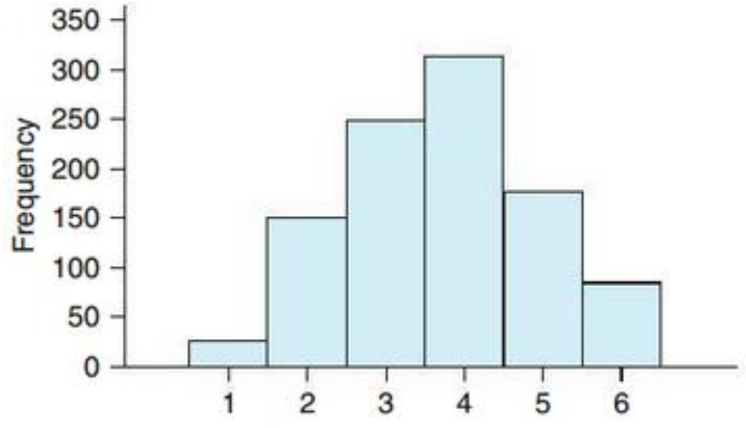
MAT111 REVIEW

Ch. 2 - 6

SUMMARIES FOR DISTRIBUTIONS

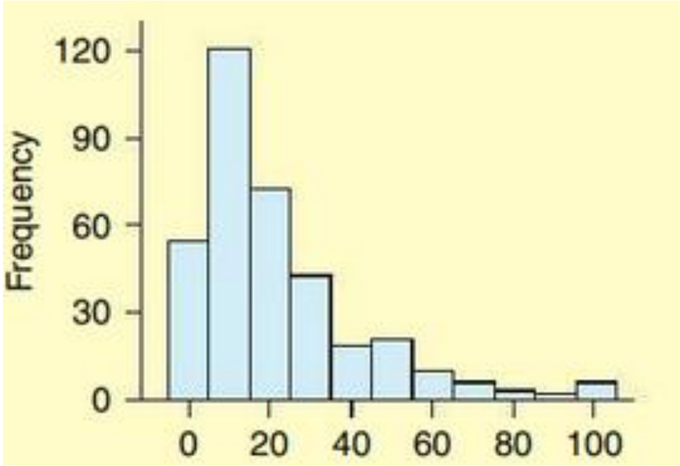
Symmetric Distributions

- Bell shaped
- The mean is a good representation for the “typical value”
- Mean = Median
- Majority of observations are less than one standard deviation from the mean.

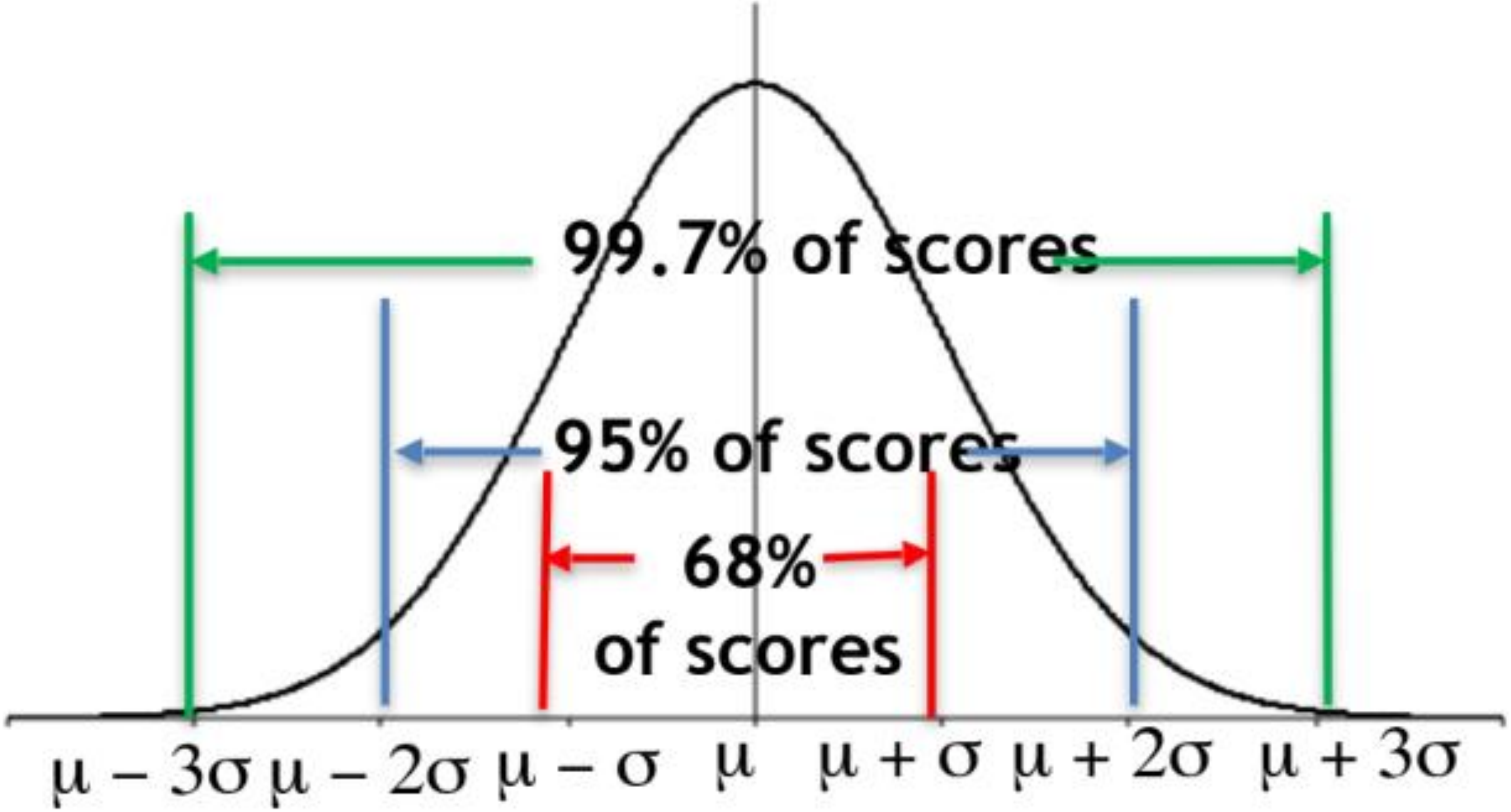


Skewed Distributions

- Higher on one end
- The median represents a “typical value” in a skewed distribution
- Skewed Left: Mean < Median
- Skewed Right: Mean > Median



EMPIRICAL RULE



Z-SCORES

- Z-Scores measure how many standard deviations away an observation is from the mean.
  - Positive Z-Score → Observed value is greater than the mean
  - Negative Z-Score → Observed value is less than the mean

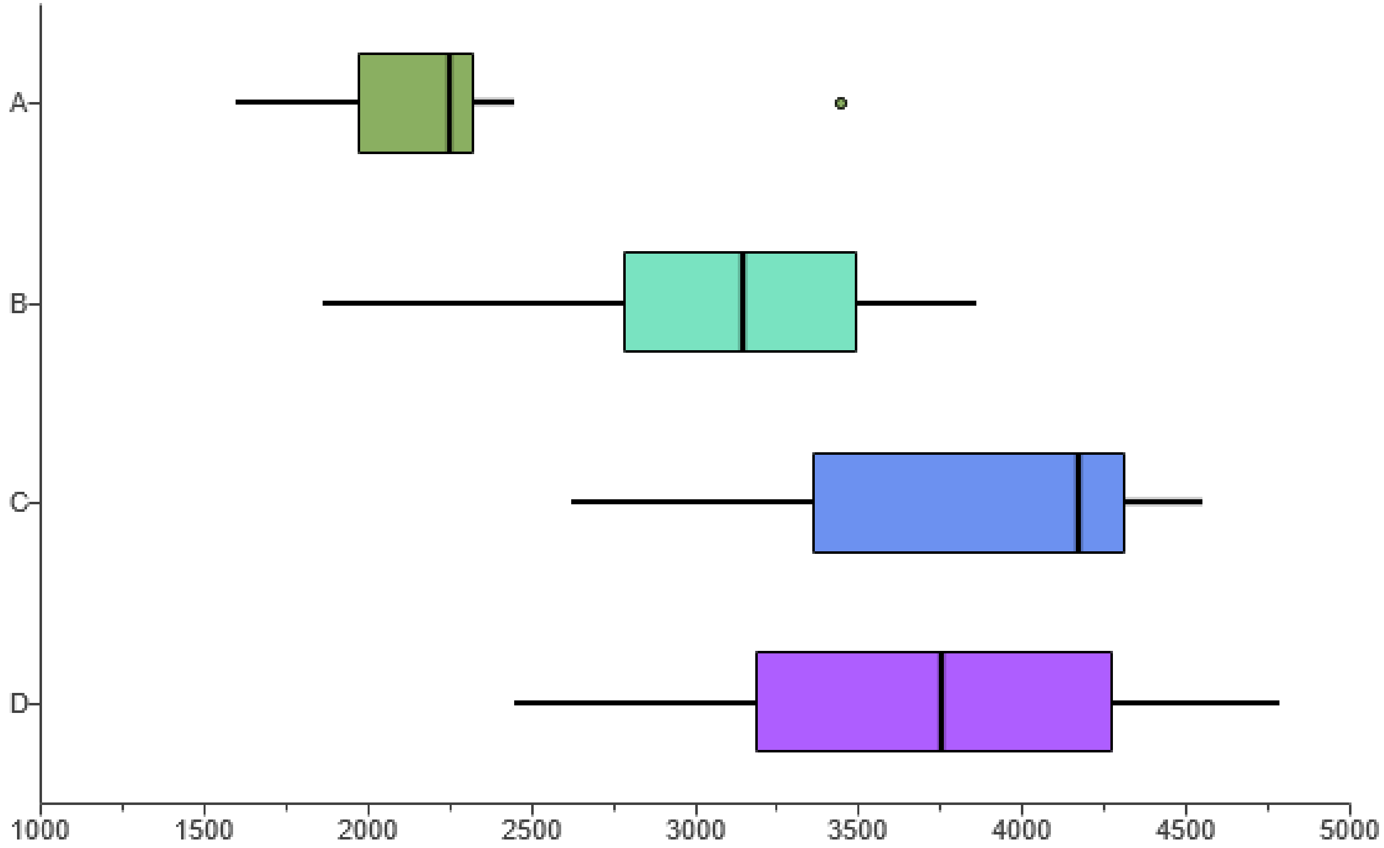
$$z = \frac{Observed\ Value\ - Mean}{Standard\ Deviation}$$

Finding Mean and Standard Deviation: StatCrunch – Stat – Summary Stats

BOXPLOTS

- 5 number Summary: Minimum, Q1, Median, Q3, Maximum
- Potential Outliers are numbers outside the "fences":
  - Lower fence: $Q1 - 1.5(IQR)$
  - Upper fence: $Q3 + 1.5(IQR)$

StatCrunch - Stat – Summary Stats



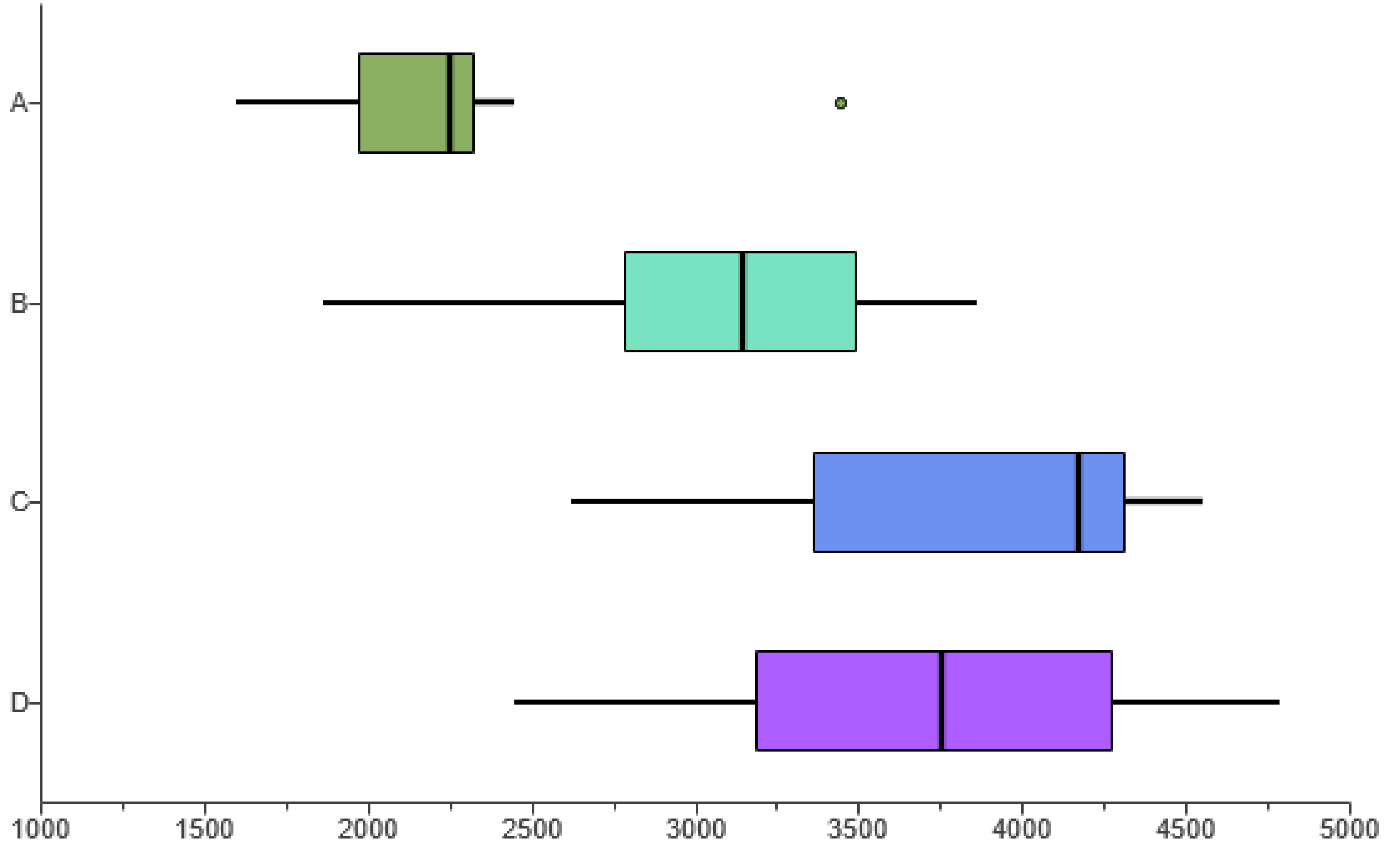
EXAMPLE: ANALYZE BOXPLOTS

- Which boxplot shows the most variation?

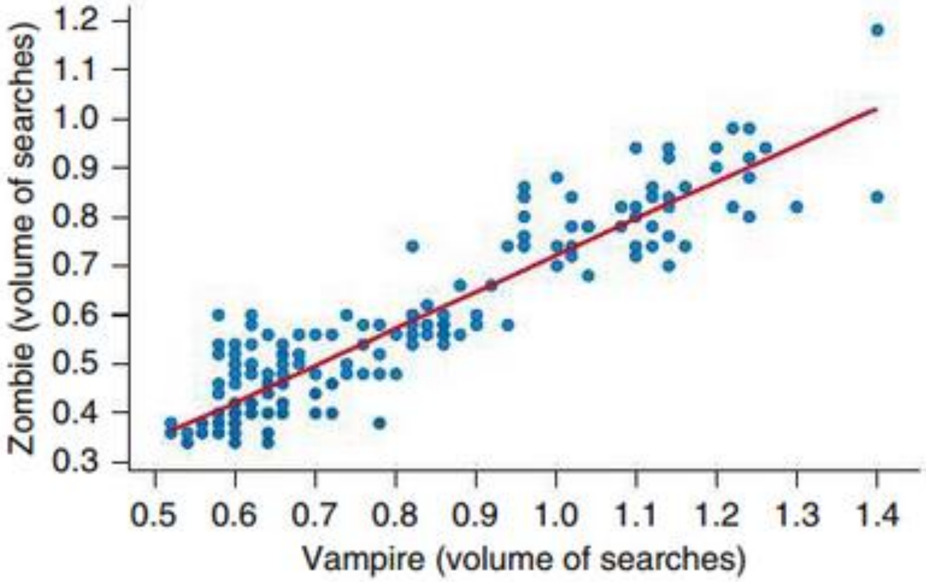
  Boxplot D

- Which boxplot shows the least variation?

