eLearning

Statistical Modeling

E-Learning Day1

📘 Lesson: Sample Surveys, Observational Studies, and Experiments

🎯 Learning Objective (SM.DPSR.3.2)

Students will be able to:
- Define and distinguish between sample surveys, observational studies, and experiments.
- Identify examples of each in real-world contexts.
- Explain when each method is appropriate.

🏳 Introduction (Hook)

“How do scientists, businesses, and policymakers gather information to make decisions?”

Example: “If a school wants to know students’ favorite lunch option, what’s the best way to find out?”

📖 Take notes in notebook (10-15 minutes)

1. Sample Surveys

Definition: .A method of collecting data from a subset of the population to make inferences about the whole population

Key feature: No manipulation, only asking questions or collecting responses.

Example: A poll of 200 students about favorite music genres.

2. Observational Studies

Definition: Researchers observe and record data without controlling variables.

Key feature: No intervention, just observation.

Example: Recording how much time students spend on homework each night.

3. Experiments

Definition: A method where researchers manipulate one or more variables and measure the effect on an outcome.

Key feature: Includes treatment groups, control groups, and random assignment.

Example: Giving one group of students extra study sessions to see if their test scores improve.

🔑 Comparison Table

Feature	Sample Survey	Observational Study	Experiment
Collect data by	Asking questions	Watching/recording	Manipulating variables
Control over variables	❌ None	❌ None	✅ Yes
Example	Student poll	Watching sleep habits	Testing effect of tutoring

📝 Quick Quiz (2-3 minutes)

1. Which method allows researchers to control variables?
   a) Survey
   b) Observational study
   c) Experiment

2. A local newspaper asks 500 people which candidate they support. This is a:
   a) Observational study
   b) Experiment
   c) Survey

🎯 Conclusion

Surveys: Ask questions

Observational studies: Watch/record

Experiments: Test with control & manipulation

https://youtu.be/5o57j4FEIpM

E-learning **Day 1**

Worksheet: Sample Surveys, Observational Studies, and Experiments

Name: _____ Date: _____ (35-40 minutes)

Instructions: Choose the best answer for each multiple-choice question. Then answer the descriptive questions at the end.

1. A teacher asks students about their favorite subject. This is an example of:

- a) Survey
- b) Observational study
- c) Experiment
- d) Case study

2. A researcher records how many hours teens spend watching TV without interfering. This is:

- a) Experiment

- b) Survey
- c) Observational study
- d) Interview

3. A scientist tests whether a new fertilizer increases plant growth by using two groups of plants. This is:

- a) Survey
- b) Observational study
- c) Experiment
- d) Simulation

4. Which method involves manipulating variables?

- a) Survey
- b) Observational study
- c) Experiment
- d) Interview

5. A newspaper asks 1,000 people who they plan to vote for. This is:

- a) Experiment
- b) Survey
- c) Observational study
- d) Random assignment

6. Researchers watch how birds behave in their natural habitat without interfering. This is:

- a) Survey
- b) Experiment

- c) Observational study
- d) Simulation

7. A psychologist gives half of the participants a memory test after meditation training. This is:

- a) Survey
- b) Observational study
- c) Experiment
- d) Case study

8. Which method relies on random assignment to reduce bias?

- a) Survey
- b) Observational study
- c) Experiment
- d) Interview

9. Which method best describes polling students to find their average bedtime?

- a) Survey
- b) Experiment
- c) Observational study
- d) Trial

10. A researcher wants to know if caffeine affects alertness. Half the students drink coffee, the other half drink water. This is:

- a) Survey
- b) Observational study
- c) Experiment

- d) Interview

11. Which is the primary advantage of experiments over surveys?

- a) Easier to conduct
- b) Can establish cause-and-effect
- c) Uses larger groups
- d) Provides opinions

12. Which method would a school use to find the most popular lunch menu?

- a) Survey
- b) Experiment
- c) Observational study
- d) Simulation

13. A wildlife biologist observes deer migration routes without altering their environment. This is:

- a) Survey
- b) Experiment
- c) Observational study
- d) Simulation

14. A pharmaceutical company gives some patients a real drug and others a placebo. This is:

- a) Survey
- b) Experiment
- c) Observational study
- d) Interview

15. Which method does not allow cause-and-effect conclusions?

- a) Experiment
- b) Survey
- c) Both survey and observational study
- d) None

16. Asking 500 people how often they exercise is an example of:

- a) Survey
- b) Experiment
- c) Observational study
- d) Simulation

17. Watching how often students check their phones during class is:

- a) Survey
- b) Observational study
- c) Experiment
- d) Simulation

18. Which method is most likely to use a control group?

- a) Survey
- b) Observational study
- c) Experiment
- d) Interview

19. Which is a disadvantage of surveys?

- a) They can be biased
- b) They establish cause-and-effect
- c) They require random assignment
- d) They involve manipulation

20. Which is a disadvantage of observational studies?

- a) They rely on interviews
- b) They cannot establish cause-and-effect
- c) They are unethical
- d) They always use random assignment

21. A teacher gives extra tutoring to one group of students to see if grades improve. This is:

- a) Survey
- b) Observational study
- c) Experiment
- d) Simulation

22. A company asks customers to rate their satisfaction on a scale of 1–10. This is:

- a) Survey
- b) Observational study
- c) Experiment
- d) Case study

23. Observing how often people wash their hands in a restroom is:

- a) Survey
- b) Observational study

- c) Experiment
- d) Interview

24. Which of the following is an experiment?

- a) Measuring rainfall
- b) Testing effect of fertilizer on crops
- c) Asking opinions on climate change
- d) Recording average height

25. Which method is best when you want to generalize about a population from a sample?

- a) Survey
- b) Observational study
- c) Experiment
- d) Case study

1. Compare and contrast sample surveys, observational studies, and experiments. Provide an example of each.

2. Explain why experiments are more reliable than surveys or observational studies when trying to establish cause-and-effect relationships.

E-learning Day 2

(SM.DPSR.3.3) Designing Surveys, Experiments, and Observational Studies

Learning Objectives:

By the end of this lesson, I will be able to:

1. Design sample surveys, experiments, and observational studies.

2. Identify key aspects of best statistical practice, including: - Formulating a clear statistical question.

Identifying the variables being investigated. - Applying random selection and/or random assignment.

3. Explain the rationale behind these design choices.

Introduction

- Real-world examples:
  - Sample Survey: “What percentage of students prefer online learning?”
  - Experiment: “Does the type of study method affect test scores?”
  - Observational Study: “Observe the correlation between hours of sleep and concentration levels in class.”

Importance of good design:

- ensures reliability
- unbiased results
- and interpretability.

2. Key Concepts

Study Type	Definition	Key Design Features
Sample Survey	Collects information from a subset of a population to estimate population characteristics	Clear statistical question, representative sample, random selection
Experiment	Actively imposes treatments to study effects on outcomes	Control group, treatment group, random assignment, replication
Observational Study	Observes outcomes without interference	Identifying variables, careful measurement, consideration of confounding factors

Best Practices in Study Design:

1. Statistical Question: Specific, measurable, and answerable.

- Example: “Does listening to classical music improve memory test scores?”

2. Variables: - Independent Variable: What is manipulated (e.g., music type).

- Dependent Variable: What is measured (e.g., memory test score).

3. Randomization: - Random Selection: Ensures sample represents population.

- Random Assignment: Reduces bias in experiments.

- Task: Design a study based on a prompt. (15-20 minutes)

- Prompt Example: “Investigate whether exercise improves concentration among high school students.”

- Instructions:

1. Identify the statistical question.

2. Identify independent and dependent variables.

3. Decide on survey, experiment, or observational study.

4. Plan random selection or assignment.

- Submission: Short paragraph or flowchart.

