STAT 2507 Assignment # 3 (Chapter 5 and 6) Fall 2014

Due in class: November 4, 2014

Last Name ________________________, **First Name** ________________________

Student # ________________________, **Lab Section #** ________________________

Total of marks=100.

Part I. Lab questions. Use only the blanks left to answer lab questions.

1. Suppose that X is a binomial random variable with parameters n and p. Assuming $n = 20$, generate $1,000$ observations from X for each case of $p = 0.1$ and 0.5. Answer the following questions:

   random 1000 c1;

   binomial 20 0.1.

   Repeat the same procedure for $p = 0.5$ and store the sample observations in c2.

   (a) [5] Based on the sample, what is the shape of the distribution of X?

   If $p = 0.1$, Skewed to right. If $p = 0.5$, Nearly symmetric.

   (b) [5] For the sample from the distribution with $p = 0.5$, use the *describe* command to obtain the sample mean $\bar{y}$ and the standard deviation s of these 1,000 observations. How many values (among the 1,000 generated) fall between $\bar{y} \pm 2s$? Approximately 950

   (c) [5] Based on the sample data stored in c1, find the lower and upper quartiles and then interquatile range.

   Q_1: 1, Q_3: 3, IQR: 2

2. Suppose that a random variable X has a Poisson distribution with mean $\mu = 25$. Use the *cdf* command that works by typing

   cdf;

   poisson 25.

   to answer the following questions.

(a) [5] $P(X < 20) =$ <u>0.13357</u>; $P(15 < X \leq 23) =$ <u>0.39388 - 0.02229=0.37159</u>

(b) [5] Generate 1,000 observations from the distribution of X and obtain the sample standard deviation. <u>Very close to 5</u>

3. Suppose that X has a hypergeometric distribution with parameters $N = 30$, $M = 10$, and $n = 5$. **(a)** [5] Use the command
cdf;
hypergeometric 30 10 5.
to find $P(X \leq 3) =$ <u>0.968759</u> and $P(X > 2) =$ <u>1-0.448781=0.551219</u>
(b) [5] Use the "inverse cdf" command which for a given number a between 0 and 1 yields the value of (x) that satisfies $P(X \leq \text{x}) = a$, and works by typing
invcdf a;
hypergeometric 30 10 5.
to find the smallest value of c in $P(X \leq c) > 0.80$. The value of c is <u>2</u>

4. Suppose that X has a normal distribution with mean μ=35 and variance σ^2=4.

(a) [5] Use the *cdf* command that gives you the value of $P(X \leq x)$ and works by typing
cdf;
normal 35 2.
to find $P(X > 32) =$ <u>1-0.066807=0.933193</u>, and $P(28 \leq X \leq 33) =$ <u>0.15865-0.000233=0.158422</u>.

(b) [5] Use the *invcdf* command which for a given number a between 0 and 1 gives the value of (x) that satisfies $P(X \leq x) = a$, and works by typing
invcdf a;
normal 35 2.
to find the value of c in $P(X \leq c) = 0.25$. The value of c is <u>33.6510</u>.

<u>ALSO do the following four questions:</u>

5. [10] Suppose that in a large population the proportion of people that have a certain disease is 0.01. Use the Poisson approximation to find the probability that in a random group of 400

people at least four people will have the disease.

Solution:
Let X be the number of people having the disease among the 400 people in the random group. Then $X \sim \text{Bin}(400, 0.01)$. Since $\mu = np = 4 (< 7)$, this distribution can be approximated by a Poisson distribution. If Y denotes a Poisson random variable with mean $\mu = 4$, then, the desired probability is $P(X \geq 4) \doteq 1 - P(Y \leq 3) = 1 - 0.433 = 0.567$.

6. Assume that head sizes (circumference) of new recruits in the Canadian armed forces can be approximated by a normal distribution with a mean of 22.8 inches and a standard deviation of 1.1 inches. **(a)** [6] What proportion of recruits have head sizes between 22 and 23 inches?

   Solution:
   Let X be the head size, then $X \sim N(22.8, 1.1^2)$ so
   $P(22 < X < 23) = P(\frac{22-22.8}{1.1} < Z < \frac{23-22.8}{1.1}) = P(-.73 < Z < 0.18) = 0.5714 - 0.2327 = 0.3387$

   (b) [6] 5% of the head sizes exceed __________________ inches?
   Solution:
   $P(Z > c) = 0.05 \rightarrow P(Z < c) = 0.95 \rightarrow c = 1.645 \rightarrow 1.645 = \frac{X-22.8}{1.1} \rightarrow X = 24.6095$

7. [12] A company manufactures washers, about 5% of which are defective. If a random sample of 100 washers are inspected, what is the approximate probability that fewer than 4 are defective?
   Solution:
   X: # of defective, so X has binomial distribution with $n = 100$, $p = 0.05$. It can be approximated by normal distribution since $np = 100(0.05) = 5$, $nq = 100(0.95) = 95$, so $X \sim N(5, 4.75)$
   $P(X < 4) = P(X \leq 3) = P(Z < \frac{3+0.5-5}{\sqrt{4.75}}) = P(Z < -0.69) = 0.2451$
   Also, since $\mu = np < 7$, Poisson approximation to Binomial is also correct $P(X < 4) = 0.265$.

8. A case of wine has 10 bottles, 2 of which contain spoiled wine. A sample of 3 bottles is randomly selected from the case.

(a) [4] What is probability that the sample contains 2 bottles of spoiled wine?

Solution:

X is # of spoiled bottles of wine in the sample which has hypergeometric distribution with $N = 10, M = 2, n = 3$, so

$P(X = 2) = \frac{C_2^2 C_1^8}{C_3^{10}} = \frac{1}{15}$

(b) [4] What is probability that all 3 of the sampled bottles are spoiled?

Solution:

$P(X = 3) = 0$ since there are just 2 spoiled bottles of wine.

(c) [4] What are the mean and variance of the number of spoiled bottles in the sample?

solution:

$\mu = n(\frac{M}{N}) = 3(\frac{2}{10}) = 0.6,\ \sigma^2 = n(\frac{M}{N})(1 - \frac{M}{N})(\frac{N-n}{N-1}) = 3(2/10)(8/10)(7/9) = 0.373$

9. A certain brand of computer disks averages 0.2 missing pulse per disk. Let the random variable X denotes the number of missing pulses.

(a) [3] What is the most appropriate distribution for X among binomial, Poisson, hypergeometric, and normal distributions? Poisson distribution. Use this distribution to answer (b) and (c).

(b) [3] Find the probability that the next inspected disk will have no missing pulse.

Solution:

$P(X = 0) = \frac{e^{-0.2} 0.2^0}{0!} = e^{-0.2} = 0.8187$

(c) [3] Find the probability neither of the next two disks inspected will contain any missing pulse. **Solution:**

$P(X = 0)P(X = 0) = e^{-0.2}e^{-0.2} = 0.6703$.

Or, let Y be the number of missing pulses of the next two disks inspected. Then, Y follows Poisson distribution with mean 0.4. Thus, $P(Y = 0) = e^{-0.4} = 0.6703$.