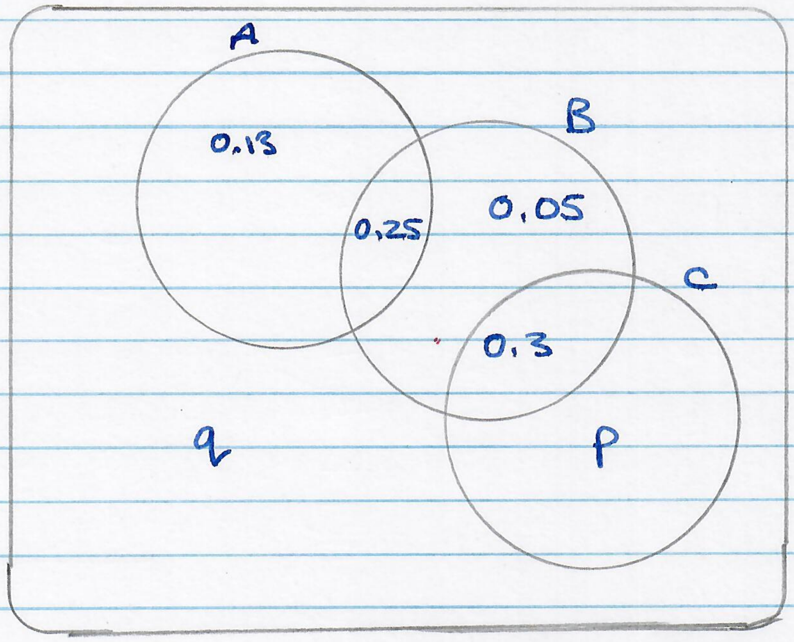
1)



a) $P(A) = 0.13 + 0.25 = \underline{\underline{0.38}}$

b) B and C are independent, so

$$P(B).P(C) = P(B \cap C)$$

so $(0.25 + 0.05 + 0.3) \cdot (0.3 + p) = 0.3$

$$0.6(0.3 + p) = 0.3$$

$$0.3 + p = 0.5$$

$$\underline{\underline{p = 0.2}}$$

Since the overall probability of $\mathcal{E}$ is 1,

$$q = 1 - (0.13 + 0.25 + 0.05 + 0.3 + 0.2)$$
$$= 1 - (0.93)$$
$$= \underline{\underline{0.07}}.$$

c) $P(A|B') = \dfrac{P(A \cap B')}{P(B')}$

$= \dfrac{0.13}{1-(0.25+0.05+0.3)}$

$= \dfrac{0.13}{0.40}$

$= \underline{\underline{0.325}}$

2) Model the no. of packets in each box of 40 that have prizes by $T \sim B(40, \frac{1}{7})$

a) A condition is that the distribution of packets into the boxes must be random - i.e. each packet's prize or not status must be independent.

b) $T \sim B(40, \frac{1}{7})$

(i) $P(T=6) = {}_{40}C_6 \left(\frac{1}{7}\right)^6 \left(\frac{6}{7}\right)^{34}$

$= \underline{\underline{0.173}}$ (BINOM.DIST(6, 40, $\frac{1}{7}$, FALSE))

(ii) $P(T<3) =$

$P(T=0) = {}_{40}C_0 \left(\frac{1}{7}\right)^0 \left(\frac{6}{7}\right)^{40} = 0.0021$

$+ P(T=1) = {}_{40}C_1 \left(\frac{1}{7}\right)^1 \left(\frac{6}{7}\right)^{39} = 0.0140$

$+ P(T=2) = {}_{40}C_2 \left(\frac{1}{7}\right)^2 \left(\frac{6}{7}\right)^{38} = \underline{0.04550}$

0.0616

c) $P(1 \text{ box has } < 3) = 0.0616 = p$

$P(1 \text{ box has } \geq 3) = 1 - 0.0616 = 0.9384 = q$

$P(\text{exactly 2 boxes have } < 3)$

$$= {}_5C_2 \; p^2 q^3$$

$$= \frac{5 \times 4}{2} (0.0616)^2 (0.9384)^3$$

$$= \underline{\underline{0.3136}}$$

d) Sample = 110 packets; 9 contain a prize

(note we've switched back from 'boxes' - which is convenient.)

Test for the significance of this result,

with
- $H_0 = p = \frac{1}{7}$, Kamil is wrong
- $H_1 = p < \frac{1}{7}$, Kamil is right
- use a <u>one-tail</u> test, since Kamil is clear about 'less than'.
- use a 5% significance level.

So we want $P(T \leq 9)$ (BINOM.DIST$(9, 40, \frac{1}{7}, V)$)

$$= \underline{\underline{0.0383}}.$$

This is less than the significance level of 0.05,

3) a) A value should be assigned to "tr" (trace) - eg 0.025.

b) $n = 184$ $\quad \sum x = 390 \quad \sum x^2 = 4336$

i) mean $\bar{x} = \dfrac{\sum x}{n} = \dfrac{390}{184} = \underline{2.120}$

ii) standard deviation $\sigma = \sqrt{\dfrac{\sum x^2}{n} - \bar{x}^2}$

$$= \sqrt{\frac{4336}{184} - 2.120^2}$$

$$= \underline{\underline{4.367}}$$

c) i) The Leeming page in the LDS only covers some months of the year - May-Oct. So it doesn't represent the whole year and we can't say Ben's results hold for the whole (annual) period.

ii) Since the LDS data covers the <u>drier</u> summer months, we would expect the estimate for the <u>whole year</u> to be <u>higher</u>: i.e. the number in b(i) is an underestimate.

4) a) $X \sim N(175.4,\ 6.8^2)$ (X = height of men from region A)

$P(X > 180) = 1 - P(X \leq 180)$

$= 1 - 0.7506$ (NORM.DIST(180, 175.4, 6.8, ✓))

$= \underline{\underline{0.2494}}$

b) This is to do with the distribution of the means of samples of the main population, which itself is a normal distribution:

$\bar{X} \sim N\left(175.4,\ \frac{6.8^2}{52}\right) = N(175.4,\ 0.889)$

$= 0.9430^2$

Use a test · H_0 mean = 175.4

· H_1 - - $\neq$ 175.4

(ie. 177.2 is sig. different.)

· sig. level 5% but a 2-tail test, so value of 0.025 prob^y.

So $P(\bar{X} > 177.2) = 1 - P(\bar{X} \leq 177.2)$

$= 1 - 0.9719$

$= 0.0281$

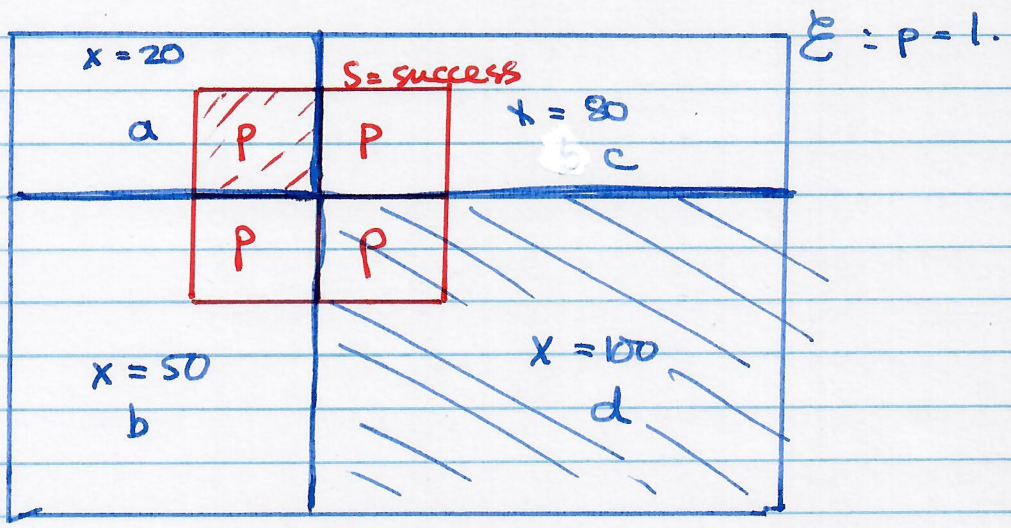
which is <u>greater than</u> 0.025

so there is not enough evidence to support H_1.

c) The p-value, total prob. for the <u>two</u> tails,

$= 2 \times 0.0281 = \underline{0.0562}$

5) This question is baffling in various ways, but it seems to come down to a probability space like:



(Note the blue space can't actually be a rectangle)

And here:

$$\frac{p}{a} = \frac{k}{20} \qquad \frac{p}{c} = \frac{k}{80}$$

$$\frac{p}{b} = \frac{k}{50} \qquad \frac{p}{d} = \frac{k}{100}$$

a) Given all these equalities, we can say

$$p = \frac{kb}{50} = \frac{kc}{80}$$

so $c = \dfrac{80}{50} b$ as required.

We can also say:

$$\frac{ka}{20} = \frac{kb}{50} \quad : \quad b = 2.5a$$

similarly $c = 4a$

$d = 5a$

And since $a+b+c+d$ must equal 1,

$$a + 2.5a + 4a + 5a = 12.5a = 1$$

So $a = \frac{2}{25}$, $b = \frac{5}{25}$, $c = \frac{8}{25}$, $d = \frac{10}{25}$

which is the probability dist^n of X as required.

c) Nav's result suggests for him that

$$\frac{p}{d} = 30\% \quad : \quad p = 0.3 = \frac{k}{100} \quad \text{so } k = 30.$$

But this would mean

$P(S \mid X = 20) = \frac{30}{20}$ which is >0

and similarly - - - -

which cannot happen.

(Nav. is just too good)

6) a) Interpolating from inside histogram bars is slightly dodgy, but since we're asked for an estimate:

10-12 hours	:	2×4.2	=	8.4
12-16 hours	:	4×4	=	16
16-20 hours	:	4×3.5	=	14
20-30 hours	:	10×1	=	10
				48.4

And the overall results in the histogram are:

0-4	:	4×2.5	=	10
4-7	:	3×3	=	9
7-12	:	5×4.2	=	21
12-16	:	4×4	=	16
16-20	:	4×3.5	=	14
20-40	:	20×1	=	20
				90

So $P(10 < T < 30) = \frac{48.4}{90} = 0.538$

b) It's doubtful, because clearly the data is skewed (not symmetric). (And the mean is not near the peak of the given curve, so a normal curve will not fit well either to left or right.)

c) $y = kxe^{-x}$ x in hours$\times 10$ $0 \leq x \leq 4$

Let $I = \int_0^n xe^{-x}\,dx.$

Using integ^n by parts:
(omitting dx)

$$\int \frac{d}{dx}\left(xe^{-x}\right) = \int -xe^{-x} + \int e^{-x}$$

$$= xe^{-x}$$

So $$\int xe^{-x}\,dx = -xe^{-x} + \int e^{-x}$$

$$= -e^{-x}(1+x)$$

So $$I = k\left[-e^{-x}(1+x)\right]_0^n$$

$$= k\left[e^{0}(1) - e^{-n}(1+n)\right]$$

$$= \underline{\underline{1-(n+1)e^{-n}}} \text{ as required.}$$

d) So for Xiang's model:

$$\int_0^4 y\,dx = k\int_0^4 xe^{-x}\,dx$$

$$= k\left(1-(n+1)e^{-n}\right)$$

$$= k\left(1-5e^{-4}\right)$$

$$= 90 \text{ (we know from (a)).}$$

$$k = \frac{90}{1-5e^{-4}} = \frac{90}{0.908} = 99.07$$

$= \underline{\underline{99}}$ to nearest integer as required.

e) i) $P(10 < T < 30)$ using $N(14.9, 9.3^2)$

$= \text{NORM.DIST}(30, 14.9, 9.3, \checkmark)$
$- \ldots \ldots (10, - - \cdot)$

$= 0.9478 - 0.2991 = \underline{\underline{0.6846}}$

ii) $P(1 < T' < 3)$ where T' is as in Xiang's model:

$I_1 \quad \int_0^1 y\,dx = 99(1 - 2e^{-1}) = 26.1599$

$I_2 \quad \int_0^3 y\,dx = 99(1 - 4e^{-3}) = 79.2842$

(could have done this simply as $\int_1^3$)

$I_2 - I_1 = 53.1244$

So $P(1 < T' < 3) = \dfrac{k(I_2 - I_1)}{90}$

$= \dfrac{99 \times 53.1244}{90}$

$= \underline{\underline{0.5844}}$

f) Xiang's model is reasonable for $0 \leq T \leq 40$, but we don't know any data for patients who stay longer than 40 hours.