```
  Course: Statistical_Inference
  Lesson: Probability2

- Class: text
  Output: "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
  Statistical_Inference/Probability.)"

- Class: text
  Output: In this lesson, we'll continue to discuss probability.

- Class: text
  Output: Recall that a random variable is a numerical outcome of an experiment. It can
  be discrete (take on only a countable number of possibilities), or continuous (take
  on any value on the real line or subset of it).

- Class: mult_question
  Output: If you had a ruler of infinite precision, would measuring the height of
  adults around the world be continuous or discrete?
  AnswerChoices: continuous; discrete
  CorrectAnswer: continuous
  AnswerTests: omnitest(correctVal='continuous')
  Hint: The ruler of infinite precision is the hint. Can you list all possible heights?

- Class: mult_question
  Output: Is the drawing of a hand of cards continuous or discrete?
  AnswerChoices:  discrete; continuous;
  CorrectAnswer: discrete
  AnswerTests: omnitest(correctVal='discrete')
  Hint: Can you enumerate the possible outcomes?

- Class: text
  Output: Continuous random variables are usually associated with measurements of time,
  distance, or some biological process since they can take on any value, often within
  some specified range. Limitations of precision in taking the measurements may imply
  that the values are discrete; we in fact consider them continuous.

- Class: text
  Output: A probability mass function (PMF) gives the probability that a discrete
  random variable is exactly equal to some value.

- Class: mult_question
  Output: For instance, suppose we have a coin which may or may not be fair. Let x=0
  represent a 'heads' outcome and x=1 represent a 'tails' outcome of a coin toss. If p
  is the probability of 'heads' which of the following represents the PMF of the coin
  toss?  The variable x is either 0 (heads) or 1 (tails).
  AnswerChoices: (p^x)*(1-p)^(1-x); (p^(1-x))*(1-p)^x
  CorrectAnswer: (p^(1-x))*(1-p)^x
  AnswerTests: omnitest(correctVal='(p^(1-x))*(1-p)^x')
  Hint: The probability p is associated with a 'heads' outcome which occurs when x=0.
  Which of the two expressions has an exponent of 1 for p when x is 0?

- Class: text
  Output: A probability density function is associated with a continuous random
  variable. To quote from Wikipedia, it "is a function that describes the relative
  likelihood for this random variable to take on a given value. The probability of the
  random variable falling within a particular range of values is given by ... the area
  under the density function but above the horizontal axis and between the lowest and
  greatest values of the range."

- Class: text
  Output: We'll repeat two requirements of a probability density function. It  must be
  nonnegative everywhere, and the area under it must  equal one."

- Class: figure
  Output: Consider this figure - a rectangle with height 1 and width 2 with a diagonal
  line drawn from the lower left corner (0,0) to the upper right (2,1). The area of the
  entire rectangle is 2 and elementary geometry tells us that the diagonal divides the
```

```
    rectangle into 2 equal areas.
    Figure: figure1.R
    FigureType: new

  - Class: mult_question
    Output: Could the diagonal line represent a probability density function for a random
    variable with a range of values between 0 and 2? Assume the lower side of the
    rectangle is the x axis.
    AnswerChoices: Yes; No
    CorrectAnswer: Yes
    AnswerTests: omnitest(correctVal='Yes')
    Hint: Is the line nonnegative? Is the area under the diagonal 1?

  - Class: figure
    Output: Now consider the shaded portion of the triangle - a smaller triangle with a
    base of length 1.6 and height determined by the diagonal. We'll answer the question,
    "What proportion of the big triangle is shaded?"
    Figure: figure2.R
    FigureType: new

  - Class: text
    Output: This proportion represents the probability that throwing a piece of cat
    kibble at the bigger triangle (below the diagonal) hits the blue portion.

  - Class: mult_question
    Output:  We have to compute the area of the blue triangle. (Since the area of the big
    triangle is 1, the area of the blue triangle is the proportion of the big triangle
    that is shaded.) We know the base, but what is its height?
    AnswerChoices: .5; .25; .8; I can't tell
    CorrectAnswer: .8
    AnswerTests: omnitest(correctVal='.8')
    Hint: The  slope of a line is the "rise" (change in height) divided by the "run"
    (change in width), so the diagonal's slope is 1/2. At x=1.6, the y value of the
    diagonal is 1/2 * 1.6.

  - Class: cmd_question
    Output: What is the area of the blue triangle?
    CorrectAnswer: .64
    AnswerTests: equiv_val(.64)
    Hint: Multiply the base by the height and divide by 2.

  - Class: cmd_question
    Output: So, what is the probability that the kibble we throw at the bigger triangle
    will hit the blue portion?
    CorrectAnswer: .64
    AnswerTests: equiv_val(.64)
    Hint: The area of the blue triangle divided by the area of the big triangle gives you
    the probability.

  - Class: text
    Output: This artificial example was meant to illustrate a simple probability density
    function (PDF). Most PDFs have underlying formulae more complicated than lines. We'll
    see more of these in future lessons.

  - Class: text
    Output: The cumulative distribution function (CDF) of a random variable X, either
    discrete or continuous, is the function F(x) equal to the probability that X is less
    than or equal to x. In the example above, the area of the blue triangle represents
    the probability that the random variable was less than or equal to the value 1.6.

  - Class: mult_question
    Output: In the triangle example from above, which of the following expressions
    represents F(x), the CDF?
    AnswerChoices: x*x/4; x*x/2; x*2x/2; x^2
    CorrectAnswer: x*x/4
    AnswerTests: omnitest(correctVal='x*x/4')
    Hint: The term 'x' is the base, x/2 is the height.  Plug these into the formula for
    the area of a triangle.

```

```
- Class: text
  Output: If you're familiar with calculus you might recognize that when you're
  computing areas under curves you're actually integrating functions.

- Class: text
  Output: When the random variable is continuous, as in the example, the PDF is the
  derivative of the CDF. So integrating the PDF (the line represented by the diagonal)
  yields the CDF. When you evaluate the CDF at the limits of integration the result is
  an area.

- Class: cmd_question
  Output: To see this in the example, we've defined  the function mypdf for you. This
  is the equation of the line represented by the diagonal of the rectangle. As the PDF,
  it is  the derivative of F(x), the CDF. Look at mypdf now.
  CorrectAnswer: mypdf
  AnswerTests: omnitest(correctExpr='mypdf')
  Hint: Type 'mypdf' at the R prompt.

- Class: cmd_question
  Output: Now use the R function integrate to integrate mypdf with the parameters lower
  equal to 0 and upper equal to 1.6. See if you get the same area (probability) you got
  before.
  CorrectAnswer: integrate(mypdf,0,1.6)
  AnswerTests: omnitest(correctExpr='integrate(mypdf,0,1.6)')
  Hint: Type 'integrate(mypdf,0,1.6)' at the R prompt.

- Class: text
  Output: The survivor function S(x) of a random variable X is defined as the function
  of x equal to the probability that the random variable X is greater than the value x.
  This is the complement of the CDF F(x), in our example, the portion of the lower
  triangle that is not shaded.

- Class: mult_question
  Output: In our example, which of the following expressions represents the survival
  function?
  AnswerChoices: 1-x*x/4; 1-x*x/2; 1-x*2x/2; 1-x^2
  CorrectAnswer: 1-x*x/4
  AnswerTests: omnitest(correctVal='1-x*x/4')
  Hint: Since areas under PDF's must be 1 and the survival function is the complement
  of the CDF, the survival function and the CDF sum to 1.

- Class: text
  Output: The quantile v of a CDF is the point x_v at which the CDF has the value v.
  More precisely, F(x_v)=v. A percentile is a quantile in which v is expressed as a
  percentage.

- Class: mult_question
  Output: What percentile is the median?
  AnswerChoices: 50-th; 25-th; 95-th; I can't tell
  CorrectAnswer: 50-th
  AnswerTests: omnitest(correctVal='50-th')
  Hint: The median is the point at which half of the outcomes are above and half are
  below.

- Class: cmd_question
  Output: What is the 50th percentile of the CDF F(x)=(x^2)/4 from the example above?
  CorrectAnswer: 1.414214
  AnswerTests: equiv_val(sqrt(2))
  Hint: Solve for the x such that x^2=4*.5=2

- Class: mult_question
  Output: What does this mean with respect to the kibble we're tossing at the triangle?
  AnswerChoices: Half of it falls to the left of 1.41; All of it falls to the left of
  1.41; All of it falls to the right of 1.41; All of it falls on the vertical line at
  1.41
  CorrectAnswer: Half of it falls to the left of 1.41
  AnswerTests: omnitest(correctVal='Half of it falls to the left of 1.41')
  Hint: Recall the meaning of median (half).
```

```
- Class: text
  Output: We'll close by repeating some important points.

- Class: text
  Output: A probability model connects data to a population using assumptions.

- Class: text
  Output: Be careful to distinguish between population medians and sample medians.

- Class: text
  Output:  A sample median is an estimator of a population median (the estimand).

- Class: text
  Output: Congrats! You've concluded this lesson on probability.

- 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: ""

```