Section 1.6 The Normal Distribution

Get Started – What is a concept map?

- How do I apply the 68-95-99.7 rule to a set of data?
- What is the relationship between the area under a normal curve and z-scores?

Get Started – What is a concept map?

Key Terms

Concept map

Summary

When making a concept map there is no right way to do it. You want to make connections that make sense to you. You might want to include examples or pictures, anything that helps you make sense of what you are studying.

Flash cards can be a great way to learn facts, but they don't usually lead to knowledge or understanding. Taking those flash cards and turning them into a concept map will help you obtain knowledge.

Steps for making a Concept Map

1. **Make a list of all the topics you are studying.** Looking through your text book for words in bold type, looking at the heading of the pages, thinking about what you've been studying are all good ways to start. Reviewing homework and quizzes (in-class and Canvas quizzes) will also help.
2. **Put each topic on a separate piece of paper.** Index cards work well but aren't required. Any scraps of paper will work.
3. **Start by trying to group them in just a few general categories.** It is OK to make a category for topics you don't know what to do with. But don't get lazy and put every topic in that category. In the example on the previous page our categories were Types of Data and Measures of Center.
4. **Pick one group and organize just those topics.** How do they connect to each other?
5. **Repeat step 4 with each group.**
6. **Try to connect each subgroup with the other groups.** You'll be surprised how many connections you can find when you look for them.
7. **Take a picture of your concept map.**

When studying for big exams making, and remaking, concept maps are a great way to review the material.

As we work through the semester you should make a concept map for each chapter and then try to link the maps for each chapter into one big map. If you can do this, you will be READY for the final exam.

Notes

How do I apply the 68-95-99.7 rule to a set of data?

Key Terms

Normal distribution   Empirical rule (68-95-99.7 rule

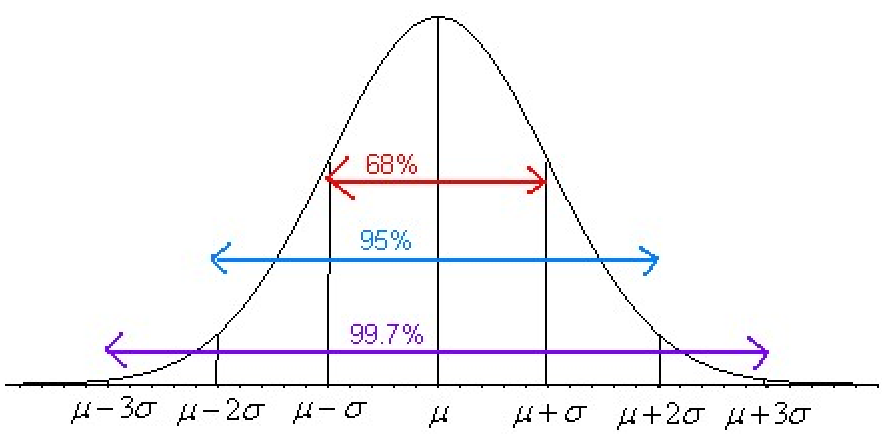
Summary

There are many different types of distributions (shapes) of quantitative data. In section 1.5 we looked at different histograms and described the shapes of them as symmetric, skewed left, and skewed right. There is a special symmetric shaped distribution called the normal distribution. It is high in the middle and then goes down quickly and equally on both ends. It looks like a bell, so sometimes it is called a bell curve. One property of the normal distribution is that it is symmetric about the mean. Another property has to do with what percentage of the data falls within certain standard deviations of the mean. This property is defined as the empirical Rule.

The Empirical Rule: Given a data set that is approximately normally distributed:

- Approximately 68% of the data is within one standard deviation of the mean.
- Approximately 95% of the data is within two standard deviations of the mean.
- Approximately 99.7% of the data is within three standard deviations of the mean.

To visualize these percentages, see the following figure.



Note: The empirical rule is only true for approximately normal distributions.

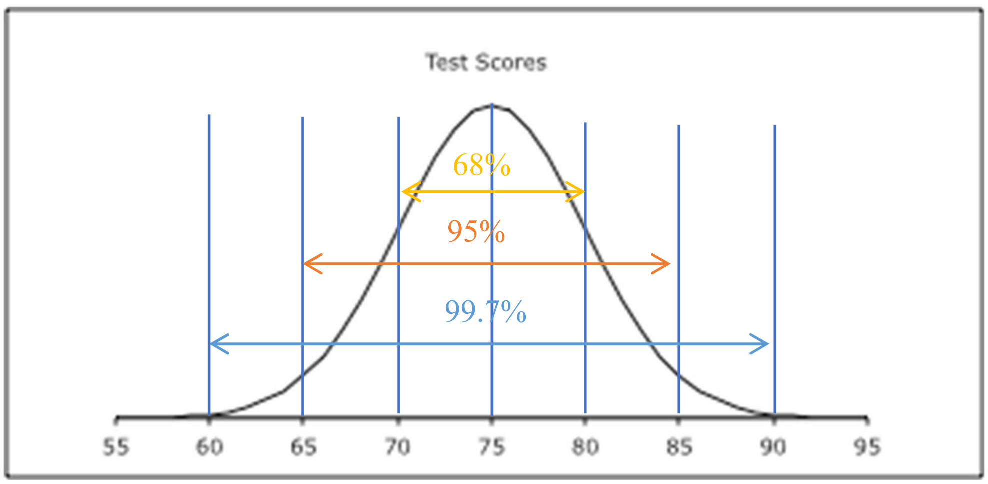
Notes

Guided Example 1

Suppose that your class took a test and the mean score was 75% and the standard deviation was 5%. If the test scores follow an approximately normal distribution, answer the following questions using the empirical rule (68-95-99.7 rule).

a. What percentage of the students had scores between 65 and 85?

Solution To solve this problem, it would be helpful to draw the normal curve that follows this situation. The mean is 75, so the center is 75. The standard deviation is 5, so for each line above the mean add 5 and for each line below the mean subtract 5. The graph looks like the following:



From the graph we can see that 95% of the students had scores between 65 and 85.

b. What percentage of the students had scores between 65 and 75?

Solution Using the graph above, the scores of 65 to 75 are half of the area of the graph from 65 to 85. Because of symmetry, that means that the percentage for 65 to 85 is ½ of the 95%, which is 47.5%.

c. What percentage of the students had scores between 70 and 80?

Solution From the graph we can see that 68% of the students had scores between 70 and 80.

d. What percentage of the students had scores above 85?

Solution For this problem we need a bit of math. If you looked at the entire curve, you would say that 100% of all the test scores fall under it. So, because of symmetry 50% of the test scores fall in the area above the mean and 50% of the test scores fall in the area below the mean. We know from part b that the percentage from 65 to 75 is 47.5%. Because of symmetry, the percentage from 75 to 85 is also 47.5%. So, the percentage above 85 is 50% - 47.5% = 2.5%.

Practice 1

A normal distribution has a mean of 25 and a standard deviation of 3. Use the empirical rule (68-95-99.7) rule to answer the questions below.

a. What is the percentage of values between 22 and 28?

b. What is the percentage of values between 22 and 25?

c. What is the percentage of values between 16 and 34?

d. What is the percentage of values below 22?

Guided Example 2	Practice 2
When consumer goods are packaged, the amount of goods in each package is not exactly what is labeled on the package. Suppose the weights (in ounces) of chips in a bag are normally distributed with a mean of 12 ounces and a standard deviation of 1 ounce. Use the empirical rule (68-95-99.7 rule) to answer each question below. a. What percentage of bags have less than 12 ounces? **Solution** Since 12 ounces is the mean, the percentage of bags below the mean will be 50%. This also means the percentage of bags above the mean is also 50% since the mean splits the normal distribution into two symmetrical halves. b. What percentage of bags have less than 10 ounces? **Solution** A weight of 10 ounces is two standard deviations below the mean. From the empirical rule, we know that 95% of bags will be between two standard deviations below the mean and two standard deviations above the mean. So, $\frac{95}{2}$ or 47.5% of bags will be between two standard deviations below the mean and the mean. The bags below two standard deviations is calculated by taking all of the bags below the mean (50%) and subtracting the bags between two standard deviation below the mean and the mean (47.5%). This tells us that 50% - 47.5% or 2.5% have less than 10 ounces.	When consumer goods are packaged, the amount of goods in each package is not exactly what is labeled on the package. Suppose the weights (in ounces) of cookies in a bag are normally distributed with a mean of 20 ounces and a standard deviation of 2 ounces. Use the empirical rule (68-95-99.7 rule) to answer each question below. a. What percentage of bags have less than 20 ounces? b. What percentage of bags have less than 18 ounces?

What is the relationship between the area under a normal curve and z-scores?

Key Terms

Raw score z score

Summary

When we look at Guided Examples 1, we realize that the numbers on the scale are not as important as how many standard deviations a number is from the mean. As an example, the number 80 is one standard deviation from the mean. The number 65 is 2 standard deviations from the mean. However, 80 is above the mean and 65 is below the mean. Suppose we wanted to know how many standard deviations the number 82 is from the mean. How would we do that? The other numbers were easier because they were a whole number of standard deviations from the mean. We need a way to quantify this. We will use a z-score (also known as a z-value or standardized score) to measure how many standard deviations a data value is from the mean. This is defined as:

$$z = \frac{x - \mu}{\sigma}$$

where x = data value (raw score)

z = standardized value (z-score or z-value)

μ = population mean

σ = population standard deviation

Note: Remember that the z-score is always how many standard deviations a data value is from the mean of the distribution.

Suppose a data value has a z-score of 2.13. This tells us two things. First, it says that the data value is above the mean, since it is positive. Second, it tells us that you must add more than two standard deviations to the mean to get to this value. Since most data (95%) is within two standard deviations, then anything outside this range would be considered a strange or unusual value. A z-score of 2.13 is outside this range so it is an unusual value. As another example, suppose a data value has a z-score of -1.34. This data value must be below the mean, since the z-score is negative, and you need to subtract more than one standard deviation from the mean to get to this value. Since this is within two standard deviations, it is an ordinary value.

An **unusual value** has a z-score < -2 or a z-score > 2

A **usual value** has a z-score between -2 and 2, that is $-2 < z - score < 2$.

You may encounter standardized scores on reports for standardized tests or behavior tests as mentioned previously.

Notes

<table>
<tr><th>Guided Example 3</th><th>Practice 3</th></tr>
<tr><td>Suspect that your class took a test the mean score was 75% and the standard deviation was 5%. If test scores follow an approximately normal distribution, answer the following questions:

a. If a student earned 87 on the test, what is that student's z-score and what does it mean?

Solution The problem tells us that $\mu = 75$, $\sigma = 5$, and $x = 87$. Put these values into the formula for z-scores and we get

$z = \frac{x-\mu}{\sigma} = \frac{87-75}{5} = 2.40$</td><td>Suppose that your class took a test the mean score was 65% and the standard deviation was 10%. If test scores follow an approximately normal distribution, answer the following questions:

a. If a student earned 92 on the test, what is that student's z-score and what does it mean?</td></tr>
</table>

This means that the score of 87 is more than two standard deviations above the mean, and so it is considered to be an unusual score. b. If a student earned 73 on the test, what is that student's z-score and what does it mean? **Solution** The problem tells us that $\mu = 75$, $\sigma = 5$, and $x = 87$. Put these values into the formula for z-scores and we get $z = \frac{x-\mu}{\sigma} = \frac{73-75}{5} = -0.40$ This means that the score of 73 is less than one-half of a standard deviation below the mean. It is considered to be a usual or ordinary score.	b. If a student earned 52 on the test, what is that student's z-score and what does it mean?

Guided Example 4 | Practice 4

Suppose that your class took a test the mean score was 75% and the standard deviation was 5%. If test scores follow an approximately normal distribution, answer the following questions: a. If a student has a z-score of 1.43, what actual score did she get on the test? **Solution** This problem involves a little bit of algebra. Do not worry, it is not that hard. Since you are now looking for *x* instead of *z*, rearrange the equation solving for *x* as follows:	Suppose that your class took a test the mean score was 65% and the standard deviation was 10%. If test scores follow an approximately normal distribution, answer the following questions: a. If a student has a z-score of 1.2, what actual score did she get on the test?

$z = \frac{x-\mu}{\sigma}$ $z \cdot \sigma = \frac{x-\mu}{\not\sigma} \cdot \not\sigma$ Multiply both sides by σ and simplify $z\sigma = x - \mu$ $z\sigma + \mu = x - \not\mu + \not\mu$ Add μ to both sides and simplify. $x = z\sigma + \mu$ Rearrange to get x on the left side. Now, you can use this formula to find *x* when you are given *z*. $x = z\sigma + \mu$ $x = 1.43 \cdot 5 + 75$ $x = 7.15 + 75$ $x = 82.15$ Thus, the *z*-score of 1.43 corresponds to an actual test score of 82.15%.	
b. If a student has a z-score of -2.34, what actual score did he get on the test? **Solution** Use the formula for *x* from part d of this problem: $x = z\sigma + \mu$ $x = -2.34 \cdot 5 + 75$ $x = -11.7 + 75$ $x = 63.3$ Thus, the z-score of -2.34 corresponds to an actual test score of 63.3%.	b. If a student has a z-score of -1.9, what actual score did he get on the test?

Looking at the Empirical Rule, 99.7% of all data is within three standard deviations of the mean. This means that an approximation for the minimum value in a normal distribution is the mean minus three times the standard deviation, and for the maximum is the mean plus three times the standard deviation. In a normal distribution, the mean and median are the same. Lastly, the first quartile can be approximated by subtracting 0.67448 times the standard deviation from the mean, and the third quartile can be approximated by adding 0.67448 times the standard deviation to the mean. All these together give the five-number summary.

In mathematical notation, the five-number summary for the normal distribution with mean μ and standard deviation σ is as follows:

Five-Number Summary for a Normal Distribution

$$\begin{aligned} \text{min} &= \mu - 3\sigma \\ Q_1 &= \mu - 0.67448\sigma \\ \text{med} &= \mu \\ Q_3 &= \mu + 0.67448\sigma \\ \text{max} &= \mu + 3\sigma \end{aligned}$$

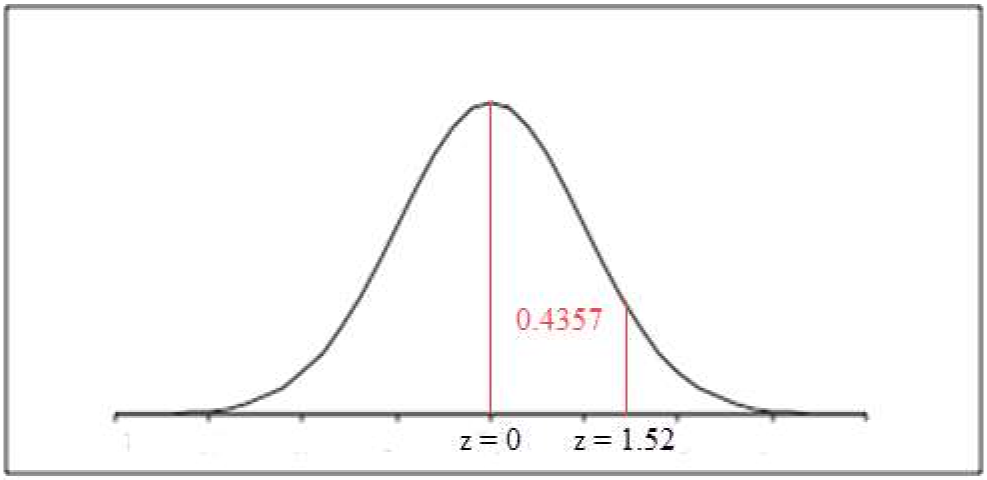
Guided Example 5	Practice 5
Suppose that your class took a test and the mean score was 75% and the standard deviation was 5%. If the test scores follow an approximately normal distribution, find the five-number summary. **Solution** The mean is $\mu = 75\%$ and the standard deviation is $\sigma = 5\%$. Thus, the five-number summary for this problem is: $\text{min} = 75 - 3(5) = 60\%$ $Q1 = 75 - 0.67448(5) \approx 71.6\%$ $\text{med} = 75\%$ $Q3 = 75 + 0.67448(5) \approx 78.4\%$ $\text{max} = 75 + 3(5) = 90\%$	Suppose that your class took a test and the mean score was 65% and the standard deviation was 10%. If the test scores follow an approximately normal distribution, find the five-number summary.

The empirical rule helps us to calculate the percentage of data values when the data values fall on one, two, or three standard deviations above or below the mean. These data values correspond to z = ±1, z = ±2, and z = ±3. However, what if a data value falls between these z-scores? In these cases, we use a table of normal curve areas to find percentages. A copy of this table is shown below:

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3304	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999

This table indicates the percentage of data values between the mean and a *z*-value in the table.

For instance, the area in the table colored red tells us two things. The row the colored cell is in indicates the z-score begins with 1.5. The second decimal comes from the column the colored cell is in, .02. Putting this together with the number in the colored cell tells us that a z-score of 1.52 corresponds to the percentage 0.4357.



This tells us the 43.57% of data values are between the mean and $z = 1.52$.

Note that since the normal distribution is symmetric, the normal curve areas can be used for positive and negative *z*-scores.

Guided Example 6	Practice 6
Use a Standard Normal Distribution table to find the percentage of values from $z = 0$ to $z = -1.82$. **Solution** Using the table, locate the row for $z = 1.8$. Now locate the column for .02. The row and column intersect at z = 1.82. Recalling that the table also corresponds to negative z values enables us to read the percentage of values from $z = 0$ to $z = -1.82$ as 0.4656 or 46.56%.	Use a Standard Normal Distribution table to find the percentage of values from $z = 0$ to $z = 0.5$.

Guided Example 7	Practice 7
Use a Standard Normal Distribution table to find the percentage of values from $z = 0.4$ to $z = 1.2$. Solution Start by finding the percentage of values using the table for $z = 1.2$. This percentage is 38.49%. This is the percentage of values from the mean to $z = 1.2$. Now find the percentage for z = 0.4. This s read from the table as 0.1554 or 15.54%. The difference between these percentages, 38.49% - 15.54% or 22.95% is the percentage of values between $z = 0.4$ and $z = 1.2$.	Use a Standard Normal Distribution table to find the percentage of values from $z = 0.5$ to $z = 0.8$.

Guided Example 8	Practice 8
Use a Standard Normal Distribution table to find the percentage of values under the standard normal curve that is above $z = 1.59$. Solution The percentage of the values above the mean is 50%. If we subtract the percentage of values from the mean to z = 1.59 from 50%, we will get the percentage above $z = 1.59$. The table gives the percentage between the mean and $z =$ 1.59 as 44.41%. So, the percentage above z = 1.59 is 50% - 44.41% or 5.59%.	Use a Standard Normal Distribution table to find the percentage of area under the standard normal curve that is below $z = -1.25$.