Course › Section... › 1.2 Intr... › Norma...

Normal Distribution: Standard Units and Z-scores

Standard units

For data that are approximately normal, *standard units* describe the number of standard deviations an observation is from the mean. Standard units are denoted by the variable z and are also known as *z-scores.*

For any value x from a normal distribution with mean μ and standard deviation σ, the value in standard units is:

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

Standard units are useful for many reasons. Note that the formula for the normal distribution is simplified by substituting z in the exponent:

$$\Pr\left(a < x < b\right) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2}\, dx$$

When $z = 0$, the normal distribution is at a maximum, the mean μ. The function is defined to be symmetric around $z = 0$.

The normal distribution of z-scores is called the *standard normal distribution* and is defined by $\mu = 0$ and $\sigma = 1$.

Z-scores are useful to quickly evaluate whether an observation is average or extreme. Z-scores near 0 are average. Z-scores above 2 or below -2 are significantly above or below the mean, and z-scores above 3 or below -3 are extremely rare.

We will learn more about benchmark z-score values and their corresponding probabilities below.

Code: Converting to standard units

The `scale` function converts a vector of approximately normally distributed values into z-scores.

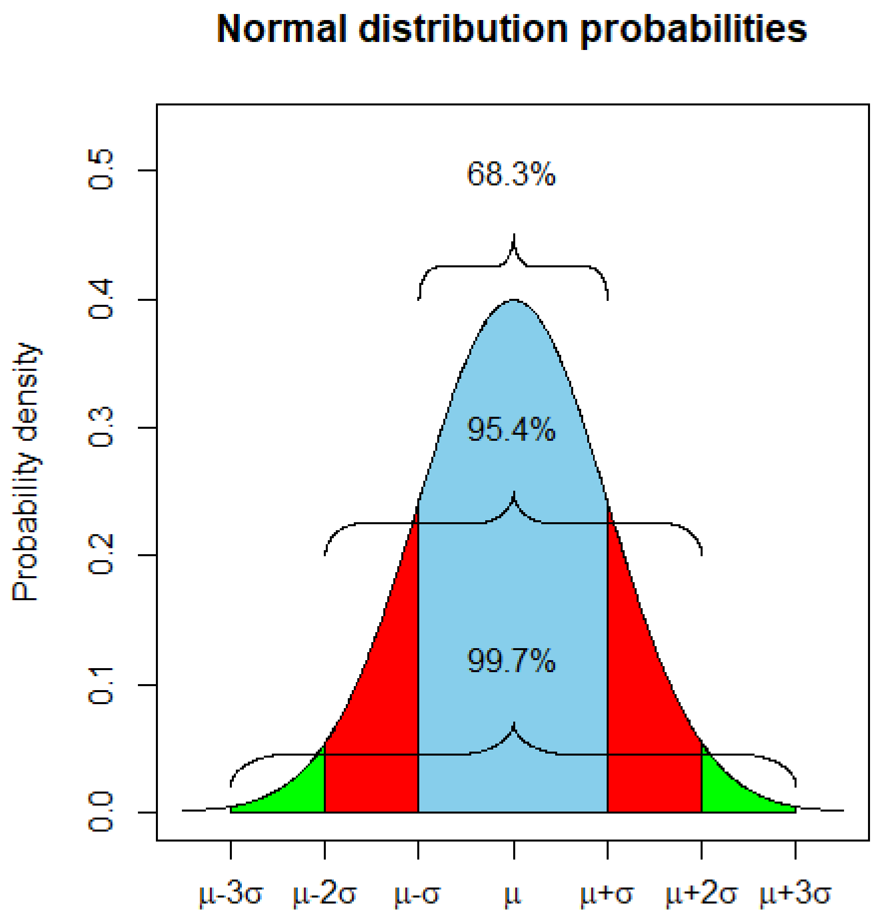
```
z <- scale(x)
```

You can compute the proportion of observations that are within 2 standard deviations of the mean like this:

```
mean(abs(z) < 2)
```

