01 EN Inferential Statistics 2

02 EN Inferential Statistics Confidence Intervals 3

03 EN Inferential Statistics Sales Exceeding a Threshold 4

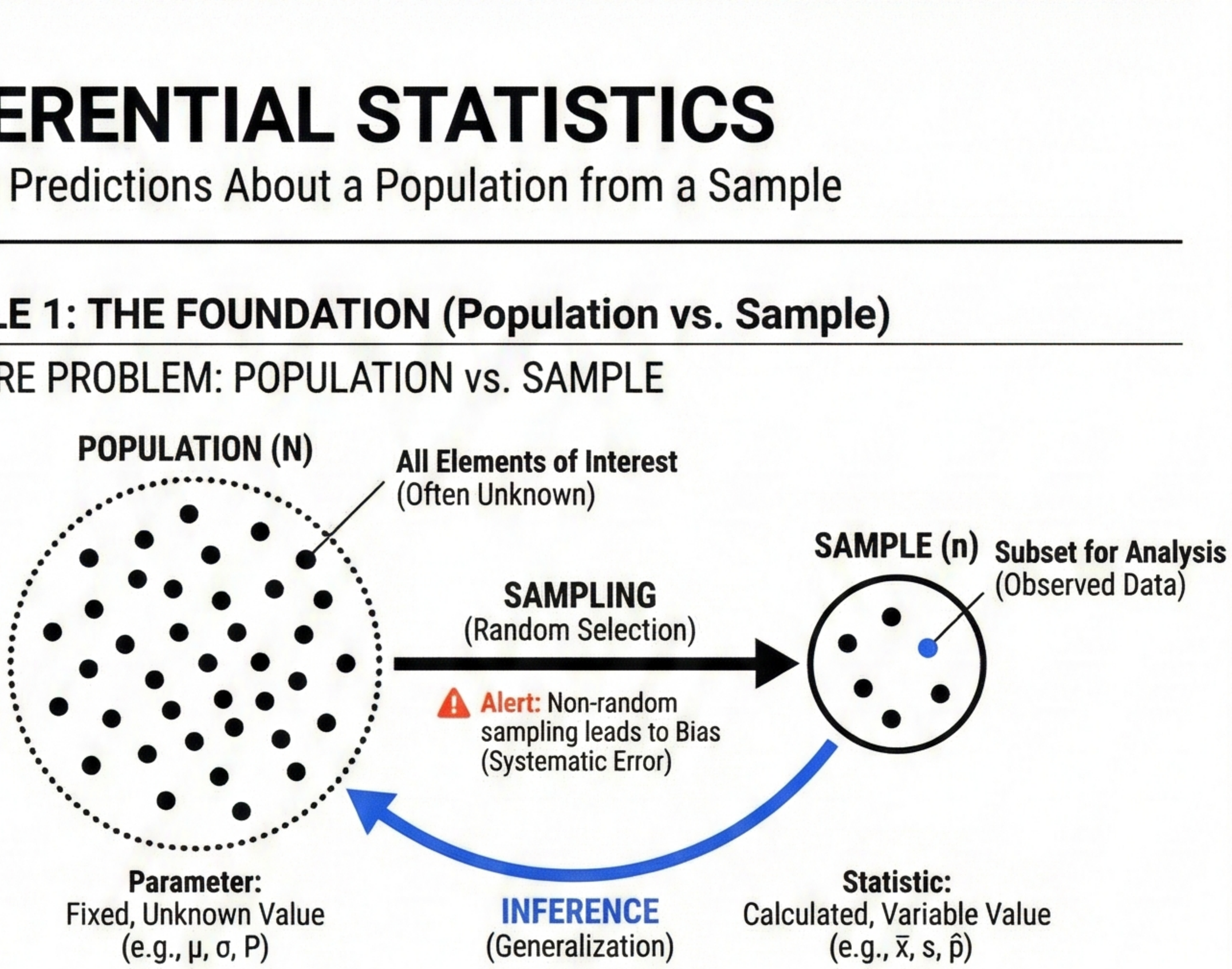
04 EN Inferential Statistics versus Predictive Statistics 5

INFERENTIAL STATISTICS

Making Predictions About a Population from a Sample

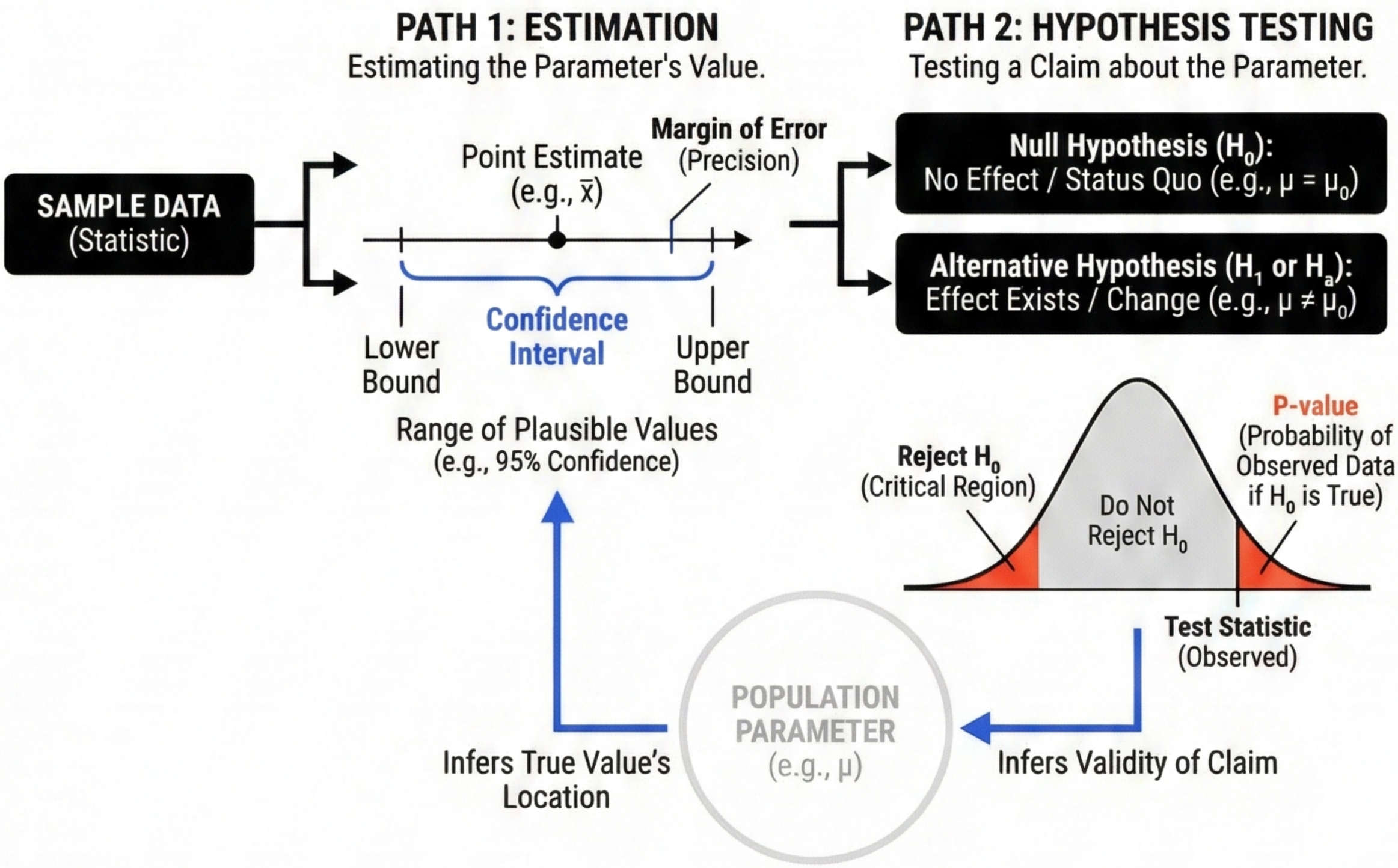
MODULE 1: THE FOUNDATION (Population vs. Sample)

THE CORE PROBLEM: POPULATION vs. SAMPLE



MODULE 2: THE PROCESS (Estimation & Testing)

THE INFERENTIAL PROCESS: TWO MAIN PATHS



MODULE 3: THE OUTCOME (Making Decisions)

THE GOAL: INFORMED DECISION-MAKING

ESTIMATION OUTCOME	HYPOTHESIS TESTING OUTCOME
Provides a likely range for the unknown parameter, quantifying uncertainty.	Provides evidence to either reject or fail to reject a specific claim about the parameter.

