Statistics Unit 1 test: Exploring one-variable data (answer key)

13 questions, 54 points

ANSWERS AT A GLANCE

1. (a) Each medium parcel is at least 2 kg and the heavy one at least 5 kg, so the total is at least 10 + 5 = 15 kg, and the light parcels add more. (b) 1.2 kg: no, the mean is more than 1.5 kg. 4 kg: yes, for example light parcels of 1.5 kg, medium of 4.8 kg and a heavy one of 10 kg.
2. (a) 10 to 20: 1; 20 to 30: 3; 30 to 40: 4; 40 to 50: 2; 50 to 60: 2.
   (b) 10 to 20: 3; 20 to 30: 4; 30 to 40: 2; 40 to 50: 2; 50 to 60: 1.
   (c) Naomi's tallest bar is 30 to 40, Daniel's is 20 to 30. With 7 of the 12 values on edges, the rule changes the counts and even the tallest bar, so a histogram must state its edge rule.
3. (a) Mean 27.5 m, median 26.5 m. (b) 18, 23, 26.5, 32, 40 m. (c) IQR 9 m; 6 throws, $\frac{1}{2}$ of the class.
4. (a) Class B scored higher on average, by 6 points. Class B's scores were much more spread out: typically about 12 points from its mean, against about 5 for Class A. (b) Jasmine: 2.4 standard deviations above her class mean; Andre: 1. Jasmine did better relative to her class. (c) A standard deviation is a typical distance from the mean, not the lowest or highest value.
5. (a) $Q_1 = 7$, $Q_3 = 9$, IQR = 2, fences 4 and 12 cm; 16 cm is the only outlier. (b) $Q_1 = 70$, $Q_3 = 90$, IQR = 20, fences 40 and 120 mm; the same pod (160 mm) is the only outlier.
6. (a) Median 3 hours, $Q_1 = 2$, $Q_3 = 5$, mean about 3.9 hours. (b) Fences −2.5 and 9.5 hours: the 16 is an outlier, with a gap from 7 to 15 hours. (c) "Isaac asked 15 classmates how many hours they spent playing video games last week. The hours are skewed right. The median is 3 hours, and the middle half of the classmates played 2 to 5 hours, an IQR of 3 hours. One classmate played 16 hours, an outlier, with a gap from 7 to 15 hours. A typical classmate played video games for about 3 hours last week." (d) The mean would drop from about 3.9 to about 3.1 hours, a change of about 0.9 hours; the median would stay 3 hours. The median is resistant, so it describes the bulk of the class whether or not the outlier is there.
7. (a) $Q_1 = 40$, $Q_3 = 45$, IQR 5, fences 32.5 and 52.5, so 10 is an outlier. (b) All 11: mean 40, median 43. Without the 10: mean 43, median 43. (c) The median, 43 sit-ups.
8. (a) Median 2, $Q_1 = 1$, $Q_3 = 3$, IQR = 2 (b) $s = \sqrt{\frac{20}{19}} \approx 1.03$ people (c) It shows 1, the divide-by-20 figure; the cars are a sample of the day's cars, so the divisor should be 19.
9. (a) $z = -2$ and $z = 1$ (b) about 18.5% (c) 31 km (d) $\mu = 44$ km
10. (a) $N(49, 8.8)$ (b) above 40.2 and 57.8 words per minute (c) 56 words per minute; 2 standard deviations above the mean before, and 66.6 is 2 above after. (d) about 0.023 faster than 66.6, and about 0.023 slower than 31.4.
11. (a) about 0.0668 (b) about 0.6247; about 750 buyers
12. (a) Field A: roughly symmetric with one peak at 150 g. Field B: skewed right, most potatoes from 100 to 150 g and a long tail out to 290 g. (b) Field B: $Q_1 = 120$, $Q_3 = 170$, IQR 50, so the cut-off is 245 g: 290 g is an outlier and 230 g is not. (c) Field A: mean 150 g, median 150 g. Field B: mean 150 g, median 130 g. Compare medians: a typical Field A potato is 20 g heavier. (d) No. The shift of 20 g is half of Field A's IQR of 40 g and less than half of Field B's 50 g, and the middle halves, 130 to 170 g and 120 to 170 g, almost coincide.
13. (a) Choir A about 33.6 years, Choir B about 45 years. (b) Choir A about 16.6 years ($Q_1 \approx 26.4$, $Q_3 \approx 43$); Choir B about 17.1 years ($Q_1 \approx 35$, $Q_3 \approx 52.1$). (c) Choir A 3 members, Choir B 12 members. (d) Choir B's members tend to be older (median about 11 years higher, B's graph lies on or below A's at every age), while the middle halves of the two choirs have about the same spread.

