A1. A bowl contains 100 Scrabble tiles: 36 A's, 30 B's, 24 C's, and 10 D's.

(a) In the first experiment, we choose a tile at random, make a note of the letter on it, and return it to the bowl. We repeat this process two more times. What is the probability that the tiles we have drawn all show the same letter of the alphabet?

Solution: We will count outcomes *taking order into account*. There are 100 different ways to choose the first tile, then 100 ways to choose the second, and 100 ways to choose the third. The total number of outcomes is $100 \times 100 \times 100 = 10^6$.

The number of ways we can draw A, A, and then A is $36 \times 36 \times 36$. Similarly, the number of ways to get B, B, B is $30 \times 30 \times 30$. There are 24^3 ways to draw C, C, C, and 10^3 ways to draw D, D, D.

The probability of getting three tiles with the same letter of the alphabet is

$$\frac{36^3 + 30^3 + 24^3 + 10^3}{100^3} = \frac{553}{6250} = 0.08848.$$

(b) In the second experiment, we randomly draw three tiles from the bowl without replacement. What is the probability that the three tiles we have drawn all show the same letter of the alphabet?

Solution: This time we will count combinations – that is, we will not take order into account. The number of possible outcomes is simply

$$C(100, 3) = 161,700.$$

The size of the event is

$$C(36, 3) + C(30, 3) + C(24, 3) + C(10, 3) = 13,344.$$

The probability of this event is

$$\frac{13344}{161700} \approx 0.082523.$$

(c) Which of the answers in (A1a) and (A1b) is larger? Can you think of a simple explanation why this should be so?

Solution: When we replace the tiles after drawing them, we have a better chance of picking the same letter of the alphabet three times. Why? If we do not replace a tile after drawing it, then we've decreased the available number of tiles with that letter, so on our subsequent draws, we will be less likely to draw the same letter.

(d) In the third experiment, we randomly draw three tiles and lay them out on the table in the order in which they were drawn. What is the probability that the three tiles are in alphabetical order?

Solution: First, the number of possible outcomes here is $P(100, 3) = 100 \times 99 \times 98 = 970,200$.

We'll assume that for three tiles to be in alphabetical order, they all have to show different letters. There are four different patterns of letters that make up this event: ABC, ABD, ACD, and BCD. We count the number of outcomes in the event as follows:

Pattern	Number of outcomes		
ABC	$36 \times 30 \times 24$	$=$	$25,920$
ABD	$36 \times 30 \times 10$	$=$	$10,800$
ACD	$36 \times 24 \times 10$	$=$	$8,640$
BCD	$30 \times 24 \times 10$	$=$	$7,200$
	Total:		$52,560$

The probability of the event is

$$\frac{52,560}{970,200} \approx 0.054174.$$

A2. Repeat parts (A1a) through (A1d) of problem A1, assuming now that the bowl contains 25 A's, 25 B's, 25 C's, and 25 D's.

Solution:

(a) In this case, the probability of AAA or BBB or CCC or DDD is

$$\frac{25^3 + 25^3 + 25^3 + 25^3}{100^3} = 0.0625.$$

(b) The probability of getting the same letter 3 times without replacement is

$$\frac{4 \times C(25,3)}{C(100,3)} \approx 0.056895.$$

(c) When we draw letters without replacement, the probability of getting the same letter three times is significantly lower. This stands to reason – when we don't replace the tile we just drew, we have fewer chances to match it.

(d) The number of outcomes in each of the four sub-events ABC, ABD, ACD, and BCD is the equal to 25^3, so the probability of drawing three letters in alphabetical order is

$$\frac{4 \times 25^3}{100 \times 99 \times 98} \approx 0.064420.$$

A3. A five-card poker hand counts as *two pair* if two of the cards have equal ranks, two of the remaining cards have equal ranks (but not equal to the rank of the first two) and the fifth card has a rank different from the ranks of the other four. Got that?

How many different two-pair poker hands are there? What is the probability of being dealt two pair?

Solution: To count the number of two-pair hands, we first choose the two ranks for the pairs (a factor of $C(13,2)$), then we choose the two suits for the first pair (a factor of $C(4,2)$), the two suits for the second pair (a factor of $C(4,2)$), and finally we choose the fifth card (a factor of $C(44,1)$). The total number of two-pair poker hands is

$$C(13,2) \times C(4,2) \times C(4,2) \times C(44,1) = 123,552.$$

There are $C(52,5) = 2,598,960$ equally-likely poker hands, so the probability of being dealt two pair is

$$\frac{123,552}{2,598,960} \approx 0.047539.$$

A4. A five-card poker hand counts as *three of a kind* if three of the cards have the same rank and the remaining cards have ranks that are different from each other and different from the first.

How many different three-of-a-kind poker hands are there? What is the probability of being dealt three of a kind?

