```yaml
Course: Statistical_Inference
Lesson: Hypothesis_Testing

- Class: text
  Output: "Hypothesis_Testing. (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/09_HT.)"

- Class: text
  Output: In this lesson, as the name suggests, we'll discuss hypothesis testing which is concerned with making decisions about populations using observed data.

- Class: text
  Output: An important concept in hypothesis testing is the NULL hypothesis, usually denoted as H_0. This is the hypothesis that represents the status_quo and is assumed true. It's a baseline against which you're testing alternative hypotheses, usually denoted by H_a. Statistical evidence is required to reject H_0 in favor of the research or alternative hypothesis.

- Class: text
  Output: We'll consider the motivating example from the slides. A respiratory disturbance index (RDI) of more than 30 events / hour is considered evidence of severe sleep disordered breathing (SDB). Suppose that in a sample of 100 overweight subjects with other risk factors for SDB at a sleep clinic, the mean RDI (X') was 32 events / hour with a standard deviation (s) of 10 events / hour.

- Class: text
  Output: We want to test the null hypothesis H_0 that mu = 30. Our alternative hypothesis H_a is mu>30. Here mu represents the hypothesized population mean RDI.

- Class: text
  Output: So we have two competing hypotheses, H_0 and H_a, of which we'll have to pick one (using statistical evidence). That means we have four possible outcomes determined by what really is (the truth) and which hypothesis we accept based on our data. Two of the outcomes are correct and two are errors.

- Class: mult_question
  Output: Which of the following outcomes would be correct?
  AnswerChoices: H_0 is TRUE and we reject it; H_a is TRUE and we accept it; H_0 is FALSE and we accept it; H_a is FALSE and we accept it
  CorrectAnswer: H_a is TRUE and we accept it
  AnswerTests: omnitest(correctVal='H_a is TRUE and we accept it')
  Hint: It's always better to ACCEPT the TRUTH.

- Class: mult_question
  Output: Which of the following outcomes would be an error?
  AnswerChoices: H_0 is TRUE and we reject it; H_a is TRUE and we accept it; H_0 is FALSE and we reject it; H_a is FALSE and we reject it
  CorrectAnswer: H_0 is TRUE and we reject it
  AnswerTests: omnitest(correctVal='H_0 is TRUE and we reject it')
  Hint: It's always a mistake to REJECT the TRUTH.

- Class: text
  Output: So it's correct to accept a true hypothesis or reject a false one. Pretty clear, right?

- Class: text
  Output: The errors are also clear - rejecting a true hypothesis or accepting a false one.

- Class: text
  Output: We distinguish between these two errors. A Type I error REJECTS a TRUE null hypothesis H_0 and a Type II error ACCEPTS a FALSE null hypothesis H_0.

- Class: mult_question
  Output: Can we ever be sure that we're absolutely right?
  AnswerChoices: Yes; No;  Always; Let's not get into philosophy now
  CorrectAnswer: No
```

```yaml
  AnswerTests: omnitest(correctVal='No')
  Hint: The key word is "absolutely." We're basing decisions on available data which may or may not be complete or representative or accurate. Who knows?

- Class: text
  Output: Since there's  some element of uncertainty in questions concerning populations, we deal with probabilities. In our hypothesis testing we'll set the probability of making errors small. For now we'll focus on Type I errors, rejecting a correct hypothesis.

- Class: text
  Output: The probabilities of making these two kinds of errors are related. If you decrease the probability of making a Type I error (rejecting a true hypothesis), you increase the probability of making a Type II error (accepting a false one) and vice versa.

- Class: mult_question
  Output: As in the slides, we'll consider an American court of law. The null hypothesis is that the defendant is innocent. If an innocent man is convicted what type of error is this?
  AnswerChoices: Type I; Type II
  CorrectAnswer: Type I
  AnswerTests: omnitest(correctVal='Type I')
  Hint: You're rejecting a true null hypothesis. Recall that a Type II error accepts a false null hypothesis.

- Class: mult_question
  Output: You might send the innocent man to jail by rejecting H_0. Suppose a guilty person is not convicted. What type of error is this?
