

Bayes Theorem and the Rose Gardern Massacre

2020-12-12

On September 25, President Trump and his administration held a ceremony in the White House's Rose Garden to announce Amy Coney Barrett's nomination to the Supreme Court. It was a contentious event since the majority of attendees disregarded social distancing and/or did not wear a face covering. This led to the ceremony being labeled a "superspreader event", a place where the virus is spread to large number of individuals. Within days from the event's date, several prominent members of the GOP such as Senator Mike Lee and Senator Thom Tillis tested positive for COVID-19. By the end of following week, President Trump, Melania Trump, Chris Christie, Kellyanne Conway, Hope Hicks and Univ. of Notre Dame President Rev. John Jenkins tested positive. This led to the event having the moniker The Rose Garden Massacre.

In a recent Wall Street Journal Article Use of Coronavirus Rapid Tests May Have Fueled White House Covid-19 Cluster, Experts Say there is a description of how the rapid test has been used in the White House.

"What seems to have been fundamentally misunderstood in all this was that they were using it almost like you would implement a metal detector," said Ashish Jha, dean of Brown University's School of Public Health.

All tests, including those processed in a lab, can produce false negatives, he and other experts say. Some studies have shown that the Abbott Now ID test, which can produce a result in minutes, has around a 91% sensitivity... meaning 9% of tests can produce false negatives.

Suppose the test used at the White House for COVID-19 has a 0.91 probability of positively (+) identifying the disease D when it is *present*. Suppose the test wrongly positively (+) identifies the disease D^c with probability 0.05 when the disease is not present. From statistical data it is known that 403.6 of 100,000 people in the population have the disease, so the prevalence of COVID-19 in the population is approximately $P(D) = 0.004$. Source: statista Further reading, Abbott defends rapid COVID-19 test with interim trial results and Abbott: ID NOW COVID-19 results more accurate with earlier testing.

Bayes Rule

Let's calculate some probabilities that we need to create our credibale intervals using Bayes Rule.

If we randomly chose an individual from our population and is given the Abbott Rapid test. The probability that an individual:

Test positive, $P(+)$. We can use the Law of Total Probability to calculate the probability of an individual testing positive.

$$P(+) = P(+|D) \cdot P(D) + P(+|D^c) \cdot P(D^c)$$

$$P(+) = 0.91 \cdot 0.004 + 0.05 \cdot 0.996$$

$$P(+) = 0.05344$$

The probability of an individual testing positive is 0.05 percent.

The individual actually suffers from the disease D if the test turns out to be positive, $P(D|+)$. We can use Bayes Rule to calculate the probability that an individual actually has COVID-19 if their test was positive.

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+|D) \cdot P(D) + P(+|D^c) \cdot P(D^c)}$$

$$P(D|+) = \frac{0.91 \cdot 0.004}{0.91 \cdot 0.004 + 0.05 \cdot 0.996}$$

$$P(D|+) = \frac{0.00364}{0.00364 + 0.0498}$$

$$P(D|+) = 0.068$$

The probability of an individual actually having COVID-19 if their test is positive is 0.07 percent!

This is surprising at first thought, but take into consideration how prevalent COVID-19 is in our population. Given that the test has a high false positive rate and the prevalence of the disease is small, it makes sense that individuals that do not have the disease make up the majority of the positive cases.

To put this in easier numbers, let's pretend that our population is 1,000 individuals. Since the prevalence rate of our disease is 0.004, that means only 4 individuals have the disease vs 9,996 that do not have it. Since the test we're using has a false positive rate of 0.05, 500 out of the 9,996 individuals that do not have the disease are going to test positive.

The individual actually does not suffer from the disease D^c if the test turns out to be positive, $P(D^c|+)$. To find the probability that an individual does not have the disease given that their test is positive, we can take the complement of our previous calculation.

$$P(D^c|+) = 1 - P(D|+)$$

$$P(D^c|+) = 1 - 0.068$$

$$P(D^c|+) = 0.932$$

If your test is positive, then you have a .93 chance of not having the disease.

We can also use simulation to calculate the probability of an individual actually having the disease if they tested positive.

```
### Conditional Probability Simulation
### Bayesian Probability
set.seed(19)

prev <- 0.004 # prevalence = P(L)
sens <- 0.91  # sensitivity = P(+|L)
spec <- 0.95  # specificity = 1 - P(+|L^c)

reps <- 10000

true.pos.test.pos <- 0 # counter for COVID-19 test catching a person that has the disease
test.pos <- 0          # counter for COVID-19 testing positive

for(i in 1:reps){
  # simulate a person test positive
  pos <- 0
  if(runif(1) < prev){
    pos <- 1
  }
  # if positive simulate if test positive or not
  if(pos == 1){
    if(runif(1) <= sens){
      test.pos <- test.pos + 1
      true.pos.test.pos <- true.pos.test.pos + 1
    }
  }
}
```

```

    }
    # if not positive simulate if test positive or not
    else{
      if(runif(1) <= (1-spec)) test.pos <- test.pos + 1
    }
  }

# simulated probability

true.pos.test.pos/test.pos

```

```
## [1] 0.07235622
```

The simulation returns a probability of 0.072 of having the disease if testing positive. Very close approximation to what we calculated using Bayes Rules.

```

set.seed(100)
library(ggplot2)

n <- 300
prob_pi <- 0.05344

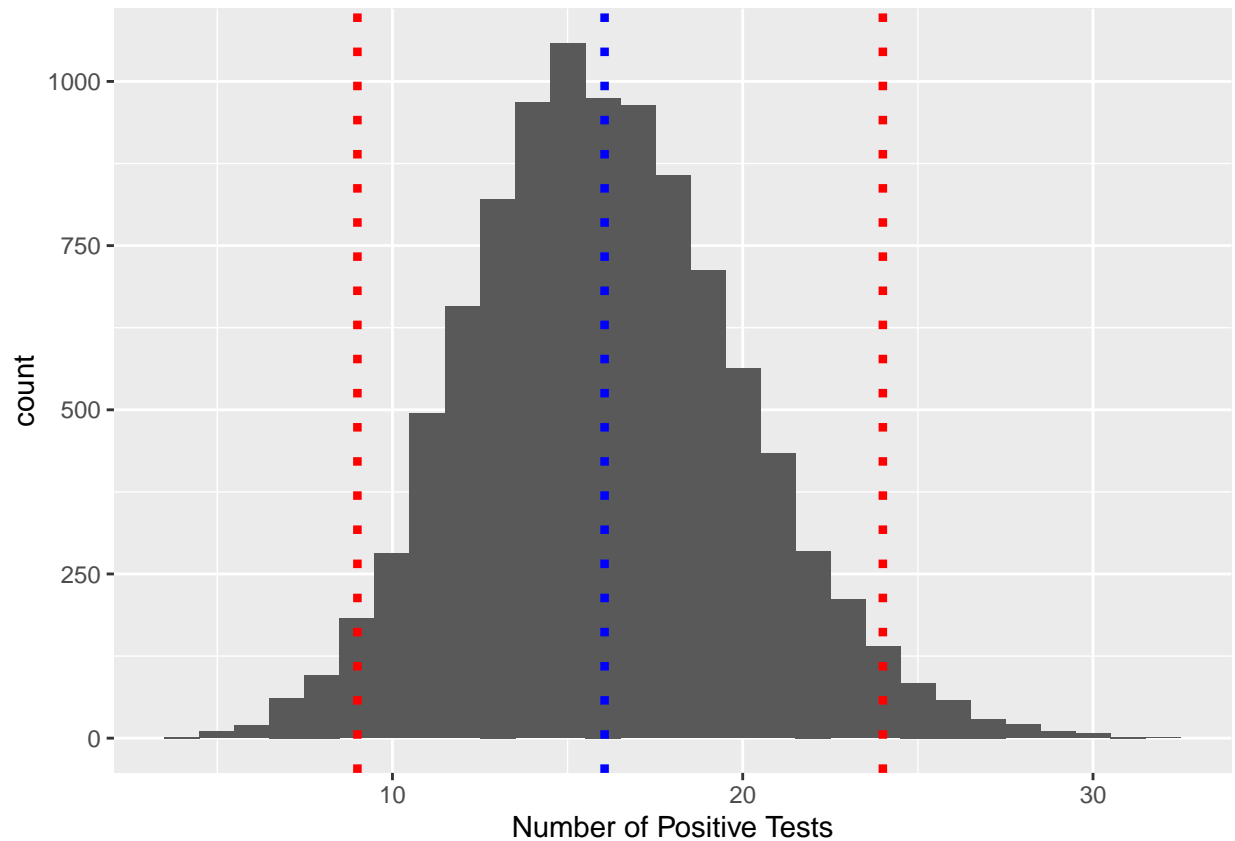
B <- 10000

x <- rbinom(B, n, prob_pi)

x_df <- data.frame(simulation = 1:B, x = x)

ggplot(x_df, aes(x = x)) +
  geom_histogram(binwidth = 1) +
  geom_vline(xintercept = mean(x), linetype="dotted", color = "blue", size=1.5)+
  geom_vline(xintercept = quantile(x, c(0.025, .975)), linetype="dotted", color = "red", size=1.5)+
  xlab("Number of Positive Tests")

```



From the Washington Post article there were 9 people listed, but it also mentions 11 people who are photographers or reports also tested positive. Is 20 people testing positive consistent with the simulations? After simulating 10,000 Rose Garden ceremonies, the average number of individuals that tested positive is 16 with 95% interval of 9 and 24. Since the median of the simulations is 16 and it has a standard deviation of 3.92. A total of 20 people that tested positive falls within one standard deviation from the mean of simulation (68 percent interval).