Probability & Statistics Workbook

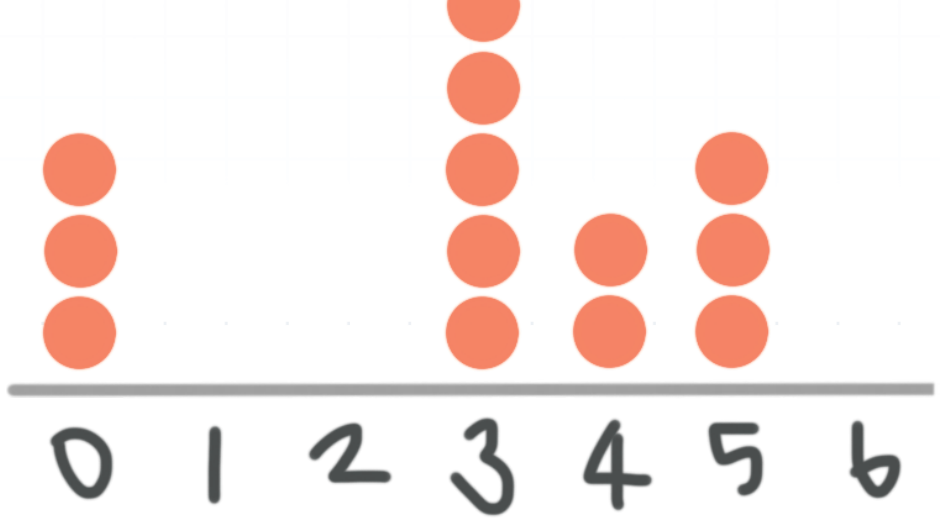
Data distributions

krista king MATH

MEAN, VARIANCE, AND STANDARD DEVIATION

1. Mrs. Bayer's students take a test on Friday. She grades their tests over the weekend and notes that the average test score is 68 points with a population standard deviation of 5 points. She decided to add 10 points to all of the tests. What are the new mean and population standard deviation?

2. What is the sample variance of the data set, rounded to the nearest hundredth?



3. Sometimes it can be helpful to calculate the standard deviation by using a table. Use the data to fill in the rest of the table and then use the table to calculate the sample standard deviation.

Data value	Data value - Mean	Squared difference
97		
110		
112		
121		
110		
98		
Total		

■ 4. The sum of the squared differences from the population mean for a data set is 212. If the data set has 25 items, what is the population standard deviation?

■ 5. For the data set $40, 44, 47, 55, 60, 60, 65, 80$, find

$$\sum_{i=1}^{n} (x_i - \bar{x})$$

What does this say about why we square the $(x_i - \bar{x})$ in the variance and standard deviation formulas?

6. Give an example of a situation where $5 could represent a large standard deviation and another where $5 could represent a small standard deviation.

FREQUENCY HISTOGRAMS AND POLYGONS, AND DENSITY CURVES

■ 1. A dog walking company keeps track of how many times each dog receives a walk. $40\,\%$ of all the dogs walked by the company received between 25 and 40 walks, and no dogs received more than 40 walks. How many dogs received between 0 and 25 walks, if the company walks 400 dogs?

■ 2. The number of crayons in each student's pencil box is

$$4,\ 1,\ 5,\ 5,\ 9,\ 11,\ 15,\ 13,\ 15,\ 14,\ 16,\ 17,\ 20,\ 16,\ 16,\ 17$$

Complete the frequency and relative frequency tables for the data and use it to create a relative frequency histogram.

Crayons	Frequency	Relative Frequency
1-5		
6-10		
11-15		
16-20		
Totals:		**100%**

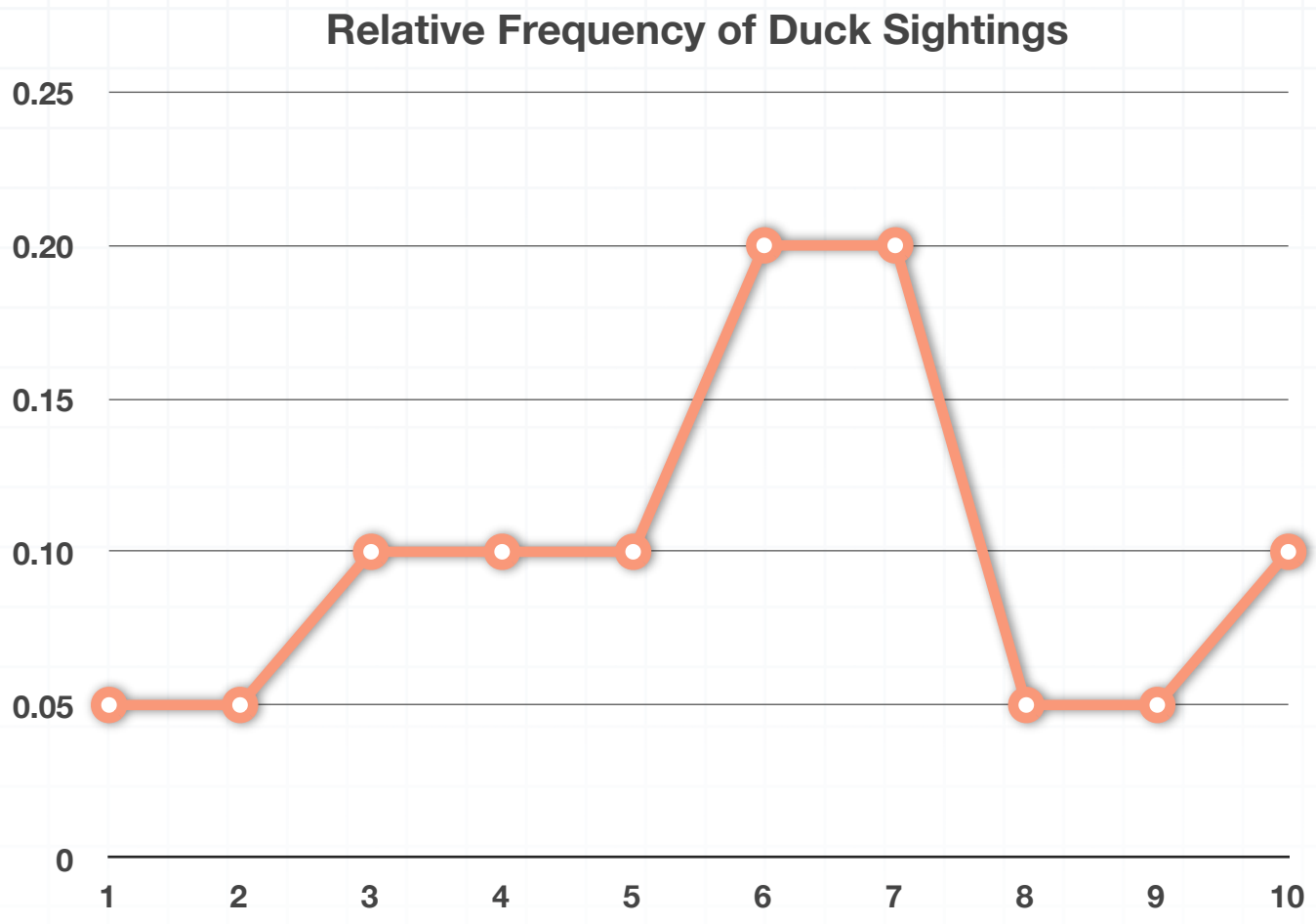
■ 3. The table shows the scores on the last history exam in Mr. Ru's class.

