cancer_test_results

April 8, 2020

1 Cancer Test Results

In [52]:
```
# load dataset
import numpy as np
import pandas as pd

df = pd.read_csv("cancer_test_data.csv")
```

In [53]:
```
# number of patients
len(df.patient_id.unique())
```

Out[53]: 2914

In [54]:
```
# number of patients with cancer
df[df["has_cancer"] == True].count()[0]
```

Out[54]: 306

In [55]:
```
# number of patients without cancer
df[df["has_cancer"] == False].count()[0]
```

Out[55]: 2608

Below we have prior probabilities

$$P(c) = 0.105$$

$$P(\ c) = 1 - P(c) = 0.8949$$

In [56]:
```
cancer_cnt = df[df["has_cancer"] == True].count()[0]
non_cancer_cnt = df[df["has_cancer"] == False].count()[0]
total_cnt = df.shape[0]
prob_cancer = cancer_cnt / total_cnt
prob_not_cancer = non_cancer_cnt / total_cnt
```

In [57]:
```
# proportion of patients with cancer
prob_cancer
```

Out[57]: 0.10501029512697323

```
In [58]: # proportion of patients without cancer
         prob_not_cancer

Out[58]: 0.89498970487302676

In [59]: assert prob_cancer + prob_not_cancer == 1
```

Here we will use joint counts and the $P(c)$

$$\frac{P(c \cap pos)}{P(c)} = P(pos|c) = testsensitity$$

$$\frac{P(c \cap neg)}{P(c)} = P(neg|c)$$

So we have derived the test sensitivity from the dataset

```
In [60]: cancer_and_positive_cnt = df[df["has_cancer"] == True] \
              .query("test_result == 'Positive'").count()[0]

         cancer_and_negative_cnt = df[df["has_cancer"] == True] \
              .query("test_result == 'Negative'").count()[0]

         total_with_cancer_cnt = df[df["has_cancer"] == True].count()[0]

         positive_given_cancer_pct = cancer_and_positive_cnt / total_with_cancer_cnt
         negative_given_cancer_pct = cancer_and_negative_cnt / total_with_cancer_cnt

In [61]: # proportion of patients with cancer who test positive
         positive_given_cancer_pct

Out[61]: 0.90522875816993464

In [62]: # proportion of patients with cancer who test positive
         negative_given_cancer_pct

Out[62]: 0.094771241830065356

In [63]: assert positive_given_cancer_pct + negative_given_cancer_pct == 1

In [64]: test_sensitivity = positive_given_cancer_pct
```

Next we will use joint probabilities and $P(\neg c)$:

$$P(\neg c \cap pos)$$

$$P(\neg c \cap neg)$$

Now we can derive the test specifity

$$\frac{P(\neg c \cap neg)}{P(\neg c)} = P(neg|\neg c) = testspecifity$$

$$\frac{P(\neg c \cap pos)}{P(\neg c)} = P(pos|\neg c)$$

```
In [65]: # proportion of patients without cancer who test positive
         false_positive_count = df[df["has_cancer"] == False] \
             .query("test_result == 'Positive'").count()[0]
         total_non_cancer_count = df[df["has_cancer"] == False].count()[0]
         prob_pos_given_not_cancer = false_positive_count / total_non_cancer_count
         prob_pos_given_not_cancer

Out[65]: 0.2036042944785276

In [66]: # proportion of patients without cancer who test negative
         true_negative_count = df[df["has_cancer"] == False] \
             .query("test_result == 'Negative'").count()[0]
         prob_neg_given_not_cancer =  true_negative_count / total_non_cancer_count
         prob_neg_given_not_cancer

Out[66]: 0.79639570552147243

In [67]: assert prob_pos_given_not_cancer + prob_neg_given_not_cancer == 1

In [68]: test_specifity = prob_neg_given_not_cancer
```

Now to calculate bayes theorem to get the posterior probability

$$P(C|pos) = \frac{P(c \cap pos)}{P(pos)}$$

$$P(pos) = P(pos|c)P(c) + P(pos|\neg c)P(\neg c)$$

$$P(neg) = P(neg|c)P(c) + P(neg|\neg c)P(\neg c)$$

$$P(c \cap pos) = P(pos|c)P(c)$$

Notice we have all the numbers to calculate posterior probabilities for all circumstances

```
In [69]: prob_pos = test_sensitivity * prob_cancer + \
                 (1-test_specifity) * prob_not_cancer
         prob_neg = (1-test_sensitivity)*prob_cancer + \
                     test_specifity * prob_not_cancer

In [70]: prob_pos, prob_neg

Out[70]: (0.27728208647906655, 0.7227179135209334)

In [71]: assert(prob_pos + prob_neg == 1)
```

Below we calculate the joint probability and from that we get the posterior probability $ P(c | pos)$

```
In [72]: joint_prob_cancer_and_pos = prob_cancer*test_sensitivity
         cancer_given_pos_test = joint_prob_cancer_and_pos / prob_pos
         cancer_given_pos_test
```

Out[72]: 0.34282178217821785

$$P(\neg c|pos) = \frac{P(\neg c \cap pos)}{P(pos)}$$

$$P(\neg c \cap pos) = P(\neg c)P(pos|\neg c)$$

is below

In [73]:
```
joint_prob_not_cancer_and_pos = prob_not_cancer * (1-test_specifity)
not_cancer_given_pos_test = joint_prob_not_cancer_and_pos / prob_pos
not_cancer_given_pos_test
```

Out[73]: 0.6571782178217821

In [74]:
```
assert not_cancer_given_pos_test + cancer_given_pos_test == 1
```

$$P(c|neg) = \frac{p(c \cap neg)}{P(neg}$$

$$P(c \cap neg) = p(neg|c)P(c)$$

In [82]:
```
prob_cancer_given_neg  = (1-test_sensitivity) *  \
    prob_cancer / prob_neg
prob_cancer_given_neg
```

Out[82]: 0.013770180436847102

$$P(\neg c|neg) = \frac{P(\neg c \cap neg)}{P(neg)}$$

$$P(\neg c \cap neg) = P(neg|\neg c)P(\neg c)$$

In [83]:
```
prob_not_cancer_given_neg = test_specifity * \
    prob_not_cancer / prob_neg
prob_not_cancer_given_neg
```

Out[83]: 0.98622981956315292

In [85]:
```
assert prob_cancer_given_neg + prob_not_cancer_given_neg == 1
```

In [ ]: