**CASE 2: TENNIS ANALYTICS**

TABLE A.1 Winning Percentages Associated with Performance Statistics

|  |  |
| --- | --- |
| MEASURE | Winning % |
| First serve percentage > 75% | 93 |
| First serve percentage $ 75% | 76 |
| Number of times at net > 30 | 91 |
| Number of times at net < 30 | 78 |

**1. The commentators are implicitly making two active predictions.**

a. Active Predictions Implied by Commentators:

The first active prediction seems to be related to the impact of the first serve percentage on winning. Commentators suggest that a first serve percentage greater than 75% is associated with a higher winning percentage.

The second prediction relates to the number of times a player is at the net. Commentators imply that being at the net more than 30 times is associated with a higher winning percentage.

b. Two Treatments and Associated Outcomes:

Treatment 1: First Serve Percentage

Treatment Level 1: First serve percentage > 75%

Outcome for Treatment Level 1: Higher winning percentage (93%)

Treatment 2: Number of Times at Net

Treatment Level 2: Number of times at net > 30

Outcome for Treatment Level 2: Higher winning percentage (91%)

c. Data-Generating Process Expression:

Let Y be the outcome (winning percentage), X1 be the first serve percentage, X2 be the number of times at the net.

For Treatment 1:

Y1=B0+B1X1 + e1

Where Y1 is the winning percentage for treatment 1, B0 is the intercept, B1 is the coefficient for first serve percentage and e1 is the unobservable error term.

For Treatment 2:

Y2=B0+B2X2+e2

Where Y2 is the winning percentage for treatment 2, B2 is the coefficient for second serve percentage and e2 is the unobservable error term.

**2. Regarding assumptions:**

a. Relevance of Historical Data:

The historical data used to derive the winning percentages associated with first serve percentage and times at the net should be relevant to the current match. If playing conditions, opponents, or other relevant factors have changed, the claims may not hold.

b. Causation vs. Correlation

The assumptions imply a causal relationship between the strategies (first serve percentage, times at the net) and winning percentage. It assumes that a higher first serve percentage or more time at the net directly contributes to a higher chance of winning.

c. Independence of Matches:

The assumptions may assume that each match is independent of others, meaning the outcome of one match does not directly influence the outcome of another. This assumption is necessary for generalizing strategies across matches.

b. Possible Violations of Assumptions:

Changing Conditions:

If match conditions have changed (e.g., different playing surface, weather conditions), the effectiveness of strategies may vary, violating the assumption of relevance.

Opponent-Specific Strategies:

If the statistics are not general but specific to an opponent, the effectiveness of strategies may not hold against other opponents. Violation of this could lead to ineffective strategies against different playing styles.

Temporal Trends:

If there are temporal trends in player performance or strategy effectiveness, relying solely on historical data may not capture these trends.

c. Impact of Opponent-Specific Data

Yes, the answer to 2a would change if the statistics in Table A.1 were specific to Serena's upcoming opponent. In this case, the assumptions would need to be modified to account for the opponent-specific nature of the data. The effectiveness of strategies may depend on the playing style and strengths/weaknesses of the specific opponent and generalizing from past matches against that opponent may be more relevant. However, this would still require assumptions about the stability of the opponent's playing style over time.

**3. If the commentators used the figures in @ Table A.1 as a guide for people betting on the match, rather than to suggest strategy to serena, would you view them differently?**

The commentators are using the figures in Table A.1 as a guide for people betting on the match rather than suggesting strategies to Serena, the context and purpose of their comments would be different. Here's how the perspective might change:

1. Information for Betting:

If the commentators are providing insights based on historical data to inform betting decisions, their role is more aligned with helping the audience understand the statistical likelihood of certain outcomes based on past performance.

1. Objective vs. Strategy:

The objective is likely to provide information for individuals who are considering placing bets on the match. In this case, the focus is on statistical trends and probabilities rather than guiding Serena's strategic decisions during the match.

1. Risk and Return:

The audience, in this case, would use the information for assessing the risk and potential return on their bets. The commentators might discuss which statistics have historically been associated with higher winning percentages, helping bettors make more informed decisions.

**4. Can you describe a set of circumstances where the information in table A.1 would lead to the predictions of the commentators? If so, detail the corresponding reasoning.**

the information in Table A.1 aligns with the predictions made by the commentators:

1. First Serve Percentage:

If a player has a first serve percentage greater than 75%, it might indicate a higher level of accuracy and effectiveness in delivering the first serve. A more accurate first serve can put pressure on the opponent, leading to a higher likelihood of winning points. The reasoning here is that a reliable first serve often results in more favorable situations for the serving player, potentially leading to a higher winning percentage.

1. Number of Times at Net:

If a player is at the net more than 30 times, it could suggest an aggressive and proactive playing style. Being at the net frequently provides opportunities for volleys and put-away shots, increasing the chances of winning points. The reasoning is that players who actively seek to finish points at the net may have a more offensive and potentially successful approach, contributing to a higher winning percentage.