MATH 10

INTRODUCTORY STATISTICS

Tommy Khoo

Your friendly neighbourhood graduate student.

Midterm Exam ٩(^◡^)۶

- In class, next week, Thursday, 26 April.

- 1 hour, 45 minutes. → try to arrive early
- 5 questions of varying lengths. → allocate 20 mins to each

- Your score will be converted to a percentage. 30% weight.
- The next slide contains nothing but good news.

Midterm Exam ٩(^ᴗ^)۶

- Formula sheet given!

Midterm Exam ٩(^ᴗ^)۶

- Formula sheet given!

- Not in midterm: chapter 8 advanced graphs, chapter 11 hypothesis testing. Difference between means in both chapter 9 and 10.

Midterm Exam ٩(^◡^)۶

- Formula sheet given!

- Not in midterm: chapter 8 advanced graphs, chapter 11 hypothesis testing. Difference between means in both chapter 9 and 10.

- Midterm exam will cover chapters 1 to 10, up to confidence intervals for mean, proportions, and t-distribution.

Week 4

- **Chapter 10 – Estimation** **← today's lecture**

Point, interval estimates. Bias, variability. Confidence interval!! → for the mean, difference between means, proportions`. t-distribution!!

- **Chapter 11 – Logic of Hypothesis Testing**
- **Chapter 8 – Advanced Graphs** **← when will we get to do this lol**

Sampling Distributions and Confidence Intervals

- Will these be in the exam and cost a lot of points? You betcha!

- While sampling distributions may be a standalone question…
- You cannot construct confidence intervals without sampling distributions.

Sampling Distributions and Confidence Intervals

- Will these be in the exam and cost a lot of points? You betcha!
- While sampling distributions may be a standalone question…
- You cannot construct confidence intervals without sampling distributions.

- We can divide these two topics into 4 major parts:

1) Normal distribution
2) t-distribution
3) Proportion
4) Difference between means → not covering today, not in midterm

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- Confidence intervals are interval estimators.
- What are, for example, 95% confidence intervals?

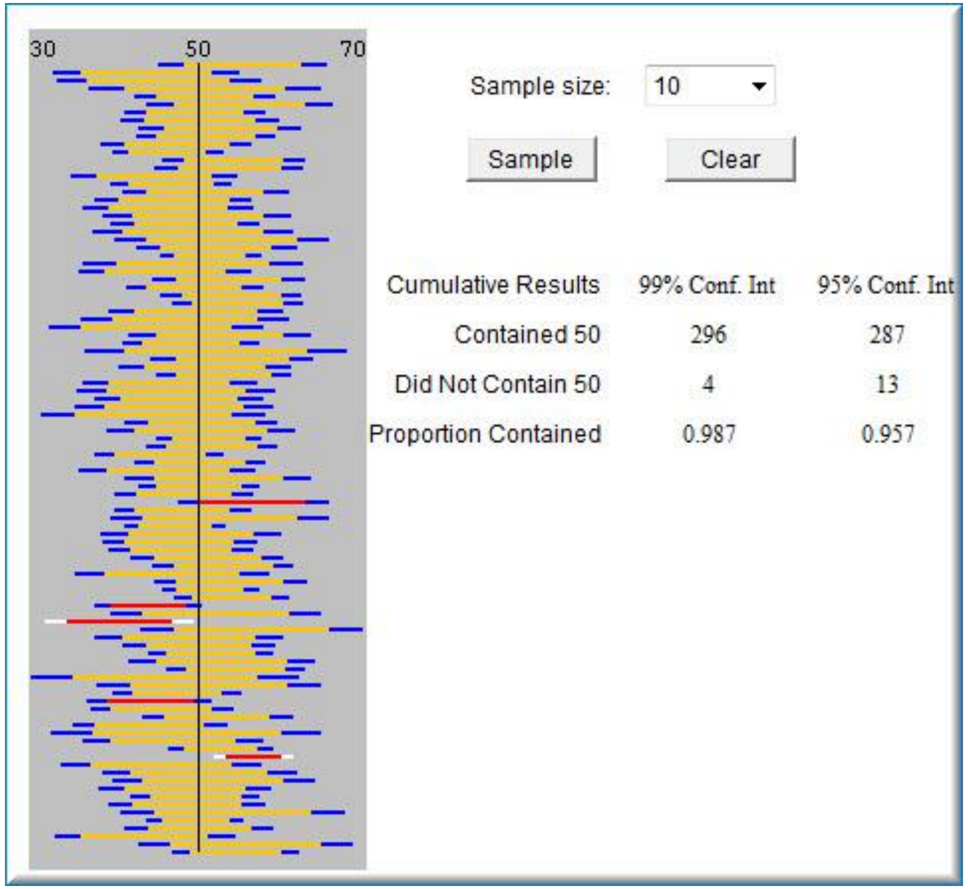
- We want to estimate the population mean.
- We take a simple random sample. Use it to calculate interval $[a, b]$.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- Confidence intervals are interval estimators.
- What are, for example, 95% confidence intervals?
- We want to estimate the population mean.
- We take a simple random sample. Use it to calculate interval $[a, b]$.

- If you repeat this procedure many times, 95% of the intervals we calculated contains the true, non-random, population mean.

- Alternatively, this procedure has a 95% chance of a producing a interval that contains the mean.



Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- How to construct a 95% confidence intervals? *Reverse engineering*.

- Take a simple random sample of size n, calculate sample mean M.
- Assuming we know the population variance σ^2.

- The sampling distribution of the sample mean can be approximated by a normal distribution with mean $\mu_M = \mu$ and variance $\sigma_M^2 = \frac{\sigma^2}{n}$.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- How to construct a 95% confidence intervals? *Reverse engineering*.

- Take a simple random sample of size n, calculate sample mean M.
- Assuming we know the population variance σ^2.

- The sampling distribution of the sample mean can be approximated by a normal distribution with mean $\mu_M = \mu$ and variance $\sigma_M^2 = \frac{\sigma^2}{n}$.

- This means that 95% of the time, the sample mean will be within 1.96 standard errors of the population mean, using the z-value table.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- The sampling distribution of the sample mean can be approximated by a normal distribution with mean $\mu_M = \mu$ and variance $\sigma_M^2 = \frac{\sigma^2}{n}$.
- This means that 95% of the time, the sample mean will be within 1.96 standard errors of the population mean.

- *Reverse engineering*: turn this around and say 95% of the time, the population mean (fixed, non-random quantity) will end up within 1.96 standard errors of the mean of a simple random sample.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- The sampling distribution of the sample mean can be approximated by a normal distribution with mean $\mu_M = \mu$ and variance $\sigma_M^2 = \frac{\sigma^2}{n}$.
- This means that 95% of the time, the sample mean will be within 1.96 standard errors of the population mean.

- *Reverse engineering*: turn this around and say 95% of the time, the population mean (fixed, non-random quantity) will end up within 1.96 standard errors of the mean of a simple random sample.

- The 95% confidence interval is $[\, M - 1.96\,\sigma_M, M + 1.96\,\sigma_M]$.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- What if we want a general $\alpha\%$ confidence interval?

- Repeat the same process.

- Then, the $\alpha\%$ confidence interval is

$$[\, M \; - Z_\alpha \, \sigma_M, M + Z_\alpha \, \sigma_M]$$

- Make sure you can do this for a general $\alpha\%$, for the exam!!!!!!!!!!!

Chapter 10, Section 9 – t distribution *FINALLY*

- What if we want a general $\alpha\%$ confidence interval?
- BUT the population variance is not known?

Chapter 10, Section 9 – t distribution *FINALLY*

- What if we want a general $\alpha\%$ confidence interval?
- BUT the population variance is not known?

- We estimate the population variance.
- We use the estimator of the population variance s^2.

Chapter 10, Section 9 – t distribution *FINALLY*

- We estimate the population variance.
- We use the estimator of the population variance s^2.

- Unfortunately, this only works if the population is **normally distributed**.

Student's t Distribution

- Has parameter: degrees of freedom, which is n – 1.
- n = sample size.
- Actually has infinite variance for degrees of freedom = 2.

History and etymology [edit]

Statistician William Sealy Gosset, known as "Student"

In the English-language literature the distribution takes its name from William Sealy Gosset's 1908 paper in *Biometrika* under the pseudonym "Student".[8][9] Gosset worked at the Guinness Brewery in Dublin, Ireland, and was interested in the problems of small samples – for example, the chemical properties of barley where sample sizes might be as few as 3. One version of the origin of the pseudonym is that Gosset's employer preferred staff to use pen names when publishing scientific papers instead of their real name, so he used the name "Student" to hide his identity. Another version is that Guinness did not want their competitors to know that they were using the *t*-test to determine the quality of raw material.[10][11]

Source: Wikipedia.

Chapter 10, Section 9 – t distribution *FINALLY*

- We estimate the population variance.
- We use the estimator of the population variance s^2.

- Unfortunately, this only works if the population is **normally distributed**.

- Then, the α% confidence interval, using degrees of freedom $df = n - 1$, is

$$[\, M - t_{\alpha,df}\, s_M, M + t_{\alpha,df}\, s_M]$$

Break time!!   \o/

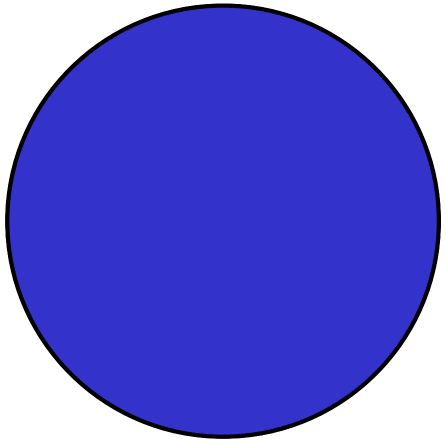
- Break starts after I hand out the exercise.

- Meant to lead into hypothesis testing, and practice constructing confidence intervals.

- Circle is a timer that becomes blue. O_o   →

(please ignore if it glitches)

12 minutes



Chapter 9, Section 9 – Sampling Distribution of p

- Population with N individuals. A proportion $\prod$ of them are of type A, and the rest are of type B.

- E.g. Type A = those who voted for candidate A, and type B = those who voted for candidate B.

- E.g. Heads and tails in large number of coin flips. (yes, casinos hire PhD mathematicians to do this for their non-coin flip games)

Chapter 9, Section 9 – Sampling Distribution of p

- Take a simple random sample of size n.
- You can see this sample as an experiment with n trials and probability of "success" $\prod$.
- Technically, we are *sampling with replacement*.

Chapter 9, Section 9 – Sampling Distribution of p

- Let p be the proportion of type A ("successes") in your sample.
- This p has sampling distribution with mean $\prod$.
- Standard error of p is $\sigma_p = \sqrt{\frac{\prod(1-\prod)}{n}}$
- IF we know what $\prod$ is.
- The sampling distribution is approximately normally distributed for large n.

Chapter 10, Section 13 – Proportions

- IN THE EXAM, we will give you situations where you don't have the population proportion $\prod$.

- The *estimator of the* standard error is,

$$S_p = \sqrt{\frac{p(1-p)}{n}}.$$

Chapter 10, Section 13 – Proportions

- IN THE EXAM, we will give you situations where you don't have the population proportion $\prod$.

- The *estimator of the* standard error is,

$$S_p = \sqrt{\frac{p(1-p)}{n}}.$$

- Just like for the t-distribution, we will use this for the standard error instead.
- However, we STILL stick to the normal distribution.

Chapter 10, Section 13 – Proportions

- **TEXTBOOK** : We use the normal approximation of the Binomal distribution, adjusting it by $\frac{0.5}{n}$
- The α% confidence interval is,

$$[\, p - Z_\alpha \, S_p - \frac{0.5}{n} \quad , \quad p + Z_\alpha \, S_p + \frac{0.5}{n}]$$

Chapter 10, Section 13 – Proportions

- **TEXTBOOK** : We use the normal approximation of the Binomal distribution, adjusting it by $\frac{0.5}{n}$
- The $\alpha\%$ confidence interval is,

$$[\, p - Z_\alpha\, S_p - \frac{0.5}{n} \quad , \quad p + Z_\alpha\, S_p + \frac{0.5}{n}]$$

- **EXAM** : We use the normal approximation of the Binomal distribution,
- The $\alpha\%$ confidence interval is,

$$[\, p - Z_\alpha\, S_p \quad , \quad p + Z_\alpha\, S_p]$$

Chapter 10, Section 13 – Proportions

Sample Exam Question

- You are a small employee in a big chain of supermarkets. Your boss wants to know if the company's image is attracting more men than women, or is it roughly 50/50.
- Population has two types: Men and Women.

Chapter 10, Section 13 – Proportions

- You are a small employee in a big chain of supermarkets. Your boss wants to know if the company's image is attracting more men than women, or is it roughly 50/50.
- Population has two types: Men and Women.
- You take a sample of size n = 10, because you're a lazy employee.
- You find that your sample proportion of men is 60% or 0.60 or 3/5.
- You want to construct a 90% confidence interval for the population proportion of men, because you took Math 10 before.

Chapter 10, Section 13 – Proportions

- You take a sample of size n = 10, because you're a lazy employee.
- You find that your sample proportion of men is 60% or 0.60 or 3/5.
- You want to construct a 90% confidence interval for the population proportion of men, because you took Math 10 before.

- How would you do it? (6 pts)