Probability Concepts
Probability Distributions

Margaret Priest (margaret.priest@wgc.school.nz)
Nokuthaba Sibanda (nokuthaba.sibanda@vuw.ac.nz)

Our Year 13 students who are most likely to want to continue with Probability and Statistics at VUW study the following Level 3 standards:

- **AS 91580 Time Series**
- **AS 91581 Bivariate Data**
- **AS 91582 Statistical Methods**
- **AS 91585 Probability Concepts**
- **AS 91586 Probability Distributions**

There is no Calculus in this programme. Very few students take both Statistics and Calculus at WGC

56 students take this course (these students are our most able statistics students)
55 students take calculus
57 students take an internally assessed course comprising: Time Series, Critical Path Analysis, Statistical Methods, Linear Programming, Bivariate Data, Experimental Design, Simultaneous Equations)

Level 3 Probability Concepts AS91585

true probability v model estimates versus experimental estimates

randomness

independence

mutually exclusive events

conditional probabilities

probability distribution tables and graphs

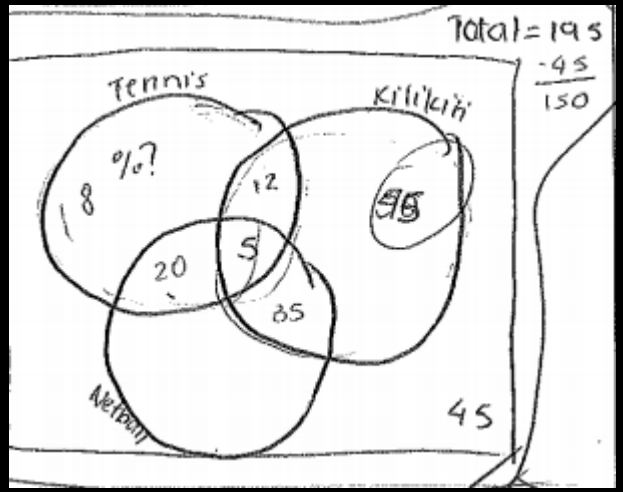
two way tables

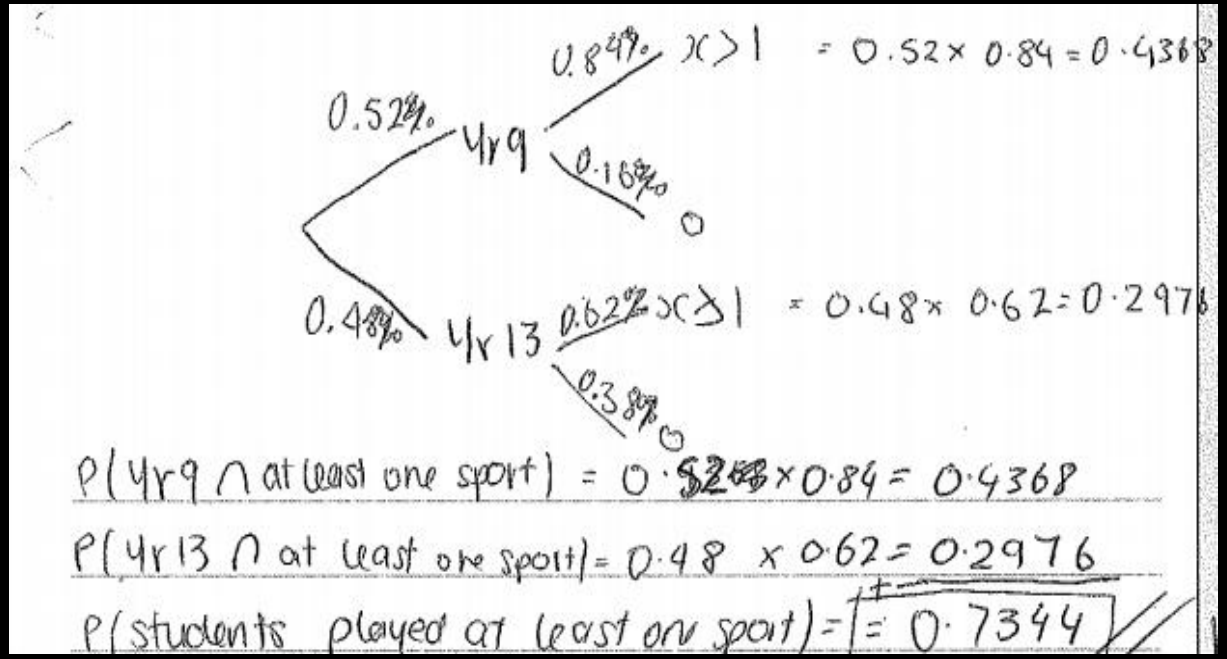
probability trees

Venn diagrams

In 2013 a student could get achieved by:

Using a Venn diagram, probability tree or table with two categories to solve problems.





For a particular sports team of 20 players:

- 14 of the players warmed up before the last game
- 5 of the players were injured during the last game
- 2 of the players did not warm up and were not injured during the last game.

Using this information, calculate the probability that a randomly chosen player from the team was injured, given that the player did not warm up before the last game.

	Injury	not	
warm up	1	13	14
Not	4	2	6
	5	15	20

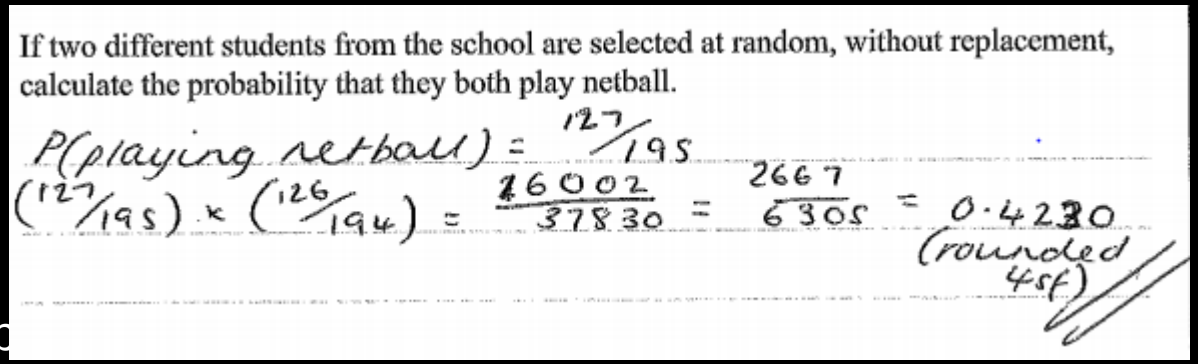
$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

P(injured given no warm up) $\frac{(5/20 \cap 6/20)}{(4/6)}$

$= 0.825$

In 2013 a student could get Excellence by:

Solving one extended abstract thinking problem in Question 1:



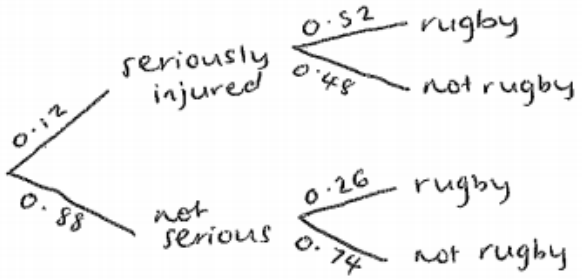
If two different students from the school are selected at random, without replacement, calculate the probability that they both play netball.

P(playing netball) = 127/195

(127/195) × (126/194) = 16002/37830 = 2667/6305 = 0.4230 (rounded 4sf)

A correct p E8

Calculate the probability of a student being injured while playing rugby during sports week.



$(0.12 \times 0.52) + (0.88 \times 0.26) = 0.0624 + 0.2288 = 0.2912$ (4sf)

P(student injured while playing rugby) $= 0.2912$ (4sf)

M

M

With regards to **probability theory**, explain why it is not possible to use this information to calculate the probability of a New Zealand adult being injured while playing tennis OR netball.

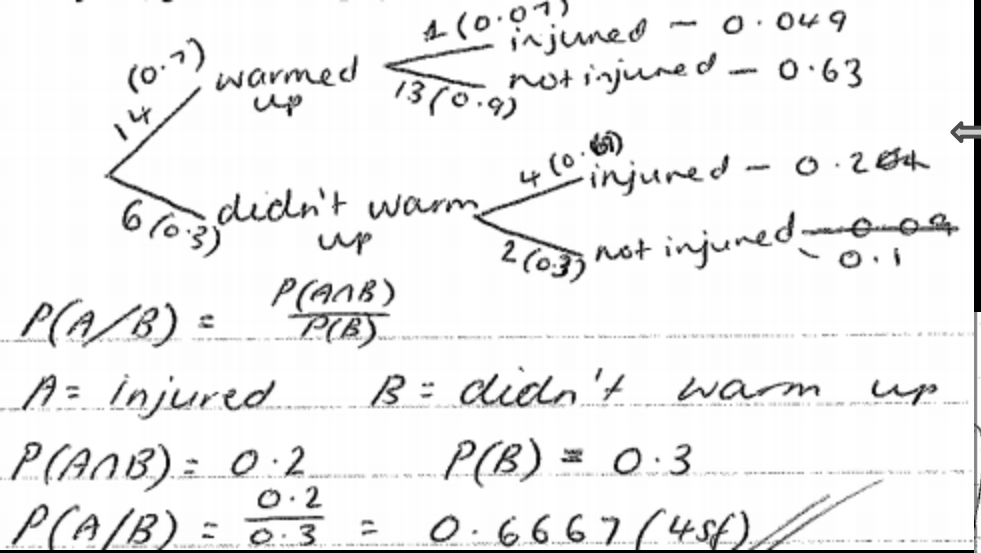
To be able to calculate the probability of a New Zealand adult being injured while played Tennis or Netball, P(A∪B) we would need to know the probability of one getting injured while playing both - P(A∩B) which this info does not give us.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ //

…..and showing four examples of relational thinking

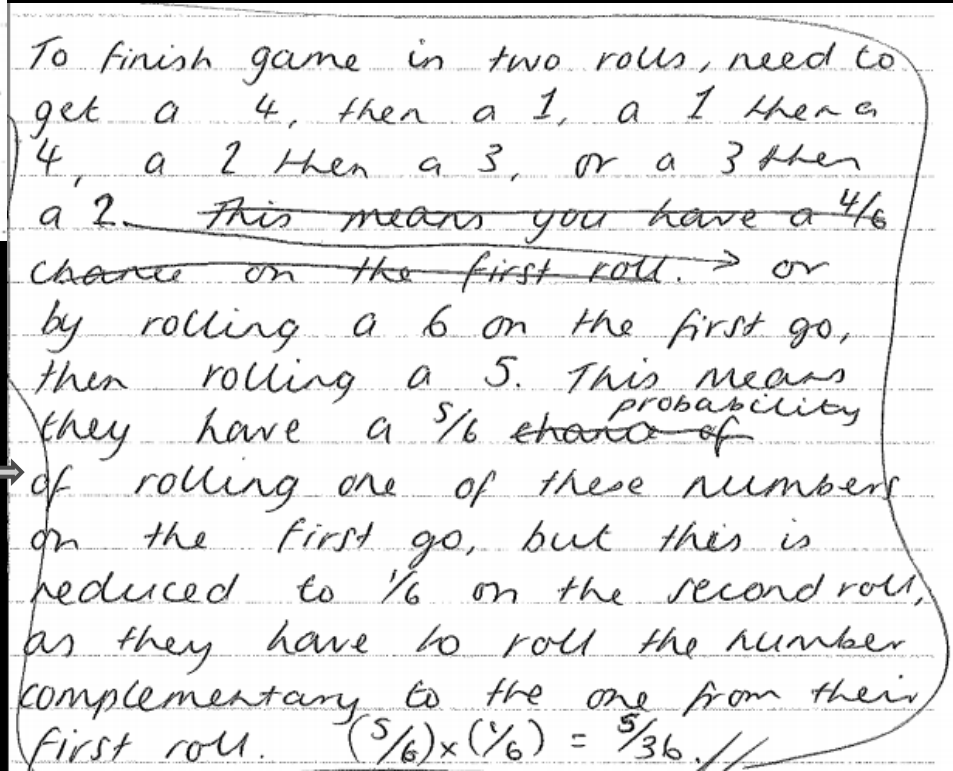
Question 2

M + M = M6



Conditional probability (M)

Justifying and explaining (M)



To finish game in two rolls, need to get a 4, then a 1, a 1 then a 4, a 2 then a 3, or a 3 then a 2. ~~This means you have a 4/6 chance on the first roll.~~ or by rolling a 6 on the first go, then rolling a 5. This means they have a 5/6 ~~chance of~~ probability of rolling one of these numbers on the first go, but this is reduced to 1/6 on the second roll, as they have to roll the number complementary to the one from their first roll. (5/6)×(1/6) = 5/36.

Question 3

M + M = M6

How do we teach Level 3 probability?

Students should carry out as many practical activities as possible to allow them to have a deep understanding of what they are doing.

By simulating and experimenting they will understand the difference between true probability, theoretical probability and experimental probability.

L3 Probability Distributions AS 91586

expected values of discrete random variables

standard deviations of discrete random variables

Poisson distribution

Binomial distribution

Normal distribution

Rectangular distribution

Triangular distribution

In 2013 a student could get an achieved grade by making three calculations

Binomial

A manufacturer of sports uniforms makes shirts for sports teams. The manufacturer knows that the machine he currently uses to make the shirts will produce defective shirts 4% of the time.

The manufacturer has received an order for 20 shirts for a sports team.

Using an appropriate distribution to model this situation, calculate the probability there will be at least two defective shirts in the order.

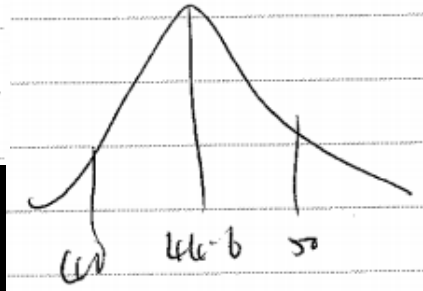
In your answer, you should justify your choice of distribution, identify the parameter(s) of this distribution, and state any assumption(s) you make.

$p = .04$

B

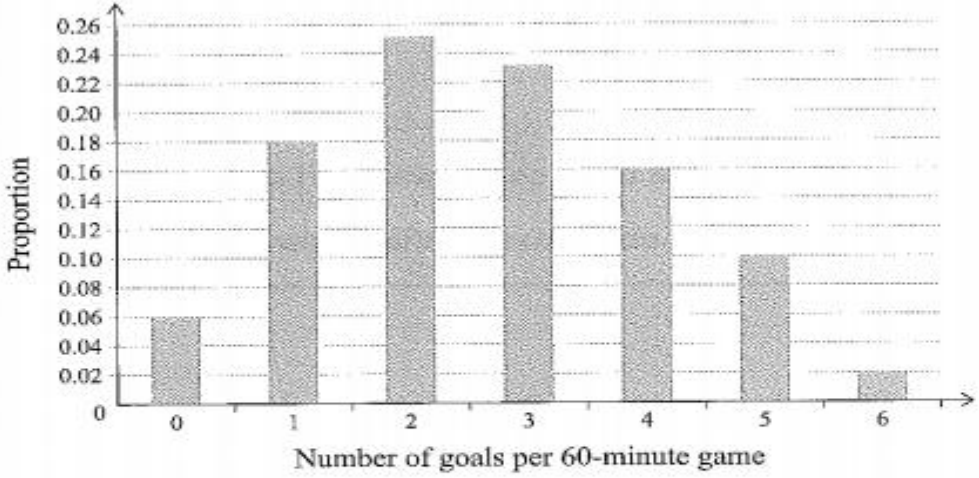
$P_{rob} = .19$

Normal



$1 - 0.50598 = 0.49404$

The observations of the number of goals scored by a team during a 60-minute ice hockey game, over a large number of games, resulted in the following graph.



(i) Describe the key features of the distribution, and obtain an estimate for the mean number of goals scored per 60-minute game (rounded to one decimal place).

