

Chapter 6: Measuring Test Accuracy

Investigation: PCR tests are very good at detecting any trace amount of SARS-CoV-2, but antigen tests may be better at detecting whether someone is still **contagious**. In 2021, researchers at the University of Illinois gathered daily antigen test results and swabs from 60 people who had known exposure to someone with SARS-CoV-2. They gathered about 8-15 days of test results from each participant for a total of **785 days** of test results. Here is a link to the [Article in Nature](#).

The researchers had each participant complete a **rapid antigen test**, and they also swabbed their nose and mouth to provide a sample to be monitored on a petri dish. The petri dish was observed over 4 days to see if there was any growth or not. Growth would mean a live virus that suggested participant may be contagious, and no growth would mean unlikely to be contagious.



Table 1. Antigen test results and actual viral status

	Live Virus (Contagious)	Non-Live Virus (Not Contagious)	Totals
Negative Antigen test	14 (False Negative)	425 (True Negative)	439
Positive Antigen test	181 (True Positive)	165 (False Positive)	346
Totals	195	590	785

- Let's start by thinking about probabilities for _____ events involving one grouping at a time.
 - Let's define "Contagious" be the event that someone truly has a live strain of SARS-CoV-2.
 - Let "+" be the event that someone gets a positive test result.

What proportion of these viral samples produced a positive test?

What proportion of viral samples were live (suggesting participant is contagious)?

$P(+)$ =

- Now, let's consider probabilities involving _____ events.
 - \cup **stands for "union."** $A \cup B$ would be every possibility where at least one of these two events A and B are true.
 - \cap **stands for "intersection,"** $A \cap B$ would be every possibility where BOTH A and B are true.

What proportion of these viral samples either produced a positive test result, or were found to be live, or both?

What proportion of these viral samples produced a positive test result and were found to be live?

$P(\text{Contagious} \cup +)$ =

- Finally, let's consider this investigation if we examine _____ probabilities.
 - This would be the probability of one event given known information about another event.
 - Notation might be something like $P(B|A)$, which would be read...
 "What is the probability of B _____ that A is true?"
 - We can represent this in a formula. $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{(\text{Probability of both A and B happening})}{(\text{Probability of A happening})}$

Test Accuracy Measures

In the context of test effectiveness, there are 4 different conditional probabilities we might examine:

- Positive Predictive Value (PPV):** What is the probability that I have this condition **given** my test result is positive?
 - Percentage of positive (+) test results that come from truly infected people.
- Negative Predictive Value (NPV):** What is the probability that I **do not** have this condition **given** my test result is negative?
 - Percentage of (-) test results that come from truly non-infected people.
- Sensitivity:** What is the probability that my test comes back positive **given** that I'm infected?
 - Percentage of infected people for whom the test will correctly indicate positive.
- Specificity:** What is the probability that my test comes back negative **given** that I'm non-infected?
 - Percentage of non-infected people for whom the test will correctly indicate negative.

Practice: For each question, identify which measure is being asked for and estimate it from the data

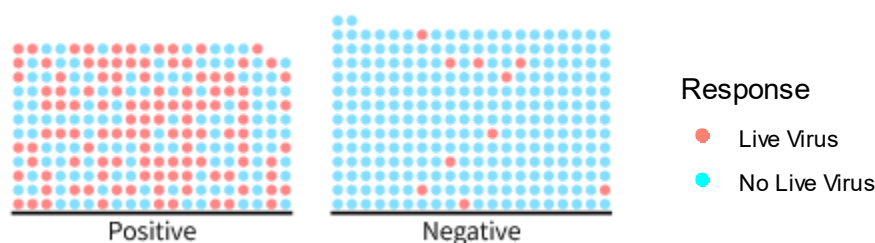
What proportion of contagious viral samples correctly came back positive?

What proportion of non-contagious samples correctly came back negative?

$P(+ | \text{contagious}) =$

Visualizing Test Accuracy

- What percentage of the positive test results are correctly **red** (dark)? _____
- What percentage of the negative test results are correctly **blue** (light)? _____
- What percentage of the live virus samples (**red**) were identified as positive? _____
- What percentage of the non-live virus samples (**blue**) were identified as negative? _____



Confidence Intervals for Test Accuracy

Each measure is technically a proportion. Therefore, we can create a confidence interval for each proportion following the same method we've used before, where n is the number of cases in our **conditional** group.

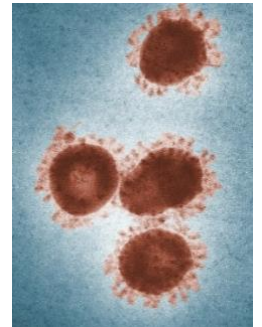
$$\text{Agresti-Coull interval for } \pi: \quad \tilde{p} \pm Z_{C\%} * SE_{\tilde{p}} \quad \text{where } SE_{\tilde{p}} = \frac{\sqrt{\tilde{p}(1-\tilde{p})}}{\sqrt{n+4}}$$

What would be our 95% margin of error for the sensitivity of the antigen test in detecting a live virus?

$$\text{Point estimate} = \tilde{p} = \frac{181+2}{195+4} = 0.9196$$

$$SE = \frac{\sqrt{0.9196(1-0.9196)}}{\sqrt{195+4}} = 0.0193$$

$$\text{Margin of error} = 1.96 * 0.0193 = \mathbf{0.0378}$$



Theoretical Models

- In the previous example, we calculated different probabilities based on estimates from data. But it's also helpful to think about these measures when building theoretical models where we make assumptions about the probabilities and notice how those measures change.
- ...But let's start with a non-technical example first!

Example: At an ice cream stand, customers have the choice of paying extra for a cone, or simply choosing a bowl. Furthermore, they have the choice of purchasing two scoops or just one scoop. [Probability Simulation](#) (or google search "Setosa conditional")



- Let's say that 40% of all customers pay extra for a cone (we'll call this event A)
- Let's say that 55% of all customers pay extra for 2 scoops (we'll call this event B)
- And Let's say that 15% of all customers pay extra for both.

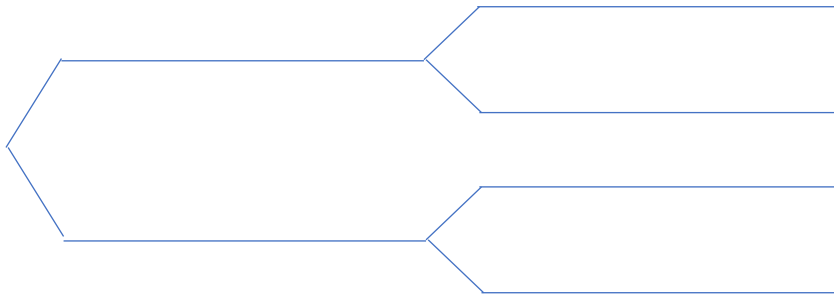
In context, what outcome(s) would be part of $A \cup B$? What color(s) are represented by that event?

In context, what outcome(s) would be part of $A \cap B$? What color(s) are represented by that event?

Given that a customer chooses to purchase a cone, what is the probability that they will also choose 2 scoops?

$$P(2 \text{ scoops} \mid \text{cone}) =$$

- **Tree Diagrams** are a great way to visualize compound events when wanting to identify conditional probabilities.
 - The first set of stems represent a simple event and its complement, and each of these probabilities should add up to ____.
 - The second set of stems represent conditional probabilities given the stem they are linked to. Each **conditional grouping** should also add up to ____.
 - *To save some extra calculation, the $P(2 \text{ scoops} \mid \text{bowl}) = 0.667$*



Now, that we've identified the various conditional probabilities, let's find the intersection probabilities and place those to the right. We can solve for those by rearranging the conditional probability formula!

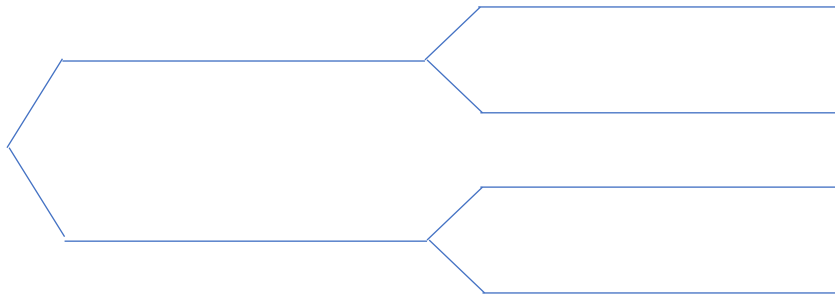
$$P(B|A) = \frac{P(A \cap B)}{P(A)} \text{ so ... } P(A \cap B) = \underline{\hspace{2cm}}$$

- **Mosaic Plot** (*if time*). A great to visualize intersections (and why we find them by multiplying as we did above) is a mosaic plot. We can show all four possible intersections, and we can represent those probabilities proportionally in the area below.

	2 scoops	1 scoop
Cone	$0.4(0.375) =$ 15%	$0.4(\underline{\hspace{1cm}}) =$ ___%
Bowl	$0.6(\underline{\hspace{1cm}}) =$ ___%	$0.6(\underline{\hspace{1cm}}) =$ ___%

Investigation revisited: Let's return to the antigen test again and think about what results we might see with the following assumed information:

- Among people **with** a live virus culture, **95%** of them will report a **positive** antigen test.
- Among people **without** a live virus culture, **20%** of them will incorrectly report a **positive** antigen test.
- Let's assume that the **true rate** of cultures tested that contained live virus strains was **25%**.
- Let's finally assume that there are no inconclusive results (all test results are identified as positive or negative).



What is the theoretical **specificity** of this test?

How **accurate** is the antigen test in detecting a live virus? (*hint: consider the two types of correct results that can occur. What do those probabilities sum to?*)

What is the **positive predictive value** of this test?

[Seeing Theory Simulation](#) (or google search “Seeing Theory” and go to the “Bayesian” section)

The example we used involved a relatively high incidence testing population (people who were exposed to the virus). What if we were instead looking at test accuracy when our tested population had a relatively low incidence. Would PPV and NPV change? Would Sensitivity or specificity change based on incidence rate?

Reflection Questions

- 6.1.** Let's say you've developed a diagnostic test to indicate whether someone has the Shingles virus. Describe what true positive, true negative, false positive, and false negative outcomes would represent.
- 6.2.** In probability, what is the difference between the *union* of two events and the *intersection* of two events?
- 6.3.** How would we read the symbols $P(B|A)$? In particular, what does the $|$ symbol mean?
- 6.4.** What is the difference between sensitivity and positive predictive value (PPV)? Try describing them in the context of testing for the presence of the shingles virus.
- 6.5.** Consider this question: "What is that probability that a chosen individual from our testing population both has the outcome of interest and tests positive?" Would this fit any of the four test accuracy measures described above? If so, which one? If not, why not?
- 6.6.** Consider if you were testing for measles. The incidence rate in your region has gone up, and you are still using the same test to test for it. Would you expect the sensitivity of your test to increase? Would you expect the PPV to increase? Would you expect the NPV to increase?

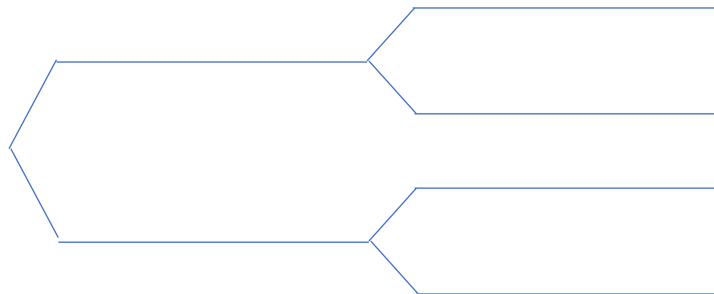
Chapter 6 Additional Practice (if you need it!)

Practice: A test is used to screen for the presence of a particular drug in one's blood system.

- Let's assume for this example that 8% of the people we are screening truly have drugs in their system.
- Let's also assume that this drug test will produce a negative for 12% of users who truly have drugs in their system.
- Let's also assume it will produce a negative for 96% of users who truly don't have drugs in their system.

Use this information to fill in the tree diagram below.

What is the test's sensitivity?



What is the test's negative predictive value *(if we assume an 8% incidence rate)*?

Now, let's assume we didn't know the true sensitivity or specificity of this test. We gather a sample of 500 people, of which 8% of them are known to have used drugs in the past 24 hours. We see the following results.

	Positive Test Result	Negative Test Result	Totals
Subject used drugs	31	9	40
Subject did not use drugs	24	436	460
Totals	55	445	500

Based on the sample data provided, what would be our estimate for specificity? And what would be the 95% margin of error on our estimate?

If the true proportion of test-takers who used drugs were to go down from 8%, would you expect the positive predictive value to get better or get worse?

