Pivot Tables

A pivot table is an Excel reporting tool that facilitates summarizing data without the use of formulas and displaying various statistics using different formats.

Below is an example of a pivot table that provides the means (arithmetic averages) for four variables (see grand totals) while also providing means for two categories (male = 1 and female = 2):

	G	H	I	J	K
3		Values			
4	Row Labels	Average of alienation	Average of isolation	Average of powerl	Average of norml
5	1	66.65277778	27.44444444	24.25694444	14.95138889
6	2	70.08333333	28.125	25.58333333	16.375
7	**Grand Total**	**67.14285714**	**27.54166667**	**24.44642857**	**15.1547619**

Constructing a Pivot Table

	A	B	C	D	E
1	gender	alienation	isolation	powerl	norml
2	1	72	26	29	17
3	1	89	33	32	24
4	1	68	26	27	15
5	1	82	30	29	23
6	1	93	38	33	22
7	1	76	29	29	18
8	1	66	30	22	14
9	1	75	28	28	19
10	1	60	25	21	14
11	1	62	28	20	14
12	1	87	33	32	22
13	1	82	34	31	17
14	1	67	23	27	17
15	1	78	29	32	17
16	1	69	27	26	16
17	1	86	37	27	22
18	1	68	31	26	11
19	1	61	25	22	14
20	1	67	29	26	12

	G	H	I	J	K
3		Values			
4	Row Labels	Average of alienation	Average of isolation	Average of powerl	Average of norml
5	1	66.65277778	27.44444444	24.25694444	14.95138889
6	2	70.08333333	28.125	25.58333333	16.375
7	**Grand Total**	**67.14285714**	**27.54166667**	**24.44642857**	**15.1547619**

	G	H	I	J
11		Column Labels		
12	Values	1	2	Grand Total
13	Average of alienation	66.65277778	70.08333333	67.14285714
14	Average of isolation	27.44444444	28.125	27.54166667
15	Average of powerl	24.25694444	25.58333333	24.44642857
16	Average of norml	14.95138889	16.375	15.1547619
17	StdDev of alienation	11.36816771	10.43787696	11.27462612
18	StdDev of isolation	4.870027897	4.326988108	4.790027273
19	StdDev of powerl	4.4482204	3.977290608	4.397548213
20	StdDev of norml	4.477339778	4.576048229	4.505469856

Open the dataset *Motivation.xlsx*. Click on the Pivot Table worksheet tab.

File available at COURSE GITHUB

TASK

Construct a pivot table using the provided datset that:

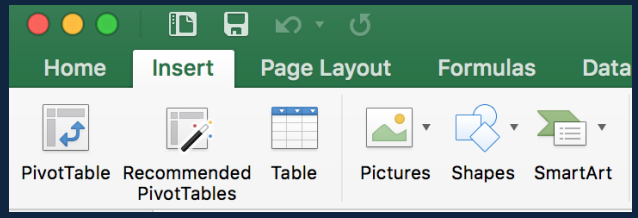
Displays values as Averages of alienation, isolation, powerlessness (powerl), and normlessness (norml).

Displays gender (i.e., 1=male and 2=female) as row labels.

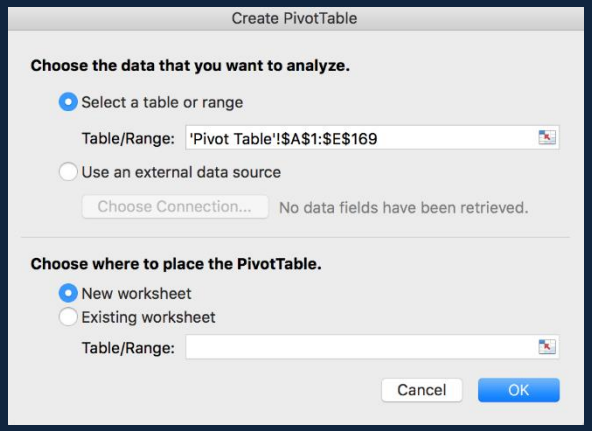
Constructing a Pivot Table

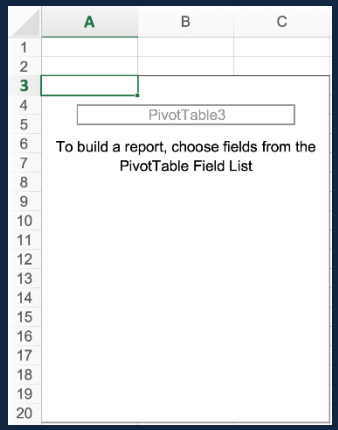
	A	B	C	D	E	F	G	H	I	J	K
1	gender	alienation	isolation	powerl	norml						
2	1	72	26	29	17						
3	1	89	33	32	24			Values			
4	1	68	26	27	15		Row Labels	Average of alienation	Average of isolation	Average of powerl	Average of norml
5	1	82	30	29	23		1	66.65277778	27.44444444	24.25694444	14.95138889
6	1	93	38	33	22		2	70.08333333	28.125	25.58333333	16.375
7	1	76	29	29	18		**Grand Total**	**67.14285714**	**27.54166667**	**24.44642857**	**15.1547619**
8	1	66	30	22	14						
9	1	75	28	28	19						
10	1	60	25	21	14						
11	1	62	28	20	14			Column Labels			
12	1	87	33	32	22		Values	1	2	Grand Total	
13	1	82	34	31	17		Average of alienation	66.65277778	70.08333333	67.14285714	
14	1	67	23	27	17		Average of isolation	27.44444444	28.125	27.54166667	
15	1	78	29	32	17		Average of powerl	24.25694444	25.58333333	24.44642857	
16	1	69	27	26	16		Average of norml	14.95138889	16.375	15.1547619	
17	1	86	37	27	22		StdDev of alienation	11.36816771	10.43787696	11.27462612	
18	1	68	31	26	11		StdDev of isolation	4.870027897	4.326988108	4.790027273	
19	1	61	25	22	14		StdDev of powerl	4.4482204	3.977290608	4.397548213	
20	1	67	29	26	12		StdDev of norml	4.477339778	4.576048229	4.505469856	

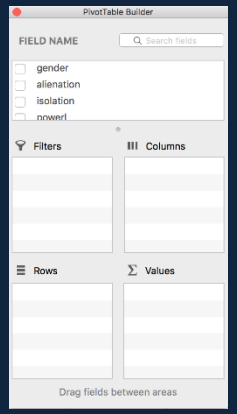
Click on a cell in the dataset to make it active, e.g., cell A1. Then proceed to the Excel Insert tab and click on the Pivot Table icon.



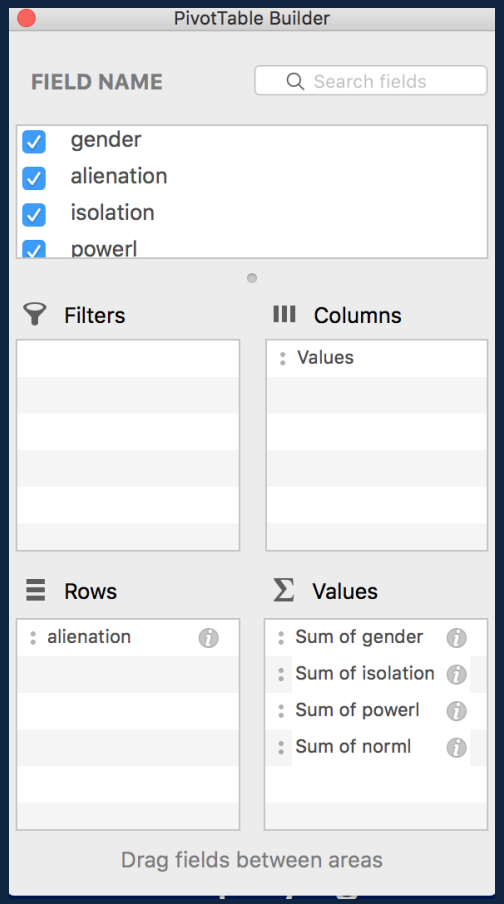
The Create Pivot Table dialog opens. Confirm/change the settings and click OK. A blank pivot table appears on the designated worksheet along with the PivotTable Builder dialog.







Constructing a Pivot Table

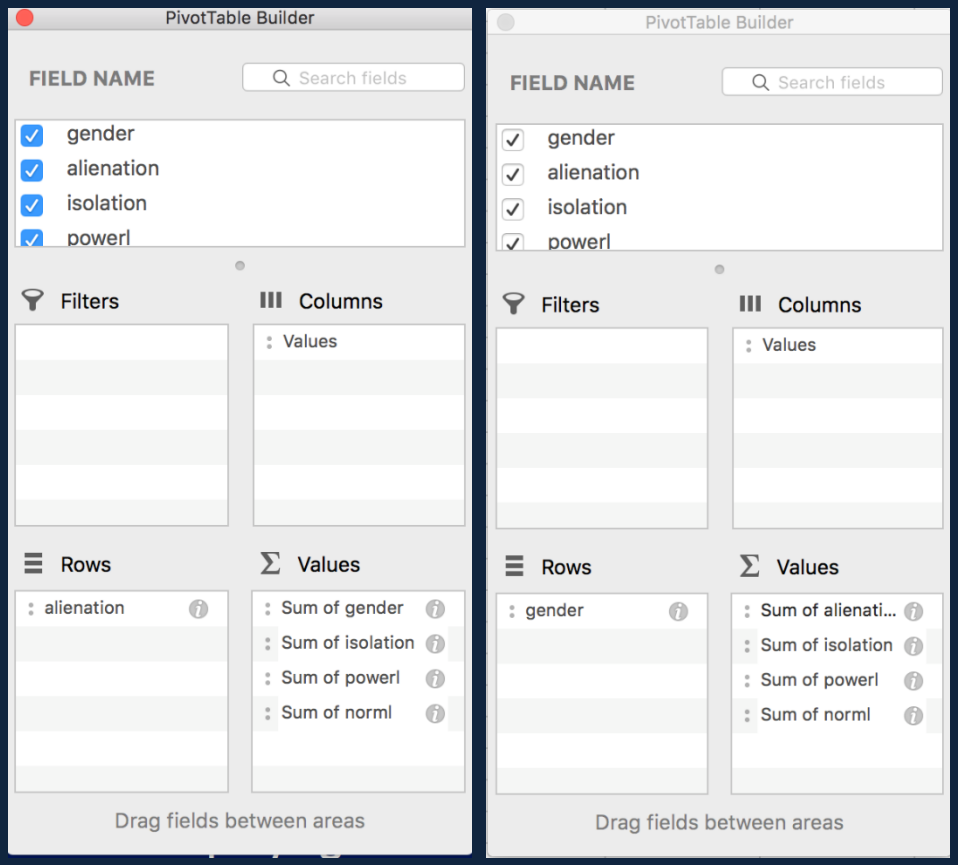


Using the Field Name panel, check gender, alienation, isolation, powerl, and norml. Notethe appearance of tiles in the Columns, Rows, and Values panel.

The pivot table changes to reflect the settings in the PivotTable Builder dialog. See below.

Row Labels	Sum of gender	Sum of isolation	Sum of powerl	Sum of norml
30	1	13	9	8
42	1	23	13	6
44	1	19	18	7
47	1	27	14	6
49	1	23	18	8
50	2	39	37	24
51	5	116	97	42
52	2	45	36	23
53	3	47	39	20
54	4	88	83	45
55	6	86	86	48
56	4	107	80	37
57	11	261	200	109
58	6	121	106	63
59	5	109	115	71
60	4	80	62	38
61	7	188	155	84
62	6	162	129	81
63	6	122	122	71
64	5	104	95	57
65	5	141	121	63
66	5	140	121	69
67	6	151	156	95
68	10	248	215	149
69	4	85	76	46
70	3	53	50	37
71	3	66	52	24
72	6	108	111	69
73	7	167	163	108
74	4	120	109	67
75	11	256	255	164
76	4	109	112	83
77	3	64	54	36
78	5	124	122	66
79	8	197	161	116
80	3	65	57	38
81	2	60	61	41
82	5	157	150	103
83	1	31	31	21
84	4	109	83	60
86	5	178	149	103
87	2	69	59	46
89	1	33	32	24
91	1	42	27	22
93	1	38	33	22
95	2	36	33	26
Grand Total	**192**	**4627**	**4107**	**2546**

Constructing a Pivot Table

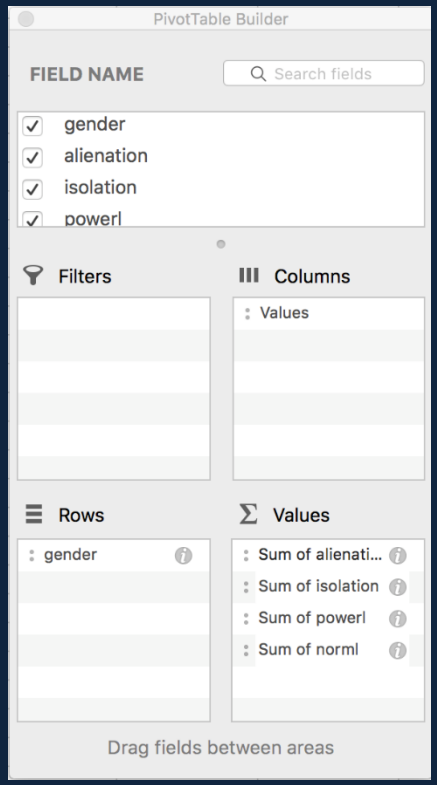


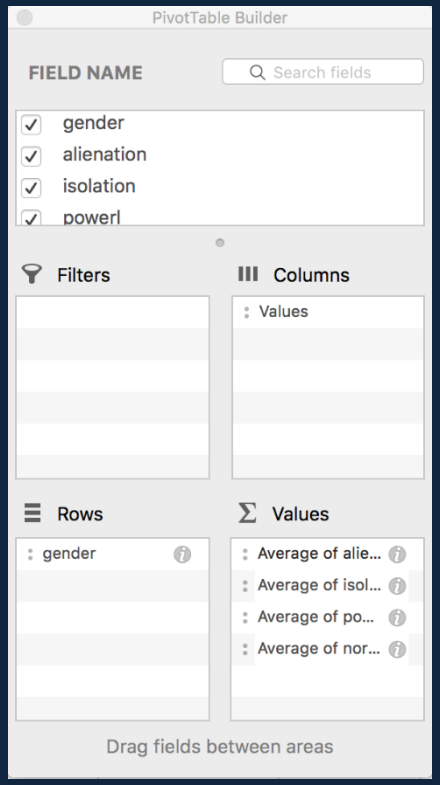
Drag the gender tile from the Values panel to the Rows panel; drag the alienation tile from the Rows panel to the Values panel.

