Department of Mathematics
The University of Toledo

Master of Science Degree
Comprehensive Examination
Applied Statistics

April 18, 2009

Instructions

Do all problems.

Show all of your computations.
Prove all of your assertions or quote appropriate theorems.
Books, notes, and calculators may be used.
This is a three hour test.

1. Suppose that we know the value of y_i for each of the $N = 8$ units in the whole population. The index set for the population is

$$U = \{1, 2, 3, 4, 5, 6, 7, 8\}.$$

The values of y_i are

i	1	2	3	4	5	6	7	8
y_i	1	2	4	4	7	7	7	8

Consider the following sampling scheme:

Sample Number	Sample s	$p(s)$
1	$\{1, 3, 5, 6\}$	$\frac{1}{8}$
2	$\{2, 3, 7, 8\}$	$\frac{1}{4}$
3	$\{1, 4, 6, 8\}$	$\frac{1}{8}$
4	$\{2, 4, 6, 8\}$	$\frac{3}{8}$
5	$\{4, 5, 7, 8\}$	$\frac{1}{8}$

(a) Find the probability of selection π_i for each unit i.

(b) Let $\bar{y}$ be the mean of the sample values. What is the sampling distribution of $\hat{\tau} = 8\bar{y}$?

(c) Find the mean and variance of the sampling distribution of $\hat{\tau} = 8\bar{y}$.

2. A botanist expected a certain plant species (black medic) to be more frequently located in areas where gopher mound density was high in the previous year. To examine this suspected relationship, data were collected from 74 10m by 10m plots sampled from a large prairie. The presence or absence of black medic in 1995 was recorded for each plot along with the number of new gopher mounds constructed in 1994 and the elevation of the plot (in meters). Presence of black medic is indicated by a one and absence by a zero.

The researchers first fit the following model to the data
$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 * mounds$
to relate the presence/absence of black medic to the number of gopher mounds constructed in 1994. We call it model 1.

1). Write down the likelihood function they maximized to obtain MLE for the parameters for the model.

2). Briefly explain how you would check the adequacy of the model.

The MLE for the parameters are given below.
$\hat{\beta}_0 = -2.38, \hat{\beta}_1 = 0.40.$

3). Use these results to obtain the MLE of the value of the number of mounds needed to achieve a 0.50 probability of the presence of black medic.

Next the researchers fit a logistic model of the form
$\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 * mounds + \beta_2 * elevation.$ We call it model 2.
The fitted parameters are $\hat{\beta}_0 = -4.50, \hat{\beta}_1 = 0.29, \hat{\beta}_2 = 1.22.$

4). According to the results above, explain how the odds of the presence of black medic change with one more mound while holding the elevation constant and how the odds of the presence of black medic change with one meter increase in elevation while holding the number of mounds constant.

5) Explain how to check whether adding the elevation term provides a significant improvement in the model. Give the test statistic and the d.f. for the test.

6). Estimate the probability of the presence of black medic for a new plot with mounds=7 and elevation=2 meters.

The researchers also fit a model with an interaction term to the data which is of the form $\log \frac{\pi}{1-\pi} = \beta_0 + \beta_1 * mounds + \beta_2 * elevation + \beta_3 * (mounds * elevation)$

7). Explain how to check whether this model significantly improves over model 2. Give the test statistic and the d.f. for the test.

The following are the AIC values for the three models

	AIC
problem 2	75.86
problem 3	60.17
problem 4	51.71

8) What do these values indicate?

3. The data on the attached sheet (both sides) is the Ht (inches)-Wt (pounds) data for the males from the Pulse Dataset in Minitab. Here you will use the data, graphs, and statistics given to answer a variety of questions about this data. You must justify each answer from the attached output. If it is in the output, you do not need to re-derive or recalculate those results.

a. The regression performed addresses the relationship between Height and Weight for the population from which this sample is drawn. What is the typical regression model for this situation? What assumptions must be satisfied for the regression model results to be valid? Does the evidence in the data indicate that these assumptions are satisfied or not?

b. What do the regression results say about this relationship? Describe the results and discuss statistical significance.

c. Two unusual observations are mentioned. In what way(s) are they unusual? How would you treat them?

d. A new observation at 67 inches was predicted. Interpret each of the statistics and intervals given. State clearly what they mean.

e. What would be the predicted weight at 0 inches using this model? State how you would interpret this answer.

f. Do an appropriate chi-square test of normality for the residuals. Hopefully the fact that the two columns of printed data are sorted in order of the residuals will make this task less difficult. Justify any choices that you make and give the statistics necessary for arriving at a conclusion.

g. Find a 95% confidence interval for the mean height of the population from which this sample was drawn. Assume that Height follows a normal distribution.

h. Find a 98% confidence interval for the standard deviation of the height of the population from which this sample was drawn. Assume that Height follows a normal distribution.

i. Test whether or not the correlation between height and weight for the population from which this sample was drawn is equal to 0. Assume that (Height,Weight) follows a multivariate normal distribution. Use $\alpha = .01$ and evaluate the appropriate test statistic. Clearly identify all of the components necessary for doing the test.

j. If we use this data as a pilot study to find the sample size required to estimate the mean height for the population with a margin of error (95% confidence limits) equal to 0.1 inch, what would that sample size be? Assume that Height follows a normal distribution.

k. Similarly, if we wanted to test the null hypothesis that the mean height is 70 inches with a power of 0.8 at 71 inches for the two-sided test with level of significance $\alpha = 0.01$, what sample size would be required?

Regression Analysis: Weight versus Height

The regression equation is
Weight = - 150 + 4.36 Height

Predictor	Coef	SE Coef	T	P
Constant	-149.93	54.92	-2.73	0.008
Height	4.3559	0.7757	5.62	0.000

S = 14.9917 R-Sq = 36.4% R-Sq(adj) = 35.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	7087.8	7087.8	31.54	0.000
Residual Error	55	12361.3	224.8		
Total	56	19449.1			

Unusual Observations

Obs	Height	Weight	Fit	SE Fit	Residual	St Resid
56	72.0	195.00	163.69	2.21	31.31	2.11R
57	72.0	215.00	163.69	2.21	51.31	3.46R

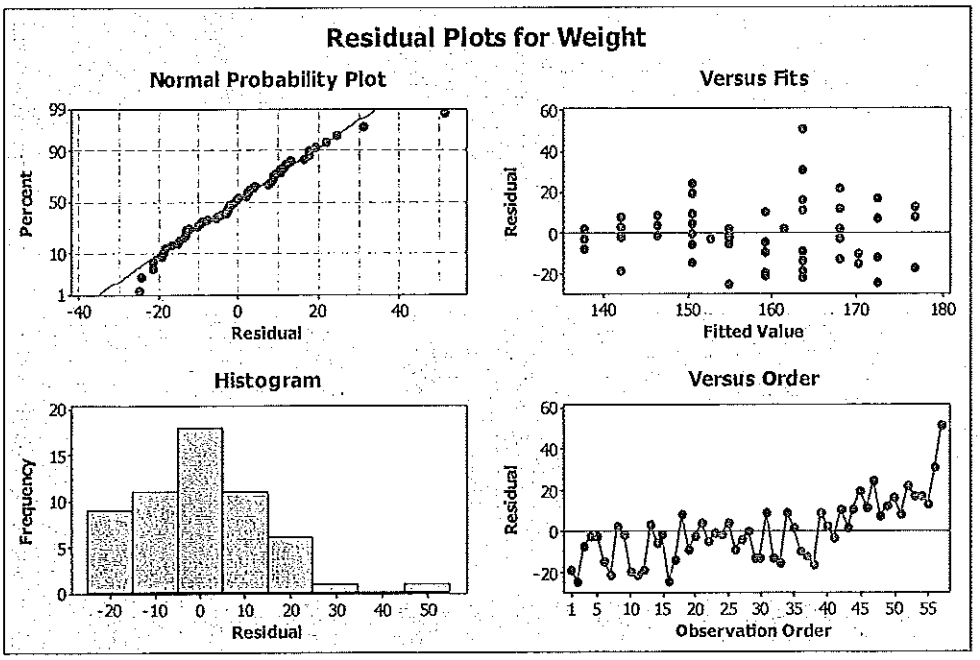
R denotes an observation with a large standardized residual.

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	141.91	3.52	(134.85, 148.97)	(111.05, 172.77)

Values of Predictors for New Observations

New Obs	Height
1	67.0



Descriptive Statistics: Height, Weight, RESI1

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Height	57	70.754	2.583	66.000	69.000	71.000	73.000	75.000
Weight	57	158.26	18.64	123.00	145.00	155.00	170.00	215.00
Residuals	57	-0.00	14.86	-24.98	-12.72	-1.27	9.06	51.31

Correlations: Height, Weight

Pearson correlation of Height and Weight = 0.604
P-Value = 0.000

Display Data

Row	Height	Weight	RESI1
1	70.0	130	-24.9772
2	74.0	148	-24.4006
3	72.0	142	-21.6889
4	71.0	138	-21.3330
5	71.0	140	-19.3330
6	67.0	123	-18.9095
7	72.0	145	-18.6889
8	75.0	160	-16.7565
9	73.5	155	-15.2227
10	69.0	136	-14.6213
11	72.0	150	-13.6889
12	73.0	155	-13.0448
13	73.0	155	-13.0448
14	73.0	155	-13.0448
15	74.0	160	-12.4006
16	73.5	160	-10.2227
17	71.0	150	-9.3330
18	72.0	155	-8.6889
19	66.0	130	-7.5537
20	69.0	145	-5.6213
21	70.0	150	-4.9772
22	71.0	155	-4.3330
23	73.0	165	-3.0448
24	69.5	150	-2.7992
25	66.0	135	-2.5537
26	66.0	135	-2.5537
27	70.0	153	-1.9772
28	67.0	140	-1.9095
29	68.0	145	-1.2654
30	69.0	150	-0.6213
31	70.0	155	0.0228
32	73.0	170	1.9552
33	70.0	157	2.0228
34	66.0	140	2.4463
35	71.5	164	2.4890
36	67.0	145	3.0905
37	68.0	150	3.7346
38	69.0	155	4.3787
39	74.0	180	7.5994
40	67.0	150	8.0905
41	75.0	185	8.2435
42	68.0	155	8.7346
43	68.0	155	8.7346
44	69.0	160	9.3787
45	71.0	170	10.6670
46	71.0	170	10.6670
47	72.0	175	11.3111
48	73.0	180	11.9552
49	75.0	190	13.2435
50	72.0	180	16.3111
51	74.0	190	17.5994
52	74.0	190	17.5994
53	69.0	170	19.3787
54	73.0	190	21.9552
55	69.0	175	24.3787
56	72.0	195	31.3111
57	72.0	215	51.3111