```yaml
    Course: Statistical_Inference
    Lesson: ConditionalProbability


  - Class: text
    Output: "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.)"

  - Class: text
    Output: In this lesson, as the name suggests, we'll discuss conditional probability.

  - Class: mult_question
    Output: If you were given a fair die and asked what the probability of rolling a 3 is, what would you reply?
    AnswerChoices: 1/6; 1/2; 1/3; 1/4; 1
    CorrectAnswer: 1/6
    AnswerTests: omnitest(correctVal='1/6')
    Hint: There are 6 possible outcomes and you want to know the probability of 1 of them.

  - Class: mult_question
    Output: 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?
    AnswerChoices: 1/6; 1/2; 1/3; 1/4; 1
    CorrectAnswer: 1/3
    AnswerTests: omnitest(correctVal='1/3')
    Hint: Given that there are 3 odd numbers on the die your possibilities have been reduced to 3 and you want to know the probability of 1 of them.

  - Class: text
    Output: The probability of this second event is conditional on this new information, so the probability of rolling a 3 is now one third.

  - Class: text
    Output: 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.

  - Class: text
    Output:   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.

  - Class: mult_question
    Output: 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?
    AnswerChoices: 1/6; 1/2; 1/3; 1/4; 1
    CorrectAnswer: 1/6
    AnswerTests: omnitest(correctVal='1/6')
    Hint: Here A is a subset of B so the probability of both A AND B happening is the probability of A happening.

  - Class: mult_question
    Output: 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?
    AnswerChoices: (1/6)/(1/2); (1/2)/(1/6); (1/3)/(1/2); 1/6
    CorrectAnswer: (1/6)/(1/2)
    AnswerTests: omnitest(correctVal='(1/6)/(1/2)')
    Hint: Here A is a subset of B so the probability of both A AND B happening is the probability of A happening. The probability of B is the reciprocal of the number of odd numbers between 1 and 6 (inclusive).

  - Class: text
    Output: 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).

  - Class: text
    Output: P(B|A) = P(B&A)/P(A) = P(A|B) * P(B)/P(A). This is a simple form of Bayes'
```

```yaml
    Rule which relates the two conditional probabilities.

  - Class: text
    Output: 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'.

  - Class: text
    Output: 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.

  - Class: text
    Output: P(B|A) = P(A|B) * P(B) / ( P(A|B) * P(B) + P(A|~B) * P(~B) )

  - Class: text
    Output: Bayes' Rule has applicability to medical diagnostic tests. We'll now discuss the example of the HIV test from the slides.

  - Class: text
    Output: 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.

  - Class: mult_question
    Output: 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.
    AnswerChoices: P(+|D) and P(-|~D); P(+|~D) and P(-|~D); P(+|~D) and P(-|D); P(+|D) and P(-|D)
    CorrectAnswer: P(+|D) and P(-|~D)
    AnswerTests: omnitest(correctVal='P(+|D) and P(-|~D)')
    Hint: The clue here is accuracy. The test is positive when the patient has the disease and negative when he doesn't.

  - Class: mult_question
    Output: 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?
    AnswerChoices: P(+|D); P(D|+);P(~D|+); P(D|-)
    CorrectAnswer: P(D|+)
    AnswerTests: omnitest(correctVal='P(D|+)')
    Hint: We've already been given the information that the test was positive '+'. We want to know whether D is present given the positive test result.

  - Class: text
    Output: By Bayes' Formula,   P(D|+) = P(+|D) * P(D) / ( P(+|D) * P(D) + P(+|~D) * P(~D) )

  - Class: text
    Output: 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!

  - Class: cmd_question
    Output: 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.)
    CorrectAnswer: .997*.001
    AnswerTests: equiv_val(0.000997)
    Hint: Multiply the test sensitivity by the prevalence.

  - Class: cmd_question
    Output: Now solve for the remainder of the denominator, P(+|~D)*P(~D).
    CorrectAnswer: (1-.985)*(1-.001)
    AnswerTests: equiv_val(.014985)
    Hint: Multiply the complement of test specificity by the complement of prevalence.
