ANLT 600 Manual May 2021

Data Analytics and Business Intelligence (Saskatchewan Polytechnic)

BUSINESS MATHEMATICS AND DATA ANALYTICS
(ANLT 600)

Saskatchewan Polytechnic
Moose Jaw, Saskatchewan

By

Sophie (Yingjie) Wei

May 2021

Contents

Chapter 1

Introduction of Business Analytics

1.1 Definitions

Analytics can be defined as a process that involves the use of statistical techniques (measures of central tendency, graphs, and so on), information system software (data mining, sorting routines), and operations research methodologies (linear programming) to explore, visualize, discover and communicate patterns or trends in data. Simply, analytics converts data into useful information. Analytics is an older term commonly applied to all disciplines, not just business. There are three types of analytics: descriptive, predictive and prescriptive.

Descriptive analytics is to identify possible trends in large data sets and to get a rough picture of what generally the data look like and what criteria might have potential for identifying trends or future business behavior. Descriptive statistics is often used as a major methodology. It includes measure of central tendency (mean, median, mode), measures of dispersion (variance or standard deviation), charts, graphs, sorting methods, frequency distributions, probability distributions, and sampling methods.

Predictive analytics is to build predictive models designed to identify and predict future trends. Operations research methods like forecasting models are also applied. Statistical methods like linear regression and ANOVA are applied. Information system methods like data mining and sorting are also used.

Prescriptive analytics is to allocate resources optimally to take advantage of predicted trends or future opportunities. Operations research methodologies like linear programming and decision theory are applied.

Business Analytics (BA) refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning. Business Analytics focuses on developing new insights and understanding of business performance based on data and statistical methods. It is used by companies committed to data-driven decision-making.

Business intelligence (BI) can be defined as a set of processes and technologies that convert data into meaningful and useful information for business purposes.

1.2 Relationship of BA and BI

One major component of BI is storing an organization's data in computer cloud storage or in data warehouses. Data warehousing is not an analytics or business analytics function, although the data can be used for analysis. In application, BI is focused on querying and reporting, but it can include reported information from a BA analysis. BI aims to answer questions such as what is happening now and where, and what business actions are needed based on prior experience. BA, on the other hand, can answer questions like why something is happening, what new trends may exist, what will happen next, and what is the best choice for the future.

In summary, BA includes the same procedures as in plain analytics but has the additional requirement that the outcome of the analytic analysis must make a measurable impact on business performance. BA includes reporting results like BI but seeks to explain why the results occur based on the analysis rather than just reporting and storing the results, as is the case with BI.

1.3 Seven Steps of the Business Analytics Process

There are different ways to define the steps of BA process. I prefer the seven steps theory, which is published by Dr. Hargreaves [2].

Step 1. Defining the business needs.

The first stage in the business analytics process involves understanding what the business would like to improve on or the problem it wants solved. Sometimes, the goal is broken down into smaller goals. Relevant data needed to solve these business goals are decided upon by the business stakeholders, business users with the domain knowledge and the business analyst. At this stage, key questions such as, “what data is available”, “how can we use it”, “do we have sufficient data”, must be answered.

Step 2. Explore the data.

This stage involves cleaning the data, making computations for missing data, removing outliers, and transforming combinations of variables to form new variables. Time series graphs are plotted as they are able to indicate any patterns or outliers. The removal of outliers from the dataset is a very important task as outliers often affect the accuracy of the model if they are allowed to remain in the data set.

Once the data has been cleaned, the analyst will try to make better sense of the data. The analyst will plot the data using scatter plots (to identify possible correlation or non-linearity). The analyst will visually check all possible slices of data and summarize the data using appropriate visualization and descriptive statistics (such as mean, standard deviation, range, mode, median) that will help provide a basic understanding of the data. At this stage, the analyst is already looking for general patterns and actionable insights that can be derived to achieve the business goal.

Step 3. Analyze the data.

At this stage, using statistical analysis methods such as correlation analysis and hypothesis testing, the analyst will find all factors that are related to the target variable. The analyst will also perform simple regression analysis to see whether simple predictions can be made. In addition, different groups are compared using different assumptions and these are tested using hypothesis testing. Often, it is at this stage that the data is cut, sliced and diced, and different comparisons are made while trying to derive actionable insights from the data.

Step 4. Predict what is likely to happen.

Business Analytics is about being proactive in decision making. At this stage, the analyst will model the data using predictive techniques that include decision trees, neural networks and logistic regression. These techniques will uncover insights and patterns that highlight relationships and "hidden evidences" of the most influential variables. The analyst will then compare the predictive values with the actual values and compute the predictive errors. Usually, several predictive models are ran and the best performing model selected based on model accuracy and outcomes.

Step 5. Optimize the model (find the best solution).

At this stage the analyst will apply the predictive model coefficients and outcomes to run what-if scenarios, using targets set by managers to determine the best solution, with the given constraints and limitations. The analyst will select the optimal solution and model based on the lowest error, management targets and his intuitive recognition of the model coefficients that are most aligned to the organization's strategic goal.

Step 6. Make a decision and measure the outcome.

The analyst will then make decisions and take action based on the derived insights from the model and the organizational goals. An appropriate period of time after this action has been taken, the outcome of the action is then measured.

Step 7. Update the system with the results of the decision.

Finally the results of the decision and action and the new insights derived from the model are recorded and updated into the database. Information such as, "was the decision and action effective?", "how did the treatment group compare with the control group?" and "what was the return on investment?", are uploaded into the database. The result is an evolving database that is continuously updated as soon as new insights and knowledge are derived.

Reference:

[1] Dr. Hargreaves, NUS-ISS quarterly e-newsletter, Issue 3 (Jul-Sept 2013).

1.4 Chapter 1 Review Questions:

1. Define the Business Analytics.
2. Describe the seven steps of Business Analytics.

Chapter 2

Simple Interest and Compound Interest

2.1 Interest

A consideration or payment for the use of invested or loaned capital is called **interest**. It is convenient and customary to express both capital and interest in units of money.

Three elements are involved in interest transactions: **principal**, **rate of interest**, and **time**. The principal is the capital originally invested. The rate of interest is the per cent of the principal which is paid as interest for its use for one period of time. The time is the number of periods during which the principal is used. One year is generally taken as the unit period of time. We use **amount** or **sum** to denote the sum of the principal and the interest.

There are two kinds of interest.

Simple interest: If interest is computed on the original principal only, it is called simple interest. A simple interest system primarily applies to short-term financial transactions, normally with a time frame of less than one year.

Compound interest: When principal is increased by interest at the end of each period and interest is thus computed on a principal which grows periodically, we call it compound interest. A compound interest system primarily applies to long-term financial transactions, normally with a time frame of one year or more.

2.2 Formulas for Simple Interest

The simple interest on a given principal at a given rate for a given time is the product of three factors: the principal, the rate of interest per unit of time, and the number of units of time. If we let P denote the principal in dollars, r the annual rate of interest, t the time in years (we always assume 12 months a year or 365 days a year), I the total interest in dollars, and S the amount in dollars at the end of t years, we may write the formulas for simple interest as follows:

$$I = Prt, \tag{2.1}$$

and, by definition, $S = P + I$, it follows that

$$S = P + I = P + Prt = P(1 + rt) \tag{2.2}$$

We will conduct calculations using Excel spreadsheet through the course. **Example 1** You deposit \$5000.00 in a bank account for one year at a rate of 4%. What is the simple interest you have earned? What is the amount?

Solution. $P = 5000.00$, $r = 4\% = 0.04$, $t = 1$ year. Substituting in formula (2.1), we have the interest

$$I = Prt = 5000.00 \times 0.04 \times 1 = \$200.00.$$

And the amount will be

$$S = P + I = 5000.00 + 200.00 = \$5200.00.$$

Example 2 Find the time required for \$4000.00 to yield \$150.00 in simple interest at 5%.

Solution. We know all the quantities except t, so we solve for t by dividing Pr to both sides of $I = Prt$. Therefore

$$t = \frac{I}{Pr} = \frac{150.00}{4000.00 \times 0.05} = 0.75 \ \text{year} = 9 \ \text{months}.$$

We will use the following example to show how to compute the time between two dates.

Example 3 Find the days from April 5 to July 12. We don't count the first day but count the last day.

April	30-5=25	days
May	31	days
June	30	days
July	12	days
Total	98	days

We can also use the built-in function **DAYS(end date, start date)** in excel. You need to format the cells to store the start and end dates as "date".

Exercises

1. Mike borrowed \$1,100.00 from Maria five months ago. When he first borrowed the money, they agreed that he would pay Maria 5% simple interest. If Mike pays her back today, how much interest does he owe her?

2. A \$3,500.00 investment earned \$70.00 of interest over the course of six months. What annual rate of simple interest did the investment earn?

3. What amount of money invested at 6% annual simple interest for 11 months earns \$2,035.00 of interest?

4. For how many months must \$95,000.00 be invested to earn \$1,187.50 of simple interest at an interest rate of 5%?

5. Find the simple interest if \$15,000.00 is invested for 4 months at

a. r=7.25%;

b) r=8%.

6. Find the amount if \$8,000.00 is invested at 12.5% for

a. 7 months;

b. Half year.

7. How much needs to be invested to earn \$1,200.00 in interest in 1 year if:

a. r=9.25%;

b. r=8%.

8. If \$20,000.00 becomes \$21,500.00 in 10 months, find r.

2.3 Present Value at Simple Interest

The **present value**, at the interest rate r, of an amount S due in t years is that principal P, which is invested now at the rate r, will amount to S in t years. Thus the present value, at 10% simple interest, of \$110 due in a year is \$100.00, since \$100.00 will amount to \$110.00 in a year at 10% simple interest.

We can also consider the present value shall be defined as being such a principal that if invested on a specified date at a specified rate will become the future value on the due date.

In general, $S = P(1 + rt)$, hence the present value, at the interest rate r, of an amount S due in t years is

$$P = \frac{S}{1 + rt} = S(1 + rt)^{-1} \tag{2.3}$$

Example 1 Jennifer signs a note that she will pay \$4,000.00 on September 18, 2021 to Monica. Jennifer won a lottery, so she wants to pay Monica sooner. How much should Jennifer pay Monica to discharge the debt on February 5, 2021 if the interest rate is 5%?

Solution There are 225 days between February 5, 2021 to September 18, 2021, and we need to convert days to years by dividing it by 365. Amount to pay is the present value of \$4000.00. Therefore,

$$P = \frac{S}{1+rt} = \frac{4000.00}{1+0.05(\frac{225}{365})} = \$3880.40.$$

Hence, Jennifer will pay \$3880.40 on February 5, 2021 to discharge the debt.

Example 2 Donald buys a car from Bill and signs a note that he will pay Bill \$5000.00 in one year. Five months later, Donald wishes to discharge his obligation. What should he pay if r=8%?

Solution Notice that on the document no interest rate is mentioned. Therefore, the \$5000.00 is a non-interest bearing debt. The 8% interest rate mentioned refers to the banks where anyone with money to invest can earn 8%. The amount that Donald should pay is such a principal that, if invested at 8% for seven months, will become \$5000.00.

$$P = \frac{S}{1+rt} = \frac{5000.00}{1+0.08(\frac{7}{12})} = \$4777.07.$$

\$4777.07 if invested by Donald or bill at 8% for 7 months will become \$5000.00.

Exercises

1. Joey signs a note that he will pay Ross $3,000.00 in 10 months. Six months later, Joey wishes to discharge his obligation. What should he pay if r=12%?

2. A retailer store signs a note on July 10 that $20,000.00 will be paid to the wholesaler in 4 months. What amount should he pay on September 15 to discharge the debt if the interest rate r=8%.

3. Jim has a bond for $2,500.00 that will mature on August 10, 2016. He sells the bond to John and John expects of 15% on his investment. What should John pay for the bond if the transaction is completed on June 6, 2016.

2.4 Present Value of Interest-Bearing Debt

The cases discussed in Section 2.3 involve debts that do not bear interest. There are cases, however, in which the debt bears interest, and the maturity value of the debt is different from the face value. The situation may be made clear by using an illustration. Suppose A borrows $100.00 from B, promising to pay it in a year with 5% simple interest. The face value of A's debt is of $100.00, while the maturity value of his debt is $105.00. However, if A has a debt of $100.00, due in a year with no interest, the face value and the maturity value of his debt are identical, i.e $100.00.

The present value (at a given interest rate) of a debt bearing the same rate of interest is the face value of the debt. Thus, the present value at 8% of a debt of $100.00 due in a year and bearing interest at 8% is $100.00. However, in case the interest rate used in finding the present value is different from the rate of interest borne by the debt, two steps are necessary in solving the problem:

Step 1. Find the maturity value of the debt, by adding the interest for the period to the face value of the debt.

Step 2. Find the present value of the maturity value of the debt, using the interest rate given for determining the present value.

Example 1 A retailer owes \$15,000.00 to a wholesaler, his debt maturing in 4 months and bearing 8.5% interest. Find the present value of the retailer's debt at the rate of 6%.

Solution

$$M.V. = 15,000.00\left[1+0.085\left(\frac{4}{12}\right)\right] = \$15,425.00,$$

$$P.V. = \frac{M.V.}{1+rt} = \frac{15,425.00}{1+0.06\left(\frac{4}{12}\right)} = \$15,122.55.$$

Example 2 Donald buys a car from Bill and signs a note that he will pay Bill \$5000.00 plus interest at 20% in one year. Seven months later, Donald wishes to discharge his obligation. What should he pay if r=8%?

Solution Once Donald has signed the note, he is contractually obliged to pay a maturity value in one year:

$$\text{Maturity Value (M.V.)} = 5000[1+0.2(1)] = \$6000.00.$$

Further, the contract has the effect of forcing Donald to pay 20% interest for the full year.

The P.V. we need to calculate is the principal that, if invested at 7 months on the time line at 8% interest, will grow to become \$6000.00 in the remaining 5 months.

$$P\left[1+0.08(\frac{5}{12})\right]=\$6000.00,$$

Hence, Donald should pay $P=\$5806.45$ to discharge the debt.

Example 3 Betty's Grocery accepts merchandise from a wholesaler and signs a note that they will pay \$12500.00 with interest at 24% in 90 days. Thirty days later the grocery store wishes to discharge the obligation. What should they pay if r=10%?

Solution

$$MV=12500.00\left[1+0.24(\frac{90}{365})\right]=\$13239.73$$

$$P=\frac{13239.73}{1+0.1(\frac{60}{365})}=\$13025.61$$

Example 4 Jason buys a TV set from Best Buy and signs that he will pay \$850.00 with interest at 25% in one year. Best Buy immediately sells the note to the Household Finance Corporation who buys it expecting a rate of return of 35% on their investment. Find the proceeds to Best Buy.

Solution

$$MV=850.00\left[1+0.25(1)\right]=\$1062.50.$$

$$P=\frac{1062.50}{1+0.35(1)}=\$787.04.$$

Exercises

1. Terry signs a note that he will pay Dale \$6,000.00 with interest at 6% in 6 months. What should Terry pay to discharge the debt one month before the debt due if the interest is 5%?

2. Five months before the maturity date of a 10 month, 15% interest bearing note with a face value of \$5,000.00, it was sold for \$5,000.00. Find the rate of return expected by the buyer.

3. Bruce signs a note to Bill that he will pay \$5,500.00 plus interest of 14.5% in 8 months. Bill sells the note immediately to John who wishes to earn 20% on their investment. What should Bill get from John?

4. A 180-day note for \$20,000.00 with interest at 8% dated May 15, 2016. Compute the value of the note on September 5, 2016 if money is worth 6%.

2.5 Compound Amount and Compound Interest

In transaction involving simple interest, the principal on which the interest is computed remains unchanged throughout the term of the loan, and the interest becomes due either at the end of the term or at the end of stated intervals during the term of the loan. If not paid when due, the interest becomes simply a non-interest-bearing debt.

In transactions of other types, it is mutually agreed that the interest for each period, instead of being paid when due, shall be added to and become part of the principal. In other words, interest is computed upon a principal which increases periodically. Consequently, the interest for a given period is more than that of the preceding period. When interest is thus added to the principal at the end of each period, it is said to be **converted into principal**, or **compounded**, or **payable**. The total amount due at the end of the last period is called the **compound amount**. The difference between the compound amount and the original principal is called the **compound interest**.

2.6 Conversion Period and Rate per Period

The accumulation or growth of a principal at compound interest depends upon two elements: (a) the number of **conversion periods** over which the investment extends, and (b) **the interest rate per conversion period.**

Interest may be converted into principal annually, semiannually, quarterly, monthly, or at any other regular periods of time. The **frequency of conversion** is a number indicating how many times interest is compounded in one year. The time between two successive conversions of interest is called the **conversion period**, or **interest period.** For example, if interest is compounded quarterly, the frequency of conversion is 4 and the conversion period is 3 months.

Regardless of the frequency of conversion, the rate of interest is usually expressed as an annual rate. When the conversion period is other than a year, the rate per conversion period is found by dividing the stated annual rate by the number of conversion periods in a year. For example, if the quoted rate is 6% compounded semiannually, the rate per conversion period is 3% (6%/2), or 0.03.

The time over which the investment extends is usually expressed in years or years and months. It is necessary to change the given time into conversion periods. Thus, if interest is compounded quarterly and the investment runs over $3\frac{1}{2}$ years, the number of conversion periods is $(3\frac{1}{2})$ multiplied by 4, i.e. 14.

2.7 The Meaning of Certain Expressions

To avoid confusion, we should pay attention to the meaning of certain expressions frequently used in succeeding discussions and problems.

(a) Unless simple interest is specified, the word "interest" will hereafter mean *compound interest*, and "amount" will mean *compound amount*.

(b) 10% c.a. = 10% compounded annually:

i) interest is to be calculated yearly;

ii) the rate of interest per year is 10%.

(c) 10% c.s.a. = 10% compounded semi-annually:

i) two interest calculations per year or two conversions per year;

ii) the rate of interest per half year is $\frac{10\%}{2} = 5\%$.

(d) 10% c.q. = 10% compounded quarterly:

i) four interest calculations per year or four conversions per year;

ii) the rate of interest per quarter year is $\frac{10\%}{4} = 2.5\%$.

(e) 15% c.m. = 15% compounded monthly:

i) 12 interest calculations per year or 12 conversions per year;

ii) the rate of interest per month is $\frac{15\%}{12} = 1.25\%$.

2.8 Formula for Compound Amount

If we are to calculate the amount of $1,000.00 invested at 10% c.a. for 3 years, we may take the following steps:

Amount in the account at the end of the first year:

$$1,000.00(1 + 0.1) = 1,000.00(1.1) = \$1,100.00.$$

Amount in the account at the end of the second year:

$$1,100.00(1 + 0.1) = 1,100.00(1.1) = \$1,210.00.$$

Amount in the account at the end of the third year:

$$1,210.00(1 + 0.1) = 1,100.00(1.1) = \$1,331.00.$$

We can also write the procedure by writing:

$$1,000.00(1 + 0.1)(1 + 0.1)(1 + 0.1) = 1,000.00(1 + 0.1)^3 = \$1,331.00.$$

This leads to the **compound amount formula**:

$$S = P(1+i)^n, \tag{2.4}$$

where

S = compound amount = sum of principal and compound interest = future value (F.V.);

P = principal = present value (P.V.);

n = total number of conversion periods;

i = rate of interest per conversion period;

$1+i$ = ratio of increase;

$(1+i)^n$ = **accumulation factor** or **compound amount of 1**.

Example 1 Find the compound amount if \$6000.00 is invested for 8 years at

a. 15% c.a.

b. 15% c.s.a.

c. 15% c.q.

d. 15% c.m.

Solutions

a.

$$S = 6000.00(1+\frac{0.15}{1})^8 = \$18354.14$$

b.

$$S = 6000.00(1+\frac{0.15}{2})^{8\times 2} = \$19084.76$$

c.

$$S = 6000.00(1 + \frac{0.15}{4})^{8\times 4} = \$19488.15$$

d.

$$S = 6000.00(1 + \frac{0.15}{12})^{8\times 12} = \$19773.08$$

Example 2 How much needs to be invested now in order to have $10000.00 in five years if the rate of interest is 9% c.q.

Solution In this problem, we are given the F.V. or S and it is the principal that must be found.

Solution

$$100000.00 = P\left(1 + \frac{0.09}{4}\right)^{5\times 4}$$

$$P = \frac{10000.00}{\left(1 + \frac{0.09}{4}\right)^{5\times 4}} = \$6408.16$$

Exercises

1. Find the:

a. interest rate per conversion period;

b. conversion period;

c. frequency of conversion;

d. total number of conversion periods;

for each of the following questions:

1) $5,000.00 invested 1.5 years at 8% c.m..

2) $25,000.00 invested 2 years at 8.75% c.q..

3) \$52,000.00 invested 2 1/4 years at 9% c.s.a..

4) \$4,000.00 invested 4 years at 11% c.a..

2. Find the compound interest and compound amount for the Question 1.

3. How much needs to be invested now in order to have \$5,000.00 in five years if the interest rate is 6% c.s.a.?

4. Mr. & Mrs. Smith invested \$2,000.00 on the date of their daughter's birth in order to provide her \$10,000.00 on her 18th birthday. On the date of her 10th birthday, they were informed by the bank that more money had to be invested. How much more should they invest on the date of her 10th birthday if r=8% c.a.?

2.9 Compound Amount at Changing Rate

We have assumed a constant rate of interest for the entire duration of an investment of compound amount formula. However, interest rates may change from time to time.

Example 1 Jim invested \$10,000.00. The investment earns 3% c.q. for the first 5 years and 4% c.s.a. for the next 8 years. Find the compound amount at the end of 13 years and the compound interest.

Solution. For the first 5 years,

$$S_1 = 10,000.00\left(1+\frac{0.03}{4}\right)^{4\times 5} = \$11,611.84.$$

This compound amount is reinvested.

$$S_2 = 11,611.84\left(1+\frac{0.04}{2}\right)^{2\times 8} = \$15,940.57.$$

This is the compound amount at the end of 13 years.

The compound interest will be:

$$I = 15,940.57 - 10,000.00 = \$5940.57.$$

Example 2 Jared invested \$20000.00 at 9% c.s.a. for two years. He decided to lock in the money for another 3 years at the best interest rate 10% c.q. at the maturity date. He continued to lock the money in for another 4 years at 13% c.m. at the maturity date. What the compound amount and compound interest?

Solution Jared invested the money for 9 years. The amount of first two years investment:

$$S_1 = 20000.00\left(1+\frac{0.09}{2}\right)^{2\times 2} = \$23850.37.$$

This compound amount is reinvested.

$$S_2 = 23850.37\left(1+\frac{0.10}{4}\right)^{3\times 4} = \$32076.10.$$

This compound amount is reinvested.

$$S_3 = 32076.10\left(1+\frac{0.13}{12}\right)^{4\times 12} = \$53802.22.$$

This is the compound amount after 9 years investment.

The compound interest is:

$$I = 53802.22 - 20000.00 = \$33802.22.$$

The calculation of the compound amount could have been compressed as follows:

$$S = 20000.00\left(1+\frac{0.09}{2}\right)^{2\times 2}\left(1+\frac{0.10}{4}\right)^{3\times 4}\left(1+\frac{0.13}{12}\right)^{4\times 12} = \$53802.22.$$

2.10 Interest for Fractional Parts of Conversion Periods

In defining compound amount and deriving a formula for it, we assumed an integral number of conversion periods in the time. When it is desired to find the compound amount for a time involving a fractional part of a conversion period, we generally use the following two methods:

(a) Compute the compound amount by $S = P(1+i)^n$ at the end of the last whole period contained in the given length of time; then, add to this amount the simple interest on it for the remaining fraction of a period.

(b) Change the time to an improper fraction and multiplied by the number of conversion periods per year. This method is more convenient that (a), we will adopt this method in this course.

Example 1 Mrs. Smith invested \$18,000.00 at 9% c.s.a. for $1\frac{2}{3}$ years. Find the compound interest.

Solution

$$S = 18,000.00\left(1+\frac{0.09}{2}\right)^{2\times\frac{5}{3}} = \$20,844.60.$$

$$\text{Compound Interest} \quad I = 20,844.60 - 18,000.00 = \$2844.60.$$

Notice that the time was changed to an improper fraction and multiplied by the number of conversion periods per year.

Example 2 Mr. Barber would like to have $25,000.00 in his account in 28 months. How much must he invest now if the rate is 5% c.s.a.?

Solution

$$P = \frac{25000.00}{\left(1+\frac{0.05}{2}\right)^{2\times\frac{28}{12}}} = \$22,278.98$$

Exercises

1. Find the compound amount of an investment of $11,500.00 for 5 years at 6% c.m and then continued for 5 years at 8% c.q..

2. Find the total interest earned if $5,000.00 is invested for four and half years at 7% c.a..

3. Find the compound amount if $35,000.00 is invested for

a. Three years at 9% c.m., and

b. Three years and three months at 10% c.s.a.

4. Jim invests $20,000.00. He earns 9% c.s.a. for 3 years, 13.5% c.a. for 1.5 years, and then 10% c.q. for 3.5 years. Find the total compound interest he earns.

5. If three investors invested \$100,000.00 separately with different combination of return, find that who did the best in the investment, and find the dollar difference between the best and worst performance (compare F.V.s).

Investor A earned 6% c.s.a. for 2 years, 10% c.q. for 3 years, and 15% c.m. for 2 years.

Investor B earned 8% c.s.a. for 3 years, 12% c.m. for 3 years, and 14% c.m. for 1 years.

Investor C earned 6% c.a. for 2 years, 8% c.s.a. for 3 years, and 13% c.q. for 2 years.

2.11 Determining the Present Value

In finding the present value of a long-term obligation, it is customary to use a compound interest rate. If money is worth i per period, the present value of S due in n periods is that principal which, invested *now* at the rate i per period, will amount to S in n periods. The word "now" refers to n periods before S is due.

We have learned that a principal P will amount to $S = P(1+i)^n$ in n periods at the rate i per period. Hence, P is the present value of S. Solving the formula for P by dividing both sides by $(1+i)^n$,

$$P = \frac{S}{(1+i)^n} = S(1+i)^{-n} \tag{2.5}$$

It should be noted that P and S represent the value of the same obligation at different dates. P is the present value of a given obligation, while S is the future value of the same obligation.

To find the present value of an interest-bearing debt, take the following two steps:

Step 1: Find the amount of the debt at maturity.

Step 2: Determine the present value of the compound amount found in step 1.

Example 1 A debt of \$4,500.00, bearing 5% c.s.a., is due in 6 years. If money is worth $4\frac{1}{2}\%$ c.a., find the present value of the debt.

Solution. Step 1: By formula, the maturity value of the debt is

$$S = 4,500.00 \left(1 + \frac{0.05}{2}\right)^{2\times 6}$$

$$= \$6,052.00.$$

Step 2: By formula, the present value of this amount is

$$P = 6,052.00(1 + 0.045)^{-6}$$

$$= \$4,647.31.$$

Hence, the present value of the debt is \$4647.31.

Example 2 If we desire \$1.00 in 4 years and the interest rate is 9% c.s.a., how much do we need to invest now?

Solution

$$1.00 = P\left(1 + \frac{0.09}{2}\right)^{8}$$

$$P = \frac{1}{(1 + \frac{0.09}{2})^{8}} = \$0.703185$$

If we desire \$10,000.00 in 4 years and the interest rate is 9% c.s.a., we would have to multiply the previously found \$0.703185 by 10000 to get \$7031.85.

