Which of these is *not* true about a survey conducted on a random sample from a population?

A It can establish a causal relationship.

B The results can be generalized to a larger population.

C It is an observational study.

D Bias can be introduced in the way questions are asked.

E It can be used by researchers to make estimates about the population.

Which of these is *not* true about a survey conducted on a random sample from a population?

(A) It can establish a causal relationship.

(B) The results can be generalized to a larger population.

(C) It is an observational study.

(D) Bias can be introduced in the way questions are asked.

(E) It can be used by researchers to make estimates about the population.

A

- **Random Sampling** – 随机抽样
  - **Sampling Frame** – 抽样框
- **Simple Random Sample** – 简单随机样本
  - **Stratified Sampling** – 分层抽样
  - **Cluster Sampling** – 集群抽样

$P(A) + P(A^c) = 1$

Complementary variable or complement

Z = a + bi **Conjugate** **z* = a-bi**

1. **Population (总体)** – The entire group being studied.
2. **Sample (样本)** – A subset of the population selected for analysis.
3. **Random Sampling (随机抽样)** – Selecting a sample so that every member of the population has an equal chance of being chosen.
4. **Bias (偏差 / 偏倚)** – Systematic errors that make a sample unrepresentative of the population.
5. **Sampling Error (抽样误差)** – The natural difference between a sample statistic and the true population parameter.

A company wants to survey its employees about job satisfaction. There are 500 employees in total. The company plans to sample 50 employees for the survey.

(a) Explain why **simple random sampling** would be the best method for selecting employees for the survey.
(b) If the company decided to use **stratified sampling**, how would the sampling process change?
(c) What potential **bias** could occur if the company used **convenience sampling** to select employees?
(d) Suppose the company uses **cluster sampling** by selecting 5 departments and surveying all employees in those departments. Explain how this sampling method works.

(a) Why simple random sampling is the best method:

- **Simple random sampling (简单随机样本)** ensures that every employee has an equal chance of being selected, reducing bias and making the sample representative of the entire population. It provides an unbiased, fair method of selection.

(b) How stratified sampling changes the process:

- **Stratified sampling (分层抽样)** would divide the employees into subgroups (strata), such as by department or job position, and then randomly sample from each subgroup. This ensures that each subgroup is proportionately represented in the survey.

(c) Potential bias with convenience sampling:

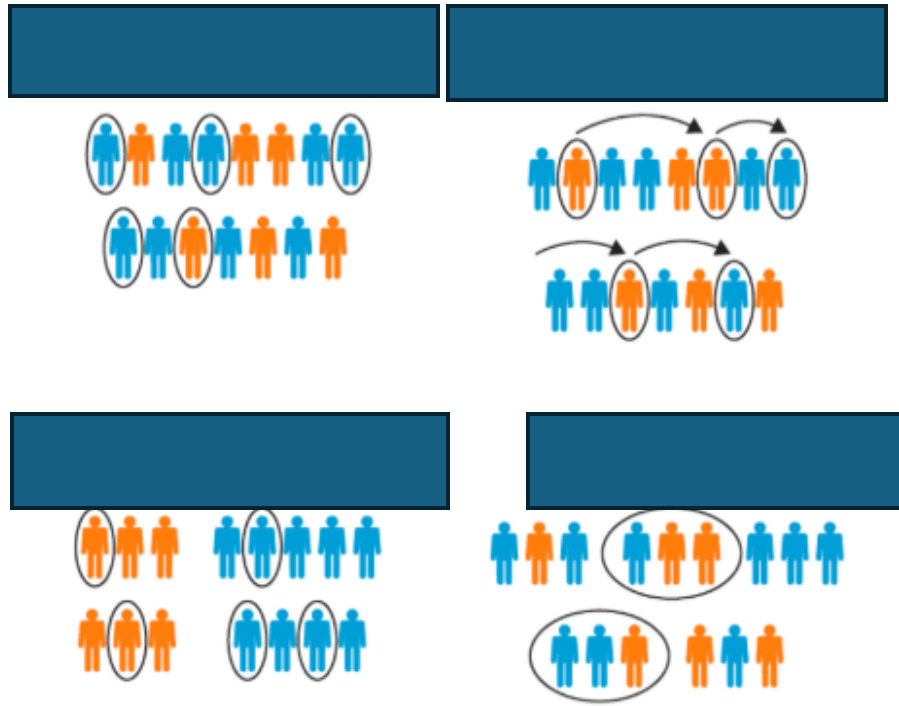
- **Convenience sampling (便利抽样)** could introduce **selection bias (选择偏差)** because employees who are easiest to reach (e.g., those in the same department or near the survey location) may not be representative of the entire workforce. This could skew the results if certain groups are over- or underrepresented.

