Eric_Hirsch_605_Assignment_5

Eric Hirsch

9/26/2021

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?

a. P(A | B) = (P(B|A)*P(A)) / P(B)

A = the probability of having the disease = .001

B = the probability of testing positive:

| P(positive test | having disease) = .001 * .96 = .00096
| P(positive test |not having disease) = .999*.02 = .01998
| .00096 + .01998 = .02094

P(B | A) = the probability of having a positive test given having the disease = .96

Thus P(A | B) = (.96*.001)/.02094 = .045845

b. Cost of treating 100k individuals:

1. How many tests do you need to get 100,000 positive individuals? = 100,000/.45845 = 2,181,263
2. Test cost = 2,181,263 * 1,000 = 2,181,263,000
3. Total cost = treatment cost + test cost = 100,000*100,000 + 2,181,263,000 = $12,181,263,000

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?

The binomial distribution gives us the number of successes in a sequence of yes/no trials.

a. Exactly 2 successes:

num of successes = 2
num of trials = 24
P(success) = .05

24!/(2!(24-2)!) * (.05)^2*(1-.05)^(24-2) = .22324 (done with Excel)

b. 2 or more successes:

To get 1 or 0 successes we can add the probabilities of each: (24!/(1!(24-1)!) * (.05)^1 * (1-.05)^(24-1)) + (24!/(1!(24-1)!) * (.05)^1*(1-.05)^(24-1)) = .369 + .292 = .661 (done with excel).

Thus, 2 or more is 1-.661 = .339

c. less than 2: .661 (see above)

d. Expected value = n*p = 1.2

e. sd = sqrt(n * p *(1-p)) = 1.068 (done with excel)

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?

The Poisson distribution expresses the probability of a given number of events occurring in a fixed interval of time or space, given a constant mean rate of occurrence. Unlike the binomial distribution, the events are continuous.

Here I will use the stats native to r:

```
#The probability that 3 arrive in an hour
pOf3 <- dpois(3,10)
pOf3
```

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

```
pOf10orLess <-0
for (i in 0:10){
  pOf10orLess = pOf10orLess + dpois(i, 10)
}

pOf11Plus = 1 - pOf10orLess
pOf11Plus
```

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

```
pof8 <- 10*8
pof8
```

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

```
#mean and variance are the same
sd = 10^.05
sd
```

```
## [1] 1.122018
```

```
#percent utilization
potenialSeenInDay <- 3*24
arriveIn8HourDay <- 8*10

percentUtilization = arriveIn8HourDay/potenialSeenInDay
percentUtilization
```

```
## [1] 1.111111
```

The average percent utlilization (clients coming to the clinic) exceeds the capacity (clients who can be seen at the clinic) by 10%. By reducing clinic hours to 7.2 or by increasing staff, the problem can be eliminated.

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?

```
#x=quantity drawn of type 1 from pop containing 2 kinds = 5
#m=number of type 1=15
#n=number of type 2=15
#k=number of items drawn = 6

pOfNurses <- dhyper(5,15,15,6,log=FALSE)
pOfNurses
```

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

a. There is a .07 chance of selecting 5 nurses, highly suspicious but possible.
b. Because there are as many nurses as non-nurses we would expect 3 nurses and 3 non-nurses.

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?

The geometric distribution represents number of failures before the first success.

pgeom(q,p) finds the probability of experiencing a certain amount of failures or less before experiencing the first success where:

q = number of failures before success p = prob of success in any given trial = .1

```
#Thus, this gives us the probability of not being injured in 1200 hours.
pNotInjured_1200 <- pgeom(1200,.001)
pInjured_1200 = 1- pNotInjured_1200
pInjured_1200
```

```
## [1] 0.3007124
```

```
# If the hours are evenly distributed, then 1200 plus (3/12)*1200 = 1500

pNotInjured_1500 <- pgeom(1500,.001)
pInjured_1500 = 1- pNotInjured_1500
pInjured_1500
```

```
## [1] 0.2227398
```

```
# The expected value= 1/p
ev <- 1/.001
ev
```

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

```
# How much the driver has already driven is not relevant.
pNotInjured_100 <- pgeom(100,.001)
pInjured_100 = 1- pNotInjured_100
pInjured_100
```

```
## [1] 0.9038874
```

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?

The question involves predicting the number of successes (generator failures) in a given set of trials. Successes are continuous so we use the Poisson distribution.

```
# The probability that 2 or less fail in an hour
pOf2OrLess <- dpois(0,1) + dpois(1,1) + dpois(2,1)
# The probability that 3 or more fail in an hour
pOf3Plus <- 1 - pOf2OrLess
pOf3Plus
```

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

```
ExpectedValue = 1^.05
ExpectedValue
```

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

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?

The continuous uniform distribution describes an experiment where there is an arbitrary outcome that lies between certain bounds. It is easy to calculate probabilities by hand because the probability simply increases uniformly from 0 to 100 over the range of 0 to 30 (in this case). Thus, the probability that a patient will wait more than 10 minutes = 1 - 10/30 = .667.

If they have already waited 10 minutes, then the range is now 20 - the probability of waiting more than 5 minutes is 1 - 5/20 = .75.

The expected waiting time would be halfway through the range - i.e. 30/2 = 15. The expected total waiting time for the patient in the second example is 20/2 = 10 minutes plus the initial wait = 20 minutes.

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?

The expected failure time (failure is a success!) is 10 years, as stated in the question. The standard deviation is the same as the mean, so 10 years.

For other questions we can use the pexp function:

pexp(q, rate, lower.tail = TRUE)

(If lower.tail =TRUE then the probabilities are P(X <= x))

Thus:

```
# Calculate lambda: E(X) = 1/lambda, 10 = 1/lambda, lambda = .1

pOf8Plus <- pexp(8, .1, lower.tail=FALSE)
pOf8Plus
```

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

```
# To get year 6 to year 8 we can subtract the function at 8 from the function at 6

pOf6OrLess <- pexp(6, .1, lower.tail=TRUE)
pOf8OrLess <- pexp(8, .1, lower.tail=TRUE)
pOf6To8 <- pOf8OrLess - pOf6OrLess
pOf6To8
```

```
## [1] 0.09948267
```