STAT 2507 Assignment # 3 Solutions Fall 2008

Note 1: Due to a typo in Long answer questions 4 and 5, these question have been withdrawn from the assignment. Consequently, the total of marks for this assignment is 76. (please refer to the marking scheme given in the assignment itself). If you get mark x out of 76 then your mark for this assignment will then be $100x/76$ out of 100. TAs, please report this last mark only.

Note 2: Some answers to lab part may vary from one student to another. In such cases they serve as guidelines only as they correspond to one replication that one instructor made.

Lab part

1. Suppose that X has a binomial distribution with parameters $n = 25$ and $p = .6$. Use minitab to simulate 18 values of X:
   random 20 c1;
   binomial 25 .6.
   i)[**2**] How many of your values are less than 17? **17**
   ii)[**2**] How many of your values are between 12 and 20 inclusive?**17**
   iii)[**4**] If Y has binomial distribution with $n = 12$ and $p = .4$, use the *cdf* command (it gives you the probability $P(Y \leq k)$ for any value k between 0 and 12) wich works by typing
   cdf;
   binomial 12 .4.
   to calculate $(Y < 9)$=**.984** and $P(6 \leq Y \leq 9$=**.331**.
   iv)[**3**] If you simulate 1,000 values of Y, what should be the number of values (among the generated 1,000 values) that are less than 9? **Close to 984**

2. Let Z be a random variable having binomial distribution with parameters $n = 16$ and $p = .5$. Generate 1,000 observations from Z and answer the following questions:
   i)[**2**] What is the shape of the distribution of X?**symmetric**
   ii) Use describe command to obtain sample mean $\bar{x}$ and standard deviation s of these 100 observations. How many values (among the 1,000 you generated) fall between $\bar{x} \pm 2s$? [4]**969**
   iii)[**3**] Based on empirical rule, at least how many should fall between $np \pm 2\sqrt{np(1-p)}$?
   at least 950

3. Suppose that X has Poisson distribution with mean $\mu = 25$. Use the *cdf* command
   cdf;
   poisson 25. to answer the following questions:
   [**3**] $P(X < 15)$ =**.0124**, [**3**] $P(9 < X \leq 14)$ =**.0124-.0002=.0122**.
   If you generate 1,000 data from X then the number of values (among the generated 1,000) should be close to [**3**] **12**.

4. Supoose Y has hypergeometric distribution with parameters $N = 20$, $M = 8$, and $n = 5$.
   i)[**4**] Use the command
   cdf 2;
   hypergeometric 20 8 5. to find $P(Y \leq 2)$ =**.7038**, and $P(Y > 2)$ =**1-.7038=.2962**
   ii)[**2**] Use the "inverse cdf" command as follows
   invcdf .942;
   hypergeometric 20 8 5. to find the value (?) in $P(Y \leq ?) = .942$. The value of (?) is **3**.

Long answer questions. For these question, completely detail your answers.

1. The random variable X has hypergeometric distribution with parameters $N = 10$, $M = 6$, and $n = 3$. Using the hypergeometric distribution formula, we get

   a.

$$P(X=2) = \frac{\binom{6}{2}\binom{4}{1}}{\binom{10}{3}} = \frac{\left(\frac{6\times5}{2}\right)4}{\left(\frac{10\times9\times8}{3\times2}\right)} = \frac{1}{2}.$$

   b.

$$\begin{aligned} P(X<2) &= P(X=0)+P(X=1) \\ &= \frac{\binom{6}{0}\binom{4}{3}}{\binom{10}{3}} + \frac{\binom{6}{1}\binom{4}{2}}{\binom{10}{3}} \\ &= \frac{4+6\times6}{\left(\frac{10\times9\times8}{3\times2}\right)} = \frac{1}{3}. \end{aligned}$$

2. Let $X(t)$ be the number of messages received during t hour period of time. Then $X(t)$ has Poisson distribution with parameter $t \times 8$.

   a. Here $t = 1$ and the Poisson parameter is simply 8.

$$P(X(1)=5) = e^{-8}\frac{8^5}{5!} = .0916$$

   b. Here $t = 1.5$ and the parameter is (1.5)8=12 and

$$P(X(1.5) \le 10) = .347 \qquad \text{directly from Poisson table.}$$

   c. Here $t = .5$ and the parameter is (.5)(8)=4 and

$$P(X(.5) < 3) = P(X \le 2) = .238 \qquad \text{directly from Poisson table.}$$

3. **a.** The probability is simply $p = .3$

   b. The number X of raining days among the selected 25 days has binomial distribution with parameters $n = 25$ and $p = .3$ (assuming independence between days). Hence $\mu = np = 25(.3) = 7.5$ and $\sigma = \sqrt{np(1-p)} = 2.3$.

   c. For $X = 10$, the z-score is

$$\frac{10-7.5}{2.3} = 1.08$$

   d. No. Observations with z-scores below 3 are not considered to be outliers. Thus observing 10 raining days does not disagree with the forecast.

4. Withdrawn.

5. Withdrawn.

6. The number X of outbreaks of E. Coli per 100,000 in a given year has Poisson distribution with parameter $\mu = (.5)(2.5) = 1.25$.

   a. We note that $\mu = 1.25$ is not listed in the Poisson table. We can directly use Minitab command "cdf" and get $P(X \le 5) = .998$. Another wasy is to compute the probability with $\mu = 1$ and with $\mu = 1.5$ and then use the middle probability as an approximation of the true probability (when $\mu = 1.25$):

With $\mu = 1$, we get $P(X \leq 5) = .999$ and with $\mu = 1.5$, $P(X \leq 5) = .996$ and hence with $\mu = 1.25$, $P(X \leq 5) = (.996 + .999)/2 = .9975$ (approximately).

b. $P(X > 5) = 1 - P(X \leq 5) = .002$ (using the first method of computing $P(X \leq 5)$).

c. Method 1. We loosely consider the Poisson distribution as symmetric, and we apply empirical rule. $\mu = 1.25$ and $\sigma = \sqrt{1.25} = 1.11$. If we use the empirical rule, 95% of observations should fall between $\mu \pm 2\sigma = 1.25 \pm 2(1.11)$=[-.98 , 3.48]. hence 95% of occurrences should involve up to 3 cases (as number of cases is integer). That is in a given year, there is 95% of chances of at most 3 outbreaks per 100,000 individuals.

Method 2. More strict. We interpolate as in a) above: At $\mu = 1$, we have $P(X \leq 3) = .981$ and at $\mu = 1.5$, $P(X \leq 3) = .934$. (.95 being between these values .934 and .981). We interpolate these two values to say that at $\mu = 1.25$, approximately $P(X \leq 3) = (.981 + .934)/2 = .9575$ which is very close to .95. Hence there are 95% of chances of up to 3 outbreaks per year in 100,000.