WORKED SOLUTIONS

1. A post office weighs 10 parcels and sorts them into classes: light (under 2 kg), medium (2 kg up to 5 kg) and heavy (5 kg and over). The clerk writes down only the classes: 4 light, 5 medium and 1 heavy. (3 points)

   (a) Explain why the total mass of the 10 parcels must be more than 15 kg.

   (b) Could the mean mass of the parcels be 1.2 kg? Could it be 4 kg? Explain each answer.

   ANSWER

   (a) Each medium parcel is at least 2 kg and the heavy one at least 5 kg, so the total is at least 10 + 5 = 15 kg, and the light parcels add more. (b) 1.2 kg: no, the mean is more than 1.5 kg. 4 kg: yes, for example light parcels of 1.5 kg, medium of 4.8 kg and a heavy one of 10 kg.

   SOLUTION

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

(a) The five medium parcels are each at least 2 kg, so together at least 5 × 2 = 10 kg. The heavy parcel is at least 5 kg. That is at least 10 + 5 = 15 kg, and the four light parcels each weigh more than 0 kg, so the total is more than 15 kg.

(b) The mean is the total shared equally among 10 parcels. The total is more than 15 kg, so the mean is more than 15 ÷ 10 = 1.5 kg. It cannot be 1.2 kg.

It can be 4 kg: the total would be 4 × 10 = 40 kg. For example, four light parcels of 1.5 kg (6 kg), five medium parcels of 4.8 kg (24 kg) and a heavy parcel of 10 kg give 6 + 24 + 10 = 40 kg, and each parcel is in its class. The class counts alone cannot tell the mean, only that it is more than 1.5 kg.

SCORING

1 point for (a) using the least mass of each medium and heavy parcel to reach 15 kg; 1 point for (b) 1.2 kg impossible because the mean is more than 1.5 kg; 1 point for (b) 4 kg possible with a set of masses that fits the classes and totals 40 kg.

2. Naomi recorded how long she spent sketching on 12 days, to the nearest 5 minutes:

30, 45, 20, 55, 15, 30, 40, 25, 50, 30, 35, 20

(a) Make a frequency table with bins 10 minutes wide from 10 to 60, using this course's rule: a value on an edge goes in the bar to its right. (b) Her brother Daniel uses the other rule: a value on an edge goes in the bar to its left, so "10 to 20" holds 15 and 20. Make his table with the same bins. (c) Which bar is tallest in each table? Explain what this shows about the edge rule. (3 points)

ANSWER

(a) 10 to 20: 1; 20 to 30: 3; 30 to 40: 4; 40 to 50: 2; 50 to 60: 2.

(b) 10 to 20: 3; 20 to 30: 4; 30 to 40: 2; 40 to 50: 2; 50 to 60: 1.

(c) Naomi's tallest bar is 30 to 40, Daniel's is 20 to 30. With 7 of the 12 values on edges, the rule changes the counts and even the tallest bar, so a histogram must state its edge rule.

SOLUTION

Sorted: 15, 20, 20, 25, 30, 30, 30, 35, 40, 45, 50, 55.

(a) Naomi's rule, $a \le x < b$: "10 to 20" holds 15 (1); "20 to 30" holds 20, 20, 25 (3); "30 to 40" holds 30, 30, 30, 35 (4); "40 to 50" holds 40, 45 (2); "50 to 60" holds 50, 55 (2). Check: 12.

(b) Daniel's rule, $a < x \le b$: "10 to 20" holds 15, 20, 20 (3); "20 to 30" holds 25, 30, 30, 30 (4); "30 to 40" holds 35, 40 (2); "40 to 50" holds 45, 50 (2); "50 to 60" holds 55 (1). Check: 12.

(c) The values 20, 20, 30, 30, 30, 40, 50 sit on edges: 7 of the 12. Each of them moves one bar to the left under Daniel's rule, so the tallest bar moves from 30 to 40 to 20 to 30. Both tables are honest counts, but a reader who did not know the rule could not tell which sketching times the bars hold. That is why the rule has to be stated, and why data rounded to the width's multiples land on edges so often.

SCORING

1 point: (a) counts 1, 3, 4, 2, 2. 1 point: (b) counts 3, 4, 2, 2, 1. 1 point: (c) tallest bars 30 to 40 and 20 to 30, with the reason that the many edge values change bins, so the rule must be stated.

3. The 12 students in Arjun's gym class each threw a softball once. The distances, in meters, were

18, 25, 31, 22, 27, 35, 20, 29, 24, 33, 26, 40. (4 points)

(a) Find the mean and the median.

(b) Find the five-number summary.

(c) Find the IQR. How many throws are longer than Q_1 and shorter than Q_3, and what fraction of the class is that?

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

ANSWER

(a) Mean 27.5 m, median 26.5 m. (b) 18, 23, 26.5, 32, 40 m. (c) IQR 9 m; 6 throws, $\frac{1}{2}$ of the class.

SOLUTION

(a) Sum: 18 + 25 + 31 + 22 + 27 + 35 + 20 + 29 + 24 + 33 + 26 + 40 = 330, so the mean is $\frac{330}{12} = 27.5$ m. Sorted: 18, 20, 22, 24, 25, 26, 27, 29, 31, 33, 35, 40. The median is $\frac{26 + 27}{2} = 26.5$ m.

(b) Lower half 18, 20, 22, 24, 25, 26: $Q_1 = \frac{22 + 24}{2} = 23$. Upper half 27, 29, 31, 33, 35, 40: $Q_3 = \frac{31 + 33}{2} = 32$. The five-number summary is 18, 23, 26.5, 32, 40 m.

(c) IQR = 32 − 23 = 9 m. Between 23 and 32 lie 24, 25, 26, 27, 29, 31: 6 throws, $\frac{6}{12} = \frac{1}{2}$ of the class. Twelve is a multiple of 4 and no throws repeat, so the middle half holds exactly half the values.

SCORING

1 point: mean 27.5 and median 26.5. 1 point: $Q_1 = 23$ and $Q_3 = 32$ from halves of six. 1 point: five-number summary in order. 1 point: IQR 9, and 6 throws, one half of the class.

4. On the same test, Class A had mean 72 points with standard deviation 5 points, and Class B had mean 78 points with standard deviation 12 points. (4 points)

(a) Compare the two classes in two sentences.

(b) Jasmine in Class A scored 84 and Andre in Class B scored 90. Who did better relative to their own class? Show the counts of standard deviations.

(c) Andre says, "Class B's standard deviation is bigger, so the lowest score in Class B must be lower than the lowest score in Class A." Explain why the standard deviations cannot show this.

ANSWER

(a) Class B scored higher on average, by 6 points. Class B's scores were much more spread out: typically about 12 points from its mean, against about 5 for Class A. (b) Jasmine: 2.4 standard deviations above her class mean; Andre: 1. Jasmine did better relative to her class. (c) A standard deviation is a typical distance from the mean, not the lowest or highest value.

SOLUTION

(a) Class B did better on average: its scores centered on 78 points against Class A's 72, a gap of 6 points. Class B's scores were also more spread out: a Class B score typically sat about 12 points from its mean, more than twice Class A's 5 points.

(b) Jasmine: $\frac{84 - 72}{5} = \frac{12}{5} = 2.4$ standard deviations above her class mean. Andre: $\frac{90 - 78}{12} = \frac{12}{12} = 1$ standard deviation above his. Both are 12 points above their own class mean, but in Class A that is far more unusual, so Jasmine did better relative to her class.

(c) The standard deviation blends all the distances from the mean into one typical distance; it does not give the lowest value of either class, or say on which side of the mean the far values lie. Class B's spread could come mostly from scores far above 78, with no score below about 70, while Class A's scores cluster near 72 apart from one low score well under 70. To compare the lowest scores you need the scores themselves, or at least each class's minimum.

SCORING

1 point: (a) states both which center is higher and which class is more spread, in points. 2 points: (b) 2.4 and 1 shown (1 point), and Jasmine (1 point). 1 point: (c) says the standard deviation is a typical distance from the mean and does not fix the lowest value.

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

5. Leah measured 11 pea pods, in centimeters, and sorted the lengths: 5, 6, 7, 7, 8, 8, 9, 9, 9, 10, 16. (4 points)

(a) Find Q_1, Q_3, the IQR and the two fences, and name any outliers.

(b) Leah's partner writes every length in millimeters (1 cm = 10 mm). Without sorting the new list, give Q_1, Q_3, the IQR and the fences in millimeters, and say whether the same pods are outliers.

ANSWER

(a) $Q_1 = 7$, $Q_3 = 9$, IQR = 2, fences 4 and 12 cm; 16 cm is the only outlier. (b) $Q_1 = 70$, $Q_3 = 90$, IQR = 20, fences 40 and 120 mm; the same pod (160 mm) is the only outlier.

SOLUTION

(a) The median is the 6th value, 8. The lower half 5, 6, 7, 7, 8 has median $Q_1 = 7$; the upper half 9, 9, 9, 10, 16 has median $Q_3 = 9$. The IQR is 9 − 7 = 2. The fences are 7 − 1.5 × 2 = 4 and 9 + 1.5 × 2 = 12. Only 16 lies beyond a fence, so it is the one outlier.

(b) Millimeters are centimeters times 10, a rescale with $k = 10$. The quartiles are multiplied by 10: $Q_1 = 70$, $Q_3 = 90$. The IQR is multiplied by 10: 20. Each fence is built from the quartiles and the IQR, so it is multiplied by 10 too: 70 − 1.5 × 20 = 40 and 90 + 1.5 × 20 = 120 mm. Every length and both fences were multiplied by the same positive number, so no length crosses a fence: the pod of 160 mm is still the only outlier.

SCORING

1 point for $Q_1 = 7$ and $Q_3 = 9$; 1 point for the fences 4 and 12 and the outlier 16; 1 point for the quartiles, IQR and fences in millimeters; 1 point for saying the same pod is the only outlier, with a reason.

6. Isaac asked 15 classmates how many whole hours each one spent playing video games last week: 2, 0, 3, 5, 4, 1, 3, 6, 2, 4, 3, 16, 5, 2, 3. (5 points)

(a) Find the median, the quartiles, and the mean to the nearest tenth of an hour.

(b) Test for outliers, and name any gap.

(c) Write a description that answers "How much time did Isaac's classmates spend playing video games last week?"

(d) If the 16 were left out, about how much would the mean change, and how much would the median change? What does that show about the pair of measures you chose?

ANSWER

(a) Median 3 hours, $Q_1 = 2$, $Q_3 = 5$, mean about 3.9 hours. (b) Fences −2.5 and 9.5 hours: the 16 is an outlier, with a gap from 7 to 15 hours. (c) "Isaac asked 15 classmates how many hours they spent playing video games last week. The hours are skewed right. The median is 3 hours, and the middle half of the classmates played 2 to 5 hours, an IQR of 3 hours. One classmate played 16 hours, an outlier, with a gap from 7 to 15 hours. A typical classmate played video games for about 3 hours last week." (d) The mean would drop from about 3.9 to about 3.1 hours, a change of about 0.9 hours; the median would stay 3 hours. The median is resistant, so it describes the bulk of the class whether or not the outlier is there.

SOLUTION

(a) Sorted: 0, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5, 5, 6, 16. With 15 values the median is the 8th, 3 hours. The lower seven, 0, 1, 2, 2, 2, 3, 3, give $Q_1 = 2$, and the upper seven, 3, 4, 4, 5, 5, 6, 16, give $Q_3 = 5$. The total is 59, and $\frac{59}{15} \approx 3.93$, so the mean is about 3.9 hours.

(b) The IQR is 5 − 2 = 3 hours and 1.5 × 3 = 4.5. The fences are 2 − 4.5 = −2.5 and 5 + 4.5 = 9.5 hours. The 16 is beyond 9.5, so it is an outlier; nothing can be below −2.5. No classmate played 7 to 15 hours: a gap.

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

(c) Fourteen values sit between 0 and 6, and one stands far off to the right; the mean is above the median. Skewed right with an outlier, so the median with the IQR.

Description: "Isaac asked 15 classmates how many whole hours they spent playing video games last week. The hours are skewed right. The median is 3 hours, and the middle half of the classmates played 2 to 5 hours, an IQR of 3 hours. One classmate played 16 hours, an outlier, with a gap from 7 to 15 hours between that classmate and the rest. A typical classmate played video games for about 3 hours last week."

(d) Without the 16 the total is 59 − 16 = 43 for 14 classmates, and $\frac{43}{14} \approx 3.07$, so the mean falls from about 3.9 to about 3.1 hours, by about 0.9 hours. The 7th and 8th of the 14 remaining values are both 3, so the median stays 3 hours. One value moved the mean almost a whole hour and the median not at all: the median and the IQR are resistant, which is why they suit data with an outlier.

SCORING

1 point: (a) median 3, $Q_1 = 2$, $Q_3 = 5$ and mean about 3.9 hours. 1 point: (b) fences −2.5 and 9.5, the 16 named as the only outlier, and the gap from 7 to 15 hours. 1 point: (c) skewed right, with the median and the IQR (not the mean or the MAD), each with its unit. 1 point: (c) context, the outlier and gap, and a last sentence answering the question. 1 point: (d) mean falls to about 3.1 (a change of about 0.9 hours), median stays 3, and the median named as resistant.

