DATA 605: Assignment 5

CUNY Spring 2021

Philip Tanofsky

28 February 2021

Problem 1

(Bayesian). A new test for multinucleoside-resistant (MNR) human immunodeficiency virus type 1 (HIV-1) variants was recently developed. The test maintains 96% sensitivity, meaning that, for those with the disease, it will correctly report "positive" for 96% of them. The test is also 98% specific, meaning that, for those without the disease, 98% will be correctly reported as "negative." MNR HIV-1 is considered to be rare (albeit emerging), with about a .1% or .001 prevalence rate. 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? If the median cost (consider this the best point estimate) 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?

Solution

Build a table based on the prevalence rate, sensitivity, and specificity estimates.

	Prev. Rate	Test Pos.	Test Neg.
Have Disease	1	.96	.04
Don't Have Diseae	999	19.98	979.02
Total	1000	20.94	979.06

The above table indicates the probability that an individual who is reported as positive by the new test actually has the disease is ~4.5% (.96 / 20.94).

For the scenario of testing 100,000 individuals, and given the above estimates, the total first-year cost would be $309,400,000. First, the 100,000 tests would cost $100,000,000, and given the expected number of positive test cases to be 2094 from the 100,000 individuals, the total for care for positive test cases is $209,400,000, thus the overall total of **$309,4000,000**.

Problem 2

(Binomial). The probability of your organization receiving a Joint Commission inspection in any given month is .05. What is the probability that, after 24 months, you received exactly 2 inspections? What is the probability that, after 24 months, you received 2 or more inspections? What is the probability that your received fewer than 2 inspections? What is the expected number of inspections you should have received? What is the standard deviation?

Solution

$$(Px) = \binom{n}{x} \times p^2 \times (1-p)^{n-x}$$

X = 2, p = .05, n = 24

$$(Px = 2) = \binom{24}{2} \times .05^2 \times (.95)^{22}$$

```
two_insp <- choose(24, 2) * .05^2 * .95^22
```

Exactly two inspections 0.2232381.

```
# For exactly 2
dbinom(2, size=24, prob=0.05)
```

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

Confirming above with an R function

```
less_than_two <- pbinom(1, size=24, prob=0.05)
two_or_more <- 1 - less_than_two
```

Two or more inspections 0.3391827 and less than two inspections 0.6608173.

```
n <- 24
prob <- .05
x <- 0:24
p <- dbinom(x,n,prob)
(p * 100)
```

```
##  [1] 2.919890e+01 3.688282e+01 2.232381e+01 8.616209e+00 2.380795e+00
##  [6] 5.012199e-01 8.353665e-02 1.130571e-02 1.264455e-03 1.183116e-04
## [11] 9.340386e-06 6.256718e-07 3.567427e-08 1.733163e-09 7.167214e-11
## [16] 2.514812e-12 7.445167e-14 1.844004e-15 3.774278e-17 6.273038e-19
## [21] 8.253997e-21 8.274684e-23 5.938768e-25 2.717972e-27 5.960464e-30
```

The result with the highest probability (or likelihood) is 1.

```
# Compute mean
mu <- sum(x*p)
# Compute variance
sigma.sq=sum(x^2*p)-mu^2
# Compute standarad deviation
sigma=sqrt(sigma.sq)
```

Mean (or expected value) is 1.2.

Variance is 1.14.

Standard deviation is 1.0677078.

Problem 3

(Poisson). You are modeling the family practice clinic and notice that patients arrive at a rate of 10 per hour. What is the probability that exactly 3 arrive in one hour? What is the probability that more than 10 arrive in one hour? How many would you expect to arrive in 8 hours? What is the standard deviation of the appropriate probability distribution? 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?

Solution

$$(Px:\mu) = \frac{e^{-\mu} \times \mu^x}{x!} where\ \mu\ is\ the\ mean$$

Exactly 3

$$(P_{3:10}) = \frac{e^{-10} \times 10^3}{3!}$$

```
prob_3 <- exp(-10) * (10^3) / 6
```

Probability of 3 in one hour: 0.0075667 (or 0.7566655%).

```
prob_3_v2 <- ppois(3, lambda=10) - ppois(2, lambda=10)
```

Confirming result using *ppois* function. Result is 0.0075667.

More than 10

```
prob_more_10 <- ppois(10, lambda=10, lower=FALSE)
```

Probability of more than 10 in one hour: 0.4169602 (or 41.696025%).

Expect in 8 hours

Answer: 80 (8 hours multiplied by rate of 10 per hour)

Standard Deviation

Standard deviation of 10 per hour rate is: 3.1622777.

Standard deviation of 80 over 8 hours: 8.9442719.

Percent Utilization

Given three family practitioners can see 24 patients each day, that's a total of 72 patients in a given day. If the expected total in an 8-hour day is 80, then percent utilization is 100%. Considering the overage of 8, 80 / 72, the calculated value is 111% percent utilization if the additional patients are seen.

Recommendation: Considering each practitioner can only see 24 patients per day, and a given day is expecting 80 patients, then I considered what are the probabilities of the patient count per day with the given rate of 80 patients per day.

```r
pprob_72 <- ppois(72, lambda=80)

pprob_80 <- ppois(80, lambda=80)

pprob_92 <- ppois(92, lambda=80)

pprob_96 <- ppois(96, lambda=80)

pprob_100 <- ppois(100, lambda=80)
```

Probability of 72 patients or fewer is 0.2023664 (or 20.2366363%), which means ~20% of the days, no patient is turned away.

Probability of 80 patients or fewer is 0.529688 (or 52.9687961%), which means ~53% of the days, no patient is turned away.

Probability of 92 patients or fewer is 0.9164425 (or 91.6442472%), which means ~92% of the days, no patient is turned away.

Probability of 96 patients or fewer is 0.9644085 (or 96.4408549%), which means ~96% of the days, no patient is turned away.

Probability of 100 patients or fewer is 0.9868311 (or 98.6831145%), which means ~99% of the days, no patient is turned away.

Given the above calculations, I would suggest the practice schedule enough family practitioners to account for 96 patients, which would mean hiring a fourth family practitioner. This would indicate that on ~96% of the days, all patients will be seen. And given 96 is a multiple of 24, I'd say it fits. This obviously doesn't take into account the cost of a family practitioner (full-time or part-time), nor the option of a family practitioner splitting time between two clinics. But given the expectation (or hope) that the clinic does not want to turn patients away, then I'd recommend 4 full-time family practitioners to ensure there's only 13 days of the year in which patients are turned away.

Problem 4

(Hypergeometric). Your subordinate with 30 supervisors was recently accused of favoring nurses. 15 of the subordinate's workers are nurses and 15 are other than nurses. As evidence of malfeasance, the accuser stated that there were 6 company-paid trips to Disney World for which everyone was eligible. The supervisor sent 5 nurses and 1 non-nurse. If your subordinate acted innocently, what was the probability he/she would have selected five nurses for the trips? How many nurses would we have expected your subordinate to send? How many non-nurses would we have expected your subordinate to send?

Solution

$$P(X = k) = \frac{\binom{K}{k} \times \binom{N-K}{n-k}}{\binom{N}{n}}$$

Where:

- K is the number of successes in the population
- k is the number of observed successes
- N is the population size
- n is the number of draws

Probability of 5 nurses

```r
K <- 15
k <- 5
N <- 30
n <- 6

result <- choose(K, k) * choose(N-K, n-k) / choose(N,n)
result
```

```
## [1] 0.07586207
```

Probability of 5 nurses selected is: 0.0758621 (or 7.5862069).

Expected count of nurses

```r
# Specify x-values for dhyper function
x_dhyper <- seq(0, 6, by = 1)

# Apply dhyper function
y_dhyper <- dhyper(x_dhyper, m = 15, n = 15, k = 6)
y_dhyper
```

```
## [1] 0.008429119 0.075862069 0.241379310 0.348659004 0.241379310 0.075862069
## [7] 0.008429119
```

Using the *dhyper* function, the most likely outcome is 3 nurses with a probability of 34.8659%. And given that 3 nurses is the most likely outcome, then 3 non-nurses is the expected result for non-nurses. Given the symmetry of the question, 30 individuals of that 15 nurses, and 15 non-nurses, the expected result is 3 of each.

Problem 5

(Geometric). The probability of being seriously injured in a car crash in an unspecified location is about .1% per hour. A driver is required to traverse this area for 1200 hours in the course of a year. What is the probability that the driver will be seriously injured during the course of the year? In the course of 15 months? What is the expected number of hours that a driver will drive before being seriously injured? Given that a driver has driven 1200 hours, what is the probability that he or she will be injured in the next 100 hours?

Solution

Given the initial percentage of .1% crashes per hour, that is the equivalent of 1 in 1000. So now, considering the one year containing 1200 hours, the calculation is 1 minus the probability of no car crash, so .999 times 1200.

```r
result_1200 <- 1 - .999^1200
```

Probability that the driver is seriously injured in 1200 hours: 0.6989866 or (69.8986571%).

Similar to above, find the probability of no car crash in 1500 hours and subtract that from 1.

```r
result_1500 <- 1 - .999^1500
```

Probability that the driver is seriously injured in 1500 hours: 0.7770372 or (77.7037236%).

Hours driven before serious injury

Finding the expected value or the mean is

$$\mu = E(x) = \frac{1}{p} = \frac{1}{.001} = 1000 \; hours$$

Injured in the next 100 hours

Given that that geometric distribution has the memoryless property, it does not matter the driver has already driven 1200, the probability that he or she will be injured in the next 100 hours is simply computing for an accident in 100 hours.

Just as above, solve for a crash not occurring in 100 hours, and then subtract from 1.

```r
result_100 <- 1 - .999^100
```

Probability that the driver is seriously injured in next 100 hours: 0.0952079 or (9.5207853%).

Problem 6

You are working in a hospital that is running off of a primary generator which fails about once in 1000 hours. What is the probability that the generator will fail more than twice in 1000 hours? What is the expected value?

Solution

Failure more than twice

```r
more_than_two <- 1 - pbinom(2, size=1000, prob=0.001)
```

Probability the generator will fail more than twice in 1000 hours: 0.0802093 or (`more_than_two * 100`%).

Expected value

```r
n <- 1000
prob <- .001

# Compute mean
mu <- x*prob
```

The mean, or expected value, for the given scenario is: 0, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.011, 0.012, 0.013, 0.014, 0.015, 0.016, 0.017, 0.018, 0.019, 0.02, 0.021, 0.022, 0.023, 0.024.

Problem 7

A surgical patient arrives for surgery precisely at a given time. Based on previous analysis (or a lack of knowledge assumption), you know that the waiting time is uniformly distributed from 0 to 30 minutes. What is the probability that this patient will wait more than 10 minutes? 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? What is the expected waiting time?

Solution

Given the wait times are 0 to 30 minutes, this indicates there are 31 possible outcomes which are uniformly distributed.

Probability that this patient will wait more than 10 minutes

More than 10 minutes indicates 20 possible outcomes of the 31.

```
result <- 20 / 31
```

Probability patient will wait more than 10 minutes: 0.6451613 (or 64.516129%).

Probability of at least 5 more minutes

Given 10 minutes have elapsed, then 20 possible outcomes remain, so waiting 5 or more minutes would be 16 of those 20 outcomes.

```
result <- 16 / 20
```

Probability patient will wait at least 5 more minutes: 0.8 (or 80%).

Expected wait time

For uniform distribution, this would be the mean of the two boundaries, 0 and 30.

```
result <- (0 + 30) / 2
```

The expected wait time: 15 minutes.

Problem 8

Your hospital owns an old MRI, which has a manufacturer's lifetime of about 10 years (expected value). Based on previous studies, we know that the failure of most MRIs obeys an exponential distribution. What is the expected failure time? What is the standard deviation? What is the probability that your MRI will fail after 8 years? Now assume that you have owned the machine for 8 years. Given that you already owned the machine 8 years, what is the probability that it will fail in the next two years?

Solution

Expected Failure Time

When μ is the historical average wait time, in this case 10 years, that means m the decay parameter is 1/10 then the mean for expected failure time is 10 years.

Standard Deviation

The standard deviation is the same as the mean for exponential distribution, so standard deviation is also 10.

Fail after 8 years

$$P(x > 8) = 1 - P(x < 8)$$

$$Since\ P(X < x) = 1 - e^{-mx},\ then\ P(X > x) = 1 - (1 - e^{-mx}) = e^{-mx}$$

$$P(x > 8) = e^{(-.1)(8)}$$

```
result <- exp(-.1 * 8)
```

Probability that your MRI will fail after 8 years: 0.449329 or (44.9328964%).

Fail in the next two years

As, exponential distribution has the memoryless property, failure in the next two years is just failure in the next two years. So taking the examples above as a template:

$$P(x < 2) = 1 - e^{(-.1)(2)}$$

```
result <- 1 - exp(-.1 * 2)
```

Probability that your MRI will fail in the next 2 years: 0.1812692 or (18.1269247%).