Video: “L11HypothesisTestingBasics” (10:52)

Video (00:00)
Statistically correct information can enable good decision making. For example, if you thought that students who took science classes in the morning than students who took science classes in the afternoon. This might influence your schedule, and ultimately your academic success, but only if it were true. We need a statistically correct method for testing our hypothesis.

(00:23)
The goal of hypothesis testing is to make an informed decision about a population based on sample data. So we formulate conclusions about morning and afternoon classes by sampling some students’ outcomes from both. Associated with hypothesis testing is a quantitative evaluation of “how reliable” or statistically significant the decision is, or how much we should trust the conclusions.

(00:51)
For a demonstration, suppose we believed all morning students’ grades followed a normal distribution with an average of 85 and a standard deviation of 5. This population is too large to measure all grades from all students, so we could sample one student from the morning science class and see if this student’s grade fits what we believe about the distribution. Based on the evidence provided by sample x, what can we conclude about the mystery population? It either refutes or supports our belief that the average is 85 with a standard deviation of 5. In a similar method, we could statistically test a belief that afternoon students have a lower average, and if this was shown statistically, then we would have evidence to support our scheduling decisions.

(1:40)
There are five steps in hypothesis testing.
First, state the null hypothesis, or what we think might be true (this is a statement, not a question). Second we state the alternative hypothesis, or what would be true to disprove the null hypothesis. Third, we set the significance level, or how rigorous our proof must be. Fourth, we evaluate the test statistic. We will use a MATLAB function to do this. Finally, we make a decision based on specified levels of significance and explain the results. We could do this to test if students in morning classes make grades of 85 on average.

(2:18)
Now I’ll demonstrate the steps of hypothesis testing using a more simple distribution than our student grade example. Even though we cannot measure all instances of the population, we can hypothesize it follows a normal distribution with a mean of 0 and a standard deviation of 1, and we can sample a data point from it.

(2:37)
The histogram of the distribution would look like this, with most points near 0 and fewer the further away from 0. Remember that N(0,1) is the normal distribution (or bell curve), with mean (mu) of 0 and unit standard deviation (sigma) of 1. 95% of the values would be between -1.96 and 1.96. That means that 5% would be outside of this range.

(3:07)
The null hypothesis is that this point that we sampled is from an N(0,1) distribution. The alternative hypothesis is that the sample point is not from an N(0,1) distribution. If the point is outside of the -1.96 to 1.96 range, we reject the null hypothesis in favor of the alternative hypothesis at the 0.05 significance level, which corresponds to the 5% of the values outside of the range.

(3:37)
Let's use some real numbers as a demonstration. Given the mystery population we hypothesized follows a normal distribution with mean of 0 and a standard deviation of 1, and we sample a value from the population and find its value is 3.0, and we remember that if the distribution is N(0,1), 95% of the values will be between -1.96 and 1.96, we accept the alternative hypothesis that this is not an N(0,1) distribution because 3 is outside of the 0.05 or 5% significance interval of -1.96, 1.96. So what is the probability of wrongly accepting the alternative hypothesis (the p-value)? We know that 0.27% of the values are outside the range -3 and 3, so the probability of picking a value outside of this range is 0.0027. This is much lower than our 0.05 cutoff point.

(4:40)
Stated formally, based on sample x = 3, we want to know whether it's likely that the unknown population has 0 mean and standard deviation of 1. First we state the null hypothesis: the population has 0 mean and standard deviation of 1. Then we state the alternative hypothesis: the population does not have 0 mean and standard deviation of 1. We specify the significance level; we want only a 5% chance of being wrong if we accept the alternative hypothesis (alpha = 0.05). We evaluate the test statistic. In this case, our sample point is 3. We make a decision: the probability of picking a value more extreme than 3 from N(0,1) is 0.0027, which is less than our cutoff of 0.05. So we reject the null hypothesis in favor of the alternative. The mystery population does not have 0 mean and a standard deviation of 1.

(5:43)
The p-value is the probability of getting a test statistic value more extreme than the value from the sample assuming the null hypothesis was actually true. This is the probability of a bad draw, or in other words, it is the probability of incorrectly rejecting the null hypothesis. We assume the null hypothesis is true - that is, that the mystery population really has 0 mean and a standard deviation of 1. The p-value is just the probability that a randomly drawn value will be outside the

range -3 to 3, which is just 0.0027. So if the p-value is less than the significance level, we reject the null hypothesis in favor of the alternative hypothesis.

(6:28)
Let’s go through another example where the sample we measure is closer to the mean of our N(0,1) hypothesis. In this case, it’s 1.0. Remember, the null hypothesis says that our mystery population is N(0,1). We know that 31.7% of the values are outside the range -1 to 1, so the probability of picking a value outside -1 to 1 is 0.317. If the mystery population was really N(0,1), we would have less than a 31.7% chance of picking a value that is more extreme than 1. This isn’t good enough to disprove our guess of N(0,1) and accept the alternative hypothesis. But, we can’t declare that the null hypothesis is true either. There are too many other possibilities. For example, if the mean of the mystery population was 2, the histogram would look like this. Notice the sample of 1 is just as close to the mean of 2 as it was to the mean of 0. This sample would not reject N(0,1) OR N(2,1). Both are equally likely given the sample, so we cannot reject the null hypothesis, but we can’t say it’s true either.

(7:49)
Again, formally: based on sample x = 1, we want to know whether the unknown population is likely to have 0 mean and standard deviation 1. We state the null hypothesis - that the population has 0 mean and a standard deviation of 1. We state the alternative hypothesis - the population does not have 0 mean and standard deviation of 1. We specify the significance level - we want only a 5% chance of being wrong if we accept the alternative hypothesis. We evaluate the test statistic. In this case the sample was 1. Finally, we make a decision. The probability of picking a value more extreme than 1 from N(0,1) is 0.317, which is greater than our cutoff of 0.05. We don’t have enough evidence to reject. We don’t have enough evidence to reject the null hypothesis and accept the alternative hypothesis, but we also have not proved the null hypothesis; only that it is possible.

(8:45)
The strategy of hypothesis testing is analogous to trial by jury, or that you’re innocent until proven guilty. So first we state the null hypothesis that the defendant is innocent. Then we state the alternative hypothesis that the defendant is guilty. Third, we specify the significance level, or what is beyond reasonable doubt. Fourth, we evaluate the test statistic - this is; the jury evaluates the evidence. Finally, we make a decision. Guilty, or that the jury is convinced beyond reasonable doubt, or not guilty: there’s not enough evidence. The key point is this: defense lawyers don’t have to prove their clients are innocent, only that there is not enough evidence to convict. Prosecutors must prove guilty beyond a reasonable doubt. Notice neither proves the defendant is innocent.

(9:38)

So, acquittal does not mean that we accept the null hypothesis. If our p-value is less than the significance level, it means that we don't have enough evidence to convict - that is, we don't have enough evidence to accept the alternative hypothesis. We don't have the ability to completely measure our mystery population; we can only find examples to disprove our hypothesis about it. We cannot find examples to prove it. Going back to our morning science class example, if we believe we have a good chance to make an 85 because we believe all morning science students' grades follow an 85, 5 distribution, then we can sample real students to see if they refute this hypothesis, but we cannot sample to prove the hypothesis. In summary, there are five steps in hypothesis testing: you state the null hypothesis, you state the alternative hypothesis, you set the significance level, you evaluate the test statistic, and then you make a decision. Steps 1 through 3 are done on paper. However, to actually make a decision, we need to evaluate the test statistic and compute a p-value. This we will with MATLAB and it is described in the next video.