HOMEWORK 2.1

Frequency Distributions and Their Graphs

Problem 1 A frequency distribution has the classes 10–19, 20–29, 30–39, 40–49, and 50–59.

a.) Find the class width. 10

$20 - 10 = 10$ (difference of consecutive lower limits)

b.) Find the midpoint of each class. 14.5, 24.5, 34.5, 44.5, 54.5

midpoint $= \frac{10 + 19}{2} = 14.5$, then add 10 each time

c.) Find the class boundaries. 9.5, 19.5, 29.5, 39.5, 49.5, 59.5

subtract 0.5 from each lower limit, and add 0.5 to the last upper limit

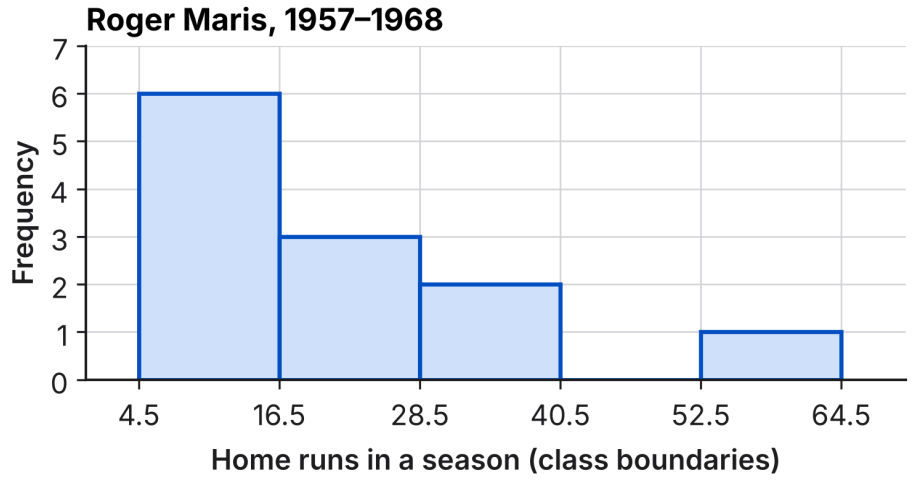
Problem 2 Roger Maris hit these home run totals from 1957 to 1968 (Baseball-Reference): 14, 28, 16, 39, 61, 33, 23, 26, 8, 13, 9, 5. Complete the frequency distribution using 5 classes. The first class is 5–16.

Range $= 61 - 5 = 56$

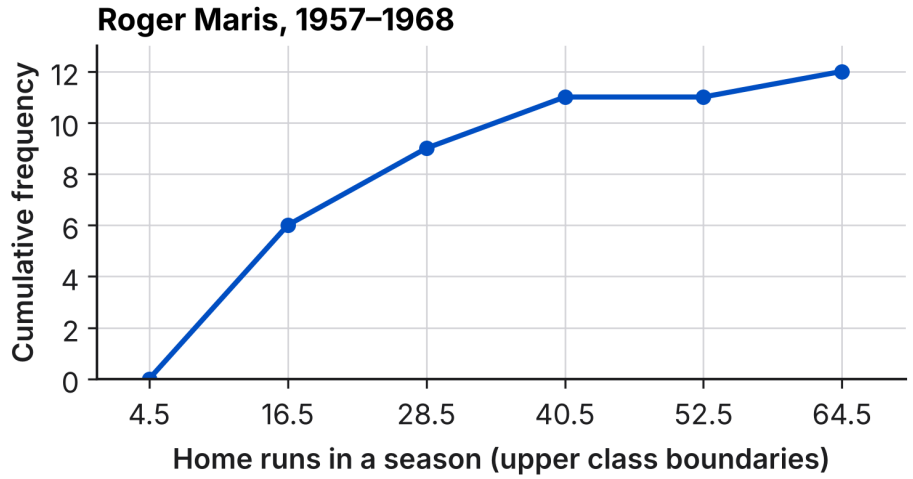
$\frac{56}{5} = 11.2 \rightarrow$ round up: class width $= 12$

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
5–16	6	10.5	0.50	6
17–28	3	22.5	0.25	9
29–40	2	34.5	0.17	11
41–52	0	46.5	0	11
53–64	1	58.5	0.08	12
	$n = 12$		$\Sigma = 1$	

Problem 3 Use your table from Problem 2 to draw a frequency histogram. Label the class boundaries on the horizontal axis.



Problem 4 Use your table from Problem 2 to draw an ogive.



Problem 5 Use your table and graphs to answer each question.

a.) Which class has the greatest frequency? 5–16

6 of Maris's 12 seasons were in 5–16.

b.) In what percent of his seasons did Maris hit 28 or fewer home runs? 75%

cumulative frequency at 28.5 is 9

$\frac{9}{12} = 0.75 = 75\%$

c.) Describe the shape of the histogram.