(d) How cluster sampling works:

- **Cluster sampling (集群抽样)** involves dividing the population into groups (clusters), such as departments, and then randomly selecting a few clusters to survey. All employees within the selected clusters are included in the survey. This is useful when the population is geographically spread out or when it's easier to survey entire groups.

Identify a Sampling Method

In statistics, when we cannot collect data from the entire population, we select a sample. The way the sample is chosen affects the reliability of conclusions. Using chance methods reduces bias and makes the sample more representative.



Population and Sample:

- **Population:** The group we want information about.
- **Sample:** The group actually studied, chosen from the population.

Common Sampling Methods:

Simple Random Sample (SRS): Every individual and every group of individuals has an equal chance of being chosen.

- Unbiased and representative if truly random.
- Gives each subset of the population equal chance.
- Good for statistical analysis and inference.
- Requires complete list of population (sampling frame).
- May be difficult or time-consuming for very large populations.

Stratified Random Sample: The population is divided into groups (strata) based on a shared characteristic, and an SRS is taken within each stratum.

- Ensures representation from important subgroups.
- Reduces variability between samples.
- More precise than a simple random sample if strata are meaningful.
- Requires knowledge of population characteristics to form strata.
- More complex and time-consuming to design.

Cluster Sample: The population is divided into clusters (often naturally occurring groups), and entire clusters are randomly selected. All individuals in chosen clusters are surveyed.

- Efficient and cost-effective for large, spread-out populations.
- Does not require a complete list of all individuals.
- Easy to implement when groups already exist (schools, neighborhoods).
- Risk of bias if chosen clusters are not representative.
- Less precise than stratified sampling.

Systematic Sample: Selects every n-th individual from a list or sequence, after a random starting point.

- Simple and quick to implement.
- Ensures sample is evenly spread across population list.
- Good for quality control and assembly-line settings.
- Can introduce bias if there is a hidden pattern in the data list.
- Not as random as SRS, but still widely used in practice.

Convenience Sample: Individuals are chosen because they are easy to reach.

- Quick, cheap, and easy to collect.
- Often used in preliminary or pilot studies.
- High risk of bias; may not represent the population well.
- Results cannot be generalized reliably.
- Common in everyday surveys but discouraged in scientific research.

Sampling Method	How It Works	Strengths / Weaknesses
[redacted]	All individuals/groups equally likely to be chosen	Unbiased; requires full list; can be impractical for large populations
[redacted]	Divide population into strata; take SRS within each	Reduces variability; ensures subgroup representation; more complex
[redacted]	Divide into clusters; randomly choose clusters; include all individuals in chosen clusters	Cheap and easy; may lead to unrepresentative samples
[redacted]	Choose every n-th person after random start	Quick; evenly spaced; biased if hidden pattern exists
[redacted]	Choose individuals easiest to reach	Fast but biased; results not reliable

Example

A teacher writes the names of all 30 students on slips of paper, puts them in a hat, and draws 5 names.

What type of sampling method is this?

Conclusion: This is a **Simple Random Sample**.

A researcher wants to survey high school students in a city. They divide schools into public and private, then randomly select students from each group.

What type of sampling method is this?

Conclusion: This is a **Stratified Random Sample**.

A researcher wants to survey high school students in a city. They divide schools into public and private, then randomly select students from each group.

What type of sampling method is this?

Conclusion: This is a **Stratified Random Sample**.

To study eating habits, a researcher randomly selects 3 classrooms in a school and surveys every student in those classrooms.

What type of sampling method is this?

Conclusion: This is a **Cluster Sample**.

Explain Why a Sampling Method Is or Is Not Appropriate

Choosing the correct sampling method is essential to make valid conclusions in statistics. Some methods are better suited to certain populations or research goals. If the wrong method is chosen, the data may be biased or unrepresentative.

Key Ideas:

- **Representativeness:** Does the method fairly reflect the population?
- **Bias:** Does the method systematically favor certain outcomes?
- **Practicality:** Is the method realistic in terms of time, cost, and effort?
- **Sample Size:** Is the number of individuals large enough to reduce variability?
- **Subgroup Representation:** Are important groups included in the sample?

Sampling Method	Appropriate When...	Not Appropriate When...
[redacted]	Every individual can be listed and chosen randomly; unbiased method needed.	Population too large to list, or too costly/time-consuming.
[redacted]	Important subgroups must be represented (e.g., gender, grade level).	Strata are hard to identify or irrelevant to the research.
[redacted]	Population naturally divided into groups (classrooms, neighborhoods).	Clusters differ greatly from each other; not representative.
[redacted]	Population list available; periodic selection is efficient.	Hidden patterns align with sampling interval, creating bias.
[redacted]	Quick, low-cost, exploratory research (not formal inference).	When results must generalize to the entire population.

Example

A college administrator wants to estimate the average GPA of students. She decides to randomly select 100 students from the entire enrollment list.

Is this sampling method appropriate?

Step 1: This is a **Simple Random Sample**, since every student had an equal chance of being selected.
Step 2: It avoids bias and fairly represents the population, as long as the enrollment list is complete.
Conclusion: Yes, this method is **appropriate** and produces reliable results.

Example

A college administrator wants to estimate the average GPA of students. She decides to randomly select 100 students from the entire enrollment list.

Is this sampling method appropriate?

Step 1: This is a **Convenience Sample** because only students near the library were included.
Step 2: Students at the library may not represent all students (they may study more than average).
Conclusion: This method is **not appropriate** for generalizing to the entire population.

A school district wants to know average internet use among students. They randomly select 5 schools in the district and survey all students in those schools.

Is this sampling method appropriate?

Step 1: This is a **Cluster Sample**, because entire schools were randomly chosen and all students in those schools were surveyed.
Step 2: If schools are similar, this is efficient and appropriate. If schools differ a lot (urban vs rural, large vs small), results may be biased.
Conclusion: The appropriateness depends on whether schools are representative of the district as a whole.

TRUE or FALSE

- Stratified sampling divides the population into groups and then randomly samples **from each group**.

TRUE or FALSE

- Stratified sampling divides the population into groups and then randomly samples **from each group**.

TRUE

TRUE or FALSE

- Increasing the **sample size** always eliminates all sources of bias.

FALSE

TRUE or FALSE

- Increasing the **sample size** always eliminates all sources of bias.

TRUE or FALSE

- Random sampling guarantees that the sample mean will equal the population mean.

TRUE or FALSE

- Random sampling guarantees that the sample mean will equal the population mean.

FALSE

STRETCH AND CHALLENGE

- **Scenario: Student Sleep Study**
- A university wants to study **average sleep hours per night** among students. They consider several sampling methods:
- **Method A:** Email a survey to the first 500 students who log into the university portal.
- **Method B:** Randomly select 50 students from each academic major.
- **Method C:** Randomly select 100 students from a list of all students.
- **Method D:** Post a survey on social media and accept responses voluntarily.
-

- **Tasks**
- **(a)** Identify which methods are **likely to produce biased samples** and explain why.
- **(b)** For Method B, explain why **stratified sampling** might give a more accurate estimate of the population mean than Method C.
- **(c)** Suppose 20% of students selected in Method C do **not respond**. Explain how **nonresponse bias** could affect the results.
- **(d) Challenge:**
  The university decides to use cluster sampling by randomly selecting 5 dormitories and surveying **all students living there**. Discuss one **advantage** and one **potential problem** with this approach.
- **(e) Synthesis:**
  Explain how random sampling, sample size, and bias interact to affect the **accuracy and validity** of conclusions.

(a) Biased: Method A (first 500 logins), Method D (voluntary response). Methods B and C are generally unbiased.

(b) Stratified sampling (Method B) ensures all majors are represented, reducing variability and giving a more precise estimate than simple random sampling.

(c) Nonresponse in Method C can bias results if nonresponders differ systematically, reducing accuracy.

(d) Cluster sampling: Advantage — cheaper and easier; Problem — dorm students may be similar, less representative.

(e) Random sampling reduces bias; larger samples reduce variability; bias (undercoverage, voluntary response, nonresponse) can invalidate conclusions.

Bias and Variability

Bias	偏差
Sampling bias	抽样偏差
Nonresponse bias	非响应偏差
Variability	变异性
Precision	精确度

- **Bias** is a systematic error causing estimates to consistently miss the true population value.
- **Sampling bias** happens when some population members are less likely to be included.
- **Nonresponse and voluntary response** can introduce bias because responders may differ from non-responders.
- **Variability** measures how much sample statistics differ between samples.
- **Increasing sample size** reduces variability but **does not remove bias**.

A city wants to estimate the average daily screen time of its high school students. Three survey methods are considered:

1. Sending an online survey to the first 200 students who log into the school portal.
2. Randomly selecting 50 students from each grade level and surveying them.
3. Posting a voluntary survey on social media.

(a) Identify which methods are **likely biased** and explain why.

(b) Explain the difference between **bias and variability** in the context of this study.

(c) If method 2 is used with a sample size of 200, what effect would **doubling the sample size** have on variability and bias?

(d) A student claims: "Increasing the sample size will make the estimate perfectly accurate." Evaluate this statement.

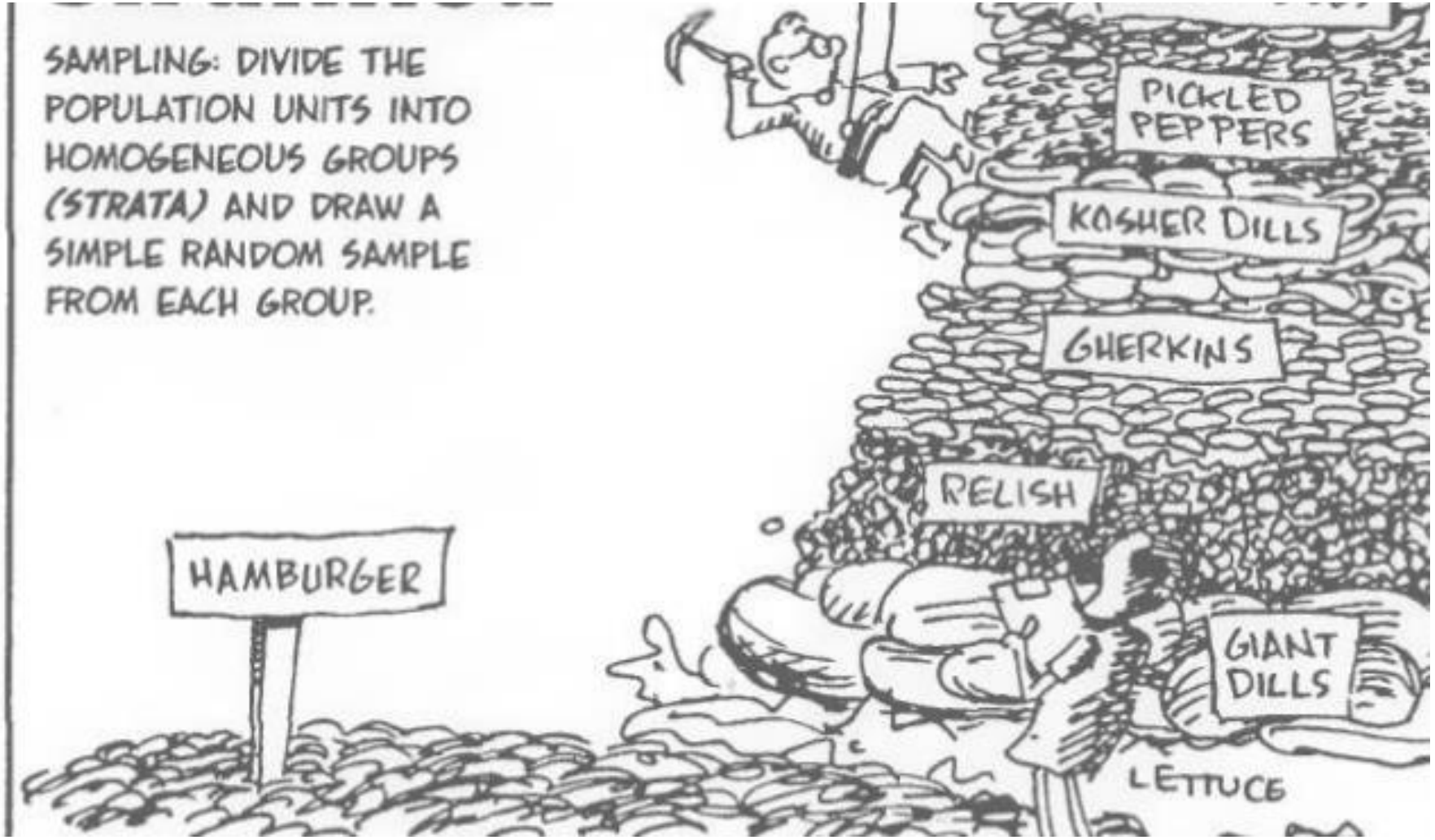
(a) Biased: Method 1 (first logins) and Method 3 (voluntary response); Method 2 is unbiased.

(b) Bias = systematic error; Variability = natural differences between samples.

(c) Doubling sample size → reduces variability (more precise); bias unchanged.

(d) Statement is incorrect → larger sample improves precision but **does not remove bias**.

Systematic sampling



Stratified. sampling

AN EXAMPLE IS A CITY HOUSING SURVEY WHICH DIVIDES A CITY INTO BLOCKS, RANDOMLY SAMPLES THE BLOCKS, AND LOOKS AT EVERY HOUSING UNIT IN EACH SAMPLED BLOCK.