Stats 1 - January 2007

(1) a) From calculator: $\Sigma x = 787$, $\Sigma x^2 = 34023$

$\bar{x} = 39.35$

$s = 12.679344...$

b) 23 people, so median $= \frac{23+1}{2} = 12$th value

$= 42$

Lower Quartile $= \frac{23+1}{4} = 6$th value

$= 31$

Upper Quartile $= \frac{3(23+1)}{4} = 18$th value

$= 55$

$IQR = 55 - 31 = 24$

c) i) Mode does not exist / there is no mode

ii) Cannot calculate Range as we don't know the maximum value

(2) a) $E \sim B(16, 0.45)$

$P(E = 5) = {}^{16}C_5 \times 0.45^5 \times 0.55^{11}$

$= 0.11228...$

b) i) $C \sim B(50, 0.25)$

$P(C \leq 12) = 0.5110$ (from tables)

ii) $N \sim B(50, 0.3)$

$P(10 < N < 20)$

can be: 11, 12 .. 19

$\rightarrow P(N \leq 19) - P(N \leq 10)$

$= 0.9152 - 0.0789 = 0.8363$

c) $B \sim B(40, 0.7)$ [0.45 + 0.25]

Mean $= np = 40 \times 0.7 = 28$

SD $= \sqrt{np(1-p)} = \sqrt{40 \times 0.7 \times 0.3} = 2.8982...$

(3) a) 0.7 (strong positive)

b) 0 (no correlation)

c) -0.7 (strong negative)

(4) a) $\bar{x} = 184$, $s = 32$, $n = 78$

z value for 90% (2 tailed) = 1.6449

90% CI for μ = $\bar{x} \pm z \times \frac{s}{\sqrt{n}}$

$= 184 \pm 1.6449 \times \frac{32}{\sqrt{78}}$

$= 184 \pm 5.9599...$

$= (£178.04,\ £189.96)$

b) i) Should be a valid assumption as performances of this play over a long time period involved

ii) Different plays have different programme prices and audience sizes, so likely to be invalid.

(5) a) $P(D', E', F') = 0.4 \times 0.3 \times 0.2 = 0.024$

b) $P(D', E', F) = 0.4 \times 0.3 \times 0.8 = 0.096$

c) $P(D, E', F') = 0.6 \times 0.3 \times 0.2 = 0.036$

$P(D', E, F') = 0.4 \times 0.7 \times 0.2 = 0.056$ +

$P(D', E', F) = 0.096$

0.188

d) $P(1 \text{ or } 2) = 1 - [P(\text{all}) + P(\text{none})]$

$P(\text{all}) = 0.6 \times 0.7 \times 0.8 = 0.336$

$\therefore P(1 \text{ or } 2) = 1 - [0.336 + 0.024]$

$= 0.64$

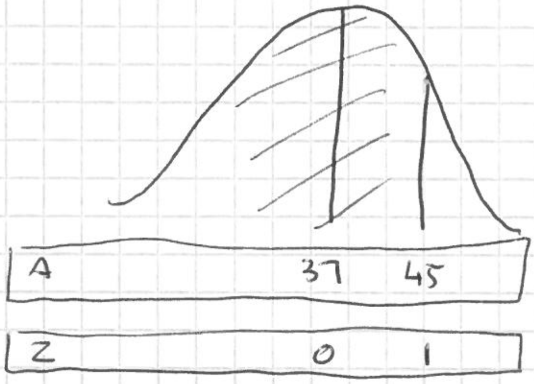
(6) a) $X \sim N(37, 8^2)$

i) $P(X < 45)$

$= P\left(Z < \frac{45 - 37}{8}\right)$

$= P(Z < 1)$

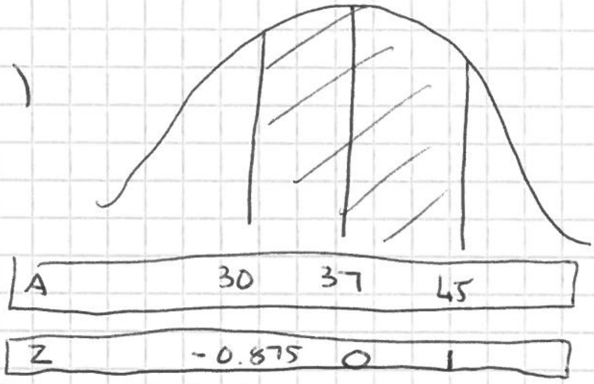
$= 0.84134$



ii) $P(30 < X < 45)$

$= P\left(\frac{30 - 37}{8} < Z < \frac{45 - 37}{8}\right)$

$= P(-0.875 < Z < 1)$



$= P(Z < 1) - P(Z < -0.875)$

↓ 0.84134

$P(Z < -0.875) = P(Z > 0.875)$

$= 1 - P(Z < 0.875)$ Look up 0.88

$= 1 - 0.81057$

$= 0.18943$

$= 0.84134 - 0.18943$

$= 0.65191$

b) $Y \sim N(40, \sigma^2)$

$P(Y > 45) = 0.12$

$\rightarrow P(Y < 45) = 0.88$

Z value for 0.88 $= 1.175$

Standardise:

$\frac{45 - 40}{\sigma} = 1.175$

$5 = 1.175\sigma \rightarrow \sigma = 5/1.175 = 4.2553...$

c) Route A: $P(X < 45) = 0.84134$

Route B: $P(Y < 45) = 1 - 0.12 = 0.88$

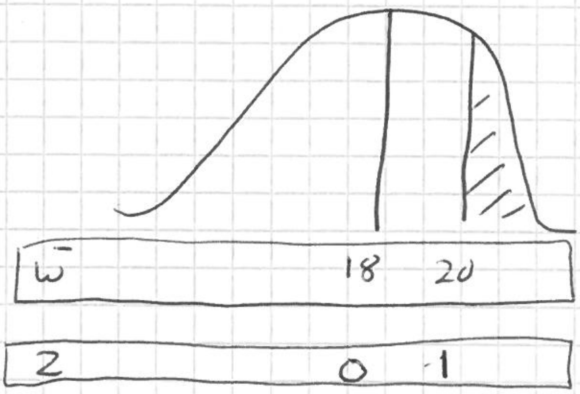
∴ Route B as there is a higher probability of being on time

d) $W \sim N(18, 12^2)$

$\bar{W} \sim N(18, 12^2/36)$

$P(\bar{W} > 20)$

$= P\left(Z > \frac{20-18}{12/\sqrt{36}}\right)$



$= P(Z > 1)$

$= 1 - P(Z < 1)$

$= 1 - 0.84134 = 0.15866$

e) In part d) we do not know if the times are normally distributed, but $n > 30$.
$\therefore$ we use CLT.

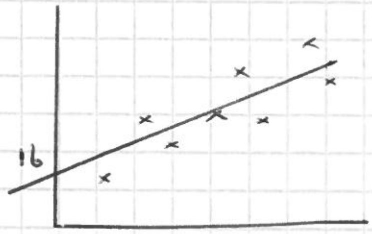
(7) a) See Mark Scheme for Scatter Diagram

b) From calculator: $a = 16.00824...$ (intercept)

$b = 0.11469...$ (gradient)

$\rightarrow y = 16.008 + 0.115x$

c) Residual =



y_H − Predicted$_H$

$= 70 - [16.008 + 0.115(480)]$

$= 70 - 71.208 = -1.2$

This is indicated on diagram as H is just below the regression line.

d) $x = 560 \rightarrow y = 16.008 + 0.115(560)$

$= 80.408$ minutes

convert to hours: $80.408 \div 60 = 1.3401...$

£12 per hour $\rightarrow$ 1.3401×12

$=$ £16.08 charge.