Applied Statistics Comprehensive Examination

In-Class Portion over STAT 519 and STAT 528

January 9, 2009, 1:00pm - 5:00pm

This exam covers material in STAT 519 *Probability Theory* and STAT 528 *Mathematical Statistics*. You are allowed one sheet of formulas and a calculator.

- Justify all of your answers and show all your work. An answer without justification is the same as no answer.
- Be as clear and complete as possible.
- Start each problem on a new page.
- Write on only one side of each page.
- Leave margins on 4 sides of each page for making copies.
- Write your name on each page.
- Do not staple your pages.
- Do as many problems as you can.

1. Three balls will be drawn at random one by one <u>with</u> replacement from a basket containing 5 red, 3 white and 2 blue balls.

   (a) Find the probability that <u>at least</u> one white ball will be drawn.

   (b) Find the probability that <u>exactly</u> two red balls will be drawn.

   (c) Find the probability that <u>more</u> red balls will be drawn than white balls.

2. Suppose that event A occurs at intervals that are IID exponential (α), and event B occurs at intervals that are IID exponential (β). Here, the parameters denote the hazard rates, the means being their reciprocals. Suppose that a timer is started when event B occurs (call it the 0-th occurrence of B) and it is stopped when event B occurs three more times. Let N be the number of times event A occurs while the timer is on. Find the distribution of N.

3. Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size 4 from a distribution having pdf $f(x) = 2x^{-3}$, $1 < x < \infty$, and zero elsewhere.

   (a) Evaluate $P(Y_4 \leq 2.0)$.

   (b) Are Y_1 and $Y_4 - Y_1$ independent? Justify.

   (c) Find (i) the expectation and (ii) the variance of $Y_1 + Y_2 + Y_3 + Y_4$.

4. I need 900 tiles to pave my 15' x 60' basement floor. The tile manufacturer tells me that typically 5% of the tiles they ship reach their destination damaged. Find the minimum number of tiles I should order so that there is a 99% probability of receiving at least 900 undamaged tiles. Clearly state all assumptions you are making and all results you are using.

5. Suppose that X_1, X_2, X_3, $\cdots$ is a random sample from a Weibull distribution having pdf $f(x;\theta) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha} e^{-(x/\theta)^\alpha}$, $0 < x < \infty$, zero elsewhere. Suppose that $\alpha > 1$ is known and $\theta > 0$ is unknown.

   (a) Find the MLE of θ and the MLE of θ^α.

   (b) Find the Rao-Cramer lower bound for the variance of unbiased estimators of θ^α .

   (c) Find a suitable function of the MLE of θ^α which is an unbiased estimator of θ^α , and find its efficiency.

6. Suppose that X_1, X_2, X_3, $\cdots$ is a random sample from a $\mathcal{N}(\mu, \theta)$ distribution, where the mean μ and the variance θ are unknown parameters. Let $\bar{X}$ be the sample mean and S^2 be the sample variance. Also let $W = \sum(X_{(i+1)} - X_{(i)})^2/S^2$.

   (a) Find the sufficient statistics for (μ, θ). Is it complete?

   (b) Are $\bar{X}$ and S^2 independent? Justify.

   (c) Show that W is independent of $(\bar{X}, S^2)$.

7. Suppose that $X_1,\ X_2,\ \cdots, X_n$ is a random sample from a beta $(\theta,\ 1)$ distribution, where $0 < \theta < \infty$ is unknown.

(a) Find a complete sufficient statistic for θ? Is it minimal sufficient?

(b) Find the likelihood ratio test for $H_0 : \theta = 2$ vs. $H_1 : \theta \neq 2$.

(c) Is this test in part (b) uniformly most powerful?

(d) Derive the sequential probability ratio test for $H_0 : \theta = 2$ vs. $H_1 : \theta = 3$, if it is desired to attain $\alpha_a = 0.10$, $\beta_a = 0.10$.

Applied Statistics Comprehensive Examination

Take-Home Portion over STAT 512, STAT 514 and STAT 524

January 9, 2009

Due: Tuesday, January 13, 2009, 4:00pm

This exam covers material in STAT 512 *Applied Regression Analysis*, STAT 514 *Design of Experiments* and STAT 524 *Applied Multivariate Analysis*. Read all the instructions carefully:

- Justify all of your answers and show all your work.
- Be as clear, complete, and succinct as possible.
- Start each problem on a new page.
- Write on only one side of each page.
- Clearly state the assumptions you are making whenever needed.
- If there are ambiguities, choose the most reasonable interpretation. However, be prepared to answer the question with alternative interpretations as well.
- Include only the most essential and enlightening computer output and no code.
- All datasets can be found at http://www.math.iupui.edu/∼fli/Comp/.
- Do not discuss any aspect of this exam with anyone else until after the completion of all oral examinations.
- Place one copy of your written report in each of the mailboxes of Ben Boukai, Samiran Ghosh, Fang Li, Hanxiang Peng, and Jyoti Sarkar before 4:00pm on Tuesday, January 13, 2009.

1. **STAT 512** *Applied Regression Problem*

The data file *http://www.math.iupui.edu/∼fli/Comp/Senic.txt* contains data on the efficacy of Nosocomial Infection Control (Senic Project). We wish to come up with some model for predicting *Length of stay* based on several variables included in the data set.

Variables in order	Description
number	1-113
Length of stay	Average Length of stay (in days)
Age	Average age of patients (in Years)
Infection risk	Average estimated probability of acquiring infection (in percent)
Routine culturing ratio	Ratio of number of cultures without signs of hospital-acquired infection, times 100
Routine chest X-ray ratio	Ratio of number of X-rays without signs of pneumonia, times 100
Number of beds	Average number of beds during study period
Medical school affiliation	1=Yes, 2=No
Region	Geographic region, where: 1=NE, 2=NC, s=S, 4=W
Average daily census	Average number of patients per day
Number of nurses	Average number of full-time equivalent registered and licensed practical nurses
Available facilities and services	Percent of 35 potential facilities and services that are provided by the hospital

(a) Read the variables given above carefully and make proper recoding for qualitative predictor variables.

(b) Determine the best regression model for predicting *Length of stay* (Y) from the other variables. Consider variable transformations, if necessary.

(c) Improve the 'good' model that you obtained from (b) through residual plots, normal probability plot and partial regression plots for testing model assumptions and detecting curvature and interaction effects. And explain how you settled on your final model.

(d) Do further model diagnostic checks for outliers, influential cases in the improved model.

(e) Interpret the values of the coefficient estimates in your model. Do they seem reasonable?

2. **STAT 514** *Design of Experiments*

An experiment was carried out to test the effect of two metals for pistons (2 s al and Teflon) on the firing time of explosives. Two variables that may affect firing time are the amount of primary initiator (5, 10 and 15 mg) and packing pressure (12K, 20K, 28K psi). Each combination of metal, initiator and pressure was used in random order. Then the whole experiment was repeated one more time in a different random order. See data in the following table.

a) Describe the statistical experimental design. Why was each combination of metal, initiator and pressure used in random order, within each of the two replicates?

b) Fit the simplest linear model that adequately explains the data in this experiment. Explain the terms in the model.

c) Study the residuals to verify the required assumptions. Are there any outliers? If so, reanalyze the data after replacing the outlier(s) by the fitted value(s). Is it necessary to transform the data in any way? If so, do it.

d) Test the hypothesis of no interaction between metal and initiator. Should you test for the main effects of metal and initiator? Why or why not?

e) For each level of initiator, test the difference between the mean firing times of the two metals and interpret the results.

f) What additional knowledge can be obtained if the experimenter's budget allowed a follow-up third replicate of this experiment?

Firing time			Rep 1						Rep 2			
Initiator	5		10		15		5		10		15	
Pressure \ metal	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
12000	61	57	62	67	42	73	56	56	56	64	58	73
20000	62	60	61	64	60	74	63	56	54	67	55	74
28000	66	59	60	71	59	72	57	63	59	65	58	73

3. **STAT 524** *Applied Multivariate Analysis*

A physical anthropologist performed a mineral of nine ancient Peruvian hairs. The results for the strontium (x_1) and chromium (x_2) levels, in parts per million (ppmm) were given in Table 1.

x_1	73.68	20.03	20.29	0.78	4.64	0.43	1.08	12.57	11.13
x_2	40.53	0.55	0.74	0.66	0.93	0.37	0.22	0.48	2.19

It is known that strontium is an indication of animal protein intake, while low levels (less than or equal to .100 ppmm) of chromium suggest the presence of diabetes.

(a) Construct and plot a 90 percent joint confidence ellipse for the population mean vector $\vec{\mu} = (\mu_1, \mu_2)^T$, assuming these nine Peruvian hairs represent a random sample from individuals belonging to a particular ancient Peruvian culture.

(b) Construct the 95 percent Bonferroni intervals for μ_1 and μ_2. Compare the two sets of intervals. What advantage, if any, do the T^2-intervals have over the Bonferroni intervals?

(c) Obtain the individual simultaneous 95 percent confidence intervals for μ_1 and μ_2. Does it appear as if this Peruvian culture has a mean strontium level of 10?

(d) Do these data appear to be bivariate normal? Discuss their status with reference to Q-Q plots and a scatter diagram. If the data are not bivariate normal, what implications does this have for the results in (a), (b) and (c)?

(e) Repeat the analysis with the obvious "outlying" observation removed. Do the inferences change? Comment.

Applied Statistics Comprehensive Examination

Schedule for the Oral Portion

Friday, January 16, 2009

Time	Name
9:30-10:15	
10:30-11:15	
11:30-12:15	
12:15-1:30	Lunch Break
1:30-2:15	
2:30-3:15	

/* Save this data as Senic.txt */

1	7.13	55.7	4.1	9.0	39.6	279	2	4	207	241	60.0
2	8.82	58.2	1.6	3.8	51.7	80	2	2	51	52	40.0
3	8.34	56.9	2.7	8.1	74.0	107	2	3	82	54	20.0
4	8.95	53.7	5.6	18.9	122.8	147	2	4	53	148	40.0
5	11.20	56.5	5.7	34.5	88.9	180	2	1	134	151	40.0
6	9.76	50.9	5.1	21.9	97.0	150	2	2	147	106	40.0
7	9.68	57.8	4.6	16.7	79.0	186	2	3	151	129	40.0
8	11.18	45.7	5.4	60.5	85.8	640	1	2	399	360	60.0
9	8.67	48.2	4.3	24.4	90.8	182	2	3	130	118	40.0
10	8.84	56.3	6.3	29.6	82.6	85	2	1	59	66	40.0
11	11.07	53.2	4.9	28.5	122.0	768	1	1	591	656	80.0
12	8.30	57.2	4.3	6.8	83.8	167	2	3	105	59	40.0
13	12.78	56.8	7.7	46.0	116.9	322	1	1	252	349	57.1
14	7.58	56.7	3.7	20.8	88.0	97	2	2	59	79	37.1
15	9.00	56.3	4.2	14.6	76.4	72	2	3	61	38	17.1
16	11.08	50.2	5.5	18.6	63.6	387	2	3	326	405	57.1
17	8.28	48.1	4.5	26.0	101.8	108	2	4	84	73	37.1
18	11.62	53.9	6.4	25.5	99.2	133	2	1	113	101	37.1
19	9.06	52.8	4.2	6.9	75.9	134	2	2	103	125	37.1
20	9.35	53.8	4.1	15.9	80.9	833	2	3	547	519	77.1
21	7.53	42.0	4.2	23.1	98.9	95	2	4	47	49	17.1
22	10.24	49.0	4.8	36.3	112.6	195	2	2	163	170	37.1
23	9.78	52.3	5.0	17.6	95.9	270	1	1	240	198	57.1
24	9.84	62.2	4.8	12.0	82.3	600	2	3	468	497	57.1
25	9.20	52.2	4.0	17.5	71.1	298	1	4	244	236	57.1
26	8.28	49.5	3.9	12.0	113.1	546	1	2	413	436	57.1
27	9.31	47.2	4.5	30.2	101.3	170	2	1	124	173	37.1
28	8.19	52.1	3.2	10.8	59.2	176	2	1	156	88	37.1
29	11.65	54.5	4.4	18.6	96.1	248	2	1	217	189	37.1
30	9.89	50.5	4.9	17.7	103.6	167	2	2	113	106	37.1
31	11.03	49.9	5.0	19.7	102.1	318	2	1	270	335	57.1
32	9.84	53.0	5.2	17.7	72.6	210	2	2	200	239	54.3
33	11.77	54.1	5.3	17.3	56.0	196	2	1	164	165	34.3
34	13.59	54.0	6.1	24.2	111.7	312	2	1	258	169	54.3
35	9.74	54.4	6.3	11.4	76.1	221	2	2	170	172	54.3
36	10.33	55.8	5.0	21.2	104.3	266	2	1	181	149	54.3
37	9.97	58.2	2.8	16.5	76.5	90	2	2	69	42	34.3
38	7.84	49.1	4.6	7.1	87.9	60	2	3	50	45	34.3
39	10.47	53.2	4.1	5.7	69.1	196	2	2	168	153	54.3
40	8.16	60.9	1.3	1.9	58.0	73	2	3	49	21	14.3
41	8.48	51.1	3.7	12.1	92.8	166	2	3	145	118	34.3
42	10.72	53.8	4.7	23.2	94.1	113	2	3	90	107	34.3
43	11.20	45.0	3.0	7.0	78.9	130	2	3	95	56	34.3
44	10.12	51.7	5.6	14.9	79.1	362	1	3	313	264	54.3
45	8.37	50.7	5.5	15.1	84.8	115	2	2	96	88	34.3
46	10.16	54.2	4.6	8.4	51.5	831	1	4	581	629	74.3
47	19.56	59.9	6.5	17.2	113.7	306	2	1	273	172	51.4
48	10.90	57.2	5.5	10.6	71.9	593	2	2	446	211	51.4
49	7.67	51.7	1.8	2.5	40.4	106	2	3	93	35	11.4
50	8.88	51.5	4.2	10.1	86.9	305	2	3	238	197	51.4
51	11.48	57.6	5.6	20.3	82.0	252	2	1	207	251	51.4
52	9.23	51.6	4.3	11.6	42.6	620	2	2	413	420	71.4

