HW1 Q4

2023-01-23

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.3.6      ✓ purrr   0.3.4
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.4.1
## ✓ readr   2.1.2      ✓ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
```

Part B

```
calculate <- function(theta_one, y_one_given_theta_one, y_one_given_theta_zero, y) {

  if(y == 1){
    result = y_one_given_theta_one*theta_one/
                  ( y_one_given_theta_one*theta_one + (y_one_given_theta_zero)*(1-theta_one))

  }

  else if ( y == 0) {
    result = (1 - y_one_given_theta_zero)*(1 - theta_one)/
                  ((y_one_given_theta_zero)*(1 - theta_one)+ y_one_given_theta_one*(theta_one))
  }
  return(result)
}
```

Part C

```
calculate(0.001, 0.95,0.05,1 )
```

```
## [1] 0.01866405
```

Part D

The value $\mathrm{P}(y = 1 \mid \theta = 1)$ indicates the probability of test giving a positive result (result indicating the individual has the disease) when he or she truly has the said disease. A value of 0.95 would mean that 95% of the time, the result will correctly tell the infected individual that they have the disease. It is called the True Positive Rate or Recall

$\mathrm{P}(\theta = 1 \mid y = 1)$ is called the Positive Predictive Value or Precision. It is the probability of individual who tested positive to actually have the disease. A value of 0.0187 would indicate that if the test was done on an individual and tested positive, there is a 1.87% that the individual actually has the disease.

The Precision and Recall are related by Bayes' Theorem.

$(= 1 y = 1) = {(y)} $

This shows that the results are no contradictory since a large false positive rate would lead to inverse results. One could work out the formula for the other way and find similar results with large false negative rate.

Part E

```
calculate( 0.001, 0.80, 0.05,1)
```

```
## [1] 0.01576355
```

```
calculate( 0.01, 0.95, 0.05,1)
```

```
## [1] 0.1610169
```

```
calculate( 0.01, 0.80, 0.05,1)
```

```
## [1] 0.1391304
```

When $\theta = 0.01$, a medical test that is more precise (testing positive when a tested individual has the disease), $\mathrm{P}(y = 1 \mid \theta = 1)$ = 0.95 will give more accurate results of whether the person truly has disease given that they tested positive. A less precise medical test with $\mathrm{P}(y = 1 \mid \theta = 1)$ = 0.80 clearly has a lower probability of correctly guessing whether a tested individual has the disease given that they tested positive as shown in part E ii) vs E iii)

Similarly with the same precision of the medical tests of giving accurate results, let's say $\mathrm{P}(y = 1 \mid \theta = 1)$ = 0.80, the rarer the prevalence of the the disease, the harder it is to accurately say they have disease when they tested positive. Refer to part E iii) vs part E i).