Video: Hypothesis Testing Basics (10:50)

(0:00):

Statistically correct information can enable good decision making. For example, if you thought students who took science classes in the mornings were better than students who took science classes in the afternoon. This might influence your schedule and ultimately your academic success only if it were true. We need a correct method for testing our hypothesis. The goal of hypothesis testing is to make an informed decision about a population based on sample data. So we formulate conclusions based on 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 data is or how much we should trust the conclusions. For a demonstration, suppose we assume all mornings 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 pull one student from the morning science classes and see if it fits our belief about the distribution. Based on the evidence provided by the sample x, what can we conclude about the mystery population? It either supports or refutes our belief that the score is a 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 we would have evidence for our scheduling decisions.

(1:35):

There are 5 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, se 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 the specified level of significance and explain the results. We could do this to test if students in morning classes make grades of 85 on average. Now I will demonstrate the steps of hypothesis testing by using a more simple distribution than the student grade example. Even though we cannot measure all instances of the population, we can hypothesize that 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. The histogram of the distribution will look like this with most points near 0 and fewer the further away from 0. Remember that N(0,1) is the normal distribution (histogram is a bell curve) with a mean of 0 and a standard deviation of 1. 95% of the values would be between -1.96 and 1.96 which means 5% of the data would be outside of this range. The null hypothesis is that this point that we sampled is from the N(0,1) distribution. The alternative hypothesis is that the sample point is not from the N(0,1) distribution. If the point is outside the of [-1.96, 1.96] range, we reject the null hypothesis in favor of the alternative at the 0.05 significance level which corresponds to the 5% of the values outside the range.

(3:37):

Let's use some real numbers as a demonstration. Given a mystery population we hypothesize that it follows a normal distribution with a mean of 0 and a standard deviation of 1 and we sample a value from the population and find its value is 3 and we remember that is the distribution is N(0,1) then 95% of the values will be between -1.96 and 1.96. We accept the alternative hypothesis since 3 is outside the 0.05 significance level of -1.96 to 1.96. What is the probability of wrongly accepting the null hypothesis, or the p value? We know that 0.27% of the values are outside the range -3 to 3, so the probability of picking a value outside the -3 to 3 range is 0.0027. This is much lower than our 0.05 cutoff point. Stated formally, based on sample x = 3, we want to know whether it's likely that the unknown population has a mean of 0 and a standard deviation of 1. First we state the null hypothesis: the population has a mean of 0 and a standard deviation of 1. Second we stated the alternative hypothesis that the population does not have a mean of 0 and a standard deviation of 1. We specify the significance level. We want only a 5% chance of being wrong if we accept the alternative hypothesis so the alpha is 0.05. We evaluate the test statistic. In this case are sample point is 3 and we make a decision. The probability that picking a value more extreme that 3 from N(0,1) is 0.0027 which is less than our cutoff 0.05. So, we reject the null hypothesis in favor of the alternative, the mystery population does not have a mean of 0 and a standard deviation of 1.

(5:44):

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 or the probability of a bad draw or in other words the probability of incorrectly rejecting the null hypothesis. We assume the null hypothesis is true that the mystery population has a mean of 0 and a standard deviation of 1. The p value is just the probability that a randomly drawn point is outside the -3 to 3 range which is just 0.0027. If the p value is less than the significance level, reject the null hypothesis in favor of the alternative hypothesis. 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, its 1. 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 because 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 that it is true either.

(7:49):

Again formally, based on the sample $x = 1$, we want to know whether the unknown population is likely to have a mean of 0 and a standard deviation of 1. We state the null hypothesis that the population has a mean of 0 and a standard deviation of 1. Then we state the alternative hypothesis that the population does not have a mean of 0 and a standard deviation of 1. We specify the significance level which is we only want 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 the N(0,1) is 0.317 which is greater than our cutoff of 0.05, so we don't have enough evidence to reject the null hypothesis. We don't have enough evidence to reject the null hypothesis and accept the alternative hypothesis, but we have also not proved the null hypothesis only that it is possible. The strategy of hypothesis testing is analogous to trial by jury or innocent until proven guilty. First we state the null hypothesis defendant is innocent. The we state the alternative hypothesis the defendant is guilty. Third we specify the significance level which is "beyond reasonable doubt". Fourth we evaluate the test statistic, evaluation of the evidence. Finally, we make a decision. Guilty is when the jury is convinced beyond a reasonable doubt and is able to convict. Not guilty is there isn't enough evidence to go beyond reasonable doubt. The key point is this: defense lawyers don't have to prove their clients are innocent, they only have to prove there is not enough evidence to convict. Prosecutors must prove guilt beyond reasonable doubt. Notice neither proves the defendant is innocent

(9:39):
Acquittal does not mean that we accept the null hypothesis. If our p value is less than the significance level it means we don't have enough evidence to convict i.e, we don't have enough evidence to accept the alternative hypothesis. Without the ability to completely measure our mystery population we can only find examples to disprove our hypothesis and 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 a N(85,5) distribution than we can sample real students to see if they refute this hypothesis, but we cannot sample to prove the hypothesis. In summary there are 5 steps to hypothesis testing. You state the null hypothesis, then you state the alternative hypothesis, then you set the significance level, then you evaluate the test statistic, and finally you make a decision. Steps 1-3 are done on paper. However, to actually make a decision we need to evaluate the test statistic and compute a p value which can be done by a computer and we will use MATLAB.