Estimating p-values

Background

For this exercise, we will begin to familiarze ourselves with a concept that will appear in the part of this course on linear regression. Suppose you are measuring the speeds of cars traveling on a particular highway, and assume that you know what they follow a normal distribution, with a mean of 60 and standard deviation 5.

You then want to know: how likely am I to see a car travelling at 70 mph or more? This value, which is called the *right tailed p-value* can be found by taking the *area under the curve* of the probability distrbution function, starting from 70 and going to infinity. This area is shown in green in the below figure. In our case, the probability distribution function is the normal curve with mean 60 and standard deviation 5.

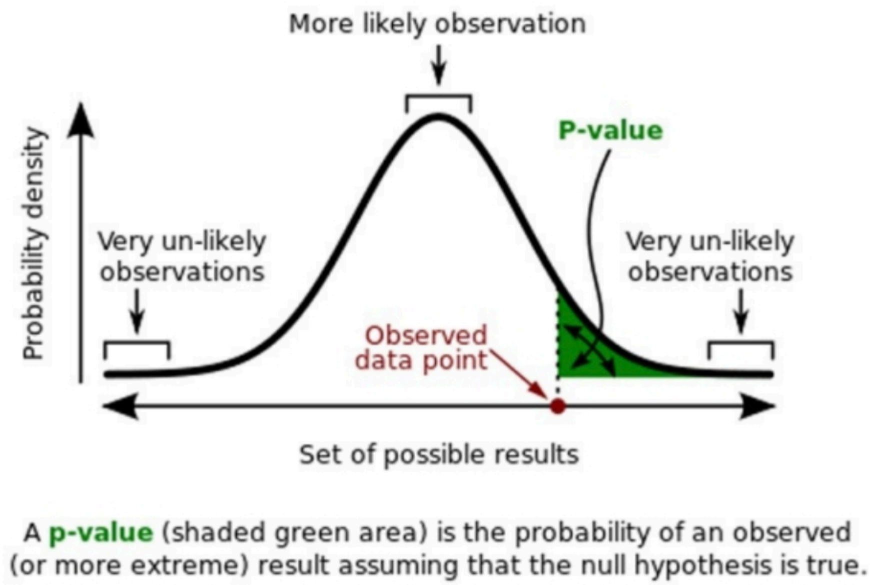

Image source [Stanford Medicine]

It turns out it is mathmatically difficult to take the area of the normal distribution function. In other words, there is no easy formula to solve for this area. However, we can make a numerical approximation.

Assignment

Part 1

Recall that the probability of an event is equal to the amount of times that event can occur, divided by the total number of possible events. Use this definition to approximate the right tailed p-value for various speeds. Your solution should contain a function that takes a speed as input and returns the p value.

What happens if you try to find the right tailed p-value for a very large speed? How do you interpret this?

Part 2

Write a second function to compute the left tailed p-value. This is the probablity of observing a speed less than or equal to a particular value. As in part 1, what happens if you try to find the left tailed p-value for a very small speed, and how do you interpret this?

> Note, you can check your answers for the left tailed p value against the values in in this table, or by using this calculator.
>
> However, you will first need to convert the speed to a standard normal variable (mean 0 and standard deviation 1). To do so, subtract the mean and divide by the standard deviation.
>
> So to convert 70 to a standard normal variable, you would take **z** = (70 - 60) / 5 = 2.
>
> Those resources contain the left tailed values. To get the right tailed values, simply subtract from 1.

Part 3

Find the two tailed p-values (sum of right and left) for the specific values we saw in lefture related to 1, 2 and 3 standard deviations. You can use the true mean and standard deviation (60 and 5) in the calculation, as opposed to the meand and standard deviation of the sample. Your results should validate the confidence levels associated with 1, 2, and 3 standard deviations.