53	11.41	61.1	7.6	16.6	97.9	535	2	3	330	273	51.4
54	12.07	43.7	7.8	52.4	105.3	157	2	2	115	76	31.4
55	8.63	54.0	3.1	8.4	56.2	76	2	1	39	44	31.4
56	11.15	56.5	3.9	7.7	73.9	281	2	1	217	199	51.4
57	7.14	59.0	3.7	2.6	75.8	70	2	4	37	35	31.4
58	7.65	47.1	4.3	16.4	65.7	318	2	4	265	314	51.4
59	10.73	50.6	3.9	19.3	101.0	445	1	2	374	345	51.4
60	11.46	56.9	4.5	15.6	97.7	191	2	3	153	132	31.4
61	10.42	58.0	3.4	8.0	59.0	119	2	1	67	64	31.4
62	11.18	51.0	5.7	18.8	55.9	595	1	2	546	392	68.6
63	7.93	64.1	5.4	7.5	98.1	68	2	4	42	49	28.6
64	9.66	52.1	4.4	9.9	98.3	83	2	2	66	95	28.6
65	7.78	45.5	5.0	20.9	71.6	489	2	3	391	329	48.6
66	9.42	50.6	4.3	24.8	62.8	508	2	1	421	528	48.6
67	10.02	49.5	4.4	8.3	93.0	265	2	2	191	202	48.6
68	8.58	55.0	3.7	7.4	95.9	304	2	3	248	218	48.6
69	9.61	52.4	4.5	6.9	87.2	487	2	3	404	220	48.6
70	8.03	54.2	3.5	24.3	87.3	97	2	1	65	55	28.6
71	7.39	51.0	4.2	14.6	88.4	72	2	2	38	67	28.6
72	7.08	52.0	2.0	12.3	56.4	87	2	3	52	57	28.6
73	9.53	51.5	5.2	15.0	65.7	298	2	3	241	193	48.6
74	10.05	52.0	4.5	36.7	87.5	184	1	1	144	151	68.6
75	8.45	38.8	3.4	12.9	85.0	235	2	2	143	124	48.6
76	6.70	48.6	4.5	13.0	80.8	76	2	4	51	79	28.6
77	8.90	49.7	2.9	12.7	86.9	52	2	1	37	35	28.6
78	10.23	53.2	4.9	9.9	77.9	752	1	2	595	446	68.6
79	8.88	55.8	4.4	14.1	76.8	237	2	2	165	182	48.6
80	10.30	59.6	5.1	27.8	88.9	175	2	2	113	73	45.7
81	10.79	44.2	2.9	2.6	56.6	461	1	2	320	196	65.7
82	7.94	49.5	3.5	6.2	92.3	195	2	2	139	116	45.7
83	7.63	52.1	5.5	11.6	61.1	197	2	4	109	110	45.7
84	8.77	54.5	4.7	5.2	47.0	143	2	4	85	87	25.7
85	8.09	56.9	1.7	7.6	56.9	92	2	3	61	61	45.7
86	9.05	51.2	4.1	20.5	79.8	195	2	3	127	112	45.7
87	7.91	52.8	2.9	11.9	79.5	477	2	3	349	188	65.7
88	10.39	54.6	4.3	14.0	88.3	353	2	2	223	200	65.7
89	9.36	54.1	4.8	18.3	90.6	165	2	1	127	158	45.7
90	11.41	50.4	5.8	23.8	73.0	424	1	3	359	335	45.7
91	8.86	51.3	2.9	9.5	87.5	100	2	3	65	53	25.7
92	8.93	56.0	2.0	6.2	72.5	95	2	3	59	56	25.7
93	8.92	53.9	1.3	2.2	79.5	56	2	2	40	14	5.7
94	8.15	54.9	5.3	12.3	79.8	99	2	4	55	71	25.7
95	9.77	50.2	5.3	15.7	89.7	154	2	2	123	148	25.7
96	8.54	56.1	2.5	27.0	82.5	98	2	1	57	75	45.7
97	8.66	52.8	3.8	6.8	69.5	246	2	3	178	177	45.7
98	12.01	52.8	4.8	10.8	96.9	298	2	1	237	115	45.7
99	7.95	51.8	2.3	4.6	54.9	163	2	3	128	93	42.9
100	10.15	51.9	6.2	16.4	59.2	568	1	3	452	371	62.9
101	9.76	53.2	2.6	6.9	80.1	64	2	4	47	55	22.9
102	9.89	45.2	4.3	11.8	108.7	190	2	1	141	112	42.9
103	7.14	57.6	2.7	13.1	92.6	92	2	4	40	50	22.9
104	13.95	65.9	6.6	15.6	133.5	356	2	1	308	182	62.9
105	9.44	52.5	4.5	10.9	58.5	297	2	3	230	263	42.9
106	10.80	63.9	2.9	1.6	57.4	130	2	3	69	62	22.9

107	7.14	51.7	1.4	4.1	45.7	115	2	3	90	19	22.9
108	8.02	55.0	2.1	3.8	46.5	91	2	2	44	32	22.9
109	11.80	53.8	5.7	9.1	116.9	571	1	2	441	469	62.9
110	9.50	49.3	5.8	42.0	70.9	98	2	3	68	46	22.9
111	7.70	56.9	4.4	12.2	67.9	129	2	4	85	136	62.9
112	17.94	56.2	5.9	26.4	91.8	835	1	1	791	407	62.9
113	9.41	59.5	3.1	20.6	91.7	29	2	3	20	22	22.9

Applied Statistics Comprehensive Examination

In-Class Portion over STAT 519 and STAT 528

August 21, 2009, 1:00pm - 5:00pm

This exam covers material in STAT 519 *Probability Theory* and STAT 528 *Mathematical Statistics*. You are allowed one sheet of formulas and a calculator.

- Justify all of your answers and show all your work. An answer without justification is the same as no answer.
- Be as clear and complete as possible.
- Start each problem on a new page.
- Write on only one side of each page.
- Leave margins on 4 sides of each page for making copies.
- Write your name on each page.
- Do not staple your pages.
- Do as many problems as you can.

1. In a certain game a participant is allowed three attempts at scoring a hit. In the three attempts she must alternate which hand is used; thus she has two possible strategies: (right, left, right) or (left, right, left). The probabilities that she scores a hit with her right and left hands are p_r and p_l, respectively. Each attempt is independent of the others and she wins if she scores at least two hits in a row. If she is right handed (and $p_r > p_l$),

   (a) which strategy maximizes her chance of winning?

   (b) If four attempts are allowed, then which strategy is better (right, left, right, left) or (left, right, left, right)?

2. Let $X_1, X_2, \cdots, X_k$ be independent Poisson random variables with means $\mu_1, \mu_2, \cdots, \mu_k$ and let $S = X_1 + X_2 + \cdots + X_k$. Given $S = s$, what are the conditional distributions of

   (a) X_1,

   (b) $X_1 + X_2$, and

   (c) (X_1, X_2)?

3. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution having density $f(x; \theta) = e^{-(x-\theta)} I_{(\theta,\infty)}(x)$.

   (a) Find the probability density function of the minimum, $Y_{(1)}$.

   (b) Show that $Y_{(1)}$ converges in distribution and specify its limiting distribution.

   (c) Can you find a consistent estimator of θ? Explain.

4. Suppose $X_1, X_2, \cdots, X_n$ is a random sample from a normal distribution with mean zero and variance $\sigma^2 > 0$.

   (a) Determine the maximum likelihood estimator (MLE) of σ^2. Is it unbiased?

   (b) What is the MLE of σ? Is it unbiased?

   (c) Construct a $100(1 - \alpha)\%$ confidence interval for σ^2.

5. Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with pdf

$$f(x) = \begin{cases} \theta(1+x)^{-(1+\theta)} & \text{for } x > 0 \\ 0 & \text{elsewhere,} \end{cases}$$

   where $\theta > 0$.

   (a) Find a sufficient statistic for θ.

   (b) Find the Cramer-Rao lower bound for unbiased estimators of θ.

(c) Find the method of moments estimator of θ, assuming $\theta > 1$. Is the estimator valid when $0 < \theta \leq 1$?

(d) Find the maximum likelihood estimator of θ.

(e) Find the maximum likelihood estimator of $1/\theta$.

(f) Find a minimum variance unbiased estimator of $1/\theta$. Justify your answer.

6. Let $X_1, X_2, \cdots, X_n$ be a random sample from an exponential distribution with pdf

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere,} \end{cases}$$

where $\beta > 0$.

(a) Using the Central Limit Theorem, find an approximate 95% confience interval for β.

(b) Using an appropriate pivotal quantity, find an exact 95% confidence interval for β.

(c) Find the most powerful level-α test for $H_0 : \beta = \beta_0$ against $H_1 : \beta = \beta_1$ where $\beta_1 > \beta_0$.

(d) Is the above test uniformly most powerful for $H_0 : \beta = \beta_0$ against $H_1 : \beta > \beta_0$?

Applied Statistics Comprehensive Examination

Take-Home Portion over STAT 512, STAT 514 and STAT 524

August 21, 2009

Due: Tuesday, August 25, 2009, 4:00pm

This exam covers material in STAT 512 *Applied Regression Analysis*, STAT 514 *Design of Experiments* and STAT 524 *Applied Multivariate Analysis*. Read all the instructions carefully:

- Justify all of your answers and show all your work.
- Be as clear, complete, and succinct as possible.
- Start each problem on a new page.
- Write on only one side of each page.
- Clearly state the assumptions you are making whenever needed.
- If there are ambiguities, choose the most reasonable interpretation. However, be prepared to answer the question with alternative interpretations as well.
- Include only the most essential and enlightening computer output and no code.
- All datasets can be found at http://www.math.iupui.edu/~fli/Comp/Data.htm.
- Do not discuss any aspect of this exam with anyone else until after the completion of all oral examinations.
- Place one copy of your written report in each of the mailboxes of Ben Boukai, Samiran Ghosh, Fang Li, Hanxiang Peng, Ryan Martin and Jyoti Sarkar before 4:00pm on Tuesday, August 25, 2009.

1. **STAT 512** *Applied Regression Problem*

The data file *http://www.math.iupui.edu/∼fli/Comp/DeathRate.txt* contains a random sample of records of death rates. We wish to come up with some regression model to represent death rate as a function of other variables included in the data set.

