

**STAT381****L.Liu****Sections 1.1-1.4****List of topics: Describing data****1. Population vs. sample**2. Given a set of 1-d data:  $x_1, x_2, \dots, x_n$ 

- **Histogram:** frequency, relative frequency

Rule of thumb: number of classes  $\approx \sqrt{n}$ 

- Describe center: **mean(average), median, mode**

- Describe dispersion:

**population variance**  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ ,**sample variance**  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ ,**population standard deviation**  $\sigma$ ,**sample standard deviation**  $s$ ,

- **Stem-and-leaf plot:** To arrange data

- **5 number summary:** min,  $Q_1$ , Med,  $Q_3$ , max

- **Inter-quartile range** =  $Q_3 - Q_1$

- **Boxplot**

- How to use TI-83/84 to compute and display data?

## 1.5

### List of topics:

1. Always draw scatter plot! Look for direction, form, strength and outlier.
2. Residual=data-model:  $e_i = y_i - \hat{y}_i$
3. Linear regression:  $\hat{y} = a+bx$ , where  $b = r\frac{s_y}{s_x}$ ,  $a = \bar{y}-b\bar{x}$
4. Correlation coefficient:  $r = \frac{\sum z_x z_y}{n - 1} =$   
$$= \frac{1}{n - 1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$
5. What can go wrong?
  - Don't fit a straight line to a nonlinear relationship;
  - Be aware of the effect of outliers, leverage, and influence;
  - Don't invert regression;
  - Be careful with extrapolation;
  - Don't infer that x causes y;