# Contents

1 Introduction: Making Decisions Under Uncertainty .................................................. 1
  1.1 Probability Concepts .................................................................................................. 1
      1.1.1 Notation .............................................................................................................. 1
      1.1.2 What exactly is a probability? ............................................................................ 2
  1.2 Axioms of Probabilities ............................................................................................ 4
  1.3 Some Useful Formulas for Computing Probabilities ............................................... 4
      1.3.1 Unions and Intersections .................................................................................... 4
      1.3.2 Complements ..................................................................................................... 7
  1.4 Conditional Probabilities and Independent Events ................................................... 8
      1.4.1 Conditional Probabilities ................................................................................... 8
      1.4.2 Independent Events .......................................................................................... 10
  1.5 Probability Tree Diagrams ....................................................................................... 12
  1.6 Bayes’ Theorem ....................................................................................................... 13
  1.7 Counting Principles ................................................................................................ 15
      1.7.1 Permutations ...................................................................................................... 15
      1.7.2 Combinations ................................................................................................... 17

2 Risk Management: Profiting From Uncertainty ............................................................ 19
  2.1 Random variables ..................................................................................................... 19
  2.2 General probability distributions for discrete random variables ................................ 20
      2.2.1 Calculating probabilities over regions ............................................................... 21
  2.3 Binomial Random Variables .................................................................................... 21
  2.4 Conditional probabilities and independent random variables .................................. 24
      2.4.1 Conditional probabilities ................................................................................... 25
      2.4.2 Independent random variables .......................................................................... 25
  2.5 Expected values and variances of random variables ................................................ 27
      2.5.1 Expected values .................................................................................................. 27
      2.5.2 Variance and Standard Deviation of a random variable ..................................... 31
  2.6 Poisson probability distributions ............................................................................. 36
  2.7 Covariance and correlation ....................................................................................... 38
      2.7.1 Covariance ......................................................................................................... 38
      2.7.2 Correlation ........................................................................................................ 42

3 Continuous Random Variables .................................................................................... 45
  3.1 General Probability Distributions ........................................................................... 45
  3.2 Expected Values and Variances of Continuous Random Variables ......................... 48
      3.2.1 Expected Values .................................................................................................. 48
## CONTENTS

3.2.2 Variance ................................................. 49
3.3 The Exponential Distribution ...................................................... 50
3.4 The Normal Distribution .......................................................... 52
  3.4.1 The Standard Normal Distribution ........................................ 55
  3.4.2 Back to the General Normal ............................................... 57
3.5 Adding Normal Random Variables .................................................. 60

### 4 What Does The Mean Really Mean? ..................................................... 63
  4.1 Sampling and Inference .................................................................. 63
    4.1.1 Expected value of $X$ .................................................. 65
    4.1.2 Variance of $X$ .............................................................. 66
  4.2 Taking a Random Sample from a Normal Population .......................... 67
  4.3 Central Limit Theorem .............................................................. 69
  4.4 Normal Approximation to the Binomial Distribution ......................... 72
    4.4.1 Calculating Probabilities for the Number of Successes, $X$ ...... 72
    4.4.2 Calculating Probabilities for the Proportion of Successes, $\hat{p}$ 74
  4.5 Distribution of $X_1 - X_2$ and $\hat{p}_1 - \hat{p}_2$ .......................... 75
    4.5.1 The distribution of $X_1 - X_2$ ........................................ 76
    4.5.2 The distribution of $\hat{p}_1 - \hat{p}_2$ ................................... 77

### 5 Estimators and Confidence Intervals ................................................. 79
  5.1 Estimators ............................................................................. 79
    5.1.1 Unbiased Estimators ....................................................... 79
    5.1.2 Minimum Variance Unbiased Estimators .............................. 81
    5.1.3 Are Unbiased Estimators Always the Best Estimators? (Not Examinable) ....... 81
  5.2 Confidence Intervals for Normal Means : $\sigma^2$ Known ............... 81
    5.2.1 What is a Confidence Interval? ........................................ 82
    5.2.2 Calculating a Confidence Interval .................................... 84
  5.3 Confidence Intervals for Normal Means : $\sigma^2$ Unknown ........... 87
    5.3.1 The $t$ distribution ....................................................... 88
    5.3.2 Using the $t$ distribution to construct confidence intervals ...... 89
  5.4 Estimating Binomial $p$ ............................................................. 91
  5.5 Sample Size ........................................................................ 93
  5.6 Confidence Intervals for $\mu_1 - \mu_2$ ......................................... 94
    5.6.1 Confidence intervals : $\sigma$ known ................................ 95
    5.6.2 Confidence intervals : $\sigma$ unknown ................................ 96
  5.7 Confidence Intervals for $p_1 - p_2$ ............................................. 97
    5.7.1 Confidence intervals for $p_1 - p_2$ ................................... 98

### 6 Hypothesis Tests ........................................................................ 101
  6.1 What is a Hypothesis Test? ......................................................... 101
  6.2 Establishing the Hypotheses ....................................................... 103
    6.2.1 Deciding on the Null and Alternative .............................. 103
    6.2.2 Some more examples ..................................................... 105
  6.3 Performing the Hypothesis Test .................................................... 105
  6.4 Type I and Type II Errors .......................................................... 106
    6.4.1 Some possible mistakes .................................................. 106
    6.4.2 Choosing $\alpha$ and $\beta$ ............................................... 107
6.4.3 The relationship between \( \alpha \) and the p-value .................................. 107
6.5 The Formal Steps in Conducting a Hypothesis Test ................................. 108
6.6 Hypotheses on \( \mu \) .............................................................................. 110
   6.6.1 \( H_0 : \mu = \mu_0 \) .......................................................... 110
   6.6.2 \( H_0 : \mu \leq \mu_0 \) or \( H_0 : \mu \geq \mu_0 \) ......................... 113
6.7 Hypotheses on \( p \) ........................................................................... 113
6.8 Hypotheses on \( \mu_1 - \mu_2 \) ............................................................... 117
6.9 Hypotheses on \( p_1 - p_2 \) ................................................................. 121
6.10 Paired Data ......................................................................................... 124
   6.10.1 Paired vs Independent Data .................................................. 124
   6.10.2 Hypothesis Tests on Paired Data ........................................... 126
   6.10.3 What are the differences? ......................................................... 127
Chapter 1

Introduction: Making Decisions Under Uncertainty

1.1 Probability Concepts

What is the probability

- of getting a red on a roulette wheel at Las Vegas?
- it will rain on the 25th of January?
- a patient will die in an operation?

We may have an intuitive idea of the answer to these questions but to answer them properly we need some definitions and notation.

1.1.1 Notation

**Definition 1** A Random Experiment or Random Trial is an operation that is repeatable under stable conditions and that results in any one of a set of outcomes; furthermore, the actual outcome cannot be predicted with certainty.

For example

- parts coming off a production line,
- selection of an invoice in an audit,
- rolling dice.

**Definition 2** A Sample Space, $S$, is the set of all possible outcomes of an experiment.

For example

- there are 38 numbers on a roulette wheel so $S = \{1, 2, 3, \ldots, 36, 0, 00\}$
- a part off a production line is faulty or not so $S = \{F, NF\}$
**Definition 3** An Event is a set of one or more outcomes of a random experiment. Or in other words a subset of $S$.

For example

- roulette ball lands on an odd number ($\{1, 3, 5, \ldots, 35\}$),
- part is faulty ($\{F\}$),
- rains on the 25th of January,
- patient lives.

We assign to each event a letter. For example

- $A = \{\text{ball lands on red}\}$
- $B = \{\text{ball lands on 1-9}\}$
- $C = \{\text{rains on the 25th of January}\}$

We then denote the probability of an event using $P$. For example, if we wished to talk about the “probability of a roulette ball landing on red” we would write

$$P(A) = P(\text{ball lands on red})$$

The obvious question then is how exactly do we calculate $P(A)$?

### 1.1.2 What exactly is a probability?

There are 3 different approaches to calculating the probability of an event.

**Relative Frequency Method**

**Definition 4** Using this method we define $P(A)$ as

$$P(A) = \lim_{n \to \infty} \frac{\text{Number of times } A \text{ occurs when the experiment is performed } n \text{ times}}{n}$$

or

$$P(A) \approx \frac{\text{Number of times } A \text{ occurs when the experiment is performed } n \text{ times}}{n} \quad (n \text{ large})$$

In other words $P(A)$ is defined as the long run fraction or frequency of the event $A$ if we repeated the experiment many times.

For example we might calculate the probability that it rains on the 25th of January using this method. Figure 1.1 provides an illustration. We can see that if we take the average over a large number of years the fraction of days that it rains on the 25th is approximately 0.3 so we would say that the probability is 0.3.
1.1. PROBABILITY CONCEPTS

Figure 1.1: The long run fraction of days that is has rained on the 25th of January over the last 1000 years (hypothetical).

Classical Method of Assigning Probabilities

To use this method we first need to define the concept of mutually exclusive events.

Definition 5 Two sets of outcomes (or equivalently events) are mutually exclusive if there is no outcome that belongs to both sets. Mathematically this means $A$ and $B$ are mutually exclusive if $A \cap B = \emptyset$.

For example

- Red and Black are mutually exclusive because it is not possible for a ball to land on Red and Black at the same time.

- However Red and Even are not mutually exclusive because it is possible for the ball to land on a Red and Even number at the same time.

Definition 6 If all the possible outcomes in an experiment are mutually exclusive and equally likely then using the Classical Method we define the probability of event $A$, $P(A)$, as

$$P(A) = \frac{\text{Number of outcomes that cause } A \text{ to happen}}{\text{Number of possible outcomes}}$$

For example if $B = \{\text{ball lands on } 1 - 9\}$ then

$$P(B) = \frac{\text{Number of ways that ball lands on } 1 - 9}{\text{Number of possible numbers the ball could land on}} = \frac{9}{38}$$

or if we are tossing a die and $C = \{1, 3\}$ i.e. the die lands with a 1 or 3 showing

$$P(C) = \frac{\text{Number of ways that die lands with } 1 \text{ or } 3 \text{ showing}}{\text{Number of possible ways the die could land}} = \frac{2}{6}$$
Subjective Probabilities

Definition 7 A subjective probability is an individual's degree of belief in the occurrence of an event.

For example a doctor may say that the probability a patient dies in an operation is 10%. The doctor will base this on prior operations performed on similar people but this number is still subjective because everyone is different.

1.2 Axioms of Probabilities

No matter which of the different methods for producing probabilities are used there are three Axioms or rules that all probabilities follow.

1. If \( A \) is an event then

\[
0 \leq P(A) \leq 1
\]

2. The sample space, \( S \), contains all possible outcomes. Thus

\[
P(S) = 1
\]

3. If event \( A \) is mutually exclusive of event \( B \), then

\[
P(A \cup B) = P(A) + P(B)
\]

For those of you that are not familiar with set notation, \( \cup \) stands for Union and can be thought of as meaning OR. So for example \( A \cup B \) is the set of all elements that are in \( A \) or \( B \) or both e.g. if \( A = \{1, 2, 3\} \) and \( B = \{3, 4, 5\} \) then \( A \cup B = \{1, 2, 3, 4, 5\} \).

While we are on the topic you should also know that \( \cap \) stands for Intersection and can be thought of as meaning AND. So for example \( A \cap B \) is the set of all elements that are in \( A \) and \( B \) e.g. if \( A = \{1, 2, 3\} \) and \( B = \{3, 4, 5\} \) then \( A \cap B = \{3\} \).

Venn Diagrams

Venn diagrams are often a useful method to help understand probabilities. They provide a pictorial method of visualizing probabilities as areas. Figures 1.2 through 1.6 provide several examples of possible Venn diagrams.

1.3 Some Useful Formulas for Computing Probabilities

1.3.1 Unions and Intersections

In order to illustrate some of the real world applications of probabilities we will consider the following case.
1.3. SOME USEFUL FORMULAS FOR COMPUTING PROBABILITIES

Figure 1.2: Here we have two sets, $A$ and $B$. The shaded region represents the intersection, $A \cap B$.

Figure 1.3: Here we have two mutually exclusive sets, $A$ and $B$.

Figure 1.4: Here the shaded region represents the union of $A$ and $B$, $A \cup B$.

Figure 1.5: Here the Venn Diagram shows that $A$ is a subset of $B$.

Figure 1.6: Here the Venn Diagram illustrates $A$ and its complement, $\overline{A}$.
Case 1 Outtel vs Microhard: Calculating the cost of a warranty

Outtel has been making computer chips for many years. They make almost all the computer chips for personal computers and have no serious competition. As a result they have never offered any kind of warranty on their chips.

Now Microhard has decided to enter the computer chip market and this will provide serious competition for Outtel. Thus Outtel has decided to offer a warranty on their chips in order to maintain their market share. However management is concerned about the cost of offering such a warranty. You have just been taken on as a summer intern and have been given the task of estimating the cost per chip. (Unfortunately you were bragging about all that you had learnt in BUAD 309.) A large part of calculating this cost is to estimate the probability that a chip will fail in the warranty period.

Unfortunately because Outtel has never cared about this before there are no records on the probability that an individual chip will fail. However your technicians have determined that there are two possible defects (call them A and B) that a chip can have, either of which will cause the chip to fail eventually (though not immediately). These defects are very hard to observe directly. There are tests for defect A, defect B and also a test for both defects. However these tests destroy the chips so only one test can be done on any given chip. Luckily you have been given a large budget to answer this question so you can do as many tests as you want.

How would you calculate the probability that a randomly chosen chip will fail? Fortunately in your BUAD 309 class you immediately realized that probability was a very important topic and paid careful attention so this question is easy to answer.

To solve this case we need to calculate $P(A \cup B)$ (Why?). From the previous section we know that if $A$ and $B$ are mutually exclusive then, $P(A \cup B) = P(A) + P(B)$, but what do we do if they are not?

It turns out that for any events, $A$ and $B$,  

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1.1)$$

This formula can be motivated by examining a Venn Diagram (see for example Figure 1.4) or using the axioms from the previous section and a formal mathematical proof.

Example one
Suppose we toss a single die and are interested in the events $A = \{\text{die is even}\}$ and $B = \{\text{number
1.3. SOME USEFUL FORMULAS FOR COMPUTING PROBABILITIES

showing is less than or equal to 4}. Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \frac{3}{6} + \frac{4}{6} - \frac{2}{6}$$
$$= \frac{5}{6}$$

Case 1 Revisited
Suppose that we define $A = \{\text{chip has defect } A\}$ and $B = \{\text{chip has defect } B\}$. If we perform tests which show $P(A) = 0.02, P(B) = 0.03, P(A \cap B) = 0.01$ then

$$P(\text{failure}) = P(A \cup B)$$
$$= P(A) + P(B) - P(A \cap B)$$
$$= 0.02 + 0.03 - 0.01$$
$$= 0.04$$

Also note that the rule for mutually exclusive events extends to more than two events e.g. if $A, B$ and $C$ are all mutually exclusive then

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

1.3.2 Complements

$A$ means, complement of $A$. In other words “everything not in $A$”. How would you calculate $P(A)$? In fact we can use the axioms from Section 1.2. By the definition of $A$ we know that

$$S = A \cup A \quad \text{and} \quad A \cap A = \emptyset$$

so $A$ and $A$ are mutually exclusive. Therefore

$$1 = P(S) \quad \text{By axiom 2}$$
$$= P(A \cup A)$$
$$= P(A) + P(A) \quad \text{By axiom 3 and mutual independence}$$

But if $1 = P(A) + P(A)$ then

$$P(A) = 1 - P(A) \quad (1.2)$$

Why do we care? Suppose for example that Outtels competitors chips have

- 0 defects with probability 0.80
- 1 defect with probability 0.05
- 2 defects with probability 0.05
- 3 defects with probability 0.05
- 4 defects with probability 0.05
and we are interested in the event \( A = \{ \text{at least one defect} \} \). Then \( P(A) = 1 - P(A) \) where \( A = \{ \text{zero defects} \} \). So 
\[
P(A) = 1 - P(\text{zero defects}) = 1 - 0.8 = 0.2
\]
The alternative is to sum up all four other probabilities i.e.
\[
P(A) = P(1) + P(2) + P(3) + P(4) = 0.2
\]

1.4 Conditional Probabilities and Independent Events

In this section we introduce the related ideas of conditional probabilities and independent events.

1.4.1 Conditional Probabilities

Often when trying to calculate the probability of a certain event occurring we need to incorporate additional information. For example suppose that we are interested in the probability of the roulette ball landing on an odd number but we are given the additional piece of information that it has landed on a red number. How does this affect the probability? Case 2 illustrates a more realistic application.

Case 2 Outtel vs Microhard: Incorporating new information into the decision process

As a result of your excellent work on the warranty calculation Outtel has taken you on as a managerial cadet. They have implemented the warranty policy and it is working well. However management is not happy with the failure rate on their chips. Outtel’s technicians have been working on ways to detect faults in the chips before they are sold.

As a result of this work a method has been found to detect chips with defect B without destroying the chip. However there is still no known way to detect defect A without destroying the chip. Nevertheless this is good news because defect B accounts for a significant proportion of the defective chips and now only chips with defect A will be sold.

Obviously this will result in a lower probability that a randomly chosen chip will be defective. Since you did such a fine job on the original probability calculation you have been asked to calculate the new probability that a chip that is sold will be defective assuming that no chips with defect B are sold.

(Recall \( P(A) = 0.02, P(B) = 0.03, P(A \cap B) = 0.01 \))

Before attempting to solve this problem we will examine a simple example. Suppose we are rolling a single die and are interested in the events
\[
A = \{ \text{number is less than or equal to 3} \} \\
B = \{ \text{the number is odd} \}
\]
and we want to know the probability that $B$ happens given (or conditional on) the fact we know $A$ has happened. In other words if somebody tells us that the die has landed with a number less than or equal to 3 showing what is the probability that the number is odd? This is called a **Conditional Probability** and is denoted as

$$P(B|A) \quad \text{(read as “Probability of } B \text{ given } A\text{”)}$$

Before considering the conditional probability what are the unconditional probabilities i.e. $P(A)$ and $P(B)$? Using the classical approach it should be clear that $P(A) = 3/6 = 1/2$ and $P(B) = 3/6 = 1/2$ (Why?) But what is $P(B|A)$?

Clearly if we know $A$ has happened then the sample space is reduced from the numbers 1, 2, 3, 4, 5, 6 to only 1, 2, 3. In other words there are now only 3 possible outcomes rather than 6. Of these three numbers two are odd. So 2 out of the 3 possible outcomes will cause $B$ to happen. Recall that using the classical approach :

$$\text{Probability} = \frac{\text{Number of outcomes that cause the event}}{\text{Number of possible outcomes}}$$

So

$$P(B|A) = \frac{2}{3}$$

This process for calculating a conditional probability is a little cumbersome but notice that

$$P(B|A) = \frac{2}{3} = \frac{2/6}{3/6} = \frac{P(A \cap B)}{P(A)}$$

This formula will hold for any conditional probability.

**Definition 8** The conditional probability of event $B$ given event $A$ is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{provided } P(A) > 0$$

Notice that this provides a method for calculating the probability of the intersection of two events i.e.

$$P(A \cap B) = P(A)P(B|A) \quad \text{or} \quad P(A \cap B) = P(B)P(A|B) \quad (1.3)$$

**Case 2 Revisited**

We are told that chips with defect $B$ will not be sold. Therefore the only way that a chip can be defective is if it has defect $A$ so we need to calculate

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

We know from Section 1.3.2 that $P(B) = 1 - P(B) = 1 - 0.03 = 0.97$. Also by referring to a Venn diagram it can be seen that
\[ P(A) = P(A \cap B) + P(A \cap \overline{B}) \quad (1.4) \]

which means

\[ P(A \cap B) = P(A) - P(A \cap \overline{B}) \]

Therefore \( P(A \cap B) = 0.02 - 0.01 = 0.01 \) and the probability of a defective chip is now

\[ P(A|B) = \frac{0.01}{0.97} = 0.0103 \]

1.4.2 Independent Events

It is often possible to take advantage of the concept of independence to simplify a probability calculation. Case 3 provides an illustration.

**Case 3 Outtel vs Microhard: Simple analysis of uncertainty; Taking advantage of independence**

Outtel now recognizes that you are a potential CEO candidate after you have solved these two impossible seeming problems and has promoted you to the level of junior manager. Outtel is looking into supplying their chips already built into the computer motherboards. Since Outtel doesn’t produce the motherboards they are considering bids from outside contractors.

The bidding has come down to the final two companies; Perfectomonono and Imperfectomonono. Both companies are willing to offer essentially the same product for the same price so it comes down to quality. Quality is being measured by probability of failure. Outtel’s technicians have calculated that the probability of a board from Perfectomonono failing is only 1% but the probability of a board from Imperfectomonono failing is 2%. On this basis Outtel is about to sign a contract with Perfectomonono.

However at the last minute while reviewing the competing bids you discover that the motherboards from Perfectomonono require the installation of a slightly different chip from Outtel’s standard. The new chip costs the same to produce as the old chip so no one else has worried about this. However the new chip has a probability of failure of 3% while the old chip only had a probability of 2% of failing. Clearly the product will fail if either the chip or the board fails so this information is important. You can’t work out why no one has thought about this before (though you are rapidly working out why you are progressing through the company so fast) but you need to work out which is really the better bid and fast! The only things you have on your side are your wonderful education and a certainty that the failure of the motherboard and the failure of the chip must be “independent” because they are produced by different companies.

We will examine a simple example of independence before attempting to solve this case.
Example 1
Suppose you deal a card face up from a deck of 52 cards and are interested in the events:

\[ A = \{ \text{Card is a spade} \} \]
\[ B = \{ \text{Picture card i.e. Jack, Queen or King} \} \]

Clearly \( P(A) = \frac{13}{52} = \frac{1}{4} \) and \( P(B) = \frac{12}{52} = \frac{3}{13} \) but what is the probability of \( B \) if we are told that the card is a spade i.e. what is \( P(B|A) \)? We know \( P(A \cap B) = \frac{3}{52} \) (Why?) so

\[ P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{3/52}{13/52} = \frac{3}{13} \]

but \( P(B) = \frac{3}{13} \) so \( P(B|A) = P(B) \). Actually this makes sense because knowing that \( A \) has happened provides no “useful information”. In other words the fact that the card is a spade should not change the probability that it is a picture card because all four suits have the same number of picture cards. If \( P(B|A) = P(B) \) then we say that \( A \) and \( B \) are independent events because knowing \( A \) does not change the probability of \( B \).

**Definition 9** \( A \) and \( B \) are independent if and only if (iff)

\[ P(B|A) = P(B) \quad \text{or equivalently} \quad P(A|B) = P(A) \]

Notice that

\[ P(B|A) = P(B) \]

\[ \Rightarrow \frac{P(A \cap B)}{P(A)} = P(B) \]

\[ \Rightarrow P(A \cap B) = P(A)P(B) \]

so an equivalent definition is

**Definition 10** \( A \) and \( B \) are independent if and only if (iff)

\[ P(A \cap B) = P(A)P(B) \]

**Case 3 Revisited**
Let \( A \) be the event that the board fails and \( B \) be the event that the chip fails. We are interested in the probability that either fails since this will cause the entire unit to fail i.e. we want to calculate \( P(A \cup B) \).

Recall from Section 1.3.1 that

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

and we know \( P(A) \) and \( P(B) \) so this only leaves \( P(A \cap B) \) to calculate. In general this may be difficult since we are not given the probability. However in this situation it is clearly reasonable to assume that the chip and board fail independently of each other (they are produced in different locations!) so we can use Definition 10. For Perfectonomo

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.01 + 0.03 - 0.01 \times 0.03 = 0.0397 \]

while for Imperfectonomo

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.02 + 0.02 - 0.02 \times 0.02 = 0.0396 \]

We see that Imperfectonomo actually has a slightly lower failure rate overall!
1.5 Probability Tree Diagrams

Probability trees are a useful method to help calculate probabilities where more than one event is involved. Figure 1.7 shows the basic structure. Probability trees are most useful when a second event (in this case $B$) depends on the outcome of the first event ($A$). We will illustrate an application of these trees through the following example.

**Example 1**
Suppose we draw two cards (without replacement) from a deck of 52 cards and we are interested in

$A = \{\text{First card is a king}\}$

$B = \{\text{Second card is a king}\}$

Then we can draw the probability tree shown in Figure 1.8. How would we use the tree to calculate $P(B)$? We already know from (1.4) that $P(B) = P(A \cap B) + P(A \cap B)$ and both these probabilities can be read off the tree i.e.,

$$P(B) = \frac{4}{52} \times \frac{3}{51} + \frac{48}{52} \times \frac{4}{51} = \frac{4}{52}.$$
1.6. BAYES’ THEOREM

Figure 1.8: The probability tree from Example 1.

Notice that we can calculate this probability directly using the following formula

\[
P(B) = P(A \cap B) + P(A \cap \bar{B})\quad \text{from (1.4)}
\]
\[
= P(A)P(B|A) + P(A)P(\bar{B}|A)\quad \text{from (1.3)}
\]
\[
= \frac{4}{52} \times \frac{3}{51} + \frac{48}{52} \times \frac{4}{51}
\]
\[
= \frac{4}{52}
\]

The formula

\[
P(B) = P(A)P(B|A) + P(A)P(\bar{B}|A) \quad (1.5)
\]

has several important uses as we will see in the next section.

1.6 Bayes’ Theorem

We saw in Section 1.4.1 that

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}
\]
However it is often the case that rather than knowing \( P(A \cap B) \) and \( P(B) \) we instead know \( P(B \mid A) \) and wish to calculate \( P(A \mid B) \). Case 4 provides a quality control application.

**Case 4 Outtel vs Microhard: Small mistakes can be huge; A quality control application**

After saving Outtel many millions of dollars with the motherboard contract you have now become the youngest senior manager in their history at the age of 24. Your new position is senior manager in charge of production quality. You have continued to work hard to improve the quality of Outtel’s chips and have now reduced the probability of any randomly chosen chip being defective to only 0.1%.

One of the technicians working under you (they are from the “other school” so it is natural for them to be working for you) has come up with an ingenious test for defective chips. This test will always identify a defective chip as defective and only “falsely” identifies a good chip as defective with probability 1%.

This sounds like an ideal test but you remember a case similar to this one which you studied in BUAD 309 and warning bells sound. You are concerned that this test sometimes rejects good chips and want to calculate the probability that a chip is indeed defective given that the test says it is. Your fellow managers ridicule you because “it is obvious that almost all the chips that the test claims are defective will be since it only makes a mistake 1% of the time”. Luckily you have received a better education than them so you do the calculation anyway.

If we let

\[ + = \{ \text{Test indicates a defect} \} \]
\[ D = \{ \text{Chip is defective} \} \]

then we are really interested in \( P(D \mid +) \) but we only know \( P(+ \mid D) \) and \( P(+ \mid \overline{D}) \). Bayes’ theorem provides an answer. It states

\[
P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(\overline{A})P(B \mid A)} \tag{1.6}
\]

Bayes’ theorem has a fairly simple derivation.

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)} \]
\[= \frac{P(A)P(B \mid A)}{P(B)} \text{ from (1.3)} \]
\[= \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(\overline{A})P(B \mid A)} \text{ from (1.5)}
\]
Despite the simplicity of its derivation, Bayes’ theorem has many important applications. Example 1 provides a health application.

Example 1
Cancer of the breast is the most frequent cancer and the leading cause of cancer death in women. It is recommended that women over the age of 50 be screened annually by X-ray mammography. However this method of screening is far from perfect. It is estimated (Moskowitz, 1983) that the annual incidence rate of breast cancer is only about 2 in 1000 women. So if a woman is randomly screened each year she has only a 0.002 probability of having breast cancer.

\[ P(\text{cancer}) = P(C) = 0.002 \]

It is possible for the test to be positive (i.e. indicate cancer) when in fact there is none. It has been estimated that

\[ P(\text{Positive test} | \text{No cancer}) = P(+|C) = 0.04 \]

It is also possible for the test to be negative when in fact cancer is present. It has been estimated that

\[ P(\text{Negative test} | \text{Cancer}) = P(-|C) = 0.36 \]

What does this suggest is the probability that a woman has breast cancer given that she tests positive i.e. \( P(C|+) \). Using Bayes’ theorem we see

\[
P(C|+) = \frac{P(C)P(+|C)}{P(C)P(+|C) + P(C)P(+|C) + P(-|C)P(-|C)} = \frac{0.002 \times 0.64}{0.002 \times 0.64 + (1 - 0.002) \times 0.04} = 0.031
\]

So a woman who tests positive has only a 3.1% probability of actually having breast cancer!

Case 4 Revisited
It is clear that \( P(D) = 0.001, P(+|D) = 1 \) and \( P(+|D) = 0.01 \). Therefore

\[
P(D|+) = \frac{P(D)P(+|D)}{P(D)P(+|D) + P(D)P(+|D) + P(-|D)P(-|D)} = \frac{0.001 \times 1}{0.001 \times 1 + (1 - 0.001) \times 0.01} = 0.091
\]

So 9 out of every 10 chips that are rejected by the test are in fact not defective!

1.7 Counting Principles

Probability and “combinatorics” are closely related ideas. In this section we will examine “permutations” and “combinations”. This will be particularly useful in the next chapter.

1.7.1 Permutations

We will illustrate this concept through an example.

Example one
Suppose that the letters A,D,E,P,S are in a hat and that we draw them out (without replacement) one at a time until they are all gone. What is the probability that the letters are drawn out in the correct order to spell SPADE? Using the classical approach we see

\[
P(\text{SPADE}) = \frac{1}{\text{The number of possible ways of drawing the five letters}}
\]
because there is only one arrangement that will spell SPADE. We call each such arrangement a permutation.

