

# Chapter 1: Measurement, Statistics, and Research

Ovande Furtado, Jr., Ph.D.

## 1 Module 1: The Foundation

This module establishes the bedrock of empirical research by turning observations into reliable data.

- We will explore:
  - The precise process of measurement
  - Systems used to ensure universal understanding
  - Classification scales that determine what we are allowed to do with our data

### ! Important

Before building the study, we need to understand the “materials” (data) and how they are created.

- Measurement transforms **abstract** observations into concrete, objective data.
- Example: Grip strength measured with a dynamometer:
  - Force output is compared to a standard unit (e.g., pounds or kilograms).
  - Result: a quantified value (e.g., 52.3 kg).
- Opening: Emphasize that “measurement” turns observations into objective, shareable numbers. Say: “We’re moving from ‘I think’ to ‘we can show’.”
- Example: Grip strength measured with a dynamometer. Describe briefly: “We use a calibrated dynamometer, read in kg or lb — e.g., 52.3 kg — and note the unit, device, and protocol (standing/sitting, dominant hand).”
- Measurement vs Statistics vs Evaluation: Walk the class through a concrete chain: “Measure VO<sub>2</sub>max (45 mL/kg/min) → summarize with mean/SD or compare groups → evaluate whether this is ‘good’ for age/sex and design training accordingly.”

- Process of measurement: For each step, give a one-sentence script to read aloud:
  - Define object: “What exactly are we measuring? (e.g., cyclist peak power over 5s).”
  - Define standard: “Which unit or device defines that measurement? (e.g., Watt via calibrated ergometer).”
  - Compare: “How do we collect the data? (protocol, calibration, number of trials).”
  - State relationship: “How will we report it? (single best value, average, best of trials).”
- Classification scales: Give quick examples for each scale and why it matters for analysis:
  - Nominal: sport type (soccer, rowing) — counts/proportions only.
  - Ordinal: race finish place or Likert rating — order but unknown gaps.
  - Interval: Celsius temperature — means make sense, but ratios do not.
  - Ratio: time, mass, force — true zero allows ratio comparisons (e.g., 100 kg = 2×50 kg).
- Teaching tip / Activity: “Have students classify three quick examples (body mass, perceived exertion, jersey number) and explain which tests are appropriate.”
- Closing: Reinforce reproducibility: “If you can’t state the object and standard clearly, your measurement can’t be reproduced.”

## 2 What is measurement?

- Measurement is the process of comparing a value to a standard, transforming **abstract** observations into concrete, objective data.
- The terms “I believe, I feel, I observe” become “The measured value is X units.”
- Kinesiology example:
  - Grip strength is measured with a dynamometer
  - The force output is compared to a standard unit (pounds or kilograms)
  - The result is a quantified value (e.g., 52.3 kg)

### ! Important

**Reproducibility:** Note and include in the report the device model, calibration date, and protocol so others can repeat the measurement.

### i Note

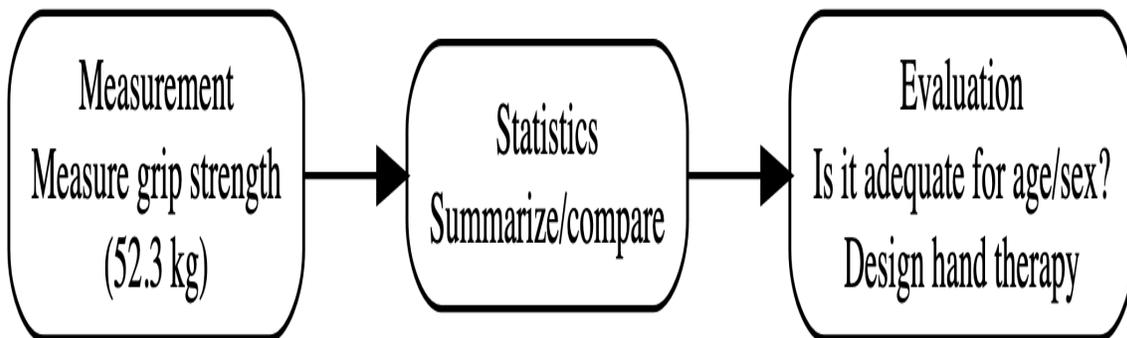
Quantification is the first step from **anecdotal** observation to **scientific inquiry**.

