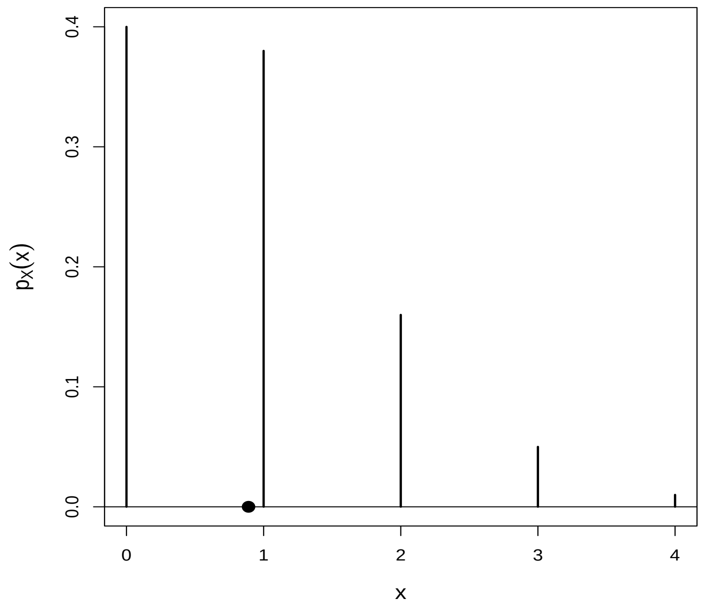
1. (a) The expected value (mean) of X is

$$\begin{aligned} E(X) &= \sum_{\text{all } x} x p_X(x) \\ &= 0(0.40) + 1(0.38) + 2(0.16) + 3(0.05) + 4(0.01) = 0.89. \end{aligned}$$

The pmf of X is shown below. $E(X)$ is shown using a solid circle.



(b) The cdf of X is

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.40, & 0 \le x < 1 \\ 0.78, & 1 \le x < 2 \\ 0.94, & 2 \le x < 3 \\ 0.99, & 3 \le x < 4 \\ 1, & x \ge 4. \end{cases}$$

2. (a) We want

$$\begin{aligned} P(X \ge 2) &= 1 - P(X \le 1) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - 0.15 - 0.08 = 0.77. \end{aligned}$$

Therefore, 77% of the students in this population are physically active for at least two days per week.

(b) We need to find the mean $\mu = E(X)$ first. We have

$$\begin{aligned} E(X) &= \sum_{\text{all } x} x p_X(x) \\ &= 0(0.15) + 1(0.08) + 2(0.10) + 3(0.11) + 4(0.10) + 5(0.12) + 6(0.07) + 7(0.27) \\ &= 3.92 \text{ days.} \end{aligned}$$

I plan to use the variance computing formula $V(X) = E(X^2) - [E(X)]^2$, so I will find $E(X^2)$ next:

$$\begin{aligned} E(X^2) &= \sum_{\text{all } x} x^2 p_X(x) \\ &= 0^2(0.15) + 1^2(0.08) + 2^2(0.10) + 3^2(0.11) + 4^2(0.10) + 5^2(0.12) + 6^2(0.07) + 7^2(0.27) \\ &= 21.82. \end{aligned}$$

The variance of X is

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= 21.82 - (3.92)^2 = 6.4536 \text{ (days)}^2. \end{aligned}$$

The standard deviation of X is

$$\sigma = \sqrt{V(X)} = \sqrt{6.4536} \approx 2.54 \text{ days}.$$

3. (a) $X \sim b(n = 150, p = 0.20)$. We are observing a fixed number of customers ("trials") and recording the number whose license is expired (number of "successes").
(b) $Y \sim \text{geom}(p = 0.20)$. We are counting the number of customers ("trials") until we find the first one with an expired license (first "success").
(c) $W \sim \text{nib}(r = 3, p = 0.20)$. We are counting the number of customers ("trials") until we find the third one with an expired license (third "success").
(d) We are conceptualizing each customer as a "trial," where a "success" means the customer's license is expired. Here are the Bernoulli trial assumptions interpreted in this context:

1. Each customer's license is expired or it is not expired.
2. The probability of an expired license $p = 0.20$ is the same for every customer.
3. The customers are independent, that is, the expired license status of any one customer is not affected by the expired license status of any other customer.

4. (a) We have

$$X = \text{the number of defective circuits} \sim b(40, 0.01).$$

This distribution arises if the Bernoulli trial assumptions hold which in this context are

1. Each circuit is either defective ("success") or not defective ("failure").
2. The probability of a defective circuit $p = 0.01$ is the same for every circuit.
3. The circuits are independent, that is, the defective status of any one circuit is not affected by the status of any other circuit.

Are these assumptions reasonable? It's hard to know for sure without having more information (the manufacturer and/or reliability engineer would likely be the expert here). The important thing to realize is that the binomial model arises if these assumptions hold. Otherwise, using the binomial distribution may give inaccurate answers.

Now onto the calculation. The product operates when there are no defective circuits, that is, when $X = 0$. We have

$$P(X=0) = \binom{40}{0}(0.01)^0(0.99)^{40} = (0.99)^{40} \approx 0.669.$$

We could also calculate this in R:

```
> options(digits=3)
> dbinom(0,40,0.01) # P(X=0)
[1] 0.669
```

(b) The mean number of defective circuits (out of 40) is

$$E(X) = np = 40(0.01) = 0.4.$$

The variance is

$$V(X) = np(1-p) = 40(0.01)(0.99) = 0.396 \text{ (circuits)}^2.$$

5. (a) Envision each LED bulb as a "trial," where $p = 0.001$ is the probability of a defective bulb (i.e., a "success") is the same for all bulbs. We want

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - P(X=0) - P(X=1) \\ &= 1 - \binom{30}{0}(0.001)^0(0.999)^{30} - \binom{30}{0}(0.001)^1(0.999)^{29} \\ &= 1 - (0.999)^{30} - 30(0.001)(0.999)^{29} = 0.000427. \end{aligned}$$

We could also calculate this in R:

```
> options(digits=3)
> 1-pbinom(1,30,0.001)
[1] 0.000427
```

(b) The calculation in part (a) was for a single light. Now, the engineers will continue to observe lights until they find the first one with 2 or more defective LED bulbs. Therefore, think of

$$\text{“observing a light with 2 or more defective bulbs”} \quad \longleftrightarrow \quad \text{“success.”}$$

We are now counting the number of lights ("trials") until we find the first success. A geometric distribution applies for

$$Y = \text{the number of lights observed to find the first one with two or more defective bulbs,}$$

and we calculated the probability of "success" in part (a). Therefore, the mean of Y is

$$E(Y) = \frac{1}{0.000427} \approx 2342.$$

We may have incurred some rounding error in this last calculation.

6. (a) A geometric distribution applies for X because we are counting the number of mosquitos ("trials") to find the first WNV-infected one ("first success"). Specifically, $X \sim \text{geom}(p = 0.0005)$. We want

$$\begin{aligned} P(X \geq 100) &= 1 - P(X \leq 99) \\ &= 1 - \sum_{x=1}^{99} (0.9995)^x (0.0005) = 1 - (0.0005) \underbrace{\sum_{x=1}^{99} (0.9995)^x}_{\text{finite geom sum}}. \end{aligned}$$

Recall the mnemonic formula for adding terms in a finite geometric sum with common ratio $0 < r < 1$ is

$$\frac{\text{“first term”} - \text{“first omitted term”}}{1 - r}.$$

Here, we recognize 0.9995 as the common ratio r in the finite geometric sum above and therefore

$$\sum_{x=1}^{99} (0.9995)^x = \frac{0.9995 - (0.9995)^{100}}{1 - 0.9995} \approx 96.6.$$

Therefore,

$$P(X \geq 100) \approx 1 - (0.0005)(96.6) = 0.952.$$

We could also calculate this in R:

```
> options(digits=3)
> 1-pgeom(99-1,0.0005)
[1] 0.952
```

(b) Define Y to be the number of mosquitos observed to find the fifth one infected with WNV. We are "waiting" for the fifth "success," so a negative binomial distribution applies. Specifically, $Y \sim \text{nib}(r = 5, p = 0.0005)$. The mean number of mosquitos that will be needed is

$$E(Y) = \frac{r}{p} = \frac{5}{0.0005} = 10000.$$

7. The hypergeometric distribution applies here. We have a finite population of $N = 40$ bags of white powder ("objects"). There are two different "classes:"

- Class 1: bags which contain cocaine ($K = 10$)
- Class 2: bags which do not contain cocaine ($N - K = 30$).

We will select $n = 4$ bags of power at random and without replacement. Let

$$X = \text{the number of bags which contain cocaine (out of 4)}.$$

Then $X \sim \text{hyper}(N = 40, n = 4, K = 10)$. In part (a), we want

$$P(X = 2) = \frac{\binom{10}{2}\binom{30}{2}}{\binom{40}{4}} \approx 0.214.$$

We could also calculate this in R:

```
> options(digits=3)
> dhyper(2,10,30,4)
[1] 0.214
```

In part (b), we want

$$\begin{aligned} P(X \leq 2) &= P(X=0)+P(X=1)+P(X=2) \\ &= \frac{\binom{10}{0}\binom{30}{4}}{\binom{40}{4}} + \frac{\binom{10}{1}\binom{30}{3}}{\binom{40}{4}} + \frac{\binom{10}{2}\binom{30}{2}}{\binom{40}{4}} \approx 0.958. \end{aligned}$$

We could also calculate this in R:

```
> options(digits=3)
> phyper(2,10,30,4)
[1] 0.958
```

8. (a) We want to calculate $P(X=0)$ and $P(X \geq 3)$. First,

$$P(X=0) = \frac{(2.5)^0 e^{-2.5}}{0!} \approx 0.0821.$$

Also,

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - \frac{(2.5)^0 e^{-2.5}}{0!} - \frac{(2.5)^1 e^{-2.5}}{1!} - \frac{(2.5)^2 e^{-2.5}}{2!} \approx 0.456. \end{aligned}$$

We could also calculate these two probabilities in R:

```
> options(digits=3)
> dpois(0,2.5) # P(X=0)
[1] 0.0821
> 1-ppois(2,2.5) # 1-P(X<=2)
[1] 0.456
```

(b) We have

$$E(C) = E(150 + 1000X + 0.1X^2) = 150 + 1000E(X) + 0.1E(X^2).$$

We know $E(X) = \lambda = 2.5$. What is $E(X^2)$? Recall $V(X) = 2.5$ and the variance computing formula

$$\begin{aligned} V(X) = E(X^2) - [E(X)]^2 &\implies E(X^2) = V(X) + [E(X)]^2 \\ &\implies E(X^2) = 2.5 + (2.5)^2 = 8.75. \end{aligned}$$

Therefore,

$$E(C) = 150 + 1000(2.5) + 0.1(8.75) = 2651.$$

The expected cost is \$2,651,000.

ADDITIONAL R CODE:

```
# Problem 1(b)
x = c(0,1,2,3,4)
prob = c(0.40,0.38,0.16,0.05,0.01)
plot(x,prob,type="h",xlab="x",ylab=expression(p[X](x)),ylim=c(0,max(prob)),
    cex.lab=1.25,lwd=2)
abline(h=0)
points(x=0.89,y=0,pch=19,cex=1.5)
```