SUPERVISOR'S USE ONLY

3

91585

915850

Draw a cross through the box (☒) if you have NOT written in this booklet

+

NZQA

Mana Tohu Mātauranga o Aotearoa
New Zealand Qualifications Authority

# Level 3 Mathematics and Statistics (Statistics) 2024

## 91585 Apply probability concepts in solving problems

Credits: Four

| Achievement | Achievement with Merit | Achievement with Excellence |
|---|---|---|
| Apply probability concepts in solving problems. | Apply probability concepts, using relational thinking, in solving problems. | Apply probability concepts, using extended abstract thinking, in solving problems. |

Check that the National Student Number (NSN) on your admission slip is the same as the number at the top of this page.

**You should attempt ALL the questions in this booklet.**

Make sure that you have the Formulae and Tables Booklet L3–STATF.

Show ALL working.

If you need more room for any answer, use the extra space provided at the back of this booklet.

Check that this booklet has pages 2–16 in the correct order and that none of these pages is blank.

Do not write in any cross-hatched area (▨). This area will be cut off when the booklet is marked.

**YOU MUST HAND THIS BOOKLET TO THE SUPERVISOR AT THE END OF THE EXAMINATION.**

Excellence TOTAL 24

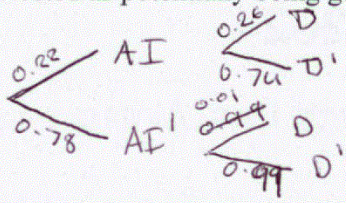
## QUESTION ONE

(a) Educators use writing detectors like Turnitin to detect if students have used AI (Artificial Intelligence) to write their assignments. When designing writing detectors, scientists focus on accuracy; if they say AI writing is present in a piece of work, they want to be pretty certain that the work is AI-generated, to ensure that students are not falsely accused of misconduct. This means that the detector may not always detect all AI writing; some may be missed.

A false positive in AI writing detection refers to incorrectly identifying fully human-written text as AI-generated. Suppose that a particular AI writing detector has a 1% false positive rate but that 74% of the time it fails to detect AI-generated writing.

Based on a confidential survey of students in one teacher's course, it is thought that 22% of pieces of student writing contain content that is generated by AI.

(i) If 120 pieces of writing are screened by this AI detector, approximately how many would be detected as potentially being generated by AI?




$P(D) = (0.22 \times 0.26) + (0.78 \times 0.01) = 0.065$

$0.065 \times 120 = 7.8 \approx 8$. Approximately 8 pieces of writing

(ii) A piece of writing is detected as potentially being generated by AI.

Comment on whether a teacher should be concerned that the student could be unfairly accused of cheating.

Support your answer with statistical reasoning.

$P(AI/D) = \frac{0.22 \times 0.26}{0.065} = 0.88$

$P(AI'/D) = \frac{0.78 \times 0.01}{0.065} = 0.12$

So the probability that this piece of writing was unfairly picked up as being AI was 12%, while there is an 88% chance it is actually AI. $\frac{0.88}{0.12} = 7.33...$ ~~the writing is 7.33 times as~~ while 12% is a small chance of the student being unfairly accused, it is large enough that the teacher should be careful with the accusation as with this probability, 1 in 10 students caught with the detector would be falsely accused.

(b) Exam supervisors are sometimes concerned that students who complete tests quickly might have cheated. Based on data collected from one school, for a particular NCEA standard with an allocated time of 60 minutes, it is known that:

- 1% of students cheat on the assessment for this standard
- 20% of students complete the assessment in less than 25% of the allocated time
- 80% of students who cheat on the assessment complete it in less than 25% of the allocated time.

(i) Comment on whether the events, 'student cheats' and 'student completes the assessment in less than 25% of the allocated time', are independent of each other.

Use statistical reasoning to support your answer.

$P(T<25\%) = 0.2$ $P(T<25\% \mid C) = 0.8$

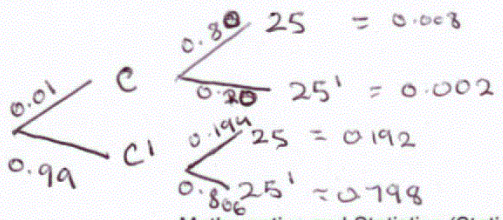
so since $P(A) \neq P(A/B)$, these two events are not independent; whether the student cheated or not does have an impact on the probability of finishing the test in less than 25% of the allocated time.

AND interpret your answer in terms of the relationship between the events 'student cheats' and 'student completes the assessment in less than 25% of the allocated time'.

This ✓lack of ✓dependence means that there is a relationship between the two events 'student cheats' and 'student completes the assessment in less than 25% of the allocated time' (i.e. the ~~events' probabilities~~ each event's probability is dependent on the outcome of the other). If the student cheated, there is an 0.8 % chance of them taking less than 25% of the time, but if they didn't cheat, there is an 0.194 chance of them taking less than 25% of the time $(0.2 - (0.8 \times 0.01))/0.99$. so cheaters ~~have a~~ are over six more times likely to finish the test in less than 25% of the time than non-cheaters $(0.8/0.194) = 4.12$

(ii) Estimate the proportion of students across New Zealand who cheat and complete the assessment for this standard in less than 25% of the allocated time.

$P(C \cap 25) = 0.01 \times 0.8 = 0.008$



(iii) Give TWO reasons why care should be taken when using this data to estimate the proportion of students being assessed for NCEA who will cheat and complete the assessment in less than 25% of the allocated time.

Reason one: This data is collected based on 1 NCEA Standard. since different standards have differing levels of difficulty, different proportions of cheaters and non-cheaters will complete the exam quickly. E.g. if an exam is really easy, more people will complete it in less than 25% of the time without more people necessarily cheating. So because this sample data only includes data on 1 standard, care should be taken extending the results to all standards, as this 1 sample standard doesn't necessarily represent all NCEA standards. If you wanted to extend the results and use this data to estimate proportions for all NCEA students being assessed, data should be collected from a much larger group of (ideally randomly selected) standards.

Reason two: This data has also been collected from 1 school. this school's population may not accurately reflect the different demographics in the population i.e. the school might be very high or very low decile, or might have different teaching methods to the other schools, which might give different results. E.g. if the teachers at this school teach a really quick method to solve the problems in the paper, a higher proportion of non-cheating students will finish in less than 25% of the time than at other schools where this method has not been taught. Therefore the data collected from this sample can't be confidently extended to the population and be used to ~~estimate~~ estimate the proportion of students being assessed for NCEA who will cheat and complete the assessment in less than 25% of the allocated time, as the sample doesn't necessarily represent the population and was not randomly selected, so there may be sampling bias.

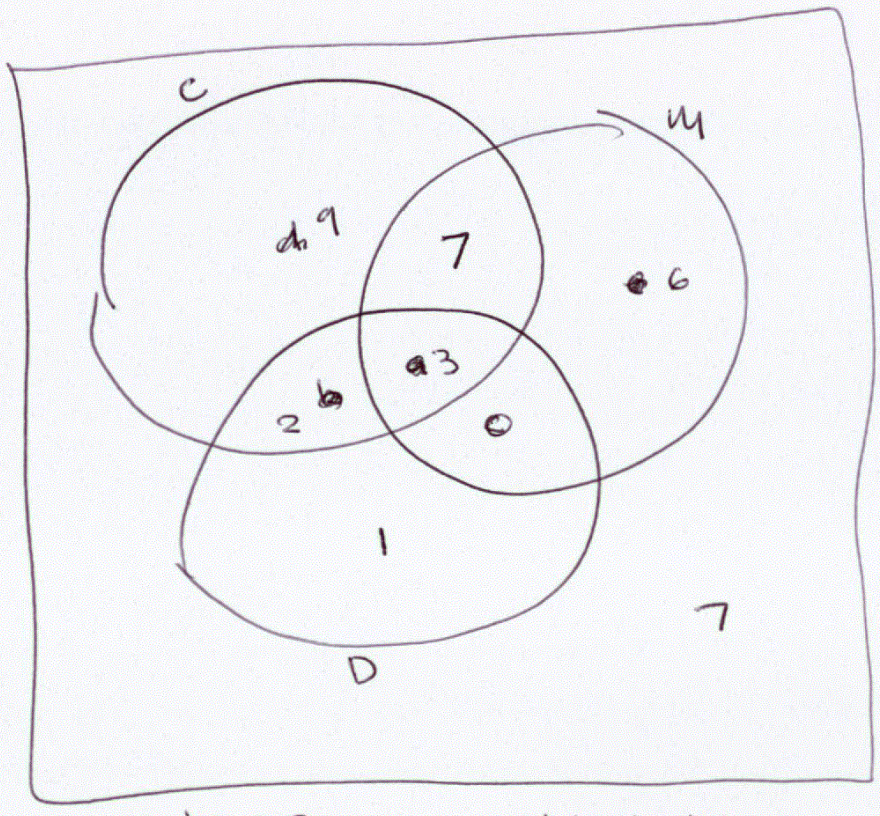
## QUESTION TWO

(a) Data science is commonly thought of as the intersection of three main skill sets: coding, mathematical and statistical knowledge, and domain (or subject-specific) understanding, e.g. finance, biology, health.

From 35 applications for a data science role, where all three skill sets were desired, the following information is known:

- 7 applicants had none of the three skill sets
- 21 applicants had coding skills
- 16 applicants had mathematical and statistical knowledge
- 6 applicants had the necessary subject-specific understanding
- 9 applicants had exactly two of the three main skill sets
- 7 applicants were good candidates for the role except that they didn't have any subject-specific knowledge
- 1 applicant had only subject-specific understanding
- out of those with only one main skill set, $\frac{3}{8}$ had only mathematical and statistical knowledge.

*space for working*




$b = 5 - a$

$a + b = 5$

OR

$\geq 0$ $a + b + c = 5$

$, c = 0$

$e = \frac{3(16)}{5}$

$e = 6$

$b = 2$

if $a + b + c + e = 9$

$a + e = 9$

$a = 3$

$b + c = 2$

$(d + e + 1) \times \frac{3}{8} = e$

$a + b + c + d + e = 20$

$3d + 3e + 3 = 8e$

$d + a + b = 14$

$a + c + e = 9$

$a + b + c = 5$

$5e + 3d = 3$

$e = \frac{3(d+1)}{5}$

$d + e = 15$

$d + \frac{3(d+1)}{5} = 15$

$5d + 3d + 3 = 75$

$8d = 72$

$d = 9$

(i) A shortlist of people with all three skill sets was created.

How many applicants were selected for the shortlist?

3 applicants

(ii) Calculate the proportion of applicants who had only one of the required skills.

$9 + 6 + 1 / 35 = 0.457$

(iii) It is claimed that, for those applicants with the required mathematical and statistical knowledge, applicants are twice as likely to have subject-specific understanding compared to coding skills.

Does this data support the claim?

Use calculations and statistical reasoning to support your answer.

$P(D/M) = 3/16 = 0.1875$

$P(C/M) = 10/16 = 0.625$

since $P(D/M) < P(C/M)$, no, the claim is not true

$\frac{P(C/M)}{P(D/M)} = 3.333.$

So applicants with the required mathematical and statistical knowledge are actually over 3 times as likely to have coding skills compared to subject specific knowledge, not the other way around.

(b) 'In 2022, 82.4% of all 18-year-olds attained the equivalent of NCEA Level 2 or above. Of those who turned 18 in 2022, 78.3% of them attained at least NCEA Level 2 or above in school, and 4.1% of them attained at least NCEA Level 2 or above post-school in a tertiary or vocational setting.'

