Covered chapters

- 2.1 - 2.7, 2.9
- 3.1 - 3.3, 3.5, 3.6
- 4.1 - 4.7
- 5.1, 5.2, 5.4, 5.6
- 6.1 - 6.4, 6.7
- 7.1 - 7.3
- 8.1 - 8.2, 8.4
- 9.1 - 9.3, 9.5
- 10.2
- 11.1 - 11.7
- 13.2

Probability Theory

These formulas deal with the fundamental concepts of probability.

Addition Rule for Two Events

- **Formula:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Explanation:** This formula calculates the probability of event A **or** event B occurring. It's the sum of their individual probabilities minus the probability of both A **and** B occurring, to avoid double-counting their intersection.
- **When to use:** When you want to find the probability that at least one of two events occurs. If events A and B are mutually exclusive (they cannot both happen at the same time), then $P(A \cap B) = 0$, and the formula simplifies to $P(A \cup B) = P(A) + P(B)$.

Addition Rule for Three Events

- **Formula:** $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
- **Explanation:** This extends the addition rule to three events. It accounts for the probabilities of single events, subtracts the probabilities of pairwise intersections, and then adds back the probability of the intersection of all three events to correct for over-subtraction.
- **When to use:** When you want to find the probability that at least one of three events occurs.

Multiplication Rule (Conditional Probability)

- **Formula:** $P(A \cap B) = P(A \mid B)P(B)$
- **Explanation:** This formula relates the probability of both A and B occurring to the conditional probability of A given B, and the probability of B. $P(A \mid B)$ is the probability of event A happening, given that event B has already happened.
- **When to use:**
  - To find the probability that two events both occur.
  - To calculate conditional probabilities if you know $P(A \cap B)$ and $P(B)$, as $P(A \mid B) = P(A \cap B)/P(B)$.

Law of Total Probability and Bayes' Theorem

- **Formulas:**
  - $P(A) = \sum_{i=1}^{p} P(A \mid B_i) P(B_i)$
  - $P(B_1 \mid A) = \frac{P(A|B_1)P(B_1)}{\sum P(A|B_i)P(B_i)}$
- **Explanation:**
  - **Law of Total Probability:** If $B_1, \ldots, B_p$ form a partition of the sample space (meaning they are mutually exclusive and collectively exhaustive), this formula states that the probability of event A can be found by summing the probabilities of A occurring under each of the B_i conditions, weighted by the probability of each B_i.
  - **Bayes' Theorem:** This formula allows you to update the probability of an event (B_1) given new evidence (event A). It calculates the posterior probability $P(B_1 \mid A)$ using the prior probability $P(B_1)$ and the likelihood $P(A \mid B_1)$, normalized by the total probability of A.
- **When to use:**
  - **Law of Total Probability:** When you need to find the overall probability of an event (A) that can occur under several different, mutually exclusive scenarios (B_i).
  - **Bayes' Theorem:** When you want to reverse the conditioning – to find the probability of a cause (B_1) given an observed effect (A), especially useful in medical diagnosis, spam filtering, and machine learning.

Random Variables and Distributions

These formulas define properties and approximations for random variables.

Cumulative Distribution Function (CDF)

- **Formula:** $F(x) = P(X \leq x) = \begin{cases} \sum_{u \leq x} f(u) & \text{discrete} \\ \int_{-\infty}^{x} f(u)du & \text{continuous} \end{cases}$
- **Explanation:** The CDF gives the probability that a random variable X takes on a value less than or equal to x.
  - For **discrete** random variables, it's the sum of the probability mass function (pmf) for all values u less than or equal to x.
  - For **continuous** random variables, it's the integral of the probability density function (pdf) from negative infinity up to x.
- **When to use:** To find probabilities of the form $P(X \leq x)$, $P(X > x) = 1 - P(X \leq x)$, or $P(a < X \leq b) = F(b) - F(a)$.

Expectation (Mean) and Variance

- **Formulas:**
  - $\mu = \begin{cases} \sum_x xf(x) & \text{discrete} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{continuous} \end{cases}$
  - $\sigma^2 = \begin{cases} \sum_x x^2 f(x) - \mu^2 & \text{discrete} \\ \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 & \text{continuous} \end{cases}$
- **Explanation:**
  - μ (expectation or mean): The average value of a random variable. It's a measure of central tendency.
  - σ^2 (variance): A measure of the spread or dispersion of a random variable around its mean. A larger variance means the values are more spread out. The second formula for variance is often called the "computational formula" and can be easier to use.
- **When to use:**
  - To calculate the expected value (average) of a random variable.
  - To quantify the variability of a random variable.

Normal Approximation to the Binomial Distribution

- **Formula:** If $X \sim \text{Binomial}(n, p)$ for large n, then $P(X = x) \approx \Phi\left(\frac{x+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x-0.5-np}{\sqrt{np(1-p)}}\right)$ where $\Phi(z)$ is the cdf of the $\text{Normal}(0, 1)$.
- **Explanation:** For a large number of trials (n), the discrete Binomial distribution can be approximated by the continuous Normal distribution. The 0.5 adjustments are called continuity corrections, which are necessary when approximating a discrete distribution with a continuous one. np is the mean and $np(1-p)$ is the variance of the Binomial distribution.
- **When to use:** When calculating probabilities for a Binomial random variable with a large n (typically $np \geq 5$ and $n(1-p) \geq 5$), as calculating individual binomial probabilities can become computationally intensive. This approximation uses the standard normal CDF (Φ) to find the probability of $X = x$. For $P(X \leq x)$, you would use $\Phi\left(\frac{x+0.5-np}{\sqrt{np(1-p)}}\right)$.

Expected Value of a Function of Two Random Variables

- **Formula:** $E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) f_{X,Y}(x, y) & \text{discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f_{X,Y}(x, y)dxdy & \text{continuous} \end{cases}$
- **Explanation:** This formula calculates the expected value of a function $h(X, Y)$ of two random variables X and Y. It's a weighted average of the function's values, where the weights are the joint probabilities (for discrete) or joint probability densities (for continuous).
- **When to use:** To find the average value of a quantity that depends on two random variables, such as $E[XY]$ or $E[X + Y]$.

Marginal Probability Distribution

- **Formula:** $f_X(x) = \begin{cases} \sum_y f_{X,Y}(x, y) & \text{discrete} \\ \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy & \text{continuous} \end{cases}$
- **Explanation:** The marginal probability distribution of X (or Y) is the probability distribution of a single random variable in a joint distribution. It is obtained by "summing out" (for discrete) or "integrating out" (for continuous) the other variable from the joint probability function.
- **When to use:** To find the probability distribution of one random variable when you are given their joint probability distribution.

Conditional Probability Distribution

- **Formula:** $f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
- **Explanation:** This formula defines the conditional probability distribution of Y given a specific value of X. It's the ratio of the joint probability distribution of X and Y to the marginal probability distribution of X. It tells you how Y behaves when X is fixed at a certain value.
- **When to use:** To understand the relationship between two random variables and how the probability of one changes given knowledge of the other.

Correlation Coefficient

- **Formula:** $\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{E(XY)-\mu_X\mu_Y}{\sqrt{V(X)V(Y)}}$
- **Explanation:** The correlation coefficient $\rho_{X,Y}$ measures the strength and direction of the linear relationship between two random variables X and Y. It ranges from -1 to 1.
  - $\rho_{X,Y} = 1$: Perfect positive linear relationship.
  - $\rho_{X,Y} = -1$: Perfect negative linear relationship.
  - $\rho_{X,Y} = 0$: No *linear* relationship (though they might have a non-linear relationship).

  - $\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y$ is the covariance, which indicates the direction of the linear relationship, but its magnitude is affected by the scales of X and Y. The correlation coefficient normalizes this by dividing by the product of the standard deviations.
- **When to use:** To quantify the linear association between two random variables.

Independent Variables Properties

- **Formula:** If X and Y are independent then $E[h(X)g(Y)] = E[h(X)]E[g(Y)]$ and $\rho_{X,Y} = 0$.
- **Explanation:** If two random variables are independent, knowing the value of one provides no information about the value of the other.
  - The expected value of the product of functions of independent variables is the product of their individual expected values.
  - Their correlation coefficient is 0 (no linear relationship). **Important:** $\rho_{X,Y} = 0$ only implies no *linear* relationship, not necessarily independence. However, if X and Y are independent, then $\rho_{X,Y} = 0$.
- **When to use:** When dealing with independent random variables, these properties simplify calculations involving expectations and allow you to conclude no linear correlation.

Expectation and Variance of a Linear Combination of Random Variables

- **Formulas:**
  - If $Y = c_0 + c_1 X_1 + \cdots + c_p X_p$ then $E(Y) = c_0 + \sum_1^p c_i \mu_i$
  - $\text{Var}(Y) = \sum_1^p c_i^2 \sigma_i^2 + 2 \sum \sum_{i<j} c_i c_j \,\text{Cov}\,(X_i, X_j)$
- **Explanation:**
  - **Expectation:** The expected value of a linear combination of random variables is the linear combination of their individual expected values. The constant c_0 is simply added.
  - **Variance:** The variance of a linear combination of random variables involves the sum of the squared coefficients times their individual variances, plus twice the sum of products of coefficients and covariances for all unique pairs of variables. If the variables are independent, all $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$, simplifying the formula.
- **When to use:** To calculate the mean and variance of a new random variable that is formed by a weighted sum of other random variables. This is very common in portfolio theory, error propagation, and statistical modeling.

Linear Combination of Independent Normal Random Variables

- **Formula:** If $X_1, \ldots, X_n$ are independent with $X_i \sim \text{Normal}\left(\mu_i, \sigma_i^2\right)$ then $Y = c_0 + c_1 X_1 + \cdots + c_n X_n$ is normal with $\mu_Y = c_0 + \sum_1^n c_i \mu_i$ and $\sigma_Y^2 = \sum_1^n c_i^2 \sigma_i^2$.
- **Explanation:** A special and very useful property of normal distributions: any linear combination of independent normal random variables is itself normally distributed. The mean and variance follow the general rules for linear combinations, with the covariance terms dropping out due to independence.
- **When to use:** When working with sums or differences of independent normally distributed quantities. This property is crucial for many statistical tests and confidence intervals that rely on normal theory.

Descriptive Statistics

These formulas are used to summarize and describe sample data.

Sample Mean and Sample Variance

- **Formulas:**
  - $\bar{x} = \frac{1}{n} \sum x_i$
  - $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum x_i^2 - n\bar{x}^2\right)$
- **Explanation:**
  - **Sample Mean** ($\bar{x}$): The average of a set of observed data points. It is an estimator for the population mean (μ).
  - **Sample Variance** (s^2): A measure of the dispersion of observed data points around the sample mean. The $(n-1)$ in the denominator (Bessel's correction) makes s^2 an unbiased estimator of the population variance (σ^2). The second form is a computational formula that can be easier to calculate by hand.
- **When to use:** To calculate descriptive statistics for a sample of data:
  - $\bar{x}$ for the central tendency.
  - s^2 for the variability.
  - $s = \sqrt{s^2}$ for the sample standard deviation.

Central Limit Theorem

- **Formula:** $X_1, \ldots, X_n$ independent with $E\left(X_i\right) = \mu$ and $V\left(X_i\right) = \sigma^2 < \infty$ then the distribution of $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ approaches a standard normal as $n \to \infty$.
- **Explanation:** The Central Limit Theorem (CLT) is a cornerstone of statistics. It states that, regardless of the original distribution of the random variables X_i, the distribution of the sample mean ($\bar{X}$) will approach a normal distribution as the sample size (n) becomes large. The standardized sample mean Z will approach a standard normal distribution ($\text{Normal}(0, 1)$).
- **When to use:**

- To justify using normal distribution theory for confidence intervals and hypothesis tests involving sample means, even if the underlying population distribution is not normal (provided n is sufficiently large, typically $n \geq 30$).
- Understanding why many natural phenomena that are the result of many small, independent effects tend to be normally distributed.

Confidence Intervals and Test Statistics

These sections provide formulas for constructing confidence intervals and calculating test statistics for various hypothesis tests. The choice of formula depends on the parameters being estimated/tested, assumptions about the population, and sample size.

General Notation

- $\sim$ means "is distributed as"
- $\dot\sim$ means "is approximately distributed as"
- $z_{\alpha/2}$ and z_α: Z-scores from the standard normal distribution corresponding to a given significance level α.
- $t_{\nu;\alpha/2}$ and t_ν: T-scores from the t-distribution with ν degrees of freedom corresponding to a given significance level α.

Normal Mean, σ^2 Known

- **Confidence Interval (CI):** $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$
- **Test Statistic:** $Z_0 = \frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{\sigma} \sim \text{Normal}(0,1)$
- **When to use:** When you are working with a sample mean from a **normally distributed population**, and you **know the population standard deviation** (σ). This is a rarer situation in practice, but forms the theoretical basis for other cases.

Non-Normal Mean, σ^2 Known, Large Sample ($n \geq 30$)

- **Confidence Interval (CI):** $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$
- **Test Statistic:** $Z_0 = \frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{\sigma} \dot\sim \text{Normal}(0,1)$
- **When to use:** When you have a **large sample** (typically $n \geq 30$), know the **population standard deviation** (**σ disputes assumption that X approx normal for large n**), and the population may **not be normally distributed**. The Central Limit Theorem allows us to use the normal approximation here.

Large Sample Mean, σ^2 Unknown

- **Confidence Interval (CI):** $\bar{x} \pm z_{\alpha/2}s/\sqrt{n}$
- **Test Statistic:** $Z_0 = \frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{S} \dot\sim \text{Normal}(0,1)$
- **When to use:** When you have a **large sample** (typically $n \geq 30$), and the **population standard deviation** (σ) **is unknown**, so you use the sample standard deviation (s) as an estimate. Due to the large sample size, we can still approximate the distribution of the test statistic as normal.

Normal Mean, σ^2 Unknown

- **Confidence Interval (CI):** $\bar{x} \pm t_{n-1;\alpha/2}s/\sqrt{n}$
- **Test Statistic:** $T_0 = \frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{S} \sim t_{n-1}$
- **When to use:** When you have a sample from a **normally distributed population**, and the **population standard deviation** (σ) **is unknown**, regardless of sample size. In this case, we use the t-distribution with $n-1$ degrees of freedom, which accounts for the extra variability introduced by estimating σ with s. This is a very common scenario.

Proportion, p (Large Sample)

- **Confidence Interval (CI):** $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$
- **Test Statistic:** $Z_0 = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \dot\sim \text{Normal}(0,1)$
- **When to use:**
  - **CI:** To estimate a **population proportion** (p) based on a sample proportion ($\hat{p}$) when the sample size is large enough (e.g., $n\hat{p} \geq 5$ and $n(1-\hat{p}) \geq 5$).
  - **Test Statistic:** To test a hypothesized value of a population proportion (p_0) using a large sample. Note that for the test statistic, p_0 is used in the denominator, while for the confidence interval, $\hat{p}$ is used.

Sample Size Determination

These formulas help determine the required sample size to achieve a desired level of precision or power for a hypothesis test.

Normal Mean, σ^2 Known (for Margin of Error)

- **Formula:** $n \geq \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$
- **Explanation:** This formula calculates the minimum sample size (n) required to estimate the population mean (μ) with a specified margin of error (E) and confidence level ($1-\alpha$), assuming the population standard deviation (σ) is known.

- **When to use:** When designing a study and you need to determine how many samples to collect to achieve a specific precision for a mean estimate, and you have some prior knowledge or estimate of σ.

Proportion, p (for Margin of Error)

- **Formula:** $n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p^* \left(1 - p^*\right)$
- **Explanation:** This formula calculates the minimum sample size (n) required to estimate a population proportion (p) with a specified margin of error (E) and confidence level $(1 - \alpha)$. p^* is a prior estimate of the population proportion. If no prior estimate is available, using $p^* = 0.5$ will yield the largest conservative sample size.
- **When to use:** When designing a study to estimate a proportion with a certain level of precision.

Normal Mean, μ (for Specified Error Probabilities)

- **Formulas:**
  - 2-tailed: $n \geq \left(\frac{(z_{\alpha/2}+z_\beta)\sigma}{\mu-\mu_0}\right)^2$
  - 1-tailed: $n \geq \left(\frac{(z_\alpha+z_\beta)\sigma}{\mu-\mu_0}\right)^2$
- **Explanation:** These formulas determine the sample size needed to achieve specified Type I error probability (α) and Type II error probability (β) for a hypothesis test about a normal mean. μ_0 is the hypothesized mean under the null hypothesis, and μ is the true mean under the alternative hypothesis. z_β is the Z-score corresponding to the Type II error probability β.
- **When to use:** When planning a hypothesis test for a mean, and you want to ensure sufficient statistical power to detect a specific difference $(\mu - \mu_0)$ with given error rates.

Two-Sample Confidence Intervals and Test Statistics

These formulas are used to compare the means of two populations.

Difference in Means $(\mu_1 - \mu_2)$, Normal, Common σ^2

- **Confidence Interval (CI):** $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,n_1+n_2-2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
- **Test Statistic:** $T_0 = \frac{\bar{X}_1-\bar{X}_2-\Delta_0}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$ where $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$
- **Explanation:** This is for comparing two population means when both populations are assumed to be normally distributed and have the **same (but unknown) variance** (σ^2). S_p^2 is the pooled sample variance, which is a weighted average of the two sample variances. Δ_0 is the hypothesized difference in means (often 0 for testing equality). The t-distribution with $n_1 + n_2 - 2$ degrees of freedom is used.
- **When to use:** When comparing two means where the normal assumption holds and you believe the population variances are equal (e.g., based on prior knowledge or an F-test for equality of variances).

Difference in Means $(\mu_1 - \mu_2)$, Normal (Unequal Variances)

- **Confidence Interval (CI):** $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- **Test Statistic:** $T_0 = \frac{\bar{X}_1-\bar{X}_2-\Delta_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \dot{\sim} t_\nu$ where $\nu = \frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1}+\frac{\left(s_2^2/n_2\right)^2}{n_2-1}}$
- **Explanation:** This is for comparing two population means when both populations are assumed to be normally distributed but have **unequal (and unknown) variances**. This is often called Welch's t-test. The degrees of freedom (ν) are calculated using a more complex formula (Satterthwaite's approximation), and will generally not be an integer.
- **When to use:** When comparing two means where the normal assumption holds, but you cannot assume equal population variances.

Difference in Means $(\mu_D = \mu_1 - \mu_2)$, Paired Data

- **Confidence Interval (CI):** $\bar{d} \pm t_{\alpha/2,n-1}\frac{s_D}{\sqrt{n}}$
- **Test Statistic:** $\frac{\bar{D}-\Delta_0}{S_D/\sqrt{n}} \sim t_{n-1}$ $(D_i = X_{1i} - X_{2i})$
- **Explanation:** This is for **paired data**, where each observation in one sample is naturally linked to an observation in the other sample (e.g., before-and-after measurements on the same subject, or measurements on identical twins). We calculate the difference (D_i) for each pair and then perform a one-sample t-test on these differences. $\bar{D}$ is the sample mean of the differences, and S_D is the sample standard deviation of the differences.
- **When to use:** When you have paired observations and want to test if there is a significant difference between the two conditions or treatments. This design effectively reduces variability.

Regression Analysis

These formulas are central to simple linear regression, which models the linear relationship between a dependent variable (Y) and an independent variable (X).

Sums of Squares for Regression

- **Formulas:**
  - $S_{xx} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

- $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- $S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$

- **Explanation:** These are fundamental sums of squares used in linear regression:
  - S_{xx}: Sum of squared deviations of x values from their mean, related to the variability of X.
  - S_{xy}: Sum of products of deviations of x and y values from their respective means, related to the covariance between X and Y.
  - S_{yy}: Sum of squared deviations of y values from their mean, related to the total variability of Y.
- **When to use:** These are intermediate calculations necessary for estimating regression coefficients and calculating other regression statistics.

Regression Coefficients

- **Formulas:**
  - $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$
  - $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- **Explanation:** These are the least squares estimates for the slope ($\hat{\beta}_1$) and y-intercept ($\hat{\beta}_0$) of the regression line ($y = \hat{\beta}_0 + \hat{\beta}_1 x$).
  - $\hat{\beta}_1$: Represents the estimated change in Y for a one-unit increase in X.
  - $\hat{\beta}_0$: Represents the estimated value of Y when $X = 0$.
- **When to use:** To find the equation of the best-fit straight line that describes the linear relationship between X and Y in your sample data.

Sums of Squares in Regression (ANOVA context)

- **Formulas:**
  - $\mathrm{SS}_T = \sum (y_i - \bar{y})^2$ (Total Sum of Squares)
  - $\mathrm{SS}_E = \sum (y_i - \hat{y}_i)^2 = S_{yy} - \hat{\beta}_1 S_{xy}$ (Error Sum of Squares or Residual Sum of Squares)
  - $\mathrm{SS}_R = \sum (\hat{y}_i - \bar{y})^2$ (Regression Sum of Squares or Explained Sum of Squares)
- **Explanation:** These sums of squares decompose the total variation in the dependent variable (Y) into parts explained by the regression model and parts due to random error.
  - SS_T: Total variation in Y.
  - SS_E: Variation in Y not explained by the regression model (residuals).
  - SS_R: Variation in Y explained by the regression model.
  - **Relationship:** $\mathrm{SS}_T = \mathrm{SS}_R + \mathrm{SS}_E$.
- **When to use:** To understand how much of the variability in the dependent variable is accounted for by the independent variable. These are used in ANOVA tables for regression.

Estimated Variance of Errors and Coefficient of Determination

- **Formulas:**
  - $\hat{\sigma}^2 = \mathrm{SS}_E/(n-2)$
  - $R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$
- **Explanation:**
  - $\hat{\sigma}^2$: The estimated variance of the random errors (ϵ_i) in the regression model. It's calculated by dividing the error sum of squares by its degrees of freedom ($n-2$, because two parameters, β_0 and β_1, are estimated).
  - R^2 (Coefficient of Determination): Represents the proportion of the total variation in the dependent variable (Y) that is explained by the independent variable (X) in the regression model. It ranges from 0 to 1. A higher R^2 indicates a better fit of the model to the data.
- **When to use:** $\hat{\sigma}^2$ is used for calculating standard errors of regression coefficients and for confidence/prediction intervals. R^2 is used to assess the overall goodness-of-fit of the regression model.

Standard Error and Inference for $\hat{\beta}_1$

- **Formulas:**
  - $\mathrm{se}\left(\hat{\beta}_1\right) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$
  - **CI:** $\hat{\beta}_1 \pm t_{\alpha/2,n-2}\,\mathrm{se}\left(\hat{\beta}_1\right)$
  - **Test Statistic:** $\frac{\hat{\beta}_1 - \beta_{1,0}}{\mathrm{se}\left(\hat{\beta}_1\right)} \sim t_{n-2}$ if $\beta_1 = \beta_{1,0}$
- **Explanation:**
  - $\mathrm{se}(\hat{\beta}_1)$: The standard error of the estimated slope coefficient. It measures the precision of the slope estimate.
  - **Confidence Interval:** Provides a range within which the true population slope (β_1) is likely to lie with a certain confidence level.
  - **Test Statistic:** Used to test a hypothesis about the true population slope (β_1), often against a null hypothesis that $\beta_1 = 0$ (i.e., no linear relationship between X and Y). The test statistic follows a t-distribution with $n-2$ degrees of freedom.
- **When to use:** To perform inference (confidence intervals and hypothesis tests) on the slope coefficient in simple linear regression.

Confidence Interval for Mean Response

- **Formula:** $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$

- **Explanation:** This interval estimates the **mean** (**average**) **value of** Y for a specific value of X, say x_0. It's narrower than the prediction interval because we are estimating a mean, not a single observation.
- **When to use:** When you want to estimate the average outcome of Y for a given X value, for example, the average blood pressure for all individuals with a certain dosage of medication.

Prediction Interval for a New Observation

- **Formula:** $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$
- **Explanation:** This interval predicts the value of a **single new observation of** Y for a specific value of X, say x_0. It's wider than the confidence interval for the mean response because it accounts for both the uncertainty in the regression line and the inherent variability of individual observations.
- **When to use:** When you want to predict a single future outcome of Y for a given X value, for example, the blood pressure of a new individual with a certain dosage.

Analysis of Variance (ANOVA)

ANOVA is used to compare means across two or more groups. The formulas here typically refer to a One-Way ANOVA.

Sums of Squares for ANOVA

- **Formulas:**
  - $\mathrm{SS}_T = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{..}\right)^2$ ($N - 1$ d.f.) (Total Sum of Squares)
  - $\mathrm{SS}_E = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{i.}\right)^2$ ($N - a$ d.f.) (Error Sum of Squares)
  - $\mathrm{SS}_{Tr} = \sum_{i=1}^{a} n\left(\bar{y}_{i.} - \bar{y}_{..}\right)^2$ ($a - 1$ d.f.) (Treatment Sum of Squares)
  - **Relationship:** $\mathrm{SS}_T = \mathrm{SS}_{Tr} + \mathrm{SS}_E$
- **Explanation:** These sum of squares decompose the total variation in the data (SS_T) into variation due to differences between group means (SS_{Tr}) and variation within groups (random error, SS_E).
  - y_{ij}: j-th observation in the i-th group.
  - $\bar{y}_{i.}$: Mean of the i-th group.
  - $\bar{y}_{..}$: Overall mean of all observations.
  - a: Number of groups (treatments).
  - n: Number of observations per group (assuming equal sample sizes per group here, $N = a \times n$).
  - N: Total number of observations.
- **When to use:** These are preliminary calculations for conducting an ANOVA test.

Pooled Variance Estimate

- **Formula:** $\hat{\sigma}^2 = \mathrm{SS}_E/(N - a) = \left(s_1^2 + \cdots + s_a^2\right)/a.$
- **Explanation:** This is the estimated common variance within all groups (σ^2), often called Mean Squared Error (MSE). It is an average of the sample variances from each group, assuming the population variances are equal across groups.
- **When to use:** In ANOVA, this is used as the denominator in the F-test statistic and for post-hoc comparisons.

F-Test Statistic for ANOVA

- **Formula:** $F_0 = \frac{\mathrm{SS}_{Tr}/(a-1)}{\mathrm{SS}_E/(N-a)} \sim F_{a-1,N-a}$ if all means are equal.
- **Explanation:** The F-statistic is the ratio of Mean Square Treatment ($\mathrm{MS}_{Tr} = \mathrm{SS}_{Tr}/(a - 1)$) to Mean Square Error ($\mathrm{MS}_E = \mathrm{SS}_E/(N - a)$). If the null hypothesis (that all group means are equal) is true, this ratio follows an F-distribution with $a - 1$ numerator degrees of freedom and $N - a$ denominator degrees of freedom. A large F-value suggests significant differences between group means.
- **When to use:** To test the null hypothesis that there are no statistically significant differences between the means of two or more groups.

Fisher's LSD Test (Least Significant Difference) and CI

- **Formulas:**
  - $\mathrm{LSD} = t_{\alpha/2,N-a}\sqrt{\mathrm{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
  - **CI:** $\bar{y}_{i.} - \bar{y}_{j.} \pm \mathrm{LSD}$
- **Explanation:** Fisher's LSD is a post-hoc test used after a significant F-test in ANOVA. It compares all possible pairs of group means to determine which specific pairs are significantly different. The LSD value provides a threshold; if the absolute difference between two group means exceeds the LSD, the difference is considered statistically significant. The confidence interval provides an estimate for the true difference between two specific group means.
- **When to use:** After conducting an ANOVA and rejecting the null hypothesis (meaning at least one group mean is different), you use post-hoc tests like LSD to identify *which* specific group means differ from each other. Note that Fisher's LSD is prone to an inflated Type I error rate when many comparisons are made; other post-hoc tests like Tukey's HSD or Bonferroni correction are often preferred for multiple comparisons.

Some Important Probability Distributions

This section summarizes common probability distributions, their probability mass/density functions, and their mean and variance.

Distribution	Probability mass or density function	$E(X)$	$V(X)$	When to Use
Binomial (n, p)	$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	np	$np(1-p)$	For the number of successes in a fixed number of independent Bernoulli trials. E.g., number of heads in 10 coin flips.
Geometric (p)	$f(x) = p(1-p)^{x-1}$, $x = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	For the number of Bernoulli trials needed to get the first success. E.g., number of attempts until first success in a game.
Negative Binomial (r, p)	$f(x) = \binom{x-1}{r-1}(1-p)^{x-r} p^r$, $x = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	For the number of Bernoulli trials needed to get the r-th success. E.g., number of sales calls until 5th successful sale.
Hypergeometric (K, N, n)	$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$, $x = \max(0, n+K-N), \ldots, \min(n, K)$	np	$np(1-p)\left(\frac{N-n}{N-1}\right)$	For the number of successes in a sample drawn *without replacement* from a finite population. E.g., number of defectives in a sample of 10 from a batch of 100 with 20 defectives.
Poisson (λ)	$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, $x = 0, 1, \ldots$	λ	λ	For the number of events occurring in a fixed interval of time or space if these events occur with a known average rate (λ) and independently of the time since the last event. E.g., number of phone calls received per hour at a call center.
Exponential (λ)	$f(x) = \lambda e^{-\lambda x}$, $x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	For the time until the next event in a Poisson process. It is a continuous distribution. E.g., time between customer arrivals.
Normal (μ, σ^2)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$, $-\infty < x < \infty$	μ	σ^2	For continuous data that clusters around a central value with symmetrical tails. Many natural phenomena are approximately normally distributed. Foundational for half of the Test 2 stuff.