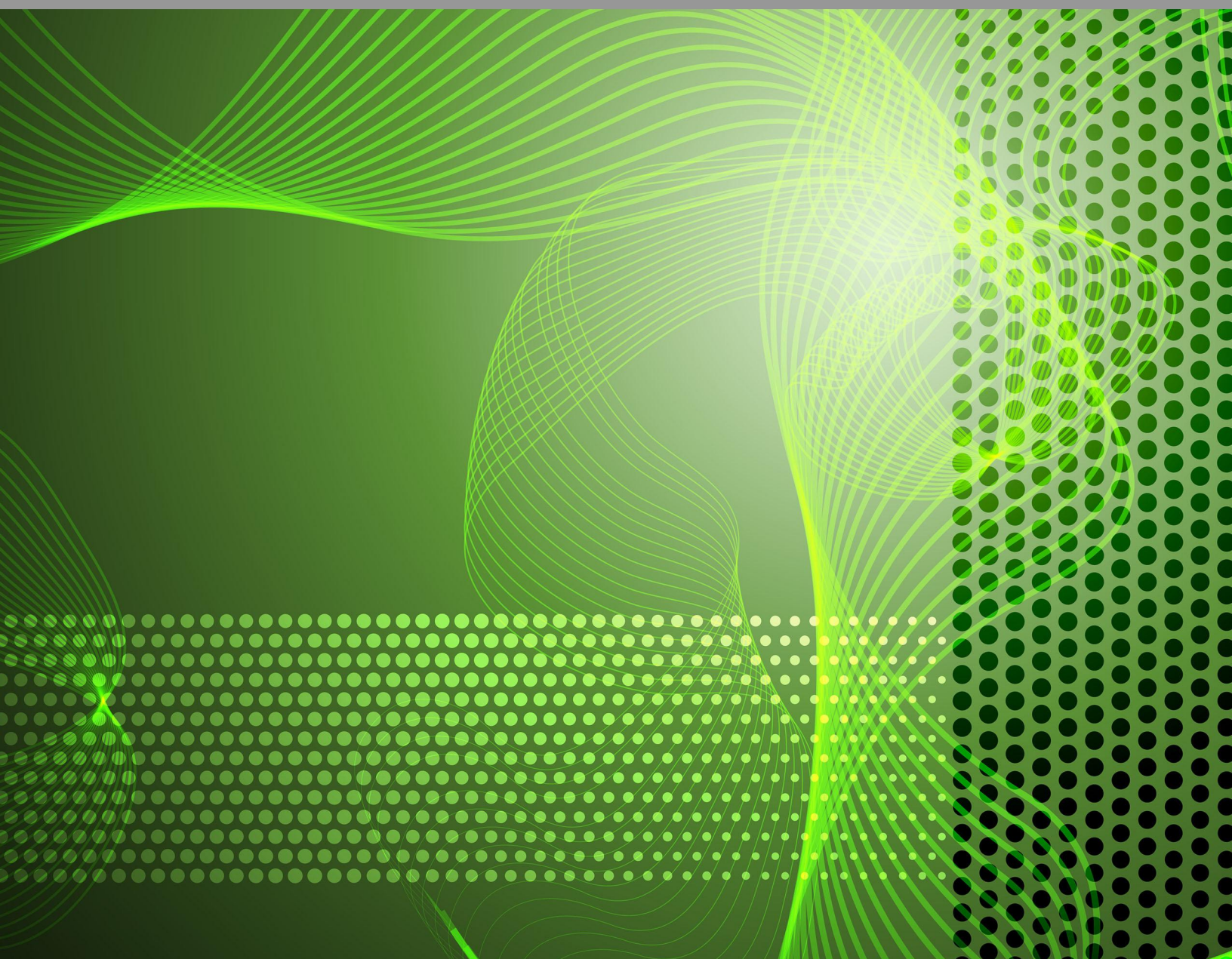


ck-12

flexbook
next generation textbooks

CK-12 Probability and Statistics

Concept Collection



CK-12 Probability and Statistics Concepts

Larame Spence

Say Thanks to the Authors

Click <http://www.ck12.org/saythanks>

(No sign in required)



CHAPTER

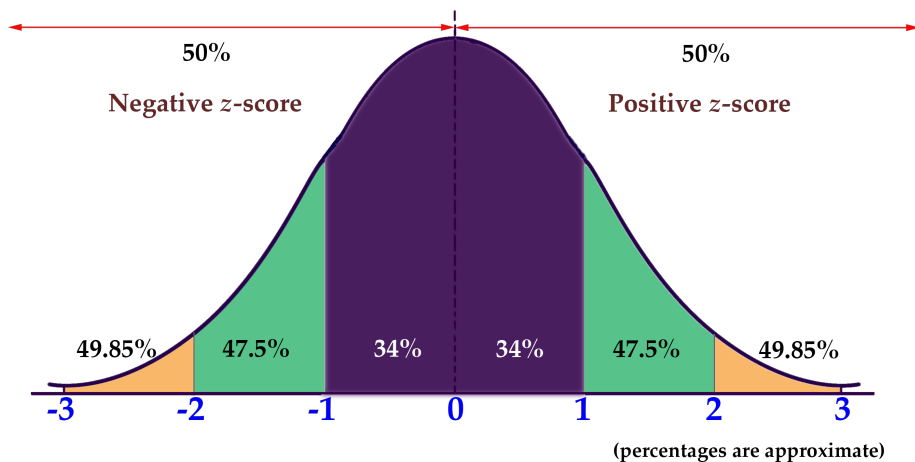
9

The Normal Distribution

Chapter Outline

- 9.1 UNDERSTANDING NORMAL DISTRIBUTION
- 9.2 THE EMPIRICAL RULE
- 9.3 Z-SCORES
- 9.4 Z SCORES II
- 9.5 Z-SCORES III
- 9.6 THE MEAN OF MEANS
- 9.7 CENTRAL LIMIT THEOREM
- 9.8 APPROXIMATING THE BINOMIAL DISTRIBUTION
- 9.9 REFERENCES

The normal distribution can be found practically everywhere, from the distribution of human heights to IQ scores. Understanding the application of the normal distribution can simplify the calculation of binomial probabilities, such as the probability of flipping heads 37 or more times out of 60. By using the Central Limit Theorem, you can evaluate the probabilities associated with many different random variables.



9.1 Understanding Normal Distribution

Objective

Here you will learn about the Normal Distribution. You will learn what it is and why it is important, and you will begin to develop an intuition for the rarity of a value in a set by comparing it to the mean and standard deviation of the data.

Concept



If you knew that the prices of t-shirts sold in an online shopping site were *normally distributed*, and had a mean cost of \$10, with a standard deviation of \$1.50, how could that information benefit you as you are looking at various t-shirt prices on the site? How could you use what you know if you were looking to make a profit by purchasing unusually inexpensive shirts to resell at prices that are more common?

Watch This



MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/67462>

<http://youtu.be/47099905> CK-12.org - Chapter6SpreadofaNormalDistributionA

Guidance

A *distribution* is an evaluation of the way that points in a data set are clustered or spread across their *range* of values. A *normal distribution* is a very specific symmetrical distribution that indicates, among other things, that exactly $\frac{1}{2}$ of the data is below the mean, and $\frac{1}{2}$ is above, that approximately 68% of the data is within 1, approximately 96% of the data is within 2, and approximately 99.7% is within 3 *standard deviations* of the mean.

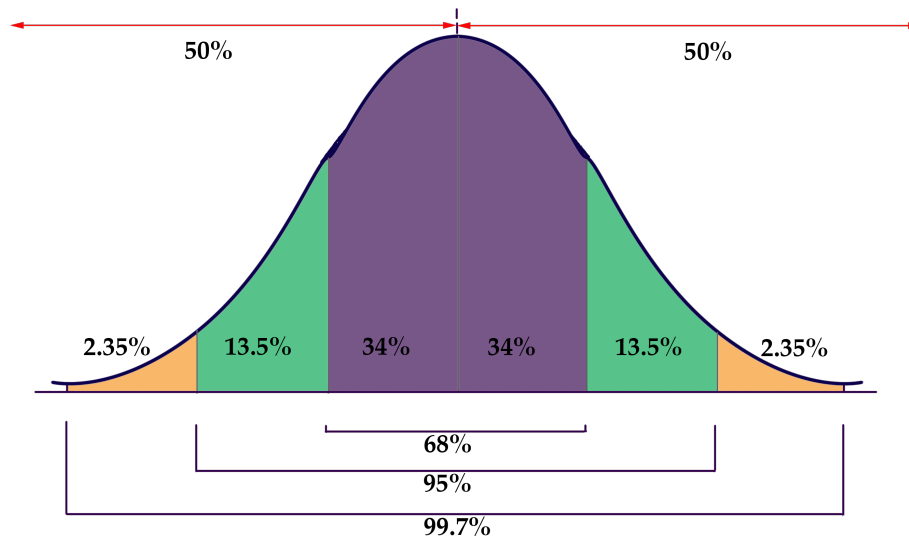
There are a number of reasons that it is important to become familiar with the normal distribution, as you will discover throughout this chapter. Examples of values associated with normal distribution:

- Physical characteristics such as height, weight, arm or leg length, etc.
- The percentile rankings of standardized testing such as the ACT and SAT
- The volume of water produced by a river on a monthly or yearly basis
- The velocity of molecules in an ideal gas

Knowing that the values in a set are exactly or approximately normally distributed allows you to get a feel for how common a particular value might be in that set. Because the values of a normal distribution are predictably clustered around the mean, you can estimate in short order the rarity of a given value in the set. In our upcoming lesson on the Empirical Rule, you will see that it is worth memorizing that normally distributed data has the characteristics mentioned above:

- 50% of all data points are above the mean and 50% are below
- Apx 68% of all data points are within 1 standard deviation of the mean
- Apx 95% of all data points are within 2 standard deviations of the mean
- Apx 99.7% of all data points are within 3 standard deviations of the mean

In this lesson, we will be practicing a 'rough estimate' of the probability that a value within a given range will occur in a particular set of data, just to develop an intuition of the use of a normal distribution. In subsequent lessons, we will become more specific with our estimates. The image below will be used in greater detail in the lesson on the Empirical Rule, but you may use it as a reference for this lesson also.



Example A

Human height is commonly considered an approximately normally distributed measure. If the mean height of a male adult in the U.S.A. is 5'10", with a standard deviation of 1.5", how common are men with heights greater than 6'2"?

Solution:

Since each standard deviation of this normally distributed data is 1.5", and 6'2" is 4" above the mean for the population, 6'2" is nearly 3 standard deviations above the mean. That tells us that men taller than 6'2" are quite rare in this population.

Example B

If the fuel mileage of a particular model of car is normally distributed, with a mean of 26 mpg and a standard deviation of 2 mpg, how common are cars with a fuel efficiency of 24 to 25 mpg?



Solution:

We know that apx 68% of the cars in the population have an efficiency of between 24 and 28 mpg, since that would be 1 SD below and 1 SD above the mean. That suggests that apx 34% have an efficiency of 24 to 26 mpg, so we can say that it is uncommon to see a car with an efficiency between 24 and 25 mpg, but not extremely so.

Example C

If the maximum jumping height of U.S. high school high jumpers is normally distributed with a mean of 5'11.5" and a SD of 2.2", how unusual is it to see a high school jumper clear 6'3"?

Solution:

If the mean is 5'11.5", then 1 SD above is 6'1.7" and 2 SD's is 6'3.9". That means that less than 2.5% of jumpers 6'3.9", so it would be pretty uncommon to see a high-school competitor exceed 6'3".

Concept Problem Revisited

*If you knew that the prices of t-shirts sold in an online shopping site were **normally distributed**, and had a mean cost of \$10, with a standard deviation of \$1.50, how could that information benefit you as you are looking at various t-shirt styles and designs on the site? How could you use what you know if you were looking to make a profit by purchasing unusually inexpensive shirts to resell at prices that are more common?*

By knowing the mean and SD of the shirt prices, and knowing that they are normally distributed, you can estimate right away if a shirt is priced at a point significantly below the norm. For instance, with this data, we can estimate that a shirt priced at \$7.00 is less expensive than apx 97.5% of all shirts on the site, and could likely be resold at a profit (assuming there is not something wrong the shirt that is not obvious from the listing).

Vocabulary

A **distribution** is an arrangement of values of a variable showing their observed or theoretical frequency of occurrence.

The **range** of values of a distribution is the difference between the least and greatest values.

The **normal distribution** is a very specific distribution that is symmetric about its mean. Half the values of the random variable are below the mean and half are above the mean. Approximately 68% of the data is within 1 **standard deviation** of the mean, apx 96% is within 2 SD's, and 99.7% within 3 SD's.

A **standard deviation** is measure of how spread out the data is from the mean. To determine if a data value is far from the mean, determine how many standard deviations it is from the mean. The SD is calculated as the square root of the variance.

Guided Practice

For questions 1-4, assume the data to be normally distributed, and describe the rarity of an event using the following scale: 0% – < 1% probability = very rare, 1% – < 5% = rare, 5% – < 34% = uncommon, 34% – < 50% = common, 50% – 100% = likely.

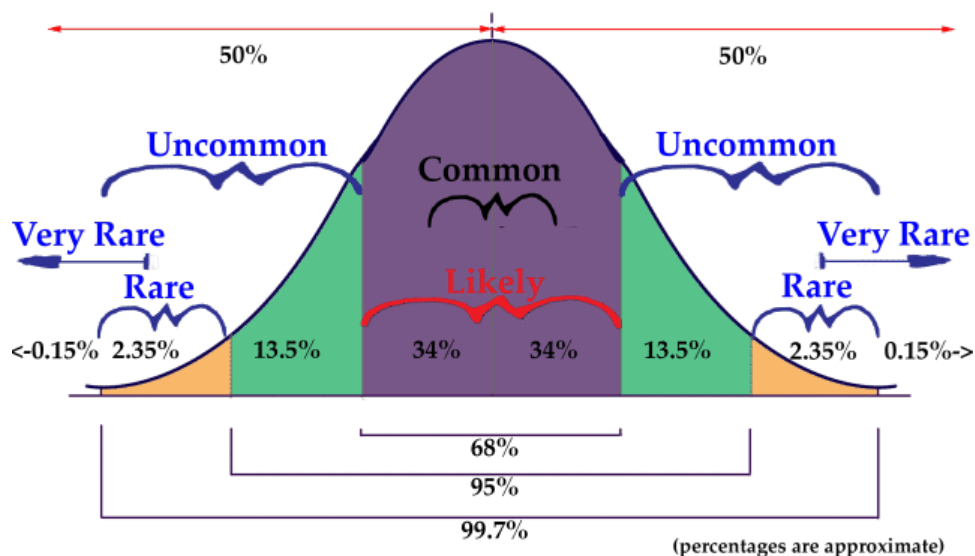
1. If the mean (μ) of the data is 75, and the standard deviation (σ) is 5, how common is a value between 70 and 75?
2. If the μ is .02 and the σ is .005, how common is a value between .005 and .01?
3. If the μ is 1280 and the σ is 70, how common is a value between 1210 and 1350?
4. If the mean defect rate at a cellphone production plant is .1%, with a standard deviation of .03%, would it seem reasonable for a quality assurance manager to be concerned about 3 defective phones in a single 1000 unit run?

Solutions:

1. A value of 70 is only 1 standard deviation below the mean, so a value between 70 and 75 would be expected approximately 34% of the time, so it would be **common**.
2. A value of .01 is 2 SD's below the mean, and .005 is 3 SD's below, so we would expect there to be about a 2.5% probability of a value occurring in that range. A value between 0.005 and 0.01 would be **rare**.
3. 1210 is 1 SD *below* the mean, and 1350 is 1 SD *above* the mean, so we would expect approximately 68% of the data to be in that range, meaning that it is **likely** that a value in that range would occur.
4. .1% translates into 1 per thousand, with a standard deviation of 3 per ten thousand. That means that 3 defects in the same thousand is nearly 7 SD's above the mean, well into the **very rare** category. While it is not impossible for random chance to result in such a value, it would certainly be prudent for the manager to investigate.

Practice

Assume all sets/populations to be approximately normally distributed, and describe the rarity of an event using the following scale: 0% – < 1% probability = very rare, 1% – < 5% = rare, 5% – < 34% = uncommon, 34% – < 50% = common, 50% – 100% = likely. You may reference the image below:



1. Scores on a certain standardized test have a mean of 500, and a standard deviation of 100. How common is a score between 600 and 700?

2. Considering a full-grown show-quality male Siberian Husky has a mean weight of 52.5 lbs, with SD of 7.5 lbs, how common are male huskies in the 37.5 - 45 lbs range?
3. A population $\mu = 125$, and $\sigma = 25$, how common are values in the 100 - 150 range?
4. Population $\mu = 0.0025$ and $\sigma = 0.0005$, how common are values between 0.0025 and 0.0030?
5. A 12 oz can of soda has a mean volume of 12 oz, with a standard deviation of .25 oz. How common are cans with between 11 and 11.5 oz of soda?
6. $\mu = 0.0025$ and $\sigma = 0.0005$, how common are values between 0.0045 and 0.005?
7. If a population $\mu = 1130$ and $\sigma = 5$, how common are values between 0 and 1100?
8. Assuming population $\mu = 1130$ and $\sigma = 5$, how common are values between 1125 and 1135?
9. The American Robin Redbreast has a mean weight of 77 g, with a standard deviation of 6 g. How common are Robins in the 59 g – 71 g range?
10. Population $\mu = \frac{3}{5}$ and $\sigma = \frac{1}{10}$, how common are values between $\frac{2}{5}$ and 1?
11. Population $\mu = 0.25\%$ and $\sigma = 0.05\%$, how common are values between 0.35% and 0.45%?
12. Population $\mu = 156.5$ and $\sigma = 0.25$, how common are values between 155 and 156?

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.1.

9.2 The Empirical Rule

Objective

Here you will learn how to use the Empirical Rule to estimate the probability of an event.

Concept

If the price per pound of USDA Choice Beef is normally distributed with a mean of \$4.85/lb and a standard deviation of \$0.35/lb, what is the estimated probability that a randomly chosen sample (from a randomly chosen market) will be between \$5.20 and \$5.55 per pound?

Watch This



MEDIA

Click image to the left or use the URL below.

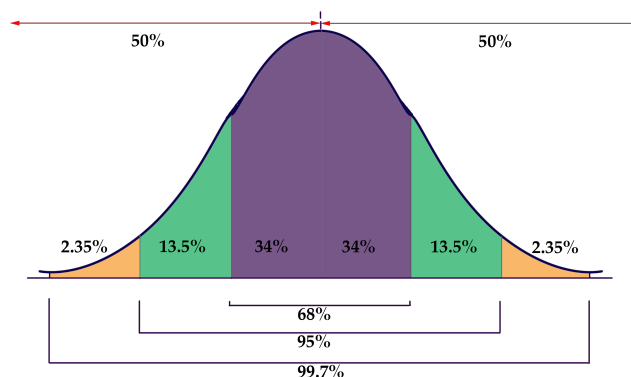
URL: <https://www.ck12.org/flx/render/embeddedobject/67465>

<http://vimeo.com/47111393> CK-12 Foundation: Chapter6EmpiricalRuleA

Guidance

This lesson on the *Empirical Rule* is an extension of the previous lesson “Understanding the Normal Distribution”. In the prior lesson, the goal was to develop an intuition of the interaction between decreased probability and increased distance from the mean. In this lesson, we will practice applying the Empirical Rule to estimate the specific probability of occurrence of a sample based on the range of the sample, measured in standard deviations.

The graphic below is a representation of the Empirical Rule:



The graphic is a rather concise summary of the 'vital statistics' of a Normal Distribution. Note how the graph resembles a bell? Now you know why the normal distribution is also called a “*bell curve*”.

- 50% of the data is above, and 50% below, the mean of the data
- Approximately 68% of the data occurs within 1 SD of the mean
- Approximately 95% occurs within 2 SD's of the mean
- Approximately 99.7% of the data occurs within 3 SD's of the mean

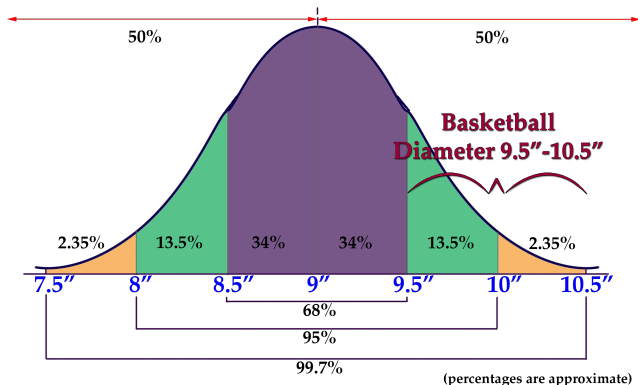
It is due to the probabilities associated with 1, 2, and 3 SD's that the Empirical Rule is also known as the “68 – 95 – 99.7 rule”.

Example A

If the diameter of a basketball is normally distributed, with a mean (μ) of 9", and a standard deviation (σ) of 0.5", what is the probability that a randomly chosen basketball will have a diameter between 9.5" and 10.5"?

Solution:

Since the $\sigma = 0.5''$ and the $\mu = 9''$, we are evaluating the probability that a randomly chosen ball will have a diameter between 1 and 3 standard deviations above the mean. The graphic below shows the portion of the normal distribution included between 1 and 3 SD's:



The percentage of the data spanning the 2nd and 3rd SD's is 13.5% + 2.35% = 15.85%

The probability that a randomly chosen basketball will have a diameter between 9.5 and 10.5 inches is 15.85%.

Example B

If the depth of the snow in my yard is normally distributed, with $\mu = 2.5''$ and $\sigma = .25''$, what is the probability that a randomly chosen location will have a snow depth between 2.25 and 2.75 inches?



Solution:

2.25 inches is $\mu - 1\sigma$, and 2.75 inches is $\mu + 1\sigma$, so the area encompassed approximately represents $34\% + 34\% = 68\%$.

The probability that a randomly chosen location will have a depth between 2.25 and 2.75 inches is 68%.

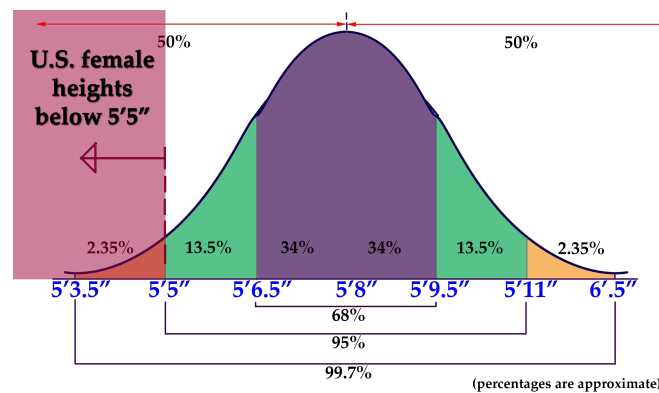
Example C

If the height of women in the U.S. is normally distributed with $\mu = 5'8''$ and $\sigma = 1.5''$, what is the probability that a randomly chosen woman in the U.S. is shorter than 5'5''?

Solution:

This one is slightly different, since we aren't looking for the probability of a limited range of values. We want to evaluate the probability of a value occurring anywhere below 5'5''. Since the domain of a normal distribution is infinite, we can't actually state the probability of the portion of the distribution on 'that end' because it has no 'end'! What we need to do is add up the probabilities that we do know and subtract them from 100% to get the remainder.

Here is that normal distribution graphic again, with the height data inserted:



Recall that a normal distribution always has 50% of the data on each side of the mean. That indicates that 50% of U.S. females are taller than 5'8'', and gives us a solid starting point to calculate from. There is another 34% between 5'6.5'' and 5'8'' and a final 13.5% between 5'5'' and 5'6.5''. Ultimately that totals: $50\% + 34\% + 13.5\% = 87.5\%$.

Since 87.5% of U.S. females are 5'5'' or taller, that leaves 12.5% that are less than 5'5'' tall.

Concept Problem Revisited

If the price per pound of USDA Choice Beef is normally distributed with a mean of \$4.85/lb and a standard deviation of \$0.35/lb, what is the estimated probability that a randomly chosen sample (from a randomly chosen market) will be between \$5.20 and \$5.55 per pound?

\$5.20 is $\mu + 1\sigma$, and \$5.55 is $\mu + 2\sigma$, so the probability of a value occurring in that range is approximately 13.5%.

Vocabulary

A **normal distribution** is a common, but specific, distribution of data with a set of characteristics detailed in the lesson above.

The **Empirical Rule** is a name for the way in which the normal distribution divides data by standard deviations: 68% within 1 SD, 95% within 2 SD's and 99.7 within 3 SD's of the mean.

The "**68-95-99.7 rule**" is another name for the Empirical Rule.

A **bell curve** is the shape of a normal distribution.

Guided Practice

1. A normally distributed data set has $\mu = 10$ and $\sigma = 2.5$, what is the probability of randomly selecting a value greater than 17.5 from the set?
2. A normally distributed data set has $\mu = .05$ and $\sigma = .01$, what is the probability of randomly choosing a value between .05 and .07 from the set?
3. A normally distributed data set has $\mu = 514$ and an unknown standard deviation, what is the probability that a randomly selected value will be less than 514?

Solutions:

1. If $\mu = 10$ and $\sigma = 2.5$, then $17.5 = \mu + 3\sigma$. Since we are looking for all data above that point, we need to subtract the probability that a value will occur *below* that value from 100%:

The probability that a value will be less than 10 is 50%, since 10 is the mean. There is another 34% between 10 and 12.5, another 13.5% between 12.5 and 15, and a final 2.35% between 15 and 17.5.

$100\% - 50\% - 34\% - 13.5\% - 2.35\% = 0.15\%$ **probability of a value greater than 17.5**

2. 0.05 is the mean, and 0.07 is 2 standard deviations above the mean, so the probability of a value in that range is $34\% + 13.5\% = 47.5\%$

3. 514 is the mean, so the probability of a value less than that is 50%.

Practice

For questions 1-15, assume all distributions to be normal or approximately normal, and calculate percentages using the 68 – 95 – 99.7rule.

1. Given mean 63 and standard deviation of 168, find the approximate percentage of the distribution that lies between -105 and 567.
2. Approximately what percent of a normal distribution is between 2 standard deviations and 3 standard deviations from the mean?
3. Given standard deviation of 74 and mean of 124, approximately what percentage of the values are greater than 198?
4. Given $\sigma = 39$ and $\mu = 101$, approximately what percentage of the values are less than 23?
5. Given mean 92 and standard deviation 189, find the approximate percentage of the distribution that lies between -286 and 470.
6. Approximately what percent of a normal distribution lies between $\mu + 1\sigma$ and $\mu + 2\sigma$?
7. Given standard deviation of 113 and mean 81, approximately what percentage of the values are less than -145?
8. Given mean 23 and standard deviation 157, find the approximate percentage of the distribution that lies between 23 and 337.
9. Given $\sigma = 3$ and $\mu = 84$, approximately what percentage of the values are greater than 90?
10. Approximately what percent of a normal distribution is between μ and $\mu + 1\sigma$?
11. Given mean 118 and standard deviation 145, find the approximate percentage of the distribution that lies between -27 and 118.
12. Given standard deviation of 81 and mean 67, approximately what percentage of values are greater than 310?
13. Approximately what percent of a normal distribution is less than 2 standard deviations from the mean?
14. Given $\mu + 1\sigma = 247$ and $\mu + 2\sigma = 428$, find the approximate percentage of the distribution that lies between 66 and 428.
15. Given $\mu - 1\sigma = -131$ and $\mu + 1\sigma = 233$, approximately what percentage of the values are greater than -495?

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.2.

9.3 Z-Scores

Objective

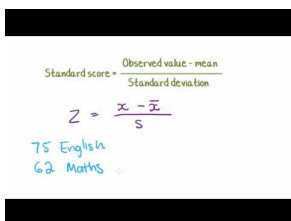
Here you will learn how z-scores can be used to evaluate how extreme a given value is in a particular set or population.

Concept

Using the Empirical Rule can give you a good idea of the probability of occurrence of a value that happens to be exactly one, two or three to either side of the mean, but how do you compare the probabilities of values that are in between standard deviations?

Watch This

The British video below is very clear and easy to follow. It is worth noting, particularly for U.S. students, that the instructor uses the notation \bar{x} rather than μ for mean, and pronounces z as “zed”.



MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/67459>

<http://youtu.be/2JjaWQZChqsvce> furthermaths - Maths Tutorial: Z scores

Guidance

Z-scores are related to the Empirical Rule from the standpoint of being a method of evaluating how extreme a particular value is in a given set. **You can think of a z-score as the number of standard deviations there are between a given value and the mean of the set.** While the Empirical Rule allows you to associate the first three standard deviations with the percentage of data that each SD includes, the z-score allows you to state (as accurately as you like), just how many SD's a given value is above or below the mean.

Conceptually, the z-score calculation is just what you might expect, given that you are calculating the number of SD's between a value and the mean. You calculate the z-score by first calculating the difference between your value and the mean, and then dividing that amount by the standard deviation of the set. The formula looks like this:

$$z\text{-score} = \frac{(\text{value} - \text{mean})}{\text{standard deviation}} = \frac{(x - \mu)}{\sigma}$$

In this lesson, we will practice calculating the z-score for various values. In the next lesson, we will learn how to associate the z-score of a value with the probability that the value will occur.

Example A

What is the z-score of a value of 27, given a set mean of 24, and a standard deviation of 2?

Solution:

To find the z-score we need to divide the difference between the value, 27, and the mean, 24, by the standard deviation of the set, 2.

$$z \text{ score} = \frac{27 - \mu}{\sigma}$$

$$\frac{27 - 24}{2}$$

$$\frac{3}{2}$$

$$z \text{ score of } 27 = +1.5$$

This indicates that 27 is 1.5 standard deviations above the mean.

Example B

What is the z-score of a value of 104.5, in a set with $\mu = 125$ and $\sigma = 6.2$?

Solution:

Find the difference between the given value and the mean, then divide it by the standard deviation.

$$z \text{ score} = \frac{104.5 - \mu}{\sigma}$$

$$\frac{104.5 - 125}{6.2}$$

$$\frac{-20.5}{6.2}$$

$$z \text{ score of } 104.5 = -3.306$$

Note that the z-score is negative, since the measured value, 104.5, is less than (below) the mean, 125.

Example C

Find the value represented by a z-score of 2.403, given $\mu = 63$ and $\sigma = 4.25$.

**Solution:**

This one requires that we solve for a missing value rather than for a missing z-score, so we just need to fill in our formula with what we know and solve for the missing value:

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

$$2.403 = \frac{x - 63}{4.25}$$

$$10.213 = x - 63$$

$$73.213 = x$$

73.213 has a z-score of 2.403

Concept Problem Revisited

Using the Empirical Rule can give you a good idea of the probability of occurrence of a value that happens to be right on one of the first three standard deviations to either side of the mean, but how do you compare the probabilities of values that are in between standard deviations?

The z-score of a value is the count of the number of standard deviations between the value and the mean of the set. You can find it by subtracting the value from the mean, and dividing the result by the standard deviation.

Vocabulary

The **z-score** of a value is the number of standard deviations between the value and the mean of the set.

Guided Practice



1. What is the z-score of the price of a pair of skis that cost \$247, if the mean ski price is \$279, with a standard deviation of \$16?
2. What is the z-score of a 5-scoop ice cream cone if the mean number of scoops is 3, with a standard deviation of 1 scoop?
3. What is the z-score of the weight of a cow that tips the scales at 825 lbs, if the mean weight for cows of her type is 1150 lbs, with a standard deviation of 77 lbs?
4. What is the z-score of a measured value of 0.0034, given $\mu = 0.0041$ and $\sigma = 0.0008$?

Solutions:

1. First find the difference between the measured value and the mean, then divide that difference by the standard deviation:

$$\frac{247 - 279}{16} = \frac{-32}{16} = -2$$

2. This one is easy: The difference between 5 scoops and 3 scoops is +2, and we divide that by the standard deviation of 1, so the **z-score is +2**.

3. First find the difference between the measured value and the mean, then divide that difference by the standard deviation:

$$\frac{825 \text{ lbs} - 1150 \text{ lbs}}{771 \text{ lbs}} = \frac{-325}{77} = -4.2207$$

4. First find the difference between the measured value and the mean, then divide that difference by the standard deviation:

$$\frac{0.0034 - 0.0041}{0.0008} = \frac{-0.0007}{0.0008} = -0.875$$

Practice

- Given a distribution with a mean of 70 and standard deviation of 62, find a value with a z-score of -1.82.
- What does a z-score of 3.4 mean?
- Given a distribution with a mean of 60 and standard deviation of 98, find the z-score of 120.76.
- Given a distribution with a mean of 60 and standard deviation of 21, find a value with a z-score of 2.19.
- Find the z-score of 187.37, given a distribution with a mean of 185 and standard deviation of 1.
- What does a z-score of -3.8 mean?
- Find the z-score of 125.18, given a distribution with a mean of 101 and standard deviation of 62.
- Given a distribution with a mean of 117 and standard deviation of 42, find a value with a z-score of -0.94.
- Given a distribution with a mean of 126 and standard deviation of 100, find a value with a z-score of -0.75.
- Find the z-score of 264.16, given $\mu = 188$ and $\sigma = 64$.
- Find a value with a z-score of -0.2, given $\mu = 145$ and $\sigma = 56$.
- Find the z-score of 89.79 given $\mu = 10$ and $\sigma = 79$.

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.3.

9.4 Z Scores II

Objective

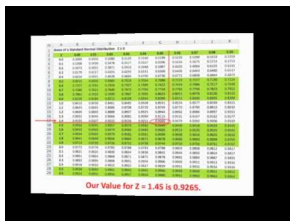
Here you will learn to evaluate z -scores as they relate to probability.

Concept

Knowing the z -score of a given value is great, but what can you do with it? How does a z -score relate to probability? What is the probability of occurrence of a z -score less than $+2.47$?

