



OFFICE OF THE DEPUTY PRINCIPAL

ACADEMICS, STUDENT AFFAIRS AND RESEARCH

UNIVERSITY EXAMINATIONS

2020 /2021 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER REGULAR EXAMINATION

**FOR THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS WITH
COMPUTING)**

COURSE CODE: STA 321

COURSE TITLE: TESTING OF HYPOTHESIS

DATE: 22/7/2021

TIME: 1300-1600HRS

INSTRUCTION TO CANDIDATES

- SEE INSIDE

THIS PAPER CONSISTS OF 5 PRINTED PAGES

PLEASE TURN OVER

REGULAR-MAIN EXAM
STA 321: TESTING OF HYPOTHESIS

STREAM: ASC

DURATION: 3 Hours

INSTRUCTION TO CANDIDATESAnswer **ALL** questions from section A and any **THREE** from section B.**SECTION A [31 Marks]. Answer ALL questions.****QUESTION ONE (16 MARKS)**

- a) Define the following terms as used under test of hypothesis
- | | |
|------------------------|----------|
| i) Type I error | [1 mark] |
| ii) Type II error | [1 mark] |
| iii) Power of the test | [1 mark] |
- b) Outline the procedure of hypothesis testing [3 marks]
- c) A counsellor wants to show that for men who are married by the time they are 30, μ = average age when the men are married *is not* 21 years old. A random sample of 10 men who were married by age 30 showed an average age at marriage of 22.2, with a sample standard deviation of 1.9 years. Assume that the age at which this population of men get married for the first time is normally distributed.
- | | |
|--|-----------|
| i) What are the appropriate null and alternative hypotheses? | [2 marks] |
| ii) Calculate the value of the test statistic | [3 marks] |
| iii) Calculate the p-value | [2 marks] |
| iv) For a significance level of $\alpha = 0.05$, are the results statistically significant? | [1 mark] |
| v) Provide an appropriate conclusion? | [2 marks] |

QUESTION TWO (15 MARKS)

- a) According to the Centers for Disease Control, the number of cigarettes smoked per day by individuals in the general population who are daily smokers is 18.1. Do retired adults who are daily smokers smoke less than the general population of daily smokers? To answer this question, a researcher obtains a random sample of 40 retired adults who are current daily smokers and records the number of cigarettes smoked on a randomly selected day. The data result in a sample mean of 16.8 cigarettes and a SD of 4.7 cigarettes.
- | | |
|--|-----------|
| i) Formulate the null and the alternative hypothesis | [2 marks] |
| ii) Is there sufficient evidence at $\alpha = 0.05$ level of significance to conclude that retired adults who are daily smokers smoke less than the general population of daily smokers? | [4 marks] |

b) The president of a large University has been studying the relationship between male/female supervisory structures in this institution and employee job satisfaction. The results of a recent survey are shown in the table below. Conduct a test at the 5% significance level to determine whether job satisfaction depends on the boss/employee gender relationship.

Level of satisfaction	Male/female	Male/female	Male/male	Female/female	Total
Satisfied	60	15	50	15	140
Neutral	27	45	48	50	170
Dissatisfied	13	32	12	55	112
Total	100	92	110	120	422

- i) Formulate the null and the alternative hypothesis [2 marks]
- ii) Conduct a test at the 5% significance level to determine whether the level of job satisfaction depends on the boss/employee gender relationship. [5 marks]
- iii) What is your conclusion? [2 marks]

SECTION B (39 MARKS)

QUESTION THREE (13 MARKS)

a) Suppose you want to test a null hypothesis of a single observation x drawn from a normal population distribution with a mean μ and std. deviation $\sigma = 5$. The hypothesis $\mu = 20$ is not rejected if the observed value $x < 26$ otherwise the alternative $H_A: \mu = 30$ is accepted. Evaluate the probability of type II error & hence compute the power of the test [6 marks]

b) Suppose that Y is a single observation (i.e., an iid sample of size $n=1$) from an exponential distribution with mean θ . using this single observation, we would like to test

$$H_0: \theta = 2,$$

$$H_A: \theta = 3.$$

Use the Neyman-Pearson lemma to find the most powerful level $\alpha = 0.05$ [7 marks]

QUESTION FOUR (13 MARKS)

a) People at high risk of sudden cardiac death can be identified using the change in a signal averaged electrocardiogram before and after prescribed activities. The current method is about 85% accurate. The method was modified, hoping to improve its accuracy. The new method is tested on 50 people and gave correct results on 45 patients (take $\alpha=0.05$)

