

6.2: Probability in Graphs and Distributions

We will see shortly that the normal distribution is the key to how probability works for our purposes. To understand exactly how, let's first look at a simple, intuitive example using pie charts.

Probability in Pie

Charts Recall that a pie chart represents how frequently a category was observed and that all slices of the pie chart add up to 100%, or 1. This means that if we randomly select an observation from the data used to create the pie chart, the probability of it taking on a specific value is exactly equal to the size of that category's slice in the pie chart.

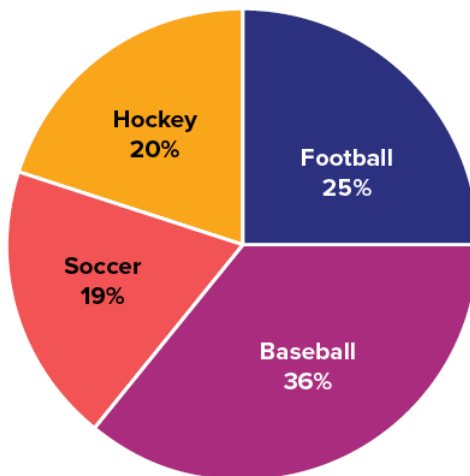


Figure 6.2.1: Favorite sports

Image Credit: Judy Schmitt, from Cote et al, 2021.

Take, for example, the pie chart in Figure 6.2.1 representing the favorite sports of 100 people. If you put this pie chart on a dart board and aimed blindly (assuming you are guaranteed to hit the board), the likelihood of hitting the slice for any given sport would be equal to the size of that slice. So, the probability of hitting the baseball slice is the highest at 36%. The probability is equal to the proportion of the chart taken up by that section.

We can also add slices together. For instance, maybe we want to know the probability to finding someone whose favorite sport is usually played on grass. The outcomes that satisfy this criteria are baseball, football, and soccer. To get the probability, we simply add their slices together to see what proportion of the area of the pie chart is in that region: $36\% + 25\% + 19\% = 81\%$. We can also add sections together even if they do not touch. If we want to know the likelihood that someone's favorite sport is not called football somewhere in the world (i.e. baseball and hockey), we can add those slices even though they aren't adjacent or continuous in the chart itself: $36\% + 20\% = 56\%$. We are able to do all of this because 1) the size of the slice corresponds to the area of the chart taken up by that slice, 2) the percentage for a specific category can be represented as a decimal (this step was skipped for ease of explanation above), and 3) the total area of the chart is equal to 100% or 1.0, which makes the size of the slices interpretable.

Probability in Normal Distributions

If the language at the end of the last section sounded familiar, that's because its exactly the language used in the last chapter to describe the normal distribution. Recall that the normal distribution has an area under its curve that is equal to 1 and that it can be split into sections by drawing a line through it that corresponds to a given z -score. Because of this, we can interpret areas under the normal curve as probabilities that correspond to z -scores.

First, let's look back at the area between $z = -1.00$ and $z = 1.00$ presented in Figure 6.2.2. We were told earlier that this region contains 68% of the area under the curve. Thus, if we randomly chose a z -score from all possible z -scores, there is a 68% chance that it will be between $z = -1.00$ and $z = 1.00$ because those are the z -scores that satisfy our criteria.

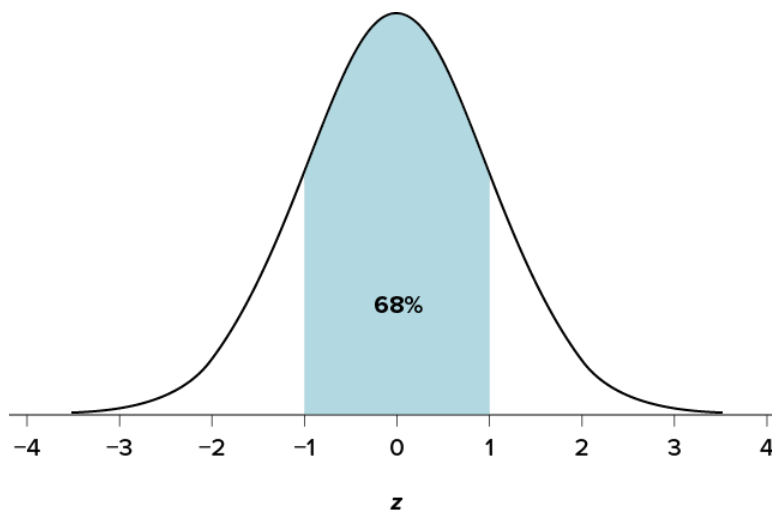


Figure 6.2.2: There is a 68% chance of selection a z -score from the blue-shaded region

Image Credit: Judy Schmitt, from Cote et al, 2021.

