Lecture 9 - M_{40} Question 1

Objective: The students will be able to compare the arithmetic computation involved in evaluating $^{n}C_{x}p^{x}q^{n-x}$ when:

(i) n is small (<25) and $p \approx 0,5$

(ii) n is large (>25) and $p \approx 0,1$.

*65 +64 Question: Consider a binomial distribution where n=3 and p=0,5. We would use the term $^{3}C_{2}p^{2}q^{1}$ in order to determine the probability of 2 successes occurring. Calculate the value of $^{3}C_{2}p^{2}q^{1}$ without using binomial tables.

Validity: 5/5

*31 +29 Instruction: After a brief introduction the students are told that it is difficult to compute binomial probabilities for large values of n. In case they had not realized this, they are to be asked two questions to illustrate this point.

Result: 42% (71%) answered the question correctly.

Analysis: Although the result shows a high vairation between groups 1 and 2, it was still considered disappointing. The objective is for the students to compare the computation involved in this question with that in question 2.

The computation involved in this question is very elementary and the students had been involved in many similar computations prior to this lecture, yet about 60% (30%) failed to compute the answer.

The point of difficulty was probably not the computation itself, but the fact that some students did not have immediate recall of the required formula. Thus it was decided to provide them with the formula immediately

+83 after the question was asked.

Result: 93% answered correctly.

Conclusion: This result supports the analysis that some students did not have immediate recall of the required formula and that the point of difficulty was not in the computation. The number of correct answers is now extremely high and the instructional sequence is now considered most satisfactory.

Lecture 6 - M_{40} Question 4

Objective: Students will be able to calculate the difference between the mean μ and the variance σ^2 for a discrete distribution.

*357 +377 Question: Consider a discrete distribution where n=100 and p=0,01. What is the value of $\mu-\sigma^2$ equal to for this situation?

Validity: 5/5

*295 +319 Instruction: The students are shown that the Poisson distribution is the limiting case of the binomial distribution. They are then shown a special property of the Poisson distribution, namely that the mean and variance (μ and σ^2) are equal.

Result: 10% (41%) answered the question correctly.

Analysis: The Poisson distribution has the property that its mean (μ) and variance (σ^2) are equal. Thus to determine whether a discrete distribution is (or is close to) a Poisson distribution, students must be able to determine whether $\mu-\sigma^2$ is equal to nought or not. Hence the objective for this question.

The insturctional sequence however stressed this property of the Poisson distribution and many students assumed the distribution in question 4 to be a Poisson distribution. They therefore stated that $\mu-\sigma^2$ was equal to nought, without evaluating the expression.

+388 It was decided to stress the point that the distribution was discrete but not necessarily a Poisson distribution. Although the special property of the Poisson distribution was that $\mu=\sigma^2$, this had to be determined in question 4 by evaluating $\mu-\sigma^2$. This was inserted immediately after question 4 was asked.

Result: 69% answered correctly.

Conclusion: The change in the instruction clarified matters as the result is substantially improved and bordering on the acceptance level. Further changes should be made to improve the clarity of the instruction.

Lecture 6 - M_{40} Question 5

Objective: Given the formula for a Poisson distribution ($\frac{\mu^x . e^{-\mu}}{x!}$) and the values of μ and x, the students must be able to evaluate the formula, using a calculator.

*450 +484 Question: Consider a Poisson probability distribution where μ=1,3. The formula for a Poisson distribution is $P(X=x) = \frac{\mu^x . e^{-\mu}}{x!}$. Use your calculator to find the probability that X will be equal to 3.

Validity: 5/5

*387 +421 Instruction: The students are shown how the graph of the Poisson distribution varies for different values of μ. They are then shown how we obtain probability values from the Poisson formula by substituting the values of μ and x into the formula.

Result: 17% (30%) answered the question correctly.

Analysis: The students have to perform two sets of behaviours or skills in order to answer this question. The first requires that they substitute the values of μ and x into the formula correctly. The second requires that they can evaluate the formula using their calculators. This second skill is the objective reflected in this question.

It was decided to perform the first skill for the students; that of
+461 substituting the values of μ and x into the formula.

Result: 40% answered the question correctly.

Conclusion: The improvement made to the instruction did not have much influence on the results. Thus the problem area lies in the second skill: the evaluation of the formula using a calculator. Further improvements to the instruction should include a detailed explanation of how to use the calculator in the evaluation of this formula.

Lecture 6 - M_{40} Question 6

Objective: Given the formula for a Poisson distribution ($\frac{\mu^x . e^{-\mu}}{x!}$) and the values of μ and x, the students must be able to evaluate the formula, using a calculator.

*554
+593 Question: Consider a Poisson probability distribution where μ=2,5. The probability that X=6 can be found from the expression $\frac{(2,5)^6 . e^{-2,5}}{6!}$. Use your calculator to evaluate this expression correct to 4 decimal places.

Validity: 5/5

*474
+517 Instruction: The students are shown how to use Poisson probability tables. It is pointed out to them that some values of μ are not tabulated and in these cases the calculator should be used to evaluate the required probability.

Result: 28% (25%) answered the question correctly.

Analysis: No further instruction has been given to the students regarding the evaluation of the formula using the calculator. Thus the result obtained here should reflect a similar percentage as that for question 5. This is the case.

Thus the change made to the instruction for question 5 should also apply here.

Result: 55% answered the question correctly.

Conclusion: Although there is quite an improvement the result is still unsatisfactory and the suggestions made in the conclusion of question 5 for further improvements also apply here.

Lecture 6 - M_{40} Question 7

Objective: Given the mean μ for a Poisson distribution, the students will be able to find the probability of a certain number of events occuring, with the aid of Poisson distribution tables.

*652
+687 Question: Consider a Poisson distribution where $\mu=3$. Use your tables to find the probability that $X=2$.

Validity: 5/5

*581
+618 Instruction: An application of the Poisson distribution to a real situation is illustrated for the students. The situation is in fact described by a binomial distribution but the Poisson distribution can be used as a very good approximation. The reason for this is explained to the students by means of two graphs. A few questions are then asked and answered for the student, based on this real situation. They are then asked a similar question.

Result: 65% (83%) answered correctly.

Analysis: The score obtained for this question by group 1 is very close to the acceptance level. Coupled to this is the belief that earlier parts of the instruction in this lecture have been improved. Therefore it was decided to leave this instructional sequence unchanged.

Result: 81% answered correctly.

Conclusion: As both groups 2 and 3 scored above the level of acceptance it would seem that the instruction was satisfactory and the decision to leave the sequence unchanged is vindicated.

Lecture 6 - M_{40} Question 8

Objective: Given the mean μ for a Poisson distribution the students will be able to find the probability of a certain number of events occuring, with the aid of Poisson distribution tables.

*676 Question: Consider a Poisson distribution where $\mu=3$. Use your tables
+709 to find the probability that $X\leq2$.

Validity: 5/5

*581
+618 Instruction: The students are presented with a problem relating to the Poisson distribution. A question is then answered on this problem as an illustration for the students before a similar question is asked for the students to solve.

Result: 17% (43%) answered the question correctly.

Analysis; Question 7 is very similar to this question and 65% (83%) of the answers were correct. On closer examination it is seen that the difference between the two questions is that only one value of the random variable X is involved in question 7, and its value is quite obvious.

In question 8, several values of X are involved and they are probably more difficult to identify. The objective is concerned with the student being able to find probabilities for certain values of X with the aid of the Poisson tables. Identifying the values of X is not part of the objective (this was covered in previous lectures).

This the values of X involved are revealed for the students immediately
+713 after the question is asked.

Result: 76% answered correctly.

Conclusion: The problem area in the instruction was correctly identified and this result now falls within the acceptance level.

Lecture 6 - M_{40} Question 9

Objective: Students will be able to solve basic problems related to practical situations by applying the Poisson probability distribution.

*714 +754 Question: Say that the average number of customers who appear at a bank teller's window is one per minute. Use the Poisson distribution where $\mu=1$ to find the probability of 4 or more customers arriving at the teller's window in any particular minute.

Validity: 5/5

*691 +733 Instruction: The students are given a problem related to a practical situation. A question on this problem is then answered for the students. They are then asked a similar question.

Result: 21% (54%) answered the question correctly

Analysis: It was felt that the instruction was adequate but the result was below the acceptance level.

There are five instructional changes in this lecture up to this point. If these changes are effective the students will have a better understanding of the content and should do better in this question.

It was decided to leave the instruction unchanged.

Result: 62% answered correctly.

Conclusion: The analysis appears to be partially correct as an improvement did arise. However the result is still unsatisfactory and instructional changes must be made.

The gap between the problem which was answered for the students and the one which they were required to answer may be too large. The narrowing of this gap might result in an improvement.

Appendix III

Transcripts of the Television Lectures

Transcripts of the six television lectures which were produced for this study are given in this appendix. Italics are used in the transcripts to indicate the parts of the original six lectures which were updated or changed. Thus one script in effect covers two lectures: the original and the updated lecture. Reference points which refer to the video cassette tapes are indicated by means of asterisks (*) for the original tapes and pluses (+) for the updated tapes.

These transcripts should be used in conjunction with the video tapes as they are transcribed verbatim and therefore will probably be difficult to follow on their own. When questions are asked on the videotapes, a pause of 30-60 seconds usually follows. This is indicated in the scripts by means of a gap.

M_{35}: Probability Distributions

*55 +58 This is the first of a series of lectures where the format is going to be slightly different. What is going to happen is that after every few minutes (or so) of lecturing, you will be asked a very simple question on the work which has just been covered beforehand. I suggest you try and answer this question just to make sure that you are following the thread of the argument or the actual subject matter which is being presented in the lecture.

Up to now we have discussed the idea of a random event in probability and we considered the addition and multiplication rules which govern the probability of a random event occuring. We later went on to consider what we called 'probability generating functions' and we saw how they generated the probability of different random events occuring.

*78 +81 Having said all that I think we will move onto our first question. This refers to a bit of revision as to what we mean by random events. This is question one. We indicate that by a question mark, zero one (?0), and the question reads as follows: We spin a coin once and what I'd like to ask you is, how many random events are there in this situation or in this experiment when a coin is spun once? Remember that random events are sometimes refered to as sample points. I could really ask you, how many random events or sample points do we have
*99 +101 when we spin a coin once? Just put down the number for the answers.

Let's go directly from question one to question two, again refering to random events. Question two reads as follows: We have a die which is thrown once. The question now will be, how many random events are there in this situation when we throw a die once? Again,
*117 +119 just put down a number for the answer.

We will go directly from question two to question three which again concerns random events and again concerns something which we have considered before in probability; not spinning a coin or throwing a die, but drawing a card from a pack of cards. We have a pack of cards, four suits, with thirteen cards in each suit. We suffle them well and draw one card from the pack. In this experiment, what I would like to know is, how many sample points or random events are there in this situation? Again, just write down the number.

*145
+147 Now we are going to discuss a very closely related concept to random events, namely random variables. As you should know, a variable is a symbol such as X which is used to stand for any one of a set of values. Now a random variable will take on values of particular random events in any experiment. These values are determined only by chance factors. Let's take an example to consider this idea of a random variable more closely.

Say a coin is spun twice. If X designates the number of heads when spinning a coin twice, X will have the values of nought, one or two. When we spin a coin twice and X is equal to the number of heads we obtain, we could obtain no heads, or one head, or of course two heads from spinning a coin twice. So the random variable X in this case (if we define it as equal to the number of heads in our experiment) can take on three possible values, namely nought, one or two.

*171
+172 Bearing all this in mind, let's proceed to the next question, which will be question four. Say we are considering an experiment where we spin a coin four times. We define our random variable X to be equal to the number of heads we obtain when spinning a coin four times. Our random variable X can take on different values. The question I want to ask here is, what is the maximum value that X can have in this situation? Write down a number as the answer.

*195 +196 Now let's go immediately from question four to question five, which refers to the same situation where we spin a coin four times. Our random variable X is again equal to the number of heads we obtain. The question I would like to ask here is, what is the minimum value of X for this experiment?

*210 +211 For each value that the random variable X can have, I think you will realize, there exists a corresponding probability of occurrence of that particular value of X (the random variable). Let us consider this idea by (once again) considering the example of spinning a coin twice.

I have here next to me the sample space which will be generated when spinning a coin twice, by actually putting down the different outcomes that we can have, with actual coins. The first possible outcome over here could be a head on the first spin and a head on the second spin. In other words, two heads from two spins of a coin. Another possibility would be a head followed by a tail on the second spin; or we could of course have a tail on the first spin and a head on the second spin; that would be another sample point in our sample space.

Finally we could have obtained a tail on the first spin followed by a head on the second; that is a fourth possibility in our total sample space as it were.

Now for this particular random event, our random variable X would take on the value of nought, because there are no heads. For this value of the random variable, X would take on the value of one because there is one head. But if you look at this next random event, X also has the value of one; so although these two outcomes are different permutations of one another, they both constitute the same result: one head and one tail, even though there are two ways of obtaining that result.

Now, finally going back to our first random event or sample space, X here will take on the value of two. For each value that X can have, nought, one or two, there exists a corresponding probability of occurence. We can write all this down in the following way.

*259
+258 Put down a heading in the first column, S.S., which refe to the sample space. Next to that we will have a column which will give the value that the random variable X will take on and in a third column we will put down the probability of X actually taking on that particular value.

The first event we considered was a head followed by a head. We will indicate that in the sample space as H.H. The random variable X will then have a value of two, and the probability of that value occuring is one in four, because that is one sample point (or random event) out of a total of four.

*272
+271 We can list the next point in our sample space (a head followed by a tail) as H.T. The value of the random variable X will then be one and its probability of occurrence, I have left for a while, because the random variable X has the value one for another event: namely a tail followed by a head i.e. T.H. Both of those events generate a X value of one. The probability of X taking on the value one is two in four or one-half. I've grouped those two together as you see there on the screen.

Finally, the last point on our sample space is tail followed by tail, or T.T. The value of our random variable X will then be zero and the corresponding probability of that value of the random variable X occuring, P(X), will be one in four once again.

A point to note here is that we have considered all the points in the sample space. Therefore we have considered all possible values of our random variable X and therefore all those probabilities of those values of X occuring must add up to a total of one. You see that to be the case on the right-hand bottom of your screen.

*295
+272 The same sample space could be written in a different manner. Let's indicate it as follows: we start off at some point in space by indicating a dot and you spin a coin once. Therefore there are two possible outcomes, a head or a tail. We draw a line upwards and indicate at the end of it a H (indicating the result of the head occuring) and a line downwards with a T (indicating the results of a tail occuring). You can see it indicated on your screen now.

That is when you spin a coin once. Now you could spin a coin a second time and again you obtain a head or a tail. This sample space now grows. We go up to the top point, namely the H, and we go upwards again to the one possible outcome, namely H, or we could have obtained a tail, so we go downwards to a tail. That is the one route. What if we had obtained a tail on the first spin? The second spin would have yielded a head or a tail as well, so we could go up to the H or down to the T.

*318
+316 If we had a look at the top line of the screen and reading it from left to right, you will read H followed by H. That is the one random event in our total sample space, head followed by head; or we will get a head followed by a tail - that is another random event or sample point. Going downwards we get a tail followed by a head or a tail followed by a tail. Reading that from left to right you will obtain your possible outcomes. If you add the ends of the lines you will get a total of four random events in your sample space.

This type of diagram is usually refered to as a 'probability tree'. I think the reason for that is quite obvious. It grows or branches out like the branches of a tree, except that we draw it on its side and not from bottom to top, as it is usually much easier to draw from left to right than from bottom to top.

*338
+335 Bearing this type of diagram in mind let's proceed to the next question which is question six. We consider a situation where we spin a coin three times. We define the random variable X to be equal to the number of heads we obtain when spinning a coin three times. The question I'd like to ask is: What is the probability of X taking on the
*347
+345 value of two for this particular situation?

Now you might be struggling a little bit to try and solve that, so let me give you a bit of help on the board. I have put up the diagram which you saw a little earlier; the probability of, or the experiment when, you spin a coin twice; so we start at some point in space where we get a head or tail in the first spin. This sort of line in the tree or this growth point in the tree corresponds to one spin of the coin. We then generated a head or a tail in the second spin via that route, or a head or tail via this route. Let's say the growth point indicating 2 spins of the coin, represents possible outcomes.

So. in this question we are considering 3 spins of the coin (which some of you might be struggling with) so we would say the tree would then grow. Take this route, say head followed by head, at this point. Here, we spin the coin a third time to obtain either a head or a tail for the third spin, via that route. If you had followed this route, head on the first spin, tail on the second, once again you could have obtained either a head or a tail. This route here: tail followed by head, a third spin would have then yielded either a head or a tail. This route here: tail followed by tail - head or tail as well.

This growth point here now corresponds to three spins of the coin and there are: 1,2,3,4,5,6,7,8 possible outcomes, or sample points or random events, in the entire sample space. For example, if you wanted to find the probability on obtaining 3 heads- head followed by head followed by head: only one of those routes out of a total of eight, giving you a probability of occurrence of one in eight.

You want to find the probability of two heads occurring so you must find the routes where you get two heads and there might not just be the one route. There might be more than one route - but I'll leave that up to you now to answer the question. We will go back to the question which was asked: What is the probability of obtaining two
+385 *heads for this situation?*

*361
+394 Let's proceed directly from question six onto question seven. Question seven is similar to question six although we are considering the situation here where we don't spin a coin three times but we spin a coin four times. We define a random variable X once again to be equal to the number of heads we obtain. The question is: What is the probability that our random variable X will be equal to four?

Once again, a little help: I'm not going to put another diagram on the board, but you should realise that you must extent your probability tree from three throws of the die to four throws. You will generate extra events and from there you should be able to find the probability of X being equal to four, out of your total number of sample
+409 *points.*

*379 +417 On the basis of the idea of a random event we can now proceed to discuss the concept of probability as a function. You might well ask, what is a function? Well, you should have been introduced to the concept of functions and relations at school. I'm going to review briefly the concept of a function so that we can go on to talk about probability as a function.

You should recall that a function is a special type of relation. I think we will indicate this by putting up a mapping where we have a domain and a range. We will choose values in the domain, say, one, three and four and values in the range of say, nine, eleven and twelve. We have a relation between the domain and the range where one is mapped onto nine, three is mapped onto eleven and four is mapped onto twelve. Some relation exists between the numbers one, three, four and the corresponding numbers, nine, eleven, twelve.

*398 +434 That mapping that you see there is a special type of mapping in that for every value in the domain there is one and only one corresponding value in the range. Not more that one value in the domain can have the same value in the range. In this case, this certainly does not happen. When we have that sort of special relation, we have one definition of a function. You will probably recall this from your school mathematics.

The probability of an event can be considered a function because corresponding to each distinct, possible value of the random variable X, there is one and only one real number (which will be between nought and one inclusive), called its probability. If we were to draw a similar mapping for the probability of an event we would get the following:

*411 +447 For the domain, we would have a random variable X and its possible values from x_1 to x_n. On the right-hand side where we have the range, we now put the probability of X occuring. We have $P(x_1)$, $P(x_2)$ down to the probability of the random variable X taking on the value x_n. The two (the domain and range) are connected by a certain relation - a very special type of relation which is a function. Let's have a look at this more closely. The value x_1 has a corresponding probability of occurence, $P(x_1)$. Similarly, the value x_2 has a corresponding probability of occurence $P(x_2)$. Now x_1 cannot have a probability of occurence $P(x_1)$ and $P(x_2)$. One event cannot have more than one probability of occurence. Therefore we will never have one value in the domain mapped to two values in the range which is, as I said before, one of the definitions of a function. Therefore we have a function mapped on the screen. This would continue until we get down to x_n. That value of the random v   ·le X has a corresponding probability of occurence $P(x_n)$.

*435
+468 As this special type of relation exists between the random variable X and the probability of occurence of those values of the random variable X, we sometimes refer to this relationship as a function.

A point to note here (now that we have introduced the idea of a function) is the notation that is sometimes used. I want to put this on the screen so that you don't get confused when you see one or the other type of notation. The probability of the random variable X taking on a certain value is sometimes indicated as $P(X)$. That can also be written as $P(X=x)$. Those two mean exactly the same thing. Finally, due to the idea of a function, when X has a certain value, the corresponding probability of occurence is written as $f(x)$. Those three mean the same thing: the probability of a particular value of the random variable X occuring.

*457
+489 Wh· ı we spin a coin twice, we could write down all the values of X and the corresponding probabilities of occurence $f(x)$ as follows: firstly we put down the possible values of the random variable X, (where X is equal to the number of heads occuring) namely nought, one and two. Below that we put down the probability of occurence, $f(x)$. For the value of nought, it is one in four: for the value of one it is one in two and for the value of two it is one in four. Notice, there are two identical values in the range of one quarter, but you cannot have one value in the domain (say for example nought) having two probabilities of occurence (say $\frac{1}{4}$ and $\frac{1}{2}$). The value in the domain can only have one possible probability of occurence, and therefore we have a function. This type of diagram or layout is usually referred to as a 'probability distribution'. It gives the possible values of the random variable X and the corresponding probabilities of occurence.

*477
+508 Let us now consider <u>question eight</u>. We consider the situation once again where we spin a coin three times. We once again define X to be a random variable equal to the number of heads we obtain. Bearing this new notation in mind that we have just refered to, what is $f(3)$ going to be?

If you are not sure what the sample sp[illegible] look[illegible] like, I suggest you refer back to the diagram that I had put [illegible] the board earlier,
+518 *where a coin was spun three times.*

*496
+526 We are next going to consider the types of probability distributions. There are two types, namely discrete and continuous probability distributions. I'd now like to point out to you the essential differences between the two. We will do this by setting up a table. On the left-hand side we will put down everything pertaining to discrete probability distributions and on the right-hand side everything pertaining to continuous probability distributions.

Firstly, the random variable involved with a discrete probability distribution is refered to as a discrete random variable. I've abbreviated it here as D.R.V. In the same way, a random variable involved in a continuous probability distribution is a continuous random variable or C.R.V. as I have indicated it here.

*511
+539 Probability distributions which are associated with discrete random variables are refered to as probability functions. Probability distributions involving continuous random variables are refered to as probability density functions. What is the essential difference between the two? The easiest way to explain it is in a very simple and uncomplicated manner which gives you the general idea of the difference between discreteness and continuity.

Discrete distributions involve discrete random variables with discrete values. By that I mean, they have particular values. There is a measureable space between the one value and the next. We could indicate this by a dotted line as you see on the screen. Examples of this are integers, one, two, three, etc. and there are measureable distances between any two integers; they are exact values.

Continuous probability distributions (or probability density functions) involve continuous random variables where there is such a small distance between two points, that the distance is immeasureable. There is an infinitesable distance between the one point and the next. We can never distinguish the one from the next. If we had to plot these values, they would be so close together that they would form a straight line. This is the concept of continuity explained in a very basic manner against the concept of discreteness.

*534
+562 In order to further illustrate the difference between discreteness and continuity and their corresponding distributions, we will look at a graph of a discrete probability distribution and then of a continuous probability distribution.

First of all then, the graph of a probability function (that's the discrete case). There we have our random variable X. In this case it can take on values of nought, one or two. The graph I have drawn here refers to the experiment where we spin a coin twice. On the Y-axis you can see f(x), the probability of occurence. For X having the value of nought it is one-quarter; for X having the value of one it is one-half and for X having the value of two it is one-quarter again. Notice that this looks like a bar chart - it refers to discrete values of X; nought or one or two.

*549
+-76 Before we look at the graph of a continuous random variable (or a probability density function) let us move onto the next question which will be question nine. Say we are considering a situation where we refer to a probability function. If we are refering to a probability function, what will our random variable be? Will it be discrete or will it be continuous? Write down one of those two words as your answer.

*563
+589 We will now have a look at the graph of a probability density function. If you can imagine the values of the random variable X in the previous graph getting closer and closer together until you cannot distinguish the one value from the next then we would not be able to distinguish the one line of the bar chart from the next if we had to graph these values. You would get a completely shaded area as you see on the screen now. We would be able to determine the shape of the distribution from the boundary line of the shaded area. This would tell you that some middle value of the random variable X on the X-axis has a greater probability of occurence (because that is where the peak occurs) than values of the (continuous) random variable to the right or left of the middle. A continuous probability distribution or a probability density function often has a shape which looks something like that when we draw a graph.

Bearing the concept of continuity in mind (where we cannot distinguish between one value of X and the next, because they are so close together) let us proceed to the next question. This is <u>question ten</u>. We are considered a probability density function. The question is: Is this a feasible statement to make: $P(X=2)=0,2$? The answer
*586 +611 will simply be yes or no.

That is the final question for this lecture. What we have done is to give you a brief introduction to random variables and their corresponding probability distributions. In the next two lectures we will consider some of the properties of probability distributions. (Answers
*596 +620 to the questionsset in this lecture will now follow.)

*597-636
+621-658

<u>ANSWERS TO QUESTIONS</u>

*13
+09 M_{36}: Expected Value of a Ramdom Variable

In this lecture we look at one of the properties of probability distributions, namely its mean, or as we called it in the title of this lecture, the expected value of a random variable. What we really do here is to start to link together descriptive statistics and probability theory and this is eventually going to lead us to what we refer to as 'statistical inference', and that is where we actually make decisions based on sample data which we have collected.

*49
+45 Before we go on to actually start looking at the mean or the expected value of a probability distribution, there are two fundamental concepts which you must have very clear ideas of in the subject of statistics, and they are what we call the 'sample and the population'.

Now in order for you to get some very clear ideas of this let's put down these two concepts in two different columns: the sample and the population. Underneath that put down some of the essential differences between the two.

First of all, starting off with the sample, I think you probably all realise that when we collect a sample of data or a sample of observations, we mean that the number of observations is of course finite or countable; n, the number of observations or pieces of data that you have collected will always be finite in a sample. However, in statistics when we refer to a population, we usually mean that the amount of data you have is almost infinite. In other words, the number of observations n, tends towards infinity, as you see indicated on the screen; n tends towards infinity with a population. We can usually not count the number of sets of data in a population.

*86
+81 If you think back to your descriptive statistics, where we actually described a sample of data as best we could, where we found measures of central tendency and measures of dispersion, the measure of central tendency which we spent most time on was the mean, and the symbol we gave to it was x bar ($\bar{x}$), which of course describes how the data in a sample clusters around one particular value, or tends towards a central value.

The symbol we use for the mean of the population is not $\bar{x}$, but the greek letter mu (μ) and we read it as we write it there, mu, and the symbol itself is as you see it on the screen now (μ).

*104 +101 Then the measure of dispersion which we spent most time with when we described sample data, which gave us of course a measure of how the different values are dispersed or scattered about the mean, was the standard deviation. The symbol we used there was capital S and we sometimes put a subscript to that, depending on whether the data was x, y or z data; in this case usually it is x, so we see it as S subscript x (S_x).

The corresponding symbol we use for the population's standard deviation is once again a greek symbol or letter which is sigma (σ) and it appears as you see on the screen now. Some of you might be a little confused at this because you have already met the greekletter sigma which did not look like that and which meant 'add together' or 'find the sum of.' Well that also is the greek letter sigma but it is the capital case of the greek letter sigma; an M on its side, as it were. This is the small case of the greek letter sigma, which refers to the standard deviation of a population.

*130 +127 You will find that all our parameters which describe a population we usually give greek symbols to or greek letters.

Bearing all this in mind now, let's proceed to the first question for this lecture. This will be <u>question number one</u>: Say we are refering to a population. Then the mean of the population is described by the following greek letter as you see on the screen. The question is, very simply: What is [illegible] name of this greek letter? Just put

*145 +141 down the word for the an[illegible]

We go from question one directly on to <u>question two</u> which is another very straightforward question. In question two we are again considering a population and one of the parameters of a population is described by the greek letter sigma. The question here is: What does sigma represent? What parameter of the population does the greek

*162 +158 letter sigma represent?

We will leave question two there, and what we are now going to do for the rest of this lecture is to examine how to find the mean of a population. We will start off by reviewing how we actually found the mean of a sample.

If you recall from your descriptive statistics we defined the mean, $\bar{x}$, to be the sum of all the x values, divided by n: the simple average. Sigma (Σ) x_i over n. By the way, there you see the capital case of the greek letter sigma. Sigma x_i means add all the x values together, divide by n, and we obtain $\bar{x}$. However, as often occurred our data was arranged in frequencies. In other words a particular value of x occurred several times and then instead of multiplying all these out we just changed our definition of $\bar{x}$ (which come to the same thing basically) to be $\bar{x}$ being equal to sigma of x_i multiplied by the number of times that value of x_i occurred. In other words times f of i (f_i), all over n,
*195
+190 where n is the total number of x values or x observations in our data.

Now I am going to ask you a very simple question on how to find the mean $\bar{x}$, but before we do that let me just give you a very simple example on the blackboard of how to find the mean $\bar{x}$ by using that bottom formula on the screen: $\bar{x}$ equal to sigma x_i times f_i over n.

Say we have a set of data given to us as a frequency distribution. We have certain values of x_i with corresponding frequencies of occurrence, f_i. Now I am just taking very small values: two occurs twice; four also occurs twice. I think you can find mentally in a matter of seconds what the mean of that is going to be. Let me use this formula, $\bar{x}$ is equal to sigma of x_i times f_i all over n, just to refresh your memories or help you use this. This means the sum of the product of these two things, x_i times f_i. Let's write this out: the first value of x was two multiplied by its frequency of occurence, which was two. We multiply the two together, and add that to the next value of x which occurs, which was four multiplied by its frequency of occurence which again is two. Divide that by n. The common mistake here is that there is not two values of x (there are two different values of x) but a total of four values of x in our sample data. Divide by n, which is four in
+225 *this case, and that comes out to three.*

Now that I have refreshed your [illegible]ies slightly, let's go on to <u>question three</u> where we give you an example to find the mean using this
*200
+229 formula. Question three will read as follows: we have data x_i and f_i listed in the following way (I think you should take this down);

the values of x are as follows: the first one is ten with frequency of occurence three; the next value of x is four, frequency of occurence three; the next value of x is six, frequency of occurrence two and the last value of x is nine and that only occurs once. The question is to find $\bar{x}$. What is $\bar{x}$ in this case?

*231
+261 We will leave question three there and lets have another look at that formula you should have been using for $\bar{x}$. Namely that $\bar{x}$ is equal to sigma of x_i times f_i over n. What I'm going to do now is to rearrange that formula slightly in the following way: write it as $\bar{x}$ being equal to sigma of x_i multiplied by f_i over n $(\frac{f_i}{n})$. As you see there I've grouped f_i over n together and that is for a special reason which should become apparent in a moment. Using this form of the formula (it will make no difference to the final answer) let's take the figures from question three and just substitute in here to see that you actually do get the same answer as you should have obtained in question three. The first value of x was ten so we could have written that as equal to ten multiplied by the relative frequency of occurrence, (f_i over n) is three out of a total of nine values. The next value of x was four. That also occurred three out of nine times, so its relative frequency of occurrence is three in nine. The next value of x was six which occurred two out of a total of nine times, so it is six times two-ninths and the fourth value was nine which occurred once out of nine times, so its relative frequency of occurrence is one in nine. The reason for me doing this will hopefully become apparent a little later because probability remember is the limiting case of the frequency of occurrence of the number of say, successes, in the total number of outcomes.

*267
+295 That would give us thirty plus twelve plus twelve plus nine, all over nine and I will leave it to you to check that you obtain the same answer as you should have obtained for the answer to question three.

*270 Now I would like to have a closer look at this formula. There we have it coming up on the screen:

$$\bar{x} = \Sigma x_i \cdot \frac{f_i}{n} \quad .$$

Let's consider this more closely. Say we would now like to find the mean of a population, and not a sample. This would mean that n becomes very, very large; in fact it tends towards infinity. In that formula, if the value of n (as you see b[illegible] indicated) does get very large and tends to infinity (it doesn't matter what the value of f is going to be) - any number divided by infinity comes out to an answer of zero or tends to zero - it means that $\frac{f_i}{n}$ will always be zero and our mean therefore will always be nought which wouldn't be at all satisfactory. Therefore we must devise some method of determining the mean for the population which doesn't always come up with the answer of nought which is obviously incorrect.

How do we do this? The answer is to consider the mean μ as the expected value of the random variable, which comes from probability theory. Before proceeding any further, let us see how this comes about. If you recall your definition of probability was $\lim_{n \to \infty} \frac{f}{n}$ or the relative frequency of occurrence divided by the total number of values. As you see being indicated, that is exactly what you have there. So $\frac{f_i}{n}$, if n gets very, very large is no longer a practical value but a theoretical value. It is a probability of occurrence of that particular ith value of x. This now gives us an alternative formula to write down for the mean of a population.

On the next line we can write:

$$\mu = \Sigma x_i \cdot P(x_i) \quad .$$

We now come back to the expression I used a moment ago, the expected value of X. μ is an average or expected value of X in a population, so we sometimes write it as E(X), which is read as the expected value of X.

$$E(X) = \Sigma x_i \cdot f(x_i) \quad .$$

One point that I might mention at this stage: for that formula to apply to a population, we are talking about discrete values or random variables because we must obtain discrete values of x to multiply them by their corresponding probabilities of occurrence. This wouldn't really apply to continuous values or continuous random variables; only discrete random variables.

This all leads us onto two more questions. The next question is

*341 going to be <u>question four</u>

+300 *Now let's have another look at this rearranged formula, namely*

$$\bar{x} = x_i \cdot \frac{f_i}{n}$$

If we wish to determine the population mean (the symbol we gave to it was μ) we must divide by the denominator n which will be infinity in the case of a population. Anything divided by infinity will come out as zero (or it is really undefined). Therefore that formula as it stands is not good to us at all. We will not be able to determine the mean of a population when we are dividing by an enormously large number for n, like infinity.

Some other device must be found to determine the population mean μ. The solution to our problem is to consider the mean μ as the so-called 'expected value' of the random variable which comes from probability theory. Before proceeding any further let's see how this actually comes about.

If you recall the definition which you had for probability: it was the number of successes out of the total number of outcomes in a sample space which was $\lim_{n \to \infty} \frac{f_i}{n}$.

