Week 6: Binomial and Normal Distributions

Stat 201: Statistics I

March 2, 2019

R Distribution Functions

R provides four functions for many commonly used distributions. These functions have names that begin with a letter (`d`, `p`, `q` or `r`) followed by the distribution name or an abbreviation of the name (i.e. `norm` for normal).

Function	Binomial	Normal
Density	`dbinom`	`dnorm`
Probability	`pbinom`	`pnorm`
Quantile	`qbinom`	`qnorm`
Random	`rbinom`	`rnorm`

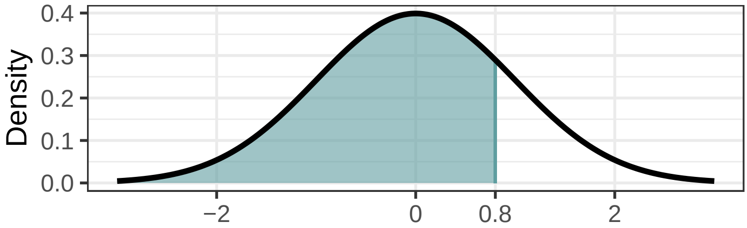
Probability functions

The `p` functions will return the probability, or proportion of, a distribution is less than or greater than a given value.

- Example: What proportion of the standard normal distribution is less than 0.8?

```r
pnorm(0.8)
```

```
## [1] 0.7881446
```



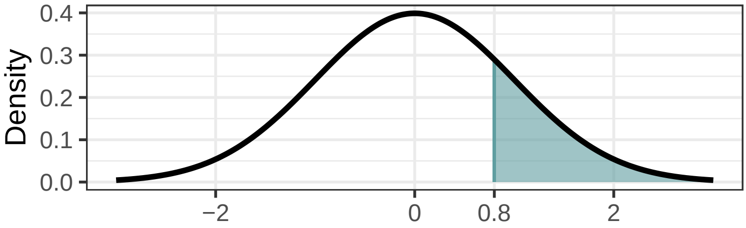
Probability functions, cont.

By default, the probability functions give proportion of a distribution that is *less than or equal to* the given value ($P(X \leq x)$). To find the proportion *greater than* a value ($P(X > x)$), include the parameter `lower.tail = FALSE` in the function call. This is equivalent to $1 - P(X \leq x)$

- Example: What proportion of the standard normal distribution is greater than 0.8?

```r
pnorm(0.8, lower.tail = FALSE)
```

```
## [1] 0.2118554
```



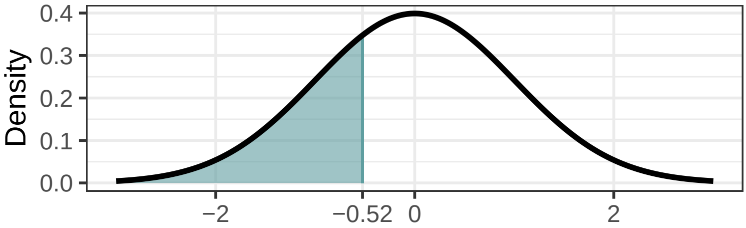
Quantile functions

The q functions will return the value the corresponds to a given quantile. In other words, it is the value that defines the given probability of a distribution.

- Example: What value is greater than 30% of a standard normal distribution?

```r
qnorm(0.3)
```

```
## [1] -0.5244005
```



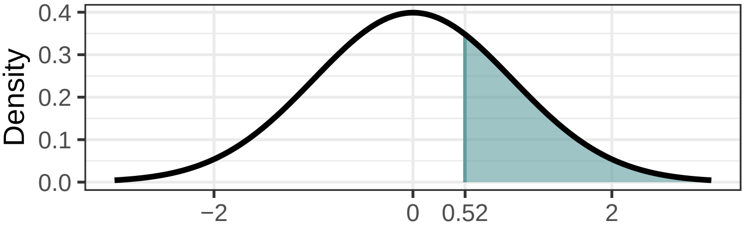
Quantile functions, cont.

