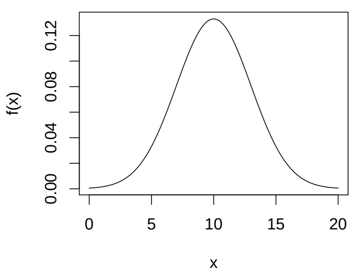
# Lecture 4: Normal Distribution
## STAT 630, Fall 2021

- The normal distribution is one of the most common and important probability distributions.
- It is symmetric, unimodal, and bell-curve shaped.
- Many phenomena in nature follow a normal distribution such as the height of individuals, the velocity in any direction of a molecule of gas, and measurement error.
- The normal distribution is characterized by two parameters: the mean $\mu$ (center of distribution) and standard deviation $\sigma$ (spread of distribution).
- The notation $X \sim N(\mu, \sigma)$ means that the random variable $X$ follows a normal distribution with mean $\mu$ and standard deviation $\sigma$.
- For example, the plot below shows the distribution $N(\mu = 10, \sigma = 3)$.




- The probability density function (pdf) for the normal distribution is given by

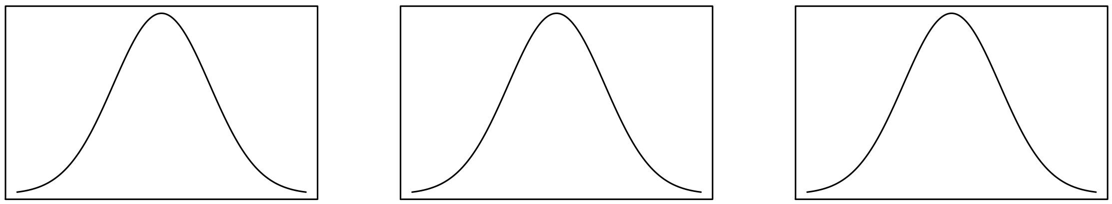
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

- Additional properties:
  - The area under the normal distribution curve is 1.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

  - The normal distribution is symmetric about the mean, $\mu$.

- Probabilities are computed as the area under the normal distribution curve.



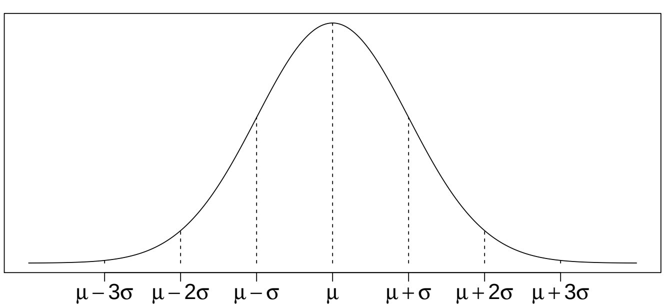
- Note that for a normally distributed random variable $P(X \geq a) = P(X > a)$. Why?

- There is no closed form solution to

$$P(X > a) = \int_a^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx,$$

so numerical approximations of the area under the curve are used to compute probabilities. In practice, we can use the R function `pnorm()` or a standard normal table to compute probabilities.

- Empirical Rule:

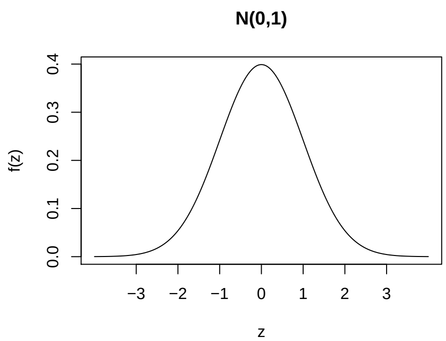



  - About 68% of the distribution is contained within 1 standard deviation of the mean.

  - About 95% of the distribution is contained within 2 standard deviations of the mean.

  - About 99.7% of the distribution is contained within 3 standard deviations of the mean.

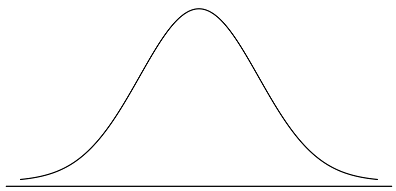
- Let $X \sim N(\mu, \sigma)$. A $Z$-score is defined as $Z = (X - \mu)/\sigma$. It can be shown that $Z \sim N(0, 1)$.
- A $z$-score can be interpreted as the number of standard deviations an observation $x$ lies away from the mean. For instance, if a student has a z-score of 2 on an exam then that student is 2 standard deviations above the average score.
- The distribution $N(0, 1)$ is called the standard normal distribution or $Z$-distribution.
- For $Z \sim N(0, 1)$ the empirical rule gives that $P(-1 < Z < 1) \approx 0.68$, $P(-2 < Z < 2) \approx 0.95$, and $P(-3 < Z < 3) \approx 0.997$.
- Computing $Z$-scores allows us to compute probabilities for any normal distribution.



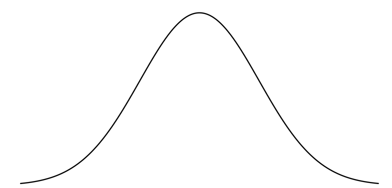

**Theorem**. Let $X \sim N(\mu, \sigma)$ and $Z = (X - \mu)/\sigma$. Show that $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

**Ex1**. The amount $X$ of the pollutant nitrogen oxide in automobiles is normally distributed with mean $\mu = 70$ ppb (parts per billion) and standard deviation $\sigma = 13$ ppb. We can write this compactly as $X \sim N(70, 13)$.

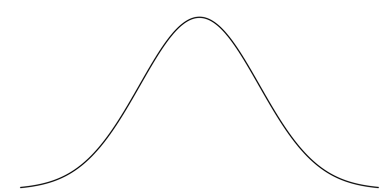
(a) What is the probability that a randomly selected vehicle will have an emission level less than 60 ppb?



(b) What is the probability that a randomly selected vehicle will have an emission level greater than 90 ppb?

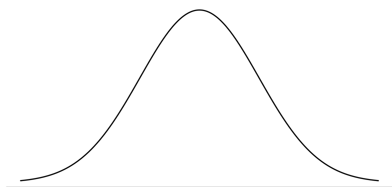


(c) What is the probability that a randomly selected vehicle will have an emission level between 60 and 90 ppb?



**Ex2**. Body temperatures are normally distributed with mean $\mu = 98.2$ and standard deviation $\sigma = 0.74$, in degrees Fahrenheit. That is, $X \sim N(98.2, 0.74)$.

(a) Find the cutoff for the lowest 5% of body temperatures (the $5^{th}$ percentile)?



(b) Find the cutoff for the highest 15% of body temperatures (the $85^{th}$ percentile)?