The Normal Distribution

MAT 152 – Statistical Methods I
Lecture 1
Instructor: Dustin Roten
Fall 2020

Revisiting an Old Example: Standard Deviation

Suppose two students, Alice and Bob, are applying for a scholarship. Their applications are nearly identical so their test scores will be used as a tiebreaker. Alice took the SAT and Brian took the ACT. These two tests are scored on different scales so a method of standardization must be used.

The **z-scores** of Alice and Brian's test scores can be calculated for comparison.

Alice's z-score: $z_A = \frac{x-\mu}{\sigma} = \frac{1345-1081}{176} = 1.5$

Brian's z-score: $z_B = \frac{x-\mu}{\sigma} = \frac{24-20.8}{5.3} \approx 0.603$

Student	Score	Mean Score	Standard Deviation
Alice	1345	1081	176
Brian	24	20.8	5.3

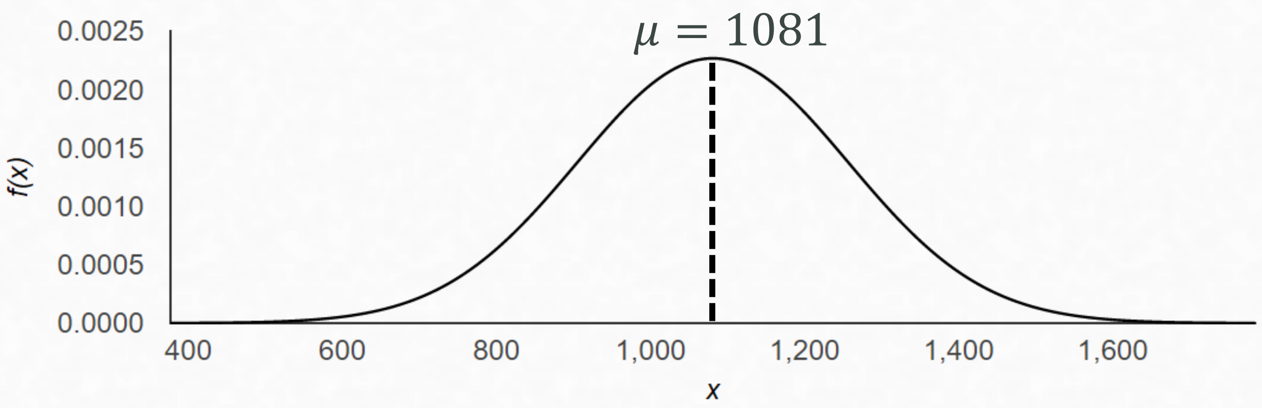
$Z_A > Z_B$

Since the z-score of Alice's test score is higher than Brian's, Alice scored better between the two.

Normal Distributions

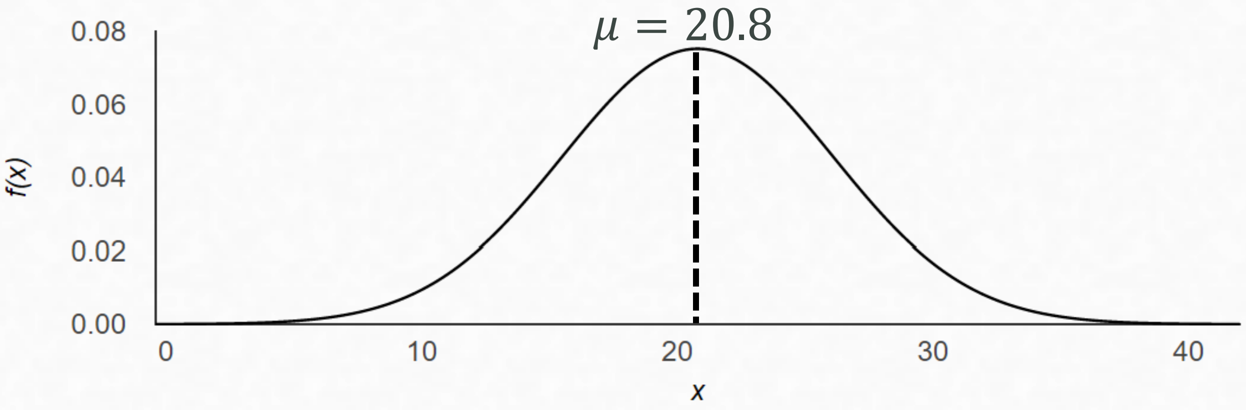
Most exam scores such as the SAT and ACT follow a Normal Distribution

Scores are clustered around a "typical" value (the mean) with fewer extreme values on either side.



