Department of Mathematics and Statistics
The University of Toledo

Master's Comprehensive Examination
Applied Statistics

April 21, 2012

Instructions:

Do all four problems;

Show all of your computations;

Prove all of your assertions or quote appropriate theorems;

This is three-hour open book examination.

1. (35 points) Suppose random vector $\mathbf{X} \sim N_2(\mu, \Sigma)$, where μ and Σ are unknown. A random sample of size $n = 4$ yields the following data matrix:

$$\begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \mathbf{x}_3' \\ \mathbf{x}_4' \end{pmatrix} = \begin{pmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{pmatrix}$$

a. What is the sample mean $\bar{\mathbf{x}}$?
b. What is the sample variance-covariance matrix $\mathbf{s}$?
c. At the significance level $\alpha = 0.05$, test

$$H_0 : \mu = \begin{pmatrix} 7 \\ 11 \end{pmatrix} \text{ vs } H_a : \mu \neq \begin{pmatrix} 7 \\ 11 \end{pmatrix}$$

d. Give a 95% confidence region for μ. Be sure to specify the center, and the half length of major and minor axes of the confidence ellipsoid.
e. Is $\mu = (8, 12)'$ in the confidence region above?

Hint: Suppose the inverse of a 2×2 matrix $\mathbf{A}$ exists. Then

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

2. (15 points) Fibrinogen (`fib`) and Globulin (`glo`) are two types of plasma proteins. The following data were collected to determine whether there is any association between the probability of an erythrocyte sedimentation rate (`esr`) reading greater than 20mm/hr and the levels of these two plasma proteins.

fib	glo	esr	fib	glo	esr
2.52	38	0	2.88	30	0
2.56	31	0	2.65	46	0
2.19	33	0	2.28	36	0
⋮	⋮	⋮	⋮	⋮	⋮

"0" denotes "esr $<$ 20" and "1" denotes "esr $>$ 20"

Using the SAS output in the appendix, answer the following questions:

a. Write down the full model.
b. Based on the SAS output, write down the most adequate model.
c. Based on the model you choose in (b), whether is there any association between the probability of an esr reading greater than 20mm/hr and the levels of the two plasma proteins? Briefly interpret the estimates of the coefficients in the model.

3. Do twins have the same IQ? To study this issue, a random sample of 6 pairs of twins was taken and their full scale IQ were recorded. Below the results are summarized in a table:

Older twin	96	89	102	104	129	98
Younger twin	89	87	103	96	125	101

a. (15 points) Use Wilcoxon signed rank test at level $\alpha = 0.05$ to decide whether there is a significant difference between twins IQ. Find the exact p-value and use it to make your decision.

b. (5 points) Use the R output to find the test statistics and p-value. Is the result the same with a?

4. Patient satisfaction. A hospital administrator wished to study the relation between patient satisfaction (y) and patient's age (x_1, in years), severity of illness (x_2, an index), and anxiety level (x_3, an index). The administrator randomly selected 46 patients and collected the data.

a. (5 points) Use C_p criterion to choose the best subset of variables in the full model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ and give this best model. (Use the best model to answer following questions)

b. (5 points) Find estimate of β values in the chosen model from C_p criterion.

c. (5 points) Obtain the studentized deleted residuals and identify any outlying y observations. Use the Bonferroni outlier test procedure with $\alpha = .10$. State the decision rule and conclusion.

d. (5 points) Obtain the diagonal elements of the hat matrix. Identify any outlying x observations.

e. (5 points) Give the estimate of mean patient satisfaction for patients who are $x_1 = 30$ years old, whose index of illness severity is $x_2 = 58$, and whose index of anxiety level is $x_3 = 2.0$. Find the variance inflation factors. Do they indicate that a serious multicollinearity problem exists here?

f. (5 points) The two largest absolute studentized deleted residuals are for cases 11 and 27. Obtain the DFFITS, DFBETAS, and Cook's distance values for this case to assess its influence. What do you conclude?

```
****************************************
*  Appendix for Problem 2
****************************************
data plasma;
input fib glo esr;
datalines;
2.52  38    0
2.56  31    0
...
2.09  44    1
3.93  32    1
;
run;
proc logistic data=plasma;
  model esr=fib glo / expb
  SELECTION=B AGGREGATE=(fib glo) scale=none;
run;
```

****

The SAS System

The LOGISTIC Procedure

Model Information

Data Set	WORK.PLASMA
Response Variable	esr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	32
Number of Observations Used	32

Response Profile

Ordered Value	esr	Total Frequency
1	0	26
2	1	6

Probability modeled is esr=0.

Backward Elimination Procedure

Step 0. The following effects were entered:

Intercept fib glo

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	32.885	28.971
SC	34.351	33.368
-2 Log L	30.885	22.971

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7.9138	2	0.0191
Score	8.2067	2	0.0165
Wald	4.7561	2	0.0927

Step 1. Effect glo is removed:

****

The SAS System

The LOGISTIC Procedure
Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	32.885	28.840
SC	34.351	31.772
-2 Log L	30.885	24.840

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	6.0446	1	0.0139
Score	6.7522	1	0.0094
Wald	4.1134	1	0.0425

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
1.8676	1	0.1718

NOTE: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	glo	1	1	1.6982	0.1925

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	24.8404	30	0.8280	0.7328
Pearson	41.8135	30	1.3938	0.0742

Number of unique profiles: 32

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)

Intercept	1	6.8451	2.7703	6.1053	0.0135	939.243
fib	1	-1.8271	0.9009	4.1134	0.0425	0.161

****

The SAS System

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
fib	0.161	0.028	0.940

Association of Predicted Probabilities and Observed Responses

Percent Concordant	71.2	Somers' D	0.429
Percent Discordant	28.2	Gamma	0.432
Percent Tied	0.6	Tau-a	0.135
Pairs	156	c	0.715

****

****
* Appendix for Problem 3
****

```
x = c(96,89,102,104,129,98)
y = c(89,87,103,96,125,101)
wilcox.test(x, y, alternative = c("two.sided"), mu = 0, paired = TRUE)
```

Wilcoxon signed rank test

data: x and y
V = 17, p-value = 0.2188
alternative hypothesis: true location shift is not equal to 0.

```
****************************************
*  Appendix for Problem 4
****************************************

data PatientSatisfaction;
input y x1 x2 x3;
datalines;
  48    50    51   2.3
  57    36    46   2.3
  ......
  92    28    46   1.8
;
run;
proc reg data=PatientSatisfaction;
model y = x1 x2 x3/selection = cp b r vif influence;
run;
quit;
```

The REG Procedure
Model: MODEL1
Dependent Variable: y

C(p) Selection Method

Number of Observations Read	46
Number of Observations Used	46

Number in Model	C(p)	R-Square	Parameter Estimates: Intercept	x1	x2	x3
2	2.8072	0.6761	145.94123	-1.20047	.	-16.74205
3	4.0000	0.6822	158.49125	-1.14161	-0.44200	-13.47016
2	5.5997	0.6550	156.67186	-1.26765	-0.92079	.
1	8.3536	0.6190	119.94317	-1.52060	.	.
2	30.2471	0.4685	181.57256	.	-1.23948	-25.14023
1	35.2456	0.4155	146.44943	.	.	-37.11667
1	42.1123	0.3635	183.07696	.	-2.40928	.

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	46
Number of Observations Used	46

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	9038.80461	4519.40231	44.88	<.0001
Error	43	4330.49973	100.70930		
Corrected Total	45	13369			

Root MSE	10.03540	R-Square	0.6761
Dependent Mean	61.56522	Adj R-Sq	0.6610
Coeff Var	16.30044		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	145.94123	11.52509	12.66	<.0001	0
x1	1	-1.20047	0.20411	-5.88	<.0001	1.48046
x3	1	-16.74205	6.08083	-2.75	0.0086	1.48046

The REG Procedure
Model: MODEL1
Dependent Variable: y

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	48.0000	47.4109	2.7560	0.5891	9.650	0.0610		0.000
2	57.0000	64.2175	1.5742	-7.2175	9.911	-0.728	*	0.004
3	66.0000	61.0899	1.6657	4.9101	9.896	0.496		0.002
4	70.0000	66.5862	3.6107	3.4138	9.363	0.365		0.007
5	89.0000	82.1923	2.8810	6.8077	9.613	0.708	*	0.015
6	36.0000	38.5662	3.4026	-2.5662	9.441	-0.272		0.003
7	46.0000	58.6889	1.8588	-12.6889	9.862	-1.287	**	0.020
8	54.0000	51.7391	1.8506	2.2609	9.863	0.229		0.001
9	26.0000	34.9648	3.4646	-8.9648	9.418	-0.952	*	0.041
10	77.0000	75.9692	2.1617	1.0308	9.800	0.105		0.000
11	89.0000	70.9466	2.7995	18.0534	9.637	1.873	***	0.099
12	67.0000	54.1400	1.6763	12.8600	9.894	1.300	**	0.016
13	47.0000	63.4908	1.5579	-16.4908	9.914	-1.663	***	0.023
14	51.0000	66.6185	1.7550	-15.6185	9.881	-1.581	***	0.026
15	57.0000	45.4837	3.6271	11.5163	9.357	1.231	**	0.076
16	66.0000	69.2402	2.1218	-3.2402	9.809	-0.330		0.002
17	79.0000	64.4705	2.5890	14.5295	9.696	1.499	**	0.053
18	88.0000	79.3177	2.5023	8.6823	9.718	0.893	*	0.018
19	60.0000	71.1674	1.8076	-11.1674	9.871	-1.131	**	0.014
20	49.0000	39.7344	3.3909	9.2656	9.445	0.981	*	0.041
21	77.0000	72.6208	2.4583	4.3792	9.730	0.450		0.004
22	52.0000	44.5685	3.5404	7.4315	9.390	0.791	*	0.030
23	60.0000	55.8142	1.7307	4.1858	9.885	0.423		0.002
24	86.0000	88.1947	3.1981	-2.1947	9.512	-0.231		0.002
25	43.0000	47.6639	2.0884	-4.6639	9.816	-0.475		0.003
26	34.0000	38.0602	3.2180	-4.0602	9.505	-0.427		0.007
27	63.0000	82.4453	2.6964	-19.4453	9.666	-2.012	****	0.105
28	72.0000	63.9968	3.2162	8.0032	9.506	0.842	*	0.027
29	57.0000	67.3452	2.3206	-10.3452	9.763	-1.060	**	0.021
30	55.0000	50.3179	2.6330	4.6821	9.684	0.483		0.006
31	59.0000	72.8416	2.0633	-13.8416	9.821	-1.409	**	0.029
32	83.0000	72.5886	3.0901	10.4114	9.548	1.090	**	0.042
33	76.0000	75.2425	2.1239	0.7575	9.808	0.0772		0.000
34	47.0000	61.0899	1.6657	-14.0899	9.896	-1.424	**	0.019
35	36.0000	35.4385	3.1960	0.5615	9.513	0.0590		0.000
36	80.0000	68.2927	1.6530	11.7073	9.898	1.183	**	0.013
37	82.0000	69.2724	3.2204	12.7276	9.505	1.339	**	0.069
38	64.0000	69.7462	2.6337	-5.7462	9.684	-0.593	*	0.009
39	37.0000	49.3381	2.0912	-12.3381	9.815	-1.257	**	0.024
40	42.0000	45.9897	2.2560	-3.9897	9.779	-0.408		0.003
41	66.0000	55.8142	1.7307	10.1858	9.885	1.030	**	0.011
42	83.0000	86.0468	3.1266	-3.0468	9.536	-0.320		0.004
43	37.0000	49.5911	2.1542	-12.5911	9.801	-1.285	**	0.027
44	68.0000	55.0875	2.2585	12.9125	9.778	1.321	**	0.031
45	59.0000	66.3655	1.7873	-7.3655	9.875	-0.746	*	0.006
46	92.0000	82.1923	2.8810	9.8077	9.613	1.020	**	0.031

The REG Procedure
Model: MODEL1
Dependent Variable: y

Output Statistics

Obs	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS Intercept	DFBETAS x1	DFBETAS x3
1	0.0603	0.0754	1.1604	0.0172	0.0009	0.0145	-0.0079
2	-0.7242	0.0246	1.0600	-0.1150	-0.0089	0.0390	-0.0261
3	0.4918	0.0276	1.0847	0.0828	0.0311	0.0313	-0.0356
4	0.3609	0.1295	1.2213	0.1392	0.1010	0.0855	-0.1258
5	0.7040	0.0824	1.1291	0.2110	0.1904	-0.0318	-0.1284
6	-0.2689	0.1150	1.2063	-0.0969	0.0811	-0.0012	-0.0710
7	-1.2968	0.0343	0.9878	-0.2444	-0.0827	-0.1365	0.1247
8	0.2267	0.0340	1.1069	0.0425	-0.0083	0.0220	-0.0019
9	-0.9508	0.1192	1.1430	-0.3498	0.2870	-0.0660	-0.2166
10	0.1040	0.0464	1.1245	0.0229	0.0117	-0.0135	-0.0005
11	1.9319	0.0778	0.9014	0.5612	-0.0777	-0.4628	0.3567
12	1.3105	0.0279	0.9789	0.2220	-0.0485	0.0727	0.0201
13	-1.6996	0.0241	0.9007	-0.2671	-0.1067	-0.0379	0.0828
14	-1.6097	0.0306	0.9250	-0.2859	-0.0193	0.1534	-0.0961
15	1.2384	0.1306	1.1085	0.4800	0.0854	0.4345	-0.2948
16	-0.3269	0.0447	1.1148	-0.0707	-0.0539	-0.0169	0.0489
17	1.5213	0.0666	0.9788	0.4062	-0.1380	-0.2884	0.3016
18	0.8913	0.0622	1.0817	0.2295	0.1929	-0.0529	-0.1156
19	-1.1351	0.0324	1.0130	-0.2079	-0.1280	0.0521	0.0586
20	0.9806	0.1142	1.1319	0.3520	-0.0361	0.3113	-0.1291
21	0.4459	0.0600	1.1256	0.1127	0.0051	-0.0899	0.0537
22	0.7879	0.1245	1.1729	0.2971	-0.2396	-0.0821	0.2581
23	0.4194	0.0297	1.0923	0.0734	0.0056	0.0380	-0.0194
24	-0.2282	0.1016	1.1900	-0.0767	-0.0622	0.0349	0.0281
25	-0.4708	0.0433	1.1041	-0.1002	0.0412	-0.0489	-0.0141
26	-0.4230	0.1028	1.1810	-0.1432	0.0377	-0.1180	0.0283
27	-2.0888	0.0722	0.8599	-0.5827	-0.3456	0.3758	0.0406
28	0.8390	0.1027	1.1378	0.2839	-0.1217	-0.2108	0.2336
29	-1.0611	0.0535	1.0473	-0.2522	0.0416	0.1843	-0.1555
30	0.4791	0.0688	1.1338	0.1303	-0.0921	-0.0344	0.1035
31	-1.4262	0.0423	0.9723	-0.2996	-0.2340	0.0154	0.1624
32	1.0929	0.0948	1.0899	0.3537	0.2990	0.1372	-0.3071
33	0.0763	0.0448	1.1230	0.0165	0.0125	-0.0040	-0.0069
34	-1.4415	0.0276	0.9547	-0.2426	-0.0911	-0.0917	0.1043
35	0.0583	0.1014	1.1940	0.0196	-0.0144	0.0074	0.0087
36	1.1884	0.0271	0.9989	0.1985	0.0740	-0.0714	-0.0022
37	1.3519	0.1030	1.0528	0.4581	-0.1260	-0.3776	0.3396
38	-0.5889	0.0689	1.1244	-0.1602	0.0235	0.1280	-0.1011
39	-1.2658	0.0434	1.0027	-0.2697	0.0464	-0.1761	0.0404
40	-0.4040	0.0505	1.1172	-0.0932	0.0560	-0.0278	-0.0373
41	1.0312	0.0297	1.0261	0.1805	0.0138	0.0934	-0.0476
42	-0.3161	0.0971	1.1801	-0.1037	-0.0529	0.0780	-0.0053
43	-1.2947	0.0461	1.0004	-0.2845	0.1798	-0.0080	-0.1653
44	1.3324	0.0507	0.9984	0.3078	0.0863	0.2249	-0.1768
45	-0.7420	0.0317	1.0658	-0.1343	-0.0841	-0.0273	0.0733
46	1.0208	0.0824	1.0866	0.3059	0.2761	-0.0461	-0.1861