40	32	40	83
95	33	87	59
32	81	46	78
91	61	55	88
40	61	82	99
72	47	83	91
101	77	65	87

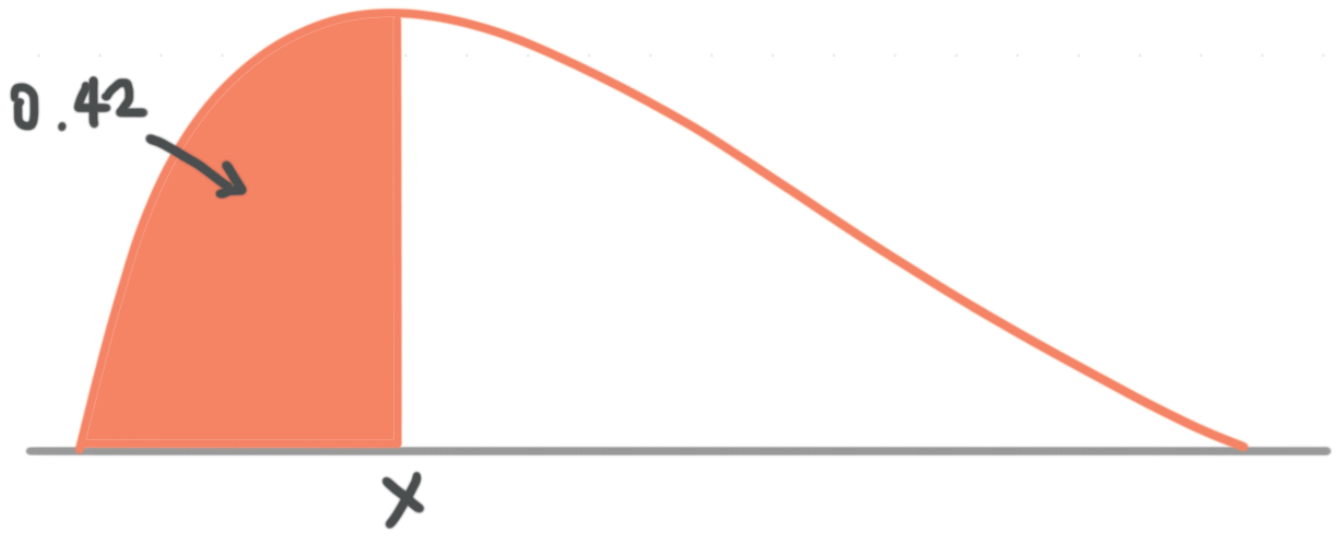
Complete the relative frequency table and create a frequency polygon for the data.

Score	Frequency	Relative Frequency
30-39		
40-49		
50-59		
60-69		
70-79		
80-89		
90-99		
100-109		
Totals:		

