Random Variables, examples

CSU, Chico Math 314

2018-09-10

outline

Probability, Uniform Distribution

What's the probability of rolling one die greater than 3?

- What's the probability mass function of rolling dice?
- We want to calculate $P(X > 3)$.

Probability, Uniform Distribution

Let $X \sim U(1,6)$. Find $P(X > 3)$.

$$P(X > 3) = \sum_{x>3} P(X = x) = 3/6.$$

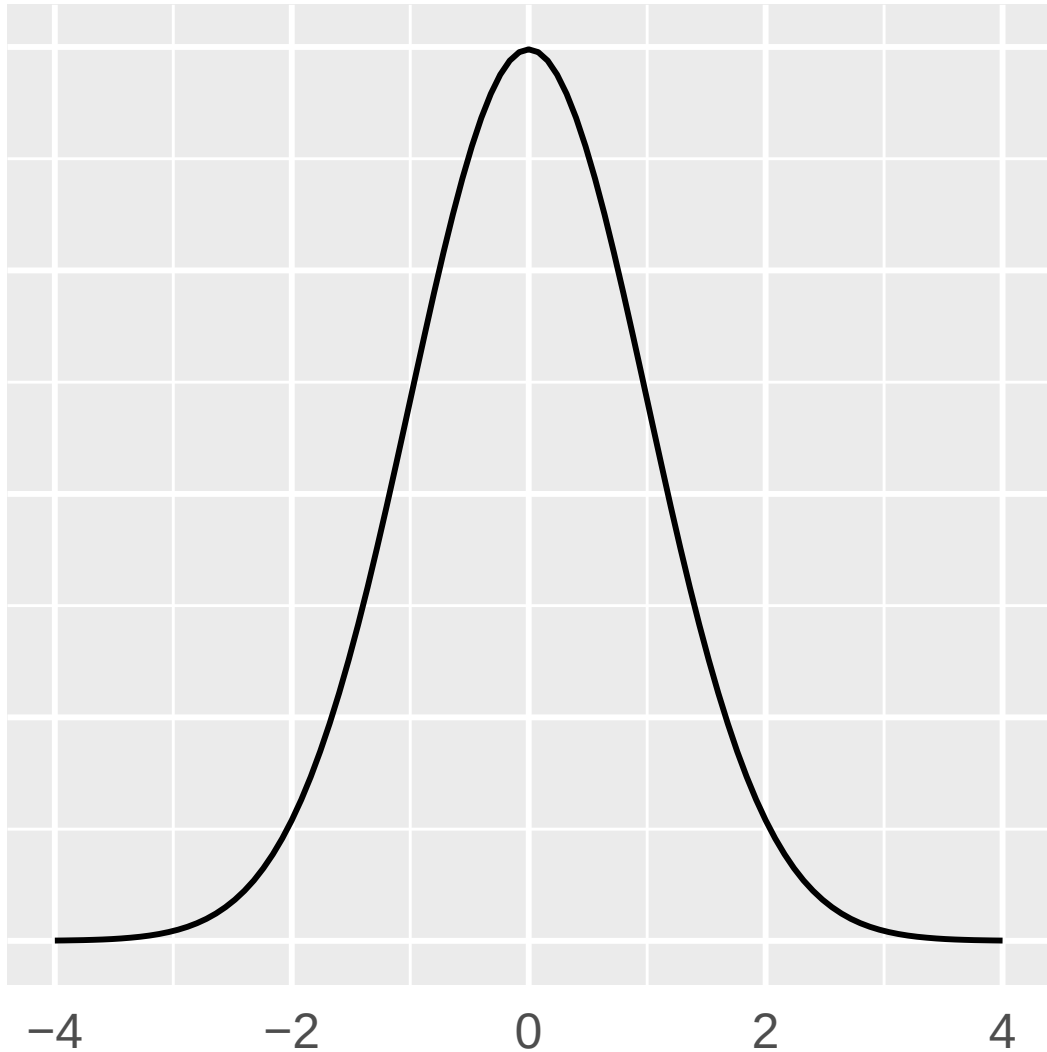
Discrete CDF

If the random variable X has the following cumulative distribution function, calculate the PMF.

x	1	2	3	4
F(x)	0.4	0.7	0.9	1.0

Normal Distribution

The **normal distribution** is ubiquitous in statistics.



Normal Distribution

The probability density function for the normal distribution is

$$f(x) = (2\sigma^2\pi)^{-1/2} \exp\left\{ \frac{-(x-\mu)^2}{2\sigma^2} \right\}.$$

