Sample Space and Probability Basics

Sample Space (Ω): Set of all possible outcomes of an experiment.

Probability of Event A: $P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total outcomes}}$

- Example: In a coin toss, the sample space is $\Omega = \{H, T\}$. If the coin is fair, $P(H) = \frac{1}{2}$.

Set Operations and Probability Laws

Complement of an Event (A^c): The event that A does not occur. $P(A^c) = 1 - P(A)$

Union of Events ($A \cup B$): The event that either A or B (or both) occurs.

- **Inclusion-Exclusion Principle**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Intersection of Events ($A \cap B$): The event that both A and B occur.

De Morgan's Laws: $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$

Inclusion-Exclusion for Three Events: $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Law of Total Probability: $P(A) = \sum_i P(A|B_i)P(B_i)$

Bayes' Theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Conditional Probability and Independence

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

- The probability of A occurring given that B has occurred.

Independence: - Events A and B are independent if $P(A \cap B) = P(A) \cdot P(B)$.

- Definition: X and Y are independent if $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x and y.
- Continuous Case: $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Combinatorics

Permutations (Ordered): $P(n, r) = \frac{n!}{(n-r)!}$

Combinations (Unordered): $C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$

Application Example: Choosing 9 players from a pool of 30 (without replacement): $\binom{30}{9} = \frac{30!}{(30-9)!\times 9!}$

PMF and PDF

Cdf of X**:** $F(x) = P(X \leq x)$, $x \in \mathbb{R}$.

Pmf of X**:** (discrete r.v.) $f(x) = P(X = x)$.

Pdf of X**:** (continuous r.v.) $f(x) = F'(x)$.

For a discrete r.v. X**:** $P(X \in B) = \sum_{x\in B} P(X = x)$.

For a continuous r.v. X **with pdf** f**:** $P(X \in B) = \int_B f(x)\,dx$.

Cdf (continuous): $F(x) = \int_{-\infty}^{x} f(u)\,du$.

- **Conditions for Validity**:

1. **Non-negativity**: $P(X = x) \geq 0$ or $f(x) \geq 0$.
2. **Normalization**: $\sum_x P(X = x) = 1$ or $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Expected Value, Variance

Expected Value (Mean):

- Discrete: $E(X) = \sum_x x \cdot P(X = x)$
- Continuous: $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

$E(X|X) = X$

$E(Y|X) = E(Y)$ if X and Y are independent

Variance (Spread of Distribution):

- Disc.: $\text{Var}(X) = E(X^2) - [E(X)]^2$
$= \sum_x (x - E(X))^2 \cdot P(X = x)$
- Cont.: $\text{Var}(X) = E(X^2) - [E(X)]^2$
$= \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)\,dx$

Covariance (Relationship between X and Y):

$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$
$= E(XY) - E(X)E(Y)$

Correlation Coefficient:

$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$

- ρ ranges from -1 (perfect negative) to +1 (perfect positive).

Properties:

- $E(X + Y) = E(X) + E(Y)$.
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.

Variance of a Linear Transformation: $\text{Var}(aX + b) = a^2\text{Var}(X)$

Properties:

1. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
2. $\text{Cov}(X, X) = \text{Var}(X)$
3. **$\text{Cov}(X, Y) = 0$ if X and Y are independent.**
4. $\text{Cov}(X + 2Y) = \text{Cov}(X) + 2\text{Cov}(Y)$

Central Limit Theorem (CLT)

- The sampling distribution of the sample mean $\bar{X}$ approaches a normal distribution $N(\mu, \sigma^2/n)$ as $n \to \infty$, regardless of the population distribution.

