Course Cheatsheet

Section 07: Probability & Statistics

Descriptive Statistics

Measures of Central Tendency

Measure	Formula	When to Use
Mean	$\bar{x} = \frac{\sum x_i}{n}$	Symmetric data, no outliers
Median	Middle value when sorted	Skewed data, outliers present
Mode	Most frequent value	Categorical data

Median calculation: - Odd n: Middle value - Even n: Average of two middle values

Measures of Spread

Measure	Formula
Range	$\text{Max} - \text{Min}$
Sample Variance	$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$
Sample Std. Dev.	$s = \sqrt{s^2}$
Population Variance	$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$

> ! Sample vs. Population
>
> Use $n - 1$ in denominator for sample variance. Use N for population variance.

Five-Number Summary

1. Minimum
2. First Quartile (Q1) - 25th percentile
3. Median (Q2) - 50th percentile
4. Third Quartile (Q3) - 75th percentile
5. Maximum

Interquartile Range (IQR): $\text{IQR} = Q3 - Q1$

Outlier Detection: - Lower fence: $Q1 - 1.5 \times \text{IQR}$ - Upper fence: $Q3 + 1.5 \times \text{IQR}$

Probability Fundamentals

Basic Probability Rules

Rule	Formula
Complement	$P(A') = 1 - P(A)$
Addition	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Multiplication	$P(A \cap B) = P(A) \cdot P(B \mid A)$

For mutually exclusive events: $P(A \cup B) = P(A) + P(B)$

For independent events: $P(A \cap B) = P(A) \cdot P(B)$

Independence Test

Events A and B are independent if and only if:

$$P(A \cap B) = P(A) \cdot P(B)$$

or equivalently:

$$P(A \mid B) = P(A)$$

Conditional Probability

Definition

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

Read as: “Probability of A given B”

Law of Total Probability

If $B_1, B_2, ..., B_n$ partition the sample space:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$$

For two partitions:

$$P(A) = P(A \mid B) \cdot P(B) + P(A \mid B') \cdot P(B')$$

Bayes’ Theorem

Formula

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Expanded Form

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid A') \cdot P(A')}$$

> 💡 Contingency Table Method
>
> For complex Bayes problems, create a hypothetical population (e.g., 10,000 people), fill in a 2×2 table using the given probabilities, then read answers directly from counts!

Medical Testing Terminology

Metric	Definition	Formula
Sensitivity	True positive rate	$P(+ \mid D)$
Specificity	True negative rate	$P(- \mid D')$
Prevalence	Disease rate	$P(D)$
PPV	Positive predictive value	$P(D \mid +)$
NPV	Negative predictive value	$P(D' \mid -)$

False positive rate: $P(+ \mid D') = 1 - \text{Specificity}$

False negative rate: $P(- \mid D) = 1 - \text{Sensitivity}$

PPV and NPV Formulas

$$\text{PPV} = P(D \mid +) = \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1 - \text{Specificity}) \times (1 - \text{Prevalence})}$$

$$\text{NPV} = P(D' \mid -) = \frac{\text{Specificity} \times (1 - \text{Prevalence})}{\text{Specificity} \times (1 - \text{Prevalence}) + (1 - \text{Sensitivity}) \times \text{Prevalence}}$$

> ⚠ PPV Depends on Prevalence
>
> Low prevalence leads to low PPV, even with high sensitivity and specificity!

Combinatorics

Factorial

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$$

Special cases: $0! = 1$, $1! = 1$

Permutations (Order Matters)

All n objects: $P(n) = n!$

r objects from n (without replacement):

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

Combinations (Order Doesn't Matter)

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

> Permutation vs. Combination
>
> - Permutation: Arranging people in a line (order matters)
> - Combination: Selecting a committee (order doesn't matter)

Counting Principles

Principle	Description
Multiplication	If task A has m ways and task B has n ways, together: $m \times n$
Addition	If choices are mutually exclusive, add the counts

Contingency Tables

Reading a 2x2 Table

	B	B'	Total
A	a	b	a+b
A'	c	d	c+d
Total	a+c	b+d	n

Probability Calculations from Tables

Probability	Formula	Name
$P(A)$	$\frac{a+b}{n}$	Marginal
$P(B)$	$\frac{a+c}{n}$	Marginal
$P(A \cap B)$	$\frac{a}{n}$	Joint
$P(A \mid B)$	$\frac{a}{a+c}$	Conditional

Probability	Formula	Name
$P(B \mid A)$	$\frac{a}{a+b}$	Conditional

Independence Test in Tables

Events are independent if:

$$P(A \cap B) = P(A) \times P(B)$$

Or equivalently, observed cell count equals expected:

$$\text{Expected} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

Binomial Distribution

Conditions

1. Fixed number of trials n
2. Two outcomes: Success (p) or Failure (1-p)
3. Independent trials
4. Constant probability p

Probability Mass Function

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Key Formulas

Measure	Formula
Expected Value	$\mu = E[X] = np$
Variance	$\sigma^2 = np(1-p)$
Standard Deviation	$\sigma = \sqrt{np(1-p)}$

Common Probability Calculations

Question	Calculation
Exactly k	$P(X = k)$
At most k	$P(X \leq k) = \sum_{i=0}^{k} P(X = i)$
At least k	$P(X \geq k) = 1 - P(X \leq k-1)$
Between a and b	$P(a \leq X \leq b) = \sum_{i=a}^{b} P(X = i)$

Complement Rule

For "at least" problems, often easier to calculate: $P(X \geq k) = 1 - P(X < k) = 1 - P(X \leq k-1)$

"At Least One" Strategy

For $P(\text{at least one success in } n \text{ trials})$, use the complement: $P(X \geq 1) = 1 - P(X = 0) = 1 - (1-p)^n$

Geometric Distribution

Definition

Probability that the first success occurs on trial n:

$$P(X = n) = (1-p)^{n-1} \cdot p$$

Key Formulas

Measure	Formula
Expected Trials	$E[X] = \frac{1}{p}$
P(First success by trial n)	$P(X \leq n) = 1 - (1-p)^n$

Normal Distribution

The 68-95-99.7 Rule

For normal distributions:

- 68% of data within $\mu \pm 1\sigma$
- 95% of data within $\mu \pm 2\sigma$
- 99.7% of data within $\mu \pm 3\sigma$

Normal Approximation to Binomial

When $np \geq 5$ and $n(1-p) \geq 5$:

$$\text{Binomial }(n,p) \approx \text{Normal }\left(\mu = np, \sigma = \sqrt{np(1-p)}\right)$$

Problem-Solving Strategies

For Bayes Problems

1. Identify what you need: Usually $P(D \mid +)$ or $P(D \mid -)$
2. Extract given information: sensitivity, specificity, prevalence

3. Calculate $\boldsymbol{P}(+)$ or $\boldsymbol{P}(-)$ using law of total probability
4. Apply Bayes' theorem
5. Interpret the result

For Binomial Problems

1. Verify binomial conditions are met
2. Identify n, p, and k
3. Translate question: "exactly", "at most", "at least"
4. Calculate using appropriate formula
5. Use complement rule when helpful

Alternative: Contingency Table Method

For Bayes problems, use a hypothetical population:

1. Choose convenient population size (e.g., 10,000)
2. Fill in table using given probabilities
3. Read probabilities directly from counts

Common Mistakes to Avoid

- Confusing $P(A \mid B)$ with $P(B \mid A)$
- Forgetting to use complement rule for "at least" problems
- Using permutations when combinations are needed (or vice versa)
- Assuming independence without checking
- Mixing up sensitivity with PPV