OVERVIEW

How do you know if your results are meaningful—or just chance This module introduces statistical significance, the tool researchers use to test whether patterns in data are real. Learn what a p-value tells you, how to interpret it, and why real-world impact always matters just as much.

Representation

What is Representation?
- **Definition:** Ensures data reflects the diversity of the population being studied.

Risks:

Type	Definition	Example
Over-representation	A group is sampled more than its actual share	Surveys at popular parks miss underserved areas
Under-representation	A group is sampled less than its actual share	Non-English speakers excluded from community feedback

Bias Reinforcement:
- **Definition:** Biased data perpetuates inequalities and leads to unfair outcomes.
- **Examples:**
  - Hiring tools favoring specific demographics.
  - Healthcare algorithms missing symptoms in underrepresented groups.

Improving Representation

Method	Description	Example
Random Sampling	Equal chance for all population members	Randomly selecting 100 people from a list
Stratified Sampling	Dividing population into	Political polling by race,

	groups for proportional sampling	gender, income

Real-World Impacts of Bias

Domain	Effect
Hiring	Tools favor applicants with specific backgrounds
Healthcare	Misses critical health trends in marginalized groups
Education	Funding decisions skewed by data from wealthier schools

Why Ethical Data Practices Matter

- **Avoid harm:** Prevent discriminatory policies or algorithms.
- **Build trust:** Ensure decisions serve the entire community.
- **Address systemic issues:** Highlight inequities for targeted solutions.

Key Takeaways

- Representative data → Fairer insights → Better decisions.
- Bias in data = Bias in outcomes.
- Proactive checks (e.g., stratified sampling, bias audits) mitigate risks.
- Ethical practices are essential for public trust and equity.