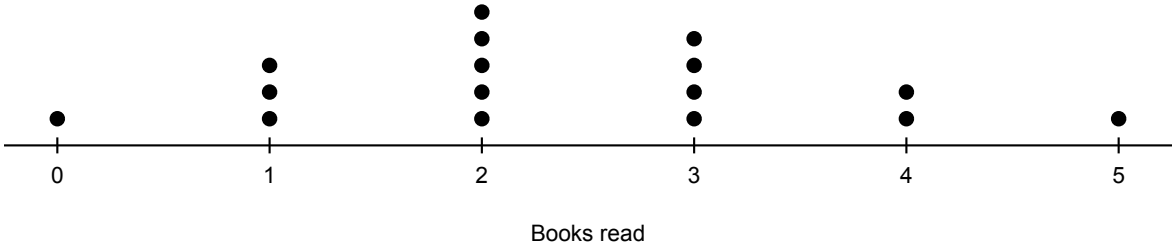
Name: _____ Date: _____

Use the repeated values

The plot records books read last month by a group of students. Each dot represents one student.



(a) How many students are represented? How many read exactly 2 books?

Set it up: Total students = sum of all frequencies. Exactly 2 means use the count above 2.

Work: Total = 1 + 3 + 5 + 4 + 2 + 1 = 16. The frequency at 2 is 5.

Answer: 16 students in total; 5 read exactly 2 books.

(b) Find the mean and median number of books. Round the mean to three decimal places.

Set it up: Mean = sum of (value times frequency) / total count. Median uses ordered student positions.

Work: Weighted total = (0 x 1) + (1 x 3) + (2 x 5) + (3 x 4) + (4 x 2) + (5 x 1) = 38. Mean = 38/16 = 2.375. Middle positions: 8 and 9. Median = 2.

Answer: Mean = 2.375 books; median = 2 books.

(c) What fraction of these students read at least 2 books? Give a decimal rounded to four places and a percent rounded to two places.

Set it up: Whole group: all students shown. Favorable group: those at the cutoff or above. Share = favorable count / total.

Work: Count at or above 2 = 12. Fraction = 12/16. Decimal = 0.7500. Percent = (12/16) x 100 = 75.00%.

Answer: 12/16 = 0.7500 = 75.00% of these students.

(d) Describe the clusters or gaps. Can you read the exact number of students at each value from this plot?

Set it up: Read the frequency of every shown value. A dot plot retains individual counts.

Work: Counts from 0 through 5: 1, 3, 5, 4, 2, 1. The highest stack is at 2; there are no empty values between the endpoints.

Answer: No empty values between 0 and 5; the highest stack is at 2. Yes, the exact count at every value is the number of dots above it.

One unusually long wait

A sample of five waiting times, in minutes: 2, 4, 6, 8, 10.

(a) Find the mean and median. State the units.

Set it up: Mean = total / number of values. Sort the list to find the median.

Work: Total = 30; n = 5. Mean = 30/5 = 6. Ordered list: 2, 4, 6, 8, 10. The middle value is 6.

Answer: Mean = 6 minutes. Median = 6 minutes.

(b) Find the mode and range.

Set it up: Mode = most frequent value, if one exists. Range = maximum minus minimum.

Work: Mode: no mode (every value occurs once). Range = 10 minus 2 = 8.

Answer: Mode: no mode (every value occurs once). Range = 8 minutes.

(c) Find the sample standard deviation, rounded to three decimal places. Explain which calculator output applies.

Set it up: Sum the squared distances from the mean, divide by n minus 1, then take the square root. This is a sample, so use Sx.

Work: Mean = 6. Distances from the mean: -4, -2, 0, 2, 4. Squared distances: 16, 4, 0, 4, 16; their sum = 40. Variance = 40/4 = 10. Standard deviation = square root of 10 = 3.162.

Answer: Sx = 3.162 minutes. It describes spread around the mean.

(d) Replace 10 minutes with 30 minutes. Recalculate the mean and median. Which measure changes less, and why?

Set it up: Use the new total for the mean. Reorder the new list for the median.

Work: New data: 2, 4, 6, 8, 30. Mean = 50/5 = 10. Median = 6. The mean rises by 4; the median changes by 0.

Answer: The median changes less: it stays at 6 minutes. It is resistant to this large value.

Waiting time set 2

Waiting times have been grouped into equal width classes. Each lower boundary is included; each upper boundary is excluded.

Time (minutes)	Frequency	Relative frequency
20 <= x < 30	15	0.5000
30 <= x < 40	8	0.2667
40 <= x < 50	4	0.1333
50 <= x < 60	2	0.0667
60 <= x < 70	1	0.0333

(a) Complete the relative frequency column as decimals rounded to four places. State the exact and rounded totals.

Set it up: Whole group: all waiting times. Relative frequency = class frequency / total frequency.

Work: n = 30. 15/30 = 0.5000; 8/30 = 0.2667; 4/30 = 0.1333; 2/30 = 0.0667; 1/30 = 0.0333. The exact fractions sum to 1; displayed rounded values sum to 1.0000.