7. In gym class, each of 11 students did as many sit-ups as possible in one minute. The counts were 43, 39, 46, 10, 41, 45, 40, 47, 43, 42, 44. The student who did 10 got a cramp and stopped early. (5 points)

(a) Use the 1.5 IQR rule to decide whether 10 is an outlier. Show the quartiles and the fences.

(b) Find the mean and the median of all 11 counts, and of the 10 counts left without the 10.

(c) The gym teacher wants one number for how many sit-ups a student in this class usually does. Which center should she report, and why?

ANSWER

(a) $Q_1 = 40$, $Q_3 = 45$, IQR 5, fences 32.5 and 52.5, so 10 is an outlier. (b) All 11: mean 40, median 43. Without the 10: mean 43, median 43. (c) The median, 43 sit-ups.

SOLUTION

(a) Sorted: 10, 39, 40, 41, 42, 43, 43, 44, 45, 46, 47. The median is the sixth value, 43. Leaving it out, the halves are 10, 39, 40, 41, 42 and 43, 44, 45, 46, 47, so $Q_1 = 40$ and $Q_3 = 45$. The IQR is 5 and the step is 1.5 × 5 = 7.5, so the fences are 40 − 7.5 = 32.5 and 45 + 7.5 = 52.5. Since 10 < 32.5, it is an outlier.

(b) All 11: the total is 440, so the mean is $\frac{440}{11} = 40$ sit-ups, and the median is 43 sit-ups. Without the 10: the total is 430, so the mean is $\frac{430}{10} = 43$ sit-ups. The sorted list 39, 40, 41, 42, 43, 43, 44, 45, 46, 47 has fifth and sixth values 43 and 43, so the median is 43 sit-ups.

(c) The median, 43 sit-ups. The outlier pulled the mean down by 3 to 40, lower than eight of the ten students who finished the minute, while the median stayed at 43 either way. The median uses only the middle place, so one cut-short count barely matters to it.

SCORING

2 points: correct quartiles (40 and 45) and fences (32.5 and 52.5), concluding 10 is an outlier (1 point if the quartiles are right but a fence is wrong). 2 points: the four values in (b), mean 40 and median 43, then mean 43 and median 43 (1 point for two or three correct). 1 point: chooses the median with a reason based on the outlier pulling the mean.

8. A toll booth handles about 3000 cars a day. One day a worker counts the people in 20 cars chosen at random. (4 points)

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

people x	**1**	**2**	**3**	**4**
cars f	8	6	4	2

(a) Find the median, the quartiles and the IQR of the 20 counts.

(b) Find the standard deviation that fits this question, to two decimal places.

(c) The worker types the 20 counts into a spreadsheet and uses STDEV.P. What does it show, and why is it not the one wanted?

ANSWER

(a) Median 2, $Q_1 = 1$, $Q_3 = 3$, IQR = 2 (b) $s = \sqrt{\frac{20}{19}} \approx 1.03$ people (c) It shows 1, the divide-by-20 figure; the cars are a sample of the day's cars, so the divisor should be 19.

SOLUTION

(a) In order, counts 1 to 8 are 1, counts 9 to 14 are 2, 15 to 18 are 3, and 19 and 20 are 4. The median is the mean of the 10th and 11th values, both 2, so it is 2. The lower half is the first 10 values, whose median (the 5th and 6th) is $Q_1 = 1$. The upper half is values 11 to 20, whose median (the 15th and 16th) is $Q_3 = 3$. The IQR is 3 − 1 = 2.

(b) The 20 cars stand in for all the day's cars, so they are a sample with $n = 20$. The total is $1 \cdot 8 + 2 \cdot 6 + 3 \cdot 4 + 4 \cdot 2 = 40$, so $\bar{x} = 2$.

x	f	d	d^2	fd^2
1	8	−1	1	8
2	6	0	0	0
3	4	1	1	4
4	2	2	4	8
sum	20			20

$s^2 = \dfrac{20}{19} \approx 1.053$ and $s = \sqrt{\frac{20}{19}} \approx 1.03$ people.

(c) STDEV.P divides by 20: $\sqrt{\frac{20}{20}} = 1$. That is the population formula. The worker wants the spread of all the day's cars from 20 of them, so the sample formula, STDEV.S, is the right one.

SCORING

1 point: median 2, $Q_1 = 1$, $Q_3 = 3$ and IQR 2. 1 point: the mean 2 and the sum of fd^2, 20. 1 point: divides by 19 and gets $s \approx 1.03$. 1 point: STDEV.P gives 1, and the cars are a sample, not the whole population.