Watch This

The video below provides a demonstration of how to use a z -score probability reference table, as we do in this lesson. The table he uses in the video is slightly different, but the concept is the same.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.45	0.9285	0.9299	0.9312	0.9325	0.9338	0.9351	0.9364	0.9377	0.9389	0.9401

MEDIA

Click image to the left or use the URL below.

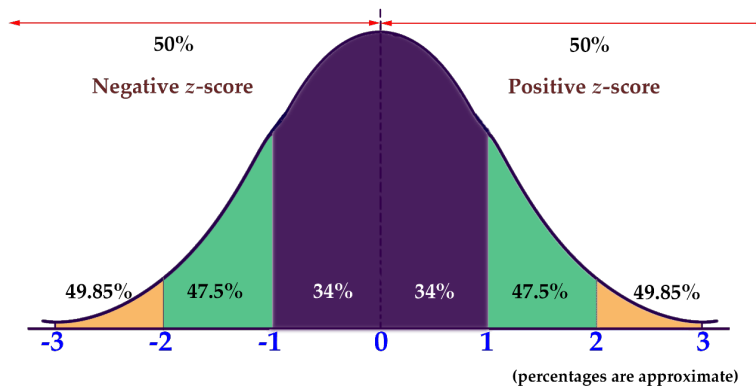
URL: <https://www.ck12.org/flx/render/embeddedobject/67464>

<http://youtu.be/rEmNUkKSpbU> TAMUC - Dr. Dawg

Guidance

Since z -scores are a measure of the number of SD's between a value and the mean, they can be used to calculate probability by comparing the location of the z -score to the area under a normal curve either to the left or right. The area can be calculated using calculus, but we will just use a table to look up the area.

I believe that the concept of comparing z -scores to probability is most easily understood with a graphic like the one we used in the lesson on the Empirical Rule, so I included one below. Be sure to review the Examples to see how the scores work.



Like the graphic we viewed in the Empirical Rule lesson, this one only provides probability percentages for integer values of z -scores (standard deviations). In order to find the values for z -scores that aren't integers, you can use a table like the one below. To find the value associated with a given z -score, you find the first decimal of your z -score on the left or right side and then the 2nd decimal of your z -score across the top or bottom of the table. Where they intersect you will find the decimal expression of the percentage of values that are less than your sample (see Ex. A).

TABLE 9.1:

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359	0.0
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.1
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.2
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517	0.3
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879	0.4
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224	0.5
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	0.6
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	0.7
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	0.8
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389	0.9
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	1.0
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883	1.1
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015	1.2
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	1.3
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1.4
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1.5
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1.6
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1.7
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1.8
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767	1.9
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	2.0
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857	2.1
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989	2.2
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	2.3
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	2.4
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	2.5
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964	2.6

TABLE 9.1: (continued)

2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974	2.7
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981	2.8
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	2.9
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999	3.0
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993	3.1
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995	3.2
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997	3.3
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	3.4
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	3.5
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.6
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.7
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.8
3.9	1	1	1	1	1	1	1	1	1	1	3.9
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z

Z-score tables like the one above describe the probability that a given value, or any value less than it, will occur in a given set. This particular table assumes you are looking to find the probability associated with a *positive* z-score. You may have additional work to do if the z-score is negative.

- To find the percentage of values **greater** than a **negative** Z score, just look up the matching positive Z score value.
- To find the percentage of values **less** than a **negative** z-score, subtract the chart value from 1.
- To find the percentage of values **greater** than a **positive** z-score, subtract the chart value from 1.

Example A

What is the probability that a value with a z-score less than 2.47 will occur in a normal distribution?

Solution:

Scroll up to the table above and find “2.4” on the left or right side. Now move across the table to “0.07” on the top or bottom, and record the value in the cell: **0.9932**. That tells us that **99.32% of values in the set are at or below a z-score of 2.47**.

Example B

What is the probability that a value with a z-score greater than 1.53 will occur in a normal distribution?

Solution:

Scroll up to the table of z-score probabilities again and find the intersection between 1.5 on the left or right and 3 on the top or bottom, record the value in the cell: **0.937**.

That decimal lets us know that 93.7% of values in the set are *below* the z-score of 1.53. To find the percentage that is *above* that value, we subtract 0.937 from 1.0 (or 93.7% from 100%), to get **0.063 or 6.3%**.

Example C

What is the probability of a random selection being less than 3.65, given a normal distribution with $\mu = 5$ and $\sigma = 2.2$?

Solution:

This question requires us to first find the z-score for the value 3.65, then calculate the percentage of values below that z-score from a reference.

1. Find the z-score for 3.65, using the z-score formula: $\frac{(x-\mu)}{\sigma}$

$$\frac{3.65 - 5}{2.2} = \frac{-1.35}{2.2} \approx -0.61$$

2. Now we can scroll up to our z -score reference above and find the intersection of 0.6 and 0.01, which should be **.7291**.

3. Since this is a negative z -score, and we want the percentage of values *below* it, we subtract that decimal from 1.0 (reference the three steps highlighted by bullet points below the chart if you didn't recall this), to get $1 - .7291 = .2709$

There is approximately a 27.09% probability that a value less than 3.65 would occur from a random selection of a normal distribution with mean 5 and standard deviation 2.2.

Concept Problem Revisited

Knowing the z -score of a given value is great, but what can you do with it? How does a z -score relate to probability? What is the probability of occurrence of a z -score less than 2.47?

A z -score lets you calculate the probability that a randomly selected value will be greater or less than a particular value in a set.

To find the probability of a z -score below +2.47, using a reference such as the table in the lesson above:

1. Find 2.4 on the left or right side
2. Move across to 0.07 on the top or bottom.
3. The cell you arrive at says: 0.9932, which means that apx 99.32% of the values in a normal distribution will occur below a z -score of 2.47.

Vocabulary

A **z -score table** associates the various common z -scores between 0 and 3.99 with the decimal probability of being less than or equal to that z -score.

Guided Practice

1. What is the probability of occurrence of a value with z -score greater than 1.24?
2. What is the probability of $z < -.23$?
3. What is $P(Z < 2.13)$?

Solutions:

1. Since this is a positive z -score, we can use the value for $z = 1.24$ directly from the table, and just express it as a percentage: **0.8925 or 89.25%**
2. This is a negative z -score, and we want the percentage of values *greater* than it, so we need to subtract the value for $z = +0.23$ from 1: $1 - 0.591 = .409$ **or 40.9%**
3. This is a positive z -score, and we need the percentage of values below it, so we can use the percentage associated with $z = +2.13$ directly from the table: **0.9834 or 98.34%**

Practice

Find the probabilities, use the table from the lesson above.

1. What is the probability of a z -score less than $+2.02$?
2. What is the probability of a z -score greater than $+2.02$?
3. What is the probability of a z -score less than -1.97 ?
4. What is the probability of a z -score greater than -1.97 ?
5. What is the probability of a z -score less than $+0.09$?
6. What is the probability of a z -score less than -0.02 ?
7. What is $P(Z < 1.71)$?
8. What is $P(Z > 2.22)$?
9. What is $P(Z < -1.19)$?
10. What is $P(Z > -2.71)$?
11. What is $P(Z < 3.71)$?
12. What is the probability of the random occurrence of a value greater than 56 from a normally distributed population with mean 62 and standard deviation 4.5?
13. What is the probability of a value of 329 or greater, assuming a normally distributed set with mean 290 and standard deviation 32?
14. What is the probability of getting a value below 1.2 from the random output of a normally distributed set with $\mu = 2.6$ and $\sigma = .9$?

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.4.

9.5 Z-scores III

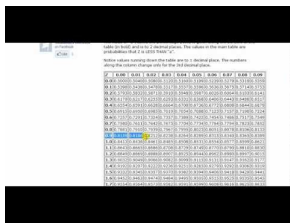
Objective

Here you will learn to calculate the probability of a z -score between two others.

Concept

Do z -score probabilities always need to be calculated as the chance of a value either above or below a given score? How would you calculate the probability of a z -score between -0.08 and $+1.92$?

Watch This



MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/67469>

<http://youtu.be/tnVr1Qbk4YU?t=9m23sAH> Metrics - Reading probabilities from the Z table

Guidance

To calculate the probability of getting a value with a z -score between two other z -scores, you can either use a reference table to look up the value for both scores and subtract them to find the difference, or you can use technology. In this lesson, which is an extension of Z-scores and Z-scores II, we will practice both methods.

Historically, it has been very common to use a z -score probability table like the one below to look up the probability associated with a given z -score:

TABLE 9.2:

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359	0.0
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.1
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.2
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517	0.3
0.4	.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879	0.4
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224	0.5
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	0.6
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	0.7
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	0.8
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389	0.9

TABLE 9.2: (continued)

1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	1.0
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883	1.1
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015	1.2
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	1.3
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1.4
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1.5
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1.6
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1.7
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1.8
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767	1.9
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	2.0
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857	2.1
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989	2.2
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	2.3
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	2.4
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	2.5
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964	2.6
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974	2.7
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981	2.8
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	2.9
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999	3.0
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993	3.1
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995	3.2
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997	3.3
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	3.4
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	3.5
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.6
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.7
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.8
3.9	1	1	1	1	1	1	1	1	1	1	3.9
Z	0.000	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z

Since the proliferation of the Internet, however, you can also use a free online calculator such as one of these three:

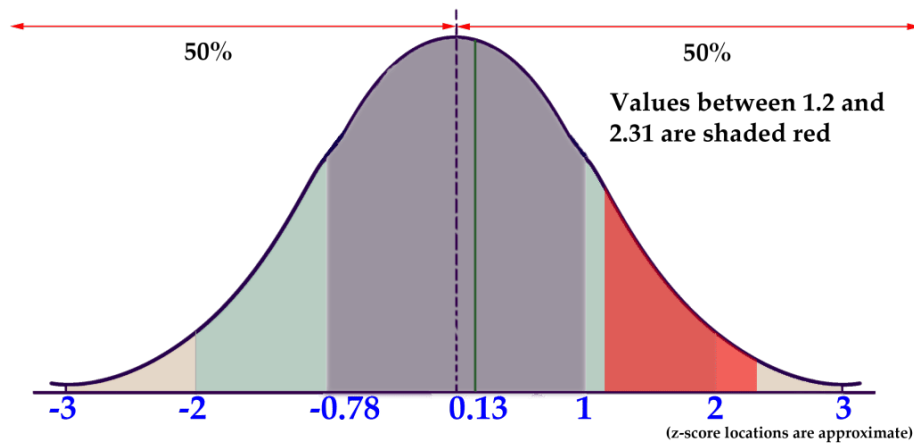
- <http://www.fourmilab.ch/rpkp/experiments/analysis/zCalc.html>
- <http://easycalculation.com/statistics/p-value-for-z-score.php>
- <http://www.mathportal.org/calculators/statistics-calculator/z-score-calculator.php>

Example A

What is the probability associated with a z -score between 1.2 and 2.31?

Solution:

To evaluate the probability of a value occurring within a given range, you need to find the probability of both the upper and lower values in the range, and subtract to find the difference.



- First find $z = 1.2$ on the z -score probability reference above: **.8849** Remember that value represents the percentage of values *below* 1.2.
- Next, find and record the value associated with $z = 2.31$: **.9896**
- Since approximately 88.49% of all values are below $z = 1.2$ and approximately 98.96% of all values are below $z = 2.31$, there are $98.96\% - 88.49\% = 10.47\%$ of values **between**.

Example B

What is the probability that a value with a z -score between -1.32 and $+1.49$ will occur in a normal distribution?

Solution:

Let's use the online calculator at mathportal.org for this one.

When you open the page, you should see a window like this:

MATHPORTAL.ORG

Z - Score Calculator

Enter cutoff points in order to find the area under normal curve. Important: The form will NOT let you enter wrong characters (like y, p, ;, ...)

Use the standard normal distribution to find one of the following probabilities:

$P(\text{ } < Z < \text{ })$

$P(\text{ } < Z)$

$P(\text{ } > Z)$

All you need to do is select the radio button to the left of the first type of probability, input “ -1.32 ” into the first box, and 1.49 into the second. When you click “Compute”, you should get the result

$$P(-1.32 < Z < 1.49) = 0.8385$$

Which tells us that there is approximately and 83.85% probability that a value with a z -score between 1.32 and 1.49 will occur in a normal distribution.

Notice that the calculator also details the steps involved with finding the answer:

1. Estimate the probability using a graph, so you have an idea of what your answer should be.
2. Find the probability of $z < 1.49$, using a reference. (0.9319)
3. Find the probability of $z < -1.32$, again, using a reference. (0.0934)
4. Subtract the values: $0.9319 - 0.0934 = 0.8385$ or 83.85%

Example C

What is the probability that a random selection will be between 8.45 and 10.25, if it is from a normal distribution with $\mu = 10$ and $\sigma = 2$?

Solution:

This question requires us to first find the z -scores for the value 8.45 and 10.25, then calculate the percentage of value between them by using values from a z -score reference and finding the difference.

1. Find the z -score for 8.45, using the z -score formula: $\frac{(x-\mu)}{\sigma}$

$$\frac{8.45 - 10}{2} = \frac{-1.55}{2} \approx -0.78$$

2. Find the z -score for 10.25 the same way:

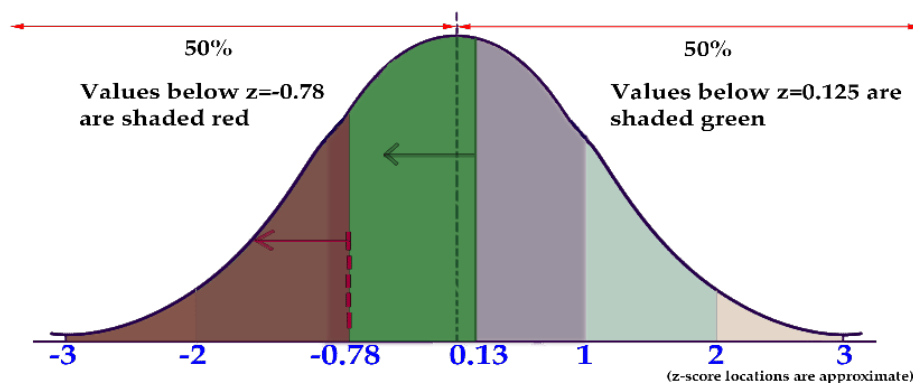
$$\frac{10.25 - 10}{2} = \frac{0.25}{2} \approx .13$$

3. Now find the percentages for each, using a reference (don't forget we want the probability of values *less* than our negative score and *less* than our positive score, so we can find the values between):

$$P(Z < -0.78) = .2177 \text{ or } 21.77\%$$

$$P(Z < .13) = .5517 \text{ or } 55.17\%$$

4. At this point, let's sketch the graph to get an idea what we are looking for:



Because the values we are interested in are between the two z -scores, we subtract the red area from the green area to get only the green "stripe" between -0.78 and 0.13.

5. Finally, subtract the values to find the difference:

$$.5517 - .2177 = .3340 \text{ or about } 33.4\%$$

There is approximately a 33.4% probability that a value between 8.45 and 10.25 would result from a random selection of a normal distribution with mean 10 and standard deviation 2.

Concept Problem Revisited

Do z-score probabilities always need to be calculated as the chance of a value either above or below a given score? How would you calculate the probability of a z-score between -0.08 and +1.92?

After this lesson, you should know without question that z-score probabilities do not need to assume only probabilities above or below a given value, the probability between values can also be calculated.

The probability of a z-score below -0.08 is 46.81%, and the probability of a z-score below 1.92 is 97.26%, so the probability between them is $97.26\% - 46.81\% = 50.45\%$.

Vocabulary

A **z-score** is a measure of how many standard deviations there are between a data value and the mean.

A **z-score probability table** is a table that associates z-scores to area under the normal curve. The table may be used to associate a Z-score with a percent probability.

Guided Practice

1. What is the probability of a z-score between -0.93 and 2.11?
2. What is $P(1.39 < Z < 2.03)$?
3. What is $P(-2.11 < Z < 2.11)$?

Solutions:

1. Using the z-score probability table above, we can see that the probability of a value below -0.93 is .1762, and the probability of a value below 2.11 is .9826. Therefore, the probability of a value between them is $.9826 - .1762 = .8064$ or 80.64%
2. Using the z-score probability table, we see that the probability of a value below $z = 1.39$ is .9177, and a value below $z = 2.03$ is .9788. That means that the probability of a value between them is $.9788 - .9177 = .0611$ or 6.11%
3. Using the online calculator at mathportal.org, we select the top calculation with the associated radio button to the left of it, enter “-2.11” in the first box, and “2.11” in the second box. Click “Compute” to get “.9652”, and convert to a percentage. The probability of a z-score between -2.11 and +2.11 is about **96.52%**.

Practice

Find the probabilities, use the table from the lesson or an online resource.

1. What is the probability of a z-score between +1.99 and +2.02?
2. What is the probability of a z-score between -1.99 and +2.02?

3. What is the probability of a z -score between -1.20 and -1.97 ?
4. What is the probability of a z -score between $+2.33$ and -0.97 ?
5. What is the probability of a z -score greater than $+0.09$?
6. What is the probability of a z -score greater than -0.02 ?
7. What is $P(1.42 < Z < 2.01)$?
8. What is $P(1.77 < Z < 2.22)$?
9. What is $P(-2.33 < Z < -1.19)$?
10. What is $P(-3.01 < Z < -0.71)$?
11. What is $P(2.66 < Z < 3.71)$?
12. What is the probability of the random occurrence of a value between 56 and 61 from a normally distributed population with mean 62 and standard deviation 4.5?
13. What is the probability of a value between 301 and 329, assuming a normally distributed set with mean 290 and standard deviation 32?
14. What is the probability of getting a value between 1.2 and 2.3 from the random output of a normally distributed set with $\mu = 2.6$ and $\sigma = .9$?

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.5.

9.6 The Mean of Means

Objective

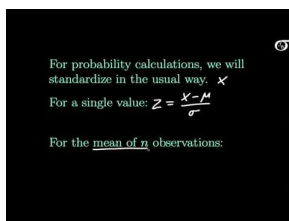
Here you will learn how to calculate the mean of means, which is the mean value of several sample means.

Concept

Suppose you have taken several samples of 10 units each from a population of 500 students, and calculated the mean of each sample. How might you use the data you now have to estimate a mean for the entire population?



Watch This



MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/67531>

<http://youtu.be/0zqNGDVNKgAJB> Statistics - The Sampling Distribution of the Sample Mean

Guidance

In statistics, you often need to take data from a small number of samples and use it to extrapolate an estimate of the parameters of the population the samples were pulled from. Since one of the more common parameters of interest is the mean, it is common to see a distribution of the means of a number of samples (I realize this may be confusing, “sample” here actually refers to the results of several individual samples) from the same population. This

distribution is called, appropriately, the “*sampling distribution of the sample mean*”. We will be investigating the sampling distribution of the sample mean in more detail in the next lesson “*The Central Limit Theorem*”, but in essence it is simply a representation of the spread of the means of several samples.

Here we will be focusing on a single value in that sampling distribution, the “*mean of means*”. The mean of means is simply the mean of all of the means of several samples. By calculating the mean of the sample means, you have a single value that can help summarize a lot of data.

The mean of means, notated here as $\mu_{\bar{x}}$, is actually a pretty straightforward calculation. Simply sum the means of all your samples and divide by the number of means.

As a formula, this looks like:

$$\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 \dots + \bar{x}_n}{n}$$

The second common parameter used to define sampling distribution of the sample means is the “*standard deviation of the distribution of the sample means*”. The only significant difference between the standard deviation of a population and the standard deviation of sample means is that you need to divide the population standard deviation by the square root of the sample size.

As a formula, this looks like:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

I recognize that the terminology in this lesson may be getting a bit scary, but the actual concept and the required calculations are actually not particularly difficult. **Work your way through the examples below**, and I think you will find that the hardest part of this lesson is getting past the wording!

Example A

Given the following sample means, what is the mean of means?

$$\bar{x}_1 = 4.35, \bar{x}_2 = 4.62, \bar{x}_3 = 4.29, \bar{x}_4 = 4.39, \bar{x}_5 = 4.55$$

Solution:

To calculate the mean of means, sum the sample means and divide by the number of samples:

$$\begin{aligned} \mu_{\bar{x}} &= \frac{4.35 + 4.62 + 4.29 + 4.39 + 4.55}{5} \\ &= \frac{22.20}{5} \\ \mu_{\bar{x}} &= 4.44 \end{aligned}$$

Example B

Brian works at a pizza restaurant, and has been carefully monitoring the weight of cheese he puts on each pizza for the past week. Each day, Brian tracks the weight of the cheese on each pizza he makes, and calculates the mean weight of cheese on each pizza for that day. If the weights below represent the mean weights for each day, what is the mean of means weight of cheese over the past week? If Brian makes 25 pizzas per day and knows the standard deviation of cheese weight per pizza is 0.5 oz, what is the standard deviation of the sample distribution of the sample means?



TABLE 9.3:

DAY	WEIGHT (OZ)
Monday	7.84
Tuesday	7.93
Wednesday	7.79
Thursday	8.03
Friday	8.14
Saturday	8.09
Sunday	7.88

Solution:

First calculate the mean of means by summing the mean from each day and dividing by the number of days:

$$\begin{aligned}\mu_{\bar{x}} &= \frac{7.84 + 7.93 + 7.79 + 8.03 + 8.14 + 8.09 + 7.88}{7} \\ &= \frac{55.7}{7} \\ \mu_{\bar{x}} &= 7.96\end{aligned}$$

Then use the formula to find the standard deviation of the sampling distribution of the sample means:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Where σ is the standard deviation of the population, and n is the number of data points in each sampling.

$$\begin{aligned}\sigma_{\bar{x}} &= \frac{.05 \text{ oz}}{\sqrt{25}} \\ &= \frac{.05}{5} \\ \sigma_{\bar{x}} &= .01\end{aligned}$$

Brian's research indicates that the cheese he uses per pizza has a mean weight of 7.96 oz, with a standard deviation of .01 oz.

Example C

Calculate $\mu_{\bar{x}}$, given the following:

$$\bar{x}_1 = 352.7$$

$$\bar{x}_2 = 351.9$$

$$\bar{x}_3 = 349.97$$

$$\bar{x}_4 = 352.33$$

$$\bar{x}_5 = 353.1$$

$$\bar{x}_6 = 349.63$$

Solution:

The $\mu_{\bar{x}}$ (mean of means) of the given data is:

$$\begin{aligned}\mu_{\bar{x}} &= \frac{352.7 + 351.9 + 349.97 + 352.33 + 353.1 + 349.63}{6} \\ &= \frac{2109.63}{6} \\ \mu_{\bar{x}} &= 351.61\end{aligned}$$

Concept Problem Revisited

Suppose you have taken several samples of 10 units each from a population of 500 students, and calculated the mean of each sample. How might you use the data you now have to estimate a mean for the entire population?

You could sum the means of each 10-unit sampling, and divide by the number of samples to get the mean of the means. You could further divide the standard deviation of the entire 500 students (if known) by $\sqrt{10}$ (since each sampling contained 10 data points), to find the standard deviation of the distribution of the sample mean.

Vocabulary

The **sampling distribution of the sample mean** is the distribution that describes the spread of the means of multiple samples from the same population.

The **mean of means** is the overall mean value of the means of several samples from the same population.

The **standard deviation of the distribution of the sample means** is a measure of the spread of the random variable \bar{x} .

Guided Practice

1. Calculate $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$, given the following data:

$$\bar{x}_1 = 251.6$$

$$\bar{x}_2 = 242.8$$

$$\bar{x}_3 = 248.79$$

$$\bar{x}_4 = 245.33$$

$$\bar{x}_5 = 253.21$$

$$\bar{x}_6 = 256.31$$

Sample size = 9, $\sigma = 5.8$

2. If the σ of a population is 2.94, and 25 samples of 12 samples each are taken, what is $\sigma_{\bar{x}}$?

3. Given the population: {1, 2, 3, 4, 5}, create a sampling distribution by finding the mean of all possible samples that include four units. How does $\mu_{\bar{x}}$ compare to μ ?

Solutions:

1. First calculate $\mu_{\bar{x}}$, using $\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 \dots + \bar{x}_n}{n}$.

$$\begin{aligned} & \frac{251.6 + 242.8 + 248.79 + 245.33 + 253.21 + 256.31}{6} \\ & \frac{1498.04}{6} \\ & \mu_{\bar{x}} = 249.67 \end{aligned}$$

Next calculate $\sigma_{\bar{x}}$, using $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$:

$$\begin{aligned} \sigma_{\bar{x}} &= \frac{5.8}{\sqrt{9}} \\ &= \frac{5.8}{3} \\ \sigma_{\bar{x}} &= 1.93\bar{3} \end{aligned}$$

2. To calculate $\sigma_{\bar{x}}$, use $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$:

$$\begin{aligned} \sigma_{\bar{x}} &= \frac{2.94}{\sqrt{12}} \\ &= \frac{2.94}{4} \\ \sigma_{\bar{x}} &= .735 \end{aligned}$$

3. This one requires a few steps, first we need to find the mean of each possible sample of four units:

\bar{x}_1 (1234) has a mean of 2.5

\bar{x}_2 (1235) has a mean of 2.75

\bar{x}_3 (1245) has a mean of 3

\bar{x}_4 (1345) has a mean of 3.25

\bar{x}_5 (2345) has a mean of 3.5

Next we calculate the mean of means, $\mu_{\bar{x}}$, using $\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 \dots + \bar{x}_n}{n}$:

$$\begin{aligned} \mu_{\bar{x}} &= \frac{2.5 + 2.75 + 3 + 3.25 + 3.5}{5} \\ &= \frac{15}{5} \\ \mu_{\bar{x}} &= 3 \end{aligned}$$

Now we need to calculate μ , using all the population data:

$$\begin{aligned}\mu &= \frac{1+2+3+4+5}{5} \\ &= \frac{15}{5} \\ \mu &= 3\end{aligned}$$

With the given data, $\mu_{\bar{x}} = \mu$

Practice

- Find $\mu_{\bar{x}}$, given $\bar{x}_1 = 21.0$, $\bar{x}_2 = 24.3$, $\bar{x}_3 = 25.0$, $\bar{x}_4 = 20.6$, $\bar{x}_5 = 22.3$, and $\bar{x}_6 = 22.3$
- Find $\mu_{\bar{x}}$, given $\bar{x}_1 = 15.1$, $\bar{x}_2 = 15.77$, $\bar{x}_3 = 15.55$, $\bar{x}_4 = 15.99$, $\bar{x}_5 = 15.42$, and $\bar{x}_6 = 15.37$
- Find $\mu_{\bar{x}}$, given $\bar{x}_1 = 341.52$, $\bar{x}_2 = 345.16$, $\bar{x}_3 = 343.66$, $\bar{x}_4 = 345.86$, and $\bar{x}_5 = 336.10$
- Find $\mu_{\bar{x}}$, given $\bar{x}_1 = 1.41$, $\bar{x}_2 = 0.59$, $\bar{x}_3 = 1.44$, $\bar{x}_4 = 0.93$, $\bar{x}_5 = 1.44$, $\bar{x}_6 = 1.01$, and $\bar{x}_7 = 0.74$
- Find $\mu_{\bar{x}}$, given $\bar{x}_1 = 218.19$, $\bar{x}_2 = 279.70$, $\bar{x}_3 = 262.86$, and $\bar{x}_4 = 243.88$
- If the σ of a population is 292.66, and samples of 17 units each are taken, what is $\sigma_{\bar{x}}$?
- If the σ of a population is 41.39, and 23 samples of 30 samples each are taken, what is $\sigma_{\bar{x}}$?
- If the σ of a population is 193.61, and samples of 19 units each are taken, what is $\sigma_{\bar{x}}$?
- If the σ of a population is 91.85, and 129 samples of 11 samples each are taken, what is $\sigma_{\bar{x}}$?
- If the σ of a population is 255.19, and 43 samples of 31 samples each are taken, what is $\sigma_{\bar{x}}$?
- Given the population: {1, 2, 3, 4, 5, 6}, create a sampling distribution by finding the mean of all possible samples that include two units. How does $\mu_{\bar{x}}$ compare to μ ?
- Given the population: {1, 2, 3, 4}, create a sampling distribution by finding the mean of all possible samples that include two units. What is $\mu_{\bar{x}}$?
- Given the population: {1, 2, 3, 4, 5}, create a sampling distribution by finding the mean of all possible samples that include two units. How does $\mu_{\bar{x}}$ compare to μ ?
- Given the population: {1, 2, 3, 4, 5}, create a sampling distribution by finding the mean of all possible samples that include three units. What is $\mu_{\bar{x}}$?
- Given the population: {1, 2, 3, 4}, create a sampling distribution by finding the mean of all possible samples that include three units. What is $\mu_{\bar{x}}$?

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.6.

9.7 Central Limit Theorem

Objective

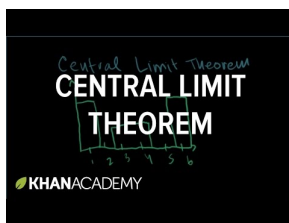
Here you will learn about one of the more remarkable theorems in all of mathematics, the Central Limit Theorem.

Concept

What is the Central Limit Theorem? How does the Central Limit Theorem relate other distributions to the normal distribution?

This lesson describes the relationship between the normal distribution and the Central Limit Theorem.

Watch This



MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/67467>

<http://youtu.be/JNm3M9cqWyc> Khan Academy - Central Limit Theorem

Guidance

The **Central Limit Theorem** is a very powerful statement in statistics, saying that as you take more and more samples from a random variable, the distribution of the means of the samples (If you completed the lesson titled “The Mean of Means”, you will recognize this as “the sampling distribution of the sample means”) will approximate a normal distribution. This is true *regardless of the original distribution of the random variable* (if the number of data points in each sample is 30 or more)! In fact, as demonstrated in the video above, even a discrete random variable with a pretty odd distribution will output an approximately normal distribution from the means of enough samples.

Formally, the CLT says:

If samples of size n are drawn at random from any population with a finite mean and standard deviation, then the sampling distribution of the sample means, \bar{x} , approximates a normal distribution as n increases.

In “normal English”:

If you collect many samples from an ordinary random variable, and calculate the mean of each sample, then the means will be distributed in an approximate bell-curve, and the “mean of means” will be the same as the mean of the population. The larger the size of the samples you collect, the more closely the distribution of their means will approximate a normal distribution.

Notes to remember:

- As long as your sample size is **30 or greater**, you may assume the distribution of the sample means to be approximately normal, meaning that you can calculate the probability that the mean of a single sample of size 30 or greater will occur by using the z -score of the mean.
- The mean of the distribution created from many sample means approaches the mean of the population. Formally: $\mu_{\bar{x}} = \mu$
- The standard deviation of the distribution of the means is estimated by dividing the standard deviation of the population by the square root of the sample size. Formally: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- Use the notation \bar{x} (x -bar) rather than the random variable x to indicate that the random variable you are describing is a sample mean.

You may use the z -score percentage reference table below as needed:

TABLE 9.4:

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359	0.0
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.1
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.2
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517	0.3
0.4	.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879	0.4
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224	0.5
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	0.6
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	0.7
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	0.8
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389	0.9
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	1.0
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883	1.1
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015	1.2
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	1.3
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1.4
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1.5
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1.6
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1.7
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1.8
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767	1.9
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	2.0
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857	2.1
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989	2.2
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	2.3
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	2.4
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	2.5
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964	2.6
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974	2.7
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981	2.8
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	2.9
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999	3.0
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993	3.1
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995	3.2
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997	3.3
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	3.4

TABLE 9.4: (continued)

3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	3.5
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.6
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.7
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	3.8
3.9	1	1	1	1	1	1	1	1	1	1	3.9
Z	0.000	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z

Example A

Mack asked 42 fellow high-school students how much they spent for lunch, on average. According to his research online, the amount spent for lunch by high school students nation wide has $\mu = \$15$, with $\sigma = \$9$. What is the probability that Mack's random sample will result within \$0.01 of the national average?

**Solution:**

There are a few important facts to note here:

- Mack's sample is 42 students, since $42 \geq 30$, he can safely assume that the distribution of his *sample* is approximately normal, according to the Central Limit Theorem.
- The range we are considering is \$14.99 to \$15.01, since that represents \$0.01 above and below the mean.
- The mean of the sample should approximate the mean of the population, in other words $\mu_{\bar{x}} = \mu$
- The standard deviation of Mack's sample, $\sigma_{\bar{x}}$, can be calculated as $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, where $n = 42$

Let's start by finding the standard deviation of the sample, $\sigma_{\bar{x}}$:

$$\begin{aligned}\sigma_{\bar{x}} &= \frac{9}{\sqrt{42}} \\ &= \frac{9}{6.48} \\ \sigma_{\bar{x}} &= 1.389\end{aligned}$$

Since Mack's sample of 42 samples can be assumed to be normally distributed, and since we now know the standard deviation of the sample, 1.39, we can calculate the z-scores of the range using $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$:

$$Z_1 = \frac{15.01 - 15.00}{1.389} = +0.01$$

$$Z_2 = \frac{14.99 - 15.00}{1.389} = -0.01$$

Finally, we look up Z_1 and Z_2 on the Z-score probability table to get a range of 50.4% to 49.6% = **0.80%**

The probability that Mack's sample will have a mean within \$0.01 of the population mean of \$15.00 is a little less than 1%.

Example B

The time it takes a student to complete the mid-term for Algebra II is a bi-modal distribution with $\mu = 1$ hr and $\sigma = 1$ hr. During the month of June, Professor Spence administers the test 64 times. What is the probability that the average mid-term completion time for students during the month of June exceeds 48 minutes?

Solution:

Important facts:

- There are more than 30 samples, so the Central Limit Theorem applies
- The mean of the sample should approximate the mean of the population, in other words $\mu_{\bar{x}} = \mu$
- The standard deviation of Professor Spence's sample, $\sigma_{\bar{x}}$, can be calculated as $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, where $n = 64$ (the number of tests/samples)
- 48 minutes is the same as $\frac{48}{60} = 0.8$ hrs, so the range we are interested in is $x > 0.8$ hrs

First calculate the standard deviation of the sample, using $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{64}}$$

$$\sigma_{\bar{x}} = 0.125$$

Since the sample is normally distributed, according to the CLT, we can use the standard deviation of the sample to calculate the z-score of the minimum value in the relevant range, 0.80 hrs:

$$Z = \frac{0.80 - 1}{0.125} = -1.60$$

Finally, we use the z-score probability reference above to correlate the z-score of -1.60 to the probability of a value greater than that

$$P(Z \geq -1.6) = .9452 \text{ or } 94.52\%$$

Example C

Evan price-checked 123 online auction sellers to record their average asking price for his favorite game. According to a major nation price-checking site, the national average online auction cost for the game is \$35.00 with a standard deviation of \$3.00. Evan found the prices less than \$34.86 on average. How likely is this result?

Solution:

Since there are more than 30 samples ($123 > 30$), we can apply the CLT theorem and treat the sample as a normal distribution.

The standard deviation of the sample is: $\sigma_{\bar{x}} = \frac{3}{\sqrt{123}} = \frac{3}{11.09} = .27$

The z -score for Evan's price point of \$34.86 is:

$$Z = \frac{34.86 - 35}{.27} = \frac{-.14}{.27} = -0.518$$

Consulting the z -score probability table, we learn that the area under the normal curve less than 0.52 is .3015 or 30.15%

The likelihood of 123 samples having a mean of \$34.86 is approximately 30.15%

Concept Problem Revisited

What is the Central Limit Theorem? How does the Central Limit Theorem relate other distributions to the normal distribution?

The Central Limit Theorem says that the larger the sample size, the more the mean of multiple samples will represent a normal distribution. Since that is true regardless of the original distribution, the CLT can be used to effect a bridge between other types of distributions and a normal distribution.

Vocabulary

The **Central Limit Theorem** states that if samples are drawn at random from any population with a finite mean and standard deviation, then the sampling distribution of the sample means approximates a normal distribution as the sample size increases beyond 30.

The **sampling distribution of the sample means** is a distribution of the means of multiple samples. It is commonly assumed to be a normal distribution, though technically it is normal only if the sample size is greater than 30.

Guided Practice



1. The time it takes to drive from Cheyenne WY to Denver CO has a μ of 1 hr and σ of 15 mins. Over the course of a month, a highway patrolman makes the trip 55 times. What is the probability that his average travel time exceeds 60 minutes?
2. Abbi polls 95 high school students for their G.P.A.. According to the school, the average G.P.A. of high school students has a mean of 3.0, and standard deviation of .5. What is the probability that Abbi's random sample will have a mean within 0.01 of the population?

3. A recipe website has calculated that the time it takes to cook Sunday dinner has a μ of 1 hr with σ of 25 mins. Over the course of a month, 172 users report their time spent cooking Saturday dinner, what is the probability that the average user reports spending less than 45 mins cooking dinner?

Solutions:

1. The sample mean, $\mu_{\bar{x}}$ is the same as the population mean: 1 hr = 60 mins.

The sample standard deviation is $\frac{15 \text{ mins}}{\sqrt{55}} = \frac{15}{7.42} = 2.02 \text{ min}$

The 55 trips made by the patrolman exceed the minimum sample size of 30 required to apply the CLT, so we may assume the sample means to be normally distributed.

The z-score of the patrolman's average time is: $\frac{60-60}{2.02} = \frac{0}{2.02} = 0$

According to the z-score percentage reference, a z-score of 0 corresponds to .50 or 50%

There is a 50% probability that the patrolman's mean travel time is greater than 60 mins.

2. The sample mean of the 95 polled G.P.A. scores is the same as the population mean: **3.0**

The sample standard deviation is $\frac{.5}{\sqrt{95}} = \frac{.5}{9.75} = .05$

The 95 sampled G.P.A.'s exceed the minimum sample size of 30, so we may apply the CLT.

The z-scores of the minimum and maximum values in the range of interest, 2.99 to 3.01 is:

$$Z_1 = \frac{2.99 - 3.00}{.05} = \frac{-.01}{.05} = -0.2$$

$$Z_2 = \frac{3.01 - 3.00}{.05} = \frac{.01}{.05} = +0.2$$

Referring to the z-score reference table, **the z-scores -0.2 and 0.2 cover a range of apx. 15.86%**

3. The sample mean, $\mu_{\bar{x}}$ is the same as the population mean: 1 hr = 60 mins.

The sample standard deviation is $\frac{25 \text{ mins}}{\sqrt{172}} = \frac{25}{13.11} = 1.91 \text{ min}$

The 172 users reporting cooking times exceed the minimum sample size of 30 required to apply the CLT, so we may assume the sample means to be normally distributed.

The z-score of the average reported cooking time is: $\frac{45-60}{1.91} = \frac{-15}{1.91} = -7.85$

According to the z-score percentage reference, a z-score of -7.85 corresponds to 0%.

There is essentially zero probability that 172 users would average only 45 mins.

Practice

- 128 randomly-sampled students reported how much they spent on a movie at the theater. If the national average amount spent at the movies has a mean of \$15 and standard distribution of \$8, what is the probability that the random sample will give a result within \$0.01 of the true value?
- The time an American family spends doing dishes in the evening has $\mu = 60 \text{ mins}$ and $\sigma = 60 \text{ mins}$. 58 Americans were polled to find the time they spend doing dishes. What is the probability that their average time exceeds 60 minutes?
- Rachel asked 65 second year college students how many credits they have taken. According to the colleges, the average number of credits taken by 2nd year students is 15, with a standard deviation of 7. How likely is it that Rachel got less than 17.17 on average?
- What do you need in order to apply the Central Limit Theorem to sample means?

5. 117 business women were asked how much they spend for lunch, on average. If the national average has a mean of \$30, and standard distribution of \$9, what is the probability that the random sample will return a result within \$0.01 of the true value?
6. According to the phone company, the daily average number of calls made by Americans is 30, with a standard deviation of 10. What is the probability that 117 Americans reported less than 30.92 calls per day, on average?
7. The time spent by the average technician repairing a laptop is governed by an exponential distribution where μ and σ are each 60 minutes. In the month of June, a technician repairs 76 laptops. How likely is it that the average repair time is greater than 77 minutes?
8. 46 teenagers were asked how many .mp3's they purchase each month. According .mp3 sales data, the average has a mean of 15, with a standard distribution of 2. How likely is it that the 46 polled teens averaged within 0.02 of the national average?
9. 44 classrooms were investigated to see how many students they contained. According to school data, the average number of students per classroom is 35, with a standard deviation of 10. How likely is it that the 44 classrooms averaged fewer than 33.49 students?
10. 100 bags of candy were counted to see how many pieces they contained. According to the company that fills the bags, the average number of candies per bag has a mean of 50, and standard distribution of 10. What is the probability that the 100 bags will have an average number within 0.02 of the production average?

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.7.

9.8 Approximating the Binomial Distribution

Objective

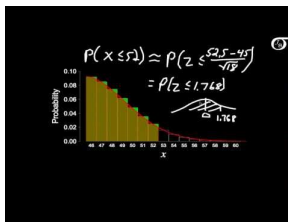
Here you will learn when it is reasonable to approximate a *binomial distribution* with a *normal distribution*, making quick work of probability calculations.

Concept

Suppose you were completing a multiple-choice test, and you are worried that you don't know the information well enough. If there are 75 questions, each with 4 answers, what is the probability that you would get at least 60 correct just by guessing randomly?

You could probably answer this question if you have completed prior lessons on binomial probability, but it would be quite a calculation, requiring you to individually calculate the probability of getting 60 correct, adding it to the probability of getting 61 correct, and so on, all the way up to 75! At the end of the lesson, we will review this question in light of the normal distribution, and see how much more efficient it can be.

Watch This



MEDIA

Click image to the left or use the URL below.

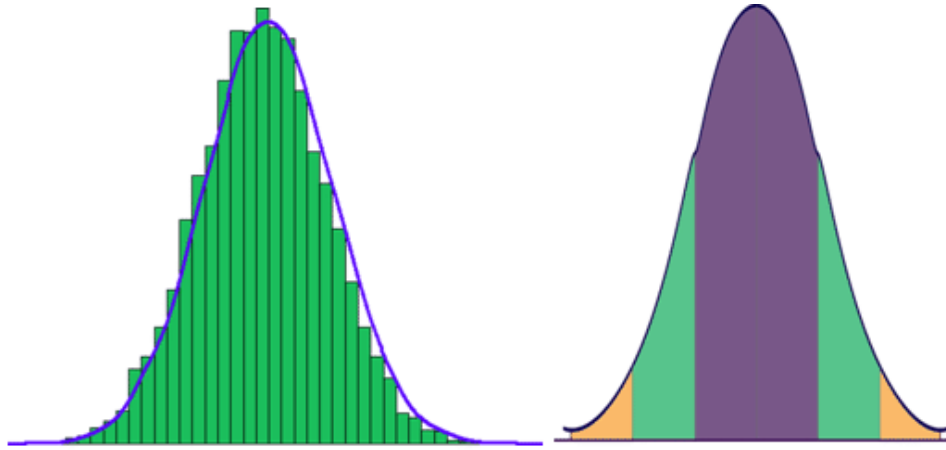
URL: <https://www.ck12.org/flx/render/embeddedobject/67471>

http://youtu.be/CCqWkJ_pqNUjb statistics - The Normal Approximation to the Binomial Distribution

Guidance

Many real life situations involve binomial probabilities, as we saw in prior lessons on binomial experiments. In fact, even many questions that don't appear binomial at first can be formatted so that they are, allowing the probability of success or failure of a given study to be calculated as a binomial probability. Unfortunately, if the probability of success spans a wide range of possible values, the calculation can become very burdensome.

The good news is that there is another way to approximate the probability of success, and you can see what it is by comparing the following graphs. The first graph displays the probability of getting various numbers of heads over 100 flips of a fair coin, in other words, the distribution of a binomial random variable with $P(\text{success}) = .50$. The second graph is a normal distribution. Notice any similarities?



They are extremely similar in shape, in fact, if you follow a “rule of thumb”, you can use a normal distribution to estimate the results of a binomial distribution with quite acceptable accuracy. The rule of thumb for knowing when the normal distribution will provide a good approximation of a binomial distribution with the same mean and standard deviation is:

$$n \times P > 10 \text{ and } n(1 - p) > 10$$

Where n is the number of trials, and p is the probability of success.

If you have determined that a given binomial distribution is a candidate for approximation using a normal distribution, you can calculate the μ and σ of the normal distribution using:

$$\begin{aligned}\mu &= np \\ \sigma &= \sqrt{np(1 - p)}\end{aligned}$$

If you are interested in the comparison between the binomial probability and normal approximation for a particular n or p value, there is an excellent Java applet at http://onlinestatbook.com/stat_sim/normal_approx/index.html that will show the actual values and a graph for any combination of n and p values.

Example A

Can the results of a binomial experiment consisting of 40 trials with a 72% probability of success of be acceptably approximated by a normal distribution?

Solution:

Here, $n = 40$, and $p = 0.72$

- First, is $n \times p > 10$?

$$\begin{aligned}n \times p &= 40 \times .72 = 28.8 \\ 21.6 &> 10 \text{ YES}\end{aligned}$$

- Second, is $n(1 - p) > 10$?

$$n(1 - p) = 40(1 - 0.72) = 40(.28) = 11.2$$

$$11.2 > 10 \text{ YES}$$

Yes, based on our rule of thumb, you could use a normal distribution to approximate the results of this binomial experiment.

Example B

If Kaile wants to estimate the probability of correctly guessing at least 9 answers out of 50 on a true/false exam, can she estimate using a normal distribution?

Solution:

Here, $n = 50$ and $p = 0.50$ (true/false):

$$n \times p = 50 \times .50 = 25$$

$$25 > 10 \text{ Yes}$$

$$n(1 - p) = 50(1 - 0.50) = 50 \times .50 = 25$$

$$25 > 10 \text{ Yes}$$

Example C

Ciere works in a production plant. Due to the balance of speed and accuracy in production, each part off the line has a 98.8% probability of defect free production.



- Can a binomial experiment based on 98.8% probability be approximated if Ciere produces 1000 parts?
- What would be the mean and standard deviation of the appropriate normal distribution?
- What is the probability that Ciere will produce at least 990 parts without a defect in a 1000 part run?

Solution:

a. Here, $n = 1000$ and $p = 0.988$, does this satisfy our 'rule of thumb'?

$$n \times p = 1000 \times 0.988 = 988 \text{ Yes}$$

$$n(1 - p) = 1000(1 - 0.988) = 1000 \times 0.012 = 12 \text{ Yes}$$

If Ciere produces 1000+ parts, this experiment may be approximated using a normal distribution.

b. The mean, μ , and standard deviation, σ , can be evaluated as follows:

$$\mu = np = 1000 \times .988 = 988$$

$$\sigma = \sqrt{np(1 - p)} = \sqrt{988(0.012)} = 3.44$$

c. To calculate this, we need the z -score of 990, which we can calculate using: $Z = \frac{(x - \mu)}{\sigma}$. Once we have the z -score, we can reference a z -score probability table to find the probability of a value above it:

$$z = \frac{(x - \mu)}{\sigma} = \frac{990 - 988}{3.44} = \frac{2}{3.44} = .52$$

Using a reference table (which you can find online or in the previous lesson), we see that the probability of a z -score greater than .52 is 34.60%

The probability of Ciere producing at least 990 parts in a row without a defect is about 34.60%.

Concept Problem Revisited

Suppose you were completing a multiple-choice test, and you are worried that you don't know the information well enough. If there are 75 questions, each with 4 answers, what is the probability that you would get at least 60 correct just by guessing randomly?

Here, since there are 75 questions, $n = 75$, and since each has 4 possible answers, $p = .25$. First we check to see if we can use the normal approximation:

$$n \times p = 75 \times .25 = 18.75$$

$$18.75 > 10 \text{ Yes}$$

Since we can use the normal distribution, we need to calculate the mean and SD of the distribution:

$$\mu = np = 75 \times .25 = 18.75$$

$$\sigma = \sqrt{np(1 - p)} = \sqrt{18.75(.75)} = 3.75$$

Now we need the z -score of our minimum number of correct guesses, 60:

$$z = \frac{(x - \mu)}{\sigma} = \frac{60 - 18.75}{3.75} = \frac{41.25}{3.75} = 11$$

Ha! We don't even need a z -score reference for this one, your chances of randomly guessing 60 or more correctly are virtually nil, since most tables only go up to $z = 3$ or $z = 3.99$. Better study more next time!

Vocabulary

A **binomial distribution** is a frequency distribution of the possible number of successful outcomes in a given number of success/failure trials, each of which has the same probability of success.

A **normal distribution** is a continuous bell-shaped distribution that is balanced about the mean, median, and mode. Sixty-eight percent of the observations fall within plus or minus one standard deviation of the mean, approximately 95% fall within plus or minus two standard deviations, and approximately 99.5% fall within plus or minus three standard deviations.

THE CONTINUITY CORRECTION FACTOR

WE WILL NOT BE DEALING WITH THIS IN OUR EXAMPLES, BUT YOU MAY FIND IN THE FUTURE THAT SINCE A BINOMIAL DISTRIBUTION IS DISCRETE, AND THE NORMAL DISTRIBUTION IS CONTINUOUS, YOU MAY NEED TO TREAT AN INTEGER VALUE OF THE BINOMIAL DISTRIBUTION AS THE INTERVAL BETWEEN .5 BELOW AND .5 ABOVE THE INTEGER, PARTICULARLY IF THE VALUE OF n IS RATHER SMALL.

Guided Practice

1. Is a binomial experiment consisting of 35 trials, each with a 55% probability of success, a good candidate for normal curve approximation?
2. A binomial random variable has a 0.821 probability of success. If data is collected from 48 trials, can the results be viably approximated with a normal distribution?
3. What is the approximate probability of correctly guessing at least 20 questions out of 50, on a true/false exam?

Solution:

1. $n = 35$ and $p = 0.55$, use the rule of thumb from the lesson to evaluate:

$$n \times p = 35 \times .55 = 19.25$$

$$19.25 > 10 \text{ Yes}$$

$$n \times (1 - p) = 35 \times .45 = 15.75$$

$$15.75 > 10 \text{ Yes}$$

Since the experiment meets both qualifications, a normal approximation should be variable.

2. $n = 48$ and $p = 0.721$, check with our rule of thumb:

$$n \times p = 48 \times 0.721 = 39.41$$

$$39.41 > 10 \text{ Yes}$$

$$n \times (1 - p) = 48 \times 0.179 = 7.05$$

$$7.05 < 10 \text{ No!}$$

3. Verify that we can use a normal approximation, using $n = 50$ and $p = 0.50$:

$$n \times p = 50 \times .5 = 25$$

$$25 > 10 \text{ Yes}$$

$$n \times (1 - p) = 50 \times .5 = 25$$

$$25 > 10 \text{ Yes}$$

Since we can use the normal approximation, we need to calculate the mean and SD, so we can get the z -score for 20 (the minimum number of correct answers we want to get).

$$\begin{aligned}\mu &= np = 50 \times .5 = 25 \\ \sigma &= \sqrt{np(1-p)} = \sqrt{25(.50)} = 3.54 \\ z &= \frac{(x-\mu)}{\sigma} = \frac{20-25}{3.54} = 1.41\end{aligned}$$

Consulting our reference, we learn that the probability that a z -score greater than 1.41 will occur is approximately 94%.

You have approximately a 94% probability of correctly guessing at least 20 questions correctly on a 50 question exam.

Practice

- Karen is playing a game of chance with a probability of success of 33%. If she plays the game 43 times, what is the probability that she will win more than 19 times?
- When approximating a binomial distribution, how do you calculate the standard deviation?
- Gregory has created a card game where you either draw a black card or a red card. If you draw a red card, you get a dollar. But if you draw a black card you owe him a dollar. The chance of drawing a red card is 61%. You decide to play against Gregory 26 times. Can you approximate this situation with a normal curve? Why or why not?
- Sue has organized her closet into summer clothing and winter clothing. She closes her eyes and reaches in her closet to pick an outfit. If she selects summer clothing she will be cold (it is winter). The chance of Sue selecting summer clothing is 41%. She decides to select 27 outfits this way. If you were to approximate this with a normal curve, what would the standard deviation be?
- Sharon can't decide between two guys that she likes. She picks a daisy from the garden and decides to play "I like Greg more, I like Stan more" with the petals. The chance of the last petal being "I like Greg more" is 67%. She decides to go through this process with 48 daisies. What is the probability that she will select Greg more than 36 times?
- Vern has to choose between two summer jobs. He painted a wheel red and blue. If the spinner lands in the red area, he works for a landscaping company. If it lands in the blue, he works for a fast food restaurant. The chance of the spinner landing in the red area is 52%. He decides to spin 33 times. What standard deviation would you use to approximate this situation with a normal curve?
- When approximating a binomial distribution, what is the mean?
- George has devised a scheme where he flips a coin to earn money. If it lands on heads you get a quarter. If it lands on tails you give him a quarter. The chance of the coin landing on heads, is 68%. You play against George 22 times. Can you approximate this situation with a normal curve? Why or why not?
- Jade has been practicing shooting a bow and arrow. Based on her target practice she has a 30% chance of hitting the bull's-eye with the bow and arrow. She shoots the bow and arrow 27 times. Can you approximate this situation with a normal curve? Why or why not?
- Steve has created a "grab the marble" game. If you grab a green marble you get a dollar, if you grab a yellow marble you get nothing. There are 31 green marbles, and 69 yellow ones. You decide to reach into the bag and grab a marble 39 times, replacing the marble you grab each time. What is the probability that you will win more than 9 dollars?
- Steve asks another friend to play "grab the marble". If you grab a green marble you get a dollar, if you grab a yellow marble you get nothing. Now, however, there are 49 green and 51 yellow marbles. You decide to reach into the bag and grab a marble 32 times, replacing the marble you grab each time. By approximating this situation with a normal curve, try to predict the expected outcome.

12. Kyle is reading a “Choose Your Own Adventure Book”. He has decided to leave his decisions to chance, so before making a choice between decision 'a' and decision 'b', he flips a coin. If it lands on heads, he selects choice 'a', if it lands on tails, he selects choice 'b'. The trick is that he uses an unfair coin with a probability of heads of 74%. He has to flip the coin 42 times to get to the end of the story. Can you approximate this situation with a normal curve? Why or why not?
13. 19 Students at a local high school are all applying to two different colleges. The chance of each of them getting into their first college of choice is 50%. Can you approximate this situation with a normal curve? Why or why not?
14. Tammy is at the circus, playing a game where she has to throw darts at pink and green balloons on a spinning dart board. If she hits a pink balloon, she earns a ticket. If she hits a green balloon, she receives nothing. If she misses entirely, the throw is not counted. There are 4 pink and 6 green balloons. She decided to play the game 29 times. Can you approximate this situation with a normal curve? Why or why not?

Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 9.8.

Students practiced the application of the normal distribution, and were introduced to the Empirical rule and the concept and calculation of z -scores. Z -score probabilities were introduced, and students practiced associating z -scores with probabilities via reference tables and technology. The concept of the sampling distribution of the sample mean was introduced, along with the Central Limit Theorem. Finally, students practiced using the normal distribution to approximate the output of binomial random variables.

9.9 References

1. . . CC BY-NC-SA
2. OpenClips. <http://pixabay.com/en/t-shirts-shirts-clothing-apparel-154312/?oq=shirt> .
3. . . CC BY-NC-SA
4. . . CC BY-NC-SA
5. stux. <http://pixabay.com/en/gasoline-prices-petrol-fuel-206098/?oq=gas%20price> .
6. . . CC BY-NC-SA
7. . . CC BY-NC-SA
8. . . CC BY-NC-SA
9. PublicDomainPictures. <http://pixabay.com/en/snow-day-georgia-back-yard-13962/?oq=snow%20yard> .
10. . . CC BY-NC-SA
11. Simon . <http://pixabay.com/en/touring-skis-ski-touring-binding-262028/?oq=skis> .
12. . . CC BY-NC-SA
13. . . CC BY-NC-SA
14. . . CC BY-NC-SA
15. . . CC BY-NC-SA
16. geralt. <http://pixabay.com/en/personal-group-silhouettes-man-112391/?oq=group> .
17. PublicDomainPictures. <http://pixabay.com/en/spinach-pizza-cheese-yellow-white-72123/?oq=cheese%20pizza> .
18. U.S. Department of Agriculture. <https://www.flickr.com/photos/usdagov/6277240734/> .
19. Highway Patrol Images. <https://www.flickr.com/photos/special-fx/6966663529/> .
20. . . CC BY-NC-SA
21. . . CC BY-NC-SA
22. ideowl. <https://www.flickr.com/photos/special-fx/6966663529/> .