Course material

Course:

ENG606 / PHYS 442

Video:

DOE_lesson2_part1

Concepts (extracted from automatically generated subtitles):

Normal distribution. Degree of freedom. Center of your distribution. Standard statistics. Given distribution. Greek letter t. So nice confidence interval. Typical measurement of the variability of a signal. Hard use of very important uses. Number of elements. Most important serums. Cumulative distribution. Probability distribution. Standard deviation. Root of the degree of freedom.

to video sequence search
(within ENG606 / PHYS 442.)

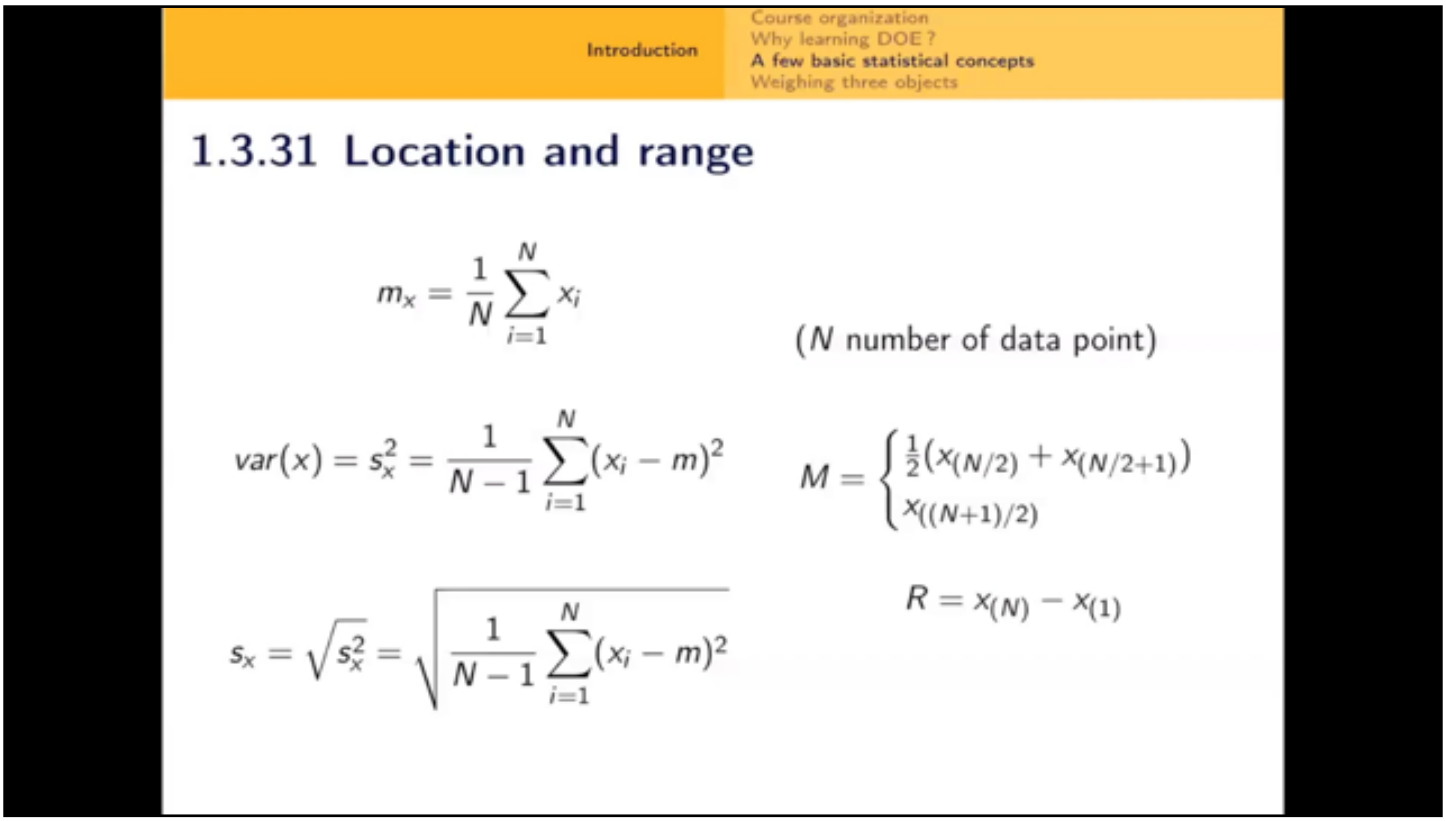
to video

...

notes

summary

0m 0s

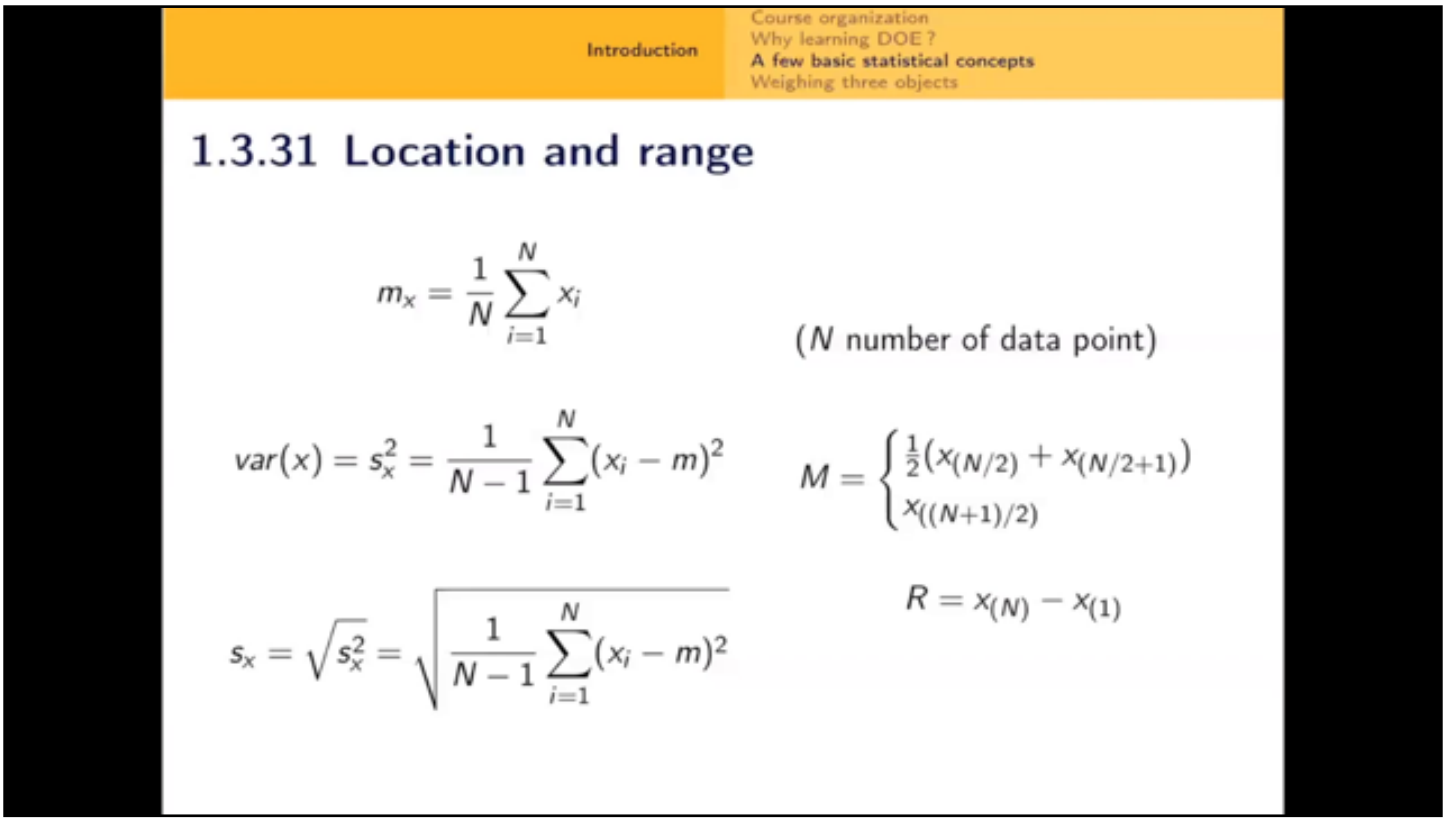


These subtitles have been generated automatically So today I will finish the introduction of the course, a few more elements because I was very, very quick at the end of the last course, so I will come back on a few statistical elements. And after I will present to you a first case, a simple case, which let us to really appreciate what DOA is doing, what is the advantage of DOA. Data, there are all the time a few things that we can calculate in all that standard statistics. So one is the mean, so the mean is the sum of the element divided by the number of elements, very evident. I don't think I need to dig into that. So something which will be very important in this course, it's important in the statistics is also an element that remember I mentioned last week I mentioned this three person, a Galton person and Feisher. So the variance, if I'm correct, the person who defend, it's appear very evident for us today, but it was a time where variance doesn't take it. So the variance is a typical measurement of the variability of a signal, the variability of a data. So if you want to sum some variability, you cannot stay with the difference with the mean because around the mean, the difference will be zero when you sum them, it's definition of the mean, so it's why you take the square. And after you sum them, and you divide it by n minus one, you have to remember in some situation, we divide it by n, some situation we divided by n minus one. So the startup way of saying that if it's a sample of your population, you have to use the numbers of data you have minus one, because you do not have all the sample, and the degree of freedom is n minus one. When

notes

summary

0m 1s



you have all the population, it's not so high and you have all, but the real way of calculating the variance should be divided by n. But this variance is really very important, we will have it all the time, we will define the, because it's left defines the uncertainty and it's what we are dealing with. And the units of the variance is a square of the unit of the measurement, so it's interesting to take the root of the variance, the standard negation, which is in the same unit as the measurement and through present range around the variance. Just be, I will come back on it, but just remember that plus minus the standard deviation, it's not a so nice confidence interval, because if you accept a normal distribution, it's just two thirds of your population, choose 64 something percent of all your range. So sometimes people publish experimental data saying plus minus standard deviation, it's okay, we all do that, but remember that it's very, very small confidence interval. And after there are two things that are also interesting to calculate. One is the median, so the position at which you have 50% of your point at left and 50% of the tribe. When you have a normal distribution, usually the median corresponds to the average, but it's a distribution, it's not symmetric, it's not the case. And you have, for calculating it, it will depend if you have an odd or a far number of data, so if you have odd, so the mid range is the one in the center, if not is the middle between the two half, etc. And you have something we call the range. And look also, it's a notation, when you have a series of number and that you put parenthesis, that means that you are ordering them, sorting them, usually it's not, usually it's from the smallest to

notes

summary

1.3.31 Location and range

$$m_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

(N number of data point)

$$var(x) = s_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - m)^2$$

$$M = \begin{cases} \frac{1}{2}(x_{(N/2)} + x_{(N/2+1)}) \\ x_{((N+1)/2)} \end{cases}$$

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - m)^2}$$

$$R = x_{(N)} - x_{(1)}$$

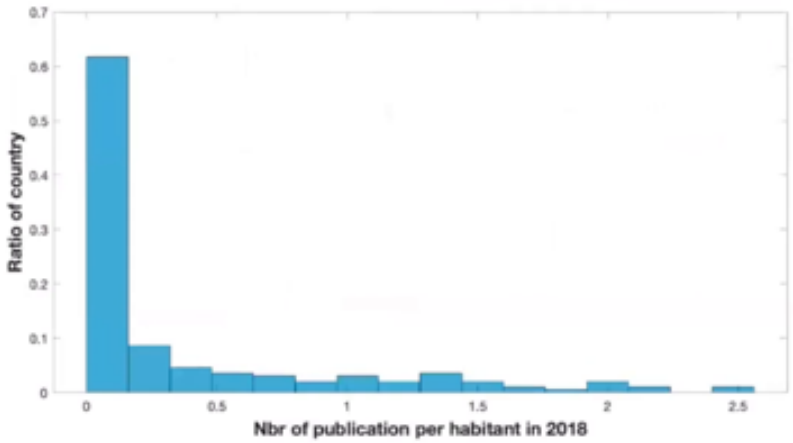
the highest,

notes

summary

1.3.32 Quartiles and histogram

	Mean	Deviation	Variance	Q25	Median	Q75
2000	0.18	0.36	0.13	0.0025	0.018	0.11
2018	0.37	0.57	0.33	0.0128	0.08	0.48



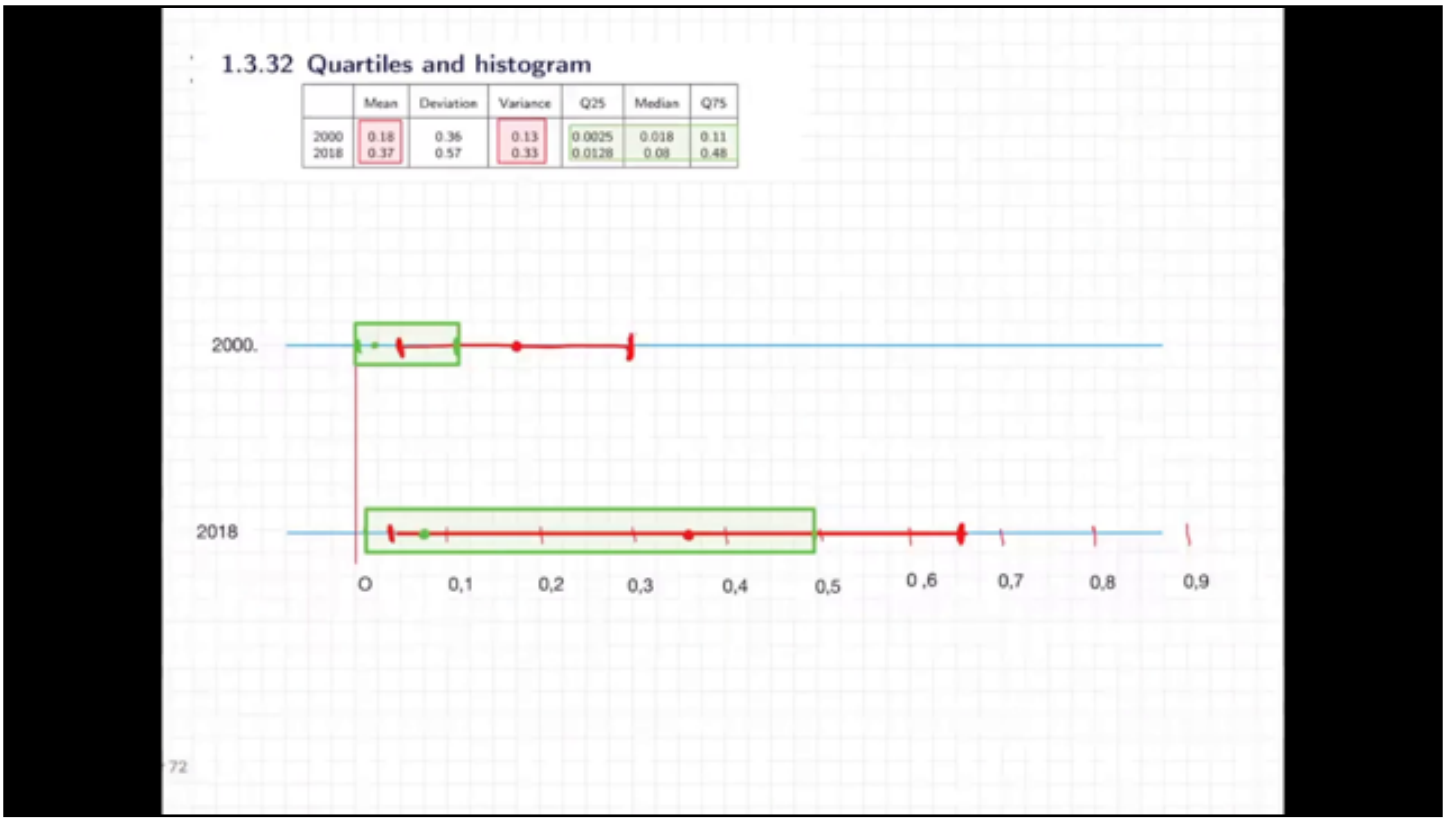
72

it could be not again, but in any case, when you have this notation, it's because you have your points that are odd. And if you play a little bit with this data about publication, you see then when you are ordering things, it starts to, some insight can be realized. No, it's here that I have to move.

notes

summary

5m 1s

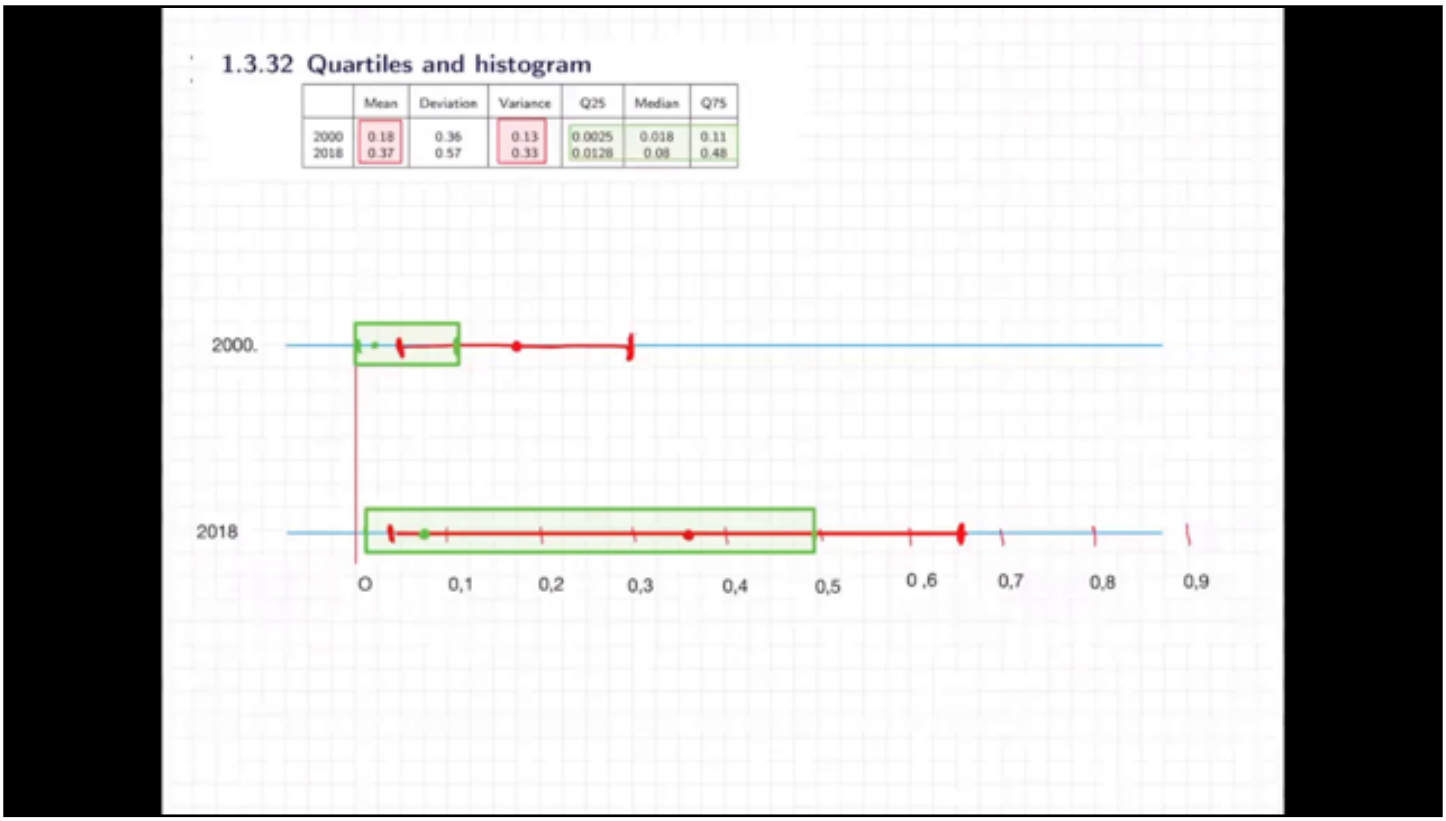


After, so the first, when you have the data, the first thing is to calculate the average, the standard deviation, the range, the mid range are things that are very interesting. After, just going, just one little bit in the statistic concept, you have what we call the quantile, so that means to try to understand what is the percentage of points that have some conditions that are at the left of the point, it's the right of the point, they are around the middle. So usually you can calculate very rapidly and is the exercise, you can make it with the category, or maybe several dataset, or several category in your dataset, mean the variance, when you have the variance, it takes the roots, and when you have also, you can be interested to calculate the median and to have so Q25 on Q75, they are the quartile, so you have your population, and that means that you are, the first quartile is from the extreme left to 25% of your population, from 25 to 50%, we call that Q25, and from the average, one more addition, one more quarter, we call it Q75, and statistician appreciate having this two value, because it's giving you between Q25 and Q75, you have 50% of your population, and you have the center of your distribution, and when you look at those things, you see that without making a graphic, we start to see something, now it's already an information, some numbers, typically I make a few, I believe I make a small analysis

notes

summary

5m 27s

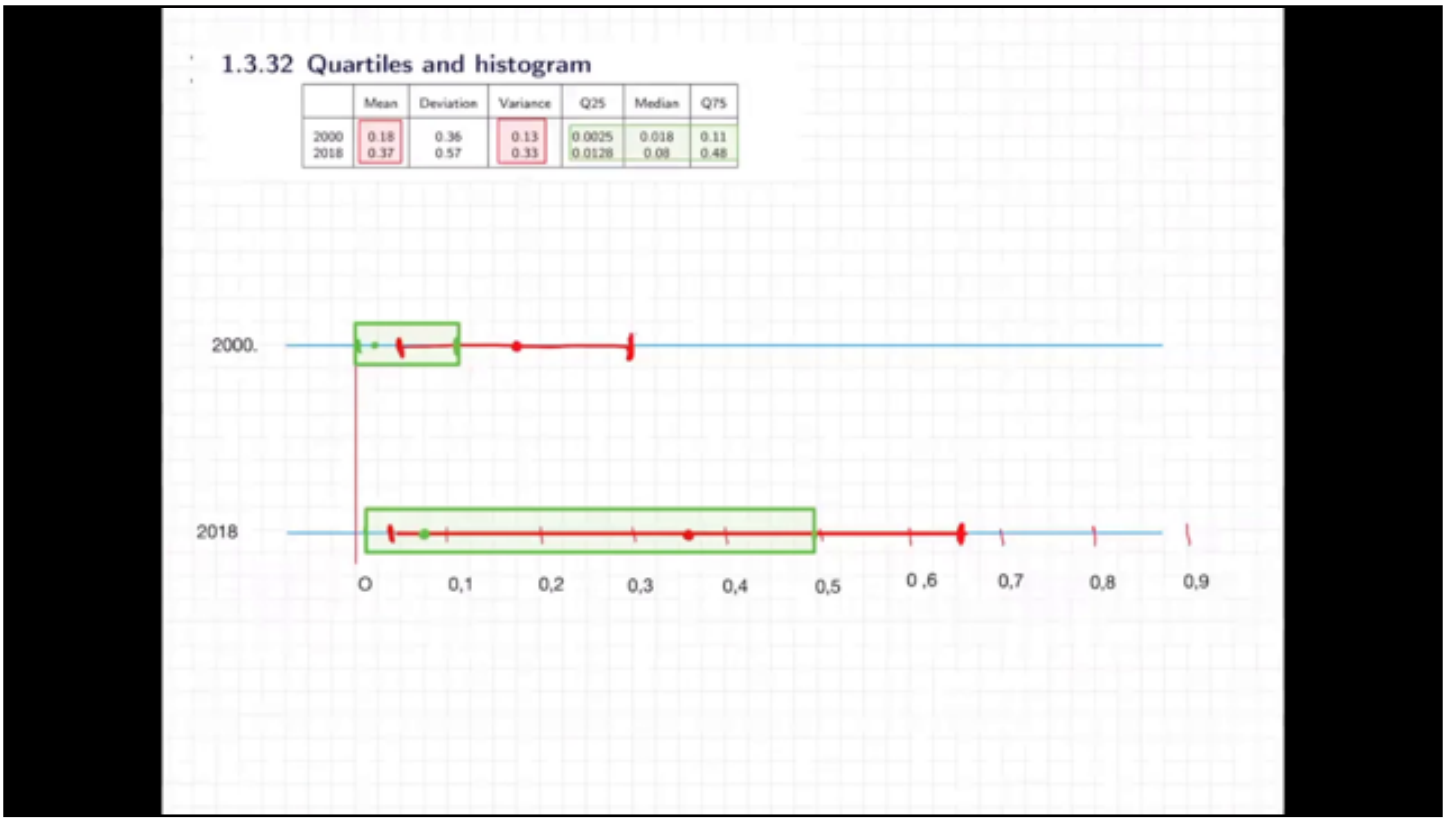


of this data, so with the means, the deviation, the standard deviation, you have this red interval, it gives you already a sort of position of your population

notes

summary

7m 28s



in the axis of the variable, and with the quantile, you see where you have the population,

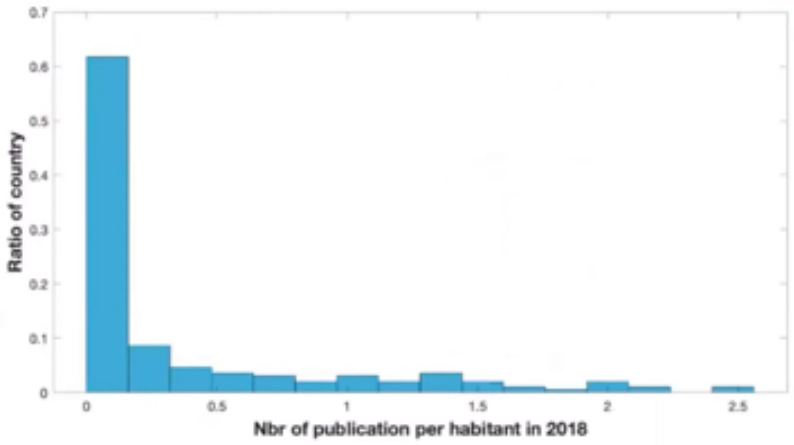
notes

summary

7m 45s

1.3.32 Quartiles and histogram

	Mean	Deviation	Variance	Q25	Median	Q75
2000	0.18	0.36	0.13	0.0025	0.018	0.11
2018	0.37	0.57	0.33	0.0128	0.08	0.48



and when you see that, typically the green frame and the red axis are not, and the equilibrium that means that you have a really non-symmetric situation, but what's the case, in our situation, you see that it was very evident in 2000, in 2018, it was a little bit more. So this is a way of rapidly, with a few numbers, see what you have, and you see that it was exactly the case here, man, on Instagram,

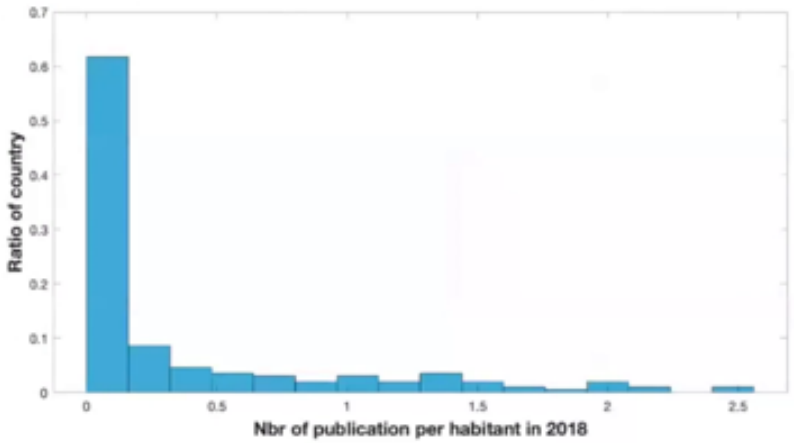
notes

summary

7m 55s

1.3.32 Quartiles and histogram

	Mean	Deviation	Variance	Q25	Median	Q75
2000	0.18	0.36	0.13	0.0025	0.018	0.11
2018	0.37	0.57	0.33	0.0128	0.08	0.48



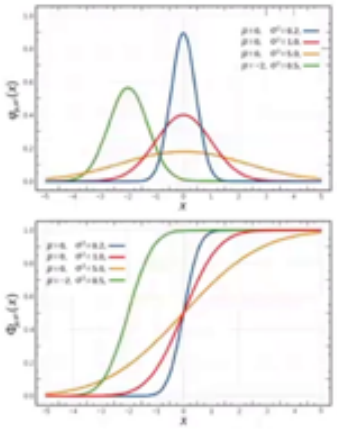
it's also another thing that's very interesting to make when you have data, is to make an Instagram of the data, so you see more of that. I forget to revise my term, but you know

notes

summary

8m 25s

1.3.33 Normal distribution $\mathrm{N}(\nu, \sigma)$



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

e^{-x^2}

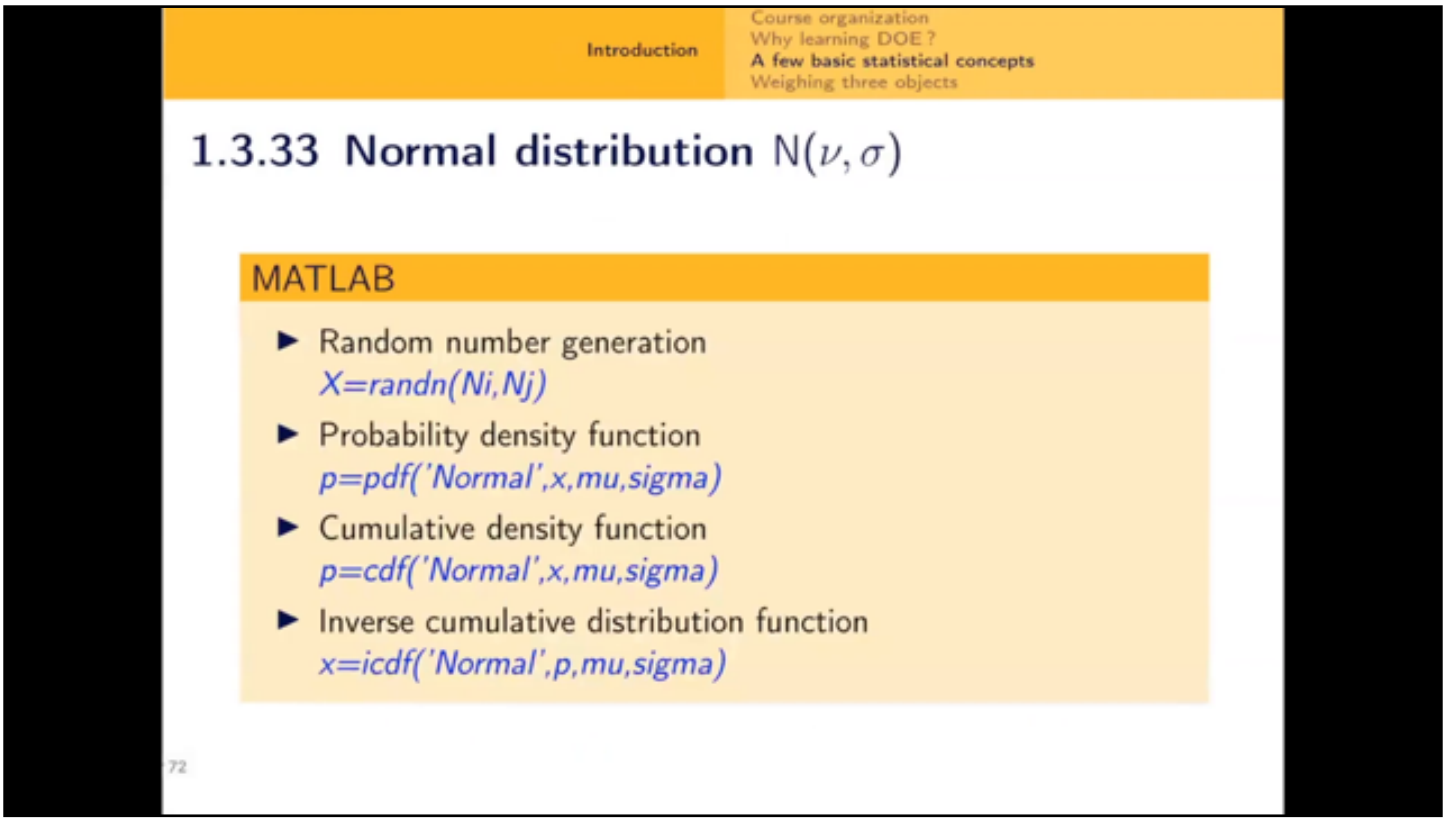
The Normal distribution is characterized by its bell-shaped curve, which is symmetric around its mean. It describes how values of a variable are distributed, with most of the data clustering around the central mean μ and decreasing in frequency as they move further away.
The spread of the distribution is determined by its standard deviation, σ.
The Normal distribution is fundamental due to its natural occurrence in many real-world phenomena and its properties, such as the central limit theorem, which states that the sum of a large number of random variables tends to follow a normal distribution, regardless of the original distribution of the variables.

that the numbers of beans, we call beans, no, the different column, are related to the number of data, because you want to have more, usually you would like to have a minimum, and the minimum of point per beans for having, for being meaningful, it's meaningful. You can check, perhaps, that it's something important. Those are different statistics that you can make on the dataset and let you see things. In science, in engineering, we make hard use of very important uses, and normal distribution, and also use a normal distribution is something. Clear imaging, the shape is normal, you already cannot write the mass-clear, please, without having seen a few times, normal distribution, but I would like to know, or I would like to insist, why, why is all the time the normal distribution? It's because the normal distribution is built on the fact that we say the error is random, there are no more information, it's just the random that provokes these changes of this distribution is built. So you know it very well, it's bell shape, it's that mean, which is position, you have widths, which is defined by the second parameter, which is sigma that corresponds

notes

summary

8m 38s

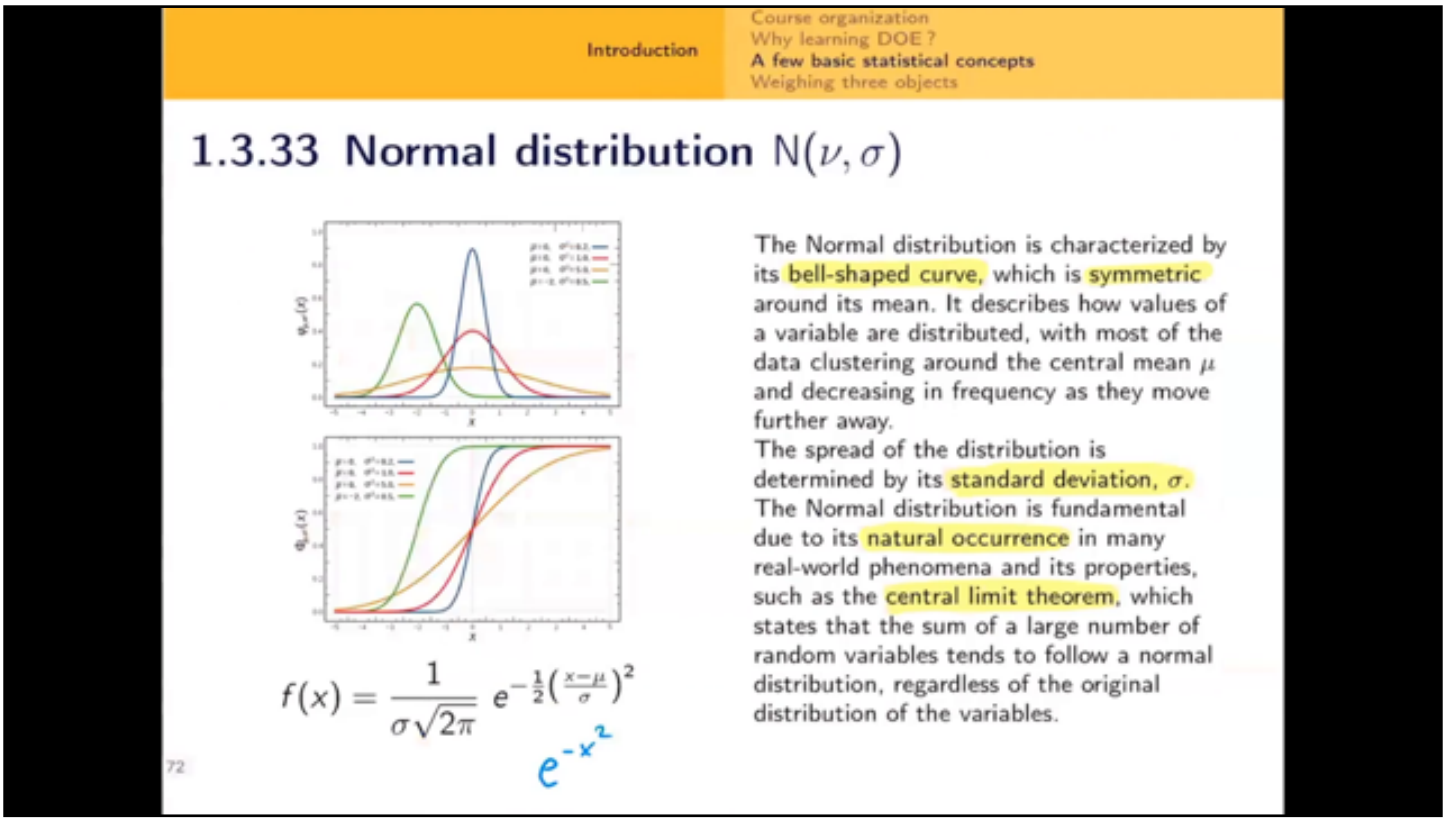


to a standard deviation, the variance usually is compared to sigma squared, and sigma is compared to standard deviation. So, and another function which comes with the normal distribution is the cumulative distribution, which is very interesting because you see why we start having population. So if you read the vertical axis, it's possible to check what is the percentage of your population, so if you want 50%, so that means that imagine that we are on the green one, so we see that for which value it's something as minus 3.5, no, minus 2.5, that we have. So the cumulative is sometimes a little bit more easy to manipulate, you should be able to go from one to see the other. What makes this distribution very important is because we have serums, central limits serums that explain that if you have uniform sampling, and each sampling is independent from the previous one, then you have distributions that can behave, if you make sufficient measurements, you will finish with a normal distribution. Even if individually, it's not, sorry, I mean it's very quick presentation of the central limit serums, so that would be one of the most important serums that is stick with like, don't want to spend too much time, so okay, normal distribution. The next slide is giving you the MATLAB function

notes

summary

10m 13s



that you can add if you have for you, so you can generate random numbers following the normal distribution, run for random and for normal, and it gives you, MATLAB gives you a matrix, so you have to build the normal development that you want vertically and horizontally. You have the probability distribution, the probability distribution function, so it's a bell curve, and then if you give, you have to say which distribution you want, so you say that it's a normal distribution, you have to give the position at which you would like to calculate probability distribution, it would be a value, it would be a matrix, it would be a vector, and after you have to define the parameter of your distribution, where is its place and what is the standard deviation. You can have the CDF cumulative distribution function, is the same type of parameter, you have to say that it's a normal distribution, if you look at the path of those functions, you will see that you can define these two other, to be something as a dozen of other possible distribution, and you have to say where you would like to calculate this function, and you have to give the parameter of the function. Some distribution have one parameter, the standard distribution has one parameter, some have more, usually the function is done that when you have defined the correct distribution that you want, you will, that will look for the right number of parameters, be careful, even if you're going to check, if you put nothing, it will put standard value like zero, what we mean as one for the standard deviation. And okay, you can also have the inverse cumulative distribution function, e, c, d, f, the same way. So what is the e, c, d, f? If we go back to this,

notes

summary

12m 16s

1.3.33 Normal distribution $\mathrm{N}(\nu, \sigma)$

MATLAB

- Random number generation
  X=randn(Ni,Nj)
- Probability density function
  p=pdf('Normal',x,mu,sigma)
- Cumulative density function
  p=cdf('Normal',x,mu,sigma)
- Inverse cumulative distribution function
  x=icdf('Normal',p,mu,sigma)

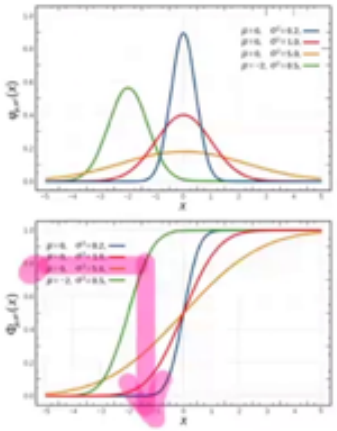
so it's a way of saying, if you see, you have to give a probability,

notes

summary

14m 44s

1.3.33 Normal distribution $\mathrm{N}(\nu, \sigma)$



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

e^{-x^2}

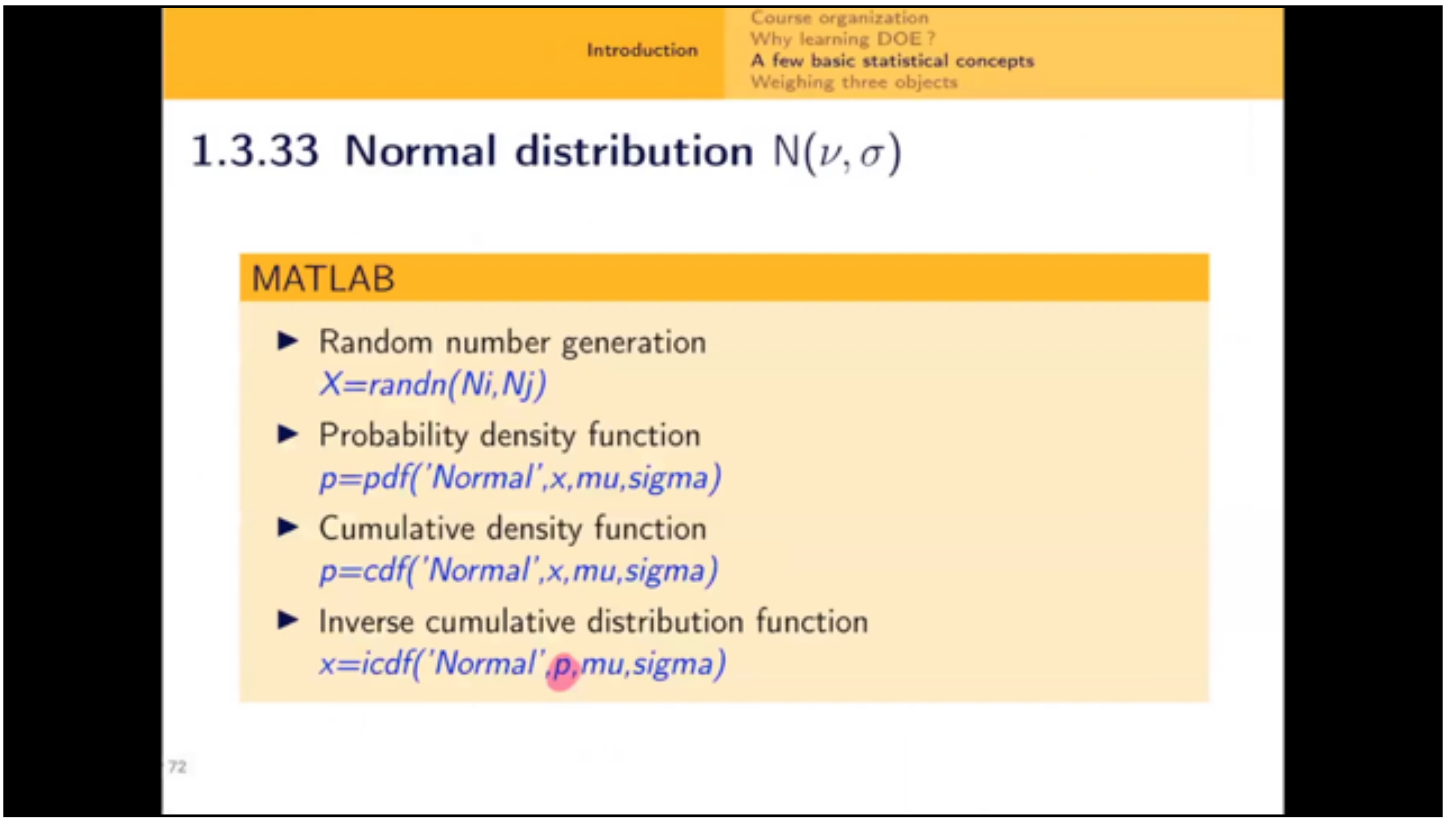
The Normal distribution is characterized by its bell-shaped curve, which is symmetric around its mean. It describes how values of a variable are distributed, with most of the data clustering around the central mean μ and decreasing in frequency as they move further away.
The spread of the distribution is determined by its standard deviation, σ.
The Normal distribution is fundamental due to its natural occurrence in many real-world phenomena and its properties, such as the central limit theorem, which states that the sum of a large number of random variables tends to follow a normal distribution, regardless of the original distribution of the variables.

so it's a way of saying, I'm choosing a value, and I would like to know at what it correspond in my access. It's very important when you calculate confidence interval, because when you calculate confidence interval, usually you are saying, okay, I would like the points that are more than the first 5%,

notes

summary

14m 50s

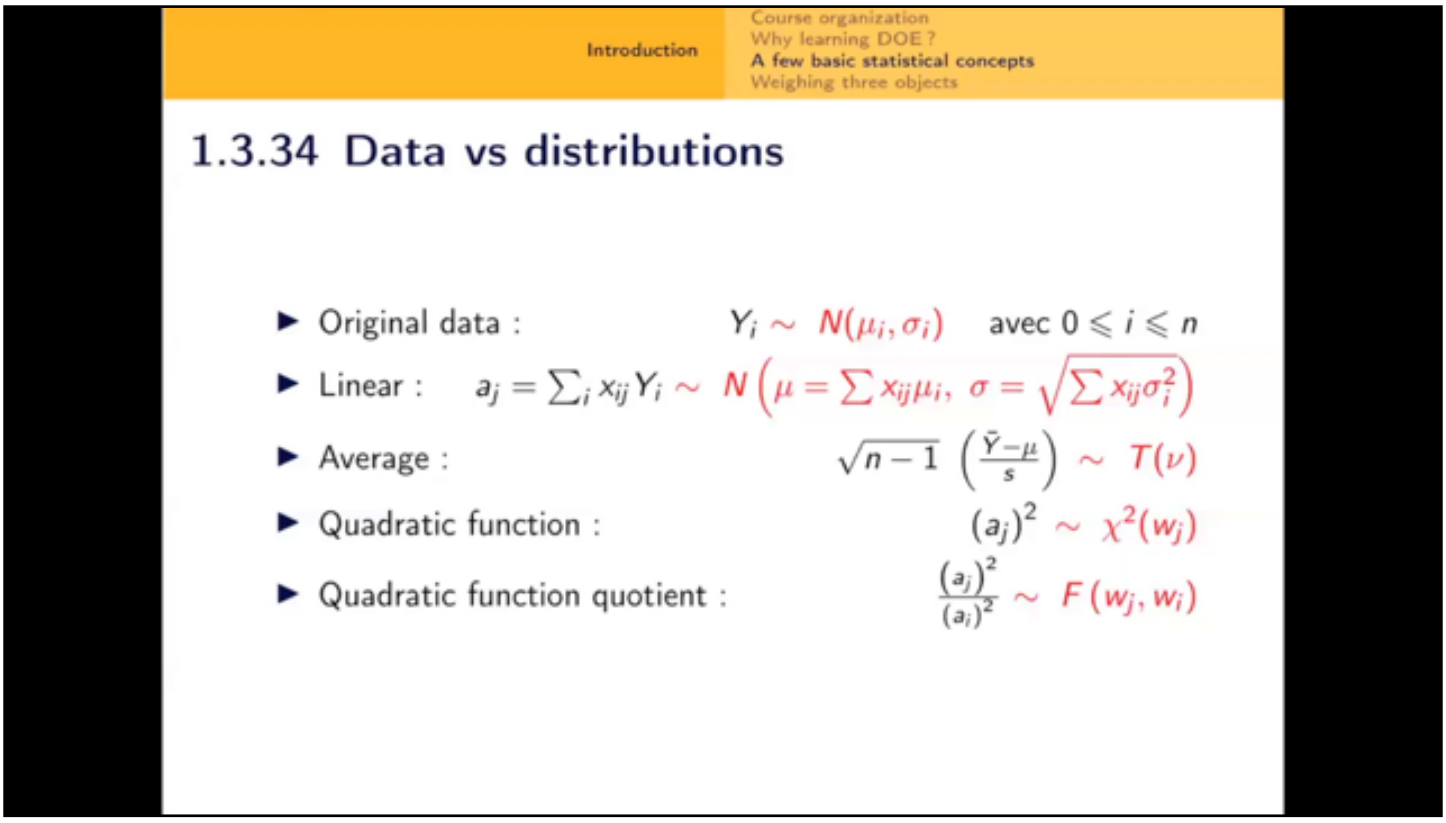


and more than the last 5% for having a 90% confidence interval, so you need to have this function. If you are not very comfortable with this distribution,

notes

summary

15m 17s

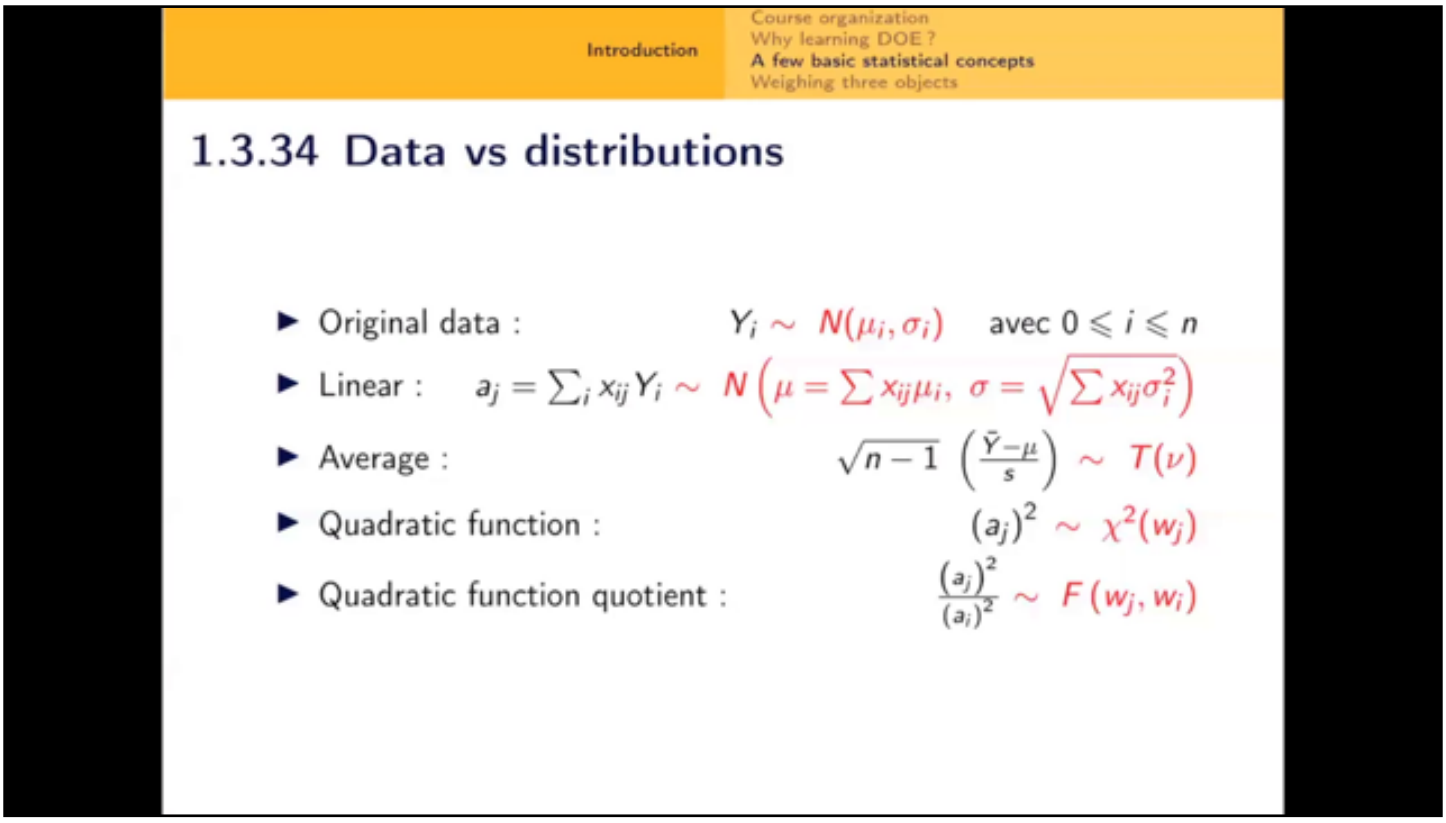


spend some time, play a little bit with this function. I would like to present you the family of the distribution, the normal family. It's a family, there are several distribution, you can actually hear about them, I don't know if you knew, they are very strongly related. We call the normal family distribution, and it's depend what you are doing with your data, so you are calculating some statistics, it's a way of talking with a certain thing that you calculate from your data, or meaning the statistic and variance, it's a statistic, and you can calculate anything with your data, and depending on your data, and depending what you are doing with your data, the statistic will follow a given distribution, so it's very important to understand that. So we start from a situation where my original data, I call it, Y is following a normal distribution, so we write that in the equation, again, I don't know if it's new for you or not, was this field sign, translation following, and after, it's a way of calling the distribution, so there are a few letters that define the distribution, capital N is defining the normal distribution, and so we have two parameters for a given measurement, a given answer, we have an average and we have, sometimes given the variance, sometimes given standard, usually with the same information, but by sigma, sometimes sigma, so you understand that here we have to define what is the statistical behavior of Y? So if we assist Y, you are taking a linear combination, is exactly what we're doing when we calculate the effect of a factor, we will compare, we will add, and subtract some results between them, so it will be a linear combination, so when you do that, that means that your A, J will also follow a normal distribution, so you are making a linear combination of

notes

summary

15m 33s



data, it will follow a normal distribution with the mean, which is following the linear combination of the mean, and with the variance that will follow also the linear combination, so the standard deviation will follow the roots of the linear combination of the variance. Now, and it's also related to the next slide, if you make an average, you are making several measurements, and you are not taking each one, but you are averaging your data, and if you make this statistic, so Y bar is for the average, it's a way of writing the average, you take it out, the real mean of your distribution, in fact, you don't know it, but you don't need to know it for making this, and you divide it by the standard deviation of your measurement, so C minus, you know it, to break letters represents the parameters of your real distribution, and the Latin letters M, and S represents, so what we call something distribution, what you really have measured, so understand it, it's funny, it will appear another time it occurs, you know Y bar, because you make your measurement, and you are able to calculate your average, you don't know MUDE, because you don't know the real average of the distribution, and you know S, so in fact, this, you cannot really calculate it, but what we say here, that the difference of your average was the real mean of your distribution, but after you have a question of the factor, divided by the variance, your experimental variance, and multiplied by the root of the degree of freedom, and the numbers of that are taken, so the M minus one is the degree of freedom of your population,

notes

summary

1.3.35 Student's T_ν distribution and CI

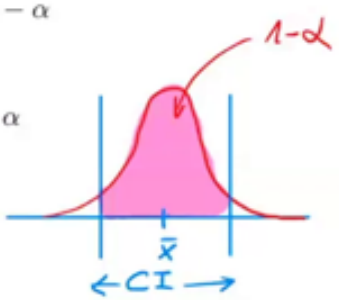
- Published by William Gosset, 1908, (Guinness)

  If the observations X_i are IID[1] then $\left(\frac{\overline{X}-\mu}{s/\sqrt{n}}\right) \sim T_{n-1}$ ($n-1 = \nu$ df)

- Confidence interval theorem :

$$P\left(x \in \left[\bar{x} - t^{n-1}_{\alpha/2}\sqrt{\frac{s^2}{n}}\ ,\ \bar{x} + t^{n-1}_{\alpha/2}\sqrt{\frac{s^2}{n}}\right]\right) = 1-\alpha$$

with t^{ν}_{α} the value of t at which $\int_0^t T_\nu(t')\,dt' = \alpha$



1. independent and identically distributed

and root of is the root of, so this is following a student's distribution, T is a student distribution, and this distribution I just presented on the next slide, as one parameter is the degree of freedom, so nu is the same thing as S and minus one. Okay, so the average is not following the normal distribution, it's following another, and probably it's not because of your mind, because more you are making measurement, more the average is precise, so if it would be the normal distribution, it would be impossible, because in the normal distribution, you do not have a, a parameter for expressing how many measurements you have done, so we need a distribution with this as, this as a factor, and the fact that we have T minus mu divided by S, free us of taking into account which normal distribution we are taking, so it's a way of simplifying the process. After if you make a quadratic function with your, your data, so it could be with your Y or with your A, if you take the square of it, they will follow another distribution that equals a T square distribution, and so it's usually represented by the Greek letter T, and we call it square because it's related to the quadratic function, and this T square is also dependent on the number of the degree of freedom, sorry we call it degree of freedom nu, in the first one is the second one, we call it w, it's the way we write it, the same thing, it's the degree of freedom, and if you make a ratio between two quadratic functions of your data, it will follow another distribution, and it's a distribution of Fisher, it was an open problem and Fisher solved it, and so the distribution has taken its name, and in this case, we need to have the degree of freedom

notes

summary

20m 25s

1.3.35 Student's T_ν distribution and CI

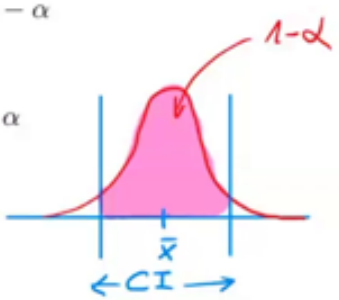
- Published by William Gosset, 1908, (Guinness)

  If the observations X_i are IID[1] then $\left(\frac{\overline{X}-\mu}{s/\sqrt{n}}\right) \sim T_{n-1}$ (ν df)

- Confidence interval theorem :

$$P\left(x \in \left[\bar{x} - t^{n-1}_{\alpha/2}\sqrt{\frac{s^2}{n}}\ ,\ \bar{x} + t^{n-1}_{\alpha/2}\sqrt{\frac{s^2}{n}}\right]\right) = 1-\alpha$$

with t^{ν}_{α} the value of t at which $\int_0^t T_\nu(t')\,dt' = \alpha$



1. independent and identically distributed

of the numerator, and we need to have the degree of freedom of the denominator where the two parameters are. So this, what is important, you can come back to this slide, so this is the logic, the rationale we'll use when we have made a statistic, and it's okay, what is the distribution that my statistic is following, so it's why it is very important, and this depends on the fact that the data was considered as a normal distribution, and so it's very practical because you have the central limit theorem, and if you do sufficient experiments, you can apply that. Let's talk a little bit about the T distribution of Fisher, the different way of writing it, so T, new T parenthesis, new, and so on, so first of thing, very funny, why it's called student distribution, and not the Gosset distribution, the historical, I don't know, it's because Mr. Gosset was a statistician that had been hired by Guinness, beginning of the 30th century, and he was not allowed to publish since, so he was not allowed to publish his work, so he has worked on the test of the beer Guinness because it was the time when they started to industrialize the production of beer, so he was working on the fact that we do for having the same, quite the same taste with all the bottles, very important for the question,

notes

summary

1.3.35 Student's T_ν distribution and CI

- Published by William Gosset, 1908, (Guinness)

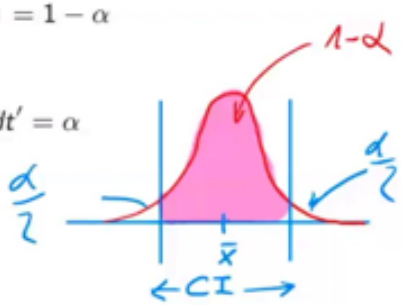
  If the observations X_i are IID [1] then $\left(\frac{\overline{X}-\mu}{s/\sqrt{n}}\right) \sim T_{n-1}$ (ν df)

- Confidence interval theorem :

$$P\left(x \in \left[\bar{x} - t^{n-1}_{\alpha/2}\sqrt{\frac{s^2}{n}},\ \bar{x} + t^{n-1}_{\alpha/2}\sqrt{\frac{s^2}{n}}\right]\right) = 1-\alpha$$

(t table)

with t^{ν}_{α} the value of t at which $\int_0^t T_\nu(t')\,dt' = \alpha$



1. independent and identically distributed

So they don't need to be normal for following the student distribution. They need to be independent and identically distributed. This is the hypothesis of the central initial. I ID. And this distribution, as I mentioned, is very important for calculating the confidence interval. There are seren or statistics seren confidence interval. When you are calculating an average, you can threaten that the for a given probability. So if you have the probability that you want in the red surface, which is one minus alpha, 95%, 90%, 99%. That means that you would like to calculate the position of the limits of the range that you are considering for your confidence interval. And so you can calculate this confidence interval having the value of your average. So the value x square and minus and plus in the two position. So here you have a minus here you have a plus and the value of the T distribution. And you see that in that sense it's a small team. When we talk about the distribution, I have no idea the solution is a big team. But when we talk about the inverse, or the cumulative distribution, it's a small team. So this student inverse cumulative distribution for and minus one degree of freedom is one of the parameter and the probability is alpha divided by two. So this value here, because we have in the in the bit which is at left we have alpha divided by two. And here we also have alpha divided by two. We calculate only one time because it's a symmetric distribution. Good idea of the distribution which is symmetry and do not depend on the mean because we have adapted the mean. So you need to check that to calculate it. You can calculate this line and calculate it is any library of statistics or usually you find these values at the end of the

notes

summary

25m 25s

1.3.35 Student's T_ν distribution and CI

- Published by William Gosset, 1908, (Guinness)

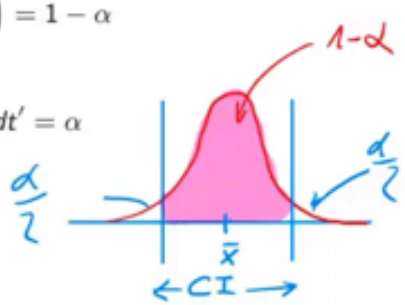
  If the observations X_i are IID[1] then $\left(\frac{\overline{X}-\mu}{s/\sqrt{n}}\right) \sim T_{n-1}$ (ν df)

- Confidence interval theorem :

$$P\left(x \in \left[\bar{x} - t_{\alpha/2}^{n-1}\sqrt{\frac{s^2}{n}},\ \bar{x} + t_{\alpha/2}^{n-1}\sqrt{\frac{s^2}{n}}\right]\right) = 1-\alpha$$

(t table)

with t_α^ν the value of t at which $\int_0^t T_\nu(t')\,dt' = \alpha$



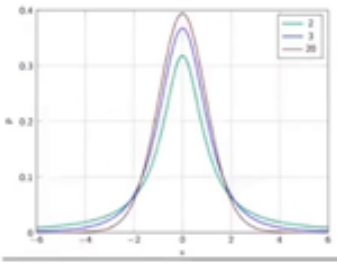
1. independent and identically distributed

statistical book. You can check already quite all the book of statistics at the end you have some tables and you find you say okay, I have 10 degree of freedom and I would like to have a 95% confidence of one minus 0.95 met 5% and the half of 5% is 2.5%. So you get this value from a table from a function. And you multiply this by the root of the sample variance because it's not sigma you don't know when you find the real variance of the distribution for the sample variance. So this is why I use the letter the Latin letter s and not sigma divided by the numbers of measurement and you take the room and this will give you the left limit of your confidence interval and if you make the same calculation. Plus, you have the right limit of your confidence interval. So it's why the student distribution is very important because a good engineer. the value so is the integral from zero to the value of t. So this is the axis of probability of the student distributions and t student distribution. So be careful when you use it in the algorithm. Sometimes it could say if you are calculating the left value or the right value

notes

summary

1.36 The Student's T_ν distribution



	Parent dist.	sampling dist. for $\bar{y}$
Mean	η	η
Variance	σ^2	$\frac{\sigma^2}{n}$
Std dev.	σ	$\frac{\sigma}{\sqrt{n}}$
Form	$\sim$ any	more nearly Normal

The Student's t-distribution is used to estimate population parameters when the sample size is small and/or when the population variance is unknown.

The t-distribution is parameterized by degrees of freedom $\nu = n - p$ (n being the sample size and p, the number of parameters) ; as the degrees of freedom increase, the t-distribution approaches the normal distribution.

It is commonly used in hypothesis testing to determine if there is a significant difference between sample means.

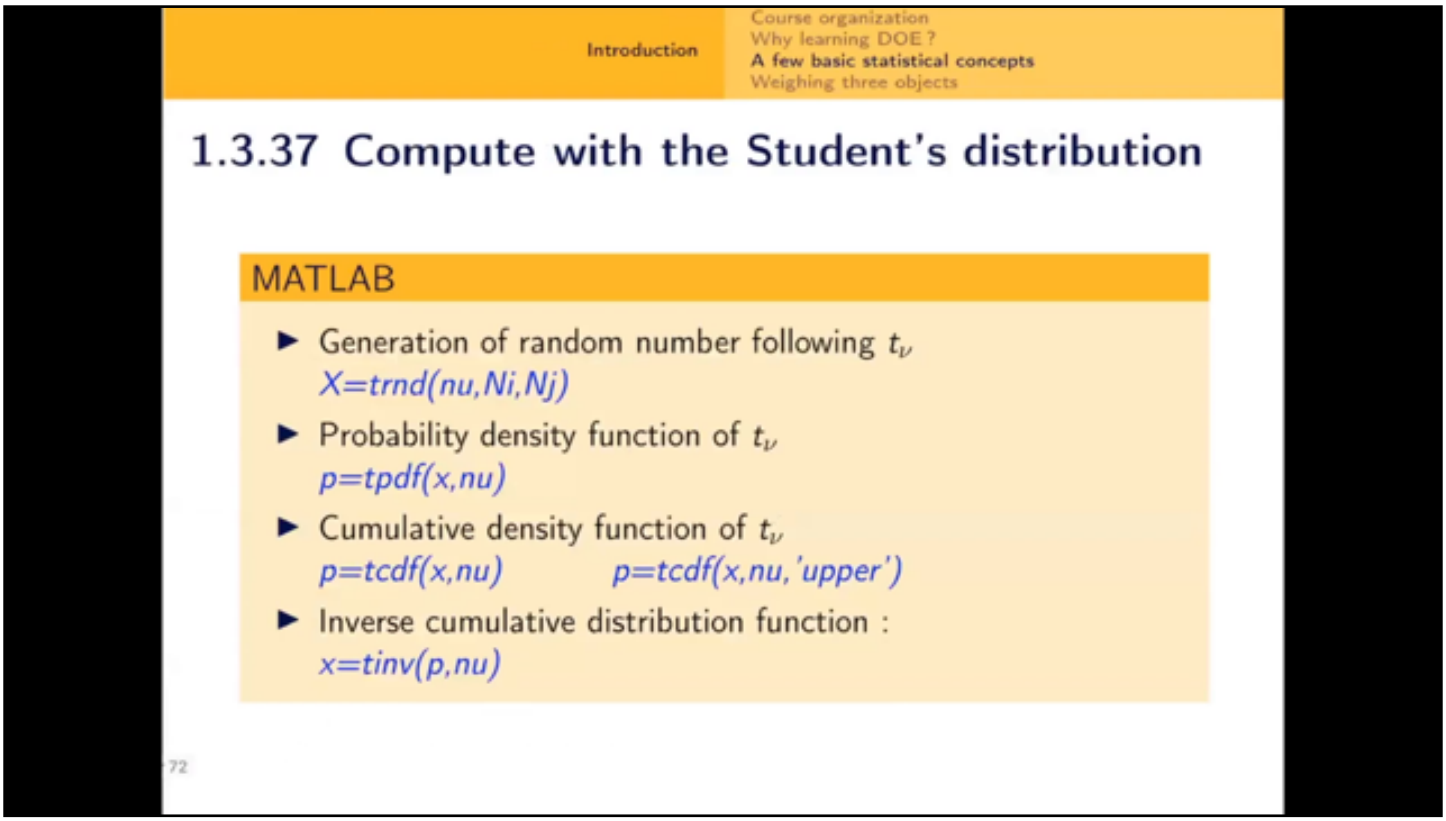
72

one just the opposite of the other. So usually we consider as the value as positive it must be. If it's negative it's because you have calculated it for a small number it just takes the absolute sometimes you have to say the lower or the higher value but so it must be a positive value in any

notes

summary

30m 37s



case. So here it's all this distribution looks so it's looked like a bell curve, a normal distribution but understand that the parameters are different. So the student distribution is all the time centered on zero and what makes its width is the number of degrees of freedom and more you make measurement smaller will be the width. It will be more and more limited. Higher peak all the time. Something you can check that if you are if you have some values of what we call the parent distribution that means the distribution of your data and you are calculating the mean of this data. So if you have the mean of your parent distribution is sigma the mean of the of the mean so that means you make several times 100 measurements okay and you are comparing the different times that you have made 100 measurements. So the other range of those different means will be the same as the mean of your distribution and that is Greek letters we are talking about the theoretical where you have something interesting to remember that for the variance if you have the variance of your parent distribution which is sigma square then the variance because you can calculate the variance is the the integral of the square of the distribution will be the original variance divided by the number of measurements. So the standard deviation is the same information but you have to take the roots of the root of n and the sigma and when we talk about the form it's what the interpretation of this iid you don't need that the parent is a normal distribution sometimes people think that it's not true it's only necessary that you are fair when you are making your measurements that it's independent that it doesn't depend on you but that your system you must have memory that really give you the

notes

summary

31m 0s

1.3.37 Compute with the Student's distribution

MATLAB

- Generation of random number following t_ν
  X=trnd(nu,Ni,Nj)
- Probability density function of t_ν
  p=tpdf(x,nu)
- Cumulative density function of t_ν
  p=tcdf(x,nu) *p=tcdf(x,nu,'upper')*
- Inverse cumulative distribution function :
  x=tinv(p,nu)

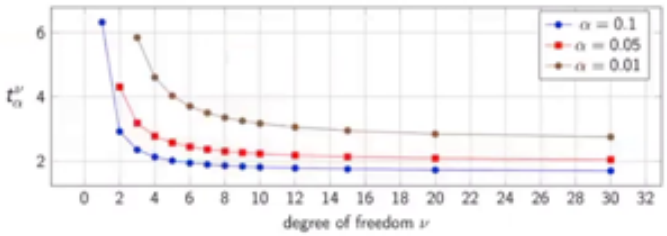
72

new value independent from the values that you have and i don't i don't think i don't change the distribution from one to don't change the system you can say system i'm statically speaking and it's what you need for having finally that's your sampling distribution for your average is nearly normal if you have a lot of the new freedom very rapidly the shape will be really a shape of a normal distribution.

notes

summary

1.3.38 Confidence and degrees of freedom



After about 10 degrees of freedom, the Student's t-distribution begins to closely resemble the normal distribution, and further increases in degrees of freedom result in only minor changes to the shape of the distribution. This means that, for practical purposes, you would need a much larger sample size (a change in the order of magnitude) to observe a significant reduction in the width of the confidence interval. Similarly, for confidence levels below 95%, the size of the confidence interval does not vary dramatically.

Therefore, using **10 degrees of freedom** and a **95% confidence** level is a reasonable standard for many statistical analyses, as it balances precision with practicality.

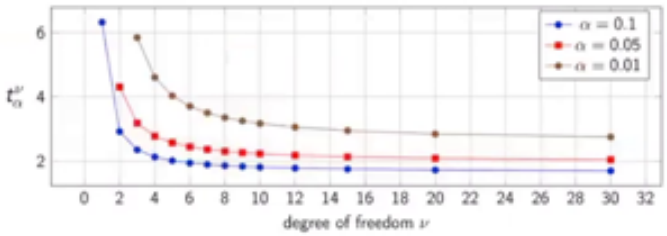
Okay so here you have the equivalent routines for MATC lamp for calculating so if you want to generate you have random but they put the t before and it was probably another program so it's t r and d i forget it all the time and you can generate a random number following the student distribution you have the probability density function again with the t before and the cumulative is also the same thing as previously quicker and this is very important function that you will need to calculate confidence interval because if you see it's exactly the t that we have in the function before we call t inverse inverse of the student distribution and you give the probability i don't know five percent for 95 percent confidence i mean it could be 97.5 percent because you have to take one minus alpha plus alpha divided by two and the degree of freedom it will give you the value that you need for calculating your confidence interval in this idea also to understand what is the influence when you are making measurement of the degree of freedom so usually it's true you said more data better as much if you have time and money please do as many experiments that you you can do because more experiment will give you a more precise result this is true but it's not uniform so you see here so first consider only one of those of those lines so consider the first one the blue one which is for 10 percent so 90 it could be for 80 percent of confidence interval of 80 percent for 10 percent on one side 10 percent on the other side so if you have only two degrees of freedom you see that this t value which is important which is multiplying the variance of your of your experiment will be around six and

notes

summary

34m 10s

1.3.38 Confidence and degrees of freedom



After about 10 degrees of freedom, the Student's t-distribution begins to closely resemble the normal distribution, and further increases in degrees of freedom result in only minor changes to the shape of the distribution. This means that, for practical purposes, you would need a much larger sample size (a change in the order of magnitude) to observe a significant reduction in the width of the confidence interval. Similarly, for confidence levels below 95%, the size of the confidence interval does not vary dramatically.

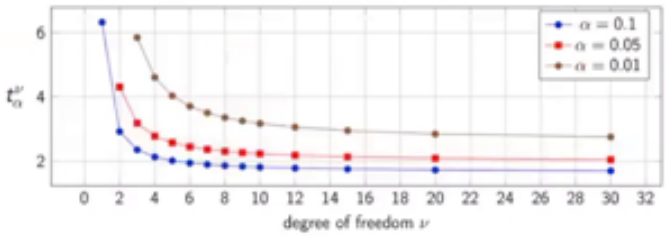
Therefore, using **10 degrees of freedom** and a **95% confidence** level is a reasonable standard for many statistical analyses, as it balances precision with practicality.

it's a lot you have to multiply six times your sigma for having confidence interval of 80 percent that after if you add some measurement very rapidly you arrive very close of two which is already six degree of freedom you are about two so for having an 80 percent confidence interval just multiply your sigma by two and it's the village but very very slowly so what you see you have something we can call a hill and so it's nice to be just after the hill this is the good balance between the efforts and the quality because if you really want to improve it's not if you have 10 it's not 12 that you have done it's 100 100 measurements that you have then you have to change the order of magnitude so having something as I could be 10 around 10 just before 10 degrees of freedom is primarily quite an optimal value of degree of freedom in a standard situation if if you are paid for making making a payment do as much experiment as you can but if you are really at a short time and you want to have a good quality without extending too much your current plan yeah eight standard real freedom would be quite a good reference and also for the accuracy of your of your confidence interval you see that you don't see it so so so much but you see that those those curves are becoming quite the same also so they are also for the primary of your confidence interval 80 perhaps is not so good so so good value but 80 the difference between 95 or 99 it's not it's not it's not so big so it's not it should be good to have something as 95 it's a good idea very good idea so it's why I I have asked strategy PT to reform related

notes

summary

1.3.38 Confidence and degrees of freedom



After about 10 degrees of freedom, the Student's t-distribution begins to closely resemble the normal distribution, and further increases in degrees of freedom result in only minor changes to the shape of the distribution. This means that, for practical purposes, you would need a much larger sample size (a change in the order of magnitude) to observe a significant reduction in the width of the confidence interval. Similarly, for confidence levels below 95%, the size of the confidence interval does not vary dramatically.

Therefore, using **10 degrees of freedom** and a **95% confidence** level is a reasonable standard for many statistical analyses, as it balances precision with practicality.

and strategy PT refinerating like that uh after about 10 degrees of freedom the student distribution begins to closely resemble to normal distribution and further increase in the degrees of freedom result in only minor change to the shape of the distribution this means that for practical purpose you would need a much larger sample size a change in the order of magnitude to observe a significant reduction in the widths of the confidence interval similarly the confidence level below 95 percent the size of the confidence interval does not dramatically change then using 10 degrees of freedom and 95 confidence level is a reasonable standard for many statistical analysis as in balanced precision with practical use well I asked strategy PT I I write something quite that was bad English strategy PT reform related and I asked him if it was agree with me he said he was agree with me I didn't check in reverse I would be very I should try to sometimes I'm suspicious that it could be every time I'm really zero and that nevertheless I appreciate because uh okay he write a nice sentence you understand well what I wanted to say and we can go one step further

notes

summary