```

```
- Class: cmd_question
  Output: 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|+).
  CorrectAnswer: .000997/(.000997+.014985)
  AnswerTests: equiv_val(.06238268)
  Hint: Divide (.997*.001) by (.997*.001 + .015*.999)

- Class: text
  Output: 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.

- Class: mult_question
  Output: 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_-?
  AnswerChoices: P(-|D) / P(-|~D); P(+|~D) / P(-|D);P(-|D) / P(+|~D); I haven't a clue.
  CorrectAnswer: P(-|D) / P(-|~D)
  AnswerTests: omnitest(correctVal='P(-|D) / P(-|~D)')
  Hint: The signs of the test in both the numerator and denominator have to agree as they did for the DLR_+.

- Class: text
  Output:  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.

- Class: mult_question
  Output: Since the numerator is close to 1 and the denominator is close to 0 do you expect DLR_+ to be large or small?
  AnswerChoices: Large; Small; I haven't a clue.
  CorrectAnswer: Large
  AnswerTests: omnitest(correctVal='Large')
  Hint: What happens when you divide a large number by a much smaller one?

- Class: mult_question
  Output: 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?
  AnswerChoices: Large; Small; I haven't a clue.
  CorrectAnswer: Small
  AnswerTests: omnitest(correctVal='Small')
  Hint: What happens when you divide by small number by a larger one?


- Class: text
  Output:  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....

- Class: text
  Output:   P(D|+) / P(~D|+) = P(+|D) * P(D) / (P(+|~D) * P(~D)) = P(+|D)/P(+|~D) * P(D)/P(~D).

- Class: mult_question
  Output: 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
  AnswerChoices: the DLR_+; the DLR_-; I haven't a clue.
  CorrectAnswer: the DLR_+
  AnswerTests: omnitest(correctVal='the DLR_+')
```

```yaml
  Hint: Do you recognize the expression P(+|D) / P(+|~D)? The '+' signs are a big clue.

- Class: text
  Output: 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.

- Class: text
  Output:  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).

- Class: mult_question
  Output:  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?
  AnswerChoices: post-test odds are greater than pre-test odds; post-test odds are less than pre-test odds; I haven't a clue.
  CorrectAnswer: the DLR_+
  AnswerTests: omnitest(correctVal='post-test odds are less than pre-test odds')
  Hint: Remember that we argued (hopefully convincingly) that DLR_- is small (less than 1). Post-test odds = Pre-test odds * DLR_- so post-test odds are a fraction of the pre-test odds.

- Class: text
  Output: Let's cover some basics now.

- Class: text
  Output: 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.

- Class: mult_question
  Output: 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?
  AnswerChoices: 1/6; 2/6; 1/36; 1/2
  CorrectAnswer: 1/36
  AnswerTests: omnitest(correctVal='1/36')
  Hint: Square the probability of rolling a single '6' since the two rolls are independent of one another.

- Class: mult_question
  Output: 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?
  AnswerChoices: 1/6; 2/6; 1/36; 1/2
  CorrectAnswer: 1/6
  AnswerTests: omnitest(correctVal='1/6')
  Hint: Now the events aren't independent. You don't care what the first roll is so that's a probability 1 event. The second roll just has to match the first, so that's a 1/6 event.

- Class: mult_question
  Output: 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?
  AnswerChoices: Yes; No
  CorrectAnswer: No
  AnswerTests: omnitest(correctVal='No')
  Hint: The events aren't independent since genetic or environmental factors likely will affect the outcome.

- Class: text
  Output: 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".

- Class: text
```

```
  Output: 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.

- Class: text
  Output: Congrats! You've concluded this lesson on conditional probability. We hope
  you liked it unconditionally.

- Class: mult_question
  Output: "Would you like to receive credit for completing this course on
    Coursera.org?"
  CorrectAnswer: NULL
  AnswerChoices: Yes;No
  AnswerTests: coursera_on_demand()
  Hint: ""
```