UNIVERSITY OF SWAZILAND

FINAL EXAMINATION PAPER 2010

TITLE OF PAPER	:	NONPARAMETRIC ANALYSIS
COURSE CODE	:	ST409
TIME ALLOWED	:	TWO (2) HOURS
REQUIREMENTS	:	CALCULATOR, GRAPH PAPER, AND STATISTICAL TABLES
INSTRUCTIONS	:	ANSWER ANY THREE QUESTIONS

Question 1 [20 marks, 10+10]

(a) Derive the distribution of *Spearman's rho* under the null hypothesis that there is no association between two variables, X and Y, for $n = 4$.

(b) A certain store advertises that the waiting times at till queues are at most five minutes and they encourage their patrons to lodge a complaint if they perceive their waiting time to be more than five minutes. The store operates under the assumption that the waiting time in queues follows an exponential distribution with parameter 0.33 (when the waiting time is measured in minutes). After various complaints have been received, the store decides to investigate the distribution of waiting times. The data below (in seconds) were observed waiting times in the store, measured at random times.

6 78 120 174 312 324 366 366 534

(i) Sketch the empirical distribution function associated with the data.

(ii) Using an suitable test the appropriate hypothesis from the customer's perspective at the 5% level of significance.

Question 2 [20 marks, 2+8+10]

Cancer cells were injected into six lab rats to study the effectiveness of a new treatment. Four rats received the new treatment, while two rats were given a placebo. After four months the size of the tumors (cm) were measured. These measurements are stated below:

Treatment (X)	0.8	0.0	0.7	1.1
Placebo (Y)	0.9	2.2		

To test if treatment is effective in reducing the size of tumors:

(a) Sketch the empirical distribution functions on the same set of axes.

(b) State the appropriate hypothesis when using a Kolmogorov-Smirnov two-sample test and then perform the test corresponging to the hypothesis at the 1% level of significance.

(c) Determine the distribution of the appropriate test statistic under the null hypothesis.

Question 3 [20 marks, 12+8]

(a) The following weight (in kilograms) were observed for a random sample of grade 7 girls in a Swazi school:

44	45	47	50	51	55	58	59	59	61
62	65	67	70	71	77	80	85	96	107

(i) Due to availability of fast food and lack of exercise, the school's governing body fears that the 66th percentile of the weight of grade 7 girls has increased since the 1980s. The school has records suggesting that the 66th percentile used to be 59kg. Test the governing body's hypothesis with the most appropriate non-parametric procedure, using p-values.

(ii) Find the 90% non-parametric confidence interval for the 66th percentile of the population of 2008 weight, stating the length and confidence level obtained for this interval with respect to the data.

(b) As part of a residence's rag activities, the house committee has taken height and weight measurements of their first year students for the past 24 years. One of the 2008 house committee members thinks that males have become taller over time, due to better nutrition. The data provided is the mean height (in cm) for males observed each year for the 24 years preceding 2008.

1984	172.9	1992	172.0	2000	173.3
1985	172.0	1993	172.2	2001	173.3
1986	173.1	1994	172.0	2002	173.0
1987	171.5	1995	171.8	2003	173.8
1988	173.0	1996	172.8	2004	172.0
1989	173.0	1997	171.8	2005	173.0
1990	173.0	1998	172.0	2006	172.8
1991	172.3	1999	172.1	2007	142.8

Test the appropriate hypothesis using a variation of the binomial test at the 5% level of significance for the data provided. Clearly state how you deal with the ties.

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

(a) As an indication of mathematical ability, learners write a test on the first day of the academic year. It is assumed that the test results follow a normal distribution, but the parameters are unknown. Twenty-five students, in a particular class (in 2009), wrote the test. The teacher wishes to know if the data does in fact follow the normal distribution. Identity the most appropriate technique for the teachers to use. Justify your answer and mention possible assumptions that are not met.

(b) Consider two continuous independent random variables, X and Y, from two possibly different populations and n and m observations for each random variable respectively. Let the Mann-Whitney test statistic, S, be the sum of the ranks associated with the random variable X. Show that under the null hypothesis the test statistic has $E(S) = \frac{n(n+m+1)}{2}$ and $Var(S) = \frac{nm(n+m+1)}{12}$.

(c) The median income for a particular profession used to be SZL185000 per annum, five years ago. Eleven such professionals were asked to disclose their current salaries and these were adjusted for inflation. Have the adjusted salaries increased at a rate similar to that of inflation? Use an appropriate test, generating the exact $p-$values to make your decision. Perform the test at the 1% level of significance.

164000	181000	189000	192500
195600	213300	231800	234900
248700	256400	267200	

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

(a) A sample of teenagers was divided into male and female on the one hand, and those that are and are not currently dieting on the other. It was hypothesized, that the proportion of dieting individuals is higher among the women than among the men, and we want to test whether any difference of proportions that we observe is significant. State the hypothesis and compute the p–value for the the test.

	Gender		
	Men	Women	Total
Dieting	1	9	10
Not dieting	11	3	12
Total	12	12	18

(b) An electronics manufacturer wants to test the thickness of the insulation which covers three competing brands of the same type of wire, using a sample of five randomly selected pieces of wire for each brand and carefully measuring the thickness of the insulation in millimitres. The results of these measurements are as follows.

Brand 1	Brand 2	Brand 3
2.4	2.0	2.6
2.3	2.2	2.6
2.2	2.5	2.6
2.0	2.6	2.0
2.6	2.7	2.7

Use a suitable procedure to test the hypothesis that the three competing brands are covered with insulation of the same thinkness at the 1% level of significance.

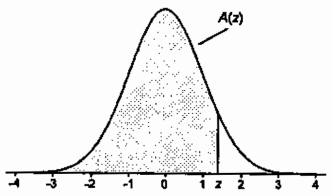
(c) Suppose we record the gender of the 15 students enrolled in an introductory statistics course as they enter the classroom.

M M F M M F M M F F M M M M F

Which test would you use to disprove the claim that there is some grouping by gender in the way students enter the classroom. State the hypothesis and mention any assumptions that would have to be satisfied for the test to be valid.

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

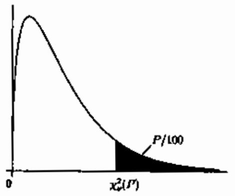
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	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 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.



	Percentage points P						
ν	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

Percentage Points of the Wilcoxon Signed Rank Distribution

This table gives the lower percentage points of W^+, the sum of the ranks of the positive observations in a ranking in order of increasing absolute magnitude of a random sample of size n from a continuous distribution which is symmetric about zero. The function tabulated $x(P)$ is the largest x such that $P(W^+ < x) \le P/100$.

	P						P				
n	5	2.5	1	0.5	0.1	n	5	2.5	1	0.5	0.1
8	6	4	2	1	0	**43**	337	311	282	262	223
9	9	6	4	2	0	**44**	354	328	297	277	236
10	11	9	6	4	1	**45**	372	344	313	292	250
11	14	11	8	6	2	**46**	390	362	329	308	264
12	18	14	10	8	3	**47**	408	379	346	323	278
13	22	18	13	10	5	**48**	427	397	363	340	293
14	26	22	16	13	7	**49**	447	416	380	356	308
15	31	26	20	16	9	**50**	467	435	398	374	324
16	36	30	24	20	12	**51**	487	454	417	391	340
17	42	35	28	24	15	**52**	508	474	435	409	356
18	48	41	33	28	19	**53**	530	495	455	428	373
19	54	47	38	33	22	**54**	551	515	474	446	390
20	61	53	44	38	27	**55**	574	537	494	466	408
21	68	59	50	43	31	**56**	596	558	515	485	426
22	76	66	56	49	36	**57**	619	580	536	505	444
23	84	74	63	55	41	**58**	643	603	557	526	463
24	92	82	70	62	46	**59**	667	626	579	547	483
25	101	90	77	69	52	**60**	691	649	601	568	502
26	111	99	85	76	59	**61**	716	673	624	590	522
27	120	108	93	84	65	**62**	742	698	647	612	543
28	131	117	102	92	72	**63**	768	722	670	635	564
29	141	127	111	101	80	**64**	794	748	694	658	585
30	152	138	121	110	87	**65**	821	773	719	682	607
31	164	148	131	119	95	**66**	848	799	743	706	629
32	176	160	141	129	104	**67**	876	826	769	730	652
33	188	171	152	139	113	**68**	904	853	794	755	675
34	201	183	163	149	122	**69**	932	880	820	780	698
35	214	196	174	160	132	**70**	961	908	847	806	722
36	228	209	186	172	142	**71**	991	937	874	832	746
37	242	222	199	183	152	**72**	1021	965	902	859	771
38	257	236	212	195	163	**73**	1051	995	929	885	796
39	272	250	225	208	174	**74**	1082	1024	958	913	822
40	287	265	239	221	186	**75**	1113	1054	987	941	848
41	303	280	253	234	198	**76**	1145	1085	1016	969	874
42	320	295	267	248	210	**77**	1177	1116	1045	998	901
43	337	311	282	262	223	**78**	1210	1148	1076	1027	928

Upper Critical Values for the Kruskal-Wallis Test

Group Sizes			Nominal size α			
			0.10	0.05	0.025	0.01
2	2	2	4.571 (.06667)	- -	- -	- -
3	2	1	4.286 (.10000)	- -	- -	- -
3	2	2	4.500 (.06667)	4.714 (.04762)	- -	- -
3	3	1	4.571 (.10000)	5.143 (.04286)	- -	- -
3	3	2	4.556 (.10000)	5.361 (.03214)	5.556 (.02500)	- -
3	3	3	4.622 (.10000)	5.600 (.05000)	5.956 (.02500)	7.200 (.00357)
4	2	1	4.500 (.07619)	- -	- -	- -
4	2	2	4.458 (.10000)	5.333 (.03333)	5.500 (.02381)	- -
4	3	1	4.056 (.09286)	5.208 (.05000)	5.833 (.02143)	- -
4	3	2	4.511 (.09841)	5.444 (.04603)	6.000 (.02381)	6.444 (.00794)
4	3	3	4.709 (.09238)	5.791 (.04571)	6.155 (.02476)	6.745 (.01000)
4	4	1	4.167 (.08254)	4.967 (.04762)	6.167 (.02222)	6.667 (.00952)
4	4	2	4.555 (.09778)	5.455 (.04571)	6.327 (.02413)	7.036 (.00571)
4	4	3	4.545 (.09905)	5.598 (.04866)	6.394 (.02476)	7.144 (.00970)
4	4	4	4.654 (.09662)	5.692 (.04866)	6.615 (.02424)	7.654 (.00762)
5	2	1	4.200 (.09524)	5.000 (.04762)	- -	- -
5	2	2	4.373 (.08995)	5.160 (.03439)	6.000 (.01852)	6.533 (.00794)
5	3	1	4.018 (.09524)	4.960 (.04762)	6.044 (.01984)	- -
5	3	2	4.651 (.09127)	5.251 (.04921)	6.004 (.02460)	6.909 (.00873)
5	3	3	4.533 (.09697)	5.648 (.04892)	6.315 (.02121)	7.079 (.00866)
5	4	1	3.987 (.09841)	4.985 (.04444)	5.858 (.02381)	6.955 (.00794)
5	4	2	4.541 (.09841)	5.273 (.04877)	6.068 (.02482)	7.205 (.00895)
5	4	3	4.549 (.09892)	5.656 (.04863)	6.410 (.02496)	7.445 (.00974)
5	4	4	4.668 (.09817)	5.657 (.04906)	6.673 (.02429)	7.760 (.00946)
5	5	1	4.109 (.08586)	5.127 (.04618)	6.000 (.02165)	7.309 (.00938)
5	5	2	4.623 (.09704)	5.338 (.04726)	6.346 (.02489)	7.338 (.00962)
5	5	3	4.545 (.09965)	5.705 (.04612)	6.549 (.02436)	7.578 (.00968)
5	5	4	4.523 (.09935)	5.666 (.04931)	6.760 (.02490)	7.823 (.00978)
5	5	5	4.560 (.09952)	5.780 (.04878)	6.740 (.02475)	8.000 (.00946)

Kolmogorov[redacted]One-Sided Test

n	0.1	0.05	0.025	0.01	0.005
1	0.9000	0.9500	0.9750	0.9900	0.9950
2	0.6838	0.7764	0.8419	0.9000	0.9293
3	0.5648	0.6360	0.7076	0.7846	0.8290
4	0.4927	0.5652	0.6239	0.6889	0.7342
5	0.4470	0.5094	0.5633	0.6272	0.6685
6	0.4104	0.4680	0.5193	0.5774	0.6166
7	0.3815	0.4361	0.4834	0.5384	0.5758
8	0.3583	0.4096	0.4543	0.5065	0.5418
9	0.3391	0.3875	0.4300	0.4796	0.5133
10	0.3226	0.3687	0.4092	0.4566	0.4889
11	0.3083	0.3524	0.3912	0.4367	0.4677
12	0.2958	0.3382	0.3754	0.4192	0.4490
13	0.2847	0.3255	0.3614	0.4036	0.4325
14	0.2748	0.3142	0.3489	0.3897	0.4176
15	0.2659	0.3040	0.3376	0.3771	0.4042
16	0.2578	0.2947	0.3273	0.3657	0.3920
17	0.2504	0.2863	0.3180	0.3553	0.3809
18	0.2436	0.2785	0.3094	0.3457	0.3706
19	0.2373	0.2714	0.3014	0.3369	0.3612
20	0.2316	0.2647	0.2941	0.3287	0.3524
21	0.2262	0.2586	0.2872	0.3210	0.3443
22	0.2212	0.2528	0.2809	0.3139	0.3367
23	0.2165	0.2475	0.2749	0.3073	0.3295
24	0.2120	0.2424	0.2693	0.3010	0.3229
25	0.2079	0.2377	0.2640	0.2952	0.3166
26	0.2040	0.2332	0.2591	0.2896	0.3106
27	0.2003	0.2290	0.2544	0.2844	0.3050
28	0.1968	0.2250	0.2499	0.2794	0.2997
29	0.1935	0.2212	0.2457	0.2747	0.2947
30	0.1903	0.2176	0.2417	0.2702	0.2899
31	0.1873	0.2141	0.2379	0.2660	0.2853
32	0.1844	0.2108	0.2342	0.2619	0.2809
33	0.1817	0.2077	0.2308	0.2580	0.2768
34	0.1791	0.2047	0.2274	0.2543	0.2728
35	0.1766	0.2018	0.2242	0.2507	0.2690
36	0.1742	0.1991	0.2212	0.2473	0.2653
37	0.1719	0.1965	0.2183	0.2440	0.2618
38	0.1697	0.1939	0.2154	0.2409	0.2584
39	0.1675	0.1915	0.2127	0.2379	0.2552
40	0.1655	0.1891	0.2101	0.2349	0.2521
> 40	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.52/\sqrt{n}$	$1.63/\sqrt{n}$

Upper Critical Values of Spearman's Rank Correlation Coefficient R_s

Note: In the table below, the critical values give significance levels as close as possible to but not exceeding the nominal α.

	Nominal α					
n	0.10	0.05	0.025	0.01	0.005	0.001
4	1.000	1.000	-	-	-	-
5	0.800	0.900	1.000	1.000	-	-
6	0.657	0.829	0.886	0.943	1.000	-
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.678	0.727	0.818
13	0.385	0.484	0.560	0.648	0.703	0.791
14	0.367	0.464	0.538	0.626	0.679	0.771
15	0.354	0.446	0.521	0.604	0.654	0.750
16	0.341	0.429	0.503	0.582	0.635	0.729
17	0.328	0.414	0.488	0.566	0.618	0.711
18	0.317	0.401	0.472	0.550	0.600	0.692
19	0.309	0.391	0.460	0.535	0.584	0.675
20	0.299	0.380	0.447	0.522	0.570	0.662
21	0.292	0.370	0.436	0.509	0.556	0.647
22	0.284	0.361	0.425	0.497	0.544	0.633
23	0.278	0.353	0.416	0.486	0.532	0.621
24	0.271	0.344	0.407	0.476	0.521	0.609
25	0.265	0.337	0.398	0.466	0.511	0.597
26	0.259	0.331	0.390	0.457	0.501	0.586
27	0.255	0.324	0.383	0.449	0.492	0.576
28	0.250	0.318	0.375	0.441	0.483	0.567
29	0.245	0.312	0.368	0.433	0.475	0.558

(Continued)