The Normal Distribution!

Stat 120

October 12 2023

Overview

Core intro stats covered: *EDA for data comprehension, estimation with confidence, and hypothesis testing via p-values.*

Upcoming focus: *Advanced inference methods, transitioning from simulations to probability models for bootstrap/randomization distributions.*

Density Curve

A density curve is a theoretical model to describe a distribution.

Distribution for

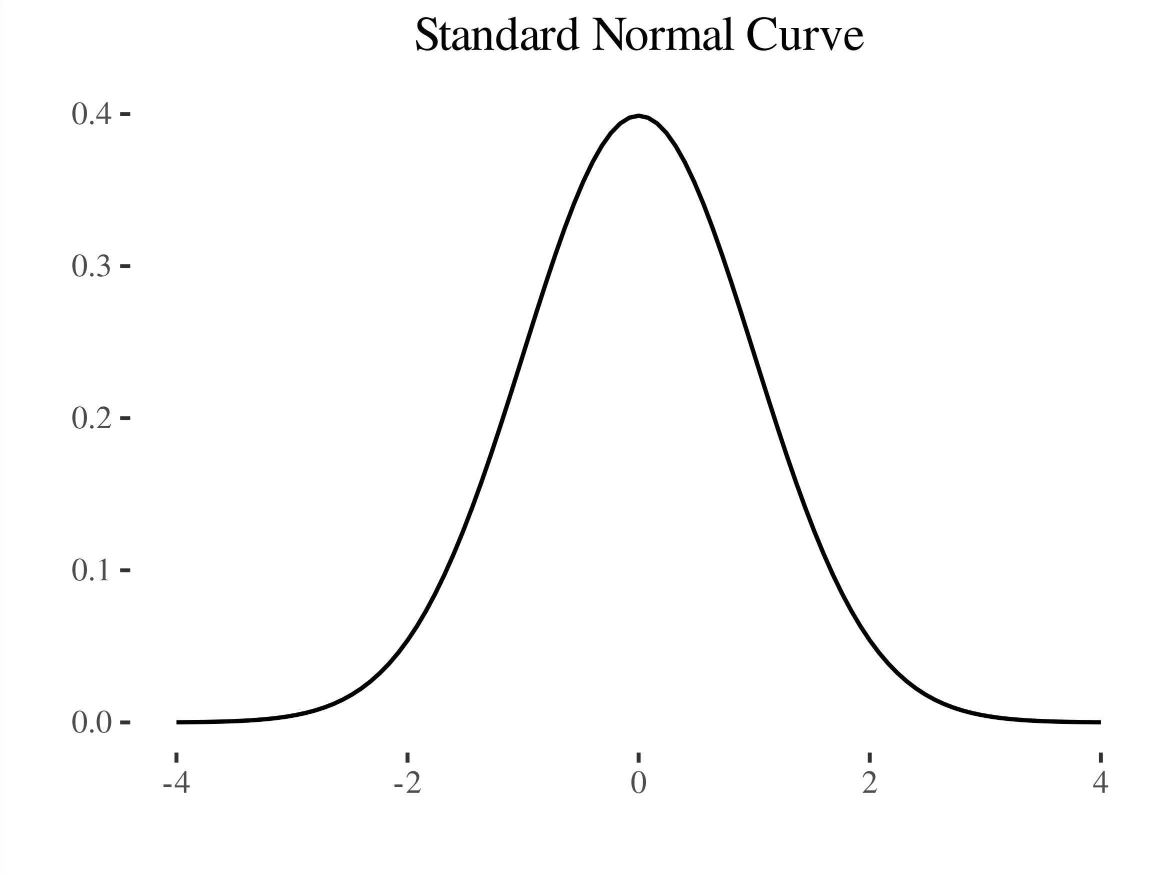
- *individual measurements in population (for a quantitative variable)*
- *Sampling distribution for a statistic*

All density curves have an area under the curve of 1 (100%)

- *give proportions/percents as areas under the curve*

Normal Distribution

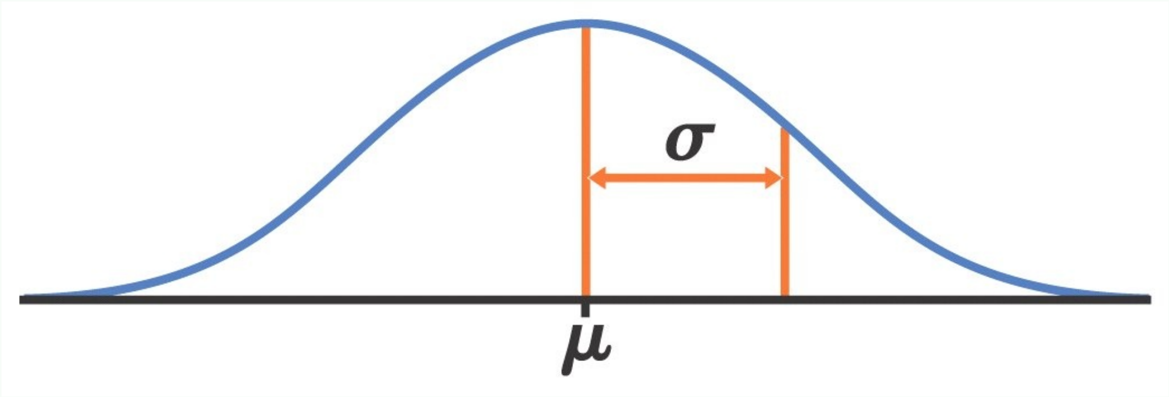
A **normal distribution** has a symmetric bell-shaped density curve.



