Cancer Test Results

In [1]:
```
import pandas as pd

df = pd.read_csv('cancer_test_data.csv')
df.head()
```

Out[1]:

	patient_id	test_result	has_cancer
0	79452	Negative	False
1	81667	Positive	True
2	76297	Negative	False
3	36593	Negative	False
4	53717	Negative	False

In [2]:
```
df.shape
```

Out[2]:
```
(2914, 3)
```

In [3]:
```
# number of patients with cancer
df.has_cancer.sum()
```

Out[3]:
```
306
```

In [4]:
```
# number of patients without cancer
(df.has_cancer == False).sum()
```

Out[4]:
```
2608
```

In [5]:
```
# proportion of patients with cancer
df.has_cancer.mean()
```

Out[5]:
```
0.10501029512697323
```

In [6]:
```
# proportion of patients without cancer
1 - df.has_cancer.mean()
```

Out[6]:
```
0.89498970487302676
```

In [7]:
```
# proportion of patients with cancer who test positive
(df.query('has_cancer')['test_result'] == 'Positive').mean()
```

Out[7]:
```
0.90522875816993464
```

In [8]:
```
# proportion of patients with cancer who test negative
(df.query('has_cancer')['test_result'] == 'Negative').mean()
```

Out[8]:
```
0.094771241830065356
```

In [9]:
```
# proportion of patients without cancer who test positive
(df.query('has_cancer == False')['test_result'] == 'Positive').mean()
```

Out[9]: 0.2036042944785276

In [10]:
```
# proportion of patients without cancer who test negative
(df.query('has_cancer == False')['test_result'] == 'Negative').mean()
```

Out[10]: 0.79639570552147243