

This document will describe how to use R to calculate probabilities associated with common distributions as well as to graph probability distributions.

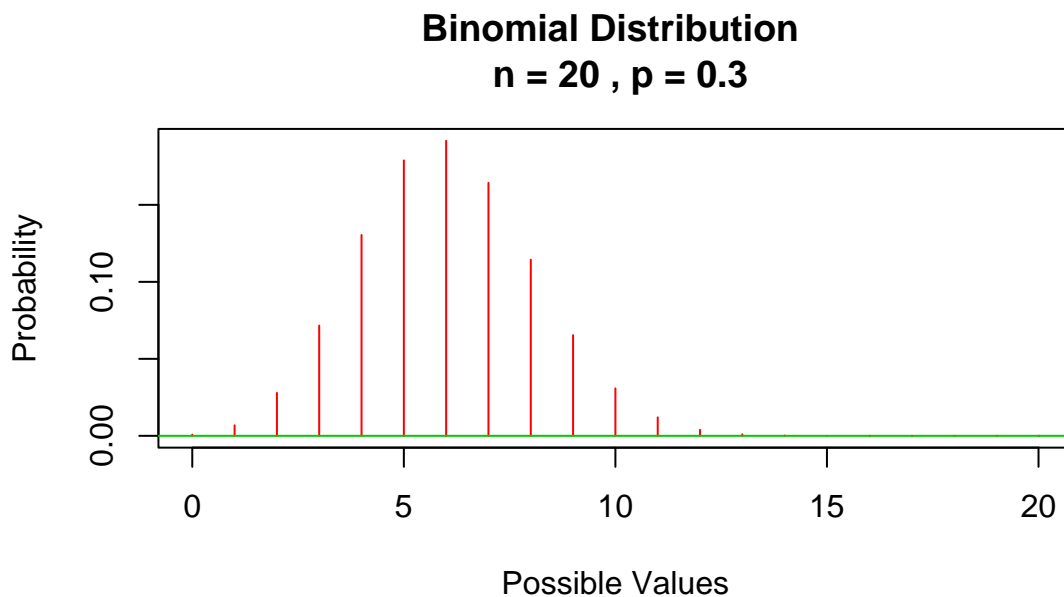
R has a number of built in functions for calculations involving probability distributions, both discrete and continuous. This semester we will see the *Binomial*, *Poisson*, and *Normal* distributions early during our introduction to probability. Later, we will use the *t*, *F*, and *chi-square* distributions in our study of statistical inference.

For each of these distributions (and others), R has four primary functions. Each function has a one letter prefix followed by the root name of the function. The names make mnemonic sense for continuous random variables but are used in both cases. For example `dnorm` is the height of the density of a normal curve while `dbinom` returns the probability of an outcome of a binomial distribution. Here is a table of these commands.

Prefix	Meaning		Distribution	Root
	Continuous	Discrete		
d	density	probability (pmf)	Binomial	<code>binom</code>
p	probability (cdf)	probability (cdf)	Poisson	<code>pois</code>
q	quantile	quantile	Normal	<code>norm</code>
r	random	random	<i>t</i>	<code>t</code>
			<i>F</i>	<code>F</code>
			Chi-square	<code>chisq</code>

