Name: _____ **Date:** _____

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?

A. Selection bias regarding socio-economic demographics
B. A failure to analyze time-series correlations
C. The use of qualitative instead of quantitative data
D. A violation of the Open Data Protocol (ODP)

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

A. True
B. False

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 _____.

A. Causal feedback loop
B. Spurious correlation
C. Linear regression error
D. Standard deviation

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?

A. The AI will automatically adjust for biological differences
B. The AI will require more data to reach a conclusion
C. The AI may provide a false-negative or under-diagnose females
D. The AI will delete the incomplete male data entries

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.

A. True
B. False

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

A. Data scrubbing
B. Informed consent
C. Metadata tagging

Name: _____ Date: _____

D. Data siloing

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

A. The data is stored in a .CSV format instead of a SQL database
B. The data lacks a dictionary or metadata explaining variable definitions
C. The dataset contains over 1,000,000 individual entries
D. The data was collected by a non-profit organization

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

A. True
B. False

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

A. Compression
B. Encryption
C. Parsing
D. Indexing

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?

A. Study A is automatically false because corporations cannot collect data
B. Study B is more likely to have used a larger computer server
C. To identify potential conflict of interest and funding bias
D. Universities are legally required to use better visual charts