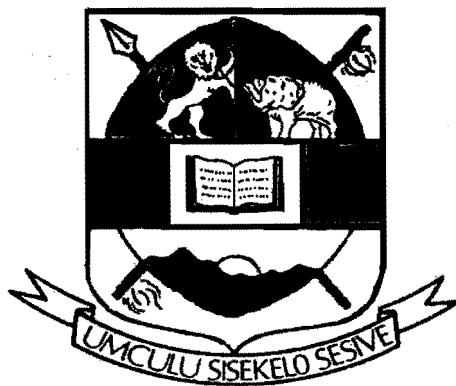


UNIVERSITY OF SWAZILAND



FINAL EXAMINATION PAPER 2017

TITLE OF PAPER: MULTIVARIATE ANALYSIS
COURSE CODE: ST 410
TIME ALLOCATED: 2 (TWO) HOURS
REQUIREMENTS: STATISTICAL TABLES AND CALCULATOR
INSTRUCTION: ANSWER ANY 3 (THREE) QUESTIONS OF YOUR CHOICE. ALL QUESTIONS CARRY THE MARKS AS INDICATED WITHIN THE PARENTHESIS

THIS PAPER IS NOT TO BE OPENED UNTIL PERMISSION HAS BEEN GRANTED BY THE INVIGILATOR.

QUESTION ONE**[7+4+1+8]**

- a) Find the matrix of variances and covariance for the following data:

X_1	X_2	X_3
4	3	10
5	2	11
8	6	15
9	6	9
10	8	5

- b) A regression model is to be developed for predicting the ability of soil to absorb chemical contaminants. Five observations have been taken on a soil absorption index (Y) and two regressors: X_1 = amount of extractable iron ore and X_2 = amount of bauxite. We wish to fit the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$. Some necessary quantities are:

$$X^T X^{-1} = \begin{bmatrix} 48.500922 & -6.016590 & 0.923963 \\ -6.016590 & 0.798618 & -0.131336 \\ 0.923963 & -0.131336 & 0.023041 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 2,047 \\ 30,989 \\ 94,400 \end{bmatrix}$$

- i. Estimate the regression coefficients in the model specified above.
- ii. What is the predicted value of the absorption index Y when $X_1 = 20$ and $X_2 = 200$?
- iii. Given that the total sum of squares for Y is 5,213.2. Test for significance of regression using $\alpha = 0.05$.

QUESTION TWO**[9+1+10]**

- a) Suppose we have three variable in each of the 3 groups with sample sizes $n_A = 3$, $n_B = 4$ and $n_C = 5$. Consider the following:

$$\mathbf{W} = \begin{bmatrix} 42.0 & 39.0 & 10.0 \\ 39.0 & 78.0 & 11.0 \\ 10.0 & 11.0 & 10.0 \end{bmatrix}, \quad \& \quad \mathbf{T} = \begin{bmatrix} 62.25 & 24.25 & 8.5 \\ 24.25 & 94.92 & 8.83 \\ 8.5 & 8.83 & 17.67 \end{bmatrix}$$

where \mathbf{W} is the within sum of square matrix and \mathbf{T} is the total sum of square matrix.

- a) Compute Wilk's Λ statistics. Use χ^2 approximation to test the equality of population mean vectors using $\alpha = 0.05$. Specify the null and alternative hypothesis.
- b) Comment on the results found in part (i).
- b) A medical researcher is interested in two particular fatty acids (A and B) found in human blood. Measurements (micrograms per gram) were taken on 16 new-born babies with Down's syndrome. The sample means were 70 and 50 for fatty acids A and B respectively, giving the sample mean vector,

$$\bar{X} = \begin{pmatrix} 70 \\ 50 \end{pmatrix}$$

and the corresponding (unbiased) sample covariance matrix was

$$S = \begin{pmatrix} 100 & 80 \\ 80 & 100 \end{pmatrix}$$

For non-Down's syndrome new-born babies the expected fatty acid levels are 80 and 65 for A and B respectively. Use the multivariate hypothesis test ($T^2 = n(\bar{X} - \mu_0)'S^{-1}(\bar{X} - \mu_0)$) to assess whether the observed data for the Down's syndrome babies are consistent with the expected values for non-Down's syndrome babies.

QUESTION THREE

[10+6+4]

- a) Discuss Factor analysis and principal component analysis with respect to the similarities and difference.
- b) The procedure for a factor analysis can be described in terms of three stages. State these three stages.
- c) The table below was found by running factor analysis using SPSS. Find the first equation of the un-rotated factor model. Assume $\lambda_1 = 3.112$

Component Matrix

	Component								
	1	2	3	4	5	6	7	8	9
X1	-.902	.032	-.341	.017	-.020	.024	-.089	.244	.002
X2	-.662	-.001	.630	.117	.292	-.111	.115	-.203	.001
X3	.434	-.581	-.614	.060	-.197	.017	.128	-.195	.001
X4	.558	-.147	-.359	.024	.720	-.115	-.033	.070	.000
X5	.391	.326	.087	.807	.052	.280	-.011	.014	.000
X6	.673	.549	.080	.174	.225	-.375	.095	.091	.001
X7	.231	.744	-.117	-.504	.111	.227	.248	-.012	.001
X8	.755	-.074	.441	-.327	-.039	.088	-.337	-.019	.001
X9	.362	-.695	.505	-.043	-.019	.079	.267	.227	.000

Extraction Method: Principal Component Analysis.

a. 9 components extracted.

QUESTION FOUR**[8+6+6]**

- a) Discuss the important of Principal Component Analysis and state its important properties.
- b) Explain the procedure for a Principal Component Analysis
- c) Consider the following table:

Eigenvalue	Eigenvectors						
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
3.111	0.512	0.375	-0.246	-0.315	-0.222	-0.382	-0.131
1.709	-0.024	0.000	0.432	0.109	-0.242	-0.408	-0.553
1.095	-0.278	0.516	-0.503	-0.292	0.071	0.064	-0.096
0.663	0.016	0.113	0.058	0.023	0.783	0.169	-0.489
0.305	0.025	-0.345	0.231	-0.854	-0.064	0.269	-0.133
0.108	-0.045	0.203	-0.028	0.208	-0.503	0.674	-0.399
0.009	0.166	-0.212	-0.238	0.065	0.014	-0.165	-0.463

- a. How many variables were there in the data set? How many components will you get?
- b. How many components will you choose? Explain why.
- c. List those selected components and interpret those in terms of original variables, X_i's.

QUESTION FIVE**[7+2+4+7]**

- a) Explain why the Discriminant Function Analysis is different from the Principle Component Analysis and the Factor Analysis.
- b) The following table shows the eigenvalues and corresponding eigenvectors of W⁻¹B:

Component	Eigenvalue	Eigenvectors			
		X ₁	X ₂	X ₃	X ₄
1	0.437	-0.0107	0.0040	0.0119	-0.0068
2	0.035	0.0031	0.0168	-0.0046	-0.0022
3	0.015	-0.0068	0.0010	0.0000	0.0247
4	0.002	0.0126	-0.0001	0.0112	0.0054

- i. How many groups and variables were considered in this problem?
- ii. List all the canonical discriminant functions.
- iii. Assuming that the ith sample size n_i = 30 for all i = 1, 2, 3, 4, 5; test whether each of these functions varies significantly from group to group.

END OF EXAMINATION

STATISTICAL TABLES

Cumulative normal distribution

Critical values of the *t* distribution

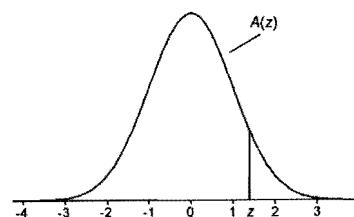
Critical values of the *F* distribution

Critical values of the chi-squared distribution

STATISTICAL TABLES

1

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							

TABLE A.3 (continued)

F Distribution: Critical Values of *F* (0.1% significance level)

v_1	25	30	35	40	50	60	75	100	150	200
v_2										
1	6.24e05	6.26e05	6.28e05	6.29e05	6.30e05	6.31e05	6.32e05	6.33e05	6.35e05	6.35e05
2	999.49	999.47	999.47	999.47	999.48	999.49	999.49	999.49	999.49	999.49
3	125.84	125.45	125.17	124.96	124.66	124.47	124.27	124.07	123.87	123.77
4	45.70	45.43	45.23	45.09	44.88	44.75	44.61	44.47	44.33	44.26
5	25.08	24.87	24.72	24.60	24.44	24.33	24.22	24.12	24.01	23.95
6	16.85	16.67	16.54	16.44	16.31	16.21	16.12	16.03	15.93	15.89
7	12.69	12.53	12.41	12.33	12.20	12.12	12.04	11.95	11.87	11.82
8	10.26	10.11	10.00	9.92	9.80	9.73	9.65	9.57	9.49	9.45
9	8.69	8.55	8.46	8.37	8.26	8.19	8.11	8.04	7.96	7.93
10	7.60	7.47	7.37	7.30	7.19	7.12	7.05	6.98	6.91	6.87
11	6.81	6.68	6.59	6.52	6.42	6.35	6.28	6.21	6.14	6.10
12	6.22	6.09	6.00	5.93	5.83	5.76	5.70	5.63	5.56	5.52
13	5.75	5.63	5.54	5.47	5.37	5.30	5.24	5.17	5.10	5.07
14	5.38	5.25	5.17	5.10	5.00	4.94	4.87	4.81	4.74	4.71
15	5.07	4.95	4.86	4.80	4.70	4.64	4.57	4.51	4.44	4.41
16	4.82	4.70	4.61	4.54	4.45	4.39	4.32	4.26	4.19	4.16
17	4.60	4.48	4.40	4.33	4.24	4.18	4.11	4.05	3.98	3.95
18	4.42	4.30	4.22	4.15	4.06	4.00	3.93	3.87	3.80	3.77
19	4.26	4.14	4.06	3.99	3.90	3.84	3.78	3.71	3.65	3.61
20	4.12	4.00	3.92	3.86	3.77	3.70	3.64	3.58	3.51	3.48
21	4.00	3.88	3.80	3.74	3.64	3.58	3.52	3.46	3.39	3.36
22	3.89	3.78	3.70	3.63	3.54	3.48	3.41	3.35	3.28	3.25
23	3.79	3.68	3.60	3.53	3.44	3.38	3.32	3.25	3.19	3.16
24	3.71	3.59	3.51	3.45	3.36	3.29	3.23	3.17	3.10	3.07
25	3.63	3.52	3.43	3.37	3.28	3.22	3.15	3.09	3.03	2.99
26	3.56	3.44	3.36	3.30	3.21	3.15	3.08	3.02	2.95	2.92
27	3.49	3.38	3.30	3.23	3.14	3.08	3.02	2.96	2.89	2.86
28	3.43	3.32	3.24	3.18	3.09	3.02	2.96	2.90	2.83	2.80
29	3.38	3.27	3.18	3.12	3.03	2.97	2.91	2.84	2.78	2.74
30	3.33	3.22	3.13	3.07	2.98	2.92	2.86	2.79	2.73	2.69
35	3.13	3.02	2.93	2.87	2.78	2.72	2.66	2.59	2.52	2.49
40	2.98	2.87	2.79	2.73	2.64	2.57	2.51	2.44	2.38	2.34
50	2.79	2.68	2.60	2.53	2.44	2.38	2.31	2.25	2.18	2.14
60	2.67	2.55	2.47	2.41	2.32	2.25	2.19	2.12	2.05	2.01
70	2.58	2.47	2.39	2.32	2.23	2.16	2.10	2.03	1.95	1.92
80	2.52	2.41	2.32	2.26	2.16	2.10	2.03	1.96	1.89	1.85
90	2.47	2.36	2.27	2.21	2.11	2.05	1.98	1.91	1.83	1.79
100	2.43	2.32	2.24	2.17	2.08	2.01	1.94	1.87	1.79	1.75
120	2.37	2.26	2.18	2.11	2.02	1.95	1.88	1.81	1.73	1.68
150	2.32	2.21	2.12	2.06	1.96	1.89	1.82	1.74	1.66	1.62
200	2.26	2.15	2.07	2.00	1.90	1.83	1.76	1.68	1.60	1.55
250	2.23	2.12	2.03	1.97	1.87	1.80	1.72	1.65	1.56	1.51
300	2.21	2.10	2.01	1.94	1.85	1.78	1.70	1.62	1.53	1.48
400	2.18	2.07	1.98	1.92	1.82	1.75	1.67	1.59	1.50	1.45
500	2.17	2.05	1.97	1.90	1.80	1.73	1.65	1.57	1.48	1.43
600	2.16	2.04	1.96	1.89	1.79	1.72	1.64	1.56	1.46	1.41
750	2.15	2.03	1.95	1.88	1.78	1.71	1.63	1.55	1.45	1.40
1000	2.14	2.02	1.94	1.87	1.77	1.69	1.62	1.53	1.44	1.38

TABLE A.4

 χ^2 (Chi-Squared) Distribution: Critical Values of χ^2

Significance level			
Degrees of freedom	5%	1%	0.1%
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	11.070	15.086	20.515
6	12.592	16.812	22.458
7	14.067	18.475	24.322
8	15.507	20.090	26.124
9	16.919	21.666	27.877
10	18.307	23.209	29.588