Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

A.Y.: 2024-25 **Class/ Sem: B.Tech/ Sem-VIII** **Sub:** Data Ethics

Bhuvi Ghosh
60009210191

Tutorial-7

Please describe following sources of bias in data analysis along with examples and their mitigation strategies:

1. Selection Bias
2. Measurement Bias
3. Confirmation Bias
4. Sampling Bias
5. Overfitting
6. Survivorship Bias
7. Observer Bias
8. Cultural or Contextual Bias
9. Algorithmic Bias
10. Reporting Bias

1. Selection Bias

Selection bias occurs when the data used in an analysis is not randomly selected from the target population, resulting in a systematic error where certain outcomes or subgroups are more likely to be included than others. This non-random selection distorts the relationship between the variables being studied and can lead to conclusions that don't generalize to the broader population.

Real-World Examples:

1. **Clinical Trial Selection:** A pharmaceutical company tests a new arthritis medication but excludes elderly patients with multiple health conditions. The drug

appears highly effective in the study but performs poorly when released to the general population where many arthritis sufferers have comorbidities.

2. **Customer Satisfaction Survey:** A company emails a satisfaction survey only to customers who have made multiple purchases. By excluding one-time customers (who may be dissatisfied), the company overestimates customer satisfaction.

3. **Berkson's Paradox:** A hospital study examines the relationship between two diseases but only includes hospitalized patients. This creates a false correlation because people with either disease alone might not be hospitalized, while those with both are more likely to be admitted.

Advanced Mitigation Strategies:

- Implement probability-based sampling methods like stratified random sampling.
- Use propensity score matching to balance characteristics across comparison groups.
- Document and analyze the characteristics of excluded participants.
- Conduct sensitivity analyses to assess how selection criteria affect results.
- Consider inverse probability weighting to adjust for non-random selection.
- Create clearly defined inclusion/exclusion criteria before data collection begins.
- When working with observational data, use instrumental variables or regression discontinuity designs.

2. Measurement Bias

Measurement bias refers to systematic errors in the collection, recording, or interpretation of data that lead to consistent deviation from the true values. This occurs when the measurement process itself introduces errors that systematically affect the data in a particular direction.

Types of Measurement Bias:

- **Instrument bias:** When measuring devices provide consistently skewed readings.
- **Observer/interviewer bias:** When the data collector influences responses.
- **Recall bias:** When participants remember past events inaccurately.
- **Social desirability bias:** When respondents answer questions to appear more favorable.
- **Leading question bias:** When question phrasing influences responses.

Real-World Examples:

1. **Health Risk Assessment:** A blood pressure device used in a hypertension study consistently reads 5 mmHg higher than actual values, causing overdiagnosis and potentially inappropriate treatment recommendations.

2. **Employment Survey:** An interviewer's tone changes when asking questions about job satisfaction to employees from different departments, unconsciously encouraging more positive responses from departments the interviewer favors.

3. **COVID-19 Case Reporting:** During the pandemic, countries using different testing methods, case definitions, and reporting standards created measurement biases that made international comparisons difficult and potentially misleading.

Mitigation Strategies:

- Implement double-blind procedures where neither participants nor researchers know group assignments.
- Use multiple measurement methods and triangulate results.
- Calibrate instruments against known standards regularly.

- Train data collectors extensively with inter-rater reliability testing.
- Use validated, standardized measurement tools with known error rates.
- Automate data collection where feasible to reduce human error.
- Implement quality control procedures like random verification of measurements.
- Use technology (audio/video recording) to verify data collection processes.
- Create detailed measurement protocols with explicit operational definitions.

3. Confirmation Bias

Confirmation bias is a cognitive bias where individuals favor information that confirms their existing beliefs and give disproportionately less consideration to alternative possibilities. In data analysis, this manifests as researchers selectively gathering or interpreting evidence to support their hypotheses while overlooking contradictory data or alternative explanations.

Manifestations in Research:

- Designing studies that can only confirm, not refute, hypotheses.
- Selectively analyzing favorable data subsets (p-hacking).
- Interpreting ambiguous results as supportive of prior beliefs.
- Applying more scrutiny to contradictory findings than confirmatory ones.
- Post-hoc rationalization of unexpected results to fit existing theories.

Real-World Examples:

1. **Pharmaceutical Research:** A researcher deeply invested in a drug's effectiveness focuses analysis on the subset of patients who showed improvement while dismissing side effects as unrelated to the treatment.

2. **Economic Policy Analysis:** An economist with strong political beliefs selects time periods and economic indicators that support their preferred policies, while

ignoring contradictory data from other periods or metrics.

3. **Machine Learning Development:** A data scientist building a fraud detection algorithm focuses on improving the detection of known fraud patterns but fails to adequately test for false positives or novel fraud techniques.