Let's just indicate that for you. On that top line we have a quantity which is

$$\frac{f_i}{n}$$

as you see being indicated for you now. Now the definition of probability was $\frac{f_i}{n}$ *as n tends to infinity. As you see on the following line we can write the mean for the population* $\bar{\bar{x}}$ *as:*

$$\mu = \Sigma x_i P(x_i) \ .$$

We now have an expression for the mean of the population. It is a theoretical type of expression because we are using a theoretical concept like probability to define the quantity $\frac{f_i}{n}$ *as n tends to infinity. What it really means is the probability of that value of x occuring times that value of x.*

Then on the third line as you see shown there, the mean of a population is usually referred to as the 'expected value of x' and we write it as:

$$E(X) = \Sigma x_i . f(x_i) \qquad \text{where } f(x_i)$$

means the same as $P(x_i)$*.*

A few new ideas are presented here but if you compare the bottom line to the top line you have essentially got the same thing; except that instead of writing $\frac{f_i}{n}$ *we write* $f(x_i)$ *for a population as* $\frac{f_i}{n}$ *is the definition of probability if n tends to infinity. Now let us go*

+367 *on to question four.*

Question four reads as follows: we have an expression $\lim_{n \to \infty} \frac{f}{n}$. This is an expression which you saw in some earlier work. What quantity

*347
+372 or what concept does that mathematical expression define?

Let us once again go directly to another question i.e. question five which will still refer to these new definitions which you have just received. Question five reads as follows: the mean of a population

$$\mu = E(X) \ .$$

How do we read the expression E(X)? We read it as the ________ value of X. What is the missing word?

*375 +406 Now let us go on and try an example where we in fact find the mean of a population using our formula we have just derived. The example we will consider is that of a fair die. We define our random variable X to be equal to the face value which shows when we throw a die. Find the mean or expected value for this situation.

Are we dealing with a population here? We are, because for the probability of any one of the faces coming up of a fair die to be one in six means that theoretically we have to throw that die an infinite number of times so that the ratio of one side coming up to any other side is one to six. We are assuming here that we have actually thrown the die an infinite number of times. This is in fact a theoretical situation but n actually has tended towards infinity.

*392 +42[illegible] The first thing to do is to write down the probability distribution for our experiment. We put down the values for the random variable X. When throwing a die there will be six values and below that we write the corresponding probabilities of occurrence f(x).

Probability Distribution

X:	1	2	3	4	5	6
f(X):	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

We now want to find the average or mean value or expected value for this theoretical situation. The formula we have for the mean μ is

$$E(X) = \Sigma X.f(x) \ .$$

Once we have the probability distribution (as we have here) it is a very easy matter to substitute in this expression. The first value of X was one; its probability of occurrence was one sixth so we write this as

$$1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6}$$

$$+ \ 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} \ .$$

*418
+446 Note here that I could have taken out a common factor of one sixth (it would have saved me the trouble of writing it down six times). However I have left it there so that you can get the idea that we are multiplying every value of the random variable X by its corresponding probability of occurrence, one in six. This is very similar to our definition of the mean of a sample. However we have assumed that the die has been thrown an infinite number of times and one in six is the ratio of the number of ones occuring to all the other values ocurring, and so on. Evaluating the expression it comes to

$$\frac{21}{6}$$

$$= 3\tfrac{1}{2} \; .$$

*428
+456 Therefore the average, mean or expected value when we throw a die is three and one half. You might immediately ask how one can obtain a value of three-and-a-half as there is no three-and-a-half value on the die. Remember that this is only an average type of value. For example, you might conduct a survey and find that six comma four people, on average, earn above a certain amount. Well, you can't get comma four of a person. You all realise that it is simply an average type of value as this is. The expected value that we would get when we throw a die is three-and-a-half; a theoretical type of figure.

Here is a question for you to try on finding the mean of a population. This will be <u>question six</u>. Say a coin is spun twice and we assume that the spinning of the coin twice is done an infinite number of times so that the ratios we get eventually fix themselves as n tends to infinity. X, our random variable we define as being equal to the number of heads involved when we spin a coin twice and we would like to
*449
+476 know what the expected value of X will be for this particular experiment.

Before you embark on this let me help you with an explanation on the board.

The first thing to do is to write down the probability distribution for this situation:

X:	0	1	2
f(x):	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$.

Now you have to find the expected value of X:

$$E(X) = \Sigma X f(x) \ ,$$

and from what I have done cn the board you should now find this answer quite easily.

*461
+503 Let's now leave question six there and now go on with some more work on the expected value of a random variable. Sometimes it happens that you might not always w nt to find the expected value of X but the expected value of some function of X. Let's work through this so as to explain what I actually mean. The expected value of X was defined as

$$E(X) = \Sigma x.f(x) \ .$$

Now say that you would like to find, not the expected value of X, but the expected value of some function of X, like X^2 or $3X+2$ or X^3+4X+3, which might very well occur. Then we would have

$$E(g(x)) = \Sigma g(x).f(x) \ .$$

Let's take a particular example. Say $g(X)=X^2$. Then $E(X^2) = \Sigma X^2.f(x)$

*480
+521 Let's actually do this for the case of our die. X^2 is now the square of the face values.

$$\therefore \quad E(X^2) = 1^2.\frac{1}{6} + 2^2.\frac{1}{6} + 3^2.\frac{1}{6} + 4^2.\frac{1}{6} + 5^2.\frac{1}{6} + 6^2.\frac{1}{6}$$

$$= \frac{91}{6}$$

$$= 15\frac{1}{5} \ .$$

Note that the expected value of X was $3\frac{1}{2}$ and the expected value of X^2 is not $(3\frac{1}{2})^2$, but $15\frac{1}{6}$. This is something we will come back to a little later when we look at the properties of expected values. Now let's do an example where you actually find the expected value of some function of X. This will be <u>question seven</u>.

*499
+539 Once again refer back to the situation where you spin a coin twice (an infinite number of times). The random variable X is once again equal to the number of heads and we want you now to find the expected value, not of X, but of X^2.

Most of you should be able to do this now but I will give you a little help on the board.

Always put down your probability distribution first. In this case the probability distribution for X^2 is:

X :	0	1	4
f(x) :	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Now you should be able to find the expected value of X^2 from that distribution quite easily. +552

*513
+562 Hopefully by now you are able to find the mean or expected value for a population. What we will do for the rest of this lecture is to look at some of the properties of operating with expected values of X or expected values of functions of X.

<u>Properties of E(X)</u>

From the previous two questions you should realize the first property I want to write down:

$$[E(X)]^2 \neq E(X^2) .$$

What is the expected value of a constant? Let us call the
*531
+576 constant c.

$$E(c) = \Sigma cf(x)$$
$$= c\Sigma f(x) \ .$$

Now what is $\Sigma f(x)$ equal to? It is the sum of all the probabilities of all the possible values of x occuring and the sum of all the possible probabilities in any experiment always adds up to one. So it is c times one which gives a [illegible] of c. The expected value of a constant is simply that constant. If you think of it intuitively, expected value means the average value or the theoretically expected value when you do something with a variable. In this case you don't have a variable, but a constant and it doesn't vary at all, so its expected value will always be its actual value, c. A property to bear in mind, which we use in later lectures.

Another property: What is the expected value of a constant multiplied by a random variable?

$$E(cX) = \Sigma(cX).f(x)$$
$$= c\Sigma X.f(x) \ .$$

*548
+596 Look at the expression on the bottom line: $\Sigma X.f(x)$. What is that the definition of? That is the simple, straightforward definition of the expected value of X. The final answer will be:

$$c.E(X) \ .$$

We have shown therefore that the expected value of a constant times the random variable X is equal to the constant c multiplied by the expected value of X. Again, something worth remembering: we can remove the constant from inside to outside the brackets for expected values.

We will stop there with these properties for the meantime. Let us now ask you a question which involves some of these properties we have looked at up to now. This will be question eight.

*563
+609 Say we are considering a particular situation where the expected value of X is equal to $4\frac{1}{2}$ or 4,5. What then is the expected value of six. i.e. E(6X)?

Let us now go on to a few more properties of expected values. What is the expected value of two random variables going to be equal to? Say we have two random variables X and Y. So the expected value of a sum will be:

$$E(X+Y) = \Sigma[Xf(x) + Yf(y)]$$
$$= \Sigma Xf(x) + \Sigma Yf(y)$$
$$= E(X) + E(Y) \ .$$

*584
+628 The expected value of the sum of two random variables is the expected value of each of those random variables added together.

We will have a look at one more property: The expected value of a constant added to a constant times a random variable

$$E(a+bX) = E(a) + E(bX)$$
$$= a + bE(X) \ .$$

Another property worth noting which we might use in later lectures.

Now for a question on these last two properties that you've just seen. This will be <u>question nine</u>. Say we are considering a situation once again where the expected value of X^2 is 3 $(E(X^2)=3)$.

*600
+646 What then will be:

$$4 + 2E(X^2)?$$

Let's now go directly to question ten which is very similar, but arranged in a different manner. <u>Question ten</u> then will be concerned with the same situation where

$$E(X^2) = 3$$

What is $E(4+2X^2)$?

*618
+662 This is the last question of this lecture. In the next lecture we will go on to look at another property of populations or rather random variables, and that will be the variance of a random variable. Until then, thanks very much. The answers for the questions will
*624
+667 come up now so you can check your answers.

*633-671
+676-712 ANSWERS TO QUESTIONS.

*017
+012 $\underline{M_{37}}$ Variance of a Random Variable

In the previous lecture attention was focused on ways of calculating the mean μ of a population. If you recall, we defined μ to be the expected value of a random variable X. In this lecture we are going to focus attention on the variance and therefore of course, standard deviation of a population.

We will start off the lecture by once again considering the essential differences between a sample and a population. Once again, by putting the two headings down and listing the differences beneath them we see that n is finite, whereas for a population, n tends towards infinity. The symbol we used for the mean of a sample was $\bar{x}$; the symbol we used for the mean of a population was the greek symbol μ. The standard deviation was S and that was a measure of dispersion: it gave us some indication of how the data was spread about the mean $\bar{x}$. Similarly for a population, the standard deviation σ will tell us something about the spread of the population data about the mean of the population μ.

*073
+070 Bearing this in mind, lets go straight on to the first question. very straightforward question. This will be <u>question one</u>. It refers to a concept which some people get a little confused with but it is very easy; I've asked you this question now so that you get the idea of the difference between the two concepts straight away, and don't confuse them as is often the case. Question one reads as follows: we have the variance of a population described by the symbol σ^2. What
*90
+84 quantity or what parameter does the symbol σ describe for a population?

*101
+97 I don't think you should have had much trouble answering question one. What we are going to do now is to have a look at how to calculate the variance of a population. We will start off by reviewing how we calculated the variance for a sample. If you review the formula that we used to find the variance of a sample, we will see if we can possibly find an expression from it for the variance of a population. If you remember, it was the mean square of the differences from the mean and it is described by that expression on the screen now:

$$S_x^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}.$$

*123
+119 We used this formula in another manner, when we were presented with a set of data with a frequency set-up. In other words one value of x occured more than once. We rearranged that formula slightly and used it in the following way:

$$S_x^2 = \frac{\Sigma(x_i-\bar{x})^2.f_i}{n}.$$

*137
+132 I am going to ask you a very simple question to find the variance using that second formula because that is the one we are going to use to develop a formula for the variance of a population later in this lecture.

*142
+137 *Before I ask you the question just let me do a very simple example on the board to refresh your memories on how to use this formula.*

If you remember, in a previous lecture, we had the data:

x_i	f_i
2	2
4	2

We found that $\bar{x}=3$. *thus the variance,*

$$S_x^2 = \frac{\Sigma(x_i-\bar{x})^2.f_i}{n}$$

$$= \frac{1.2 + 1.2}{4}$$

$$= 1 .$$

The standard deviation, which is the square root of this, will also be one.

I think that is enough to refresh your memory on how to use this formula. Now for a question where I want you to use the same formula to find the variance for a set of sample data.

*142
+173 This will be <u>question number two.</u>

We have data listed under x_i and f_i:

x_i	f_i
10	3
4	3
6	2
9	1

$\bar{x}=7$

What is the variance (σ^2) equal to? If your answer works out to an improper fraction, leave it as an improper fraction. Don't worry about working it out as a decimal fraction.

*184 +211 What we are now going to do is to use this formula that you should be using for the variance in this question and see how we can rearrange it, so that we can use it in order to find the variance of a population. This is really what this whole lecture is all about.

*190 +218 Let us have another look at that formula for S_x^2:

$$S_x^2 = \frac{\Sigma(x_i-\bar{x})^2 . f_i}{n} .$$

Now, in a similar way as was done in the previous lecture, when we had the mean $\bar{x}$, let us rearrange that formula slightly:

$$S_x^2 = \Sigma(x_i-\bar{x})^2 . \frac{f_i}{n} .$$

*198 +226 I have separated the $\frac{f_i}{n}$, as it were, because that is the quantity which I want to refer to once again. We are faced with the same problem as we were before. If we are now considering data for a population, n is tending towards infinity and in that expression we then have to divide by infinity; which is undefined or which gives us an answer of nought, which is not necessarily the case. The fact that n tends towards infinity for a population presents us with a problem, when using this formula as a means of calculation.

*207 +235 Again, we find a solution to this problem by using our probability theory. When you are conducting any experiment, if you conduct it an infinite number of times, the ratio of the number of successes (call them f) to your total number of sample points or random events (call then n) is actually your probability when n tends towards infinity. In other words, the probability of an event occuring is defined as the limit as n tends towards infinity of that quantity on the end, $\frac{f_i}{n}$. There you see it high-lighted. Remember that $\frac{f_i}{n}$ is the definition of probability if n tends towards infinity. Therefore, we can now write down a formula which will enable us to calculate the variance of a population, because $\frac{f_i}{n}$ can be written as the probability of that particular value of X occuring. So the variance

$$\sigma^2 = \Sigma(x_i-\mu)^2.P(X).$$

*241 +266 The variance of a population is not always written as σ^2. We often write it in another way as

$$Var(X) = \Sigma(x_i-\mu)^2.f(x_i) .$$

That is, I think the most commonly used formula for the variance of X, although σ^2 is sometimes used. The standard deviation of course will be the square root of that quantity.

*256 +281 Bearing some of these new symbols and notations in mind, lets proceed to the next question which will be question three.

*259 +285 Once again we are refering to a population. The variance of that population, we write using the symbol σ^2. The question here is: how else can we write σ^2 or how else do we abbreicate the variance?

*281 +305 I think we will leave question three there. If you referred back to those equations which you should have taken down, there should have been no difficulty in answering this question. Let us now refer to an actual example where we find the variance of a population.

*285 +310 Let us consider once again the case where we throw a fair die an infinite number of times. We define the random variable X as being equal to the face value of the die. Remember that theoretically we have to throw the die an infinite number of times in order to obtain the ratios of one sixth or one in six for each of the faces coming up, in relation to each other. So theoretically, we are here considering a population situation where n has tended towards infinity. We found in a previous lecture that the mean value μ came to 3,5, even though that value does not appear as one of the face values on the die. Nevertheless I think you understood that this is a mean or average value which you expect to obtain (the expected value of X). What would the variance, σ^2, be in this case? Lets put down the definition:

$$\sigma^2 = \Sigma(x_i-\mu)^2.f(x_i) \ .$$

*304
+327 Now lets substitute in values (I am not going to refer to a probability distribution for this. You should realize that the probability distribution will be Y values going from one to six and all their probabilities of occurrence will be one in six). The first value of X was one so we get

$$(1-3\tfrac{1}{2})^2.\ \frac{1}{6} + \ldots\ldots$$

$$\ldots\ldots\ldots\ +(6-3\tfrac{1}{2})^2.\ \frac{1}{6}\ .$$

*314
+337 If you work this out, in other words, you perform all those subtractions; square each value and then you add up all those numerators, you will find it will come out to be $\frac{35}{12}$.

That was for the case where we threw a die an infinite number of times. Now let's give you a question where you can try to find the variance of a population. This will be <u>question number four</u>.

*321
+344 Say we spin a coin twice (again, theoretically, we have spun the coin twice, an infinite number of times so that our probabilities of occurrence are as we expected). We define our random variable X to be the number of heads which occur. If you remember, you had the same question in a previous lecture where you were asked to find the mean μ, which in this case came out to be one. (I think this coincides with your intuition; when you spun a coin twice you would expect to get one head). The question here is: What is the variance for this
*333
+355 situation?

Before you go ahead with this, let me illustrate what you should do by helping you on the blackboard. Always put down the probability distribution. The possible values of our random variable X are nought, one and two:

f:	0	1	2
f(x):	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

The variance $\sigma^2 = \Sigma(x_i-\mu)^2.f(x_i)$.