The Normal Model: $X \sim N(\mu, \sigma)$

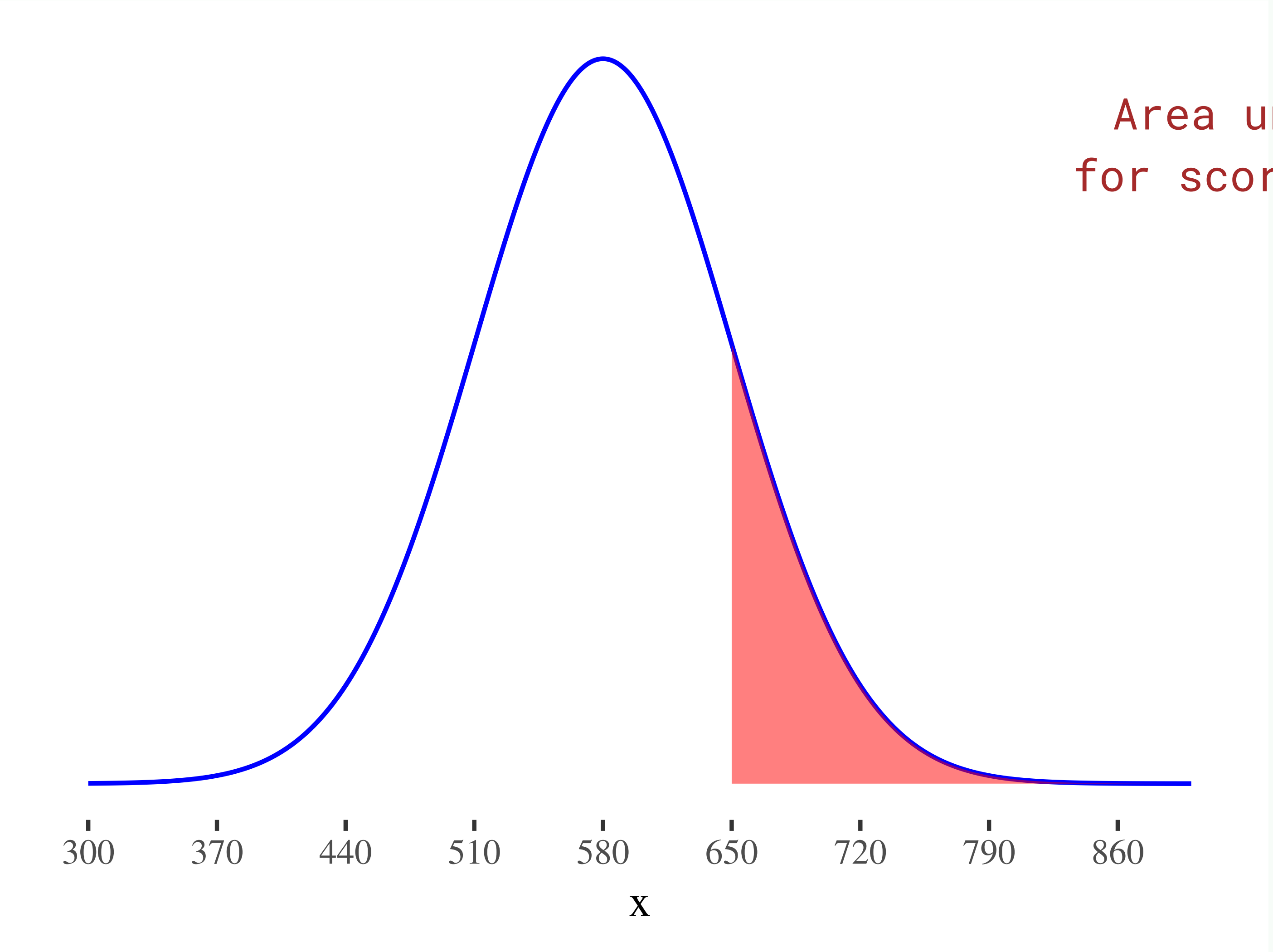
The mean and SD determine how a normal density curve looks. The normal model parameters are

- μ = model mean (center)
- $\sigma =$ model SD (variability)



Verbal SAT $\sim N(580, 70)$

What proportion of people score above 650?



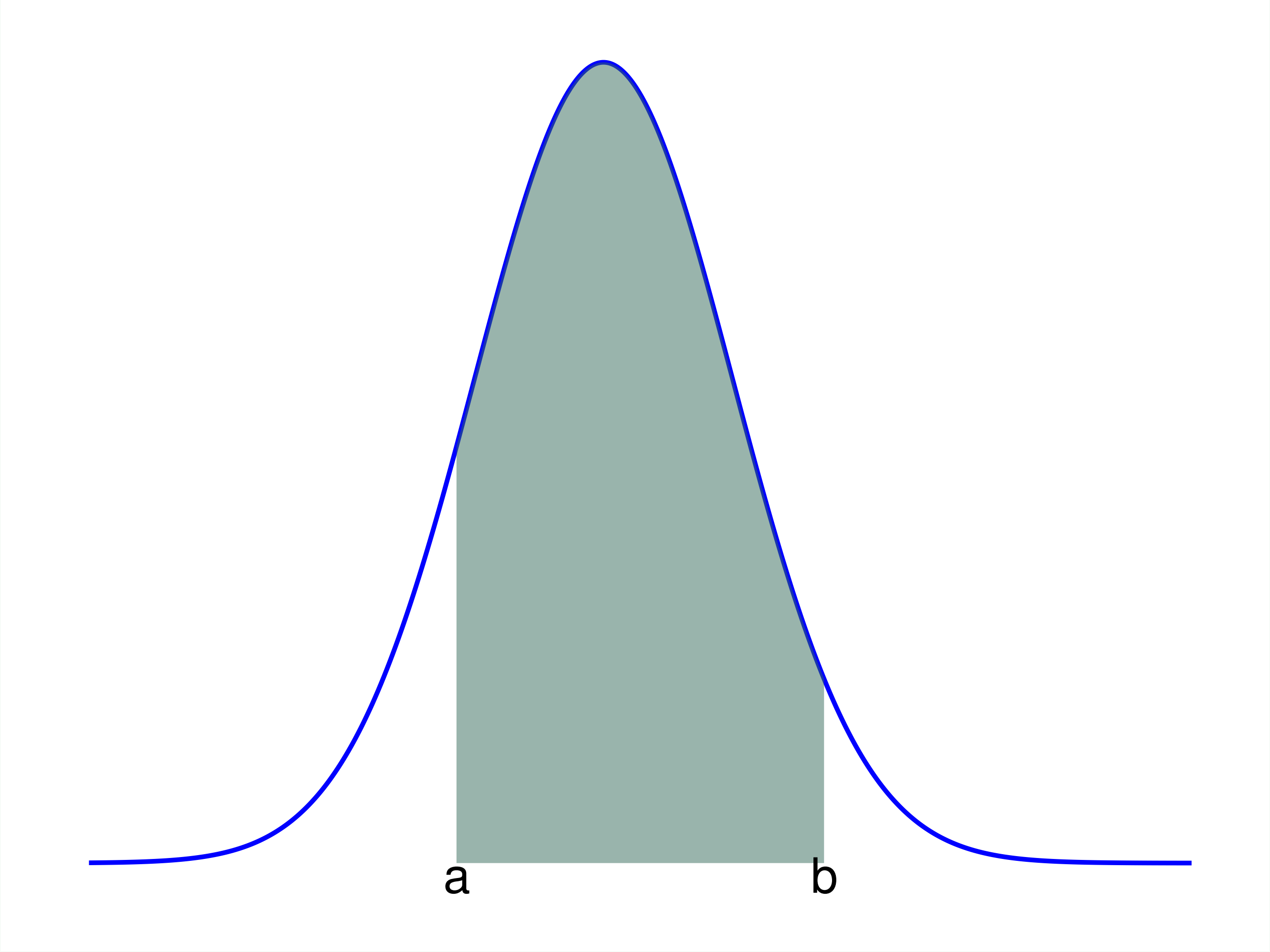
Area under the curve
for scores 650 & higher

How can we find areas under a normal density?

