Probability Tutorial Additional Answers 2021

Note that the final answers are in bold.

Answer to Question 1:

a) How many different arrangements can be formed from the letters of the work CHEMISTRY?
The number of permutations of a list of n items is n!
The number of different arrangements is 9!
=362,880

b) How many different ways can the letters be arranged if:

i. The first letter is a vowel and the last letter is Y

E or I _ _ _ _ _ _ _ Y

2! * 7! * 1!
=2*5040*1
=10,080

ii. The letters C and H must be together
C, H or H, C _ _ _ _ _ _ _ [CH box plus 7 other letters =8 boxes]
It's like we are working with only 8 letters when C and H are together
When C and H are together it could be CH or HC so there are 2 permutations.
=2! * 8!
=2 * 40,320
=80,640

iii. The letters C and H must not be together
=the total number of permutations minus the number of permutations where C and H are together
=362,880 - 80,640
=282,240

c) In how many ways can 6 numbers be chosen from 42?
N, the total number of objects to choose from is 42
R, the total number of objects that are being chosen is 6

Combinations = nCr
42 C 6 = **5,245,786**

Alternatively
nCr= n! / [(r!)*(n-r)!]
42C6 = $\frac{42!}{6!*(42-6)!}$
=5,245,786

d) In a 20 team league, how many matches can be organised if:

i. Each team plays each other exactly once

20 C 2 = **190** (using the function on the calculator nCr)

ii. Each team plays each other home and away

Each team plays twice

(20 C 2) * 2 = 190 * 2 = **380**

Answers to Question 2:

A bag contains 3 white beads, 4 blue beads and 3 black beads.

Two beads are chosen at random and not replaced.

a) Find the probability that the beads chosen are:

i. Both black

There are 10 beads in total. The beads aren't replaced so these are not independent events.

P($black_1$ and $black_2$)= P($black_1$) * P($black_2$|$black_1$)

= (3/10) * (2/9)

= 1/15

ii. Black and blue in that order

P(black and blue) = P(black) * P(blue)

= (3/10) * (4/9)

= **2/15**

iii. Black and blue in any order

P(black and blue) or P(blue and black)

=2/15 (see part ii) + (4/10) * (3/9)

=2/15 + 2/15

=4/15

b) Find the probability that at least one of the beads is white.

P(both white) or P(one white, other non-white)+ P(non-white, white)

= (3/10)*(2/9) + (3/10)*(7/9) + (7/10)*(3/9)

=8/15

Answer to Question 3:

The random variable X has a discrete distribution. The probability that it takes a value other than 13, 14, 15 or 16 is negligible.

a) Complete the probability distribution table below and hence calculate E(X), the expected value of X.

x	13	14	15	16
P(X=x)	0.383	0.575	**0.038**	0.004

The probability that x=13 or 14 or 15 or 16 is approximately 1 given that the probability of X taking over values is negligible.
P(X=13) + P(X=14) +P(X=14) +P(X=15)+ P(X=16)=1
0.383 +0.575 +P(X=15) +0.004=1
P(X=15) = 1 -0.383 -0.575 -0.004
P(X=15) = 0.038

E(X) = ∑x*P(X=x)
E(X) = 13* P(X=13) +14* P(X=14) +15* P(X=15) +16* P(X=16)
E(X) = 13*0.383 + 14*0.575 + 15*0.038 + 16*0.004
E(X) = 13.663

b) If X is the age, in complete years, on 1 January 2013 of a student selected at random from among all second year students in Irish schools, explain what E(X) represents.
E(X) is the mean age of all second year students.

c) If ten students are selected at random from this population, find the probability that exactly six of them were 14 years old on 1 January 2013. Give your answer correct to three significant figures.
= P(X=14)^6 * P(X is not 14)^4
= (10 C 6) * (0.575)^6 * (1-0.575)^4   Where (10 C 6) is a combination
=210* 0.001179132
=0.248

Answer to Question 4:

Question 1 **(25 marks)**

(a) Explain each of the following terms:

(i) Sample space

The set of all possible outcomes of an experiment.

(ii) Mutually exclusive events

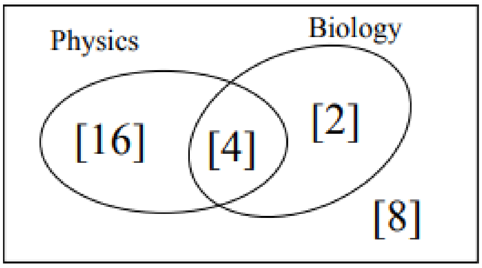
Events E and F are mutually exclusive if they have no outcomes in common.
i.e $P(E \cup F) = P(E) + P(F)$

(iii) Independent events

Two events are independent if the outcome of one does not depend on the outcome of the other.
i.e $P(E \cap F) = P(E).P(F)$ or $P(E|F) = P(E)$ or $P(F|E) = P(F)$

(b) In a class of 30 students, 20 study Physics, 6 study Biology and 4 study both Physics and Biology.

(i) Represent the information on the Venn Diagram.



A student is selected at random from this class.
The events E and F are:

E: The student studies Physics
F: The student studies Biology.

(ii) By calculating probabilities, investigate if the events E and F are independent.

$P(E \cap F) = \frac{4}{30}$

$P(E) \times P(F) = \frac{20}{30} \times \frac{6}{30} = \frac{4}{30}$

$P(E \cap F) = P(E) \times P(F) \Rightarrow$ E and F are independent events

Answer to Question 5:

Q1	Model Solution – 25 Marks	Marking Notes
(a)	$\frac{4}{5}\times\frac{1}{5}\times\frac{4}{5}\times\frac{1}{5}\times\frac{4}{5}\times\frac{1}{5}\times\frac{4}{5}=\frac{256}{78125}$ **or** $=0{\cdot}0032768$	**Scale 10C (0, 4, 5, 10)** *Low Partial Credit:* • $\frac{4}{5}$ • $(\frac{1}{5})^3$ *High Partial Credit:* • $\frac{4}{5}\times\frac{1}{5}\times\frac{4}{5}\times\frac{1}{5}\times\frac{4}{5}\times\frac{1}{5}\times\frac{4}{5}$ in any order
(b)	$\binom{6}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)$ $=\frac{1280}{78125}$ **or** $\frac{256}{15625}$ **or** $0{\cdot}016384$	**Scale 5D (0, 2, 3, 4, 5)** *Low Partial Credit:* • $\binom{6}{3}$ or $\left(\frac{1}{5}\right)^3$ or $\left(\frac{4}{5}\right)^3$ • $\frac{1}{5}$ for last day *Mid Partial Credit:* • $\binom{6}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^3$ and stops or continues • $\binom{7}{4}\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^3$ and continues *High Partial Credit:* • $\binom{6}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)$
(c)	$1-\left(\frac{4}{5}\right)^n$	**Scale 5B (0, 3, 5)** *Partial Credit:* • 1 or $\left(\frac{4}{5}\right)^n$ • any correct term from the expansion
(d)	$1-\left(\frac{4}{5}\right)^n>0{\cdot}99$ $\left(\frac{4}{5}\right)^n<0{\cdot}01$ $\left(\frac{4}{5}\right)^{20{\cdot}6377}\approx 0{\cdot}01000000517$ $n=21$	**Scale 5C (0, 2, 4, 5)** *Low Partial Credit:* • Ans **(c)** $>0{\cdot}99$ *High Partial Credit:* • viable solution to inequality • n = 20·6377 and stops

Q6

(a)

	A	B	C	D
The data are skewed to the left	×	×	√	×
The data are skewed to the right	√	×	×	×
The mean is equal to the median	×	√	×	√
The mean is greater than the median	√	×	×	×
There is a single mode	√	√	√	×

Note on the mean v the median:

A is skewed right. Taking pairs of values about the median, working from the most extreme values inwards, the values to the right have greater leverage about the median so the mean is greater than the median. Another way to think about this would be to cut out the histogram on cardboard and consider at what point it would balance on your finger: this is the mean and would be to the right of the median.

(b) D has the largest standard deviation.

The standard deviation is the square root of the variance. The variance is the average value of $(x - \mu)^2$ where x is the variable and μ is its mean. D is symmetrical so its mean is in the centre. Most of the values in D are spread about this, the largest frequencies being at the extremes. The spread is clearly greater in D than in the other histograms so its variance and standard deviation are therefore largest.

Q7

(a) $P(Z \le 2.43) = 0.9925$ from tables

(b) $P(Z \ge 1.96) = 1 - P(Z \le 1.96)$

$= 1 – 0.975$ from tables

$= 0.025$

(c) $P(Z \ge -0.5) = P(Z \le 0.5)$

$= 0.6915$ from tables

(d) $P(Z \le -1.2) = P(Z \ge 1.2)$

$= 1 - P(Z \le 1.2)$

$= 1 – 0.8849$ from tables

$= 0.1151$

Q8

(a) Retired people, full-time students

(b) The **median** is the middle value when the data are placed in order.
Total persons at work in ascending order:

1773.4 1784.8 1803.4 1813.4 1828.6 1867.0 1903.3 1954.9 2049.6 2060,4

Since we have an even number we take half the sum of the two middle values:

Median = ½ (1828.6 + 1867.0) = 1847.8

The **interquartile range** is the difference between
the point which cuts off the first quarter of the data and
the point which cuts off the last quarter of the data:

Interquartile range = 1954.9 – 1803.4 = 151.5

(c) (i) Percentage of persons aged 15 and over in 2011

at work: $100 \times \frac{1813{,}4}{3599.1} = 50.4\%$

unemployed: $100 \times \frac{364.1}{3599.1} = 10.1\%$

not in the labourforce $100 \times \frac{1421.6}{3599.1} = 39.5\%$

We can check that these sum to 100%.

(d) Percentage of the total population in 2011

at work: $100 \times \frac{1813{,}4}{3599.1+979.6} = 39.6\%$

unemployed: $100 \times \frac{364.1}{3599.1+979.6} = 8.0\%$

not in the labour force $100 \times \frac{1421.6+979.6}{3599.1+979.6} = 52.4\%$

We can check that these sum to 100%.

Do you agree with the statement: Yes.

Only persons at work pay income tax. The number of persons at work fell from 1954.9 thousand in 2006 to 1813.4 thousand in 2011, so the Government's revenue from this source fell (before any changes in tax rates, etc.).

The Government has to pay unemployment benefit. The number of persons unemployed more than doubled from 118.7 thousand in 2006 to 260.7 thousand in 2011.

Q9

10 years ago the proportion of students who lived within 2.5 km of the school was

$$p = .26$$

If this is still the true proportion, for samples of students of size n = 60, the mean and standard deviation of the sample proportions are

p and $\sqrt{\left(\frac{p(1-p)}{n}\right)}$ respectively (see Formulae page 34)

mean = .26 standard deviation = $\sqrt{\left(\frac{.26(1-.26)}{60}\right)}$ = .0566

For n = 60 the distribution of the sample proportions is approximately normal.

(a) Null hypothesis (H_0): The proportion of students living within 2.5 km of the school is still .26.
Alternative hypothesis (H_1): The proportion of students living within 2.5 km of the school has changed.

Assuming H_0, corresponding to the observed sample proportion .20, the standard normal variable is

$$z = \frac{.20-.26}{.0566} = -1.06$$

The standard normal distribution has a standard deviation of 1 and 95% of the distribution is within 2 standard deviations of the mean. The observed value of -1.06 is well within this range so the probability of getting a proportion as extreme as .20 in a sample of 60 while the proportion for all students is .26 is greater than 5%. We can't reject H_0 based on this result.

(b) Since we are asked to use the sample data, I will use the observed sample standard deviation $\sqrt{\left(\frac{.20(1-.20)}{60}\right)}$ = .0516 in constructing the confidence interval for p, the true proportion.

From the Tables page 37 we note that

$P(Z \leq 1.96) = .9750$

Hence $P(Z \geq 1.96) = .025$ so $P(\frac{.20-p}{.0516} \geq 1.96) = .025$

Also $P(Z \leq -1.96) = .025$ so $P(\frac{.20-p}{.0516} \leq -1.96) = .025$

Hence $P(.20 - 1.96 \times .0516 \leq p \leq .20 + 1.96 \times .0516) = .95$

$P(.10 \leq p \leq .30) = .95$

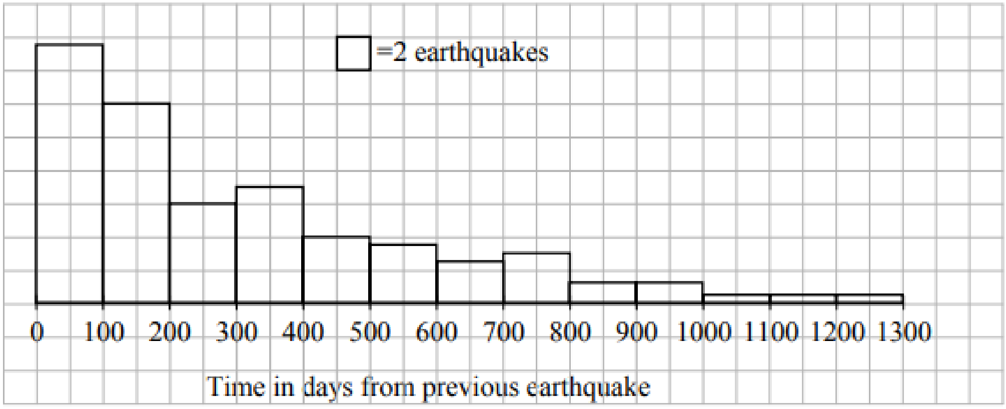
The required confidence interval is $.10 \leq p \leq .30$, i.e. 10% to 30%.

Q10

(i) Create a suitable graphical representation of the distribution.

Histogram.
First, divide up unequal intervals, and estimate an allocation of the data (optional step).

800 – 900	900 – 1000	1000 – 1100	1100 – 1200	1200 – 1300
2·5	2·5	1	1	1



(ii) Describe the distribution. Your description should refer to the shape of the distribution and should include an estimate of the median.

> The distribution is skewed to the right. (Or, e.g., there is a lot of data to the left, and it tails off to the right, etc.)
>
> The median is approximately 220 days.

(iii) The mean time between these earthquakes is 309 days and the standard deviation is 277 days. Suppose that such an earthquake has just occurred and that we want to find the probability that the time to the next one will be between 100 and 200 days. Explain why it would **not** be correct to use standard normal distribution tables (z-tables) to do this.

> Because the distribution is not normal.

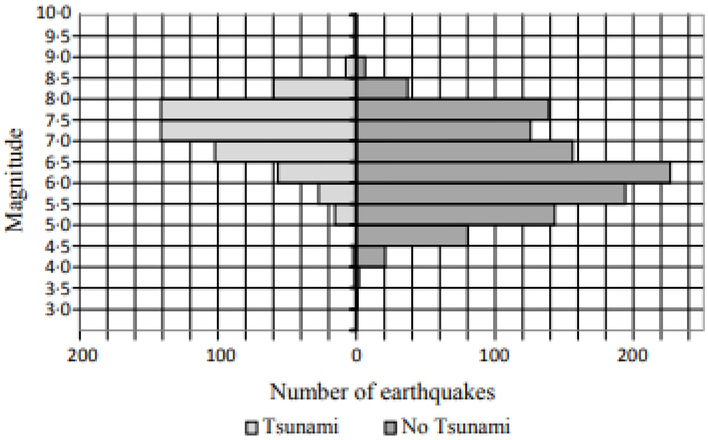
(iv) Based on the information presented in this question so far, what is the best estimate for the probability described in part **(iii)** above? Explain your reasoning.

> The best estimate is to assume the probability is as reflected in the proportion of such intervals in the historical data.
>
> $$\frac{24}{115} \approx 0{\cdot}2$$

(v) As stated at the beginning, the students chose to analyse earthquake timings by looking at the time intervals between the occurrences of a particular type of earthquake. Suggest a different way that they could have looked at the data in the database in order to try to find patterns in the timing of earthquakes.

> - They could have looked at the number of earthquakes each year, or some other interval of time (e.g. distribution of earthquakes per decade, per year, etc.)
> - They could have redefined serious earthquakes as earthquakes greater than a certain magnitude; earthquakes in less populated areas are not included.
> - The data set could have been broadened to include less serious earthquakes. This could result in a different pattern.

(b) The students heard a reporter saying that "strong earthquakes will cause large destructive ocean waves called tsunamis, while weaker ones will not". They decide to check this. They draw two histograms back to back, one showing the magnitudes of the earthquakes that caused tsunamis, and the other showing the magnitudes of those that did not. They use all of the suitable data from the 20th century that were recorded in this particular database.



(i) Comment on the reporter's statement, using information from the diagram to support your answer, and suggest a more accurate statement.

> The statement is too deterministic – strong earthquakes don't always cause tsunamis and weak ones sometimes do. A better statement would be "Strong earthquakes are more likely to cause tsunamis than weaker ones."

(ii) By taking suitable readings from the diagram, estimate the probability that an earthquake of magnitude between 6·5 and 7·0 will cause a tsunami.

> About 103 of these did and about 156 didn't. So probability is $\frac{103}{259} \approx 0{\cdot}4$

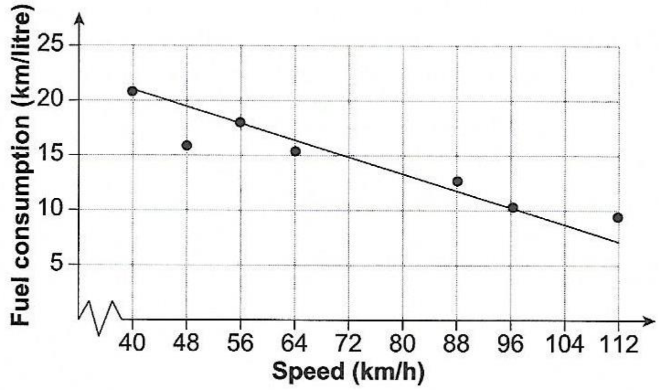
(iii) Consider the next six earthquakes of magnitude at least 7·5. Find an estimate for the probability that at least four of them will cause a tsunami, assuming that these six events are independent of each other.

> Tsunami : 142 + 60 + 8 =210
>
> No tsunami: 139 + 36 + 7 = 182
>
> $p \approx \frac{210}{392} \approx 0{\cdot}54$ *
>
> $\binom{6}{4}(0{\cdot}54)^4(0{\cdot}46)^2 + \binom{6}{5}(0{\cdot}54)^5(0{\cdot}46) + (0{\cdot}54)^6 = 0{\cdot}421$

Q11

(a) Using a calculator, the correlation coefficient is -.957.

(b)



(c) As speed increases by 1 km/h the average distance travelled on 1 litre of fuel decreases by 0·15km.

(d) (i) Time taken by Mary = $\frac{260}{96}$ = 2.708 hours

Time taken by Jane = $\frac{260}{112}$ = 2.321 hours

Mary took .387 hours longer, i.e. 23 minutes.

(ii) Based on the data in the table, at 96 km/hour the fuel consumption was 11 km/litre

Mary needed $\frac{260}{11}$ = 23.64 litres

Based on the data in the table, at 112 km/hour the fuel consumption was 9 km/litre

Jane needed $\frac{260}{9}$ = 28.89 litres

Jane needed 5.25 more litres, costing 5.25 × 1.329 = €6.98.