Now lets go back to the question and I will give you a minute or
+373 *so to answer it.*

*351
+390 Now we are going to look at a few different ways of writing this formula for the variance of a population. You might recall that the population mean μ was written as the expected value of a random variable X as you see on the screen now:

$$E(X) = \Sigma x.f(x) .$$

The variance was written as

$$Var(X) = \Sigma(x-\mu)^2.f(x) .$$

*362
+399 What I want you to do now is to examine the two right-hand sides of those two equations on the screen. Notice that they are exactly the same except that the first equation has an x, the second equation has a $(x-\mu)^2$. In other words the second equation represents the definition of the expected value of, not X, but a certain function of X, namely $(X-\mu)^2$. In actual fact, the variance of X can also be written as an 'expected value'. We could write that as:

$$E(X-\mu)^2 = \Sigma(x-\mu)^2.f(x) .$$

This is often the case. The variance of a population is sometimes
*377 written as the expected value of $(x-\mu)^2$.
+415 *Just let me stress here that σ^2, $Var(X)$ and $E(X-\mu)^2$ all mean the same thing. They are the variance of a population.*

*378
+415 Hopefully you have assimilated that and let's proceed now to the next question on this very point which will be <u>question five</u>.
*380
+412 Say we are considering a population where the standard deviation σ is equal to two. What would then be the expected value of $(X-\mu)^2$ for this population?

*393
+433 You have probably noticed that a fair amount of arithmetic calculation is usually involved in finding the variance of a population because there is so much subtracting and then squaring and adding to do. This was true in the case of sample data as well and we rearranged our formula for the variance so that the arithmetic was more simplified. We do a similar rearrangement of the formula for the variance of a population, or the variance of a random variable. Let see how we per-
*403
+422 form this rearrangement. We start then with

$$
\begin{aligned}
& E(X-\mu)^2 \\
&= E(X^2-2X\mu + \mu^2) \\
&= E(X^2) - E(2X\mu) + E(\mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(x^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}
$$

*441
+479 Looking at the top line and then the bottom line you might think that we have not simplified things at all. But, if you consider the arithmetic involved when you are using that top expression compared to the arithmetic involved in this bottom expression: we already have the mean μ or $E(X)$; all you have to do is square it. Then, all you have to do is find $E(X^2)$, which means only squaring the possible values of the random variable X, and multiplying them by their probabilities of occurence to get $E(X^2)$. Usually this formula represents a far easier way (arithmetically) of finding the variance of a population and therefore worth taking note of.
*451
+488 Now lets see if you can use this new formula in <u>question six</u>. Say we are considering a population and

$$E(X^2) = 7$$

and $$E(X) = 2 .$$

If you are given these facts, what will the variance of X be (Var(X))? As a little clue here, this shouldn't involve you in
*462 +499 mammoth calculations at all.

*467 +504 We will leave question six there. Let us now use this new formula to calculate the variance of a population once again for the situation of a fair die. Remember that E(X) was $3\frac{1}{2}$. To use this new formula we also have to know what the expected value of X squared is. That is easy enough because we simply take all the values of X and square them.
*472 +510 So we can write that as

$$E(X^2) = 1^2 . \frac{1}{6} + 2^2 . \frac{1}{6}$$

$$+ ....+ 6^2 . \frac{1}{6}$$

$$= \frac{91}{6} .$$

*478 +515 So once we have E(X), this is the only extra bit of calculation you have to do i.e. to find $E(X^2)$. Then the variance of X

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \frac{91}{6} - (3\tfrac{1}{2})^2$$

which gives us a final answer of $\frac{35}{12}$, which is what we obtained earlier in this lecture (I think you might agree) with far less tedius arithmetic involved.

*487 +523 Now for <u>question seven</u>. Once again we spin a coin twice. Say X is the number of heads we obtain as we had before. What you did in the previous question was to calculate the variance. If you had to use your new formula, all you would need to know to calculate the variance is also the expected value of X squared. So in this question, all
*494 +529 that I want you to do is to find $E(X^2)$.

Again, let us just look at the blackboard very briefly before you embark on this question. Here I have the distribution for X, but now I would like the distribution for X^2:

X^2 :	0	1	4
f(x) :	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$.

I think you will realise now that it is a fairly easy thing to calculate $E(X^2)$ from the probability distribution. I will leave you
+536 *for half a minute to actually do that.*

*498
+543 So far in this lecture we have found ways of finding the variance of a population and in a previous lecture we found ways of finding the mean of a population. We have also, in this lecture, shown you some easier formulae to use in order to find the variance of a population in terms of expected values.

Now what we are going to do in order to end off this lecture is to apply these ideas of expected value and variance of a random variable to a more practical type of problem to see if you can apply the theory.

*507
+551 Say a lottery is to be held. One thousand tickets are sold at three rand each. Three prizes are offered. The first prize is 500 Rand, the second prize 300 Rand and the third prize 200 Rand. Let us now define X as our random variable, being equal to the prize money. Could we draw up a probability distribution in this case? If we consider this to be a probabilistic situation consisting of independent events i.e. when any ticket is drawn, it is then replaced so that there are always a thousand tickets. Although n is not infinite here, it is rather large and we could draw up a probability distribution.
*522
+566 The probability distribution would look as follows:

X:	500	300	200	0
f(x):	0,001	0,001	0,001	0,997

*538
+580 Now we are going to apply some of our theory to this situation which will form question number eight.

Question eight reads as follows: What is the expected value of X for this situation? In other words, what is the mean value of X? Simply apply your formula to that probability distribution in order to
*545 +587 find E(X).

*559 +600 Hopefully you have been able to calculate the theoretical mean or expected value of X for this situation. Let us now ask a further question. This will be <u>question nine</u>. Say that you actually purchase one, three-Rand ticket. So that is what you have spent on this venture. From the previous question you should know what you expect to obtain from this venture or outlay. Therefore, considering that you have outlayed three Rand and you know what to expect to bring in, what is your expected loss or gain in this venture? If it is a loss, just put it as minus a certain amount or minus so many Rand; if it's a gain, indicate
*572 +613 it as plus whatever amount it is.

Let us proceed now to a third question on this same situation. This will be <u>question ten</u>. What is the variance going to be for this situation? Remember, the easiest formula for Var(X) is equal to $E(X\ ) - [E(X)]^2$. You should already know what E(X) is from question eight. It remains for you to find $E(X^2)$ which is a farily simple
*587 +627 matter using this probability distribution.

*601 +640 You should find that although you expect to obtain very little, your variance is going to be rather large. I think this is what probably encourages gambling: there is a chance of making a lot of money but most of the people most of the time make very little money or lose large quantities of money. We have now covered how to treat probability distributions in general and we can now proceed to examine some particular types of probability distributions which is what we will be doing in the
*609 +648 next few lectures. Thank you very much.

*613-656 +652-692 <u>ANSWERS TO QUESTIONS</u>

*016
+008 M_{38}: <u>The Binomial Probability Distribution</u>

*030
+023 In this lecture we are going to go on to a particular probability distribution and you are going to have to use your binomial tables to answer some of the questions we are going to ask you. Before we go any further I suggest you get these tables ready so that you can use them when some of these questions come up during the lecture.

The probability distributions which we have examined so far have fallen into two categories - one being discrete and the other continuous probability distributions. We have looked at properties of these: finding the expected value and the variance of these probability distributions. What we are going to do now is to consider some particular probability distributions which are fairly often encountered. We will look at two discrete distributions, namely the Binomial and the Poisson distributions and then once we have done that we will look at one continuous distribution, called the normal distribution.

*059
+051 There are several discrete distributions which are based on what is known as the Bernoulli process, so I think to start the lecture we will consider what we mean by this Bernoulli process.

Jacques Bernoulli was a Swiss mathematician who was a pioneer in the study of probability in the late sixteen hundreds. A Bernoulli process is what we call an event which considers two mutually exclusive and exhaustive events. What does that all mean? Well, let me try and explain this: take, for example, when we spin a coin once. This would be a mutually exclusive event because of the fact that you will either get a head to the exclusion of getting a tail; or if you obtain a tail it will exclude the possibility of obtaining a head. Other examples of this are answers to questions that are either yes or no; they would be mutually exclusive answers; an object in a factory on a production line could be defective or not defective, the sex of a person could be either male or female. These are all examples of mutually exclusive events. By exhaustive, we mean that the two events completely exhaust the sample space or form the whole of the sample space.

*095
+088 If we were to draw up a sample space for a Bernoulli process we would get the following: starting at some point in space we would move off to two sample points, the one being the outcome of one event which

we would consider to be a successful outcome, and the other one we would consider to be an unsuccessful outcome or failure. The probability of success we usually indicate by the letter p and the probability of failure occuring we usually indicate by the letter q. Now because these are two mutually exclusive and exhaustive events (they are the only two events in the sample space) p+q must add up to one.

*114 +107 The Bernoulli random variable X takes on two values, namely the values of nought and one, and by convention, one corresponds to success and nought corresponds to failure. Therefore if we had to write down the Bernoulli probability distribution we would obtain the following:

X:	0	1
f(x):	q	p

*134 +127 Some other points which I might just mention about the Bernoulli probability distribution are that it is sometimes also called the Bernoulli probability model or the two-point probability model. You can see why we refer to it as a two-point probability model.

Let's go on to the first question of this lecture. This will be <u>question one</u>. Say we are considering a Bernoulli process. Say the probability of success p is equal to 0,8. The question here is:

*152 +145 What is the value of q?

*165 +157 What I want you to do now is to have another look at the Bernoulli probability distribution. I will put up the distribution on the board:

X:	0	1
f(x):	1-p	p

Instead of writing q, I am going to write 1-p. The reason for that of course is that p plus q must be equal to one. So q must be equal to one minus p. The probability of success (or the random variable X having a value of one) will just be p.

Bearing this type of probability distribution in mind (with these probabilities) let's go on to the next questions which will be <u>question</u>

number two. Question two reads as follows: we have a Bernoulli process once again and we want you to find the expected value of X. What is E(X) or what is its mean going to be?

Just to remind you E(X) = ΣX.f(x) i.e. the sum of all the X values times their corresponding probabilities of occurrence. To answer this question, apply that formula and look at the probability distribution
+194 *which I have just put up on the board.*

*215
+211 Let's go on new and consider, not the mean or expected value, but the variance for a Bernoulli process. The variance we write as

$$\begin{aligned} &\sigma^2 \\ &= E(X-\mu)^2 \\ &= E(X^2) - [E(X)]^2 . \end{aligned}$$

*230
+225 Now looking at that second term in the square brackets: we have already determined E(X) (or you should have from question two which you have just done), but how do we determine $E(X^2)$? In order to this, lets go back to the blackboard and I will explain it to you. First of all we should put down the probability distribution involved in this case:

X^2:	0	1
f(x):	1-p	p .

Therefore

$$\begin{aligned} E(X\ ) &= \Sigma X^2.f(x) \\ &= 0.(1-p) + 1.p \\ &= p. \end{aligned}$$

Now that we have determined that lets go back to the equation that we had, namely:

$$E(X^2) - [E(X)]^2$$

$$= p - [p]^2$$

$$= p(1-p)$$

*272 $= p.q.$

+266

At this stage, let me just summarize very briefly what we have done. We have considered a Bernoulli process with its two possible outcomes (or two mutually exclusive and exhaustive events). We have found the mean or expected value is equal to p, the probability of success, and the variance, σ^2, has come out to p.q, the probability of success multiplies by the probability of failure.

*275

+275 Let us now go on to one more questions. This will be question three.

In a particular clothing manufacturing company seventy-five percent of the employees are women and twenty-five percent of these employees are men. X is a Bernoulli random variable that takes on a value of one when an employee randomly selected is a woman and a value of nought when a man is selected.

*288 *Remember that p in this case (the probability of success) we will*

+288 *consider to be 0,75, because X being equal to one is our successful outcome. X being equal to nought (that's our unsuccessful outcome) will have a probability of failure q of 0,25.*

The question that I want to ask you is: What is the expected value

*291

+297 of X in this case (μ)?

Remember that you should use some of the facts we have just developed for a Bernoulli Process.

*303
+309 Now we are going to ask you another question on these facts we have just developed on the Bernoulli process. This will be question four. Consider the same situation as in question three. What is the variance in this case (σ^2)?

Don't plunge yourself into a long calculation. Use some of the facts about the variance for a Bernoulli process which have just been developed.

*325
+332 We are going to move on now from the Bernoulli process, which is the basis of the so-called Binomial distribution which is the discrete distribution we want to have a look at in this lecture.

If a Bernoulli process is repeated n times where each of these processes or each of these trials has the same probability of a successful outcome and therefore of course, the same probability of an unsuccessful outcome, we have what we call the Binomial probability function. In other words a Binomial distribution is a repeated number of Bernoulli processes. Let us take a very simple example to illustrate this.

*339
+345 Say we spin a coin, not once, but twice. The sample space will be as follows: There won't be two possible outcomes (as was the case with the Bernoulli process) but there will be four.

HH HT or TH TT

We group the two middle outcomes together because it is the same combination although they are different permutations. You will still end up with one head and one tail. So we say there are three different outcomes and not four. The ratio of the outcomes will be: 1 2 1. Now if you would like to find the probability of each of those three outcomes occuring you would find it a very easy thing to do. If we are going to consider it in terms of p's and q's (this coin might not be unbiased and the probability of success might not be one-half) then the probability of two heads would be p multiplied by p, in other words, p squared. For completeness lets write it as p^2q^0–two successes, no failures. Then what about the middle two events which we have grouped together as one? We have one failure and one success: p^1q^1. But there are two possibilities that will give you this outcome, so we multiply by two: $2p^1q^1$. The final process or outcome will be no successes and two failures so the term there will be p^0q^2. Those three terms will then generate the probabilities of our three outcomes:

$$(p^2q^0 \qquad 2p^1q^1 \qquad p^0q^2) \ .$$

*371 +376 However, those three terms are simply the terms we obtain when we expand the binomial term: $(p+q)^2$.

If you look at that, a Bernoulli process was $(p+q)^1$. We have now repeated this twice $(p+q)^2$. So that actually generates three terms which give us the probabilities of the three possible outcomes. Hopefully you can see my meaning here: a Binomial distribution is simply a repeated Bernoulli process.

If we had to have three trials or three spins of the coin the Binomial distribution function would be: $(p+q)^3$ because there are three Bernoulli processes involved. The probabilities of the possible outcomes would simply be the terms we obtain when we expand that Binomial function:

$$^3C_0p^0q^3 + {}^3C_1p^1q^2 + {}^3C_2p^2q^1 + {}^3C_3p^3q^0 \ .$$

Notice that there are four outcomes here, although there would be a total of six because it is three Bernoulli trials. But two of those would be grouped together because they would be the same combination.

*401 +406 Now let us write down the general case when we have n trials and not just two or three trials (say n spins of the coin). In this case the probabilities of each of these values of our random variable X occuring we can write in functional notation $f(x)$. In other words X can have values of nought or one or two or three up to n (if we have n trials we could have n successes. That is also written as: $P(X=x)$ and the general term for the probability of success (for whatever number of successes we are considering) would be equal to $^nC_xp^xq^{n-x}$, and x could go from nought, to one, to two up to n. We just substitute for n in that expression (we know what n is as we have got the number of trials we are considering) and we will generate all the probabilities for all the different values of X, from that general term or expression.

*422
+426 The value of n and that of p, the probability of success, specify a particular distribution amongst all possible Binomial distributions. Therefore, if X is a Binomial random variable, n and p will be its parameters. Often, because of this fact, we write the Binomial distribution in a more concise way as follows:

$$B(n;p) \; .$$

Some texts write it in a different way:

$$b(x;\; n;\; p) \; .$$

X is our random variable. It is also sometimes called the variate.

As we have moved on from the Bernoulli process to the Binomial distribution let us now ask a few questions. The next question will be <u>question five</u>.

*442
+447 Say you are given a Binomial distribution. It is written as B(15; 0,7). That now specifies a particular Binomial distribution. The question is: What is q equal to (the probability of failure)?

*454
+460 Let's move on from question five and consider a particular example which involves a Binomial process or distribution. Say we are considering a certain university where eighty percent of the students are men. We take, from the student population, a random sample of twenty students. We define our random variable X to be equal to the number of men that we obtain in our sample. So X could vary from nought to twenty. We would like to find the probability that the sample that we have taken actually contains fifteen women.

*469
+473 Our random variable X is equal to the number of men that we obtain - not the number of women. Fifteen women means that we will have five men in our sample of twenty students. Therefore that is the same

as saying: What is the probability of our random variable having the value of five?

$$P(15\text{ women})$$
$$= P(X=5)\ .$$

This is the case where n is twenty; the probability of success p is 0,8 and X is five. We could find this by using the general binomial term of ${}^{20}C_5p^5q^{15}$. Some of you might think that you could look this up in your tables. It is possible, but the probability of finding fifteen women in your sample, when there are eight percent men at the university, is very small indeed; to four figures you will actually obtain an answer of nought in your tables. The answer here comes to, correct to eight decimal places: 0,00000017. It is a very small probability indeed.

We will spend a little time with the same problem and ask you a few questions on it. Let's now move onto question six which is based on the same information.

*491
+494 I want you to determine the probability of there being eighteen men in this sample.

Before you attempt to answer question six let me give you some assistance on the board. You are asked to find:

$$P(18\text{ men})\ .$$

Our random variable X was equal to the number of men in that sample so this could be written as:

$$P(X=18)\ .$$

We could look this value up in our tables; n is twenty so you turn to your tables where n=20 and we look up where X=18, with a probability of success of 0,8. If you have managed to turn to the relevant page in your tables, you'll find that p only goes up to 0,5. We must consider then, not the probability of eighteen successes but two failures. I could rewrite this as being equal to:

$$P(\bar{X}=2)\quad \text{where } p = 0{,}2.$$

because we are now considering our failures to be our successful events. Now it is quite a simple matter for you to look this value up and put
+518 *down the answer.*

You should find, if you use your tables, that your answer will only be correct to four decimal places.

*504
+530 We will now go onto another question on the same situation. I will do this one for you and then ask you a question a little later on the same set-up.

In this case we want to find, not an exact number, but the probability of finding at most, five women in our sample. At most means five women, or four, or three, or two, or one, or even no women in the sample. Our random variable X applies to men. At most five women means at least fifteen men or perhaps more; sixteen and so on. We would change this then to be

$$P(X=15,16,17,\ldots 20).$$

We can look this up in our tables. Remember that p=0,8, so we will have to make some change here. I'm not going to do it for you in this particular case. You could check this later. You add up all the relevant probabilities concerned and the final answer you should come to is 0,8042 - something that you can check after the lecture.
*524
+548 Now let's go onto a similar type of question which you will now answer. This is <u>question seven</u>. In this case we want to find the probability of obtaining at least ten women; at least ten women meaning ten or more.

Before you launch into this answer let me just give you a little aid on the board.

We want to know the probability of obtaining at least ten women in the sample. First of all, convert to men, because our random variable X refers to the number of men:

$$P(\text{at least } 10w)$$
$$= P(X=10,9,8,\ldots 0)\ .$$

We could look this up in our tables if our tables had values of p going up as far as 0,8; which they don't do. We have to change to the non-event of X, $\bar{X}$.

$$P(\bar{X} = 10,11,12....,20).$$

But now our probability of success or failure (however you want to look at it) will change from 0,8 to 0,2. Some of you might feel that you could have written down this line immediately from the first line, but it is always advisable to put down the values corresponding to your random variable X, and then see if it is necessary to change in order to look up these values in your tables. In this case it was, so we changed to $\bar{X}$ *and we are considering p=0,2 now and not 0,8. You have to look up all those values in your tables, add them together, and once you have done that you will come to the final answer for*
+583 *this question.*

*541
+596 Now we will go onto another two questions on the same situation. The first one I will do with you.

Say we would like to find the probability of obtaining between twelve and sixteen men inclusive. We could write that down as:

$$P(12 \le X \le 16)$$
$$= P(X=12 \text{ or } 13... \text{ or } 16) .$$

However, we have the problem here that we cannot actually look up those values in the tables because p=0,8. We change to the non-event of X:

$$P(\bar{X}=8 \text{ or } 7 \text{ or}... \text{ or } 4) .$$

We look these values up where p=0,2. The final answer comes to (once again you can check on this): 0,5786.

*560
+613 Now let's do a similar question which I want you to answer. This will be <u>question eight</u>. What is the probability of obtaining between thirteen and fifteen men inclusive in the sample. Writing that in

terms of random variable X:

$$P(13 \leq X \leq 15).$$

Once again I'll try and help you. Perhaps its not necessary as I think most of you would be quite happy to answer this. Our random variable X goes from thirteen to fifteen inclusive which means:

$$P(X=13,14,15)\,p=0,8.$$

Once again p does not appear in your tables so we have to change to:

$$P(\bar{X}=7,6,5)\,p=0,2.$$

You hav. to add those three values together so I'll give you a minute
+630 *to answer the question.*

*577
+642 I'm sure that most of you have found in this lecture that you often have to add these values together in your Binomial tables. Therefore it is often useful to have what is called cumulative probabilities. In other words, the values are added from X equal to nought, one, two etc. in a cumulative fashion until finally, when you get to your last value of X, your cumulative total must be one. You do often find cumulative Binomial tables but the ones we use are not cumulative. When you have a situation where you've got X going from nought to six say, you have to add up all those seven values.

In the next lecture we are going to consider some more properties
*586
+650 of the Binomial distribution.

*586-620
+655-686 ANSWERS TO QUESTIONS

*14
+13 $\underline{M_{39}}$: <u>Properties of the Binomial Distributions</u>

*31
+29 In this lecture we are going to consider some of the properties of the Binomial distribution. The Binomial distribution, in common with most other distributions has three significant attributes. These three attributes are the arithmetic mean, the variance and the shape or pattern of the distribution.

At the beginning of the previous lecture we considered the Bernoulli process or Bernoulli distribution. We considered two of these attributes for the Bernoulli process, namely its mean and variance. We found the mean was equal to p, the probability of success and the variance was equal to p multiplied by q, the probability of success multiplied by the probability of failure.

*43
+53 In this lecture we are going to look at these three attributes for the Binomial distribution. We will start off with its mean or expected value. To start things off we will go straight to a question. This is <u>question one</u>; reviewing some of the work from the previous lecture.

Say we are considering a Bernoulli process. In a Bernoulli process remember, we are considering the case where n is equal to one (there is only one trial) and because we are considering mutually exclusive and exhaustive events, p plus q must be equal to one. The question is: what is the mean or expected value for a Bernoulli process

*61
+71 equal to (μ)?

*71
+81 The reason for asking you this very straightforward question is that we are going to use this fact (the mean of the Bernoulli process) when we derive the mean for a Binomial distribution. We spoke of the Binomial distribution in the previous lecture as a Bernoulli process repeated n times. Bearing this in mind, let us now write down the definition for the mean or expected value for a probability distribution:

$$
\begin{aligned}
\mu &= E(X) \\
&= \Sigma X f(x) \\
&= \Sigma[x_1 f(x_1) + x_2 f(x_2) + \ldots + x_n f(x_n)] \\
&= \Sigma x_1 f(x_1) + \Sigma x_2 f(x_2) + \ldots + \Sigma x_n f(x_n) \;.
\end{aligned}
$$

Let us now look at the first term on that line $\Sigma x_1 f(x_1)$. That is the definition of the expected value of x_1. If you look at your top two lines $E(X)=\Sigma X.f(x)$ and we have $\Sigma x_1.f(x_1)$. So we could write as:

$$E(x_1) + E(x_2)+\ldots+E(x_n) \; .$$

*142 +153 Consider the first term again rather carefully. We are thinking of a Binomial distribution as repeated Bernoulli processes (or Bernoulli trials repeated n times). Therefore the expected value of x_1 is the expected value of the first trial. The expected value of x_2 is the expected value of the second trial, and so on. We know that the expected value of X or the mean for a Bernoulli process is equal to p, the probability of success. The expected value of the first trial is equal to p; $E(x_1) = p$. The same argument would hold for the expected value of x_2, which would also be p. Therefore we would have

$$\begin{aligned} & p + p + \ldots\ldots + p \\ = \; & n.p \; . \end{aligned}$$

*168 +198 We have show here that the mean for a Binomial distribution is n multiplied by p, whereas for a Bernoulli process it was simply p. Let us show you that this does work by taking a very simple example of spinning a coin twice.

If you want to find the expected value of X, where X say, is equal to the number of heads we obtain, the first thing to do is to write down the probability distribution:

X:	0	1	2
f(x):	q^2	2pq	p^2 .

With this probability distribution let us now find the expected value of X via our normal formula: $\mu=E(X)$

$$\begin{aligned} &= 0.q^2 + 1.2pq + 2p^2 \\ &= 2pq + 2p^2 \\ &= 2p(q+p) \end{aligned}$$

*207 +217 Consider the bracket (q+p).(q+p) must add up to one. So the final answer will be. 2p. We have shown that when you spin a coin twice, the mean is 2 times p - as you would expect. We have just derived a formula for a Binomial distribution where the mean is n times p. Well, n in this case is two if we spin a coin twice, so the answer then is as we expected: two multiplied by p.

One other point: in the previous lecture we found the actual mean for spinning a coin twice where the coin was unbiased, so the probability of a head was one-half. Well, if that was the case here, p would be one-half and the mean μ would come out to one; the result that we did obtain in the previous lecture.

*227 +236 We have now determined a formula for the mean of a Binomial distribution. Let us now go on to its variance. First of all let's ask you another question. This is <u>question two</u>.

Once again we go back to the Bernoulli process. We are considering one trial so n is equal to one. Its a mutually exclusive and exhaustive event so p plus q is equal to one. In a previous lecture we derived an expression for the variance of a Bernoulli process, and that is the
*241 +249 question here. What is the variance (σ^2)?

Once again, the reason for asking this question is that we are going to use the variance of a Bernoulli process when finding the variance of a Binomial distribution. Let us now consider the variance of a Binomial distribution.

*256 +264 Perhaps you would like to have an intuitive guess at what the variance of the Binomial distribution is going to be. Remember that the mean for a Bernoulli process was p and the mean for a Binomial distribution was n times p, because we are considering a Binomial process to be n repeated Bernoulli processes. From your answer for the variance of a Bernoulli process, you could perhaps guess what the variance of a Binomial distribution is going to be. Let us not guess it but derive it as we did for the mean of a Binomial distribution.

The variance (σ^2) we can write as:

$$\begin{aligned}
\sigma^2 &= E(X-\mu)^2 \\
&= \Sigma(X-\mu)^2.f(x) \\
&= \Sigma[(x_1-\mu)^2.f(x_1)+\ldots+(x_n-\mu)^2.f(x_n)] \\
&= \Sigma(x_1-\mu)^2.f(x_1)+\ldots+\Sigma(x_n-\mu)^2.f(x_n) \;.
\end{aligned}$$

*305
+312 Consider carefully the first term on that line with the top two lines on the screen. The definition of the variance is $\Sigma(X-\mu)^2.f(x)$. If you now look at that term on the bottom line, that is the exact definition: $\Sigma(x_1-\mu)^2.f(x_1)$. That would be the variance of x_1. The next term would be the variance of x_2, etc. Below that we write:

$$\sigma^2_{x_1} + \sigma^2_{x_2} + \ldots + \sigma^2_{x_n} \quad .$$

What is the variance of each one of these single Bernoulli trials. This is what we are considering the Binomial distribution to be: n of these Bernoulli trials. If you answered the second question correctly the variance for a Bernoulli trial is p multiplied by q. For the first trial the variance would be p times q. For the second trial the variance would be p times q, and for the third, and so on. We would have a whole string of p times q terms:

$$\begin{aligned} &p.q+\ldots\ldots+p.q \\ &= n.p.q \quad . \end{aligned}$$

The standard deviation will be the square root of that quantity.
*335
+343 Again let us verify this result for the situation where we spin a coin twice. The variance is equal to:

$$\begin{aligned} &E(X-\mu)^2 \\ &= E(X^2) - [E(X)]^2 \\ &= 0^2.q^2 + 1^2.2pq + 2^2.p^2 - 4p^2 \\ &= 2pq + 4p^2 - 4p^2 \\ &= 2pq. \end{aligned}$$

*366
+373 We have shown that the variance, when we spin a coin twice is 2pq, as we would expect. We have just shown a formula for the variance of the Binomial distribution is npq In this case n was two and our an-
*371
+378 swer came out to 2pq; so we have verified it to some extent.

Before we continue, let me summerize on the board, what we have shown so far for a Binomial distribution. For a Binomial distribution (B.D.), we have shown that the mean or expected value $\mu=n.p$ n is the number of trials; p your probability of success for each trial. The variance we have shown to be equal to:

$$\sigma^2 = n.p.q \ ,$$

where q is the probability of failure. For the Bernoulli process it was just p and p.q; now it is np and npq. The standard deviation $\sigma = \sqrt{n.p.q}$. *That is a summary of the properties for the mean and the variance of the Binomial distribution.*

*372 +398 Let us now proceed to another question. This will be <u>question three</u>. Say we have a random variable X, where X is equal to the number of heads involved when we spin a coin. It is a fair coin so the probability of obtaining a head or the probability of success is one-half or 0,5. Say we spin the coin fifty times, so n=50 in this case; there are fifty Bernoulli trials. The question here is: what is the expected value of X for this situation?

*384 +409 *I suggest you use some of the new formulae which we have developed.*

Let's go directly to another question; <u>question four</u>. What is the variance of X going to be for the same situation as in question three? Make use of the formulae we have just developed for a Binomial
*401 +426 distribution.

We have examined two attributes of the Binomial distribution, namely the mean and the variance. We have derived two very useful formulae for finding the mean and the variance for a Binomial distribution. We are now going to consider the third attribute which is the shape of the Binomial distribution.

The shape of the Binomial distribution depends on n and p, the parameters of the Binomial distribution. We are going to look at two different distributions and change one or the other of these two parameters.

*420 +444 First of all, we are going to fix the parameter p at 0,5. We are going to have the other parameter n being equal to five, then 15, then 30 and then eventually increase n up to infinity. I am going to draw histograms for these distributions rather than bar diagrams to give you a better idea of the outline or shape of the distribution.

Secondly, we are going to change our parameter p from 0,5 to 0,1 and again vary n from five to fifteen to thirty and then to infinity. Hopefully you will get an idea of how the changes in the two parameters n and p affect the shape of the Binomial distribution.

*434 +459 The very first graph will look like this: p is 0,5 so the graph is entirely symmetrical about a mean value. X can have values of nought, one, two, three, four or five so I should have drawn straight lines directly above those numbers, but I have drawn a rectangle about that number. I have used up all the space about the numbers nought to five so that I have a histogram effect. I think the outline of the histogram gives you a better idea of the shape than a bar chart. On the Y-axis we have the probabilities of X occuring. We are now going to increase n from five up to fifteen and notice how the shape will change.

It is very similar. There are more possible values of X. Notice once again that the graph is entirely symmetrical about a mean value. The scale has changed on the Y-axis. Nevertheless, it is a similar type of situation. Now we increase n from fifteen to thirty. Let's have a look at the shape we get.

*452 +477 Again my axes have to change their scales; there are more possible values of X. Notice that the outline of the histogram is getting very much like a bell-shaped curve. Or am I begging the question? What happens when we increase n from thirty to infinity?

The possible values of X will be so close together that they will be indistinguishable. I can't draw all those lines so I have simply drawn the outline. They will be so close together that I will have a smooth curve as n tends to or gets close to infinity. That is the type of shape I will end up with.

Now we are going to change our parameter p from 0,5 to 0,1 and go through the same process. The first graph: we have p fixed at 0,1 and n as small as five. That graph will look like this. Notice it is skewed to the right. Some of you might disagree and say it is skewed to the left- remember when we say it is skewed to the right we take the point of view of the histogram. If you were standing there, facing us, it would be your right-hand side which has moved off to its right and you have a tail on the left. As you look at it, it is the other way around. We say it is skewed to the right, so the mean is somewhere to the right as well, in this case. It is certainly not symmetrical as was the case when p was 0,5. Let's have a look at this asymmetrical or skewed graph or distribution as we increase n ; we now go from five up to fifteen.

*477
+500 There is our change. It is still skewed to the right. However, not as badly skewed as in the first case. Now we increase n from fifteen up to thirty. Note what happens.

It is still skewed to the right, but even less; n is getting so large that we are getting an equal distribution about the mean or middle value. You can see more or less where you would expect the mean to be. What happens as we increase n from thirty up to infinity?

We have so many possible values that the skewness disappears completely and we get a smooth curve as you see now. However the mean would be more to the right than in the case where p was equal to 0,5.

Hopefully, by looking at those graphs, you now have an idea of how the parameters of the binomial distribution, n and p, affect the shape. What I'd like to do now is to demonstrate to you that this is not just theoretical reasoning. I have a practical demonstration to show that these are the shapes of the Binomial distribution that we will get.

*495
+518 I have here, on the desk, what is known as a 'hexstat' or a probability demonstrater. Let me briefly describe how it works. You have a whole lot of beads up here and they fall down these chutes. In each case they have a probability of 0,5 of going to one side or the other. This pattern continues as they meet another of these points. They can go to the right or to the left. It is called a hexstat because all these little shapes are hexagonal. We get here a ratio of one to one. On the next line we get an increased probability of it coming into this chute and the ratio there is one to two to one; the binomial coefficients when n is two. Then it is one to three to three to one. This generates as we go down, the steps in the Binomial distribution as n goes from one up to, I think, sixteen. Therefore in these chutes down at the bottom, we should obtain a type of distribution corresponding to the Binomial distribution which you have just had a look at. Will

this happen? Let me get all the beads back to the top and let them run down. I will keep the base on the desk so that the probability of going from either to the right or left is 0,5. There should not be any skewness involved.

I have got them all into the top chute and I will place the base on the desk squarely and let them run down. The base was flat on the desk so I should get an even distribution. That is exactly what I do get. Hopefully, this is demonstrating to you that we get a situation as we had before. n is not infinity, but a finite number here. If we had to increase it, we would get even more symmetry.

*524
+546 Let's take a case where we don't have p equal to 0,5. I don't keep the base level on the desk but I tip it; I skew the graph. Let's see what happens in that case. I obtain this by doing the same thing, but holding this at an angle as the beads fall down. You should find that we get skewness in this case to the right. The peak is more to the right than to the middle as it was before.

Let us now go on to another question. This is <u>question five</u>. Consider a graph where n is thirty. Note carefully what the distribution looks like. The question is: what is the value of p in this si-
*541
+562 tuation for this Binomial distribution?