Example 3 Investor Amanda buys an investment certificate for \$15,000.00 bearing interest at 12% c.s.a. due principal and interest in 10 years. Amanda sells the certificate to Brenda after one year during which interest rate expectations dropped down to 8% c.s.a.. Brenda sells two years later to Carmela who expected to earn 5% c.s.a. Find all selling prices.

Solution Amanda purchases the certificate for \$15,000.00 expecting to collect a maturity value in 10 years:

$$M.V. = 15000.00\left(1+\frac{0.12}{2}\right)^{2\times 10} = \$48,107.03$$

Only 1 year later she sells to Brenda. Brenda will pay such an amount that, if invested at 8% c.s.a., will accumulate to \$48,107.03 in 9 years:

$$P = 48,107.03\left(1+\frac{0.08}{2}\right)^{-2\times 9} = \$23,746.98$$

Brenda holds the certificate for 2 years and then sells to Carmela. Carmela will pay such an amount that, if invested at 5% c.s.a., will accumulate to \$48,107.03 in 7 years:

$$P = 48,107.03\left(1+\frac{0.05}{2}\right)^{-2\times 7} = \$34,046.65.$$

Exercises

1. Bob needs \$8,000.00 to change a new furnace in 2 and half years. How much should he invest if the interest rate is 8.5% c.m.?

2. Adam holds an investment bond for $20,000.00 bearing interest at 10% c.q. due principal and interest in 8 years. He sells the bond to Bill after two years, and the interest rate is expected as 11.5% c.q.. Bill sells it to Cory one year later and Cory expects to earn 12.9% c.q.. One and half years later, Cory sells it to Don who expects to earn 14.5% c.q.. Find all the selling prices.

3. How much needs to be invested now to mature to $18,000.00 in 2 and half years at 8% c.s.a.?

4. Find the present value of $5,000.00 due in 32 months if money is worth 12% c.s.a..

5. Amy signs a note of $6,500.00 with interest rate 12% c.s.a due in two and half years to Beth. Beth sells the note to Celine one year later, and the interest rate is expected as 9% c.q.. Find the selling price to Celine.

6. Mark received compound amount of $32,150.88 after investing his money for 3 years and 8 months with 15% c.a.. Find how much he invested.

2.12 Chapter 2 Review Questions

1. Mark borrowed $12,500.00 from Mary six months ago. When he first borrowed the money, they agreed that he would pay Mary 4.5% simple interest. If Mark pays her back today, how much interest does he owe her?

2. A $5,500.00 investment earned $170.00 of interest over the course of eight

months. What annual rate of simple interest did the investment earn?

3. What amount of money invested at 5.6% annual simple interest for 10 months earns \$1,050.00 of interest?

4. For how long must \$15,000.00 be invested to earn \$187.50 of simple interest at an interest rate of 5%?

5. Tom signs a note that he will pay Roger \$5,000.00 in eight months. Five months later, Tom wishes to discharge his obligation. What should he pay if r=11.5%?

6. A convenient store signs a note on August 2 that \$15,000.00 will be paid to the wholesaler in 3 months. What amount should he pay on September 30 to discharge the debt if the interest rate r=8%.

7. Jack has a bond for \$12,000.00 that will mature on September 1, 2016. He sells the bond to John and John expects of 10% on his investment. What should John pay for the bond if the transaction is completed on May 30, 2016.

8. George signs a note to John that he will pay \$45,000.00 plus interest of 12.5% in 10 months. John sells the note immediately to Bob who wishes to earn 15% on the investment. What amount should John get from Bob?

9. A 150-day note for \$10,000.00 with interest at 8.5% dated May 1, 2018. Compute the value of the note on September 6, 2018 if money is worth 9.6%.

The following questions are based on compound interest:

10. How much needs to be invested now in order to have \$50,000.00 in five years

if the interest rate is 6.5% c.s.a.?

11. Dr. Brown invested \$5,000.00 on the date of their daughter's birth in order to provide her \$20,000.00 on her 18th birthday for education expense. On the date of her 8th birthday, they were informed by the bank that the interest rate is lower than they expected and more money had to be invested. How much more should they invest on the date of her 8th birthday if E.R.=7% c.a.?

12. Jobs invests \$10,000.00. He earns 5% c.q. for 3 years, 7.5% c.a. for 2 years, and then 10% c.s.a. for 3.5 years. Find the total compound interest he earns.

13. Bill received compound amount of \$52,150.00 after investing his money for 3 years and 6 months at 10% c.q.. Find how much he invested.

14. Meg signs a note of \$5,000.00 with interest rate 10.5% c.a due in two and half years to Betty. Betty sells the note to Mary two year later, and the interest rate is expected as 11% c.s.a.. Find the selling price to Mary.

15. A investment received \$50,450.00 after 3 years and eight months at 7.5% c.q.. How much has been invested originally?

Chapter 3

Cost-Benefit Analysis and Return On Investment (ROI)

3.1 Cost-Revenue-Net Income Analysis

3.1.1 Types of Costs

A cost is an outlay of money required to produce, acquire, or maintain a product, which includes both physical goods and services. Costs can come in three forms:

1. A **fixed cost** is a cost that does not change with the level of production or sales (call this "output" for short). In other words, whether the business outputs nothing or outputs 10,000 units, these costs remain the same. For example, rent, insurance, property taxes, salaries unrelated to production (such as administration and management), production equipment, office furniture, and much more. Total

fixed costs are the sum of all fixed costs that a business incurs.

2. A **variable cost** is a cost that changes with the level of output. In other words, if the business outputs nothing there is no variable cost. However, if the business outputs one unit (or more) then a variable cost appears. It may include material costs of products, production labour (hourly or piecework wages), sales commissions, repairs, maintenance, and more. Total variable costs are the sum of all variable costs that a business incurs at a particular level of output.

3. A **blended cost** is a cost that comprises both fixed cost and variable cost components. In other words, a portion of the total cost remains unchanged while another portion depends on the output. For calculation purposes, you must separate a blended cost into its fixed and variable cost components. A few examples will illustrate the concept of blended costs:

- Residential natural gas bills from SaskEnergy include a fixed charge per month of $23.20 plus charges for cubic metres of actual consumption based on transportation, distribution, and primary and supplemental gas rates. In this situation, the $23.20 is a fixed cost while the actual consumption of natural gas is a variable cost.

- A cellphone bill includes a fixed charge for the phone service plus any additional charges for usage, such as long distance, text, or data.

- If employees are paid a salary plus commission, then their salaries represent fixed costs while their commissions are a variable cost.

3.1.2 Calculation of Costs

We will introduce some notations first:

TVC is **Total Variable Cost**: The total variable costs in dollars that were incurred at a particular level of output. In the simple average formula, this is represented by the symbol $\sum x$, which stands for the total of all x values.

n is the Level of Output: In the simple average formula, n represented the number of data. For this chapter, the definition is further specified to represent the total number of units produced or sold or the total output that incurred the total variable costs.

UVC is **Unit Variable Cost**: The unit variable cost is an adapted simple average formula (3.6) with specific definitions for the data and the quantity. The end result of the calculation is the typical or average variable cost associated with an individual unit of output. Being a dollar cost, the unit variable cost is rounded to two decimals.

$$UVC = \frac{TVC}{n} \tag{3.6}$$

Follow these steps to calculate the unit variable cost:

Step 1: Identify all fixed, variable, and blended costs, along with the level of output. For variable costs, understand any important elements of how the cost is structured. For blended costs, separate the costs into variable and fixed components.

Step 2: Calculate the total variable cost (TVC) by totaling all variable costs at the indicated level of output. This involves taking any known unit variable costs and multiplying each by the level of output.

Step 3: Divide the total variable cost by the total level of output by applying Formula (3.6).

Example 1 Assume a company produces 10,000 units and wants to know its unit variable cost. It incurs production labour costs of \$3,000.00, material costs of \$1,875.00, and other variable costs totaling \$1,625.00.

Solution

Step 1: In this case, all costs are variable costs (production labour and material costs are always variable). The level of output is n = 10,000 units.

Step 2: Total all variable costs together to get

$$TVC = \$3,000.00 + \$1,875.00 + \$1,625.00 = \$6,500.00.$$

Step 3: Apply Formula 3.6 to arrive at

$$UVC = \frac{6,500.00}{10,000} = \$0.65.$$

This means that, on average, the variable cost associated with one unit of production is \$0.65.

Example 2 Mr. Smith runs a rental property and his costs in January 2021 are as follows: Property tax: \$157.00; Condo fee: \$350.00; Insurance cost: \$135.00;

Internet: \$77.65; Electricity: Basic monthly charges \$22.79 plus the cost of the usage of 540 kW. h and the unit cost is \$0.14 per kW. h; Gas: Basic monthly charges \$23.20 plus the cost of the usage of 196.952 m^3 and the unit cost is \$0.0993 per m^3. What's his total fixed cost and variable cost.

Exercise

1. **(Knowing your cost)** You are considering starting your own home-based Internet business. After a lot of research, you have gathered the following financial information: Generating and fulfilling sales of 430 units involves 80 hours of

Table 3.1: Costs Involved

Dell computer	\$214.48 monthly lease payments
Office furniture (desk and chair)	\$186.67 monthly rental
Shaw high-speed Internet connection	\$166.88 per month
Your wages	\$30.00 per hour
Utilities	\$13.00 per month plus \$0.20 per hour usage
Software (and ongoing upgrades)	\$20.00 per month
Business licences and permits	\$27.00 per month
Google click-through rate	\$10.00 per month + \$0.01 per click payable as total clicks per sale

work per month. Based on industry response rates, your research also shows that to achieve your sales you require a traffic volume of 34,890 Google clicks. On a monthly basis, calculate the total fixed cost, total variable cost, and unit variable cost.

For questions 2–7, solve for the unknown variables (identified with a "?") based on the information provided (For some questions, you may leave it for review if you cannot solve it now).

	Total Fixed Costs	Total Variable Costs	Unit Variable Cost	Selling Price	Total Revenue	Level of Output	Net Income	Contribution Rate	Unit Contribution Margin
2.	\$5,000	\$6,600	?	\$13	?	?	\$4,000	?	\$7.50
3.	\$2,000	?	\$5	\$10	\$10,000	?	?	?	?
4.	?	?	?	\$75	\$60,000	?	\$14,500	35%	?
5.	\$18,000	\$45,000	?	?	\$84,600	1,800	?	?	?
6.	?	?	?	?	\$78,000	3,000	\$18,000	?	\$13
7.	?	\$94,050	\$75.24	?	?	?	−\$19,500	38%	?

Figure 3.1: Questions 2-7

8. Classify each of the following costs as fixed costs, variable costs, or blended costs. If a cost is blended, separate it into its fixed and variable components.

a. Natural gas bill for \$15.00 per month plus \$0.33/m^3 of consumption.

b. A chief executive officer salary of \$240,000.00 per year.

c. An author earning a royalty of 5% of sales.

d. Placing a commercial on television for \$300,000.00.

e. A cellphone bill for \$40.00 per month plus \$0.25/minute for long distance.

f. Hourly production worker wages of \$18.00/hr.

g. Sales staff who are compensated at a salary of \$1,000.00 per month plus 15% of sales.

3.1.3 Net Income Using a Total Revenue and Total Cost Approach

We will calculate the net income based on total revenues and total costs using the following model:

$$\begin{aligned}\texttt{Net Income} &= \texttt{Total Revenue-Total Cost}\\ &= \texttt{Total Revenue} - (\texttt{Total Fixed Cost} + \texttt{Total Variable Cost}),\end{aligned}$$

i.e.

$$NI = nS - (TFC + n(UVC)). \tag{3.7}$$

Net Income (NI) is the amount of money left after all costs are deducted from all revenues. If the number is positive, then the business is profitable. If the number is negative, then the business suffers a loss since the costs are exceeding the revenues. Note that many companies use the terms *net earnings* or *net profit* instead of the term *net income*. Net income is based on a certain level of output. This model assumes that the number of units that are produced or purchased (for resale) by the company exactly matches the number of units that are output or sold by the company. Therefore, the model does not consider inventory and its associated costs.

n is Level of Output: The number of units produced or sold or the output that incurred all of the variable costs.

S is Unit Selling Price: The unit selling price of the product.

Total Revenue (TR or nS) in the formula calculates how much money or gross income the sale of the product at a certain output level brings into the organization. Total revenue is the entire amount of money received by a company for selling its product, calculated by multiplying the quantity n sold by the selling price S.

TFC is Total Fixed Costs: The total of all costs that are not affected by the level of output.

UVC is Unit Variable Cost: From Formula 3.6, this is the average variable cost associated with an individual unit of output.

$TFC + n(UVC)$ is Total Cost. This term in the formula calculates how much money is spent to generate the revenue. Total cost is the sum of all costs for the company, including both the total fixed costs and total variable costs. Two terms make up the costs: Total fixed cost (TFC) is a constant since these costs do not change with the level of output; total variable cost, represented mathematically by $n(UVC)$, is the level of output multiplied by the unit variable cost.

We will follow the following steps to calculate the net income:

Step 1: Calculate the total revenue. This requires identifying the unit selling price of the product and multiplying it by the level of sales or output.

Step 2: Calculate your total costs. This requires identifying and separating costs into fixed and variable components. Fixed costs are totaled to arrive at the total fixed cost. Total variable costs are either known or can be calculated through multiplying the unit variable cost by the level of sales or output.

Step 3: Calculate the net income by applying Formula (3.7).

Example 1 Assume that in February a company incurred total variable costs of $10,000.00 in the course of producing 1,000 units. For next month it forecasts total fixed costs of $5,000.00 and all variable costs remaining unchanged. Projected production for March is 1,200 units selling for $25.00 each. You want to estimate the net income in March.

Step 1: Using Formula (3.7), you calculate total revenue from nS, or the total level of output multiplied by the price of the product. If you project sales of 1,200 units (n) at $25.00 each (S), then the total predicted revenue is 1,200($25) = $30,000.00.

Step 2: Total fixed cost (TFC) is $5,000.00. To get the total variable cost, you must resolve $n(UVC)$. You calculate the unit variable cost (UVC), with Formula (3.6). Using the current month data, you see $UVC = 10{,}000.00/1{,}000 =$ $10.00. If the projected level of output is 1,200 units, then the total variable costs are 1,200($10.00) = $12,000.00.

Step 3: Applying Formula (3.7) you have NI=Total Revenue-Total Costs=30,000.00 – (5,000 + 12,000) = $13,000.

Based on the numbers, you forecast net income of $13,000.00 for next month.

We will solve it using Excel.

Exercises

1. **(Start your internet business)** Recall from Exercise 1 in Section 3.1.2 that

you are considering starting your own home-based Internet business. The following information is known:

$TFC = \$638.03$, $UVC = \$6.43$, predicted as $n = 430$.

Based on these numbers, calculate:

1). The predicted net income if your price per unit is \$10.00.

2). The dollar change in net income if you decide to pay yourself a higher wage of \$35.38 per hour instead of \$30.00 per hour while still working 80 hours. Note the total variable costs excluding wages were \$364.90.

3). The dollar change in net income if sales are 30% lower than your initial forecast.

3.1.4 Net Income Using a Total Contribution Margin Approach

In Exercise 1 in Section 3.1.3, you calculated that if you sell the projected 430 units of product for your Internet business, the total net income is \$897.07. What if you sold 431 units of the product? How would your net income change? Clearly it should rise, but by how much? One approach to answering this question is to rerun the numbers, revising the total revenues and total variable costs to calculate a new net income. This new net income can then be compared against the existing net income to determine how it changed. This is a multistep approach and involves a lot of work.

An alternative approach explored in this section involves using a *unit contribution margin* to calculate the net income. The benefit of this approach is that with minimal calculations you can easily assess the impact of any change in the level of output. In accounting and marketing, the **contribution margin** is the amount that each unit sold adds to the net income of the business. This approach allows you to understand the impact on net income of each unit sold, and we will explain its further benefit of allowing for a straightforward calculation of a break-even point.

The contribution margin determines on a per-unit basis how much money is left after unit variable costs are removed from the price of the product. This leftover money is then available to pay for the fixed costs. Ultimately, when all fixed costs have been paid for, the leftover money becomes the profits of the business. If not enough money is left to pay for the fixed costs, then the business has a negative net income and loses money.

We will introduce a new definition, Unit Contribution Margin.

Unit Contribution Margin (UCM**)** is the amount of money that remains available to pay for fixed costs once the unit variable cost is removed from the selling price of the product. We can calculate using the following formula:

$$UCM = S - UVC, \tag{3.8}$$

where S is the selling price and UVC is the unit variable cost.

If you have no units sold, your net income is negative and equal to the total fixed costs associated with your business, since there is no offsetting revenue to pay for those costs. With each unit sold, the contribution margin of each product is available to pay off the fixed costs.

Therefore, we can get a new formula to calculate the net income:

$$NI = n(UCM) - TFC, \tag{3.9}$$

where n is the number of units produced or sold or the output that incurred all of the variable costs, $n(UCM)$ is Total Contribution Margin (calculates how much money is left to pay the total fixed costs), and TFC is the Total Fixed Costs.

We can calculate the net income using a contribution margin approach by following steps:

Step 1: If unit information is known, apply Formula (3.8) and calculate the unit contribution margin by subtracting the unit variable cost from the selling price. This may or may not require you to use Formula (3.6) to calculate the unit variable cost.

Step 2: Calculate the total contribution margin by multiplying the unit contribution margin with the associated level of output.

Step 3: Calculate the total fixed cost by adding all fixed costs.

Step 4: Based on the level of output, calculate the net income by applying Formula (3.9).

Example 1 Using the contribution margin approach, calculate the net income for a product that sells for \$75.00, has unit variable costs of \$31.00, total fixed costs of \$23,000.00, and total sales of 800 units.

Step 1: The unit contribution margin is calculated from Formula (3.8). If the product sells for \$75.00 and has unit variable costs of \$31.00, then UCM=75.00-31.00 = \$44.00. This means that every unit sold has \$44.00 left over to contribute toward fixed costs.

Step 2: Now convert that into a total contribution margin. The first part of Formula 3.9 calculates total contribution margin through n(UCM). With sales of 800 units, the total contribution margin is 800(\$44.00)=\$35,200.00.

Step 3: Total fixed costs are known: TFC=\$23,000.00.

Step 4: Apply Formula 3.9, which translates to Net Income=Total contribution margin-Total fixed costs=35,200.00-23,000.00=\$12,200.00.

Example 2 Frank runs a designer candle-making business out of his basement. He sells the candles for \$15.00 each, and every candle costs him \$6.00 to manufacture. If his fixed costs are \$2,300.00 per month, what is his projected net income or loss next month, for which he forecasts sales of 225 units?

Example 3 A college print shop leases an industrial Xerox photo copier for \$1,500.00 per month plus 1.5 cents for every page. Additional printing costs are estimated at 2 cents per page, which covers toner, paper, labour, and all other incurred costs. If copies are made for students at 10 cents each, determine the following:

a. How does net income change with every 100 copies sold?

b. What is the monthly net income if, on average, the shop makes 25,000 copies for students each month?

Exercises

1. **(The contribution margin for your internet business)** Continuing with exercises of the internet business, calculate the unit contribution margin and net income using the contribution margin approach. From the previous exercises, recall the unit variable cost is \$6.43, unit selling price is \$10.00, total fixed costs are \$638.03, and the projected sales are 430 units. (Calculate the unit contribution margin (UCM) followed by the net income (NI) using the contribution margin approach.)

3.1.5 Contribution Rates

It is difficult to compare different products and their respective dollar amount contribution margins if their selling prices and costs vary widely. For example, how do you compare a unit contribution margin of \$1,390.00 (selling price of \$2,599.99) on a big screen television to a unit contribution margin of \$0.33 on a chocolate bar (selling price of \$0.67)? On a per-unit basis, which contributes relatively more to fixed costs? To facilitate these comparisons, the products must be placed on equal terms, requiring you to convert all dollar amount contribution margins into percentages. **A contribution rate** is a contribution margin expressed as a percentage of the selling price.

If unit information is known, including the unit variable cost and unit selling price, then calculate the contribution rate using unit information as expressed in Formula (3.10):

$$CR = \frac{UCM}{S} \times 100\%. \tag{3.10}$$

If any unit information, including the unit variable cost or unit selling price, is unknown or unavailable, then you cannot apply Formula (3.10). Sometimes only aggregate information is known. When total revenue and total variable costs for any product are known or can at least be calculated, then you must calculate the contribution rate from the aggregate information as expressed in Formula (3.11).

$$CR = \frac{TR - TVC}{TR} \times 100\%. \tag{3.11}$$

where TR is the total revenue.

Steps to calculate a contribution rate:

Step 1: Identify the required variables and calculate the unit contribution margin, if needed.

Step 2: Calculate the contribution rate.

Example 1 The television sells for \$2,599.99 and has a unit contribution margin of \$1,390.00; a chocolate bar sells for \$0.67 and has a unit contribution margin of \$0.33. Calculate contribution rates of the television and chocolate bar.

Notice that the information being provided is on a per unit basis, we will use Formula (3.10).

Step 1: The contribution margins are known. For the television, UCM=\$1,390.00, and for the chocolate bar, UCM=\$0.33.

Step 2: Applying Formula 3.10, the television $CR = \frac{1{,}390.00}{2{,}599.99} \times 100\% = 53.4617\%$, while the chocolate bar $CR = \frac{0.33}{0.67} \times 100\% = 49.2537\%$. It is now evident from the contribution rate that 4.208% more of the television's selling price is available to pay for fixed costs as compared to the chocolate bar's price.

Example 2 The television's total revenue is \$129,999.50 and associated total variable costs are \$60,499.50. The chocolate bar has total revenue of \$3,886.00 with total variable costs of \$1,972.00. Based on this information, you are to determine the product with the higher contribution rate.

Step 1: The aggregate numbers are known for both products. For the television, TR=\$129,999.50 and TVC=\$60,499.50.

For the chocolate bar, TR=\$3,886.00 and TVC=\$1,972.00.

Step 2: Applying Formula 3.11, the television $CR = \frac{129{,}999.50-60{,}499.50}{129{,}999.50} \times 100\% = 53.4617\%$, while the chocolate bar $CR = \frac{3{,}886-1{,}972}{3{,}886} \times 100\% = 49.2537\%$.

We have reached the same conclusion as above.

Example 3 Last year, A Child's Place franchise had total sales of \$743,000.00. If its total fixed costs were \$322,000.00 and net income was \$81,000.00, what was its contribution rate?

Example 4 Monsanto Canada reported the following on its income statement for one of its divisions:

Sales	\$6,000,000
Total Fixed Costs	\$2,000,000
Total Variable Costs	\$3,200,000
Total Costs	\$5,200,000
Net Income	\$800,000

Calculate the total contribution margin in dollars and the contribution rate for this division.

Exercises

1. In 2018, A Child's Place franchise had total sales of \$743,000.00. If its total fixed costs were \$322,000.00 and net income was \$81,000, what was its contribution rate?

3.2 Break-Even Analysis

Break-even analysis is the analysis of the relationship between costs, revenues, and net income with the sole purpose of determining the point at which total revenue equals total cost. This **break-even point** is the level of output (in units or dollars) at which all costs are paid but no profits are earned, resulting in a net income equal to zero. To determine the break-even point, you can calculate a break-even analysis in two different ways, involving either the number of units sold or the total revenue in dollars.

3.2.1 Break-Even Analysis in Units

In this method, our goal is to determine the level of output that produces a net income equal to zero. This method requires unit information, including the unit selling price and unit variable cost. It is helpful to see the relationship of total cost and total revenue on a graph.

Example 1 Assume that a company has the following information:

TFC=\$400.00, S=\$100.00, UVC=\$60.00.

The graph shows dollar information on the y-axis and the level of output on the x-axis. Here is how you construct such a graph:

1. Plot the total costs:

a. At zero output you incur the total fixed costs of \$400.00. Denote this as Point 1 (0, \$400.00).

b. As you add one level of output, the total cost rises in the amount of the unit variable cost. Therefore, total cost is TFC+$n(UVC)$ = 400.00 + 1(60.00)=460.00. Denote this as Point 2 (1, \$460.00).

c. As you add another level of output (2 units total), the total cost rises once again in the amount of the unit variable cost, producing 400.00+2(60.00)=\$520.00. Denote this as Point 3 (2, \$520.00).

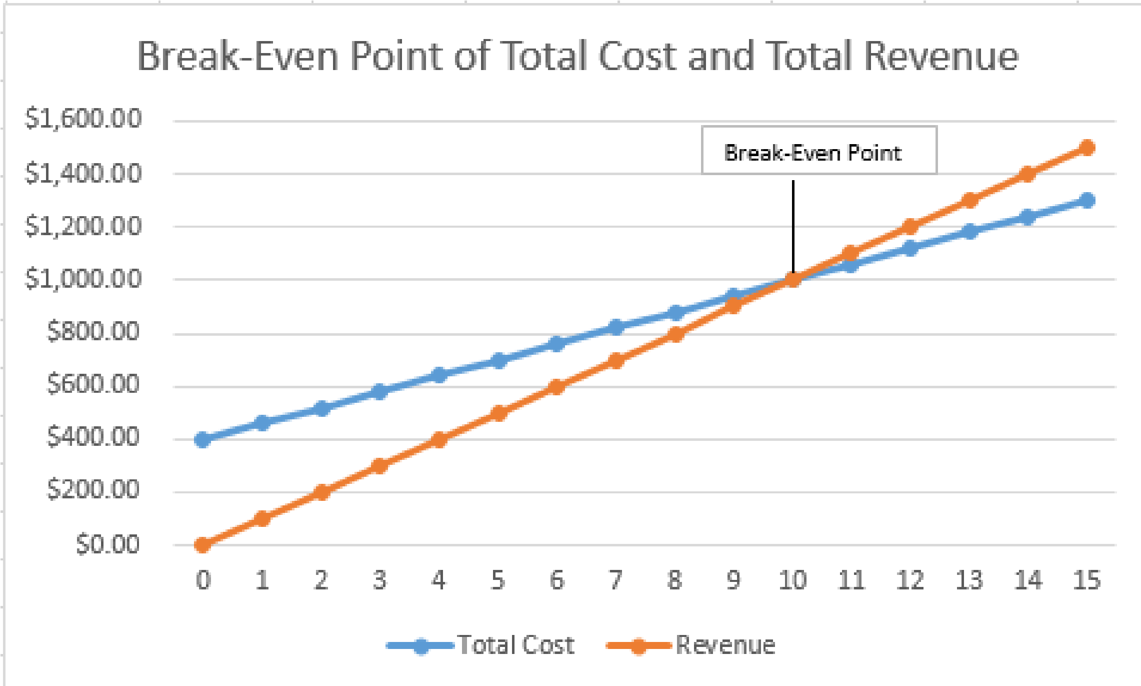


Figure 3.2: Break-Even Point

d. Repeat this process for each subsequent level of output and plot it onto the figure. The blue line plots these total costs at all levels of output.

2. Plot the total revenue:

a. At zero output, there is no revenue. Denote this as Point 1′ (0, \$0.00).

b. As you add one level of output, total revenue rises by the selling price of the product. Therefore, total revenue is nS=1(100.00)=\$100.00. Denote this as Point 2′ (1, \$100.00).

c. As you add another level of output (2 units total), the total revenue rises once again in the amount of the selling price, producing 2(100.00)=\$200.00. Denote this as Point 3′ (2, \$200.00).

d. Repeat this process for each subsequent level of output and plot it onto the figure. The orange line plots the total revenue at all levels of output.

We can draw it through Excel using the formula.

Our goal is to find n by setting the net income to zero, resulting in

$$0 = nS - (TFC + n(UVC)).$$

Rearranging and solving this formula for n gives the following:

$$0 = nS - TFC - n(UVC),$$

$$TFC = nS - n(UVC),$$

$$TFC = n(S - UVC),$$

$$n = \frac{TFC}{S - UVC}, \tag{3.12}$$

or

$$n = \frac{TFC}{UCM}. \tag{3.13}$$

When you calculate the break-even units, the formula may produce a number with decimals. A partial unit cannot be sold, so the rule is always to round the n up to the next integer, regardless of the decimal.

We can calculate the break-even point in units by following steps:

Step 1: Calculate or identify the total fixed costs (TFC).

Step 2: Identify the unit contribution margin (UCM), or the selling price (S) and the unit variable cost (UVC).

Step 3: Apply Formula (3.12) or (3.13) based on the information we are given.

Solving this example,

Step 1: Total fixed costs are known, TFC=\$400.00.

Step 2: The selling price S=\$100.00 and the unit variable cost UVC=\$60.00. (The unit contribution margin is UCM=100.00-60.00=\$40.00.)

Step 3: Applying Formula (3.12) or (3.13) results in

$$n = \frac{TFC}{S - UVC} = \frac{400.00}{100.00 - 60.00} = 10 \quad \texttt{units}, \quad or \quad n = \frac{TFC}{UCM} = \frac{400.00}{40.00} = 10 \quad \texttt{units},$$

and this agrees with the result from graph.

Example 2 If the break-even point is 15,000 units, the selling price is \$95.00, and the unit variable cost is \$75.00, what are the company's total fixed costs?

Example 3 Louisa runs a secretarial business part time in the evenings. She takes dictation or handwritten minutes and converts them into printed word-processed documents. She charges \$5.00 per page for her services. Including labour, paper, toner, and all other supplies, her unit variable cost is \$2.50 per page. She invested \$3,000.00 worth of software and equipment to start her business. How many pages will she need to output to break even?

Example 4 Cheryl is starting a deluxe candy apple business. The cost of producing one candy apple is \$4.50. She has total fixed costs of \$5,000.00. She is thinking of selling her deluxe apples for \$9.95 each.

a. Determine her unit break-even point at her selling price of \$9.95.

b. Cheryl thinks her price might be set too high and lowers her price to $8.95. Determine her new break-even point.

c. An advertising agency approaches Cheryl and says people would be willing to pay the $9.95 if she ran some "upscale" local ads. They would charge her $1,000.00. Determine her break-even point.

d. If she wanted to maintain the same break-even units as determined in a, what would the price have to be to pay for the advertising?

Exercise

1. (The Break-Even Units for Your Planned Internet Business)

Recall that the Internet business explored throughout four exercises from section 3.1. Now let's determine the break-even point in units. As previously calculated, the total fixed costs are $638.03 and the unit contribution margin is $3.57.

2. High Five manufactures a line of children's pet toys. If it sells the toy to distributors for $2.30 each while variable costs are $0.75 per toy, how many toys does it need to sell to recover the fixed cost investment in these toys of $510,000.00? What total revenue would this represent?

3.2.2 Break-Even Analysis in Dollars

The income statement of a company does not display unit information. All information is aggregate, including total revenue, total fixed costs, and total variable costs. Typically, no information is listed about unit selling price, unit variable costs, or the level of output. Without this unit information, it is impossible to apply Formula (3.12) or (3.13). This leads us to develop a new method for calculating the break-even point relies strictly on aggregate information. As a result, we cannot calculate the break-even point in units. Instead, we calculate the break-even point in terms of aggregate dollars expressed as total revenue.

To derive the break-even point in dollars, once again start with the Net Income Formula, where total revenue at break-even point less total fixed costs and total variable costs must equal a net income of zero:

$$NI = TR - (TFC + TVC),$$

$$0 = TR - (TFC + TVC).$$

Rearranging this formula for total revenue gives:

$$0 = TR - TFC - TVC,$$

hence

$$TR = TFC + TVC.$$

Thus, at the break-even point the total revenue must equal the total cost. Substituting this value into the numerator of Formula 3.11 gives us:

$$CR = \frac{TR - TVC}{TR} \times 100\%,$$

$$CR = \frac{(TFC + TVC) - TVC}{TR} \times 100\%,$$

$$CR = \frac{TFC}{TR} \times 100\%,$$

therefore, a final rearrangement results in Formula (3.14), which expresses the break-even point in terms of total revenue dollars.

$$TR = \frac{TFC}{CR}. \tag{3.14}$$

Steps to calculate the break-even point in total revenue dollars:

Step 1: Calculate or identify the total fixed costs (TFC).

Step 2: Calculate the contribution rate (CR), by applying any needed techniques or formulas. If not provided, typically the CR is calculated using Formula (3.11), which requires aggregate information only.

Step 3: Apply Formula (3.14) to calculate the break-even point in dollars.

Example 1 Assume that you are looking at starting your own business. You determine that your total fixed costs TFC=$420,000.00. You are not sure of your variable costs but need to calculate your break-even point. Many of your competitors are publicly traded companies, so you go online and analyze their financial statements, and you find that your competitors have a contribution rate of 35%, or CR=0.35, on average. What is your estimate of your breakeven point in dollars?

Step 1: Total fixed costs are TFC=$420,000.00.

Step 2: The estimated contribution rate is CR=0.35.

Step 3: Applying Formula (3.14)results in

$$TR = 420,000.00/0.35 = \$1,200,000.00.$$

If you average a similar contribution rate, you require total revenue of $1,200,000.00 to cover all costs, which is your breakeven point in dollars.

Example 2 If your organization has a contribution rate of 45% and knows the break-even point is $202,500.00, what are your organization's total fixed costs?

Example 3 What is the unit contribution margin on a product line that has fixed costs of $1,800,000.00 with a break-even point of 360,000 units?

Example 4 Robert is planning a wedding social for one of his close friends. Costs involve $865.00 for the hall rental, $135.00 for a liquor licence, $500.00 for the band, and refreshments and food from the caterer cost $10.00 per person. If he needs to raise $3,000.00 to help his friend with the costs of his wedding, what price should he charge per ticket if he thinks he can fill the social hall to its capacity of 300 people?

Example 5 Boston Beer Company, the brewer of Samuel Adams, reported the following financial information to its shareholders:

Total Revenue	$388,600,000
Total Variable Costs	$203,080,000
Total Fixed Costs	$182,372,000
Total Costs	$385,452,000
Net Income	$3,148,000

If this represented sales of 2,341,000 barrels of beer, determine its break-even point in units and dollars.

Exercises

1. In the annual report to shareholders, ABC Manufacturing reported total gross sales of \$7,200,000.00, total variable costs of \$4,320,000.00, and total fixed costs of \$2,500,000.00. Determine ABC's break-even point in dollars.

2. You are thinking of starting your own business and want to get some measure of feasibility. You have determined that your total fixed costs would be \$79,300.00. From annual business reports and competitive studies, you estimate your contribution rate to be 65%. What is your break-even point in dollars?

3.3 Return On Investment (ROI)

Return on Investment (ROI) is a performance measure used to evaluate the efficiency of an investment or compare the efficiency of a number of different investments. ROI tries to directly measure the amount of return on a particular investment, relative to the investment's cost. To calculate ROI, the benefit (or return) of an investment is divided by the cost of the investment. The result is expressed as a percentage or a ratio.

The return on investment formula:

$$\texttt{ROI} = \frac{\texttt{Current Value of Investment} - \texttt{Cost of Investment}}{\texttt{Cost of Investment}} \tag{3.15}$$

In the above formula, "Current Value of Investment" refers to the proceeds obtained from the sale of the investment. Because ROI is measured as a percentage, it can be easily compared with returns from other investments, allowing one to measure a variety of types of investments against one another.

For example, suppose Adam invested \$10,000.00 in Evergreen Printing Corp. in 2015 and sold his stock shares for a total of \$12,000.00 one year later. To calculate his return on his investment, he would divide his profits (\$12,000.00-\$10,000.00=\$2,000.00) by the investment cost (\$10,000), for a *ROI* of 20%.

With this information, he could compare his investment in Evergreen Printing Corp. with his other projects. Suppose Adam also invested \$20,000.00 in AAA Construction Inc. in 2012 and sold his shares for a total of \$28,00.000 in 2015. The *ROI* on Adam's holdings in AAA would be $\frac{8{,}000}{20{,}000} = 40\%$.

The *ROI* of Adam's second investment was twice that of his first investment. However, the time between Adam's purchase and sale was one year for his first investment and three years for his second one.

We could adjust the *ROI* of his multi-year investment accordingly. Since his total *ROI* was 40%, to obtain his average annual ROI, he could divide 40% by 3 to yield 13.33%. With this adjustment, it appears that although Adam's second investment earned him more profit, his first investment was actually the more efficient investment.

There are many benefits to using the return on investment ratio: first, it is simple and easy to calculate; second, it is a universally understood concept.

However, there are some limitations of the *ROI* formula: On one hand, the *ROI* formula disregards the factor of time. A higher ROI number does not always mean a better investment option. The above example reflects this fact. The investor needs to compare two instruments under the same time period and same circumstances. On the other hand, the ROI formula is susceptible to manipulation. An ROI calculation will differ between two people depending on what ROI formula is used in the calculation. A marketing manager can use the property calculation explained in the example section without accounting for additional costs such as maintenance costs, property taxes, sales fees, stamp duties, and legal costs. An investor needs to look at the true ROI, which accounts for all possible costs incurred when each investment increases in value.

As mentioned above, one of the drawbacks of the traditional return on investment metric is that it doesn't take into account time periods. For example, a return of 25% over 5 years is expressed the same as a return of 25% over 5 days. But obviously, a return of 25% in 5 days is much better than 5 years!

To overcome this issue we can calculate an annualized ROI formula.

$$\texttt{Annualized ROI} = (1 + \texttt{ROI})^{\frac{1}{\texttt{number of years}}} - 1. \tag{3.16}$$

or

$$\texttt{Annualized ROI} = \left(\frac{\texttt{Ending Value}}{\texttt{Beginning Value}}\right)^{\frac{1}{\texttt{number of years}}} - 1. \tag{3.17}$$

We will learn how to calculate ROI with an Excel spreadsheet.

Example 1 (Net Income Method) AAA company invested \$1,000,000.00 for two years and the net gain is \$125,500.00 from this investment. What is the regular return on this investment?

Example 2 (Capital Gain Method) Cheryl purchased a stock at \$12.50 per share and she sold it at \$15.20 per share. What's the ROI of this investment?

Example 3 (Total Return Method) Brenda bought 1000 shares of a stock at \$12.50 per share May 13, 2018, and has received \$1.25 dividends per share on July 1, 2019, and she sold it at \$15.20 per share on January 15, 2020. What's the return on this investment? What's the Annualized ROI?

Example 4 Dianna bought 2000 shares of a stock at \$12.50 per share January 5, 2017, and she sold it at \$15.20 per share on August 24, 2017. What's the return on this investment? What's the Annualized ROI?

Example 5 Rhoda has invested to buy four stocks three years ago: A: \$120,000, B: \$10,000, C: \$50,000, and D: \$80,000, and their current values are: \$135,050, \$10,890, \$52,000 and \$79,000. Calculate ROI and Annualized ROI for each stock.

Exercises

1. An investor buys a stock on January 15th, 2017 for \$12.50 and sells it on September 24, 2017 for \$16.25. What is the regular and annualized return on this investment?

2. An investor purchases property A, which is valued at \$500,000.00. Two years later, the investor sells the property for \$1,000,000.00. What is the regular and

annualized return on investment?

Reference

[1] J. Olivier, Business Math, Creative Commons License (CC BY-NC-SA).

3.4 Chapter 3 Review Questions

1. Logitech manufactures a surround sound multimedia speaker system for personal computers. The average selling price is $120.00 per unit, with unit variable costs totaling $64.00. Fixed costs associated with this product total $980,000.00. Calculate the net income or loss if 25,000 units are sold.

2. Microsoft sells its wireless laser desktop mouse and keyboard for $70.00. Unit variable costs are $45.60 and fixed costs associated with this product total $277,200.00.

a. What is the break-even point in units?

b. What is the net income or loss if 40,000 units are sold?

3. Frances has a home-based Avon business. She purchases makeup from Avon for an average price of $12.00 and sells it to her customers on average for $20.00. She assigns $100.00 of her monthly computer leasing costs to this product line and spends $260.00 per month advertising this product line. In hosting Avon parties, she estimates her labour cost at $2.00 per unit sold.

a. What is her net income or loss if 100 units are sold?

b. What is her net income or loss if 35 units are sold?

c. What is the break-even point in both units and dollars?

d. What is the new unit break-even point if Frances increases her advertising costs by 50% per month?

4. Brynne is developing a business plan. From market research she has learned that her customers are willing to pay $49.95 for her product. Unit variable costs are estimated at $23.75, and her monthly total fixed costs are $12,000.00. To earn a net income of at least $5,000.00, what minimum level of monthly sales (in units) does she need?

5. Advanced Microcomputer Systems had annual sales of $31,979,000.00 with total variable costs of $20,934,000.00 and total fixed costs of $6,211,000.00.

a. What is the company's contribution rate?

b. What annual revenue is required to break even?

c. In the following year, a competitor dealt a severe blow to sales and total revenues dropped to $15,000,000.00. Calculate the net income.

6. In the North American automotive markets, Toyota and Honda are two of the big players. The following financial information (all numbers in millions of dollars) for each was reported to its shareholders:

Compare the break-even points in total dollars between the two companies based on these reports.

	Toyota	Honda
Total Revenue	247,734	101,916
Total Variable Costs	204,135	75,532
Total Fixed Costs	20,938	24,453
Total Costs	225,073	99,985
Net Income	22,661	1,931

7. Lagimodiere Industry's unit variable costs are 45% of the unit selling price. Annual fixed costs are $1.5 million.

a. What total revenue results in a net income of $250,000.00?

b. Calculate the break-even dollars for Lagimodiere Industry.

c. If revenues exceed the break-even dollars by 20%, determine the net income.

8. The following annual information is known about a company:

Material inputs to manufacturing	$17,400,000.00
Management salaries	$2,750,000.00
Supplies needed for manufacturing	$8,345,000.00
Utilities used in non-production activities	$2,222,000.00
Number of units sold	75,132
Marketing and advertising	$3,500,000.00
Production worker wages	$9,990,000.00
Utilities used in production activities	$6,235,000.00
Property taxes	$2,100,000.00
Unit price	$749.00

a. Calculate the unit variable cost, total fixed costs, unit contribution margin, contribution rate, net income, and break-even points in both units and dollars.

b. The marketing research manager shows that if the company lowers its price by 5%, unit sales will rise by 5%. Should the company lower its price? Justify your decision.

c. Assume that the price remains the same. Due to economic pressures, fixed costs are forecasted to rise by 6% and variable costs rise by 4% next year. At the same level of output, by what percentage must the selling price rise to maintain the same level of net income achieved in (a)?

9. Dennis buys a stock on August 7th, 2017 for $22.50 and sells it on March 1, 2019 for $25.20. What is the regular and annualized return on this investment?

10. An real estate investor purchases a property, which is valued at $590,000.00. Three years later, the investor sells the property for $980,000.00. What is the regular and annualized return on investment?

Chapter 4

Descriptive Analysis

4.1 Variables and Data

Statistics is a branch of mathematics dealing with data collection, organization, analysis, interpretation and presentation. We begin by introducing some definitions that you need to know.

A **variable** is a characteristic that changes or varies over time/or for different individuals or objects under consideration.

For example, body temperature is a variable that changes over time within a single individual; it also varies from person to person. Income, height, age, and political affiliation are all variables.

An **experimental unit** is the individual or object on which a variable is measured. A single **measurement** or data value results when a variable is actually measured on an experimental unit.

A **population** is the set of all measurements of interest to the investigator.

A **sample** is the subset of measurements selected from the population of interest.

If a measurement is generated for every experimental unit in the entire collection, the resulting data set constitutes the population of interest. Any smaller subset of measurements is a sample.

Univariate data result when a single variable is measured on a single experimental unit.

Bivariate data result when two variables are measured on a single experimental unit.

Multivariate data result when more than two variables are measured.

Table 4.1: Data from a clinic

Patient	**Systolic blood pressure**	**Body mass**	**Age**	**Gender**	**Married**	**Smoke**
1	185	32.70	54	F	N	Y
2	174	18.00	53	F	Y	Y
3	147	15.90	50	M	Y	N
4	174	12.30	34	F	Y	N
5	130	20.60	28	M	N	Y

Table 4.1 is a sample containing 5 experimental units (patients), and there are 6 variables. Hence this is a multivariate data set, and the experimental unit is a particular patient. For example, the measurement of patient 5 is (130, 20.60, 28, M, N, Y).

Example 1 Tablet PC Comparison provide a wide variety of information about tablet computers. Their website enables consumers to easily compare different tablets using factors such as cost, type of operating system, display size, battery life, and CPU manufacturer. A sample of 10 tablet computers is shown in the following table:

Table 4.2

Product Infromation for 10 Tablet Computers

Tablet	Cost ($)	Operating System	Diplay Size (inches)	Battery Life (hours)	CPU Manufacturer
Acer Iconia W510	599	Windows	10.1	8.5	Intel
Amoazon Kindle Fire HD	299	Android	8.9	9	TI OMAP
Apple iPad 4	499	iOS	9.7	11	Apple
HP Envy X2	860	Windows	11.6	8	Intel
Lenovo ThinkPad Tablet	668	Windows	10.1	10.5	Intel
Microsoft Surface Pro	899	Windows	10.6	4	Intel
Motorola Droid Xyboard	530	Android	10.1	9	TI OMAP
Samsung Ativ Smart PC	590	Windows	11.6	7	Intel
Samsung Galaxy Tab	525	Android	10.1	10	Nvidia
Sony Tablet S	360	Android	9.4	8	Nvidia

a. How many experimental units are in this data set?

b. How many variables are in this data set?

c. Which variables are qualitative (nominal or ordinal) and which variables are quantitative (continuous or discrete)?

Exercises:

1. Identify the experimental unit on which the following variables are measured:

a. Gender of a student Each student

b. Number of errors on a midterm exam Each exam from the student

c. Age of a cancer patient Each cancer patient

d. Number of flowers on an azalea plant Each azalea plant

e. Colour of a car entering the parking lot Each car entering the parking lot

4.2 Types of Variables

Variables can be classified into two categories: **qualitative** and **quantitative**.

Qualitative variables measure a quality or characteristic on each experimental unit.

Qualitative variables produce data that can be categorized according to similarities or differences in kind; hence, they are called categorical data. For example:

- Letter grades: A, B, C, D, F
- Service ranking: excellent, good, fair, poor
- Political affiliation: Conservatives, Liberals, NDP, Green Parties, Independent

There are two types of qualitative variable:

- **Nominal** - blood type, gender, occupation (can not be ranked)
- **Ordinal** - letter grades, ranking of performance (ranked)

Quantitative variables measure a numerical quantity or amount on each experimental unit.

There are two types of quantitative variables: **discrete** and **continuous**.

A **discrete variable** can assume only a finite or countable number of values.

A **continuous variable** can assume the infinitely many values corresponding to the points on a line interval.

Example 1 Identify the following variables:

a. Colors of skittles in the average package - Qualitative, nominal.

b. Income of a randomly selected professor - Quantitative, discrete.

c. Setting on a toaster (light, medium, dark, crispy) - Qualitative, ordinal.

d. Religions of people in Canada - Qualitative, nominal.

e. The anxiety level that rates you from -10 (almost none) to 10 (extremely) - Qualitative, ordinal.

f. Number of universities in Canada - Quantitative, discrete.

g. The average temperature in December in Regina - Quantitative, continuous.

Exercises

1. Identify each variable as quantitative or qualitative:

a. Amount of time it takes to assemble a wardrobe Quantitative, continuous

b. Number of student in a stat class Quantitative, discrete

c. Rating a service (excellent, good, fair, poor) Qualitative, ordinal

d. Province or territory from which a student come Qualitative, nominal

2. Identify the following quantitative variables as discrete or continuous:

a. Population in Saskatchewan Discrete

b. Weight of sugar consumed one day in Tim Horton's continuous

c. Time to complete a test continuous

d. Number of customers in superstore one day discrete

3. **Parking on Sask Polytech Campus** Five vehicles are selected from the vehicles that have been issued campus parking permits, and the following data are recorded:

Vehicle	Type (Qual)	Brand (Qual)	Carpool (Qual)	Commute Distance (kms) (Quan)	Age of Vehicle (Years) (Quan)
1	Car	Honda	No	50.5	4
2	SUV	Toyota	No	15.8	2
3	SUV	Honda	Yes	150.2	3
4	Truck	Ford	No	10.4	6
5	Van	Dodge	No	145.5	8

a. What are the experimental units? Each vehicle that has been issued campus parking permit

b. What are the variables being measured? What types of varibles are they?

c. Is this univariate, bivariate, or multivariate data? multivariate

4. A data set consists of the years that 100 cancer patients survived after they were diagnosed.

a. Is this set of measurements a population or a sample?

b. What are the variables being measured? What types of variable are they?

c. Is this univariate, bivariate, or multivariate data?

4.3 Graphs for Categorical Data

After the data have been collected, they can be consolidated and summarized to show the following information:

- What values of the variable have been measured?
- How often each value has occurred?

We can measure "how often" in the following ways:

- The **frequency**, i.e. the number of measurements in each category.
- The **relative frequency**, i.e. proportion of measurements in each category.
- The **percentage** of measurements in each category.

Let n be the total number of measurements in the set, we can calculate the relative frequency and percentage using these formulas:

$$\texttt{Relative frequency} = \frac{\texttt{Frequency}}{n}, \tag{4.18}$$

and

$$\texttt{Percentage} = \texttt{Relative frequency} \times 100\%. \tag{4.19}$$

It is easy to get that the sum of the frequencies is always n, the sum of the relative frequencies is 1, and the sum of the percentages is 100%.

The criteria to choose categories for a qualitative variable are:

- A measurement will belong to one and only one category.
- Each measurement has a category to which it can be assigned.

Once the measurements have been categorized and summarized in a statistical table, we can use either a pie chart or a bar chart to display the distribution of the data.

A **pie chart** is the circular graph that shows how the measurements are distributed among the categories.

A **bar chart** shows the same distribution of measurements in categories, with the height of the bar measuring how often a particular category was observed.

Example 1 A bag of M&M candies contains 25 candies, including 3 reds, 6 blues, 4 greens, 5 oranges, 3 browns and 4 yellows. Construct a pie chart and bar chart for this set of data.

Solution To construct a pie chart, assign one sector of a circle to each category. The angle of each sector should be proportional to the proportion of measurements (or relative frequency) in that category. Since a circle contains 360°, you can use this equation to find the angle:

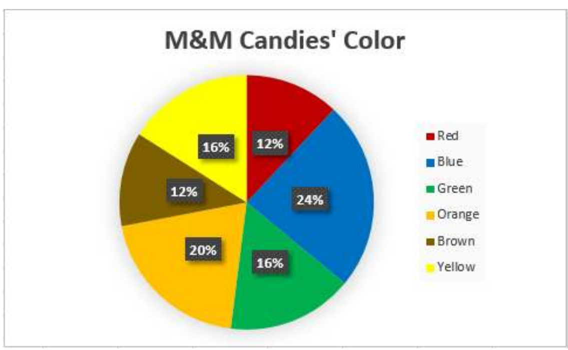
Angle = Relative frequency $\times$ 360°.

Table 4.3 shows the colors along with the frequency, relative frequency, percentages, and sector angles necessary to construct the pie chart.

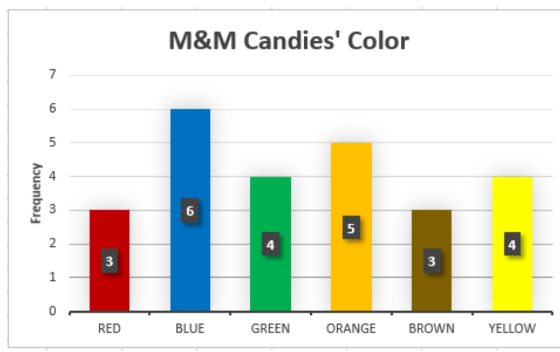
Table 4.3: Calculation Data of the M & M Candies

Color	Frequency	Relative Freq	Percent	Angle
Red	3	3/25=.12	12%	$0.12\times 360° = 43.2°$
Blue	6	6/25=.24	24%	$0.24\times 360° = 86.4°$
Green	4	4/25=.16	16%	$0.16\times 360° = 57.6°$
Orange	5	5/25=.20	20%	$0.20\times 360° = 72.0°$
Brown	3	3/25=.12	12%	$0.12\times 360° = 43.2°$
Yellow	4	4/25=.16	16%	$0.16\times 360° = 57.6°$
Total	25	1	100%	360°

Figure 4.1 (a) shows the pie chart constructed from the values in the table. While pie charts use percentages to determine the relative sizes of the "pie slices", bar charts usually plot frequency against the categories. A bar chart for these data is shown in Figure 4.1 (b).



(a) Pie Chart



(b) Bar Chart

Figure 4.1: Pie Chart and Bar Chart of the Color of 25 Candies

We will learn how to calculate pie chart data, how to create a pie chart and a bar chart using spreadsheet.

Exercises

1. Fifty people are grouped into four categories - A, B, C, and D - and the number of people who fall into each category is shown in the table:

Category	Frequency
A	11
B	14
C	20
D	5

a. What is the experimental unit?

b. What is the variable being measured? Is it qualitative or quantitative?

c. Construct a pie chart to describe the data.

d. Construct a bar chart to describe the data.

e. Does the shape of the bar chart in part d change depending on the order of presentation of the four categories? Is the order of presentation important?

f. What *proportion* of the people are in category B, C, or D?

g. What *percentage* of the people are *not* in category B?

2. A manufacturer of dining table has plants in Quebec (QC), Ontario (ON), and Manitoba (MB). A group of 30 tables is randomly selected from the computerized database, and the province in which each is produced is recorded:

ON QC QC MB ON QC QC MB MB MB QC QC ON QC MB ON QC MB MB MB ON QC QC ON ON QC MB MB ON QC

a. What is the experimental unit?

b. What is the variable being measured? Is it qualitative or quantitative?

c. Construct a pie chart to describe the data.

d. Construct a bar chart to describe the data.

e. What proportion of the table are made in Ontario?

f. Which province produced the most tables in the group?

4.4 Graphs for Quantitative Data

We still can use pie chart and bar chart to describe data collected for a quantitative variable measured on different segments of the population, or for different categories of classification. The **pie chart** displays how the total quantity is distributed among the categories, and the **bar chart** uses the height of the bar to display the amount in a particular category.

Example 1 Describe the Canadian defence expenses data [1] (posted on Brightspace) using bar chart and pie chart.

Exercises

1. The following tables are the complaints against Air Canada and major U.S. airlines [2] in 2003.

Airline	June 2003	Dec. 2003	Canada Total	U.S. Total	Grand Total	Passengers (millions)
Air Canada	310	176	486	36	522	20.1
Airtran Airways	0	0	0	97	97	11.7
American West Airlines	0	0	0	168	168	20.1
American Airlines	0	5	5	781	786	88.8
Continental Airlines	0	0	0	371	371	38.9
Delta	4	3	7	656	663	84.3
Northwest Airlines	1	4	5	492	497	52
Southwest Airlines	0	0	0	106	106	74.8
United Airlines	6	2	8	548	556	66.2
US Airways	1	5	6	373	379	41.3

a. Describe the grand total number of complaints by airline using a pie chart.

b. Describe the number of passengers by airline using a pie chart.

4.5 Line Charts

When a quantitative variable is recorded over time at equally spaced intervals (such as daily, weekly, monthly, quarterly, or yearly), the data set forms a **time series**. Time series data are most effectively presented on a **line chart** with time as the horizontal axis. The idea is to try to discern a pattern or trend that will likely continue into the future, and then to use that pattern to make accurate predictions for the immediate future.

Example 1 The Dow Jones Industrial Average was monitored at the opening of trading for 10 days in 2019, with the results shown in the Table [3].

Day	1	2	3	4	5	6	7	8	9	10
DJIA	26,033	26,091	26,000	26,087	26,280	26,516	26,725	26,684	26,794	26,745

Construct the line chart.

Exercises

1. The following table is the projections for the age group 65-69 given by Statistics Canada.

Year	2006	2011	2016	2021	2026	2031
Populations (thousands)	1227.3	1513.1	1942.1	2184.7	2466.6	2527.6

Construct a line chart to illustrate the data. What is the effect of stretching and shrinking the vertical axis on the line chart?

2. The following data set is from a grocery store two weeks sales (thousand dollars):

Mon	Tue	Wed	Thu	Fri	Sat	Sun
15.3	16.2	17.8	14.6	18.8	23.2	24.7
16.2	15.9	17.3	15.2	19.7	22.3	25.7

Describe the data with a line chart.

4.6 Dotplots

For a small set of measurements, we can simply plot the measurements as points on a horizontal axis.

Example 1 Draw a dotplot for dataset 2, 6, 9, 3, 7, 6 3.

Exercises

1. A discrete variable can only take the values 1, 2, or 3. A set of 25 measurements on this variable is shown: 1 2 1 3 2 1 1 3 3 1 2 3 3 1 1 1 2 1 3 2 1 1 3 3 1
Draw a dotplot to describe the data.

4.7 Relative Frequency Histograms

A relative frequency histogram resembles a bar chart, but it is used to graph quantitative rather than qualitative data.

A relative frequency histogram for a quantitative data set is a bar graph in which the height of the bar shows "how often" (measured as a proportion or relative frequency) measurements fall in a particular class or sub-interval. The classes or subintervals are plotted along the horizontal axis.

As a rule of thumb, the number of classes should range from 5 to 12; the more data available, the more classes you need. The classes must be chosen so that each measurement falls into one and only one class. By using the **method of left inclusion**, and including the left class boundary point but not the right boundary point in the class, we eliminate any confusion about where to place measurement that happens to fall on a class boundary point.

Steps to construct a relative frequency histogram:

1. Choose the number of classes, usually between 5 and 12. The more data you

have, the more class you should use.

2. Calculate the approximate class width by dividing the difference between the largest and smallest values by the number of classes.

3. Round the approximate class width up to a convenient number.

4. If the data are discrete, you might assign one class for each integer value taken on by the data. For a large number of integer values, you may need to group them into classes.

5. Locate the class boundaries. The lowest class must include the smallest measurement. Then, add the remaining classes using the left inclusion method.

6. Construct a statistical table containing the classes, their frequencies, and their relative frequencies.

7. Construct the histogram like a bar graph, plotting class intervals on the horizontal axis and relative frequencies as the heights of the bars.

Example 1 (Pulse data posted at Brightspace) A group of 50 biomedical students records their pulse rates by counting the number of beats for 30 seconds and multiplying by 2. Construct a relative frequency histogram for the data.

Exercises

1. The ages in months at which 60 children were first enrolled in a preschool are listed:

38 47 32 55 42 40 35 34 39 50 30 34 41 33 37 35 43 30 32 39 39 41 46 32 33 40 36 35

45 45 48 41 40 42 38 36 43 30 41 46 31 48 46 36 36 36 40 37 50 31 38 47 32 55 42 40 35 34 39 50

Construct a relative frequency histogram to describe the data. Start the lower boundary of the first class at 30 and use a class width of 5 months.

2. The waiting time (minutes) in a medical clinic was recorded for 50 patients:

15 20 23 50 32 51 23 24 39 48 18 20 26 50 32 51 23 24 39 48 36 43 53 65 49 27 35 39 43 56 53 46 48 41 57 51 23 24 39 23 24 39 55 59 41 35 28 24 28 34

Construct a relative frequency histogram to describe the data. Start the lower boundary of the first class at 15 and use a class width of 10 minutes.

4.8 Steps of Spreadsheet Application

Microsoft Excel is a spreadsheet program in the Microsoft Office system. It is designed for a variety of analytical applications, including statistical applications. The current version of *Excel* in this note is *Excel* 2016, used in the Windows 7 environment. When the program opens, a **spreadsheet** appears (see Figure 4.2), containing rows and columns into which you can enter data. Tabs at the bottom of the screen identify the four spreadsheets available for use; when saved as a collection, these spreadsheets are called a **workbook**.

Pie charts, bar charts, line charts and **histograms** can all be created in *excel.* Data is entered into an *Excel* spreadsheet, including labels if needed. Highlight the

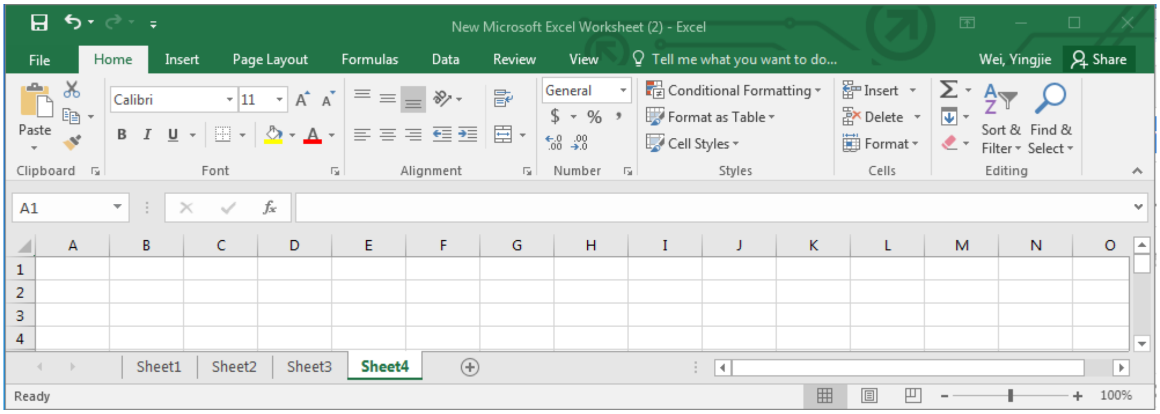


Figure 4.2: General View

data to be graphed, and then click the chart type that you want on the **Insert** tab in the **Charts** group. Once the chart has been created, it can be edited in a variety of ways to change its appearance.

Example 1 Pie and Bar Charts Construct the pie and bar charts for Example 1 in Section 4.3 using *Excel* spreadsheet.

1. Enter the categories into column A of the first spreadsheet and the frequencies into column B. You should have two columns of data, including the labels.

2. Highlight the data, using your left mouse button to *click-and drag* from cell A1 to cell B7 (normally written as **A1:B7**). Click the **Insert** tab and select **Pie** in the **Charts** group. In the drop-down list,you will see a variety of styles to choose from. Select the first option to produce the pie chart. Double-click on the title "Frequency" and change the title to " M&M Candies' Color."

3. **Editing the pie chart**: Once the chart has been created, use your mouse

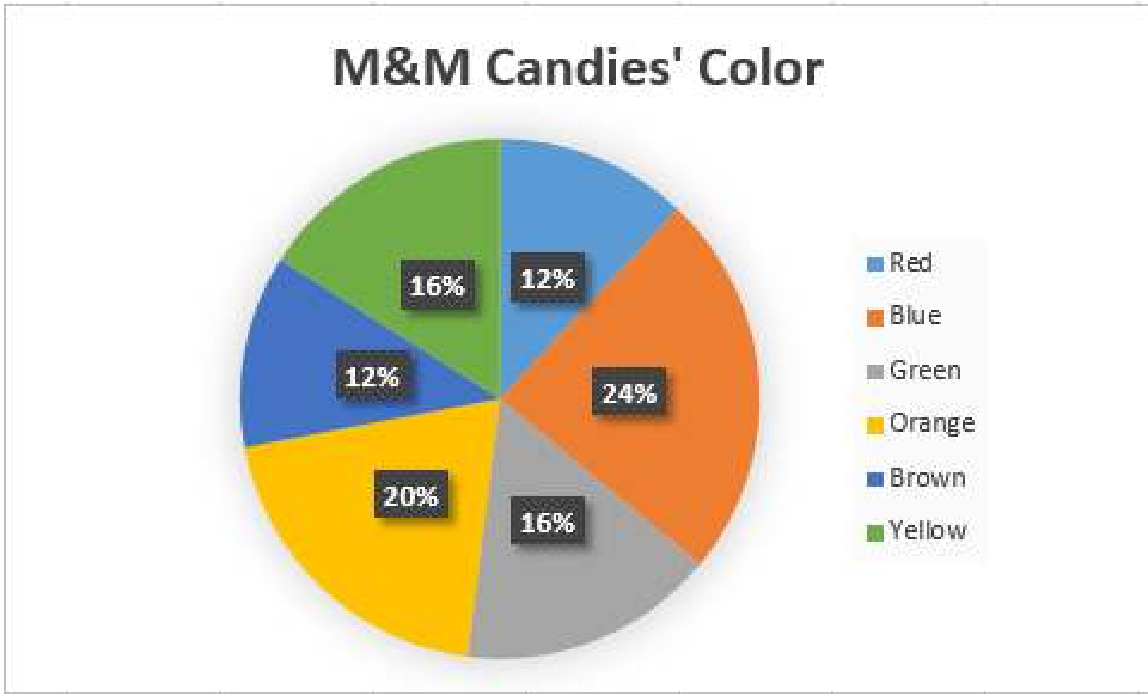


Figure 4.3: Original Pie Chart

to make sure that the chart is selected. You should see a green area above the tabs marked "Chart Tools." Click the **Design** tab, and look at the drop-down lists in the **Chart Layout** and **Chart Styles** groups. These lists allow you to alter the appearance of your chart. In Figure 4.3, the pie chart has been changed so that the percentages are shown in the appropriate sectors. By clicking on the legend, we have dragged it so that it is closer to the pie chart. The colors of the pie chart may not match with the colors of candies. When you let your cursor point to one area, e.g. the blue area, you will see: Series "Frequency" Point "Red" Value: 3(12%). You can select each sector to change it by right-click the area and fill it with the right color. In Figure 4.4, colors have been changed to match the colors of candies.

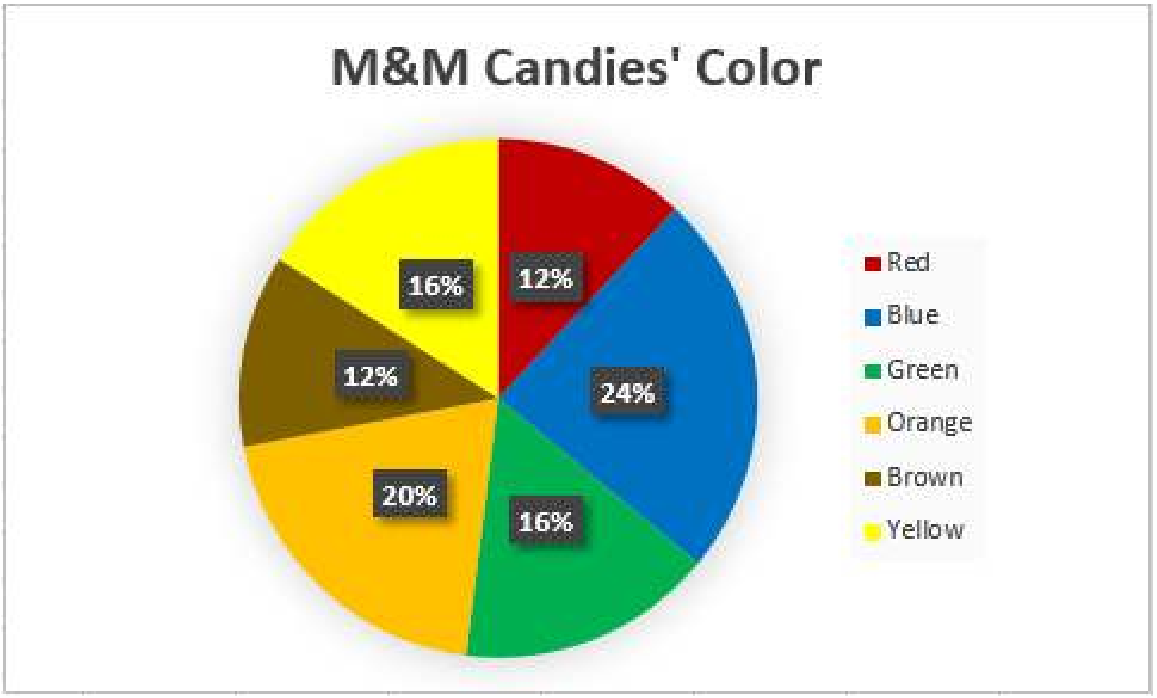


Figure 4.4: Modified Pie Chart

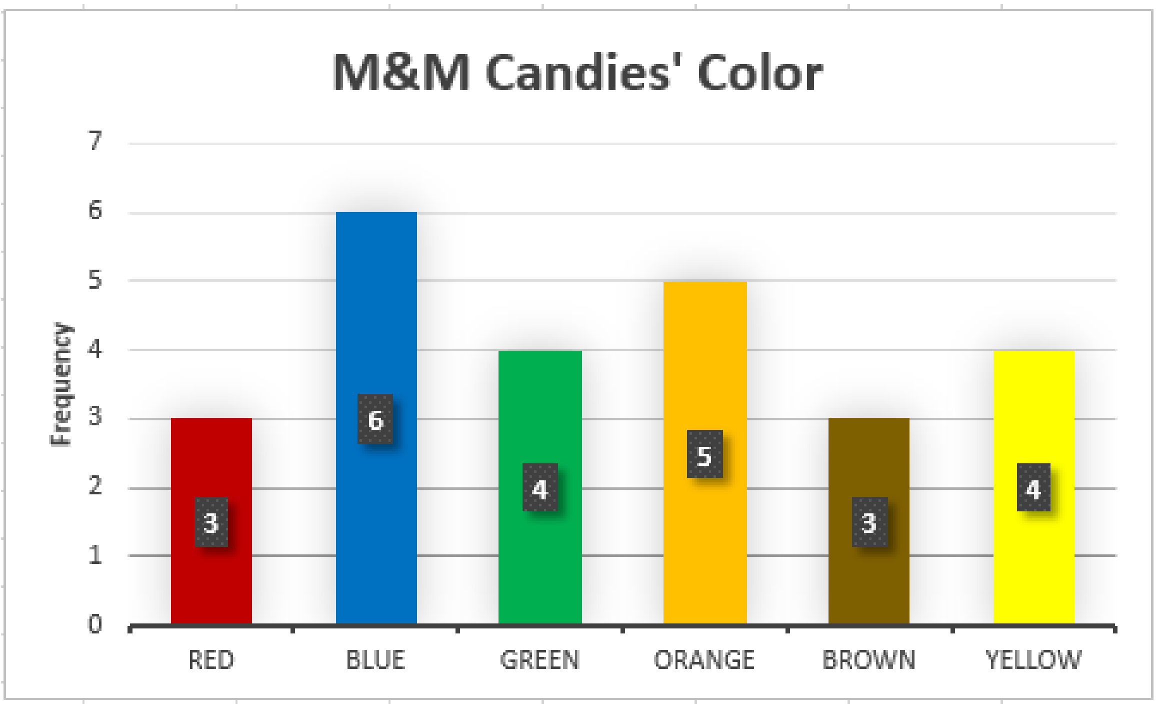


Figure 4.5: Bar Chart

4. Click on various parts of the pie chart (legend, chart area, sector) and a box with round and/or square handles will appear. Double-click, and a dialogue box will appear. You can adjust the appearance of the selected object or region in this box and click **OK** to confirm the adjustment. Click **Cancel** to exit the dialogue box without changes.

5. Still in the **Design** section, but in the **Type** group, click on **Change Chart Type** and choose the simplest Column type. Click **OK** to create a bar chart for the same data set, shown in Figure 4.5.

6. **Editing the bar chart:** Again, you can experiment with the various options in the **Chart Layout** and **Chart Styles** groups to change the look of the chart. You can click the entire bar chart ("Chart area") or the interior "Plot area" to stretch the chart. You can change colors by double-clicking on the appropriate region. We have chosen a recommended design and have deleted the legend. We also have added the axis title "Frequency" by clicking the **Chart Layouts** tab in the **Chart Tools**, and selecting **Axis Titles** ▹ **Primary Vertical**, and have entered "Frequency" in the text box.

Example 2 Line Charts The Dow Jones Industrial Average was monitored at the close of trading for 10 days in a recent year, with the results shown in Table 4.4. Construct the line chart.

1. Click the tab at the bottom of the screen marked "Sheet 2." Enter **Day** into

Table 4.4: Dow Jones Industrial Average

Day	1	2	3	4	5	6	7	8	9	10
DJIA	10,636	10,680	10,674	10,653	10,698	10,644	10,378	10,319	10,303	10,302

column A of this sheet and the **DJIA** into Column B. You should have two columns of data, including the labels.

2. Highlight the **DJIA** data in column B, using your left mouse button to *click-and-drag* from cell **B1** to cell **B11**(**B1:B11**). Click the **Insert** tab and select **Line** in the **Charts** group. In the drop-down list, you will see a variety of styles to choose from. Select the first option to produce the line chart.

3. **Editing the line chart:** You can experiment with the various options in the **Chart Layout** and **Chart Styles** groups to change the look of the chart. We have deleted the chart title and have added the axes titles. The line chart is shown in Figure 4.6

Example 3 Dotplots Draw the dotplots of dataset 2, 6, 9, 3, 7, 6, 3 using spreadsheet.

1. Count the frequency of each number and enter into excel spread sheet:

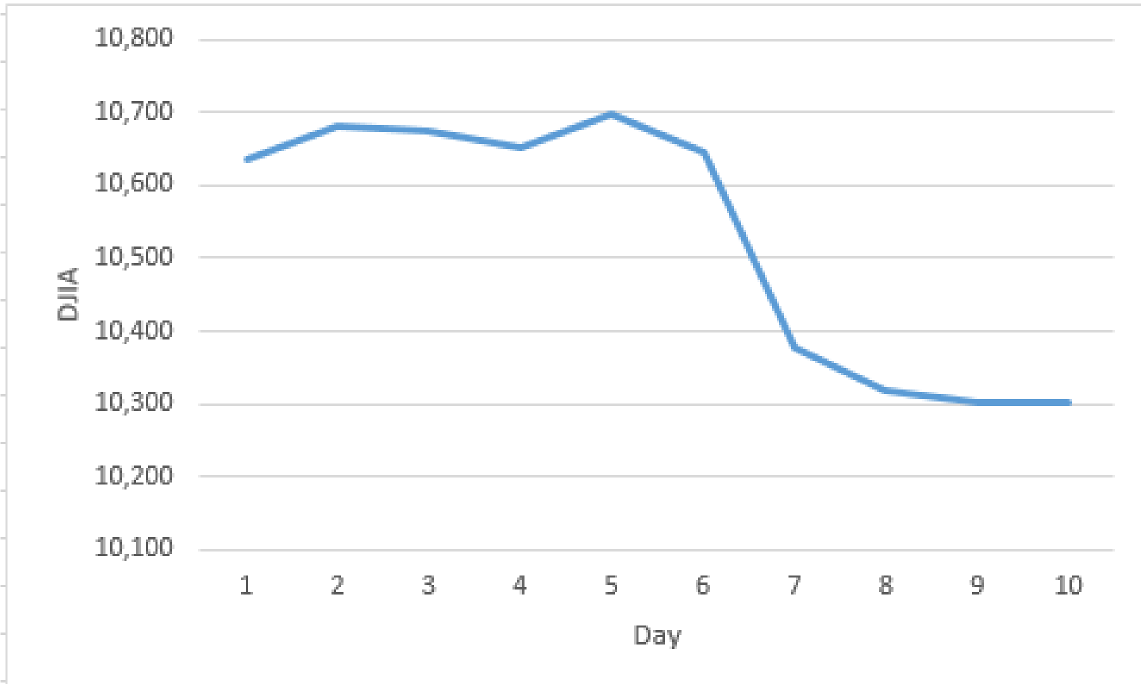
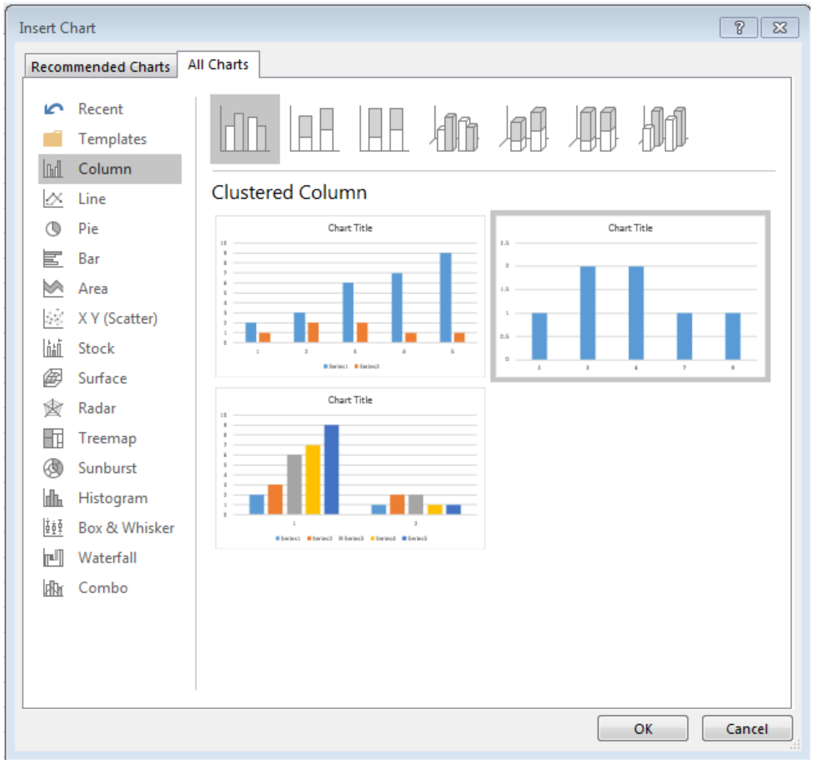


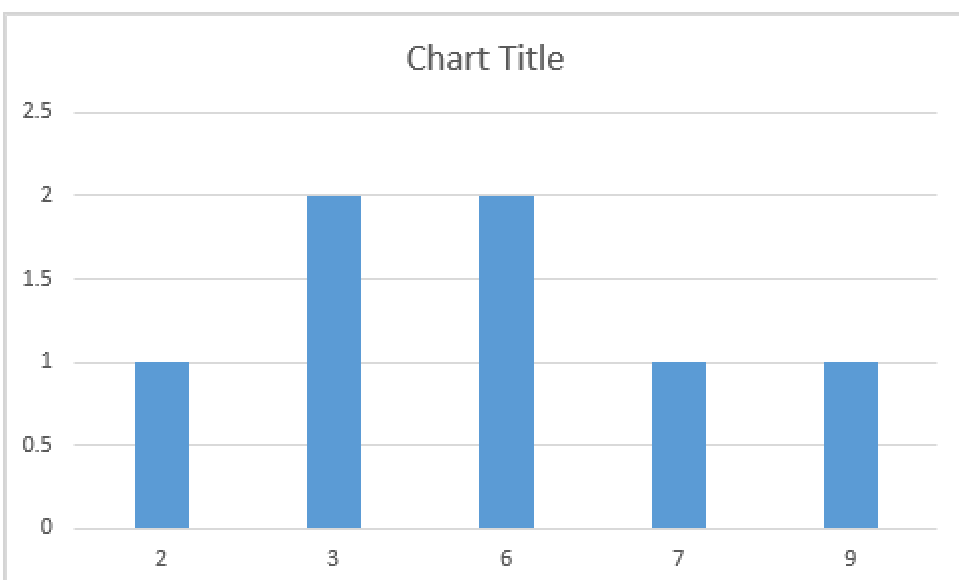
Figure 4.6: Dow Jones Industrial Average

	A	B	C
1	Numbers	Count	
2	2	1	
3	3	2	
4	6	2	
5	7	1	
6	9	1	
7			

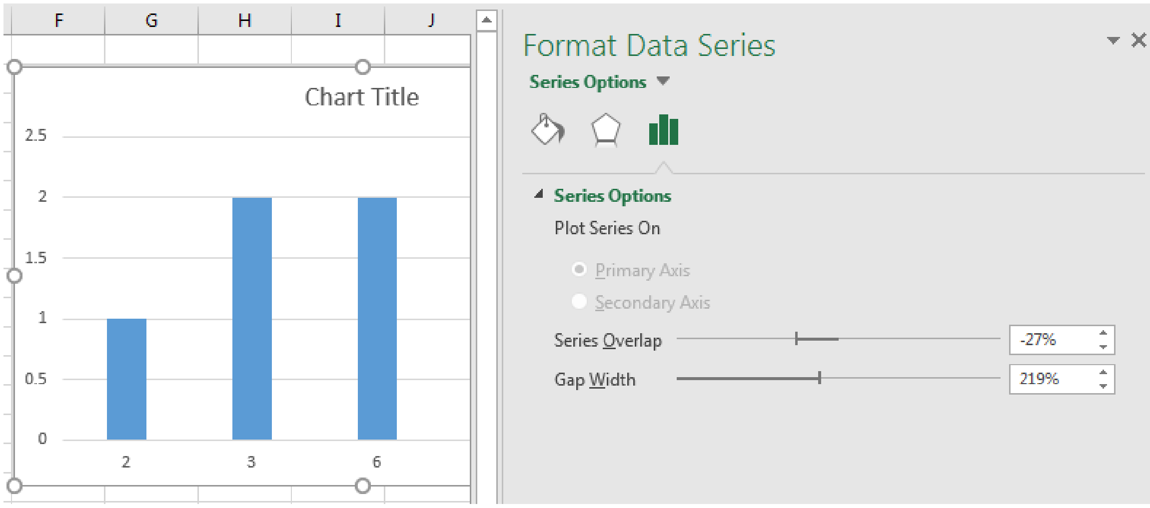
2. Select data area and click **Insert**, then put your cursor to Bar chart, and scroll down to **More Column Charts**, you will see a single bar chart:



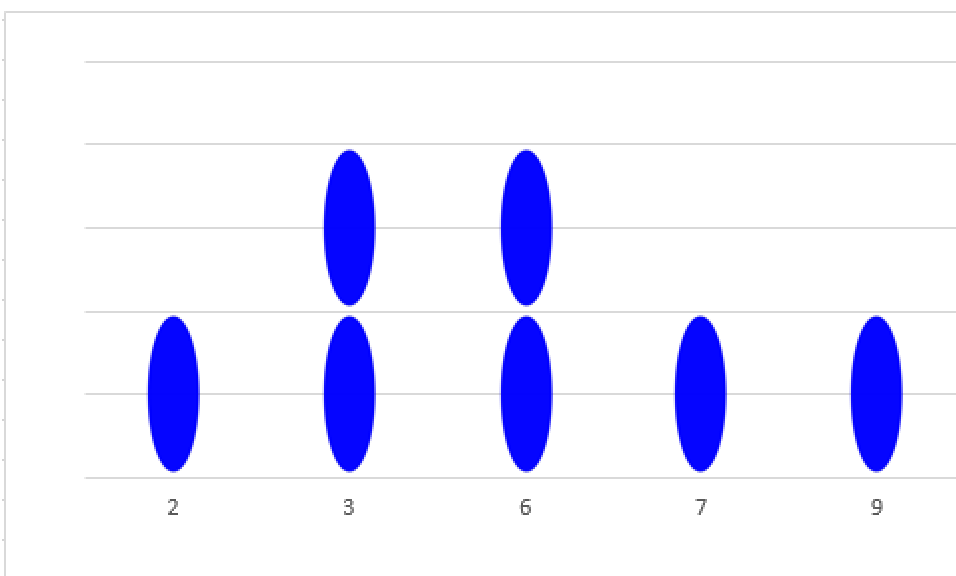
Select it and click **OK**, then you will get a bar chart:



3. Double click one bar, **Format Data Series** will show at the right of the excel:



Click **Fill & Line**, choose **Picture or texture fill**, click **Online**, type "dot" at the **Bing Image Search** bar and search, choose the image you want, and **Insert**, then choose **Stack and Scale with**, **Units/Picture 1**. You will get the dotplot:



You can change the **Series Overlap** and **Gap Width** under **Series Options**. You can also modify it under **Design** tab by clicking **Add Chart Element**. You can edit or delete the **Chart Title**.

Example 4 Frequency Histograms A group of 50 biomedical students records their pulse rates by counting the number of beats for 30 seconds and multiplying by 2.

80 70 88 70 84 66 84 82 66 42 52 72 90 7 96 84 96 86 62 78 60 82 88 54 66 66 80 88 56 104 84 84 60 84 88 58 72 84 68 74 84 72 62 90 72 84 72 110 100 58

Construct a relative frequency histogram for the data.

The **Analysis ToolPak** is a Microsoft Excel data analysis add-in, available in all modern versions of Excel beginning with Excel 2007. However, this add-in is not loaded automatically on Excel start, so you would need to load it first.

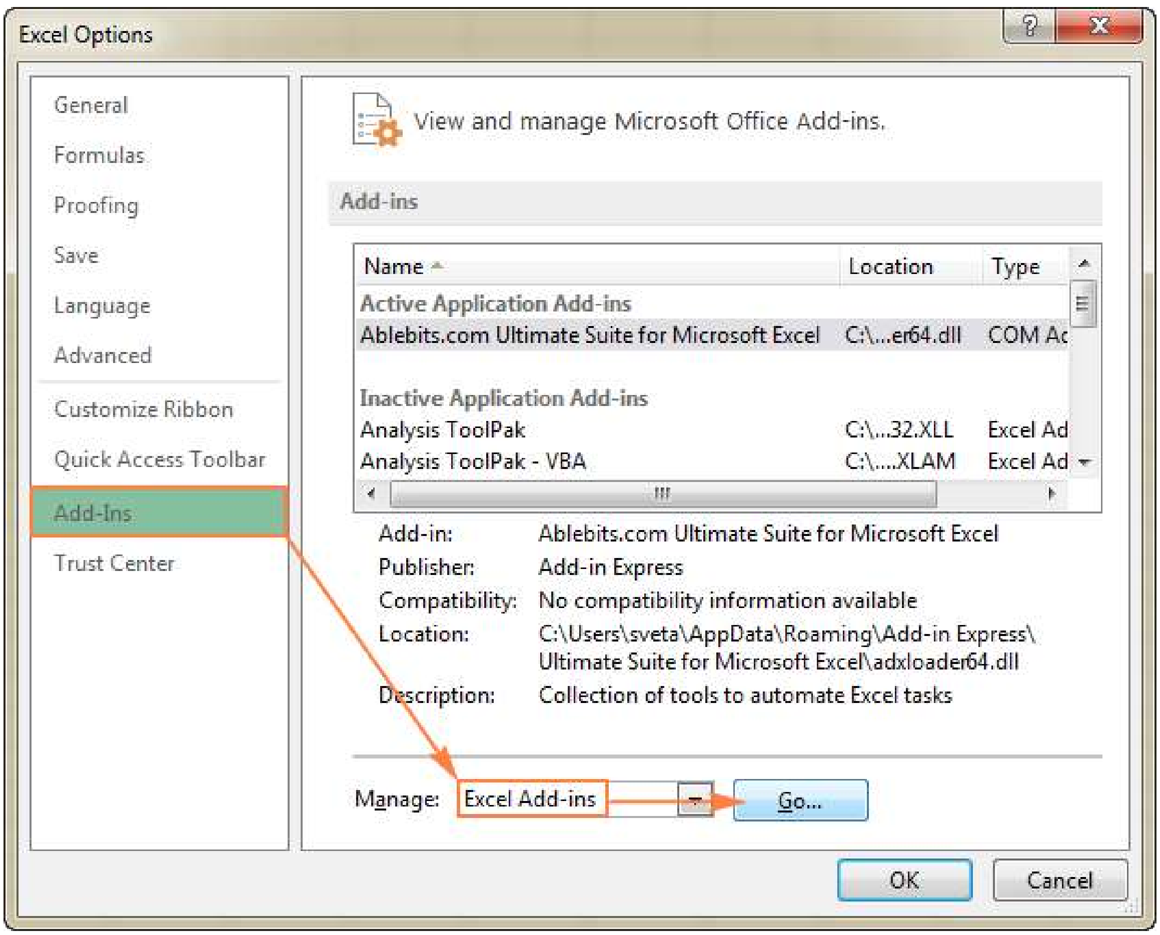
To add the **Data Analysis add-in** to your Excel, perform the following steps:

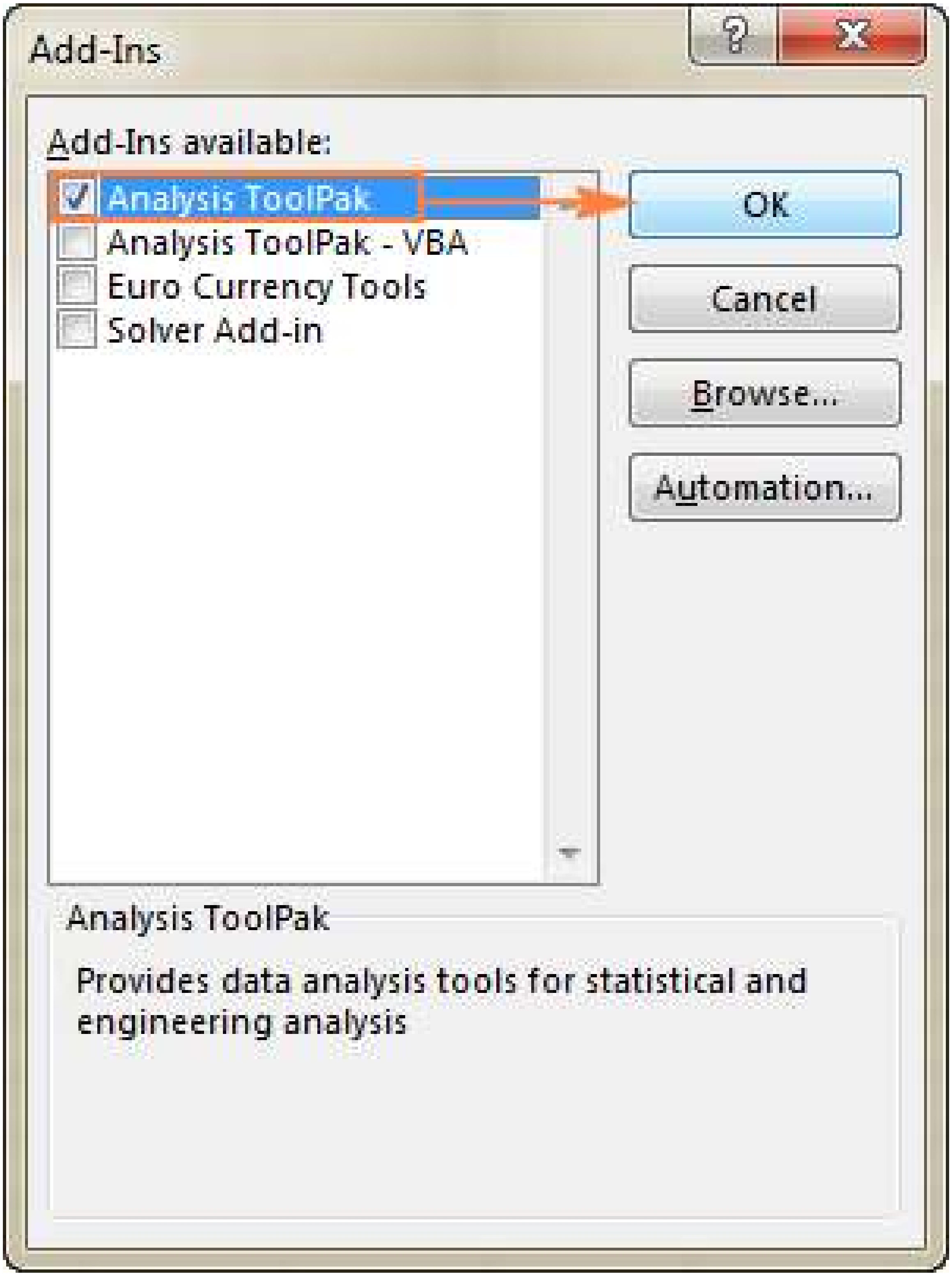
1. In Excel 2010, Excel 2013, Excel 2016, and Excel 2019, click **File** ▷ **Options**. In Excel 2007, click the **Microsoft Office** button, and then click **Excel Options**.

2. In the **Excel Options** dialog, click **Add-Ins** on the left sidebar, select **Excel Add-ins** in the **Manage** box, and click the **Go** button.

3. In the Add-Ins dialog box, check the **Analysis ToolPak** box, and click **OK** to close the dialog.

If Excel shows a message that the **Analysis ToolPak** is not currently installed on your computer, click **Yes** to install it.


Excel Options
General
Formulas
Proofing
Save
Language
Advanced
Customize Ribbon
Quick Access Toolbar
Add-Ins
Trust Center
View and manage Microsoft Office Add-ins.
Add-ins
Name
Location
Type
Active Application Add-ins
Ablebits.com Ultimate Suite for Microsoft Excel
C:\...er64.dll
Inactive Application Add-ins
Analysis ToolPak
C:\...32.XLL
Analysis ToolPak - VBA
C:\....XLAM
Add-in:
Ablebits.com Ultimate Suite for Microsoft Excel
Publisher:
Add-in Express
Compatibility:
No compatibility information available
Location:
C:\Users\sveta\AppData\Roaming\Add-in Express\
Ultimate Suite for Microsoft Excel\adxloader64.dll
Description:
Collection of tools to automate Excel tasks
Manage:
Excel Add-ins
Go...
OK
Cancel


Add-Ins
Add-Ins available:
Analysis ToolPak
Analysis ToolPak - VBA
Euro Currency Tools
Solver Add-in
OK
Cancel
Browse...
Automation...
Analysis ToolPak
Provides data analysis tools for statistical and engineering analysis

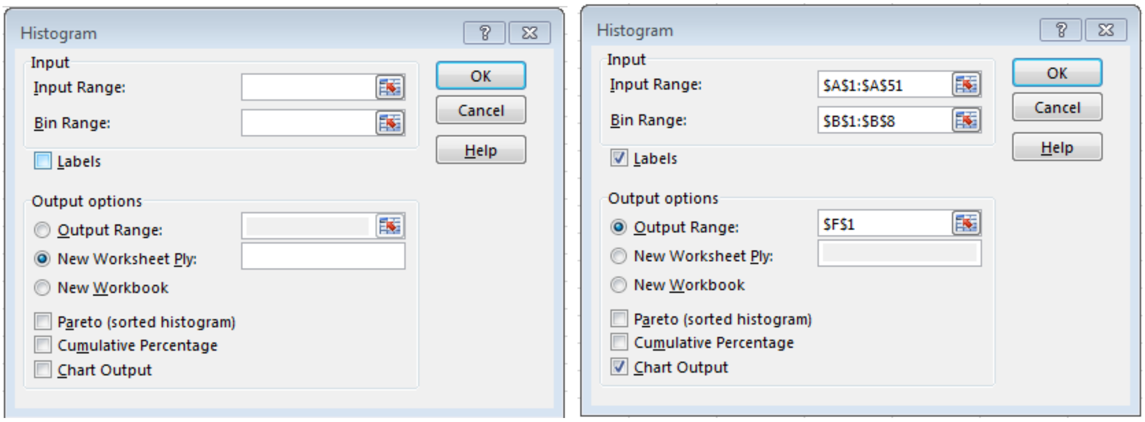
Now, the Analysis ToolPak is loaded in your Excel, and its command is available in the **Analysis** group on the **Data** tab.

4. Click the tab at the bottom of the screen marked "Sheet 3." Enter the data into the first column of this spreadsheet and include the label "Pulse " in the first cell "A1."

5. *Excel* refers to the maximum value for each class interval as a **bin**. This means that *Excel* is using a **Method of Right Inclusion**, which is slightly different from the method we presented before. For this example, we choose to use the class intervals >40-50, >50-60, >60-70, >70-80, >80-90, >90-100, >100-110. Enter the *bin values* (50, 60, $\cdots$, 110), which is a collection of upper boundaries of each interval in ascending order, into the second column of the spreadsheet, labeling them as "Pulse Class" in cell **B1**.

6. Select **Data** $\triangleright$ **Data Analysis** $\triangleright$ **Histogram** and click **OK**. The Histogram dialogue box will appear, as shown in Figure 4.7 (a).

7. Highlight or type in the appropriate **Input Range** and **Bin Range** for the data. Notice that you can click the minimize button on the right of the box before you click-and-drag to highlight. Click the minimize button again to see the entire dialogue box. The **Input Range** will appear as $A$1:$A$51, with the dollar sign indicating an absolute cell range. Make sure to click the "Labels" and "Chart Output" check boxes. Pick a convenient cell location for the output (we picked F1), shown in

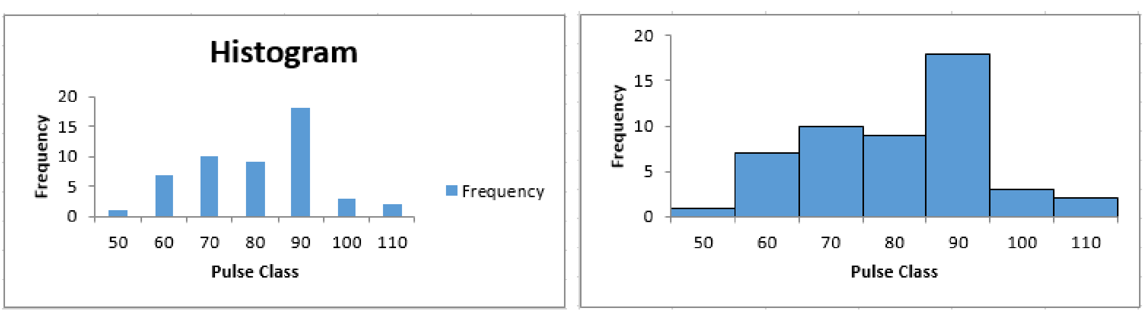


(a) Blank (b) Filled

Figure 4.7: Information needed by a histogram

Figure 4.7 (b), and click **OK**. The frequency table and histogram will appear in the spreadsheet. The histogram, Figure 4.8 (a), doesn't appear like we wanted.

8. **Editing the histogram**: Click on the frequency legend entry and press the **Delete** key. Then select the Data Series by *double-clicking* on a bar. In the Series Options box that appears, change the **Gap Width** to **0** and click **Close**, Stretch the graph by dragging the lower right corner, and edit the colors, titles, and labels if necessary to finish your histogram, as shown in Figure 4.8 (b).



(a) Original (b) Modified

Figure 4.8: Histogram of 50 students pulse rates data

9. You can save your Excel workbook for use at a later time using **File** ▹ **Save** or **File** ▹ **Save As** and naming it, for example, "Chapter 4."

Data Sources

[1] National Defence and Canadian Forces Budget-Budget 2002-2003 National Defence. Reproduced with permission of the Minister of the Public Works and Government Services, 2009. http:/www.forces.gc.ca/site/about/$budget_e$.asp

[2] Canadian complaints against U.S. airlines. http://www.cta-otc.gc.ca/cta-otc2000/report-rapport/2004/findings1_e.html; Canadian complaints against Air Canada: http://www.cta-otc.gc.ca/cta-otc2000/report-rapport/2004/findins_e.html#2; U.S. airline passengers: http://airconsumer.ost.dot.gov/reports/2004/0402atcr.doc; Air Canada passengers: http://www.statca.ca/english/freepub/51-004-XIB/0050451-004-XIB.pdf

[3] https://www.marketwatch.com/investing/index/djia/charts (Dow Jones)

4.9 Chapter 4 Review Questions

1. Classify the following for their measurement scale: Qualitative(nominal, ordinal), Quantitative (discrete, continuous)

a. Temperature of all indoor pools in the city of Regina.

b. Ages of students enrolled in a fitness course at the YMCA.

c. Rankings of a comedian's performance.

d. Marital status of the full-time staff at University of Regina.

2. The value of a quantitative variable is measured once a year for a 10-year period. The data are as follows:

Year	Measurement	Year	Measurement
1	61.5	6	58.2
2	62.3	7	57.5
3	60.7	8	57.5
4	59.8	9	56.1
5	58.0	10	56.0

a. Create a line chart to describe the variable as it changes over time.

b. Describe the measurements using the chart construct in part a.

3. The data below represent the ages (Current year - Year on penny) of a set of 50 pennies:

5 1 9 1 2 20 0 25 0 17 1 4 4 3 0 25 3 3 8 28 5 21 19 9 0 5 0 2 1 0 0 1 19 0 2 0 20 16 22 10 19 36 23 0 1 17 6 0 5 0

Draw a relative frequency histogram to describe the distribution of penny ages. Start the lower boundary of the first class at 0 and use a class width of 4 years.

4. The following data are a sample of blood types were taken in a two hour time frame at a Canadian Blood Services branch:

O O AB A B O A A B O A A B O B O A B O O A A AB B O A A O A A O B O AB O A O A

a. Create a bar chart to describe it.

b. Create a pie chart to describe it.

5. Consumer Reports evaluates products for consumers. The file CompactSUV contains the data for 15 compact sports utility vehicles (SUVs) from the 2018 model line (Consumer Reports website):

Table 4.5

Consumer Reports Data Set for 15 Compact Sports Utility Vehicles

Make	Model	Overall Score	Recommended	Owner Satisfaction	Overall Miles Per Gallon	Acceleration (0-60) Sec
Subaru	Forestser	84	Yes	+	26	8.7
Honda	CRV	83	Yes	++	27	8.6
Toyota	Rav4	81	Yes	++	24	9.3
Nissan	Rogue	73	Yes	+	24	9.5
Mazda	CX-5	71	Yes	++	24	8.6
Kia	Sportage	71	Yes	+	23	9.6
Ford	Escape	69	Yes	0	23	10.1
Volkswagen	Tiguan Limited	67	No	0	21	8.5
Volkswagen	Tiguan	65	No	+	25	10.3
Mitsubishi	Outlander	63	No	0	24	10.0
Chevrolet	Equinox	63	No	0	31	10.1
Hyundai	Tuscon	57	No	0	26	8.4
GMC	Terrain	57	No	0	22	7.2
Jeep	Cherokee	55	No	-	22	10.9
Jeep	Compass	50	No	0	24	9.8

Make – manufacturer

Model – name of the model

Overall score – awarded based on a variety of measures, including those in this data set

Recommended – Consumer Reports recommends the vehicle or not

Owner satisfaction – satisfaction on a five-point scale based on the percentage of owners who would purchase the vehicle again (- -, -, 0, +, + +).

Overall miles per gallon – miles per gallon achieved in a 150-mile test trip

Acceleration (0-60 sec) – time in seconds it takes vehicle to reach 60 miles per hour from a standstill with the engine idling

a. How many variables are in the data set?

b. Which of the variables are qualitative nominal and qualitative ordinal, and which are quantitative discrete and quantitative continuous?

c. What percentage of these 15 vehicles are recommended?

d. What is the average of the overall miles per gallon across all 15 vehicles?

e. For owner satisfaction, construct a bar chart to demonstrate it.

f. Show the frequency distribution for acceleration using the following method: start the lower boundary of first class at 7.0, and the class width of 1.0.

Chapter 5

Analyze Data with Numerical Measures

Graphs can help you describe the basic shape of a data distribution. However, there are limitations. On one hand, you will need some equipments. On the other hand, graphs are somewhat imprecise for use in statistical inference. To overcome these limitations, we can use numerical measurements, which can be calculated for either a sample or a population of measurements. These measures are called parameters when associated with the population, and they are called statistics when calculated from sample measurements.

Numerical descriptive measures associated with a population of measurements are called **parameters**; those calculated from sample measurements are called **statistics**.

5.1 Measures of Centre

We have introduced the histograms to describe the distribution of a set of measurements on a quantitative variable X. The horizontal axis displays the values of X, and the data are distributed along this horizontal line. We will introduce an important numerical measure, **measure of centre**, a measure along the horizontal axis that locates the centre of the distribution. The arithmetic average of a set of measurements is a very common and useful measure of centre. This measure is often called the **arithmetic mean**, or the **mean**, of a set of measurements. To distinguish between the mean for the sample and the mean for the population, we will denote a sample mean as $\bar{x}$ and the population mean as μ.

The **arithmetic mean** or **mean** of a set of n measurements is equal to the sum of the measurements divided by n.

Suppose there are n measurements on the variable X, $x_1, x_2, \cdots, x_n$. To add n measurement together, we use the shorthand notation:

$$\sum_{i=1}^{n} x_i \texttt{ which means } \quad x_1 + x_2 + \cdots + x_n$$

or

$$\sum x_i \texttt{ which means the sum of all the X measurements.}$$

Using this notation, the formula for the mean can be written as:

Sample mean:

$$\bar{x} = \frac{\sum x_i}{n}, \tag{5.20}$$

where n is sample size.

Population mean:

$$\mu = \frac{\sum x_i}{N}, \tag{5.21}$$

where N is the population size.

The **median** m of a set of n measurements is the value of x that falls in the middle position when the measurements are ordered from smallest to largest.

The value **0.5(n+1)** indicates the **position of the median** in the ordered data set. If the position of the median is a number that ends in the value **0.5**, we need to average the two adjacent values.

Example 1 For the $n = 5$ measurements 3, 9, 6, 8, 5. Find the sample mean and median.

Solution Sample mean: $\bar{x} = \frac{\sum x_i}{n} = \frac{3+9+6+8+5}{5} = 6.2$.

Median: rank the $n = 5$ measurements from smallest to largest: 3, 5, 6, 8, 9. The middle observation is in the centre of the set, or the position of the median, $0.5(5 + 1) = 3$, i.e. the median is the 3rd ordered number, $m = 6$.

Example 2 For the $n = 6$ measurements 3, 9, 6, 8, 5, 11. Find the sample mean and median.

Solution Sample mean: $\bar{x} = \frac{\sum x_i}{n} = \frac{3+9+6+8+5+11}{6} = 7$.

Median: rank the $n = 6$ measurements from smallest to largest: 3, 5, 6, 8, 9, 11. The position of the median is $0.5(n+1) = 0.5(6+1) = 3.5$. The middle observation is in the middle of the 3rd and 4th data. To average these two values to get the median, $m = \frac{6+8}{2} = 7$. It happens that mean and median are the same in this dataset, but it not always the case.

The **mode** is the measurement that occurs most frequently.

Example 3 The data set: 2, 4, 9, 8, 8, 5, 3.

The mode is 8, which occurs twice.

Example 4 The set: 2, 2, 9, 8, 8, 5, 3.

There are two modes—8 and 2, we call it bimodal set.

Example 5 The set: 2, 4, 9, 8, 5, 3.

There is no mode (each value is unique).

The mode is generally used to describe large data sets, whereas the mean and median are used for both large and small data sets.

Exercises

1. The following data set is the number of liters of milk purchased by 25 households:

0 0 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 4 4 4 5

Find the mean, median, and mode (if it has).

2. You are given 8 measures: 3, 2, 4, 6, 5, 4, 7, 9. Find mean and median.

3. The following table (NASDAQ Salmon Index - Prices Per Weight Class [1]) shows the prices of salmon based on different classes:

Weight Class(kg)	Price (NOK/kg)
1-2	42.45
2-3	49.59
3-4	56.86
4-5	60.80
5-6	65.05
6-7	68.26
7-8	68.50
8-9	67.83
9+	63.44

Find the average price and the median price of salmon.

4. In an experiment, the time on task was recorded for 10 subjects under a 5-minute time constraint. These measurements are in seconds:

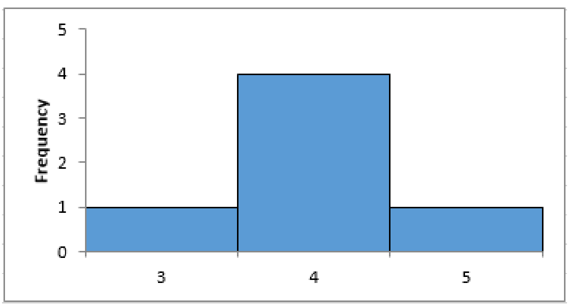
199 222 253 293 162 183 193 296 284 228

a. Find the average time on task.

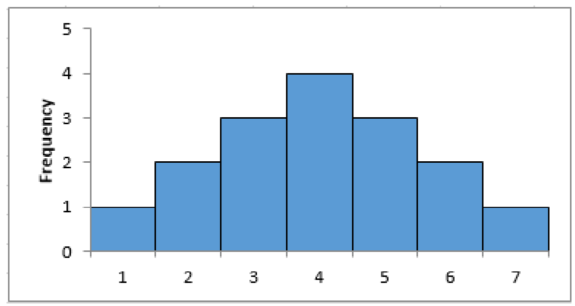
b. Find the median time on task.

5.2 Measures of Dispersion

Different data sets may have the same centre but look different because of the way the numbers spread out from the centre. Consider the two distributions shown in Figure 5.1. Both distributions are centred at x=4, but there is a big difference in the way the measurements spread out, or vary. The measurements in Figure 5.1(a) vary from 3 to 5; in Figure 5.1(b), the measurements vary from 1 to 7.



(a) Data 1



(b) Data 2

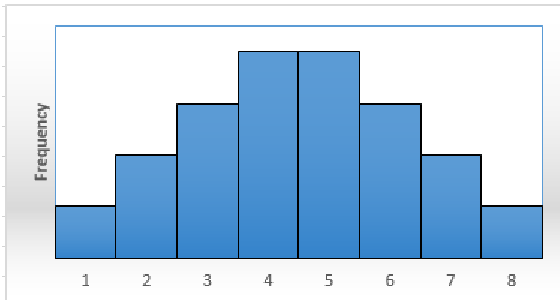
Figure 5.1: Variability or Dispersion of Data

Dispersion or **variability** is a very important characteristic of data. For example, if you are manufacturing bearing, extreme variation in the diameters would cause a high percentage of defective products.

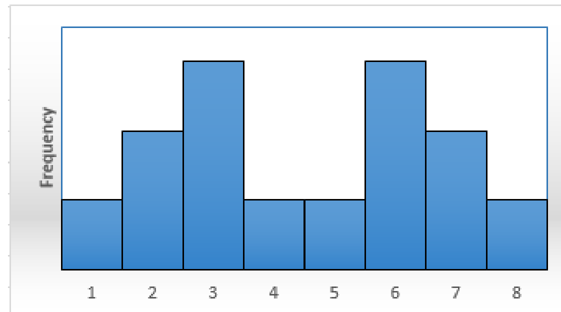
Measures of variability can help you create a mental picture of the spread of the data. The simplest measure of variation is the **range**.

The **range**, R, of a set of n measurements is defined as the difference between the largest and smallest measurements.

The range is easy to calculate, easy to interpret, and is an adequate measure of variation for small data sets. But, for large data sets, the range is not an adequate measure of variability.



(a) Data 1



(b) Data 2

Figure 5.2: Distribution with equal range and unequal variability

For example, the two relative frequency distributions in Figure 5.2 have the same range but very different shapes and variability.

Is there a measure of variability that is more sensitive than the range?

Consider the sample measurements 4, 6, 2, 7, 5, displayed as a dotplot in Figure 5.3. The mean of these five measurements is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{24}{5} = 4.8$$

as indicated on the dotplot.

The horizontal distance between each dot (measurement) and the mean $\bar{x}$ will help to measure the variability. If the distances are large, the data are more spread out or variable than if the distances are small. If x_i is a particular dot, then the deviation of that measurement from the mean is $x_i - \bar{x}$. Measurements to the right of the mean

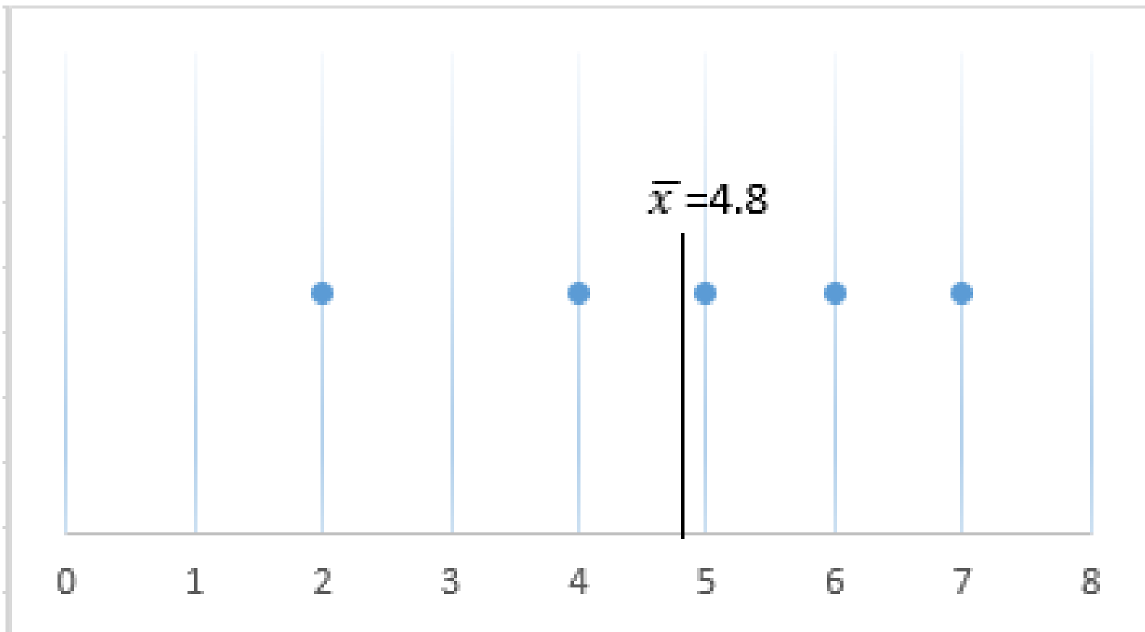


Figure 5.3: Dotplot showing the deviations of points

produce positive deviations, and those to the left produce negative deviations. The sum of all the deviations will be always 0. Therefore, we can not use the average of the deviations to evaluate it. We overcome the difficulty caused by the signs of the deviations by working with their sum of squares.

The **variance of a population** of N measurements is the average of the squares of the deviations of the measurements about their mean μ. The population variance is denoted by σ^2 and is given by the formula

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}. \tag{5.22}$$

Normally, we will not have all the population measurements available but will need to calculate the **variance of a sample** of n measurements.

The **variance of a sample** of n measurements is the sum of the squared deviations of the measurements about their mean $\bar{x}$ divided by $n-1$. The sample variance is denoted by s^2 and is given by the formula

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}. \tag{5.23}$$

The **standard deviation** of a set of measurements is equal to the positive square root of the variance.

The measures used in sample and population:

n: Number of measurements in the sample	N: Number of measurements in the population
s^2: Sample variance	σ^2: Population variance
$s = \sqrt{s^2}$: Sample standard deviation	$\sigma = \sqrt{\sigma^2}$: Population standard deviation

Table 5.1: Basic computations of a variance

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
4	-0.8	0.64
6	1.2	1.44
2	-2.8	7.84
7	2.2	4.84
5	0.2	0.04
$\sum x_i$=24	$\sum(x_i - \bar{x})$=0	$\sum(x_i - \bar{x})^2$=14.8

For the set of $n = 5$ sample measurements presented in Table 5.1, the square of the deviation of each measurement is recorded in the third column. We obtain

$$\sum (x_i - \bar{x})^2 = 14.8$$

and the sample variance is

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{14.8}{4} = 3.7.$$

Taking the square root of the variance, we obtain the **standard deviation**, 1.92, which returns the measure of variability to the original units of measurement.

Normally we use the following shortcut method to calculate σ^2 and s^2.

$$\sigma^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2, \tag{5.24}$$

and

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}, \tag{5.25}$$

where

$\sum x_i^2$ = Sum of the squares of the individual measurements.

$(\sum x_i)^2$ = Squares of the sum of the individual measurements.

The sample standard deviation, s, is the positive square root of s^2. The population standard deviation, σ, is the positive square root of σ^2.

You should keep the following points in mind:

- The value of s (σ) is always greater than or equal to zero.

- The larger the value of s^2 (σ^2) or s (σ), the greater the variability of the data set.

- If $s^2(\sigma^2)$ or $s(\sigma)$ is equal to zero, all the measurements must have the same value.

- In order to measure the variability in the same units as the original observation, we compute the standard deviation $s = \sqrt{s^2}$ ($\sigma = \sqrt{\sigma^2}$).

Exercises

1. You are given a set of sample measurements: 2, 6, 3, 1, 4.

a. Find the range.

b. Calculate the mean $\bar{x}$.

c. Calculate the sample variance s^2 using the formula.

d. Calculate the sample standard deviation s.

2. You are given a set of measurements: 1, 2, 2, 1, 3, 1, 4, 3.

a. Find the range.

b. Calculate the mean $\bar{x}$.

c. Calculate the sample variance s^2 using the formula.

d. Calculate the sample standard deviation s.

3. **Electricity Bill in Regina** The monthly electricity bills for a household in Regina were recorded for 12 consecutive months starting in January 2018:

Month	Amount($)	Month	Amount($)
Jan	204.94	Jul	276.70
Feb	180.00	Aug	309.70
Mar	178.23	Sept	312.40
Apr	176.43	Oct	238.66
May	165.12	Nov	225.47
Jun	236.72	Dec	222.23

a. Find the range of the electricity bills for the year 2018.

b. Calculate the average monthly eclectricity bill for the year 2018.

c. Calculate the variance of the 2018 electricity bills.

d. Calculate the standard deviation for the 2018 electricity bills.

4. **Gas Bill in Moose Jaw** The monthly gas bills for a household in Moose Jaw were recorded for 12 consecutive months starting in January 2018:

Month	Amount($)	Month	Amount($)
Jan	74.94	Jul	48.70
Feb	68.50	Aug	39.75
Mar	68.23	Sept	42.42
Apr	56.43	Oct	38.66
May	45.12	Nov	75.47
Jun	46.72	Dec	82.23

a. Find the range of the gas bills for the year 2018.

b. Calculate the average monthly gas bill for the year 2018.

c. Calculate the variance of the 2018 gas bills.

d. Calculate the standard deviation for the 2018 gas bills.

Data Sources [1] Nasdaq: https://salmonprice.nasdaqomxtrader.com/public/report?0

5.3 Chapter 5 Review Questions

1. For the following sample data: 0 1 3 4 4 5 6 6 6 7 7 8

Find the sample mean, median, mode, sample standard deviation, and the range.

2. The following data are the daily sales (in thousands dollars) of a grocery store:

Days	1	2	3	4	5	6	7
Sales	23.4	25.5	24.5	24.9	26.8	31.7	33.2

a. Find the sample mean and median.

b. Calculate the sample variance s^2 and standard deviation s.

c. Find the range.

3. The scores for a business math test are as follows:

87 89 76 66 85 88 92 94 77 96 79 69 95 18 58 82 88 83 91 50

Compute the mean, median, mode, and standard deviation for this data.

4. The following data give the number of hotdogs consumed in a hotdog-eating-contest by a sample of 10 contestants.

21 17 32 8 20 15 17 23 9 18

a. Find the sample mean, median.

b. Calculate sample variance and sample standard deviation.

c. Calculate the range.

Chapter 6

Linear Regression

Sometimes the data that are collected consist of observations for two variables on the same experimental unit. Special techniques that can be used in describing these variables will help you identify possible relationships between them.

6.1 Graphs for Bivariate Data

When you have one qualitative and one quantitative variable that have been measured in two different populations or groups, you will have several ways to display the data. You can use two side-by-side pie charts, or a bar chart in which the bars for the two populations are placed side by side, or you can use a stacked bar chart, in which the bars for each category are stacked on top of each other.

Example 1 Table 6.1 shows the status of students in an introductory statistics class at University of Regina(UofR) and University of Saskatchewan (UofS).

Table 6.1: Status of Students in a Statistics Class at UofR and UofS

	1st Year	2nd Year	3rd Year	4th Year	Grad Student
Frequency(UR)	5	23	32	35	10
Frequency(US)	10	35	24	25	16

a. Create side-by-side pie charts to describe these data.

b. Create a side-by-side bar chart to describe these data.

c. Draw a stacked bar chart to describe these data.

Exercises

1. **Service Satisfaction** Male and female respondents to a questionnaire about the service satisfaction level are categorized into three groups according to their answers:

	Beyond Expectation	Meet Expectation	Not Meet Expecation
Men	30	58	12
Women	22	52	24

a. Create side-by-side pie charts to describe these data.

b. Create a side-by-side bar chart to describe these data.

c. Draw a stacked bar chart to describe these data.

2. The price of living in Canada is continually increasing, as demonstrated by the consumer price indexes (CPIs) for food and shelter. These CPIs are listed in the table for the years 2009 - 2018 [1].

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Food	121.4	123.1	127.7	130.8	132.4	135.5	140.5	142.6	142.7	145.3
Shelter	121.6	123.3	125.6	127.1	128.7	132.2	133.7	135.8	138.1	140.9

a. Create a side-by-side bar chart to describe CPIs over time.

b. Draw two line charts on the same set of axes to describe the CPIs over time.

c. What conclusions can you draw using the two graphs in parts a and b? Which is the most effective?

6.2 Scatterplots for Two Quantitative Variables

When both variables to be displayed on a graph are quantitative, one variable (x) is plotted along the horizontal axis and the second (y) along the vertical axis. Hence, the graph takes the form of a plot on the (x, y) axes. Each pair of data values is plotted as a point on this two-dimensional graph, called a **scatterplot**. It is the two-dimensional extension of the dotplot we used to graph one quantitative variable.

You can describe the relationship between two variables, x and y, using the patterns shown in the scatterplot.

- **What type of pattern do you see?** Is there a constant upward or downward trend that follows a straight-line pattern? Is there a curved pattern? Is there no pattern at all, but just a random scattering of points?

- **How strong is the pattern?** Do all of the points follow the pattern exactly? Or is the relationship only weakly visible?

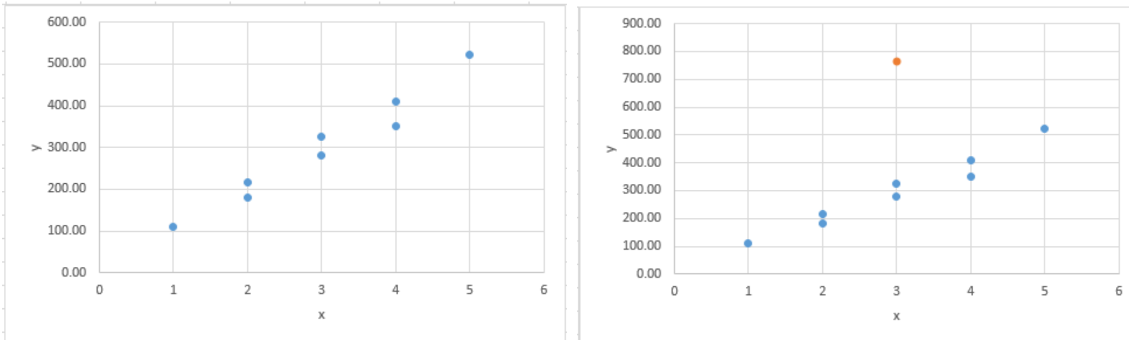
- **Are there any unusual observations?** An outlier is a point that is far from the cluster of the remaining points. Do the points cluster into groups? If so, is there an explanation for the observed groupings?

Example 1 The number of household members, x, and the amount spent on groceries per week, y, are recorded for eight households in Regina. Draw a scatterplot of theses eight data points.

x	2	2	3	4	4	3	1	5
y	181.40	215.85	289.76	350.50	378.90	305.50	110.58	521.64

Solution Label the horizontal axis x and vertical axis y. Plot the points using the coordinates (x, y) for each of the eight pairs. The scatterplot in Figure 6.1 (a) showed eight pairs marked as dots. You can see a pattern with these eight data pairs. The cost of weekly groceries increases with the number of household members in an apparent straight-line relationship.

Suppose that you have a ninth household with three members spent \$765.24 that week. This observation is shown as orange dot in Figure 6.1 (b). This household may have a special celebration, such as a party, in that week.



(a) Eight Households (b) Nine Households

Figure 6.1: Scatterplots for Weekly Groceries

6.3 Numerical Measures for Quantitative Bivariate Data

A constant rate of increase or decrease is perhaps the most common pattern found in bivariate scatterplots. It is a straight line with the data points lying both above and below the line and within a fixed distance from the line. We say that the two variables exhibit a *linear relationship*.

Example 1 The data in Table 6.2 are the size of the living area (in m^2), x, and the selling price ($thousands), y, of 5 residential properties. The scatterplot in Figure 6.2 shows a linear pattern.

Table 6.2: Living Area and Selling Price of 5 Properties

Properties	x (m^2)	y ($thousands)
1	126.3	178.5
2	134.7	188.6
3	137.5	168.8
4	144.0	229.8
5	148.6	205.2

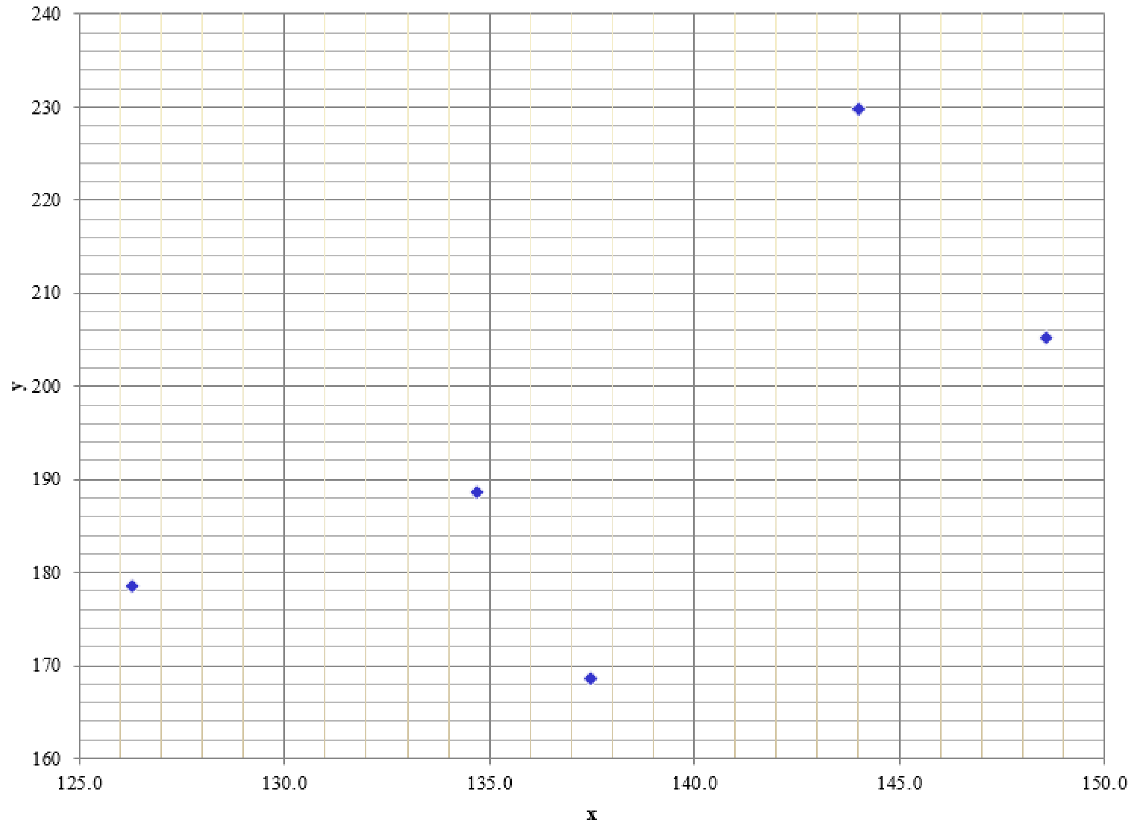


Figure 6.2: Scatterplot of Living Area and Selling Price

For the data in **Example 1**, you could describe each variable, x and y, individually using descriptive measures, such as the means ($\bar{x}$ and $\bar{y}$) or the standard deviations (s_x and s_y). However, theses measures do not describe the relationship between x and y for a particular residence, i.e. how the size of the living space affects the selling price of the house. A measure, the **correlation coefficient**, denoted by r, can solve this problem. It is defined as

$$r = \frac{s_{xy}}{s_x s_y}. \tag{6.26}$$

The quantities s_x and s_y are the standard deviations for the variables x and y, respectively. The quantity s_{xy} is called the **covariance** between x and y, and is defined as

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}. \tag{6.27}$$

The Table 6.3 provides how to interpret the correlation of two variables according to the value of r:

Table 6.3: Interpretation of r

$-1 \leq r \leq 1$	`Sign of` r `indicates direction of the linear relationship`
$r \approx 0$	`Weak relationship; random scatter of points`
$r \approx 1$ or $r \approx -1$	`Strong relationship; either positive or negative`
$r = 1$ or $r = -1$	`All points fall exactly on a straight line`

We usually compute the covariance using the following formula:

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1}, \tag{6.28}$$

where $\sum x_i y_i$ is the sum of the products $x_i y_i$ for each of the n pairs of measurements.

Example 2 Find the correlation coefficient for the number of square metres of living area and the selling price of a home for the data in **Example 1**.

Solution First, calculate $\sum x = 691.1$, $\sum y = 970.9$, $\sum xy = 134742.9$, $\sum x^2 = 95820.0$, and $\sum y^2 = 190840.73$.

Second, calculate s_{xy}, s_x^2, s_x, s_y^2 and s_y.

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1} = \frac{670988.99 - \frac{(691.1)(970.9)}{5}}{4} = 136.27,$$

$$s_x^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{95820.0 - \frac{(691.1)^2}{5}}{4} = 74.04,$$

therefore, $s_x = \sqrt{s_x^2} = 8.60$.

$$s_y^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1} = \frac{190840.73 - \frac{(970.9)^2}{5}}{4} = 577.84,$$

therefore, $s_y = \sqrt{s_y^2} = 24.04$.

Then, calculate correlation coefficient r.

$$r = \frac{s_{xy}}{s_x s_y} = \frac{136.27}{(8.60)(24.04)} = 0.659.$$

Two variables, x and y, may be related in a particular way. For example, the value of y depends on the value of x, i.e. the value of x in some way explains the value of y. In the above example, the cost of a home (y) may depend on its amount of area (x). In this situation, we call y the **dependent variable**, while x the **independent variable**. If the data exhibit a straight-line pattern, it is possible to describe the relationship relating y to x using a straight line given by the equation

$$y = a + bx,$$

where a is called the **y-intercept**, and b is called the **slope** of the line, which determines whether the line is increasing ($b > 0$), decreasing ($b < 0$), or horizontal ($b = 0$).

When plotting the (x, y) points for two variables x and y, the points generally do not fall exactly on a straight line, but they may show a trend that could be described as a linear pattern. We will describe this trend by fitting a line as best we can through the points. This best-fitting line related y to x, called the **regression line** or **least-squares line**, is found by minimizing the sum of the squared differences between the data points and the line itself. We compute b and a using the following formulas:

$$b = r\frac{s_y}{s_x}, \tag{6.29}$$

and

$$a = \bar{y} - b\bar{x}. \tag{6.30}$$

Since s_x and s_y are both positive, b and r have the same sign, so that

- When r is positive, so is b, and the line is increasing with x.
- When r is negative, so is b, and the line is decreasing with x.
- When r is close to 0, then b is close to 0.

The **coefficient of determination** can be interpreted as the percentage of the total variation that can be explained by using the regression equation. For simple linear regression, the coefficient of determination is the square of the coefficient of correlation, r, it is denoted as r^2. In regression, the coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. Normally, the coefficient of determination r^2 ranges from 0 to 1. If r^2 is close to 1, it indicates that the regression predictions perfectly fit the data. A coefficient of determination of 0 indicates that none of the variation in the dependent variable is due to the variation in the independent variable; while a coefficient of determination of 1 indicates that 100% of the variation in the dependent variable is due to the variation in the dependent variable. Therefore, r^2 measures the proportion of variation in y that is explained by the variation in x.

Example 3 Find the regression line for the number of square metres of living area and the selling price of a home for the data in **Example 1**. Find the coefficient of determination and explain it.

Solution First, calculate the value of r. In this question, we can use the result from Example 2, $r = 0.659$.

Second, find the slope, $b = r\frac{s_y}{s_x} = 0.659(\frac{24.04}{8.60}) = 1.84$, and the y-intercept, $a = \bar{y} - b\bar{x} = -60.23$.

Then, write the regression line by substituting the values for a and b into the equation:

$y = a + bx = -60.23 + 1.84x.$

The coefficient of determination: $r^2 = 0.434$. The variation in number of square metres of living area explains 43.4% of the variation in the selling price of a home.

Summary of calculation steps:

The steps to calculate the correlation coefficient:

1. First, create a table to find $\sum x$, $\sum y$, and $\sum xy$.

sum of x-squares and sum of y squares

2. Calculate the covariance, s_{xy}.
3. Use the computing formula to calculate s_x and s_y.
4. Calculate $r = \frac{s_{xy}}{s_x s_y}$.

The steps to calculate the regression line:

1. First, calculate $r = \frac{s_{xy}}{s_x s_y}$.
2. Find the slope, $b = r\frac{s_y}{s_x}$ and the y-intercept, $a = \bar{y} - b\bar{x}$.
3. Write the regression line by substituting the values for a and b into the equation: $y = a + bx$.

Exercises

1. When nicotine is absorbed in the body, cotinine is produced. The level of cotinine in the blood is proportionate to the amount of exposure to tobacco smoke. A measurement of cotinine in the body should be a good indicator of how much a

person smokes. The values below are randomly selected from a National Health Examination Survey. The number of reported cigarettes smoked per day are listed with the corresponding amounts of cotinine (measured in ng/mL) found in each person.

x(# of cigarettes/day)	20	10	10	50	4	15	1	20	8	10
y(cotinine)	344	226	209	408	65.5	185	9.5	350	2.5	270

a. Draw a scatterplot.

b. Calculate the linear correlation coefficient, r.

c. Find the regression line and add it to your scatterplot.

d. Estimate the amount of cotinine found in a person, given he/she smoked 6 cigarettes a day.

e. Find the coefficient of determination and explain it.

2. A study was made by a retailer to determine the relation between weekly advertising expenditures and sales. The following data were recorded:

a. Draw a scatterplot diagram for the above data.

b. Find the linear regression equation to predict weekly sales from advertising expenditures. Add the regression line to your scatterplot diagram.

c. Predict the weekly sales when advertising costs are $32.

d. Calculate the linear coefficient of correlation, r.

e. Find the coefficient of determination and explain it.

x, Advertising Costs ($)	y, Sales ($)
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

6.4 Analyzing Bivariate Data in Excel and Making Forecasts

Excel provides different graphical techniques for qualitative and quantitative bivariate data, as well as commands for obtaining bivariate descriptive measures when the data are quantitative.

Example 1 Comparative Line and Bar Charts Table 6.4 shows the status of 105 students in an introductory statistics class at University of Regina(UofR) and 110 students from University of Saskatchewan (UofS).

Table 6.4: Status of Students in a Statistics Class at UofR and UofS

	1st Year	2nd Year	3rd Year	4th Year	Grad Student
Frequency(UofR)	5	23	32	35	10
Frequency(UofS)	10	35	24	25	16

1. **Enter the data** into an *Excel* spreadsheet *just as it appears in the table,*

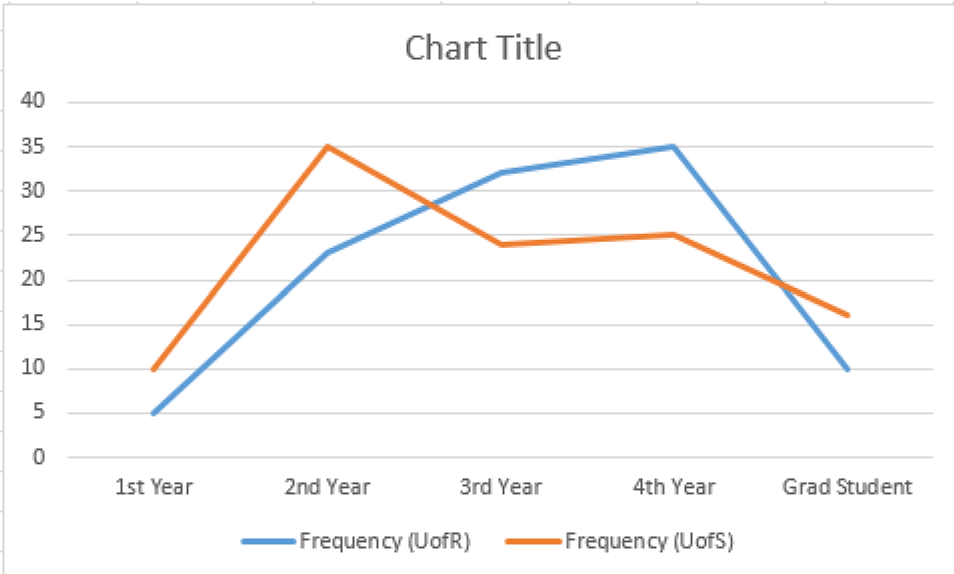


Figure 6.3: Original Line Charts

including the labels. Highlight the data in the spreadsheet, click the **Insert** tab and select **Line** in the **Charts** group. In the drop-down list, you will see a variety of styles to choose from. Select the first option to produce the line chart, Figure 6.3.

2. **Editing the line chart**: Again, you can experiment with the various options in the **Chart Layout** and **Chart Styles** groups to change the look of the chart. We have added a title on the vertical axis; we have added the title and have changed the "line style" of the UofR to a "dashed" style, by double-clicking on that line and choosing the **Dash type**. The line chart is shown in Figure 6.4.

3. Once the line chart has been created, right-click on the chart area and select

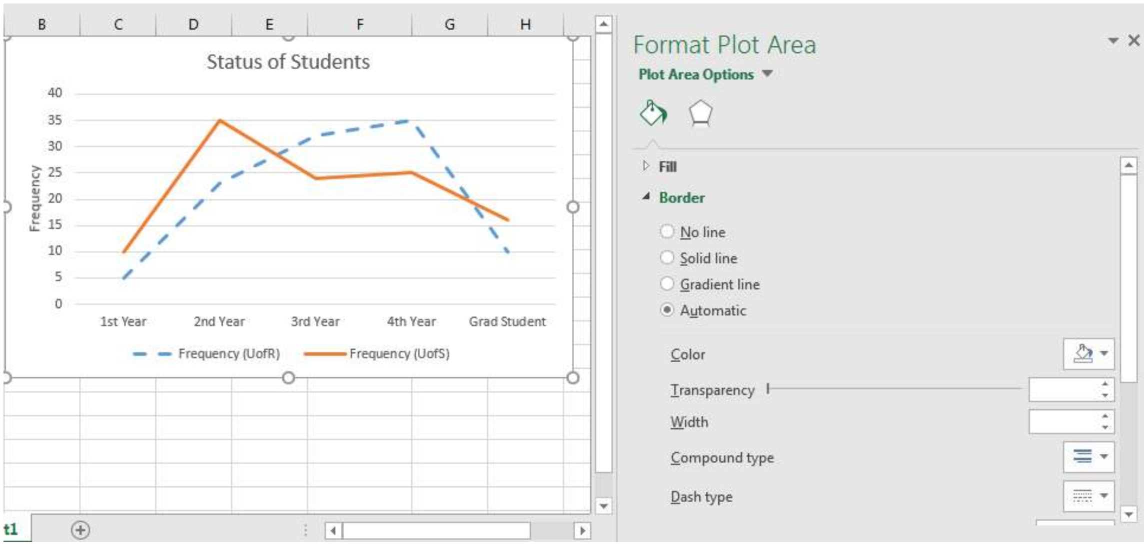


Figure 6.4: Edited Line Charts

Change Chart Type. Then choose either **Stacked Column** or **Clustered Column**. The comparative bar chart (a clustered bar chart), with the same editing that you chose for the line chart, will appear as shown in Figure 6.5.

Linear Regression analysis is a statistical technique for estimating the relationships among variables. In other words how do the sales figures change over time? If the goal is prediction, or forecasting, linear regression can be used to fit a predictive model to an observed data set of y and x values or known actual data (y) over time (x) (time series data). After developing such a model, if an additional value of x (a ne w period) is then given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.

In Excel the linear regression can be calculated using the **Forecast** function, the **Trend** function, the **Fill-Handle**, by calculating the equation: $y = a + bx$, and by

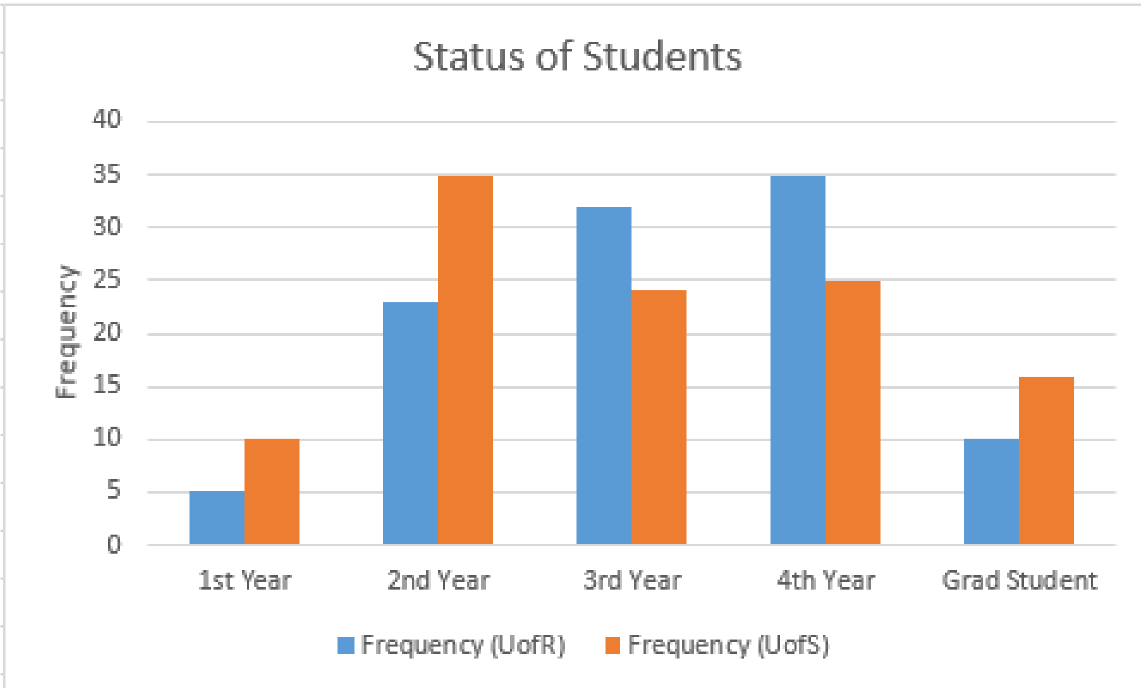


Figure 6.5: Clustered Column Charts

adding a **Trendline** to a chart.

The Data Analysis Tool **Regression** is an analysis tool to return important information working with linear regression such as the **Slope**, the **Y-Intercept**, **R-square**, and other statistical useful information.

R-square in the Excel is the coefficient of determination r^2, a statistic that will give some information about the goodness of fit of a model.

You can also calculate the Slope and the y Interceptor using the functions **Slope** and **Intercept**.

Example 2 (Regression and Forecasting) The data in Table 6.5 are the size of the living area (in m^2), x, and the selling price, y, of 12 residential properties.

Table 6.5: Living Area and Selling Price of 12 Properties

Properties	x (m^2)	y(\$thousands)
1	126.3	178.5
2	180.2	275.7
3	162.6	239.5
4	144.0	229.8
5	166.3	195.3
6	162.6	210.3
7	207.2	360.5
8	148.6	205.2
9	134.7	188.6
10	173.7	265.7
11	205.3	325.3
12	137.5	168.8

Get the scatterplot using *Excel* and find if it shows a linear pattern. If it is, draw a regression line for the data, calculate the mean, median, slope, intercept and estimate the price for 220 m^2 residential property.

1. Enter the data into *Excel* spreadsheet. Highlight Column A and Column B, click the **Insert** tab and select **Scatter** in the Charts group, and select the first option in the drop-down list. The scatterplot appears as in Figure 6.6, and will need to be edited.

2. **Editing the scatterplot:** With the scatterplot selected, look in the drop-down list in the **Chart Layouts** group. Find a layout that you want (we chose layout 1) and select it. Label the axes if you would like. Retitle the chart as "Scatterplot of Living Area vs. Selling Price", Figure 6.7.

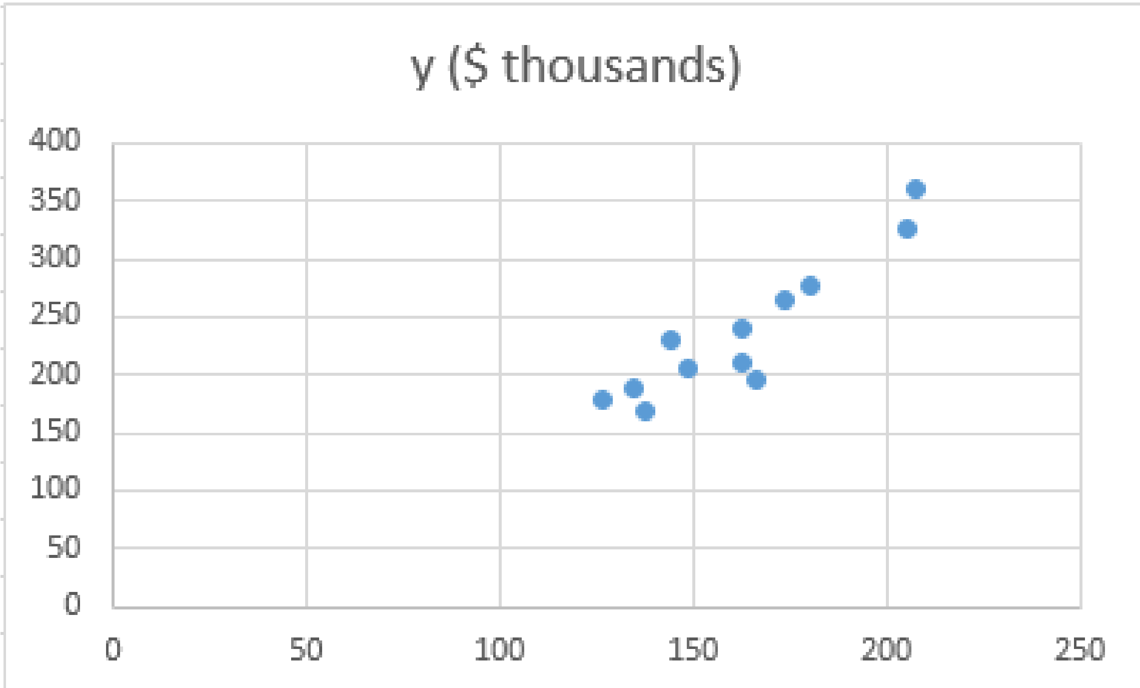


Figure 6.6: Original: Scatterplot of Living Area and Selling Price

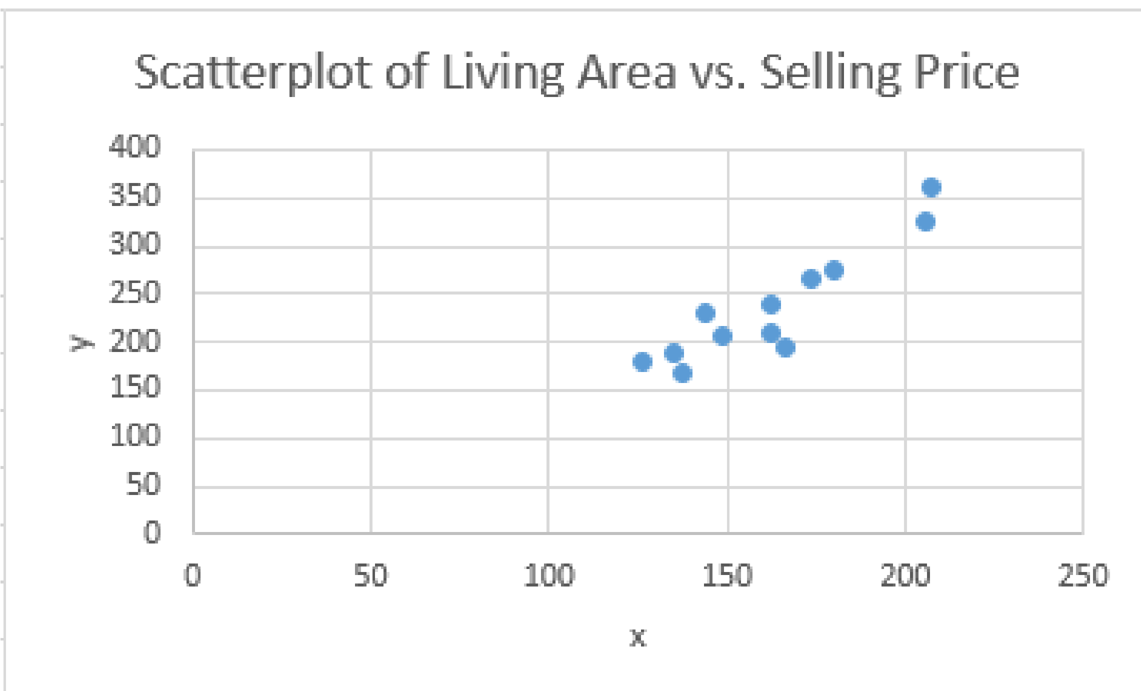


Figure 6.7: Adjusted: Scatterplot of Living Area and Selling Price

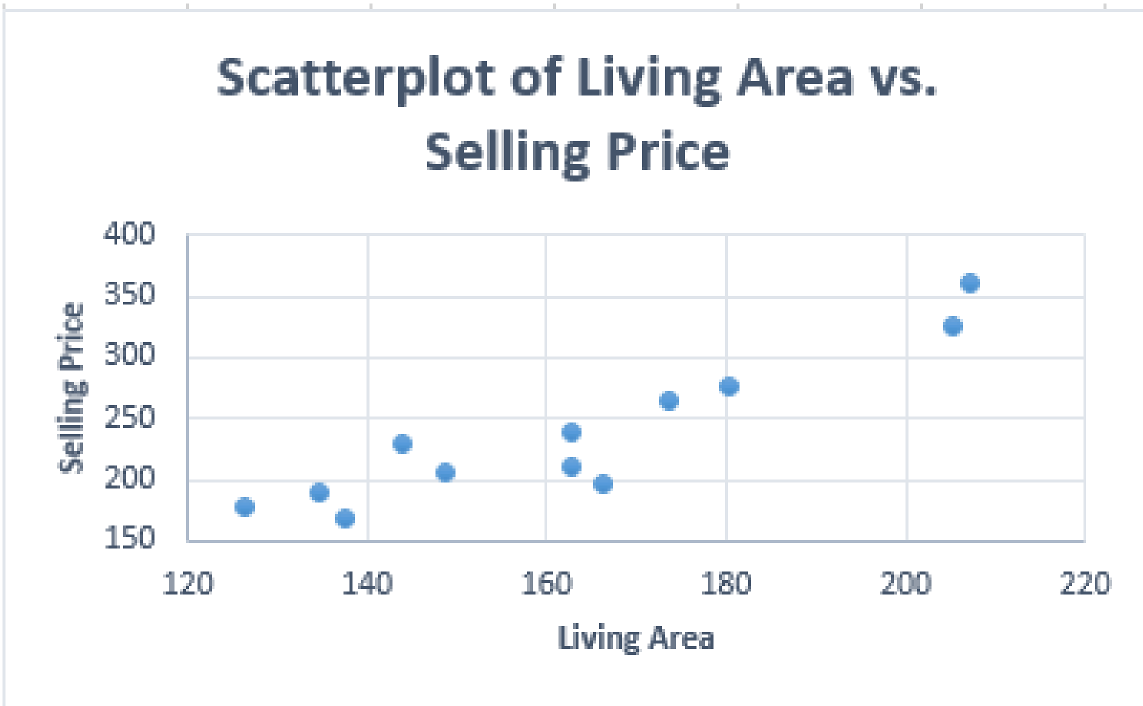


Figure 6.8: Finalized: Scatterplot of Living Area and Selling Price

The plot is still not optimal, since Excel chooses to use zero as the lower limit of the vertical scale, causing the points to cluster at the top of the plot. To adjust this, double-click on the vertical axis. In the **Format Axis** dialogue box, change the **Minimum** to **Fixed**, type (**150**) in the box, and click **Close**. You can make a similar adjustment to the horizontal axis if needed, we set it as 120 in this example. The scatterplot is shown in Figure 6.8.

3. To plot the regression line, simply right-click on one of the data points and select **Add Trendline**. In the dialogue box that opens, make sure that the radio button marked "Linear" is selected, and check the boxes marked "Display Equation on Chart" and "Display R-squared value on Chart"'. The final scatterplot is shown

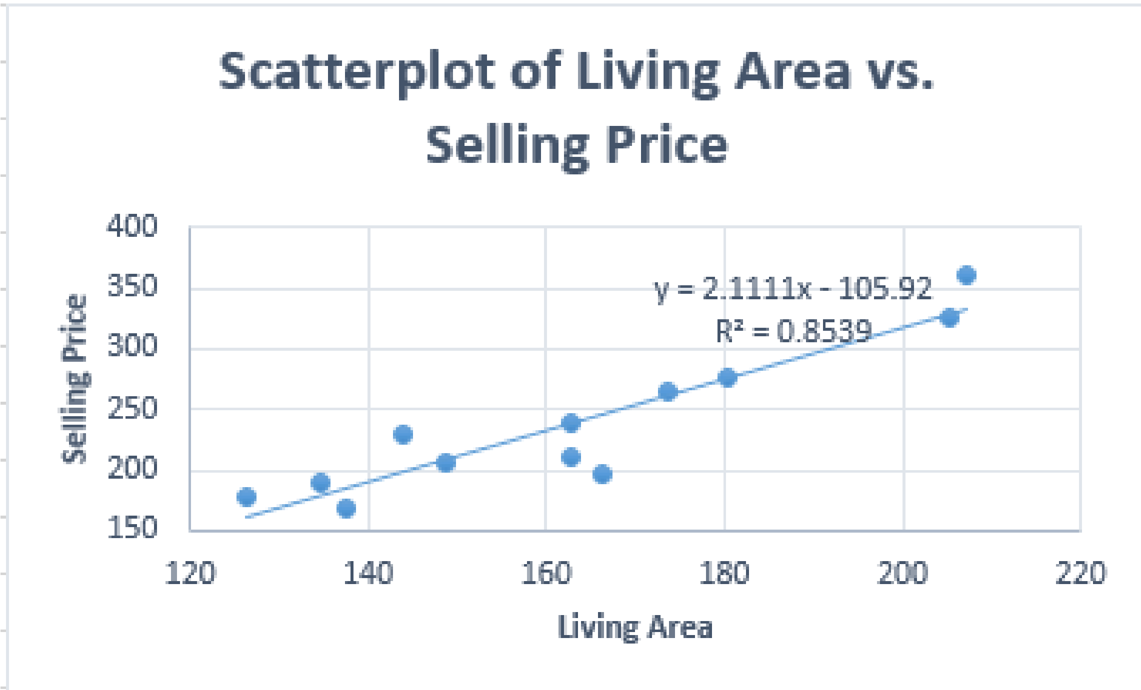


Figure 6.9: Scatterplot of Living Area and Selling Price with Regression Line

	x (m2)	*y ($ thousands)*
x (m2)	1	
y ($ thousands)	0.924086227	1

Figure 6.10: Correlation Matrix of Living Area and Selling Price

in Figure 6.9.

4. To find the sample correlation coefficient, r, you can use the command **Data** ▷ **Data Analysis** ▷ **Correlation**, selecting the two appropriate columns for the Input Range, clicking "Labels in First Row", and selecting an appropriate Output Range. When you click **OK**, the correlation matrix will appear in the spreadsheet, see Figure 6.10.

5. (**Alternate Procedure**) You can also place your cursor in the cell in which

F1 | =CORREL(B2:B13,C2:C13)

	A	B	C	D	E	F	G
1	Properties	x (m^2)	y ($ thousands)		Correlation, Area with Price	0.92	=CORREL(B2:B13,C2:C13)
2	1	126.3	178.5				
3	2	180.2	275.7		Standard deviation of price	59.76	=STDEV.S(C2:C13)
4	3	162.6	239.5		Standard deviation of area	26.16	=STDEV.S(B2:B13)
5	4	144	229.8				
6	5	166.3	195.6		Mean price	236.96	=AVERAGE(C2:C13)
7	6	162.6	210.3		Mean area	162.42	=AVERAGE(B2:B13)
8	7	207.2	360.5				
9	8	148.6	205.2		Median price	220.05	=MEDIAN(C2:C13)
10	9	134.7	188.6		Median area	162.60	=MEDIAN(B2:B13)
11	10	173.7	265.7				
12	11	205.3	325.3		Regression coefficient	2.11	=F1*F3/F4
13	12	137.5	168.8		Intercept	-105.92	=F6-F12*F7
14							
15					Forecast price given area of 220	358.52	=F12*220+F13
16							
17					Slope	2.11	=SLOPE(C2:C13,B2:B13)
18					Intercept	-105.92	=INTERCEPT(C2:C13,B2:B13)
19							

Figure 6.11: Statistical Information of Living Area and Selling Price

you want the correlation coefficient to appear. Select **Formulas** ▷ **More Functions** ▷ **Statistical** ▷ **CORREL** or click the "Insert Function" icon at the top of the spreadsheet, choosing **CORREL** from the **Statistical** category. Highlight or type the cell ranges for the two variables in the boxes marked "Array 1" and "Array 2" and click **OK**. For this example, the value is $r = 0.92$. Or you can use the built-in function "CORREL(Array1, Array2)". Similarly, you can get the standard deviation, mean and median of the price and living area, the regression coefficient, intercept and the forecast price, shown in Figure 6.11.

6. When $x = 220$, substituting it to the regression line

$$y = 2.1111x - 105.92 = 2.1111(220) - 105.92 = 358.52.$$

It means the estimate value of a 220 square meters residential property will be 358.52 thousands dollars.

You should keep in mind that you can only predict the value of near future or very close to the observed data.

Exercises

1. A group of products are categorized according to a certain attribute - A, B, C and according to the country in which they are produced:

	A	B	C
Canada	20	5	5
USA	10	10	5

a. Create a comparative side-by-side bar chart to compare the numbers of products of each type made in Canada and USA.

b. Create a stacked bar chart to compare the numbers of products of each type made in the two countries.

2. The owner of a one-bedroom apartment was undecided what to charge per month. The apartment was located 1.5 kilometers from a rapid transit station. The owner obtained the following information pertaining to one-bedroom apartments in the city. The variable x represents the distance from a rapid transit station, and y is the rent.

Distance in Kms (x)	Rent in $hundreds ($y$)
0.3	8.5
0.5	8.0
0.7	8.2
1.1	7.1
1.2	7.6
2.3	6.8
2.9	7.0
3.0	6.8

a. Draw a scatterplot. Does it appear to be a positive or negative linear relationship between x and y?

b. Find the regression line equation for estimating rent given the distance from the rapid transit station. Add the regression line to your scatterplot diagram.

c. Estimate the rent charged for an apartment 2.0 kilometers from the rapid transit station.

d. Find the linear correlation coefficient, r.

e. Find the coefficient of determination r^2 and explain it.

Data Source

[1] Statistics Canada. Table 18-10-0005-01 Consumer Price Index, annual average, not seasonally adjusted. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000501

6.5 Chapter 6 Review Questions

1. Assume the correlation coefficient was computed to be -.62. Which of the following values of r represents a stronger relationship than -.62?

a) +0.35 b) -.75 c) 10.4 d) 0

2. What is the possible range for the correlation coefficient?

3. In the equation $y = a + bx$, the letter b stands for the:

a) coefficient of correlation

b) slope of the regression line

c) y-intercept of the regression line

d) none of the above

4. Consider the data on breaking distance (in m) for a vehicle driven at various speeds (in km/h).

Speed	40	50	60	70	80	90
Distance	5	11	20	33	51	73

a. Compute the correlation coefficient

b. Find the equation of the regression line.

c. Predict the breaking distance of this vehicle traveling at 120 km/h.

d. Find the coefficient of determination and explain it.

Chapter 7

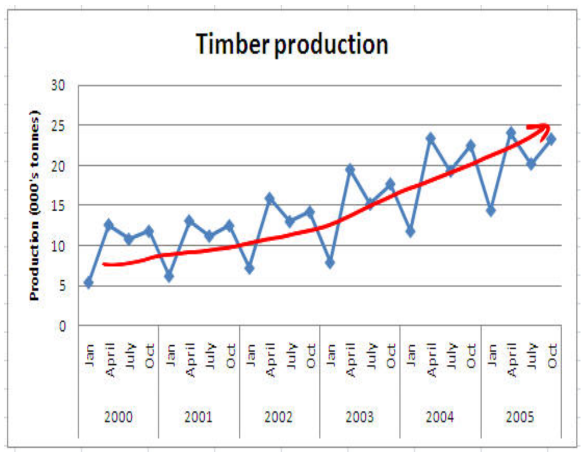
Time Series Forecasting

A time series is a sequence of observations taken sequentially in time. Many sets of data appear as time series: a monthly sequence of the sales of goods in a grocery store, a weekly series of the number of traffic accidents, a daily sequence data of server utilization, and so on. Time series data abound in economics, business, engineering, the natural sciences, and the social sciences. An intrinsic feature of a time series is that adjacent observations are dependent. The nature of this dependence allows us to forecast the future values of a time series from current and past values.

An observed time series can be decomposed into three components: the trend (long term direction), the seasonality (systematic, calendar related movements) and the irregularity (unsystematic, short term fluctuations).

7.1 Trend

The trend is defined as the "long term" movement in a time series without calendar related and irregular effects, and is a reflection of the underlying level. It is the result of influences such as population growth, price inflation and general economic changes. The following graph depicts a series in which there is an obvious upward trend over time:



Trend line is not for seasonal variations, is applicable to data with random variation and trend. It can make short term and medium term forecasting. For example, it has been used in linear regression, $y = a+bx$, where, y stands for forecast, x is the independent variable (time), a is the intercept, also called base level, b is the slope, which is the trend. There are different types of trendlines: exponential, linear, logarithmic, polynomial, power and moving average. We will demonstrate it in Excel.

Example Predict the number of airline takeoffs in India for 2012 using the following data:

	A	B
1	Airline takeoffs	from India
2	Year	Takeoffs
3	2003	263870
4	2004	302790
5	2005	330484
6	2006	453921
7	2007	569033
8	2008	592292
9	2009	601977
10	2010	634062
11	2011	707754

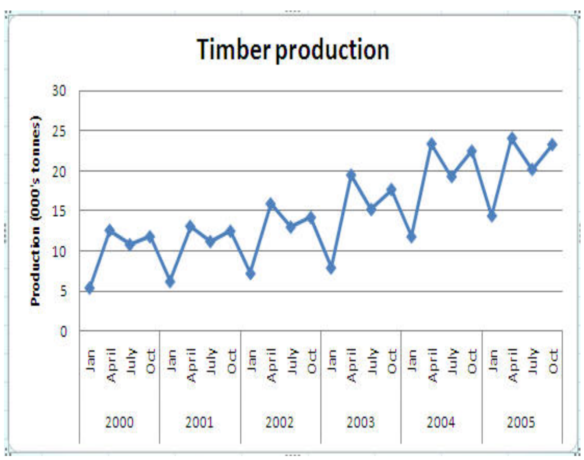
Exercise Predict the number of broadband internet subscriber per 100 people for 2012 using the following data:

	A	B	C
1	Broadband Internet Subscribers per 100 People		
2	Year	Code	Broadband Penetration
3	2002	1	0.01
4	2003	2	0.01
5	2004	3	0.02
6	2005	4	0.12
7	2006	5	0.2
8	2007	6	0.27
9	2008	7	0.44
10	2009	8	0.64
11	2010	9	0.9
12	2011	10	1.08

7.2 Seasonality

A seasonal effect is a systematic and calendar related effect. In many business scenarios, we would like to take a seasonality pattern into account in the forecast. Some examples include the sharp escalation in most retail series which occurs around December in response to the Christmas period, or an increase in ice cream consumption in summer due to warmer weather.

Seasonality in a time series can be identified by regularly spaced peaks and troughs which have a consistent direction and approximately the same magnitude every year, relative to the trend. The following diagram depicts a strongly seasonal series.



In this example, the magnitude of the seasonal component increases over time, as does the trend.

In some cases we know what the seasonality length is, but in other cases we do not. This forecasting method looks for seasonality patterns in the historical data and tries to determine the pattern that best matches the data. In order to work properly, the more repeating cycles the historical data contains the better. It is recommended to have at least 2–3 full seasonal cycles in the historical data. It works for data with seasonal variation and it can be used for short term forecasting. It is usually tested by inserting a line chart to the data in Excel. We will show how to calculate the seasonal indices in Excel with data examples. First, you need to insert a line chart to see if the seasonality exists. If it has seasonality, then you calculate the indices of each period, for example, each quarter.

Example Calculate the seasonal indices using the following data and make forecast for each quarter for year 4, if the total sales forecast for year 4 is 561 ($1000).

	A	B	C
1	Year	Quarter	Sales ($1000)
2	Year 1	1	145
3		2	185
4		3	132
5		4	94
6	Year 2	1	140
7		2	190
8		3	135
9		4	90
10	Year 3	1	145
11		2	188
12		3	130
13		4	95

Steps:

1. Calculate the total and quarter average of each year.

2. Each quarter's sale divided by the average of that year accordingly.

3. Average each quarter's data obtained from Step 2 to get the seasonal indices for each quarter.

4. Make each quarter's forecast by multiplying the quarter average of that year with each quarter's seasonal index.

5. Add the forecast data line to the graph.

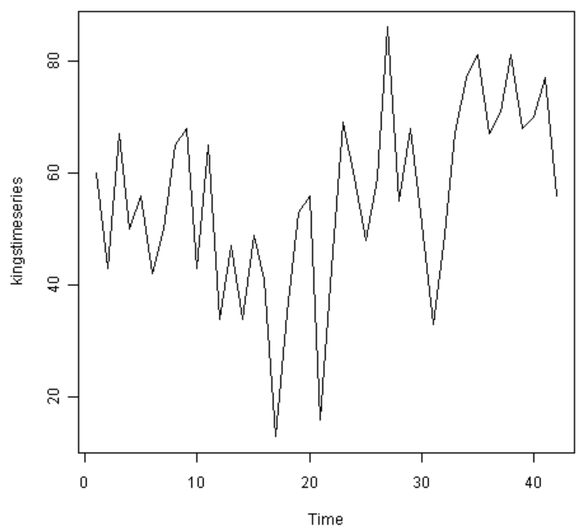
Exercise Calculate the seasonal indices using the following data and make forecast for each quarter for year 4, if the total sales forecast for year 4 is 612 ($1000).

	A	B	C
1	Year	Quarter	Sales ($1000)
2	Year 1	1	95
3		2	164
4		3	148
5		4	175
6	Year 2	1	102
7		2	159
8		3	152
9		4	181
10	Year 3	1	101
11		2	172
12		3	148
13		4	190

7.3 Irregularity

The irregular component (it's also known as the residual) is what remains after the seasonal and trend components of a time series have been estimated and removed.

It results from short term fluctuations in the series which are neither systematic nor predictable. In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality. The following graph is of a highly irregular time series:



7.4 Decomposition of the Time Series

Decomposition models are typically additive or multiplicative, but can also take other forms such as pseudo-additive.

Additive Decomposition

In some time series, the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls. In such cases, an additive model is appropriate.

In the additive model, the observed time series Y_t is considered to be the sum of three independent components: the trend T_t, the seasonality S_t, and the irregularity I_t, i.e., $Y_t = T_t + S_t + I_t$. We will not practice additive decomposition method. In stead, we focus on the multiplicative decomposition.

Multiplicative Decomposition

In many time series, the amplitude of both the seasonal and irregular variations increase as the level of the trend rises. In this situation, a multiplicative model is usually appropriate.

In the multiplicative model, the original time series is expressed as the product of trend, seasonal and irregular components, i.e., $Y_t = T_t \times S_t \times I_t$.

7.5 Moving Average

The idea of a moving average is straightforward. We generally apply a moving average to a series of observations taken over time. The average moves forward with the data series. For example, moving average with 4 days data, the first moving average includes Days 1 through 4, the second moving average includes Days 2 through 5, and so on. Any data series has an average value (or mean) and individual values that vary around the mean. A data series that consists of the sales from last week has an average sales amount, measured in currency. It also has individual sales amounts, measured in currency, that vary around the mean. Some individual sales amounts are higher than the mean and some lower. In some data series, the individual amounts steadily increase or decrease over time, pulling the mean up or down accordingly. In other series, the individual amounts vary randomly around a static mean value. One of the questions that we would like to answer with forecasting is whether the mean value of the data series is increasing, decreasing, or merely remaining at the same level over time. We will use data examples to show two ways to make forecast using Moving Average in Excel.

Example Make forecast using 3 month and 4 month moving average based on the following data:

	A	B
1	Time	Actual
2	1	134
3	2	143
4	3	144
5	4	130
6	5	135
7	6	125
8	7	140
9	8	137
10	9	143
11	10	126
12	11	132
13	12	139
14	13	136
15	14	132
16	15	124
17	16	137
18	17	128
19	18	134
20	19	145
21	20	146

We will use two methods and get the following forecasts:

	A	B	C	D
1	Time	Actual	3 month moving	4 month moving
2	1	134		
3	2	143		
4	3	144		
5	4	130	140.3	
6	5	135	139.0	137.8
7	6	125	136.3	138.0
8	7	140	130.0	133.5
9	8	137	133.3	132.5
10	9	143	134.0	134.3
11	10	126	140.0	136.3
12	11	132	135.3	136.5
13	12	139	133.7	134.5
14	13	136	132.3	135.0
15	14	132	135.7	133.3
16	15	124	135.7	134.8
17	16	137	130.7	132.8
18	17	128	131.0	132.3
19	18	134	129.7	130.3
20	19	145	133.0	130.8
21	20	146	135.7	136.0

	A	B	C	D	E
1	Time	Actual	3 month moving	4 month moving	center
2	1	134			
3	2	143	140.3	137.8	
4	3	144	139.0	138.0	137.9
5	4	130	136.3	133.5	135.8
6	5	135	130.0	132.5	133.0
7	6	125	133.3	134.3	133.4
8	7	140	134.0	136.3	135.3
9	8	137	140.0	136.5	136.4
10	9	143	135.3	134.5	135.5
11	10	126	133.7	135.0	134.8
12	11	132	132.3	133.3	134.1
13	12	139	135.7	134.8	134.0
14	13	136	135.7	132.8	133.8
15	14	132	130.7	132.3	132.5
16	15	124	131.0	130.3	131.3
17	16	137	129.7	130.8	130.5
18	17	128	133.0	136.0	133.4
19	18	134	135.7	138.3	137.1
20	19	145	141.7		
21	20	146			

7.6 Weighted Moving Average

One drawback of forecasting with moving averages is that all the data used in the look-back period is weighted equally. As the name suggests, with the weighted moving average technique, historical data is weighted, with some observations being more important than others. The difficulty with weighted moving average is that both a look-back period and weights have to be selected. One way is to put more weight on recent data and less on past data. Another way is to apply the Excel function “Solver”. We need to select an error measure to minimize and then with the help of Solver using the standard GRG non-linear method we can let Excel try many values for weight until it finds the best combination.

Example Make forecast using weighted moving average for the moving average data of last section. The weights are: A_{m-1}: 0.5, A_{m-2}: 0.3, A_{m-3}: 0.2, where m is the period to forecast, A_i are actual data, for period i.

7.7 Exponential Smoothing

Exponential smoothing is a rule of thumb technique for smoothing time series data using the exponential window function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Exponential smoothing is often used for analysis of time series data.

The simplest form of exponential smoothing is given by the formula:

$$F_n = F_{n-1} + \alpha(A_{n-1} - F_{n-1})$$

Or equivalently

$$F_n = (1 - \alpha)F_{n-1} + \alpha A_{n-1}$$

where α is the smoothing factor, and $0 \leq \alpha \leq 1$, n is the period to forecast, F_i are forecast data, A_i are actual data, for period i. The first forecast data could be the first actual data or the average of all actual data.

Example Apply the exponential smoothing to the moving average data.

7.8 Application of Multiplicative Decomposition

Since most of data affected by seasonality, trend and irregularity, it is impossible to use one single method to make reasonable forecast. The common way is to identify these factors and use the combination of trend line, seasonality indices and moving average. The main steps to use the combination and multiplicative decomposition in forecast:

Step 1: In a worksheet, enter time series and sales series (Y_t) that correspond to each other.

Step 2: Insert a line chart with markers to see if there are seasonality and trend.

Step 3: Insert a column $t(1, 2, 3, \cdots)$.

Step 4: Calculate the moving average (MA), for example, we will choose moving average 4 data (MA(4)) if it shows a quarterly seasonality.

Step 5: Calculate the centered moving average (CMA) if necessary.

Step 6: Add the trend line.

Step 7: Isolate the product of seasonality (S_t) and irregularity (I_t): $S_t I_t = \frac{Y_t}{CMA}$.

Step 8: Calculate the seasonal component S_t. For quarterly data, we calculate it by average $S_t I_t$ for each quarter to get the seasonality components.

Step 9: Deseasonalize the data by calculating $\frac{Y_t}{S_t}$.

Step 10: Find the trend component T_t (**Data**▷**Data Analysis**▷**Regression**▷**OK**▷ **Input Y Range**▷**Input X range**▷`Select` **Labels**▷`Select` **Output Range**▷

OK). We will need the coefficients data to calculate: $T_t = \texttt{intercept} + \texttt{slope} \times t$.

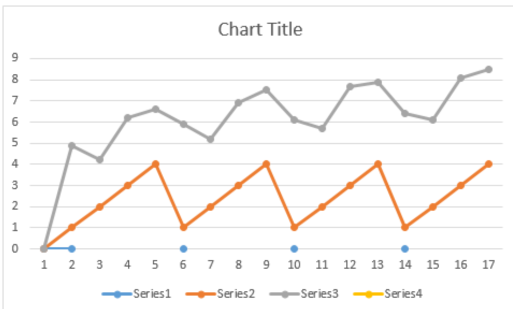
Step 11: Make forecast (forecast=the product of seasonal and trend component), and add the predicted data to the graph.

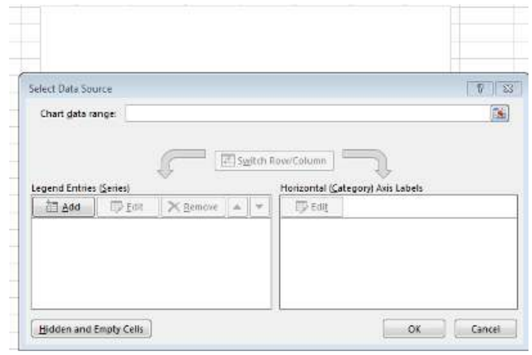
Example 1 Four-year quarterly car sales data (QuarterlySales.xlsx) are posted at Brightspace. Predict the quarterly sales of Year 5 using moving average and multiplicative decomposition method.

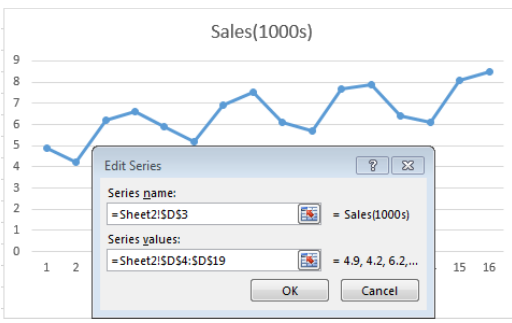
The 9 graphs are steps we may experience, and the first one is data entry:

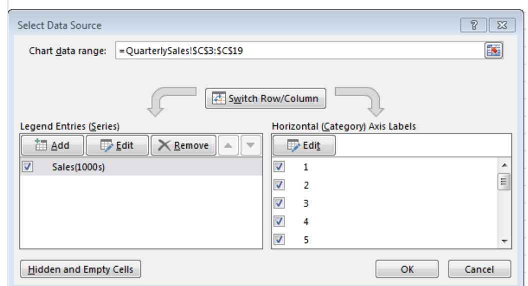
	A	B	C	D
1	Quarterly Data for Cars Sales			
2				
3	Year	Quarter	Sales(1000s)	
4	Year 1	1	4.9	
5		2	4.2	
6		3	6.2	
7		4	6.6	
8	Year 2	1	5.9	
9		2	5.2	
10		3	6.9	
11		4	7.5	
12	Year 3	1	6.1	
13		2	5.7	
14		3	7.7	
15		4	7.9	
16	Year 4	1	6.4	
17		2	6.1	
18		3	8.1	
19		4	8.5	

We need to edit the line chart:

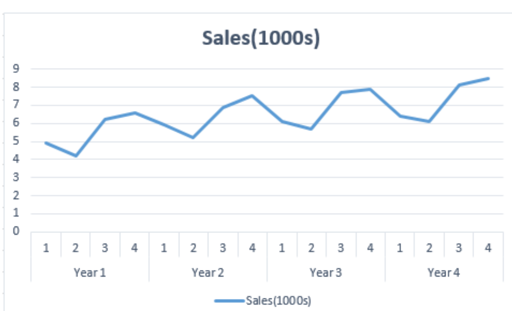


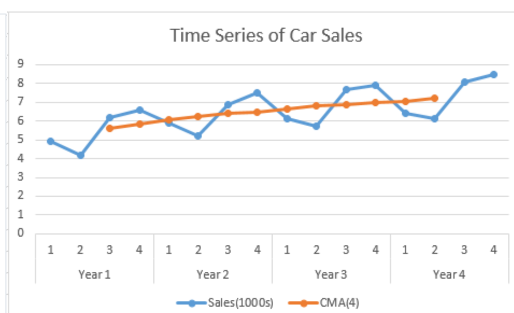




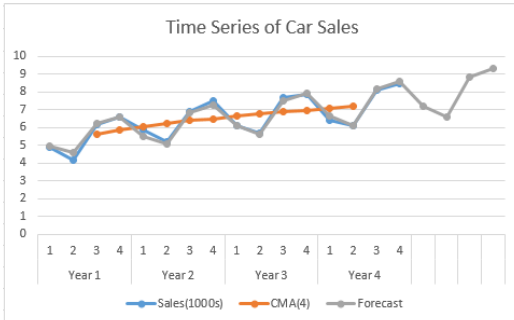


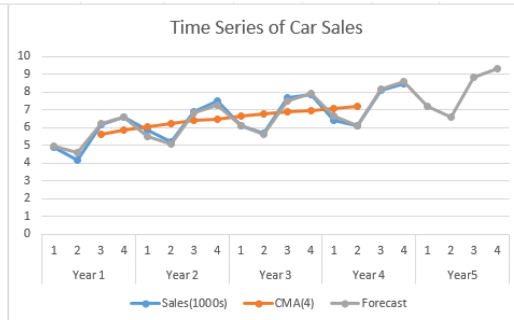
Add the trend line:





Make forecast (forecast=the product of seasonal and trend component), add the predicted data to the table and graph, and edit the horizontal axis by adding four quarters of Year 5 .





Exercises

1. Three-year quarterly TV sales data (QuarterTVSale.xlsx) are posted at Brightspace. Make a forecast of Year 4 by applying moving average and multiplicative decomposition method.

7.9 Apply Built-In Function in Excel to Make Forecasts

Often we use Excel to analyze time series data to find recurring seasonality patterns and trends. In Excel 2016, new forecasting sheet functions and one-click forecasting helps you to explain the data and understand future trends.

Exponential Smoothing

We can apply a built-in function, Exponential Smoothing or ETS, in Excel 2016. Exponential Smoothing methods are a popular way to forecast and are among the leading methods that have become industry standards.

The main advantages of using the ETS method are the ability to detect seasonality patterns and confidence intervals.

Confidence Intervals

Apart from predicting future values for the input time series, the ETS forecast can also return a confidence interval.

The confidence interval is the range surrounding each predicted value in which 95 percent of future points are expected to fall based on the forecast (with normal distribution). The confidence interval helps you figure out the accuracy of the prediction. A smaller interval implies more confidence in the prediction for the specific point. The default level of 95 percent confidence can be changed using the up or down arrows and can be used in two ways:

You can deduct from the width of the confidence interval to understand the accuracy of the prediction. You can experiment with several of the advanced options (such as how to account for missing points, seasonality, etc.) and observe if the previewed confidence interval got thinner or wider. This provides an indication of how well the underlying model fits the historical data.

Example 1 30 days car sales data (CarSales.xlsx) are posted at Brightspace. Make a forecast of another seven days sales using Create Forecast Worksheet, using 95% confidence interval, detecting seasonality automatically, filling missing points using interpolation. Is the trend increasing or decreasing?

Steps to create a forecast:

Step 1: In a worksheet, enter two data series that correspond to each other:

- A series with date or time entries for the timeline
- A series with corresponding values

These values will be predicted for future dates.

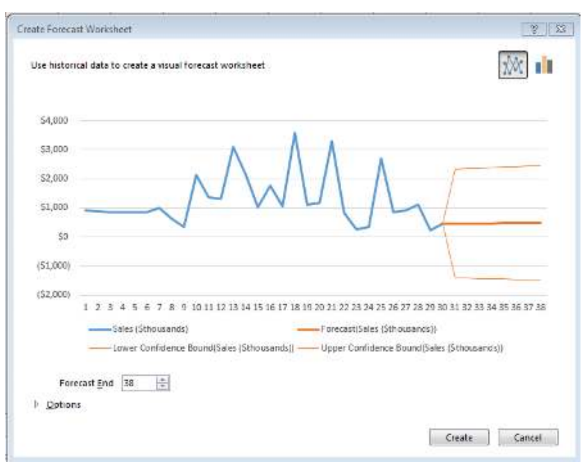
	A	B
1	Day	Sales ($thousands)
2	1	$901
3	2	$890
4	3	$855
5	4	$865
6	5	$845
7	6	$848
8	7	$1,001
9	8	$637

Step 2: Select both data series.

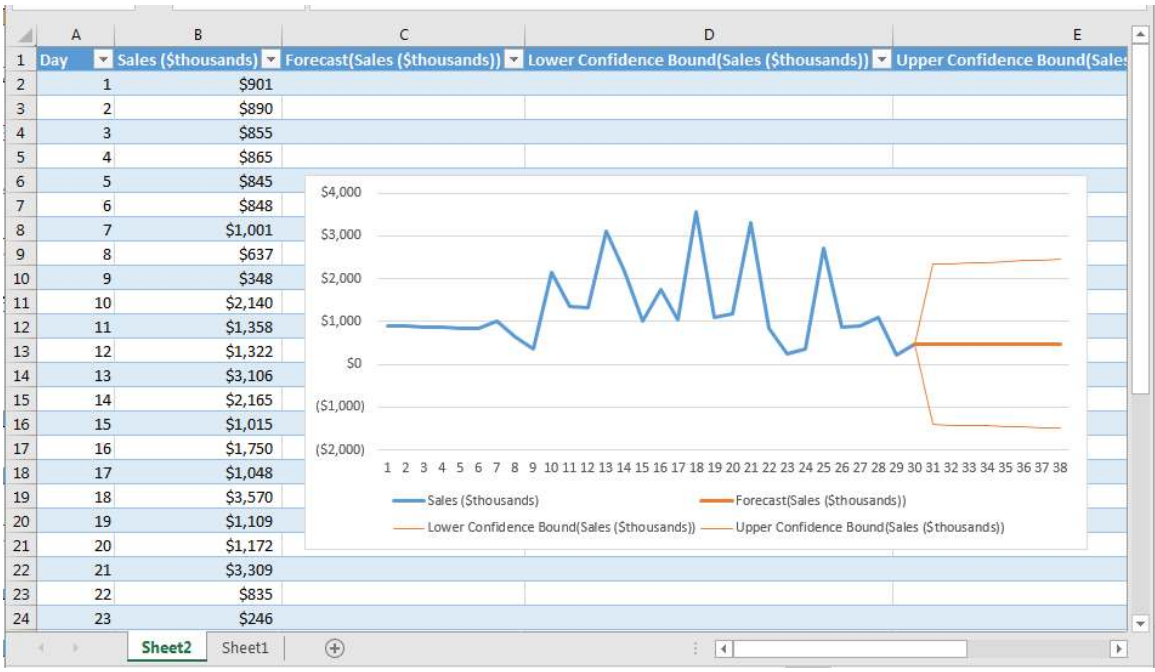
Step 3: On the **Data** tab, in the **Forecast** group, click **Forecast Sheet**.

Step 4: In the **Create Forecast Worksheet** box, pick either a line chart or a column chart for the visual representation of the forecast.

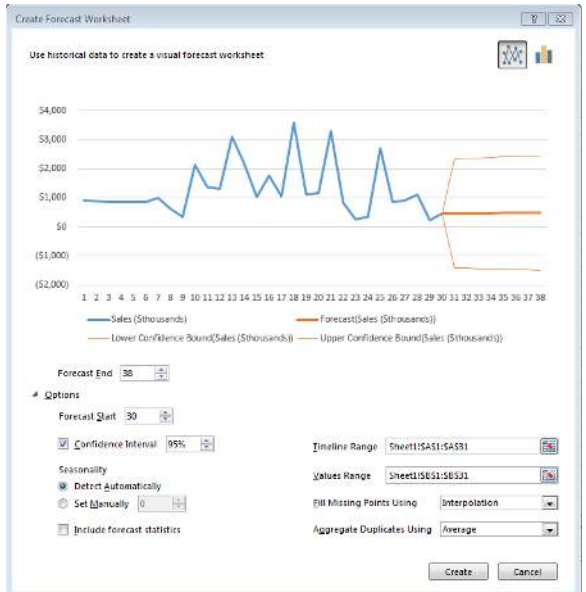
Step 5: In the **Forecast End** box, pick an end date, and then click **Create**.



Excel creates a new worksheet that contains both a table of the historical and predicted values and a chart that expresses this data. You'll find the new worksheet just to the left of ("in front of") the sheet where you entered the data series.



If you want to change any advanced settings for your forecast, click **Options**. You'll find information about each of the options in the following table.



The automatically detected value in the **Seasonality** section can be found in the **Create Forecast** dialog under **Options**. In case the seasonal data was not significant enough to be detected, or you know which seasonality you expect, you can manually override the automatically detected value by selecting **Set Manually**.

Exercises

1. A grocery store's daily retail sales data (DailyRetailSales.xlsx) are posted at Brightspace. Please predict the first two weeks daily sales of March using the **Forecast Sheet**, using 95% confidence interval, detecting seasonality automatically, filling missing points using interpolation. Is the trend increasing or decreasing?

7.10 Chapter 7 Review Questions

1. What are the three components of an observed time series?

2. A electronic appliance store plans to make a forecast for next year sales. Three-year quarterly computer sales data (QuarterlyComputerSales.xlsx) are posted at Brightspace. Make a forecast of the quarterly sales for next year by applying moving average and multiplicative decomposition method.

3. A retail store's daily sales data (DailyRetailSales.xlsx) are posted at Brightspace. Please predict the next 10 days daily sales using the **Forecast Sheet**, using 95% confidence interval, detecting seasonality automatically, filling missing points using interpolation. Is the trend increasing or decreasing?

Appendix A

Formula Book

Chapter 2

Simple interest:

$$I = Prt$$

Amount:

$$S = P + I = P + Prt = P(1 + rt)$$

Present Value:

$$P = \frac{S}{1 + rt} = S(1 + rt)^{-1}$$

Compound Amount:

$$S = P(1 + i)^n,$$

Present Value:

$$P = \frac{S}{(1 + i)^n} = S(1 + i)^{-n}$$

Chapter 3

Unit Variable Cost:

$$UVC = \frac{TVC}{n}$$

Net Income using a total revenue and total cost approach:

$$NI = n(S) - (TFC + n(UVC))$$

Unit Contribution Margin:

$$UCM = S - UVC$$

Net Income using a total contribution margin approach:

$$NI = n(UCM) - TFC$$

Contribution Rate:

$$CR = \frac{UCM}{S} \times 100\%$$

or

$$CR = \frac{TR - TVC}{TR} \times 100\%$$

Break-even Analysis in units:

$$n = \frac{TFC}{S - UVC} = \frac{TFC}{UCM}$$

Break-even Analysis in dollars:

$$TR = \frac{TFC}{CR}$$

Return On Investment:

$$\texttt{ROI} = \frac{\texttt{Current Value of Investment} - \texttt{Cost of Investment}}{\texttt{Cost of Investment}}$$

Annualized ROI:

$$\texttt{Annualized ROI} = (1 + \texttt{ROI})^{\frac{1}{\texttt{number of years}}} - 1$$

or

$$\texttt{Annualized ROI} = \left(\frac{\texttt{Ending Value}}{\texttt{Beginning Value}}\right)^{\frac{1}{\texttt{number of years}}} - 1$$

Chapter 4

$$\texttt{Relative frequency} = \frac{\texttt{Frequency}}{n}$$

$$\texttt{Percentage} = \texttt{Relative frequency} \times 100\%$$

Chapter 5

Sample Mean:

$$\bar{x} = \frac{\sum x_i}{n}$$

Population mean:

$$\mu = \frac{\sum x_i}{N}$$

Position of the Median:

$$\texttt{Position of the Median} = 0.5(n + 1)$$

Variance of a Population:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Calculating formula for Variance of a Population:

$$\sigma^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2,$$

Variance of a Sample:

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

Calculating formula for Variance of a Sample:

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$$

Chapter 6

Correlation Coefficient:

$$r = \frac{s_{xy}}{s_x s_y}$$

where s_{xy} is called the **covariance** between x and y:

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Computing formula for covariance:

$$s_{xy} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1}$$

The slope of the linear regression line $y = a + bx$:

$$b = r\frac{s_y}{s_x}$$

The y-intercept of the linear regression line $y = a + bx$:

$$a = \bar{y} - b\bar{x}$$

Chapter 7

Multiplicative Decomposition of Time Series:

$$Y_t = T_t \times S_t \times I_t$$

Exponential Smoothing:

$$F_n = F_{n-1} + \alpha(A_{n-1} - F_{n-1})$$

Or

$$F_n = (1 - \alpha)F_{n-1} + \alpha A_{n-1}$$