Last Name: _____**ANSWERS**_____

First Name: _____

Student Number: _____

Instructions:

1. Print your name and student number above AND on the MC answer sheet. A test or MC answer sheet without a name and student number won't be marked. A page (except the front page) without a name at the top right, **PARTICULARLY PAGE 4**, where shown, won't be marked.
2. Use only a pencil when filling in the MC answer sheet for the multiple choice questions. Circle the correct answers on your question paper first and only when you are certain of your answer fill in the MC answer sheet. Only the answers found on the MC answer sheet will be used when marking the multiple choice questions.
3. Check that you have all 5 PAGES before beginning the exam.
4. Pace yourself – you have ~ 45 minutes.
5. Use the blank spaces on exam pages for rough work. No scrap paper is permitted.
6. Simple calculators (not cell phones or tablets) are allowed.
7. If you have a cell phone or any electronic device (other than a pacemaker) with you – be sure it is turned off now, and stored in a safe place away from your desk.
8. Hand in BOTH this exam booklet AND the MC answer sheet. Taking an exam booklet from the exam room will result in an automatic grade of 'F' for this course.
9. This test is worth 17.5% of your final mark.

I have read, understood, and will comply with all of the above instructions:

_____ _____

sign your full name here date

Choose the **best** answer in each of the following multiple-choice questions**. 1 mark each**
Formulas and expressions are those from the course book used in class. ***Formulas are at the end.***

1. What is a difference between interpolation and extrapolation?
a. extrapolation is within the range of data points, interpolation is outside
b. interpolation is within the range of data points, extrapolation is outside
c. interpolation is between data points, extrapolation is exactly at data points
d. extrapolation is between data points, interpolation is exactly at data points
e. there is no difference - they are just two terms that mean the same thing

2. A fact about the Lagrange Interpolation function y = f(x) is
a. it must be a converging series b. it's a series of polynomials of powers of y
c. it uses the derivatives at x = 0 d. it's a series of polynomials of powers of x e. it uses the derivatives at x = a

3. For three points in a data set, Lagrange Interpolation would generate a function of the forma
a. ax^3+bx^2+c b. ax^2+bx+c c. ax^2+b d. ax^3+bx^2+cx+d e. ax^3+bx+c

4. Using the formula at the end for the j^{th} term in the Lagrange Interpolation function, which of the following is the basis function of the 3^{rd} term?
a. $(x-x_2)(x-x_1)/(x_3-x_1)(x_3-x_2)$ b. $(x-x_1)(x-x_2)/(x_3-x_1)(x_3-x_2)$ c. $(x-x_2)(x-x_1)/(x_1-x_2)(x_1-x_3)$
d. $(x-x_3)(x-x_2)/(x_3-x_1)(x_3-x_2)$ e. $(x-x_3)(x-x_1)/(x_2-x_1)(x_3-x_1)$

5. Which of the following statements about the Lagrange interpolation function f(x) is not true?
a. each basis function passes exactly through its control point
b. each basis function is zero at all control points except its own
c. the interpolation function passes exactly through all control points
d. the highest power of x is one less than the number of control points
e. the number of basis functions is one less than the number of control points

6. What is a problem with Lagrange interpolation f(x)?
a. the method needs more than 3 data points b. the method cannot be implemented in code
c. the method doesn't work with negative values of x
d. the data points must be equally spaced in x for the method to work
e. data with fluctuations introduces fluctuations in the interpolation function where there shouldn't be any

The following data is a summary of weekly hours Internet usage for a sample of 50 users. It is used in questions 7 to 10. The frequency and relative frequency for 20-24 hours are blanked out for use in questions 8 and 9.

Estimated Hours on Internet	*Frequency*	*Relative Frequency (%)*
0–4	2	4
5–9	9	18
10–14	19	38
15–19	11	22
20–24	6	12
25–29	3	6
Total	50	100

7. The range in each of the rows of the "*Estimated Hours on Internet*" column is called
a. standard deviation b. class width c. variance d. mean e. residual

8. What is the frequency for internet hours 20-24
a. 3 b. 2 c. 1 d. 0 e. none of the previous

9 What is the relative frequency (%) for internet hours 20-24
d. 50 c. 12 b. 6 a. 3 e. none of the previous

10. Where fr means relative frequency and f means frequency, what is the value of the ratio $\frac{\Sigma f}{\Sigma fr}$
e. 100 d. 50 c. 2 b. 1 a. 1/2

11. The number corresponding to the horizontal position of the middle of each rectangle in a histogram is called
a. frequency b. class mark c. class width d. population e. sample

12. If f_i represents the frequency of class i, which of the following is the value of the expression $\sum f_i$?
a. size of the population b. size of the sample c. 0 d. 0.5 e. 1.0

13. Where f means frequency, what is the value of the ratio $\frac{\sum xf}{\sum f}$
a. standard deviation b. total sample size c. total population d. arithmetic mean
e. it has no special meaning because it's wrong

14. A set of statistical data has a population standard deviation of 100.0.
What sample size would you choose so that sample means have a standard deviation of 20.0?
a. 100 b.25 c. 16 d. 10 e. 5

15 What is not true in the following?
a. the mean of residuals is always zero b. the sum of the residuals is always zero
c. the population standard deviation is always smaller than the sample standard deviation
d. a sample is smaller than the population e. the standard deviation can be either positive or negative

16 Which of the following is an expansion of the expression $\sum(x-\mu)^2$
a. $(x_1^2-\mu^2) + (x_2^2-\mu^2) + \ldots$ b. $(x_1^2-\mu_1^2) + (x_2^2-\mu_1^2) + \ldots$ c. $(x_1-\mu_1)^2 + (x_2-\mu_2)^2 + \ldots$ d. $(x_1-\mu)^2 + (x_2-\mu)^2 + \ldots$
e. $(x-\mu_1)^2 + (x-\mu_2)^2 + \ldots$

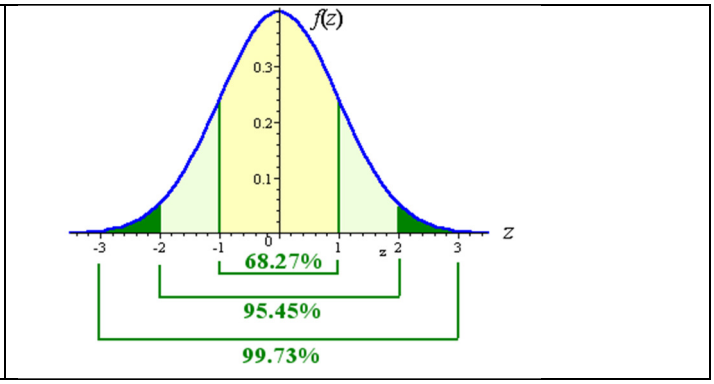
The following data sample is used in questions 17 and 18

x	-2.0	0.0	2.0	3.0	4.0

17. The value of the expression $\sum x$ is
a. 15 b. 9 c. 7 d. 0 e. none of the previous

18. The mean of the sample is closest to which of the following?
a. 3.0 b. 2.5 c. 2.0 d. 1.5 e. 0.5

The standard normal distribution on the right is used in questions 19 and 20. The symbols A, B and C represent the areas under the curve as percentages.



19. Which of the following represents the value of area A?
a. 34% b. 68% c. 48% d. 49.9% e. 96%

20. Which of the following represents the value of area B?
a. 34% b. 68% c. 48% d. 49.9% e. 96%

Short Questions

21. [10 marks]. Interpolate the following data at x = 3.0 using Lagrange polynomials:

x	y
0.0	3.0
1.0	0.0
4.0	1.0

The Lagrange polynomial is:

$$y = f(x) = \frac{(x-1)(x-4)}{(0-1)(0-4)}3.0 + \frac{(x-0)(x-4)}{(1-0)(1-4)}0.0 + \frac{(x-0)(x-1)}{(4-0)(4-1)}1.0$$

at x = 3

$$y = f(x) = \frac{(3-1)(3-4)}{(0-1)(0-4)}3.0 + \frac{(3-0)(3-4)}{(1-0)(1-4)}0.0 + \frac{(3-0)(3-1)}{(4-0)(4-1)}1.0$$

= -3/2 + 0 +1/2 = -1.0

22. [10 marks] *After analysing data for a long period of time, it was determined that for samples of 1000 the lifetimes of a computer keyboard are distributed normally, with a mean lifetime $\mu = 10.0$ years and a sample standard deviation $\sigma = 2.0$ years.*

22.1.[5 marks] In a sample, how many keyboards are expected to last more than 11 years?
For 11.0, z = (11.0 – 10.0)/2.0 = 1/2 for which A = 0.1915

So total area above this = 0.5 - 0.1915 = 0.3085
Hence for a sample of 1000, the number in this range = 1000*0.3085 = 309

22.2.[5 marks] In a sample, how many keyboards last between 9 and 11 years?
For 11.0, z = (11.0 – 10.0)/2.0 = 1/2 for which A = 0.1915
For 9.0, z = (9.0 – 10.0)/2.0 = -1/2 for which A = 0.1915

Total area above = 2*0.0.1915 = 0.383
Hence for a sample of 1000, the number in this range = 1000*0.383 = 383

Formulas

Lagrange Interpolation

$$f(x) = \frac{(x-x_2)(x-x_3)\ldots(x-x_N)}{(x_1-x_2)(x_1-x_3)\ldots(x_1-x_N)}y_1 + \frac{(x-x_1)(x-x_3)\ldots(x-x_N)}{(x_2-x_1)(x_2-x_3)\ldots(x_2-x_N)}y_2 +$$
$$\ldots\ldots\ldots\ldots + \frac{(x-x_1)(x-x_2)\ldots(x-x_{N-1})}{(x_N-x_1)(x_N-x_2)\ldots(x_N-x_{N-1})}y_N$$

The jth term in the Lagrange Interpolation function

$$P_j(x) = y_j \prod_{k=1k\neq j}^{n} \frac{x-x_k}{x_j-x_k}$$

Statistics

Mean $\mu = \frac{\sum_1^n x_i}{N}$

Standard deviation of population N = $\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{N}}$

Standard deviation of sample N = $\sqrt{\frac{\Sigma(x_i-\mu)^2}{N-1}}$

Standard deviation of sample mean = $\frac{\sigma}{\sqrt{N}}$

Area under a Standard Normal Distribution

Standard score z (or z score) is defined as $z = \frac{x-\mu}{\sigma}$

Table 22.1 Standard Normal (z) Distribution

z	Area	z	Area	z	Area
0.0	0.0000	1.0	0.3413	2.0	0.4772
0.1	0.0398	1.1	0.3643	2.1	0.4821
0.2	0.0793	1.2	0.3849	2.2	0.4861
0.3	0.1179	1.3	0.4032	2.3	0.4893
0.4	0.1554	1.4	0.4192	2.4	0.4918
0.5	0.1915	1.5	0.4332	2.5	0.4938
0.6	0.2257	1.6	0.4452	2.6	0.4953
0.7	0.2580	1.7	0.4554	2.7	0.4965
0.8	0.2881	1.8	0.4641	2.8	0.4974
0.9	0.3159	1.9	0.4713	2.9	0.4981
1.0	0.3413	2.0	0.4772	3.0	0.4987