ECEN321 Engineering Statistics

(Lec 0 Introduction)

Yau Hee KHO

Victoria University of Wellington

yauhee.kho@vuw.ac.nz

with ZHENG Wei Heng 郑蔚恒

Xiamen University of Technology

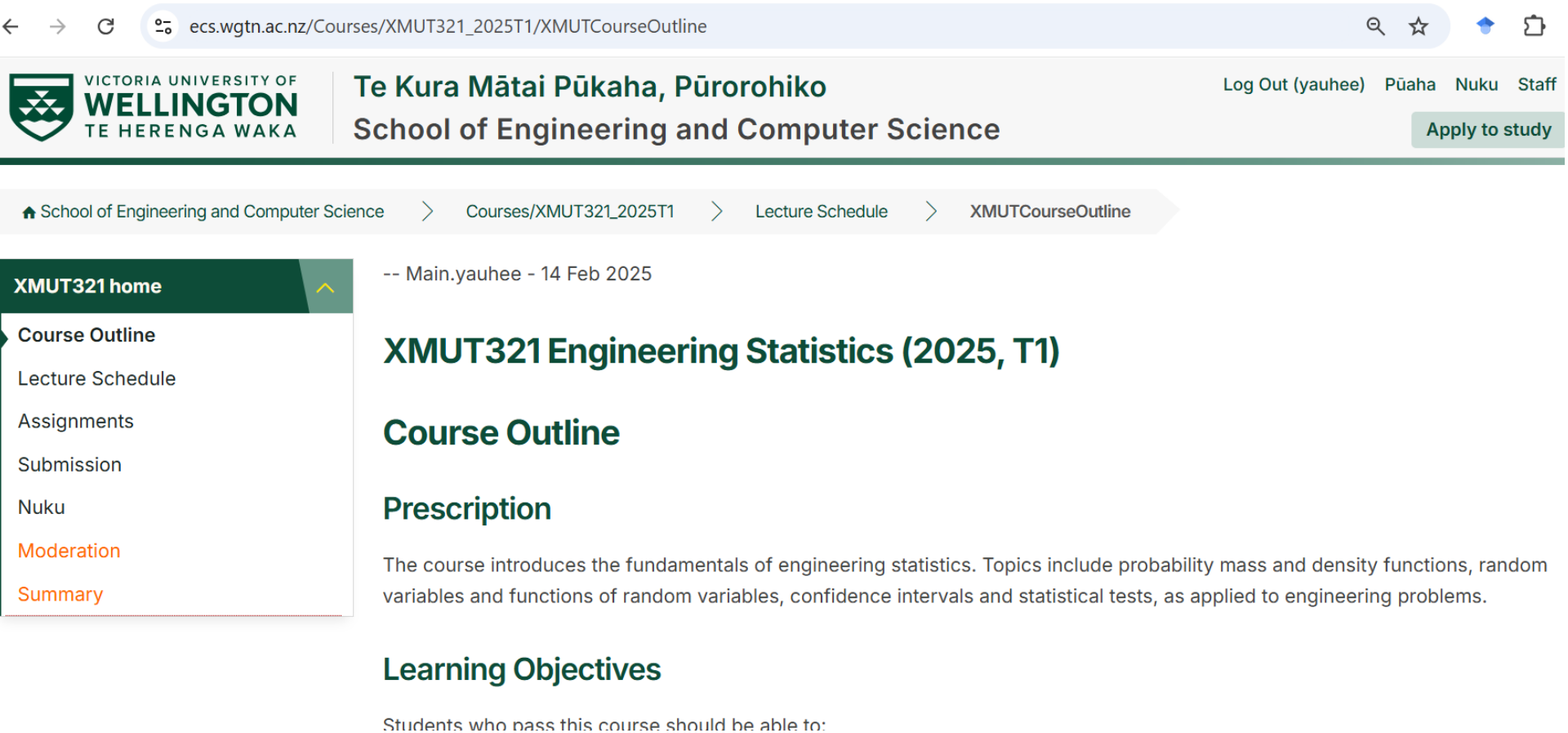
VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA

School of Engineering and Computer Science

Te Kura Mātai Pūkaha, Pūrorohiko

Course website

- https://ecs.wgtn.ac.nz/Courses/XMUT321_2025T1/WebHome



Engineering Statistics Course Overview

- This is a 1 semester undergraduate course, covering the basics of Engineering Statistics:

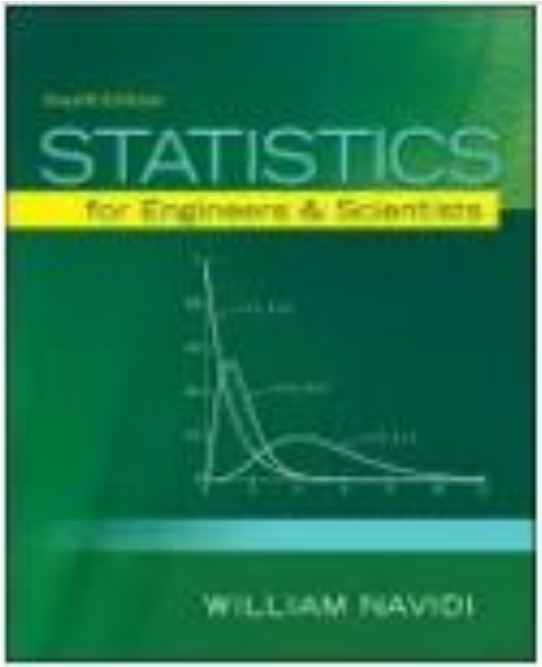
  - ☐ Sampling and Descriptive Statistics
  - ☐ Probability
  - ☐ Propagation of Error
  - ☐ Commonly Used Distributions
  - ☐ Confidence Intervals
  - ☐ Hypothesis Testing
  - ☐ (If time permits, Correlation and Simple Linear Regression)

  All LABS will be conducted by ZHENG Wei Heng

Textbook

- Any book on Statistics is equally good.
- But we will refer to this book:

"Statistics for Engineers and Scientists" by William Navidi, published by Mc Graw Hill, 4th edition.



Assessment Criteria (subject to confirmation)

- One end-of-semester final exam 40%.
- 1 mid-semester Test at 10%.
- 3 Labs to be done. Only 2 labs will be marked (10% each, giving a total of 20%).
- 4 Homework Assignments. (5% each, with a total 20%).
- Attendance in Lectures and Labs (total 10%)

On top of the above, you need to manage on your own:

- Tutorials
- Reading

Grade Computation

Your grade for the course will be based on a combined mark for the assignments, the tests, and the exam:

Item	Weight
4 Homeworks	5% each = 20% in total
2 Labs	10% each = 20% in total
1 Test	10%
Examination	40%
Attendance	10%

Teaching Schedule: 8 weeks!

Correct as of 14th February 2025. Subject to change. Please check regularly.

Teaching schedule

Week	Lecture topic (6 periods/week)	Labs (4 periods/week)	Homework	Task Due
1	Introductory Ideas	Lab 1_1 Intro to Matlab	HW 1	
2	Probability	Lab 1_2 Using Matlab for Statistics	HW 2	
3	Propagation of Errors	Lab 1_2 (continue)	HW 3	HW 2
4	Probability Distributions	Lab 2 Probability Distributions		HW 3
5	Probability Distributions	Lab 2 (continue)		Lab2 report? + **Test**
6	Confidence Intervals	Lab 3 Noise and Averaging	HW 4	
7	Confidence Intervals	Lab 3 (continue)	HW 5	HW 4, Lab3 report?
8	Hypothesis Testing	Extra time for Lab	HW 6	HW 5
9	**Final Exams**			TBA

