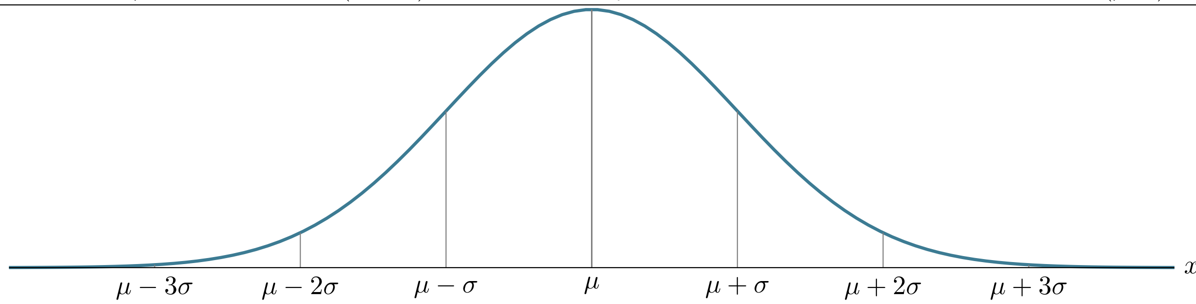
STATISTICS DISTRIBUTIONS

NORMAL DISTRIBUTION

A **Normal Distribution** is described by a symmetric bell-shaped density curve (as below). It is centered at the mean μ and the spread (width) is determined by the standard deviation σ. Written as $N(\mu, \sigma)$.



THE 68-95-99.7 RULE

With a Normal Distribution with mean μ and standard deviation σ:

- Approximately **68%** of the data is within σ of μ. Between $(\mu - \sigma, \mu + \sigma)$
- Approximately **95%** of the data is within 2σ of μ. Between $(\mu - 2\sigma, \mu + 2\sigma)$
- Approximately **99.7%** of the data is within 3σ of μ. Between $(\mu - 3\sigma, \mu + 3\sigma)$

STANDARDIZING AND Z-SCORE

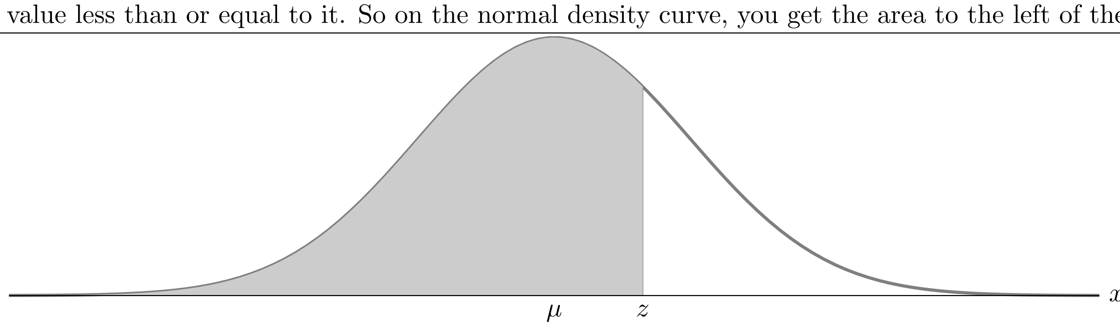
The **standard Normal distribution** $N(0, 1)$ has mean 0 and standard deviation 1. Given a Normal distribution $N(\mu, \sigma)$ we can standardize an observation x by calculating

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

The standardized value is called the **z-score**. The z-score tells us how many standard deviations the x value is away from the mean μ.

STANDARD NORMAL TABLE

Once you calculate the z-score, you can use the standard Normal table to find the probability that you get a value less than or equal to it. So on the normal density curve, you get the area to the left of the z value.



The area to the left of z is the probability any observation is less than or equal to it, which is $\boldsymbol{P(x \leq z)}$. If you want the probability that an observation is greater than or equal to z, $\boldsymbol{P(x \geq z)}$, take 1 minus the table value which is $\boldsymbol{1 - P(x \leq z)}$.

STATISTICS

DISTRIBUTIONS

BINOMIAL DISTRIBUTION

A distribution X is a **binomial distribution** when we want to count the number of successes of n trials and the probability of success is p for **each** trial. You can only use the binomial distribution if each trial is independent, that knowing the result of one trial doesn't change the probability of the next trial.

BINOMIAL PROBABILITY

If x has a binomial distribution with n trials and probability p of success. The possible values of X are $0, 1, 2, \ldots, n$. If k is one of those values, the probability we get exactly k successes is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{where} \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The **mean**, μ, and the **standard deviation**, σ, of X is

$$\mu = np \qquad \text{and} \qquad \sigma = \sqrt{np(1-p)}$$

NORMAL APPROXIMATION TO BINOMIAL DISTRIBUTIONS

Given a binomial distribution X with n trials and probability p of success, if n is large, the distribution is approximately Normal with μ and σ as above. So we get $\boldsymbol{N\left(np, \sqrt{np(1-p)}\right)}$
In general, you can use a Normal approximation when n is large that $np \geq 10$ and $n(1-p) \geq 10$.

SAMPLING DISTRIBUTION ON $\bar{X}$

The **population mean** is represented by μ whereas when we take a SRS, the **sample mean** is represented by $\boldsymbol{\bar{x}}$ is used to estimate μ. The **sampling distribution** of $\boldsymbol{\bar{x}}$ describes the possible values that $\bar{x}$ can take from samples of the same size from the same population. The **mean** of $\bar{x}$ is $\boldsymbol{\mu}$ and the **standard deviation** is $\boldsymbol{\sigma/\sqrt{n}}$.

SAMPLING DISTRIBUTION ON A NORMAL POPULATION

If the individual observations have a Normal distribution $N(\mu, \sigma)$ then the sample mean $\bar{x}$ of an SRS of size n is also Normal with $\boldsymbol{N(\mu, \sigma/\sqrt{n})}$. This is because the more samples we take (the bigger n is) the more likely $\bar{x}$ is closer to μ, which makes the standard deviation smaller.

CENTRAL LIMIT THEOREM

Given **any** distribution with mean μ and standard deviation σ, when n is large, the sampling distribution of $\bar{x}$ becomes approximately Normal with $\boldsymbol{N(\mu, \sigma/\sqrt{n})}$.

SAMPLING DISTRIBUTION ON $\widehat{P}$

The actual **population proportion** is **p** whereas when we take a SRS, the **sample proportion** $\boldsymbol{\widehat{p}}$ estimates p. The **mean** of $\widehat{p}$ is $\boldsymbol{p}$ and the **standard deviation** is $\boldsymbol{\sqrt{p(1-p)/n}}$. Also, when n is large, np and $n(1-p)$ are large and when the population is at least 20 times larger than the sample, the sampling distribution becomes approximately Normal with $\boldsymbol{N(p, \sqrt{p(1-p)/n})}$