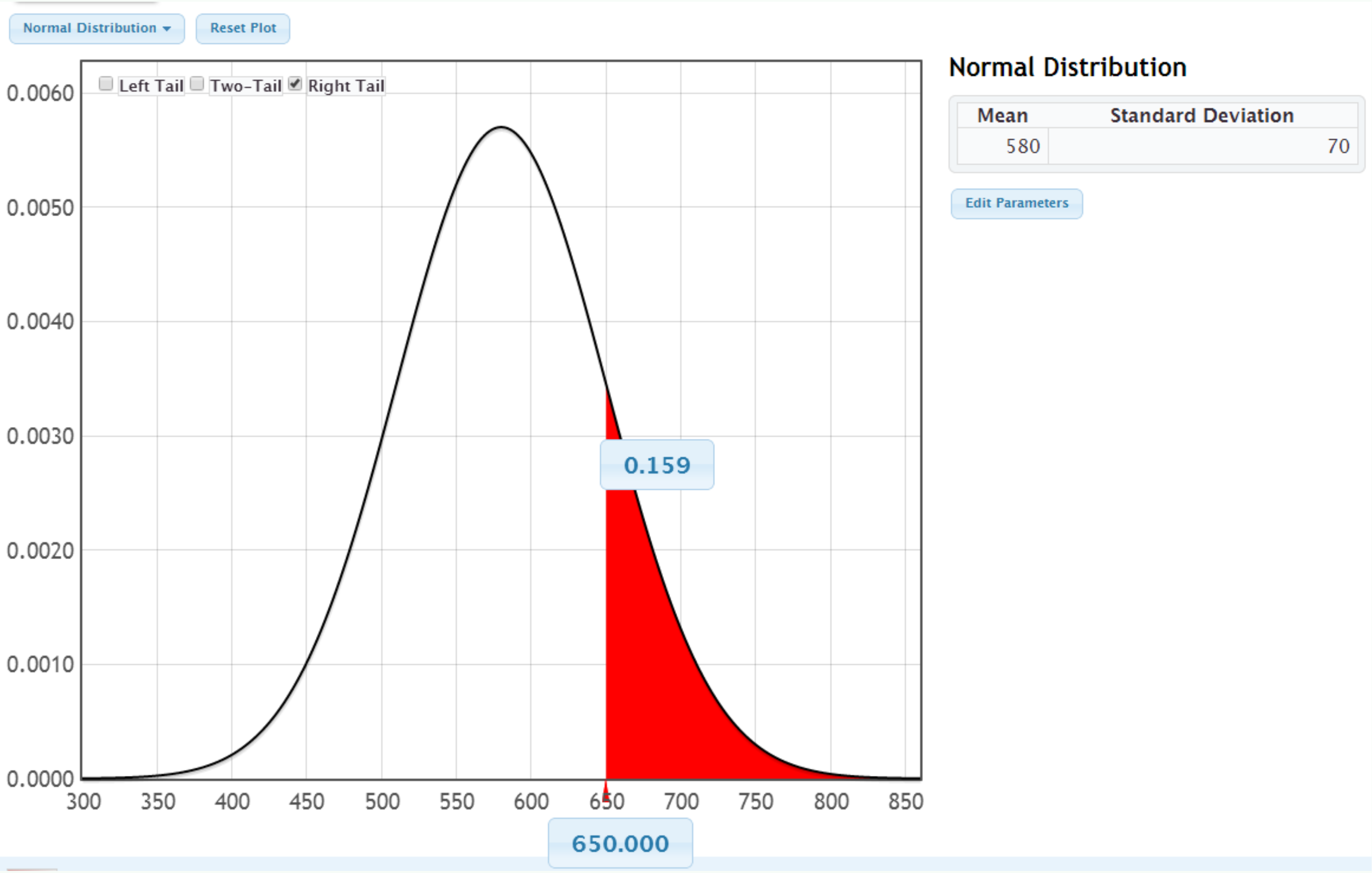
- The curve represents the normal distribution, denoted by $N(\mu, \sigma)$.
- (CALCULUS!) Calculating the exact area requires integration, as given by the formula:

  Area $= \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

- We'll just utilize technological tools.



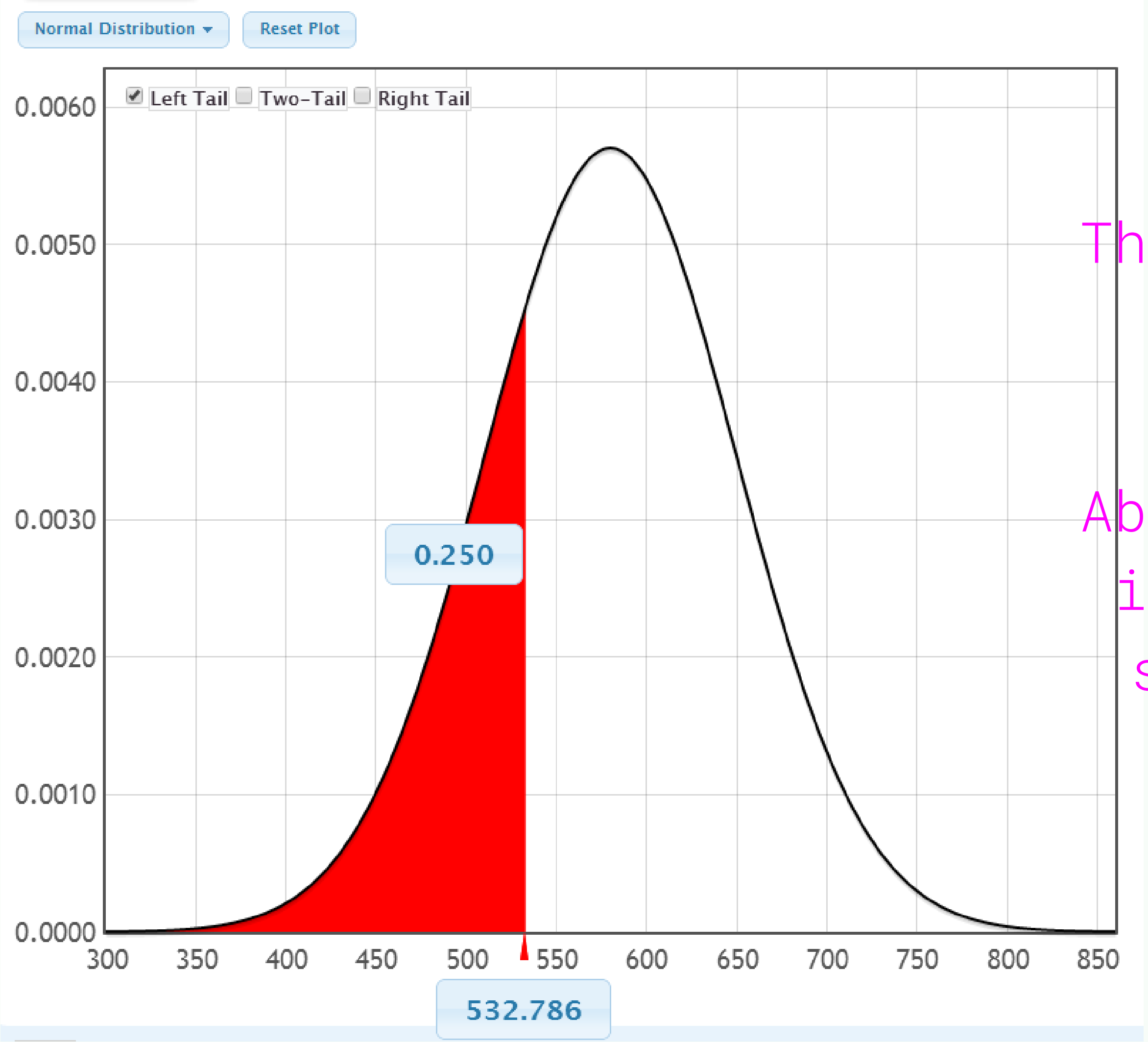
StatKey – Verbal SAT population



Some observations!

- *About 16% of individuals scored 650 or above.*
- *The 25^{th} percentile score lies between 440 (2 SD below 580) and the median of 580 .*
- *Utilize **Statkey** to fine-tune the left-tail area to 0.25 for precise percentile calculation.*

StatKey – Verbal SAT scores



The 25th percentile is 533.

About 75% of people in the population score above 533.

Example: Verbal SAT scores Using R

What percent of the population scored 650 or higher?

```
1 - pnorm(650, 580,70)
[1] 0.1586553
```

Alternatively:

```
pnorm(650, 580,70, lower.tail = FALSE)
[1] 0.1586553
```

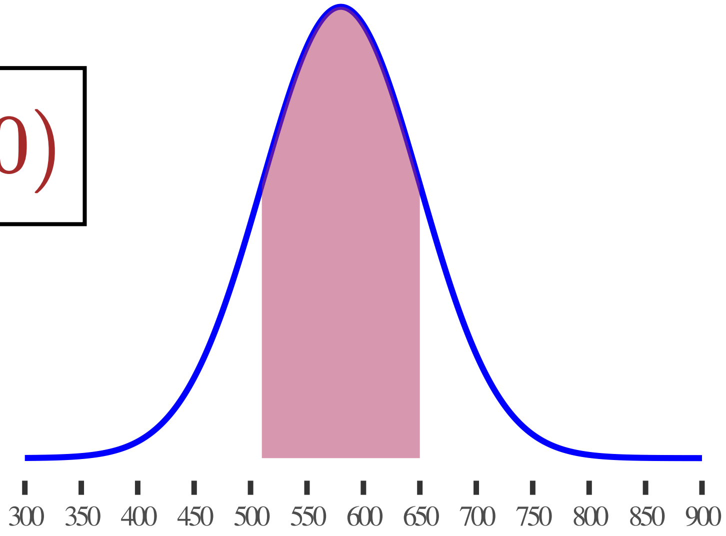
What score is the 25^{th} percentile?

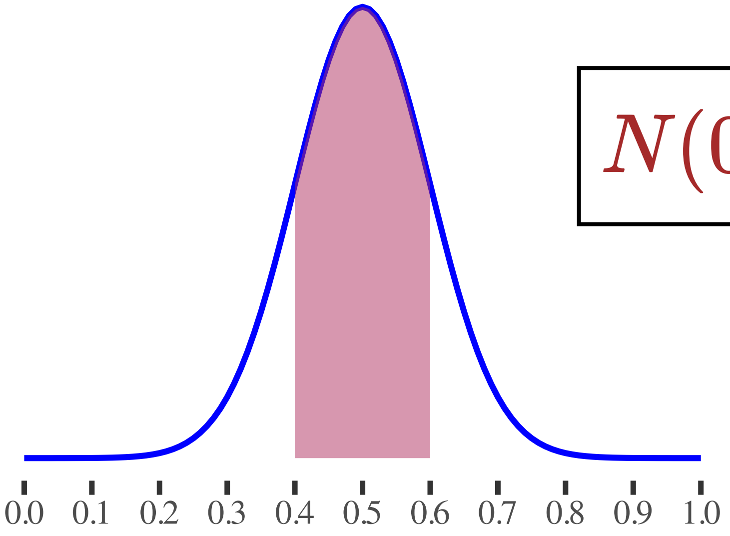
```
qnorm(.25,580,70)
[1] 532.7857
```

What score is the 75^{th} percentile?

```
qnorm(.75,580,70)
[1] 627.2143
```

$N(580, 70)$

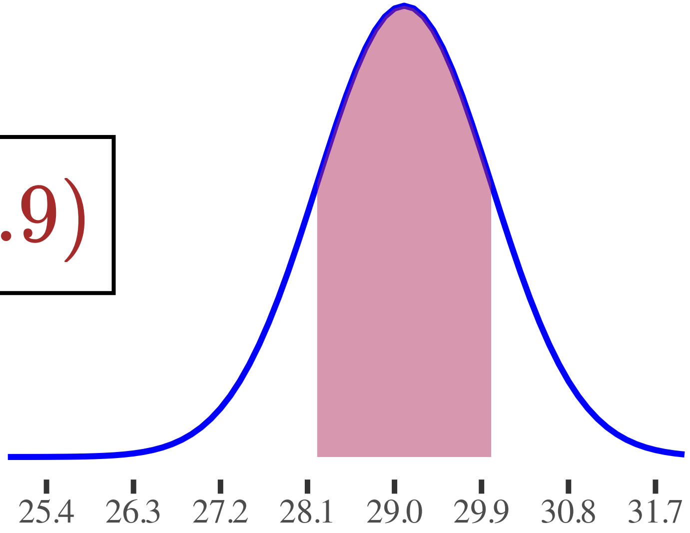


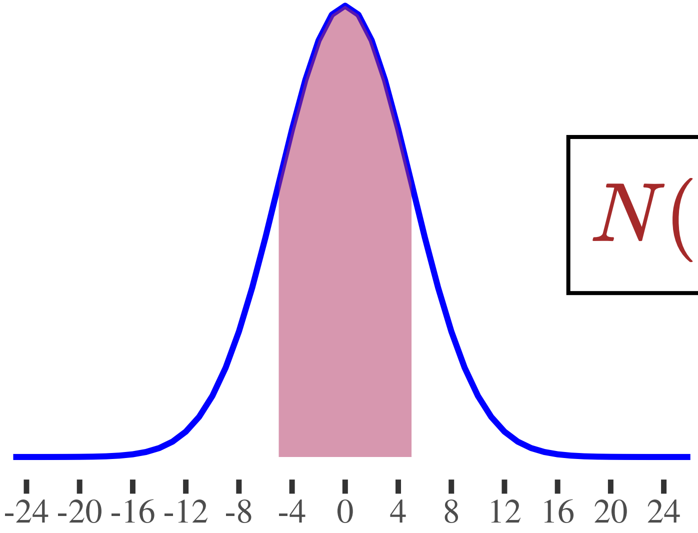


$N(0.5, 0.1)$

Observation: about 68% should be within one standard deviation of the mean

$N(29.1, 0.9)$

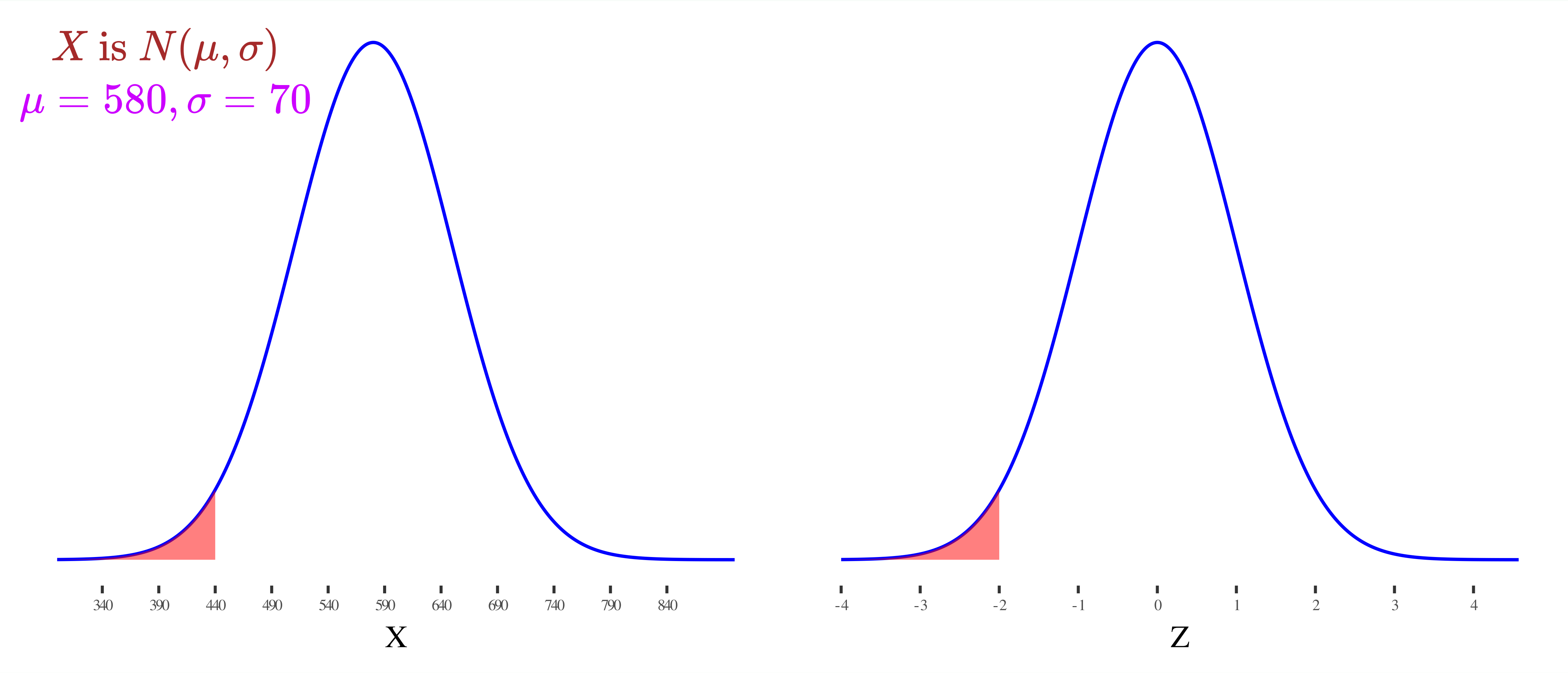




$N(0, 5)$

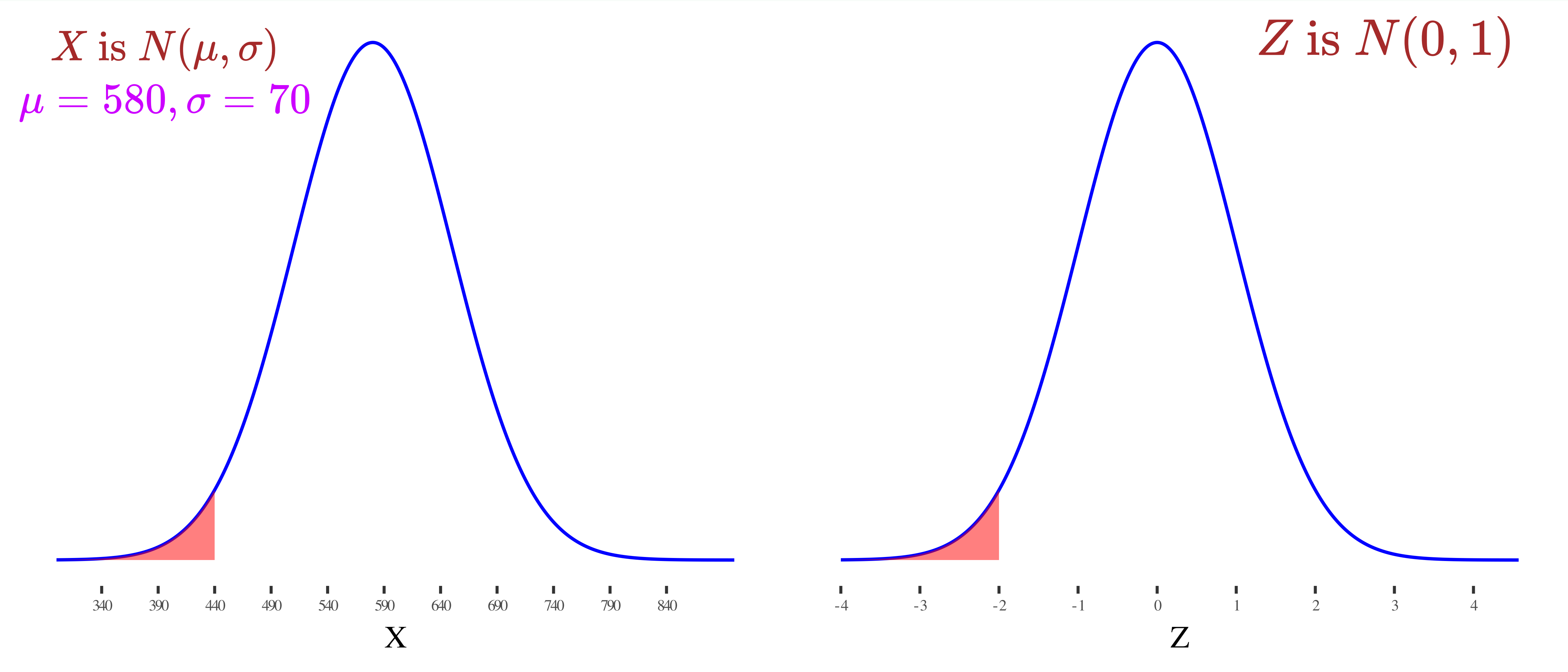
Connecting any Normal model to the standard normal model

