# Lecture 22 - Sensitivity, Specificity, and Decisions

Sta 102

June 16, 2016

Colin Rundel & Mine Çetinkaya-Rundel

# Sensitivity and Specificity

## (An old) Example - *House*

If you've ever watched the TV show *House* on Fox, you know that Dr. House regularly states, "It's never lupus."

Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffer from this disease.

The test for lupus is very accurate if the person actually has lupus, however is very inaccurate if the person does not. More specifically, the test is 98% accurate if a person actually has the disease. The test is 74% accurate if a person does not have the disease.

Is Dr. House correct even if someone tests positive for Lupus?

# (An old) Example - *House*


$$
\begin{aligned}
P(\text{Lupus}|+) &= \frac{P(+, \text{Lupus})}{P(+, \text{Lupus}) + P(+, \text{No Lupus})} \\
&= \frac{0.0196}{0.0196 + 0.2548} = 0.0714
\end{aligned}
$$

# Testing for lupus

It turns out that testing for Lupus is actually quite complicated, a diagnosis usually relies on the outcome of multiple tests, often including: a complete blood count, an erythrocyte sedimentation rate, a kidney and liver assessment, a urinalysis, and or an antinuclear antibody (ANA) test.

It is important to think about what is involved in each of these tests (e.g. deciding if complete blood count is high or low) and how each of the individual tests and related decisions plays a role in the overall decision of diagnosing a patient with lupus.

# Testing for lupus (cont.)

At some level we can view a diagnosis as a binary decision (lupus or no lupus) that involves the complex integration of various explanatory variables.

The example does not give us any information about how a diagnosis is made, but what it does give us is just as important - the *sensitivity* and the *specificity* of the test(s). These values are critical for our understanding of what a positive or negative test result actually means.

# Sensitivity and Specificity

*Sensitivity* - measures a tests ability to identify positive results.

$$P(\text{Test } + \mid \text{Conditon } +) = P(+ \mid \text{lupus}) = 0.98$$

*Specificity* - measures a tests ability to identify negative results.

$$P(\text{Test } - \mid \text{Condition } -) = P(- \mid \text{no lupus}) = 0.74$$

# Sensitivity and Specificity

*Sensitivity* - measures a tests ability to identify positive results.

$$P(\text{Test } + \mid \text{Conditon } +) = P(+ \mid \text{lupus}) = 0.98$$

*Specificity* - measures a tests ability to identify negative results.

$$P(\text{Test } - \mid \text{Condition } -) = P(- \mid \text{no lupus}) = 0.74$$

It is illustrative to think about the extreme cases - what is the sensitivity and specificity of a test that always returns a positive result?

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$

$$\textit{Specificity} = P(\text{Test } - \mid \text{Condition } -)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$

$$\textit{Specificity} = P(\text{Test } - \mid \text{Condition } -) = TN/(FP + TN)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$

$$\textit{Specificity} = P(\text{Test } - \mid \text{Condition } -) = TN/(FP + TN)$$

$$\textit{False negative rate } (\beta) = P(\text{Test } - \mid \text{Condition } +)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$

$$\textit{Specificity} = P(\text{Test } - \mid \text{Condition } -) = TN/(FP + TN)$$

$$\textit{False negative rate } (\beta) = P(\text{Test } - \mid \text{Condition } +) = FN/(TP + FN)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$

$$\textit{Specificity} = P(\text{Test } - \mid \text{Condition } -) = TN/(FP + TN)$$

$$\textit{False negative rate } (\beta) = P(\text{Test } - \mid \text{Condition } +) = FN/(TP + FN)$$

$$\textit{False positive rate } (\alpha) = P(\text{Test } + \mid \text{Condition } -)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$

$$\textit{Specificity} = P(\text{Test } - \mid \text{Condition } -) = TN/(FP + TN)$$

$$\textit{False negative rate } (\beta) = P(\text{Test } - \mid \text{Condition } +) = FN/(TP + FN)$$

$$\textit{False positive rate } (\alpha) = P(\text{Test } + \mid \text{Condition } -) = FP/(FP + TN)$$

