STAT 211: Normal Distribution

Computing probabilities and quantiles from this important distribution

Darren Homrighausen

```
Loading required package: ggplot2
```

Joint Distributions: Continuous

Probability for bivariate pdf's

Example:

Consider the joint pdf:

$$f_{X,Y}(x,y) = \begin{cases} \frac{12}{7}(x^2 + xy) & \text{for } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Find $P(X \leq 0.6 \text{ and } Y \leq 0.4)$

Computing probabilities: old vs. new

In times of yore, this meant using tables:

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA t

Z	.00	.01	.02	.03	.04	.05	.06
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.5239
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.5635
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.6025
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.6405
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.6772
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.7122
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.7453
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.7763
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.8051
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.8314
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.8554
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.8769
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.8961
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.9130
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.9278
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.9406
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.9515
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.9608
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.9685
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.9750
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.9803
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.9846
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.9880
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.9908
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.9930
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.9947
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Now, we can just use a computer:

```r
pnorm(1.8,0,1)
```

```
[1] 0.9640697
```

```r
pnorm(1.89,0,1)
```

```
[1] 0.970621
```

Computing Probabilities and Values

Computing probabilities

We will always write $X \sim N(\mu, \sigma^2)$ for the normal

R syntax will be specified with μ and $\sigma = \sqrt{\sigma^2}$ instead

Example

If $X \sim N(5, 9)$, then the probability X is less than 4 is:

```r
x       = 4
mu      = 5
sigmaSq = 9
pnorm(x, mu, sqrt(sigmaSq))
```

```
[1] 0.3694413
```

Computing values (quantiles)

Continuing with $X \sim N(5, 9)$:

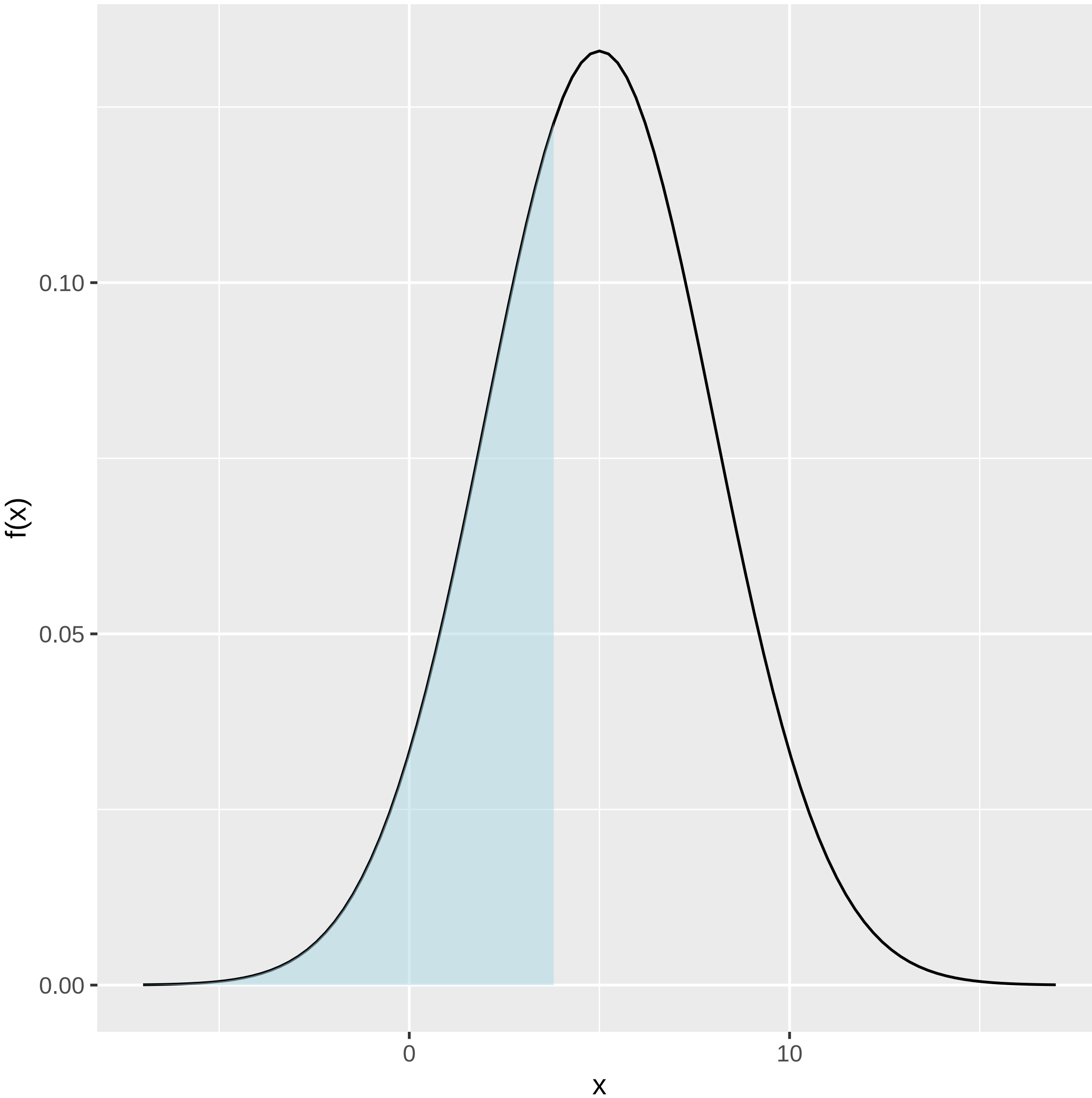
(that is, normal with expected value 5 and variance 9)

