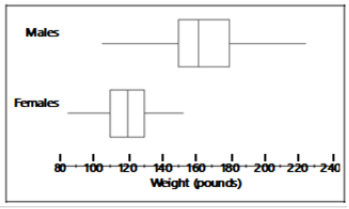
Multiple Choice Questions (10 marks)

1	Samson collated all the exam marks and analysed the results. Suppose that the exam marks, X, is normally distributed with mean of 75 and standard deviation σ = 10. Empirically, what is the probability if the score is lower than 55 marks i.e. P(X≤55)? a) 5% b) 2.275% c) 4.55% d) 2.5%	(D)
2	Suppose that a normal model describes the acidity (pH) of rainwater, and that water tested after last week's storm had a z-score of 1.8. This means that the acidity of the rain a) had a pH 1.8 higher than average rainwater. b) varied with standard deviation 1.8. c) had a pH 1.8 standard deviations higher than that of average rainwater. d) had a pH 1.8 times that of average rainwater.	(C)
3	SAE MST has 5 multiple choice questions with five choices with two correct answer each. If we just randomly guess on each of the 5 questions, what is the probability that you get exactly 4 questions correct? a) 0.0768 b) 0.4096 c) 0.2592 d) 0.0064	(A)
4	The weights of male and female students in a class are summarized in the boxplots below. Which of the following statements is not true?  a) About 50% of male students weigh between 150 and 180 pounds. b) About 25% of female students weigh more than about 130 pounds. c) The median weight of male students is about 162 pounds. d) Male students' weights have less variability than female students' weights.	(D)
5	The distribution of the population of household incomes in Singapore is skewed to the right. Which of the following best describes what happens to the sampling distribution of the sample mean when the size of a random sample increases from 10 to 100? a) Its mean gets closer to the population mean, its standard deviation gets closer to the population standard deviation, and its shape gets closer to the population's shape. b) Its mean gets closer to the population mean, its standard deviation gets smaller, and its shape gets closer to normal. c) Its mean stays constant, its standard deviation gets closer to the population standard deviation, and its shape gets closer to the population's shape. d) Its mean stays constant, its standard deviation gets smaller, and its shape gets closer to normal.	(D)

Question 1 (25 marks)

1 A study was conducted to analyse the attributes of male workers who work in physically demanding jobs. 147 male workers working in physically demanding jobs were recruited to participate in the study. 5 variables on each worker were recorded.
The data collected can be found in the file "Revision Data–MST.xlsx" in the tab labelled "Data".
The descriptions of the variables are as follow:

Variable	Meaning
Grip strength	Maximum force (in Newton) of grip.
Arm strength	Maximum force (in Newton) of arm.
Age	Age of worker in years.
Dominant hand	The worker's preference for use of a hand (left, right or mixed).
Rating	Job performance rating of the worker given by his supervisor on a scale of 1 to 60. Higher rating indicates better job performance.

a Identify the population and the sample in which the study is based on. — 2 marks

Population	All male workers in physically demanding jobs
Sample	147 male workers in physically demanding jobs who participated in the study

b What is the type of data for each of the variables listed in the table below? — 8 marks

Variable	Qualitative / Quantitative	Ordinal / Nominal / Discrete / Continuous
Grip strength	Quantitative	Continuous
Age	Quantitative	Discrete
Dominant hand	Qualitative	Nominal
Rating	Qualitative	Ordinal

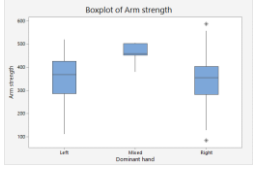
c Fill in the following information. — 8 marks

Mean Arm Strength: 350.77 N (2 dec pl)		
Interquartile range of Grip Strength: 135.70 N (2 dec pl)		
Average Age: 35.3	Shape of distribution of Age: Symmetrical	
Percentage of workers with:		
Right dominant hand 85%	Left dominant hand 10.2%	Mixed dominant hand 4.8%
Mode Rating of workers with mixed dominant hand: 50		

d Find the correlation coefficient r between Grip Strength and Arm Strength. Hence, comment on the relationship between the two variables. — 3 marks
r = 0.627
Positive and moderate

e Which dominant hand (left, right or mixed) group has the highest Arm Strength on average? — 1 mark
Mixed dominant hand

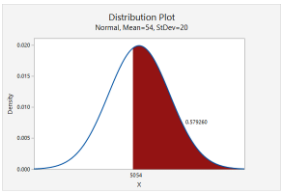
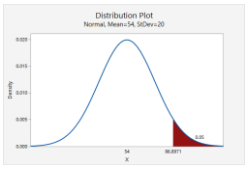
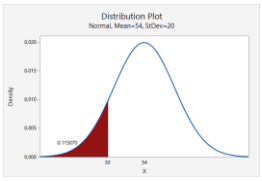
f Based on your answer in part (e), is there visual evidence that this group has significantly higher Arm Strength than workers in the other 2 groups? Explain. — 3 marks
Yes, the box of the mixed dominant hand group does not overlap with the "boxes" of the left dominant hand of the group or the right dominant hand of the group



Summary Statistics

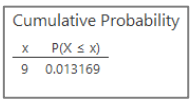
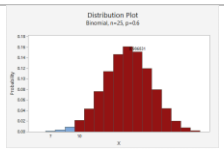
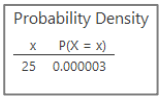
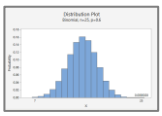
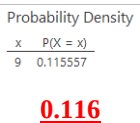
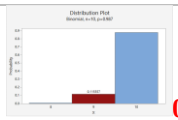
Dominant hand	N	Minimum	Q1	Median	Q3	Maximum	95% Median CI
Left	15	111.20	286.90	369.20	427.00	520.40	(295.19, 426.18)
Mixed	7	380.10	451.50	460.40	502.60	504.80	(432.46, 503.19)
Right	125	84.500	283.550	355.800	404.750	587.100	(316.253, 373.600)

Question 2 (20 marks)

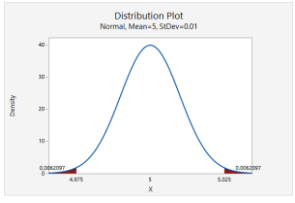
2	The marks of 500 candidates in an examination are normally distributed with a mean of 54 marks and a standard deviation of 20 marks. The pass mark for the examination is 50 marks.	
a	If a candidate is chosen at random, find the probability that the candidate passes the examination $P(X>50)$ $P(X > \frac{50-54}{20})$ $P(Z > -0.2)$ $0.5+0.0793$ **0.579** or  **0.579**	4 marks
b	Estimate the number of candidates who passed the examination. Round your answer to the nearest whole number. Estimated number of candidates who passed = 500 x 0.5793 = 290	2 marks
c	If 5% of the candidates obtain a distinction by scoring x marks or more, estimate the value of x. Round your answer to the nearest whole number. $P(X>x) = 0.05$ $P(Z > \frac{x-54}{20}) = 0.05$ From z table, $\frac{x-54}{20} = 1.645$ x=**87** or  **87**	4 marks
d	Is it rare that a candidate scores less than 30 marks for this examination? Explain why. $P(X<30)$ $P(X < \frac{30-54}{20})$ $P(Z < -1.2)$ $0.5-0.3849$ **0.115** or  **0.115** Since $P(X<30)$ is mire than 5%, it is not rare that a student scores less than 30 marks	5 marks
e	Estimate the interquartile range of the distribution. **Show your workings clearly.** Round your answer to the nearest whole number. **Students have to show the workings clearly and MINITAB express answers not allowed since the question mentioned "Show your workings clearly"** $P(X<x_1) = 0.25$ $P(X < \frac{x_1-54}{20}) = 0.25$ From z table, $\frac{x-54}{20} = -0.67$ $x_1=40.6$ $P(X<x_2) = 0.75$ $P(X < \frac{x_1-54}{20}) = 0.75$ From z table, $\frac{x-54}{20} = 0.67$ $X_2=67.4$ IQR = 67.4-40.6 = **27**	5 marks

Question 3 (20 marks)

3	A study by the Insurance Association concluded that 60% of all travellers from Singapore are covered by travel insurance. A random sample of 25 travellers were surveyed.	
a(i)	What is the probability that 13 of the 25 travellers surveyed are covered by travel insurance? **Show your working clearly.** No marks awarded if you use MINITAB Express $P(X=13)= {}^{25}C_{13}(0.6)^{13}(0.4)^{12}$ $= \mathbf{0.114}$	4 marks
a(ii)	What is the probability that at least 10 of the 25 travellers surveyed are covered by travel insurance?	4 marks

	$P(X \geq 10)$ $= 1 - P(X \leq 9)$ $= 1 - 0.013169$ $= \underline{\mathbf{0.987}}$ or  $\underline{\mathbf{0.987}}$ or  $\underline{\mathbf{0.987}}$	
a(iii)	Is it rare that all 25 travellers surveyed are covered by travel insurance? Explain. $P(X=25)$ $= {}^{25}C_{25}(0.6)^{25}(0.4)^{0}$ $= \underline{\mathbf{0.00000284 \approx 0}}$ or  $\underline{\mathbf{0.000003 \approx 0}}$ or  $\underline{\mathbf{0.000003 \approx 0}}$ Yes, its is rare for all 25 travellers to be covered by travel insurance as the probability is less than 0.05(close to zero)	4 marks
a(iv)	What is the mean number of travellers surveyed who are covered by travel insurance? (show your workings clearly by using a formula) Mean number of travellers covered by the travel insurance = np = 25x0.6 = $\underline{\mathbf{15}}$	2 marks
	The Insurance Association wanted to research further and did the study with 10 focus groups. Each focus group consisted of a random sample of 25 travellers. What is the probability that exactly 9 focus groups will each have at least 10 travellers covered by travel insurance? Derive your answer by filling the blanks below.	
b(i)	Random variable Y: Let Y be the number of focus groups that will each have at least 10 travellers covered by travel insurance	1 mark
b(ii)	Binomial expression of Y: $Y \sim B(10, 0.987)$	1 mark
b(iii)	$P(Y=9)$ $= {}^{10}C_{9}(0.987)^{9}(0.013)^{1}$ $= 0.116$ or  $\underline{\mathbf{0.116}}$ or  $\underline{\mathbf{0.116}}$	4 marks

Question 4 (25 marks)

4	A manufacturing process produces cylindrical component parts for the automotive industry. It is critical that the process produces parts that have a mean diameter of 5.0 mm. A quality control engineer conducts an experiment in which 100 cylindrical component parts produced by the process are randomly selected and the diameter of each part measured. It is known that the population standard deviation, σ is 0.1 mm.	
a(i)	Describe the sampling distribution of the sample mean, if n = 100. Explain if it is Normally distributed. $\mu_{\bar{X}} = 5.0, \sigma_{\bar{X}} = \frac{0.1}{\sqrt{100}} = 0.01$ Since n=100 is large, then by Central Limit Theorem, $\bar{X} \sim N(5.0, 0.01^2)$	4 marks
a(ii)	What is the probability that the sample mean diameter of 100 cylindrical component parts differ from the population mean diameter by at least 0.025 mm? $P(\lvert\bar{X} - \mu\rvert \geq 0.025)$ $P(\bar{X} - \mu \leq -0.025 \text{ or } \bar{X} - \mu \geq 0.025)$ $2 \times P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \geq \frac{0.025}{0.01}\right)$ $2 \times P(Z \geq 2.5)$ $2(0.5-0.4938)$ $\underline{\mathbf{0.0124}}$ or  $0.0062097 \times 2 = \underline{\mathbf{0.0124}}$	6 marks
a(iii)	The mean diameter of the sample taken by the quality control engineer is computed to be 5.027 mm. Using the answer from part (b), what can the quality control engineer conclude about the mean diameter of all the cylindrical component parts produced? Since the probability from part a(ii) is close to zero, it is rare to get a sample mean diameter that differs from the population mean diameter by at least 0.025mm. The engineer's	5 marks

	sample mean diameter is 5.027mm, which is more than 0.025mm away from the population mean diameter. Thus, his claim does not seem correct.	
	Donna, a new quality control worker at a potato chips plant, is investigating the packaging of potato chips marked with a net weight of 28.3 grams. She is tasked to ensure that the machine in the potato chips plant is continuously packing correct weight of potato chips into the bags. Donna carefully weighed the contents of six bags of potato chips. She recorded the weights (in grams), as follows: 29.3 | 28.2 | 29.1 | 28.7 | 28.9 | 28.5 She then keyed the data in Minitab Express and based on the same set of data, she generated two confidence intervals shown in Figure 1 and Figure 2. **1-Sample t: Weight (g)** Descriptive Statistics N Mean StDev SE Mean 95% CI for μ 6 28.7833 0.4021 0.1641 (28.3614, 29.2053) *μ: mean of Weight (g)* **Figure 1** **1-Sample Z: Weight (g)** Descriptive Statistics N Mean StDev SE Mean 95% CI for μ 6 28.7833 0.4021 0.1642 (28.4616, 29.1051) *μ: mean of Weight (g)* *Known standard deviation = 0.4021* **Figure 2** Donna is unsure which confidence interval she should refer to, so she approached you for help.	
b(i)	Explain to Donna under which situation would she refer to the confidence intervals constructed in Figure 1 and Figure 2. CI in Figure 1 – when σ is unknown, sample size is small, and the parent population is Normally distributed CI in Figure 2 – when σ is known or when sample size is large.	1 mark
b(ii)	If Donna refers to the confidence interval constructed in Figure 1, what is the assumption she has to make. The population distribution of weights of potato chips in all the bags is Normally distributed.	1 mark
b(iii)	From part b(ii), explain in context what the confidence interval means. Donna can be 95% confident that the true mean weight of the potato chips in the bag is between 28.3614g and 29.2053g	2 marks
b(iv)	Hence, what can Donna say about the stated net weight of 28.3 grams? It appears that the machine may be packing slightly more chips into the bags than the stated net weight of 28.3g	2 marks
b(v)	Another worker, Darren, approached you with the following interpretation of the confidence interval in Figure 2. ***"95% of all samples will have a mean weight between 28.4633 and 29.1034 grams"*** Is Darren's interpretation correct? Please elaborate. No, because Darren's interpretation seems to indicate a fixed range for CI. But actually the CI varies from sample to sample constructed in the same manner.	2 marks
b(vi)	How will the width of the confidence interval change if Donna had constructed a 90% confidence interval instead? Justify your answer. The width of CI will decrease Lower confidence level → lower margin of error → narrower CL	2 marks