

The Normal Distribution

Diana Mindrila, Ph.D.

Phoebe Baletnyne, M.Ed.

Based on Chapter 3 of The Basic Practice of Statistics (6th ed.)

Concepts:

- Density Curves
- Normal Distributions
- The 68-95-99.7 Rule
- The Standard Normal Distribution
- Finding Normal Proportions
- Using the Standard Normal Table
- Finding a Value When Given a Proportion

Objectives:

- Define and describe density curves
- Measure position using percentiles
- Measure position using z-scores
- Describe Normal distributions
- Describe and apply the 68-95-99.7 Rule
- Describe the standard Normal distribution
- Perform Normal calculations

References:

Moore, D. S., Notz, W. I., & Fligner, M. A. (2013). *The basic practice of statistics* (6th ed.). New York, NY: W. H. Freeman and Company.

Density Curves

Exploring Quantitative Data

1. Always plot data first: make a graph.
2. Look for the overall pattern (shape, center, and spread) and for striking departures such as outliers.
3. Calculate a numerical summary to briefly describe center and spread.
4. Sometimes the overall pattern of a large number of observations is so regular that it can be described by a smooth curve.

- When describing data, always start with a graphical representation.
- Graphs help identify the overall distribution pattern. Looking at a graph makes it visually clear how spread a variable is, which values occur most frequently, and whether or not the distribution is skewed.
- Next, obtain more precise information by providing a numerical summary of the data using the mean, median, range, five-number summary, and any other appropriate information.
- Some distributions are so regular that they can be described by a smooth curve. Real data are represented in a histogram. Curves represent a symbol, or an abstract version of a distribution.

A **density curve** is a curve that:

- is always on or above the horizontal axis
- has an area of exactly 1 underneath it

A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values on the horizontal axis is the proportion of all observations that fall in that range.

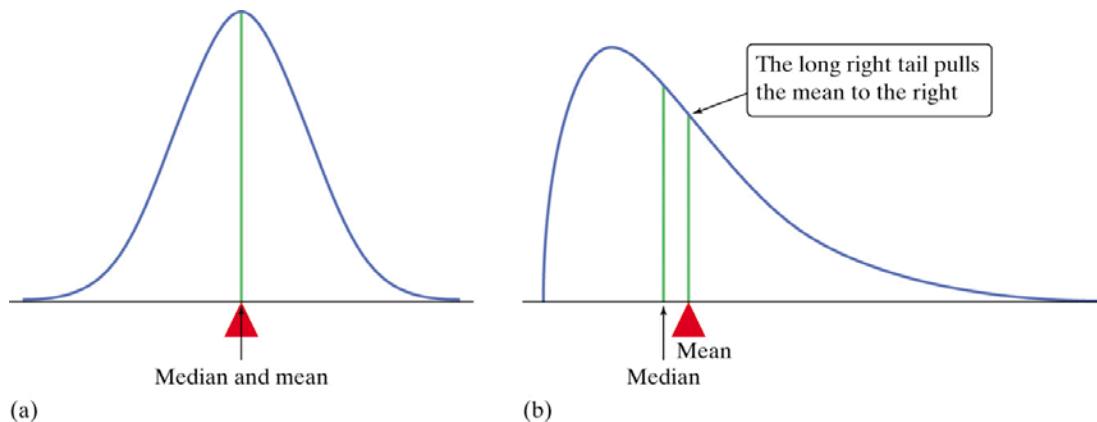
- **Density curves** are lines that show the location of the individuals along the horizontal axis and within the range of possible values.
- They help researchers to investigate the distribution of a variable.
- Some density curves have certain properties that help researchers draw conclusions about the entire population.

Density Curves

- Measures of center and spread apply to density curves as well as to actual sets of observations.