  AnswerChoices: Type I; Type II
  CorrectAnswer: Type II
  AnswerTests: omnitest(correctVal='Type II')
  Hint: You're accepting a null hypothesis (innocence) that is false. Recall that a Type I error rejects the truth.

- Class: text
  Output: Back to sleep (example)! A reasonable strategy would reject the null hypothesis if our sample mean X' was larger than some constant C. We choose C so that the probability of a Type I error, alpha, is .05 (or some other favorite constant). Many scientific papers use .05 as a standard level of rejection.

- Class: text
  Output:  This means that  alpha, the Type I error rate, is the probability of rejecting the null hypothesis when, in fact, it is correct. We don't want alpha too low because then we would never reject the null hypothesis even if it's false.

- Class: cmd_question
  Output: Recall that the standard error of a sample mean is given by the formula s/sqrt(n). Recall in our sleep example we had a sample of 100 subjects, our mean RDI (X') was 32 events / hour with a standard deviation (s) of 10 events / hour. What is the standard error of the mean in this example?
  CorrectAnswer: 1
  AnswerTests: equiv_val(1)
  Hint: Divide s by sqrt(n).

- Class: text
  Output: Under H_0, X' is normally distributed with mean mu=30 and variance 1. (We're estimating the variance as the square of the standard error which in this case is 1.) We want to choose the constant C so that the probability that X is greater than C given H_0 is 5%. That is, P(X > C| H_0) is 5%. Sound familiar?

- Class: figure
  Output: Here's a plot to show what we mean. The shaded portion represents 5% of the area under the curve and those X values in it are those for which the probability that X>C is 5%.
  Figure: conf_5pct.R
  FigureType: new

- Class: mult_question
```

```
    Output: The shaded portion represents 5% of the area under this normal density curve. Which expression represents the smallest value X for which the area is shaded, assuming this is standard normal?
    AnswerChoices: qnorm(.95); rnorm(.95); dnorm(.95); qt(.95,99)
    CorrectAnswer: qnorm(.95)
    AnswerTests: omnitest(correctVal='qnorm(.95)')
    Hint: The shading begins at the 95th percentile and the smallest value X for which the area is shaded represents the 95th quantile.

  - Class: text
    Output: The 95th percentile of a standard normal distribution is 1.645 standard deviations from the mean, so in our example  we have to set C to be 1.645 standard deviations MORE than our hypothesized mean of 30, that is,  C = 30 +  1.645 * 1 = 31.645 (recall that the variance and standard deviation equalled 1).

  - Class: text
    Output: This means that if our OBSERVED (sample) mean X' >= C, then it's only a 5% chance that a random draw from this N(30,1) distribution is larger than C.

  - Class: mult_question
    Output: Recall that our observed mean X' is 32 which is greater than C=31.645, so it falls in that 5% region. What do we do with H_0?
    AnswerChoices: reject it; fail to reject it; give it another chance
    CorrectAnswer: reject it
    AnswerTests: omnitest(correctVal='reject it')
    Hint: The observed sample mean X' falls in the region of rejection so we toss out H_0.

  - Class: text
    Output: So the rule "Reject H_0 when the sample mean X' >= 31.645" has the property that the probability of rejecting H_0 when it is TRUE is 5% given the model of this example - hypothesized mean mu=30, variance=1 and n=100.

  - Class: text
    Output:  Instead of computing a constant C as a cutpoint for accepting or rejecting the null hypothesis, we can simply compute a Z score, the number of standard deviations the sample mean is from the hypothesized mean. We can then  compare it to quantile determined by alpha.

  - Class: text
    Output: How do we do this? Compute the distance between the two means (32-30) and divide by the standard error of the mean, that is  (s/sqrt(n)).

  - Class: cmd_question
    Output: What is the Z score for this example? Recall the Z score is X'-mu divided by the standard error of the mean. In this example X'=32, mu=30 and the standard error is 10/sqrt(100)=1.
    CorrectAnswer: 2
    AnswerTests: equiv_val(2)
    Hint: Divide 2 by 1.

  - Class: mult_question
    Output: The Z score is  2. The quantile is 1.645, so since 2>1.645. What do we do with H_0?
