4} b_i = 10 + 26 = 36$

- Similarly, $\sum_{i=1}^{n} a_i - b_i = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n} b_i$

- **Note:** $\sum a_i b_i \neq \sum a_i \sum b_i$

Mean

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ "sum all the values and divide by the number of values you have"

Ex. $x = \{3, 1, 8, 4, 6, 3, 8, 3\}$ Find the mean of x.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3 + 1 + 8 + 4 + 6 + 3 + 8 + 3}{8}$$
$$= \frac{36}{8} = 4.5$$

Trimmed Mean

Finding the mean after removing some percent of large values and small values. Make sure data is first sorted. The purpose of trimming is to remove possible outliers
Ex. $x = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ Find the 10% trimmed mean of x.

1. The sample size is 9, as there are 9 values in x.
2. 10% of 9 is 0.9, we round to the nearest integer, so we remove 1 item from the high end and 1 item from the small end.
3. After trimming, the set is {2,3,4,5,6,7,8}
4. Calculate the mean: $\frac{2 + 3 + 4 + 5 + 6 + 7 + 8}{7} = 35/7$

Median

The middle value. The median is a robust statistic, meaning it is resistant to outliers and changes in small portions data.
This measure is preferred over the mean when there are outliers or skewed data.
Ex. For $x = \{1, 2, 3, 4, 5\}$, the median is 3.

For $x = \{1, 2, 3, 4, 5, 6\}$, the median is the average of the two "middle terms", 3.5

Quartiles
Divide the data into fourths, Q1 = 25th percentile, Q2 = median = 50th percentile, Q3 = 75th percentile
IQR = Q3 – Q1
Mild outlier: falls outside the range of $Q1 - 1.5 * IQR$ and $Q3 + 1.5 * IQR$
Extreme outlier: falls outside the range of $Q1 - 3 * IQR$ and $Q3 + 3 * IQR$

Range:
Measure of variation, max – min

Standard Deviation:
$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

The sample variance is denoted as s^2
Note that $s_{xx} = \sum(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$
Also note: $\sum_i x_i^2 \neq (\sum_i x_i)^2$

z-score:
$$\text{z score} = \frac{\text{value} - \text{mean}}{\text{standarddeviation}}$$

Calculating z-scores is also called standardization

Parameter vs Statistic:
A parameter is a value about the population, and a statistic is a value about the sample.

- Mean: μ vs $\bar{x}$
- Variance: σ^2 vs s^2
- Proportion: p vs $\hat{p}$