Distinguishing the Median and Mean of a Density Curve

- The median of a density curve is the equal-areas point, the point that divides the area under the curve in half.
 - The mean of a density curve is the balance point, at which the curve would balance if made of solid material.
 - The median and the mean are the same for a symmetric density curve. They both lie at the center of the curve. The mean of a skewed curve is



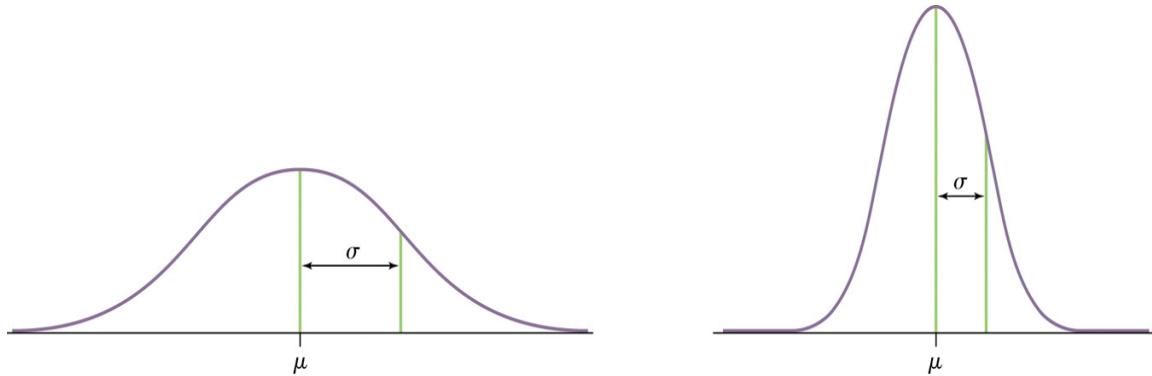
- The mean, median, and mode can also be represented on density curves.
 - When a distribution is symmetric or Normal, the mean and median overlap. The actual recorded values may be slightly different, but they are very close.
 - The mode will always be located at the highest point on the curve, because it shows the value that occurs most frequently.
 - The median shows the point that divides the area under the curve in half, whereas the mean, which is drawn toward the extreme observations, shows the balance point.

Density Curves

- The mean and standard deviation computed from actual observations (data) are denoted by \bar{x} and s , respectively
- The mean and standard deviation of the actual distribution represented by the density curve are denoted by μ ("mu") and σ ("sigma"), respectively.
 - The mean and standard deviation (\bar{x} and s) are called *statistics*, and they can be computed based on observations in the sample.
 - The mean and standard deviation of the density curves (μ and σ) are called *parameters*. They describe the entire population and are only estimated. With very few exceptions, the real value of the population is unknown and the values must be estimated, with a certain degree of confidence, based on observations from the sample.

Normal Distributions

- One particularly important class of density curves are the Normal curves, which describe Normal distributions.
- All Normal curves are symmetric, single-peaked, and bell-shaped.
- A Specific Normal curve is described by giving its mean μ and standard deviation σ .



- Density curves are used to illustrate many types of distributions.
- The Normal distribution, or the bell-shaped distribution, is of special interest. This distribution describes many human traits. All Normal curves have symmetry, but not all symmetric distributions are Normal.
- Normal distributions are typically described by reporting the mean, which shows where the center is located, and the standard deviation, which shows the spread of the curve, or the distance from the mean.
 - When the standard deviation is large, the curve is wider like the example on the left.
 - When the standard deviation is small, the curve is narrower like the example on the right.
- One example of a variable that has a Normal distribution is IQ. In the population, the mean IQ is 100 and its standard deviation, depending on the test, is 15 or 16. If a large enough random sample is selected, the IQ distribution of the sample will resemble the Normal curve. The larger the sample, the more clear the pattern will be.

