

Variable data and data comparison

AP STATISTICS DUMP SHEET



One variable data:

- Mean: $\bar{x} = \frac{\sum x_i}{n}$
- Median: the middle value when the data are ordered.
- Mean: $max - min$
- Interquartile range (IQR): $IQR = Q_3 - Q_1$
- Standard deviation: $S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$
- Variance: $S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$
 - $Q_1 = 25\text{th Percentile}$
 - $Q_2 = 50\text{th Percentile (median)}$
 - $Q_3 = 75\text{th Percentile}$
- Variance: numerical measures
- Qualitative/Categorical: categories

Two variable data:

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

Data collection:

- Simple random sampling:** random number generator or drawing names from a hat.
- Stratified random sampling:** dividing the population into homogeneous groups (strata) and then simple random sampling from each group (stratum) to ensure representation from all groups.
- Cluster sampling:** divide the population into heterogeneous groups (clusters) and randomly then randomly select clusters and sample all in chosen clusters.
- Systematic sampling:** start from a random place and select every n th individual.
- Nonresponse bias:** people who don't respond to the survey may be different from those who responded.
- Response bias:** questions answered inaccurately or untruthfully
- Undercoverage bias:** some groups in the population are not included or are underrepresented in the sample.