    AnswerChoices: reject it; fail to reject it; give it another chance
    CorrectAnswer: reject it
    AnswerTests: omnitest(correctVal='reject it')
    Hint: Since the Z score exceeded the quantile the observed sample mean X' falls in the region of rejection so we toss out H_0.

  - Class: text
    Output:  The general rule for rejection is if sqrt(n) * ( X' - mu) / s > Z_{1-alpha}.

  - Class: text
    Output: Our test statistic is (X'-mu) / (s/sqrt(n)) which is standard normal.

  - Class: mult_question
    Output: This means that our test statistic has what mean and standard deviation?
    AnswerChoices: 0 and 1; 1 and 0; 0 and 0; 1 and 1
    CorrectAnswer: 0 and 1
```

```
  AnswerTests: omnitest(correctVal='0 and 1')
  Hint: The standard normal is centered around 0 and has a standard deviation of 1.

- Class: text
  Output: Let's review and expand. Our null hypothesis is that the population mean mu equals the value mu_0 and alpha=.05. (This is the probability that we reject H_0 if it's true.) We can have several different alternative hypotheses.

- Class: text
  Output: Suppose our first alternative, H_a, is that mu < mu_0. We would reject H_0 (and accept H_a) when our observed sample mean is significantly less than mu_0. That is, our test statistic  (X'-mu) / (s/sqrt(n)) is  less than Z_alpha. Specifically, it is more than 1.64 standard deviations to the left of the mean mu_0.

- Class: figure
  Output: Here's a plot to show what we mean. The shaded portion represents 5% of the area under the curve and those X values in it are those which are at least 1.64 standard deviations less than  the mean. The probability of this is 5%. This means that if our sample mean fell in this area, we would reject a true null hypothesis, mu=mu_0, with probability 5%.
  Figure: conf_5pct_left.R
  FigureType: new

- Class: mult_question
  Output: We already covered the alternative hypothesis H_a that mu > mu_0 but let's review it. We would reject H_0 (and accept H_a) when our sample mean is what?
  AnswerChoices: significantly greater than mu_0; significantly less than mu_0; equal to mu_0; huh?
  CorrectAnswer: significantly greater than mu_0
  AnswerTests: omnitest(correctVal='significantly greater than mu_0')
  Hint: If we accept H_a, that the true mu is greater than the H_0 value mu_0 we would want our sample mean to be greater the mu_0.

- Class: mult_question
  Output: This means that our test statistic (X'-mu) / (s/sqrt(n)) is what?
  AnswerChoices: at least 1.64 std dev greater than mu_0; at least 1.64 std dev less than mu_0; equal to mu_0; huh?
  CorrectAnswer: at least 1.64 std dev greater than mu_0
  AnswerTests: omnitest(correctVal='at least 1.64 std dev greater than mu_0')
  Hint: If we accept H_a, that the true mu is greater than the H_0 value mu_0 we would want our test statistic to be greater than 1.64 standard deviations from the mean.

- Class: figure
  Output: Here again is the plot to show this. The shaded portion represents 5% of the area under the curve and those X values in it are those which are at least 1.64 standard deviations greater than the mean. The probability of this is 5%. This means that if our observed mean fell in this area we would reject a true null hypothesis, that mu=mu_0, with probability  5%.
  Figure: conf_5pct.R
  FigureType: new

- Class: text
  Output: Finally, let's consider the alternative hypothesis H_a that mu is simply not equal to mu_0, the mean hypothesized by the null hypothesis H_0.  We would reject H_0 (and accept H_a) when our sample mean is significantly different than mu_0, that is, either less than OR greater than mu_0.

- Class: text
  Output: Since we want to stick with a 5% rejection rate, we divide it in half and consider values at both tails, at the .025 and the .975 percentiles.  This means that our test statistic  (X'-mu) / (s/sqrt(n)) is  less than .025, Z_(alpha/2), or greater than .975, Z_(1-alpha/2).

- Class: figure
  Output: Here's the plot. As before, the shaded portion represents the 5% of the area composing the region of rejection. This time, though, it's composed of two equal pieces, each containing 2.5% of the area under the curve. The X values in the shaded portions are values which are at least 1.96 standard deviations away from the hypothesized mean.