Mitigation Strategies:

- Pre-register study protocols, hypotheses, and analysis plans before collecting data.
- Actively seek disconfirming evidence and alternative explanations.
- Engage team members with diverse perspectives to challenge assumptions.
- Implement adversarial collaboration with researchers holding opposing views
- Use structured analytical techniques like Analysis of Competing Hypotheses (ACH).
- Separate exploratory from confirmatory analyses, clearly labeling each.
- Conduct blind analysis where researchers analyze anonymized data without knowing which group is which.
- Establish decision rules and thresholds before seeing results.
- Create a "red team" tasked specifically with finding weaknesses in your analysis.

4. Sampling Bias

Sampling bias occurs when the process of selecting a sample from a population causes certain members to be less likely to be included than others, resulting in a non-representative sample. Unlike selection bias (which often occurs after initial selection), sampling bias happens during the initial sampling procedure itself.

Types of Sampling Bias:

- **Convenience sampling bias:** Selecting easily accessible participants.
- **Voluntary response bias:** Using only participants who volunteer.
- **Undercoverage bias:** Inadequately representing certain population segments.
- **Time bias:** Sampling during restricted time periods that influence who's available.
- **Geographic bias:** Limiting samples to specific locations.

Real-World Examples:

1. **Election Polling Failure:** The 1936 Literary Digest poll predicted Alf Landon would defeat Franklin Roosevelt by sampling people from telephone directories and vehicle registrations—methods that overrepresented wealthy Americans during the Great Depression. Roosevelt won in a landslide.

2. **Medical Research:** A study on sleep disorders recruits participants from a university student population, resulting in data that fails to represent older adults, people with different work schedules, or those from diverse socioeconomic backgrounds.

3. **Marketing Survey:** A company conducts in-person surveys about shopping preferences only during business hours on weekdays, systematically excluding working professionals and capturing a disproportionate number of retirees, stay-at-home parents, and shift workers.

Mitigation Strategies:

- Implement probability sampling methods (simple random, stratified, cluster).
- Use respondent-driven sampling for hard-to-reach populations.
- Apply post-stratification weights to correct for known sampling imbalances.
- Conduct power analyses to determine appropriate sample sizes.
- Use sampling frames that comprehensively cover the target population.
- Employ mixed-mode data collection (online, phone, in-person).

- Document detailed characteristics of your sample and compare to population parameters.
- Conduct non-response follow-up studies to understand who isn't participating.
- Consider temporal sampling to capture cyclical variations.

5. Overfitting

Overfitting occurs when a statistical model or machine learning algorithm captures noise or random fluctuations in the training data rather than the underlying pattern. An overfit model performs exceptionally well on training data but fails to generalize to new, unseen data. This represents a fundamental trade-off between model complexity and generalizability.

Technical Indicators of Overfitting:

- Large gap between training and validation performance metrics.
- Model complexity far exceeding the available data points.
- Perfect or near-perfect performance on training data.
- Erratic or implausible parameter estimates.
- High sensitivity to small changes in training data.

Real-World Examples:

1. **Financial Forecast Model:** An investment firm builds a stock prediction model with hundreds of parameters using only a few years of historical data. The model identifies spurious correlations and seasonal patterns that won't recur, leading to poor investment decisions.

2. **Healthcare Prediction:** A researcher develops a disease risk prediction algorithm using 50 variables but only 100 patient records. The resulting model achieves 99% accuracy on training data but performs no better than chance on new

patients.

3. **Marketing Segmentation:** A company creates an extremely detailed customer segmentation model that perfectly categorizes existing customers but fails to correctly classify new customers because it's built around idiosyncrasies in the original dataset rather than stable purchasing patterns.

Mitigation Strategies:

- Implement regularization techniques (L1, L2, elastic net).
- Use k-fold cross-validation to assess model stability.
- Employ early stopping in iterative training processes.
- Apply pruning techniques for decision trees and neural networks.
- Reduce model complexity through feature selection or dimensionality reduction.
- Use ensemble methods that average multiple models.
- Implement Bayesian approaches with informative priors.
- Ensure sufficient data relative to model parameters (10:1 rule of thumb).
- Monitor learning curves to identify the point of diminishing returns.
- Use information criteria (AIC, BIC) to balance fit and complexity.

6. Survivorship Bias

Survivorship bias is a logical error that occurs when we focus only on entities that "survived" or passed some selection process while overlooking those that did not. This creates a false perception about what factors lead to success or failure because we're analyzing an incomplete, non-representative sample that excludes the "non-survivors."

Common Contexts for Survivorship Bias:

- Business and investment analysis.

- Product performance evaluation.
- Career success studies.
- Historical analyses.
- Performance benchmarking.
- Innovation research.

Real-World Examples:

1. **Mutual Fund Performance:** Investment platforms advertise funds with strong 10-year performance records. This naturally excludes funds that performed poorly and were closed or merged, artificially inflating the apparent average return of long-term investments.

2. **World War II Aircraft Armor:** Statistician Abraham Wald noticed that military analysts were examining returning aircraft to determine where to add armor based on damage patterns. He pointed out they should reinforce areas where returning planes showed no damage—those were the critical areas where planes that got hit didn't survive to return.

3. **Entrepreneurship Research:** Books highlighting common traits of successful entrepreneurs often study only businesses that succeeded, missing that failed ventures may have exhibited many of the same characteristics. This creates misleading "success formulas" that overemphasize visible factors while missing critical contextual elements.

Mitigation Strategies:

- Create and maintain comprehensive databases that track both successes and failures.
- Use time-series approaches that capture entities at various stages before outcomes are known.

- Implement "dead case" analysis to specifically study failures.
- Examine entire cohorts from inception rather than sampling based on current status.
- Apply statistical techniques to estimate and correct for missing data.
- Document and account for entities that leave the sample over time.
- Construct counterfactual scenarios based on known failure mechanisms.
- Use matched-pair designs comparing successful and unsuccessful cases with similar starting conditions.
- Be transparent about the scope of survivorship bias in your analyses and its potential impact.

7. Observer Bias

Observer bias (also called experimenter or researcher bias) occurs when the beliefs, expectations, or perceptions of the researcher unconsciously influence how they collect, interpret, or report data. This bias can manifest at any stage of research, from study design to publication, and often operates subconsciously.

Manifestations of Observer Bias:

- Subtle cueing of participants toward expected responses.
- Inconsistent application of subjective measurements.
- Selective recording of observations.
- Differential attention to confirming versus disconfirming evidence.
- Unconscious influence on participant behavior (Pygmalion/Rosenthal effect).
- Subjective interpretation of ambiguous results.

Real-World Examples:

1. **Psychological Assessment:** A clinician conducting diagnostic interviews has a preliminary hypothesis that a patient has depression. During the interview, the

clinician unconsciously spends more time on questions about depressive symptoms while minimizing attention to symptoms that might suggest alternative diagnoses.

2. **Animal Behavior Research:** A zoologist studying aggressive behaviors in a species counts and records incidents differently based on unconscious expectations about which gender exhibits more aggression, creating artificial gender differences in the data.

3. **Educational Research:** Teachers participating in a study of a new teaching method are told which students are expected to show the most improvement. Despite identical starting abilities, these students do show greater gains because teachers unconsciously give them more attention and positive reinforcement.

Mitigation Strategies:

- Implement triple-blind procedures where possible (participant, experimenter, and analyst all unaware of group assignments).
- Use automated data collection systems when feasible.
- Develop detailed, objective coding schemes before data collection begins.
- Employ multiple independent observers and calculate inter-rater reliability.
- Rotate data collectors across conditions randomly.
- Separate roles so that those collecting data are different from those analyzing it.
- Use standardized scripts for interactions with participants.
- Video record sessions for quality control or secondary coding.
- Train observers specifically on bias awareness and mitigation.
- Implement computational approaches to identify patterns in data collection that might indicate observer effects.

8. Cultural or Contextual Bias

Cultural or contextual bias emerges when research methods, measurements, or interpretations are based on assumptions from one cultural context but applied to others without appropriate adaptation. This bias leads to misinterpretation of data and potentially invalid conclusions because cultural factors fundamentally shape how people perceive, process, and respond to information.

Key Dimensions Affected:

- Language and communication norms.
- Values and belief systems.
- Social norms and expectations.
- Cognitive styles and decision-making approaches.
- Emotional expression and regulation.
- Time orientation and concepts.
- Individual vs. collectivistic frameworks.

Real-World Examples:

1. **Mental Health Assessment:** A depression screening tool developed in Western contexts asks about feelings of sadness and worthlessness, but in some Asian cultures, depression more commonly manifests as physical complaints. The tool systematically underdiagnosed depression in these populations due to culturally biased symptom expectations.

2. **Educational Testing:** A standardized test includes reading comprehension passages about baseball, skiing, and suburban lifestyles. Students from cultures or contexts unfamiliar with these concepts score lower not because of reading ability but because of contextual knowledge gaps.

3. **User Experience Research:** A tech company conducts usability testing for a global app using protocols developed in Silicon Valley. The research misses

critical usability issues because it assumes universal comfort with digital privacy trade-offs, linear navigation patterns, and individualistic user goals that don't apply across all cultures.

Mitigation Strategies:

- Include researchers from diverse cultural backgrounds on research teams.
- Conduct preliminary ethnographic research before designing measurement tools.
- Use mixed-methods approaches that capture both quantitative metrics and qualitative context.
- Develop and validate measures separately for different cultural contexts.
- Employ cognitive interviewing to understand how questions are interpreted across cultures.
- Use local collaborators for data collection and interpretation.
- Implement cross-cultural validation studies for assessment instruments.
- Train researchers in cultural humility and awareness of their own cultural lenses.
- Use emic (culture-specific) and etic (universal) approaches in complementary ways.
- Consider cultural adaptation rather than mere translation of research materials.

9. Algorithmic Bias

Algorithmic bias refers to systematic errors in computational systems that create unfair or discriminatory outcomes for certain groups or individuals. These biases emerge from various sources: biased training data, flawed algorithm design, inappropriate problem formulation, or misaligned optimization objectives. As algorithms increasingly influence decisions in healthcare, finance, employment, and criminal justice, these biases can amplify existing social inequities.

Sources of Algorithmic Bias:

- Training data reflecting historical discrimination.
- Proxy variables that correlate with protected characteristics.
- Representation disparities in training datasets.
- Inappropriate feature selection.
- Misaligned optimization metrics.
- Inappropriate problem formulation.
- Feedback loops that amplify initial biases.

Real-World Examples:

1. **Criminal Justice Risk Assessment:** COMPAS, an algorithm used to predict recidivism risk, was found to falsely flag Black defendants as future criminals at nearly twice the rate as white defendants, while white defendants were more likely to be incorrectly labeled as low risk.

2. **Healthcare Resource Allocation:** An algorithm used by hospitals to identify patients needing extra medical care systematically underestimated the health needs of Black patients because it used healthcare costs as a proxy for health needs, but historical inequities meant Black patients had generated lower costs despite equal needs.

3. **Facial Recognition:** Multiple commercial facial recognition systems have shown significantly higher error rates for women with darker skin tones compared to lighter-skinned males, leading to potential discrimination in systems using this technology for identification.

Mitigation Strategies:

- Implement fairness constraints in algorithm design (demographic parity, equal opportunity, etc).
- Use techniques like adversarial debiasing to reduce discriminatory outputs.
- Conduct regular algorithmic audits across different demographic groups.
- Employ counterfactual fairness testing to evaluate decisions under different scenarios.
- Develop diverse and representative training datasets.
- Consider causal approaches to prevent algorithmic reliance on spurious correlations.
- Implement transparency mechanisms to explain algorithmic decisions.
- Establish multi-disciplinary oversight committees including ethicists and domain experts.
- Use techniques like data augmentation to balance representation in training data.
- Develop context-specific fairness metrics relevant to the problem domain.
- Consider human-in-the-loop systems for high-stakes decisions.

10. Reporting Bias:

Reporting bias occurs when the dissemination of research findings is influenced by the nature and direction of results rather than their scientific merit. This systematic distortion creates a public record that over-represents certain outcomes (typically positive or statistically significant ones) while underrepresenting others. Reporting bias encompasses publication bias (what gets published) and outcome reporting bias (which results are highlighted).

Types of Reporting Bias:

- **Publication bias:** Journals preferentially publishing positive results.
- **Outcome reporting bias:** Selectively reporting favorable outcomes from multiple measured.

- **Time-lag bias:** Delayed publication of negative findings.
- **Citation bias:** Preferential citing of studies with desirable results.
- **Language bias:** Over-reliance on English-language studies.
- **Multiple publication bias:** Publishing positive results multiple times.
- **Funding bias:** Influence of financial interests on reporting.

Real-World Examples:

1. **Antidepressant Effectiveness:** A 2008 study in the New England Journal of Medicine found that 94% of published antidepressant trials showed positive results, but when unpublished trials were included, only 51% were positive—suggesting a significant publication bias that overstated drug efficacy.

2. **Surgical Technique Evaluation:** Researchers investigating a new minimally invasive surgery measured 20 different outcomes. Their published paper reports only the three outcomes showing statistically significant improvements while omitting the 17 outcomes showing no benefit or potential harm.

3. **Nutritional Research:** A food industry-funded study examining the health effects of a specific sweetener reports cardiovascular benefits but doesn't mention observed negative effects on insulin sensitivity, creating a distorted picture of the overall health impact.

Mitigation Strategies:

- Mandate pre-registration of studies and analysis plans in public registries.
- Implement registered reports where journals accept papers based on methodology before results are known.
- Develop comprehensive reporting guidelines specific to different research types (CONSORT, PRISMA, etc).
- Create result-blind peer review processes.

- Support journals dedicated to publishing null or negative results.
- Use statistical methods to detect and correct for publication bias in meta-analyses (funnel plots, trim-and-fill).
- Implement open data policies requiring sharing of complete datasets.
- Conduct systematic follow-up of unpublished studies identified in registries.
- Require conflict of interest disclosures that include non-financial influences.
- Create incentives in academic systems for transparency rather than just positive findings.
- Develop machine learning tools to identify selective reporting by comparing registrations to publications.