

I. Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

II. Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

III. Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: $\text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$
Chi-square statistic: $\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

*Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

3. Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.

Study Finds Strong Association Between Smoking and Alzheimer's

The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

- (a) Identify the explanatory and response variables in the study.

Explanatory variable:

Response variable:

- (b) Is the study described in the article an observational study or an experiment? Explain.
- (c) Exercise status (regular weekly exercise versus no regular weekly exercise) was mentioned in the article as a possible confounding variable. Explain how exercise status could be a confounding variable in the study.