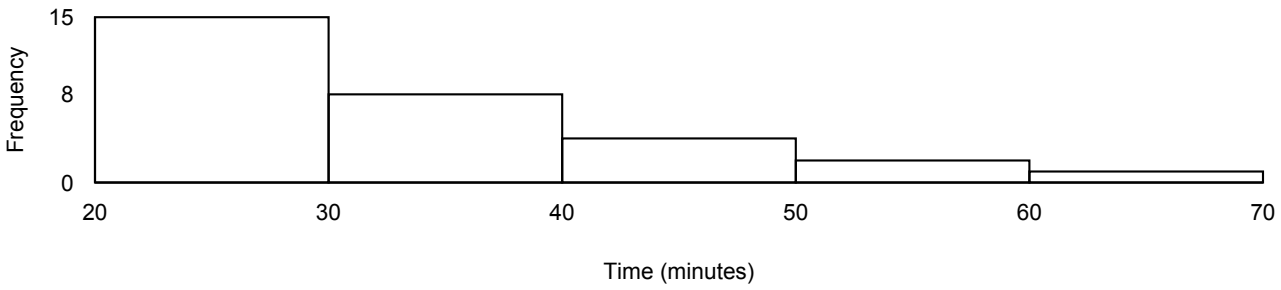
Answer: Relative frequencies: 0.5000, 0.2667, 0.1333, 0.0667, 0.0333. Exact total = 1. Rounded total = 1.0000.

(b) Draw a frequency histogram with touching bars. Label both axes.

Set it up: Horizontal axis: time classes. Vertical axis: frequency. Heights equal the class counts.

Work: Class boundaries: 20, 30, 40, 50, 60, 70. Bar heights: 15, 8, 4, 2, 1.

Answer: The completed histogram has one touching bar for each class, using the stated frequency as its height.



(c) Describe the shape. Which class contains a waiting time of 40 minutes?

Set it up: A class includes its lower boundary and excludes its upper boundary. Shape depends on where the tail extends.

Work: Most counts are at the lower times, with a tail toward larger times. The value 40 satisfies 40 <= x < 50.

Answer: Right skewed. 40 belongs in the class 40 <= x < 50.

(d) Estimate the mean with class midpoints. Can you recover the exact raw mean or exact median from this table alone?

Set it up: Estimated mean = sum of (class midpoint times frequency) / total count.

Work: Midpoints: 25, 35, 45, 55, 65. Weighted total = (25 x 15) + (35 x 8) + (45 x 4) + (55 x 2) + (65 x 1) = 1010. Estimate = 1010/30 = 33.667.

Answer: Estimated mean = 33.667 minutes. Neither the exact raw mean nor the exact median can be recovered. For this skewed distribution, the median is a suitable resistant center, though its exact value is unavailable here.

A zero in the table

Participants attended a workshop at one of four centers. Use the same table for every part. Select one participant at random from the group named in each prompt.

Center	Completed	Did not complete	Total
North	8	0	8
South	3	7	10
East	4	6	10
West	5	7	12
Total	20	20	40

(a) A participant is chosen from ALL 40 participants. What is the probability that the participant is from North AND completed the workshop? Round to four decimal places.

Set it up: Whole group: all participants. Joint probability = overlap count / grand total.

Work: Overlap = 8. Grand total = 40. P(North and completed) = 8/40 = 0.2000.

Answer: 0.2000.

(b) A participant is chosen ONLY from North. What is the probability that the participant completed the workshop? Round to four decimal places.

Set it up: Whole group: participants from this center only. Conditional probability = completed in the row / row total.

Work: North total = 8; completed = 8. P(completed given North) = 8/8 = 1.0000.

Answer: 1.0000.

(c) A participant is chosen ONLY from those who completed. What is the probability that the participant is from North? Round to four decimal places.

Set it up: Whole group: all participants who completed. Reverse conditional probability = this center and completed / completed total.

Work: Completed total = 20; overlap = 8. P(North given completed) = 8/20 = 0.4000.

Answer: 0.4000.

(d) A participant is chosen from ALL 40 participants. What is the probability that the participant is from North OR completed, or both? Round to four decimal places.

Set it up: Whole group: all participants. Union probability = (row total plus completed total minus overlap) / grand total.

Work: P(North or completed) = (8 + 20 minus 8)/40 = 20/40 = 0.5000.

Answer: 0.5000.

Across class sections

A college has 900 introductory statistics students in 30 class sections of 30 students each. Researchers randomly select 3 students from EVERY section and record each selected student's usual study method and next quiz score. Researchers do not assign study methods. Some invited students do not respond.

(a) Classify the study and identify the population, intended sample size, explanatory variable and response.

Set it up: Observational studies record existing conditions. Count selected students across all sections.

Work: Population = 900. Intended sample = 30 x 3 = 90 students. Explanatory variable = usual study method; response = next quiz score.

Answer: Observational study. Population: 900 students. Intended sample: 90 students. Explanatory variable: usual study method. Response: next quiz score. Actual respondent count is not given.

(b) Is the sampling method stratified, cluster, or a simple random sample of all students? Explain.

Set it up: Stratified sampling selects within every stratum. Cluster sampling selects whole groups.

Work: Each class section is a stratum, and 3 students are randomly selected within each one.

Answer: Stratified sampling by class section. It is not cluster sampling.

(c) What bias could nonresponse create? Can the results prove that a method caused higher scores?

Set it up: Nonresponse bias can arise if response is related to relevant traits. Causal inference needs an appropriate assignment design.

Work: Students who answer may differ in study habits or scores from those who do not. No study method was assigned.

Answer: Nonresponse could distort the sample; its direction is unknown. This observational design cannot by itself establish a causal effect.