

UNIVERSITY OF SWAZILAND



SUPPLEMENTARY EXAMINATION PAPER 2012

TITLE OF PAPER : TOPICS IN STATISTICS
(STATISTICAL MODELLING)

COURSE CODE : ST 405

TIME ALLOWED : THREE (3) HOURS

REQUIREMENTS : CALCULATOR AND STATISTICAL TABLES

INSTRUCTIONS : ANSWER ANY FIVE QUESTIONS

Question 1

In surgery, it is desirable to give enough anaesthetic so that patients do not move when an incision is made. It is also desirable not to use much more anaesthetic than necessary. In an experiment, patients are given different concentrations of anaesthetic. The response variable is whether or not they move at the time of incision 15 minutes after receiving the drug.

	Concentration					
	0.8	1.0	1.2	1.4	1.6	2.5
Move	6	4	2	2	0	0
No move	1	1	4	4	4	2
Total	7	5	6	6	4	2
Proportion	0.17	0.20	0.67	0.67	1.00	1.00

- (a) Suggest an appropriate model to explain the impact of anaesthetic on the response variable. (6 Marks)
- (b) Write an R program which reads these data into R data set called `ane`. The program should then produce a contingency table and a glm analysis (8 Marks)
- (c) From the glm analysis below, what can you conclude between concentrations of anaesthetic and movement of patients?

```
      coef.est coef.se
(Intercept) -6.469   2.418
conc         5.567   2.044
```

n = 30, k = 2

residual deviance = 27.8, null deviance = 41.5 (difference = 13.7)

(6 Marks)

Question 2

The table below is a randomized clinical trial of a respiratory disorder.

Treatment	Outcome		Total
	Favourable	Unfavourable	
Placebo	16	48	64
Test	40	20	60
Total	56	68	124

Use the Randomization Chi-square, Pearson's Chi-Square, and Mantel-Haenszel Chi-Square to determine if the new treatment is effective for respiratory disorders.

(20 Marks)

Question 3

The table below displays a 4-way cross-classification of data related to complaints of symptoms of a respiratory disease, byssinosis, which occurs among textile mill workers.

Table. Frequency table of Byssinosis Complaints

WORKPLACE CONDITIONS	YEARS EMPLOYMENT	SMOKING	COMPLAINTS	
			yes	no
Dusty	<10	yes	30	203
Dusty	<10	no	7	119
Dusty	>=10	yes	57	161
Dusty	>=10	no	11	81
Not Dusty	<10	yes	14	1340
Not Dusty	<10	no	12	1004
Not Dusty	>=10	yes	24	1360
Not Dusty	>=10	no	10	986

- a) Create 2×2 tables stratified by years of employment and test for independence between Workplace conditions and complaints of byssinosis by comparing the Mantel- Haenszel Chi-Square test and Mantel- Haenszel Odds Ratio test.

(20 Marks)

Question 4

In the 2002 Winter Olympic Games held at Salt Lake City there was concern that figure skating judges may have judged with bias for certain skaters according to geopolitical preferences. Consider Table B which presents the results of the 9 judges of the "long program" in women's figure skating. All judges rated these two skaters first or second; the first place rating, or preferred skater, is reported in the table. Each judge is placed into a region based upon her/his country of origin: EE refers to Belarus, Russia and Slovakia; WE refers to Denmark, Finland, Italy, and Germany; and NA refers to Canada and the United States.

TABLE B. Number of First place ratings

Region	Figure Skater		Total
	Slutskaya	Hughes	
EE	3	0	3
WE	1	3	4
NA	0	2	2
Total	4	5	9

Briefly, summarize your results for a report to the International Olympic Committee (IOC). What do you conclude about the charge of geopolitically based bias in figure skating judging? Explain any assumptions of the statistical methods that led you to this conclusion (the IOC will consult a statistician to independently review your findings.)

(20 Marks)

Question 5

A marketing research firm was engaged by an automobile manufacturer to conduct a pilot study to examine the feasibility of using logistic regression for ascertaining the likelihood that a family will purchase a new car during the next year. A random sample of 33 suburban families was selected. Data on annual family income and the current age of the oldest family automobile were obtained. A follow-up interview conducted 12 months later was used to determine whether the family actually purchased a new car or did not purchase a new car. The model in Appendix I was fitted;

- a) State the response function.
(3 Marks)
- b) Using the output in Appendix I (coefficients) advise appropriately.
(10 marks)
- c) What is the estimated probability that a family with annual income of E50,000 and an oldest car of 3 years will purchase a new car next year?
(3 marks)
- d) Using the output Appendix II, state whether the two-factor interaction effect between annual family income and age of oldest automobile should be added to the regression model containing family income and age of oldest automobile as first-order terms; use $\alpha = 0.05$. What is the approximate p-value?
(4 marks)

Question 6

A cohort of subjects, some non-smokers and others smokers, was observed for several years. The number of cases of cancer of the lung diagnosed among the different categories was recorded. Data regarding the number of years of smoking were also obtained from each individual. For each category the person-years of observation were calculated. The investigators wish to address the question of the relative risks of smoking. In the observed data the average number of cigarettes smoked per day represents the daily dose, and the years of smoking together with the average number of cigarettes smoked daily represents the total dose inhaled over time. The results of the analysis are given in Appendix III. Justify the method of analysis, state the model, interpret all relevant estimates and write a short report.

APPENDICES

APPENDIX I

```
> data <- read.table("car_table.txt", header=T)
> attach(data)
> glm1 <- glm(purchase~income+age, family="binomial")
> summary(glm1)
```

Call:

```
glm(formula = purchase ~ income + age, family = "binomial")
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6189	-0.8949	-0.5880	0.9653	2.0846

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.73931	2.10195	-2.255	0.0242 *
income	0.06773	0.02806	2.414	0.0158 *
age	0.59863	0.39007	1.535	0.1249

APPENDIX II

```
> glm3<-update(glm1, .~.+age:income)
> summary(glm3)
```

Call:

```
glm(formula = purchase ~ income + age + income:age, family = "binomial")
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6096	-0.8222	-0.5334	0.8731	1.9924

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.372993	2.862477	-0.829	0.407
income	0.001326	0.064770	0.020	0.984
age	-0.303860	0.890512	-0.341	0.733
income:age	0.028860	0.026493	1.089	0.276

APPENDIX III

Response variate: CASES
Distribution: Poisson
Link function: Log
Fitted terms: Constant, PERSONYR, CIGS_DAY, SMOKING_

*** Summary of analysis ***

	d.f.	deviance	mean deviance	deviance ratio
Regression	3	63.168816931	21.056272310	21.06
Residual	31	74.122027311	2.391033139	
Total	34	137.290844242	4.037966007	

Change -3 -63.168816931 21.056272310 21.06
* MESSAGE: ratios are based on dispersion parameter with value 1

*** Estimates of regression coefficients ***

	estimate	s.e.	t(*)
Constant	-4.669	0.988	-4.72
PERSONYR	0.000410	0.000104	3.94
CIGS_DAY	0.0559	0.0100	5.58
SMOKING_	0.0888	0.0166	5.34

* MESSAGE: s.e.s are based on dispersion parameter with value 1

STATISTICAL TABLES

Cumulative normal distribution

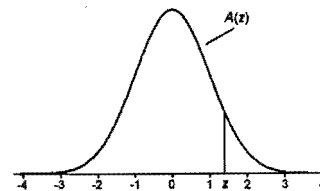
Critical values of the *t* distribution

Critical values of the *F* distribution

Critical values of the chi-squared distribution

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
1	6.24605	6.26603	6.28603	6.29605	6.30605	6.31605	6.32605	6.33605	6.34605	6.35603
2	999.46	999.47	999.47	999.47	999.48	999.48	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	2.13	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

Degrees of freedom	Significance level		
	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