Part A: Multiple Choice Questions (25 Questions) (30-35 minutes)

1. Which of the following is a key feature of a well-designed experiment?

a. Observing participants without intervention

b. Random assignment of treatments

c. Selecting participants based on convenience

d. Ignoring variables

2. A statistical question should be:

a. Vague and general

b. Specific and measurable

c. Unrelated to data

d. Based only on opinion

3. Which of the following is an independent variable?

a. Test scores

b. Type of study method

c. Number of students

d. Average age

4. An observational study:

a. Randomly assigns treatments

b. Observes without manipulating variables

c. Always uses a control group

d. Ignores confounding variables

5. Random selection ensures that:

a. The sample represents the population

b. Treatments are applied fairly

c. Only the best students are included

d. Data collection is biased

6. Random assignment is used in:

a. Surveys

b. Experiments

c. Observational studies

d. Case studies

7. Which is an example of a survey question?

a. “Does caffeine affect alertness?”

b. “How many hours do you sleep on school nights?”

c. “Compare two teaching methods on test scores.”

d. “Assign students to treatment groups randomly.”

8. Confounding variables are:

a. Variables that are controlled

b. Variables that may affect results unintentionally

c. Variables that are dependent

d. Variables that are always independent

9. In an experiment testing the effect of study time on grades, grades are the:

a. Independent variable

b. Dependent variable

c. Random variable

d. Confounding variable

10. Which study type cannot prove causation?

a. Experiment

b. Observational study

c. Randomized controlled trial

d. Laboratory experiment

11. Which is an example of random selection?

a. Choosing students from a single class only

b. Selecting students using a lottery system

c. Picking friends for a survey

d. Assigning treatments arbitrarily

12. Best statistical practice includes:

a. Ignoring variability

b. Clearly identifying the statistical question

c. Avoiding randomization

d. Using only small samples

13. A control group in an experiment:

a. Receives no treatment

b. Receives the experimental treatment

c. Is selected after the experiment

d. Is optional in observational studies

14. Which is NOT a variable type?

a. Independent

b. Dependent

c. Confounding

d. Randomized

15. The purpose of a sample survey is to:

a. Test causation

b. Estimate population characteristics

c. Assign treatments

d. Manipulate variables

16. In an experiment on study methods, which is a confounding variable?

a. Test scores

b. Prior knowledge

c. Study method

d. Random assignment

17. Why is random assignment important?

a. To increase bias

b. To reduce the effect of confounding variables

c. To select the most intelligent participants

d. To eliminate dependent variables

18. Which statement is an example of a statistical question?

a. “Do students like chocolate?”

b. “What is the average number of hours students sleep per night?”

c. “Who is the best student in class?”

d. “Is math hard?”

19. Observational studies are best used when:

a. Random assignment is impossible or unethical

b. Treatments can be safely assigned

c. The researcher wants full control

d. The sample size is 1

20. Which study type uses both random selection and random assignment?

a. Survey

b. Experiment

c. Observational study

d. Case study

21. To reduce bias, a study should:

a. Ignore variability

b. Include randomization

c. Avoid control groups

d. Only study volunteers

22. Which of the following is NOT part of best practice in experiment design?

a. Clear research question

b. Identification of variables

c. Random assignment of treatments

d. Using a single participant only

23. A survey asks: “How often do students exercise each week?” The variable being measured is:

a. Student names

b. Number of exercise sessions

c. Gender

d. Height

24. When an experiment cannot be performed, researchers should:

a. Guess results

b. Conduct an observational study

c. Ignore the question

d. Use the control group only

25. Which is an example of a well-phrased statistical question?

a. “Do boys or girls run faster?”

b. “What is the average number of hours of homework completed by high school students per week?”

c. “Who is the smartest student?”

d. “Do people like pizza?”

E-learning **Day 3**

Review

SM.DPSR.1.1Calculating and Interpreting Z-Scores

Learning objectives: Students will be able to :

Calculate and interpret z-scores as a measure of relative standing to standardize units

Apply the concept to solve real life problems

1. What is a Z-Score?

A **z-score** is a measure that tells us **how many standard deviations a value is from the mean**.

It allows us to **standardize different data points**, making them comparable even if they come from different scales or units.