Area below x = Area below z



Connecting any Normal model to the standard normal model

Area below x = Area below z



Connecting any Normal model to the standard normal model

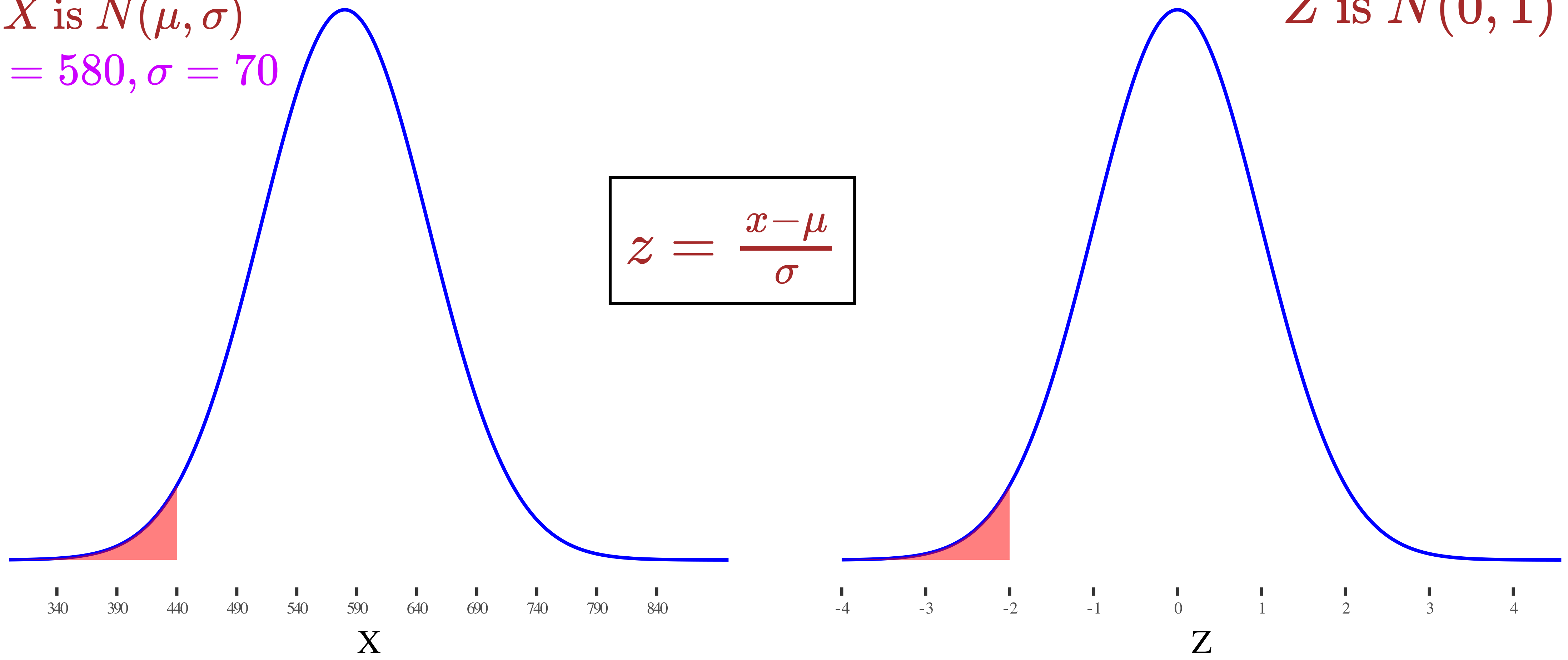
Area below x = Area below z

X is $N(\mu, \sigma)$

$\mu = 580, \sigma = 70$

Z is $N(0, 1)$

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



Big picture

When have we already been using normal models??

- *Bootstrap distributions – get confidence intervals if a bootstrap distribution is roughly bell-shaped*
- *Randomization distributions – many of these are bell-shaped.*
- *Normal models play a huge role in statistical inference.*
- *If we know the (bootstrap/randomization) standard error then we can just use a normal model rather than a resampling model (which requires more computational effort)*

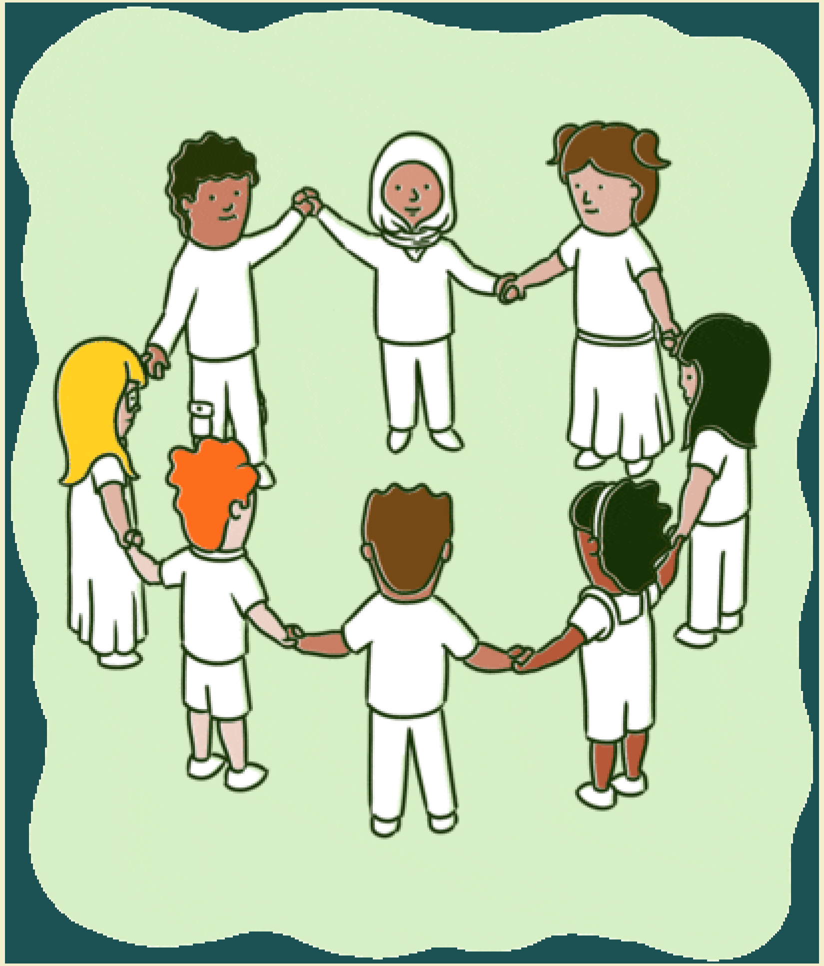
Big Idea for Normal Models: All we need is a z-score.

Standard Normal:

$$\mu = 0, \sigma = 1 \quad \implies Z \sim N(0, 1)$$

YOUR TURN 1

15:00



- Let's go over to the course helper page
- Please do the class activity and let me know if you have any questions
- Feel free to talk to your neighbor