NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY OF HEALTH AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

QUALIFICATION: BACHELOR OF INFORMATICS BACHELOR OF COMPUTER SCIENCE	
QUALIFICATION CODE: 07BAIF 07BACS	**LEVEL:** 6
COURSE CODE: ASP612S	**COURSE NAME:** APPLIED STATISTICS & PROBABILITY FOR IT
SESSION: JULY 2019	**PAPER:** THEORY
DURATION: 3 HOURS	**MARKS:** 100

SUPPLEMENTARY / SECOND OPPORTUNITY EXAMINATION QUESTION PAPER	
EXAMINERS:	**MR. A. ROUX**
MODERATOR:	**DR. I.K.O. AJIBOLA**

INSTRUCTIONS

1. Answer **ALL** the questions.
2. Number the answers **clearly** and show **clearly ALL** the steps used in the calculations.
3. All written work **MUST** be done in blue or black ink.

PERMISSIBLE MATERIAL

1. Non-Programmable Calculator without the cover.

ATTACHMENTS

1. Formulae sheet.
2. Statistical tables (Z and T).
3. 1 X A4 GRAPH PAPER

THIS QUESTION PAPER CONSISTS OF 6 PAGES (Including this front page)

QUESTION 1 [10x2 = 20]

1.1. Any characteristic of a population distribution may properly be referred to as a
a.) standard deviation.
b.) standard score.
c.) raw score.
d.) standard error.
e.) parameter.

1.2.Characteristics of a population are called _____, while those of a sample are termed _____.
a.) statistics; measures
b.) parameters; statistics
c.) statistics; variables
d.) statistics; parameters
e.) none of these

1.3. A population is:
a.) a number or measurement collected as a result of observation
b.) a subset of a population
c.) a characteristic of a population which is measurable
d.) a complete set of individuals, objects, or measurements having some common observable characteristics
e.) none of these

1.4 Your statistics class
a.) is a representative sample of your college student body
b.) is not a representative sample of your college student body
c.) is not a sample of your college student body
d.) none of the above

1.5. Inferential statistics
a.) refers to the process of drawing inferences about the sample based on the characteristics of the population
b.) is the same as descriptive statistics
c.) refers to the statistical methods used to draw inferences about a population based on sample information
d.) is the same as a census
e.) none of the above answers is correct.

1.6. Which of the following is NOT a valid reason for selecting a sample instead of studying the whole population?
a.) The cost of studying an entire population may be too high.
b.) The population may be at least partially destroyed in the process of studying it.
c.) Studying the entire population might be too time consuming.
d.) All of the above are valid reasons for sampling.

1.7. Suppose we sample by selecting every fifth invoice in a file after randomly obtaining a starting point. What type of sampling is this?

a.) simple random sampling
b.) cluster random sampling
c.) stratified random sampling
d.) systematic random sampling
e.) None of the above

1.8. The _____ sampling method typically will require a larger sample size than other methods; however, the close proximity of sample elements can be cost-effective.

a.) simple random
b.) cluster
c.) stratified
d.) systematic
e.) None of the above.

1.9. All possible samples of size *n* are selected from a population, and the mean of each sample is determined. The mean of the sample means is _____.

a.) Exactly the same as the population mean
b.) Larger than the population mean
c.) Smaller than the population mean
d.) Cannot be estimated in advance
e.) None of the above

1.10. The _____ tells us that the distribution of all possible sample means will be approximately normal for reasonably large sample sizes.

a.) Central Limit Theorem
b.) Mean Limit Theorem
c.) Combination Theorem
d.) Estimation Theorem
e.) None of the above is correct.

QUESTION 2 [33]

The data below represents the annual rainfall (mm) recorded over forty farms across Namibia in 2018,

250	600	553	295	210	389	400	625	850	723
157	423	300	239	487	535	762	532	672	678
522	435	628	456	239	863	764	433	677	245
342	296	456	586	349	421	568	825	924	598

2.1 Summarize the data in a frequency distribution with classes of equal width of 100 rand, starting at 100 - < 200 ; 200 - < 300 ; ext.. (6)

2.2Use the data obtained in 2.1 to draw a histogram and a polygon (6 + 4 = 10)

2.3 Use the grouped data set produced in 2.1 to calculate and interpret the

2.3.1 mean (5)

2.3.2 median (5)

2.3.3 mode (5)

2.3.4 Based upon your measures of central tendencies calculated, comment of the shape of the distribution (2)

QUESTION 3 [30]

3.1) The Ministry of Education revealed that in a random sample of two hundred new Grade 1 learners, exactly one hundred and thirty eight of the are fully bilingual.

Write down ONLY the letter corresponding to your choice next to the question number.

3.1.1) What part of this sample of Grade 1 learners is fully bilingual
a) 0.96 b) 0.69 c) 1.38 d) none of the provided (2)

When constructing a confidence interval estimate for the single unknown population proportion { π} of Grade 1 learners who is fully bilingual:-

3.1.2) What critical value will be used?
a) t b) z c) χ d) none of the provided (1)

3.1.3) Compute the Standard Error of estimate
a) 0.2139 b) 1.0695 c) 0.0327 d) none of the provided (2)

If you construct a 90 % degree of confidence interval estimate for the population proportion of successes.

3.1.4) What critical value will be used?
a) 1.645 b) 1.96 c) 2.575 d) none of the provided (1)

3.1.5) What will be the lower limit (LL) for this confidence interval estimate?
a) 0.05379 b) 0.63620 c) 0.69 d) none of the provided (2)

3.1.6) What will be the upper limit (UL) for this confidence interval estimate?

a) 0.69 b) 0.05379 c) 0.7438 d) none of the provided (2)

3.2 In a multiple choice question, there are five different answers, of which only one is correct. The probability that a student will know the correct answer is 0.6. If a student does not know the answer, he guesses an answer at random.

3.2.1) What is the probability that the student gives the correct answer? (4)

3.2.2) If the student gives the correct answer, what is the probability that he guessed? (6)

3.3 A lab orders 100 rock samples a week for each of the 52 weeks in the year for experiments that the testing centre conducts. Prices for 100 rock samples follow the following distribution:

Price	(x):	$10.00	$12.50	$15.00
Probability	p(x):	0.35	0.40	0.25

Write down ONLY the letter corresponding to your choice next to the question number.

3.3.1) Find P (x ≤ N$ 12.50) (2)

a) 0.35 b) 0.40 c) 0.25 d) 0.65 e) None of the provided

3.3.2) Find the average price for the rock samples (3)

a) N$ 12.50 b) N$ 12.15 c) N$ 12.25 d) N$ 13.25 e) None of the provide

3.3.3) Find the variance for this distribution (3)

a) 3.74 b) 4.25 c) 3.48 d) 3.69 e) None of the provided

3.3.4) Find the variance for this distribution (2)

a) 1.93 b) 1.92 c) 1.87 d) 1.92 e) None of the provided

QUESTION 4 [17]

The following table shows the information of House sales given in quarters.

	Period	House sales
2003	Q1	54
	Q2	58
	Q3	94
	Q4	70
2004	Q1	55
	Q2	61
	Q3	87
	Q4	66
2005	Q1	49
	Q2	55
	Q3	95
	Q4	74
2006	Q1	60
	Q2	64
	Q3	99
	Q4	80

5.1 Use the data provided to construct a scatter plot (5)

5.2 Use the least squares regression to compute the estimated straight line trend equation starting with x=1 at 2003 - Q1. (9)

5.3 Use the trend line to estimate the value house sales for Q1 of 2007. (3)

XXXXXXXXXXXXXXXXXXXXXXXX END OF EXAMINATION XXXXXXXXXXXXXXXXX

APPENDIX A

Population mean, raw data

$$\mu = \frac{\sum x}{N}$$

Sample mean, raw data

$$\bar{X} = \frac{\sum x}{n}$$

Weighted mean

$$\overline{X_w} = \frac{w_1 X_1 + w_2 X_2 + \ldots + w_n X_n}{w_1 + w_2 + \ldots + w_n}$$

Geometric mean

$$GM = \sqrt[n]{(X_1)\ (X_2)\ (X_3)\ \ldots\ (X_n)}$$

Geometric mean rate of increase

$$GM = \sqrt[n]{\frac{\text{Value at end of period}}{\text{Value at start of period}}} - 1.0$$

Sample mean grouped data

$$\overline{X} = \frac{\sum fx}{n}$$

Median of grouped data

$$\text{Median} = L + \frac{\frac{n}{2} - CF}{f}\ (\text{Class width})$$

Mean deviation

$$MD = \frac{\sum |\ X - \bar{X}\ |}{n}$$

Linear regression equation

$$Y = a + bX$$

Sample variance for raw data

$$S^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

Sample variance, raw data computational form

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}$$

Sample standard deviation, raw data

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}}$$

Sample standard deviation, grouped data

$$S = \sqrt{\frac{\sum fX^2 - \frac{(\sum fx)^2}{n}}{n - 1}}$$

Coefficient of variation

$$CV = \frac{S}{X}\ (100)$$

Location of percentile

$$L_p = (n + 1)\ \frac{P}{100}$$

Pearson's Correlation coefficient

$$r = \frac{n\ (\sum XY) - (\sum X)\ (\sum Y)}{\sqrt{[n\ (\sum X^2) - (\sum X)^2]\ [n\ (\sum Y^2) - (\sum Y)^2]}}$$

Correlation test of hypothesis

$$t = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$$

Population standard deviation for raw data

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

Population variance for raw data

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Slope of regression line

$$b = \frac{n\ (\sum XY) - (\sum X)\ (\sum Y)}{n\ (\sum X^2) - (\sum X)^2}$$

Intercept of a regression line

$$a = \frac{\sum Y}{n} - b\left(\frac{\sum X}{n}\right)$$

The Range

$$\text{Range} = \text{highest} - \text{lowest}$$

APPENDIX B: ADDITIONAL FORMULAE

$$\text{Mode} = L + \left(\frac{d_1}{d_1 + d_2}\right) \times c$$

$$\text{position } Q_j = \frac{jn}{4} \qquad \text{value } Q_j = L + \frac{\left(\frac{jn}{4} - F\right) \times c}{f_{Q_j}}$$

$$\text{position } P_j = \frac{jn}{100} \qquad \text{value } P_j = L + \frac{\left(\frac{jn}{100} - F\right) \times c}{f_{P_j}}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(x) = \frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x} \qquad \text{P (x)} = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$z = \frac{x - \mu}{\sigma} \qquad z_{calc} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad t_{calc} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$z_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n-1)s_1^2 + (n-1)s_2^2}{n_1 + n_2 - 1}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

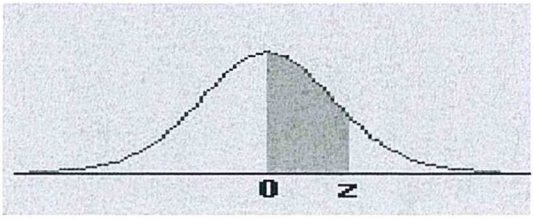
$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \qquad z_{calc} = \frac{p_A - p_B}{\sqrt{(p \times q)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \qquad p = \frac{n_A p_B + n_B p_A}{n_A + n_B} \qquad q = 1 - p$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$F_V = P_V(1 + in) \qquad F_V = P_V(1+i)^n \qquad r = (1+i)^m - 1 \qquad D = B(1-i)^n$$

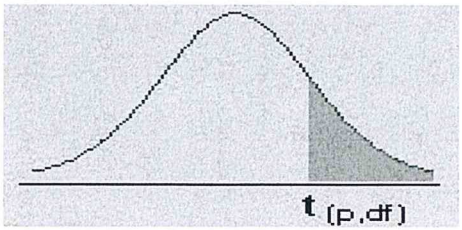
$$P = \frac{A}{(1+i)^n} \qquad PV = \frac{P(1+i)^n}{(1+j)^n} \qquad IRR = \frac{N_1 I_2 - N_2 I_1}{N_1 - N_2}$$

APPENDIX A: The Standard Normal Distribution



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

APPENDIX B: The t-distribution



df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905