The pivot table changes to reflect the settings in the PivotTable Builder dialog. See below.

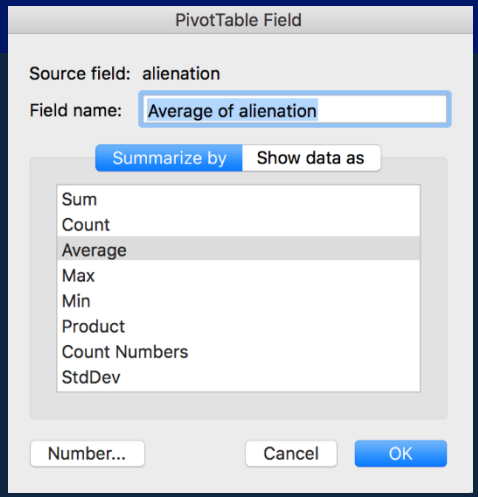
	A	B	C	D	E	F
1						
2						
3	Row Labels	Sum of alienation	Sum of isolation	Sum of powerl	Sum of norml	
4	1	9598	3952	3493	2153	
5	2	1682	675	614	393	
6	**Grand Total**	**11280**	**4627**	**4107**	**2546**	
7						
8						

Constructing a Pivot Table



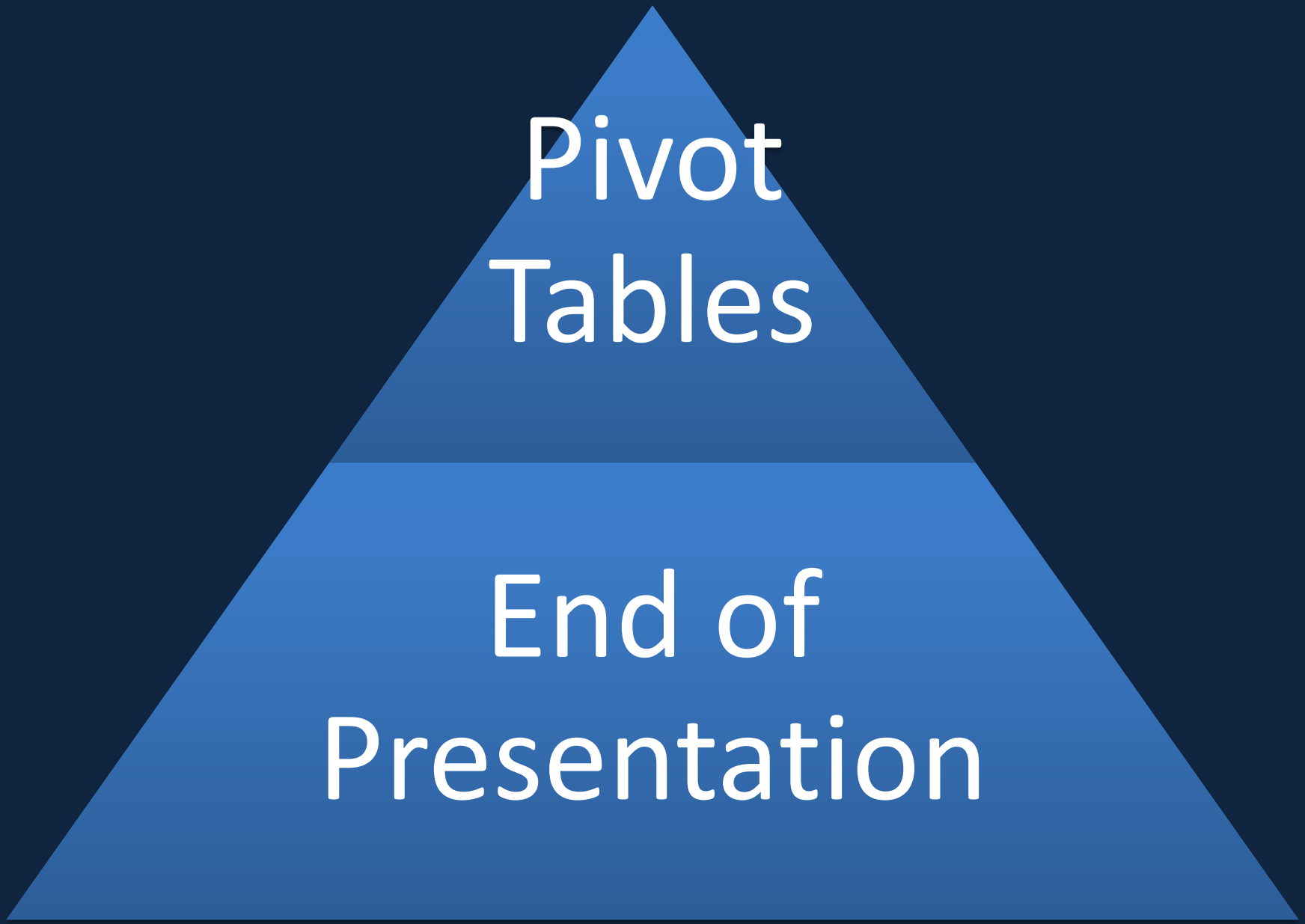


Click the "I" icon on each tile in the Values panel (in turn). Change each variable from Sum (default) to Average.



The pivot table changes to reflect the settings in the PivotTable Builder dialog. See below.

	A	B	C	D	E
3	Row Labels	Average of alienation	Average of isolation	Average of powerl	Average of norml
4	1	66.65277778	27.44444444	24.25694444	14.95138889
5	2	70.08333333	28.125	25.58333333	16.375
6	**Grand Total**	**67.14285714**	**27.54166667**	**24.44642857**	**15.1547619**
7					


Pivot Tables
End of Presentation

Descriptive Statistics

- Statistics
  - Summary measures calculated for a sample dataset.
- Parameters
  - Summary measures calculated for a population dataset.
- Used to describe variables
  - Measures of central tendency, e.g., mean, median, mode
  - Measures of dispersion, e.g., standard deviation, variance, range
  - Measures of relative position, e.g., percentile, quartile
  - Graphs and charts, e.g., scatterplot, column chart, histogram

Measures of Central Tendency

Designed to give information concerning the typical score of a large number of scores.

Researchers typically report the best measures of central tendency and dispersion for each variable. The best measure to report varies based on the shape of a variable's distribution and scale of measurement.

- Interval/ratio data – mean, median, and mode can be calculated and reported, as appropriate.
- Ordinal data - median can and should be reported; use of the mean is wrong.
- Nominal data - mode can and should be reported; use of either mean or median is wrong.

	Measures of Central Tendency
Nominal data	Mode
Ordinal data	Median, Mode
Interval data	Mean, Median, Mode
Ratio data	Mean, Median, Mode

Calculating Measures of Central Tendency

	A	B	C	D	E	F
1	gender	age	ethnicity	gpa	p_learning	c_community
2	1	2	2	1.58	7	23
3	1	2	2	1.87	7	22
4	1	2	2	2	5	23
5	1	2	2	2	7	23
6	1	3	2	2.1	5	22
7	1	3	2	2.4	8	32
8	1	4	2	2.5	7	24
9	1	2	2	2.5	6	22
10	1	3	4	2.5	5	28
11	1	2	2	2.5	5	25
12	1	3	2	2.55	7	22
13	1	2	2	2.6	6	23
14	1	3	2	2.6	9	33
15	1	3	2	2.62	2	19

Open the dataset *Motivation.xlsx*.
Click the worksheet Descriptive Statistics tab (at the bottom of the worksheet).

File available at
http://www.watertreepress.com/stats

TASK
Calculate the count, mean, median, and mode of the classroom community (c_community) variable.

Count (Sample Size; *N* or *n*)

- The count (*N, n*) is a statistic that reflects the number of cases selected in the dataset. It is often used to represent sample (*N*) or sub-sample (*n*) size. It is an important statistic in any research study. The mathematical consist of counting the number of cases represented by the data.

$$N = x_1 + x_2 ... + x_k$$

- Excel functions:

COUNT(value1,value2,...). Counts the numbers in the range of values, e.g., the formula =COUNT(A1:A30) counts the number of values in cells A1 through A30. The result will be 30 in this example.

COUNTA(value1,value2,...). Counts the cells with non-empty values in the range of values. This functions counts cells with text as well as cells with numbers.

Example of Count

Measurements
x
7
7
5
7
5
8
7
6
5

$N = 9$

	C	D
1	Count	=COUNT(A2:A170)
2	Mean	=AVERAG 2:A170)
3	Standard Error of the Mean	=STDE 70)/SQRT(COUNT(A2:A170))
4	Median	=MEDIA 170)
5	Mode	
6		
7	Varia	
8	Stand	
9	Rang	
10	Interq	0,1)
11	Maximum	=MAX(A2:A170)
12	Minimum	=MIN(A2:A170)
13		
14	Skewness Coefficient	=SKEW(A2:A170)
15	SE Skewness	=SQRT(6/COUNT(A2:A170))
16	Standard Coefficient of Skewness	=D14/D15
17	Kurtosis Coefficient	=KURT(A2:A170)
18	SE Kurtosis	=SQRT(24/COUNT(A2:A170))
19	Standard Coefficient of Kurtosis	=D17/D18

TASK

Enter the following formula in cell D1 to calculate the sample size used to measure c_community:

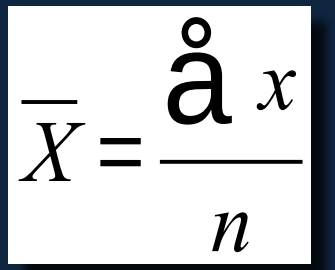
=COUNT(A2:A170)

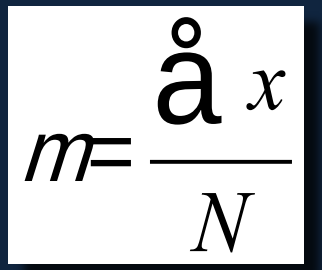
	C	D	E
1	Count	169	
2	Mean	28.84023	
3	Standard Error of the Mean	0.478	
4	Median		
5	Mode		
6			
7	Variance		
8	Standard Devia		
9	Range		
10	Interquartile Range	10	
11	Maximum	40	
12	Minimum	15	
13			
14	Skewness Coefficient	0.073045168	
15	SE Skewness	0.188422288	
16	Standard Coefficient of Skewness	0.387667347	
17	Kurtosis Coefficient	-1.044172509	
18	SE Kurtosis	0.376844576	
19	Standard Coefficient of Kurtosis	-2.770830673	

Excel displays the count as 169. This sample statistic is typically reported as *N* = 169 in the results section of a research paper, as appropriate. This statistic means that there are 169 cases in the measured sample.

Mean (Arithmetic Average; *M* , μ)

- Determines the sample mean or estimates an unknown population mean.
  - Sample mean is denoted by *M* or x-bar.
  - Population mean is denoted by the Greek letter μ (mu)
- Used with interval and ratio scale variables.
- Best measure to describe normal unimodal distributions. Unlike the median and the mode, it is not appropriate to use the mean only to describe a highly skewed distribution.
- Always located toward the skewed (tail) end of skewed distributions in relation to the median and mode.
- Formulas





- Excel function:

AVERAGE(number1,number2,...). Returns the arithmetic mean, where numbers represent the range of numbers.

Example of Mean

Measurements	Deviation
x	x − mean
7	1
7	1
5	-1
7	1
5	-1
8	2
7	1
6	0
5	-1
Sum	0

The first column represents the raw score and the second column represent the raw score minus the mean (i.e., the deviation from the mean).