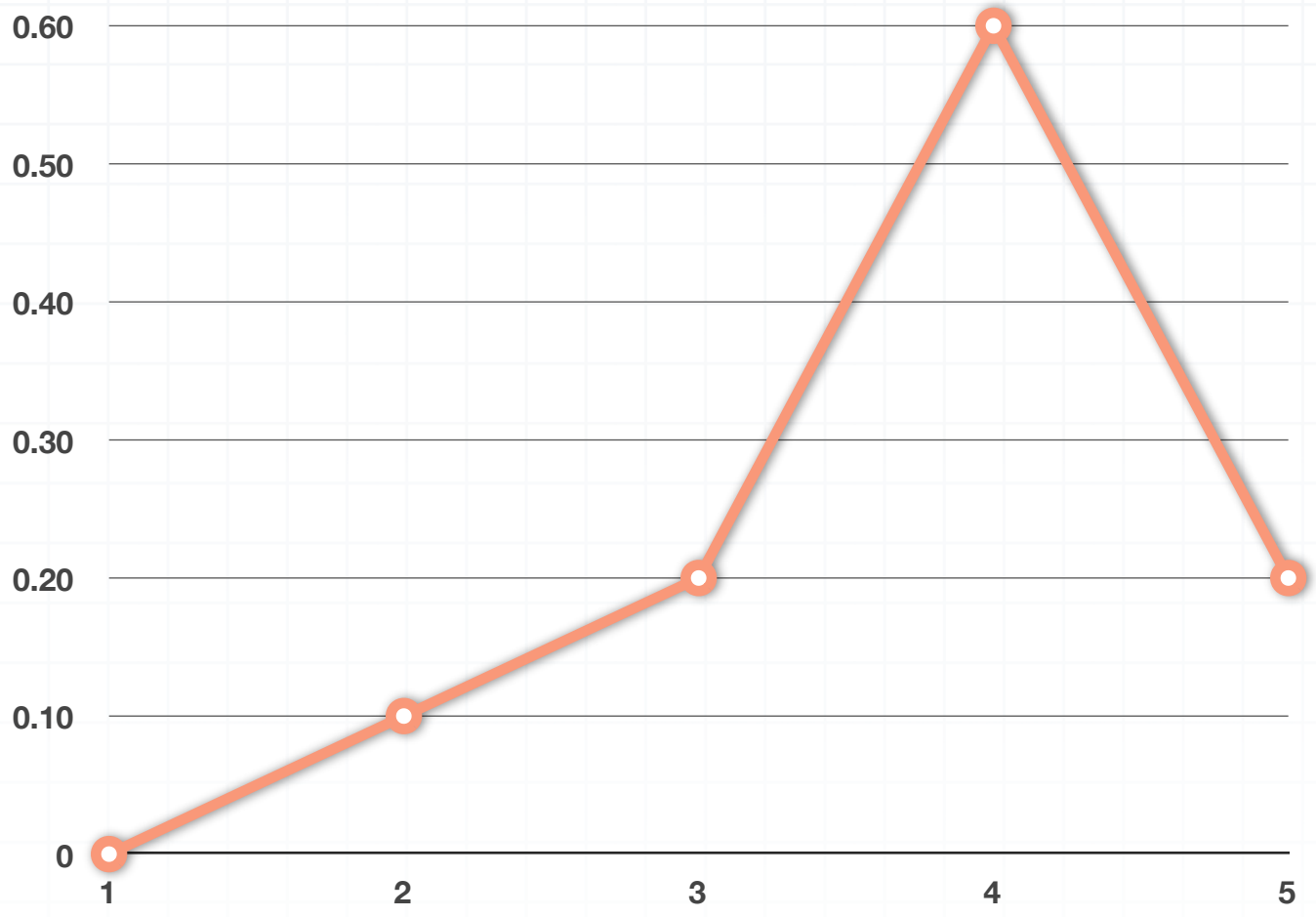
■ 4. Becky kept track of the number of ducks she saw at her neighborhood pond at $6:30$ a.m. every morning for 365 days. On how many days did Becky see more than 5 ducks?



5. What percentage of the population is greater than x for the density curve?



6. What percentage of the area in the density curve is between 3 and 5?



SYMMETRIC AND SKEWED DISTRIBUTIONS AND OUTLIERS

1. Which type of distribution is modeled in the box plot (symmetric, negatively skewed, or positively skewed)?

2. Which type of distribution is modeled in the box plot (symmetric, negatively skewed, or positively skewed)?

3. The ages (in months) that babies spoke for the first time are

6, 8, 9, 10, 10, 11, 11, 12, 12, 13, 15, 15, 18, 19, 20, 20, 21

Are there outliers in the data set? If so, state what they are. What is the best measure of central tendency for the data? What is the best measure of spread?