Recall:

For a value $x = 4$, $P(X \leq 4)$:

```r
pnorm(4,5,3)
```

```
[1] 0.3694413
```



Now, let's ask the **opposite** question!

For a fixed probability p, what is the value x so that $P(X \leq x) = p$?

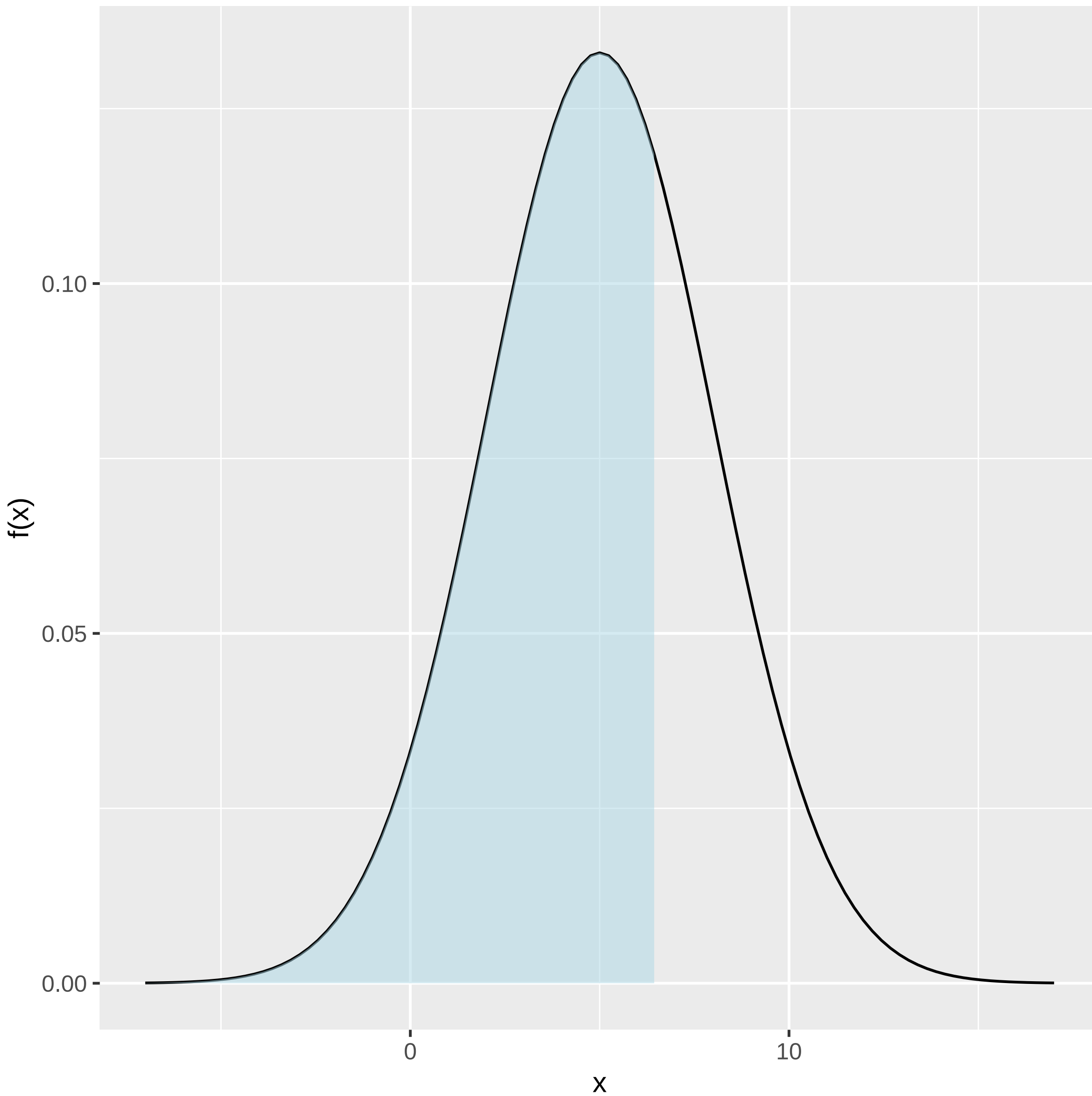
This value is called a quantile

Example:

What x makes $P(X \leq x) = .7$?

```r
qnorm(.7,5,3)
```

```
[1] 6.573202
```



Using pnorm and qnorm

Concept Check: What would each of the following codes produce?

```r
qnorm( pnorm(3,5,3), 5, 3)
```

. . .

```
[1] 3
```

. . .

```r
pnorm( qnorm(.2,5,3), 5, 3)
```

. . .

```
[1] 0.2
```

Computing probabilities and values

Suppose $X \sim N(-3.2, 100)$

that is, normal with expected value -3.2 and variance 100

Probability: What is the probability X equals 0?

. . .

This is always zero!

. . .

Probability: What is the probability X is less than 0?

. . .

```r
pnorm(0,-3.2,sqrt(100))
```

```
[1] 0.6255158
```

. . .

Values: at what value x does $P(X \leq x) = .05$?

. . .

```
qnorm(.05,-3.2,sqrt(100))
```

```
[1] -19.64854
```

Computing probabilities: SATs

Example

The scores on the SAT math section are $X \sim N(520, 3600)$

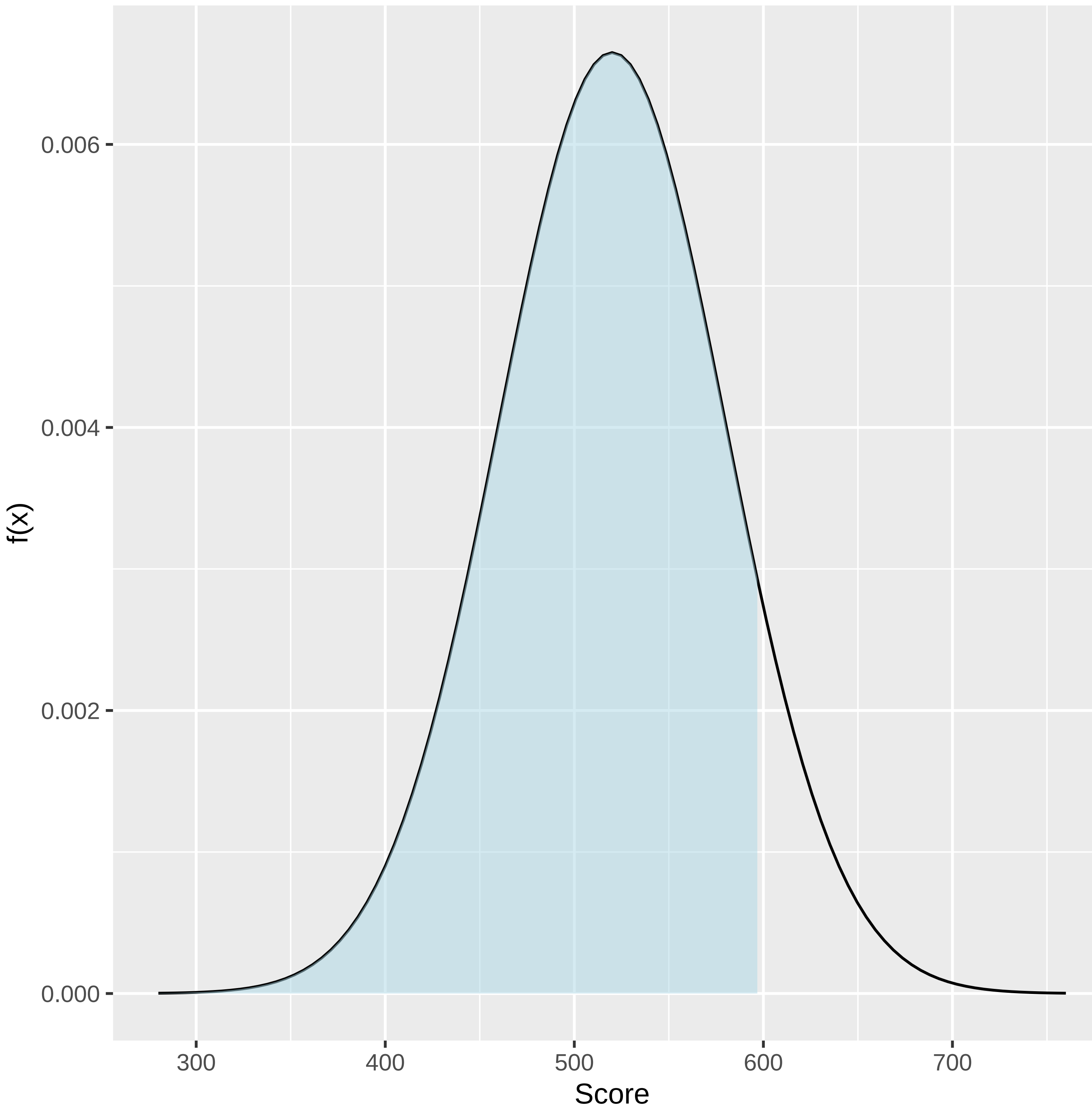
(that is, normal with expected value 520 and variance 3600)

What is the probability someone scores less than 600?

$P(X \leq 600) =$ what probability?

```
x       = 600
mu      = 520
sigmaSq = 3600
pnorm(x, mu, sqrt(sigmaSq))
```

```
[1] 0.9087888
```


0.006
0.004
0.002
0.000
f(x)
300
400
500
600
700
Score

Computing values: SATs

Example

The scores on the SAT math section are $X \sim N(520, 3600)$

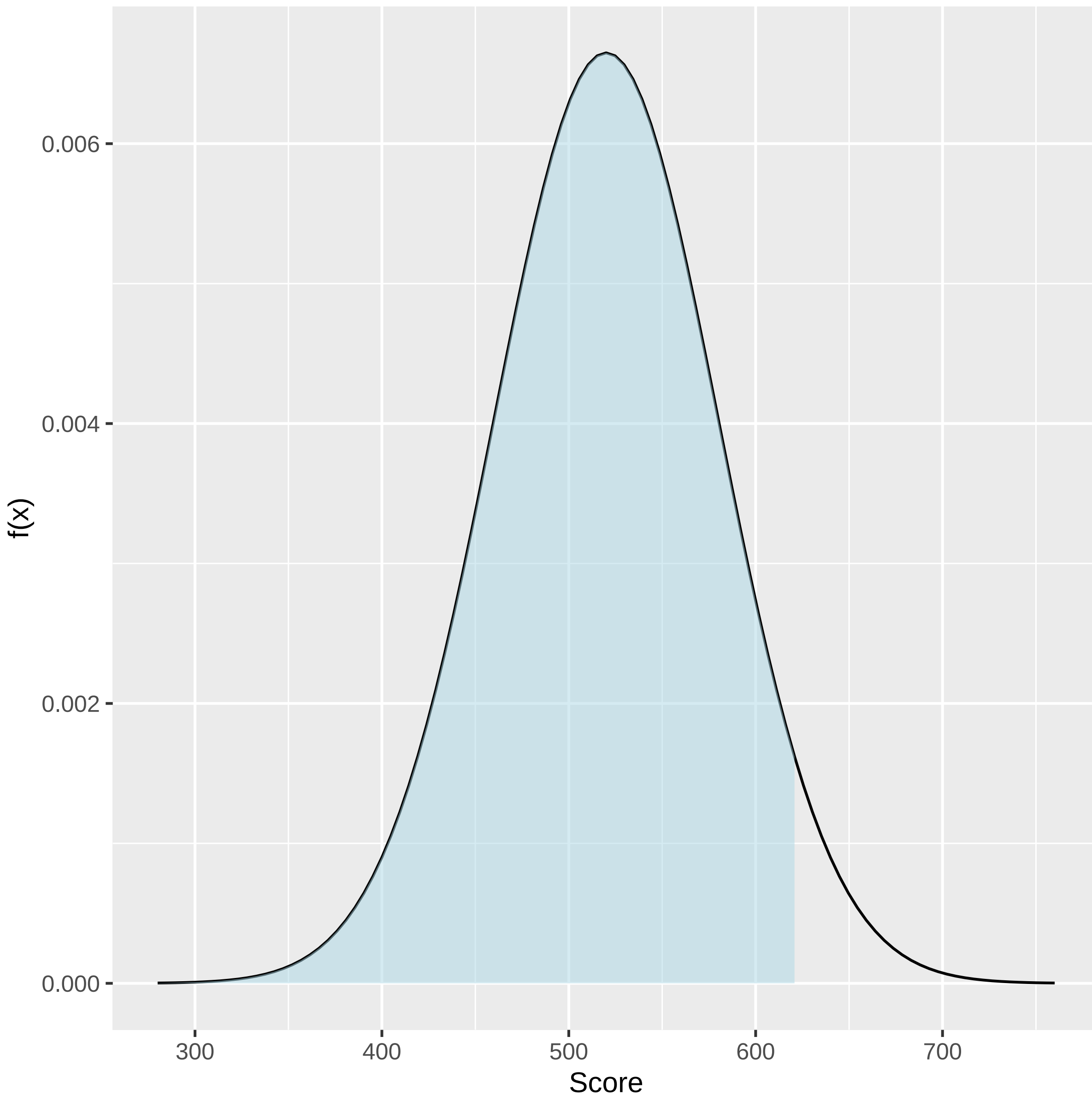
The stats dept. admits students scoring above 96^{th} percentile

What is the cutoff score (x) for recruitment by the stats dept.?

$P(X \leq x) = 0.96$

```
prob     = .96
mu       = 520
sigmaSq = 3600
qnorm(prob, mu, sqrt(sigmaSq))
```

```
[1] 625.0412
```



Transformations

Transforming Normals: $aX + b$

Suppose again $X \sim N(\mu, \sigma^2)$

If we multiply/add constants a, b to X to form a new rv Y:

$$Y = aX + b$$

then $Y \sim N(a\mu + b, a^2\sigma^2)$

(This should remind you of our $E[X]$ and $Var[X]$ results)

Example:

If $X \sim N(150, 100)$ then

$$Y = 3X + 30 \sim N(480, 900)$$

Transforming Normals: inches to cm

We are measuring a machined part using a ruler on a .01" scale

If several people measure, they will get some random amount of measurement error:

$$X \sim N(.875, .00005)$$

What is the probability that that a measurement will be within $\pm.01$ centimeters?

(An inch is 2.54 cm)

$Y = 2.54X$, therefore $Y \sim N(2.54 * .875, 2.54^2 * .00005) = N(2.2225, 0.00032)$

$$\begin{aligned} P(2.2225 - .01 \leq Y \leq 2.2225 + .01) &= \quad (2) \\ = P(2.2125 \leq Y \leq 2.2325) & \quad (3) \\ = F(2.2325) - F(2.2125) & \quad (4) \end{aligned}$$

```r
mu      = 2.54*.875
sigmaSq = 2.54^2*.00005
pnorm(mu + .01, mu, sqrt(sigmaSq)) -
    pnorm(mu - .01, mu, sqrt(sigmaSq))
```

```
[1] 0.4223202
```

The Standard Normal Distribution

When $\mu = 0$ and $\sigma = 1$, we call it the standard normal

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \ -\infty < x < \infty.$$

(Once again, you don't need to know this formula for this class)

The standard normal will be notated as: $Z \sim N(0, 1)$

Standardization

Going from a general normal to a standard normal is known as standardization

$$X \sim N(\mu, \sigma^2) \longrightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

The opposite (actually, the inverse) is true

$$Z \sim N(0, 1) \longrightarrow X = Z\sigma + \mu \sim N(\mu, \sigma^2)$$

(this is using the results in Transforming Normals)

Z-scores: How to compare values

Z-scores: Comparing values

The z-score can be computed:

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

(For a value x from a distribution with mean μ and variance σ^2)

Question: Which of the two values from two different distributions is more unusual?

Answer: Whichever z-score has larger magnitude

(that is, largest |z-score|)

Example: We want to compare olympic records in men's and women's sprinting. Which is more unusual?

Let's look at a table:

Table 1: Table of Average and Standard Deviation of Olympic Sprinting times.

Category	Mens	Womens
average	9.85	10.83
variance	.0057	.0049
record	9.63	10.61

Let's compute the two z-scores:

```
zScoreM = (9.63 - 9.85)/sqrt(.0057)
zScoreM
```

```
[1] -2.913971
```

```
zScoreW = (10.61 - 10.83)/sqrt(.0049)
zScoreW
```

```
[1] -3.142857
```

The women's result is more unusual

The z-score for men is -2.914 while for women it is -3.143

Z-scores: Another application

Another use of z-scores: what value x would be equally as unusual?

Answer: Choose the x so that the z-scores are equal

Example: What time for a men's sprinter would be equivalent to the female record?

Table 2: Table of Average and Standard Deviation of Olympic Sprinting times.

Category	Mens	Womens
average	9.85	10.83
variance	.0057	.0049
record	9.63	10.61

We found the female z-score $= \frac{10.61-10.83}{\sqrt{.0049}}$ = -3.143

To find the equivalent time, we need to unstandardize

$$\text{z-score} * \sigma + \mu = x$$

Here, use

- z-score from the female sprinters
- the μ and σ from the male sprinters ($\mu = 9.85$, $\sigma^2 = .0057$)

```
x = zScoreW * sqrt(.0057) + 9.85
x
```

```
[1] 9.612719
```

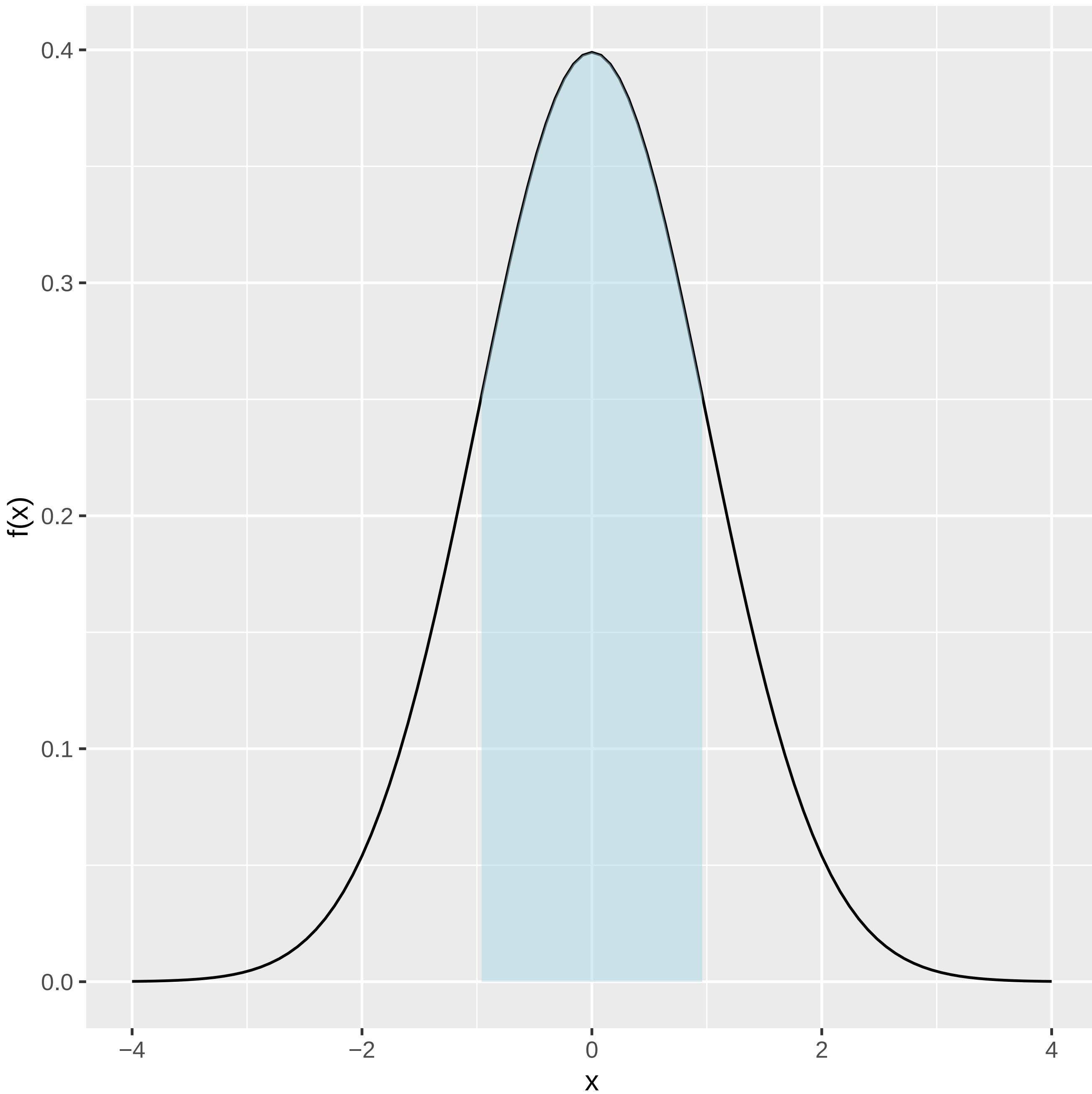
The male record would need to be 9.613 seconds to be as unusual as the female record.

Empirical rule

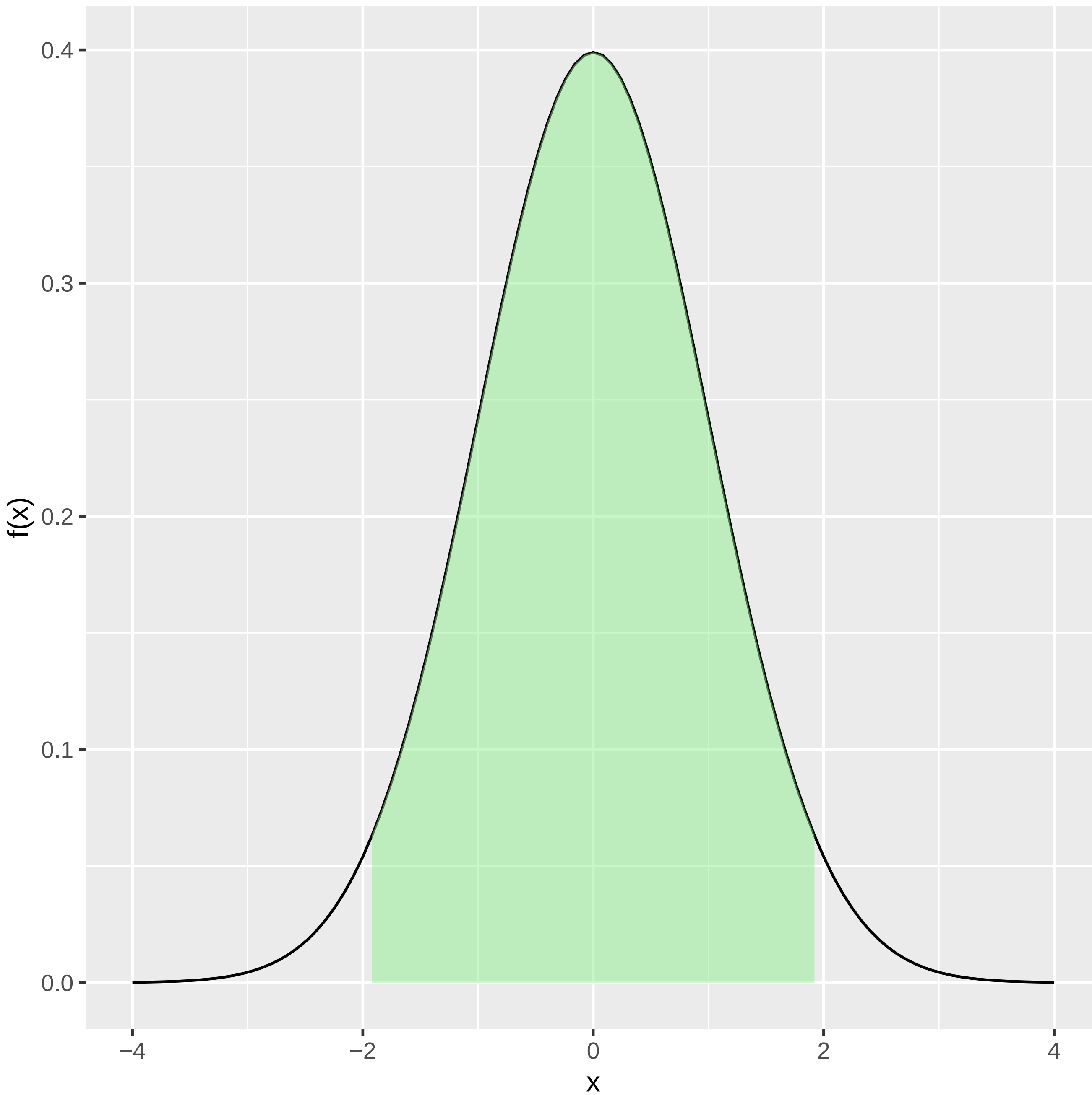
Values from normal distributions contain predictable amounts of probability

This is the empirical rule:

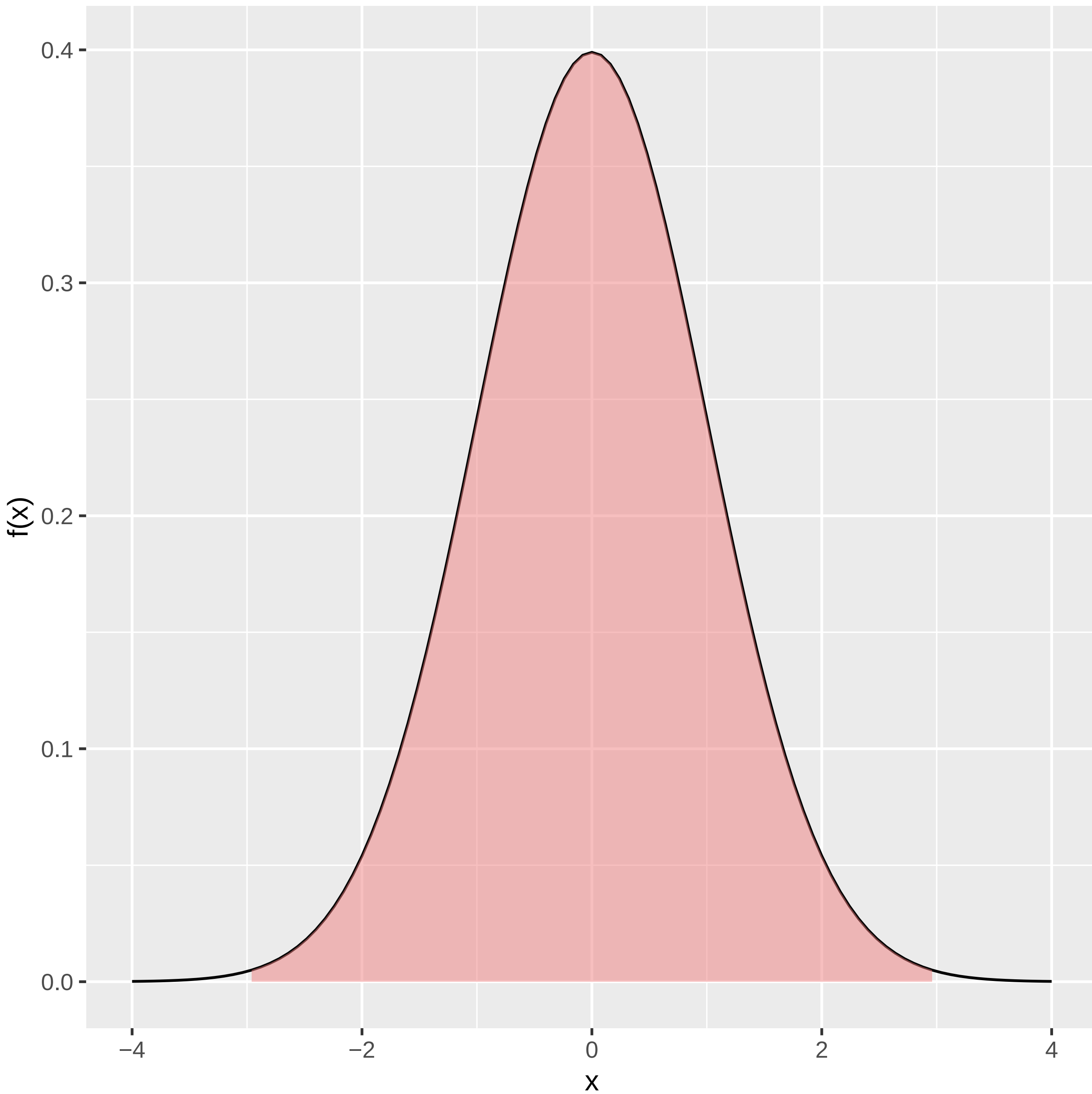
- $P(|Z| \leq 1) = 0.683$ ($\sim \frac{2}{3}$ obs. within 1 σ)
- $P(|Z| \leq 2) = 0.954$ ($\sim \frac{19}{20}$ obs. within 2 σ)
- $P(|Z| \leq 3) = 0.997$ ($\sim \frac{99}{100}$ obs. within 3 σ)



$P(|Z| \leq 1) = 0.68$



$P(|Z| \leq 2) = 0.95$

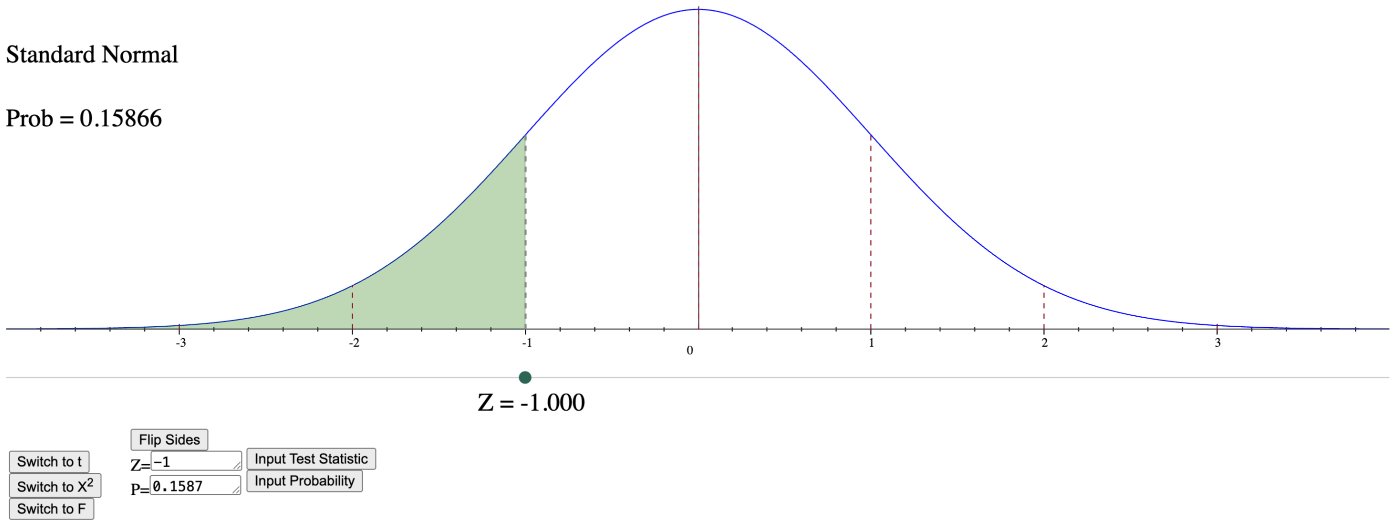


$P(|Z| \leq 3) = 0.997$

A (potentially) helpful applet

On exams, there will be a link to an applet: Probability applet

Note: This applet is totally optional and unnecessary if you use R



Returning to SAT example:

```
x       = 600
mu      = 520
sigmaSq = 3600
pnorm(x, mu, sqrt(sigmaSq))
```

```
[1] 0.9087888
```

```
qnorm(.96, mu, sqrt(sigmaSq))
```

```
[1] 625.0412
```

To use the applet

- Probability: we need the z-score = 1.3333333
- Value: We need to get standardized z-score from applet and then unstandardize to get value.

```
1.751*sqrt(sigmaSq) + mu
```

```
[1] 625.06
```

(here, 1.751 came from the applet by using the probability .96 from SAT example)

We'll be back...

The normal plays a central role in probability and statistics

We will return to it later during sampling distributions...