Introduction to Business Statistics Problem Ancillary Materials: Yukon Edition

Introduction to Business Statistics Problem Ancillary Materials: Yukon Edition

Lisa Kanary

Contents

Chapter 1. Sampling and Data

Chapter 2. Descriptive Statistics

Chapter 3: Probability Topics

Chapter 4: Discrete Random Variables

Chapter 5. Continuous Random Variables

Chapter 6. The Normal Distribution

Chapter 7. The Central Limit Theorem

Chapter 8. Confidence Intervals

Chapter 13. Linear Regression and Correlation

Accessibility Statement

BCcampus Open Education believes that education must be available to everyone. This means supporting the creation of free, open, and accessible educational resources. We are actively committed to increasing the accessibility and usability of the resources we produce.

Accessibility of This Resource

The web version of Introductory Business Statistics Ancillaries (https://opentextbc.ca/busstatsancillary/) has been designed to meet Web Content Accessibility Guidelines 2.0 (https://www.w3.org/TR/WCAG20/), level AA. In addition, it follows all guidelines in Appendix A: Checklist for Accessibility (https://opentextbc.ca/accessibilitytoolkit/back-matter/appendix-checklist-for-accessibility-toolkit/) of the *Accessibility Toolkit – 2nd Edition* (https://opentextbc.ca/accessibilitytoolkit/). It includes:

- **Easy navigation**. This resource has a linked table of contents and uses headings in each chapter to make navigation easy.

- **Accessible math equations**. Many of the equations in this resource have been written in LaTeX and rendered with MathJax, which makes them accessible to people using screen readers that are set up to read MathML. The rest of the equations are rendered as images with appropriate alternative text.

- **Accessible images**. All images in this resource that convey information have alternative text. Images that are decorative have empty alternative text.

- **Accessible links**. All links use descriptive link text.

Accessibility Checklist

Element	Requirements	Pass?
Headings	Content is organized under headings and subheadings that are used sequentially.	Yes
Images	Images that convey information include alternative text descriptions. These descriptions are provided in the alt text field, in the surrounding text, or linked to as a long description.	Yes
Images	Images and text do not rely on colour to convey information.	Yes
Images	Images that are purely decorative or are already described in the surrounding text contain empty alternative text descriptions. (Descriptive text is unnecessary if the image doesn't convey contextual content information.)	Yes
Tables	Tables include row and/or column headers that have the correct scope assigned.	Yes
Tables	Tables include a title or caption.	Yes
Tables	Tables do not have merged or split cells.	Yes
Tables	Tables have adequate cell padding.	Yes
Links	The link text describes the destination of the link.	Yes
Links	Links do not open new windows or tabs. If they do, a textual reference is included in the link text.	Yes
Links	Links to files include the file type in the link text.	Yes
Formulas	Formulas have been created using LaTeX and are rendered with MathJax.	Yes
Formulas	If LaTeX is not an option, formulas are images with alternative text descriptions.	Yes
Font	Font size is 12 point or higher for body text.	Yes
Font	Font size is 9 point for footnotes or endnotes.	Yes
Font	Font size can be zoomed to 200% in the webbook or eBook formats.	Yes

Known Accessibility Issues and Areas for Improvement

There are currently no known accessibility issues.

Let Us Know if You are Having Problems Accessing This Book

We are always looking for ways to make our resources more accessible. If you have problems accessing this resource, please contact us to let us know so we can fix the issue.

Please include the following information:

- The name of the resource

- The location of the problem by providing a web address or page description.

- A description of the problem

- The computer, software, browser, and any assistive technology you are using that can help us diagnose and solve your issue (e.g., Windows 10, Google Chrome (Version 65.0.3325.181), NVDA screen reader)

You can contact us one of the following ways:

- Web form: BCcampus Open Ed Help (https://collection.bccampus.ca/contact/)

- Web form: Report an Error (https://open.bccampus.ca/browse-our-collection/reporting-an-error/)

This statement was last updated on January 20, 2023.

The Accessibility Checklist table was adapted from one originally created by the Rebus Community (https://press.rebus.community/the-rebus-guide-to-publishing-open-textbooks/back-matter/accessibility-assessment/) and shared under a CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).

Acknowledgements

For Students: How to Access and Use this Resource

This resource is available in the following formats:

- **Online webbook**. You can read this book online on a computer or mobile device in one of the following browsers: Chrome, Firefox, Edge, and Safari.

- **PDF**. You can download this book as a PDF to read on a computer (Digital PDF) or print it out (Print PDF).

- **Mobile**. If you want to read this book on your phone or tablet, you can use the EPUB (eReader) file.

- **HTML**. An HTML file can be opened in a browser. It has very little style so it doesn't look very nice, but some people might find it useful.

For more information about the accessibility of this book, see the Accessibility Statement.

You can access the online webbook and download any of the formats for free here: *Introductory Business Statistics Ancillaries (https://opentextbc.ca/busstatsancillary/)*. To download the book in a different format, look for the "Download this book" drop-down menu and select the file type you want.

How can I use the different formats?

Format	Internet required?	Device	Required apps	Accessibility Features	Screen reader compatible
Online webbook	Yes	Computer, tablet, phone	An Internet browser (Chrome, Firefox, Edge, or Safari)	WCAG 2.0 AA compliant, option to enlarge text, and compatible with browser text-to-speech tools	Yes
PDF	No	Computer, print copy	Adobe Reader (for reading on a computer) or a printer	Ability to highlight and annotate the text. If reading on the computer, you can zoom in.	Unsure
EPUB	No	Computer, tablet, phone	An eReader app	Option to enlarge text, change font style, size, and colour.	Unsure
HTML	No	Computer, tablet, phone	An Internet browser (Chrome, Firefox, Edge, or Safari)	WCAG 2.0 AA compliant and compatible with browser text-to-speech tools.	Yes

Tips for Using This Resource

- **Search the book**.

- If using the online webbook, you can use the search bar in the top right corner to search the entire book for a key word or phrase. To search a specific chapter, open that chapter and use your browser's search feature by hitting **[Cntr] + [f]** on your keyboard if using a Windows computer or **[Command] + [f]** if using a Mac computer.

- The **[Cntr] + [f]** and **[Command] + [f]** keys will also allow you to search a PDF, HTML, and EPUB files if you are reading them on a computer.

- If using an eBook app to read this book, the app should have a built-in search tool.

- **Navigate the book**.

  - This book has a table of contents to help you navigate through the book easier. If using the online webbook, you can find the full table of contents on the book's homepage or by selecting "Contents" from the top menu when you are in a chapter.

- **Annotate the book**.

  - If you like to highlight or write on your books, you can do that by getting a print copy, using the Digital PDF in Adobe Reader, or using the highlighting tools in eReader apps.

About BCcampus Open Education

Open educational resources (OER) are teaching, learning, and research resources that, through permissions granted by the copyright holder, allow others to use, distribute, keep, or make changes to them. Our open textbooks are openly licensed using a Creative Commons licence (https://creativecommons.org/licenses/) and are offered in various eBook formats free of charge, or as printed books that are available at cost.

For more information about open education in British Columbia, please visit the BCcampus Open Education (https://open.bccampus.ca/) website. If you are an instructor who is using this book for a course, please fill out our Adoption of an Open Textbook (https://open.bccampus.ca/use-open-textbooks/tell-us-youre-using-an-open-textbook/) form.

Chapter 1. Sampling and Data

1.1 Definitions of Statistics, Probability, and Key Terms

The following activities and questions relate to material covered in Chapter 1.1 Definitions of Statistics, Probability, and Key Terms (https://openstax.org/books/introductory-business-statistics/pages/1-1-definitions-of-statistics-probability-and-key-terms) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 1.1 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/1.1-Data-Sets.xlsx).

Questions

1. Consider the following and choose the correct answer: X = The number of times a student missed a class over zoom. In this instance, X is:

    a. Statistic

    b. Population

    c. Variable

    d. Sample

2. Mount Sima Ski Resort is interested in the average age that children take their first ski lessons. They need this information to plan their ski classes. Identify the

    a. population

    b. sample

    c. parameter

    d. statistic

    e. variable, and

    f. an example of variable data.

3. The Council of Yukon First Nations (CYFN) is interested to know about the mean age of First Nation people who are living in Yukon, Canada. Identify the

    a. population

    b. sample

    c. parameter

    d. statistic

   e. variable, and

   f. an example of variable data.

4. A Yukon statistician is interested in the proportion of eligible Yukon dogsled teams that compete in the Yukon Quest. Identify the

   a. population

   b. sample

   c. parameter

   d. statistic

   e. variable, and

   f. an example of variable data.

5. The city of Whitehorse wants to know how much time it takes for people to travel downtown from their place of residence. The following chart displays the surrounding areas and time taken to travel downtown in minutes according to Google Maps 2021. Find the average time and round to one and two decimal places.

**Time taken to travel downtown Whitehorse
from surrounding areas**

Place Name	Time (minutes)
Whistle Bend	9
Riverdale	3
Porter Creek	11
Takhini	5
Copper Ridge	10
Mt. Sima	10
Wolf Creek	14
Grey Mountain	9

Whitehorse cliffs

Solutions

1. c. Variable

2.
   a. The population is all people who ski at Mount Sima Ski.

   b. The sample is a selected group of people who ski at Mount Sima Ski.

   c. The parameter is the mean age of when all people who ski at Mount Sima Ski took their first lesson.

   d. The statistic is the mean age of when the sample of people who ski at Mount Sima Ski took their first lesson.

   e. The variable is the age of when someone skiing at Mount Sima Ski took their first lesson.

   f. Example: data values can be 1 year, 10.5 yrs.

3.
   a. The population is all the First Nation People living in Yukon Territory.

   b. The sample is a selected group of First Nation people living in the Yukon.

   c. The parameter is the mean age of all First Nation people living in the Yukon.

   d. The statistic is the mean age of the sample of First Nation people living in Yukon.

   e. The variable is the age of a First Nation person living in Yukon.

   f. Example: data values can be 4 months, 25 yrs.

4.
   a. The population is the total number of dogsled teams in Yukon eligible for competition.

b. The sample is a selected group of dogsled teams in Yukon eligible for competition.

c. The parameter is the proportion of all eligible Yukon dogsled teams that compete in the Yukon Quest.

d. The statistic is the proportion of the sample of eligible Yukon dogsled teams that compete in the Yukon Quest.

e. The variable is whether an eligible dogsled team competed in the Yukon Quest.

f. Example: data will be either: (1) Yes, the team did compete, or (2) No, the team did not compete.

5. Sum of all travel time = 9 + 3 + 11 + 5 + 10 + 10 + 14 + 9 = 71
Number of places = 8
Mean travel time = $\dfrac{\text{Sum of all travel time}}{\text{Number of places}} = \dfrac{71}{8}$ = 8.875 minutes
Rounded to one decimal place = 8.9 minutes
Rounded to two decimal places = 8.88 minutes

Media Attributions

- Whitehorse cliffs by Lisa Kanary is licensed under a CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).

1.2 Data, Sampling, and Variation in Data and Sampling

The following activities and questions relate to material covered in Chapter 1.2 Data, Sampling, and Variation in Data and Sampling (https://openstax.org/books/introductory-business-statistics/pages/1-2-data-sampling-and-variation-in-data-and-sampling) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Questions

Name the sampling method used in each of the following situations:

a. At the Beaver Creek Public Library, a librarian wants to determine what proportion of the library users are children. The librarian has a tally sheet on which she marks whether books are checked out by an adult or a child. She records this data for every fourth patron who checks out books.

b. A Yukon University instructor wants to know if her students are doing the required take-home exercises, so she randomly selects rows 1 and 3 in the (virtual) classroom and calls on all students in those rows to present the solutions to the take-home exercises to the rest of the class.

c. An Air North employee is at the airport handing out questionnaires to travelers asking them to evaluate their experience with the Yukon airline company. They do not want to disturb anyone looking like they are in a rush, so focus on asking travelers who are calmly sitting at their gates or look to not be too busy.

d. The regional manager of Petro-Canada wants information about the ages of the customers who buy gas from them. Over the next two weeks, at each station location all over the Yukon, 50 randomly selected customers are given questionnaires to fill out asking for information about their age.

e. Yukon Health & Social Services wants to know what people's reactions to the Covid-19 vaccination roll-out information package are. The day after the schedule was released, the department team surveys 1,200 randomly selected Yukoners.

Solutions

a. Systematic sampling
b. Cluster sampling
c. Convenience sampling
d. Stratified sampling
e. Simple Random sampling

1.3 Levels of Measurements

The following activities and questions relate to material covered in Chapter 1.3 Levels of Measurements (https://openstax.org/books/introductory-business-statistics/pages/1-3-levels-of-measurement) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel:1.3 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/1.3-Data-Sets.xlsx).

Questions

1. Between 1953 and 2020, 38 hurricanes have made landfall on Canadian soil. The hurricanes are often downgraded by the time they reach Canada; however, at their strongest force, they receive a category rating to determine the maximum size and intensity of the storm of either Tropical, 1, 2, 3, 4, or 5. Five represents the largest and most ferocious storms, with the title of Tropical given to downgraded storms that more commonly impact Canada.[1] The table below depicts hurricanes at their peak strength before impacting Canada. Use the given data to create a table with columns for frequency, relative frequency, and cumulative frequency for each category of hurricane.

1. Data source: Wikipedia. (n.d.) *List of hurricanes in Canada*. https://en.wikipedia.org/wiki/List_of_Canada_hurricanes

List of hurricanes in Canada 1950-1999

Year	Name	Category
1953	Carol	5
1954	Hazel	4
1955	Connie	4
1958	Helene	4
1959	Cindy	1
1959	Escuminac	1
1962	Daisy	2
1963	Ginny	2
1964	Gladys	4
1971	Beth	1
1973	Alice	1
1975	Blanche	1
1989	Dean	2
1990	Bertha	1
1991	1991 "Perfect Storm"	1
1995	Luis	4
1996	Bertha	3
1996	Huron	Tropical
1996	Hortense	4
1999	Floyd	4

List of hurricanes in Canada 2000-2020

Year	Name	Category
2000	Michael	2
2001	Karen	1
2002	Gustav	2
2003	Juan	2
2005	Ophelia	1
2007	Noel	1
2008	Kyle	1
2009	Bill	4
2010	Earl	4
2010	Igor	4
2011	Irene	3
2011	Maria	1
2011	Ophelia	4
2012	Leslie	1
2012	Sandy	3
2014	Arthur	2
2016	Matthew	5
2019	Dorian	5
2020	Isaias	1
2020	Teddy	4

Solutions

1.

Category	Frequency of Hurricanes	Relative Frequency	Cumulative Frequency
1	14	0.350	0.350
2	7	0.175	0.525
3	3	0.075	0.600
4	12	0.300	0.900
5	3	0.075	0.975
Tropical	1	0.025	1

Sampling and Data - Excel Tools Instruction

Download Sampling and Data – Excel Instruction Data Sets [Excel] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2022/10/Sampling-and-Data-–-Excel-Instruction-Data-Sets.xlsx).
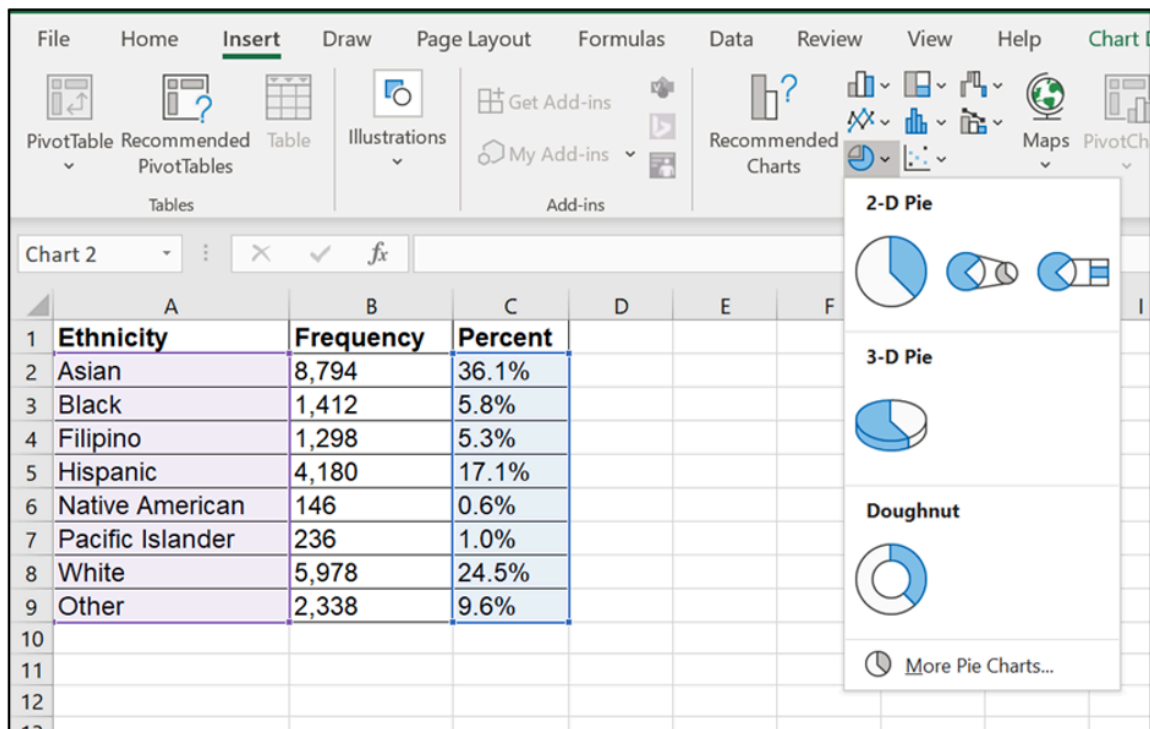
Pie Charts and Bar Graphs

- In a **pie chart**, categories of data are represented by wedges in a circle and are proportional in size to the percent of individuals in each category.
- In a **bar graph**, the length of the bar for each category is proportional to the number or percent of individuals in each category. Bars may be vertical or horizontal.

Pie Charts in Excel

1. Highlight columns of cells (hold 'Ctrl' button if columns not adjacent)
2. Click Insert, Charts and the first 2-D pie
3. Click on the new pie chart and icon + to make changes
4. Right click on wedge, choose 'Add Data Labels'

Table 1.4 Ethnicity of Students at De Anza College Fall Term 2007 (Census Day)

Ethnicity	Frequency	Percent
Asian	8,794	36.1%
Black	1,412	5.8%
Filipino	1,298	5.3%
Hispanic	4,180	17.1%
Native American	146	0.6%
Pacific Islander	236	1.0%
White	5,978	24.5%
Other	2,338	9.6%

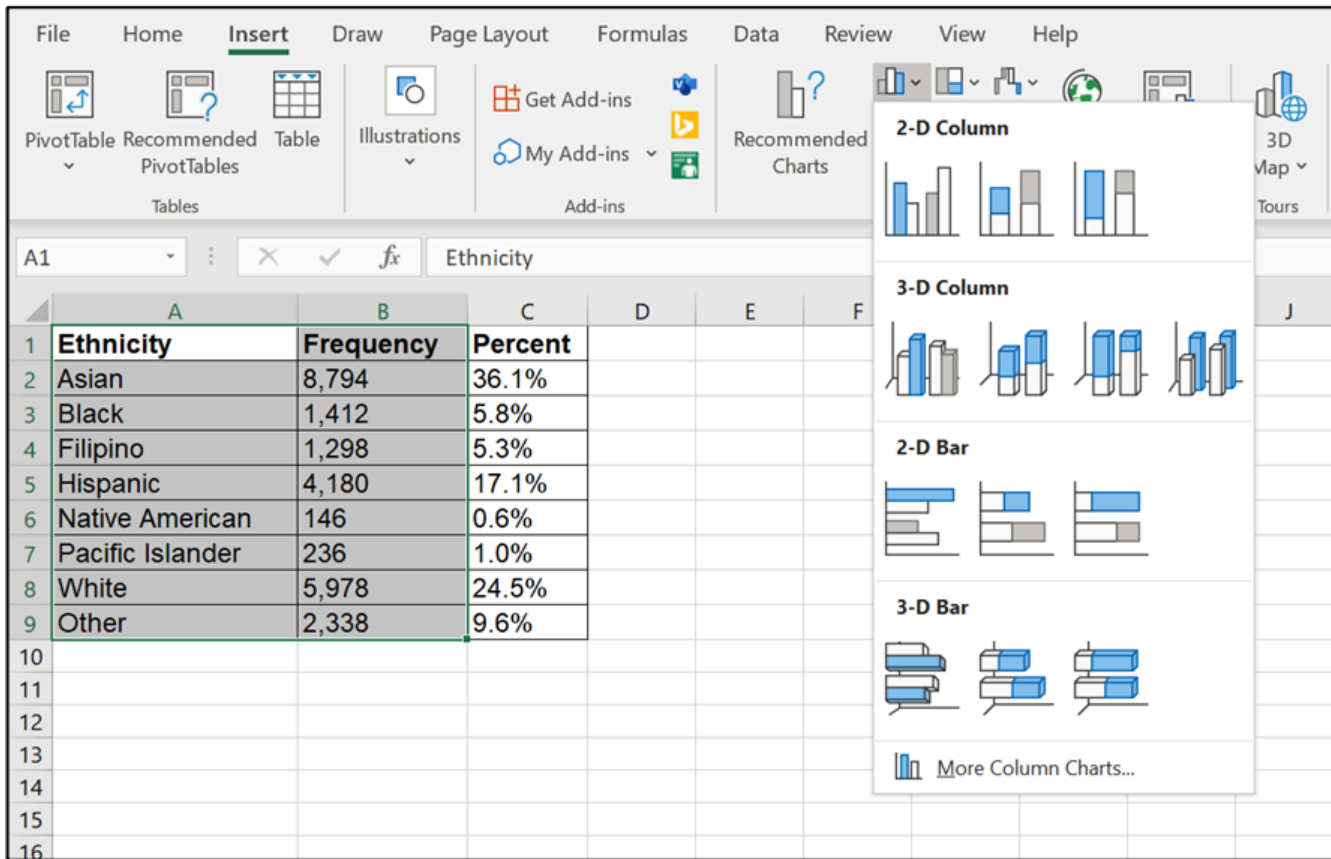*Step 1 & 2: Highlight column of cells and create 2-D Pie*
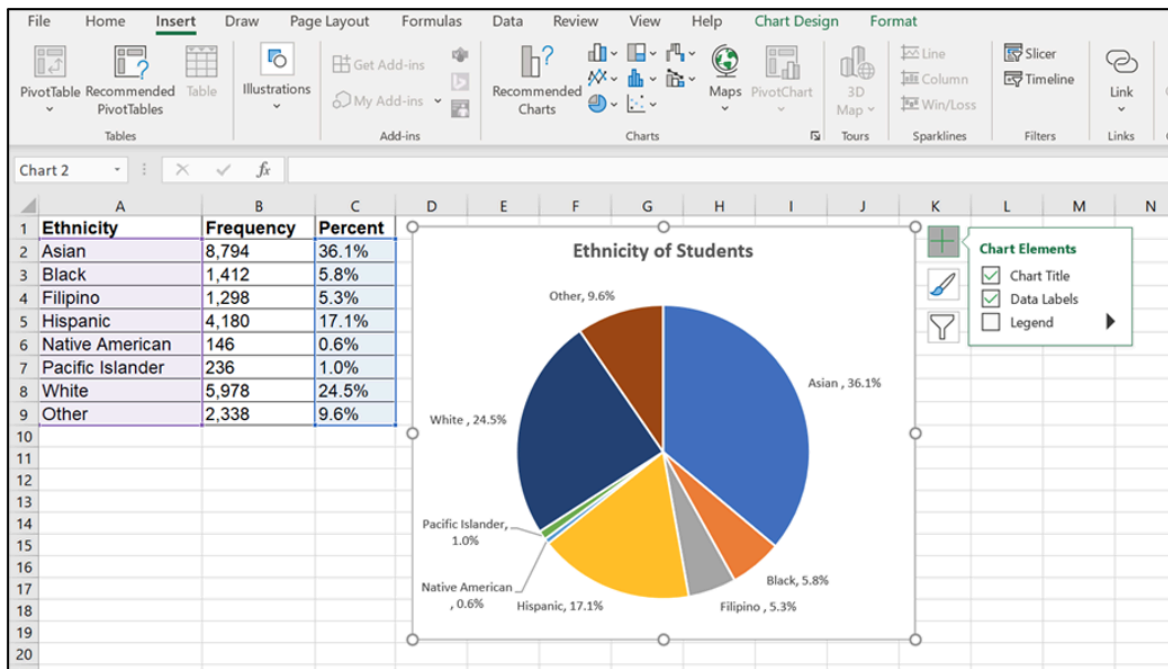
Step 3 & 4: Click icon + and add Data Labels

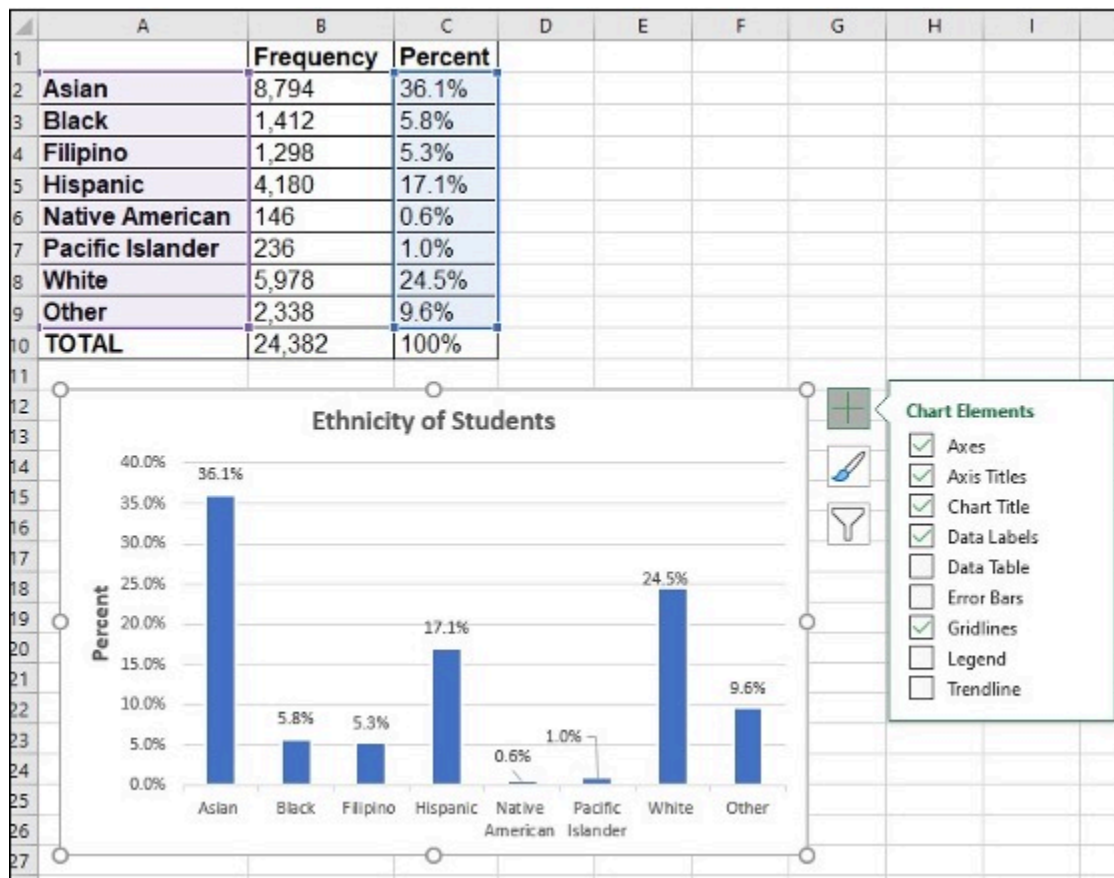## Bar graphs in Excel

1. Highlight columns of cells

2. Click Insert, Charts and the first 2-D Column

3. Click on the new bar graph and icon + to make changes

4. Right click on wedge, choose 'Add Data Labels'

Step 1 & 2: Highlight column of cells and insert 2-D Chart

	A	B	C	D	E	F	G	H	I
1		Frequency	Percent						
2	Asian	8,794	36.1%						
3	Black	1,412	5.8%						
4	Filipino	1,298	5.3%						
5	Hispanic	4,180	17.1%						
6	Native American	146	0.6%						
7	Pacific Islander	236	1.0%						
8	White	5,978	24.5%						
9	Other	2,338	9.6%						
10	TOTAL	24,382	100%						

Step 3 & 4: Click on New Bar Graph + icon to make changes and Add Data Labels

Frequency, Relative Frequency and Cumulative Relative Frequency

- A **frequency** is the number of times a value of the data occurs.
- A **relative frequency** is the ratio (fraction or proportion) of the number of times a value of the data occurs in the set of all outcomes to the total number of outcomes. Relative frequencies can be written as fractions, percents, or decimals.
- A **cumulative relative frequency** is the accumulation of the previous relative frequencies.

Frequency in Excel

Twenty students were asked how many hours they worked per day. Their responses, in hours, are as follows: 5; 6; 3; 3; 2; 4; 7; 5; 2; 3; 5; 6; 5; 4; 4; 3; 5; 2; 5; 3.

Enter all data in column A. Create a column for data categories (column B).

1. Enter =COUNTIF in cell C2, click column A, comma and B2
2. Click Enter
3. Click C2, move the mouse to the right bottom corner, a little + appears

4. Click on the little + and drag down to repeat cell

Step 1 & 2: Enter =COUNTIF, select column A, column B, then click Enter

Step 3 & 4: Click C2 and use + in bottom right corner to drag down and repeat cell

Relative Frequency in Excel

Twenty students were asked how many hours they worked per day. Their responses, in hours, are as follows: 5; 6; 3; 3; 2; 4; 7; 5; 2; 3; 5; 6; 5; 4; 4; 3; 5; 2; 5; 3.

1. In cell D2, enter =, click cell C2, enter /, and click C8 F4 (F4 holds a cell constant)

2. Click Enter, 0.15 shows in D2

3. Click D2, move the mouse to the right bottom corner, a little + shows up

4. Click on the little + and drag down to repeat cell

	File	Home	Insert	Page Layout	Formulas	Data	Review

C8 · ⋮ × ✓ ƒ× =C2/$C$8

	A	B	C	D	E
1	Data	Data Value	Frequency	Relative Frequency	Cumulative Relative Frequency
2	5	2	3	=C2/$C$8	
3	6	3	5		
4	3	4	3		
5	3	5	6		
6	2	6	2		
7	4	7	1		
8	7	Total	20		
9	5				

Step 1: In cell D2, enter =, click cell C2, enter /, and click C8 F4

	A	B	C	D	E
1	Data	Data Value	Frequency	Relative Frequency	Cumulative Relative Frequency
2	5	2	3	0.15	
3	6	3	5	0.25	
4	3	4	3	0.15	
5	3	5	6	0.3	
6	2	6	2	0.1	
7	4	7	1	0.05	
8	7	Total	20	1	
9	5				

Step 2, 3 & 4: Click Enter. Then click D2, click and drag + in bottom right corner to repeat cells

Cumulative Relative Frequency in Excel

Twenty students were asked how many hours they worked per day. Their responses, in hours, are as follows: 5; 6; 3; 3; 2; 4; 7; 5; 2; 3; 5; 6; 5; 4; 4; 3; 5; 2; 5; 3.

1. In cell E2, enter =, click D2 and Enter
2. In cell E3, enter =SUM($D$2:D3)
3. Click Enter, 0.4 shows in E3
4. Click E3, move the mouse to the right bottom corner, a little + shows up
5. Click on the little + and drag down to repeat cell

E3		⋮	✕ ✓	fx	=SUM($D$2:D3)

	A	B	C	D	E
1	Data	Data Value	Frequency	Relative Frequency	Cumulative Relative Frequency
2	5	2	3	0.15	0.15
3	6	3	5	0.25	=SUM($D$2:D3)
4	3	4	3	0.15	
5	3	5	6	0.3	
6	2	6	2	0.1	
7	4	7	1	0.05	
8	7	Total	20	1	
9	5				

Step 1 & 2: In cell E2, enter =, click D2 and Enter, then in cell E3, enter =SUM($D$2:D3)

	A	B	C	D	E
1	Data	Data Value	Frequency	Relative Frequency	Cumulative Relative Frequency
2	5	2	3	0.15	0.15
3	6	3	5	0.25	0.4
4	3	4	3	0.15	0.55
5	3	5	6	0.3	0.85
6	2	6	2	0.1	0.95
7	4	7	1	0.05	1
8	7	Total	20	1	
9	5				

Step 3, 4 & 5: Click Enter, click E3, then click and drag + in bottom right corner to repeat cells

Pivot Table

- A Pivot Table helps to arrange and summarize complex data.

Pivot Table for Frequency in Excel

Twenty students were asked how many hours they worked per day. Their responses, in hours, are as follows: 5; 6; 3; 3; 2; 4; 7; 5; 2; 3; 5; 6; 5; 4; 4; 3; 5; 2; 5; 3.

1. Enter all data in column A
2. Click Insert, Pivot Table
3. Click on the area of Table/Range, highlight all data
4. Click on the area of Location, click on cell C1 on where the pivot table will display
5. Click OK

Step 1 & 2: Enter all data in column A then Click Insert,
Pivot Table

Step 3, 4 & 5: Click on the area of Table/Range, highlight all data then click
Location, click on cell C1 where pivot table will display, click OK

6. Click on Working hours, hold and drag down to Rows area

7. Click on Working hours, hold and drag down to Values area

8. Click on the drop-down icon, Value Field Setting

9. Choose Count Numbers, click OK
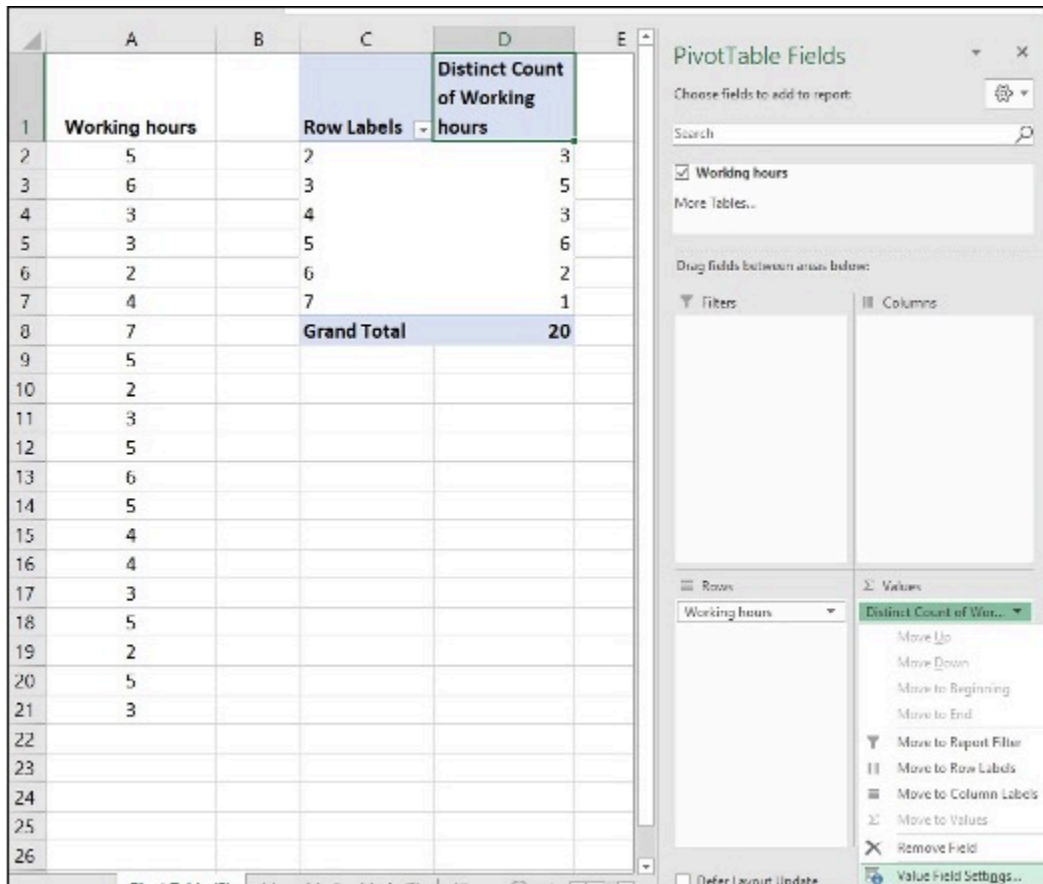
Step 6, 7 & 8: Click on Working hours, hold and drag down to Rows area, repeat the same for Values area then click on the drop-down icon, Value Field Setting
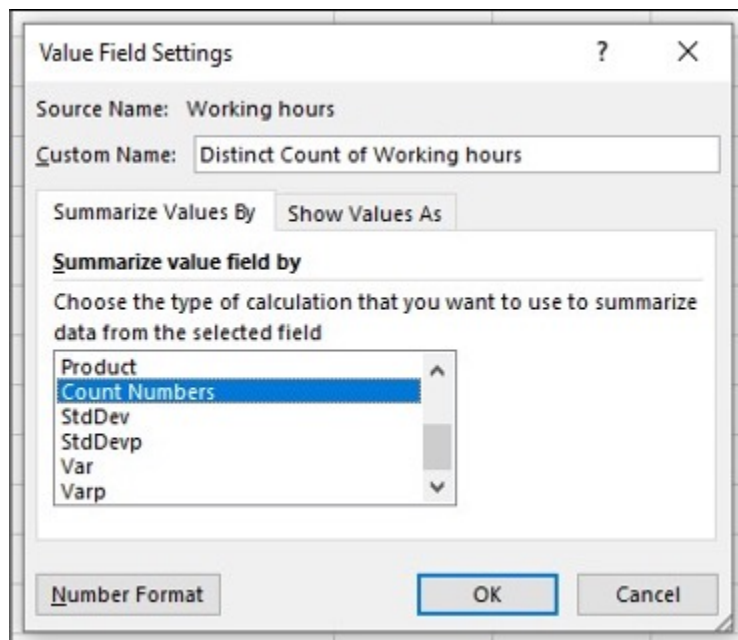
Step 9: Choose Count Numbers, click OK

Media Attributions

- Screenshots of Excel are used with permission from Microsoft (https://www.microsoft.com/en-us/legal/intellectualproperty/copyright/permissions).

Chapter 2. Descriptive Statistics

2.1 Display Data

The following activities and questions relate to material covered in Chapter 2.1 Display Data (https://openstax.org/books/introductory-business-statistics/pages/2-1-display-data) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 2.1 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2019/07/2.1-Data-Sets.xlsx).

Questions

1. The annual Canadian CO_2 emissions from 2003 to 2009 are as follows.

Year	CO_2 Emissions (kt)
2003	553,101
2004	553,358
2005	561,425
2006	571,993
2007	571,846
2008	561,781
2009	532,192

   a. Organize the data into a line graph.
   b. Calculate the sample mean.

2. The table below is the number of police officers and crime rate (each per 100,000 citizens) in Canada during the period from 2009 to 2019.

Number of police officers and crime rate (each per 100,000 citizens) [1]

Year	Number of Police Officers	Crime Rate
2009	200	6461
2010	203	6159
2011	202	5780
2012	200	5638
2013	197	5206
2014	194	5061
2015	193	5232
2016	191	5297
2017	189	5375
2018	185	5514
2019	183	5874

    a. Construct a double time series graph using a common x-axis for both sets of data and different y-axes.

    b. How did variables change for both data sets? Explain.

3. Construct a histogram of the ages of Prime Ministers at the time of their being sworn into the position by the Governor General using the table below:

**Ages of Prime Ministers at the time of being sworn into the position** [2]

Age of Canadian Prime Minister (yrs.)	Frequency
38.5 – 44.5	2
44.5 – 40.5	7
50.5 – 56.5	4
56.5 – 62.5	4
62.5 – 68.5	3
68.5 – 74.5	2
74.5 – 80.5	1

1. Data source: Statistics Canada. (2020). *Table 3 Trends in police personnel, Canada, 1962 to 2019.* https://www150.statcan.gc.ca/n1/pub/85-002-x/2020001/article/00015/tbl/tbl03-eng.htm. Statistics Canada. (2020). *Table 7 Police-reported crime rate, Canada, 2009 to 2019.* https://www150.statcan.gc.ca/n1/pub/85-002-x/2020001/article/00010/tbl/tbl07-eng.htm
2. Data source: Wikipedia. (n.d.). *List of prime ministers of Canada.* https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Canada

4. Assume a sample of 66 people in northern Canada were asked how often they go on vacations a year.

    a. Create a table with columns for frequency, relative frequency, and cumulative frequency for each number category of vacations taken (i.e., 0, 1, 2…).

    b. Construct a bar chart/histogram of the data.

    c. Describe two features of the bar chart.

Solutions

1.   a. Line graph of Canadian CO_2 Emissions from $2003 - 2009$:

Canadian CO2 Emissions from 2003 - 2009
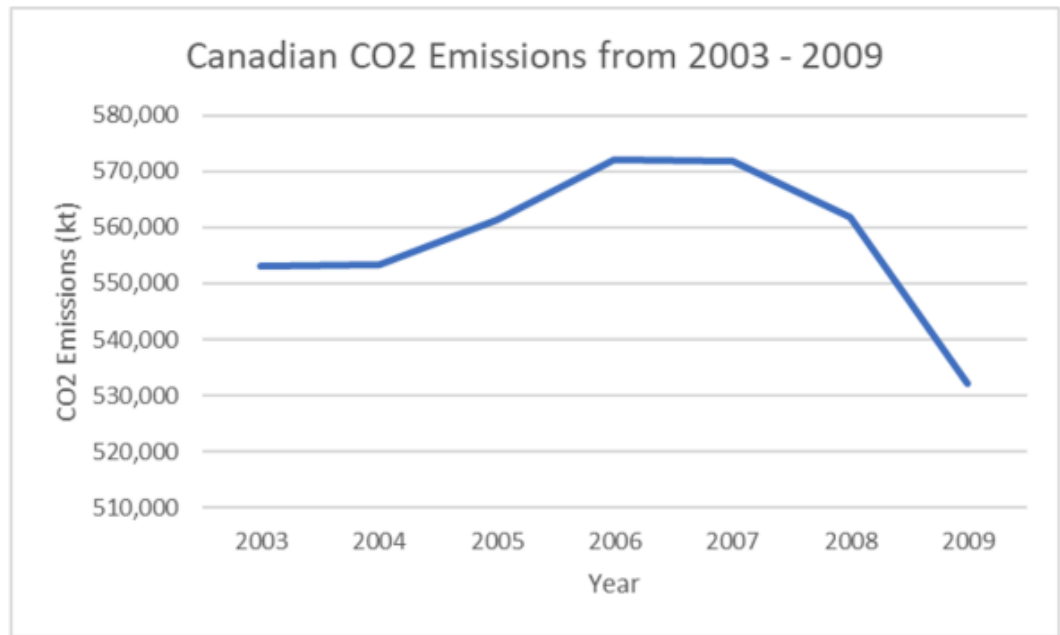
    b. $\text{Sample mean} = \dfrac{553101 + 553358 + 561425 + 571993 + 571846 + 561781 + 532192}{7}$
$$= 557,956 \text{ kt}$$

2.   a. Double time series graph of number of police officers and crime rate in Canada:
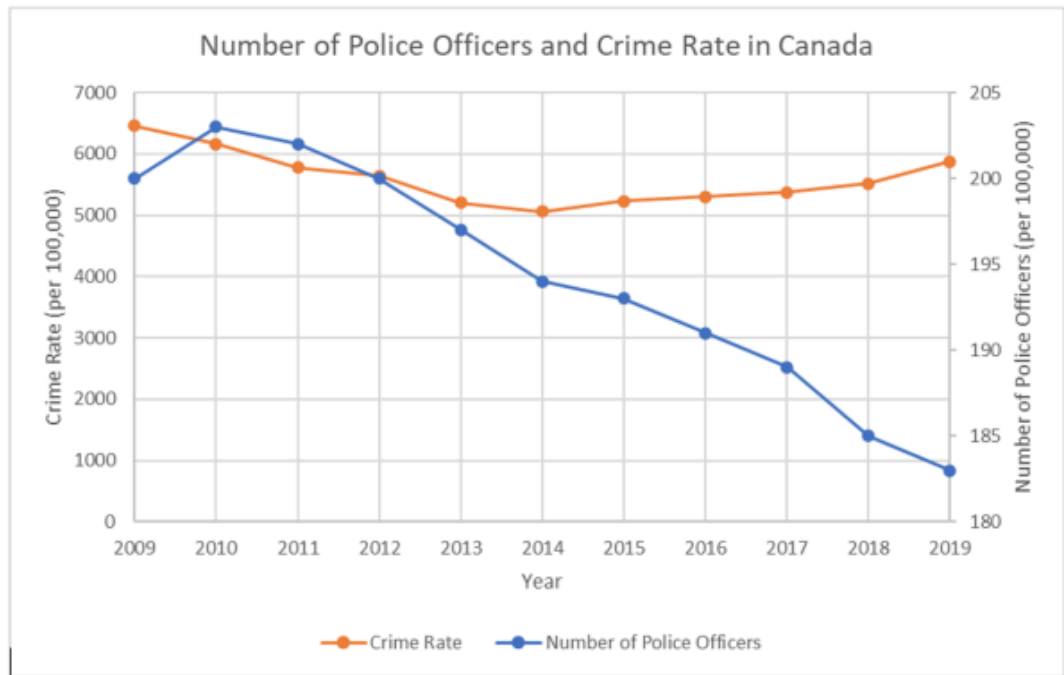
b. Crime rate decreased between 2009 and 2014, then increased until 2019. The number of police officers steadily decreased between 2010 and 2019.

3. Histogram of age of Canadian Prime Minister at the beginning of their first term:
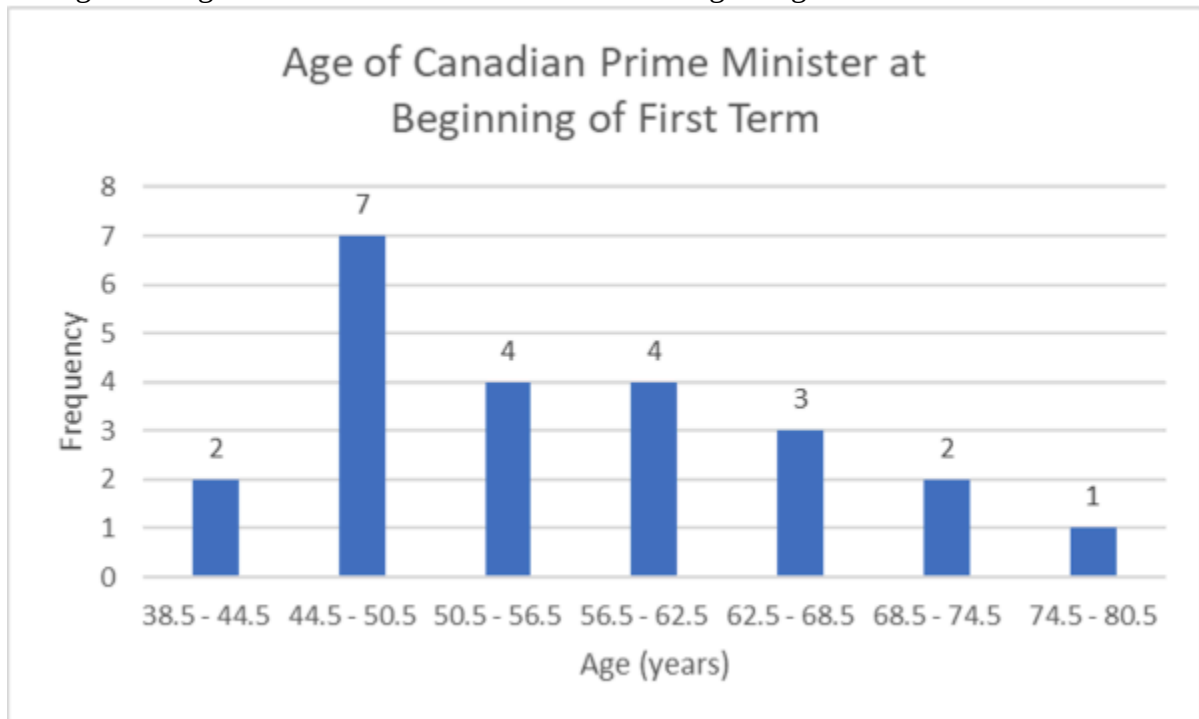
4. a.

Number of Annual Vacations	Frequency	Relative Frequency
0	2	3.03%
1	6	9.09%
2	11	16.67%
3	27	40.91%
4	12	18.18%
5	8	12.12%

b. Histogram of Yukoners' Annual Vacations:
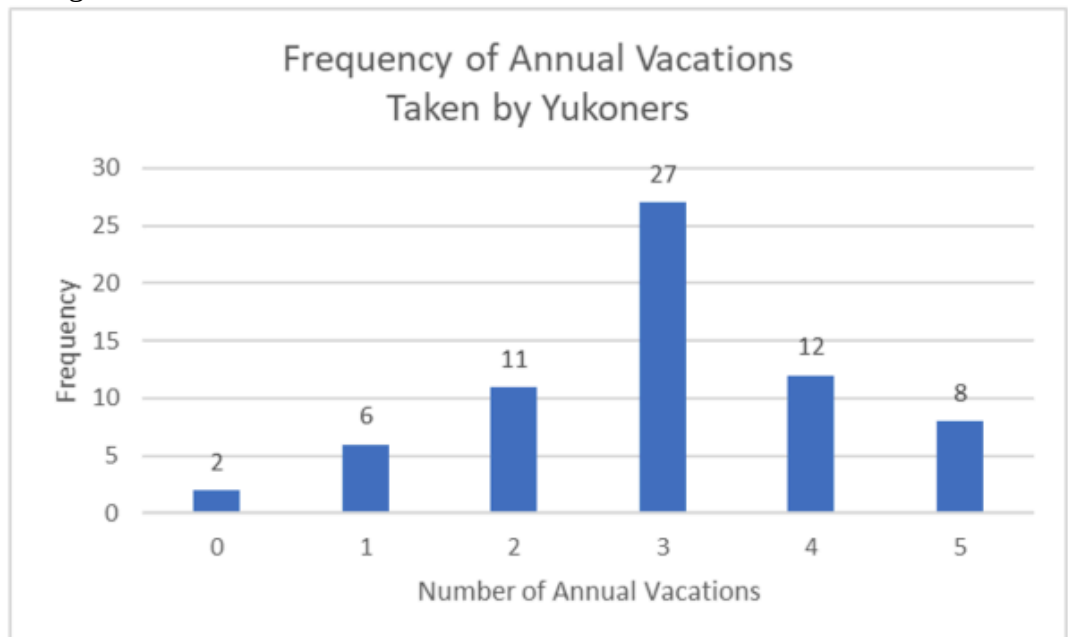
c. Yukoners are most likely to take 3 annual vacations. The data is negatively skewed.

2.2 Measures of the Location of the Data

The following activities and questions relate to material covered in Chapter 2.2 Measures of the Location of the Data (https://openstax.org/books/introductory-business-statistics/pages/2-2-measures-of-the-location-of-the-data) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 2.2 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/2.2-Data-Sets.xlsx).

Questions

1. This is a table of monthly ticket sales at the Yukon Beringia Interpretive Centre in 2018.

Monthly Ticket Sales at Yukon Beringia Interpretive Centre [1]

Month	Beringia 2018 Ticket Sales	Relative Frequency	Cumulative Frequency
January	673		
February	814		
March	1670		
April	1303		
May	1347		
June	3519		
July	4745		
August	4065		
September	1808		
October	1170		
November	1475		
December	1309		

1. Data source: Government of Yukon. (2018). *Tourism Yukon 2018 Year-End Report*. https://yukon.ca/sites/yukon.ca/files/tc/tc-tourism-yukon-2018-year-end-report.pdf

     a. Complete the table.

     b. What proportion of 2018 ticket sales were sold in July and August?

     c. What proportion of ticket sales in the first 8 months of 2018?

2. Use the 2020 Yukon population by age group table below to:

Age Group (yrs.)	Frequency	Relative Frequency
0-9	4692	
10-19	4355	
20-29	5435	
30-39	7266	
40-49	5938	
50-59	5913	
60-69	5459	
70-79	2574	
80+	875	

     a. Create a bar chart.

     b. Complete the table.

     c. Which age category has the highest frequency and what proportion of the total population is that category?

     d. Use 2 sentences to describe features of the bar chart

3. Forty randomly selected people in Yellowknife, NWT were asked the number of nights they saw northern lights last year. The results are as follows:

Number of Nights Northern Lights Viewed	Frequency	Relative Frequency	Cumulative Frequency
0	0		
1	15		
2	10		
3	2		
4	2		
5	2		

     a. Complete the columns of the chart.

    b. Use a bar chart to graph the frequencies.

    c. What is the probability someone chosen randomly viewed at least one night of northern lights?

4. In a survey collecting data about the salaries earned by recent Yukon University graduates, Christy found that her salary was in the 85th percentile. Should Christy be pleased by this result? Explain.

5. Thirty part-time grocery clerks who work at grocery store in northern Canada were asked, "What is your average weekly salary?" The results are in the table below.

Weekly Salary ($)	Frequency
No response	7
200-250	5
350-300	6
300-350	6
350-400	3
400-450	1
450-500	2

    a. What percentage of the survey did not give an answer?

    b. What percentage says their weekly salary is from $200.00 to $350.00?

    c. Construct a bar graph of the data.

6. The annual precipitation in Atlin, BC between 2005 and 2014 is as follows:

Annual Precipitation in Altlin, BC [2]

Year	Annual Atlin Precip. (mm)
2005	307.3
2006	24.1
2007	50.0
2008	143.8
2009	66.4
2010	54.2
2011	85.9
2012	259.7
2013	330.8
2014	279.4

a. Organize the data into a line graph.

b. Construct a histogram of the data.

c. Calculate the sample mean.

d. Find the interquartile range.

2. Data source: Weather Dashboard for Atlin. (n.d.). *Total precipitation - Annual data for Atlin*. https://atlin.weatherstats.ca/charts/precipitation-yearly.html

Solutions

1.   a.

Monthly Ticket Sales at Yukon Beringia Interpretive Centre

Month	Beringia 2018 Ticket Sales	Relative Frequency	Cumulative Frequency
January	673	2.82%	2.82%
February	814	3.41%	6.22%
March	1670	6.99%	13.21%
April	1303	5.45%	18.66%
May	1347	5.64%	24.30%
June	3519	14.73%	39.02%
July	4745	19.86%	58.88%
August	4065	17.01%	75.89%
September	1808	7.57%	83.45%
October	1170	4.90%	88.35%
November	1475	6.17%	94.52%
December	1309	5.48%	100%

b. In July and August of 2018, (19.86%+17.01% =) 36.87% of 2018's tickets were sold.

c. In 2018, 75.89% of 2018's tickets were sold in the first 8 months of the year.

2.   a. Bar chart of 2020 Yukon population by age group

2020 Yukon Population by Age Group
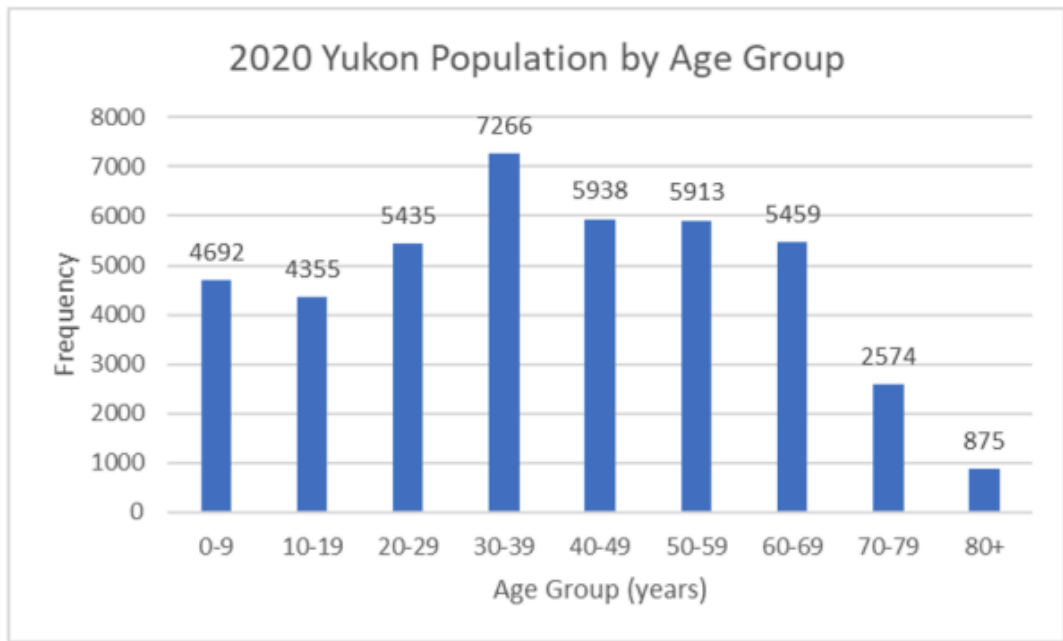
b.

| Age Group (yrs.)	Frequency	Relative Frequency
0-9	4692	11.04%
10-19	4355	10.25%
20-29	5435	12.79%
30-39	7266	17.09%
40-49	5938	13.97%
50-59	5913	13.91%
60-69	5459	12.84%
70-79	2574	6.06%
80+	875	2.06%

c. The age category with the highest frequency is Yukoners between 30-39 years of age occupying 17.09% of the total Yukon population.

d. There is a lower probability of choosing a random Yukoner between 10 and 19 years of age than any other age group. There are almost as many Yukoners between the ages of 0 and 39 as there are Yukoners older than 40 yrs. old.

3.   a.

| Number of Nights Northern Lights Viewed	Frequency	Relative Frequency	Cumulative Frequency
0	0	22.50%	22.50%
1	15	37.50%	60.00%
2	10	25.00%	85.00%
3	2	5.00%	90.00%
4	2	5.00%	95.00%
5	2	5.00%	100.00%

b. Bar chart of the frequency of viewing northern lights.
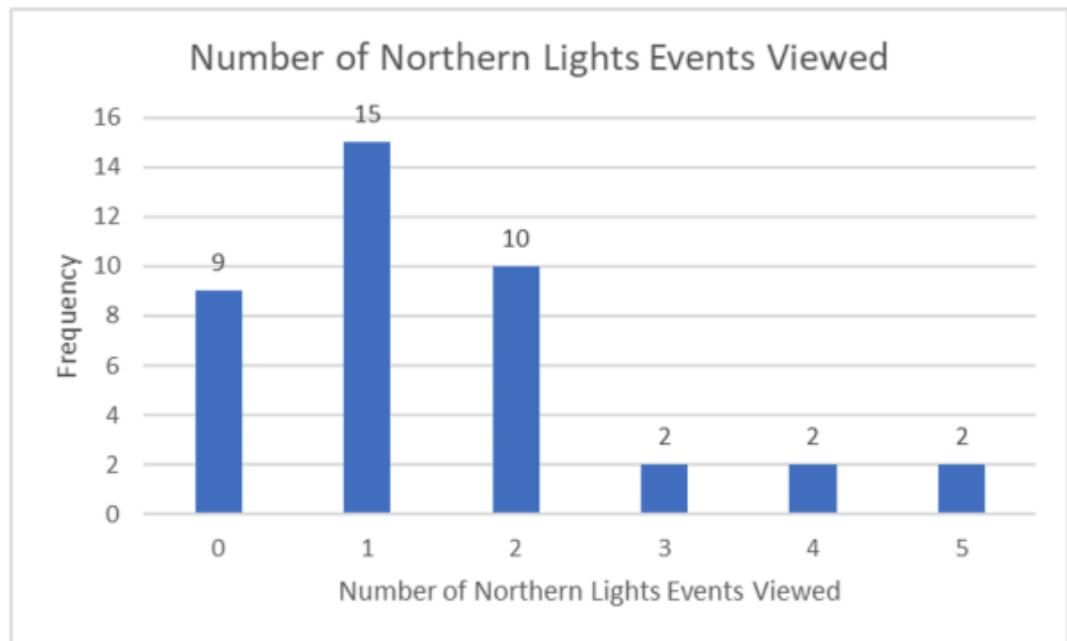
c. The probability a randomly chosen person in Yellowknife viewed northern lights last year is 77.5%.

4. Christy should be pleased considering this means her salary is higher than 85% of people surveyed, or rather it is in the top 15% of people surveyed.

5.
   a. The percentage of people who did not provide a response 7/30 = 0.2333 = 23.33%. 23.37% of grocery clerk refused to give answers.

   b. 16.67% + 20% + 20% = 56.67%. 56.67% of the grocery clerks says their weekly salary is from $200.00 to $350.00.

   c. Bar graph of part-time grocery employee weekly salary frequency.

Part-Time Grocery Employee Weekly Salary Frequency
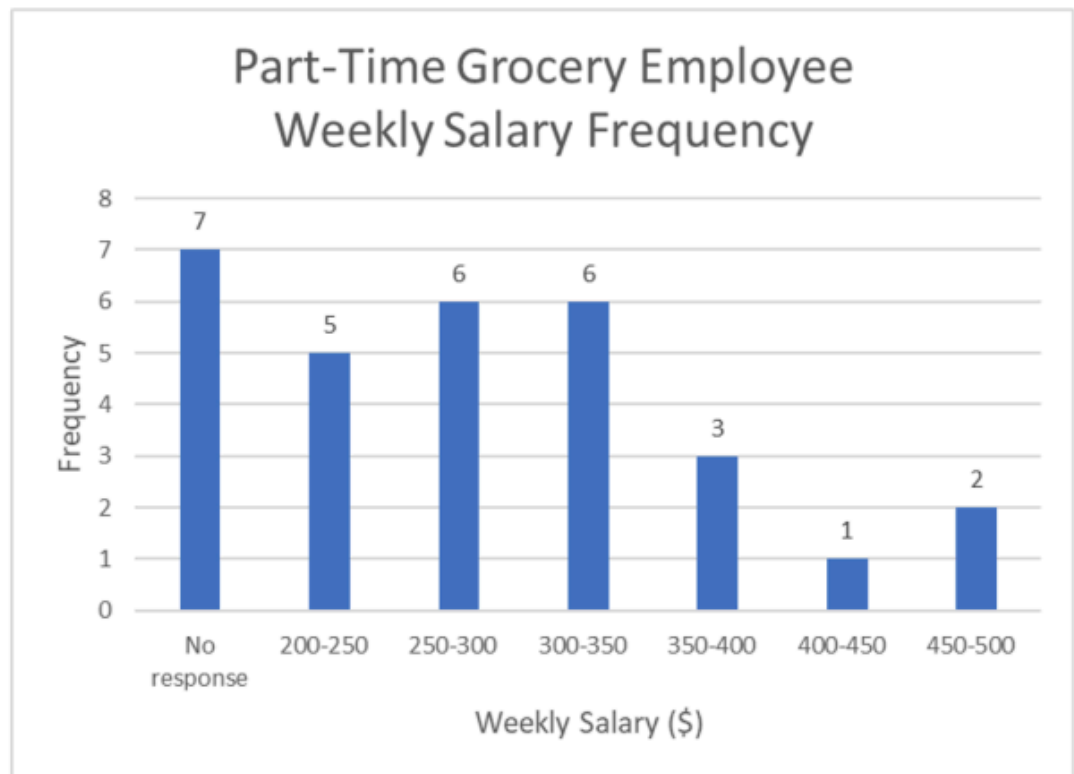
6. a. Line graph of annual precipitation in Atlin, BC.

Annual Atlin Precip. (mm)
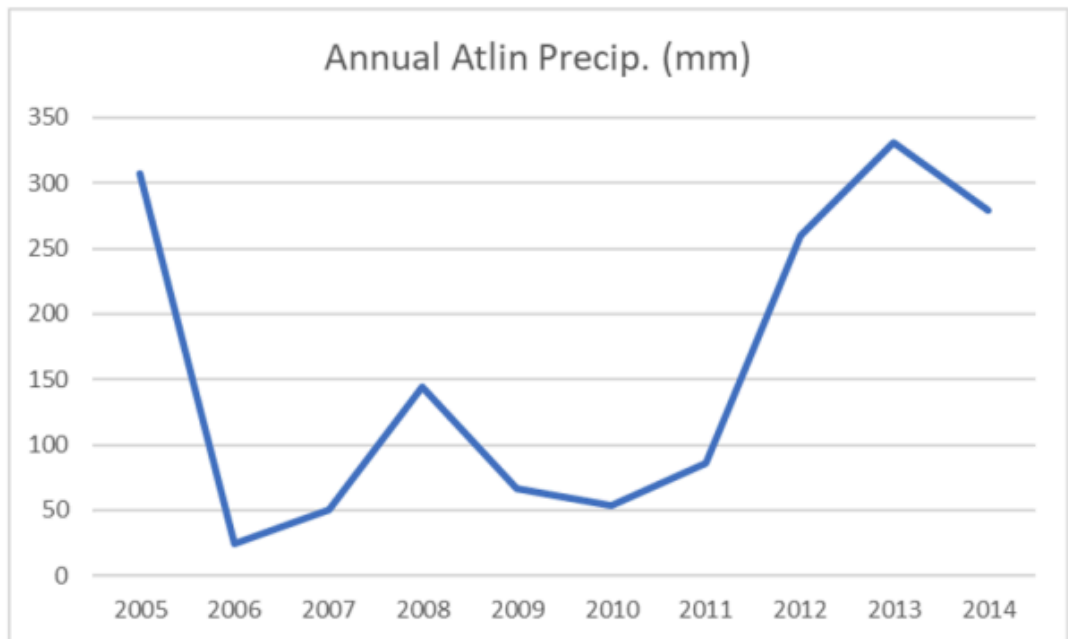
b. $\text{Width of bins} = \dfrac{(\text{max} - \text{min})}{\text{No. of bins}} = \dfrac{(330.80 - 24.10)}{6} = 51.12$, use 50 mm for the width.

Annual Precipitation Frequency Atlin, BC
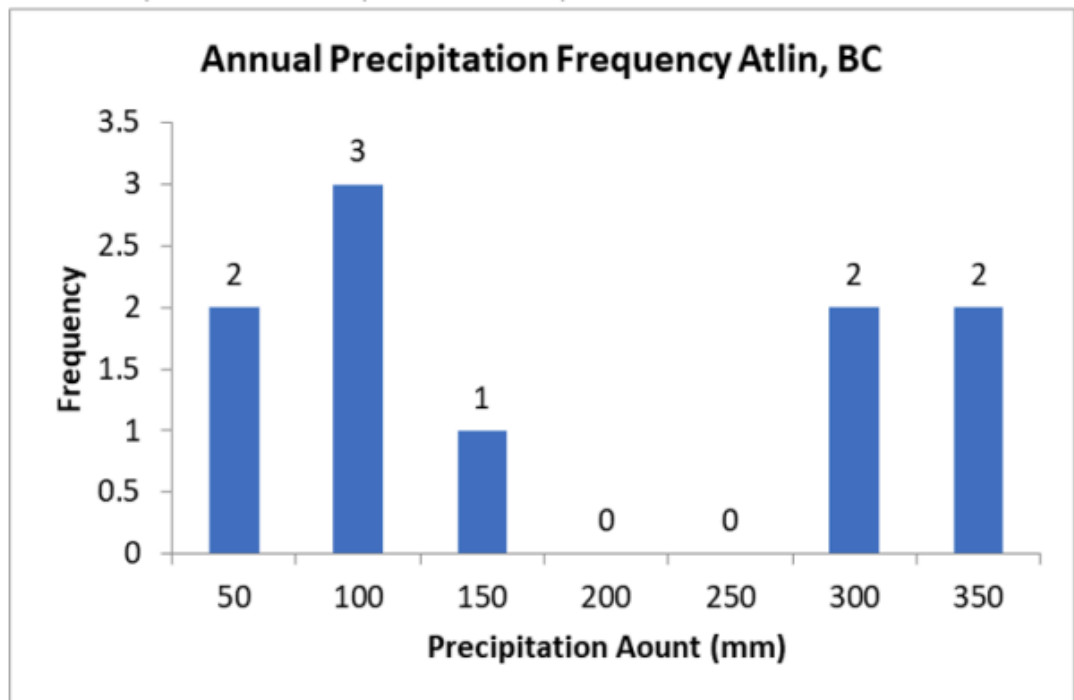
c. Sample mean = 160.15 mm

d. Interquartile Range = $Q_3 - Q_1$ = 286.375mm − 53.075 mm = 233.3 mm

2.3 Measures of the Center of the Data

The following activities and questions relate to material covered in Chapter 2.3 Measures of the Center of the Data (https://openstax.org/books/introductory-business-statistics/pages/2-3-measures-of-the-center-of-the-data) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 2.3 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2019/07/2.3-Data-Sets.xlsx).

Questions

1. Eleven selected Northern Canadian Adventure salespersons were asked the number of guests they sold the full package winter adventure to in one week during the pandemic in northern Canada January 2021: four salespeople answered 20 bookings; two salespeople answered 6 bookings; five salespeople answered 11 bookings. What is the sample mean, median and mode?

2. Twelve car-selling agencies in northern Canada were asked the number of cars they generally sell in a week. Two of them answered that they generally sell two cars; three of them said that they generally sell three cars; three answered four cars; and the four said that they sell five cars a week. Calculate the sample mean, median, and mode of this data.

3. The following data show the lengths (in feet) of boats that sold in northern British Columbia and Yukon. The data are ordered from smallest to largest: 16; 17; 19; 20; 20; 21; 23; 24; 25; 25; 25; 26; 26; 27; 27; 27; 27; 29; 30; 32; 33; 33; 34; 35; 37; 39; 40.

    a. What is the sample mean, median and mode?

    b. Create a histogram to display the data distribution.

4. Ten random Yukon University Bachelor of Business Administration students were given a general business-related test during their first year. In their final year, the same students were again tested with the same test. The questionnaire score totaled 100. Both scores listed below:

Student	Beginning Score	Ending Score
1	61	81
2	62	83
3	67	85
4	68	91
5	68	91
6	68	95
7	72	95
8	73	95
9	74	95
10	79	96

    a. What is the average improvement?

    b. Does it matter if the means are subtracted, or the individual values are subtracted?

Solutions

1. The sample mean is $\dfrac{(4 \times 20) + (2 \times 6) + (5 \times 11)}{11} = \dfrac{147}{11}$ = 13.36 sales

   The median is 11 sales.
   The mode is 11 sales.

2. Sample order: 2 2 3 3 3 4 4 4 5 5 5 5

   Sample mean = $\dfrac{2 \times 2 + 3 \times 3 + 3 \times 4 + 4 \times 5}{12}$ =3.75 car sales

   Median = $\dfrac{4 + 4}{2}$ = 4 car sales
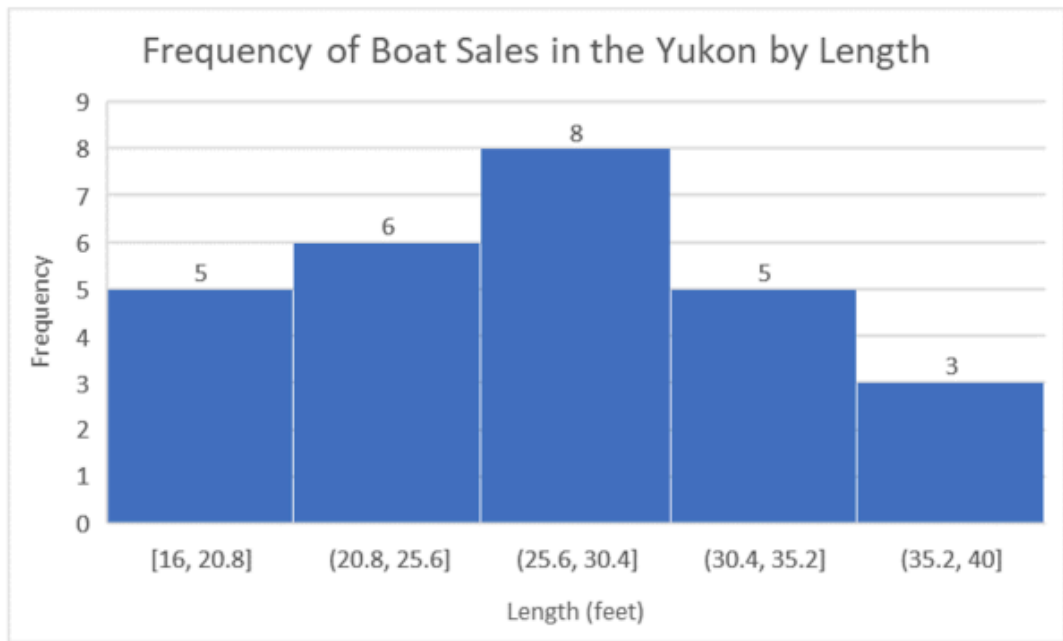
   Mode = 5 car sales

3.   a. The sample mean is:

   $$\frac{16 + 17 + 19 + 20 + 20 + 21 + 23 + 24 + 25 + 25 + 26 + 26 + 27 + 27 + 27 + 27 + 29 + 30 + 32 + 33 + 33 + 34 + 35 + 37 + 39 + 40}{27} = \frac{73}{27}$$

   = 27.3 feet.
   The median is 27 feet.
   The mode is 27 feet.

       b. Histogram of frequency of boat sales in Yukon

Frequency of Boat Sales in the Yukon by Length

4.

    a. The average beginning score is 69.2% and the average ending score is 90.7%, so the overall average improvement is 21.5%.

    b. To answer the previous question, it does not matter if the differences are taken after averages or averages are taken after individual differences.

2.7 Measures of the Spread of the Data

The following activities and questions relate to material covered in Chapter 2.7 (https://openstax.org/books/introductory-business-statistics/pages/2-7-measures-of-the-spread-of-the-data) (https://openstax.org/books/introductory-business-statistics/pages/2-7-measures-of-the-spread-of-the-data)Measures of the Spread of the Data (https://openstax.org/books/introductory-business-statistics/pages/2-7-measures-of-the-spread-of-the-data) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 2.7 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/2.7-Data-Sets.xlsx).

Questions

Janine lives in Whitehorse, YT and she just graduated from high school. She is considering a career as a baker and would like to know what the Canadian salary outlook is for this career. She collected data from Jobillico (table below) and would like to calculate the following values:

   a. What is the sample size?
   b. What is the sample mean?
   c. What is the sample variance?
   d. What is the sample standard deviation?

Average Salary for Baker in Canada [1]

Province	Average salary for Baker in Canada
Alberta	$41,589
British Columbia	$58,497
Manitoba	$44,321
New Brunswich	$65,756
Newfoundland	$67,803
Northwest Territories	$48,484
Nova Scotia	$64,981
Nunavut	$72,280
Ontario	$45,858
Prince Edward Island	$67,754
Quebec	$40,858
Saskatchewan	$39,773
Yukon	$37,440

Solutions

a. Sample size n = 13

b. Sample mean x = $53,491.85

c. Sample variance s2 = 165,274,453.5

d. Sample standard deviation s = $12,855.91

1. Data source: Jobillico. (n.d). *Average salary for baker in Canada*. https://www.jobillico.com/salary/job/Baker

Descriptive Statistics - Excel Tools Instruction

Download Descriptive Statistics – Excel Instructions Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2022/10/Descriptive-Statistics-–-Excel-Instructions-Data-Sets.xlsx).
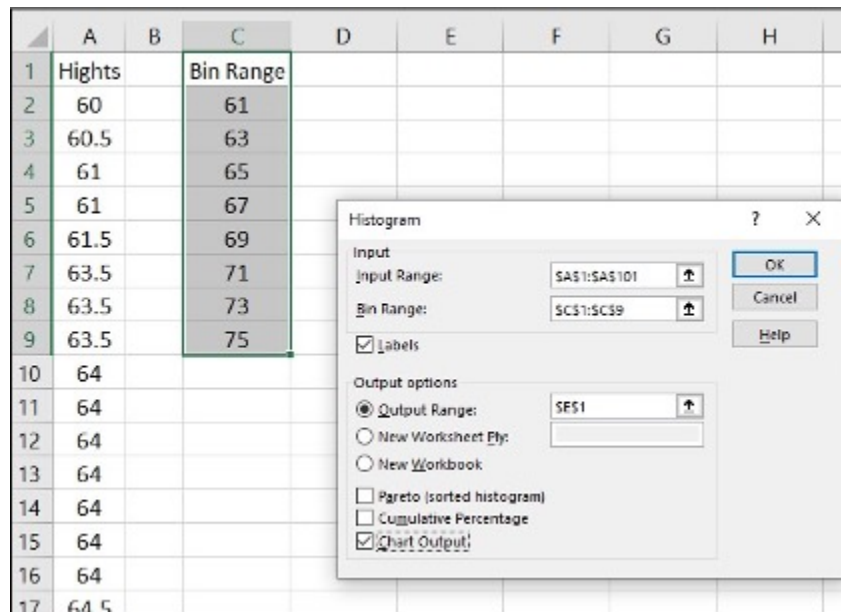
Histogram

- A **histogram** consists of contiguous (adjoining) boxes. It has both a horizontal axis and a vertical axis. The horizontal axis is labeled with what the data represents (for instance, distance from your home to school). The vertical axis is labeled either **frequency** or **relative frequency** (or percent frequency or probability). The graph will have the same shape with either label. The histogram (like the stemplot) can give you the shape of the data, the center, and the spread of the data.

Histogram in Excel

The following data are the heights (in inches to the nearest half inch) of 100 male semiprofessional soccer players. The heights are continuous data, since height is measured.

60; 60.5; 61; 61; 61.5; 63.5; 63.5; 63.5; 64; 64; 64; 64; 64; 64; 64; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5 66; 66; 66; 66; 66; 66; 66; 66; 66; 66; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67.5; 67.5; 67.5; 67.5; 67.5; 67.5; 67.5; 68; 68; 69; 69; 69; 69; 69; 69; 69; 69; 69; 69; 69.5; 69.5; 69.5; 69.5; 69.5; 70; 70; 70; 70; 70; 70; 70.5; 70.5; 70.5; 71; 71; 71; 72; 72; 72; 72.5; 72.5; 73; 73.5; 74

1. Enter the data into column A. Create bins Range into column C
2. Click Data, Data Analysis, Histogram and OK
3. Specify Input Range ($A$1:$A$101), Bin Range ($C$1:$C$9), and Output Range ($E$1)
4. Click Labels, Chart Output and OK
5. Make changes for the Histogram (i.e. delete Frequency, More on the right side)
6. Click on one blue rectangle, right click, click Format Data Series

Step 1, 2, 3 & 4: Enter data into column A, create bins Range into column C. Click Data, Data Analysis, Histogram and click OK. Specify Input Range ($A$1:$A$101), Bin Range ($C$1:$C$9), and Output Range ($E$1), then click Labels, Chart Output and OK
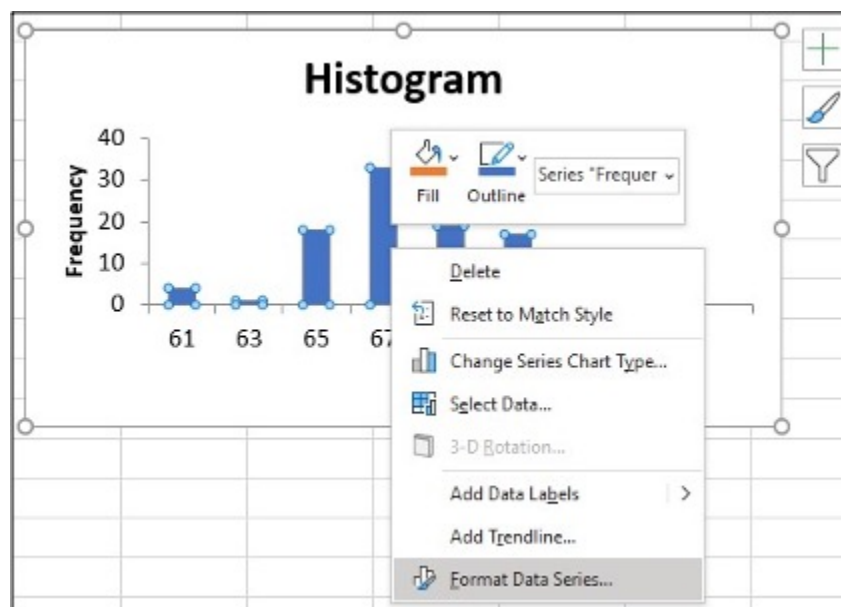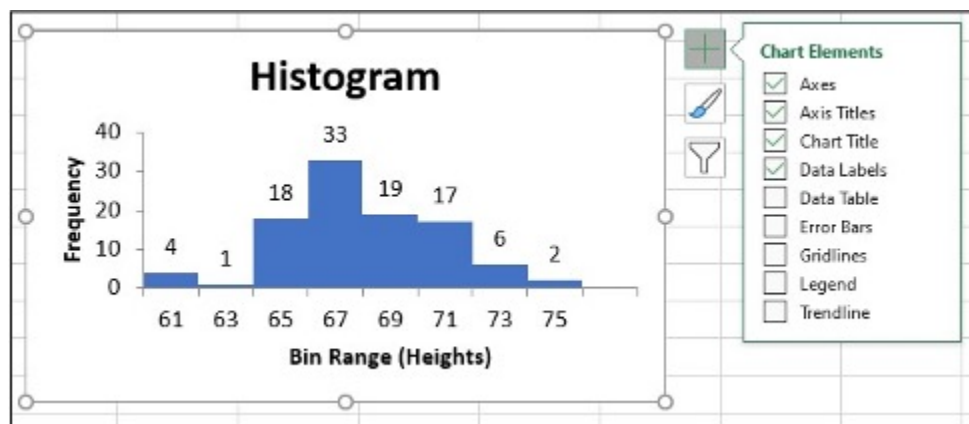

Step 5 & 6: Make changes for the Histogram, click on one blue rectangle, right click, click Format Data Series

7. To change the 'gap width', click on symbol ı under Gap Width slide line, hold and slide it to 0%

8. Click on the Histogram and icon + to make changes
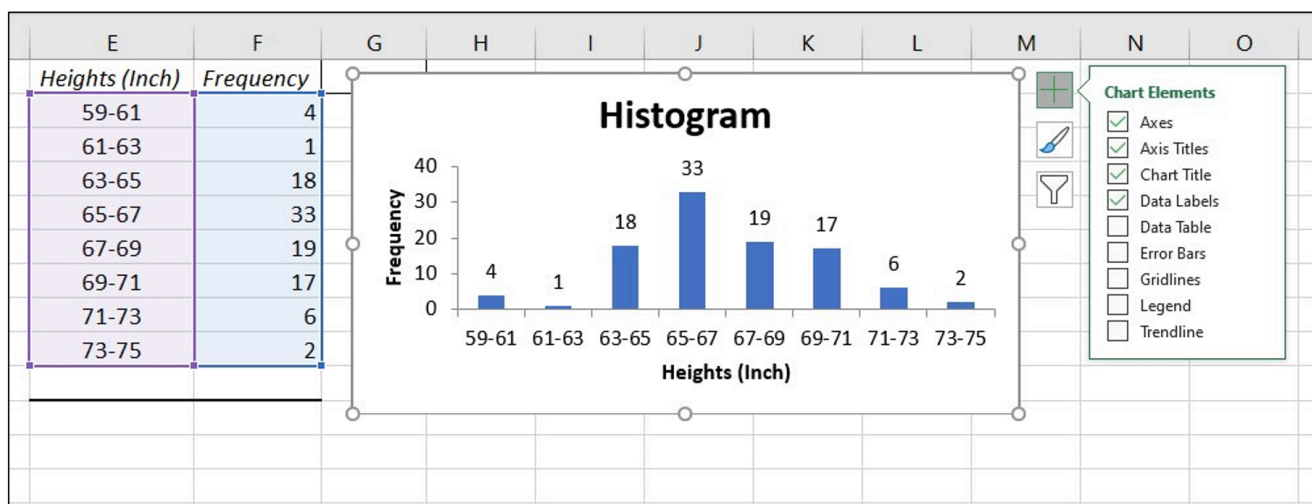
9. Change Axis Titles
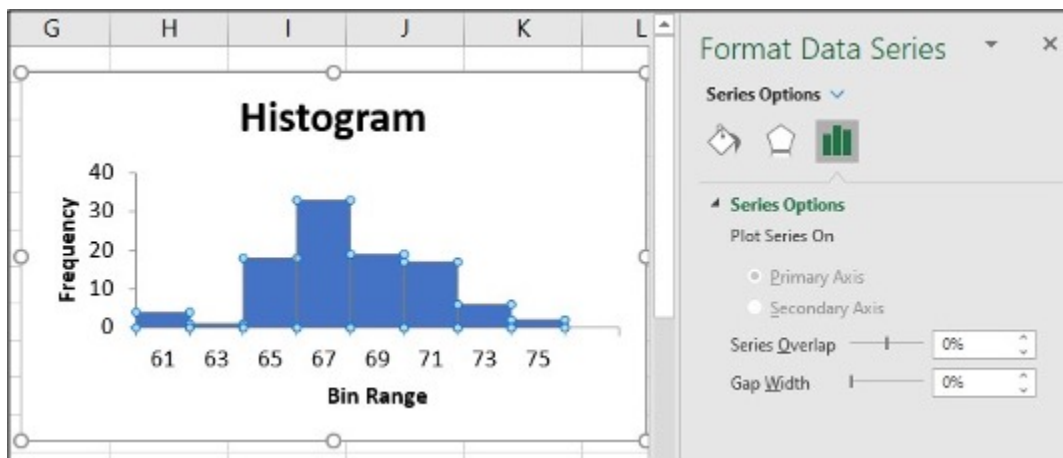
Step 7: To change the 'gap width', click on symbol ı under Gap Width slide line, hold and slide it to 0%

Step 8 & 9: Click on the Histogram and icon + to make changes, Change Axis Titles

To change the bin range on the histogram table, change the values in the X-Axis data.

Line Chart

- A **line chart** is often used to represent a set of data values in which a quantity varies with time. These graphs are useful for finding trends. That is, finding a general pattern in data sets including temperature, sales, employment, company profit or cost over a period of time.

Line chart in Excel

1. Enter the data (Year and Annual) into column A, B
2. Highlight the columns of data, click Insert, Line Chart
3. Click "More Line Charts"
4. Choose the graph with a single line.

Year	Aug	Sep	Oct	Nov	Dec	Annual
2003	184.6	185.2	185.0	184.5	184.3	184.0
2004	189.5	189.9	190.9	191.0	190.3	188.9
2005	196.4	198.8	199.2	197.6	196.8	195.3
2006	203.9	202.9	201.8	201.5	201.8	201.6
2007	207.917	208.490	208.936	210.177	210.036	207.342
2008	219.086	218.783	216.573	212.425	210.228	215.303
2009	215.834	215.969	216.177	216.330	215.949	214.537
2010	218.312	218.439	218.711	218.803	219.179	218.056
2011	226.545	226.889	226.421	226.230	225.672	224.939
2012	230.379	231.407	231.317	230.221	229.601	229.594

Construct a line chart for the Annual Consumer Price Index data only

Table 2.19

Table 2.19 Construct a line chart for the Annual Consumer Price Index data only

Step 1, 2 & 3: Enter the data (Year and Annual) into column A, B, then highlight the columns of data, click Insert, Line Chart. Click "More Line Charts"
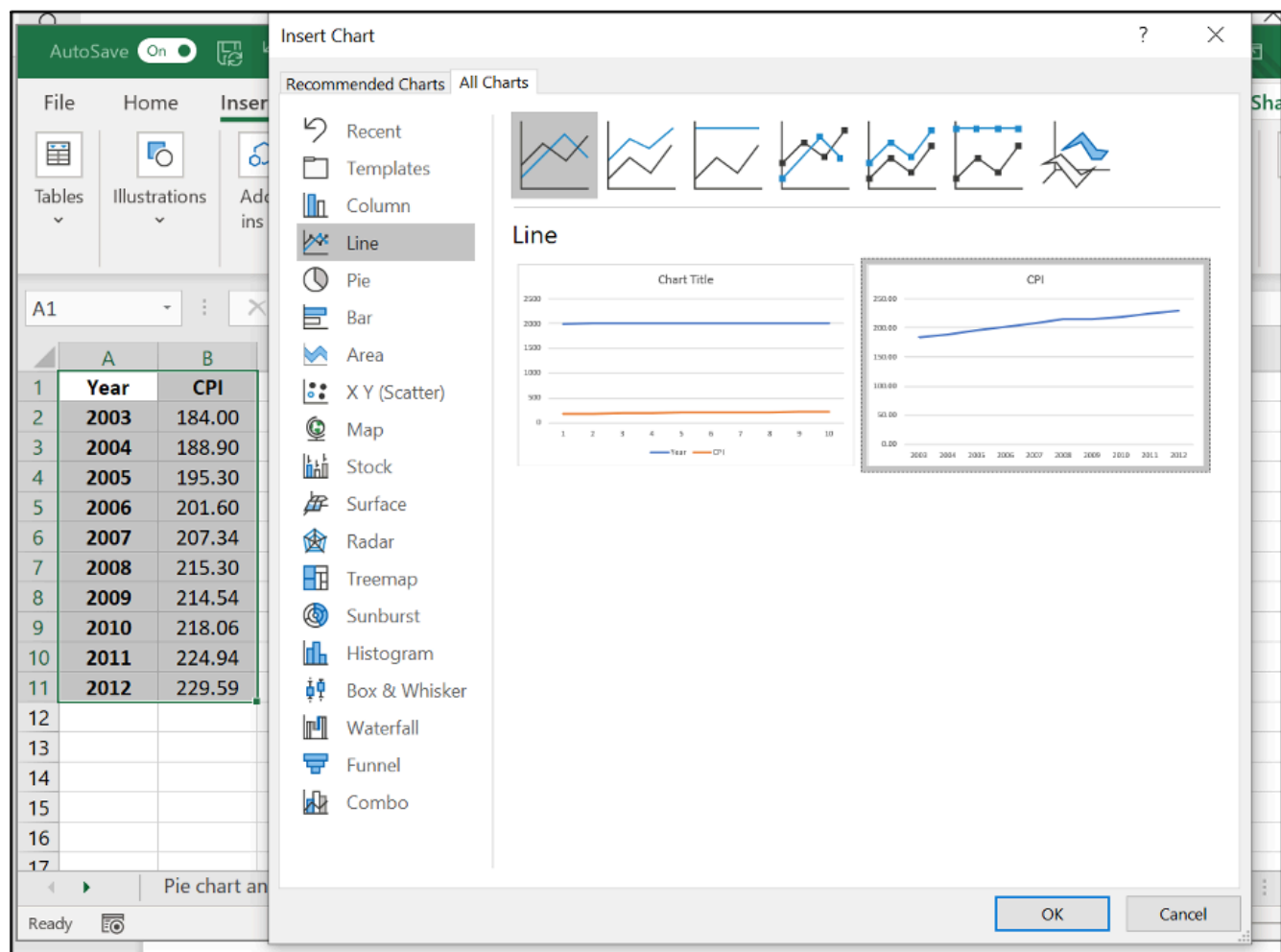
Step 4: Choose the graph with a single line

5. Click on Y-axis data, right click, Format Axis

6. Change Minimum and Maximum values which suit your data best, click Enter

7. Click on the new bar graph and icon + to make Axis title changes
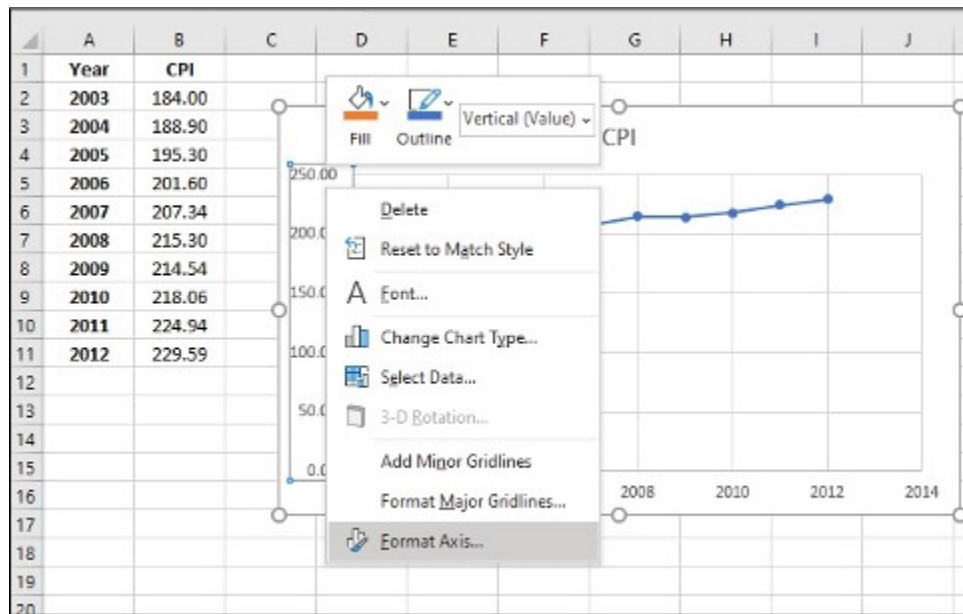
Step 5: Click on Y-axis data, right click, Format Axis

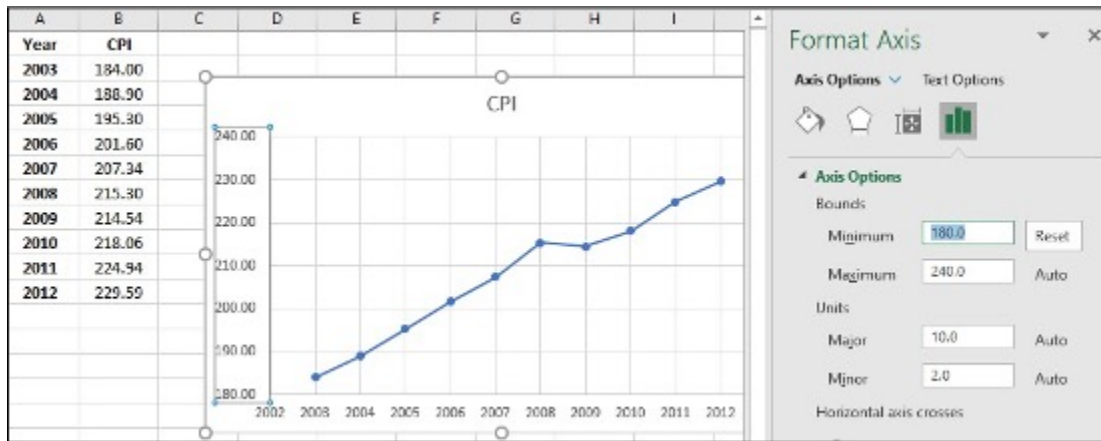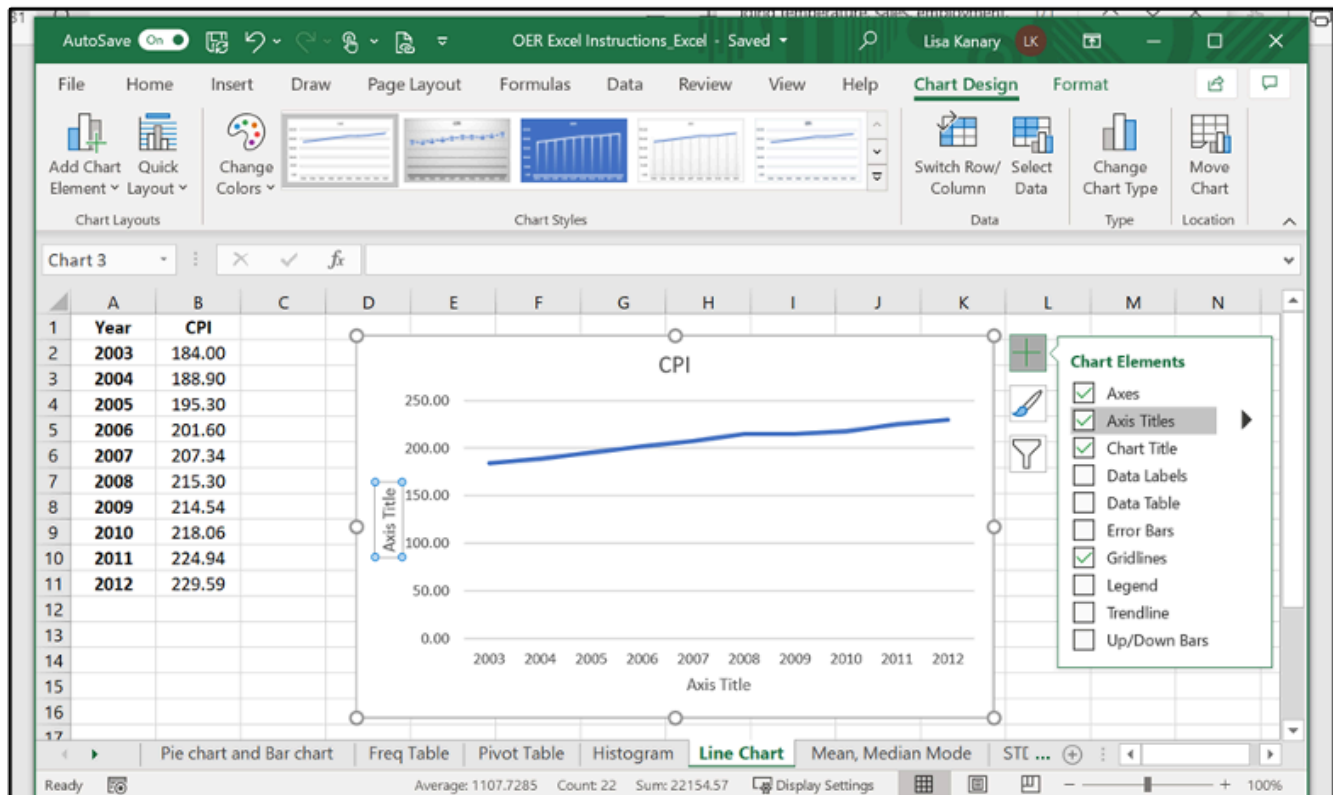*Step 6: Change Minimum and Maximum values which suit your data best, click Enter*

Step 7: Click on the new bar graph and icon + to make Axis title changes

Mean, Median & Mode

- **Mean:** a number that measures the central tendency of the data; a common name for mean is 'average'.

- **Median:** a number that separates ordered data into halves; half the values are the same number or smaller than the median and half the values are the same number or larger than the median. The median may or may not be part of the data.

- **Mode:** the value that appears most frequently in a set of data.

Mean, Median & Mode in Excel (Formula tool)

Use the following information to answer the next three exercises: The following data show the lengths of boats moored in a marina. The data are ordered from smallest to largest:

16; 17; 19; 20; 20; 21; 23; 24; 25; 25; 25; 26; 26; 27; 27; 27; 28; 29; 30; 32; 33; 33; 34; 35; 37; 39; 40

1. Enter the data into column A
2. Create a table for Mean, Median and Mode
3. Enter =AVERAGE(A:A) in cell D2, click Enter
4. Enter =MEDIAN(A:A) in cell D3, click Enter

5. Enter =MODE(A:A) in cell D4, click Enter

Step 1, 2 & 3: Enter the data into column A. Create a table for Mean, Median and Mode. Enter =AVERAGE(A:A) in cell D2, click Enter

Step 4: Enter =MEDIAN(A:A) in cell D3, click Enter

Step 5: Enter =MODE(A:A) in cell D4, click Enter

Mean, Median & Mode in Excel (Data Analysis tool)

Use the following information to answer the next three exercises: The following data show the lengths of boats moored in a marina. The data are ordered from smallest to largest:

16; 17; 19; 20; 20; 21; 23; 24; 25; 25; 25; 26; 26; 27; 27; 27; 28; 29; 30; 32; 33; 33; 34; 35; 37; 39; 40

1. Enter the data into column A
2. Click Data, Data Analysis
3. Click Descriptive Statistics, OK
4. Specify Input Range ($A$1:$A$28), Output Range ($C$1)
5. Click Labels in first row, Summary statistics and OK
6. Find Mean, Median and Mode in the Summary statistics table
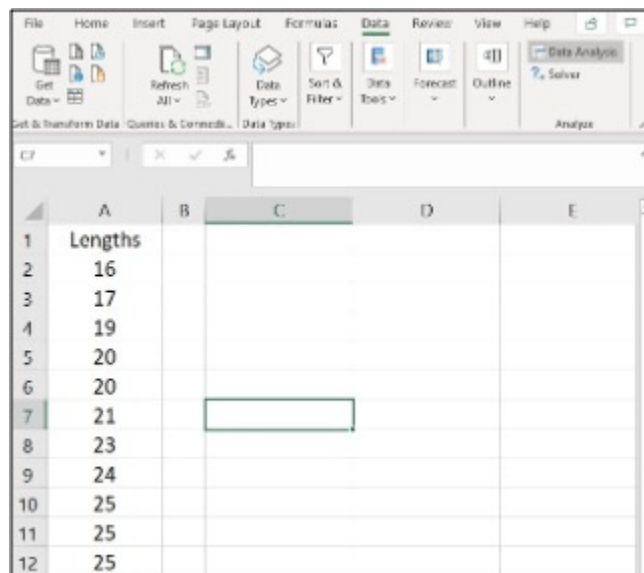
Step 1 & 2: Enter the data into column A and Click Data, Data Analysis
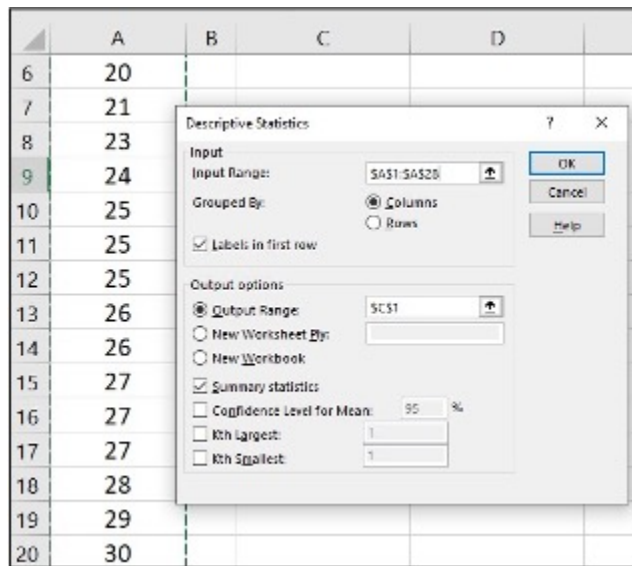
*Step 3, 4 & 5: Click Descriptive Statistics, click OK.
Specify Input Range ($A$1:$A$28), Output Range
($C$1). Click Labels in first row, Summary statistics and
OK*

*Step 6: Find Mean, Median and Mode in the
Summary statistics table*

Variance & Standard Deviation

- The **variance** is the **average of the squares of the deviations**.

- If x is a number, then the difference "x minus the mean" is called its **deviation**. The standard deviation is a number that is equal to the square root of the variance and measures how far data values are
  from their mean.

- Notation: s for sample standard deviation and σ for population standard deviation.

Variance and Standard Deviation in Excel (Formula tool)

In a fifth-grade class, the teacher was interested in the average age and the sample standard deviation of the ages of her students. The following data are the ages for a SAMPLE of n = 20 fifth grade students. The ages are rounded to the nearest half year: 9; 9.5; 9.5; 10; 10; 10; 10; 10.5; 10.5; 10.5; 10.5; 11; 11; 11; 11; 11; 11; 11.5; 11.5; 11.5;

1. Enter the data into column A
2. Create a table for Variance and Standard deviation
3. Enter =VAR.S(A:A) in cell D2, click Enter
4. Enter =STDEV.S(A:A) in cell D3, click Enter

Step 1, 2 & 3: Enter the data into column A and create a table for Variance and Standard deviation. Enter =VAR.S(A:A) in cell D2, click Enter

Step 4: Enter =STDEV.S(A:A) in cell D3, click Enter

Variance and Standard Deviation in Excel (Data Analysis tool)

In a fifth-grade class, the teacher was interested in the average age and the sample standard deviation of the ages of her students. The following data are the ages for a SAMPLE of n = 20 fifth grade students. The ages are rounded to the nearest half year: 9; 9.5; 9.5; 10; 10; 10; 10; 10.5; 10.5; 10.5; 10.5; 11; 11; 11; 11; 11; 11; 11.5; 11.5; 11.5;

1. Enter the data into column A
2. Click Data, Data Analysis
3. Click Descriptive Statistics, OK
4. Specify Input Range ($A$1:$A$21), Output Range ($C$6)
5. Click Labels in first row, Summary statistics and OK
6. Find Variance and Standard Deviation in the Summary statistics table
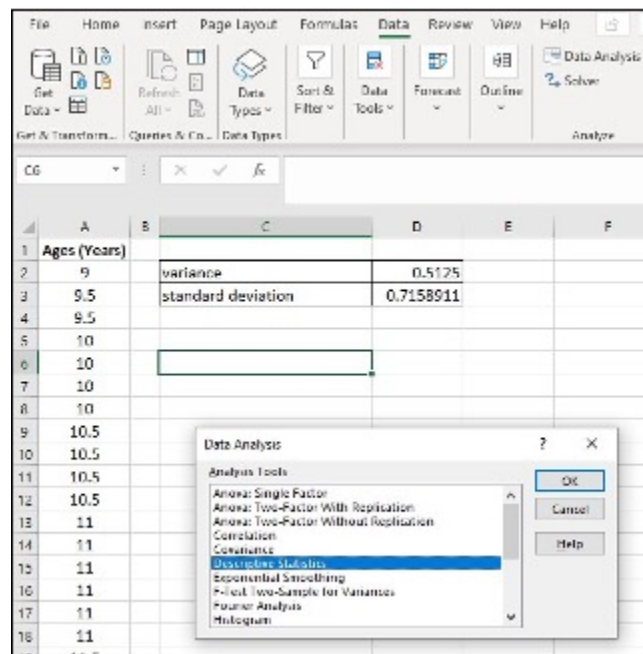
Step 1 & 2: Enter the data into column A and Click Data,
Data Analysis
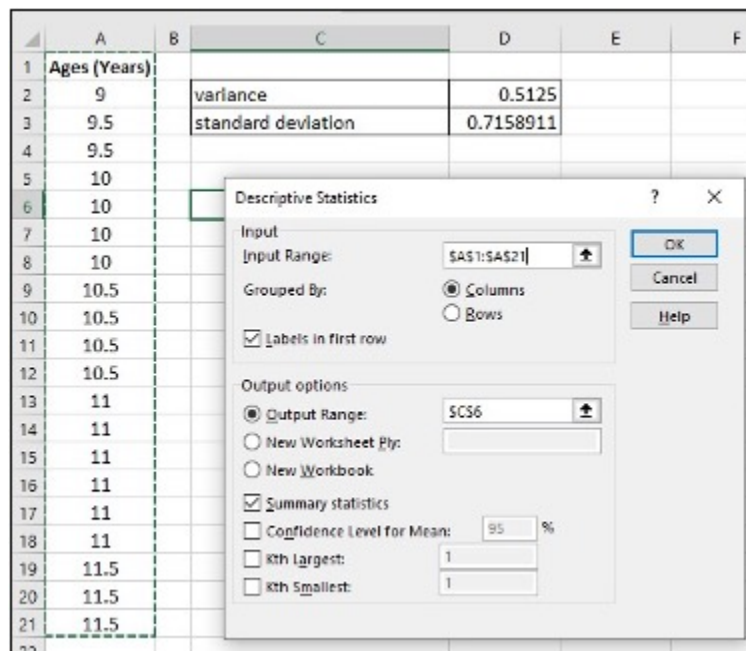
Step 3, 4 & 5: Click Descriptive Statistics, click OK. Specify Input
Range ($A$1:$A$21), Output Range ($C$6). Click Labels in first
row, Summary statistics and OK

Ages (Years)	
Mean	10.525
Standard Error	0.160078106
Median	10.5
Mode	11
Standard Deviation	0.715891053
Sample Variance	0.5125
Kurtosis	-0.53129464
Skewness	-0.45516018
Range	2.5
Minimum	9
Maximum	11.5
Sum	210.5
Count	20

Step 6: Find Variance and Standard Deviation in the Summary statistics table

Quartiles

- **Quartiles** are the numbers that separate the data into quarters; quartiles may or may not be part of the data. The second quartile is the median of the data.

Quartiles in Excel

Use the following data (first exam scores) from Susan Dean's spring pre-calculus class:

33; 42; 49; 49; 53; 55; 55; 61; 63; 67; 68; 68; 69; 69; 72; 73; 74; 78; 80; 83; 88; 88; 88; 90; 92; 94; 94; 94; 94; 96; 100

1. Enter the data into column A, and sort them
2. Create a table for Q1, Q2, Q3 and Q3-Q1
3. Enter =QUARTILE.EXC($A:$A,1) in cell D2, click Enter
4. Enter =QUARTILE.EXC($A:$A,2) in cell D3, click Enter
5. Enter =QUARTILE.EXC($A:$A,3) in cell D4, click Enter
6. Enter = in cell D5, click on cell D4, enter –, clike on cell D2, click Enter

Step 1, 2 & 3: Enter the data into column A, and sort them. Create a table for Q1, Q2, Q3 and Q3-Q1. Enter =QUARTILE.EXC($A:$A,1) in cell D2, click Enter

Step 4: Enter =QUARTILE.EXC($A:$A,2) in cell D3, click Enter

Step 5: Enter =QUARTILE.EXC($A:$A,3) in cell D4, click Enter

Step 6: Enter = in cell D5, click on cell D4, enter -, clike on cell D2, click Enter

Media Attributions

- Screenshots of Excel are used with permission from Microsoft (https://www.microsoft.com/en-us/legal/intellectualproperty/copyright/permissions).

Chapter 3: Probability Topics

3.1 Terminology

The following activities and questions relate to material covered in Chapter 3.1 Terminology (https://openstax.org/books/introductory-business-statistics/pages/3-1-terminology) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 3.1 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/3.1-Data-Sets.xlsx).

Questions

1. In 2020, the following survey information was provided by 177 students at Yukon University asking them to define which school services are most needed or most important to that student.

Important Services in Yukon University Survey Result [1]

Most Important Services	Frequency
Academic support	30
Study space	22
Communication with Instructors	22
Learning Commons	21
Mental health supports/counselling	11
Student-to-student connection	10
Other	

Let:

- AS = the event of randomly selecting a student that responded, "Academic support".
- SS = the event of randomly selecting a student that responded, "Study space".
- CI = the event of randomly selecting a student that responded, "Communication

1. Data source: Macdonald, D. (2020). *Student survey 2020: Results*. Yukon University Institutional Research and Planning. https://www.yukonu.ca/sites/default/files/inline-files/StudentSurveyResults2020_Report_0.pdf

with Instructors".

- ◦ LC = the event of randomly selecting a student that responded, "Learning Commons".

- ◦ MH = the event of randomly selecting a student that responded, "Mental health supports/counselling".

- ◦ SC = the event of randomly selecting a student that responded "Student-to-student connection".

- ◦ O = the event of randomly selecting a student that responded with various other answers.

Find:

a. P(AS)

b. P(SS)

c. P(CI)

d. P(LC)

e. P(MH)

f. P(SC)

g. P(O)

2. The 39th Yukon General Election was held on April 21, 2021. Data was retrieved from Yukon Elections:

2021 Yukon General Election Results [2]

Electoral District	Number of Votes
Copperbelt North	1381
Copperbelt South	1274
Klondike	1114
Kluane	782
Lake Laberge	1277
Mayo Tatchun	632
Mount Lorne – Southern Lakes	1144
Mountainview	1052
Pelly Nisutlin	713
Porter Creek Centre	1681
Porter Creek North	1143
Porter Creek South	655
Riverdale North	1124
Riverdale South	1056
Takhini Kopper King	1205
Vuntut Gwitchin	156
Watson Lake	550
Whitehorse Centre	1059
Whitehorse West	1003

Find the probabilities a randomly selected voter (someone who voted in the election) had voted in each of the electoral districts below:

    a. P(Whitehorse Centre)

    b. P(Riverdale North)

    c. P(Porter Creek Centre)

    d. P(Mountainview)

    e. P(Takhini Kopper King)

    f. P(Copperbelt North)

2. Data source: Election Yukon. (2021). *Unofficial election results 2021*. https://electionsyukon.ca/en/content/
unofficial-election-results-2021

Kathleen Lake

Solutions

1.
   a. $P(AS) = \dfrac{30}{177} = 0.1695 = 16.95\%$

   b. $P(SS) = \dfrac{22}{177} = 0.1243 = 12.43\%$

   c. $P(CI) = \dfrac{22}{177} = 0.1243 = 12.43\%$

   d. $P(LC) = \dfrac{21}{177} = 0.1186 = 11.86\%$

   e. $P(MH) = \dfrac{11}{177} = 0.0621 = 6.21\%$

   f. $P(SC) = \dfrac{10}{177} = 0.0565 = 5.65\%$

   g. $P(VA) = \dfrac{(177 - 30 - 22 - 22 - 21 - 11 - 10)}{177} = \dfrac{61}{177} = 0.3446 = 34.46\%$

2.

    a. P(Whitehorse Centre) = $\dfrac{1059}{19001}$ = 5.57%

    b. P(Riverdale North) = $\dfrac{1124}{19001}$ = 5.92%

    c. P(Porter Creek Centre) = $\dfrac{1681}{19001}$ = 8.85%

    d. P(Mountainview) = $\dfrac{1052}{19001}$ = 5.54%

    e. P(Takhini Kopper King) = $\dfrac{1205}{19001}$ = 6.34%

    f. P(Copperbelt North) = $\dfrac{1381}{19001}$ = 7.27%

Media Attributions

- Kathleen Lake by Lisa Kanary is licensed under a CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).

3.3 Two Basic Rules of Probability

The following activities and questions relate to material covered in Chapter 3.3 Two Basic Rules of Probability (https://openstax.org/books/introductory-business-statistics/pages/3-3-two-basic-rules-of-probability) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 3.3 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/3.3-Data-Sets.xlsx).

Questions

1. According to the Yukon Bureau of Statistics, the total population of the Yukon territory in the year 2016 was 35,874 people. [1]

   ◦ From the 2016 census we see that 32,538 people spoke English most often at home.

   ◦ 3,336 people reported speaking a language other than English most often at home.

   ◦ Of those who speak another language at home, 0.2% spoke an aboriginal language most often at home.

   Let:

   ◦ E = speaks English most often, at home.

   ◦ E'= Speaks another language most often, at home.

   ◦ A = Speaks an aboriginal language most often, at home.

   Finish each probability statement. Check your work after answering b. by making sure P(E) + P(E') = 1.

   a. P(E')

   b. P(E)

   c. P(A ∩ E′)

   d. P(A | E′)

2. Yukon University is looking to ask the local café to cater some desserts for a faculty meeting.

---

1. Data source: Yukon Bureau of Statistics. (2016). *Yukon Census historical population, 1901 to 2016*. https://yukon.ca/sites/yukon.ca/files/ybs/fin-yukon-census-historical-population-1901-2016.pdf.

The café makes cookies with chocolate (36%) and nuts (12%) and, of those, 8% contain both chocolate and nuts. Let's imagine an instructor attending the meeting is allergic to both chocolate and nuts but wants a safe cookie to eat from this selection.

    a. Find the probability that a cookie contains chocolate or nuts (they cannot eat the cookies).

    b. Find the probability that a cookie does not contain chocolate or nuts (they can eat the cookie)

3. The table below is the roster from newly renamed Edmonton Elks (formerly "Eskimos").

Weight of Edmonton Elks Roster [2]

Jersey Number (J) / Weight (W)	≤200 lbs	201 – 231 lbs	232 – 262 lbs	≥263 lbs
1 – 33	9	6	2	0
34 – 66	4	3	4	7
67 – 99	9	1	1	4

For the following, suppose that you randomly selected one player from the Elks roster.

    a. Find the probability that the jersey number is from 1 – 33.

    b. Find the probability that the player weighs at most 200 pounds.

    c. Find the probability that their jersey number is from 1 – 33 AND the player weighs at most 200 pounds.

    d. Find the probability that their jersey number is from 1 – 33 OR the player weighs at most 200 pounds.

    e. Find the probability that their jersey number is from 1 – 33 GIVEN that the player weighs at most 200 pounds.

Solutions

1.

    a. $P(E') = \dfrac{3,336}{35,874} = 0.093$ or 9.3%

    b. $P(E) = \dfrac{32,538}{35,874} = 0.907$ or 90.7%.

    Check your work by making sure P(E) + P(E') = 1!

    c. $P(A \cap E') = \dfrac{7}{35,874} = 0.0002$ or 0.02%

2. Data source: Edmonton Elks. (n.d.). *Roster*. https://www.goelks.com/roster/

    d. P (A | E′) = P(A ∩ E′) / P(E') = $\dfrac{0.0002}{0.093}$ = 0.0022

2. Let:

C = the event that the cookie contains chocolate.

N = the event that the cookie contains nuts.

    a. P(C ∪ N) = P(C) + P(N) – P(C ∩ N) = 0.36 + 0.12 – 0.08 = 0.40

    b. P(NEITHER chocolate NOR nuts) = 1 – P(C ∪ N) = 1 – 0.40 = 0.60

3.

    a. P(J = 1-33) = $\dfrac{17}{50}$ = 34%

    b. P(W ≤ 200) = $\dfrac{22}{50}$ = 44%

    c. P(J = 1-33 ∩ W ≤ 200) = $\dfrac{9}{50}$ = 18%

    d. P(J = 1-33 ∪ W ≤ 200) = $\dfrac{17}{50} + \dfrac{22}{50} - \dfrac{9}{50} = \dfrac{30}{50}$ = 60%

    e. P(J = 1-33 | W ≤ 200) = $\dfrac{P(J - 33 \cap W \le 200)}{P(W \le 200)} = \dfrac{0.18}{0.44}$ = 40.9%

Chapter 4: Discrete Random Variables

4.1 Hypergeometric Distribution

The following activities and questions relate to material covered in Chapter 4.1 (https://openstax.org/books/introductory-business-statistics/pages/4-1-hypergeometric-distribution) Hypergeometric Distribution (https://openstax.org/books/introductory-business-statistics/pages/4-1-hypergeometric-distribution) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 4.1 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/4.1-Data-Sets.xlsx).

Questions

1. A member of the YT Elites Volleyball Club has a volleyball practice four days a week. She practices for all of the four days 87% of the time, three days 6% of the time, two days 4% of the time, one day 2% of the time, and no days 1% of the time. Suppose one week is randomly selected.

    a. Define the random variable X.

    b. Construct a probability distribution table for the data.

    c. We know that for a probability distribution function to be discrete, it must have two characteristics. One is that the sum of the probabilities is one. What is the other characteristic?

2. Valerie volunteers at the local territorial farmer's market no more than 7 hours and no less than 1 hour, per week. She volunteers 7 hours 42% of this summer, 6 hours 29% of this summer, 5 hours 10% of this summer, 4 hours 7% of this summer, 3 hours 6% of this summer, 2 hours 4% of this summer, and 1 hour 2% of this summer.

    a. Define the random variable X.

    b. What values does X take on?

    c. Construct a table to organize the data.

    d. Find the probability that Valerie volunteers for more than 5 hours, per week.

    e. Find the probability that Valerie volunteers for at most 3 hours, per week.

3. Over the last 13 years, Yukon University, has averaged 5,574 students enrolled, per year. Of these enrollments 11.35% are full-time credit students, 10.15% are part-time credit students, and 78.49% are non-credit students. One student is selected at random. Data source: Yukon

University. *Fast facts, Credit program enrolment*. (2021). https://www.yukonu.ca/about-us/publications-plans-reports/institutional-research/fast-facts

   a. Define the Random Variable, is it discrete or continuous?

   b. Use the data to construct a probability distribution table.

   c. What is the probability that the student will be enrolled in credit courses?

   d. What are the two characteristics required for a probability distribution function to be discrete?

Solutions

1.   a. Let x be the number of days volleyball member attends the practice per week.

   b.

x	P(x)
0	0.01
1	0.02
2	0.04
3	0.06
4	0.87

   c. Each probability is between zero (0) and one (1), inclusive.

2.   a. Let X = the number of hours Valerie volunteers at the market.

   b. The values that X takes on are the number of hours Valerie volunteers at the market. X = 1, 2, 3, 4, 5, 6, 7

   c.

X	P(X) (%)
1	2
2	4
3	6
4	7
5	10
6	29
7	42

   d. P(X > 5) = 29% + 42% = 71%.
The probability that Valerie volunteers for more than 5 hours per week is 71%.

    e. $P(X < 3) = 2\% + 4\% + 6\% = 12\%$.

       The probability that Valerie volunteers for at most 3 hours per week is 12%.

3.

    a. The discrete random variable is the type of program the student is enrolled in. (FT, PT, NC)

b.

X	P(X)
Full-time	0.1135
Part-time	0.1015
Non-credit	0.7849

    c. There is a 11.35% + 10.15% = 21.51% chance the randomly selected student will be enrolled in credit courses.

    d. Each probability is between 0 and 1, inclusive. The sum of probabilities is 1.

4.2 Binomial Distribution

The following activities and questions relate to material covered in Chapter 4.2 Binomial Distribution (https://openstax.org/books/introductory-business-statistics/pages/4-2-binomial-distribution) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 4.2 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/4.2-Data-Sets.xlsx).

Questions

1. When a patient calls the medical advice line saying that they have the flu, the probability of him or her testing positive for the flu is 4%. Of the next 25 patients who call in claiming to have the flu, we are curious to see how many test positive for the flu. Find the probability that at exactly four of the 25 patients test positive flu.

2. People visiting a northern outdoor gear rental shop during the summer often rent camping and backpacking gear. The probability distribution for gear rentals per customer is given in the table below. There is a five-gear rental limit per customer at this shop.

X	P(x)
0	0.05
1	0.40
2	0.25
3	
4	0.10
5	0.07

    a. Describe the random variable X in words.

    b. Find the probability that a customer rents three gear items.

    c. Find the probability that a customer rents at least four gear items.

    d. Find the probability that a customer rents at most 2 gear items.

3. In a northern Alberta community, 46% of the children are involved in an organized sport. Suppose you randomly pick 20 children.

    a. Find the probability that at least 12 are enrolled in an organized sport.

    b. Is it more likely that 11 or that 12 will be enrolled in an organized sport?

4. Suppose about 70% of youth (age 12 – 19 years) living in a northern Canadian town attend local music festivals. A group of 15 youths is randomly chosen.

    a. In words, define the random variable X.

    b. List the values that X may take on.

    c. Give the distribution of X. X ~ _____ (___, ___)

    d. Find the probability that 7 or 8 will attend.

Solutions

1. This is binominal distribution as there are two outcomes; test positive for the flu (success), test negative for the flu (failure).

$$p(x = 4, n = 25, p = 0.04) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

$$= \frac{25!}{4!25-4!}0.04^4(1-0.04)^{25-4}$$

$$= 0.0137$$

$$= 1.37\%$$

There is a 1.37% probability that exactly 4 out of 25 patients who call in claiming the flu, test positive for the flu.

2.
    a. The random variable X is an example of discrete random variables. X has a countable number of possible values. In the above problem X represents the number of items a customer will rent.
X = the number of northern outdoor gear items a customer will rent.

    b. 1− 0.05 + 0.4 + 0.25 + 0.10+ 0.07 = 0.13
The probability that a customer rents three gear items is 13%.

    c. P (X ≥ 4) = P (0.10) + P (0.07) = 0.17
The probability that a customer rents at least four gear items is 17%.

    d. P (X ≤ 2) = P (0.05) + P (0.40) + P (0.25) = 0.70
The probability that a customer rents at least 2 gear items is 70%.

3.
    a. P(x > = 12) = 0.1511 = 15.11%

    b. P(x = 11) = 0.1280, P(x = 12) = 0.0817. It is more likely that 11, rather than 12, children will be enrolled in an organized sport.

4.
    a. Let X = the number of youths who would attend a local music festival

b. X = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

c. X ~ B(15, 0.70)

d. P(X = 7 or X = 8) = 0.0348 + 0.0811 = 0.1159 = 11.59%

Discrete Random Variables - Excel Tools Instruction

Download Discrete Random Variables – Excel Instructions Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2022/10/Discrete-Random-Variables-–-Excel-Instructions-Data-Sets.xlsx).

Binomial Distribution

- A more valuable probability density function with many applications is the **binomial distribution**. This distribution will compute probabilities for any binomial process.

Binomial Distribution in Excel

A fair coin is flipped 15 times. Each flip is independent. What is the probability of getting more than ten heads?

Let X = the number of heads in 15 flips of the fair coin. X takes on the values

1. Enter the value of n and p. Create a table for the probability and cumulative of 0, 1, 2, 3, …, 15.

2. Enter =BINOM.DIST(A5,$B$1,$B$2,FALSE) in cell B5, click Enter

3. Click B5, move the mouse to the right bottom corner, a little + shows up

4. Click on the little + and drag down

Step 1 & 2: Enter the value of n and p. Create a table for the probability and cumulative of 0, 1, 2, 3, …, 15. Enter =BINOM.DIST(A5,$B$1,$B$2,FALSE) in cell B5, click Enter

Step 3 & 4: Click B5, move the mouse to the right bottom corner, a little + shows up. Click on the little + and drag down

5. Enter =BINOM.DIST(A5,$B$1,$B$2,TRUE) in cell C5, click Enter

6. Click C5, move the mouse to the right bottom corner, a little + shows up

7. Click on the little + and drag down

Step 5: Enter =BINOM.DIST(A5,$B$1,$B$2,TRUE) in cell C5, click Enter

Step 6 & 7: Click C5, move the mouse to the right bottom corner, a little + shows up. Click on the little + and drag down

Media Attributions

- Screenshots of Excel are used with permission from Microsoft (https://www.microsoft.com/en-us/legal/intellectualproperty/copyright/permissions).

Chapter 5. Continuous Random Variables

5.2 The Uniform Distribution

The following activities and questions relate to material covered in Chapter 5.2 The Uniform Distribution (https://openstax.org/books/introductory-business-statistics/pages/5-2-the-uniform-distribution) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Questions

1. The White Pass & Yukon Route Railway is a well-known tourist attraction bringing people between Skagway Alaska and Carcross Yukon. The time that the train takes is a uniformly distributed random variable that lies between 2.4 hours and 2.6 hours to complete a round trip. Imagine that you are in Skagway and have just missed the first train of the day and decide to wait for the next one.[1] What is the probability that you will wait fewer than 150 minutes?
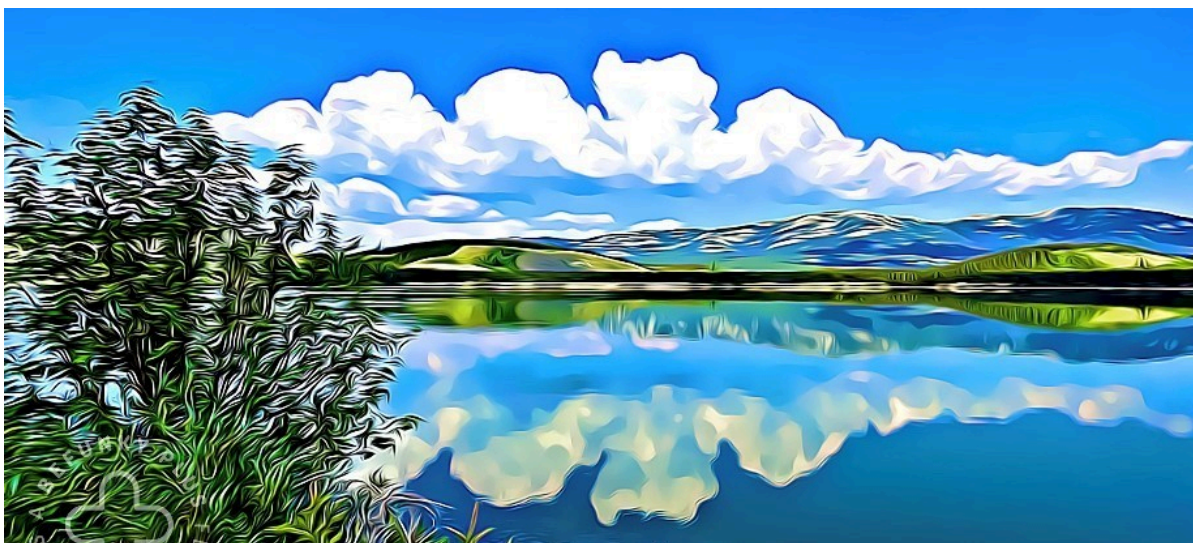
Carcross

Solutions

1. Let X = the number of minutes you must wait for the train
   a = 2.4 hours = 144 minutes
   b = 2.6 hours = 156 minutes

---

1. Data source: ALASKA.ORG. (n.d.). *White pass & Yukon route railroad*. https://www.alaska.org/detail/white-pass-yukon-route-railroad

$$P(135 < x < 155\ ) = \frac{(150 - 144)}{(156 - 144)} = \frac{6}{12} = 50\%$$

There is a 50% chance that you will wait fewer than 150 minutes for the next train.

Media Attributions

- Carcross by Lisa Kanary is licensed under a CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).

Chapter 6. The Normal Distribution

6.1 The Standard Normal Distribution

The following activities and questions relate to material covered in Chapter 6.1 The Standard Normal Distribution (https://openstax.org/books/introductory-business-statistics/pages/6-1-the-standard-normal-distribution) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Questions

1. A Yukon Rent Survey in October 2020 showed that the average unit rent was $1315, per month. Assuming the rental standard deviation σ = $ 150. The rental price is normally distributed. X ~ N (1315, 150). [1] Calculate the z-score for a rental price of $1000 and interpret it using a complete sentence.

Solutions

1. $$Z = \frac{X - \mu}{\sigma} > \frac{1000 - 1315}{150} = -2.1$$
   The rental price of $1000 is 2.1 standard deviations below the mean of $1315, per month.

1. Data source: Yukon Bureau of Statistics. (October, 2020). *Yukon rent survey October 2020*. https://yukon.ca/sites/yukon.ca/files/ybs/rent_oct20.pdf.

6.2 Using the Normal Distribution

The following activities and questions relate to material covered in Chapter 6.2 Using the Normal Distribution (https://openstax.org/books/introductory-business-statistics/pages/6-2-using-the-normal-distribution) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Questions

1. The average salary for a baker in Whitehorse, YT, in the year 2020, was $35,360, approximately $22/hour. Assume that the population of baker salaries in Whitehorse is normally distributed with a population mean of $35,360 and a population standard deviation of $1,500. [1] What is the probability of a randomly selected Whitehorse baker making more than $38,000 a year?

Solutions

1. $$P(X > \$38000) = P\left(\frac{X - \mu}{\sigma} > \frac{38000 - 35360}{1500}\right) = 1 - P(Z < 1.76) = 3.92$$

   The probability of a randomly selected Whitehorse baker making more than $38,000 a year is 3.92%.

---

1. Data source: Canadian Visa | Canadian Immigration Consultants. (n.d.). *Top jobs in Yukon Canada*. https://canadianvisa.org/blog/jobs/top-jobs-in-yukon-canada.

The Normal Distribution - Excel Tools Instruction

Download The Normal Distribution – Excel Instructions Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2022/10/The-Normal-Distribution-–-Excel-Instructions-Data-Sets.xlsx).

Normal Distribution

- The **normal distribution**, which is continuous, is the most important of all the probability distributions. Its graph is bellshaped. This bell-shaped curve is used in almost all disciplines. Since it is a continuous distribution, the total area under the curve is one. The parameters of the normal are the mean μ and the standard deviation σ. A special normal distribution, called the standard normal distribution is the distribution of z-scores. Its mean is zero, and its standard deviation is one.

Normal Distribution in Excel

The final exam scores in a statistics class were normally distributed with a mean of 63 and a standard deviation of five.

    a. Find the probability that a randomly selected student scored more than 65 on the exam.

    b. Find the probability that a randomly selected student scored less than 85.

1. Find out what is x, mean, standard deviation for question a and b.
2. For question a, to compute cumulative normal probabilities P(X>65)
3. Enter P(X>65) in cell A2, enter=1-NORMDIST(65,63,5,TRUE) in cell B2, click Enter
4. For question b, to compute cumulative normal probabilities P(X<85)
5. Enter P(X<85) in cell A4, enter=NORMDIST(85,63,5,TRUE) in cell B4, click Enter

Step 1 & 2: Find out what is x, mean, standard deviation for question a. Compute cumulative normal probabilities P(X>65)

▲	A	B
1		
2	P(X>65)	=1-NORMDIST(65,63,5,TRUE)

Step 3: Enter P(X>65) in cell A2,
enter=1-NORMDIST(65,63,5,TRUE) in cell B2, click Enter

▲	A	B	C	D
1				
2	P(X>65)	0.3446		
3				
4	P(X<85)	=NORMDIST(		
5		NORMDIST(x, mean, standard_dev, cumulative)		

Step 4: Find out what is x, mean, standard deviation for
question b. Compute cumulative normal probabilities
P(X<85)

▲	A	B	C
1			
2	P(X>65)	0.3446	
3			
4	P(X<85)	=NORMDIST(85,63,5,TRUE)	

Step 5: Enter P(X<85) in cell A4,
enter=NORMDIST(85,63,5,TRUE) in cell B4, click Enter

For a reference to calculate a different value with typing a different formula

	A	B	C
1	To calculate a different value, type a different formula		
2			
3	Value	Answer	Formula
4	P(X>65)	0.3446	=1-NORMDIST(65, 63, 5, TRUE)
5			
6	P(X<85)	1.0000	=NORMDIST(85, 63, 5, TRUE)
7			
8	P(55 < X < 85)	0.9452	=NORMDIST(85, 63, 5, TRUE) -NORMDIST(55, 63, 5, TRUE)
9			
10	P(Z < -2.16)	0.02	=NORM.S.DIST(-2.16, TRUE)
11			
12	P(Z > -1.24)	0.89	=1-NORM.S.DIST(-1.24, TRUE)
13			
14	$Z_{0.03}$	1.88	=NORMSINV(1-0.03)
15			
16	P(Z<-z0.02)	4689	=NORMINV(0.02, 5100, 200)
17			

Media Attributions

- Screenshots of Excel are used with permission from Microsoft (https://www.microsoft.com/en-us/legal/intellectualproperty/copyright/permissions).

Chapter 7. The Central Limit Theorem

7.3 The Central Limit Theorem for Proportions

The following activities and questions relate to material covered in Chapter 7.3 The Central Limit Theorem for Proportions (https://openstax.org/books/introductory-business-statistics/pages/7-3-the-central-limit-theorem-for-proportions) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Questions

1. A Yukon-based farm usually butchers chicken in mid-July and mid-August. If a chicken weighs less than 6 pounds, the farmer will keep them. The farmer weighed and butchered 220 chickens last year. One in forty chickens was not fit to keep. What is the standard deviation of the mean of the sampling distribution of sample proportions?

Solutions

1. $$\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\frac{1}{40}\left(1 - \frac{1}{40}\right)}{220}} = 0.0105$$

   The standard deviation of the mean of the sampling distribution of sample proportions is 0.0105.

7.4 Finite Population Correction Factor

The following activities and questions relate to material covered in Chapter 7.4 Finite Population Correction Factor (https://openstax.org/books/introductory-business-statistics/pages/7-4-finite-population-correction-factor) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 7.4 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/7.4-Data-Sets.xlsx).

Questions

1. A local territorial brewery would like to know if the morning shift and the afternoon shift are equally efficient at brewing beer. The average brewing time is normally distributed with a mean of 2.38 hours days and a standard deviation of 0.36 hours. A study is completed to sample 10 shifts from the morning, and 13 shifts from the afternoon. The data was collected as below:

Shift	Number of Hours to Process 100 Bottles of Beer
Morning	2
Morning	2.2
Morning	2.3
Morning	2
Morning	3
Morning	2.6
Morning	2.2
Morning	2.6
Morning	2
Morning	2.6

Shift	Number of Hours to Process 100 Bottles of Beer
Afternoon	2.4
Afternoon	2.4
Afternoon	2.1
Afternoon	2.8
Afternoon	3.4
Afternoon	2.2
Afternoon	2.4
Afternoon	2
Afternoon	2
Afternoon	2.8
Afternoon	2.2
Afternoon	2
Afternoon	2.6

Complete the chart below and write a sentence to assess whether the morning or afternoon shift is more efficient at brewing beer.

Shift	Sample Size	Average Number of Hours to Process 100 Bottles of Beer	Sample Standard Deviation	z-score
Morning Shift				
Afternoon Shift				

Solutions

1.

Shift	Sample Size	Average Number of Hours to Process 100 Bottles of Beer	Sample Standard Deviation	z-score
Morning Shift	10	2.35	0.34	0.28
Afternoon Shift	13	2.41	0.41	−0.24

Because the afternoon shift has a negative z-score (meaning they take less than the average time), the afternoon shift is more efficient at brewing beer compared to the morning shift.

Chapter 8. Confidence Intervals

8.1 A Confidence Interval for a Population Standard Deviation, Known or Large Sample Size

The following activities and questions relate to material covered in Chapter 8.1 A Confidence Interval for a Population Standard Deviation, Known or Large Sample Size (https://openstax.org/books/introductory-business-statistics/pages/8-1-a-confidence-interval-for-a-population-standard-deviation-known-or-large-sample-size) in *Introductory Business Statistics (OpenStax)* (https://openstax.org/details/books/introductory-business-statistics).

Data sets for the following questions are available in Excel: 8.1 Data Sets [XLSX] (https://opentextbc.ca/busstatsancillary/wp-content/uploads/sites/426/2023/01/8.1-Data-Sets.xlsx).

Questions

1. The Yukon Bureau of Statistics tracked the self-serve fuel price data in Whitehorse from May 2018 to May 19, 2021, (see table below). Assume that the fuel price population distribution of fuel price is normal, and the population standard deviation is known to be $6.50. Construct a 90% confidence interval for the population mean of the fuel price. [1] State the confidence interval.

1. Data source: Yukon Bureau of Statistics. (2021). *Yukon fuel price survey May 2021*. https://yukon.ca/sites/yukon.ca/files/ybs/fuel_may21.pdf

Whitehorse, Yukon Fuel Price 2018-2019

Date	Fuel Price
May' 2018	141.6
Jun' 2018	144.0
Jul' 2018	143.9
Aug' 2018	143.7
Sep' 2018	143.5
Oct' 2018	143.6
Nov' 2018	141.9
Dec' 2018	124.0
Jan' 2019	124.5
Feb' 2019	124.6
Mar' 2019	131.5

Whitehorse, Yukon Fuel Price 2019-2020

Date	Whitehorse
Apr' 2019	137.5
May' 2019	137.4
Jun' 2019	137.1
Jul' 2019	140.9
Aug' 2019	141.4
Sep' 2019	142.0
Oct' 2019	141.8
Nov' 2019	142.2
Dec' 2019	142.1
Jan' 2020	142.1
Feb' 2020	142.0

2. Monthly sales of the motor vehicle and parts dealers in Yukon in 2020 are list in the table below ($1,000). Assume that the population distribution of monthly sales is normal, and the population standard deviation is known to be 2500 in $1,000. Construct a 95% confidence interval for the population mean of monthly sales. [2] State the confidence interval.

Monthly motor vehicle and parts sales in Yukon, 2020

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
10136	10875	12094	9614	15156	16032	17898	15011	13285	15706	13110	12672

Solutions

1. $\bar{x} \pm z_{\frac{a}{2}} \dfrac{\sigma}{\sqrt{n}} = 138.80 \pm 1.645 \dfrac{6.50}{\sqrt{22}} = 138.80 \pm 2.28$

   LCL = 136.52
   UCL = 141.08
   The population mean of fuel price in Whitehorse from May 2018 to May 19, 2021 is between (LCL) $136.52 and (UCL) $141.08 with 90% confidence.

2. Sample mean:

---

2. Data source: Yukon Bureau of Statistics. (2020). *Yukon retail sales 2020*. https://yukon.ca/sites/yukon.ca/files/ybs/retail2020.pdf

$$\bar{x} = \frac{10136+10875+12094+9614+15156+16032+17898+15011+13285+15706+13110+12672}{12}$$

$$= 13465.75$$

$$\bar{x} \pm z_{\frac{a}{2}} \frac{\sigma}{\sqrt{n}} = 13465.75 \pm 1.96 \frac{2500}{\sqrt{12}} = 13456.75 \pm 1414.51$$

LCL = $12,051,240

UCL = $14, 880,260

In 2020, the population mean of monthly sales of the motor vehicles and parts dealers in Yukon is between (LCL) $12,051,240 and (UCL) $14, 880,260 with 95% confidence.

Chapter 13. Linear Regression and Correlation

# Linear Regression and Correlation - Excel Tools Instruction

Scatter Plot

A **scatter plot graph** presents the data set where it is hypothesized that Y is dependent upon the single independent variable X.
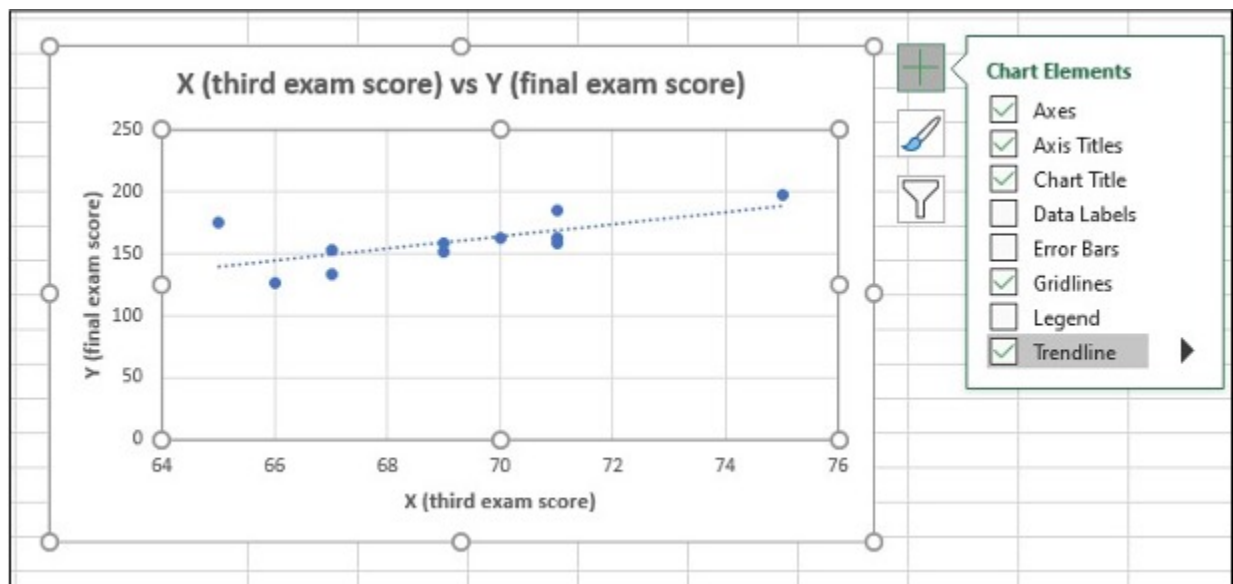
Scatter Plot in Excel

1. Highlight columns of cells
2. Click Insert, Scatter and the one with only Markers
3. Click on the new Scatter plot to change Title, and click icon + to make other changes

x (third exam score)	y (final exam score)
65	175
67	133
71	185
71	163
66	126
75	198
67	153
70	163
71	159
69	151
69	159
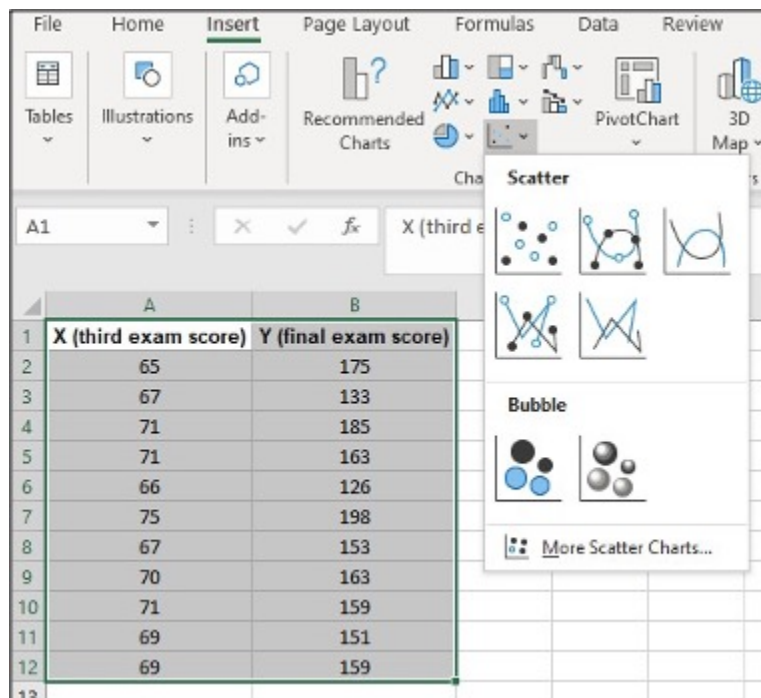
Step 1 & 2: Highlight columns of cells. Click Insert, Scatter and the one with only Markers

Step 3: Click on the new Scatter plot to change Title, and click icon + to make other changes

Media Attributions

- Screenshots of Excel are used with permission from Microsoft (https://www.microsoft.com/en-us/legal/intellectualproperty/copyright/permissions).

Versioning History

This page provides a record of edits and changes made to this book since its initial publication. Whenever edits or updates are made in the text, we provide a record and description of those changes here. If the change is minor, the version number increases by 0.01. If the edits involve substantial updates, the version number increases to the next full number.

The files posted by this book always reflect the most recent version. If you find an error in this book, please fill out the Report an Error (https://open.bccampus.ca/browse-our-collection/reporting-an-error/) form.

Version	Date	Change	Details
1.00	March 29, 2023	Book published.	