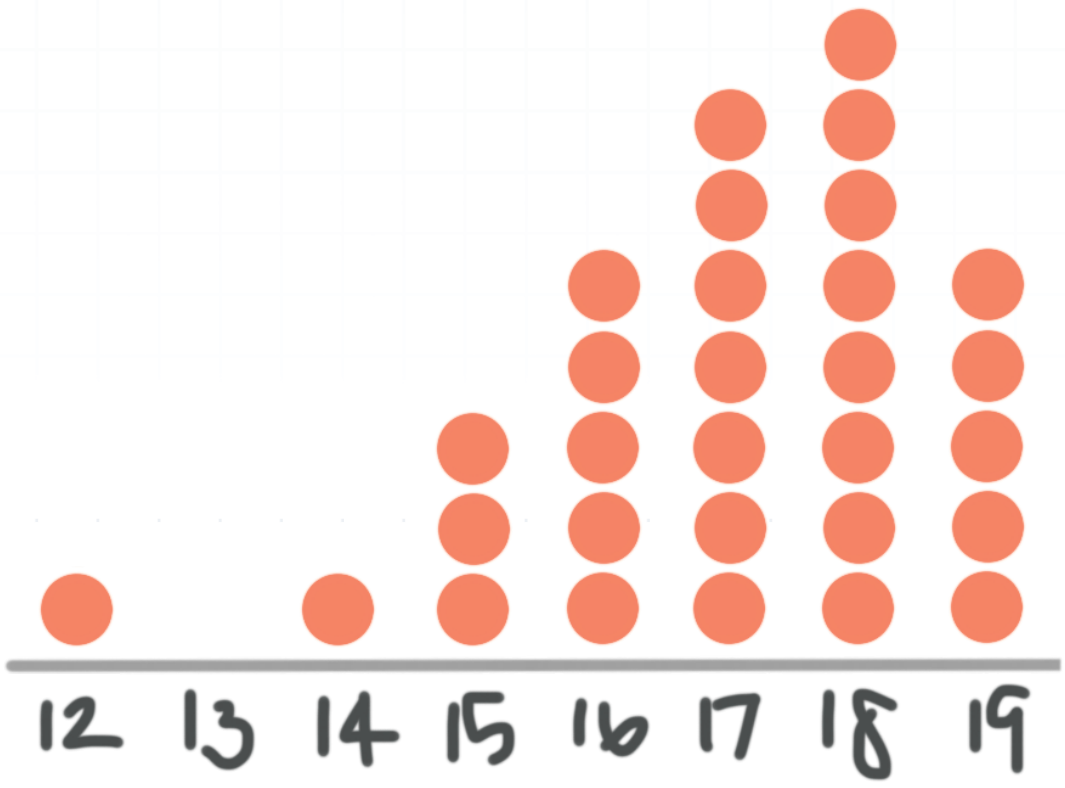
4. The number of text messages sent each day by Lucy's mom is

0, 18, 19, 20, 20, 20, 21, 23, 23, 23, 24, 24,

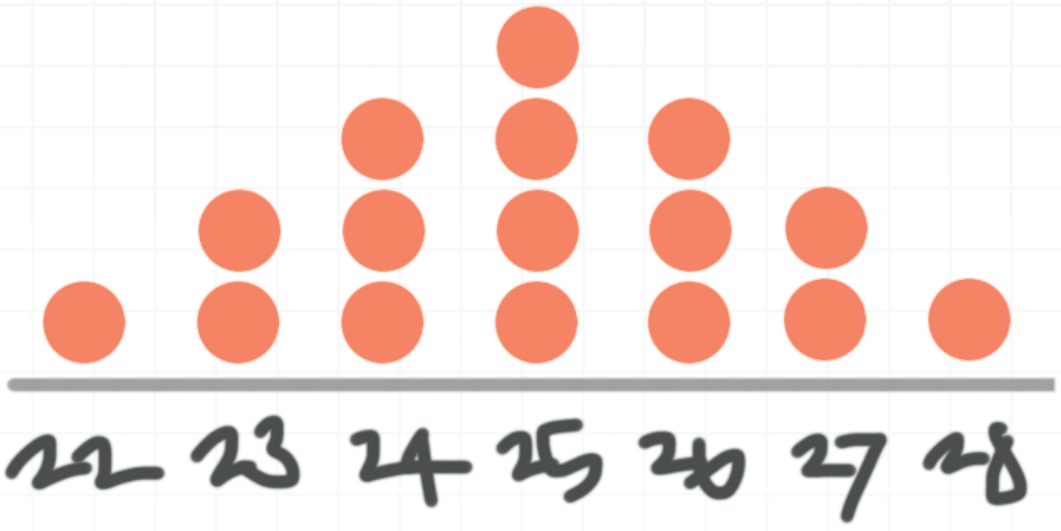
24, 24, 24, 25, 25, 25, 25, 25, 25, 30, 30, 31