2·3

In 2013 a student could get Excellence by:

Normal dist^n is appropriate because data in histogram is

- continuous (time of matches)
- fairly symmetrical
- approx bell shaped.

For data using given mean & sd

$P(30 < x < 45) = 0.62334$

$P(50 < x < 60) = 0.12125$

histogram

$P(30 < x < 45) = 0.62$

$P(50 < x < 60) = 0.09$

Both sets of values seem quite close and we tested 2 different parts of histogram data so normal dist^n seems to be appropriate

compares theoretical & experimental distributions

justifies distribution by making two comparative calculations

E8

(ii) Using an appropriate distribution, find the probability that the team will score at least three goals in at least two of the next five different 60-minute games.

In your answer, identify the distribution and its parameter(s), and state any assumption(s) you make.

Poisson $\lambda = 0.8$ $P(X \geq 3) = 1 - P(X \leq 2)$

$= 0.0474$ (4SF)

Binomial $p = 0.0474$ $n = 5$

$P(X \geq 2) = 1 - P(X \leq 1)$

$= 1 - 0.9796$

$= 0.0204$ (4SF)

Used binomial because

- there are 2 outcomes scored ≥ 3 goals, or not
- fixed number of trials $= 5$
- A fixed probability of ~~succe~~ "at least 3 goals in a game" – assumed didn't change.

It is an assumption as this probability depends on the opposition – chance of ≥ 3 goals is higher against bottom team in competition & less against top team

Question grade

Correct calculation and assumption

E8

The manufacturer of a brand of squash rackets has received a complaint about a particular racket it produced.

According to the complainant, the racket logo is more likely to be facing up than down after the racket is spun, suggesting that the racket is unbalanced.

The manufacturer states that it is equally likely for the logo to be facing up or down after the racket is spun.

The complainant has recorded data over 20 spins, which is shown in the table below:

	Racket logo	
	Facing up	Facing down
20 spins	13	7

Apply an appropriate probability distribution to investigate whether the complaint is justified.

You should support your answer with statistical reasoning and calculations.

identified Binomial
correct calculation
correct interpretation in relation to the problem

E7

$P(X \geq 13) = 1 - P(X \leq 12)$

$= 0.1316$

$P(X = 13) = 0.0739$

The probability for $X \geq 13$ is not that small (13%)
so 13 (or more) logo up comes quite often (13%)
we expect variation in the number of heads after spinning squash racquet and 13 seems to fit well into this possibility.
so the complaint is not really justified.

How do we teach probability distributions?

With lots of experiments. Almost always the questions are given in a practical situation

For Poisson we had scoops of hokey pokey ice-cream and counted the lumps of hokey pokey

For Binomial we used different types of dice

For expected value and variance we used spinners

http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S8-4

What are the differences between Year 13 and first year at VUW?

Year 13	VUW
Differentiated learning http://www.youtube.com/watch?v=01798frimeQ	Lectures: mix of note-taking, worked examples, powerpoint slides; Tutorials: Work through set examples
Practical activities	Introduction of R demos in MATH 177
Write-on work books	-
Class sizes ~ 25	Lecture numbers ~ 300, tutorials optional
Students form strong relationships with their teachers/have fun	Not possible in first year due to large classes
Very few students take Statistics and Calculus	Math 177 (Probability and Decision Modelling) entry requirement is Differentiation and Integration or Math 141 (Introduction to Calculus) - the entry requirement would stop most of our students taking this course although all of our highest achieving statistics students take Level 2 Differentiation and Integration
Use calculators, iNZight, generally no statistical tables	Use calculators, statistical tables and R (for demonstration)

Probability Concepts and Distributions @ Vic

MATH 177: Probability and Decision Modelling

Entry Requirements: 16 credits NCEA level 3 Mathematics or Statistics, including AS 3.6 (differentiation, AS91578) and 3.7 (integration, AS91579), or MATH 141 (Introduction to Calculus)

STAT 193: Statistics for the Natural and Social Sciences

Entry Requirements: Assumes no previous knowledge of Statistics, but Mathematics to Year 12 preferred

MATH 177

Course content – covered in 30 lectures

- properties of probability
- conditional probabilities, Bayes' rule
- discrete and continuous probability distributions (eg Binomial, Poisson, Chi-square, Normal, Beta, Exponential)
- Joint distributions, functions of random variables
- estimation and goodness-of-fit check
- value of information in decision making, decision trees
- modelling of queueing systems

MATH 177

Assessment

- 10 Assignments (10%)
- Test (15%) – 50 minutes
- Exam (75%) – 3 hours. Six questions - expected to attempt all.

Students have to get an overall mark of at least a C- (50%) to pass the course

MATH 177

Course Resources

- Set of course notes - cost: $12.50
- Recommended text
- Statistical tables (provided in exam)
- Calculator – scientific calculator (graphics calculator not required)
- R software package used to demonstrate ideas of joint distributions, sampling and the central limit theorem

MATH 177

Assumed knowledge:

- Differentiation
- Integration
- Basic algebra

Useful to have

- Knowledge of random variables and distributions
- Probability rules and associated tools

Additional support

- Tutorials – enrolment required, attendance optional
- Student Learning Support Service – one-on-one appointments
- MATH Helpdesk – for calculus-related problems
- Peer support via Facebook – not monitored
- Staff office hours
- Mentoring available through Āwhina or Te Pūtahi Atawhai

Calculus in MATH 177

Example in course notes

Example 4.16 The duration, X, in minutes of phone calls from company business phones is a continuous r.v. with pdf

$$f_X(x) = \begin{cases} \frac{1}{6}e^{-x/6} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(a) Derive the cdf, and show that this function is, in fact, a satisfactory cdf.
(b) Calculate the probability that a call will last between 3 and 6 minutes.
(c) Find the expected length of a call.
(d) Determine the variance of call length.

Solution

(a) $F(x) = P(X \leq x) = \int_0^x \frac{1}{6}e^{-u/6}\,du$

$$= \int_0^x -e^{-u/6}\,d(-u/6) \qquad \text{or, let } v = -u/6 \text{ so that } F(x) = \int_0^{-x/6} -e^v\,dv$$

$$= \left[-e^{-u/6}\right|_0^x$$

$$= -e^{-x/6} - (-1) = 1 - e^{-x/6} \qquad x \geq 0$$

We can confirm that this is a satisfactory cdf:

Calculus in MATH 177

Exam question

5. (a) The weekly repair cost, X (measured in hundreds of dollars), for a certain machine has a probability density function given by

$$f(x) = \begin{cases} A\,x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

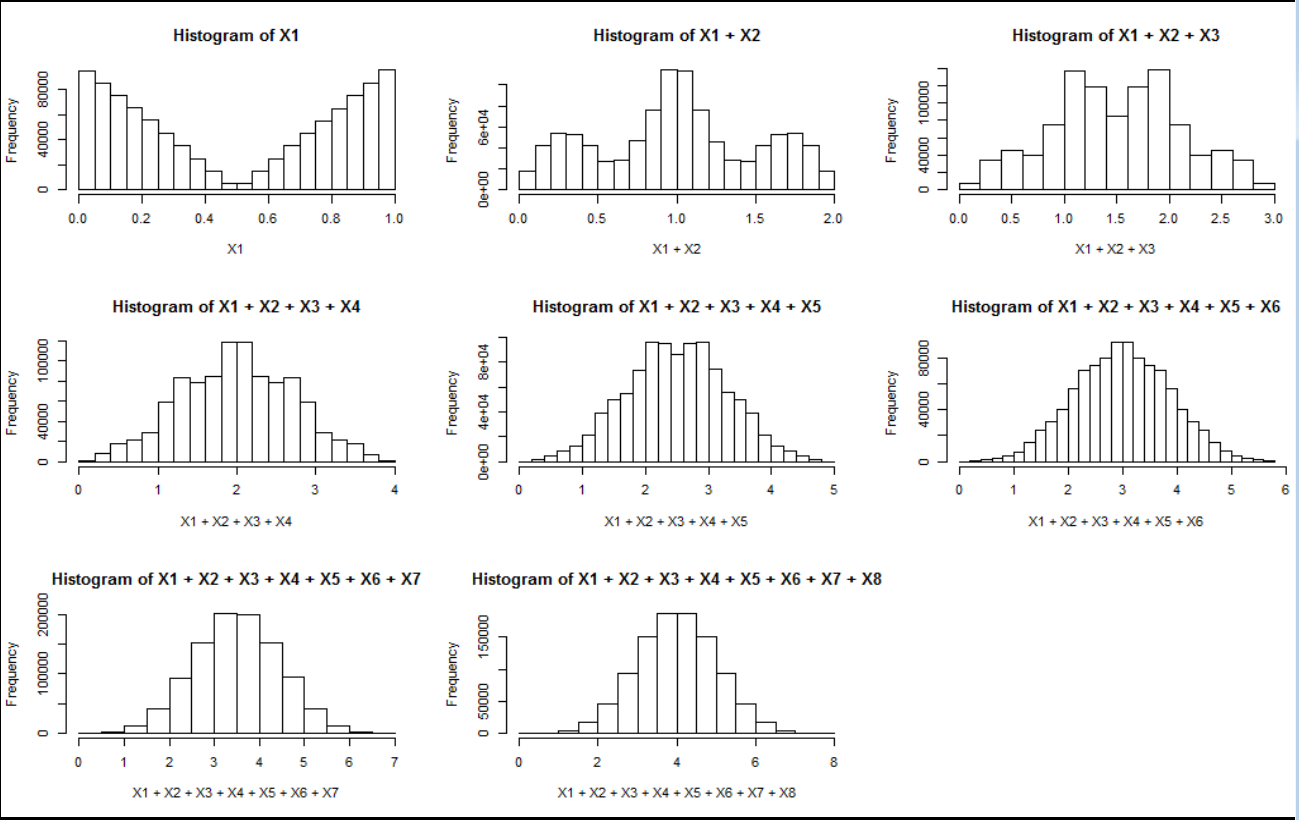
i. Show that $A = 6$.

ii. Sketch the graph of $f(x)$.

iii. What is the probability that repair costs will exceed \$75 during a week?

iv. Given that mean weekly repair cost is \$50, determine the variance of X.