Another question: <u>question six</u>. This refers to another graph of a Binomial distribution. That is the graph I want to refer to for question six. Once again, n is equal to thirty. The question I want to ask: is p equal to q for this situation?

*561
+582 We will now go onto some other examples where we work with expected values and the formulae we have developed for the mean and variance of a Binomial distribution.

Say we are considering a Binomial distribution where X is the binomial random variable and is distributed as follows:

$$b(x;\ 10;\ 0{,}6) \quad .$$

That means we have a Binomial distribution where n is 10 and p is 0,6. Another way of writing that is:

$$X \text{ is } B(10;\ 0{,}6)\ .$$

What would the expected value of (X-6) be for this distribution? We have a formula for the expected value of X (the mean) of a Binomial distribution, but not for the expected value of (X-6). We can answer this by operating with expected values:

$$\begin{aligned} &E(X-6) \\ &= E(X) - E(6)\ . \end{aligned}$$

E(X) for a Binomial distribution is n multiplied by p and E(6) is simply six. Below that we will have:

$$\begin{aligned} &n.p-6 \\ &= 10(0{,}6) - 6 \\ &= 0\ . \end{aligned}$$

The expected value of (X-6) in this case is nought.

*584 +604 A similar question for you to try: <u>question seven</u>. We are considering a Binomial distribution:

$$b(x;\ 10;\ 0{,}6)\ .$$

What is the expected value of (X+3) going to be [E(X+3)]?

*599 +618 Let us now try another of these questions. Say we are again considering a Binomial distribution where X is a binomial random variable; n is ten and p is 0,6. What is the expected value of $(\frac{X}{10})$. We treat the division by ten as X multiplied by one-tenth. When we are multiplying by a constant we can remove the constant to outside the expected value. We could rewrite that as:

$$E(\frac{X}{10})$$

$$= \frac{1}{10}.E(X) \quad .$$

E(X) for a Binomial distribution, we now know, is equal to n times p; so that will be:

$$\frac{1}{10}.n.p$$

$$= 0,6 \quad .$$

*613
+632 Once again, we will do a very similar problem, but I will leave it up to you to get the answer. This is question eight.

We are considering a random variable X which is binomially distributed, where n is ten and p is 0,6. What is the expected value of (3X+2) in this situation?

*629
+649 Let us go on from these theoretical questions to a more practical type of problem. The problem we are going to consider is the following:

Say that sixty percent of the residents of Johannesburg favour cinemas opening on Sundays. Let X be equal to a binomial random variable equal to the number of people who favour cinemas opening on a Sunday. If a random sample of twenty residents is taken in Johannesburg, lets find the probability of certain outcomes.

The sample size n is equal to twenty. We are considering a Binomial distribution because the answer is yes, they are in favour or no, they are not. The probability of success I will leave for you to find out from the sixty percent who favour Sunday cinemas. We now go onto question nine:

What is the expected value of X (the mean) for this sample of twenty?

*651
+671 This answer is going to tell you how many people you expect, out of that sample of twnety, to say that they do favour cinemas opening on a Sunday. What sort of variance is going to be involved about this mean value?

This is <u>question ten</u>; based on the same situation as in question nine. What is the variance of X?

*656
+675 *Use some of the formulae we have developed in this lecture - there is no need for long, involved calculations.*

We have considered in this lecture three main attributes of the Binomial distribution, namely its mean (which we found was equal to n times p), secondly its variance (which is equal to n times p times q) and thirdly we had a look at how the parameters of the Binomial distri-
*671
+690 bution, namely n and p, actually change or affect the shape.

*674-718
+693-737 <u>ANSWERS TO QUESTIONS</u>

*12 +12 $\underline{M_{40}}$: The Poisson Probability Distribution

*31 +29 In this lecture we are going to examine another discrete probability distribution, namely the Poisson probability distribution. In order to introduce this distribution, let me first refer back to the previous discrete distribution which we looked at, namely the Binomial probability distribution.

The first point that I am going to make is one that you might have realized already. When we are considering calculating the probabilities of certain events occuring in a Binomial distribution, it is very easy to calculate these probabilities when n is small. If n is large it is usually difficult or it is quite an involved arithmetic expression to find the probability of a certain event occuring. We are not considering using tables, as they are only useful when we have them tabulated for specific values of n. We can't have tables for all values of n, especially when n gets rather large. I think your tables go up only as far as n equal to twenty.

*59 +58 I think we will start off by emphasizing this point because some of you might not have realized this. I'm going to ask you two questions in connection with this fact.

This will be question one. Say we are considering a Binomial distribution and say p is equal to 0,5; n is equal to three. We would like to find the probability of two successes. We would then use the term:

$$^3C_2.p^2.q^1 .$$

*80 +79 The question is: calculate the value of that expression without using your tables.

Before you do this, let me give you a little revision on the board. Some of you might have forgotten how to calculate combination terms. We are considering a Binomial distribution and we are trying to find the probability that X is equal to a particular value. A general term for this was:

$$P(X=x)$$
$$= {}^{n}C_{x}p^{x}q^{n-x} \quad .$$

How do we calculate ${}^{n}C_{x}$? Well, ${}^{n}Cx$ is defined as:

$${}^{n}C_{x} = \frac{n!}{x!(n-x)!}$$

The term up on the screen was ${}^{3}C_{2}$; hopefully you can calculate that now from what I've just given you. You know p is (0,5) and you should know what q is. Most of you should be able to calculate the value of that term quite simply now.

*115 +146 To make my point about this expression for a Binomial distribution being rather tedius when n is large, I'm going to increase n to a very large number indeed. You should be able to see from this that the calculation involved is going to be far more tedius and strenuous than what you have just done for question one. This will be <u>question two</u>.

Say we are considering a Binomial probability distribution where n is equal to one hundred - rather a large number. Your tables don't give values of the Binomial distribution for as high a number as that. p, the probability of success is 0,1. Say we wanted to find the probability of X being equal to thirty-two. You should all realise that the arithmetic involved in this question would be very tedius indeed. What I'm going to ask is: would this be the expression that you would use in order to calculate the probability of X bing equal to

*149 +180 thirty-two: $\frac{100!}{32!68!} \cdot (0,1)^{32}(0,9)^{68}$? Simply answer yes or no.

*162
+192 From these first two questions you should realize that it does become far more difficult to calculate the probability of an event occuring in the Binomial distribution. You might now well ask: isn't there perhaps another formula we can use when n does get very, very large. Let's take this argument a step further and ask ourselves what happens when n gets extremely large; in fact when n tends to infinity. Will it be at all possible to calculate the probability of a certain number of successes in an infinite number of events taking place? The values we would have to calculate would be calculated from the general
*179 term of the Binomial expansion with n tending to infinity, and it would
+209 look like this:

$$^{\infty}C_{x}p^{x}q^{\infty-x} .$$

How on earth do we evaluate that expression when we are dealing with infinity? I think most of you will realize it is impossible to evaluate that expression when n tends to infinity. You might well ask: will this ever occur in the real world, where we are considering an infinite number of events in some situation?

The answer to this is yes, this situation does arise in the real world fairly often. For example, say we are considering the number of defects in a certain length of wire. There could be an infinite number of defects in a length of electric wire, so n in fact tends to infinity. But usually there are only a very small number, so the probability of success p is very small. Or, perhaps the number of defects in bolts or nails that you might be making. You might produce millions of these per year so n again would get very, very large. The breakdowns of some machinery per day: n could be large - you could have a breakdown every second or every tenth of a second. Or, in the same way, the number of telephone calls you might expect at a certain exchange. There is usually only so many per minute, but the possibility of there being, say a million, in a particular second, does exist.

*211
+23[illegible] All these types of processes are characterised by the number of successes per unit of time or sometimes per unit of space, rather than simply in n trials as was the case in the Binomial distribution. We are actually moving off from the Binomial distribution to another type of discrete distribution. This is the distribution that exists when n tends towards infinity and p gets very, very small; in fact it tends

towards nought. Then the probability distribution that the random variable X is going to follow is known as the Poisson probability distribution or the Poisson distribution.

This is named after Poisson himself who was a Swiss mathematician. He first described this distribution in eighteen thirty seven. Let us now state the distribution mathematically. The Poisson distribution applies to the case where n tends to infinity and p tends to nought. The probability of success gets so small that we expect very few successes in an infinite number of possible events or outcomes or rather in a unit of time or space. Then the probability of our random variable X being equal to a particular value of X is written as:

$$P(X=x) = \frac{\mu^{x}.x^{-\mu}}{x!} \quad .$$

The value of the constant e is an irrational number equal to 2,71828... .
*256 We usually round it off to about five figures as you see on the screen.

+281 The important thing about this distribution is that it only has one parameter, and then it is completely defined. That parameter is μ, the mean. In the Binomial distribution we had two parameters, n and p.

In actual fact one of the early pioneers in statistics who helped promote the use of the Poisson distribution did so by showing that the number of deaths from the kick of a horse per year per army cavalary corps in the Prussian army during the period 1875-1894 followed a Poisson distribution with high precision. This is the sort of situation we are considering. In that period of time, almost an infinite number of troops, or a very large number of troops could have been kicked to death by their horses. But the probability of this occuring was very, very small indeed; n tended towards infinity, p tended towards zero and the actual distribution of the number of men dying from the kick of
*276 a horse followed a Poisson distribution with high precision.

+300 Let's go onto the next question. This is <u>question three</u>. Say we are considering a Poisson probability distribution and we would like to find the probability of X being equal to a particular value, x. The formula we would use to find this would be the following:

$$\frac{(?)^{x}.e^{-(?)}}{x!} \quad .$$

Put down the name of the quantity which should be in those brackets where we have placed a question mark. A very easy question to find out if you have taken note of what the formula is for the Poisson distribution.

*295
+319 The Poisson distribution is still however a discrete probability distribution, as is the Binomial distribution. Another way of thinking of the Poisson distribution is as the limiting case of the Binomial probability distribution. (i.e. the limit as n tends to infinity and p tends to nought). Say we are considering a Binomial distribution: the general expression would be

$$^{n}C_{x}p^{x}.q^{n-x} .$$

It can be shown mathematically that if we start with that formula, and we let n tend to infinity and p tend to nought, then we will come out with an expression:

$$\frac{\mu^{x}.e^{-\mu}}{x!} .$$

*310
+334 Therefore, taking the general term of the Binomial distribution and letting n tend to infinity and p tend to nought we will end up with an expression which is equal to the Poisson distribution; so it is the limiting case of the Binomial distribution.

Now let us examine some of the more important properties of the Poisson probability distribution. The three most important ones being its mean, its variance and its shape. We will start off by looking at the mean and variance of the Poisson probability distribution.

In common with the Binomial probability distribution (because it is also a discrete distribution), its mean μ is also equal to n multiplied by p, even though n tends to infinity and p tends to nought. Its variance σ^2 will also be equal to n multiplied by p multiplied by q. There is however, a special property involved between the mean and variance of a Poisson probability distribution. I will show you now what this property is.

*329
+351 Let us rewrite that line on the screen npq as being equal to n times p and instead of writing q, we can write (1-p). Let's just put that down on the board. I think it will be a little easier.

$$\sigma^2 = npq$$
$$= np(1-p) \ .$$

If p tends to nought, what is the value of this expression in the brackets going to be equal to? The answer is that if that gets so small that it is almost equal to nought, this expression will eventually be equal to one and then the variance will be equal to n times p. You might well ask, won't this also be zero if p is nought? We are talking about a limiting case and without getting too involved in the mathematics, we can show that if p tends to nought we will get an expression for σ^2 of n times p which is exactly the same as the mean μ.

The special property of the Poisson distribution is that the mean and the variance are both the same (equal to n times p). This is something that you should accept without going into the mathematics of it. This often acts as a very useful check to find out whether a distribution is a Poisson distribution or not. Let's now go onto another question somehow related to this.

*357
+377 This is <u>question four</u>. Say we are considering a discrete distribution; n is equal to one hundred and p is equal to 0,01. The question is: what is the value of $\mu-\sigma^2$?

Once again, just before you answer that let me give you a little help on the board. I've just shown you a certain property of the Poisson distribution, but we don't know what this distribution is. To check, we still use the formula: $\mu = n.p$, because it is a discrete distribution and $\sigma^2 = npq$. If those two are equal we realize that this is a Poisson distribution; n was equal to one hundred and p was equal to 0,01. Very large, very small. Will these things be equal? I'm not asking you whether they are equal or not. I want you to find the difference
+401 *between the two. What is the value of $\mu-\sigma^2$?*

*387
+421 We will come back to the significance of this answer in a moment. Firstly, let's have a look at the third property of the Poisson distribution, its shape. What does it actually look like?

We are going to consider two graphs. The first will be where the mean μ, its only parameter, is equal to two. Secondly, where its mean μ is equal to ten, and see how they differ for different values of the only parameter , μ. Remember that μ is the expected value of X or the mean of the distribution which is equal to n times p for a discrete distribution. For μ to be small when n is large, p must be very small and for μ to be equal to ten when n is large, p will then not be that small. Remember that n is tending towards infinity while p is getting rather small. Lets see how a change from μ equal to two to μ equal to ten changes the shape of the graph.

The first graph is going to look like this. I've plotted points above the discrete values of X of nought to six. Seven, eight and nine have almost zero probability of occuring. For μ to be equal to two when n is very large means that p is very small. The probability of success is very small, so we would expect, on average, nought, one, two or three successes at the most- very seldom more than that. The graph is a skew distribution; skewed to the right.

*414
+446 We now increase the parameter μ from two up to ten. Let's see how this affects the graph. The second graph will look like this. Notice it is not skewed, but almost symmetrical. The distance between the adjacent points has become a lot smaller, because my X scale has changed; n and now μ are sufficiently large so that the points on the graph appear to almost touch each other. If p gets larger, so μ gets *bigger*. If μ gets larger than ten, we would get what would almost appear as a normal distribution, because those points would be very close together indeed. That is something we might come to a little later on when we consider the normal distribution. The graph however, is symmetrical because the mean is more of a middle value. We expect more successes than we had before when we had a small value of μ.

Now we are going to consider some applications of the Poisson distributions. Before we proceed onto these applications let me point out to you how we obtain values for the Poisson distribution. We will do the first few via calculation.

Say we have a situation where μ is equal to 1,2 (a Poisson distribution) and we would like to find the probability of obtaining two successes (in other words the probability of our random variable X being equal to two). We make use of the Poisson formula:

$$\frac{\mu^x . e^{-\mu}}{x!}$$

$$= \frac{(1,2)^2 . e^{-1,2}}{2!} \; .$$

Now that is rather easy to work out on your calculator because your calculator should give you the value of e. Raise it to the power: -1,2 (raise it to 1,2 and invert); multiply by $(1,2)^2$ and divide by 2! (which is just two). If you do this on your calculator you should obtain an answer of 0,2169 - something I suggest you do check at a later stage.

*450 +484 Here is a similar question that you can try and answer. This is question five. Say we have a Poisson probability distribution and μ is equal to 1,3. What is the probability of obtaining three successes (the probability of X being equal to three)? We want to calculate this to four decimal places. I suggest you all get out your calculators which you will need. If you don't know how to find the value of e on your calculator, approximate it to four decimal places as 2,7183.

Again, let me give you a clue as to how you should go about this on the board. We are trying to find:

$$P(X=3)$$

$$= \frac{\mu^x . e^{-\mu}}{x!}$$

+503

$$= \frac{(1,3)^3 . e^{-1,3}}{3!} \; .$$

*474 +513 You should find that, that expression is not that difficult to evaluate when you use a calculator. However, it is not always necessary to use a calculator to find these probabilities, because we do have tables for these probabilities of a Poisson distribution. At the back of your text you do have these tables available for certain values of μ. I would like you to take out your texts and turn to page 300, where you will find the Poisson probability distribution tables.

You will find that the probabilities are tabulated correct to four decimal places for certain values of the only parameter μ. Let's have a look at what these tables look like. If we go in to the top left hand corner where they specify the values of μ. μ goes down the left hand column; depending on which value of X you want to determine, you go from left to right. Let's go down the left hand column.

*494 +536 You can see that μ starts at 0,1; 0,2 and continues down to 1; then it increases in tenths from 1,0 to 2,0. From 2,0 onwards it goes up in jumps of 0,2; 2,0 to 2,2 to 2,4 etc. These larger values of μ are not that common. If you go from there and look at page 301, the same thing happens to μ from 3 to 4, but from there on we go up in integer values: 4 to 5 to 6 up to finally 10. Values of μ greater than 10 are very uncommon.

Let's go back to the top left-hand corner and move from left to right so that we can indicate to you the probabilities of different values of X. Say we are considering a μ value of 0,1 and if you want to find the probability of 0 successes you look in that first column; 1 success, 2 successes and so on. Notice that when we get to 3 successes, its very small: 0,0002. If we stop at 4 successes, the probability of obtaining 4 successes when μ is 0,1 is, to four decimal places 0,0000. For 5 successes, and from there on, they haven't bothered to tabulate the values, as you see being indicated. As μ increases, the probability of getting more successes increases, and so this would continue. If you look at page 301, when μ gets to 10, they go up to 24 successes, which is tabulated in a separate section on the bottom of page 301.

*519 +560 Let's go back to the left-hand corner and check the answers we have obtained for the last two problems, to four decimal places. In the first question that I did with you, we wanted to calculate the probability of two successes when μ was 1,2. We go down the left-hand column to the value of 1,2. We now move to the right until we get to the column, X=2, which is the third column. You can see the value there: 0,2169 - eactly as we did obtain. The question I asked you to do, question five, μ was equal to 1,3. Let's go back to the left-hand column to identify the point where μ is 1,3. Move along that row until we get to X=3, so that we get three successes. That is the fourth column: 0,0998. If you answered question five correctly on your calculator, that is the answer you should have obtained. Where you have values of μ which are tabulated in your text, you needn't work it out on your calculator, just look it up in your tables. There is the problem of not having a tabulated value.

*541 +580 Let's go back to the left-hand column. Say we have a μ value of 2,5. Go down the left-hand column and identify the value 2,5. You will find there is no 2,5; there is only a 2,4 and 2,6. How do we find probabilities for μ equal to 2,5?

One way of doing this is to interpolate between the values of 2,4 and 2,6. Say we wanted to find the probability of X=0; that's the first column there. We would find the geometric mean between 0,0907 and 0,0743. Thats one way of doing it. To find the geometric mean we have to multiply those two values together and then find the square root. There is a certain amount of calculation on the calculator. I think it is easier to use the formula and work it out. It will certainly be the correct value if we use the formula.

*554 +593 Let's do this for <u>question six</u> so that you have a little more practice in using your calculator and the formula, to find these probabilities. We will make this the situation where μ is equal to 2,5. We would like to find the probability of X being equal to six. We could go to the column in our tables where X is equal to six and interpolate between those two values or we have the choice of substituting these values in the formula. We will then get:

$$\frac{(2,5)^6 . e^{-2,5}}{6!} .$$

*568 +606 Work the answer out correct to four decimal places.

*581 +618 Let us consider some applications of the Poisson probability distribution, now that you know how to look up or calculate the values of certain probabilities. Suppose that one percent of the bolts produced by a certain machine are defective. Say we take a random sample of three hundred bolts and we would now like to find the probability of a certain number of these being defective where the mean or expected value μ is equal to n times p. We have to calculate μ from this: n is 300, p is 0,01 and we get a value of 3.

The question now is, can we use the Poisson probability distribution to calculate the probability of a certain number of defectives occuring in this sample? In order to use a Poisson distribution μ and σ^2 must be the same. What is σ^2, the variance? It is equal to n times p times q so it is 300 times 0,01 times 0,99 which gives you 2,97. The mean and the variance are not the same so this is not a Poisson probability distribution but a Binomial probability distribution. However, we don't have Binomial tables for n equal to three hundred and the term that we have to calculate is going to be difficult to calculate arithmetically, as you should have realized from the first two questions in this lecture.

*601
+637 A point to note here is that, although the mean and variance are not the same, they are almost the same. There is a difference of 0,03 in 3, that is one percent. There is only a one percent difference between the mean and the variance. The question might now arise: can't we use the Poisson distribution (which will be very easy to look up in the tables because μ is equal to three) as on approximation to this situation? This is really a Binomial situation. How accurate would we be if we used the Poisson as an approximation to the Binomial? The answer is that we would find it to be a very good approximation. I want to demonstrate this point by refering to two graphs.

We are going to have a look at two graphs where there is a difference between the mean and variance of more than one percent and we will show you how close the two graphs still are - the one being a graph of the Binomial distribution, the other a graph of the Poisson distribution.

*614
+650 Let's start off by looking at the figures we are going to consider. Here we are considering a case where n is 20, p is 0,2; that gives us a mean μ of 4 and the variance will be 4 times 0,8 which is 3,2. So there is quite a significant difference between the mean and the variance. It is 0,8 in 4, which is twenty percent. A large difference, you could say, between the mean and the variance. If we use the Poisson as on approximation to the Binomial distribution, how accurate would we be, or how much accuracy would we lose? For these

figures I have plotted a histogram o the Binomial distribution and it appears as you see it on the screen now. Using the same figures I could draw a histogram over the discrete values of X and compare the two. This is what we are going to do now.

*627 +662 We will start from the left-hand side of the graph and draw the histogram corresponding to the Poisson distribution for a mean of 4 and not 3,2. For a mean of 4 the Poisson distribution will look as you see now. You will find that it is very close indeed to the histogram we already have. Only in the middle there will you lose some accuracy and coming along as we do now it is close to what the histogram for the Binomial distribution is. If we used those values of the Poisson distribution as an approximation to the Binomial we would be rather close to the true values. That is a case where there is a twenty percent difference between the two. In our example there was only a one percent difference, so if we now used the Poisson distribution as an approximation to the Binomial distribution we would lose no real accuracy at all. Certainly not really to four figures. Let's now use the Poisson distribution (although it is not really a Poisson distribution) to answer questions on this problem.

*641 +675 We had one percent defective bolts, n was three hundred, and we are taking the mean μ to be three. What would the probability be of obtaining no defectives at all? In other words, all the bolts in our sample of three hundred are good. If we use our random variable X we could say:

$$
\begin{aligned}
&P(\text{all good}) \\
&= P(X=0) \ .
\end{aligned}
$$

To find this value you could go to your Poisson tables, where μ is three and in the column X=0, if you check on that, you will find it comes to 0,0498. If some of you would like to, you can use the Binomial expansion term, and you would find you get a very similar answer indeed. You would lose no accuracy at all.

*652 +687 Now let's do another question on the same situation. This one you can answer. This is <u>question seven</u>. Considering the same situation where the mean μ is three, what is the probability of obtaining two defective bolts in this sample of three hundred? What is P(2 defectives)? I don't expect you to work this out as you can find the value in the tables.

*665
+698 Let's move onto some more questions on the same situation. Say we wanted to find the probability of obtaining two or more defective bolts in the sample. We have two or more here, so it is rather tedius to add all these probabilities together. Lets rather change that to

$$= 1 - P(0 \text{ or } 1 \text{ defective})$$
$$= 1 - P(X=0 \text{ or } 1),$$

by our probability rule of complementation. This means we have to look up the value when X is 0 and when X is one, along the row of μ equal to three in your tables. If you do that you will find that will be equal to:

$$1 - (0{,}0498+0{,}1494)$$
$$= 0{,}8008.$$

Let's go onto a similar question which you once again can do. This is question eight. We would like to find the probability of obtaining two or less defectives. In other words, at the most two de-
*679
+713 fectives.

Before you try and answer this let me give you a hand on the board. We want to find:

$$P(2 \text{ or less})$$
$$= P(X \leq 2)$$
$$= P(X = 0,1,2) \ .$$

Therefore you have to look up three values (where μ is equal to three)
+721 *and add them together in order to answer that question .*

*6[illegible]1
+733 What I want to do now is to go onto another class of problem which can often be solved with the aid of the Poisson distribution. These problems are known as queueing problems. They concern themselves with

determining how many customers, say, are likely to be waiting in line for service, and how long they must wait before being served. Without delving into queueing theory (which we don't concern ourselves with in this course), take an example where the average number of customers who appear at a bank teller's window is say, one per minute . This is a typical example of where we use the Poisson distribution because you could in fact have an infinite number of customers waiting there per minute - there is a possibility that the whole of Johannesburg could turn up at the teller's window in that particular minute. The probability, of course, of this occuring is infinitely small, but nevertheless it is the number of events or successes occuring per unit of time so we assume this follows the Poisson distribution. It has in fact been shown to do this. We use the Poisson distribution here; we are not using it as an approximation to the Binomial distribution.

*705 +745 Say we would now like to find the probability of no customers arriving, in one particular minute or in any particular minute , at the teller's window. We bear in mind that our μ value is one and we want to find the probability of our random variable X being equal to nought. If you go to your tables where μ equals one and look up the value where X is nought you will find it is 0,3679. There is a thirty six percent chance in this situation that in any particular minute during the day, no customer's will arrive at the teller's window or be queueing at the teller's window where the average is one per minute during a whole day.

Let's go onto a question that you can answer. This is question nine concerning the same situation; μ is equal to one. What is the probability of four or more customers arriving at the teller's window in any particular minute? *717 +758

*726 +766 That brings us to the end of this lecture. Thank you very much.

*726-746 +766-785 ANSWERS TO QUESTIONS

Appendix IV

Lesson Programmes and Computer Reports

The lesson programmes for the six television lectures, M_{35}-M_{40}, are given in this appendix. Their structure is fully described in section 5.3.1.

In addition to these lesson programmes, brief examples are given of a student response report, a detail report and a student report, as described in section 5.3.2.

Lesson Programme for Lecture - M_{35}

```
   @EHI
00 M35
@I
01 2\TWO
@I
02 6\SIX
@I
03 52\FIFTY TWO
@I
04 X=4\4\FOUR\X=FOUR
@I
05 0\ZERO\X=0\X=ZERO\NOUGHT
@I
06 3/8\0,375\0.375
@I
07 1/16\0,0625\0.0625
@I
08 1/8\0,125\0.125\F(3)=1/8\F(3)=0,125\F(3)=0.125
@I
09 DISCRETE\DIS
@I
10 NO
```

Lesson Programme for Lecture 2 - M_{36}

```
 @EHI
00 M36
@I
01 MU
@I
02 STANDARD DEVIATION\STD DEVIATION\S D\SD\S\SX\THE STANDARD DEVIATION
@I
03 7\SEVEN
@I
04 PROBABILITY\P
@I
06 1\ONE\E(X)=1\E(X)=ONE
@I
07 1.5\1,5\ONE AND ONE HALF\11/2\1 1/2
@I
08 27\TWENTY SEVEN\E(6X)=27\E(6X)=TWENTY SEVEN
@I
09 10\TEN
@I
10 10\TEN
```

Lesson Programme for Lecture 3 - M_{37}

```
@EHI
00 M37
@I
01 STANDARD DEVIATION\THE STANDARD DEVIATION\S D\SD\S\SX
@I
02 60/7\SIXTY/SEVEN
@I
03 VAR(X)\VAR (X)
@I
04 0,5\0.5\1/2\ONE HALF
@I
05 4\FOUR
@I
06 3\THREE\VAR(X)=3\VAR (X)=3\VAR(X)=THREE\VAR (X)=THREE
@I
07 1,5\1.5\3/2\ONE AND A HALF
@I
08 1\R1\1R\1 RAND\ONE RAND\ONE R
@I
09 -2R\- 2R\-R2\- R2\-TWO RAND\-TWO R\-TWO
@I
10 379\VAR(X)=379\VAR (X)=379
```

Lesson Programme for Lecture 4 - M_{38}

```
@EHI
00 M38
@I
01 ,2\0,2\1/5
@I
02 P
@I
03 0,75\,75\3/4
@I
04 0,1875\,1875\3/16
@I
05 0,3\,3\3\10
@I
06 0,1369\,1369
@I
07 0,0026\,0026
@I
08 0,3382\,3382
```

Lesson Programme for Lecture 5 - M_{39}

```
@EHI
00 M39
@I
01 P
@I
02 PQ\N PQ\ 1 PQ\ P*Q\N * P* Q\ 1 * P* Q
@I
03 25
@I
04 12, 5\ 121/2\ 12 1/2\ 12 AND 1/2
@I
07 9
@I
08 20
@I
09 12
@I
10 4, 8\ 4.8
```

Lesson Programme for Lecture 6 - M_{40}

```
@EHI
00 M40
@I
01 0, 375\, 375\3/8\ 0, 3750\, 3750
@I
02 YES
@I
03 NO
@I
04 0, 01\, 01\ 1/100
@I
05 0, 0998\, 0998
@I
06 0, 0278\, 0278
@I
07 0, 224\, 224\ 0, 2240\, 2240
@I
08 0, 4232\, 4232
@I
09 0, 019\, 019\ ., 0190\, 0190
```

A Student Reponse Report

A student response report displays the responses on the V.D.U. as they are being made by the students. In the report given below for example, the student named Moffat is seated in row A, seat number 2 and has responded: NO. Gilchrist, in A1, has not as yet made a response.

```
CC M35
AC LEDGER B             ---NO
A1 GILCHRIST R          ---
A2 MOFFAT A             ---NO
A3 ISMAIL M             ---
A4 LAZARUS I R          ---NO
```

A Detail Report

In the detail report given below, it can be seen that Ledger, in seat A0, has answered question 7 incorrectly while he has answered the other nine questions correctly. Ismail, in seat A3, has answered questions 1,3,5 and 10 incorrectly and questions 2,4,6,7 and 8 correctly. He has not attempted question 9. If a student answered a question correctly at a second or later attempt, a L would appear. This does not occur in the example below.

```
DETAIL REPORT

M35
DATE:
TEACHER:
CLASS:
GROUP:0, STUDENTS NO.=005
   0000000001 NAMES
   1234567890
A0 RRRRRRXRRR LEDGER B
A1 RRRRRRRRRX GILCHRIST R
A2 RRXRXRXRXR MOFFAT A
A3 XRXRXRRR-X ISMAIL M
A4 XXRXXRXXXR LAZARUS I R
```

A Student Report

The student report gives the names of the students in the extreme right-hand column and their seat locations in the extreme left-hand column. The middle four columns are headed L%, R%, A% and M%. The L% column gives the student's final mark after his last or final attempts at answering all the questions have been recorded. The R% column gives the student's final mark based on the initial answers entered by the student. Notice that these two columns display identical values as a second attempt was not permitted in the example below. The A% column gives the percentage of questions attempted by the student and the M% column reveals the student's final mark out of the number of questions attempted (i.e. the questions which were not attempted are ignored). The means and standard deviations are given for the columns at the end of the example.

STUDENT REPORT

M35
DATE:
TEACHER:
CLASS:
GROUP:0, STUDENTS NO.=005

ST	L%	R%	A%	M%	AE	NAMES
A0	90	90	100	90	0	LEDGER B
A1	90	90	100	90	0	CILCHRIST R
A2	60	60	100	60	0	MOFFAT A
A3	50	50	90	56	0	ISMAIL M
A4	30	30	100	30	0	LAZARUS I R
MEAN	64	64	98	65	0	
S.D.	23	23				

Appendix V

The Pre-Test

A copy of the pre-test, which is referred to in section 6.3.2, is given in this appendix.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG.

DEPARTMENT OF APPLIED MATHEMATICS.

MATHS AND STATS 1B

Time: 30 minutes. Twenty Questions

1. $(a-b)^2$ is equal to:
(a) a^2-b^2
(b) $(a-b)(a+b)$
(c) $a^2-2ab+b^2$
(d) a^2+b^2
(e) $a^2+2ab-b^2$

2. The factors of x^2-5x-6 are:
(a) $(x-3)(x+2)$
(b) $(x+6)(x-1)$
(c) $(x-6)(x+1)$
(d) $(x-3)(x-2)$
(e) none of these.

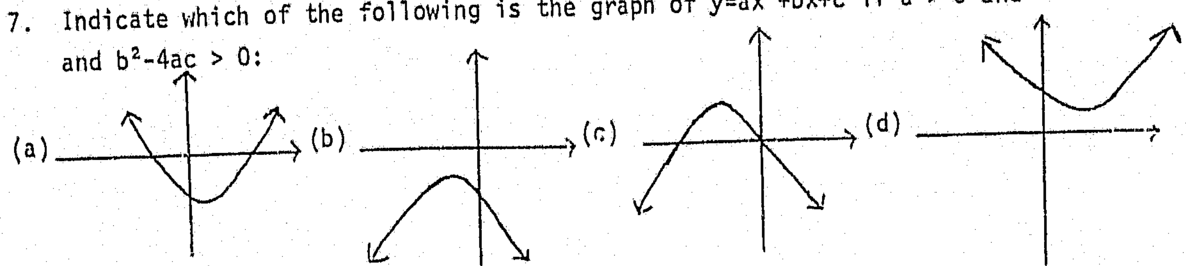
3. The number $27^{-\frac{2}{3}}$ is equal to
(a) $\frac{1}{9}$
(b) 9
(c) $\frac{1}{18}$
(d) 18
(e) none of these.

4. $3x^{-4}$ may also be written as
(a) $-12 \log x$
(b) $\frac{1}{3}x^4$
(c) $\frac{1}{\sqrt[4]{3x}}$
(d) $\frac{3}{x^4}$
(e) none of these.

5. If log 4 = 0,602 then log 40 is equal to
(a) 6,02
(b) 10,602
(c) 1,602
(d) 0,702
(e) none of these.

6. A candidate has to use logarithm tables to simplify $\sqrt{(13,3)^3+\sqrt[3]{91,5}}$.
Indicate in which step he is making an error if he proceeds as follows:
(a) multiply the logarithm of 13,3 by 3
(b) divide the logarithm of 91,5 by 3
(c) add the results of (a) and (b)
(d) divide the result of (c) by 2 and find the antilogarithm.
(e) none of the above steps.

7. Indicate which of the following is the graph of $y=ax^2+bx+c$ if $a > 0$ and $b^2-4ac > 0$:



(e) none of these

8. The forth term of the sequence 4; -6; 9;... is equal to

(a) -13
(b) $-13\frac{1}{2}$
(c) $-12\frac{1}{2}$
(d) -12
(e) none of these.

9. The simplified value of $\frac{\log 9+\log 27}{\log 9-\log 3}$ is

(a) -9 (Do not use log tables)
(b) 6
(c) -3
(d) 5
(e) none of these.

10. If $S = A(1 + \frac{r}{100})^n$, then r is equal to:

(a) $100(\frac{\log S}{n \log A} - 1)$

(b) $100(\sqrt[n]{\frac{S}{A} - 1})$

(c) $100(\frac{S.n}{A} - 1)$

(d) $(\sqrt[n]{100-1}).\frac{S}{A}$

(e) none of these.

11. The 20th term of the sequence 4; 9; 14;... is

(a) 61
(b) [illegible]$^{n}-1$
(c) 99
(d) $4(5^{19})$
(e, none of these.

12. The value of $\sum_{n=1}^{6} 3.(-2)^{n-1}$ is:

(a) 63
(b) -66
(c) 189
(d) -63
(e) none of these.

13. $\log_2 y - \log_2 x = 1$. If $y = 7$, then x is equal to:

(a) $\log_2 6$
(b) $1 - \log 3{,}5$
(c) $3{,}5$
(d) $(\log_2 7-1)^2$
(e) none of these.

14. k is the arithmetic mean of two numbers. If one number is x, then the other number is:

(a) $k-2x$
(b) $k - \frac{x}{2}$
(c) $2k-x$
(d) $\frac{k}{2} - x$
(e) $\frac{k-x}{2}$

15. $\log 3$; $\log 3r$; $\log 3r^2$; $\log 3r^3$ is a

(a) Geometric sequence with common ratio of r
(b) Arithmetic sequence with common difference of $\log r$
(c) Geometric sequence with common ratio of $\log r$
(d) Arithmetic sequence with common difference of r
(e) none of these.

16. x; $\sqrt{x}$ is a Geometric sequence. The third term of the sequence is:

(a) x
(b) $\sqrt{x^3}$
(c) 1
(d) $\frac{1}{\sqrt{x}}$
(e) $2\sqrt{x}-x$

17. $\frac{1}{\sqrt{2}-\sqrt{3}}$ is identical to

(a) $\sqrt{2} + \sqrt{3}$
(b) $\sqrt{2} - \sqrt{3}$
(c) $-\sqrt{2} - \sqrt{3}$
(d) -1
(e) $\frac{1}{5}$

18. $\log x + \log x^2 + \log x^4$ is equal to:

(a) $\log(x+x^2+x^4)$
(b) $\log x^6$
(c) $7 \log x$
(d) $\log 7x$
(e) $\log \frac{1}{x}$

19. If $t = \frac{ax}{p} + s$ then

(a) $x = \frac{p(t+s)}{a}$

(b) $x = \frac{t}{sp} - a$

(c) $x = \frac{p(t-s)}{a}$

(d) $x = \frac{pt-s}{a}$

(e) none of these.

20. If $\log x = \bar{3},124$ then $\log x^{\frac{1}{2}}$ is equal to:

(a) $\bar{6},248$
(b) $1,562$
(c) $\bar{1},562$
(d) $1,062$
(e) $\bar{2},562$

Bibliography

Allen, W.H. Instructional media research: Past, present and future. Audiovisual Communication Review, 1971, 19, 5-18.

Baker, E.L. The Technology of Instructional Development. In R.M.W. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973. Pp245-285.

Basler, D.I. and Smawley, R.B. Tukey's compact versus classic tests. Journal of Experimental Education vol 36(3), 1968. Pp86-88.

Birley, D. Planning and Education. London: Routledge and Keegan Paul: 1972.

Blalock, H.M. and Blalock, A.B. (Eds.) Methodology in Social Research. McGraw-Hill, 1968, p340.

Bligh, D. Teaching Students. Devon: Exeter University Teaching Services, 1975.

Bloom, B.S. et al, Taxonomy of Educational Objectives. London: Longman. Vol. I, 1956. Vol. II, 1964.

Booth, G.D. and Miller, H.R. Effectiveness of monochrome and color presentations in facilitating affective learning. Audiovisual Communication Review 1974, 22, 409-422.

Bull, S.G. The role of questions in maintaining attention to texual material. Review of Educational Research, 1973, 43, 83-87.

Burkhart, R.C. (Ed.) The assessment revolution: New viewpoints for teacher evaluation. New York State Education Department, Albany. Buffalo, N.Y.: State University of New York, Division of Teacher Education and Certification, 1969.

Carnegie Commission on Educational Television. Public Television: a program for action; the report and recommendations. New York: Harper and Row, 1967.

Carver, R. The Coleman Report: Using Inappropriately Designed Achievement Tests. American Educational Research Journal, 1975, 12(1), 77-86.

Chao, L.L. Statistics: Methods and Analyses. McGraw-Hill, 1974.

Chervany, N.L., Collier. R.O., Fienberg, S.E., Johnson, P.E. and Neter, J.A. Framework for the Development of Measurement Instruments for Evaluating the Introductory Statistics Course. The American Statistician, 1977, 31(1), 17-23.

Chu, G.B. and Schramm, W. Learning from television: What the research says. Stanford, California: Stanford University Press, 1967.

Cohen, A.A. An Evaluation of Instructional Television. In R. Eberlein, P.J. van Zyl and B.J. Venter (Ed.), Instructa 78: A collection of papers read at the Educational Technology Symposium, Johannesburg. Butterworth 1979(1). Pp338-344.

Cohen, A.A. Cognitive Learning and Media: The Effects of T.V. Viewing on Mental Skills. In R. Eberlein, P.J. van Zyl and B.J. Venter (Eds.), Instructa 78: A collection of papers read at the Educational Technology Symposium, Johannesburg. Butterworth 1979(2). Pp219-225.

Cohen, A.M. and Brawer, F.B. Measuring faculty performance. ERIC Clearinghouse for Junior Colleges, Monograph No. 4. Washington, D.C.: American Association of Junior Colleges, 1969. ED 031222.

Cohen, A.M., Rose, C. and Trent, J.W. Teaching Technology and Methods. In R.M.W. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973. Pp1023-1035.

Cohen, A.M. and Trent, J.W. Research on Teaching in Higher Education. In R.M.W. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973. Pp997-1067.

Coles, C. The Present State and Future Expectations of Educational Technology: A Provocative Review. In R. Eberlein, P.J. van Zyl and B.J. Venter (Eds.). Instructa 78: A collection of papers read at the Educational Technology Symposium, Johannesburg. Butterworth: 1979. Pp409-415.

Coulson, D.B. and Hambleton, R.K. Some validation methods for domain-referenced test items. Laboratory of Psychometric and Evaluative Research. Report No. 7 Amherst, Mass: The University of Massachusetts, 1974.

Cronback, L.J., and Suppes, P. (Eds.) Research for Tomorrow's Schools - Disciplined Inquiry for Education. Report of the Committee on Educational Research of the National Academy of Education. London: Macmillan Callien Macmillan, Limited, 1969.

Dessart, D.J. and Frandsen, H. Role of Media in Mathematics Instruction. In R.M.W. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973. Pp1185-1187.

Dwyer, F.M. Exploratory studies i the effectiveness of visual illustrations. Audiovisual Communication Review, 1970, 18, 235-249.

Dwyer, F.M. The effect of I.Q. level on the instructional effectiveness of black-and-white and color illustrations. Audiovisual Communication Review 1976, 24, 49-62.

Ebel, R.L. Measuring Educational Achievement. Prentice-Hall, 1965.

Elton, L.R.B. Is it possible to assess teaching? Evaluating Teaching in Higher Education. University Teaching Methods Unit. London, 1975.

Lsson, M.E. Psychological Foundations of Education. New York: Holt, Rinehart and Winston, 1972.

French, J.S. and Michael, W.M. Standards for Educational and Psychological Tests and Manuals. Washington D.C.: American Psychological Association, 1966.

Gagné, R.M. Learning hierarchies. Educational Psychologist, 1968, 6,1-9.

Gagné, R.M. The conditions of learning. (3rd ed.) New York: Holt, Rinehart and Winston, 1977.

Gagné, R.M. and Paradise, N.E. Abilities and learning sets in knowledge acquisition. Psychological Monographs: General and Applied, 1961, 75(14), No. 518.

Glaser, R. Instructional Technology and the Measurement of Learning Outcomes-Some Questions. American Psychologist, 1963, 18, 519-521.

Glaser, R. and Cooley, W.W. Instrumentation for Teaching and Instructional Management. In R.M.W. Travers (Ed.) Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973. Pp832-857.

Gordon, G.N. Educational Television. New York: The Library of Education, 1965.

Gordon, G.N. Classroom Television-New Frontiers in I.T.V. Communication Arts Books; Hastings House, 1970.

Gropper, G.L. Does 'programmed' television need active responding. Audiovisual Communication Review, 1967, (15), p5-22.

Gropper, G.L. and Lumsdaine, A.A. An Experimental Comparison of a Conventional T.V. Lesson with a Programmed T.V. Lesson Requiring Active Student Response, Report No. 2. Studies in Televised Instruction: The Use of Student Response to Improve Televised Instruction. Pittsburgh, Pa.: American Institutes for Research, March 1961.

Hatfield, L.L. Computers in mathematics instruction. In J.W. Wilson and L.R. Carry (Eds.), Studies in mathematics, Vol. 19. Reviews of recent research in mathematics education. Stanford, Calif.: School Mathematics Study Group, 1969. Pp129-152.

Judd, W.A. and Glaser, R. Response latency as a function of training method, information level, acquisition, and overlearning. Journal of Educational Psychology Monograph, 1969, 60(4), Part 2.

Kanner, J.H. The instructional effectiveness of colour in television: A review of the evidence. Stanford Calif.: Stanford University, 1968.

Kogan, L.S. Statistics. In Annual Review of Psychology vol 11. Palo Alto: Annual Reviews Inc. 1960. p208.

Ksobiech, K.J. The importance of perceived task and type of presentation in student response to instructional television. Audiovisual Communication Review, 1976, 24, 401-412.

Levie, W.H. and Dichie, K.E. The Analysis and Application of Media. In R.M.W. Travers (Ed.). Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973. Pp858-882.

Lomax, D. The systemsapproach to education. In R. Eberlein, P.J. van Zyl and B.J. Venter (Eds.). Instructa 78: A collection of papers read at the Educational Technology Symposium, Johannesbueg. Butterworth, 1979(1). Pp8-14.

Lomax, D. Computer-based teaching strategies: the Cybercom system. The South African Journal of Science. 1979(2), Vol. 75, 249-252.

Lomax, D. and Bayne , S. Computer Assessment of student attitudes Journal of Education, University of Natal. Durban and Pietermaritzburg. 1978, 10, 55-58.

Lucas, R.J. The relationship of training and experience variables to teachers' skills in judging the effectiveness of ITV programmes. Audiovisual Communication Review, 1974(22). Pp79-89.

Lumsdaine, A.A. Instruments and media of instruction. In N.L. Gage (Ed.). Handbook of Research on Teaching. Chicago: Rand McNally, 1963. Pp583-682.

Macdonald-Ross, M. The Quality of Course Materials. In Evaluating Teaching in Higher Education: A collection of conference papers. London: University Teaching Methods Unit, 1975.

Mager, R.F. Preparing Instructional Objectives. California: Fearon, 1962.

Meyer, J.H.F. Faculty evaluation and development - putting theory into practice. In R. Eberlein, P.J. van Zyl and B.J. Venter (Eds.). Instructa 78: A collection of papers read at the Educational Technology Symposium, Johannesburg. Butterworth 1979. Pp329-337.

Moldstad, J.A. Selective Review of Research Studies Showing Media Effectiveness: A Primer for Media Directors. Audiovisual Communication Review, 1974, 22(4), 387-407.

Murphy, J. and Gross, R. Learning by Television. New York: Fund for the Advancement of Education, 1966.

Nedelsky, L. Science Teaching and Testing. Harcourt, Brace and World, 1965.

Norman, C. and Rieber, M. Facilitation of concept formation in children by the use colour cues. Journal of Experimental Psychology, 1[illegible], 76, 460-463.

Ormell, [illegible]: Blooms Taxonomy and the Objecti[illegible] [illegible] Education. Educational Research, 1974, 17(1).

Pearson, E.S. (Ed.). [illegible]ometrica Tables for Statisticians. Third Edition. Cambridge U.P. Vol. 1, 1966.

Piper D.W. Educational Objectives. In Improving Teaching in Higher Education. University Teaching Methods Unit. Leicester: Cavendish Press, 1978.

Popham, W.J. Development of a performance test of teaching proficiency. Final report, U.S. Office of Education Project No. 5-0566-2-12-1. Washington D.C.: U.S. Office of Education, 1967.

Popham W.J. Performance tests of teaching proficiency: Rationale, development, and validation. American Educational Research Journal, 1971, 8, 105-117.

Popham, W.J. Criterion Referenced Instruction. Belmont: Fearon, 1973.

Popham, W.J. Criterion-Referenced Measurement. Englewood Cliffs: Prentice-Hall, 1978.

Rothkopf, E.Z. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-336.

Roueche, J.E. and Pitman, J.C. A modest proposal: Students [illegible]an learn. San Francisco: Jossey-Bass, 1972.

Rovinelli, R.J. and Hambleton, R.K. On the use of content specialists in the assessment of criteri[illegible]n-referenced test item validity. Laboratory of Pscychrometric and Evaluative Research Report No. 24. Amherst, Mass.: The University of Massachusetts, 1976.

Rowntree, D. Educational Technology in Curriculum Development. London: Harper and Row, 1974.

Ryans, D.G. Characteristics of teachers, their description, comparison, and appraisal. Washington, D.C.: American Council on Education, 1960.

Salomon, G. Sesame Street in Israel: Its instructional and psychological effect on children. New York: Children's Television Workshop, 1976.

Sceiford, M. Television for learning: Future research opportunities. Educational Communication and Technology Journal. 1978, Vol. 26(1). Pp55-63.

Schramm, W. The research on programmed instruction. An annotated bibliography. U.S. Office of Education Bulletin No. 35. Washington D.C.: Government Printing Office, 1964.

Shulman, L.S. Psychology and mathematics education. In E.G. Begle (Ed.), Mathematics Education. The sixty-ninth Yearbook of the National Society for the Study of Education, Part I. Chicago: NSSE, 1970. Pp23-71.

Simpson, E. Educational Objectives in the Psychomotor Domain, 1967. In M.B. Kapfer (Ed.). Behavioural Objectives in Curriculum Development. Englewood Cliffs: Educational Technology Publications, 1971.

Skinner, B.F. Science and Human Behaviour. New York: Free Press, 1953.

Snowberg, R.L. Bases for the selection of background colors for transparencies. Audiovisual Communication Review, 1973, 21, 191-207.

Solomon, D., Rosenberg, L., and Bezdek, W. Teacher behaviour and student learning. Journal of Educational Psychology, 1964 Vol. 55. Pp23-30.

Sorenson, G., and Gross, C.F. Teacher Appraisal- a matching process. An Occasional Report. Los Angeles: University of California, Los Angeles, Graduate School of Education, Center for the Study of Evaluation, 1967.

Stephens, M.D. and Roderick, G.W. (Eds.). Teaching Technique in Adult Education. Newton Abbot: David and Charles, 1971.

Stickel, D.W. A critical review of the methodology and results of research comparing televised and face-to-face instruction. In R.M.W. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973. P860.

Tosti, D.T. and Ball, J.R. A 'behavioural approach' to instructional design and media selection. Audiovisual Communication Review, 1969, 17, 5-25.

Tukey, J.W. A quick compact, two-sample test to Duckworth's specifications. Technometrics, vol 1(1), 1959. Pp31-48.

Tyler, R.W. Basic Principles of Curriculum and Instruction. Chicago: University of Chicago Press, 1949.

Tyler, R.W. How to improve instruction on the basis of evaluation. In W.J. Popham: Criterion Referenced Instruction, p83. Englewood Cliffs: Prentice-Hall, 1978.

University Teaching Methods Unit. Improving Teaching in Higher Education. Leicester; Cavendish Press 1978.

Whitaker, P.D. The Evolution in the Use of Media in British and European Universities. In R. Eberlein, P.J. van Zyl and B.J. Venter (Ed.). Instructa 78: A collection of papers read at the Educational Technology Symposium, Johannesburg. Butterworth, 1979. PP241-250.

Winer, B.J. Statistical Principles in Experimental Design. McGraw-Hill, 1971.

Young, D.A. 'The Effectiveness of Television Instruction in Service Course Teaching': Internal Report. Department of Applied Mathematics, University of the Witwatersrand, Johannesburg, 1978(1).

Young, D.A. The Effect of Immediate Feedback on Achievement in Televised Statistics Lectures: Internal Report. Department of Applied Mathematics, University of the Witwatersrand, 1978(2).

Young, D.A. 'What does a student learn in a statistics lecture? The belief and the truth'. Unpublished paper. University of the Witwatersrand, Johannesburg, 1979.

Zawels, J. The Cybercom Multimode Education Communication System. Mineo. Department of Education, University of the Witwatersrand, 1976.

Zeaman, D. and House, B.J. The role of attention in retardate discrimination learning. In N.R. Ellis (Ed.), Handbook of Mental Deficiency. New York: McGraw-Hill, 1963. Pp159-223.