Performance evaluation of computer systems

Prof. Krishna Moorthy Sivalingam

Department of Computer Science and Engineering

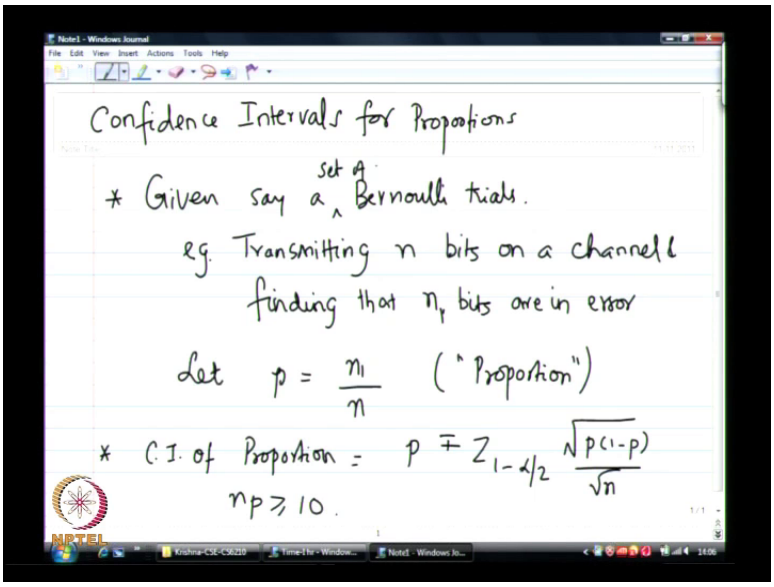
Indian Institute of Technology, Madras

Lecture No. # 33

Confidence interval for propotions and introduction to experimental design

So, last day we were looking at the comparing two alternatives, and then we saw we can find out the confidence intervals for paired observation and unpaired observation. So, sometimes we also have done experiments, where we measure whether we are successful or not like a set of only Bernoulli trials. So, we are measuring for a example bit error rates on a channel right. So, you measure a sample of bits, 10000 bits and report that have a 10000 bits 100 bits were in error, and therefore the bit error rate equals 100 byte or 1000 bytes right.

(Refer Slide Time: 00:43)

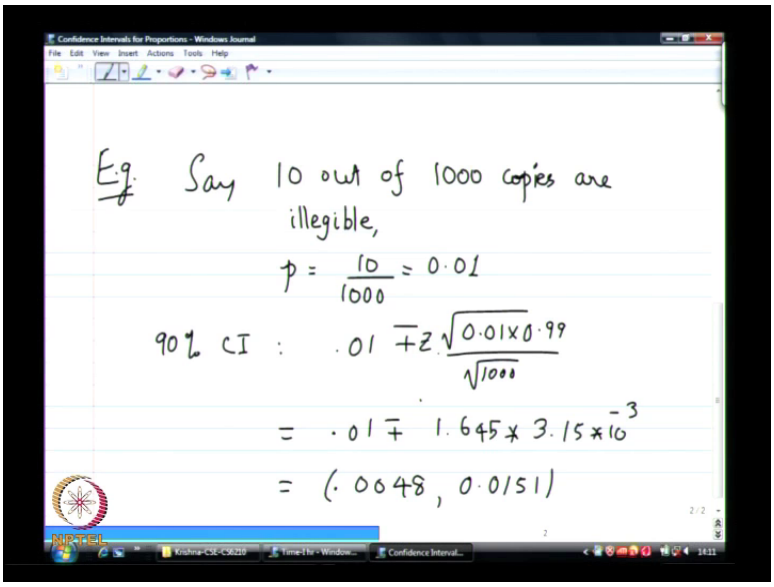


So, let us define this form. So, given say example transmitting n bits on a channel and finding that n 1 bits are in error. So, normally say mean bit error rate is simply n 1 by n that is the mean; so that definition for it is good, but the sometimes want to know the probability of what the confidence intervals. So, let us p equals n1 by n, so this is the so

called proportion. So, if you want to have something more statically serious to say about the bit error rate of the channel or something else, then we talk about the confidence interval of this proportion. So, that is P is is plus or minus plus or minus alpha by 2.

They are standard one, this standard deviation is Bernoulli trials P into 1 minus P at the variance is right the standard deviation is square root of here, and since we get n experiments right. The book says at use the dynamical derives this that I find the use of thing is set of Bernoulli trials choice right ultimately and set to that, but anyway so this is the confidence intervals for a proportion, but we will requires is that n P is greater than or equal to 10. And the book does not go and explain why and I have also not bothered looking in to it, but the book has example which has n P less than 10, one of the example is shown here does not repeat this this condition, so somewhere they discrepancy.

(Refer Slide Time: 03:57)



So far this is again very simple steps, these are the two mark question for final exam right. If you are you know going to Gurunath to making copies to find out 10 out of every 1000 paper right gets messed up and get stuck in the printer right. So, say write 10 out of 1000 right. So, the mean error is 10 out 1000 right, so that is 0.01. So, your P is 0.01 but then 90 percent confidence interval. What was the 90 percent confident interval? 2.3, all other do some other things this is the 0.01 plus or minus 0.01 into 0.99 divided by square root of 1000, the corresponding the z values.

Let the t distribution is half.

Only Z, t is not allowed again books simply say that only Z.

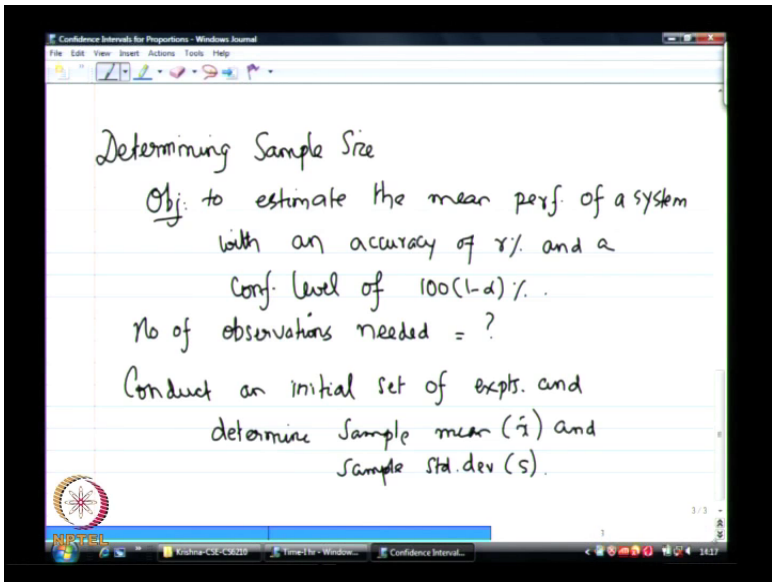
More than 30 we take.

I know, but only we always take, but any way we take the way you take the Z distribution at the unit the normal, but not t. So, here this is the t is not used here. 1.645 is 0.10 plus or minus 1.645 into. So, what is that? Into 3.15 into 10 to the power minus 3. It is enough, this defines that is the full. So, that I put it in plane 0.0048 and then.

So, that is the 90 percent confidence. So, whenever we see the trials on CNN dot com or IBN live dot com, and trials assume suppose the poles the online poles, now this is 1000 people interviewed in say that in the people approve of Bush, and do not approve of Bush. And then just you will ask what is the confidence interval initially say plus or minus 5 percent along that 1 bottom that say plus or minus 5 percent from that mean is observed. So, that was what this would be. Take the ratio Z into the divided mean is given the deviation with respect to mean, you never can trust those numbers. I am just asking one question and people, say like Bush or do not like Bush. Approval ratings, now its Obama. So, large the samples size then closer the smaller user.

So, now on somebody says the question is how often do or what is the fractions of Nano cars that burn up? That is what people when you sell Nano cars call you the question I have heard several Nano cars that burnt up spontaneously, in the question how many actually? So, 3 out of some 100 1000 cars that were sold, but only here about the negative points; so anyway that is how we look at the proposal that we work on our theses. So, that is conference in tern for proportion. The next the relevant topic is thus say we want we have choice of the conducting as experiment we want for example, application that we saw in stimulation runs and so on we control over m, there we saw what should be the number of number of replication that necessary to meet certain width and confidence interval. So, that is what we will try to find out.

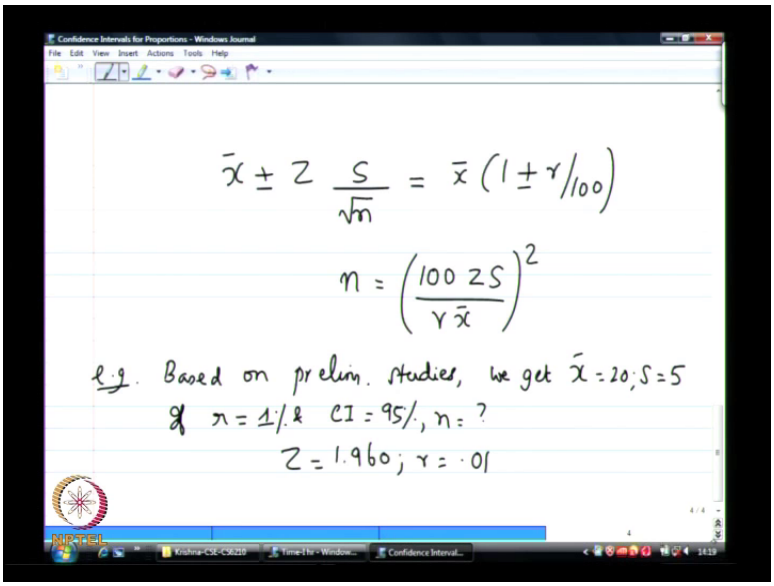
(Refer Slide Time: 09:17)



So, the objective is to estimate the mean performance of the system with an accuracy of r percent and the confidence level of… This is what we did (( )) the answer that suppose to measure the 0.05; for example, it is an accuracy 5 percent that is not something that you will state, so the number of observations needed is what? So, I want to conduct experiment several times in number of times that the only condition said that the ratio of mean, and in the bit of interval plus or minus that should be bounded by the observation.

It is not entire width, half the width. We can say the width if you want, but it is half the width. How the conference interval should be bounded by r. So, how do we do that? So, starting off with nothing we cannot say anything. So, we have to conduct some basic experiments say 1000 trials and then compute the x bar the mean and the samples standard deviation and based on that we can try to do. We need conduct an initial set of experiments. See your choice of (( )) depend up on the initial set of experimental values you get.

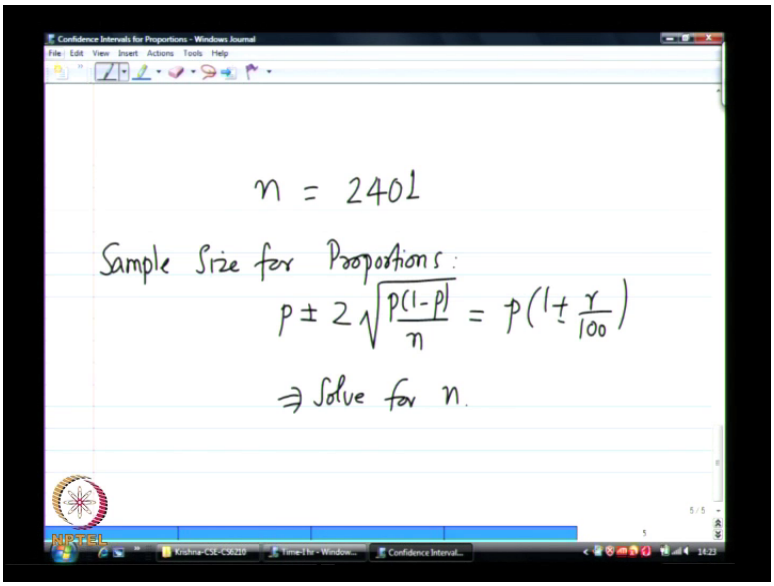
(Refer Slide Time: 12:37)



So, basically it is x bar plus Z whatever you are looking at in to S divided by root n. So, this is the actual the interval plus or minus and this we want this to the equal to. So, anyway this is no (( )) you can simply solve for both sides in end of n sequence. So far example based on some prelim, we get x bar is 20 and S equals 5. x bar is the response time. So, if r equal to 1 and the confidence interval is 95 percent, what is n? The book uses a smaller value this is r is the different person right. So, r is very small 0.1.

So, for 95 percent your Z is 1.960. So, Z is 1 .960, r is 0.01 and the rest is already given. We can simply plug this in.

(Refer Slide Time: 15:18)



So, what is n? We need 2401 and 1 observation. So, let us see initially. Our final mean will not be this.

r is 1 you know.

r is 1 percent.

Should count 0.01

Yes r is 0.01-

But in the formula already accounted for the percentage, yeah sorry yeah

I jumped there that we have. So, it will be simply 1.

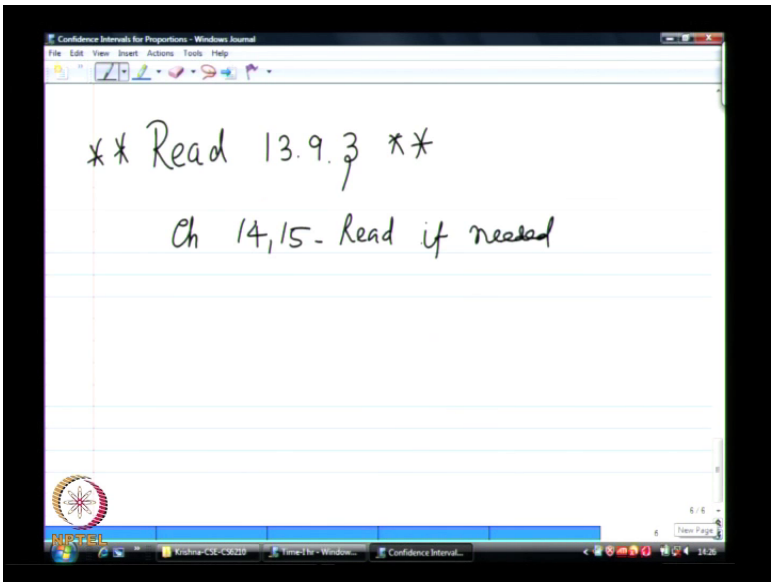
The question is also r is 1.

It is basically 100 into 2 into 5. So, that is 1000 divided by 20, that is 50 squared, that is 2401 all. So, 2.01, so we can do the same thing for proportion also which will not elaborate much of a samples size.

So, again the same terms as before r is here half with accuracy. So, now, it is P plus or minus Z. (No audio 17:24 to 17:47) The book has an annoying habit of switching from the percentage in the previous word no percentage is next one. So, the book will not have

r by 100 but you know what to do that is the simply type r is then take with what you want not a say. Let us see here we have actually ratio, if it is P, then is little bit if it is respect the proportional to the little bit, if it is simply a raw value that is given then you simply take P is the plus or minus 1 right. So, we can again solve for which we skip for now. That is end of the chapter 13 and there is one more little section of comparing to alternatives and give that us 3D materials for a exam.

(Refer Slide Time: 18:57)



So, that brings us, we sort of went back to the book right well begin front of the book. The other chapter is the earlier section which is skipped over work load characterization, which will never get attain to this semester. What is coming up next is, this is just window in you want compare to different is the terms two or more system probability compare that and what are the confidence interval and so on. Next we will look at actually designing experiments, and then try to find out that when I have a set of factors, multiple factors that impact the performance of the system. For examples, CPU speed, memory size, cache size and so on.

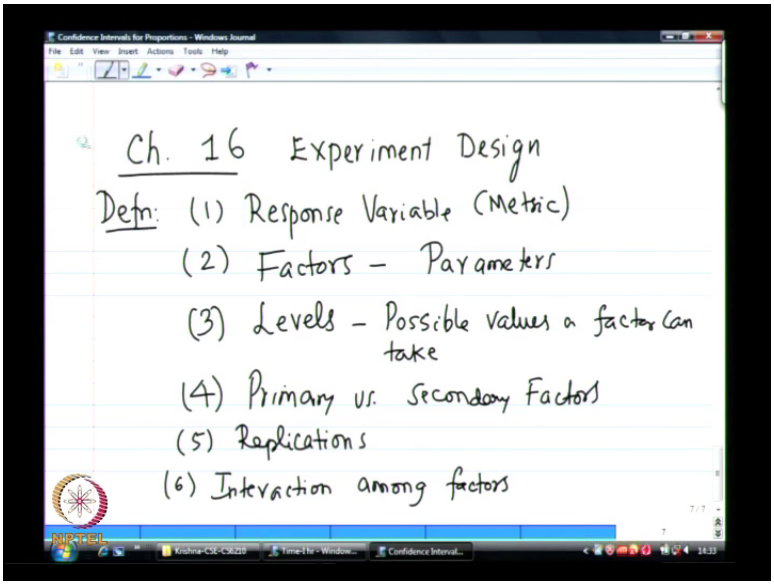
So, how do these different factors affect the performance which is more significant than the other right. That is what we will step in to next, but before that there are two chapters on simple linear regression model which most of you would have seen right linear regression models fitting something to the straight line y equal to m s plus b all the stuff. right if not this if you are interested you can look at that. Many times we will only have

the plots general fear, the results, analysis and so on. So, a right you want to see if it fits a linear models and that is usual of predication right you can predict for some unknown values are explored are attested with some different parameter combinations, so that I will leave for self enthusiastic reading.

Formula that is you use as x a to determine n. So, I did not understand idea behind that. So, you are assuming that explore with estimate through this actual S.

That is not correct, yeah. This is the approximate here. So, if we actually run the experiment that many times will you really get that you are looking for I am not sure, but then give the idea formally you should try to, and that time you still have to compute this r right. And but if all the samples are if the represents say from the same distribution then assumption is the x bar represents the samples, the population makes right that the where was defined different x bar different s to start with then the values bit right. So, next is looking at experiment design and analysis of factors.

(Refer Slide Time: 21:45)



So, that is the next part of these entire books through. So, often then are not ones you finish you are program or whatever system you are building software then we have to actually produce performance results, which is always the hardest thing to do right. We did not really like that in the lab program assignment too that that finish the codes, and is done for, so many combination. And what is the point of all the program was going to all the other things, but that is the whole point of performance analysis right is to look at all

the parameter combinations and so on. So, that is what this part is all about. So, some of the definition that you will need, let us just that here right, this are things all ready we have seen, but just as a reminder.

So, for any system use the term response variable to measure some part of system parts it is also called metric like the delay through put queue length all those response variables right the depend up on. We talk about right the earlier part of the said you specify the system parameters, configuration, right the number of servers, the queue length of the number of buffers, the CPU processor in capability all those of you are factors.

So, the factors own that you control that change the performance of the system right depending on which and studying factors is important, because you want to come up with the set of operational parameter that will you give the best performance right. But if you are seeking some performance goal in mind like want to service 10 million transactions per second you should have an idea of how many system you need or how many CPU's you need, virtual boxes you need. So, the factors on that you play with and the response time is what you measure and again these are all fairly rudimentary stuff and different values the each factor can take this the level. So, you are factor can be qualitative or quantitative right if you simply (( )) or Motorola 68 k, then there is nothing connecting there.

It is simply qualitative to the two level sets or something from the factor can take CPU right it could simply the number of CPU of course which is 1 2, what about 8 or 16 and so on; so a case is the numerical value or discrete value right and if it continues and so on. So, and is their level definition other things. Then we have so called primary factors and secondary factors.

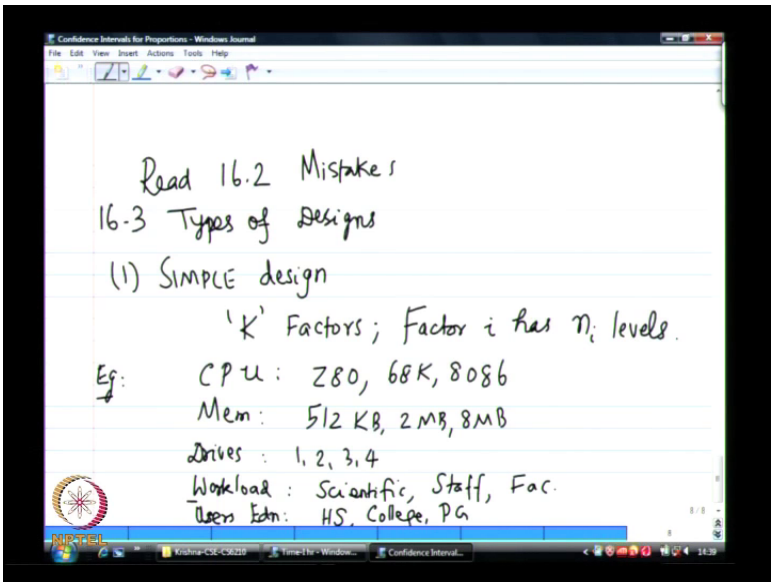
So, the primary factor what you want identify right you want find which have this make that have the most important of the system, if I want change something what should be. That usually comes when you try to redefine in a system right if you have deployed your wireless cellular network on the field as (( )) works done in the olden days and we are still doing right. If you have to decide how many server should put in the back bone to get all this all the traffic and process it. So, you need to find out what is the most important right should you have more server simply add more memory to the existing

system, right same thing we are looking at lab too. We have 13, 15 machines and then students are cribbing that not enough the machines, because all want to use it at the same time they are mostly asleep. When this April May time frame all want to use it at the same time. So, should I simply bump the memory on them, add more CPU, cards so on. So, that is those primary verses secondary factors.

So, that is what is the purpose of this experiment, it is to find out. Then of course you know about the replication right, so we look at experimental design. So, we look at experimental design and analyzing the important of the different factors with and without replication depending on time. Chapter 17 is without replication and 18 is with replication and they do both running running till next Wednesday and is like chapter 20 to cover and there are other design and other items which need to design experiment design and I will skip all that and more important things interaction among factors. So, two factors, one is memory size, by memory size and other is cache memory size right and these are two independent factors, I can vary them independently. When you find that increase memory size right yes proportionally for the given cache size the performance of the system increases.

But when you find that some increase both cache as well as memory right that the increase is not simply linear, but it is something (( )) linear where we get much more performance in if you both say cache size and as well as your main memory size. That is what is called interaction right between these two factors. So, independently they would give you better performance going from one level to next level, but there are cases where they combination of these two factors would simply do very well. It is like Sachin and Ganguly right by themselves they play well, but when actually the team up and partnership for the same, they end up doing betters, that is they help each other. That is what these two factors also try to do when there is a big list of common mistakes made in experimental design, which is all of you should read.

(Refer Slide Time: 28:12)



This is where slides are useful. As in as a slide and unfortunately, I have to get his permission before using the slides otherwise simply you can flip through the slides and find out, by the way all of these chapters are Raj Jain webpage in case you are not. So, end mistakes mistakes in the books are also in the slides as if not is not fixed those right. So all right now, what are different types of the experimental deign right. So, one is called the simple design.

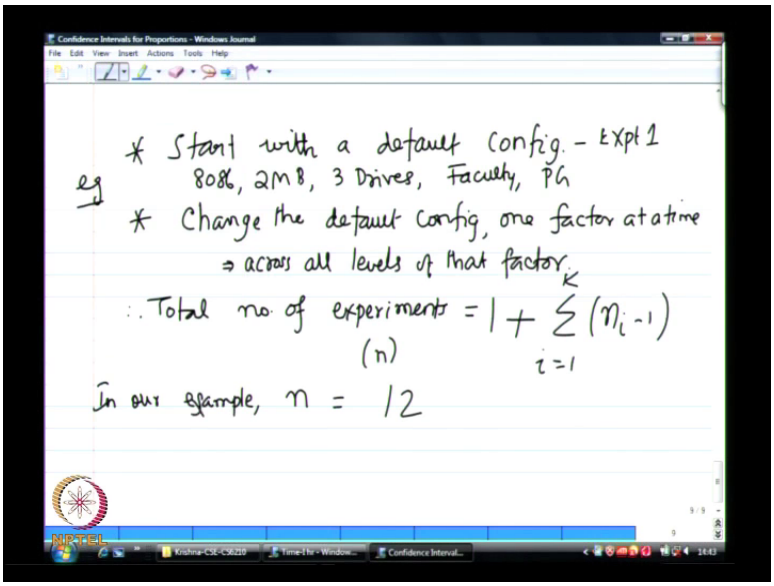
So, we will always talk about k factors. There are the k factors that affect the system performance which want to design experiments for in each factors i has and n i levels. So, for the book example is just use as it is. So, I am trying to design the computer system right and I just what to find out performance of the system for various word flow types the different types of users something like that and I know why the user type matters for high school education matter, but any way that would this example right. So, let us say that you know I have CPU is the factor, I have three factors and again I sent a mail to Raj Jain two weeks ago that he should update this book, but he did not reply to that mail. And when I sent a list of errors, he replied immediately, said book needs update, because some of you might not have heard of the date is still around I think. So, you give the date you have the date is something is not heard of within occurred of without 68 k the 68 k unknown. So, I did assemble the program it was (( )) and all it was memorize was half codes for no reason just figure.

You like we get time memorizing it. We really did not have much time to actually player on within, but we architecture is very simple you got it. It is a nice architecture. In fact what you did in 7 semester computer system design let us kind of what we did. I have this huge board where from scratch all the components will be painfully run from graph for the entire hour in simply copy that go that is was only check all your components of the system right that was actually see the entire CPU it will all its components in action. So, the memory, so memory again so if I tell this again you will all laugh right 512 k. So, we will just assume these are the smart phones right as many as first generation saw some Nokia here to 2 MB the and this you can say. So, these are your different factors and then work load. So, work load is either the scientific experimentation that is being run right. So, it also says secretarial change staff it will be staff the work load in state editing document. So, supporting materials and the third is the faculty of students. So, what ever want. The faulty must be staring at the pdf file.

And then last one I will just users education levels which is not relevant to study, but perhaps in those days they did not have all those windows right. We have to actually like those commands. Have you seen while checking into the air port, what they do, they type sort of long series of command right something slash slash slash. So, that is from the 1980s just user for checking in this not when we log in and check on in web, but is different we have this slides back in and. So, I still remember see folks using ancient codes right to check in and out and so on so on. So, to do that, your level of education matters right.

So, if it is a highly complex system that interface, then this is high school, all of this is irrelevant, but anyways that gives you an idea as to what you mean by levels right. So, now you have to design a set of experiments right that the trace the characteristics of the system performance for all these type of work loads to scientific work load and high school educated person, PG educated person and so on. At the end of you what have this big comparison to say whether this what are right set up parameter try to at the system is shooter.

(Refer Slide Time: 34:24)



So, the simple design is what we do is to start with a basic configuration, the default configuration is some default value for each some parameters right. So, you will always say if you see in the paper the wood papers, they will say this is are all different values, but if I do not specify the default value this right, for each factor you will say this is the default level. For example, CPU if you do not on say anything memory is 8086 and memory is 8MB and so on, so always set the default values for every parameter, and that is always listed in some stay will form.

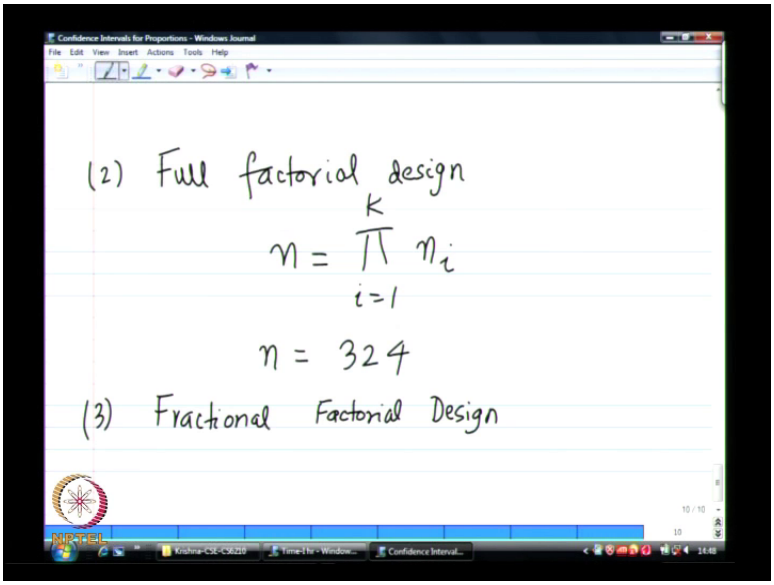
So, we start the default configuration, then that is you are first right this is you are first experiment and most of the times it ends here, one default configuration here are the results, some that and that report, summiting you are report. And I do not need anything else right, but it the simples derive start know the default configuration you want the simpler design then what we do is right, then change one factor at a time. And see this is you are experiment one this is the design of what should be the parameter combination that is all; this time to be right nothing high-tech here. So, take the default say CPU level right.

So, for example, in our case that says 8086, 2 MB, 3 drives and faculty are low and then PG that is your default next. So, take these things and vary the CPU between next two types right one factor time right. So, simply take the default and then change one factor at a time right running its across all the possible values are the particular factor all

possible represents the particular factors. So, the total number of experiments is 1 plus sigma because as for all 1 basic system right at every time take already includes the 1 possible values for the factors.

So, that only n i minus 1 level left for each factors, so simply 1 plus. So, in our case how many experiments will you have 2 plus 2, 4 plus 3. So, you have total of 12 experiments which different in parameter combination. So, this is still better than simply running it for one it saying that done, but it still is not complete because what it does not capture is, right some of the inter actions are not captured. Some of the inter action in the factors are fully captured. So, that is better step, then just having one of set up right parameter combination.

(Refer Slide Time: 38:36)



So, this is your so called simple design say what is simple beginning this is at least this would be expected. If you really want to be thorough then you need to do something more, say what is the next alternative, which full factor in design right self factorial is appearing here, but I do not see this factorial as well the factorial well if you want look at product factorial well.

So, full factorial design is yes right all combination and that is what is expected in a group, a non lab group. I want to put it in the website next time if you want to do project with me, factorial design, if you are willing to do that, then you can join the group. So, if you finish the end of the April, one week before the viva are, and say I am done with the

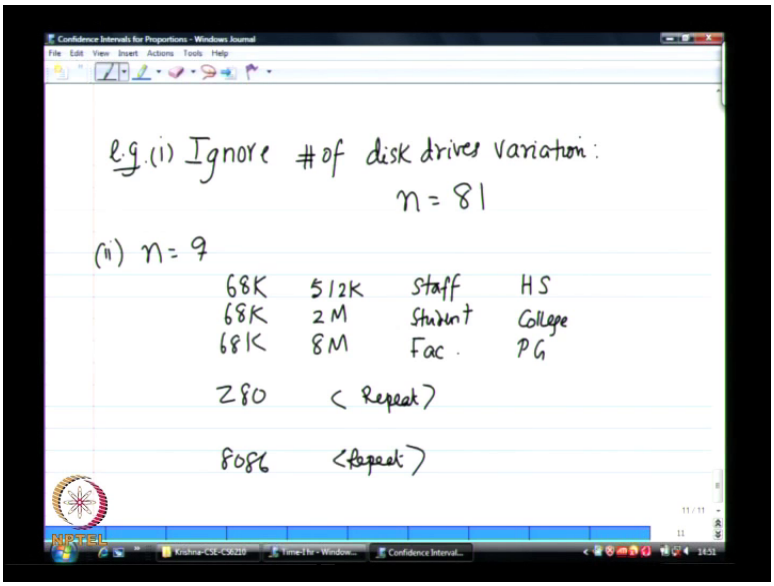
implementation everything working sample, does not go to full factorial, if you would not come back till May or until May, which mean it is serious about the system. If you want think of your project something, you are related on take out or talking with you, and then making our product company. And sell it then will be serious about figure good ideas, this is place to try to any have this is n equals to 324, and what is even more fun is and I talked about replication right.

So, then there is a factor of r. So, each of the one parameter combination, and then each I would like at least 30 times. So, not we can get the z values right. So, that was the reason why people take 3 and half years to complete their MS as they get stuck. What to do you replications too, but it the say otherwise you just one especially the random the system you have to replication we have to. So, this is the you can really find out the right what all combination then the reason for doing this not just sake of exacting we getting, we can also predict right performance of the system by looking at the impact of the factor if you 5 factors you can… And so what is the important of significant important of the factor impact to other factor, and how is interaction between the two different factor right. If I two these factor put to gather the change the system the significant way compare to; so independent of each other and so on.

So, thus why you like to do that, but naturally time consuming right and the expendably and the code keep changing frequently if I save you know find a bug the just before the due date you go back to the all of this remember types. It is good, but it is enough too exhaustive and very time consuming. So, that is why something called the fractional factorial. So, I think both fractional and factorial same set of characteristics except set of n in fractional. So, this is sometime called, this is two right.

So, book talks about fractional factor how to design the set of experiments right in the same set; for example, if you simply do not keep the number of traces fixed, then you can cut down number of experiment to 81 right. We do not vary the number traces I am going to use for the traces, and therefore I can simply have only 81 parameter, which is much better than 324 experiment and so on. So, what we are going to do and next this is end of chapter 16 finished. It is very simple chapter, now we have to the model part right. So, look at full factorial, what is fractional factorial for relevant to skip one.

(Refer Slide Time: 43:01)



The ignore what is the fractional factorial right that I just ignore about this combination, because all ready know the intuition that is there no point only one desire, you know end of the system say and this going to say for sake of and pretty much decided that (( )) right. So, that is one, here is another one, and this is the 68 k do not write 27, it is 9 and this is was staff secretarial staff. So, this is the result of the design, how do you design with coming up later, which will not have time to go through.

So, we have to after all the exams are over, go through the chapters in the book, you want a time actually figure out of fact of how it is done. It fairly straight forward, but we have no time do that this semester. So, here I fix the CPU right when I am going here memory right for type combination of the three each are those then each other exactly ones then you repeat the same thing for (( )). So, percent eighty this same thing is repeated Z 80 same thing is repeated and then 8086 this expect the next simple design right; so this less than the simple design. So, font you can throw that describe also I am ignoring.

If you want you can include this one, I am saying you described combination see after something more than example of the fractional factor, this some there only some of combination that derive, but arbitrarily. But based on one table that generate this was what is the fraction factor is the subset of the combination, wherever slightly compare to our you are simple design. But it try to keep through this combination and try to ensure

that some of these factor can be identified easily. In fact, if you look at book sample this talk about confounded factors. So, two factors combine together on confounded that, because you do not have those values, you only have the combined co-efficient of the impact system performance. So, that is.