Just like a pie chart is broken up into slices by drawing lines through it, we can also draw a line through the normal distribution to split it into sections. Take a look at the normal distribution in Figure 6.2.3 which has a line drawn through it as $z = 1.25$. This line creates two sections of the distribution: the smaller section called the tail and the larger section called the body. Differentiating between the body and the tail does not depend on which side of the distribution the line is drawn. All that matters is the relative size of the pieces: bigger is always body.

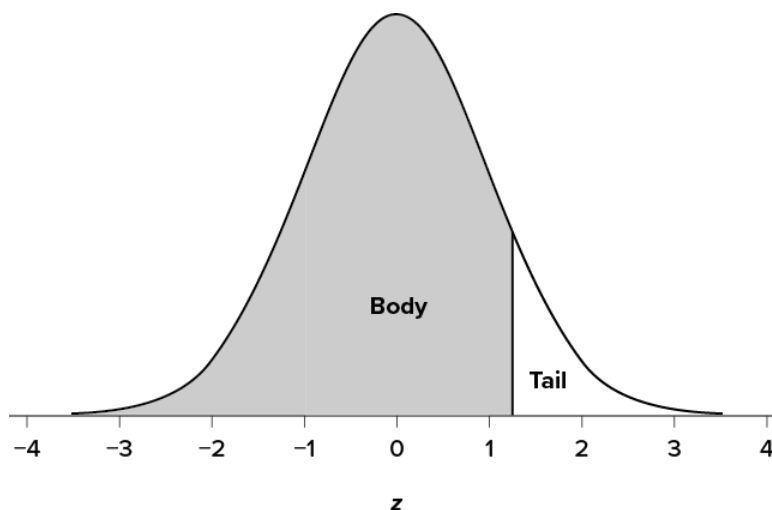


Figure 6.2.3: Body and tail of the normal distribution

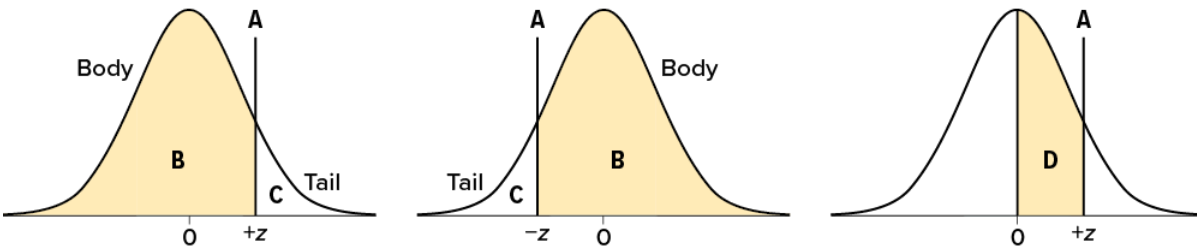
Image Credit: Judy Schmitt, from Cote et al, 2021.

As you can see, we can break up the normal distribution into 3 pieces (lower tail, body, and upper tail) as in Figure 6.2.2 or into 2 pieces (body and tail) as in Figure 6.2.3. We can then find the proportion of the area in the body and tail based on where the line was drawn (i.e. at what z -score). Mathematically this is done using calculus. Fortunately, the exact values are given you to you in the Standard Normal Distribution Table, also known at the z -table. Using the values in this table, we can find the area under the normal curve in any body, tail, or combination of tails no matter which z -scores are used to define them.

The z -table presents the values for the area under the curve to the left of the positive z -scores from 0.00-3.00 (technically 3.09), as indicated by the shaded region of the distribution at the top of the table. To find the appropriate value, we first find the row corresponding to our z -score then follow it over until we get to the column that corresponds to the number in the hundredths place of our z -score. For example, suppose we want to find the area in the body for a z -score of 1.62. We would first find the row for 1.60 then follow it across to the column labeled 0.02 ($1.60 + 0.02 = 1.62$) and find 0.9474 (see Table 5.1). Thus, the odds of

randomly selecting someone with a z -score less than (to the left of) $z = 1.62$ is 94.74% because that is the proportion of the area taken up by values that satisfy our criteria.

TABLE 5.1. Standard normal distribution table (z table).



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion between Mean and z
1.60	.9452	.0548	.4452	1.80	.9641	.0359	.4641
1.61	.9463	.0537	.4463	1.81	.9649	.0351	.4649
1.62	.9474	.0526	.4474	1.82	.9656	.0344	.4656
1.63	.9484	.0516	.4484	1.83	.9664	.0336	.4664
1.64	.9495	.0505	.4495	1.84	.9671	.0329	.4671
1.65	.9505	.0495	.4505	1.85	.9678	.0322	.4678
1.66	.9515	.0485	.4515	1.86	.9686	.0314	.4686
1.67	.9525	.0475	.4525	1.87	.9693	.0307	.4693
1.68	.9535	.0465	.4535	1.88	.9699	.0301	.4699
1.69	.9545	.0455	.4545	1.89	.9706	.0294	.4706
1.70	.9554	.0446	.4554	1.90	.9713	.0287	.4713
1.71	.9564	.0436	.4564	1.91	.9719	.0281	.4719
1.72	.9573	.0427	.4573	1.92	.9726	.0274	.4726
1.73	.9582	.0418	.4582	1.93	.9732	.0268	.4732
1.74	.9591	.0409	.4591	1.94	.9738	.0262	.4738
1.75	.9599	.0401	.4599	1.95	.9744	.0256	.4744
1.76	.9608	.0392	.4608	1.96	.9750	.0250	.4750
1.77	.9616	.0384	.4616	1.97	.9756	.0244	.4756
1.78	.9625	.0375	.4625	1.98	.9761	.0239	.4761
1.79	.9633	.0367	.4633	1.99	.9767	.0233	.4767

("z Table Curves" by Judy Schmitt is licensed under [CC BY-NC-SA 4.0.](https://creativecommons.org/licenses/by-nc-sa/4.0/))

Table 5.1: Using the z -table to find the area in the body to the left of $z = 1.62$

Image Credit: Judy Schmitt, from Cote et al, 2021.

The z -table only presents the area in the body for positive z -scores because the normal distribution is symmetrical. Thus, the area in the body of $z = 1.62$ is equal to the area in the body for $z = -1.62$, though now the body will be the shaded area to the right of z (because the body is always larger). When in doubt, drawing out your distribution and shading the area you need to find will always help. The table also only presents the area in the body because the total area under the normal curve is always equal to 1.00, so if we need to find the area in the tail for $z = 1.62$, we simply find the area in the body and subtract it from 1.00 ($1.00 - 0.9474 = 0.0526$).

Let's look at another example. This time, let's find the area corresponding to z -scores more extreme than $z = -1.96$ and $z = 1.96$. That is, let's find the area in the tails of the distribution for values less than $z = -1.96$ (farther negative and therefore more extreme) and greater than $z = 1.96$ (farther positive and therefore more extreme). This region is illustrated in Figure 6.2.5.

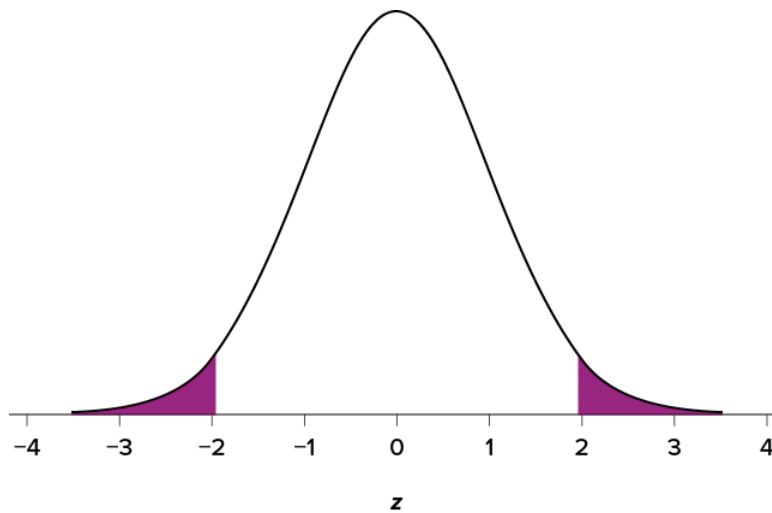


Figure 6.2.5: Area in the tails beyond $z = -1.96$ and $z = 1.96$

Image Credit: Judy Schmitt, from Cote et al, 2021.

Let's start with the tail for $z = 1.96$. If we go to the z -table we will find that the body to the left of $z = 1.96$ is equal to 0.9750. To find the area in the tail, we subtract that from 1.00 to get 0.0250. Because the normal distribution is symmetrical, the area in the tail for $z = -1.96$ is the exact same value, 0.0250. Finally, to get the total area in the shaded region, we simply add the areas together to get 0.0500. Thus, there is a 5% chance of randomly getting a value more extreme than $z = -1.96$ or $z = 1.96$ (this particular value and region will become incredibly important in Unit 2).

Finally, we can find the area between two z -scores by shading and subtracting. Figure 6.2.6 shows the area between $z = 0.50$ and $z = 1.50$. Because this is a subsection of a body (rather than just a body or a tail), we must first find the larger of the two bodies, in this case the body for $z = 1.50$, and subtract the smaller of the two bodies, or the body for $z = 0.50$. Aligning the distributions vertically, as in Figure 6, makes this clearer. From the z -table, the area in the body for $z = 1.50$ is 0.9332 and the area in the body for $z = 0.50$ is 0.6915. Subtracting these gives us $0.9332 - 0.6915 = 0.2417$.

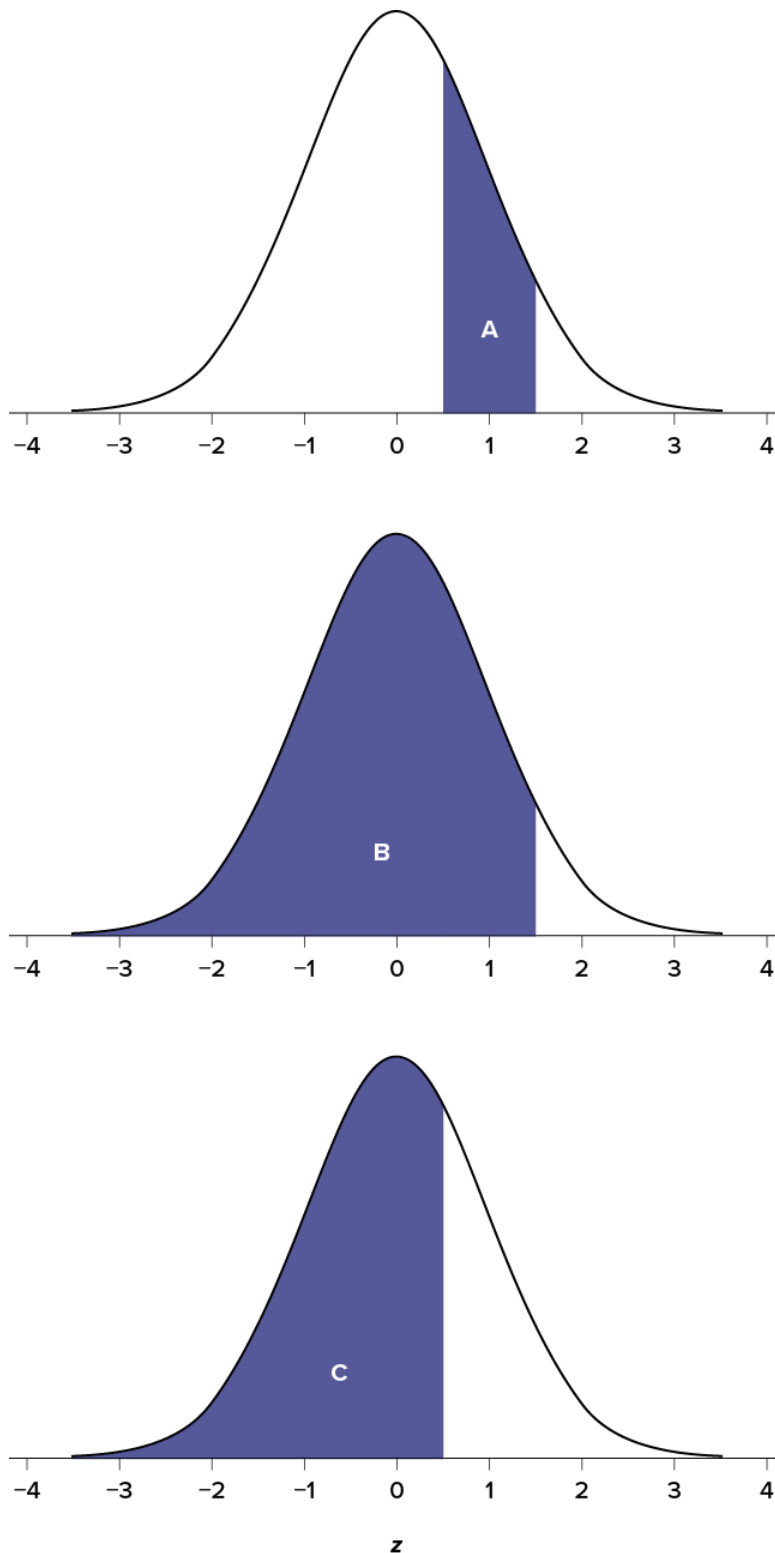


Figure 6.2.6: Area between $z = 0.50$ and 1.50 , along with the corresponding areas in the body

Image Credit: Judy Schmitt, from Cote et al, 2021.

This page titled [6.2: Probability in Graphs and Distributions](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Foster et al. \(University of Missouri's Affordable and Open Access Educational Resources Initiative\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **5.2: Probability in Graphs and Distributions** by Foster et al. is licensed CC BY-NC-SA 4.0. Original source: <https://irl.umsl.edu/oer/4>.