SAT

$\mu = E(X) = 1081 \quad \sigma = SD(X) = 176 \quad \sigma^2 = Var(X) = 30976$



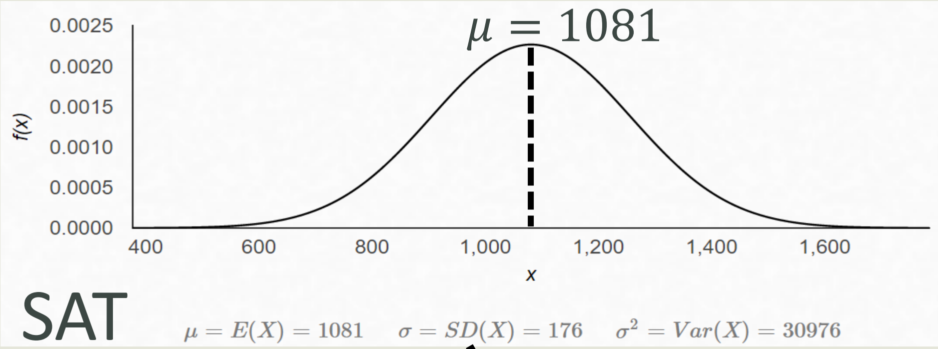
ACT

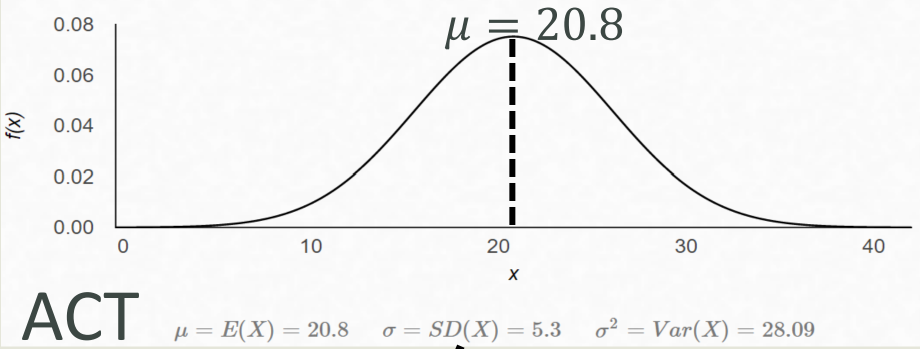
$\mu = E(X) = 20.8 \quad \sigma = SD(X) = 5.3 \quad \sigma^2 = Var(X) = 28.09$

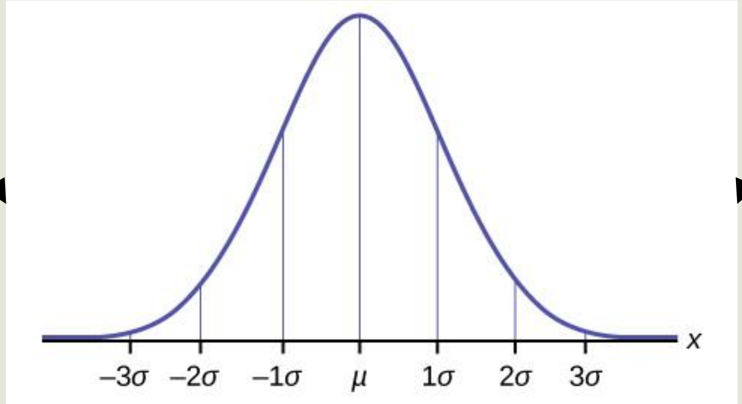
Standard Normal Distribution: Z-Scores

Calculating z-scores allows for comparisons between datasets.

