

5.2: The Normal Distribution

Figure 5.2.1: The Normal Distribution

Note that the the tails go to $\pm\infty\pm\infty$. In addition, the density of a distribution over the range of x is the key to hypothesis testing. With a normal distribution, $\sim 68\%\sim 68\%$ of the observations will fall within 1 standard deviation of the mean, $\sim 95\%\sim 95\%$ will fall within 2 standard deviations, and $\sim 99.7\%\sim 99.7\%$ within 3 standard deviations. This is illustrated in Figures 5.2, 5.3, 5.4.

Figure 5.2.1: The Normal Distribution

Note that the the tails go to $\pm\infty\pm\infty$. In addition, the density of a distribution over the range of x is the key to hypothesis testing. With a normal distribution, $\sim 68\%\sim 68\%$ of the observations will fall within 1 standard deviation of the mean, $\sim 95\%\sim 95\%$ will fall within 2 standard deviations, and $\sim 99.7\%\sim 99.7\%$ within 3 standard deviations. This is illustrated in Figures 5.2, 5.3, 5.4.

For purposes of statistical inference, the normal distribution is one of the most important types of probability distributions. It forms the basis of many of the assumptions needed to do quantitative data analysis, and is the basis for a wide range of hypothesis tests. A standardized normal distribution has a mean, μ , of 0 and a standard deviation (s.d.), σ , of 1. The distribution of an outcome variable, Y , can be described:

$$Y \sim N(\mu_Y, \sigma_Y^2) \quad (5.1) \quad Y \sim N(\mu_Y, \sigma_Y^2)$$

Where \sim stands for “distributed as”, N indicates the normal distribution, and mean μ_Y and variance σ_Y^2 are the parameters. The probability function of the normal distribution is expressed below:

The Normal Probability Density Function: The probability density function (PDF) of a normal distribution with mean μ and standard deviation σ :

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

The Standard Normal Probability Density Function: The standard normal PDF has a $\mu=0$ and $\sigma=1$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Using the standard normal PDF, we can plot a normal distribution in R .

```
x <- seq(-4,4,length=200)
y <- 1/sqrt(2*pi)*exp(-x^2/2)
plot(x,y, type="l", lwd=2)
```

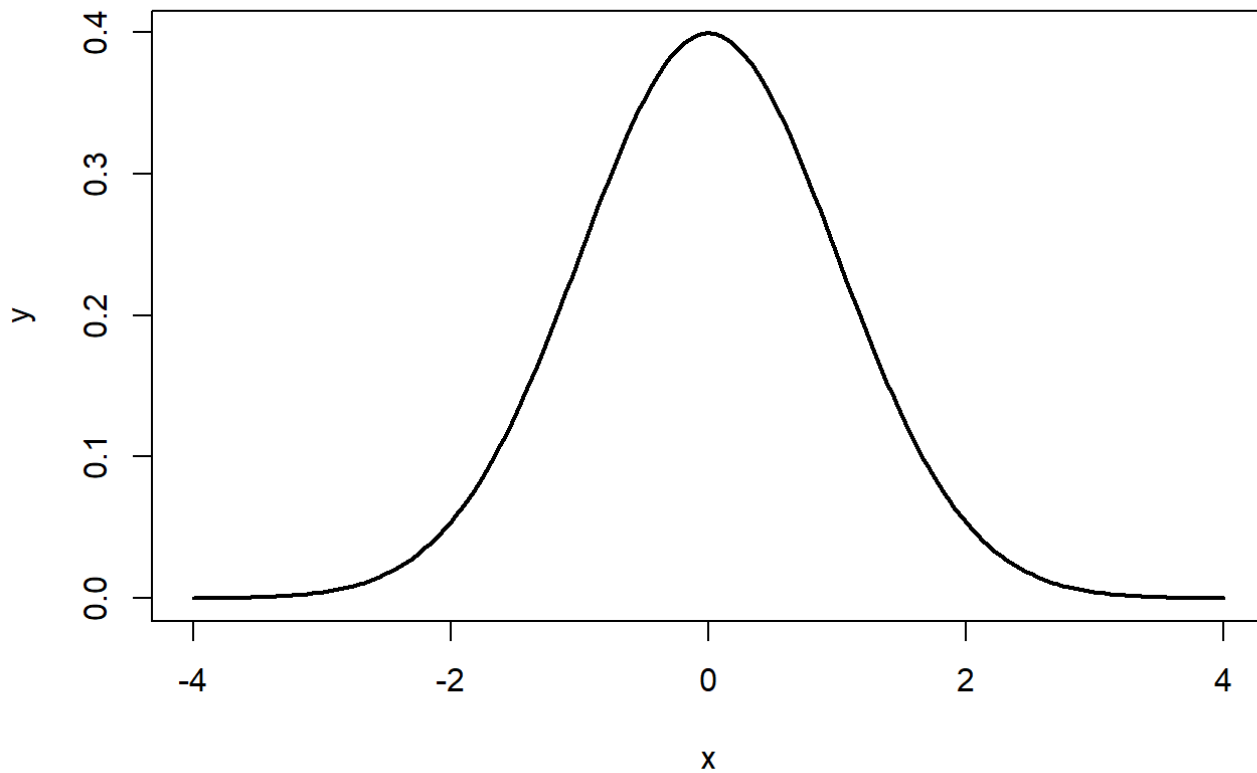


Figure 5.2.1: The Normal Distribution

Note that the the tails go to $\pm\infty\pm\infty$. In addition, the density of a distribution over the range of x is the key to hypothesis testing With a normal distribution, $\sim 68\%\sim 68\%$ of the observations will fall within 1 standard deviation of the mean, $\sim 95\%\sim 95\%$ will fall within 2 standard deviations, and $\sim 99.7\%\sim 99.7\%$ within 3 standard deviations. This is illustrated in Figures 5.2, 5.3, 5.4.

