Understanding p-value

This is not a formal or complete text.

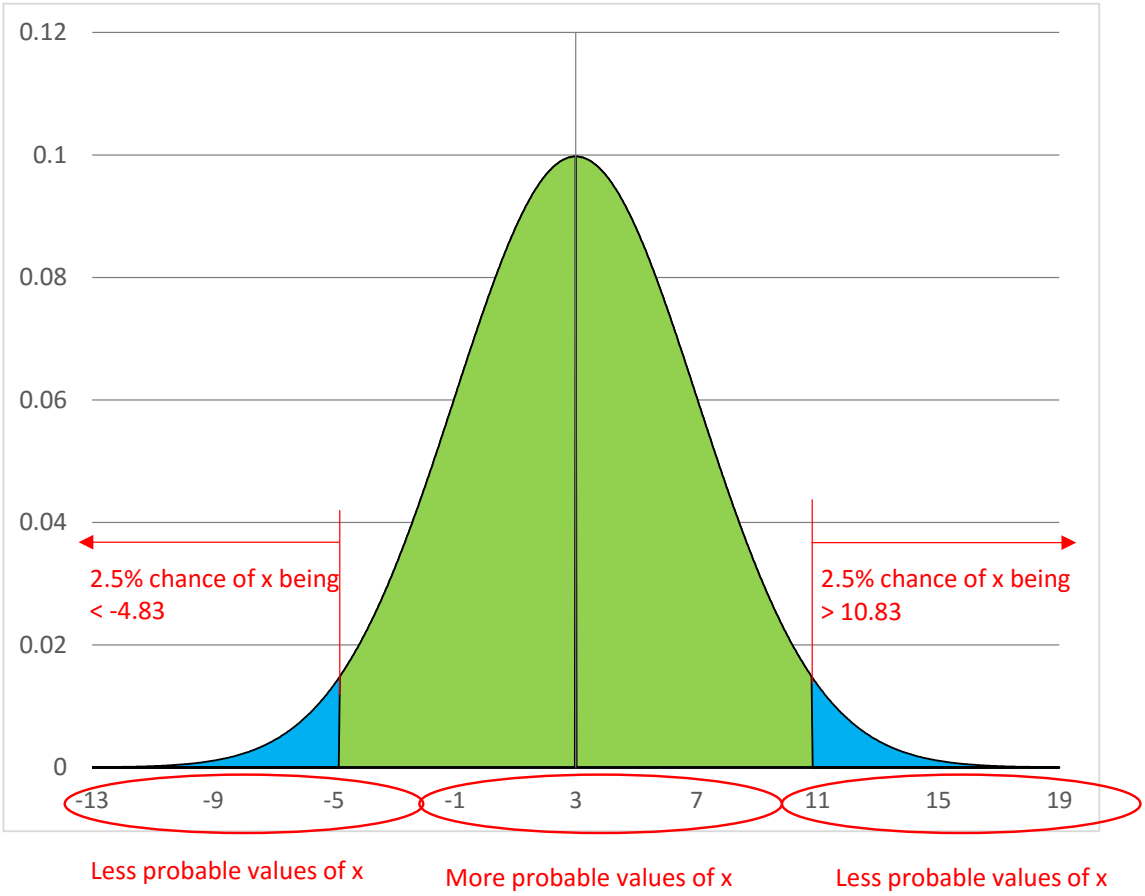
To be read for understanding purpose only in addition to a main formal text explaining the terms.

It just tries to help understand the concept.

- Amit Bhola, *amit_aromatic@yahoo.com*
March 2017

Understanding p-value

Suppose that you know that some variable (say x) is having some known distribution, say e.g. normal distribution. Then if we know it's defining parameters, in this case, mean and standard deviation, we can calculate any sort of probabilities. Let us assume, it is known that mean is 3 and std dev. is 4.
Usual criteria of testing hypothesis is at 5% alpha value, i.e. 2.5% each side for a 2 way test. Knowing the distribution parameters, we can know that the values of x which define lowest and highest 2.5% values of all possible values of x are -4.83 and 10.83 (These are seen as near -5 and 11 below) :-



