Statistics Diagnostic

August 30, 2013

NAME: ____

Work on all six problems. Write clearly and state any assumptions you make. Show what you know—partial credit is generously given.

Problem #1

Consider the following distribution for a discrete random variable:

$$p_X(k) = \alpha \cdot k \qquad k = 1, 2, \ldots, M$$

$$p_X(k) = 0 \qquad \text{otherwise}$$

a) Sketch this PDF.
b) What is the CDF for X?
c) Write an expression for α as a function of M.
d) Suppose we have a set of observations, $X_1, X_2, ..., X_n$ drawn from this distribution. Write the likelihood and log likelihood functions that you would use to estimate M.
e) What is the maximum likelihood estimator for M? Is it biased or unbiased?

For the remaining questions let $M = 4$.

f) What is $E(X)$?
g) What is $VAR(X)$?
h) Suppose we take the mean of 100 observations drawn from this distribution. What is the approximate PDF of the result?

Hint: $\sum_{k=1}^{m} k = \frac{m(m+1)}{2}$

Problem #2

The Runway Occupancy Times (ROT) for runway 28R at SFO of 60 randomly chosen large (L) jets and 60 randomly chosen heavy (H) jets are measured (in seconds). The following results are obtained:

$$\sum_{i=1}^{60} L_i = 5040 \qquad \sum_{i=1}^{60} H_i = 3950$$

$$\sum_{i=1}^{60} L_i^2 = 532000 \qquad \sum_{i=1}^{60} H_i^2 = 390000$$

$$\sum_{i=1}^{60} \left(L_i - \bar{L}\right)^2 = 110500 \qquad \sum_{i=1}^{60} \left(H_i - \bar{H}\right)^2 = 81000$$

Assuming that ROT's follow a normal distribution:

a. What is the MLE of the mean ROT for large jets?
b. What is the MLE for the variance of ROT for large jets?
c. What is an unbiased estimator for the variance of ROT for large jets?
d. What is the 95% CI for the average ROT for large jets?
e. What is the 95% CI for the variance of ROT for large jets?
f. Test whether the variances of ROT for large and heavy jets are equal, at the 0.05 level of significance.
g. Test whether the mean ROT's for large and heavy jets are equal, at the 0.05 level of significance.

Problem #3

The structural soundness of aircraft is determined using alignment checks, which involve measuring distances between different points on the aircraft and comparing them with specifications. When a measurement is outside of a prescribed tolerance, it is termed an "alignment error." The number of alignment errors is related to the number of missing rivets on the aircraft. Data for four aircraft is presented below.

Number of Missing Rivets, x	Number of Alignment Errors, y
13	7
15	7
10	5
22	12

From the above data, the following quantities were calculated:

$\bar{x} = 15$ $\bar{y} = 8$

$s_x = 5.1$ $s_y = 2.7$

$\sum_{i=1}^{4} x_i^2 = 978$

$\sum_{i=1}^{4} y_i^2 = 278$

$\sum_{i=1}^{4} (x_i - \bar{x})^2 = 78$

$\sum_{i=1}^{4} (y_i - \bar{y})^2 = 22$

$\sum_{i=1}^{4} (x_i - \bar{x}) * (y_i - \bar{y}) = 40$

a) What are the OLS estimates of the slope (β) and intercept (α) of the regression line relating x and y? Show your calculations.
b) Estimate the variance of the residual of this model.
c) How would you measure the "fit" of the OLS equation to your data? Perform the calculation.
d) What is the 95% confidence interval for the slope?
e) Test the hypothesis that $\beta = 0.5$ against the alternative hypothesis that $\beta \neq 0.5$, using a significance level of 0.05.

Problem #4

An air traffic control tower is located in an active earthquake region. When an earthquake occurs, the probability that there will be fatality depends on the magnitude of the earthquake and also on number of controllers staffing the tower at the time of shaking of the ground. Assume that the tower is fully staffed or partially staffed with probabilities 0.2 and 0.8 respectively. The earthquake magnitude may be assumed to be either strong or weak with relative frequencies 1 to 19. When a strong earthquake occurs, the tower will definitely collapse and fatalities are unavoidable. However, there is a 2/3 probability that everyone in the tower will survive a weak earthquake if the tower is partially staffed. If the tower is fully staffed during a weak earthquake, the chance of fatality will be 50-50.

Let us define the following events:

F: the tower is fully staffed
H: the tower is partially staffed
S: strong earthquake
W: weak earthquake
C: there is an earthquake-induced fatality in the control tower

Assume staffing level of the tower and earthquake magnitude are independent. Answer the following questions:

a) $P(F) = ?\ P(H) = ?\ P(S) = ?\ P(W) = ?$
b) $P(C \mid WH) = ? \quad P(C \mid WF) = ? \quad P(\bar{C} \mid WH) = ? \quad P(\bar{C} \mid WF) = ?$
c) What is the probability that there will be fatality in the event of an earthquake?
d) If some people in the tower die in the event of an earthquake, what is the probability that the tower was fully staffed when the earthquake occurred, i.e., P(F|C) = ?

Problem #5

A researcher is interested in whether BART travelers pay attention to the schedule in planning their arrival at their origin station. During a period when there are 20 minute headways and only one line is operating, she observes the following times (t) of arrival at the platform (assume the station has separate platforms for each direction, like El Cerrito Plaza), measured in minutes before the next train is scheduled to arrive:

0.3	1.7	4.6	8.4	14.4
0.7	2.4	4.8	8.7	16.3
1.2	3.4	5.1	11.7	17.3
1.5	3.5	5.1	12.5	18.2
1.6	3.7	7.3	12.7	19
1.7	4.3	8.1	12.9	19.1

The researcher decides to test the hypothesis that travelers arrive at the station at random, by testing whether these data come from a distribution with a mean of 10.

a) Perform a t-test of this hypothesis, using a significance level of 0.1.
b) The t-test is not, strictly speaking, appropriate for this problem. Why not?

Hints: $\bar{t} = 7.7$; $\overline{t^2} = 96.2$

Problem #6

A small rental car company has five vehicles for rent. Assume that the number of hybrid vehicles in its fleet is equally likely to have any value between two and five inclusive. Two cars, both hybrids, are seen leaving from the company's lot.

a) If these cars are a random sample of the fleet, what is the probability that they would both be hybrids, if the fleet contains just two hybrids?
b) Repeat a), assuming the fleet contains three, four, and five hybrids.
c) What is the probability that the fleet contains four hybrids, given that the two cars seen leaving the lot were hybrids?