"Essential Mathematics & Statistics for Science" Second Edition by Dr G Currell & Dr A A Dowman (John Wiley & Sons)

Answers to 'Q' Questions

8 Distributions & Uncertainty

To navigate, use the Bookmarks in the PDF file

Q8.1

Read the probability values from Figure 8.1:

i)	80 - 81	= 0.1915
ii)	81 - 83	= 0.1499 + 0.0918 = 0.2417
iii)	>80	= 0.5000
iv)	78 - 82	= 2 × (0.1915 + 0.1499) = 0.6828
v)	76 - 84	= 2 × (0.1915 + 0.1499 + 0.0918 + 0.0441) = 2 × (0.4773) = 0.9546
vi)	74 - 86	= 2 × (0.4773 + 0.0165 + 0.0049) = 2 × (0.4987) = 0.9974
vii)	72 - 88	= 2 × (0.4987 + 0.0011 + 0.0002) = 2 × (0.5000) = 1.0000 (within an accuracy of 0.0001)

Q8.2

i) A good estimate of the mean value gives μ = 1.5

ii) The extreme tails of the distribution go from about -2.0 to +5.0.
This is a distance of 7.0.
Hence 6 standard deviations will equal about 7.0
Hence 1 standard deviation, $\sigma \approx 7/6 = 1.17$
(which is quite close to the true value of 1.2!)

iii) The width at half height of the distribution goes from about 0.1 to about 2.9.
This is a distance of 2.8
Hence $2.4\times\sigma \approx 2.8$, and
$\sigma \approx 2.8 / 2.4 = 1.17$
(which again is quite close to the true value of 1.2!)

Q8.3

True mean value μ = 50, standard deviation σ = 4

i) 46 – 54 is 50 ± 4 which is $\mu \pm 1.00\times\sigma$
Using Table 8.1, this is 68.3%
68.3% of 100,000 = (68.3/100) × 100000 = 68,300

ii) 42.16 – 57.84 is 50 ± 7.84 7.84/4 = 1.96

So 42.16 – 57.84 is 50 ± 1.96×σ 95% from Table 8.1
95% of 100,000 = (95/100) × 100000 = 95,000

iii) 38-62 is 50 ± 12 or 50 ± 3.00×σ
99.7% lie within this range so 100-99.7 = 0.3% lie outside.
0.3% of 100,000 = (0.3/100) × 100000 = 300

iv) >54
From i) 68.3% lie within 46-54 so 100-68.3 = 31.7% lie outside. This is in 2 symmetrical 'tails' so the percentage in one 'tail' is (31.7)/2 = 15.85%
15.85% of 100,000 = (15.85/100) × 100000 = 15850

Q8.4

i) Possible sources of variation:
A) In the measurement of cadmium in the iron, the variation between experimental results will be due to variations in the measurement process itself, e.g. human variations in preparation of samples, instrumental variations in recording measured values.
B) The dominant source of variation will be between the different individual soya bean plants - there will be relatively little uncertainty in the actual measurement process.

ii) The mean value of the results will be the best estimate of the single true value of the cadmium concentration. Repeating the experimental results has enabled the metallurgist to estimate, and compensate for, the uncertainty in the measurement process.

iii) The mean value of the results will be the best estimate of the true population mean value that would be obtained if the entire soya bean crop had been measured. In this case the sampling of the crop has enabled the biologist to obtain a good estimate for the entire crop without the need to measure every plant.

Q8.5

95% of experimental results, x, would fall in the range: μ - 1.96×σ to μ + 1.96×σ

We have values for the mean and standard deviation:
μ = 23
σ = 2
Hence, we can expect that:
95% of experimental results would fall in the range:
from 23 - 1.96×2 to 23 + 1.96×2
i.e.
from 19.08 ppb to 26.92 ppb

Q8.6

We know the values of the population mean, μ = 23, ppb,
population standard deviation, σ = 2 ppb, sample size, n = 4

i) As we know the population standard deviation, σ, we use the following expression for Standard Error:

$$SE = \frac{\sigma}{\sqrt{n}}$$

giving

SE = 2 / √4 = 2 / 2 = 1 ppb

ii) 95% of mean values, $\bar{x}$, would fall in the range:
μ - {1.96× σ $\sqrt{n}$} to μ + {1.96× σ $\sqrt{n}$}
Substituting values, the 95% range becomes:
23 - {1.96 × $\tilde{\sigma}$ √4} to 23 + {1.96 × $\tilde{\sigma}$ √4}
23 - 1.96 to 23 + 1.96
which is from 21.04 to 24.96

Note that, by taking an average of 4 readings, the uncertainty in the average (mean) value is less than the uncertainty in a single reading - compare with question Q8.5.

Q8.7

i) Mean value of the 3 measurements, $\bar{x}$ = 19.33.
Sample standard deviation, s = 1.026

Sample size = 3,
Degrees of Freedom = 3 - 1 = 2
Giving $t_{2,\,0.05,\,2}$ = 4.3

95% Confidence Interval of the true mean,

$$CI(\mu l\,95\%) = \bar{x} \pm \left\{ t_{2,0.05,2} \times \frac{s}{\sqrt{n}} \right\} = 19.33 \pm \left\{ 4.3 \times \frac{1.026}{\sqrt{3}} \right\} = 19.33 \pm 2.55$$

The student can say with 95% confidence that the true value will lie between

16.78 to 21.88

and should then round the *range* to values with a reasonable number of significant figures:

16.7 ppb to 21.9 ppb

ii) Mean value of the 6 measurements, $\bar{x}$ = 18.93.
Sample standard deviation, s = 0.987

Sample size = 6,
Degrees of Freedom = 6 - 1 = 5
Giving $t_{2,0.05,5}$ = 2.57

95% Confidence Interval of the true mean,

$$CI(\mu|95\%) = \bar{x} \pm \left\{ t_{2,0.05,5} \times \frac{s}{\sqrt{n}} \right\} = 18.93 \pm \left\{ 2.57 \times \frac{0.987}{\sqrt{6}} \right\} = 18.93 \pm 1.04$$

The student can say with 95% confidence that the true value will lie between

17.89 to 19.97

and should then round the *range* to values with a reasonable number of significant figures:

17.8 ppb to 20.0 ppb

Note that, by taking a few more measurements, the student achieves a far greater degree of accuracy in quoting the final results.
An increased sample size reduces uncertainty due to the effect of the '√*n*' factor and also due to the reduced size of the '*t*' factor.

iii) The best-estimate of the true value is 18.2 ppb.
However, the student has only made one measurement, and has NO idea about the uncertainty in the measurement.

The student would not be able to quote any Confidence Interval for his results.

Q8.8

CV = 5% = $100 \times \frac{\sigma}{\bar{x}}$ and $\bar{x}$ is 4.5 ppm

So $\sigma = 4.5 \times \frac{5}{100} = 0.225$ ppm

Standard uncertainty $u(x) = \frac{0.225}{\sqrt{8}} = 0.08$

Q8.9

$$Ru(x) = 100 \times \frac{u(x)}{\bar{x}} \qquad \rho = \frac{m}{V}$$

Variable	*Value, x*	*u(x)*	⇒ ⇐	*Ru(x)*

Mass, m, (g)	4.3	0.3	$\Rightarrow$	=100×0.3/4.3=7.0 %
Volume, V, (cm^3)	2.3	0.2	$\Rightarrow$	=100×0.2/2.3=8.7 %
Density, ρ, (g cm^{-3})	4.3/2.3 = 1.87	=1.87×11.12/100 = 0.21	$\Leftarrow$	= $\sqrt{7.0^2 + 8.7^2}$ = 11.12%

Q8.10

Using the same data as Example 8.7, calculate the value and uncertainty for the diagonal of the rectangle, $h = \sqrt{a^2 + b^2}$ (using Pythagoras), given that:

a = 8.4 mm with a standard deviation uncertainty, $u(a)$ = 0.5 mm,
b = 6.7 mm with a standard deviation uncertainty, $u(b)$ = 0.5 mm,

