Statistics Unit 1 worksheet: Types of data: categorical, numerical, discrete, continuous (answer key)

ANSWERS AT A GLANCE

1. (a) categorical (b) numerical, discrete (c) numerical, continuous (d) categorical (e) numerical, continuous (f) numerical, discrete
2. (a) label, categorical (b) amount, numerical (c) label, categorical (d) amount, numerical (e) label, categorical (f) amount, numerical
3. (a) wheat 8, white 5, sourdough 4, rye 3 (b) wheat (c) 15%
4. (a) at least 3.35 kg and less than 3.45 kg (b) 3 kg
5. (a) short 3, medium 6, long 3 (b) $\frac{1}{4}$ (c) 42 minutes; no.
6. (a) Screen and showing code: they name a screen and a showing, they are not amounts. (b) 96 tickets per showing (c) Screen 5, with 263 tickets.
7. 12 have a library card and 8 do not.
8. (a) The labels only put the heats in order; they are not amounts of heat. (b) The labels are categorical, so their mean is not an amount; a sensible summary is the count: 2 mild and 2 hot (none medium).

WORKED SOLUTIONS

1. Carlos runs a bike repair shop. For each bike brought in, he records:

   (a) the kind of bike (road, mountain or kids'), (b) the number of spokes he replaces, (c) the time the repair takes, in minutes, (d) the color of the frame, (e) the mass of the bike, in kilograms, (f) the number of days the bike waits before it is picked up.

   Sort each variable as categorical or numerical. Say whether each numerical variable is discrete or continuous.

   ANSWER

   (a) categorical (b) numerical, discrete (c) numerical, continuous (d) categorical (e) numerical, continuous (f) numerical, discrete

   SOLUTION

   (a) The kind of bike names a group, so it is categorical.

   (b) Spokes are counted one by one, so the number replaced is numerical and discrete.

   (c) Time is measured, so it is numerical and continuous, even if Carlos writes it to the nearest minute.

   (d) A color names a group: categorical.

   (e) Mass is measured on a scale: numerical and continuous.

   (f) Days waited, as recorded, are counted in whole days: numerical and discrete.

2. A school office keeps a record for each student. Each number below is written with digits. Say whether it is an amount (numerical) or a label for a group or an individual (categorical).

   (a) the student's ID number, (b) the number of days the student was absent this term, (c) the number of the student's homeroom, (d) the student's height, in centimeters, (e) the number on the student's locker, (f) the number of books the student borrowed from the library this term.

   ANSWER

   (a) label, categorical (b) amount, numerical (c) label, categorical (d) amount, numerical (e) label, categorical (f) amount, numerical

   SOLUTION

   Ask whether adding or averaging the values gives an amount of something.

   (a) An ID number names one student. Adding two ID numbers gives no amount of anything: categorical.

   (b) Days absent are counted: numerical (discrete).

   (c) A homeroom number names a room, and so a group of students. Its digits are a name, so it is categorical.

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

(d) Height is measured in centimeters: numerical (continuous).

(e) A locker number names a locker: categorical.

(f) Books are counted: numerical (discrete).

3. A sandwich shop writes down the bread chosen for its first 20 orders on Monday:

wheat, white, wheat, rye, sourdough, wheat, white, wheat, sourdough, rye, wheat, white, sourdough, wheat, white, rye, wheat, sourdough, white, wheat.

(a) Count each kind of bread. (b) Which bread is the mode? (c) What percent of the orders were on rye?

ANSWER

(a) wheat 8, white 5, sourdough 4, rye 3 (b) wheat (c) 15%

SOLUTION

(a) Bread type is categorical, so count each category: wheat 8, white 5, sourdough 4, rye 3. Check: $8 + 5 + 4 + 3 = 20$.

(b) Wheat has the largest count, so it is the mode.

(c) 3 of 20 orders were on rye: $\frac{3}{20} = \frac{15}{100} = 15\%$.

4. A hospital scale shows a newborn baby's mass to the nearest tenth of a kilogram. Mei's baby brother reads 3.4 kg.

(a) What could his true mass be?

(b) The clinic's old scale shows mass to the nearest whole kilogram. What would it show for him?

ANSWER

(a) at least 3.35 kg and less than 3.45 kg (b) 3 kg

SOLUTION

(a) A mass rounds to 3.4 kg when it is at least $3.4 - 0.05 = 3.35$ kg and less than $3.4 + 0.05 = 3.45$ kg.