Skewed right: most seasons are on the left, with a long tail to the right (the 61).

Problem 6 What is the midpoint of the class 40–49?

(A) 40
(B) 44.5 midpoint $= \frac{40 + 49}{2} = 44.5$
(C) 45
(D) 49.5

Problem 7 A relative frequency histogram has a bar of height 0.30 over the class 20–29. The data set has 40 entries. How many entries are in the class 20–29?

$f = (\text{relative frequency})(n) = 0.30(40) = 12$ entries

Problem 8 A data set has a minimum of 12 and a maximum of 87. You want 6 classes.

a.) Find the class width. 13

$\frac{87 - 12}{6} = 12.5 \rightarrow$ round up: 13

b.) List the lower class limits. 12, 25, 38, 51, 64, 77

start at the minimum, 12, and add 13 each time

Problem 9 Which graph plots cumulative frequencies at the upper class boundaries?

(A) frequency histogram
(B) frequency polygon
(C) ogive
(D) relative frequency histogram

An ogive (cumulative frequency graph) uses upper boundaries and starts at 0.

HOMEWORK 2.2

More Graphs and Displays

Problem 1 Aaron Judge hit these home run totals from 2017 to 2025 (Baseball-Reference): 52, 27, 27, 9, 39, 62, 37, 58, 53. Complete the stem-and-leaf plot.

Stem	Leaves
0	9
1	
2	7 7
3	7 9
4	
5	2 3 8
6	2

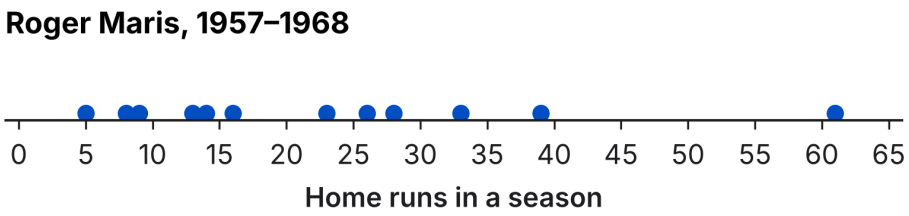
Key: 5 | 2 = 52 home runs

a.) On which stem do the most leaves fall? 5

b.) Why do stems 1 and 4 have no leaves? no seasons in the 10s or 40s

Judge never hit 10–19 or 40–49 home runs.
A stem with no leaves still stays in the plot.

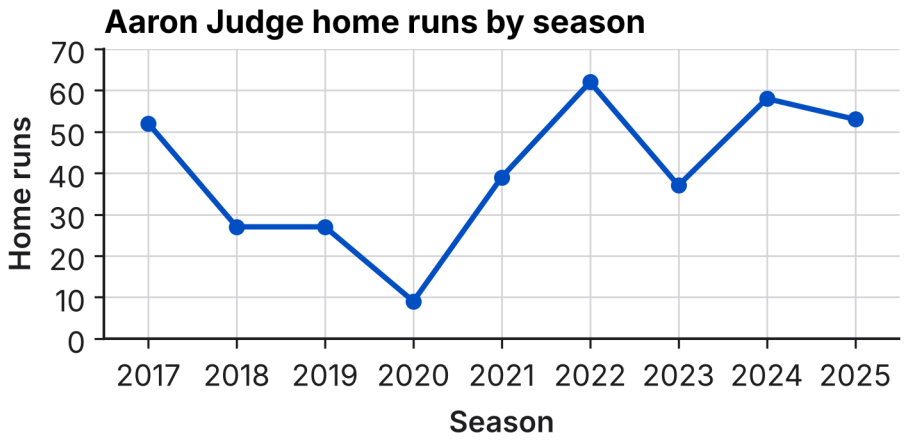
Problem 2 Make a dot plot of Roger Maris's home runs, 1957–1968: 14, 28, 16, 39, 61, 33, 23, 26, 8, 13, 9, 5.



Which data entry stands out from the rest? 61

61 sits far to the right, with a gap of 22 after the next-highest entry, 39.

Problem 3 Make a time series chart of Judge's home runs, 2017–2025.



Which season is lowest? What might explain it? 2020

2020 was a 60-game season, and Judge played only 28 games.

Problem 4 A survey asked 50 students to name their favorite sport to watch. The results are below. Find the relative frequency and the central angle of each sector of a pie chart.

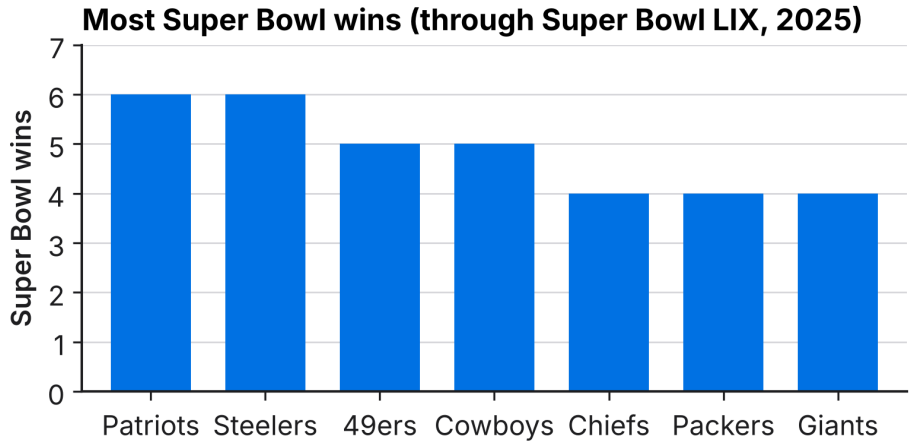
Sport	Frequency	Relative frequency	Angle
Football	22	0.44	158.4°
Basketball	15	0.30	108°
Baseball	9	0.18	64.8°
Other	4	0.08	28.8°

Football: $\frac{22}{50} = 0.44$ angle $= 0.44(360°) = 158.4°$

Which sport would be the first bar in a Pareto chart? Why? Football

A Pareto chart puts bars in order of decreasing frequency. Football has the highest, 22.

Problem 5 Use the Pareto chart of Super Bowl wins from the notes (through Super Bowl LIX, 2025).



a.) How many more Super Bowls have the Steelers won than the Chiefs? 2

$6 - 4 = 2$

b.) What fraction of these seven teams' 34 wins belong to the 49ers and Cowboys together? $10/34 \approx 0.29$

$\frac{5+5}{34} = \frac{10}{34} \approx 0.29$

Problem 6 Which display shows a relationship between two quantitative variables, such as games played and home runs?

(A) pie chart
(B) scatter plot
(C) Pareto chart
(D) dot plot

A scatter plot graphs paired data (x, y).

Problem 7 Which display keeps every original data entry?

(A) histogram
(B) pie chart
(C) stem-and-leaf plot
(D) ogive

Each leaf rebuilds an entry, like 5 | 2 = 52.

HOMEWORK 2.3

Measures of Central Tendency

Problem 1 Find the mean, median, and mode of Aaron Judge's home runs, 2017–2025: 52, 27, 27, 9, 39, 62, 37, 58, 53.

a.) Mean about 40.4

$\bar{x} = \frac{\Sigma x}{n} = \frac{364}{9} \approx 40.4$

b.) Median 39

ordered: 9, 27, 27, 37, 39, 52, 53, 58, 62

9 entries: the middle is the 5th → 39

c.) Mode 27

27 occurs twice (2018 and 2019).

Problem 2 The scores on a short quiz are 6, 7, 7, 8, 9, 9, 9, 10, 10, 10. Find the mean, median, and mode.

Mean 8.5 Median 9 Mode 9 and 10

$\bar{x} = \frac{85}{10} = 8.5$

$\text{median} = \frac{9+9}{2} = 9$

9 and 10 each occur 3 times: two modes (bimodal)

Problem 3 Decide which measure of central tendency best describes each data set: mean, median, or mode. Explain.

a.) The favorite NFL team of each student in a class mode

qualitative data: only the mode works

b.) Home prices in a town where a few homes cost several million dollars median

the expensive homes are outliers; the median resists them

c.) The heights of the students in a class, with no outliers mean

no outliers: the mean uses every entry

Problem 4 A student takes 4 classes. Find the student's grade point average (GPA). Use A = 4, B = 3, C = 2.

Class	Grade	Credit hours
Statistics	A	4
English	B	3
History	C	3
Gym	A	1

GPA about 3.18

$$\bar{x} = \frac{\Sigma(x \cdot w)}{\Sigma w} = \frac{4(4) + 3(3) + 2(3) + 4(1)}{4 + 3 + 3 + 1}$$

$$\bar{x} = \frac{35}{11} \approx 3.18$$

Problem 5 Use the frequency distribution of Maris's home runs from Homework 2.1 to estimate his mean.

Midpoint, x	Frequency, f	x · f
10.5	6	63
22.5	3	67.5
34.5	2	69
46.5	0	0
58.5	1	58.5
	n = 12	$\Sigma = 258$

Estimated mean 21.5

$\bar{x} = \frac{\Sigma(x \cdot f)}{n} = \frac{258}{12} = 21.5$

(actual mean ≈ 22.9)

Problem 6 Describe the most likely shape of each distribution.

a.) mean 80, median 85 skewed left

mean < median: the tail pulls the mean left

b.) mean 50, median 50 symmetric

mean = median

c.) mean 35,000, median 28,000 skewed right

mean > median: the tail pulls the mean right

Problem 7 Which measure of central tendency can be used for qualitative data?

(A) mean
(B) median

ANSWER KEY

(C) mode
(D) weighted mean

Categories cannot be added or ordered, but you can find the most common one.

Problem 8 One entry in a data set is changed from 40 to 400. Which measure changes the most?

(A) the mean
(B) the median
(C) the mode
(D) none of them

The mean uses every value, so a huge entry pulls it up. The median resists outliers.

HOMEWORK 2.4

Measures of Variation

Problem 1 Aaron Judge hit 62, 37, 58, and 53 home runs from 2022 to 2025 (Baseball-Reference). Treat these 4 seasons as a population.

x	x − μ	$(x - \mu)^2$
62	9.5	90.25
37	−15.5	240.25
58	5.5	30.25
53	0.5	0.25
Σx = 210	Σ = 0	SS = 361

a.) Range 25

$62 - 37 = 25$

b.) Mean 52.5

$\mu = \frac{210}{4} = 52.5$

c.) Population variance 90.25

$\sigma^2 = \frac{SS}{N} = \frac{361}{4} = 90.25$

d.) Population standard deviation 9.5

$\sigma = \sqrt{90.25} = 9.5$ home runs

Problem 2 Now treat Judge's 2022–2025 seasons from Problem 1 as a sample.

a.) Sample variance about 120.3

$s^2 = \frac{SS}{n-1} = \frac{361}{3} \approx 120.3$

b.) Sample standard deviation about 11.0

$s = \sqrt{120.33} \approx 11.0$ home runs

Problem 3 Use a TI-84 and all 9 of Judge's seasons, 2017–2025: 52, 27, 27, 9, 39, 62, 37, 58, 53. Treat the data as a sample.

a.) $\bar{x} \approx 40.4$ $s \approx 17.4$

STAT ▸ CALC ▸ 1-Var Stats, List: L1

$\bar{x} \approx 40.44$, $Sx \approx 17.42$

b.) Find the coefficient of variation. about 43.1%

$CV = \frac{s}{\bar{x}} \cdot 100\% = \frac{17.42}{40.44} \cdot 100\% \approx 43.1\%$

c.) Ruth's CV was about 25.6%. Whose home run totals varied more, relative to the mean? Explain.

Judge. His CV is larger, so his seasons were spread farther from his mean, relative to its size.

Problem 4 Suppose the weights of a group of NFL offensive linemen are bell-shaped with a mean of 315 pounds and a standard deviation of 15 pounds. Use the Empirical Rule.

a.) About what percent weigh between 300 and 330 pounds? 68%

$315 \pm 15 \rightarrow$ within 1 SD: about 68%

b.) About what percent weigh between 285 and 345 pounds? 95%

$315 \pm 2(15) \rightarrow$ within 2 SD: about 95%

c.) About what percent weigh between 315 and 345 pounds? 47.5%

half of 95%: $\frac{95\%}{2} = 47.5\%$

d.) About what percent weigh more than 345 pounds? 2.5%

outside 2 SD: 100% − 95% = 5%

upper tail: $\frac{5\%}{2} = 2.5\%$

e.) In a group of 200 such linemen, about how many weigh between 300 and 330 pounds? 136

0.68(200) = 136 linemen

Problem 5 Use Chebyshev's Theorem with Judge's 2017–2025 data: $\bar{x} \approx 40.4$, $s \approx 17.4$.

a.) At least what percent of the data lie within 2 standard deviations of the mean? 75%

$1 - \frac{1}{2^2} = 1 - \frac{1}{4} = 75\%$

b.) Find the interval within 2 standard deviations. 5.6 to 75.2

40.4 − 2(17.4) = 5.6

40.4 + 2(17.4) = 75.2

c.) What percent of Judge's 9 seasons actually lie in the interval? 100%

All 9 entries, from 9 to 62, are in the interval.

$\frac{9}{9} = 100\%$ (at least 75%: true)

d.) Why should you use Chebyshev's Theorem, not the Empirical Rule, for these data? not bell-shaped

The Empirical Rule needs bell-shaped data.

Chebyshev works for any shape.

Problem 6 Two data sets each have a mean of 50. Set A has s = 4. Set B has s = 12. Which statement is true?

(A) Set A is more spread out.
(B) Set B is more spread out.
(C) The sets have the same spread.
(D) Set B has more entries.

larger standard deviation → more spread out

Problem 7 For any data set, at least what percent of the entries lie within 3 standard deviations of the mean?