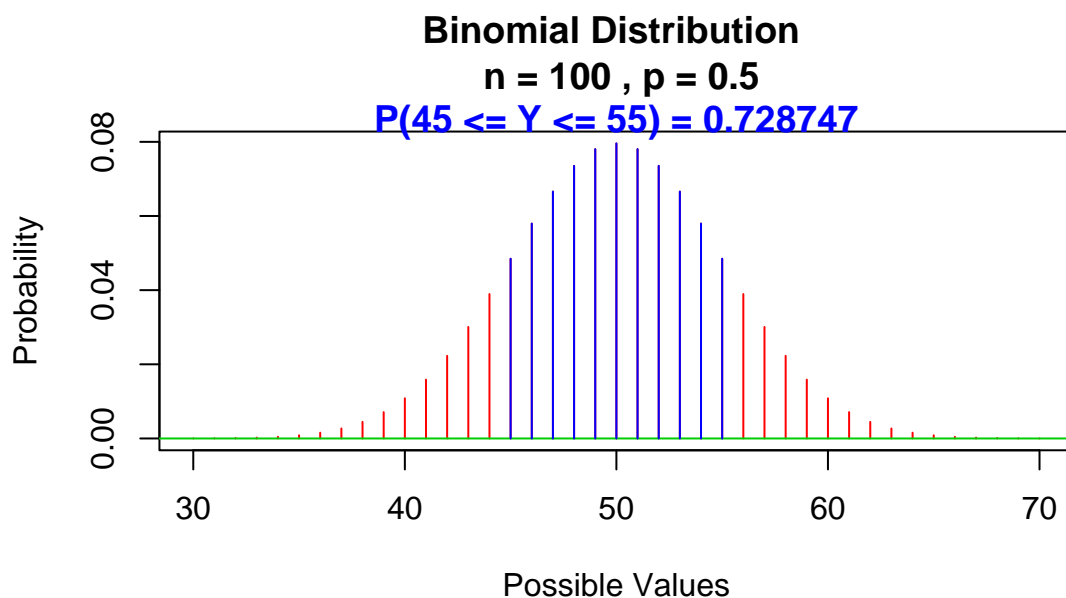
Graphing Probability Distributions. The file `prob.R` contains function that may be used to graph and visualize the binomial and normal distributions. Here are some examples of their use.

```
> source("prob.R")
> gbinom(20, 0.3)
```



This plot will help visualize the probability of getting between 45 and 55 heads in 100 coin tosses.

```
> gbinom(100, 0.5, a = 45, b = 55, scale = T)
```



The Binomial Distribution. The binomial distribution is applicable for counting the number of outcomes of a given type from a prespecified number n independent trials, each with two possible outcomes, and the same probability of the outcome of interest, p . The distribution is completely determined by n and p . The probability mass function is defined as:

$$\Pr\{Y = j\} = \binom{n}{j} p^j (1-p)^{n-j}$$

where

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}$$

is called a binomial coefficient. (Some textbooks use the notation ${}_n C_j$ instead.) In R, the function `dbinom` returns this probability. There are three required arguments: the value(s) for which to compute the probability (j), the number of trials (n), and the success probability for each trial (p).

For example, here we find the complete distribution when $n = 5$ and $p = 0.1$.

```
> dbinom(0:5, 5, 0.1)
```

```
[1] 0.59049 0.32805 0.07290 0.00810 0.00045 0.00001
```

If we want to find the single probability of exactly 10 successes in 100 trials with $p = 0.1$, we do this.

```
> dbinom(10, 100, 0.1)
```

```
[1] 0.1318653
```

The function `pbinom` is useful for summing consecutive binomial probabilities. With $n = 5$ and $p = 0.1$, here are some example calculations.

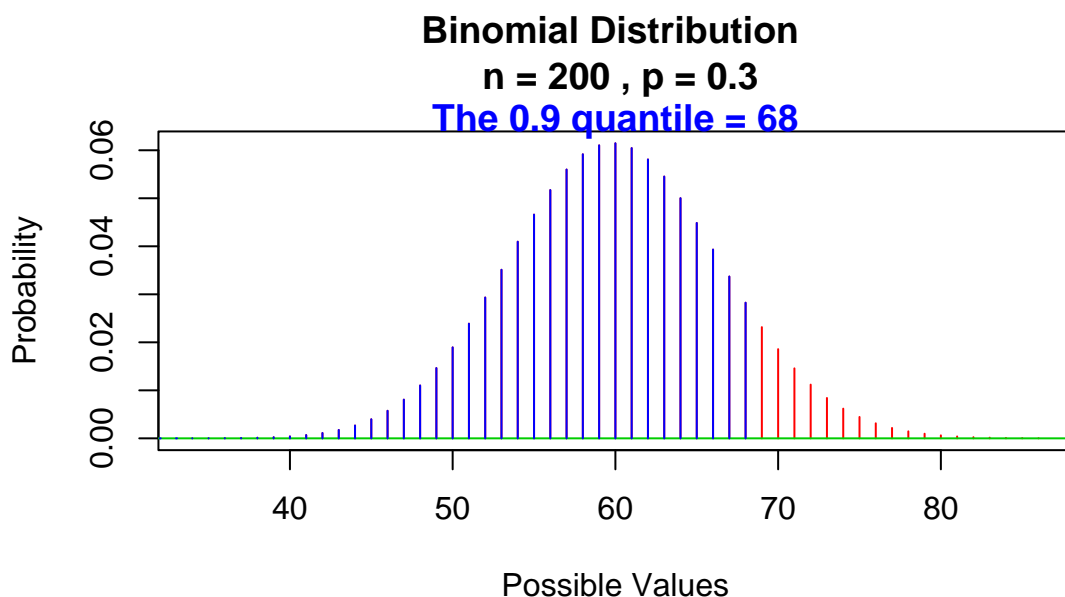
$$\begin{aligned}\Pr\{Y \leq 2\} &= \text{pbinom}(2,5,0.1) \doteq 0.99144 \\ \Pr\{Y \geq 3\} &= 1 - \Pr\{Y \leq 2\} = 1 - \text{pbinom}(2,5,0.1) \doteq 0.00856 \\ \Pr\{1 \leq Y \leq 3\} &= \Pr\{Y \leq 3\} - \Pr\{Y \leq 0\} = \text{pbinom}(3,5,0.1) - \text{pbinom}(0,5,0.1) \doteq 0.40905\end{aligned}$$

We can also find the quantiles of a binomial distribution. For example, here is the 90th percentile of a binomial distribution with $n = 200$ and $p = 0.3$. The function `qbinom` finds the quantile. The function `gbinom` with option `quantile = 0.9` graphs the distribution.

```
> qbinom(0.9, 200, 0.3)
```

```
[1] 68
```

```
> gbinom(200, 0.3, scale = T, quantile = 0.9)
```



The last function for the binomial distribution is used to take random samples. Here is a random sample of 20 binomial random variables drawn from the binomial distribution with $n = 10$ and $p = 0.5$.

```
> rbinom(20, 10, 0.5)
```

```
[1] 6 7 3 5 3 6 7 6 5 8 5 5 6 4 5 7 5 3 5 6
```

Normal Distribution Normal distributions have symmetric, bell-shaped density curves that are described by two parameters: the mean μ and the standard deviation σ . The two points of a normal density curve that are the steepest—at the “shoulders” of the curve—are precisely one standard deviation above and below the mean.

Heights of individual corn plants may be modeled as normally distributed with a mean of 145 cm and a standard deviation of 22 cm (Samuels and Witmer, third edition, exercise 4.29). Here are several example normal calculations using R. The commands using `gnorm` allow you to visualize the answers.

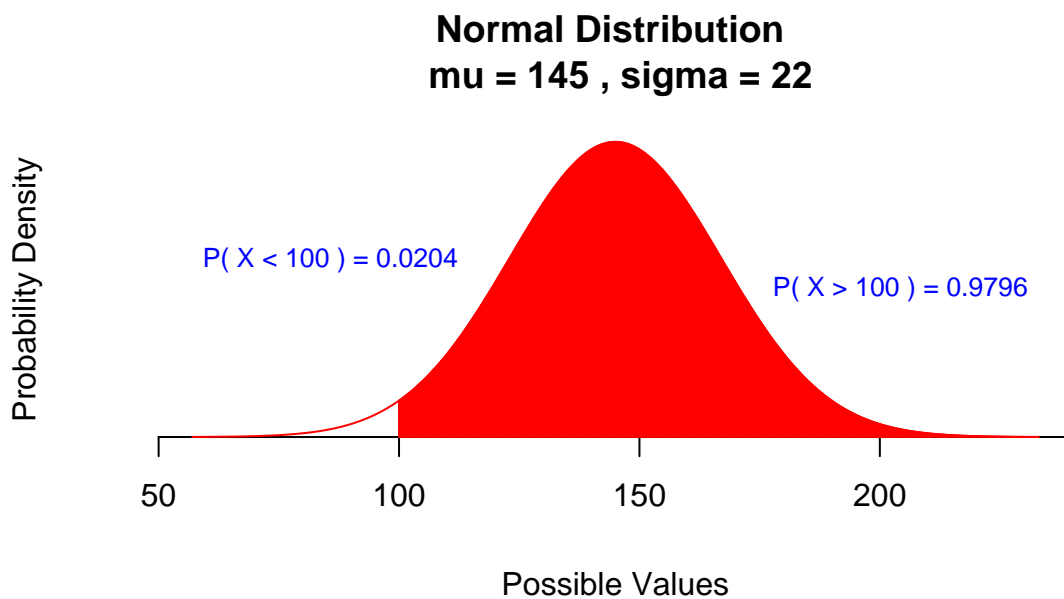
Find the proportion of plants:

... larger than 100cm;

```
> 1 - pnorm(100, 145, 22)
```

```
[1] 0.979595
```

```
> gnorm(145, 22, a = 100)
```



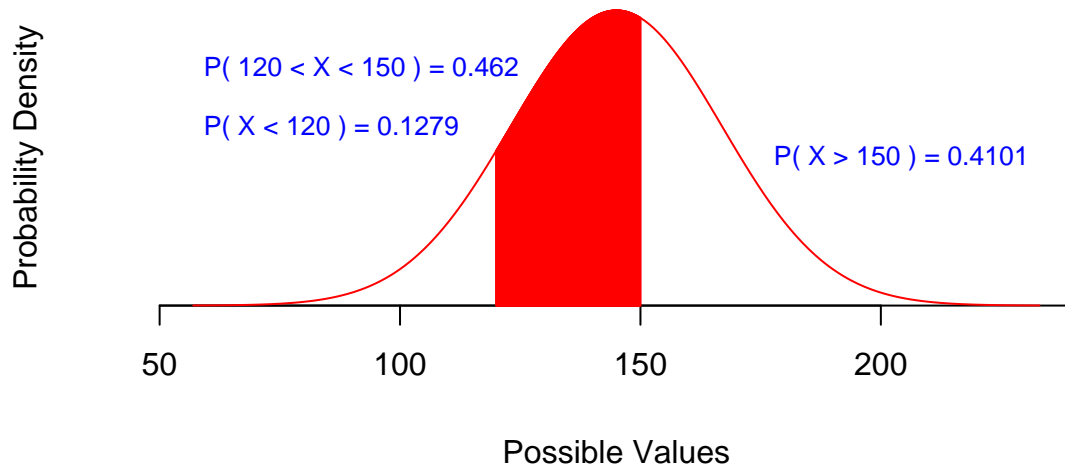
... between 120cm and 150cm:

```
> pnorm(150, 145, 22) - pnorm(120, 145, 22)
```

```
[1] 0.461992
```

```
> gnorm(145, 22, a = 120, b = 150)
```

Normal Distribution mu = 145 , sigma = 22



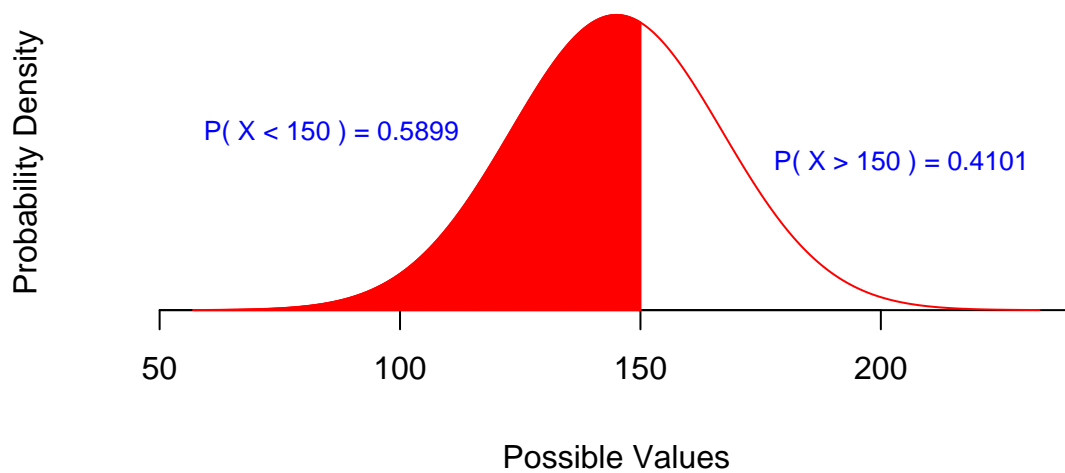
... 150cm or less:

```
> pnorm(150, 145, 22)
```

```
[1] 0.5898942
```

```
> gnorm(145, 22, b = 150)
```

Normal Distribution mu = 145 , sigma = 22



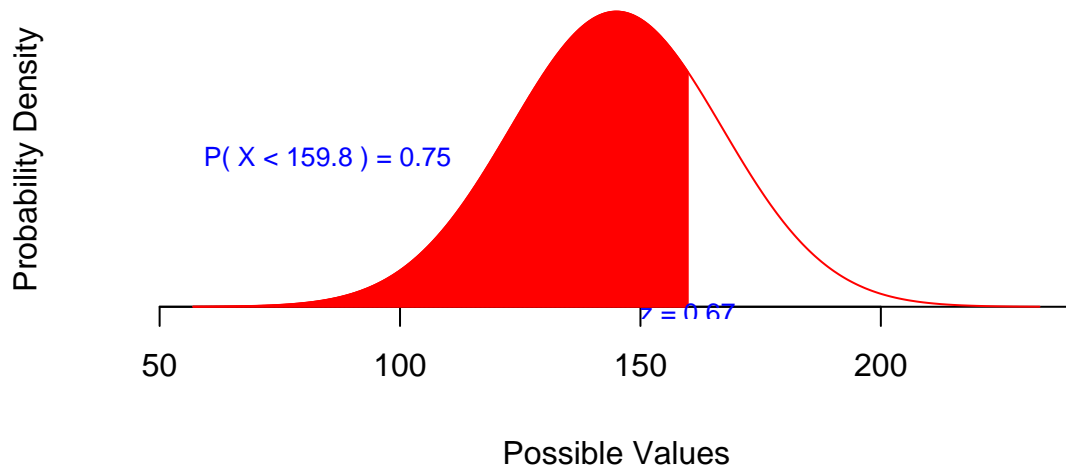
Find the 75th percentile.

```
> qnorm(0.75, 145, 22)
```

```
[1] 159.8388
```

```
> gnorm(145, 22, quantile = 0.75)
```

Normal Distribution mu = 145 , sigma = 22



Find the endpoints of middle 95% of the distribution.

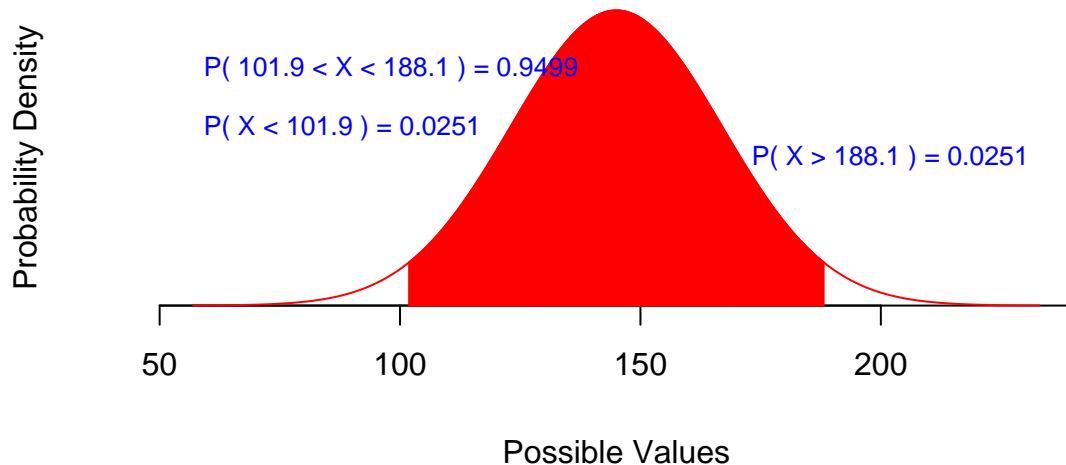
```
> ab = qnorm(c(0.025, 0.975), 145, 22)
```

```
> ab
```

```
[1] 101.8808 188.1192
```

```
> gnorm(145, 22, a = round(ab[1], 1), b = round(ab[2], 1))
```

Normal Distribution mu = 145 , sigma = 22



Other Distributions Other distributions work in a similar way, except that I have not yet written analogous graphing functions. Details on how to express the parameters for different probability distributions can be found from the help files. For example, to learn about find Poisson probabilities, type `?dpois`.