9. The distances that one model of electric scooter travels on a full charge are modeled by $N(40, 3)$, in kilometers. Use the empirical rule: for a normal model, about 68%, 95% and 99.7% of values lie within one, two and three standard deviations of the mean. (4 points)

(a) Find the z-scores of distances of 34 km and 43 km.

(b) Estimate the share of scooters that either travel less than 34 km or travel more than 43 km on a full charge.

(c) The maker wants to advertise a range that only about 0.15% of its scooters fall short of. What range should it advertise?

(d) A new battery keeps the standard deviation at 3 km but moves the mean, so that only about 2.5% of scooters travel less than 38 km. Find the new mean.

ANSWER

(a) $z = -2$ and $z = 1$ (b) about 18.5% (c) 31 km (d) $\mu = 44$ km

SOLUTION

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

(a) $\frac{34-40}{3} = \frac{-6}{3} = -2$ and $\frac{43-40}{3} = \frac{3}{3} = 1$.

(b) Below $z = -2$ is half of 100% − 95%, which is 2.5%. Above $z = 1$ is half of 100% − 68%, which is 16%. The two regions do not overlap, so together they hold about 2.5% + 16% = 18.5%.

(c) The tail below $z = -3$ is half of 100% − 99.7%, which is 0.15%. So the range is the distance with $z = -3$: $x = 40 + (-3)(3) = 31$ km.

(d) A share of 2.5% below is the tail below $z = -2$, so 38 km has $z = -2$ on the new model: $38 = \mu - 2(3)$, and $\mu = 44$ km.

SCORING

(a) 1 point for both z-scores, −2 and 1. (b) 1 point for about 18.5%, as 2.5% + 16%. (c) 1 point for 31 km, from $z = -3$. (d) 1 point for $\mu = 44$ km, from $z = -2$. Follow through from a wrong (a) in (b).

10. Before a typing course, the typing speeds of a large group of students are modeled by $N(40, 8)$, in words per minute. Suppose the course turns every speed x into $1.1x + 5$. (4 points)

(a) Give the model for the speeds after the course.

(b) Where are the inflection points of the new curve?

(c) What speed before the course turns into 66.6 words per minute after it? How many standard deviations above the mean is it before the course, and how many is 66.6 after it?

(d) Before the course, the model gives a share of about 0.023 of students typing faster than 56 words per minute. What share does the new model give faster than 66.6 words per minute, and what share slower than 31.4 words per minute? Explain.

ANSWER

(a) $N(49, 8.8)$ (b) above 40.2 and 57.8 words per minute (c) 56 words per minute; 2 standard deviations above the mean before, and 66.6 is 2 above after. (d) about 0.023 faster than 66.6, and about 0.023 slower than 31.4.

SOLUTION

(a) A change $a + bx$ turns $N(\mu, \sigma)$ into $N(a + b\mu, |b|\sigma)$. Here the mean becomes 1.1 × 40 + 5 = 49 and the standard deviation 1.1 × 8 = 8.8: the model is $N(49, 8.8)$. Adding 5 moves the center only; multiplying by 1.1 stretches both.

(b) 49 − 8.8 = 40.2 and 49 + 8.8 = 57.8 words per minute.

(c) Solve $1.1x + 5 = 66.6$: $1.1x = 61.6$, so $x = 56$ words per minute. Before: $\frac{56-40}{8} = 2$ standard deviations above the mean. After: $\frac{66.6-49}{8.8} = \frac{17.6}{8.8} = 2$ standard deviations above. The change moves every feature of the curve along with the values.

(d) The change $1.1x + 5$ keeps order, so a student faster than 66.6 after the course is exactly a student who was faster than 56 before it: the share is the same, about 0.023. The new curve is symmetric about 49, and 31.4 = 49 − 17.6 is the mirror image of 66.6 = 49 + 17.6, so the share slower than 31.4 is also about 0.023.

SCORING

1 point: $N(49, 8.8)$. 1 point: 40.2 and 57.8. 1 point: 56 words per minute, 2 standard deviations above on both curves. 1 point: 0.023 for both, with the same students (order kept) and the mirror image in 49 as reasons.

11. The times taken to put together a flat-pack desk are modeled by $N(52, 8)$, in minutes. Use the standard normal table or a calculator's normal function, and give shares to four decimal places. (4 points)

(a) What share of people take more than 64 minutes?

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

(b) A furniture store sells 1,200 of these desks in a year, and each buyer puts one together. Find the share of buyers who take between 48 and 64 minutes, and about how many of the 1,200 buyers the model expects to take that long. Round the count to the nearest whole buyer.

