Learn, Visualize, & Analyze

LAFOREST LAB R WORKSHOP

May 12-14th, 2025

DAY3: LEARNING GOALS

1. Understand Statistical Philosophies

- Compare **Frequentist**, **Bayesian**, and **Likelihood** approaches

2. Parametric vs. Non-Parametric Tests

- Choose tests based on data **distribution** and **type**
- Practice checks for normality (Shapiro-Wilk, Levene)

3. Perform Key Statistical Tests in R

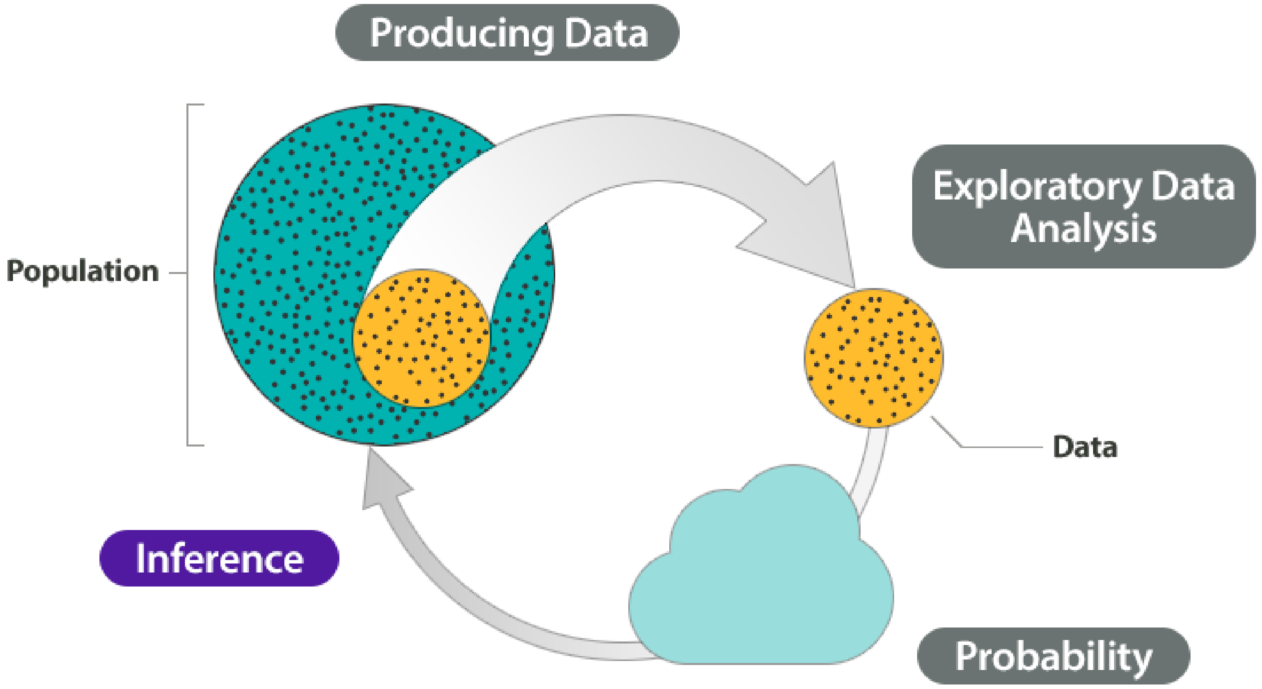
- **t-tests**
- **ANOVA** (one-way, post-hoc tests)
- **Correlation** (Pearson, Spearman)

4. Build & Interpret Linear Models

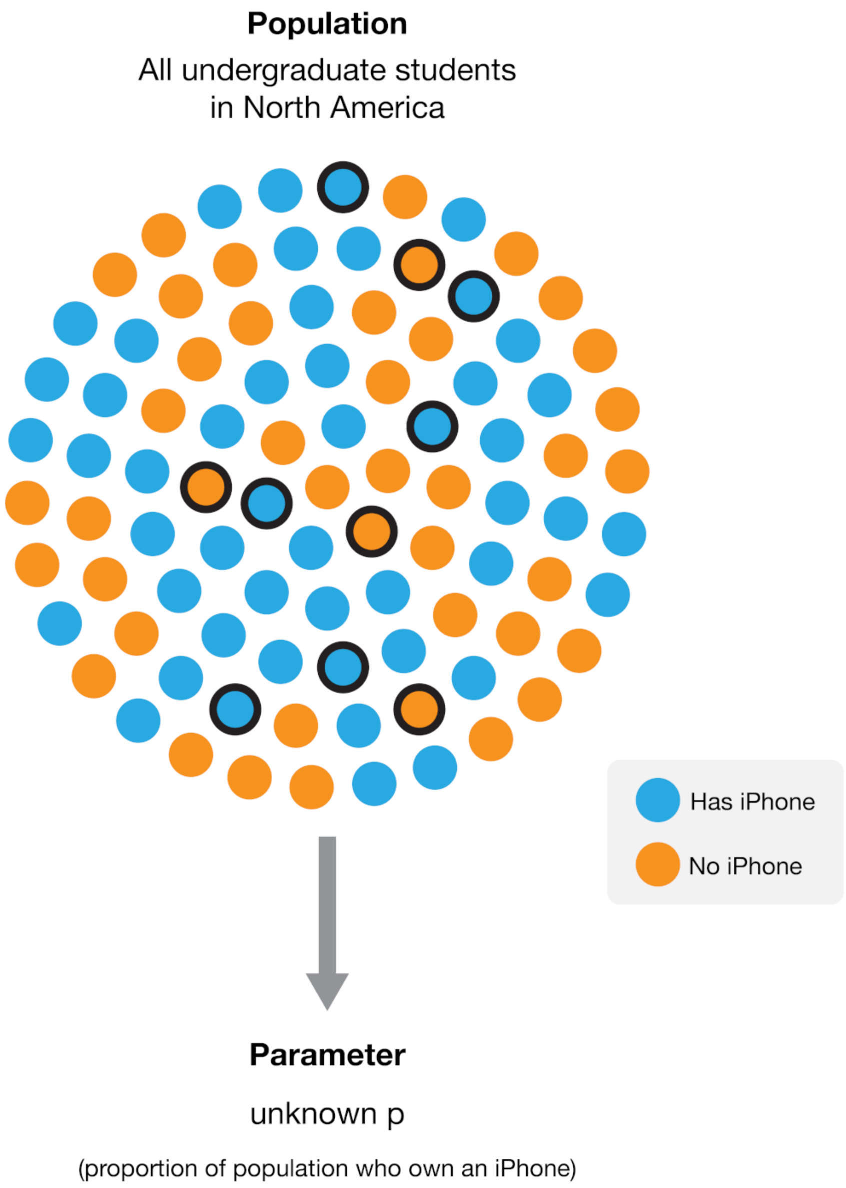
- Simple/multiple **linear regression** (lm())
- **Linear mixed-effects models** (lmer() or nlme) for nested data
- Validate assumptions (residual plots)

LAFOREST LAB

Statistical Inference

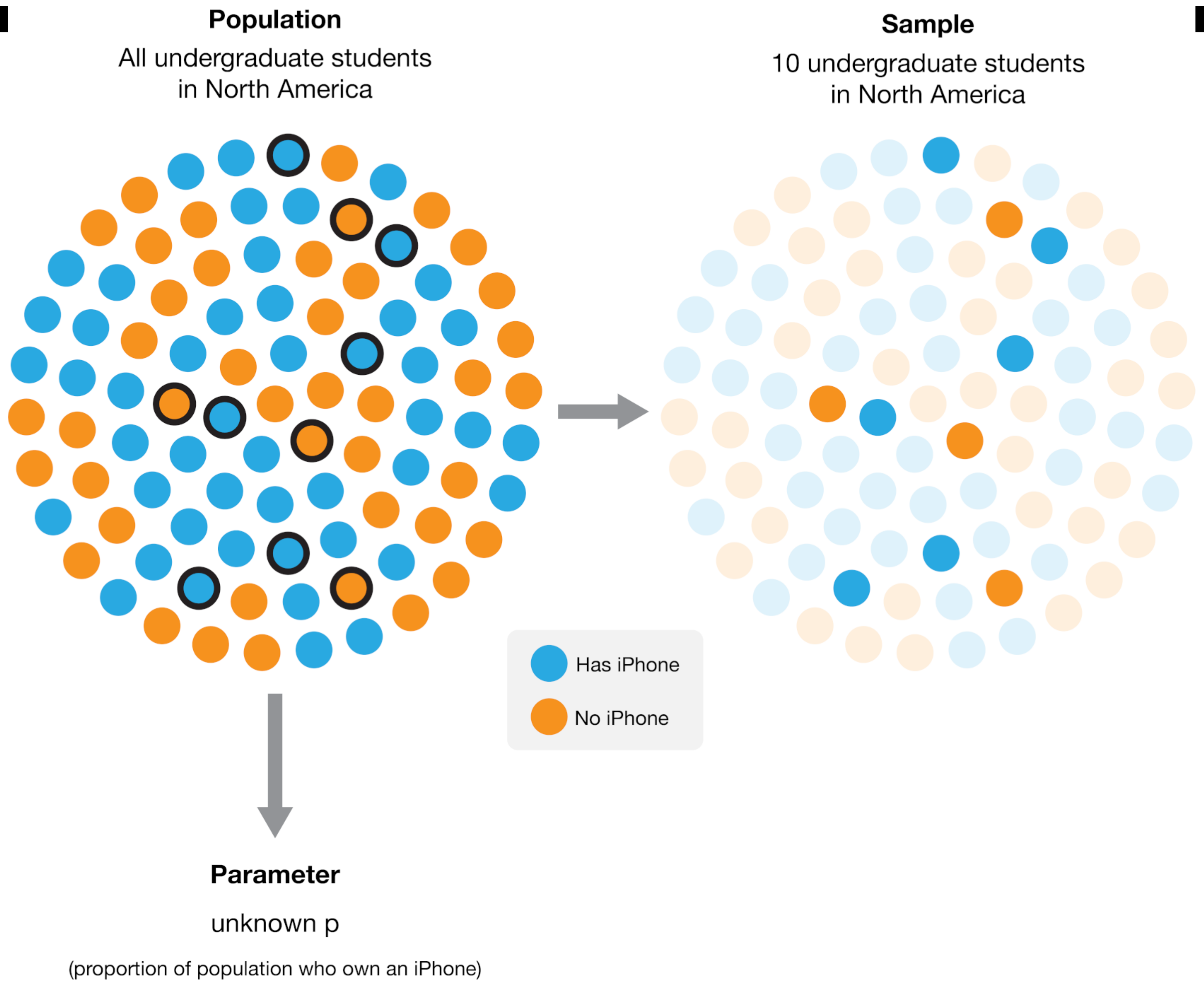


Statistical Inference



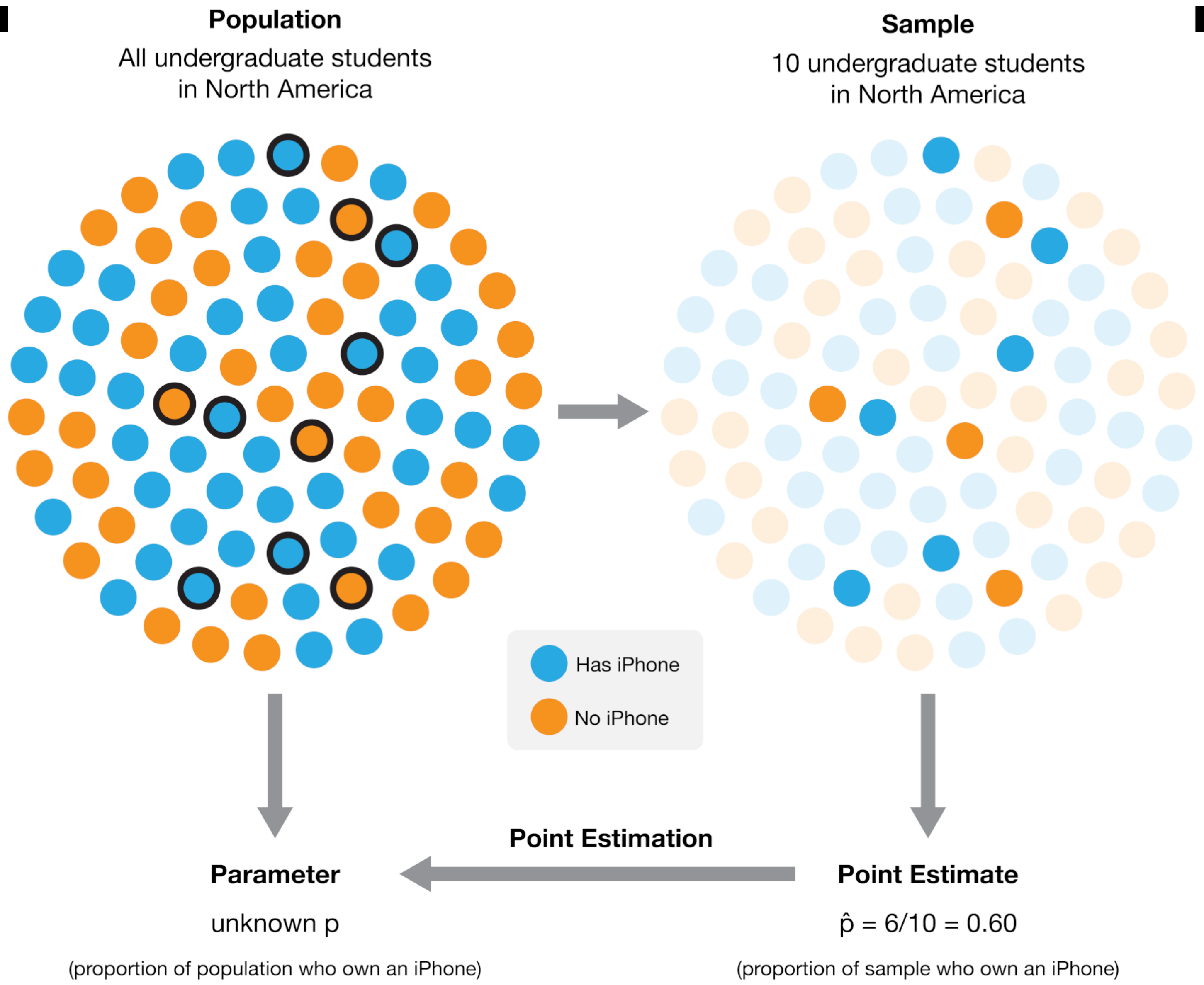
https://python.datasciencebook.ca/inference.html

Statistical Inference



https://python.datasciencebook.ca/inference.html

Statistical Inference



https://python.datasciencebook.ca/inference.html

Statistical Inference: 3 Key Approaches

1. Frequentist Statistics

- **Core Idea**: Probability = long-run frequency of events.
- **Focus**: *P(data | hypothesis)* (e.g., p-values, confidence intervals).
- **Tools**: Hypothesis tests (t-tests, ANOVA), Null Hypothesis Significance Testing.

- **Strengths:** Objective, widely used, standardized.
- **Limitations:** Ignores prior knowledge; misinterpreted p-values.

Statistical Inference: 3 Key Approaches

1. Frequentist Statistics

- **Core Idea**: Probability = long-run frequency of events.
- **Focus**: *P(data | hypothesis)* (e.g., p-values, confidence intervals).
- **Tools**: Hypothesis tests (t-tests, ANOVA), Null Hypothesis Significance Testing.

- **Strengths:** Objective, widely used, standardized.
- **Limitations:** Ignores prior knowledge; misinterpreted p-values.

2. Bayesian Statistics

- **Core Idea**: Probability = degree of belief (updated with data).
- **Focus**: *P(hypothesis | data)* using Bayes' Theorem.
- **Tools**: Posterior distributions, credible intervals, MCMC.

- **Strengths:** Incorporates prior knowledge; intuitive interpretations.
- **Limitations:** Computationally intensive; subjective priors.

LAFOREST LAB

Statistical Inference: 3 Key Approaches

1. Frequentist Statistics

- **Core Idea**: Probability = long-run frequency of events.
- **Focus**: *P(data | hypothesis)* (e.g., p-values, confidence intervals).
- **Tools**: Hypothesis tests (t-tests, ANOVA), Null Hypothesis Significance Testing.

- **Strengths:** Objective, widely used, standardized.
- **Limitations:** Ignores prior knowledge; misinterpreted p-values.

2. Bayesian Statistics

- **Core Idea**: Probability = degree of belief (updated with data).
- **Focus**: *P(hypothesis | data)* using Bayes' Theorem.
- **Tools**: Posterior distributions, credible intervals, MCMC.

- **Strengths:** Incorporates prior knowledge; intuitive interpretations.
- **Limitations:** Computationally intensive; subjective priors.

3. Likelihood-Based Inference

- **Core Idea**: Focus on the *likelihood function* (support for hypotheses given data).
- **Focus**: Compare models via likelihood ratios (no priors).
- **Tools**: MLE, AIC/BIC, profile likelihoods.

- **Strengths:** Flexible; bridges Frequentist and Bayesian ideas.
- **Limitations:** Less intuitive for complex models.

LAFOREST LAB

Statistical Inference: 3 Key Approaches

1. Frequentist Statistics

- **Core Idea**: Probability = long-run frequency of events.
- **Focus**: *P(data | hypothesis)* (e.g., p-values, confidence intervals).
- **Tools**: Hypothesis tests (t-tests, ANOVA), Null Hypothesis Significance Testing.

- **Strengths:** Objective, widely used, standardized.
- **Limitations:** Ignores prior knowledge; misinterpreted p-values.

2. Bayesian Statistics

- **Core Idea:** Probability = degree of belief (updated with data).
- **Focus**: *P(hypothesis | data)* using Bayes' Theorem.
- **Tools**: Posterior distributions, credible intervals, MCMC.

- **Strengths:** Incorporates prior knowledge; intuitive interpretations.
- **Limitations:** Computationally intensive; subjective priors.

3. Likelihood-Based Inference

- **Core Idea**: Focus on the *likelihood function* (support for hypotheses given data).
- **Focus:** Compare models via likelihood ratios (no priors).
- **Tools**: MLE, AIC/BIC, profile likelihoods.

- **Strengths**: Flexible; bridges Frequentist and Bayesian ideas.
- **Limitations**: Less intuitive for complex models.

LAFOREST LAB

Parametric vs. non-parametric tests

Feature	Parametric Tests	Non-Parametric Tests
Assumptions	Normality, equal variance, independence	Fewer assumptions (ordinal/any distribution)
Data Types	Continuous, normally distributed	Ordinal, skewed, or small samples
Power	Higher power (when assumptions met)	Robust but less powerful
Examples	t-tests, ANOVA, Pearson’s r	Wilcoxon, Kruskal-Wallis, Spearman’s ρ

Basic parametric tests

Test	Predictor (X)	Outcome (Y)	Answer	R Function
t-test	2 groups	Continuous	"Are means different?"	t.test(y ~ group, data)
ANOVA	3+ groups	Continuous	"Which groups differ?"	aov(y ~ group, data)
Correlation	Continuous	Continuous	"How strong is the linear link?"	cor.test(data$x, data$y)
Regression	1+ continuous/cat.	Continuous	"How does X affect Y?"	cor.test(data$x, data$y)

Basic non-parametric tests

Parametric Test	**Non-Parametric Alternative**	**When to Use It**	**R Function**
Independent t-test	**Mann-Whitney U test**	Compare 2 independent groups (ordinal/skewed)	wilcox.test(y ~ group)
Paired t-test	**Wilcoxon signed-rank test**	Compare paired measurements (non-normal)	wilcox.test(y1, y2, paired=TRUE)
One-way ANOVA	**Kruskal-Wallis test**	Compare 3+ independent groups	kruskal.test(y ~ group)
Pearson correlation	**Spearman’s rank correlation**	Assess monotonic relationships	cor.test(x, y, method="spearman")
Repeated-measures ANOVA	**Friedman test**	Compare 3+ paired groups	friedman.test(y ~ group | subject)

Let’s code!