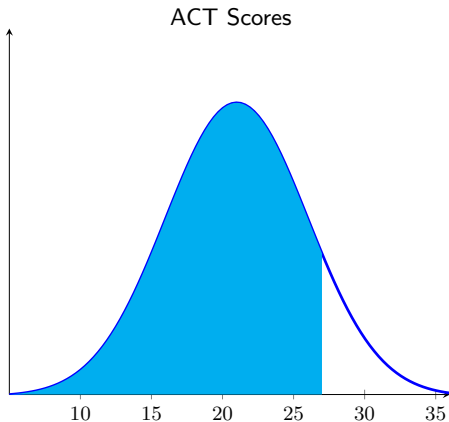# Introductory Statistics – Day 15

Introduction to the Normal Distribution

Big ideas about normal distributions

In a probability distribution, the total area represented in the graph is always 1 (or 100%) and the probability of being in a given section of the graph is the same as the area of that section of the graph.

ACT Scores

This area represents 88.87% of the data, so if a data point is picked at random, there is an 88.87% chance of a data point ending up in this area.

Many real world situations result in probability distributions in the shape of a bell curve / mound shaped distribution. Examples:

- Heights of 5 year olds
- Weights of parakeets
- Scores on standarized tests
- Sampling distributions

In each case, there are many minor independent causes for change underlying the variable, which causes most data to clump toward the center.

Curves of this type are so common in statistics, that this curve is called the *normal distribution*.

- The normal distribution can be built with any mean and standard deviation you like.
- The mean is denoted μ (mu) and the standard deviation is σ (sigma). The normal distribution can be noted in shorthand as $N(\mu, \sigma)$. e.g. If we say a distribution is $N(400, 100)$, then it is a normal distribution with mean 400 and standard deviation 100.
- If you pick the mean of $\mu = 0$ and the standard deviation of $\sigma = 1$, then this is even more special and its called the *standard normal distribution*, i.e. $N(0, 1)$.

All normal distributions are the same shape, except for scale.

Whatever the mean and standard deviation, for a normal distribution, we always have

- about 68% of data within 1 standard deviation σ of the mean μ
- about 95% of data within 2 standard deviations σ of the mean μ
- about 99.7% of data within 3 standard deviation σ of the mean μ
- If your data is more than 3 standard deviations from the mean, it is really rare / weird.
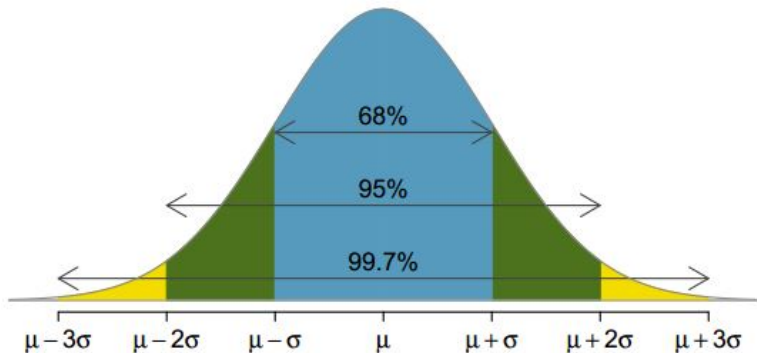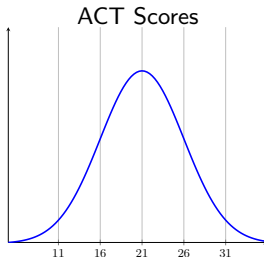

68%, 95%, 99.7% Rule, image from OpenIntro Statistics by Diez et al.

Bell curves in ACTs, SATs, and IQ scores

Consider the scores on two different standardized tests: the ACTs and SATs.

If we were to plot everyone who took the ACT in a given year we would get a bell curve with a mean of 21 points and a standard deviation of about 5 points.

ACT Scores

If we were to do the same with the SAT scores we would get a bell curve with a mean of 1500 and a standard deviation of about 300.

