UNIVERSITY OF SWAZILAND

SUPPLEMENTARY EXAMINATION PAPER 2016

TITLE OF PAPER	:	INFERENTIAL STATISTICS
COURSE CODE	:	ST 220
TIME ALLOWED	:	TWO (2) HOURS
REQUIREMENTS	:	CALCULATOR AND STATISTICAL TABLES
INSTRUCTIONS	:	THIS PAPER HAS FIVE (5). ANSWER ANY THREE (3) QUESTIONS.

Question 1 [20 marks, 10+7+3]

(a) Random samples are taken from two populations with distributions $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$ (i.e. their variances are the same). The summary statistics for the two samples are shown in the following Table:

	Sample Size n	Sample Mean m	Sample Variance s^2
x-data	19	7.0	1.69
y-data	25	5.1	2.56

Compute a 95% confidence interval for the difference $\mu_X - \mu_Y$ between the two population means. Does the result support the view that there is no true difference between the population means? (Explain your reasoning!)

(b) A market research company has conducted a survey of adults in two large towns, either side of an international border, in order to judge attitudes towards a controversial internationally broadcast celebrity television programme. The following table shows some of the information obtained by the survey:

	Town A	Town Z
Sample size	40	40
Sample number approving of the programme	24	22

Conduct a formal hypothesis test, at the 5% significance level, of the claim that the population proportions approving the programme in the two towns are equal.

(c) Every morning, Duncan bakes 30 scones to sell in his cafe'. If any scones are unsold at the end of the day, Duncan throws them away. The number of scones requested during a day may be modelled by a Poisson distribution with mean 27. Estimate the probability that Duncan does not have enough scones to satisfy all the requests on a particular day (*Use normal approximation to the Poisson distribution*).

Question 2 [20 marks, 6+4+4+6]

(a) A short-stay car park in a shopping area has spaces marked out for 90 cars. A local councillor notices that there are always some vacant spaces. He puts forward a plan to create a garden and seating area using part of the car park. This would reduce the number of parking spaces to 78.

(i) From a random sample of 33 users of the car park, 26 say that the car park will be too small if this plan is carried out. Carry out a test, at the 5% significance level, to determine whether more than half of the users of the car park think it will be too small.

(ii) The number of occupied spaces, x, in the car park is recorded on each of 16 randomly chosen occasions during shopping hours. The results may be summarised as follows:

$$\bar{x} = 59.9 \qquad s = 7.83$$

Construct a 95% confidence interval for the mean, μ, of the number of spaces occupied in the car park during shopping hours. Assume that the sample is drawn from a normal population.

(iii) The councillor claims that the value of μ is no more than 65. It is found that the number of occupied spaces during shopping hours is best modelled by a Poisson distribution with mean μ. Taking μ to be 65, use a distributional approximation to find the probability that more than 78 spaces are occupied in the car park at any one time.

(b) A standard pack of 52 playing cards consists of 4 suits (clubs, diamonds, hearts and spades), each consisting of 13 cards numbered 2, 3, 4, $\cdots$, 10, Jack, Queen, King, Ace (their face values). In the game of poker, a hand of 5 cards is drawn without replacement from a well-shuffled pack. A poker hand consisting of a pair of cards with the same face value and three other cards with the same face value (different from that of the pair) is called a *full house.* Find the probability that a poker hand drawn from a well-shuffled pack is a full house. Express your answer either as a fraction in lowest terms or as a decimal correct to 3 significant figures.

Question 3 [20 marks, 8+7+1+4]

Students on an environmental science course are investigating nitrate pollution in a river in an agricultural region. The level of pollution becomes a cause for concern when the mean concentration of nitrate exceeds 30 milligrams per litre of water.

The river is divided into a large number of sections of equal length.

(a) One student takes samples of water at 8 randomly chosen locations along one of these sections and analyses the samples for nitrate concentration. Her results, in milligrams of nitrate per litre of water, are

30 34 34 37 28 30 34 35

Carry out a test to investigate whether the nitrate pollution in this section of the river is a cause for concern. Assume that the data are drawn from a normal population, and use the 1% significance level.

(b) The students carry out similar investigations to that in part (a) on 42 sections. Their tests indicate that the mean concentration of nitrate exceeds 30 milligrams per litre of water in 16 sections.

(i) Carry out a test, at the 1% significance level, to determine whether the level of nitrate concentration is a cause for concern in less than 60 per cent of sections of this river.

(ii) State one assumption that must be made for your conclusion in part (b)(i) to be valid.

(iii) Construct the 95% confidence interval for the proportion of sections which are not polluted.

