# MSIS 5503 – Statistics for Data Science – Fall 2021 - Assignment 3

1. It has been estimated that only about 30% of California residents have adequate earthquake supplies. Suppose we are interested in the number of California residents we must survey until we find a resident who does not have adequate earthquake supplies.

   a. In words, define the random variable X.
   *X= Number of residents with adequate earthquake supplies we survey (Number of "failures"), before we find one without adequate insurance.*
   *Here "success" will be" resident who does not have adequate earthquake supplies" and p, the probability of "success" is 0.7.*
   *The number of residents surveyed until we find the first one without adequate insurance ("success") n = x + 1*

   b. List the values that X may take on.
   *X= 0,1,2,…,m ,where m is the number of California Residents.*

   c. Give the distribution of X. X ~ ***Geometric****(n, 1, 0.70) where n = x+1*

   d. How many California residents do you expect to need to survey until you find a California resident who does not have adequate earthquake supplies?

   *Expected value(X) = (1-p)/p =03/0.7 = 0.4285 is the expected number of failures i.e., expected number of residents surveyed with adequate insurance before we find one without.*

   *The Expected number of residents surveyed until you find one without adequate insurance is 1+0.4285 = 1.4285.*

   I am giving credit for either answer.

   **Show using R (practice and check with calculator)**

   e. What is the probability that we must survey just one or two residents *before* we find a California resident who does not have adequate earthquake supplies?
   *Pdf =* $(1\text{-}p)^x p$
   $P(X=1) = 0.3^1 * 0.70 = 0.21$
   $P(X=2) = 0.30^2 * 0.70 = 0.\ 0.063$
   $P(X=1 \text{ OR } X=2) = 0.2273$

```
> print(dnbinom(1, 1, 0.7) + dnbinom(2, 1, 0.7))
[1] 0.273
```

   f. What is the probability that we must survey at least three California residents *before* we find a California resident who does not have adequate earthquake supplies?

*P(X >= 3) = 1 – P(X=0, X=1 or X=2) =1 – (0.7 + 0.21 + 0.063) = 0.027.*

```
> print(1- pnbinom(2, 1, 0.7))
[1] 0.027
```

2. A bridge hand is defined as 13 cards selected at random and without replacement from a deck of 52 cards. In a standard deck of cards, there are 13 cards from each suit: *hearts, spades, clubs,* and *diamonds*. We want to answer the question: what is the probability of being dealt a hand that does not contain a heart?

   *a.* What is the group of interest? *The hand that is dealt (i.e., a 13-card hand).*

   *b.* How many are in the group of interest? *13 cards.*

   *c.* How many are in the other group? *39 cards.*

   d. Let X = __________. What values does X take on?
   *X= **Number of hearts** in a 13 hand card; X=0,1,2,…,13.*

   *The distribution is **Hypergeometric(K, N, n)**: P (X = x) =* $f_{hypgeo} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
   *Here K = 13, N-K = 39, N=52, n = number of hearts = 13 and k=X.*

   e. Find the (i) mean and (ii) standard deviation of X.
   *Mean = n *(K/N) = (13)*(13/52) = 3.25*
   *Variance = n *(K/N)*((N-K)/N)*((N – n)/ (N – 1)) =*
   *(13)*(13/52)*(39/52)*(39/51) = 1.864*
   *Standard Deviation = 1.365*

   **Show using R (practice and check with calculator)**

   f. Find the probability in question (hand does not contain a heart).

   The probability question is P(*X=k=0*) *i.e., probability of being dealt a hand that does not contain a heart.*

   $P(X=0) = \frac{\binom{13}{0}\binom{39}{13}}{\binom{52}{13}} = 0.01279$

```
> print(round(dhyper(0, 13, 39, 13, log = FALSE),4))
[1] 0.0128
```

   g. Find the probability of having no more than 4 hearts in a hand.

```
> print(round(phyper(4, 13, 39, 13, log = FALSE),4))
[1] 0.8237
```

3. Suppose we draw 13 cards from a deck with 52 cards. Each time we draw a card, *we replace the card in the deck before picking another card*. We would like to look at the probability of drawing a specific number of each of the four suits (*spades*, *hearts*, *diamonds*, *clubs*). For example, we may be interested in the probability of having 3 spades, 4 hearts, 4 diamonds and 2 clubs.

   *a.* Let X = {$x_s$, $x_h$, $x_d$, $x_c$} where each $x_i$ is the number of cards of that particular suit. What is the distribution of X?

   *X ~* ***Multinomial*** *(k, n, $p_k$) where k = 4, n = 13, and each $p_k$ = 13/52 or ¼.*

   $$f_{multinom} = \frac{n!}{x_1!x_2!,\ldots,x_k!}\, p_1^{x_1} p_2^{x_2},\ldots,\, p_k^{x_k}$$

   b. What is the expected number of the number of cards of each suit (*spades*, *hearts*, *diamonds*, *clubs*) in any hand of 13 cards?

   *Mean = Expected Value $E(X_k) = np_k$;* So, Expected value of each suit is 13/4 = 3.25.

   c. What is the standard deviation of the number of cards of each suit (*spades*, *hearts*, *diamonds*, *clubs*) in any hand of 13 cards?

   *Variance = $np_k\,(1-p_k)$; Standard Deviation = $\sqrt{(3.25)(0.75)}$ = 1.5612*

   **<u>Show using R</u>**

   d. What is the probability of having a hand with 1 heart and equal number of the other 3 suits?

```
> x_vec <- c(4,1,4,4)
> n_size = 13
> prob_x <- c(0.25, 0.25, 0.25, 0.25)

> print(dmultinom(x_vec, n_size, prob_x))
[1] 0.006712228
```

4. A baseball hitter hits a home run about once every 10 times at bat. We are interested in the number of hits *before* the first home run happens.

The question should have said ("number of times at bat") not ("number of hits"). Everyone who attempted gets one point.

a. In words, define the random variable X.
*X is the number of times at bat before first home run.*
*"Success" is a Home Run and p = probability of success = 0.1.*

b. List the values that X may take on.
*X= {0, 1, 2,…,m} where m is the number of times he goes to bat before he hits the first home run*

c. Give the distribution of X including parameters, if any. i.e., X ~ ***Geometric****(n,0.1) or Negative Binomial(n,1,0.1)*

d. How many times does he bat on the average before the first home run?
*Expected Value = (1-p)p = 0.9/0.1 = 9*

e. Suppose we are interested in the number of hits *before* the *third* home run. Give the distribution of X including parameters, if any. i.e., X ~ ________()

*X ~* ***Negative Binomial*** *(n,3,0.1) where n = x+3*

**<u>Show using R (practice and check with calculator)</u>**

f. What is the probability that the he has 4 hits *before* the first home run?

$P(X=4) = (0.9)^4 * 0.1 = 0.06561$

```
> print(dnbinom(4,1,0.1))
[1] 0.06561
```

g. What is the probability that he has 7 hits *before* the third home run?

$f_{negbin} = \binom{x+r-1}{r-1}(1-p)^x p^r$

$P(X=7) = \binom{9}{2}(0.9)^7(0.1)^3 = 0.01722$

```
> print(dnbinom(7,3,0.1))
[1] 0.01721869
```

5. The maternity ward at Dr. Jose Fabella Memorial Hospital in Manila in the Philippines is one of the busiest in the world with an average of 60 births per day. Let X = the number of births in an hour.
   a. List the values that X may take on.

      *X = {0,1,2,3…}*

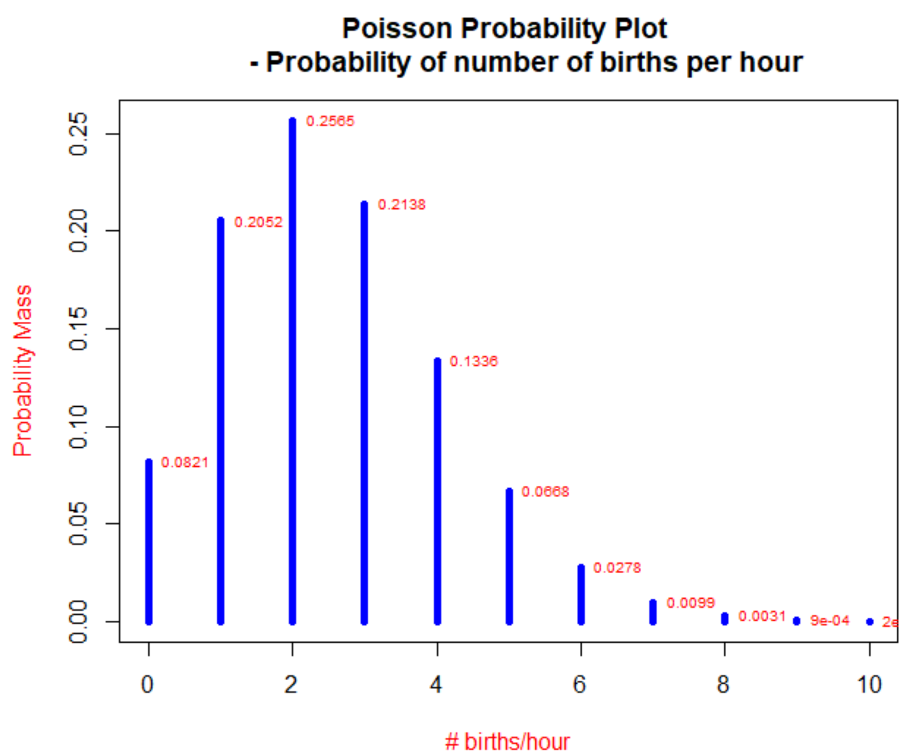
   b. Give the distribution of X including parameters, if any. i.e., X ~ ________()

      *X ~ Poisson(2.5)*

   c. What is the mean and standard deviation of X?

      *Mean = 2.5 and Standard Deviation = $\sqrt{2.5}$ = 1.5811*

   d. **Using R**, Sketch a graph of the probability distribution of X.




**<u>Show using R (practice and check with calculator)</u>**

   e. What is the probability that the maternity ward will deliver three babies in one hour?

$Pmf\ f_{pois} = P\ (X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$, $k = 1,2,3,\ldots$

$P(X = 3) = \frac{e^{-2.5}2.5^3}{3!} = 0.2138$

```
> print(dpois(3,2.5, log = FALSE))
[1] 0.213763
```

f. What is the probability that the maternity ward will deliver at most three babies in one hour?

$P(X \leq 3) = \frac{e^{-2.5}2.5^0}{0!} + \frac{e^{-2.5}2.5^1}{1!} + \frac{e^{-2.5}2.5^2}{2!} + \frac{e^{-2.5}2.5^3}{3!} = 0.0821 + 0.2052 + 0.2565 + 0.2138 = 0.7576$

```
> print(ppois(3,2.5, log = FALSE))
[1] 0.7575761
```

g. What is the probability that the maternity ward will deliver more than five babies in one hour?

$P(X > 5) = 1 - P(X \leq 5) = 1 - (\frac{e^{-2.5}2.5^0}{0!} + \frac{e^{-2.5}2.5^1}{1!} + \frac{e^{-2.5}2.5^2}{2!} + \frac{e^{-2.5}2.5^3}{3!} + \frac{e^{-2.5}2.5^4}{4!} + \frac{e^{-2.5}2.5^5}{5!})$

```
> print(1 - ppois(5,2.5, log = FALSE))
[1] 0.04202104
```