SKJ Education

LC HL MATHS

FOUNDATION PROGRAM: WEEK 7

PROBABILITY

Steven James
steven@skjeducation.com
www.skjeducation.com

LC HL MATHEMATICS – FOUNDATION PROGRAM

Week 7: Probability & Statistics

Learning Objectives

- **7.1:** To calculate conditional probabilities using tree diagrams and two-way tables and differentiate between independent and mutually exclusive events.
- **7.2:** To calculate and interpret measures of central tendency (mean, median, mode) and spread (standard deviation, interquartile range).
- **7.3:** To calculate probabilities using Normal Distribution and standard z-scores.
- **7.4:** To construct and interpret confidence intervals for a population mean.
- **7.5:** To perform a full hypothesis test for a population mean (one-sample z-test), stating hypotheses, test statistic, p-value, and conclusion in context.

Key Terms - Week 7

- **Conditional Probability:** The probability of an event occurring given that another event has already occurred, denoted as $P(A|B)$.
- **Independent and Mutually Exclusive Events:** Events are independent if the occurrence of one does not affect the probability of the other, while mutually exclusive events cannot occur simultaneously.
- **Measures of Central Tendency and Spread:** The mean, median, and mode describe the centre of a dataset, while standard deviation and interquartile range describe the spread.
- **Normal Distribution:** A continuous probability distribution with a characteristic bell-shaped curve, defined by its mean and standard deviation.
- **Standard z-Score:** A z-score is a measure of how many standard deviations an observation is from the mean, calculated as $z = \frac{X-\mu}{\sigma}$.
- **Confidence Interval:** A range of values within which a population parameter is likely to lie, constructed using sample data and a specified confidence level.
- **Hypothesis Test:** A statistical test used to determine whether a hypothesis about a population parameter is supported by sample data.

Weekly *Challenge*: A company is testing a new manufacturing process to improve the average weight of their product. Using a sample of products, perform a one-sample z-test to determine whether the new process has resulted in a significant change in the average weight. ***State the hypotheses, calculate the test statistic and p-value, and interpret the results in context***.

WEEK 7 STUDY PLAN

Day	**Activities & Time Commitment**	✓	**Rating (1-10)**
Monday	- Review Learning Objectives (5 min) - Rank your current ability (5 min) - Review Key Terms (10 min) - Watch video (10-20 min) *Focus: PREPARATION*		
Tuesday	- Complete Exercises A1 & A2 (60 min) *Focus: EXPLORING*		
Wednesday	- Complete Exercise A3 (30 min) - Correct Exercises A1-3 and reattempt difficult questions (45 min) *Focus: PROCESSING*		
Thursday	- 1-hour online lesson (60 min) *Focus: QUESTIONING*		
Friday	- Complete Exercise B (40 min) *Focus: ERROR ANALYSIS*		
Saturday	- Complete Exam Question Assessment (C) (60 min) *Focus: EXECUTION*		
Sunday	- Correct assessment (30 min) - Complete self-reflection (15 min) - Plan next week (15 min) *Focus: REFLECTION & RECHARGING*		

Study Tips for Success

- **Active Recall:** After studying, close your notes and write down **everything you remember**. Force your brain to grow.
- **Spaced Repetition:** Review concepts **multiple times** over several days.
- **Mathematics in Action:** Look for **real-world examples** of the concepts you're learning.
- **Ask Questions:** Don't hesitate to ask for help when concepts are unclear. Reach out via *Google Classroom* or email; *steven@skjeducation.com*.
- **Celebrate Progress:** **Acknowledge your improvements**, no matter how small.

A1. Proficiency Drills (Week 7)

Learning Focus: Building foundational skills in **probability and statistics**, covering conditional probability, descriptive statistics, the Normal distribution, confidence intervals, and hypothesis testing.

Part 1: Conditional Probability and Events

Key Concepts

Conditional Probability: The probability of event A occurring given that event B has occurred, denoted $P(A|B)$. Formula: $P(A|B) = \frac{P(A\cap B)}{P(B)}$.

- **Independent Events:** The occurrence of one event does not affect the probability of the other. $P(A|B) = P(A)$ or $P(A \cap B) = P(A)P(B)$.
- **Mutually Exclusive Events:** Two events that cannot occur at the same time. $P(A \cap B) = 0$.

Tools: Two-way tables and tree diagrams are excellent for organizing information in probability problems.

Task #1: A survey of 100 students found that 60 take Physics (P), 50 take Chemistry (C), and 20 take both.

1. Create a two-way table to represent this data.
2. A student is chosen at random. Find the probability that the student takes Physics, given that they take Chemistry.
3. Are the events "takes Physics" and "takes Chemistry" independent? Justify your answer.

Part 2: Descriptive Statistics - Measures of Center and Spread

Essential Summary Statistics

Measures of Central Tendency:

- **Mean ($\bar{x}$):** The average of the data.
- **Median:** The middle value when the data is ordered.
- **Mode:** The most frequently occurring value.

Measures of Spread:

- **Standard Deviation (s):** A measure of how spread out the data is from the mean.
- **Interquartile Range (IQR):** The range of the middle 50% of the data. IQR $= Q_3 - Q_1$.

Task #2: For the following dataset of test scores: $\{2, 4, 6, 8, 10\}$, calculate the:

1. Mean, Median, and Mode.
2. Interquartile Range (IQR).
3. Sample Standard Deviation.

Part 3: The Normal Distribution

The Bell Curve

The Normal Distribution is a continuous probability distribution that is symmetric about the mean.

- **Standardization (z-score):** We convert any normal variable X with mean μ and standard deviation σ to the standard normal variable $Z \sim N(0, 1)$ using the formula:
$$Z = \frac{X - \mu}{\sigma}$$
- **Probabilities:** We use a Z-table or calculator to find the area under the standard normal curve, which corresponds to the probability.

Task #3: The scores on a standardized test are normally distributed with a mean (μ) of 100 and a standard deviation (σ) of 15.

1. Find the probability that a randomly selected student scores more than 120.
2. Find the probability that a randomly selected student scores between 90 and 115.

Part 4: Inferential Statistics - Estimation and Testing

Making Inferences About a Population

We use sample data to make educated guesses about a population.

- **Confidence Interval for a Mean (known σ):** A range of plausible values for the true population mean μ. Formula: $\bar{x} \pm Z^* \frac{\sigma}{\sqrt{n}}$.
- **Hypothesis Testing (One-sample z-test):** A formal procedure to test a claim about a population mean μ.
  1. State Hypotheses: Null (H_0) and Alternative (H_a).
  2. Calculate Test Statistic: $Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.
  3. Find p-value: The probability of observing a result as extreme as the sample, assuming H_0 is true.
  4. Conclude: If p-value $< \alpha$ (significance level), reject H_0. Otherwise, fail to reject H_0.

Task #4: A machine is designed to fill bags with 500g of coffee. It is known that the standard deviation of the fill weight is $\sigma = 4$g. A random sample of 64 bags shows a sample mean weight of $\bar{x} = 501.5$g.

1. Construct and interpret a 95% confidence interval for the true mean fill weight. (For 95% confidence, $Z^* = 1.96$).
2. Perform a hypothesis test at the $\alpha = 0.05$ significance level to determine if the machine is working correctly (i.e., if the mean weight is different from 500g).

[Answers]

- **Task #1:**
  1.

	Chem	No Chem	Total
Physics	20	40	60
No Physics	30	10	40
Total	50	50	100

  2. $P(P|C) = \frac{P(P\cap C)}{P(C)} = \frac{20/100}{50/100} = \frac{20}{50} = 0.4$.
  3. To be independent, $P(P|C)$ must equal $P(P)$. Here, $P(P|C) = 0.4$ and $P(P) = 60/100 = 0.6$. Since $0.4 \neq 0.6$, the events are **not independent**.
- **Task #2:**
  1. Mean $\bar{x} = \frac{2+4+6+8+10}{5} = 6$. Median = 6. Mode = None.
  2. Ordered data: $\{2, 4, 6, 8, 10\}$. $Q_1 = 3$, $Q_3 = 9$. IQR $= 9 - 3 = 6$.
  3. Variance $s^2 = \frac{(2-6)^2+(4-6)^2+(6-6)^2+(8-6)^2+(10-6)^2}{5-1} = \frac{16+4+0+4+16}{4} = 10$. Standard Deviation $s = \sqrt{10} \approx 3.16$.
- **Task #3:**
  1. $Z = \frac{120-100}{15} \approx 1.33$. $P(X > 120) = P(Z > 1.33) = 1 - P(Z < 1.33) = 1 - 0.9082 = 0.0918$.
  2. $Z_1 = \frac{90-100}{15} \approx -0.67$. $Z_2 = \frac{115-100}{15} = 1$. $P(90 < X < 115) = P(-0.67 < Z < 1) = P(Z < 1) - P(Z < -0.67) = 0.8413 - 0.2514 = 0.5899$.
- **Task #4:**
  1. CI $= 501.5 \pm 1.96 \times \frac{4}{\sqrt{64}} = 501.5 \pm 1.96 \times 0.5 = 501.5 \pm 0.98$. The 95% CI is $[500.52, 502.48]$. We are 95% confident that the true mean fill weight is between 500.52g and 502.48g.
  2.
     - **Hypotheses:** $H_0 : \mu = 500$; $H_a : \mu \neq 500$.
     - **Test Statistic:** $Z = \frac{501.5-500}{4/\sqrt{64}} = \frac{1.5}{0.5} = 3.0$.
     - **p-value:** For a two-tailed test, p-value $= 2 \times P(Z > 3.0) = 2(1 - 0.99865) = 0.0027$.
     - **Conclusion:** Since the p-value (0.0027) is less than α (0.05), we reject H_0. There is sufficient evidence to conclude that the true mean fill weight is different from 500g.