Cumulative Distribution Function (CDF)

- **Definition**: $F_X(x) = P(X \leq x)$
- Discrete: Sum of probabilities up to x
- Continuous: Integral of $f(x)$ from $-\infty$ to x, $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$
- **Finding PDF from CDF**: $f(x) = \frac{d}{dx}F_X(x)$

Transformations and Quantile Functions

- **Transformation**: For $Y = g(X)$, use the change of variables to find $f_Y(y)$.
- **Linear Example**: If $Y = aX + b$, then $f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$.
- **Non-Linear Example**: For $Y = X^2$ where $X \sim N(0, 1)$, use the Jacobian approach, splitting $f_X(x)$ for $x > 0$ and $x < 0$.
- **Monotonic Functions**: For monotonic $g(X)$, $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$ where $y = g(x)$.
- **Min and Max of Two Variables**: For $X = \min(U, V)$ and $Y = \max(U, V)$, determine distributions by analyzing the values of X and Y.

- **Quantile Function**: The inverse of the CDF, representing the value below which a proportion p of observations fall.
- Defined as $Q(p) = F^{-1}(p)$.
- **Application**: Used for calculating percentiles, such as finding the 90th percentile when $p = 0.90$.

Chi-Square and F-Distributions

- χ^2-distribution (degrees of freedom): Often used for tests involving variance, such as goodness-of-fit or independence tests.
- **Properties**: If $X_i \sim N(0, 1)$, then $\sum X_i^2 \sim \chi^2$ with n degrees of freedom.
- F-distribution: Used in ANOVA and comparison of variances.
- **Properties**: Ratio of two chi-square distributions; if $X_1 \sim \chi^2_{d_1}$ and $X_2 \sim \chi^2_{d_2}$, then $F = \frac{X_1/d_1}{X_2/d_2} \sim F(d_1, d_2)$.
- Test statistic $\chi^2 = \sum\frac{(O-E)^2}{E}$ for independence in categorical data.
- Purpose: Test if two categorical variables are independent.
- Test Statistic: $\chi^2 = \sum\frac{(O-E)^2}{E}$
- O: Observed frequency, E: Expected frequency under independence.
- Degrees of Freedom: $df = (r - 1)(c - 1)$, where r and c are the number of rows and columns.

Moment Generating Functions (MGF)

Definition: $M_X(s) = E(e^{sX})$, useful for finding moments.

- $E(X) = M'_X(0)$, $\text{Var}(X) = M''_X(0) - [M'_X(0)]^2$

Properties:

1. Linear Transformation: $M_Y(s) = e^{bs}M_X(as)$
2. Uniqueness: Identical MGFs imply identical distributions.
3. Moment Calculation: $E(X^n) = \left.\frac{d^nM_X(s)}{ds^n}\right|_{s=0}$

Benford's Law for Leading Digits

- **Formula**: $P(D = d) = \log_{10}\left(\frac{d+1}{d}\right)$, where $d \in \{1, 2, \ldots, 9\}$.
- **Example**: In financial datasets, the first digit distribution often adheres to Benford's Law, making deviations suspect for audit purposes.

Normal Distribution

- **Standard Normal (Z)**: $Z = \frac{X-\mu}{\sigma}$, $Z \sim N(0, 1)$
- **68-95-99.7 Rule**: Approx 68% of data falls within 1 standard deviation, 95% within 2, and 99.7% within 3.
- **Sum of Normals**: The sum of independent normal random variables is also normally distributed. If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.
- **App**: Useful for naturally occurring phenomena such as heights, test scores, and measurement errors.

Regression Models

- **Single Variable Model**: $Y = \beta_0 + \beta_1X + \epsilon$.
- **Multiple Regression Model**: $Y = \beta_0 + \beta_1X_1 + \cdots + \beta_pX_p + \epsilon$.
- Coefficients β_j represent the change in Y for a unit increase in X_j, holding other variables constant.
- **Interpreting Slope**: β_1 represents the change in Y per unit increase in X.

Regression Assumptions

1. **Linearity**: The relationship between X and Y should be linear.
2. **Independence**: Observations must be independent.
3. **Homoscedasticity**: Residuals should have constant variance across levels of X.
4. **Normality of Residuals**: Residuals should be approx normally distributed.

- **Residuals**: $e_i = y_i - \hat{y}_i$.