ANSWER

(a) about 0.0668 (b) about 0.6247; about 750 buyers

SOLUTION

(a) $z = \dfrac{64 - 52}{8} = \dfrac{12}{8} = 1.5$, so $P(X > 64) = 1 - \Phi(1.5) \approx 1 - 0.9332 = 0.0668$.

(b) The cuts as z-scores are $\dfrac{48 - 52}{8} = -0.5$ and 1.5. The share between is

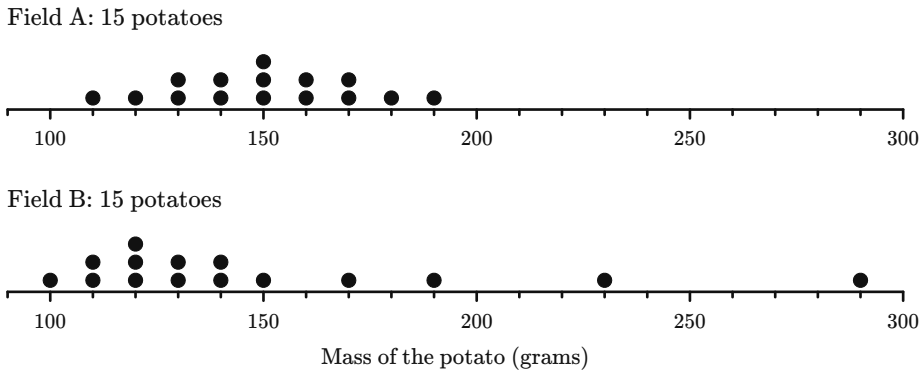
$$P(48 < X < 64) = \Phi(1.5) - \Phi(-0.5) \approx 0.9332 - 0.3085 = 0.6247.$$

Of 1,200 buyers the model expects about 1,200 × 0.6247 = 749.64, so about 750 buyers.

SCORING

(a) 2 points: 1 for $z = 1.5$, 1 for 0.0668 (an area to the right, not 0.9332). (b) 2 points: 1 for 0.6247 from both z-scores, −0.5 and 1.5 (a student who uses $\Phi(1.5) - \Phi(0.5)$ gets 0.2417 and loses this point); 1 for about 750 buyers from their share.

12. Mateo dug random samples of 15 potatoes from each of two large fields and weighed each one, to the nearest 10 grams. The dot plots show the masses.



Field A: 110, 120, 130, 130, 140, 140, 150, 150, 150, 160, 160, 170, 170, 180, 190

Field B: 100, 110, 110, 120, 120, 120, 130, 130, 140, 140, 150, 170, 190, 230, 290 (5 points)

(a) Describe the shape of each distribution.

(b) Use the 1.5 IQR rule to decide whether either of Field B's two heaviest potatoes is an outlier.

(c) Find the mean and the median of each field. Which pair should be compared, and what does the comparison say?

(d) Do the samples give good reason to think the potatoes in Field A tend to be heavier? Explain by comparing the shift with the spread.

ANSWER

(a) Field A: roughly symmetric with one peak at 150 g. Field B: skewed right, most potatoes from 100 to 150 g and a long tail out to 290 g. (b) Field B: $Q_1 = 120$, $Q_3 = 170$, IQR 50, so the cut-off is 245 g: 290 g is an outlier and 230 g is not. (c) Field A: mean 150 g, median 150 g. Field B: mean 150 g, median 130 g. Compare medians: a typical Field A potato is 20 g heavier. (d) No. The shift of 20 g is half of Field A's IQR of 40 g and less than half of Field B's 50 g, and the middle halves, 130 to 170 g and 120 to 170 g, almost coincide.

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

SOLUTION

(a) Field A's dots stack highest at 150 g and fall away evenly on both sides: roughly symmetric with one peak. Field B's dots crowd from 100 to 150 g and then trail away one potato at a time to 290 g: skewed right.

(b) Field B has 15 values, so the median is the 8th, 130 g, left out of both halves. Lower half: 100, 110, 110, 120, 120, 120, 130, so $Q_1 = 120$. Upper half: 140, 140, 150, 170, 190, 230, 290, so $Q_3 = 170$. The IQR is 50 g and 170 + 1.5 × 50 = 245. So 290 g is an outlier and 230 g is not.