The quantile functions, like the probability functions, by default will find the value that the given proportion is *less than*. To find the value that defines an upper proportion, use the `lower.tail = FALSE` parameter.

- Example: What value is less than 30% of a standard normal distribution?

```
qnorm(0.3, lower.tail = FALSE)
```

```
## [1] 0.5244005
```



Relationship of probability and quantile functions

The probability and quantile functions are inverses, or opposites, of each other. That is, if you pass the result of one into the other, you will get the original value back (there might be slight differences due to rounding).

```r
q <- qnorm(0.3)
pnorm(q)
```

```
## [1] 0.3
```

```r
p <- pnorm(0.8)
qnorm(p)
```

```
## [1] 0.8
```

Other functions

We will not use the other two types of distribution functions much, if at all, in this class.

- Density functions (`d`) give the value of a distribution's density function for a given point in the distribution. They are used to create the distribution plots in these slides.

```r
dnorm(0.8)
```

```
## [1] 0.2896916
```

- Random functions (`r`) will generate random numbers drawn from the distribution. They can be used to create data for simulation studies.

```r
rnorm(5)
```

```
## [1]  1.3709584 -0.5646982  0.3631284  0.6328626  0.4042683
```

Binomial distributions in R

- Binomial distributions represent the number of successes in a set number of trials with exactly two possible outcomes (success/failure).
- A particular distribution is defined by two parameters, the number of trials and the probability of success for each trial.
- When using the binomial distribution functions in R, the parameters are specified by adding `size=` and `prob=` to the function call.
- Recall, the possible values of a binomial process are between 0 and the number of trials (inclusive).

Flipping coins example

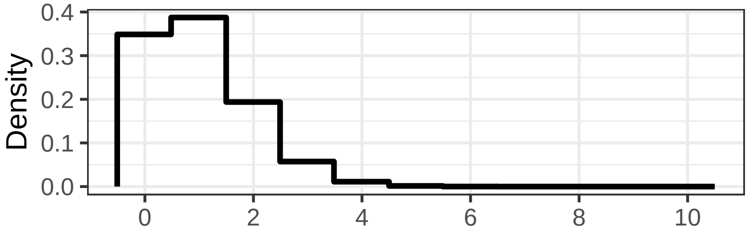
Consider flipping a fair coin three times, counting the number of times that the flip resulted in "heads".

- For this process, the number of trials is 3, the probability of success is 0.5. Possible values are 0, 1, 2, or 3.
- Then, to utilize this distribution in R, we should include the parameters `size = 3` and `prob = 0.5` in the appropriate function calls.

Cancer screening example

Consider the cancer screening example from class. We are interested in the number out of 10 randomly selected subjects who get a positive result from the screening test. The probability of testing positive is 10%.

- For this process, the number of trials is 10, the probability of success is 0.1. Possible values are 0 through 10.
- To utilize this distribution in R, we should include the parameters `size = 10` and `prob = 0.1` in the appropriate function calls.



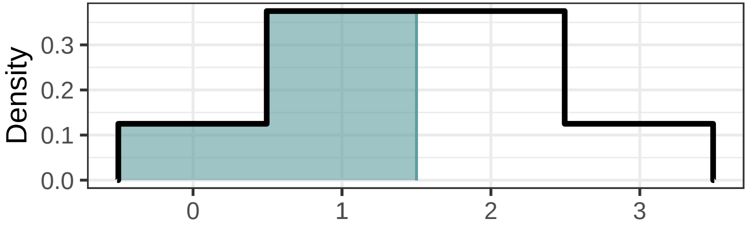
Binomial probabilities of a range

To calculate binomial probabilities, we use the `pbinom` function. This function will return the probability of getting a value less than *or equal to* the given value by default.

- What is the probability of getting one or fewer (0 or 1) heads out of three coin flips?

```
pbinom(1, size=3, prob=0.5)
```

```
## [1] 0.5
```



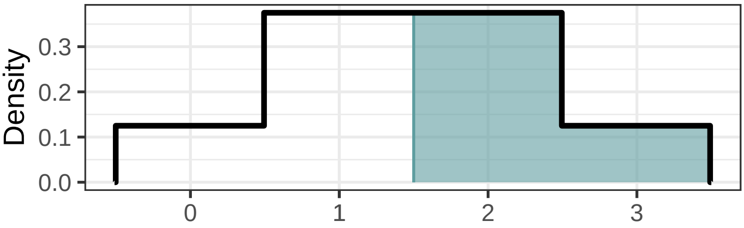
Binomial probabilities greater than a value

To calculate binomial probabilities greater than a value (not greater than *or equal to*), add the `lower.tail = FALSE` to the function call.

- What is the probability of getting more than one (2 or 3) heads out of three coin flips?

```
pbinom(1, size=3, prob=0.5, lower.tail = FALSE)
```

```
## [1] 0.5
```



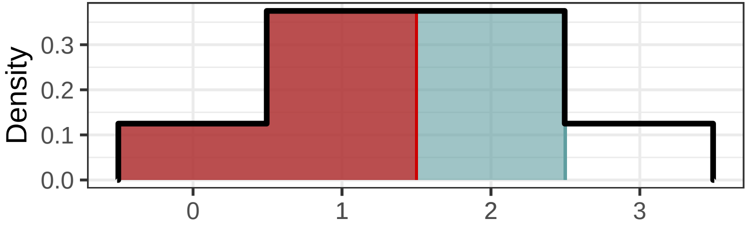
Binomial probabilities of a single value

There are two methods to calculate the probability of a single value in a binomial distribution. The first method is find the probability less than or equal the value and subtract the probability less than or equal to the value minus one. What will be left is the probability of getting exactly the single value.

- What is the probability of getting exactly two heads out of three coin flips?

```
pbinom(2, size=3, prob=0.5) - pbinom(1, size=3, prob=0.5)
```

```
## [1] 0.375
```



Binomial probabilities of a single value, cont.

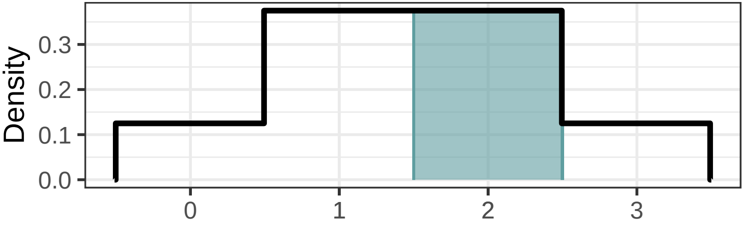
Alternatively, the density function can be used to find the probability of a single value.

[Note: this method can only be used for discrete distributions.]

- What is the probability of getting exactly two heads out of three coin flips?

```r
dbinom(2, size = 3, prob = 0.5)
```

```
## [1] 0.375
```



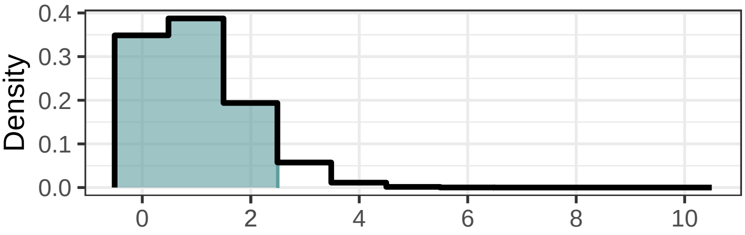
Binomial quantiles

Finding quantiles in a binomial distribution is not a common problem, but if necessary, they can be found with `qbinom`.

- What number or less of positive screening tests out of 10 will occur 75% of the time?

```r
qbinom(0.75, size = 10, prob = 0.1)
```

```
## [1] 2
```

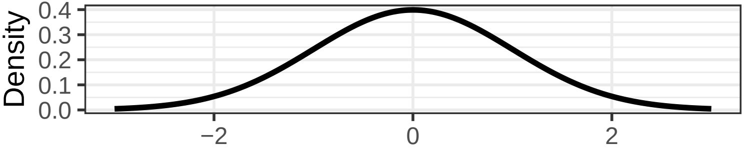


Normal distributions in R

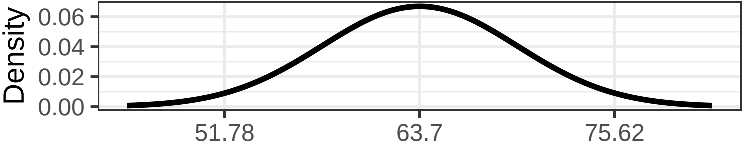
- Normal distributions are continuous distributions representing a many kinds naturally occurring variables.
- A normal distribution is defined by it's mean and standard deviation.
- The standard normal distribution (the z distribution) has a mean of 0 and standard deviation of 1.
- The parameters that define a distribution are specified in R functions by including `mean=` and `sd=` in the function call.
- Because R defines default values for `mean` and `sd` of 0 and 1, respectively, when we want to do calculations from the standard normal distribution, we don't have to include those values.

Normal distributions, examples

- The standard normal distribution (`mean=0` and `sd=1`)



- The heights of adult women (inches) are normally distributed with a mean of 63.7 and standard deviation of 5.96.



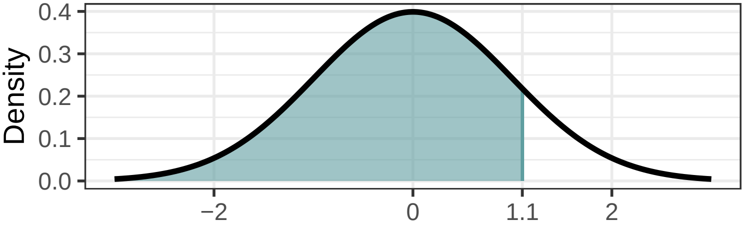
Normal probabilities

To find probabilities from normal distributions, we use the function `pnorm`. By default, this function gives probabilities less than a value. [In continuous distributions $P(X = x) = 0$, so $P(X \leq x) = P(X < x)$.]

- What proportion values from normal distributions are less than 1.1 standard deviations above the mean ($P(z < 1.1)$)?

```r
pnorm(1.1)
```

```
## [1] 0.8643339
```



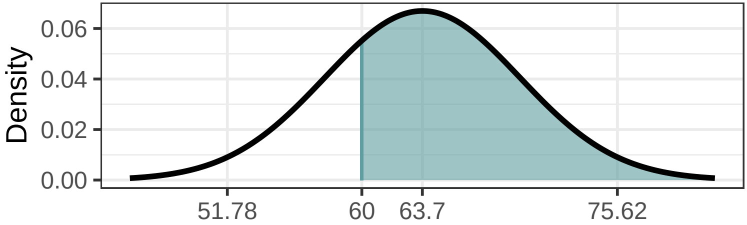
Upper normal probabilities

To find probabilities *greater than* a value, we include the `lower.tail=FALSE` parameter.

- What is the probability of a randomly selected woman being taller than 5 feet (60 inches)?

```
pnorm(60, mean = 63.7, sd = 5.96, lower.tail = FALSE)
```

```
## [1] 0.7326362
```



Probabilities of ranges

- While there is no sense in finding the probability of a single value in a continuous distribution, there is often the need to find probabilities of ranges of values.
- In a similar manner as finding the probability of a single value in a binomial, the probability of a range of values can be found by subtracting the probability less than the lower bound from the probability lass than the upper value.
- This is equivalent to $P(a < x < b) = P(x < b) - P(x < a)$

Example:

- What is the proportion of values within one standard deviation of the mean?

```
pnorm(1) - pnorm(-1)
```

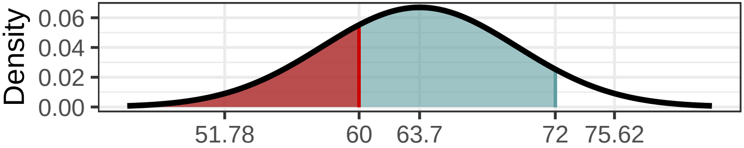
```
## [1] 0.6826895
```

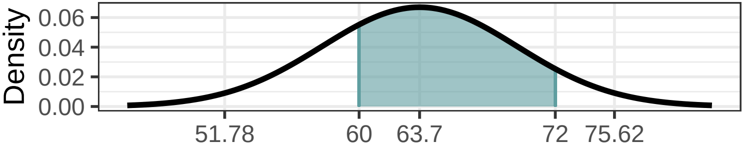
Probabilities of ranges, example

- What is the probability of a randomly selected woman being between 5 and 6 feet tall?

```r
pnorm(72, mean = 63.7, sd = 5.96) - pnorm(60, mean = 63.7, sd = 5.96)
```

```
## [1] 0.6507684
```





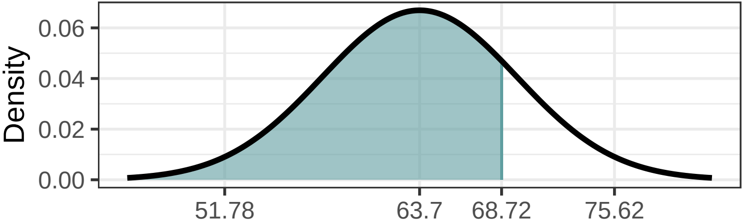
Quantiles in normal distributions

To find a quantile of a normal distribution, we use the `qnorm` function.

- 80% of women are shorter than what height?

```r
qnorm(0.8, mean = 63.7, sd = 5.96)
```

```
## [1] 68.71606
```



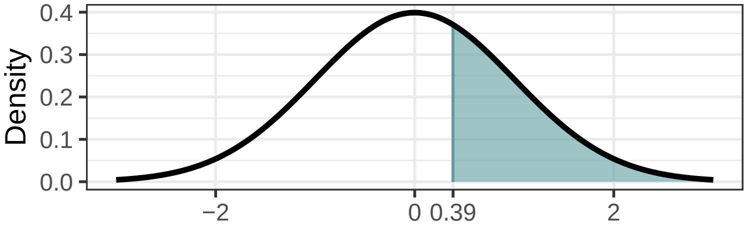
Upper quantiles

Use the `lower.tail = FALSE` option to find quantiles corresponding to upper probabilities.

- What value separates the upper 35% from the lower 65% in a standard normal distribution?

```r
qnorm(0.35, lower.tail = FALSE)
```

```
## [1] 0.3853205
```

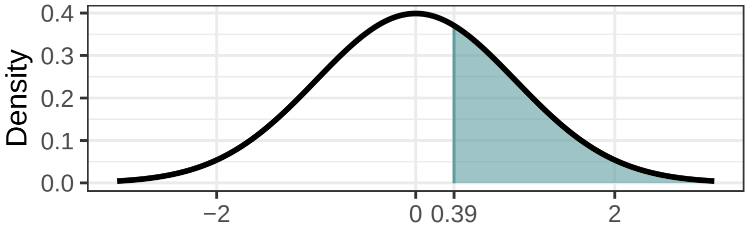


Upper quantiles, alternate

A value that is less than p proportion of a distribution is also greater than $1 - p$ proportion of the distribution [this is a form of the complement rule, i.e. if $P(X > x) = p$, then $P(X < x) = 1 - p$]. This can be used as another way to find the quantile of an upper proportion.

- What value separates the upper 35% from the lower 65% in a standard normal distribution?

```
qnorm(1 - 0.35)
## [1] 0.3853205
```



Critical values

Recall, a critical value z_α is the value of a standard normal distribution that delineates an upper probability of α from the rest of the distribution. This is a quantile of an upper proportion.

- What are the critical values z_α and $z_{\alpha/2}$ for $\alpha = 0.05$?

```
alpha <- 0.05
qnorm(alpha, lower.tail = FALSE)
```

```
## [1] 1.644854
```

```
qnorm(1 - alpha/2)
```

```
## [1] 1.959964
```