The Normal Distribution

The Normal distribution breaks into five question types.

Properties of the Normal Curve

You should be able to answer questions about the normal curve from a multiple choice or true/false framework.

A normal or approximately normal distribution has the following characteristics:

- The distribution is **bell-shaped, symmetric** and **unimodal.**
- We can denote that a variable, X, comes from a normal distribution with notation:

$$\mathbf{X \sim N(\mu, \sigma)}$$

Z-Scores

If you want to compare values from two distributions, you can simply find and compare their Z-scores. Taking Z-scores allows us to put the values on the same scale (**the standard normal distribution**). You should know how to compare values using Z-scores as well as the basic properties and definition of a Z-score.

The Standard Normal Distribution

This distribution is just a special case of the normal distribution – one with a mean of 0 and a standard deviation of 1. This is what our normal table is made with and your calculator uses.

- We denote variables from the standard normal with a Z:

$$\mathbf{Z \sim N(0, 1)}$$

Z-Scores

A Z-score gives us the number of standard deviations away from the mean a value is, as well as if that value is below or above the mean. You can find the Z-score for a value with the given formula:

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

- If you "standardized" your last test score and found you had a Z-score of 2, that would mean that your score is 2 standard deviations above the mean. A Z-score of -1 would mean that your score was 1 standard deviation below the mean.
- Scores further away from zero are more unusual, regardless of sign.
- Be careful about direction if asking who performed "better" – for a test, a higher Z-score would be better since scoring higher on the test is better; however, for a race, a lower Z-score is better since a smaller time is actually a better performance.

Example: Assume Stat119 midterm scores are normally distributed. The first midterm exam had a mean of 70.2 and a standard deviation of 9.3. The second midterm had a mean score of 76.4 and a standard deviation of 7.8. If Nicole scored a 75 on the first midterm and an 80 on the second midterm, which midterm did she perform better on?

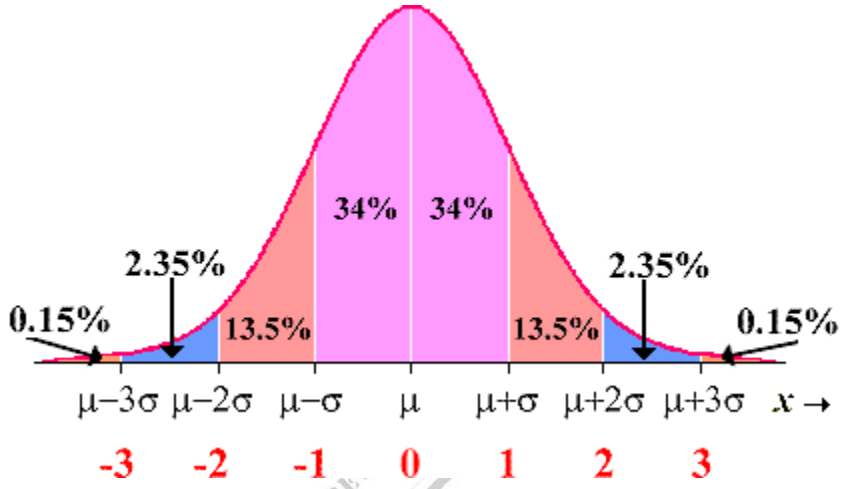
A. The first because the Z-score is greater
B. The second because the Z-score is greater
C. The first because the Z-score is closer to 0
D. The second because the Z-score is closer to 0

The Empirical Rule

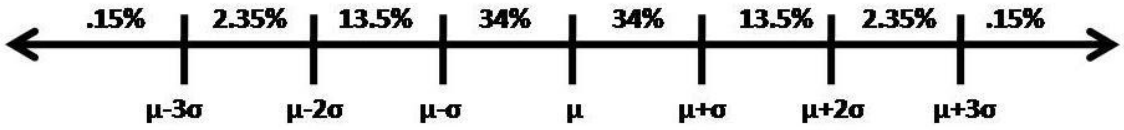
The Empirical Rule provides an estimate of the percent of data falling between certain values on a normal curve. **Do NOT use the Empirical Rule unless the question expressly states to solve using the Empirical Rule.**

Memorize the FOUR values corresponding to the amounts inside each section of half the curve.

The **first** thing to do when a question says "using the Empirical Rule" is to draw the curve with the FOUR inside percents putting the values at the bottom.



Remember, you don't necessarily need to draw the curve. A+ Review "dumbs" this down by simply writing the values in a rectangle instead. If you like this idea, use a number line. Same concept and ease, but actually correct since the data goes to infinity on either direction.



After that, it's as simple as adding up the percentages that correspond to area you're being asked about.

Hints and Tips:

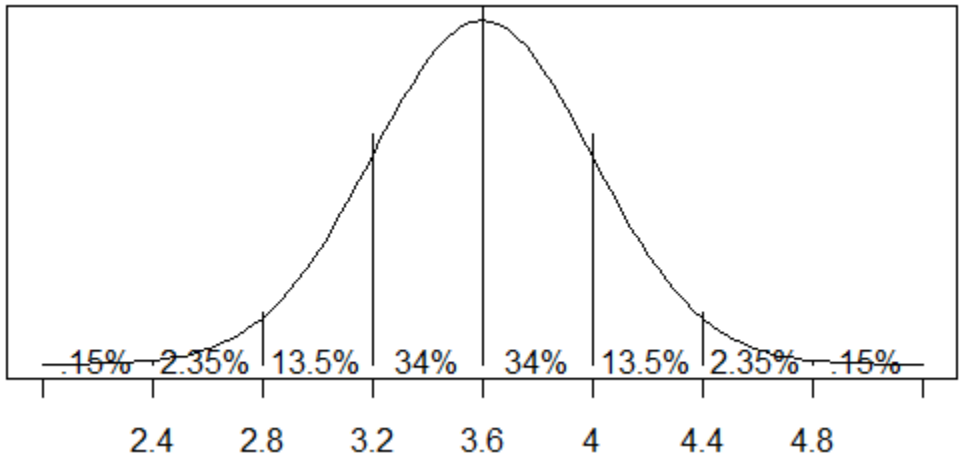
- If you really hate the Empirical Rule (I do!), then you could always solve these problems as if they were a regular normal problem. Your answer will be just slightly off from the estimate provided by the Empirical Rule. (Only works on multiple choice.)
- If you are looking for values that aren't the ones you've written at the bottom of the curve, you've done something wrong. Either you're using the Empirical Rule when the question didn't say to OR you made a simple math error when constructing your curve.
- For questions that give you a percentage and ask what the value is, first determine if they are talking about a value on the right side or the left side (there will be a keyword letting you know where the percent is). Then tackle it in the forward direction using the three values on that side of the curve.

- You may still be asked a question about the other three values: 68-95-99.7%, but it would only be in a multiple choice or true/false question.

Example: True or false. Approximately 95% of the data in a normal distribution falls within 2 standard deviations of the mean.

Example: The weights of adorable, fluffy kittens are normally distributed with a mean of 3.6 pounds and a standard deviation of 0.4 pounds. Answer the following questions, **using the Empirical Rule.**

First, draw your Empirical curve with the 4 percentages!



What percent of adorable, fluffy kittens weigh between 2.8 and 4.8 pounds?

What percent of adorable, fluffy kittens weigh less than 2.4 pounds?

What value corresponds to a 97.5th percentile of kitten weights?

16% of kittens weigh less than what amount?

The Normal Distribution

In normal questions, we can be asked one of two things. We'll either be given a value and asked to find the percentage above or below that amount (or between two values) OR we'll be given a percentage and asked to find the associated value (or possibly the mean or standard deviation).

Your first step on a question that mentions that data is **normally** distributed, is to determine which question is being asked.

Your second step is to **DRAW A PICTURE.** This will keep you from making silly mistakes.

Direct Calculations

This is when we're given a value and asked to find a probability or percent above or below that value (or between two values). Normally this is as easy as looking at the question, did they ask you: "what is the **probability**…" or "what **percent**"?

These problems boil down to the following sequence:

$$\mathbf{X} \rightarrow \mathbf{Z} \rightarrow \%$$

We will take our X, convert it to a Z, then use that Z to find a percent using the table.
This is the formula to convert your cutoff value (x) to a Z-score.

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

Now, it's time to draw our curve:

Like in the Empirical Rule questions, you'll draw your curve centered at the mean. Then you will draw in the give X value(s) and shade the part of the curve the question is asking you for.

Example: Let's revisit those adorable, fluffy kittens. The weights of adorable, fluffy kittens are normally distributed with a mean of 3.6 pounds and a standard deviation of 0.4 pounds.

What is the probability a kitten weighs more than 3 pounds?

Reading this problem we see some keywords:

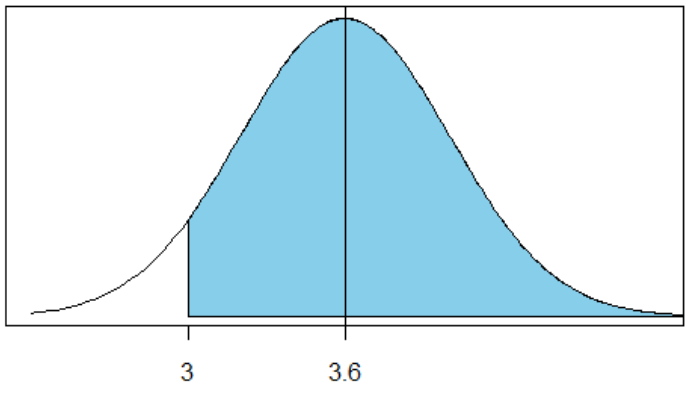
*"The weights of adorable, fluffy kittens are **normally** distributed with a mean of 3.6 pounds and a standard deviation of 0.4 pounds.*

***What is the probability** a kitten weighs **more** than 3 pounds?"*

So, we know it's a normal question. And we see that it's a direct question: $\boldsymbol{X \rightarrow Z \rightarrow \%}$

Next, let's translate what we've been given: $\boldsymbol{\mu = 3.6,\ \sigma = 0.4,\ X = 3}$

Now, we can draw the picture: *The key word of **more** lets us know we should shade the right side of the graph. And once we draw our picture, we see that we should expect to get an answer larger than 50%. If we don't, we know we made a mistake.*



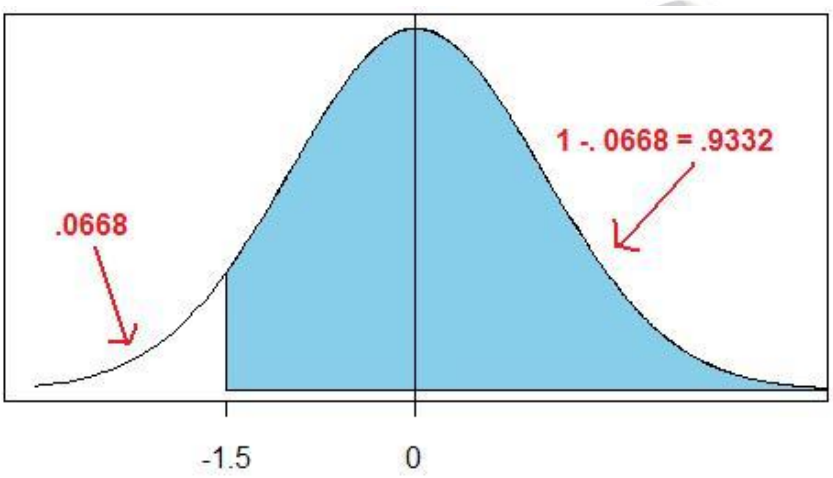
Now that we've drawn a picture to make sure we don't make a silly mistake , we can work through the steps. First, we need to convert X to Z:

$$z = \frac{x - \mu}{\sigma} = \frac{3 - 3.6}{0.4} = -1.5$$

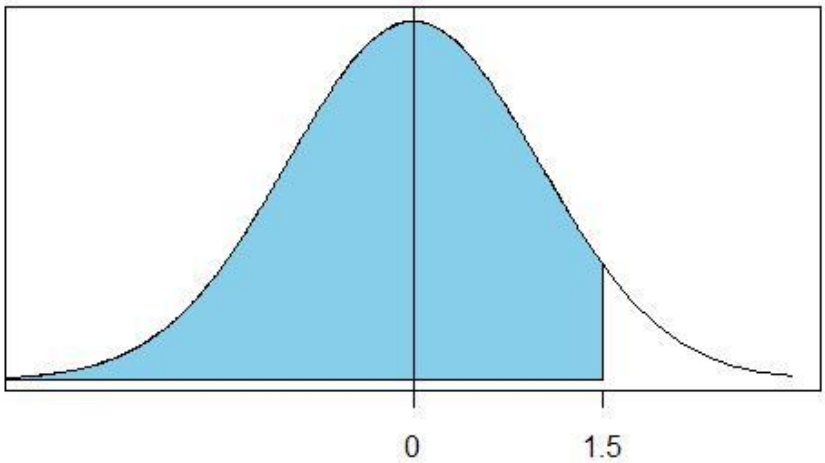
Second, we need to convert this Z into a percent (or probability):

For the right hand probability, we have two methods to calculate this. I suggest using whichever your TA uses to keep from getting confused. I will present both. (Or whichever you like, so long as you practice it on your own!)

Method 1: *Look up the Z-score you got in the first step in your Z table. Subtract this from 1 to get the right side since your table gives the left side and the whole curve is 1 (100%). If you use this method, the drawing is CRUCIAL!*



Method 2: *Since the Normal curve is symmetric, we can simply look up the opposite Z-score and we'll get the correct answer! (Look up 1.5 instead)*



Note: ONLY use Method 2 on GREATER THAN questions. For all other questions, you will look up the Z-scores as calculated!

We're done. Our answer using either method is .9332.

Example: Suppose that speeds that cars drive on a freeway with a speed limit of 65 (without traffic) are approximately normally distributed with a mean of 65 mph and a standard deviation of 4.5 mph. What is the probability that a car chosen at random will be driving between 62 and 70 mph?

Hints and Tips:

- Use the keywords to determine the type of question and where you should shade.
- Draw your curve. It will save you from wrong answer traps on multiple choice questions!
- On the between questions, you will get two Z-scores, look them both up and subtract the smaller one from the larger one. No matter what, your answer is a probability, which means it needs to be between 0 and 1.

Inverse Calculations

If at any point in the question they give you a percentage, percentile or quartile, you're doing an inverse question.

These problems boil down to the following sequence:

$$\% \rightarrow \mathbf{Z} \rightarrow \mathbf{X}$$

We will take our %, use it to find a Z, then use that Z to in the inverse formula to find X (or the mean or the standard deviation). This is the formula you should use for these questions, it is given:

$$\mathbf{x = z(\sigma) + \mu}$$

Now, it's time to draw our curve:

You first need to determine if the percentage is on the left, right or middle of your curve.

Here are some examples of translating these questions:

Example: Birth weights are normally distributed with a mean of 109 oz and a standard deviation of 13oz. Answer the following questions using this information.

A hospital puts any infant in the **lightest 2%** into the NICU, what is the cutoff for putting a baby in the NICU? *Left side*

Your friend had a baby that was in the **95th percentile** for weight, how much did the baby weigh? *Left side*

The **heaviest 10%** of babies weigh more than what amount? *Right side*

If a baby is considered normal if its weight falls within the **middle 80%**, what must a normal baby weigh between? *Middle*

What is **Q3** for the distribution of baby weights? *Left side, remember a Q3 is the 75th percentile!*

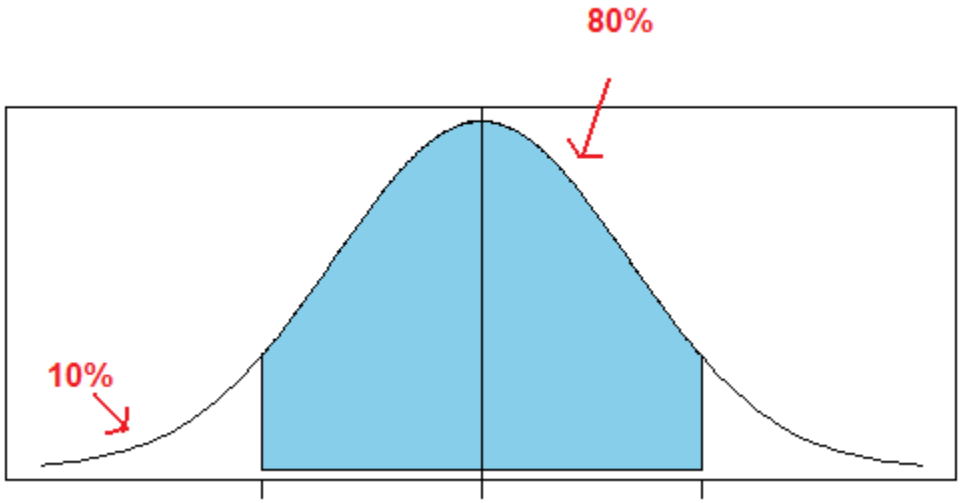
Let's try answering one of these using our steps.

Example: Birth weights are normally distributed with a mean of 109 oz and a standard deviation of 13oz. If a baby is considered normal if its weight falls within the **middle 80%**, what must a normal baby weigh between?

We recognize this is an inverse question because they wrote a percentage! So, we know we're doing the following steps: $\% \rightarrow \boldsymbol{Z} \rightarrow \boldsymbol{X}$

Translate what was given to us: $\boldsymbol{\mu = 109},\ \boldsymbol{\sigma = 13},\ \boldsymbol{middle\ 80\ \%}$

Let's draw our picture:



Since 80% was in the middle and the whole curve represents 100% of the data and is symmetric, there must be 10% in the left tail (which our table and our calculator use.)

Now that we have the left-hand percent, we can get a Z from our table or calculator.

Table: *Hunt in the middle for .1, once you find the closest thing to it, determine the Z that gives that value.*

TI-83 and TI-84: *Z = invNorm(.1)*

Since it's a middle question and the graph is symmetric, you will also want to use the positive version of Z as an answer.

Second step, convert the Z (or in this case, Zs) to an X using the formula.

Hints and Tips:

- Again, draw the curve. Most students make a mistake by using the opposite Z, if you draw your picture by looking for keywords, you won't make this mistake.
- Always use the inverse formula for these problems, even if you're solving for the mean or standard deviation instead.

