SUMMARY STATISTICS

STEP-BY-STEP EXAMPLES

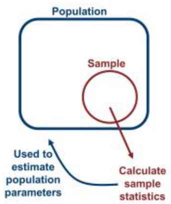
Let's do some examples

Location
Spread
Shape

Estimating data values

We usually care about the **parameters** of a **population** (all the individuals of interest), but we *usually* can't measure them all.

More often, we measure the values for a subset and calculate the **statistics** of a **sample** to **estimate** the population parameters.



Statistics of location, spread, and shape

Location: Mean, Median, Mid-range, Mode

Spread: Range, Interquartile range, Sum of squares, Variance, Standard deviation, Coefficient of variation

Shape: Skewness, Kurtosis

See the companion video linked below, and at the end, for more detail about what these represent

← Slightly different equations for population parameters and sample statistics (due to estimation biases for variance)

Data set: numbers of birds in 8 trees (Location)

1, 4, 5, 7, 3, 9, 7, 12

Best order to do these: median, mode, mid-range, mean.

First step. Reorder all the values from smallest to largest.

Original data set: 1 4 5 7 3 9 7 12

Ordered data set:

Data set: numbers of birds in 8 trees

1, 4, 5, 7, 3, 9, 7, 12

Best order to do these: median, mode, mid-range, mean.

First step. Reorder all the values from smallest to largest.

Original data set: 1 4 5 7 3 9 7 12

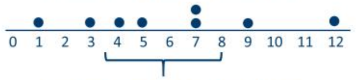
Ordered data set: 1 3 4 5 7 7 9 12

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12

IQR: this is the difference between Q3 and Q1.

$$IQR = Q3 - Q1 = 8 - 3.5 = \mathbf{4.5}$$



IQR = width of middle 50% of data

Data set: numbers of birds in 8 trees

1, 4, 5, 7, 3, 9, 7, 12

Step 1. divide data set into bottom and top halves.

Ordered data set: 1 3 4 5 | 7 7 9 12

Divided data set: 1 3 4 5   7 7 9 12

Compute medians of each half.

$$Q1 = \frac{3+4}{2} = \frac{7}{2} = 3.5 \qquad Q3 = \frac{7+9}{2} = \frac{16}{2} = 8$$

Ordered data set: numbers of birds in 8 trees (Spread)

1, 3, 4, 5, 7, 7, 9, 12

Range: this is difference between largest and smallest. $range = 12 - 1 = 11$

Interquartile range: this is difference between Q1 and Q3 so we have to compute the quartiles first.

Step 1: divide data set into bottom and top halves
Step 2: compute medians for each half, Q1 and Q3
Step 3: compute difference between Q3 and Q1

Ordered data set: numbers of birds in 8 trees

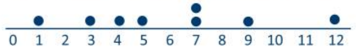
1, 3, 4, 5, 7, 7, 9, 12

Mean: this is the sum divided by the number of values, n=8.

$$Mean = \frac{1+3+4+5+7+7+9+12}{8} = \frac{48}{8} = 6$$

Median = 6   Mid-range = 6.5   Mode = 7   Mean= 6

All similar, suggests no extreme asymmetry or outliers.



Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12

Median: the data set has an even number of values, n=8, so we find the middle two and take their mean.
Mode: since the data set is ordered, we look for the longest string of consecutive identical values.
Mid-range: this is halfway between the smallest and largest value (the mean of the two values)

1, 3, 4, 5, 7, 7, 9, 12

$$Mode = 7$$

$$Median = \frac{5+7}{2} = \frac{12}{2} = 6$$

$$Mid-range = \frac{1+12}{2} = \frac{13}{2} = 6.5$$

A note on calculating quartiles (Odd n)

A note on calculating quartiles. If the data set has an odd number of values, the middle value is the median and there are 3 simple options for calculating quartiles.

Option 1. Include the median in both the bottom and top half of the data.

Option 2. Do not include the median in either the bottom and top half of the data.

Option 3. Calculate Q1 and Q3 using a weighted average of the data points

Calculating quartiles with an odd number of values. (Odd n)

Option 1. Include the median in both the bottom and top half of the data.

Ordered data set: 1 3 4 5 6 7 7 9 12

Divided data set: 1 3 4 5 6 | 6 7 7 9 12

Compute medians of each half.

$$Q1 = 4 \qquad Q3 = 7$$

$$IQR = 7 - 4 = 3$$

Calculating quartiles with an odd number of values. (Odd n)

Option 2. Do not include the median in either the bottom or top half of the data.

Ordered data set: 1 3 4 5   7 7 9 12

Divided data set: 1 3 4 5 | 7 7 9 12

Compute medians of each half.

$$Q1 = \frac{3+4}{2} = \frac{7}{2} = 3.5 \qquad Q3 = \frac{7+9}{2} = \frac{16}{2} = 8$$

$$IQR = 8 - 3.5 = 4.5$$

Calculating quartiles with an odd number of values. (Odd n)

Option 3. Calculate Q1 and Q3 using a weighted average of the data points

For 4n+1 values: Q1 is 0.25 n^{th} value + 0.75 $(n+1)^{th}$ value
Q3 is 0.75 $(3n+1)^{th}$ value + 0.25 $(3n+2)^{th}$ value

For 4n+3 values: Q1 is 0.75 $(n+1)^{th}$ value + 0.25 $(n+2)^{th}$ value
Q3 is 0.25 $(n+2)^{th}$ value + 0.75 $(n+3)^{th}$ value

Ordered data set: 1 3 4 5 6 7 7 9 12

9 values so n=2

$$Q1 = 0.25(2^{nd}) + 0.75(3^{rd}) = 0.25(3) + 0.75(4) = 3.75$$

$$Q3 = 0.75(7^{th}) + 0.25(8^{th}) = 0.75(7) + 0.25(9) = 7.5$$

$$IQR = 7.5 - 3.75 = \mathbf{3.75}$$

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12

Sum of squares: sum of squared differences from the mean, $\bar{x} = 6$.

$$SS = \sum_{i=1}^{8}(x_i - \bar{x})^2 = (1-6)^2 + (3-6)^2 + (4-6)^2 + (5-6)^2 + (7-6)^2 + (7-6)^2 + (9-6)^2 + (12-6)^2$$

$$= (-5)^2 + (-3)^2 + (-2)^2 + (-1)^2 + (1)^2 + (1)^2 + (3)^2 + (6)^2$$

$$= 25 + 9 + 4 + 1 + 1 + 1 + 9 + 36 = \mathbf{86}$$

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12   If the data set is a ***population***

$$Kurtosis = \frac{\left(\frac{\sum(x_i-\bar{x})^4}{n}\right)}{\sigma^4} = \frac{\left(\frac{\sum(x_i-6)^4}{8}\right)}{3.279^4} = \frac{1}{8(3.279^4)}\sum(x_i-6)^4$$

$$= \frac{1}{8(3.279^4)}[(1-6)^4 + (3-6)^4 + (4-6)^4 + (5-6)^4 + (7-6)^4 + (7-6)^4 + (9-6)^4 + (12-6)^4]$$

$$= \frac{1}{8(3.279^4)}[(-5)^4 + (-3)^4 + (-2)^4 + (-1)^4 + (1)^4 + (1)^4 + (3)^4 + (6)^4]$$

$$= \frac{1}{8(3.279^4)}[625 + 81 + 16 + 1 + 1 + 1 + 81 + 1296] = \frac{2102}{942.817} = \mathbf{2.229}$$

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12   If the data set is a ***population***

$$Skewness = \frac{\left(\frac{\sum(x_i-\bar{x})^3}{n}\right)}{\sigma^3} = \frac{\left(\frac{\sum(x_i-6)^3}{8}\right)}{3.279^3} = \frac{1}{8(3.279^3)}\sum(x_i-6)^3$$

$$= \frac{1}{8(3.279^3)}[(1-6)^3 + (3-6)^3 + (4-6)^3 + (5-6)^3 + (7-6)^3 + (7-6)^3 + (9-6)^3 + (12-6)^3]$$

$$= \frac{1}{8(3.279^3)}[(-5)^3 + (-3)^3 + (-2)^3 + (-1)^3 + (1)^3 + (1)^3 + (3)^3 + (6)^3]$$

$$= \frac{1}{8(3.279^3)}[-125 - 27 - 8 - 1 + 1 + 1 + 27 + 216] = \frac{84}{282.042} = \mathbf{0.298}$$

Ordered data set: numbers of birds in 8 trees (Shape)

1, 3, 4, 5, 7, 7, 9, 12

If the data set is a ***population***

Skewness

$$Skewness = \frac{\left(\frac{\sum(x_i-\bar{x})^3}{n}\right)}{\sigma^3}$$

Kurtosis

$$Kurtosis = \frac{\left(\frac{\sum(x_i-\bar{x})^4}{n}\right)}{\sigma^4}$$

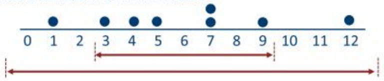
The skewness and kurtosis are rarely studied for their own sake. They are usually calculated to see if the distribution is normal.

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12

$\sigma = 3.279$
$s = 3.505$

Useful property of the standard deviation:
For normal distributions, approximately 66% of the values are within 1 standard deviation of the mean.
For normal distributions, approximately 95% of the values are within 2 standard deviations of the mean.



Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12

	Population	Sample
Variance:	SS divided by n	SS divided by n-1
Standard deviation	Square root of variance	
Coefficient of variation	Standard deviation/mean x 100	

$$\sigma^2 = \frac{SS}{n} = \frac{86}{8} = \mathbf{10.75} \qquad s^2 = \frac{SS}{n} = \frac{86}{7} = \mathbf{12.286}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{10.75} = \mathbf{3.279} \qquad s = \sqrt{s^2} = \sqrt{12.286} = \mathbf{3.505}$$

$$CV = \frac{3.279}{6} \times 100 = \mathbf{54.65\%} \qquad CV = \frac{3.505}{6} \times 100 = \mathbf{58.42\%}$$

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12

If the data set is a ***sample***

we can use the sums of cubes and 4th powers from previous $\sum(x_i-6)^3 = 84$, $\sum(x_i-6)^4 = 2102$

$$Skewness = \frac{\left(\frac{\sum(x_i-\bar{x})^3}{n}\right)}{s^3} \quad or \quad \frac{\sqrt{n(n-1)}}{n-2}\frac{\left(\frac{\sum(x_i-\bar{x})^3}{n}\right)}{\sigma^3}$$

$$Excess\ Kurtosis = \frac{(n+1)n}{(n-1)(n-2)}\frac{\left(\frac{\sum(x_i-\bar{x})^4}{n}\right)}{s^4} - 3\frac{(n-1)^2}{(n-2)(n-3)}$$

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12   If the data set is a ***sample***

$$Skewness = \frac{\left(\frac{\sum(x_i-\bar{x})^3}{n}\right)}{s^3} = \frac{\left(\frac{84}{8}\right)}{3.505^3} = \mathbf{0.244}$$

The sums of cubes from previous $\sum(x_i-6)^3 = 84$

$$Skewness = \frac{\sqrt{n(n-1)}}{n-2}\frac{\left(\frac{\sum(x_i-\bar{x})^3}{n}\right)}{\sigma^3} = \frac{\sqrt{8(7)}}{6}\frac{\left(\frac{84}{8}\right)}{3.279^3} = \mathbf{0.371}$$

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12   If the data set is a ***sample***

$$Excess\ Kurtosis = \frac{(n+1)n}{(n-1)(n-2)}\frac{\left(\frac{\sum(x_i-\bar{x})^4}{n}\right)}{s^4} - 3\frac{(n-1)^2}{(n-2)(n-3)}$$

$$= \frac{(9)8}{(7)(6)}\frac{\left(\frac{2102}{8}\right)}{3.505^4} - 3\frac{(7)^2}{(6)(5)} = \mathbf{-1.915}$$

The sums of the 4th powers from previous $\sum(x_i-6)^4 = 2102$

Ordered data set: numbers of birds in 8 trees

1, 3, 4, 5, 7, 7, 9, 12

If the data set is a ***population***

$Skewness = 0.298$

$Kurtosis = 0.229$

$Excess\ Kurtosis = -0.771$

If the data set is a ***sample***

$Skewness = 0.244$

$Skewness = 0.371$

$Excess\ Kurtosis = -1.915$

				If population parameters		If sample statistics
Location	Mean	6				
	Median	6				
	Mid-range	6.5				
	Mode	7				
Spread	Range				13	
	Interquartile range				8.5	
	Sum of squares				86	
	Variance			10.75		12.286
	Standard deviation			3.279		3.505
	Coefficient of variation			54.65%		58.42%
Shape	Skewness			0.298		0.244 or 0.371
	Kurtosis (excess)			2.229 (-0.771)		(-1.915)

StatsExamples.com