Learning Probability and Statistics Through Tutorial Questions

Qian Kemao, CCDS, NTU (drafted when doing tutorials for 2025S2 FCE1/FCEE groups)

Note: (1) For internal use and teaching purpose only, as I do not have the copyright of a few images at this moment; (2) For your reference only. Yet to refine; (3) If there is anything wrong, do let me know. Many thanks.

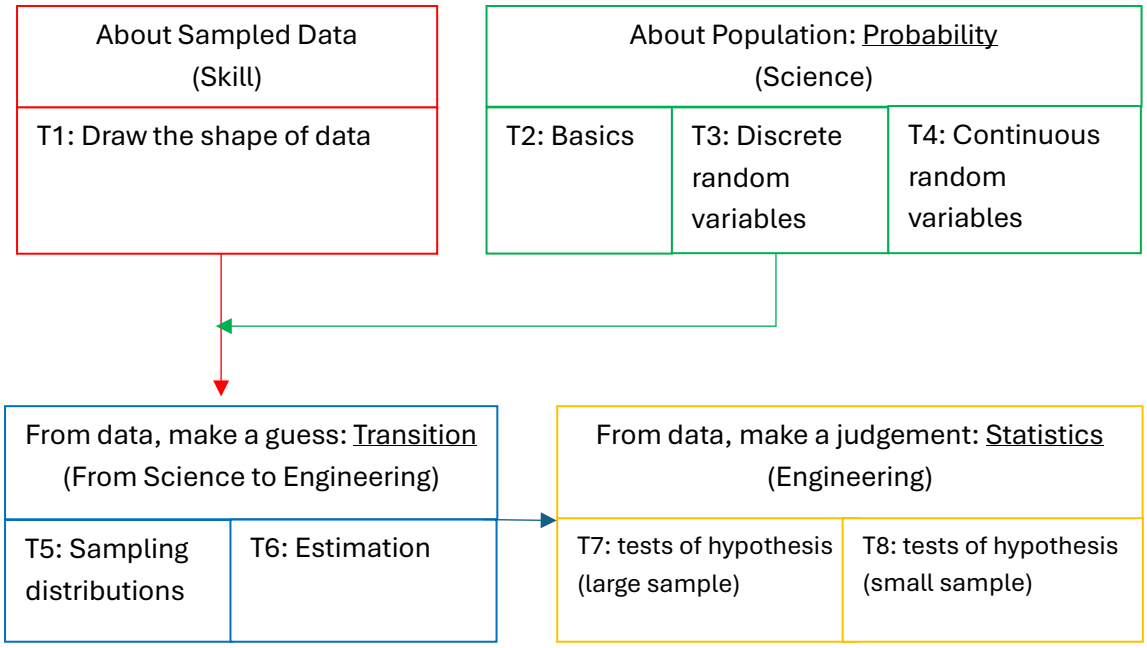
Table of Contents

Introduction

The best approach of learning, in my opinion, is reading a textbook, understanding the concepts, and practising a few (not many) exercises. Probability and Statistics is a fundamental and useful course – we should not fully rely on AI to tell us answers.

Since I am teaching tutorial classes instead of lectures, I prepare this file with a humble hope that at least it is better than none. As a tutor is not allow to give solutions directly, I will just try to do some reasoning and give some hints.

I know you are busy with recorded videos, lecture slides, lecture summaries, consultations, and tutorials. You can safely ignore this file.



Tutorial 1: Draw the shape of data

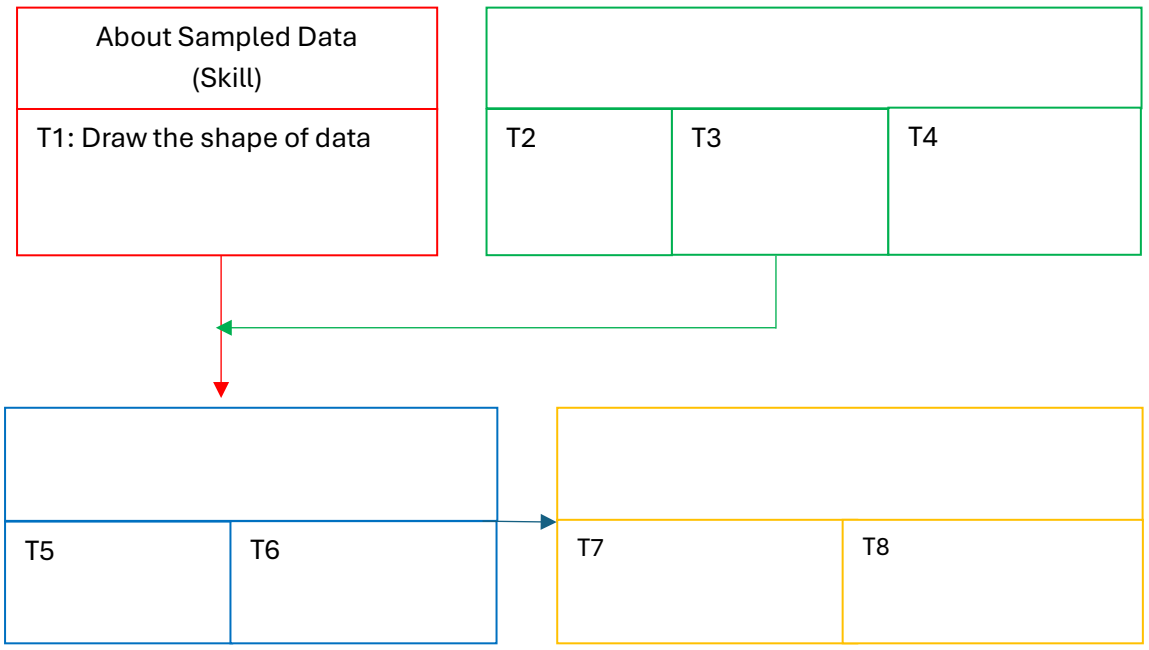
Data are important, but not useful, until the information behind the data is revealed – this is data science (what is our college's name?).

Unfortunately, given data, especially big data, we may just feel dizzy. What can we do then? The most straightforward way is to draw the shape of data, which is the content of this tutorial. Artists draw the shape of wind, romantically. We draw the shape of data, understand the world, improve the life quality, bravely embrace the future – scientists and engineers can be romantic as well, just in a different way – you can share this with your boyfriend/girlfriend.

Let's create something together for the given data.

- Get an impression of how data gather and disperse: average (mean) and variance (or standard deviation)
- Order data from small to large, and cut the number range into groups/bins, check how many data in each group/bin: histogram
- Again, order data from small to large, and check the cut-off lines at 75%, 50%, 25%, roughly you can guess how good your GPA is in your cohort :). We make it a little more comprehensive: percentile and box plot

Isn't it simple? As if we can also invent them. If you think so, congratulations, you have good mastering of this tutorial.



1.1: type of variables

Question 1: Types of Variables

Categorise each of the following variables as **qualitative** or **quantitative**, and state the **level of measurement** (nominal, ordinal, interval, or ratio).

(a) Hair colour

(b) Number of television sets in a private home

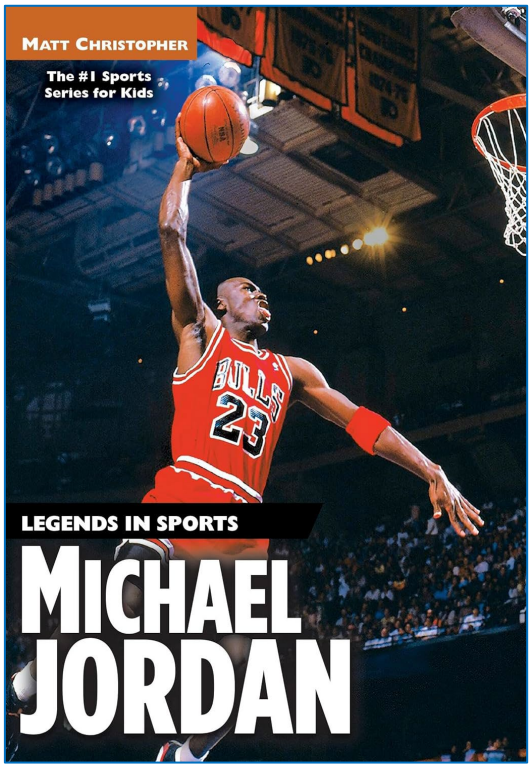
(c) Responses to a questionnaire item: strongly agree, agree, neutral, disagree, strongly disagree

(d) Intelligence score on a scale from 0 to 100
Briefly justify your answer.

(e) Country of birth

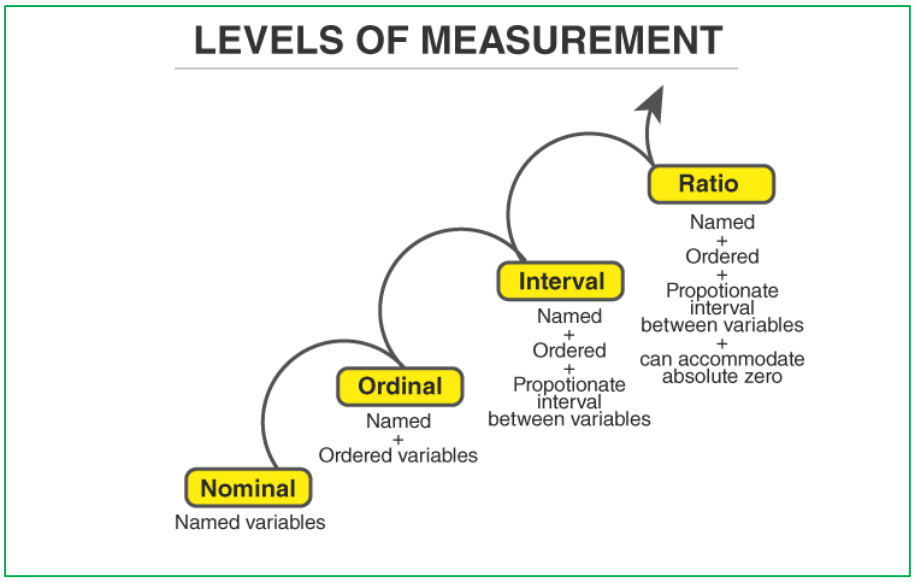
1.1.1 Qualitative vs quantitative

As a rule of thumb, qualitative is about naming, while quantitative is about amount. For example, Michael Jordan's sport shirt number is 23, which is equivalent to his name, so it is qualitative. It does not provide any sense of counting.



1.1.2 Nominal vs ordinal vs interval vs ratio

- Nominal is for naming; [Q1(a), Q1(e)]
- Ordinal means we can "further" order them; [Q1(c)]
- Interval means we can "further" count them; [Q1(d), no true zero]
- Ratio means that we "further" have a true base/zero for counting. [Q1(b)]



(Image source: byjus.com)

1.2 Describe your data

Question 2: Fuel Droplet Size Data

The following data represent measurements of fuel droplet sizes:

2.1	2.2	2.2	2.3	2.3	2.4	2.5	2.5	2.5	2.8
2.8	2.9	2.9	3.0	3.1	3.1	3.2	3.3	3.3	3.3
3.4	3.5	3.6	3.6	3.6	3.7	3.7	4.0	4.2	4.5
4.9	5.1	5.3	5.3	5.7	6.0	6.1	7.1	7.8	7.9
7.9	7.9	8.9							

(a) Construct a frequency table using class intervals of width 1:

$$[2,3),[3,4),[4,5),\ldots$$

Then draw a frequency histogram and comment on the shape of the distribution.

(b) Construct a stem-and-leaf diagram.
Use the integer part as the stem and the first decimal place as the leaf.

(c) Compute the following summary statistics:

- Mean
- 25th percentile
- 50th percentile (median)
- 75th percentile

Use the percentile definition taught in lecture:

$$R = \frac{P}{100}(N-1)+1$$

(d) Construct a box plot for the data. Clearly label:

- Minimum and maximum
- Median
- Lower and upper hinges
- Any outliers (if present)

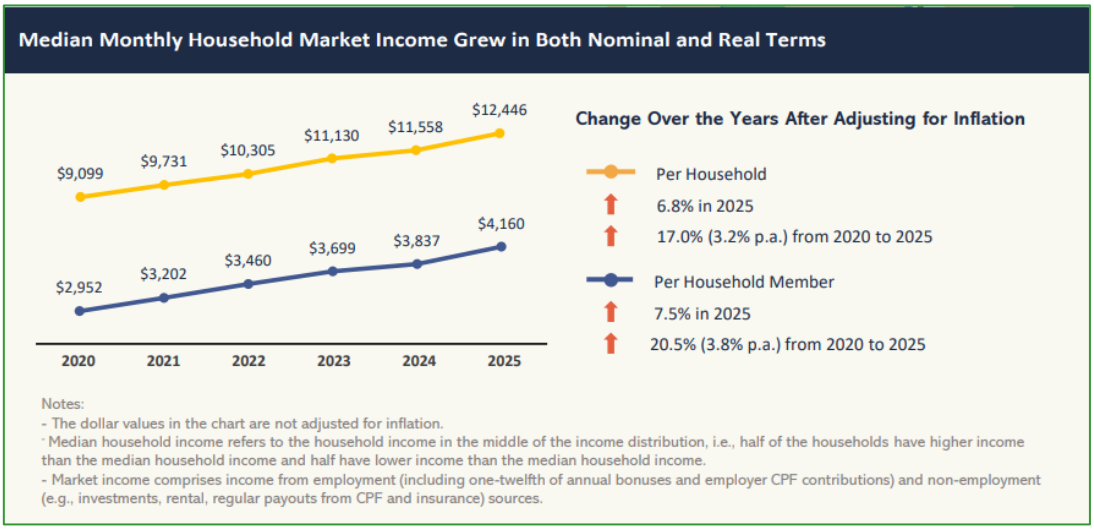
1.2.1 Histogram vs stem-and-leaf

This is a good question, which contains many important operations and descriptors.

- Ordering: Note that the data have been ordered. If not, you should be aware that you need to order them.
- Grouping data and drawing the histogram: that is question 2(a).

- Stem-and-leaf diagram: this is essentially very similar to the histogram (can you see this point?) However, with the popularity of computers, histogram is dominantly used.

1.2.2 Mean vs median



(Image source: https://www.singstat.gov.sg/)

- As a fact, you will often see "median income", rarely "mean income". Why? [Consider 10 people, 9 earn 3K per month, and 1 earns 3M per month.]
- In the future when you learn signal/image analysis, you will use both mean filter and median filter, with respective advantages and disadvantages. Curious about it? Remember to take Computer Vision (SC4061).

1.2.3 Percentile (How to turn our known knowledge into new knowledge?)

Let's start with an example. Ther are N students, each has a GPA, ordered from lowest to highest, i.e., 1st, 2nd, ..., N-th. The 1st student has the lowest GPA, so it is top 0%, sorry :(. The N-th student has the highest GPA, so it is top 100%, congratulations :). I want to find where is the cut-point for top P%.

1-st	R-th	N-th
0%	P%	100%

We call the first row as absolute position, and second row as relative position. We build a simple linear relationship between them:

$$\frac{R-1}{P-0}=\frac{N-1}{100-0}. \qquad \text{Eq. (1.1)}$$

We immediately get

$$R = \frac{P}{100}(N-1) + 1\,. \qquad \text{Eq. (1.2)}$$

Take an example of R=76. It means that the 76-th student is top P%, and his GPA is the top P% cut-off line of the GPA.

Unfortunately, after calculation, R is usually a fraction number, say, R=76.1. Then how to determine the cut-off? We do not have a student whose rank is 76.1. That is easy. The cut-off should be a little higher than 76-th student's GPA, but lower than 77-th's student' GPA. It is somewhere in-between. How to find the exact value?

We use linear mapping again, for simplicity. We denote the integer part of R as I_R, and the fraction part of R as F_R, then $R = I_R + F_R$, say 76.1=76 + 0.1. We now build up the table below:

(I_R)-st (Eg. 76-th)	R-th (Eg. 76.1)	(I_R +1)-th (Eg. 77-th)
GPA of (I_R -th)	x=GPA of R-th	GPA of (I_R +1)-th

We play the same linear trick again,

$$\frac{R-I_R}{x-(\text{GPA of } I_R)} = \frac{(I_R+1)-I_R}{(\text{GPA of } (I_R+1))-(\text{GPA of } I_R)}\,. \qquad \text{Eq. (1.3)}$$

We then obtain

$$x = (\text{GPA of } I_R) + F_R \times [(\text{GPA of } I_R + 1) - (\text{GPA of } I_R)]\,. \qquad \text{Eq. (1.4)}$$

This is exactly the same as the following material, but we obtained it by ourselves. Now we do not need to memorize it and more importantly, we discover it so that we can demystify it. Learning is not to memorize, it is to demystify, especially in the AI era.

Calculation of P^{th} Percentile for a set of N data:

1. Compute the rank R = P/100 x (N + 1)
2. Let I_R = Integer part of R and F_R = Fractional part of R
3. P^{th} Percentile = Data at rank I_R + (Data at rank (I_R+1) – Data at rank I_R) x F_R

Finally, you may wonder why the linear mapping is selected. This is because a linear relationship is simplest to use. In many applications, if the linear relationship is not

satisfactory, we then consider to elevate it to a quadratic relationship or even a higher order relationship, which is called, in general, a nonlinear relationship. In the old days, nonlinear relationships are difficult to handle, making the linear relationship often the first choice. Tayler expansion is to express a function as a sum of many polynomials with different orders.

Amazingly, the current deep learning easily establishes a complicated and nonlinear relationship between the network input and output through learning. That is essential why deep learning flourishes today. The input-output relationship is expressed by a network structure and its many weights, instead of by a traditional/simple/elegant equation. Many people see this as regretful, but it is actually pretty normal – we just use the power of the numerical digits.

Similar things actually have already happened much earlier. For example, in image processing, we do intensity stretching so that the image looks more vibrant and clearer. Such stretching can be realized by linear/quadratic functions. Histogram equalization is more widely used, and it should be in your handphone camera App. Histogram equalization does not look like a function. Instead, it is a small but effective algorithm.

Why I say so much on linear mapping? First, it is how we start to create a solution to a problem. Second, from a function in mathematics to a network in deep learning, there is a transition. Mathematics without computing is weak; computing without mathematics is superficial. We are learning both!

1.2.4 Box plot details

The left and right figures are copied from two resources. Do you spot any differences in the definitions? Which is correct? (Answer: left)

Upper Hinge = 75th Percentile (3rd quartile)	20
Lower Hinge = 25th Percentile (1st quartile)	17
H-Spread = Upper Hinge - Lower Hinge (IQR)	3
Step = 1.5 x H-Spread	4.5
Upper Inner Fence = Upper Hinge + 1 Step	24.5
Lower Inner Fence = Lower Hinge - 1 Step	12.5

Upper Outer Fence = Upper Hinge + 2 Steps		29
Lower Outer Fence = Lower Hinge - 2 Steps		8
Upper Adjacent = Largest value ≤ Upper Inner Fence		24
Lower Adjacent = Smallest value ≥ Lower Inner Fence		14
Outside Value	Outlier = Value > upper adjacent or < lower adjacent	29
Far Out Value		none

Quantity	Definition	Value
Upper Hinge	75th percentile (Q_3)	5.2
Lower Hinge	25th percentile (Q_1)	2.85
H-Spread	$Q_3 - Q_1$	2.35
Step	1.5× H-Spread	3.525
Upper Inner Fence	$Q_3 + 1.5\times$ H-Spread	8.725
Lower Inner Fence	$Q_1 - 1.5\times$ H-Spread	0*
Upper Outer Fence	$Q_3 + 2\times$ H-Spread	12.25
Lower Outer Fence	$Q_1 - 2\times$ H-Spread	0*
Upper Adjacent	Largest value below UIF	7.9
Lower Adjacent	Smallest value above LIF	2.1
Outside Value	Beyond inner fence only	8.9
Far Out Value	Beyond outer fence	None

1.3 Grouped data

Question 3: Grouped Data

A frequency distribution of the lengths of telephone calls monitored at an office switchboard is given below:

Length of Calls (minutes)	Number of Calls
0 and under 2	10
2 and under 4	25
4 and under 6	20
6 and under 8	40
8 and under 10	5
Total	100

(a) Use class midpoints to estimate the mean call duration.

(b) State the modal class and give the approximate mode.

Within one group, say [0, 2), we may have different values like 0.1, 0.7, 1.8 etc. They are all "treated" as equal and represented by the midpoint of [0, 2), which is 1. Apparently, it is not good to arbitrarily modify the data. However, people doing statistics can have a much easier job.

Since we are computing students, we usually do not need such convenience, because we have a computer and we know how to program.

1.4 Just for fun

Question 4: Variance and Feasibility

A dataset has the following summary values:

$$n = 10, \quad \sum x = 50, \quad \sum x^2 = 500$$

(a) Compute the sample standard deviation.

(b) Explain why it is impossible for a dataset to have

$$n = 10, \quad \sum x = 50, \quad \sum x^2 = 100$$

1.4.1 Homework

There is a hint to use standard deviation to show (b) is impossible, which is a smart solution. However, I think using the following inequality for (b) is more elegant. (Hint: take **u** as the vector with the data in the question and **v** as a unit vector.)

The Cauchy–Schwarz inequality states that for all vectors $\mathbf{u}$ and $\mathbf{v}$ of an inner product space

$$|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \cdot \langle \mathbf{v}, \mathbf{v} \rangle,$$

1.5 Two datasets: Are they correlated?

Question 5: Bivariate Data

The following paired data are observed:

X	2.5	3.4	5.6	6.7	7.9
Y	10.2	11.8	13.7	19.7	20.6

(a) Compute the Pearson correlation coefficient between X and Y.

(b) Comment on the strength and direction of the relationship.

1.5.1 The concept of correlation

In this example, when X increases, Y decisively follows the increase. They should have a high correlation. In fact, the correlation value is as high as 0.955 (the highest possible is 1).

Let's take a look at the definition:

- Pearson Correlation ρ

  - An indicator on the strength of the linear relationship between two variables.

Definition: $\rho = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$ Covariance of X and Y, denoted as cov(XY)

$$= \frac{E[XY]-\mu_X\mu_Y}{\sqrt{E[X^2]-(\mu_X)^2}\sqrt{E[Y^2]-(\mu_Y)^2}}$$

$$= \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}}$$

If $\mu_X = \mu_Y = 0$, then $\rho = \frac{\sum XY}{\sqrt{\sum X^2}\sqrt{\sum Y^2}}$

Think for a while until the definition is clear and natural to you.

- The simplest possible definition is just E(XY) to check whether Y is following X. Why?
- In the numerator, the mean values of X and Y are deducted, why? [Answer: By doing so, X+a and Y+b (where a and b are constants) have the same correlation level as X and Y];
- In the denominator, the standard deviations are divided, why? [Answer: By doing so, c*X and d*Y have the same correlation level as X and Y.];
- The correlation value will be within [-1 1];
- If you understand the above points, now you should be able to write them out on a paper easily. Try!

1.5.2 Application in convolutional neural network

Think: if two datasets are the same or very similar, what is the correlation value? [Answer: 1 or nearly 1].

What does it mean? It means that, if you have a dataset A, from many other datasets, you can easily find the one is similar to dataset A by computing the correlation value! And correlation and convolution are essentially the same, which will be explained in the course SC4061.

Now we generate as small image of a nose, and call it image A. We take another large image B. We take a patch in image B, and do correlation with image A, we can decide whether there is a nose in the patch. We repeat it by correlating A with any possible patches in image B, we can decide whether there is a nose in image B. Isn't it amazing? In fact, this is the most fundamental component of a convolutional neural network! It should be interesting for you to note that, first, the small image is call a kernel; and second, the kernel is not pre-defined; rather, it will be generated through learning.

1.6 Two datasets: Merge into one

6. Combining Two Samples (Mean & Standard Deviation)

In an attempt to find the mean number of hours his tutorial classmates spent per day preparing for tutorials, John collected data from 10 of his friends in the tutorial group and found that the mean is $\bar{x}_1 = 2.4$ hours with a **sample** standard deviation $s_1 = 0.8$ hours.

A day later he felt that the sample size is too small. So he collected data from another 5 of his friends and found that the mean is $\bar{x}_2 = 2.0$ hours with a **sample** standard deviation $s_2 = 1.2$ hours.

Find the mean and the **sample** standard deviation when these 2 sets of data are combined.

(Clarification: Treat s_1 and s_2 as sample standard deviations, i.e. computed with denominator $n - 1$.)

Such a question can be easily tested in quiz so you must practice it. Do not look for a formula! Do not arbitrarily create a formula by yourself without thinking! (Unfortunately, many students told me that they did this.) Use definitions!!!

Dataset1: mean1 and std1 (known): express them by definition please. (denoted as A)

Dataset2: mean2 and std2 (known): express them by definition please. (denoted as B)

Dataset3 (combined): mean3 and std3 (unknown): express them by definition please. (denoted as C)

Now, just rearrange the expressions in C, and try to utilize something from A and B. You can do it!

The highlight of this question from me is the fundamental understanding. Formulas have little value, but understanding them is invaluable.

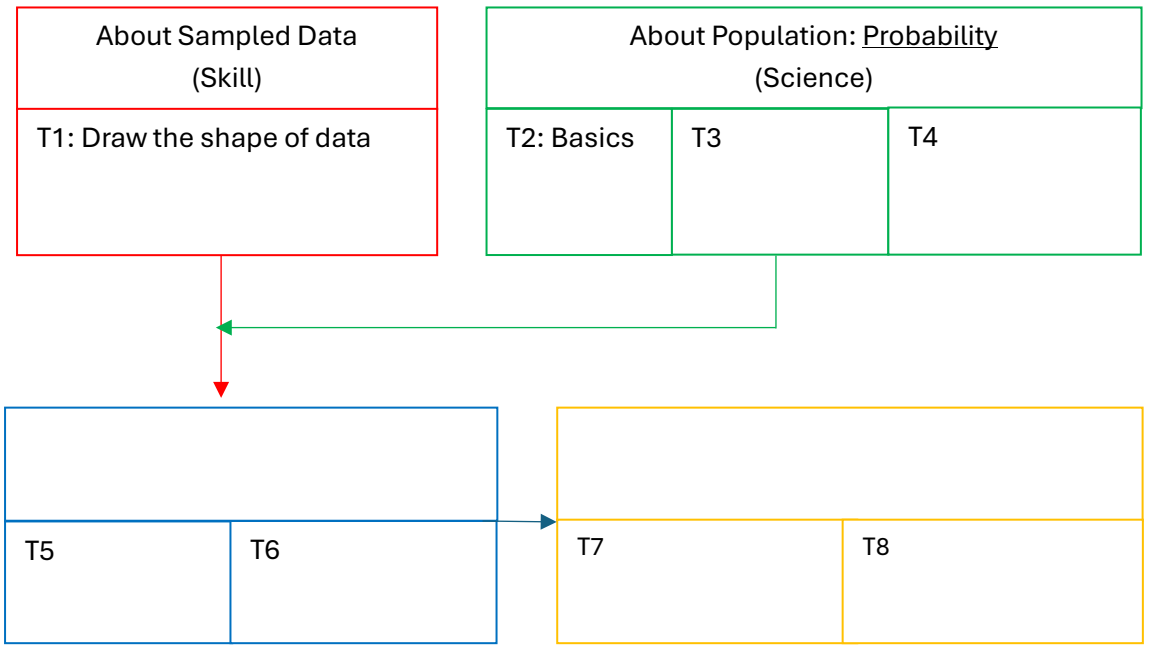
Tutorial 2: Probability Basics

In tutorial 1, we do it directly and *passively* – we just illustrate the data. In this tutorial, we will understand data in a more *active* way.

This was how we started in tutorial 1: "Given data, …" Wait! How are the data generated? What is the set of data about? Where and under what condition did you acquire them? What is the essential property of data?

Under our context, data include two parts, one is deterministic, and the other is random.

- Deterministic/predictable: For example, we have learnt Ohm's law, V (Voltage) = I (Current) × R (Resistance). Once I and R are avaialbe, I can calculate V. Revealing such theorems is the task of physics or other courses.
- Random/unpredictable: Let's directly measure the voltage V for 10 times. You will obtain 10 different readings. Why? Because there will be inevitable distances. Studying the uncertainty is the task of this module.
- "Unpredictable" does not "not understandable". The famous "bell curve" of the examination result is a good example. Such interesting things are worth observing, summarizing, analysing, and utilizing. The probability is our focus. We now look at population, i.e., all data. We may not be able to have all data, but we just assume that we have them. This part consists of three tutorials and has a strong mathematical flavour.



2.1 Permutation and Combination

Question 1: Sample Space and Counting

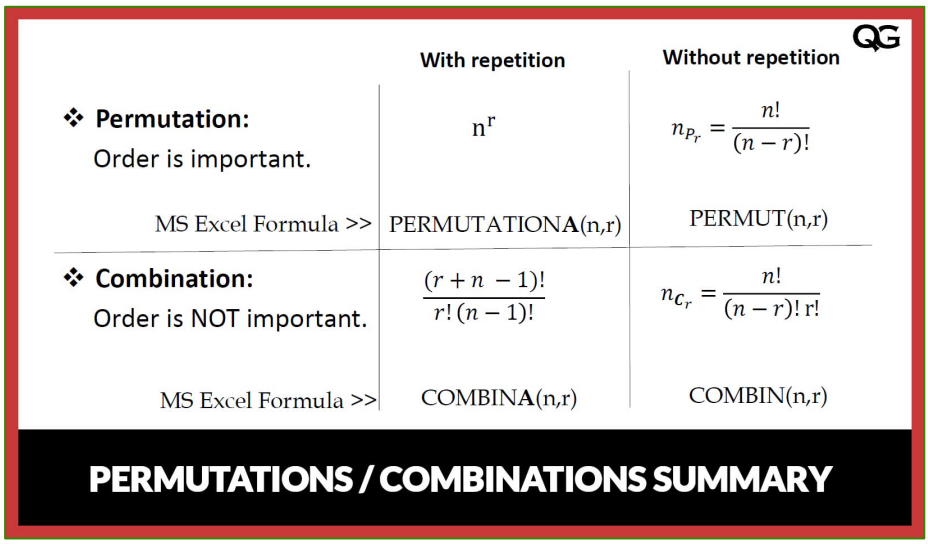
A jar contains four distinct coins: a nickel (5¢), a dime (10¢), a quarter (25¢), and a half-dollar (50¢). Three coins are randomly selected **without replacement**.

(a) Consider outcomes as **ordered triples** (i.e. the order of selection matters). Describe the sample space S and compute its size $|S|$.

(b) Let A be the event that the total value of the three selected coins is at least 60¢. Compute $\mathbb{P}(A)$.

(c) Now consider outcomes as **unordered sets** (i.e. only which 3 coins are chosen matters). How many outcomes are there now? Briefly explain.

- When we think about probability, a dice pops up from our minds, from books, from everywhere. Permutation and combination follow. We learned those before, please:
  - Review them by yourself;
  - come up with a summary table regarding permutation/combination and with/without replacement. I copy a sample table from internet. You need to come up with your own version;
  - Do some questions. Only doing this question here is far from enough.



(Image source: www.qualitygurus.com)

- Q1(a) is about permutation and Q1(c) is about combination. Doing Q1(b) based on Q1(c) is easier than doing Q1(b) based on Q1(a), why?
- Doing Q1(a) based on Q1(c) is also possible, how?

2.2 Sample space

Question 2: Limited Supply and Random Choices

A mother prepares 9 popsicles: 3 orange (O), 3 cherry (C), and 3 grape (G). Four children each independently choose one flavour from $\{O, C, G\}$, with each flavour chosen with equal probability.

If more children choose a particular flavour than the number of available popsicles of that flavour, the excess children do not receive a popsicle.

(a) What is the probability that all four children successfully receive a popsicle?

(b) Explain briefly why this probability is relatively large.

2.2.1 Definitions (super-important!)

Definitions

- A sample space is the set S of all possible outcomes of an experiment.
- An event is a set of one or more (favorable) outcomes in the sample space.
- Two events are mutually exclusive if they have no outcomes in common.
- They are exhaustive if they cover all possible outcomes.
- Two events are independent if the probability that one occurs is not affected by whether or not the other has occurred.

2.2.2 Solving the question simply by definition

Think:

- What is the concerned event in this question? (Answer: Children's wishes that come true)
- What is the complement of the above event? (Answer: Children's wishes that do not come true, or failed)
- What is the sample space? (Answer: put both the above outcomes together: successful wishes + failed wishes = all wishes)

Do (Now it is very easy):

- How many wishes in the sample space? (Answer: 3*3*3*3=81)
- How many wishes in the complement event? (3)
- How many wishes in the concerned event? (81-3)=78.

2.3 Conditional probability

Question 3: Mutually Exclusive and Independent Events

Recall the following facts:

- Events A and B are **mutually exclusive** if they cannot occur together, i.e.

$$P(A \cap B) = 0 \quad \text{and} \quad P(A|B) = 0.$$

  In this case, the occurrence of one event affects the probability of the other.

- Events A and B are **independent** if the occurrence of one event does not affect the probability of the other, i.e.

$$P(A|B) = P(A) \quad \text{and} \quad P(A \cap B) = P(A)P(B).$$

Using the relationships above, complete the following table.

$P(A)$	$P(B)$	Condition	$P(A\|B)$	$P(A \cap B)$	$P(A \cup B)$
0.3	0.4	mutually exclusive			
0.3	0.4		0.3		
0.1	0.5				0.6
0.2	0.5			0.1	

2.3.1 Conditional probability

Conditional probability is an important concept. Think:

- Suppose that I airdrop you at Jurong Point. Open your eyes, what is the probability that the first person you see is a female? Why? [Answer: 0.5]
- Now I airdrop you at an NTU basketball court. Open your eyes, what is the probability that the first person you see is a female? [Answer: <<0.5 according to my observation when I lived in campus. More boys play basketball than girls.]
- Why the two probabilities are different? Because the conditions are different – one is at Jurong Point, and the other is at an NTU basketball court. We call this as conditional probability. This is how we express them:
  - P(female | JP) = 0.5
  - P(female | NTU basketball court) << 0.5
- An important rule:

$$P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(A|B) \quad \text{Eq. (2.1)}$$

Explanation: the following mean the same:

- $P(A \cap B)$: Both events A and B happen;
- $P(A) \times P(B|A)$: A happens, and under the condition that A has already happened, B also happens;
- $P(B) \times P(A|B)$: B happens, and under the condition that B has already happened, A also happens.

2.3.2 Mutually exclusive

- What does it mean? One happens, the other will then not happen. So they will not "both happen". So

$$P(A \cap B) = 0 \quad \text{Eq. (2.2)}$$

- Under the condition that A has already happened, then B will not happen. Or B happened, then A will not happen,

$$P(B|A) = P(A|B) = 0 \quad \text{Eq. (2.3)}$$

2.3.3 Independent

- What does it mean? Whether one happens or not is irrelevant to the other. So

$$P(A|B) = P(A),\ P(B|A) = P(B) \quad \text{Eq. (2.4)}$$

- By referring back to Eq. (2.1), we have

$$P(A \cap B) = P(A) \times P(B) \quad \text{Eq. (2.5)}$$

These concepts and equations should be as natural as you like bubble tea! Yes, math is as sweet and enjoyable as bubble tea!

2.4 Bayes theory

The original version of the question

A blood disease is found in 2% of the persons in a certain population. A new blood test will correctly identify 96% of the persons with the disease and 94% of the persons without the disease.

(a) What is the probability that a person who is called positive by the blood test actually has the disease?
(b) What is the probability that a person who is called negative by the blood test actually does not have the disease?
(c) Comment on the results obtained in part (a) & (b).

The new version of the question

Question 4: Bayes' Theorem and Medical Testing

A blood disease is present in 2% of a certain population. A diagnostic blood test has the following properties:

- Sensitivity: $P(T^+ \mid D) = 0.96$
- Specificity: $P(T^- \mid D^c) = 0.94$

Here, D denotes the event that a person has the disease, and T^+ (T^-) denotes a positive (negative) test result.

(a) What is the probability that a person who tests positive actually has the disease?

(b) What is the probability that a person who tests negative actually does not have the disease?

(c) Comment briefly on the results obtained in parts (a) and (b).

The question has been rephrased.

2.4.1 Be able to translate words into mathematical expressions

Compare the two versions of the question, the translation has been done for you. However, you need to learn to do the translation by yourself. One step further, if you see a real-life problem, can you model it into a mathematical problem – this is a super-important capability you should learn to have. The purpose of learning math is to solve real-life problems.

2.4.2 Why Bayes?

There are two conditional probabilities (I continue to use the symbols I used in class: D for disease; Po for positive testing result; P for probability; an overhead bar to indicate complement):

$P(Po \mid D)$: Under the condition that a person has already been diagnosed the disease earlier, what is the probability that the current test is positive. This probability is easy to obtain (go to hospital, recruit volunteer patients who have already been diagonalized, do the test on them). The volunteers do not care about the result, as they have already been diagonalized. Test kit manufacturers and hospitals care about the results.

$P(D \mid Po)$: Under the condition that the test is positive, what is the probability the person indeed has the disease. Apparently, this probability is difficult to get, but the patients really care about it!!!

Let's summarize the discussions: $P(Po \mid D)$ is easy to get but we don't care; $P(D \mid Po)$ is difficult to get but we care. So is it possible that we obtain $P(Po \mid D)$ first, and use it to exchange for $P(D \mid Po)$? Yes, that is Bayes.

2.4.3 How to do Bayes?

We have known from the above discussion that Bayes is all about $P(Po \mid D)$ and $P(Po \mid D)$, a game of swapping. However, $P(D \mid P) \neq P(P \mid D)$, but recall that $P(D \cap P) = P(D \cap P)$. That is sufficient for us to re-invent Bayes.

$P(D \mid P)$

$= \frac{P(D \cap P)}{P(P)}$ [Using a property of conditional probability]

$= \frac{P(P \cap D)}{P(P \cap D) + P(P \cap \bar{\bar{D}})}$ [1, now we swap in the numerator; 2, try to understand what I do in the denominator, 3, an overhead bar indicates complement]

$= \frac{P(P \mid D)P(D)}{P(P \mid D)P(D) + P(P \mid \bar{\bar{D}})P(\bar{\bar{D}})}$ [Using the property of conditional probability again]

I skip Q5, because it is EXACTLY the same! Just remember the following steps:

(1) carefully do the translation;

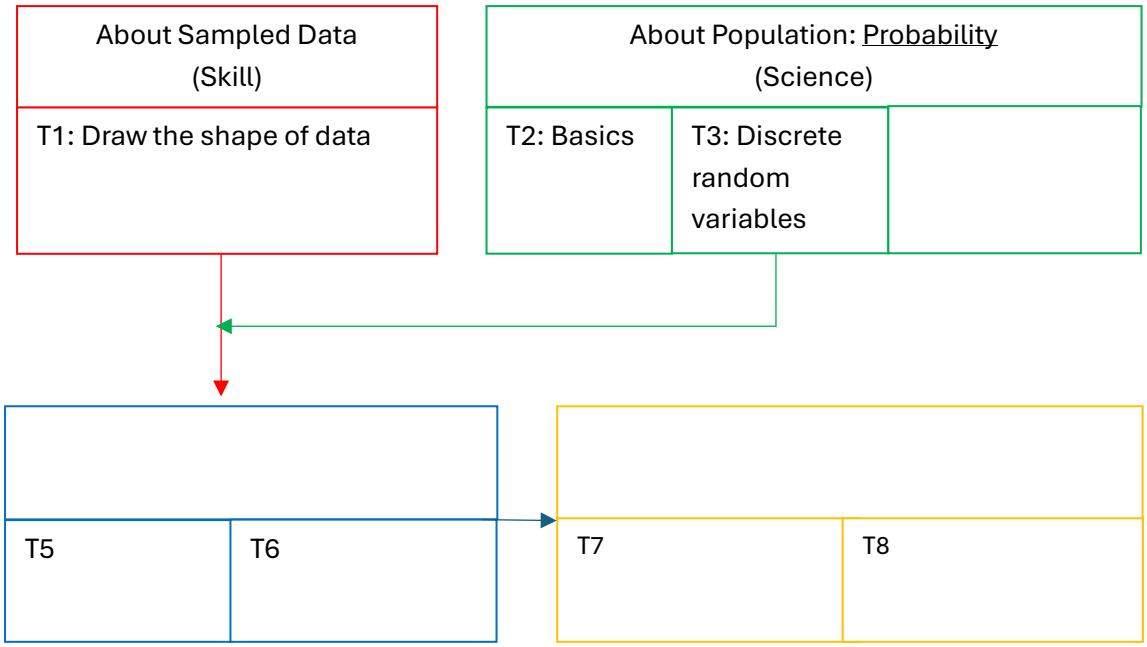
(2) carefully identify the probability (whether it is general or conditional);

(3) play the swapping game (try to derive the equation step by step, as what I did just now).

Tutorial 3: Discrete Random Variables

We need to identify the concerned random variable (RV). I just saw news that one lady passenger, who had an medical appointment with a renowned doctor in a different city, was not able to board an airplane although she bought the ticket - The airplane company over-sold tickets. Many people thought and commented that overselling was wrong, but actually overselling is actually a common practice. Some guests will not take the flight after buying the tickets due to unexpected urgent arrangement (eg. he/she is sick, or has an urgent meeting etc). The final passenger number is uncertain and cannot be predicted, and is a discrete RV. How to oversell scientifically is a probability question. How to take care of the small probability that the final number of passenger exceeds the limit is an administration problem.

Usually, after knowing the RV, we want to further understand its probability distribution. Histogram is about a particular dataset or a sample set (we can manage to know it), probability distribution is about the whole population (usually only God knows). But it does not mean that we cannot do anything. We can do reasoning or accumulate experiences to get a good estimation. For example, if the dice is fair, the number we draw is uniformity distributed.



3.1 Some important definitions for pupation

Question 1: Discrete Probability Distribution

Let X denote the number of wills filed on a given day at Kusu Island. Suppose that the probability distribution of X is given by:

$$P(X = x) = \begin{cases} 0.2, & x = 0, \\ 0.4, & x = 1, \\ 0.3, & x = 2, \\ 0.1, & x = 3. \end{cases}$$

(a) Find the probability that at least two wills are filed on a given day.

(b) Find the expected number of wills filed per day.

(c) Find the variance of the number of wills filed per day.

1. Random variable (a variable, but its values are taken with randomness): number of wills here.
2. Some students asked me about the definition of expected value of a discrete random variable – it is so fundamental that you should find it out by yourself. However, due to its importance, I give it here with explanations:

$$\mu = E(X) = \sum_x x f(x)\,. \qquad \text{Eq. (3.1A)}$$

Where f(x) is the probability distribution.

- The expected value (or population mean) is very similar to the sample mean. So let's start from the sample mean:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \qquad \text{Eq. (3.1B)}$$

- Let's have a sample data of {1,2,2,3,3,3,4,4,4,4}, we can immediately calculate the sample mean as follows:

$$\bar{x} = \frac{1+2+2+3+3+3+4+4+4+4}{10} = \frac{30}{10} = 3 \qquad \text{Eq. (3.1C)}$$

- Can we have a smarter way to do this? Well, we have one 1's, two 2's, three 3's and four 4's, right? Actual your counting is the histogram (please recall).

$$\bar{x} = \frac{1*1+2*2+3*3+4*4}{10} = \frac{1+4+9+16}{10} = \frac{30}{10} = 6 \qquad \text{Eq. (3.1D)}$$

Here, the numbers in red, namely, 4, 3, 2, 1, are called frequencies.

- Now, we do it differently. Don't blink your eyes!

$$\bar{x} = 1 * (\frac{1}{10}) + 2 * (\frac{2}{10}) + 3 * (\frac{3}{10}) + 4 * (\frac{4}{10}) = 3 \qquad \text{Eq. (3.1E)}$$

Here, $(\frac{4}{10})$, $(\frac{3}{10})$, $(\frac{2}{10})$, $(\frac{1}{10})$ are often called probability. However, I think this is not strict. Probability is used for population, but here we are dealing with a sample. Nevertheless, the last equation sounds appealing.

- Now, imagine that we work on the population, not a sample. This time, the probability is strictly the probability in our course title. If we express the probability as f(x), we can move from Eq. (3.2E) to Eq. (3.1A). Accordingly, we are doing population mean, or called expected value. Please remember this definition!

3. Variance is defined as follows, which is more intuitive. Please understand the meaning:

$$V = E\{[X - E(X)]^2\} \qquad \text{Eq. (3.2A)}$$

Now, derive into the following form, which is more convenient for computing (I did it in class. Please try):

$$V = E(X^2) - [E(X)]^2 \qquad \text{Eq. (3.2B)}$$

4. After knowing these concepts, Q1 is just a piece of cake for you.

3.2 A trivial property of probability

Question 2: Discrete Probability Function

A discrete random variable X takes values $x = 0, 1, 2, 3, 4$ and has probability mass function

$$f(x) = \frac{k}{2^x}.$$

(a) Find the value of the constant k.

(b) Tabulate the cumulative distribution function $F(x) = P(X \leq x)$.

All the probabilities adds up to 1 (why?):

$$\sum_x f(x) = 1. \qquad \text{Eq. (3.2B)}$$

With this, you have the second piece of cake today :)

3.3 Three probability distributions

Question 3: Binomial Distribution

A biased die is rolled 50 times, and the number of times the outcome **"2"** appears is 10. Assume that the probability of obtaining a **"2"** remains constant for each independent roll and the relative frequency observed is equal to the actual probability that the outcome is **"2"**.

Let X denote the number of times a **"2"** appears in the next 10 rolls.

(a) Find the probability that a **"2"** appears exactly three times.

(b) Find the expected value of X.

(c) Find the variance of X.

Question 4: Poisson Distribution

The number of calls arriving per minute at a hotel reservation center follows a Poisson distribution with mean 3.

(a) Find the probability that no calls arrive in a given one-minute period.

(b) Assuming that the numbers of calls arriving in different minutes are independent, find the probability that at least two calls arrive in a two-minute period.

Question 5: Geometric Distribution

The probability that a student fails Subject A in a given attempt is 0.05. If the student fails an attempt, he must re-take the subject in the following semester.

Let X denote the number of attempts required for the student to pass the subject for the first time.

(a) Determine the probability distribution of X and name the distribution.

(b) Find the probability that the student passes the subject with no more than 2 attempts.

(c) Find the expected number of attempts required to pass the subject.

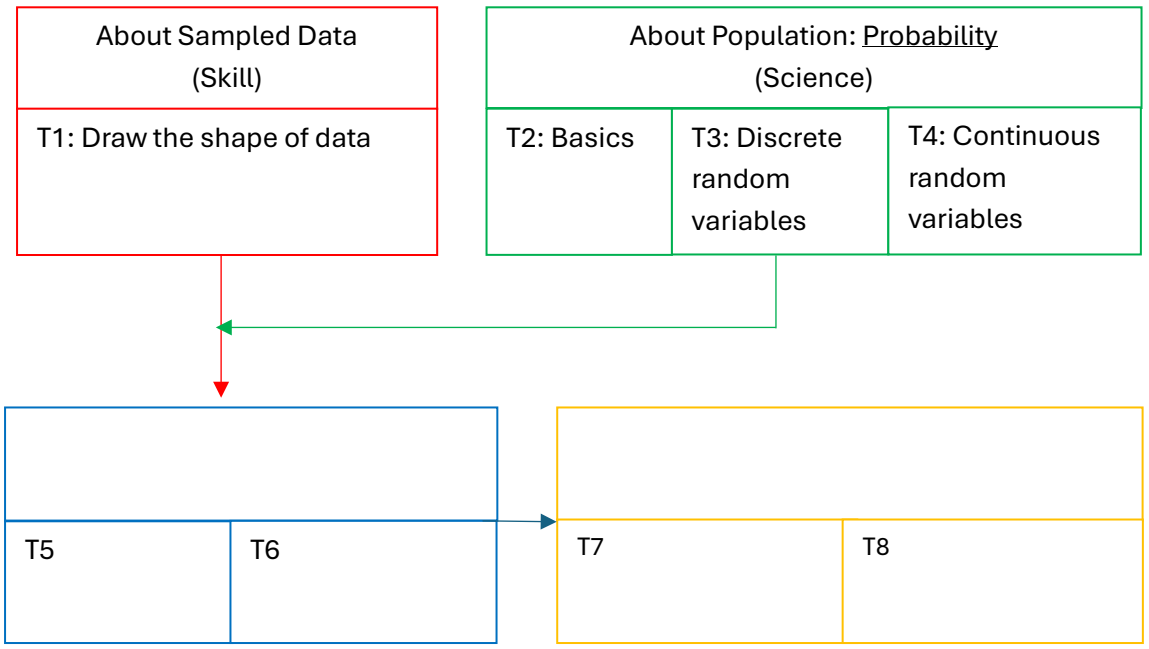
Now, we know f(x) is important. However, if there are many functions f(x), it is still troublesome. Interestingly, many RVs share probability functions governed by only one or two parameters. The three common functions are listed in the Table below, where the properties of each distribution are highlighted. If there are anything wrong in the table, do let me know immediately.

	Binomial	Geometric	Poisson
Condition or properties	1, Bernoulli trial: the result is success/failure. 2, repeated Bernoulli trials: (a) independent; (b) same probability.		1, the probability of a given number of events occurring in a fixed interval of time or area if these events occur with a known constant mean rate and independently of the different time slots or areas; 2, when the interval is short, the occurring probability is proportional to the length of interval; 3, when the interval is very short, the probability of occurring is zero.
RV	The number of successes in n trials	The number of trials x, when the x-th trial is the first success	The number of outcomes occurring in one time interval
Distribution	$P(X=x)=\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x}$	$\mathrm{P}(X=x)=(1-p)^{x-1}\ p$	$P(X=x)=\frac{e^{-\mu}\mu^x}{x!}$
Mean	np	1/p	μ
Variance	npq (note: q=1-p)	q/p^2	μ

Tutorial 4: Continuous Probability Distributions

This chapter is an extension of the previous one, and thus it is easy. Let's clarify just a few points.

- "Discrete" and "continuous" are used to refer to RVs. For example, the car number in a company is discrete, while a person's height is continuous.
- Question: Is exam score a discrete RV or continuous RV?
- For a continuous RV X, given a value x0, P(X=x0) = 0, i.e., it is impossible that your height is exactly 1.7m. At beginning, you may feel it is counter-intuitive and ridiculous. The reason is, since X is a continuous random variable, X can take 1.71, or 1.702, or 1.7003, 1.70004, etc. There are infinite number of values to take. Each of them has a zero chance. Instead, if we say that your height is somewhere from 1.69 to 1.71, it is possible. So for continuous probability distribution, we look at probability *density* function (PDF), and the probability is the integration of PDF in an interval. If you understand this point, you have already mastered this tutorial.



4.1 Three PDFs

Question 1: Uniform Distribution

Let $X \sim \text{Uniform}(2, 10)$.

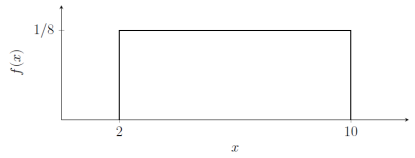


Figure 1: PDF of $X \sim \text{Uniform}(2, 10)$.

(a) Find $P(X < 7)$.

(b) Find $\mathbb{E}[X]$.

(c) Find $\text{Var}(X)$.

Question 2: Exponential Distribution

The waiting time (in minutes) for one to be served in a queueing system follows an exponential distribution with mean 4 minutes. Let X be the waiting time.

(a) State the parameter λ and find $\text{Var}(X)$.

(b) What is the probability that one has to wait for at least 10 minutes before being served?

Question 4: Normal Distribution

Let $X \sim N(\mu = 16.2, \sigma^2 = 1.5625)$. Using the standard normal distribution table:

(a) Find $P(X > 16.8)$.

(b) Find $P(13.6 < X < 18.8)$.

Please summarize a table. I have done one below for your reference only.

	Uniform	**Exponential**	**Normal (Gaussian)**
Condition or Properties	Reflecting equal possibility.	1, If the number of arrivals during an interval is Poisson distributed, then the interarrival times are exponentially distributed 2, Play an important role in both queuing theory and reliability problems	1, The most important PDF in the entire field of statistics; 2, PDF is a bell-shaped curve; 3, central limit theorem: many non-normal random variables adding together could be approximated as a normal distribution; 4, because it is so important, we build a table, which you should know how to use it.
RV	Rare in real life but fundamental in math, and simple and fun to play with.	Time between arrivals at service facilities, and time to failure of component parts and electrical systems	Physical measurements, measurement errors etc are often more than adequately explained with a normal distribution.
Distribution	$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$f(x) = \lambda e^{-\lambda x}$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Mean	$\frac{a+b}{2}$	$\frac{1}{\lambda}$	μ
Variance	$\frac{(a-b)^2}{12}$	$\frac{1}{\lambda^2}$	σ^2

4.2 Normal approximation to binomial

Question 5: Normal Approximation to Binomial

Studies have shown that 22% of all patients taking a certain antibiotic will get a headache. Among $n = 50$ patients taking this antibiotic, let X be the number who get a headache. Use the normal approximation (with continuity correction) to find the probability that:

(a) at least 10 patients will get a headache

(b) at most 15 patients will get a headache.

- The normal distribution is so important that we build a table. Once we have the table, the computing becomes table-searching, which is easy and fast. Thus, we approximate binomial by normal.
- Remember Binomial is for a discrete RV while Normal is for a continuous RV.
- We just find the values for the RV to take and the complement part for the RV NOT to take, and then cut a centre line between two parts. [Do not attempt to remember the "rules", which is difficult and unnecessary. Use my way.]
- Example (a): If RV takes "at least 10 patients", it means that RV takes {10, 11, 12, ….}, and will not take {0, 1, 2, …, 7, 8, 9}. Where do we cut? 9.5!!! So we calculate P(X>=9.5).
- Example (b): We concern "at most 15 patients", which means {0, 1, 2, …, 13, 14, 15}. The complement part is {16, 17, 18, …}. We cut at 15.5. So we calculate P(X<=15.5). Isn't it easy?!

Tutorial 5: From Data, Make a Guess: Sampling Distribution

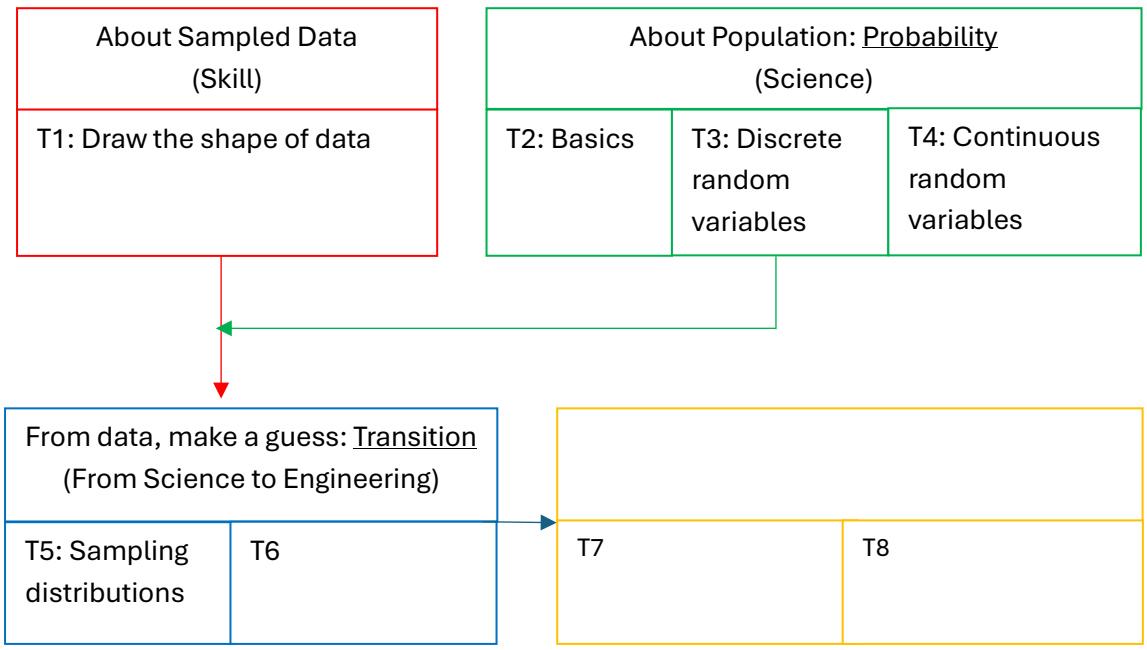
Now we are experts on probability of population. However, properties of population are difficult to know, because the population, which usually has a very large size, is difficult to get. For us, it is much easier to get a subset of population, which is often called a sample. Recall that in Tutorial 1, we can only draw the shape of data at that time. We can now do more.

Here is the first important concept to link up a sample and the population: from a population, we can take many samples. If we calculate one characteristic of a sample x_i $(i = 1, 2, \ldots . n)$, say, sample mean $\bar{x}$, then this sample mean will vary with respect to different samples. Thus, the sample mean $\bar{x}$ is a RV. This concept is called sampling distribution.

The second important concept: central limit theorem. Interestingly, $\bar{x}$ obeys an approximately normal distribution when n is large, with mean μ and standard deviation $\sigma/\sqrt{n}$. Now, you can use $\bar{x}$ to guess (estimate) μ – isn't it amazing and cool?

This tutorial is to answer: (1) what is the RV? (2) under what conditions? (3) what is the distribution? Your job is just to pick the right one to use, which is supposed to be easy.

As you may have seen, our focus is now moved from population to sample. This and next tutorials (T5 and T6) serve as a transition, with which, the last two tutorials (T7 and T8) will be practical.



5.1 Solving questions slowly, step by step

Problem 1

To determine whether a bottling machine is working satisfactorily, a production line manager randomly samples ten 12-ounce bottles every hour and measures the amount of beverage in each bottle. The mean $\bar{x}$ of the 10 fill measurements is used to decide whether to readjust the amount of beverage delivered per bottle by the filling machine.

If records show that the amount of fill per bottle is normally distributed, with a standard deviation of .2 ounce, and if the bottling machine is set to produce a mean fill per bottle of 12.1 ounces, what is the approximate probability that the sample mean $\bar{x}$ of the 10 test bottles is less than 12 ounces?

I have summarized a general five-step procedure to solve most of the questions in T7 and T8.

- Step 1: Read the questions carefully, which is not easy. Read it at least twice.
- Step 2: What is the random variable RV? Sample mean (see the red box).
- Step 3: Under what conditions? Amount per bottle is normally distributed, with the given population mean of 12.1, and population standard deviation of 0.2 (important: σ is known). The sample size is 10 (small). (see the green boxes)
- Step 4: What is the distribution of $\bar{x}$? It is normally distributed, told by the central limit theorem (although the sample size is small, but the original population distribution is normal, the sample mean is also normal). The mean is 12.1, but the standard deviation here is $0.2/\sqrt{10}$.
- Step 5: We compute

$P(\bar{x} < 12)$

$= P\left(Z < \frac{12-12.1}{0.2/\sqrt{10}}\right)$ [Convert to standard normal distribution. Important! Why conversion? There are infinite numbers of normal distributions. You cannot build infinite numbers of tables. There is only one that is standard normal distribution, and luckily, other normal distributions can be easily converted to standard normal distribution.]

$= P(Z < -1.581)$ [Simple calculation]

$= 0.0057$ [get this number from the standard normal distribution table]

The above procedure applies to all the questions in this tutorial, so I just skip all of them. You may build a table for different random variable/conditions/distributions. Below is for your reference.

RV	Condition	RV PDF
Sample mean $\overline{x}$	(Size: $n \geq 30$) & Population PDF: less restrictive OR (Size: $n < 30$) & Population PDF: normal with known σ	Distribution: Normal Mean: population μ Standard deviation: $\sigma/\sqrt{n}$
Proportion $\hat{p}$	Trial: binomial Size: $np > 5$ and $nq > 5$	Distribution: Normal Mean: population p Standard deviation: $\sqrt{pq/n}$

5.2 Add-on Discussions

5.2.1 Number vs Proportion

Q2 is about the total number of acceptance, and the variance is npq; Q3 is about proportion, and the variance is pq/n. Why? [Answer: If an RV X has a variance of V, then another RV Y=X/n has a variance of V/n^2.]

5.2.2 Why Q2 applies correction, while Q3 does not?

In fact, correction should be considered in both cases. In our course, we apply the correction for Q2, since we apparently convert from a discrete RV to a continuous RV. We do not highlight this point for the proportion. Why? [Personally answer: We are concerning proportion (continuous). It is not necessarily to convert it to an integer number of "success" (discrete), and then do correction to convert back to continuous.

5.2.3 Where is the condition (np>5 and nq>5) needed?

(1) When do convert a binomial distribution to normal distribution, this condition is needed, because under this condition, the binomial and normal distributions have a similar shape, so that the conversion is valid;

(2) When we model the proportion distribution as normal, this condition is also needed, with the same reason.

Tutorial 6: From Data, Make a Guess: Estimation

In Tutorial 5, under certain conditions, our RV is normally distributed – this is a strong result, enable us to do an interesting thing called estimation. For example, we assume the population mean μ. If we obtain a large sample with its sample mean as $\bar{x}$, we know that $\bar{x}$ has a high probability to be close to μ.

We now move one step further: it also mean that μ is also close to $\bar{x}$, with a high probability. Do you agree? If yes, we then can use $\bar{x}$ to estimate μ! This is point estimation. If we consider the probability, we then use an interval to estimate μ, make the estimation with a certain level of confidence – usually large interval means your estimation is more conservative, less useful, but more confident. Let me elaborate a little bit. In the last tutorial, we can write

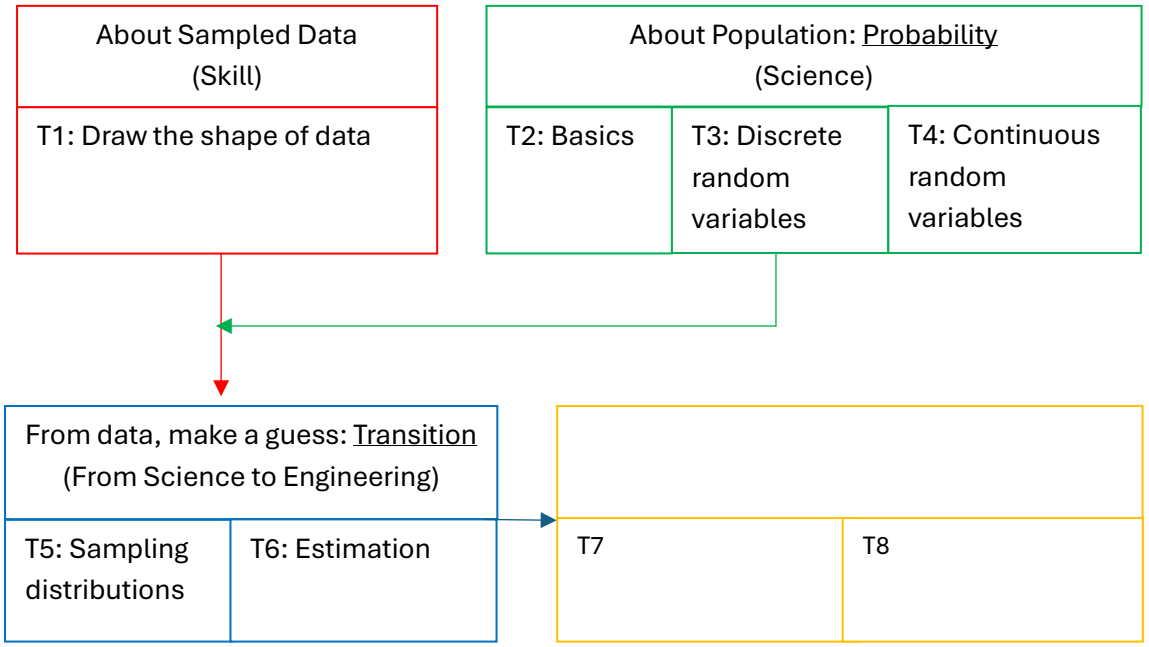
$$\bar{x} = \mu \pm r, \qquad \text{(Meaning: Population determines the sample property)} \quad \text{(Eq. 6.1)}$$

where r is the margin of error. We can now derive that (so amazing and fun that you must try!)

$$\mu = \bar{x} \pm r. \qquad \text{(Meaning: Sample estimates the population property)} \quad \text{(Eq. 6.2)}$$

They look similar but they have different meanings!

This is the only concept we need to know. This chapter is an extension of the last tutorial, so please have a quick review. We will follow the same five-step procedure in solving all the tutorial questions.



6.1 Solving questions slowly, step by step

Problem 4

A survey is designed to estimate the proportion of sports utility vehicles (SUVs) being driven in the state of California. A random sample of 500 registrations is selected from a Department of Motor Vehicles database, and 68 are classified as SUVs. Use a 95% confidence interval to estimate the proportion of SUVs in California.

In tutorial 5, we solve a question on sample mean. In this tutorial, we solve a question on proportion. The rest of the questions are similar ad thus skipped.

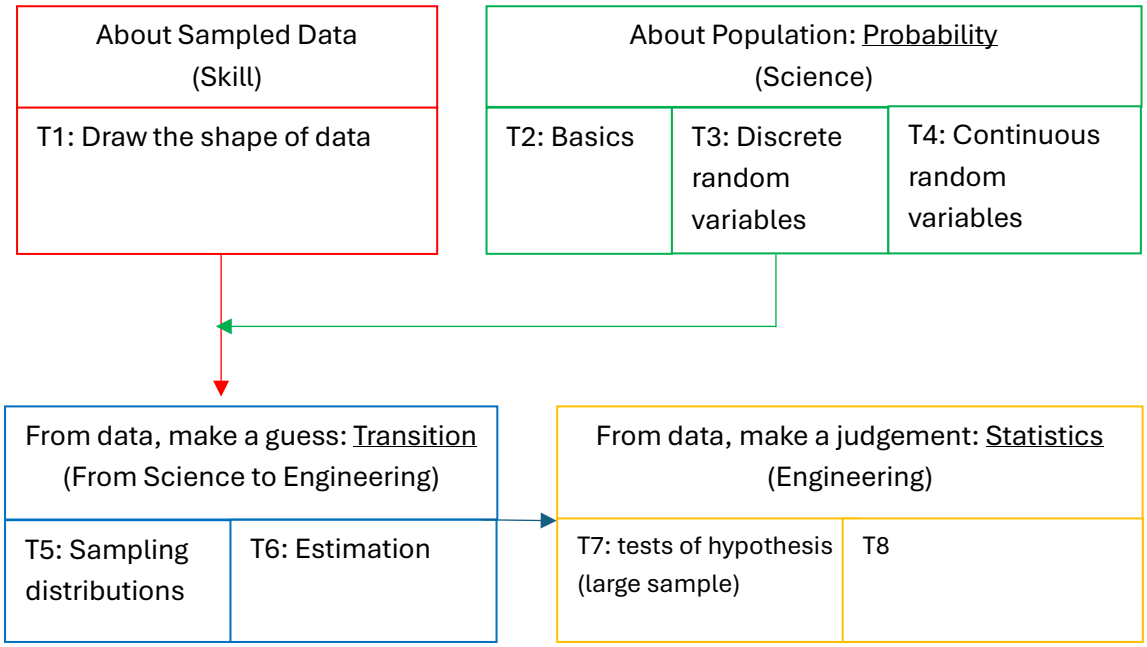
We use the same five-step procedure:

- Step 1: Read the questions carefully.
- Step 2: What is the random variable? Sample proportion. (see the red box. However, it is important to know that in the red box, it is referring to the population mean, which is just a number.).
- Step 3: Under what conditions? The sample size is 500. $\hat{p} = \frac{68}{500} = 0.136$, $n\hat{p} = 68 > 5$, $n\hat{q} = 432 > 5$ (see the green box)
- Step 4: What is the distribution of $\hat{p}$? Normal.
- Step 5: We compute the estimation for p is
  $p = \hat{p} \pm \text{SE}$ [by definition]
  $= \hat{p} \pm 1.96\sqrt{\hat{p}\hat{q}/n}$ [(1) 1.96 is due to 95% confidence. Use 1.645 for 90%, 1.96 for 95%, 2.58 for 99%; (2) we use $\hat{p}\hat{q}$, not pq, because we do not know p and q.
  $= 0.136 \pm 1.96\sqrt{0.136 * 0.864/500}$
  $= 0.136 \pm 0.030$

Tutorial 7: Making a Decision: Tests of Hypothesis from Large Sample(s)

In Tutorial 6, We only do some estimation through computation. We now move on to make a decision, which is important in our daily life. For example, in Semi-conductor industry, we make some *measurements*, and get a dataset, which is considered as a *sample*. From Tutorial 6, we can have some guess of the population. However, our ultimate goal is very simple: can the products (as population) pass the quality check? Yes or no?

There are two requirements for making the decision: (1) it should be scientific, so that the decision is reasonable; (2) it should be practiced as a routine, so that an engineer (not a mathematician) can use it. The first point is guaranteed by T5 and T6; the second point is addressed in this and next tutorials, T7 and T8.



7.0: Null hypothesis testing

In this chapter, only one thing needs your attention: why do we do null testing (i.e., there is a null hypothesis and an alternative hypothesis)? Why not just directly making a decision based on the computation results from T6?

My answer is that, we usually have a default state, and then this state is challenged/suspected. For example, when we cut the metal to pieces with a length of 10cm, and we adjust the machine to this nominal value – this is the default. You can challenge/suspect the length to be longer (one-sided), or shorter (one-sided), or simply different (two-sided). In addition, following this, the null hypothesis can serve as the role of population property, and then it is easier for us to form a 5-step routine:

Step 1: Read the question carefully. Recognize the concerned RV; form the null hypothesis (H0) and the alternative hypothesis (Ha);

Step 2: Set the significance level [this step is difficult to set in real life, but easy to set in this course: if given in the question, use it; if not, use 5%];

Step 3: Find the proper statistic for your problem [Build a summary table of the statistics by yourself for different scenarios];

Step 4: Do the computation [if specified, follow the question; if not specified, I recommend to compute and use p-value];

Step 5: based on Step 4 and Step 2, make a decision.

7.1: An example on proportion

Problem 1

A four-sided (tetrahedral) die is tossed 1000 times, and 290 fours are observed. Is there evidence to conclude that the die is biased, that is, say, that more fours than expected are observed? (Use $\alpha = 0.05$)

We use the five-step routine:

- Step 1: (1) Read the questions carefully. (2) The default state is that the fours appear 25% in the tosses. (3) However, our experiment shows that the proportion of fours is 290/1000=29%. (4) Well, we suspect that something is wrong. Fours appear more! We want to find out whether it is true. So the situation is one-sided; (5) We form Ho and Ha. Ho: p=25%; Ha: p>25%. [Note: you only need to write point (5) on your paper.]
- Step 2: The significance level is $\alpha = 0.05$; [If not specified in the question, use 0.05].

- Step 3: Find the proper statistic to use. (1) RV: sample proportion; (2) Condition: large sample, $\hat{p} = \frac{290}{1000} = 0.29, n\hat{p} = 290 > 5, n\hat{q} = 710 > 5$; (3) Distribution: normal, with the mean of $p = 0.25$ and standard deviation $\sqrt{pq/n}$; [Important notes: (1), the mean in this normal distribution is 0.25, instead of 0.29, because of the null hypothesis; (2), in the standard deviation, we use $\sqrt{pq/n}$, instead of $\sqrt{\hat{p}\hat{q}/n}$, because we know p and q from the null hypothesis.]
- Step 4: Computation: we use p-value:
  p-value
  $=P(\hat{p} > 0.29)$ [The probability of the right tail (one-sided)]
  $= P\left(Z > (0.29 - 0.25)/\sqrt{0.25 \times 0.75/1000}\right)$ [Convert to standard normal]
  $= P(Z > 2.92)$ [Simple computation]
  =0.0018 [Search the standard normal distribution table]
- Step 5: p-value $< \alpha$, indicating that getting $\hat{p} \geq 0.29$ is very unlikely, but it happened!!! So our decision is to reject the null hypothesis and take the alternative hypothesis.

7.2: An example on difference of means

Problem 4

To compare customer satisfaction levels of two competing cable television companies, 174 customers of Company 1 and 355 customers of Company 2 were randomly selected and were asked to rate their cable companies on a five-point scale, with 1 being least satisfied and 5 most satisfied. The survey results are summarized in the following table

Company 1	Company 2
$n_1 = 174$	$n_2 = 355$
$\bar{x}_1 = 3.51$	$\bar{x}_2 = 3.24$
$s_1 = 0.51$	$s_2 = 0.52$

Test at the 1% level of significance whether the data provide sufficient evidence to conclude that Company 1 has a higher mean satisfaction rating than Company 2.

We use the five-step testing again (a little boring, isn't it?):

- Step 1: (1) Read the questions carefully. (2) There is a word "rate". Do not be cheated. It is not about proportion. It is about mean (mean of rate)! More specifically, it is about difference of means. The default state is that two companies have not difference regarding the mean. (3) However, the survey result shows that the means are different (3.51 vs 3.24). (4) Is this difference significant? Does it mean that Company 1 has a higher mean rate? We want to find out. The situation is one-sided; (5) We form Ho and Ha. Ho: $\mu_1 - \mu_2 = 0$; Ha: $\mu_1 - \mu_2 > 0$. [Note: you only need to write point (5) on your paper.]
- Step 2: The significance level is $\alpha = 0.01$; [Given by the question].

- Step 3: Find the proper statistic to use. (1) RV: difference of sample means, $\bar{x}_1 - \bar{x}_2$; (2) Condition: $n_1 \geq 30, n_2 \geq 30$. Large samples; (3) Distribution: normal, with the mean of $\mu_1 - \mu_2 = 0$, and the standard deviation of $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (if σ1 and σ2 are unknown, use s1 and s2 instead, i.e., $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$; [This part is your homework, see the Section 7.4.]
- Step 4: Computation: we use p-value:
  p-value
  $= P(\bar{x}_1 - \bar{x}_2 > 5.51 - 3.24 = 0.27)$ [one-sided]
  $= P\left(Z > (0.27 - 0)/\sqrt{\frac{0.51^2}{174} + \frac{0.52^2}{355}}\right)$ [Convert to standard normal]
  $= P(Z > 5.68)$ [Simple computation]
  ≈0 [Search the standard normal distribution table. Actually no need, it's small.]
- Step 5: p-value < α, so our decision is to reject the null hypothesis and take the alternative hypothesis.

7.3: Importance of Intuition

When you read the questions in Step 1, you may quickly form an intuitive decision. If your final result is against your intuitive decision, then careful check your work.

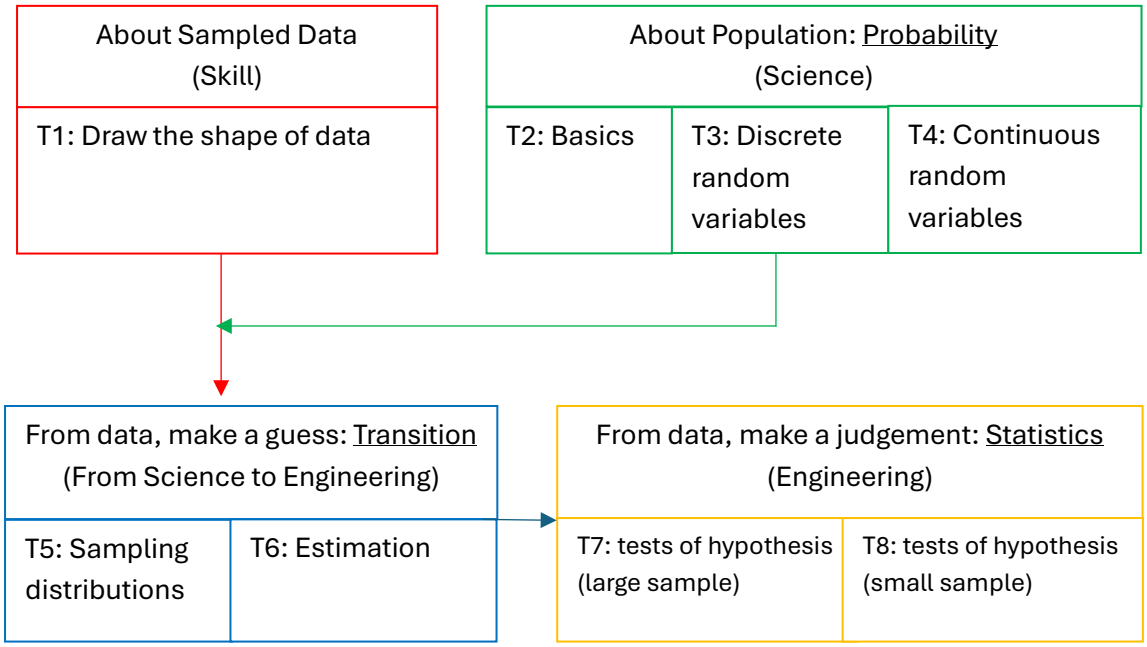
Use Q4 as an example, you see s1 and s2. Remember, for sample means, you need to divide them by sqrt(sample size), i.e., $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. You don't need to compute, and just estimate, the standard deviation will be around 0.02. But the difference of means is $\bar{x}_1 - \bar{x}_2 = 0.27$, which is more than 10 times of standard deviation. The difference is too large and unlikely possible! So we can guess that we will reject Ho. Even without computation, I already have a rough idea for my decision! If you make a careless mistake in your formal computation, your intuitive decision can alert you to check it.

7.4: Summary of statistics for large samples (Homework)

RV	Condition	Distribution/Statistic
Sample mean	The sample is large (n>30)	Normal (if σ known, use it) $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Difference of sample means	Both samples are large (n1>30, n2>30)	Normal (if σ1 and σ1 known, use them) $z \approx \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\text{SE}} = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
Proportion	The sample is large $(np > 5, nq > 5)$	Normal $z = \frac{\hat{p} - p}{\sqrt{pq/n}}$
Difference of proportions	Both samples are large $(n_1p_1 > 5, n_1q_1 > 5, n_2p_2 > 5, n_2q_2 > 5.)$	Normal $z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}$

Tutorial 8: Making a Decision: Tests of Hypothesis from Small Sample(s)

This is our last tutorial and is a simple tutorial. We keep using the five-step routine for hypothesis testing. The only difference is that, now the sample size is small, so we need to use different test statistics. I do not have much to say in this tutorial, which is consistent to what I said in class: the hypothesis testing has been made simple for everyone to use.



8.0: About t-distribution

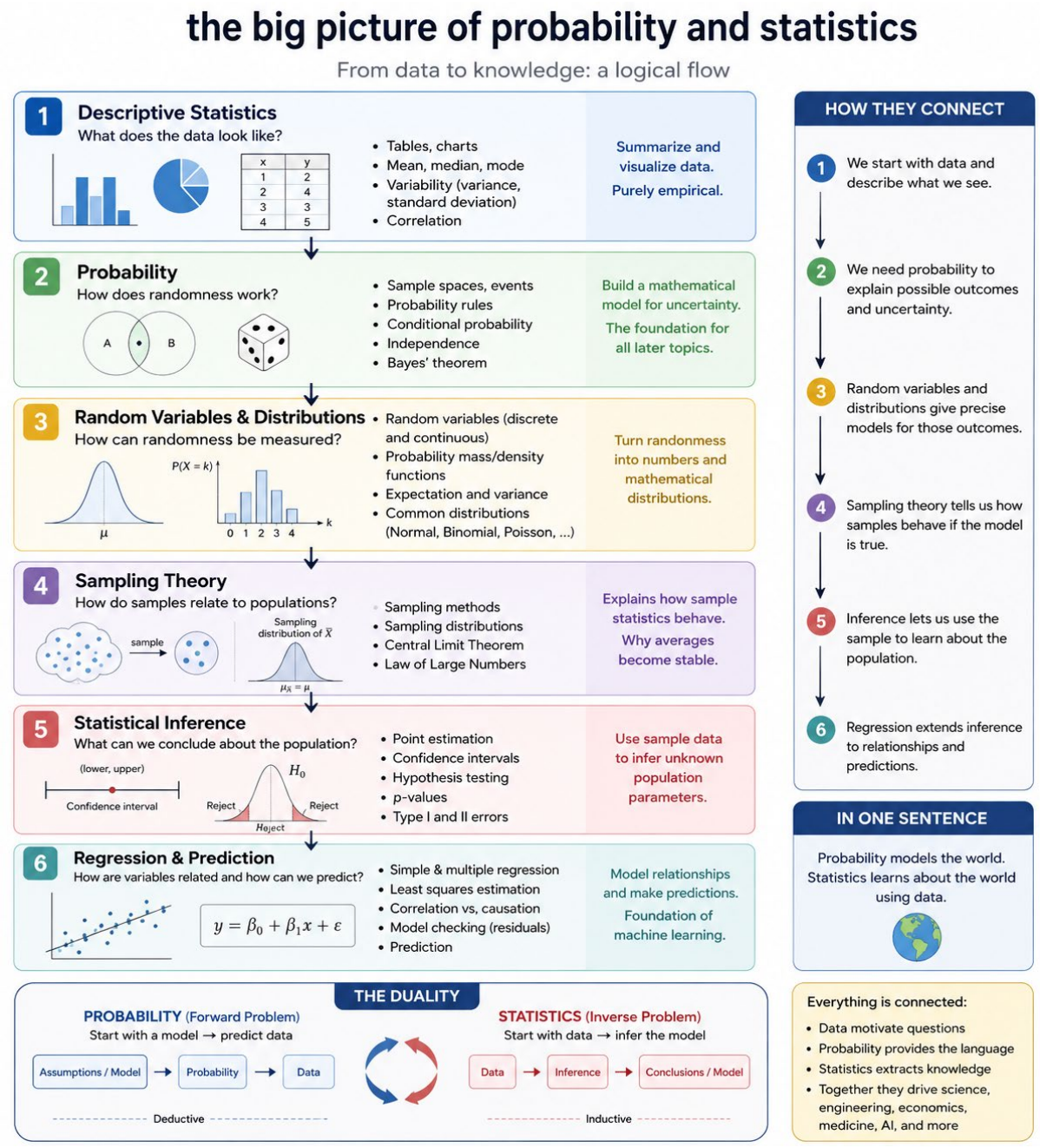
For sample mean, when the sample size is small, and the population standard deviation is unknown (so we use sample standard deviation s instead), the sample mean obeys t-distribution.

8.1: Summary of statistics for small samples (Homework)

RV	Condition	Distribution/Statistic
Sample mean	- Small sample (n<30), - σ unknown - if σ is known, you use normal distribution, but in real life hypothesis testing, σ is usually unknown.	t-distribution with a degree of freedom of n-1 $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Difference of sample means	- Small sample (at least one is <30), - σ1=σ2=σ, but they are unknown	t-distribution with a degree of freedom of n1+n2-2 $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

8.2: The ending (1)

On Saturday morning on my way to a food court, I suddenly realized that I forgot to consult AI. I took out my handphone and asked ChatGPT with my simple question, without extra refined prompts, and this what it gave me:



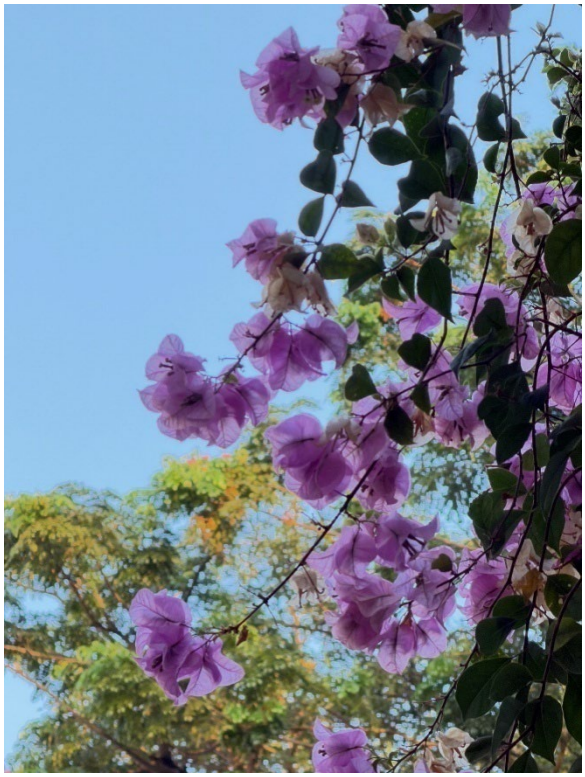
You can see that it is about the same as my structure, but it is beautiful and pleasant. Here are the hints at this moment:

- AI is indeed super-powerful, and you can use this power during your learning;

- Do not rely on what AI gives you, which does not enhance your understanding and increase your knowledge.

8.3: The ending (2)

If you are here, you are great! With a high probability, you have mastered the course!



The random distribution and the deterministic beauty (Phone taken by me at NTU)

Thank you!