Probability Distributions

Bernoulli Trial

In many experiments there are only two outcomes. We call such an experiment a Bernoulli trial, and refer to the two outcomes—often arbitrarily—as *success* and *failure*.

$$X = \begin{cases} 0 & \text{with } P(0) = 1 - \pi \\ 1 & \quad\quad\; P(1) = \pi \end{cases}$$

Mean and variance

$$\mathrm{E}(X) = \pi$$
$$\mathrm{Var}(X) = \pi(1 - \pi)$$

Examples of Bernoulli Trials

- Outcome for a single coin flip
- Outcome of a penalty shot in hockey
- Determine if a school meets AYP
- Determine if a birth is a boy or girl

In 2015, 60.2% of all Minnesota students that were tested under NCLB met the proficiency standard.

```r
# Sample a single student from MN
> sample(
   x = c("proficient", "not proficient"),
   size = 1,
   replace = TRUE,
   prob = c(0.602, 0.398)
   )

[1] "proficient"


# Use {0,1} rather than {failure, success}
> sample(
   x = c(0, 1),
   size = 1,
   replace = TRUE,
   prob = c(0.398, 0.602)
   )

[1] 0
```

In 2015, 60.2% of all Minnesota students that were tested under NCLB met the proficiency standard.

```
# Sample 10 students from MN
> sample(
   x = c(0, 1),
   size = 10,
   replace = TRUE,
   prob = c(0.398, 0.602)
   )

[1] 0 1 1 1 0 1 1 1 1 1
```

Binomial Distribution

Binomial is the sum of n independent Bernoulli trials. Each Bernoulli trial is

$$x = \begin{cases} 0 & \text{with } P(0) = 1 - \pi \\ 1 & P(1) = \pi \end{cases}$$

Mean and variance

$$E(X) = n\pi$$
$$\text{Var}(X) = n\pi(1 - \pi)$$

When a sequence of independent Bernoulli trials are carried out, and the random variable X contains the **count of the number of successes**, X is referred to as a **binomial random variable**.

Examples of what can be modeled using the Binomial distribution

- Flip a coin 100 times; X = number of heads
- Many penalty shots in hockey are taken; X = number of goals scored
- Of the 955 public schools in MN; X = the number that meet AYP
- Of 100,000 births; X = number of girls

In 2015, 60.2% of all Minnesota students that were tested under NCLB met the proficiency standard. Choose 100 random students. How many are proficient?

```
# Sample 100 students from MN
> X = sample(
   x = c(0, 1),
   size = 10,
   replace = TRUE,
   prob = c(0.398, 0.602)
   )


# Sum the results
> sum(X)
[1] 4
```

We can also use the `rbinom()` function. This function randomly (*r*) generates a binomial (*binom*) distribution with given parameters.

In 2015, 60.2% of all Minnesota students that were tested under NCLB met the proficiency standard. Choose 100 random students. How many are proficient?

```
# Sample 100 students from MN (one time) with probability of success = 0.602
> rbinom(1, size = 10, prob = 0.602)
[1] 63
```

We can also use the `rbinom()` function to carry out a Bernoulli trial.

```
# Sample a single student from MN (ten times) with probability of success = 0.602
> rbinom(10, size = 1, prob = 0.602)

[1] 1 0 1 0 0 1 1 0 0 0
```

Now we will use the `rbinom()` function to randomly generate 100 students (with 60.2% success), and count the number that are proficient, many times.

```
# Sample 100 students from MN (25 times) with probability of success = 0.602
> rbinom(25, size = 100, prob = 0.602)

[1] 53 65 64 57 57 63 59 57 66 58 61 58 66 65 58 52 69 65 67 70 62 52 56 62 68
```

Now we will use the `rbinom()` function to randomly generate 100 students (with 60.2% success), and count the number that are proficient, many, many, many,…, many times.

```
# Sample 100 students from MN (1,000,000 times) with probability of
# success = 0.602
> y = rbinom(1000000, size = 100, prob = 0.602)

> mean(y)
[1] 60.20407

> var(y)
[1] 24.01326
```