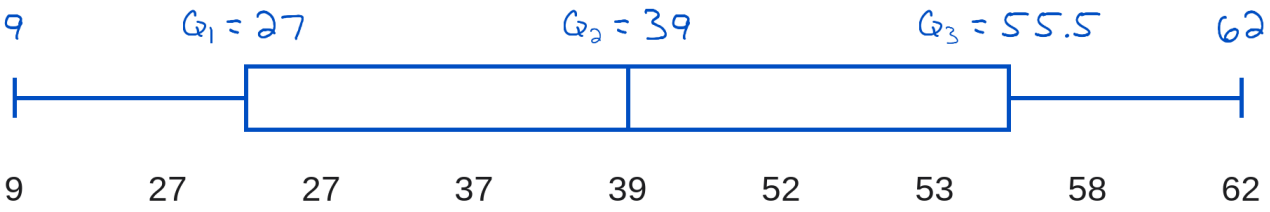
(A) 68%
(B) 75%
(C) 88.9%
(D) 99.7%

$1 - \frac{1}{3^2} = 1 - \frac{1}{9} \approx 88.9\%$ (Chebyshev, any shape)

HOMEWORK 2.5

Measures of Position

Problem 1 Use Aaron Judge's home runs, 2017–2025, in order: 9, 27, 27, 37, 39, 52, 53, 58, 62.



a.) Q_2 39 Q_1 27 Q_3 55.5

9 entries: Q_2 is the 5th entry = 39

lower 4: 9, 27, 27, 37 → $Q_1 = \frac{27+27}{2} = 27$

upper 4: 52, 53, 58, 62 → $Q_3 = \frac{53+58}{2} = 55.5$

b.) IQR 28.5

$IQR = 55.5 - 27 = 28.5$

c.) Find the fences. Are there any outliers? no

$1.5(IQR) = 1.5(28.5) = 42.75$

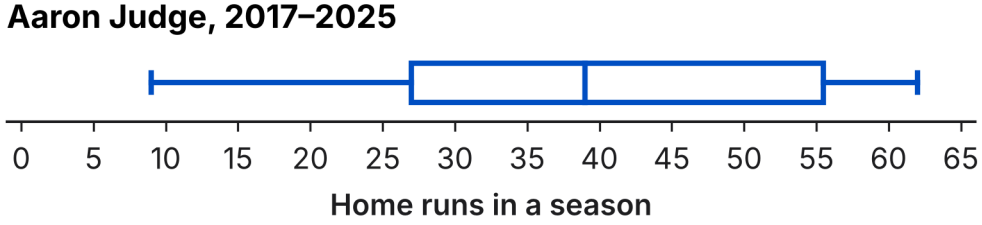
lower fence: $27 - 42.75 = -15.75$

upper fence: $55.5 + 42.75 = 98.25$

All entries are between the fences: no outliers.

d.) Write the five-number summary. 9, 27, 39, 55.5, 62

Problem 2 Draw a box-and-whisker plot of Judge's home runs.



Compare Judge's box plot with Ruth's (five-number summary 22, 35, 46, 54, 60). Whose middle 50% is more spread out?

Judge. His IQR is 28.5, while Ruth's is 19.
Judge's box is wider.

Problem 3 Find the percentile of Judge's 58 home runs (2024) among his 2017–2025 seasons. 78th

entries less than 58: 9, 27, 27, 37, 39, 52, 53 → 7 entries

percentile = $\frac{7}{9} \cdot 100 \approx 78$ → 78th percentile

Problem 4 Use Judge's sample statistics $\bar{x} \approx 40.44$ and $s \approx 17.42$.

a.) Find the z-score of his 62 home runs in 2022. about 1.24

$$z = \frac{x - \bar{x}}{s} = \frac{62 - 40.44}{17.42} \approx 1.24$$

b.) Find the z-score of his 9 home runs in 2020. about −1.80

$$z = \frac{9 - 40.44}{17.42} \approx -1.80$$

c.) Are either of these seasons unusual? Explain.

No. Both z-scores are between −2 and 2.

d.) Maris's 61 had a z-score of 2.38. Whose record season stood out more from his own career, Judge's 62 or Maris's 61?

Maris's 61: 2.38 > 1.24. It is farther above
his own mean, in standard deviations.

Problem 5 A student scores 85 on a statistics test with mean 75 and standard deviation 5. The same student scores 90 on a history test with mean 82 and standard deviation 8. On which test did the student do better compared with the class?

stats: $z = \frac{85 - 75}{5} = 2$

history: $z = \frac{90 - 82}{8} = 1$

Statistics: the score is 2 SDs above the mean.

Problem 6 In a box-and-whisker plot, what does the vertical line inside the box show?

(A) the mean
(B) the median
(C) the mode
(D) the IQR

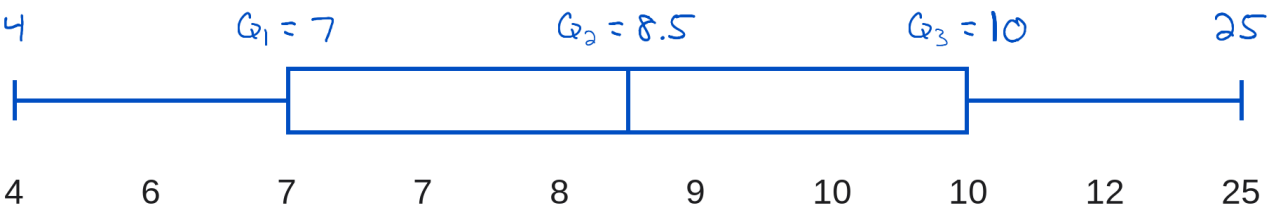
The box runs from Q_1 to Q_3; the line is Q_2.

Problem 7 A data value has a z-score of –2.5. Which statement is true?

(A) It is 2.5 units below the mean.
(B) It is 2.5 standard deviations above the mean.
(C) It is 2.5 standard deviations below the mean, so it is unusual.
(D) It is at the 25th percentile.

negative: below the mean; $|z| > 2$: unusual

Problem 8 The numbers of points a player scored in 10 games are 4, 6, 7, 7, 8, 9, 10, 10, 12, 25.



a.) Q_1 7 Q_3 10 IQR 3

lower 5: 4, 6, 7, 7, 8 → $Q_1 = 7$
upper 5: 9, 10, 10, 12, 25 → $Q_3 = 10$
IQR = 10 − 7 = 3

b.) Is 25 an outlier? Explain. yes

upper fence: $10 + 1.5(3) = 14.5$

$25 > 14.5$, so 25 is an outlier.

c.) In a modified box plot, where does the right whisker end?

At 12, the largest entry that is not an outlier.

25 is marked with a star.

Test 1 Review: Descriptive Statistics

Part I · Multiple Choice

SECTION A · NO CALCULATOR

Choose the best answer. Explain your choice in a few words.

Problem 1 A data set has a mean of 64 and a median of 71. What is the most likely shape of its distribution?

(A) symmetric
(B) uniform
(C) skewed left
(D) skewed right

mean < median → tail on the left

Problem 2 A data set is bell-shaped with a mean of 70 and a standard deviation of 6. About what percent of the data lie between 64 and 76?

(A) 34%
(B) 68%
(C) 95%
(D) 99.7%

64 and 76 are 1 SD from 70 → 68%

Problem 3 Which z-score is **unusual**?

(A) –1.45
(B) 0.82
(C) 2.31
(D) 1.99

2.31 > 2

unusual: $z < -2$ or $z > 2$

Problem 4 In a box-and-whisker plot, the box shows

(A) the middle 50% of the data — the box runs from Q_1 to Q_3: middle 50%
(B) the range
(C) the mean and standard deviation
(D) only the outliers

Problem 5 What is the midpoint of the class 20–29?

(A) 24
(B) 24.5 — $\text{midpoint} = \frac{20 + 29}{2} = 24.5$
(C) 25
(D) 29.5

Problem 6 Which graph is best for showing how one team's attendance changed from season to season?

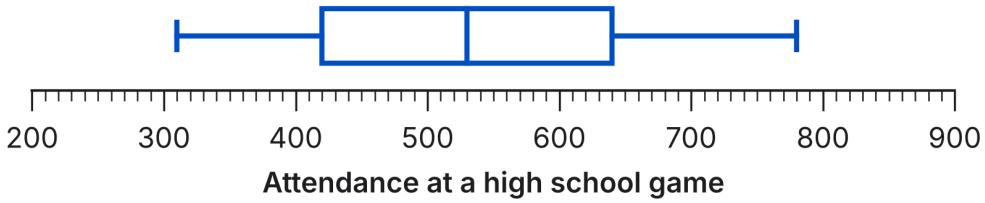
(A) pie chart
(B) Pareto chart
(C) time series chart — data measured over time → time series
(D) stem-and-leaf plot

Problem 7 Which measure of central tendency should you use for the favorite NFL team of each student in a class?

(A) mean
(B) median
(C) mode
(D) range

qualitative data → mode

Use the box-and-whisker plot for Problems 8 and 9.



Attendance at a high school game

Problem 8 What is the interquartile range?

(A) 110
(B) 220
(C) 470
(D) 1060

$IQR = Q_3 - Q_1 = 640 - 420 = 220$

Problem 9 About what percent of the data lie between the median and the maximum?

(A) 25%
(B) 50%
(C) 75%
(D) 100%

median to max: the right half of the box plus the right whisker = 50% of the data

Problem 10 A student's score is at the 80th percentile. What does this mean?

