Day 32 DPQR Functions: Calculations with the Normal Model

12/7/2019 rev 1/5/2020

Random Number Functions

In this section you will learn about:

- the `r--`, `d--`, `p--`, `q--` functions that go with random variables.
- How to use these four functions to analyze values from the perspective of a random fluctuation.
- How to use these to generate random numbers

Random Variables

In statistics we deal with an idea called a *random variable*. These are different from the variables that you have dealt with in other math classes in important ways, and you shouldn't try to use them in the way that you have in other math classes. Most importantly, you *don't solve for a random variable*. Random variables, often written with capital letters such as X, stand for a random process. So to fully describe a random variable we need to tell both what values it can take, and a way to determine the likelihood that those values will occur. The easiest forms of random variables are the *discrete random variables* which we worked with last term in the Probability unit. Now we are working with continuous random variables.

Continuous Normal Random Variable A continuous random variable whose probabilities are described by the normal distribution. Basically, a variable that we can model with the Normal model, like SAT scores, height, weight, IQ, etc.

In `R` every kind of continuous random variable comes with 4 functions. The `d--`, `p--`, `q--`, and `r--`.

- `d--` - gives the **d**ensity (or relative likelihood) of a value. This is used for graphing.
- `p--` - gives the **p**robability of getting a value less than or equal to the given one
- `q--` - gives the value that corresponds to a given **q**uantile.
- `r--` - gives **r**andom values that follow the desired rule.

We are dealing with Normal random variables here so we will have "norm" after each letter:

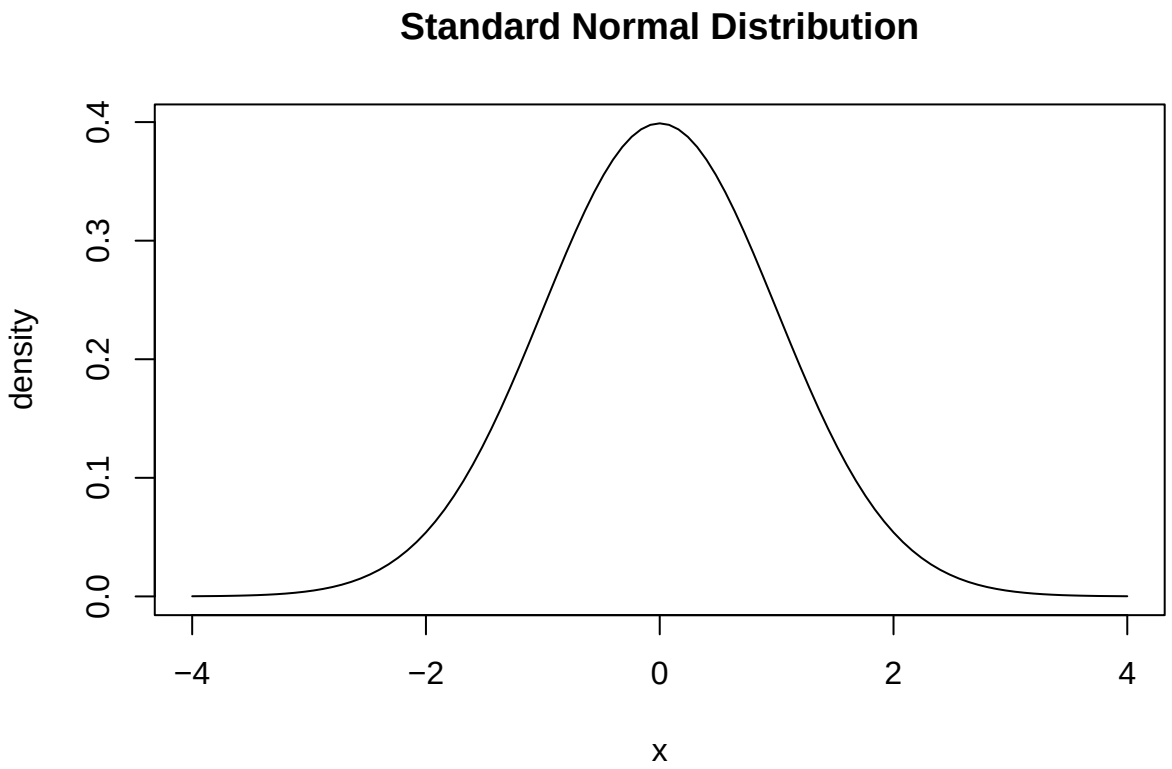
- `dnorm`
- `pnorm`
- `qnorm`
- `rnorm`

What are these functions good for?

`dnorm`

In general the `d--` functions are mostly used for drawing probability density graphs (which means, essentially, probability distributions). This graph is of the standard normal distribution:

```
curve( dnorm(x, mean=0, sd=1),
       xlim=c(-4,4), #pick your xlims by taking  mean - 4sd and mean +4sd
       main="Standard Normal Distribution",
       ylab="density")
```

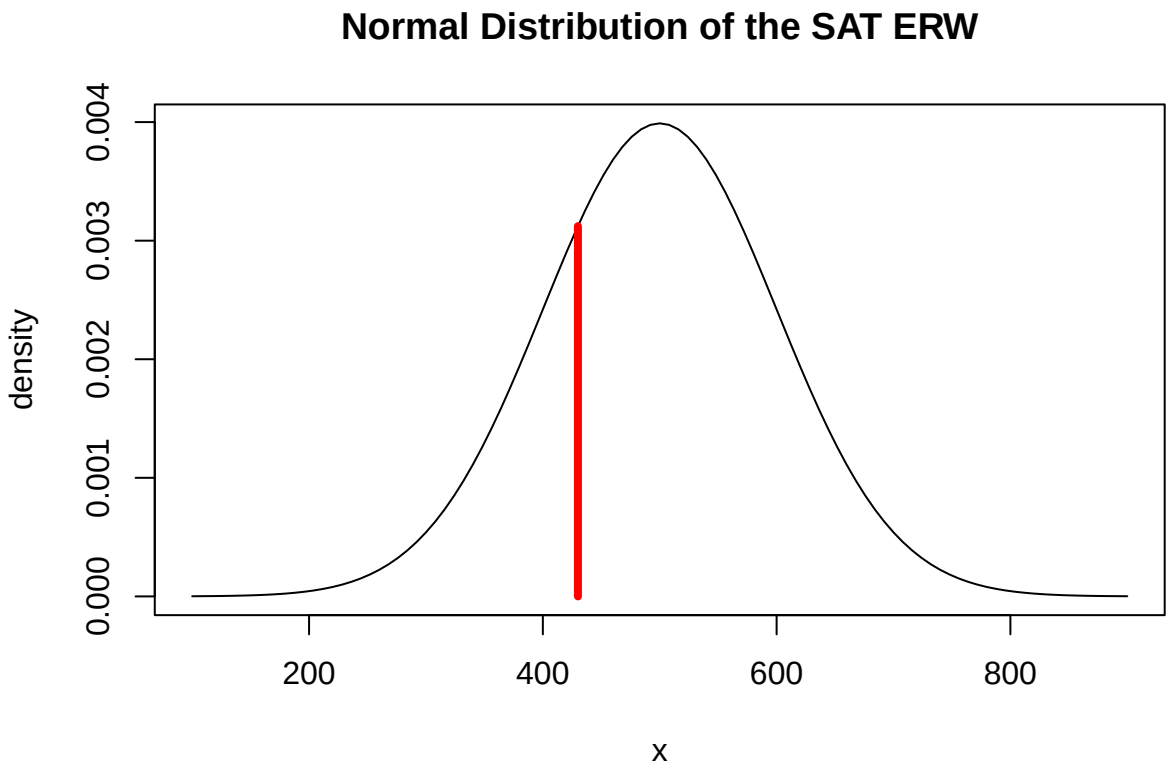


If we want to draw the Normal distribution for the Evidence-Based Reading and Writing section (ERW) of the SAT which has a mean of 500 and a standard deviation of 100 (so it is N(500, 100)), it would look like this:

If we want we can draw a line on the graph to indicate a value of note. Imagine that you scored a 430 on the ERW section of the SAT. You could represent your score as follows:

```
#draw the Normal distribution
curve( dnorm(x, mean=500, sd=100),
       xlim=c( 100, 900),  #pick your xlim by taking the mean - 4sd and mean +4sd
       main="Normal Distribution of the SAT ERW",
       ylab="density")

#add a segment at our value of note
segments( 430, 0,
          430, dnorm(430, mean=500, sd=100),
          #these are the coordinates (430,0) and (430, f(430) )
          col = "red",
          lwd = 4)
```

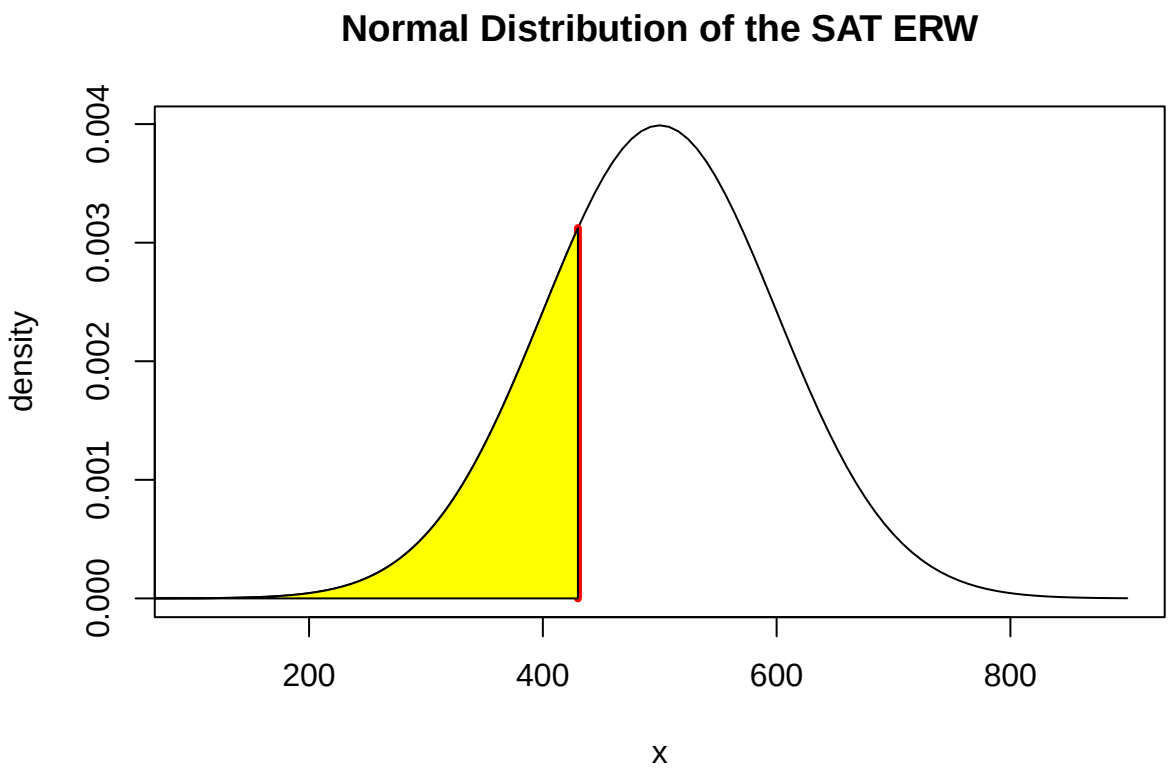


You can also shade in a portion of the graph using the polygon function. If we want to see visually what percentile of the population we did better than we could use the following code:

```
#draw the Normal distribution
curve( dnorm(x, mean=500, sd=100),
       xlim=c( 100, 900),  #pick your xlim by taking the mean - 4sd and mean +4sd
       main="Normal Distribution of the SAT ERW",
       ylab="density")

#draw segment for our value of note
segments( 430, 0,
          430, dnorm(430, mean=500, sd=100),
          #these are the coordinates (430,0) and (430, f(430) )
          col = "red",
          lwd = 4)

#use polygon function to shade in area to the left of segment
polygon(x=c(1,
            seq(from=1, to=430, by=1),
            430),
        y=c(0,
            dnorm(seq(from=1, to=430, by=1),mean=500, sd=100),
            0),
        col="yellow")
```



Your turn: Graph the Normal distribution of the Delorean's accelerations. Make sure you adjust the x-axis appropriately. Draw a segment at the speed where the DeLorean achieves time travel.

pnorm The `p--` functions are used to get values, not graphs. We can use them to tell us *the probability of getting a value less than or equal to that value*. This means it tells us the percentile of that value. We give the function the value we want to know the percentile of, as well as the mean and sd of the distribution, and pnorm will give us back the percentile (aka what percent of the data is less than or equal to the value).

The following code tells us what percentile a score of 430 on the SAT ERW section is.

```
pnorm( 430, mean = 500, sd = 100)
```

```
## [1] 0.2419637
```

Your turn: What percent of scores are less than or equal to a 550?

```
pnorm(550, 500, 100)
```

```
## [1] 0.6914625
```

Many times we will want to find the probability of getting something between two numbers. The easiest way to do this is by taking the difference in pnorms. For instance, if I want to determine the likelihood of getting scores between 550 and 640 then I would do the following:

Note that this is the same as asking the question "what percent of SAT ERW scores are between 550 and 640?"

```
pnorm( 640, mean = 500, sd = 100) - pnorm( 550, mean = 500, sd = 100)
```

```
## [1] 0.2277809
```

Your turn: What percent of SAT ERW scores are between a 600 and 650?

```
1 - pnorm(700, 500, 100)
```

```
## [1] 0.02275013
```

qnorm The quantile function qnorm for the Normal distribution is used for finding specific value in the distribution, given a percentile. For instance, if I wanted to find what score I would need to get to do better than 80% of SAT ERW test takers I would do the following:

```
qnorm(0.8, mean=500, sd=100)
```

```
## [1] 584.1621
```

Your turn: What score do you need to earn to do better than 30% of SAT ERW test takers?

Your turn: Find the score that represents the 99th percentile:

rnorm This is the simplest function. rnorm just generates random values that follow the given Normal distribution. If we want to simulate the test scores of 300 random students who took the SAT ERW we would do the following:

```
rnorm(300, mean=500, sd=100)
```

```
##   [1] 567.9606 528.1517 535.1211 453.6613 590.7944 313.3533 458.2955 522.1476
##   [9] 460.3933 538.3555 462.8646 592.5170 412.9581 647.2082 372.6261 479.8862
##  [17] 592.8948 525.0594 356.0773 654.7901 617.2994 535.3368 520.5853 530.4533
##  [25] 441.5206 606.9439 530.0002 523.2893 479.1646 489.8625 320.3831 480.5738
##  [33] 383.2284 428.3738 498.1606 586.1015 457.1466 624.5538 292.0654 464.5104
##  [41] 539.2624 527.4597 547.1193 719.7582 498.4245 524.9041 461.9525 492.5914
##  [49] 407.1453 427.8088 519.2000 346.8963 765.7463 559.2105 334.3745 469.9297
##  [57] 510.8272 527.0372 580.3428 464.8414 612.3745 553.9493 346.0053 606.5627
##  [65] 391.6668 536.8868 510.1753 475.5629 535.2320 330.7504 390.2499 535.6833
##  [73] 603.4181 226.0269 590.4446 393.2631 422.9264 404.4684 456.3636 440.8867
##  [81] 591.5868 544.0081 323.3923 597.0286 453.8913 414.2589 538.0410 454.0430
##  [89] 616.7352 497.7990 479.4633 309.9934 486.8805 574.7785 602.1636 494.4137
##  [97] 392.2589 289.0215 402.6938 414.2819 414.9384 612.1129 326.2279 669.3286
## [105] 428.0680 614.9776 351.3155 466.6728 550.6167 484.9313 641.5821 439.9196
## [113] 670.4297 521.6470 562.7301 537.4667 459.2315 577.5940 418.5488 431.5462
## [121] 385.6069 595.0885 282.7707 563.4253 394.1058 426.5499 411.2162 720.2118
## [129] 513.2455 377.9763 601.5138 526.8368 525.4769 648.8699 742.5929 380.3188
## [137] 623.7539 546.5970 747.6698 361.2795 516.6616 478.3329 544.1310 534.4830
## [145] 481.9018 368.3613 537.4142 494.6996 556.5233 536.1362 331.3570 569.6220
## [153] 631.5525 499.5968 415.7082 337.1607 538.0865 360.6091 355.0038 710.3280
## [161] 512.0996 531.8368 419.8395 629.7210 613.7429 508.4552 572.8844 658.5149
## [169] 558.4377 418.6325 424.4158 374.4400 381.3272 537.5649 422.6947 333.1237
## [177] 490.9431 634.7093 593.4443 601.3196 504.7430 342.0887 668.1363 405.5796
## [185] 325.6045 337.2797 467.5034 355.3374 533.8650 364.8176 470.6827 559.7078
## [193] 510.1433 613.8961 474.0519 480.5221 520.9220 568.6342 710.7433 394.4221
## [201] 479.7874 536.7679 342.4826 494.7654 652.9047 585.6211 672.1744 505.1237
## [209] 367.7645 537.0138 351.9071 447.7176 500.3846 466.9727 409.6567 531.1193
## [217] 582.9923 584.4514 227.8206 638.3830 306.8314 620.0529 493.4263 511.7274
## [225] 443.8314 557.2213 588.9434 520.1820 463.1049 558.8874 531.0920 545.4167
## [233] 564.3076 463.6097 434.2810 579.9097 418.1459 407.9597 658.0977 473.0639
## [241] 608.3135 589.3018 333.8740 760.6367 520.8298 520.8648 521.0364 430.8979
## [249] 472.1254 358.8795 633.1461 460.1165 513.8636 349.2753 423.1291 544.9360
## [257] 550.8564 525.2493 621.6988 576.5302 542.4614 408.0029 496.6900 546.8898
## [265] 316.9074 486.6342 470.7899 401.2715 314.7176 487.6806 612.0165 524.5080
## [273] 525.5179 445.8662 397.5380 438.4145 656.6038 566.5664 517.5136 525.3635
## [281] 608.1336 543.8564 447.0235 377.0969 369.9573 506.9231 586.0081 540.3546
```

```
## [289] 530.1541 401.0750 555.7475 730.8254 500.1199 320.3057 677.6340 416.2109
## [297] 461.5893 531.8148 543.9577 352.8558
```

Just in case you are interested we can make the scores look like real SAT scores.

```
round( rnorm(300, mean=500, sd=100), -1)
```

```
##   [1] 350 470 560 400 390 370 440 510 340 440 450 520 660 490 210 390 470 570
##  [19] 480 600 370 560 420 650 400 540 390 500 650 490 540 520 380 500 530 460
##  [37] 550 420 620 580 420 470 440 470 270 450 540 430 380 420 410 360 330 530
##  [55] 350 490 640 580 370 460 430 520 490 580 560 580 620 520 540 450 510 470
##  [73] 420 490 530 390 470 470 580 550 480 380 550 390 600 490 580 530 690 440
##  [91] 490 470 620 610 240 380 450 390 430 540 420 370 500 330 500 470 440 500
## [109] 610 570 290 370 560 430 400 450 530 650 550 600 620 550 380 570 450 490
## [127] 370 540 510 390 530 620 400 610 410 460 460 490 420 580 580 510 380 530
## [145] 560 410 520 370 510 570 540 440 490 670 420 450 540 480 700 410 490 540
## [163] 550 380 700 540 520 460 500 590 640 660 650 670 440 670 500 420 530 520
## [181] 490 540 510 560 530 530 460 590 590 660 410 700 580 590 590 450 440 560
## [199] 550 390 530 380 540 400 480 400 570 560 460 610 670 440 610 590 460 480
## [217] 360 470 460 590 540 420 560 490 380 420 510 560 760 610 590 450 510 330
## [235] 630 620 520 530 740 540 450 580 470 520 540 460 440 560 230 340 570 570
## [253] 420 580 430 360 300 500 480 310 560 360 390 570 450 440 450 400 730 490
## [271] 600 260 460 660 420 480 450 390 430 440 600 480 480 660 470 380 410 380
## [289] 370 500 350 530 390 420 640 440 380 480 410 590
```

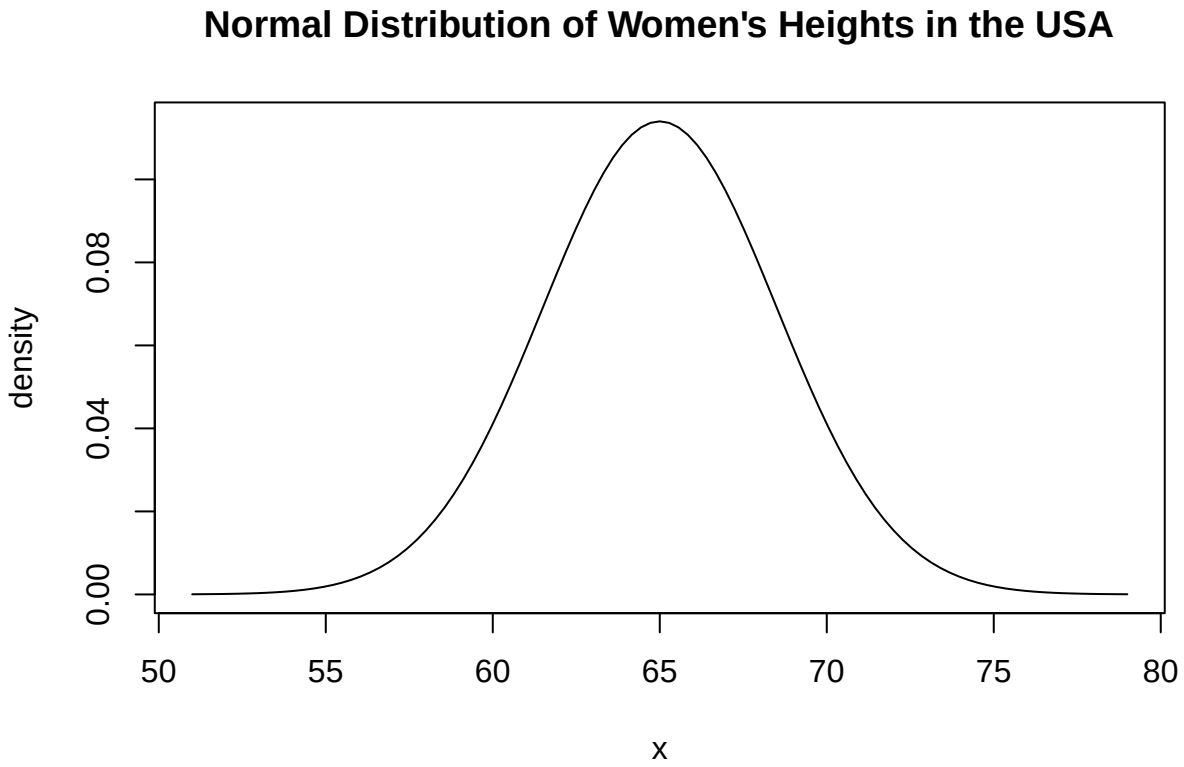
Practice Exercises

First, complete the Back to the Future worksheet, using R to do the calculations.

Then, for these next problems use the following information. The average height of a woman in the United States is 65 inches with a standard deviation of 3.5 inches. Assume that the distribution of heights of women follows the normal distribution N(65, 3.5).

Create a graph that shows the relative likelihoods of womens heights in the USA.

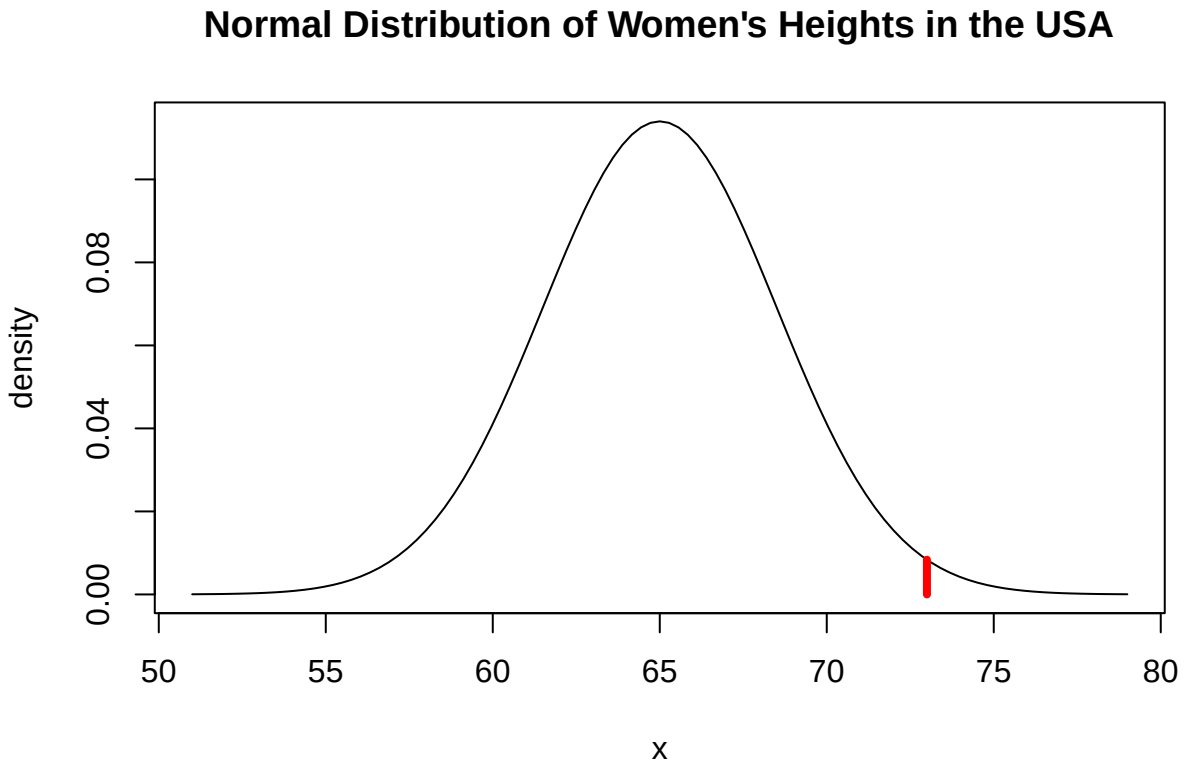
```
curve(dnorm(x, mean=65, sd=3.5),
      xlim=c(51, 79),
      main="Normal Distribution of Women's Heights in the USA",
      ylab="density")
```



What if we want to see better what this graph says about the relative likelihood of being 73 inches tall? Redraw the graph with a red line indicating the height 73 inches tall.

```
curve(dnorm(x, mean=65, sd=3.5),
      xlim=c(51, 79),
      main="Normal Distribution of Women's Heights in the USA",
      ylab="density")

#draw segment for our value of note
segments( 73, 0,
          73, dnorm(73, mean=65, sd=3.5),
          col = "red",
          lwd = 4)
```



What percent of US women are less than 73 inches tall?

```
pnorm(73, mean=65, sd=3.5)
```

```
## [1] 0.9888645
```

What proportion of adult women in the USA are between 63 and 70 inches tall?

```
pnorm(70, mean=65, sd=3.5) - pnorm(63, mean=65, sd=3.5)
```

```
## [1] 0.6395817
```

How tall do you have to be so that you are taller than 90% of women in the USA?

```
qnorm(0.9, mean=65, sd=3.5)
```

```
## [1] 69.48543
```

Generate a list of 100 numbers that could represent the heights of 100 adult women in the USA.

```
rnorm(100, mean=65, sd=3.5)
```

```
##   [1] 64.96771 67.43924 61.79187 67.08458 69.25607 63.48244 64.80386 62.96203
##   [9] 63.82058 73.39871 63.89950 66.04592 62.11398 64.79420 65.19160 70.64619
##  [17] 60.57808 69.27744 70.42006 66.64016 67.66082 65.74354 60.69081 64.69401
##  [25] 61.98402 66.17574 63.97933 65.41020 65.07777 61.12674 64.11331 68.19325
##  [33] 67.84233 62.21611 63.71136 70.67575 63.02633 63.96613 59.41820 63.94517
##  [41] 62.91004 61.58375 55.38122 67.37870 63.74492 65.42522 69.65629 72.53620
##  [49] 67.73179 61.37745 61.90327 59.42020 65.56362 72.81291 66.37674 63.73175
##  [57] 58.28360 71.16481 63.71669 63.41842 62.36197 63.09014 64.26359 73.25794
##  [65] 60.56973 63.37950 57.02120 66.56563 63.76248 64.64507 63.97020 66.54840
##  [73] 62.04324 59.79034 64.19167 64.09524 68.96209 65.45767 64.82590 63.90926
##  [81] 67.70662 63.37057 66.15913 67.56098 64.06682 65.49252 65.20684 61.78696
##  [89] 62.91718 61.81178 70.61599 63.74748 64.91337 63.63019 66.81233 61.53928
```

```
##  [97] 65.91816 66.49831 59.79582 69.20320
```