STATISTICS

A STATISTICAL INVESTIGATION PROCESS

A.1 A STEP-BY-STEP INVESTIGATION

A.1.1 IDENTIFYING THE STEPS OF AN INQUIRY

MCQ 1: Read the statement: "The girls' average score in math is 87, while the boys' average is 75. Since $87 > 75$, this data suggests that, on average, the girls in this group performed better."
Which step of the statistical process does this describe?

- ☐ Step 1: Pose the Question
- ☐ Step 2: Collect Data
- ☐ Step 3: Analyze Data
- ☐ Step 4: Represent Data (e.g., create charts)
- ☒ Step 5: Interpret the Results

Answer: **Step 5: Interpret the Results**. This statement draws a conclusion by comparing two statistics (the averages) to answer a potential question about performance.

MCQ 2: Read the statement: "Our investigation will explore whether students in our school prefer science over math."
Which step of the statistical process does this describe?

- ☒ Step 1: Pose the Question
- ☐ Step 2: Collect Data
- ☐ Step 3: Analyze Data
- ☐ Step 4: Represent Data
- ☐ Step 5: Interpret the Results

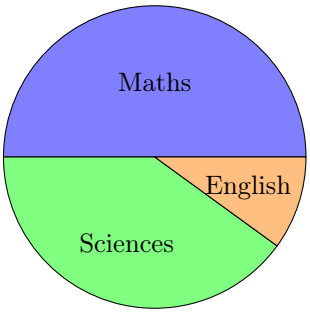
Answer: **Step 1: Pose the Question**. This sentence defines the specific problem that the investigation aims to answer.

MCQ 3: Read the statement: "We asked every student in the school to fill out a survey about their favorite subject."
Which step of the statistical process does this describe?

- ☐ Step 1: Pose the Question
- ☒ Step 2: Collect Data
- ☐ Step 3: Analyze Data
- ☐ Step 4: Represent Data
- ☐ Step 5: Interpret the Results

Answer: **Step 2: Collect Data**. This sentence describes the action of gathering information from the population.

MCQ 4: Read the statement: "We made a pie chart showing the proportion of students who chose each subject."



Which step of the statistical process does this describe?

- ☐ Step 1: Pose the Question
- ☐ Step 2: Collect Data
- ☐ Step 3: Analyze Data
- ☒ Step 4: Represent Data
- ☐ Step 5: Interpret the Results

Answer: **Step 4: Represent Data**. This statement describes creating a visual display (a pie chart) to show the data.

MCQ 5: Read the statement: "The relative frequency of students choosing 'Math' as their favorite subject is 32%."
Which step of the statistical process does this describe?

- ☐ Step 1: Pose the Question
- ☐ Step 2: Collect Data
- ☒ Step 3: Analyze Data
- ☐ Step 4: Represent Data
- ☐ Step 5: Interpret the Results

Answer: **Step 3: Analyze Data**. This statement describes the calculation of a descriptive statistic (relative frequency) to summarize a key feature of the data.

B STEP 1: POSING THE QUESTION

B.1 KEY CONCEPTS OF THE INQUIRY

B.1.1 IDENTIFYING POPULATIONS

MCQ 6: **Your Question:** "How many hours do **kids** spend playing outside each day?"
Which group is the correct **population** to survey?

- ☐ All adults in a city.
- ☒ All kids in a school.
- ☐ All dogs in a neighborhood.
- ☐ All teachers in a country.

Answer: The question is about **kids**.

- Adults, dogs, and teachers are not kids.
- "All kids in a school" is a group of kids. This is the correct population.

MCQ 7: Your Question: "How many **families** own a pet in **our town**?"
Which group is the correct **population** to survey?

☐ All kids in a playground.

☐ All birds in a forest.

☐ All workers in a factory.

☒ All families in our town.

Answer: The question is about **families in our town**.

- Kids, birds, and workers are not "all families in our town."
- "All families in our town" perfectly matches the group we want to study.

MCQ 8: Your Question: "How many books do **students** borrow from the **school library** each month?"
Which group is the correct **population** to survey?

☐ All librarians in a state.

☒ All students in the school.

☐ All books in a bookstore.

☐ All parents in a neighborhood.

Answer: The question asks about **students in the school**.

- Librarians, books, and parents are not the students we are studying.
- "All students in the school" is the correct group to ask about borrowing books from the school library.

MCQ 9: Your Question: "How tall are the **oak trees** in a **national park**?"
Which group is the correct **population** to study?

☒ All oak trees in the national park.

☐ All rivers in a country.

☐ All clouds in the sky.

☐ All rocks on a mountain.

Answer: The question specifically asks about **oak trees in a national park**.

- Rivers, clouds, and rocks are not oak trees.
- "All oak trees in the national park" is the exact group we need to measure.

B.1.2 CLASSIFYING DATA TYPES

MCQ 10: Data: Favorite subject (e.g., Math, Science, English).
Think: Is this a category or a number you can measure?

☐ Quantitative variable

☒ Qualitative variable

Answer: **Qualitative variable**. The answers are words that describe categories (Math, Science). You cannot measure "Math" with a number.

MCQ 11: Data: Number of siblings.
Think: Is this a category or a number you can count?

☒ Quantitative variable

☐ Qualitative variable

Answer: **Quantitative variable**. The answer is a number that you can count (e.g., 0, 1, 2).

MCQ 12: Data: Type of vehicle (e.g., car, bicycle, bus).
Think: Is this a category or a number you can measure?

☐ Quantitative variable

☒ Qualitative variable

Answer: **Qualitative variable**. The answers are words that describe different types. You cannot measure "car".

MCQ 13: Data: Height of students (in cm).
Think: Is this a category or a number you can measure?

☒ Quantitative variable

☐ Qualitative variable

Answer: **Quantitative variable**. The answer is a number that you get by measuring (e.g., 145 cm).

MCQ 14: Data: Level of education (e.g., high school, bachelor's, master's).
Think: Is this a category or a number you can measure?

☐ Quantitative variable

☒ Qualitative variable

Answer: **Qualitative variable**. The answers are labels for different stages of education. They are categories, not measurements.

MCQ 15: Data: Annual income (in dollars).
Think: Is this a category or a number you can count or measure?

☒ Quantitative variable

☐ Qualitative variable

Answer: **Quantitative variable**. The answer is a number representing an amount of money (e.g., $30,000).

B.2 SURVEY DESIGN AND QUESTION FORMULATION

B.2.1 DESIGNING QUESTIONS FOR DATA TYPES

Ex 16: Goal: Collect **quantitative (numerical)** data about music.
Write a survey question that will give you answers that are numbers.

Answer: To collect quantitative data, your question must ask for a number that can be counted or measured. It should start with "How many" or "How much."
A good question is: **How many hours do you listen to music each day?**

- **Why it works:** The answer will be a number (like "2 hours" or "0 hours").

Other good examples:

- "How many songs do you know by heart?"
- "How many musical instruments can you play?"

Ex 17: Goal: Collect **qualitative (categorical)** data about music.
Write a survey question that will give you answers that are words or categories.

Answer: To collect qualitative data, your question must ask for a word, a type, or a category. It often starts with "What is your favorite..." or "Which type..."
A good question is: **What is your favorite genre of music?**

- **Why it works:** The answers will be categories (like "Pop," "Rock," or "Classical").

Other good examples:

- "Who is your favorite singer?"
- "Which instrument has the best sound?"

Ex 18: Goal: Collect **qualitative (categorical)** data about food.
Write a survey question that will give you answers that are words or categories.

Answer: To get categorical data, your question should ask for a type, a preference, or a name.
A good question is: **What is your favorite vegetable?**

- **Why it works:** The answers are categories that you can group and count (like "Carrots," "Broccoli," "Corn").

Other good examples:

- "Which topping do you prefer on pizza?"
- "What is your favorite kind of breakfast cereal?"

Ex 19: Goal: Collect **quantitative (numerical)** data about food.
Write a survey question that will give you answers that are numbers.

Answer: To collect numerical data, your question must ask for an amount that can be counted.
A good question is: **How many times did you eat fruit yesterday?**

- **Why it works:** The answer will be a number (like "3 times" or "1 time").

Other good examples:

- "How many different vegetables are in your fridge right now?"
- "How many cookies do you eat in a week?"

B.2.2 EVALUATING SURVEY QUESTIONS

MCQ 20: **Scenario:** A cafeteria manager wants to survey students to improve the menu.
Proposed Question: "How often do you buy lunch at school? (a) Daily, (b) 2–3 times a week, (c) Rarely, (d) Never"
Classify this question.

☒ Structured.

☐ Unstructured.

Answer: The question is **structured** because it provides a fixed list of options. However, it is flawed because the option "(c) Rarely" is **ambiguous**. One person's definition of "rarely" might be once a week, while for another it might be once a month. The categories are not precisely defined.

MCQ 21: **Scenario:** A cafeteria manager wants to survey students to improve the menu.
Proposed Question: "Do you like the current lunches and why?"
Classify this question.

☐ Structured, and well-formulated.

☐ Structured, but it is a leading question.

☐ Unstructured, and well-formulated.

☒ Unstructured, but it is a "double-barreled" question.

Answer: The question is **unstructured** as it allows a free response. However, it is a poorly formulated **"double-barreled" question** because it asks two things at once: a "yes/no" question about liking the lunches, and an open-ended question for the reason. This makes the data difficult to analyze. A better approach would be to ask two separate questions.

MCQ 22: **Scenario:** A cafeteria manager wants to survey students.
Proposed Question: "Don't you think the lunches are too expensive and unhealthy?"
What is the primary flaw of this question?

☐ It is too personal.

☒ It is a leading and double-barreled question.

☐ The options are not exhaustive.

☐ It is ambiguous.

Answer: The question is **leading** (the phrase "Don't you think..." encourages agreement) and **double-barreled** (it asks about two different issues, price and health, at the same time). A student might think the food is healthy but expensive, and would not be able to answer accurately. This question is designed to produce biased results.

MCQ 23: **Scenario:** A cafeteria manager wants to survey students.
Proposed Question: "On a scale of 1 to 5, how satisfied are you with the variety, taste, and price of the lunches?"
This question is structured, but what is its primary flaw?

☒ It is a "triple-barreled" question.

☐ It is a leading question.

☐ The scale is not clearly defined.

☐ It is too personal.

Answer: The question is **"triple-barreled."** It asks the respondent to give a single rating for three separate concepts: variety, taste, and price. A student might be very satisfied with the taste (5/5) but very unsatisfied with the price (1/5). It is impossible to give a single, meaningful rating, making the data invalid. The solution is to ask three separate rating-scale questions.

B.2.3 CRITIQUING AND REWRITING SURVEY QUESTIONS

Ex 24: Scenario: A cafeteria manager wants to survey students about the price of lunches.
Proposed Question: "Don't you think the lunches are too expensive?"
Critique this question and propose a better structured (closed-ended) and unstructured (open-ended) version.

Answer: **Critique:** This is a **leading question**. The phrasing "Don't you think..." encourages the respondent to agree, which introduces bias and may not reflect their true opinion.
Improved Versions:

- **Structured Rewrite:** "How would you rate the price of the lunches?" with a balanced scale like {Very Inexpensive, Inexpensive, Fair Price, Expensive, Very Expensive}.
- **Unstructured Rewrite:** "What are your thoughts on the price of the lunches?"

Ex 25: Scenario: A cafeteria manager wants to measure student satisfaction.
Proposed Question: "How satisfied are you with the quality and the price of the lunches?"
Critique this question and rewrite it properly.

Answer: **Critique:** This is a **double-barreled question**. It asks for a single response about two separate concepts (quality and price). A student could be satisfied with the quality but unsatisfied with the price, making a single answer meaningless.
Improved Rewrite: The question must be split into two distinct questions.

- "On a scale of 1 to 5, how would you rate the **quality** of the lunches?"
- "On a scale of 1 to 5, how would you rate the **price** of the lunches?"

Ex 26: Scenario: A cafeteria manager wants feedback on a new menu.
Proposed Question: "How great is the new eco-friendly menu? Choose one: Excellent, Very good, Good."
Critique this question and rewrite it properly.

Answer: **Critique:** This question has multiple flaws:

- It uses **leading and subjective language** ("how great," "eco-friendly"), which creates bias.
- The response options are **unbalanced**, as they only allow for positive feedback. There is no way to express a neutral or negative opinion.
- The options are not **exhaustive** as there is no opt-out choice (e.g., "I haven't tried it").

Improved Rewrite: "How would you rate the new menu?" with a balanced, neutral scale like: {Very Poor, Poor, Average, Good, Very Good, I haven't tried it}.

Ex 27: Scenario: A cafeteria manager wants to understand student dissatisfaction.
Proposed Question: "Why do you hate the cafeteria food?"
Critique this question and rewrite it properly.

Answer: **Critique:** This is a highly **leading and presumptive question**. It assumes the respondent "hates" the food, which may not be true. This phrasing discourages neutral or positive responses and will lead to biased, purely negative feedback.
Improved Rewrite: An unstructured but neutral question would be far more effective: "What are your thoughts on the cafeteria food? Please provide any feedback you have."

C STEP 2: COLLECTING DATA

C.1 CENSUS VS. SAMPLING

C.1.1 CHOOSING THE RIGHT DATA COLLECTION METHOD

MCQ 28: Scenario: You need to elect the Grade 7 class representative.
Which method should be used to ensure a fair result?

☐ Survey

☒ Census

Answer: **Census**. For an election to be fair, every person in the population (the entire class) must have the opportunity to vote. Surveying only a sample would be undemocratic.

MCQ 29: Scenario: You want to estimate the average amount of soda Grade 7 students drink per week across the entire country.
Which method is more practical?

☒ Survey

☐ Census

Answer: **Survey**. The population (all Grade 7 students in the country) is extremely large. A census would be nearly impossible due to the high cost and time required. A survey of a representative sample is the only practical option.

MCQ 30: Scenario: A teacher wants to know the exact number of students in their own class who have a pet.
Which method should be used for an accurate count?

☐ Survey

☒ Census

Answer: **Census**. The population (one class) is small and easily accessible. A census is quick, easy, and will provide a completely accurate result, not an estimate.

MCQ 31: Scenario: Researchers want to estimate the average number of hours Grade 7 students sleep per night in a large city.
Which method is more practical?

☒ Survey

☐ Census

Answer: **Survey**. As with the country-wide example, the population of all Grade 7 students in a large city is too big to realistically survey every member. A well-chosen sample will provide a reliable estimate.

MCQ 32: Scenario: A large food company wants to find out the most popular new snack flavor among teenagers across the country before launching it.
Which method should the company use?

☒ Survey

☐ Census

Answer: **Survey**. A census of all teenagers in a country would be extremely expensive and time-consuming. Market research companies specialize in conducting surveys with carefully selected samples to estimate the preferences of the entire population.

C.2 SAMPLING METHODS AND POTENTIAL FOR ERROR

C.2.1 IDENTIFYING SAMPLING METHODS

MCQ 33: Scenario: A researcher wants to survey 100 students in a high school of 1000. The school population is divided by grade: 9th (250), 10th (300), 11th (250), and 12th (200). The researcher randomly selects 25 students from 9th grade, 30 from 10th, 25 from 11th, and 20 from 12th.
What type of sampling method is being used?

☐ Simple random sampling

☐ Systematic sampling

☐ Convenience sampling

☒ Stratified sampling

☐ Quota sampling

Answer: **Stratified sampling**.
The key features are:

- The population is divided into distinct subgroups (strata), in this case, by grade level.
- A random sample is drawn from *each* subgroup.
- The sample size from each stratum is proportional to the size of the stratum in the population (e.g., 10th grade is 30% of the school and makes up 30% of the sample).

MCQ 34: Scenario: A researcher wants to survey 100 students in a high school of 1000. The researcher assigns each student a unique number and uses a random number generator to select 100 students to survey.
What type of sampling method is being used?

☒ Simple random sampling

☐ Systematic sampling

☐ Convenience sampling

☐ Stratified sampling

☐ Quota sampling

Answer: **Simple random sampling**.
This is the most basic form of probability sampling. Every student has an equal and independent chance of being selected, and every possible group of 100 students is equally likely to be chosen.

MCQ 35: Scenario: A researcher wants to survey 100 students in a high school of 1000. The researcher obtains an alphabetized list of all students and selects every 10th student, starting from the 5th person on the list.
What type of sampling method is being used?

☐ Simple random sampling

☒ Systematic sampling

☐ Convenience sampling

☐ Stratified sampling

☐ Quota sampling

Answer: **Systematic sampling**.
The key feature is the selection of elements at a regular interval (every 10th student) from an ordered list. While it is a probability method, it can be biased if the ordering of the list has a hidden pattern.

MCQ 36: Scenario: A researcher wants to survey 100 students in a high school. The researcher stands in the cafeteria and surveys the first 100 students who agree to participate.
What type of sampling method is being used?

☐ Simple random sampling

☐ Systematic sampling

☒ Convenience sampling

☐ Stratified sampling

☐ Quota sampling

Answer: **Convenience sampling**.
This is a non-probability method. The sample is chosen based on ease of access (whoever is available and willing in the cafeteria) rather than random selection. It is highly likely to be biased, as students who eat in the cafeteria at that specific time may not be representative of the entire school population.

C.2.2 APPLYING SYSTEMATIC SAMPLING

Ex 37: **Scenario:** A factory produces 10,000 bottles per day and wants to test a 2% sample for quality. The operator uses a systematic sample, starting from the 23rd bottle.

1. Calculate the total size of the sample.

   200 bottles

2. List the numbers of the first three bottles to be sampled.

   23, 73, 123

Answer:

1. **Calculate Sample Size (n):** The sample is 2% of the population ($N = 10,000$).

$$n = 0.02 \times 10,000 = 200 \text{ bottles}$$

2. **Calculate Sampling Interval (k):**

$$k = \frac{N}{n} = \frac{10,000}{200} = 50$$

3. **Determine the Sampled Items:** The starting point is 23. We add the interval k repeatedly.
   - 1st bottle: 23
   - 2nd bottle: $23 + 50 = 73$
   - 3rd bottle: $73 + 50 = 123$

Ex 38: **Scenario:** A factory produces 80,000 chocolate bars per day and wants to test a 0.5% sample. The operator uses a systematic sample, starting from the 17th bar.

1. Calculate the total size of the sample.

$$\boxed{400} \text{ bars}$$

2. List the numbers of the first three bars to be sampled.

$$\boxed{17}, \boxed{217}, \boxed{417}$$

Answer:

1. **Calculate Sample Size (n):** $n = 0.005 \times 80,000 = 400$ bars.
2. **Calculate Sampling Interval (k):** $k = \frac{80,000}{400} = 200$.
3. **Determine the Sampled Items:** Start at 17 and add 200.
   - 1st: 17
   - 2nd: $17 + 200 = 217$
   - 3rd: $217 + 200 = 417$

Ex 39: **Scenario:** A publisher prints 5,000 books per day and wants to check a 4% sample. They use a systematic sample, starting from the 10th book.

1. Calculate the total size of the sample.

$$\boxed{200} \text{ books}$$

2. List the numbers of the first three books to be sampled.

$$\boxed{10}, \boxed{35}, \boxed{60}$$

Answer:

1. **Calculate Sample Size (n):** $n = 0.04 \times 5,000 = 200$ books.
2. **Calculate Sampling Interval (k):** $k = \frac{5,000}{200} = 25$.
3. **Determine the Sampled Items:** Start at 10 and add 25.
   - 1st: 10
   - 2nd: $10 + 25 = 35$
   - 3rd: $35 + 25 = 60$

Ex 40: **Scenario:** A car manufacturer assembles 500 cars per day and wants to safety test a 10% sample. They use a systematic sample, starting from the 5th car.

1. Calculate the total size of the sample.

$$\boxed{50} \text{ cars}$$

2. List the numbers of the first three cars to be sampled.

$$\boxed{5}, \boxed{15}, \boxed{25}$$

Answer:

1. **Calculate Sample Size (n):** $n = 0.10 \times 500 = 50$ cars.
2. **Calculate Sampling Interval (k):** $k = \frac{500}{50} = 10$.
3. **Determine the Sampled Items:** Start at 5 and add 10.
   - 1st: 5
   - 2nd: $5 + 10 = 15$
   - 3rd: $15 + 10 = 25$

C.2.3 EVALUATING SAMPLING METHODS

Ex 41: Scenario: A marketing manager wants to test whether a new product will appeal to the public. He surveys only the employees who developed the product.
Conclusion: "The product will be a success with all consumers."

1. Identify the sampling method.
2. Discuss the main problems with this sampling method and conclusion.
3. Suggest a more appropriate sampling method.

Answer:

1. **Method:** This is a **convenience sample**, a type of non-probability sampling.
2. **Problems:**
   - **Selection Bias:** The sample is not representative of "all consumers." The employees are personally and professionally invested in the product, which introduces a strong bias towards positive feedback.
   - **Invalid Generalization:** The conclusion is overgeneralized. Findings from a small, biased sample of internal employees cannot be used to make a valid claim about the entire consumer population.
3. **Better Method:** A more valid approach would be to use a **simple random sample** of the target consumer demographic. An even better method would be **stratified random sampling**, ensuring the sample includes proportional representation from different key demographic groups (e.g., age, gender, geographic location).

Ex 42: Scenario: A city planner surveys a large, diverse, and randomly selected group of residents, stratified by age, about a proposed new park.
Conclusion: "The majority of residents support the new park."

1. Identify the sampling method used.
2. Explain why this method is appropriate.
3. What potential sources of non-sampling error could still affect the conclusion?

Answer:

1. **Method: Stratified random sampling**. The population is divided into strata (age groups), and a random sample is taken from each.
2. **Appropriateness:** This is a strong probability sampling method. It reduces selection bias and ensures that different age groups are represented proportionally, making the sample more likely to be representative of the entire city's population.
3. **Potential Non-Sampling Errors:**
   - **Non-response bias:** If a significant number of selected residents refuse to participate, the results could be biased if the non-respondents share a common opinion.
   - **Measurement error:** The wording of the survey question could be leading (e.g., "Wouldn't you agree that a beautiful new park would benefit our city?"), influencing responses.

Ex 43: Scenario: A food festival averages 4,200 visitors. The manager wants to estimate the proportion of visitors who will spend more than $30. He decides to survey the first 50 people who enter.

1. Identify the sampling method.
2. Discuss the problems with this method.
3. Suggest a better sampling method.

Answer:

1. **Method: Convenience sampling**.
2. **Problems:**
   - **Selection Bias:** This method is not random. The first 50 visitors are likely not representative of all 4,200 attendees. They might be more enthusiastic, arrive at a specific time of day (e.g., before lunch), or have different spending habits than those who arrive later.
   - **Small Sample Size:** A sample of 50 is very small relative to a population of 4,200, leading to high potential for sampling error.
3. **Better Method: Systematic random sampling** would be far more effective. The manager could sample every k-th person. For example, to get a sample of 210, he would calculate $k = 4200/210 = 20$, and then survey every 20th person to enter throughout the entire day.

Ex 44: Scenario: A nutritionist tests a new diet's effectiveness by selecting clients from her own clinic. They all report significant weight loss.
Conclusion: "The diet works for everyone."

1. Identify the sampling method and key threats to the study's validity.
2. Critique the conclusion.

Answer:

1. **Method and Validity Threats:**
   - **Method:** This is a **convenience sample**.
   - **Selection Bias:** The sample is not representative. Clients of a nutrition clinic are likely more motivated and health-conscious than the general population.
   - **Confounding Variables:** The weight loss cannot be attributed solely to the diet. The nutritionist's professional support, the clients' motivation, and the psychological effect of being in a study (the Hawthorne effect) are all confounding factors.
   - **No Control Group:** There is no comparison group that did not follow the diet, so it is impossible to know if the weight loss would have happened anyway.
2. **Critique of Conclusion:** The conclusion "works for everyone" is an invalid generalization. The sample is small and heavily biased, and the study lacks a control group. Therefore, the results cannot be generalized to the wider population, and the cause of the weight loss cannot be confidently determined.

C.2.4 IDENTIFYING FLAWS IN STATISTICAL REASONING

MCQ 45: Scenario: A teacher wants to gauge how students feel about his lessons. He surveys only the 5 students who sit in the front row.
Conclusion Drawn: "All of my students enjoy my teaching."
Which statements are true about this conclusion? Check all that apply:

☐ The conclusion is reliable and accurate.

☒ The conclusion is likely flawed due to a small sample size.

☒ The conclusion is likely flawed due to selection bias.

