

One-variable data

- ✓ **Mean:** $\bar{x} = \frac{\sum x_i}{n}$
- ✓ **Median:** the middle value when the data are ordered
- ✓ **Range:** $\max - \min$
- ✓ **Interquartile range:** $IQR = Q_3 - Q_1$
- ✓ **Standard deviation:** $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$
- ✓ **Variance:** $s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$
- ✓ **Quartiles:**
 - Q_1 = 25th percentile
 - Q_2 = 50th percentile (the median)
 - Q_3 = 75th percentile
- ✓ **Quantitative (numerical) variables:** numerical measures
- ✓ **Qualitative (categorical) variables:** categories

Two-variable data

- ✓ **Correlation coefficient:** $r = \frac{1}{n - 1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$, where $-1 \leq r \leq 1$
- ✓ **Regression line:** $\hat{y} = a + bx$, where $b = r \frac{s_y}{s_x}$ and $a = \bar{y} - b\bar{x}$
- ✓ **Residual:** $e_i = y_i - \hat{y}_i$
- ✓ **Standard deviation of the residuals:** $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}}$

Data collection

- ✓ **Simple random sampling:** e.g., a random number generator or drawing names from a hat
- ✓ **Stratified random sampling:** divide the population into homogeneous groups (strata), then take a simple random sample from each, so every group is represented
- ✓ **Cluster sampling:** divide the population into heterogeneous groups (clusters), randomly select clusters, and survey everyone in them
- ✓ **Systematic sampling:** start at a random place and select every n th individual
- ✓ **Nonresponse bias:** people who don't respond may differ from those who do
- ✓ **Response bias:** questions answered inaccurately or untruthfully
- ✓ **Undercoverage bias:** some groups in the population are left out of the sample or underrepresented in it