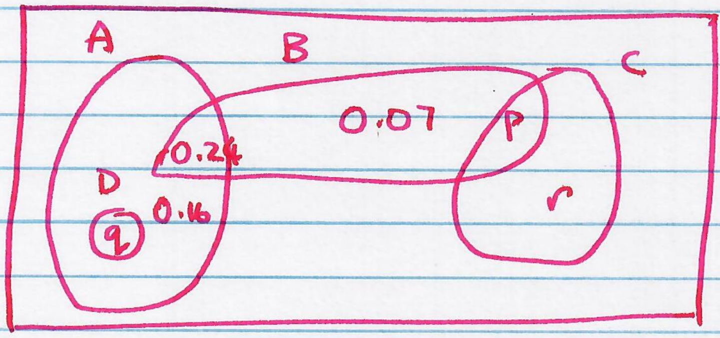
1)



a) Pairs of mutually exclusive events:

A, C
D, B
D, C

b) $P(B) = 0.4 = 0.24 + 0.07 + p = 0.31 + p$

So $p = 0.09$

c) A and B are independent, so

$P(A) \times P(B) = P(A \cap B)$:

$(0.24 + 0.16 + q) \times 0.4 = 0.24$

$q + 0.4 = \frac{0.24}{0.4} = 0.6$

So $q = 0.2$

d) $P(B'|C) = \frac{r}{r+p} = \frac{r}{r+0.09} = 0.64$ (given)

So $r = 0.64r + 0.64 \times 0.09$
$0.36r = 0.64 \times 0.09$ $r = 0.16$

Overall sum of probs. $= 1$,
So $\cancel{0.24}$ $0.24 + 0.16 + 0.07 + 0.09 + 0.16 + s = 1$
$s = 0.08$

Perth

2) a) The correlation (such as it is) appears to be negative, since a 'line of best fit' would clearly have a negative gradient.

b) Given that this is Perth in Australia and the months are June/July (winter), the y-axis could well be rainfall, measured in mm.

Heathrow.

c) we're asked about 'a' correlation, so do a 2-tailed test at 5% sig (2.5% either end) to test

H_0 there is no sig. correlation

H_1 there is ...

Critical value for corr^n on 30 samples is:

2.5% : ± 0.3610

−0.377 is lower than −0.3610,

so we regard the result as significant.

there is evidence for H_1, hypothesis that there is a correlation (and btw it's negative).

d) high humidity (97%) ⇒

(heuristically) very muggy, so not much real sunshine

(statistically) negative corr^n suggest not much sunshine.

3)

a) The range of the data encompasses all of the values, so it's $68-7 = \underline{\underline{61}}$

b) The interquartile range is given by the two ends of the box:

$$25 - 14 = \underline{\underline{11}}$$

c) $\sum x = 607.5 \qquad \sum x^2 = 17623.25$

(evidently we're including the 2 outliers.)

The mean $= \dfrac{\sum x}{n} = \dfrac{607.5}{27} = \underline{\underline{22.5}}$

d) $SD = \sqrt{\dfrac{\sum x^2}{n} - \bar{x}^2}$

$$= \sqrt{\frac{17623.25}{27} - 22.5^2}$$

$$= \underline{\underline{12.10}}$$

e) From the diagram, T identifies 2 outliers. Precisely, T would identify any points more than 3σ from the mean μ - i.e. $22.5 + 3 \times 12.1 = 22.5 + 36.3 = 58.8$ - so T actually identifies only 1.

f) If the median (half-way) value increases then both a and b must be to the right of 20: i.e. $a, b > 20$.

Mean doesn't change - so a and b must be equidistant from the mean 22.5.

So a could be 24 and b 21.

g) Both new values will be within 1 SD of the mean, so the new overall SD will be smaller. (They're close to the mean, so they weight the data towards the mean.)

4 a) $\frac{k}{10} + \frac{k}{20} + \frac{k}{30} + \frac{k}{40} + \frac{k}{50}$

$= \frac{60 + 30 + 20 + 15 + 12}{600} k$

$= \frac{137}{600} k.$

This sums to $= 1$ so $k = \frac{600}{137}.$

$D_1 + D_2$ 80 so possible combinations are:

D_1	30	40	50
D_2	50	40	30

Probability of this $= \frac{k^2}{1500} + \frac{k^2}{1600} + \frac{k^2}{1500}$

$= \frac{1}{100} \cdot \frac{600^2}{137^2} \left(\frac{2}{15} + \frac{1}{16} \right)$

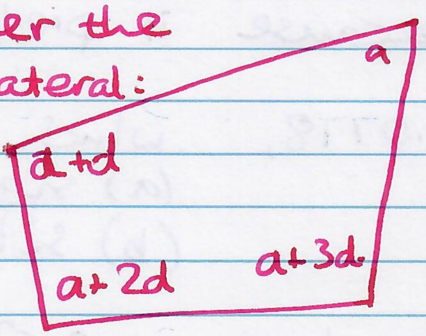
$= \frac{6}{137^2} \times 600 \left(\frac{32 + 15}{240} \right)$

$= \frac{6}{137^2} \times \frac{600}{240} \times 47.$

$=$ ~~which is wrong~~

$= \underline{\underline{0.0376.}}$

Consider the quadrilateral:



$\underline{4a + 6d = 360°}$ (angles of a quadrilateral)

It could be $\boxed{\begin{array}{l} a = 0: \\ a = 90 \end{array}}$ $\begin{array}{l} d = 60 \\ d = 0. \end{array}$

if $a > 50$

$4a > 200$

so $6d < 160$ (amount 'leftover')

$d < \frac{160}{6} = \underline{26.666}.$

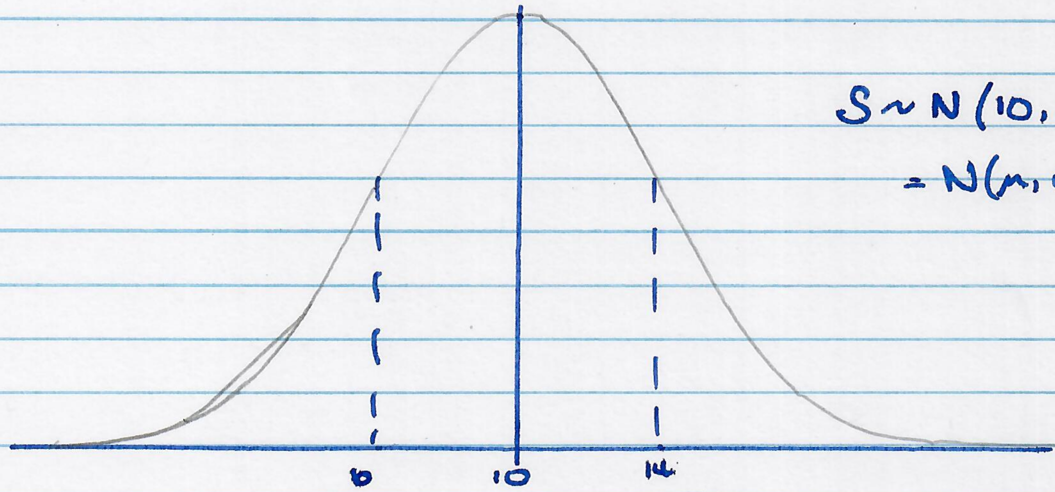
So we need $P(d < 26.666)$ for a to be > 50

– i.e. $P(d = 10$ or $d = 20)$.

$= \frac{600}{137}\left(\frac{1}{10} + \frac{1}{20}\right) = \frac{600}{137}\left(\frac{3}{20}\right)$

$= \frac{90}{137}$ (exact) $= \underline{\underline{0.6569.}}$

5)



$S \sim N(10, 4^2)$
$= N(\mu, \sigma^2)$

Standardise as $\bar{S} = \dfrac{S-\mu}{\sigma} = \dfrac{S-10}{4}$

Then $\bar{S} \sim N(0,1)$

a) for value $S = 15$, $\bar{S} = \dfrac{15-10}{5} = 1.25$

So we need $P(\bar{S} > 1.25)$ which is $\underline{\underline{0.1056}}$
(1- NORM.S.DIST (1.25, ✓))

b) Sample of 20 with mean found to be $\mu' = 11.5$

Is this significant? Use the sample distribution

$$S' = \frac{\mu' - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{11.5-10}{4/\sqrt{20}} = \frac{1.5 \times \sqrt{20}}{4}$$

$$= \underline{\underline{1.6770}}$$

So to test for significance we consider

$P(S' > 1.6770)$

and use this on 2 hypotheses:

H_0 there isn't a problem
H_1 there is a problem with the 10 min claim.

Because the suspicion is known as over-run, we use a 1-tail test with sig. level 5%.

And $P(S' > 1.6770) = \underline{0.0468}$
(1- NORM.S.DIST (1.6770))

This is below 5% so we regard it as significant: there is evidence to support H_1.

c) i) Dentist: $T \sim N(5, 3.5^2)$

Again $\bar{T} \sim N(0,1) = \frac{T-5}{3.5}$

So $P(T<2) = P\left(\bar{T} < \frac{-3}{3.5}\right) = P(\bar{T} < -0.8571)$

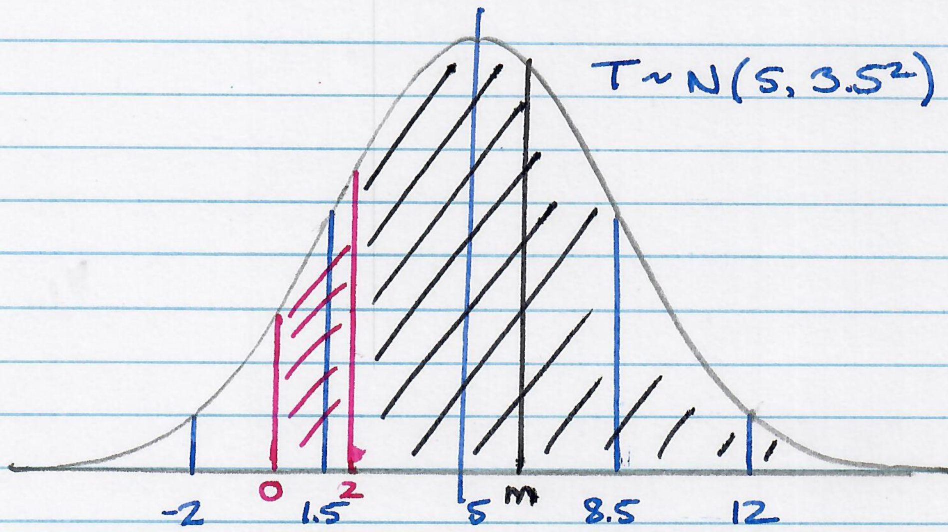
$= \underline{0.1957}$

ii) $P(T<2 \mid T>0)$ Why would this be different?

- Because $T \sim N(5, 3.5^2)$ allows for T to be negative!

so $P(T<0) = P\left(\bar{T} < \frac{0-5}{3.5}\right) = P(\bar{T} < -1.4286)$

$= 0.0778$ (8% or about 1 in 12 - significantly wrong).



So $P(T>0) = 0.9222$

And $P(T>0 \text{ and } T<2) = 0.1957 - 0.0778$

$= 0.1179$ (red area)

So $P(T<2 \mid T>0) = \dfrac{0.1179}{0.9222}$

$= \underline{\underline{0.1278}}$

iii) The current model suggests a significant (8%) number of negative appointment lengths.

d) This is asking what's the median m (0.5 prob) in the area shaded black.

We know $P(T<2) = 0.1957$

so $P(T>2) = 0.8043$

So the median m is where $P(T>m) = \dfrac{0.8043}{2} = 0.4022$

$= \underline{\underline{5.866}}$ (NORM.INV((1-0.4022), 5, 3.5)