- i) Formulate the null and the alternative hypothesis [2 marks]
- ii) Is this convincing evidence that the new method is more accurate? [3 marks]
- iii) If the new method has 95% accuracy, what power does a sample of 50 has, to demonstrate that the new method is better? [5 marks]

b) Suppose a customer advocacy group would like to conduct a survey to find the proportion p of consumers who bought the newest generation of computers were happy with their purchase. How large a sample should they take to estimate p with 2% margin of error and 95% confidence? [3 marks]

QUESTION FIVE (13 MARKS)

Two machines are used for filling plastic bottles with a net volume of 16.10 ml. The fill volume can be assumed normal, with a standard deviation of $\sigma_1 = 0.020$ ml and $\sigma_2 = 0.025$ ml for machines 1 and 2 respectively. A member of the quality engineering staff suspects that both machines fill to the same mean net volume, whether or not this volume is 16.10mL. A random sample of 10 bottles is taken from the output of each machine. Take $\alpha = 0.05$.

Machine 1	16.012	16.014	16.012	16.015	16.013	16.013	16.015	16.013	16.012	16.016
Machine 2	16.009	16.006	16.004	16.007	16.005	16.007	16.005	16.004	16.008	16.005

- a) Formulate the null & the alternative hypothesis [2 marks]
- b) Do you think that the engineer is correct? [5 marks]
- c) What is the p-value for this test? [3 marks]
- d) Calculate a 95% confidence interval on the difference in means. Provide a practical interpretation of this interval [3 marks]

QUESTION SIX (13 MARKS)

a) Statistics students attempting to establish a relationship between hours of study in the week immediately preceding a major midterm exam and the marks (test score) received on the exam to prepare themselves for future exams. They gathered the data listed below from last years in class survey.

Hours of study	25	12	18	26	19	20	23	15	22	8
Exam score	93	57	55	90	82	95	95	80	85	61

- i) Make a sketch of the scatter plot to examine the relationship [2 marks]
- ii) Estimate the regression model to help predict the exam [4 marks]
- iii) Interpret the slope coefficient [2 marks]
- iv) Predict the exam score when the hours of study is 35 [2 marks]
- v) Write a R code that inputs the data and conducts the regression analysis. [3 marks]

b) An article in the Journal of Agricultural Science investigated means of wheat grain crude protein content (CP) and Hagburg falling number (HFN) surveyed in the UK. The analysis used a variety of nitrogen fertilizer applications (kg N/ha), temperature ($^{\circ}$ C), and total monthly rainfall (mm). The data shown below describe temperatures for wheat grown at Harper Adams Agricultural college between 1982 and 1993. The temperatures measured in June were obtained as follows: {15.2, 14.2, 14.0, 12.2, 14.4, 12.5, 14.3, 14.2, 13.5, 11.8, 15.2}. Construct a 99% two-sided confidence interval on the mean temperature. [3 marks]

QUESTION SEVEN (13 MARKS)

To test the hypothesis that the mean systolic blood pressure in a certain population equals 140 mmHg. The standard deviation has a known value of 20 and a data set of 55 patients is available (Table 1).

Table 1 Blood Pressure data

DATA: Systolic blood pressure (mmHg) of 25 healthy subjects (status=0) and 30 subjects with hypertension (status=1).

No.	Status	mmHg	No.	Status	mmHg
1	0	120	29	1	127
2	0	115	30	1	141
3	0	94	31	1	149
4	0	118	32	1	144
5	0	111	33	1	142
6	0	102	34	1	149
7	0	102	35	1	161
8	0	131	36	1	143
9	0	104	37	1	140
10	0	107	38	1	148
11	0	115	39	1	149
12	0	139	40	1	141
13	0	115	41	1	146
14	0	113	42	1	159
15	0	114	43	1	152
16	0	105	44	1	135
17	0	115	45	1	134
18	0	134	46	1	161
19	0	109	47	1	130
20	0	109	48	1	125
21	0	93	49	1	141
22	0	118	50	1	148
23	0	109	51	1	153
24	0	106	52	1	145
25	0	125	53	1	137
26	1	150	54	1	147
27	1	142	55	1	169
28	1	119			

Dataset name: blood_pressure

- a) State all the three possible hypotheses [3 Marks]
- b) Write a code to input the data and bind the variables into a single data frame named blood_pressure [2 Marks]
- c) Using the data provided, write a R code to
 - i) Calculate sample mean and total sample size [2 Marks]
 - ii) Set mean value under the null hypothesis and known sigma [2 Marks]
 - iii) Calculate test statistic and p-values [4 Marks]
