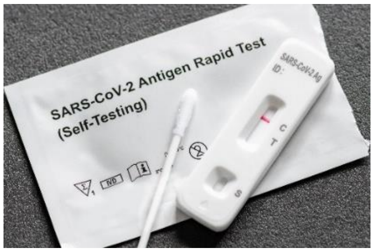
Chapter 6: Measuring Test Accuracy

Investigation: Different tests for the SARS-CoV-2 virus (the virus that leads to COVID-19) indicate presence of the virus in different ways. One particularly helpful metric is the accuracy of these tests to indicate whether someone is **contagious**. In 2021, researchers at the University of Illinois was able to get daily test results and viral samples 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, a nasal swab PCR test, and a saliva sample PCR test. Let's focus on the accuracy of the **rapid antigen test**. The results of the antigen test are shown below, along with the researchers' examination of the viral sample, which was observed over 4 days to see if there was any growth or not. Growth would mean a live virus that could be transmitted, and no growth would mean no threat of transmission.

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?

P(+) =

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

- Now, let's consider probabilities involving ________________ events.
  - ∪ **stands for "union."** A ∪ B would be every possibility where at least one of these two events A and B are true.
  - ∩ **stands for "intersection,"** A ∩ 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?

P(Contagious ∪ +) =

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

- 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):** Percentage of (+) tests that correctly indicate condition of interest.
  - Probability of a true result **given** the test result is _____________.
- **Negative Predictive Value (NPV):** Percentage of (-) tests that correctly indicate condition of interest.
  - Probability of a true result **given** the test result is _____________.
- **Sensitivity:** Percentage of people who have the "condition" who correctly test positive
  - Probability of a true result **given** the person _____ condition of interest.
- **Specificity:** Percentage of people who don't have the "condition" who correctly test negative
  - Probability of a true result **given** the person ____________ have the condition of interest.

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

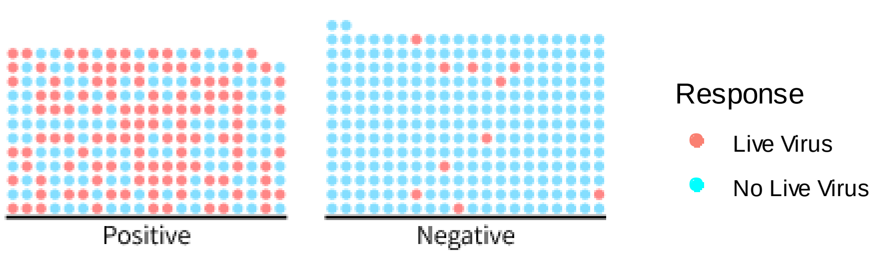
What proportion of contagious viral samples correctly came back positive?

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

P(+ | contagious) =

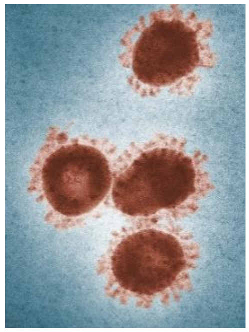
Visualizing Test Accuracy

- 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.



Z-Interval for π: $\hat{p} \pm Z_{C\%} * SE_{\hat{p}}$ where $SE_{\hat{p}} \approx \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$

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

Point estimate:

SE:

Margin of error:

Remember that assumptions still apply for z-intervals here. In particular…

✓ We are assuming that the distribution of possible sample proportions is normally distributed. **The 10/10 Rule**: This is a reasonable assumption to make if our sample has **at least 10 of each response (e.g., >=10 "yes" and >=10 "no" responses)**

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.

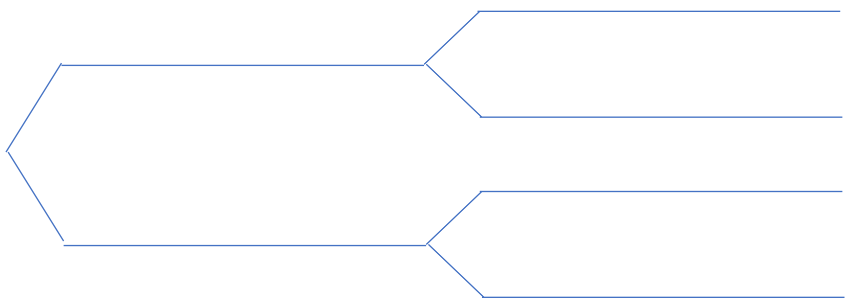
In context, what outcome(s) would be part of A ∪ B? What color(s) are represented by that event?

In context, what outcome(s) would be part of A ∩ 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?

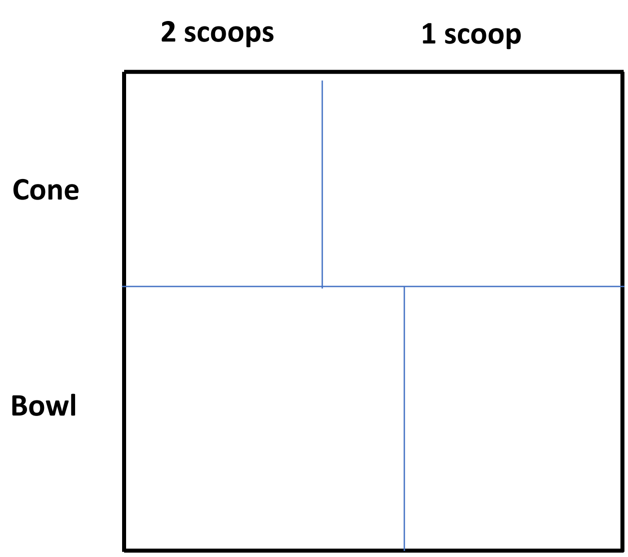
P(2 scoops | cone) =

- **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 **grouping** should represent possible conditions and should also add up to ___.
  - *To save some extra calculation, the P(2 scoops | bowl) = 0.583*



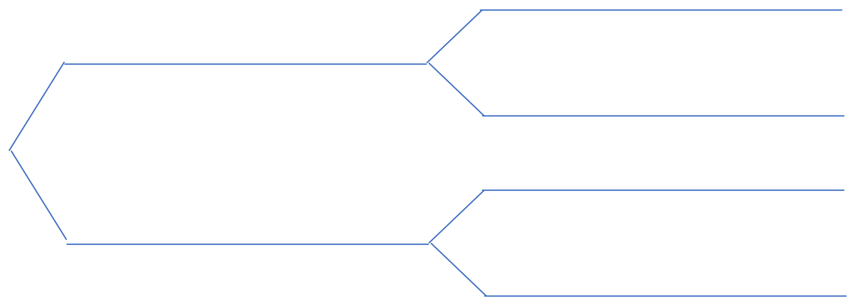
- Now, multiply down each branch to find the intersection probabilities.

- **Mosaic Plot** *(if time).* Everyone can be categorized as being in exactly one of four possible intersections, and we can represent those probabilities proportionally in the area below.



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?

Chapter 6 Reflection Questions

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 look like

In probability, what is the difference between the *union* of two events and the *intersection* of two events?

How would we read the symbols P(B|A)? In particular, what does the | symbol mean?

Can you describe what the conditional probability formula is calculating and make sense of it?

How would you describe what sensitivity, specificity, positive predictive value, and negative predictive value each represent? Could you do it in the context of testing for the presence of the shingles virus?

Consider this question: “What is that probability that a chosen individual from our testing population both has the outcome of interest and tests positive?” Why would this **not** be any of the four test accuracy measures described above?

In a tree diagram, which probabilities would we describe as conditional probabilities?

How would the positive and negative predictive value change if the incident rate of those being tested grows *higher*?

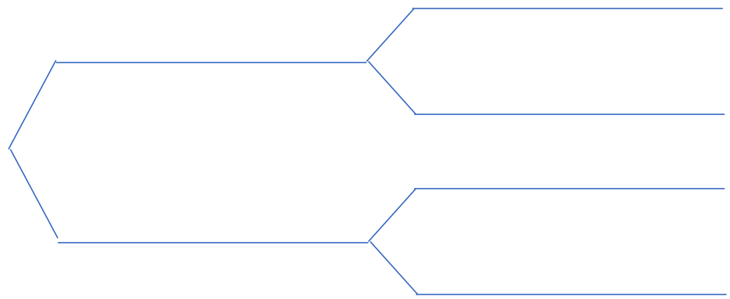
Chapter 6 Additional Practice (if you need it!)

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 specificity?

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?