(b) Every mass from 3.35 kg up to 3.45 kg is nearer to 3 kg than to 4 kg, so the old scale shows 3 kg whatever his true mass in that range. Mass is still continuous: each scale only rounds it more or less finely.

5. Ryan's class spends a morning sorting donated books for the town library's book sale. The teacher writes down how many minutes each student worked before taking a break:

15, 42, 28, 35, 8, 50, 22, 30, 45, 18, 25, 38.

She sorts the times into classes: short (under 20 minutes), medium (20 minutes up to 40 minutes) and long (40 minutes and over).

(a) Count the students in each class.

(b) What fraction of the students worked for a long time?

(c) How many more minutes did the longest worker work than the shortest worker? Could you find this from the class counts alone? Explain.

ANSWER

(a) short 3, medium 6, long 3 (b) $\frac{1}{4}$ (c) 42 minutes; no.

SOLUTION

(a) Short: 15, 8, 18, so 3. Medium: 28, 35, 22, 30, 25, 38, so 6. Long: 42, 50, 45, so 3. Check: $3 + 6 + 3 = 12$.

(b) $\frac{3}{12} = \frac{1}{4}$ of the students.

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

(c) 50 − 8 = 42 minutes. From "3 short, 6 medium, 3 long" alone, no: the shortest worker could have worked anything under 20 minutes and the longest anything from 40 minutes up. Sorting minutes into classes turns a numerical variable into a categorical one, and subtracting needs the exact amounts.

6. A movie theater lists its Saturday showings. Every column is written with digits.

Screen	Showing code	Tickets sold	Film length (min)
2	214	84	95
5	309	120	128
1	127	56	102
5	452	143	128
3	380	97	110
2	118	76	95

(a) Which columns are categorical? Give a reason.

(b) Find the mean number of tickets sold per showing, and say what it means.

(c) Which screen sold the most tickets in all on Saturday, and how many?

ANSWER

(a) Screen and showing code: they name a screen and a showing, they are not amounts. (b) 96 tickets per showing (c) Screen 5, with 263 tickets.

SOLUTION

(a) A screen number names a room and a showing code names a showing. Averaging them or subtracting them gives no amount of anything, so both are categorical. Tickets sold is a count (numerical, discrete) and film length is a measured time (numerical, continuous).

(b) 84 + 120 + 56 + 143 + 97 + 76 = 576 and 576 ÷ 6 = 96. If the 576 tickets were shared equally among the six showings, each would have 96.

(c) The screen number is used to sort the showings into groups. Screen 5: 120 + 143 = 263. Screen 2: 84 + 76 = 160. Screen 3: 97. Screen 1: 56. Screen 5 sold the most, 263 tickets.

7. Leah asks the 20 students in her reading club "Do you have a library card?" She codes each answer 1 for yes and 0 for no. The mean of the codes is $\frac{3}{5}$. How many of the students have a library card, and how many do not?

ANSWER

12 have a library card and 8 do not.

SOLUTION

Adding the codes adds a 1 for each yes and nothing for each no, so the sum of the codes is the number of yes answers. The sum is the mean times the number of students: $\frac{3}{5} \times 20 = 12$.

So 12 students have a card and 20 − 12 = 8 do not. The mean $\frac{3}{5}$ is the fraction who said yes.

8. A taco truck labels its sauces by heat: 1 for mild, 2 for medium and 3 for hot.

(a) Andre says a hot sauce is three times as hot as a mild one, because 3 is three times 1. Explain why the labels do not show this.

(b) Andre buys four sauces labeled 1, 1, 3 and 3. He says their average heat is 2, so on average his sauces are medium. Explain what is wrong, and give a summary of his four sauces that does make sense.

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

ANSWER

(a) The labels only put the heats in order; they are not amounts of heat. (b) The labels are categorical, so their mean is not an amount; a sensible summary is the count: 2 mild and 2 hot (none medium).

SOLUTION

(a) Heat level is a categorical variable with an order. The labels say hot comes after medium, which comes after mild, but not by how much. The truck could have labeled them 1, 5 and 10 with the same sauces in the jars, and then "hot" would be "ten times" mild. A ratio that changes when only the labels change says nothing about the sauces.

(b) $1 + 1 + 3 + 3 = 8$ and $8 \div 4 = 2$, but with labels 1, 5 and 10 the same sauces would average $(1 + 1 + 10 + 10) \div 4 = 5.5$, and that is not "medium" either. The mean describes the labels, not the sauces. Andre has no medium sauce at all. Count each category instead: 2 mild and 2 hot, so half his sauces are mild and half are hot.

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