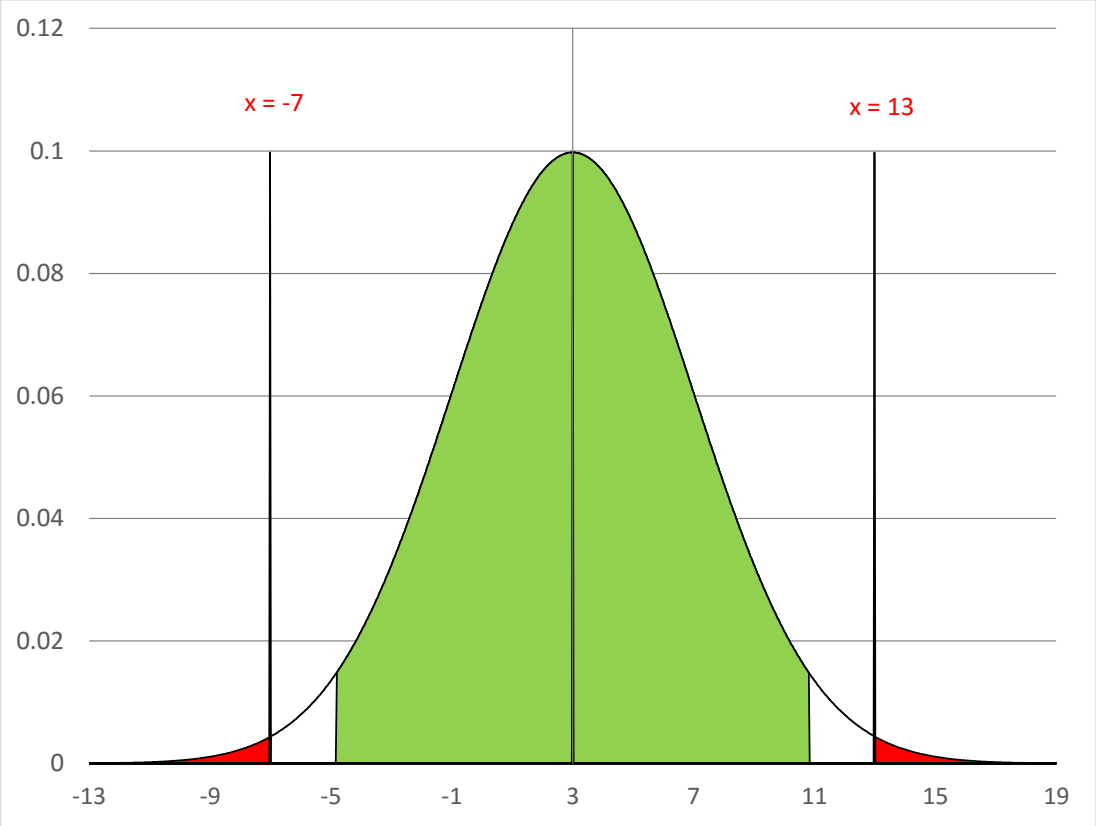
< Well this text assumes that you know the concept of probability distributions and hypothesis testing. So it does not explain the depth of these. >

< Focus is only on interpreting the meaning or significance of "p-value" which is a usual outcome of the a hypothesis test. >

Next, suppose that in some two tailed hypothesis test, we get a p-value as 0.0124 What does it mean?

It means that the observed scenario (situational data) that is being tested is one of the following two vertical bars :-

Either x= -7 OR x = 13



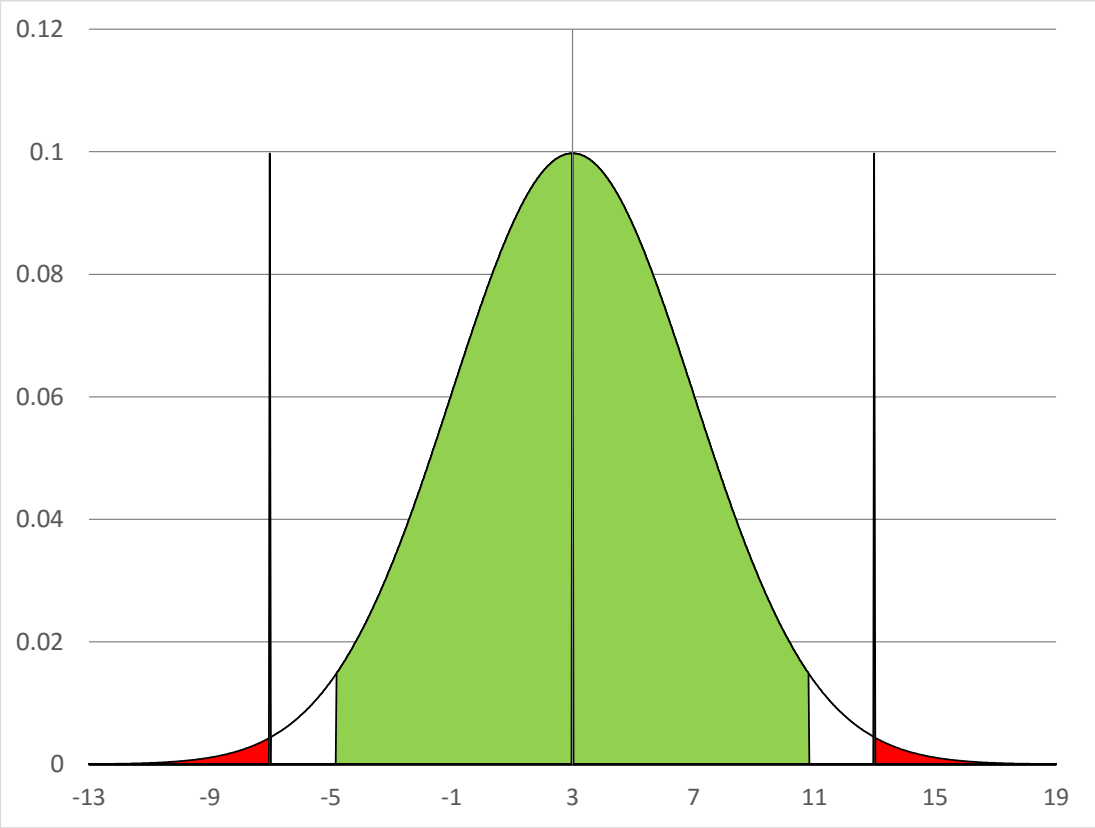
Why so ? Why only these two x values ?

Because, -7 is such value of x below which lie 0.62% of all possible values of x
and 13 is such value of x above which lie 0.62% of all possible values of x

Why only 0.62% ? Where does this figure come from ?

Because, p / 2 = 0.0124/2 = 0.0062 , or in other words 0.62%

=> p value is the probability indicated by the red zone above



Okay, p-value is red zone. So what ?

Clearly it means, that the observations are in one of the the extreme zones. The extreme zones are zones of very less probability. So the observations are rare. It means that the actual underlying phenomenon is not following the assumed distribution of (a) type : Normal with hypothysed (b) mean = 3 and (c) Std dev = 4 *very nicely/goodly*

Hmmm…

Should you be happy or sad about it?

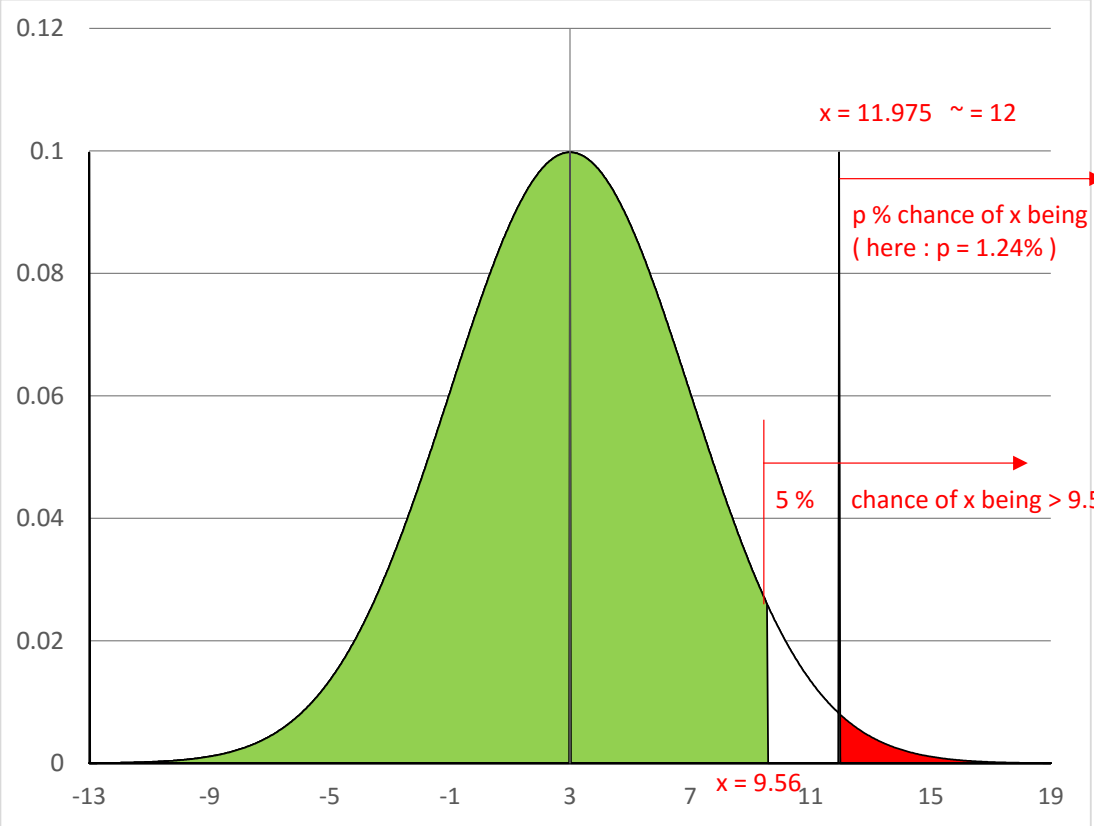
That depends what are you testing.

Imaging that you are the designer of two bearings of Design 1 and Design 2.
You claim by your design that the Design 2 bearing is better than Design 1 on some performance parameter, say bearing life, by 3 units (e.g. 3000 hours). Next ***assume*** that you ***know*** that the bearing life *'difference'* in general follows the normal dist. Curve as shown above of mean = 3 and Std. dev. = 4 (say)
< This is just to explain the thing. Assume for time being that this 'assumed to be known model' is a universal law >

Next you test some bearings and perform some hypothesis test like test for difference of two means or ANOVA.
And the test yields the p-value as above = 0.0124.

Then, should you be happy or sad?
The result (red zone) indicates that the observed difference is so high or so less that the so believed universal law is being followed with *very low probability.*

< let us include a case correction here >
More suitably (since here you know that Design 2 will yield a higher life always),
you performed a one way test, then same p value would be interpreted like (5% alpha and p% zone are on one side):-



Now should you be happy or sad?

The observed values are still rarer than the *considered hypothesis test cut-off* of alpha = 5% (of the extreme values)

So it means any or both of the following :-

1) The tested design 2 parts were rare good gene which outperformed design 1 better than normally expected
or alternatively that the design 1 parts were rare bad genes. They both conform to the universal law, just that they (i.e. the actual situational data tested) happens to be the *rarity within the possibility.*

2) The believed universal law about the difference of the performances is probably not really universal
More verification is required about the assumed underlying mathematical model.

But should you be happy or sad?

That is really a contextual discretion. How you want to interpret the results and how much confidence you have on (a) the measurements of experiments (statement 1 above) or (b) the belief in truthness of the law, i.e. the hypothesis (statement 2 above)

The testing output "p-value" is just a support to you to think further. It is not a binary yes/no or good/bad output! One can feel tempted to make an accept/reject decision by comparing p-value with alpha, but in this regard also note, that considering alpha = 5% too is merely just a convention!