Diagnostics for Assumptions:

- **Residuals**: $e_i = y_i - \hat{y}_i$ (difference between observed and predicted values), with $\sum e_i = 0$ for least squares.
- **Diagnostic Tools**:

1. **Residual Plot**: Checks for homoscedasticity and linearity.
2. **Normal Q-Q Plot**: Assesses normality of residuals.
3. **Leverage and Influence Measures**: Identifies influential data points.

Margin of Error and Sample Size for Proportions

- **Margin of Error (ME)**:

$ME = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

- **Sample Size Calculation**:

$n = \frac{(z^*)^2\hat{p}(1-\hat{p})}{ME^2}$

- Use ME and confidence level z^* for precision.

F-Test for Model Significance

- Purpose: Tests if at least one predictor is significantly related to Y.
- Test Statistic: $F = \frac{\text{MS}_{\text{Regression}}}{\text{MS}_{\text{Residual}}}$ with DF: p and $n-p-1$.

Confidence Interval for Predicted Values

- CI for Predicted Value at $X = X_0$:

$\hat{Y}_0 \pm t^* \cdot \sqrt{\text{SE}^2(\hat{Y}_0) + \sigma^2}$, where $\text{SE}(\hat{Y}_0) = \sqrt{\frac{1}{n} + \frac{(X_0-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$.

One-Way ANOVA

- Purpose: Tests if means across multiple groups differ significantly.
- Hypotheses:
- H_0: All group means are equal.
- H_a: At least one group mean is different.
- **Conditions**: Independent observations, approx normal data within groups, homogeneity of variances (use Levene's test if needed).

ANOVA Test Statistic

- $F = \frac{\text{MS}_{\text{Between}}}{\text{MS}_{\text{Within}}} = \frac{\text{SS}_{\text{Between}}/\text{df}_{\text{Between}}}{\text{SS}_{\text{Within}}/\text{df}_{\text{Within}}}$.
- DF for Between: $k - 1$; DF for Within: $N - k$.

R-Squared in ANOVA

- $R^2 = \frac{\text{SS}_{\text{Between}}}{\text{SS}_{\text{Total}}}$: Proportion of total variability explained by the factor.

Standard Deviation in ANOVA

- Residual SD $\sigma_{\text{residual}} = \sqrt{\frac{\text{SS}_{\text{Within}}}{\text{df}_{\text{Within}}}}$.

Post-Hoc Tests

- Used to identify which specific groups differ after a significant F-test. Common methods include Tukey's HSD and Bonferroni correction.

ANOVA vs. Regression: Predictors

- **Categorical in ANOVA**: Treats factor levels as discrete.
- **Continuous in Regression**: Treats predictors as continuous variables.

Differentiation and Integration Rules

Differentiation Rules:

- **Power Rule:** $\frac{d}{dx}x^n = nx^{n-1}$
- **Product Rule:** $\frac{d}{dx}[u \cdot v] = u'v + uv'$
- **Quotient Rule:** $\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{u'v-uv'}{v^2}$
- **Chain Rule:** $\frac{d}{dx}f(g(x)) = f'(g(x)) \cdot g'(x)$

Basic Integration Rules:

- **Power Rule:** $\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$ (for $n \neq -1$)
- **Exponential Rule:** $\int e^{ax}\,dx = \frac{e^{ax}}{a} + C$
- **Logarithmic Rule:** $\int \frac{1}{x}\,dx = \ln|x| + C$

Trigonometric Integrals:

- $\int \sin(ax)\,dx = -\frac{\cos(ax)}{a} + C$
- $\int \cos(ax)\,dx = \frac{\sin(ax)}{a} + C$

Special Techniques:

- **Integration by Parts:** $\int u\,dv = uv - \int v\,du$
- **Substitution:** Let $u = g(x)$, then $\int f(g(x))g'(x)\,dx = \int f(u)\,du$

Common Confidence Interval Calculations

- **Two-Tailed Confidence Intervals**: Used when estimating a range for a parameter with both upper and lower limits.
  - **90% CI:** $z = 1.645$
  - **95% CI:** $z = 1.96$
  - **99% CI:** $z = 2.576$
- **One-Tailed Confidence Intervals:** Applied when focusing on either an upper or lower bound (e.g., "greater than" or "less than" hypotheses).
  - **90% CI:** $z = 1.28$.
  - **95% CI:** $z = 1.645$.
  - **99% CI:** $z = 2.33$.

For sample sizes $n < 30$ or unknown population standard deviation, use t-distribution values ($t_{\alpha/2,n-1}$) based on degrees of freedom (df):

Calculating Confidence Intervals

For a sample mean $\bar{X}$:

- **Normal Distribution (large** n**)**: $\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.
- **t-Distribution (small** n**)**: $\bar{X} \pm t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}$.

Standard Errors

- $\text{se}(\bar{x}) = \frac{s}{\sqrt{n}}$
- $\text{se}(\bar{x} - \bar{y}) = s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$
- Pooled sample variance: $s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}$
- $\text{se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- $\text{se}(\hat{p}_x - \hat{p}_y) = \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$

Interpreting p-values for Hypothesis Testing

The strength of evidence against the null hypothesis H_0 is typically interpreted based on the p-value:

- **Strong Evidence:** p-value < 0.01 — very strong evidence against H_0, likely leading to rejection.
- **Moderate Evidence:** $0.01 \leq p$-value < 0.05 — moderate evidence against H_0, often justifying rejection.
- **Weak Evidence:** $0.05 \leq p$-value < 0.1 — weak evidence against H_0, cautious rejection may be considered.
- **Inconclusive Evidence:** p-value ≥ 0.1 — inconclusive, insufficient evidence to reject H_0.

#	Parameter	Condition	CI Formula	Test Statistic	Degrees of Freedom (df)
1	Mean	Known σ^2	$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	N/A
2	Mean	Unknown σ^2	$\bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$	$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	$n-1$
3	Two Means	Known variances	$(\bar{X}-\bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}$	$Z = \frac{\bar{X}-\bar{Y}-\delta_0}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}$	N/A
4	Two Means	Unknown equal variances	$(\bar{X}-\bar{Y}) \pm t_{\alpha/2,df}\cdot S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}$	$T = \frac{\bar{X}-\bar{Y}-\delta_0}{S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}}$	n_X+n_Y-2
5	Two Means	Unknown unequal variances	$(\bar{X}-\bar{Y}) \pm t_{\alpha/2,\nu}\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}$	$T = \frac{\bar{X}-\bar{Y}-\delta_0}{\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}}$	$\nu = \min(n_X-1, n_Y-1)$
6	Proportion	Single	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	N/A
7	Proportion	Two (pooled)	$(\hat{p}_X-\hat{p}_Y) \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_X}+\frac{1}{n_Y}\right)}$	$Z = \frac{\hat{p}_X-\hat{p}_Y}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_X}+\frac{1}{n_Y}\right)}}$	N/A
8	Proportion	Two (non-pooled)	$(\hat{p}_X-\hat{p}_Y) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}+\frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}$	$Z = \frac{\hat{p}_X-\hat{p}_Y}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}+\frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}}$	N/A

- $\bar{X}, \bar{Y}$: Sample means for samples X and Y.
- $\sigma, \sigma_X, \sigma_Y$: Known population standard deviations (assumes normal distribution for small n or large sample size $n > 30$ for approximation).
- S, S_X, S_Y: Sample standard deviations (used when population standard deviation is unknown).
- n, n_X, n_Y: Sample sizes for samples X and Y; if $n > 30$, normal approximation is valid. Random sampling and independence within samples are assumed.
- $z_{\alpha/2}$: Critical value from the standard normal distribution for confidence level $1-\alpha$; used when sample size is large or population standard deviation is known.
- $t_{\alpha/2,df}$: Critical value from the t-distribution with df degrees of freedom for confidence level $1-\alpha$; used for smaller sample sizes or unknown population standard deviation.
- δ_0: Hypothesized difference between two population means in two-sample tests.
- $\hat{p}, \hat{p}_X, \hat{p}_Y$: Sample proportions for single or two samples.
  For two-sample tests, ensure $n \cdot \min(\hat{p}_X, 1-\hat{p}_X) > 8$ and $n \cdot \min(\hat{p}_Y, 1-\hat{p}_Y) > 8$ for normal approximation validity.
- p_0: Hypothesized proportion in a single proportion test.
- S_p: Pooled standard deviation for two samples with equal variances; only used if variances are assumed equal.
- ν: Conservative degrees of freedom, this approach provides a lower bound on the degrees of freedom.

Summary of Important Distributions

Distribution	PMF/PDF	Mean	Variance	MGF $M_X(s)$	Support
Bernoulli Ber(p)	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$1-p+pe^s$	$\{0,1\}$
Binomial Bin(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$	$(1-p+pe^s)^n$	$\{0,1,\ldots,n\}$
Poisson Poi(λ)	$\frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ	$e^{\lambda(e^s-1)}$	$\{0,1,\ldots\}$
Geometric Geom(p)	$p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^s}{1-(1-p)e^s}$ for $s < \ln\left(\frac{1}{1-p}\right)$	$\{1,2,\ldots\}$
Uniform U$[a,b]$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bs}-e^{as}}{(b-a)s}$	$[a,b]$
Exponential Exp(λ)	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-s}$ for $s < \lambda$	$\mathbb{R}^+$
Normal N(μ,σ^2)	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu s+\frac{\sigma^2 s^2}{2}}$	$\mathbb{R}$

- k: Number of successes in binomial trials.
- n: Number of trials in the binomial distribution.
- p: Probability of success in a trial (Bernoulli, Binomial, and Geometric).
- λ: Rate parameter for Poisson and Exponential distributions.
- μ: Mean of the normal distribution.
- σ^2: Variance of the normal distribution.
- a, b: Bounds for the uniform distribution.
- x: Value at which the probability or density function is evaluated.

For hypothesis testing about parameters θ_1 and θ_2 of two populations, the p-value is calculated with respect to the alternative hypothesis. Suppose $H_0 : \theta_1 - \theta_2 = \theta_0$. Then,

- If $H_1 : \theta_1 - \theta_2 > \theta_0$, the p-value is $P_{H_0}(T \geq t)$.
- If $H_1 : \theta_1 - \theta_2 < \theta_0$, the p-value is $P_{H_0}(T \leq t)$.
- If $H_1 : \theta_1 - \theta_2 \neq \theta_0$, the p-value is $2\min\{P_{H_0}(T \leq t), P_{H_0}(T \geq t)\}$.

Joint, Marginal, and Conditional Distributions, and Expectations

Joint Distributions:

Joint PMF (Discrete): $P(X = x, Y = y)$

Joint PDF (Continuous): $f_{X,Y}(x,y)$, where $P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$

Expressing Joint PDF in Terms of Conditional and Marginal Distributions:

$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$

Marginal Distributions:

Discrete: Obtained by summing over y: $P(X = x) = \sum_y P(X = x, Y = y)$

Continuous: Obtained by integrating over y: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

Conditional Distributions:

Conditional PMF (Discrete): $P(Y = y|X = x) = \frac{P(X=x,Y=y)}{P(X=x)}$

Conditional PDF (Continuous): If $f_X(x) > 0$, then

$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad y \in \mathbb{R}$

Independence of Random Variables: X and Y are independent if, for all x and y,

$P(X = x, Y = y) = P(X = x)P(Y = y) \quad$ or $\quad f_{X,Y}(x,y) = f_X(x)f_Y(y)$

Expectation and Variance:

Conditional Expectation of Y given $X = x$:

$E[Y|X = x] = \int y f_{Y|X}(y|x)\,dy$

Expectation of Y^2:

$E[Y^2] = \int_{-\infty}^{\infty} y^2 f_Y(y)\,dy$

Expected Sum: $E(aX + bY) = aE[X] + bE[Y]$

Expected Product (if X and Y are independent): $E[XY] = E[X]E[Y]$

Law of Total Expectation:

$E(Y) = E[E(Y|X)]$

Markov Inequality: For a non-negative random variable X,

$P(X > x) \leq \frac{E[X]}{x}$

Moment Generating Function (MGF)

- When it exists, for $t \in I \subset \mathbb{R}$:
  $M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx}f(x)\,dx$.