SAT Scores
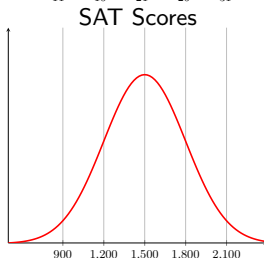
Voting questions:

An IQ test is normally distributed with mean 100 and standard deviation 15, $N(100, 15)$. Sketch a normal distribution curve for the IQ test and label the x axis with appropriate values.
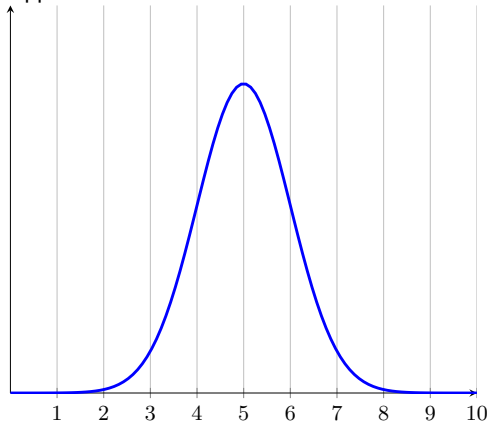

Earning a 21 on the ACT is similar to earning a 1500 on the SAT. What SAT score is similar to a 16 on the ACT?

What ACT score is similar to a 2100 on the SAT?

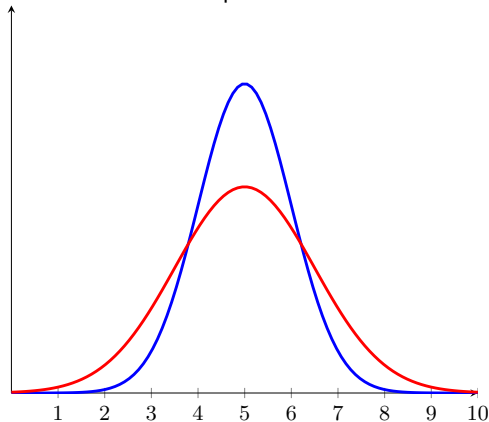How rare is it to earn a 31 or higher on the ACT?

Voting questions:

Approximate the standard deviation on the blue normal distribution.

(A) $\mu = 5$ and $\sigma = 1$

(B) $\mu = 5$ and $\sigma = 2$

(C) $\mu = 5$ and $\sigma = 3$
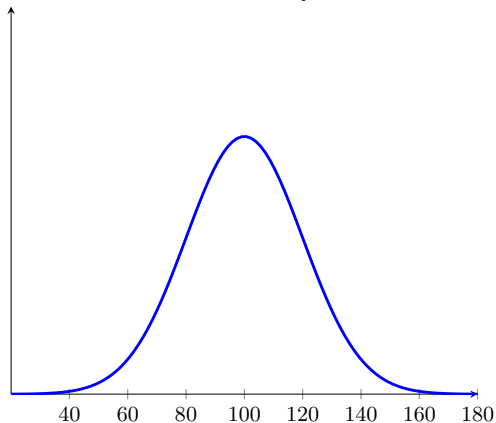
(D) $\mu = 5$ and $\sigma = 4$

Voting questions:

Pick the correct comparison

(A) $\sigma_{red} = \sigma_{blue}$

(B) $\sigma_{red} > \sigma_{blue}$

(C) $\sigma_{red} < \sigma_{blue}$

(D) Not enough information

Voting questions:

In the following normal distribution we have $\mu = 100$ and $\sigma = 20$.
68% of the data described by this distribution is between which two values?

(A) 99 and 101
(B) -80 and 120
(C) 80 and 120
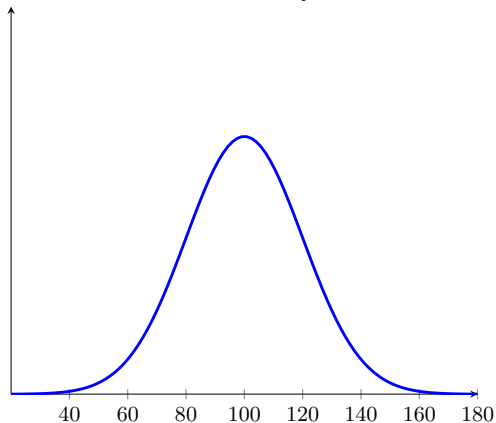(D) 60 and 140

Voting questions:

In the following normal distribution we have $\mu = 100$ and $\sigma = 20$.
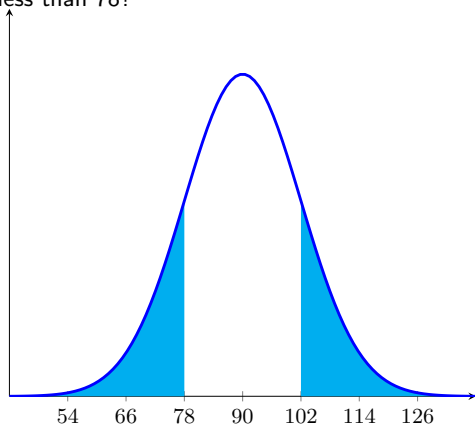95% of the data described by this distribution is between which two values?

(A) 99 and 101

(B) -80 and 120

(C) 80 and 120

(D) 60 and 140

Voting questions:

In the following normal distribution we have $\mu = 90$ and $\sigma = 12$.
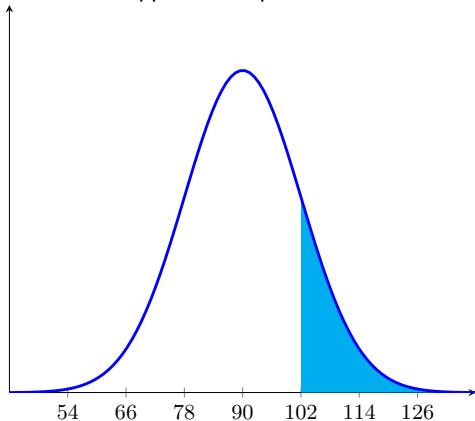What is the approximate percent of the data that is either greater than 102 or less than 78?

(A) 68%

(B) 95%

(C) 32%

(D) 5%

(E) Not enough information

Voting questions:

In the following normal distribution we have $\mu = 90$ and $\sigma = 12$.
What is the approximate percent of the data that is greater than 102?

(A) 32%

(B) 68%

(C) 16%
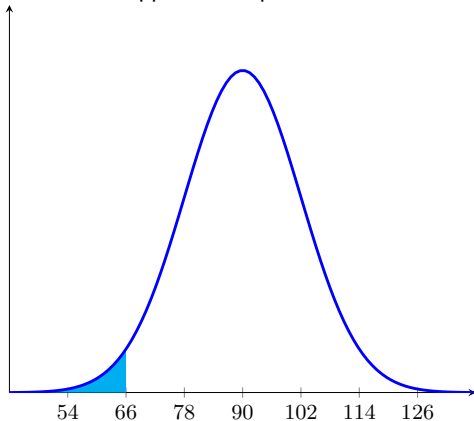
(D) 5%

(E) Not enough information

Voting questions:

In the following normal distribution we have $\mu = 90$ and $\sigma = 12$.
What is the approximate percent of the data that is less than 66?

(A) 95%

(B) 5%

(C) 2.5%

(D) 10%

(E) Not enough
    information

Voting questions:

In the following normal distribution we have $\mu = 5$ and $\sigma = ??$.
If 97.5% of the data falls below 8 then what is the approximate standard deviation for the distribution?

97.5%
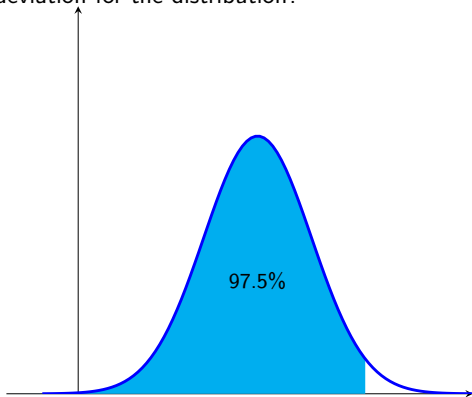
(A) $\sigma \approx 1$
(B) $\sigma \approx 1.5$
(C) $\sigma \approx 2$
(D) $\sigma \approx 2.5$
(E) $\sigma \approx 3$

Definition

The *z-score* for a value from a normal distribution is the number of standard deviations the value is away from the mean.

In a normal distribution with mean $\mu = 6$ and standard deviation $\sigma = 2$, what is the z score for $x = 5.5$?

- (A) $z = -0.5$
- (B) $z = -0.25$
- (C) $z = 0.5$
- (D) $z = 0.25$

Note: The distance from the mean $= x - \mu$, while the number of standard deviations from the mean is $z = \frac{x - \mu}{\sigma}$

Voting questions:

A number 1.5 standard deviations below the mean has a z score of

(A) 1.5

(B) -1.5

(C) 3

(D) -3

(E) not enough information

Draw the associated normal distribution plot.

Some Questions:

What proportion of the data is 1.5 σ above μ ?

What proportion of the data is 2.34 σ below μ?

Fortunately, we can use Excel to find these exact proportions.

Excel With Normal Distributions: Given Score Find Area

Mathematical Question	Excel Command
Prob. getting less than x	=NORM.DIST(x,mean,stdev,1)
Prob. getting greater than x	=1-NORM.DIST(x,mean,stdev,1)
Prob. getting exactly x	not possible

Activity 1.

The distribution of heights of American women aged 18 to 24 is approximately normally distributed with a mean of 65.5 inches and a standard deviation of 2.5 inches. What percent of these women is less than 5'8" (68 inches)?

(A) $P(x < 68) \approx 0.841$

(B) $P(x < 68) \approx 0.159$

(C) $P(x < 68) \approx 0.097$

(D) $P(x < 68) \approx 0.903$

Draw the associated normal distribution plot.

Activity 1, continued.

The distribution of heights of American women aged 18 to 24 is approximately normally distributed with a mean of 65.5 inches and a standard deviation of 2.5 inches. What percent of these women is greater than 5'8" (68 inches)?

(A) $P(x > 68) \approx 0.841$

(B) $P(x > 68) \approx 0.159$

(C) $P(x > 68) \approx 0.097$

(D) $P(x > 68) \approx 0.903$

Draw the associated normal distribution plot.

Excel With Normal Distributions: Given Area Find Score

Mathematical Question	Excel Command
Prob. getting less than x is P. Find x	`=NORM.INV(P,mean,stdev)`
Prob. getting greater than x is P. Find x	`=NORM.INV(1-P,mean,stdev)`

Activity 2.

A group of students at Carroll takes a statistics quiz. The distribution is normal with a mean of 25 and a standard deviation of 4. Everyone who scores in the top 30% of the distribution gets a certificate. What is the lowest score someone can get and still earn a certificate?

(A) 29

(B) 25

(C) 27

(D) 23

Draw the associated normal distribution plot.

Activity 2, continued.

A group of students at Carroll takes a statistics quiz. The distribution is normal with a mean of 25 and a standard deviation of 4. The top 5% of the scores get to compete in a statewide statistics contest. What is the lowest score someone can get and still go on to compete with the rest of the state?

 (A) 31

 (B) 32

 (C) 18

 (D) 19

Draw the associated normal distribution plot.

Activity 3.

Assume a normal distribution with a mean of 70 and a standard deviation of 12. What limits would include the middle 65% of the cases?

(A) Bottom Score = `norm.inv(58,70,12)`,
    Top Score = `norm.inv(82,70,12)`

(B) Bottom Score = `norm.inv(5,70,12)`,
    Top Score = `norm.inv(135,70,12)`

(C) Bottom Score = `norm.inv(.175,70,12)`,
    Top Score = `norm.inv(.825,70,12)`

(D) Bottom Score = `norm.inv(.475,70,12)`,
    Top Score = `norm.inv(.975,70,12)`

(E) not enough information

Draw the associated normal distribution plot.