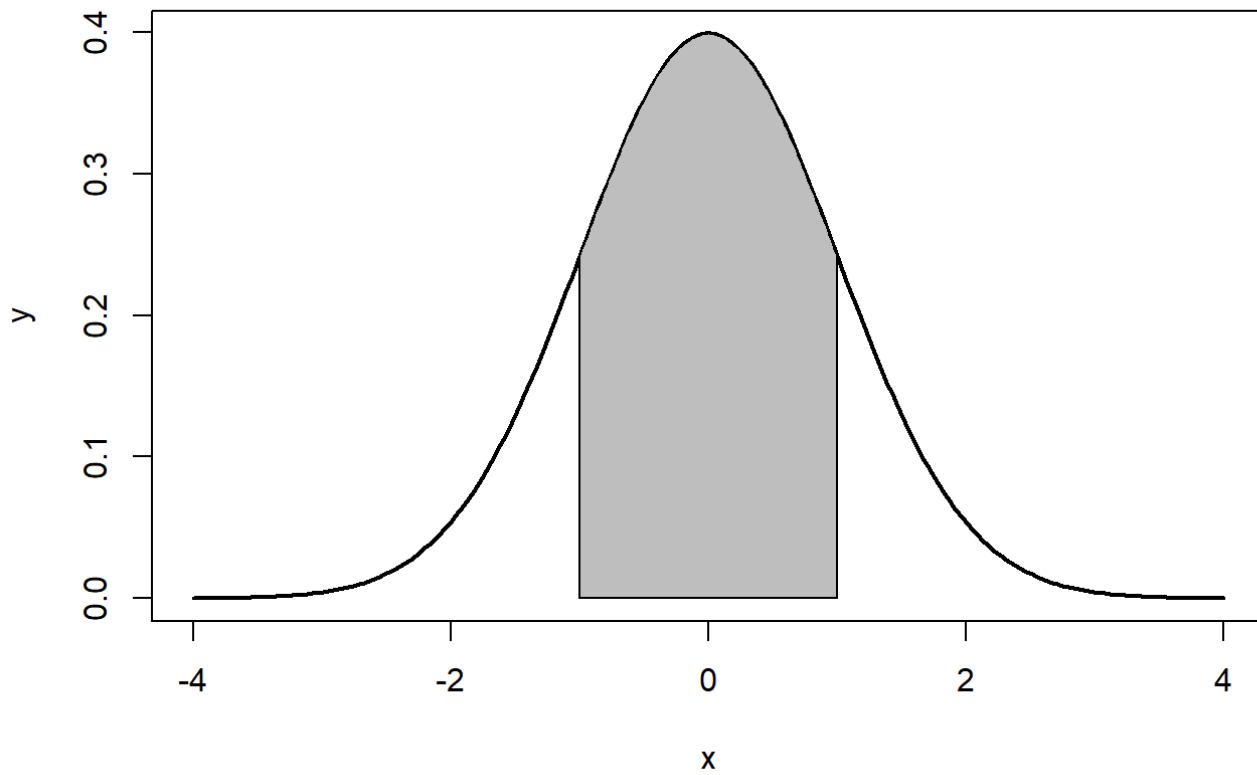


Figure 5.2.2: ~68%: 1 standard deviation

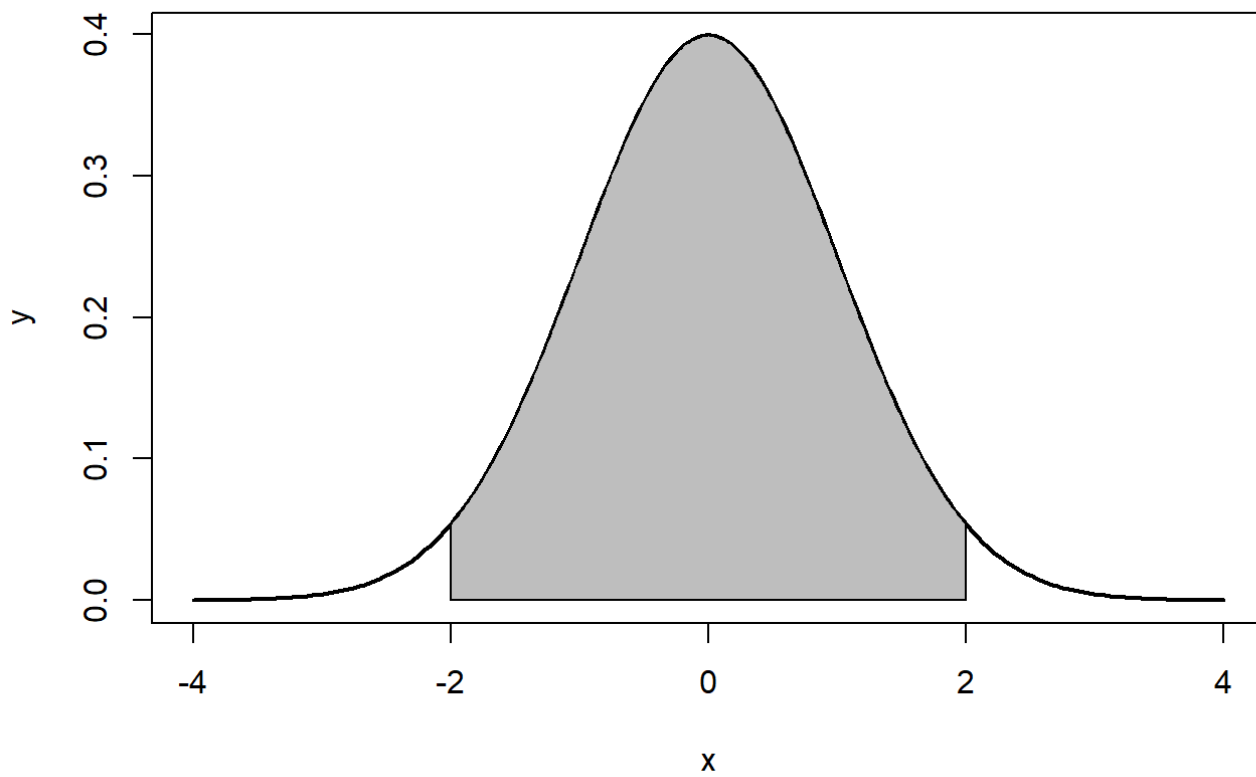


Figure 5.2.3: ~95%: 2 standard deviations

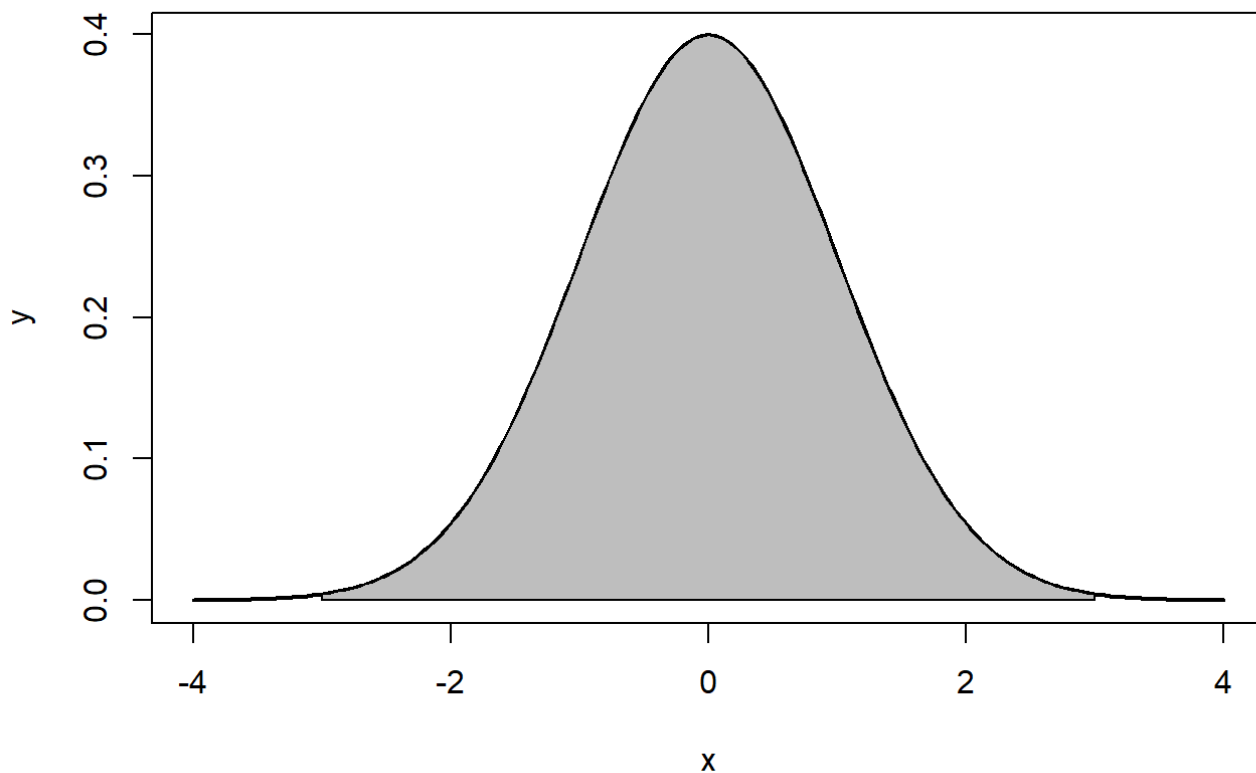


Figure 5.2.4: ~99.7%: 3 standard deviations

The normal distribution is characterized by several important properties. The distribution of observations is symmetrical around the mean μ ; the frequency of observations is highest (the mode) at μ , with more extreme values occurring with lower frequency (this can be seen in Figure ??); and only the mean and variance are needed to characterize data and test simple hypotheses.

The Properties of the Normal Distribution

- It is symmetrical around its mean and median, μ
- The highest probability (aka “the mode”) occurs at its mean value
- Extreme values occur in the tails
- It is fully described by its two parameters, μ and σ^2

If the values for μ and σ^2 are known, which might be the case with a population, then we can calculate a Z-score to compare differences in μ and σ^2 between two normal distributions or obtain the probability for a given value given μ and σ^2 . The Z-score is calculated:

$$Z = \frac{Y - \mu}{\sigma} \quad (5.2)$$

Therefore, if we have a normal distribution with a μ of 70 and a σ^2 of 9, we can calculate a probability for $i=75$. First we calculate the Z-score, then we determine the probability of that score based on the normal distribution.

```
z <- (75-70)/3
z
```

```
## [1] 1.666667
```

```
p <- pnorm(1.67)
p
```

```
## [1] 0.9525403
```

```
p <- 1-p
p
```

```
## [1] 0.04745968
```

As shown, a score of 7575 falls just outside two standard deviations ($>0.95>0.95$), and the probability of obtaining that score when $\mu=70$ and $\sigma^2=9$ is just under 5%.

5.2.1 Standardizing a Normal Distribution and Z-scores

A distribution can be plotted using the raw scores found in the original data. That plot will have a mean and standard deviation calculated from the original data. To utilize the normal curve to determine probability functions and for inferential statistics we will want to convert that data so that it is standardized. We standardize so that the distribution is consistent across all distributions. That standardization produces a set of scores that have a mean of zero and a standard deviation of one. A standardized or Z-score of 1.5 means, therefore, that the score is one and a half standard deviations about the mean. A Z-score of -2.0 means that the score is two standard deviations below the mean.

As formula (4.4) indicated, standardizing is a simple process. To move the mean from its original value to a mean of zero, all you have to do is subtract the mean from each score. To standardize the standard deviation to one all that is necessary is to divide each score the standard deviation.

5.2.2 The Central Limit Theorem

An important property of samples is associated with the **Central Limit Theorem** (CLT). Imagine for a moment that we have a very large (or even infinite) population, from which we can draw as many samples as we'd like. According to the CLT, as the n -size (number of observations) within a sample drawn from that population increases, the more the distribution of the means taken from samples of that size will resemble a normal distribution. This is illustrated in Figure 5.2.5. Also note that the population does not need to have a normal distribution for the CLT to apply. Finally, a distribution of means from a normal population will be approximately normal at any sample size.

This page titled [5.2: The Normal Distribution](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Jenkins-Smith et al. \(University of Oklahoma Libraries\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.