Probability statements about the normal distribution take the form

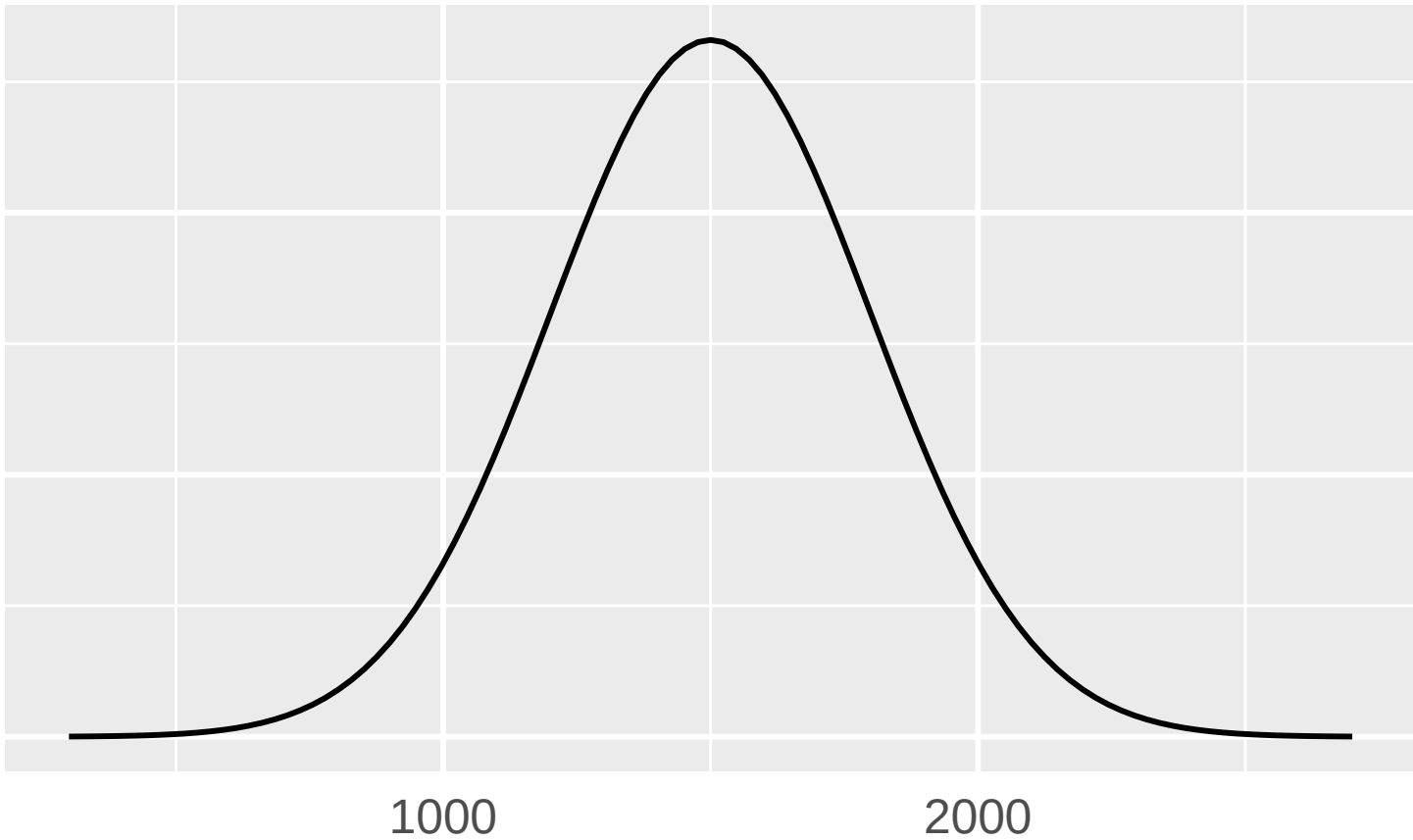
$$P(a < X < b) = \int_a^b f(x)dx.$$

Normal Distribution, example

A well known example of the normal distribution is SAT scores. Let $X \sim N(1500, 300^2)$ be the distribution of SAT scores with population mean $\mu = 1500$ and population variance $\sigma^2 = 300^2$.

Normal Distribution, example

SAT scores: $X \sim N(1500, 300^2)$.



Z-score

Z-score

The Z-score of an observation is the number of standard deviations the observation lies above or below the mean. We compute the Z-score for an observation x that follows a distribution with mean μ and standard deviation σ using