Answer: The conclusion is flawed for two main reasons:

- **Small sample size:** 5 students is not enough to represent an entire class.
- **Selection bias:** Students in the front row are not a random sample. They might be more engaged or motivated, and their opinions may not reflect those of students in the back.

MCQ 46: Scenario: A company wants to test if a new product will be popular. The manager surveys only the employees who helped develop the product.
Conclusion Drawn: "The product will be a success with all consumers."
Which statements are true about this conclusion? Check all that apply:

☐ The manager's conclusion is reliable and accurate.

☐ The manager's conclusion could be flawed due to a small sample size.

☒ The manager's conclusion is flawed due to selection bias.

Answer: The conclusion is flawed primarily due to:

- **Selection bias:** The employees who created the product are not a representative sample of the general public. They have a personal investment in the product's success and are likely to have a much more positive opinion than an average consumer.

MCQ 47: Scenario: A city planner surveys a large, diverse, and randomly selected group of residents from all neighborhoods and age groups about a new park.
Conclusion Drawn: "The majority of residents support the new park."
Which statement best describes this conclusion?

☒ The planner's conclusion is likely reliable and accurate.

☐ The planner's conclusion is flawed due to a small sample size.

☐ The planner's conclusion is flawed due to selection bias.

Answer: The planner's conclusion is likely reliable.

- **Good sampling technique:** The sample is described as large, diverse, and random. This methodology is designed to minimize both sampling error and selection bias, making the results likely to be representative of the entire city's population.

MCQ 48: Scenario: A nutritionist tests a new diet on a small group of highly motivated clients from her clinic. They all report significant weight loss.
Conclusion Drawn: "This diet works for everyone."
Which statements are true about this conclusion? Check all that apply:

☐ The nutritionist's conclusion is reliable and accurate.

☒ The conclusion is likely flawed due to a small sample size.

☒ The conclusion is flawed due to selection bias.

Answer: The conclusion is flawed for two main reasons:

- **Small sample size:** A "small group" is not large enough to generalize to "everyone."
- **Selection bias:** Clients at a nutrition clinic are already highly motivated to improve their health. They do not represent the general population, which includes people with varying levels of motivation.

MCQ 49: Scenario: During WWII, engineers studied aircraft that returned from missions. They saw that the wings and fuselage were often full of bullet holes, but the engines and cockpit were rarely damaged.
Conclusion Drawn: "We should add more armor to the wings and fuselage."
Which statement best evaluates this conclusion?

☐ The conclusion is correct.

☒ The conclusion is flawed due to a type of selection bias known as survivorship bias.

☐ The engineers need a larger sample of planes to conclude anything.

Answer: The conclusion is flawed due to **survivorship bias**.

- The engineers were only looking at the planes that *survived* to return. They were missing data from the planes that were shot down.
- The fact that returning planes had damage to the wings and fuselage means that those areas could withstand damage. The reason they saw no damage to the engines or cockpit is that planes hit in those critical areas did not return.
- Therefore, they should add armor to the places that had no damage on the surviving planes: the engines and cockpit.

D STEP 3: ANALYZING DATA

D.1 FREQUENCIES

D.1.1 DISTINGUISHING A STATISTIC FROM DATA

MCQ 50: "Su averages 14.6 points per game."
Is this statement a statistic or a single piece of data?

☒ A statistic

☐ A single piece of data

Answer: **A statistic**. The average (14.6 points) is a single number that summarizes Su's performance over multiple games.

MCQ 51: "John's height is 180 cm."
Is this statement a statistic or a single piece of data?

☐ A statistic

☒ A single piece of data

Answer: **A single piece of data**. It is one measurement for one individual and does not summarize information from a larger group.

MCQ 52: "The average temperature in July is 25°C."
Is this statement a statistic or a single piece of data?

☒ A statistic

☐ A single piece of data

Answer: **A statistic**. The average temperature is a summary value calculated from the temperature data of many different days in July.

MCQ 53: "Emily's favorite color is blue."
Is this statement a statistic or a single piece of data?

☐ A statistic

☒ A single piece of data

Answer: **A single piece of data**. It is one person's preference and does not summarize the preferences of a group.

MCQ 54: "On average, students in the class scored 85%on the exam."
Is this statement a statistic or a single piece of data?

☒ A statistic

☐ A single piece of data

Answer: **A statistic**. The average score (85%) is a single number that summarizes the performance of the entire class.

MCQ 55: "The median income in the city is $50,000."
Is this statement a statistic or a single piece of data?

☒ A statistic

☐ A single piece of data

Answer: **A statistic**. The median is a summary value that describes the central point of the income data for the entire city's population.

D.1.2 CALCULATING RELATIVE FREQUENCIES

Ex 56: A class of 25 students was surveyed about their sex.
Complete the relative frequency column.

Sex	Frequency	Relative Frequency (%)
Girls	13	52 %
Boys	12	48 %
Total	25	100%

Answer: The calculations are:

Sex	Frequency	Relative Frequency (%)
Girls	13	$\frac{13}{25} \times 100\% = 52\%$
Boys	12	$\frac{12}{25} \times 100\% = 48\%$

Ex 57: 40 students were asked about their favorite pet.
Complete the relative frequency column (round to 1 decimal place).

Pet	Frequency	Relative Frequency (%)
Cats	18	45 %
Dogs	14	35 %
Hamsters	5	12.5 %
Fish	3	7.5 %
Total	40	100%

Answer: The calculations are:

Pet	Frequency	Relative Frequency (%)
Cats	18	$\frac{18}{40} \times 100\% = 45\%$
Dogs	14	$\frac{14}{40} \times 100\% = 35\%$
Hamsters	5	$\frac{5}{40} \times 100\% = 12.5\%$
Fish	3	$\frac{3}{40} \times 100\% = 7.5\%$

Ex 58: 80 students were surveyed about their primary mode of transportation.
Complete the relative frequency column (round to 1 decimal place).

Mode of Transportation	Frequency	Relative Frequency (%)
Bus	35	43.8 %
Bicycle	25	31.3 %
Walking	15	18.8 %
Car	5	6.3 %
Total	80	100%

Answer: The calculations are:

Mode of Transportation	Frequency	Relative Frequency (%)
Bus	35	$\frac{35}{80} \times 100\% \approx 43.8\%$
Bicycle	25	$\frac{25}{80} \times 100\% \approx 31.3\%$
Walking	15	$\frac{15}{80} \times 100\% \approx 18.8\%$
Car	5	$\frac{5}{80} \times 100\% \approx 6.3\%$

D.2 CENTRAL TENDENCY

D.2.1 FINDING THE MODE IN TABLES AND CHARTS

Ex 59: This frequency table shows the marks students received on a test.

Mark	Frequency
A	10
B	22
C	19
D	15
E	6

What is the mode of this dataset?

B

Answer: The mode is the category with the highest frequency.

1. The highest frequency in the table is **22**.
2. The category corresponding to this frequency is the mark **B**.

Therefore, the mode is B.

Ex 60: This frequency table shows the primary mode of transport for a group of people.

Mode of Transport	Frequency
Bus	18
Bicycle	12
Car	8
Walking	14
Train	6

What is the mode of this dataset?

Bus

Answer: The mode is the category with the highest frequency.

1. The highest frequency is **18**.
2. The category corresponding to this frequency is **Bus**.

Therefore, the mode is Bus.

Ex 61: This frequency table shows students' favorite fruits.

Fruit	Frequency
Apple	14
Banana	20
Orange	12
Grapes	10
Mango	16

What is the mode of this dataset?

Banana

Answer: The mode is the category with the highest frequency.

1. The highest frequency is **20**.
2. The category corresponding to this frequency is **Banana**.

Therefore, the mode is Banana.

Ex 62: A group of students was surveyed about their favorite music genre. The results are shown in this relative frequency table.

Music Genre	Relative Frequency (%)
Pop	45%
Rock	25%
Hip Hop	20%
Classical	10%

What is the mode of this dataset?

$$\boxed{\textbf{Pop}}$$

Answer: The mode is the category with the highest relative frequency.

1. The highest relative frequency in the table is **45**%.
2. The category corresponding to this frequency is **Pop**.

Therefore, the mode is Pop.

D.2.2 CALCULATING THE MEAN

Ex 63: Over the last 5 basketball games, a player scored the following points: $15, 20, 10, 2, 5$.
Find the mean number of points scored per game.

$\boxed{10.4}$ points

Answer: To find the mean, we follow two steps:

1. **Sum the values:** $15 + 20 + 10 + 2 + 5 = 52$.
2. **Divide by the number of values:** There are 5 games, so we divide by 5.

$$\text{Mean} = \frac{52}{5} = 10.4$$

The mean score is 10.4 points per game.

Ex 64: Over the last 5 days, a waiter earned the following tips: \$12, \$18, \$15, \$22, \$28.
Find the mean daily tip.

$\boxed{19}$ dollars

Answer: To find the mean, we follow two steps:

1. **Sum the values:** $12 + 18 + 15 + 22 + 28 = 95$.
2. **Divide by the number of values:** There are 5 days, so we divide by 5.

$$\text{Mean} = \frac{95}{5} = 19$$

The mean tip is \$19 per day.

Ex 65: Over the last 7 days, a student read the following number of pages each day: $30, 25, 35, 40, 20, 15, 45$.
Find the mean number of pages read per day.

$\boxed{30}$ pages

Answer: To find the mean, we follow two steps:

1. **Sum the values:** $30 + 25 + 35 + 40 + 20 + 15 + 45 = 210$.
2. **Divide by the number of values:** There are 7 days, so we divide by 7.

$$\text{Mean} = \frac{210}{7} = 30$$

The mean number of pages read is 30 per day.

Ex 66: For 6 days, the daily cost of lunch was: \$8, \$12, \$10, \$15, \$9, \$11.
Find the mean cost of a lunch.

\$$\boxed{10.83}$

Answer: To find the mean, we follow two steps:

1. **Sum the values:** $8 + 12 + 10 + 15 + 9 + 11 = 65$.
2. **Divide by the number of values:** There are 6 days, so we divide by 6.

$$\text{Mean} = \frac{65}{6} \approx 10.83$$

The mean cost is approximately \$10.83.

D.2.3 CALCULATING THE MEDIAN

Ex 67: A café tracked its hourly customers:

$$12, 8, 15, 10, 14, 11, 9$$

Calculate the median number of customers.

$\boxed{11}$ customers

Answer:

1. **Order the data:** $8, 9, 10, 11, 12, 14, 15$.
2. **Find the middle value:** There are 7 data points (an odd number), so the median is the single value in the middle, which is the 4th value.

$$8, 9, 10, \underline{11}, 12, 14, 15$$

The median is **11** customers.

Ex 68: A fitness group recorded their daily exercise minutes:

$$25, 40, 30, 45, 35$$

Find the median exercise time.

$\boxed{35}$ minutes

Answer:

1. **Order the data:** $25, 30, 35, 40, 45$.
2. **Find the middle value:** There are 5 data points (an odd number), so the median is the 3rd value.

$$25, 30, \underline{35}, 40, 45$$

The median is **35** minutes.

Ex 69: A family's savings (in $) over 6 months were:

$$120, 80, 150, 90, 200, 110$$

Determine the median savings.

$ $\boxed{115}$

Answer:

1. **Order the data:** $80, 90, 110, 120, 150, 200$.
2. **Find the middle value:** There are 6 data points (an even number), so the median is the average of the two middle values (the 3rd and 4th).

$$80, 90, \underline{110, 120}, 150, 200$$

$$\text{Median} = \frac{110 + 120}{2} = 115$$

The median savings is $**115**.

Ex 70: A group of students reported the number of books they read in a month:

$$1, 3, 4, 2, 5, 3, 6, 4, 3, 2$$

Determine the median of this dataset.

$\boxed{3}$ books

Answer:

1. **Order the data:** $1, 2, 2, 3, 3, 3, 4, 4, 5, 6$.
2. **Find the middle value:** There are 10 data points (an even number), so the median is the average of the two middle values (the 5th and 6th).

$$1, 2, 2, 3, \underline{3, 3}, 4, 4, 5, 6$$

$$\text{Median} = \frac{3 + 3}{2} = 3$$

The median is **3** books.

D.3 SPREAD

D.3.1 CALCULATING THE RANGE

Ex 71: The math marks (out of 20) for a group of students are:

$$4, 12, 9, 7, 11, 15, 8, 6, 14$$

Find the range of the marks.