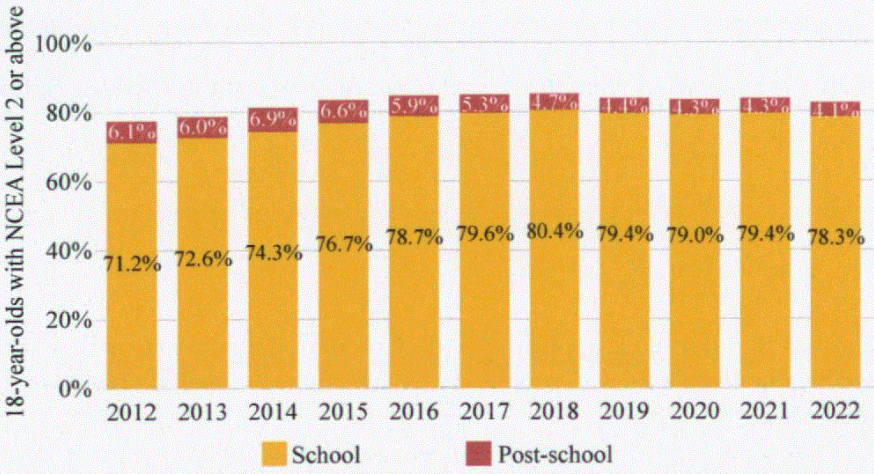



**Figure 1: Most 18-year-olds with NCEA Level 2 or above attained their qualification in school**

Source: www.educationcounts.govt.nz/statistics/18-year-olds-with-level-2-or-equivalent

(i) Based on the data, is the suggestion that the total proportion of 18-year-olds with NCEA Level 2 has increased between 2012 and 2022 correct?

Support your answer with statistical reasoning.

P(L2+ in 2012) = 77.3%.

P(L2+ in 2022) = 82.4%.

82.4% > 77.3% so yes, based on this data, the total proportion of 18 year olds with NCEA Level 2 has increased between 2012 and 2022, by 5.1%. However, this increase has not been consistent over the years between 2012 and 2022. The 'increase between 2012 and 2022' looks only at the two end years. Actually, the peak proportion of 18 year olds with NCEA Level 2 or higher was in 2018, when ~~5~~ 85.1% of 18 year olds had L2 or higher.

(ii) Comment on the contribution of post-school qualifications to the total proportion of 18-year-olds with NCEA Level 2 or above between 2012 and 2022.

Support your answer with statistical reasoning.

In 2012, 6.1% of 18 year olds had L2 or higher from a post-school setting. This was $\frac{0.061}{0.712+0.061} =$ $0.0789 = 7.89\%$ of total L2+ qualified 18 year olds achieved after school.

In 2022, 4.1% of 18 year olds had L2 or higher from a post school setting. $\frac{0.041}{0.783+0.041} =$ $0.0498 = 4.98\%$ of total level 2 or above qualified 18 year olds achieved this post school.

So $\frac{7.89}{4.98} = 1.58$. So in 2012, post-schad qualifications contributed to the total proportion over 1.5 times as much as in 2022. Again, this decreasing proportion of post-school qualifications wasn't consistent over the 10 years, this comparison only looks at the years 2012-2022. Actually, the peak contribution post-school qualifications made to the total proportion was in 2014, when these post-school qualifications made up 8.5% of the total proportion of 18 year olds with NCEA Level 2 or above.

So post-school qualifications make up a significantly smaller proportion of the total proportion of 18 year olds with NCEA Level 2 or above in 2022 compared with 2012.

## QUESTION THREE

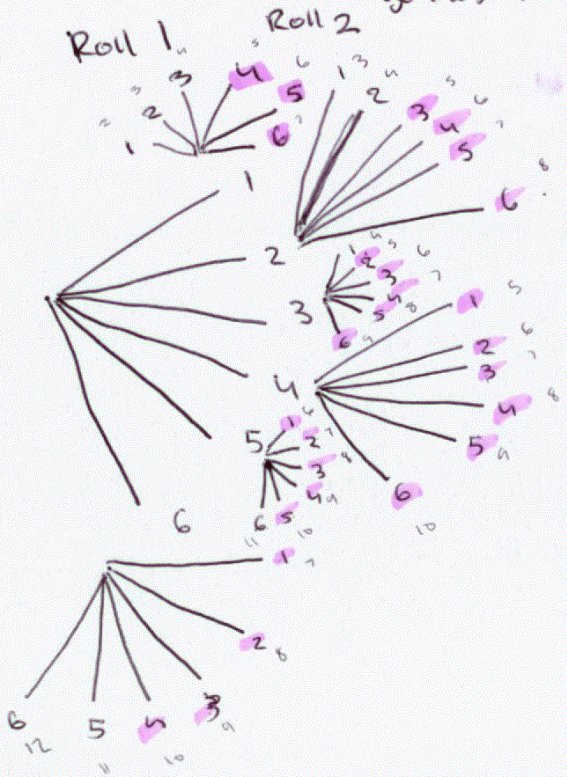
(a) A player plays two different games, A and B, by rolling a pair of dice.

(i) For game A, if the total of two dice rolls is between 5 and 10 (inclusive), then the player wins.

Calculate the probability that a player wins at least once when playing 3 times.

27/36 rolls will win. Each roll is equally likely (assuming fair dice), so ~~0.75~~ P(win) = 0.75

P(at least once in 3) = 1 - P(no times in 3). P(loss) = 0.25 $(0.25)^3 = 0.0156$ P(at least once in 3) = 1 - 0.0156 = 0.9844. This assumes that each game is independent (i.e. you're not more or less likely to lose because you lost last game)




Viable 'winning' rolls:

| | | |
|---|---|---|
| 1, 4 | 2, 3 | 3, 2 |
| 1, 5 | 2, 4 | 3, 3 |
| 1, 6 | 2, 5 | 3, 4 |
| | 2, 6 | 3, 5 |
| | | 3, 6 |
| 4, 1 | 5, 1 | 6, 1 |
| 4, 2 | 5, 2 | 6, 2 |
| 4, 3 | 5, 3 | 6, 3 |
| 4, 4 | 5, 4 | 6, 4 |
| 4, 5 | 5, 5 | |
| 4, 6 | | |

So 27/36 total outcomes are 'winning' rolls.

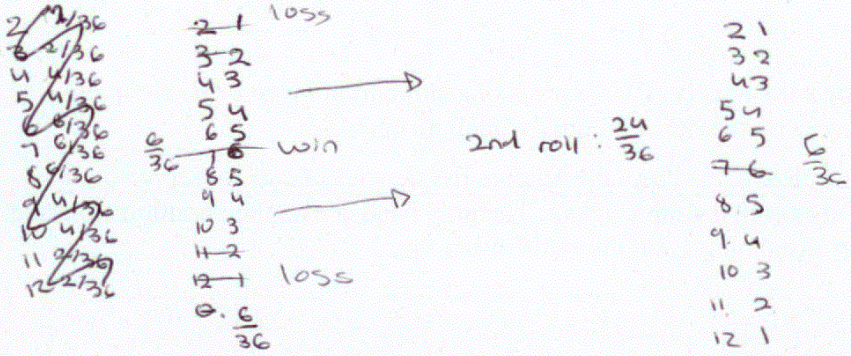
(ii) For game B, a version of the game called Hazard is played. The rules of this game are as follows:

- When a player rolls the dice for the first time, any combination of the two dice that adds up to 7 is a winner.
- On the first roll, any dice total that equals 2, 3, 11, or 12 is an immediate loser.
- If the first roll has not produced a winner or a loser, the total of the dice becomes known as the point.
- For all successive rolls, the player will win a game if the point is rolled again. However, if a 7 is rolled before the point is rolled, the player loses.

Calculate the probability that the game is lost before the player has to roll the dice for a third time.

Roll 1:

possible outcomes:




So P(loss within 2) = $\frac{6}{36} + \left(\frac{24}{36} \times \frac{6}{36}\right) = 0.278$

P(Loss within 2 rolls) = P(1/36 + 2/36 + 2/36 + 1/36) + P( P(Loss within 1 roll) = $\frac{1}{36} + \frac{2}{36} + \frac{2}{36} + \frac{1}{36} = 0.1667$

(i.e probability of a 2nd roll being necessary)

P(no win or loss) = $1 - 0.667 - \frac{6}{36} = 0.6667$

P(loss on 2nd roll) = $0.6667 \times \frac{6}{36} = 0.1111$

So p (loss within 2 rolls) = 0.278

The probability that the game is lost within the first two rolls is 0.278

*Question Three continues on the next page.*

(b) The player is concerned that one of their dice is biased. The outcomes from 1000 rolls of this die are summarised in the table below.

| Outcome | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Totals** | 138 | 189 | 197 | 143 | 179 | 154 |

(i) For these 1000 rolls, which is more likely? Rolling 3 or less OR rolling 4 or more?

Support your answer with statistical calculations.

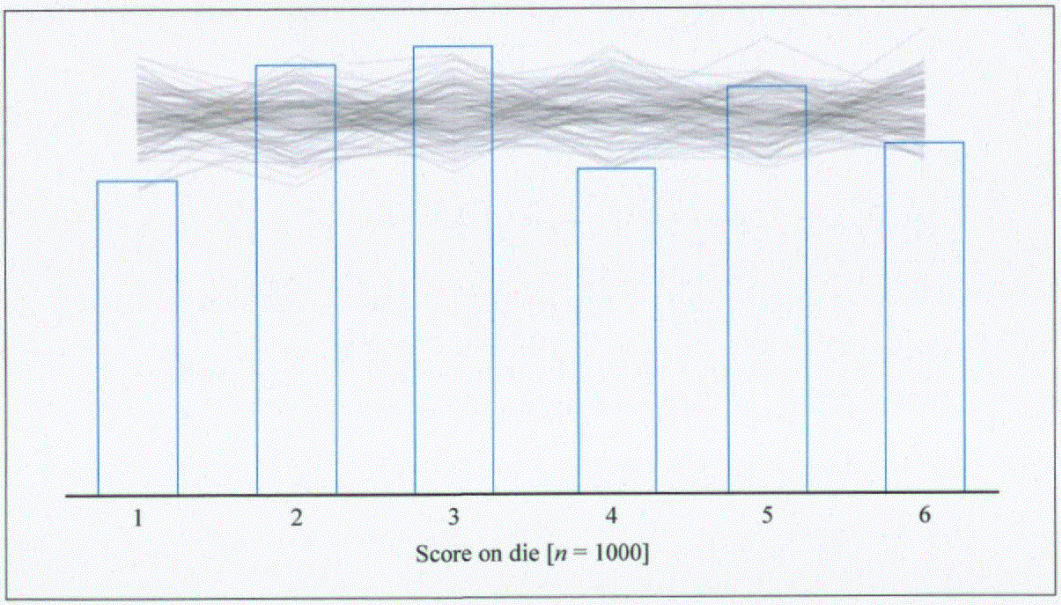
$P(X \leq 3) = \frac{138+189+197}{1000} = 0.524$

$P(X \geq 4) = \frac{143+179+154}{1000} = 0.476$

$0.524 > 0.476$ so for these 100 rolls, it is more likely to roll 3 or less than to roll 4 or more.

The diagram below shows the results of 1000 trials of a simulation model. The simulation assumed that each outcome on the die was equally likely to occur.

The height of the blue vertical bars shows the relative frequencies of each observed digit outcome on the die, as shown in the table above. The grey band shows the variation expected for each outcome, based on simulating 1000 throws of a fair die.



(ii) Should the player be concerned that one of their dice is biased?

Use the results of the simulation model shown in the diagram on the previous page and the outcomes of the 1000 rolls given in the table, and refer to experimental, theoretical, and true probability as part of your answer.

A simulation is when results are generated many times (here 1000) based on a theoretical probability. A theoretical probability is the probability based on a theoretical model (here the model is that each number is equally likely to roll, so the theoretical probability of each number rolling is $\frac{1}{6}$.) The simulation generates random dice throws based on each number having a $\frac{1}{6}$ chance of being thrown, then plots the frequency of each number (the grey lines). This is done 1000 times. This takes into account sampling variability as, because each throw dice is independant, we can't expect the observed probability to exactly match the theoretical one (observed or experimental probability is the probability found by actually carrying an experiment out. cont. on pg 14.

(iii) Explain what this result means for the chance of throwing a total of 2 using this particular die as one of the pair of dice in game B.

While not concrete evidence that this dice rolls ones less often than it should, this simulation does imply that the dice may not be fair. ~~In this case~~ If this was the case, then the probability of throwing A 2 in game B would be decreased from $1/36$, since 2 ones must be thrown to make a 2, and a 1 is less likely than $1/6$ to be thrown. Using the experimental probability instead of the theoretical probability, $P(2) = \left(\frac{138}{1000}\right)^2 = \frac{138}{1000} \times \frac{1}{6}$ (assuming the other dice is fair $P = 0.023$ $(0.138 < 0.0278)$.

Extra space if required.
Write the question number(s) if applicable.

QUESTION NUMBER

3bii) So a simulation in this case allows us to see the distribution of outcomes we would see with a fair dice, telling us how likely the outcomes (number frequencies) we got in our experimental probability are, if the dice is fair. So the grey lines represent expected variation using a fair dice (theoretical probability), and the blue columns represent the experimental (observed) probability with the dice we are testing. Since the experimental probability of rolling a 1 and a 4 is so low they are just inside the grey lines (showing that though possible with a fair dice, these results are very unlikely with ours) and the experimental probability of rolling a 3 is so high it is above the grey band (showing that the frequency of a 3 for this dice is above expected variation with a fair dice), this student should be concerned. While comparing the results of this simulation (theoretical probability) and the observed probability of the dice here doesn't give concrete evidence that this dice is not fair (after all, no frequency is far outside the limits of expected variation), it is enough that I think that the student should be concerned that the dice is not fair and should investigate further.

~~Theoretic~~ The true probability is the actual chance of rolling each number. Here, the true probability cannot be known, though we try to model it with both the theoretical probability and the experimental probability. Since the experimental probabilities are based on a large number (1000) of throws, I would expect them to be fairly close to the ~~experimental~~ true probability, though, due to sampling variability, this is not necessarily the case.

# Excellence

**Subject:** Statistics

**Standard:** 91585

**Total score:** 24

| Q | Grade score | Marker commentary |
|---|---|---|
| One | E8 | 1(a)(i) – The candidate correctly interpreted the question and used the provided information to calculate the probability of being detected as AI generated and used this probability to calculate the number of pieces of work, rounded to the nearest whole number.<br><br>1(a)(ii) – The candidate correctly calculated the conditional probability that the work is written by a human given that it is detected as potentially being generated by AI and commented that the teacher should be concerned.<br><br>1(b)(i) – The candidate gave a full and correct explanation that the events are not independent and indicated that if students are cheating, it is more likely for them to complete the assessment in under 25% of the time compared to students who are not cheating.<br><br>1(b)(ii) – The candidate correctly estimated the proportion of students in NZ who cheat and complete the assessment for this standard in under 25% of the time.<br><br>1(b)(iii) – Both reasons were identified and clearly explained in context. |
| Two | E8 | The candidate correctly processed the information given into a Venn diagram and used this to answer the three parts of question 2(a).<br><br>2(b)(i) – The candidate made a comparison of the proportions for 2012 and 2022 and concluded that the claim can be supported. They also indicated that the change was consistently positive year to year.<br><br>2(b)(ii) – The candidate correctly calculated the relative proportions for 2012 and 2022 of 18 year old students gaining their NCEA Level 2 or above from post school education and commented that it has decreased over that time period. |
| Three | E8 | The candidate correctly processed the information given, using a tree diagram in 3(a)(i) and a list of outcomes in 3(a)(ii), to calculate the required probabilities.<br><br>3(b)(ii) – The candidate provided a clear discussion of how the simulation results allow one to take into account sampling variation to make a decision to make a decision that the player should be concerned that the die may be biased. The discussion has |

| | | |
|---|---|---|
| | | included a clear explanation of experimental, theoretical and true probability in this context with reference to the sample size.<br><br>3(b)(iii) – the candidate indicated that a total of 2 could only come from rolling two 1's and if the die is biased fewer 1's are occurring than expected which means the probability of getting a total of 2 is reduced. While this candidate provided some calculations, this was not required for the excellence grade. |