$$z = \frac{x - \mu}{\sigma}.$$

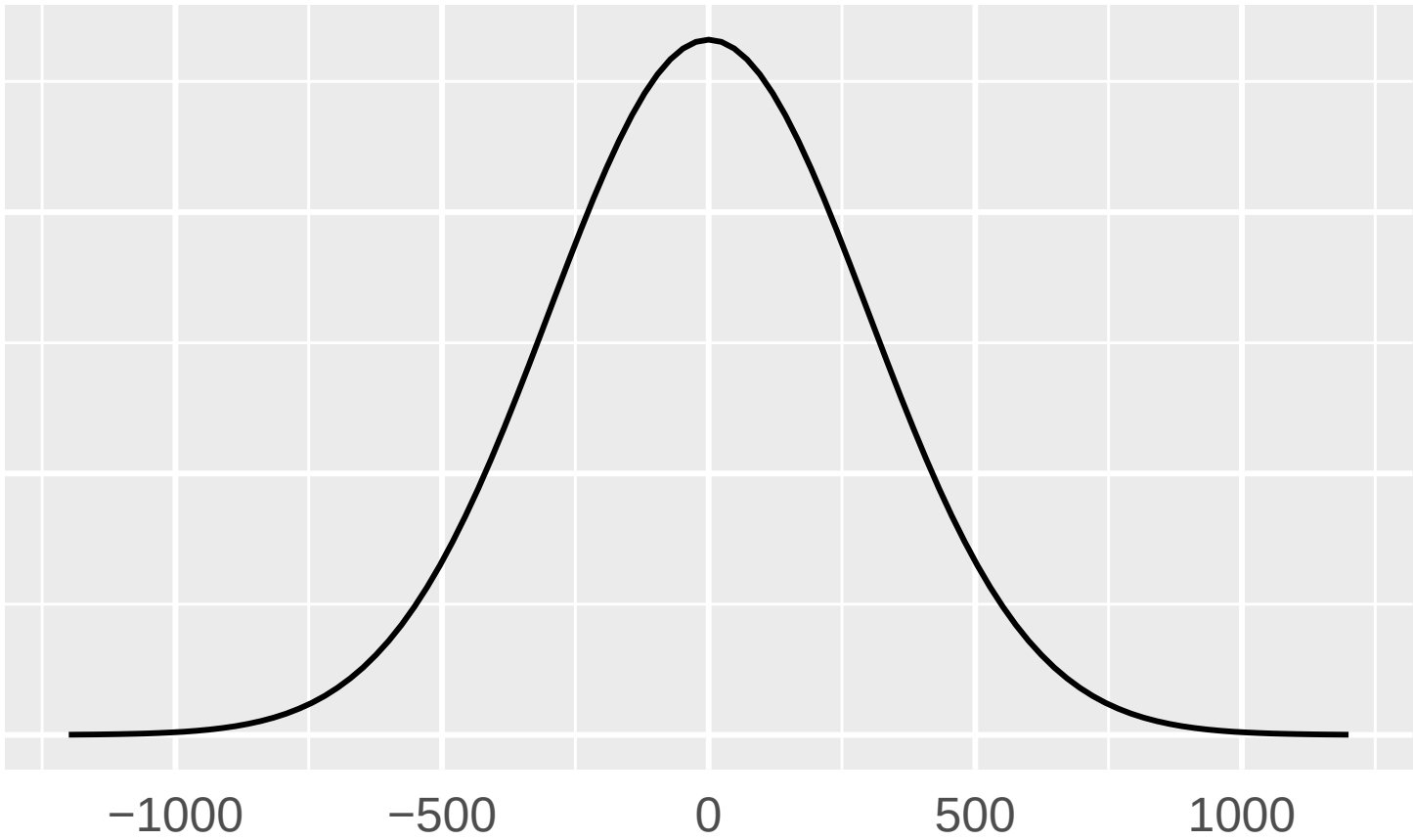
We'll often use the Z-score with data that are nearly normal, but this is a technique used for any distribution.

Z-score, notes

The Z-score simply shifts and scales the distribution so that z follows a $N(0, 1)$ distribution. Why?

Z-score, shift

If SAT scores, X, are distributed $N(1500, 300^2)$, then $X - \mu = X - 1500$ is distributed as $N(0, 300^2)$.



Z-score, scale

If $(X - 1500) \sim N(0, 300^2)$, then $\frac{x-1500}{300} = z \sim N(0, 1)$.

Z-score, notes

The Z-score is a unitless number. Why?

Z-score, idea

By definition, if an observation is 1 standard deviation above its mean, the Z-score is 1. If an observation is 1.5 standard deviations below the mean, its Z-score is -1.5.

