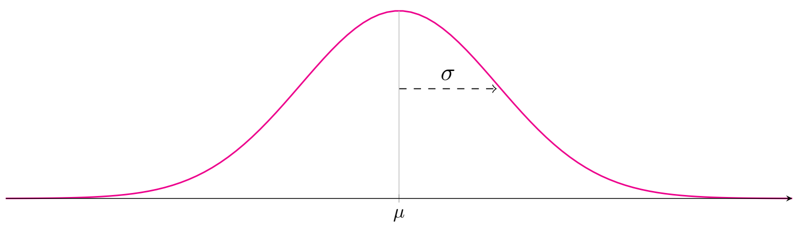
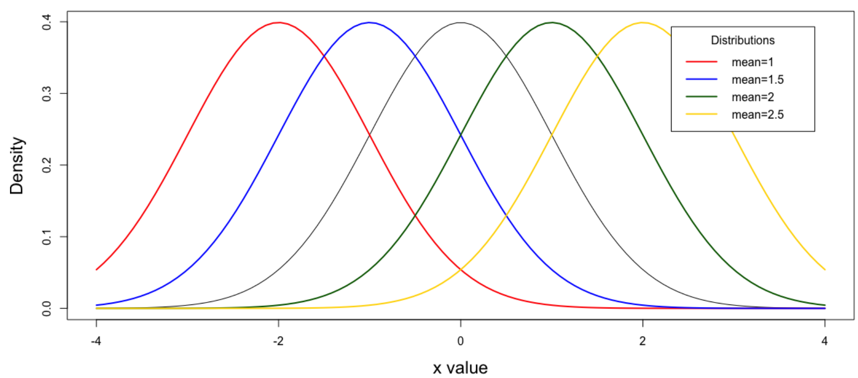
3.6.2 Normally Distributed Random Variables

Normally distributed random variables of the most widely used. The normal or Gaussian disttribution is a familiar bell curve. Let's consider a normally distributed random variable Y. This can be denoted as $Y \sim \text{Normal}(\mu, \sigma)$.

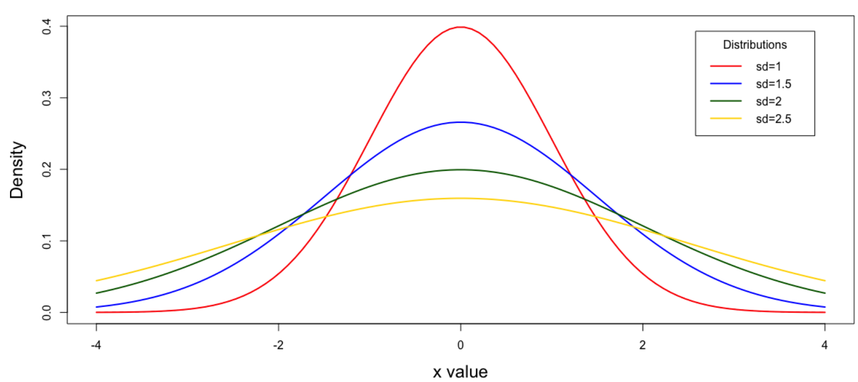
$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$



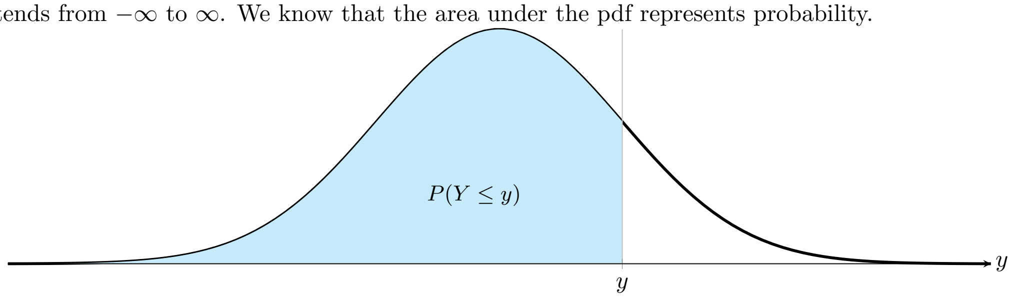
This is a relatively complex equation. We will not explore it in much deeper detail but we will notice that it has two parameters: μ and σ. μ is the mean or expected value for a normal random variable. It is characterized as the center of the distribution. Shown below are several normal distributions with different values of μ (holding σ constant).



σ is the standard deviation of a normal random variable. Shown below is a normal distribution with various values of σ (holding μ constant). Also recall that σ^2 is Variance.



So $E(Y) = \mu$, and $Var(Y) = \sigma^2$, but how do we calculate probability? The normal distribution extends from $-\infty$ to ∞. We know that the area under the pdf represents probability.



This can also be shown using calculus

$$P(Y \leq y) \quad = \quad \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

However, it turns out there is no closed form expression for this integral! The best way to find the probability for a normal distribution is using a computer or probability distribution tables. It would be silly to make tables for normal distributions with all values of μ and σ, so tables only describe the standard normal distribution.

3.6.3 The Standard Normal Distribution

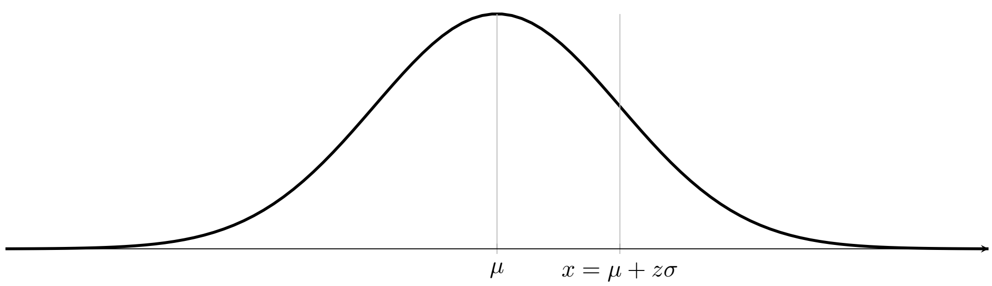
Standard normal random variables are simply normal random variables with $\mu = 0$ and $\sigma = 1$. Z is reserved to denote standard normal random variables. $Z \sim \text{Normal}(0, 1)$. It turns out we can easily turn any normal distribution into a standard normal distribution using a simple transformation. Suppose $X \sim \text{Normal}(\mu, \sigma)$.

$$Z = \frac{X - \mu_X}{\sigma_X}$$

But what are z-scores?

A z-score describes the the number of standard deviations an observation falls above the mean. For a normally distributed random variable X a value of $X = x$ can be described as $x = \mu + z\sigma$. With some simple algebra we can find the z-score for x.

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

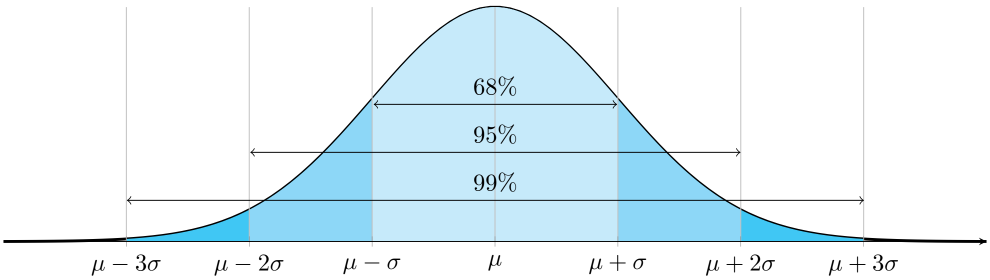


We like standardized scored because they make things easier to interpret then values in a distribution with arbitrary μ and σ. A z-score of 0.2 means the observation is relatively close to the mean, while a z-score of 4 is extraordinarily far from the mean.

3.6.4 Empirical Rule for the Normal Distribution

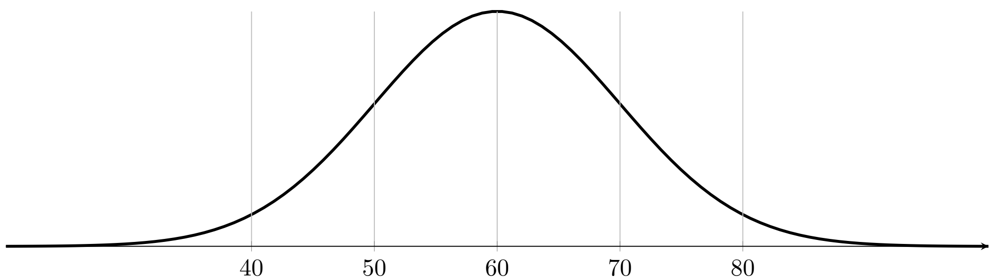
The empirical rule is also called the *68-95-99* rule. Suggesting for **normally distributed data**

- 68% of the data is captured within 1 standard deviation from the mean
- 95% of the data is captured within 2 standard deviations from the mean
- 99% or almost all the data is captured within 3 standard deviations of the mean



Example 4: Imperial officers must first pass the Empire's standardized entrance test to join work on the Death Star. The scores for the entrance examination are normally distributed with mean 60 and standard deviation 10.

(a) Sketch the distribution indicating the one, and two standard deviations from the mean.



(b) What's the probability an officer fails the entrance exam and is cast into the cold dark vacuum of space?

Notice that 50 is exactly 1 standard deviation below the mean of 60. Using the empirical rule we know about 34% of the distribution is between 50, and 60. We also know 50% of the distribution is above 60. So here there is a 1-0.84=0.16, or 16% chance the officer fails.

We could have also found the z-score of $\frac{50-60}{10} = -1$, and used a table or our calculator to find $P(Z < -1) = 0.1586553$.

Alternatively, we could have let X be the scores. $X \sim \text{Normal}(60, 10)$, so we can also find the probability without standardizing using calculators or a computer.

(c) What's the probability an officer passes the entrance exam and joins the empire?

In the last question we found that the probability of failure is 16%, using complimentary probability we are able to find the probability of passing, 1-0.16=0.84 or 84%.

(d) Whats the probability an officer scores between 75 and 80 on their entrance examination.

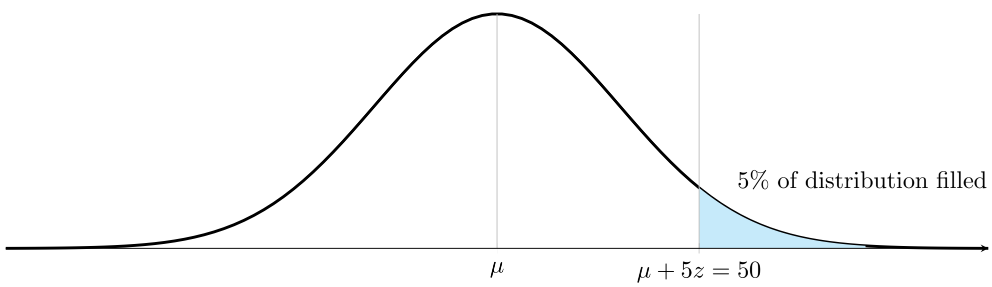
First off, as the distribution of the scores, X, is given we can find $P(75 \leq X \leq 80)$ easily using a calculator, or computer. However let's find the respective z-scores so we know how it works.

$$\begin{aligned} z_{75} &= \frac{75-60}{10} \\ &= 1.5 \\ z_{80} &= \frac{80-60}{10} \\ &= 2 \end{aligned}$$

So we can find the probability by finding $P(1.5 \leq Z \leq 2) = 0.04405707$

Example 5: An ice-cream machine can be programmed so that it serves an average of μ ml per cone. If the ml of ice cream are normally distributed with standard deviation 5 ml, give the setting for μ so that 50 ml of ice-cream or more will be served 5% of the time.

Let's start by drawing out the picture



Using the quantile function on our calculator (or a computer) we can determine that $P(Z > 1.644854) = 0.95$, so in our picture $z = 1.644854$.

$$\begin{aligned} \mu + 5z &= 50 \\ \mu &= 50 - 5z \\ &= 50 - 5(1.644854) \\ &= 41.77573 \end{aligned}$$

We would program the mean as 41.77573 ml.

3.6.5 Sums and Differences of Normal Random Variables

We are also introduced in the distribution that is created when we combine normal random variables. Something quite lovely happens here. Suppose we have X_1, X_2, ..., X_n and $X_i \sim \text{Normal}(\mu_i, \sigma_i)$. Then consider the distribution of

$$U = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

We can easily find the distribution for U

$$\begin{aligned} U &\sim \text{Normal}\left(a_1\mu_1 + a_2\mu_2 + \cdots a_n\mu_n\ ,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2\right) \\ &\sim \text{Normal}\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right) \end{aligned}$$

Example 6: Let $X \sim \text{Normal}(5, 3)$, and $Y \sim \text{Normal}(2, 2)$. What is the distribution of $2X + 3Y$?

$2X + 3Y \sim \text{Normal}\,(2(5) + 3(2), 4(3) + 9(2))$, so $2X + 3Y \sim \text{Normal}\,(16, 32)$.