$\boxed{11}$

Answer:

1. **Identify the minimum value:** The lowest mark is 4.
2. **Identify the maximum value:** The highest mark is 15.
3. **Calculate the range:** Range = Maximum - Minimum = $15 - 4 = \mathbf{11}$.

Ex 72: The average monthly temperatures (in °C) in Montréal over a year are:

$$-10, -7, 0, 7, 14, 19, 22, 21, 16, 9, 2, -5$$

Find the range of the temperatures.

$\boxed{32}$

Answer:

1. **Identify the minimum value:** The lowest temperature is -10°C.
2. **Identify the maximum value:** The highest temperature is 22°C.
3. **Calculate the range:** Range = Maximum - Minimum = $22 - (-10) = 22 + 10 = \mathbf{32}$.

The range in temperature is 32°C.

Ex 73: The speeds (in km/h) recorded on a highway are:

$$88.4, 91.0, 95.7, 102.3, 89.6, 100.0, 97.5, 92.1, 94.3, 90.8, 93.2, 96.6$$

Find the range of the speeds.

$\boxed{13.9}$

Answer:

1. **Identify the minimum value:** The lowest speed is 88.4 km/h.
2. **Identify the maximum value:** The highest speed is 102.3 km/h.
3. **Calculate the range:** Range = Maximum - Minimum = $102.3 - 88.4 = \mathbf{13.9}$.

The range of the speeds is 13.9 km/h.

Ex 74: The weights (in kg) of 10 packages are:

$$4.2, 3.8, 5.5, 6.1, 4.9, 3.6, 4.4, 5.2, 6.7, 3.9$$

Find the range of the weights.

$\boxed{3.1}$

Answer:

1. **Identify the minimum value:** The lowest weight is 3.6 kg.
2. **Identify the maximum value:** The highest weight is 6.7 kg.
3. **Calculate the range:** Range = Maximum - Minimum = $6.7 - 3.6 = \mathbf{3.1}$.

The range of the weights is 3.1 kg.

D.3.2 CALCULATING THE INTERQUARTILE RANGE

Ex 75: The average monthly temperatures (in °C) in Montréal over a year are:

$$-10, -7, 0, 7, 14, 19, 22, 21, 16, 9, 2, -5$$

Find the interquartile range of the temperatures.

20

Answer:

1. **Order the data:** $-10, -7, -5, 0, 2, 7, 9, 14, 16, 19, 21, 22$.
2. **Find the median (Q2):** There are 12 values (even). The median is the average of the 6th and 7th values.
$$-10, -7, -5, 0, 2, \underline{7, 9}, 14, 16, 19, 21, 22 \quad \rightarrow \quad Q_2 = \frac{7+9}{2} = 8$$
3. **Find the lower quartile (Q1):** Find the median of the lower half: $(-10, -7, -5, 0, 2, 7)$.
$$-10, -7, \underline{-5, 0}, 2, 7 \quad \rightarrow \quad Q_1 = \frac{-5+0}{2} = \textbf{-2.5}$$
4. **Find the upper quartile (Q3):** Find the median of the upper half: $(9, 14, 16, 19, 21, 22)$.
$$9, 14, \underline{16, 19}, 21, 22 \quad \rightarrow \quad Q_3 = \frac{16+19}{2} = \textbf{17.5}$$
5. **Calculate the IQR:**
$$\text{IQR} = Q_3 - Q_1 = 17.5 - (-2.5) = 17.5 + 2.5 = \textbf{20}$$

Ex 76: The speeds (in km/h) recorded for 11 cars are:

$$88, 95, 102, 91, 87, 98, 105, 93, 89, 100, 92$$

Find the interquartile range of the speeds.

11

Answer:

1. **Order the data:** $87, 88, 89, 91, 92, 93, 95, 98, 100, 102, 105$.
2. **Find the median (Q2):** There are 11 values (odd). The median is the 6th value.
$$87, 88, 89, 91, 92, \underline{93}, 95, 98, 100, 102, 105 \quad \rightarrow \quad Q_2 = 93$$
3. **Find the lower quartile (Q1):** Find the median of the lower half: $(87, 88, 89, 91, 92)$. The middle value is the 3rd one.
$$87, 88, \underline{89}, 91, 92 \quad \rightarrow \quad Q_1 = \textbf{89}$$
4. **Find the upper quartile (Q3):** Find the median of the upper half: $(95, 98, 100, 102, 105)$. The middle value is the 3rd one.
$$95, 98, \underline{100}, 102, 105 \quad \rightarrow \quad Q_3 = \textbf{100}$$
5. **Calculate the IQR:**
$$\text{IQR} = Q_3 - Q_1 = 100 - 89 = \textbf{11}$$

Ex 77: The weights (in kg) of 10 packages are:

$$4.2, 3.5, 6.1, 5.0, 4.8, 3.9, 6.7, 5.5, 4.4, 5.2$$

Find the interquartile range of the weights.

1.3

Answer:

1. **Order the data:** $3.5, 3.9, 4.2, 4.4, 4.8, 5.0, 5.2, 5.5, 6.1, 6.7$.
2. **Find the median (Q2):** There are 10 values (even). The median is the average of the 5th and 6th values.
$$3.5, 3.9, 4.2, 4.4, \underline{4.8, 5.0}, 5.2, 5.5, 6.1, 6.7 \quad \rightarrow \quad Q_2 = \frac{4.8+5.0}{2} =$$
3. **Find the lower quartile (Q1):** Find the median of the lower half: $(3.5, 3.9, 4.2, 4.4, 4.8)$. The middle value is the 3rd one.
$$3.5, 3.9, \underline{4.2}, 4.4, 4.8 \quad \rightarrow \quad Q_1 = \textbf{4.2}$$
4. **Find the upper quartile (Q3):** Find the median of the upper half: $(5.0, 5.2, 5.5, 6.1, 6.7)$. The middle value is the 3rd one.
$$5.0, 5.2, \underline{5.5}, 6.1, 6.7 \quad \rightarrow \quad Q_3 = \textbf{5.5}$$
5. **Calculate the IQR:**
$$\text{IQR} = Q_3 - Q_1 = 5.5 - 4.2 = \textbf{1.3}$$

Ex 78: The marks (out of 20) for 9 students are:

$$1, 19, 10, 2, 18, 11, 5, 15, 10$$

Find the interquartile range of the marks.

13

Answer:

1. **Order the data:** $1, 2, 5, 10, 10, 11, 15, 18, 19$.
2. **Find the median (Q2):** There are 9 values (odd). The median is the 5th value.
$$1, 2, 5, 10, \underline{10}, 11, 15, 18, 19 \quad \rightarrow \quad Q_2 = 10$$
3. **Find the lower quartile (Q1):** Find the median of the lower half: $(1, 2, 5, 10)$.
$$1, \underline{2, 5}, 10 \quad \rightarrow \quad Q_1 = \frac{2+5}{2} = \textbf{3.5}$$
4. **Find the upper quartile (Q3):** Find the median of the upper half: $(11, 15, 18, 19)$.
$$11, \underline{15, 18}, 19 \quad \rightarrow \quad Q_3 = \frac{15+18}{2} = \textbf{16.5}$$
5. **Calculate the IQR:**
$$\text{IQR} = Q_3 - Q_1 = 16.5 - 3.5 = \textbf{13}$$

D.3.3 CALCULATING STANDARD DEVIATION

Ex 79: Over the last 5 basketball games, a player scored: $15, 20, 10, 2, 5$.**Find the standard deviation of the scores, rounded to one decimal place.**

$$\sigma \approx \boxed{6.5} \text{ points}$$

Answer:

1. **Find the mean ($\bar{x}$):**
$$\bar{x} = \frac{15 + 20 + 10 + 2 + 5}{5} = \frac{52}{5} = 10.4$$
2. **Calculate the squared differences from the mean:** $(15 - 10.4)^2 = 21.16$, $(20 - 10.4)^2 = 92.16$, $(10 - 10.4)^2 = 0.16$, $(2 - 10.4)^2 = 70.56$, $(5 - 10.4)^2 = 29.16$.
3. **Find the variance (the average of the squared differences):**
$$\text{Variance} = \frac{21.16 + 92.16 + 0.16 + 70.56 + 29.16}{5} = \frac{213.2}{5} = 42.64$$
4. **Take the square root to find the standard deviation (σ):**
$$\sigma = \sqrt{42.64} \approx 6.5$$

The standard deviation is approximately **6.5 points**. This means the player's scores are, on average, about 6.5 points away from their mean score.

Ex 80: A student's marks on their last 5 exams were: $78, 85, 62, 90, 75$.**Find the standard deviation of the marks, rounded to two decimal places.**

$$\sigma \approx \boxed{9.57} \text{ marks}$$

Answer:

1. **Find the mean ($\bar{x}$):**
$$\bar{x} = \frac{78 + 85 + 62 + 90 + 75}{5} = \frac{390}{5} = 78$$
2. **Calculate the squared differences:** $(78-78)^2 = 0$, $(85-78)^2 = 49$, $(62-78)^2 = 256$, $(90-78)^2 = 144$, $(75-78)^2 = 9$.
3. **Find the variance:**
$$\text{Variance} = \frac{0 + 49 + 256 + 144 + 9}{5} = \frac{458}{5} = 91.6$$
4. **Find the standard deviation (σ):**
$$\sigma = \sqrt{91.6} \approx 9.57$$

The standard deviation is approximately **9.57 marks**.

Ex 81: The daily temperatures (°C) in a city over 5 days were: $22, 25, 19, 30, 24$.**Find the standard deviation, rounded to two decimal places.**

$$\sigma \approx \boxed{3.63} \text{ °C}$$

Answer:

1. **Find the mean ($\bar{x}$):**
$$\bar{x} = \frac{22 + 25 + 19 + 30 + 24}{5} = \frac{120}{5} = 24$$
2. **Calculate the squared differences:** $(22-24)^2 = 4$, $(25-24)^2 = 1$, $(19 - 24)^2 = 25$, $(30 - 24)^2 = 36$, $(24 - 24)^2 = 0$.
3. **Find the variance:**
$$\text{Variance} = \frac{4 + 1 + 25 + 36 + 0}{5} = \frac{66}{5} = 13.2$$
4. **Find the standard deviation (σ):**
$$\sigma = \sqrt{13.2} \approx 3.63$$

The standard deviation is approximately **3.63°C**.

Ex 82: A small business recorded these weekly sales over 5 weeks: $\$1500, \$2000, \$1800, \$2200, \$1700$.**Find the standard deviation, rounded to the nearest dollar.**

$$\sigma \approx \$ \boxed{242}$$

Answer:

1. **Find the mean ($\bar{x}$):**
$$\bar{x} = \frac{1500 + 2000 + 1800 + 2200 + 1700}{5} = \frac{9200}{5} = 1840$$
2. **Calculate the squared differences:** $(1500 - 1840)^2 = 115600$, $(2000 - 1840)^2 = 25600$, $(1800 - 1840)^2 = 1600$, $(2200 - 1840)^2 = 129600$, $(1700 - 1840)^2 = 19600$.
3. **Find the variance:**
$$\text{Variance} = \frac{115600 + 25600 + 1600 + 129600 + 19600}{5} = \frac{292000}{5}$$
4. **Find the standard deviation (σ):**
$$\sigma = \sqrt{58400} \approx 241.66$$

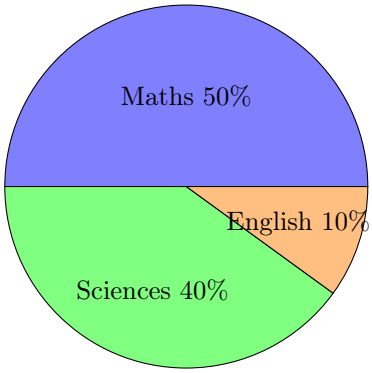
Rounded to the nearest dollar, the standard deviation is \$ **242**. This indicates that the weekly sales typically fluctuate by about \$242 from the average.

E STEP 4: REPRESENTING DATA

E.1 VISUALIZING FREQUENCIES

E.1.1 INTERPRETING CHARTS AND GRAPHS

Ex 83: 30 students were asked to name their favorite school subject. The results are displayed in this graph.

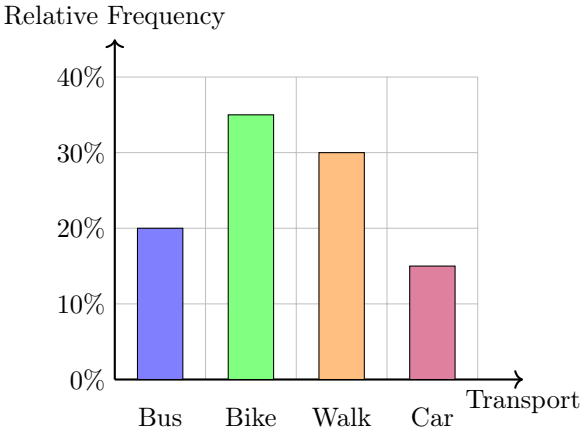


1. What type of graph is this? **Pie chart**
2. Which was the most popular subject? **Maths**
3. What percentage of students chose Sciences? 40 %
4. What percentage chose either Maths or Sciences? 90 %

Answer:

1. This is a **pie chart** because it shows data as slices of a circle.
2. The most popular subject is **Maths** because it has the largest slice (50%).
3. The percentage for Sciences is stated in its slice: **40**%.
4. To find the total for Maths or Sciences, we add their percentages: $50\% + 40\% = \mathbf{90\%}$.

Ex 84: 200 students were asked how they travel to school. The results are shown in the graph.

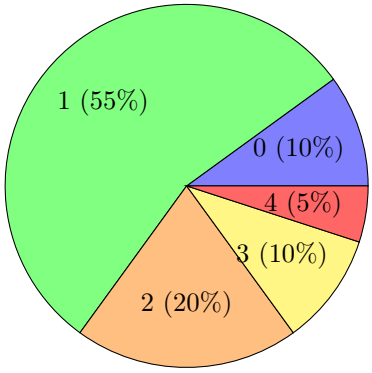


1. What type of graph is this? **Bar chart**
2. What is the most common mode of transport? **Bike**
3. What percentage of students travel by bike? 35 %
4. What percentage travel by either bus or bike? 55 %

Answer:

1. This is a **bar chart** because it uses rectangular bars to show frequencies.
2. The most common mode is **Bike** because it has the tallest bar.
3. The bar for "Bike" reaches the line for **35**%.
4. To find the total for "Bus or Bike," we add their percentages: $20\% + 35\% = \mathbf{55\%}$.

Ex 85: This pie chart shows the number of siblings for 30 students.

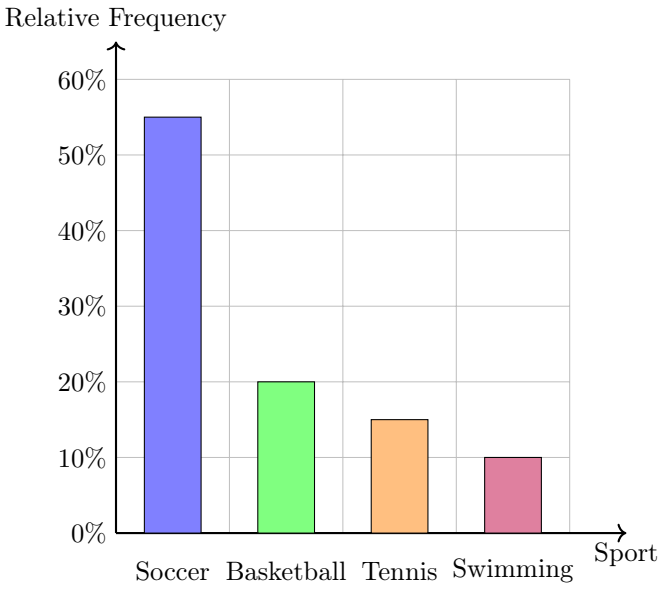


1. What is the mode for the number of siblings? **1 Sibling**
2. What percentage of students have 2 siblings? 20 %
3. What percentage of students have at least 1 sibling? 90 %

Answer:

1. The mode is **1 Sibling** because it has the largest slice (55%).
2. The percentage for "2 Siblings" is written in its slice: **20**%.
3. "At least 1 sibling" means everyone except those with 0 siblings. We can calculate this in two ways:
   - Add the other categories: $55\% + 20\% + 10\% + 5\% = \mathbf{90\%}$.
   - Subtract the "0 Siblings" group from the total: $100\% - 10\% = \mathbf{90\%}$.

Ex 86: A survey asked students to name their favorite sport. The results are in the graph.



1. Which was the most popular sport? **Soccer**
2. What percentage of students chose Basketball? 20 %
3. What percentage chose either Soccer or Basketball? 75 %

Answer:

1. The most popular sport is **Soccer**, because its bar is the tallest (55%).

2. The bar for Basketball reaches the line for **20**%.
3. To find the total for "Soccer or Basketball," we add their percentages: $55\% + 20\% = \mathbf{75\%}$.

E.1.2 CONSTRUCTING STATISTICAL GRAPHS

Ex 87: **Goal:** Represent categorical data as a pie chart.
A survey asked 20 students to choose their favorite season. The results are in the table below.

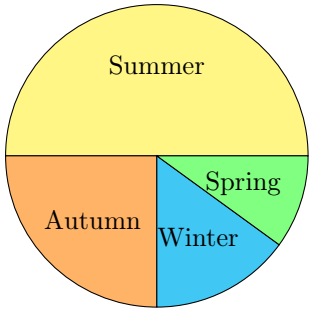
Season	Frequency
Summer	10
Autumn	5
Winter	3
Spring	2
Total	**20**

First, calculate the angle for each season's slice. Then, use a protractor to draw and label a pie chart representing this data.

Answer: To draw a pie chart, we first need to calculate the angle for each slice. The total number of students is 20.

- **Step 1: Calculate the Angles**
  - **Summer:** $\frac{10}{20} \times 360° = 0.5 \times 360° = 180°$
  - **Autumn:** $\frac{5}{20} \times 360° = 0.25 \times 360° = 90°$
  - **Winter:** $\frac{3}{20} \times 360° = 0.15 \times 360° = 54°$
  - **Spring:** $\frac{2}{20} \times 360° = 0.1 \times 360° = 36°$

  (Check: $180° + 90° + 54° + 36° = 360°$)
- **Step 2: Draw the Pie Chart**



Ex 88: Goal: Represent categorical data as a bar chart.
A survey asked 25 students to choose their favorite type of movie. The results are in the table below.

Movie Genre	Frequency
Comedy	8
Action	7
Fantasy	6
Sci-Fi	4
Total	**25**

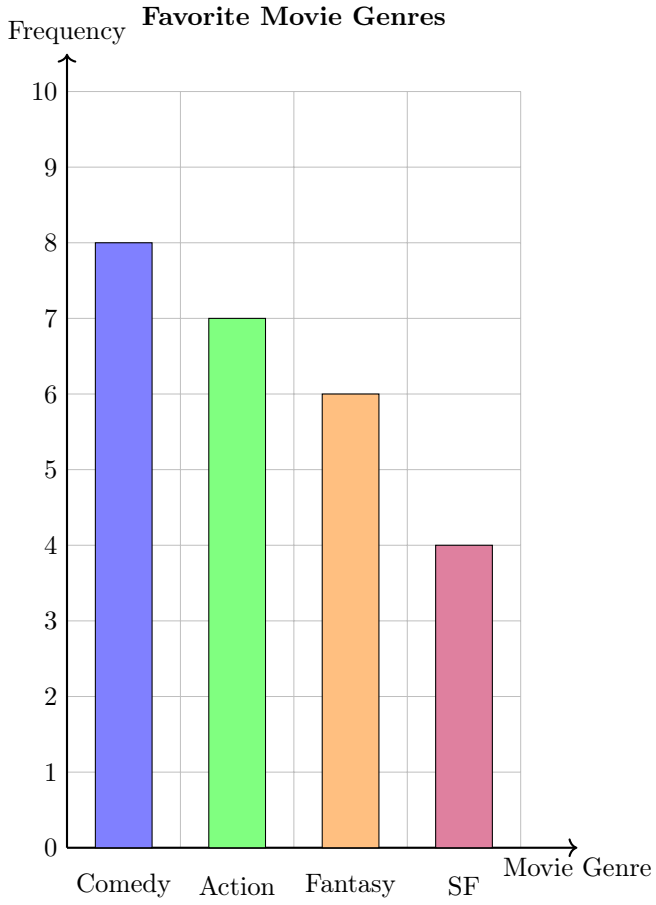
Draw and label a bar chart to represent this data. Make sure to include a title, labels for both axes, and an appropriate scale for the frequency axis.

Answer: To draw a bar chart, we need to set up our axes and draw a bar for each category with the correct height.

- **Step 1: Set up the Axes and Scale**
  - The horizontal axis (x-axis) will show the categories: Comedy, Action, Fantasy, Sci-Fi.
  - The vertical axis (y-axis) will show the frequency. The highest frequency is 8, so our scale should go up to at least 8. A scale from 0 to 10 is a good choice.
- **Step 2: Draw and Label the Bars**
  - Draw a bar for "Comedy" up to the number 8.
  - Draw a bar for "Action" up to the number 7.
  - Draw a bar for "Fantasy" up to the number 6.
  - Draw a bar for "Sci-Fi" up to the number 4.

  Don't forget to add a title for the chart!

Completed Bar Chart:

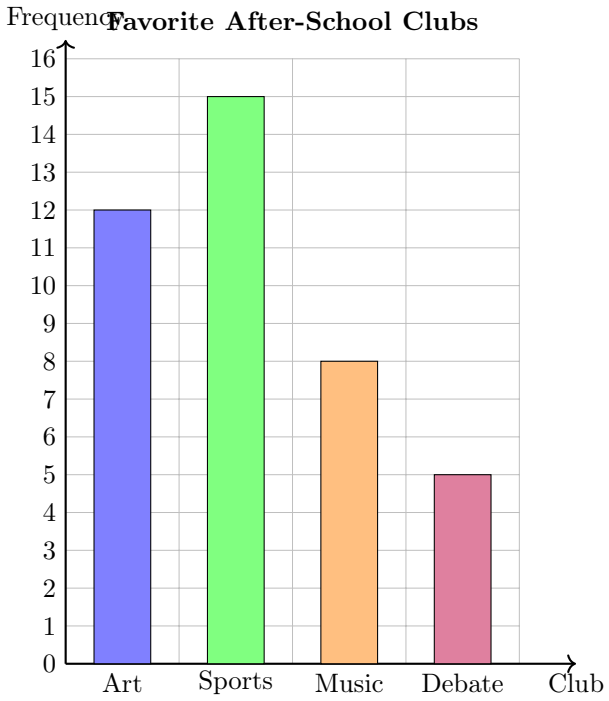


Ex 89: **Goal:** Represent categorical data as a bar chart.
The frequency table shows the results of a survey on students' favorite after-school clubs.

Club	Frequency
Art	12
Sports	15
Music	8
Debate	5
Total	**40**

Draw and label a bar chart to represent this data. Ensure all components of the chart are included.

Answer: **Completed Bar Chart:**



Ex 90: **Goal:** Represent categorical data as a pie chart. In a survey, 50 people were asked for their favorite hot drink.

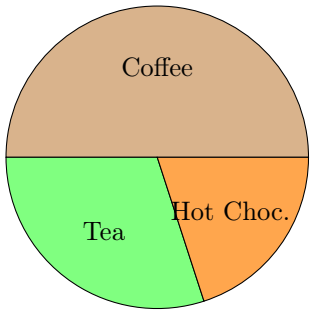
Drink	Frequency
Coffee	25
Tea	15
Hot Chocolate	10
Total	**50**

Calculate the angle for each drink's slice, then draw and label the pie chart.

