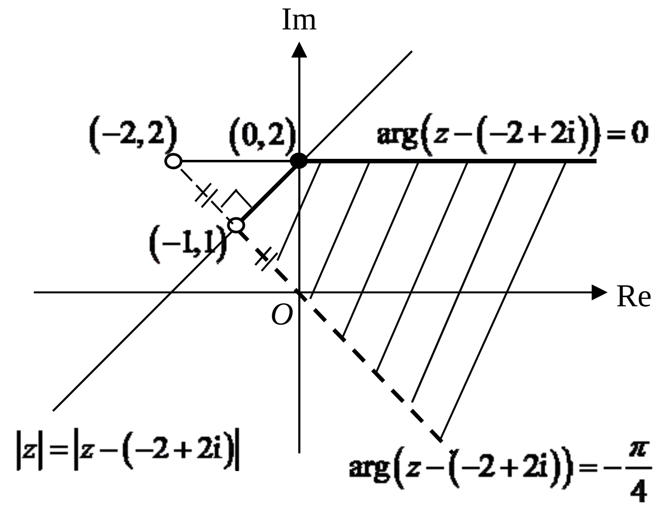
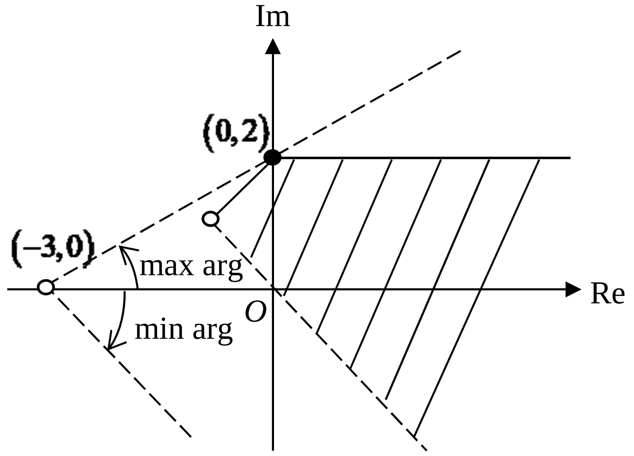
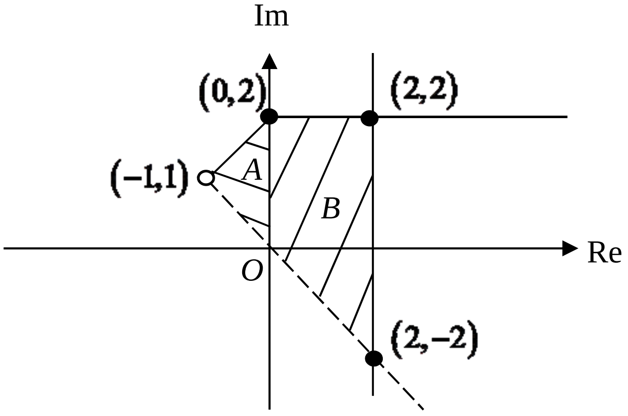
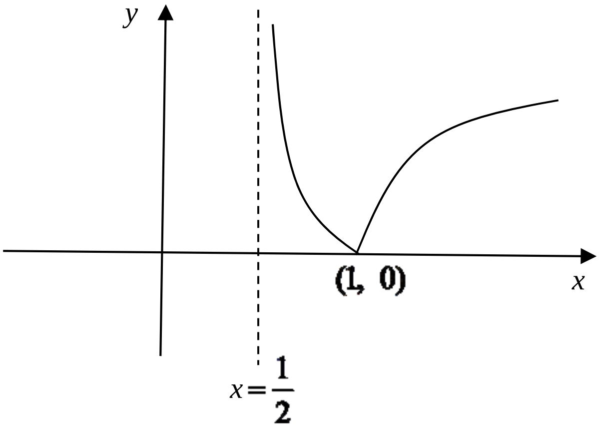
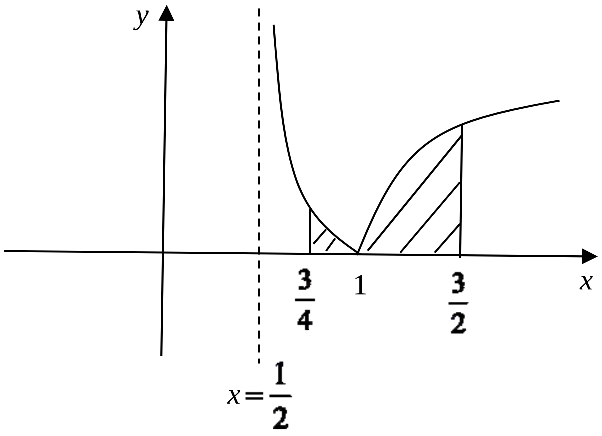
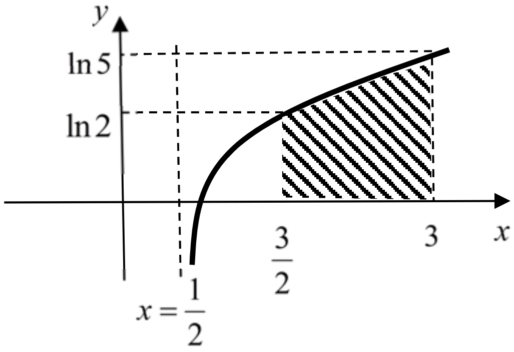
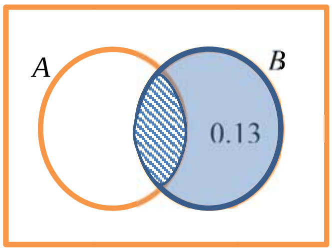
2015 H2 Mathematics Prelim Paper 2 Solutions

Qn	Solutions
1. **(a)**	Locus of *R* is a straight line passing through the origin and parallel to $\overrightarrow{AB}$.
1. **(b)**	$\mathbf{a}$ and $\mathbf{b}$ are parallel $\mathbf{a}=\left(2\mathbf{a}\cdot\mathbf{b}\right)\mathbf{b}$ $\lvert\mathbf{a}\rvert=2\lvert\mathbf{a}\rvert\lvert\mathbf{b}\rvert^2\lvert\cos\theta\rvert$ $\lvert\mathbf{b}\rvert=\dfrac{1}{\sqrt{2}}$
1. **(c)**	Using ratio theorem, $\overrightarrow{ON}=\dfrac{1}{3}\mathbf{a}+\dfrac{2}{3}\mathbf{b}$ and $\overrightarrow{OM}=\dfrac{1}{2}\left(\mathbf{b}+\mathbf{c}\right)$ Since *O* is the midpoint of *MN*, $\overrightarrow{ON}=-\overrightarrow{OM}$ $\dfrac{1}{3}\mathbf{a}+\dfrac{2}{3}\mathbf{b}=-\dfrac{1}{2}\left(\mathbf{b}+\mathbf{c}\right)$ $2\mathbf{a}+4\mathbf{b}=-3\mathbf{b}-3\mathbf{c}$ $2\mathbf{a}+7\mathbf{b}+3\mathbf{c}=\mathbf{0}$
2. **(i)**	$\mathrm{f}\left(r\right)=\left(3r-2\right)\left(3r+1\right)=9r^2-3r-2$ $\displaystyle\sum_{r=1}^{n}\mathrm{f}\left(r\right)=\sum_{r=1}^{n}\left(9r^2-3r-2\right)$ $\displaystyle=9\sum_{r=1}^{n}r^2-3\sum_{r=1}^{n}r-\sum_{r=1}^{n}2$ $=\dfrac{3}{2}n\left(n+1\right)\left(2n+1\right)-\dfrac{3}{2}n\left(n+1\right)-2n$ $=3n^3+3n^2-2n$ $=n\left(3n^2+3n-2\right)$ Hence, $a=3$, $b=3$ and $c=-2$.
2. **(ii)**	$S_1=\dfrac{1}{\left(3-2\right)\left(3+1\right)}=\dfrac{1}{4}$ $S_2=\dfrac{1}{4}+\dfrac{1}{\left(6-2\right)\left(6+1\right)}=\dfrac{2}{7}\ \left(\text{qed}\right)$ $S_3=\dfrac{2}{7}+\dfrac{1}{\left(9-2\right)\left(9+1\right)}=\dfrac{3}{10}$ $S_4=\dfrac{3}{10}+\dfrac{1}{\left(12-2\right)\left(12+1\right)}=\dfrac{4}{13}$

Qn	**Solutions**
2. **(iii)**	Conjecture: $S_n = \dfrac{n}{3n+1}$ Let $\mathrm{P}(n)$ be the proposition that "$S_n = \dfrac{n}{3n+1}$ for all $n \in \mathbb{Z}^+$". Consider $\mathrm{P}(1)$, $\text{LHS} = S_1 = \dfrac{1}{4} = \text{RHS}$. Therefore $\mathrm{P}(1)$ is true. Assume $\mathrm{P}(k)$ is true for some $k \in \mathbb{Z}^+$, i.e., $S_k = \dfrac{k}{3k+1}$. Want to show $\mathrm{P}(k+1)$ is true, i.e., $S_{k+1} = \dfrac{k+1}{3(k+1)+1}$. Consider $\mathrm{P}(k+1)$, $S_{k+1} = \displaystyle\sum_{r=1}^{k+1} \frac{1}{\mathrm{f}(r)}$ $= \dfrac{k}{3k+1} + \dfrac{1}{\big(3(k+1)-2\big)\big(3(k+1)+1\big)}$ $= \dfrac{k(3k+4)+1}{(3k+1)(3k+4)}$ $= \dfrac{(3k+1)(k+1)}{(3k+1)(3k+4)}$ $= \dfrac{k+1}{3(k+1)+1}$ $\therefore \mathrm{P}(k+1)$ is true. Since $\mathrm{P}(1)$ is true and $\mathrm{P}(k)$ is true $\Rightarrow \mathrm{P}(k+1)$ is true, by mathematical induction, $\mathrm{P}(n)$ is true for all $n \in \mathbb{Z}^+$.
3. **(i)**	$\dfrac{3\pi}{4} < \arg(-2+2\mathrm{i}-z) \le \pi$ $\dfrac{3\pi}{4} < \arg(-1) + \arg(z+2-2\mathrm{i}) \le \pi$ $-\dfrac{\pi}{4} < \arg\big(z-(-2+2\mathrm{i})\big) \le 0$ **-----(*)**

Qn	Solutions
	
3. (ii)	 $\max\arg(z+3)=$ angle of $(0,2)$ from $(-3,0)=\tan^{-1}\left(\frac{2}{3}\right)$ $-\frac{\pi}{4}<\arg(z+3)\le\tan^{-1}\left(\frac{2}{3}\right)$
3. (iii)	 Area = triangle *A* + trapezium *B* $=\frac{1}{2}(1)(2)+\frac{1}{2}(2)(2+4)=7$

Qn	Solutions
4. **(i)**	$\int \ln\left(2x-1\right)\mathrm{d}x$ $= x\ln\left(2x-1\right)-\int x\frac{2}{2x-1}\,\mathrm{d}x$ $= x\ln\left(2x-1\right)-\int 1+\frac{1}{2x-1}\,\mathrm{d}x$ $= x\ln\left(2x-1\right)-x-\frac{1}{2}\ln\left(2x-1\right)+c$ $= \left(x-\frac{1}{2}\right)\ln\left(2x-1\right)-x+c$
4. **(ii)**	
4. **(iii)**	 Area of shaded region $=\int_{3/4}^{1}-\ln(2x-1)\mathrm{d}x+\int_{1}^{3/2}\ln(2x-1)\mathrm{d}x$ $=-\left[x\ln(2x-1)-x-\frac{1}{2}\ln(2x-1)\right]_{3/4}^{1}$ $+\left[x\ln(2x-1)-x-\frac{1}{2}\ln(2x-1)\right]_{1}^{3/2}$

Qn	Solutions
	$=-\left[(0-1-0)-\left(\frac{3}{4}\ln\frac{1}{2}-\frac{3}{4}-\frac{1}{2}\ln\frac{1}{2}\right)\right]$ $+\left[\left(\frac{3}{2}\ln 2-\frac{3}{2}-\frac{1}{2}\ln 2\right)-(0-1-0)\right]$ $=\frac{3}{4}\ln 2-\frac{1}{4}$ square units
4. (iv)	 Volume = $\pi(3)^2(\ln 5)-\pi\left(\frac{3}{2}\right)^2\ln 2-\pi\int_{\ln 2}^{\ln 5} x^2\,\mathrm{d}y$ $=\pi(3)^2(\ln 5)-\pi\left(\frac{3}{2}\right)^2\ln 2-\pi\int_{\ln 2}^{\ln 5}\left(\frac{\mathrm{e}^y+1}{2}\right)^2\mathrm{d}y$ = 26.927 = 26.9 cubic units (3s.f.)
5(i)	In order to use stratified sampling, we need to know the complete composition of the participants according to strata, e.g. race, gender or age-groups. However, it is difficult to have this complete information due to absentees, incomplete registration information, etc. Hence it would be difficult to use a stratified sample.
5(ii)	We assign an index number to each participant when they arrive at the conference. Let the total number of participants to arrive at the conference be *x*. Sampling interval = $\frac{x}{0.04x}=25$. Pick a random number from 1 to 25, then pick every 25th participant thereafter. E.g. If the number 5 was selected, we sample the participants with numbers 5, 30, 55, … until 4% of the participants are sampled.
6 (ai)	$\mathrm{P}(A\mid B)=0.675=\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}$ $\Rightarrow \mathrm{P}(A\cap B)=0.675\left(\mathrm{P}(A\cap B)+0.13\right)$ $\Rightarrow \mathrm{P}(A\cap B)=0.675\mathrm{P}(A\cap B)+0.08775$ $\Rightarrow \mathrm{P}(A\cap B)=0.27$ 

Qn	Solutions
	Alternative 1 $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$ $0.62 = 0.49 + \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(A \mid B)} - \mathrm{P}(A \cap B)$ $\Rightarrow \mathrm{P}(A \cap B) = 0.27$ **Alternative 2** $\mathrm{P}(A' \mid B) = \dfrac{\mathrm{P}(A' \cap B)}{\mathrm{P}(B)} \Rightarrow \mathrm{P}(B) = 0.4$ $\mathrm{P}(A \cap B) = \mathrm{P}(B)\mathrm{P}(A \mid B) = 0.27$
6 (aii)	$\mathrm{P}(A \mid B) = 0.675$ $\mathrm{P}(A) = 1 - 0.13 - 0.38 = 0.49$ $\mathrm{P}(A \mid B) \neq \mathrm{P}(A)$ Therefore the two events *A* and *B* are not independent Or $\mathrm{P}(A \cap B) = 0.27$ $\mathrm{P}(A) \times \mathrm{P}(B) = 0.49 \times 0.4 = 0.196$ $\mathrm{P}(A \cap B) \neq \mathrm{P}(A) \times \mathrm{P}(B)$ Therefore the two events *A* and *B* are not independent
6 (b)	Required Probability $= \mathrm{P}(CC) + \mathrm{P}(CDCC) + \mathrm{P}(CDCDCC) + \ldots + \mathrm{P}(DCC) + \mathrm{P}(DCDCC) + \mathrm{P}(DCDCDCC) + \ldots$ $= p^2 + p^2(1-p)^2 + p^2(1-p)^4 + \ldots. + (1-p)^2 p + (1-p)^4 p + (1-p)^6 p + \ldots.$ $= p^2\left[\dfrac{1}{1-(1-p)^2}\right] + (1-p)^2 p\left[\dfrac{1}{1-(1-p)^2}\right]$ $= \dfrac{p\left[p + (1-p)^2\right]}{2p - p^2}$ $= \dfrac{1 - p + p^2}{2 - p}$ (shown)
7(a)	$9! \times {}^{10}C_3 \times 3! = 261273600$

Qn **Solutions**

7(b)

[All Universities]=[Any six-person]
– [University *A*&*B*]
– [University *B*&*C*]
– [University *A*&*B*]

Case 1: University *A*&*B* ${}^{9}C_{6}$

Case 2: University *B*&*C* ${}^{8}C_{6}$

Case 3: University *A*&*C* ${}^{7}C_{6}$

So the answer is ${}^{12}C_{6} - {}^{9}C_{6} - {}^{8}C_{6} - {}^{7}C_{6} = 805$

<u>Alternate Method</u> [Not advisable because of too many cases]
Consider number of candidates from each university.

If we have x from A, y from B and z from C, the number of possible ways is ${}^{4}C_{x}\,{}^{5}C_{y}\,{}^{3}C_{z}$

A(4)	1	1	1	2	2	2	3	3	4
B(5)	2	3	4	1	2	3	1	2	1
C(3)	3	2	1	3	2	1	2	1	1
Ans	40	120	60	30	180	180	60	120	15

So the total number of possible way is 805.

7(c)

All possible groupings: $\dfrac{{}^{12}C_{4}\,{}^{8}C_{4}\,{}^{4}C_{4}}{3!} = 5775$

<u>Method 1 (Direct Method)</u>

Case 1: Candidates from University *A* grouped in 2,2,0
(4 from *A* versus 8 from ‘the rest’)
Number of ways = ${}^{4}C_{2}\,{}^{8}C_{2} \times {}^{2}C_{2}\,{}^{6}C_{2} \div 2! = 1260$

Case 2: Candidates from University *A* grouped in 2,1,1
Number of ways = ${}^{4}C_{2}\,{}^{8}C_{2} \times {}^{2}C_{1}\,{}^{6}C_{3} \div 2! = 3360$

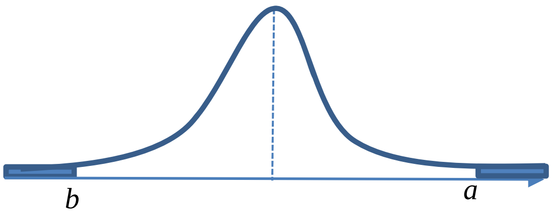
So the probability is $\dfrac{1260+3360}{5775} = 0.8$

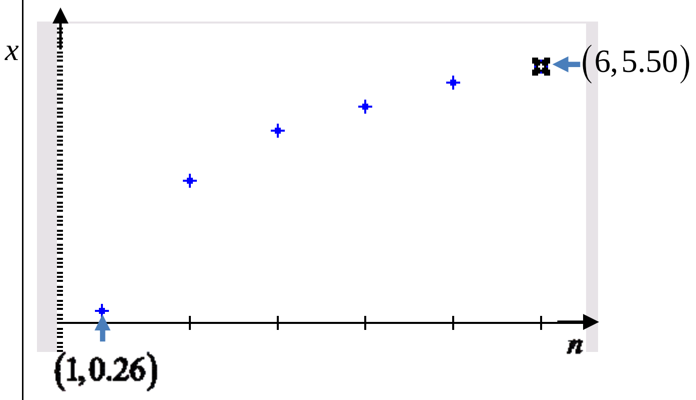
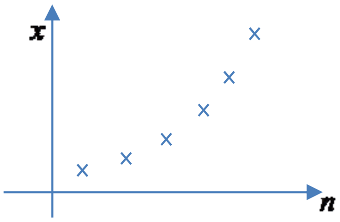
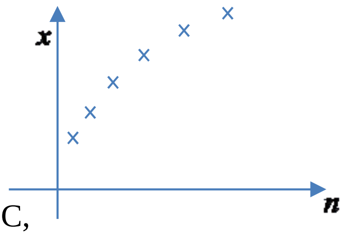
<u>Method 2 (Method of Complementation –cases where a group has more than 2 candidates from University *A*)</u>

Case 1: Candidates from University *A* grouped in 3,1,0
Number of ways = ${}^{4}C_{3}\,{}^{8}C_{1} \times {}^{8}C_{4} \div 2! = 1120$

Case 2: Candidates from University *A* grouped in 4,0,0
Number of ways = ${}^{4}C_{4} \times {}^{8}C_{4} \div 2! = 35$

So the probability is $1 - \dfrac{1120+35}{5775} = 0.8$

8(a)	$2X-Y\sim \mathrm{N}\left(2\mu_1-\mu_2,2^2\left(4\sigma^2\right)+9\sigma^2\right)$ i.e. $2X-Y\sim \mathrm{N}\left(2\mu_1-\mu_2,25\sigma^2\right)$ $9X-8Y\sim \mathrm{N}\left(9\mu_1-8\mu_2,9^2\left(4\sigma^2\right)+8^2\left(9\sigma^2\right)\right)$ i.e. $9X-8Y\sim \mathrm{N}\left(9\mu_1-8\mu_2,900\sigma^2\right)$ $\mathrm{P}(2X>Y)=\mathrm{P}(9X<8Y)$ $\mathrm{P}(2X-Y>0)=\mathrm{P}(9X-8Y<0)$ $\Rightarrow \mathrm{P}\left(\frac{(2X-Y)-(2\mu_1-\mu_2)}{\sqrt{25\sigma^2}}>\frac{0-(2\mu_1-\mu_2)}{\sqrt{25\sigma^2}}\right)$ $=\mathrm{P}\left(\frac{(9X-8Y)-(9\mu_1-8\mu_2)}{\sqrt{900\sigma^2}}>\frac{0-(9\mu_1-8\mu_2)}{\sqrt{900\sigma^2}}\right)$ $\Rightarrow \mathrm{P}\left(Z>\frac{0-(2\mu_1-\mu_2)}{\sqrt{25\sigma^2}}\right)=\mathrm{P}\left(Z<\frac{0-(9\mu_1-8\mu_2)}{\sqrt{900\sigma^2}}\right)$ $\Rightarrow \mathrm{P}\left(Z>\frac{\mu_2-2\mu_1}{5\sigma}\right)=\mathrm{P}\left(Z<\frac{8\mu_2-9\mu_1}{30\sigma}\right)$  By symmetry of the standard normal distribution, So $\frac{\mu_2-2\mu_1}{5\sigma}=-\frac{8\mu_2-9\mu_1}{30\sigma}$ $\Rightarrow 6\mu_2-12\mu_1=-8\mu_2+9\mu_1$ $\Rightarrow 14\mu_2=21\mu_1$ $\therefore \frac{\mu_1}{\mu_2}=\frac{14}{21}=\frac{2}{3}$
8 (bi)	$Y\sim \mathrm{N}(10,9)$ $\mathrm{P}(Y<16)=0.97725=0.977$ (3 sf)
8 (bii)	Let S be the number of observations with $Y\geq 16$ Hence $S\sim \mathrm{B}(100,0.02275)$ Since $n=100$ is sufficiently large, $np=2.275<5$, $S\sim \mathrm{Po}(2.275)$ approximately $\therefore \mathrm{P}(100-S\geq 95)=\mathrm{P}(S\leq 5)=0.971$ (3 s.f.)

9(i)	$\bar{x}=\dfrac{17.21+k}{6}$, $\bar{n}=3.5$ $(\bar{n},\bar{x})$ lies on the regression line, so $x=0.943n+0.484$ $\dfrac{17.21+k}{6}=0.943(3.5)+0.484$ $k=5.50(2\text{ dp})$
9(ii)	
9 (iii)	From the scatter plot, x and n have a curvilinear relationship. Therefore **a linear model** is inappropriate even though the product moment correlation coefficient is relatively high (i.e. 0.915).
9 (iv)	The graph of $x=a+bn^2$ is concave upwards (or increase at increasing rate) similar to the scatter plot  The graph of $x=a+b\ln n$ is concave downwards (or increase at decreasing rate) .  From GC, $r_B=0.8212061688\approx 0.821(3\text{ s.f.})$ $r_A=0.9859197289\approx 0.986(3\text{ s.f.})$ which is closer to 1 and hence suggested a relatively stronger linear relationship between x and $\ln n$ as compared to x and n^2. Therefore **Model (A)** is more appropriate.

9 **(iv)**	Using GC, the regression line of x on $\ln n$ is $x = 0.638570346 + 2.869411323 \ln n$ $b = 2.869411323 \approx 2.87$ $a = 0.638570346 \approx 0.639$ For $x \geq 10$, $2.869411323 \ln n + 0.638570346 \geq 10$ $\ln n \geq 3.262491362$ $n \geq 26.1$ The number of reported cases is at least 1,000 in the 27th month
9 **(v)**	The value of a represents the **estimated** number of cases **(in hundreds)** of the virus reported in the **first month**.
10 **(i)**	Let X_1, X_2 be the number of customers on Day 1 & Day 2 respectively Then $X_1 \sim \text{Po}(20), X_2 \sim \text{Po}(20)$ Since $\lambda = 20 > 10$, $X_1 \sim \text{N}(20, 20)$ approximately, $X_2 \sim \text{N}(20, 20)$ approximately. So $X_1 - X_2 \sim \text{N}(0, 40)$ $\text{P}(\lvert X_1 - X_2 \rvert \leq 2) = \text{P}(-2 \leq X_1 - X_2 \leq 2)$ $\xrightarrow{c.c.} \text{P}(-2.5 \leq X_1 - X_2 \leq 2.5) = 0.307$
10 **(ii)**	Let Y be the number of customers arrived in n hours So, $Y \sim \text{Po}\left(\frac{20}{8} n\right)$ i.e. $Y \sim \text{Po}(2.5n)$ where $n \in \mathbb{Z}^+$, $0 \leq n \leq 8$. $\text{P}(Y \leq 2) < 0.01$ $\Rightarrow \text{P}(Y = 0) + \text{P}(Y = 1) + \text{P}(Y = 2) < 0.01$ $\Rightarrow \text{e}^{-2.5n} \left[\frac{(2.5n)^0}{0!} + \frac{(2.5n)^1}{1!} + \frac{(2.5n)^2}{2!} \right] < 0.01$ $\Rightarrow \frac{1 + 2.5n + 3.125n^2}{\text{e}^{2.5n}} < 0.01$ When n = 3, $\text{P}(Y \leq 2) = 0.02026 > 0.01$ When n = 4, $\text{P}(Y \leq 2) = 0.00277 < 0.01$ So the set of values of n is $\{4, 5, 6, 7, 8\}$ or $\{n \in \mathbb{Z}^+ : 4 \leq n \leq 8\}$
10 **(iii)**	1. The arrival of customers may not be independent, for example, relatives/friends visit the restaurant together. 2. The mean number of customers per unit time may not be constant throughout the day. For example, lunch time we may expect more customers.

11 **(ai)**	Let μ be the population mean time (in seconds) for a school boy in the school to complete a 4×10 m shuttle run. H_0 : $\mu = 10.8$ H_1 : $\mu \neq 10.8$
11 **(aii)**	**Assumption**: Since sample size is small and population variance is unknown, we need to assume that the time, in seconds, for a school boy to complete a 4×10 m shuttle run follows a normal distribution. $\overline{x} = \frac{\sum x}{9} = \frac{93.96+k}{9}$ Sample variance = 1.09^2 $s^2 = \frac{9}{8}(1.09)^2 = 1.3366125$ Under H_0, **Test Statistic** $T = \frac{\overline{X} - 10.8}{\sqrt{\frac{1.3366125}{9}}} \sim t\,(8)$ **Test Statistic value**, $t = \frac{\overline{x} - 10.8}{\sqrt{\frac{1.3366125}{9}}}$ Reject H_0 at 5% level of significance, $\mathrm{P}\left(\lvert T\rvert \geq t\right) \leq 0.05$ $\left\lvert \frac{\overline{x} - 10.8}{\sqrt{\frac{1.3366125}{9}}} \right\rvert \geq 2.306004133$ $\frac{\left(\frac{93.96+k}{9}\right) - 10.8}{\sqrt{\frac{1.3366125}{9}}} \leq -2.306004133$ $k \leq -4.758049639$ (rejected since $k > 0$)

	or $\dfrac{\left(\dfrac{93.96+k}{9}\right)-10.8}{\sqrt{\dfrac{1.3366125}{9}}} \ge 2.306004133$ $k \ge 11.2$ (3 sf) $\left\{k : k \in \square ,\quad k \ge 11.2\right\}$
11 (b)	The unbiased estimate for population mean is $\bar{y}$ $\quad \bar{y} = \dfrac{\sum y}{50} = \dfrac{517.49}{50}$ $\bar{y} = \dfrac{51749}{5000} = 10.3498$ The unbiased estimate for population variance is s^2 $s^2 = \dfrac{\sum\left(y-\bar{y}\right)^2}{n-1}$ $s^2 = \dfrac{122.32}{49} = \dfrac{3058}{1225} = 2.496326531$ $H_0 : \mu = 10.8$ $H_1 : \mu < 10.8$ Under H_0, **Test Statistic** $Z = \dfrac{\overline{X} - 10.8}{\sqrt{\dfrac{\left(\dfrac{3058}{1225}\right)}{50}}} \sim \mathrm{N}\left(0,1\right)$ approximately by Central Limit Theorem since *n* (=50) is large From G.C., *p*-value = $0.0219608 \approx 0.0220$ (3 s.f.) Reject H_0, $\alpha\% \ge 2.20\%$ and thus $\alpha = 2.20$. Since *n* (=50) is large, the mean timing for the boys to complete a 4×10 m shuttle run will be approximated to a normal distribution by Central Limit Theorem. Therefore **not necessary** to assume population follows normal distribution.
11 (c)	Given *n* is large, by Central Limit Theorem, $\overline{X} \sim \mathrm{N}\left(\mu, \dfrac{2.8^2}{n}\right)$. $\mathrm{P}(\overline{X} - \mu \ge 0.8) \le 0.001$ $\mathrm{P}\left(Z \le \dfrac{0.8}{\dfrac{2.8}{\sqrt{n}}}\right) \ge 0.999$

	$\frac{0.8}{\frac{2.8}{\sqrt{n}}} \geq 3.090232308$ $n \geq 116.98$ $\therefore n = 117$