Probability Practice

Kristen Lowe and Siboney Cardoso

2024-08-07

Part A

Visitors to your website are asked to answer a single survey question before they get access to the content on the page. Among all of the users, there are two categories: Random Clicker (RC), and Truthful Clicker (TC). There are two possible answers to the survey: yes and no. Random clickers would click either one with equal probability. You are also giving the information that the expected fraction of random clickers is 0.3. After a trial period, you get the following survey results: 65% said Yes and 35% said No. What fraction of people who are truthful clickers answered yes? Hint: use the rule of total probability.

```
# P(RC) = 0.3
# P(TC) = 0.7
# P(yes|RC) = 0.5
# P(no|RC) = 0.5
# P(yes) = 0.65
# P(no) = 0.35
# P(yes|TC) = ?
# P(no|TC) = ?

prob_RC <- 0.3
prob_TC <- 0.7
prob_yes_given_RC <- 0.5
prob_no_given_RC <- 0.5
prob_yes <- 0.65
prob_no <- 0.35

prob_yes_and_RC <- prob_RC * prob_yes_given_RC # 0.15
prob_no_and_RC <- prob_RC * prob_no_given_RC # 0.15

prob_yes_and_TC <- prob_yes - prob_yes_and_RC # 0.5
prob_no_and_TC <- prob_no - prob_no_and_RC # 0.2

prob_yes_given_TC <- prob_yes_and_TC / prob_TC # 0.7142857
prob_no_given_TC <- prob_no_and_TC / prob_TC # 0.2857143

prob_yes_given_TC
```

```
## [1] 0.7142857
```

71.4% of people who are truthful clickers answered yes.

Part B

Imagine a medical test for a disease with the following two attributes:

- The sensitivity is about 0.993. That is, if someone has the disease, there is a probability of 0.993 that they will test positive.
- The specificity is about 0.9999. This means that if someone doesn't have the disease, there is probability of 0.9999 that they will test negative.
- In the general population, incidence of the disease is reasonably rare: about 0.0025% of all people have it (or 0.000025 as a decimal probability).

Suppose someone tests positive. What is the probability that they have the disease?

```
# P(test positive | disease) = 0.993
# P(test negative | no disease) = 0.9999
# P(disease) = 0.000025
# P(disease | test positive) = P(disease and test positive) / P(test positive) = ?

prob_disease <- 0.000025
prob_no_disease <- 1 - prob_disease # P(no disease) = 0.999975

prob_test_pos_given_disease <- 0.993
prob_test_neg_given_no_disease <- 0.9999

prob_test_neg_given_disease <- 1 - prob_test_pos_given_disease # P(test negative | disea
se) = 0.007
prob_test_pos_given_no_disease <- 1 - prob_test_neg_given_no_disease # P(test positive |
no disease) = 0.0001

prob_disease_and_test_pos <- prob_disease * prob_test_pos_given_disease # P(disease and
test positive) = 2.4825e-05
prob_no_disease_and_test_pos <- prob_no_disease * prob_test_pos_given_no_disease # P(no
disease and test positive) = 9.99975e-05

prob_test_pos <- prob_disease_and_test_pos + prob_no_disease_and_test_pos # P(test posit
ive) = 0.0001248225

prob_disease_given_test_pos <- prob_disease_and_test_pos / prob_test_pos
prob_disease_given_test_pos
```

```
## [1] 0.1988824
```

There is a 19.9% chance that someone has the disease given that they test positive.