  Boxplot A



REGRESSION ANALYSIS

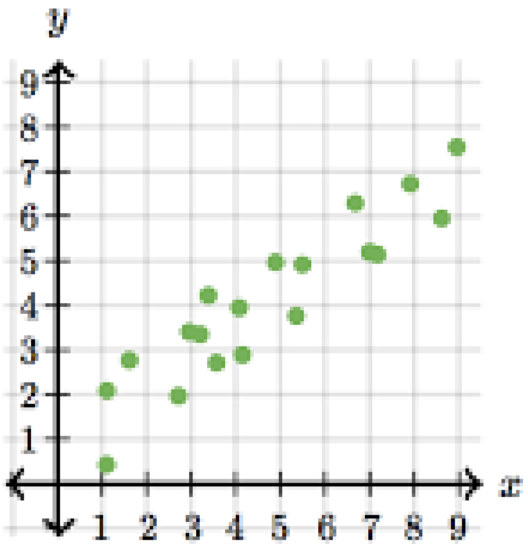


- Correlation Coefficient (r):
  Always between -1 and 1
  - A strong correlation is closer to -1 or 1
  - A weak correlation is closer to 0
- Regression Line (also Line of Best Fit or Least Squares): For making predictions about future observed values
  - x – explanatory, predictor, independent
  - y – response, predicted, dependent

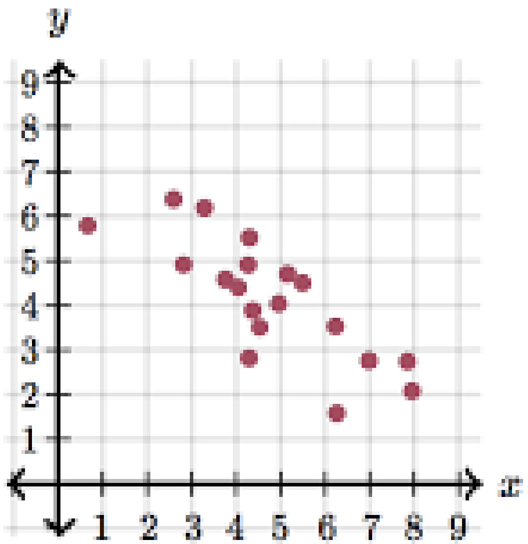
StatCrunch – Stat – Regression – Simple Linear – Select x and y

EXAMPLE: MATCH SCATTERPLOTS TO CALCULATED CORRELATIONS

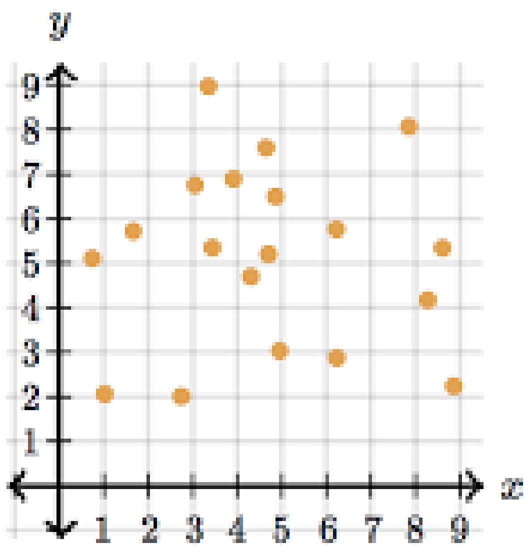
- Three scatterplots are shown below. The calculated correlations are 0.915, -0.782, and 0.037. Determine which correlation goes with which scatterplot.



0.915



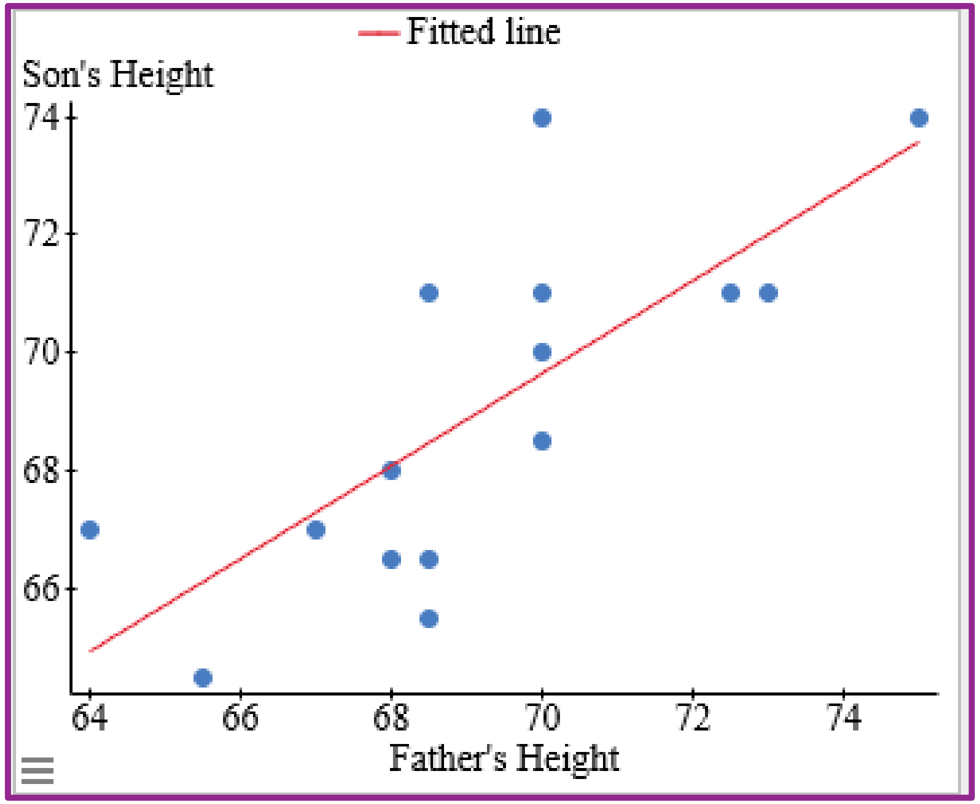
-0.782



0.037

EXAMPLE: REGRESSION ANALYSIS

- The accompanying table shows some data from a sample of heights of fathers and their sons. The scatterplot suggests a linear trend.



Father	Son
75	74
72.5	71
73	71
70	74
70	68.5
70	70
68	68
68	66.5
70	71
68.5	66.5
68.5	65.5
68.5	71
67	67
65.5	64.5
64	67

EXAMPLE: REGRESSION ANALYSIS

Father	Son
75	74
72.5	71
73	71

a. Find and report the regression equation for predicting the son's height from the father's height.

StatCrunch - Stat –
Regression – Simple Linear
Choose x and y variables

Son's Height=14.59 + 0.79 Father's Height

Simple linear regression results:
Dependent Variable: Son's Height
Independent Variable: Father's Height
Son's Height = 14.586898 + 0.78641938 Father's Height
Sample size: 15
R (correlation coefficient) = 0.75396424
R-sq = 0.56846207
Estimate of error standard deviation: 2.0060827

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	14.586898	13.167181	≠ 0	13	1.1078224	0.288
Slope	0.78641938	0.19003836	≠ 0	13	4.1382139	0.0012

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	68.916552	68.916552	17.124815	0.0012
Error	13	52.316781	4.0243678		
Total	14	121.23333			

EXAMPLE: REGRESSION ANALYSIS

b. Interpret the slope in the context of the problem.

Son's Height=14.59 + 0.79 Father's Height

For every additional inch on the Father's height, on average, the Son's height goes up by 0.79 inches.

Father	Son
75	74
72.5	71
73	71
70	74
70	68.5
70	70
68	68
68	66.5
70	71
68.5	66.5
68.5	65.5
68.5	71
67	67
65.5	64.5
64	67

EXAMPLE: REGRESSION ANALYSIS

c. Using the regression line, predict the height of a son whose father is 74 inches tall.

Son's Height=14.59 + 0.79 Father's Height

Son's Height=14.59 + 0.79 (74)

We predict the son will be 72.78 inches tall.

Note: Don't extrapolate!
Don't make predictions beyond the range of the observed data, because we are not sure that the linear trend will continue beyond the range of the data.

Father	Son
75	74
72.5	71
73	71
70	74
70	68.5
70	70
68	68
68	66.5
70	71
68.5	66.5
68.5	65.5
68.5	71
67	67
65.5	64.5
64	67

PROBABILITY

Theoretical Probability

- Long run relative frequencies – what would occur after infinitely many repetitions
  - Rules:

$$P(x) = \frac{Number\ of\ outcomes\ in\ x}{Number\ of\ outcomes\ possible}$$

$$0 \leq P(x) \leq 1$$

Empirical Probability

- Relative frequencies based on an experiment or on observations of a real life process
  - The Law of Large Numbers: The larger the number of repetitions, the closer the empirical probability will be to the theoretical probability

EXAMPLE: PROBABILITY

A person was trying to figure out the probability of getting two heads when flipping two coins. He flipped two coins 20 times, and in 4 of these 20 times, both coins landed heads. On the basis of this outcome, he claims that the probability of two heads is 4/20, or 20%.

Is this an example of an empirical probability or a theoretical probability? Explain.

This is an example of empirical probability because it is based on an experiment.

EXAMPLE: PROBABILITY

	Female	Male	All
No	41	51	92
Unsure	89	96	185
Yes	596	404	1000
All	726	551	1277

A poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table above shows a summary of the responses.

a. What is the probability that the person from the table is male?

$$\frac{551}{1277} = \mathbf{0.431}$$

EXAMPLE: PROBABILITY

	Female	Male	All
No	41	51	92
Unsure	89	96	185
Yes	596	404	1000
All	726	551	1277

A poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table above shows a summary of the responses.

b. What is the probability that the person said Yes?

$$\frac{1000}{1277} = \mathbf{0.783}$$

EXAMPLE: PROBABILITY

	Female	Male	All
No	41	51	92
Unsure	89	96	185
Yes	596	404	1000
All	726	551	1277

A poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table above shows a summary of the responses.

c. Are the event being male and the event saying Yes mutually exclusive? Why or why not?

The events are not mutually exclusive because a person chosen could be male and say yes.

EXAMPLE: PROBABILITY

	Female	Male	All
No	41	51	92
Unsure	89	96	185
Yes	596	404	1000
All	726	551	1277

A poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table above shows a summary of the responses.

d. What is the probability that a person is male and said Yes?

$$\frac{404}{1277} = \mathbf{0.316}$$

EXAMPLE: PROBABILITY

	Female	Male	All
No	41	51	92
Unsure	89	96	185
Yes	596	404	1000
All	726	551	1277

A poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table above shows a summary of the responses.

e. What is the probability that a person is male or said Yes?

To find the probability that a person is male or said yes, why should you subtract the probability that a person is male and said Yes from the sum as shown below?
P(male or Yes)=P(male)+P(Yes)-P(male and Yes)

$$\frac{551}{1277}+\frac{1000}{1277}-\frac{404}{1277}=\mathbf{0.898}$$

EXAMPLE: PROBABILITY

	Female	Male	All
No	41	51	92
Unsure	89	96	185
Yes	596	404	1000
All	726	551	1277

A poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table above shows a summary of the responses.

f. What is the probability that a randomly chosen person was male given that the person said Yes. In other words, what percentage of the people who said Yes were male?

$$\frac{404}{1000} = \mathbf{0.404}$$

$$= \mathbf{40.4\%}$$

EXAMPLE: PROBABILITY

	Female	Male	All
No	41	51	92
Unsure	89	96	185
Yes	596	404	1000
All	726	551	1277

A poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table above shows a summary of the responses.

g. Find the probability that a randomly chosen person who reported being Unsure was female. In other words, what percentage of the people who were Unsure were female?

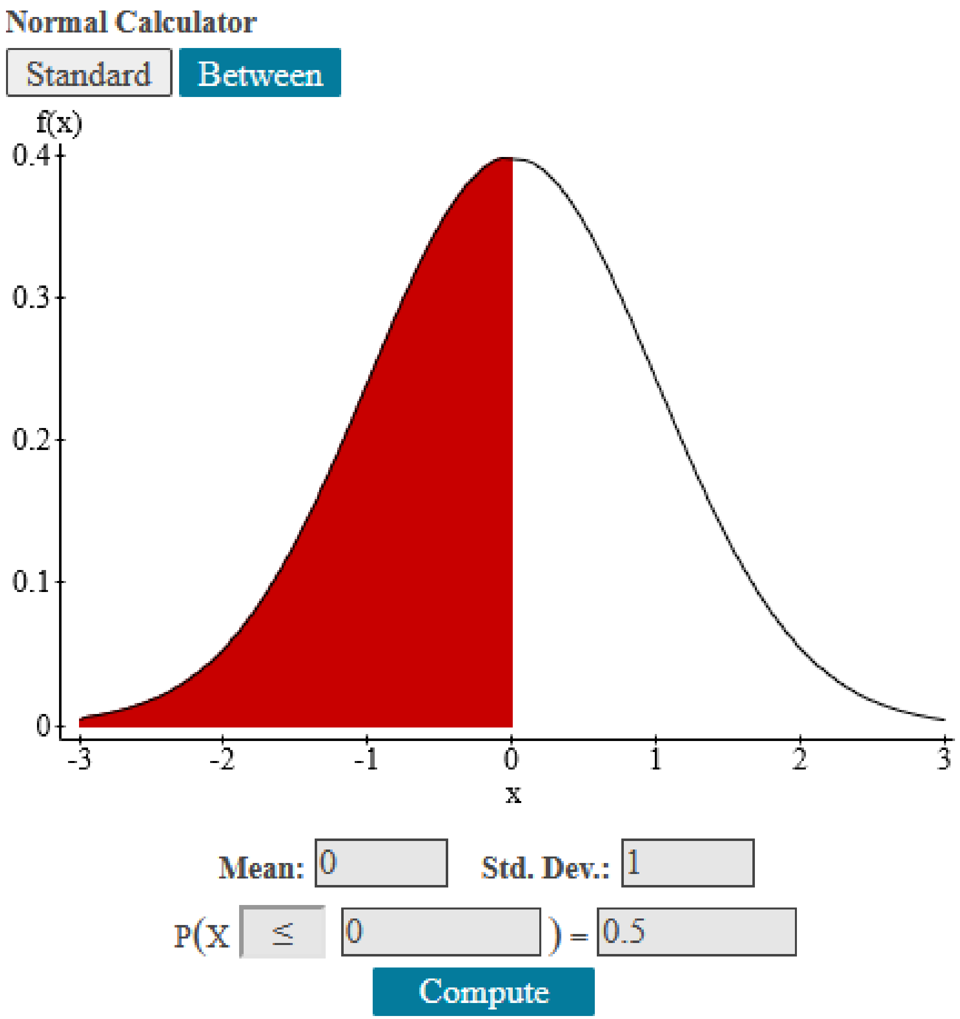
$$\frac{89}{185} = \mathbf{0.481}$$

$$= \mathbf{48.1}\%$$

NORMAL DISTRIBUTION

- Finding probabilities by finding the area under the Normal Curve.
- Percentiles: Area based on percentage – used to work backwards
  - 90th Percentile means 90% of the data is below that value or
    90% of the area under the Normal curve is to its left.

StatCrunch – Stat – Calculators – Normal

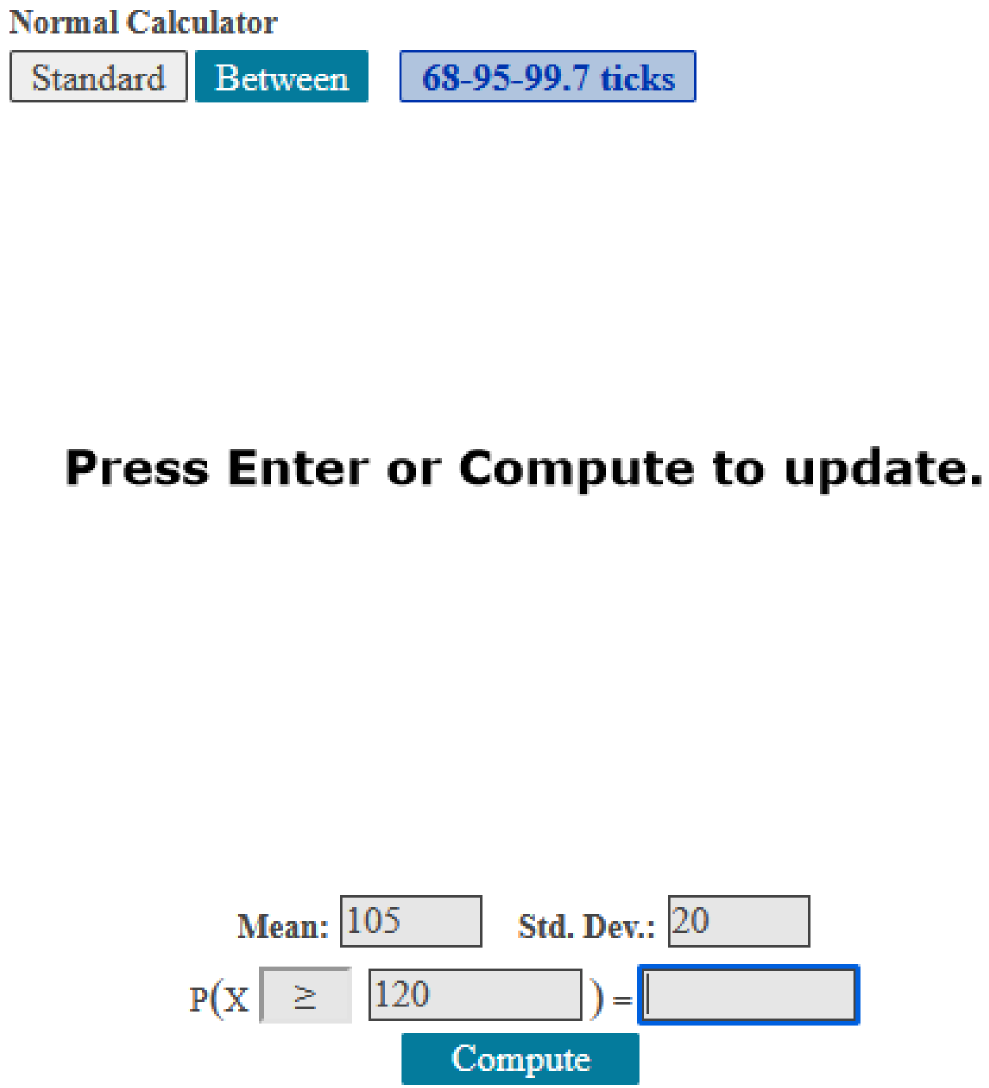


EXAMPLE: NORMAL DISTRIBUTIONS

Assume that adults have IQ scores that are Normally distributed with a mean of $\mu = 105$ and a standard deviation $\sigma = 20$. Find the probability that a randomly selected adult has an IQ of 120 or above.

StatCrunch – Stat – Calculators – Normal

Fill in mean, standard deviation, and score.

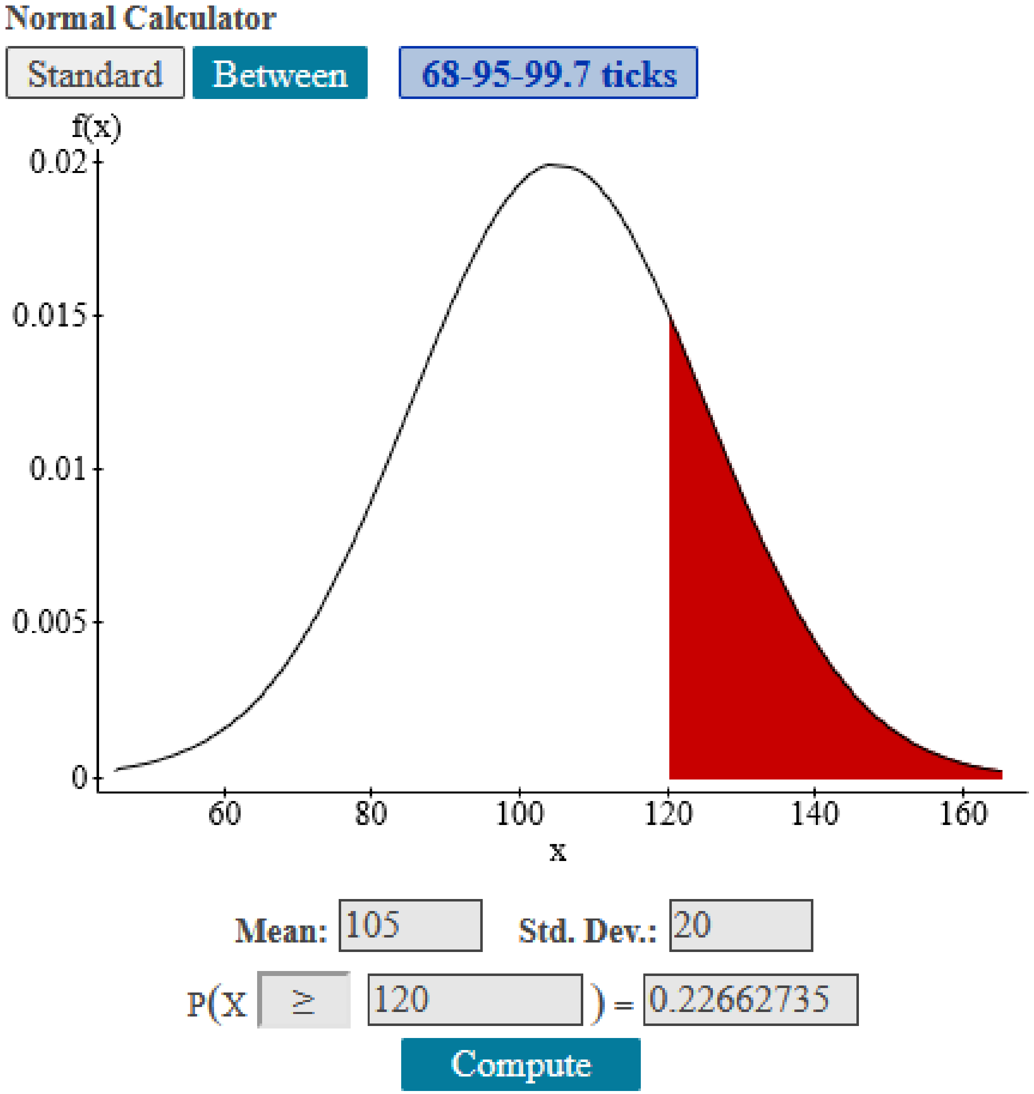


EXAMPLE: NORMAL DISTRIBUTIONS

Assume that adults have IQ scores that are Normally distributed with a mean of $\mu = 105$ and a standard deviation $\sigma = 20$. Find the probability that a randomly selected adult has an IQ of 120 or above.

StatCrunch – Stat – Calculators – Normal

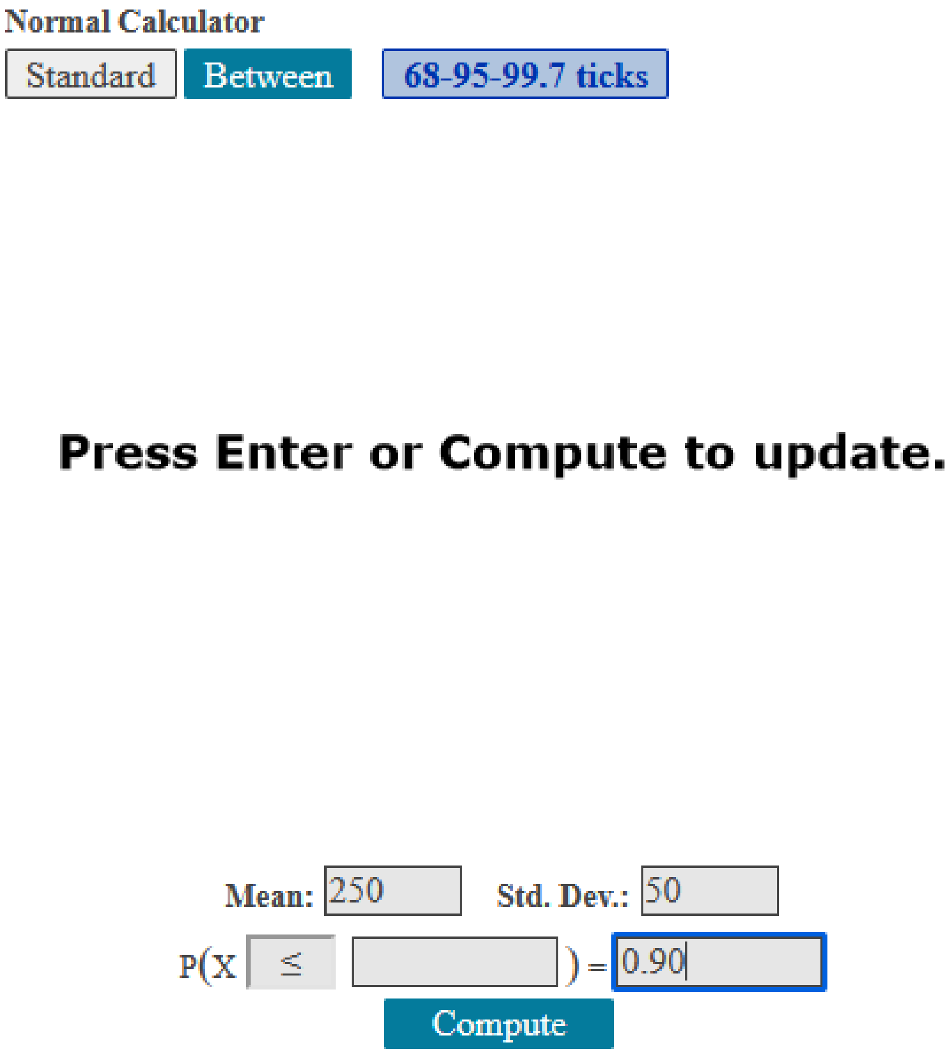
The probability that a randomly selected adult has an IQ of 120 or more is 0.227.



EXAMPLE: NORMAL DISTRIBUTIONS

The average birth weight of elephants is 250 pounds. Assume that the distribution of birth weights is Normal with a standard deviation of 50 pounds. Find the birth weight of elephants at the 90th percentile

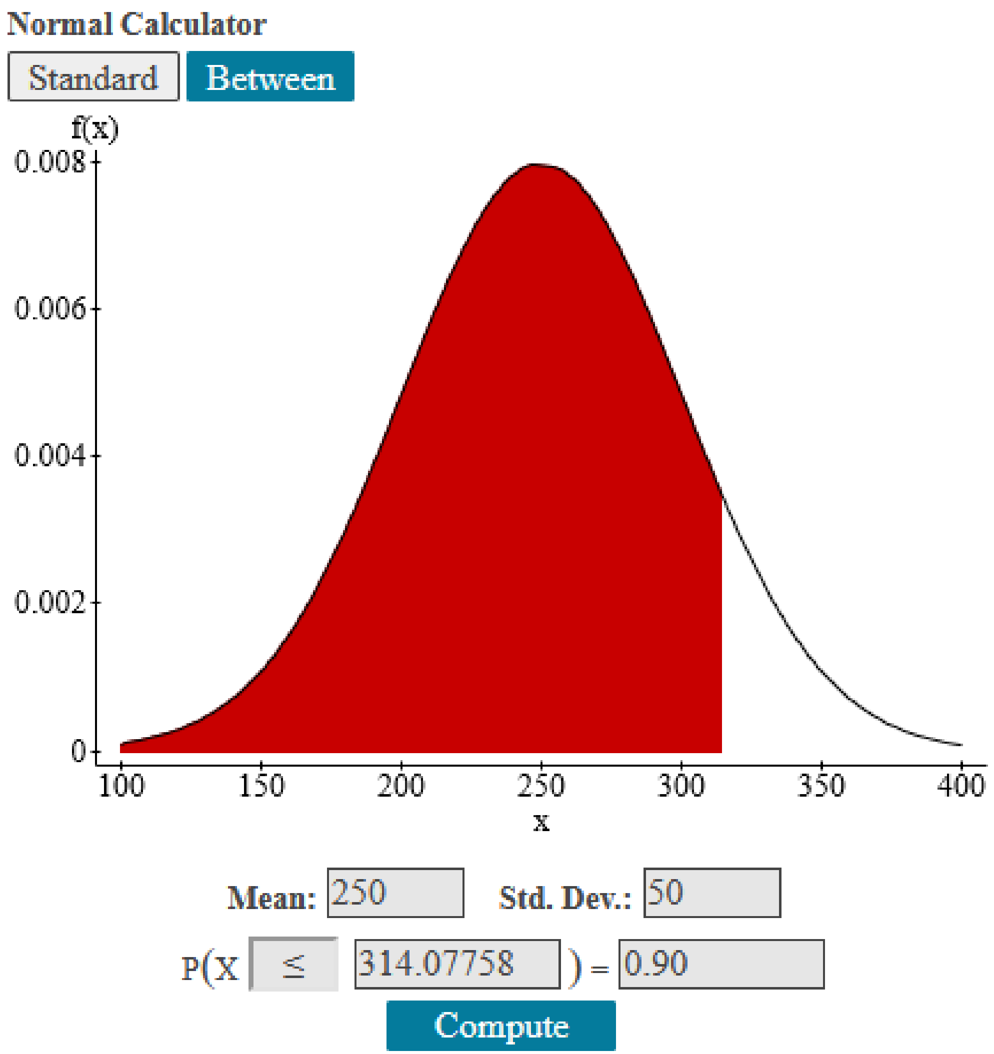
StatCrunch – Stat – Calculator – Normal
Fill in mean, standard deviation, and probability.



EXAMPLE: NORMAL DISTRIBUTIONS

The average birth weight of elephants is 250 pounds. Assume that the distribution of birth weights is Normal with a standard deviation of 50 pounds. Find the birth weight of elephants at the 90th percentile

StatCrunch – Stat – Calculator – Normal
Fill in mean, standard deviation, and probability.



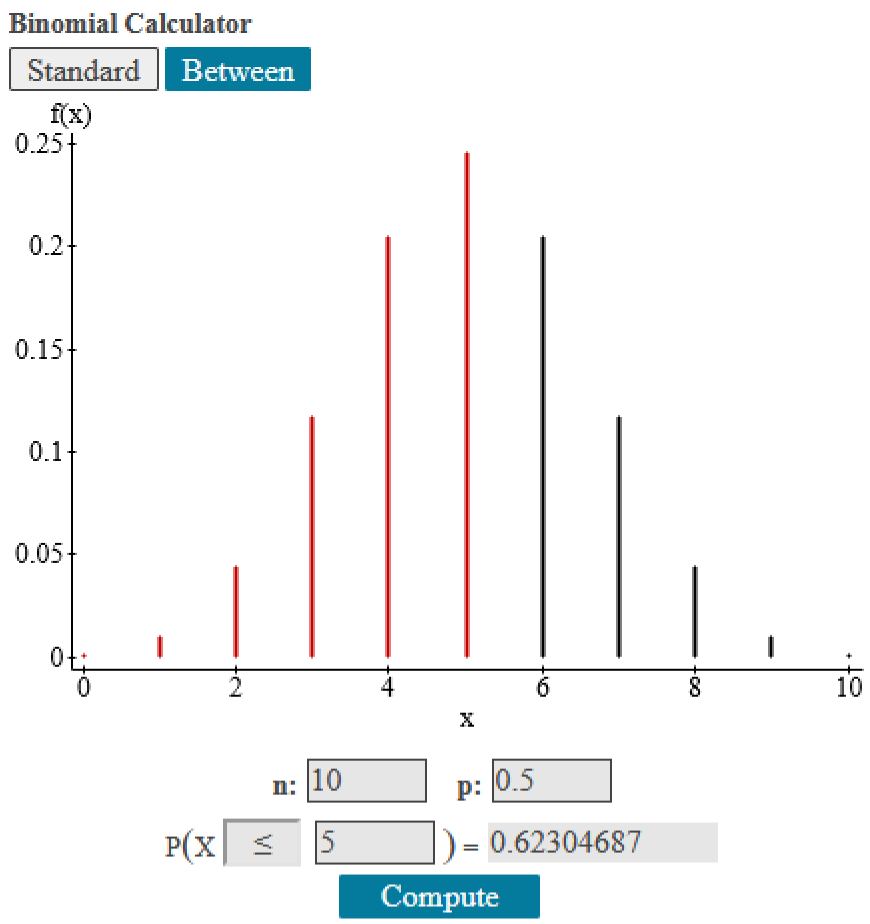
The birth weight of elephants at the 90th percentile is 314 pounds.

BINOMIAL MODEL

- Requires:
  - Fixed number of trials (n)
  - Only two possible outcomes for each trial (success vs failure)
  - Probability of success (p) is the same for each trial
  - Trials are independent
  - x is the number of successes out of n trials

  - Expected Value of Binomial Distribution:
    $\mu = np$
  - Standard Deviation of Binomial Distribution:
    $\sigma = \sqrt{np(1-p)}$

StatCrunch – Stat – Calculators – Binomial



BINOMIAL MODEL

- Because the binomial probability distribution models probability of discrete random variables, we have to pay attention to the wording!
  - "Exactly 6" $\rightarrow P(X = 6)$
  - "More than 6" $\rightarrow P(X > 6)$
  - "At least 6" $\rightarrow P(X \geq 6)$
  - "6 or more" $\rightarrow P(X \geq 6)$
  - "Less than 6" $\rightarrow P(X < 6)$
  - "Fewer than 6" $\rightarrow P(X < 6)$
  - "At most 6" $\rightarrow P(X \leq 6)$

EXAMPLE: BINOMIAL MODEL

A recent poll indicated that about 74% of U.S. households had access to a high-speed Internet connection.

a. Suppose 100 households were randomly selected from the United States. How many of the households would you expect to have access to a high-speed Internet connection?

$$Expected\ value = np \rightarrow 100(0.74) = 74$$

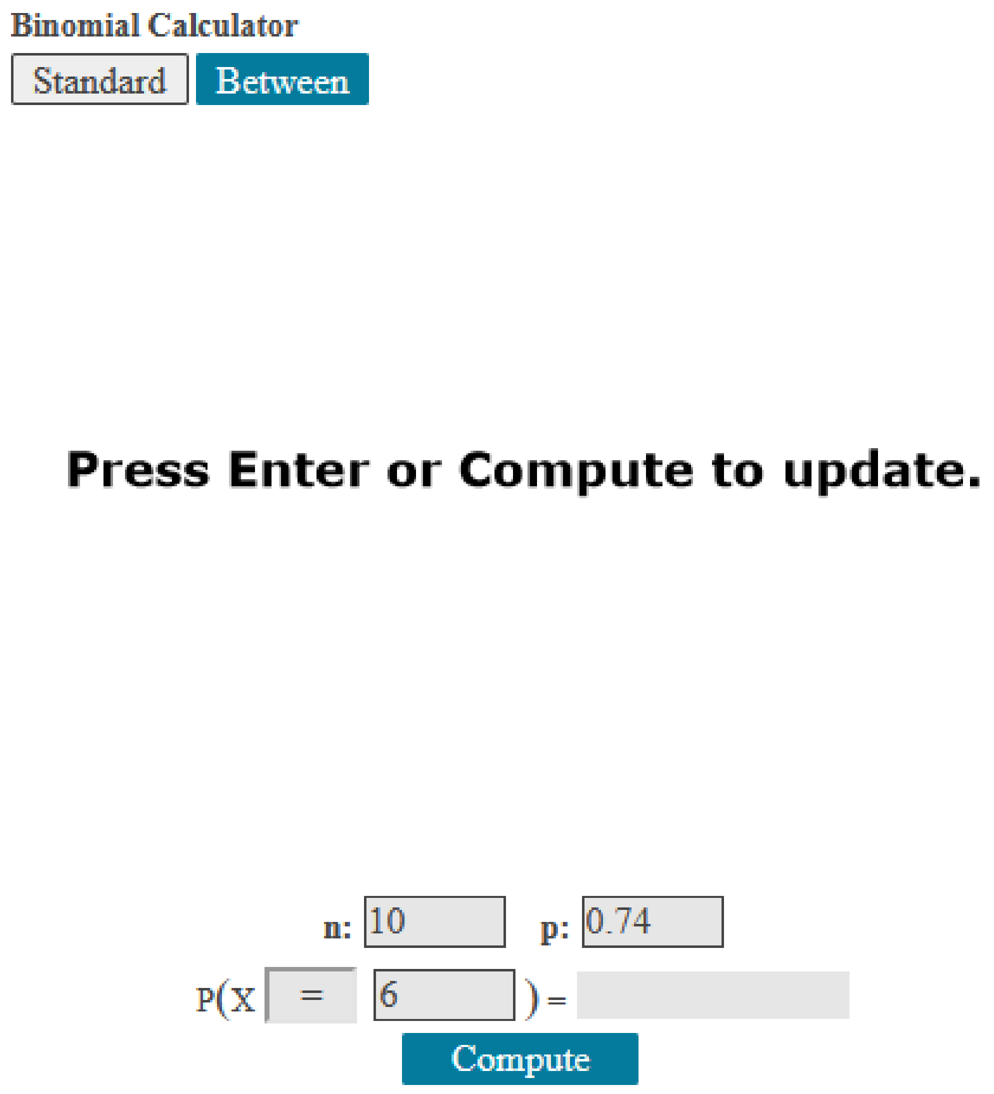
You would expect 74 households to have access to high-speed internet connection.

EXAMPLE: BINOMIAL

A recent poll indicated that about 74% of U.S. households had access to a high-speed Internet connection.

b. If 10 households are selected randomly, what is the probability that exactly 6 have high-speed access?

StatCrunch – Stat – Calculators – Binomial

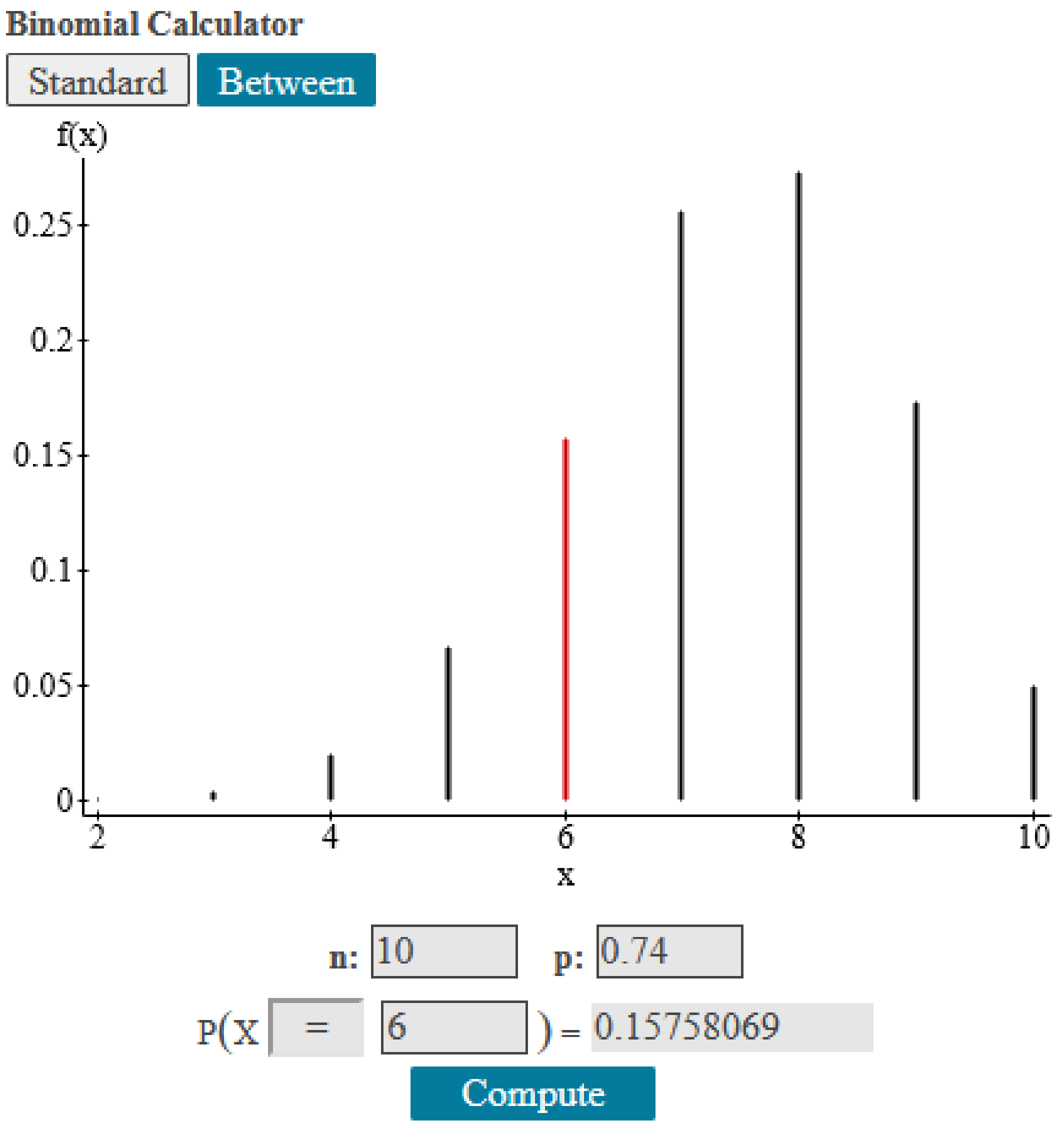


EXAMPLE: BINOMIAL

A recent poll indicated that about 74% of U.S. households had access to a high-speed Internet connection.

b. If 10 households are selected randomly, what is the probability that exactly 6 have high-speed access?

