Data 605 HW 5

Jean Jimenez

2023-09-26

#1

Question

(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?

Work and Answer

First, I defined the information we were given. To answer this question, I used Bayes theorem. Since we had the sensitivity, specificity, and prevalence rate, we are able to plug them into bayes theorem to calculate the probability of an individual testing positive being positive.

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Given the costs and the population size, I calculated the total first year cost by first calculating the true number of positive tests in this population. I did this by multiplying the population size by the prevalence rate and sensitivity rate. I calculated the total cost by adding the costs per test for the whole population with the cost per positive test for the whole population.

Run code to see answer

```
#givens

sens = 0.96
spec = 0.98
prev = 0.001
```

```
pop = 100000
cost_per_pos = 100000
cost_per_test = 1000

#bayes theorem
pos_prob = (sens * prev) / (sens * prev + (1 - spec) * (1 - prev))

# true positives

true_pos = pop * prev * sens

total_cost = true_pos * cost_per_pos + pop * cost_per_test


cat("The probability of an individual who tested positive actually having the disease is",pos_prob,"or"
```

The probability of an individual who tested positive actually having the disease is 0.04584527 or 0.0

```
cat("The total first year cost of treating 100,000 individuals is $",total_cost,"\n")
```

The total first year cost of treating 100,000 individuals is $ 109600000

#2

Question

(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?

Work and Answer

This question represents a binomial distribution. Luckily, there is a package binom which helps us do binomial distributions. First I defined the probability of receiving an inspection as p. This corresponds to p in the binomial distribution expression. Then, I defined the number of months as n. dbinom() gives us the density of the probability. I used this to calculate the probability of having exactly 2 inspections in 24 months. I then used the pbinom() function to calculate the probability of 2 or more or fewer than 2 inspections. pbinom() returns the probability distribution function. To calculate the expected number of inspections, we just multiply the p by n. The standard deviation can be calculated by getting the sqrt() of the expected inspections times (1-p).

$$P_x = \binom{n}{x} p^x q^{n-x}$$

Run code to see answer

```r
library(binom)

#givens

#prob of recive inspection
p = 0.05

#num months
n=24

#prob of 2 inspections after 24 months

prob_2=dbinom(2,size=n,prob=p)

#prob 2 or more after 24 months

prob_2_more=1-pbinom(1, size=n, prob=p)

#prob fewer than 2 inspections after 24months

prob_less_2=pbinom(1, size=n, prob=p)

#expected number of inspections

expected_num=n*p

std_dev_24=sqrt(n*p*(1-p))

cat("The probability of reviving 2 inspections after 24 months is",prob_2,"or",prob_2*100,"%  \n")
```

```
## The probability of reviving 2 inspections after 24 months is 0.2232381 or 22.32381 %
```

```r
cat("The probability of reviving 2 or more inspections after 24 months is", prob_2_more,"or",prob_2_more
```

```
## The probability of reviving 2 or more inspections after 24 months is 0.3391827 or 33.91827 %
```

```r
cat("The probability of reviving fewer than 2 inspections after 24 months is", prob_less_2,"or",prob_les
```

```
## The probability of reviving fewer than 2 inspections after 24 months is 0.6608173 or 66.08173 %
```

```r
cat("The expected number of inspections in 24 months is",expected_num,'inspections \n')
```

```
## The expected number of inspections in 24 months is 1.2 inspections
```

```r
cat("The standard deviation calculated is +/-",std_dev_24,"\n")
```

```
## The standard deviation calculated is +/- 1.067708
```

#3

Question

(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?

Work and Answer

The above question is modeled after the poisson distribution. In the poisson equation, x is the probability of an event happening in a given time frame. Lambda represents the average rate of occurrence. Since the patients arrive at a rate of 10 per hour, I defined lambda as 10. To calculate the probability of 3 patients in an hour, I used dpois() which is the poison distribution function. I passed 3 as x and 10 as lambda. To get the probability of more than 10 in an hour, I subtracted 1 from the result of the probability of getting exactly 10 patients. Then, I calculated the expected number of patients in 8 hours by multiplying 8 by the average of 10 per hour to get 80. The standard deviation of the probability distribution is the standard deviation of lambda. If there are 3 providers, and they can take on 24 patients each per day, then they can see 72 patients total. The number of expected patients in 8 hours is 80. To calculate the percent utilization, I divided the expected number of patients by the total capacity at the clinic.

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Run code to see answer

```
#givens
#10 px / hr = lambda
lambda=10

#poisson

#prob 3 px in 1 hr

prob_3=dpois(3, lambda=lambda)

#prob more 10 px

prob_more_10= 1- ppois(10, lambda = lambda)

#expected num of px in 8 hrs

exp_8_hrs=8*lambda

std_dev_px=sqrt(lambda)


#24 px x 3 providers = 72
```

```
# 10 px x 8 hrs = 80

total_cap=24*3
exp_px=exp_8_hrs
percent_util=(exp_px/total_cap)*100

cat("Probability that exactly 3 patients arrive is", prob_3,"or",prob_3*100, "% \n")
```

Probability that exactly 3 patients arrive is 0.007566655 or 0.7566655 %

```
cat("Probability that more than 10 patients arrive is", prob_more_10,'or',prob_more_10*100, "% \n")
```

Probability that more than 10 patients arrive is 0.4169602 or 41.69602 %

```
cat("Expected number of patients in 8 hours", exp_8_hrs, "\n")
```

Expected number of patients in 8 hours 80

```
cat("Standard deviation", std_dev_px, "\n")
```

Standard deviation 3.162278

```
cat("Percent utilization", percent_util, "%\n")
```

Percent utilization 111.1111 %

The percent utilization is 111.111% meaning all resources of the clinic is being utilized. My recommendation is to get another provider to get to the patients that are not being seen.

If you have 5 providers, the total capacity would be increased to 120 patients. If the clinic extends its hours from 8 hours to 12 hours, and still theoretically experiences patients at the same rate, then there can be a 100% utilization rate.

#4

Question

(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?

Work and Answer

The above question is a hypergeometric distribution. To solve, first I defined all the givens. There are 15 nurses, and 30 total workers. So far, there have been 6 trips. To get the probability of the subordinate innocently selecting 5 nurses, we can replace k in the equation below with 5. K is the total number of nurses 15. N is the total number of workers, and n is the number of trips. To calculate this, I used the dhyper() function and defined each of the variables. To get the expected number of nurses that would be fair to go on the trip, I multiplied the number of trips by the proportion of nurses to workers. To calculate the expected number of non nurses on the trip, its number of trips minus the number of expected nurse.

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

Run code to see answer

```
#givens
tot_nurse=15
tot_workers=30
num_trips=6
num_nurse_trip=5

#prob selecting 5 nurses

prob_5_nurse=dhyper(num_nurse_trip,tot_nurse,tot_workers-tot_nurse,num_trips)

#exp num nurses

exp_nurse=(tot_nurse/tot_workers)*num_trips

#num non nurse

exp_not_nurse=num_trips-exp_nurse

cat("The probability of innocently selecting 5 nurses is", prob_5_nurse,"or",prob_5_nurse*100, "%\n")
```

```
## The probability of innocently selecting 5 nurses is 0.07586207 or 7.586207 %
```

```
cat("The expected number of nurses is", exp_nurse, "\n")
```

```
## The expected number of nurses is 3
```

```
cat("The expected number of non-nurses is", exp_not_nurse, "\n")
```

```
## The expected number of non-nurses is 3
```

There is a chance that the person IS favoring the nurses over the other employees.

#5

Question

(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?

Work and Answer

The above question is a geometric distribution. We were given the probability rate per hour of serious injury. To calculate the probability that a driver will be seriously injured in the course of those 1200 hours, we do 1-(1- the hourly injury rate)^1200 hours. This gives the cumulative probability of being injured per 1200 hours of driving. To calculate the probability in the course of 15 months, I first calculated 1800 hours in those 15 months by creating a proportion and cross multiplying. Then, I did the same as above but with the exponent 1800. The expected number of hours before seriously injured was calculated by getting the reciprocal of the hourly injury rate. Since 1200 hrs already happened with no injury, to calculate the probability of not being injured after 100 hours more will be the same as above, just plugging in 100 in the exponent.

$$P(X = k) = (1 - p)^{(k-1)} \times p$$

Run code to see answer

```
prob_inj_hr=0.001

prob_inj_1200= 1-(1-prob_inj_hr)^1200

#15 months is 1800 hrs

prob_inj_1800=1-(1-prob_inj_hr)^1800

exp_hrs=1/prob_inj_hr

#if 1200 hrs happend prob of 100 mr hrs injury

prob_100_hrs_more=1-(1-prob_inj_hr)^100


cat("The probability of being injured in 1200 hours is", prob_inj_1200,"or", prob_inj_1200*100,"%\n")
```

```
## The probability of being injured in 1200 hours is 0.6989866 or 69.89866 %
```

```
cat("The probability of being injured in 1800 hours is", prob_inj_1800,"or",prob_inj_1800*100, "%\n")
```

```
## The probability of being injured in 1800 hours is 0.8348499 or 83.48499 %
```

```r
cat("The expected number of hours before being seriously injured is", exp_hrs, "hours \n")
```

```
## The expected number of hours before being seriously injured is 1000 hours
```

```r
cat("The probability of being injured in the next 100 hours after 1200 hours is", prob_100_hrs_more, "o
```

```
## The probability of being injured in the next 100 hours after 1200 hours is 0.09520785 or 9.520785 %
```

#6

Question

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?

Work and Answer

I used the poisson distribution for this question because there is a fixed period of time and and an event occuring. Since the generator fails once every 1000 hours, I set lambda for this question as 1. The probability that the generator will fail twice in two hours was calculated by plugging in 2 to the ppois() function, which will be similar to saying x=2 in the equation below. The expected value is already given.

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Run code to see answer

```r
#given
lambda_2=1

#prob that generator will fail twice in 1000 hrs
prob_twice_fail_1000= 1- ppois(2, lambda = lambda_2)
cat("The probability that the generator will fail more than twice is", prob_twice_fail_1000, "or",prob_
```

```
## The probability that the generator will fail more than twice is 0.0803014 or 8.03014 %
```

```r
cat("The expected number of failures in 1000 hours is", lambda_2, "\n")
```

```
## The expected number of failures in 1000 hours is 1
```

#7

Question

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?

Work and Answer

Since this question states that the time is uniformly distributed between the 0 and 30 minutes, the probability of waiting more than 10 minutes is based on the remaining time range (10-30) divided by the total range (0 -30). If they already waited 10 minutes, the remaining time is distributed from 0-20. The probability of waiting 5 more minutes is calculated like the other probability above. The expected wait time is the average of the upper and lower limits.

Run code to see answer

```
lower_limit= 0

upper_limit=30

#prob more than 10 min

prob_10_more_min= (upper_limit-10)/(upper_limit)

#Prob if waited 10 wait 5 more min

prob_5_more_min=(20-5)/(20-0)

#exp wait time

expected_min=upper_limit/2


cat("The probability of waiting more than 10 minutes is", prob_10_more_min,"or",prob_10_more_min*100, ")
```

```
## The probability of waiting more than 10 minutes is 0.6666667 or 66.66667 %
```

```
cat("The probability of waiting at least another 5 minutes after 10 minutes is", prob_5_more_min,"or",p
```

```
## The probability of waiting at least another 5 minutes after 10 minutes is 0.75 or 75 %
```

```
cat("The expected waiting time is", expected_min, "minutes\n")
```

```
## The expected waiting time is 15 minutes
```

#8

Question

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 MRI's 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?

Work and Answer

This question is a geometric distribution. The expected failure time is already given. It is the mean lifetime of the MRI. Also, the standard deviation of failure is the same (being 10 years). The probability that the MRI will fail after 8 years is calculated by using the exp() function, which uses the exponential distribution formula. The probability that it will fail in the next two years given that you already owned it from 8 years is calculated the same way as above. If the MRI machine were new the probability would be the same.

Run code to see answer

```
#givens

rate=1/10

#prob fail 8 years

prob_fail_8yrs=exp(-rate*8)

#prob fail 2 years after 8

prob_fail_2yrs=exp(-rate*2)

print("The expected failure time is 10 years")
```

```
## [1] "The expected failure time is 10 years"
```

```
print("The standard deviation is 10 years")
```

```
## [1] "The standard deviation is 10 years"
```

```
cat("The probability that MRI will fail after 8 years is", prob_fail_8yrs,"or",prob_fail_8yrs*100, "%\n
```

```
## The probability that MRI will fail after 8 years is 0.449329 or 44.9329 %
```

```
cat("The probability that MRI will fail in the next 2 years after 8 years is", prob_fail_2yrs,"or",prob
```

```
## The probability that MRI will fail in the next 2 years after 8 years is 0.8187308 or 81.87308 %
```