- Key point: Measurement turns observation into objective, comparable data.
- Example to say aloud: “Grip strength measured with a calibrated dynamometer — report the value and unit (e.g., 52.3 kg), the device, and the testing protocol (position, hand, trial count).”
- Emphasize reproducibility: note device model, calibration date, and protocol so others can repeat the measurement.
- Teaching prompt: ask students to list what details they would record when measuring grip strength.

### 3 Measurement vs statistics vs evaluation

These are sequential, not interchangeable:

- **Measurement:** produces data via comparison to a standard
- **Statistics:** organizes and interprets measured data
- **Evaluation:** assigns meaning or worth to the statistical results
- Example:
  - Measure a client’s VO2 max
  - Use statistics to summarize or compare (45 mL/kg/min)
  - Evaluate whether the value is “good” for age and sex, then design training



**i** Note

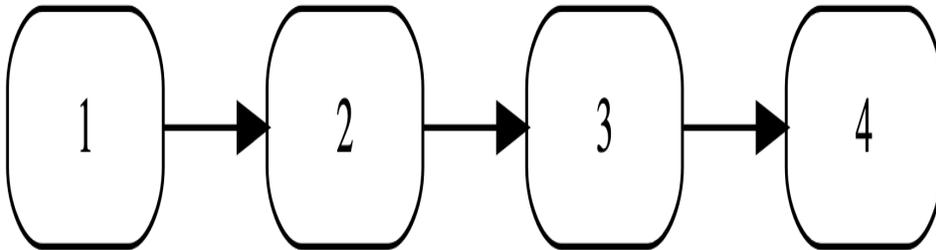
The ethical and scientific risk is confusing objective data collection with subjective interpretation.

- Point to make: Measurement produces data, but interpretation is where bias and values enter.
- Example script: “We record VO2max (a number). How we summarize and judge it is a separate, subjective step.”

- Quick prompt: Ask students how a coach might misinterpret a single measurement.

## 4 Process of measurement (4 steps)

- Assuming that you have a clear research question, the measurement process involves four key steps:



1. **Define object:** Identify and define what is measured (example: athlete’s vertical jump height)
2. **Define standard:** Choose and define the standard (example: centimeters)
3. **Compare:** Collect data by comparing object to standard (example: jump mat or Vertec device)
4. **State relationship:** Quantify the result (example shown: 65 cm)

### **i** Note

If you cannot clearly state the object and standard, your measurement is not reproducible.

- Emphasize reproducibility: define object and standard precisely (device, settings, posture).
- Example cue: “If I don’t state device model and calibration, you can’t repeat my result.”
- Ask: “What three protocol details would you record for a power test?”

## 5 Test your knowledge: Measurement process

- You want to measure the **peak power output** of cyclists during a **5-second sprint** on a **cycle ergometer**.
- Outline the four steps of the measurement process for this scenario.

Using the ClassShare App, submit your answers.

Answer

1. **Define object:** Peak power output during a 5-second sprint
2. **Define standard:** Watts (W)
3. **Compare:** Use a calibrated cycle ergometer to measure power output during the sprint
4. **State relationship:** Quantify the result (example shown: 1,150 Watts)

## 6 Classification of data

- Once a dataset is collected, it must be classified according to its properties.
- This **classification** determines which **statistical analyses** are appropriate.
- Nominal is the most basic level; ratio is the most advanced.
- Each scale builds on the previous one by adding more properties - see next slide.
- Understanding scale is essential for valid data analysis.
- Four primary scales of measurement:

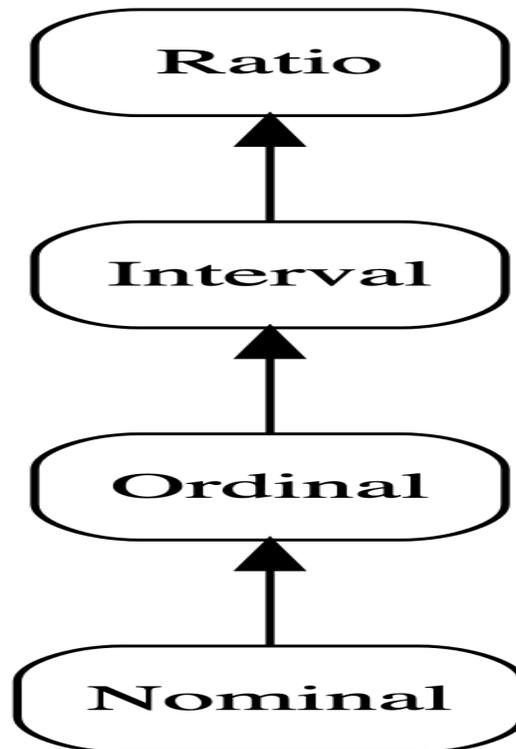


Figure 1: Hierarchy of measurement scales in kinesiology research

### **i** Note

This is a major decision point. Misclassifying scale can lead to invalid analysis choices.

- Explain consequences: wrong scale → wrong statistics (e.g., means on ordinal data).
- Example: Treating Likert responses as interval can overstate precision.
- Activity: Have students convert examples to appropriate tests (chi-square, median test, t-test).

## 7 Test your knowledge: scale classification

- For each of the following examples, identify the measurement scale (nominal, ordinal, interval, or ratio):
  1. Sport type (e.g., soccer, basketball, swimming)
  2. Finish place in a race (1st, 2nd, 3rd)
  3. Body temperature in Celsius
  4. Body mass in kilograms

Using the ClassShare App, submit your answers.

Answer

1. Nominal
2. Ordinal
3. Interval
4. Ratio

## 8 Measurement scales at a glance

Scale	Core idea	What you can do	Example (kinesiology)
Nominal	Categories only	counts, proportions	sport type, group labels
Ordinal	Rank order	medians, ranks	finish place, Likert-type ratings
Interval	Equal intervals, no true zero	add/subtract, means	Celsius temperature
Ratio	Equal intervals and true zero	all arithmetic, ratios	mass, time, distance, power

### **i** Note

If 0 means none of the quantity, it is ratio. If 0 is arbitrary, it is interval.

- Clarify: True zero enables multiplicative statements (ratios). Use examples: mass vs temperature.
- Teaching prompt: “Give one measure that’s interval and one that’s ratio in sports science.”

## 9 Scale pitfalls and examples

- **Nominal**: labels only; numbers are arbitrary (example: “1” gymnasts, “2” weightlifters)
- **Ordinal**: ordered, but gaps are unknown (example: 1st vs 2nd place could differ by milliseconds or seconds)
- **Interval**: equal steps, but 0 is not absence (example: 0°C does not mean “no heat”)
- **Ratio**: includes a true zero; ratio statements make sense (example: 100 kg is twice 50 kg)

### **i** Note

Using parametric tests that assume interval or ratio data on ordinal-only data is a methodological flaw that can invalidate conclusions.

- Reinforce: Match test assumptions to data scale before choosing analyses.
- Example note: For ordinal Likert scales, consider medians, nonparametric tests, or justified interval treatment.

---

## 10 Module 2: The Blueprint

- Research Designs and Variable Roles
  - A research design is the strategic blueprint that dictates how a scientific question will be answered.
- The choice of design is not arbitrary. It must match the problem you intend to solve.
- Primary design styles:
  - Historical
  - Observational
  - Experimental

### **i** Note

Move from “materials” (measurement) to the architectural plan (design).

- Transition script: “Now that we have data, how do we structure a study to answer a question?”
- Example question to pose: “Do we want causation (experiment) or description (observational)?”

## 11 Research design and statistical analysis

- **Historical**
  - Investigates the past to describe and understand events
  - **Example:** analyzing training logs from the 1960s to understand marathon preparation trends
  - Provides context, but cannot establish cause and effect
- **Observational** (descriptive)
  - Describes existing phenomena without manipulation
  - **Example:** survey weekly physical activity levels of office workers
  - Useful for identifying correlations (example: sitting time and low back pain), but does not prove causation
- **Experimental**
  - Manipulates a variable to test cause and effect
  - **Example:** sports drink vs placebo, then measure anaerobic power output (Wingate test)
  - Provides the strongest evidence for causal claims
  - Requires careful control of confounding variables
  - Statistical analyses depend on design:
    - \* Historical: qualitative summaries, thematic analysis
    - \* Observational: correlation, regression
    - \* Experimental: t-tests, ANOVA, regression

### **i** Note

Selecting the right design is the most critical strategic decision because it determines the strength of the evidence you can claim.

- Key point: Design determines inference strength — choose design to match the causal claim you seek.
- Classroom prompt: List one research question and the best matching design.

## 12 Independent and dependent variables

In experimental research, variable roles clarify the cause-effect relationship:

- **Independent variable (cause):** manipulated or controlled by the researcher  
Example: beverage type (sports drink vs placebo)
- **Dependent variable (effect):** measured outcome expected to change  
Example: anaerobic power output during a Wingate test

Read the abstract of this study and identify the independent and dependent variables.  
<https://pubmed.ncbi.nlm.nih.gov/17507739/>

Answer

This is an **experimental** study that manipulates the **independent variable** (carbohydrate vs placebo consumption) and **measures** the **dependent variable** (anaerobic power output on the Wingate Anaerobic Test).

### **i** Note

We will use DV for dependent variable and IV for independent variable throughout the course.

- Teaching script: Define IV (manipulated) and DV (measured outcome) with the sports-drink example.
- Quick check: “Identify IV and DV in a simple training study you know.”

## 13 Predictor and Criterion variables

In observational studies (no manipulation):

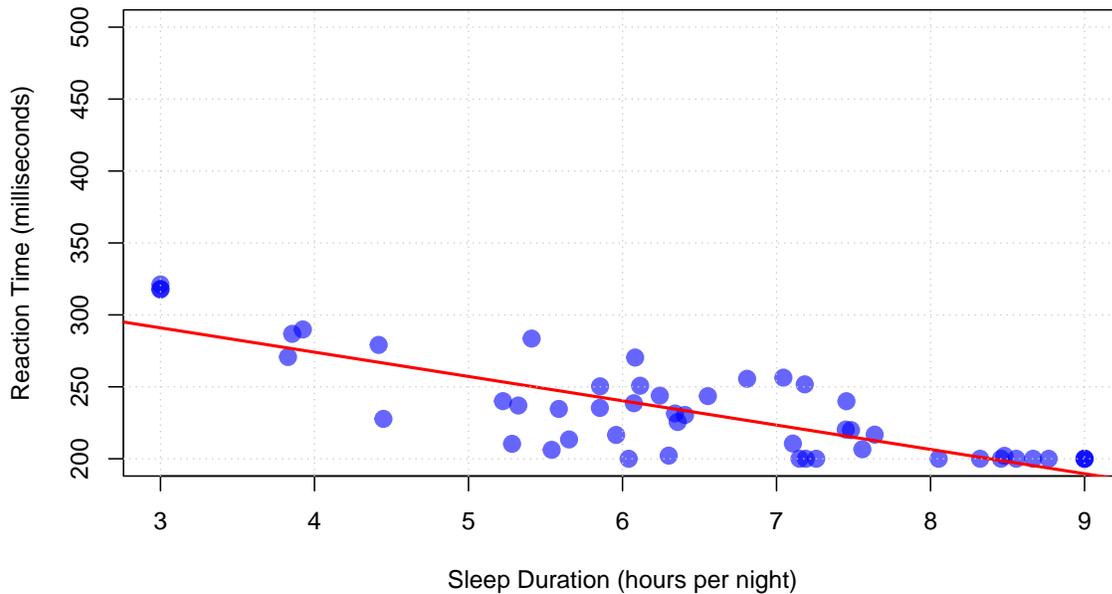
- **Predictor variable:** used to predict an outcome
- **Criterion variable:** the outcome being predicted

Read this study and identify the predictor and criterion variables. <https://pubmed.ncbi.nlm.nih.gov/38782723/>

Answer

- Predictor: chronic short sleep duration (hours of sleep per night)
- Criterion: reaction time

## Sleep Duration and Reaction Time



### **i** Note

**Best practice for scatter plots:** Predictor variable on x-axis, criterion variable on y-axis. This follows the causal flow (predictor → criterion) and makes regression relationships intuitive to interpret.

## 14 Module 3: The Integrity Check

- Evaluating Research Validity
- A blueprint can look elegant, but it is worthless if the resulting structure is unsound.
- Validity refers to the truthfulness and appropriateness of the research process.
- We focus on:
  - Internal validity
  - External validity
  - Common threats that compromise validity

### **i** Note

Validity is not just a technical detail, it is trustworthiness.

- Teaching script: “A beautiful research design is useless if the results are not valid. Validity is about truthfulness—does your study really answer the question?”
- Example: Compare a well-designed but poorly executed study to a house with a beautiful blueprint but a cracked foundation.
- Prompt: Ask students to name a threat to validity they’ve seen in published research or the news.

## **15 Internal and external validity**

- Internal validity: Emphasize control—“Did the intervention, not something else, cause the result?”
- Example: Pre/post test with a control group helps rule out time effects.
- External validity: Generalizability—“Do results apply outside the lab?”
- Example: Lab treadmill test vs. real-world running.
- Prompt: “Which is more important for your own research: internal or external validity? Why?”

### **15.1 Internal validity**

- Degree of control within the experiment
- Confidence that changes are due to the independent variable, not confounds
- Example: a 12-week strength training study with a control group (pre and post tests, no training) helps rule out time and test learning effects

### **15.2 External validity**

- Ability to generalize findings beyond the specific sample and setting
- Example: treadmill oxygen consumption measured in a lab may not generalize perfectly to outdoor competitive running with many uncontrolled factors
- There is often a trade-off between internal and external validity

### **i** Note

Design studies to be both methodologically sound and practically relevant so findings are both true and useful.

## 16 Real-world examples of validity in research

- Use these studies to show how validity is addressed in real research.
- Internal: List common threats (maturation, selection bias, attrition) and how to control them.
- External: Show how validation in real-world settings increases impact.
- Prompt: “What would you do to improve validity in your own project?”

### 16.1 Internal validity example

**Study: “Internal Validity in Resistance Training Research” (Makaruk et al., 2022)**  
<https://pubmed.ncbi.nlm.nih.gov/36281664/>

This review of 340 randomized controlled trials (RCTs) in resistance training identifies threats to internal validity like maturation, history, testing effects, instrumentation, selection bias, and attrition. It provides recommendations for control groups, randomization, standardized protocols, and blinding to strengthen causal inferences in exercise science research.

### 16.2 External validity example

**Study: “Decision-Making Skills in Youth Basketball Players: Diagnostic and External Validation of a Video-Based Assessment” (Rösch et al., 2021)**  
<https://pubmed.ncbi.nlm.nih.gov/33673427/>

This study validated a video-based decision-making assessment for youth basketball players by correlating results with real-game performance data (assists and turnovers). Significant associations showed the tool predicts on-court behavior, ensuring generalizability from lab to competitive sports for talent identification.

## 17 Threats to internal validity

- Script: “Three main threats: confounds, instrument error, and investigator error.”
- Example: Diet study with uncontrolled exercise habits (confound), uncalibrated scale (instrument), inconsistent encouragement (investigator).
- Activity: Have students brainstorm one example for each threat from their field.

Three common categories:

### 1. Intervening variables

1. extraneous variables: these influence the DV but are not related to the IV (e.g., weather)

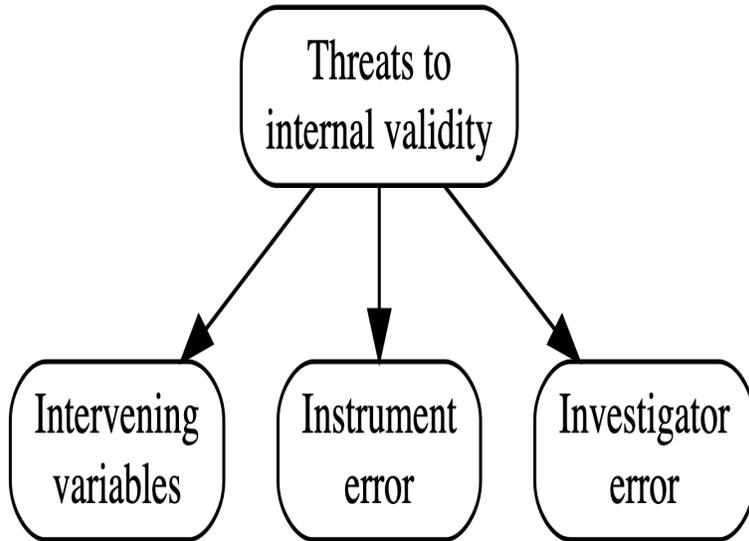
2. confounding variables: these influence the DV and are related to the IV (e.g., prior training status)

**2. Instrument error**

1. example: scale not zeroed, force plate miscalibrated

**3. Investigator error**

1. example: inconsistent instructions, subjective scoring without blinding



**i** Note

Students should understand act like “validity defenders” who anticipate threats before data collection begins.

## 18 Threat 1: Intervening variables

- Emphasize: Confounding variables can mask or mimic effects.
- Example: Sleep, stress, or prior training status in a diet study.
- Prompt: “What is one confound you would control in a hydration study?”
  
- Intervening variables are factors outside the planned design that influence the dependent variable.

Example:

- A diet plan study on body composition that does not control participants' resistance training habits

Mitigation strategies:

- Control groups
- Inclusion and exclusion criteria
- Clear protocols and monitoring of key behaviors

## 19 Knowledge check: threat 1

- You are conducting a study to evaluate the effect of a new hydration strategy on cycling performance.
- Identify one potential confounding variable and describe how you would control for it.

Using the ClassShare App, submit your answers.

Answer

- Potential confounding variable: Ambient temperature during cycling tests
- Control strategy: Conduct all performance tests in a climate-controlled lab environment to ensure consistent temperature

## 20 Threat 2: Instrument error

- Key point: Calibration and standardization are essential.
- Example: Force plate not zeroed, scale not tared.
- Prompt: "How would you document calibration in your methods section?"
- Instrument error occurs when measurement tools are faulty or uncalibrated.

Examples:

- A scale or force plate not properly zeroed
- Faulty equipment settings that create systematic inaccuracies

Mitigation strategies:

- Calibration routines
- Quality checks and logs
- Standard operating procedures for setup and measurement

## 21 Knowledge check: threat 2

- You are using a force plate to measure ground reaction forces during a jump test.
- Describe one step you would take to minimize instrument error.

Using the ClassShare App, submit your answers.

Answer

- Step to minimize instrument error: Perform a calibration check before each testing session by using known weights to ensure the force plate readings are accurate and consistent.

## 22 Threat 3: Investigator error

- Emphasize: Blinding and standardization reduce bias.
- Example: Subjective scoring, inconsistent instructions.
- Prompt: “How can you train raters to be more reliable?”
- Investigator actions can unintentionally introduce bias into data.

Examples:

- Inconsistent verbal encouragement during performance tests
- Subjective scoring of movement quality without blinding

Mitigation strategies:

- Blinding when feasible
- Scripts and standardized instructions
- Training and reliability checks for scoring

## 23 Knowledge check: threat 3

- You are conducting a study where participants perform a series of balance tests, and you are scoring their performance based on movement quality.
- Describe one strategy you would implement to reduce investigator error.

Using the ClassShare App, submit your answers.

Answer

- Strategy to reduce investigator error: Use video recordings of the balance tests and have multiple blinded raters independently score the performances to ensure consistency and reduce subjective bias.

## 24 Module 4: Scaling the Model

- Teaching script: “Once your study is valid, you want to generalize—this is where sampling and inference come in.”
- Example: Using a small, biased sample can undermine even the best experiment.
- Prompt: “What is the difference between a parameter and a statistic?”
  
- Statistical Inference and Sampling
- After a study is built and its integrity is checked, the next challenge is scaling conclusions.
- Statistical inference is the bridge:
  - Making educated generalizations about a population based on a sample

Sampling focus:

- Random sampling
- Stratified sampling
- Parameter vs statistic

## 25 Parameters and statistics

- Definitions: Population vs. sample, parameter vs. statistic.
- Example: Surveying only gym users overestimates activity.
- Prompt: “How would you define the population for your own research question?”

Definitions:

- **Population:** the full group of interest
- **Sample:** the measured subset
- **Parameter:** numerical characteristic of the population
  - Example: average VO2 max of all college athletes
- **Statistic:** numerical characteristic of the sample
  - Example: average VO2 max of surveyed college athletes

Sampling bias example:

- Surveying only university gym students likely overestimates physical activity compared to the broader student body

## 26 Random and stratified sampling

- Random sampling: Equal chance for all, reduces bias.
- Stratified: Ensures subgroups are represented.
- Example: Sampling by sport type or gender.
- Prompt: “Why might you use stratified sampling in a study of college athletes?”

### 26.1 Random sampling

- Every population member has an equal and independent chance of selection
- Helps avoid systematic bias
- With sufficient size, tends to reflect population characteristics in natural proportions

### 26.2 Stratified sampling

- Used when population has distinct subgroups
- Steps:
  1. Divide population into strata (example: endurance, power, team sports)
  2. Randomly sample within each stratum proportional to its size
- Helps ensure no subgroup is over or under represented
- Can improve representativeness and precision

## 27 Module 5: The Stress Test

- Teaching script: “Now we test our structure—can our theory generate a testable hypothesis, and can we use statistics to evaluate it?”
- Example: Theory → hypothesis → test.
- Prompt: “What is the difference between a theory and a hypothesis?”

Hypotheses and Theoretical Frameworks

Now the finished structure undergoes a final test.

This module covers:

- How broad theories generate testable hypotheses
- How hypotheses are tested against statistical reality
- The logic of null hypothesis significance testing using probability

## 28 Theories and hypotheses

- Theory: Broad, explanatory; hypothesis: specific, testable.
- Example: Visualization theory → hypothesis about free-throw accuracy.
- Prompt: “Write a hypothesis for a theory you know.”
  
- **Theory**: broad conceptual framework explaining a phenomenon, informed by prior research and observation  
Example: mental visualization in motor learning suggests imagined practice can improve performance
- **Hypothesis**: specific, testable prediction derived from theory  
Example hypothesis shown:  
“Basketball players who engage in 20 minutes of daily mental visualization of free throws will show greater improvement in free-throw accuracy than players who do not.”

Key idea:

- Theories are usually too broad to test directly
- Hypotheses are the testable units that accumulate evidence for or against a theory

### **i** Note

Page 15 synthesis note: translating a broad idea into a precise testable question is a core research skill and a core source of scientific creativity.

## 29 Hypothesis testing (H0 and H1)

- H0: No effect; H1: Effect exists.
- Example: Null—no difference in free-throw improvement.
- Emphasize: We test H0, not H1 directly.
- Prompt: “Why is ‘fail to reject H0’ not the same as ‘prove H0’?”
  
- **Research hypothesis (H1)**: an effect exists
- **Null hypothesis (H0)**: no difference or no relationship

Example null statement:

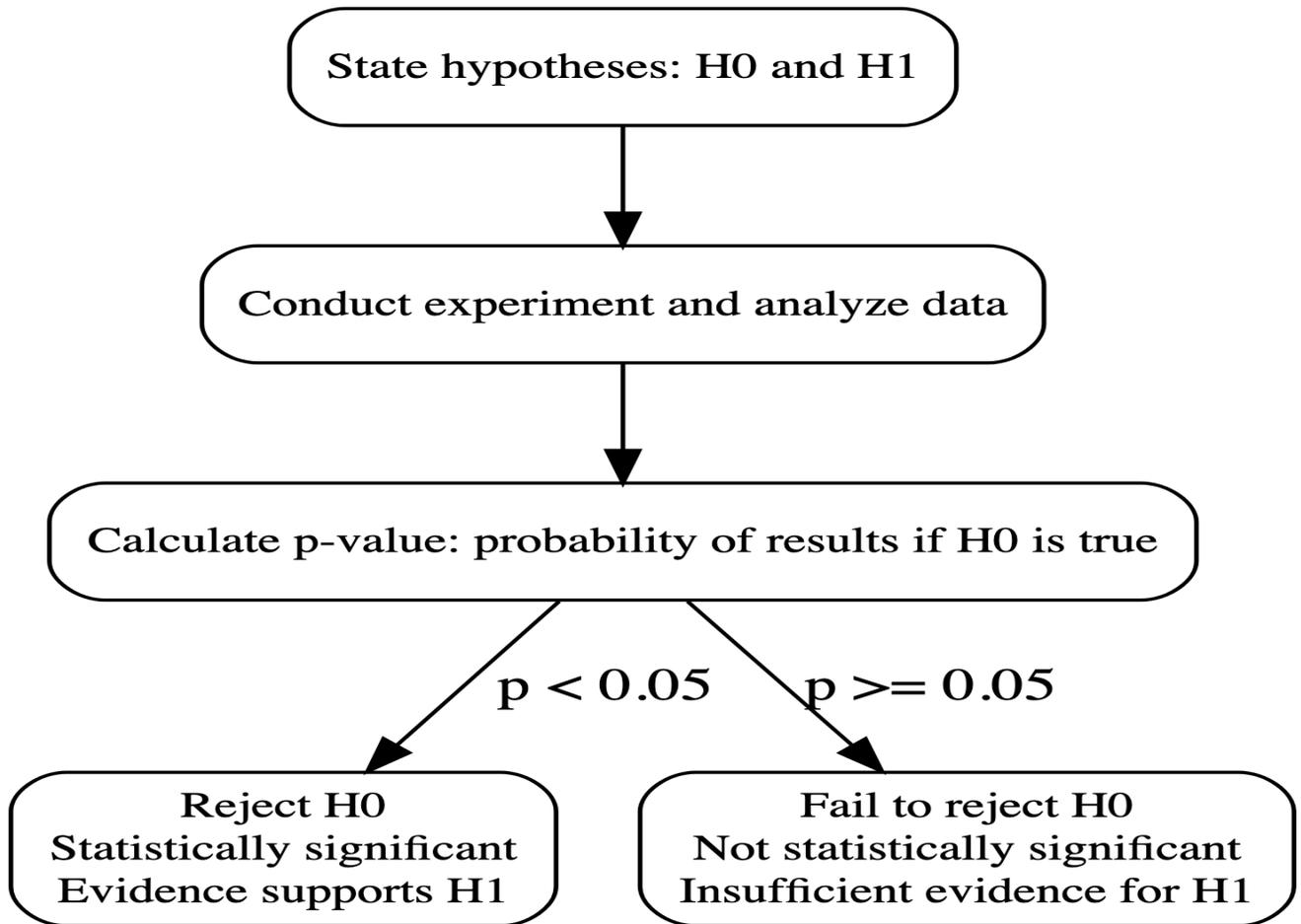
- “There is no difference in free-throw accuracy improvement between the visualization group and the control group.”

Key logic:

- We do not prove H1 directly
- We evaluate how consistent the data are with H0

### 30 Hypothesis testing workflow and the p-value

- p-value: Probability of observed results if H0 is true.
- Example:  $p < 0.05$  means results are unlikely by chance.
- Emphasize: Statistical significance practical importance.
- Prompt: “What does a p-value actually tell us?”



Interpretation aligned with the chapter:

- The p-value is the probability of observing the results (or more extreme results) if H0 were true
- If  $p < 0.05$ , results are unlikely due to random chance or sampling error, so we reject H0 and conclude a statistically significant effect