Exercise 8: 95% Confidence Interval – Using σ Instead of s

Question: A sample of $n = 36$ has mean $\bar{x} = 82$ and sample standard deviation $s = 12$. Construct a 95% confidence interval for the population mean.

Incorrect Calculation: Margin $= 1.96 \times \frac{\sigma}{\sqrt{n}} = 1.96 \times \frac{12}{6} = 3.92$. CI $= (82 \pm 3.92)$.

Flawed Thinking	Correct Approach
Error Analysis:	**Correction:**

Exercise 9: One–Sample z–Test – Hypotheses Reversed

Question: A machine is claimed to fill bottles with $\mu = 500$ ml. A sample of $n = 40$ bottles has $\bar{x} = 494$ ml, $\sigma = 20$ ml. Test at $\alpha = 0.05$ if the mean volume differs from 500 ml (two–sided).

Incorrect Statement: $H_0: \ \mu \neq 500, \quad H_1: \ \mu = 500$.

Flawed Thinking	Correct Approach
Error Analysis:	**Correction:**

Exercise 10: p–Value Interpretation

Question: In a one–sample z–test the p–value is 0.08 with $\alpha = 0.05$. State the conclusion.

Incorrect Conclusion: Because $0.08 < 0.05$, we reject H_0 and conclude the result is significant.

Flawed Thinking	Correct Approach
Error Analysis:	**Correction:**

C. Weekend Assessment – Past Exam Questions

Learning Focus: Applying learning to past exam questions under exam conditions.

Practical Advice: To tackle the questions in this assessment, consider the following tips:

- For probability distributions (Q1(a)), set up the expected value equation carefully.
- When working with mutually exclusive and independent events (Q1(b), Q1(d), Q5(a)), understand the definitions and relationships.
- For normal distribution problems (Q2, Q6, Q7), use z-scores and the empirical rule appropriately.
- When comparing performances across different distributions (Q2(b)(i)), convert to standardized scores.
- For counting problems (Q3(a)), consider positions and arrangements systematically.
- When working with confidence intervals (Q4(a), Q6(a)(i)), use the correct formulas and interpretations.
- For hypothesis testing (Q4(b), Q6(b), Q7(b)), state hypotheses clearly and interpret p-values correctly.
- When dealing with combinatorial probability (Q7(c)), consider sampling with and without replacement carefully.

General Tips:

- **Show your work** - partial credit is available for correct methods
- **Check domains/restrictions** - especially in probability calculations
- **Label answers clearly** - especially when questions have multiple parts
- **Use appropriate precision** - note where decimal places are required
- **Verify your answers** - do they make sense in the context of the problem?

Question 1 (30 marks)
(a) The table below shows the prizes, in euro, that a player can win in a game, as well as the probability of winning each prize. The player wins at most one prize each time that she plays. Some of the prizes are given in terms of $x \in \mathbb{R}$.

Prize (€)	None	2	$x - 10$	x
Probability	30%	40%	28%	2%

It costs €10 to play the game once.
The game is fair – that is, the expected value of the winnings, taking the cost into account, is €0.
Work out the value of x.
(b) A and B are mutually exclusive events.

$$P(A) = 0.1 \quad \text{and} \quad P(B) = 0.4.$$

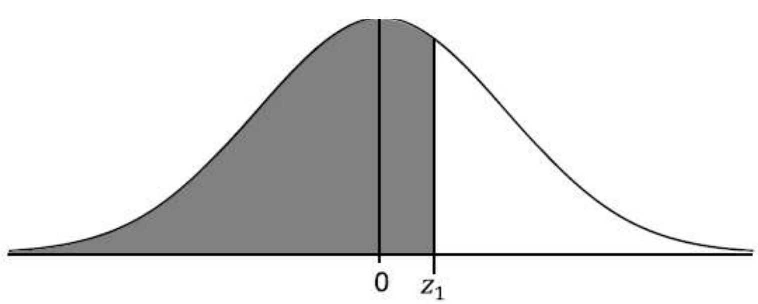
Write down the value of each of the following:

$$P(A \cap B) =$$

$$P(A \cup B) =$$

(c) C and D are two other events, with universal set U. $P(C) = 0.5$ and $P(D) = 0.7$. Find the maximum value of $P[(C \cup D)']$.
Note: $(C \cup D)'$ is the complement of the set $C \cup D$ in the set U.
(d) E is the event that it will be raining tomorrow morning. F is the event that I will wear a coat going outside tomorrow morning.
Explain why it would not be reasonable to assume that E and F are independent events.

Question 2 (25 marks)
(a) The diagram shows the standard normal curve. The shaded area represents 67% of the data. Find the value of z_1.



(b) The percentage results in a Maths exam for a class had a mean mark of 70 with a standard deviation of 15. The percentage results in an English exam for the same class had a mean mark of 72 with a standard deviation of 10. The results in both exams were normally distributed.
(i) Mary got 65 in Maths and 68 in English. In which exam did Mary do better relative to the other students in the class? Justify your answer.
(ii) In English the top 15% of students were awarded an A grade. Find the least whole number mark that merited the award of an A grade in English.
(iii) Using the empirical rule, or otherwise, estimate the percentage of students in the class who scored between 52 and 82 in the English test.

Question 3 (25 marks)
(a) A security code consists of six digits chosen at random from the digits 0 to 9. The code may begin with zero and digits may be repeated.
For example

0 7 1 7 3 7

is a valid code.
(i) Find how many of the possible codes will end with a zero.
(ii) Find how many of the possible codes will contain the digits 2 0 1 8 together and in this order.
(b) Find a, b, c, and d, if

$$\frac{(n+3)!(n+2)!}{(n+1)!(n+1)!} = an^3 + bn^2 + cn + d,$$

where a, b, c, and $d \in \mathbb{N}$.

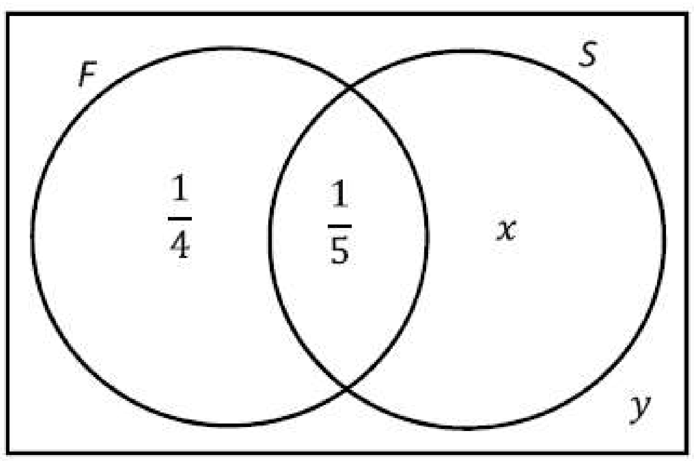
Question 4 (30 marks)
(a) A survey on remote learning was carried out on a random sample of 400 students. 135 of the students preferred remote learning over in-person learning. For parts (a)(i), (a)(ii), and (a)(iii), give all solutions as decimals, correct to 4 decimal places.
(i) Work out the proportion of the sample that preferred remote learning.
(ii) Use the margin of error $\left(\frac{1}{\sqrt{n}}\right)$ to create a 95% confidence interval for the proportion of the population that preferred remote learning.
(iii) Using the proportion from part (a)(i), create a 95% confidence interval for this population proportion that is more accurate than the 95% confidence interval based on the margin of error.