# Sensitivity and Specificity (cont.)

| | Condition Positive | Condition Negative |
|---|---|---|
| Test Positive | True Positive | False Positive (Type I error) |
| Test Negative | False Negative (Type II error) | True Negative |

$$\textit{Sensitivity} = P(\text{Test } + \mid \text{Condition } +) = TP/(TP + FN)$$
$$\textit{Specificity} = P(\text{Test } - \mid \text{Condition } -) = TN/(FP + TN)$$
$$\textit{False negative rate } (\beta) = P(\text{Test } - \mid \text{Condition } +) = FN/(TP + FN)$$
$$\textit{False positive rate } (\alpha) = P(\text{Test } + \mid \text{Condition } -) = FP/(FP + TN)$$

$$\textit{Sensitivity} = 1 - \textit{False negative rate} = \text{Power}$$
$$\textit{Specificity} = 1 - \textit{False positive rate}$$

# So what?

Clearly it is important to know the Sensitivity and Specificity of a test (and or the false positive and false negative rates). Along with the incidence of the disease, e.g. $P(\text{lupus})$, these values are necessary to calculate important quantities like $P(\text{lupus}|+)$.

Additionally, our foray into power analysis after the first midterm should also give you an idea about the trade offs that are inherent in minimizing false positive and false negative rates (increasing power required either increasing $\alpha$ or $n$).

How do we use this information when we are trying to come up with a decision?

# ROC curves

## Back to Spam

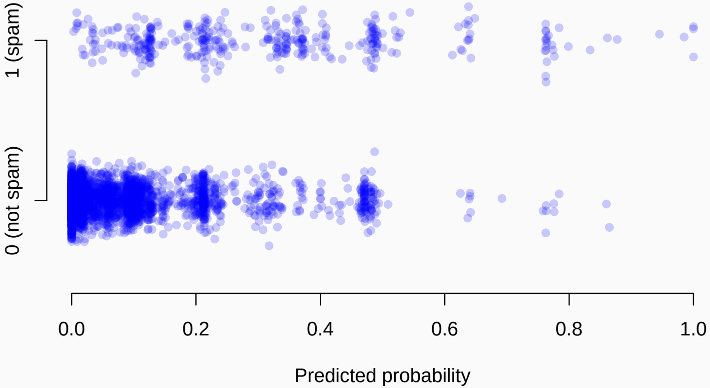
we will now examine a data set of emails where we are interested in identifying spam messages. We will examine several different logistic regression models, however these models only predict the probability an incoming message is spam. If we were designing a spam filter this would only be half of the battle, we also need to design a decision rule about which emails get flagged as spam (e.g. what probability should we use as out cutoff?)

# Back to Spam

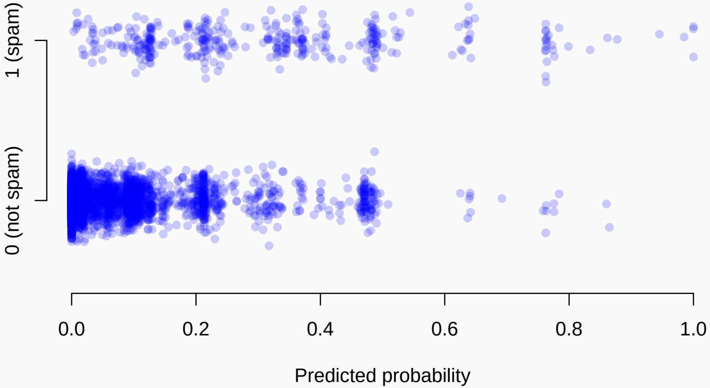
we will now examine a data set of emails where we are interested in identifying spam messages. We will examine several different logistic regression models, however these models only predict the probability an incoming message is spam. If we were designing a spam filter this would only be half of the battle, we also need to design a decision rule about which emails get flagged as spam (e.g. what probability should we use as out cutoff?)

While not the only possible solution, we will consider a simple approach where we choose a single threshold probability and any email that exceeds that probability is flagged as spam.

# Picking a threshold

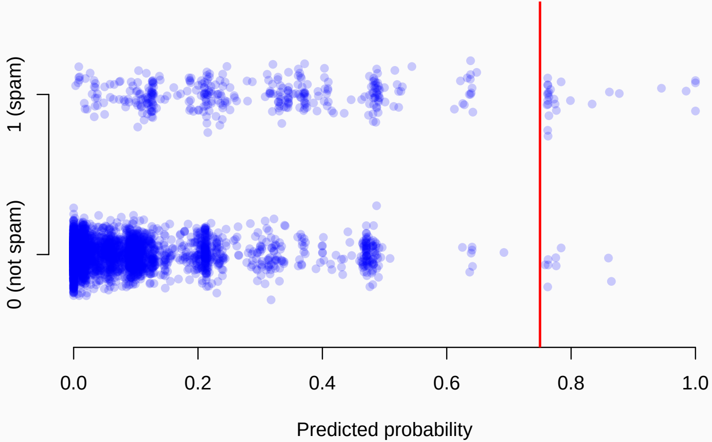


# Picking a threshold



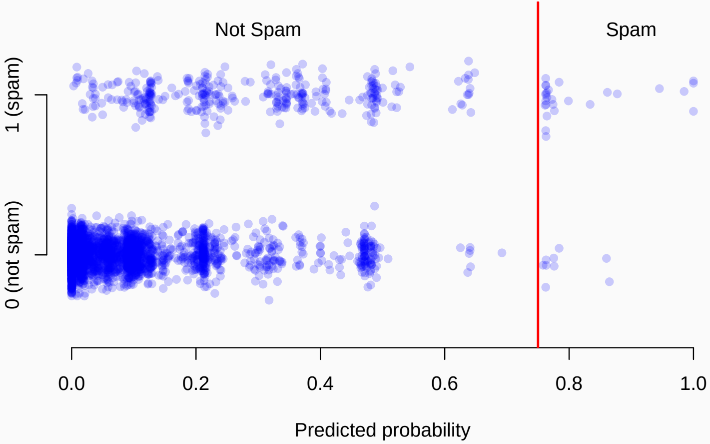

Lets see what happens if we pick our threshold to be *0.75*.

# Picking a threshold



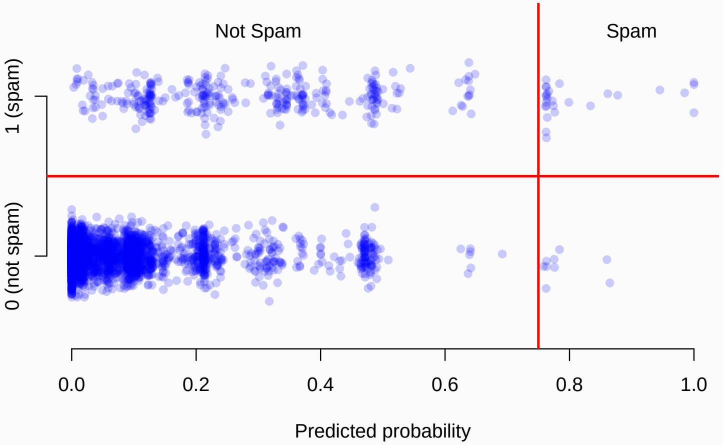

Lets see what happens if we pick our threshold to be *0.75*.

# Picking a threshold



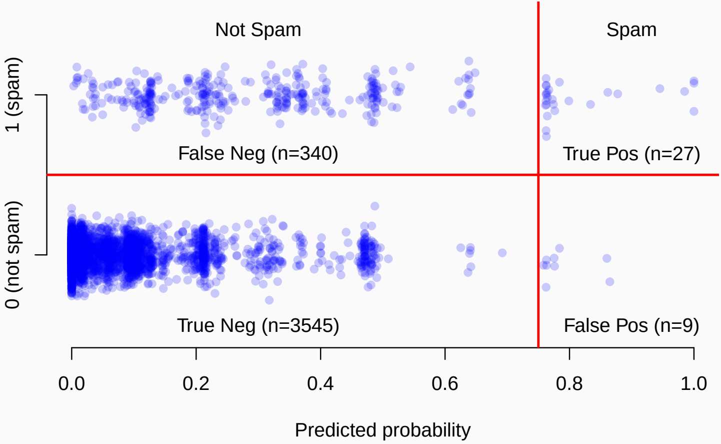

Lets see what happens if we pick our threshold to be *0.75*.

# Picking a threshold




Lets see what happens if we pick our threshold to be *0.75*.

# Picking a threshold




Lets see what happens if we pick our threshold to be *0.75*.

# Consequences of picking a threshold

For our data set picking a threshold of 0.75 gives us the following results:

$$FN = 340 \qquad TP = 27$$
$$TN = 3545 \qquad FP = 9$$

# Consequences of picking a threshold

For our data set picking a threshold of 0.75 gives us the following results:

$$FN = 340 \qquad TP = 27$$
$$TN = 3545 \qquad FP = 9$$

What are the sensitivity and specificity for this particular decision rule?

# Consequences of picking a threshold

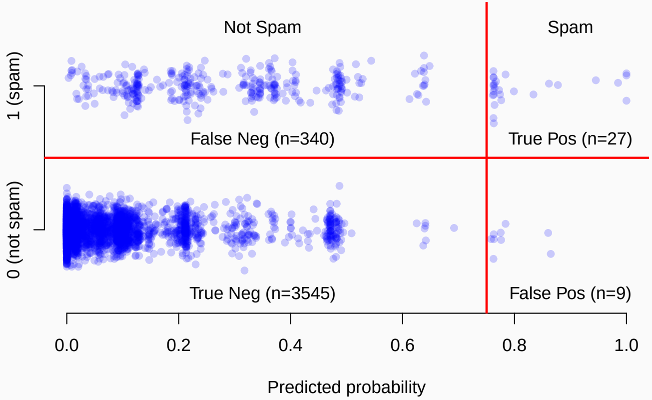
For our data set picking a threshold of 0.75 gives us the following results:

$$FN = 340 \qquad TP = 27$$
$$TN = 3545 \qquad FP = 9$$

What are the sensitivity and specificity for this particular decision rule?

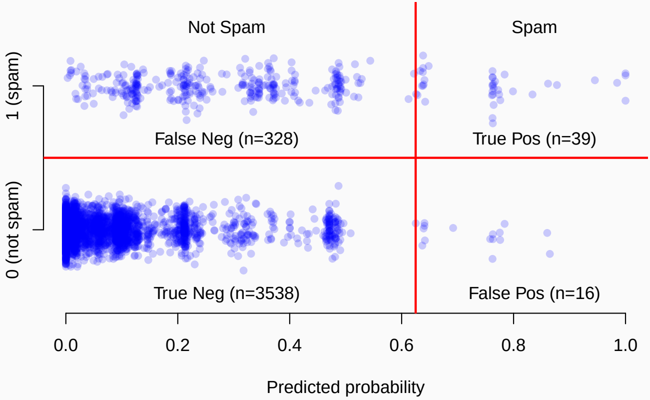
$$Sensitivity = TP/(TP + FN) = 27/(27 + 340) = 0.073$$
$$Specificity = TN/(FP + TN) = 3545/(9 + 3545) = 0.997$$

# Trying other thresholds



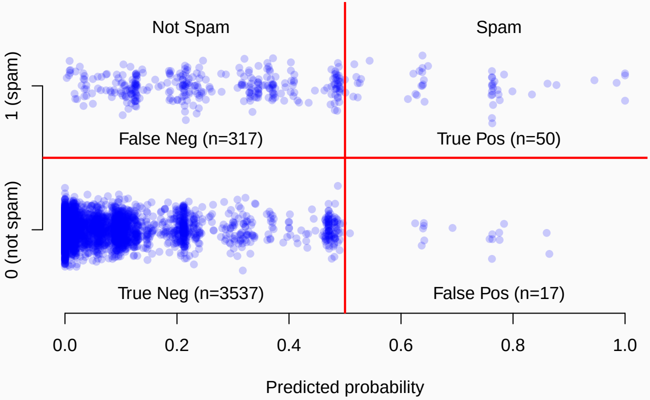

| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | | | | |
| Specificity | 0.997 | | | | |

# Trying other thresholds



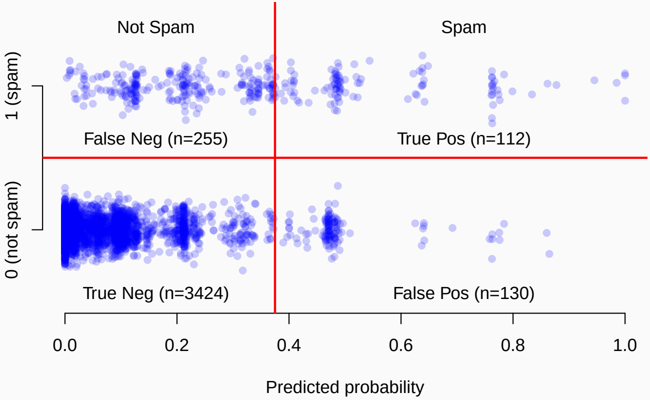

| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | 0.106 | | | |
| Specificity | 0.997 | 0.995 | | | |

# Trying other thresholds



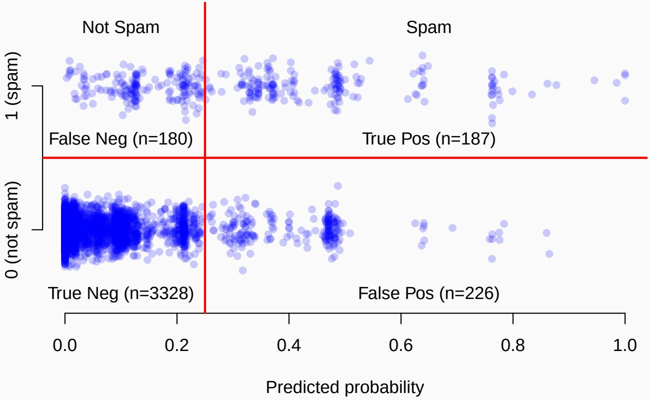

| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | 0.106 | 0.136 | | |
| Specificity | 0.997 | 0.995 | 0.995 | | |

# Trying other thresholds




| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | 0.106 | 0.136 | 0.305 | |
| Specificity | 0.997 | 0.995 | 0.995 | 0.963 | |

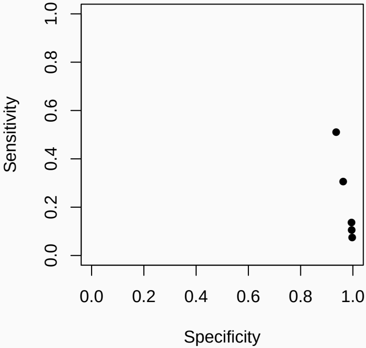
# Trying other thresholds




| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | 0.106 | 0.136 | 0.305 | 0.510 |
| Specificity | 0.997 | 0.995 | 0.995 | 0.963 | 0.936 |

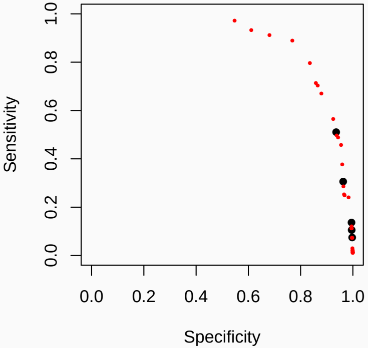
# Relationship between Sensitivity and Specificity

| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | 0.106 | 0.136 | 0.305 | 0.510 |
| Specificity | 0.997 | 0.995 | 0.995 | 0.963 | 0.936 |



