MATH 10

INTRODUCTORY STATISTICS

Tommy Khoo

Your friendly neighbourhood graduate student.

Homework

- Please hand in your homework.
- Answers to these 3 homework will be posted online.
- No homework this week! ヽ(ヅ)ノ

Midterm Exam ٩(^ᗜ^)۶

- In class, Thursday, 26 April.
- Seat one space apart please.

- 1 hour, 45 minutes. → come early for favorite seat
- 5 questions of varying lengths. → e.g. ~20 mins each

- 40 points total. 2 short, 3 long questions.
- Your score will be converted to the 30% weight.

Exam Syllabus

- Chapters 1 to 10.

- Up to confidence intervals for mean, proportions, and t-distribution.

- Not in midterm:

chapter 8 advanced graphs

anything about difference between means in both chapter 9 and 10.

Exam Policy

- Formula sheet given!
- Our policy: we will not explain things on the formula during the exam.
- You are responsible for understanding what those things mean.
- It is not meant to be a comprehensive explanation/textbook.

Practice Materials

- Practice materials: https://math.dartmouth.edu/~m10s18/exam.php
- Please prioritize the "2018 Midterm Practice".

- If you can do the 8 sample exam questions in the beginning in under 20 minutes each, you are in very good shape!

- Older 2017 materials → contains many things not in our course.

Quick Exam Advice

- The points on each question/part is a good indicator of how simple/straightforward or hard/complicated the question is.

- You can come up and ask me to clarify questions that are unclear.
- You can tell me your interpretation of what the question is asking, and I will tell you if it is correct.

- You should be getting nice numbers or fractions.
- If you think you got something ugly, feel free to ask me if you have simplified enough.
- Do the calculations without rounding. Only apply the X significant figures to the final answer.

Today's Review Topics

This might be a short lecture.

I will highlight things you should know for the exam…and more (to keep you on your toes).

These are what *I think* will get you 100% on the exam but everyone is different.

The topics are as follows…

Today's Review Topics

1. Summary statistics
2. Graphing distributions
3. Bivariate data and Pearson's Correlation Coefficient
4. Probability, Binomial Distribution, Bayes Theorem BREAK
5. Normal distribution, using the z-table, normal approximation to binomial distribution.
6. Sampling distributions and confidence intervals: the 3 cases.
7. True/False and "subjective" questions.

Summary Statistics, Graphing Distributions

- Know how to calculate mean, median, mode for very simple cases.
- Distribution shapes: uniform, symmetric, positive skewed, negative skewed.
- Symmetric, uniform: Mean = median = mode (not unif) usually.
- Skewed: mean can differ from median. Eyeball the mode.

Summary Statistics, Graphing Distributions

- Variability: population variance, population standard deviation, estimate of the variance, estimate of the standard deviation.
- Note: understand population vs sample statistics.
- Symmetric, uniform: mean and variance good summary statistics (e.g. homework 1).
- Skewed: not good. Solution: report as many as possible. Report median and interquartile range if available.

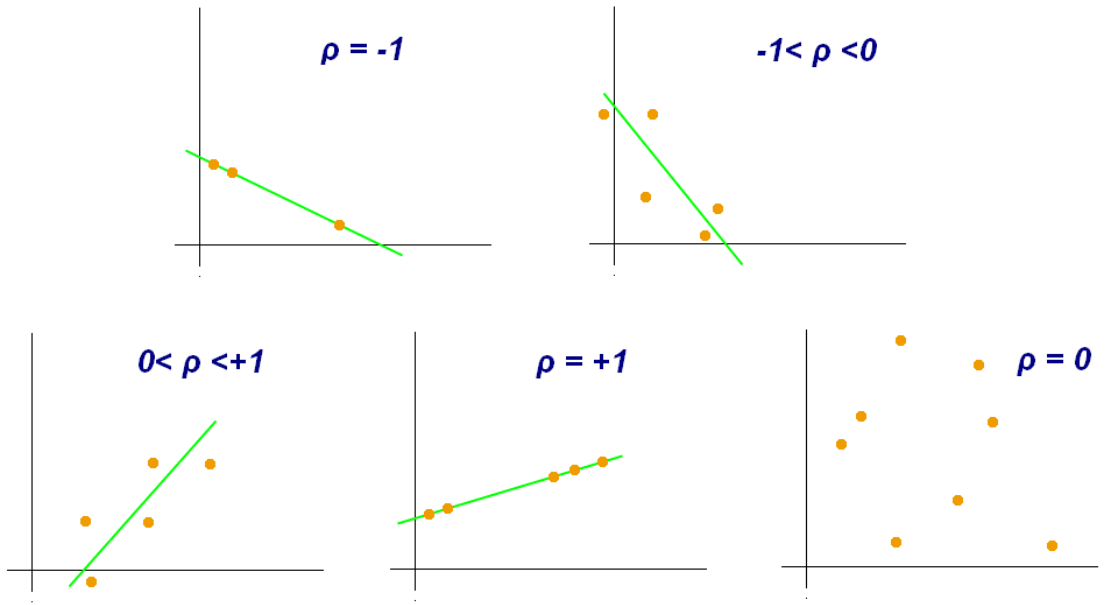
Summary Statistics, Graphing Distributions

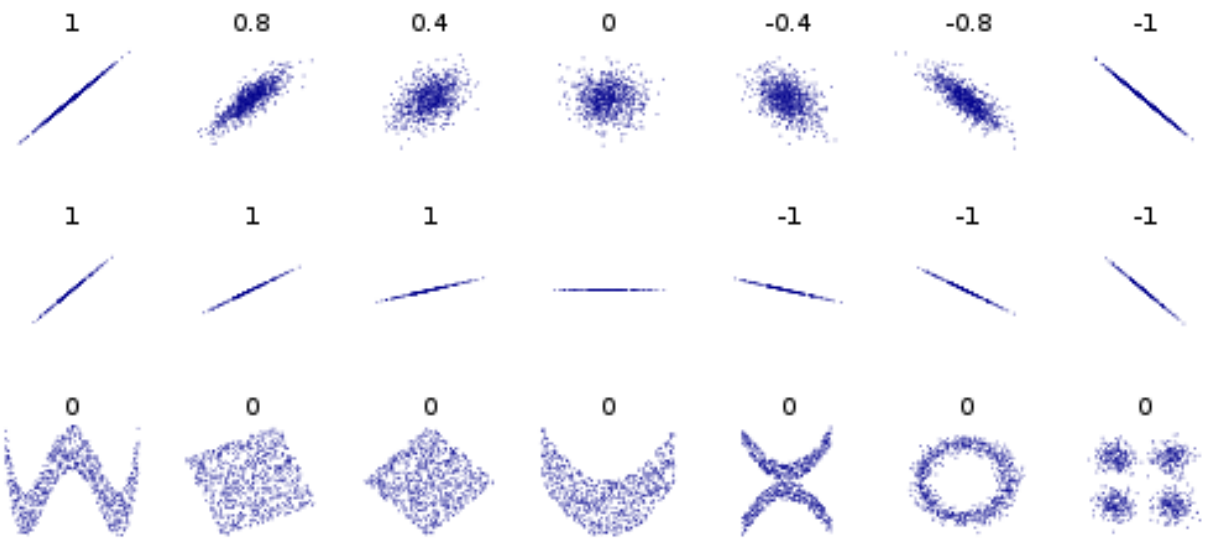
- Percentiles: given in formula sheet.

- Interquartile range: $75^{th} - 25^{th}$ percentile (given as well).
- Median: 50^{th} percentile.

Bivariate data and Pearson's Correlation Coefficient r

- Guessing Pearson's r from scatter plot : if you understand these figures from Wikipedia, you're good to go.
- Note: just pretend "rho" is r in the figure below.
- Data must be paired (X,Y). i.e. Bivariate data.





Bivariate data and Pearson's Correlation Coefficient r

- Properties of r :

1. Symmetric.
2. Invariant under change of units. i.e. aX + b where a > 0.
3. Cannot be calculated if variance of X or Y is zero.

- When to use r :

1. r is a descriptive statistic. Use to describe what you believe are real linear relations.
2. Recipe: believe a real linear relationship exists → can use r to describe it. (can be used to search for real linear relationships → not in exam)
3. It is not to be used for inference. Correlation r alone does not imply anything!!

Probability, Binomial Distribution, Bayes Theorem

$P(A)$

$P(\text{not } A) = 1 - P(A)$

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

$P(A \text{ and } B) = P(A)P(B)$, independence

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Probability, Binomial Distribution, Bayes Theorem

Binomial distribution → given in formula sheet.

Make sure you can do questions such as the ones below!

3. Suppose the probability of being admitted to a college is 0.50 for every student who applies. If 3 students applied, what is the probability that 0 or 2 students got admitted? Give a simplified numerical answer and show your work.

4. If instead, the probability of being admitted is 0.30 and 10 students applied, write down a numerical expression for the probability that betwee 1 and 3 students got admitted. You do not have to simplify your answer.

Probability, Binomial Distribution, Bayes Theorem

Bayes Theorem → formula given too.

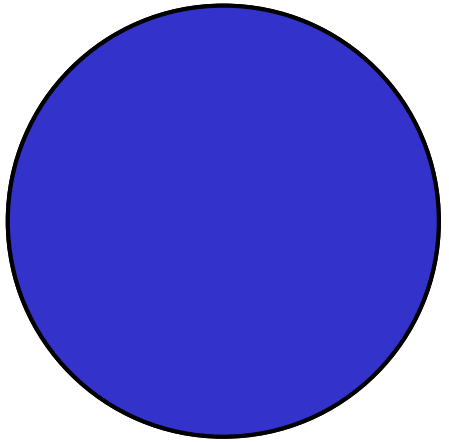
Sample Exam Question 3

Break time!! \o/

12 minutes

- Circle is a timer that becomes blue. O_o →

(please ignore if it glitches)



Normal Distribution

How to use the z-table.

The linear transformation that turns any normal random variable X into standard normal Z.

Sample exam question 5.

Normal approximation to the binomial distribution → can always use unless we tell you not to.

Sampling distributions and Confidence intervals

1. Central limit theorem, normal distribution → sample exam qns 6

2. 2. t-distribution → sample exam qns 8

3. 3. Proportion → sample exam qns 7

True/False and "Subjective" Questions

- Does data/statistic X **alone** proves that statement Y is true?
- Based on statistical evidence X, is statement Y **likely/probably** true?

1. Do not use information outside of data/statistic X.
2. Focus on "is statement Y true?". X may not say anything abt "is statement Y false?" → very common in hypothesis testing
3. This is about good use of statistics, not what things "really are".

*P.S. Stats **alone** cannot prove anything, so "proves" → look for an obvious problem/flaw. Use probabilistic statements. Likely/probably.*

True/False and "Subjective" Questions

E.g. I give you statistical evidence X from a treatment and control group. Double-blind drug trial. Patients randomly assigned to groups. Use of placebo. Everything performed to perfection.

Is X statistical evidence for drug efficacy? Or is X statistical evidence that the drug *probably* has an effect?

Ans: yes, everything has been done correctly. So the drug probably has an effect. Or this is evidence that it is *likely* to have an effect.

E.g. Statistical evidence X from a treatment/control group drug trial with an obvious flaw. Is X good statistical evidence that the drug work?

Ans: no. due to the obvious flaw.

True/False and "Subjective" Questions

If you think a question is too vague, open-ended or subjective on the exam…

Let us know and we will explain.

…or tell you to go back to your seat. :3

The end! Good luck with the midterms.