Name: _____ **Date:** _____

Answer Key: Data Literacy and Algorithmic Bias for 9th Grade

Critical analysis of 10 complex scenarios covering statistical anomalies and ethical data management beyond simple chart reading.

---

1. A urban planning committee uses a 'heat map' of smartphone GPS pings to decide where to install new bike lanes. Which data literacy oversight is most likely occurring here?

Answer: A) Selection bias regarding socio-economic demographics

Smartphone ownership and data usage patterns vary by income and age; relying solely on GPS pings may lead to under-representing low-income residents or the elderly who may use bikes but not tracking apps.

2. In the context of data ethics, 'de-identified' data can often be 're-identified' by cross-referencing it with other publicly available datasets.

Answer: A) True

This is known as a linkage attack. Even if names are removed, specific combinations of ZIP codes, birthdates, and gender can uniquely identify individuals when compared to other data sources.

3. When a researcher finds that two variables (like ice cream sales and sunburns) move together, but one does not cause the other, this is known as a _____.

Answer: B) Spurious correlation

A spurious correlation occurs when two variables appear related but are actually both influenced by a third 'lurking' variable (in this case, hot weather).

4. A healthcare AI was trained on historical data where doctors primarily treated male patients for heart disease. What is the most likely algorithmic outcome when it assesses female patients?

Answer: C) The AI may provide a false-negative or under-diagnose females

This is a classic example of training data bias. If the algorithm lacks representative data for a group, its predictive accuracy for that group will be significantly lower.

5. Using a 'Creative Commons Zero (CC0)' license means that the data creator has waived all copyright and placed the work in the public domain.

Answer: A) True

Name: _____ **Date:** _____

CC0 is a public domain Dedication, allowing anyone to use, modify, and distribute the data for any purpose without restriction.

6. The ethical practice of ensuring that individuals are aware of how their data is collected and used is called _____.

Answer: B) Informed consent

Informed consent is a fundamental pillar of data ethics, requiring transparency between the data collector and the subject.

7. Which of the following is the most significant indicator that a dataset might be unreliable for a long-term sociological study?

Answer: B) The data lacks a dictionary or metadata explaining variable definitions

Without a data dictionary or metadata, researchers cannot understand what the labels or values actually represent, leading to misinterpretation and invalid analysis.

8. A 'P-value' of 0.05 is the universal proof that a data trend is 100% true and cannot be attributed to chance.

Answer: B) False

A P-value of 0.05 suggests there is a 5% chance the results occurred by luck; it is a threshold for statistical significance, not a guarantee of absolute truth.

9. To protect a database from being easily read if stolen, administrators use _____, which scrambles data into unreadable code.

Answer: B) Encryption

Encryption is a critical management skill that ensures data integrity and privacy by making data inaccessible to unauthorized parties.

10. You are evaluating two studies on car safety. Study A is funded by a car manufacturer. Study B is funded by a university research grant. Why is this distinction important for data literacy?

Answer: C) To identify potential conflict of interest and funding bias

Evaluating the source's motivation is key. A car manufacturer has a financial interest in specific outcomes, which might influence how they interpret or display their data.