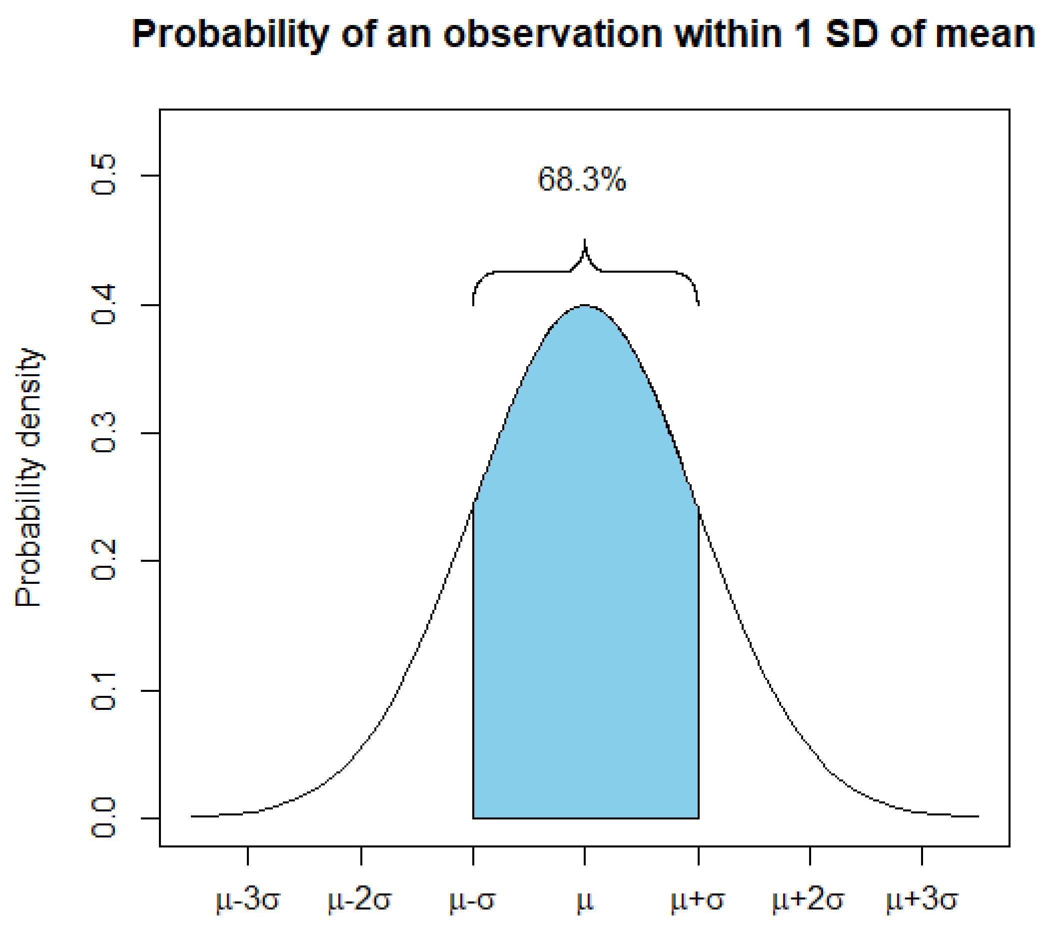
The 68-95-99.7 Rule

The normal distribution is associated with the 68-95-99.7 rule. This rule describes the probability of observing events within a certain number of standard deviations of the mean.



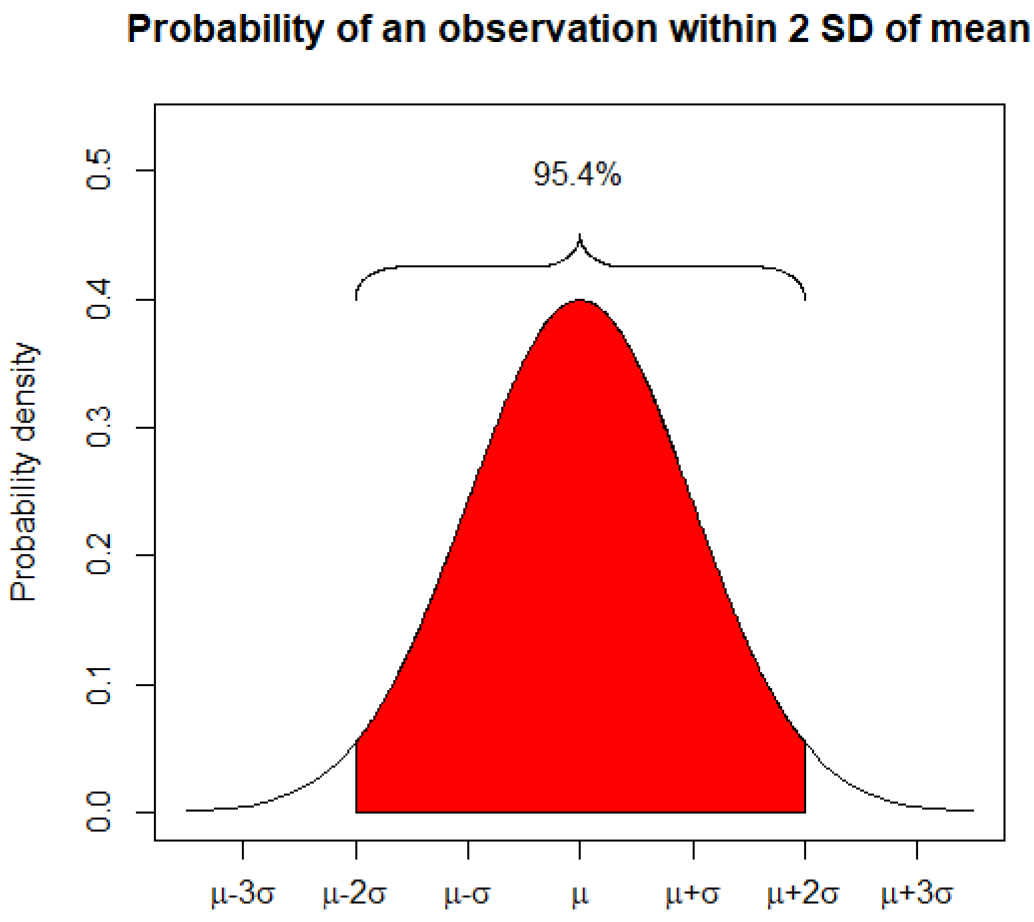
The probability distribution function for the normal distribution is defined such that:

- About 68% of observations will be within one standard deviation of the mean ($\mu \pm \sigma$). In standard units, this is equivalent to a z-score of $|z| \leq 1$.

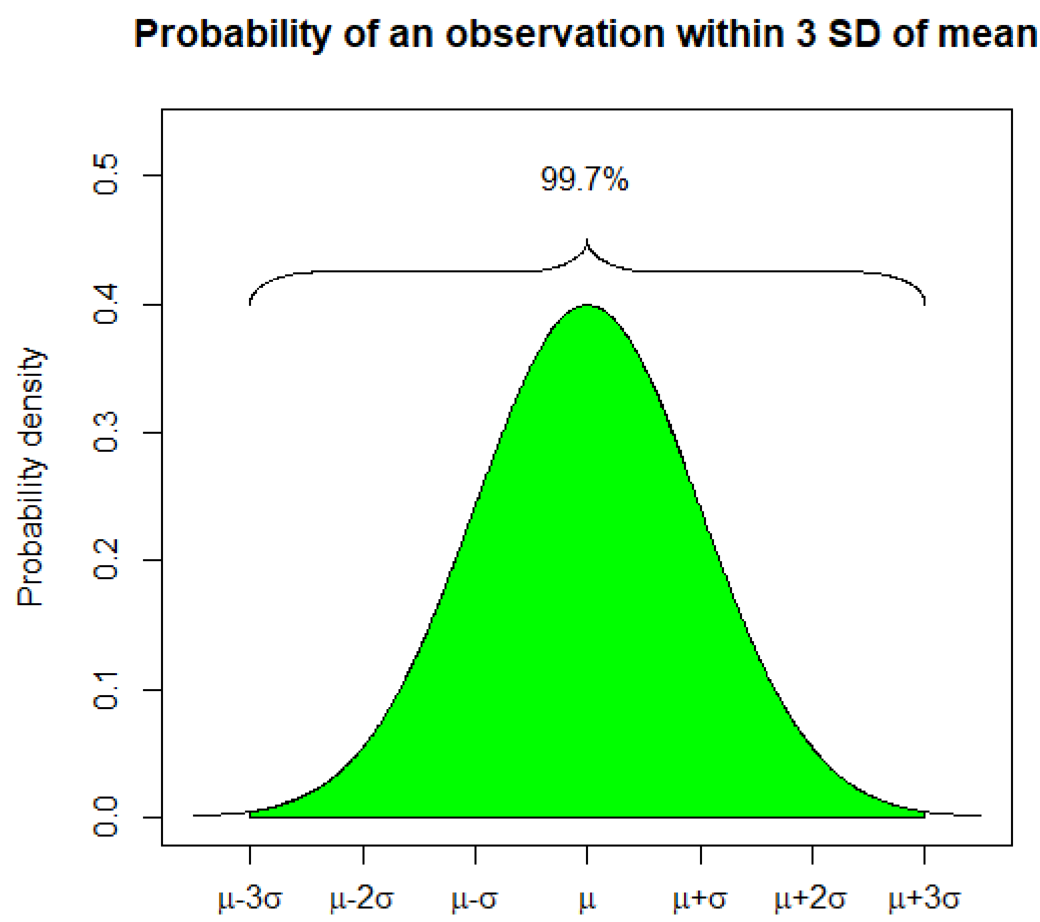


- About 95% of observations will be within one standard deviation of the mean ($\mu \pm 2\sigma$). In standard units, this is equivalent to a z-score of $|z| \leq 2$.

Sorry we could not search the store for annotations



- About 99.7% of observations will be within one standard deviation of the mean ($\mu \pm 3\sigma$). In standard units, this is equivalent to a z-score of $|z| \leq 3$.



- We will learn how to compute these exact probabilities in a later section, as well as probabilities for other intervals.

Learn About Verified Certificates