Answer: **Step 1: Calculate the Angles**

- **Coffee:** $\frac{25}{50} \times 360° = 180°$
- **Tea:** $\frac{15}{50} \times 360° = 108°$
- **Hot Chocolate:** $\frac{10}{50} \times 360° = 72°$

Step 2: Draw the Pie Chart



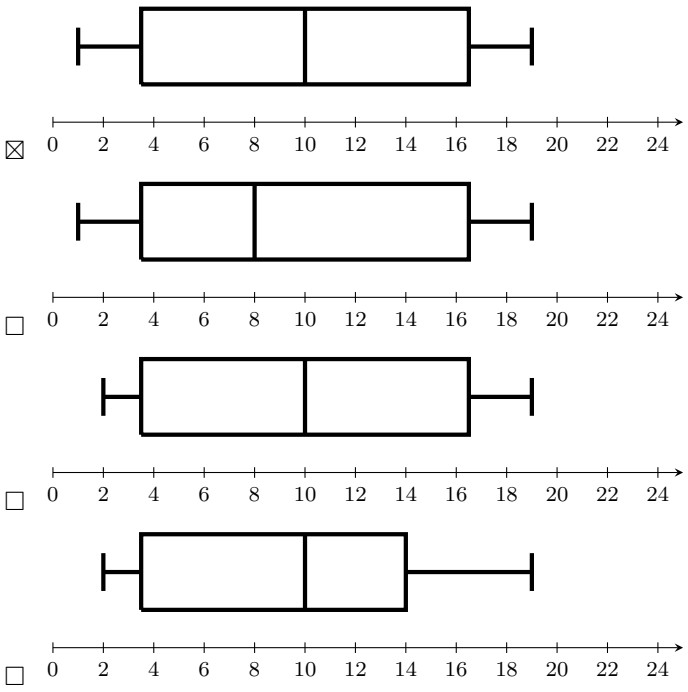
E.2 VISUALIZING CENTRAL TENDENCY AND DISPERSION

E.2.1 MATCHING A FIVE-NUMBER SUMMARY TO A BOX PLOT

MCQ 91: The five-number summary for a dataset is:

Min $= 1$, $Q_1 = 3.5$, Median $= 10$, $Q_3 = 16.5$, Max $= 19$

Select the correct box plot that represents this summary.



Answer: To find the correct box plot, we check each part of the five-number summary against the options.

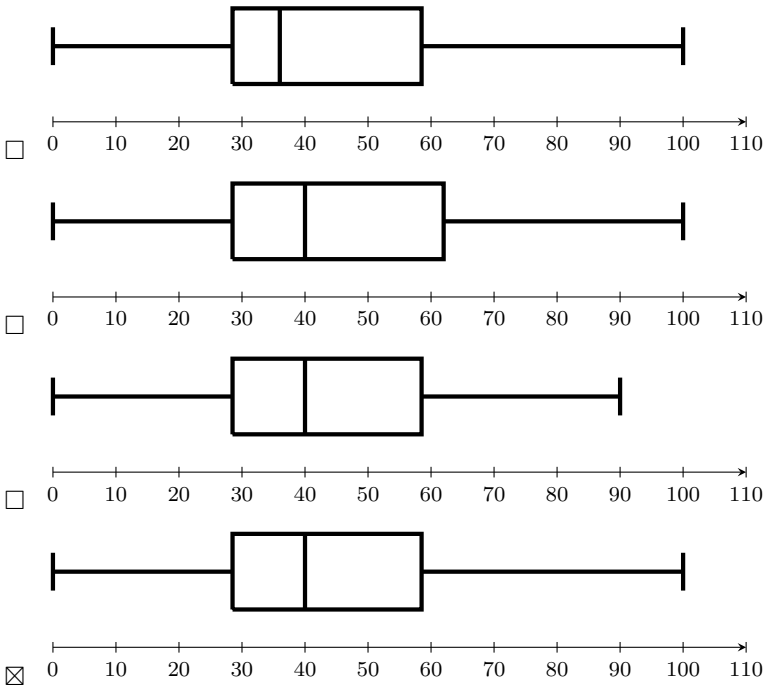
- **Minimum $= 1$**: The left whisker must start at 1. This eliminates options 3 and 4.
- **Maximum $= 19$**: The right whisker must end at 19. Both options 1 and 2 are correct.
- **Median $= 10$**: The line inside the box must be at 10. This eliminates option 2, where the median is 8.
- **Q1 $= 3.5$** and **Q3 $= 16.5$**: The box must span from 3.5 to 16.5. This is true for option 1.

Only the first option matches all five summary points.

MCQ 92: The five-number summary for a dataset is:

Min $= 0$, $Q_1 = 28.5$, Median $= 40$, $Q_3 = 58.5$, Max $= 100$

Select the correct box plot that represents this summary.



Answer: Let's check each value against the options:

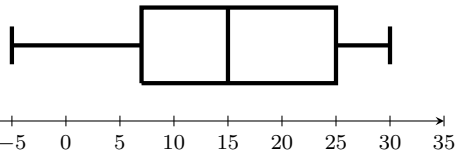
- The **Median** must be 40. This eliminates option 1 (median is 36).
- The **Maximum** must be 100. This eliminates option 3 (maximum is 90).
- **Q3** must be 58.5. This eliminates option 2 (Q3 is 62).
- Option 4 has all five correct values: Min=0, Q1=28.5, Median=40, Q3=58.5, and Max=100.

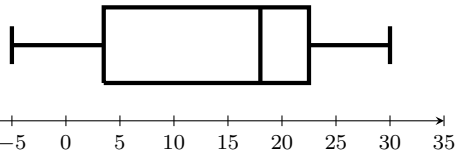
The correct choice is the last one.

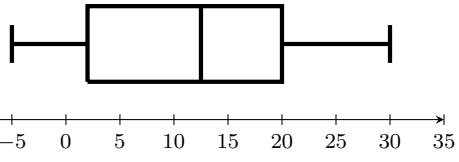
MCQ 93: The five-number summary for average monthly temperatures (°C) is:

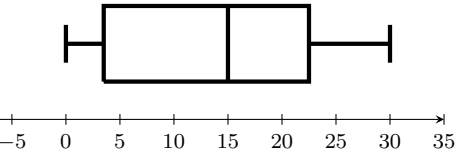
$$\text{Min} = -5, \quad Q_1 = 2, \quad \text{Median} = 12.5, \quad Q_3 = 20, \quad \text{Max} = 30$$

Select the correct box plot that represents this summary.

☐ 

☐ 

☒ 

☐ 

Answer: Let's check each value against the options:

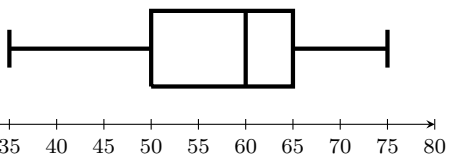
- The **Minimum** must be -5. This eliminates option 4 (minimum is 0).
- **Q1** must be 2. This eliminates options 1 (Q1=7) and 2 (Q1=3.5).
- Option 3 has all five correct values: Min=-5, Q1=2, Median=12.5, Q3=20, and Max=30.

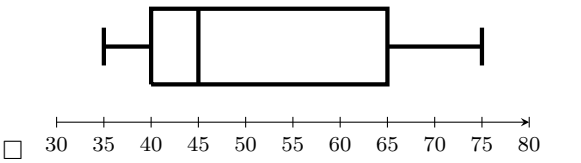
The correct choice is the third one.

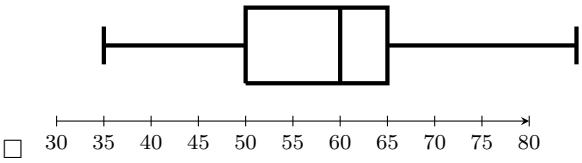
MCQ 94: The five-number summary for an air quality index (AQI) is:

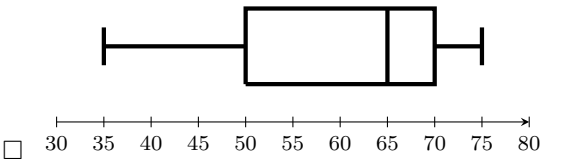
$$\text{Min} = 35, \quad Q_1 = 50, \quad \text{Median} = 60, \quad Q_3 = 65, \quad \text{Max} = 75$$

Select the correct box plot that represents this summary.

☒ 

☐ 

☐ 

☐ 

Answer: Let's check each value against the options:

- The **Median** must be 60. This eliminates option 2 (median=45) and option 4 (median=65).
- The **Maximum** must be 75. This eliminates option 3 (maximum=85).
- Option 1 has all five correct values: Min=35, Q1=50, Median=60, Q3=65, and Max=75.

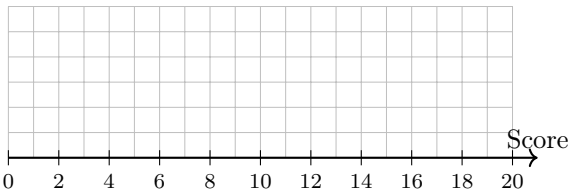
The correct choice is the first one.

E.2.2 CONSTRUCTING BOX PLOTS

Ex 95: The five-number summary for a basketball player's scores in a season is:

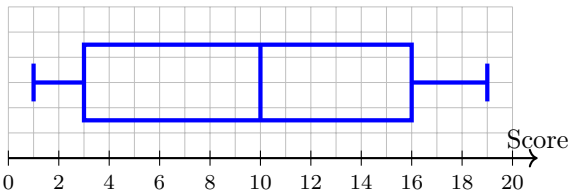
$$\text{Minimum} = 1, \quad Q_1 = 3, \quad \text{Median} = 10, \quad Q_3 = 16, \quad \text{Maximum} =$$

Plot the corresponding box plot on the grid below.



Answer: To draw the box plot:

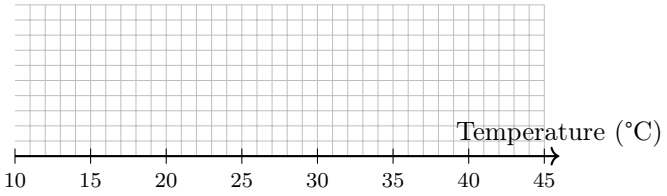
1. Draw a box from **Q1 (3)** to **Q3 (16)**.
2. Draw a vertical line inside the box at the **Median (10)**.
3. Draw a whisker from the left of the box to the **Minimum (1)**.
4. Draw a whisker from the right of the box to the **Maximum (19)**.



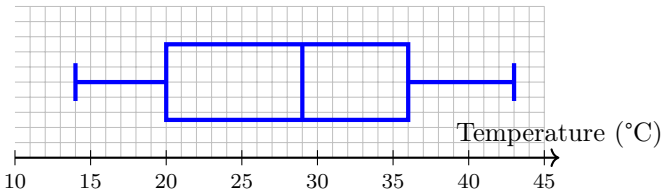
Ex 96: The five-number summary for yearly temperatures (°C) in Dubai is:

$$\text{Minimum} = 14, \quad Q_1 = 20, \quad \text{Median} = 29, \quad Q_3 = 36, \quad \text{Maximum}$$

Plot the corresponding box plot on the grid below.



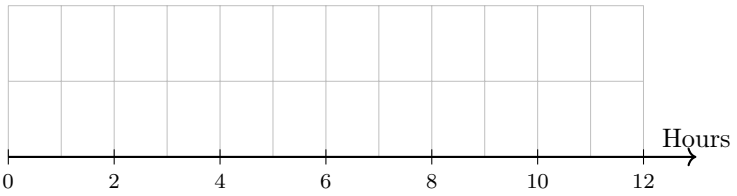
Answer:



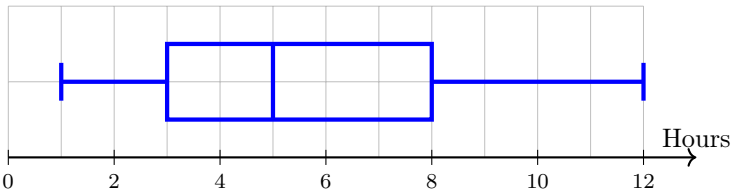
Ex 97: The five-number summary for hours spent reading per week is:

Minimum $= 1$, $Q_1 = 3$, Median $= 5$, $Q_3 = 8$, Maximum $= 12$

Plot the corresponding box plot on the grid below.

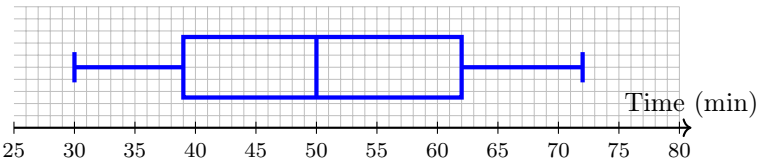


Answer:



E.2.3 INTERPRETING BOX PLOTS

Ex 98: This box plot shows the number of minutes passengers spent waiting in an airport departure lounge.



Answer the following questions based on the box plot:

1. What is the median waiting time? $\boxed{50}$ minutes
2. What is the range of the waiting times? $\boxed{42}$ minutes
3. What is the interquartile range (IQR)? $\boxed{23}$ minutes
4. 75% of passengers waited longer than what amount of time? $\boxed{39}$ minutes

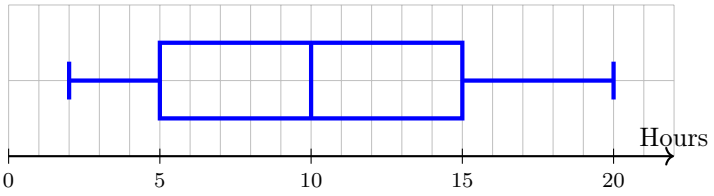
Answer: First, read the five-number summary from the plot:

- Minimum $= 30$
- Q1 $= 39$
- Median (Q2) $= 50$
- Q3 $= 62$
- Maximum $= 72$

Now, we can answer the questions:

1. The median is the line inside the box, which is at **50 minutes**.
2. The range is Maximum - Minimum $= 72 - 30 =$ **42** minutes.
3. The IQR is Q3 - Q1 $= 62 - 39 =$ **23** minutes.
4. Q1 marks the point where 25% of the data is lower and 75% is higher. Therefore, 75% of passengers waited longer than **39 minutes**.

Ex 99: This box plot shows the hours students spent on an online learning platform in one week.



What is the interquartile range (IQR) of the hours spent on the platform?

$\boxed{10}$ hours

Answer: First, read the quartiles from the plot:

- The lower quartile (Q1), the left edge of the box, is at 5 hours.
- The upper quartile (Q3), the right edge of the box, is at 15 hours.

The interquartile range is the difference between these two values:

$$\text{IQR} = Q_3 - Q_1 = 15 - 5 = \mathbf{10} \quad \textbf{hours}$$

F STEP 5: INTERPRETING DATA

F.1 DRAWING CONCLUSIONS AND EVALUATING CLAIMS

F.1.1 INTERPRETING RELATIVE FREQUENCY TABLES

MCQ 100: This table shows the relative frequency of children's favorite beverages.

Beverage	Relative Frequency (%)
Water	55%
Juice	30%
Soda	10%
Milk	5%

Based on the table, check all the true statements:

☒ Water is the most popular beverage.

☒ Milk is the least popular beverage.

☐ Soda is more popular than Juice.

☐ Milk is the most popular beverage.

☒ Water accounts for more than half of the choices.

☒ Juice and Soda combined are less popular than Water alone.

Answer:

- **True**. Water has the highest percentage (55%).
- **True**. Milk has the lowest percentage (5%).
- **False**. Soda (10%) is less popular than Juice (30%).
- **False**. Milk is the least popular, not the most.
- **True**. More than half means more than 50%, and Water is 55%.
- **True**. Juice (30%) + Soda (10%) = 40%, which is less than Water's 55%.

MCQ 101: This table shows how students get to school.

Transportation	Relative Frequency
Bus	15%
Walking	40%
Bicycle	30%
Car	15%

Based on the table, check all the true statements:

☒ Walking is the most popular way to get to school.

☒ Car and Bus are equally popular.

☒ Bicycle is more popular than Bus.

☐ More students take the Bus than walk.

☒ Bicycle and Walking combined account for more than half the students.

☒ Bus is one of the least popular ways to get to school.

Answer:

- **True**. Walking has the highest percentage (40%).
- **True**. Both are at 15%.
- **True**. Bicycle (30%) is greater than Bus (15%).
- **False**. Bus (15%) is less than Walking (40%).
- **True**. 30% + 40% = 70%, which is more than half (50%).
- **True**. Bus is tied with Car for the lowest percentage (15%).

MCQ 102: This table shows the relative frequency of students' favorite subjects.

Subject	Relative Frequency (%)
Maths	46%
Science	44%
English	10%

Based on the table, check all the true statements:

☒ Maths is the most popular subject.

☒ English is the least popular subject.

☒ Maths and Science are almost equally popular.

☐ Students get good grades in Maths.

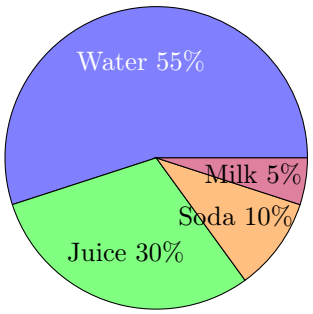
☐ English is the most popular subject.

Answer:

- **True**. At 46%, Maths has the highest percentage.
- **True**. English has the lowest percentage at 10%.
- **True**. There is only a small difference between 46% and 44%.
- **False**. This table shows what students *like* (preference), not how well they perform (grades). We cannot make a conclusion about grades from this data.
- **False**. English is the least popular, not the most.

F.1.2 INTERPRETING VISUAL DATA

Ex 103: This pie chart shows the favorite beverages of a group of children.



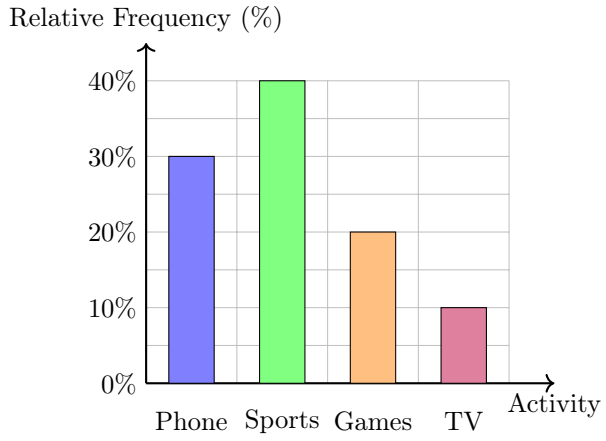
Answer these questions based on the pie chart:

1. Which beverage is the mode? **Water**
2. Which beverage is the least popular? **Milk**
3. Is it true that more children prefer Juice than Soda? **Yes**

Answer:

1. The mode is **Water**. It has the largest slice, representing the highest frequency (55%).
2. The least popular beverage is **Milk**. It has the smallest slice, representing the lowest frequency (5%).
3. **Yes**. More children prefer Juice (30%) than Soda (10%).

Ex 104: This bar chart shows how students spend their free time.



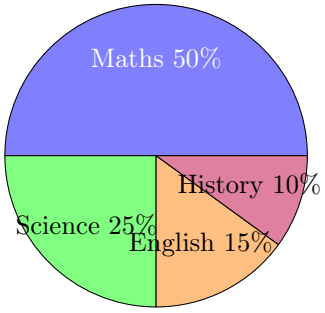
Answer these questions based on the bar chart:

1. What is the most popular activity (the mode)? **Sports**
2. What is the least popular activity? **TV**
3. Do more students play games than use their phones? **No**

Answer:

1. The most popular activity is **Sports**, because it has the tallest bar (40%).
2. The least popular activity is **TV**, because it has the shortest bar (10%).
3. **No**. Fewer students prefer Games (20%) than Phone (30%).

Ex 105: This pie chart shows the proportion of time students spend studying different subjects.



Answer these questions based on the pie chart:

1. Which subject receives the most study time? **Maths**
2. Which subject receives the least study time? **History**
3. Do students spend more time on English than Science? **No**

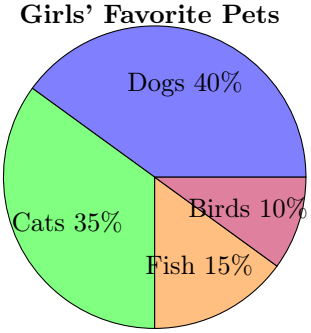
Answer:

1. **Maths** receives the most study time, as it has the largest slice (50%).
2. **History** receives the least study time, as it has the smallest slice (10%).
3. **No**. Students spend less time on English (15%) than on Science (25%).

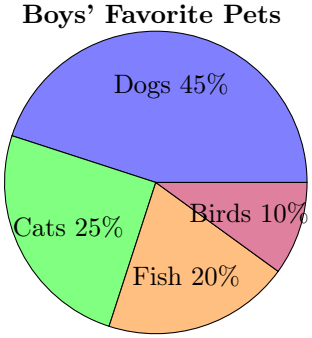
F.1.3 COMPARING DATA SETS WITH PIE CHARTS

MCQ 106: The pie charts below show the favorite pets for two different groups: Girls and Boys.

Girls' Favorite Pets



Boys' Favorite Pets



Based on a comparison of the two charts, check the true statements:

☒ Dogs are the most popular pet for both girls and boys.

☒ A higher percentage of girls like cats than boys do.

☐ A lower percentage of boys like fish than girls do.

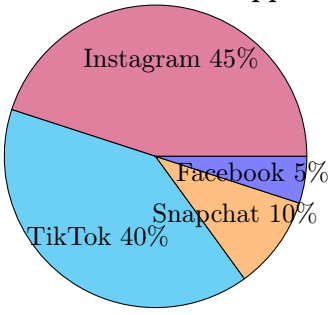
☒ Birds are equally popular among girls and boys.

Answer:

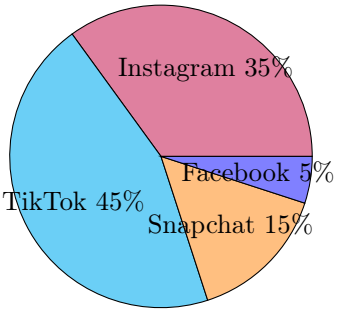
- **True**. Dogs have the largest slice for both Girls (40%) and Boys (45%).
- **True**. The percentage for Cats is higher for Girls (35%) than for Boys (25%).
- **False**. The percentage for Fish is higher for Boys (20%) than for Girls (15%).
- **True**. The percentage for Birds is the same for both groups (10%).

MCQ 107: These pie charts show the favorite social media apps for girls and boys.

Girls' Favorite Apps



Boys' Favorite Apps



Based on a comparison of the two charts, check the true statements:

☐ Instagram is the most popular app for both girls and boys.

☒ A higher percentage of boys like TikTok than girls do.

☐ A higher percentage of girls like Snapchat than boys do.

☒ Facebook is the least popular app for both groups.

Answer:

- **False**. Instagram is the most popular for Girls (45%), but TikTok is the most popular for Boys (45%).
- **True**. Boys' preference for TikTok (45%) is higher than Girls' (40%).
- **False**. Girls' preference for Snapchat (10%) is lower than Boys' (15%).
- **True**. Facebook has the smallest slice for both Girls (5%) and Boys (5%).

F.1.4 MAKING COMPARISONS WITH THE MEAN

Ex 108: The mean score for girls on a math test is 87, while the mean score for boys is 75.
Based on this data, what can be concluded about the performance of girls compared to boys?

Answer: **Conclusion:** On average, the girls in this group performed better than the boys.
Evidence: The mean score for girls (87) is higher than the mean score for boys (75).

Ex 109: The mean annual salary at Company A is \$65,000, while at Company B, it is \$58,000.
Which company pays a higher salary on average?

Answer: **Conclusion:** Company A pays a higher salary on average.
Evidence: The mean salary at Company A (\$65,000) is greater than the mean salary at Company B (\$58,000).

Ex 110: The mean summer temperature in City P is 26°C, while in City Q, it is 29°C.
Which city is hotter on average during the summer?

Answer: **Conclusion:** City Q is hotter on average during the summer.
Evidence: The mean temperature in City Q (29°C) is higher than in City P (26°C).

Ex 111: The mean household income in Neighborhood A is \$82,000, while in Neighborhood B it is \$68,500.
Which neighborhood has a higher central tendency for income?

Answer: **Conclusion:** Neighborhood A has a higher central tendency for income.
Evidence: The mean is a measure of central tendency. Since the mean income in Neighborhood A (\$82,000) is greater than in Neighborhood B (\$68,500), its central tendency is higher.

F.1.5 EVALUATING STATISTICAL CLAIMS

MCQ 112: **Scenario:** Company A reports a mean salary of \$50,000, while Company B reports a mean salary of \$55,000.
Can we conclude that the mean salary is higher in Company A?

☐ Yes

☒ No

☐ The data is insufficient to answer

Answer: **No**. The question asks if the mean salary is higher in Company A. The data shows that the mean for Company A (\$50,000) is *lower* than for Company B (\$55,000).

MCQ 113: **Scenario:** In 2023, the mean temperature was 22°C. In 2024, it was 24°C.
Can we conclude that temperatures were more variable in 2024?

☐ Yes

☐ No

☒ The data is insufficient to answer

Answer: **The data is insufficient to answer**. The question asks about variability (spread), but we are only given the mean (a measure of central tendency) for each year. To compare variability, we would need a measure of spread, such as the range or IQR.

MCQ 114: **Scenario:** Store A and Store B both have a mean daily sale of \$1,500. Store A's sales range from \$1,000 to \$2,000. Store B's sales range from \$1,400 to \$1,600.
Were the sales more variable in Store A than in Store B?

☒ Yes

☐ No

☐ The data is insufficient to answer

Answer: **Yes**. The question is about variability. We can calculate the range for each store.

- Range for Store A = \$2,000 - \$1,000 = \$1,000.
- Range for Store B = \$1,600 - \$1,400 = \$200.

Since Store A has a larger range, its sales were more variable.

MCQ 115: **Scenario:** In a study, the mean height of girls was 160 cm, and the mean height of boys was 162 cm.
Are girls taller than boys on average?

☐ Yes

☒ No

☐ The data is insufficient to answer

Answer: **No**. The question asks if girls are taller on average. The data shows that the mean height for girls (160 cm) is *less than* the mean height for boys (162 cm).

F.1.6 INTERPRETING AND COMPARING STATISTICAL MEASURES

Ex 116: In Country X, the interquartile range (IQR) of salaries was \$20,000 in 2022 and \$25,000 in 2023. Does this indicate greater salary inequality in 2023?

Answer: **Conclusion:** Yes, this suggests greater salary inequality in 2023.
Evidence: The IQR is a measure of spread. Since the IQR increased from \$20,000 to \$25,000, it means the income of the middle 50

Ex 117: Two schools had the same mean score of 14/20 on a math exam. However, School A had an IQR of 4, while School B had an IQR of 7.
Which school had more variability in its student results?

Answer: **Conclusion:** School B had more variability in its results.
Evidence: Although the means were the same, the IQR measures the spread of the middle 50

Ex 118: In City X, the mean income is \$40,000 with an IQR of \$10,000. In City Y, the mean income is \$45,000 with an IQR of \$18,000.
Which city shows more income disparity?

Answer: **Conclusion:** City Y shows more income disparity.
Evidence: Income disparity is measured by spread, not the average. City Y has a much larger IQR (\$18,000) compared to City X (\$10,000), which means the incomes in City Y are more varied.

Ex 119: Investment A has a mean annual return of \$5,000 with an IQR of \$2,000. Investment B has a mean annual return of \$6,000 with an IQR of \$4,000.
If you are only seeking the highest average return, which investment is more attractive?

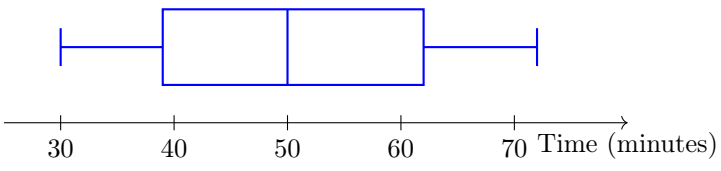
Answer: **Conclusion:** Investment B is more attractive based on average return.
Evidence: The question asks to only consider the mean. Since the mean return of Investment B (\$6,000) is higher than that of Investment A (\$5,000), it is the more attractive option for this specific goal.

Ex 120: Using the same data as the previous question: Investment A has a mean return of \$5,000 (IQR \$2,000), and Investment B has a mean return of \$6,000 (IQR \$4,000).
If you prefer a safer, more predictable investment, which one should you choose?

Answer: **Conclusion:** Investment A is the safer and more predictable choice.
Evidence: Predictability and safety are related to the spread of the data. A smaller IQR means less variability and more consistent returns. Investment A has a much smaller IQR (\$2,000) than Investment B (\$4,000).

F.1.7 INTERPRETING BOX PLOTS IN CONTEXT

MCQ 121: **Scenario:** This box plot shows the number of minutes passengers spent waiting in an airport lounge.



Evaluate the following statements. Check all that are true.

☒ "Half the passengers wait more than 50 minutes."

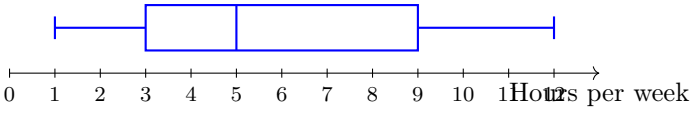
☒ "25% of passengers wait more than 62 minutes."

☐ "The shortest wait time was 72 minutes."

Answer:

- **True**. The median is 50 minutes, which means exactly half (50%) of the passengers waited longer than this time.
- **True**. The upper quartile (Q3) is 62 minutes. By definition, 25% of the data lies above Q3.
- **False**. The value 72 is the maximum (longest) wait time, not the minimum (shortest), which was 30 minutes.

MCQ 122: **Scenario:** This box plot shows the hours per week students in a music school practice their instrument.



Evaluate the following statements. Check all that are true.

☒ "The middle 50% of students practice between 3 and 9 hours per week."

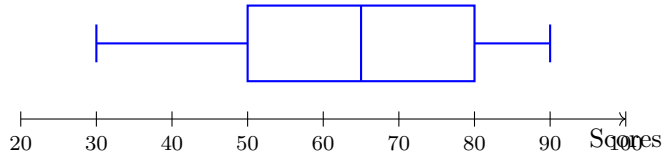
☐ "Every student practices at least 3 hours per week."

☒ "75% of students practice 9 hours or less per week."

Answer:

- **True**. The "box" part of the plot represents the middle 50% of the data, and it extends from Q1 (3 hours) to Q3 (9 hours).
- **False**. The minimum value (the end of the left whisker) is 1 hour, which means at least one student practices less than 3 hours.
- **True**. The upper quartile (Q3) is 9 hours. By definition, 75% of the data falls at or below Q3.

MCQ 123: **Scenario:** This box plot shows students' scores on a reading test (out of 100).



Evaluate the following statements. Check all that are true.

☒ "Half the students scored 65 or higher."

☒ "25% of students scored below 50, which might indicate they need extra reading help."

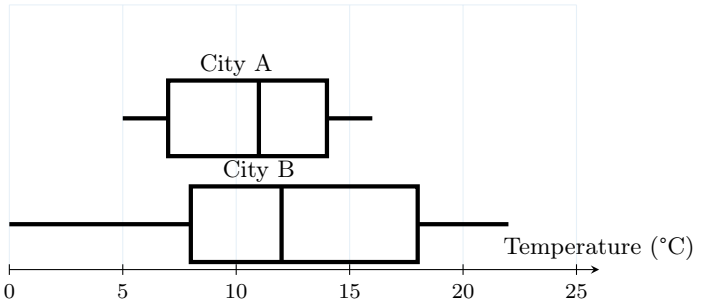
☐ "The lowest score was 90."

Answer:

- **True**. The median score is 65. By definition, 50% of the students scored at or above the median.
- **True**. The lower quartile (Q1) is 50. By definition, 25% of the students scored at or below this value.
- **False**. 90 is the maximum score (the highest). The minimum score (the lowest) was 30.

F.1.8 COMPARING DISTRIBUTIONS WITH BOX PLOTS

MCQ 124: **Scenario:** These box plots show the daily temperatures (°C) in City A and City B over a year.



Based on the plots, check the true statements.

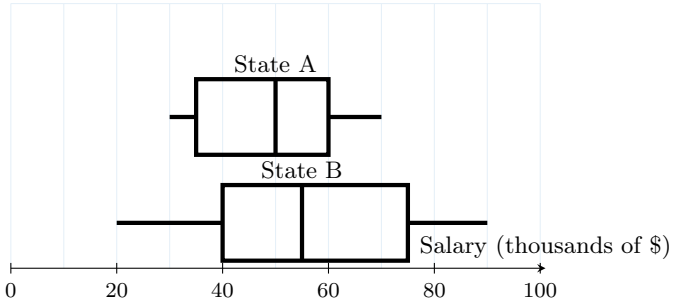
☒ "City B's temperatures are more spread out than City A's."

☐ "City A has colder minimum temperatures than City B."

Answer:

- **True**. We can compare the ranges. The range for City B is $22 - 0 = 22°C$, while the range for City A is $16 - 5 = 11°C$. Since City B has a larger range, its temperatures are more spread out.
- **False**. The minimum temperature for City A is 5°C, while the minimum for City B is 0°C. Therefore, City B experiences colder minimum temperatures.

MCQ 125: **Scenario:** These box plots show yearly salaries (in thousands of $) in State A and State B.



Based on the plots, check the true statements.

☐ "The range of salaries in State A is wider than in State B."

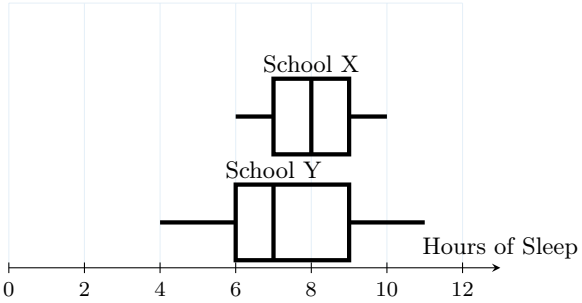
☒ "The median salary in State B is higher than in State A."

☒ "State B has greater income inequality than State A."

Answer:

- **False**. The range for State A is \$70k − \$30k = \$40k. The range for State B is \$90k − \$20k = \$70k. State B has a wider range.
- **True**. The median salary for State B (\$55k) is higher than the median for State A (\$50k).
- **True**. Income inequality is a measure of spread. Since State B has a larger range (\$70k vs. \$40k) and a larger IQR (\$75k-\$40k = \$35k vs. \$60k-\$35k = \$25k), it shows greater income inequality.

MCQ 126: **Scenario:** These box plots show the hours of sleep per night for teens in School X and School Y.



Based on the plots, check the true statements.

☒ "The sleep hours for teens in School X are more consistent than in School Y."

☒ "The median teen in School X gets more sleep than the median teen in School Y."

Answer:

- **True**. "Consistent" means less spread out. The range for School X is $10 - 6 = 4$ hours, and the IQR is $9 - 7 = 2$ hours. The range for School Y is $11 - 4 = 7$ hours, and the IQR is $9 - 6 = 3$ hours. Since School X has a smaller range and a smaller IQR, its sleep hours are more consistent.
- **True**. The median for School X (8 hours) is higher than the median for School Y (7 hours).

F.1.9 INTERPRETING AVERAGES VS. ABSOLUTE STATEMENTS

Ex 127: Claim: "The average science score is higher for Class 1 than for Class 2. Therefore, every student in Class 1 is better at science than every student in Class 2."
Evaluate this statement and explain whether the conclusion is statistically valid.

Answer: The conclusion is **not statistically valid**.

- The **mean** is a measure of central tendency; it describes the typical performance of a group as a whole.
- It does not provide information about the spread or the individual values within the group. It is highly likely that high-performing students in Class 2 scored better than some students in Class 1.
- The statement makes an invalid overgeneralization from a group average to every individual.

Ex 128: Claim: "The slowest runner in Team A was faster than the fastest runner in Team B. Therefore, every runner in Team A is faster than every runner in Team B."
Evaluate this statement and explain whether the conclusion is statistically valid.

Answer: The conclusion is **statistically valid**.

- This statement is not based on an average but on a comparison of the **extrema** (minimum and maximum values) of the two datasets.
- Let T_A be the set of times for Team A and T_B for Team B. The claim states that $\max(T_A) < \min(T_B)$ (a faster time is a smaller number).

- If the worst time in Team A is better than the best time in Team B, it logically follows that all times in Team A are better than all times in Team B. There is no overlap between the ranges of the two datasets.

Ex 129: Claim: "The average score in the piano competition was higher for participants from School X than from School Y. Therefore, every pianist from School X played better than every pianist from School Y."
Evaluate this statement and explain whether the conclusion is statistically valid.

Answer: The conclusion is **not statistically valid**. This is the same logical flaw as in the first exercise. The mean describes the overall performance of the group, not the performance of each individual. A higher mean for School X does not prevent a talented pianist from School Y from outscoring a less-skilled pianist from School X.

Ex 130: Claim: "The tallest player on Team Red is shorter than the shortest player on Team Blue. Therefore, every player on Team Blue is taller than every player on Team Red."
Evaluate this statement and explain whether the conclusion is statistically valid.

Answer: The conclusion is **statistically valid**. The statement compares the maximum value of one dataset (Team Red's heights) to the minimum value of another (Team Blue's heights). If the highest value in the first group is smaller than the lowest value in the second group, then all values in the second group must be greater than all values in the first.