(A) The student got 80% of the questions right.
(B) About 80% of the scores are at or below this score.
(C) About 80% of the scores are above this score.
(D) The score is 80 points.

percentile = percent at or below

Part I · Multiple Choice

SECTION B · CALCULATOR

Choose the best answer. Use your TI-84 where it helps.

Problem 11 A sample of test scores has a mean of 61.3, a median of 58, and a variance of 1854.76. What is the standard deviation?

(A) 927.38
(B) 43.07 $s = \sqrt{\text{variance}} = \sqrt{1854.76} \approx 43.07$
(C) 21.53
(D) 7.83

Problem 12 A data set has a minimum of 12 and a maximum of 58. You want 6 classes. What class width should you use?

(A) 7
(B) 8 $\frac{58 - 12}{6} \approx 7.67 \rightarrow$ round up: 8
(C) 46
(D) 7.67

Problem 13 A team scored 21, 28, 14, 35, 17 points in five games. Treat the data as a sample. What is the sample standard deviation?

(A) 7.62
(B) 8.51 1-Var Stats: $Sx \approx 8.51$
(C) 72.50 ($\sigma x \approx 7.62$ is the population SD)
(D) 23.00

Problem 14 A test has a mean of 80 and a standard deviation of 6. What is the z-score of a score of 92?

(A) –2
(B) 2 $z = \frac{92 - 80}{6} = 2$
(C) 12
(D) 0.5

Problem 15 A course grade is 40% tests, 30% homework, and 30% final exam. A student has a test average of 90, a homework average of 84, and a final exam score of 75. What is the course grade?

(A) 83.0 $\bar{x} = \frac{\Sigma(x \cdot w)}{\Sigma w}$
(B) 83.7
(C) 86.2 $\bar{x} = 90(0.4) + 84(0.3) + 75(0.3) = 83.7$
(D) 80.6

Part II · Free Response

SECTION A · CALCULATOR

Show all work. Round to one decimal place where needed.

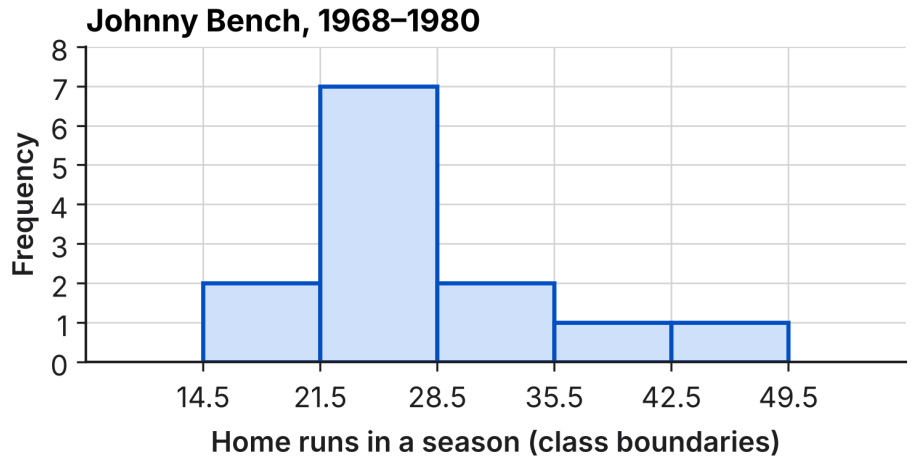
Problem 16 Johnny Bench hit these home run totals from 1968 to 1980 (Source: Baseball-Reference):

15, 26, 45, 27, 40, 25, 33, 28, 16, 31, 23, 22, 24

a.) Complete the frequency distribution. Use 5 classes. The first class is 15–21, and the class width is 7.

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
15–21	2	18	0.15	2
22–28	7	25	0.54	9
29–35	2	32	0.15	11
36–42	1	39	0.08	12
43–49	1	46	0.08	13
	$n = 13$		$\Sigma \approx 1$	

b.) Draw a frequency histogram.



c.) Describe the shape of the distribution. skewed right

Most seasons are in 22–28; the tail of high seasons (40, 45) stretches right.