Question 4 [20 marks, 8+8+4]

(a) Researchers in the USA investigated the effects of alcohol consumption on brain size. A sample of 10 non-drinkers and 8 heavy drinkers (more than 14 units of alcohol per week) was obtained and each person in the sample had a brain scan.The ratios of brain volume to skull size obtained for the eighteen people in the sample are given in the table.

Non-drinkers	0.794	0.798	0.799	0.802	0.803	0.804	0.805	0.806	0.809	0.810
Heavy drinkers	0.785	0.787	0.789	0.791	0.792	0.796	0.797	0.801		

Assuming that the sample is random, carry out a test, at the 5% level of significance, to investigate the claim that heavy drinkers have a smaller average ratio of brain volume to skull size than non-drinkers. Interpret your conclusion in context.

(b) In a diet test, each of four diet programs is applied to a sample of people. At the end of three weeks, the amount of pounds people lost are shown below.

Diet Program			
1	**2**	**3**	**4**
12	19	16	28
6	10	20	17
18	13	26	22
23	20	19	16
	25		20

Test to determine if there is enough evidence at the 5% significance level to infer that at least two population locations differ. State the hypothesis, critical region(s) and conclusions. Show all calculations.

(c) A blended wine is intended to comprise two parts of Sauvignon to one part of Merlot. The amounts dispensed to make up a nominal 75cl bottle of this wine are X cl of Sauvignon and Y cl of Merlot, where X and Y are assumed to be independent Normally distributed random variables with respective means 52 and 26 cl and respective variances 1 and 0.5625. Find the probability that the actual volume of wine dispensed into a bottle is less than the nominal volume.

Question 5 [20 marks, 6+8+6]

(a) A study was carried out into the life of mammals. The maximum recorded life span, w years, the mean gestation time, x days, and the mean daily sleep time, y hours, were obtained for a random sample of 11 species of mammal. The results are given in the table.

The results are given in the table.

Species	w	x	y
A	38	645	3.3
B	14	60	12.5
C	69	624	3.9
D	27	180	9.8
E	19	35	19.7
F	50	230	14.5
G	30	281	9.7
H	40	365	3.9
I	28	400	3.1
J	4	16	14.4
K	39	252	12.0

The product moment correlation coefficient between x and y is -0.853, correct to three significant figures. Carry out a hypothesis test, at the 5% level of significance, to determine whether the value of the product moment correlation coefficient between x and y indicates an association between this pair of variables.

(b) In a first phase of a health study in a city, a random sample of size 2000 is to be obtained. The city is comprised (broadly) of five different ethnic subpopulations that make up 40%, 30%, 10%, 10% and 10% of the city population respectively.

A commercial company is employed to obtain the random sample, with the instruction that the sample should reflect the ethnic composition of the city. The sample they return is summarized in the following table.

	Ethnic Subpopulation				
	1	**2**	**3**	**4**	**5**
Number in Sample	822	638	210	157	173

Using a Chi-squared test for this one-way layout, comment on whether the company have fulfilled their remit to produce a sample that reflects the ethnic composition of the city.

(c) The number of flaws, X, in a standard length of yarn is assumed to be Poisson distributed with probability mass function

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, 3, \cdots$$

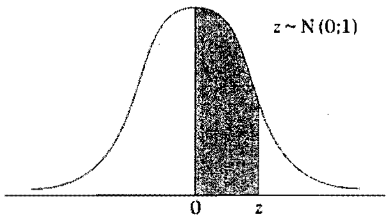
where λ is a positive parameter. A textile manufacturer buys yarn from suppliers P, Q and R in the long-run proportions $\frac{1}{6}$, $\frac{1}{3}$ and $\frac{1}{2}$ respectively. It is known from experience that the numbers of flaws in lengths of yarn from these suppliers are independently Poisson distributed with respective parameter values $\lambda_P = 3$, $\lambda_Q = 2$ and $\lambda_R = 1$. An unlabelled length of yarn is found to have 2 flaws. Is it more likely to have come from supplier Q or supplier R?

APPENDIX 1: LIST OF STATISTICAL TABLES

TABLE 1

The standard normal distribution (z)

This table gives the area under the standard normal curve between 0 and z
i.e. $P[0 < Z < z]$

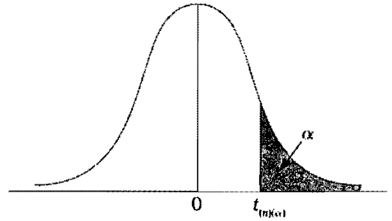


Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	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	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998

TABLE 2

The t distribution

This table gives the value of $t_{(n)(\alpha)}$ where n is the degrees of freedom
i.e. $\alpha = P[t \geq t_{(n)(\alpha)}]$



α / df	0.100	0.050	0.025	0.010	0.005	0.0025
1	3.078	6.314	12.706	31.821	63.657	127.321
2	1.886	2.920	4.303	6.965	9.925	14.089
3	1.638	2.353	3.182	4.541	5.841	7.453
4	1.533	2.132	2.776	3.747	4.604	5.598
5	1.476	2.015	2.571	3.365	4.032	4.773
6	1.440	1.943	2.447	3.143	3.707	4.317
7	1.415	1.895	2.365	2.998	3.499	4.029
8	1.397	1.860	2.306	2.896	3.355	3.833
9	1.383	1.833	2.262	2.821	3.250	3.690
10	1.372	1.812	2.228	2.764	3.169	3.581
11	1.363	1.796	2.201	2.718	3.106	3.497
12	1.356	1.782	2.179	2.681	3.055	3.428
13	1.350	1.771	2.160	2.650	3.012	3.372
14	1.345	1.761	2.145	2.624	2.977	3.326
15	1.341	1.753	2.131	2.602	2.947	3.286
16	1.337	1.746	2.120	2.583	2.921	3.252
17	1.333	1.740	2.110	2.567	2.898	3.222
18	1.330	1.734	2.101	2.552	2.878	3.197
19	1.328	1.729	2.093	2.539	2.861	3.174
20	1.325	1.725	2.086	2.528	2.845	3.153
21	1.323	1.721	2.080	2.518	2.831	3.135
22	1.321	1.717	2.074	2.508	2.819	3.119
23	1.319	1.714	2.069	2.500	2.807	3.104
24	1.318	1.711	2.064	2.492	2.797	3.091
25	1.316	1.708	2.060	2.485	2.787	3.078
26	1.315	1.706	2.056	2.479	2.779	3.067
27	1.314	1.703	2.052	2.473	2.771	3.057
28	1.313	1.701	2.048	2.467	2.763	3.047
29	1.311	1.699	2.045	2.462	2.756	3.038
30	1.310	1.697	2.042	2.457	2.750	3.030
31	1.309	1.696	2.040	2.453	2.744	3.022
32	1.309	1.694	2.037	2.449	2.738	3.015
33	1.308	1.692	2.035	2.445	2.733	3.008
34	1.307	1.691	2.032	2.441	2.728	3.002
35	1.306	1.690	2.030	2.438	2.724	2.996
36	1.306	1.688	2.028	2.434	2.719	2.990
37	1.305	1.687	2.026	2.431	2.715	2.985
38	1.304	1.686	2.024	2.429	2.712	2.980
39	1.304	1.685	2.023	2.426	2.708	2.976
40	1.303	1.684	2.021	2.423	2.704	2.971
45	1.301	1.679	2.014	2.412	2.690	2.952
50	1.299	1.676	2.009	2.403	2.678	2.937
60	1.296	1.671	2.000	2.390	2.660	2.915
70	1.294	1.667	1.994	2.381	2.648	2.899
80	1.292	1.664	1.990	2.374	2.639	2.887
90	1.291	1.662	1.987	2.368	2.632	2.878
100	1.290	1.660	1.984	2.364	2.626	2.871
110	1.289	1.659	1.982	2.361	2.621	2.865
120	1.289	1.658	1.980	2.358	2.617	2.860
140	1.288	1.656	1.977	2.353	2.611	2.852
160	1.287	1.654	1.975	2.350	2.607	2.846
180	1.286	1.653	1.973	2.347	2.603	2.842
200	1.286	1.653	1.972	2.345	2.601	2.839
∞	1.282	1.645	1.960	2.326	2.576	2.807

MEASURES OF DISPERSION AND SKEWNESS

Range Range = Maximum value – Minimum value + 1 $= x_{max} - x_{min} + 1$ 3.9

Variance *Mathematical – ungrouped data*

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{(n-1)}$$ 3.10

Computational – ungrouped data

$$s^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{(n-1)}$$ 3.11

Standard deviation $s = \sqrt{s^2}$ 3.12

Coefficient of variation $CV = \frac{s}{\bar{x}} \times 100\%$ 3.13

Pearson's coefficient of skewness

$$sk_p = \frac{n\Sigma(x_i - \bar{x})^3}{(n-1)(n-2)s^3}$$ 3.14

$$sk_p = \frac{3\,(\text{Mean} - \text{Median})}{\text{Standard deviation}}$$ (approximation) 3.15

PROBABILITY CONCEPTS

Conditional probability $P(A/B) = \frac{P(A \cap B)}{P(B)}$ 4.2

Addition rule *Non-mutually exclusive events*

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 4.3

Mutually exclusive events

$P(A \cup B) = P(A) + P(B)$ 4.4

Multiplication rule *Statistically dependent events*

$P(A \cap B) = P(A/B) \times P(B)$ 4.5

Statistically independent events

$P(A \cap B) = P(A) \times P(B)$ 4.6

***n*! = *n* factorial** $n \times (n-1) \times (n-2) \times (n-3) \times \ldots \times 3 \times 2 \times 1$ 4.8

Permutations ${}_nP_r = \frac{n!}{(n-r)!}$ 4.10

Combinations ${}_nC_r = \frac{n!}{r!\,(n-r)!}$ 4.11

PROBABILITY DISTRIBUTIONS

Binomial distribution $P(x) = {}_nC_x\, p^x (1-p)^{(n-x)}$ for $x = 0, 1, 2, 3, \ldots, n$ 5.1

$P(x \textit{ successes}) = \frac{n!}{x!\,(n-x)!}\, p^x(1-p)^{(n-x)}$ for $x = 0, 1, 2, 3, \ldots, n$

Binomial descriptive measures

Mean $\mu = np$

Standard deviation $\sigma = \sqrt{np(1-p)}$ 5.2

Poisson distribution $P(x) = \frac{e^{-a} a^x}{x!}$ for $x = 0, 1, 2, 3 \ldots$ 5.3

Poisson descriptive measures

Mean $\mu = a$

Standard deviation $\sigma = \sqrt{a}$ 5.4

Standard normal probability $z = \frac{x - \mu}{\sigma}$ 5.6

CONFIDENCE INTERVALS

Single mean — n *large; variance known*

$$\bar{x} - z\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z\frac{\sigma}{\sqrt{n}} \qquad 7.1$$

(lower limit) (upper limit)

n *small; variance unknown*

$$\bar{x} - t_{(n-1)}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{(n-1)}\frac{s}{\sqrt{n}} \qquad 7.2$$

(lower limit) (upper limit)

Single proportion

$$p - z\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z\sqrt{\frac{p(1-p)}{n}} \qquad 7.3$$

(lower limit) (upper limit)

HYPOTHESES TESTS

10

Single mean — *Variance known*

$$z\text{-}stat = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad 8.1$$

Variance unknown; n *small*

$$t\text{-}stat = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \qquad 8.2$$

Single proportion

$$t\text{-}stat = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \qquad 8.3$$

Difference between two means — *Variances known*

$$z\text{-}stat = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad 9.1$$

Variances unknown; n_1 and n_2 small

$$t\text{-}stat = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad 9.2$$

Paired *t*-test

$$t\text{-}stat = \frac{\bar{x}_d - \mu_d}{\frac{s_d}{\sqrt{n}}} \qquad 9.5$$

where $\mu_d = (\mu_1 - \mu_2)$

and $s_d = \sqrt{\frac{\Sigma(x_d - \bar{x}_d)^2}{n-1}}$

Differences between two proportions

$$z\text{-}stat = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where } \hat{\pi} = \frac{x_1 + x_2}{n_1 + n_2};\ p_1 = \frac{x_1}{n_1};\ p_2 = \frac{x_2}{n_2} \qquad 9.8$$

Chi-Squared

$$\chi^2\text{-}stat = \Sigma\frac{(f_o - f_e)^2}{f_e} \qquad 10.1$$

Overall mean

$$\bar{\bar{x}} = \frac{\Sigma\Sigma x_{ij}}{N} \qquad 11.2$$

Total sum of squares (SSTotal)

$$= \sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2 \qquad 11.3$$

SST

$$= \sum_j^k n_j(\bar{x}_j - \bar{\bar{x}})^2 \qquad 11.4$$

SSE

$$= \sum_j \sum_i (x_{ij} - \bar{x}_j)^2 \qquad 11.5$$

SSTotal

$$= \text{SST} + \text{SSE} \qquad 11.6$$

MSTotal

$$= \frac{\text{SSTotal}}{N-1} \qquad 11.7$$

MST

$$= \frac{\text{SST}}{k-1} \qquad 11.8$$

MSE

$$= \frac{\text{SSE}}{N-k} \qquad 11.9$$

***F*-stat**

$$= \frac{\text{MST}}{\text{MSE}} \qquad 11.10$$