Best value of $h = \sqrt{8.4^2 + 6.7^2} = 10.75$ mm

from [8.9] $Ru_a = 100 \times \dfrac{u(a)}{a} = 100 \times \dfrac{0.5}{8.4} = 5.95\%$

$Ru_b = 100 \times \dfrac{u(b)}{b} = 100 \times \dfrac{0.5}{6.7} = 7.46\%$

Let $x = a^2 = 8.4^2 = 70.56$ mm^2 and $y = b^2 = 6.7^2 = 44.89$ mm^2

$a^2 + b^2 = x + y = z = 70.56 + 44.89 = 115.45$ and

$h = \sqrt{z} = \sqrt{115.45} = 10.744$

[8.16] $Ru_x = n \times Ru_a$ so in this case $Ru_x = 2 \times Ru_a = 2 \times 5.95 = 11.9\%$
$Ru_y = n \times Ru_b$ so in this case $Ru_y = 2 \times Ru_b = 2 \times 7.46 = 14.92\%$

Using [8.9], $u_x = \dfrac{x \times Ru_x}{100} = \dfrac{70.56 \times 11.9}{100} = 8.40$ mm

$u_y = \dfrac{y \times Ru_y}{100} = \dfrac{44.89 \times 14.92}{100} = 6.70$ mm

[8.14] $u_z = \sqrt{u_x^2 + u_y^2} = \sqrt{8.40^2 + 6.70^2} = \sqrt{70.56 + 44.89} = 10.745$

Ru_z = (10.745/115.45) ×100 = 9.307%

Finally using [8.16], remembering in $h = \sqrt{z}$, n = 0.5 for the square root

$R_{u_h} = 0.5 \times R_{u_z} = 0.5 \times 9.307 = 4.65\%$

So $u_h = \dfrac{4.65 \times 10.745}{100} = 0.50$ mm

Q8.11

Using the binomial probability equation

$$p(r) = {}_nC_r \times p^r \times (1 - p)^{(n-r)}$$

for r = 0, 1, 2, 3, 4 and 5:

$p(0) = {}_5C_0 \times 0.25^0 \times (1 - 0.25)^{(5-0)} = 1 \times 0.75^5 = 0.2373$
$p(1) = {}_5C_1 \times 0.25^1 \times (1 - 0.25)^{(5-1)} = 5 \times 0.25 \times 0.75^4 = 0.3955$
$p(2) = {}_5C_2 \times 0.25^2 \times (1 - 0.25)^{(5-2)} = 10 \times 0.25^2 \times 0.75^3 = 0.2637$
$p(3) = {}_5C_3 \times 0.25^3 \times (1 - 0.25)^{(5-3)} = 10 \times 0.25^3 \times 0.75^2 = 0.0879$
$p(4) = {}_5C_4 \times 0.25^4 \times (1 - 0.25)^{(5-4)} = 5 \times 0.25^4 \times 0.75 = 0.0146$
$p(5) = {}_5C_5 \times 0.25^5 \times (1 - 0.25)^{(5-5)} = 1 \times 0.25^5 = 0.0010$

The total of all the probabilities is 1 as one of them must result.

Q8.12

This is a binomial distribution with a total of n = 10 children, and the probability that each person could be cured, p = 0.7.
The probability that just r people will be cured is given by the binomial probability:

$$p(r) = {}_nC_r \times p^r \times (1 - p)^{(n-r)}$$

i) For no people to be cured, r = 0, giving:

$p(0) = {}_{10}C_0 \times 0.7^0 \times (1 - 0.7)^{(10-0)}$
$p(0) = {}_{10}C_0 \times 1 \times (0.3)^{10} = 5.9 \times 10^{-6}$

ii) For 1 person to be cured, r = 1, giving:

$p(1) = {}_{10}C_1 \times 0.7^1 \times (1 - 0.7)^{(10-1)}$
$p(1) = {}_{10}C_1 \times 0.7 \times (0.3)^9 = 1.4 \times 10^{-4}$

iii) For 8 people to be cured, r = 8, giving:

$p(8) = {}_{10}C_8 \times 0.7^8 \times (1 - 0.7)^{(10-8)}$
$p(8) = {}_{10}C_8 \times 0.7^8 \times (0.3)^2 = 0.2335$

iv) For 9 people to be cured, r = 9, giving:

$p(9) = {}_{10}C_9 \times 0.7^9 \times (1 - 0.7)^{(10-9)}$
$p(9) = {}_{10}C_9 \times 0.7^9 \times (0.3)^1 = 0.1211$

v) For 10 people to be cured, $r = 10$, giving:

$$p(10) = {}_{10}C_{10} \times 0.7^{10} \times (1 - 0.7)^{(10-10)}$$
$$p(10) = {}_{10}C_{10} \times 0.7^{10} \times (0.3)^{0} = 0.0282$$

vi) The probability of at least 8 people will be cured can be obtained by adding the probabilities $p(8) + p(9) + p(10)$ and these were obtained in iii) iv) and v) above

$p(8) + p(9) + p(10) = 0.2335 + 0.1211 + 0.0282 = 0.3828$

vii) The probability of no more than 7 people being cured can be obtained by adding $p(0)$ to $p(7)$, or by finding $1-\{ p(8) + p(9) + p(10)\}$, which from vi)

Probability of no more than 7 = 1 – 0.3828 = 0.6172

Q8.13

The probability, p, that a particular person will die is 50 per 100,000:

$$p = 50 / 100{,}000 = 0.0005$$

The low probability and random incidence will give a Poisson distribution.

In a group size of 1000 people, the deaths, μ, *on average*, that can be predicted is given by:

$$\mu = 1000 \times 0.0005 = 0.5 \text{ deaths.}$$

The probability that just r people will die is given by the Poisson probability:

$$P(r) = \frac{e^{-\mu} \times \mu^{r}}{r!}$$

The values for $P(r)$ can be calculated by working out the above expression, or by using the POISSON function in EXCEL.

i) For no deaths, $r = 0$, giving:

$$P(0) = e^{-0.5} \times 0.5^{0} / 0! = e^{-0.5} = 0.6065$$

ii) For 1 death:

$$P(1) = e^{-0.5} \times 0.5^{1} / 1! = e^{-0.5} \times 0.5^{1} = 0.3033$$

iii) For 2 deaths:

$$P(2) = e^{-0.5} \times 0.5^{2} / 2! = e^{-0.5} \times 0.5^{2} / 2 = 0.0758$$

iv) For more than 2 deaths:

$$P(>2) = 1- P(2 \text{ or less}) = 1- \{P(0)+P(1)+P(2)\}$$
$$= 1 - \{0.6065 + 0.3032 + 0.0758\} = 1 - 0.9856 = 0.0144$$

Q8.14

37 frogs out of 50 are female

In a sample of n = 50 frogs, the mean number of female frogs, $\bar{r}$ = 37.

Using [8.22] the best-estimate for the true mean of 50 frogs:

$$CI \approx \bar{r} \pm \left\{ t_{2,\alpha\infty} \times \sqrt{\bar{r} \times \left(1 - \frac{\bar{r}}{n}\right)} \right\} = 37 \pm \left\{ 1.96 \times \sqrt{37 \times \left(1 - \frac{37}{50}\right)} \right\}$$

=37 ± 1.96 × 3.11 = 37 ± 6.08

Hence the best-estimate for the *proportion* of female frogs will be

$$P = \frac{37}{50} \pm \frac{6.08}{50} = 0.74 \pm 0.12$$

Hence the 95% confidence interval for the proportion will be from 0.62 to 0.86

Q8.15

i) CD = Variance/mean for a Poisson distribution so
$CD_A = 1.76^2/3.13 = 0.99$, $CD_B = 2.04^2/3.13 = 1.33$

ii) For a Poisson distribution, CD ~ 1, so A is a Poisson

iii) For B, CD>1 so clumping is possible.