Introduction

Statistics

- Statistics is the mathematical science behind the problem "what can I know about a population if I'm unable to reach every member?"

Statistics

- If we could measure the height of every resident of Australia, then we could make a statement about the average height of Australians at the time we took our measurement.
- This is where random sampling comes in.

Statistics

- If we take a reasonably sized random sample of Australians and measure their heights, we can form a statistical inference about the population of Australia.
- Probability helps us know how sure we are of our conclusions!

Data

What is Data?

- Data = the collected observations we have about something.
- Data can be continuous:
  "What is the stock price?"
- or categorical:
  "What car has the best repair history?"

Why Data Matters

- Helps us understand things as they are:

"What relationships if any exist between two events?"

"Do people who eat an apple a day enjoy fewer doctor's visits than those who don't?"

Why Data Matters

- Helps us predict future behavior to guide business decisions:

  "Based on a user's click history which ad is more likely to bring them to our site?"

Visualizing Data

- Compare a table:

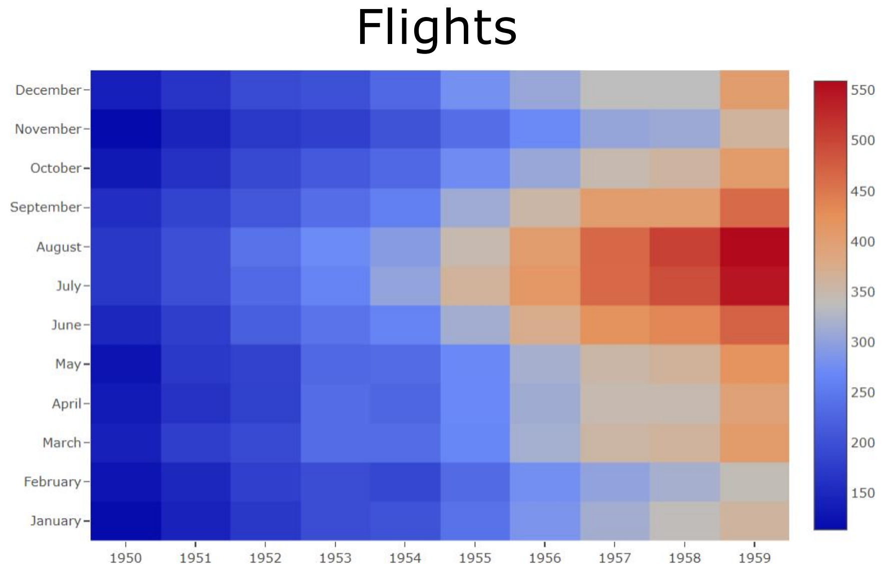
Flights

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	**year**	**month**	**passengers**		**year**	**month**	**passengers**		**year**	**month**	**passengers**		**year**	**month**	**passengers**
2	1950	January	115		1952	July	230		1955	January	242		1957	July	465
3	1950	February	126		1952	August	242		1955	February	233		1957	August	467
4	1950	March	141		1952	September	209		1955	March	267		1957	September	404
5	1950	April	135		1952	October	191		1955	April	269		1957	October	347
6	1950	May	125		1952	November	172		1955	May	270		1957	November	305
7	1950	June	149		1952	December	194		1955	June	315		1957	December	336
8	1950	July	170		1953	January	196		1955	July	364		1958	January	340
9	1950	August	170		1953	February	196		1955	August	347		1958	February	318
10	1950	September	158		1953	March	236		1955	September	312		1958	March	362
11	1950	October	133		1953	April	235		1955	October	274		1958	April	348
12	1950	November	114		1953	May	229		1955	November	237		1958	May	363
13	1950	December	140		1953	June	243		1955	December	278		1958	June	435
14	1951	January	145		1953	July	264		1956	January	284		1958	July	491
15	1951	February	150		1953	August	272		1956	February	277		1958	August	505
16	1951	March	178		1953	September	237		1956	March	317		1958	September	404
17	1951	April	163		1953	October	211		1956	April	313		1958	October	359
18	1951	May	172		1953	November	180		1956	May	318		1958	November	310

Not much can be gained by reading it.

Visualizing Data

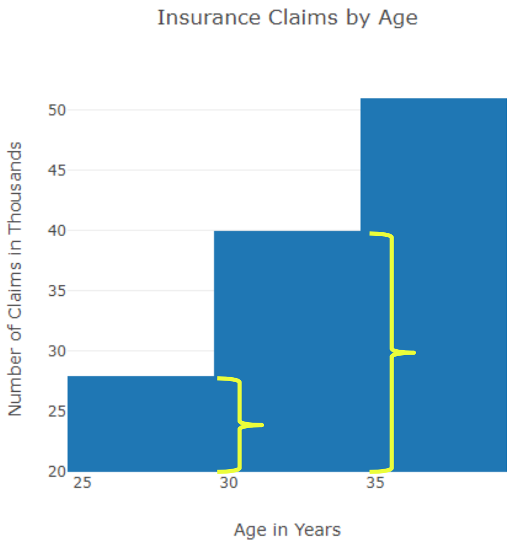
- to a graph:

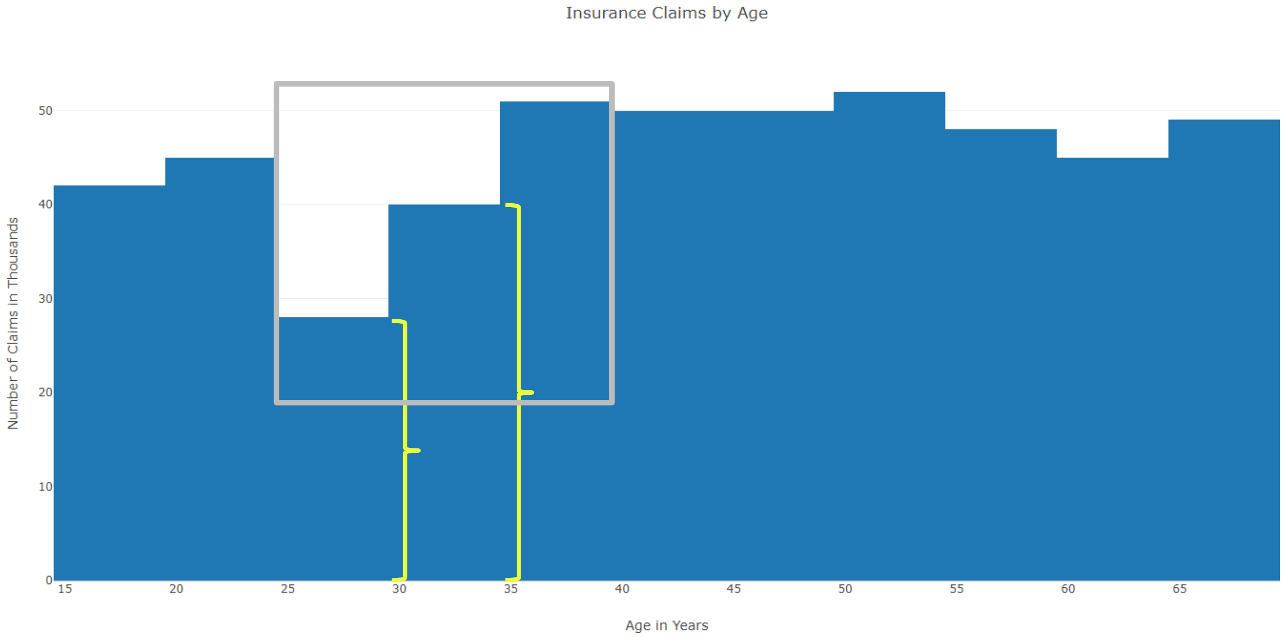


The graph uncovers two distinct trends - an increase in passengers flying over the years and a greater number of passengers flying in the summer months.

Analyze Visualizations Critically!

- Graphs can be misleading:





Measuring Data

Levels of Measurement

Nominal

- Predetermined categories
- Can’t be sorted

Animal classification (*mammal fish reptile*)

Political party (*republican democrat independent*)

Levels of Measurement

Ordinal

- Can be sorted
- Lacks scale

Survey responses

Levels of Measurement

Interval

- Provides scale
- Lacks a “zero” point

Temperature

Levels of Measurement

Ratio

- Values have a true zero point

Age, weight, salary

Population vs. Sample

- Population = every member of a group
- Sample = a subset of members that time and resources allow you to measure

Mathematical Symbols & Syntax

Symbol/Expression	Spoken as	Description
x^2	x squared	x raised to the second power $x^2 = x \times x$
x_i	x-sub-i	a subscripted variable (the subscript acts as a label)
$x!$	x factorial	$4! = 4 \times 3 \times 2 \times 1$
$\bar{x}$	x bar	symbol for the sample mean
μ	"mew"	symbol for the population mean (Greek lowercase letter mu)
Σ	sigma	syntax for writing sums (Greek capital letter sigma)

Exponents

$$\boldsymbol{x}^{\mathbf{5}} = \underset{1}{x} \times \underset{2}{x} \times \underset{3}{x} \times \underset{4}{x} \times \underset{5}{x}$$

EXAMPLE: $3^4 = 3 \times 3 \times 3 \times 3 = 81$

Exponents – special cases

$$x^{-3} = \frac{1}{x \times x \times x}$$

EXAMPLE: $2^{-3} = \frac{1}{2\times2\times2} = \frac{1}{8} = 0.125$

$$x^{\left(\frac{1}{n}\right)} = \sqrt[n]{x}$$

EXAMPLE: $8^{\left(\frac{1}{3}\right)} = \sqrt[3]{8} = 2$

Factorials

$$\boldsymbol{x}! \;= x \times (x-1) \times (x-2) \times \cdots \times 1$$

EXAMPLE: $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$

EXAMPLE: $\frac{5!}{3!} = \frac{5 \times 4 \times \cancel{3 \times 2 \times 1}}{\cancel{3 \times 2 \times 1}} = 5 \times 4 = 20$

Simple Sums

$$\sum_{x=1}^{n} x = 1 + 2 + 3 + \cdots + n$$

EXAMPLE: $\sum_{x=1}^{4} x = 1 + 2 + 3 + 4 = 10$

EXAMPLE: $\sum_{x=1}^{4} x^2 = 1 + 4 + 9 + 16 = 30$

Series Sums

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

EXAMPLE: $x = \{5,3,2,8\}$

$$n = \#\, elements\ in\ x = 4$$

$$\sum_{i=1}^{4} x_i = 5 + 3 + 2 + 8 = 18$$

Equation Example

- Formula for calculating a sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Read out loud:

“$\boldsymbol{x}$ bar (the symbol for the sample mean) is equal to the sum (indicated by the Greek letter sigma) of all the $\boldsymbol{x}$-sub-$\boldsymbol{i}$ values in the series as $\boldsymbol{i}$ goes from 1 to the number $\boldsymbol{n}$ items in the series divided by $\boldsymbol{n}$."

Equation Example

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

1. Start with a series of values:

   {7 8 9 10}

2. Assign placeholders to each item

   {7 8 9 10}

   1 2 3 4 n=4

3. These become x_1 x_2 etc.

   $x_1 = 7 \quad x_2 = 8 \quad x_3 = 9 \quad x_4 = 10$

Equation Example

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

4. Plug these into the equation:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 \ldots + x_n}{n}$$

$$= \frac{7 + 8 + 9 + 10}{4} = \frac{34}{4} = 8.5$$

Measurement Types
Central Tendency

Measurements of Data

- “What was the average return?”

  Measures of Central Tendency

- “How far from the average did individual values stray?”

