Chapter 3: Numerical Summaries of Data

Sample mean:
\[ \bar{x} = \frac{\sum x}{n} \]

Population mean:
\[ \mu = \frac{\sum x}{N} \]

Range:
Range = largest value − smallest value

Population variance:
\[ \sigma^2 = \frac{\sum (x - \mu)^2}{N} \]

Sample variance:
\[ s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \]

Coefficient of variation:
\[ CV = \frac{\sigma}{\mu} \]

z-score:
\[ z = \frac{x - \mu}{\sigma} \]

Interquartile range:
IQR = \[Q_3 - Q_1 = \text{third quartile} - \text{first quartile}\]

Lower outlier boundary:
\[ Q_1 - 1.5 \cdot \text{IQR} \]

Upper outlier boundary:
\[ Q_3 + 1.5 \cdot \text{IQR} \]

Chapter 4: Probability

General Addition Rule:
\[ P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \]

Multiplication Rule for Independent Events:
\[ P(A \text{ and } B) = P(A)P(B) \]

Addition Rule for Mutually Exclusive Events:
\[ P(A \text{ or } B) = P(A) + P(B) \]

Rule of Complements:
\[ P(A^c) = 1 - P(A) \]

General Method for Computing Conditional Probability:
\[ P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)} \]

General Multiplication Rule:
\[ P(A \text{ and } B) = P(A) P(B \mid A) = P(B) P(A \mid B) \]

Permutation of \( r \) items chosen from \( n \):
\[ n^P_r = \frac{n!}{(n - r)!} \]

Combination of \( r \) items chosen from \( n \):
\[ n^C_r = \frac{n!}{r!(n - r)!} \]
Chapter 5: Discrete Probability Distributions

Mean of a discrete random variable:
\[ \mu_X = \sum [x \cdot P(x)] \]

Variance of a discrete random variable:
\[ \sigma_X^2 = \sum [(x - \mu_X)^2 \cdot P(x)] = \sum [x^2 \cdot P(x)] - \mu_X^2 \]

Standard deviation of a discrete random variable:
\[ \sigma_X = \sqrt{\sigma_X^2} \]

Mean of a binomial random variable:
\[ \mu_X = np \]

Variance of a binomial random variable:
\[ \sigma_X^2 = np(1 - p) \]

Standard deviation of a binomial random variable:
\[ \sigma_X = \sqrt{np(1 - p)} \]

Chapter 6: The Normal Distribution

z-score:
\[ z = \frac{x - \mu}{\sigma} \]

Convert z-score to raw score:
\[ x = \mu + z\sigma \]

Standard deviation of the sample mean:
\[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \]

z-score for a sample mean:
\[ z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \]

Standard deviation of the sample proportion:
\[ \sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}} \]

z-score for a sample proportion:
\[ z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} \]

Chapter 7: Confidence Intervals

Confidence interval for a mean, standard deviation known:
\[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \]

Sample size to construct an interval for \( \mu \) with margin of error \( m \):
\[ n = \left( \frac{z_{\alpha/2} \cdot \sigma}{m} \right)^2 \]

Confidence interval for a mean, standard deviation unknown:
\[ \bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \]

Confidence interval for a proportion:
\[ \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]

Sample size to construct an interval for \( p \) with margin of error \( m \):
\[ n = \frac{\hat{p}(1 - \hat{p})}{m^2} \]

if a value for \( \hat{p} \) is available

\[ n = 0.25 \left( \frac{z_{\alpha/2}}{m} \right)^2 \]

if no value for \( \hat{p} \) is available
Chapter 8: Hypothesis Testing

Test statistic for a mean, standard deviation known:
\[ z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \]

Test statistic for a mean, standard deviation unknown:
\[ t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \]

Test statistic for a proportion:
\[ z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \]

Chapter 9: Inferences on Two Samples

Test statistic for the difference between two means, independent samples:
\[ t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2_1/n_1 + s^2_2/n_2}} \]

Confidence interval for the difference between two means, independent samples:
\[ \bar{x}_1 - \bar{x}_2 - t_{\alpha/2} \sqrt{s^2_1/n_1 + s^2_2/n_2} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\alpha/2} \sqrt{s^2_1/n_1 + s^2_2/n_2} \]

Test statistic for the difference between two proportions:
\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{(\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2}))}} \]

where \( \hat{p} \) is the pooled proportion \( \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \)

Chapter 10: Tests with Qualitative Data

Chi-square statistic:
\[ \chi^2 = \sum \frac{(O - E)^2}{E} \]

Expected frequency for goodness-of-fit:
\[ E = np \]

Expected frequency for independence or homogeneity:
\[ E = \frac{\text{Row total} \cdot \text{Column total}}{\text{Grand total}} \]
Chapter 11: Correlation and Regression

Correlation coefficient:
\[ r = \frac{1}{n-1} \sum \left( \frac{x - \bar{x}}{s_x} \right) \left( \frac{y - \bar{y}}{s_y} \right) \]

Equation of least-squares regression line:
\[ \hat{y} = b_0 + b_1 x \]

Slope of least-squares regression line:
\[ b_1 = \frac{s_y}{s_x} \]

y-intercept of least-squares regression line:
\[ b_0 = \bar{y} - b_1 \bar{x} \]

Residual standard deviation:
\[ s_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}} \]

Standard error for \( b_1 \):
\[ s_{b_1} = \frac{s_e}{\sqrt{\sum (x - \bar{x})^2}} \]

Confidence interval for slope:
\[ b_1 - t_{(n-2, \alpha/2)} \cdot s_{b_1} < b_1 < b_1 + t_{(n-2, \alpha/2)} \cdot s_{b_1} \]

Test statistic for slope \( b_1 \):
\[ t = \frac{b_1}{s_{b_1}} \]

Confidence interval for the mean response:
\[ \hat{y} \pm t_{(n-2, \alpha/2)} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2}} \]

Prediction interval for an individual response:
\[ \hat{y} \pm t_{(n-2, \alpha/2)} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2}} \]