Selection: 1

| Please choose a lesson, or type 0 to return to course menu.

1: Introduction 2: Probability1 3: Probability2

4: ConditionalProbability 5: Expectations 6: Variance

7: CommonDistros 8: Asymptotics 9: T Confidence Intervals

10: Hypothesis Testing 11: P Values 12: Power

13: Multiple Testing 14: Resampling

Selection: 4

| | | 0%

| Conditional Probability. (Slides for this and other Data Science courses may be found

| at github https://github.com/DataScienceSpecialization/courses/. If you care to use

| them, they must be downloaded as a zip file and viewed locally. This lesson

| corresponds to 06\_Statistical\_Inference/03\_Conditional\_Probability.)

...

| |== | 2%

| In this lesson, as the name suggests, we'll discuss conditional probability.

...

| |==== | 5%

| If you were given a fair die and asked what the probability of rolling a 3 is, what

| would you reply?

1: 1/4

2: 1

3: 1/2

4: 1/3

5: 1/6

Selection: 5

| That's the answer I was looking for.

| |====== | 7%

| Suppose the person who gave you the dice rolled it behind your back and told you the

| roll was odd. Now what is the probability that the roll was a 3?

1: 1/2

2: 1/3

3: 1/4

4: 1

5: 1/6

Selection: 2

| You are amazing!

| |======= | 10%

| The probability of this second event is conditional on this new information, so the

| probability of rolling a 3 is now one third.

...

| |========= | 12%

| We represent the conditional probability of an event A given that B has occurred with

| the notation P(A|B). More specifically, we define the conditional probability of

| event A, given that B has occurred with the following.

...

| |=========== | 14%

| P(A|B) = P(A & B)/ P(B) . P(A|B) is the probability that BOTH A and B occur divided

| by the probability that B occurs.

...

| |============= | 17%

| Back to our dice example. Which of the following expressions represents P(A&B), where

| A is the event of rolling a 3 and B is the event of the roll being odd?

1: 1/3

2: 1/2

3: 1/4

4: 1

5: 1/6

Selection: 5

| You are amazing!

| |=============== | 19%

| Continuing with the same dice example. Which of the following expressions represents

| P(A&B)/P(B), where A is the event of rolling a 3 and B is the event of the roll being

| odd?

1: (1/3)/(1/2)

2: (1/6)/(1/2)

3: 1/6

4: (1/2)/(1/6)

Selection: 2

| Keep up the great work!

| |================= | 21%

| From the definition of P(A|B), we can write P(A&B) = P(A|B) \* P(B), right? Let's use

| this to express P(B|A).

...

| |=================== | 24%

| P(B|A) = P(B&A)/P(A) = P(A|B) \* P(B)/P(A). This is a simple form of Bayes' Rule which

| relates the two conditional probabilities.

...

| |==================== | 26%

| Suppose we don't know P(A) itself, but only know its conditional probabilities, that

| is, the probability that it occurs if B occurs and the probability that it occurs if

| B doesn't occur. These are P(A|B) and P(A|~B), respectively. We use ~B to represent

| 'not B' or 'B complement'.

...

| |====================== | 29%

| We can then express P(A) = P(A|B) \* P(B) + P(A|~B) \* P(~B) and substitute this is

| into the denominator of Bayes' Formula.

...

| |======================== | 31%

| P(B|A) = P(A|B) \* P(B) / ( P(A|B) \* P(B) + P(A|~B) \* P(~B) )

...

| |========================== | 33%

| Bayes' Rule has applicability to medical diagnostic tests. We'll now discuss the

| example of the HIV test from the slides.

...

| |============================ | 36%

| Suppose we know the accuracy rates of the test for both the positive case (positive

| result when the patient has HIV) and negative (negative test result when the patient

| doesn't have HIV). These are referred to as test sensitivity and specificity,

| respectively.

...

| |============================== | 38%

| Let 'D' be the event that the patient has HIV, and let '+' indicate a positive test

| result and '-' a negative. What information do we know? Recall that we know the

| accuracy rates of the HIV test.