(b) In 2019, people with a pre-pay mobile phone plan spent an average (mean) of €20.79 on their mobile phone each month (source: www.comreg.ie).
In 2021, some students carried out a survey to see if this figure had changed. They surveyed a random sample of 500 people with pre-pay mobile phone plans. For this sample, the mean amount spent per month was €22.16 and the standard deviation was €8.12.
Carry out a hypothesis test at the 5% level of significance to see if this shows a change in the mean monthly spend on mobile phones for people with a pre-pay plan. State your null hypothesis and your alternative hypothesis, state your conclusion, and give a reason for your conclusion.

Question 5 (25 marks)
(a) Two independent events F and S are represented in the Venn diagram shown below.

$$P(F \setminus S) = \frac{1}{4}, \quad P(F \cap S) = \frac{1}{5}, \quad P(S \setminus F) = x, \text{ and } P(F \cup S)' = y, \text{ where } x, y \neq 0.$$

Find the value of x and the value of y.



(b) In a club there are German, Irish and Spanish children only. There are 10 Spanish children. There are twice as many Irish children as German children.
They are all in a group waiting to get on a swing. One child will be selected at random to go first and will not re-join the group. Then a second child will be selected at random to go next.
The probability that the first child selected will be German and that the second child selected will not be German is $\frac{1}{6}$ Find how many children are in the club.

Question 6 (45 marks)
(a) A motoring magazine collected data on cars on a particular stretch of road. Certain details on 800 cars were recorded.
(i) The ages of the 800 cars were recorded. 174 of them were new (less than 1 year old). Find the 95% confidence interval for the proportion of new cars on this road. Give your answer correct to 4 significant figures.
(ii) The data on the speeds of these 800 vehicles is normally distributed with an average speed of 87.3 km per hour and a standard deviation of 12 km per hour. What proportion of cars on this stretch of road would you expect to find travelling at over 95 km per hour?
(iii) The driver of a car was told that 70% of all the speeds recorded were higher than his speed. Find the speed at which this driver was recorded. Give your answer correct to the nearest whole number.
(b) (i) A road safety programme was carried out in the area using posters, signs and radio slots. After the programme the motoring magazine recorded the speeds of 100 passing cars. The magazine carried out a hypothesis test, at the 5% level of significance, to determine whether the average speed had changed. The p-value of the test was 0.024. What can the magazine conclude based on this p-value? Give a reason for your answer.
(ii) The magazine found that the average speed of this sample was lower than the previously established average speed of 87.3 km per hour. Find the average speed of the cars in this sample, correct to 1 decimal place.

Question 7 (50 marks)

(a) In a school all First Years sat a common maths exam. The results, in integer values, were normally distributed with a mean of 176 marks and a standard deviation of 36 marks. The top 10% of students will go forward to a county maths competition.

(i) Find the minimum mark needed on the exam to progress to the county stage.

(ii) The school awarded a Certificate of Merit to any student who achieved between 165 marks and 210 marks. Find the percentage of First Years who received the Certificate of Merit.

(b) A news report claimed that 6th year students in Ireland studied an average of 21 hours per week, outside of class time. A Leaving Cert class surveyed 60 students in 6th year, chosen at random, from different schools. It found that the average study time was 19.8 hours and the standard deviation was 5.2 hours.

(i) Find the test statistic (the z-score) of this sample mean.

(ii) Find the p-value of this test statistic. Comment on what can be concluded from its value, in a two-tailed hypothesis test at the 5% level of significance, in relation to the news report claim.

(c) The school caretaker has a box with 23 room keys in it. 12 of the keys are for general classrooms, 6 for science labs and 5 for offices.

(i) Four keys are drawn at random from the box. What is the probability that the 4th key drawn is the first office key drawn? Give your answer correct to 4 decimal places.

(ii) All the keys are returned to the box. Then 3 keys are drawn at random from the box one after the other, without replacement. What is the probability that one of them is for a general classroom, one is for a science lab and one is for an office? Give your answer correct to 4 decimal places.

Self-Assessment

After completing the assessment:

- Grade your work honestly
- Identify areas needing improvement
- Scan and submit via Google Classroom
- Reflect on your performance in your weekly reflection

Another excellent week of work completed - ***well done!*** You are another step closer to *smashing your exams*, and another week closer to your summer holdiays!

Weekly Reflection Zone

What worked well this week?

What challenges did I face?

What surprised me the most this week?

Key mathematics concepts I want to review:

Goals for next week: