# Lecture Summaries

Alex White

July 29, 2025

# 1 Lecture 1: Introduction to Probability and Data Analysis

## 1.1 Descriptive Statistics

**Types of Variables and Data:**

- **Quantitative Variables:** Represent numerical data and include:
  - **Continuous** - Data that can take any value within a range (e.g., height, weight).
  - **Discrete** - Data that takes fixed values, often counts (e.g., number of students).
- **Categorical Variables:** Represent categories or groups, including:
  - **Binary** - Two categories (e.g., yes/no).
  - **Nominal** - Categories without a natural order (e.g., gender, race).
  - **Ordinal** - Categories with a meaningful order (e.g., ratings: poor, average, good).

## 1.2 Summarizing Data

**Five-Number Summary:**

- Components: Minimum, Q1 (First Quartile), Median, Q3 (Third Quartile), and Maximum.
- Provides a summary of the distribution's range and center.

**Mean and Standard Deviation:**

- **Mean:** The average of the data, a measure of central tendency.
- **Standard Deviation:** Describes the dispersion or spread of the data around the mean.

## 1.3 Visualizing Data

**Common Plot Types:**

- **Boxplot:** Represents the five-number summary; useful for identifying outliers.
- **Histogram:** Displays the frequency distribution of a single variable; helpful for understanding the shape of the data.
- **Scatter Plot:** Shows the relationship between two quantitative variables; useful for identifying correlation.

## 1.4 Data Analysis and Probability Models

**Case Study - Eruption Times:**

- Analysis of eruption times for a geyser with a five-number summary and interpretation of the distribution.
- Identified potential patterns and discussed variability in data.

**Survey Data - Handedness and Eye Color:**

- Calculated proportions of certain combinations (e.g., blue-eyed left-handers, brown-eyed right-handers).
- Discussed how data grouping reveals insights into variable distributions.

## 1.5 Review and Example Questions

**Key Questions from the Lecture:**

- Classify variables (e.g., age, response type, sleep time) as either categorical or quantitative.
- Identifying study types:
  - **Observational Study:** Observes individuals without intervention.
  - **Designed Experiment:** Applies specific treatments to study effects.

## 2 Lecture 2: Understanding Randomness, Counting, and Probability Laws

### Random Experiments and Probability

- **Random Experiments**:
  - Experiments with uncertain outcomes, like tossing coins or rolling dice.
  - Examples include coin tosses, dice rolls, and random tweets.
- **Probability vs. Statistics**:
  - **Probability** is theoretical and model-based; it deals with calculating likelihoods based on assumptions.
  - **Statistics** is empirical, analyzing data to make conclusions (e.g., checking if a coin is fair based on observed results).
- **Probability Model Components**:
  - **Sample Space** ($\Omega$): The set of all possible outcomes.
  - **Event**: A subset of the sample space, either simple (one outcome) or compound (multiple outcomes).
  - **Probability Measure**: Assigns likelihoods to events based on three main rules:
    1. Probability of an event is between 0 and 1.
    2. Probability of the whole sample space is 1.
    3. For disjoint events, the probability of their union is the sum of individual probabilities.
- **Properties of Probability Measures**:
  - **Complements**: $P(A^c) = 1 - P(A)$.
  - **Subsets**: If $A \subset B$, then $P(A) \leq P(B)$.
  - **Union of Events**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

### Counting Techniques

- **Equally Likely Outcomes**:
  - When outcomes are equally likely, the probability of an event $A$ is $P(A) = \frac{|A|}{|\Omega|}$.
  - Examples include rolling dice and determining probabilities based on counting outcomes.
- **Fundamental Principle of Counting**:
  - **Stages**: If an experiment has multiple stages, the total outcomes are the product of possibilities at each stage.
  - **Example**: Completing a 20-question exam with 4 answer choices each has $4^{20}$ possible completions.
- **Counting with Replacement**:
  - For ordered arrangements (permutations), all choices are available in each stage.
  - **Example**: Arranging 9 players from a team of 9, without re-selection, has 9! possibilities.
- **Counting without Replacement**:
  - Selections where order doesn't matter (combinations) divide out the repetitive arrangements.
  - **Example**: Choosing 9 players from a pool of 30, order doesn't matter: $\frac{30!}{(30-9)! \times 9!}$.

### Laws of Probability

- **Law of Total Probability**: Given a partition $\{B_1, B_2, \ldots, B_n\}$ of the sample space $\Omega$, the probability of any event $A$ can be expressed as:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

- **Bayes' Rule**: For events $B_1, B_2, \ldots, B_n$ that partition $\Omega$,

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

  This rule is fundamental in updating probabilities based on new information.

**Example: Monty Hall Problem**

- **Problem Setup**: A contestant picks one of three doors. Behind one door is a prize; the other two contain nothing. Monty Hall opens one of the remaining doors without a prize and offers the contestant a choice: stick with the initial door or switch to the other unopened door.
- **Probability Calculation**:
  - Let $A_i$ be the event that the prize is behind door $i$.
  - Let $B_j$ be the event that Monty opens door $j$.
  - The initial probabilities are $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$.
  - Using Bayes' Rule:
  $$P(A_1|B_3) = \frac{P(B_3|A_1)P(A_1)}{P(B_3|A_1)P(A_1) + P(B_3|A_2)P(A_2)}$$
  - After calculations, it is determined that switching doors gives the contestant a $\frac{2}{3}$ chance of winning, compared to a $\frac{1}{3}$ chance if they stick with the original choice.

**Probability Calculations Using Counting**

- **Permutations and Combinations**:
  - **Permutations (Ordered Selection)**: $P(n, r) = \frac{n!}{(n-r)!}$.
  - **Combinations (Unordered Selection)**: $C(n, r) = \frac{n!}{r!(n-r)!}$.
- **Example Calculations**:
  - Examples include calculating lineups for teams, both ordered (permutations) and unordered (combinations).

## Lecture 3: Random Variables and Their Distribution

### 3.1 Introduction to Random Variables

- **Definition**: A random variable $X$ is a function that assigns a real number to each outcome $\omega$ in the sample space $\Omega$. This provides a numeric representation of experimental outcomes for analysis.
  - **Example**: Rolling a die and recording the face value, or counting heads in coin tosses.
- **Events and Random Variables**: Events involving random variables can be defined with conditions on $X$. For example, the event that $X \geq 1$ in a coin toss corresponds to a set of outcomes.
- **Cumulative Distribution Function (CDF)**: The CDF $F(x) = P(X \leq x)$ represents the probability of $X$ being less than or equal to $x$. Key properties of the CDF include being increasing, right-continuous, and bounded by 0 and 1.
- **Discrete vs. Continuous Random Variables**:
  - **Discrete**: Takes on countable values, represented by a probability mass function (pmf).
  - **Continuous**: Takes on an uncountable range, represented by a probability density function (pdf), with a continuous CDF.
- **Probability Density Function (PDF)**: For a continuous variable $X$, the pdf $f(x)$ allows for calculating probabilities over intervals. The pdf integrates to 1 over the domain of $X$.

### 3.2 Expectation and Variance

- **Expectation**:
  - **Discrete**: $E(X) = \sum xP(X = x)$.
  - **Continuous**: $E(X) = \int xf(x)\,dx$.
  - **Example**: Rolling two dice and calculating the expectation of the highest face value.
- **Functions of Random Variables**: The expected value of a function $g(X)$ can be calculated as $E(g(X))$.

  - **Law of the Unconscious Statistician (LOTUS)**:
    * **Discrete**: $E(g(X)) = \sum g(x)f(x)$.
    * **Continuous**: $E(g(X)) = \int g(x)f(x)\,dx$.

- **Linearity of Expectation**: For any constants $a$ and $b$, $E(a \cdot g(X) + b \cdot h(X)) = a \cdot E(g(X)) + b \cdot E(h(X))$.
  - **Example**: Converting temperature from Fahrenheit to Celsius by applying the linearity property.

- **Variance**:
  - Defined as $\text{Var}(X) = E((X - \mu)^2) = E(X^2) - \mu^2$.
  - **Nonlinearity**: The variance of $aX + b$ is $a^2\text{Var}(X)$.

- **Applications**: Expectation and variance help quantify average values and variability in data, such as temperature measurements or dice rolls, enabling probabilistic interpretations.

### 3.3 Moment Generating Functions (MGFs)

#### Definition

The moment generating function (MGF) of a random variable $X$, denoted as $M_X(s)$, is defined by:

$$M_X(s) = E(e^{sX})$$

where $s$ is within an interval containing zero, ensuring that $M_X(s)$ is finite.

#### Properties

- **Uniqueness**: MGFs uniquely determine distributions. If two random variables have the same MGF, their distributions are identical.

- **Moment Calculation**: For a random variable $X$ with an MGF $M_X(s)$ finite around $s = 0$, the $n$-th moment $E(X^n)$ can be derived as:
$$E(X^n) = \frac{d^n M_X(s)}{ds^n}\bigg|_{s=0}$$

- **Transformation Property**: For constants $a$ and $b$, the MGF of $aX + b$ is:
$$M_{aX+b}(s) = e^{bs} M_X(as)$$

#### Example

For a random variable $X$ with MGF $M(s) = \frac{p}{1-e^{s(1-p)}}$ (valid for $s < -\ln(1-p)$), the expectation $E(X)$ can be found by differentiating $M(s)$ with respect to $s$ and evaluating at $s = 0$:

$$E(X) = \frac{dM(s)}{ds}\bigg|_{s=0}$$

## Lecture 4: Common Probability Distributions

### 4.1 Types of Probability Distributions

- **Discrete Distributions**: Defined by a probability mass function (pmf) and suitable for variables that take on countable values.

- **Continuous Distributions**: Defined by a probability density function (pdf) for variables with an uncountable range of values.

### 4.2 Discrete Distributions

- **Bernoulli Distribution**
  - A random variable $X$ is Bernoulli distributed with probability $p$ of "success" and $1 - p$ of "failure."
  - Notation: $X \sim \text{Ber}(p)$
  - Properties:
    - **Expectation** $E(X) = p$
    - **Variance** $\text{Var}(X) = p(1 - p)$
    - **Moment Generating Function** $M(s) = 1 - p + pe^s$
- **Binomial Distribution**
  - The sum of $n$ independent Bernoulli trials with success probability $p$.
  - Notation: $X \sim \text{Bin}(n, p)$
  - Probability mass function: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$
  - Properties:
    - **Expectation** $E(X) = np$
    - **Variance** $\text{Var}(X) = np(1 - p)$
    - **Moment Generating Function** $M(s) = (1 - p + pe^s)^n$
- **Poisson Distribution**
  - Used for modeling the number of events in a fixed interval with a constant mean rate $\lambda$.
  - Notation: $X \sim \text{Poisson}(\lambda)$
  - Probability mass function: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$
  - Properties:
    - **Expectation** $E(X) = \lambda$
    - **Variance** $\text{Var}(X) = \lambda$
    - **Moment Generating Function** $M(s) = e^{\lambda(e^s - 1)}$
  - **Binomial Approximation**: When $n$ is large and $p$ is small, $\text{Bin}(n, p) \approx \text{Poisson}(np)$.

### 4.3 Continuous Distributions

- **Uniform Distribution**
  - Continuous random variable $X$ on $[a, b]$ with equal probability density over the interval.
  - Notation: $X \sim U[a, b]$
  - Probability density function: $f(x) = \frac{1}{b-a}$
  - Cumulative distribution function: $F_X(x) = \frac{x-a}{b-a}$
  - Properties:
    - **Expectation** $E(X) = \frac{a+b}{2}$
    - **Variance** $\text{Var}(X) = \frac{(b-a)^2}{12}$
- **Exponential Distribution**
  - Models the time until the occurrence of an event with a rate $\lambda$.
  - Notation: $X \sim \text{Exp}(\lambda)$
  - Probability density function: $f(x) = \lambda e^{-\lambda x}$, for $x \geq 0$
  - Cumulative distribution function: $F_X(x) = 1 - e^{-\lambda x}$
  - Properties:
    - **Expectation** $E(X) = \frac{1}{\lambda}$
    - **Variance** $\text{Var}(X) = \frac{1}{\lambda^2}$
    - **Moment Generating Function** $M(s) = \frac{\lambda}{\lambda - s}$
  - **Memoryless Property**: $P(X > x + y | X > x) = P(X > y)$ for $x, y \geq 0$

## Lecture 5: Multiple Random Variables

### 5.1 Joint Distributions

- **Multiple Random Variables**: Many scenarios involve more than one random variable:
  - Example: Rolling two dice with outcomes $X$ and $Y$.
  - Example: Twitter activity where $X$ is the number of people followed, and $Y$ is the followers.
- **Joint Cumulative Distribution Function (CDF)**:
  - For two random variables $X_1$ and $X_2$, the joint CDF $F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$.
  - More generally, the joint CDF for $X_1, \ldots, X_n$ is $F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$.
- **Joint Probability Mass Function (PMF) for Discrete Variables**:
  - For discrete variables $X_1$ and $X_2$, the joint PMF $f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$.
  - Example calculation: For events $(X_1, X_2) \in B$, $P((X_1, X_2) \in B) = \sum_{(x_1,x_2)\in B} f(x_1, x_2)$.
- **Marginal PMF**:
  - The marginal PMF of $X$ is obtained by summing out $Y$: $f_X(x) = \sum_y f_{X,Y}(x, y)$.
  - Marginals cannot determine the joint PMF without additional information.
- **Joint Probability Density Function (PDF) for Continuous Variables**:
  - For continuous variables $X$ and $Y$, $f(x, y) \geq 0$ with total integral one.
  - The probability over region $B$ is calculated as $P((X, Y) \in B) = \int_{(x,y)\in B} f(x, y)\, dx\, dy$.
  - Marginal PDFs are calculated by integrating out other variables, e.g., $f_X(x) = \int f(x, y)\, dy$.

### 5.2 Conditional Distributions

- **Discrete Case**: The conditional PMF of $Y$ given $X = x$ is $P(Y = y|X = x) = \frac{P(X=x,Y=y)}{P(X=x)}$.
- **Continuous Case**: The conditional PDF of $Y$ given $X = x$ is $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$.
- **Conditional Expectation**:
  - For discrete variables, $E(Y|X = x) = \sum_y y \cdot P(Y = y|X = x)$.
  - For continuous variables, $E(Y|X = x) = \int y \cdot f_{Y|X}(y|x)\, dy$.

### 5.3 Independence of Random Variables

- **Definition**: Two random variables $X$ and $Y$ are independent if $P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$.
- **Discrete Independence**: For discrete variables, $X$ and $Y$ are independent if $f(x, y) = f_X(x) \cdot f_Y(y)$.
- **Continuous Independence**: For continuous variables, independence is shown by $f(x, y) = f_X(x) \cdot f_Y(y)$.
- **Example**: For random variables $X$ and $Y$ with joint PMF or PDF, check if the product of marginal probabilities equals the joint probability.
- **Functions of Independent Variables**: If $X$ and $Y$ are independent, any functions $g(X)$ and $h(Y)$ are also independent.

## Lecture 6: Multiple Random Variables

### 6.1 Joint Distributions and Expectations

- **Law of the Unconscious Statistician (LOTUS)**:
  - For a discrete random vector $(X_1, \ldots, X_n)$ with pmf $f$:
  $$E[h(X_1, \ldots, X_n)] = \sum_{x_1, \ldots, x_n} h(x_1, \ldots, x_n) f(x_1, \ldots, x_n)$$
  - For a continuous random vector $(X_1, \ldots, X_n)$ with joint pdf $f$:
  $$E[h(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$
- **Expectation of Linear Combinations**: For constants $a, b_1, \ldots, b_n$:
$$E[a + b_1 X_1 + \cdots + b_n X_n] = a + b_1 \mu_1 + \cdots + b_n \mu_n$$

### 6.2 Covariance and Correlation

- **Covariance**: Measures how two random variables $X$ and $Y$ vary together:
$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$$
- **Properties of Covariance**:
  - Symmetry: $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
  - Variance of Sum: $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$
- **Correlation Coefficient** $\rho(X, Y)$: Scaled covariance:
$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$
with $\rho \in [-1, 1]$.

### 6.3 Conditional Distributions and Expectations

- **Conditional Probability**: For discrete random variables $X$ and $Y$:
$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$
- **Conditional Probability Mass Function (pmf)**:
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$
- **Law of Total Probability for Expectations** (Tower Property):
$$E(E(X|Y)) = E(X)$$

### 6.4 Example Problems

- **Conditional Expectation Example**: Given $f_Y(y) = 3y^2$ for $y \in [0, 1]$, and $X|Y = y \sim \text{Uniform}(0, y)$, calculate:
$$E(X|Y = y) = \int_0^y x \cdot \frac{1}{y} \, dx = \frac{y}{2}$$
and
$$E(X) = \int_0^1 \frac{y}{2} \cdot 3y^2 \, dy = \frac{3}{8}$$

## Lecture 7: Estimators and Confidence Intervals

### 7.1 Estimators and Point Estimates

- **Estimators and Estimates**:
  - Given a sample $X_1, X_2, \dots, X_n$ from a population, the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an estimator for the population mean $\mu$.
  - An estimate is a specific value from a realisation of the estimator.
- **Uncertainty in Point Estimates**:
  - Different samples yield different estimates due to sampling variability.
  - Larger sample sizes generally reduce variability in the estimator, providing more accurate point estimates.

### 7.2 Confidence Intervals (CI)

- **Definition of Confidence Interval**:
  - A CI is an interval estimate that reflects the uncertainty in a point estimate. It indicates the range within which the population parameter lies with a specified level of confidence.
  - The length of the CI depends on sample size and variability of the data.
- **Example with Normal Distribution**:
  - For a normally distributed sample $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$, the CI for $\mu$ (assuming known variance) is:
  $$\left[\bar{X} - z^* \frac{\sigma}{\sqrt{n}}, \bar{X} + z^* \frac{\sigma}{\sqrt{n}}\right]$$
  where $z^*$ is the critical value from the standard normal distribution for a given confidence level.
- **Interpreting Confidence Intervals**:
  - A 95% CI means that if we repeated the sampling process many times, approximately 95% of the CIs would contain the true mean $\mu$.
  - Common misunderstandings include mistaking the CI for a probability interval of a specific sample.
- **Effect of Sample Size and Confidence Level**:
  - Increasing sample size reduces the width of the CI.
  - Increasing the confidence level (e.g., from 95% to 99%) increases the CI width due to a higher critical value $z^*$.

### 7.3 Student's t-Distribution for Unknown Variance

- **Confidence Interval with Unknown Variance**:
  - When population variance $\sigma^2$ is unknown, use the sample standard deviation $S$ and the t-distribution:
  $$\left[\bar{X} - t^* \frac{S}{\sqrt{n}}, \bar{X} + t^* \frac{S}{\sqrt{n}}\right]$$
  where $t^*$ is the critical value from the t-distribution with $n-1$ degrees of freedom.
- **Properties of Student's t-Distribution**:
  - Symmetric and bell-shaped, like the normal distribution, but with heavier tails.
  - As $n$ increases, the t-distribution approaches the standard normal distribution.
- **Example Calculation**:
  - For a sample of 25 students with mean video game time of 6.96 hours and $S = 7.42$, the 95% CI for the population mean is:
  $$\left[6.96 - 1.96 \times \frac{7.42}{\sqrt{25}}, 6.96 + 1.96 \times \frac{7.42}{\sqrt{25}}\right] = [4.05, 9.87]$$

### 7.4 Confidence Interval for Proportions

- **Confidence Interval for a Population Proportion**:
  - For a sample proportion $\hat{p} = X/n$, the approximate CI for a population proportion $p$ is:

$$\left[\hat{p} - z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

    provided $n\hat{p} \geq 5$ and $n(1-\hat{p}) \geq 5$.

- **Example Calculation for Proportion**:
  - For 103 female students, 22 responded they did not play video games. The proportion is $\hat{p} = 22/103$.
  - Using $z^* = 1.96$ for 95% confidence, the CI is:

$$\left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{103}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{103}}\right]$$

## Lecture 8: Hypothesis Testing for One and Two Populations

### 8.1 Introduction to Hypothesis Testing

- **Hypothesis Testing** is a statistical approach used to make inferences or decisions about a population parameter based on sample data.
- Involves two competing hypotheses:
  - **Null Hypothesis** ($H_0$): Represents the status quo or no effect.
  - **Alternative Hypothesis** ($H_1$): Contradicts $H_0$ and represents the effect we want to support.
- The objective is to assess the evidence in the sample data to decide between $H_0$ and $H_1$.
- **p-value**: The probability of obtaining a test statistic as extreme or more extreme than the one observed, assuming $H_0$ is true.
- **Decision Rules**:
  - Reject $H_0$ if the p-value is smaller than a chosen significance level $\alpha$.
  - Fail to reject $H_0$ if the p-value is greater than or equal to $\alpha$.

### 8.2 Test Statistic and Types of Tests

- **Test Statistic** is computed to measure the degree of consistency between the sample data and $H_0$. For hypothesis tests, the test statistic is often:

$$\text{Test Statistic} = \frac{\text{Estimator} - \text{Hypothesized Value}}{\text{Standard Error of Estimator}}$$

- Types of Hypothesis Tests:
  - **One-Sided Test**: Tests whether the parameter is greater or less than a hypothesized value.
  - **Two-Sided Test**: Tests whether the parameter is different from a hypothesized value.

### 8.3 Hypothesis Testing for Two Populations

- Extends single-population hypothesis testing methods to compare parameters across two population distributions.
- **Assumptions**:
  1. $X_1, \ldots, X_{n_X}$ are i.i.d., representing one population.
  2. $Y_1, \ldots, Y_{n_Y}$ are i.i.d., representing a second population.
  3. All $X$ and $Y$ samples are independent of one another.

- **Test Statistic for Comparing Two Means**: For testing $H_0 : \mu_X = \mu_Y$,

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n_X + S_Y^2/n_Y}}$$

where $S_X^2$ and $S_Y^2$ are sample variances. The test can be adjusted based on equal or unequal variances.

**Example: Two-Sample Hypothesis Test**

- A study investigated whether Minecraft improves spatial reasoning. Thirty-four participants were randomly allocated to Group A (Minecraft training) or Group B (no training).
- **Hypotheses**:

$$H_0 : \mu_A = \mu_B \quad \text{vs.} \quad H_1 : \mu_A > \mu_B$$

- Calculate the test statistic using pooled variance if variances are equal or Welch's correction if variances are unequal.
- Conclusion: Based on the p-value, determine if Minecraft training significantly improves spatial reasoning.

**Comparing Two Proportions**

- Suppose $X_1, \ldots, X_{n_X}$ is a sample from a Bernoulli distribution with success probability $p_X$, and $Y_1, \ldots, Y_{n_Y}$ is a sample from a Bernoulli distribution with success probability $p_Y$.
- Test Statistic:

$$T = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$$

where $\hat{p} = \frac{\sum X + \sum Y}{n_X + n_Y}$ is the pooled proportion under $H_0 : p_X = p_Y$.

**Example: Comparing Proportions for Ear Infections**

- A study tested xylitol's effectiveness in preventing ear infections. Two groups received either xylitol or a placebo, and infection rates were compared.
- **Hypotheses**:

$$H_0 : p_{\text{Xylitol}} = p_{\text{Placebo}} \quad \text{vs.} \quad H_1 : p_{\text{Xylitol}} < p_{\text{Placebo}}$$

- Conclusion: Based on the calculated p-value, determine if xylitol significantly reduces ear infection risk.

## Lecture 9: Linear Regression and Inference

### 9.1 Introduction to Regression Analysis

- Regression analysis explores relationships between variables, enabling inference about one variable using the values of another. Example: predicting college GPA based on high school GPA.
- **Key Terms**:
  - $Y$: Response variable.
  - $X$: Predictor or explanatory variable.

### 9.2 Simple Linear Regression Model

- The simplest relationship between variables $y$ and $x$ is linear: $y = \beta_0 + \beta_1 x$.
- For random variables, this relationship becomes probabilistic:

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$ represents random deviation.

- **Estimation of Parameters**: $\beta_0$ and $\beta_1$ minimize the sum of squared errors, $\sum(y_i - (\beta_0 + \beta_1 x_i))^2$.

- **Least Squares Estimates**:

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

- **Mean Squared Error (MSE)** estimates $\sigma^2$:

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

### 9.3 Inference for Linear Regression

- **Assumptions**:
  - Independence: $Y_i$'s are independent.
  - Normality: $Y_i$'s follow a normal distribution.
  - Linearity: Mean of $Y$ is a linear function of $X$.
  - Homoscedasticity: Constant variance $\sigma^2$ for $Y_i$ across all $x$.
- With assumptions met, the estimators have distributions:

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right)\right), \quad \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum(x_i - \bar{x})^2}\right)$$

- **Hypothesis Testing**:

$$t = \frac{\hat{\beta}_i - \beta_i}{\text{se}(\hat{\beta}_i)} \sim t_{n-2}, \quad i = 0, 1$$

- **Confidence Interval** for $\beta_i$:

$$\hat{\beta}_i \pm t^* \cdot \text{se}(\hat{\beta}_i)$$

- **Prediction** of $Y$ given $x$:

$$\hat{\mu}_x = \hat{\beta}_0 + \hat{\beta}_1 x$$

  with variance

$$\text{Var}(\hat{\mu}_x) = \sigma^2\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)$$

## Lecture 10: Multiple Regression and Case Study

### 10.1 Introduction to Multiple Regression

- **Definition**: A multiple linear regression model includes more than one explanatory variable. The general form of the model is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1}x_{p-1} + \epsilon$$

  where $\epsilon \sim N(0, \sigma^2)$.
- **Interpretation of Coefficients**: Each coefficient $\beta_i$ represents the expected change in $Y$ associated with a 1-unit increase in $x_i$, holding all other variables constant.

### 10.2 Models with Interaction Terms

- **Interaction Terms**: Used to model relationships where the effect of one predictor variable on $Y$ depends on another predictor. Common forms include:
  - **First-order with interaction**: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$
  - **Second-order with interaction**: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon$

### 10.3 Multiple Regression with Categorical Variables

- **Incorporating Categorical Variables**: Use dummy variables to include categorical predictors in regression models. For example:

$$x = \begin{cases} 1 & \text{if category A} \\ 0 & \text{otherwise} \end{cases}$$

- Categorical variables with more than two categories require $c - 1$ dummy variables.

### 10.4 Estimation for Multiple Linear Regression

- **Objective**: Find coefficients $\beta$ that minimize the sum of squared errors (SSE).
- **Normal Equation**: The estimates are given by solving:

$$X^T X\beta = X^T y$$

where $X$ is the matrix of predictor values and $y$ is the response vector.

### 10.5 Inference on Predictors

- Each coefficient $\beta_i$ has an associated standard error, allowing us to test hypotheses (e.g., $H_0 : \beta_i = 0$).
- For testing groups of predictors, fit both full and reduced models and use an F-test:

$$F = \frac{(\text{SSE}_{\text{reduced}} - \text{SSE}_{\text{full}})/(p - l)}{\text{SSE}_{\text{full}}/(n - p)}$$

### 10.6 Case Study: Fish Catch Data

- A regression analysis was conducted on fish caught in Lake Längelmävesi, Finland.
- Multiple predictors (size, type, etc.) were used to explain variability in the catch quantity.

## Lecture 11: Analysis of Variance (ANOVA)

### Introduction to ANOVA

- **Definition of ANOVA**: A statistical method used to compare the means of three or more populations, assuming all are normally distributed with the same standard deviation.
- **Connection to Regression**: ANOVA is a specific case of multiple regression models with categorical explanatory variables and a continuous response variable.
- **Application**: Often applied in factorial experiments to assess how a response variable is influenced by one or more factors tested at several levels.

### Single-Factor ANOVA Example: Stem Cell Survival Study

- **Experiment Details**: Mouse embryonic stem cells were cultured in 16 wells with fibrin scaffolds, counting live and dead cells after three days.
- **Total Variability**: Total variability in data is quantified by the sample variance and sum of squares ($SST = 4237.8$) with 15 degrees of freedom.

### Survival and NT3 Concentration: Hypothesis Testing

- **Hypothesis**: Testing if different levels of NT3 concentration impact survival rates. Defined treatment levels for NT3 concentrations are $0, 10, 20,$ and 30 ng/mL.
- **ANOVA Table Construction**:
  - Calculate sums of squared errors (SSE) across groups and compare to total sum of squares (SST).
  - Example ANOVA table summarizing results:

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| NT3-factor | 3 | 2779.2 | 926.4 | 7.62 |
| Residuals | 12 | 1458.6 | 121.6 | |
| Total | 15 | 4237.8 | | |

**ANOVA Model as Multiple Linear Regression**

- **ANOVA Model Formulation**:

$$Y = \mu_1 x_1 + \mu_2 x_2 + \mu_3 x_3 + \mu_4 x_4 + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

  where each $x_i$ is an indicator variable for NT3 concentration.

- **Full and Reduced Models**:
  - Full Model: $Y = \mu_1 + (\mu_2 - \mu_1)x_2 + (\mu_3 - \mu_1)x_3 + (\mu_4 - \mu_1)x_4 + \epsilon$.
  - Reduced Model: $Y = \mu_1 + \epsilon$.

**F-Statistic and Hypothesis Testing**

- **Calculation of F-Statistic**:

$$F = \frac{(SST - SSE)/(d-1)}{SSE/(n-d)}$$

  where $SST$ is the total sum of squares and $SSE$ is the sum of squared errors.

- **P-Value Interpretation**: Using the calculated F-statistic and the associated degrees of freedom, the p-value determines if the null hypothesis (no effect) is rejected.

**Assumptions and Diagnostics for ANOVA**

- **Assumptions**: Assumes normality, independence, and equal variances across groups.
- **Diagnostics**: Residual plots, normal Q-Q plots, and checking homoscedasticity are essential to validate model assumptions.

**Multiple Comparisons and Pairwise Comparisons**

- **When $H_0$ is Rejected**: If ANOVA detects differences among means, further analysis identifies which means differ.
- **Bonferroni Method**: Adjusts for multiple hypothesis tests by setting each test's significance level to $\alpha/k$ to control overall error rate.

## Lecture 12: Two-Factor ANOVA and Goodness of Fit Tests

### 12.1 Two-Factor ANOVA

- **Two-Factor ANOVA**: Extends single-factor ANOVA to test effects from two factors (categorical variables) simultaneously, each with multiple levels. Useful when both main effects and their interaction are of interest.
- **Example - NT3 and PDGF on Cell Survival**: NT3 (four levels) and PDGF (two levels) were tested to observe cell survival. Observations suggested an interaction between NT3 and PDGF in cell survival.
- **Model Representation**:

$$Y = \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \mu_{jk} x_{jk} + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

  where $x_{jk}$ indicates factor levels and $\mu_{jk}$ the mean response.

- **Sums of Squares**:
  - SST $= \sum_{i,j,k} (Y_{ijk} - \bar{Y})^2$ - Total variability.
  - SSF1, SSF2 - Variability between levels of each factor.
  - SSF12 - Variability due to interaction between factors.
  - SST = SSE + SSF1 + SSF2 + SSF12.

- **Hypothesis Testing in Two-Way ANOVA**:

$$H_0^{(1)} : \text{No effect of Factor 1}, \quad H_0^{(2)} : \text{No effect of Factor 2}, \quad H_0^{(12)} : \text{No interaction effect}$$

- **Example ANOVA Table**:

| Source | df | SS | MS | F |
|---|---|---|---|---|
| NT3 | 3 | 2779.23 | 926.41 | 9.60 |
| PDGF | 1 | 43.23 | 43.23 | 0.45 |
| Interaction | 3 | 643.21 | 214.40 | 2.22 |
| Error | 8 | 772.13 | 96.52 | |
| Total | 15 | 4238 | | |

### 12.2 Goodness of Fit for Categorical Data

- **Goodness of Fit Concept**: Assesses how well a model explains observed data; discrepancies between observed and expected values indicate the model's accuracy.
- **Example - Blood Types**:
  - Tested theoretical probabilities for blood types A, B, AB, and O in a sample of Californians.
  - **Null Hypothesis $H_0$**: Theoretical probabilities are correct.
  - **Chi-Square Test Statistic**:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \quad \chi^2 \sim \chi^2_{k-1}$$

- **Chi-Square Test - Example Results**:

$$p\text{-value} = P(\chi^2_3 \geq 20.37) \approx 1.42 \times 10^{-4}$$

- **Making Babies by the Flip of a Coin?**:
  - Tested a binomial model ($X \sim \text{Binomial}(2, 0.5)$) for two-child families.
  - Chi-square test used to compare observed and expected counts, yielding:

$$\chi^2 = 57.28, \quad p\text{-value} \approx 3.64 \times 10^{-13}$$

- **Test of Independence**: For contingency tables, tests association between rows and columns.
- **Example - Green Roof Survey**:
  - Tested association between age groups and interest in green roof systems using a chi-square test.
  - Resulted in strong evidence of association.

## Lecture 13: Frequentist vs. Bayesian Inference

### Frequentist vs. Bayesian Inference

- Frequentist inference focuses on confidence intervals and hypothesis testing. This approach defines probabilities in terms of relative frequency over repeated trials.
- Bayesian inference provides an alternative by interpreting probability as a subjective degree of belief, incorporating prior beliefs with observed data.

### Frequentist vs. Bayesian Hypothesis Testing

- **Frequentist Hypothesis Testing**: Calculates the $p$-value as $P(\text{observed results} \mid H_0 \text{ is true})$. This does not give $P(H_0 \text{ is true} \mid \text{observed results})$.
- **Bayesian Hypothesis Testing**: Uses Bayes' Theorem to calculate $P(H_0 \mid \text{observed results})$, updating the prior probability of $H_0$ based on evidence:

$$P(H_0 \mid \text{observed results}) = \frac{P(\text{observed results} \mid H_0) P(H_0)}{P(\text{observed results})}$$

**Mental Experiments in Hypothesis Testing**

- **Example 1: Neighbor's Newborn (Boy or Girl)**
  - Hypothesis: $H_0$ (Newborn is a girl) with prior $P(H_0) = \frac{1}{2}$.
  - If 60% of girls and 10% of boys have pink rooms:

$$P(\text{girl} \mid \text{pink room}) = \frac{P(\text{pink room} \mid \text{girl})P(\text{girl})}{P(\text{pink room})} = \frac{0.6 \times 0.5}{0.6 \times 0.5 + 0.1 \times 0.5} \approx 0.86$$

- **Example 2: Neighbor's Newborn (Puppy or Not)**
  - Hypothesis: $H_0$ (Newborn is a puppy) with prior $P(H_0) = 10^{-10}$.
  - If 80% of households with a puppy have a kennel while 5% without a puppy have one:

$$P(\text{puppy} \mid \text{kennel}) = \frac{P(\text{kennel} \mid \text{puppy})P(\text{puppy})}{P(\text{kennel})} \approx 1.6 \times 10^{-9}$$

**Life and Death Example: Sally Clark Case**

- Sally Clark was wrongfully convicted in 1996 based on a misinterpretation of probability, which led to an incorrect frequentist calculation.
- Frequentist approach: $P(\text{Two SIDS} \mid \text{Innocence}) = 1$ in 73 million.
- Bayesian approach: $P(\text{Innocence} \mid \text{Two SIDS}) \geq 0.66$, highlighting the importance of prior beliefs in the interpretation.

**Differences in Interpretation of Probability**

- **Frequentist Interpretation**: Probability as the relative frequency over many trials.
- **Bayesian Interpretation**: Probability as the subjective degree of belief, incorporating prior information and observed evidence.

**Philosophical Debate and Conclusion**

- **Frequentist View**: Probability is objective and does not assign probabilities to hypotheses.
- **Bayesian View**: Allows for subjective priors, acknowledging the role of belief in science.