1: P(+|~D) and P(-|D)

2: P(+|D) and P(-|~D)

3: P(+|~D) and P(-|~D)

4: P(+|D) and P(-|D)

Selection: 2

| You are quite good my friend!

| |================================ | 40%

| Suppose a person gets a positive test result and comes from a population with a HIV

| prevalence rate of .001. We'd like to know the probability that he really has HIV.

| Which of the following represents this?

1: P(D|+)

2: P(+|D)

3: P(~D|+)

4: P(D|-)

Selection: 1

| Keep up the great work!

| |================================= | 43%

| By Bayes' Formula, P(D|+) = P(+|D) \* P(D) / ( P(+|D) \* P(D) + P(+|~D) \* P(~D) )

...

| |=================================== | 45%

| We can use the prevalence of HIV in the patient's population as the value for P(D).

| Note that since P(~D)=1-P(D) and P(+|~D) = 1-P(-|~D) we can calculate P(D|+). In

| other words, we know values for all the terms on the right side of the equation.

| Let's do it!

...

| |===================================== | 48%

| Disease prevalence is .001. Test sensitivity (+ result with disease) is 99.7% and

| specificity (- result without disease) is 98.5%. First compute the numerator,

| P(+|D)\*P(D). (This is also part of the denominator.)

> .997\*.001

[1] 0.000997

| That's correct!

| |======================================= | 50%

| Now solve for the remainder of the denominator, P(+|~D)\*P(~D).

> (1-.985)\*(1-.000)

[1] 0.015

| Not quite right, but keep trying. Or, type info() for more options.

| Multiply the complement of test specificity by the complement of prevalence.

> (1-.985)\*(1-.001)

[1] 0.014985

| You got it right!

| |========================================= | 52%

| Now put the pieces together to compute the probability that the patient has the

| disease given his positive test result, P(D|+). Plug your last two answers into the

| formula P(+|D) \* P(D) / ( P(+|D) \* P(D) + P(+|~D) \* P(~D) ) to compute P(D|+).

> (.997\*.001)/((.997\*.001) + (1-.985)\*(1-.001))

[1] 0.06238268

| You are quite good my friend!

| |=========================================== | 55%

| So the patient has a 6% chance of having HIV given this positive test result. The

| expression P(D|+) is called the positive predictive value. Similarly, P(~D|-), is

| called the negative predictive value, the probability that a patient does not have

| the disease given a negative test result.

...

| |============================================= | 57%

| The diagnostic likelihood ratio of a positive test, DLR\_+, is the ratio of the two +

| conditional probabilities, one given the presence of disease and the other given the

| absence. Specifically, DLR\_+ = P(+|D) / P(+|~D). Similarly, the DLR\_- is defined as a

| ratio. Which of the following do you think represents the DLR\_-?

1: P(+|~D) / P(-|D)

2: I haven't a clue.

3: P(-|D) / P(+|~D)

4: P(-|D) / P(-|~D)

Selection: 4

| All that practice is paying off!

| |============================================== | 60%

| Recall that P(+|D) and P(-|~D), (test sensitivity and specificity respectively) are

| accuracy rates of a diagnostic test for the two possible results. They should be

| close to 1 because no one would take an inaccurate test, right? Since DLR\_+ = P(+|D)

| / P(+|~D) we recognize the numerator as test sensitivity and the denominator as the

| complement of test specificity.

...

| |================================================ | 62%

| Since the numerator is close to 1 and the denominator is close to 0 do you expect

| DLR\_+ to be large or small?

1: Small

2: Large

3: I haven't a clue.

Selection: 2

| That's the answer I was looking for.

| |================================================== | 64%

| Now recall that DLR\_- = P(-|D) / P(-|~D). Here the numerator is the complement of

| sensitivity and the denominator is specificity. From the arithmetic and what you know

| about accuracy tests, do you expect DLR\_- to be large or small?

1: Small

2: Large

3: I haven't a clue.

Selection: 1

| That's the answer I was looking for.

| |==================================================== | 67%

| Now a little more about likelihood ratios. Recall Bayes Formula. P(D|+) = P(+|D) \*

| P(D) / ( P(+|D) \* P(D) + P(+|~D) \* P(~D) ) and notice that if we replace all

| occurrences of 'D' with '~D', the denominator doesn't change. This means that if we

| formed a ratio of P(D|+) to P(~D|+) we'd get a much simpler expression (since the

| complicated denominators would cancel each other out). Like this....

...

| |====================================================== | 69%

| P(D|+) / P(~D|+) = P(+|D) \* P(D) / (P(+|~D) \* P(~D)) = P(+|D)/P(+|~D) \* P(D)/P(~D).

...

| |======================================================== | 71%

| The left side of the equation represents the post-test odds of disease given a

| positive test result. The equation says that the post-test odds of disease equals the

| pre-test odds of disease (that is, P(D)/P(~D) ) times

1: the DLR\_-

2: I haven't a clue.

3: the DLR\_+

Selection: 3

| Excellent work!

| |========================================================== | 74%

| In other words, a DLR\_+ value equal to N indicates that the hypothesis of disease is

| N times more supported by the data than the hypothesis of no disease.

...

| |=========================================================== | 76%

| Taking the formula above and replacing the '+' signs with '-' yields a formula with

| the DLR\_-. Specifically, P(D|-) / P(~D|-) = P(-|D) / P(-|~D) \* P(D)/P(~D). As with

| the positive case, this relates the odds of disease post-test, P(D|-) / P(~D|-), to

| those of disease pre-test, P(D)/P(~D).

...

| |============================================================= | 79%

| The equation P(D|-) / P(~D|-) = P(-|D) / P(-|~D) \* P(D)/P(~D) says what about the

| post-test odds of disease relative to the pre-test odds of disease given negative

| test results?

1: I haven't a clue.

2: post-test odds are less than pre-test odds

3: post-test odds are greater than pre-test odds

Selection: 2

| All that practice is paying off!

| |=============================================================== | 81%

| Let's cover some basics now.

...

| |================================================================= | 83%

| Two events, A and B, are independent if they have no effect on each other. Formally,

| P(A&B) = P(A)\*P(B). It's easy to see that if A and B are independent, then

| P(A|B)=P(A). The definition is similar for random variables X and Y.

...

| |=================================================================== | 86%

| We've seen examples of independence in our previous probability lessons. Let's review

| a little. What's the probability of rolling a '6' twice in a row using a fair die?

1: 1/6

2: 1/2

3: 1/36

4: 2/6

Selection: 3

| Great job!

| |===================================================================== | 88%

| You're given a fair die and asked to roll it twice. What's the probability that the

| second roll of the die matches the first?

1: 2/6

2: 1/2

3: 1/36

4: 1/6

Selection: 4

| You're the best!

| |======================================================================= | 90%

| If the chance of developing a disease with a genetic or environmental component is p,

| is the chance of both you and your sibling developing that disease p\*p?

1: No

2: Yes

Selection: 2

| One more time. You can do it!

| The events aren't independent since genetic or environmental factors likely will

| affect the outcome.

1: Yes

2: No

Selection: 1

| Not quite, but you're learning! Try again.

| The events aren't independent since genetic or environmental factors likely will

| affect the outcome.

1: Yes

2: No

Selection: 2

| You're the best!

| |======================================================================== | 93%

| We'll conclude with iid. Random variables are said to be iid if they are independent

| and identically distributed. By independent we mean "statistically unrelated from one

| another". Identically distributed means that "all have been drawn from the same

| population distribution".

...

| |========================================================================== | 95%

| Random variables which are iid are the default model for random samples and many of

| the important theories of statistics assume that variables are iid. We'll usually

| assume our samples are random and variables are iid.

...

| |============================================================================ | 98%

| Congrats! You've concluded this lesson on conditional probability. We hope you liked

| it unconditionally.

...

| |==============================================================================| 100%

| Would you like to receive credit for completing this course on Coursera.org?

1: Yes

2: No

Selection: 1

What is your email address? sweeyean@gmail.com

What is your assignment token? NMBuE5aiZTceP1Pu

Grade submission succeeded!

| Keep working like that and you'll get there!

| You've reached the end of this lesson! Returning to the main menu...

| Please choose a course, or type 0 to exit swirl.

1: Statistical Inference

2: Take me to the swirl course repository!

Selection: