

GSE Honors Algebra II

Unit 6: Data Collection and Design and Using The Normal Distribution Curve

Standards Addressed

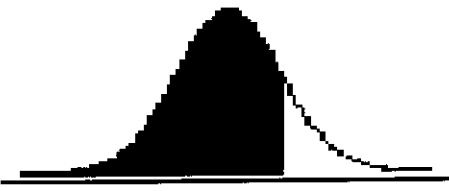
S.ID.2: Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

S.IC.4: Use data from a sample survey to estimate a population mean or proportion and develop a margin of error through the use of simulation models for random sampling.

S.IC.5: Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant

S.IC.6: Evaluate Reports Based on Data

Date	Learning Objective/EQ	Classwork	Homework
April	-Populations, samples & their use in research -What Does a Sample say about a Population? -Appropriate Sampling Methods and Bias -Mean, Median, Mode -IQR, Box & Whisker Plots, 5-Number summary, Variance and Standard Deviation	NOTES: Populations, samples and research NOTES: Conclusions from Data and Bias NOTES: How to make a boxplot Pages 1-4	Page 5
April	-Variance and Standard Deviation	Standard Deviation WS Pages 7-10	Page 11
	-Determine the Center and Spread of Data -Calculate Variance and S.D. to make comparisons -Stats on Calculator	-Measures of Center and Spread WS -Math Award Learning Task Pages 12-13	Page 14
May	QUIZ		Catch up on work
May	Probability Distributions and Histograms	Probability Distribution & Histograms Notes [SMART] Pages 15-16	Pages 17-18
May	Intro to the Empirical Rule	Empirical Rule Handout & Normal Distribution Notes Pages 19-21	Pages 22-23
May	Modeling a normal distribution	Empirical Rule WS Pages 24-27	Finish pages 24-27



The table shows the area to the left of a z-score:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Unit 1: Inferences and Conclusions from Data

Making yourself an intelligent consumer of information

Jul 31-3:12 PM

Definitions

Population:

Sample:

Data:

Jul 31-3:12 PM

Statistic:

Parameter:

Inference:

Aug 4-6:55 AM

Example 1:

In the following study identify the population, sample, and statistic.

The Greyhound Tracks School Newspaper decides to run a story on how many hours per week the average Pope student spent at a pool over the summer. They set up a table in 4th, 5th and 6th lunch and ask every 20th student who enters the cafeteria how many hours they were at the pool between May 26th, 2014 and August 3rd, 2014. After collecting 84 responses, they conclude that the average Pope student spent 5.3 hours per week at a pool.

Population:

Sample:

Statistic:

*Using this study, what could we infer about the time the average Pope student spent at a pool?

*Using this study, could you estimate the number of students who ate lunch that day?

*Using this study, could you estimate the total numbers of hours Pope students spent at the pool per week?

Food for thought - a *statistic* and a *parameter* are very similar. *Statistics* describe samples, *parameters* describe populations. Can you think of a situation where a statistic would be a better measure? Can you think of a situation where a parameter would be a better measure?

Jul 31-3:24 PM

In what ways can we collect data?

3 Primary methods of obtaining data

- Survey.

- Experiment.

- Observational Study.

In what ways can our collection be flawed?

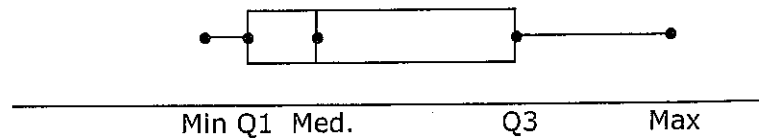
- Bias.

Jul 31-3:34 PM

Jul 31-7:51 PM

How to make a Boxplot (AKA, Box-and-Whisker Plot):

- Put data in numerical order from least to greatest.
- Find the 5-number summary: Min, Q1, Median(Q2), Q3, Max
 - The median is the middle number in the ordered set of data (or the average of the two middle numbers).
 - The lower quartile (Q1) is the median of the lower half of the data (excluding the median).
 - The upper quartile (Q3) is the median of the upper half of the data (excluding the median).
- A boxplot (box-and-whisker plot) is drawn using the 5-number summary as shown below. (The number line below the boxplot should be scaled evenly.)
- The inter-quartile range is $Q3 - Q1$.



In one personality assessment test, a group of questions relate to self-acceptance. A random sample of 15 scores on the self-acceptance portion are

5	20	22	27	30	17	12	15
18	13	12	28	19	9	16	

- Compute the five-number summary and the interquartile range
- Make a boxplot

Practice B

Data Gathering

Decide whether the sampling method could result in a biased sample.

Explain your reasoning.

1. A survey at a school is conducted by asking 100 students chosen randomly from the school population how many times a week they eat in the school cafeteria.

2. A survey of the community is conducted by asking the first 25 persons passing through the entrance gate to the football field if the school should drop the sports programs.

3. A survey is conducted by asking every fifth person boarding a commuter train for one day how many times per week they use public transportation to get to work.

Decide whether the results of the survey are likely to be representative of the population. Explain.

4. The manager of a store that sells DVDs wants to know which types of movies teenagers like to buy. He asks the first 25 teenagers who come into his store to rate comedies, drama, action, and thriller movies on a scale of 1 to 10.

5. A national health and nutrition store chain wants to know if a new line of vitamins would be bought by adults over the age of 21. It takes a survey of 1000 families randomly selected from a nation-wide telephone directory.

6. The campaign manager for a state senator wants to know how many voters are likely to vote for his candidate. He programs the computer to make 100 random phone calls to constituents and asks them to name their choice for senator.

One hundred students out of 1200 at a school have been surveyed. The results are recorded in each problem below. Predict the number of students in the population that would answer similarly.

7. Forty said they attended one sporting event per week.

8. Seventy-five said they ate lunch in the cafeteria.

9. Fourteen said they had an after-school job.

10. Eighty-four said they planned to study math for four years.

Center and Spread of Data

Find the mode, median, mean, lower quartile, upper quartile, interquartile range, and each data set.

1) Test Scores

37	42	48	51	52	53	54
54	55					

2) Mens Heights (Inches)

62	64	69	70	70	71	72
73	74	75	77			

3) Age Assumed Office

Senator	Age	Senator	Age	Senator	Age	Senator	Age	Senator	Age
Patrick Leahy	34	Carl Levin	44	Tammy Baldwin	50	John Barrasso	54	Mike Johanns	58
Mark Pryor	39	Rand Paul	47	Barbara Boxer	52	Kay Hagan	55	John Boozman	60
Brian Schatz	40	John Cornyn	50	Claire McCaskill	53	Jerry Moran	56	Jim Risch	65
John Thune	43								

4) Sales Tax

State	Percent	State	Percent	State	Percent	State	Percent
Colorado	2.9	New Mexico	5.125	Maryland	6	Washington	6.5
Louisiana	4	Maine	5.5	South Carolina	6	Indiana	7
Wyoming	4	Florida	6	Kansas	6.15	New Jersey	7
Oklahoma	4.5	Idaho	6	Massachusetts	6.25	Rhode Island	7
North Dakota	5						

Measures of Central Tendency - Variance and Standard Deviation

Variance is the average squared deviation from the mean of a set of data. It is used to find the **standard deviation**.

Variance and Standard Deviation are more accurate than M.A.D.

To find variance:

- 1) find the mean (average) of the data
- 2) find the difference from EACH data point from the mean. This is called the deviation from the mean.
- 3) square each deviation from the mean
- 4) find the sum of all the squared deviations
- 5) divide the total by (n-1), where n represent the number of data points in your sample.

Aug 8-10:53 AM

Variance Formula

The formula for Variance includes the Sigma Notation (Σ) which represents the sum of all items to the right of it.

$$\frac{\Sigma(x - \bar{x})^2}{n - 1}$$

The sample mean is represented by x-bar and n represents the number of data.

Standard Deviation

Standard Deviation shows the variation in data. *If the data is closely grouped, the standard deviation will be small. If the data is spread out, the standard deviation will be large*

Standard Deviation is denoted using the Greek letter sigma δ or a lower case s.

Aug 8-11:06 AM

To find standard deviation:

- 1) find the variance
- 2) take the square root of the variance

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$$

To calculate variance and/or standard deviation by hand, use this table. We will use the data set {10, 8, 6} with mean = 8

Data Point (x)	x - mean	(x - mean) ²	Sum
10	10 - 8 = 2	2 ² = 4	4
8	8 - 8 = 0	0 ² = 0	4 + 0
6	6 - 8 = -2	-2 ² = 4	4 + 0 + 4
			= 8

$$\text{Variance} = 8/2 = 4$$

$$\text{Standard Deviation} = \sqrt{4} = 2$$

Aug 8-11:13 AM

Example 1:

Find the standard deviation for a group of friends who all just took the same math test.

Their grades were: 92, 88, 80, 68 and 52

Mean:

Deviation for each entry:

Square each deviation:

Sum up all the squared deviations:

Divide the sum of squared deviations by (n-1) - this is the variance.

Find the square root of the variance. This is the S.D.

Aug 8-11:16 AM

What does spread tell us about the data?

Consider another data set from a different class that took the same test. Their test scores were 92, 92, 92, 52 and 52

What is the mean score for this class? What is the standard deviation for this data set?

Can you use what you know to estimate the standard deviation for another data set? Say a third class that took the test had test scores of 77, 76, 76, 76 and 75. What would you predict the standard deviation to be? What could this tell you about the 3 different classes?

Aug 10-2:54 PM

Standard Deviation WS

(and Measures of Central Tendency)

Find the mean, median, variance,, standard deviation, and interquartile range of each set of data using the formulas presented. Graph a boxplot and histogram for each set of data.

1) 6, 22, 4, 15, 14, 8, 8

Mean:

Median:

Variance:

Standard Deviation:

Interquartile Range:

Box Plot:

2) 15, 11, 18, 14, 14, 13, 17, 18

Mean:

Median:

Variance:

Standard Deviation:

Interquartile Range:

Box Plot:

Practice B

Measures of Central Tendency and Variation

Find the mean, median, and mode of each data set.

1. { 12, 11, 17, 3, 9, 14, 16, 2 }

a. Mean _____

b. Median _____

c. Mode _____

2. { 6, 9, 9, 20, 4, 5, 9, 13, 10, 1 }

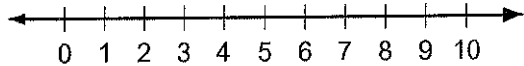
a. Mean _____

b. Median _____

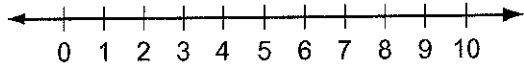
c. Mode _____

Make a box-and-whisker plot of the data. Find the interquartile range.

3. { 3, 7, 7, 3, 10, 1, 6, 6 }



4. { 1, 2, 3, 5, 3, 5, 8, 2 }



Find the variance and standard deviation.

5. { 7, 4, 3, 9, 2 }

6. { 35, 67, 21, 16, 24, 51, 18, 32 }

7. { 19, 23, 17, 20, 25, 19, 15, 22 }

8. { 5, 12, 10, 13, 8, 11, 15, 12 }

Solve.

9. The probability distribution for the amount of rain that falls on Boston in May each year is given below. Find the expected amount of rain for Boston in May.

Inches of Rain, n	5	6	7	8
Probability	0.05	0.10	0.64	0.21

10. A biologist is growing bacteria in the lab. For a certain species of bacteria, she records these doubling times: 41 min, 45 min, 39 min, 42 min, 38 min, 88 min, 43 min, 40 min, 44 min, 39 min, 42 min, and 40 min.

- a. Find the mean of the data. _____
- b. Find the standard deviation. _____
- c. Identify any outliers. _____
- d. Describe how any outlier affects the mean and the standard deviation. _____

Math Award Learning Task: Standard Deviation

Your teacher has a problem and needs your input. She has to give one math award this year to a deserving student, but she can't make a decision. Here are the test grades for her two best students:

Bryce: 90, 90, 80, 100, 99, 81, 98, 82

Brianna: 90, 90, 91, 89, 91, 89, 90, 90

Part 1: Write down which of the two students should get the math award and discuss why they should be the one to receive it.

The *mean* is the average of the data values.

The *mean deviation* (or *mean absolute deviation*) is the average of the differences of each data value from the mean.

The *variance* is the average of the squared differences from the mean.

The *standard deviation* is the square root of the variance and is a measure of how spread out the data values are.

$$\text{mean deviation: } \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

$$\text{variance: } \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

$$\text{standard deviation: } \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

Part 2: Calculate the mean, mean deviation, variance, and standard deviation of both students' distributions. The formulas for mean deviation, variance, and standard deviation are below. Fill out the tables on the back to help you calculate them by hand.

Bryce: 90, 90, 80, 100, 99, 81, 98, 82
Brianna: 90, 90, 91, 89, 91, 89, 90, 90

Calculate the mean of Bryce's distribution.

X_i for Bryce	$X_i - \bar{X}$	$ X_i - \bar{X} $	$(X_i - \bar{X})^2$
90			
90			
80			
100			
99			
81			
98			
82			

Mean deviation for Bryce:

Variance for Bryce:

Standard deviation for Bryce:

What do these measures of spread tell you?

Calculate the mean of Brianna's distribution.

X_i for Brianna	$X_i - \bar{X}$	$ X_i - \bar{X} $	$(X_i - \bar{X})^2$
90			
90			
91			
89			
91			
89			
90			
90			

Mean deviation for Brianna:

Variance for Brianna:

Standard deviation for Brianna:

What do these measures of spread tell you?

Find the mean, median, mode, standard deviation, and 5-number summary using a calculator. Draw a boxplot and histogram using a graphing calculator.

1. In Mrs. Speer's 1st period class, 10 test scores were randomly selected, and the following results were obtained:

74 73 77 77 71 68 65 77 67 66

2. In Mrs. Speer's 2nd period class, 10 test scores were randomly selected, and the following results were obtained:

42 100 77 54 93 85 67 77 62 58

3. The blood alcohol concentrations of 15 drivers involved in fatal accidents and then convicted with jail sentences are given below:

0.27 0.17 0.17 0.16 0.13 0.24 0.29 0.24
0.14 0.16 0.12 0.16 0.21 0.17 0.18

4. The following are the ages of motorcyclists at the time they were fatally injured in traffic accidents:

17 38 27 14 18 34 16 42 28 24 40 20
23 31 37 21 30 25 17 28 33 25 23 19
51 18 29

5. Given below are the ages of US presidents when they were inaugurated:

57 61 57 57 58 57 61 54 68 51 49 64
50 48 65 52 56 46 54 49 51 47 55 55
54 42 51 56 55 51 54 51 60 62 43 55
56 61 52 69 64 46

6. Weights (in pounds) of plastic discarded by households in one week:

2.19 2.10 1.41 0.63 0.92 1.40 1.74 2.87

7. Weights (in pounds) of paper discarded by households in one week:

11.42 15.09 13.61 16.39 8.08 9.46 10.58 11.03

8. Weights (in milligrams) of Bufferin tablets:

672.2 679.2 669.8 672.6 672.2 662.2
662.7 661.3 654.2 667.4 667.0

Probability Distributions and Histograms Notes

Probability Distribution:

Identify which of the following are examples of probability distributions.

Ex. 1:

Number of items on a large pizza

Number of items	0	1	2	3	4
Probability	0.3	0.4	0.2	0.06	0.04

Ex 2:

X	0	2	4	6
P(x)	0.34	0.15	0.41	0.09

Mar 22-9:05 AM

Ex 3:

During the summer students kept track of how early they woke up. Complete the chart to create a probability distribution for what time a student might wake up on summer break.

Time to wake up	# of times recorded by students	P(x)
Before 10:00	12	
Between 10:00 and 11:00	13	
Between 11:00 and 12:00	15	
Between 12:00 and 1:00	10	
After 1:00	7	

- a) What is the probability that a student wakes up before 10:00?
- b) What is the probability that a student wakes up after 11:00?

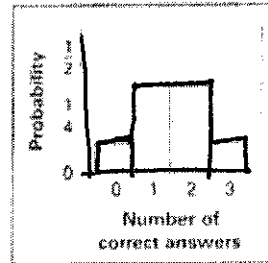
Mar 22-9:07 AM

Probability Histogram:

Ex 4:

The following histogram represents the probability distribution of the number of true/false questions that students guessed correctly

X (number correct)	0	1	2	3
Outcomes	1	3	3	1
P(X)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$



- a) What is the probability a student guessed 1 correct answer?
- b) What is the probability a student guessed at least one correct answer?

Mar 22-9:08 AM

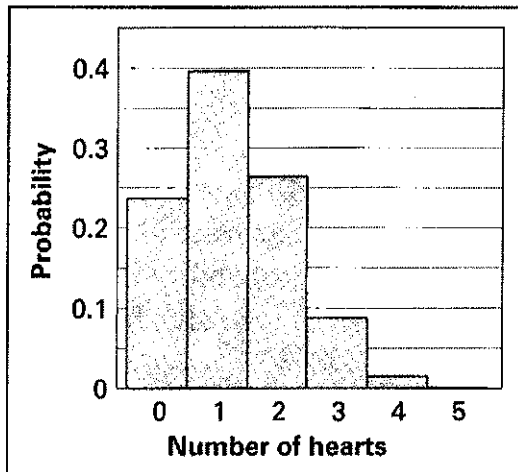
Ex 5:

Use the data to construct a table and histogram showing the probability distribution of the number of cell phones per house.

X (number of cell phones)	0	1	2	3
Number of households	19	28	37	16

Mar 22-9:09 AM

1. A student conducted an experiment using a standard deck of 52 cards. The student chose a card at random, noted if it was a heart then replaced it. The student conducted this experiment 5 times. The resulting histogram is below.



a) What is the probability that the student selected two hearts?

b) What is the probability that the student selected at least three hearts?

In the following problems, determine whether a probability distribution is given. In those cases where a probability distribution is not described, identify the requirement that is not satisfied. In those cases where a probability distribution is described, create a histogram of the probability model.

2. A vendor supplies refreshments at a baseball stadium and must plan for the possibility of a World Series contest. In the accompanying table, x represents the number of baseball games required to complete a World Series contest.

x	4	5	6	7
$P(x)$	0.120	0.253	0.217	0.410

3. The Young Fun Company operates a national toy store chain that depends heavily on the size of the child population, and research is being conducted to learn more about the number of children in families. In the accompanying table, x represents the number of children in families.

x	0	1	2	3
$P(x)$	0.48	0.21	0.19	0.08

4. To settle a paternity suit, two different people are given blood tests. If x is the number having group A blood, then x can be 0, 1, or 2 and the corresponding probabilities are 0.36, 0.48, and 0.16, respectively.

5. Car headlight manufacturers are concerned about failure rates. One headlight failure is an inconvenience, but if both lights fail, you can't drive at night. Assume that the probabilities of 0, 1, and 2 failures are 0.96, 0.036, and 0.004, respectively.

6. United Airlines Flight 470 from Denver to St. Louis has an on-time performance described as follows: For 4 independent flights, the probabilities for 0, 1, 2, 3, and 4 on-time flights are 0.026, 0.345, 0.346, 0.154, and 0.129, respectively.

7. The Baltimore Computer House finds that the probabilities of selling 0, 1, 2, 3, and 4 microcomputers in one day are .245, .370, .210, .095, and .080, respectively.