Moment Property

- The n-th moment: $E[X^n] = M_X^{(n)}(0)$, where $M_X^{(n)}(0)$ denotes the n-th derivative of the MGF evaluated at $t = 0$.
- If X and Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Properties of Linear Combinations of Normals

- If $X_i \sim N(\mu_i, \sigma_i^2)$ are independent, then $a + \sum_{i=1}^n b_i X_i \sim N\left(a + \sum_{i=1}^n b_i\mu_i, \sum_{i=1}^n b_i^2\sigma_i^2\right)$.

Multivariate Normal Distribution

- The pdf of a multivariate Normal distribution $N(\mu, \Sigma)$ for a random vector $Z \in \mathbb{R}^n$:

$$f_Z(z) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left(-\frac{1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu)\right),$$

where Σ is the covariance matrix and μ is the mean vector.

- If $X \sim N(\mu, \Sigma)$ and $Y = a + BX$, then $Y \sim N(a + B\mu, B\Sigma B^T)$.

Central Limit Theorem

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x),$$

where Φ is the cdf of the standard Normal distribution.

Normal Approximation to Binomial

If $X \sim$ Binomial(n,p), for large n, $P(X \leq k) \approx P(Y \leq k)$, where $Y \sim N(np, np(1-p))$.

Statistics: Tests and Confidence Intervals

- **Test statistic:** $\frac{\text{estimate}-\text{hypothesized value}}{\text{se(estimate)}}$.
- **Confidence interval:** estimate $\pm$ (critical value) $\times$ se(estimate).

Other Mathematical Formulas

- **Factorial:** $n! = n(n-1)(n-2)\cdots 1$.
- **Binomial coefficient:** $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
- **Newton's binomial theorem:** $(a+b)^n = \sum_{k=0}^n \binom{n}{k}a^k b^{n-k}$.
- **Geometric sum:** $1 + a + a^2 + \cdots + a^n = \frac{1-a^{n+1}}{1-a}$ for $a \neq 1$.
- **Logarithms:**
  1. $\log(xy) = \log x + \log y$.
  2. $e^{\log x} = x$.
- **Exponential:**
  1. $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$.
  2. $e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n$.
  3. $e^{x+y} = e^x e^y$.
- Multinomial Dist.: $P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$
  $= \frac{n!}{x_1!\,x_2!\ldots x_k!}p_1^{x_1}\,p_2^{x_2}\,\ldots\,p_k^{x_k}$

Differentiation

- $(f+g)' = f' + g'$
- $(fg)' = f'g + fg'$
- $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$
- $\frac{d}{dx}x^n = nx^{n-1}$
- $\frac{d}{dx}e^x = e^x$
- $\frac{d}{dx}\log(x) = \frac{1}{x}$

Chain Rule

$(f(g(x)))' = f'(g(x))g'(x)$.

Integration

$\int_a^b f(x)\,dx = [F(x)]_a^b = F(b) - F(a), \quad$ where $F' = f$.

Integration by Parts

$$\int_a^b f(x)G(x)\,dx = [F(x)G(x)]_a^b - \int_a^b F(x)g(x)\,dx, \quad \text{where } F' = f \text{ and } G' = g.$$

Source of Variation	df	SS	Mean Square	F	$P[F > f]$
Treatment	$k-1$	SS_F	$MS_F = \frac{SS_F}{k-1}$	$\frac{MS_F}{MS_E}$	$P-value$
Error	$N-k$	SS_E	$MS_E = \frac{SS_E}{N-k}$	-	-
Total	$N-1$	SS_T	-	-	-