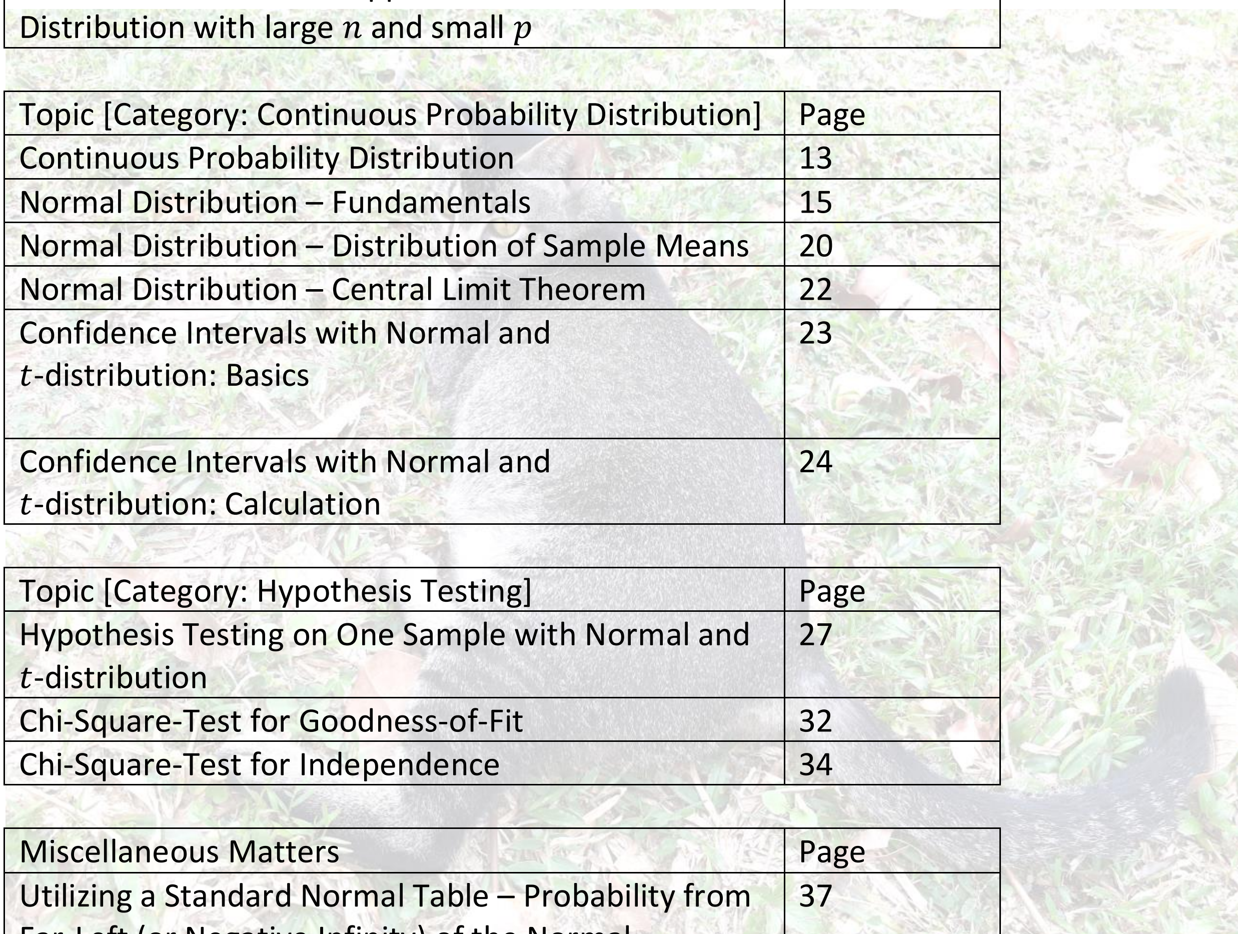
Title	JC and Polytechnic Mathematics Material Compilation – Statistics
Editor	Lee Jian Lian
Date	12/4/2019

Topic [Category: Discrete Probability Distribution]	Page
Discrete Probability Distribution	2
Binomial Distribution	3
Poisson Distribution – Fundamentals	8
Poisson Distribution as Approximation to Binomial Distribution with large n and small p	11

Topic [Category: Continuous Probability Distribution]	Page
Continuous Probability Distribution	13
Normal Distribution – Fundamentals	15
Normal Distribution – Distribution of Sample Means	20
Normal Distribution – Central Limit Theorem	22
Confidence Intervals with Normal and t-distribution: Basics	23
Confidence Intervals with Normal and t-distribution: Calculation	24

Topic [Category: Hypothesis Testing]	Page
Hypothesis Testing on One Sample with Normal and t-distribution	27
Chi-Square-Test for Goodness-of-Fit	32
Chi-Square-Test for Independence	34

Miscellaneous Matters	Page
Utilizing a Standard Normal Table – Probability from Far-Left (or Negative Infinity) of the Normal Distribution to the z-score	37



Applicable to Courses of the Following Nature

- Information Technology
- Engineering
- Applied Science/Science
- Business/Business Management
- GCE ‘A’ Level H1 Mathematics
- GCE ‘A’ Level H2 Mathematics & ‘A’ Level H2 Further Mathematics

Title	Discrete Probability Distribution
Author	Liu Hui Ling, Ngee Ann Polytechnic
Date	17/10/2018

Discrete Probability Distribution has the following properties.

- Takes in discrete variables (Whole number values k , where $k \geq 0$)
- Countable number of values involved
- Takes in random variables (Sum of all probabilities must be equal to 1)

Example

Number of Events	0	1	2	3
Probability	0.5	0.25	0.10	0.15

In the case of the Binomial Distribution, as represented by the formula below,

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

The following limitation is imposed, as any values that doesn't comply to the following limitation is undefined.

$0 \leq k \leq n$

In the case of Poisson Distribution, as represented by the formula below,

$P(X = k) = e^{-\mu} \left(\frac{\mu^k}{k!}\right)$

Where $0 \leq k < \infty$

The inequality $0 \leq k < \infty$, implies the number of events you are performing the probability calculations for can be any finite whole number greater than or equal to 0.

While there is no upper limit to the value of k, a theorem guarantees the value of all probability will sum up to 1:

As the probability within a Binomial Distribution approaches 0 and the number of trials approaches infinity. The Binomial Distribution will converge to the Poisson Distribution.

This implies that the Poisson Distribution is just a special case of Binomial Distribution, which means the probability will still sum up to 1 anyway.

Title	Polytechnic and A Level H2 Mathematics (Statistics) Binomial Distribution
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	9/6/2018

Applicable to the following levels

- ✓ School of Information Technology Students (Computing Mathematics)
- ✓ School of Engineering (Engineering Mathematics – Statistical Analysis)
- ✓ School of Business Management (Statistics – Business Statistics)
- ✓ School of Chemical and Life Sciences – Biostatistics
- ✓ JC/MI Students – H2 Mathematics – Statistics

Due to my school's syllabus, it may or may not cover everything required for H2 Mathematics. JC/MI students should see referring to this guide as a last resort if you still don't know the basics.

To use the binomial distribution, the following requirements must be met.

- There will only be 2 possible outcomes (Success/Failure, Yes/No, etc.)
- Each trial is an independent event (that is, will not affect the subsequent trial or be affected by past trial)

You must also know the following information or able to derive the following details

- You know the probability of each trial
- You are given the total number of trials and the number of trials the probability is being calculated for, which will be shown in notation form in the next few pages.

Formula for Binomial Distribution Probability Given as Follows

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

[It may be written slightly differently in other textbooks. But they should mean the same thing.]

Notation	Meaning
$P(X = k)$	Probability of obtaining an outcome, given the variable or the number of trials being calculated for will be exactly equal to k Or simply put, the number of trials the outcome is being calculated for
$\binom{n}{k}$	Total number of combinations the 2 outcomes can be rearranged n refers to the total number trials k refers to the number of trials the outcome is being is being calculated for
p^k	The probability the outcome you are finding for after k number of independent trials. (Example, the outcome can be Yes or Success)
$(1-p)^{n-k}$	The probability of obtaining the alternate outcome after $n-k$ number of independent trials. (Example, if your outcome is Yes or Success, then the alternate corresponding outcome are No or Failure respectively.)
$X \sim B(n, p)$	The random variable X is to follow a binomial distribution, over n number of independent trials, which trial shall have a p probability of obtain the outcome mentioned in question.

Formula List for Analyzing a Binomial Distribution	
Formula for Mean (μ) (Also called Expected Value)	$\mu = np$
Formula for Variance (σ^2)	$\sigma^2 = np(1-p)$
Formula for Standard Deviation (σ)	$\sigma = \sqrt{np(1-p)}$

Binomial Distribution Questions and Example
[Section I]: Basic Calculation
Q1: Given the following binomial distribution and information.

$$X \sim B(5,0.3)$$

Evaluate the following
(a) $P(X = 2)$
(b) $P(X < 2)$
(c) $P(X < 3)$
(d) $P(X \geq 2)$

Q1(a)

$$P(X = 2) = \binom{5}{2} 0.3^2(1 - 0.3)^{5-2} = 10(0.09)(0.343) = 0.3087$$

Q1(b)

$$P(X < 2) = P(X = 0) + P(X = 1)$$

$$P(X = 0) = \binom{5}{0} 0.3^0(1 - 0.3)^{5-0} = 1(0.3)^0(0.7)^{5-0} = 0.16807$$

$$P(X = 1) = \binom{5}{1} 0.3^1(1 - 0.3)^{5-1} = 5(0.3)^1(0.7)^{5-1} = 0.36015$$

$$P(X < 2) = 0.16807 + 0.36015 = 0.52822$$

Q1(c)

$$P(X < 3) = 1 - [P(X = 3) + P(X = 4) + P(X = 5)]$$

[Values of all probabilities in binomial distribution must sum up to 1]
**Use the method that require the least number of calculation.

$$P(X = 3) = \binom{5}{3} 0.3^3(1 - 0.3)^{5-3} = 10(0.027)(0.7)^2 = 0.1323$$

$$P(X = 4) = \binom{5}{4} 0.3^4(1 - 0.3)^{5-4} = 5(0.0081)(0.7) = 0.02835$$

$$P(X = 5) = \binom{5}{5} 0.3^5(1 - 0.3)^{5-5} = 1(0.00243)(1) = 0.00243$$

$P(X = 3) + P(X = 4) + P(X = 5) = 0.16308$

$P(X < 3) = 1 - 0.16308 = 0.83692$
Q1(d) [From Answers Derived in Q1(b)]
$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - (0.16807 + 0.3015)$
$= 1 - 0.46458$
$= 0.53542$

Section II (Application of Binomial Distribution)
Q2
A survey indicates that 60% of the school's student population is interested to participate in an event. You randomly selected 7 students who had participated in the survey.

(a) Is binomial distribution suitable for this question, please justify your answer.
(b) Find the probability that exactly 4 students are interested in the event.
(c) Find the probability that at most 3 students are interested in the event.
(d) Find the expected value, standard deviation and variance of the distribution.

(a) Yes. Every student's interest in the event can be regarded as independent. There are only two possible outcomes, either a "YES" or a "NO".

(b)
$$P(X = 4) = \binom{7}{4}(0.6)^4(1 - 0.6)^{7-4} = 35(0.1296)(0.064) = 0.290304$$

(c)
$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$P(X = 0) = \binom{7}{0}(0.6)^0(1 - 0.6)^{7-0} = 0.00164$$

$$P(X = 1) = \binom{7}{1}(0.6)^1(1 - 0.6)^{7-1} = 0.01720$$

$$P(X = 2) = \binom{7}{2}(0.6)^2(1 - 0.6)^{7-2} = 0.00741$$

$$P(X = 3) = \binom{7}{3}(0.6)^3(1 - 0.6)^{7-3} = 0.19354$$

$P(X \leq 3) = 0.00164 + 0.01720 + 0.00741 + 0.19354 = 0.21979$

(d)
$\mu = np$
$\mu = 0.6(7) = 4.2$

$\sigma^2 = np(1-p)$
$\sigma^2 = 4.2(1-0.6) = 1.62$

$\sigma = \sqrt{1.62} = 1.2728$

Q3 [Question Taken from Nanyang Polytechnic Computing Mathematics 2 Exam Paper]
Given that the mean and variance of a Binomial Distribution X is 5 and $\frac{15}{8}$ respectively, find the value of n and p in the Binomial Distribution of X.

Mean $\mu = np = 5$
Variance $=\sigma^2 = np(1-p) = \frac{15}{8}$

Equation 1: $np = 5$
Equation 2: $np(1-p) = \frac{15}{8}$

$$1 - p = \frac{np(1-p)}{np} = \frac{\left(\frac{15}{8}\right)}{5} = 0.375$$

$p = 1 - 0.375 = 0.625$

$$n = \frac{np}{p} = \frac{5}{0.625} = 8$$

Title	Polytechnic and JC H2 Mathematics – Poisson Distribution
Author	Liu Hui Ling, Ngee Ann Polytechnic (Assisted by Chen Xin Yi)
Date	10/6/2018

Applicable to the following levels and types of education institution

- ✓ JC/MI – H2 Mathematics (Statistics)
- ✓ Engineering, Physics, Chemistry and Biology – Statistical Calculations
- ✓ Information Technology – Data Analytics

Apart from studying Business related modules, we also do some Business Statistics Module which drew my interest in this topic of probability distribution. I think we can just get straight to the point and explain what are the prerequisite and reason for using of this type of probability distribution.

Purpose of Poisson Distribution is to

- Calculate the probability of an event happening in the subsequent intervals when the mean rate of occurrences per unit of interval is given.

Information needed

- Mean occurrence rate
- Unit of Intervals (Unit of interval is the key word here. Without this unit of intervals, it is highly likely that the use of Poisson Distribution cannot be justified. Unit of intervals can come in terms of the time-interval, area-interval, volume-interval and etc.)

Requirements

- Multiple events cannot happen simultaneously
- All events must be <u>independent</u> (i.e. Unaffected by past events and will not affect subsequent events)

Formula used (Explanation given at the next page)

$$P(X = k) = e^{-\mu}\left(\frac{\mu^k}{k!}\right)$$

$P(X = k)$	The probability that the number of events occurrence within the unit interval being exactly equal to k.
μ	Mean occurrence of event per unit interval. (i.e. Expected Mean or Expected Value)
e	Euler Constant (Approximately 2.71828, rounded off to 6 significant figures) Modern scientific calculators should have this functionality, you just need to locate e^x or e.

Question 1.
The number of sick leaves taken by students in a class per week is known to follow a Poisson distribution with a mean of 1.8.

Find the probability that
(a) There are no sick leaves taken by students in the class in a one-week period.
(b) At least 4 sick leaves are taken by students in the class in a one-week period.

(a)

$$P(X = 0) = e^{-1.8}\left(\frac{1.8^0}{0!}\right) = 0.165298$$

(b)

$$P(X = 0) = e^{-1.8}\left(\frac{1.8^0}{0!}\right) = 0.165298$$

$$P(X = 1) = e^{-1.8}\left(\frac{1.8^1}{1!}\right) = 0.297538$$

$$P(X = 2) = e^{-1.8}\left(\frac{1.8^2}{2!}\right) = 0.267784$$

$$P(X = 3) = e^{-1.8}\left(\frac{1.8^3}{3!}\right) = 0.160671$$

$P(X = 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)] = 0.108709$

Q2 [Taken from NYP Exam Paper] [Added with help from Anonymous Engineering Student from Nanyang Polytechnic who refuses to disclose his/her name]

(a) The IT department experiences an average of 2.4 Local Area Network (LAN) errors in a day. Assuming that these LAN errors experience in a day follows a Poisson distribution, find the probability that on any given day:

(i) zero network error will occur?

(2 marks)

(ii) two or more network errors will occur?

(3 marks)

(iii) there are more than one network errors in a 7-day work-week?

(3 marks)

Give your answers correct to 4 decimal places.

(a)

(i)

$$P(X = 0) = e^{-2.4}\left(\frac{2.4^0}{0!}\right) = 0.09071795 = 0.0907(4dp)$$

(ii)

$$P(X = 1) = e^{-2.4}\left(\frac{2.4^1}{1!}\right) = 0.2177231$$

$$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 0.6916\ (4dp)$$

(iii)Since there are 2.4 Network Errors a day, we can argue that in a 7-day-work week, there should be 7×2.4 Network Errors which is an average of 16.8 network errors per week. In this case, the mean occurrence rate per week is $\mu = 16.8$.

$$P(X = 0) = e^{-16.8}\left(\frac{16.8^0}{0!}\right) = 5.0565313 \times 10^{-8}$$

$$P(X = 1) = e^{-16.8}\left(\frac{16.8^0}{1!}\right) = 8.4949727 \times 10^{-7}$$

$$P(X > 1\ ) = 1 - [P(X = 0) + P(X = 1)] = 1 - [P(X = 0) + P(X = 1)]$$
$$= 1.0000\ (4dp)$$

Title	Poisson Distribution Approximation for Binomial Distribution for large values of n and small values of p
Author	[Anonymous], Student from School of Engineering, Nanyang Polytechnic
Date	4/3/2019

Let's look at the Poisson Limit Theorem closely.
"Given p approaches 0 and the value of n approaches ∞ (infinity) in a Binomial Distribution, the distribution will approach the Poisson Distribution."

As a result, the following are the requirements for any Binomial Distribution to be approximated by a Poisson Distribution:

If $n > 50, p < 0.1$ such that $np < 5$
$X{\sim}B(n, p) \approx\ X{\sim}Po(np)$

Example 1. [Questions Obtained from School of Information Technology Examination Papers – Computing Mathematics 2, with help from friends]
SIT/2019/January

A stamping machine produces components at a rate of 300 per day. It is known that 1% of the output is defective. Assuming this rate is approximated by a Poisson Distribution.

(a) Estimate the mean of the Poisson Distribution
(b) Find the probability that no defective output is produced in any given day
(c) Find the probability that at least 1 and at most 10 defective outputs are produced in any given day

(a)
$X{\sim}B(n, p) \approx X{\sim}Po(\mu)$
$\mu = np = 1\%(300) = 3$

(b)
$X{\sim}Po(3)$

$$P(X = k) = e^{-u}\left(\frac{\mu^k}{k!}\right)$$

$$P(X = 0) =\ e^{-0}\left(\frac{3^0}{0!}\right) = 0.0497870683$$

(c)

$$P(X=1)=e^{-3}\left(\frac{3^1}{1!}\right)$$

$$P(X=2)=e^{-3}\left(\frac{3^2}{2!}\right)$$

$$P(X=3)=e^{-3}\left(\frac{3^3}{3!}\right)$$

$$P(X=4)=e^{-3}\left(\frac{3^4}{4!}\right)$$

$$P(X=5)=e^{-3}\left(\frac{3^5}{5!}\right)$$

$$P(X=6)=e^{-3}\left(\frac{3^6}{6!}\right)$$

$$P(X=7)=e^{-3}\left(\frac{3^7}{7!}\right)$$

$$P(X=8)=e^{-3}\left(\frac{3^8}{8!}\right)$$

$$P(X=9)=e^{-3}\left(\frac{3^9}{9!}\right)$$

$$P(X=10)=e^{-3}\left(\frac{3^{10}}{10!}\right)$$

Summing up all the probabilities $P(X=0)\ to\ P(X=10)$, we get the following values
$P(X\leq 10)=0.999707663$

$P(1\leq X\leq 10)=0.999707663-0.049787068=0.9499$(4 decimal places)

Title	Continuous Probability Distribution
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	4/3/2019

In this topic we focus on

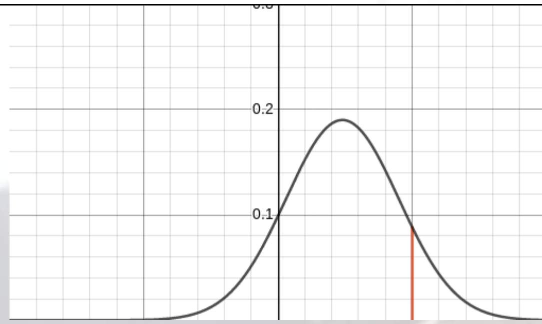
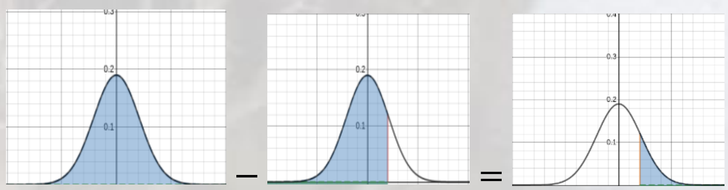
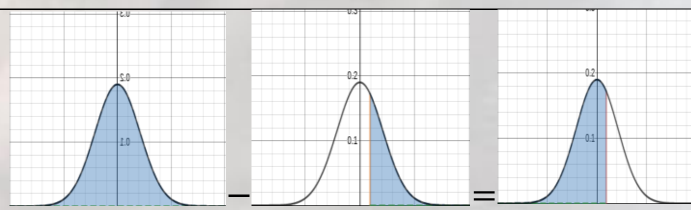
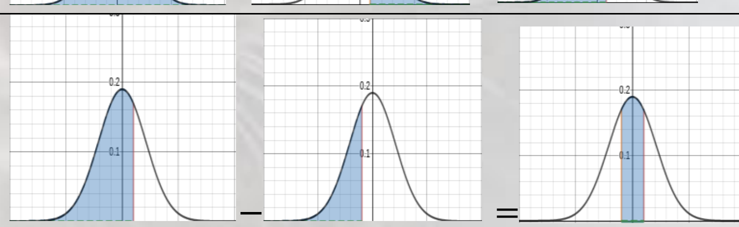
- Continuous random variables
- Basic Concepts of Area Under Curve as Probability Value

Continuous random variables occur in many areas in statistics, they can take on uncountable number of variables in contrast with Discrete random variables which takes on countable number of variables.

A continuous probability distribution takes on continuous random variables, where the probability distribution is typically represented by a graph, which area under curve from the left all the way to the right of the probability distribution is exactly equal to 1.

Example includes

- Height of students
- Test scores
- Weight of bobcats
- Time students spend studying and revising for exams

Basic Operations Involving Continuous Probability Distribution	Illustration using a graph (Normal Distribution as an Example)
$P(X = k) = 0$ Consequently $P(X > k) = P(X \geq k)$ $P(X < k) = P(X \leq k)$	 The red line refers to the variable $X = k$ and it extremely obvious that you can't produce an area under curve with just a point on the X-axis, you need at least 2 points or a range of values to be able to define what is the area under curve is.
$P(X > k\ ) = 1 - P(X < k)$	
$P(X < k) = 1 - P(X > k)$	
$P(a < X < b) =$ $P(X < b) - P(X < a)$	

Title	Polytechnic and A Level H2 Mathematics (Statistics) – Normal Distribution
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	15/6/2018

Applicable to the following levels

- ✓ School of Information Technology Students (Computing Mathematics)
- ✓ School of Engineering (Engineering Mathematics – Statistical Analysis)
- ✓ School of Business Management Students (Statistics – Business Statistics)
- ✓ School of Chemical and Life Science – Biostatistics
- ✓ JC/MI Students – H2 Mathematics – Statistics

Items needed to start the topic

- ✓ Standard Normal Table

(Recommended, print a Standard Normal Table to refer to while doing your homework and assignments, while there are literally thousands of them on the internet, best is get from your school teacher and keep it. I also recommend you upload a copy to a cloud disks, just in case you lose the Standard Normal Table, you restore them quickly and reprint them.)

SEAB do have a copy of Standard Normal Table on their website. With enough searching you should be able to find it.

My school also issues its own version of the standard normal table
(I have seen Standard Normal Table issued by other schools before, they have different way of expressing the value of area under curve and different numerical accuracy requirements.)

Table of Notation	
$\boldsymbol{X \sim N(\mu, \sigma^2)}$	This is how a Normally Distributed Variable should be written. This literally means, The variable X is to be normally distributed, with a mean of μ, and a variance of σ^2. (Replace the symbols with values as specified in the questions you are going to answer)
$\boldsymbol{Z \sim N(0, 1)}$	Standard Normal Distribution. With mean as 0 and a variance of 1. Since $\sqrt{1} = 1$, the standard deviation of the distribution is also 1 in the case of a Standardized Normal Distribution. In this case, Z is the number of standard deviations away from the mean, also called the Z-score.

Table of Formula
Formula for Standardization
$Z \sim N(0,1) = \frac{X - \mu}{\sigma}$

Properties of a Normal Distribution Curve.

- Mean, Median and Mode are all on the same value
- Symmetrical at mean*, implying the left side of the Normal Distribution has a total area of 0.5 and the right side of the Normal Distribution has a total area of 0.5 as well.

(This is important to know as I am aware that some standard normal table out there are not as straightforward, I have seen other schools' standard normal table that shows value of area under curve from the mean to the Z-score, the most common types, however, shows area from the left of the distribution to the mean and shows the area from the left of the distribution all the way to the right of the distribution.)

Example Questions
Example 1:
(Taken from Oxford University Lecture Notes)

The marks of 500 candidates in an examination are normally distributed with a mean of 45 marks and a standard deviation of 20 marks.
If 20% of the candidates obtained a distinction by scoring x marks or more, estimate the value of x.

Written in Normal Distribution Notation
$X \sim N(45, 20^2)$

$P(20\% \; of \; the \; candidates \; socring \geq x \; marks)$
$= P(80\% \; of \; candidates \; scoring \leq x \; marks)$

Within the Standard Normal Table, I will look for the probability value closest to 0.800
(In this case, the standard normal table doesn't have a value exactly equal to 0.800.)

It turned out, the standard normal table probability value closest to 0.800 is 0.7995, under $z = 0.84$

Given
$z = \frac{x-\mu}{\sigma}$
Applying Standardization Formula

$0.84 = \frac{x-45}{20}$

$20(0.84) = x - 45$

$16.8 = x - 45$

$x = 16.8 + 45 = 61.8$

Example 2 (Taken from Online Sources)
The daily revenue of a small restaurant is approximately normally distributed with a mean of \$530 and a standard deviation of \$120. To be in profit, the restaurant must receive at least \$350.

Find the probability that the restaurant will be in profit on any given day.

Given $z = \frac{x-\mu}{\sigma}$

Applying Standardization Formula

$$z = \frac{350 - 530}{(120)} = -1.5$$

Looking for z-$score = 1.5$ in the standard normal table, it turns out the probability value is 0.9332, thus the probability of the restaurant getting $\geq \$350$, is 0.9332

Example 3 (Taken from NYP Computing Mathematics 2 Paper)

(a) Suppose K is normally distributed with mean 15 and variance 4, find

(i) $P(10 < K < 20)$, (4 marks)

(ii) $P(K > 18)$. (2 marks)

Rewritten in Normal Distribution Notation, we get this
$X \sim N(\mu, \sigma^2) = K \sim N(15, 4)$
Implying the standard deviation $\sigma = \sqrt{4} = 2$

Example 3

(a) (i)

$Z(K = 10) = \frac{10-15}{2} = -\frac{5}{2} = -2.5$

$$Z(K = 20) = \frac{20 - 15}{2} = 2.5$$

$P(Z < -2.5) = 0.0062$
$P(Z < 2.5) = 0.9938$
$P(-2.5 < Z < 2.5) = 0.9938 - 0.0062 = 0.9876$

(ii)

$Z(K = 18) = \frac{18-15}{2} = \frac{3}{2} = 1.5$

$P(Z < 1.5) = 0.9332$
$P(Z > 1.5) = P(K > 18) = 1 - 0.9332 = 0.0668$

Title	Normal Distribution – Distribution of Sample Mean
Date	13/8/2018
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]

Applicable to

- Nanyang Polytechnic – School of Chemical and Life Sciences (Biostatistics)
- Nanyang Polytechnic – School of Engineering (Engineering Mathematics)
- Nanyang Polytechnic – School of Information Technology (Computing Mathematics – Statistics)

Assumptions

- You already understood normal distribution and how to read the standard normal table. (Do read up on Normal Distribution if you don't understand Normal Distribution as the notations and calculations used in this topic are rather similar.)

(It is unclear if this topic would apply to 'A' Level students)

Purpose of this topic

- Determining probability of obtaining a certain range of mean values from a defined sample, given a normally distributed or approximately normally distributed population.

Notation	Meaning
μ	Population Mean
$\bar{X}$	Sample Mean
σ	Population standard deviation
$\sigma_{\bar{X}}$	Standard Deviation of Sampling Distribution (Also referred to as standard error)
n	Sample Size (Number of subjects you are performing the analysis on)

Formula List

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$Z - score = \frac{X-\mu}{\sigma_{\bar{X}}}$$

Question 1. (Taken from SCL Notes)
In a certain population of swordtail fish, the length of individual fish follows an approximately normal distribution, with a mean of 52.0 mm and standard deviation of 6.0mm. Find the probability that a random sample of 25 swordtail fishes with have an average length of

a) Less than 48.6 mm
b) Between 52.4mm and 54.4mm

Population Normal Distribution to Be Written as Follows
$X \sim N(52.0, 6.0^2)$

Sample Normal Distribution Values to Be Written as Follows
$\bar{X} \sim N(\mu, \sigma_x^2)$
Computation of σ_x as follows
$\mu = 52.0$
$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{6.0}{\sqrt{(25)}} = \frac{6.0}{5} = 1.2$
Rewritten as: $\bar{X} \sim N(52.0, 1.2^2)$

Answering Question 1(a)
$Z(\bar{X} = 48.6) = \frac{(48.6-52.0)}{1.2} = -2.83$

$P(Z < -2.83) = 0.0023$

Answer: 0.0023

Answering Question 1(b)
$Z(\bar{X} = 54.4) = \frac{54.4-52.0}{1.2} = 2.00$

$Z(\bar{X} = 52.4) = \frac{52.4-52.0}{1.2} = 0.33$

$P(\bar{X} < 54.4) = 0.9772$

$P(\bar{X} < 52.4) = 0.6293$

$P(54.4 > \bar{X} > 52.4) = 0.9772 - 0.6293 = 0.3479$
Answer: 0.3479

Title	Normal Distribution – Central Limit Theorem
Editor	-
Date	6/4/2019

Formal Statements of Central Limit Theorem as Follows

The central limit theorem states that if you have a population of mean μ *and take sufficiently large random sample (size* $n \geq 30$) *from the population with replacement, the distribution of the sample means will be approximately normally distributed.*

If the population is normally distributed or approximately normally distributed to start with and random samples are taken from the population, regardless of the sample size, the distribution of sample mean will also be normally or approximately normally distributed.

Why bother with central limit theorem?

- t −distribution for large degrees of freedom approximates the Normal Distribution.
- Chi-Square distribution for large degree of freedom is also approximately normally distributed

Examination which lots of candidates participate in uses the Normal Distribution to conduct grading, data reporting and data analysis as such examination will have many candidates and therefore, invoke the Central Limit Theorem.

Title	Normal Distribution and t distribution – Construction of Confidence Interval – Basics Theory
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	22/8/2018

Applicable to the following schools of Nanyang Polytechnic

- School of Information Technology – Computing Mathematics 2
- School of Engineering – Engineering Mathematics 2B
- School of Chemical and Life Sciences – Biostatistics
- School of Business Management – Business Statistics

Applicable to A Level Syllabus

- H2 Further Mathematics

The following table best illustrate the prerequisite to using either z-score method (z-test) or t-score method (t-test) to compute the confidence interval of a sample.

Situation	Action Taken
Question asks you to construct confidence interval with population standard deviation σ known	Construct Confidence Interval Using Standard Normal Distribution ($z -$ score)
Question asks you to construct confidence interval with small sample size but stating the sample is approximately normally distributed	Construct Confidence Interval Using Standard Normal Distribution ($z -$ score)
Question asks you to construct confidence interval of a sample with large sample sizes ($n \geq 30$)	By Central Limit Theorem, the distribution is approximately normal and thus we use the Standard Normal Distribution ($z -$score) to construct confidence interval
Question ask you to construct confidence interval of small sample sizes ($n < 30$), population standard deviation σ unknown	Construct Confidence Interval Using $t -$score values from the $t -$distribtuion.

Title	Normal and t-Distribution – Construction of Confidence Interval – Calculation Phase
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	14/9/2018

**This guide assumes you have already read my previous guide on accessing which method is the most suitable for construction of confidence interval in various situations.

Margin of Error → • Population Standard Deviation (σ) Known • Question Mentions the Sample is Normally or Approximately Normally Distributed • Large Sample Size of $n \geq 30$	$$E = Z_C\left(\frac{\sigma}{\sqrt{n}}\right)$$ E refers to the margin of error Z_c refers to the $z - score$ of the confidence interval in question σ refers to the population standard deviation n refers to the sample size
Margin of Error → • Population Standard Deviation (σ) not known and small sample size of $n < 30$	$$E = t_c\left(\frac{s}{\sqrt{n}}\right)$$ E refers to the margin of error t_c refers to the $t -$ score of the confidence interval in question s refers to the standard deviation of the sample n refers to the sample size
Confidence Interval Formula →	$\bar{X} \pm E$
Degrees of Freedom →	$n - 1$ Where n is the sample size in the question

Reason for this topic:
Confidence interval serves as a robust and analytical approach to determine how much will actual value deviate from observed value. The idea of confidence interval is that it is a range of value where we are reasonably sure our population mean lies in.

A 0.95 or 95% confidence interval has a 0.95 probability of containing the population mean under the curve of the distribution.

(Taken from University of Texas at Dallas Website)
Q1. A sample size of $n = 100$ produced a sample mean of $\bar{X} = 16$. Assuming the population standard deviation $\sigma = 3$, compute the 95% confidence interval for the population mean μ.

Since the population standard deviation is known, use z $-$score method.

From my school's Standard Normal Table, the z-score of the confidence interval mentioned in the question is 1.960.

Standard Error

$$\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{100}} = 0.3$$

Margin of Error = $E = z_c \left(\frac{\sigma}{\sqrt{n}}\right)$

$$E = 0.3(1.960) = 0.588$$

Confidence Interval 16 ± 0.588
Confidence Interval at Between 15.412 and 16.588

Q2. To access the accuracy of a laboratory scale, a standard weight known to weigh 1 gram is repeatedly weighed 4 times. The resulting measurements are (In grams): 0.95, 1.02, 1.01, 0.98.

Compute the Confidence Interval for μ.

Since $n \leq 30$ and the population standard deviation σ is not known, we answer the question using the t-score method.

Sample Mean $\bar{X} = \frac{0.95+1.02+1.01+0.98}{4} = 0.99$

Standard Deviation of Sample $s = 0.03162$

Standard Error = $\frac{0.03162}{\sqrt{4}} = 0.01581$

At Degrees of Freedom $= 3$ and 0.95 Confidence Interval
Margin of Error $= t_c\left(\frac{s}{\sqrt{n}}\right) = 3.182(0.01581) = 0.05030742$

Confidence Interval at 0.99 ± 0.05030742
Confidence Interval at Between 0.940 and 1.040

Title	Normal and t −distribution – Hypothesis Testing (One Sample)
Date	27/9/2018
Author(s)	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club] Liu Hui Ling, Ngee Ann Polytechnic

H_0: The person's claim is valid. We do not reject the null hypothesis.
H_a: The person's claim is not valid. We reject the null hypothesis in favor of an alternative hypothesis.

	Left-Tail	Two-Tail	Right-Tail
Symbol Used for Null Hypothesis	$\mu \geq x$	$\mu = x$	$\mu \leq x$
Symbol Used for Alternative Hypothesis	$\mu < x$	$\mu \neq x$	$\mu > x$
Objective of H_0:	H_0:Testing if value is above a certain minimum threshold.	H_0: Testing if value is within a certain acceptable range of values.	H_0:Testing if the value is below a certain maximum threshold.

Once again, the same concepts from other topics will apply. I want to elaborate in the context of this topic in this case.

If the question specifies, the distribution is approximately normal, normal, population standard deviation σ known or sample size $n \geq 30$, use the $z - score$ method to solve the question.

If question doesn't specify a known standard deviation and sample size $n < 30$, use the $t - score$ method to solve the question.

Steps to hypothesis testing as follows:

1. State null and alternative hypothesis
2. Determine nature of test and write down criteria for rejecting null hypothesis
3. Compute the standard error in question
4. Compute test statistics ($z - score$ or $t - score$)
5. Make your decision and justify why you fail to reject or rejected your null hypothesis

Standard Error	
Sample Size $n \geq 30$, Normally Distributed, Approximately Normally Distributed or σ known	Standard Error = $\frac{\sigma}{\sqrt{n}}$ σ is the population standard deviation.
Sample Size $n < 30$, σ not known	Standard Error = $\frac{s}{\sqrt{n}}$ s is the sample standard deviation
Formula for Test Statistics	
$z - score$	$z = \frac{sample\ mean - hyphothesized\ mean}{standard\ error}$
$t - score$	$t = \frac{sample\ mean - hyphothesized\ mean}{standard\ error}$

(Questions 1 and Question 2 Obtained from NYP SCL Notes)
Question 1.
A report claims that an adult has an average of 130 Facebook friends. A random sample of 50 adults revealed that the average number of Facebook friends is 142 with a standard deviation of 38.2. At 5% significance level, is there enough evidence to reject the claim?

Since question doesn't specify words that imply "more than" or "less than", the test is said to be <u>two-tailed</u> in nature, the null and alternative hypothesis will follow.
$H_0: \mu = 130$
$H_a: \mu \neq 130$, implying $\mu > 130\ OR\ \mu < 130$

We will also need to set the criteria for not rejecting and rejecting the null hypothesis. Since $n \geq 30$, we will use $z - score$ to perform the test. (As inferred from the standard normal table issued by my school.)

$H_0: -1.960 \leq z \leq 1.960$
$H_a: z > 1.960\ OR\ z < -1.960$

Calculate Standard Error
$$\frac{38.2}{\sqrt{50}} = 5.402\,296$$

Compute Test Statistics
$$z = \frac{142 - 130}{\frac{38.2}{\sqrt{50}}} = 2.221$$

Since the test statistics falls in the rejection region I mentioned above,
H_0 is to be rejected, as there is a lack of evidence to support the claim.

Question 2.
The management of a weight loss club claims it's members lose an average of 3 kg or more within the first month after joining the club. A consumer agency that wanted to check this claim took a random sample of 36 members of this club and they lost an average of 2.9 kg with standard deviation of 0.6 within the first month of membership. Test at 10% significance level if the management's claim is true.

The question stated the claim as "3 kg or more", implying the objective of the test is reject the hypothesis should the value falls below a certain threshold, the test is left-tailed in nature.

$H_0: \mu \geq 3$
$H_a: \mu < 3$

Since the test is left tailed nature, involving sample size $n \geq 30$, the following information is required to test the claim.
$H_0: z \geq -1.282$
$H_\alpha: z < -1.282$

Calculate Standard Error

$$\frac{0.6}{\sqrt{36}} = 0.1$$

Compute test statistics

$$z = \frac{2.9 - 3.0}{\frac{0.6}{\sqrt{36}}} = -1$$

Since test statistics doesn't fall within the rejection region as specified earlier,
We will not reject the claim as there is enough evidence to support the management's claim.

Question 3. [Question Created by Hui Ling herself.]
A report from XYZ Clinics claims that the waiting time for each patient from registration to consultation is 25 minutes or less. A civil servant from the Ministry of Health was tasked to check if the claim is valid and took a random sample of 15 patients and found out the average waiting time for each patient is 26.5 minutes, with a standard deviation of 8 minutes. Given the test is to be performed at $\alpha = 0.05$, what conclusion should that civil servant come to?

The claim specifies 25 minutes or less, implying the aim of the test is to reject the claim should the value fall above a certain threshold, which further implies a right tailed test is to be conducted.

$H_0: \mu \leq 25$
$H_\alpha: \mu > 25$

Since sample size is small and question did not mention "normally", "approximately normally distributed" or the population standard deviation, we use the $t - score$ method to approach the question. At degrees of freedom of $n - 1 = 14$, the following information is obtained.

$H_0: t \leq 1.761$
$H_\alpha: t > 1.761$

Compute Standard Error

$$\frac{8}{\sqrt{15}} = 2.065591$$

Compute test statistics

$$t = \frac{Sample\ Mean - Hyphothesized\ Mean}{Standard\ Error} = \frac{26.5 - 25}{\frac{8}{\sqrt{15}}} = 0.726$$

Since the test statistics doesn't falls in the rejection region, we conclude the following:

Since $t \leq 1.761$, the civil servant should not reject the claim mentioned in the report of XZY Clinics.

Title	Chi-Squared-Test for Goodness-of-Fit
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	19/10/2018

***You will need a Chi-Squared table (for assignments) or a software (for projects) in order to be able to calculate or obtain the Chi-Squared Critical Values.

The purpose of Chi-Squared-Test in general is to provide a robust, mathematical and analytical approach towards the following goals

- Determine how much categorical variables differ in terms of hypothesized value and observed value
- Determine whether two categorical variables are independent.

In this topic, we will focus mainly on the first goal, to measure the difference between the hypothesized value and observed value and from there, we will arrive at a decision on whether the hypothesized value is considered reliable.

Formula List as given:

$$\chi^2 = \sum \frac{(Observed\ Value - Expected\ Value)^2}{Expected\ Value}$$

χ^2 refers to the chi-squared value.
$Observed\ Value$ refers to the value under each category as obtained from the sample.
$Expected\ Value$ refers to the value of the respective category as hypothesized.
The above formula literally implies:
You compute the sum of $\frac{(Observed\ Value - Expected\ Value)^2}{Expected\ Value}$ for every category in the question to obtain the overall χ^2 value.

Degrees of Freedom = $Number\ of\ Categories - 1$

(χ^2 test for Goodness-of-Fit are always right-tailed in nature. You reject the null hypothesis should the χ^2 value goes beyond a certain threshold as obtained in your $\chi^2\ table$. That value is sometimes called the “critical value”.)

Example 1 (Obtained from NYP SCL Biostatistics Notes):
A recruitment agency's manager says that 22% of the undergraduates do not work, 26% work 1 to 20 hours per week, 18% work 21 to 34 hours, and 34% work 35 or more hours per week. You randomly selected 120 undergraduates and gather the results shown in the table. At $\alpha = 0.01$, can you reject the manager's claim?

Response	Frequency
Do not work	29
Work 1 to 20 hours	26
Work 21 to 34 hours	25
Work 35 or hours	40

Step 1. Propose a null and alternative hypothesis.
H_0: The manager's claim is reliable.
H_α: The manager's claim is not reliable.

Step 2. Set Rejection Criteria (As obtained from Chi-Square Table)
Under degrees of freedom = 3 and $\alpha = 0.01$
$\boldsymbol{H_0: \chi^2 < 11.345}$
$\boldsymbol{H_\alpha: \chi^2 \geq 11.345}$

Step 3. Compute Expected Values in Question:

Response	Frequency (Expected Values)
Do not work	26.4 = 22% × 120
Work 1 to 20 hours	31.2 = 26% × 120
Work 21 to 34 hours	21.6 = 18% × 120
Work 35 or hours	40.8 = 34% × 120

Step 4. Compute $\boldsymbol{\sum \frac{(Observed - Expected)^2}{Expected}}$ to get $\boldsymbol{\chi^2}$ value.

$$\chi^2 = \sum \frac{(Observed\ Value - Expected\ Value)^2}{Expected\ Value}$$

$$\chi^2 = \frac{(29 - 26.4)^2}{26.4} + \frac{(26 - 31.2)^2}{31.2} + \frac{(25 - 21.6)^2}{21.6} + \frac{(40 - 40.8)^2}{40.8} = 1.6736$$

Step 5. Make a decision to reject or not reject $\boldsymbol{H_0}$.
Since the χ^2 value < 11.345 , we have to conclude the following:
We do not reject H_0.

Title	Chi-Square Test of Independence
Author	Lim Wang Sheng, School of Information Technology, Nanyang Polytechnic [CCA: NYP Mentoring Club]
Date	21/10/2018

As mentioned in my previous topic, Chi-Square test is also used to determine if two categorical variables are dependent or independent of each other. What happen in this scenario is, you will be given a table (i.e. contingency table) where the rows represent a categorical variable, the columns represent another categorical variable. The aim of such test is to determine if row is independent of the column.

Despite similarities in formula, some major difference is to be noted.

Formula for Degrees of Freedom on a Contingency Table

$$d.f. = (Number\ of\ Rows - 1)(Number\ of\ Columns - 1)$$

Formula for Grand Total

$$Grand\ Total = \sum Value\ of\ Every\ Cell$$

Formula for Expected Value for Each Cell

$$Expected\ Value = \frac{(Row\ Total)(Column\ Total)}{Grand\ Total}$$

Formula for Chi-Square Statistics Value

$$\chi^2 = \sum \frac{(Observed\ Value - Expected\ Value)^2}{Expected\ Value}$$

The above formula means, you compute $\frac{(Observed\ Value - Expected\ Value)^2}{Expected\ Value}$ for every cell and add them up to get the Chi-Square Statistics Value.

Null Hypothesis

H_0: The 2 categorical variables in the question are independent.

H_α: The 2 categorical variables in the question are dependent.

Example 1. (Taken from NYP SCL Biostatistics Notes)
A health club manager wants to determine whether the number of days per week that students spent exercising is dependent of gender. A random sample of 275 students is selected and the results are shown as classified in the table. At 5% level, is there enough evidence to conclude that the number of days spent exercising per week is dependent of gender?

Gender	Days spent per week exercising			
	0-1	2-3	4-5	6-7
Male	40	53	26	6
Female	34	68	37	11

Step 1: Define Null and Alternative Hypothesis
H_0: The number of days spent exercising per week is independent of gender.
H_α: The number of days spent exercising per week is dependent of gender.

Step 2: Identify Degrees of Freedom
$d.f. = (Number\ Rows - 1)(Number\ of\ Columns - 1) = (4-1)(2-1) = 3$

Step 3: Set rejection criteria
At $\alpha = 0.05$
$H_0: \chi^2 < 7.815$
$H_\alpha: \chi^2 \geq 7.815$

Step 4: Calculate Row Total, Column Total and Grand Total
Row Total in Green Parenthesis
Column Total in Blue Parenthesis

Gender	Days spent per week exercising				
	0-1	2-3	4-5	6-7	Row Totals ↓
Male	40	53	26	6	(125)
Female	34	68	37	11	(150)
Column Totals →	(74)	(121)	(63)	(17)	

Grand Total = 275

Step 5: Compute Expected Value for Every Cell Using the Formula

$$Expected\ Value = \frac{(Row\ Total)(Column\ Total)}{Grand\ Total}$$

The following table is the result of calculation of the Expected Value using the above formula as shown.

(Expected Values Table) Gender	Days spent per week exercising			
	0-1	2-3	4-5	6-7
Male	$\frac{370}{11}$	55	$\frac{315}{11}$	$\frac{85}{11}$
Female	$\frac{444}{11}$	66	$\frac{378}{11}$	$\frac{102}{11}$

Step 6:

Compute Chi Square Statistic Value by Applying the Following Formula

$$\chi^2 = \sum \frac{(Observed\ Value - Expected\ Value)^2}{Expected\ Value}$$

$$\chi^2 = \frac{\left(40 - \frac{370}{11}\right)^2}{\frac{370}{11}} + \frac{(53-55)^2}{55} + \frac{\left(26 - \frac{315}{11}\right)^2}{\frac{315}{11}} + \frac{\left(6 - \frac{85}{11}\right)^2}{\frac{85}{11}} + \frac{\left(34 - \frac{444}{11}\right)^2}{\frac{444}{11}} + \frac{(68-66)^2}{66} + \frac{\left(37 - \frac{378}{11}\right)^2}{\frac{378}{11}} + \frac{\left(11 - \frac{102}{11}\right)^2}{\frac{102}{11}}$$

$$\chi^2 = 3.493$$

Step 7:

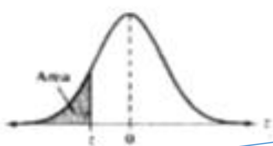
Make a conclusion

Since $\chi^2 < 7.815$, we do not reject the null hypothesis, that number of days spent exercising per week is independent of gender.

Title	Utilizing a Standard Normal Table - Probability as Area from Far Left of the Normal Distribution to the Z-Score
Author	-
Date	10/4/2019

Probability as Area from Far Left of the Normal Distribution to the Z-Score

(There are many types of Standard Normal Table out there, check before proceeding. For a different type of Standard Normal Table, consult your teachers, professors or lecturers for help as I cannot accommodate to all the possible types with limited resources.)



2nd decimal place

Whole Number and 1st Decimal Place

z	.09	.08	.07	.06	.05	.04	.03	.02	.01	.00
−3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
−3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
−3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
−3.1	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
−3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
−2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
−2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
−2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
−2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047
−2.5	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062
−2.4	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082
−2.3	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107
−2.2	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139
−2.1	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179
−2.0	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228
−1.9	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287
−1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359
−1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446
−1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548
−1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668
−1.4	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808
−1.3	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968
−1.2	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151
−1.1	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357
−1.0	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587
−0.9	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841
−0.8	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119
−0.7	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420
−0.6	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743
−0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085
−0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446
−0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821
−0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207
−0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602
−0.0	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000

Objective	Instructions
Probability for Which Z-score ≤ 0	Look up the whole number and first decimal place, then, look up the second decimal place and take the probability as shown in the table. Example: To find the probability value of z-score ≤ -1.5, look up the row “-1.5” and look up the column “0.00” for the z-score value, which turns out to be 0.0668
Probability for Which Z-score > 0	Look up the whole number and first decimal place of the negative counterpart, then, look up the second decimal place. Deduct the value from 1 to get the probability value. Example: To compute the probability value of z-score $\leq$1.33, you search for the probability value for -1.33 which is 0.0918 and deduct the value from 1 to get 0.9082
Z_c of Confidence Interval	Obtain the confidence interval value and the corresponding probability value, then find the value of z_c (z-score of confidence interval) Example: If question wants 95% confidence interval. The corresponding p-value is $0.95 + \frac{1-0.95}{2} = 0.95 + 0.025 = 0.975$ General Formula $p = \ c + \frac{1-c}{2}$ Deduct p-value from 1 to get, 0.025 $p - value\ of\ 0.025$ on the standard normal table corresponds to $z = -1.96$ Therefore $z_c = 1.96$

$z - score$ for 2-tailed test	Compute $\frac{\alpha}{2}$ and find the z-score for $\frac{\alpha}{2}$ If the question wants a significance level of $\alpha = 0.05$ Compute $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ Find the z-score corresponding to $p = 0.0025$ Value turns out to be -1.96, therefore $H_0: -1.96 \leq z \leq 1.96$ $H_\alpha: z < -1.96\ OR\ Z > 1.96$