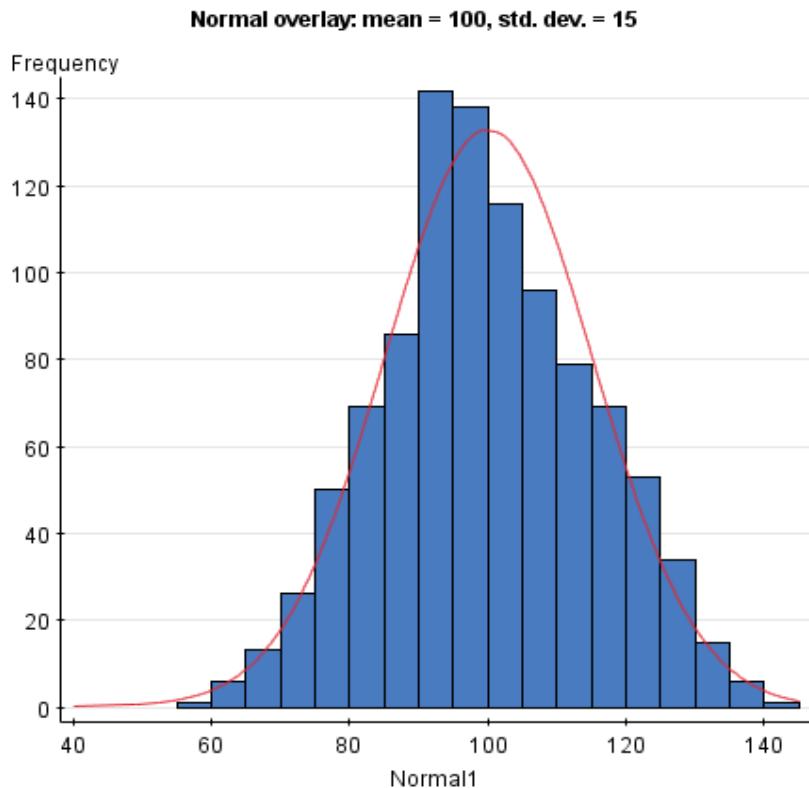
Normal Distributions

- A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers: its mean μ and its standard deviation σ .
 - The mean of a Normal distribution is the center of the symmetric **Normal curve**.
 - The standard deviation is the distance from the center to the change-of-curvature points on either side.
 - The Normal distribution is abbreviated with mean μ and standard deviation σ as $N(\mu, \sigma)$

Normal Curve

Example: IQ score distribution based on the Standford-Binet Intelligence Scale

The smooth curve drawn over the histogram is a **mathematical model** for the distribution.



- The histogram in this image represents a distribution of real IQ scores as measured by the Standford-Binet Intelligence Scale.
- The blue bars represent the number of individuals who recorded IQ scores within a certain 5-point range.
- The main purpose of a histogram is to illustrate the general distribution of a set of data.
- This variable has a mean of 100 and a standard deviation of 15.
- The curve that is drawn over the histogram is the Normal curve, and it summarized the distribution of the recorded scores.

Normal Curve

The areas of the shaded bars in this histogram represent the proportion of scores in the observed data that are less than or equal to 90.

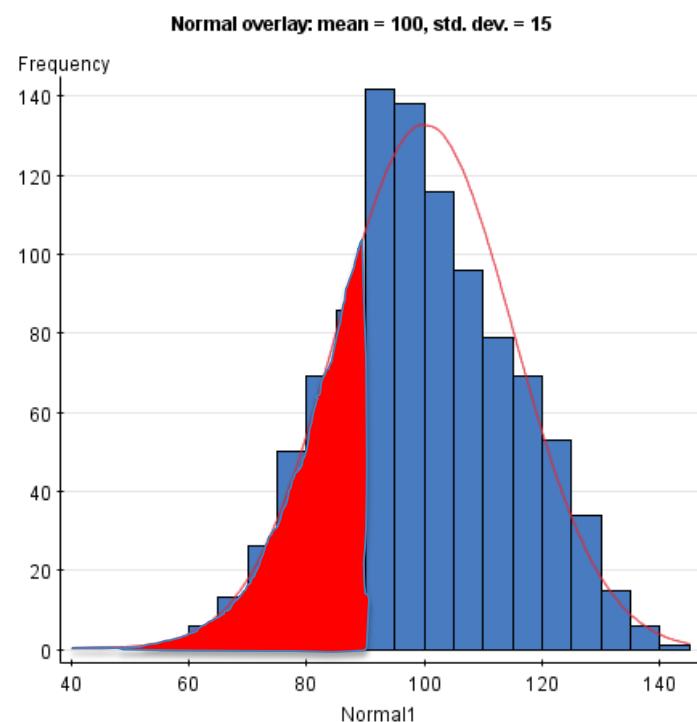
Total: N = 1015

IQ<90: N = 256 (25.22%)

Now the area under the smooth curve to the left of 90 is shaded. If the scale is adjusted so the total area under the curve is exactly 1, then this curve is called a **density curve**.

Total Area = 1

Shaded Area = 0.2546



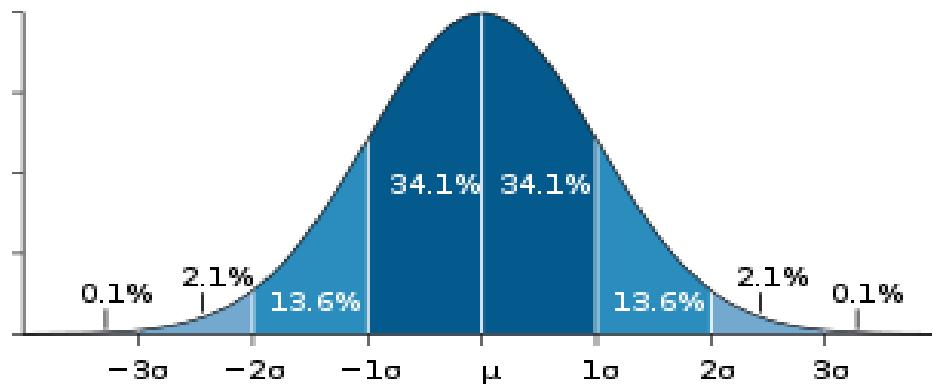
- The entire area under the curve represents all the individuals in the sample.
- If only part of the area is shaded, this represents the proportion of individuals who scored below a certain point.
- In this above example, the area under the curve represents all the individuals in the sample. In this case, they add up to 1,015. This number represents 100% of the sample.
- The shaded area in the above example represents the individuals who had an IQ score below 90. This group consists of 256 individuals.
- To find the percentage, divide the number in the group by the total number, and then multiply by 100. In this case, 256 divided by 1015 times 100 results in a percentage of 25.22. This means that 25.22% of the individuals in this sample had an IQ score below 90.
- The Normal curve is used to find proportions from the entire population, rather than just from the sample. The values for the entire population are often unknown, but if the variable has a Normal distribution, the proportion can be found using only the population mean and standard deviation for that variable.
- Rather than using percentages, statisticians use decimals. Therefore, the entire area under the curve is 1. Using the properties of the Normal curve, the shaded area in the above example is 0.2546. This will be explained in greater detail later.

The 68-95-99.7 Rule

The 68-95-99.7 Rule

In the Normal distribution with mean μ and standard deviation σ :

- Approximately **68%** of the observations fall within σ of μ .
- Approximately **95%** of the observations fall within 2σ of μ .
- Approximately **99.7%** of the observations fall within 3σ of μ .

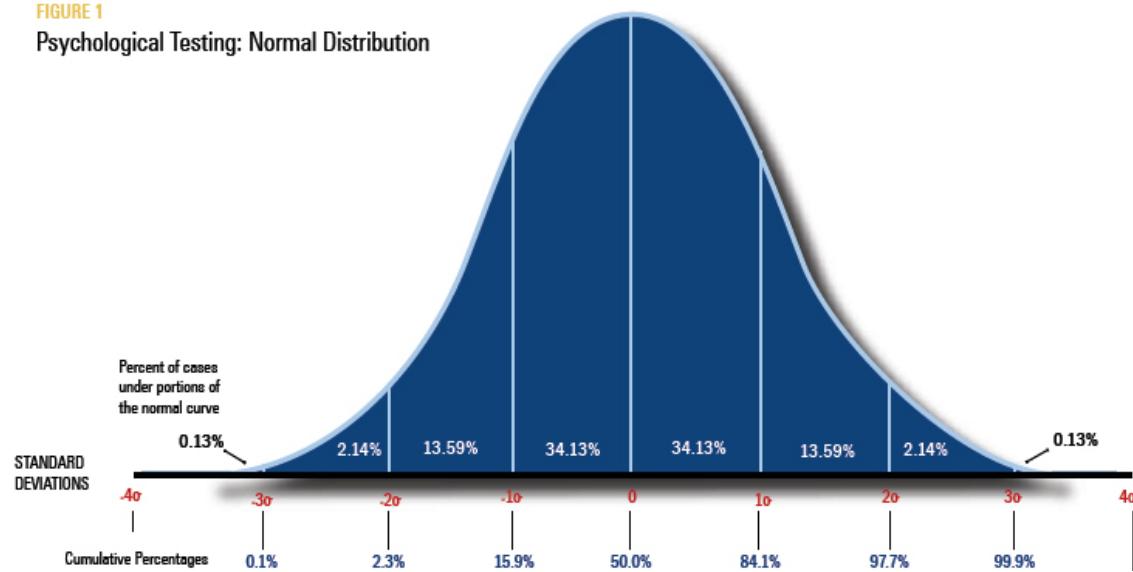


- Normal curves enable researchers to calculate the proportions of individuals who are located within certain intervals. With Normal curves, some intervals are already calculated. This is called the 68-95-99.7 Rule.
- If the population mean and standard deviation for a particular variable are known, the location of the majority of individuals can be quickly found.
- The majority of individuals are located in the highest area of the curve, which is around the mean.
- The intervals within one standard deviation of the mean each account for 34.1% of the population.
- Therefore, approximately 68% of the population is located within one standard deviation above or below the mean.
- The intervals between one and two standard deviations away from the mean in either direction each account for 13.6% of the population.
- Therefore, after adding the percentages in all four intervals, approximately 95% of the population is located within two standard deviations above or below the mean.
- The intervals between two and three standard deviations away from the mean in either direction each account for 2.1% of the population.
- Therefore, approximately 99.7% of the population is located within three standard deviations from the mean.

- Technically, the two tails of the Normal curve extend to positive or negative infinity, but these numbers would be limited for certain variables like IQ, which cannot be smaller than zero.
- The proportion of individuals who are located more than three standard deviations above or below the mean is extremely small: only 0.3%.

The 68-95-99.7 Rule Example

FIGURE 1
Psychological Testing: Normal Distribution



- Figure 1 illustrates how to apply the 68-95-99.7 Rule to the distribution of IQ scores.
- In this example, the population mean is 100 and the standard deviation is 15.
- Based on the 68-95-99.7 Rule, approximately 68% of the individuals in the population have an IQ between 85 and 115. Values in this particular interval are the most frequent.
- Approximately 95% of the population has IQ scores between 70 and 130.
- Approximately 99.7% of the population has IQ scores between 55 and 145.
- Only approximately 0.3% of the population has IQ scores outside of this interval (less than 55 or higher than 145).

The Standard Normal Distribution

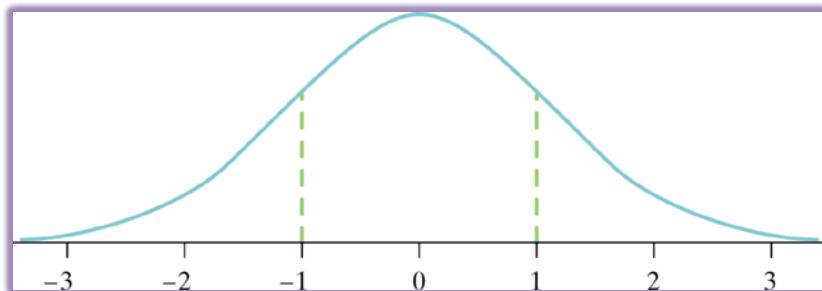
- All Normal distributions are the same if they are measured in units of size σ from the mean μ as center.

The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1.

If a variable x has any Normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ , then the standardized variable

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

has the standard Normal distribution, $N(0,1)$.

