




**Total Quiz Score:**

- 5 points.

Notes:

- For “**Multiple-Choice**” questions with just one correct answer:
  - The “” symbol next to a bolded option represents the *correct answer*.
- For “**Select All That Apply**” questions:
  - The “” symbol next to a bolded option indicates a *correct selection*.
  - The “” symbol next to an option indicates an *incorrect selection*.

**Question 1.A (1 point)**

A study exploring the effects of political corruption on tax fraud found that in countries with less political corruption, private firms were less likely to engage in tax fraud.

What type of relationship is described in the study?

Question 1.A options:

- ☐ No relationship (0 points)
- ☐ Negative relationship (0 points)
- ☒ **Positive relationship** (+1 point)
- ☐ Non-linear relationship (0 points)

**Justification:**

By definition, there is a positive relationship between two variables if they tend to move together in the same direction. A positive association can manifest in one of two ways: (a) if one variable decreases, the other variable also tends to decrease; (b) if one variable increases, the other variable also tends to increase. In this case, because countries with less political corruption are observed to have lower rates of tax fraud, both variables decrease together. Conversely, this implies that as political corruption increases, tax fraud would also tend to increase. This consistent co-movement confirms a positive relationship between political corruption and tax fraud.

**Question 1.B (1 point)**

A study examines how changes in the minimum wage affect employment levels in the fast-food industry. Researchers compare employment data from states that increased their minimum wage with those that did not, analyzing changes over time.

Based on this information about the study, select all that apply:

(Note: Incorrect selections will discount partial credit from this question)

Question 1.B options:

- ☒ **The response variable is employment levels in the fast-food industry** (+0.25 if selected)
- ☒ The study is experimental (+0.25 if NOT selected; -0.25 if selected)
- ☒ **The study is observational** (+0.25 if selected)
- ☒ **The explanatory variable is the minimum wage level** (+0.25 if selected)

**Justification:**

The study is not experimental, as the researchers do not directly manipulate any of the variables. Thus, the study fits the definition of an observational study, in which researchers take variables' measures without any direct manipulation. According to the respective definitions, the explanatory variable is the minimum wage level because it is the variable that is used to predict or explain changes in employment. Conversely, the response variable is employment levels in the fast-food industry because it represents the outcome being measured. This is the outcome as we seek to understand how it changes in response to variations in the explanatory variable.

### Question 1.C (1 point)

Consider the voting intentions for primary candidates as expressed through a survey where respondents must choose the candidate they most prefer among a closed set of options (e.g., Candidate A, Candidate B, Candidate C). Respondents' answers are recorded in a column of a dataset.

Which kind of variable type is this?

Question 1.C options:

- ☒ **Categorical Nominal** (1 points)
- ☐ Numerical Discrete (0 points)
- ☐ Numerical Continuous (0 points)
- ☐ Categorical Ordinal (0 points)

#### Justification:

The information measured by this variable reflects people's electoral preferences among a set of candidate options. Since each candidate represents a complex set of information without a unique numeric description or ordering, we classify this variable as non-numeric. The values of this variable correspond to different candidates, each representing a distinct category. Since these categories lack an inherent order for comparison in intensity or value, this is a categorical nominal variable.

### Question 1.D (1 point)

What is the key distinction between simple random sampling and stratified sampling?

Question 1.D options:

- ☐ Stratified sampling is only used for small populations, while simple random sampling is preferred for large populations. (0 points)
- ☒ **Simple random sampling selects individuals directly from the population with equal probability, while stratified sampling first divides the population into distinct subgroups based on a given characteristic and then selects random samples from each.** (1 points)
- ☐ Simple random sampling guarantees proportional representation of all demographic groups, while stratified sampling risks underrepresenting minority subgroups. (0 points)
- ☐ Stratified sampling does not use randomization, whereas simple random sampling relies entirely on random selection. (0 points)

#### Justification:

- "Stratified sampling is only used for small populations, while simple random sampling is preferred for large populations" → This is **incorrect** since simple random sampling can also be used with small populations.
- "Simple random sampling guarantees proportional representation of all demographic groups, while stratified sampling risks underrepresenting minority subgroups." → **Incorrect.**
  - Doing simple random sampling with a small sample size might result in a distribution of features in the sample that deviates from the population characteristics because chance alone may lead to an overrepresentation of some groups and an underrepresentation of others.
  - In contrast, stratified sampling specifically ensures proportional representation by dividing the population into distinct subgroups and selecting a random sample from each.
- "Stratified sampling does not use randomization, whereas simple random sampling relies entirely on random selection" → **Incorrect.** Stratified sampling incorporates randomization, but unlike simple random sampling, it first divides the population into distinct strata based on specific characteristics. Once the strata are defined and the target sample size for each is determined, a random sample is taken from each stratum.
- The correct option (☒) matches the definitions discussed in class and in the textbook.

### Question 1.E (1 point)

From the following options, select all the correct statements:

(Note: Incorrect selections will discount partial credit from this question)

#### Question 1.E options:

- ☒ A data set includes all collected observations, typically organized in a structured format (+0.25 if selected)
- ☐ The population includes only those individuals who are surveyed in a study (+0.25 if NOT selected; -0.25 if selected)
- ☒ An observation refers to an individual unit of data collected in a study (+0.25 if selected)
- ☒ A sample consists of a subset of the population that is analyzed to draw conclusions (+0.25 if selected)

#### Justification:

- **First option: Correct.** A dataset is a structured collection of data points, usually organized in rows (observations) and columns (variables), representing *all collected information in a study*.
- **Second option: Incorrect.** The *population* refers to the entire group of interest, while the sample consists of the individuals surveyed. A study surveys only a subset of the population, not the entire group.
- **Third option: Correct.** An observation is a *single* unit of analysis within a dataset, such as one respondent's survey answers or one recorded event.
- **Fourth option: Correct.** A *sample* is a smaller group selected from the population to infer conclusions about the whole population.