UNIVERSITY OF SWAZILAND

FINAL EXAMINATION PAPER 2014

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, 8+4+4+4]

(a) A study was carried out to determine the effect of the presence or absence of a company safety programme on the number of work hours lost due to work-related accidents. Fifty companies were selected from a particular business sector that contains a large number of companies. A lost work hours ratio was calculated for each company by dividing the total number of hours lost over a given period by the number of employees. The data (given for convenience in order of size) are shown below.

```
Companies with no safety programme (sample A)

0.0083 0.0089 0.0091 0.0094 0.0116 0.0133 0.0133 0.0145
0.0153 0.0166 0.0168 0.0169 0.0169 0.0173 0.0179 0.0183
0.0186 0.0189 0.0195 0.0202 0.0204 0.0211 0.0217 0.0223
0.0223 0.0230 0.0233 0.0234

Companies with a safety programme (sample B)

0.0029 0.0070 0.0072 0.0085 0.0092 0.0095 0.0099 0.0106
0.0106 0.0111 0.0123 0.0128 0.0138 0.0142 0.0143 0.0153
0.0155 0.0164 0.0189 0.0198 0.0213 0.0245
```

(i) Use the summary data given below to carry out a two-sample t test to determine whether there is evidence that a safety programme reduces the lost work hours ratio. State carefully the null and alternative hypotheses.

	Number of observations	Sample mean	Sample standard deviation
Sample A	28	0.01711	0.00462
Sample B	22	0.01298	0.00513

(ii) State the assumptions required for validity of the test in (i).

(b) A survey is to be conducted amongst students at a particular university to estimate the mean number of hours spent in part-time work per week.

Previous studies of employment patterns of students at universities have established 9.3 hours as the standard deviation of hours worked. Calculate how many students, approximately, should be sampled in order to estimate the mean to within 1 hour, at a 95% level of confidence. You may assume in your calculation that the population standard deviation is equal to the above figure.

(c) A random sample of twenty bottles of a medicine that comes in liquid form is taken and the contents of each bottle are accurately weighed. The sample mean weight (in grams) of the contents of a bottle is 99.43 and the sample variance is 0.4678. You may assume that the weights are Normally distributed. Calculate a 95% confidence interval for the underlying population mean weight of medicine in a bottle.

Question 2 [20 marks, 8+6+6]

(a) The number of telephone enquiries received by an operative at a call centre during each minute is modelled by a Poisson distribution with mean λ, and the numbers of enquiries received in non-overlapping intervals are independent.

The numbers of enquiries received by an operative over a period of 150 non-overlapping one-minute intervals are shown in Table below.

Number of enquiries per minute	Frequency
0	39
1	66
2	29
3	10
4	4
5	2
6 or more	0

Calculate the mean number of enquiries received per minute. Carry out an appropriate statistical test to determine whether or not a Poisson distribution is a suitable model for these data.

(b) In the Swaziland Household survey of 2006, data were collected on family structure during childhood. One question asked "Did you live with both your biological mother and biological father from the time you were born until you were 16?" Those who responded 'yes' are classed as living in an intact family until the age of 16; those who responded 'no' are classed as living in a non-intact family before the age of 16. The responses from those born between 1984 and 2000 are tabulated below.

	Lived in an intact family until age 16	
Sex	**No**	**Yes**
Female	133	460
Male	155	418

(i) Compute and briefly interpret an approximate 95% confidence interval for the difference between the proportions of males and females born between 1984 and 2000 who lived in a non-intact family before the age of 16.

(ii) The confidence interval found in (i) is of width approximately 0.1. How large a sample size would you need to reduce the width to approximately 0.05?

Question 3 [20 marks, 10+10]

(a) The following coded pairs of measurements were taken of the temperature (X) and thrust (Y) of a jet engine while it was being tested under uniform operating conditions.

x	15	20	25	26	30	33	34	35	38	39	41	46	49	52	57
y	1.4	1.2	1.9	1.6	2.5	2.1	2.4	1.5	2.3	2.7	1.8	2.2	2.8	3.4	3.2

You are given that $\sum x = 540$, $\sum x^2 = 21412$, $\sum y = 33.0$, $\sum y^2 = 78.54$, $\sum xy = 1276.6$.

Calculate r for these data, and test at the 5% significance level the hypothesis of zero correlation against the alternative that thrust and temperature are positively correlated. State clearly (but do not prove) any formulae that you have used, and list the assumptions you have made in the test.

(b) Two examiners, A and B, independently mark student projects, in each case assigning a mark out of 100 for the project. The marks assigned by the examiners for each of 12 randomly selected student projects are recorded below.

Student	Examiner A	Examiner B
1	53	42
2	54	44
3	39	42
4	47	56
5	72	65
6	61	63
7	64	58
8	51	59
9	68	89
10	59	87
11	58	72
12	53	79

The issue has been raised as to whether one of the examiners tends to give higher marks than the other. Noting any assumptions made, carry out an appropriate parametric test and draw conclusions

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

(a) The weights in grams of a random sample of 10 cherry tomatoes taken from the plants in a large greenhouse are 37, 39, 40, 33, 41, 39, 36, 46, 35 and 44.

(i) Calculate the mean and variance of the observed weights.

(ii) Assuming that the distribution underlying the data may be taken as approximately Normal, give 95% confidence intervals for the mean of the weights of cherry tomatoes in the greenhouse. State clearly any formulae you use for these calculations.

(iii) In order to be graded as acceptable for sale, the cherry tomatoes should not be too small or very variable in size. Test, in each case at the 5% level of significance, (a) the hypothesis that the mean weight is 42 g against the hypothesis that it is less than 42 g,

(b) A die is rolled 100 times. Use a suitable approximation to find the probability that the sum of the scores on the uppermost face is more than 365, assuming the die to be fair. [You are given that the variance of the score on the uppermost face for each rolling of a fair die is $\frac{35}{12}$.]

Question 5 [20 marks, 8+12]

(a) The question of interest is whether the subject matter of the postgraduate courses is relevant to the level of female participation. The MSc courses are either finance or management oriented. The

MBA is predominantly a management course. The MSc students are generally younger than the MBA students, who have work experience and are often sponsored by their employers. Subdividing the MScs into finance and management gives the following table.

	MSc (finance)	MSc (management)	MBA (management)
Male	611	105	169
Female	363	103	51

Analyse these data in a manner that allows you to comment on the relationship between the type of Master's course studied and the proportion of female students, and also to comment on any influence of the type of subject matter. Write a short report (5 or 6 sentences) to summarise your findings.

(b) A trial is undertaken to investigate the effect on fuel economy of 3 fuel additives **A**, **B** and **C**, where **A** and **B** are new and **C** is the current standard additive. The same driver drives the same car on a fixed test route during 20 working days. The additive used on each day is randomly assigned so that **A** and **B** are each used for 5 days and **C** is used for 10 days. The response variable measured each day is Y, the number of miles per gallon (mpg) achieved.

The results are shown in the following table.

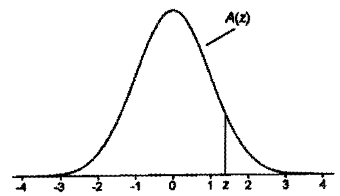
Additive	y	Total
A	39, 35, 37, 36, 38	$\sum y_A = 185$
B	36, 41, 39, 40, 39	$\sum y_B = 195$
C	37, 33, 30, 34, 36, 34, 31, 36, 34, 35	$\sum y_C = 340$

You are given that the sum of squares of the observations is 26078.

Carry out an analysis of variance to test for differences between the effects on Y of the additives. State clearly your null and alternative hypotheses and present your conclusions.

TABLE A.1

Cumulative Standardized Normal Distribution



$A(z)$ is the integral of the standardized normal distribution from $-\infty$ to z (in other words, the area under the curve to the left of z). It gives the probability of a normal random variable not being more than z standard deviations above its mean. Values of z of particular importance:

z	$A(z)$	
1.645	0.9500	Lower limit of right 5% tail
1.960	0.9750	Lower limit of right 2.5% tail
2.326	0.9900	Lower limit of right 1% tail
2.576	0.9950	Lower limit of right 0.5% tail
3.090	0.9990	Lower limit of right 0.1% tail
3.291	0.9995	Lower limit of right 0.05% tail

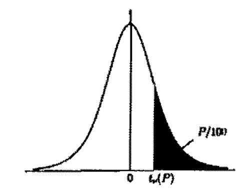
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

Percentage Points of the t-Distribution

This table gives the percentage points $t_\nu(P)$ for various values of P and degrees of freedom ν, as indicated by the figure to the right.

The lower percentage points are given by symmetry as $-t_\nu(P)$, and the probability that $|t| \geq t_\nu(P)$ is $2P/100$.

The limiting distribution of t as $\nu \to \infty$ is the normal distribution with zero mean and unit variance.



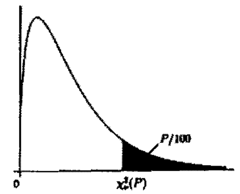
ν	Percentage points P 10	5	2.5	1	0.5	0.1	0.05
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
70	1.294	1.667	1.994	2.381	2.648	3.211	3.435
100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Percentage Points of the χ^2-Distribution

This table gives the percentage points $\chi^2_\nu(P)$ for various values of P and degrees of freedom ν, as indicated by the figure to the right.

If X is a variable distributed as χ^2 with ν degrees of freedom, $P/100$ is the probability that $X \geq \chi^2_\nu(P)$.

For $\nu > 100$, $\sqrt{2X}$ is approximately normally distributed with mean $\sqrt{2\nu - 1}$ and unit variance.



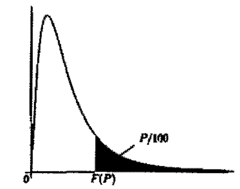
ν	Percentage points P						
	10	5	2.5	1	0.5	0.1	0.05
1	2.706	3.841	5.024	6.635	7.879	10.828	12.116
2	4.605	5.991	7.378	9.210	10.597	13.816	15.202
3	6.251	7.815	9.348	11.345	12.838	16.266	17.730
4	7.779	9.488	11.143	13.277	14.860	18.467	19.997
5	9.236	11.070	12.833	15.086	16.750	20.515	22.105
6	10.645	12.592	14.449	16.812	18.548	22.458	24.103
7	12.017	14.067	16.013	18.475	20.278	24.322	26.018
8	13.362	15.507	17.535	20.090	21.955	26.124	27.868
9	14.684	16.919	19.023	21.666	23.589	27.877	29.666
10	15.987	18.307	20.483	23.209	25.188	29.588	31.420
11	17.275	19.675	21.920	24.725	26.757	31.264	33.137
12	18.549	21.026	23.337	26.217	28.300	32.909	34.821
13	19.812	22.362	24.736	27.688	29.819	34.528	36.478
14	21.064	23.685	26.119	29.141	31.319	36.123	38.109
15	22.307	24.996	27.488	30.578	32.801	37.697	39.719
16	23.542	26.296	28.845	32.000	34.267	39.252	41.308
17	24.769	27.587	30.191	33.409	35.718	40.790	42.879
18	25.989	28.869	31.526	34.805	37.156	42.312	44.434
19	27.204	30.144	32.852	36.191	38.582	43.820	45.973
20	28.412	31.410	34.170	37.566	39.997	45.315	47.498
25	34.382	37.652	40.646	44.314	46.928	52.620	54.947
30	40.256	43.773	46.979	50.892	53.672	59.703	62.162
40	51.805	55.758	59.342	63.691	66.766	73.402	76.095
50	63.167	67.505	71.420	76.154	79.490	86.661	89.561
80	96.578	101.879	106.629	112.329	116.321	124.839	128.261

5 Percent Points of the F-Distribution

This table gives the percentage points $F_{\nu_1,\nu_2}(P)$ for $P = 0.05$ and degrees of freedom ν_1, ν_2, as indicated by the figure to the right.

The lower percentage points, that is the values $F'_{\nu_1,\nu_2}(P)$ such that the probability that $F \leq F'_{\nu_1,\nu_2}(P)$ is equal to $P/100$, may be found using the formula

$$F'_{\nu_1,\nu_2}(P) = 1/F_{\nu_1,\nu_2}(P)$$



ν_2	ν_1								
	1	2	3	4	5	6	12	24	∞
2	18.513	19.000	19.164	19.247	19.296	19.330	19.413	19.454	19.496
3	10.128	9.552	9.277	9.117	9.013	8.941	8.745	8.639	8.526
4	7.709	6.944	6.591	6.388	6.256	6.163	5.912	5.774	5.628
5	6.608	5.786	5.409	5.192	5.050	4.950	4.678	4.527	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.000	3.841	3.669
7	5.591	4.737	4.347	4.120	3.972	3.866	3.575	3.410	3.230
8	5.318	4.459	4.066	3.838	3.687	3.581	3.284	3.115	2.928
9	5.117	4.256	3.863	3.633	3.482	3.374	3.073	2.900	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	2.913	2.737	2.538
11	4.844	3.982	3.587	3.357	3.204	3.095	2.788	2.609	2.404
12	4.747	3.885	3.490	3.259	3.106	2.996	2.687	2.505	2.296
13	4.667	3.806	3.411	3.179	3.025	2.915	2.604	2.420	2.206
14	4.600	3.739	3.344	3.112	2.958	2.848	2.534	2.349	2.131
15	4.543	3.682	3.287	3.056	2.901	2.790	2.475	2.288	2.066
16	4.494	3.634	3.239	3.007	2.852	2.741	2.425	2.235	2.010
17	4.451	3.592	3.197	2.965	2.810	2.699	2.381	2.190	1.960
18	4.414	3.555	3.160	2.928	2.773	2.661	2.342	2.150	1.917
19	4.381	3.522	3.127	2.895	2.740	2.628	2.308	2.114	1.878
20	4.351	3.493	3.098	2.866	2.711	2.599	2.278	2.082	1.843
25	4.242	3.385	2.991	2.759	2.603	2.490	2.165	1.964	1.711
30	4.171	3.316	2.922	2.690	2.534	2.421	2.092	1.887	1.622
40	4.085	3.232	2.839	2.606	2.449	2.336	2.003	1.793	1.509
50	4.034	3.183	2.790	2.557	2.400	2.286	1.952	1.737	1.438
100	3.936	3.087	2.696	2.463	2.305	2.191	1.850	1.627	1.283
∞	3.841	2.996	2.605	2.372	2.214	2.099	1.752	1.517	1.002