

# Jan 08 S2 Worked answers

I (i)  $x$  is independent,  $y$  is dependent

because growth ( $y$ ) is affected by hormone concentration ( $x$ ), not vice versa

$$(ii) \bar{x} = \frac{\sum x}{n} = \frac{30}{12} = 2.5, \bar{y} = \frac{967.6}{12} = 80.63$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 90 - \frac{30^2}{12} = 15 \quad (\text{or } \sum x^2 - n\bar{x}^2 = 90 - 12(2.5)^2)$$

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 2530.3 - \frac{(30)(967.6)}{12} = 111.3 \quad (\text{or } \sum xy - n\bar{x}\bar{y})$$

$$\Rightarrow b = \frac{S_{xy}}{S_{xx}} = \frac{111.3}{15} = 7.42$$

regression line is

$$y - \bar{y} = b(x - \bar{x}) \Rightarrow y - 80.63 = 7.42(x - 2.5)$$

$$\Rightarrow \underline{y = 7.42x + 62.08}$$

$$(iii) x = 1.2 \Rightarrow y = 7.42 \times 1.2 + 62.08 = \underline{71.0}$$

$$x = 4.3 \Rightarrow y = 7.42 \times 4.3 + 62.08 = \underline{94.0}$$

$x = 1.2$  is within the range for  $x$ , and values for  $x = 1$  are very close together, so it is probably quite a reliable estimate

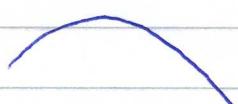
$x = 4.3$  is outside the range (though only just), and the values of  $y$  for  $x = 4$  are more scattered, so it might be a less reliable estimate.

$$(iv) x = 3 \Rightarrow y = 7.42 \times 3 + 62.08 = 84.34$$

$$\Rightarrow \text{residual} = 80 - 84.34 = \underline{-4.34}$$

(v) Clearly, this does not fit what we'd expect from the regression line. It may be that the linear relationship only works for lower concentrations, or that the actual relationship is not linear - it may be a curve

e.g.



2 (i)  $X \sim B(94, 0.1)$   $X = \text{no. of no shows in one night}$

(ii)  $n$  large,  $p$  small

$$(iii) \lambda = np = 94 \times 0.1 = 9.4 \Rightarrow X \sim P_0(9.4)$$

$$(A) P(X=4) = e^{-9.4} \frac{9.4^4}{4!} = \underline{0.0269}$$

$$(B) \text{ want } P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.0160 = \underline{0.984}$$

$$(iv) (0.984)^3 = \underline{0.607}$$

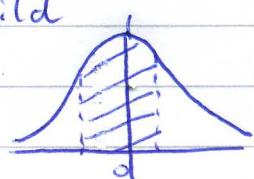
(v) (A)  $X \sim B(2914, 0.1)$   $X = \text{no. of no shows in August}$

(B) Need to use Normal approximation i.e.  $Y \sim N(np, npq)$  ( $p=0.1, q=0.9$ )  
 $\Rightarrow Y \sim N(291.4, 262.26)$

$$P(X \leq 300) \approx P(Y < 300.5) = P\left(Z < \frac{300.5 - 291.4}{\sqrt{262.26}}\right)$$

$$= P(Z < 0.562) = \underline{0.713}$$

3.(i)  $X \sim N(56, 6.5^2)$   $X = \text{DBP of randomly chosen child}$

$$\begin{aligned} P(52.5 < X < 57.5) &= P\left(\frac{52.5-56}{6.5} < Z < \frac{57.5-56}{6.5}\right) \\ &= P(-0.538 < Z < 0.231) = \end{aligned}$$


$$= 0.5914 - (1 - 0.7046)$$

$$= \underline{0.296}$$

(ii)  $P(X < 62) = P(Z < \frac{62-56}{6.5}) = P(Z < 0.923) =$



$$= 0.8220$$

$Y \sim N(68, 10^2)$ ,  $Y = \text{DBP of randomly chosen adult}$

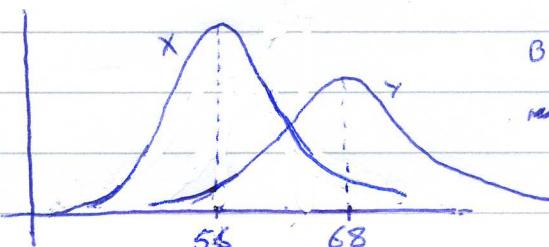
$$\Rightarrow P(Y < 62) = P(Z < \frac{62-68}{10}) = P(Z < -0.6) =$$


$$= 1 - 0.7257 = 0.2743$$

$$\begin{aligned} \therefore \text{ans} &= P(X > 62) \times P(Y < 62) + P(X < 62) \times P(Y > 62) \\ &= (1 - 0.822) \times 0.2743 + 0.8220 \times (1 - 0.2743) \\ &= 0.1780 \times 0.2743 + 0.8220 \times 0.7257 \end{aligned}$$

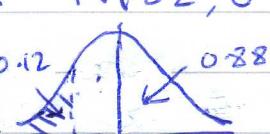
$$= \underline{0.645}$$

(iii)



Because of smaller  $\sigma$ ,  $X$  must have higher relative distribution to keep areas equal

(iv)  $P(X < 62) = 0.12$ ,  $X \sim N(82, \sigma^2)$

$$\Rightarrow P\left(Z < \frac{62-82}{\sigma}\right) = 0.12 \Rightarrow$$


$$\Rightarrow \frac{-20}{\sigma} = -1.175 \Rightarrow \sigma = \frac{20}{1.175} = \underline{17.0}$$

Jan 08 S2

4.(a)  $H_0$ : no association between sex and subject     $H_1$ : some association between sex & subject

O					E
38	19	6	32	95	30.4
42	55	9	49	155	49.6
80	74	15	81	250	57.7

$$\frac{(O-E)^2}{E} \quad 1.9 \quad 2.958 \quad 0.016 \quad 0.048 \\ \sum \frac{(O-E)^2}{E} \quad 1.165 \quad 1.813 \quad 0.010 \quad 0.030$$

$$\Rightarrow \chi^2 = \sum \frac{(O-E)^2}{E} = 7.94$$

$$v = (4-1)(2-1) = 3 \Rightarrow \text{critical value} = 7.815$$

since  $7.94 > 7.815$ , we reject  $H_0$

$\therefore$  sufficient evidence of an association between the student's sex and the choice of subjects

$$(b) \quad H_0: \mu = 67.4 \quad \text{if } H_0 \text{ is true, } \bar{X} \sim N(67.4, \frac{8.9^2}{12}) \quad \mu = \text{population mean for all students taught by new method}$$

$$\bar{X} = 68.3 \Rightarrow P(\bar{X} > 68.3) = P\left(Z > \frac{68.3 - 67.4}{\sqrt{8.9/12}}\right)$$

$$= P(Z > 0.350)$$

$$= 1 - 0.6368 = 0.363 \quad (\rightarrow 0.05)$$

$\therefore$  accept  $H_0$

There is not sufficient evidence that the new method has been successful.