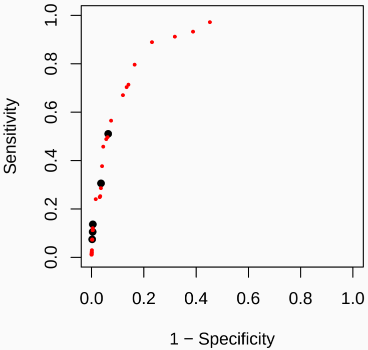
# Relationship between Sensitivity and Specificity

| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | 0.106 | 0.136 | 0.305 | 0.510 |
| Specificity | 0.997 | 0.995 | 0.995 | 0.963 | 0.936 |

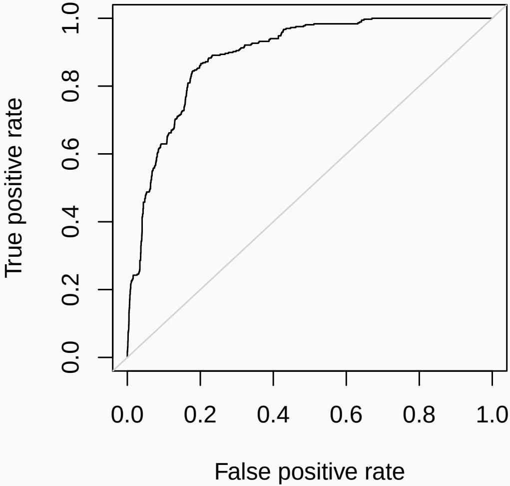


# Relationship between Sensitivity and Specificity

| Threshold | 0.75 | 0.625 | 0.5 | 0.375 | 0.25 |
|---|---|---|---|---|---|
| Sensitivity | 0.074 | 0.106 | 0.136 | 0.305 | 0.510 |
| Specificity | 0.997 | 0.995 | 0.995 | 0.963 | 0.936 |



# Receiver operating characteristic (ROC) curve



# Receiver operating characteristic (ROC) curve (cont.)

Why do we care about ROC curves?

- Shows the trade off in sensitivity and specificity for all possible thresholds.
- Straight forward to compare performance vs. chance.
- Can use the area under the curve (AUC) as an assessment of the predictive ability of a model.

# Refining the Spam model

```
refined = glm(spam ~ to_multiple+cc+image+attach+winner
                    +password+line_breaks+format+re_subj
                    +urgent_subj+exclaim_mess,
              data=email, family=binomial)
summary(refined)
```

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---:|---:|---:|---:|---:|
| (Intercept) | -1.7594 | 0.1177 | -14.94 | 0.0000 |
| to_multipleyes | -2.7368 | 0.3156 | -8.67 | 0.0000 |
| ccyes | -0.5358 | 0.3143 | -1.71 | 0.0882 |
| imageyes | -1.8585 | 0.7701 | -2.41 | 0.0158 |
| attachyes | 1.2002 | 0.2391 | 5.02 | 0.0000 |
| winneryes | 2.0433 | 0.3528 | 5.79 | 0.0000 |
| passwordyes | -1.5618 | 0.5354 | -2.92 | 0.0035 |
| line_breaks | -0.0031 | 0.0005 | -6.33 | 0.0000 |
| formatPlain | 1.0130 | 0.1380 | 7.34 | 0.0000 |
| re_subjyes | -2.9935 | 0.3778 | -7.92 | 0.0000 |
| urgent_subjyes | 3.8830 | 1.0054 | 3.86 | 0.0001 |
| exclaim_mess | 0.0093 | 0.0016 | 5.71 | 0.0000 |

# Comparing models