Stats 1 – June 2005

① a) i) From calculator: $r = 0.7970129....$

$(\Sigma x^2 = 1725)$

ii) Strong, positive, linear correlation between time spent in store & value of items purchased.

b) 0.7970129 (Linear coding has no effect on r)

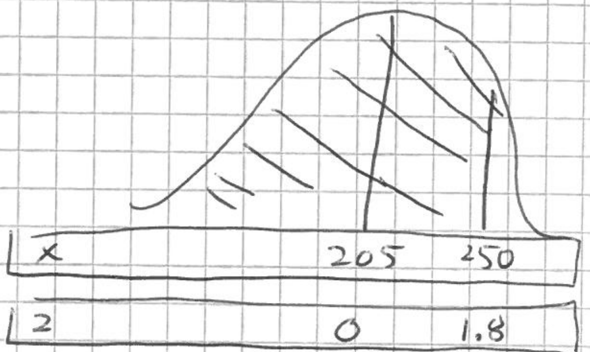
② $X \sim N(205, 25^2)$

a) i) $P(X < 250)$

$= P\left(Z < \frac{250-205}{25}\right)$

$= P(Z < 1.8)$

$= 0.96407$



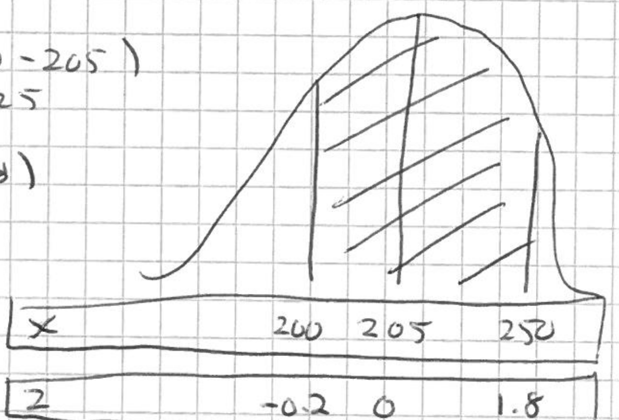
ii) $P(200 < X < 250)$

$= P\left(\frac{200-205}{25} < Z < \frac{250-205}{25}\right)$

$= P(-0.2 < Z < 1.8)$

$= P(Z < 1.8) - P(Z < -0.2)$

$= 0.96407$



$= P(Z > 0.2)$

$= 1 - P(Z < 0.2)$

$= 1 - 0.57926 = 0.42074$

$= 0.96407 - 0.42074 = 0.54333$

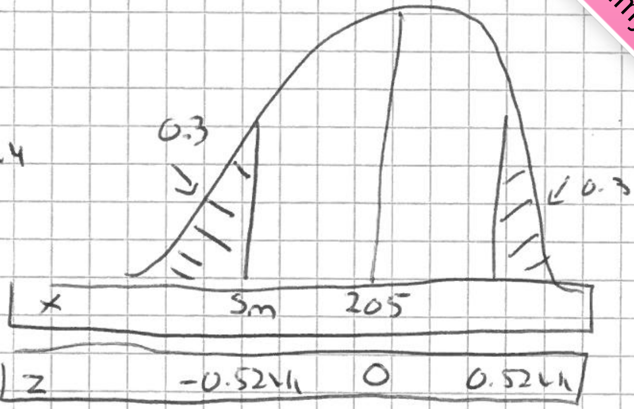
b) $\boxed{\text{SMALL}}$

$P(X < S_m) = 0.3$

Look up 0.7 → Z value of 0.5244

Standardise:

$\frac{S_m - 205}{25} = -0.5244$



$\rightarrow S_m = 25 \times (-0.5244) + 205$

$= 191.89\text{g}$ or 191.9 (1dp)

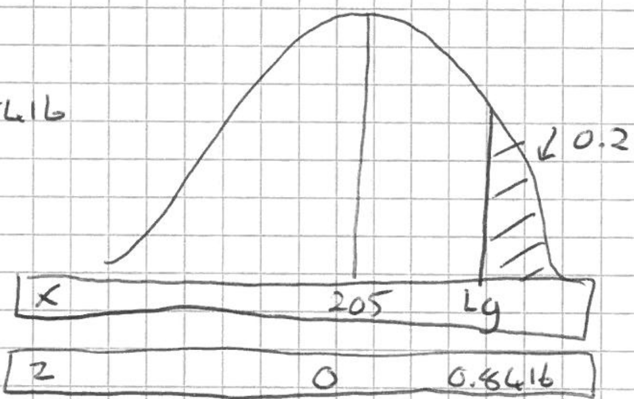
$\boxed{\text{MEDIUM}}$ (find minimum Large weight)

$P(X < L_g) = 0.8$

Look up 0.8 → Z value of 0.8416

Standardise:

$\frac{L_g - 205}{25} = 0.8416$



$\rightarrow L_g = 25 \times 0.8416 + 205$

$= 226.04\text{g}$ or 226.0 (1dp)

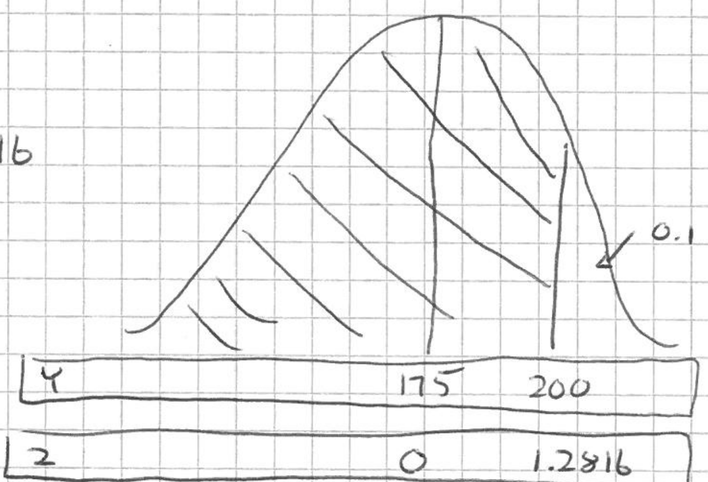
c) $Y \sim N(175, \sigma^2)$

$P(Y < 200) = 0.9$

Look up 0.9 → Z value of 1.2816

Standardise:

$\frac{200 - 175}{\sigma} = 1.2816$



$\rightarrow \sigma = \frac{25}{1.2816}$

$= 19.5068\ldots$

3 a) i) $P(F \cap D) = 0.8 \times 0.9 = 0.72$

ii) $P(F' \cap D') = 0.2 \times 0.6 = 0.12$

b) i) $P(F \cap D \cap M) = 0.72 \times 0.7 = 0.504$

ii) $P(F \cap D \cap M') = 0.72 \times 0.3 = 0.216$

$P(F \cap D' \cap M) = 0.8 \times 0.1 \times 0.7 = 0.056$

$P(F' \cap D \cap M) = 0.2 \times 0.4 \times 0.7 = 0.056$

$\underline{0.328}$

4 a) From calculator:

$a = 1.72555\ldots$ (intercept)

$b = 0.084801\ldots$ (gradient)

$\rightarrow y = 1.726 + 0.0848x$

b) i) Residual = actual y − predicted y

3 $\quad 3.2 - [1.726 + 0.0848(23)]$

$= 3.2 - 3.6764 = -0.4764$

7 $\quad 4.6 - [1.726 + 0.0848(38)]$

$= 4.6 - 4.9484 = -0.3484$

ii) The residuals are small relative to y values

This suggests the linear regression line is appropriate.

c) Total Time = Scan + Transmit = $y + z$

i) $x = 15 \rightarrow [1.726 + 0.0848(15)] + [0.8 + 0.05(15)]$

$= 4.548$

Should be reliable due to small residuals

$x = 75 \rightarrow [1.726 + 0.0848(75)] + [0.8 + 0.05(75)]$

$= 12.636$

Not reliable as $x = 75$ way outside data range & unlikely to get 75 lines of print on A4 page.

(5) a) i) $F \sim B(17, 0.07)$

$P(F=2) = {}^{17}C_2 \times 0.07^2 \times 0.93^{15}$

$= 0.22438$

ii) $F \sim B(50, 0.07)$

$P(F \leq 5) = 0.8650$ (from tables)

b) $F \sim B(50, 0.55)$

$\rightarrow F' \sim B(50, 0.45)$

$P(F \geqslant 30) = P(F' \leq 20)$

F	30	31	32 ...	49	50
F'	20	19	18 ...	1	0

$P(F' \leq 20) = 0.2862$ (from tables)

c) i) $p = 10/50 = 0.2$

ii) variance $= np(1-p)$

$= 50 \times 0.2 \times 0.8 = 8$

$sd = \sqrt{8} = 2.8284...$

iii) recorded sd $= 6.8$, and estimated $= 2.8$

$\therefore$ Not a reliable assumption that X can be modelled by a Binomial Distribution.

(6) a) Need to use midpoints:

x	freq
5.5	13
15.5	33
23	17
28	12
33	8
38	5
45.5	5
75.5	7

From calculator:

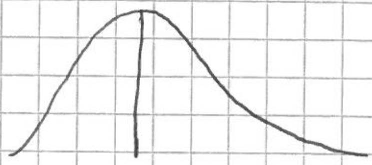
$\Sigma x = 2520$

$\Sigma x^2 = 92907.5$

$\bar{x} = 25.2$

$s = 17.2338...$

b) Data seems to be positively skewed:



OR/ Mean − 2...

= 25.2 − 2 × 17....

= Negative!

c) i) As sample is > 30,
we can use Central Limit Theorem

ii) mean = μ

variance = $\sigma^2/100$

d) $\bar{x} = 25.2$ $s = 17.233...$ $n = 100$

99% z multiplier (2 tailed) = 2.5758

$\Rightarrow \mu = \bar{x} \pm z_m \times \frac{s}{\sqrt{n}}$

$= 25.2 \pm 2.5758 \times \frac{17.233}{\sqrt{100}}$

$= 25.2 \pm 4.4389...$

$= (20.761, 29.639)$

e) $\boxed{>30}$ 30 lies above range of confidence interval

$\therefore$ reject this claim

$\boxed{>50}$ from the data table, $7/100$ or 7% are > 50

$\therefore$ reject this claim as well!