```yaml
Course: Statistical_Inference
Lesson: CommonDistros

- Class: text
  Output: "Common Distributions. (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/06_CommonDistros.)"

- Class: mult_question
  Output: Given the title of this lesson, what do you think it will cover?
  AnswerChoices: Common Distributions; Rare Distributions; Common Bistros; I haven't a clue
  CorrectAnswer: Common Distributions
  AnswerTests: omnitest(correctVal='Common Distributions')
  Hint: Part of the title is an abbreviation for another word we've seen several times in earlier lessons.

- Class: text
  Output: The first distribution we'll examine is the Bernoulli which is associated with experiments which have only 2 possible outcomes. These are also called (by people in the know) binary trials.

- Class: mult_question
  Output: It might surprise you to learn that you've probably had experience with Bernoulli trials. Which of the following would be a Bernoulli trial?
  AnswerChoices: Drawing a card from a deck; Tossing a die; Flipping a coin; Spinning a roulette wheel
  CorrectAnswer: Flipping a coin
  AnswerTests: omnitest(correctVal='Flipping a coin')
  Hint: Which of the choices has only two possible outcomes?

- Class: mult_question
  Output: For simplicity, we usually say that Bernoulli random variables take only the values 1 and 0. Suppose we also specify that the probability that the Bernoulli outcome of 1 is p. Which of the following represents the probability of a 0 outcome?
  AnswerChoices: p; 1-p; p^2; p(1-p)
  CorrectAnswer: 1-p
  AnswerTests: omnitest(correctVal='1-p')
  Hint: Recall that the sum of the probabilities of all the outcomes is 1.

- Class: mult_question
  Output: If the probability of a 1 is p and the probability of a 0 is 1-p which of the following represents the PMF of a Bernoulli distribution? Recall that the PMF is the function representing the probability that X=x.
  AnswerChoices: p^x * (1-p)^(1-x); p^(1-x) *(1-p)*(1-x); p*(1-p); x*(1-x)
  CorrectAnswer: p^x * (1-p)^(1-x)
  AnswerTests: omnitest(correctVal='p^x * (1-p)^(1-x)')
  Hint: When x=1, which of the given expressions yields p?

- Class: mult_question
  Output: Recall the definition of the expectation of a random variable. Suppose we have a Bernoulli random variable and, as before, the probability it equals 1 (a success) is p and probability it equals 0 (a failure) is 1-p. What is its mean?
  AnswerChoices: p; 1-p; p^2; p(1-p)
  CorrectAnswer: p
  AnswerTests: omnitest(correctVal='p')
  Hint: Add the two terms x*p(x) where x equals 0 and 1 respectively.

- Class: mult_question
  Output: Given the same Bernoulli random variable above, which of the following represents E(X^2)
  AnswerChoices: p(1-p); p^2; (1-p)^2; p; 1-p
  CorrectAnswer: p
  AnswerTests: omnitest(correctVal='p')
  Hint: Add the two terms x^2*p(x) where x equals 0 and 1 respectively.

- Class: mult_question
  Output: Use the answers of the last two questions to find the variance of the
```

```
  Bernoulli random variable. Recall Var = E(X^2)-(E(X))^2
  AnswerChoices: p(1-p); p^2-p; p(p-1); p^2*(1-p)^2
  CorrectAnswer: p(1-p)
  AnswerTests: omnitest(correctVal='p(1-p)')
  Hint: E(X^2)=p and E(X)=p, so Var=p-p^2. Rewrite this expression by factoring out the p.

- Class: text
  Output: Binomial random variables are obtained as the sum of iid Bernoulli trials. Specifically, let X_1, ..., X_n be iid Bernoulli(p) random variables; then X = X_1 + X_2 + ... X_n is a binomial random variable. Binomial random variables represent the number of successes, k, out of n independent Bernoulli trials. Each of the trials has probability p.

- Class: mult_question
  Output: The PMF of a binomial random variable X is the function representing the probability that X=x. In other words, that there are x successes out of n independent trials. Which of the following represents the PMF of a binomial distribution? Here x, the number of successes, goes from 0 to n, the number of trials, and  choose(n,x) represents the binomial coefficient 'n choose x' which is the number of ways x successes out of n trials can occur regardless of order.
  AnswerChoices: choose(n,x) * p^x * (1-p)^(n-x); choose(n,x) * p^(n-x) * (1-p)^x; p^x; choose(n,x) * p*x*(1-p)*(1-x)
  CorrectAnswer: choose(n,x) * p^x * (1-p)^(n-x)
  AnswerTests: omnitest(correctVal='choose(n,x) * p^x * (1-p)^(n-x)')
  Hint: To take the value x, the random variable X must have x 'successes'. Each of these occurs with probability p. It also must have n-x 'failures', each of which occurs with probability (1-p). We don't care about the order in which the successes and failures occur so we have to multiply by choose(n,x).

- Class: cmd_question
  Output: Suppose we were going to flip a biased coin 5 times. The probability of tossing a head is .8 and a tail .2. What is the probability that you'll toss at least 3 heads.
  CorrectAnswer: 0.94208
  AnswerTests: equiv_val(0.94208)
  Hint: You'll have to add together 3 terms each of the form, choose(5,x)*(.8)^x*(.2)^(5-x) for x=3,4,5 .

- Class: cmd_question
  Output: Now you can verify your answer with the R function pbinom. The quantile is 2, the size is 5, the prob is .8 and the lower.tail is FALSE. Try it now.
  CorrectAnswer: pbinom(2,size=5,prob=.8,lower.tail=FALSE)
  AnswerTests: omnitest(correctExpr='pbinom(2,size=5,prob=.8,lower.tail=FALSE)')
  Hint: Type pbinom(2,size=5,prob=.8,lower.tail=FALSE) at the R prompt.

- Class: text
  Output: Another very common distribution is the normal or Gaussian. It has a complicated density function involving its mean mu and variance sigma^2. The key fact of the density formula is that when plotted, it forms a bell shaped curve, symmetric about its mean mu. The variance sigma^2 corresponds to the width of the bell, the higher the variance, the fatter the bell. We denote a normally distributed random variable X as X ~ N(mu, sigma^2).

- Class: text
  Output:  When mu = 0 and sigma = 1 the resulting distribution is called the standard normal distribution and it is often labeled Z.

- Class: figure
  Output: Here's a picture of the density function of a standard normal distribution. It's centered at its mean 0 and the vertical lines (at the integer points of the x-axis) indicate the standard deviations.
  Figure: plotNormal.R
  FigureType: new

- Class: figure
  Output:  Approximately 68%, 95% and 99% of the normal density lie within 1, 2 and 3 standard deviations from the mean, respectively. These are shown in the three shaded areas of the figure. For example, the darkest portion (between -1 and 1) represents
```

```
    68% of the area.
  Figure: stddev1.R
  FigureType: new

- Class: cmd_question
  Output: The R function qnorm(prob) returns the value of x (quantile) for which the area under the standard normal distribution to the left of x equals the parameter prob. (Recall that the entire area under the curve is 1.) Use qnorm now to find the 10th percentile of the standard normal. Remember the argument prob must be between 0 and 1. You don't have to specify any of the other parameters since the default is the standard normal.
  CorrectAnswer: qnorm(.10)
  AnswerTests: omnitest(correctExpr='qnorm(.10)')
  Hint: Type qnorm(.1) at the R prompt.

- Class: figure
  Output:  We'll see this now by drawing the vertical line at the quantile -1.281552.
  Figure: plotQuantile.R
  FigureType: new

- Class: mult_question
  Output: Which of the following would you expect to be the 1st percentile?
  AnswerChoices: -2.33; -1.0; 0; 2.33; -1.28
  CorrectAnswer: -2.33
  AnswerTests: omnitest(correctVal='-2.33')
  Hint: Since 1 is smaller than 10 the quantile for the 1st percentile should be smaller than the quantile for 10th percentile.

- Class: cmd_question
  Output: By looking at the picture can you say what the 50th percentile is?
  CorrectAnswer: 0
  AnswerTests: equiv_val(0)
  Hint: What point x marks the halfway point of the graph?

- Class: mult_question
  Output: We can use the symmetry of the bell curve to determine other quantiles. Given that 2.5% of the area under the curve falls to the left of x=-1.96, what is the 97.5 percentile for the standard normal?
  AnswerChoices: 1.96; 2.33; -1.28; 2
  CorrectAnswer: 1.96
  AnswerTests: omnitest(correctVal='1.96')
  Hint: 2.5% of the area falls to the right of the quantile of the 97.5 percentile.

- Class: text
  Output: Here are two useful facts concerning normal distributions. If X is a normal random variable with mean mu and variance sigma^2, i.e., X ~ N(mu,sigma^2),

- Class: text
  Output: then the random variable Z defined as Z = (X -mu)/sigma is normally distributed with mean 0 and variance 1, i.e., Z ~ N(0, 1). (Z is standard normal.)

- Class: text
  Output: The converse is also true. If Z is standard normal, i.e., Z ~ N(0,1), then the random variable X defined as  X = mu + sigma*Z is normally distributed with mean mu and variance sigma^2, i.e., X ~ N(mu, sigma^2)

- Class: text
  Output: These formulae allow you to easily compute quantiles (and thus percentiles) for ANY normally distributed variable if you know its mean and variance. We'll show how to find the 97.5th percentile of a normal distribution with mean 3 and variance 4.

- Class: cmd_question
  Output: Again, we can use R's qnorm function and simply specify the mean and standard deviation (the square root of the variance). Do this now. Find the 97.5th percentile of a normal distribution with mean 3 and standard deviation 2.
  CorrectAnswer: qnorm(.975,mean=3,sd=2)
  AnswerTests: omnitest(correctExpr='qnorm(.975,mean=3,sd=2)')
  Hint: Type qnorm(.975,mean=3,sd=2) at the R prompt.
```

```yaml
- Class: cmd_question
  Output: Let's check it using the formula above, X = mu + sigma*Z. Here we'll use the 97.5th percentile for the standard normal as the value Z in the formula. Recall that we previously calculated this to be 1.96. Let's multiply this by the standard deviation of the given normal distribution (2) and add in its mean (3) to see if we get a result close to the one qnorm gave us.
  CorrectAnswer: 6.92
  AnswerTests: equiv_val(6.92)
  Hint: Type 1.96*2 + 3 at the R prompt.

- Class: cmd_question
  Output: Suppose you have a normal distribution with mean 1020 and standard deviation of 50 and you want to compute the probability that the associated random variable X > 1200. The easiest way to do this is to use R's pnorm function in which you specify the quantile (1200), the mean (1020) and standard deviation (50). You also must specify that the lower.tail is FALSE since we're asking for a probability that the random variable is greater than our quantile. Do this now.
  CorrectAnswer: pnorm(1200,mean=1020,sd=50,lower.tail=FALSE)
  AnswerTests: omnitest(correctExpr='pnorm(1200,mean=1020,sd=50,lower.tail=FALSE)')
  Hint: Type pnorm(1200,mean=1020,sd=50,lower.tail=FALSE) at the R prompt.

- Class: cmd_question
  Output: Alternatively, we could use the formula above to transform the given distribution to a standard normal. We compute the number of standard deviations the specified number (1200) is from the mean with Z = (X -mu)/sigma. This is our new quantile. We can then use the standard normal distribution and the default values of pnorm. Remember to specify that lower.tail is FALSE.  Do this now.
  CorrectAnswer: pnorm((1200-1020)/50,lower.tail=FALSE)
  AnswerTests: omnitest(correctExpr='pnorm((1200-1020)/50,lower.tail=FALSE)')
  Hint: Type pnorm((1200-1020)/50,lower.tail=FALSE) at the R prompt.

- Class: cmd_question
  Output: For practice, using the same distribution, find the  75% percentile. Use qnorm and specify the probability (.75), the mean (1020) and standard deviation (50). Since we want to include the left part of the curve we can use the default lower.tail=TRUE.
  CorrectAnswer: qnorm(.75,mean=1020,sd=50)
  AnswerTests: omnitest(correctExpr='qnorm(.75,mean=1020,sd=50)')
  Hint: Type qnorm(.75,mean=1020,sd=50) at the R prompt.

- Class: cmd_question
  Output: Note that R functions pnorm and qnorm are inverses. What would you expect pnorm(qnorm(.53)) to return?
  CorrectAnswer: .53
  AnswerTests: equiv_val(.53)
  Hint: Type pnorm(qnorm(.53)) at the R prompt.

- Class: cmd_question
  Output: How about qnorm(pnorm(.53))?
  CorrectAnswer: .53
  AnswerTests: equiv_val(.53)
  Hint: Type qnorm(pnorm(.53)) at the R prompt.

- Class: text
  Output: Now let's talk about our last common distribution, the Poisson. This is, as Wikipedia tells us, "a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event."

- Class: text
  Output: In other words, the Poisson distribution models counts or number of event in some interval of time. From Wikipedia, "Any variable that is Poisson distributed only takes on integer values."

- Class: text
  Output: The PMF of the Poisson distribution has one parameter, lambda. As with the other distributions the PMF calculates the probability that the Poisson distributed random variable X takes the value x. Specifically, P(X=x)=(lambda^x)e^(-lambda)/x!.
```

```
  Here x ranges from 0 to infinity.

- Class: text
  Output: The mean and variance of the Poisson distribution are both lambda.

- Class: text
  Output: Poisson random variables are used to model rates such as the rate of hard
  drive failures. We write X~Poisson(lambda*t) where lambda is the expected count per
  unit of time and t is the total monitoring time.

- Class: cmd_question
  Output: For example, suppose the number of people that show up at a bus stop is
  Poisson with a mean of 2.5 per hour, and we want to know the probability that at most
  3 people show up in a 4 hour period. We use the R function ppois which returns a
  probability that the random variable is less than or equal to 3. We only need to
  specify the quantile (3) and the mean (2.5 * 4). We can use the default parameters,
  lower.tail=TRUE and log.p=FALSE. Try it now.
  CorrectAnswer: ppois(3,2.5 * 4)
  AnswerTests: ANY_of_exprs('ppois(3,2.5 * 4)','ppois(3,4*2.5)')
  Hint: Type  ppois(3,2.5 * 4) at the R prompt.

- Class: text
  Output: Finally, the Poisson distribution approximates the binomial distribution in
  certain cases. Recall that the binomial distribution is the discrete distribution of
  the number of successes, k, out of n independent binary trials, each with probability
  p. If n is large and p is small then the Poisson distribution with lambda equal to
  n*p is a good approximation to the binomial distribution.

- Class: cmd_question
  Output: To see this, use the R function pbinom to estimate the probability that
  you'll see at most 5 successes out of 1000 trials each of which has probability .01.
  As before, you can use the default parameter values (lower.tail=TRUE and log.p=FALSE)
  and just specify the quantile, size, and probability.
  CorrectAnswer: pbinom(5,1000,.01)
  AnswerTests: omnitest(correctExpr='pbinom(5,1000,.01)')
  Hint: Type  pbinom(5,1000,.01)  at the R prompt.

- Class: cmd_question
  Output: Now use the function ppois with quantile equal to 5 and lambda equal to n*p
  to see if you get a similar result.
  CorrectAnswer: ppois(5,1000*.01)
  AnswerTests: omnitest(correctExpr='ppois(5,1000*.01)')
  Hint: Type  ppois(5,1000*.01)  at the R prompt.

- Class: text
  Output: See how they're close? Pretty cool, right? This worked because n was large
  (1000) and p was small (.01).

- Class: text
  Output: Congrats! You've concluded this uncommon lesson on common distributions.

- 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: ""
```