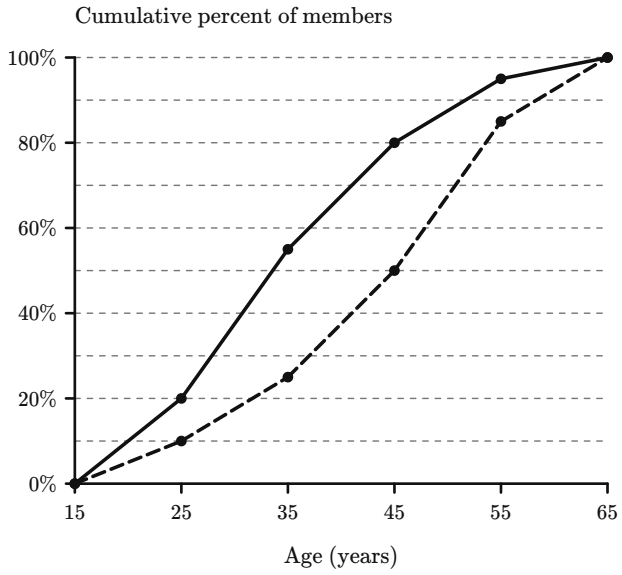
(c) Field A: the sum is 2,250, so the mean is $\frac{2{,}250}{15} = 150$ g, and the 8th value is 150 g. Field B: the sum is 2,250, so the mean is also 150 g, and the median is 130 g. Field B is skewed right with an outlier, so compare medians (and interquartile ranges) for both fields: "In these samples, the median Field A potato, 150 g, is 20 g heavier than the median Field B potato, 130 g." The equal means hide this: the tail of heavy Field B potatoes pulls its mean 20 g above its median.

(d) Field A's quartiles are 130 and 170 g, an IQR of 40 g; Field B's IQR is 50 g. The shift of 20 g is small against both, and the middle halves, 130 to 170 g and 120 to 170 g, lie mostly on top of each other. Another pair of random samples could shrink or reverse the gap, so the samples do not justify saying Field A's potatoes tend to be heavier.

SCORING

1 point for both shapes; 1 point for the quartiles 120 and 170, the cut-off 245 and deciding 290 is an outlier and 230 is not; 2 points for both means and medians (1) and choosing medians because of Field B's skew or outlier, with the 20 g comparison (1); 1 point for comparing the shift of 20 g with the IQRs and concluding no.

13. The graph shows the cumulative relative frequency of the ages of the members of two community choirs: Choir A has 60 members and Choir B has 80. Each plotted point lies on a faint line or halfway between two.
(5 points)



Solid line: Choir A. Dashed line: Choir B.

(a) Estimate the median age in each choir, to the nearest tenth of a year.

(b) Estimate the interquartile range of the ages in each choir, to the nearest tenth of a year.

(c) How many members of each choir are 55 or older?

(d) Compare the two age distributions in context, using the graph.

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

ANSWER

(a) Choir A about 33.6 years, Choir B about 45 years. (b) Choir A about 16.6 years ($Q_1 \approx 26.4$, $Q_3 \approx 43$); Choir B about 17.1 years ($Q_1 \approx 35$, $Q_3 \approx 52.1$). (c) Choir A 3 members, Choir B 12 members. (d) Choir B's members tend to be older (median about 11 years higher, B's graph lies on or below A's at every age), while the middle halves of the two choirs have about the same spread.

SOLUTION

(a) Choir A passes 50% between 20% at 25 and 55% at 35, 30 points into a climb of 35; Choir B passes 50% exactly at 45:

$$M_A \approx 25 + \frac{30}{35} \cdot 10 \approx 33.6 \text{ years}, \qquad M_B \approx 45 \text{ years}.$$

(b) Choir A: 25% lies between 20% and 55%, and 75% between 55% (at 35) and 80% (at 45):

$$Q_1 \approx 25 + \frac{5}{35} \cdot 10 \approx 26.4,$$
$$Q_3 \approx 35 + \frac{20}{25} \cdot 10 = 43,$$

so the interquartile range is about 43 − 26.43 ≈ 16.6 years. Choir B: 25% is reached exactly at 35, and 75% lies between 50% at 45 and 85% at 55:

$$Q_3 \approx 45 + \frac{25}{35} \cdot 10 \approx 52.1,$$

so the interquartile range is about 52.14 − 35 ≈ 17.1 years.

(c) 55 is a boundary. Choir A is at 95% there, so 5% of 60, 3 members, are 55 or older. Choir B is at 85%, so 15% of 80, 12 members, are.

(d) At every age the share of Choir B below that age is no more than Choir A's, so B's graph lies to the right of A's: B's members tend to be older. B's median, about 45 years, is about 11 years above A's, about 33.6. The spreads of the middle halves are about the same, interquartile ranges of about 16.6 and 17.1 years. B also has more older members, 15% against 5% aged 55 or over.

SCORING

1 point for both medians; 2 points for both interquartile ranges (1 each, quartiles shown); 1 point for 3 and 12 members; 1 point for a comparison in context naming center (B older) and spread (similar).

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