**Definition 11** A permutation of a set of objects is an arrangement of these objects in a definite order.

The question then becomes: How many possible permutations are there? Suppose we were only drawing out one letter from the hat. Clearly there would be 5 possible orderings (i.e. S,P,A,D or E). Now suppose we draw out two letters there will then be 5 possible orderings for the first letter but only 4 more for the second (e.g. if we draw an S out first then it is only possible to draw a P,A,D or E on the second). Therefore there are $5 \times 4 = 20$ possible permutations. If we now draw 5 letters there are 5 orderings for the first, 4 for the second etc. down to only 1 for the last. This gives

$$5 \times 4 \times 3 \times 2 \times 1 = 120$$

permutations. So

$$P(\text{SPADE}) = \frac{1}{120}$$

Note that $5 \times 4 \times 3 \times 2 \times 1 = 5!$ we define $n!$ as

$$n! = n \times (n - 1) \times (n - 2) \times \cdots \times 3 \times 2 \times 1 \quad (0! = 1) \quad (1.7)$$

**Example two**

Suppose we only draw out 3 letters from the hat. What is the probability that we spell SPA. Again

$$P(\text{SPA}) = \frac{1}{\text{Number of possible permutations}}$$

Using the same logic as for the previous example we see that the number of permutations is $5 \times 4 \times 3$ but notice

$$5 \times 4 \times 3 = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = \frac{5!}{2!} = 60$$

This suggests a general rule. Namely, the number of permutations of $x$ out of $n$ objects ($x \leq n$) is

$$n \times (n - 1) \times (n - 2) \times \cdots \times (n - x + 1) = \frac{n \times (n - 1) \times (n - 2) \times \cdots \times 3 \times 2 \times 1}{(n - x) \times (n - x - 1) \times \cdots \times 3 \times 2 \times 1} = \frac{n!}{(n - x)!}$$

We write the number of permutations of $x$ out of $n$ objects as $n^P_x$ so

$$n^P_x = \frac{n!}{(n - x)!} \quad (1.8)$$
1.7. Counting Principles

1.7.2 Combinations

We now know how to calculate the number of possible orderings when we care about order. However it is often the case that we only care about the objects that we draw and not the order they are drawn in.

**Definition 12** A combination of a set of objects is a subset of the objects disregarding their order.

**Example one**

We have five potential committee members, Alison, Bert, Catherine, David and Erin and we wish to choose 3 for the committee. How many different committees or combinations are possible?

One possible answer is \( _5 P_3 = \frac{5!}{2!} = 60 \). But this counts \{Alison, Bert, Catherine\} and \{Bert, Alison, Catherine\} as different committees. Clearly we don’t care about the order members are chosen.

This method is over counting.

Let \( k = \) number of distinct committees of 3 people. Then

\[
_5 P_3 = k \times \text{(number of orderings of 3 people)}
\]

\[
\Rightarrow \frac{5!}{2!} = k \times 3!
\]

\[
\Rightarrow k = \frac{5!}{3!2!} = 10
\]

In general if we have \( n \) objects and choose \( x \) without regard to order then

\[
_n P_x = k \times \text{(number of orderings of } x \text{ objects)}
\]

\[
\Rightarrow \frac{n!}{(n-x)!} = k \times x!
\]

\[
\Rightarrow k = \frac{n!}{x!(n-x)!}
\]

Instead of using \( k \) we generally write the number of possible combinations as \( \binom{n}{x} \) (read as \( n \) choose \( x \)) where

\[
\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad (1.9)
\]

**Example two**

What is the probability of winning the grand prize in the California State Lottery? There are 51 balls numbered 1 - 51 and you must guess the 6 numbers that are drawn (but not the order that they are drawn in). Therefore the probability is

\[
\frac{1}{\text{Number of combinations}} = \frac{1}{\binom{51}{6}} = \frac{1}{18,009,460}
\]

(If you did care about the ordering the probability would be 1/12,966,811,200!! It is often said that lotteries are a tax on people that don’t understand probabilities!)
Example three
Suppose you toss a coin 5 times. How many ways are there of getting 3 heads? This is exactly the same as asking: How many ways are there of picking 3 balls out of 5? i.e.

\[
H, H, H, T, T = \{1, 2, 3\} \\
H, H, T, H, T = \{1, 2, 4\} \quad \text{etc.}
\]

Note that the order is not important. For example \{1, 2, 3\}, \{2, 3, 1\} and \{2, 1, 3\} all result in \(H, H, H, T, T\). Therefore the number of possible orderings is

\[
\binom{5}{3} = \frac{5!}{3!2!} = 10
\]

In general if we toss a coin \(n\) times the number of ways of getting \(x\) heads is

\[
\binom{n}{x}
\]
Chapter 2

Risk Management: Profiting From Uncertainty

2.1 Random variables

In this section we will introduce the idea of a “random variable”. First we will provide a formal definition.

**Definition 13** A random variable (r.v.) is a variable whose numerical value is determined by the outcome of a random trial. Random variables are often denoted by a capital letter from the end of the alphabet i.e. $X, Y, Z$

Note that not all variables are random.

**Example one**

Let $l =$ the length of the side of a square. Until I give you the square $l$ is undetermined but it is not random. Once I give you the square $l$ will always be the same number.

On the other hand if we let $X =$ the sum of a pair of dice this is random. Even if you are given the dice the number is not fixed. One time $X$ could be 7 and the next 3 etc. Note $X$ is random before the experiment is performed. Obviously once the dice are thrown $X$ is fixed for that experiment.

There are two types of random variables:

- discrete
- and continuous.

**Definition 14** A discrete random variable is able to take on a countable number of values.

For example

- the sum of two dice,
- the number of heads in 5 tosses of a coin,
- or the number of defective computer chips on a production line.

**Definition 15** A continuous random variable is assumed to take any value in an interval.
For example

- weight,
- height,
- and temperature etc.

In this chapter we will concentrate on discrete random variables and cover continuous random variables in the next.

### 2.2 General probability distributions for discrete random variables

Suppose we let \( X \) = the sum of 2 dice. Then \( X \) is a discrete random variable. What do we know about \( X \)? To start with we know that \( X \) must be an integer between 2 and 12. But \( X \) could be any one of these numbers. Can we say anything else? Yes. For example even though \( X \) could be any one of these numbers it is far more likely to be 7 than 2. We can make this statement more precise using the probability framework that we developed in the previous chapter i.e.

\[
P(X = 2) = P(\text{first die is a one and second die is a one})
\]

\[
= P(\{1,1\})
\]

\[
= \frac{1}{36}
\]

But

\[
P(X = 7) = P(\{1,6\} \cup \{2,5\} \cup \{3,4\} \cup \{4,3\} \cup \{5,2\} \cup \{6,1\})
\]

\[
= P(\{1,6\}) + P(\{2,5\}) + P(\{3,4\}) + P(\{4,3\}) + P(\{5,2\}) + P(\{6,1\})
\]

\[
= \frac{1}{36} + \cdots + \frac{1}{36}
\]

\[
= \frac{6}{36} = \frac{1}{6}
\]

Using the same method we could calculate \( P(X = x) \) for \( x = 2,3,\ldots,11,12 \). This would give us the “probability mass function” of \( X \).

**Definition 16** The set of all possible probabilities is called the probability mass function of \( X \) (pmf). This is also referred to as the distribution of \( X \). It tells you everything there is to know about \( X \).

Often we write the pmf (or distribution) as a table.

For example the distribution of \( X \), where \( X \) is the sum of the two dice, is

<table>
<thead>
<tr>
<th>( x )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X = x) )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{2}{36} )</td>
<td>( \frac{3}{36} )</td>
<td>( \frac{4}{36} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{4}{36} )</td>
<td>( \frac{3}{36} )</td>
<td>( \frac{2}{36} )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{2}{36} )</td>
<td>( \frac{1}{36} )</td>
</tr>
</tbody>
</table>

All probability mass functions (pmfs) must

1. be non-negative i.e. \( P(X = x) \geq 0 \) for all \( x \)
2. and sum to one i.e. \( \sum_x P(X = x) = 1 \)
2.2.1 Calculating probabilities over regions

Let \( X \) be the sum of two dice as above. Then recall

\[
\begin{array}{c|cccccccccccc}
X & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
\hline
P(X = x) & \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \\
\end{array}
\]

Suppose we want to calculate the probability of \( X \) falling in a region e.g., \( P(3 \leq X \leq 5) \). How would we do this?

\[
P(3 \leq X \leq 5) = P(\{X = 3\} \cup \{X = 4\} \cup \{X = 5\})
= P(\{X = 3\}) + P(\{X = 4\}) + P(\{X = 5\})
= \frac{2}{36} + \frac{3}{36} + \frac{4}{36}
= \frac{9}{36}
\]

It should be clear that this method can be extended to any sized interval. In general

\[
P(a \leq X \leq b) = \sum_{x=a}^{b} P(X = x) \quad (2.1)
\]

2.3 Binomial Random Variables

There are certain types of random variables that occur so frequently they are given a name to identify them. One such example is a “Binomial random variable”. Case 5 provides an example.

Case 5 Outtel vs Microhard: Winning contracts and making friends

Outtel has given up worrying about your extreme youth and has now promoted you to the position of vice president at the age of 25. You of course know that all your success is thanks to the Marshall School of Business and in particular BUAD 309 so you are about to write a fat check to the Alumni Association! However the people at Outtel are unaware of this fact and believe you must just be smart!

One of your new responsibilities as VP is to negotiate contracts with large organizations. There is one particularly large contract that you are working on which will be very lucrative if you can get it. Unfortunately Microhard is also in the bidding so it has got quite competitive. Never the less you have used all your many charms to get the inside running. You have been offered the contract with one condition.

The contract is for bulk shipments of the new “Bentium 500” chips. Each shipment is made in boxes of 50 chips. The key condition of the contract is that no more than 1 out of every 200 boxes may contain more than 2 defective chips. There is a very stiff penalty for not meeting this requirement so it is very important that you do an accurate calculation. The failure rate on the new
How do we know that this is an example of a Binomial? There are 3 conditions that an experiment must fulfill before a Binomial will result. These being:

1. There are a fixed number of trials (denoted by \( n \)) with only 2 possible outcomes: “success” or “failure”.
2. The probability of success on each trial, \( p \), remains constant. The probability of failure is \( 1 - p \).
3. The trials are independent of each other.

**Definition 17** Let \( X = \) the number of successes in the \( n \) trials. Then if the above 3 conditions hold \( X \) will be a binomial random variable.

We often write

\[
X \sim Bin(n, p)
\]

where the \( \sim \) means “has the distribution of” or “is distributed like” and \( Bin(n, p) \) means a Binomial with \( n \) trials and probability \( p \) of success on each one. To get a feel for handling these random variables we will try some easy examples.

**Example one**
Suppose a coin is flipped 3 times. Let \( X = \) the number of heads. Then \( X \) is a binomial random variable i.e. \( X \sim Bin(n = 3, p = 0.5) \). This is because there are a fixed number of trials (we toss the coin 3 times), there are two possible outcomes (head or tail), the probability of success (head) is 0.5 for every trial (toss), and the tosses of the coin are all independent.

**Example two**
Digital Industries uses a manufacturing process to produce memory chips for computers. The probability that a chip is defective (prior to quality control) is 1/6. The problems that cause the defects occur at random. The defects also occur independently from one chip to another. Digital industries has selected three chips at random from its production process and is interested in the number that are defective. If we let \( Y = \) the number of defective chips, then

\[
Y \sim Bin(n = 3, p = 1/6)
\]

We know that the possible values for \( Y \) are 0, 1, 2, 3 but what are the various probabilities i.e. \( P(Y = 0) = P(\text{no defects}) \) etc. To calculate these probabilities we can make use of a probability tree such as in Figure 2.1. Here \( S \) denotes a success (i.e. chip is defective) and \( F \) denotes a failure (i.e. chip is not defective).

From the tree we can see, for example,

\[
P(Y = 3) = P(\{S, S, S\}) = \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0 = 0.005
\]
2.3. **BINOMIAL RANDOM VARIABLES**

<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Sequence</th>
<th>Number</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>S</td>
<td>SSS</td>
<td>3</td>
<td>((\frac{1}{6}))^3 ((\frac{5}{6}))^0</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>SF</td>
<td>SSF</td>
<td>2</td>
<td>((\frac{1}{6}))^2 ((\frac{5}{6}))^1</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>SFS</td>
<td>SFS</td>
<td>2</td>
<td>((\frac{1}{6}))^2 ((\frac{5}{6}))^1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>SFF</td>
<td>1</td>
<td>((\frac{1}{6}))^1 ((\frac{5}{6}))^2</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>SF</td>
<td>FSS</td>
<td>2</td>
<td>((\frac{1}{6}))^2 ((\frac{5}{6}))^1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>SFF</td>
<td>1</td>
<td>((\frac{1}{6}))^1 ((\frac{5}{6}))^2</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>F</td>
<td>FFS</td>
<td>1</td>
<td>((\frac{1}{6}))^1 ((\frac{5}{6}))^2</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>FFF</td>
<td>0</td>
<td>((\frac{1}{6}))^0 ((\frac{5}{6}))^3</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.1: A Probability Tree

\[
P(Y = 2) = P(\{S, S, F\} \cup \{S, F, S\} \cup \{F, S, S\}) = (\frac{1}{6})^2 \left(\frac{5}{6}\right)^1 + (\frac{1}{6})^2 \left(\frac{5}{6}\right)^1 + (\frac{1}{6})^2 \left(\frac{5}{6}\right)^1 = 0.069
\]

\[
P(Y = 1) = P(\{S, F, F\} \cup \{F, S, F\} \cup \{F, F, S\}) = (\frac{1}{6})^1 \left(\frac{5}{6}\right)^2 + (\frac{1}{6})^1 \left(\frac{5}{6}\right)^2 + (\frac{1}{6})^1 \left(\frac{5}{6}\right)^2 = 0.347
\]

and

\[
P(Y = 0) = P(\{F, F, F\}) = (\frac{1}{6})^0 \left(\frac{5}{6}\right)^3 = 0.579
\]

Note that 0.069 + 0.069 + 0.347 + 0.579 = 1 as it must!

This method works OK for a small number of trials such as this case which has 3. However what if there are 10 trials i.e. 10 different chips are selected? We can still calculate

\[
P(Y = 0) = P(\{F, F, F, F, F, F, F, F, F, F\}) = (\frac{1}{6})^{10} \left(\frac{5}{6}\right)^0
\]

but what about \(P(Y = 5)\) for example. For \(Y\) to equal 5 we need to have 5 successes and 5 failures so it should be clear that

\[
P(Y = 5) = ? \left(\frac{1}{6}\right)^5 \left(\frac{5}{6}\right)^5
\]
where \( r \) is the number of different possible ways of getting 5 successes in 10 trials. At first glance it seems like this may be a difficult number to calculate but in fact this is exactly the same as asking for the number of ways of getting 5 heads out of 10 tosses. We solved this problem in Section 1.7, where we showed that in fact this is simply

\[
\binom{10}{5} = \frac{10!}{5!5!} = 252
\]

so

\[
P(Y = 5) = \binom{10}{5} \left( \frac{1}{6} \right)^5 \left( \frac{5}{6} \right)^5
\]

We can apply this reasoning to calculate the probability for any value of \( y \) i.e.

\[
P(Y = y) = \binom{10}{y} \left( \frac{1}{6} \right)^y \left( \frac{5}{6} \right)^{10-y} \quad y = 0, 1, \ldots, 10
\]

In general if \( X \sim Bin(n, p) \) where \( n \) is the number of trials and \( p \) is the probability of success on an individual trial then

\[
P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \ldots, n \quad (2.2)
\]

This formula gives the probability mass function (or distribution) of a binomial random variable.

**Case 5 Revisited**

Let \( X \) be the number of defective chips. If we check the conditions for a binomial we see that \( X \sim Bin(50, 0.005) \). We want to calculate \( P(X > 2) \).

\[
P(X > 2) = 1 - P(X \leq 2) \quad \text{using the complement rule}
\]

\[
= 1 - P(\{X = 0\} \cup \{X = 1\} \cup \{X = 2\})
\]

\[
= 1 - (P(X = 0) + P(X = 1) + P(X = 2))
\]

\[
= 1 - \left( \binom{50}{0} 0.005^0 \times 0.995^{50} + \binom{50}{1} 0.005^1 \times 0.995^{49} + \binom{50}{2} 0.005^2 \times 0.995^{48} \right)
\]

\[
= 1 - (0.7783 + 0.1956 + 0.024)
\]

\[
= 0.0021
\]

This is less than 1 in every 200 so we can accept the contract.

### 2.4 Conditional probabilities and independence of random variables

Just as we defined conditional probabilities and independence for events in the previous chapter we can also provide definitions for random variables. First we will examine the concept of a conditional probability for a random variable.
2.4. CONDITIONAL PROBABILITIES AND INDEPENDENT RANDOM VARIABLES

2.4.1 Conditional probabilities

Consider the following example. Suppose we toss two dice. Let \( X \) be die one, \( Y \) be die two and \( Z \) be the sum (i.e. \( Z = X + Y \)). Then we saw in Section 2.2 that

<table>
<thead>
<tr>
<th>( z )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(Z = z) )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{2}{36} )</td>
<td>( \frac{3}{36} )</td>
<td>( \frac{4}{36} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{4}{36} )</td>
<td>( \frac{3}{36} )</td>
<td>( \frac{2}{36} )</td>
<td>( \frac{1}{36} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

So we know all the probabilities associated with \( Z \). Suppose however that we are told that the first dice was a 3 (i.e. \( X = 3 \)). Now what is the probability that the sum is 5 (i.e. \( Z = 5 \))? Mathematically this is written as \( P(Z = 5 | X = 3) \) which should be read as “the probability that \( Z = 5 \) given that we know \( X = 3 \).” Clearly in this case, if the first dice is 3, the only way that the sum can be 5 is if the second dice is 2 \((3 + 2 = 5)\) so

\[
P(Z = 5 | X = 3) = P(Y = 2) = \frac{1}{6}
\]

However we can’t always rely on the problem being as simple as this. Recall from the previous chapter that for any events \( A \) and \( B \),

\[
P(A | B) = \frac{P(A \cap B)}{P(B)}
\]

If we let \( A = \{Z = 5\} \) and \( B = \{X = 3\} \) then

\[
P(Z = 5 | X = 3) = P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(Z = 5, X = 3)}{P(X = 3)}
\]

where the comma denotes “and” or “intersection”. (Note that \( P(Z = 5, X = 3) \) is called the “joint probability of \( Z \) and \( X \)). So

\[
P(Z = 5 | X = 3) = \frac{P(Z = 5, X = 3)}{P(X = 3)} = \frac{P(Y = 2, X = 3)}{P(X = 3)} = \frac{1/36}{1/6} = \frac{1}{6}
\]

which corresponds to the direct calculation. In general we define a conditional probability for a random variable in the following way.

**Definition 18** The conditional probability that \( Y = y \) given that \( X = x \) is

\[
P(Y = y | X = x) = \frac{P(Y = y, X = x)}{P(X = x)} \quad \text{provided that } P(X = x) > 0
\]

2.4.2 Independent random variables

Recall that \( A \) and \( B \) are said to be “independent events” if

\[
P(A \cap B) = P(A)P(B) \quad \text{or} \quad P(A | B) = P(A)
\]

What if we are dealing with random variables rather than events? In fact there is a very similar definition.
Definition 19  \( X \) and \( Y \) are said to be “Independent random variables” if
\[
P(X = x, Y = y) = P(X = x)P(Y = y) \quad \text{for all } x \text{ and } y
\]
or equivalently
\[
P(X = x \mid Y = y) = P(X = x) \quad \text{for all } x \text{ and } y
\]

This definition can be used in two ways. We can verify that two random variables are independent or we can use the fact that they are independent to calculate a “joint probability”.

Example one
Suppose we are told that \( X \sim Bin(10, 0.25) \), \( Y \sim Bin(20, 0.5) \) and that \( X \) and \( Y \) are independent. Then what is \( P(X = 3, Y = 5) \)?
\[
P(X = 3, Y = 5) = P(X = 3)P(Y = 5) \quad \text{from the definition of independence}
\]
\[
= \binom{10}{3} (0.25)^3 (1 - 0.25)^{10-3} \binom{20}{5} (0.5)^5 (1 - 0.5)^{20-5} \quad \text{binomial distribution}
\]
\[
= 0.2503 \times 0.01479
\]
\[
= 0.0037
\]

Example two
Suppose that we are given the joint distribution of \( X \) and \( Y \) in the form of the following table.

<table>
<thead>
<tr>
<th></th>
<th>( Y )</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1 2</td>
</tr>
<tr>
<td>( X )</td>
<td>0</td>
<td>0 1/4 0 1/4</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0 1/4 1/4 2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1/4 0 0 1/4</td>
</tr>
</tbody>
</table>

The table gives the joint probabilities (for example \( P(X = 0, Y = 0) = 0 \) while \( P(X = 0, Y = 1) = 1/4 \)). However by summing the rows and columns it also gives the individual probabilities for \( X \) and \( Y \) (for example \( P(Y = 0) = 1/4 \) and \( P(X = 1) = 1/2 \)).

We can then ask the question; Are \( X \) and \( Y \) independent? To show that they are independent we would need to verify that
\[
P(X = x, Y = y) = P(X = x)P(Y = y)
\]
for all the possible values of \( x \) and \( y \) (in this case 9 combinations). However to show that they are not independent we need only show
\[
P(X = x, Y = y) \neq P(X = x)P(Y = y)
\]
for one \( x \) and \( y \). In fact we can see from the table that
\[
P(X = 0, Y = 0) = 0, \quad P(X = 0) = 1/4, \quad P(Y = 0) = 1/4
\]
so
\[
0 = P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0) = 1/16
\]
and we see that $X$ and $Y$ are not independent. Note that

$$P(X = 1, Y = 1) = 1/4, \quad P(X = 1) = 1/2, \quad P(Y = 1) = 1/2$$

so

$$1/4 = P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = 1/4$$

It is possible for the equation to hold for some values of $x$ and $y$ even if the random variables are not independent.

## 2.5 Expected values and variances of random variables

Probabilities are not the only quantity of interest related to random variables. Often we are also interested in the expected value and variance of a random variable. Case 6 provides an example.

### Case 6 Keeping contracts and friends

It is now a year since your coup in winning the Bentium chip contract. After your spectacular success the president of Outtel stepped down so that the “better person” could run the business (don’t expect this to happen in the real world). You are now the president of Outtel!

Outtel has prospered and expanded under your benign dictatorship! The Bentium contract has played a large part in this success. However now the Bentium contract has come up for renewal and as usual there are problems. The buyer now wants boxes containing 100 chips. This is not a problem but they also insist that the average number of defective chips must be no more than 1 per box and the variability (using some standard measure) must be less than 0.75 per box.

You look at the demands and think to yourself that they must have “got in a beep beep statistician”. No one else at Outtel even has a good idea what we mean by average or variability. However this doesn’t throw you because (you guessed it) you know exactly what this means from BUAD 309.

*Recall the probability of an individual chip failing is 0.5%*

We will examine the idea of an expected value first.

### 2.5.1 Expected values

The “expected value” of a random variable, $X$, (written as $Ex$ or $\mu$) is the weighted average of the possible values of $X$, the weights being the probabilities of these values. Formally

**Definition 20** *The expected value of $X$ is defined as*

$$Ex = \mu = \sum_{i=1}^{k} x_i P(X = x_i)$$

$$= x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots + x_k P(X = x_k)$$
where \( x_1, \ldots, x_k \) are the possible values that \( X \) can take on.

Notice that if \( X \) is equally likely to take on any of \( k \) possible values (e.g. if \( X \) is the value of a die) then \( P(X = x_i) = \frac{1}{k} \) and

\[
EX = \frac{1}{k} \sum_{i=1}^{k} x_i = \text{“ordinary average of } x_i \text{”}
\]

One can also think of \( EX \) as the “long run average”. In other words if you performed the experiment many, many times, recorded the value of \( X \) each time and took the average it would equal \( EX \).

**Example one**

Suppose \( X \) is the value of the throw of a die. Then \( P(X = x) = \frac{1}{6} \) for \( x = 1, 2, 3, 4, 5, 6 \) and

\[
EX = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5
\]

So if we tossed the die many, many times the average of all the tosses would be 3.5.

Notice that the term expected value is somewhat misleading because \( X \) may never take on its expected value. In this case \( X \) will never equal 3.5! However it will “average out” to 3.5.

**Example two**

Suppose \( X \) equals the number of cars arriving at a tollbooth in a one minute period and that it has the following distribution:

| \( x \) | 1 \( 2 \) \( 3 \) |
|-------|---|---|
| \( P(X = x) \) | 0.25 | 0.40 | 0.35 |

Then

\[
EX = 1 \times 0.25 + 2 \times 0.40 + 3 \times 0.35 = 2.10
\]

Again we notice that it is not possible to have 2.10 cars arrive in a one minute period but if you recorded the number of cars arriving over several days and divided that by the number of minutes it would average out to about 2.10 cars per minute.

An alternative way of visualizing the expected value is as the balance point on a beam. If we put weights proportional to the probabilities at the possible values of \( X \) then \( EX \) would correspond to the balance point of the beam.

**Some useful results for expected values**

It is possible to calculate the expected value for combinations of random variables. Here we list some useful results.
1. If $a$ and $b$ are constants then
   \[ E(aX + b) = aEX + b \]

2. If $X$ and $Y$ are random variables then
   \[ E(X + Y) = EX + EY \]
   and
   \[ E(X - Y) = EX - EY \]
   also in general, for any constants $a$ and $b$
   \[ E(aX + bY) = aEX + bEY \]

3. Similarly if $X_1, X_2, \ldots, X_n$ are random variables then
   \[ E\left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} EX_i \]

4. If $X$ and $Y$ are independent then
   \[ E(XY) = (EX)(EY) \]

We will prove the first and last results.

**Proof of 1**
\[
E(aX + b) = \sum_{i=1}^{k} (ax_i + b)P(X = x_i) \\
= a \sum_{i=1}^{k} x_i P(X = x_i) + b \sum_{i=1}^{k} P(X = x_i) \\
= aEX + b
\]

**Proof of 4**
\[
E(XY) = \sum_{x} \sum_{y} xyP(X = x, Y = y) \\
= \sum_{x} \sum_{y} xyP(X = x)P(Y = y) \text{ by independence} \\
= \sum_{x} xP(X = x) \sum_{y} yP(Y = y) \\
= (EX)(EY)
\]

**Example one**
Suppose that in the previous example there are 3 different toll booths $(X_1, X_2, X_3)$ and we are
interested in the average “total number of cars to all 3 booths in a one minute period”. If \( Y = X_1 + X_2 + X_3 \) then we want \( EY \). If we wished to calculate this directly from the definition we would need to calculate all the various possible values for \( Y \) and the corresponding probabilities. However if we use the third “useful result” we see that

\[
EY = E(X_1 + X_2 + X_3) = EX_1 + EX_2 + EX_3 = 2.1 + 2.1 + 2.1 = 6.3
\]

**Example two**
A plant manufacturing calculators has fixed costs of \( $300,000 \) per year. The gross profit from each calculator sold (i.e. the price per unit less the variable cost per unit) is \( $5 \). The random variable \( X \) denotes the number of calculators that the plant sells per year and past information indicates that \( EX = 300,000 \). What is the expected value of the plant’s annual profit?

The plant’s annual profit will be equal to the number of calculators sold \( (X) \) multiplied by \( $5 \) (the profit for each one) less \( $300,000 \) (the fixed cost even if none are sold). If \( Y \) is the annual profit then

\[
Y = 5X - 300,000
\]

If we use the first useful result this tells us

\[
EY = E(5X - 300,000) = 5EX - 300,000 = 5 \times 300,000 - 300,000 = 1,200,000
\]

**Example three**
Suppose we toss two dice \((X, Y)\). What is the expected value of the sum, \( Z = X + Y? \)

Again we could calculate this directly using the formula for expectation and the table of probabilities from Section 2.4.1. However since we already know that \( EX = 3.5 \) and \( EY = 3.5 \) an easier way is to use the “second useful result” and note

\[
EZ = E(X + Y) = EX + EY = 3.5 + 3.5 = 7
\]

**Expected value for a Binomial random variable**

If we know that \( X \sim Bin(n, p) \) what is \( EX? \)

We know that \( X \) must take on one of the values \( 0, 1, 2, \ldots, n \) so from the definition of expected value

\[
EX = 0P(X = 0) + 1P(X = 1) + 2P(X = 2) + \cdots + nP(X = n)
\]

In Section 2.3 we showed that \( P(X = x) = \binom{n}{x} p^x(1-p)^{n-x} \) so

\[
EX = 0 \binom{n}{0} p^0(1-p)^n + 1 \binom{n}{1} p^1(1-p)^{n-1} + \cdots + n \binom{n}{n} p^n(1-p)^0 = \sum_{x=0}^{n} x \binom{n}{x} p^x(1-p)^{n-x}
\]

After some (messy) algebra we can show that this sum simplifies to \( np \) so:

\[
\text{If } X \sim Bin(n, p) \text{ then } \quad EX = np \quad (2.3)
\]
Case 6 Revisited
We are now ready to answer part of the question from Case 6. One of the requirements is that the average number of defective chips per box must not exceed 1. If \( X \) is the number of defects in a randomly chosen box we need to calculate \( EX \). In general this may be difficult since there are 101 different values \( X \) can take on (why?). However if we realize that the conditions for a binomial random variable are fulfilled so that \( X \sim Bin(n = 100, p = 0.005) \) we can use the formula that we just derived which tells us that

\[
EX = np = 100 \times 0.005 = 0.5
\]

Clearly \( EX < 1 \) so the first condition is satisfied.

2.5.2 Variance and Standard Deviation of a random variable

Notice that the expected value only tells us what the random variable is on average. Often \( X \) will be larger or smaller than \( EX \). Consider for example the temperatures in two US cities, one in Southern California and the other in the Mid West.

<table>
<thead>
<tr>
<th></th>
<th>Summer</th>
<th>Autumn</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern California</td>
<td>72F</td>
<td>65F</td>
<td>58F</td>
<td>65F</td>
</tr>
<tr>
<td>Mid West</td>
<td>95F</td>
<td>70F</td>
<td>25F</td>
<td>70F</td>
</tr>
</tbody>
</table>

Both cities have the same average temperature i.e.

\[
65 = \frac{1}{4}(72 + 65 + 58 + 65) = \frac{1}{4}(95 + 70 + 25 + 70) = 65
\]

but clearly this does not indicate that they have similar climates. The difference is that the Southern California city is always close to 65F but the Mid West city has temperatures that vary wildly over the year. This “variability” around the average value is also important. How might we give a numerical measure of variability?

