Statistics Unit 1 worksheet: Cumulative frequency graphs and percentiles (answer key)

ANSWERS AT A GLANCE

1. (a) Cumulative counts 4, 13, 28, 40, 47, 50; cumulative percents 8%, 26%, 56%, 80%, 94%, 100% at 4, 8, 12, 16, 20, 24 minutes. (b) (0, 0), (4, 4), (8, 13), (12, 28), (16, 40), (20, 47), (24, 50). (c) The segment from 8 to 12 minutes, slope 3.75 calls per minute. (d) 22 calls, exact.
2. (a) About 37.0 percent. (b) About 90.7 percent. (c) About 29 hailstones between 19 and 29 mm, and about 5 above 29 mm.
3. (a) About 267.9 minutes. (b) $Q_1 \approx 242.1$, $Q_3 \approx 297.6$, interquartile range about 55.5 minutes: the middle half of the runners finished within about 55.5 minutes of each other.
4. (a) 12, 36, 36, 24, 12 trees in 10 to 20, 20 to 30, 30 to 40, 40 to 50, 50 to 60 cm. (b) About the 80th percentile. (c) About 36.7 cm.
5. (a) (210, 12) and (330, 184) disappear; the other points stay at the same heights. (b) About 206.1 minutes. (c) About 10 runners. (d) The merged class spreads 46 runners evenly over 180 to 240, but only 12 of them finished before 210; trust the six-class estimate, about 217.1 minutes.
6. (a) No. Both medians lie in 10 to 20, but the graph cannot say where: if all 18 middle values were 11, the median would be 11; if all were 19, it would be 19. (b) No; it says only that the largest value is at least 20 and less than 30. (c) No; the height at 20 counts values less than 20. It is first counted at (30, 30).

WORKED SOLUTIONS

1. A help desk logged the lengths of 50 phone calls, in minutes. A call on a class boundary goes in the class to its right.

Call length (minutes)	0 to 4	4 to 8	8 to 12	12 to 16	16 to 20	20 to 24
Count	4	9	15	12	7	3

(a) Make a table of the cumulative frequency and the cumulative percent at each upper boundary.

(b) List every point of the cumulative frequency graph, and sketch the graph.

(c) Which segment of the graph is steepest, and what is its slope, with units?

(d) How many calls lasted 12 minutes or more? Is your answer exact or an estimate?

ANSWER

(a) Cumulative counts 4, 13, 28, 40, 47, 50; cumulative percents 8%, 26%, 56%, 80%, 94%, 100% at 4, 8, 12, 16, 20, 24 minutes. (b) (0, 0), (4, 4), (8, 13), (12, 28), (16, 40), (20, 47), (24, 50). (c) The segment from 8 to 12 minutes, slope 3.75 calls per minute. (d) 22 calls, exact.

SOLUTION

(a) Each running total adds the next count; each percent divides by 50:

Upper boundary	4	8	12	16	20	24
Cumulative count	4	13	28	40	47	50
Cumulative percent	8%	26%	56%	80%	94%	100%

The last total is 50, the whole data set, which checks the adding.

(b) The graph starts at (0, 0), since no call is shorter than 0 minutes, and then has one point above each upper boundary: (0, 0), (4, 4), (8, 13), (12, 28), (16, 40), (20, 47), (24, 50), joined by straight segments.

(c) Every class is 4 minutes wide, so the steepest segment belongs to the largest count, 15 calls in 8 to 12 minutes: a slope of $\frac{15}{4} = 3.75$ calls per minute.

(d) 12 is a boundary, and the graph's height there says exactly 28 calls were shorter than 12 minutes. The other 50 − 28 = 22 calls, 44%, lasted 12 minutes or more. No estimate is needed at a boundary.

Free sample from math.wiki for classroom use. Do not post. math.wiki

2. After a hailstorm, a weather office measured the diameters of 54 hailstones, in millimeters. A diameter on a class boundary goes in the class to its right.

Diameter (mm)	10 to 15	15 to 20	20 to 25	25 to 30	30 to 35
Count	4	20	13	15	2

Use the cumulative frequency graph of these counts, with straight segments between its points.

(a) Estimate the percentile rank of a diameter of 19 mm.

(b) Estimate the percentile rank of a diameter of 29 mm.

(c) Estimate how many hailstones measured between 19 and 29 mm, and how many measured more than 29 mm.

Give percentile ranks to the nearest tenth of a percent.

ANSWER

(a) About 37.0 percent. (b) About 90.7 percent. (c) About 29 hailstones between 19 and 29 mm, and about 5 above 29 mm.

SOLUTION

The running totals are 4 at 15 mm, 24 at 20 mm and 37 at 25 mm, out of 54.

(a) 19 mm is $\frac{4}{5}$ of the way through the 15 to 20 class, which holds 20 hailstones above the 4 below 15 mm. Spreading them evenly:

$$4 + \frac{4}{5} \cdot 20 = 20 \text{ hailstones}, \qquad \frac{20}{54} \approx 37.0\%.$$

(b) 29 mm is $\frac{4}{5}$ of the way through the 25 to 30 class, which holds 15 hailstones above the 37 below 25 mm:

$$37 + \frac{4}{5} \cdot 15 = 49 \text{ hailstones}, \qquad \frac{49}{54} \approx 90.7\%.$$

(c) A count between two values is the difference of two readings: 49 − 20 = 29 hailstones. Above 29 mm lie about 54 − 49 = 5. Both are estimates, since they rest on the evenly spread assumption inside a class.

3. The finishing times of the 200 runners in a city marathon, in minutes, are grouped below. A time on a class boundary goes in the class to its right.

Time (minutes)	180 to 210	210 to 240	240 to 270	270 to 300	300 to 330	330 to 360
Runners	12	34	58	50	30	16

(a) Estimate the median finishing time.

(b) Estimate Q_1, Q_3 and the interquartile range, and say what the interquartile range tells you about the runners.

Give each time to the nearest tenth of a minute.

ANSWER

(a) About 267.9 minutes. (b) $Q_1 \approx 242.1$, $Q_3 \approx 297.6$, interquartile range about 55.5 minutes: the middle half of the runners finished within about 55.5 minutes of each other.

SOLUTION

The running totals are 12, 46, 104, 154, 184 and 200 at 210, 240, 270, 300, 330 and 360 minutes.

(a) The median has target $0.50 \cdot 200 = 100$ runners, between 46 (at 240) and 104 (at 270), so it is 54 runners into the 58 of the 240 to 270 class:

Free sample from math.wiki for classroom use. Do not post. math.wiki

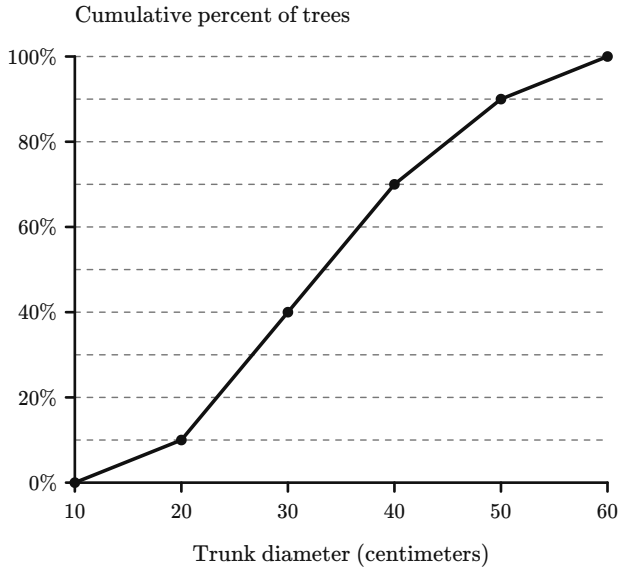
$$M \approx 240 + \frac{54}{58} \cdot 30 \approx 267.9 \text{ minutes.}$$

(b) Q_1 has target $0.25 \cdot 200 = 50$, between 46 and 104: 4 runners into the 58 of the 240 to 270 class. Q_3 has target 150, between 104 and 154: 46 runners into the 50 of the 270 to 300 class:

$$Q_1 \approx 240 + \frac{4}{58} \cdot 30 \approx 242.1,$$
$$Q_3 \approx 270 + \frac{46}{50} \cdot 30 = 297.6.$$

The interquartile range is about 297.6 – 242.07 ≈ 55.5 minutes: the middle half of the runners finished within a span of about 55.5 minutes.

4. A forest survey measured the trunk diameters of 120 trees in one plot. Its report shows them only as the cumulative relative frequency graph below.



(a) Find the number of trees in each 10 cm class.

(b) Estimate the percentile rank of a trunk 45 cm across.

(c) Estimate the diameter that 60 percent of the trunks are narrower than, to the nearest tenth of a centimeter.

ANSWER

(a) 12, 36, 36, 24, 12 trees in 10 to 20, 20 to 30, 30 to 40, 40 to 50, 50 to 60 cm. (b) About the 80th percentile. (c) About 36.7 cm.

SOLUTION

(a) The graph passes through 0%, 10%, 40%, 70%, 90% and 100% at 10, 20, 30, 40, 50 and 60 cm. The climb across each class is its share, and each share of 120 is a count:

$$10\%,\ 30\%,\ 30\%,\ 20\%,\ 10\% \text{ of } 120$$
$$= 12,\ 36,\ 36,\ 24,\ 12 \text{ trees,}$$

which add to 120.

Free sample from math.wiki for classroom use. Do not post. math.wiki

(b) 45 cm is halfway through the class that climbs from 70% to 90%, so the reading is $70 + \frac{5}{10} \cdot 20 = 80$ percent: about the 80th percentile.

(c) 60% lies between 40% at 30 cm and 70% at 40 cm, 20 points into a climb of 30:

$$30 + \frac{20}{30} \cdot 10 \approx 36.7 \text{ cm}.$$

5. A marathon's 200 finishing times were first grouped into six 30 minute classes from 180 to 360 minutes, with counts 12, 34, 58, 50, 30 and 16. From that table the 10th percentile is about 217.1 minutes. The race website regroups the same times into four classes:

Time (minutes)	180 to 240	240 to 270	270 to 300	300 to 360
Runners	46	58	50	46

(a) Which points of the six-class cumulative frequency graph disappear on the four-class graph? Do the other points move?

(b) Estimate the 10th percentile from the four-class table, to the nearest tenth of a minute.

(c) The race gives a medal to every runner under your time in (b). Use the six-class table to estimate how many runners that is.

(d) Explain why the answer to (c) is far from 20 runners, and which estimate of the 10th percentile you trust more.

ANSWER

(a) (210, 12) and (330, 184) disappear; the other points stay at the same heights. (b) About 206.1 minutes. (c) About 10 runners. (d) The merged class spreads 46 runners evenly over 180 to 240, but only 12 of them finished before 210; trust the six-class estimate, about 217.1 minutes.

SOLUTION

(a) The running totals of the four-class table are 46, 104, 154 and 200 at 240, 270, 300 and 360 minutes, the same heights as on the six-class graph at those boundaries. Merging removes the boundaries at 210 and 330 minutes, so the points (210, 12) and (330, 184) disappear and the graph runs straight from (180, 0) to (240, 46) and from (300, 154) to (360, 200).

(b) The target is $0.10 \cdot 200 = 20$ runners, inside the first class, which holds 46 over 60 minutes:

$$P_{10} \approx 180 + \frac{20}{46} \cdot 60 \approx 206.1 \text{ minutes}.$$

(c) On the six-class graph 206.1 minutes lies in the 180 to 210 class, 26.1 minutes into its 30, which holds 12 runners:

$$\frac{26.1}{30} \cdot 12 \approx 10.4,$$

so about 10 runners get a medal.

(d) The four-class graph assumes the 46 runners of 180 to 240 minutes are spread evenly, which puts 23 of them before 210 minutes. The finer table shows only 12 there. So the coarse graph climbs too fast at first and reaches 20 runners too early. The six-class estimate, about 217.1 minutes, rests on narrower classes, which leave the evenly spread assumption less room to be wrong.

6. Two data sets of 30 values each are both grouped into the classes 0 to 10, 10 to 20 and 20 to 30, with counts 6, 18 and 6 in each data set. A value on a boundary goes in the class to its right.

Free sample from math.wiki for classroom use. Do not post. math.wiki

(a) The two data sets have the same cumulative frequency graph. Must they have the same median? Explain, with an example.

(b) Can the graph tell you the largest value in either data set? Explain.

(c) One data set contains a value of exactly 20. Is it counted in the height of the graph at 20? At which plotted point is it first counted?

ANSWER

(a) No. Both medians lie in 10 to 20, but the graph cannot say where: if all 18 middle values were 11, the median would be 11; if all were 19, it would be 19. (b) No; it says only that the largest value is at least 20 and less than 30. (c) No; the height at 20 counts values less than 20. It is first counted at (30, 30).

SOLUTION

(a) The graph records only the running totals 6, 24 and 30 at the boundaries 10, 20 and 30. The 15th and 16th values are both in 10 to 20, since $6 < 15$ and $16 \leq 24$, so each median lies in that class, but the graph says nothing about where the 18 values sit inside it. If they are all 11, the median is 11; if they are all 19, the median is 19; both data sets have the same table and so the same graph. The graph's estimate, $10 + \frac{9}{18} \cdot 10 = 15$, is the same for both and right for neither.

(b) No. Above 30 the graph is flat at 30, and the climb from (20, 24) to (30, 30) only shows that 6 values lie in 20 to 30. The largest value is at least 20 and less than 30; the table does not say more, so neither does the graph.

(c) No. The height at 20 is the number of values less than 20, and a value of exactly 20 goes in the class 20 to 30. It is first counted in the running total at the upper boundary of that class, the point (30, 30).

Free sample from math.wiki for classroom use. Do not post. math.wiki