Are there outliers in the data set? If so, state what they are. What is the best measure of central tendency for the data? What is the best measure of spread?

■ 5. Describe the shape, center, and spread of the data. State if there are outliers and what they are if they exist.



■ 6. Describe the shape, center and spread of the data. State if there are outliers and what they are if they exist.



NORMAL DISTRIBUTIONS AND Z-SCORES

■ 1. A population has a mean of 62 and a standard deviation of 5. What is the z-score for a value of 50?

■ 2. What percentile is a z-score of -1.68?

■ 3. A population has a mean of 170 centimeters and a standard deviation of 8 centimeters. What percentage of the population has a value less than 154 centimeters?

■ 4. The mean diameter of a North American Native Pine tree is $18''$ with a standard deviation of $4''$. What is the approximate diameter for a tree in the 21st percentile for this distribution? Assume an approximately normal distribution.

■ 5. The mean diameter of a North American Native Pine tree is $18''$ with a standard deviation of $4''$. According to the Empirical Rule, $68\,\%$ of North American Native Pines have a diameter between which two values? Assume an approximately normal distribution.

6. IQ scores are normally distributed with a mean of 100 and a standard deviation of 16. What percentage of the population has an IQ score between and 120 and 140?

CHEBYSHEV'S THEOREM

1. If the Empirical Rule tells us that $95\,\%$ of the area under the normal distribution falls within two standard deviations of the mean, what will Chebyshev's Theorem say about the same number of standard deviations?

2. A basket of strawberries has a mean weight of 2 ounces with a standard deviation of 0.35 ounces. What percentage of the strawberries in the basket have a weight between 1.5 and 2.5 ounces?

3. A pod of 580 migrating whales travels a mean distance of $2{,}000$ miles each year, with a standard deviation of 175 miles. How many whales in the pod travel between $1{,}600$ and $2{,}400$ miles?

4. A hockey team of 20 boys have a mean height of 73 inches, with a standard deviation of 1.8 inches. Find the height range for the central $90\,\%$ of team members.

5. A university with $40{,}000$ students accepts an average of $10{,}000$ new students each year, with a standard deviation of 500 students. Find the values that make up the middle $75\,\%$ of the yearly acceptance range.

6. A pack of 26 wolves have a mean weight of 100 pounds, with a standard deviation of 24 pounds. Find the weight range for the central $82\,\%$ of the wolves.

COVARIANCE

■ 1. A bakery records sales and number of customers for a sample of hours throughout the week. Calculate the covariance of customers and sales.

Customers	4	7	12	2	3	9	15
Sales	45.75	36.00	58.5	20.00	15.80	39.95	123.45

■ 2. The cost of the stock of two unrelated companies over five days is recorded in the table. Calculate the covariance of the stocks.

Company X	13	13.75	12.70	13.15	14.80
Company Y	21.05	21.55	20.95	21.75	21.50

■ 3. The following table represents temperatures, in Celsius, during a sample of 5 days in two cities with a distance of 50 miles between them. Calculate the covariance.

City X	25	23	24.5	20	18
City Y	23	24	21	18	22

■ 4. David prepares for his annual math and physics exams and decides to take four practice tests for each subject. Calculate the covariance for his test scores for math and physics.

Math, X	85	89	89	93
Physics, Y	92	93	89	90

■ 5. Mark and John exercise daily and record their minutes of daily exercise over 10 days. Calculate the covariance.

Mark	53	57	63	55	45	50	65	60	59	70
John	65	55	60	53	30	45	25	65	57	50

■ 6. An annual return on investment of two stocks over the last 7 years is recorded in the table. Calculate the covariance.

Stock X	3.5	2.4	1.4	-0.5	0.7	1.1	0.5
Stock Y	2.4	1.7	2.1	1.8	2.1	-0.4	0.8

CORRELATION COEFFICIENT

■ 1. Calculate the correlation coefficient of the newborns‘ weight and body length, and then interpret the result.

Weight, kg	Body length, cm
3.55	51
4.01	54
3.05	50
5.35	60
4.22	52
6.12	61
7.45	63
5.95	59
6.35	68
6.98	74

■ 2. Oliver is wondering whether there’s a correlation between the number of hours his classmates studied to prepare for the exam and their exam scores. He surveyed five classmates and recorded the data in a table. Calculate the correlation coefficient.

Study hours	6	2	11	7	5
Exam score	85	79	84	89	91

■ 3. Calculate the value of the Pearson correlation coefficient for the age, in years, and blood glucose levels, in mg/dL, then interpret the result.

Age	28	35	58	42	21	63	46
Blood glucose	101	93	95	105	93	89	100

■ 4. Maria likes discovering interesting correlations. She decides to choose random six days and record the data for shark attacks and ice cream sales in her coastal city. How should she interpret the correlation coefficient.

Shark attacks	4	2	8	11	5	9
Ice cream sales	38	30	55	61	38	42

■ 5. Calculate and interpret the correlation coefficient of the variables.

Hand length, cm	Height, cm
12	158
15	160
11	157
13	164
9	150
18	178
16	169
17	156

6. Calculate and interpret the value of the correlation coefficient for the correlation between systolic blood pressure, in mmHg, and weight, in lbs.

SBP, mmHg	Weight, lbs
138	167
125	153
145	149
156	165
132	170
148	175
160	180
135	140
150	190
155	155

WEIGHTED MEANS AND GROUPED DATA

■ 1. An investor purchases shares of a particular stock on the same date every month for 12 months. He records the price and number of shares each month. Calculate the mean share price.

Stock price	Shares
8	30
10	12
14	10
9	25
6	35
13	15
18	10
21	5
25	7
27	10
28	8
31	4

■ 2. A chemistry course teacher weights class discussions at 0.05, quizzes at 0.10, and group projects at 0.40. Given the grades for one student in the table below, calculate her final grade.

Assignment	Grade	Weight
Quiz 1	88	0.10
Discussion 1	92	0.05
Quiz 2	93	0.10
Discussion 2	90	0.05
Quiz 3	85	0.10
Discussion 3	94	0.05
Quiz 4	97	0.10
Discussion 4	80	0.05
Group Project	85	0.40

■ 3. Given the dataset $\{12, 15, 8, 21, 25, 14, 16, 18, 10\}$, divide it into four groups and find the sample variance and standard deviation.

■ 4. A sample of book club members record the number of books they read last year. Calculate the mean number of books per member.

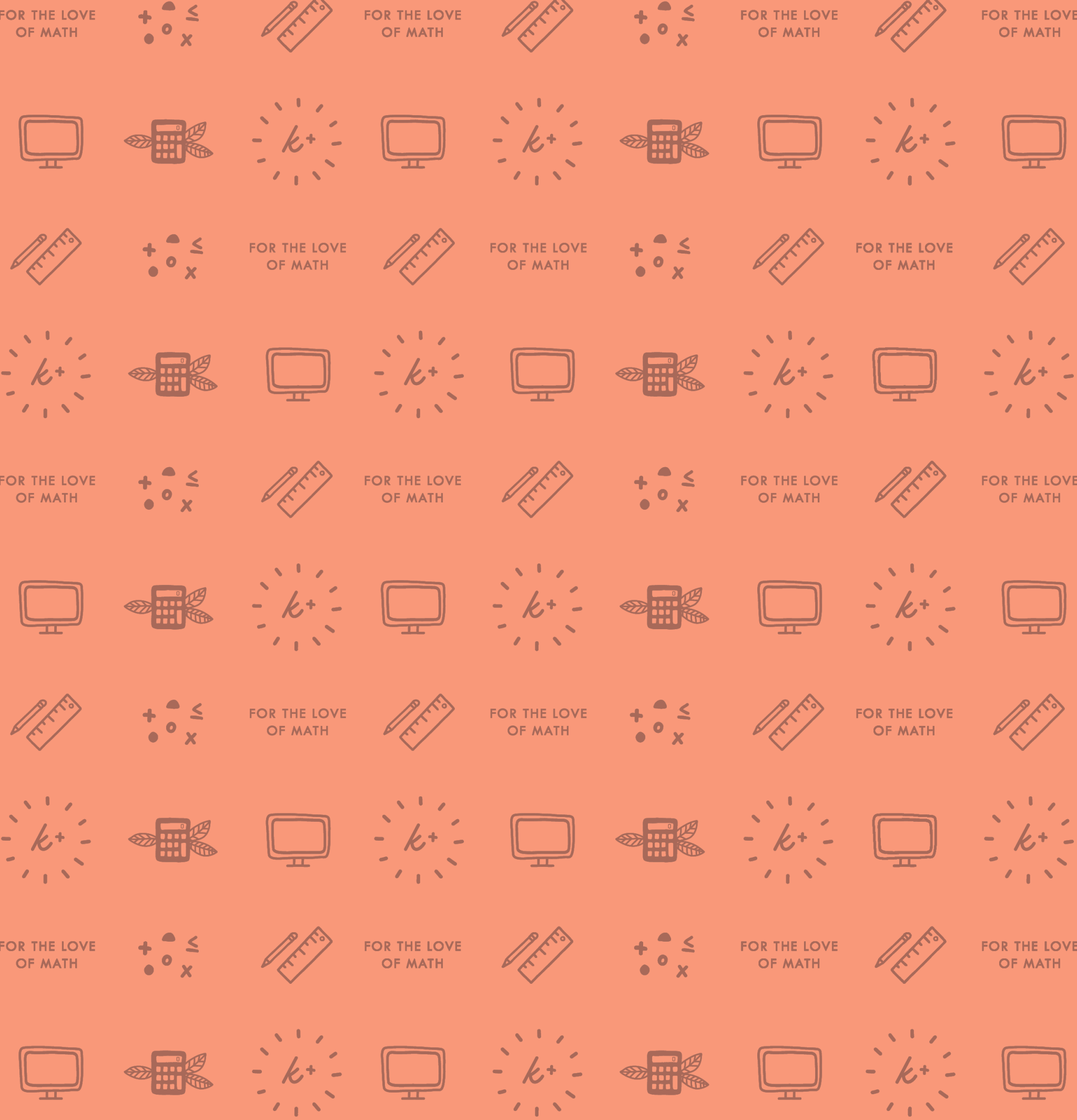
Number of books	0	1	2	3	4	5
Number of people	25	15	18	5	12	3

■ 5. The frequency distribution represents the number of pizza orders a local pizza restaurant received each day over the last 20 days. Calculate the weighted sample mean, variance, and standard deviation.

Orders	Number of days
5 - 7	5
8 - 10	4
11 - 13	4
14 - 16	3
17 - 19	3
20 - 22	1

6. Use the sample data to find the mean, variance, and standard deviation of commute time.

Commute time	Number of people
1 - 5	1
6 - 10	4
11 - 15	6
16 - 20	3
21 - 25	10
26 - 30	13


FOR THE LOVE
OF MATH