Solution: To count the number of three-of-a-kind hands, we choose a rank for the triple (a factor of $C(13,1)$) and then three suits for the triple (a factor of $C(4,3)$). The remaining two cards have to have different ranks (a factor of $C(12,2)$), but can have any suits (a factor of 4^2). The total number of three-of-a-kind hands is

$$C(13,1) \times C(4,3) \times C(12,2) \times 4^2 \quad = \quad 54,912.$$

The probability of being dealt three of a kind is

$$\frac{54,912}{2,598,960} \quad \approx \quad 0.021129.$$

Choose one of the cryptograms on the handout **Cryptograms #8 – Vigenère ciphers**.

B1. Calculate the index of coincidence for your cryptogram. You will probably want to use an Excel spreadsheet to do this. Print out your spreadsheet and hand it in. If you entered formulas in any of the Excel cells, write them on the printout.

Also, write a paragraph or so explaining how you carried out the calculation.

I chose the first cryptogram. I used the frequency counting program on the website to count the letter frequencies, which I then pasted into the first two columns of an Excel spreadsheet. I entered formulas into the third and fourth columns to compute $f_i - 1$ and $f_i(f_i - 1)$ for each letter, then used the Excel `sum` function to compute the index of coincidence, which is about 0.043367.

Here is the table from the Excel sheet:

A	18	17	306	sum(p_i(p_i-1))=	23460
B	33	32	1056	n(n-1)=	540960
C	34	33	1122	I=	0.043367347
D	30	29	870	k is about	5.411764706
E	50	49	2450		
F	43	42	1806		
G	14	13	182		
H	12	11	132		
I	32	31	992		
J	19	18	342		
K	39	38	1482		
L	18	17	306		
M	22	21	462		
N	30	29	870		
O	49	48	2352		
P	15	14	210		
Q	8	7	56		
R	33	32	1056		
S	39	38	1482		
T	35	34	1190		
U	31	30	930		
V	35	34	1190		
W	12	11	132		
X	33	32	1056		
Y	34	33	1122		
Z	18	17	306		
sum=	736	735	540960		

I used these formulas:

- At the top of the third column, `=B1-1`, copied down the column.
- At the bottom of the second column, `=sum(B1:B26)`.
- At the top of the third column, `=B1*C1`, copied down the column.
- In cell F1, `=sum(D1:D26)`.
- In cell F2, `=B27*C27`.
- In cell F3, `=F1/F2`.

- In cell F4, `=0.0265*B27/((0.065-F3)+B27*(F3-0.0385))`.

B2. Use the index of coincidence to estimate the keyword length (using the Friedman test).

From the index of coincidence and the formula on p. 138 of Barr, I came up with the value $k \approx 5.4$ for the length of the keyword.

B3. Now apply the Kasiski test to your cryptogram. Identify all repeated strings of four characters or more and find out how many characters separate the repetitions. Make a list of possible keyword lengths. (If you don't find enough repeated strings of four characters, try throwing in a few three-character strings.)

Solution: Here are the repeated strings and the distances at which they occur:

SOTYO	at	8
AISO	at	364
ENFR	at	300
HXRE	at	384
MUOI	at	20
ODVB	at	308
ZIEQ	at	68
XETY	at	217
KKMKVBT	at	132
VBTA	at	44

From the SOTYO string, we guess that the keyword length is 1, 2, 4, or 8. Except for XETY at 217, all the remaining string-separation numbers are divisible by 4, but only one of them is divisible by 8. We conclude that the keyword length must be 1, 2, or 4. From the result of the Friedman test, it seems most likely that they keyword length is 4.

B4. Guess the length of your keyword, and use the techniques from the last problem set to decrypt your cryptogram.

The first coset suggests that the first letter of the keyword is B.

The second coset suggests that the second letter of the keyword is A.

The third coset suggests that the third letter of the keyword is R.

The fourth coset suggests that the fourth letter of the keyword is K.

Using the keyword BARK, we decrypt the following message:

"Windtalkers" tells a luke-warm story about some unknown heroes of World War II: Navajo radio operators whose encoded transmissions, based on their ancient language, helped win the war in the Pacific. It features a handful of convincing battle scenes under the sure hand of director John Woo and a lead actor, Nicolas Cage, who does his best to wrest a full-bodied performance from an undernourished script. Cage may go down with the script, but he goes down trying. He grunts, sweats, and bleeds. He talks slowly, as if trying to impart some underlying anguish. It's a strenuous effort and the audience feels the strain. He plays Enders, a marine assigned to protect the Navajo code, even if that means killing codetalkers about to fall into enemy hands. Unfortunately, Enders as written is a cipher, so his moral dilemma lacks emotional force or meaning. We come away from "Windtalkers" thinking, "Fine, but who cares?"