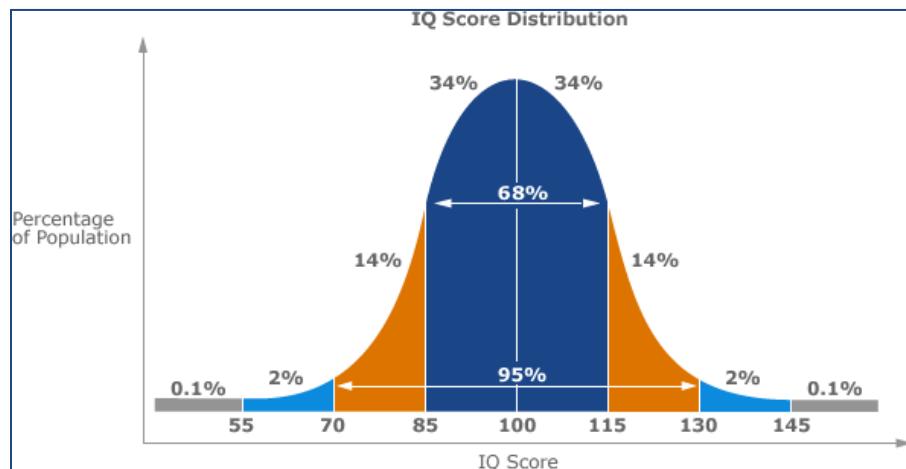


- The Normal curve can be used to describe the distribution of many variables.
- Sometimes researchers want to compare scores that have been measured on different scales. Comparisons are meaningless if scores are not on the same scale. Therefore, to be able to make comparisons across variables, variables that have a Normal distribution can be standardized, which simply means that they are put onto the same scale.
- There are many types of standardized scales. One type of standardized score that researchers use frequently is z scores.
- Z scores are used with variables that have a Normal distribution. Z scores change the values so that the distribution has a mean of 0 and a standard deviation of 1. In theory, z scores can range from negative infinity to positive infinity.
- Z scores can be calculated for every individual in the data set using a simple formula:
 - 1) Compute the difference between the individual's score and the population mean.
 - 2) Divide this difference by the standard deviation.
- If an individual's score is lower than the mean, the z score will be negative. If the individual's score is higher than the mean, the z score will be positive.

Normal Distributions Example

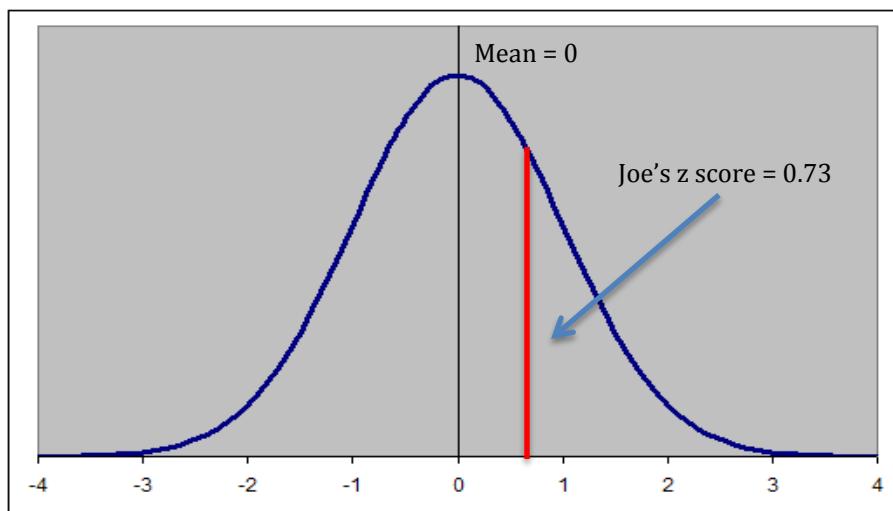
Example:

Joe: IQ = 111
 Sigma = 15
 Pop. Mean = 100



Joe's IQ on the z distribution:

$$\begin{aligned} z &= (111 - 100) / 15 \\ z &= 11 / 15 \\ z &= 0.73 \end{aligned}$$



- In this example, an individual score on a Normal distribution is given.
- The top image shows the IQ score distribution. The bottom image shows the curve from this distribution transformed into z scores.
- To find the z score for Joe's IQ:
 - 1) Subtract the mean from the score ($\text{score} - \text{mean}$) = $(111 - 100) = 11$
 - 2) Divide the difference by the standard deviation $11 / 15 = 0.73$
- Now that a z score has been obtained, it would be helpful to find out the proportion of individuals who have an IQ below 111, or a z score below 0.73. In other words, what is the area under the curve on the left side of this specific score?

The Standard Normal Table

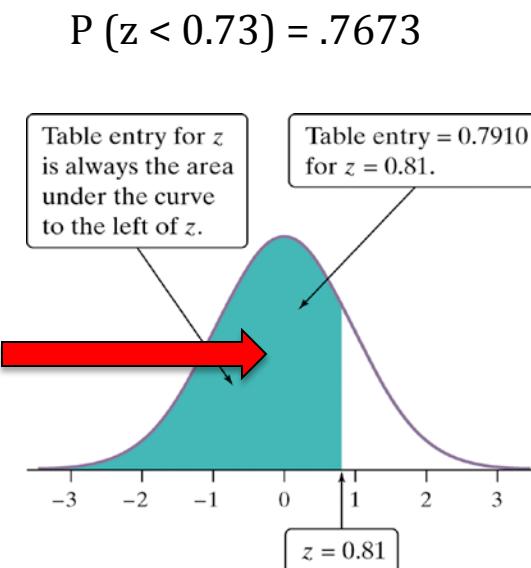
- All Normal distributions are the same when they have been turned into z scores.
- Therefore, areas under any Normal curve can be found using a single table.

The Standard Normal Table

Table A is a table of areas under the standard Normal curve. The table entry for each value z is the area under the curve to the left of z .

To find the proportion of observations from the standard Normal distribution that are less than 0.73, use table A:

Z	.00	.01	.02	.03
0.6	.7257	.7291	.7324	.7357
0.7	.7580	.7611	.7642	.7673
0.8	.7881	.7910	.7939	.7967

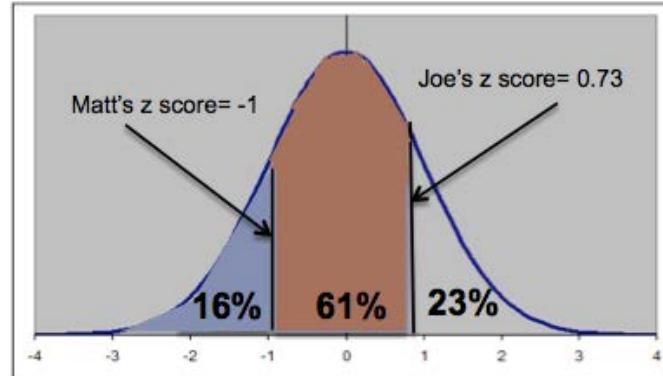


- For every z score, areas on the left side of the curve have already been computed and are listed in a probability table. Statistics textbooks generally store these tables in the appendices.
- This table lists the first two digits of the z score vertically and the last digit horizontally.
- In this example, to find the area under the curve for a z score of 0.73, start by finding 0.7 on the left. Then find 0.03 at the top. Finally, find the cell where this row and column meet. The value in this cell (0.7673) is the area under the curve for a z score of 0.73.
- This value means the probability of a z score being lower than this one is 0.7673.
- Simply put, 76.73% of the population has a z score at or below 0.73.
- In this case, 76.73% of the population has an IQ equal to or lower than 111.

Normal Calculations

- The Normal curve can be used to compute proportions, not only from one standard deviation to another, but also for specific values of interest.

Joe: IQ = 111, z = 0.73
 Matt: IQ = 85, z = -1



- Example: Find the proportion of observations from the standard Normal distribution that are between IQ = 85 and IQ = 111.
 - Find the probability of having an IQ score that is 111 (Joe's score) or lower.
 - Transform Joe's score (111) into a z score ($z = 0.73$)
 - Use the z score table to find the area to the left of this score (0.7673)
 - The probability of having an IQ score that is 111 or lower is approximately 77%
 - Find the probability of having a score that is 85 (Matt's score) or lower.
 - Transform Matt's score (85) into a z score ($z = -1$)
 - Use the z score table to find the area to the left of this score (0.1587)
 - The probability of having an IQ score that is 85 or lower is approximately 16%
 - Find the percentage of the population that has IQ scores between 85 and 111.
 - Take the probability of having an IQ score that is 111 or lower and subtract the probability of having an IQ score that is 85 or lower to find the probability of being in between these two scores.
 - $77\% - 16\% = 61\%$ or $0.7673 - 0.1587 = 0.6086$
 - Approximately 61% of individuals have IQ scores between 111 and 85
- To find the percentage of individuals who score **above** a certain z score, simply subtract the percentage to the left of that score from 100%.
 - Example: Find the percentage of individuals with an IQ score *higher* than 111
 - Find the percentage of individuals who score below 111 (77%)
 - Subtract this percentage from 100% ($100\% - 77\% = 23\%$)
 - Approximately 23% of individuals score above 111

Normal Distributions

- The Normal distribution is very useful for comparing variables that are measured on different scales.
- Example: a graduate student has a score of 25 on a quiz and a score of 56 on the final exam. On which assessment did the student perform better? This depends on the distribution of the two variables.
 - ◆ If the quiz is out of 30 points and the exam is out of 100 points, it may seem clear that the quiz performance was better.
 - ◆ However, using the Normal curve, the most important information is the mean and the standard deviation.
 - ◆ Standardized scores (e.g. z scores) can help to compare scores measured on different scales.

Quiz

$$x_1 = 25$$

$$\text{Mean} = 20$$

$$\text{St. Dev.} = 5$$

$$Z_1 = (25-20)/5$$

$$= 5/5$$

$$= 1$$

Exam

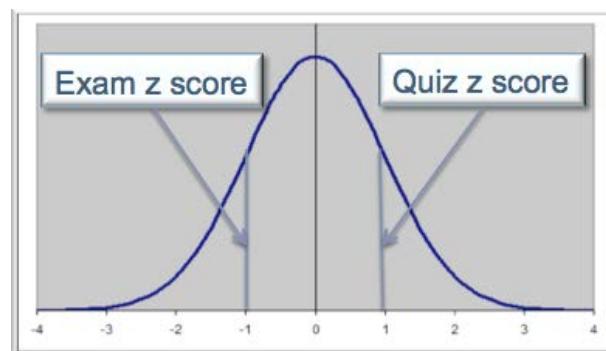
$$x_2 = 56$$

$$\text{Mean} = 68$$

$$\text{St. Dev.} = 12$$

$$Z_2 = (56-68)/12$$

$$= -12/12$$



- ❖ After calculating the z scores for both assessments, it can be concluded that this student performed better on the quiz, even though the raw score of 25 was lower than the raw score of 56.
- ❖ The student's performance on the quiz was one standard deviation *above* the mean and the student's performance on the exam was one standard deviation *below* the mean, resulting in a higher performance on the quiz.
- This example makes it clear that if it is necessary to compare scores that are on different scales, the scores must be standardized or put on the same scale.
- In this case, the standardized scores are z scores, but there are many other kinds of standardized scores.
- If raw scores only are compared, the results can be misleading, as this example demonstrated.
- It is important to note that these comparisons are based on the assumption that the two variables have a Normal distribution in the population.

Normal Calculations

How to Solve Problems Involving Normal Distributions

State: Express the problem in terms of the observed variable x .

Plan: Draw a picture of the distribution and shade the area of interest under the curve.

Do: Perform calculations.

- **Standardize** x to restate the problem in terms of a standard Normal variable z .
- **Use Table A** and the fact that the total area under the curve is 1 to find the required area under the standard Normal curve.

Conclude: Write the conclusion in the context of the problem.