  Measures of Dispersion

Measures of Central Tendency (mean, median, mode)

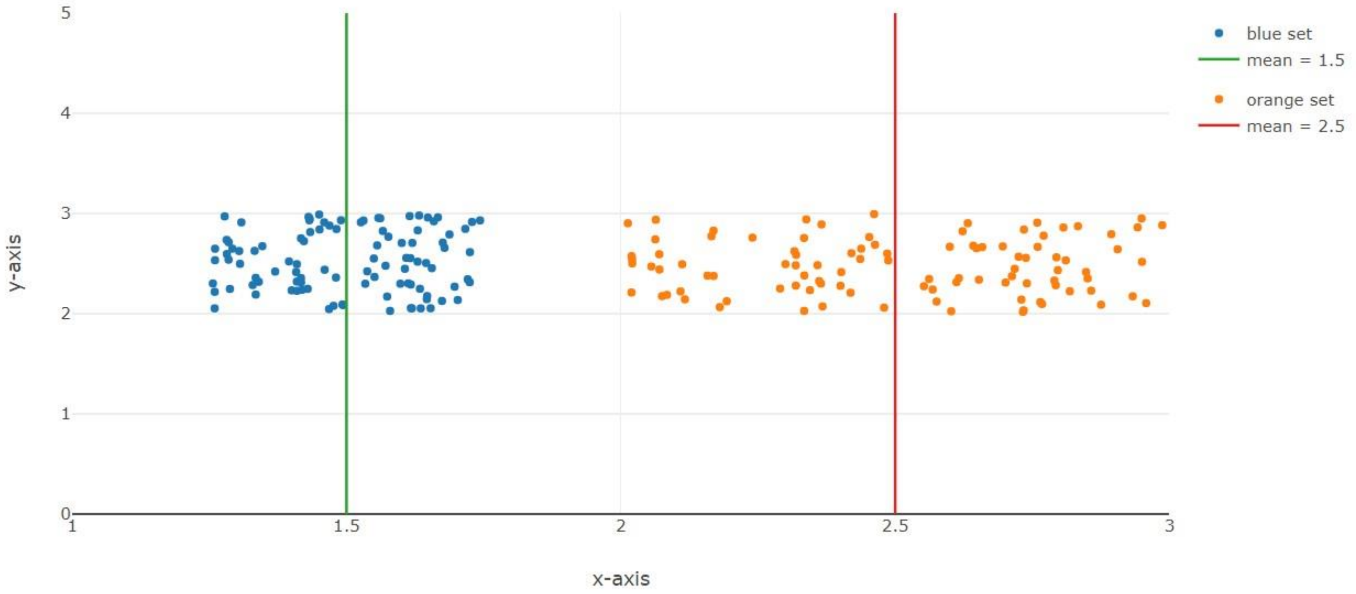
- Describe the "location" of the data
- Fail to describe the "shape" of the data

mean ="calculated average"

median ="middle value"

mode ="most occurring value"

Mean



- Shows “location” but not “how spread out”

Median – *odd number of values*

9 10 10 11 13 15 16 19 19 21 23 28 30 33 34 36 44

= 19

Median - *even number of values*

10 10 11 13 15 16 19 19 21 23 28 30 33 34 36 44

$$\frac{19+21}{2} = 20$$

Mean vs. Median

- The mean can be influenced by *outliers*.
- The mean of {2,3,2,3,2,12} is 4
- The median is 2.5
- The median is much closer to most of the values in the series!

Mode

10 10 11 13 15 16 16 16 21 23 28 30 33 34 36 44

= 16

Measurement Types
Dispersion

Measures of Dispersion (range, variance, standard deviation)

9 10 11 13 15 16 19 19 21 23 28 30 33 34 36 39

- In this sample the mean is 22.25
- How do we describe how “spread out” the sample is?

Range

9 10 11 13 15 16 19 19 21 23 28 30 33 34 36 39

$$Range = max - min$$

$$= 39 - 9$$

$$= 30$$

Variance

- Calculated as the sum of square distances from each point to the mean
- There's a difference between the SAMPLE variance and the POPULATION variance
- subject to Bessel's correction ($\boldsymbol{n-1}$)

Variance

SAMPLE VARIANCE: $$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

POPULATION VARIANCE: $$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

Sample Variance

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$$

4 7 9 8 11

$$\bar{x} = \frac{4 + 7 + 9 + 8 + 11}{5} = \frac{39}{5} = 7.8$$

sample mean

$$s^2 = \frac{(4-7.8)^2+(7-7.8)^2+(9-7.8)^2+(8-7.8)^2+(11-7.8)^2}{5-1}$$

$$= 6.7$$

sample variance

Standard Deviation

- square root of the variance
- benefit: same units as the sample
- meaningful to talk about "*values that lie within one standard deviation of the mean*"

Sample Standard Deviation

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

Sample:

4	7	9	8	11

$$\bar{x} = \frac{4 + 7 + 9 + 8 + 11}{5} = \frac{39}{5} = 7.8$$

sample mean

$$s = \sqrt{\frac{(4 - 7.8)^2 + (7 - 7.8)^2 + (9 - 7.8)^2 + (8 - 7.8)^2 + (11 - 7.8)^2}{5 - 1}}$$

$$= \sqrt{6.7} = 2.59$$

sample standard deviation

Population Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$

Population:

4	7	9	8	11

$$\mu = \frac{4+7+9+8+11}{5} = \frac{39}{5} = 7.8$$

population mean

$$\sigma = \sqrt{\frac{(4-7.8)^2 + (7-7.8)^2 + (9-7.8)^2 + (8-7.8)^2 + (11-7.8)^2}{5}}$$

$$= \sqrt{5.36} = 2.32$$

population standard deviation

Measurement Types
Quartiles

Quartiles and IQR

- Another way to describe data is through quartiles and the interquartile range (IQR)
- Has the advantage that every data point is considered, not aggregated!

Quartiles and IQR

- Consider the following series of 20 values:

9 10 10 11 13	15 16 19 19 21	23 28 30 33 34	36 44 45 47 60

1st quartile

2nd quartile or median

3rd quartile

1. Divide the series
2. Divide each subseries
3. These become quartiles

Quartiles and IQR

- Consider the following series of 20 values:

9 10 10 11 13	15 16 19 19 21	23 28 30 33 34	36 44 45 47 60

1st quartile

2nd quartile or median

3rd quartile

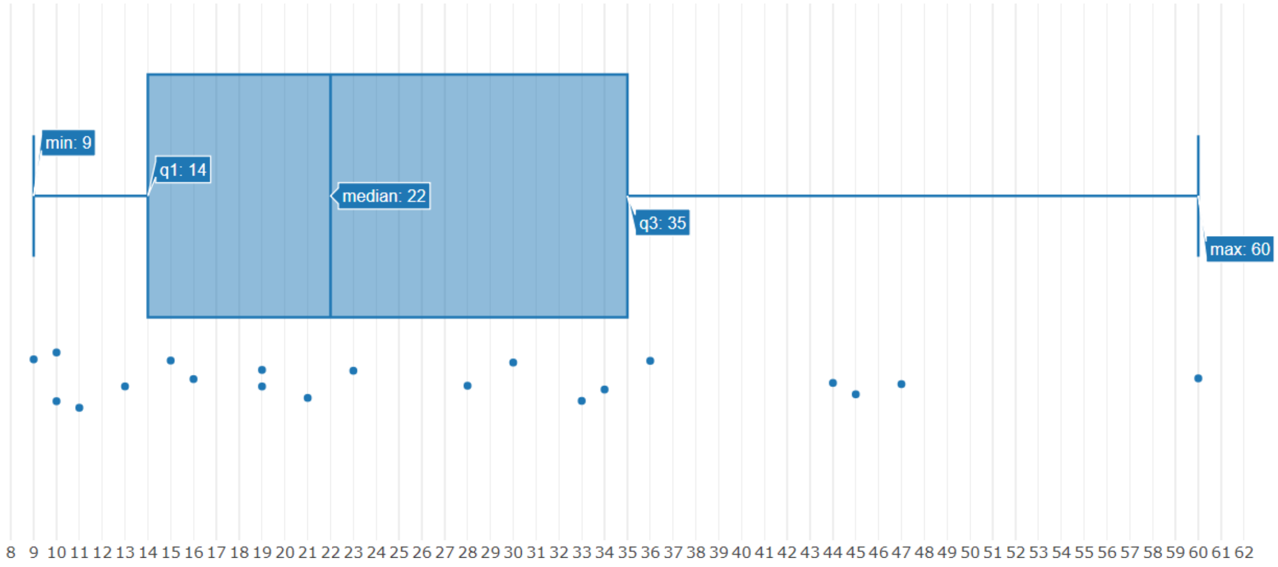
1st quartile = 14

2nd quartile = 22

3rd quartile = 35

Plot the Quartiles

9 10 10 11 13	15 16 19 19 21	23 28 30 33 34	36 44 45 47 60



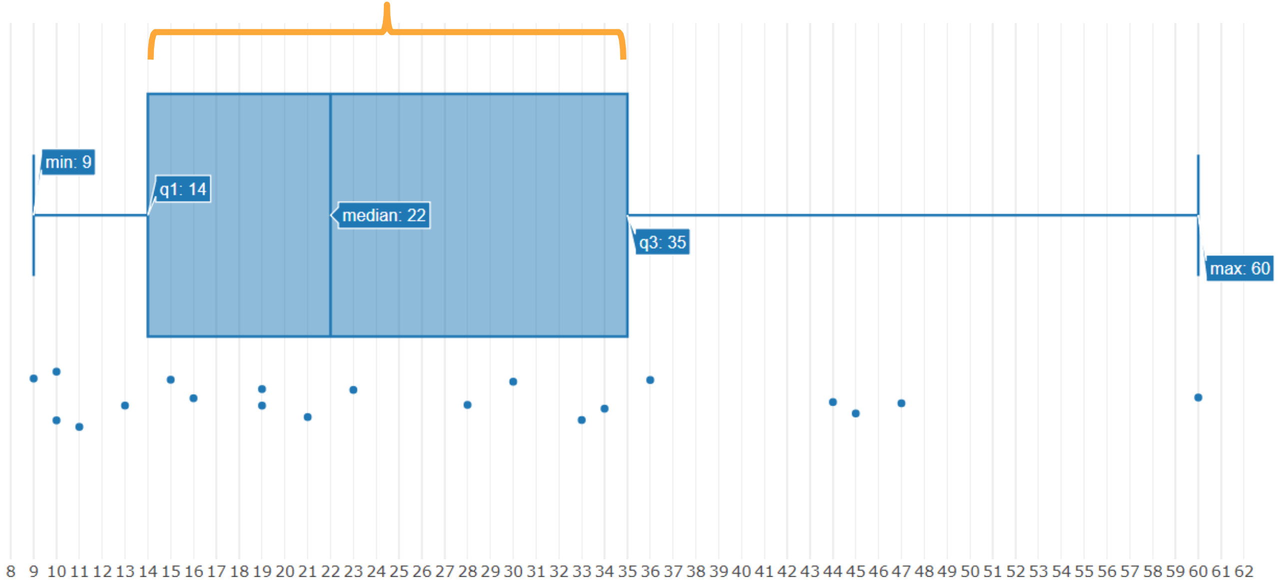
Quartile ranges are seldom the same size!

Fences & Outliers

- What is considered an "outlier"?
- A common practice is to set a "fence" that is 1.5 times the width of the IQR
- Anything outside the fence is an outlier
- This is determined by the *data*, not an arbitrary percentage!

Fences & Outliers

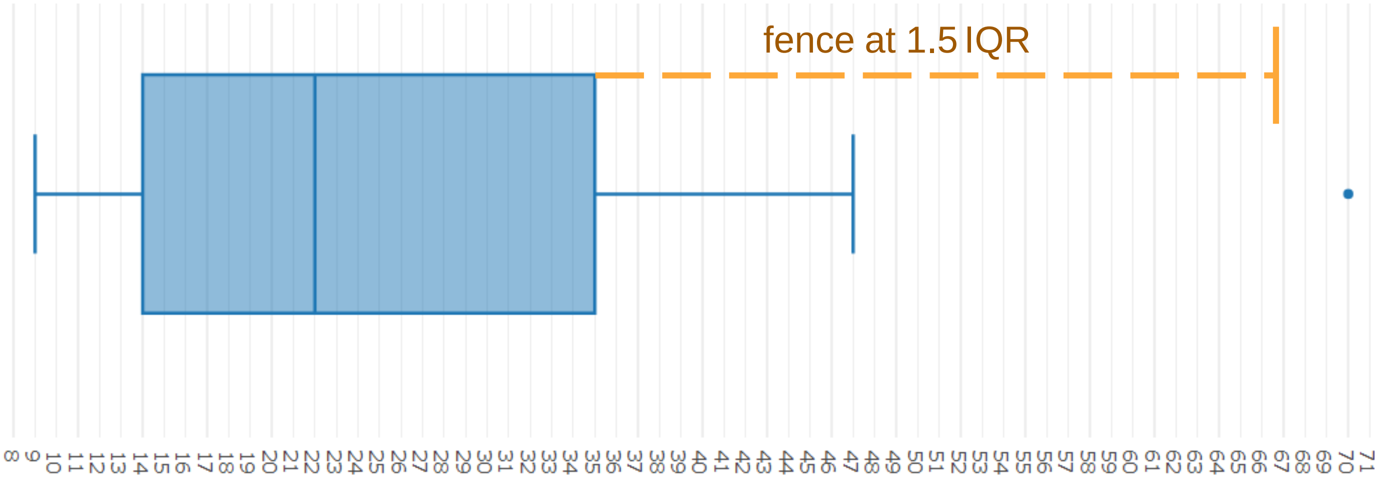
1IQR   1.5 IQR



In this set, 60 is *not* an outlier, but 70 would be

Fences & Outliers

9 10 10 11 13 | 15 16 19 19 21 | 23 28 30 33 34 | 36 44 45 47 70



Here 70 is a true outlier

- When drawing box plots, the whiskers are brought inward to the outermost values inside the fence.

Bivariate Data

Bivariate Data

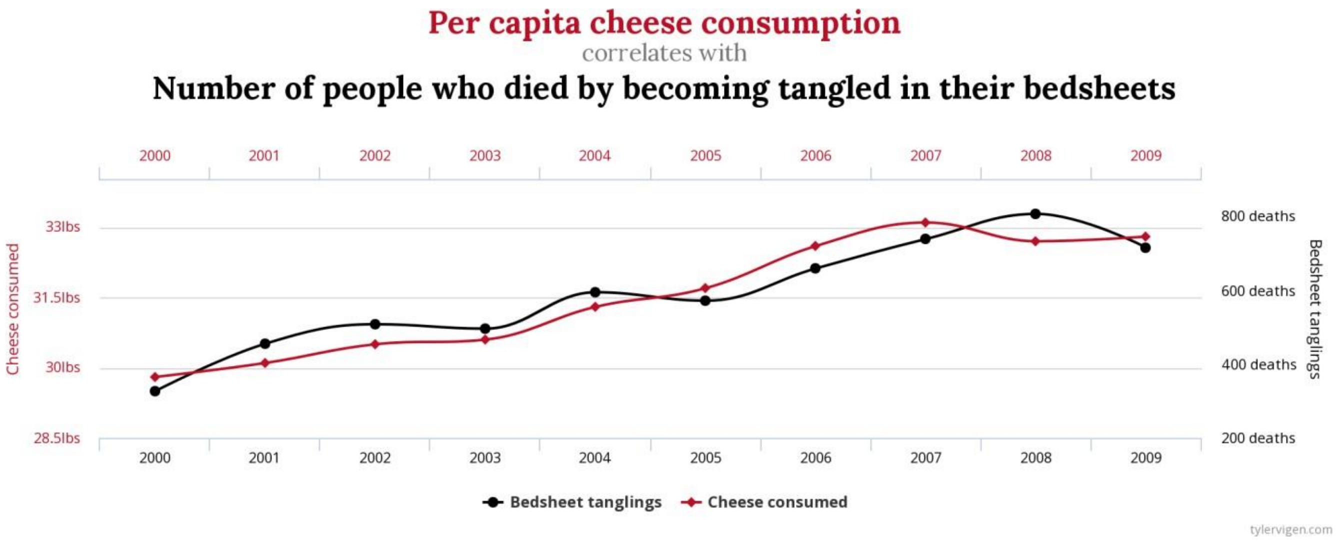
- Compares two variables
- By convention, the x-axis is set to the independent variable
- The y-axis is set to the dependent variable, or that which is being measured relative to x.

Bivariate Data

- Scatter plots may uncover a correlation between two variables
- They *can't* show causality!

Bivariate Data

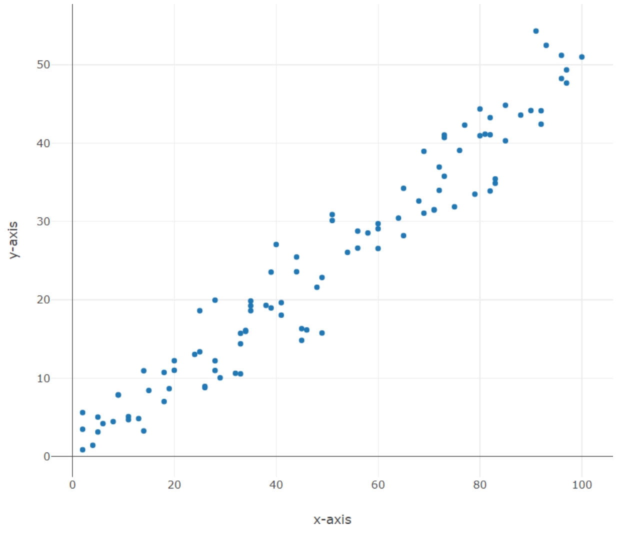
- Correlation between two variables
- Doesn't prove causality!



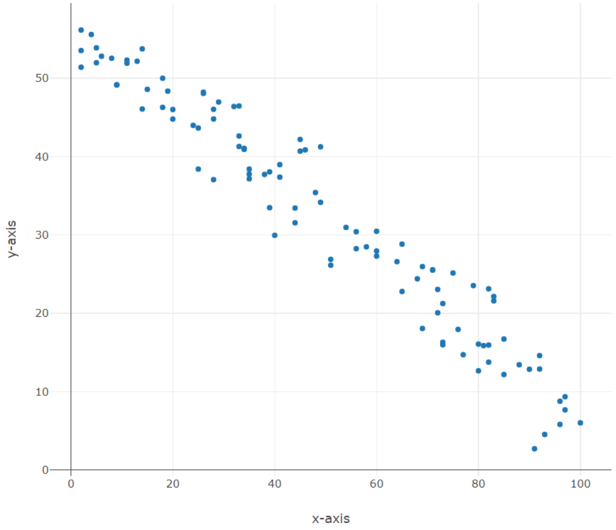
Bivariate Data

- More statistical analysis is needed to determine causality!
- For example: "Does increasing number of police officers decrease crime?"
- We would look at correlation, and do further analysis to understand causality.

Bivariate Data



Positive correlation



Negative or Inverse correlation

Covariance

- A common way to compare two variables is to compare their variances – how far from each item's mean do typical values fall?
- The first challenge is to match scale. Comparing height in inches to weight in pounds isn't meaningful unless we develop a standard score to normalize the data.

Covariance

- For simplicity, we'll consider the *population covariance:*

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

Covariance Exercise

- Consider the following two tables:

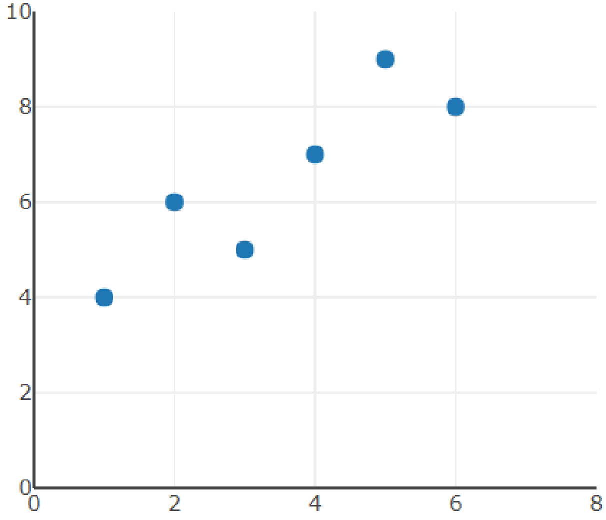
x	y
1	4
2	6
3	5
4	7
5	9
6	8

x	y
1	5
2	9
3	7
4	4
5	8
6	6

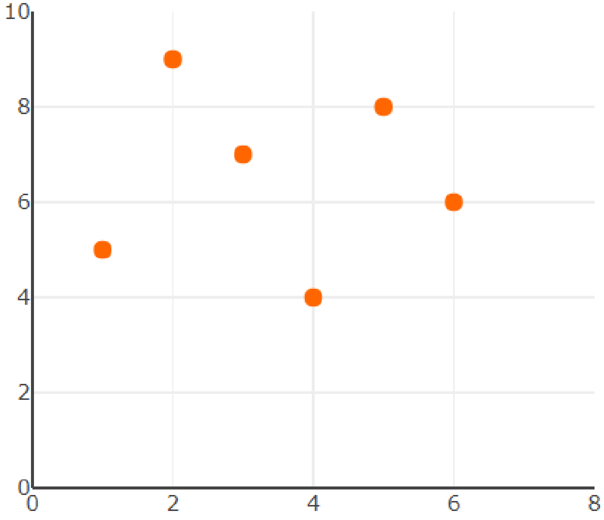
Covariance Exercise

- Plot them:

x	y
1	4
2	6
3	5
4	7
5	9
6	8



x	y
1	5
2	9
3	7
4	4
5	8
6	6



Covariance Exercise

$\boxed{\bar{x} = 3.5, \bar{y} = 6.5}$

- Calculate mean values:

x	y
1	4
2	6
3	5
4	7
5	9
6	8

$$\bar{x} = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\bar{y} = \frac{4+6+5+7+9+8}{6} = 6.5$$

x	y
1	5
2	9
3	7
4	4
5	8
6	6

$$\bar{x} = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\bar{y} = \frac{5+9+7+4+8+6}{6} = 6.5$$

Covariance Exercise

$\bar{x} = 3.5, \bar{y} = 6.5$

- Calculate $(x - \bar{x})$ and $(y - \bar{y})$:

x	y	$(x - \bar{x})$	$(y - \bar{y})$
1	4	-2.5	-2.5
2	6	-1.5	-0.5
3	5	-0.5	-1.5
4	7	0.5	0.5
5	9	1.5	2.5
6	8	2.5	1.5

x	y	$(x - \bar{x})$	$(y - \bar{y})$
1	5	-2.5	-1.5
2	9	-1.5	2.5
3	7	-0.5	0.5
4	4	0.5	-2.5
5	8	1.5	1.5
6	6	2.5	-0.5

Covariance Exercise

$\bar{x} = 3.5, \bar{y} = 6.5$

- Calculate $(x - \bar{x})(y - \bar{y})$:

x	y	(x - x̄)	(y - ȳ)	(x - x̄)(y - ȳ)
1	4	-2.5	-2.5	6.25
2	6	-1.5	-0.5	0.75
3	5	-0.5	-1.5	0.75
4	7	0.5	0.5	0.25
5	9	1.5	2.5	3.75
6	8	2.5	1.5	3.75

x	y	(x - x̄)	(y - ȳ)	(x - x̄)(y - ȳ)
1	5	-2.5	-1.5	3.75
2	9	-1.5	2.5	-3.75
3	7	-0.5	0.5	-0.25
4	4	0.5	-2.5	-1.25
5	8	1.5	1.5	2.25
6	6	2.5	-0.5	-1.25

Covariance Exercise

$$\bar{x} = 3.5, \bar{y} = 6.5$$

- Calculate sums:

x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
1	4	-2.5	-2.5	6.25
2	6	-1.5	-0.5	0.75
3	5	-0.5	-1.5	0.75
4	7	0.5	0.5	0.25
5	9	1.5	2.5	3.75
6	8	2.5	1.5	3.75
			Σ	**15.5**

x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
1	5	-2.5	-1.5	3.75
2	9	-1.5	2.5	-3.75
3	7	-0.5	0.5	-0.25
4	4	0.5	-2.5	-1.25
5	8	1.5	1.5	2.25
6	6	2.5	-0.5	-1.25
			Σ	

Covariance Exercise

$\bar{x} = 3.5, \bar{y} = 6.5$

- Calculate covariance:

x	y
1	4
2	6
3	5
4	7
5	9
6	8

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{15.5}{6} = \mathbf{2.583}$$

Σ	15.5

x	y
1	5
2	9
3	7
4	4
5	8
6	6

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

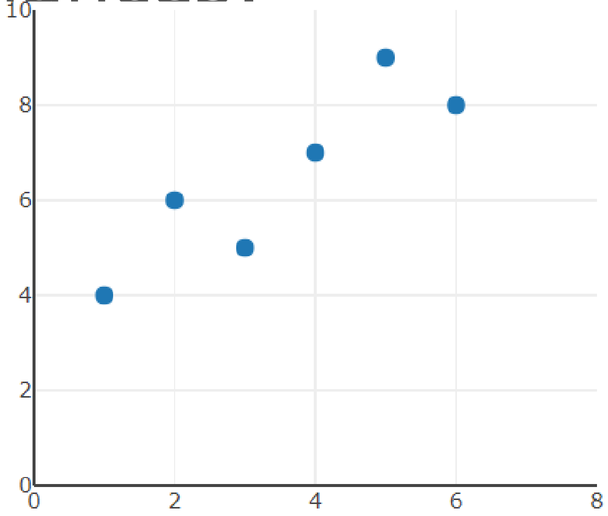
$$= \frac{-0.5}{6} = \mathbf{-0.083}$$

Σ	-0.5

Covariance Exercise

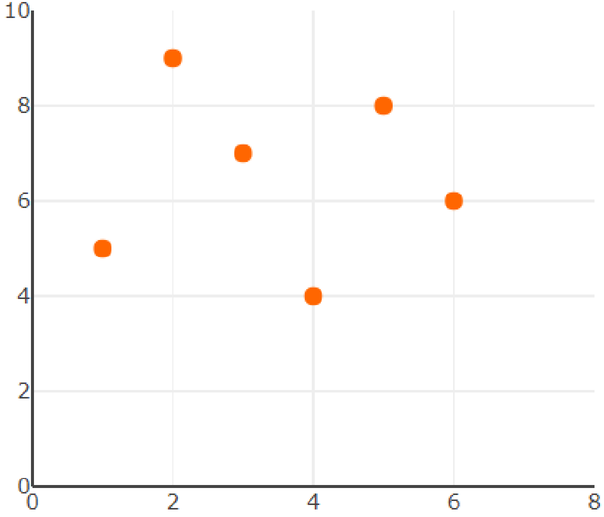
- Compare covariances:

x	y
1	4
2	6
3	5
4	7
5	9
6	8



cov(x,y) = 2.583

x	y
1	5
2	9
3	7
4	4
5	8
6	6



cov(x,y) = -0.083

Pearson Correlation Coefficient

Pearson Correlation Coefficient

- In order to normalize values coming from two different distributions, we use:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{n}\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}\sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$\rho =$ Greek letter "rho"

$cov =$ covariance

$\sigma =$ standard deviation

$\bar{x} =$ mean of X

Pearson Correlation Coefficient

- Values fall between $+1$ and -1, where

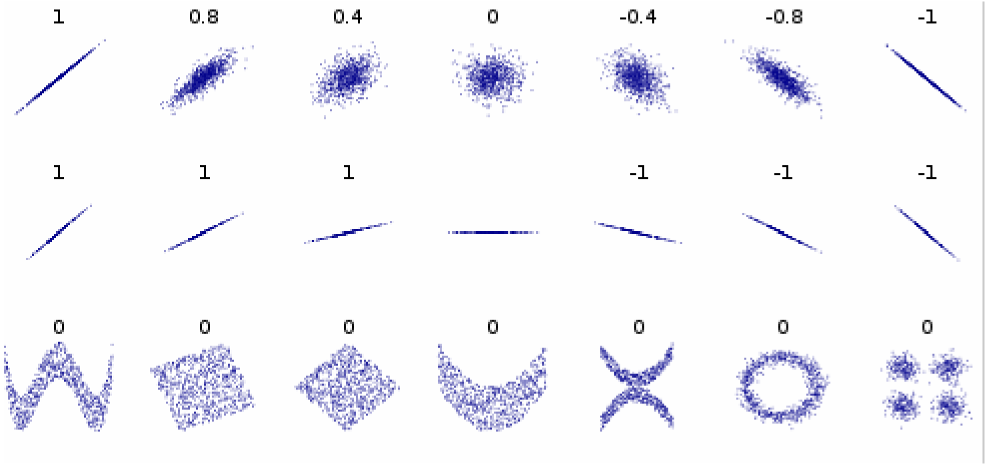
  1 = total positive linear correlation

  0 = no linear correlation

  -1 = total negative linear correlation

Pearson Correlation Coefficient

- Several sets of (x, y) points, with the correlation coefficient for each set:



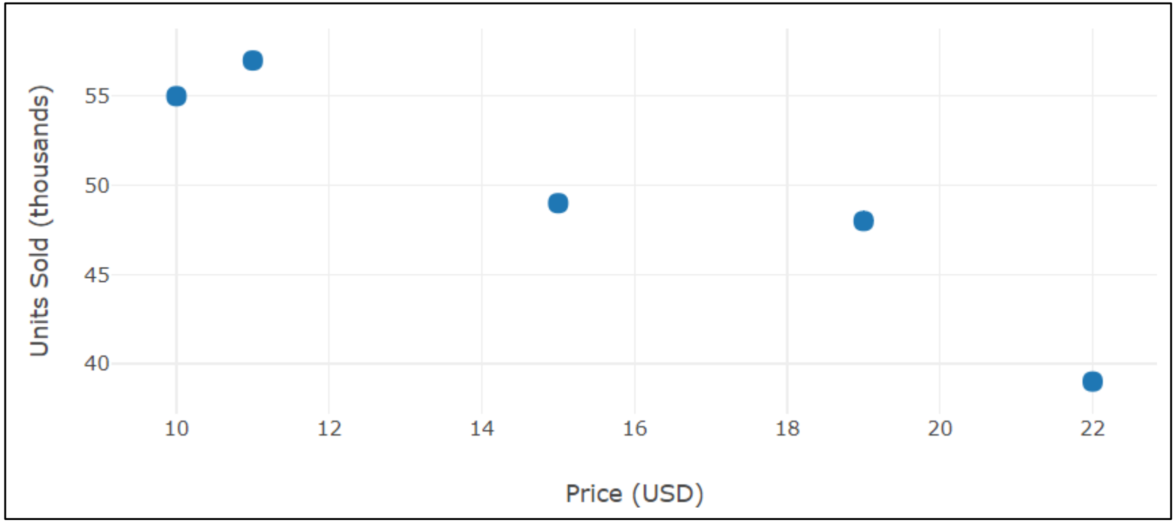
Correlation Exercise

- A company decides to test sales of a new product in five separate markets, to determine the best price point.
- They set a different price in each market and record sales volume over the same 30 day period.

Correlation Exercise

- These are the results
- Plot the results

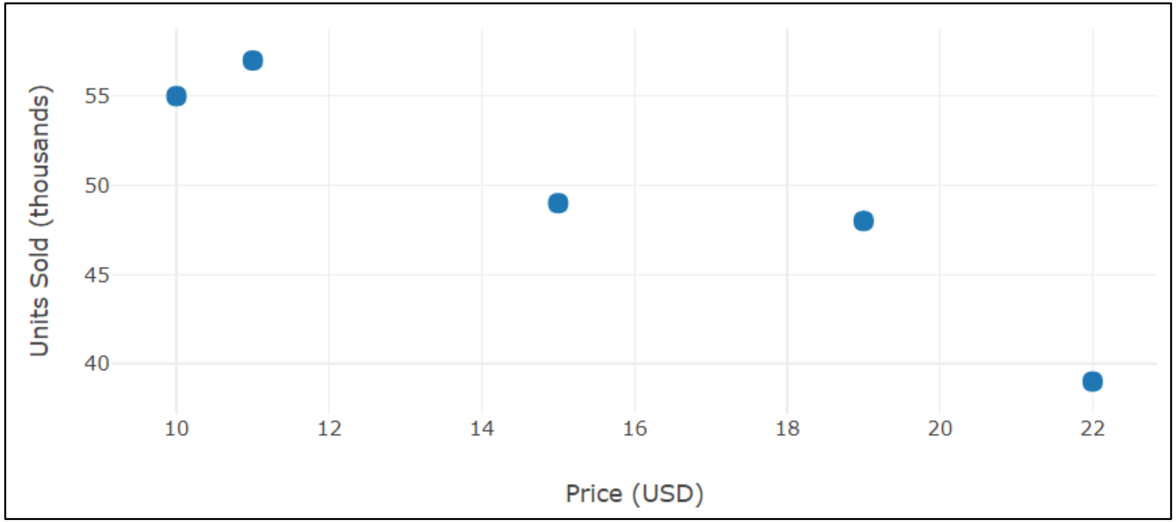
Price (USD)	Units Sold (thousands)
10	55
11	57
15	49
19	48
22	39



Correlation Exercise

- There appears to be a strong correlation, but how strong?

Price (USD)	Units Sold (thousands)
10	55
11	57
15	49
19	48
22	39



Correlation Exercise

1. Recall the simplified correlation formula:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

Price (USD)	Units Sold (thousands)
10	55
11	57
15	49
19	48
22	39

2. Find the mean of x and y:

$$\bar{x} = \frac{10+11+15+19+22}{5} = 15.4$$

$$\bar{y} = \frac{55+57+49+48+39}{5} = 49.6$$

Correlation Exercise

$$\rho_{X,Y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$\bar{x} = 15.4 \quad \bar{y} = 49.6$

3. Calculate $(x-\bar{x})$ and $(y-\bar{y})$:

Price (USD)	Units Sold (thousands)	$(x-\bar{x})$	$(y-\bar{y})$
10	55	-5.4	5.4
11	57	-4.4	7.4
15	49	-0.4	-0.6
19	48	3.6	-1.6
22	39	6.6	-10.6

Correlation Exercise

$$\rho_{X,Y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$\bar{x} = 15.4 \quad \bar{y} = 49.6$

4. Calculate $(x-\bar{x})(y-\bar{y})$:

Price (USD)	Units Sold (thousands)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
10	55	-5.4	5.4	-29.16
11	57	-4.4	7.4	-32.56
15	49	-0.4	-0.6	0.24
19	48	3.6	-1.6	-5.76
22	39	6.6	-10.6	-69.96

Correlation Exercise

$$\rho_{X,Y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$\bar{x} = 15.4 \quad \bar{y} = 49.6$

5. Calculate $(x-\bar{x})^2$ and $(y-\bar{y})^2$:

Price (USD)	Units Sold (thousands)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	55	-5.4	5.4	-29.16	29.16	29.16
11	57	-4.4	7.4	-32.56	19.36	54.76
15	49	-0.4	-0.6	0.24	0.16	0.36
19	48	3.6	-1.6	-5.76	12.96	2.56
22	39	6.6	-10.6	-69.96	43.56	112.36

Correlation Exercise

$$\rho_{X,Y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$\bar{x} = 15.4 \quad \bar{y} = 49.6$

6. Compute the sums:

Price (USD)	Units Sold (thousands)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	55	-5.4	5.4	-29.16	29.16	29.16
11	57	-4.4	7.4	-32.56	19.36	54.76
15	49	-0.4	-0.6	0.24	0.16	0.36
19	48	3.6	-1.6	-5.76	12.96	2.56
22	39	6.6	-10.6	-69.96	43.56	112.36
			Σ	**-137.2**	**105.2**	**199.2**

Correlation Exercise

$$\rho_{X,Y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$\bar{x} = 15.4 \quad \bar{y} = 49.6$

7. Plug these into the original formula:

Price (USD)	Units Sold (thousands)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	55	-5.4	5.4	-29.16	29.16	29.16
11	57	-4.4	7.4	-32.56	19.36	54.76
15	49	-0.4	-0.6	0.24	0.16	0.36
19	48	3.6	-1.6	-5.76	12.96	2.56
22	39	6.6	-10.6	-69.96	43.56	112.36
			Σ	**-137.2**	**105.2**	**199.2**

Correlation Exercise

$$\rho_{X,Y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$$\bar{x} = 15.4 \quad \bar{y} = 49.6$$

7. Plug these into the original formula:

$$\rho_{X,Y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{-137.2}{\sqrt{105.2}\sqrt{199.2}}$$

$$= \frac{-137.2}{10.26 \times 14.11} = \frac{-137.2}{144.8} = \mathbf{-0.948}$$

Σ	**-137.2**	**105.2**	**199.2**

Correlation Exercise

- $\rho_{X,Y} = -0.948$ shows a *very* strong negative correlation!