Problem 17 Use Johnny Bench's data from Problem 16. Treat the data as a sample.

a.) Mean 27.3 Median 26 Mode none

$$\bar{x} = \frac{\Sigma x}{n} = \frac{355}{13} \approx 27.3$$

13 entries: median = 26

mode: no entry repeats

b.) Sample standard deviation 8.5

STAT ▶ CALC ▶ 1-Var Stats: $S_x \approx 8.49$

c.) Five-number summary 15, 22.5, 26, 32, 45 IQR 9.5

lower half: 15, 16, 22, 23, 24, 25 → $Q_1 = 22.5$

upper half: 27, 28, 31, 33, 40, 45 → $Q_3 = 32$

$IQR = 32 - 22.5 = 9.5$

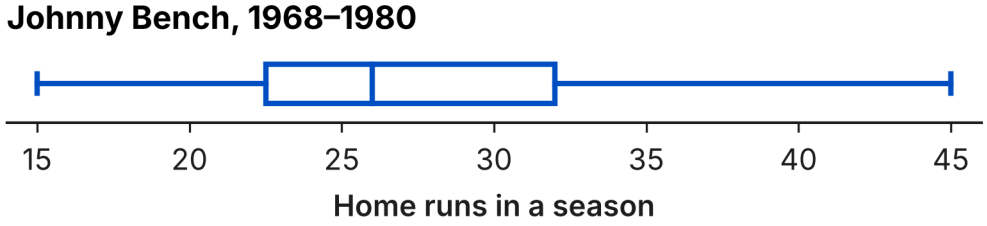
d.) Use the fences to decide whether there are any outliers. none

$1.5(IQR) = 1.5(9.5) = 14.25$

fences: $22.5 - 14.25 = 8.25$ and $32 + 14.25 = 46.25$

no entries outside the fences

e.) Draw a box-and-whisker plot.



f.) Johnny Bench hit 45 home runs in 1970, his first MVP season. Find the z-score. Is it unusual? 2.08

$$z = \frac{x - \bar{x}}{s} = \frac{45 - 27.31}{8.49} \approx 2.08$$

$|z| > 2$: unusual

Part II · Free Response

SECTION B · NO CALCULATOR

Show all work.

Problem 18 A high school quarterback completed these numbers of passes in five games: 12, 15, 18, 21, 24. Complete the table, then find each value.

x	Deviation	Squared deviation
12	-6	36
15	-3	9
18	0	0
21	3	9
24	6	36
$\Sigma x = 90$	$\Sigma = 0$	$SS = 90$

a.) Mean 18

$$\bar{x} = \frac{\Sigma x}{n} = \frac{90}{5} = 18$$

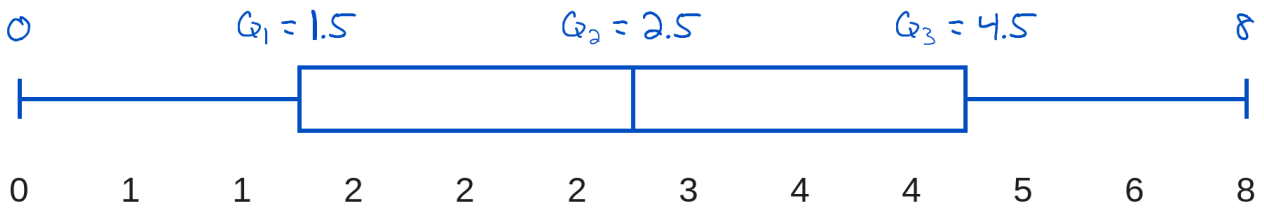
b.) Sample variance 22.5

$$s^2 = \frac{SS}{n-1} = \frac{90}{4} = 22.5$$

c.) Population variance, if the 5 games are the whole population 18

$$\sigma^2 = \frac{SS}{N} = \frac{90}{5} = 18$$

Problem 19 A basketball player made these numbers of 3-pointers in 12 games:



a.) Median 2.5 Mode 2 Range 8

12 entries: median $= \frac{2+3}{2} = 2.5$

mode: 2 occurs 3 times

range $= 8 - 0 = 8$

b.) Q_1 1.5 Q_3 4.5 IQR 3

lower 6: 0, 1, 1, 2, 2, 2 $\rightarrow Q_1 = \frac{1+2}{2} = 1.5$

upper 6: 3, 4, 4, 5, 6, 8 $\rightarrow Q_3 = \frac{4+5}{2} = 4.5$

$IQR = 4.5 - 1.5 = 3$

c.) Find the percentile of 5. 75th

entries less than 5: 9

percentile $= \frac{9}{12} \cdot 100 \approx 75 \rightarrow$ 75th percentile

Problem 20 Suppose the heights of a group of basketball players are bell-shaped with a mean of 79 inches and a standard deviation of 3.5 inches.

a.) About what percent lie between 72 and 86 inches? 95%

$79 \pm 2(3.5) \rightarrow$ within 2 SD: about 95%

b.) About what percent lie above 82.5 inches? 16%

outside 1 SD: $100\% - 68\% = 32\%$

upper tail: $\frac{32\%}{2} = 16\%$

c.) Find the z-score of 87.75 inches. Is it unusual? Explain. 2.5

$z = \frac{x - \mu}{\sigma} = \frac{87.75 - 79}{3.5} = 2.5$

$|z| > 2$: unusual