**This section contains multiple choice questions. There is only one correct answer for each question. Write the correct answer in the space provided.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q1:** | Which of the following is **NOT** an example of an application of descriptive statistics? | **Mark (1)** |
|  | Plotting a box plot from a given numerical data set | |
|  | Computing the median of a numerical data set | |
|  | Performing an ANOVA test | |
|  | Computing the standard deviation of a numerical data set | |

Bottom of Form

Top of Form



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q2:** | A numerical data set (sorted in ascending order) is given in Table A1 as follows:   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 1 | 2 | 4 | 4 | 5 | 8 | 9 | 11 | 13 | 13 | | 13 | 14 | 14 | 16 | 17 | 19 | 19 | 19 | 19 | 20 | | 21 | 21 | 22 | 22 | 23 | 24 | 24 | 24 | 24 | 25 | | 25 | 25 | 26 | 27 | 27 | 28 | 29 | 29 | 31 | 34 | | 36 | 39 | 41 | 45 | 48 | 48 | 50 | 51 | 52 | 53 |   Table A1  The data set is summarised in the partially completed box plot in Figure A2.  C:\Users\17046589\AppData\Roaming\Republic Poly\eQuest\_assessmentimages\_assessmentimg_94128856_-1139403939.jpeg  Figure A2 (Not drawn to scale)  Based on the box plot, how many **outliers** are there in the data set? | **Mark (1)** |
|  | More than 2 | |
|  | 0 | |
|  | 2 | |
|  | 1 | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q3:** | Which of the following is one of the parameters of a **binomial** distribution? | **Mark (1)** |
|  | The variance of the distribution. | |
|  | The mean number of occurrences of the event per unit time, *λ*. | |
|  | The total number of trials, *n*. | |
|  | The standard deviation of the distribution. | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q4:** | Which of the following is **NOT** a condition that should be fulfilled before using a **Poisson** distribution to model a practical situation? | **Mark (1)** |
|  | The number of repeated trials are fixed. | |
|  | Events occur independently in any disjoint and small space or time interval, i.e. the process has no ‘memory’. | |
|  | The mean number of occurrences of the event in the time or space interval is constant and known, and proportional to the length of the interval. | |
|  | Events occur randomly within the continuous space or time interval. | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q5:** | Which of the following is a difference between a **binomial** distribution and a **Poisson** distribution? | **Mark (1)** |
|  | The mean and variance of a binomial distribution can be computed with its parameters, the mean and variance of a Poisson distribution cannot be computed with its parameters. | |
|  | A binomial distribution has a finite number of possible values, whereas a Poisson distribution has countably infinite number of possible values. | |
|  | The graph of a binomial distribution can either be symmetrical, left-skewed or right-skewed, whereas the graph of a Poisson distribution is definitely symmetrical. | |
|  | A binomial distribution is a discrete probability distribution whereas a Poisson distribution is a continuous probability distribution. | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q6:** | Which of the following is **NOT** true about a normal distribution? | **Mark (1)** |
|  | The parameter a normal distribution are its mean and its variance. | |
|  | A normal curve is bell-shaped and symmetrical about its mean. | |
|  | The mean, median and mode of a normal distribution are equal. | |
|  | A normal distribution is a discrete probability distribution. | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q7:** | Which of the following is most representative of a **binomial** random variable? | **Mark (1)** |
|  | The service time at a bubble tea shop. | |
|  | The mathematics examination scores of students from a school. | |
|  | The number of whale sightings during a cruise trip. | |
|  | The number of defective electronic component from a batch of 20 components. | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q8:** | Which of the following is most representative of a **Poisson** random variable? | **Mark (1)** |
|  | The weight of male citizens of a country from a certain age group. | |
|  | The lifespan of a light bulb. | |
|  | The number of ‘heads’ obtained out of 10 tosses of a fair coin. | |
|  | The number of cracks on a glass panel for mobile phone screens. | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q9:** | Which of the following is most representative of an **exponential** random variable? | **Mark (1)** |
|  | The number of bull’s eye struck by an archery player out of 10 shots. | |
|  | The height of female citizens of a district from a certain age group. | |
|  | The number of accidents along an expressway for a one-month period. | |
|  | The inter-arrival time of taxis at a taxi stand. | |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q10:** | Which of the following statements about the Central Limit Theorem is **NOT** true? | **Mark (1)** |
|  | Regardless of the population distribution, with a sample size of at least 30, the distribution of sample means will be approximately normal. | |
|  | For a non-normal population, the sample size needs to be at least 30 for the distribution of sample means to be approximately normal. | |
|  | For a normal population, the sample size needs to be at least 30 for the distribution of sample means to be approximately normal. | |
|  | For a non-normal population, the sample size needs to be at least 30 for the distribution of sample sums to be approximately normal. | |

Bottom of Form

**Section: B**

**Write your answers in the blanks provided. Your working need not be shown. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q11:** | Given a numerical data set, % of all data values are **lower** than the **third quartile**. | **Mark (1)** |

Bottom of Form

Top of Form



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q12:** | Identify whether each of the descriptions in Table B1 is a discrete or continuous random variable by filling in the blanks.   |  |  | | --- | --- | | **Description of random variable** | **Discrete or Continuous?** | | The weight of a randomly chosen student from a school. |  | | The lifespan of a certain species of butterfly. |  | | The number of pins knocked down in a game of bowling. |  |   Table B1 | **Mark (3)** |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q13:** | The distributions of normal variables *X* and *Y* are as follows:  *X* ~ N(15, 9), *Y* ~ N(25, 16)  Determine the **mean** and **variance** of normal variable *X* + 2*Y*.  Mean = , Variance = | **Mark (2)** |

Bottom of Form

**Section: C**

**Write your answers in the blanks provided. Your working need not be shown. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q14:** | A numerical data set is given as follows: 36, 42, 49, 51, 54, 56, 62, 64, 68, 87  Compute the 80th percentile of the data set.  Answer: | **Mark (2)** |

Bottom of Form

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q15:** | The time between occurrences of accidents at a factory follows an exponential distribution with a mean of 10 days. What is the probability of no accidents occuring in a 12-day period, given that no accidents occurred over the first 6 days? Round your answer to 3 decimal places.  Answer: | **Mark (2)** |

Bottom of Form

**Section: D**

**Write your answers in the blanks provided. Your working need not be shown. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q16:** | Collin recently started to record the number of steps he has taken each day. After a period of time, Collin wishes to find out whether the mean number of steps taken on each day of the week are the same or significantly different on some days.  He extracted the data from his phone onto a spread sheet and carried out an **ANOVA** test with a 5% significance level. The hypotheses are as follows:  H0: The mean number of steps taken on each day of the week by Collin are the same.  H1: Not all days of the week have the same mean number of steps taken by Collin.  The partial ANOVA output table is shown in Table D1.  C:\Users\17046589\AppData\Roaming\Republic Poly\eQuest\_assessmentimages\_assessmentimg_1941466250_1907155616.jpeg  Table D1: Partial ANOVA output table    a)             How many days of data are recorded and used for the ANOVA test?  Answer:  b)             What is the total degrees of freedom?  Answer:  c)              Compute the value of SS(Tr).  Answer:  d)             Compute the value of MSE.  Answer:  e)             Using a 5% significance level, what is the decision rule for this test? Choose the option that represents the correct decision.  Option A: Reject H0, since the p-value is smaller than the significance level  Option B: Reject H0, since the p-value is larger than the significance level  Option C: Do not reject H0, since the p-value is smaller than the significance level  Option D: Do not reject H0, since the p-value is larger than the significance level  Answer: Option | **Mark (10)** |

Bottom of Form

**Section: E**

**Write your answers in the space provided. Label your answer to each part clearly. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q17:** | A manufacturer of electronic calculators experimented with a new type of circuit for a new calculator model. The desired response time for a typical mathematical computation is 30 milliseconds.  The manufacturer wishes to test whether the mean response time for this new calculator will be more than 30 milliseconds, and uses a **single-sample t-test for population mean**. He conducted a trial of 10 responses and the response times, recorded in milliseconds, are shown in Table E1.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | 21 | 22 | 38 | 25 | 37 | | 24 | 40 | 36 | 32 | 33 |   Table E1: Calculator response times in milliseconds  a)             State one assumption (in the context of the question) needed for this test.  b)             State the null and alternative hypotheses. Let *µ* denote the population mean response time (in milliseconds) of the calculator.  c)              Calculate the value of the test statistic.  d)             Using a 5% significance level, calculate the critical value and state the critical region.  e)             Is the manufacturer able to reject the null hypothesis? Explain. | **Mark (10)** |
|  |  | |
|  | Word Count: 95 | Max Words: 8000 |

Bottom of Form

**Section: F**

**Write your answers in the space provided. Label your answer to each part clearly. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q18:** | The manager of a company observed that certain photocopiers in the company, among the total of 5 photocopiers (A, B, C, D and E), tend to break down rather frequently. Due to the difference in years of usage, he believed that the breakdown frequencies of photocopiers A, B, C, D and E are in the ratio of **3:2:1:1:1**.  He tasked one of his staff to record and collate the breakdown frequency of each photocopier over a period of one year. The collated data is shown in Table F1.   |  |  | | --- | --- | | **Photocopier** | **Number of breakdowns recorded** | | A | 15 | | B | 9 | | C | 8 | | D | 10 | | E | 8 |   Table F1: Breakdown frequencies of photocopiers in the company  The manager conducted a **Chi-square goodness-of-fit test** with the following hypotheses:  H0: The breakdown frequencies of photocopiers A, B, C, D and E are in the ratio of **3:2:1:1:1**.  H1: The breakdown frequencies of photocopiers A, B, C, D and E are not in the ratio of **3:2:1:1:1**.  a)             State one assumption needed to conduct the Chi-square goodness-of-fit test.  b)             Compute the expected breakdown frequencies of photocopiers A, B, C, D and E.  c)              State the degree of freedom for the test.  d)             Calculate the Chi-square test statistic.  e)             Calculate the p-value for the test.  f)               Using a 10% significance level, can the null hypothesis be rejected? Explain. | **Mark (10)** |
|  |  | |
|  | Word Count: 142 | Max Words: 8000 |

Bottom of Form

**Section: G**

**Write your answers in the space provided. Label your answer to each part clearly. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q19:** | A recent study shows that 40% of households in a certain country experienced at least one break-in a year. Fiona, who works in the research department of the police force, wishes to test whether the household break-in proportion for one of the districts is significantly **higher** than that of the national proportion.  She proceeded to carry out a **single-sample z-test for population proportion**. Out of 50 households in the district interviewed, 28 experienced at least one break-in over the last one year.  a)             State the null and alternative hypotheses. Let *p* denote the population proportion of households in the district that experienced at least one break-in a year.  b)             Calculate the value of the test statistic.  c)              Using a 5% significance level, calculate the critical value and state the critical region.  d)             Can the null hypothesis be rejected?  e)             Suppose that another 50 households in the district are interviewed. What is the range of values of the number of households that experienced at least one break-in over the last one year for the decision to be different from that in part (d)? Show your working clearly. | **Mark (10)** |
|  |  | |
|  | Word Count: 135 | Max Words: 8000 |

Bottom of Form

**Section: H**

**Write your answers in the space provided. Label your answer to each part clearly. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q20:** | The human resource (HR) manager of a Food and Beverage (F&B) company would like to assess whether the employees of two different outlets, A and B, have similar customer service feedback ratings. A **two-sample pooled t-test for the difference between two population means** was conducted using a 10% significance level. The Excel output table for a two-sample t-test assuming equal variances, is shown in Table H1.  C:\Users\17046589\AppData\Roaming\Republic Poly\eQuest\_assessmentimages\_assessmentimg_-1876985550_-563560251.jpeg  Table H1: Excel output table for two-sample t-test assuming equal variances  a)             State 2 reasons why applying a two-sample **Z-**test for the difference between two population means would be **inappropriate** for this case.  b)             State the null and alternative hypotheses. Let *µ*A – *µ*B denote the difference between the two population means, where *µ*A and *µ*B represent the population mean customer service feedback rating of employees from Outlet A and Outlet B, respectively.  c)              Using the critical region method, which two values from Table H1 are to be used for the decision rule (that is, to decide whether or not to reject the null hypothesis)?  d)             Using a 10% significance level, can the null hypothesis be rejected?  One assumption needed for the two-sample pooled t-test for the difference between two population means is that the two population variances are equal.  e)             Using a general rule of thumb, explain why this assumption is valid for the test carried out in this section.  f)               What is a suitable hypothesis test to conduct before deciding whether a two-sample pooled t-test for the difference between two population means or a two-sample t-test with unequal variances is more appropriate? | **Mark (10)** |
|  |  | |
|  | Word Count: 134 | Max Words: 8000 |

Bottom of Form

**Section: I**

**Write your answers in the space provided. Label your answer to each part clearly. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q21:** | a)             State two differences between the graphs of the F-distribution and the t-distribution.  b)             State two assumptions needed to conduct the two-sample F-test on population variances.  c)              What are the two parameters of an F-distribution?  Herman is a production engineer of a company that produces mineral water. The company is considering to purchase a new mineral water filling machine. Herman is tasked to study the performance of the new machine as compared to an existing one.  He proceeded to conduct a **two-sample F-test on population variances** to test whether the variance of the volume of water per bottle filled by the new machine is higher than that of the existing machine, using a 5% significance level. The hypotheses are as follows, where σE2 and σN2 denote the population variance of the volume of water per bottle filled by the existing machine and by the new machine, respectively.  H0: σE2 = σN2  H1: σE2 < σN2  He obtained a trial sample using the new machine and a sample using the existing machine. The Excel output table is shown in Table I1.  **C:\Users\17046589\AppData\Roaming\Republic Poly\eQuest\_assessmentimages\_assessmentimg_2047540304_2087616287.jpeg**  Table I1: Excel output table for two-sample F-test on population variances    d)             Using a 5% significance level, can the null hypothesis be rejected? Explain.  e)             Based on the decision in part (d), should the company purchase the new filling machine? Why or why not? | **Mark (10)** |
|  |  | |
|  | Word Count: 134 | Max Words: 8000 |

Bottom of Form

**Section: J**

**Write your answers in the space provided. Label your answer to each part clearly. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |
| --- | --- | --- |
| **Q22:** | As a process engineer, Ian is tasked to look into the current production processes and identify areas for improvement. He suspected that the current production line setup leads to longer production time and proposed a new setup to **reduce** the production time.  With the current production line setup, he recorded 10 operators’ production time for one particular product. Three months after the new setup, he recorded the 10 operators’ production time for the same product.  Ian conducted a **paired t-test for the population mean of differences** using a 5% significance level. The Excel output table is shown in Table J1, where ‘**Current**’ represents the production time recorded with the **current** setup and ‘**New**’ represents the production time recorded with the **new** setup.  **C:\Users\17046589\AppData\Roaming\Republic Poly\eQuest\_assessmentimages\_assessmentimg_650416741_1709559256.jpeg**  Table J1: Excel output table for paired t-test for the population mean of differences    a)             State one reason why applying a paired t-test for the population mean of differences is appropriate for this case.  b)             State the null and alternative hypotheses. Use *µ*D to denote the population mean of differences in production time. In this test, we denote the production time difference for each operator as “**Production time recorded with the current setup – Production time recorded with the new setup**”.  c)              State the p-value for this test.  d)             Using a 5% significance level, can the null hypothesis be rejected?  e)             Write down a formal conclusion in the context of the question.  f)               State whether type I or type II error is relevant to the conclusion of the test. State the error in the context of the question. | **Mark (10)** |
|  |  | |
|  | Word Count: 109 | Max Words: 8000 |

Bottom of Form

**Section: K**

**Write your answers in the space provided. Label your answer to each part clearly. Give non-exact numerical answers correct to 2 decimal places unless a different level of accuracy is specified in the question.**

Top of Form



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q23:** | a)             State two advantages of using a sign test in hypothesis testing.  Jerome, a student from a polytechnic engineering course, was assigned to a local airport’s check-in department for his internship. Besides the daily check-in counter duties, he was given a project to review the current work processes and suggest areas for improvements.  Based on his observations, at certain periods of the day, the queue for check-in tends to be quite long. He wishes to conduct a study on the service time at the check-in counters. From his observations, he estimated that the median service time at the check-in counters is **longer** than 5 minutes.  Jerome proceeded to conduct a **single-sample sign test** with the following hypotheses, where *m* denotes the median service time (in minutes) at the check-in counters.  H0: *m* = 5  H1: *m* > 5  He obtained a sample of 12 service times, which is shown in Table K1.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 4 min 43 sec | 4 min 19 sec | 6 min 15 sec | 6 min 37 sec | 5 min 25 sec | 6 min 23 sec | | 5 min 19 sec | 5 min 48 sec | 5 min 42 sec | 4 min 32 sec | 6 min 9 sec | 5 min 32 sec |   Table K1: A sample of service times    b)             State the total number of signs, *n*.  c)              State the number of positive signs, *r*.  d)             Compute the p-value.  e)             Using a 5% significance level, can the null hypothesis be rejected? Explain.  f)               Suppose the final service time in the sample is amended from 5 min 32 sec to 8 min 20 sec. Will this impact the conclusion of the test in part (e)? Why or why not? | **Mark (10)** |
|  |  | |

Bottom of Form