Variable in order	Description
Index	1-60
A1	the average annual precipitation
A2	the average January temperature
A3	the average July temperature
A4	the size of the population older than 65
A5	the number of members per household
A6	the number of years of schooling for persons over 22
A7	the number of households with fully equipped kitchens
A8	the population per square mile
A9	the size of the nonwhite population
A10	the number of office workers
A11	the number of families with an income less than $3000
A12	the hydrocarbon pollution index
A13	the nitric oxide pollution index
A14	the sulfur dioxide pollution index
A15	the degree of atmospheric moisture
B	the death rate

(a) Determine the best regression model for explaining the death rate from the other variables using different model-building criteria and automatic selection procedures. Consider variable transformations, if necessary.

(b) Improve the 'good' model that you obtained from (a) through residual plots, normal probability plot and partial regression plots for testing model assumptions and detecting curvature and interaction effects. And explain how you settled on your final model.

(c) Do further model diagnostic checks for outliers, influential cases and multicollinearity in the improved model.

(d) Interpret the values of the coefficient estimates in your model. Do they seem reasonable?

2. **STAT 514** *Design of Experiments* Vacuum tubes are used in the audio equipment. To determine the effect of exhaust index (in seconds) and pump heater voltage (in volts) on the pressure inside a vacuum tube (in micrometers of mercury), three exhaust indexes and two voltages were chosen at fixed levels. It was decided to run two experiments at each experimental condition. In this experiment complete randomization was employed. Given below is the data for this experiment,

Pump Heater	Exhaust		Index
Voltage	60	90	120
127	48	28	7
	58	33	15
220	62	14	9
	54	10	6

(a) Identify the design and write down the model for this experiment along with assumptions.

(b) Write down the possible hypothesis you want to test.

(c) Complete the ANOVA table and state your conclusion in plain English

(d) Suppose now on the basis your findings at part (c) you are allowed to run one more experiment at a new exhaust index level of your choice to further improve your result. What possible exhaust level you should go for in that experiment and why?

3. **STAT 524** *Applied Multivariate Analysis*

A physical anthropologist performed a mineral of nine ancient Peruvian hairs. The results for the strontium (x_1) and chromium (x_2) levels, in parts per million (ppmm) were given in Table 1.

Table 1:

x_1	73.68	20.03	20.29	.78	4.64	.43	1.08	12.57	11.13
x_2	40.53	0.55	0.74	.66	0.93	.37	0.22	0.48	2.19

It is known that strontium is an indication of animal protein intake, while low levels (less than or equal to .100 ppmm) of chromium suggest the presence of diabetes.

(a) Construct and plot a 90 percent joint confidence ellipse for the population mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$, assuming these nine Peruvian hairs represent a random sample from individuals belonging to a particular ancient Peruvian culture.

(b) Construct the 95 percent Bonferroni intervals for μ_1 and μ_2. Compare the two sets of intervals. What advantage, if any, do the T^2-intervals have over the Bonferroni intervals?

(c) Obtain the individual simultaneous 95 percent confidence intervals for μ_1 and μ_2. Does it appear as if this Peruvian culture has a mean strontium level of 10?

(d) Do these data appear to be bivariate normal? Discuss their status with reference to Q-Q plots and a scatter diagram. If the data are not bivariate normal, what implications does this have for the results in (1), (2) and (3)?

(e) Repeat the analysis with the obvious "outlying" observation removed. Do the inferences change? Comment.

Applied Statistics Comprehensive Examination

Schedule for the Oral Portion

Friday, August 28, 2009

Time	Name
10:30-11:15	
11:30-12:15	

```
/* Save this data as DeathRate.txt */

 1  36  27  71   8.1  3.34  11.4  81.5  3243   8.8  42.6  11.7   21   15
59  59   921.8
 2  35  23  72  11.1  3.14  11.0  78.8  4281   3.6  50.7  14.4    8   10
39  57   997.8
 3  44  29  74  10.4  3.21   9.8  81.6  4260   0.8  39.4  12.4    6    6
33  54   962.3
 4  47  45  79   6.5  3.41  11.1  77.5  3125  27.1  50.2  20.6   18    8
24  56   982.2
 5  43  35  77   7.6  3.44   9.6  84.6  6441  24.4  43.7  14.3   43   38
206  55  1071.2
 6  53  45  80   7.7  3.45  10.2  66.8  3325  38.5  43.1  25.5   30   32
72  54  1030.3
 7  43  30  74  10.9  3.23  12.1  83.9  4679   3.5  49.2  11.3   21   32
62  56   934.7
 8  45  30  73   9.3  3.29  10.6  86.0  2140   5.3  40.4  10.5    6    4
4  56   899.5
 9  36  24  70   9.0  3.31  10.5  83.2  6582   8.1  42.5  12.6   18   12
37  61  1001.9
10  36  27  72   9.5  3.36  10.7  79.3  4213   6.7  41.0  13.2   12    7
20  59   912.3
11  52  42  79   7.7  3.39   9.6  69.2  2302  22.2  41.3  24.2   18    8
27  56  1017.6
12  33  26  76   8.6  3.20  10.9  83.4  6122  16.3  44.9  10.7   88   63
278  58  1024.8
13  40  34  77   9.2  3.21  10.2  77.0  4101  13.0  45.7  15.1   26   26
146  57   970.4
14  35  28  71   8.8  3.29  11.1  86.3  3042  14.7  44.6  11.4   31   21
64  60   985.9
15  37  31  75   8.0  3.26  11.9  78.4  4259  13.1  49.6  13.9   23    9
15  58   958.8
16  35  46  85   7.1  3.22  11.8  79.9  1441  14.8  51.2  16.1    1    1
1  54   860.1
17  36  30  75   7.5  3.35  11.4  81.9  4029  12.4  44.0  12.0    6    4
16  58   936.2
18  15  30  73   8.2  3.15  12.2  84.2  4824   4.7  53.1  12.7   17    8
28  38   871.7
19  31  27  74   7.2  3.44  10.8  87.0  4834  15.8  43.5  13.6   52   35
124  59   959.2
20  30  24  72   6.5  3.53  10.8  79.5  3694  13.1  33.8  12.4   11    4
11  61   941.1
21  31  45  85   7.3  3.22  11.4  80.7  1844  11.5  48.1  18.5    1    1
1  53   891.7
22  31  24  72   9.0  3.37  10.9  82.8  3226   5.1  45.2  12.3    5    3
10  61   871.3
23  42  40  77   6.1  3.45  10.4  71.8  2269  22.7  41.4  19.5    8    3
5  53   971.1
24  43  27  72   9.0  3.25  11.5  87.1  2909   7.2  51.6   9.5    7    3
10  56   887.4
25  46  55  84   5.6  3.35  11.4  79.7  2647  21.0  46.9  17.9    6    5
1  59   952.5
26  39  29  76   8.7  3.23  11.4  78.6  4412  15.6  46.6  13.2   13    7
33  60   968.6
```

27	35	31	81	9.2	3.10	12.0	78.3	3262	12.6	48.6	13.9	7	4	4	55	919.7
28	43	32	74	10.1	3.38	9.5	79.2	3214	2.9	43.7	12.0	11	7	32	54	844.0
29	11	53	68	9.2	2.99	12.1	90.6	4700	7.8	48.9	12.3	648	319	130	47	861.8
30	30	35	71	8.3	3.37	9.9	77.4	4474	13.1	42.6	17.7	38	37	193	57	989.2
31	50	42	82	7.3	3.49	10.4	72.5	3497	36.7	43.3	26.4	15	10	34	59	1006.4
32	60	67	82	10.0	2.98	11.5	88.6	4657	13.6	47.3	22.4	3	1	1	60	861.4
33	30	20	69	8.8	3.26	11.1	85.4	2934	5.8	44.0	9.4	33	23	125	64	929.1
34	25	12	73	9.2	3.28	12.1	83.1	2095	2.0	51.9	9.8	20	11	26	50	857.6
35	45	40	80	8.3	3.32	10.1	70.3	2682	21.0	46.1	24.1	17	14	78	56	961.0
36	46	30	72	10.2	3.16	11.3	83.2	3327	8.8	45.3	12.2	4	3	8	58	923.2
37	54	54	81	7.4	3.36	9.7	72.8	3172	31.4	45.5	24.2	20	17	1	62	1113.1
38	42	33	77	9.7	3.03	10.7	83.5	7462	11.3	48.7	12.4	41	26	108	58	994.6
39	42	32	76	9.1	3.32	10.5	87.5	6092	17.5	45.3	13.2	29	32	161	54	1015.0
40	36	29	72	9.5	3.32	10.6	77.6	3437	8.1	45.5	13.8	45	59	263	56	991.2
41	37	38	67	11.3	2.99	12.0	81.5	3387	3.6	50.3	13.5	56	21	44	73	893.9
42	42	29	72	10.7	3.19	10.1	79.5	3508	2.2	38.3	15.7	6	4	18	56	938.5
43	41	33	77	11.2	3.08	9.6	79.9	4843	2.7	38.6	14.1	11	11	89	54	946.1
44	44	39	78	8.2	3.32	11.0	79.9	3768	28.6	49.5	17.5	12	9	48	53	1025.5
45	32	25	72	10.9	3.21	11.1	82.5	4355	5.0	46.4	10.8	7	4	18	60	874.2
46	34	32	79	9.3	3.23	9.7	76.8	5160	17.2	45.1	15.3	31	15	68	57	953.5
47	10	55	70	7.3	3.11	12.1	88.9	3033	5.9	51.0	14.0	144	66	20	61	839.7
48	18	48	63	9.2	2.92	12.2	87.7	4253	13.7	51.2	12.0	311	171	86	71	911.7
49	13	49	68	7.0	3.36	12.2	90.7	2702	3.0	51.9	9.7	105	32	3	71	790.7
50	35	40	64	9.6	3.02	12.2	82.5	3626	5.7	54.3	10.1	20	7	20	72	899.2
51	45	28	74	10.6	3.21	11.1	82.6	1883	3.4	41.9	12.3	5	4	20	56	904.1
52	38	24	72	9.8	3.34	11.4	78.0	4923	3.8	50.5	11.1	8	5	25	61	950.6
53	31	26	73	9.3	3.22	10.7	81.3	3249	9.5	43.9	13.6	11	7	25	59	972.4

54	40	23	71	11.3	3.28	10.3	73.8	1671	2.5	47.4	13.5	5	2	11	60	912.2
55	41	37	78	6.2	3.25	12.3	89.5	5308	25.9	59.7	10.3	65	28	102	52	967.8
56	28	32	81	7.0	3.27	12.1	81.0	3665	7.5	51.6	13.2	4	2	1	54	823.7
57	45	33	76	7.7	3.39	11.3	82.2	3152	12.1	47.3	10.9	14	11	42	56	1003.5
58	45	24	70	11.8	3.25	11.1	79.8	3678	1.0	44.8	14.0	7	3	8	56	895.6
59	42	83	76	9.7	3.22	9.0	76.2	9699	4.8	42.2	14.5	8	8	49	54	911.8
60	38	28	72	8.9	3.48	10.7	79.8	3451	11.7	37.5	13.0	14	13	39	58	954.4

Applied Statistics Comprehensive Examination

In-Class Portion over STAT 519 and STAT 528

August 15, 2008, 1:00pm - 5:00pm

This exam covers material in STAT 519 *Probability Theory* and STAT 528 *Mathematical Statistics*. You are allowed one sheet of formulas and a calculator.

- Justify all of your answers and show all your work. An answer without justification is the same as no answer.
- Be as clear and complete as possible.
- Start each problem on a new page.
- Write on only one side of each page.
- Leave margins on 4 sides of each page for making copies.
- Write your name on each page.
- Do not staple your pages.
- Do as many problems as you can.

1. A professor has three identical copies of a calculus textbook on her desk. Each class day she picks one of the copies at random to take to class. Let N be the first class day each copy has been picked up at lease once.

   (a) Find the probability mass function of N.

   (b) Find the mean and the variance of N.

   (c) What is the probability that copy C is picked up on day N?

   (d) Given that copy C is picked up on day N, what is the probability that copies A and B were picked up equally often (Prior to day N)?

2. Let $X_1, \cdots$ be i.i.d. $\mathcal{N}(0, \theta)$, where $\theta > 0$. For fixed sample size n

   (a) find an MVUE of θ^2.

   (b) find a UMP for $H_0 : \theta = 1$ v.s. $H_1 : \theta > 1$, if it exists, or prove that it does not exist.

   (c) develop a sequential probability ratio test for $H_0 : \theta = 4$ v.s. $H_1 : \theta = 9$ that attains $\alpha = 0.05$ and $\beta = 0.10$.

3. Let $Y_1, Y_2, \cdots, Y_n$ denote a random sample from the probability density function

$$f(y \mid \theta) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1; \quad \theta > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

   (a) Show that this density function is in the exponential family and that $\sum_{i=1}^{n} -ln(Y_i)$ is sufficient for θ.

   (b) If $W_i = -ln(Y_i)$, show that W_i has an exponential distribution with mean $1/\theta$.

   (c) Show that $2\theta \sum_{i=1}^{n} W_i$ has a χ^2 distribution with $2n$ degrees of freedom.

   (d) Show that

$$E\left(\frac{1}{2\theta \sum_{i=1}^{n} W_i}\right) = \frac{1}{2(n-1)}.$$

   (e) What is the MVUE (Minimum variance unbiased estimator) for θ?

4. Let the random variables Y_1 and Y_2 are independent exponentially distributed random variables, both with mean β.

   (a) Compute $E[Y_1 \,|\, Y_1 < Y_2]$.

   (b) Compute $E[Y_1 \,|\, Y_1 > Y_2]$.

   (c) Compute $E[Y_1 \wedge Y_2]$

   (d) Compute $E[Y_1 \vee Y_2]$

5. (a) Derive the mgf (moment generating function) of a standard normal random variable.

   (b) Let X be a random variable with mgf $M_X(t)$ and let a and b be constants. Show that the mgf of $aX + b$ is $e^{tb}M_X(at)$. Use this result to write down the mgf of a $N(\mu, \sigma^2)$ random variable.

   (c) Suppose that $W_1, \cdots, W_m$ are independent random variables such that $W_i \sim N(\mu_i, \sigma^2)$. Find the distribution of $\sum_{i=1}^{m} a_i W_i$ where $a_1, \cdots, a_m$ are known constants.

   (d) Suppose that the random variables $Y_1, \cdots, Y_n$ satisfy

$$Y_i = \beta x_i + \varepsilon_i, \qquad i = 1, \cdots, n,$$

   where $x_1, \cdots, x_n$ are known constants and $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. $N(0, \tau^2)$. Find a two dimensional sufficient statistic for the unknown parameter (β, τ^2).

   (e) Find the MLE of β, call it $\hat{\beta} = \hat{\beta}(Y_1, \cdots, Y_n)$. Show that $\hat{\beta}$ is unbiased.

   (f) Find the distribution of $\hat{\beta}$.

   (g) Find the distribution of the alternative estimator of β given by

$$\tilde{\beta} = \tilde{\beta}(Y_1, \cdots, Y_n) = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}.$$

   (h) Which of these estimators is better?

Applied Statistics Comprehensive Examination

Take-Home Portion over STAT 512, STAT 514 and STAT 524

August 15, 2008

Due: Tuesday, August 19, 2008, 4:00pm

This exam covers material in STAT 512 *Applied Regression Analysis*, STAT 514 *Design of Experiments* and STAT 524 *Applied Multivariate Analysis*. Read all the instructions carefully:

- Justify all of your answers and show all your work.
- Be as clear, complete, and succinct as possible.
- Start each problem on a new page.
- Write on only one side of each page.
- Clearly state the assumptions you are making whenever needed.
- If there are ambiguities, choose the most reasonable interpretation. However, be prepared to answer the question with alternative interpretations as well.
- Include only the most essential and enlightening computer output and no code.
- All datasets can be found at http://www.math.iupui.edu/∼fli/Comp/.
- Do not discuss any aspect of this exam with anyone else until after the completion of all oral examinations.
- Place one copy of your written report in each of the mailboxes of Ben Boukai, Samiran Ghosh, Fang Li, Hanxiang Peng and Jyoti Sarkar before 4:00pm on Tuesday, August 19, 2008.

1. **STAT 512** *Applied Regression Problem*

   The data file `diamond.txt` contains data on diamond sales. We wish to come up with some sensible pricing model for diamond stones based on several variables included in the data set.

Variable in order	Description
Carat	Weight of diamond stones in carat units
Size	Low($<$ .05 carat), Median ($\geq$ .05 and $<$ 1) or Large ($\geq$ 1)
Colour	D, E, F, G, H or I
Clarity	IF, VVS1, VVS2, VS1 or VS2
Certification Body	GIA, IGI or HRD
Price	Price of the diamond stone in Singapore $

   (a) Read the variables given above carefully and make proper recoding for qualitative predictor variables.

   (b) Determine the best regression model for explaining price of the diamond stone (Price) from the other variables. Consider variable transformations, if necessary.

   (c) Improve the 'good' model that you obtained from (c) through residual plots, normal probability plot and partial regression plots for testing model assumptions and detecting curvature and interaction effects. And explain how you settled on your final model.

   (d) Do further model diagnostic checks for outliers, influential cases in the improved model.

   (e) Interpret the values of the coefficient estimates in your model. Do they seem reasonable? Give a 95% prediction interval for the price of a diamond with .56 carat, colour E, clarity VVS1 and certification body GIA.

2. **STAT 514** *Design of Experiments*

In a study of the effect of four vitamin supplements on weight gain in a laboratory, five separately caged animals were used for each diet group. For each animal, the caloric intake (c) and weight gain (W) were recorded. See data in the following table. Answer each of the following questions by proposing a suitable model, formulating the hypotheses to be tested and carrying out the test. Also, for each test, verify the model assumptions and comment on their validity.

(a) Does diet influence caloric intake?

(b) Does caloric intake affect weight gain?

(c) Does diet affect weight gain other than through caloric intake?

(d) Is the expected weight gain per unit increase in caloric intake the same for all diet groups?

(e) Suppose that the fifth animal on diet group 4 was a sickly animal and hence its weight gain (W=34) was atypically low. (It is probably a typo). Drop this datum altogether. Then fit a suitable model to the rest of the data and construct a 95% predictive interval for the weight gain of an animal on diet four with a calorie intake of 43.

(f) Suppose that in a follow-up experiment, five litters of animals, each having two males and two females, are available. How should the experimenter assign the animals to the diet groups?

╲ Diet	1		2		3		4	
Animals	W	c	W	c	W	c	W	c
1	48	35	65	40	79	51	59	53
2	67	44	49	45	52	41	50	52
3	78	44	37	37	63	47	59	52
4	69	51	73	53	65	47	42	51
5	53	47	63	42	67	48	34	43

3. **STAT 524** *Applied Multivariate Analysis*

The cranial capacity C of skulls is related to glabellar-occipital length L, the maximum parietal breadth B and the basio-bregmatic height H according to the following formula $C \propto L^{\beta_1} B^{\beta_2} H^{\beta_3}$. Based on these variables measured on several well-preserved specimens, it is desired to predict the cranial capacity of some broken or damaged skulls for which L, B and H can still be measured. Some additional variables are also measured: Flower's ophryo-occipital length LO, glabeller-projective length LP, least forehead breadth BF, basion Frankfort height HF, craniophor auricular height HA, basion-nasion length LN. See the data file `skulldata.txt`.

(a) What transformation on each variable would you need in order to write a linear regression model?

(b) Obtain the desired prediction equation. Give a 90% confidence interval for the cranial capacity of a skull which has $L = 190.5$, $B = 148.0$ and $H = 136.5$.

(c) If one wishes to use up to three "best" linear combinations of all covariates, instead of the default variables L, B and H, in order to predict cranial capacity C, how much improvement can be made in the prediction? Interpret the selected linear combinations.

(d) How serious is the problem of multicollinearity in this problem?

(e) Are the model assumptions satisfied? Are there any outliers?

Applied Statistics Comprehensive Examination

Schedule for the Oral Portion

Friday, August 22, 2008

Time	Name
9:00-9:45	
10:00-10:45	
11:00-11:45	
12:00–12:45	

```
/* Save this data at diamond.txt */

Carat Size  Colour      Clarity     CB    Price
0.3   Low   D     VS2   GIA   1302
0.3   Low   E     VS1   GIA   1510
0.3   Low   G     VVS1  GIA   1510
0.3   Low   G     VS1   GIA   1260
0.31  Low   D     VS1   GIA   1641
0.31  Low   E     VS1   GIA   1555
0.31  Low   F     VS1   GIA   1427
0.31  Low   G     VVS2  GIA   1427
0.31  Low   H     VS2   GIA   1126
0.31  Low   I     VS1   GIA   1126
0.32  Low   F     VS1   GIA   1468
0.32  Low   G     VS2   GIA   1202
0.33  Low   E     VS2   GIA   1327
0.33  Low   I     VS2   GIA   1098
0.34  Low   E     VS1   GIA   1693
0.34  Low   F     VS1   GIA   1551
0.34  Low   G     VS1   GIA   1410
0.34  Low   G     VS2   GIA   1269
0.34  Low   H     VS1   GIA   1316
0.34  Low   H     VS2   GIA   1222
0.35  Low   E     VS1   GIA   1738
0.35  Low   F     VS1   GIA   1593
0.35  Low   G     VS1   GIA   1447
0.35  Low   H     VS2   GIA   1255
0.36  Low   F     VS1   GIA   1635
0.36  Low   H     VVS2  GIA   1485
0.37  Low   F     VS2   GIA   1420
0.37  Low   H     VS1   GIA   1420
0.4   Low   F     VS1   GIA   1911
0.4   Low   H     VS1   GIA   1525
0.41  Low   F     VS1   GIA   1956
0.43  Low   H     VVS2  GIA   1747
0.45  Low   I     VS1   GIA   1572
0.46  Low   E     VVS2  GIA   2942
0.48  Low   G     VVS2  GIA   2532
0.5   Median      E     VS1   GIA   3501
0.5   Median      E     VS1   GIA   3501
0.5   Median      F     VVS2  GIA   3501
0.5   Median      F     VS1   GIA   3293
0.5   Median      G     VS1   GIA   3016
0.51  Median      F     VVS2  GIA   3567
0.51  Median      G     VS1   GIA   3205
0.52  Median      D     VS2   GIA   3490
0.52  Median      E     VS1   GIA   3635
0.52  Median      F     VVS2  GIA   3635
0.52  Median      F     VS1   GIA   3418
0.53  Median      D     VS1   GIA   3921
0.53  Median      F     VVS2  GIA   3701
0.53  Median      F     VS1   GIA   3480
0.53  Median      G     VVS2  GIA   3407
0.54  Median      E     VS1   GIA   3767
```

0.54	Median	F	VVS1	GIA	4066
0.55	Median	E	VVS2	GIA	4138
0.55	Median	F	VS1	GIA	3605
0.55	Median	G	VVS2	GIA	3529
0.56	Median	F	VS1	GIA	3667
0.56	Median	I	VVS2	GIA	2892
0.57	Median	G	VVS2	GIA	3651
0.59	Median	G	VVS2	GIA	3773
0.6	Median	F	VS1	GIA	4291
0.62	Median	E	VVS1	GIA	5845
0.63	Median	G	VVS2	GIA	4401
0.64	Median	G	VVS1	GIA	4759
0.66	Median	H	VVS1	GIA	4300
0.7	Median	F	VS1	GIA	5510
0.7	Median	G	VS1	GIA	5122
0.7	Median	H	VVS2	GIA	5122
0.7	Median	I	VS2	GIA	3861
0.71	Median	F	VVS2	GIA	5881
0.71	Median	F	VS1	GIA	5586
0.71	Median	F	VS2	GIA	5193
0.71	Median	H	VVS2	GIA	5193
0.72	Median	F	VS2	GIA	5263
0.8	Median	I	VVS2	GIA	5441
0.82	Median	I	VS2	GIA	4948
0.84	Median	H	VS2	GIA	5705
0.85	Median	F	VS2	GIA	6805
0.86	Median	H	VVS2	GIA	6882
0.89	Median	H	VS1	GIA	6709
0.9	Median	I	VVS2	GIA	6682
0.5	Median	E	VS1	GIA	3501
0.5	Median	G	VVS1	GIA	3432
0.51	Median	F	VVS1	GIA	3851
0.55	Median	H	IF	GIA	3605
0.56	Median	E	VS1	GIA	3900
0.57	Median	H	VVS1	GIA	3415
0.6	Median	H	IF	GIA	4291
0.63	Median	E	IF	GIA	6512
0.7	Median	E	VS1	GIA	5800
0.7	Median	F	VVS1	GIA	6285
0.7	Median	F	VS2	GIA	5122
0.7	Median	F	VS2	GIA	5122
0.7	Median	G	VS1	GIA	5122
0.7	Median	H	VVS2	GIA	5122
0.71	Median	D	VS1	GIA	6372
0.71	Median	E	VS1	GIA	5881
0.71	Median	H	VVS2	GIA	5193
0.72	Median	E	VS1	GIA	5961
0.72	Median	H	VVS1	GIA	5662
0.73	Median	E	VS2	GIA	5738
0.73	Median	H	VS1	GIA	5030
0.73	Median	H	VS1	GIA	5030
0.73	Median	I	VVS1	GIA	4727
0.73	Median	I	VS1	GIA	4221
0.74	Median	G	VVS2	GIA	5815

0.74	Median	H	VS2	GIA	4585
0.75	Median	D	VVS2	GIA	7368
0.75	Median	I	VVS2	GIA	4667
0.75	Median	I	VS1	GIA	4355
0.76	Median	D	IF	GIA	9885
0.77	Median	F	VVS1	GIA	6919
0.78	Median	H	VS1	GIA	5386
0.8	Median	I	VS2	GIA	4832
0.83	Median	E	VS2	GIA	7156
0.9	Median	F	VS1	GIA	7680
1	Large	D	VVS1	GIA	15582
1	Large	D	VS1	GIA	11419
1	Large	E	VS1	GIA	10588
1	Large	E	VS2	GIA	9757
1	Large	F	IF	GIA	13913
1	Large	F	VVS2	GIA	10588
1	Large	F	VS1	GIA	10713
1	Large	F	VS2	GIA	9480
1	Large	G	VVS2	GIA	9896
1	Large	G	VS1	GIA	9619
1	Large	G	VS2	GIA	9169
1	Large	G	VS2	GIA	9203
1	Large	H	VS2	GIA	8788
1	Large	I	VS1	GIA	8095
1	Large	I	VS2	GIA	7818
1.01	Large	D	VVS1	GIA	16008
1.01	Large	E	VS1	GIA	10692
1.01	Large	E	VS2	GIA	9853
1.01	Large	F	VS1	GIA	10272
1.01	Large	F	VS2	GIA	9573
1.01	Large	H	VS1	GIA	9153
1.01	Large	H	VS2	GIA	8873
1.01	Large	I	VVS1	GIA	8873
1.01	Large	I	VVS2	GIA	8455
1.01	Large	I	VS2	GIA	7895
1.02	Large	F	VS1	GIA	10372
1.02	Large	F	VS2	GIA	9666
1.02	Large	G	VVS2	GIA	10090
1.03	Large	E	VS1	GIA	10900
1.04	Large	F	VS1	GIA	10571
1.04	Large	I	IF	GIA	9563
1.05	Large	I	VVS2	GIA	8781
1.06	Large	G	VS2	GIA	9743
1.06	Large	H	VS2	GIA	9302
1.07	Large	I	VVS2	GIA	8945
1.1	Large	H	VS2	GIA	9646
0.18	Low	F	VVS1	IGI	823
0.18	Low	F	VVS2	IGI	765
0.18	Low	G	IF	IGI	803
0.18	Low	G	IF	IGI	803
0.18	Low	G	VVS2	IGI	705
0.18	Low	H	IF	IGI	725
0.19	Low	D	VVS2	IGI	967
0.19	Low	E	IF	IGI	1050

0.19	Low	F	IF	IGI	967	
0.19	Low	F	VVS1	IGI	863	
0.19	Low	F	VVS2	IGI	800	
0.19	Low	G	IF	IGI	842	
0.19	Low	G	VVS1	IGI	800	
0.19	Low	H	IF	IGI	758	
0.2	Low	D	VS1	IGI	880	
0.2	Low	G	IF	IGI	880	
0.2	Low	G	VS1	IGI	705	
0.2	Low	G	VS2	IGI	638	
0.21	Low	D	VS1	IGI	919	
0.21	Low	E	IF	IGI	1149	
0.21	Low	F	IF	IGI	1057	
0.21	Low	G	IF	IGI	919	
0.22	Low	E	IF	IGI	1198	
0.23	Low	E	IF	IGI	1248	
0.23	Low	F	IF	IGI	1147	
0.23	Low	G	IF	IGI	995	
0.24	Low	H	IF	IGI	1108	
0.25	Low	F	IF	IGI	1485	
0.25	Low	G	IF	IGI	1283	
0.25	Low	H	IF	IGI	1149	
0.25	Low	I	IF	IGI	1082	
0.26	Low	F	IF	IGI	1539	
0.26	Low	F	VVS1	IGI	1365	
0.26	Low	F	VVS2	IGI	1260	
0.26	Low	I	IF	IGI	1121	
0.27	Low	F	IF	IGI	1595	
0.27	Low	H	IF	IGI	1233	
0.28	Low	I	IF	IGI	1199	
0.29	Low	G	IF	IGI	1471	
0.29	Low	I	IF	IGI	1238	
0.3	Low	E	VVS2	IGI	1580	
0.3	Low	F	VVS2	IGI	1459	
0.3	Low	G	VVS1	IGI	1459	
0.3	Low	H	VVS2	IGI	1218	
0.3	Low	I	IF	IGI	1299	
0.31	Low	E	VVS2	IGI	1628	
0.31	Low	F	VVS1	IGI	1628	
0.31	Low	I	IF	IGI	1337	
0.32	Low	H	IF	IGI	1462	
0.33	Low	H	IF	IGI	1503	
0.34	Low	F	VVS1	IGI	1773	
0.34	Low	F	VVS2	IGI	1636	
0.35	Low	F	VVS1	IGI	1821	
0.35	Low	G	VVS2	IGI	1540	
0.4	Low	G	IF	IGI	2276	
0.41	Low	I	VVS1	IGI	1616	
0.41	Low	I	VVS2	IGI	1506	
0.47	Low	F	VVS2	IGI	2651	
0.48	Low	F	VS1	IGI	2383	
0.5	Median		G	IF	IGI	3652
0.51	Median		E	VVS2	IGI	3722
0.51	Median		F	VVS1	IGI	3722

0.52	Median	I	IF	IGI	3095
0.55	Median	F	VVS2	IGI	3706
0.56	Median	E	VVS2	IGI	4070
0.56	Median	G	VVS2	IGI	3470
0.58	Median	E	VVS1	IGI	4831
0.58	Median	F	VVS1	IGI	4209
0.58	Median	G	VVS1	IGI	3821
0.7	Median	G	VVS1	IGI	5607
0.7	Median	G	VVS2	IGI	5326
0.71	Median	D	VS1	IGI	6160
0.76	Median	F	VVS2	IGI	6095
0.78	Median	G	VVS2	IGI	5937
1	Large H	VVS2	IGI	9342	
1.01	Large G	VS1	IGI	9713	
1.01	Large H	VS2	IGI	8873	
1.01	Large I	VS1	IGI	8175	
0.5	Median	F	VVS1	HRD	3778
0.5	Median	G	VVS1	HRD	3432
0.51	Median	F	VVS1	HRD	3851
0.52	Median	E	VS2	HRD	3346
0.52	Median	H	VVS1	HRD	3130
0.53	Median	F	VVS1	HRD	3995
0.53	Median	F	VVS2	HRD	3701
0.55	Median	G	VVS2	HRD	3529
0.56	Median	F	VS1	HRD	3667
0.56	Median	F	VS2	HRD	3202
0.57	Median	F	VS2	HRD	3256
0.57	Median	H	VVS1	HRD	3415
0.58	Median	H	IF	HRD	3792
0.6	Median	G	VS1	HRD	3925
0.6	Median	G	VS2	HRD	3421
0.6	Median	H	VVS1	HRD	3925
0.61	Median	H	VVS2	HRD	3616
0.62	Median	I	VVS2	HRD	3615
0.64	Median	H	VVS2	HRD	3785
0.65	Median	I	VVS2	HRD	3643
0.66	Median	H	VVS1	HRD	4300
0.7	Median	E	VVS1	HRD	6867
0.7	Median	E	VVS2	HRD	6285
0.7	Median	G	VVS1	HRD	5800
0.7	Median	G	VVS2	HRD	5510
0.7	Median	H	VS2	HRD	4346
0.71	Median	G	IF	HRD	6372
0.71	Median	H	VVS2	HRD	5193
0.72	Median	H	VVS1	HRD	5662
0.73	Median	F	VS2	HRD	5333
0.73	Median	G	VVS1	HRD	6041
0.74	Median	H	VVS1	HRD	5815
0.8	Median	F	IF	HRD	8611
0.8	Median	F	VS1	HRD	6905
0.8	Median	G	VVS2	HRD	6905
0.8	Median	H	VVS2	HRD	6416
0.8	Median	H	VS1	HRD	6051
0.81	Median	E	VVS1	HRD	8715

0.81	Median	E	VS2	HRD	6988
0.81	Median	F	VS1	HRD	6988
0.81	Median	G	VS1	HRD	6495
0.81	Median	H	IF	HRD	7358
0.82	Median	F	VS2	HRD	6572
0.82	Median	G	VVS2	HRD	7072
0.85	Median	F	VVS1	HRD	8359
0.85	Median	F	VS2	HRD	6805
0.85	Median	G	VVS1	HRD	7711
0.86	Median	H	VS2	HRD	5835
1	Large	D	VVS2	HRD	13775
1	Large	E	VVS1	HRD	14051
1	Large	E	VVS2	HRD	11419
1	Large	E	VS1	HRD	10588
1	Large	F	VVS1	HRD	11696
1	Large	F	VVS2	HRD	10588
1	Large	G	VVS1	HRD	10450
1	Large	G	VVS2	HRD	9896
1	Large	G	VS2	HRD	9203
1	Large	H	VVS1	HRD	9480
1	Large	H	VS1	HRD	9065
1	Large	H	VS2	HRD	8788
1	Large	I	VVS1	HRD	8788
1	Large	I	VVS2	HRD	8372
1	Large	I	VS1	HRD	8095
1	Large	I	VS2	HRD	7818
1.01	Large	D	VVS2	HRD	13909
1.01	Large	E	VVS2	HRD	11531
1.01	Large	E	VS1	HRD	10692
1.01	Large	F	VVS1	HRD	11811
1.01	Large	F	VS1	HRD	10272
1.01	Large	G	VVS2	HRD	9993
1.01	Large	G	VS2	HRD	9293
1.01	Large	H	VVS2	HRD	9433
1.01	Large	H	VS1	HRD	9153
1.01	Large	I	VVS1	HRD	8873
1.01	Large	I	VS1	HRD	8175
1.02	Large	F	VVS2	HRD	10796
1.06	Large	H	VVS2	HRD	9890
1.02	Large	H	VS2	HRD	8959
1.09	Large	I	VVS2	HRD	9107

```
/* Save this data as skulldata.txt*/
/* SKULL DATA SET of Hooke (1926)*/
/* variables are:  id c lo lp l b bf hf h ha lb */
2  1512.5   181.5  187   186.5  145    96     137   136.5 117    103
3  1619.5   192.5  197   195    145    99.5   137   135.5 110.5  108.5
4  1595.5   199    201   202    149.5  97     133   132   111.4  104.5
5  1602     188    195   193    142.5  97.5   137.5 137.5 117.5  104
6  1497.5   194.5  197   196    141.5  96     136   135   112    101.5
7  1492.5   190.5  192   191.5  144    105.5  137   134   118.5  102.5
8  1605.5   194    199   197    154    101.5  133.5 134   121    110
9  1697     191.5  191   192    148    101.5  128   128   110    97.5
10 1579     187    187   187.5  147    97     126.5 127   109.3  97.5
11 1500     176    177   176.5  140    101.5  129   127.5 111.5  91
13 1481     191.5  196.5 196.5  149    106.5  128.5 126   114    103
14 1432.5   181    183.5 183.5  141    95.5   140   139   114.5  104.5
15 1670.5   186.5  183   186.5  154.5  106    140.5 138.5 113.8  104.5
16 1476     187    187   188    142    94     132.5 132   110.5  97
17 1590.5   190.5  195   192.5  145    95     139.5 139.5 116.3  105.5
18 1725     194    199   198    146.5  100    139.5 137   116    104.5
19 1613     189    188.5 189.5  142    99     134   134   111    103
20 1524     193    196   195.5  146    101    123.5 122   111    98
21 1345     179    179   180    144    102    127   125.5 115.5  99.5
22 1557.5   195.5  188.5 189    146.5  102    128   127   115.7  99.5
23 1632.5   184    187   185    149    104    130.5 130   113    97.5
24 1406     185    188   187.5  141.5  95     128   127.5 110.4  98
25 1726.5   191    193   191.5  145.5  94     136.5 136   120.5  102.5
27 1477     198.5  202.5 201.5  136    97     131.5 130.5 106.3  103.5
28 1296.5   182.5  191   190    134.5  85.5   128.5 127.5 107.2  102
30 1385.5   182    192   190.5  138    93     129.5 127.5 103.5  104
31 1359     180.5  181   181    140    100.5  129   127.5 106    90.5
32 1600.5   190    193   193.5  143.5  96.5   129   129   111.5  97.5
33 1521.5   182    186.5 185    144.5  99.5   132   131   110.3  101.5
34 1411     186.5  193   192.5  135.5  90.5   129.5 130.5 112.4  100
35 1502.5   185    193.5 193    140.5  95     128.5 128.5 107    99
36 1314     184    189.5 188.5  130.5  93.5   134   132   112.4  103.5
37 1309     182.5  186   185    137.5  100.5  130.5 131   109.5  105
38 1627     194    200   197.5  142.5  103    138.5 138   115.3  103
40 1478.5   185.5  184.5 185.5  142.5  102    133   133.5 114.4  104
41 1470.5   184    187.5 187    146    105    126   124.5 108.3  104.5
42 1519     185.5  184.5 184    148    97     136   137   114.7   97
43 1398     179.5  183   183.5  148    96.5   132.5 131.5 109.5  100.5
44 1292.5   180.5  188   187    137.5  87.5   122.5 122   105.3  104.5
47 1524     185.5  186   186.5  144    87.5   128.5 128.5 107.5  100
48 1600.5   192.5  199   198    148    97     122.5 121.5 102    98
49 1501.5   180    185   183    139    90.5   138   138   111.5  101
52 1554.5   187    186.5 187.5  149    92.5   135   133.5 110.5  97.5
53 1476     185    189   190    143    98.5   125   126   106    99.5
58 1229     177.5  182   182    136.5  93.5   122   122.5 102.5  95.5
63 1425     186.5  184.5 186    137    95     132   132   108.5  93
64 1241.5   171.5  175   174.5  141.5  95     123.5 122.5 104.4  101.5
67 1561     187.5  190   190.5  146    94.5   128.5 127.5 114.4  102.5
69 1549.5   187    190.5 191    148.5  101.5  128.5 127.5 97.6   95.5
70 1376.5   189    189   190.5  131    101    122.5 123.5 107    98.5
71 1248     187    188   188.5  129    98     122.5 122.5 101.3  101.5
```

```
74 1418.5   178    181   181    140    92     130.5 130   107.5  100
77 1476     190    189   191    138.5  97     127.5 127.5 112.7  96
83 1387     179.5  185   183    134    91.5   132   131   107    97.5
84 1568.5   180.5  183   182    147    96     138.5 138.5 115.2  94.5
85 1458     182.5  184   185.5  135    90.5   132.5 131.5 106.3  102
87 1283.5   178.5  183   182    130    89.5   123.5 123   99.5   96.5
88 1403.5   181.5  183   183    135    101    128   127.5 111.5  97.5
91 1359     184.5  189   188    141.5  93     126   126   104.7  93
92 1395.5   171    175   176    138    85     125   123.5 104.3  89.5
93 1581.5   192    197   196.5  137.5  101    134   133.5 111.5  107.5
94 1565     189.5  191   190.5  146    98     134   133.5 111.5  102.5
95 1212.5   178    184   185.5  133.5  94     116.5 118.5 93.5   100
/* end of data file */
```