

STATISTICS WITH APPLIED PROBABILITY

Custom eBook for STA258

Nishan Mudalige

Nurlana Alili

Bryan Su

CANADA

Statistics with Applied Probability

Custom eBook for STA258

Nishan Mudalige

Department of Mathematical and Computational Sciences
University of Toronto Mississauga

Nurlana Alili

University of Toronto Mississauga

Bryan Xu

University of Toronto Mississauga

© 2025 N. Mudalige, N. Alili, B. Xu
All rights reserved.

This work may not be copied, translated, reproduced or transmitted in any form or by any means — graphic, electronic or mechanical including but not limited to photocopying, scanning, recording, microfilming, electronic file sharing, web distribution or information storage systems — without the explicit written permission of the authors.

Every effort has been made to trace ownership of all copyright material and to secure permission from copyright holders. In the event of any question arising as to the use of copyright material, we will be pleased to make necessary corrections in future publications.

First edition: August 2025

Mudalige, M.; Alili, N.; Xu, B.

University of Toronto Mississauga,
Mississauga, Ontario, Canada

Contents

0	Overview	1
1	Descriptive Statistics and an Introduction to R	2
1.1	Introduction	2
1.2	Descriptive Statistics	4
1.3	Graphical Techniques	10
1.3.1	Histograms	10
1.3.2	Box-Plots	12
1.4	Introduction to R	13
2	Sampling Distributions Related to a Normal Population	3
2.1	Normal Distribution	3
2.2	Gamma and Chi-square Distribution	4
2.3	Student's t-Distribution and F-Distribution	7
3	The Central Limit Theorem	2
8	One Sample Confidence Intervals On a Proportion	2

Chapter 0

Overview

Uncertainty is an inherent part of everyday life. We all face questions regarding uncertainty such as whether classes will go ahead as planned on any given day; will a flight leave on time; will a student pass a certain course? Uncertainties might also change depending on other factors, such as whether classes will still go ahead as planned when there is a snow warning in effect; if a flight is delayed can a person still manage to make their connection; will a student pass their course considering that the instructor is known to be a tough grader?

The ability to quantify uncertainty using rigorous mathematics is a powerful and useful tool. Calculating uncertainty on an intuitive level is something that is hard-wired in our DNA, such as the decision to fight or flight depending on a given set of circumstances. However we cannot always make such intuitive decisions based purely on hunches and gut feelings. Fortunes have been lost based on someone having a good feeling about something. If we have information available, we should make the best prediction possible using this information. For instance if we wanted to invest a lot of money in a company, we should use all available data such as past sales, market and industry trends, leadership ability of the CEO, forward looking statements etc. and with all this information we can then predict whether our investment will be profitable.

In order for companies to survive and remain competitive in today's environment it is essential to monitor industry trends and read markets properly. Companies that don't adapt and stick to an outdated business model tend to pay the price. At the other end of the spectrum, companies that understand the needs of the consumer, build their product around the consumer and keep evolving their product offerings based on consumer trends tend to perform well and remain competitive.

Statistics is the science of uncertainty and it is clearly a very useful subject for business. In this book you will be given an introduction to statistics and you will learn the framework as well as the language required at the introductory level. The material may be daunting at times, but the more you get familiar with the subject the more comfortable you will become with it. As business students, doing well in a statistics course will give you a competitive edge since the ability to interpret and perform quantitative analytics are skills that are highly desired by many employers.

Chapter 1

Descriptive Statistics and an Introduction to R

1.1 Introduction

Intuitively, statistics can be considered the science of uncertainty. Formally,

Definition 1.1 (Statistics).

Statistics is the science of collecting, classifying, summarizing, analyzing and interpreting data.

Population, Sample, Parameter

In statistics, researchers need to observe behavior, pattern, trends and other types of data to give a conclusion. To make the conclusion more persuasive, researchers require huge amount of data to support them, that's why study statistics need population.

Definition 1.2 (Population).

In statistics, a population is a set of similar observations which is of interest for some experimental questions. It can be a set of existing objects such as all people in Canada, or hypothetical group of existing objects such as the set of all possible hands in a game of poker.

However, data collection from population is a lot work. Usually, researchers select a finite number of observations to study.

Definition 1.3 (Sample).

It refers to a selection of a subset from population that researchers use it to estimate population characteristics.

Now, we have already chosen a sample, but how do we use it to estimate population characteristics? This is the point where parameter comes to play.

Definition 1.4 (Parameter Statistics).

A parameter is a quantity of statistical population which summerizes characteristics of the population. For example, mean, variance and standard deviation.

Descriptive and Inferential Statistics

Now, we have set everything we need. A population, a chosen sample in that population with its parameters. Next step is studying. There are two major types of analysis: descriptive and Inferential statistics. In this section, we are only going to give you a rough idea about what they are, more detailed materials will be introduced in later chapters.

Definition 1.5 (Descriptive Statistics).

It refers to the summation of all quantitive values that describe characteristics of the population. Usually, we use descriptive statistics to summerize characteristics of a data set.

Furthermore, we use inferential statistics to do statistical analysis.

Definition 1.6 (Inferential Statistics).

It refers to the process of using data analysis to indicate properties of a population. For example, testing hypothesis and confidence interval (both will be introduced in later chapters).

Qualitative and Quantitative Data

At this point, assume that we have finished all procedures such as obtaining parameters and analyzing properties. Now, another important thing is illustrating all the discovery.

Definition 1.7 (Qualitative Data).

This type of illustration refers to showing categorical data. For example, lecture notes from a course, open-question survey.

To illustrate numerical data, we use quantitative data.

Definition 1.8 (Quantitative Data).

Unless the previous type of illustration, quantitative data is represented numerically, including anything that can be counted, measured, or given a numerical value. For example, STA258 final mark score range from 100 different students who have taken this course.

1.2 Descriptive Statistics

Previously, we defined descriptive statistics. Now, let's introduce what exact they are.

Sample Mean, Variance and Standard Deviation

Sample mean (or sample average) is the average value of a sample which is selected from an interested population of an experiment. Usually, the sample mean is used to estimate population mean. In other words, we say that the sample mean is an estimator of population mean.

Definition 1.9 (Sample Mean).

Let $x_1, x_2, x_3, \dots, x_n$ be a sample of data points. We define sample mean of the sample data points (\bar{x}) as the following:

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

Also, we define sample variance of the sample data points (s^2) as:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Moreover, the standard deviation of the sample of data points (s) is:

$$s = \sqrt{s^2}, \quad \text{for } s > 0.$$

Now, let's move to variance. It refers to the expected value of the squared deviation from the mean of a random variable in a population. Similarly, we do have sample variance as well, which is the expected value of the squared deviation from the mean of a random variable in a selected sample. At this point, we can still use sample variance to estimate population variance with adjustment, because the sample variance may differ significantly based on what data points are chosen from that population.

Definition 1.10 (Sample Variance).

Let $x_1, x_2, x_3, \dots, x_n$ be a sample of data points, we define sample variance of the sample data points (s^2) as:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{where } \bar{x} \text{ is the sample mean of the data points.}$$

Next is standard deviation. It is a measure of the amount of variation of the values of a variable about its mean. If standard deviation is relatively larger, then data points are

widely spread out from the mean. Otherwise, data points stay close from the mean. Also, standard deviation is obtained by taking squared root from variance which is dependent on the choices of data points as well. To use sample standard deviation as an estimator to population standard deviation, we still need to adjust it.

Definition 1.11 (Sample Standard Deviation).

Let $x_1, x_2, x_3, \dots, x_n$ be a sample of data points. The standard deviation of the sample of data points (s) is:

$$s = \sqrt{s^2}, \quad \text{for } s > 0.$$

Median and Mode

The median and mode are two important measures of central tendency used in statistics to summarize and understand data. The median represents the middle value in a sorted dataset, giving a sense of the center that is not affected by extreme values or outliers. In contrast, the mode is the value that appears most frequently in a dataset, making it useful for identifying common or repeated observations.

Definition 1.12 (Median).

Let: $x_1, x_2, x_3, \dots, x_n$ be a collection of data points which is arranged in ascending order from the smallest value to the largest value (or descending order from the largest value to the smallest value in that collection). The median of the given collection of data points is the middle value in that collection, which equally spreads the collection into two parts. Half of all the collection values are above the median value and the rest of the values in the collection is below the median value.

- *Case 1: when n is an odd number. (i.e. 1, 3, 11, 237, ...). Then, the median M is defined as:*

$$M = \frac{n+1}{2}, \text{ where } n \text{ represents the } n^{\text{th}} \text{ position.}$$

- *Case 2: when n is an even number (i.e. 2, 6, 100, 500, ...). Then, the median M is: the average value of $\frac{n}{2}$'s and $\frac{n+2}{2}$'s position, where n represents the n^{th} position.*
-

Now, let's introduce mode.

Definition 1.13 (Mode).

It refers to a value that appears the most frequent than the appearance of all other values in a given dataset.

Percentile and Quartile

Percentiles and quartiles are statistical measures used to describe the distribution of data. A percentile indicates the value below which a given percentage of observations fall, helping to understand relative standing within a dataset. Quartiles, a specific type of percentile, divide the data into four equal parts (Q1, Q2/median, and Q3), providing insights into the spread and central tendency.

Definition 1.14 (Percentile and Quartile).

Let: x_1, x_2, \dots, x_n be a collection of data points in either ascending order. Percentile is denoted as: p^{th} , which indicates $p\%$ of observations are below to a such value. Quartiles, are special cases of percentile which equally spread the collection of data into four parts. Each part contains 25% of the entire collection. More specifically, we define quartiles as the following:

- Q_1 : the 25 percentile (or 25^{th}), which shows that 25% of the data points are below the value Q_1 .
- Q_2 : the 50 percentile (or 50^{th}), which shows that 50% of the data points are below the value Q_2 .
- Q_3 : the 75 percentile (or 75^{th}), which shows that 75% of the data points are below the value Q_3 .
- Q_2 is qual to median.

Moreover, we use $Q_3 - Q_1$ to calculate interquartile range (I.P.R), which shows the spread of the whole data set.

Skewness and Symmetry

The two terms 'skewness' and 'symmetry' are used to describe the shape of probability distribution. There are two types of skewness: left (or negative) skew and right (or positive) skew. In real life, a famous distribution highly used in hypothesis testing which is χ_n^2 with n degrees of freedom, is right skewed probability distribution function. Another example regarding to symmetry is normal distribution such that its probability under its curve greater than μ is same as the probability below than μ . Now let's introduce the proper definition of skewness and symmetry.

Definition 1.15 (Skewness).

Skewness refers to such a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive, zero, negative or undefined.

Now, let's break down the main definition of skewness and symmetry:

Definition 1.16 (Left (or Negative) Skew).

By observing given probability distribution curve, if the left tail of the curve is longer than the right tail the mass of the distribution is concentrated on the right of the figure, then we say that probability distribution is left skew or negative skew. (See figure below)

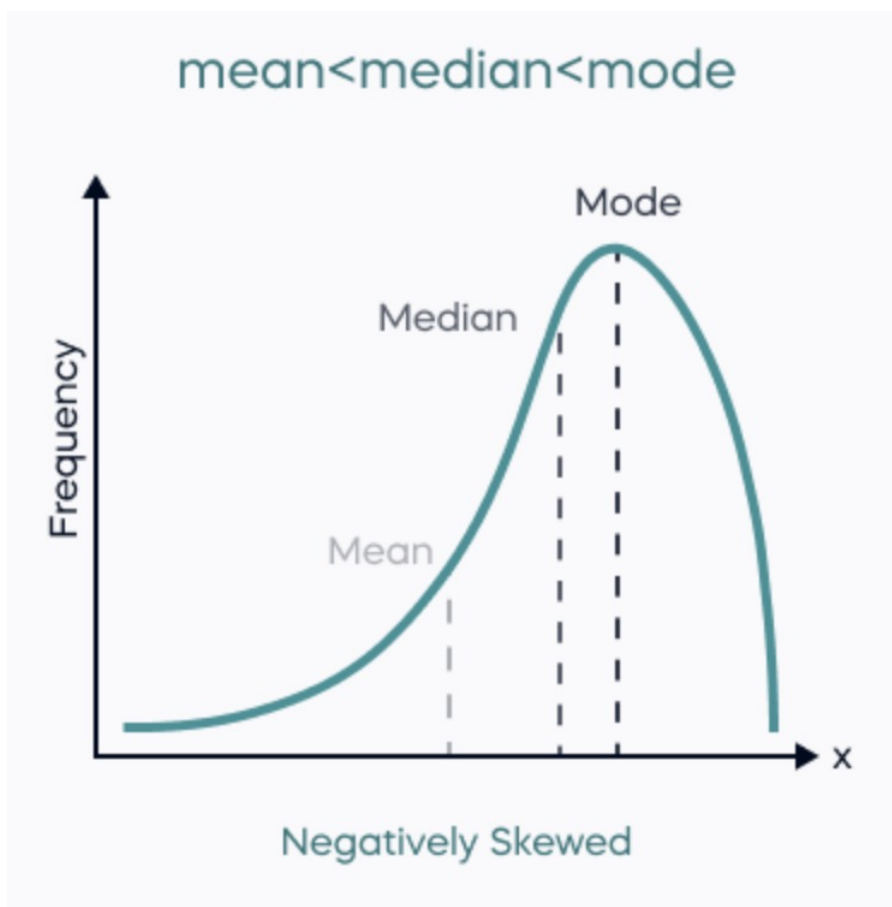


Figure 1.1: Visualization of left skew probability distribution

Definition 1.17 (Right (or Positive) Skew).

By observing given probability distribution curve, if the right tail of the curve is longer than the left tail the mass of the distribution is concentrated on the left of the figure, then we say that probability distribution is right skew or positive skew. (See figure below)

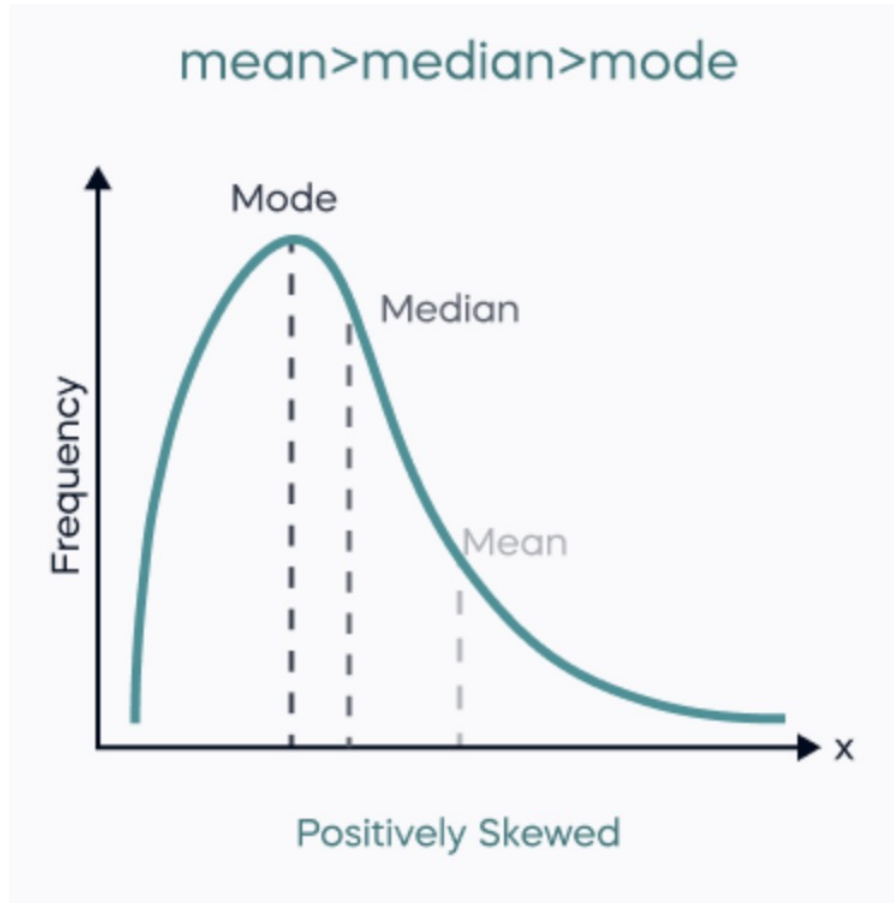


Figure 1.2: Visualization of right skew probability distribution

Symmetry is a special case of skewness when the value of skewness is 0.

Definition 1.18 (Symmetry). —————

In statistics, symmetry is a probability distribution is reflected around a vertical line at some value of the random variable represented by the distribution. Probability under the curve below that value is equal to probability under the curve greater than that value. (see figure below)

Since symmetry is a special case, so that it has a unique property as the following:

Theorem 1.1 (Empirical Rule (or 68 – 95 – 99.7 Rule)). *For any symmetric (bell-shaped) curve, let μ be its mean and σ be its standard deviation, the following probability set function is true:*

- 1. $P(\mu - \sigma < X < \mu + \sigma) = 68.27\%$;
- 2. $P(\mu - 2\sigma < X < \mu + 2\sigma) = 95.45\%$;
- 3. $P(\mu - 3\sigma < X < \mu + 3\sigma) = 99.73\%$.

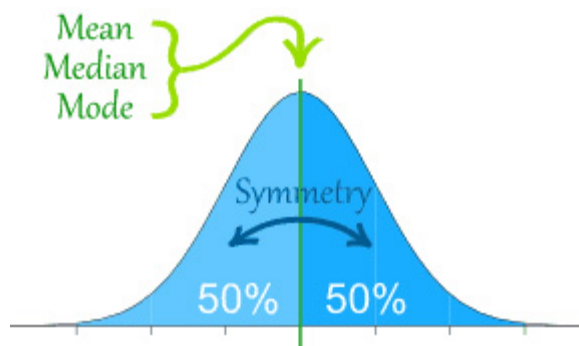


Figure 1.3: Visualization of symmetric probability distribution

Practice Example

Example 1.1.

[Calculating Sample Mean, Variance and Standard Deviation] Let: $x_1 = 1, x_2 = 3$ and $x_3 = 7$. Calculate the sample mean, sample variance and sample standard deviation for this collection of data points.

Solution (all results are kept in four digits):

By Definition 1.9, 1.10, 1.11, sample mean:

$$\bar{x} = \frac{1 + 3 + 7}{3} \approx 3.6667.$$

Then, we use sample mean to calculate sample variance:

$$s^2 = \frac{1}{3-1} \times [(1 - 3.6667)^2 + (3 - 3.6667)^2 + (7 - 3.6667)^2] \approx 9.3333.$$

Finally, we take the square root of sample variance to get sample deviation, and remember that $s > 0$:

$$s = \sqrt{s^2} \approx 3.0551.$$

Example 1.2.

[Median Calculation] Given two distinct collections of data points: $S_1 = \{2, 4, 6\}$ and $S_2 = \{1, 5, 16, 28\}$. Calculate the median of both two sets.

Solution:

For S_1 , since $n = 3$ which is an odd number, so by *Definition 1.3*, $M_{S_1} = 4$. For S_2 , $n = 4$ in this case, so that we need to calculate the average of $\frac{n}{2}$ and $\frac{n+1}{2}$. Then,

$$M_{S_2} = \frac{5 + 16}{2} = 10.5.$$

Example 1.3.

Consider the data set $S = \{4, 25, 30, 30, 30, 32, 32, 35, 50, 50, 50, 55, 60, 74, 110\}$. Calculate its median and Q_1 (25^{th}).

Solution:

Simply counting the number of data points, $n = 15$, such that $M_S = \frac{15+1}{2} = 8$. Thus, the 8^{th} value in the set which is 35.

Since we know the median of this collection of data points, we just need to find the median of the lower half of this data, which is exactly going to be 25 percentile (25^{th}). In the lower half of the given collection (all values below the median), $n_{lower} = 7$. By *Definition 1.3*, then median of the lower half (25^{th}) is going to be:

$$25^{th} = \frac{7+1}{2} = 4, \text{ the } 4^{th} \text{ position in the data set.}$$

Thus, Q_1 (25^{th}) = 30. To find Q_3 (75^{th}), apply the same strategy will guide you to find the correct answer, and we leave this as an exercise to you.

1.3 Graphical Techniques

In statistics, there are lots of types of graph to illustrate data, for example histograms and box-plots. This technique is used in the field of statistics for data visualization. Our objective is to both be able to identify some classical types of graph and interpret key statistical values (descriptive statistical values) from it.

1.3.1 Histograms

Introduction to Histograms

Histogram is a graphical representation of data that uses bars to display the frequency distribution of a dataset. Unlike bar graphs, which represent categorical data, histograms group numerical data into intervals (bins) and show how many values fall into each range. This makes histograms ideal for visualizing the shape, spread, and central tendency of continuous data, helping identify patterns such as symmetry, skewness, and outliers.

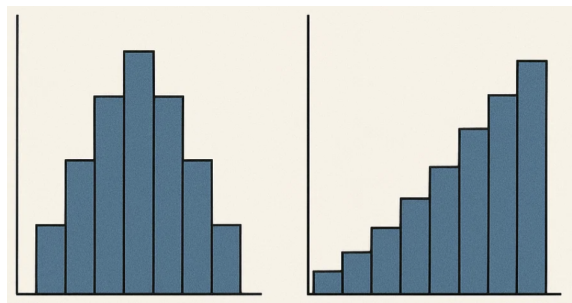


Figure 1.4: Visualization of histograms

Advantages and Disadvantages of Histograms

- Advantages of Histograms:
 1. Histograms are easily to used for visualise data (relatively). It allows us to get the idea of the "shape" of distribution (i.e. skewness which will be discussed late in this section).
 2. It is also flexible that people are able to modify bin widths.
- Disadvantages of Histograms:
 1. It is not suitable for small data sets.
 2. The values from histograms close to breaking points are likely similar, in fact they need to be classified into different bins (i.e. Student A and B scores 79 and 80 respectively in STA258, we consider a breaking point between 79 and 80. The two students have similar score, but student A is *B+* and student B is *A-* in GPA from).

Histograms with Skewness and Symmetry

A histogram visually represents the distribution of numerical data, making it a useful tool for assessing skewness and symmetry. It is quite straightforward to estimate the skewness of histograms by simply drawing a curve above bins on the histogram.

For a histogram to have a left (or negative) skew probability distribution:

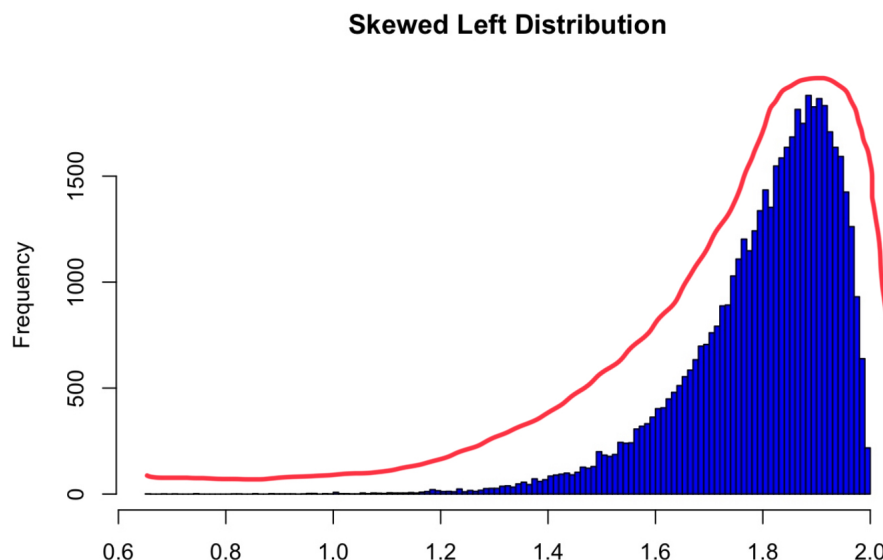


Figure 1.5: Visualization of a histogram has a left (or negative) skew probability distribution

For a histogram to have a right (or positive) skew probability distribution:

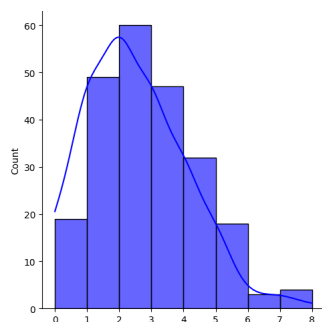


Figure 1.6: Visualization of a histogram has a right (or positive) skew probability distribution

For a histogram to have a symmetric probability distribution:

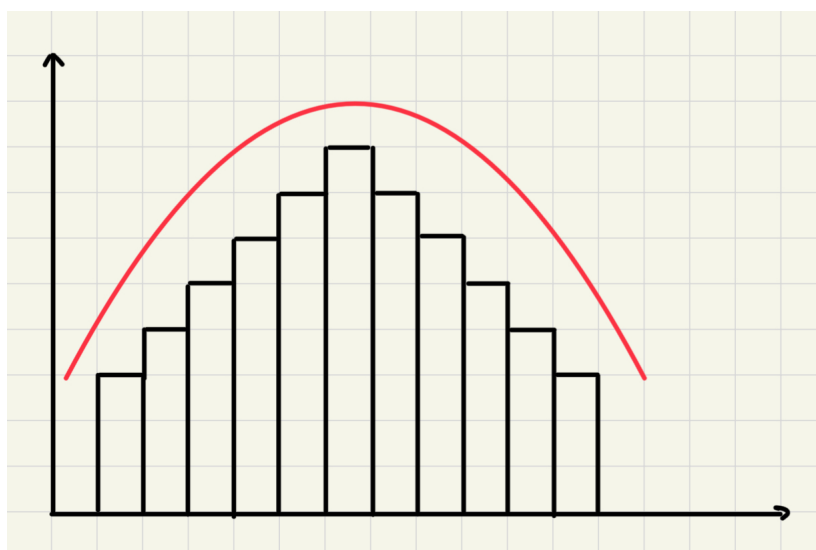


Figure 1.7: Visualization of a histogram has a symmetric probability distribution

1.3.2 Box-Plots

A boxplot (or box-and-whisker plot) is a standardized way to display data distribution based on a five-number summary: minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. The box represents the interquartile range (IQR), while the whiskers show variability outside Q1 and Q3. Outliers are plotted as individual points. Boxplots efficiently compare distributions and highlight skewness, spread, and outliers. (See figure below)

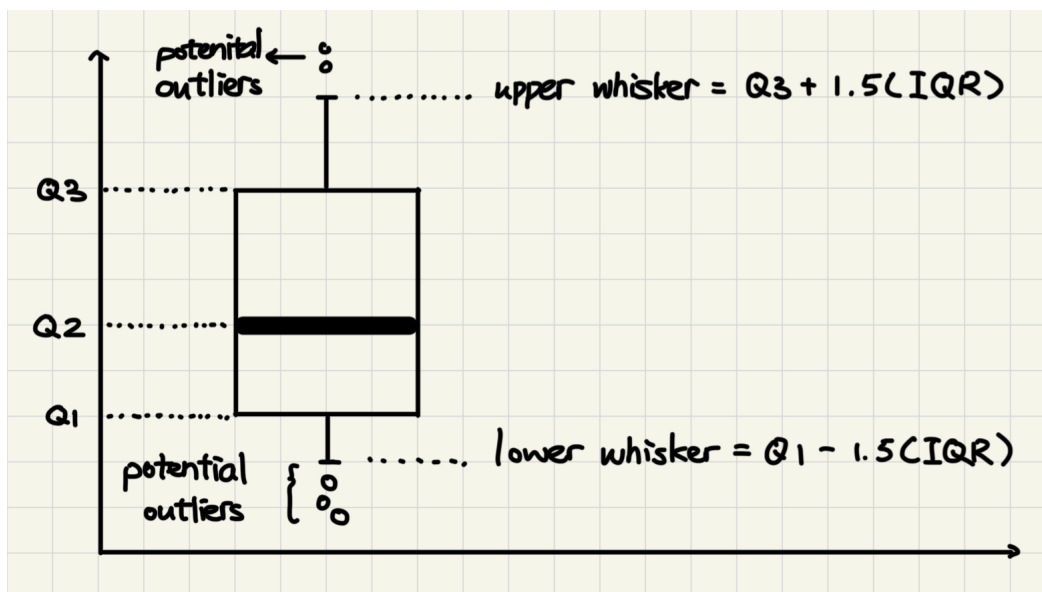


Figure 1.8: Visualization of a box-plot

Similar to histograms, we can still obtain information about skewness and symmetry, by observing the cut from the line of $Q2$.

If the median ($Q2$) cuts the box with upper area smaller than lower area, then we say that box-plot with left skew probability distribution. Or, if the median ($Q2$) cuts the box with upper area larger than lower area, then we say that box-plot with right skew probability distribution.

Otherwise, if the median ($Q2$) cuts the box with upper area equal to lower area, then we say that box-plot with symmetric probability distribution.

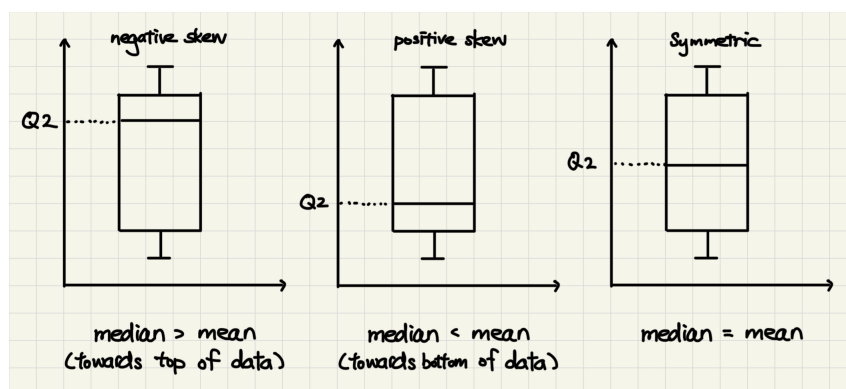


Figure 1.9: Visualization of a box-plot with skew and symmetric probability distribution

1.4 Introduction to R

R is used for data manipulation, statistics, and graphics. It is made of: operations ($+$, $-$, $<$) which is for calculations on vectors, arrays and matrices; a huge collection of functions; fa-

cilities for making unlimited types quality graphs; user contributed packages (sets of related functions); the ability to interface with procedures written in C, C+, or FORTRAN and to write additional primitives. R is also an open-source computing package which has seen a huge growth in popularity in the last few years (Please use this website: <https://cran.r-project.org>, to download R).

What is R-studio?

RStudio is a relatively new editor specially targeted at R. RStudio is cross-platform, free and open-source software (Please use: <https://www.rstudio.com>, to download Rstudio).

Make a Histogram Using R-studio

This is just a demonstration of how to start and use R-studio.

1. First of all, we need to know which dataset are we going to make into a histogram. In this case, as an example, we are going to use the waiting time in faithful in R-studio.
2. For any dataset, use the code: `names(faithful)` to get it. (inside the parentheses, type the names of variables you want in faithful dataset)
3. Then, we proceed with the code: `hist(faithful$waiting)` to get a basic plot.

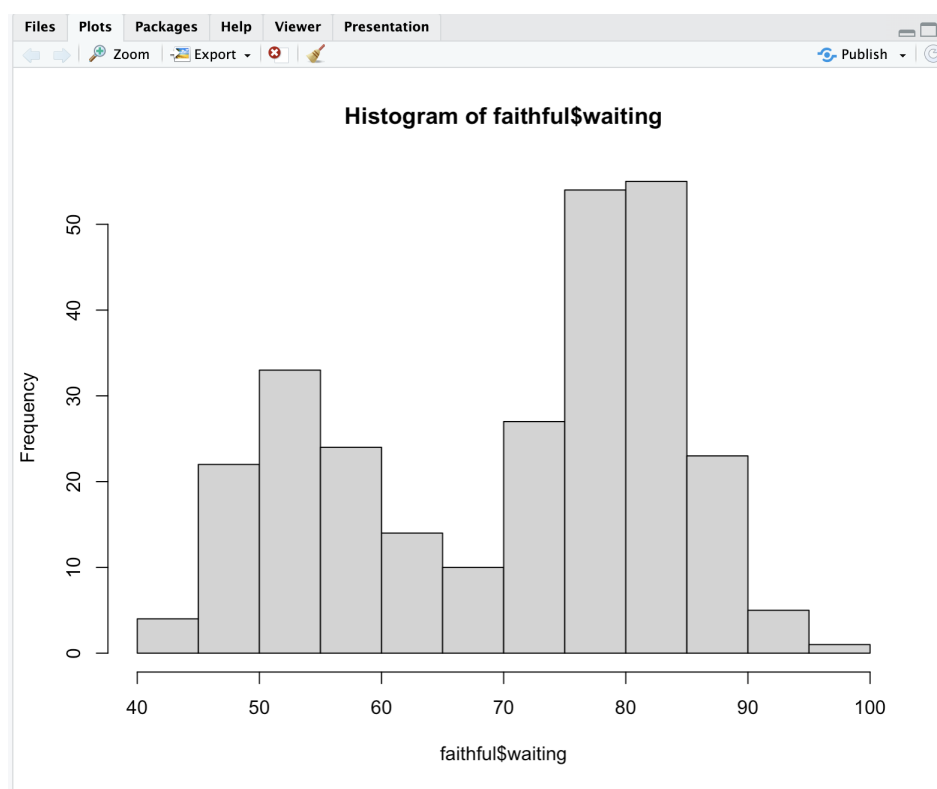


Figure 1.10: R-studio first three steps (by following the instructions, you should get this histogram)

4. Furthermore, we can also get more information. For example, by keep proceeding with the code: `hist(faithful$waiting,plot=FALSE)$breaks`, R-studio will show you all the breaking points between histogram cells.

```
> hist(faithful$waiting)
> hist(faithful$waiting,plot=FALSE)$breaks
[1] 40 45 50 55 60 65 70 75 80 85 90 95 100
```

Figure 1.11: R-studio the forth step(by following the instructions, you should get this histogram)

Chapter 2

Sampling Distributions Related to a Normal Population

Previously, we have introduced lots of definitions and given you a rough idea about what really statistics it and what people do in statistics. Now, we are going to proceed statistical distributions.

2.1 Normal Distribution

In probability theory and statistics, normal distribution also called Gaussian distribution which is discovered by a famous German mathematician Johann Carl Friedrich Gauss in 1809. It is one of the most important distribution that used to approximate other types of probability distribution, such as binomial, hypergeometric, inverse (or negative) hypergeometric, negative binomial and Poisson distribution. Generally, it is denote as $N(\mu, \sigma^2)$ with probability density function as the following:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Formally, let's begin with its definition:

Definition 2.1 (Normal Distribution).

Suppose a random variable $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$ and $Var(X) = \sigma^2$. And $-\infty < \mu < \infty, \sigma^2 > 0$. Moreover, X has probability density function as:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ for } -\infty < x < \infty \text{ (same as above).}$$

The only special case of normal distribution is standard normal distribution, such that a random variable $Y \sim N(\mu = E(Y) = 0, \sigma^2 = Var(Y) = 1)$, then Y has probability density function as:

$$f(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}.$$

2.2 Gamma and Chi-square Distribution

The Chi-square and Gamma distributions are two fundamental probability distributions widely used in statistical theory and applications. The Gamma distribution is a continuous distribution characterized by its shape and scale parameters, making it versatile for modeling waiting times and various positively skewed data. The Chi-square distribution, a special case of the Gamma distribution, arises naturally in the context of hypothesis testing and confidence interval estimation, especially in tests involving variance and categorical data.

Gamma Distribution

Definition 2.2 (Gamma Distribution).

Suppose a random variable X is Gamma distributed with $\alpha > 0$ (shape parameter) and $\beta > 0$ (scale parameter) if and only if the probability density function of X is

$$f(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}, \text{ for } 0 < x < \infty.$$

Then, $E(X) = \alpha\beta$, $Var(X) = \alpha\beta^2$ and its moment generating function is $M_X(t) = \frac{1}{(1-\beta t)^\alpha}$, for $t < \frac{1}{\beta}$.

Now, let's introduce some properties of Gamma function:

- Gamma function (**not a distribution**):

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \text{ for } x > 0.$$

- Properties

- 1. $\Gamma(x) = x \cdot \Gamma(x-1)$;
- 2. For all $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$;
- 3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Chi-square Distribution

Here is its formal definition:

Definition 2.3 (Chi-square Distribution).

A random variable X has a Chi-squared distribution with n degrees of freedom (χ_n^2) if and only if X is a random variable with a Gamma distribution with parameters $\alpha = \frac{n}{2}$ and $\beta = 2$. Then, the probability density function of X is given by

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}.$$

Moreover, $E(X) = n$, $Var(X) = 2n$ and moment generating function of X is $M_X(t) = (1 - 2t)^{-\frac{n}{2}}$, for $t < \frac{1}{2}$.

We claim that Chi-square distribution is a special case of Gamma distribution with $\alpha = \frac{n}{2}$ and $\beta = 2$. Now, let's prove it by using moment generating function.

The proof is quite straightforward as the following shows:

Proof. Suppose $X \sim \text{Gamma}(\alpha = \frac{n}{2}, \beta = 2)$.

Then the following moment generating function holds for X :

$$M_X(t) = (1 - 2t)^{-\frac{n}{2}}, \text{ for } t < \frac{1}{2}.$$

Compare the moment generating function of X under Gamma distribution with Chi-square distribution, we can conclude that $X \sim \chi_n^2$. \square

Obtaining Chi-square Distribution by Normal Distribution

Previously, we showed how to use Gamma distribution to get Chi-square distribution by moment generating function method. Now, let's do something interestingly, to use normal distribution to get Chi-square distribution. We will begin with a theorem, then prove it.

Theorem 2.1.

Suppose a random variable Z is standard normally distributed, such that $Z \sim N(0, 1)$. Then, Z^2 is Chi-square distributed with 1 degree of freedom, so that $Z^2 \sim \chi_1^2$.

The proof of Theorem 2.1 isn't that trivial to see. We still need moment generating function, but in a different way. Before we get into the proper proof, let's grab everything we need:

- 1. Recall STA256 about how to get moment generating function for a given continuous random variable that:

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx.$$

- 2. We also need Gaussian integral:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}; \tag{2.2.1}$$

$$\int_{-\infty}^{\infty} e^{-kx^2} dx = \sqrt{\frac{\pi}{k}}, \text{ for } k > 0; \tag{2.2.2}$$

$$\int_{-\infty}^{\infty} e^{kx^2} dx = \sqrt{\frac{\pi}{-k}}, \text{ for } k < 0. \tag{2.2.3}$$

Proof. Suppose that $Z \sim N(0, 1)$, then $f_Z(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$.

$$\text{Next, computing M.G.F for } Z^2: M_{Z^2}(t) = \mathbb{E}(e^{tz^2}) = \int_{-\infty}^{\infty} e^{tz^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz \quad (2.2.4)$$

$$\text{After rearranging: } M_{Z^2}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(\frac{1}{2}-t)z^2} dz \quad (2.2.5)$$

$$\text{Let: } u = z\sqrt{\frac{1}{2}-t}, \text{ and then } dz = \frac{1}{\sqrt{\frac{1}{2}-t}} du \quad (2.2.6)$$

$$\text{By substitution: } M_{Z^2}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2} \cdot \frac{1}{\sqrt{\frac{1}{2}-t}} du \quad (2.2.7)$$

$$\text{Rearranging and using 2.2.1 from above: } M_{Z^2}(t) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\frac{1}{2}-t}} \cdot \sqrt{\pi} \quad (2.2.8)$$

$$\text{Then: } M_{Z^2}(t) = (1-2t)^{-\frac{1}{2}}, \text{ which is the M.G.F of } \chi_{n=1}^2. \quad (2.2.9)$$

□

Now, we can do another proof by using Theorem 2.1.

Theorem 2.2.

Suppose $Z_1, Z_2, \dots, Z_n \stackrel{i.i.d.}{\sim} N(0, 1)$, then the sum of n independent Z^2 is going to be Chi-square distributed with n degrees of freedom, as the following:

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2.$$

We need Theorem 2.1 to prove this, but it going to be easier.

Proof. Assume that $Z_1, Z_2, \dots, Z_n \stackrel{i.i.d.}{\sim} N(0, 1)$, then we compute the M.G.F for the entire summation as well, and we let $\delta = \sum_{i=1}^n Z_i^2$, which as the following shows:

$$M_\delta(t) = E[e^{t\delta}] = E[e^{t(Z_1^2 + \dots + Z_n^2)}] = E[e^{tZ_1^2} \dots e^{tZ_n^2}].$$

Since Z_1, Z_2, \dots, Z_n are independent and identically distributed, so that:

$$M_\delta(t) = E[e^{tZ_1^2}] \dots E[e^{tZ_n^2}].$$

By Theorem 2.1, we have: $Z^2 \sim \chi_1^2$, then:

$$M_\delta(t) = E[e^{t\delta}] = E[e^{tZ_1^2} \dots e^{tZ_n^2}] = [(1 - 2t)^{-\frac{1}{2}}] \dots [(1 - 2t)^{-\frac{1}{2}}].$$

Since there are n individuals of Z^2 , hence:

$$M_\delta(t) = (1 - 2t)^{-\frac{n}{2}}, \text{ which is the M.G.F of } \chi_n^2, \text{ as required.}$$

□

Here is the last theorem for Chi-square and normal distribution, but we won't show you the proof due to its complexity. For people who are interested in that, please see STA260 lecture notes or power point slide to figure out.

Theorem 2.3.

Let n be sample size, s^2 be sample variance and σ^2 be population variance, then $\frac{(n-1)s^2}{\sigma^2}$ is Chi-square distributed with $n - 1$ degrees of freedom. As the following:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

2.3 Student's t-Distribution and F-Distribution

The t-distribution and F-distribution are essential tools in inferential statistics, particularly in the context of hypothesis testing and variance analysis. The t-distribution, which resembles the normal distribution but with heavier tails, is primarily used when estimating population means in situations where the sample size is small and the population standard deviation is unknown. On the other hand, the F-distribution is used to compare variances between two populations and plays a central role in analysis of variance (ANOVA) and regression analysis.

Student's t-Distribution

Definition 2.4 (Student's t-Distribution).

Suppose X is t-distributed with n degrees of freedom, then the probability density function of X is given by:

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Alternatively, define a new variable T is the following:

$$T = \frac{W}{\sqrt{\frac{V}{r}}}, \text{ for } W \sim N(0, 1) \text{ and } V \sim \chi_r^2.$$

Or suppose $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. Thus,

$$T = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}.$$

Same as normal distribution, student's t-distribution is also symmetric. Also, as the degrees of freedom of t-distribution getting larger, the curve of student's t-distribution getting closer to standard normal distribution.

F-Distribution

Definition 2.5.

We define a new variable F as the following shows:

$$F = \frac{\left(\frac{W_1}{v_1}\right)}{\left(\frac{W_2}{v_2}\right)} \sim F_{v_1, v_2}; \text{ for } W_1 \sim \chi_{v_1}^2 \text{ and } W_2 \sim \chi_{v_2}^2; \text{ also both } W_1 \text{ and } W_2 \text{ are independent.}$$

Alternatively, we select two samples (with same population variance) with size n and m , and also sample variance s_x and s_y respectively. Then, F -distribution is:

$$F = \frac{\left[\frac{\left(\frac{n-1}{2}\right)\sigma^2 s_x^2}{n-1}\right]}{\left[\frac{\left(\frac{m-1}{2}\right)\sigma^2 s_y^2}{m-1}\right]} \sim F_{n-1, m-1}.$$

Both student's t-distribution and F-distribution are highly used in inferential statistics, until confidence interval, testing hypothesis and ANOVA analysis, these two distributions will come to play a lot. At this point, just guarantee that you know how to obtain those distribution from random given information is sufficient.

Chapter 3

The Central Limit Theorem

The Central Limit Theorem (CLT) is one of the most important results in probability and statistics. It states that, given a sufficiently large sample size, the distribution of the sample mean of independent and identically distributed (i.i.d.) random variables approaches a normal distribution, regardless of the shape of the original distribution. Real-life Application of Central Limit Theorem in Financial Analysis. The CLT is often used by financial experts to examine stock market results.

Now, let's discuss Central Limit Theorem with more details. Suppose we have a finite number of populations and each population follows a distribution with population mean μ and population variance σ^2 . Then we take samples of same size n from each population, such that we have $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ from population group 1 to m , respectively. Next, we make a histogram using the large collection of sample taken from each population group. Then, what we are doing right now is sampling distribution of \bar{x} . As a result, \bar{x} follows a normal distribution with mean $\mu_{\bar{x}} = \mu$ and variance $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$, which is denoted as the following:

$$\bar{x} \sim N(\mu_{\bar{x}} = \mu, \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}).$$

There is a figure shows the entire procedure from above paragraph.

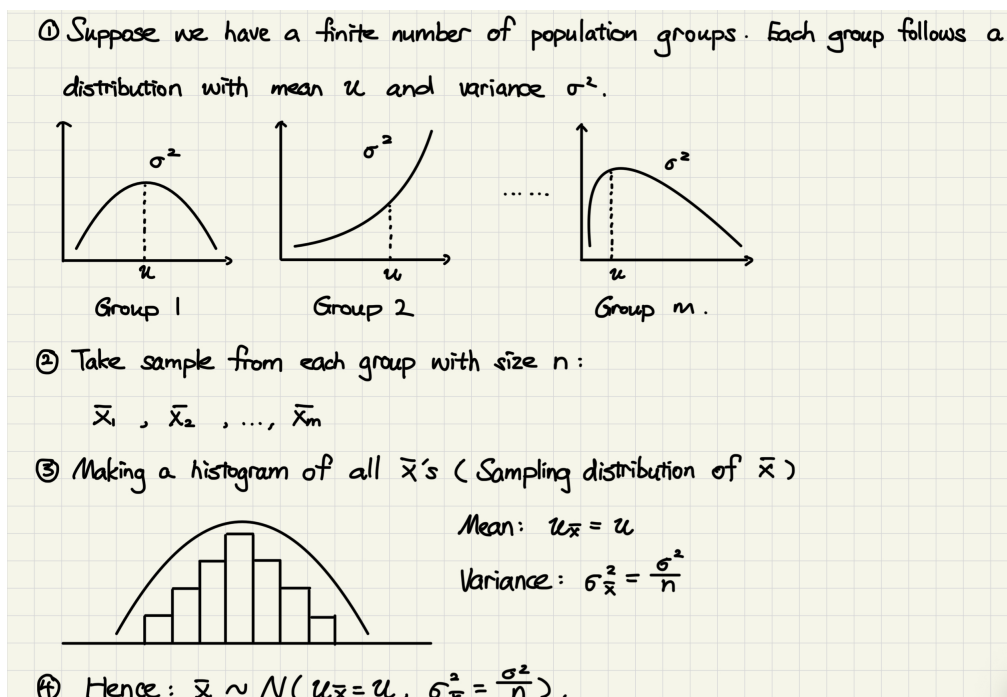


Figure 3.1: Procedure of the Central Limit Theorem

Now, let's begin with the proper definition of Central Limit Theorem.

Definition 3.1 (Central Limit Theorem). —————

Let X_1, X_2, \dots, X_n be independent and identically distributed random variables with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$. Then, we define the following:

$$U_n = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(\mu = 0, \sigma^2 = 1), \text{ where } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Then the distribution function of U_n converges to the standard Normal distribution function as $n \rightarrow \infty$. That is,

$$\lim_{n \rightarrow \infty} P(U_n \leq u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt; \text{ for all } u.$$

For this course in particular, we do not need to pay that much attention to the proving part of the definition above. However, we use Central Limit Theorem to approximate distributions. Here are the two important approximations:

•

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right);$$

-

$$T = \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2).$$

A reminder that the distribution of U_n in definition 3.1 and the two approximation of distribution above are extremely important in this course, until later chapters you may see some materials that are similar.

Chapter 8

One Sample Confidence Intervals On a Proportion

Perviously, we have introduced two types of confidence interval based on known and unknown variance. Moreover, confidence intervals are also applied to an unknown population proportion. For example, suppose we are interested the proportion of total number of left-handed students among all students who are currently studying at University of Toronto Mississauga. The question is: how do we know such the parameter which estimates the proportion of left-handed students at UTM? While, it is impossible to proceed it directly by counting both the total number of students and all left-handed students at UTM, due to the complexity and the total workload of that task. Then, we have to work with confidence intervals.

Firstly, we take a random sample of students at UTM, then we calculate how many students are left-handed by dividing total number of left-handed students in that sample with total number of students in it, and denote the proportion as \hat{p} . Next we begin our confidence interval calculation to get a range of number with a certain level of confidence.

Now let's begin with the proper definition of confidence interval on proportion.

Definition 8.1 (One Sample Confidence Intervals On a Proportion). —————
*We select a random sample of size n from a population with **unknown** proportion p of success. An approximate confidence interval for p is:*

$$p = \hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \text{ where } \hat{p} = \frac{\text{number of observations satisfying the criteria}}{n}.$$

In addition, n is the sample size.

To apply this confidence interval, there are 3 conditions that we need to guarantee:

- 1. Random sample;
- 2. Independent and identically distributed Bernoulli trials;
- 3. We have a large chosen sample size ($n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$).

A good way to understand confidence interval is visualization. Now, suppose we have a valid estimation \hat{p} . After the entire procedure of confidence interval, our population proportion (p) should be as the following number line shows:



Remember that your final answer of the range of p must be between 0 and 1, since we are working with proportion.

Summary about One Sample Confidence Intervals

We have introduced one sample confidence interval under three different cases: given population variance, unknown population variance and unknown population proportion. All the material of one sample confidence interval comes from chapter 6, 7, 8, which seems like you to remember a lot. However, the reason why we give this summary is to help you to remember the basic skeleton of one sample confidence interval. Let's revisit the three distinct types of confidence interval:

- 1. For given population variance, we have: $\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$;
- 2. For unknown population variance, we have: $\bar{x} \pm t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$;
- 3. To estimate population proportion, we have: $p = \hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

The question is: what is the similarity between the three distinct types of confidence interval? You may have already noticed that the confidence intervals above are all follow such a skeleton that \bar{x} plus or minus its margin of error (different between each type of C.I.).

If we keep questioning ourselves that how the margin of error comes from, you will catch the pattern. The margin of error contains reference distribution and the standard deviation of \bar{x} under its reference distribution. Let's analyze each type of confidence interval to prove my statement is true:

The first type (given population variance) confidence interval is quite easy to recognize. Recall chapter 3: The Central Limit Theorem, we state that $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$, which is how reference distribution comes from with given population variance. To get the standard deviation of \bar{x} , we simply take the square root of the variance, then $s_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Finally, we multiply reference distribution at the point $\frac{\alpha}{2}$ with the standard deviation of \bar{x} under normal distribution to get the margin of error.

The second type (unknown population variance) confidence interval is similar to the first type, but the reference distribution is t-distribution instead of normal distribution. In chapter 7, we introduced the calculation of sample variance (s^2) to estimate population variance (σ^2), and s^2 is an unbiased estimator of σ^2 (the proof of this statement is in STA260, in

this course we can assume it freely). In chapter 2, we defined that $T = \frac{\bar{x} - \mu}{(\frac{s}{\sqrt{n}})} \sim t_{n-1}$, then the term $\frac{s}{\sqrt{n}}$ (verify this by yourself as an extra exercise) is the standard deviation of \bar{x} under t-distribution with $n - 1$ degrees of freedom. Finally, we multiply multiply reference distribution at the point $\frac{\alpha}{2}$ with the standard deviation of \bar{x} under t-distribution to get the margin of error.

The third type (estimate population proportion) confidence interval is slightly harder to identify. Recall chapter 4 that we can approximate binomial distribution by normal distribution. Suppose that a random variable $X \sim \text{Binomial}(n, p)$, then the random variable $X \sim N(np, np(1-p))$. From chapter 4, we know that $\hat{p} \sim N(\mu_p = p, \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n})$. Trivially, the standard deviation of \hat{p} is $\sqrt{\frac{p(1-p)}{n}}$, since we don't know the value p and use \hat{p} to estimate that. Then, standard deviation of \hat{p} is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Finally, by multiplying the reference distribution and standard deviation, we get the margin of error of the confidence interval.

Conclusion About One Sample Confidence Intervals

If you follow the skeleton below, one sample confidence interval will become easier:

- 1. Identify what type of one sample confidence interval to use from given information;
- 2. Construct your confidence interval that fits the circumstance which you are facing, it either going to be $\bar{x} \pm M.E.(\bar{x})$ or $\hat{p} \pm M.E.(\hat{p})$;
- 3. Check the validity of your final answer. For example, the range of p must between 0 and 1.
- 4. Clearly state your final conclusion: we have a certain percentage of confidence to guarantee that the value of the chosen sample is between its lower bound and upper bound.

Index

Introduction, [2](#), [3](#)

Overview, [1](#)

Statistics

Definition, [2](#)

Introduction, [1](#)