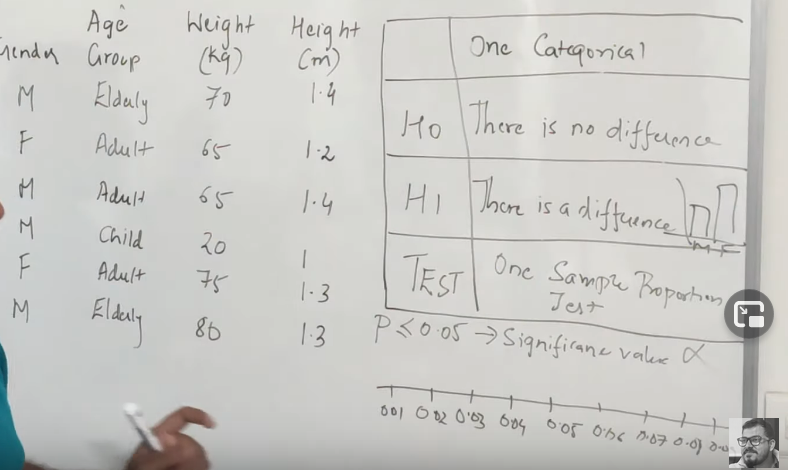
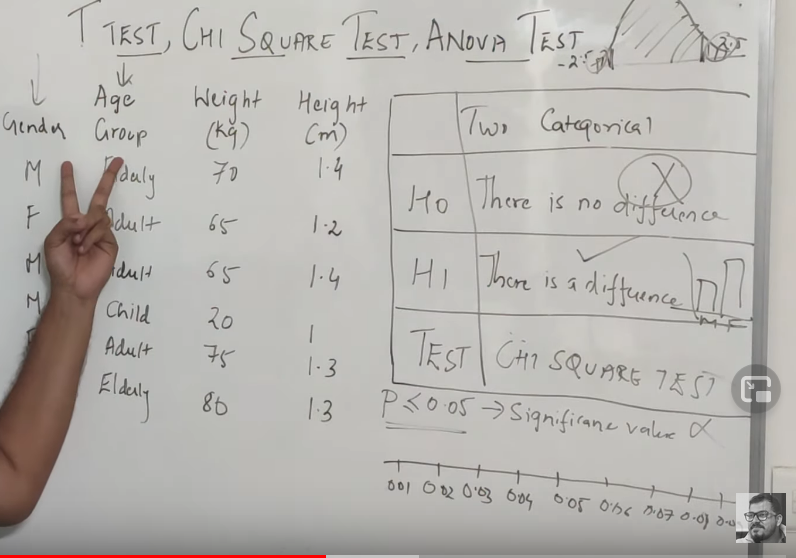
**One Catgorical Data – One Sample Proportion Test** :

1. Considered gender column alone

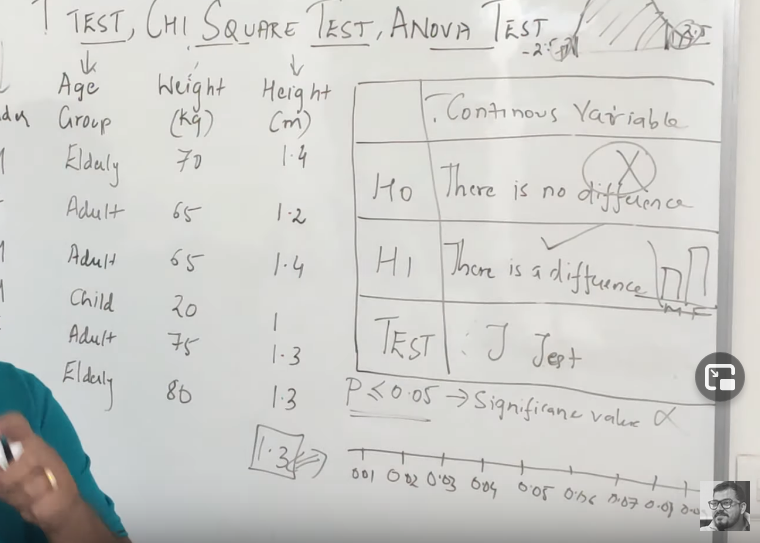


Two Categorical Feature

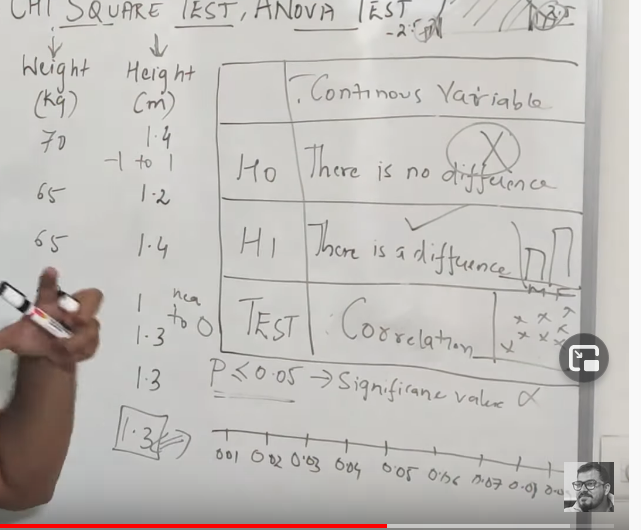
1. Gender and Age is considered



One Continuous Variable :



Two numerical features :



More than 2 Categories :

Inside a particular column if it has more than 2 variables than ANOVA Test.

Discrete Uniform Distributions :

Definition:

A random variable X has a discrete uniform distribution if each of the n values in its range: x1,x2,x3....xn has equal probability:

(*𝑥𝑖*)=1/*𝑛*

The mean is simply the max and min value divided by two, just like the mean of two numbers.

*𝜇*=(*𝑏*+*𝑎*)/2

With a variance of:

*𝜎*2=(*𝑏*−*𝑎*+1)2−112

Continuous Uniform Distributions :

A continous random variable X with a probability density function is a continous uniform random variable when:

(*𝑥*)=1(*𝑏*−*𝑎*)*𝑎*<=*𝑥*<=*𝑏*

This makes sense, since for a discrete uniform distribution the f(x)=1/n but in the continous case we don't have a specific n possibilities, we have a range from the min (a) to the max (b)!

The mean is simply the average of the min and max:

(*𝑎*+*𝑏*)2

The variance is defined as:

*𝜎*2=(*𝑏*−*𝑎*)2/12

####That's it for Uniform Continuous Distributions. Here are some more resource for you:

1.)[http://en.wikipedia.org/wiki/Uniform\_distribution\_%28continuous%29](http://en.wikipedia.org/wiki/Uniform_distribution_%28continuous%29" \t "_blank)

2.)[http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.uniform.html](http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.uniform.html" \t "_blank)

3.)[http://mathworld.wolfram.com/UniformDistribution.html](http://mathworld.wolfram.com/UniformDistribution.html" \t "_blank)

Poisson distrubtion :

A poisson distribution focuses on the number of discrete events or occurrences over a specified interval or continuum (e.g. time,length,distance,etc.). We'll look at the formal definition, get a break down of what that actually means, see an example and then look at the other characteristics such as mean and standard deviation.

Formal Definition: A discrete random variable X has a Poisson distribution with parameter λ if for k=0,1,2.., the probability mass function of X is given by:

*𝑃𝑟*(*𝑋*=*𝑘*)=*𝜆𝑘𝑒*−*𝜆𝑘*!

where e is Euler's number (e=2.718...) and k! is the factorial of k.

The Poisson Distribution has the following characteristics:

1.) Discrete outcomes (x=0,1,2,3...)

2.) The number of occurrences can range from zero to infinity (theoretically).

3.) It describes the distribution of infrequent (rare) events.

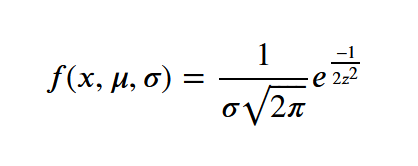
4.) Each event is independent of the other events.

5.) Describes discrete events over an interval such as a time or distance.

6.) The expected number of occurrences E(X) are assumed to be constant throughout the experiment.

Normal Distribution :

The distribution is defined by the probability density function equation:



Where:

*𝑧*=(*𝑋*−*𝜇*)*𝜎*

where: μ=mean , σ=standard deviation , π=3.14... , e=2.718... The total area bounded by curve of the probability density function equation and the X axis is 1; thus the area under the curve between two ordinates X=a and X=b, where a<b, represents the probability that X lies between a and b. This probability can be expressed as:

*𝑃𝑟*(*𝑎*<*𝑋*<*𝑏*)

Let's look at the curve. The normal distribution has several characteristics:

1.) It has a lower tail (on the left) and an upper tail (on the right)

2.) The curve is symmetric (for the theoretical distribution)

3.) The peak occurs at the mean.

4.) The standard deviation gives the curve a different shape:

-Narrow and tall for a smaller standard deviation.

-Shallower and fatter for a larger standard deviation.

5.) The area under the curve is equal to 1 (the total probaility space)

6.) The mean=median=mode.

For the normal distribution, we can see what percentage of values lie between +/- a standard deviation. 68% of the values lie within 1 TSD, 95% between 2 STDs, and 99.7% between 3 STDs. The number of standard deviations is also called the z-score, which we saw above in the PDF.

# Bayes Theorem

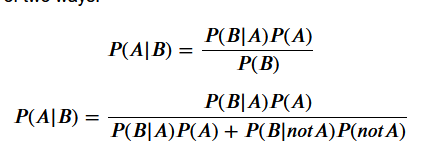
In this Statistics Notebook we will be going over Bayes Theorem and an example of how to use it. This theorem will be important to understand in many situations and we will be using it for Machine Learning in certain lectures.

We will start by describing the equation for Bayes Theorem and then complete an example problem. There won't be a SciPy portion to this lecture because this is really more of a math overview than a discussion on a distribution.

### Overview

In probability theory and statistics, Bayes' theorem describes the probability of an event, based on conditions that might be related to the event. Bayes Theorem allows us to use previously known information to asess likelihood of another related event.

You will usually see Bayes Theorem displayed in one of two ways:



Here P(A|B) stands for the Probability that A happened given B has occured. These are both the same, in the second case, you would use this form if you don't directly have the Probability of B occuring on its own.