Stats 1 - May / June 2008

① a) From calculator: $\Sigma x^2 = 3986$

$a = 53.0673...$ (intercept)

$b = -1.00336...$ (gradient)

$\rightarrow y = 53.07 - 1.00x$

b) $x = 21 \rightarrow y = 53.07 - 1(21) = 32.1$ mins

② a) $P(B) = 160/400$

Tip: Put totals on table to help calculations

b) $P(\text{marker}) = 280/400$

c) $P(B \cup \text{Marker}) = \dfrac{160 + 280 - 119}{400} = \dfrac{321}{400}$

d) $P(G \mid M) = \dfrac{42}{120}$

e) $P(\text{non-}M \mid R) = 21/40$

③ a) From calculator: $\Sigma x^2 = 681.575$

$r = 0.806554...$

b) Strong, positive, linear correlation between lengths and widths of ceramic plates

c) See Mark Scheme

d) A to F: zero linear correlation $\rightarrow r = 0$
G to L: zero linear correlation $\rightarrow r = 0$

④ a) In order: 0 0 13 28 35 40 47 51 63 77 a

Median $= \dfrac{11 + 1}{2} = 6$th value $= 40$

LQ $= \dfrac{11 + 1}{4} = 3$rd value $= 13$

UQ $= \dfrac{3(11 + 1)}{4} = 9$th value $= 63$

IQR $= 63 - 13 = 50$

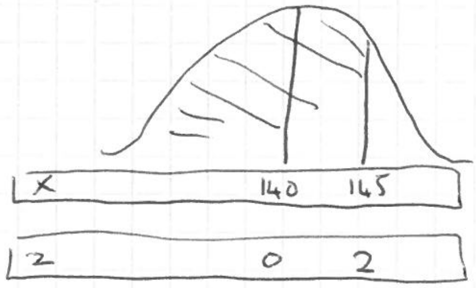
b) i) MODE - gives 0 which is not representative of data

ii) RANGE - cannot calculate as we don't know a

(5) $X \sim N(140, 2.5^2)$

a) i) $P(X < 145)$

$= P\left(Z < \frac{145-140}{2.5}\right)$

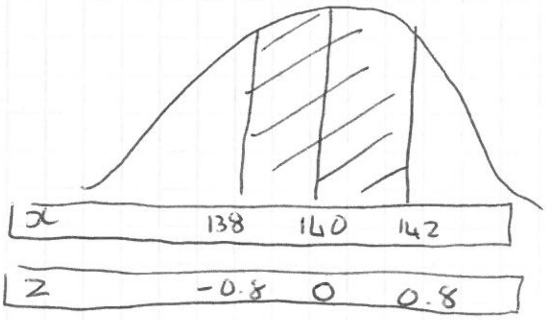


$= P(Z < 2)$

$= 0.97725$

ii) $P(138 < x < 142)$

$= P\left(\frac{138-140}{2.5} < Z < \frac{142-140}{2.5}\right)$



$= P(-0.8 < Z < 0.8)$

$= P(Z < 0.8) - P(Z < -0.8)$

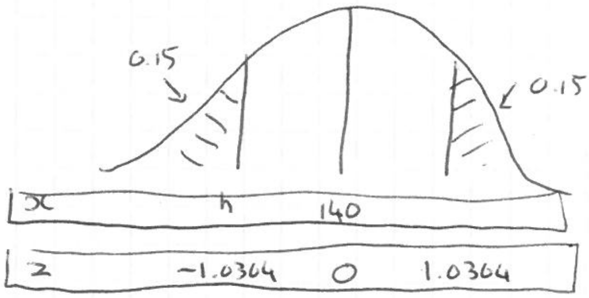
$\downarrow$ 0.78814

$= P(Z > 0.8)$
$= 1 - P(Z < 0.8)$
$= 1 - 0.78814 = 0.21186$

$\rightarrow 0.78814 - 0.21186 = 0.57628$

b) $P(X > h) = 0.85$

Z value for $0.85 = 1.0364$



Standardise:

$\frac{h - 140}{2.5} = -1.0364$

$h = 2.5(-1.0364) + 140$

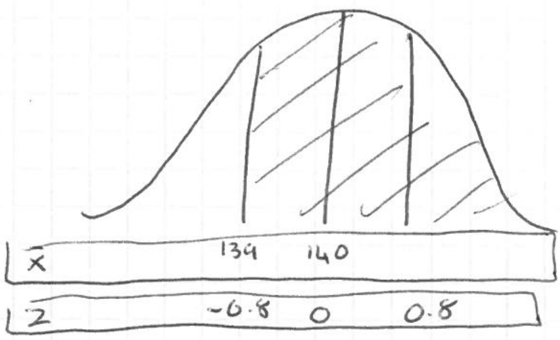
$= 137.409$

c) $\bar{X} \sim N(140, 2.5^2/4)$

$P(\bar{X} > 139)$

$= P\left(Z > \frac{139-140}{2.5/\sqrt{4}}\right)$

$= P(Z > -0.8)$



$= P(Z < 0.8) = 0.7881$

(6) a) $M \sim B(40, 0.35)$

i) $P(M \le 15) = 0.6946$ (from tables)

ii) $P(10 < M < 20)$

can be: 11, 12 .. 19

$\rightarrow P(M \le 19) - P(M \le 10)$

$= 0.9637 - 0.1215 = 0.8422$

b) $W \sim B(10, 0.29)$

$P(W = 3) = {}^{10}C_3 \times 0.29^3 \times 0.71^7$

$= 0.26618...$

c) i) $W_{20} \sim B(20, 0.71)$

MEAN: $np = 20 \times 0.71 = 14.2$

VARIANCE: $np(1-p) = 20 \times 0.71 \times 0.29 = 4.118$

ii) Both the means and the variances are different.

This suggests the claim that they are not random samples is justified.

(7) a) i) From calculator: $\Sigma x^2 = 1452$ $\Sigma x = 190$

$\bar{x} = 1.9$ $s = 3.31966...$

use midpoints for x

ii) 60 mins have been subtracted

$\therefore$ Mean $= 1.9 + 60 = 61.9$

Standard deviation = unchanged = 3.31966..

b) $\bar{x} = 61.9$ $s = 3.31966$ $n = 100$

Z value for 98% (2 tailed) = 2.3263

98% CI for $\mu = \bar{x} \pm z \times \frac{s}{\sqrt{n}}$

$= 61.9 \pm 2.3263 \times \frac{3.31966}{\sqrt{100}}$

$= 61.9 \pm 0.77225\ldots$

$= (61.128, \quad 62.672)$

ii) Mean and Standard Deviation based on grouped data with midpoints used

c) "More than 1 hour on average" : I agree as 60 mins is below lower bound of confidence interval

"More than 1 hour more often than not" : I agree as 74/100 in our sample were more than 60 mins