Applied Statistics Comprehensive Examination

Methods I & II

April 21, 2007

1. (a) (20 Points) In testing hypotheses about a population mean, there are situations in which (i) the standard Normal distribution, (ii) the t distribution, and (iii) neither the standard Normal nor the t distribution may be used. Explain when each of these three situations may occur.

   (b) Two types of gas boiler used in central heating are being checked for the dust deposits left during operation. Random samples of each type of boiler were compared under identical operating conditions. The quantities of dust (grams) deposited in each boiler over the test period were as follows:

Boiler Type 1	25	31	13	39	43	37	18	36	26	32
Boiler Type 2	27	32	18	22	17	8	13	23	20	

   Using an appropriate statistical test, investigate the hypothesis that under the given operating conditions, the deposits from Boiler Type 1 are, on average, at least 5 grams more than those from Boiler Type 2. Be sure to state any assumptions that are needed to ensure the validity of the above test.

2. (20 Points) Suppose that the defective rate for items supplied by a vendor is believed to be 1% and we are testing $H_o : p = 0.01$ vs. $H_a : p > 0.01$, where p is the population proportion of defective items in the current batch. The fate of each batch is based on an inspection of a random sample of 100 items drawn from it. The decision rule is: Reject the null hypothesis if the number of defective items in a sample of 100 is three or more.

   (a) What is the Type I error rate for this test?

   (b) If the defective rate for this batch is actually 3%, what is the power of the study?

3. (30 Points) A family doctor is interested in examining the relationship between a patient's age and total cholesterol. Fourteen female patients are randomly selected and presented below:

Subject	1	2	3	4	5	6	7
Age, x	25	25	28	32	32	32	38
Cholesterol, y	180	195	186	180	210	197	239
Subject	8	9	10	11	12	13	14
Age, x	42	48	51	51	58	62	65
Cholesterol, y	183	204	221	243	208	228	269

The following calculations have already been made for you:

$\Sigma x = 589 \quad \Sigma x^2 = 27253 \quad \Sigma y = 2943 \quad \Sigma y^2 = 628055 \quad \Sigma xy = 127276$

$\Sigma (x - \bar{x})^2 = 2472.9 \quad \Sigma (y - \bar{y})^2 = 9394.4 \quad \Sigma (x - \bar{x})(y - \bar{y}) = 3459.8$

Mean Square Error = 379.5

(a) Create a scatterplot of the data and comment on the adequacy of a linear relationship between age and total cholesterol.

(b) Calculate and interpret the correlation coefficient and the coefficient of determination.

(c) Test the significance of the linear relationship between age and total cholesterol.

4. (30 Points) Some common strategies for treating hypertensive patients by non-pharmacologic methods include weight reduction and trying to get the patient to relax more by meditational or other techniques. Eighty hypertensive patients are randomized to the following four groups:

In:	Patients Receive:
Group 1	counseling for both weight reduction and meditation
Group 2	counseling only for weight reduction
Group 3	counseling only for meditation
Group 4	no counseling at all

The change in diastolic blood pressure (calculated as before - after) is noted in these patients after a one-month period. The results are given in the following table:

Group	Sample Size	Mean Change	Standard Deviation
1	20	8.6	6.2
2	20	5.3	5.4
3	20	4.9	7.0
4	20	1.1	6.5

(a) State an appropriate mathematical model to describe the data and describe each component of the model. Be sure to state the assumptions of the model.

(b) Conduct a test of the hypothesis that there are no differences in the mean change in diastolic blood pressure across the four groups.

(c) Conduct a test of the hypothesis that the mean change in diastolic blood pressure across the groups receiving only one form of counseling is different from the mean change in diastolic blood pressure for the group receiving no counseling.