```

```
  Figure: conf_5pct_both.R
  FigureType: new

- Class: text
  Output: Notice that  if we reject H_0, either it was FALSE (and hence our model is wrong and we are correct to reject it) OR H_0 is TRUE and we have made an error (Type I). The probability of this is 5%.

- Class: text
  Output: Since our tests were based on alpha, the probability of a Type I error, we say that  we "fail to reject H_0" rather than we "accept H_0". If we fail to reject H_0,  then H_0 could be true OR we just might not have enough data to reject it.

- Class: text
  Output: We have not fixed the probability of a type II error (accepting H_0 when it is false), which we call beta. The term POWER refers to the quantity 1-beta and it represents the probability of rejecting H_0 when it's false. This is used to determine appropriate sample sizes in experiments.

- Class: mult_question
  Output: What do you think we call the region of values for which we reject H_0?
  AnswerChoices: the rejection region; the shady tails; the abnormal region; the region of interest; the waggy tails
  CorrectAnswer: the rejection region
  AnswerTests: omnitest(correctVal='the rejection region')
  Hint: Which choice has the word 'reject' in it?

- Class: text
  Output: Note that so far we've been talking about NORMAL distributions and implicitly relying on the CENTRAL LIMIT THEOREM (CLT).

- Class: mult_question
  Output: Remember the CLT. For a distribution to be approximated by a normal what does the sample size have to be?
  AnswerChoices: large; small; abnormal; normal;
  CorrectAnswer: large
  AnswerTests: omnitest(correctVal='large')
  Hint: As the sample size gets bigger the distribution looks normal.

- Class: text
  Output: No need to worry. If we don't have a large sample size, we can use the t distribution which conveniently uses the same test statistic (X'-mu) / (s/sqrt(n)) we used above.  That means that all the examples we just went through would work exactly the same EXCEPT instead of using NORMAL quantiles, we would use t quantiles and n-1 degrees of freedom.

- Class: text
  Output: We said t distributions were very handy, didn't we?

- Class: text
  Output: Let's go back to our sleep disorder example and suppose our sample size=16 (instead of 100). As before,  (sample mean) X'=32,  (standard deviation) s=10.  H_0 says the true mean mu=30, and H_a is that mu>30. With this smaller sample size we use the t test, but our test statistic is computed the same way, namely (X'-mu)/(s/sqrt(n))

- Class: cmd_question
  Output: What is the value of the test statistic (X'-mu)/(s/sqrt(n)) with sample size 16?
  CorrectAnswer: .8
  AnswerTests: equiv_val(.8)
  Hint: Type (32-30)/(10/4) at the command prompt.

- Class: cmd_question
  Output: How many degrees of freedom do we have with a sample size of 16?
  CorrectAnswer: 15
  AnswerTests: equiv_val(15)
  Hint: Recall that the number of degrees of freedom is one less than the sample size. Here the sample size is 16.
```

```
- Class: cmd_question
  Output: Under H_0, the probability that the test statistic is larger than the 95th percentile of the t distribution is 5%. Use the R function qt with the arguments .95 and the correct number of degrees of freedom to find the quantile.
  CorrectAnswer: qt(.95,15)
  AnswerTests: omnitest(correctExpr='qt(.95,15)')
  Hint: Type qt(.95,15) at the command prompt.

- Class: text
  Output: So the test statistic (.8) is less than 1.75, the  95th percentile of the t distribution with 15 df. This means that our sample mean (32) does not fall within the region of rejection since H_a was that mu>30.

- Class: mult_question
  Output: This means what?
  AnswerChoices: we reject H_0; we fail to reject H_0; we reject H_a;
  CorrectAnswer: we fail to reject H_0
  AnswerTests: omnitest(correctVal='we fail to reject H_0')
  Hint: The test statistic is outside the region of rejection so we fail to reject H_0.

- Class: text
  Output: Now let's consider a two-sided test. Suppose that we would reject the null hypothesis if in fact the sample mean was too large or too small. That is, we want to test the alternative H_a that mu is not equal to 30. We will reject if the test statistic, 0.8, is either too large or too small.

- Class: text
  Output: As we discussed before, we want the probability of rejecting under the null to be 5%, split equally as 2.5% in the upper tail and 2.5% in the lower tail. Thus we reject if our test statistic is larger than qt(.975, 15) or smaller than qt(.025, 15).

- Class: mult_question
  Output: Do you expect qt(.975,15) to be bigger or smaller than qt(.95,15)?
  AnswerChoices: bigger; smaller
  CorrectAnswer: bigger
  AnswerTests: omnitest(correctVal='bigger')
  Hint: You're looking for a smaller area under the curve so that means the quantile will be farther out on the right tail so it should be bigger.

- Class: mult_question
  Output: Since the test statistic was smaller than qt(.95,15) will it be bigger or smaller than qt(.975,15)?
  AnswerChoices: bigger; smaller
  CorrectAnswer: smaller
  AnswerTests: omnitest(correctVal='smaller')
  Hint: If A<B and B<C it follows that A<C, right?

- Class: mult_question
  Output: Now for the left tail, qt(.025,15). What can we say about it?
  AnswerChoices: it is less than 0; it is greater than 0; it is bigger than qt(.975,15); we don't know anything about it
  CorrectAnswer: it is less than 0
  AnswerTests: omnitest(correctVal='it is less than 0')
  Hint: Any quantile of a percentile less than .5 will be less than 0 by the symmetry of the distribution.

- Class: text
  Output: Bottom line here is if you fail to reject the one sided test, you know that you will fail to reject the two sided.

- Class: mult_question
  Output: So the test statistic .8 failed both sides of the test. That means we ?
  AnswerChoices: fail to reject H_0; reject H_0; reject H_a; huh?
  CorrectAnswer: fail to reject H_0
  AnswerTests: omnitest(correctVal='fail to reject H_0')
  Hint: Again, the test statistic is close to the hypothesized mean so we fail to reject it.
```

```yaml
- Class: text
  Output: Now we usually don't have to do all this computation ourselves because R provides the function t.test which happily does all the work! To prove this, we've provided a csv file with the father_son height data from John Verzani's UsingR website (http://wiener.math.csi.cuny.edu/UsingR/) and read it into a data structure fs for you. We'll do a t test on this paired data to see if fathers and sons have similar heights (our null hypothesis).

- Class: cmd_question
  Output: Look at the dimensions of fs now using the R function dim.
  CorrectAnswer: dim(fs)
  AnswerTests: omnitest(correctExpr='dim(fs)')
  Hint: Type dim(fs) at the command prompt.

- Class: cmd_question
  Output: So fs has 1078 rows and 2 columns. The columns, fheight and sheight, contain the heights of a father and his son. Obviously there are 1078 such pairs. We can run t.test on this data in one of two ways. First, we can run it with just one argument, the difference between the heights, say fs$sheight-fs$fheight. OR we can run it with three arguments, the two heights plus the paired argument set to TRUE. Run t.test now using whichever way you prefer.
  CorrectAnswer: t.test(fs$sheight-fs$fheight)
  AnswerTests: ANY_of_exprs('t.test(fs$sheight-fs$fheight)','t.test(fs$fheight-fs$sheight)','t.test(fs$sheight,fs$fheight,paired=TRUE)','t.test(fs$fheight, fs$sheight,paired=TRUE)')
  Hint: Type t.test(fs$sheight-fs$fheight) at the command prompt.

- Class: mult_question
  Output: The test statistic is what?
  AnswerChoices: 2.2e-16; 11.7885; .8310296; 0.9969728
  CorrectAnswer: 11.7885
  AnswerTests: omnitest(correctVal='11.7885')
  Hint: We've been using the t statistic and doing t tests.

- Class: text
  Output: So the test statistic is 11.79 which is quite large so we REJECT the null hypothesis that the true mean of the difference  was 0 (if you ran the test on the difference sheight-fheight) or that the true difference in means was 0 (if you ran the test on the two separate but paired columns).

- Class: mult_question
  Output: The test statistic tell us what?
  AnswerChoices: the number of estimated std errors between the sample and hypothesized means; the sample mean; the true mean; the true variance
  CorrectAnswer: the number of estimated std errors between the sample and hypothesized means
  AnswerTests: omnitest(correctVal='the number of estimated std errors between the sample and hypothesized means')
  Hint: The test statistic tells us how many standard deviations the sample mean is from the hypothesized one. Remember t=(X'-mu)/(s/sqrt(n))

- Class: cmd_question
  Output: We can test this by multiplying the t statistic (11.7885) by the standard deviation of the data divided by the square root of the sample size. Specifically run 11.7885 * sd(fs$sheight-fs$fheight)/sqrt(1078).
  CorrectAnswer: 11.7885 * sd(fs$sheight-fs$fheight)/sqrt(1078)
  AnswerTests: omnitest(correctExpr='11.7885 * sd(fs$sheight-fs$fheight)/sqrt(1078)')
  Hint: Type 11.7885 * sd(fs$sheight-fs$fheight)/sqrt(1078) at the command prompt.

- Class: text
  Output: This should give you a close match to the mean of x which t.test gave you, 0.9969728.

- Class: text
  Output: Note the 95% confidence interval, 0.8310296 1.1629160, returned by t.test. It does not contain the hypothesized population mean 0 so we're pretty confident we can safely reject the hypothesis. This tells us that either our hypothesis is wrong or we're making a mistake (Type 1) in rejecting it.

```

```yaml
- Class: text
  Output: You've probably noticed the strong similarity between the confidence intervals we studied in the last lesson and these hypothesis tests. That's because they're equivalent!

- Class: text
  Output: If you set alpha to  some value (say .05) and ran many tests checking alternative hypotheses against H_0, that mu=mu_0, the set of all possible values for which you fail to reject H_0 forms the  (1-alpha)% (that is 95%) confidence interval for mu_0.


- Class: text
  Output: Similarly, if a (1-alpha)% interval contains mu_0, then we fail to reject H_0.

- Class: text
  Output: Let's see how hypothesis testing works with binomial distributions by considering the example from the slides. A family has 8 children, 7 of whom are girls and none are twins. Let the null hypothesis be that either gender is equally likely, like an iid coin flip.

- Class: text
  Output: So our H_0 is that p=.5, where p is the probability of a girl. We want to see if we should reject H_0 based on this sample of size 8. Our  H_a is that p>.5, so we'll do a one-sided test, i.e., look at only the right tail of the distribution.

- Class: text
  Output: Let's set alpha, the level of our test, to .05 and  find the probabilities associated with different rejection regions, where a rejection region i has at least i-1 girls out of a possible 8.

- Class: cmd_question
  Output: We've defined for you a 9-long vector, mybin, which shows nine probabilities, the i-th of which is the probability that there are at least i-1 girls out of the 8 possible children. Look at mybin now.
  CorrectAnswer: mybin
  AnswerTests: omnitest(correctExpr='mybin')
  Hint: Type mybin at the command prompt.

- Class: cmd_question
  Output: So mybin[1]=1.0, meaning that with probability 1 there are at least 0 girls, and mybin[2]=.996 is the probability that there's at least 1 girl out of the 8, and so forth. The probabilities decrease as i increases. What is the least value of i for which the probability is less than .05?
  CorrectAnswer: 8
  AnswerTests: equiv_val(8)
  Hint: mybin[7]=.144 and mybin[8]=.035.

- Class: mult_question
  Output: So  mybin[8]=.03 is the probability  of having at least 7 girls out of a sample of size 8 under H_0 (if p actually is .5) which is what our sample has. This is less than .05 so  our sample falls in this region of rejection. Does that mean we accept or reject H_0, (that either gender is equally likely) based on this sample of size 8?
  AnswerChoices: accept H_0; reject H_0
  CorrectAnswer: reject H_0
  AnswerTests: omnitest(correctVal='reject H_0')
  Hint: Our sample had 7 daughters which is in the region of rejection. H_0 is OUT.

- Class: text
  Output:  Finally, we note that a 2-sided test would mean that our alternative hypothesis is that p is not equal to .5, and it's not obvious how to do this with a binomial distribution. Don't worry, though, because the next lesson on p-values will make this clearer. It's interesting that for discrete distributions such as binomial and Poisson, inverting 2-sided tests is how R calculates exact tests. (It doesn't rely on the CLT.)

- Class: text
  Output: Congrats! We confidently hypothesize that you're happy to have finished this
```

```
    lesson. Can we test this?

- 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: ""
```