StatCrunch – Stat – Calculators – Binomial



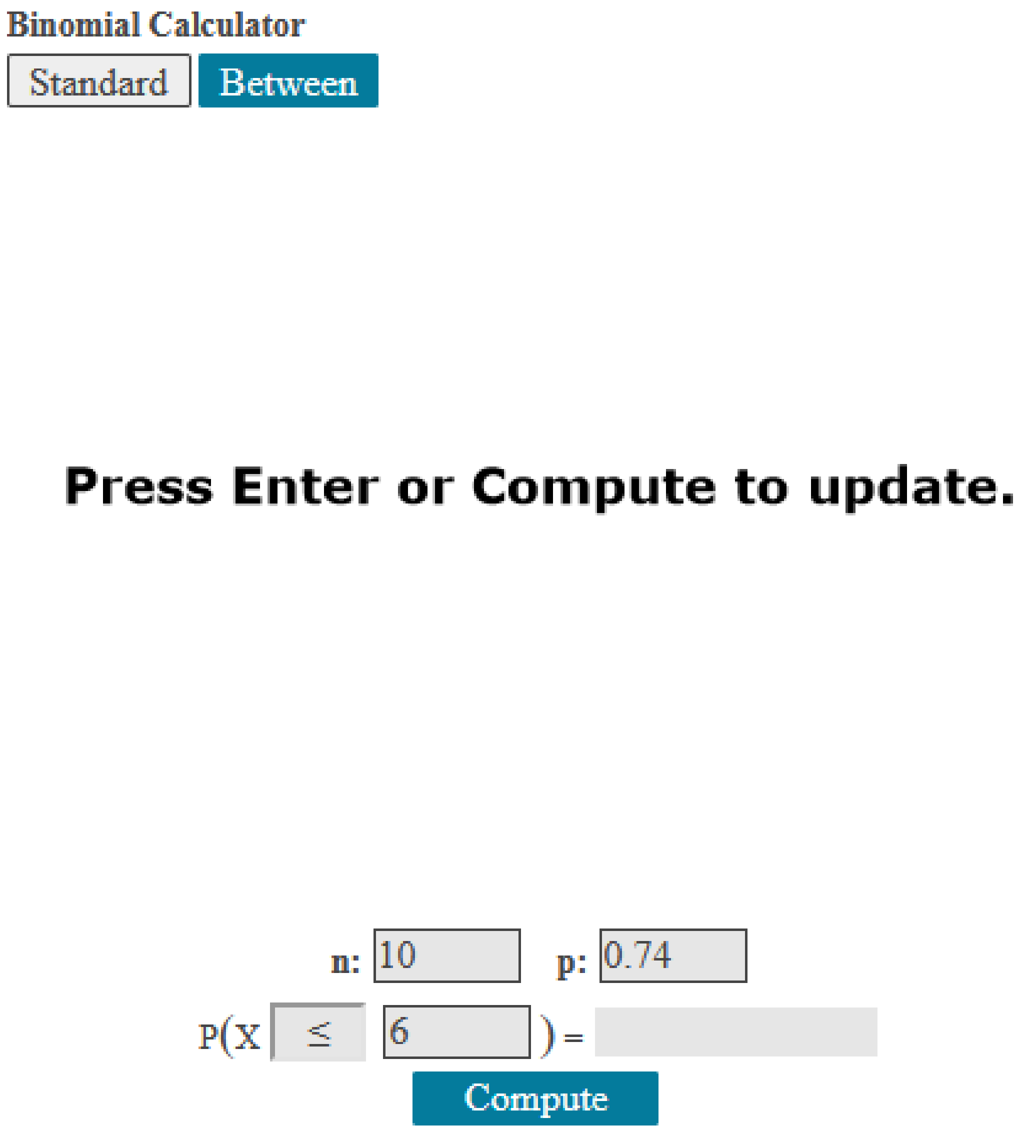
The probability that exactly 6 households have high-speed access is 0.158

EXAMPLE: BINOMIAL

A recent poll indicated that about 74% of U.S. households had access to a high-speed Internet connection.

b. If 10 households are selected randomly, what is the probability that 6 or fewer have high-speed access?

StatCrunch – Stat – Calculators – Binomial



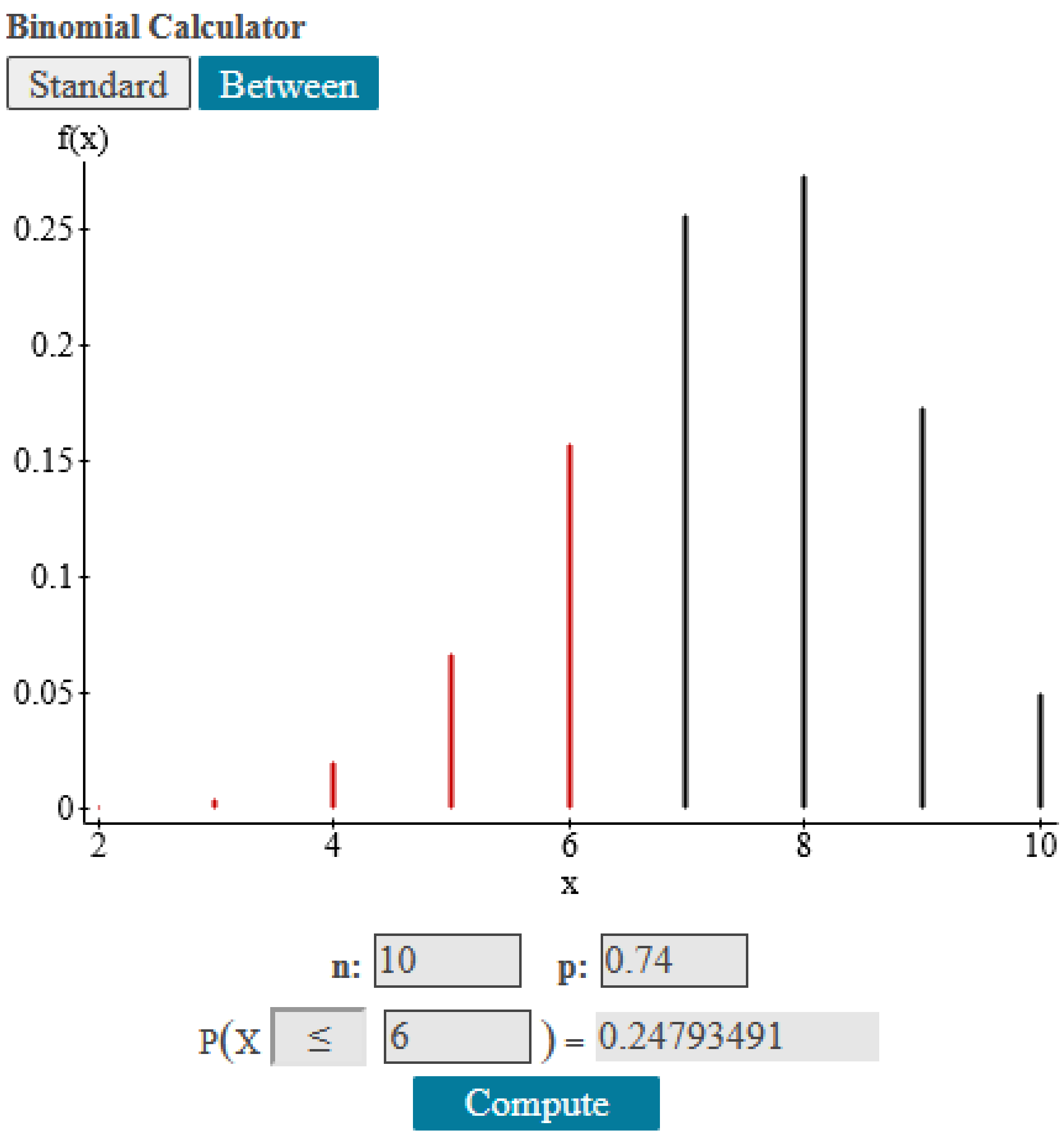
EXAMPLE: BINOMIAL

A recent poll indicated that about 74% of U.S. households had access to a high-speed Internet connection.

b. If 10 households are selected randomly, what is the probability that 6 or fewer have high-speed access?

StatCrunch – Stat – Calculators – Binomial

The probability that exactly 6 or fewer households have high-speed access is 0.248



CENTER FOR ACADEMIC SUPPORT

Hearnes 213
816-271-4524