Statistics Unit 1 quiz 1: Types of data: categorical, numerical, discrete, continuous, Making a histogram and choosing the bin width, Describing a distribution (answer key)

7 questions, 21 points

ANSWERS AT A GLANCE

1. (a), (b) and (d) are wrong; (c) and (e) are right. (a) categorical, with an order (b) categorical: a label (d) numerical and continuous: height is measured.
2. (a) mean house number 11; mean people 4 (b) mean house number 6; mean people 4 (c) The mean number of people: it stayed 4 when nothing about the houses changed, while the mean house number changed with the labels.
3. (a) at least 86.5 cm and less than 87.5 cm (b) 0.8 cm: yes; 1.2 cm: no (c) Yes.
4. (a) The mean and the MAD: the data is roughly symmetric with no outliers. (b) The median and the IQR: the mean is far above the median, so the data is probably skewed right or has a high outlier. (c) 960 dollars, from the mean: $15 \times 64 = 960$. The median does not depend on the largest amounts, so it cannot give the total.
5. (a) Both are right: median 13 days, $Q_1 = 11.5$, $Q_3 = 14$, IQR 2.5 days. (b) "Skewed right" is wrong: the values pile up against the ceiling of 14 and the tail points left, so the data is skewed left. "No outliers" is wrong: the lower fence is 7.75 days, so 6 is an outlier. (c) "Jack asked 13 classmates on how many of the 14 days of spring break they read for at least 20 minutes. The counts are skewed left. The median is 13 days, and the middle half of the students read on 11.5 to 14 days, an IQR of 2.5 days. Five students read on all 14 days. One student, who read on 6 days, is an outlier. A typical student read on about 13 of the 14 days."
6. (a) 10 to 20: 5; 20 to 30: 13; 30 to 40: 9; 40 to 50: 3.
   (b) Yes: 5 to 15: 2; 15 to 25: 9; 25 to 35: 12; 35 to 45: 6; 45 to 55: 1.
   (c) No: a bin 2 m wide such as 10 to 12 is only part of Maya's bin 10 to 15, and the table does not say which of those 2 throws are below 12. A bin cannot be split once the separate values are gone.
7. (a) The bars for 4 to 6, 6 to 8 and 8 to 10 are wrong: Liam put the values on edges (6.0, 6.0 and 8.0) in the bar to their left.
   (b) 4 to 6: 3; 6 to 8: 6; 8 to 10: 4; 10 to 12: 1; 12 to 14: 1.

WORKED SOLUTIONS

1. Jin sorts the variables his gymnastics club records for each member: (3 points)

(a) level (bronze, silver or gold): numerical, because the levels come in an order;

(b) membership number: numerical and discrete, because it is a whole number;

(c) number of classes attended this month: numerical and discrete;

(d) height, recorded to the nearest centimeter: numerical and discrete, because every recorded value is a whole number;

(e) longest time holding a handstand, in seconds: numerical and continuous.

Find each of Jin's mistakes. For each one, give the right type and the reason.

ANSWER

(a), (b) and (d) are wrong; (c) and (e) are right. (a) categorical, with an order (b) categorical: a label (d) numerical and continuous: height is measured.

SOLUTION

(a) Wrong. A level names a group, so it is categorical. The levels have an order, but an order alone is not an amount: it says which level comes first, not by how much.

(b) Wrong. A membership number names a member. Adding or averaging membership numbers gives no amount of anything, so it is categorical, even though it is a whole number.

(c) Right: classes are counted.

(d) Wrong. Height is measured, so it is numerical and continuous. The whole numbers come from rounding to the nearest centimeter, not from the heights themselves.

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

(e) Right: time is measured.

SCORING

1 point for each of (a), (b) and (d) marked wrong with the right type and a reason. Take off 1 point (to no less than 0) if (c) or (e) is marked wrong.

2. Five houses on Elm Road are numbered 3, 7, 11, 15 and 19. They have 4, 2, 5, 3 and 6 people living in them, in that order. (3 points)

(a) Find the mean house number and the mean number of people per house.

(b) The town renumbers the same five houses 2, 4, 6, 8 and 10. Nobody moves. Find both means again.

(c) Use your answers to explain which mean describes the houses.

ANSWER

(a) mean house number 11; mean people 4 (b) mean house number 6; mean people 4 (c) The mean number of people: it stayed 4 when nothing about the houses changed, while the mean house number changed with the labels.

SOLUTION

(a) House numbers: $3 + 7 + 11 + 15 + 19 = 55$ and $55 \div 5 = 11$. People: $4 + 2 + 5 + 3 + 6 = 20$ and $20 \div 5 = 4$.

(b) House numbers: $2 + 4 + 6 + 8 + 10 = 30$ and $30 \div 5 = 6$. People: still $20 \div 5 = 4$.

(c) The houses and the people in them are exactly the same as before, yet the mean house number moved from 11 to 6. A summary that changes when nothing about the houses changed cannot describe the houses: a house number is a label, so it is categorical. The mean of 4 people is an amount: 20 people shared equally among 5 houses.

SCORING

1 point for both means in (a); 1 point for both means in (b); 1 point for explaining that the house-number mean changed only because the labels changed, so only the mean number of people describes the houses.

3. A lumber yard records the length of each board to the nearest centimeter. Board A is recorded as 87 cm. (3 points)

(a) What could the true length of board A be?

(b) Board B is also recorded as 87 cm. Could the true lengths of boards A and B differ by 0.8 cm? Could they differ by 1.2 cm? Explain.

(c) Board C is recorded as 88 cm. Is board C surely longer than board A? Explain.

ANSWER

(a) at least 86.5 cm and less than 87.5 cm (b) 0.8 cm: yes; 1.2 cm: no (c) Yes.

SOLUTION

(a) A length rounds to 87 cm when it is at least $87 - 0.5 = 86.5$ cm and less than $87 + 0.5 = 87.5$ cm.

(b) Both boards lie from 86.5 cm up to, but not including, 87.5 cm, a stretch 1 cm long. Board A could be 86.6 cm and board B 87.4 cm, which differ by $87.4 - 86.6 = 0.8$ cm, so yes. Two lengths in that stretch always differ by less than 1 cm, so they cannot differ by 1.2 cm.

(c) Board C is at least $88 - 0.5 = 87.5$ cm long, and board A is less than 87.5 cm. So board C is surely the longer, even though they might differ by only a tiny amount.

SCORING

1 point for the range in (a) with the right ends; 1 point for both answers in (b) with a reason using the 1 cm stretch; 1 point for (c) yes, comparing 87.5 cm with both ranges.

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

4. For (a) and (b), say which center and spread a description should give, and why. (3 points)

(a) The amounts of juice in 18 juice boxes from one carton are roughly symmetric, with a mean of 201 mL, a median of 200.5 mL and no outliers.

(b) The money 15 kids earned pet-sitting last month has a mean of 64 dollars and a median of 40 dollars.

(c) The same 15 kids put all their pet-sitting money together for a class trip. How much do they have? Say which center you used, and why the other could not give it.

ANSWER

(a) The mean and the MAD: the data is roughly symmetric with no outliers. (b) The median and the IQR: the mean is far above the median, so the data is probably skewed right or has a high outlier. (c) 960 dollars, from the mean: 15 × 64 = 960. The median does not depend on the largest amounts, so it cannot give the total.

SOLUTION

(a) For roughly symmetric data with no outliers the mean is a fair center, and the MAD, which measures distances from the mean, goes with it. (The median with the IQR would also be honest here, but the mean and the MAD are usual, since they use every value.)

(b) The mean is 64 − 40 = 24 dollars above the median, so a few kids who earned a lot pull the mean up: the data is probably skewed right, or has a high outlier. Use the resistant pair, the median with the IQR, which stays with the bulk of the kids.

(c) The mean is the total shared out equally, so the total is 15 × 64 = 960 dollars. The median is only the middle amount: changing what the top seven kids earned would not change it, so it cannot tell the total. When the question needs a total, the mean is needed whatever the shape.

SCORING

1 point: (a) mean with MAD, because the data is symmetric with no outliers. 1 point: (b) median with IQR, because the mean well above the median points to a right skew or a high outlier. 1 point: (c) 960 dollars from 15 × 64, naming the mean and saying the median does not carry the total.

5. Jack asked 13 classmates on how many of the 14 days of spring break they read for at least 20 minutes: 14, 14, 13, 14, 12, 13, 11, 14, 6, 13, 10, 12, 14. Maya wrote:

"The 13 students read on 6 to 14 days. The data is skewed right, because most of the values are at the high end. The median is 13 days and the IQR is 2.5 days. There are no outliers." (3 points)

(a) Check Maya's median and IQR.

(b) Which of her claims are wrong? Explain each.

(c) Write an improved description.

ANSWER

(a) Both are right: median 13 days, $Q_1 = 11.5$, $Q_3 = 14$, IQR 2.5 days. (b) "Skewed right" is wrong: the values pile up against the ceiling of 14 and the tail points left, so the data is skewed left. "No outliers" is wrong: the lower fence is 7.75 days, so 6 is an outlier. (c) "Jack asked 13 classmates on how many of the 14 days of spring break they read for at least 20 minutes. The counts are skewed left. The median is 13 days, and the middle half of the students read on 11.5 to 14 days, an IQR of 2.5 days. Five students read on all 14 days. One student, who read on 6 days, is an outlier. A typical student read on about 13 of the 14 days."

SOLUTION

(a) Sorted: 6, 10, 11, 12, 12, 13, 13, 13, 14, 14, 14, 14, 14. With 13 values the median is the 7th, 13 days. The lower six, 6, 10, 11, 12, 12, 13, give $Q_1 = \frac{11 + 12}{2} = 11.5$ and the upper six, 13, 14, 14, 14, 14, 14,

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

give $Q_3 = 14$. The IQR is 14 − 11.5 = 2.5 days. Both numbers check.

(b) Shape: a skew is named for the side of the long tail, not the side where the values crowd. The counts pile up at 13 and 14, and no one can read on more than 14 days, while a few values trail off to the left as far as 6. Skewed left. The mean, $\frac{160}{13} \approx 12.3$ days, below the median, agrees.

Outliers: 1.5 × 2.5 = 3.75, so the lower fence is 11.5 − 3.75 = 7.75 days. The 6 is below it, so it is an outlier, and "no outliers" is false. Because of it, the range "6 to 14 days" stretches over a value that is not typical, without saying so.

(c) "Jack asked 13 classmates on how many of the 14 days of spring break they read for at least 20 minutes. The counts are skewed left. The median is 13 days, and the middle half of the students read on 11.5 to 14 days, an IQR of 2.5 days. Five students read on all 14 days. One student, who read on only 6 days, is an outlier. A typical student read on about 13 of the 14 days."

SCORING

1 point: (a) median 13 and IQR 2.5 confirmed from the sorted list. 1 point: (b) "skewed right" corrected to skewed left with the tail as the reason, and "no outliers" refuted with the lower fence 7.75. 1 point: (c) a description with context and units, skewed left, median with IQR, the 6 named as an outlier, and a last sentence answering the question.

6. Maya measured how far each of the 30 students in her gym class threw a softball, in meters. She kept only this frequency table, made the lesson's way (first edge the largest multiple of 5 at or below the shortest throw, a value on an edge going in the bar to its right):

Distance (m)	Count
10 to 15	2
15 to 20	3
20 to 25	6
25 to 30	7
30 to 35	5
35 to 40	4
40 to 45	2
45 to 50	1

(a) Make the frequency table for bins 10 m wide starting at 10. (b) Omar wants bins 10 m wide starting at 5 instead. Can he get the counts from Maya's table alone? If so, give them; if not, explain. (c) Can anyone make a table with bins 2 m wide from Maya's table alone? Explain. (3 points)

ANSWER

(a) 10 to 20: 5; 20 to 30: 13; 30 to 40: 9; 40 to 50: 3.

(b) Yes: 5 to 15: 2; 15 to 25: 9; 25 to 35: 12; 35 to 45: 6; 45 to 55: 1.

(c) No: a bin 2 m wide such as 10 to 12 is only part of Maya's bin 10 to 15, and the table does not say which of those 2 throws are below 12. A bin cannot be split once the separate values are gone.

SOLUTION

Each new bin can be found by adding counts only if it is made of whole bins of the old table, that is, if its edges are old edges.

(a) "10 to 20" is "10 to 15" and "15 to 20": 2 + 3 = 5. Likewise 6 + 7 = 13, 5 + 4 = 9 and 2 + 1 = 3. Check: 5 + 13 + 9 + 3 = 30.

(b) The new edges are 5, 15, 25, 35, 45, 55. The inner ones, 15, 25, 35 and 45, are all edges of Maya's table, so each new bin is made of whole old bins. No throw is below 10 (the first edge is at or below the shortest

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

throw and the first bin starts at 10), so "5 to 15" holds only the 2 throws of "10 to 15". No throw is 50 or more (bins stop at the one holding the longest), so "45 to 55" holds only the 1 throw of "45 to 50". The others: 3 + 6 = 9, 7 + 5 = 12, 4 + 2 = 6. Check: 2 + 9 + 12 + 6 + 1 = 30.

(c) Bins 2 m wide have edges such as 12 and 14, which fall inside Maya's bin 10 to 15. Its 2 throws could be anywhere from 10 up to 15, so the table cannot say how many lie from 10 to 12. Neighboring bins can be merged by adding counts, but a bin cannot be split.

SCORING

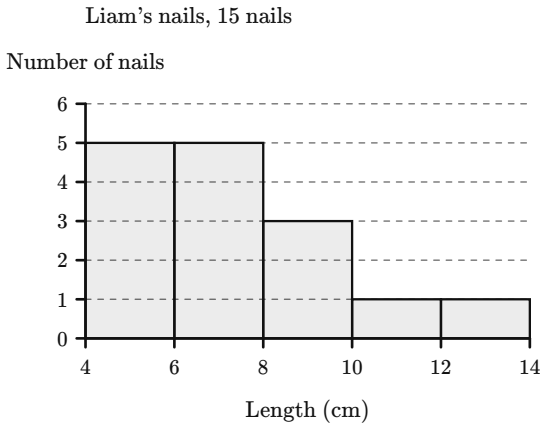
1 point: (a) counts 5, 13, 9, 3. 1 point: (b) yes, with counts 2, 9, 12, 6, 1 (the first and last bins each taking one old bin). 1 point: (c) no, because the new edges fall inside old bins and a bin cannot be split.

7. Liam measured the 15 nails in a jar in his grandfather's workshop, in centimeters:

4.5, 6.0, 7.2, 5.8, 8.0, 9.4, 6.6, 10.1, 7.9, 8.5, 6.0, 12.3, 9.0, 7.5, 5.1

He drew the histogram below with bins 2 cm wide from 4 cm. In this course a value on an edge goes in the bar to its right.

(a) Which bars are the wrong height, and what mistake did Liam make? (b) Give the correct height of every bar. (3 points)



ANSWER

(a) The bars for 4 to 6, 6 to 8 and 8 to 10 are wrong: Liam put the values on edges (6.0, 6.0 and 8.0) in the bar to their left.

(b) 4 to 6: 3; 6 to 8: 6; 8 to 10: 4; 10 to 12: 1; 12 to 14: 1.

SOLUTION

Sorted: 4.5, 5.1, 5.8, 6.0, 6.0, 6.6, 7.2, 7.5, 7.9, 8.0, 8.5, 9.0, 9.4, 10.1, 12.3.

With the edge rule: "4 to 6" holds 4.5, 5.1, 5.8 (3); "6 to 8" holds 6.0, 6.0, 6.6, 7.2, 7.5, 7.9 (6); "8 to 10" holds 8.0, 8.5, 9.0, 9.4 (4); "10 to 12" holds 10.1 (1); "12 to 14" holds 12.3 (1). Check: 3 + 6 + 4 + 1 + 1 = 15.

(a) Liam's bars are 5, 5, 3, 1, 1. His first bar has the two 6.0 nails added, his second has lost them but gained the 8.0 nail, and his third has lost the 8.0 nail. So he counted each edge value in the bar to its left. His total is still 15, which is why the mistake is easy to miss.

(b) Correct heights: 3, 6, 4, 1, 1.

SCORING

1 point: names the three wrong bars (4 to 6, 6 to 8, 8 to 10). 1 point: says the mistake is putting the edge values 6.0, 6.0 and 8.0 in the bar to their left. 1 point: correct heights 3, 6, 4, 1, 1.

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