Week 05 R Workshop

Elmer V Villanueva

October 7, 2019

New functions and packages

- `choose()`
- `sum()`
- `dbinom()` calculate P(probability) of x
- `pbinom()` calculate F(cumulative probability) of x
- `dpois()`
- `ppois()`

Set your working directory

```
setwd("D:/git/DPH101-xjtlu/Y3/week05_lec_10.7")
```

Load GLOW500 Data

```
GLOW500_WORK <- read.csv("GLOW500.csv")
str(GLOW500_WORK)
```

```
## 'data.frame':    500 obs. of  15 variables:
##  $ SUB_ID   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ SITE_ID  : int  1 4 6 6 1 5 5 1 1 4 ...
##  $ PHY_ID   : int  14 284 305 309 37 299 302 36 8 282 ...
##  $ PRIORFRAC: int  0 0 1 0 0 1 0 1 1 0 ...
##  $ AGE      : int  62 65 88 82 61 67 84 82 86 58 ...
##  $ WEIGHT   : num  70.3 87.1 50.8 62.1 68 68 50.8 40.8 62.6 63.5 ...
##  $ HEIGHT   : int  158 160 157 160 152 161 150 153 156 166 ...
##  $ BMI      : num  28.2 34 20.6 24.3 29.4 ...
##  $ PREMENO  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ MOMFRAC  : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ ARMASSIST: int  0 0 1 0 0 0 0 0 0 0 ...
##  $ SMOKE    : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ RATERISK : int  2 2 1 1 2 2 1 2 2 1 ...
##  $ FRACSCORE: int  1 2 11 5 1 4 6 7 7 0 ...
##  $ FRACTURE : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
head(GLOW500_WORK)
```

```
##   SUB_ID SITE_ID PHY_ID PRIORFRAC AGE WEIGHT HEIGHT      BMI PREMENO
## 1      1       1     14         0  62   70.3    158 28.16055       0
```

```
## 2      2       4    284         0  65   87.1    160 34.02344       0
## 3      3       6    305         1  88   50.8    157 20.60936       0
## 4      4       6    309         0  82   62.1    160 24.25781       0
## 5      5       1     37         0  61   68.0    152 29.43213       0
## 6      6       5    299         1  67   68.0    161 26.23356       0
##   MOMFRAC ARMASSIST SMOKE RATERISK FRACSCORE FRACTURE
## 1       0         0     0        2         1        0
## 2       0         0     0        2         2        0
## 3       1         1     0        1        11        0
## 4       0         0     0        1         5        0
## 5       0         0     0        2         1        0
## 6       0         0     1        2         4        0
```

Let's focus on the variable `PRIORFRAC`. It's still a numeric variable and we need to transform it into a factor variable.

```
GLOW500_WORK$PRIORFRAC.F <- factor(GLOW500_WORK$PRIORFRAC,
                                   labels=c("No", "Yes"))
table(GLOW500_WORK$PRIORFRAC.F, GLOW500_WORK$PRIORFRAC)
```

```
##
##         0   1
##   No  374   0
##   Yes   0 126
```

Question 1: What is the probability of a history of prior fracture in this study population?

```
prop.table(table(GLOW500_WORK$PRIORFRAC.F))
```

```
##
##    No   Yes
## 0.748 0.252
```

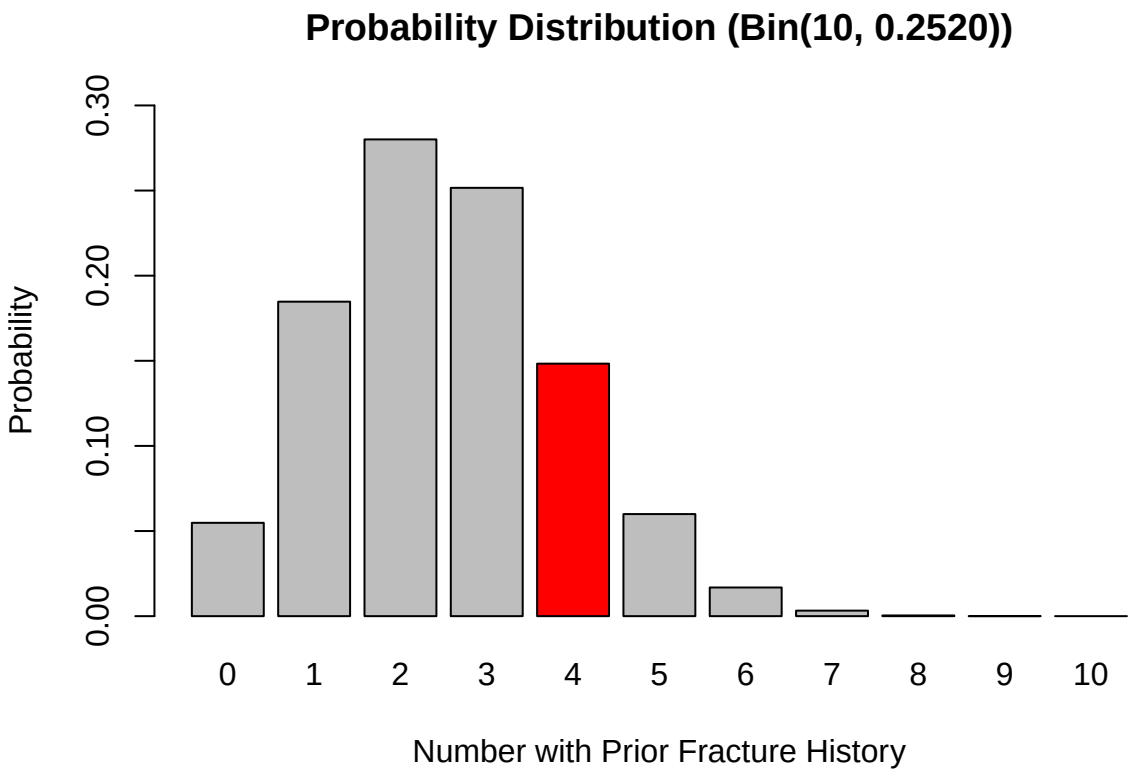
Binomial Distribution

Question 2: Say we select 10 women at random from the study population. What is the probability that exactly four of them have a history of prior fracture?

Prior fracture is a binomial process. First, the probability of prior fracture is the same for every woman. Second, the prior fracture history of one woman is not related to any other women.

The history of prior fracture is given as $126/500 = 0.2520$. In addition, we have selected 10 women. These two pieces of information give use out parameters, p and n, respectively.

Let us visualise this problem.



Finally, to calculate the probability, we can use the probability expression $P(X = x) = C\binom{n}{x}p^x(1-p)^{(n-x)}$.

```
P <- 126/500; N <- 10; X <- 4
choose(N, X) * P ^ X * (1-P) ^ (N-X)
```

```
## [1] 0.1483307
```

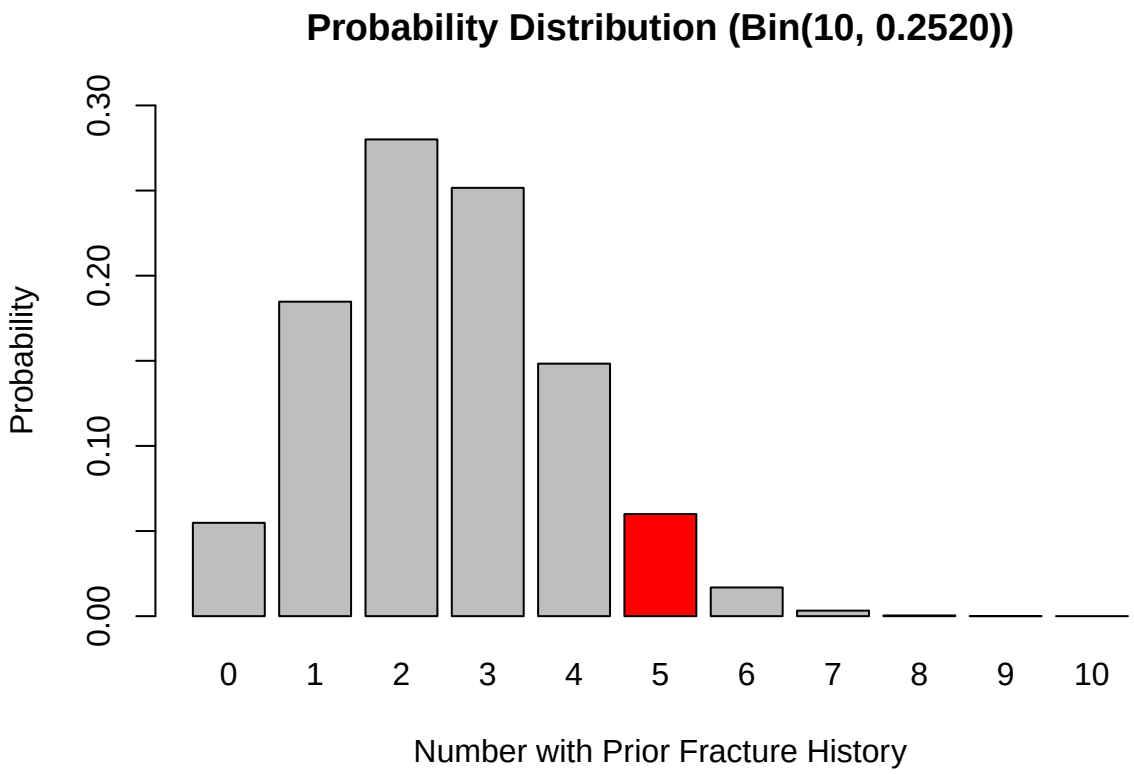
Another way to calculate this is to use the `dbinom` function, which gives the *density* or $P(X = x)$.

```
dbinom(4, 10, 0.2520) # success, all trial, probability
```

```
## [1] 0.1483307
```

Question 3: What is the probability that exactly five out of 10 have a prior fracture history?

Let's visualise the problem.

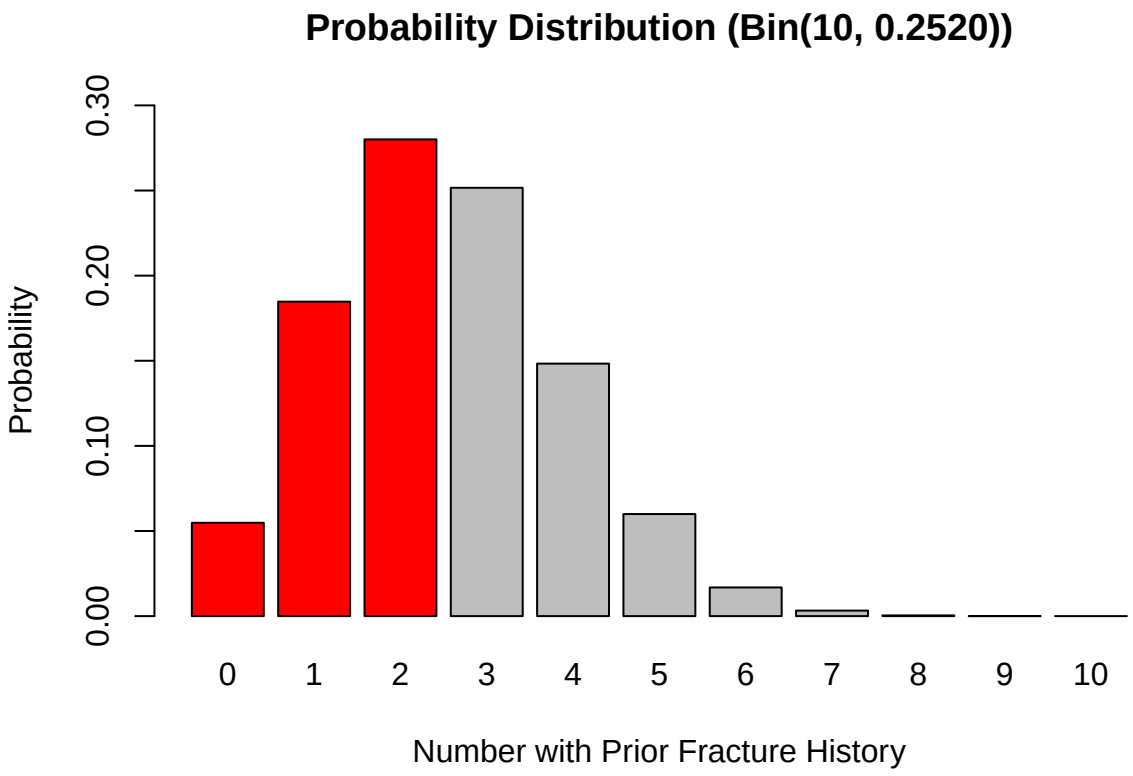


To solve,

```
dbinom(5, 10, 0.2520)
```

```
## [1] 0.05996685
```

Question 4: What is the probability that fewer than three out of 10 have a prior history of fracture?



We can solve this problem a number of ways. First, we can summate using brute force:

```
dbinom(0, 10, 0.2520) + dbinom(1, 10, 0.2520) + dbinom(2, 10, 0.2520)
```

```
## [1] 0.5195942
```

Second, we can use the `sum` function with a sequence operator:

```
sum(dbinom(0:2, 10, 0.2520))
```

```
## [1] 0.5195942
```

Third, we can use the 'pbinom' function, which gives the *distribution function* or $P(X \leq x)$.

```
pbinom(2, 10, 0.2520)
```

```
## [1] 0.5195942
```

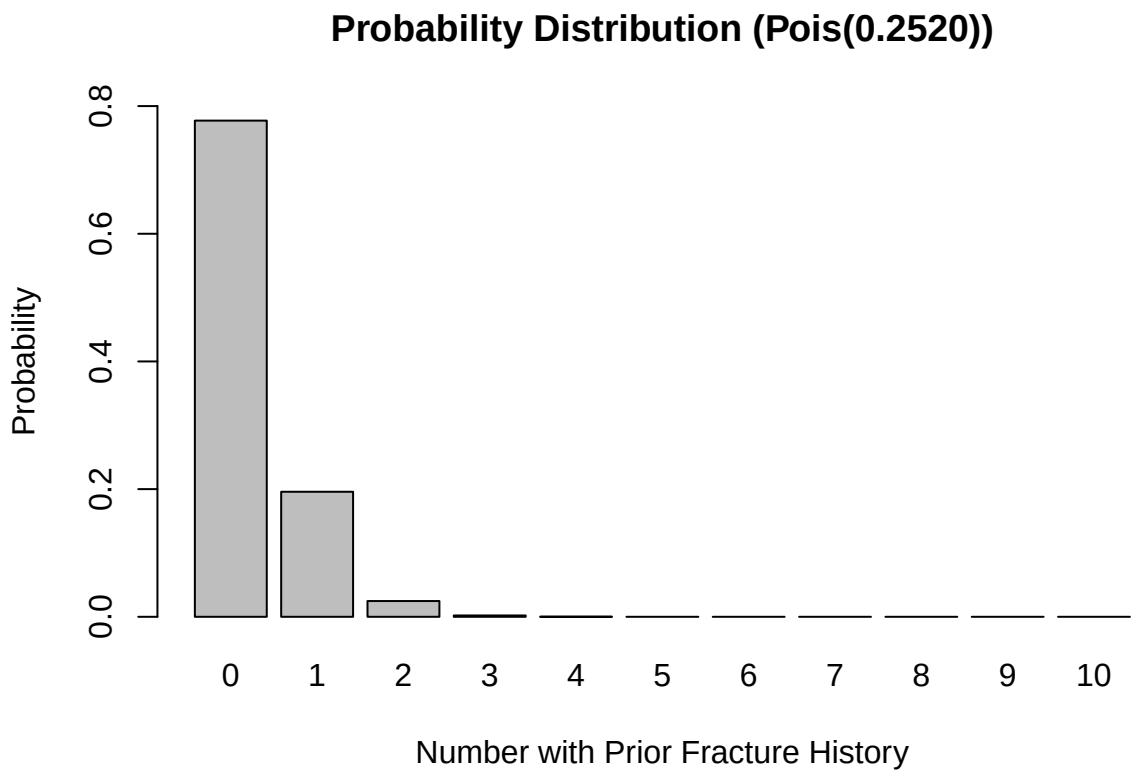
Question 5: What is the probability that two out of fourteen randomly selected women have a prior history of fracture?

Poisson Distribution

Question 6: What is the probability that exactly four randomly selected women have a prior fracture?

In this question, we can assume that the events are distributed as a Poisson process and with parameter $\lambda = 0.2520$.

Let us visualise this problem.



The probability is calculated as $P(X = x) = e^{-\lambda}\lambda^x/x!$. In this problem, $x = 4$.

```
X <- 4; LAMBDA <- 0.2520
exp(-LAMBDA) * LAMBDA ^ X / factorial(X)
```

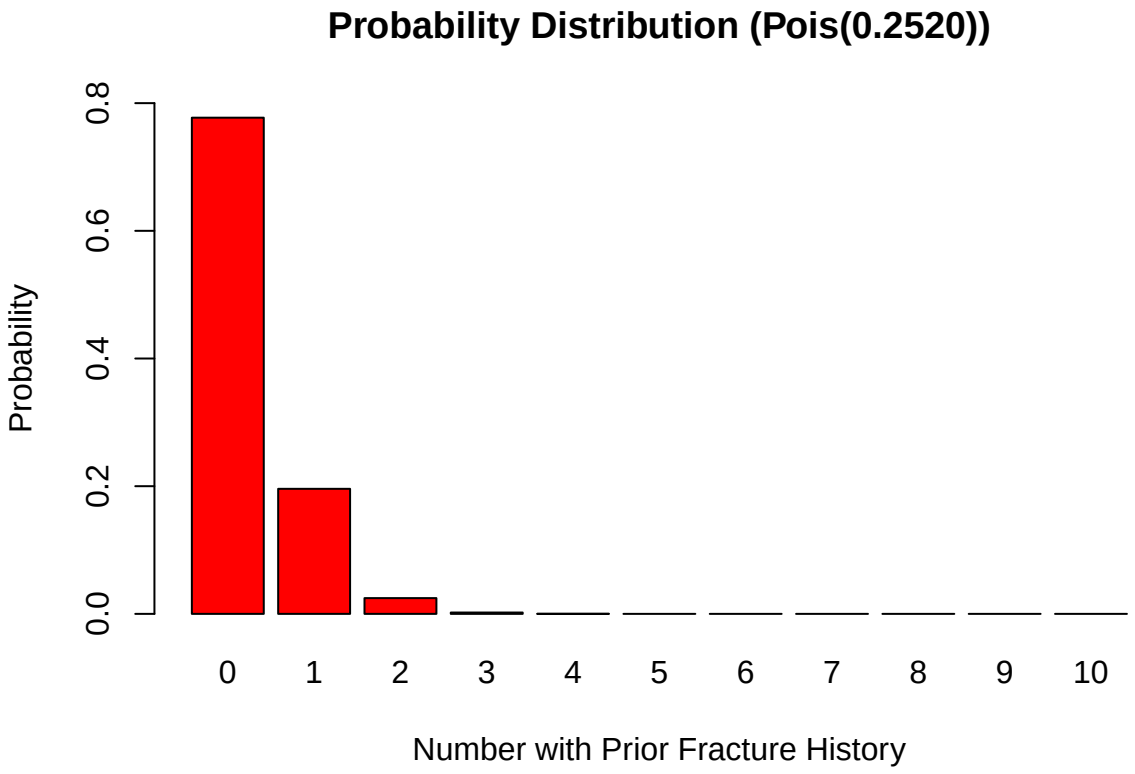
```
## [1] 0.0001306017
```

Similar to the binomial density function, `dpois` provides a Poisson density function:

```
dpois(4, 0.2520)
```

```
## [1] 0.0001306017
```

Question 7: What is the probability that fewer than three randomly selected women have a prior fracture?




```
dpois(0, 0.2520) + dpois(1, 0.2520) + dpois(2, 0.2520)
```

```
## [1] 0.9977895
```

```
sum(dpois(0:2, 0.2520))
```

```
## [1] 0.9977895
```

```
ppois(2, 0.2520)
```

```
## [1] 0.9977895
```

Question 8: What is the probability that fourteen randomly selected women have a prior history of fracture?

THE END