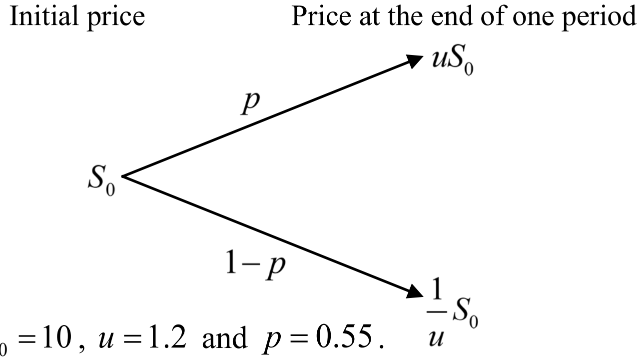
RAFFLES INSTITUTION
H2 Mathematics 9758
2025 Year 6 Term 3 Revision (Zeta)
Topic(s): (C) Statistics
(Solutions)

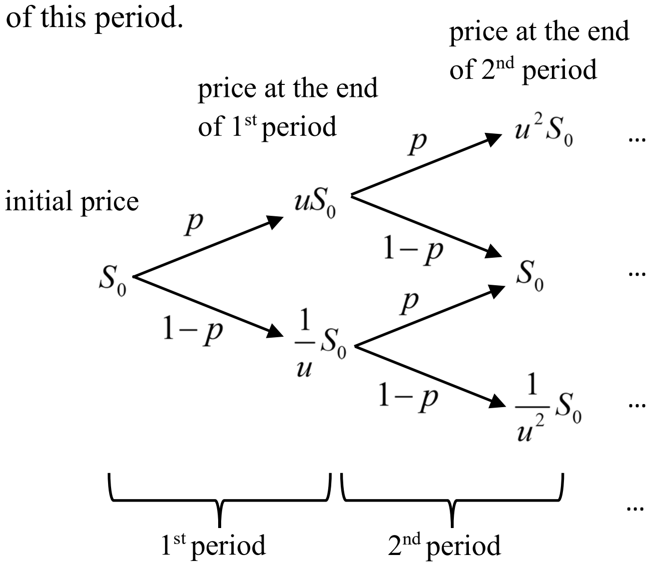
Source of Question: NJC Prelim 9758/2023/02/Q11

1 **(a)** In a one-period binomial model, the stock price initially at S_0, either rises to uS_0 with a probability p or drops to $\frac{1}{u}S_0$ with a probability $(1-p)$ at the end of one period, where $u>1$.



Suppose that $S_0=10$, $u=1.2$ and $p=0.55$.
Show that the expected value of the stock price S at the end of the period is 10.35 and find the variance of the stock price. [3]

(b) The above model can be extended to a n-period binomial model. For each period, the stock price S either rises to uS with a probability p or goes drops to $\frac{1}{u}S$ with a probability $1-p$ at the end of this period.



(i) In the first quarter, the initial price of the stock at the start of the first period is S_0. Suppose the stock price rises 4 times and drops twice over 6 periods. Find the stock price in terms of S_0 and u at the end of the 6 periods. [1]

(ii) Show that the probability for part **(b)(i)** to happen is $15p^4(1-p)^2$. [1]

(iii) In the second quarter, the initial price of the stock at the start of the first period is S_1. Find the probability that the stock price is higher than S_1 at the end of 6 periods in terms of p. [2]

(c) A *contract* that gives investors the right to buy a stock at a specified *price* at the end of a specified period is termed "call option". The specified price is called "strike price" and the amount needed to buy this contract is called "premium".

Each unit of a stock's call option has a return value V at the end of the period. Its value is given by

$$V = \begin{cases} S - S_{\text{strike}} & \text{if } S > S_{\text{strike}}, \\ 0 & \text{if } S \le S_{\text{strike}}, \end{cases}$$

where S_{strike} is the strike price and S is the stock price at the end of the period.

Stock K, currently priced at \$10, is modelled by the n-period binomial model in part **(b)** with $n = 6$, $u = 1.2$ and $p = 0.55$. A call option of stock K has a strike price of \$10.

(i) Fill in the missing numbers in the following probability distribution table of V for this call option, giving the values to a suitable degree of accuracy. [2]

Number of times for the stock price to rise	3 or below	4	5	6
v	0		10.736	19.85984
$\mathrm{P}(V = v)$			0.13589	0.02768

The call option currently has a premium of \$2.50 per unit. Peter plans to invest \$10 000 on purchasing 4000 units of the call option and sell them at the end of the 6 periods.

(ii) Find the expected return value of the investment, giving your answer to the nearest dollar. [2]

(iii) Comment on whether Peter should proceed with this investment plan. [1]

Solution:

(a)	$\mathrm{E}(S) = 0.55(1.2 \times 10) + 0.45\left(\frac{1}{1.2} \times 10\right) = 10.35$ $\mathrm{E}(S^2) = 0.55(1.2 \times 10)^2 + 0.45\left(\frac{1}{1.2} \times 10\right)^2 = 110.45$ $\mathrm{Var}(S) = \mathrm{E}(S^2) - [\mathrm{E}(S)]^2 = 110.45 - 10.35^2 = 3.3275$

(b)(i)	$S_0\left(u^4\right)\left(\frac{1}{u}\right)^2=u^2S_0$
(b)(ii)	The required probability $=\binom{6}{4}p^4\left(1-p\right)^2=15p^4\left(1-p\right)^2$
(b)(iii)	If the stock price goes up 3 times and goes down 3 times, the end price is exactly S_0, so there must be at least 4 "up"s. Let X denote the number of "Up"s in the 6 periods. $X\sim \mathrm{B}\left(6,p\right)$ $P\left(X\geq 4\right)=\binom{6}{4}p^4\left(1-p\right)^2+\binom{6}{5}p^5\left(1-p\right)^1+\binom{6}{6}p^6$ $=15p^4\left(1-p\right)^2+6p^5\left(1-p\right)+p^6$
(c)(i)	$10\times 1.2^2-10=4.4$ $X\sim \mathrm{B}\left(6,0.55\right)$ $\mathrm{P}\left(X\leq 3\right)=0.55848$ and $\mathrm{P}\left(X=4\right)=0.27795$ <table><tr><td>3 or below</td><td>4</td><td>5</td><td>6</td></tr><tr><td>0</td><td>4.4</td><td>10.736</td><td>19.85984</td></tr><tr><td>0.55848</td><td>0.27795</td><td>0.13589</td><td>0.02768</td></tr></table>
(c)(ii)	$\mathrm{E}\left(V\right)=4.4\times 0.27795+10.736\times 0.13589+19.85984\times 0.02768$ $=3.2316154$ Alternatively, $\mathrm{E}\left(V\right)=\left(10\times 1.2^2-10\right)\mathrm{P}\left(X=4\right)+\left(10\times 1.2^4-10\right)\mathrm{P}\left(X=5\right)$ $+\left(10\times 1.2^6-10\right)\mathrm{P}\left(X=6\right)$ $=\sum_{r=1}^{3}\left(1.2^{2r}-1\right)\mathrm{P}\left(X=r+3\right)=3.23159461$ The expected return is $4000\left(3.2316154\right)=12926.46=12926$ (nearest dollar).
(iii)	Acceptable answer 1: Good investment as the expectation is higher than 10000. Acceptable answer 2:

	By GC, $\sigma_V = 4.6387546$ The return $4000V$ has a standard deviation $4.6387546 \times 4000 = 18555$ (nearest dollar) Risky investment as the standard deviation (or variance) is large (in comparison / relative to the expectation). Acceptable answer 3: Calculate the expectation of using \$10000 to buy the stock directly. $10000\sum_{r=0}^{6}\left(1.2^{2r-6}-1\right)\mathrm{P}\left(X=r\right)=12293$ (nearest dollar) Good investment as the expectation is higher than buying the stock directly.

Source of Question: NJC Prelim 9758/2022/02/Q7

2 On average, 7.8% of markers produced by a factory are faulty. Each day, the factory's quality manager picks a random sample of n markers and inspects the number of faulty markers found in the sample. The number of faulty markers found in the sample is the random variable X and whether the markers are faulty are independent of one another.

(i) Explain what is meant by a random sample in this context. [1]

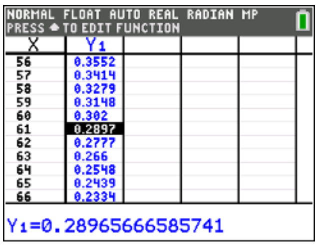
(ii) Given that the probability that more than $(n-4)$ markers in the random sample are found to be non-faulty is less than 0.3, find the least possible value of n. [2]

Assume for the remainder of this question that $n = 10$.

The samples in 50 randomly chosen days are collected and the number of faulty markers in each sample is recorded.

(iii) Find the probability that exactly 35 of the samples each contains at least one faulty marker, given that none of the 50 samples contain more than 2 faulty markers each. [4]

Solution:

2 (i)	A random sample is one such that every marker produced has equal probability of being selected in this sample and the markers are selected independently of one another.
(ii)	Let X be the number of markers, out of n, that are faulty. $X \sim \mathrm{B}\left(n, 0.078\right)$ If $n-X$ markers are not faulty, X markers are faulty. Thus, $\mathrm{P}\left(n-X > n-4\right) < 0.3$ $\mathrm{P}\left(X < 4\right) < 0.3$ $\mathrm{P}\left(X \le 3\right) < 0.3$ 

Using GC,

$n=60,\ \mathrm{P}(X \le 3)=0.302>0.3$

$n=61,\ \mathrm{P}(X \le 3)=0.290<0.3$

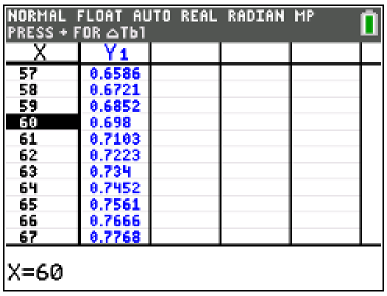
Least value of n is 61.

Alternatively, we can define W to be the number of markers, out of n, that are NOT faulty.
Then $W \sim \mathrm{B}(n,1-0.078)=\mathrm{B}(n,0.922)$.

$\mathrm{P}(W>n-4)<0.3$

$1-P(W \le n-4)<0.3$

$\mathrm{P}(W \le n-4)>0.7$



Using GC,

When $n=60,\ \mathrm{P}(W \le n-4)=0.698<0.7$

When $n=61,\ \mathrm{P}(W \le n-4)=0.7103>0.7$

Hence least n is 61.

(iii) Let the number of faulty markers found in the sample of 10 be Y.

$Y \sim \mathrm{B}(10,0.078)$

$\mathrm{P}(Y=0)=0.4439246$

$\mathrm{P}(Y \le 2)=0.962450$

$\mathrm{P}(1 \le Y \le 2)=\mathrm{P}(Y \le 2)-\mathrm{P}(Y=0)=0.518526$

Thus, the required probability is

$$\frac{\text{P(all 50 samples each have less than or equal to 2 faulty markers and 35 of the samples have at least 1 faulty marker each)}}{\text{P(all 50 samples each have less than or equal to 2 faulty markers)}}$$

$$=\frac{\binom{50}{35}\left(\mathrm{P}(1 \le Y \le 2)\right)^{35}\left(\mathrm{P}(Y=0)\right)^{15}}{\left(\mathrm{P}(Y \le 2)\right)^{50}}$$

$$=\frac{\binom{50}{35}(0.518526)^{35}(0.4439246)^{15}}{(0.962450)^{50}}$$

$=0.0081291=0.00813$ (to 3 s.f.)

Source of Question: TMJC Prelim 9758/2022/02/Q11

3 **In this question you should state clearly all the distributions that you use, together with the values of the appropriate parameters.**

In a factory, oil is stored in barrels of different sizes. It is given that the volumes of oil in the barrels can be modelled using normal distributions with means and standard deviations as shown in the table.

	Mean volume	Standard deviation

	(in litres)	(in litres)
Light oil	110	2.5
Heavy oil	145	3.5

(i) Find the probability that the volume of light oil in a randomly chosen barrel is between 104 litres and 116 litres. [1]

(ii) Sketch the distribution for the volume of light oil, indicating clearly the probability found in part **(i)**. [2]

(iii) A random sample of seven barrels of heavy oil is chosen. Find the probability that exactly four barrels of heavy oil have volume between 142 and 150 litres each and exactly one barrel of heavy oil has volume more than 150 litres. [3]

(iv) A random sample of n barrels of heavy oil is chosen and it is given that the probability that the mean volume of these n barrels of heavy oil exceeding k litres is at least 0.3. Find an inequality, expressing k in terms of n. [3]

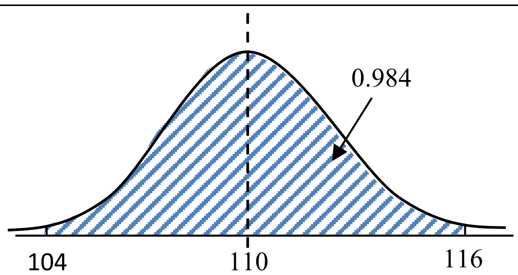
A lorry with a maximum laden mass of 6800 kilograms is used to transport 25 barrels of light oil and 30 barrels of heavy oil. It is given that the densities of light oil and heavy oil are 0.83 kilograms per litre and 0.94 kilograms per litre respectively, and the empty barrels for light oil weigh 5 kilograms each and the empty barrels for heavy oil weigh 8 kilograms each.

(v) Find the probability that the load of the lorry exceeds its maximum laden mass. [4]

[Density is defined as $\dfrac{\text{Mass}}{\text{Volume}}$.]

(vi) State an assumption needed for your calculations in part **(v)**. [1]

Solution:

3(i)	Let X and Y be the volume of oil in a randomly chosen barrel of light and heavy oil respectively. $X \sim \mathrm{N}\left(110,\ 2.5^2\right)$ $\quad Y \sim \mathrm{N}\left(145,\ 3.5^2\right)$ $\mathrm{P}\left(104 < X < 116\right) = 0.984$.
(ii)	 Your sketch should indicate clearly that the mean of the distribution of X is at 110, and that the required limits, 104 and 116, are symmetric with respect to the mean.
(iii)	Required Probability $= \mathrm{P}\left(142 < Y < 150\right)^4 \times \mathrm{P}\left(Y > 150\right) \times \mathrm{P}\left(Y < 142\right)^2 \times \dfrac{7!}{4!2!} = 0.0863$
(iv)	Let $\overline{Y} = \dfrac{Y_1 + Y_2 + \ldots + Y_n}{n}$ and $\overline{Y} \sim \mathrm{N}\left(145, \dfrac{3.5^2}{n}\right)$

	$\mathrm{P}\left(\overline{Y} > k\right) \geq 0.3$ $\mathrm{P}\left(Z > \dfrac{k-145}{\frac{3.5}{\sqrt{n}}}\right) \geq 0.3$ $\dfrac{k-145}{\frac{3.5}{\sqrt{n}}} \leq 0.52440$ $k \leq 145 + \dfrac{1.84}{\sqrt{n}}$
(v)	Let $T = 0.83\left(X_1 + X_2 + \cdots X_{25}\right) + 0.94\left(Y_1 + Y_2 + \cdots Y_{30}\right)$ $\mathrm{E}(T) = 0.83(110 \times 25) + 0.94(145 \times 30) = 6371.5$ $\mathrm{Var}(T) = 0.83^2\left(2.5^2 \times 25\right) + 0.94^2\left(3.5^2 \times 30\right) = 432.363625$ $T \sim \mathrm{N}(6371.5, 432.363625)$ $\mathrm{P}\left(T + 25(5) + 30(8) > 6800\right)$ $= \mathrm{P}(T > 6435) = 0.00113$
(vi)	The distributions of the volume of **all types of oil** are **independent** of one another.

Source of Question: HCI Prelim 9758/2023/02/Q8

4 A biased 4-sided die gives the scores 1, 2, 3, and 4 with the probabilities shown in the table, where a and b are constants. The random variable S denotes the score on the die.

Score (s)	1	2	3	4
$\mathrm{P}(S = s)$	a	a	a	b

It is given that the mean of the score is 2.56.

(a) Show that $b = \dfrac{7}{25}$. [2]

(b) S_1 and S_2 are two independent observations of S.

(i) Find the probability that the sum of the two independent observations of S is at least 6, given that at least one of the observed scores is 3, [3]

Let $Y = \left|S_1 - S_2\right|$. Without finding the probability distribution table of Y,

(ii) find $\mathrm{Var}\left(2S - \mathrm{E}(Y)\right)$. [3]

(iii) state the value of $\mathrm{E}\left(Y - \mathrm{E}(Y)\right)$. [1]

Solution:

(a)	$3a+b=1 \quad \text{-- } (*)$ $\mathrm{E}(S)=6a+4b=2.56 \quad \text{-- } (*)$ $3a+2b=1.28$ Subtracting one eqn from the other, $b=1.28-1=0.28=\dfrac{7}{25}$
(b)(i)	From part (a), $a=\dfrac{6}{25}$ $\mathrm{P}(S_1+S_2\geq 6 \mid \text{at least one of the scores is 3})$ $=\dfrac{\mathrm{P}(S_1=3,S_2=4)+\mathrm{P}(S_1=4,S_2=3)+\mathrm{P}(S_1=3,S_2=3)}{1-\mathrm{P}(\text{both not 3})}$ $=\dfrac{2\left(\frac{6}{25}\right)\left(\frac{7}{25}\right)+\left(\frac{6}{25}\right)^2}{1-\left(\frac{19}{25}\right)^2}=\dfrac{120}{264}=\dfrac{5}{11}=0.455 \text{ (3 s.f.)}$
(b)(ii)	From part (a), $a=\dfrac{6}{25}$ $\mathrm{Var}(2S-\mathrm{E}(Y))$ $=\mathrm{Var}(2S) \quad (\because \mathrm{E}(Y) \text{ is a constant and } \mathrm{Var}(\text{constant})=0)$ $=4\mathrm{Var}(S)$ $=4\left(\mathrm{E}(S^2)-2.56^2\right)$ $=4\left(a+4a+9a+16b-2.56^2\right)$ $=4(14a+16b-6.5536)=5.1456$
(b)(iii)	$\mathrm{E}(Y-\mathrm{E}(Y))=\mathrm{E}(Y)-\mathrm{E}(\mathrm{E}(Y))$ $=\mathrm{E}(Y)-\mathrm{E}(Y)=0$

Source of Question: EJC Prelim 9758/2021/02/Q9

5 A company produces light bulbs with a brightness rating of 470 lumens. It is known that the brightness of these light bulbs is normally distributed with a standard deviation of 10 lumens. A production manager wishes to test whether the mean brightness is less than 470 lumens. He collects a random sample of 40 light bulbs and finds that their mean brightness is 466.8 lumens.

(i) Explain what is meant by the term "random sample" in the context of this question. [1]

(ii) Carry out a test at the 5% level of significance for the production manager, clearly stating the *p*-value of this test. Explain what *p*-value means in the context of this question. [5]

(iii) Suppose that the production manager had tested if the mean brightness differed from 470 lumens at the 5% level of significance. Without performing another hypothesis test, determine whether your conclusion in **part (ii)** would be affected. [2]

The company decides to produce a new line of 'lowlight' bulbs with a brightness rating of 350 lumens. A random sample of 50 'lowlight' bulbs is taken and the brightness level, y lumens, are summarised as follows.

$$n = 50 \qquad \sum(y-350) = -125 \qquad \sum(y-347.5)^2 = 4704$$

(iv) Calculate unbiased estimates of the population mean and variance for the brightness of the 'lowlight' bulbs. [2]

(v) Given instead that the standard deviation of the brightness of the 'lowlight' bulbs is known to be σ, the company found insufficient evidence from this sample to conclude that the mean brightness was less than 350 lumens at the 1% level of significance. Determine the range of values of σ. [2]

Solution:

(i)	The term 'random sample' means that each light bulb had an equal probability of being selected for the sample, and the selection of light bulbs were independent of each other.
(ii)	Let X lumens be the brightness of a randomly chosen light bulb with population mean μ lumens. $H_0 : \mu = 470$ $H_1 : \mu < 470$ Perform a 1-tail test at 5% level of significance. Under H_0, $\overline{X} \sim N\left(470, \frac{100}{40}\right)$ Using a *z*-test, *p*-value = $P\left(\overline{X} \le 466.8\right) \approx 0.021492 = 0.0215$ (3sf) Since 00215< 0.05, we reject H_0 and conclude that there is sufficient evidence at 5% level of significance that the population mean brightness is less than 470 lumens. *p*-value refers to the probability of obtaining a sample mean brightness of 40 light bulbs that is less than or equal to the observed **<u>sample mean value</u>** of 466.8 lumens if the **<u>population mean brightness</u>** is 470 lumens.
(iii)	Performing a 2-tail test instead, *p*-value = $2P\left(\overline{X} \le 466.8\right) \approx 0.0430 < 0.05$ We would still reject H_0 at 5% level of significance, the conclusion of the test in part (ii) will not be affected.

(iv)

An unbiased estimate of population mean, $\bar{y} = \dfrac{-125}{50} + 350 = 347.5$

An unbiased estimate of population variance, $s_y^{\ 2} = \dfrac{4704}{49} = 96$

(v)

$\mathrm{H_0} : \mu_y = 350 \quad \text{vs} \quad \mathrm{H_1} : \mu_y < 350$

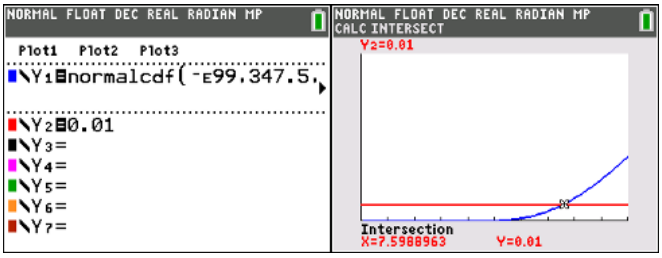
Perform a 1-tail test at 1% level of significance.

Under $\mathrm{H_0}$, since $n = 50$ is large, $\overline{Y} \sim \mathrm{N}\left(350, \dfrac{\sigma^2}{50}\right)$ approximately by Central Limit Theorem.

Using a z-test, p-value $= \mathrm{P}(\overline{Y} \leq 347.5)$

$\mathrm{H_0}$ is not rejected, i.e. $\mathrm{P}(\overline{Y} \leq 347.5) \geq 0.01$

<u>Method 1:</u>



From GC, $\sigma \geq 7.60$ (3 s.f)

<u>Method 2:</u>

$\mathrm{P}(\overline{Y} \leq 347.5) \geq 0.01$

$$\mathrm{P}\left(Z \leq \frac{347.5 - 350}{\sigma/\sqrt{50}}\right) \geq 0.01$$

$$-\frac{2.5\sqrt{50}}{\sigma} \geq -2.3263$$

$-2.5\sqrt{50} \geq -2.3263\sigma$

$\sigma \geq 7.60$

Source of Question: HCI Prelim 9758/2023/02/Q10

6 **In this question you should, where applicable, state clearly all the distributions that you use, together with the values of the appropriate parameters.**

A supermarket produces its own packs of white sesame seeds and fennel seeds for sale under its house brand. The masses (in grams) of a randomly chosen pack of white sesame seeds and that of a randomly chosen pack of fennel seeds, denoted by W and F respectively, have independent normal distributions. The means and standard deviations of these distributions are shown in the following table.

Mass of a pack of	Mean (g)	Standard deviation (g)
white sesame seeds (W)	250	σ
fennel seeds (F)	300	2.5

It is given that 5% of the packs of white sesame seeds have a mass of less than 245.

(a) Show that the value of σ is 3.0398, correct to 5 significant figures. [2]

(b) Find the probability that the total mass of six randomly chosen packs of white sesame seeds is less than five times the mass of a randomly chosen pack of fennel seeds by not more than 20 grams. [3]

(c) n packs of white sesame seeds and n packs of fennel seeds are chosen at random. Find the smallest value of n such that the probability that the mean mass of these $2n$ packs being at least 278 grams is less than 0.015. [4]

Solution:

(a)	$W \sim \text{N}(250,\ \sigma^2)$ $Z = \dfrac{W-250}{\sigma} \sim \text{N}(0,\ 1)$ $\text{P}\left(W < 245\right) = 0.05$ $\text{P}\left(Z < \dfrac{245-250}{\sigma}\right) = 0.05$ $\dfrac{-5}{\sigma} = -1.644853626$ $\sigma = 3.039784161$ $= 3.0398$ (to 5 s.f.) (shown)
(b)	$W_1 + W_2 + \cdots + W_6 \sim \text{N}\left(1500,\ 55.44172647\right)$ $F \sim \text{N}(300,\ 2.5^2)$ $5F \sim \text{N}(1500,\ 156.25)$ Let $D = 5F - \left(W_1 + W_2 + \cdots + W_6\right)$ $D \sim \text{N}(0,\ 211.6917265)$ $\text{P}\left(0 < D \le 20\right) = 0.4153730399 = 0.415$ (to 3 s.f.)
(c)	Let $M = \dfrac{\left(W_1 + W_2 + \cdots + W_n\right) + \left(F_1 + F_2 + \cdots + F_n\right)}{2n}$

$$\mathrm{E}(M) = \frac{250n + 300n}{2n} = 275$$

$$\mathrm{Var}(M) = \frac{(3.039784161)^2 n + (2.5)^2 n}{4n^2} = \frac{3.872571936}{n}$$

$$M \sim \mathrm{N}\left(275,\ \frac{3.872571936}{n}\right)$$

$$Z = \frac{M - 275}{\sqrt{\dfrac{3.872571936}{n}}} \sim \mathrm{N}(0,\ 1)$$

Method 1:

$$\mathrm{P}(M \geq 278) < 0.015$$

$$\mathrm{P}\left(Z \geq 1.524479216\sqrt{n}\right) < 0.015$$

$$\mathrm{P}\left(Z < 1.524479216\sqrt{n}\right) > 0.985$$

$$1.524479216\sqrt{n} > 2.170090375 --- (*)$$

$$n > 2.026341439$$

$$n \geq 3$$

$\therefore$ smallest value of $n = 3$

Method 2:

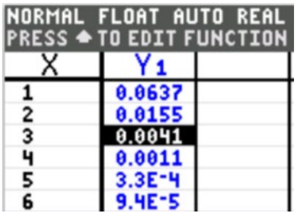
$\mathrm{P}(M \geq 278) < 0.015$

From GC,

When $n = 2$, $\mathrm{P}(M \geq 278) = 0.0155 > 0.015$

When $n = 3$, $\mathrm{P}(M \geq 278) = 0.0041 < 0.015$

$\therefore$ smallest value of $n = 3$



NORMAL FLOAT AUTO REAL
PRESS ♦ TO EDIT FUNCTION

X	Y_1	
1	0.0637	
2	0.0155	
3	0.0041	
4	0.0011	
5	3.3E-4	
6	9.4E-5	

Source of Question: HCI Prelim 9758/2023/02/Q11

7 Researchers at University S claimed that *Fungus A*, a strain of fungi that is typically found in soil and plants, was able to break down polypropylene, a plastic that is hard to recycle, that had been pre-treated with either UV light or heat by at least 27 grams per 100 grams of polypropylene over 90 days.

Researchers at University M believe that the researchers at University S overstated their claim and decided to perform a hypothesis test on a random sample of 120 batches of polypropylene, each of mass 100 grams. The researchers recorded the amount x grams of polypropylene for each batch that *Fungus A* was able to break down over a 90-day period. The results are summarised as follows:

$$\sum(x-27)=-81,\ \sum(x-27)^2=2070.8$$

(a) Calculate the unbiased estimates of the population mean and variance. [3]

(b) Test, at 5% level of significance, the claim that the population mean mass of polypropylene per 100 grams that were broken down is at least 27 grams. You should state your hypotheses and define the symbols you use. [4]

(c) State, in the context of the question, what you understand by the expression 'at 5% level of significance' found in part **(b)**. [1]

(d) Explain if it is necessary for University M to assume a normal distribution for the mass of polypropylene per 100 grams that *Fungus A* was able to break down for this test to be valid. [1]

University S responded by experimenting on 50 batches of polypropylene, each of mass 100 grams, and recorded the amount y grams of polypropylene for each batch that *Fungus A was* able to break down over a 90-day period. The null and alternative hypotheses remain the same as the ones stated by University M in part **(b)**.

Based on the research findings from University S, the population standard deviation of the mass of polypropylene per 100 grams that *Fungus A* was able to break down is known to be 4 grams.

(e) Find the set of values of sample mean mass, $\overline{y}$ grams, of polypropylene per 100 grams from University S such that the null hypothesis is not rejected in favour of the alternative hypothesis at 5% level of significance. [3]

Solution:

(a)	Let $m = x - 27$, $\sum m = -81$ and $\sum m^2 = 2070.8$. An unbiased estimate of the population mean, $\bar{x} = \bar{m} + 27 = \dfrac{-81}{120} + 27 = 26.325$ g An unbiased estimate for the population variance, $s_x^{\ 2} = \dfrac{1}{119}\left(\sum m^2 - \dfrac{\left(\sum m\right)^2}{120}\right) = \dfrac{1}{119}\left(2070.8 - \dfrac{(-81)^2}{120}\right) = 16.94222689 = 16.9\text{ g}^2$ (to 3 s.f.)
(b)	Let μ be the population mean of X. $H_0 : \mu = 27$ vs $H_1 : \mu < 27$ Perform 1 tail test at 5% level of significance. Under H_0, since $n = 120$ is large, then by Central Limit Theorem, $\bar{X} \sim N\left(27, \dfrac{16.942}{120}\right)$ approximately. p-value = $P(\bar{X} < 26.325) = 0.0362$ (to 3 s.f.). Since p-value = 0.0362 < 0.05, we reject H_0 and conclude that at the 5% level of significance, there is sufficient evidence that the population mean mass of polypropylene per 100g that were broken down is less than 27g.
(c)	It means that there is a probability of 0.05 of wrongly concluding that the **<u>population mean mass</u>** of polypropylene per 100g that was broken down by Fungus A was less than 27g when the **<u>population mean mass</u>** is, in fact, 27 g.
(d)	Since the sample size of 120 is large enough, University M can use Central Limit Theorem to approximate the distribution of the sample mean mass ($\bar{X}$) of polypropylene per 100g that were broken down to be a normal distribution, hence University M does not need to know the distribution of the mass of polypropylene (X) per 100g that were broken down.
(e)	Under H_0, since $n = 50$ is large, by Central Limit Theorem, $\bar{Y} \sim N\left(27, \dfrac{4^2}{50}\right)$ approximately. Given that H_0 is not rejected at 5% level of significance, $P(\bar{Y} < \bar{y}) > 0.05$ $\dfrac{\bar{y} - 27}{\sqrt{\dfrac{4^2}{50}}} > -1.644853626$ $\bar{y} - 27 > -0.9304697224$ $\bar{y} > 26.06953028$ $\therefore$ set of values = $\left\{\bar{y} \in \mathbb{R} : 26.1 \le \bar{y} \le 100\right\}$ (to 3 s.f.)