2. Z-Score Formula

$z=(x-\mu)/\sigma$

Where:

- x = the value of interest
- μ= mean of the dataset
- σ= standard deviation of the dataset

. Step-by-Step Example

Suppose a test has:

- Mean (μ) = 80
- Standard deviation (σ) = 5
- Your score (x) = 90

Step 1: Subtract the mean from your score:

$x - \mu = 90 - 80 = 10$

Step 2: Divide by the standard deviation:

$z = 10/5 = 2$

$z = 2$

✅ **Z-score =**

4. Interpretation of Z-Scores

- **Positive z-score** → value is **above the mean**.
- **Negative z-score** → value is **below the mean**.
- **Magnitude** → number of standard deviations away from the mean.

Example:

- $z = 2$ means your score is **2 standard deviations above the mean**.
- In a normal distribution, this score is approximately in the **top 2.5%**.

5. Uses of Z-Scores

1. **Standardize different units** → Compare data measured in different units (e.g., height vs. weight).
2. **Compare across tests** → Compare scores from different exams.
3. **Identify outliers** → Values with $|z| > 2$ or $|z| > 3$ are unusual.
4. **Calculate probabilities** → Find percentiles using normal distribution tables.

Topic: Z-Scores and Standardization

Instructions:
- Choose the correct answer for multiple-choice questions.
- Show all your work for descriptive questions.

Multiple Choice Questions (MCQs) – 25 Questions (35-40minutes)

1. A score of 85 in a test has a mean of 80 and standard deviation of 5. What is the z-score?
a) 0.5
b) 1
c) -1
d) 1.5

2. If a z-score is 0, it means the score is:
a) Above the mean
b) Equal to the mean
c) Below the mean
d) Cannot determine

3. Which of the following represents a negative z-score?
a) 2.5
b) -1.3
c) 0
d) 1

4. The mean of a dataset is 100 and the standard deviation is 10. A score of 120 has a z-score of:
a) 1
b) 2
c) 1.5
d) 2.5

5. Z-scores are useful for:
a) Measuring central tendency only
b) Standardizing scores
c) Changing the units of scores arbitrarily
d) None of the above

6. If a score is 2 standard deviations above the mean, its z-score is:
a) -2
b) 0
c) 2
d) 1

7. Which z-score corresponds to the lowest relative standing?
a) 0
b) -3
c) 1
d) 2

8. A z-score tells us:
a) The exact value of the score
b) The position of the score relative to the mean
c) The mean of the data
d) The variance of the data

9. The mean of z-scores in a dataset is always:
a) 0
b) 1
c) Equal to the original mean
d) Cannot determine

10. The standard deviation of z-scores is always:
a) 0
b) 1
c) Equal to the original standard deviation
d) Cannot determine

11. Which of the following scores is the most unusual?
a) $z = 0.5$
b) $z = -0.2$
c) $z = 3$
d) $z = 1$

12. If a test score has $z = -1.5$, it is:
a) Above the mean
b) Below the mean
c) Equal to the mean
d) Cannot determine

13. Convert a z-score of 2 to the original score if mean = 50 and SD = 5.
a) 55
b) 60
c) 45
d) 52

14. A z-score of -2 indicates the score is:
a) 2 SD above the mean
b) 2 SD below the mean
c) Equal to the mean
d) Cannot determine

15. For a normal distribution, about 68% of data lie within:
a) ±1 z
b) ±2 z
c) ±3 z
d) ±0.5 z

16. A score of 90 in a dataset with mean = 100 and SD = 10 has z-score:
a) 1
b) -1
c) 0
d) 2

17. Z-scores are used to compare:
a) Scores from different distributions
b) Only raw scores
c) Variance of data
d) Median values

18. If all scores in a dataset increase by 5, the z-scores:
a) Increase by 5
b) Decrease by 5
c) Remain unchanged
d) Become negative

19. The formula for z-score is:
a) $z = (x - \mu)/\sigma$
b) $z = x \times \mu$
c) $z = \sigma - x$
d) $z = \sigma/x$

20. A negative z-score means the score is:
a) Above mean
b) Below mean
c) Equal to mean
d) Cannot determine

21. A score of 70 in a dataset with mean = 80 and SD = 5 has z-score:
a) -2
b) -1
c) 1
d) 2

22. Which z-score represents a score higher than the mean?
a) -0.5
b) 0
c) 1.2
d) -1.5

23. If mean = 50, SD = 4, what is the z-score of 58?
a) 2
b) 1
c) -2
d) 0

24. The primary purpose of calculating z-scores is:
a) Identify exact scores
b) Compare relative positions
c) Change units
d) Calculate variance

25. A student has z = -0.8. The score is:
a) Slightly above mean
b) Slightly below mean
c) Much above mean
d) Much below mean

Descriptive Questions

1. A test has a mean of 75 and a standard deviation of 8. Calculate the z-scores for the following scores: 60, 80, 90. Interpret each z-score.

2. The weights of a sample of students have a mean of 65 kg and a standard deviation of 7 kg. If a student weighs 50 kg, calculate the z-score and explain whether this weight is unusual or not.

E-learning **Day 4**

Topic: Confidence Intervals and Margin of Error

Standard: SM.DPSR.5.4 – Interpret confidence intervals and understand how sampling variability affects estimates.

Objectives: By the end of this lesson, I will be able to:

1. Understand sampling variability and why different samples give slightly different results.
2. Define confidence interval and margin of error.
3. Interpret a confidence interval in the context of a real-world situation.

Key Concepts:

1. Sampling Variability:
   - Different samples from the same population can give slightly different results.
   - Example: Surveying 50 students about their favorite sport might give different percentages each time.

2. Summary Statistic:
   - A number summarizing the sample (e.g., sample mean, sample proportion).

3. Margin of Error (MOE):

- Indicates how much the sample estimate might differ from the true population parameter.
- Forms the basis for the confidence interval.

4. Confidence Interval (CI):

- Range around the sample estimate likely to contain the true population parameter.

- Example: A poll says 60% of students like online learning ±5%.

- CI: 55% to 65%

- Interpretation: "We are confident that the true proportion of students who like online learning is between 55% and 65%."

Steps to Interpret a Confidence Interval:

1. Identify the sample statistic (e.g., 60%).

2. Identify the margin of error (e.g., ±5%).

3. Calculate the interval: lower bound = sample − MOE, upper bound = sample + MOE.

4. Interpret in context: Explain in words what the interval tells you about the population.

Example:

Scenario: A survey of 100 students finds that 40% prefer school uniforms, with a margin of error of ±6% at 95% confidence.

Confidence Interval: 40% − 6% = 34%,

40% + 6% = 46%

Interpretation: "We are 95% confident that between 34% and 46% of all students prefer school uniforms."

E-learning **Day 4**

Worksheet: Confidence Intervals & Margin of Error **(25-30 Minutes)**

Instructions: Choose the correct answer for each question.

Multiple Choice Questions

5. A survey finds that 60% of students like online learning with a margin of error of ±4%. What is the confidence interval?

a. 56%–64%

b. 60%–64%

c. 56%–60%

d. 55%–65%

6. A sample shows 45% of students bring lunch from home with ±5% margin of error. The interval is:

a. 40%–50%

b. 45%–50%

c. 40%–45%

d. 50%–55%

7. If the confidence level increases, the margin of error:

a. Decreases

b. Increases

c. Stays the same

d. Becomes zero

8. A survey of 100 students reports 30% ± 6%. The lower bound of the CI is:

a. 24%

b. 30%

c. 36%

d. 34%

9. Which statement is correct about a 95% confidence interval?

a. It always contains the true population value

b. 95% of sample statistics will fall in this interval

c. We are 95% confident the true population value lies within the interval

d. None of the above

10. A poll reports 50% ± 3%. The true proportion is likely:

a. 47%–53%

b. 45%–55%

c. 50%–53%

d. 48%–52%

11. Sampling variability refers to:

a. Differences in populations

b. Differences between sample statistics from different samples

c. Measurement errors

d. None of the above

12. The margin of error depends on:

a. Sample size

b. Population variability

c. Confidence level

d. All of the above

13. Which of the following is a valid 90% confidence interval for a sample estimate of 70% with ±5% MOE?

a. 65%–75%

b. 60%–80%

c. 70%–75%

d. 66%–74%

14. If a sample of 50 students gives 60% ± 6%, increasing the sample size generally:

a. Increases MOE

b. Decreases MOE

c. MOE stays the same

d. Cannot tell

15. A CI is wider when:

a. Confidence level is lower

b. Confidence level is higher

c. Sample size increases

d. Population size decreases

16. What does ±4% in a survey result mean?

a. The estimate is wrong by 4%

b. True value is within 4% above or below the estimate

c. 4% of respondents are unreliable

d. None of the above

17. A 99% confidence interval is generally:

a. Narrower than 95% CI

b. Wider than 95% CI

c. Same as 95% CI

d. Half the width of 95% CI

18. Sample statistic of 55% ± 5% indicates:

a. True proportion is exactly 55%

b. Interval 50%–60% likely contains true proportion

c. Only 5% chance estimate is wrong

d. None of the above

19. Which factor does NOT affect MOE?

a. Sample size

b. Confidence level

c. Population size (for large populations)

d. Population mean

20. A CI for mean height of students is 160–170 cm. Interpretation:

a. All students are between 160–170 cm

b. We are confident the average height is between 160–170 cm

c. Sample mean is 170 cm

d. None of the above

21. Margin of error decreases when:

a. Sample size increases

b. Population variability increases

c. Confidence level increases

d. None of the above

22. True or False: A 95% CI guarantees 95% of individuals in population fall in the interval.

a. True

b. False

23. If MOE is ±3% and sample proportion is 62%, the CI is:

a. 65%–68%

b. 59%–65%

c. 62%–65%

d. 60%–63%

24. A 90% CI for proportion of students who like pizza is 0.40–0.50. Which is correct?

a. True proportion is exactly 0.45

b. We are 90% confident true proportion is between 0.40–0.50

c. 40% of students like pizza

d. None of the above

E-Learning **Day5**

Review

SM.DPSR.4.1 Quantitative and Categorical Data with Displays and Summaries

Learning Objective: Students will be able to:

Identify types of displays that are appropriate for categorical data (pie chart, bar chart, pareto chart) versus quantitative data (histograms, stem plots, box plots, dot plots).

Use tables (include relative frequency table), graphical displays (include histograms, modified box plots), and numerical summary statistics

Understanding Data Types

Categorical Data

- Describes qualities, labels, or categories.
- Examples: Favorite color, type of pet, gender, car brand.

Quantitative Data

- Describes numerical measurements or counts.
- Examples: Age, height, test scores, number of siblings.

2. Displays for Data

Categorical Data → Appropriate Displays

- **Pie Chart** → shows parts of a whole.
- **Bar Chart** → compares frequency across categories.
- **Pareto Chart** → a bar chart with categories ordered from most to least frequent.

Quantitative Data → Appropriate Displays

- **Histogram** → shows distribution of data in intervals (bins).

- **Stem-and-Leaf Plot** → preserves original data values while showing distribution.
- **Box Plot (Modified)** → shows median, quartiles, and outliers.
- **Dot Plot** → shows individual data points along a number line.

3. Tables: Frequency Table Example (Categorical Data: Favorite Fruit)

Fruit	**Frequency**	**Relative Frequency**
Apple	10	0.25 (25%)
Banana	8	0.20 (20%)
Orange	12	0.30 (30%)
Grapes	10	0.25 (25%)
Total	**40**	**1.00 (100%)**

4. Numerical Summary Statistics (Quantitative Data)

- **Measures of Center:** Mean, Median, Mode
- **Measures of Spread:** Range, Variance, Standard Deviation, Interquartile Range (IQR)
- **Five-Number Summary:** Minimum, Q1, Median, Q3, Maximum