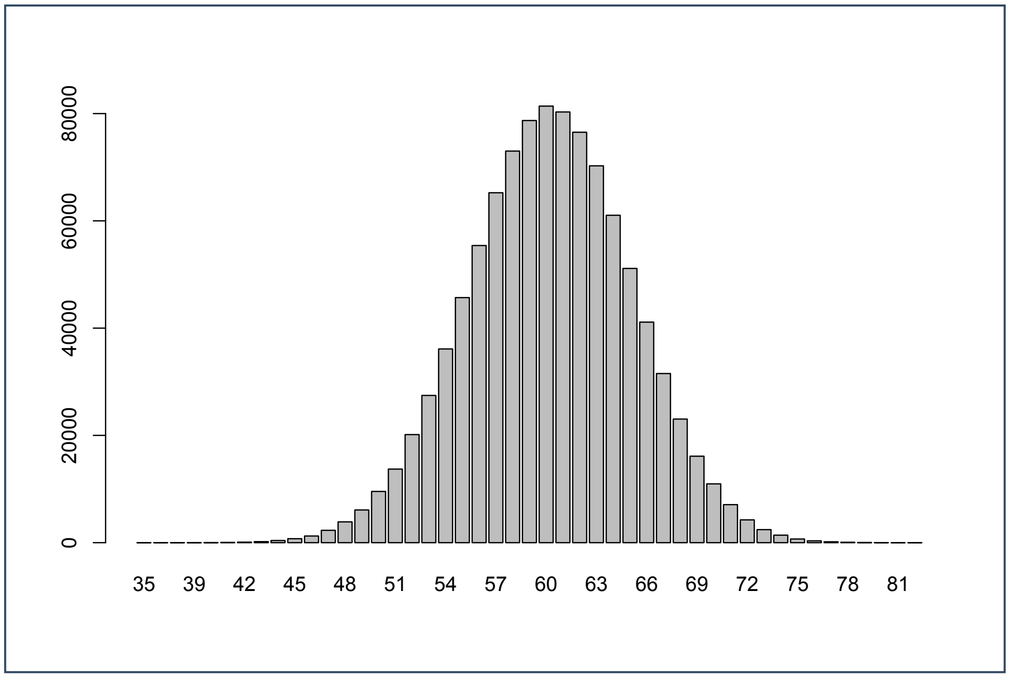
Theoretical mean and variance

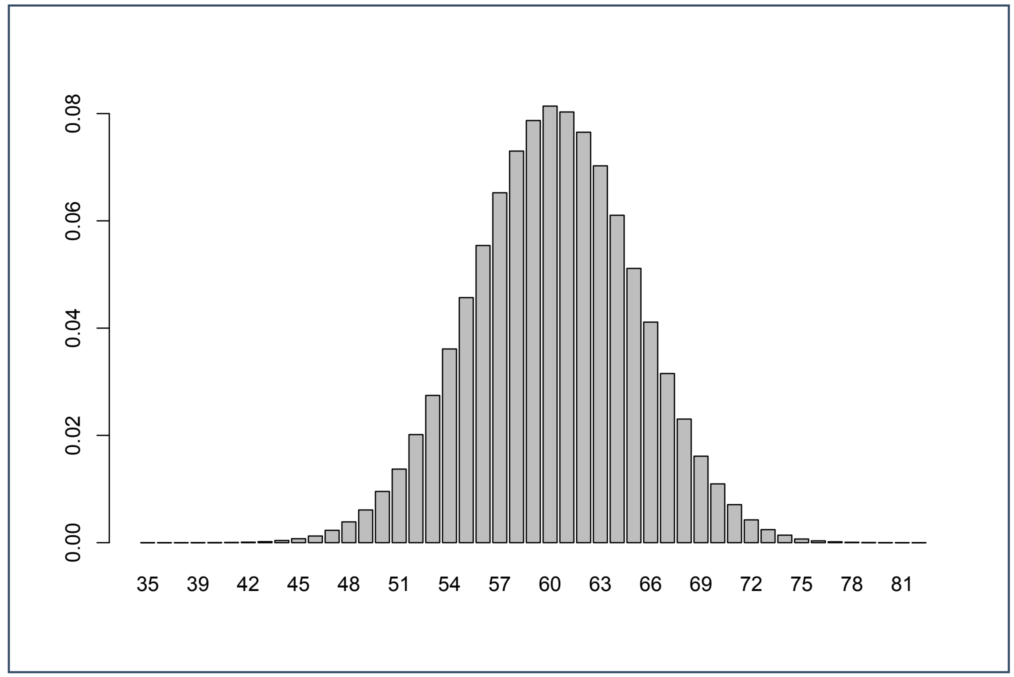
$$
\begin{aligned}
\mathrm{E}(X) &= n\pi \\
&= 100(.602) \\
&= 60.2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(X) &= n\pi(1-\pi) \\
&= 100(.602)(.398) \\
&= 23.9596
\end{aligned}
$$

The empirical (simulated) mean and variance computed from the random number generation are very close to the theoretical values. In practice, the difference between the theoretical and simulated values are generally negligible.

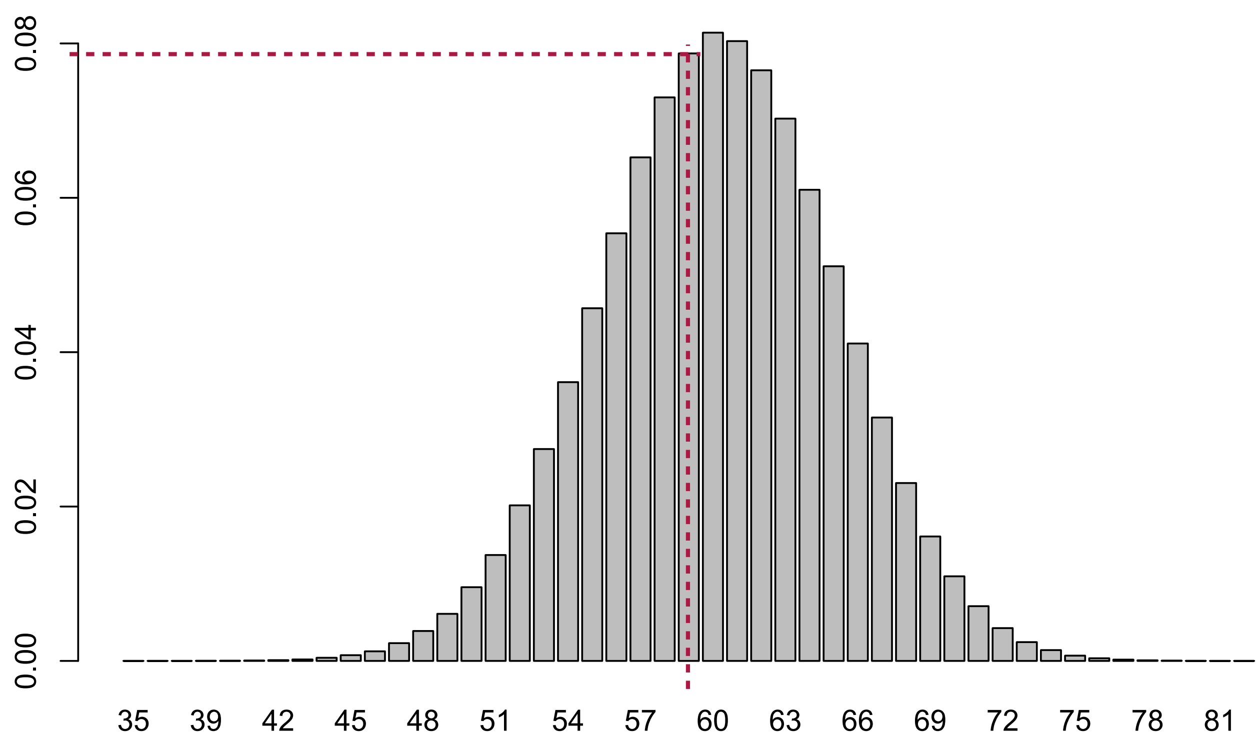
```
# Plot the frequency distribution (1) obtain counts and (2) plot
> counts = table(y)
> barplot(counts)

# Plot the (relative) probability distribution (1) obtain proportions from counts
# and (2) plot
> prop = counts / sum(counts)
> barplot(prop)
```





This is the empirical probability distribution. It is binomially distributed (not normally distributed) with p = .602.

This is the empirical probability distribution. It is binomially distributed (not normally distributed) with probability for success = .602.



$p(x = 59) \sim 0.079$

To theoretically calculate the probability of x, from n independent Bernoulli trials we compute:

$$p(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

The theoretical calculation of $p(x = 59)$ is

$$p(59) = \binom{100}{59} 0.602^{59} (1 - 0.602)^{100-59}$$

$$p(59) = \binom{100}{59} 0.602^{59} (0.398)^{41}$$

```
> choose(100, 59) * 0.602^59 * 0.398^41

[1] 0.07851454
```

Poisson Distribution

The Poisson distribution is defined as:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

for $x = 0, 1, 2, 3, \ldots$ and $\lambda > 0$

Mean and variance

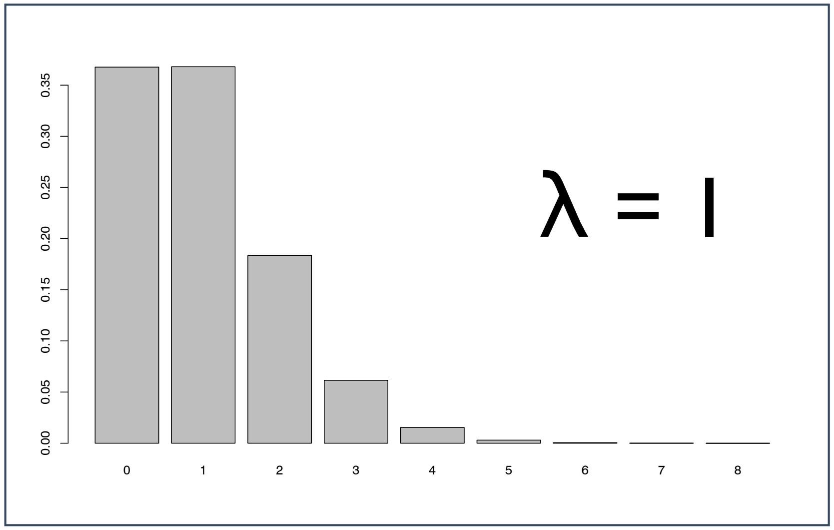
$$\mathrm{E}(X) = \lambda$$
$$\mathrm{Var}(X) = \lambda$$

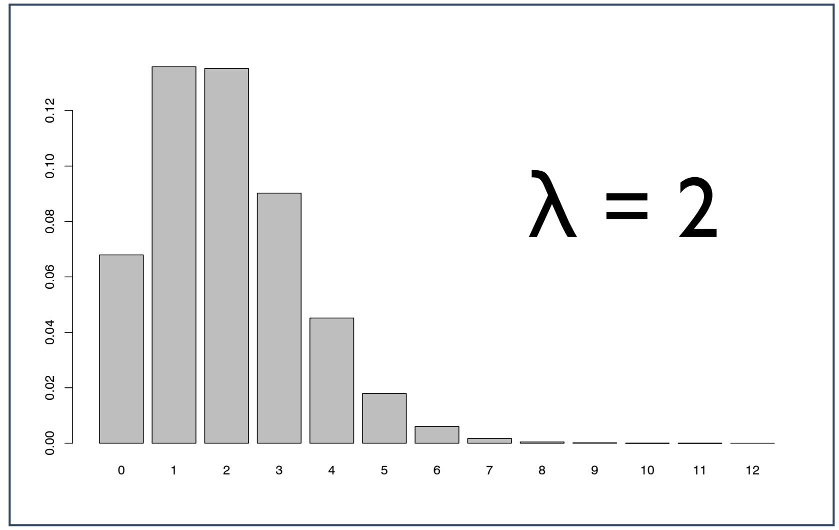
Originally introduced as an approximation of the binomial distribution. We often use this distribution when modeling counts of things in a fixed time period.

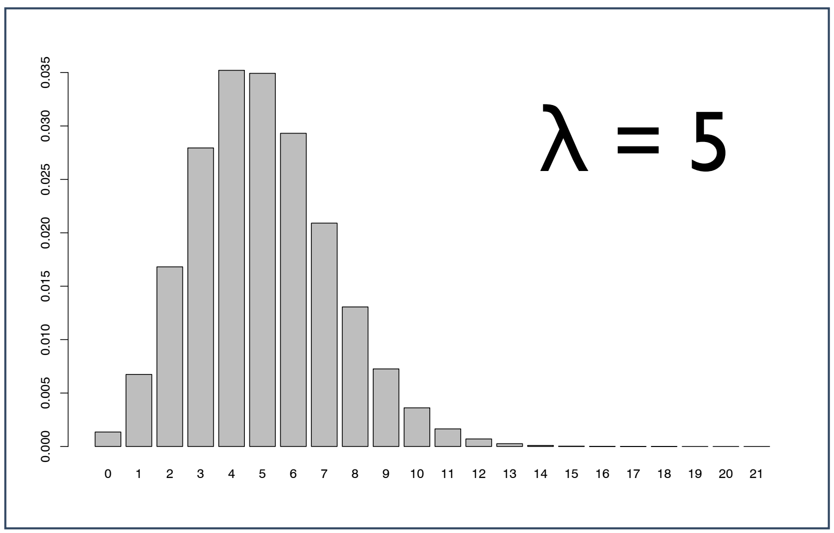
Examples of what can be modeled using the Poisson distribution

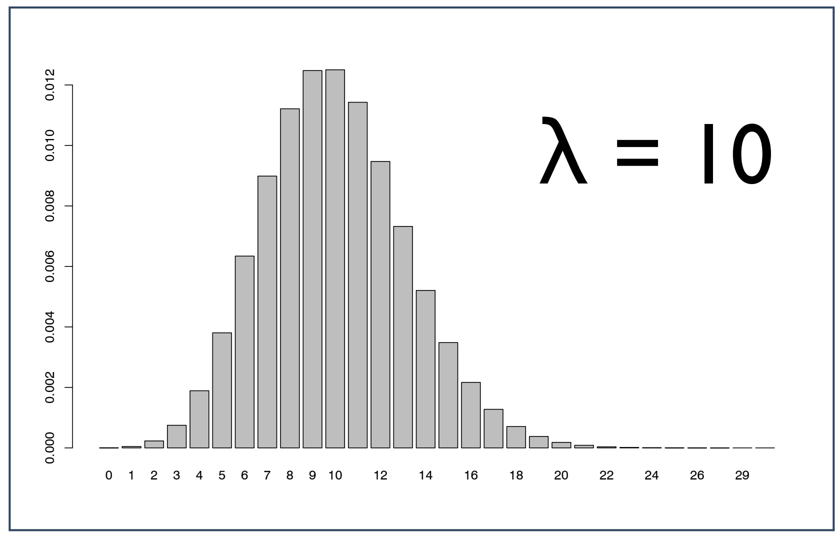
- Number of deaths by horse kicking in the Prussian army (the first application)
- Number of car accidents over a period of time
- Failure rates of electronics
- Number of aggressive incidents during a recess period

The parameter λ is called the rate parameter.









Since the mean is equal to lambda, the rate parameter is the average number of events in the time period.

Twenty students are observed on a playground, and the number of aggressive incidents over one-hour are recorded for each student. The frequency of the number of incidents is shown at right.

What is the probability of selecting a student at random (from the population) who had 3 or more aggressive incidents?

Number of Incidents	Frequency
0	4
1	3
2	5
3	2
4	4
5	1
6	1

```
# Compute the mean number of incidents
> (0*4 + 1*3 + 2*5 + 3*2 + 4*4 + 5*1 + 6*1) / 20
[1] 2.3
```

Now, we can simulate many, many, many randomly generated students using the Poisson distribution with a rate parameter of 2.3

We use the `rpois()` function. This function randomly (*r*) generates a Poisson (*pois*) distribution with a given lambda parameter.

```
# Simulate 1,000,000 observations from Poisson distribution with lambda = 2.3
> y = rpois(1000000, lambda = 2.3)

> mean(y)
[1] 2.301871

> var(y)
[1] 2.309067

# Compute the frequency of incidents
> table(y)
y
     0      1      2      3      4      5      6      7      8      9     10     11     12     13     14
100472 230025 265372 203183 116636  53899  20976   6753   1970    549    138     21      4      1      1

# Compute the frequency of incidents >= 3
> table(y >= 3)
 FALSE   TRUE
595869 404131

# Compute the relative frequency of incidents >= 3
> table(y >= 3) / 1000000
   FALSE     TRUE
0.595869 0.404131
```

The simulation results give an empirical probability of 0.404 of selecting a student at random (from the population) who had 3 or more aggressive incidents.

To solve this using probability theory,

$$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

where

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

```
# P(X = 0)
> p0 = (2.3^0 * exp(-2.3)) / factorial(0)

# P(X = 1)
> p1 = (2.3^1 * exp(-2.3)) / factorial(1)

# P(X = 2)
> p2 = (2.3^2 * exp(-2.3)) / factorial(2)


# Compute probability of X >= 3
> 1 - (p0 + p1 + p2)
[1] 0.4039612
```

Come up with 2–3 different examples (in your field) where we would use the **multinomial distribution** to model events.

Come up with 2–3 different examples (in your field) where we would use the **negative binomial distribution** to model events.

Normal Distribution

The normal distribution is defined as:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

for $-\infty < x < \infty$

Mean and variance

$$\mathrm{E}(X) = \mu$$
$$\mathrm{Var}(X) = \sigma^2$$

Common Notation: $X \sim N(\mu, \sigma^2)$

The normal distribution was also introduced as an approximation to the binomial distribution.

Standard (Unit) Normal Distribution

Mean and variance

$$E(X) = \mu = 0$$
$$\mathrm{Var}(X) = \sigma^2 = 1$$

The standard normal distribution is defined as:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

for $-\infty < x < \infty$

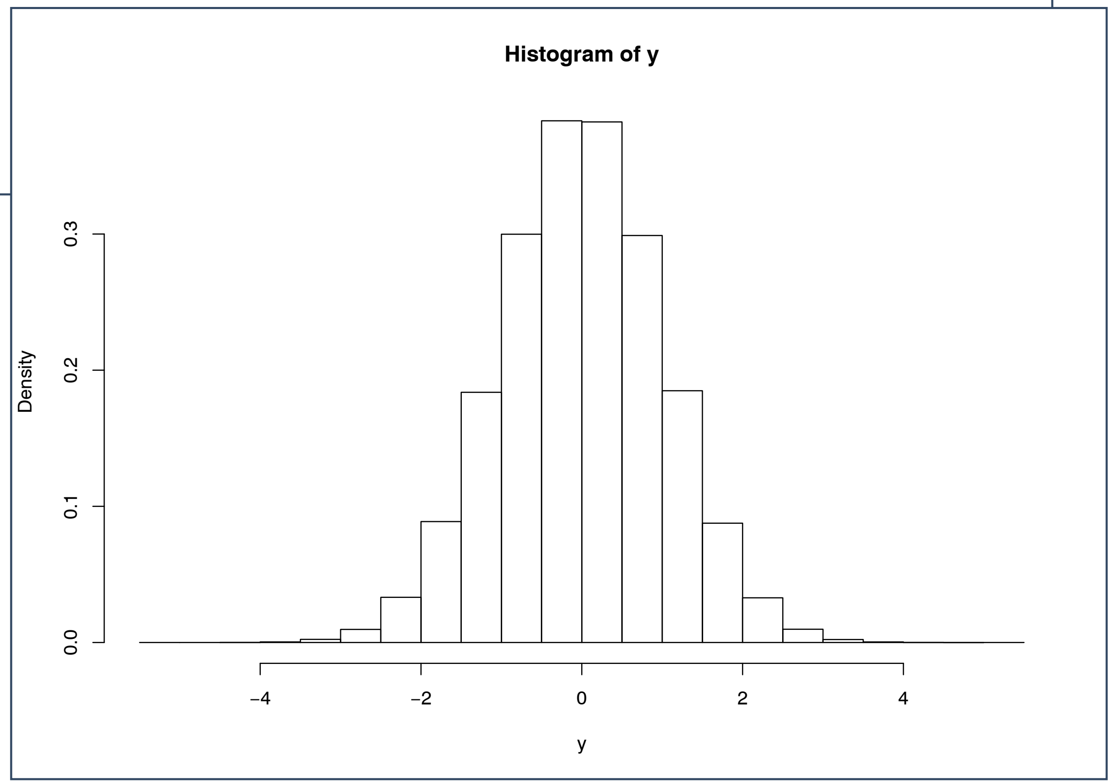
but now $\mu = 1$ and $\sigma^2 = 1$

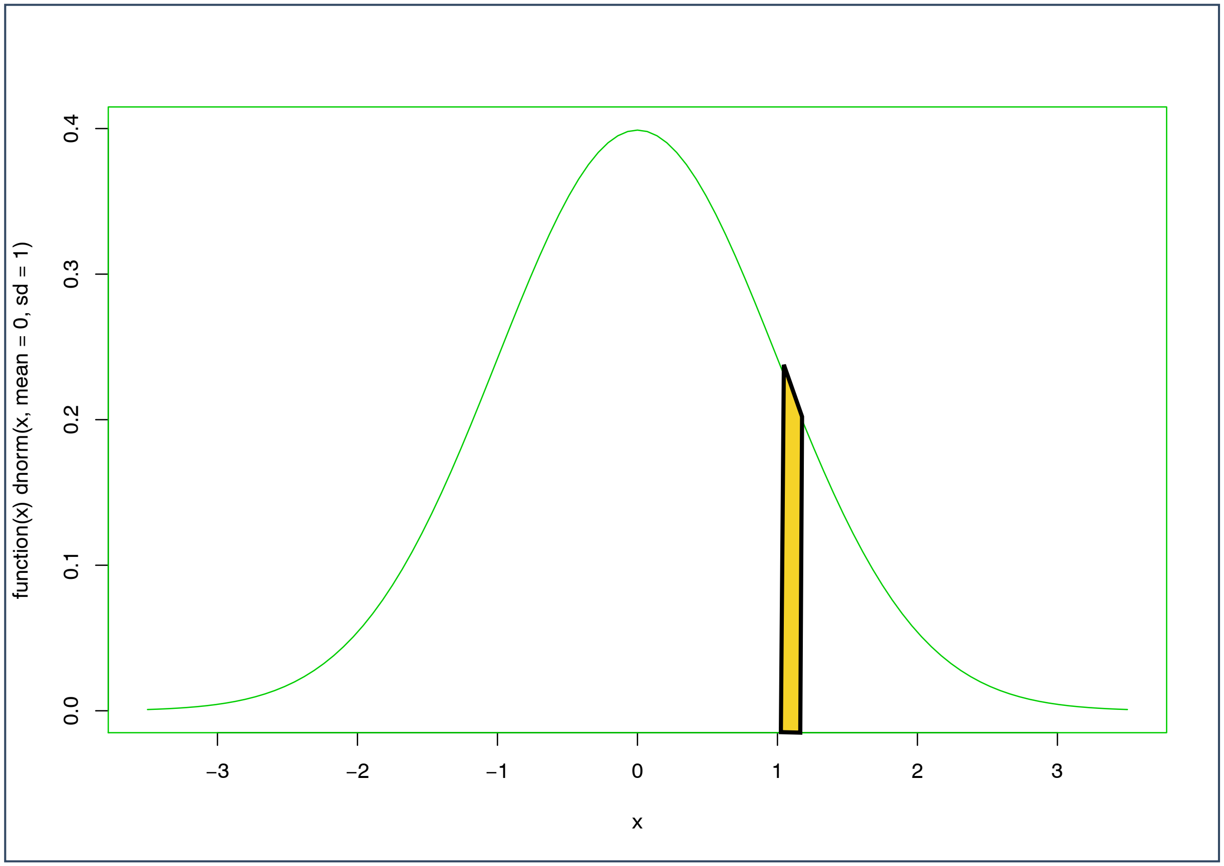
$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(z)^2}{2}\right]$$

Common Notation: $z \sim N(0, 1)$

We use the `rnorm()` function to randomly (*r*) generate a normal (*norm*) distribution with a given mean and standard deviation.

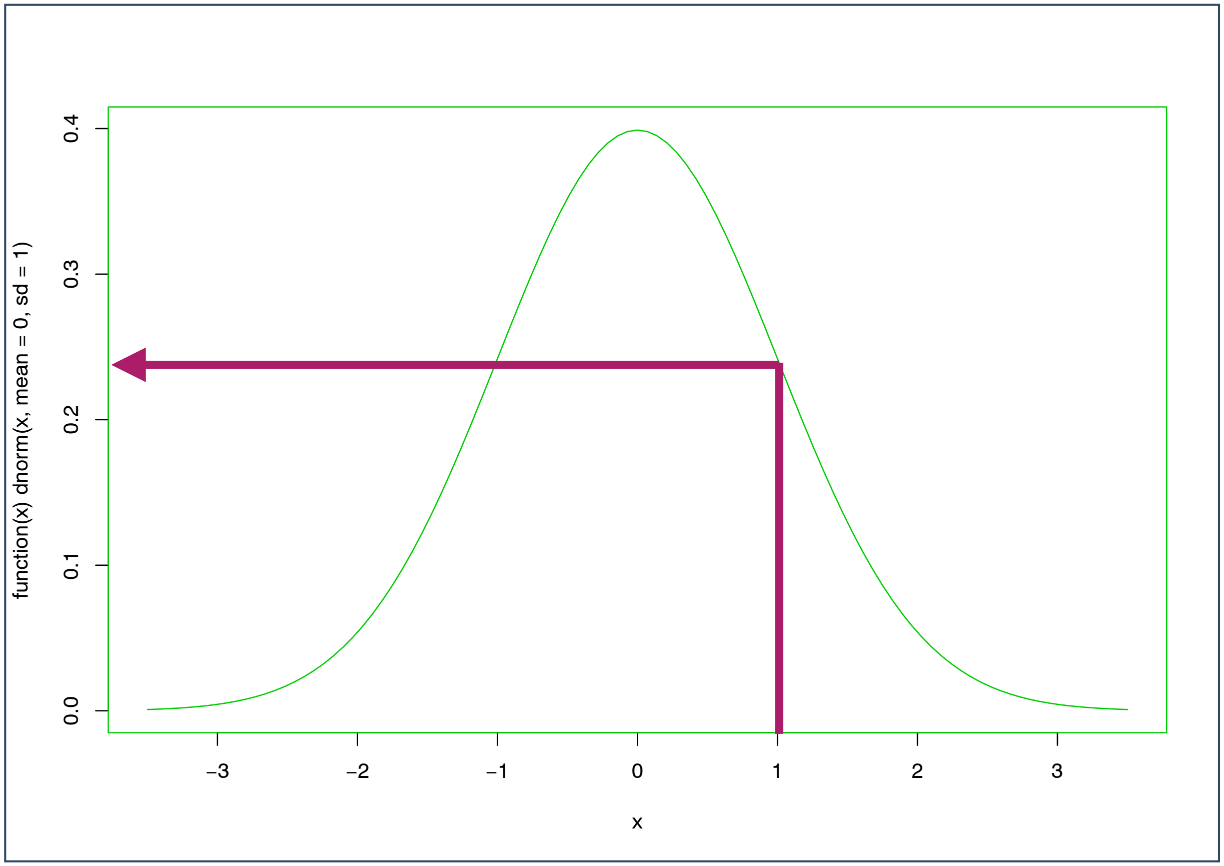
```
# Simulate 1,000,000 observations from the standard normal distribution
> y = rnorm(1000000, mean = 0, sd = 1)

> mean(y)
[1] -0.001268725

> var(y)
[1] 0.9995501

# Plot the probability distribution
> hist(y, freq = FALSE)
```





In continuous distributions, we can no longer compute the probability of a single value X. This is because there are an infinite number of x's, so the denominator is infinitely large. Instead we compute a **probability density**, or just **density**, which is the probability of measuring x in the interval $[x, x+dx]$. This is computed as: $p(x)\, dx$

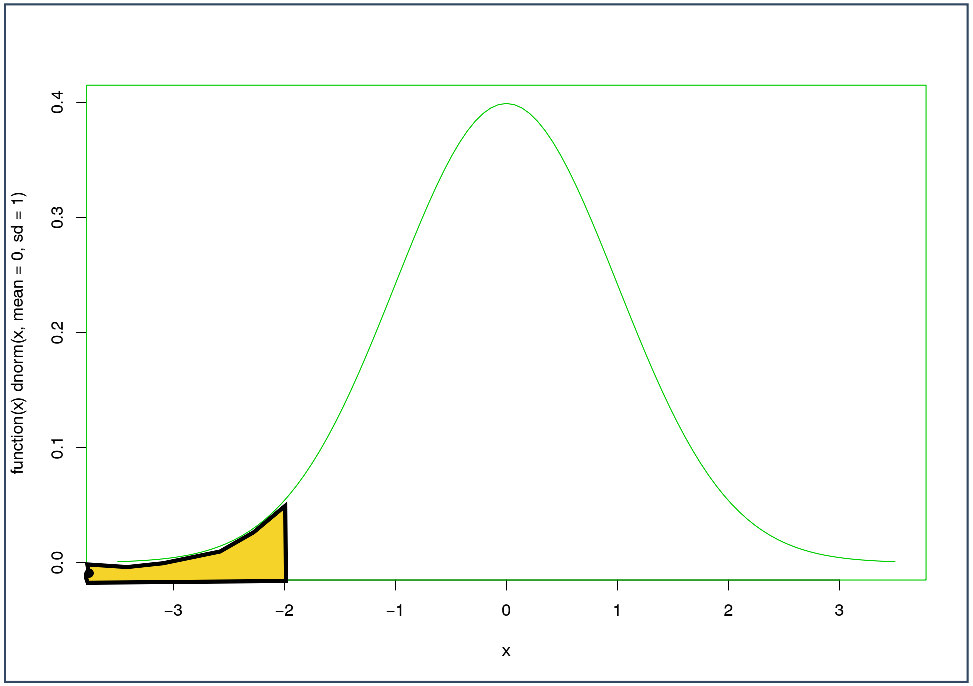
The yellow fill indicates how we might compute the probability density at $x = 1$. dx is the width of the yellow box, and $p(x)$ is the height. The product, $p(x)\, dx$, represents an area.

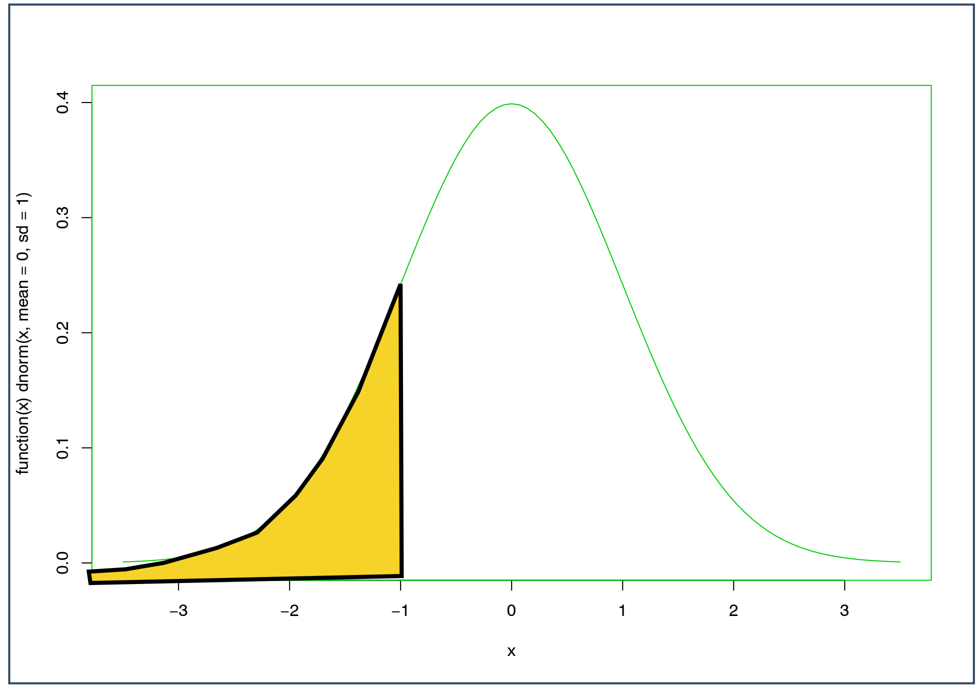


We use the `dnorm()` function to compute the density (*d*) in a normal (*norm*) distribution with a given mean and standard deviation.

```
# Density for x = 1 using the standard normal distribution
> dnorm(1, mean = 0, sd = 1)
[1] 0.2419707
```

\





The cumulative density is the area under the density curve from $-\infty$ to x.

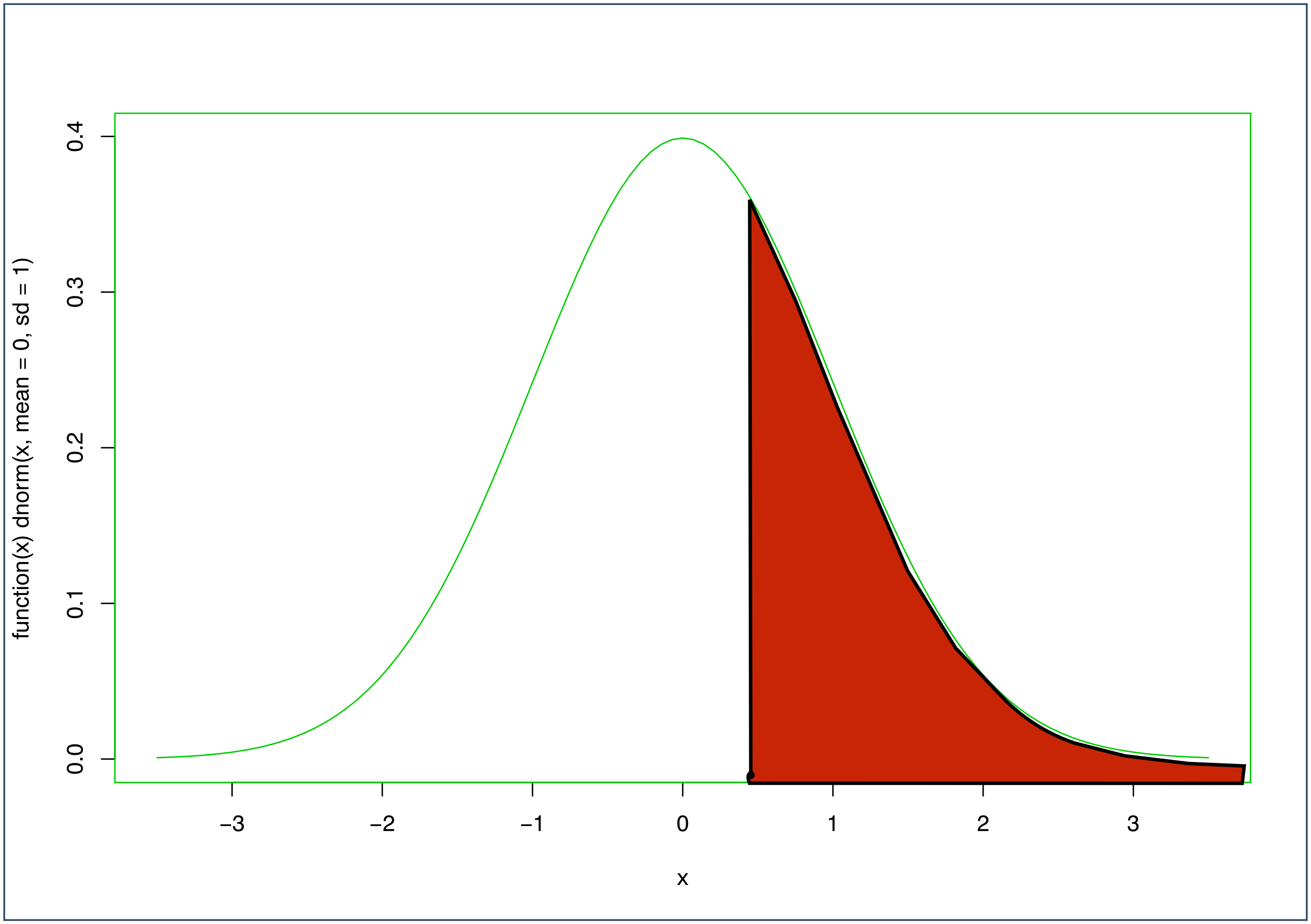
For the plot on the left, the yellow fill indicates $P(-\infty < z \le -2)$.

For the plot on the right, the yellow fill indicates $P(-\infty < z \le -1)$.

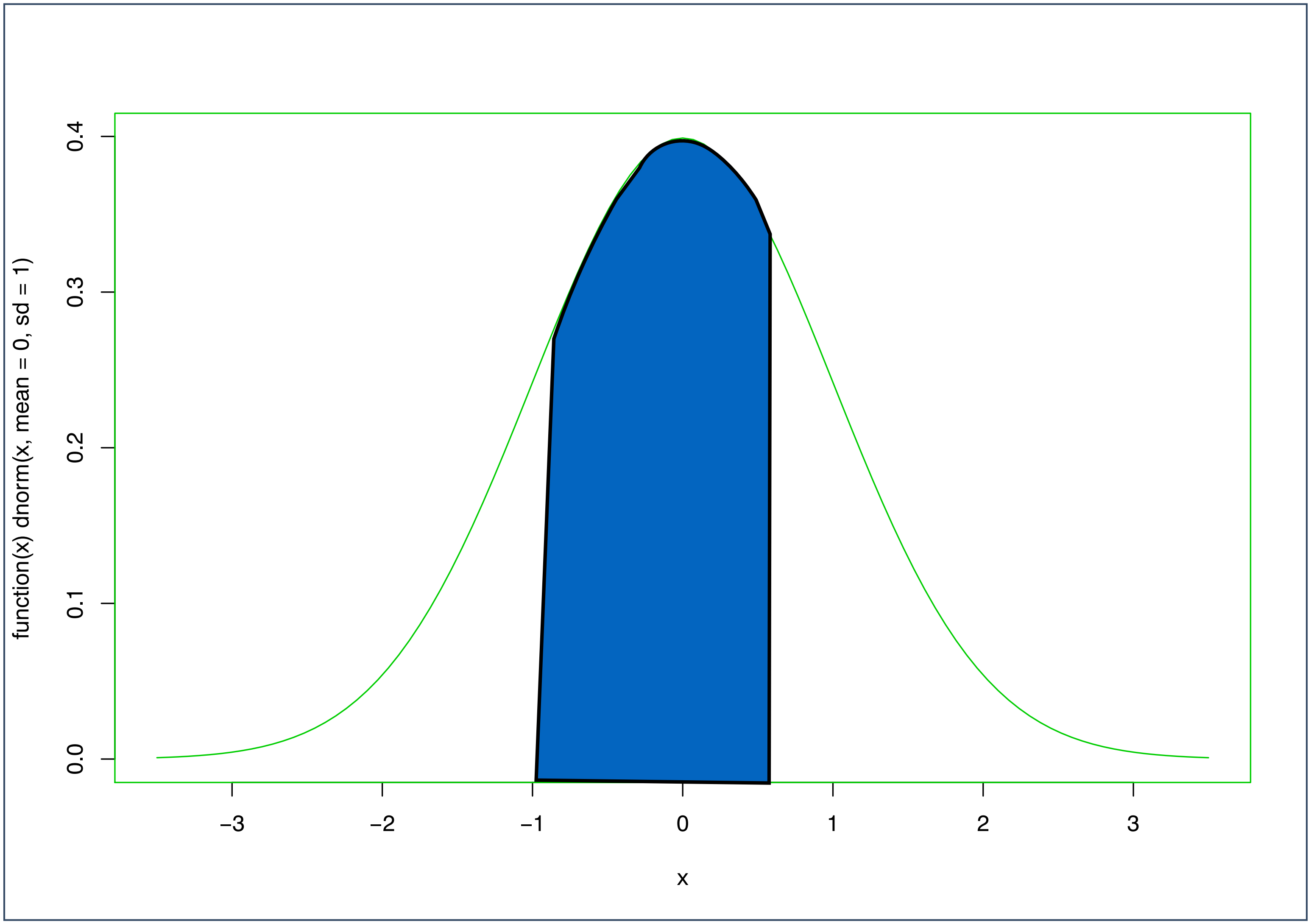
We use the `pnorm()` function to compute the cumulative probability (p) in a normal (*norm*) distribution with a given mean and standard deviation.

```
# Cumulative density for -2 using the standard normal distribution
> pnorm(-2, mean = 0, sd = 1)
[1] 0.02275013
```

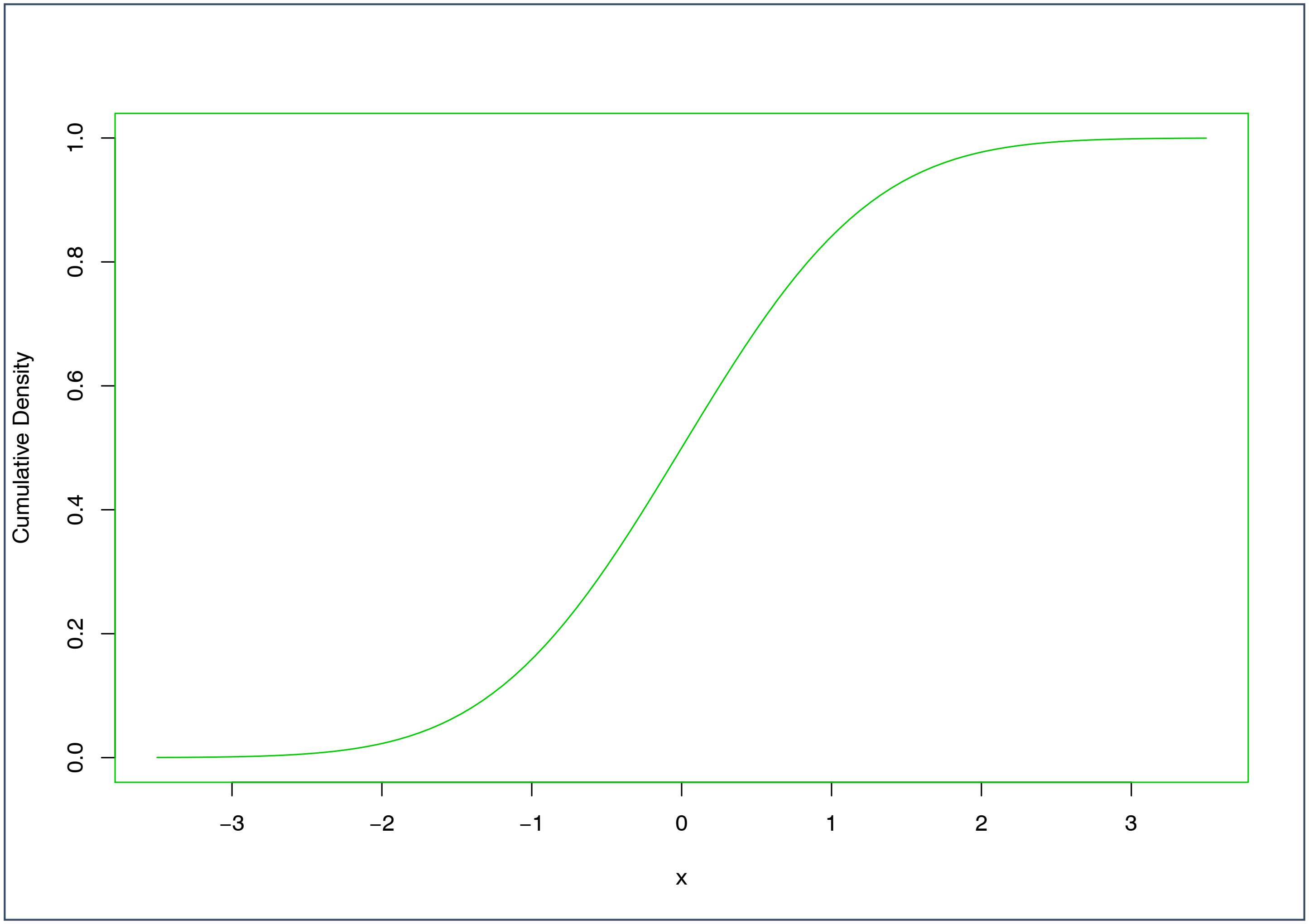
Find P(x > 0.5)
(red area)



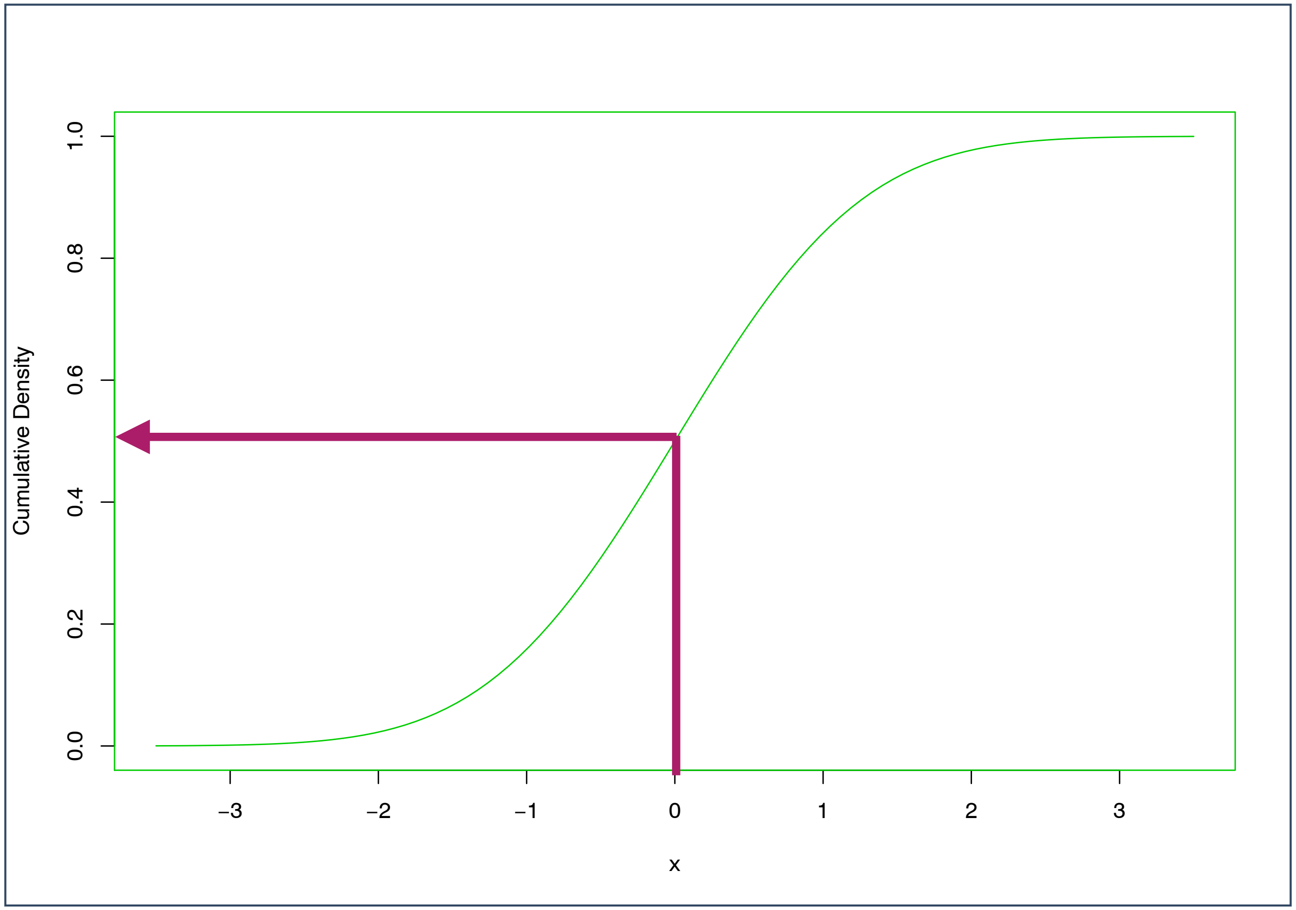
Find P(−1 < x ≤ 0.5)
(blue area)



Cumulative Density Function

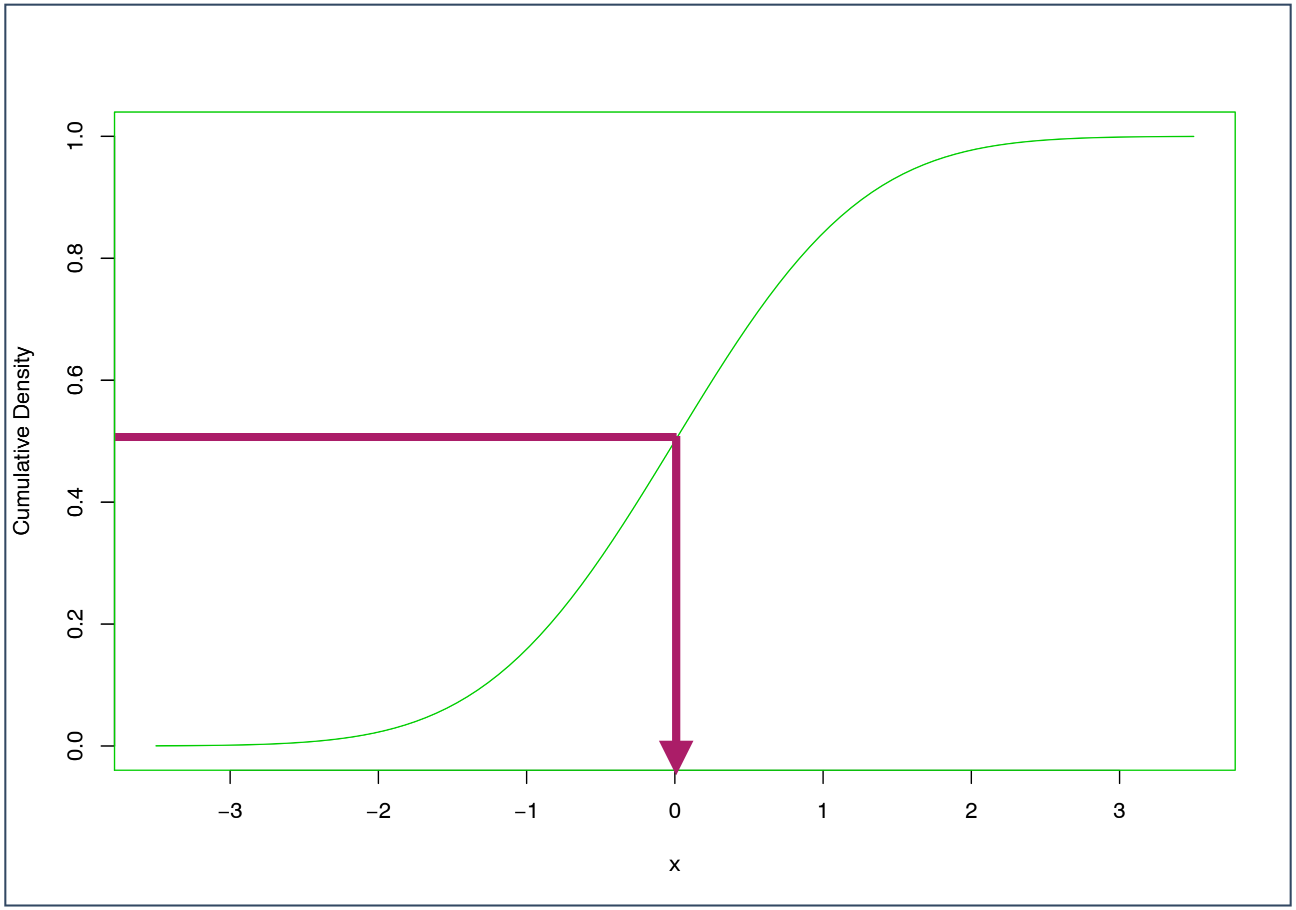


The **cumulative density function** (CDF) is a plot of the cumulative density for $-\infty < x < \infty$. This is the CDF for the standard normal distribution. Sometimes a plot of the CDF is referred to as an **ogive**.



The pnorm() function computes the cumulative probability (the value on the y-axis of the CDF) given a value of x (a value on the x-axis, and of course, the mean and standard deviation.

```
# Cumulative density for 0 using the standard normal distribution
> pnorm(0, mean = 0, sd = 1)
[1] 0.5
```



The qnorm() function computes the value of x (a value on the x-axis) given a cumulative probability (the value on the y-axis of the CDF), and of course, the mean and standard deviation.

```
# Compute quantile (x) given a cumulative density using the standard normal
distribution
> qnorm(0.5, mean = 0, sd = 1)
[1] 0
```

R Function	Purpose
`dnorm()`	Compute the probability density for a particular quantile (x)
`pnorm()`	Computes the cumulative density / probability for a particular quantile (x)
`qnorm()`	Computes the quantile for a given cumulative density / probability
`rnorm()`	Randomly generates an observation (x) from the normal distribution specifies

The `d`, `p`, `q`, and `r` functions exist for a variety of probability distributions. The optional arguments correspond to how the particular family of distributions is parameterized. For example, the `norm` functions require the arguments `mean=` and `sd=`, whereas the `pois` functions require `lambda=`.

Some Common Distributions and Their R Family Names

R Function	Distribution
`[d,p,q,r]beta`	Beta
`[d,p,q,r]binom`	Binomial
`[d,p,q,r]chisq`	Chi-Square
`[d,p,q,r]f`	*F*
`[d,p,q,r]multinom`	Multinomial
`[d,p,q,r]nbinom`	Negative Binomial
`[d,p,q,r]norm`	Normal
`[d,p,q,r]pois`	Poisson
`[d,p,q,r]t`	Student's *t*
`[d,p,q,r]unif`	Uniform

There are many other probability distributions included in the base R packages and also in other add-on packages.

To determine the parameters that need to be specified for each, examine the help menu for any of that distributions functions, e.g.,

```
> help(rnorm)
> help(pnorm)
```