A natural way to measure variation of a random variable is to consider the average difference of \( X \) from its mean \( \mu \) or \( EX \). In other words the average value of \( (X - EX) \).

\[
E(X - EX) = \sum_{i=1}^{k} (x_i - EX)P(X = x_i)
\]

Unfortunately this quantity is always equal to zero!

\[
E(X - EX) = EX - E(EX) = EX - EX = 0
\]

The positive and negative deviations cancel out. A more sensible measure is the average distance rather than the average difference i.e.

\[
E|X - \mu| = \sum_{i=1}^{k} |x_i - \mu|P(X = x_i)
\]

This is known as the mean deviation. However it is not commonly used because it has poor mathematical properties (absolute values are a pain!). On the other hand, squared differences do have nice mathematical properties so we use the following definition
Definition 21 The “Variance” of a random variable, \( X \), (written as \( \sigma^2 \) or \( \text{Var}(X) \)) is defined as

\[
\sigma^2 = \text{Var}(X) = E(X - \mu)^2 = E(X - EX)^2 = \sum_{i=1}^{k} (x_i - \mu)^2 P(X = x_i)
\]

Example one
Let \( X \) be the temperature of the Southern California city and \( Y \) be the temperature of the Mid West city. Then

\[
\text{Var}(X) = (72 - 65)^2 \frac{1}{4} + (65 - 65)^2 \frac{1}{4} + (58 - 65)^2 \frac{1}{4} + (65 - 65)^2 \frac{1}{4}
\]

\[= 24.5\]

while

\[
\text{Var}(Y) = (95 - 65)^2 \frac{1}{4} + (70 - 65)^2 \frac{1}{4} + (25 - 65)^2 \frac{1}{4} + (70 - 65)^2 \frac{1}{4}
\]

\[= 462.5\]

It is clear that the temperature in the Mid West has much greater variability.

Example two
Recall that the number of cars arriving at a toll booth, in a one minute period, are distributed as follows

<table>
<thead>
<tr>
<th>( x )</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X = x) )</td>
<td>0.25</td>
<td>0.40</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Then

\[
\text{Var}(X) = (1 - 2.1)^2 P(X = 1) + (2 - 2.1)^2 P(X = 2) + (3 - 2.1)^2 P(X = 3)
\]

\[= (-1.1)^2 \times 0.25 + (0.1)^2 \times 0.4 + (0.9)^2 \times 0.35\]

\[= 0.59\]

Notice that this formula requires using \( EX^2 \) (2.1) three separate times. There is an alternative formula for the variance which is often easier to use.

\[
\text{Var}(X) = E((X - EX)^2)
\]

\[= E(X^2 - 2XEX + (EX)^2)\]

\[= EX^2 + E(-2XEX) + E[(EX)^2]\]

\[= EX^2 - 2EXEX + (EX)^2\]

\[= EX^2 - 2(EX)^2 + (EX)^2\]

\[= EX^2 - (EX)^2\]

What do we mean by \( EX^2 \)? If \( g \) is a function of \( X \) then we define

\[
Eg(X) = \sum_{i=1}^{k} g(x_i) P(X = x_i) \tag{2.4}
\]
and in particular

\[ EX^2 = \sum_{i=1}^{k} x_i^2 P(X = x_i) \]  \hspace{1cm} (2.5)

So

\[ Var(X) = EX^2 - (EX)^2 \]  \hspace{1cm} (2.6)

where \( EX^2 = \sum_{i=1}^{k} x_i^2 P(X = x_i) \)

In this example

\[ EX^2 = 1^2 \times 0.25 + 2^2 \times 0.4 + 3^2 \times 0.35 = 5 \]

so

\[ Var(X) = EX^2 - (EX)^2 = 5 - 2.1^2 = 0.59 \]

which is the same as we calculated using the original formula.

Some useful results for variance of a random variable

1. For any random variable, \( X \)

\[ Var(X) \geq 0 \]

2. If \( a \) and \( b \) are constants then

\[ Var(aX + b) = a^2 Var(X) \]

This means that adding a constant \( b \) to a random variable leaves the variance unchanged and multiplying by a constant \( a \) causes the variance to be multiplied by \( a^2 \).

3. If \( X \) and \( Y \) are independent then

\[ Var(X + Y) = Var(X) + Var(Y) \]

and

\[ Var(X - Y) = Var(X) + Var(Y) \]

4. If \( X_1, X_2, \ldots, X_n \) are all independent random variables, then

\[ Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots Var(X_n) \]
We will prove the second result and leave the rest as exercises.

**Proof of second result**

\[
\text{Var}(aX + b) = E[(aX + b - E(aX + b))^2] \\
= E[(aX + b - aEX - b)^2] \\
= E[(a(X - EX))^2] \\
= a^2E(X - EX)^2 \\
= a^2\text{Var}(X)
\]

**Example one**

A plant manufacturing calculators has fixed costs of $300,000 per year. The gross profit from each calculator sold (i.e. the price per unit less the variable cost per unit) is $5. The random variable \(X\) denotes the number of calculators that the plant sells per year and past information indicates that \(EX = 300,000\) and \(\text{Var}(X) = 1,000,000\). What is the variance of the plant's annual profit? Recall that if \(Y\) is the annual profit then

\[
Y = 5X - 300,000
\]

so

\[
\text{Var}(Y) = \text{Var}(5X - 300,000) = 5^2\text{Var}(X) = 25,000,000
\]

**Standard Deviation of a random variable**

There is a problem with the variance. The variance of a random variable is measured in units squared. For example if \(X\) is the number of cars arriving at a toll booth then \(\text{Var}(X)\) is measured in cars\(^2\)! So the fact that the variance is 0.59 cars\(^2\) is hard to interpret. Therefore we often use the “standard deviation” as a measure of spread because it measures variability in the correct units.

**Definition 22** The standard deviation of a random variable, \(X\), (written as \(SD(X)\) or \(\sigma\)) is

\[
\text{SD}(X) = \sigma = \sqrt{\text{Var}(X)}
\]

In other words the standard deviation is just the square root of the variance.

**Example one**

The variance for the toll booth example was 0.59 so the standard deviation is \(\sqrt{0.59} = 0.768\). Note that this is 0.768 cars per minute so is in the correct units.

**Example two**

The variance for the calculator example was 25,000,000 so the standard deviation is 5,000.

**Useful results for standard deviation**

Note that all these results can be derived by taking square roots of the previous results for variance.
2.5. **EXPECTED VALUES AND VARIANCES OF RANDOM VARIABLES**

1. For any constants $a$ and $b$
   \[
   SD(aX + b) = \sqrt{Var(aX + b)} = \sqrt{a^2 Var(X)} = |a|SD(X)
   \]

2. If $X$ and $Y$ are independent random variables then
   \[
   SD(X + Y) = \sqrt{Var(X) + Var(Y)}
   \]
   and
   \[
   SD(X - Y) = \sqrt{Var(X) + Var(Y)}
   \]

3. If $X_1, X_2, \ldots, X_n$ are all independent random variables, then
   \[
   SD(X_1 + \cdots + X_n) = \sqrt{Var(X_1) + \cdots + Var(X_n)}
   \]

**Variance and Standard Deviation for a Binomial random variable**

Just as we did for the expected value we can produce a simple formula for the variance and standard deviation of a binomial random variable i.e.

\[
Var(X) = EX^2 - (EX)^2 = \sum_{x=0}^{n} x^2 P(X = x) - (np)^2 = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1 - p)^{n-x} - n^2 p^2
\]

Again with some messy algebra we can show that this sum simplifies to $Var(X) = np(1 - p)$. Therefore

If $X \sim Bin(n, p)$ then

\[
Var(X) = np(1 - p)
\]  

(2.7)

and

\[
SD(X) = \sqrt{np(1 - p)}
\]  

(2.8)

Now we are ready to finish off Case 6

**Case 6 Revisited**

We have already seen that the number of defects is Binomial i.e. $X \sim Bin(n = 100, p = 0.005)$. Therefore

\[
Var(X) = np(1 - p) = 100 \times 0.005 \times 0.995 = 0.4975
\]

and

\[
SD(X) = \sqrt{0.4975} = 0.7053
\]

The standard deviation is less than 0.75 so the second condition is fulfilled.
2.6 Poisson probability distributions

Case 7 provides an illustration of an application of the Poisson distribution.

Case 7  Dealing with customer complaints

After reaching the top of Outtel you discover that all the fun was fighting your way there and now there is nothing to do all day but make heaps of money! You decide to give it all away so you can start at the bottom and work your way up (this happens often in practice NOT). After all you are only 26!

You join Microhard! They want to take you on as their new president but you insist on starting in the customer relations department (as you might guess this department is a real mess at Microhard) so you can make your way to the top again. As part of their new “customer orientation” they are setting up an 800 number for dealing with customer complaints. As such it is considered very important that there are enough operators so that a customer almost never has to wait for more than a few minutes to speak to someone. You have been put in charge of determining the number of operators.

It has been calculated that each operator can handle 2 customers per 10 minute period and that on average there will be 12 calls per hour. It is important that in any 10 minute period there is no more than a 0.01 probability of there being more calls than the operators can handle.

We will return to this case after learning more about the Poisson distribution.

Many random variables involving “counts” have a probability distribution called the “Poisson”. Some possible examples include:

- Number of arrivals at a checkout counter during a five minute period.
- Number of flaws in a pane of glass.
- Number of telephone calls received at a switchboard during a one minute period.
- Number of touchdowns scored in a football game.
- Number of machine breakdowns during a night shift.

Definition 23 If a random variable, X, has a Poisson distribution with \( \lambda = EX \) (written \( X \sim \text{Poisson}(\lambda) \)) then its probability mass function (or distribution) is

\[
P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \ldots
\]
In other words

\[
P(X = 0) = e^{-\lambda} \\
P(X = 1) = \frac{e^{-\lambda}\lambda^1}{1!} \\
P(X = 2) = \frac{e^{-\lambda}\lambda^2}{2!}
\]

etc.

Example one
If \(X \sim \text{Poisson}(\lambda = 2)\) what is the probability that \(X\) is less than or equal to 2?

\[
P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-2} + e^{-2}\lambda + e^{-2}\frac{\lambda^2}{2!}
\]

\[
= e^{-2}(1 + 2 + \frac{2^2}{2}) = 0.6767
\]

Example two
The number of bubbles in plate glass produced by Guardian industries is Poisson with an average of 0.004 bubbles per square foot. What is the distribution of the number of bubbles in a 20 \times 5 foot window and what is the probability of no bubbles?

\[
\lambda = \text{average number of bubbles in a 20 \times 5 foot window}
= 20 \times 5 \times 0.004 = 0.4
\]

Therefore if \(X\) is the number of bubbles in a 20 \times 5 foot window then

\(X \sim \text{Poisson}(\lambda = 0.4)\)

and

\[
P(X = 0) = e^{-\lambda} = e^{-0.4} = 0.6703
\]

Interesting property of Poisson
The Poisson distribution has an interesting property that if \(X \sim \text{Poisson}(\lambda)\) then

\[
EX = \lambda \quad \text{and} \quad Var(X) = \lambda
\]

So for example if \(X \sim \text{Poisson}(\lambda = 9)\) then \(EX = 9, Var(X) = 9\) and \(SD(X) = 3\).

Case 7 Revisited
Let \(X\) be the number of calls in a random 10 minute period. We will assume that \(X \sim \text{Poisson}(\lambda)\) but what is \(\lambda\)?

\[
\lambda = \text{average calls per 10 minutes} = \text{average per hour}/6 = 12/6 = 2
\]
How many operators should we choose?
If we choose 1 then only 2 calls can be handled in a 10 minute period so we would want \( P(X \leq 2) \geq 0.99 \)

\[
P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\
= e^{-2} + e^{-2}2 + e^{-2}2^2/2!
= 0.6767
\]

This is not high enough so we need to try 2 operators. Now we can handle up to 4 calls in a 10 minute period so we need \( P(X \leq 4) \geq 0.99 \)

\[
P(X \leq 4) = P(X \leq 2) + P(X = 3) + P(X = 4) \\
= 0.6767 + e^{-2}2^3/3! + e^{-2}2^4/4! \\
= 0.9473
\]

This is not high enough so we need to try 3 operators. Now we can handle up to 6 calls in a 10 minute period so we need \( P(X \leq 6) \geq 0.99 \)

\[
P(X \leq 6) = P(X \leq 4) + P(X = 5) + P(X = 6) \\
= 0.9473 + e^{-2}2^5/5! + e^{-2}2^6/6! \\
= 0.9955
\]

Therefore we need at least 3 operators.

## 2.7 Covariance and correlation

So far we have learnt about the expected value of a random variable (a measure of its average) and the variance (a measure of how close the random variable usually is to its expected value). In this section we will talk about two related concepts, namely covariance and correlation.

### 2.7.1 Covariance

**Case 8 Important stuff like making money**

*Rumors are running through Microhard of a possible killing to be made in the stock market. Two companies Almost There and Even Closer are both on the verge of discovering a cure for the common cold! Obviously which ever company makes the breakthrough first will be worth a fortune. However as soon as the formula is discovered it will be patented and the other company will get nothing. Both companies have invested heavily in research so the losing company will undoubtable go bankrupt and the shares will be worth nothing.*

*Both companies shares are trading at $2 at present but the winning companies shares will increase to $10 while the losing companies shares will be worth nothing. The two companies seem to be neck and neck so they have an equal chance of discovering the cure first. You are keen to invest your life savings ($10,000) as you can see big profits to be had but you are unsure which company to invest in. If you choose the wrong one you will also be bankrupt!*
What should you do? You call up your favourite 309 Professor. After he balls you out for forgetting everything you learnt in his class he gives you the solution (for a small fee!)

We will return to this case after defining what we mean by covariance.

**Definition 24** The Covariance between two random variables, \(X\) and \(Y\), is defined as

\[
Cov(X, Y) = E[(X - EX)(Y - EY)]
\]

or equivalently

\[
Cov(X, Y) = E(XY) - (EX)(EY)
\]

Covariance is a measure of how two random variables vary together.

- \(Cov(X, Y) > 0\) means “if \(X\) is large then \(Y\) tends to be large also”. \(X\) and \(Y\) are said to have a positive relationship.

- \(Cov(X, Y) < 0\) means “if \(X\) is large then \(Y\) tends to be small”. \(X\) and \(Y\) are said to have a negative relationship.

- \(Cov(X, Y) = 0\) means “there is no clear trend”.

**Example one**
Suppose \(X\) and \(Y\) have the following joint distribution

<table>
<thead>
<tr>
<th></th>
<th>(-1)</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-1)</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td></td>
</tr>
</tbody>
</table>

Then \(EX = 0\), \(EY = 0\) and \(E(XY) = 2/3\) so \(Cov(X, Y) = 2/3 - 0 \times 0 = 2/3\). Therefore \(X\) and \(Y\) will have a positive relationship.

**Example two**
Suppose \(X\) and \(Y\) have the following joint distribution

<table>
<thead>
<tr>
<th></th>
<th>(-1)</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-1)</td>
<td>0</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td></td>
</tr>
</tbody>
</table>
Then $EX = 0$, $EY = 0$ and $E(XY) = -2/3$ so $Cov(X, Y) = -2/3 - 0 \times 0 = -2/3$. Therefore $X$ and $Y$ have a negative relationship.

**Example three**

Suppose $X$ and $Y$ have the following joint distribution

<table>
<thead>
<tr>
<th></th>
<th>-1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>-1</td>
<td>1/9</td>
<td>1/9</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1/9</td>
<td>1/9</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1/9</td>
<td>1/9</td>
</tr>
<tr>
<td></td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
</tbody>
</table>

Then $EX = 0$, $EY = 0$ and $E(XY) = 0$ so $Cov(X, Y) = 0 - 0 \times 0 = 0$. Therefore $X$ and $Y$ have no clear relationship.

**Useful results for covariance**

The covariance between two random variables is especially useful for calculating the variance of a sum of random variables. This is commonly used in finance for “calculating the risk of a portfolio”.

1. For any random variables $X$ and $Y$

   $$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

   (Recall $Var(X + Y) = Var(X) + Var(Y)$ if $X$ and $Y$ are independent)

2. For any random variables $X$ and $Y$

   $$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

3. If $X$ and $Y$ are independent then

   $$Cov(X, Y) = 0$$

   However note that the fact $Cov(X, Y) = 0$ does not mean $X$ and $Y$ are independent!

We will prove the first result.

**Proof of first result**

$$Var(X + Y) = E((X + Y)^2) - [E(X + Y)]^2$$

$$= E(X^2 + 2XY + Y^2) - [EX + EY]^2$$

$$= EX^2 + 2E(XY) + EY^2 - (EX)^2 - (EY)^2 - 2(EX)(EY)$$

$$= [EX^2 - (EX)^2] + [EY^2 - (EY)^2] + 2E(XY) - (EX)(EY)$$

$$= Var(X) + Var(Y) + 2Cov(X, Y)$$
Notice that since $Cov(X,Y)$ can be negative this means that $Var(X + Y)$ can be zero even if $Var(X)$ and $Var(Y)$ are greater than zero!

**Example one**

Imagine you have two fair coins.

\[
X = \begin{cases} 
0 & \text{if coin 1 lands tails} \\
1 & \text{if coin 1 lands heads} 
\end{cases}
\]

\[
Y = \begin{cases} 
0 & \text{if coin 2 lands tails} \\
1 & \text{if coin 2 lands heads} 
\end{cases}
\]

Let $Z$ be the number of heads showing on the two coins (i.e. $Z = X + Y$). What is the variance of $Z$? If the coins are independent then

\[
Var(Z) = Var(X + Y) = Var(X) + Var(Y) = 1/4 + 1/4 = 1/2
\]

this is greater than zero as you may expect because $Z$ could be 0, 1 or 2. However now imagine that you glue the coins together, side by side, so that one head and one tail is always showing. Then either coin can land heads or tails but there will always be exactly one head showing. Thus $Var(X) > 0, Var(Y) > 0$ but $Var(X + Y) = 0$.

**Case 8 Revisited**

Let $X$ be the stock price of “Almost There” and $Y$ be the stock price of “Even Closer”. Then the joint distribution of $X$ and $Y$ is

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
</tr>
</tbody>
</table>

Therefore $EX = 0 \times 1/2 + 10 \times 1/2 = 5$ so the expected profit of investing in Almost There is $3 per share. However it is possible we could lose all our money if they go broke!

$EY = 0 \times 1/2 + 10 \times 1/2 = 5$ so the expected profit of investing in Even Closer is also $3 per share. Again we could lose all our money if they go broke!

What would happen if we invested half our money in each? i.e. $\frac{1}{2}X + \frac{1}{2}Y$

Then $E(\frac{1}{2}X + \frac{1}{2}Y) = \frac{1}{2}EX + \frac{1}{2}EY = 5$ so this method has the same expected profit. What about the risk i.e $Var(\frac{1}{2}X + \frac{1}{2}Y)$?

\[
Var(\frac{1}{2}X + \frac{1}{2}Y) = Var(\frac{1}{2}(X + Y)) = \frac{1}{4}Var(X + Y)
\]

but

\[
Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)
\]
Now
\[ \text{Var}(X) = EX^2 - (EX)^2 = \frac{1}{2} \times 0^2 + \frac{1}{2} \times 10^2 - 5^2 = 25 \]
and similarly
\[ \text{Var}(Y) = 25 \]
Lastly we need to calculate the covariance
\[ \text{Cov}(X, Y) = EXY - EXEY = \frac{1}{2} \times 10 \times 0 + \frac{1}{2} \times 10 \times 0 - 5 \times 5 = -25 \]
Therefore
\[ \text{Var}(X + Y) = 25 + 25 - 2 \times 25 = 0 \]
So this method has the same expected return but involves no risk i.e. we are guaranteed to make a profit no matter what happens! In finance this procedure is called diversifying your portfolio.

### 2.7.2 Correlation

There is one obvious problem with covariance. It is not scale invariant.

**Example one**

Suppose I am measuring the relationship between heights of fathers in meters \((X)\) and heights of sons in meters \((Y)\) and that \(\text{Cov}(X, Y) = 1.2\). If I decide to measure heights in centimeters instead with \(S\) the height of fathers and \(T\) the height of sons I get \(S = 100X\) and \(T = 100Y\) so
\[ \text{Cov}(S, T) = \text{Cov}(100X, 100Y) = 100 \times 100 \times \text{Cov}(X, Y) = 12,000 \]
The covariance has increased by a factor of 10,000 yet the relationship is exactly the same. This means that it is impossible to tell whether the covariance you have observed is large or not because it is entirely dependent on the scale that is used.

To avoid this problem we often use correlation instead.

**Definition 25** The correlation between two random variables, \(X\) and \(Y\) is denoted as \(\text{Cor}(X, Y)\) or \(\rho\) and is equal to
\[ \text{Cor}(X, Y) = \rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \]
Correlation has the following properties

1. For any random variables, \(X\) and \(Y\),
   \[ -1 \leq \rho \leq 1 \]
2. If \(\rho > 0\) then \(X\) and \(Y\) have a positive relationship. In particular if \(\rho = 1\) then \(Y = aX + b\) for some constants \(a\) and \(b\) \((a > 0)\).
3. If \(\rho < 0\) then \(X\) and \(Y\) have a negative relationship. In particular if \(\rho = -1\) then \(Y = aX + b\) for some constants \(a\) and \(b\) \((a < 0)\).
4. \(\rho\) is scale invariant. In other words no matter what units we use to measure \(X\) and \(Y\), \(\rho\) will remain unchanged.
This last property is very important. It means that if $\rho$ is large (i.e. close to 1 or $-1$) then that implies a strong relationship no matter what units we are using.

**Example one continued**

Recall $X = \text{heights of fathers measured in meters}$ and $Y = \text{heights of sons measured in meters}$. Suppose that $\text{Var}(X) = \text{Var}(Y) = 2$ and $\text{Cov}(X, Y) = 1.2$. Then

$$
\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{1.2}{\sqrt{2} \times 2} = 0.6
$$

This indicates a “fairly” strong positive relationship.

Also recall that $S = \text{heights of fathers measured in centimeters}$ and $T = \text{heights of sons measured in centimeters}$. We have already seen that $\text{Cov}(S, T) = 12,000$. We can also calculate

$$
\text{Var}(S) = \text{Var}(100X) = 100^2\text{Var}(X) = 20,000
$$

and similarly $\text{Var}(T) = 20,000$. Therefore

$$
\rho = \frac{\text{Cov}(S, T)}{\sqrt{\text{Var}(S)\text{Var}(T)}} = \frac{12,000}{\sqrt{20,000 \times 20,000}} = 0.6
$$

The fact that the units have changed has had no effect on $\rho$.

Correlations will be used heavily in BUAD 310 when you study regression.
Chapter 3

Statistical Gossip: What To Do When Your Data Aren’t “Discrete” (A “Normal” State Of Affairs)

In this chapter we will discuss the second type of random variables, namely “Continuous”.

3.1 General Probability Distributions

Recall a continuous random variable takes on any value in an interval. For example

- weight,
- height,
- or temperature.

For a discrete random variable we described its distribution using the probability mass function i.e. the list of probabilities for each possible outcome. This doesn’t work for a continuous random variable because if \( X \) is continuous

\[
P(X = x) = 0 \quad \text{for any } x !!!!
\]

i.e. \( P(X = 2.0) = 0, P(X = 3.2815) = 0 \) etc. This seems counterintuitive. Why is this the case? The following example may shed some light.

Example one

Let \( X \) be the weight of a randomly chosen “one pound bag of sugar”. Even though the bag is nominally one pound in reality this will only approximate the true weight so \( X \) will be a random variable (i.e. every bag will have a slightly different weight).

Now suppose we have a scale that reads weights up to one decimal place. Then

\[
P(\text{Scale reads one pound}) = P(0.95 < X < 1.05) = 0.9 \text{ (say)}
\]

Now suppose we get a more accurate scale that reads weights up to two decimal places. Then

\[
P(\text{Scale reads one pound}) = P(0.995 < X < 1.005) = 0.1 \text{ (say)}
\]
In theory we can continue to get more accurate scales and the probability will continue to decline.

So if we get a scale that reads weights accurate to one hundred decimal places then

\[ P(\text{Scale reads one pound}) = P(0.99 \cdots 95 < X < 1.00 \cdots 05) = 0.00 \cdots 01 \text{ (say)} \]

So no matter what probability you give that the bag will weigh exactly one pound I can find a scale that is accurate enough so that the probability is less than that number. Thus the only possible probability is zero.

So listing all the probabilities makes no sense for a continuous random variable. How do we describe such a random variable?

Suppose \( X \) is the height of a randomly chosen person (in centimeters). Clearly

\[ P(175 \leq X \leq 200) > 0 \]

and in general

\[ P(a \leq X \leq b) > 0 \quad \text{(for certain values of } a \text{ and } b) \]

so we could list these probabilities for various different values of \( a \) and \( b \). The problem is that there are an infinite number of possible values for \( a \) and \( b \). How can we do this for every possible value?

The way we get around this problem is to draw a curve such that the area under the curve between any two points \( a \) and \( b \) corresponds to the probability between \( a \) and \( b \). We call this curve a “probability density function”.

**Definition 26** A “probability density function” (or curve) (pdf for short) of a random variable, \( X \), is a function, \( f \), such that

\[
P(a \leq X \leq b) = \int_a^b f(x) \, dx = \text{The area under } f \text{ between } a \text{ and } b
\]

Consider for example Figure 3.1. Here we have a curve for a particular continuous random variable. To calculate the probability that this random variable falls between \( a \) and \( b \) we simply compute the area under \( f \) between \( a \) and \( b \). This is the shaded region in the figure. We can do this for any two points.

All pdf’s have three properties

1. \( f(x) \geq 0 \) for all \( x \)

2. \( \int_{-\infty}^{\infty} f(x) \, dx = 1 \) i.e. the area under the whole curve is one.

3. For every \( a \) and \( b \)

\[
P(a \leq X \leq b) = \int_a^b f(x) \, dx = \text{Area under } f \text{ between } a \text{ and } b
\]
Example one
Suppose $X$ is a continuous random variable which is always between 0 and 2 and has a density function as given in Figure 3.2. Note that the density function is $f(x) = \frac{1}{2} x$ for $0 \leq x \leq 2$.

Is $f$ a density? Since $f$ is clearly positive we need only check that the area under it is one. But

$$\int_0^2 f(x) \, dx = \int_0^2 \frac{x}{2} \, dx = \left[ \frac{x^2}{4} \right]_0^2 = 1$$

therefore $f$ is a density.

What is $P\left(\frac{1}{2} \leq X \leq \frac{3}{2}\right)$? (i.e. the area of the shaded region)

$$P\left(\frac{1}{2} \leq X \leq \frac{3}{2}\right) = \int_{\frac{1}{2}}^{\frac{3}{2}} \frac{x}{2} \, dx = 1/2$$

What about $P\left(\frac{1}{2} < X < \frac{3}{2}\right)$?

$$P\left(\frac{1}{2} < X < \frac{3}{2}\right) = P\left(\frac{1}{2} \leq X \leq \frac{3}{2}\right) - P\left(X = \frac{1}{2}\right) - P\left(X = \frac{3}{2}\right) = P\left(\frac{1}{2} \leq X \leq \frac{3}{2}\right)$$

since the end points have zero probability. This brings up an important point. Namely, if $X$ is a continuous random variable, then for any $a$ and $b$

$$P(a < X < b) = P(a \leq X \leq b)$$

so the endpoints are not important.
3.2 Expected Values and Variances of Continuous Random Variables

3.2.1 Expected Values

Recall that for a discrete random variable

\[ EX = \sum_{i=1}^{k} x_i P(X = x_i) \]

For a continuous random variable we have a very similar definition.

**Definition 27** If \( X \) is a continuous random variable with pdf, \( f \), then

\[ EX = \int_{-\infty}^{\infty} x f(x) \, dx \]

Notice how similar this is to the definition for a discrete random variable. The only difference is that the sum is replaced by an integral and the probabilities are replaced by the density curve.

**Example one (continued)** If \( X \) is a random variable with density curve

\[ f(x) = \frac{x}{2}, \quad 0 \leq x \leq 2 \]
3.2. EXPECTED VALUES AND VARIANCES OF CONTINUOUS RANDOM VARIABLES

what is $EX$?

$$EX = \int_{-\infty}^{\infty} xf(x)dx$$
$$= \int_{0}^{2} \frac{x}{2} dx$$
$$= \frac{1}{6}[x^{3}]_{0}^{2}$$
$$= 4/3$$

Example two
Suppose we spin a pointer and read off the angle that it stops at. If $X$ is the angle then $X$ is equally likely to be any number between 0 and 360. This means $X$ has a flat density i.e.

$$f(x) = \frac{1}{360} \quad 0 \leq x \leq 360$$

What is $EX$?

$$EX = \int_{0}^{360} \frac{1}{360} dx$$
$$= \frac{1}{360}[x^{2}]_{0}^{360}$$
$$= 180$$

3.2.2 Variance

We also defined variance for discrete random variables as:

$$Var(X) = E[(X - EX)^{2}] = \sum_{i=1}^{k} (x_{i} - EX)^{2} P(X = x_{i})$$

For a continuous random variable we use

Definition 28 The variance of a continuous random variable with density $f$ is defined as

$$Var(X) = E[(X - EX)^{2}] = \int_{-\infty}^{\infty} (x - EX)^{2} f(x)dx$$

or equivalently

$$Var(X) = EX^{2} - (EX)^{2} = \int_{-\infty}^{\infty} x^{2} f(x)dx - (EX)^{2}$$

Example one (continued)
Recall

$$f(x) = \frac{x}{2} \quad 0 \leq x \leq 2$$

and $EX = 4/3$. What is $Var(X)$?

$$EX^{2} = \int_{0}^{2} x^{2} \frac{x}{2} dx$$
$$= \frac{1}{8}[x^{4}]_{0}^{2}$$
$$= 2$$
So

\[ \text{Var}(X) = E X^2 - (EX)^2 = 2 - (4/3)^2 = 2/9 \]

and

\[ SD(X) = \sqrt{2/9} \]

**Example two (continued)**

Recall for the pointer example

\[ f(x) = \frac{1}{360} \quad 0 \leq x \leq 360 \]

and \( EX = 180 \). What is \( \text{Var}(X) \)?

\[
EX^2 = \int_{0}^{360} x^2 \frac{1}{360} dx \\
= \frac{1}{360} \int_{0}^{360} x^3 dx \\
= 43,200
\]

So

\[ \text{Var}(X) = EX^2 - (EX)^2 = 43,200 - 180^2 = 10,800 \]

and

\[ SD(X) = \sqrt{10,800} = 103.9 \]

### 3.3 The Exponential Distribution

We have already learnt about the Binomial and Poisson distributions. In this section we will consider a special continuous distribution called the “Exponential”. Case 9 provides an application of the exponential which we will discuss later in the section.

**Case 9 Back to warranties**

The warranty policy that you helped implement (all those classes ago) at Outtel has been a complete success. As a result Microhard has been unable to make any impact in the micro chip market. They have decided that the only way they are going to make any inroads is if they can produce a chip with a very long warranty. You have become famous in the industry for your success at Outtel so you are called on for this problem.

Management at Microhard want to offer a 10 year warranty on their chips. However they are concerned that such a long warranty may cost them a great deal of money. They have calculated that every chip that has to be replaced under warranty costs $10. The engineers at Microhard have come up with a chip with an average lifetime of 20 years. This sounds very impressive and it seems like almost all chips should last 10 years at least!

However before they will OK this new policy management wants to know exactly what the average cost per chip will be. In particular they insist that the average cost per chip must be no greater than $1. You set your formidable mind to the problem!
Many random quantities that involve the study of lifetimes (e.g. transistors, light bulbs etc.) can be approximated by the Exponential distribution.

**Definition 29** If a random variable, \( X \), has an Exponential distribution with parameter \( \lambda \) (written \( X \sim \text{Exp}(\lambda) \)) then its probability density function is

\[
f(x) = \begin{cases} 
\lambda e^{-\lambda x}, & x > 0 \\
0, & \text{otherwise}
\end{cases}
\]

Figure 3.3 provides an illustration of the density curve for an exponential random variable.

![Density Curve](image)

Based on the density curve we can calculate probabilities, expected values and variances for an exponential. For example

\[
P(X \leq x) = \int_0^x \lambda e^{-\lambda t} \, dt \\
= \left[-e^{-\lambda t}\right]_0^x \\
= 1 - e^{-\lambda x}
\]

and

\[
EX = \int_0^\infty t f(t) \, dt \\
= \int_0^\infty t\lambda e^{-\lambda t} \, dt \\
= \frac{1}{\lambda} \quad \text{(using integration by parts)}
\]

Similarly we can show that

\[
Var(X) = \frac{1}{\lambda^2}
\]
Example one
Suppose that the lifetimes of a brand of light bulbs are approximately exponentially distributed with a mean of 1000 hours.

1. What is $\lambda$?

$$\quad 1000 = EX = \frac{1}{\lambda} \quad \text{therefore} \quad \lambda = \frac{1}{1000}$$

2. What is $\text{Var}(X)$?

$$\quad \text{Var}(X) = \frac{1}{\lambda^2} = 1,000,000$$

3. What is $P(X > 1200)$?

$$\quad P(X > 1200) = 1 - P(X \leq 1200) = 1 - (1 - e^{-1200/1000}) = e^{-1.2} = 0.301$$

4. What is $P(500 \leq X \leq 900)$?

$$\quad P(500 \leq X \leq 900) = P(X \leq 900) - P(X \leq 500)$$
$$\quad = (1 - e^{-900/1000}) - (1 - e^{-500/1000})$$
$$\quad = e^{-0.9} - e^{-0.5} = 0.200$$

Case 9 Revisited

If we assume that the lifetimes are exponential then $\lambda = 1/EX = 1/20$ and the probability that a chip will fail inside 10 years is

$$\quad P(X \leq 10) = 1 - e^{-10\lambda} = 1 - e^{-0.5} = 0.394$$

So the expected cost per chip is $\$10 \times 0.394 = \$3.94$ This is much larger than the $\$1 limit that was set. For the cost to be only $\$1 we would need $P(X \leq 10) = 0.1$ which implies

$$\quad 1 - e^{-10\lambda} = 0.1 \quad \Rightarrow \quad e^{-10\lambda} = 0.9$$
$$\quad \Rightarrow \quad \lambda = 0.0105$$
$$\quad \Rightarrow \quad EX = 1/0.0105 = 94.91 \text{ years!}$$

3.4 The Normal Distribution

In this section we are going to talk about the single most important random variable in statistics, namely the “Normal”. Case 10 provides an introduction.
Case 10 A little outside consulting

A friend of yours in the auto manufacturing industry asks for your advice and you agree to help her out (a friend in need is a friend in deed and more importantly she pays well!). She is in charge of quality control for her plant and they are currently in the process of redesigning the axle system for their latest model. There is a certain amount of randomness in the axle production process. An ideal axle is exactly 2 meters long. However any axle in the range 1.995 to 2.005 meters will be acceptable. Her engineers have calculated that the best they can do with their current plant setup is to produce axles with an average length of 2 meters and a standard deviation of 0.0025 meters.

Your friend wants to know what proportion of axles produced will need to be scrapped under the current setup. She would also like to know how low they would need to get the variability to scrap only 1% of the axles.

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

Figure 3.4: An example of a density curve for a normal random variable.

Many, many random quantities turn out to be approximately “normally distributed”. For example

1. peoples heights and weights,

2. length of axels,

3. costs and prices.

Definition 30 If a random variable, X, has a Normal distribution with expected value of \(\mu\) and variance of \(\sigma^2\) (written \(X \sim N(\mu, \sigma^2)\)) then its probability density function is

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty
\]
CHAPTER 3. CONTINUOUS RANDOM VARIABLES

Figure 3.5: An example of normal curves with different variances.

\[ \mu = E(X) \] and \[ \sigma^2 = Var(X) \].

Figure 3.4 provides an illustration of the density curve for a Normal random variable. Notice the bell shape. This is sometimes referred to as the bell curve. Figure 3.5 illustrates two normals, both with a mean of 0 but different variances. Notice that the larger the variance the more peaked the curve is. Also notice that the curve is always symmetric about the mean (0 in this case). Figure 3.6 illustrates two normals with different means. Notice that changing the mean causes the curve to shift but does not alter its shape.

How might we use the normal distribution?

**Example one**

Let \( X \) be the weight of a randomly chosen “1 pound” bag of sugar. We will assume that the weights of the bags are approximately normally distributed with a mean of 1 pound and a variance of 0.01. All bags are weighed and bags weighing less than 0.8 pounds are rejected (i.e. not sold). What proportion of bags will be rejected or equivalently what is \( P(X < 0.8) \)?

We know that for a continuous random variable we calculate probabilities by computing the area under the density curve. Figure 3.7 shows the density for this random variable. The shaded region is the area below 0.8 and represents the probability we are interested in. The question then becomes; How do we calculate the area of the shaded region? One possibility is to use calculus i.e.

\[
P(X < 0.8) = \int_{-\infty}^{0.8} f(x) \, dx
\]

\[
= \int_{-\infty}^{0.8} \frac{1}{\sqrt{2\pi} \times 0.01} e^{-\frac{(x-1)^2}{2 \times 0.01}} \, dx
\]

\[
= ???
\]

It turns out that there is no closed form solution to this problem (i.e. it can’t be solved analytically). We will try an easier problem and come back to this one.
3.4. THE NORMAL DISTRIBUTION

![Diagram of normal distribution curves with different means]

Figure 3.6: An example of normal curves with the same variance but different means.

3.4.1 The Standard Normal Distribution

We define the standard normal in the following way.

**Definition 31** The standard normal distribution is a normal with \( \mu = 0 \) (mean zero) and variance (and standard deviation) of 1 (\( \sigma^2 = 1 \)).

We usually use the letter \( Z \) to denote a standard normal, i.e.

\[
Z \sim N(0,1)
\]

Note that the density curve of \( Z \) is

\[
f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty
\]

This function looks a little easier to deal with than that of the general normal curve, see (3.4). Can we calculate probabilities for the standard normal? For example \( P(Z < 0) \). The shaded region in Figure 3.8 illustrates the area we wish to calculate. From the symmetry of the normal curve we see that \( P(Z < 0) = P(Z > 0) \) and since the total area must be one we can conclude that

\[
P(Z < 0) = \frac{1}{2}
\]

What about \( P(Z \leq 1) \)? We could use calculus again i.e.

\[
P(Z \leq 1) = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dx = ???
\]

Unfortunately there is still no closed form solution to this problem. However we do have tables in the text book which give us areas under the standard normal curve.
If $Z \sim N(0, 1)$ then the tables give

$$P(0 \leq Z \leq z)$$

Figure 3.9 provides an illustration. From the tables we see $P(0 \leq Z \leq 1) = 0.3413$ so

$$P(Z \leq 1) = P(Z \leq 0) + P(0 \leq Z \leq 1) = 0.5 + 0.3413 = 0.8413$$

Some more examples

1.

$$P(1 \leq Z \leq 2) = P(0 \leq Z \leq 2) - P(0 \leq Z \leq 1)$$

$$= 0.4772 - 0.3413 \quad \text{from tables}$$

$$= 0.1359$$

2.

$$P(Z \geq 2) = 1 - P(Z \leq 2)$$

$$= 1 - (P(Z \leq 0) + P(0 \leq 2))$$

$$= 1 - (\frac{1}{2} + 0.4772) \quad \text{from tables}$$

$$= 0.0228$$

3.

$$P(-1 \leq Z \leq 1) = P(0 \leq Z \leq 1) + P(-1 \leq Z \leq 0)$$

$$= P(0 \leq Z \leq 1) + P(0 \leq Z \leq 1) \quad \text{by symmetry}$$

$$= 0.3413 + 0.3413 \quad \text{from tables}$$

$$= 0.6826$$
4. Suppose we want to find $z$ such that $P(Z \geq z) = 0.025$. This is equivalent to finding $Z$ such that $P(0 \leq Z \leq z) = 0.475$ (why?). From the tables we see that

$$P(0 \leq Z \leq 1.96) = 0.475$$

so $z = 1.96$.

### 3.4.2 Back to the General Normal

In the previous section we saw that calculating probabilities for a standard normal is relatively easy but this does not seem to help us unless $\mu = 0$ and $\sigma^2 = 1$. It is not possible to have tables for every possible value of $\mu$ and $\sigma^2$ so what do we do?

It turns out that normal random variables have some very useful properties.

1. If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ (a and b constants) then

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

i.e. $Y$ is normal with expected value $a\mu + b$ and variance $a^2\sigma^2$.

2. If $X$ and $Y$ are independent with $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ then

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$
Now suppose that $X \sim N(\mu, \sigma^2)$ and we let

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$

What is the distribution of $Z$? From the above results we know that $Z$ must be normal with expected value $a\mu + b$ and variance $a^2\sigma^2$. In our case $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$. Therefore

$$EZ = a\mu + b = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0$$

and

$$Var(Z) = a^2\sigma^2 = \left(\frac{1}{\sigma}\right)^2\sigma^2 = 1$$

So

$$Z \sim N(0, 1)$$

Therefore if you take any normal, subtract off the mean and divide by the standard deviation you end up with a standard normal. This allows us to calculate probabilities for any normal using the standard normal tables. For example suppose $X \sim N(1, 100)$. What is $P(X \geq 2)$?

$$P(X \geq 2) = P\left(\frac{X - \mu}{\sigma} \geq \frac{2 - 1}{\sqrt{100}}\right)$$

$$= P\left(Z \geq \frac{1}{10}\right)$$

$$= 0.4602$$
Figure 3.10: The left hand graph is a general normal. We wish to calculate $P(a \leq X \leq b)$ which is Area I. The right hand graph is obtained by transforming the general normal to a standard normal. Here we are calculate $P\left(\frac{x-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$ which is Area II. Area I and Area II are equal.

In general if $X \sim N(\mu, \sigma^2)$ then

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right)$$

$$= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \quad Z \sim N(0, 1)$$

Figure 3.10 illustrates this result. The shaded region in the left hand figure is the area or probability we wish to compute. The shaded region in the right hand figure is the corresponding area under the standard normal curve. The two areas are equal and we can calculate the second one from our tables.

So to recap. If we wish to calculate the area for a general normal we use the following formula

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \quad Z \sim N(0, 1) \quad (3.5)$$

Example one (continued)
Recall for the sugar example we had bags of sugar with mean weight of 1 pound and variance of
0.01. We wanted to calculate the fraction with weight less than 0.8 pounds since these would be rejected. $X \sim N(1, 0.01)$ so

$$P(X < 0.8) = P\left( \frac{X - \mu}{\sigma} \leq \frac{0.8 - \mu}{\sigma} \right)$$

$$= P\left( \frac{X - \mu}{\sigma} \leq \frac{0.8 - 1}{0.1} \right)$$

$$= P(Z \leq -2)$$

$$= 0.0228$$

We will reject 2.28% of the bags of sugar.

**Case 10 revisited**

If $X$ is the length of a randomly produced axle then under the current setup

$$X \sim N(2, 0.0025^2)$$

An axle is OK provided $1.995 \leq X \leq 2.005$ so we want to calculate

$$P(1.995 \leq X \leq 2.005) = P\left( \frac{1.995 - 2}{0.0025} \leq \frac{X - \mu}{\sigma} \leq \frac{2.005 - 2}{0.0025} \right)$$

$$= P(-2 \leq Z \leq 2)$$

$$= 1 - 2 \times 0.0228$$

$$= 0.954$$

Therefore under the current setup we will reject about 4.6% of all axles. To calculate how low $\sigma$ would need to be before the rejection rate dropped to 1% we need to work backwards. From the tables we know

$$P(-2.6 \leq Z \leq 2.6) = 0.99$$

so we need to find $\sigma$ such that

$$2.6 = \frac{2.005 - 2}{\sigma} \quad \Rightarrow \quad \sigma = \frac{2.005}{2.6} = 0.0019$$

We need to reduce the standard deviation down to 0.0019 to get the rejection rate down to 1%.

### 3.5 Adding Normal Random Variables

In the previous section we discussed in some detail calculating probabilities for normal random variables. In this chapter we will discuss some further useful properties of the normal distribution. Case 11 provides a motivating example.

**Case 11 A little more outside consulting**

As is usually the case when people talk to you your friend is very impressed. She comes back with a harder problem! Her plant is also redesigning the engine piston manufacturing process. The
engineers believe they can manufacture engine pistons with diameters that are approximately Normally distributed with a mean of 30.00 mm and a standard deviation of 0.05 mm. They already produce cylinders which are approximately Normally distributed with a mean of 30.10 mm and a standard deviation of 0.02 mm.

Under these specs what is the probability that a random piston will fit inside a random cylinder? You are starting to wonder why some one in charge of quality control can’t answer these questions her self.

We will use the following extremely useful properties of the normal to answer this question.

<table>
<thead>
<tr>
<th>1. If $X$ and $Y$ are independent normal random variables and $Z = aX + bY$</th>
</tr>
</thead>
<tbody>
<tr>
<td>then $Z$ is also normal i.e.</td>
</tr>
<tr>
<td>$Z \sim N(aEX + bEY, a^2\text{Var}(X) + b^2\text{Var}(Y))$</td>
</tr>
</tbody>
</table>

2. If $X_1, X_2, \ldots, X_n$ are all independent normal random variables with mean $\mu$ and variance $\sigma^2$ then $\sum_{i=1}^{n} X_i$ is also normal i.e. $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

Example one
Suppose $X \sim N(1, 2), Y \sim N(3, 5)$ and $X$ and $Y$ are independent. Then what is the distribution of:

1. $X + Y$.
   $X + Y \sim N(1 + 3, 2 + 5) = N(4, 7)$

2. $2X + 3Y$.
   $2X + 3Y \sim N(2 \times 1 + 3 \times 3, 2^2 \times 2 + 3^2 \times 5) = N(11, 53)$

3. $X - Y$.
   $X - Y = X + (-1)Y \sim N(1 - 3, 2 + 5) = N(-2, 7)$

4. $3X - Y$.
   $3X - Y \sim N(3 \times 1 - 3, 3^2 \times 2 + 5) = N(0, 23)$

Example two
A commuter plane carries 12 people. Each person carries on luggage. Let $X_i$ be the weight of the
ith person's luggage. Assume that
\[ X_i \sim N(35, 100) \]
and the \( X_i \) are all independent. The plane can not take off if the total luggage weighs more than 500 pounds. What is the probability that the plane can not take off?

Let \( Y \) be the total weight of the luggage. Then
\[ Y = X_1 + X_2 + \cdots + X_{12} \]
so
\[ Y \sim N(12 \times 35, 12 \times 100) = N(420, 1200) \]
We are interested in \( P(\text{plane can't take off}) = P(Y \geq 500) \).

\[
P(Y \geq 500) = P \left( \frac{Y - \mu}{\sigma} \geq \frac{500 - \mu}{\sigma} \right) \\
= P \left( \frac{Y - \mu}{\sigma} \geq \frac{500 - 420}{\sqrt{1200}} \right) \\
= P(Z \geq 2.31) \\
= 0.0104
\]

Case 11 revisited

Let \( C \) be the diameter of a randomly chosen cylinder and \( P \) be the diameter of a randomly chosen piston. Then
\[ C \sim N(30, 0.10^2) \quad \text{and} \quad P \sim N(30.00, 0.05^2) \]
We need to calculate \( P(C > P) \) (why?). This is a hard calculation because we need to find the probability that one random variable is greater than another. However we can rewrite this probability into a much simpler form i.e.

\[
P(C > P) = P(C - P > 0) \\
= P(Y > 0)
\]

where \( Y = C - P \). Now this is just the probability that a single random variable is greater than 0. Further more we know that
\[ Y \sim N(EC - EP, Var(C) + Var(P)) = N(30.10 - 30.00, 0.02^2 + 0.05^2) = N(0.10, 0.0029) \]
Therefore

\[
P(Y > 0) = P \left( \frac{Y - EY}{\sqrt{Var(Y)}} > \frac{0 - EY}{\sqrt{Var(Y)}} \right) \\
= P \left( Z > \frac{0 - 0.1}{\sqrt{0.0029}} \right) \\
= P(Z > -1.857) \\
= 0.969
\]
Chapter 4

What Does The Mean Really Mean?

In this chapter we will introduce the idea of “sampling from a population” and how to use this sample to make ‘inferences’ about the population.

4.1 Sampling and Inference

We will start by defining exactly what we mean by a population.

Definition 32 A population is a set or collection of objects that we are interested in.

For example

• heights of USC students,
• potential customers for a new product,
• amount of liquid in Coke cans,
• or lifetimes of light bulbs.

There are any number of possible examples.

Generally a population is very large so it is impossible to “view” it all at once. Instead we try to find a way of summarizing the “important information”. This is often achieved by calculating “summary statistics”. One common summary statistic is the mean or expected value of the population ($\mu$).

Of course the only way that we can calculate $\mu$ exactly is to count every member of the population (i.e. take a census) and take the average. This is often not possible because:

• it may be too expensive
• it may destroy the population
• the population may be infinite
• it may not be possible to find all members of the population (e.g. homeless)
Instead we investigate a part or sample of the population and use it to estimate $\mu$. For the purposes of this course we will assume that we are always taking a "random sample" i.e. we randomly choose objects from the population to sample and every object is equally likely to be chosen.

Suppose we choose a random sample of $n$ objects from a population of size $N$ ($n << N$). We use the following notation to denote the values of the $n$ different objects.

\[
X_1 = \text{Value of first object or person chosen} \\
X_2 = \text{Value of second object or person chosen} \\
\vdots \\
X_n = \text{Value of $n$th object or person chosen}
\]

$X_1, X_2, \ldots, X_n$ are random variables because every object was chosen at random. In other words if the experiment was repeated (a new sample was taken) $X_1, X_2, \ldots, X_n$ would change. They are all independent random variables because each object is chosen independently from the others. Also, they all come from the same distribution because they were all taken from the same population.

Therefore an important question is: How do we use this random sample $(X_1, X_2, \ldots, X_n)$ to estimate the population mean ($\mu$)? Let

\[
X = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

then $X$ is called the sample mean. We use $X$ to estimate $\mu$.

**Example one**

Suppose we produce "1 pound" bags of sugar and we want to know the true average weight of the bags. Then the population is the weights of all bags of sugar we produce and we are interested in $\mu = \text{average weight of all bags}$. To estimate $\mu$ we take a random sample of 3 bags (say) with weights $X_1 = 0.9, X_2 = 0.95, X_3 = 1.05$ then

\[
X = \frac{1}{3} (X_1 + X_2 + X_3) = \frac{1}{3} (0.9 + 0.95 + 1.05) = 0.9667
\]

Therefore our estimate (or best guess) for $\mu$ is 0.9667. However note that $X$ is random because $X_1, X_2, \ldots, X_n$ are random so for example if we took another sample of three bottles we might get $X = 1.02$ and then our estimate for $\mu$ would be 1.02. Note that $\mu$ is probably neither of these values. They are only our best guess for $\mu$!

**Summary**

<table>
<thead>
<tr>
<th>Population mean</th>
<th>Sample Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$ (constant, what we want but unknown)</td>
<td>$X$ (random, what we know but not what we want)</td>
</tr>
</tbody>
</table>

**Example two**

Suppose the population we are interested in is the years of existence of 5 small businesses that have
been assisted by the Small Business Development Corporation. The years are 2, 4, 6, 8 and 10 so
\[ \mu = \frac{1}{5}(2 + 4 + 6 + 8 + 10) = 6 \]

Obviously with a population this small there is no need to take a sample to estimate \( \mu \) since we could just count the five members. However we will assume that we take a sample to illustrate the distribution of \( X \). Suppose that we take a random sample of size 2 “without replacement” i.e., we don’t replace the first observation before choosing the second. There are \( \binom{5}{2} = 10 \) possible outcomes i.e.,
\[ (2, 4), (2, 6), \ldots, (8, 10) \]
and corresponding values for \( X \) i.e.,
\[
\begin{array}{c|cccccccccc}
\text{Sample} & (2, 4) & (2, 6) & (2, 8) & (2, 10) & (4, 6) & (4, 8) & (4, 10) & (6, 8) & (6, 10) & (8, 10) \\
X & 3 & 4 & 5 & 6 & 6 & 7 & 7 & 8 & 9 \\
\end{array}
\]

So
\[
P(X = 3) = P(2, 4) = \frac{1}{10}
\]
\[
P(X = 5) = P(2, 8) + P(4, 6) = \frac{2}{10}
\]

etc. Just like any other random variable we can represent the possible values of \( X \) and the corresponding probabilities in the form of a table,
\[
\begin{array}{c|cccccccc}
\text{ } & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\text{ } & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{2}{10} & \frac{2}{10} & \frac{2}{10} & \frac{1}{10} \\
\end{array}
\]

Also just like any other random variable \( X \) has an expected value,
\[
EX = 3 \times \frac{1}{10} + 4 \times \frac{1}{10} + 5 \times \frac{2}{10} + 6 \times \frac{2}{10} + 7 \times \frac{2}{10} + 8 \times \frac{1}{10} + 9 \times \frac{1}{10} = 6
\]

### 4.1.1 Expected value of \( \bar{X} \)

Notice that in the previous example \( EX = \mu \). This is not a coincidence. This will be the case for any population. Here we present the proof.
\[
EX = E\left(\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right)
\]
\[
= \frac{1}{n}E(X_1 + X_2 + \ldots + X_n)
\]
\[
= \frac{1}{n}(EX_1 + EX_2 + \ldots + EX_n)
\]
\[
= \frac{1}{n}(\mu + \mu + \ldots + \mu)
\]
\[
= \mu
\]

So for any population, the expected value of \( X \) is the population mean.
\[
EX = \mu \quad \text{(4.1)}
\]
4.1.2 Variance of $\bar{X}$

Just like any other random variable we are not just interested in the mean of $X$ but also its variance. It turns out that provided the population we are sampling from is infinite (or very large if you prefer) the variance can be calculated in a similar way.

$$
Var(X) = Var\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right) \\
= \frac{1}{n^2}Var(X_1 + X_2 + \cdots + X_n) \\
= \frac{1}{n^2}[Var(X_1) + Var(X_2) + \cdots + Var(X_n)] \\
= \frac{1}{n^2}[\sigma^2 + \sigma^2 + \cdots + \sigma^2] \\
= \frac{n\sigma^2}{n^2} \\
= \frac{\sigma^2}{n}
$$

$$
Var(X) = \frac{\sigma^2}{n} \quad (4.2)
$$

The formula $Var(X) = \sigma^2/n$ is very useful because it tells us how variable our estimate of $\mu$ is. However to use it we need to know $\sigma^2$ (the population variance). Often this is unknown so we need to estimate it. Let

$$
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2
$$

We use $S^2$ to estimate $\sigma^2$. It can be shown that

$$
E S^2 = \sigma^2
$$

$S^2$ is called the sample variance.

Summary

$X$ is a random variable with

$$
EX = \mu = \text{population mean} \\
Var(X) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \text{population variance}/n \\
SD(X) = \sqrt{\sigma^2/n} = \sigma/\sqrt{n} = \text{population s.d.}/\sqrt{n}
$$

The distribution of $X$ is often called the sampling distribution because it is the distribution of the mean of a random sample. The standard deviation of $X$ ($\sigma_{\bar{X}}$) is often called the standard error.
4.2 Taking a Random Sample from a Normal Population

If the population that we take the random sample from has a normal distribution (i.e., bell shaped) then the sample mean ($X$) will have a number of desirable properties. Case 12 provides a possible scenario where these properties may be useful.

Case 12 A cure for the common cold

This consulting thing is paying so well that you decide to set up a private consulting business. Your first customer is Even Closer. It turns out that they were the first company to discover the cure for the common cold and are now mass producing it.

After many months of work they are on the verge of getting FDA approval for the drug (they are marketing it under the name going going gone cold remedy). However they need to pass one last test. They are producing tablets with 20 mg of active ingredient and the FDA wants to check that the average really is 20 mg. You know that each tablet has a random amount of active ingredient which is normally distributed with mean 20 mg and standard deviation of 0.1 mg.

The FDA will take a sample of $n$ tablets and accept them provided the average amount of active ingredient is somewhere between 19.99 and 20.01 mg. However you get to provide the tablets so you can decide how large $n$ should be. Even Closer wants at least a 99% probability that the mean will be in the required range.

We will examine some of the properties of $X$ and then return to this case.

Suppose we take a random sample, of size $n$, from a normal population. Then

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$$

What is the distribution of $X$? Recall from Chapter 3 that if $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ then

$$X_1 + X_2 + \cdots + X_n \sim N$$

and that a normal divided by a constant is still normal. Therefore

$$X = \frac{(X_1 + X_2 + \cdots + X_n)}{n} \sim \text{Normal}/n = N(EX, Var(X))$$

We have just calculated that $EX = \mu$ and $Var(X) = \sigma^2/n$ so

$$X \sim N(\mu, \sigma^2/n) \quad (4.3)$$

Example one

Suppose we take a random sample of size 100 from a normal population with mean 10 and variance 100 ($N(10,100)$).
• What is the distribution of $X$?

\[ X \sim N\left(\mu, \sigma^2/n\right) \]

\[ = N\left(10, 100/100\right) \]

\[ = N\left(10, 1\right) \]

• What is the probability that $X > 11$?

\[
P(X > 11) = P\left(\frac{X - EX}{\sqrt{Var(X)}} > \frac{11 - EX}{\sqrt{Var(X)}}\right) \]

\[ = P\left(\frac{X - EX}{\sqrt{Var(X)}} > \frac{11 - \mu}{\sigma/\sqrt{n}}\right) \]

\[ = P\left(Z > \frac{10 - 11}{10/10}\right) \]

\[ = P(Z > 1) \]

\[ = 0.1587 \]

**Example two**

Suppose we know that the amount of liquid in bottles of coke has a normal distribution with a mean of $\mu$ (unknown) and a variance of 0.01 (i.e. $N(\mu, 0.01)$). We would like to be filling the bottles to an average of one litre (i.e. we hope $\mu = 1$). We take a random sample of size $n = 25$ and get $X = 1.04$. What is the probability of getting $X$ this large or larger if $\mu$ really is 1?

If $\mu = 1$ then

\[ X \sim N\left(1, 0.01/25\right) = N\left(1, 0.0004\right) \]

so

\[
P(X > 1.04) = P\left(\frac{X - EX}{\sqrt{Var(X)}} > \frac{1.04 - EX}{\sqrt{Var(X)}}\right) \]

\[ = P\left(\frac{X - EX}{\sqrt{Var(X)}} > \frac{1.04 - \mu}{\sigma/\sqrt{n}}\right) \]

\[ = P\left(Z > \frac{1.04 - 1}{0.1/5}\right) \]

\[ = P(Z > 0.04/0.02) \]

\[ = P(Z > 2.0) \]

\[ = 0.0228 \]

This is a fairly small probability so we may be suspicious that the mean is not really one litre.

**Case 12 Revisited**

Let $X$ be the amount of active ingredient in a randomly selected tablet. Then

\[ X \sim N\left(20, 0.01\right) \]
and

\[ X \sim N(\mu, \sigma^2/n) = N(20, 0.01/n) \]

We need to find the value of \( n \) which will give

\[ P(19.99 \leq X \leq 20.01) = 0.99 \]

Therefore

\[
P(19.99 - 20 \leq X - EX \leq 20.01 - 20) \Rightarrow P\left( \frac{-0.01}{\sqrt{0.01/n}} \leq \frac{X - EX}{\sqrt{Var(X)}} \leq \frac{0.01}{\sqrt{0.01/n}} \right) \Rightarrow P(-0.1 \leq Z \leq 0.1) = 0.99 \quad Z \sim N(0, 1)\]
\[
P(0 \leq Z \leq 0.1\sqrt{n}) = 0.495 \Rightarrow 0.1\sqrt{n} = 2.575 \quad \text{from normal tables} \Rightarrow n = 25.75^2 = 663
\]

### 4.3 Central Limit Theorem

We have seen in the previous section that if we take a sample from a normal population then both the sample mean, \( X \), and the sum of the observations, \( \sum X_i \), also have a normal distribution. This allowed us to perform probability calculations on \( X \). However there are many populations that are not normal. What do we do in this case? Case 13 provides an example of a non normal population.

**Case 13 Want a safe business - try a Casino!**

*Your reputation is spreading fast now. Your latest client is a concerned investor. They are considering investing in a casino but are concerned that this seems like a very risky business since it is based on gambling. What if the casino got on an unlucky run and lost a lot of money? You try to convince him that a casino is actually a very safe investment but he is not convinced. So you collect some data for him. You find out that every day about 100,000 bets are placed at roulette with an average bet of $10. To simplify the calculations we will assume that every bet is for exactly $10 and all bets are on red or black.

Based on these numbers you calculate that the average net winnings for the casino are $52,631 per day on roulette alone! However your client is still not convinced. He has done some statistics and realizes that on any given day the net winnings could be lower than this. You go away and calculate that even though the winnings could vary the probability that they are lower than $45,000 is about 0.0078! Finally he is happy and leaves you in peace. How did you do the calculations?*

Clearly the population in this case is not normal so it is not clear where the probability calculations come from. In fact we made use of the most important result in statistics. Namely the Central Limit Theorem.
Central Limit Theorem (Version one)

Let $X_1, X_2, \ldots, X_n$ be iid with $EX_i = \mu$ and $Var(X_i) = \sigma^2$. (Note the $X_i$ are not necessarily normal). Then if $n$ is large

$$Z = \frac{X - \mu}{\sigma/\sqrt{n}} \approx N(0, 1) \quad (4.4)$$

or equivalently

$$X \approx N(\mu, \sigma^2/n) \quad (4.5)$$

Note that this result holds for a random sample from any population. Not just a normal population. As a rough rule of thumb $n = 30$ is considered large.

**Example one**

Suppose we have a random sample $X_1, X_2, \ldots, X_{100}$ from a population with unknown distribution and $EX_i = 10, Var(X_i) = 9$. What is $P(9.5 \leq X \leq 10.5)$?

$$P(9.5 \leq X \leq 10.5) = P\left(\frac{9.5 - 10}{\sqrt{9}/100} \leq \frac{X - \mu}{\sqrt{\sigma^2/n}} \leq \frac{10.5 - 10}{\sqrt{9}/100}\right)$$

$$\approx P\left(\frac{-0.5}{0.3} \leq Z \leq \frac{0.5}{0.3}\right) \quad \text{(by CLT)}$$

$$= P(-1.67 \leq Z \leq 1.67)$$

$$= 0.905$$

**Example two**

Suppose we have a random sample $X_1, X_2, \ldots, X_{100}$ from a population with an exponential distribution with $\lambda = 2$. What is $P(X > 0.6)$?

First we calculate

$$EX_i = \frac{1}{\lambda} = \frac{1}{2}$$

and

$$Var(X_i) = \frac{1}{\lambda^2} = \frac{1}{4}$$

Therefore

$$Z = \frac{X - \frac{1}{2}}{\sqrt{\frac{1}{4}/100}} \approx N(0, 1)$$

and

$$P(X > 0.6) = P\left(\frac{X - \mu}{\sqrt{\sigma^2/n}} > \frac{0.6 - 1/2}{1/20}\right)$$

$$\approx P(Z > 2) \quad \text{(by CLT)}$$

$$= 0.0228$$
We can see from the previous two examples that the CLT has many applications for calculating probabilities for sample means, $X$. However, it can also be used for dealing with sums, $\sum X_i$. Notice the CLT tells us
\[
\frac{X - \mu}{\sigma/\sqrt{n}} \approx N(0, 1)
\]
but
\[
\frac{X - \mu}{\sigma/\sqrt{n}} = \frac{n(X - \mu)}{n\sigma/\sqrt{n}} = \frac{\sum X_i - n\mu}{\sigma\sqrt{n}}
\]
which suggests the second version of CLT

**Central Limit Theorem (Version two)**

Let $X_1, X_2, \ldots, X_n$ be iid with $EX_i = \mu$ and $Var(X_i) = \sigma^2$. (Note the $X_i$ are not necessarily normal). Then if $n$ is large

\[
Z = \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \approx N(0, 1) \quad (4.6)
\]
or equivalently

\[
\sum X_i \approx N(n\mu, n\sigma^2) \quad (4.7)
\]

**Case 13 Revisited**

Let $X_i$ be the casino winnings on the $i$th game. Then

\[
X_i = \begin{cases} 
10 & \text{with probability } 20/38 \\
-10 & \text{with probability } 18/38 
\end{cases}
\]

so

\[
EX_i = 10 \times \frac{20}{38} - 10 \times \frac{18}{38} = \frac{10}{19}
\]

and

\[
Var(X_i) = EX_i^2 - (EX_i)^2 = 10^2 \times \frac{20}{38} + (-10)^2 \times \frac{18}{38} - \left(\frac{10}{19}\right)^2 = 99.72
\]

Therefore

\[
E(\text{Winnings}) = E\left(\sum X_i\right) = nEX = 100,000 \times \frac{10}{19} = 52,631
\]

\[
P(\text{Overall Profit}) = P\left(\sum X_i > 0\right) = P\left(\frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}} > \frac{0 - 52,631}{\sqrt{99.72 \times 100,000}}\right) \approx P(Z > -16.7) = 1
\]
and
\[ P(\text{Profit} > 45,000) = P\left( \sum X_i > 45,000 \right) \]
\[ = P\left( \frac{\sum X_i - np}{\sqrt{np(1-p)}} > \frac{45,000 - 52,631}{\sqrt{99.72 \times 100,000}} \right) \]
\[ \approx P(Z > -2.42) = 0.9922 \]

4.4 Normal Approximation to the Binomial Distribution

The Central Limit Theorem has further applications. One of them is to calculating Binomial probabilities.

4.4.1 Calculating Probabilities for the Number of Successes, X

Case 14 illustrates a Binomial calculation where the CLT is useful.

Case 14 Keeping your planes full

It is common in the airline industry to overbook seats on a flight because it is rare for everyone to turn up. By overbooking they can fill the flight and still get money for the people that didn’t turn up! However if more people turn up than there are seats for the flight then the airline must pay people not to take the flight so obviously they do not want to overbook a flight by too much! How many extra bookings should an airline take? In this case we will see how to use statistics to help answer this question.

Suppose that Disunified is trying to decide how many reservations to accept for their 747 flights. A 747 can hold 400 people and from past experience the airline knows that about 95% of people that book on the flight will turn up for it. It costs money to convince people to switch flights so Disunified doesn’t want more than a 1% chance that too many people will turn up. If we assume that people turn up independently of each other, how many bookings should they accept?

Suppose \( X \sim Bin(n = 100, p = 0.5) \) and we want to know \( P(40 \leq X \leq 60) \). We know from Chapter 2 that
\[ P(40 \leq X \leq 60) = P(X = 40) + P(X = 41) + \cdots + P(X = 60) \]
In theory we could calculate all these probabilities using the binomial formula but it would be a lot of work. However recall that
\[ X = \sum_{i=1}^{100} X_i \]
where
\[ X_i = \begin{cases} 1 & \text{if } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases} \]
and the \( X_i \) are independent because the trials are independent. We can calculate the expected value and variance of these \( X_i \),

\[
EX_i = 1 \times p + 0 \times (1 - p) = p
\]

and

\[
Var(X_i) = EX_i^2 - p^2 = 1^2 p + 0^2 (1 - p) - p^2 = p(1 - p)
\]

Therefore

\[
Z = \frac{\sum X_i - np}{\sqrt{np(1 - p)}} \approx N(0, 1)
\]

but the Central Limit Theorem tells us

\[
Z = \frac{\sum X_i - np}{\sqrt{np(1 - p)}} \approx N(0, 1)
\]

This is called the Normal approximation to the Binomial distribution. To summarize. If \( X \) is binomial and \( n \) is large then

\[
Z = \frac{X - np}{\sqrt{np(1 - p)}} \approx N(0, 1) \quad \text{(4.8)}
\]

or

\[
P(a \leq X \leq b) \approx P \left( \frac{a - np}{\sqrt{np(1 - p)}} \leq Z \leq \frac{b - np}{\sqrt{np(1 - p)}} \right) \quad Z \sim N(0, 1) \quad \text{(4.9)}
\]

**Example one**

If \( X \sim Bin(n = 100, p = 0.5) \) then

\[
P(40 \leq X \leq 60) \approx P \left( \frac{40 - 100 \times 0.5}{\sqrt{100 \times 0.5 \times 0.5}} \leq Z \leq \frac{60 - 100 \times 0.5}{\sqrt{100 \times 0.5 \times 0.5}} \right) = P(-2 \leq Z \leq 2) = 0.9544
\]

**Example two**

Suppose that we take a random sample of 200 people and ask them if they believe that the stock market will go down in 1999. Suppose that 45\% of Americans believe it will. What is the probability that more than half of those surveyed (i.e. 100 people) say yes?

Let \( X \) be the number of people that say yes. Then \( X \sim Bin(n = 200, p = 0.45) \) (Why?) and

\[
P(X > 100) = P \left( \frac{X - np}{\sqrt{np(1 - p)}} > \frac{100 - 200 \times 0.45}{\sqrt{200 \times 0.45 \times 0.55}} \right)
\]

\[
\approx P(Z \geq \frac{10}{7.04})
\]

\[
= P(Z \geq 1.42) = 0.0778
\]
CHAPTER 4. WHAT DOES THE MEAN REALLY MEAN?

Case 14 Revisited

Let $X$ be the number of people that turn up. Then

$$X \sim Bin(n, p = 0.95)$$

and the question is; What is $n$? We know

$$P(X > 400) = 0.01$$

Therefore

$$P(X > 400) = 0.01$$

$$\Rightarrow P \left( \frac{X - np}{\sqrt{np(1 - p)}} > \frac{400 - n \times 0.95}{\sqrt{n \times 0.95 \times 0.05}} \right) = 0.01$$

$$\Rightarrow P \left( Z > \frac{400 - n \times 0.95}{\sqrt{n0.218}} \right) \approx 0.01$$

$$\Rightarrow 2.33 = \frac{400 - n \times 0.95}{\sqrt{n0.218}}$$

$$\Rightarrow (\sqrt{n})^2 \times 0.95 + 0.5078\sqrt{n} - 400 = 0$$

We can use the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

with $x = \sqrt{n}, a = 0.95, b = 0.5078$ and $c = -400$. This gives

$$\sqrt{n} = 20.25 \text{ or } n \approx 410$$

4.4.2 Calculating Probabilities for the Proportion of Successes, $\hat{p}$

Sometimes rather than the number of successes, $X$, we are interested in the proportion or fraction of successes,

$$\hat{p} = \frac{X}{n}$$

If $X \sim Bin(n, p)$ then

$$E\hat{p} = E\left(\frac{X}{n}\right) = \frac{1}{n}EX = \frac{1}{n}np = p$$

and

$$Var(\hat{p}) = Var\left(\frac{1}{n}X\right) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}np(1 - p) = \frac{p(1 - p)}{n}$$

which means

$$SD(\hat{p}) = \sqrt{\frac{p(1 - p)}{n}}$$

Therefore if

$$Z = \frac{X - np}{\sqrt{np(1 - p)}} \approx N(0, 1)$$

then

$$\frac{\frac{X}{n} - \frac{np}{n}}{\frac{1}{n}\sqrt{np(1 - p)}} = \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} \approx N(0, 1)$$

So the Central Limit Theorem tells us that
\[ Z = \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} = \frac{\hat{p} - E\hat{p}}{\sqrt{\text{Var}(\hat{p})}} \approx N(0, 1) \quad (n \text{ large}) \quad (4.10) \]

or equivalently

\[ \hat{p} \approx N \left( p, \frac{p(1-p)}{n} \right) \quad (n \text{ large}) \quad (4.11) \]

**Example one**
Suppose 35% of people will buy a new product we are marketing. What is the probability that at least 30% of a group of 500 people will buy the product?

\[
P(\hat{p} \geq 0.3) = P\left( \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} > \frac{0.3 - 0.35}{\sqrt{0.35 \times 0.65/500}} \right) 
\approx P(Z > -2.34) 
= 0.9904
\]

Notice that this is the same as asking for the probability that at least 150 out of 500 buy the product.

\[
P(X \geq 150) = P\left( \frac{X - np}{\sqrt{np(1 - p)}} \geq \frac{150 - 500 \times 0.35}{\sqrt{500 \times 0.35 \times 0.65}} \right) 
\approx P(Z \geq -2.34) 
= 0.9904
\]

**Example two**
Suppose that \( X \sim Bin(n = 300, p = 0.5) \)

What is \( P(0.45 \leq \hat{p} \leq 0.55) \)?

\[
P(0.45 \leq \hat{p} \leq 0.55) = P\left( \frac{0.45 - 0.5}{\sqrt{0.5 \times 0.5/300}} \leq \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} \leq \frac{0.55 - 0.5}{\sqrt{0.5 \times 0.5/300}} \right) 
\approx P(-1.73 \leq Z \leq 1.73) 
= 0.9164
\]

### 4.5 Distribution of \( \bar{X}_1 - \bar{X}_2 \) and \( \hat{P}_1 - \hat{P}_2 \)

We will begin by examining the distribution of \( X_1 - X_2 \).
4.5.1 The distribution of $\bar{X}_1 - \bar{X}_2$

Suppose that we have two samples from two different normal populations. In other words

$$X_{1,1}, X_{1,2}, \ldots, X_{1,n_1} \sim N(\mu_1, \sigma_1^2)$$
$$X_{2,1}, X_{2,2}, \ldots, X_{2,n_2} \sim N(\mu_2, \sigma_2^2)$$

We will further assume that all the observations are independent from each other. Often we are interested in using the two random samples to compare the populations. For example we might ask if both populations have the same mean i.e. $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$. A natural way to answer this question is to calculate

$$X_1 - X_2$$

and see if the difference is close to zero. However $X_1 - X_2$ is a random variable because it is based on random data. Therefore in order to tell whether $X_1 - X_2$ is “close” to zero we need to know its distribution. We will start by calculating its expected value and variance.

$$E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2$$
$$Var(X_1 - X_2) = Var(X_1) + Var(X_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$
$$SD(X_1 - X_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

We also know that $X_1$ and $X_2$ are both normal so their difference must also be normal. This tells us that

$$X_1 - X_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

or in other words

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1) \quad (4.12)$$

Note that even if the two populations are not normal we can still say that

$$Z = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0, 1)$$

by the central limit theorem, provided that $n_1$ and $n_2$ are large.

**Example one**

Suppose

$$X_{1,1}, X_{1,2}, \ldots, X_{1,90} \sim N(10, 2)$$
$$X_{2,1}, X_{2,2}, \ldots, X_{2,100} \sim N(12, 4)$$
4.5 DISTRIBUTION OF $X_1 - X_2$ AND $\hat{P}_1 - \hat{P}_2$

What is the distribution of $X_1 - X_2$?

$$X_1 - X_2 \sim N \left( 10 - 12, \frac{2}{90} + \frac{4}{100} \right)$$

$$= N(-2, 0.062)$$

What is $P(X_1 > X_2)$?

$$P(X_1 > X_2) = P(X_1 - X_2 > 0)$$

$$= P \left( \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \frac{0 + 2}{\sqrt{0.062}} \right)$$

$$= P(Z > 8.03)$$

$$\approx 0$$

Example two

Suppose that we are operating two mines and the daily production for each is approximately $N(150, 400)$ and $N(125, 625)$ respectively. If we take a random sample of five days from each mine what is the probability that the average for the first mine is less than that from the second? i.e.

$$P(X_1 \leq X_2) = P(X_1 - X_2 \leq 0)$$

We know that

$$X_1 - X_2 \sim N \left( \mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)$$

$$= N \left( 150 - 125, \frac{400}{5} + \frac{625}{5} \right)$$

$$= N(25, 205)$$

Therefore

$$P(X_1 - X_2 \leq 0) = P \left( \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \frac{0 - 25}{\sqrt{205}} \right)$$

$$= P(Z \leq -1.75)$$

$$= 0.0401$$

4.5.2 The distribution of $\hat{p}_1 - \hat{p}_2$

Suppose we have two independent binomial random variables, $X_1 \sim Bin(n_1, p_1)$ and $X_2 \sim Bin(n_2, p_2)$ and we are interested in whether they both have the same probability of success on each trial i.e. $p_1 = p_2$ or $p_1 - p_2 = 0$. Just as in the previous section a natural way to answer this question is to calculate $\hat{p}_1 - \hat{p}_2$ and see if the difference is close to zero. These are also random variables so to decide what we mean by close we must calculate the distribution. We will start with the expected value and variance.

$$E(\hat{p}_1 - \hat{p}_2) = E\hat{p}_1 - E\hat{p}_2$$

$$= p_1 - p_2$$

$$Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(\hat{p}_2)$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$
Also recall that provided \( n_1 \) and \( n_2 \) are large enough the central limit theorem tells us that both \( \hat{p}_1 \) and \( \hat{p}_2 \) will be approximately normal. Therefore their difference will be approximately normal i.e.

\[
\hat{p}_1 - \hat{p}_2 \approx N \left( p_1 - p_2, \frac{p_1 (1 - p_1)}{n_1} + \frac{p_2 (1 - p_2)}{n_2} \right)
\]
or in other words

\[
Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 (1 - p_1)}{n_1} + \frac{p_2 (1 - p_2)}{n_2}}} \approx N(0, 1) \quad \text{provided } n_1 \text{ and } n_2 \text{ are large enough} \quad (4.13)
\]

**Example one**

Suppose \( X_1 \sim Bin(50, 0.5) \) and \( X_2 \sim Bin(50, 0.4) \). What is \( P(\hat{p}_1 > \hat{p}_2) \)? We know that

\[
\hat{p}_1 - \hat{p}_2 \approx N \left( 0.5 - 0.4, \frac{0.5 \times 0.5}{50} + \frac{0.6 \times 0.4}{50} \right)
\]

\[
= N(0.1, 0.0098)
\]

Therefore

\[
P(\hat{p}_1 > \hat{p}_2) = P(\hat{p}_1 - \hat{p}_2 > 0)
\]

\[
= P \left( \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 (1 - p_1)}{n_1} + \frac{p_2 (1 - p_2)}{n_2}}} > \frac{0 - 0.1}{\sqrt{0.0098}} \right)
\]

\[
\approx P(Z > -1.01) \quad Z \sim N(0, 1)
\]

\[
= 0.8438
\]

**Example two**

Suppose we get parts from two different suppliers. We reject parts from the two suppliers at the rate of 8 out of every 100 and 5 out of every 100 respectively. Each day we use \( n_1 = 150 \) parts from the first supplier and \( n_2 = 300 \) from the second supplier. What proportion of days will the difference in proportion of parts rejected from the two suppliers (i.e. \( \hat{p}_1 - \hat{p}_2 \)) be 1% or less?

This question is asking us to calculate

\[
P(-0.01 \leq \hat{p}_1 - \hat{p}_2 \leq 0.01) \quad (\text{Why?})
\]

We know that \( p_1 = 0.08 \) and \( p_2 = 0.05 \) so

\[
\hat{p}_1 - \hat{p}_2 \approx N \left( 0.08 - 0.05, \frac{0.08 (1 - 0.08)}{150} + \frac{0.05 (1 - 0.05)}{300} \right)
\]

\[
= N(0.03, 0.000649)
\]

Therefore

\[
P(-0.01 \leq \hat{p}_1 - \hat{p}_2 \leq 0.01) = P \left( \frac{-0.01 - 0.03}{\sqrt{0.000649}} \leq \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 (1 - p_1)}{n_1} + \frac{p_2 (1 - p_2)}{n_2}}} \leq \frac{0.01 - 0.03}{\sqrt{0.000649}} \right)
\]

\[
\approx P(-1.57 \leq Z \leq -0.785)
\]

\[
= P(0.785 \leq Z \leq 1.57)
\]

\[
= 0.1566
\]
Chapter 5

How Good Is That Answer Anyway?  
Measuring The Accuracy Of Your Estimate

In the previous four chapters we were assuming that we knew all about the population of interest and then using that information to calculate probabilities. For example we might assume that the population was normal with a mean of \( \mu = 2 \) and ask for the probability that the sample mean was greater than 0. However, usually we do not know everything about the population and would like to use our random sample to learn more. In this chapter we will use the knowledge we have gained in the previous chapters to make guesses or “inferences” about the population based on our random samples. We will also provide methods for deciding how good these guesses are.

5.1 Estimators

Recall that sample statistics are numbers we calculate from a sample \( X_1, X_2, \ldots, X_n \) e.g. \( X \) and \( S^2 \). On the other hand population parameters are (generally) unknown values that we would like to know e.g. \( \mu = E X \) or \( \sigma^2 = Var(X) \). In this chapter we will learn how to use sample statistics to estimate population parameters. We will begin with a few definitions.

**Definition 33** An estimator is a sample statistic that is used to estimate an unknown population parameter. On the other hand an estimate or point estimate is an actual numerical value obtained from an estimator.

For example \( X \) is an estimator for \( \mu \). If we take a random sample and the sample mean is 10.31 (i.e. \( X = 10.31 \)), we say that 10.31 is our estimate, or point estimate, for \( \mu \). (Note that this does not mean that \( \mu = 10.31 \).) There are many possible estimators for any given population parameter e.g. we could use the sample mean, the median or the mode as estimators of \( \mu \). Usually some estimators are “better” than others. How do we decide what is a good estimator?

5.1.1 Unbiased Estimators

One desirable property is that on average the estimator is correct i.e. its expected value is equal to the population parameter. In this case the estimator is said to be an “unbiased estimator”.

79
CHAPTER 5. ESTIMATORS AND CONFIDENCE INTERVALS

Figure 5.1: Both estimators are unbiased but one has a smaller variance.

**Definition 34** An estimator is unbiased if

\[ E(\text{Estimator}) = \text{Population parameter} \]

**Example one**
In the previous chapter we showed that provided we had a random sample from the population then

\[ EX = \mu \]

So if we are interested in estimating \( \mu \) we know that \( X \) is an unbiased estimator for \( \mu \).

**Example two**
We have also briefly talked about the sample variance

\[ S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

If the population variance, \( \sigma^2 \), is unknown we can use \( S^2 \) as an estimator for it. It can be shown that

\[ ES^2 = \sigma^2 \]

so that \( S^2 \) is an unbiased estimator for \( \sigma^2 \).

However there are often many unbiased estimators. For example if the population is normal then both \( X \) and the sample median are unbiased estimators for the population mean, \( \mu \). There is a second important criteria that we need to take into account.
5.1.2 Minimum Variance Unbiased Estimators

Consider Figure 5.1. It provides an illustration of the distributions of two different estimators which we have labeled “Estimators One and Two”. Both estimators are centered around $\mu$ so both are unbiased estimators. Is one better than the other? While the second estimator is centered about $\mu$ we notice that it is often a long way below $\mu$ and often a long way above. However the first estimator is generally very close to $\mu$. Therefore we would prefer Estimator One over Estimator Two because it is generally closer to $\mu$. Estimator One is better because it has a lower variance. It is possible that there is a third estimator which is also unbiased and has even lower variance. Ideally we would like to find the unbiased estimator with minimum variance.

Definition 35 An estimator is said to be a “Minimum Variance Unbiased Estimator” if it is unbiased and has smaller variance than any other unbiased estimator.

For example, if we take a random sample from a normal population then both $X$ and the sample median will be unbiased estimators of $\mu$. However

$$Var(X) = \frac{\sigma^2}{n} \quad \text{while} \quad Var(\text{Median}) \approx \frac{\pi \sigma^2}{2n} = \frac{1.57 \sigma^2}{n} > Var(X)$$

so $X$ is a better estimator.

It is possible to spend an entire course discussing how to find Minimum Variance Unbiased Estimators but most of the concepts are beyond the scope of this course. All that is required here is the fact that, if the population is normal, then $X$ is the Uniform Minimum Variance Estimator for $\mu$. By this definition at least it is “best”.

5.1.3 Are Unbiased Estimators Always the Best Estimators? (Not Examinable)

Consider Figure 5.2. Again it illustrates the distributions of two estimators. One is unbiased and the other is not. Is the unbiased estimator better? Not necessarily. In this example the unbiased estimator has a much higher variance which means it is often a long way away from $\mu$. On the other hand the biased estimator is generally fairly close to $\mu$. In this example the biased estimator is probably better. In this course will will only deal with unbiased estimators.

5.2 Confidence Intervals for Normal Means : $\sigma^2$ Known

In the previous section we learnt that we can use $X$ as an estimator for $\mu$ and that in some sense it is the best estimator. However $X$ is still random so it is unlikely to be exactly correct i.e. it is unlikely to be exactly equal to $\mu$. Therefore we would like to have some way of deciding how good our guess for $\mu$ is. Case 15 provides a possible business application.

Case 15 How sure is sure?

One day you are sitting in your office enjoying the fruits of success when a representative from your old firm Outtel turns up. It seems that Outtel has fallen on hard times since you left (you and your statistics professor were single handedly keeping it running!)
They have been working very hard to turn the company around. As is always the case for high technology firms a large investment in research is required to stay competitive. Outtel is currently considering investing heavily in producing the Bentium 1000 chip. Outtel has many thousand computer manufacturers that would potentially be interested in purchasing this new chip. However management has calculated that the cost of investment is only justifiable if on average each manufacturer will order at least 300 Bentiums per month.

The Outtel market research department has conducted a simple survey of 9 of their potential customers and asked them how many Bentiums they would purchase per month. The average response is 325. This sounds good but management is concerned that this does not necessarily reflect the true population mean. They want to be at least 99% sure that the true mean is 300 or higher. How can this question be answered if we assume from past experience that $\sigma = 30$?

Again we will try out a simpler example and then return to this case.

5.2.1 What is a Confidence Interval?

Example one
Suppose that we have a random sample from a normal population with a variance of $\sigma^2 = 1$ but an unknown mean $\mu$, i.e.

$$X_1, X_2, \ldots, X_n \sim N(\mu, 1)$$

We would like to estimate what $\mu$ is. If we take a random sample of size 16 and get $X = 0.5$ then we know from the previous section that our best guess for $\mu$ is 0.5. But how sure are we? In other
words are we pretty sure that \( \mu \) is very close to 0.5 or could we be a long way off?

Recall from the previous chapter that

\[
X \sim N(\mu, \sigma^2/n) = N(\mu, 1/16)
\]

so \( \sigma_x = 1/4 \). We also know that for a normal random variable it falls within one standard deviation of its mean about 68% of the time, within two standard deviations about 95% of the time and within three standard deviations about 99% of the time. Figure 5.3 provides an illustration.

Suppose that \( \mu \) is really 1.5. Then \( X \sim N(1.5, 1/16) \). Figure 5.4 illustrates this distribution. We see that 0.5 is not even in the picture. In fact if \( \mu = 1.5 \) then we saw \( X \) fall FOUR standard deviations from its mean. This would almost never happen so it is almost certain that \( \mu < 1.5 \).

What about \( \mu = 1.25 \)? Then \( X \) fell three standard deviations from its mean. This is also very unlikely so it seems unlikely that \( \mu = 1.25 \).

What about \( \mu = 1.00 \)? Then \( X \) fell two standard deviations from its mean. This happens sometimes so it seems possible that \( \mu = 1.00 \).

What about \( \mu = 0.75 \)? Then \( X \) fell one standard deviation from its mean. This is reasonably likely so it seems quite plausible that \( \mu \) could equal 0.75.

What about \( \mu = 0.25 \)? Then \( X \) fell one standard deviation below its mean. This is reasonably likely so it seems quite plausible that \( \mu \) could equal 0.25.

What about \( \mu = 0 \)? Then \( X \) fell two standard deviations from its mean. This happens sometimes so it seems possible that \( \mu = 0 \).
What about $\mu = -0.25$? Then $X$ fell three standard deviations from its mean. This is also very unlikely so it seems unlikely that $\mu = -0.25$.

Therefore it seems that, even though we have made a guess for $\mu$ of 0.5, it is possible that $\mu$ could be anywhere between 0 and 1. We call the region

$$[0, 1]$$

a “Confidence Interval” for $\mu$. It is simply a list of all “reasonable” values that $\mu$ could be.

### 5.2.2 Calculating a Confidence Interval

Now that we know what a confidence interval is we need to know how to construct it. The method we used in the previous section required a lot of work and was rather ad hoc. However before constructing the interval we need to decide on the “confidence level”. In other words if we construct a confidence interval with a lower bound of $L$ and an upper bound of $U$

$$[L, U]$$

then how sure do we want to be that

$$L \leq \mu \leq U$$

It is common to use a 95% confidence level. In other words, find $L$ and $U$ so that

$$P(L \leq \mu \leq U) = 0.95$$

Another way of thinking about this is that if we took 100 samples and constructed 100 different confidence intervals from the samples about 95 would be correct. Once we have decided on the
5.2. CONFIDENCE INTERVALS FOR NORMAL MEANS : $\sigma^2$ KNOWN

...confidence level we need to calculate $L$ and $U$.

Suppose we want a 95% confidence interval. Then we need to find $L$ and $U$ so that

$$P(L \leq \mu \leq U) = 0.95$$

This is where our work in the previous chapters comes in handy. Recall that

$$\frac{X - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Now if $Z \sim N(0,1)$ then we know

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

Therefore

$$P \left( -1.96 \leq \frac{X - \mu}{\sigma/\sqrt{n}} \leq 1.96 \right) = 0.95$$

$$\Rightarrow P(-1.96 \sigma/\sqrt{n} \leq X - \mu \leq 1.96 \sigma/\sqrt{n}) = 0.95$$

$$\Rightarrow P(X - 1.96 \sigma/\sqrt{n} \leq \mu \leq X + 1.96 \sigma/\sqrt{n}) = 0.95$$

$$\Rightarrow P(L \leq \mu \leq U) = 0.95$$

where $L = X - 1.96 \sigma/\sqrt{n}$ and $U = X + 1.96 \sigma/\sqrt{n}$

A 95% confidence interval for $\mu$ is

$$[X - 1.96 \sigma/\sqrt{n}, X + 1.96 \sigma/\sqrt{n}] \quad (5.1)$$

What if we wanted to construct a confidence interval with a different confidence level? For example 99%. This works in exactly the same way except that instead of using 1.96 we use 2.57 i.e.

$$P \left( -2.57 \leq \frac{X - \mu}{\sigma/\sqrt{n}} \leq 2.57 \right) = 0.99$$

so a 99% confidence interval for $\mu$ is

$$[X - 2.57 \sigma/\sqrt{n}, X + 2.57 \sigma/\sqrt{n}]$$

What about a $100(1 - \alpha)$% confidence interval $(0 < \alpha < 1)$? i.e

$$P(L \leq \mu \leq U) = 1 - \alpha$$

Let $Z_\alpha$ be the point such that

$$P(Z > Z_\alpha) = \alpha$$

eg $Z_{0.025} = 1.96$ and $Z_{0.005} = 2.57$. Then

$$P \left( -Z_{\alpha/2} \leq \frac{X - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2} \right) = 1 - \alpha$$

$$\Rightarrow P(X - Z_{\alpha/2} \sigma/\sqrt{n} \leq \mu \leq X + Z_{\alpha/2} \sigma/\sqrt{n}) = 1 - \alpha$$
A 100(1 − α)% confidence interval for μ is

\[ [X - Z_{α/2}σ/\sqrt{n}, X + Z_{α/2}σ/\sqrt{n}] \]  

(5.2)

**Example one**

Suppose we take a sample of \( n = 100 \) from a normal population with \( σ = 2 \) and get \( X = 9.8 \). What are 90% and 95% confidence intervals for \( μ \)? First we will give the 95% interval.

\[
\begin{align*}
[X - 1.96σ/\sqrt{n}, X + 1.96σ/\sqrt{n}] &= [9.8 - 1.96 \times 2/\sqrt{100}, 9.8 + 1.96 \times 2/\sqrt{100}] \\
&= [9.408, 10.192]
\end{align*}
\]

Next we give the 90% interval. In this case \( α = 0.1 \) so \( α/2 = 0.05 \). This means \( Z_{0.05} = 1.645 \) because \( P(Z > 1.645) = 0.05 \).

\[
\begin{align*}
[X - 1.645σ/\sqrt{n}, X + 1.645σ/\sqrt{n}] &= [9.8 - 1.645 \times 2/\sqrt{100}, 9.8 + 1.645 \times 2/\sqrt{100}] \\
&= [9.471, 10.129]
\end{align*}
\]

**Example two**

Suppose we take a random sample of 10 bottles of coke and measure the level in each bottle. We get the following 10 measurements (in litres).

\[ 0.97, 0.93, 0.99, 1.01, 1.03, 1.02, 1.05, 0.90, 0.95, 1.05 \]

Assume from past experience we know that the population is approximately normal with \( σ = 0.04 \). We are interested in estimating the average amount of coke in each bottle. From this data we can calculate that

\[ X = 0.99 \]

so our best guess for \( μ \) is 0.99. However it is unlikely that \( μ \) is exactly equal to 0.99 so we want to construct a confidence interval to give us an idea of the possible values. We will construct 90% and 99% intervals, i.e. \( α = 0.1 \) \( (Z_{0.05} = 1.645) \) and \( α = 0.01 \) \( (Z_{0.005} = 2.57) \). First the 90% interval

\[
\begin{align*}
[0.99 - 1.645 \times 0.04/\sqrt{10}, 0.99 + 1.645 \times 0.04/\sqrt{10}] &= [0.969, 1.011]
\end{align*}
\]

and the 99% interval

\[
\begin{align*}
[0.99 - 2.57 \times 0.04/\sqrt{10}, 0.99 + 2.57 \times 0.04/\sqrt{10}] &= [0.957, 1.023]
\end{align*}
\]

Notice that as we increase the confidence level we also increase the width of the interval. This will always be the case. If you want to be more certain about being correct you need to have a larger range of possible values.
5.3. CONFIDENCE INTERVALS FOR NORMAL MEANS : $\sigma^2$ UNKNOWN

Case 15 Revisited

With this example we were told that $X = 325$, $\sigma = 30$ and $n = 9$. We will start by constructing a 95% confidence interval for $\mu$

$$[325 - 1.96 \times 30 / \sqrt{9}, 325 + 1.96 \times 30 / \sqrt{9}]$$

$$= [305.4, 344.6]$$

This looks good because the interval does not contain 300. It looks like $\mu$ is greater than 300. Now we will calculate a 99% interval

$$[325 - 2.57 \times 30 / \sqrt{9}, 325 + 2.57 \times 30 / \sqrt{9}]$$

$$= [299.3, 350.7]$$

This interval indicates that $\mu$ could be as low as 299.3. Roughly speaking we can be 95% sure that the mean is above 300 but not 99% sure.

5.3 Confidence Intervals for Normal Means : $\sigma^2$ Unknown

The problem with the previous confidence intervals was that they assumed that we know $\sigma$. Usually if we don’t know $\mu$ we don’t know $\sigma$ either. What should we do in this case? Case 16 provides an example of the problem.

Case 16 Let's get realistic

Having decided that the previous survey is inconclusive, the management at Outtel performs another one. This time they take a sample of 16 computer manufacturers in the hope that they will give a more accurate confidence interval.

The sorted answers are

$$275, 277, 296, 305, 309, 312, 313, 330, 332, 335, 340, 345, 350, 362, 373, 426$$

If we no longer make the (unrealistic) assumption that we know $\sigma$, how do we estimate $\mu$ and construct a confidence interval?

If $\sigma$ is unknown we must estimate it. Usually in this situation we estimate $\sigma$ using

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - X)^2}$$

but can we then say that

$$[X - 1.96S / \sqrt{n}, X + 1.96S / \sqrt{n}]$$

is a 95% confidence interval for $\mu$? The answer is NO. A confidence interval constructed in this manner will tend to be wrong a lot more than 5% of the time. What is the problem?
In the previous section we used the fact that

\[ P \left( -1.96 \leq \frac{X - \mu}{\sigma/\sqrt{n}} \leq 1.96 \right) = 0.95 \]

because

\[ \frac{X - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \]

But is it true that

\[ P \left( -1.96 \leq \frac{X - \mu}{S/\sqrt{n}} \leq 1.96 \right) = 0.95 \]

The answer is no because

\[ \frac{X - \mu}{S/\sqrt{n}} \neq N(0,1) \]

What is the distribution of \( \frac{S-\mu}{S/\sqrt{n}} \)?

5.3.1 The \( t \) distribution

**Definition 36** If \( X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2) \) then

\[ t = \frac{X - \mu}{S/\sqrt{n}} \sim t_{n-1} \]

is said to have a \( t \) distribution with \( n - 1 \) “degrees of freedom”.

The \( t \) distribution has a similar shape to the normal but has fatter tails. Figure 5.5 illustrates the \( t \) distribution with 1 and 2 degrees of freedom compared to the standard normal. As the “degrees of freedom” get large the \( t \) looks like a normal. For \( \text{dof} > 30 \) they are almost identical.
5.3. CONFIDENCE INTERVALS FOR NORMAL MEANS : $\sigma^2$ UNKNOWN

Let $t_{\alpha, \nu}$ be the point where

$$P(t_{\nu} > t_{\alpha, \nu}) = \alpha$$

where $\alpha$ the probability level and $\nu$ is the degrees of freedom, e.g.

$$
\begin{align*}
t_{0.025,1} &= 12.706 \\
t_{0.025,2} &= 4.303 \\
t_{0.025,10} &= 2.228 \\
t_{0.025,30} &= 2.042 \\
t_{0.025,\infty} &= 1.96 = Z_{0.025}
\end{align*}
$$

Note $P(t_{\nu} > t_{\alpha, \nu}) = P(t_{\nu} < -t_{\alpha, \nu})$.

5.3.2 Using the $t$ distribution to construct confidence intervals

Since

$$\frac{X - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

we know that

$$P \left( -1.96 \leq \frac{X - \mu}{S/\sqrt{n}} \leq 1.96 \right) \neq 0.95$$

but

$$P \left( -t_{0.05/2, n-1} \leq \frac{X - \mu}{S/\sqrt{n}} \leq t_{0.05/2, n-1} \right) = 0.95$$

Therefore

$$P \left( -t_{0.025, n-1} S/\sqrt{n} \leq X - \mu \leq t_{0.025, n-1} S/\sqrt{n} \right) = 0.95$$

$$\Rightarrow P \left( X - t_{0.025, n-1} S/\sqrt{n} \leq \mu \leq X + t_{0.025, n-1} S/\sqrt{n} \right) = 0.95$$

Therefore a 95% confidence interval for $\mu$ is

$$[X - t_{0.025, n-1} S/\sqrt{n}, X + t_{0.025, n-1} S/\sqrt{n}]$$

In general

A 100(1 - $\alpha$)% confidence interval for $\mu$ is

$$[X - t_{\alpha/2, n-1} S/\sqrt{n}, X + t_{\alpha/2, n-1} S/\sqrt{n}] \quad (5.3)$$

Example two (continued)

Recall that we take a random sample of 10 bottles of coke and measure the level in each bottle. We get the following 10 measurements (in litres).

$$0.97, 0.93, 0.99, 1.01, 1.03, 1.02, 1.05, 0.90, 0.95, 1.05$$
This gives $X = 0.99$. Previously we assumed that we knew $\sigma = 0.04$. However this is probably not realistic. Instead we will estimate $\sigma$ using $S$.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - X)^2$$

$$= \frac{1}{9} ((0.97 - 0.99)^2 + \cdots + (1.05 - 0.99)^2)$$

$$= 0.00264$$

so $S = \sqrt{0.00264} = 0.0514$. Therefore a 95% confidence interval for $\mu$ is

$$[X - t_{0.025,9}S/\sqrt{n}, X + t_{0.025,9}S/\sqrt{n}]$$

$$= [0.99 - 2.262 \times 0.0514/\sqrt{10}, 0.99 + 2.262 \times 0.0514/\sqrt{10}]$$

$$= [0.953, 1.027]$$

and a 99% confidence interval for $\mu$ is

$$[X - t_{0.005,9}S/\sqrt{n}, X + t_{0.005,9}S/\sqrt{n}]$$

$$= [0.99 - 3.250 \times 0.0514/\sqrt{10}, 0.99 + 3.250 \times 0.0514/\sqrt{10}]$$

$$= [0.937, 1.043]$$

**Example three**

A personal computer retailer is interested in predicting their computer sales for the following month. Over the last 12 months sales of personal computers have been

463, 522, 290, 714, 612, 401, 403, 510, 498, 715, 673, 691

Give an estimate for the future months sales and a 95% confidence interval.

From the above sales we can calculate that $X = 541$ and $S = 140.15$. Therefore our best estimate for $\mu$ is 541 and a 95% confidence interval is

$$[X - t_{0.025,11}S/\sqrt{n}, X + t_{0.025,11}S/\sqrt{n}]$$

$$= [541 - 2.201 \times 140.15/\sqrt{12}, 541 + 2.201 \times 140.15/\sqrt{12}]$$

$$= [452.0, 630.0]$$

What are some possible problems with our estimate and the confidence interval?

- **Non independence.** We assumed that each months sales were independent from the previous month. This may not be realistic.

- **Non constant mean for each month.** It is possible that there are seasonal effects so that sales are higher at certain times of the year which our calculation does not take account of. This is called a time series problem and you will learn more about it in BUAD 310.

- **We are assuming that the number of sales is not increasing or decreasing over time.** It is possible that even though sales in the past year have averaged about 540 computers per month the average will be quite different next month.

- **Non normality.** In constructing the confidence interval we assumed that the underlying population was normal. This may not be realistic. However the $t$ distribution is fairly robust to non normality provided we have a distribution that is “close” to normal.
Case 16 Revisited

From our latest sample we calculate that \( X = 330 \) and \( S = 38 \) with \( n = 16 \). Therefore a 95% confidence interval for \( \mu \) is

\[
\left[ X - t_{0.025,15} S / \sqrt{n}, X + t_{0.025,15} S / \sqrt{n} \right] = [330 - 2.131 \times 38 / \sqrt{16}, 330 + 2.131 \times 38 / \sqrt{16}] = [309.76, 350.24]
\]

which roughly speaking indicates that we are 95% certain the mean is above 300. A 99% confidence interval for \( \mu \) is

\[
\left[ X - t_{0.005,15} S / \sqrt{n}, X + t_{0.005,15} S / \sqrt{n} \right] = [330 - 2.947 \times 38 / \sqrt{16}, 330 + 2.947 \times 38 / \sqrt{16}] = [302.00, 358.00]
\]

Which indicates we are 99% certain the mean is above 300. Outtel should go ahead and market the chip.

5.4 Estimating Binomial \( p \)

Just as we may be interested in estimating the mean for a population, \( \mu \), we may also be interested in estimating \( p \) for a Binomial population. Case 17 provides an example.

Case 17 To invest or not? That is the question.

Often the result of an election can have important economic implications. As a result being able to predict the result ahead of time is obviously important. This case illustrates one possible example.

Suppose Proposition 1029 has been put on the ballot. This proposition would ban all commercial felling of trees in California. United Tree Eaters holds 80% of all the felling rights in California. Obviously if this proposition passes the company will be devastated. However at the moment it is running very well so if the proposition fails to pass they will remain profitable.

Despite the severity of Proposition 1029, support for it is running close to 50%. The shareholders of United Tree Eaters are obviously nervous and the share price has dropped a great deal in the lead up to the election. If the proposition fails to pass the share price will return to normal levels and a large profit could be made. If not the share price will plummet and a large loss could be made. You decide to use your statistics training to see if you should invest in the company.

You take a random sample of 1000 eligible voters and ask them if they will vote for Proposition 1029. 460 or 46% say they will vote for it. Should you invest in the company?

Suppose we observe \( X \) from a \( Bin(n, p) \) distribution and we want to estimate \( p \). For example we might receive a shipment of 100 parts and we observe \( X = 3 \) are defective. We might then
be interested in estimating \( p = \) proportion of defective parts that the manufacturer produces. An obvious seeming estimate is the proportion of the parts we received that were defective i.e. 
\[ \frac{X}{n} = \frac{3}{100} = 0.03. \]
Recall from the previous chapter that we call this estimator
\[ \hat{p} = \frac{X}{n} \]
In the previous chapter we showed that
\[ E\hat{p} = p \]
so \( \hat{p} \) is an unbiased estimator for \( p \). However, just as with \( X \), \( \hat{p} \) is unlikely to be exactly equal to \( p \).
We would like to find a confidence interval for \( p \). Recall we showed in the previous chapter that because of the Central Limit Theorem
\[ Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0,1) \]
Also since \( \hat{p} \approx p \)
\[ Z = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \approx N(0,1) \]
provided \( n \) is large. We can use this information to give us a confidence interval to gauge the accuracy of our estimate of \( p \).
\[
\begin{align*}
\Pr\left(-Z_{a/2} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq Z_{a/2}\right) &= 1 - \alpha \\
\Rightarrow \quad \Pr\left(-Z_{a/2}\sqrt{\hat{p}(1-\hat{p})/n} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq Z_{a/2}\sqrt{\hat{p}(1-\hat{p})/n}\right) &= 1 - \alpha \\
\Rightarrow \quad \Pr\left(\hat{p} - Z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) &= 1 - \alpha
\end{align*}
\]
A 100(1 - \( \alpha \))% confidence interval for \( p \) is
\[
\left[ \hat{p} - Z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]
\] (5.4)

**Example one**
Suppose

\[ X \sim Bin(n = 300, p) \]
with \( p \) unknown and we get an estimate for \( p \) of \( \hat{p} = 0.2 \). Then a 99% confidence interval for \( p \) is
\[
\left[ \hat{p} - Z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]
\]
\[
= \left[ 0.2 - 2.57\sqrt{\frac{0.2 \times 0.8}{300}}, 0.2 + 2.57\sqrt{\frac{0.2 \times 0.8}{300}} \right]
\]
\[
= \left[ 0.141, 0.259 \right]
\]
Example two
Suppose we get in a large shipment of widgets. We want to know what proportion are defective so we take a random sample of 250 and find 30 are defective. Estimate the proportion defective and give a 90% confidence interval.

Our estimate for $p$ is
$$\hat{p} = \frac{30}{250} = 0.12$$
and the 90% confidence interval is
$$\left[ \hat{p} - Z_{0.05} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{0.05} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$
$$= \left[ 0.12 - 1.645 \sqrt{\frac{0.12 \times 0.88}{250}}, 0.12 + 1.645 \sqrt{\frac{0.12 \times 0.88}{250}} \right]$$
$$= [0.086, 0.154]$$

Case 17 Revisited
Let $p$ be the proportion who support proposition 1029. We want to invest provided $p < 0.5$. From our random sample of $n = 1000$ people we got $\hat{p} = 0.46$ so our best estimate for $p$ is 0.46 but how sure are we about our estimate? If we calculate a 95% confidence interval we get
$$\left[ \hat{p} - Z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$
$$= \left[ 0.46 - 1.96 \sqrt{\frac{0.46 \times 0.54}{1000}}, 0.46 + 1.96 \sqrt{\frac{0.46 \times 0.54}{1000}} \right]$$
$$= [0.429, 0.491]$$
So we are roughly 95% sure the proposition will not pass. However, if we calculate a 99% confidence interval we get
$$\left[ \hat{p} - Z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$
$$= \left[ 0.46 - 2.57 \sqrt{\frac{0.46 \times 0.54}{1000}}, 0.46 + 2.57 \sqrt{\frac{0.46 \times 0.54}{1000}} \right]$$
$$= [0.419, 0.501]$$
so we can not be 99% sure that the proposition will not pass.

5.5 Sample Size
Often we are able to decide how large we would like our sample size ($n$) to be before we take the random sample. Obviously the larger $n$ is the better but it may be expensive to make $n$ too large so we must decide how small a sample we can “get away with”. There have been books written on
this subject. Here we will just briefly touch on the issue.

One common question is the following. How large a sample do I need so that the width of my confidence interval is no larger than $W$? (The width of a confidence interval is just the difference between the upper and lower bounds). For example in Case 15 $\sigma$ was 30 and we took a sample of size $n = 9$. This gave a 95\% confidence interval with a width of 39.2. Suppose we wanted to reduce that width to only 10. How large would $n$ need to be? To answer this question we can calculate a formula for $W$ and solve for $n$, i.e.

\[
W = X + 1.96\sigma/\sqrt{n} - (X + 1.96\sigma/\sqrt{n})
= 2 \times 1.96\sigma/\sqrt{n}
\Rightarrow \sqrt{n} = 2 \times 1.96\sigma/W
\Rightarrow n = 4 \times 1.96^2\sigma^2/W^2
\]

Therefore

\[
n = 4 \times 1.96^2 \times 30^2/10^2 \approx 138
\]

For a general $100(1 - \alpha)$\% confidence interval

\[
W = 2Z_{\alpha/2}\sigma/\sqrt{n}
\]

and

\[
n = \frac{4Z_{\alpha/2}^2\sigma^2}{W^2} \quad (5.5)
\]

There are many similar formulas where, for example, $\sigma$ is not assumed known and $S$ is used instead (see the text book for further examples). We will not cover them in this course. Note that if we halve $W$ then

\[
n = \frac{4Z_{\alpha/2}^2\sigma^2}{(W/2)^2} = 4 \left( \frac{4Z_{\alpha/2}^2\sigma^2}{W^2} \right)
\]

so to halve the width of our confidence interval we need to take a sample four times as large!

5.6 **Confidence Intervals for $\mu_1 - \mu_2$**

Often we are interested not just in the mean for a single population but in the difference in means between two populations.

**Example one**

In manufacturing companies many days per year can be lost though worker injuries. Not only are these injuries bad for the workers but they can cost a company a great deal of money in lost productivity. Many of the injuries are caused by bad safety procedures that could be fixed relatively simply. Imagine we wish to decide whether a new safety awareness course helps prevent injuries. We take a random sample of 100 workers and give them the training course. We then observe
the number of days lost through injuries in the following year for each of the 100 workers. These numbers are labeled

\[ X_{1,1}, X_{1,2}, \ldots, X_{1,100} \]

We also randomly pick 100 workers that have not taken the course and record the number of days lost through injuries for each of the 100 workers. These numbers are labeled

\[ X_{2,1}, X_{2,2}, \ldots, X_{2,100} \]

We are interested in whether the average number of days lost is less for the first group than the second. We will assume that we have

1. Independent random samples.
2. Normal populations i.e. the population of workers who have taken the course and the population of workers who haven’t are both normal.
3. The populations have the same variance \((\sigma_1^2 = \sigma_2^2 = \sigma^2)\).

i.e.

\[
X_{1,i} \sim N(\mu_1, \sigma^2) \\
X_{2,i} \sim N(\mu_2, \sigma^2)
\]

Then we are interested in the difference between the means for each population i.e.

\[
\mu_1 - \mu_2
\]

If the difference is zero then the training course is having no effect. If it is negative then the training course is reducing the number of days lost through injuries. If it is positive then the training course is having a detrimental effect! Suppose that

\[
X_1 = 5, \quad X_2 = 7, \quad S_1^2 = 8, \quad S_2^2 = 9, \quad n_1 = n_2 = 100
\]

where \(X_1\) is the sample mean for the first group and \(X_2\) is the mean for the second.

### 5.6.1 Confidence intervals : \(\sigma\) known

In general to estimate \(\mu\) we use the sample mean \(X\). Therefore a reasonable estimate for \(\mu_1 - \mu_2\) would be

\[
X_1 - X_2
\]

So for the above example our best estimate for \(\mu_1 - \mu_2\) is \(5 - 7 = -2\) which suggests that the course may be helping. However this is just a guess. How sure are we about it? Again a natural approach is to calculate a confidence interval for the difference. Recall from the previous chapter that

\[
X_1 - X_2 \sim N \left( \mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right) = N \left( \mu_1 - \mu_2, \sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \right)
\]

so

\[
\frac{X_1 - X_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1)
\]
Therefore

\[
P \left( -1.96 \leq \frac{X_1 - X_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \leq 1.96 \right) = 0.95
\]

\[\Rightarrow P \left( X_1 - X_2 - 1.96 \sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \leq \mu_1 - \mu_2 \leq X_1 - X_2 + 1.96 \sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right) = 0.95\]

so a 95\% confidence interval for \( \mu_1 - \mu_2 \) is

\[
\left[ X_1 - X_2 - 1.96 \sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \ X_1 - X_2 + 1.96 \sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right]
\]

5.6.2 Confidence intervals: \( \sigma \) unknown

Of course \( \sigma^2 \) is usually unknown. In this case we estimate it using

\[
S^2_p = S^2_{pooled} = \frac{(n_1 - 1)S^2_1 + (n_2 - 1)S^2_2}{n_1 + n_2 - 2} \tag{5.6}
\]

\( S^2_p \) is a weighted average of the two sample variances. Of course since we are estimating \( \sigma^2 \) we need to use the \( t \) distribution.

A 100\((1 - \alpha)\%\) confidence interval for \( \mu_1 - \mu_2 \) is

\[
\left[ X_1 - X_2 - t_{\alpha/2,n_1+n_2-2} \sqrt{S^2_p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \ X_1 - X_2 + t_{\alpha/2,n_1+n_2-2} \sqrt{S^2_p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right] \tag{5.7}
\]

Notice that the degrees of freedom are now \( n_1 + n_2 - 1 = n_1 + n_2 - 2 \).

Example one revisited

In order to compute a confidence interval we first need to calculate \( S^2_p \)

\[
S^2_p = \frac{(100 - 1)S^2_1 + (100 - 1)S^2_2}{100 + 100 - 2} = \frac{(100 - 1) \times 8 + (100 - 1) \times 9}{100 + 100 - 2} = 8.5
\]

Therefore a 99\% confidence interval for \( \mu_1 - \mu_2 \) is

\[
\left[ X_1 - X_2 - t_{0.005,100+100-2} \sqrt{S^2_p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \ X_1 - X_2 + t_{0.005,100+100-2} \sqrt{S^2_p \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right]
\]

\[
= \left[ 5 - 7 - 2.576 \sqrt{8.5 \left( \frac{1}{100} + \frac{1}{100} \right)}, \ 5 - 7 + 2.576 \sqrt{8.5 \left( \frac{1}{100} + \frac{1}{100} \right)} \right]
\]

\[
= \left[ -3.062, -0.938 \right]
\]
Since this interval does not include the value zero it appears that $\mu_1 < \mu_2$ so the course is helping.

**Example two**
Suppose we take a random sample and get the following summary statistics

\[
X_1 = 2, \quad S_1^2 = 3, \quad n_1 = 4 \\
X_2 = 3, \quad S_2^2 = 4, \quad n_2 = 6
\]

Give a 95\% confidence interval for $\mu_1 - \mu_2$.

Here

\[
S_p^2 = (3 \times (4 - 1) + 4 \times (6 - 1))/(4 + 6 - 2) = 3.625
\]

so a 95\% confidence interval is of the form

\[
\left[ X_1 - X_2 - t_{0.025,4+6-2} \sqrt{\frac{S_p^2}{n_1} + \frac{1}{n_2}}, X_1 - X_2 + t_{0.025,4+6-2} \sqrt{\frac{S_p^2}{n_1} + \frac{1}{n_2}} \right]
\]

\[
= \left[ 2 - 3 - 2.306 \sqrt{3.625 \left( \frac{1}{4} + \frac{1}{6} \right)}, 2 - 3 + 2.306 \sqrt{3.625 \left( \frac{1}{4} + \frac{1}{6} \right)} \right]
\]

\[
= [-3.834, 1.834]
\]

This interval contains the value zero so it is possible that the two populations are the same. i.e. have the same mean.

### 5.7 Confidence Intervals for $p_1 - p_2$

Just as we may be interested in the difference between two population means we may also be interested in estimating the difference between two proportions from Binomial populations, $p_1 - p_2$.

**Example one**
Suppose we receive two shipments from two different suppliers and we are interested in whether there is a difference in the proportion of defective components for each. If $X_1$ is the number of defective components for the first supplier and $X_2$ is the number of defective components for the second then

\[
X_1 \sim Bin(n_1, p_1), \quad X_2 \sim Bin(n_2, p_2)
\]

Therefore we are interested in estimating $p_1 - p_2$. If the difference is zero then the suppliers are identical. Otherwise one is better than the other. Suppose we got in shipments of size $n_1 = 120$ and $n_2 = 180$ and the number of defectives for each were $X_1 = X_2 = 5$. We estimate $p_1$ using $\hat{p}_1 = X_1/n_1$ and $p_2$ using $\hat{p}_2 = X_2/n_2$ so a natural estimate for the difference is

\[
\hat{p}_1 - \hat{p}_2 = \frac{5}{120} - \frac{5}{180} = 0.0417 - 0.0278 = 0.0139
\]

It seems that the second supplier may be better but this is only a guess. How sure are we?
5.7.1 Confidence intervals for $p_1 - p_2$
Recall from the previous chapter that if $n_1$ and $n_2$ are large then

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \approx N(0,1)$$

Also since $\hat{p}_1 \approx p_1$ and $\hat{p}_2 \approx p_2$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{S_{\hat{p}_1-\hat{p}_2}} \approx N(0,1)$$

where

$$S_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Therefore

$$P \left( -Z_{\alpha/2} \leq \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{S_{\hat{p}_1-\hat{p}_2}} \leq Z_{\alpha/2} \right) = 1 - \alpha$$

$$\Rightarrow P(\hat{p}_1 - \hat{p}_2 - Z_{\alpha/2}S_{\hat{p}_1-\hat{p}_2} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2}S_{\hat{p}_1-\hat{p}_2}) = 1 - \alpha$$

Therefore

A 100(1 - $\alpha$)% confidence interval for $p_1 - p_2$ is

$$[\hat{p}_1 - \hat{p}_2 - Z_{\alpha/2}S_{\hat{p}_1-\hat{p}_2}, \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2}S_{\hat{p}_1-\hat{p}_2}]$$

(5.8)

Example one continued
In order to calculate a confidence interval we need to compute $S_{\hat{p}_1-\hat{p}_2}$

$$S_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.0417 \times 0.9583}{120} + \frac{0.0278 \times 0.9722}{180}} = 0.0220$$

Therefore a 90% confidence interval for $p_1 - p_2$ is

$$[\hat{p}_1 - \hat{p}_2 - Z_{0.05}S_{\hat{p}_1-\hat{p}_2}, \hat{p}_1 - \hat{p}_2 + Z_{0.05}S_{\hat{p}_1-\hat{p}_2}]$$

$$= [0.0417 - 0.0278 - 1.645 \times 0.0220, 0.0417 - 0.0278 + 1.645 \times 0.0220]$$

$$= [-0.022, 0.0501]$$

Since this interval contains zero it is possible that both suppliers produce the same proportion of defective products (or even that supplier 1 is better!)

Example two
We take a random sample of 500 women and 400 men and ask them “Do you support tougher gun controls?” Suppose 55% of women and 50% of men say yes. We want to estimate the difference in
proportions (if any) between the male and female populations. Here \( \hat{p}_w = 0.55 \) and \( \hat{p}_m = 0.5 \) so our best estimate for the difference is
\[
\hat{p}_w - \hat{p}_m = 0.55 - 0.5 = 0.05
\]

Also
\[
S_{\hat{p}_w - \hat{p}_m} = \sqrt{\frac{0.55 \times 0.45}{500} + \frac{0.5 \times 0.5}{400}} = 0.0335
\]

Therefore a 95\% confidence interval for \( p_1 - p_2 \) is
\[
[\hat{p}_w - \hat{p}_m - Z_{0.025}S_{\hat{p}_w - \hat{p}_m}, \hat{p}_w - \hat{p}_m + Z_{0.025}S_{\hat{p}_w - \hat{p}_m}]
= [0.05 - 1.96 \times 0.0335, 0.05 + 1.96 \times 0.0335]
= [-0.0156, 0.1156]
\]

Therefore there may in fact be no difference between genders on this issue.
Chapter 6

Lies, Damn Lies And Statistics;
Convincing People That What You Say Is True

This chapter deals with the concept of “Hypothesis Tests”. Students often find this topic confusing so we will begin by spending some time on the intuition behind them before we move onto the “formulas”. We will use Case 18 as a motivating example and return to it later in the chapter.

Case 18  The boy who cried wolf

One day you are sitting in your office when Honest Jim walks in. You are a little taken aback because you have heard some rumors about the accuracy of his claims. He is very agitated. He has developed a diet pill which he is calling “Honest Jim’s really good for you diet pills”. However he is having a hard time convincing people that they really work.

He admits that he probably exaggerated their effect when he claimed that the average weight loss on the pills was 25 pounds. However he is certain that they do cause some weight loss. He took a random sample of 16 people taking the pill and found that the average weight loss was 20 pounds with a sample standard deviation of $s = 10$ pounds. He wants to know what sort of analysis he could do to convince people that the drug really works.

6.1 What is a Hypothesis Test?

The basic idea behind hypothesis testing is the following. You have some theory that tells you what the value of a population parameter (e.g., the population mean) should be. You want to decide whether that theory is correct or not. In other words there are two possible situations. You make a decision as to which is correct based on the available data. The following are examples of hypothesis testing situations:

Example one
I have a coin which I claim is fair. I toss the coin 100 times and you observe the number of heads.
Based on the observed number of heads, you decide whether or not my coin is really fair. This is a hypothesis testing situation because there are two “hypothoses” (“the coin is fair” or “the coin is not fair”) and we are trying to decide which is correct. If \( p \) is the probability that the coin lands heads we can write the two hypotheses out mathematically i.e.

\[
p = 0.5 \quad \text{(the coin is fair)}
\]
\[
\text{vs} \quad p \neq 0.5 \quad \text{(the coin is not fair)}
\]

It is important that you are comfortable expressing the hypotheses both in English and mathematically. The English is required to explain to a “non expert” what you are doing and the mathematics is required so that you can actually perform the test (i.e. decide which hypothesis is correct). Speaking of which how might we decide which is correct? That of course is where the statistics comes in.

If the coin is truly fair, you expect to see around 50 heads. That is about half the tosses should come up heads. If you observed 51 heads, or 48, it probably wouldn’t make you suspicious about the fairness of my coin. However, if I got 100 heads in a row, you definitely wouldn’t believe the coin was fair. Why not? Because, if the coin is fair, I am extremely unlikely to get 100 heads, whereas 48 heads or 51 heads is a perfectly reasonable number. The difference from 50 is probably just random fluctuation. This is the basic idea of hypothesis testing. Start with a theory e.g. the coin is fair. Look at some data. If the data are consistent with the theory, you accept the theory. If the data are very unlikely assuming the theory is true, you reject the theory.

Example two
Management wants to improve the speed with which workers on an assembly line perform a certain task. They take a random sample of \( n = 25 \) workers, and record how long it takes the workers to perform the task. Then they administer a training program and measure the time to complete the task again. Here the two hypotheses are that “the training program does not help” (or makes things worse) vs “the training program reduces the time required”. If \( \mu_1 \) and \( \mu_2 \) are respectively the average times required before and after the program we can write these hypotheses mathematically as

\[
\mu_1 \leq \mu_2 \quad \text{(the training program does not help or makes things worse)}
\]
\[
\text{vs} \quad \mu_1 > \mu_2 \quad \text{(the training program does help)}
\]

We will decide that the training helps if “on average” workers take a “lot less” time to perform the task after training. Again we will use statistics to tell us what a “lot less” means. Note that this situation is a bit different from Example one. There either a very high or very low value discredited the original theory. Here only a very low value (i.e big reduction in time) convinces you that the training helps.

There are many types of hypothesis testing situations. Phrases which should tip you off include “test the hypothesis at level \( \alpha \ldots \)”, “is the data consistent with \ldots \”, “does treatment A work better than treatment B”, “is group A different from group B...” etc.
6.2 Establishing the Hypotheses

In the next few sections we will discuss the various steps that are required to actually perform the test i.e. make a final decision. The first step is to identify the two alternatives. For instance in the coin example, the alternatives were that the coin was fair or that it wasn’t. In the assembly line example, training either helped or didn’t help.

6.2.1 Deciding on the Null and Alternative

The next step is to decide which of these alternatives is the “Null” and which the “Alternative”. Generally, one of the two alternatives is more interesting or important than the other. First we choose the null hypothesis. This is denoted $H_0$. There are several ways to decide which should be the null hypothesis. It can be

- The thing you want to disprove.
- The status quo—that is that nothing has changed from the past.
- Your default position—i.e. the thing you would assume unless someone provided strong evidence to the contrary.
- The less interesting or important situation, or the one that does not require taking any action.

The other hypothesis is called the alternative hypothesis and is denoted $H_A$. There are also several ways to characterize the alternative. It can be

- The thing you want to prove.
- That what was true in the past is no longer true.
- That something interesting or important or requiring action has occurred.

Note that you start by assuming the null hypothesis, the less interesting or important hypothesis, is true! We either reject $H_0$ i.e. decide that $H_0$ is not true, or fail to reject $H_0$ i.e. decide we don’t have enough evidence to change our minds. This may seem backwards if what you want to prove is that the alternative hypothesis is right. The reason for doing this is that it is much easier to prove something wrong than to prove something right. A single counter-example can prove a statement is false, but no single example will prove a statement is true.

Here we will illustrate some possible null and alternative hypotheses.

Example one

Suppose we have a food testing trial. A subject is given three portions of food. Two are identical and the other has a slightly different taste. The subject is then asked to identify which is the different food. Let $p$ be the probability that the person is correct for a given test (or the average proportion of the time they are correct). Then there are two possibilities, $p = 1/3$ (i.e. the person is randomly guessing) and $p > 1/3$ (i.e. the person has some ability to tell the difference). If we are interested in proving that the person has some ability our hypotheses would be

\[
H_0 : p = 1/3 \\
H_A : p > 1/3
\]
Example two
The state of California has mandated that average smog levels in Los Angeles be under 100 units by the year 2000. The city measures smog levels at various points and times to see if they are in compliance with the mandate. \( \mu \) is the average smog level. If \( \mu \leq 100 \) the city is in compliance. Otherwise, if \( \mu > 100 \) the city will need to implement new smog reduction techniques. In the past the city has not been in compliance but it now believes it is and wants to prove that it is meeting the standards. Therefore, the null hypothesis is that they are not meeting the standards and the alternative is that they are meeting the standards. In symbols, we would have

\[
H_0 : \mu \geq 100 \\
H_A : \mu < 100
\]

If a government inspector was trying to prove that the city was in violation of the policy, the roles of the two hypotheses would be switched.

Example three
Suppose we are not sure whether a new cold drug is helping or not. We give the drug to one group and a “placebo” to another group. We then measure the number of days they are sick with a cold during one year. The company is interested in proving that the drug is helping. If \( \mu_1 \) and \( \mu_2 \) are the average number of days sick on the drug and placebo respectively, then

\[
H_0 : \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0 \\
H_A : \mu_1 < \mu_2 \text{ or } \mu_1 - \mu_2 < 0
\]

If it is possible that the drug could make things worse and we are just interested in whether it has any effect (i.e. good or bad) we would have

\[
H_0 : \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0 \\
H_A : \mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0
\]

Example four
A company produces computer chips. A certain percentage, \( p \), of their chips are defective. In the past \( p \) has been 5% i.e. 5% of the chips have been defective. However the company is concerned that the machinery may be out of adjustment which would mean that \( p > 5\% \). If the machinery is running properly, nothing needs to be done. Since \( p \) has been 5% in the past the null hypothesis is that the machines are working fine (\( p = 5\% \)), and the alternative is that they are not (\( p > 5\% \)). In symbols, we have

\[
H_0 : p = .05 \\
H_A : p > .05
\]

Note: if the company was trying to prove to an inspector that its process was working properly then the null and alternative hypotheses would be reversed. What you need to establish plays a big role in determining which is the null and which is the alternative.

Example five
A company produces parts at two different plants. Historically both plants have produced the same proportion of defective products. The company wishes to test whether this is still the case.
Let the proportion of defectives from the two plants be \( p_1 \) and \( p_2 \) respectively. Then the null and alternative hypotheses will be

\[
H_0 : \quad p_1 = p_2 \quad \text{or} \quad p_1 - p_2 = 0
\]
\[
H_A : \quad p_1 \neq p_2 \quad \text{or} \quad p_1 - p_2 \neq 0
\]

### 6.2.2 Some more examples

**Example:** In an election, a candidate wins if they get more than 50% of the vote. Suppose a poll was taken before an election to see what proportion of people supported Candidate A. The value of special interest to Candidate A is \( p = .5 \). Specifically Candidate A wants to know whether \( p \leq .5 \)-i.e. the race is a dead heat, or he is losing-or whether \( p > .5 \)-i.e. he is going to win.

In this case the candidate wants to establish that he or she is ahead. Therefore the null hypothesis is that he or she is behind, and the alternative is that he or she is ahead. In symbols, we have \( H_0 : \quad p \leq .5 \) versus \( H_A : \quad p > .5 \).

**Example:** The FDA requires that canned food contain fewer than 5 micrograms of toxic substances. To see whether a company is in compliance with the regulations, the FDA tests 100 cans for toxic materials. They want to determine whether \( \mu \leq 5 \)-i.e. the company is in compliance or whether \( \mu > 5 \) in which case the company will be fined and the product removed from the shelves. The company wants to prove that it is meeting the standards. Therefore, the null hypothesis is that they are not meeting the standards and the alternative is that they are meeting the standards. In symbols, we have \( H_0 : \mu \geq 5 \) versus \( H_A : \mu < 5 \). If a government inspector was trying to prove that the company was in violation of the policy, the roles of the two hypotheses would be switched.

**Example:** In the past, the average household has purchased 5.5 quarts of laundry detergent per year. A government board which monitors consumption of various products wants to know if the amount of laundry detergent used by Americans has changed in the last 20 years. The board wishes to test whether \( \mu = 5.5 \)-that is consumption levels have remained unchanged—or whether \( \mu \neq 5.5 \)-i.e. consumption levels have changed. The null hypothesis is that the status quo is being maintained—people still buy an average of 5.5 quarts of laundry detergent. The alternative is that detergent consumption has changed. In symbols, \( H_0 : \mu = 5.5 \) versus \( H_A : \mu \neq 5.5 \).

### 6.3 Performing the Hypothesis Test

We now know how to decide what our two hypotheses are i.e. the null and alternative. We are finally ready to actually perform a test i.e. decide which hypothesis we believe. There are two different ways to imagine conducting the test. We will illustrate the two approaches through an example.

**Example two (continued)**

Suppose the pollution level in LA is sampled on 100 random days and the sample mean \( X \) is 93 with a sample standard deviation of \( S = 25 \). Recall we wish to decide between

\[
H_0 : \quad \mu \geq 100
\]
\[
H_A : \quad \mu < 100
\]

We start by assuming that \( H_0 \) is true so we will only reject it if we have clear evidence it is false.
What are the possible values for $\mu$? One approach is to calculate a 95\% confidence interval i.e.  
$$[X - Z_{0.025}S/\sqrt{n}, X + Z_{0.025}S/\sqrt{n}] = [93 - 1.96 \times 25/\sqrt{100}, 93 + 1.96 \times 25/\sqrt{100}] = [88.1, 97.9]$$
This interval does not cover $\mu = 100$. Recall we are 95\% sure that the interval is correct or in other words we are 95\% sure that $\mu < 100$. Therefore we reject $H_0$ on the basis of this interval.

A second approach is to calculate the probability that we would see $X$ this low if $\mu = 100$ i.e.  
$$P(X \leq 93 | \mu = 100) = P\left( \frac{X - 100}{25/\sqrt{100}} \leq \frac{93 - 100}{25/\sqrt{100}} | \mu = 100 \right) \approx P(Z \leq -2.8) = 0.0026$$
This probability is very low. It says that if $\mu = 100$ then we would only see a number this low 1 in every 400 times. This tells us that $\mu$ must be less than 100 because we DID SEE $X = 93$. Again we would reject $H_0$. This probability is called a p-value. When we conduct a hypothesis test we will always be calculating the p-value and rejecting $H_0$ if it is small.

### 6.4 Type I and Type II Errors

In the previous section we illustrated the basic idea for performing a hypothesis test. That is, decide whether the data is consistent with the null hypothesis being true and if it is not then reject $H_0$. What do we mean by consistent? In the above example we showed two different ways that the data was not consistent with $H_0$. First we calculated a confidence interval which did not include $H_0 : \mu = 100$. This indicated the data was inconsistent with $\mu = 100$. Second we calculated the probability of observing $X$ as low as 93 if $\mu$ really was 100. This probability was extremely low so that again the data was inconsistent with $\mu$ being 100.

In this example it was easy to decide to reject $H_0$. The probability was only 0.0026 so we were almost certain that $\mu$ was not 100. However the decision is not always that clear cut. Suppose for example the probability had been 0.1 or 0.2. Those are still fairly low probabilities but not completely inconsistent with $H_0$. In other words they suggest that $\mu = 100$ may be unlikely but certainly not impossible. How low does the probability have to be before we reject $H_0$? In order to answer that question we need to understand the “consequences of our actions” i.e. what happens if we are wrong.

#### 6.4.1 Some possible mistakes

First off how could we be wrong? In fact there are two ways that we could make a mistake. The table below illustrates the four possible situations.

<table>
<thead>
<tr>
<th>Truth</th>
<th>$H_0$</th>
<th>$H_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>OK</td>
<td>Type I error</td>
</tr>
<tr>
<td>$H_A$</td>
<td>Type II error</td>
<td>OK</td>
</tr>
</tbody>
</table>

1. We could fail to reject $H_0$ and $H_0$ is in fact the truth. This is good i.e. we have not made a mistake.
2. We could reject $H_0$ when in fact $H_0$ is the truth. This is bad i.e. we have made a mistake. This sort of mistake is called a “Type I error”.

3. We could fail to reject $H_0$ when in fact $H_0$ is not the truth. This is also bad. This is called a “Type II error”.

4. We could reject $H_0$ when in fact $H_0$ is not true. This is good.

Therefore it is the Type I and II errors that we are concerned about.

**Definition 37** We define $\alpha$ (alpha) and $\beta$ (beta) as

$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{ is true})$$

$$\beta = P(\text{Type II error}) = P(\text{Fail to reject } H_0 | H_0 \text{ is false})$$

Both Type I and II errors are a problem so we would like to find a hypothesis test that had $\alpha = \beta = 0$ i.e. we never make a mistake. It is easy to make $\alpha = 0$ i.e. just have a procedure that never rejects $H_0$. Unfortunately this makes $\beta = 1$! On the other hand it is easy to make $\beta = 0$ i.e. just have a procedure that always rejects $H_0$. Unfortunately this makes $\alpha = 1$! In general as we make $\alpha$ smaller we will make $\beta$ larger so it is not easy to make both of them small. (Note that we define the “power” of a test as $1 - \beta$ so a test with high power is good.)

### 6.4.2 Choosing $\alpha$ and $\beta$

Generally a Type I error is considered a bigger problem than Type II error so we find a method that makes $\alpha$ small (e.g. $\alpha = 0.05$) and then try to make $\beta$ as small as possible subject to this constraint. Exactly how small we make $\alpha$ depends on the consequences of a Type I error. Obviously if someone’s life depends on not making a Type I error (as happens in medical studies) we would make $\alpha$ very small (perhaps 1% or smaller). A common value for $\alpha$ is 5% but any value less than or equal to 10% is possible. The value that we choose to set $\alpha$ equal to is called the “significance level” of the test.

**Definition 38** We define the significance level of the hypothesis test as the level that we set $\alpha$ equal to. In other words the probability of making a Type I error. So for example if we were willing to accept a 5% chance of a Type I error we would set $\alpha = 0.05$ and the significance level for the test would be 5%.

### 6.4.3 The relationship between $\alpha$ and the p-value

**Definition 39** The p-value (denoted by $p^*$) is the probability of $X$ taking on a value as extreme or more extreme than the observed value of $X$ if the null hypothesis is true. It is a measure of how likely it is we would observe this $X$ if $H_0$ is true. Therefore if it is low we suspect that $H_0$ is not true. If it is high there is no reason to believe that it is not true.

Notice that there is a relationship between the probability that we calculated in the Smog example (the p-value) and $\alpha$. Both probabilities are calculated assuming that $H_0$ is true. Suppose we only reject $H_0$ when the p-value we calculate is less than 5%, then there is only a 5% chance of us rejecting $H_0$ when it is true (i.e. making a Type I error). This means that $\alpha = 5\%$. On the other hand if we only reject $H_0$ when the p-value is less than 1% then $\alpha = 1\%$. Therefore we should only reject $H_0$ if the p-value is less than $\alpha$. 
We will reject $H_0$ if and only if
\[ p^* < \alpha \] (6.1)

6.5 The Formal Steps in Conducting a Hypothesis Test

Now that we have gone through all the important steps for a hypothesis test we will summarize them.

1. Decide on the two possible hypotheses and which will be the null and which the alternative. You need to be able to write these both in English and mathematically.

2. Decide on the significance level, $\alpha$, for the test.

3. Calculate the p-value, $p^*$, based on the available data.

4. Reject $H_0$ if
\[ p^* < \alpha \]
otherwise fail to reject $H_0$.

We will illustrate these steps through an example.

Example one
Suppose a company is marketing a new cold drug. They wish to prove that the drug reduces the average number of days that people are ill with a cold. Without the drug people are ill with a cold on average 5 days per year. The company gives the drug to a random sample of 9 people and records the number of days they are ill from a cold. They are

3, 2, 4, 5, 4, 5, 6, 6

From these numbers we see that the sample mean $\bar{X}$ is 4.33 and the sample standard deviation $S$ is 1.323.

Step one
First we must decide what the null and alternative hypotheses are. If $\mu$ is the average number of days ill we know that in the past $\mu$ has been 5. Therefore our null hypothesis will be “no change” i.e. $\mu = 5$. The company wishes to prove that the drug is helping i.e. that for people taking the drug $\mu$ has decreased. Therefore the alternative hypothesis should be $\mu < 5$.

\[
H_0 : \mu = 5 \\
H_A : \mu < 5
\]

Step two
We can choose any significance level that we like. We will set $\alpha = 0.05$. 

Step three
Next we calculate the p-value. That is the probability of observing \( X \) as low or lower than 4.33 given that \( \mu = 5 \).

\[
p^* = P(X \leq 4.33 | \mu = 5)
\]

\[
= P\left( \frac{X - 5}{1.323/\sqrt{9}} \leq \frac{4.33 - 5}{1.323/\sqrt{9}} | \mu = 5 \right)
\]

\[
= P(t_s \leq -1.511)
\]

\[
= 0.0846
\]

The last step in the probability calculation requires a computer program such as JMP to calculate the exact probability. However using the \( t \) tables we can see that \( 0.05 \leq p^* \leq 0.1 \).

Step four
The last step is very easy. Since

\[
0.0846 = p^* > .05 = \alpha
\]

we fail to reject \( H_0 \). In other words the data does not provide enough evidence for us to be certain that \( \mu < 5 \). It is important that you can also explain your conclusion in English! We are now ready to return to Case 18.

Case 18 Revisited
Step one
Let \( \mu \) be the average weight loss by people taking the pill. Here Honest Jim wants to prove that the pill is helping so this would be the alternative hypothesis i.e. \( H_A : \mu > 0 \). The null hypothesis would be that the pill does not help i.e. \( H_0 : \mu = 0 \).

\[
H_0 : \mu = 0
\]

\[
H_A : \mu > 0
\]

Step two
We will choose \( \alpha = 0.01 \). Again there is nothing special about this number.

Step three
Now we need to calculate the p-value, \( p^* \). Here \( X \) was 20 so we want to calculate the probability of getting \( X \) this large or larger given that \( \mu = 0 \).

\[
p^* = P(X \geq 20 | \mu = 0)
\]

\[
= P\left( \frac{X - 0}{10/\sqrt{16}} \geq \frac{20 - 0}{10/\sqrt{16}} | \mu = 0 \right)
\]

\[
= P(t_{15} \geq 8)
\]

\[
= 0.0000004
\]

Again a computer package is required to get the exact probability but from the tables we see that \( p^* < 0.005 \).

Step four
\( p^* < \alpha \) so we reject \( H_0 \) and conclude that \( H_A \) is true. In other words we believe that the pills are
causing a weight loss.

Notice that the value of \( \alpha \) was not really very important for this test. For any reasonable value of \( \alpha \) we would have rejected \( H_0 \). This illustrates another use of the p-value i.e. it tells us exactly how sure we are about our decision. If \( p^* \) was 0.008 we would still reject \( H_0 \) but we would not be nearly as sure about our decision as in this case.

### 6.6 Hypotheses on \( \mu \)

It should be clear from the previous examples that the only difficult parts in performing a hypothesis test are to choose the null and alternative hypotheses and to calculate the p-value. The choice of the hypotheses and calculation of the p-value depend on the situation. We will spend the rest of this chapter examining some of them. There are many possible situations in which you may wish to perform a hypothesis test. We will only cover some of the more common ones in this course. Probably the most common situation in which you would want to perform a hypothesis test is on the population mean \( \mu \).

#### 6.6.1 \( H_0 : \mu = \mu_0 \)

Suppose we have a random sample \( X_1, X_2, \ldots, X_n \) from a normal population i.e. \( N(\mu, \sigma^2) \) and we want to test \( H_0 : \mu = \mu_0 \) (where \( \mu_0 \) is just some number e.g. 100) vs an alternative. Then there are three common alternatives you may be interested in.

1. \( H_A : \mu > \mu_0 \) (one sided alternative)
2. \( H_A : \mu < \mu_0 \) (one sided alternative)
3. \( H_A : \mu \neq \mu_0 \) (two sided alternative)

The first two are called “one sided alternatives” because we are only interested in values of \( \mu \) in one direction (i.e. \( \mu > \mu_0 \) or \( \mu < \mu_0 \)). The last is called a “two sided alternative” because we are interested in values of \( \mu \) in both directions (i.e. \( \mu \neq \mu_0 \)). We will examine separately the three possible hypotheses. We will use the drug testing trial to illustrate the calculations. Recall \( X = 4.33 \) and \( \mu_0 = 5 \).

\( H_A : \mu > \mu_0 \)

For this hypothesis we are interested in large values of \( X \) because that will provide evidence for \( H_A \). Hence our p-value will be

\[
p^* = P(X \geq 4.33 | \mu = 5) = P \left( \frac{X - 5}{1.323/\sqrt{9}} \geq \frac{4.33 - 5}{1.323/\sqrt{9}} \bigg| \mu = 5 \right) = P(t_s \geq -1.511) = 0.9154
\]

Notice that for this alternative hypothesis the p-value is very large. We will definitely not reject \( H_0 \). Here

\[
t_{obs} = \frac{4.33 - 5}{1.323/\sqrt{9}} = -1.511
\]

is called the “observed test statistic”. Notice that all we really need to do is compute the observed test statistic, \( t_{obs} \), and then calculate

\[
p^* = P(t_s \geq t_{obs}) = P(t_s \geq -1.511)
\]
6.6. HYPOTHESES ON $\mu$

So in general we will use the following formula to calculate the p-value.

If the alternative hypothesis is $H_A : \mu > \mu_0$ then

$$p^* = P(t_{n-1} \geq t_{obs}) \quad \text{where} \quad t_{obs} = \frac{X - \mu_0}{S/\sqrt{n}} \quad (6.2)$$

$H_A : \mu < \mu_0$

For this hypothesis we are interested in low values of $X$ because that will provide evidence for $H_A$. Hence our p-value will be

$$p^* = P(X \leq 4.33|\mu = 5) = P\left(\frac{X - 5}{1.323/\sqrt{9}} \leq \frac{4.33 - 5}{1.323/\sqrt{9}} | \mu = 5\right) = P(t_s \leq -1.511) = 0.0846$$

Again

$$t_{obs} = \frac{4.33 - 5}{1.323/\sqrt{9}} = -1.511$$

is called the “observed test statistic”. Notice that all we really need to do is compute the observed test statistic, $t_{obs}$, and then calculate

$$p^* = P(t_s \leq t_{obs}) = P(t_s \leq -1.511)$$

So in general we will use the following formula to calculate the p-value.

If the alternative hypothesis is $H_A : \mu < \mu_0$ then

$$p^* = P(t_{n-1} \leq t_{obs}) \quad \text{where} \quad t_{obs} = \frac{X - \mu_0}{S/\sqrt{n}} \quad (6.3)$$

$H_A : \mu \neq \mu_0$

For this hypothesis we are interested in low values and high values of $X$ because either will provide evidence for $H_A$. So we want the probability of getting a value of $X$ as far or further away from 5. Hence not only do we want $P(X \leq 4.33)$ but also $P(X \geq 5.67)$. Hence the p-value is

$$p^* = P(X \leq 4.33|\mu = 5) + P(X \geq 5.67|\mu = 5)$$

$$= P\left(\frac{X - 5}{1.323/\sqrt{9}} \leq \frac{4.33 - 5}{1.323/\sqrt{9}} | \mu = 5\right) + P\left(\frac{X - 5}{1.323/\sqrt{9}} \geq \frac{5.67 - 5}{1.323/\sqrt{9}} | \mu = 5\right)$$

$$= P(t_s \leq -1.511) + P(t_s \geq 1.511)$$

$$= 2P(t_s \geq |t_{obs}|) = 0.1692$$

Notice that the only difference that this makes to the calculation is that we use $|t_{obs}|$ and multiply the probability by 2. So in general we will use the following formula to calculate the p-value.
If the alternative hypothesis is \( H_A : \mu \neq \mu_0 \) then

\[
p^* = 2P(t_{n-1} \geq |t_{cbs}|) \quad \text{where} \quad t_{cbs} = \frac{X - \mu_0}{S/\sqrt{n}} \tag{6.4}
\]

**Example one**

It is often thought that companies that have been achieving relatively poor returns on shareholders’ capital are those most likely to attract takeover bids. The accounting measure “abnormal returns” standardizes the rate of return so that it averages zero over all companies. Thus a company with poorer than average return on investment has a negative value of “abnormal returns”. Krummer and Hoffmeister [(1978) Journal of Finance, 33, p565-516] report on a random sample of 88 businesses that had attracted takeover bids. For returns reported prior to the bids, the mean “abnormal return” for the companies was \(-0.0029\) and the standard deviation was 0.0169. Perform a hypothesis test to see if the initial conjecture is true.

Let \( \mu \) be the average abnormal return among companies that have attracted a takeover bid. We wish to prove that \( \mu < 0 \). Therefore \( H_A : \mu < 0 \). The null hypothesis will be that there is no difference from other companies so \( H_0 : \mu = 0 \).

\[
H_0 : \mu = 0 \quad H_A : \mu < 0
\]

This tells us that we need to use equation (6.3) with \( \mu_0 = 0 \) to calculate the p-value. From the question we see that \( X = -0.0029 \) and \( S = 0.0169 \) so

\[
t_{cbs} = \frac{X - \mu_0}{S/\sqrt{n}} = \frac{-0.0029 - 0}{0.0169/\sqrt{88}} = -1.610
\]

and the p-value is

\[
p^* = P(t_{87} < -1.610) = 0.0555
\]

Again you would have to use a computer to calculate this probability exactly. However since \( n \) is so large \( t_{87} \approx Z \) so we could estimate the p-value as

\[
p^* = P(t_{87} < -1.610) \approx P(Z < -1.610) = 0.0537
\]

which you can see is very close to the exact probability. (This approximation is valid provided \( n \) is greater than about 30.) Therefore we would reject \( H_0 \) provided \( \alpha > 0.0555 \). To explain this result in English to someone that knows nothing about hypothesis testing you would probably say that there is marginal evidence to support the conjecture.

Suppose instead of conjecturing that only companies with bad performances attracted takeover bids we also conjectured that companies with very good performances might attract takeover bids. How would this change our analysis? In this case \( H_A : \mu \neq 0 \) rather than \( H_A : \mu < 0 \). So we get

\[
H_0 : \mu = 0 \quad H_A : \mu \neq 0
\]
This means that we have a two sided hypothesis test so we need to use equation (6.4) to calculate the p-value. Note that we still have \( t_{\text{obs}} = -1.61 \) so

\[
p^* = 2P(t_{s7} > |t_{\text{obs}}|) = 2P(t_{s7} > 1.610) = 2 \times 0.0555 = 0.111
\]

This tells us that even if \( \alpha \) is as large as .10 we will fail to reject \( H_0 \). This is only weak evidence for our new conjecture.

### 6.6.2 \( H_0 : \mu \leq \mu_0 \) or \( H_0 : \mu \geq \mu_0 \)

Sometimes it may make more sense to use the null hypothesis \( H_0 : \mu \leq \mu_0 \) or \( H_0 : \mu \geq \mu_0 \) rather than \( H_0 : \mu = \mu_0 \). For example in the LA smog example our null hypothesis was \( H_0 : \mu \geq 100 \). How do we calculate a p-value for this situation? It turns out that we use exactly the same formula as for \( H_0 : \mu = 100 \). This is because we are always giving the benefit of the doubt to \( H_0 \). If \( X = 93 \) then the p-value is

\[
p^* = P(X \leq 93 | \mu)
\]

The larger that \( \mu \) is the smaller the probability will be e.g.

\[
P(X \leq 93 | \mu = 100) = P \left( Z \leq \frac{93 - 100}{25/\sqrt{100}} \right) = 0.0026
\]
\[
P(X \leq 93 | \mu = 101) = P \left( Z \leq \frac{93 - 101}{25/\sqrt{100}} \right) = 0.0007
\]
\[
P(X \leq 93 | \mu = 102) = P \left( Z \leq \frac{93 - 102}{25/\sqrt{100}} \right) = 0.0002
\]

Yet we know that if \( H_0 \) is true \( \mu \geq 100 \). Therefore we make \( \mu \) as small as possible i.e. \( \mu = 100 \) so that we get as large a p-value as possible.

To summarize.

<table>
<thead>
<tr>
<th>If ( H_0 : \mu \geq \mu_0 ) or ( H_0 : \mu \leq \mu_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>calculate the p-value as if ( H_0 : \mu = \mu_0 )</td>
</tr>
</tbody>
</table>

### 6.7 Hypotheses on \( p \)

The next possible situation we may be interested in is performing a hypothesis test on \( p \), the proportion from a binomial distribution. Case 19 provides a motivating example.
Case 19 To invest or not? Part 2

In this case we are going to reexamine Case 17 using hypothesis testing. Recall in Case 17 we had the following setup:

Proposition 1029 has been put on the ballot. This proposition would ban all commercial felling of trees in California. United Tree Eaters holds 80% of all the felling rights in California. Obviously if this proposition passes the company will be devastated. However at the moment it is running very well so if the proposition fails to pass they will remain profitable.

Despite the severity of Proposition 1029, support for it is running close to 50%. The shareholders of United Tree Eaters are obviously nervous and the share price has dropped a great deal in the lead up to the election. If the proposition fails to pass the share price will return to normal levels and a large profit could be made. If not the share price will plummet and a large loss could be made. You decide to use your statistics training to see if you should invest in the company.

You take a random sample of 1000 eligible voters and ask them if they will vote for Proposition 1029. 540 or 54% say they will vote against it.

How would we use a hypothesis test to help us make a decision?

Let's try a simpler example before we attempt to answer the case study.

Example one
Suppose our old supplier of parts for our production process produced 3% defective parts. We receive a batch of 500 parts from a new supplier, which we hope is better, and find 10 defective parts. Do we have enough evidence to conclude that the new supplier is better?

$H_A : p < p_0$

First we must decide on the null and alternative hypotheses. In fact this situation is very similar to Example 4 in Section 6.2.1. Let $p$ be the long run proportion of defective parts from our new supplier. We are hopeful that the new supplier is better so the alternative hypothesis will be $H_A : p < 0.03$ and the null hypothesis will be that nothing has changed i.e. $H_0 : p = 0.03$.

$$H_0 : p = 0.03$$
$$H_A : p < 0.03$$

The next step is to calculate the p-value. We know how to calculate p-value for means ($\mu$) but how do we do it for proportions? The basic idea is identical. We still want to calculate the probability of our data being as extreme or more extreme given $H_0$ is true. The only difference is that instead of using $X$ we use $\hat{p}$. In this example $\hat{p} = 10/500 = 0.02$. The lower $\hat{p}$ is the more extreme the evidence against $H_0$. Therefore the p-value is

$$p^* = P(\hat{p} \leq 0.02|p = 0.03)$$
6.7. HYPOTHESES ON P

How do we calculate this probability? Recall from Chapter 4 that if $n$ is large then

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0,1)$$

Therefore

$$p^* = P(\hat{p} \leq 0.02|p = 0.03)$$
$$= P\left(\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \leq \frac{0.02 - 0.03}{\sqrt{0.03(1-0.03)/500}}\right)$$
$$= P(Z \leq -1.31)$$
$$= 0.0951$$

This is a fairly large p-value so there is only weak evidence to reject the null hypothesis i.e. there is only weak evidence that the new supplier is better.

$$Z_{obs} = \frac{0.02 - p_0}{p_0(1-p_0)/n} = \frac{0.02 - 0.03}{0.03(1-0.03)/500} = -1.31$$

is called the “observed test statistic. Notice that all we really need to do is compute $Z_{obs}$ and then calculate

$$p^* = P(Z \leq Z_{obs}) = P(Z \leq -1.31)$$

$H_A : p > p_0$

What if the alternative hypothesis was that the new supplier was worse i.e.

$$H_A : p > 0.03$$

In this case very large values of $\hat{p}$ (rather than very small values) would give evidence for $H_A$ so the p-value would be

$$p^* = P(\hat{p} \geq 0.02|p = 0.03)$$
$$= P\left(\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \geq \frac{0.02 - 0.03}{\sqrt{0.03(1-0.03)/500}}\right)$$
$$= P(Z \geq -1.31)$$
$$= 0.9049$$

This is a very large number so there is certainly no evidence to support this hypothesis. Notice again that all we really needed to do was compute $Z_{obs}$ and then calculate

$$p^* = P(Z \geq Z_{obs}) = P(Z \geq -1.31)$$

$H_A : p \neq p_0$

Finally we may be interested in whether there is any difference (either better or worse) between the suppliers. Than the alternative hypothesis would be

$$H_A : p \neq 0.03$$
In this case very large or very small values of $\hat{p}$ would give evidence for $H_A$ so the p-value would be the probability of being as far or further from 0.03 i.e.

$$p^* = P(\hat{p} \leq 0.02 | p = 0.03) + P(\hat{p} \geq 0.04 | p = 0.03)$$

$$= P \left( \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \leq \frac{0.02 - 0.03}{\sqrt{0.03(1-0.03)/500}} \right) + P \left( \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \geq \frac{0.04 - 0.03}{\sqrt{0.03(1-0.03)/500}} \right)$$

$$= P(Z \leq -1.31) + P(Z \geq 1.31)$$

$$= 2P(Z \geq |Z_{obs}|)$$

$$= 0.1902$$

Now we don’t even have weak evidence to reject $H_0$. Notice again that we just needed to compute $Z_{obs}$ and then calculate

$$p^* = 2P(Z \geq |Z_{obs}|) = 2P(Z \geq 1.31)$$

So in summary for the three possible alternative hypotheses we calculate the p-values in the following way.

$$\text{If } H_A : p < p_0$$

$$p^* = P(Z \leq Z_{obs}) \quad \text{where} \quad Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \quad (6.5)$$

$$\text{If } H_A : p > p_0$$

$$p^* = P(Z \geq Z_{obs}) \quad \text{where} \quad Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \quad (6.6)$$

$$\text{If } H_A : p \neq p_0$$

$$p^* = 2P(Z \geq |Z_{obs}|) \quad \text{where} \quad Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \quad (6.7)$$

**Example two**

For good or ill, statistical evidence has become very important in this country for civil rights and discrimination cases. The Supreme Court has made some important rulings which effect all subsequent trials in lower courts. In *Casteneda vs. Partida* (decided March 1977) the court noted that 79% of the population of Hidalgo County had Spanish sounding last names. Of the 870 jurors selected to jury panels by the county only 339 (approximately 39%) had Spanish sounding last names. Is there evidence of discrimination?

Let $p$ be the overall probability of a person selected for jury service having a Spanish sounding last name. We would like to prove that there is discrimination which would mean $p < 0.79$ so $H_A : p < 0.79$. The null hypothesis is that nothing is going on so $H_0 : p = 0.79$.

$$H_0 : p = 0.79$$

$$H_A : p < 0.79$$
Next we calculate $Z_{obs}$

$$Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.39 - 0.79}{\sqrt{0.79(1 - 0.79)/870}} = -29.0$$

and finally the p-value

$$p^* = P(Z < Z_{obs}) = P(Z < -29.0) = 0$$

We are absolutely certain that we should reject $H_0$ i.e. the probability of a person with a Spanish sounding last name being selected for jury service is much lower than one would expect if they were randomly chosen.

**Case 19 Revisited**

Let $p$ be the true proportion who will vote against the proposition. Here we would like to prove that the proposition will fail so $H_A : p > 0.5$ and the null will be that is passes $H_0 : p \leq 0.5$.

$$H_0 : \quad p \leq 0.5$$
$$H_A : \quad p > 0.5$$

How do we deal with the fact that $H_0 : p \leq 0.5$ rather than $H_0 : p = 0.5$? Just as with hypotheses on $\mu$ we can treat this exactly the same way as $H_0 : p = 0.5$. Therefore $\hat{p} = 540/1000 = 0.54$ and

$$Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.54 - 0.5}{\sqrt{0.5 \times 0.5/1000}} = 2.53$$

and from equation (6.6) the p-value is

$$p^* = P(Z \geq Z_{obs}) = P(Z \geq 2.53) = 0.0057$$

For any value of $\alpha > 0.0057$ we will reject $H_0$. There is very strong evidence that the proposition will fail.

### 6.8 Hypotheses on $\mu_1 - \mu_2$

The third possible situation in which you may wish to perform a hypothesis test is in comparing the means of two different populations. Case 20 (our final case) provides an example.

**Case 20 Lies, Damn Lies and Statistics**

*It is common for companies to make claims about their products e.g. our chip runs twice as fast as their chip or our weed killer kills 20% more weeds than their weed killer. These claims are usually backed up by some study that has been done. However, often the claims are exaggerated or just plain wrong. Therefore it is important that you know the statistics behind them so you can make an informed decision about their accuracy. This case will examine one possible situation.*

*A representative of Healthy Health Foods has come to see you. She is very concerned about the claims that one of their competitors, Honest Jim’s Really Healthy Health Foods, are making. Healthy Health Foods has been producing low fat cookies for many years. They are very popular*
and have a significant proportion of the low fat cookie market. However Honest Jim has recently entered the market and is claiming that his cookies have less fat than those from Healthy Health Foods.

Healthy Health Foods vigorously denies this and claims that the average fat content is identical. Honest Jim has based his claim on a study that he performed where he took a random sample of 32 cookies from Healthy Health Foods and 32 of his own. He measured the fat content for each of the cookies and found that the average fat content in his cookies was 5.5 grams per cookie while the average for Healthy Health Foods was 6.0. His conclusion was that his cookies have less fat on average. You manage to find the original data and calculate that the sample standard deviation for Honest Jim’s cookies is 1.75 and for Healthy Health Foods is 2.25. How would we use a hypothesis test to investigate Honest Jim’s Claim?

Again we will consider an easier example before returning to this case.

**Example one**

There is a problem with the hypothesis test that we performed for the new cold drug. Often when people are given a pill and told that it will make them better they get better even if the pill does nothing! This is called the “placebo effect”. The problem is that it is then very difficult to tell whether the drug is really helping. To get around this problem drug companies often use what is called a “double blind randomized trial”. In this trial a group of people are randomly allocated to one of two groups. One group is given a placebo pill (i.e. a pill that does nothing) and the other is given the drug. However no one is told which they have. Then the two groups are compared. The placebo effect should be the same for both groups so it is then possible to see whether the drug itself is helping.

Suppose that we re perform the drug trial using this strategy. Let $\mu_1$ be the average days sick of people on the new drug and $\mu_2$ be the average days sick of people on the placebo. We wish to prove that the drug is helping so the alternative hypothesis is $H_A : \mu_1 < \mu_2$ and the null hypothesis will be $H_0 : \mu_1 = \mu_2$.

\[
H_0 : \mu_1 = \mu_2 \quad \text{or} \quad H_0 : \mu_1 - \mu_2 = 0 \\
H_A : \mu_1 < \mu_2 \quad \text{or} \quad H_0 : \mu_1 - \mu_2 < 0
\]

Suppose we perform the trial and get the following summary statistics.

\[
X_1 = 5, \quad X_2 = 7, \quad S_1^2 = 8, \quad S_2^2 = 9, \quad n_1 = n_2 = 100
\]

We will assume that we have

1. Independent random samples.
2. Normal populations.
3. Both populations have the same variance.
How do we calculate the p-value. Intuitively it seems like we want to reject $H_0$ if $X_1 - X_2$ is very small. For our experiment we got $X_1 - X_2 = 5 - 7 = -2$ so the p-value will be

$$p^* = P(X_1 - X_2 \leq -2 | \mu_1 - \mu_2 = 0)$$

How do we calculate this probability? Recall that

$$t = \frac{X_1 - X_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

so if $H_0 : \mu_1 - \mu_2 = 0$ is true then

$$t = \frac{X_1 - X_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

In our case

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{99 \times 8 + 99 \times 9}{100 + 100 - 2} = 8.5$$

Therefore

$$p^* = P(X_1 - X_2 \leq -2 | \mu_1 - \mu_2 = 0) = P \left( \frac{X_1 - X_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq \frac{-2}{\sqrt{8.5 \left( \frac{1}{100} + \frac{1}{100} \right)}} | \mu_1 - \mu_2 = 0 \right) = P(t_{198} \leq -4.85) = 0.000001$$

This is an extremely small p-value so we would reject $H_0$ for any reasonable value of $\alpha$. The observed test statistic in this case is

$$t_{obs} = \frac{X_1 - X_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{-2}{\sqrt{8.5 \left( \frac{1}{100} + \frac{1}{100} \right)}} = -4.85$$

so again calculating the p-value involves

$$p^* = P(t_{198} \leq t_{obs}) = P(t_{198} \leq -4.85)$$

We can apply the same reasoning to the other 2 possible alternative hypotheses (i.e. $H_A : \mu_1 - \mu_2 > 0$ or $H_A : \mu_1 - \mu_2 \neq 0$) and we get the following.
CHAPTER 6. HYPOTHESIS TESTS

If $H_A : \mu_1 - \mu_2 < 0$

$$p^* = P(t_{n_1+n_2-2} \leq t_{obs}) \quad \text{where} \quad t_{obs} = \frac{X_1 - X_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.8)$$

If $H_A : \mu_1 - \mu_2 > 0$

$$p^* = P(t_{n_1+n_2-2} \geq t_{obs}) \quad \text{where} \quad t_{obs} = \frac{X_1 - X_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.9)$$

If $H_A : \mu_1 - \mu_2 \neq 0$

$$p^* = 2P(t_{n_1+n_2-2} \geq |t_{obs}|) \quad \text{where} \quad t_{obs} = \frac{X_1 - X_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.10)$$

Example two
Suppose we wished to test

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 \neq \mu_2$$

and got the following summary statistics

$$X_1 = 3, \quad X_2 = 2, \quad S_1^2 = 1.5, \quad S_2^2 = 1, \quad n_1 = 12, \quad n_2 = 10$$

First we calculate

$$S_p^2 = \frac{(12 - 1) \times 1.5 + (10 - 1) \times 1}{12 + 10 - 2} = 1.275$$

Next we calculate

$$t_{obs} = \frac{3 - 2}{\sqrt{1.275 \left( \frac{1}{12} + \frac{1}{10} \right)}} = 2.068$$

Finally we compute the p-value using equation 6.10.

$$p^* = 2P(t_{12+10-2} \geq |t_{obs}|) = 2P(t_{20} \geq 2.068) = 0.0518$$

You would need a computer to calculate this probability exactly but from the tables we see that $0.05 < p^* < .10$. We have only moderate evidence to reject $H_0$. We would only reject $H_0$ if $\alpha > 0.05$.

Case 20 Revisited
Let $\mu_1$ be the average fat content of Healthy Health Foods and $\mu_2$ be the average fat content of Honest Jim’s Really Healthy Health Foods. Honest Jim is trying to claim that his cookies have less
fat i.e. \( \mu_1 > \mu_2 \) so the alternative hypothesis should be \( H_A : \mu_1 - \mu_2 > 0 \). The null hypothesis will be that nothing is going on i.e. \( H_0 : \mu_1 - \mu_2 = 0 \).

\[
H_0 : \mu_1 - \mu_2 = 0 \\
H_A : \mu_1 - \mu_2 > 0
\]

Our summary statistics are

\[
X_1 = 6.0, \quad X_2 = 5.5, \quad S_1 = 1.75, \quad S_2 = 2.25, \quad n_1 = n_2 = 32
\]

First we calculate

\[
S_p^2 = \frac{(32 - 1) \times 1.75^2 + (32 - 1) \times 2.25^2}{32 + 32 - 2} = 4.0625
\]

Next we calculate

\[
t_{obs} = \frac{6.0 - 5.5}{\sqrt{4.0625 \left( \frac{1}{32} + \frac{1}{32} \right)}} = 0.99
\]

Finally we compute the p-value using equation 6.9.

\[
p^* = P(t_{32+32-2} \geq t_{obs}) = P(t_{62} \geq 0.99) = 0.1630
\]

Again you would need a computer to calculate this probability exactly but we can use the normal tables (since \( n \) is large) and we get an approximate p-value of 0.1611 which is pretty close to the exact answer. This p-value is pretty large so we have no real evidence to reject \( H_0 \). In other words Honest Jim has no significant evidence to back up his claim.

### 6.9 Hypotheses on \( p_1 - p_2 \)

The fourth possible situation is where you are comparing the proportions from two different Binomial populations. We will illustrate this situation through an example.

**Example one**

Suppose we have two random samples of parts from two different potential suppliers. We are interested in whether the quality of parts from the first supplier is better than that from the second. Suppose we find 21 defective parts from supplier one and 32 defective parts from supplier two and we had a total of 1000 parts from each supplier. Is the probability of a defective part from supplier one lower than that from supplier two?

First we need to define the null and alternative hypotheses. Let \( p_1 \) be the probability of a defective part from supplier one and \( p_2 \) be the probability of a defective part from supplier two. We are interested in showing that supplier one is better so \( H_A : p_1 < p_2 \) and the null hypothesis will be no difference i.e. \( H_0 : p_1 = p_2 \).

\[
H_0 : p_1 = p_2 \quad \text{or} \quad p_1 - p_2 = 0 \\
H_A : p_1 < p_2 \quad \text{or} \quad p_1 - p_2 < 0
\]

Next we need to calculate the p-value. We estimate \( p_1 \) using \( \hat{p}_1 = 21/1000 = 0.021 \) and \( p_2 \) using \( \hat{p}_2 = 32/1000 = 0.032 \). Therefore we will have strong evidence for \( H_A \) if \( \hat{p}_1 - \hat{p}_2 \) is a very large negative value so the p-value will be

\[
p^* = P(\hat{p}_1 - \hat{p}_2 \leq 0.021 - 0.032|p_1 - p_2 = 0) = P(\hat{p}_1 - \hat{p}_2 \leq -0.011|p_1 - p_2 = 0)
\]
How do we calculate this probability? Recall that if \( n_1 \) and \( n_2 \) are large then

\[
Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{S_{\hat{p}_1 - \hat{p}_2}} \approx N(0, 1)
\]

where

\[
S_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

However if \( H_0 \) is true then \( p_1 = p_2 = p \) so we need to come up with a single estimate for \( p \). We use the following estimate

\[
\hat{p}_{pool} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}
\]

which gives the following estimate for \( S_{\hat{p}_1 - \hat{p}_2} \).

\[
S_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_{pool}(1 - \hat{p}_{pool})}{n_1} + \frac{\hat{p}_{pool}(1 - \hat{p}_{pool})}{n_2}} = \sqrt{\hat{p}_{pool}(1 - \hat{p}_{pool}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}
\]

Hence if \( H_0 \) is true

\[
Z = \frac{\hat{p}_1 - \hat{p}_2}{S_{\hat{p}_1 - \hat{p}_2}} \approx N(0, 1)
\]

In this example

\[
\hat{p}_{pool} = \frac{21 + 32}{1000 + 1000} = 0.0265
\]

and

\[
S_{\hat{p}_1 - \hat{p}_2} = \sqrt{0.0265(1 - 0.0265) \left( \frac{1}{1000} + \frac{1}{1000} \right)} = 0.00718
\]

Therefore the p-value is

\[
p^* = P(\hat{p}_1 - \hat{p}_2 \leq -0.011 | p_1 - p_2 = 0)
\]

\[
= P \left( \frac{\hat{p}_1 - \hat{p}_2}{S_{\hat{p}_1 - \hat{p}_2}} \leq \frac{-0.011}{0.00718} | p_1 - p_2 = 0 \right)
\]

\[
= P(Z \leq -1.53)
\]

\[
= 0.063
\]

This is a fairly large p-value so unless \( \alpha > 0.063 \) we will fail to reject \( H_0 \). We have very marginal evidence that supplier one is better than supplier two. The observed test statistic in this case is

\[
Z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{S_{\hat{p}_1 - \hat{p}_2}} = \frac{0.021 - 0.032}{0.00718} = -1.53
\]

So we could calculate the p-value as

\[
p^* = P(Z \leq Z_{obs}) = P(Z \leq -1.53)
\]

We can use the same approach for the other two possible alternatives (i.e. \( H_A : p_1 - p_2 > 0 \) and \( H_A : p_1 - p_2 \neq 0 \)).
6.9. Hypotheses on $P_1 - P_2$

If $H_A : p_1 - p_2 < 0$

$$p^* = P(Z \leq Z_{obs}) \quad \text{where} \quad Z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{S_{\hat{p}_1 - \hat{p}_2}}$$

(6.13)

If $H_A : p_1 - p_2 > 0$

$$p^* = P(Z \geq Z_{obs}) \quad \text{where} \quad Z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{S_{\hat{p}_1 - \hat{p}_2}}$$

(6.14)

If $H_A : p_1 - p_2 \neq 0$

$$p^* = 2P(Z \geq |Z_{obs}|) \quad \text{where} \quad Z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{S_{\hat{p}_1 - \hat{p}_2}}$$

(6.15)

**Example two**

Suppose that in the previous example we were only interested in whether there was a difference between the suppliers. Then the alternative hypothesis would be $H_A : p_1 - p_2 \neq 0$.

$$H_0 : \quad p_1 - p_2 = 0$$
$$H_A : \quad p_1 - p_2 \neq 0$$

Our observed test statistic is still $Z_{obs} = -1.53$ so the p-value (from equation 6.15) is

$$p^* = 2P(Z \geq |Z_{obs}|) = 2P(Z \geq 1.53) = 2 \times 0.063 = 0.126$$

Now there is only very weak evidence to reject $H_0$. In other words unless $\alpha > 0.126$ we do not have strong enough evidence to conclude that there is a difference between suppliers.

**Example three**

Suppose we wish to test whether there is a difference between men and women on gun control. We survey 500 women and 55% support tougher gun controls. We also survey 400 men and 50% support tougher gun controls. Do we have significant evidence that there is a difference between men and women?

The alternative hypothesis is $H_A : p_w - p_m \neq 0$ and the null is no difference i.e. $H_0 : p_w - p_m = 0$.

$$H_0 : \quad p_w - p_m = 0$$
$$H_A : \quad p_w - p_m \neq 0$$

Next we need to calculate $\hat{p}_{pool}$.

$$\hat{p}_{pool} = \frac{0.55 \times 500 + 0.5 \times 400}{500 + 400} = \frac{275 + 200}{900} = 0.528$$

and

$$S_{\hat{p}_w - \hat{p}_m} = \sqrt{0.528(1 - 0.528) \left( \frac{1}{400} + \frac{1}{500} \right)} = 0.033$$
then the observed test statistic

\[ Z_{\text{obs}} = \frac{\hat{p}_w - \hat{p}_m}{\sqrt{\frac{p_w(1-p_w)}{n_w} + \frac{p_m(1-p_m)}{n_m}}} = 1.49 \]

Finally we need to calculate the p-value

\[ p^* = 2P(Z \geq Z_{\text{obs}}) = 2P(Z > 1.49) = 2 \times 0.0681 = 0.1362 \]

This is fairly large so there is no real evidence of a difference between men and women i.e. for any \( \alpha < 0.136 \) we would fail to reject \( H_0 \).

6.10 Paired Data

6.10.1 Paired vs Independent Data

The last possible situation that we will examine in this course is very similar to that in Section 6.8 where we were examining the difference between two population means. The following example illustrates the differences.

Example one

Suppose we wish to test whether heat treatment reduces the bacteria counts in milk. We record the bacteria counts on 12 bottles of milk before treatment \((X_1)\) and after treatment \((X_2)\).

<table>
<thead>
<tr>
<th>Sample, ( i )</th>
<th>Before treatment ((X_{1i}))</th>
<th>After treatment ((X_{2i}))</th>
<th>Difference (D_i = X_{1i} - X_{2i})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.98</td>
<td>6.95</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>7.08</td>
<td>6.94</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>8.34</td>
<td>7.17</td>
<td>1.17</td>
</tr>
<tr>
<td>4</td>
<td>5.30</td>
<td>5.15</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>6.26</td>
<td>6.28</td>
<td>-0.02</td>
</tr>
<tr>
<td>6</td>
<td>6.77</td>
<td>6.81</td>
<td>-0.04</td>
</tr>
<tr>
<td>7</td>
<td>7.03</td>
<td>6.59</td>
<td>0.44</td>
</tr>
<tr>
<td>8</td>
<td>5.56</td>
<td>5.34</td>
<td>0.22</td>
</tr>
<tr>
<td>9</td>
<td>5.97</td>
<td>5.98</td>
<td>-0.01</td>
</tr>
<tr>
<td>10</td>
<td>6.64</td>
<td>6.51</td>
<td>0.13</td>
</tr>
<tr>
<td>11</td>
<td>7.03</td>
<td>6.84</td>
<td>0.19</td>
</tr>
<tr>
<td>12</td>
<td>7.69</td>
<td>6.99</td>
<td>0.70</td>
</tr>
<tr>
<td>Average</td>
<td>6.721</td>
<td>6.463</td>
<td>0.258</td>
</tr>
</tbody>
</table>

This data gives us the following summary statistics.

\[ X_1 = 6.721, \quad X_2 = 6.463, \quad S_1^2 = 0.736, \quad S_2^2 = 0.434, \quad S_p^2 = \frac{11S_1^2 + 11S_2^2}{12 + 12 - 2} = 0.585 \]

Let \( \mu_1 \) be the average bacteria count before treatment and \( \mu_2 \) the average after treatment. We are interested in whether there is evidence that the heat treatment reduces the amount of bacteria so we could test \( H_A : \mu_1 > \mu_2 \) vs \( H_0 : \mu_1 = \mu_2 \).

\[ H_0 : \quad \mu_1 = \mu_2 \]

\[ H_A : \quad \mu_1 > \mu_2 \]
We would then use Equation 6.9 to calculate the p-value. First we calculate $t_{obs}$

$$t_{obs} = \frac{X_1 - X_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{6.721 - 6.463}{\sqrt{0.585 \left( \frac{1}{12} + \frac{1}{12} \right)}} = 0.826$$

and then the p-value.

$$p^* = P(t_{n_1 + n_2 - 2} \geq t_{obs}) = P(t_{22} \geq 0.826) = 0.2088$$

This is a pretty large p-value so we would fail to reject $H_0$. In other words we do not have sufficient evidence to conclude that the heat treatment is helping.

This is not very interesting because we have failed to prove anything! However there is a problem with the calculation. $X_{i1}$ and $X_{i2}$ are not independent because both observations come from the same bottle. If the bacteria count is high before treatment it tends to be high after treatment. On the other hand if the count is low before treatment it tends to be low after treatment also. The numbers are related! This means that

$$Var(X_1 - X_2) \neq \sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$$

and

$$S_{\bar{X}_1 - \bar{X}_2}^2 \neq S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$$

so that

$$t_{obs} = \frac{X_1 - X_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \neq t_{n_1 + n_2 - 2}$$

This means that the p-value we calculated is incorrect. How do we fix the problem and calculate the correct value? First we need to work out exactly what $Var(X_1 - X_2)$ is equal to.

$$Var(X_1 - X_2) = Var \left[ \frac{1}{12} (X_{1,1} + X_{2,1} + \cdots + X_{12,1}) - \frac{1}{12} (X_{1,2} + X_{2,2} + \cdots + X_{12,2}) \right]$$

$$= Var \left[ \frac{1}{12} [(X_{1,1} - X_{1,2}) + (X_{2,1} - X_{2,2}) + \cdots + (X_{12,1} - X_{12,2})] \right]$$

$$= Var \left[ \frac{1}{12} (D_1 + D_2 + \cdots + D_{12}) \right] \quad (D_i = X_{i1} - X_{i2})$$

$$= Var(D)$$

$$= Var(D)/12$$

In other words to estimate the variance of $Var(X_1 - X_2)$ we need to estimate the variance of the differences $(D_i = X_{i1} - X_{i2})$. This is very easy. All we need to do is to take the differences $D_1, D_2, \ldots, D_{12}$, treat them as a single sample and calculate the sample variance

$$S_D^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (D_i - \bar{D})^2$$
This gives us a new test statistic

\[ t_{\text{obs}} = \frac{X_1 - X_2}{\sqrt{\frac{S_D^2}{n}}} = \frac{D}{\sqrt{\frac{S_D^2}{n}}} \]

With this new test statistic, provided \( H_0 \) is true,

\[ t_{\text{obs}} \sim t_{n-1} \]

In our case

\[ D = \frac{1}{12}(0.03 + 0.14 + \cdots + 0.70) = 0.258 \]

and

\[ S_D^2 = \frac{1}{12 - 1}[ (0.03 - 0.258)^2 + \cdots + (0.70 - 0.258)^2] = 0.1275 \]

Therefore

\[ t_{\text{obs}} = \frac{D}{\sqrt{\frac{S_D^2}{n}}} = \frac{0.258}{\sqrt{0.1275/12}} = 2.50 \]

This gives a new p-value of

\[ p^* = P(t_{n-1} \geq t_{\text{obs}}) = P(t_{11} \geq 2.50) = 0.0148 \]

This is much smaller than the previous (incorrect) p-value. Now we have strong evidence to reject \( H_0 \) i.e. we are pretty sure that the heat treatment does help.

### 6.10.2 Hypothesis Tests on Paired Data

The previous example illustrates a situation where we have “paired data”. In other words instead of having two independent samples that we are comparing we have paired observations where the two data points within each pair are related to each other.

In general when we have paired data we calculate the differences

\[ D_i = X_{i,1} - X_{i,2} \]

and then treat them as if we had a single sample (as in Section 6.6). Usually we wish to test

\[ H_0 : \mu_D = 0 \]
\[ H_A : \mu_D < 0 \quad (or \ \mu_D > 0 \quad or \ \mu_D \neq 0) \]

so we would calculate the p-value in the following manner. (Note that this is identical to the method used for a hypothesis on the mean for a single sample.)
If $H_A : \mu_D < 0$

$$p^* = P(t_{n-1} \leq t_{obs}) \quad \text{where} \quad t_{obs} = \frac{D}{\sqrt{S_D^2 / n}}$$ (6.16)

If $H_A : \mu_D > 0$

$$p^* = P(t_{n-1} \geq t_{obs}) \quad \text{where} \quad t_{obs} = \frac{D}{\sqrt{S_D^2 / n}}$$ (6.17)

If $H_A : \mu_D \neq 0$

$$p^* = 2P(t_{n-1} \geq |t_{obs}|) \quad \text{where} \quad t_{obs} = \frac{D}{\sqrt{S_D^2 / n}}$$ (6.18)

### 6.10.3 What are the differences?

Why did we get a different conclusion when we used pairing?

<table>
<thead>
<tr>
<th>Without pairing</th>
<th>With pairing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test statistic</strong></td>
<td>$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$</td>
</tr>
<tr>
<td><strong>Mean difference</strong></td>
<td>$X_1 - X_2 = 0.258$</td>
</tr>
<tr>
<td><strong>Estimated standard deviation</strong></td>
<td>$\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = 0.312$</td>
</tr>
<tr>
<td><strong>Observed test statistic</strong></td>
<td>$t_{obs} = \frac{0.258}{0.312} = 0.826$</td>
</tr>
<tr>
<td><strong>Distribution</strong></td>
<td>$t_{obs} \sim t_{n_1+n_2-2} = t_{22}$</td>
</tr>
</tbody>
</table>

**Advantage of pairing**

A reduced estimate of standard deviation (i.e. 0.103 vs 0.312)

**Disadvantage of pairing**

Fewer degrees of freedom (i.e. 11 vs 22)

Generally we will want to use the paired hypothesis test if the data are paired. It usually results in a more “powerful” (i.e. smaller $\beta$) test than assuming two independent groups.