

# Data 605 Assignment 5

Peter Gatica

2022-09-25

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.

Bayes' Theorem  $P(A|B) = P(B|A)P(A)/P(B)$

[https://video.search.yahoo.com/yhs/search?fr=yhs-Lkry-SF01&ei=UTF-8&hsimp=yhs-SF01&hspart=Lkry&param1=mT\\_AdEqifFR1H216HH0tRtFczBsbR2tS4Ml1-yd\\_57x9p5y5yiRTJ9QLZz71OceMq\\_GxXX3-OJVf88rnvrhKiZ\\_jMWpvrtl\\_dH1vj3PFCNT3u41E6nxn9Z5ZVePjOZm\\_TyERk0RYzztLHtlOsoAhquBBxtzqAfr1YGnAWWh40DNTcki2IjHeKj9718xtvxrq8I7-gyTz2Lw6koK9j12TAEUhgU6ZuDb4GfNhCBgxBdl2qqT\\_Idj96HZUxmi1ilXkLtljuzPVSKzvjjLASn7UTZtNYQ%2C%2C&p=bayesian+probability&type=ANYS\\_D3ZXG\\_set\\_bfr\\_%2452462\\_000000%24#id=2&vid=9ef4555e31a0976928a9d8b3f8b19f5c&action=view](https://video.search.yahoo.com/yhs/search?fr=yhs-Lkry-SF01&ei=UTF-8&hsimp=yhs-SF01&hspart=Lkry&param1=mT_AdEqifFR1H216HH0tRtFczBsbR2tS4Ml1-yd_57x9p5y5yiRTJ9QLZz71OceMq_GxXX3-OJVf88rnvrhKiZ_jMWpvrtl_dH1vj3PFCNT3u41E6nxn9Z5ZVePjOZm_TyERk0RYzztLHtlOsoAhquBBxtzqAfr1YGnAWWh40DNTcki2IjHeKj9718xtvxrq8I7-gyTz2Lw6koK9j12TAEUhgU6ZuDb4GfNhCBgxBdl2qqT_Idj96HZUxmi1ilXkLtljuzPVSKzvjjLASn7UTZtNYQ%2C%2C&p=bayesian+probability&type=ANYS_D3ZXG_set_bfr_%2452462_000000%24#id=2&vid=9ef4555e31a0976928a9d8b3f8b19f5c&action=view)

**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?**

$P(\text{disease}|\text{positive}) = (P(\text{positive}|\text{disease})P(\text{disease})) / (P(\text{positive}|\text{disease})P(\text{disease}) + P(\text{positive}|\text{non-disease})P(\text{non-disease}))$

```
# sensitivity = P(positive/disease)
pos_sensitivity = .96

neg_specificity = .98

#false_positive = P(positive/non-disease)
false_positive = 1 - neg_specificity

# prevalence = P(disease)
disease_prevalence = .001

# disease_free = P(non-disease)
disease_free = 1 - disease_prevalence

(positive_prob = (pos_sensitivity * disease_prevalence) / ((pos_sensitivity * disease_prevalence) + (false_

## [1] 0.04584527
```

```

num_individuals = 100000
median_cost = 100000
test_cost = 1000

(positive_prob)

```

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?

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

```
(num_positive = num_individuals * positive_prob)
```

```
## [1] 4584.527
```

```
(cost_to_test = test_cost * num_individuals)
```

```
## [1] 1e+08
```

```
(tot_treatment_cost_1yr = (num_positive * median_cost) + cost_to_test)
```

```
## [1] 558452722
```

If the probability is that 4.584527% of 100,000 individuals (4,585) will test positive then the total cost of testing and treating those individuals will equal \$458,500,000 plus \$100,000,000 = \$558,500,000.

2. (Binomial). The probability of your organization receiving a Joint Commission inspection in any given month is .05.

### Binomial Formula

$$b(n, p, j) = \binom{n}{j} p^j q^{n-j} \text{ where } q = 1 - p$$

What is the probability that after 24 months you received exactly 2 inspections?

```

n = 24 # number of months
j = 2 # number of inspections
p = .05 # probability of receiving an inspection

(prob_2_inspects = (choose(n, j)) * (p^j) * ((1-p)^(n-j)))

```

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

**Answer: 22.3% probability that you receive 2 inspections after 24 months** What is the probability that, after 24 months, you received 2 or more inspections?

```
j <- 1
(one_inspect <- (choose(n,j))*(p)*((1-p)^(n-j)))
```

```
## [1] 0.3688282
```

```
j <- 0
(zero_inspect <- (choose(n,j))*(p)*((1-p)^(n-j)))
```

```
## [1] 0.01459945
```

```
(prob_1_or_0 <- one_inspect - zero_inspect)
```

```
## [1] 0.3542288
```

```
(two_plus_inspect <- 1 - prob_1_or_0)
```

```
## [1] 0.6457712
```

**Answer: 92.3% probability that you receive 2 or more inspections after 24 months** What is the probability that you received fewer than 2 inspections?

```
(prob_1_or_0)
```

```
## [1] 0.3542288
```

**Answer: 92.3% probability that you receive 2 fewer than 2 inspections?** What is the expected number of inspections you should have received?

$$E(X) = n \times p$$

or

$$E[X] = (np)(p + (1-p)^{n-1})$$

```
# Two ways to calculate the expected value of binomial distribution.
(expected_value <- n * p)
```

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

```
(exp_value <- (n*p)*(p+(1-p)^(n-1)))
```

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

**Answer: The expected number of inspections you should have received = 1.2.** What is the standard deviation? standard deviation formula for a binomial distribution is

$$\sigma = \sqrt{npq} \text{ where } q = 1 - p$$

```
(std_dev <- sqrt(n*p*(1-p)))
```

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

**Answer: The standard deviation formula for this binomial distribution is 1.07.**

3. (Poisson) You are modeling the family practice clinic and notice that patients arrive at a rate of 10 per hour.

Poisson formula

$$p_x = \frac{e^{-m} m^x}{x!}$$

What is the probability that exactly 3 arrive in one hour?

```
m <- 10 # average number of events given a time interval
x <- 3  # probability that the number of people arrive in an hour

((exp(1)^(-m)*m^(x))/factorial(x))
```

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

**Answer: The probability that exactly 3 arrive in one hour is 0.0076.** What is the probability that more than 10 arrive in one hour?

```
m_range <- c(0:10)
poisson_values <- c()
for (i in m_range){
  x <- (exp(1)^(-m)*m^(i))/factorial(i)
  poisson_values <- append(poisson_values, x)
}
(1 - sum(poisson_values))
```

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

**Answer: The probability that more than 10 arrive in one hour is 0.417.** How many would you expect to arrive in 8 hours? ##### Answer: If the average rate is 10 per hour, I would expect 80 people to arrive in 8 hours.

What is the standard deviation of the appropriate probability distribution?

```
print(sqrt(m))
```

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

**Answer: The standard deviation of this Poisson distribution is 3.16.** 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?

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.

Hypergeometric Formula

$$h(N, k, n, x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$

$$h(n, a, k) = ((\text{choose}(k, x)) * (\text{choose}((n-k), (a-x)))) / (\text{choose}(n, a))$$

N = population k = number of nurses N - k = number of other than nurses n = sample size x = selects nurses a = number of company paid trips

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

```
N = 30
k = 15
a = 6
x = 5
((h_prob <- choose(k,x))*(choose((N-k), (a-x))))/(choose(N,a))
```

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

**Answer: The probability he/she would have selected five nurses for the trips is 0.076.** How many nurses would we have expected your subordinate to send? How many non-nurses would we have expected your subordinate to send?

**Answer: My recommendation would to send 3 nurses and 3 non-nurses and of course would be based on each person's performance.**

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.

Geometric formula:

$$P(T = j) = q^{j-1}p$$

What is the probability that the driver will be seriously injured during the course of the year?

```
p <- .001
j <- 1200
(((1-p)^(j-1))*p)
```

```
## [1] 0.0003013147
```

**Answer: The probability that the driver will be seriously injured during the course of the year 0.0003.** In the course of 15 months?

```
j <- 1500
((1-p)^(j-1))*p)
```

```
## [1] 0.0002231859
```

**Answer:** The probability that the driver will be seriously injured during the course of 1500 hours of driving is 0.0002. What is the expected number of hours that a driver will drive before being seriously injured?

```
(mean <- (1-p)/p)
```

```
## [1] 999
```

**Answer:** The mean (expected value) of this geometric distribution is 999 hours. Given that a driver has driven 1200 hours, what is the probability that he or she will be injured in the next 100 hours?

```
p <- .001
j <- 1200
(1-((1-p)^(j-1))*p)
```

```
## [1] 0.9996987
```

**Answer:** The probability that he or she will be injured in the next 100 hours after driving 1200 hours is 0.9997.

6. You are working in a hospital that is running off of a primary generator which fails about once in 1000 hours.

```
(exp(1)^(-1)*1^(x))/factorial(x)
```

What is the probability that the generator will fail more than twice in 1000 hours?

```
m <- 1
# solution will be to find the values when x =0, 1, 2. Then subtract the values from 1.
m_range <- c(0:2)
poisson_values <- c()
for (i in m_range){
  x <- (exp(1)^(-m)*m^(i))/factorial(i)
  poisson_values <- append(poisson_values, x)
}

(1 - sum(poisson_values))
```

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

What is the expected value?

**Answer: The average as stated in the problem is 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?

**Answer: If the wait time distributed uniformly from 0 minutes to 30 minutes then there is a 2/3 chance that the 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?

**Answer: Since 15 minutes is 1/2 of 30 minutes then I would say that the probability that the patient waits for longer than 15 minutes is 50%.** What is the expected waiting time?

**Answer: The mean (avg waiting time) is 15 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.

Exponential distribution  $m = 1 / \mu$  pdf =  $m * (\exp(1)^{-mx})$

What is the expected failure time? ##### Answer: The expected failure time is the stated expected value of 10 years.

What is the standard deviation?

```
(sqrt(1/(10^2)))
```

```
## [1] 0.1
```

**Answer: The standard deviation is 0.1.** What is the probability that your MRI will fail after 8 years?

```
m = .1
x = 8
(m*(exp(1)^((-m)*x)))
```

```
## [1] 0.0449329
```

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?