Calculating a z-score "transforms" any normal distribution into the **standard normal distribution**. ($\mu = 0; \sigma = 1$)







Transforming normal distributions to the **standard normal distribution** puts them on the same scale.

The Normal Distribution

The normal distribution is widely applied in many fields.

Two parameters are necessary: the mean and the standard deviation.

If a continuous random variable has a normal distribution, it is notated by $X \sim N(\mu, \sigma)$.

The probability density function (pdf):

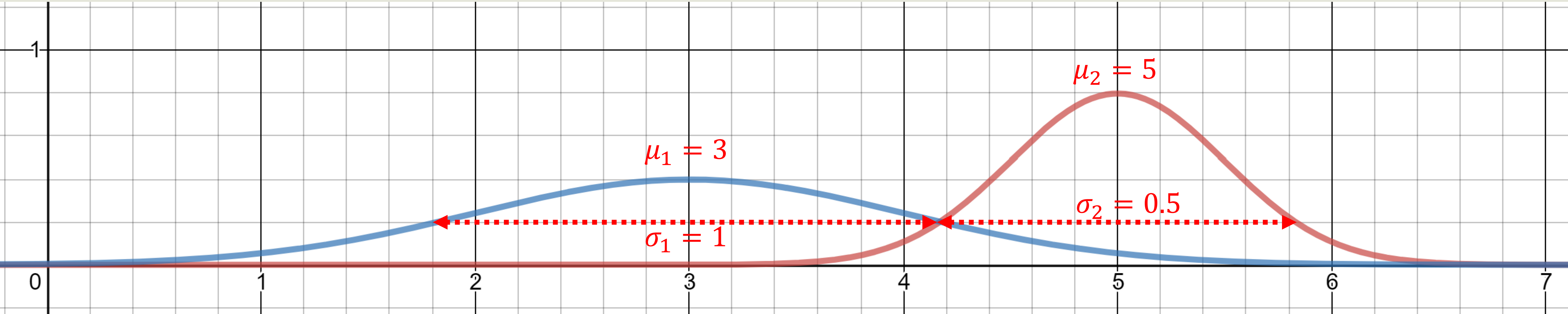
$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The curve is symmetric about a mean. In theory, the mean is the same as the median.

The Normal Distribution

The value of the mean determines WHERE the normal distribution is located.

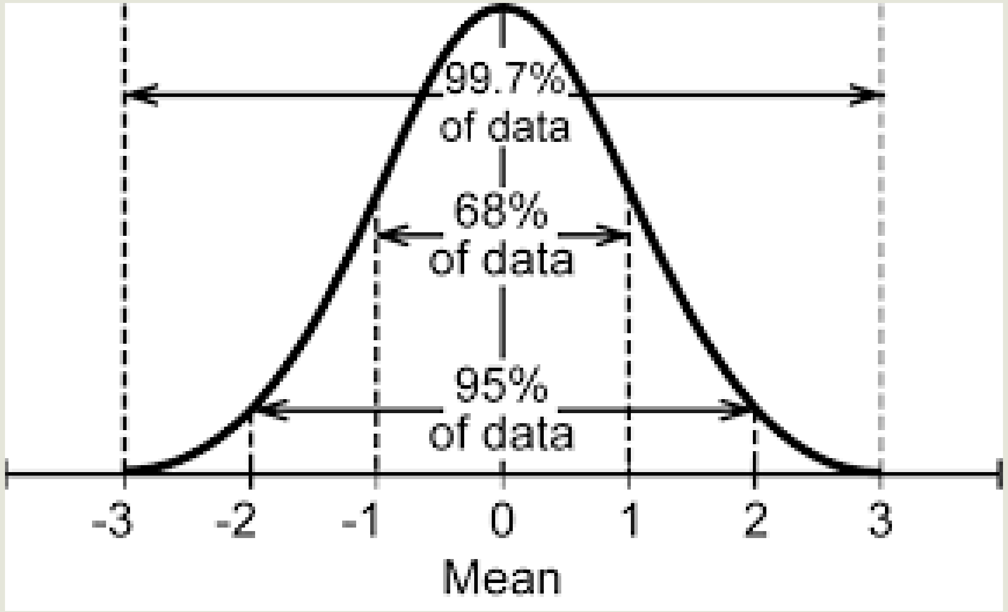
The value of the standard deviation determines the SPREAD of the normal distribution



The Empirical Rule

If $X \sim N(\mu, \sigma)$ then the **Empirical Rule** states the following:

- About 68% of values lie between -1σ and $+1\sigma$ of the mean. (Within one standard deviation of the mean)
- About 95% of values lie between -2σ and $+2\sigma$ of the mean. (Within two standard deviations of the mean)
- About 99.7% of values lie between -3σ and $+3\sigma$ of the mean. (Within three standard deviations of the mean)



Example

The weights of peaches (in ounces) packaged at a particular farm follow the distribution: $X \sim N(6.2, 0.8)$. Determine the following:

1) If the weight of a particular peach is one standard deviation below the mean, what is its weight?

   $z = \frac{x-\mu}{\sigma}$ $\quad -1 = \frac{x-6.2}{0.8}$ $\quad x = -1 \cdot (0.8) + 6.2 = 5.4$ $\quad$ 5.4 ounces

2) If the weight of a particular peach is one standard deviation above the mean, what is its weight?

   $z = \frac{x-\mu}{\sigma}$ $\quad 1 = \frac{x-6.2}{0.8}$ $\quad x = 1 \cdot (0.8) + 6.2 = 7.0$ $\quad$ 7.0 ounces

3) In a sample of 500 peaches, approximately how many of them have a weight between these values?

   Using the Empirical Rule, we know that ~68% of data lies between -1σ and 1σ (between z=-1 and z=1). So, roughly 68% of the 500 peaches lie within this range.
   $500 \cdot 0.68 = 340$ peaches.

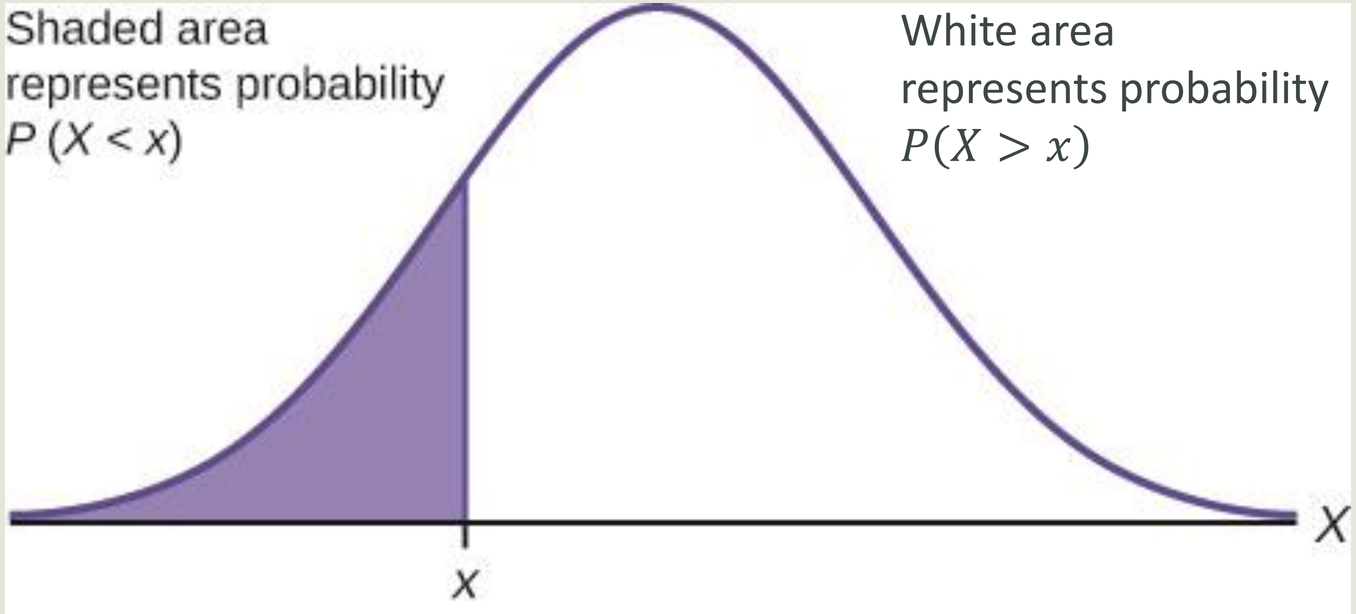
A Quick Review of Continuous Distributions

$P(X < x)$ represents the area to the left of x.

$P(X > x)$ represents the area to the right of x.

$P(X \leq x) = P(X < x)$

$P(X \geq x) = P(X > x)$

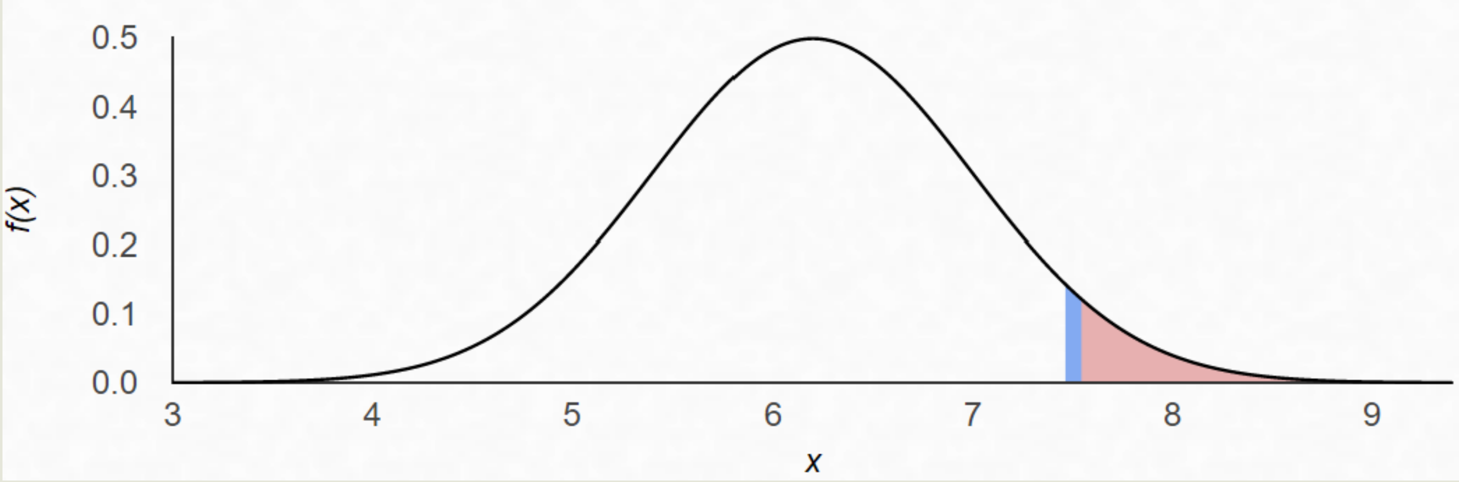


Example

The weights of peaches (in ounces) packaged at a particular farm follow the distribution: $X \sim N(6.2, 0.8)$.

Find the probability that a randomly selected peach weighs more than 7.5 ounces.

$P(X > 7.5) = 0.05208$ (calculated using technology)



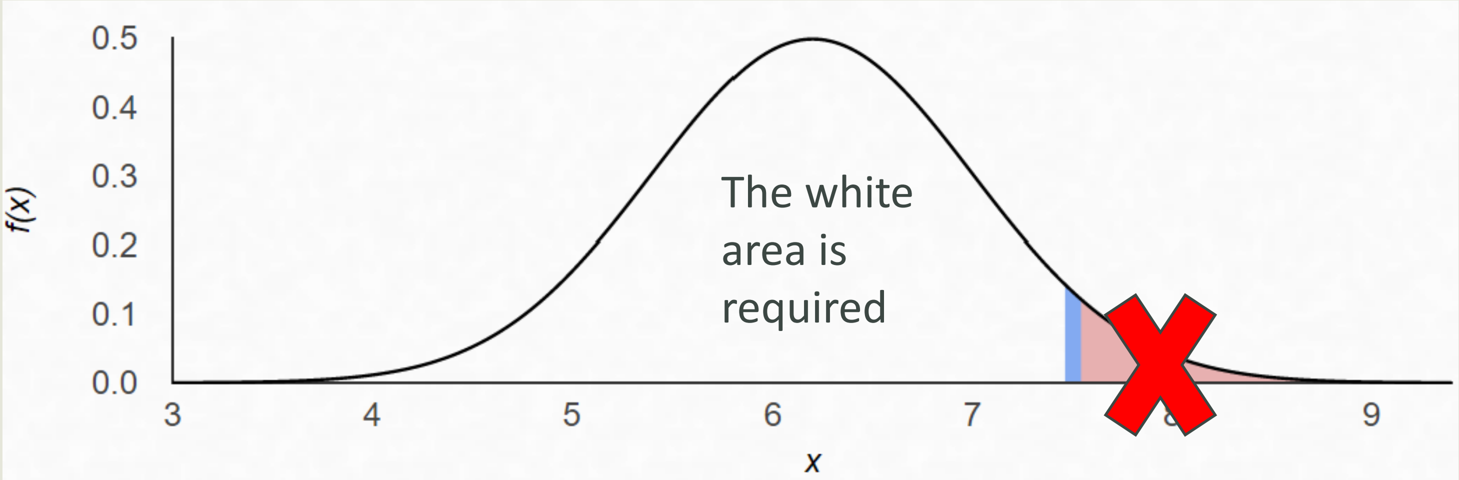
Example

The weights of peaches (in ounces) packaged at a particular farm follow the distribution: $X \sim N(6.2, 0.8)$.

Find the probability that a randomly selected peach weighs less than 7.5 ounces.

$P(X < 7.5) = 1 - P(X > 7.5) = 1 - 0.05208 = 0.94792$

(calculating the complement)



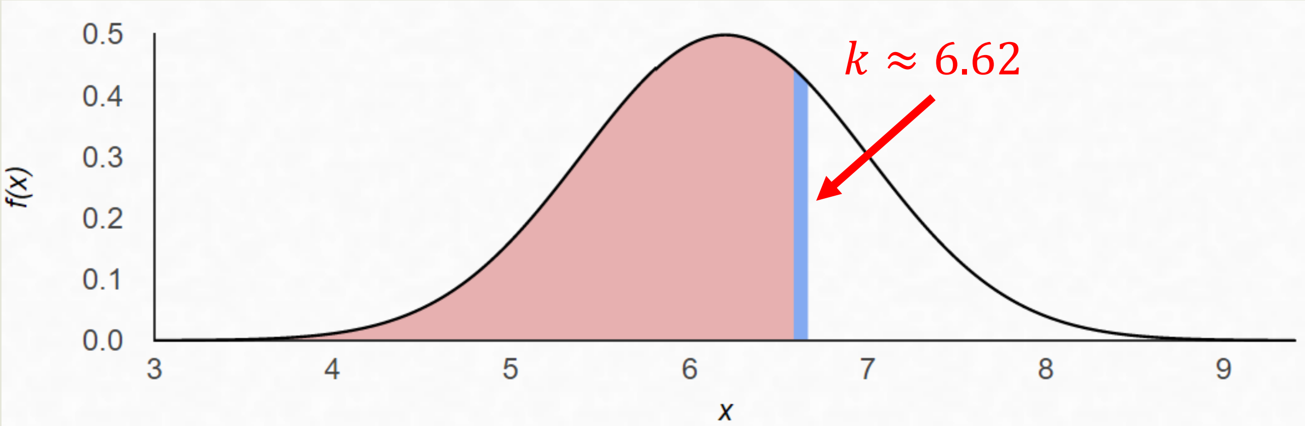
Example

The weights of peaches (in ounces) packaged at a particular farm follow the distribution: $X \sim N(6.2, 0.8)$.

Find the 70$^{\text{th}}$ percentile of weights.

$P(X < k) = 0.7$ $\qquad$ $k \approx 6.62$

(calculated using technology)



Roughly 70% of the peaches weigh less than 6.62 ounces

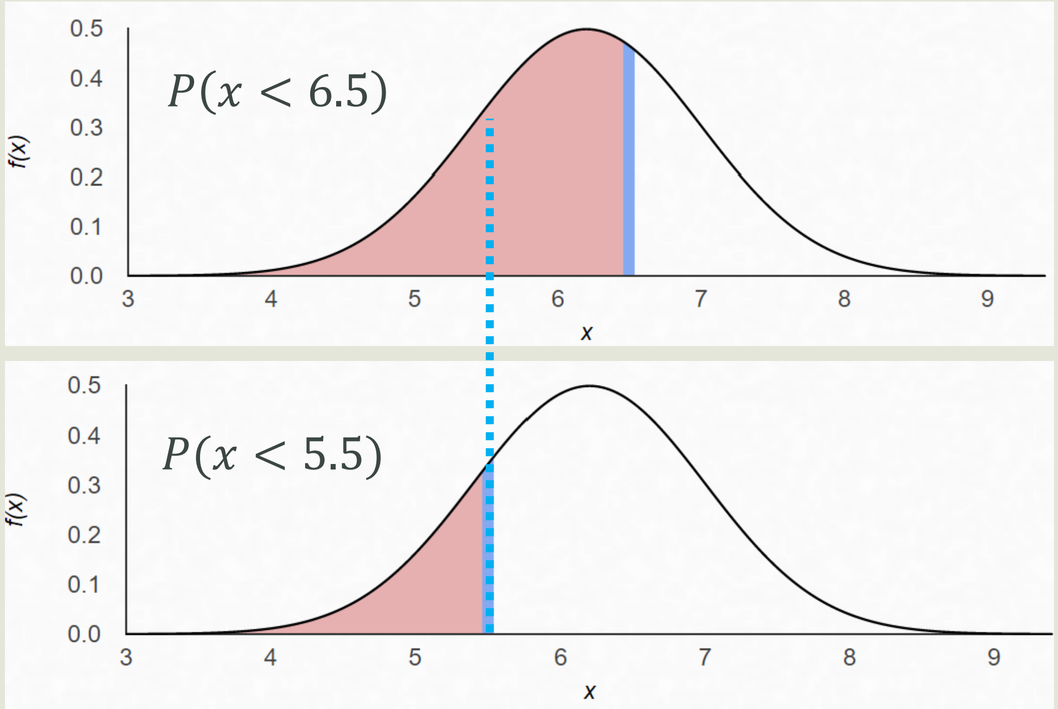
Roughly 30% of the peaches weigh more than 6.62 ounces.

Example

The weights of peaches (in ounces) packaged at a particular farm follow the distribution: $X \sim N(6.2, 0.8)$.

Find the probability that the weight of a randomly selected peach falls between 5.5 and 6.5 ounces.

$P(5.5 < x < 6.5)$



$P(5.5 < x < 6.5) = P(x < 6.5) - P(x < 5.5)$
$P(5.5 < x < 6.5) = 0.64617 - 0.19079 = 0.45538$

Remove the "unwanted" probability

A Quick Review

The normal distribution is characterized by many values grouped around a “typical” value (the mean) with fewer values further away.

The normal distribution has a bell shape. It is also called a Gaussian curve.

Calculating a z-score “standardizes” a normal distribution.

The normal distribution follows the Empirical Rule.