Z-score, example

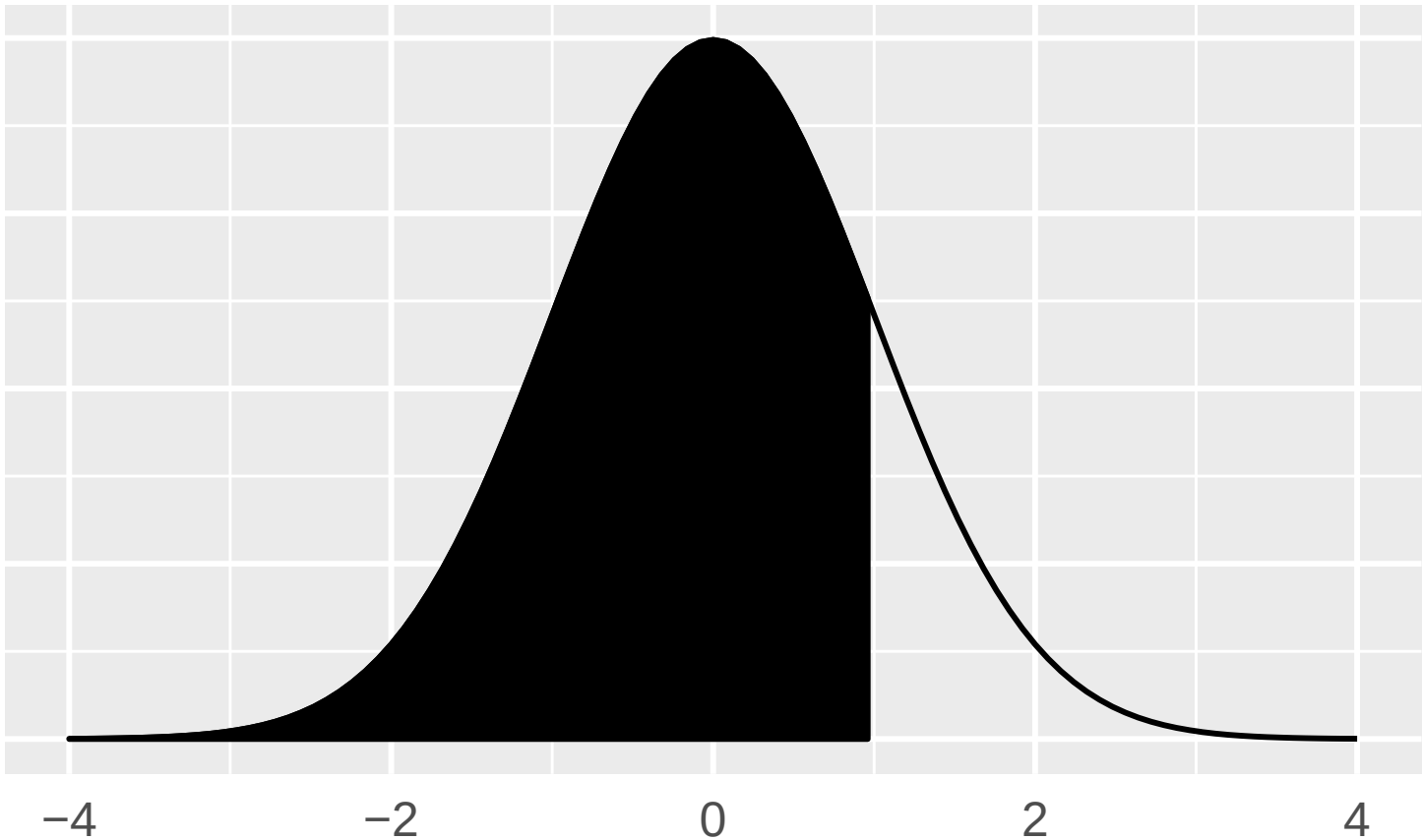
Suppose SAT scores follow a normal distribution, $X \sim N(1500, 300^2)$ and that Ann got 1800. What proportion of SAT takers did worse than Ann?

Steps

1. draw picture
2. Z-score: $z = \frac{1800-1500}{300} = 1$
3. $P(Z < 1)$?

Z-score, example

Picture of proportion of SAT takers that did worse than Ann: $P(Z < 1)$



Z-score, finding probability

Use R to find $P(Z < 1)$.

```
pnorm(1) # = P(Z<1)

## [1] 0.8413447
```

pnorm

The function `pnorm` in `R` calculates for us the area under the curve that's less than the value provided.

```
?pnorm
```

What is the stats name of this function?

Arithmetic with Densities

Since all the area under the normal curve is equal to one, we can easily manipulate the output of `pnorm` to find areas greater than some value, or in between two values.

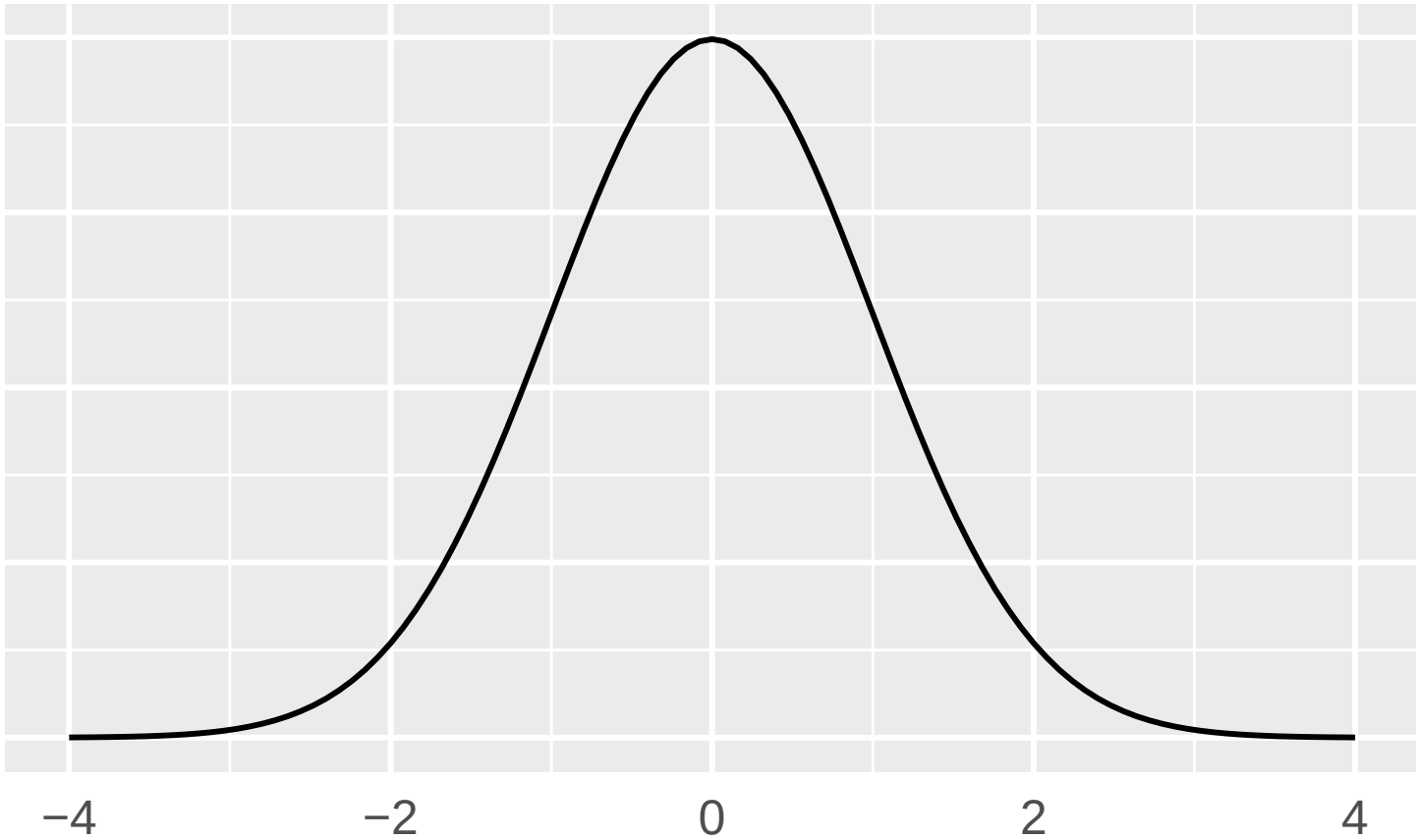
Z-score, example 2

Shannon is a randomly selected SAT taker, and nothing is known about Shannon's SAT aptitude. What is the probability Shannon scores at least 1630 on her SATs?

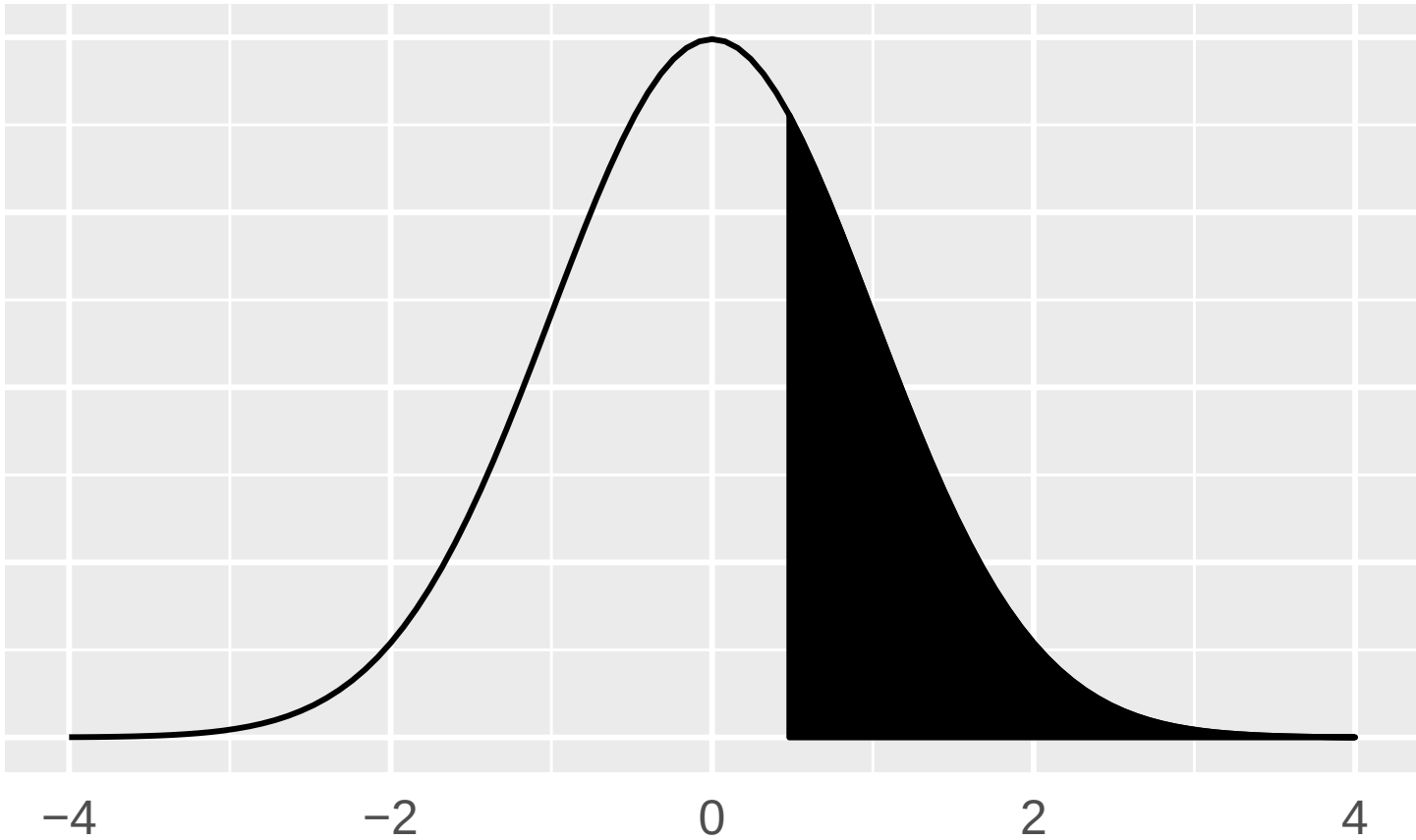
Steps

1. draw picture.
2. Z-score: $z =$
3. $P(Z > z)$?

Z-score, example 2



Z-score, example 2



Z-score, finding probability

We get $z = 0.43$. Use R to find $P(Z > 0.43)$.

```
1 - pnorm(0.43) # = P(Z>0.43)

## [1] 0.3335978
```

Z-score, example 3

What's the probability that Edward scores in between Ann and Shannon? Picture?

Z-score, finding probability

$P(0.43 < z < 1)$?

```
pnorm(1) - pnorm(0.43)
```

```
## [1] 0.1749426
```

This is R's version of

$$P(a < Y < b) = \int_a^b f_Y(y)dy = F(b) - F(a).$$

Z-score, another example

Because the $Z - score$ is unitless, we can compare the likelihood of different events even if they are measured in different units.

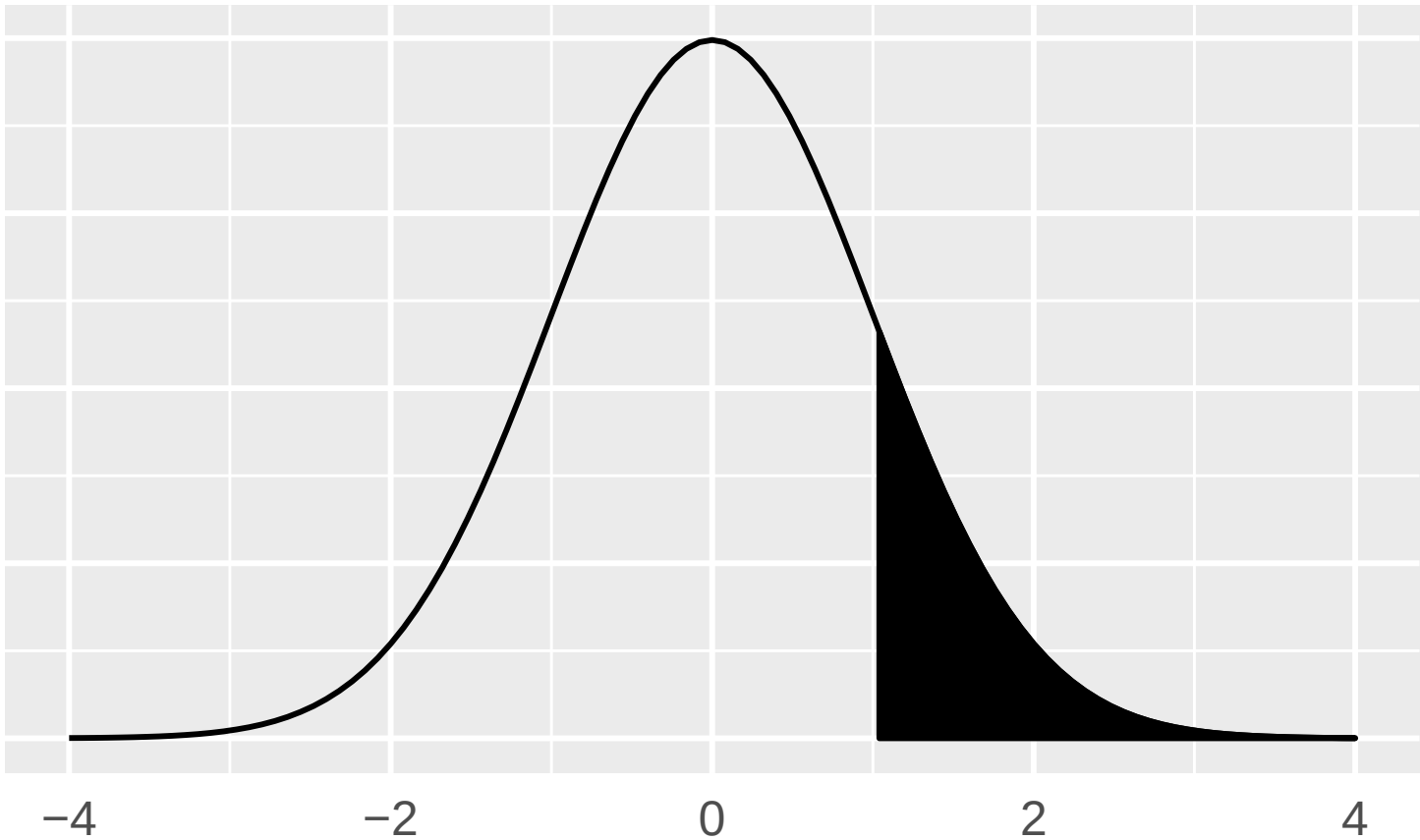
Z-score, another example

Suppose SAT scores are distributed $X \sim N(1500, 300^2)$ and ACT scores are distributed $N(21, 5^2)$. If Ann scored 1800 on the SAT and Tom scored 24 on the ACT, who performed better?

1. draw pictures.
2. Z-scores: $z_A = 1$ and $z_T = \frac{24-21}{5} = 0.6$.
3. Want to find $P(Z > 1)$ and $P(Z > 0.6)$.

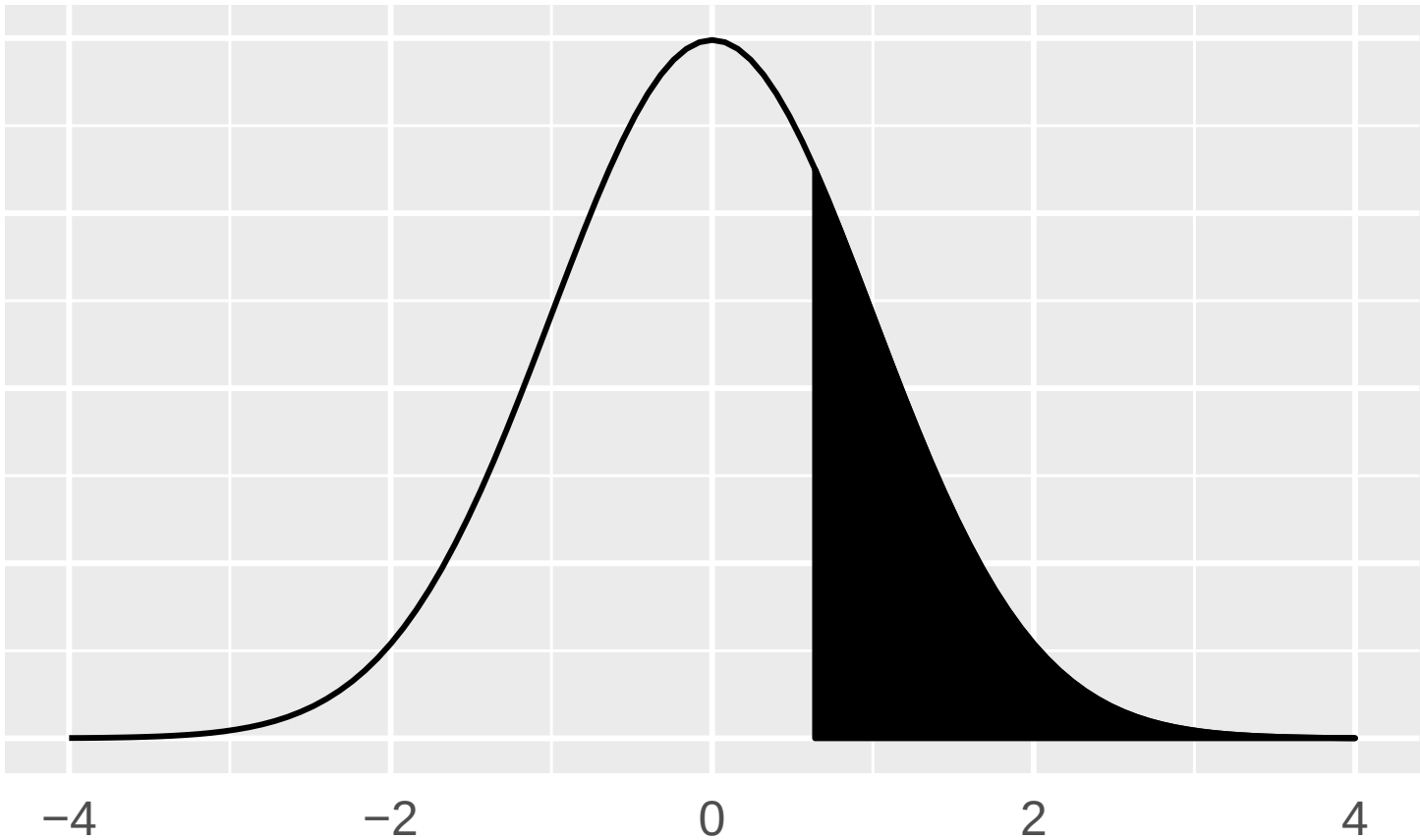
Z-score, example

Proportion who scored better than Ann



Z-score, example

Proportion who scored better than Tom



Z-score, finding probability

Use R to find $P(Z > 1)$ and $P(Z > 0.6)$.

```
1 - pnorm(1) # = P(Z>1), Ann
```

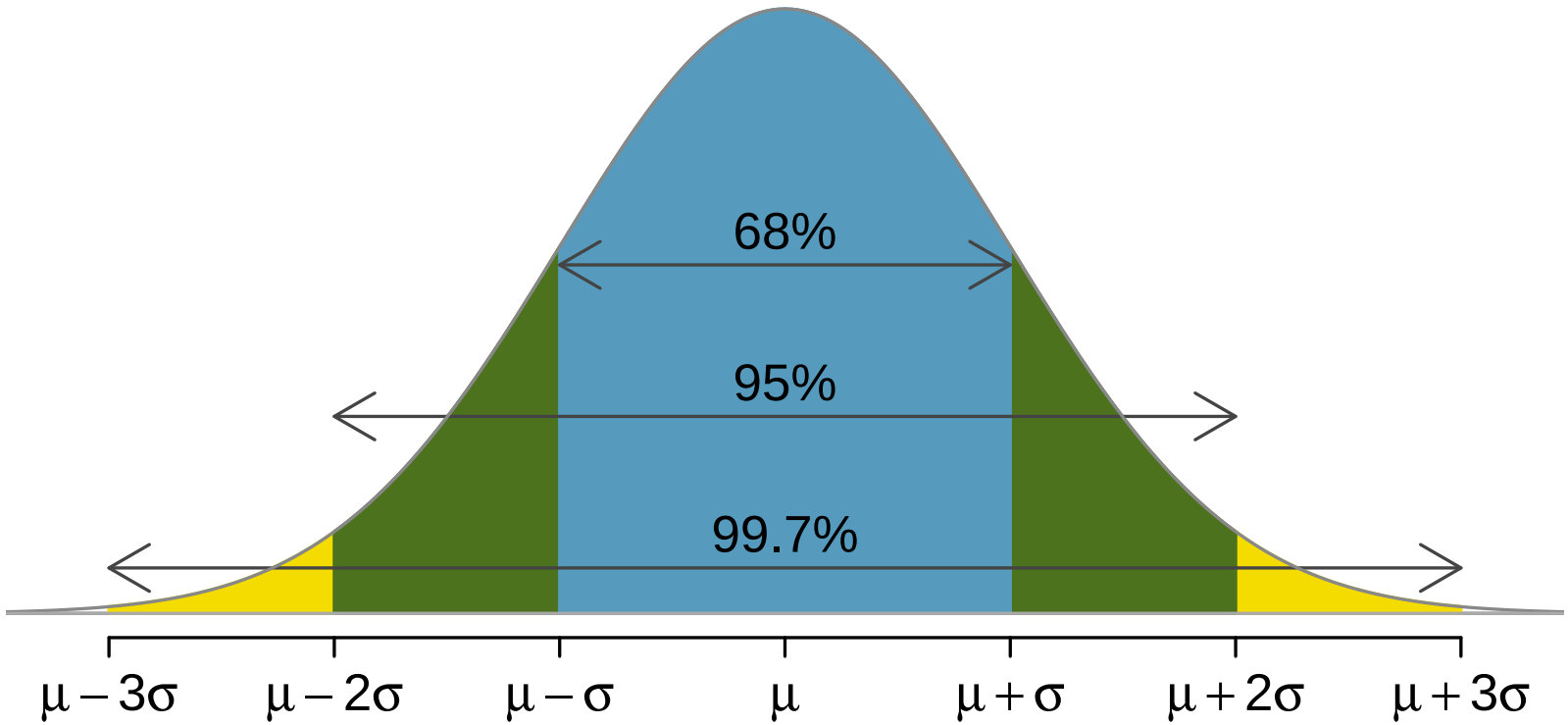
```
## [1] 0.1586553
```

```
1 - pnorm(0.6) # = P(Z>0.6), Tom
```

```
## [1] 0.2742531
```

Conclusion?

Normal Distribution, 68, 95, 99.7 Rule



Percentiles

Edward earned a 1400 on his SAT. What is his percentile?
Steps

1. draw picture.
2. Z-score: $z = (1400 - 1500)/300 = -0.33$.
3. $P(Z < -0.33)$?

Percentiles

The probability $P(Z < -0.33)$ can also be thought of as a percentile.

```
pnorm(-0.33)

## [1] 0.3707
```

Edward is at the 37th percentile.

Percentiles

We can essentially ask the previous questions backwards.

Percentiles, example

Atahualpa's SAT score is at the 90th percentile. What score did he receive?

Steps

1. Draw picture.
2. find Z-score.
3. solve for SAT score.

Percentiles, example

Find Z-score

```
qnorm(0.9) # finds Z-score z that puts 90% <

## [1] 1.281552
```

Percentiles, example

With $z = 1.28$, we can solve for Atahualpa's SAT score.

$$z = 1.28 = \frac{x - 1500}{300}.$$

qnorm

The function `qnorm` in `R` calculates for us the Z-score z that puts the supplied probability less than z.

```
?qnorm
```

Distribution Functions in R

Notice that all random variables in R follow the same naming convention

```
?pexp
?pnorm
?punif
?pbinom # to get the Bernoulli use size=1
```

Exponential Distribution, integrates to 1

Recall, all distribution functions sum or integrate to one. Let's try one. Calculate

$$\int_0^\infty \lambda \exp\left(-\lambda x\right) dx.$$

references I