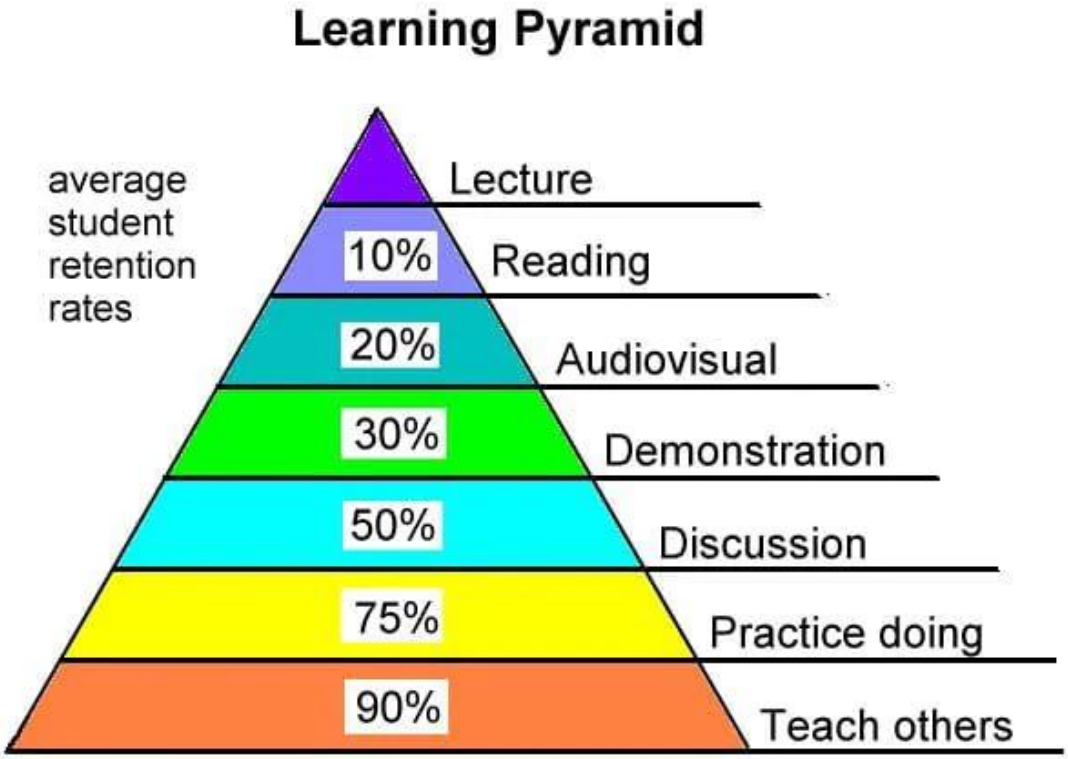
Week 1 : Feb 17, 2025

How to study Engineering Statistics?

Tell me and I forget,

Show me and I remember,

Involve me and I learn.



Source: National Training Laboratories, Bethel, Maine

Expected hours of study

- 8 periods of lectures per week,
- 4 periods of Labs per week,
- You also need self-directed study; how many hours per week will be required?
- At VUW this is a 15-point course; 1 point ~ 10 hours, so this means a total of 150 total hours for whole course.
- We have 8 weeks, so ~ **20 hours per week**!
- These include class/lab times, doing homework, tutorials, assignments, write lab reports, reading etc

General Information

- Difference between the style of teaching and learning at VUW and XMUT
- Plagiarism – do not copy from your classmates
- Attendance – make sure you have enough attendance to sit final exam.
- Mark scaling between VUW and XMUT
- Resit exam – course marks will be calculated.
- Recommendation letters – show me how good you are for me to write good about you

Different systems and teaching styles

- “Easy in Difficult out” vs “Difficult in Easy out”,
- Students treated as mature adults, not higher secondary students,
- Training you to be an independent professional with critical thinking and problem solving skills,
- More hands-off => you explore and teach yourself,
- Open ended questions, reports etc,
- Tests and Exams are usually not the repeat of homeworks.

Plagiarism video @ https://www.bilibili.com/video/BV1nb4y117QN/

- Meaning to copy others' works and submit as your own. This is NOT OK.
- You can discuss and seek help from classmates, BUT
- DO NOT just COPY => you must work on your own.
- This includes NOT just copy and translate words from webpages/textbooks => you must re-write in your own words.
- This includes diagrams and figures!
- In group lab work, indicate all members' names and IDs.

Attendance

- By XMUT's rules, you need to have a minimum level of attendance rate in classes.
- Otherwise, you will not be allowed to take the final exam => you will fail the course
- There is also 10% of marks given to attendance.

Mark scaling

- Different pass marks: 50 marks in VUW and 60 marks in XMUT.
- Does this mean it is easier to pass in VUW?
- NO, due to an inverse relationship => a lower pass mark means assessments are usually more difficult.
- As long as you get 50 marks in VUW's courses, you would have passed!
- When reporting to XMUT, we will apply a piecewise linear scaling so that 50 in VUW becomes 60 in XMUT.

Resit exams

- There is NO resit exams in VUW!
- But we have adapted to XMUT's system, by offering resit exams BUT with a major difference: your normal assessments and tests marks will still be calculated.
- This is different from XMUT where only the resit exam marks will be considered.
- This means, if your normal assessment marks are too low, you will still not pass the resit exam => in some cases even if you get 100% in the resit exam!
- Please take note!

Recommendation letters

- Some of you may choose to go overseas for further studies,
- I have provided many recommendation letters to your seniors,
- And I am happy to write one for you BUT
- It is your responsibility to provide me with good stuff to write about you.
- To me, academic results is one aspect, and your study/personal attitude is another that I will consider.
- My reputation is at stake when recommending you, so
- Do not get offended if I refuse to write one for you ☺

Feedback from your senior...

Mr Yau Hee, I think there is no problem with your teaching method.I know that most students get a bad score in this exam, but it's not your fault. Frankly, most questions in this exam are easy, and i believe that we can complete this exam with high school knowledge only.

I have some explanations with this phenomenon.

First, some students don't complete homework by themselves. They ask help from others. Obviously,for a poor man, fishing gear is better than fish, so good students will teach them how to fo this question.However, most them care about the answer only.

Therefore, they cannot do Same kind of question in exam.

5:35 AM

Second, some students focus on postgraduate qualifying examination, and they attended Off-campus training courses. Therefore, they don't have enough time to do your course well.

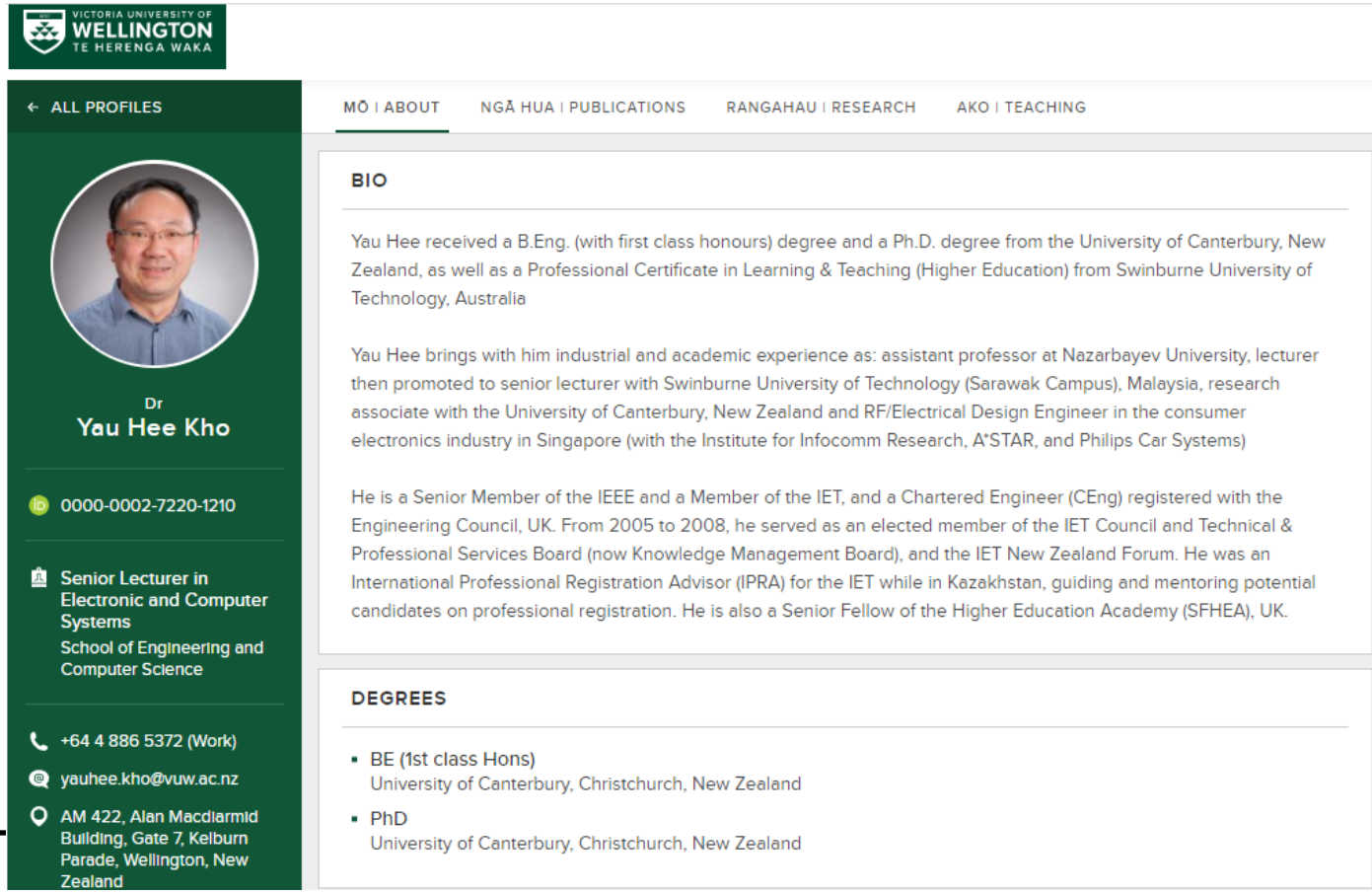
6:57 AM

Many students solve problems in the wrong way, they don't care about how to get the answers, they only care about the process of solving the problems. In other words, they don't understand knowledge,they only remember knowledge. so they can finish the same type of problems well, but for different types of problems, they don't know how to do.

Yau Hee Kho：Oh, do you mean they hoping the same questions from homew...

Introduction

- Myself https://people.wgtn.ac.nz/YauHee.Kho
- XMUT teacher
- Our class…



ECEN321 What is Statistics?

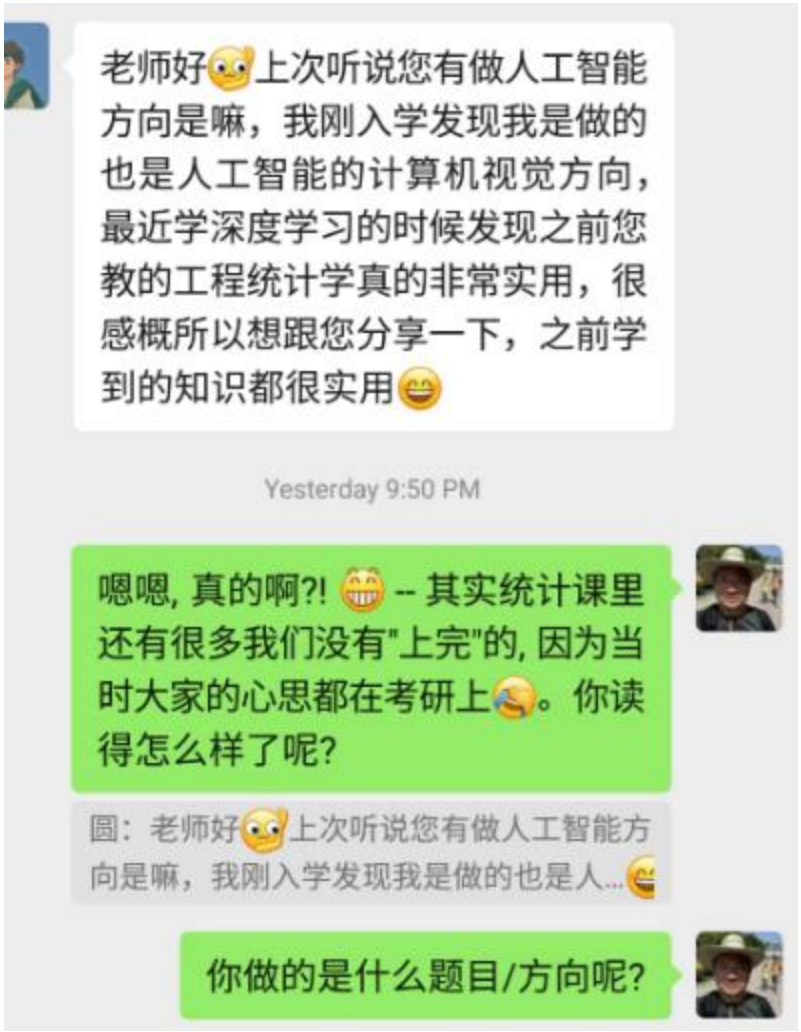
- Statistics is a branch of mathematics dealing with random processes.
- Through observing a part of the random process, we then decide on a conclusion of the overall, actual process.
- It's built on the theory of probability, starting with data sampling, and analysis, to arrive at the distributions and conclusions.
- Dealing with data and randomness is common in engineering, so statistics is very important.
- Require basic calculus: integration and differentiation.

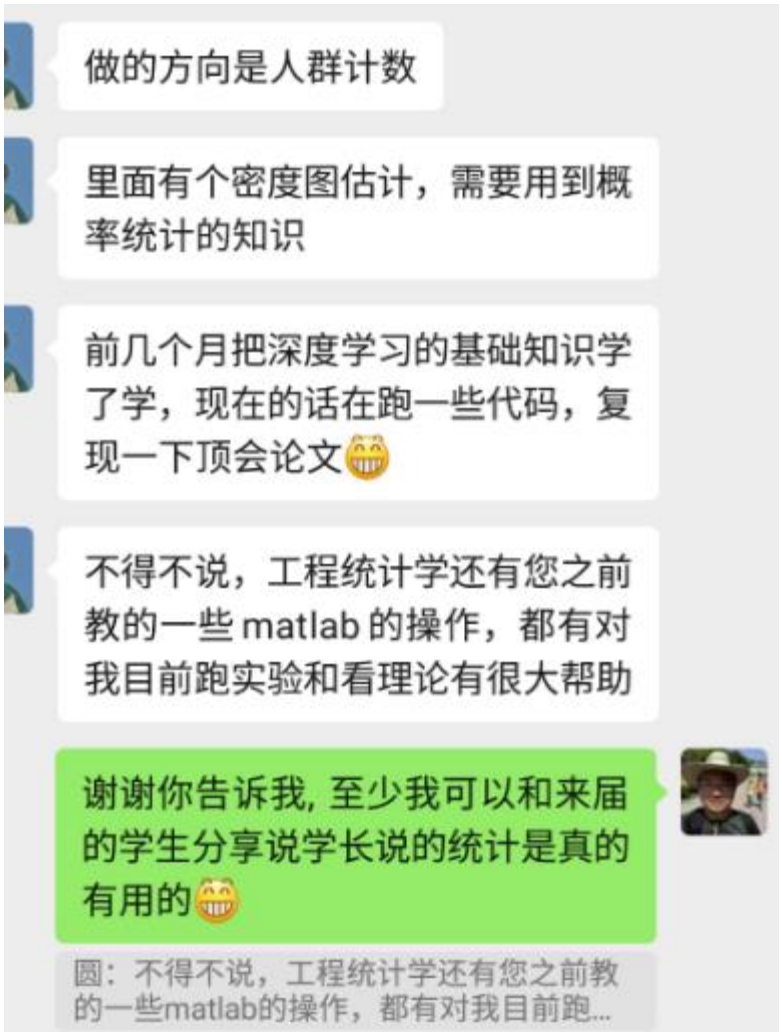
Usage of Statistics

- Communication theory (5G / 6G: Probability of correctly receiving a transmission in the presence of noise)
- Computer network data traffic (queueing theory, arrival rate, dropped data packets etc)
- Image processing (histogram equlisation),
- Signal processing (estimation and detection),
- Big data (eg personalised study materials),
- Artificial Intelligence (facial recognition, false positive, false negative) etc.