Zbigniew Galar

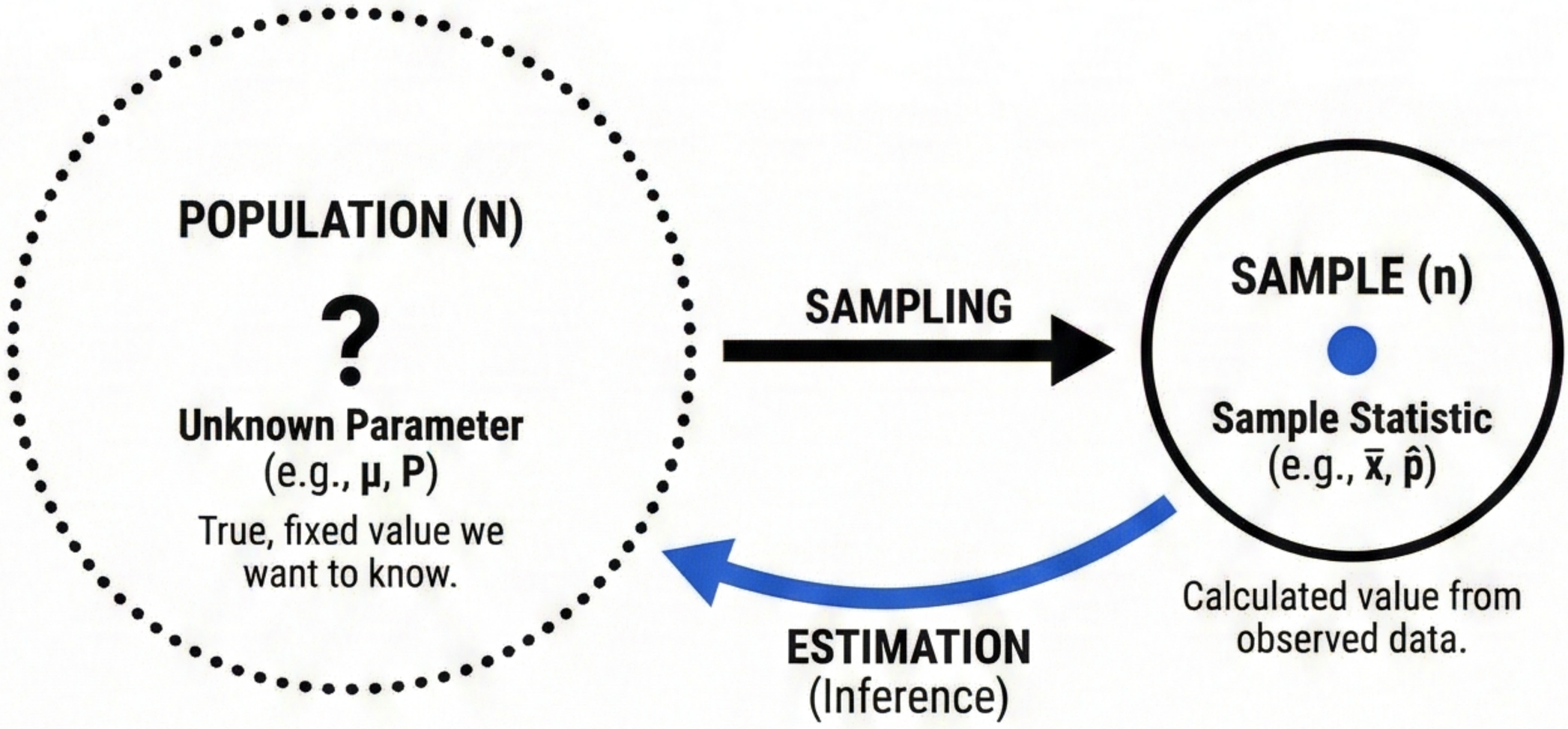
CONFIDENCE INTERVALS

Estimating an Unknown Population Parameter with a Stated Level of Confidence

---

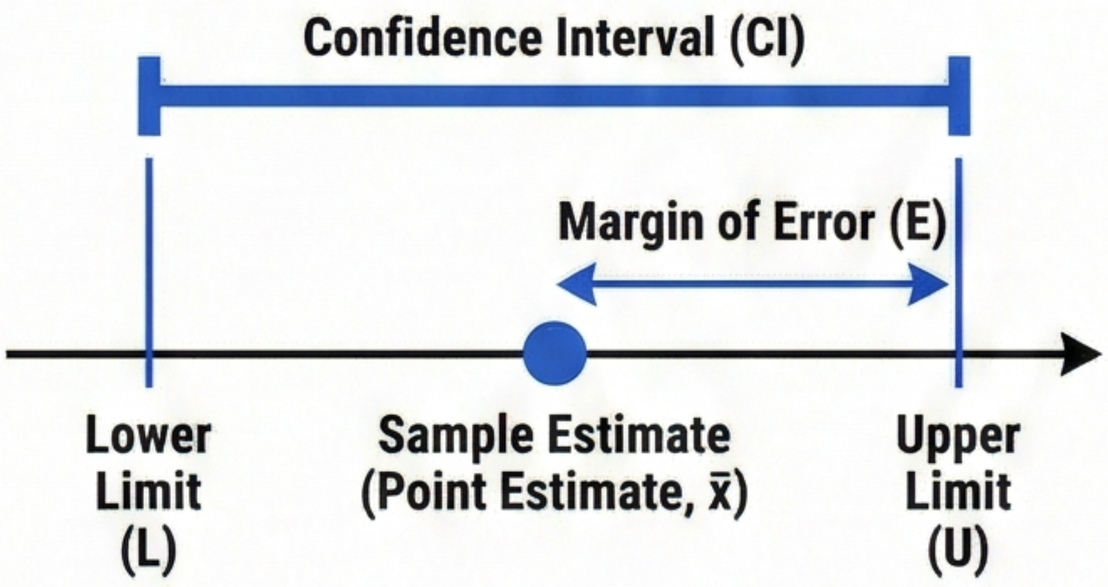
MODULE 1: THE PROBLEM (Population vs. Sample)

THE INFERENCE GAP: ESTIMATING THE UNKNOWN



MODULE 2: THE SOLUTION
(The Interval)

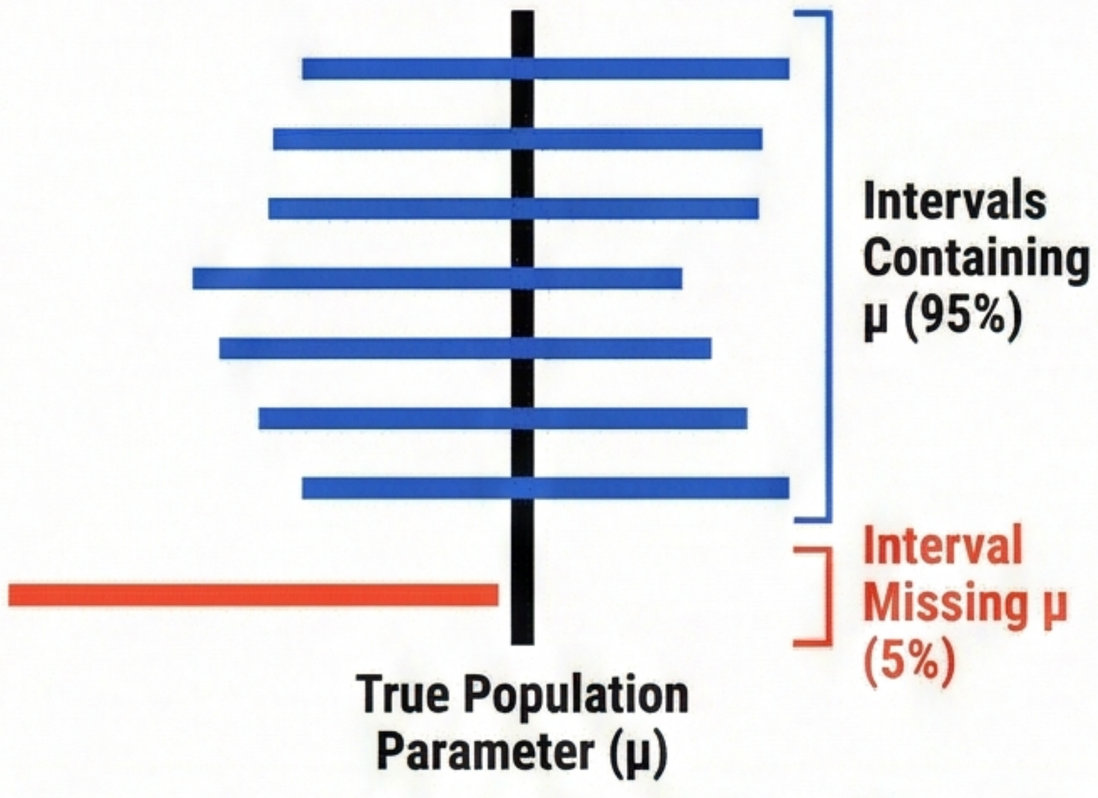
THE CONFIDENCE INTERVAL: A RANGE OF PLAUSIBLE VALUES



CI = Point Estimate ± Margin of Error

MODULE 3: THE INTERPRETATION
(Confidence Level)

INTERPRETING CONFIDENCE (e.g., 95% Level)

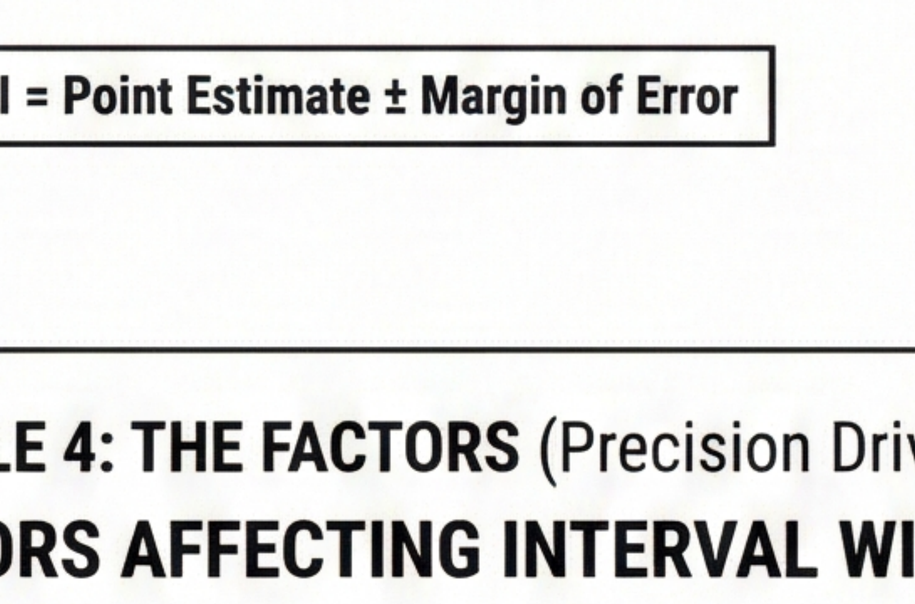


Correct Interpretation: If we repeated the sampling process many times, 95% of the calculated intervals would contain the true population parameter. It is **NOT** a probability that a specific interval contains μ.

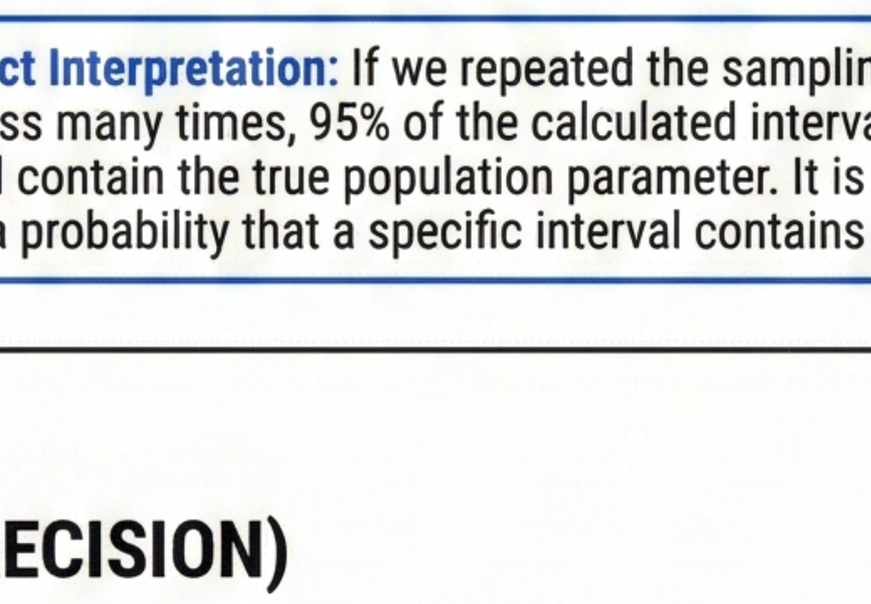
MODULE 4: THE FACTORS (Precision Drivers)

FACTORS AFFECTING INTERVAL WIDTH (PRECISION)

SAMPLE SIZE (n)



VARIABILITY (σ or s)



Zbigniew Galar

HYPOTHESIS TESTING: SALES EXCEEDING A THRESHOLD

Quantifying the probability of a single sale amount surpassing a specific target value using a Gamma distribution model.

1. THE RESEARCH QUESTION & HYPOTHESES

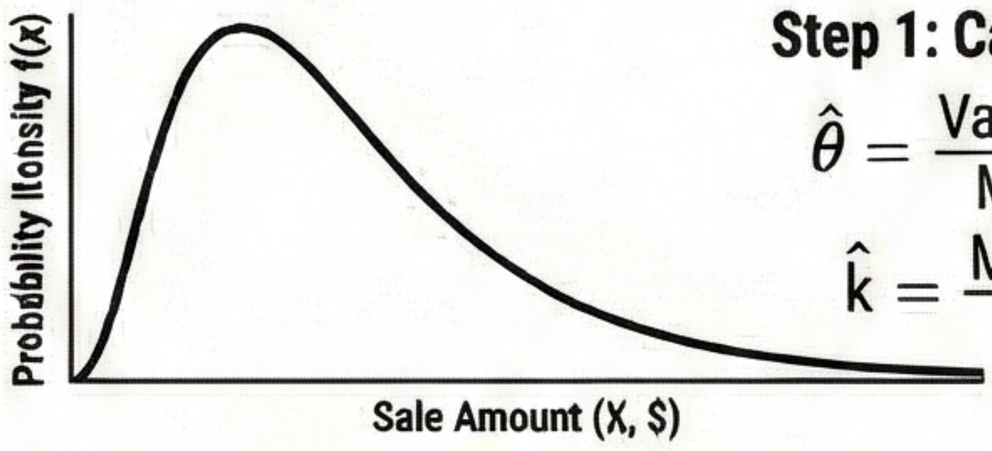
What is the probability of a single sale exceeding a threshold (e.g., $250)?

NULL HYPOTHESIS (H_0: STATUS QUO)	ALTERNATIVE HYPOTHESIS (H_1: THE TEST)
The sale amount (X) is less than or equal to the threshold value.	The sale amount (X) is strictly greater than the threshold value.
H_0: $X \leq \$250$	H_1: $X > \$250$

2. THE STATISTICAL MODEL (MODEL (GAMMA DISTRIBUTION)

SCENARIO: Analyst analyzing daily transaction values.

Mean = $100,
Standard Deviation = $50
(Variance = 2500).



Step 1: Calculate Parameters:

$$\hat{\theta} = \frac{\text{Variance}}{\text{Mean}} = \frac{2500}{100} = 25$$

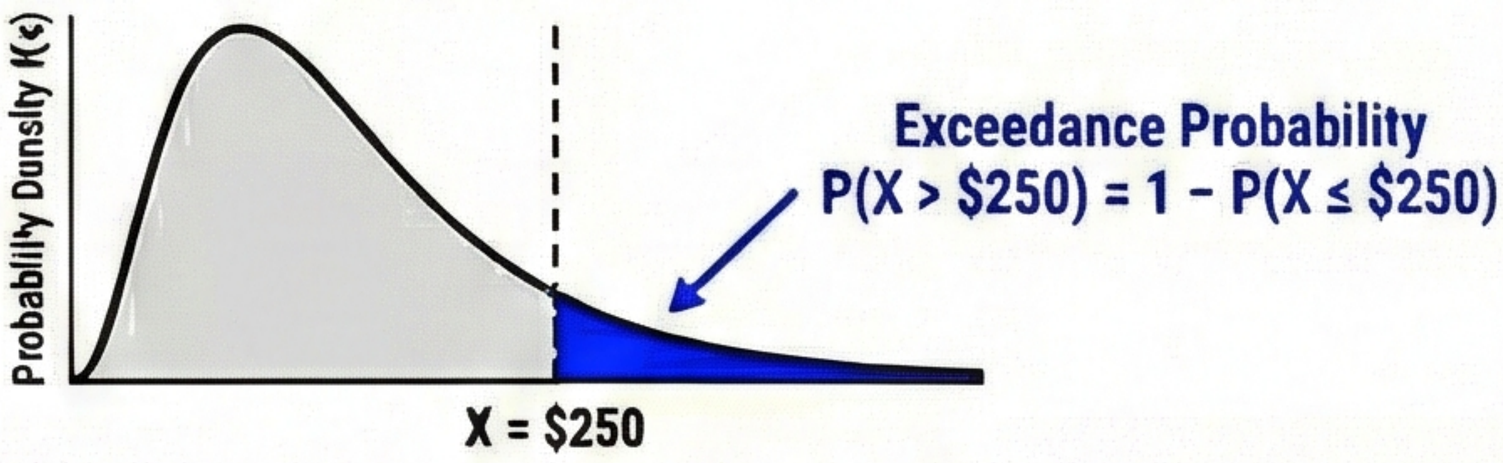
$$\hat{k} = \frac{\text{Mean}}{\hat{\theta}} = \frac{100}{25} = 4$$

SHAPE PARAMETER (α)	SCALE PARAMETER (β)
$\alpha\ (\hat{k}) = 4$	$\beta\ (\hat{\theta}) = 25$

MODULE 2A: BUSINESS INTERPRETATION OF PARAMETERS

BUSINESS INTERPRETATION: DISTRIBUTION SHAPE

Since **α (k) = 4**, the distribution is humped and skewed (not exponential, not normal). This suggests a healthy mix of low and high value tiers, with most sales around the mode, but a significant tail of larger



EXCEL FORMULA & RESULT (THE EXAMPLE)

= 1 - GAMMA.DIST(250, 4, 25, TRUE)

≈ 1 − 0.9897 = **0.0103** (or **1.03%**)

Interpretation: There is a **1.03% probability** that a single sale will exceed $250, based on this model.

4. CONCLUSION & DECISION RULE (MANAGEMENT CONTEXT)

The calculated probability (p-value context) informa management decisions regarding high-value sales.

LOW PROBABILITY (e.g., < 5%)	HIGH PROBABILITY (e.g., > 10%)
If the probability is low, high-value sales are rare but potentially significant anomalies. **Action:** Investigate for unique drivers or potential errors. ALERT	If the probability is relatively high (like 1.03%), high-value sales are a recurrent part of the process. **Action:** Incorporate into regular forecasting and resource planning.

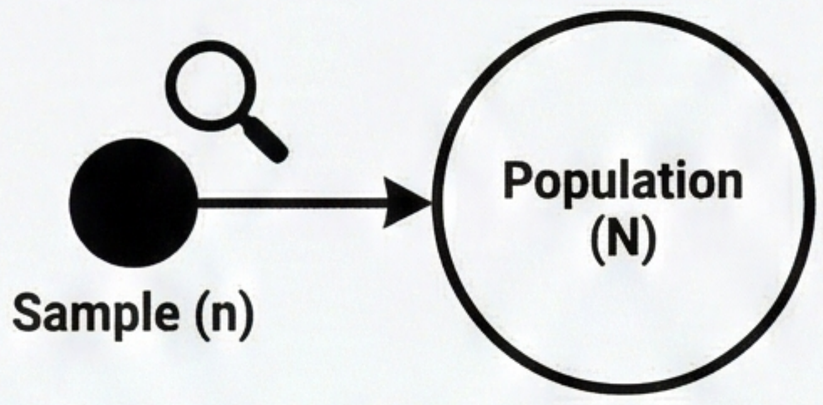
Note: Always validate the distributional assumptions before relying on the model's output.

Zbigniew Galar

INFERENTIAL VERSUS PREDICTIVE STATISTICS

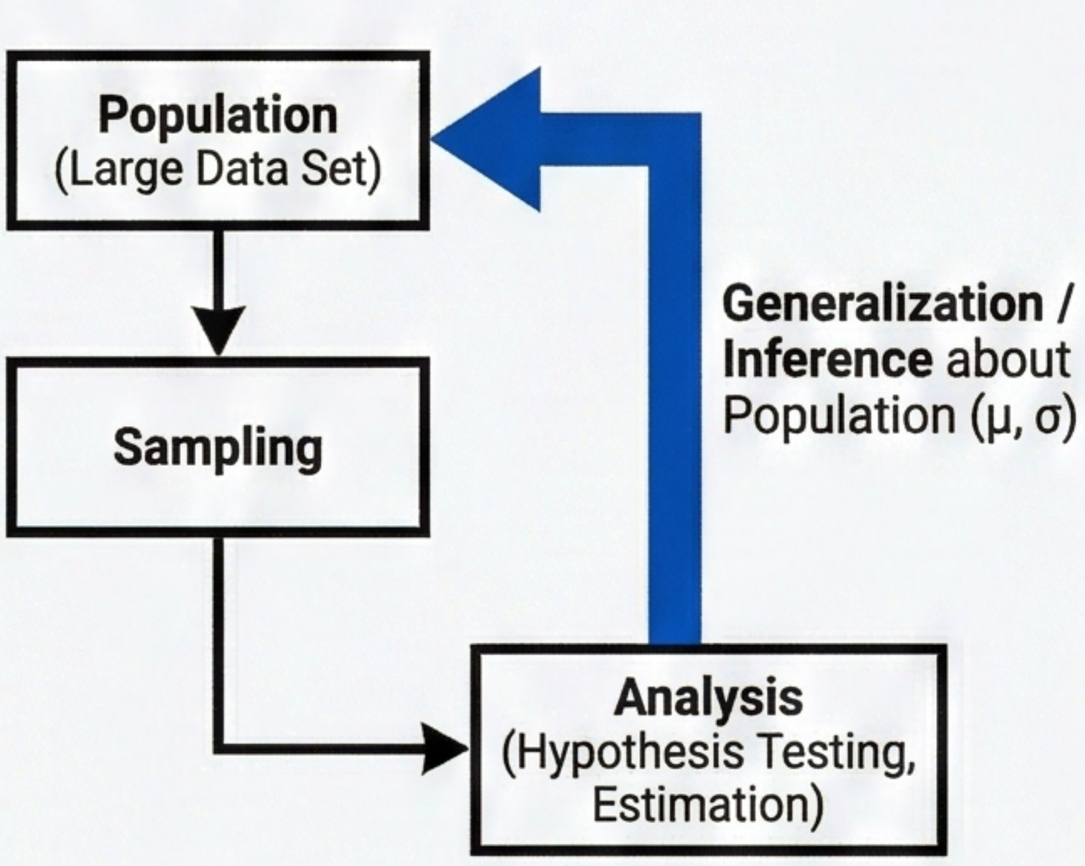
INFERENTIAL STATISTICS

GOAL & FOCUS



Understand population parameters from sample data.
Focus on relationships and causality.

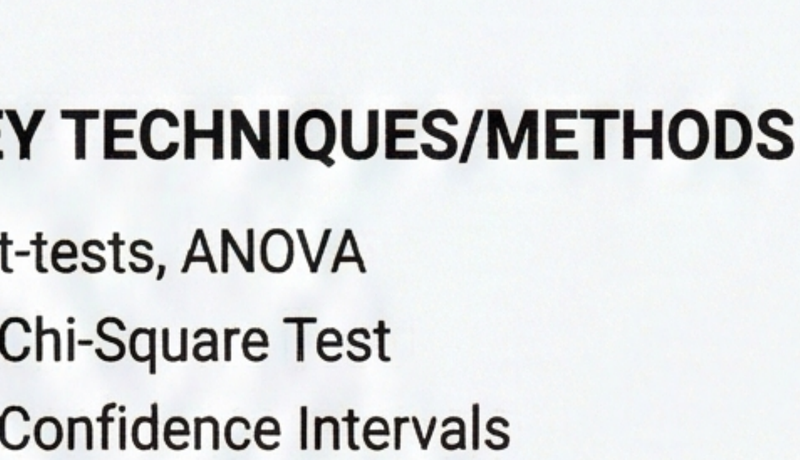
DATA FLOW & PROCESS



KEY TECHNIQUES/METHODS

- t-tests, ANOVA
- Chi-Square Test
- Confidence Intervals
- Linear Regression (for relationship)

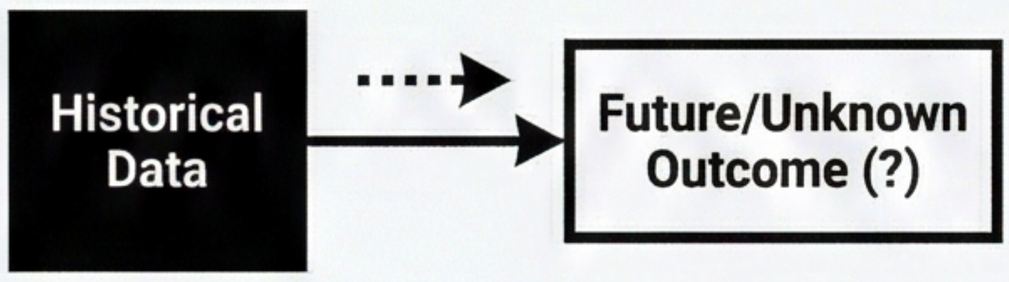
OUTPUT & INTERPRETATION



Estimates with uncertainty.
Explains *_why_* a phenomenon occurs.

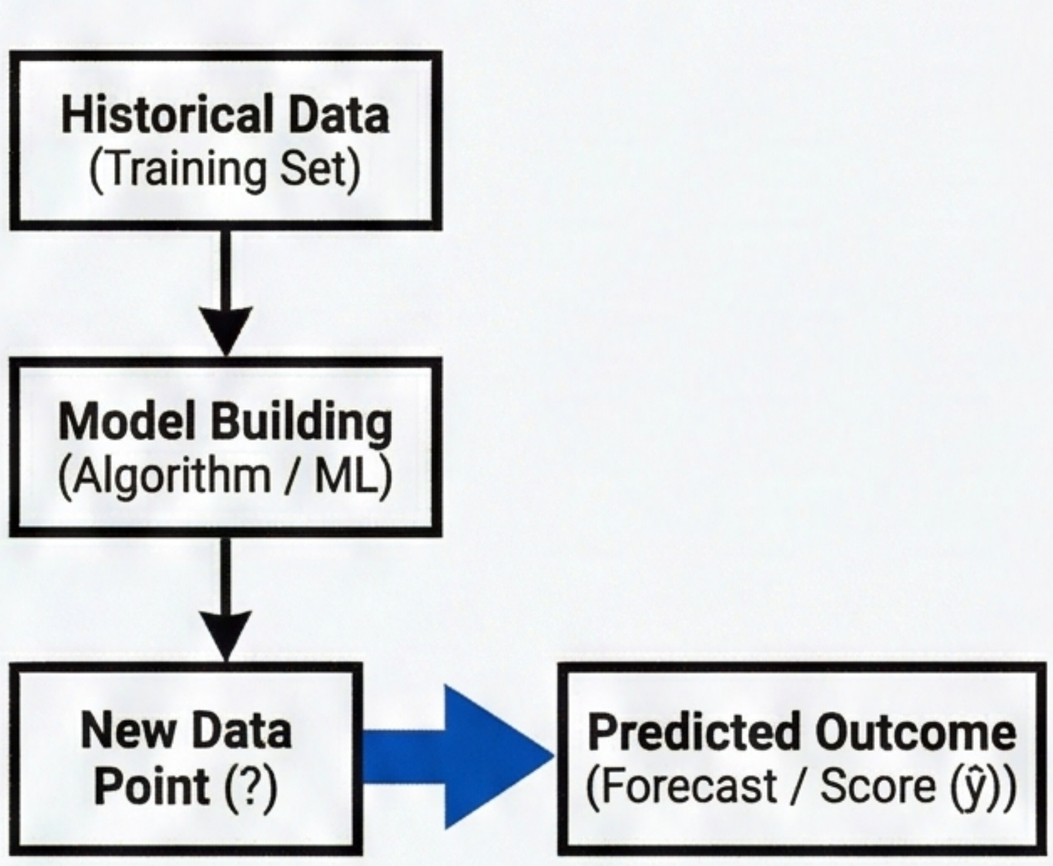
PREDICTIVE STATISTICS

GOAL & FOCUS



Forecast future outcomes for new observations.
Focus on individual data point accuracy.

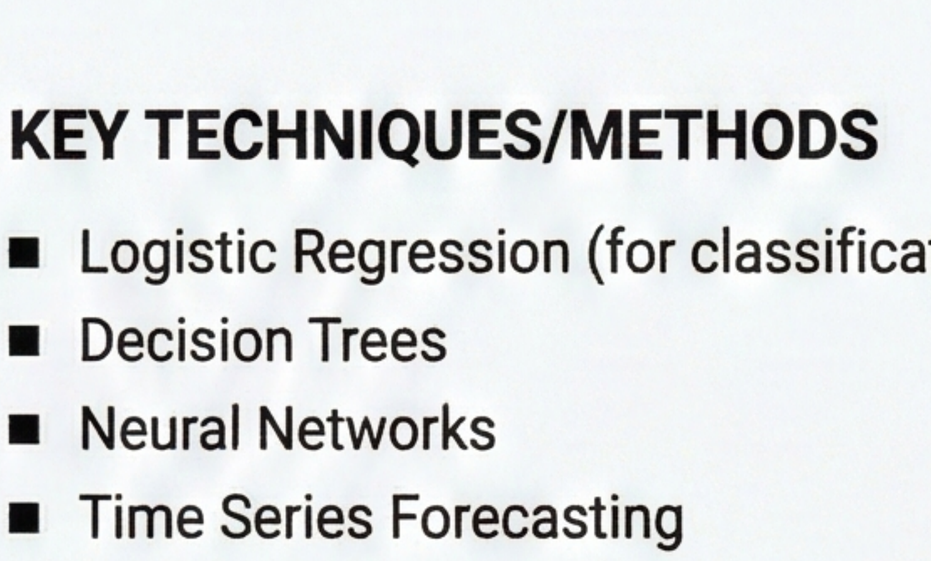
DATA FLOW & PROCESS



KEY TECHNIQUES/METHODS

- Logistic Regression (for classification)
- Decision Trees
- Neural Networks
- Time Series Forecasting

OUTPUT & INTERPRETATION



Specific value or class.
Metrics: Accuracy, RMSE.
Focuses on *_what_* will happen.

Zbigniew Galar