Mean = 6.33

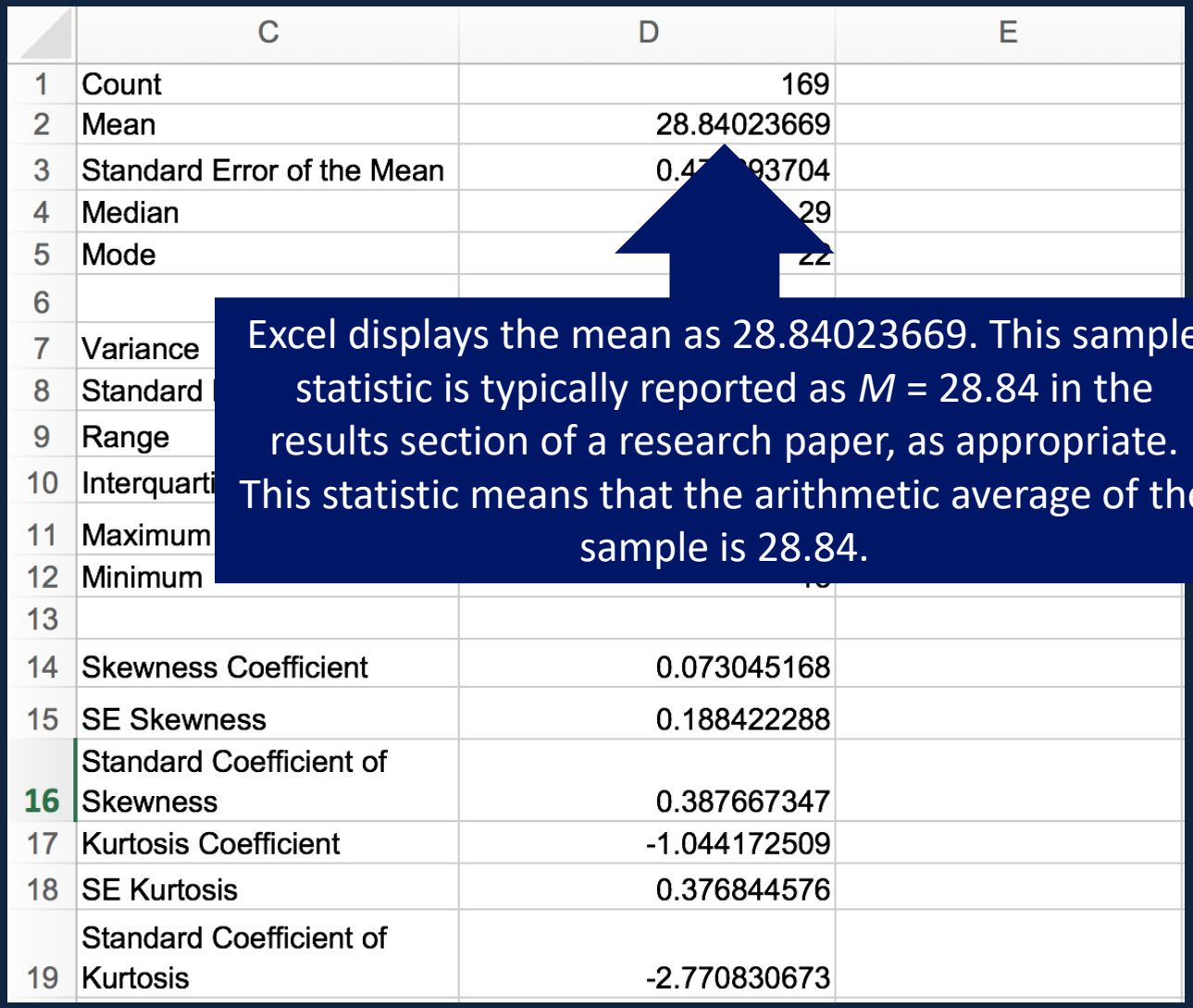
Sum of deviations from the mean = 0

	C	D
1	Count	=COUNT(A2:A170)
2	Mean	=AVERAGE(A2:A170)
3	Standard Error of the Mean	=STDEV :A170)/SQRT(COUNT(A2:A170))
4	Median	=MED 70)
5	Mode	=MODE (A2:A170)
6		
7	Varian	
8	Standa	
9	Range	
10	Interqu	70,1)
11	Maximum	=MAX(A2:A170)
12	Minimum	=MIN(A2:A170)
13		
14	Skewness Coefficient	=SKEW(A2:A170)
15	SE Skewness	=SQRT(6/COUNT(A2:A170))
16	Standard Coefficient of Skewness	=D14/D15
17	Kurtosis Coefficient	=KURT(A2:A170)
18	SE Kurtosis	=SQRT(24/COUNT(A2:A170))
19	Standard Coefficient of Kurtosis	=D17/D18

TASK

Enter the following formula in cell D2 to calculate the mean of variable c_community:

=AVERAGE(A2:A170)



	C	D	E
1	Count	169	
2	Mean	28.84023669	
3	Standard Error of the Mean	0.4[illegible]93704	
4	Median	[illegible]29	
5	Mode	[illegible]22	
6			
7	Variance		
8	Standard		
9	Range		
10	Interquarti		
11	Maximum		
12	Minimum		
13			
14	Skewness Coefficient	0.073045168	
15	SE Skewness	0.188422288	
16	Standard Coefficient of Skewness	0.387667347	
17	Kurtosis Coefficient	-1.044172509	
18	SE Kurtosis	0.376844576	
19	Standard Coefficient of Kurtosis	-2.770830673	

Median (*Mdn*)

- The median is the score that divides the distribution into two equal halves (it represents the 50th percentile).
  - It is the midpoint of the distribution when the distribution has an odd number of scores.
  - It is the number halfway between the two middle scores when the distribution has an even number of scores.
- Not sensitive to outliers.
- Used with the ordinal scale or when the distribution is skewed
- If the distribution is normally distributed (i.e, symmetrical and unimodal), the mode, median, and mean coincide.
- Excel function:

MEDIAN(number1,number2,...). Returns the median of a range of numbers.

Example of Median

Measurements	Ranked Data
x	x
7	5
7	5
5	5
7	6
5	7
8	7
7	7
6	7
5	8

Median = 7

The median is the mid value of ranked data when there are an odd number of cases

	C	D
1	Count	=COUNT(A2:A170)
2	Mean	=AVERAGE(A2:A170)
3	Standard Error of the Mean	=STDEV.P(A2:A170)/SQRT(COUNT(A2:A170))
4	Median	=MEDIAN(A2:A170)
5	Mode	=MODE.SNGL(A2:A170)
6		
7	Variance	=VAR.P(A2:A170)
8	Standa	
9	Range	
10	Interqu	70,1)
11	Maxim	
12	Minimu	
13		
14	Skewness Coefficient	=SKEW(A2:A170)
15	SE Skewness	=SQRT(6/COUNT(A2:A170))
16	Standard Coefficient of Skewness	=D14/D15
17	Kurtosis Coefficient	=KURT(A2:A170)
18	SE Kurtosis	=SQRT(24/COUNT(A2:A170))
19	Standard Coefficient of Kurtosis	=D17/D18

TASK

Enter the following formula in cell D4 to calculate the median of variable c_community:

=MEDIAN(A2:A170)

	C	D	E
1	Count	169	
2	Mean	28.84023669	
3	Standard Error of the Mean	0.478693704	
4	Median	29	
5	Mode		
6			
7	Variance	38.7259	
8	Standard Deviation		
9	Range		
10	Interquartile Range		
11	Maximum		
12	Minimum		
13			
14	Skewness Coefficient	0.073045168	
15	SE Skewness	0.188422288	
16	Standard Coefficient of Skewness	0.387667347	
17	Kurtosis Coefficient	-1.044172509	
18	SE Kurtosis	0.376844576	
19	Standard Coefficient of Kurtosis	-2.770830673	

Excel displays the median as 29. This statistic means 29 is halfway into the dataset when the data are rank-ordered.

Mode (*Mo*)

- Most frequently occurring score
- A distribution is called unimodal if there is only one major peak in the distribution of scores when displayed as a histogram
- If the distribution is normally distributed (i.e, symmetrical and unimodal), the mode, median, and mean coincide
- The mode is useful in describing nominal variables and in describing a bimodal or multimodal distribution (use of the mean or median only can be misleading)
  - Major mode = most common value, largest peak
  - Minor mode(s) = smaller peak(s)
  - Unimodal (i.e., having one major peak or mode)
  - Bimodal (i.e., having two major peaks or modes)
  - Multimodal (i.e., having two or more major peaks or modes)
  - Rectangular (i.e., having no peaks or modes)
- Excel function:

MODE.SNGL(number1,number2,...). Returns the most frequently occurring value of the range of data

Example of Mode

Measurements
x
7
7
5
7
5
8
7
6
5

Major mode: 7

Minor mode: 5

	C	D
1	Count	=COUNT(A2:A170)
2	Mean	=AVERAGE(A2:A170)
3	Standard Error of the Mean	=STDEV.P(A2:A170)/SQRT(COUNT(A2:A170))
4	Median	=MEDIAN(A2:A170)
5	Mode	=MODE.SNGL(A2:A170)
6		
7	Variance	=VAR
8	Standard Deviation	=STDEV
9	Rang	
10	Interq	0,1)
11	Maxir	
12	Minim	
13		
14	Skewness Coefficient	
15	SE Skewness	=SQRT(6/COUNT(A2:A170))
16	Standard Coefficient of Skewness	=D14/D15
17	Kurtosis Coefficient	=KURT(A2:A170)
18	SE Kurtosis	=SQRT(24/COUNT(A2:A170))
19	Standard Coefficient of Kurtosis	=D17/D18

TASK

Enter the following formula in cell D5 to calculate the major mode of variable c_community:

=MODE.SNGL(A2:A170)

	C	D	E
1	Count	169	
2	Mean	28.84023669	
3	Standard Error of the Mean	0.478693704	
4	Median	29	
5	Mode	22	
6			
7	Variance	38.72595	
8	Standard Deviation		
9	Range		
10	Interquartile Range		
11	Maximum		
12	Minimum	15	
13			
14	Skewness Coefficient	0.073045168	
15	SE Skewness	0.188422288	
16	Standard Coefficient of Skewness	0.387667347	
17	Kurtosis Coefficient	-1.044172509	
18	SE Kurtosis	0.376844576	
19	Standard Coefficient of Kurtosis	-2.770830673	

Excel displays the mode as 22. This means that 22 is the most frequently occurring value in the dataset.

Descriptive Statistics – Measures of Central Tendency

End of Presentation

Descriptive Statistics

- Statistics
  - Summary measures calculated for a sample dataset.
- Parameters
  - Summary measures calculated for a population dataset.
- Used to describe variables
  - Measures of central tendency, e.g., mean, median, mode
  - Measures of dispersion, e.g., standard deviation, variance, range
  - Measures of relative position, e.g., percentile, quartile
  - Graphs and charts, e.g., scatterplot, column chart, histogram

Measures of Dispersion

Designed to give information concerning the amount of dispersion of scores about a central value.

Researchers typically report the best measures of central tendency and dispersion for each variable. The best measure to report varies based on the shape of a variable's distribution and scale of measurement.

- Interval/ratio data – standard deviation, variance, and range can be calculated and reported, as appropriate.
- Ordinal/nominal data - range can and should be reported; use of the standard deviation or variance is wrong.

	Measures of Dispersion
Nominal data	Range
Ordinal data	Range
Interval data	Standard Deviation, Variance, Range
Ratio data	Standard Deviation, Variance, Range

Standard Deviation (*S, SD,* σ)

- Indicates how much scores deviate below and above the mean
- For normally distributed data
  - 68.2% of the distribution falls within ± 1 *SD* of the mean
  - 95.4% of the distribution falls within ± 2 *SD* of the mean
  - 99.6%of the distribution falls within ± 3 *SD* of the mean
- Formulas

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N}} \qquad s = \sqrt{\frac{\sum (X - m)^2}{N}}$$

(Note: dividing by (N – 1) rather than N for sample standard deviation results in an unbiased estimate of population standard deviation.)

- Excel functions:

STDEV.S(number1,number2,...). Returns the unbiased estimate of population standard deviation, where numbers represent the range of numbers

STDEV.P (number1,number2,...). Returns the population standard deviation, where numbers represent the range of numbers

Example of Standard Deviation

Measurements	Deviations	Squared deviations
x	x − mean	
7	0.67	0.4444444
7	0.67	0.4444444
5	-1.33	1.7777778
7	0.67	0.4444444
5	-1.33	1.7777778
8	1.67	2.7777778
7	0.67	0.4444444
6	-0.33	0.1111111
5	-1.33	1.7777778
57	0.00	10.00

$$\sum (X - \bar{X})^2$$

$$\bar{X} = \frac{\sum X}{N} = \frac{57}{9} = 6.33$$

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N}} = \sqrt{\frac{10}{9}} = 1.05$$

For an unbiased estimate of the population standard deviation, $N - 1$ is used in the formula in place of N, otherwise the formula will underestimate the population sum of squares.

	C	D
1	Count	=COUNT(A2:A170)
2	Mean	=AVERAGE(A2:A170)
3	Standard Error of the Mean	=STDEV.P(A2:A170)/SQRT(COUNT(A2:A170))
4	Median	=MEDIAN(A2:A170)
5	Mode	=MODE.SNGL(A2:A170)
6		
7	Variance	=VAR.P(A2:A170)
8	Standard Deviation	=STDEV.P(A2:A170)
9	Range	=MAX(A2:A170)-MIN(A2:A170)
10	Interquartile Range	=QUARTILE.INC(A2:A170,3)-QUARTILE.INC(A2:A170,1)
11	Maximum	
12	Minimum	
13		
14	Skewness C	
15	SE Skewnes	
16	Standard Coefficient of Skewness	=D14/D15
17	Kurtosis Coe	
18	SE Kurtosis	
19	Standard Co	

TASK

Enter the following formula in cell D8 to calculate the standard deviation for c_community:
=STDEV.P(A2:A170)

Note: this measure is not an unbiased estimate of the population *SD*. If an unbiased estimate of the population *SD* is desired use the formula =STDEV.S(A2:A170).

Excel displays the *SD* as 6.223018156. This sample statistic is typically reported as *SD* = 6.22 in the results section of a research paper, as appropriate. It represents a measure of the spread of the distribution.

	C	
1	Count	
2	Mean	
3	Standard Error	
4	Median	
5	Mode	2
6		
7	Variance	38.72[illegible]497
8	Standard Deviation	6.223018156
9	Range	25
10	Interquartile Range	10
11	Maximum	40
12	Minimum	15
13		
14	Skewness Coefficient	0.073045168
15	SE Skewness	0.188422288
16	Standard Coefficient of Skewness	0.387667347
17	Kurtosis Coefficient	-1.044172509
18	SE Kurtosis	0.376844576
19	Standard Coefficient of Kurtosis	-2.770830673

Variance (S^2, σ^2)

- Variance is the average of each score's squared difference from the mean.
- Not a very useful as a descriptive statistic. Important value used in certain techniques (e.g., the analysis of variance or ANOVA)
- The formula for the population and sample variances are given below.

$$S^2 = \frac{\sum (X - \overline{X})^2}{N} \qquad s^2 = \frac{\sum (X - m)^2}{N}$$

(Note: dividing by (N – 1) rather than N for sample variance results in an unbiased estimate of population variance.)

- Excel functions:

VAR.S(number1,number2,...). Returns the unbiased estimate of population variance, with numbers representing the range of numbers.

VAR.P (number1,number2,...). Returns the population variance, with numbers representing the range of numbers.

Example of Variance

Measurements	Deviations	Squared deviations
x	x − mean	$(x - \bar{X})^2$
7	0.67	0.4444444
7	0.67	0.4444444
5	-1.33	1.7777778
7	0.67	0.4444444
5	-1.33	1.7777778
8	1.67	2.7777778
7	0.67	0.4444444
6	-0.33	0.1111111
5	-1.33	1.7777778
57	0.00	10.00

$$S^2 = \frac{\sum (X - \bar{X})^2}{N} = \frac{10}{9} = 1.11$$

For an unbiased estimate of the population standard deviation, $N - 1$ is used in the formula in place of N, otherwise the formula will underestimate the population sum of squares.

	C	D
1	Count	=COUNT(A2:A170)
2	Mean	=AVERAGE(A2:A170)
3	Standard Error of the Mean	=STDEV.P(A2:A170)/SQRT(COUNT(A2:A170))
4	Median	=MEDIAN(A2:A170)
5	Mode	=MODE.SNGL(A2:A170)
6		
7	Variance	=VAR.P(A2:A170)
8	Standard Deviation	=STDEV.P(A2:A170)
9	Range	=MAX(A2:A170)-MIN(A2:A170)
10	Interquartile Range	=QUARTILE.INC(A2:A170,3)-QUARTILE.INC(A2:A170,1)
11	Maximum	
12	Minimum	
13		
14	Skewness	
15	SE Skewness	
16	Standard C	
17	Kurtosis C	
18	SE Kurtosi	
19	Standard C	

TASK

Enter the following formula in cell D8 to calculate the variance of variable c_community:

=VAR.P(A2:A170)

Note: this measure is not an unbiased estimate of the population variance. If an unbiased estimate of the population variance is desired use the formula =VAR.S(A2:A170).

Excel displays the variance as 38.72595497. Like the SD, it It represents a measure of the spread of the distribution.

	C	
1	Count	
2	Mean	
3	Standard Error of th	
4	Median	9
5	Mode	
6		
7	Variance	38.72595497
8	Standard Deviation	6.223018156
9	Range	25
10	Interquartile Range	10
11	Maximum	40
12	Minimum	15
13		
14	Skewness Coefficient	0.073045168
15	SE Skewness	0.188422288
16	Standard Coefficient of Skewness	0.387667347
17	Kurtosis Coefficient	-1.044172509
18	SE Kurtosis	0.376844576
19	Standard Coefficient of Kurtosis	-2.770830673

Range

- The range of a distribution is calculated by subtracting the minimum score from the maximum score.

$$Range = X_{Max} - X_{Min}$$

- The range is not very stable (reliable) because it is based on only two scores. Consequently, outliers have a significant effect on the range of a variable.
- Excel formula:

=MAX(number1,number2,...)–MIN(number1,number2,...)

Note: MAX(number1,number2,...) returns the maximum value in a set of numbers and MIN(number1,number2,...) returns the minimum value in a set of numbers.

Example of Range

Measurements	Ranked Data
x	x
7	5
7	5
5	5
7	6
5	7
8	7
7	7
6	7
5	8

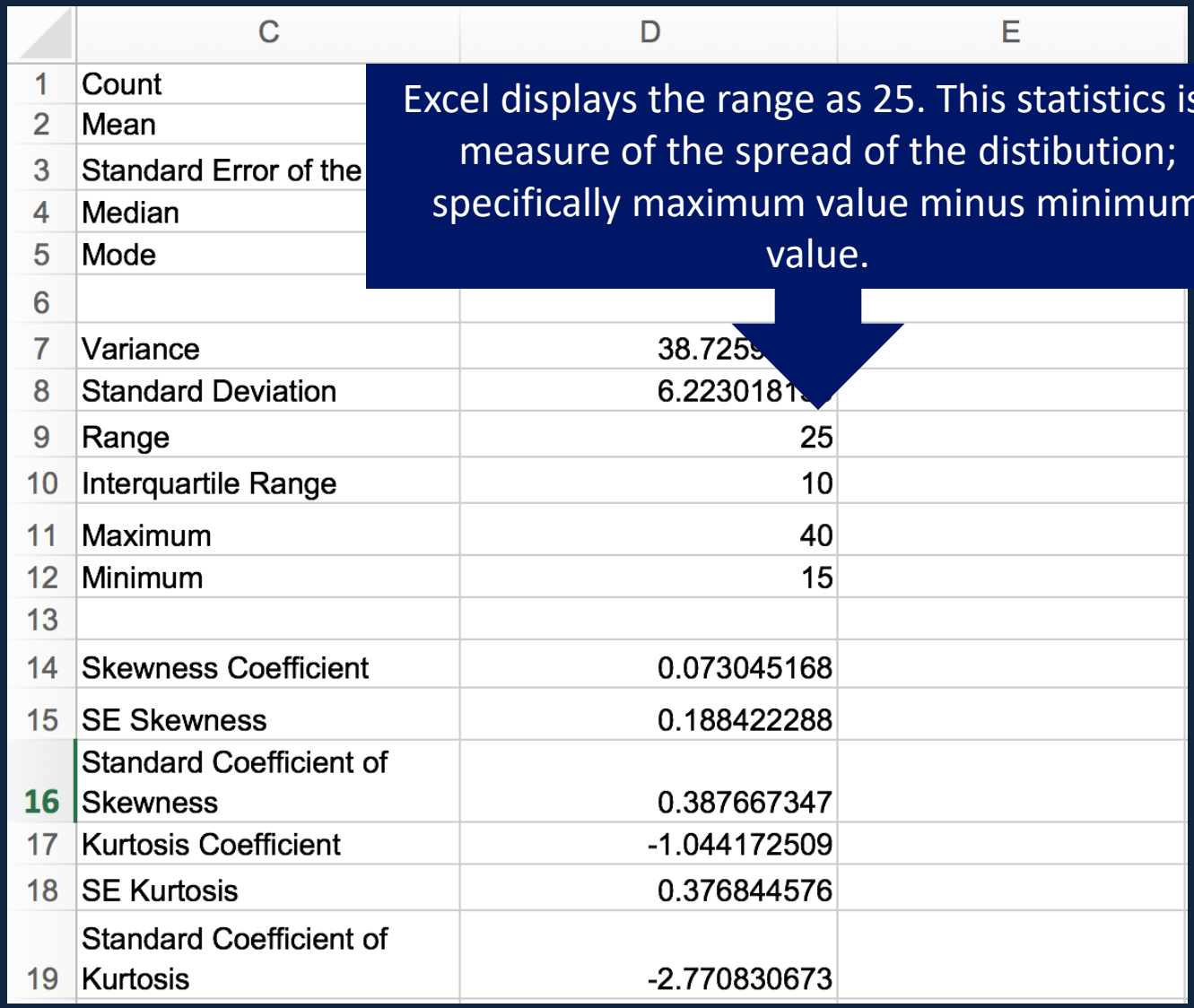
Range = maximum value – minimum value = 8 – 5 = 3

TASK

Enter the following formula in cell D9 to calculate the range of variable c_community:

=MAX(A2:A170)-MIN(A2:A170)

	C	D
1	Count	
2	Mean	
3	Standard Error c	
4	Median	
5	Mode	
6		
7	Variance	=VAR.P(A2:A
8	Standard Deviation	=STDEV.P(A2:A
9	Range	=MAX(A2:A170)-MIN(A2:A170)
10	Interquartile Range	=QUARTILE.INC(A2:A170,3)-QUARTILE.INC(A2:A170,1)
11	Maximum	=MAX(A2:A170)
12	Minimum	=MIN(A2:A170)
13		
14	Skewness Coefficient	=SKEW(A2:A170)
15	SE Skewness	=SQRT(6/COUNT(A2:A170))
16	Standard Coefficient of Skewness	=D14/D15
17	Kurtosis Coefficient	=KURT(A2:A170)
18	SE Kurtosis	=SQRT(24/COUNT(A2:A170))
19	Standard Coefficient of Kurtosis	=D17/D18



	C	D	E
1	Count		
2	Mean		
3	Standard Error of the		
4	Median		
5	Mode		
6			
7	Variance	38.725	
8	Standard Deviation	6.2230181	
9	Range	25	
10	Interquartile Range	10	
11	Maximum	40	
12	Minimum	15	
13			
14	Skewness Coefficient	0.073045168	
15	SE Skewness	0.188422288	
16	Standard Coefficient of Skewness	0.387667347	
17	Kurtosis Coefficient	-1.044172509	
18	SE Kurtosis	0.376844576	
19	Standard Coefficient of Kurtosis	-2.770830673	

Descriptive Statistics - Dispersion

End of Presentation

Descriptive Statistics

- Statistics
  - Summary measures calculated for a sample dataset.
- Parameters
  - Summary measures calculated for a population dataset.
- Used to describe variables
  - Measures of central tendency, e.g., mean, median, mode
  - Measures of dispersion, e.g., standard deviation, variance, range
  - Measures of relative position, e.g., percentile, quartile
  - Graphs and charts, e.g., scatterplot, column chart, histogram

Measures of Relative Position

- Measures of relative position indicate how high or low a score is in relation to other scores in a distribution.
  - Answers the question: Where is this value with respect to the other values in the population or in the sample?
- A percentile (P) is a measure that tells one the percent of the total frequency that scored at or below that measure.
  - The kth percentile (P_k) of a set of data is a value such that k percent of the observations are less than or equal to the value.
- A quartile (Q) divides the data into four equal parts based on their statistical ranks and position from the bottom.
  - Q_1 has 25% of the data at or below it.
  - Q_2 (median) has 50% of the data at or below it; it is equal to the median.
  - Q_3 has 75% of the data at or below it.
  - Interquartile range (IQR) = $Q_3 - Q_1$; the range of the middle 50% of the data.
- Percentiles and quartiles are cutoff scores and not ranges of values.
- Standardized scores (e.g., z-scores).

Measures of Relative Position

- Excel functions:

PERCENTILE.INC(array,k). Returns the kth percentile in a range of numbers.

QUARTILE.INC(array,quart). Returns the specified quartile, in a range of numbers.

Note: *k* = the percentile value in the range 0 to 1, inclusive; quart = 0 returns the minimum value, quart = 1 returns Q_1, quart = 2 returns Q_2 (median), quart = 3 returns Q_3, quart = 4 returns the maximum value.

Calculating Measures of Relative Position

	C	D	E
20			
21	90th percentile	=PERCENTILE.INC(A2:A170,0.9)	
22	10th percentile	=PERCENTILE.INC(A2:A170,0.1)	
23	1st quartile	=QUARTILE.INC(A2:A170,1)	
24	2nd quartile	=QUARTILE.INC(A2:A170,2)	
25	3rd quartile	=QUARTILE.INC(A2:A170,3)	
26			

TASK

Enter the formulas in cells D12:D25 as shown on the worksheet to calculate P_{90}, P_{10}, Q_1, Q_2, and Q_3.

Calculating Measures of Relative Position

	C	D	E
20			
21	90th percentile	37.2	
22	10th percentile	21	
23	1st quartile	24	
24	2nd quartile	29	
25	3rd quartile	34	
26			

Excel displays percentiles and quartiles, as shown. These statistics can be interpreted as follows:

90% of c_community scores are at or below a score of 37.2
10% of c_community scores are at or below a score of 21
25% of c_community scores are at or below a score of 24
50% of c_community scores are at or below a score of 29
75% of c_community scores are at or below a score of 34

Note: interpretations assume c_community is normally distributed

z-Scores

- A standard score is a general term referring to a score that has been transformed for reasons of convenience, comparability, etc.
- The basic type of standard score, known as a *z*-score, is a measure of a score's distance from the mean in standard deviation units. For example…
  - If $z = 0$, it's on the mean.
  - If $z = 1.5$, it's 1.5 standard deviations above the mean.
  - If $z = -1$, it's 1.0 standard deviations below the mean.
- A *z*-score distribution is the standard normal distribution, $N(0,1)$, with mean = 0 and standard deviation = 1. The formula for calculating *z*-scores from raw scores is

$$z = \frac{X - \bar{X}}{SD}$$

where X = raw score, $\bar{X}$ = raw score mean, and SD = raw score standard deviation.

- Most other standard scores are linear transformations of *z*-scores, with different means and standard deviations. For example, *T*-scores, used in the Minnesota Multiphasic Personality Inventory (MMPI), have $M = 50$ and $SD = 10$, and SAT scores have $M = 500$ and $SD = 100$.

Why *z*-Scores?

- Transforming raw scores to *z*-scores facilitates making comparisons, especially when using different scales.
- A *z*-score provides information about the relative position of a score in relation to other scores in a sample or population.
  - A raw score provides no information regarding the relative standing of the score relative to other scores.
  - A *z*-score tells one how many standard deviations the score is from the mean. It also provides the approximate percentile rank of the score relative to other scores. For example, a *z*-score of 1 is 1 standard deviation above the mean and equals the 84.1 percentile rank (50% of occurrences fall below the mean and 34.1% of the occurrences fall between 0 and 1; 50% + 34.1% = 84.1%).

Calculating *z*-Scores from Raw Scores

X	$\bar{X}$	$X - \bar{X}$	SD	$z = \frac{X - \bar{X}}{SD}$
23	28.84	-5.84	6.24	-.94
22	28.84	-6.84	6.24	-1.10
33	28.84	4.16	6.24	.67
19	28.84	-9.84	6.24	-1.58

A raw score of 23 equals a *z*-score of – .94, indicating both scores are .94 standard deviations below the mean.

Calculating Raw Scores from *z*-Scores

z	SD	zSD	$\bar{X}$	$X = zSD + \bar{X}$
-.94	6.22	-5.85	28.84	22.99
-1.10	6.22	-6.84	28.84	22
.67	6.22	4.17	28.84	33.01
-1.58	6.22	-9.83	28.84	19.01

Differences (±.01) in calculated raw scores and actual raw scores are the result of rounding.

Open the dataset *Motivation.xlsx*.
Click the worksheet Descriptive Statistics tab (at the bottom of the worksheet).

File available at
http://www.watertreepress.com/stats

	A	B	C	D	E	F
1	c_community		Count	169		z-scores
2	23		Mean	28.84023669		-0.938489418
3	22		Mean	0.478693704		-1.099183148
4	23		Median	29		-0.938489418
5	23		Mode	22		-0.938489418
6	22					-1.099183148
7	32		Variance	38.72595497		0.507754153
8	24		Standard Deviation	6.223018156		-0.777795688
9	22		Range	25		-1.099183148
10	28		Interquartile Range	10		-0.135020767
11	25		Maximum	40		-0.617101958
12	22		Minimum	15		-1.099183148
13	23					-0.938489418
14	33		Skewness Coefficient	0.073045168		0.668447883
15	19		SE Skewness	0.188422288		-1.581264338

TASK
Convert classroom community (c_community) raw scores into *z*-scores.

Calculating *z*-Scores from Raw Scores

	F
1	z-scores
2	=STANDARDIZE(A2,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
3	=STANDARDIZE(A3,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
4	=STANDARDIZE(A4,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
5	=STANDARDIZE(A5,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
6	=STANDARDIZE(A6,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
7	=STANDARDIZE(A7,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
8	=STANDARDIZE(A8,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
9	=STANDARDIZE(A9,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
10	=STANDARDIZE(A10,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
11	=STANDARDIZE(A11,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
12	=STANDARDIZE(A12,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
13	=STANDARDIZE(A13,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
14	=STANDARDIZE(A14,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))
15	=STANDARDIZE(A15,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170))

	F
1	z-scores
2	-0.938489418
3	-1.099183148
4	-0.938489418
5	-0.938489418
6	-1.099183148
7	0.507754153
8	-0.777795688
9	-1.099183148
10	-0.135020767
11	-0.617101958
12	-1.099183148
13	-0.938489418
14	0.668447883
15	-1.581264338

Excel includes the following function that converts raw scores to z-scores:

STANDARDIZE(number,AVERAGE(number1,number2,...),STDEV.P(number1,number2,...)). Returns a standardized value.

Enter the following formula in cell F2:
=STANDARDIZE(A2,AVERAGE(A$2:A$170),STDEV.P(A$2:A$170)).
Click on cell F2, hold the Shift key down, and click on cell F170 in order to select the range F2:F170.
Using the Excel Edit menu, select Fill Down. The *z*-scores are displayed in column F.

Calculating z-Scores from Raw Scores

	A	B	C	D	E	F
1	c_community		Count	169		z-scores
2	23		Mean	28.84023669		-0.938489418
3	22		Mean	0.47 3704		-1.099183148
4	23		Median	29		-0.938489418
5	23		Mode			-0.938489418
6	22					-1.099183148
7	32		Variance			0.507754153
8	24		Standard Deviation			-0.777795688
9						
10						
11						
12						
13						
14						
15						

An alternative method is to use the z-score mathematical formula

$Z = (X - \bar{x})/SD$.

First, calculate the c-community mean in cell D2 using the formula

=AVERAGE(A2:A170).

The mean is 28.84.

Calculating z-Scores from Raw Scores

	A	B	C	D	E	F
1	c_community		Count	169		z-scores
2	23		Mean	28.84023669		-0.938489418
3	22		Mean	0.478693704		-1.099183148
4	23		Median	29		-0.938489418
5	23		Mode	22		-0.938489418
6	22					-1.099183148
7	32		Variance	38.72595497		0.507754153
8	24		Standard Deviation	6.223018156		-0.777795688
9	22		Range	25		-1.099183148
10	28		Interquartile Range	10		-0.135020767
11	25		Maximum	40		-0.617101958
12						
13						
14						
15						

Next, calculate the c-community standard deviation in cell D8 using the formula =STDEV.P(A2:A170).
The standard deviation is 6.22.

Calculating *z*-Scores from Raw Scores

	D	E	F
1	=COUNT(A2:A170)		z-scores
2	=AVERAGE(A2:A170)		=(A2-D$2)/D$8
3	=STDEV.P(A2:A170)/SQRT(COUNT(A2:A170))		=(A3-D$2)/D$8
4	=MEDIAN(A2:A170)		=(A4-D$2)/D$8
5	=MODE.SNGL(A2:A170)		=(A5-D$2)/D$8
6			=(A6-D$2)/D$8
7	=VAR.P(A2:A170)		=(A7-D$2)/D$8
8	=STDEV.P(A2:A170)		=(A8-D$2)/D$8
9	=MAX(A2:A170)-MIN(A2:A170)		=(A9-D$2)/D$8
10	=QUARTILE.INC(A2:A170,3)-QUARTILE.INC(A2:A170,1)		=(A10-D$2)/D$8
11	=MAX(A2:A170)		=(A11-D$2)/D$8
12	=MIN(A2:A170)		=(A12-D$2)/D$8
13			=(A13-D$2)/D$8
14	=SKEW(A2:A170)		=(A14-D$2)/D$8
15	=SQRT(6/COUNT(A2:A170))		=(A15-D$2)/D$8

	A	B	C	D	E	F
1	c_community		Count	169		z-scores
2	23		Mean	28.84023669		-0.938489418
3	22		Mean	0.478693704		-1.099183148
4	23		Median	29		-0.938489418
5	23		Mode	22		-0.938489418
6	22					-1.099183148
7	32		Variance	38.72595497		0.507754153
8	24		Standard Deviation	6.223018156		-0.777795688
9	22		Range	25		-1.099183148
10	28		Interquartile Range	10		-0.135020767
11	25		Maximum	40		-0.617101958
12	22		Minimum	15		-1.099183148
13	23					-0.938489418
14	33		Skewness Coefficient	0.073045168		0.668447883
15	19		SE Skewness	0.188422288		-1.581264338

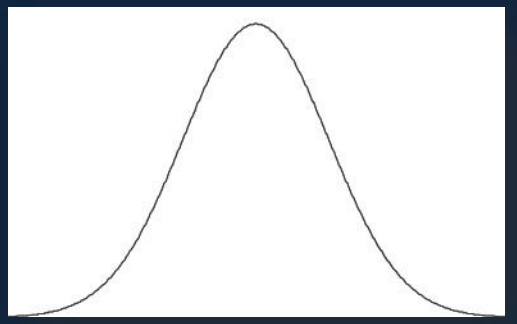
Enter the *z*-score formula in cell F2:
=(A2-D$2)/D$8
Where D2 is the mean and D8 is the standard deviation.
(Note the use of absolute addresses for cells D2 and D8.)

Click on cell F2, hold the Shift key down, and click on cell F170 in order to select the range F2:F170.
Using the Excel Edit menu, select Fill Down. The *z*-scores are displayed in column F.

Measures of Relative Position

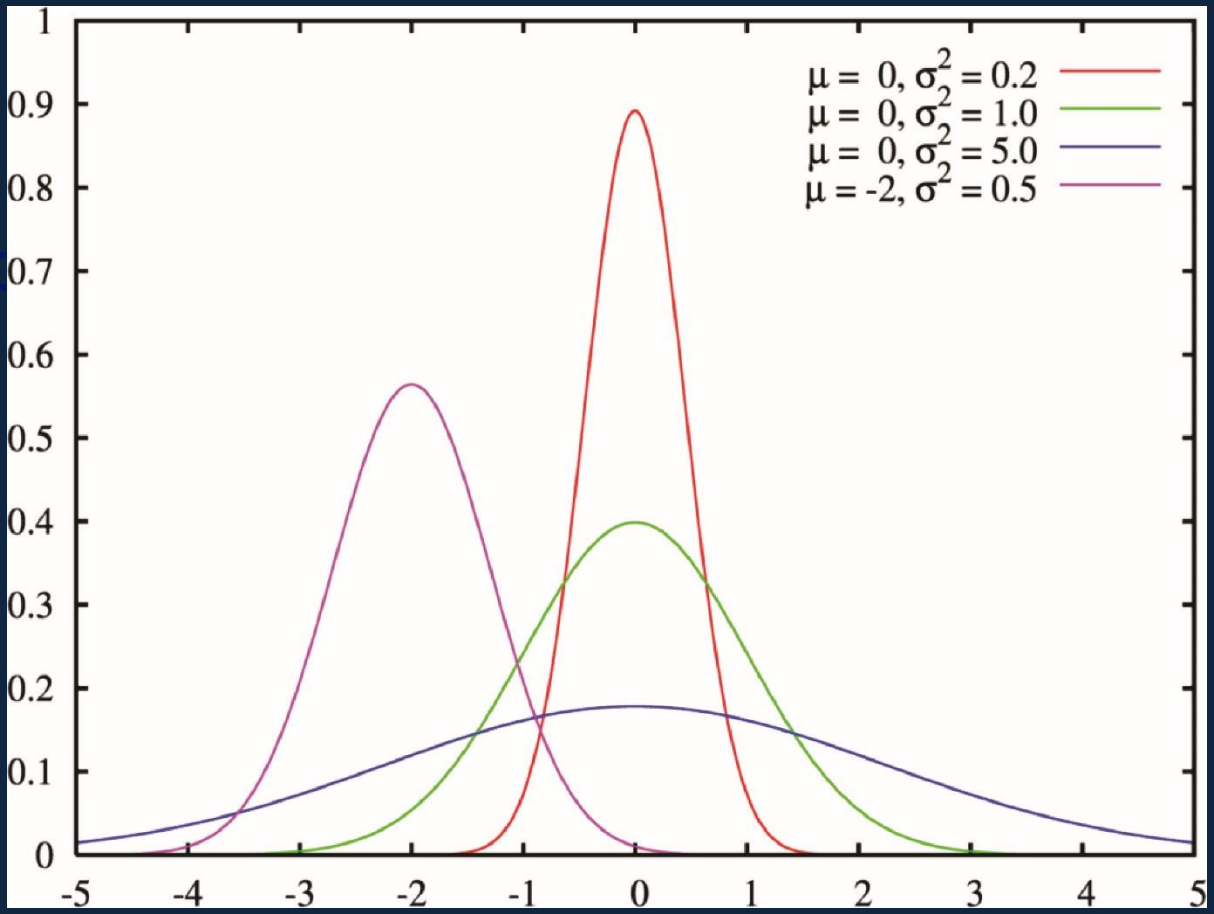
End of Presentation

Normal Curve



- The normal or Gaussian curve is a family of distributions.
- It is a smooth curve and is referred to as a probability density curve for a random variable, *x*, rather than a frequency curve as one sees in a histogram.
  - The area under the graph of a density curve over some interval represents the probability of observing a value of the random variable in that interval.
- The family of normal curves has the following characteristics:
  - Bell-shaped
  - Symmetrical about the mean (the line of symmetry)
  - Tails are asymptotic (they approach but do not touch the x-axis)
  - The total area under any normal curve is 1 because there is a 100% probability that the curve represents all possible occurrences of the associated event (i.e., normal curves are probability density curves)
  - Involve a large number of cases

Normal Curve



Various normal curves are shown above. The line of symmetry for each is at μ (the mean). The curve will be peaked (skinnier or leptokurtic) if the σ (standard deviation) is smaller and flatter or platykurtic if it is larger.

Empirical Rule

In a normal distribution (or approximately normal distribution) with mean μ and standard deviation σ, the approximate areas under the normal curve are as follows:

- 34.1% of the occurrences will fall between μ and 1σ
- 13.6% of the occurrences will fall between 1σ & 2σ
- 2.15% of the occurrences will fall between 2σ & 3σ

If one adds percentages, approximately:

- 68% of the distribution lies within $\pm$ one σ of the mean.
- 95% of the distribution lies within $\pm$ two σ of the mean.
- 99.7% of the distribution lies within $\pm$ three σ of the mean.

These percentages are known as the *empirical rule*.

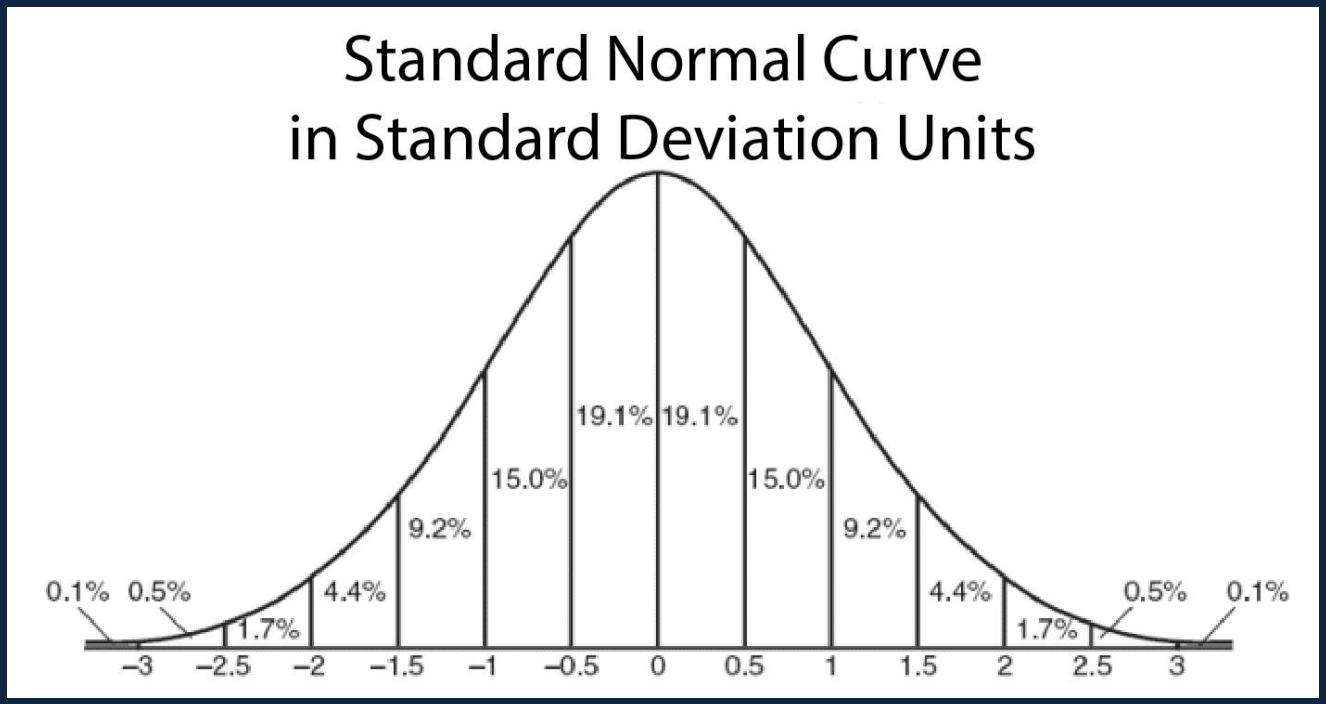
Example: Given a normal curve (i.e., a density curve), if , $\mu = 10$ and $\sigma = 2$, the probability that x is between 8 and 12 is .68.

Chebyshev's Theorem

The empirical rule does not apply to distributions that are not normal or approximately normal.

For all distributions (including non-normal distributions) Chebyshev's theorem applies and states:

- At least 75% of all scores will fall within 2 standard deviations above and below the mean.
- At least 89% of all scores will fall within 3 standard deviations above and below the mean.



When $\mu = 0$ and $\sigma = 1$, the distribution is called the standard normal distribution.

- 34.1% of the occurrences will fall between 0 and 1
- 13.6% of the occurrences will fall between 1 & 2
- 2.15% of the occurrences will fall between 2 & 3

Univariate Normality

- Univariate refers to one variable. Normality refers to the shape of a variable's frequency distribution.
  - Symmetrical and shaped like a bell-curve.
- Parametric tests assume normality.
  - The dependent variable (DV) is approximately normally distributed.
- The perfectly normal univariate distribution has standardized kurtosis and skewness statistics equal to zero and mean = mode = median
- The assumption of univariate normality does not require a perfectly normal shape.
  - Many parametric procedures, e.g., one-way ANOVA, are robust in the face of light to moderate departures from normality.

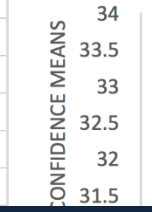
Procedures

Several tools are available in Excel to evaluate univariate normality

- Create a histogram to observe the shape of the distribution and to conduct a preliminary evaluation of normality
- Calculate standard coefficients of skewness and kurtosis to determine if the shape of the distribution differs from that of a normal distribution
- Calculating the presence of extreme outliers
- Use the Kolmogorov-Smirnov test to determine if the data come from a population with a normal distribution

Evaluating Univariate Normality

	A	B	C	D	E
1	gender	comconf1	comconf2	comconf3	
2	1	32	35	35	comconf1
3	1	38	40	39	comconf2
4	1	23	30	32	comconf3
5	1	31	34	36	
6	1	32	34	35	
7	1	34	37	38	
8	1	29	26	28	
9	1	31	32	35	
10	1	37	35	30	
11	1	39	37	36	
12	1	33	32	33	
13	1	32	28	31	
14	1	30	35	35	
15	1	30	27	30	

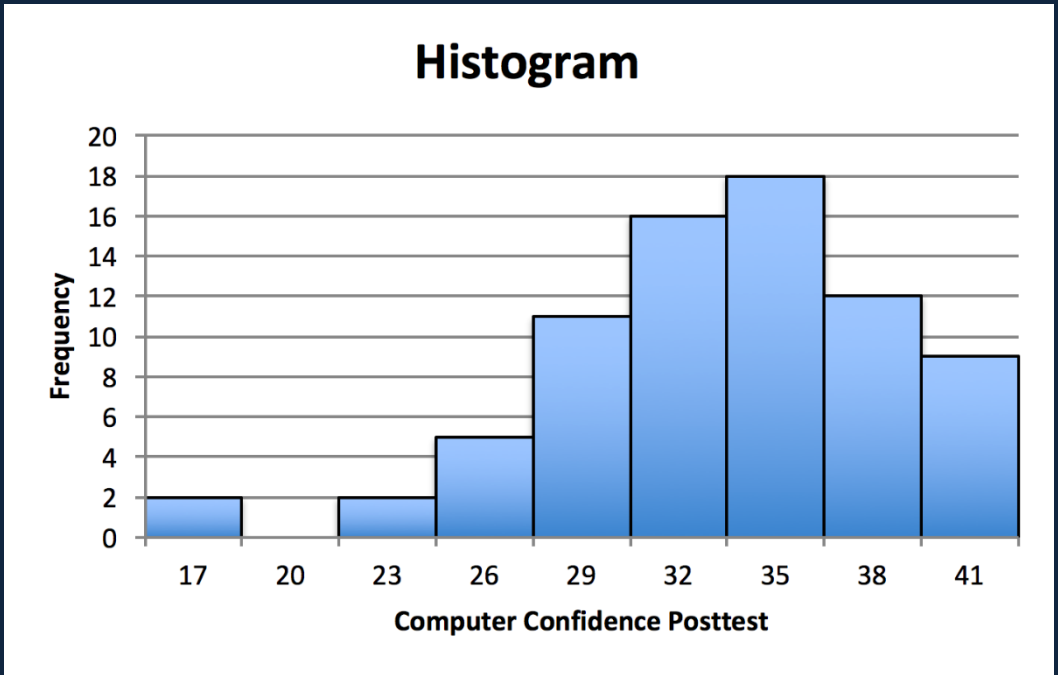


Open the dataset *Computer Anxiety.xlsx*.
Click the worksheet Charts tab (at the bottom of the worksheet).

File available at
http://www.watertreepress.com/stats

TASK
Evaluate computer confidence posttest (comconf2) for univariate normality.

Creating a Histogram



Create a histogram that displays computer confidence posttest (comconf2) in accordance with the procedures described in the textbook and the Histograms Power Point presentation. It reveals a non-symmetric, negatively-skewed shape that approximates a bell shape.

The issue now is to determine whether or not univariate normality is tenable. That is, to determine the extent of the deviations from normality.

Calculating Standard Coefficient of Skewness

Skewness is based on the third moment of the distribution, or the sum of cubic deviations from the mean. It measures deviations from perfect symmetry.

- Positive skewness indicates a distribution with a heavier positive (right-hand) tail than a symmetrical distribution.
- Negative skewness indicates a distribution with a heavier negative tail.

Excel function:

SKEW(number1,number2,...). Returns the skewness statistic of a distribution.

The standard error of skewness (SES) is a measure of the accuracy of the skewness coefficient and is equal to the standard deviation of the sampling distribution of the statistic.

$$SES = \sqrt{\frac{6}{N}}$$

Normal distributions produce a skewness statistic of approximately zero. The skewness coefficient divided by its standard error can be used as a test of normality. That is, one can reject normality if this statistic is less than –2 or greater than +2.

C2:C76 is the address of comconf2 values.

Enter the formulas displayed in cells B80:B82 to calculate the skewness coefficient, the standard error of skewness, and the standard coefficient of skewness using the Charts tab.

	A	B	C
80	Skewness	=SKEW(C2:C76)	
81	SE skewness	=SQRT(6/COUNT(C2:C76))	
82	Standard coefficient of skewness	=B80/B81	
83	Kurtosis	=KURT(C2:C76)	
84	SE kurtosis	=SQRT(24/COUNT(C2:C76))	
85	Standard coefficient of kurtosis	=B83/B84	

Data | Charts | Independent t-Test | Dependent t-Test | Chi-Square Contingency Table | Mann-Whitney U T

Ready

The standard coefficient of skewness for the computer confidence posttest data (-3.45) indicates a non-normal, negatively-skewed distribution because it is lower than -2.

	A	B	C
80	Skewness	-0.976007795	
81	SE skewness	0.282842712	
82	Standard coefficient of skewness	-3.450708652	
83	Kurtosis	1.458045582	
84	SE kurtosis	0.565685425	
85	Standard coefficient of kurtosis	2.577484795	

Data | Charts | Independent t-Test | Dependent t-Test | Chi-Square Contingency Table | Mann-Whitney U Tes

Ready

Calculating Standard Coefficient of Kurtosis

Kurtosis is derived from the fourth moment (i.e., the sum of quartic deviations). It captures the heaviness or weight of the tails relative to the center of the distribution. Kurtosis measures heavy-tailedness or light-tailedness relative to the normal distribution.

- A heavy-tailed distribution has more values in the tails (away from the center of the distribution) than the normal distribution, and will have a negative kurtosis.
- A light-tailed distribution has more values in the center (away from the tails of the distribution) than the normal distribution, and will have a positive kurtosis.

Excel function:

KURT(number1,number2,...). Returns the kurtosis statistic of a distribution.

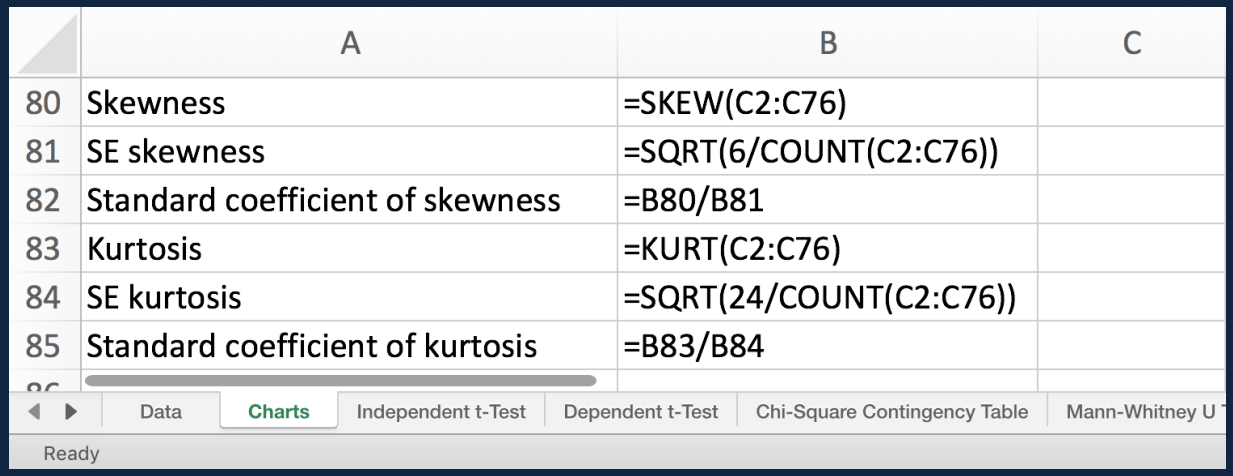
The standard error of kurtosis is a measure of the accuracy of the kurtosis coefficient and is equal to the standard deviation of the sampling distribution of the statistic.

$$SEK = \sqrt{\frac{24}{N}}$$

Normal distributions produce a kurtosis statistic of approximately zero. The kurtosis coefficient divided by its standard error can be used as a test of normality. That is, one can reject normality if this ratio is less than –2 or greater than +2.

C2:C76 is the address of comconf2 values.

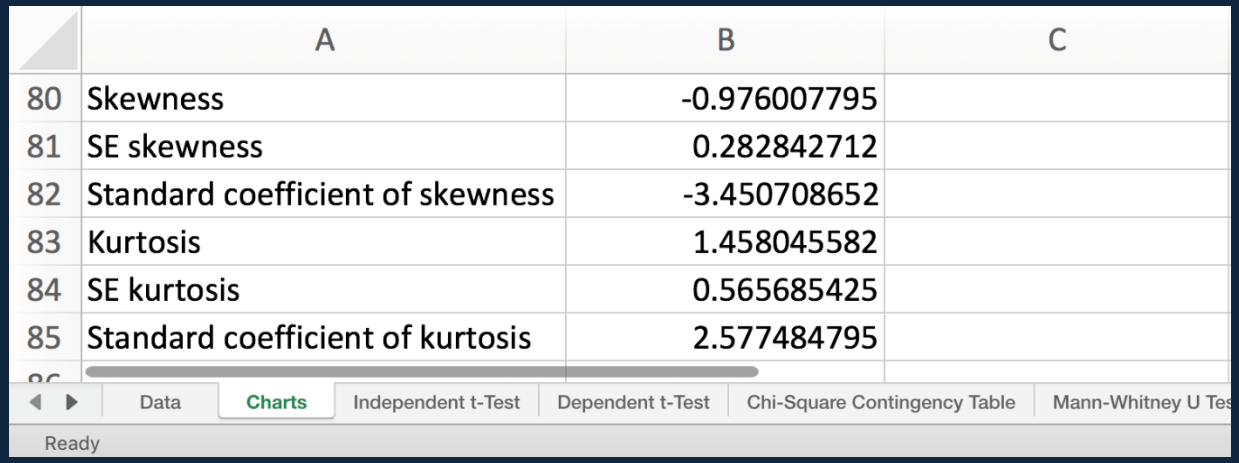
Enter the formulas displayed in cells B83:B85 to calculate the kurtosis coefficient, the standard error of kurtosis, and the standard coefficient of kurtosis using the Charts tab.



	A	B	C
80	Skewness	=SKEW(C2:C76)	
81	SE skewness	=SQRT(6/COUNT(C2:C76))	
82	Standard coefficient of skewness	=B80/B81	
83	Kurtosis	=KURT(C2:C76)	
84	SE kurtosis	=SQRT(24/COUNT(C2:C76))	
85	Standard coefficient of kurtosis	=B83/B84	

Data | Charts | Independent t-Test | Dependent t-Test | Chi-Square Contingency Table | Mann-Whitney U T

Ready



	A	B	C
80	Skewness	-0.976007795	
81	SE skewness	0.282842712	
82	Standard coefficient of skewness	-3.450708652	
83	Kurtosis	1.458045582	
84	SE kurtosis	0.565685425	
85	Standard coefficient of kurtosis	2.577484795	

Data | Charts | Independent t-Test | Dependent t-Test | Chi-Square Contingency Table | Mann-Whitney U Tes

Ready

The standard coefficient of kurtosis for the computer confidence posttest data (2.58) indicates a non-normal, peaked (as opposed to flat) distribution because it is higher than 2.

Calculating Extreme Outliers

Outliers are anomalous observations that have extreme values with respect to a single variable.

- Reasons for outliers vary from data collection or data entry errors to valid but unusual measurements.
- Normal distributions do not include extreme outliers.
- It is common to define extreme univariate outliers as cases that are more than three standard deviations above the mean of the variable or less than three standard deviations from the mean.

Open the dataset *Computer Anxiety 3dEd.xlsx*

File available at
http://www.watertreepress.com/stats

	P
1	z-scores
2	=STANDARDIZE(J2,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
3	=STANDARDIZE(J3,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
4	=STANDARDIZE(J4,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
5	=STANDARDIZE(J5,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
6	=STANDARDIZE(J6,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
7	=STANDARDIZE(J7,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
8	=STANDARDIZE(J8,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
9	=STANDARDIZE(J9,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))
10	=STANDARDIZE(J10,AVERAGE(J$2:J$87),STDEV.S(J$2:J$87))

Calculate z-scores for variable computer confidence posttest (comconf2). Enter the formula shown on the worksheet in cell P2. Then select cell P2, hold down the Shift key, and click on cell P87 in order to select the range P2:P87. Then use the Excel Edit menu and Fill Down to replicate the formula.

	P	Q	R
1	z-scores		
2	0.462701274		
3	1.396793047		
4	1.023156338		
5	0.462701274		
6	1.023156338		
7	-0.28457214		
8	1.209974692		
9	-0.4713905		
10	1.023156338		

Extreme outliers have z-scores below – 2 and above +2.

Scan the z-scores to note the following extreme low outliers:
Case 61: -3.09
Case 73: -3.46

Conducting the Kolmogorov-Smirnov Test

	E	F	G	H
1		N	Mean	Deviation
2	c_community	169	28.84024	6.24151156
3				
4	D	2.230745693		
5	D critical	0.104615385		
6				
7	Kolmogorov-Smirnov Test			
8		N	Value	Critical Value
9	D	169	2.230746	0.10461538
10				

Conduct the Kolmogorov-Smirnov Test in accordance with the procedures described in the textbook in order to evaluate the following null hypothesis:

H_o: There is no difference between the distribution of computer confidence posttest data and a normal distribution.

Test results are significant since *D* (2.23) > the critical value (0.10) at the .05 significance level. Therefore, there is sufficient evidence to reject the null hypothesis and assume normality is not tenable.

Conclusion

Univariate normality is not tenable for posttest computer confidence

The histogram reveals a non-symmetrical negatively-skewed shape

The standard coefficient of skewness of -3.45 indicates a non-normal negatively-skewed distribution

The standard coefficient of kurtosis of 2.58 indicates a non-normal leptokurtic (peaked) distribution

There are two low extreme outliers, $z < -2$

The Kolmogorov-Smirnov test results are statistically significant at the .05 level indicating a non-normal distribution

The Normal Curve & Univariate Normality

End of Presentation

Charts

Imagery is the key to understanding statistics

Helps clarify complex data and summary statistics

Supplements the written narrative

One can create and edit a variety of charts using Excel that convey the imagery that promotes meaning and understanding of statistical results. Often, selection of chart type is subjective.

Most charts share the following characteristics

- Two, axes that are drawn at a right angle
- The horizontal axis is the abscissa, or x-axis
- The vertical axis is the ordinate, or y-axis
- The independent variable is plotted on the x-axis, and the dependent variable is plotted on the y-axis

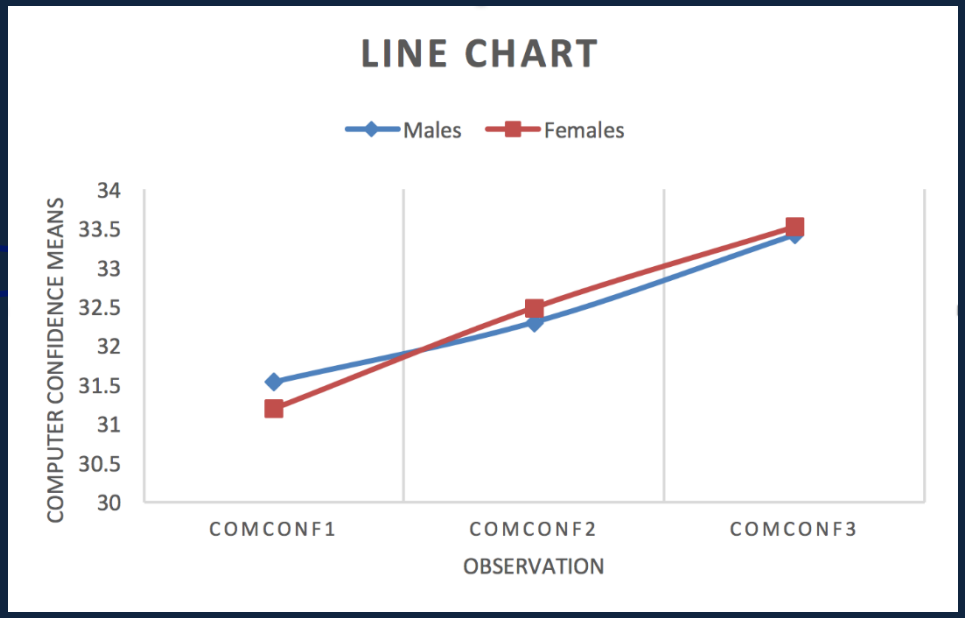
Creating a Chart

1. Select the data to chart (first column includes data for the x-axis; second column includes data for y-axis)
2. Highlight the data
3. Select the type of chart, e.g., column, bar, line, pie, area, scatter, other
4. Insert the chart type, e.g., 2-D column, 3-D column, cylinder, cone, pyramid, other
5. Select the chart and modify the chart layout, as desired

Major Types of Charts

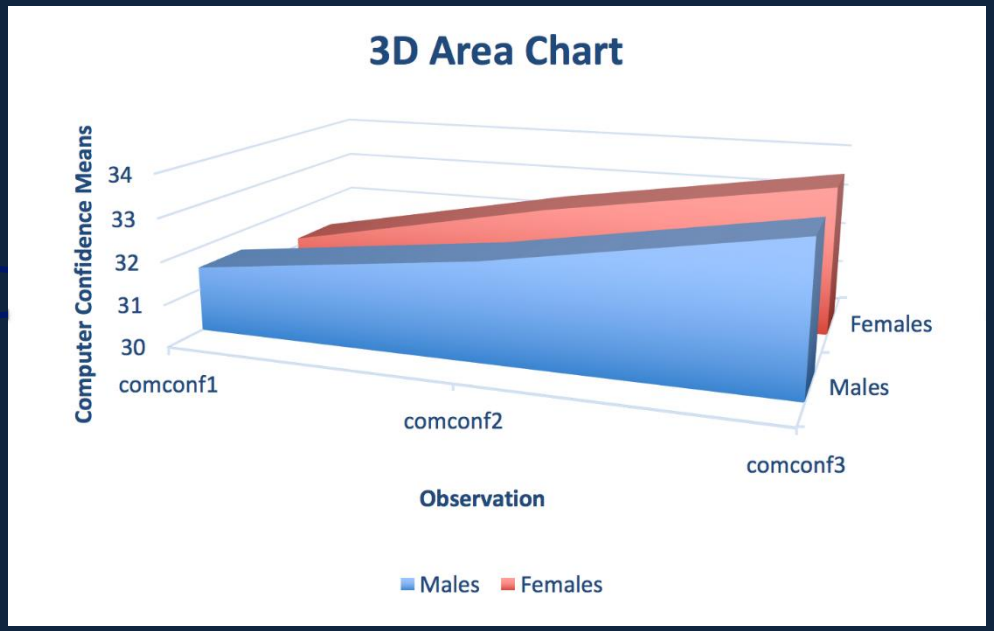
Line & Area Charts	•**Line charts and area charts** are most often used to present longitudinal or time series data, arranged to display change over time, e.g., year 1, year 2, year 3, year 4.
Column & Bar Charts	•**Column and bar charts** contain columns or bars with lengths proportional to the values that they represent. They are used to compare discrete data (i.e., various categories), with each category represented by a single bar or column.
Scatterplots	•**Scatterplots** are used to display the strength and direction of relationship between two continuous variables.
Histograms	•**Histograms** are used to display the frequency distribution of a continuous variable. Histograms are drawn so that the range of the data is split into equal-sized bins and plotted on the x-axis from lowest to highest values. Frequency counts for each bin are plotted on the y-axis. Histograms are frequently used to evaluate normality.
Pie Charts	•**Pie charts** are used to display percentage values as slices of a pie and to illustrate numerical proportions.

Line Charts



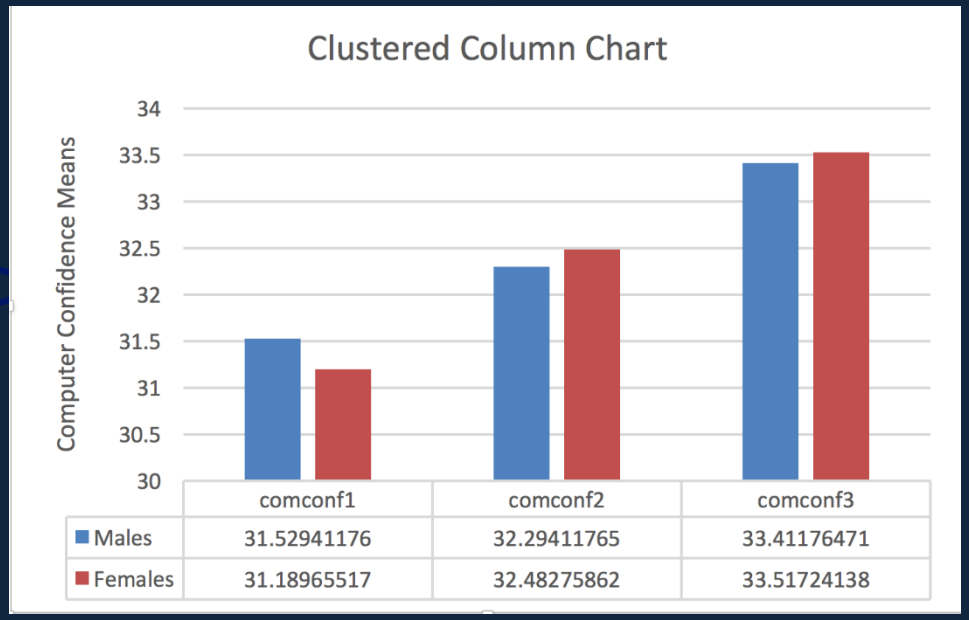
- A line chart is used to display related categorical information as a series of data points called markers connected by straight **line** segments. Line charts are useful in determining trends over time that include multiple observations.
- Line charts can be used to extrapolate beyond known data values (i.e., forecasting).
- The above chart shows how computer confidence means change over three observations (measurements) for male and female students enrolled in an undergraduate computer literacy course. Computer confidence increases for all students during the course.

Area Charts



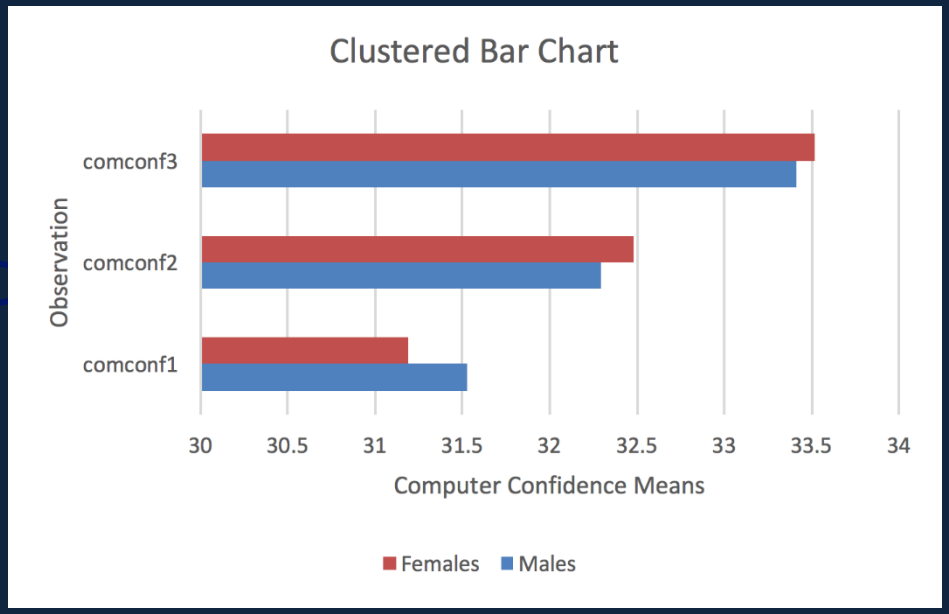
- An area chart, based on the line chart, is used to display related categorical information as a series of data points connected by straight **line** segments.
- Selection of an area chart over a line chart is based on the personal of the author.
- The above chart shows how computer confidence means change over three observations (measurements) for male and female students enrolled in an undergraduate computer literacy course.

Column Charts



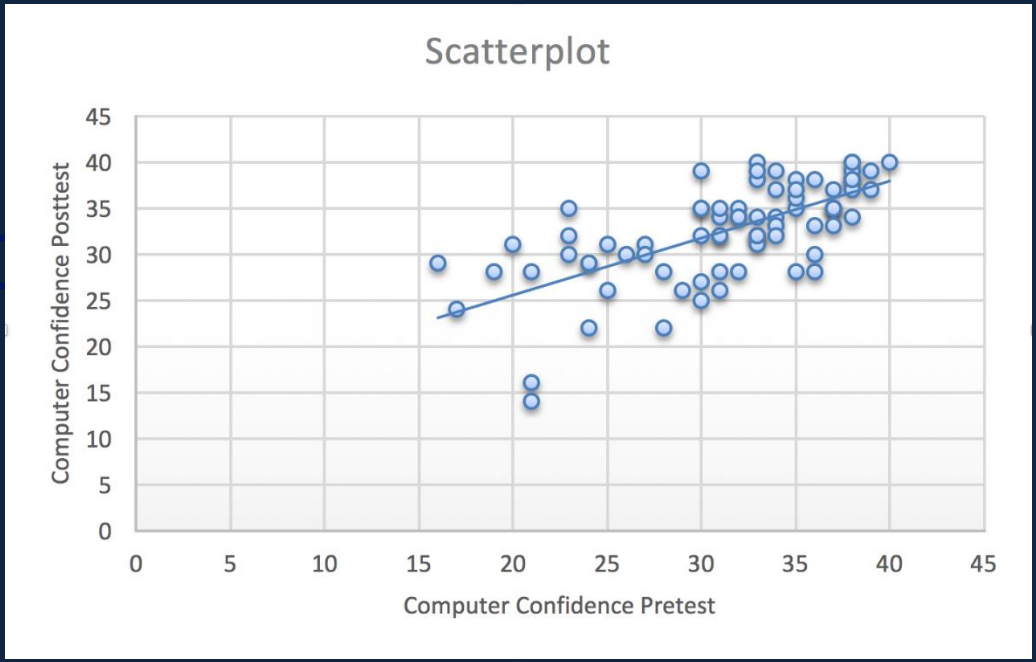
- A column chart is used to compare categories of a categorical variable on some metric using a vertical orientation. The height of the column represents the measurement shown on the y-axis.
- The above chart shows how computer confidence means for male and female students change over three observations (measurements). Included in this chart is an optional data table.

Bar Charts



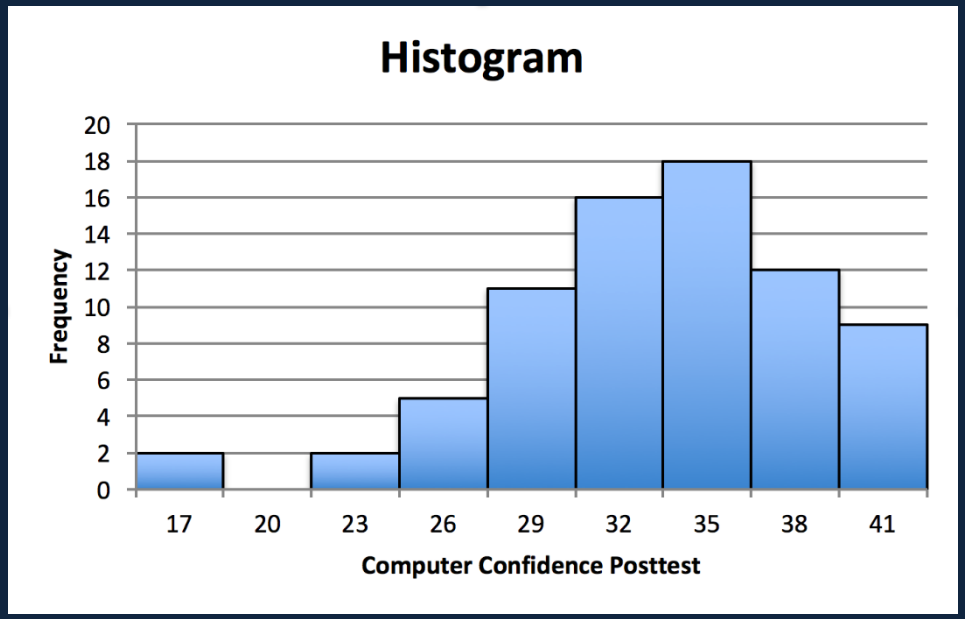
- A bar chart, like a column chart, is used to compare categories of a categorical variable on some metric.
- A bar chart is the horizontal version of a column chart. Use the bar chart to display large text labels.
- The above chart shows how female and male computer confidence scores vary over three observations (measurements).

Scatterplots



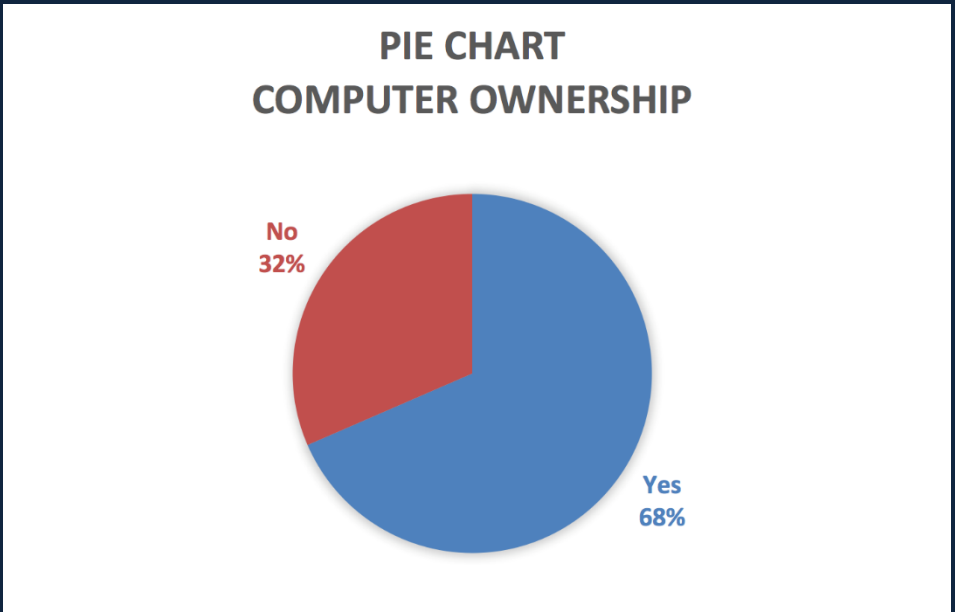
- Scatterplots show the relationship between two continuous variables by graphing a collection of ordered pairs (x,y). Each dot on a scatterplot represents a case. The dot is placed at the intersection of each case's scores on the x and y axes.
- The above scatterplot shows that computer confidence pretest and posttest are related because the dots are clustered together and don't form a random or shotgun pattern.
- Included is an optional trendline that shows the relationship is linear because the major axis of all the dots appears to be a straight line.
- Finally, the relationship is positive – as computer confidence pretest values increase, so do computer confidence posttest values.

Histogram

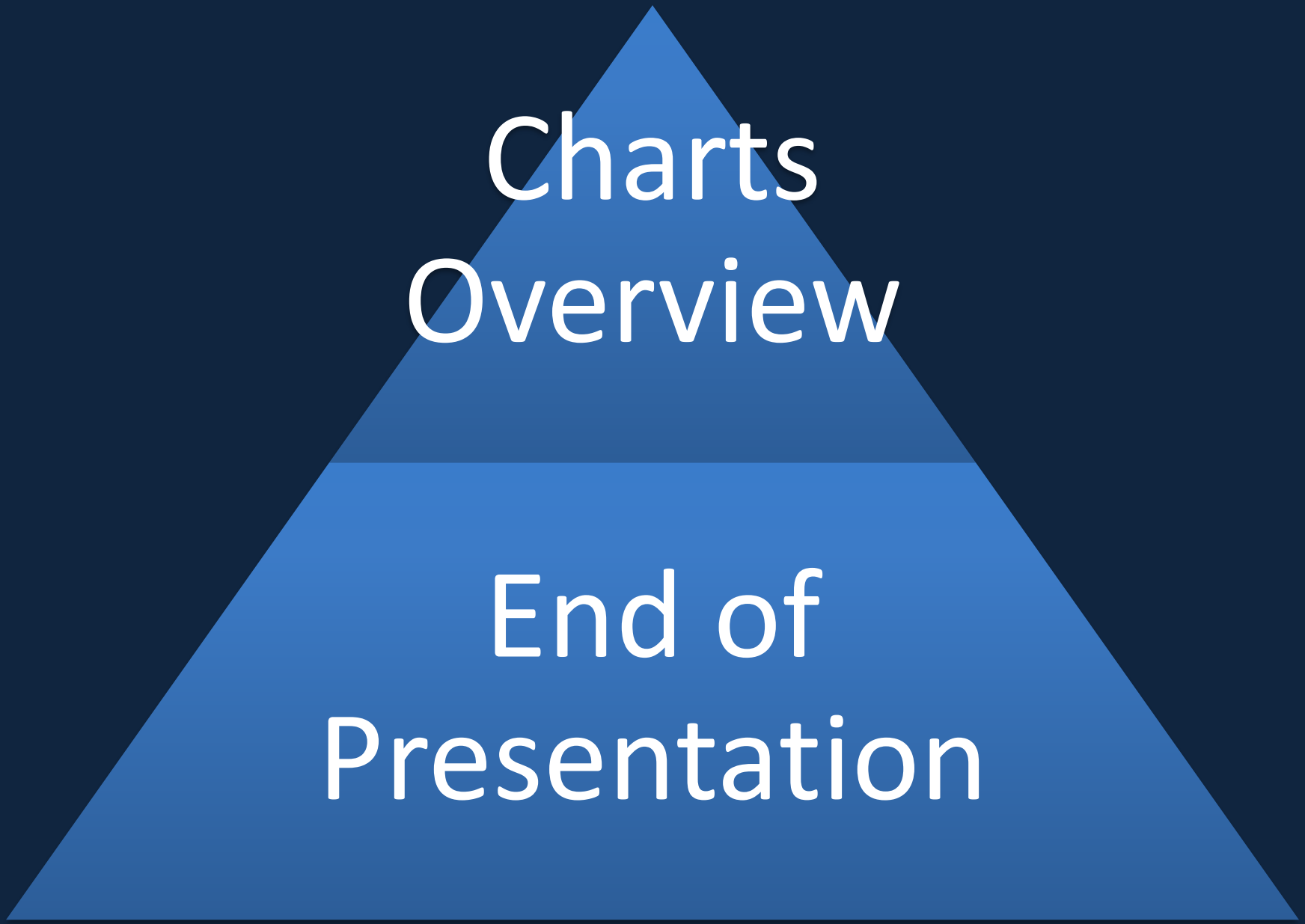


- A histogram is used to evaluate the shape of a distribution of a continuous variable.
- The x-axis depicts the range of the variable from minimum to maximum scores in fixed intervals.
- The y-axis depicts the number of values in each bin (column). For example, there are two scores in the first bin (scores 17 and lower) and there are zero scores in the second bin (scores higher than 17 and no higher than 20).
- Histograms, unlike column charts, have no spaces between bins (columns).
- The above histogram shows that computer confidence posttest scores are negatively skewed because the negative (left) tail is longer (heavier) than the right tail.

Pie Chart



- A pie chart is a circular chart that is divided into sectors or slices to show approximate proportional relationships (i.e., relative size of data) to the whole at a specific point in time.
- Pie charts are useful for comparing proportions and showing how data are distributed. However, a weakness of pie charts is that angles are harder to estimate for people than distances.
- The above pie chart shows that 32% of respondents to a survey responded “no” to owning a computer and 68% responded with a “yes.”


Charts
Overview
End of
Presentation