

Stats 1 - June 2005

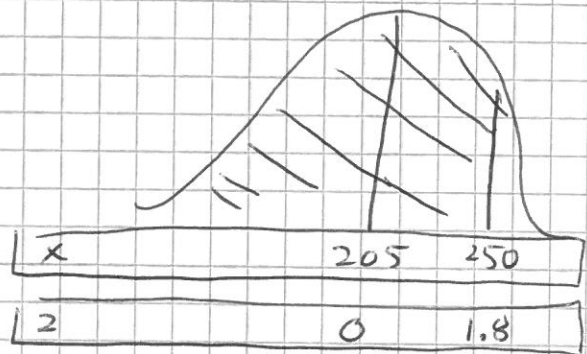
① a) i) From calculator: $r = 0.7970129 \dots$
($2x^2 = 1725$)

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

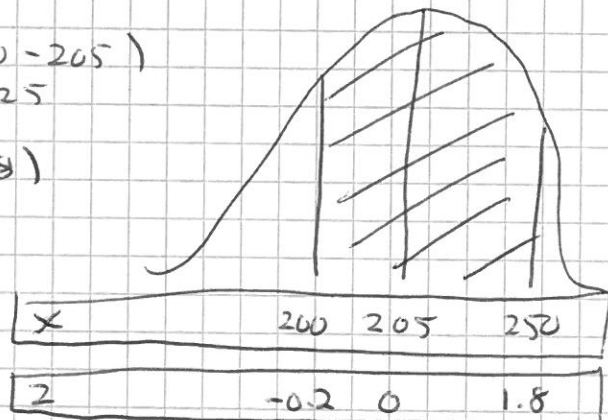
b) 0.7970129 (Linear coding has no effect on r)

② $X \sim N(205, 25^2)$

a) i) $P(X < 250)$
 $= P(Z < \frac{250 - 205}{25})$
 $= P(Z < 1.8)$
 $= 0.96407$



ii) $P(200 < X < 250)$
 $= P(\frac{200 - 205}{25} < Z < \frac{250 - 205}{25})$
 $= P(-0.2 < Z < 1.8)$
 $= P(Z < 1.8) - P(Z < -0.2)$
 $= 0.96407$



$= P(Z > 0.2)$
 $= 1 - P(Z < 0.2)$
 $= 1 - 0.57926 = 0.42074$

$= 0.96407 - 0.42074 = 0.54333$

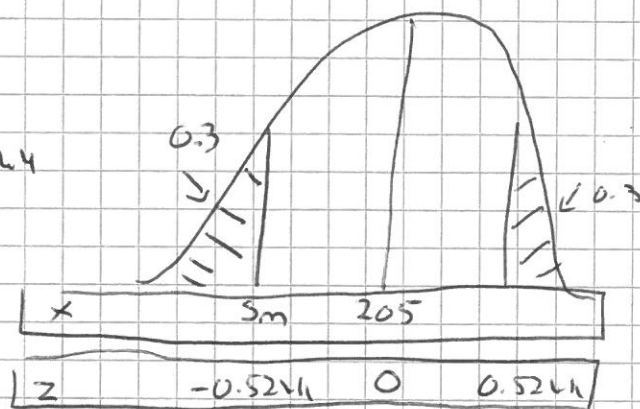
b) SMALL

$$P(X < S_m) = 0.3$$

Look up 0.7 → Z value of 0.5244

Standardize:

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



$$\begin{aligned} \rightarrow S_m &= 25 \times (-0.5244) + 205 \\ &= 191.89 \text{ g} \quad \text{or } 191.9 \text{ (1dp)} \end{aligned}$$

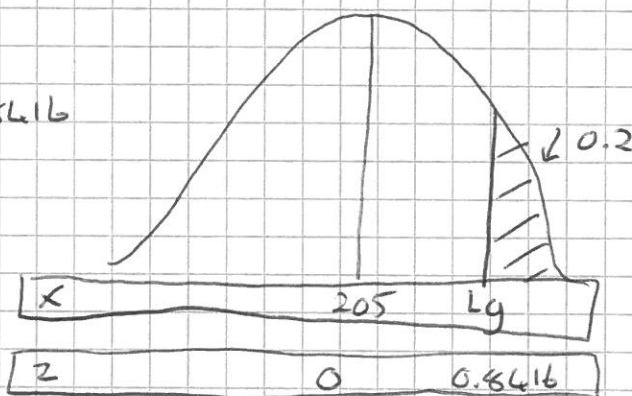
MEDIUM (find minimum large weight)

$$P(X < L_g) = 0.8$$

Look up 0.8 → Z value of 0.8416

Standardize:

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



$$\begin{aligned} \rightarrow L_g &= 25 \times 0.8416 + 205 \\ &= 226.04 \text{ g} \quad \text{or } 226.0 \text{ (1dp)} \end{aligned}$$

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

$$P(Y < 200) = 0.9$$

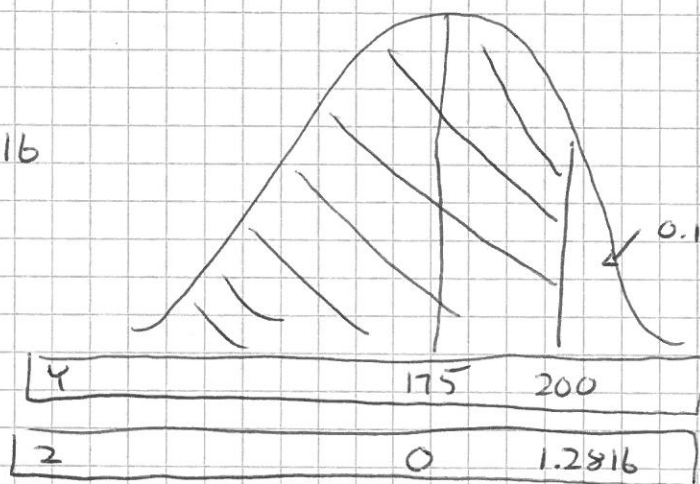
Look up 0.9 → Z value of 1.2816

Standardize:

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

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

$$= 19.5068 \dots$$



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

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

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

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

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

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

$$0.328$$

(4) a) From calculator:

$$a = 1.72555 \dots \quad (\text{intercept})$$

$$b = 0.084801 \dots \quad (\text{gradient})$$

$$\rightarrow y = 1.726 + 0.0848x$$

b) i) Residual = actual y - predicted y

$$\begin{aligned} (3) \quad 3.2 - [1.726 + 0.0848(23)] \\ = 3.2 - 3.6764 = -0.4764 \end{aligned}$$

$$\begin{aligned} (7) \quad 4.6 - [1.726 + 0.0848(38)] \\ = 4.6 - 4.9484 = -0.3484 \end{aligned}$$

ii) The residuals are small relative to y values

This suggests the linear regression line is appropriate.

c) Total Time = Scan + Transmit = $y + 2$

$$\begin{aligned} i) \quad (50=15) \rightarrow [1.726 + 0.0848(15)] + [0.8 + 0.05(15)] \\ = 4.548 \end{aligned}$$

Should be reliable due to small residuals

$$\begin{aligned} (50=75) \rightarrow [1.726 + 0.0848(75)] + [0.8 + 0.05(75)] \\ = 12.636 \end{aligned}$$

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

$$P(F \leq 5) = 0.8650 \quad (\text{from tables})$$

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

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

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

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

$$P(F' \leq 20) = 0.2862 \quad (\text{from tables})$$

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) rounded sd = 0.8, and estimated = 2.8

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

⑥ a) Need to use midpoints:

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

From calculator:

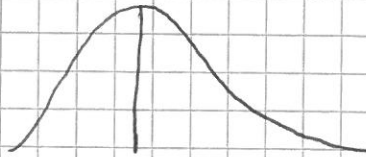
$$\sum x = 2520$$

$$\sum x^2 = 92907.5$$

$$\bar{x} = 25.2$$

$$s = 17.2338...$$

b) Data seems to be positively skewed.



$$\begin{aligned} \text{OR/ Mean} &= 2 \times \text{SD} \\ &= 25.2 - 2 \times 17.233... \\ &= \text{Negative!} \end{aligned}$$

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

$$\begin{aligned} \text{ii) mean} &= \mu \\ \text{variance} &= \frac{\sigma^2}{100} \end{aligned}$$

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

$$99\% \text{ z multiplier (2 tailed)} = 2.5758$$

$$\begin{aligned} \rightarrow N &= \bar{x} \pm z_{\alpha} \times \frac{s}{\sqrt{n}} \\ &= 25.2 \pm 2.5758 \times \frac{17.233}{\sqrt{100}} \\ &= 25.2 \pm 4.4389... \\ &= (20.761, 29.639) \end{aligned}$$

e) $\boxed{730}$ 30 lies above range of confidence interval

$\therefore$ reject H_0 claim

$\boxed{750}$ from the data table, $\frac{7}{100}$ or 7% are > 50

$\therefore$ reject H_0 claim as well.