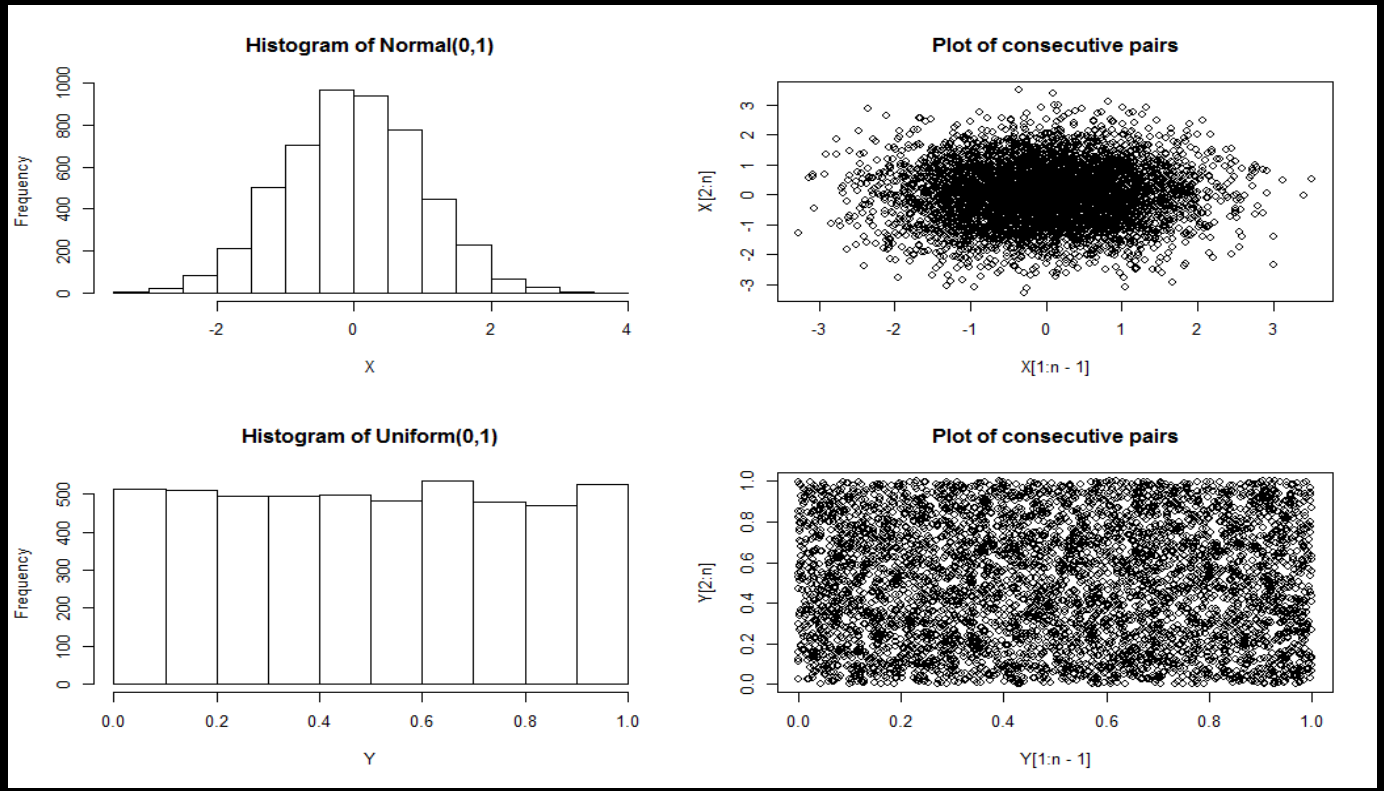
MATH177 R Demos



File Edit View Misc Packages Windows Help

```
> # Central Limit Theorem
> # V (triangle) distribution
>
> rtriangle <- function(N) {
+   r = runif(N,0,1)
+   i = (r<0.5)
+   x = r
+   x[i] = 0.5 - sqrt(1-2*r[i])/2
+   x[!i] = 0.5+sqrt(2*r[!i]-1)/2
+   return(x)
+ }
>
> n = 1000000
> X1 = rtriangle(n)
> X2 = rtriangle(n)
> X3 = rtriangle(n)
> X4 = rtriangle(n)
> X5 = rtriangle(n)
> X6 = rtriangle(n)
> X7 = rtriangle(n)
> X8 = rtriangle(n)
>
> par(mfrow=c(3,3))
> hist(X1)
> hist(X1+X2)
> hist(X1+X2+X3)
> hist(X1+X2+X3+X4)
> hist(X1+X2+X3+X4+X5)
> hist(X1+X2+X3+X4+X5+X6)
> hist(X1+X2+X3+X4+X5+X6+X7)
> hist(X1+X2+X3+X4+X5+X6+X7+X8)
>
```

MATH177 R Demos



Probability concepts and Distributions in STAT 193

- Probability trees, concepts of independence, mutually exclusive events
- Calculation of probabilities, mean and variance for Binomial distribution
- Calculation of probabilities for Normal random variables
- Use of Binomial, Normal, t, chi-square and F distributions in hypothesis testing
- Central limit theorem applied to means and proportions

STAT 193

Assessment

- 5 weekly assignments - mandatory, but no contribution to final grade
- 2 project assignments - 15%
- Test (10%) – 50 minutes
- Exam (75%) – 3 hours. Eight questions - expected to attempt six.

Students have to get an overall mark of at least a C- (50%) to pass the course

https://teachingcommons.stanford.edu/teaching-services/new-junior-faculty-assistance/smart-teaching

STAT 193

Course Resources

- Set of course notes - free to download at end of each week
- Recommended text
- Lectures video-recorded and available to students
- Calculator – graphics calculator required
- Statistical tables (provided in exam)

Additional support

- Tutorials – enrolment required, attendance optional
- Student Learning Support Service – one-on-one appointments
- Drop-in Help Sessions
- Staff office hours
- Mentoring available through Āwhina or Te Pūtahi Atawhai