DATA 605 Assignment 5

Daniel Moscoe

2/26/2021

1A. "Given the prevalence rate, sensitivity, and specificity estimates, what is the probability that an individual who is reported as positive by the new test actually has the disease?"

```
P_PT_given_P <- 0.96 #P(positive test | pos)
P_NT_given_N <- 0.98 #P(negative test | neg)
P_P <- 0.001 #P(pos)

P_P_given_PT <- #P(pos | positive test)
  (P_PT_given_P * P_P) / ((P_PT_given_P * P_P) + (1 - P_NT_given_N) * (1 - P_P))
P_P_given_PT
```

```
## [1] 0.04584527
```

1B. "If the median cost… is about \$100,000 per positive case total and the test itself costs \$1000 per administration, what is the total first-year cost for treating 100,000 individuals?"

The total cost is the sum of costs for: *Administering 100,000 tests;* Administering confirmatory second tests to anyone with a positive first result; *Administering second tests to anyone with a false-negative first result (assuming they eventually display symptoms and seek an additional test);* Treating all positive cases.

```
count_NT_and_N <- #number of neg individuals testing neg
  100000 * (1 - P_P) * P_NT_given_N
count_PT_and_P <- #number of pos individuals testing pos
  100000 * P_P * P_PT_given_P
count_NT_and_P <- #number of neg individuals testing pos
  100000 * P_P * (1 - P_PT_given_P)
count_PT_and_N <- #number of pos individuals testing neg
  100000 * (1 - P_P) * (1 - P_NT_given_N)

total_cost <- 100000 * 1000 + 1000 *(count_PT_and_P + count_NT_and_P + count_PT_and_N) + (100000 * P_P)

total_cost
```

```
## [1] 102098100
```

2A. "What is the probability that, after 24 months, you received exactly 2 inspections?"

```
p_2 <- 0.05
n_2 <- 24
#X = number of inspections in 24 months
```

```r
P_X_is_2 <- dbinom(2, n_2, p_2) #P(X = 2)
P_X_is_2
```

```
## [1] 0.2232381
```

2B. "What is the probability that, after 24 months, you received 2 or more inspections?"

```r
P_X_geq_2 <- 1 - pbinom(1, n_2, p_2) #1 - P(X <= 1)
P_X_geq_2
```

```
## [1] 0.3391827
```

2C. "What is the probability that you received fewer than 2 inspections?"

```r
P_X_lt_2 <- pbinom(1, n_2, p_2) #P(X <= 1)
P_X_lt_2
```

```
## [1] 0.6608173
```

2D. "What is the expected number of inspections that you should have received?"

```r
ans_2D_sim <- mean(rbinom(10000, n_2, p_2)) #Simulated E(X)
ans_2D_calc <- n_2 * p_2 #Calculated E(X)

ans_2D_sim
```

```
## [1] 1.2076
```

```r
ans_2D_calc
```

```
## [1] 1.2
```

2E. "What is the standard deviation?"

```r
ans_2E_sim <- sqrt(var(rbinom(10000, n_2, p_2))) #Simulated sd(X)
ans_2E_calc <- sqrt(n_2 * p_2 * (1 - p_2)) #Calculated sd(X)

ans_2E_sim
```

```
## [1] 1.060994
```

```r
ans_2E_calc
```

```
## [1] 1.067708
```

3A. "What is the probability that exactly 3 [patients] arrive in one hour?"

```r
lambda_3 <- 10 #lambda for problem 3
P_X_is_3 <- dpois(3, lambda_3)

P_X_is_3
```

```
## [1] 0.007566655
```

3B. "What is the probability that more than 10 arrive in one hour?"

```r
P_X_gt_10 <- 1 - ppois(10, lambda_3) #1 - P(X <= 10)

P_X_gt_10
```

```
## [1] 0.4169602
```

3C. "How many [patients] would you expect to arrive in 8 hours?"

```r
ans_3C_sim <- #Simulated 8 * E(X)
  mean(8 * rpois(10000, lambda_3))

ans_3C_calc <- #Calculated 8 * E(X)
  8 * lambda_3

ans_3C_sim
```

```
## [1] 79.972
```

```r
ans_3C_calc
```

```
## [1] 80
```

3D. "What is the standard deviation of the appropriate probability distribution?"

```r
ans_3D_sim <- #Simulated sd(X)
  sqrt(var(rpois(10000, lambda_3)))

ans_3D_calc <- #Calculated sd(X)
  sqrt(lambda_3)

ans_3D_sim
```

```
## [1] 3.125897
```

```r
ans_3D_calc
```

```
## [1] 3.162278
```

3E. "If there are three family practice providers that can see 24 templated patients each day, what is the percent utilization and what are your recommendations?"

My plan is to determine the fraction of days for which every patient arriving could be seen.

```r
P_X_leq_72 <- #Probability that at most 24 * 3 = 72 patients arrive on a given day
  ppois(72, 8 * lambda_3)

P_X_leq_72
```

```
## [1] 0.2023664
```

Then I'll determine the number of available appointments per day that would be required so that the clinic turns away patients on no more than 10% of days. This is the 90th percentile of a Poisson distribution with lambda = 8 * 10 = 80.

```r
P_90 <- qpois(0.90, 8 * lambda_3)

P_90
```

```
## [1] 92
```

In order to meet demand on at least 90% of days, the clinic needs capacity to handle 92 patients per day. That means one additional provider should be hired. Under the current staffing regime, the clinic will turn away patients on almost 80% of days.

4A. "If your subordinate acted innocently, what was the probability he/she would have selected five nurses for the trips?"

#####Explanation

```r
m_4 <- 15 #Number of nurses
n_4 <- 15 #Number of non-nurses
k_4 <- 6 #Number of staff members selected
#X = Number of nurses selected

P_X_geq_5 <- #1 - P(X < 5)
  1 - phyper(4, m_4, n_4, k_4) #####

P_X_geq_5
```

```
## [1] 0.08429119
```

4B, 4C. "How many nurses would we have expected your subordinate to send? How many non-nurses would we have expected your subordinate to send?"

Because the number of nurses and non-nurses are equal, the expected number of nurses sent equals the expected number of non-nurses sent.

```r
ans_4BC_sim <- mean(rhyper(10000, m_4, n_4, k_4))
ans_4BC_calc <- (m_4 * k_4) / (m_4 + n_4)

ans_4BC_sim
```

```
## [1] 2.9847
```

```r
ans_4BC_calc
```

```
## [1] 3
```

5A. "What is the probability that the driver will be seriously injured during the course of the year?"

```r
p_5 = 0.001
n_5 = 1200
#X = Hour in which first serious injury occurs
P_X_leq_1200 <- #P(X <= 1200)
  pgeom(n_5, p_5)

P_X_leq_1200
```

```
## [1] 0.6992876
```

5B. "[What is the probability that the driver will be seriously injured] in the course of 15 months?"

```r
P_X_leq_1500 <- #P(X <= 1500)
  pgeom(1500, p_5)

P_X_leq_1500
```

```
## [1] 0.7772602
```

5C. "What is the expected number of hours that a driver will drive before being seriously injured?"

```r
ans_5C_sim <- mean(rgeom(10000, p_5))
ans_5C_calc <- 1 / p_5

ans_5C_sim
```

```
## [1] 998.5845
```

```r
ans_5C_calc
```

```
## [1] 1000
```

5D. "Given that a driver has driven 1200 hours, what is the probability that he or she will be injured in the next 100 hours?"

Because the probability of injury in each hour is independent, the probability of injury in any 100 hour span is constant.

```r


P_X_leq_1300_g_1200 <- #P(X <= 1300 | X > 1200)
  pgeom(100, p_5)

P_X_leq_1300_g_1200
```

Why are these numbers different?

```
## [1] 0.09611265
```

```r
100 * (0.999^99) * 0.001
```

```
## [1] 0.09056978
```

6A. "What is the probability that the generator will fail more than twice in 1000 hours?"

```r
#X = Number of failures in a given time period.
P_X_gt_2 <- 1 - ppois(2, 1) #1 - P(X <= 2)
P_X_gt_2
```

```
## [1] 0.0803014
```

6B. "What is the expected value of [failures in 1000 hours]?"

```r
ans_6B_sim <- mean(rpois(10000, 1))
ans_6B_calc <- 1

ans_6B_sim
```

```
## [1] 1.0087
```

```r
ans_6B_calc
```

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

7A. "What is the probability that this patient will wait more than 10 minutes?"

```r
#X = Number of minutes the patient will wait

P_X_gt_10 <- #1 - P(X <= 10)
  2/3

P_X_gt_10
```

```
## [1] 0.6666667
```

7B. "If the patient has already waited 10 minutes, what is the probability that he/she will wait at least another 5 minutes prior to being seen?"

```r
#P(X >= 15 | X > 10) =
  #(1 - P(X < 15)) / (1 - P(X <= 10))
ans_7B <- (1/2) / (2/3)

ans_7B
```

```
## [1] 0.75
```

7C. "What is the expected waiting time?"

```r
ans_7C_sim <- mean(runif(10000, min = 0, max = 30))
ans_7C_calc <- 30 / 2

ans_7C_sim
```

```
## [1] 14.93548
```

```r
ans_7C_calc
```

```
## [1] 15
```

8A. "What is the expected failure time?"

```r
#X = Number of years until failure
ans_8A_sim <- mean(rexp(10000, 0.1))
ans_8A_calc <- 10

ans_8A_sim
```

```
## [1] 9.934534
```

```r
ans_8A_calc
```

```
## [1] 10
```

8B. "What is the standard deviation?"

```r
ans_8B_sim <- sqrt(var((rexp(10000,0.1))))
ans_8B_calc <- 10

ans_8B_sim
```

```
## [1] 9.697828
```

```r
ans_8B_calc
```

```
## [1] 10
```

8C. "What is the probability that your MRI will fail after 8 years?"

```r
P_X_gt_8 <- #1 - P(X <= 8)
  1 - pexp(8, 0.1)

P_X_gt_8
```

```
## [1] 0.449329
```

8D. "Given that you already owned the machine 8 years, what is the probability that it will fail in the next two years?"

Because the exponential distribution is "memoryless," the probability of failure in any two-year period is constant.

```r
P_X_leq_2 <- #P(X <= 10 | X > 8)
  pexp(2, 0.1)

P_X_leq_2
```

```
## [1] 0.1812692
```