1) a) 96% have an innie.

P(person has an innie) = $\frac{96}{100}$ = 0.96.

i) Using a binomial model,

P(no student has an ~~innie~~ outie)

= P(all 30 students have an innie)

= $(0.96)^{30}$ = $\underline{\underline{0.2939}}$.

ii) The binomial model requires

a) there be a pre-decided no. of trials

b) the trials be ind^{t}

c) trials have only 2 possible outcomes

d) P(success) is known and constant.

iii) 66% have fluff.

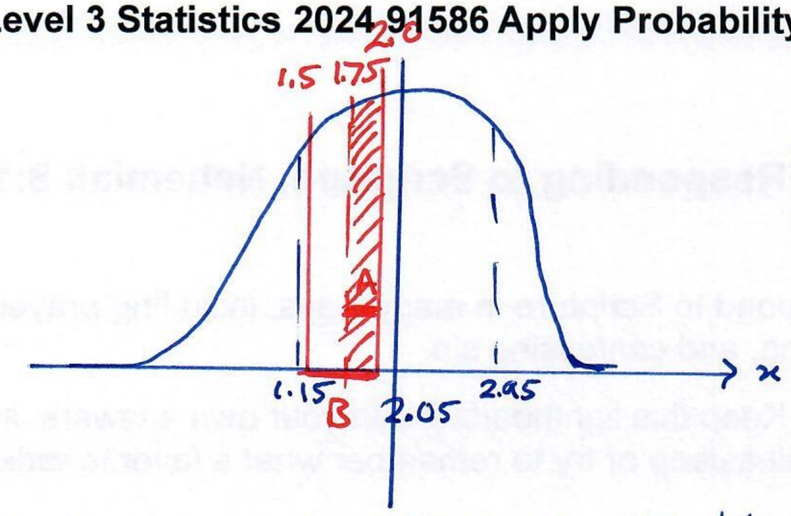
Using the same binomial model:

P(1) has an outie = $30 \times 0.04 \times 0.96^{29}$
= 0.3673

P(has fluff) = 0.3673×0.66

= $\underline{\underline{0.2424}}$

b)



$\mu = 2.05$

$\sigma = 0.9$

The actual normal distribution is blue, but this is mainly an exercise in calculating the cumulative probabilities up to the red values, using the 'standard' normal distn. Values for z equivalent to x are given by

$$z = \frac{(x - \mu)}{\sigma}$$

Given we know the size lies between 1.5 and 2.0 this means we need to know the prob (1.70 or above) as a fraction of prob (1.5 – 2.0), or $\frac{A}{B}$ in the diagram.

And just to add to the pain, the prob. table is drawn upwards of from $x=0$, so we'll have to invert everything.

Converting to standard dist^n:

$$Z(x = 1.50) = -0.611$$

$$Z(x = 1.75) = -0.333$$

$$Z(x = 2.00) = -0.056$$

Inverting around 0 gives values and sets the question as:

$$\frac{P(0.056 < Z < 0.333)}{P(0.056 < Z < 0.611)}.$$

Cumulative probabilities are:

$$P(0 < Z < 0.056) = 0.0223$$

$$P(0 < Z < 0.333) = 0.1304$$

$$P(0 < Z < 0.611) = 0.2294$$

So we need:

$$\frac{0.1304 - 0.0223}{0.2294 - 0.0223}$$

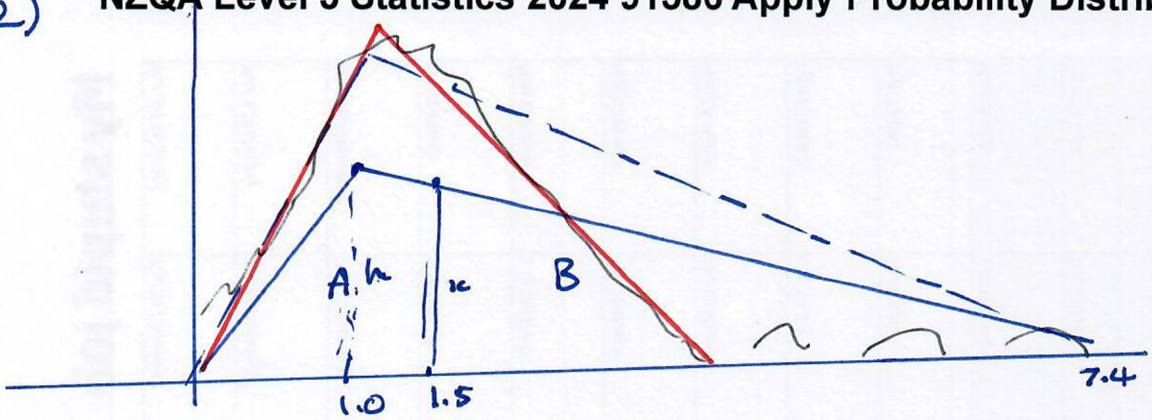
$$= \frac{0.1081}{0.2071} = \underline{\underline{0.5220}}$$

ii) The model may be OK for belly buttons within 1 SD of the mean (ie a diameter between 1.15cm and 2.95cm.)

But at 3 SD of the mean this would mean limits of −0.65 to 4.75.

- The first of these is nonsense (though arguably there'd be very few people near that measurement anyway)
- The second seems perhaps too large, though again very few people would be near it anyway.

(The % lying outside 3 SDs should be
$0.5 - 0.4987 = 0.0013$ or 0.13%,
so this concern would be irrelevant in most cases)



a) i) The 'answer' gives the blue solid line, but there's no particular reason given for its choice of height. Happily the calculation required works regardless of the height. (Personally I'd go for the dotted blue line ... though they're both clearly wrong.)

ii) What's required is the ratio $\frac{A}{A+B}$:

all of these depend on the height h, though it cancels out.

Using areas of Δs and trapezium,

$$A = \frac{1.h}{2} + \frac{h+x}{2} \times 0.5$$

$$x = h \times \frac{5.9}{6.4}$$

$$B = \frac{x.5.9}{2}$$

$$A = \frac{h}{2}\left(1 + 0.5 + \frac{5.9}{6.4} \times 0.5\right)$$

$$= \frac{h}{2}\left(1.9610\right)$$

$$= 0.9805\,h.$$

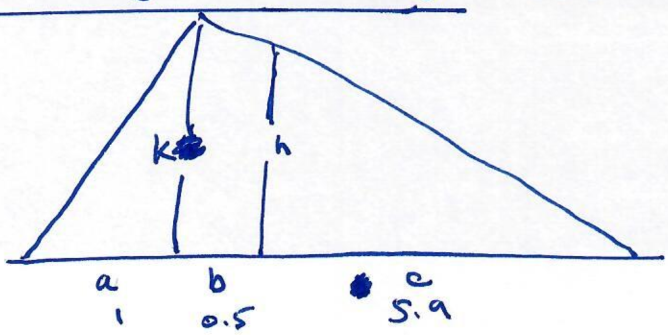
$$B = h.\ \frac{5.9 \times 5.9}{6.4 \times 2} = 2.7195\,h.$$

So $$\frac{A}{A+B} = \frac{0.9805}{2.7195} = \underline{\underline{0.2650}}$$

The answer given agrees with this though it's not obvious why they calculate n ... or why they get a value for it as 0.2492.

I've written about this on the next page - though I can't work it out.

Further analysis (puzzling)



$h = \frac{2(c)}{(a+b+c)\,a}$ why? when h has been picked arbitrarily.

Could say: $k = \frac{N}{2(a+b+c)}$, which would

at least make the area under the curve N.

but they don't

I'd say: $h = \frac{k \times c}{b+c}$. and you don't

need to know h.

Is it that they're saying prob overall = 1.

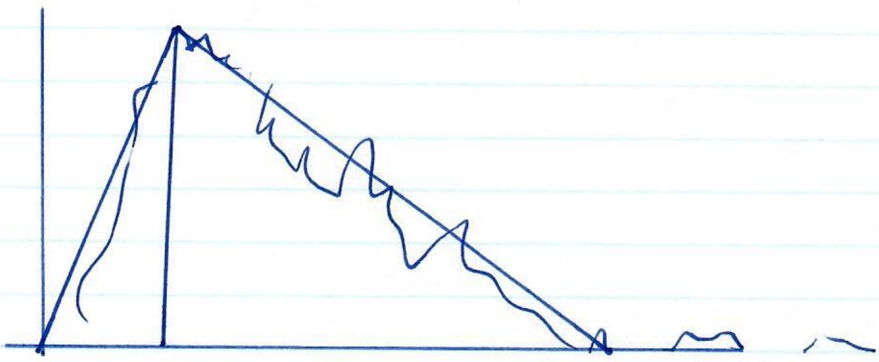
That would mean $k = \frac{1}{7.4} \times 2$

$= 0.2702.$

which would make $h = 0.2492.$

but that's not the answer they give.

iii) A better triangular distⁿ would be to consider the upper values as outliers and construct the main triangle between the fre weight values of 0 and something like 4.5, with a mode 0.9.



0 is min because obviously no value <0.
4.5 is max - - - this fits the data.
0.9 is mode because the greatest freq. is between 0.8 and 1.0
Values beyond 4.5 are considered outliers.

b) i) Mean (average) $= \sum t \cdot P(T=t)$

$= 2.031$

So 2 would be a good figure for the average daily use.

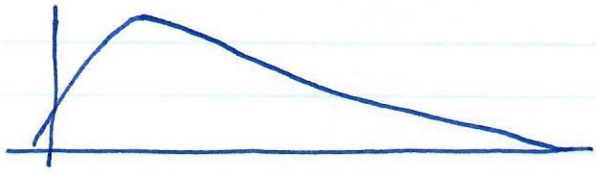
ii) $SD(C) = 0.5$
By calculation, $SD(T) = 1.405$

So $SD(C)$ is less than $SD(T)$.

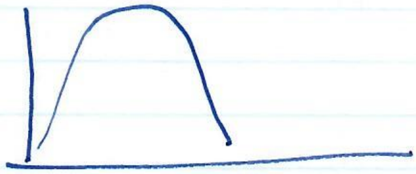
(The 'this day' language is confusing - it actually means 'any day', with C as the 'ordinary' daily value.)

Explain the difference ... well, they're two different dist^n's so that might explain it anyway (or it might not) but basically it looks like

T:



C:



C is much less spread out, a tighter narrower dist^n. So it's SD will be lower.

(That said, T doesn't look like a normal dist^n anyway, so I'm not sure the idea of SD is valid.)

iii) SD (T+C) = 1.754. Var (= SD^2) = 3.0765

If T and C are independent then

this should equal var (T) + var (C)

$= (1.405)^2 + 0.5^2$

$= 2.224$

so the vars are not equal, and T, C are <u>not</u> independent.

This suggests (not surprisingly) that if people change their clothes more they prob. have a shower more often.

3) Number who wash at least daily

$= freq(7) + freq(9) = 32 + 1 = 33$

As a proportion of 138 this is < 25%. ($\frac{33}{138} = 0.2391$)
so the headline is correct.

b) Poisson Distribution

$$P(X = x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \qquad \lambda = 2.8$$

where x is the number of occurrences.

We've just been told $\lambda = 2.8$, but the 'answers' calculate a new $\lambda = \frac{2.8}{7} = 0.4$. I've no idea why, unless they're 'standardising' in some way.

i) $P(x = 7)$ (or more) = (from the tables) 0.0163
+ 0.0057
0.0018
0.0005
0.0001

= 0.0244.

Compare this with the graph given – but it's nothing like a Poisson Dist^n for $x = 7$.

(and of course 7 washes might not occur 'everyday').

ii) the whole argument 'answers' seem misjudged.

iii) 1000 runs of the simulated set of 138 results.

iv) It's not remotely right (as was obvious) Particularly the value for $x=7$ fails to take in this (obviously high, and believable) figure.

It's possible treating this as an outlier or special case would allow a better PD with λ more like 1; but at the moment the values for $x=0$ and 1 are also hopeless.

Indeed everything is hopeless.

v) 'Answers' suggest a uniform model - but this would be no better.

Suggest a mixed uniform and discrete model like:

0	51/138	=	0.3696
1 – 6	(138-32-51)/6/138	=	0.0664 (for each x)
7	32/138	=	0.2319
8 and above		=	0.0001