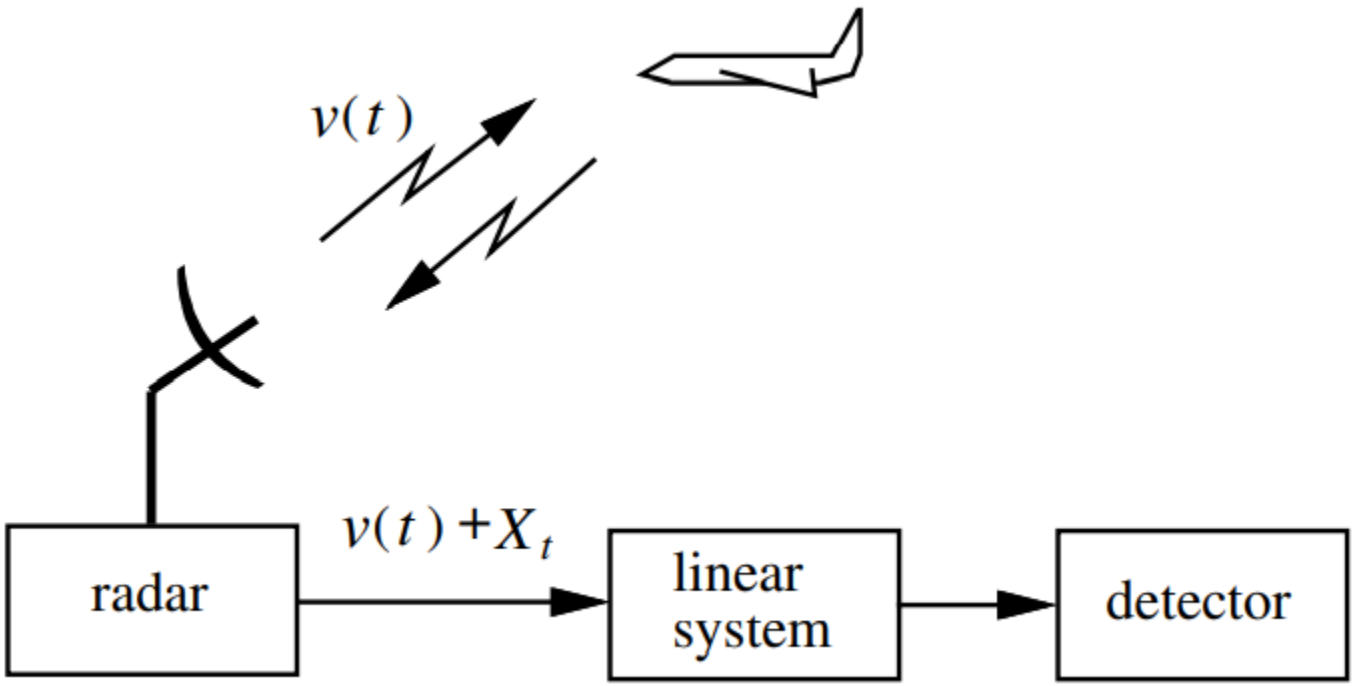
and many more. Let's look at some situations/applications:

工程统计，学了有用吗？


老师好上次听说您有做人工智能方向是嘛，我刚入学发现我是做的也是人工智能的计算机视觉方向，最近学深度学习的时候发现之前您教的工程统计学真的非常实用，很感概所以想跟您分享一下，之前学到的知识都很实用
Yesterday 9:50 PM
嗯嗯, 真的啊?! -- 其实统计课里还有很多我们没有"上完"的, 因为当时大家的心思都在考研上。你读得怎么样了呢?
圆：老师好上次听说您有做人工智能方向是嘛，我刚入学发现我是做的也是人...
你做的是什么题目/方向呢?

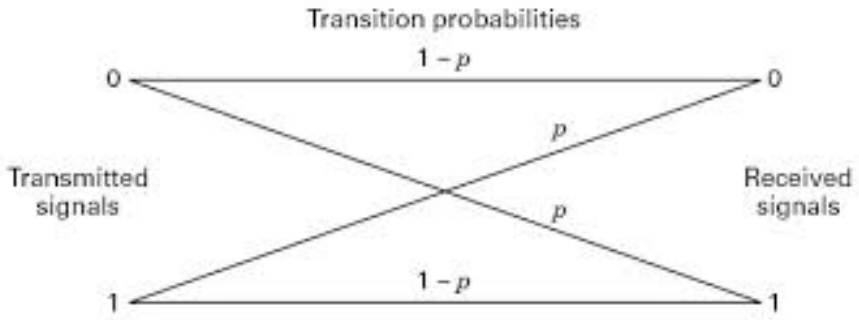

做的方向是人群计数
里面有个密度图估计，需要用到概率统计的知识
前几个月把深度学习的基础知识学了学，现在的话在跑一些代码，复现一下顶会论文
不得不说，工程统计学还有您之前教的一些 matlab 的操作，都有对我目前跑实验和看理论有很大帮助
谢谢你告诉我, 至少我可以和来届的学生分享说学长说的统计是真的有用的
圆：不得不说，工程统计学还有您之前教的一些matlab的操作，都有对我目前跑...

Signal Processing: radar detection system

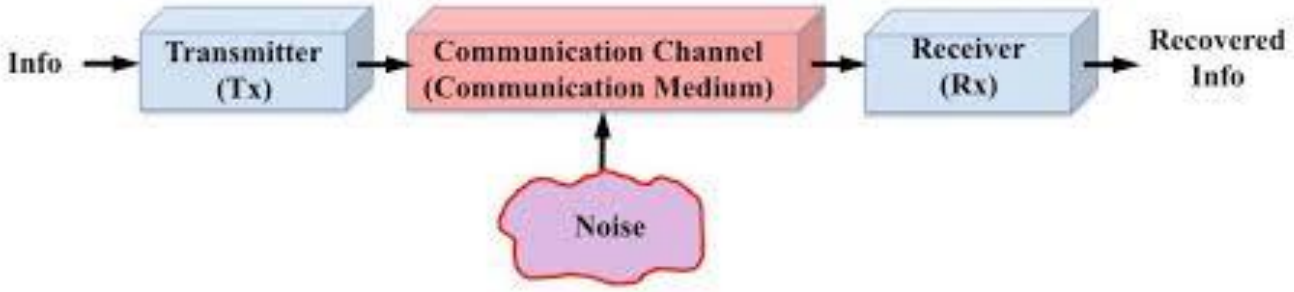


Communication Engineering: receiving bits

- We transmit 100101… , what is the probability of receiving all bits correctly?
- Due to random noise and interference, error occurs with a probability of *p*;
- Need to know the statistical distribution of noise…
- Assume equal probability => Binary symmetric channel
- Question: if we received 100101…, how confidence are we that this is what has been actually transmitted?
- Need to use Bayes' theorem, hypothesis testing etc…



Communication channel: estimation



- Given a Maximum Likelihood (ML) estimate of channel impulse response, h,
- the likelihood of the received info, p(r/h), is maximised.

Can use Bayes' rule to express Maximum a posteriori (MAP) estimation

p(h/r) = p(r/h) times p(h) divided by p(r)

MAP estimation

Likelihood
How Probable is the evidence(data) given that the hypothesis(parameters) is true

Prior
Our belief in hypothesis(parameters) before observing the evidence(data)

$$P(H \mid e) = \frac{P(e \mid H)\,P(H)}{P(e)}$$

Marginal
Distribution of evidence(data), mostly we are not interested in this as this is a constant term

General scenario: Which bucket will you choose?

If randomly picking a red ball earns you $10,

- Bucket A contains 9 blue balls and 1 red ball.
- Bucket B contains 91 blue balls and 9 red balls.

Which bucket will you choose to pick from?

Numerical illusion?

- If you need a surgery, the doctor can either tell you that:

"1 out of 1000 persons will not survive the surgery,"

or

"999 out of 1000 persons will survive the surgery".

Which statement do you prefer your doctor to tell you?

We have the same birthday, yippie!

How often do you meet someone who has the same birthday (day and month) as you?

How big the class needs to be before the probability of getting 2 persons sharing the same birthday will be 0.5?

Our class has 76 students, what is the probability to find 2 students with the same birthday?

How about 3 students? 4 students? etc

Accuracy of Medical Testing

Covid-19!

False positive vs False negative – which one is more critical?

If you are tested once, and are called for a second confirmation test, do you have to worry?

For this, we need to know the infection rate of Covid-19. Let's assume it is 0.4%, i.e. in 10,000 people, it infects 40 people.

Let's assume the accuracy of the Covid-19 test is 90%. Then we can draw the following "tree diagram":

Tree diagram:

中国的执业医师考试笔试

PI

3．用两种测量肺活量的仪器，测量18名女生的肺活量，为对比两仪器的测量差异应该用的统计学方法是（ ）。

A．q检验

B．配对，Z检验

C．配对，t检验

D．分组，Z检验

E．分组，t检验

这是一道统计题，作为畅销书《临床回顾性研究实用指南》的主编，丁香公开课排名第二的《零基础发表临床回顾性研究SCI》主讲，这种题我想都不想就选出了答案——

C。

这是课上讲的例题啊，用两组仪器测试同一组对象，测出的数据必然是配对资料，而肺活量是计量资料，样本量又不大，用t检验就可以，所以但是配对t检验。

结果我又对了一下答案。

【答案】E

【解析】
t检验可用于成组设计的两样本均数差异比较的独立样本。本题用两种测量肺活量的仪器，测量18名女生的肺活量，为对比两仪器的测量差异，应在分组后进行t检验。

Time to ponder about plagiarism:

You are given these information:

- The probability that a student plagiarises is 1%.
- If the student has cheated, the probability of a positive software detection is 90%.
- If the student does not cheat, the probability that software detection still indicate cheating is 9%.

Which of the following statements best describes the chances that a student with a positive software detection has actually cheated:

A. The probability that the student has cheated is 81%.

B. Out of 10 students with a positive software detection, about 9 have cheated.

C. Out of 10 students with a positive software detection, about 1 has cheated.

D. The probability that the student has cheated is 1%.

Which is your answer?

Another tree diagram: