

National College of Ireland

**Higher Diploma in Data Analytics – Year 1 – HDSDASEP1/HDSDASEP2/HDSDASEPOL**

**BSc in Technology Management – Year 4 – BSHTM**

**BSc in Computing – Year 4 – BSHCDA/BSHCDAE**

**BSc in Business Information Systems – Year 4 – BSHBIS**

**Semester One Examinations – 2018/19**

**10th January, 2019**

**2.00pm - 4.00pm**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Business Data Analysis**

Dr Andrew McCarren,

Dr Shazia Afzal,

Dr Stephen O’Shaughnessy,

Mr Brian Buckley,

Mr David O’Dwyer,

Dr Eugene O’Loughlin.

Answer any four questions.

**Duration of exam:** 2 hours

**Attachments:** None

Calculators are permitted in this exam.

1. Portugal has a number of popular tourist destinations during the summer months, including the Algarve, in the far south, and Cascais, just west of Lisbon. A travel agent has visited the websites of several Portuguese estate agents and compiled information about the daily rent for 12 holiday properties in the Algarve, and 11 in Cascais. The daily rent for each property is as follows:

|  |  |
| --- | --- |
| **Algarve** | **Cascais** |
| €64 | €62 |
| €63 | €77 |
| €75 | €69 |
| €66 | €72 |
| €83 | €51 |
| €80 | €65 |
| €53 | €64 |
| €73 | €53 |
| €79 | €36 |
| €56 | €57 |
| €68 | €60 |
| €62 |  |

Using the sample data above, answer the following questions:

1. Specify the null and alternate hypotheses to determine if there is a significant   
   difference in the mean daily rents between the Algarve and Cascais.

*(3 Marks)*

1. Specify and justify an appropriate probability of committing a Type I error (α).

*(2 Marks)*

1. Calculate the test statistic to test your hypotheses and report your result.

*(10 Marks)*

1. Report your decision and clearly explain your result.

*(10 Marks)*

*(Total 25 marks)*

1. Many high schools in the United States prepare their University bound students for the Scholastic Aptitude Test (SAT) with test-preparation classes in which students take timed practice exams in the hope of reducing the time taken to complete the tests. In one school, the Principal is considering adopting the latest cutting-edge test-preparation software to assist students as they get ready for their SATs. However, before she purchases the new software, she wants to know if it really makes a difference in how the students perform.

A random sample of 12 students is selected from the populations of all students who are preparing for the SAT test in the school. Each student takes a practice test: first using the old software, and then a week later using the new software. The time in minutes for each student to complete the two tests are recorded as follows:

|  |  |
| --- | --- |
| **Time Old Software** | **Time New Software** |
| 60 | 58 |
| 52 | 48 |
| 41 | 43 |
| 58 | 54 |
| 62 | 59 |
| 58 | 53 |
| 54 | 50 |
| 64 | 61 |
| 67 | 64 |
| 64 | 63 |
| 72 | 69 |
| 53 | 51 |

Using the sample data above, answer the following questions:

1. Specify the null and alternate hypotheses to determine if there is a significant   
   difference in mean test scores between the old and the new software.

*(3 Marks)*

1. Specify and justify an appropriate probability of committing a Type I error (α).

*(2 Marks)*

1. Calculate the test statistic to test your hypotheses and report your result.

*(10 Marks)*

1. Report your decision and clearly explain your result.

*(10 Marks)*

*(Total 25 marks)*

1. A review of animal-assisted therapy in childhood mental health found that in several studies there was a positive effect on happiness when children were given a puppy to play with.

Assume that you have conducted an experiment to test if this finding is true with adults. You divide 15 randomized adults into three groups: a control group where no puppies are provided, a second group whose members each get a puppy for 30 minutes, and a third group whose members each get a puppy for 60 minutes. Happiness in each individual is measured on a scale of 0 (extremely unhappy) to 10 (extremely happy). The results were recorded as follows:

|  |  |  |
| --- | --- | --- |
| **Control** | **30 Minutes** | **60 Minutes** |
| 3 | 5 | 7 |
| 2 | 2 | 4 |
| 1 | 4 | 5 |
| 1 | 2 | 3 |
| 4 | 3 | 6 |

Using the sample data above, answer the following questions:

1. Specify the null and alternate hypotheses to determine if there is a significant   
   difference between the mean happiness of the three puppy therapy groups.

*(3 Marks)*

1. Specify an appropriate probability of committing a Type I error (α).

*(2 Marks)*

1. Calculate the test statistic to test your hypotheses and report your result in an   
   ANOVA table.

*(10 Marks)*

1. Report your decision and clearly explain your result.

*(10 Marks)*

*(Total 25 Marks)*

1. a)  
     
   On average, the expected mix of each variety of M&Ms Milk Chocolate sweets in a packet follows   
   the following colour percentage breakdown:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Brown** | **Yellow** | **Red** | **Orange** | **Green** | **Blue** |
| 30% | 20% | 20% | 10% | 10% | 10% |

Assume that you have purchased a packet of M&Ms and observed that there were 58 M&Ms   
in the packet, with the colour breakdown as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Brown** | **Yellow** | **Red** | **Orange** | **Green** | **Blue** |
| 15 | 14 | 13 | 8 | 4 | 4 |

Based on these data, are these different enough to conclude a difference exists between the   
observed and expected percentage colours of M&Ms?

* 1. Specify the null and alternate hypotheses to test if a difference exists.

*(2 marks)*

* 1. Calculate the test statistic to measure the discrepancy (if any) between the   
     observed and the expected results.

*(6 marks)*

* 1. Interpret and explain your results.

*(7 marks)*

*(Sub-total 15 Marks)*

b)  
  
According to the Irish Nurses and Midwives Organization the daily total of the number of patients waiting on trollies in the Republic of Ireland for the first two weeks of October 2018 (excluding weekends) is as follows:

|  |  |
| --- | --- |
| **Date & Time** | **Patients on Trolleys** |
| 01 October 2018 | 322 |
| 02 October 2018 | 349 |
| 03 October 2018 | 377 |
| 04 October 2018 | 370 |
| 05 October 2018 | 341 |
| 08 October 2018 | 430 |
| 09 October 2018 | 525 |
| 10 October 2018 | 518 |
| 11 October 2018 | 483 |
| 12 October 2018 | 397 |

Assume that you have been tasked with making a forecast of the number of patients on trollies for   
15th October (exclude weekend of 13/14 October) based on these figures. Calculate each of the following:

1. Use a four-point simple moving average to estimate the number of patients waiting on trollies for 15th October, 2018.

(1 Mark)

1. Use a three-point weighted moving average to estimate the number of patients   
   waiting on trollies for 15th October, 2018 (use weightings of 4 for the most recent   
   day, 3 for the day before, and 2 for the remaining day).

(1 Mark)

1. Use Exponential Smoothing to estimate number of patients waiting on trollies for   
   15th October, 2018. Use a smoothing constant (α) of 0.5, assume a forecasted value   
   of 320 for 1st October, 2018.

(4 Marks)

1. Interpret and briefly explain your results.

(4 Marks)

(Sub-total 10 Marks)

(Total 25 Marks)

5

1. Assume that you are conducting a study to investigate the relationship between monthly e-commerce   
   sales and online advertising costs for seven online stores in the last year. The sales and advertising   
   costs for the seven stores are as follows:

|  |  |  |
| --- | --- | --- |
| **Online Store** | **Monthly E-commerce Sales (1000s)** | **Online Advertising (1000s)** |
| 1 | €368 | €1.7 |
| 2 | €340 | €1.5 |
| 3 | €665 | €2.8 |
| 4 | €954 | €5.0 |
| 5 | €331 | €1.3 |
| 6 | €556 | €2.2 |
| 7 | €376 | €1.3 |

Using the Simple Linear Regression model, calculate the following for the above data:

1. The correlation coefficient (r).

(9 marks)

1. The slope of a line of best-fit for this data.

(2 marks)

1. The y intercept.

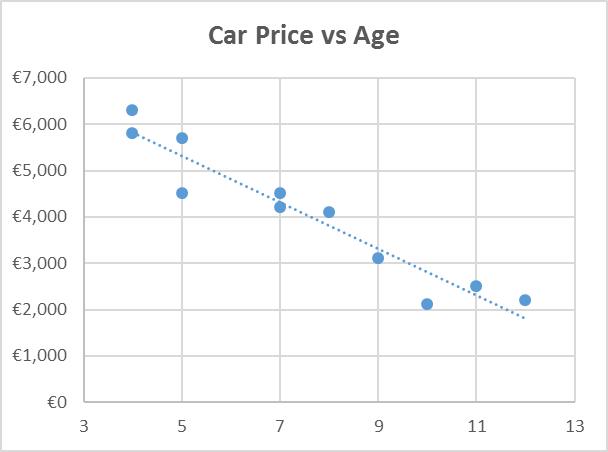
(2 marks)

1. The value of y (Sales in €000s)) given a value of x (Advertising in €000s) = €2.5.

(2 marks)

(Sub-total 15 marks)

1. You have to examine the relationship between the age and price for used cars sold in the   
   last year by a car dealership company. The following table and scatter chart shows the age   
   and price achieved for 11 cars:



|  |  |
| --- | --- |
| **Car Age (years)** | **Car Price** |
| 4 | €6,300 |
| 4 | €5,800 |
| 5 | €5,700 |
| 5 | €4,500 |
| 7 | €4,500 |
| 7 | €4,200 |
| 8 | €4,100 |
| 9 | €3,100 |
| 10 | €2,100 |
| 11 | €2,500 |
| 12 | €2,200 |

The simple linear regression equation for these data is y = 7836.3 - 502.42x and the correlation coefficient is r = -0.955.

What can you tell about the relationship between these two variables?

(Sub-total 10 marks)

(Total 25 Marks)

**Statistical Formulas & Tables**

|  |  |
| --- | --- |
| Sample variance |  |
| Sample Standard Deviation |  |
| Z statistic |  |
| Test statistic (F) for the difference between two population variances | F = larger sample variance (s2)  smaller sample variance (s2) |
| Pooled Variance |  |
| t statistic (Unpaired/Independent test) – Equal Variance | Degrees of Freedom = |
| t statistic (Unpaired/Independent test) – Unequal Variance | Degrees of Freedom = |
| t statistic (Paired/Dependent test) | or  Degrees of Freedom |
| ANOVA - Within Groups Mean Sum of Squares |  |
| ANOVA - Between Groups Mean Sum of Squares |  |
| ANOVA - F statistic |  |
| Chi-squared Formula | or |
| Correlation Coefficient (r) | or |
| Linear Regression Function | or |
| Slope (b) of Regression Line | or |
| Y-Intercept (a) of Regression Line | or |
| Simple Moving Average | Ft = At-1 + At-2 + At-3  n |
| Weighted Moving Average | Ft = w1At-1 + w2At-2 + w3At-3  Sum of Weights |
| Exponential Smoothing | *Ft* = α *At*−1 + (1 − α)*Ft*−1 |

**Statistical Tables**

**ANOVA Table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | Df | SS | MS | F | Sig(P) |
| Treatment |  |  |  |  |  |
| Residual error |  |  |  |  |  |
| Total |  |  |  |  |  |

