SC-RR-68-473
August 1968

Research Report

A CONTINUOUS ACCEPTANCE SAMPLING PROCEDURE FOR VARIABLES BASED ON CUMULATIVE SUMS

Richard R. Prairie
Statistics and Computing Division 1643
William J. Zimmer
Statistics Research Division 1723
Sandia Laboratories, Albuquerque

When printing a copy of any digitized SAND Report, you are required to update the markings to current standards.

SANDIA LABORATORIES

OPERATED FOR THE U. S. ATOMIC ENERGY COMMISSION BY SANDIA CORPORATION | ALBUQUERQUE, NEW MEXICO; LIVERMORE, CALIFORNIA

Issued by Sandia Corporation
a prime contractor to the
United States Atomic Energy Commission

LEGAL NOTICE

This report was prepared as an account of Government-sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.

SC-RR-68-473

A CONTINUOUS ACCEPTANCE SAMPLING PROCEDURE FOR VARIABLES BASED ON CUMULATIVE SUMS

Richard R. Prairie
Statistics and Computing Division 1643
William J. Zimmer
Statistics Research Division 1723
Sandia Laboratories, Albuquerque

August 1968

ABSTRACT

A procedure is presented for the acceptance sampling of continuously produced product where the response taken from the product is a variables measurement following a normal distribution with unknown mean and variance. The procedure, which is of the type originally proposed by Beattie, and which is amenable to destructive testing, is essentially a cumulative sum chart with acceptance and rejection zones. The criterion of effectiveness is the discrimination between two qualities. The probability of acceptance for a given quality is

$$P_A = \frac{L(0)}{L(0) + L^*(0)}$$

where L(0) is the Average Run Length (ARL) in the accept zone and L*(0) is the ARL in the reject zone. The equations for L(0) and L*(0) are derived and the numerical technique used for solving these equations is described. Figures from which values of L(0) and L*(0) can be obtained are given, and a group of procedures is presented for various levels of quality and sample size.

TABLE OF CONTENTS

	Page
Introduction	5
The Procedure	6
Computation of the Average Run Length	10
Results	11
List of References	15

A CONTINUOUS ACCEPTANCE SAMPLING PROCEDURE FOR VARIABLES BASED ON CUMULATIVE SUMS

Introduction

Consider the situation where it is desired to accept or reject continuously produced product based on the proportion of the product that yields a response beyond a one-sided specification limit. The response taken from the product is a variables measurement following a normal distribution with unknown mean and variance and is perhaps determined by means of a destructive test.

For example, thermal batteries are produced at, say, a rate of 50 per day. The test to assess the quality of the battery is a destructive test with a principal output being rise time. It is desired to discriminate between two qualities, where quality is measured by the proportion of rise times above a specification limit of 1.5 seconds, and to accept or reject that day's product by means of a small sample taken daily. It can be reasonably assumed that the rise time of a thermal battery follows a normal distribution. The usual continuous sampling plans cannot be used because the test is destructive and because the desire to discriminate between two qualities exists. The majority of the continuous plans control the Average Outgoing Quality Limit and do so by requiring that the defective items be replaced by assured nondefective items, which is not possible with destructive testing. In addition, continuous plans generally entail variable inspection rates, which is often undesirable. A continuous plan that does allow destructive testing has been proposed by Hillier.[1] His procedure is designed to minimize the total expected cost. However, the plan has disadvantages such as requiring a priori information and rather complex calculations to implement the plan.

Another procedure that has been used in this situation is to choose arbitrarily a number of days, say 20, and call the group of batteries produced during that period a <u>lot</u>. After the batteries from these days have been accumulated, a random sample

is selected and a decision made to accept or reject the lot. There are several objections to this type of procedure:

1. The time delay in making a decision on the product is too long

2. Several bad days of production could cause a large quantity of good product from other days to be rejected

3. The definition of a lot is arbitrary.

Procedures are given in this paper which can be applied where:

1. The responses follow a normal distribution with unknown mean and variance

2. The proportion of the product exceeding a one-sided specification limit determines quality

3. The testing is destructive.

These procedures have the following properties:

1. They permit the product to be accepted or rejected on a continuing basis

2. They require only simple calculations

3. They permit a fixed sampling rate

4. They discriminate between qualities with a given power.

The Procedure

The procedure given here is based on cumulative sums and is a type proposed by Beattie.[2] This type of procedure has many advantages in this situation. A small sample is chosen regularly from the product, and the acceptance or rejection decision is also made regularly on the basis of this sample and preceding ones. The

overall cumulative sum chart allows changes in quality to be detected easily, and the bound for which rejection takes place controls the quality of the output.

Beattie's procedure is to set up a cumulative sum chart which has an accept zone and a reject zone, and to accept or reject product appropriately. When the sampling rate is the same in both zones, the probability of acceptance (P_A) for a given quality is:

$$P_A = \frac{L(0)}{L(0) + L^*(0)} \tag{1}$$

where L(0) is the average run length (ARL) for that quality of product, that is, the average number of samples in the accept zone, and L*(0) is the ARL for that quality of product in the reject zone. Using an equation for L(0) and nomograms for its solutions given by Evan and Kemp,[3] Beattie gives procedures in the case where the assumption of a normal distribution with known variance is made.

In the situation of interest here, it is necessary to control the process mean and variance simultaneously. That is, the responses are assumed to be normally independently distributed (μ, σ^2) with the unknown mean μ and the unknown variance σ^2, and a sampling procedure is desired where $P_A = (1 - \alpha)$ when p, the proportion of the product beyond a specification limit A is p_1, and $P_A = \beta$ when p is p_2. Under these assumptions and assuming an upper limit A,

$$Y = \frac{\sqrt{n}(A - \overline{X})}{s} \sim \text{Non-central } t \left(\text{parameter } \delta = \frac{\sqrt{n}(A - \mu)}{\sigma}\right)$$

where

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}} .$$

If the statistic

$$D_i = k - Y_i = k - \frac{\sqrt{n}\,(A - \overline{X}_i)}{s_i} \tag{2}$$

is accumulated from samples of n items taken at regular intervals, and K is chosen such that $\sum D_i$ decreases if $p = p_1$ and increases if $p = p_2$, the decision lines on the accept and reject zones can be horizontal. The procedure is as follows:

1. Start the cumulation at 0.

2. Accept product as long as $\sum D_i < h$. (When $\sum D_i < 0$, return cumulation to 0.)

3. Reject product when $\sum D_i > h$; restart cumulation at $h + h^*$ and continue rejecting product until $\sum D_i < h$. (When $\sum D_i > h + h^*$, return cumulation to $h + h^*$.)

4. When h is crossed from above, accept product and restart cumulation at 0.

Values of n, h, h*, and k can be found such that product will be accepted $100(1 - \alpha)\%$ of the time when $p = p_1 \left[P_A(p_1) = 1 - \alpha\right]$ and rejected $100(1 - \beta)\%$ of the time when $p = p_2 \left[P_A(p_2) = \beta\right]$.

It might be noted that if there is no restriction that the sampling rate be held constant, then it might be reasonable to increase the sampling rate in the reject zone. However, this decision should be based on the economics and time restrictions involved in sampling and a consideration of the usefulness of setting aside this material for later 100-percent inspection or using it to determine causes of deterioration of quality. In some cases, the rejected product could more economically be reused as raw material or reworked.

To obtain L(0) and L*(0) and thus P_A, for values of p, n, h, h*, and k, consider L(z), the ARL in the accept zone of the procedure for which the first cumulation starts at a point z units from the 0 boundary. L(z) can be written recursively in terms of a weighted average of one plus the ARL from u for all u where the weights are the probabilities of moving from z to u on the next cumulation; that is,

$$L(z) = \int_{-\infty}^{\infty} \left[1 + L(u)\right] f_y(k + z - u)du .$$

Under the restrictions on the cumulations made by the procedure, this becomes

$$L(z) = 1 + L(0)\left[1 - F_y(k + z)\right] + \int_0^h L(u) f_y(k + z - u)du, \tag{3}$$

where f_y and F_y are the non-central t density and distribution functions, respectively.

The ARL in the reject zone, $L^*(0)$, that is, the ARL necessary for $\sum D_i$ to move from $h + h^*$ to h, is thus equivalent to the ARL necessary for $\left(-\sum D_i\right)$ to move from 0 to h^*. Then $L^*(z)$ can be written:

$$L^*(z) = \int_{-\infty}^{\infty}\left[1 + L^*(u)\right] f_y\,(u + k - z)du$$

and under the restrictions on the cumulations, this becomes

$$L^*(z) = 1 + L^*(0)F_y(k - z) + \int_0^{h^*} L^*(u) f_y(u + k - z)du. \tag{4}$$

Using the results of setting z = 0 in Equations (3) and (4), solving for L(0) and $L^*(0)$, and substituting back into (3) and (4), respectively, one can now write these equations as

$$L(z) = 1 + \frac{1 - F_y(k + z)}{F_y(k)} + \int_0^h L(u)\left\{f_y(k + z - u) + \frac{1 - F_y(k + z)}{F_y(k)}\, f_y(k - u)\right\}du \tag{5}$$

and

$$L^*(z) = 1 + \frac{F_y(k - z)}{1 - F(k)} + \int_0^{h^*} L(u)\left\{f_y(u + k - z) + \frac{F_y(k - z)}{1 - F_y(k)}\, f_y(u + k)\right\}du. \tag{6}$$

Equations (5) and (6) are now seen to be Fredholm integral equations of the second kind in which the kernel is a function of the non-central t density and distribution functions. These equations must be solved for L(0) and $L^*(0)$ in order to yield the desired procedures.

Computation of the Average Run Length

In order to solve (5) for L(0) as a function of p, n, k, and h, and (6) for L*(0) as a function of p, n, k, and h*, it was decided to replace the integral equation by a finite system of linear algebraic equations. (See Reference 4, Chapter II, paragraph 1.) Thus, a formula for approximate integration is used and Equation (5) becomes:

$$L(z) = 1 + \frac{1 - F_y(k + z)}{F_y(k)} + \sum_{i=1}^{N} B_i L(u_i) \left[f_y(k + z - u_i) + \frac{1 - F_y(k + z)}{F_y(k)} f_y(k - u_i) \right] + p(z) \quad (7)$$

where p(z) is the error. In particular, successively setting $z = 0, u_1, \ldots, u_N$ in Equation (7), results in a system of equations which, if p(z) is ignored, can be solved for L(0).

It is obvious that the accuracy of the result obtained by replacing (5) by the system of linear equations will be higher if less error is made in replacing the integral by the sum. Since enlarging the number of ordinates of the function to be applied increases the difficulty of solving the system of equations, it is generally desirable to use the most exact of the quadrature formulas; therefore, the method of Gaussian quadrature with constant weight function was used. Thus, in the system of Equation (7) for $z = 0, u_1, u_2, \ldots, u_N$:

$$B_i = \frac{h}{2} A_i, \quad u_i = \frac{h}{2} \left[1 + x_i \right], \quad i = 1, 2, \ldots, N$$

where A_i and x_i are given in Krylov.[5] A similar approximation can be made to L*(z).

The solution of (7) for L(0) requires the evaluation of the non-central t density and distribution functions. For the distribution function $F_y(t; \delta, d)$, where δ is the non-centrality parameter and d is the degrees of freedom, a program was used that was based on formula (9) in Reference 6. For the density function $f_y(t; \delta, d)$, different methods, depending upon the degrees of freedom, were used. For $d \leq 20$, the

polynomials given by Resnikoff and Lieberman on page 26 of Reference 7, were used. For $d > 20$, the following expression (also found in Reference 7) was used:

$$f_y(t; \delta, d) = \frac{d!}{2^{\frac{d-1}{2}} \Gamma\left(\frac{d}{2}\right) \sqrt{\pi d}} e^{-\left(\frac{1}{2} \frac{d\delta^2}{d+t^2}\right)} \left(\frac{d}{d+t^2}\right)^{\frac{d+1}{2}} H_d\left(\frac{-\delta t}{\sqrt{d+t^2}}\right)$$

where

$$H_d(a) = \frac{1}{d!} b^d e^{-\frac{1}{2}(b+a)^2} \sqrt{\frac{2\pi b^2}{d+b^2}} \left[1 - \frac{3d}{4\left(d+b^2\right)^2} + \frac{5d^2}{6\left(d+b^2\right)^3}\right]$$

and

$$b = \frac{-a + \sqrt{a^2 + 4d}}{2} .$$

Using these methods, the accuracy of the density is adjudged to be at least five decimal places. By checking results for L(0) at different values of p, n, k, and h using different numbers of intervals in the quadrature, it is believed that the results given are accurate to within ±0.1 for values of $L(0) < 1000$.

Results

Figures 1 through 21 present L(0) and L*(0) as functions of h and h*, respectively, for various values of p, n, and k. The p values are 0.005, 0.01, 0.02, 0.03, 0.05, 0.07, and 0.10, and the n values are 5, 10, and 20. The combinations of these values made up the 21 figures, each of which contains an L(0) figure and an L*(0) figure. Each figure contains six curves, corresponding to six different k values. Each k value on a given figure corresponds exactly to a k value on another figure with the same sample size but a different p. In this case the k value is

$$k = \frac{\delta_{p_1} + \delta_{p_2}}{2},$$

where δ_{p_1} is the parameter of the non-central t distribution with a proportion p_1 beyond a specification limit A, and δ_{p_2} is defined similarly. Other k values may be approximately matched for two p values, but there are indications that

$$k = \frac{\delta_{p_1} + \delta_{p_2}}{2}$$

is optimum.

On some of the figures there are curves which are starred. This is done when the curve for a specific k value cannot be plotted on its proper graph because of the scale on the ordinate, i.e., L(0) must be plotted on the L*(0) graph and L*(0) must be plotted on the L(0) graph. For this reason it was necessary to represent these curves on the opposite figures; thus, the starred curve on the L(0) and h figure is actually giving values of L*(0) and h*, and the starred curve on the L*(0) and h* figure is giving values of L(0) and h.

Using these figures, it is possible to obtain procedures for discriminating between any two of the given p values. The probabilities of acceptance for the two qualities, that is, p values, can be restricted by appropriate choices from the L(0) and L*(0) values. Examples of procedures for various levels of quality and various sample sizes are given in Table I. The procedures may be obtained from the figures by choosing an h and an h* which give L(0) and L*(0) at the acceptable quality, p_1, such that

$$P_A = \frac{L(0)}{L(0) + L^*(0)} = 1 - \alpha$$

and then checking the P_A at the rejectable quality, p_2. The values h and h* can then be adjusted until

$$P_A(p_1) \cong 1 - \alpha, \quad P_A(p_2) \cong \beta$$

as required.

TABLE I

Procedures for Given Levels of Quality
($\alpha = 0.1$, $\beta = 0.1$)

Qualities		Sample Size	Chart Boundaries			Average Run Lengths in Undesirable Situations	
p_1	p_2	n	h	h*	k	L*(0) at p_1	L(0) at p_2
0.005	0.05	5	2.25	9.80	4.72	5.40	3.65
0.01	0.07	5	2.15	9.00	4.25	5.25	3.72
0.01	0.10	5	1.40	4.60	4.03	3.10	2.28
0.02	0.10	5	2.05	8.00	3.73	5.30	3.78
0.005	0.03	10	2.35	5.88	7.05**	4.30	3.38
0.005	0.03	10	1.70	7.90	~6.66	4.40	3.45
0.01	0.05	10	1.90	4.30	6.28**	3.55	2.82
0.01	0.05	10	2.72	3.80	~6.66	3.80	3.20
0.01	0.10	10	0.60	1.30	5.70	1.55	1.40
0.02	0.05	10	3.53	10.50	5.85	9.40	7.60
0.02	0.07	10	2.10	5.21	5.60	4.50	3.50
0.03	0.07	10	3.21	8.80	5.31	8.50	7.05
0.005	0.02	20	2.40	4.45	10.35	3.70	3.11
0.005	0.03	20	1.18	2.10	9.97	2.00	1.75
0.01	0.03	20	2.51	4.62	9.41**	4.20	3.60
0.01	0.03	20	1.81	8.55	~8.84	5.30	4.18

**This value of k, that is

$$k = \frac{\delta_{p_1} + \delta_{p_2}}{2},$$

is the recommended value.

In computing the example procedures, three examples were computed with the designed k and also with another approximate k which could be obtained from the figures. In each case, it appears that the designed k is better in the sense that it provides a procedure with the same protection as the procedure based on the second k

value but with smaller expected run lengths in undesirable situations, that is, $L^*(0)$ at p_1 and $L(0)$ at p_2.

As an example of the use of the procedures, consider the following situation. An item is produced at the rate of 200 a day and a sample of size n = 10 per day is to be selected. It is important that the proportion of the responses from the items above an upper specification limit, A, be kept small. As long as this proportion is less than or equal to 0.01, the items should be accepted with a high probability, say 0.90. If the proportion of the responses above A becomes as large as 0.05, then the items should be rejected with a high probability, also say 0.90.

Hence $p_1 = 0.01$, $p_2 = 0.05$, $\alpha \cong 0.10$, and $\beta \cong 0.10$. If k is selected so that $k = (\delta_{p_1} + \delta_{p_2})/2$, then k = 6.28. For n = 10, $p_1 = 0.01$, Figure 9 is used. For n = 10, $p_2 = 0.05$, Figure 12 is used. With some trial and error it is found, for h = 1.9 and h* = 4.3, that L(0) = 33, $L^*(0) = 3.5$ at p_1 and L(0) = 2.9 and $L^*(0) = 24$ at p_2. Therefore, at p_1,

$$P_A = \frac{33}{33 + 3.5} \sim 0.9$$

and at p_2,

$$P_A = \frac{2.9}{2.9 + 24} \sim 0.1 \ .$$

Hence, if a cumulative sum chart with boundaries 0 and 1.90 for the accept zone and 1.90 and 6.20 for the reject zone is used, and the statistic

$$D_i = 6.28 - \frac{\sqrt{10}(A - \overline{X}_i)}{s_i}$$

is accumulated, then such a procedure will have $P_A \sim 0.90$ at p = 0.01 and $P_A \sim 0.10$ at p = 0.05.

LIST OF REFERENCES

1. Hillier, Frederick S., "Continuous Sampling Plans with Destructive Testing," Journal of the American Statistical Association, Vol. 59, Nr. 306, June 1964, pp. 376-401.

2. Beattie, D. W., "A Continuous Acceptance Sampling Procedure Based Upon a Cumulative Sum Chart for the Number of Defectives," Applied Statistics, Vol. XI, No. 3, November 1962, pp. 137-147.

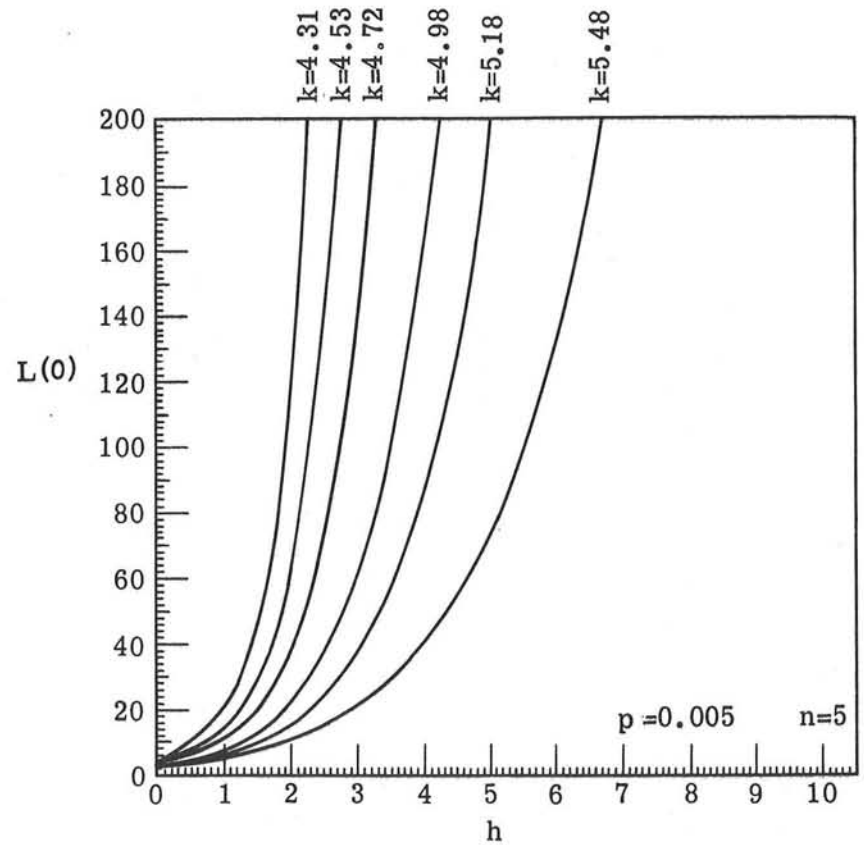
3. Ewan, W. O. and Kemp, K. W., "Sampling Inspection of Continuous Processes with no Autocorrelation Between Successive Results," Biometrika, Vol. 47 (1960), pp. 363-380.

4. Kantrovich, L. V. and Krylov, V. I., Approximate Methods of Higher Analysis, translated by Curtis D. Benster, Interscience Publishers Inc., New York, 1958.

5. Krylov, V. I., Approximate Calculation of Integrals, translated by Arthur H. Stroud, the Macmillan Co., New York, 1962.

6. Amos, D. E., "Representations of the Central and Non-Central t Distributions," Biometrika, Vol. 51 (1964), pp. 451-458.

7. Resnikoff, George J. and Lieberman, Gerald J., Tables of the Non-Central t Distribution, Tech Report No. 32, Stanford University, April 1, 1957.


k=4.31
k=4.53
k=4.72
k=4.98
k=5.18
k=5.48
L(0)
h
p =0.005
n=5

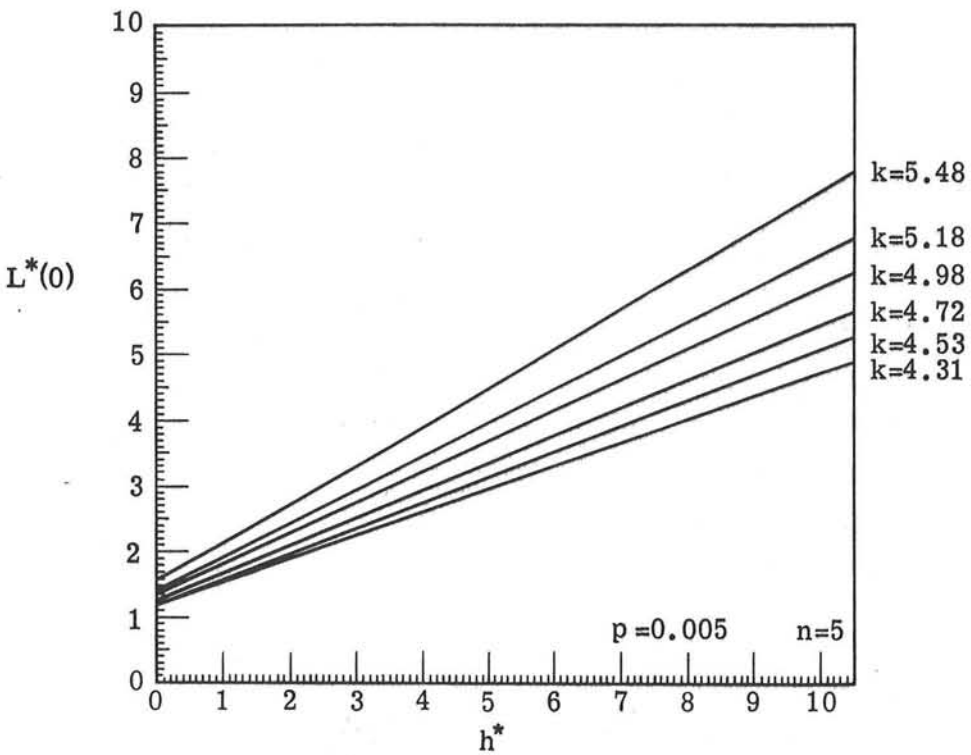
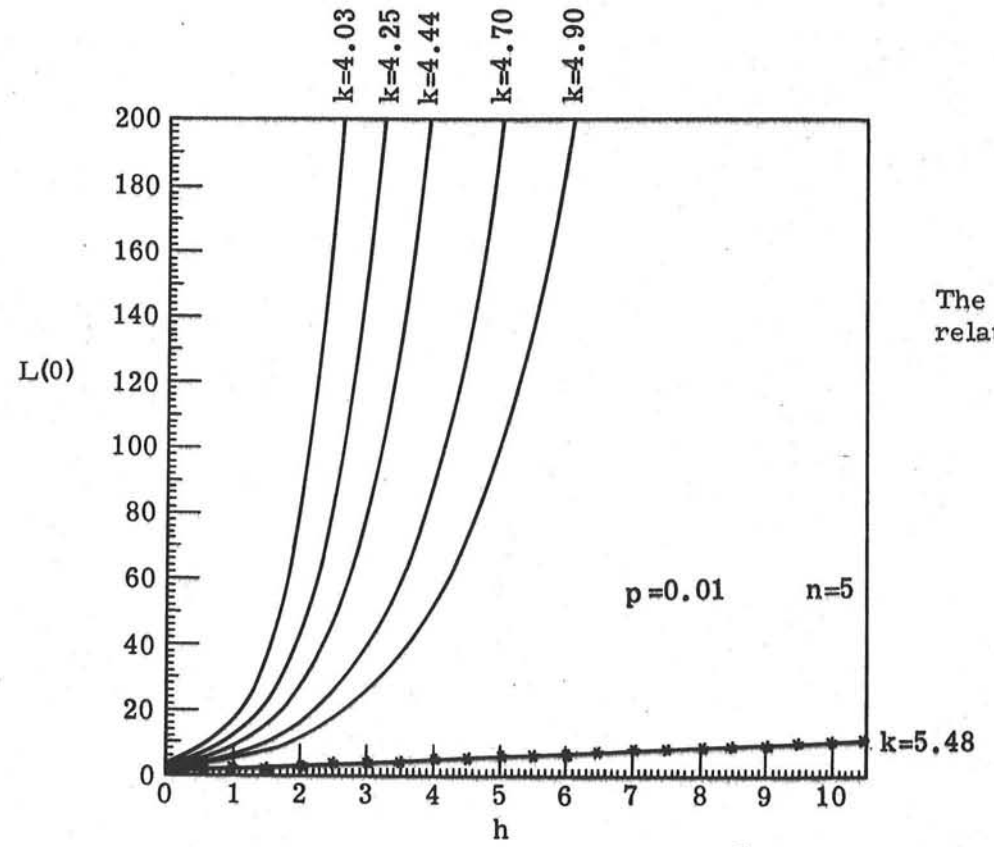
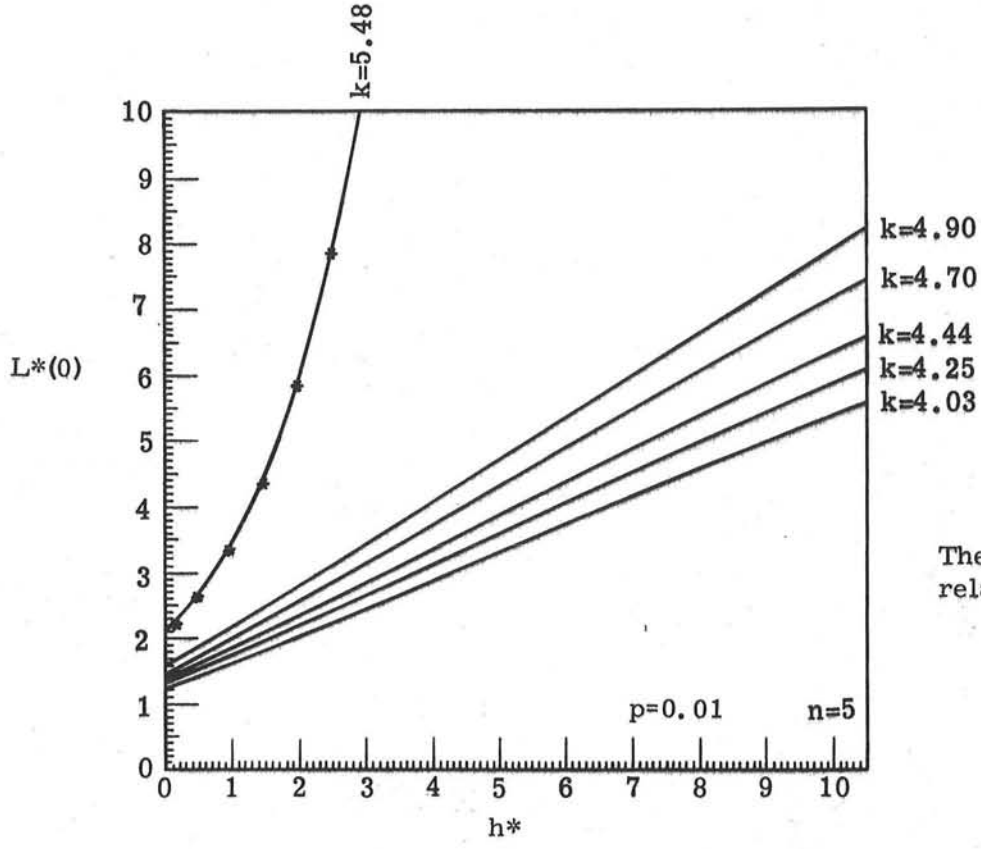

L*(0)
h*
k=5.48
k=5.18
k=4.98
k=4.72
k=4.53
k=4.31
p =0.005
n=5

Figure 1.

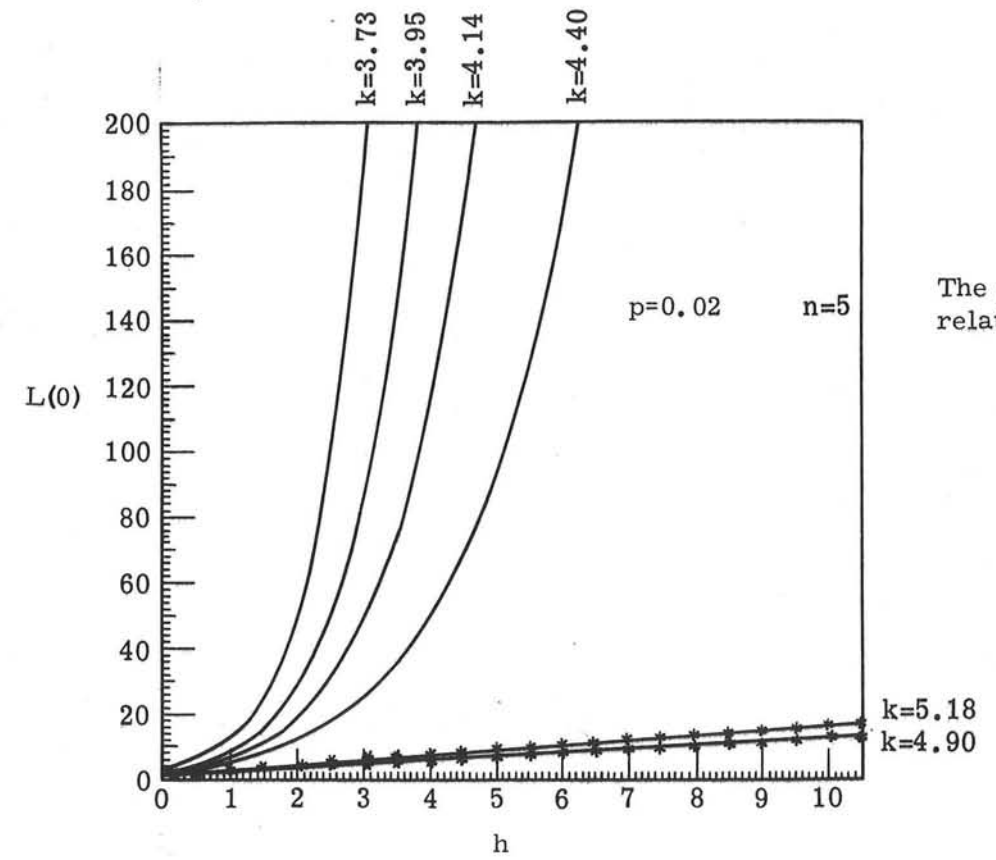


The starred curve relates L*(0) and h*.

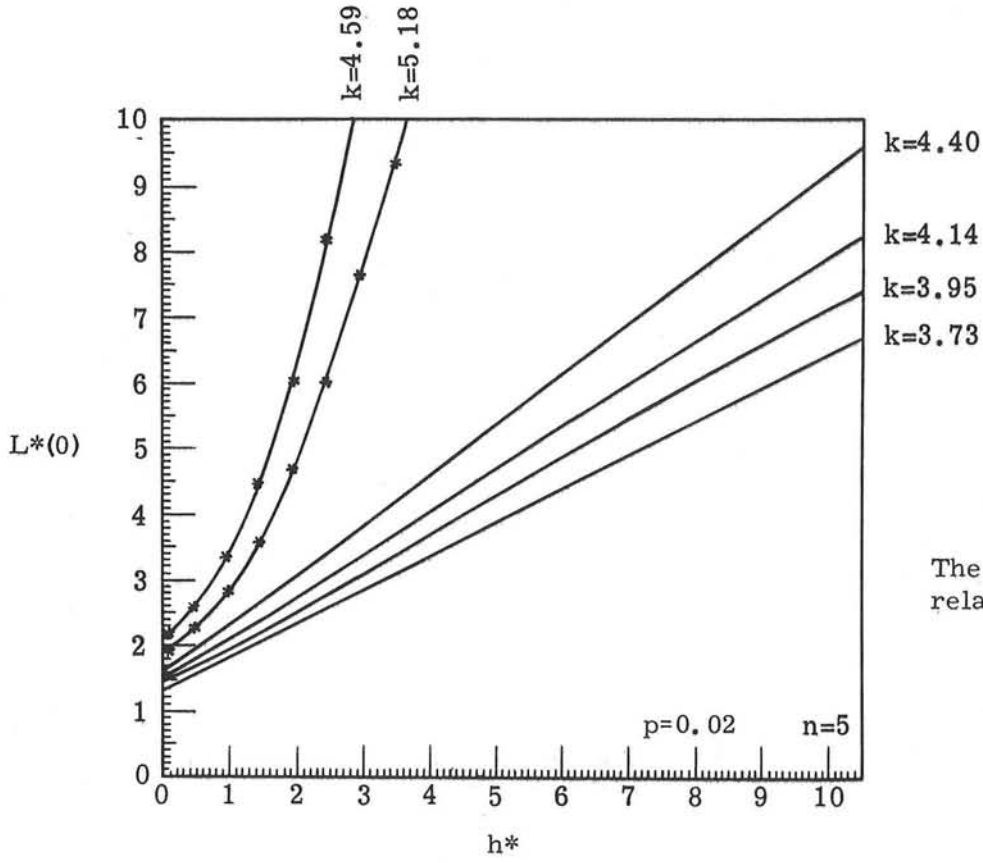


The starred curve relates L(0) and h.

Figure 2.

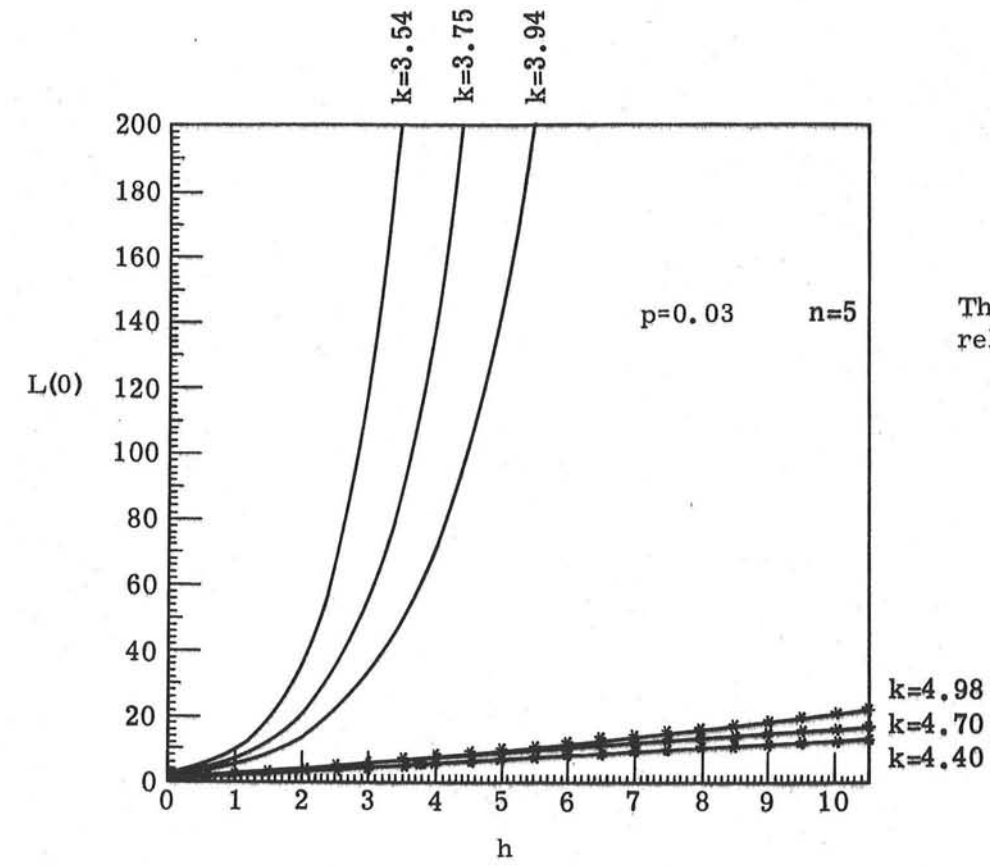


The starred curves related L*(0) and h*.

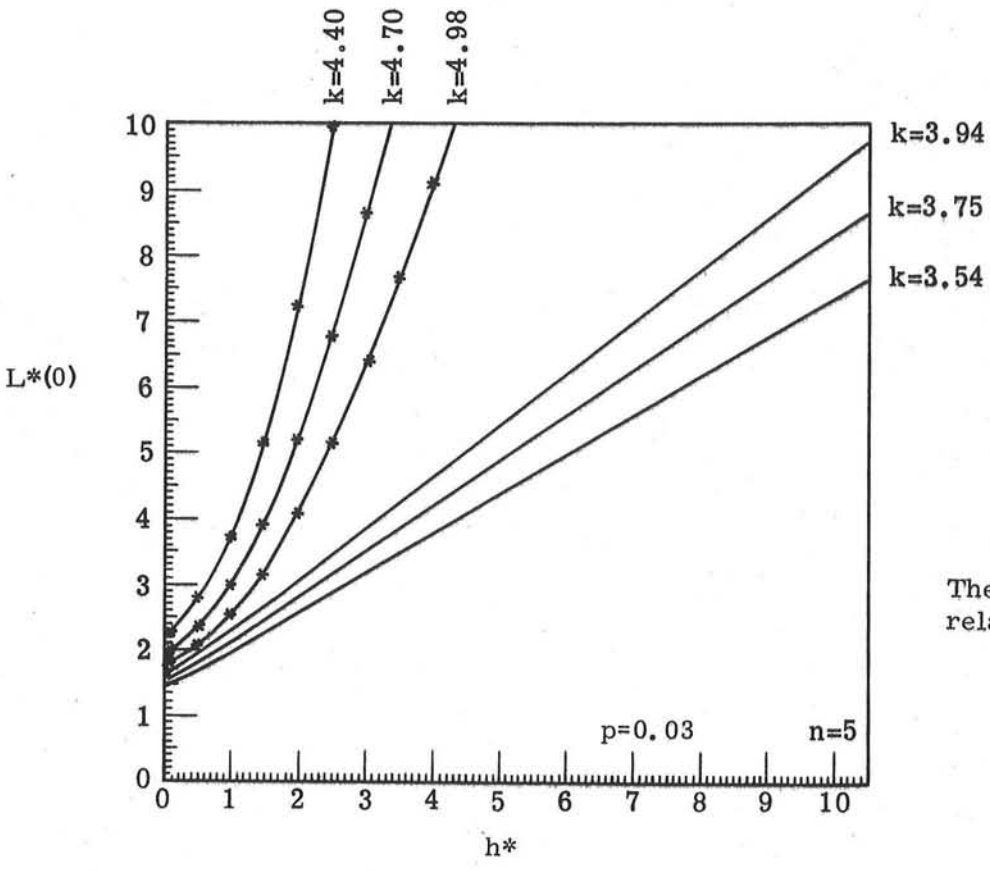


The starred curves relate L(0) and h.

Figure 3.

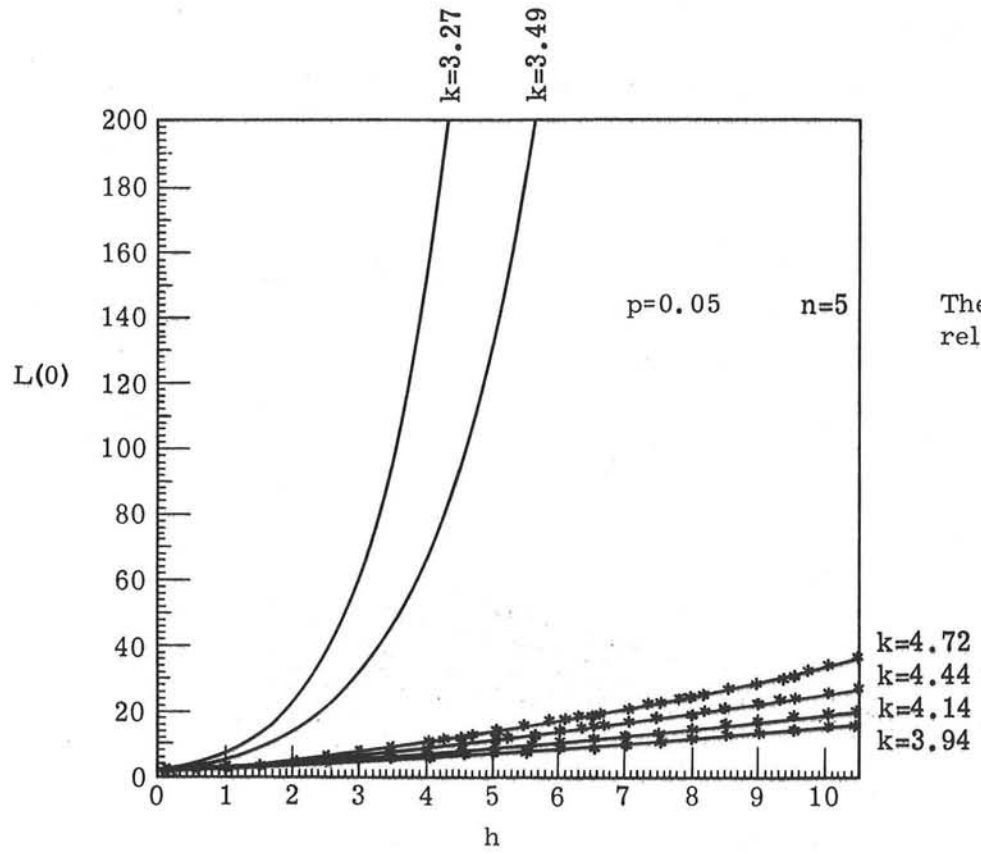


The starred curves relate L*(0) and h*.

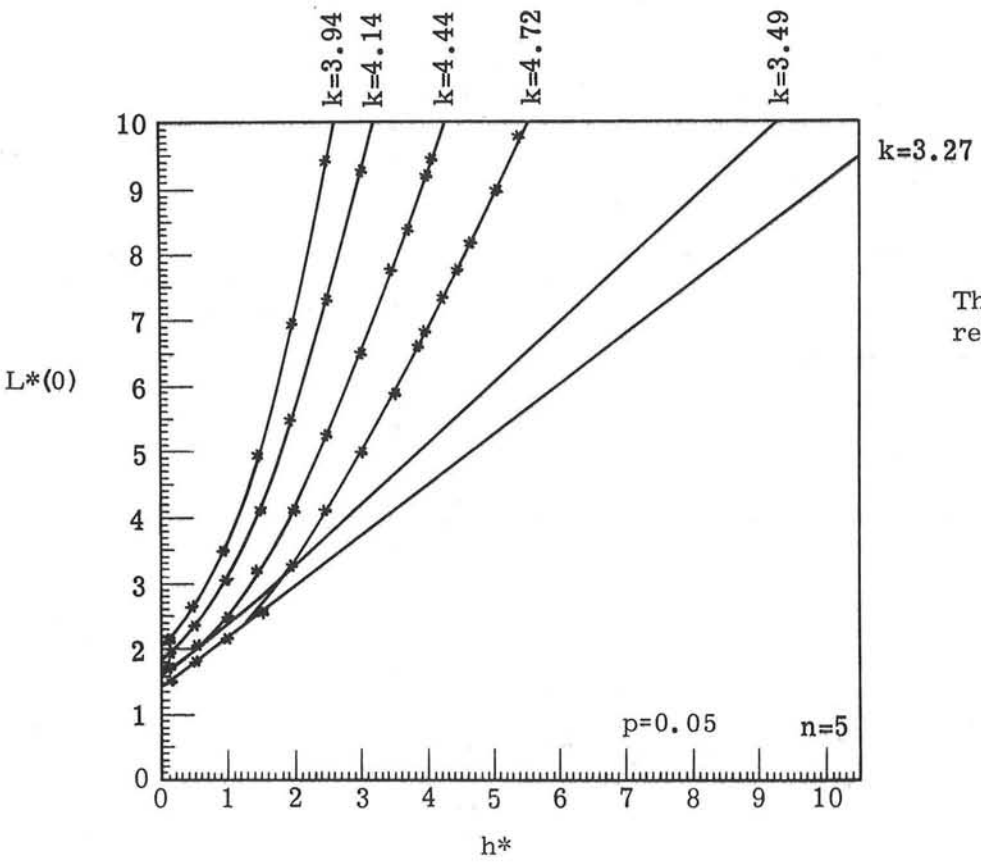


The starred curves relate L(0) and h.

Figure 4.

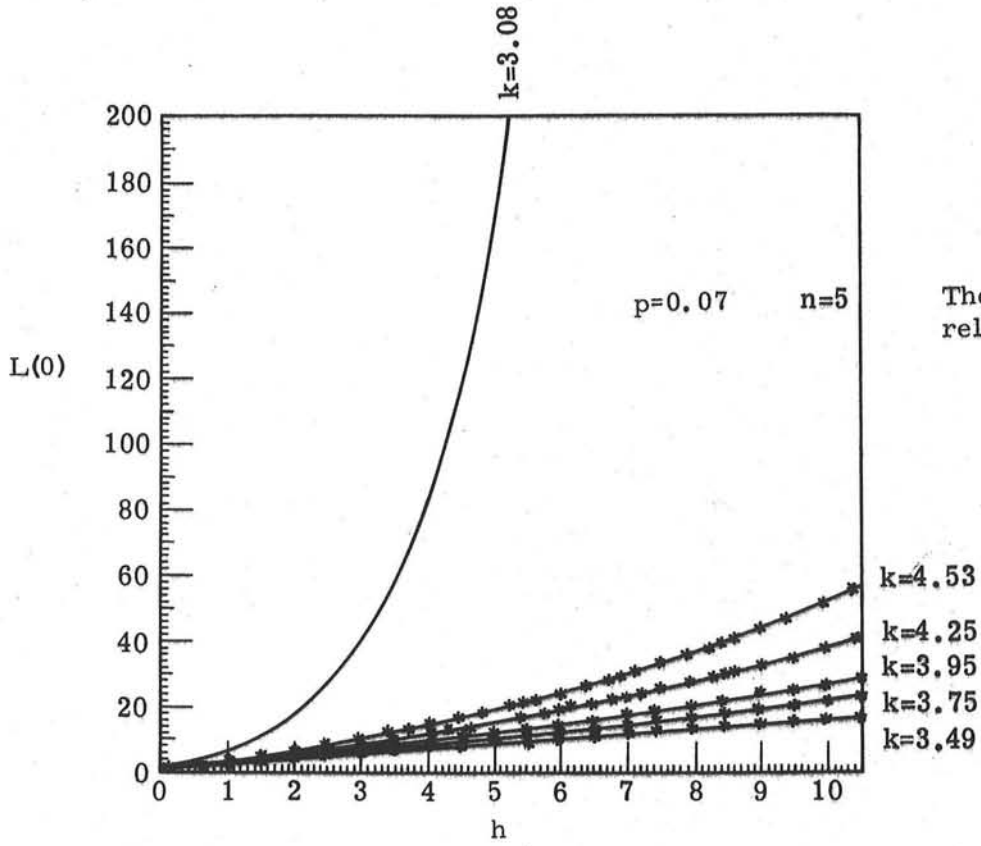


The starred curves relate L*(0) and h*.

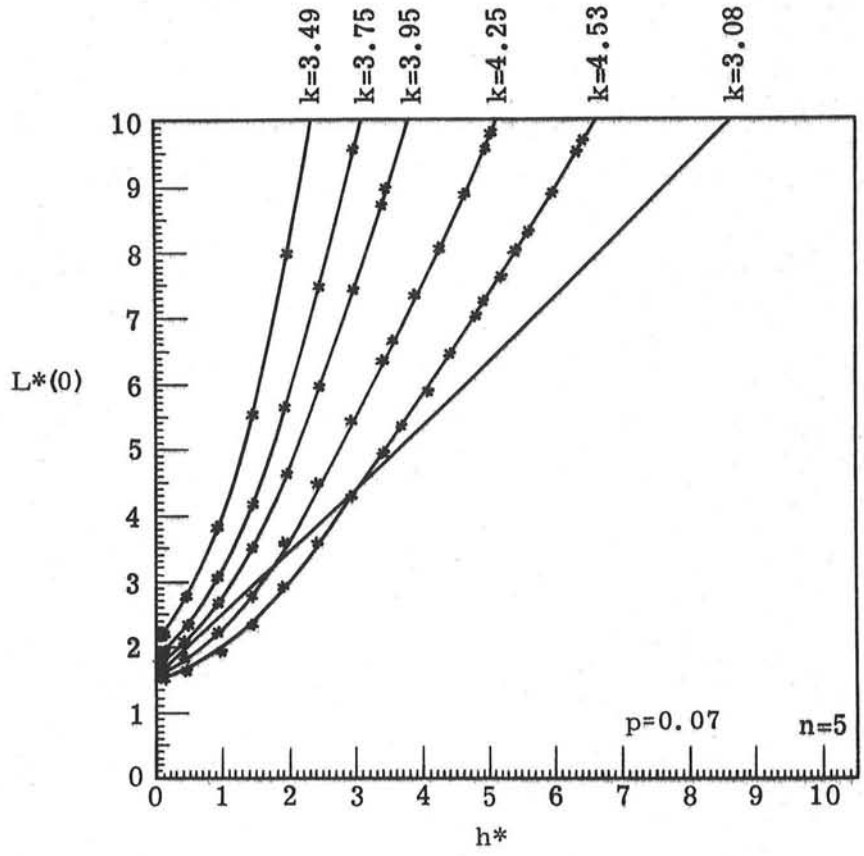


The starred curves relate L(0) and h.

Figure 5.

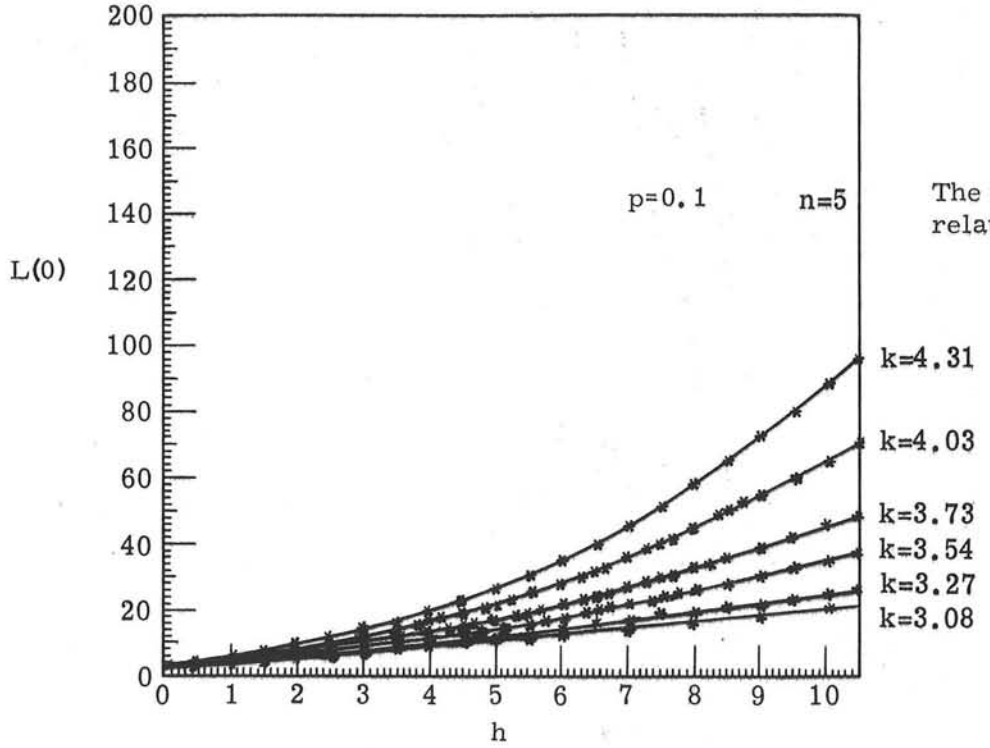


The starred curves relate L*(0) and h*.

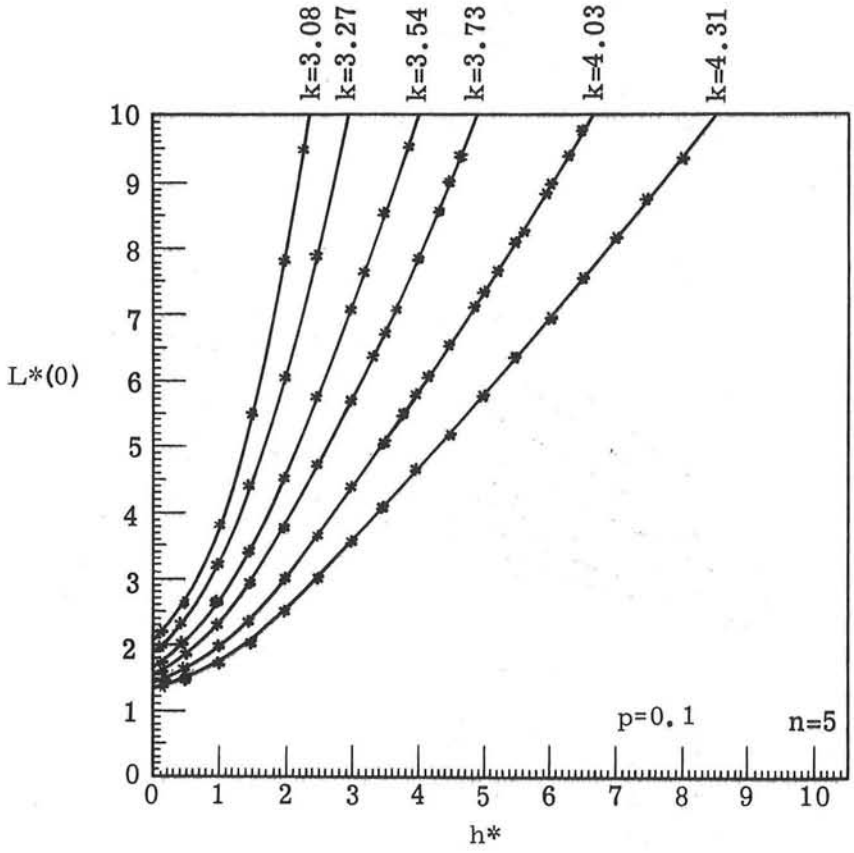


The starred curves relate L(0) and h.

Figure 6.

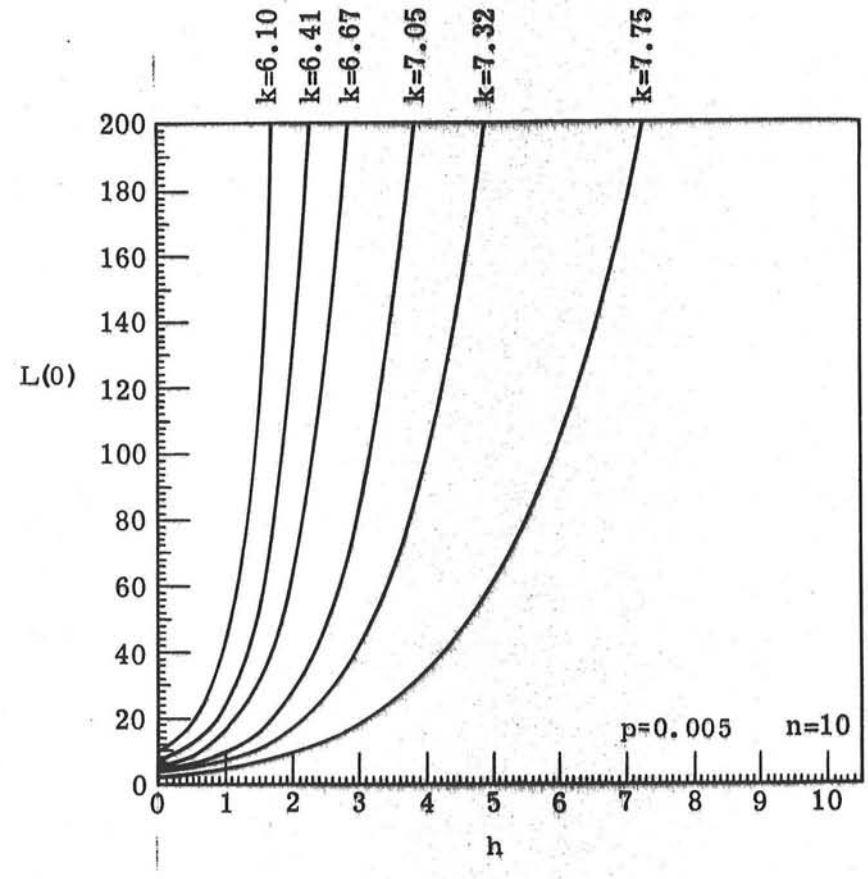


The starred curves relate L*(0) and h*.



The starred curves relate L(0) and h.

Figure 7.


k=6.10
k=6.41
k=6.67
k=7.05
k=7.32
k=7.75
L(0)
h
p=0.005
n=10

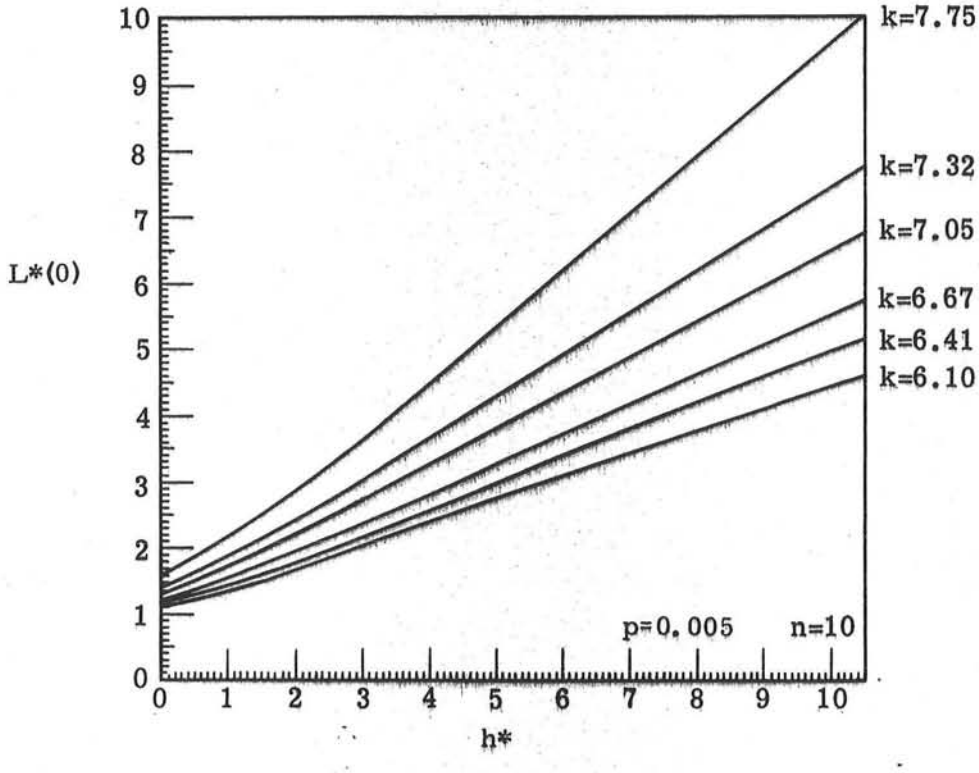
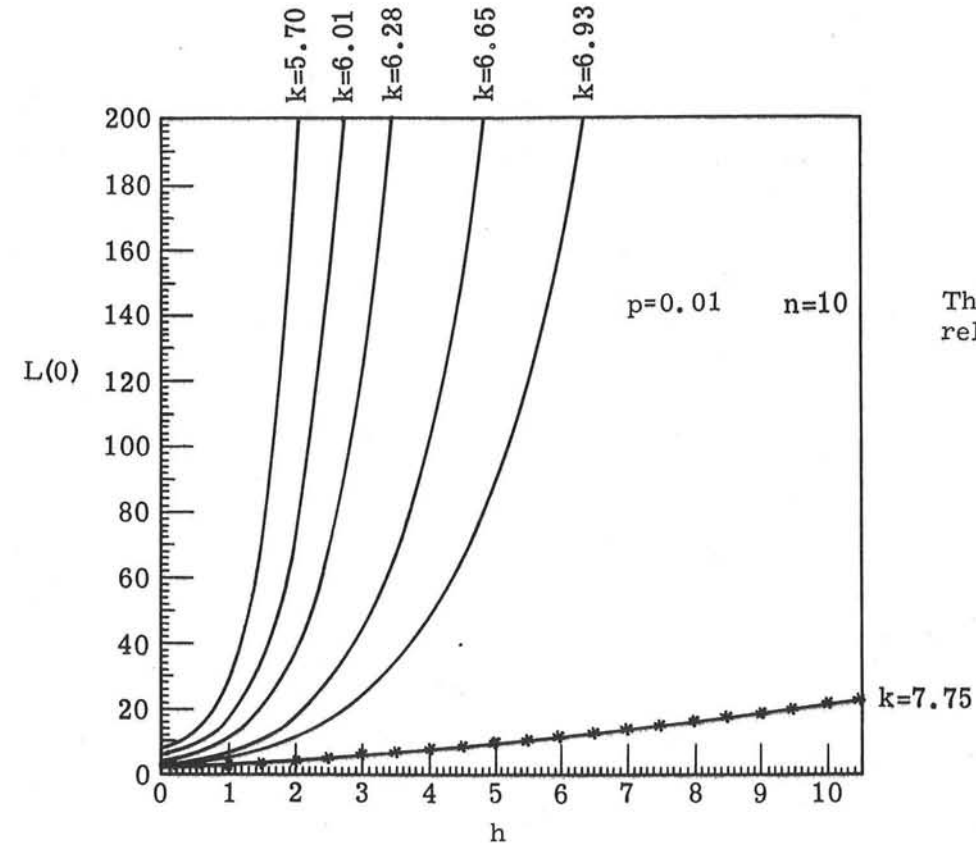
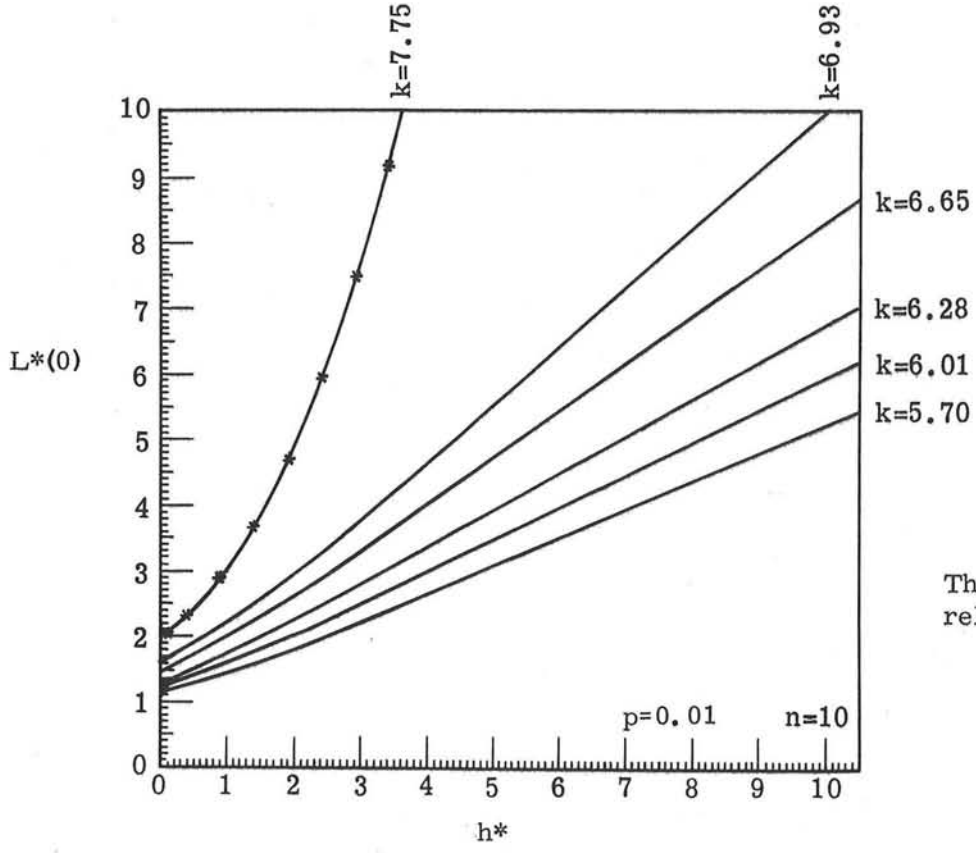

k=7.75
k=7.32
k=7.05
k=6.67
k=6.41
k=6.10
L*(0)
h*
p=0.005
n=10

Figure 8.

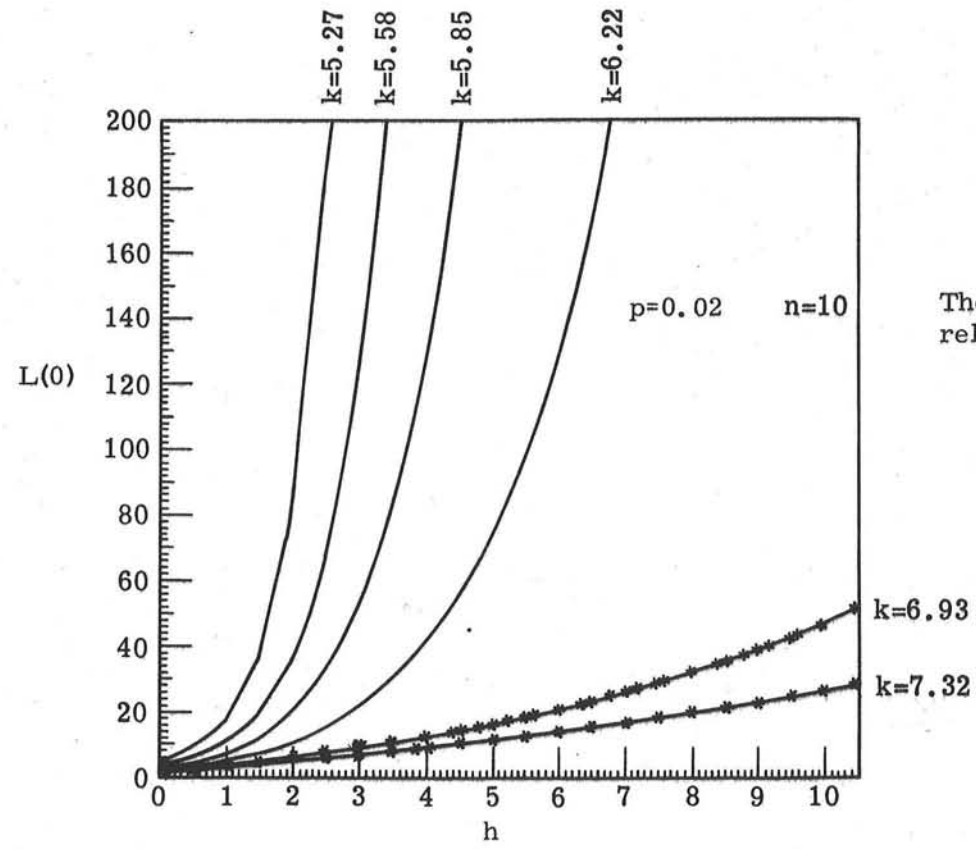


The starred curve relates L*(0) and h*.

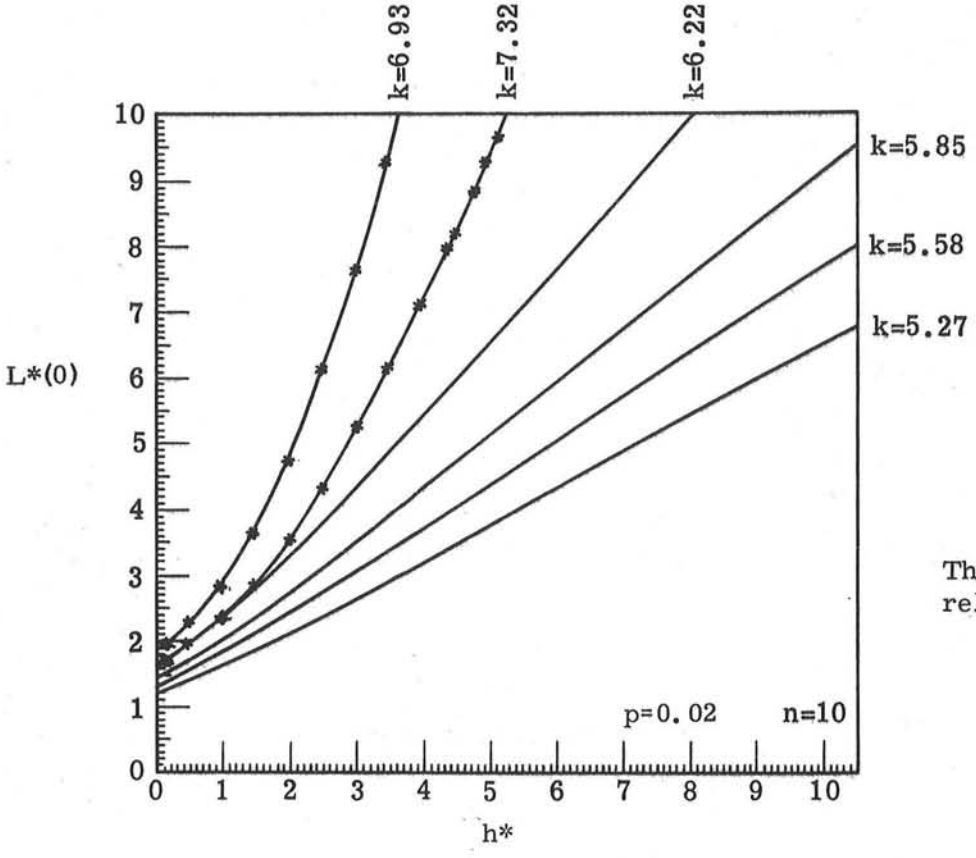


The starred curve relates L(0) and h.

Figure 9.

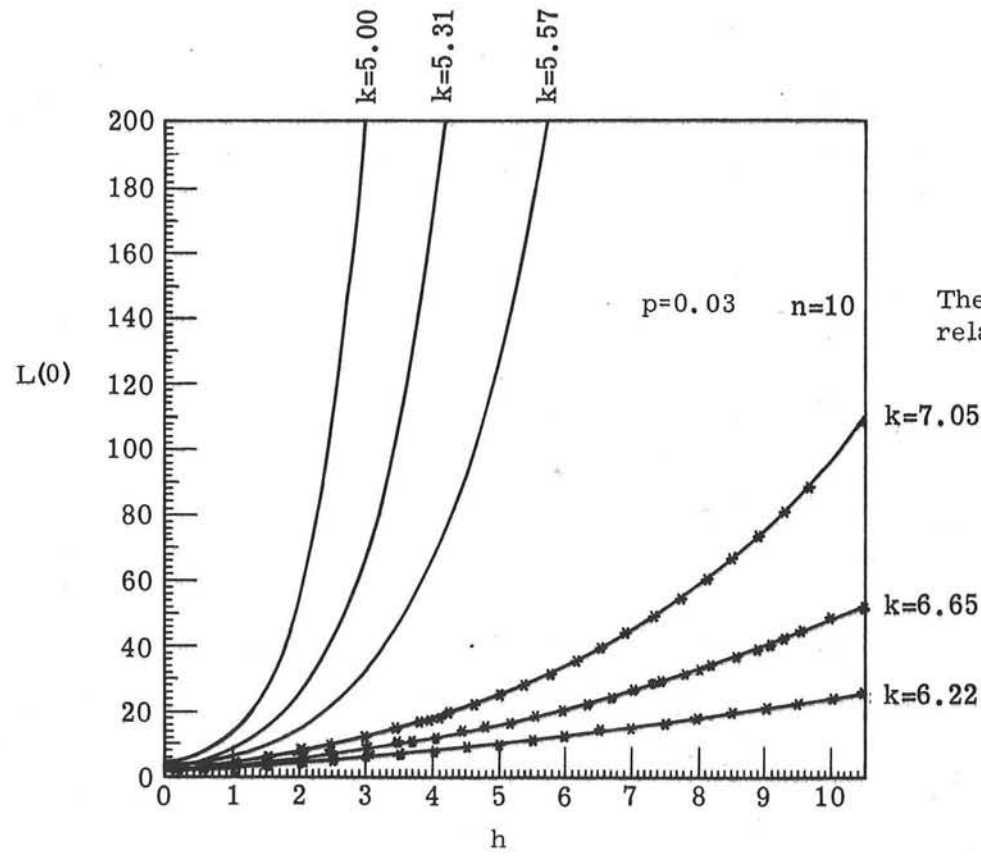


The starred curves relate L*(0) and h*.

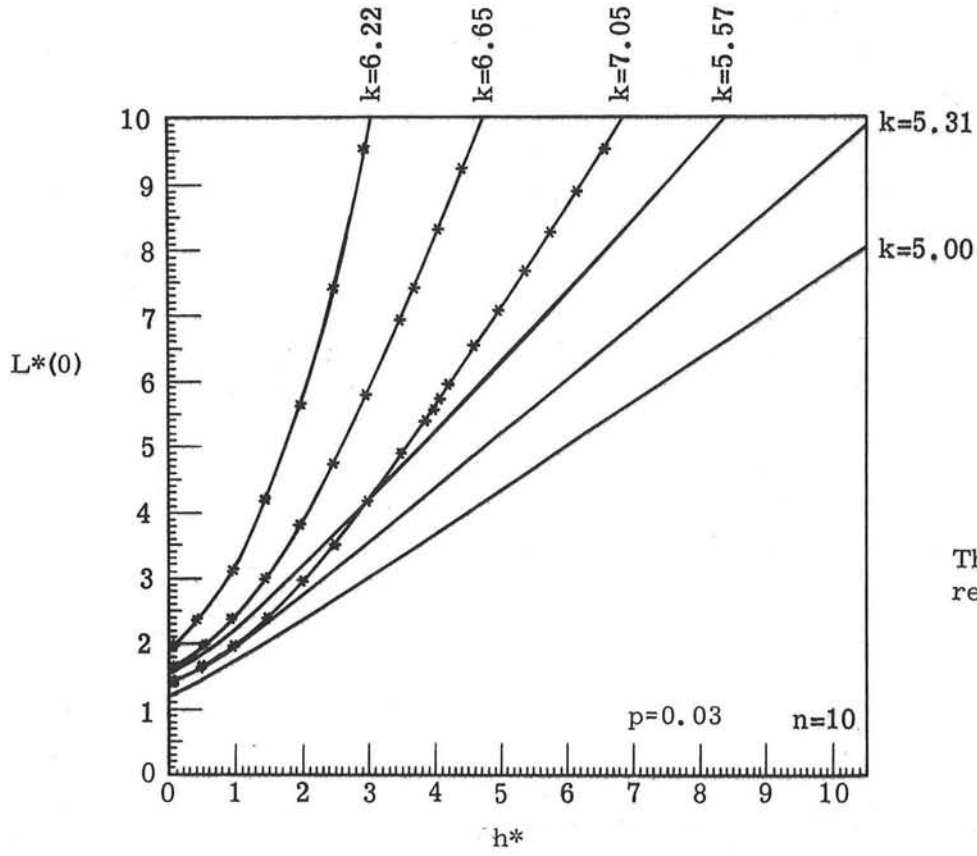


The starred curves relate L(0) and h.

Figure 10.

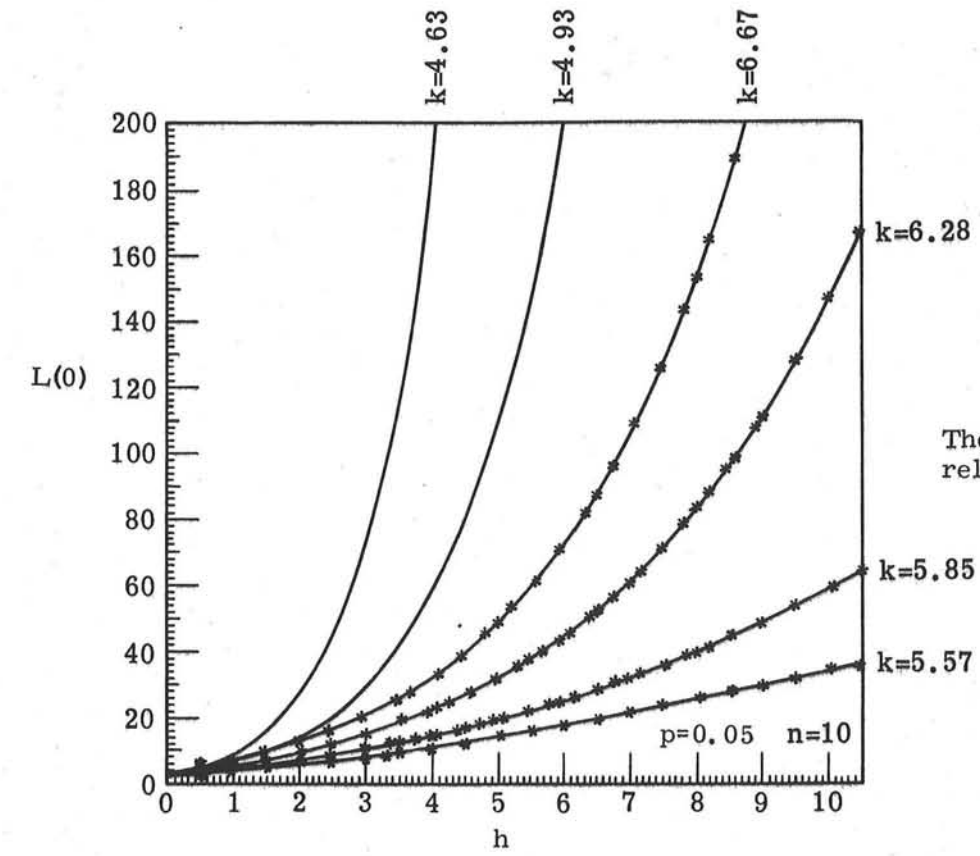


The starred curves relate L*(0) and h*.

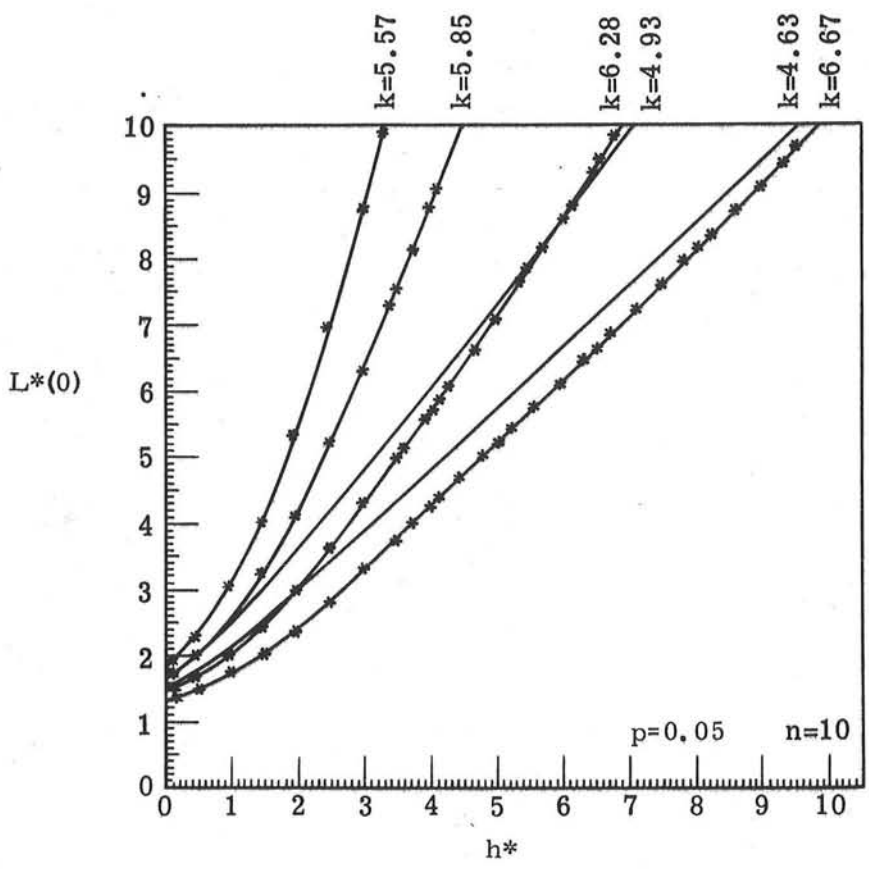


The starred curves relate L(0) and h.

Figure 11.



The starred curves relate L*(0) and h*.



The starred curves relate L(0) and h.

Figure 12.

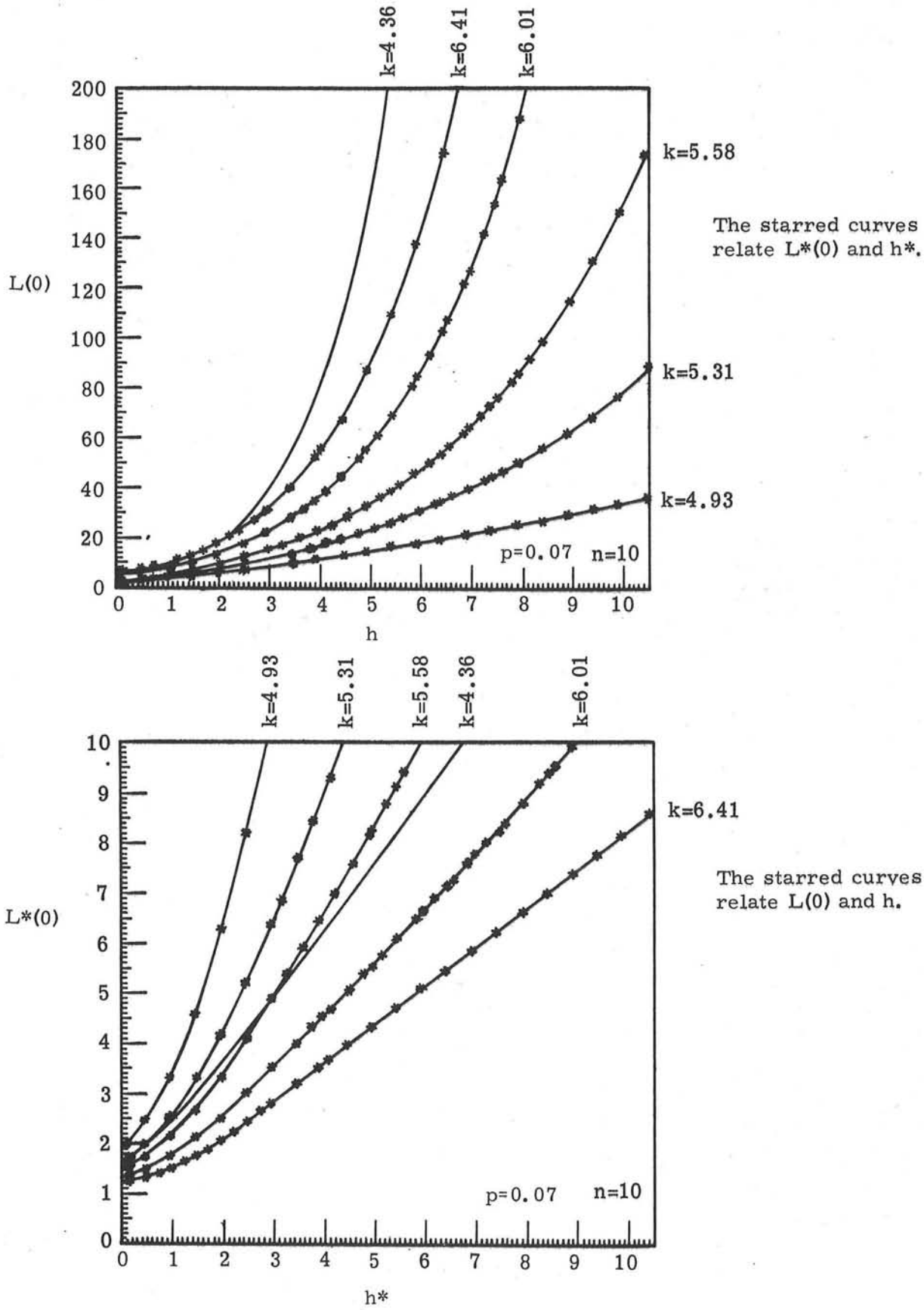
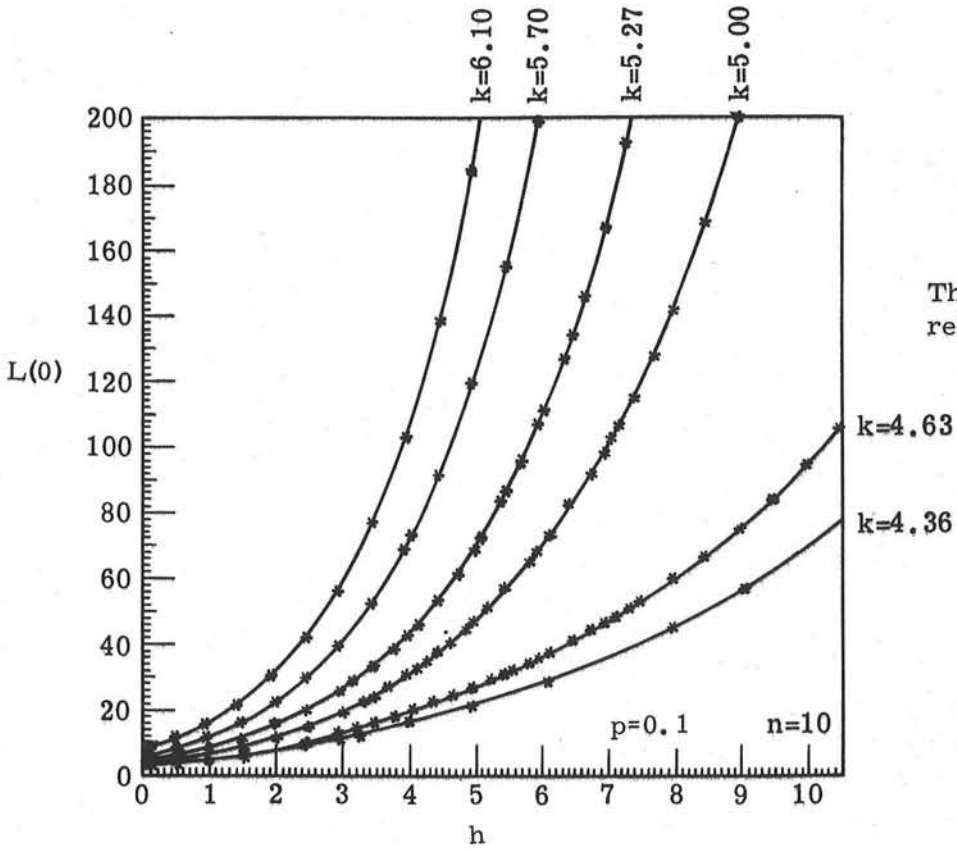
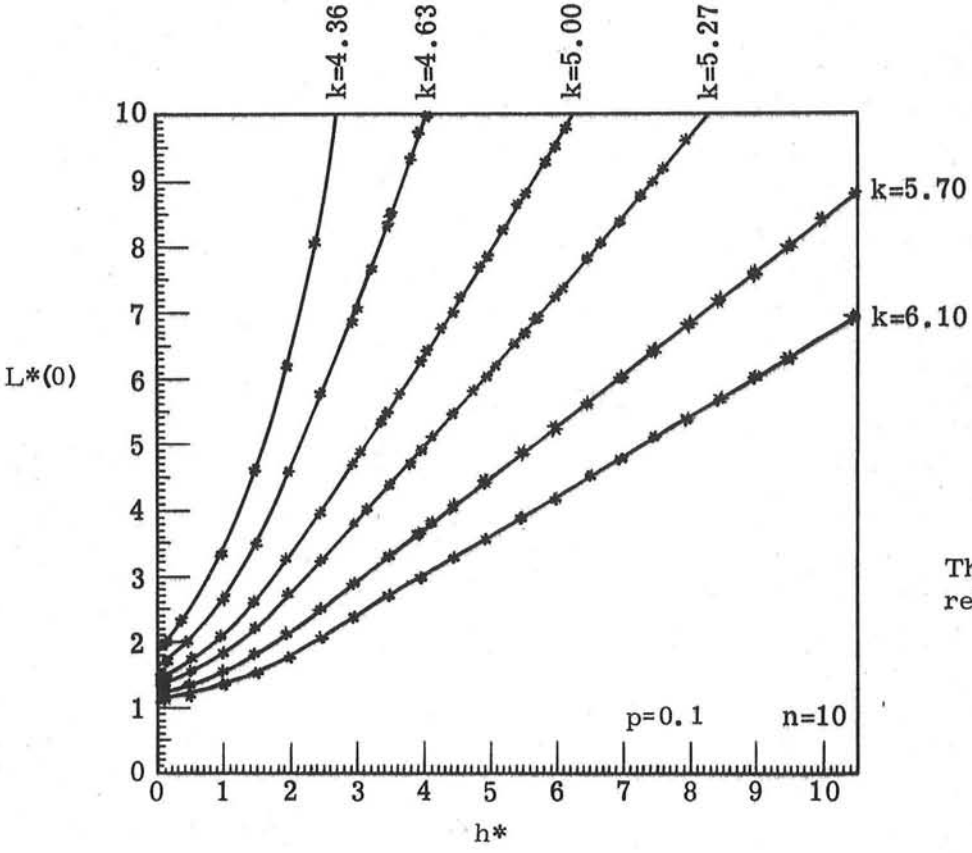

k=4.36
k=6.41
k=6.01
k=5.58
k=5.31
k=4.93
L(0)
h
p=0.07 n=10
The starred curves relate L*(0) and h*.
k=4.93
k=5.31
k=5.58
k=4.36
k=6.01
k=6.41
L*(0)
h*
p=0.07 n=10
The starred curves relate L(0) and h.

Figure 13.



The starred curves relate L*(0) and h*.



The starred curves relate L(0) and h.

Figure 14.

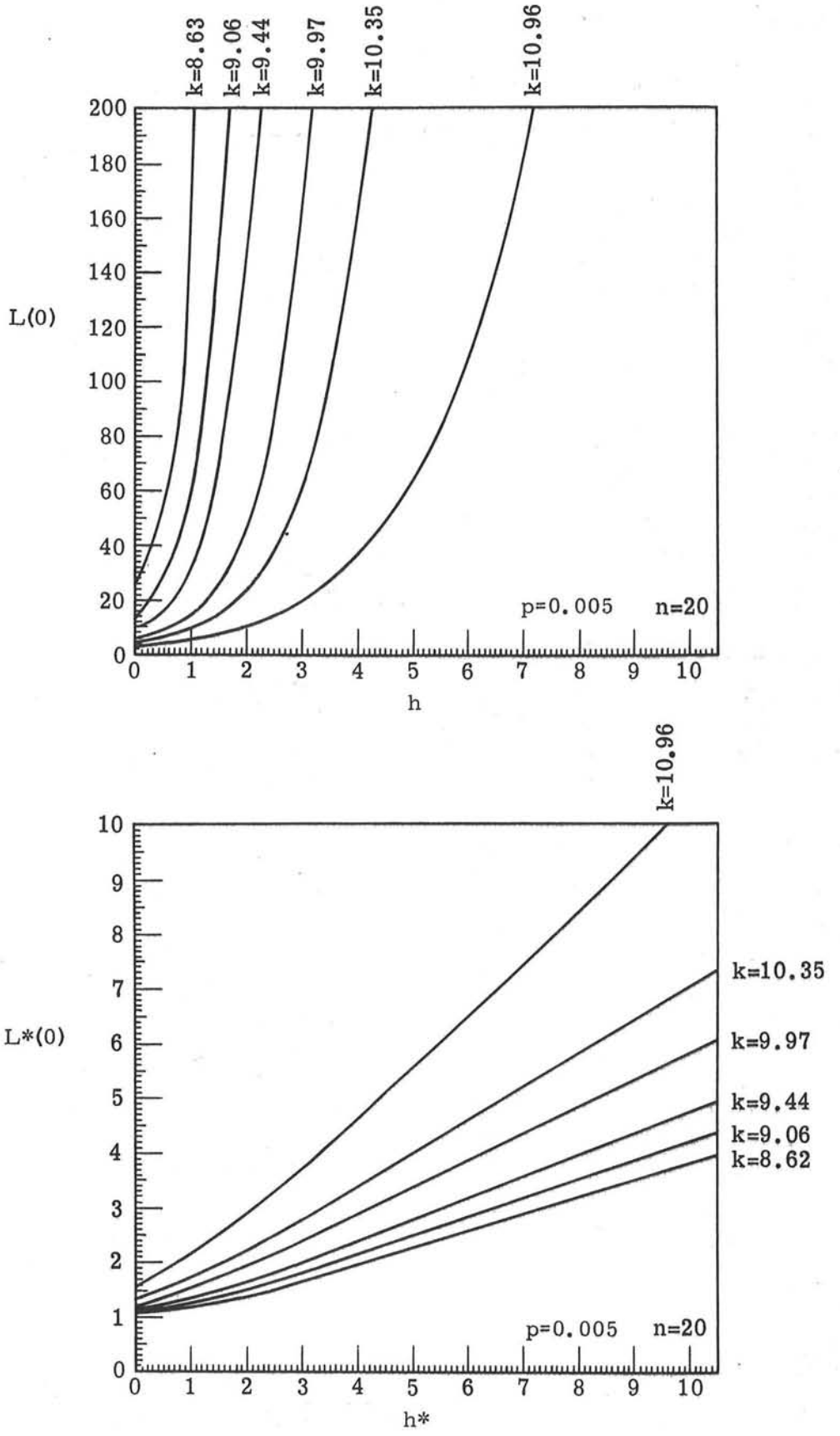
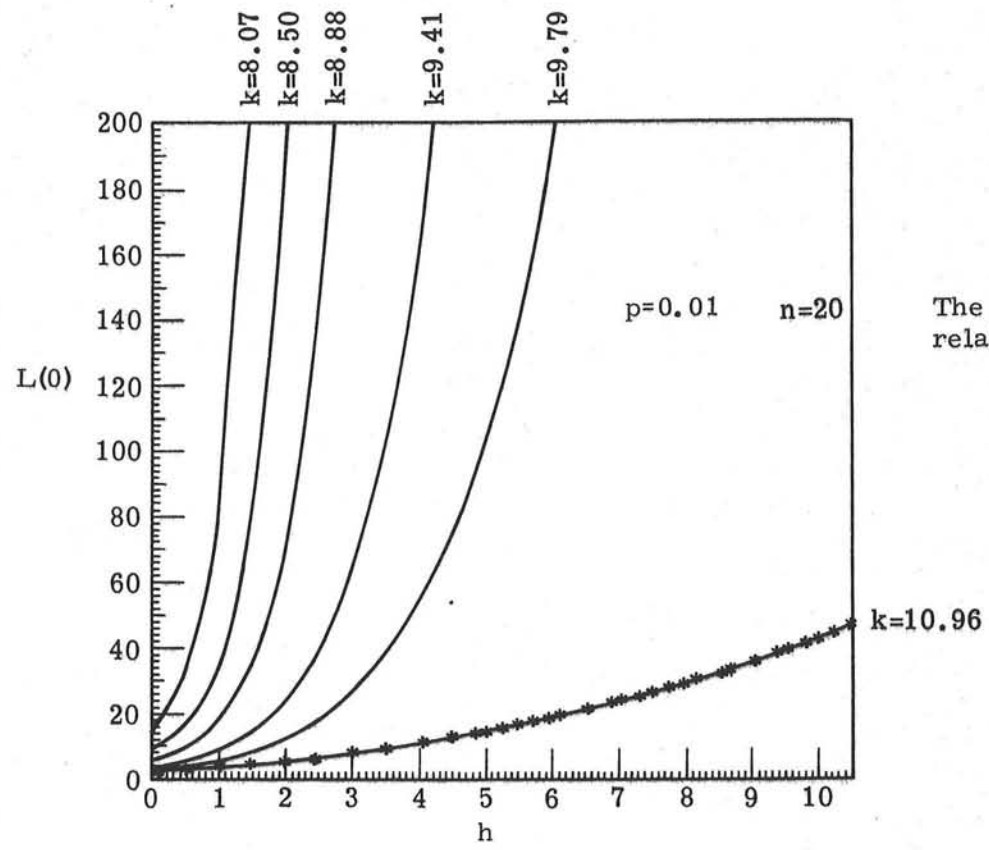
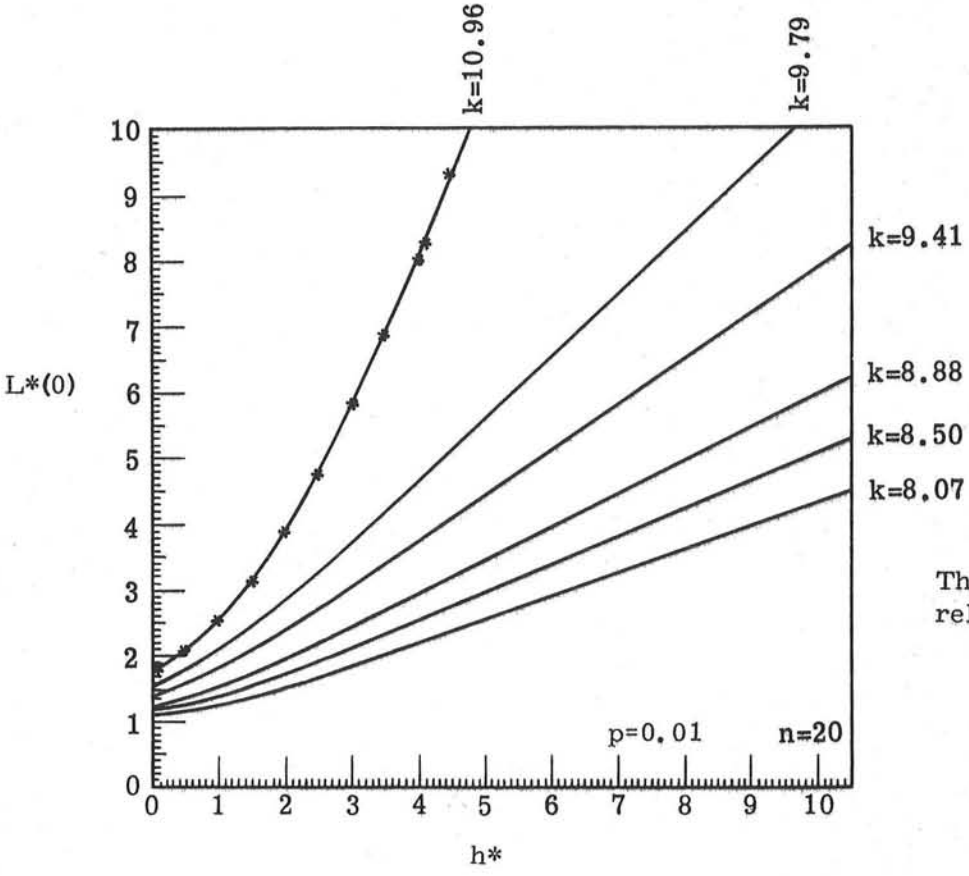

k=8.63
k=9.06
k=9.44
k=9.97
k=10.35
k=10.96
L(0)
h
p=0.005 n=20
k=10.96
k=10.35
k=9.97
k=9.44
k=9.06
k=8.62
L*(0)
h*
p=0.005 n=20

Figure 15.

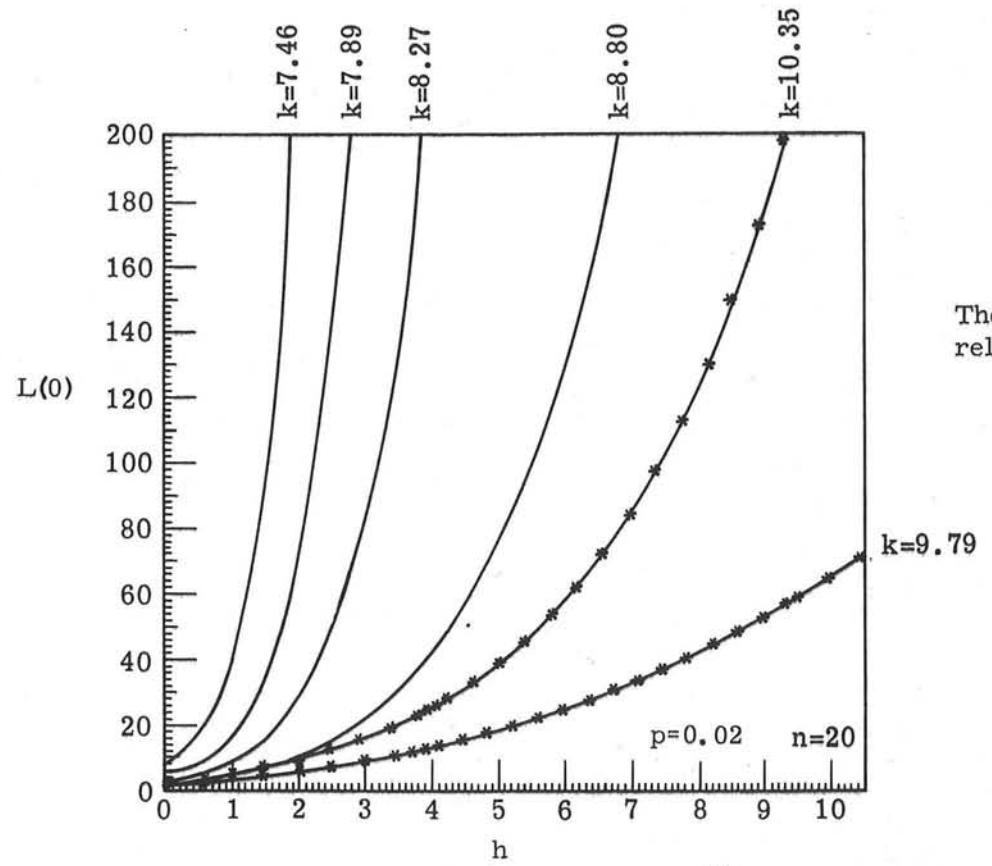


The starred curve relates L*(0) and h*.

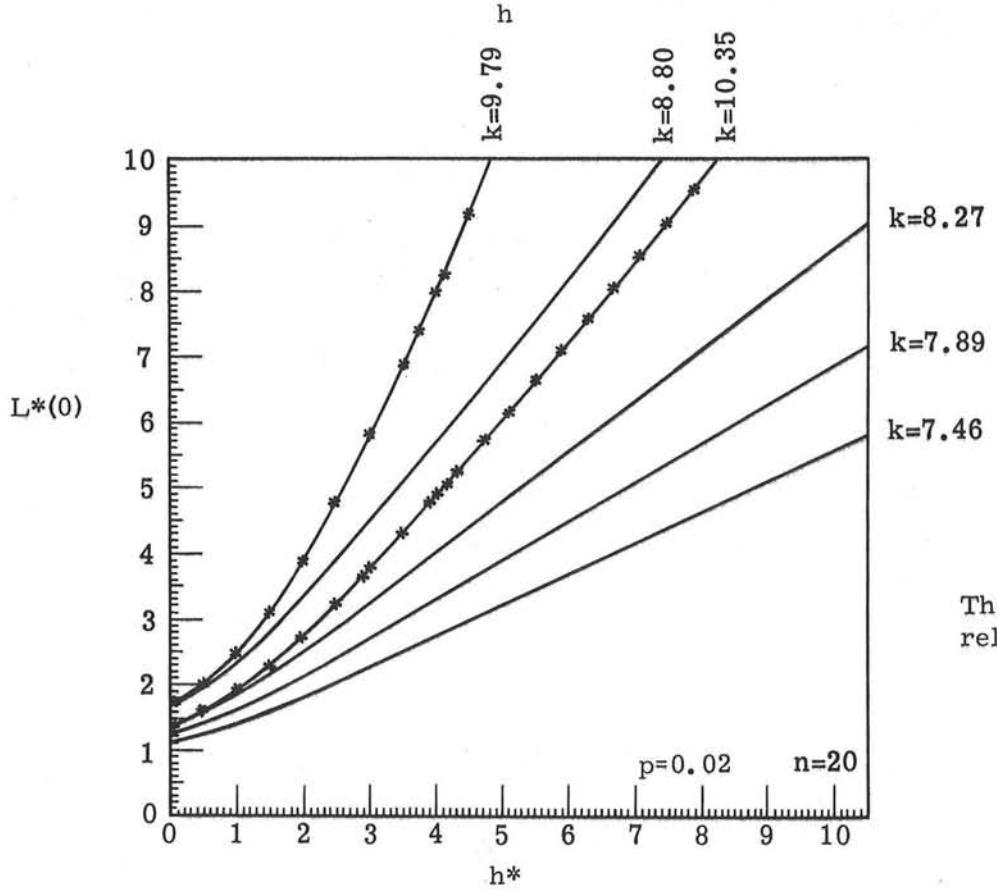


The starred curve relates L(0) and h.

Figure 16.

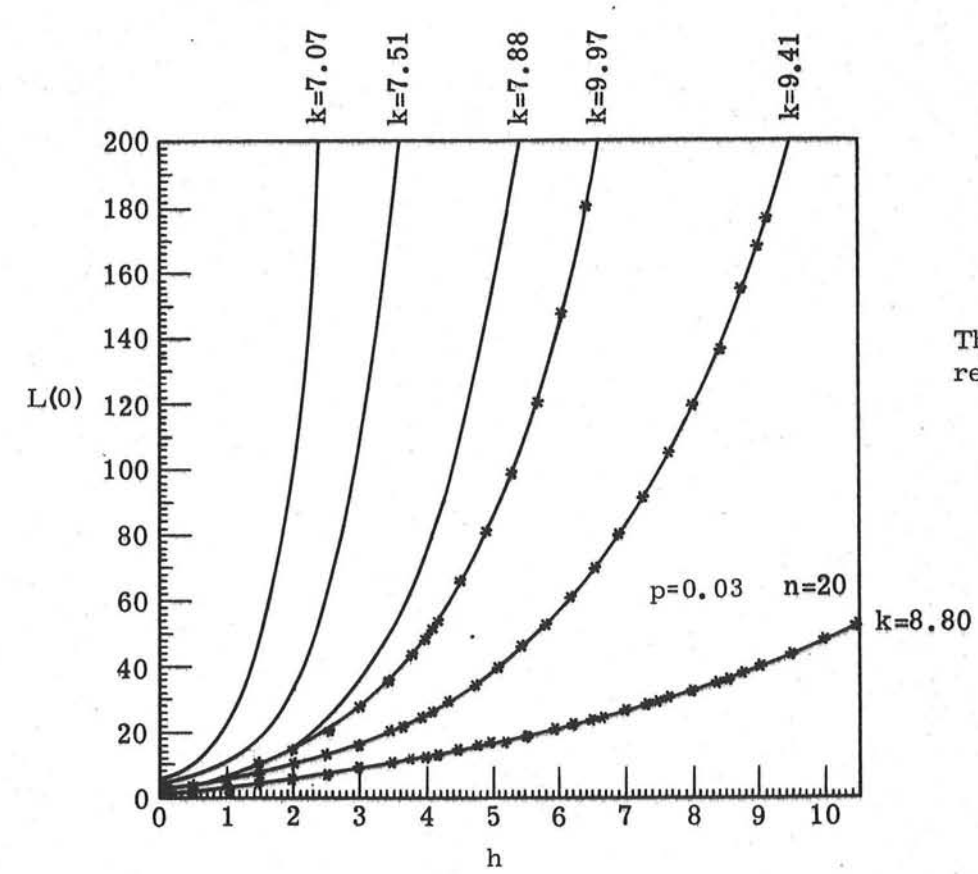


The starred curves relate L*(0) and h*.

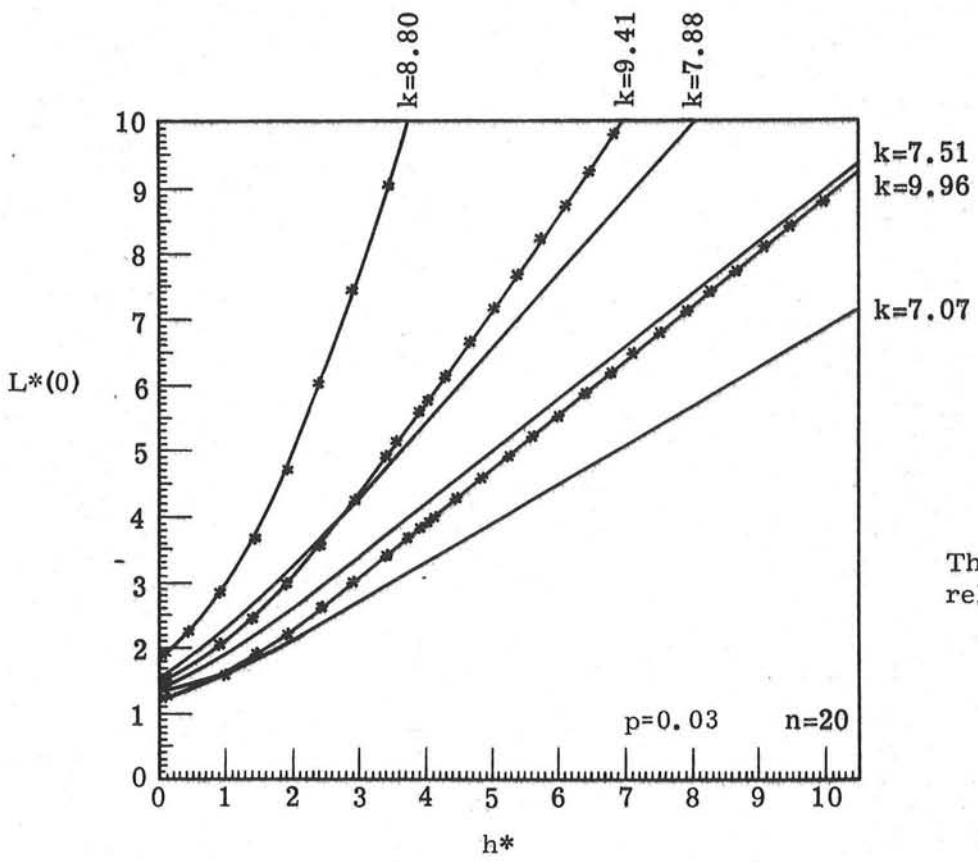


The starred curves relate L(0) and h.

Figure 17.

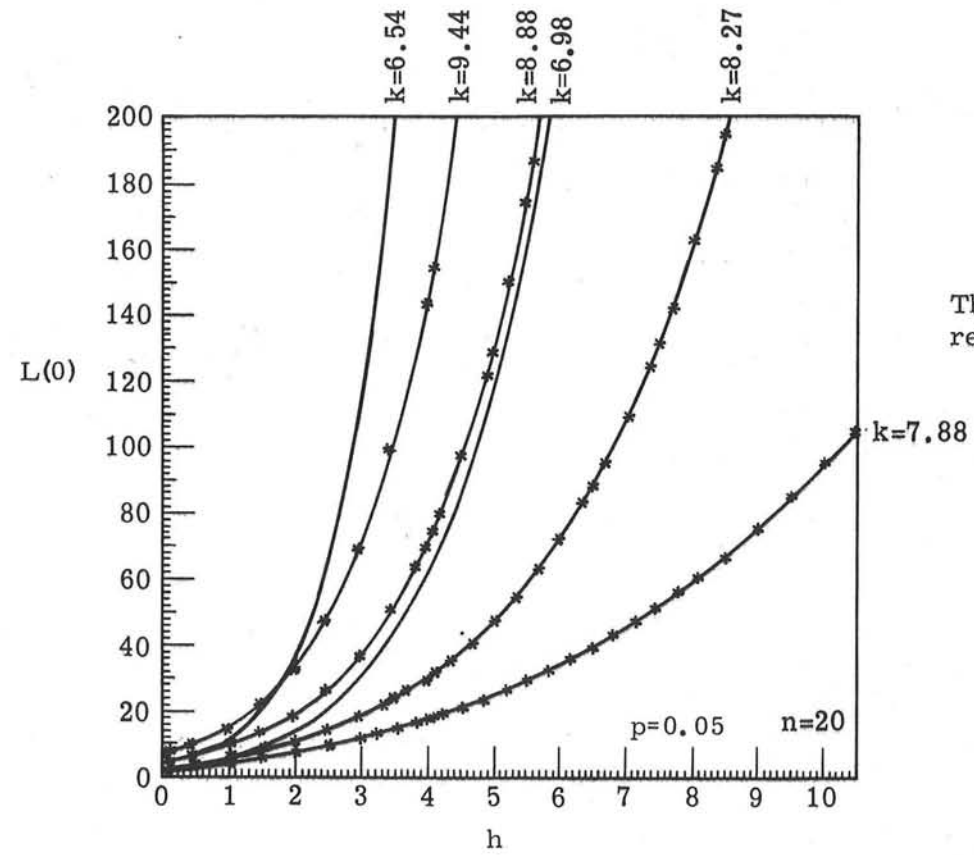


The starred curves relate L*(0) and h*.

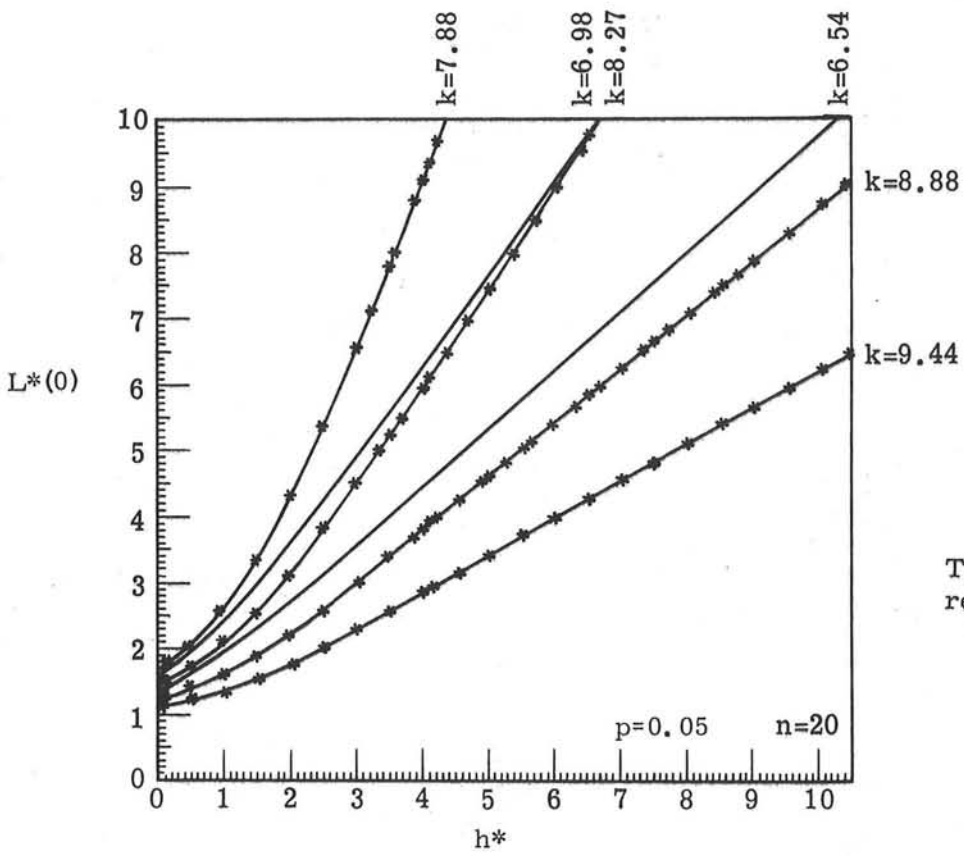


The starred curves relate L(0) and h.

Figure 18.

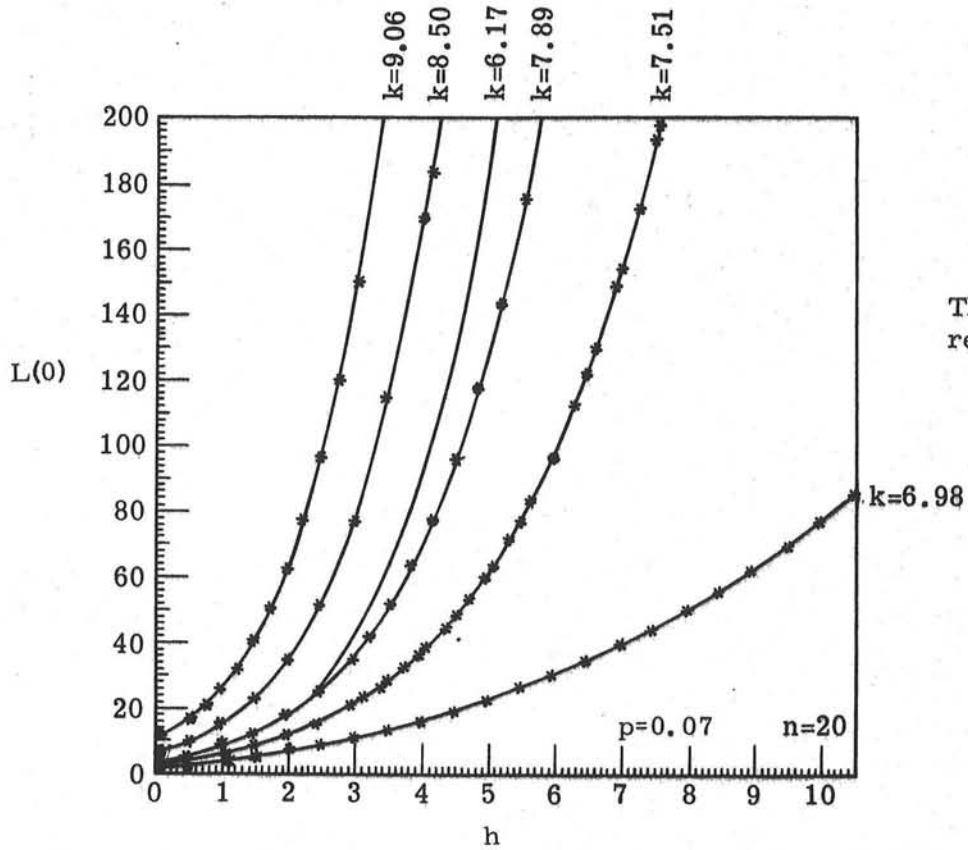


The starred curves relate L*(0) and h*.

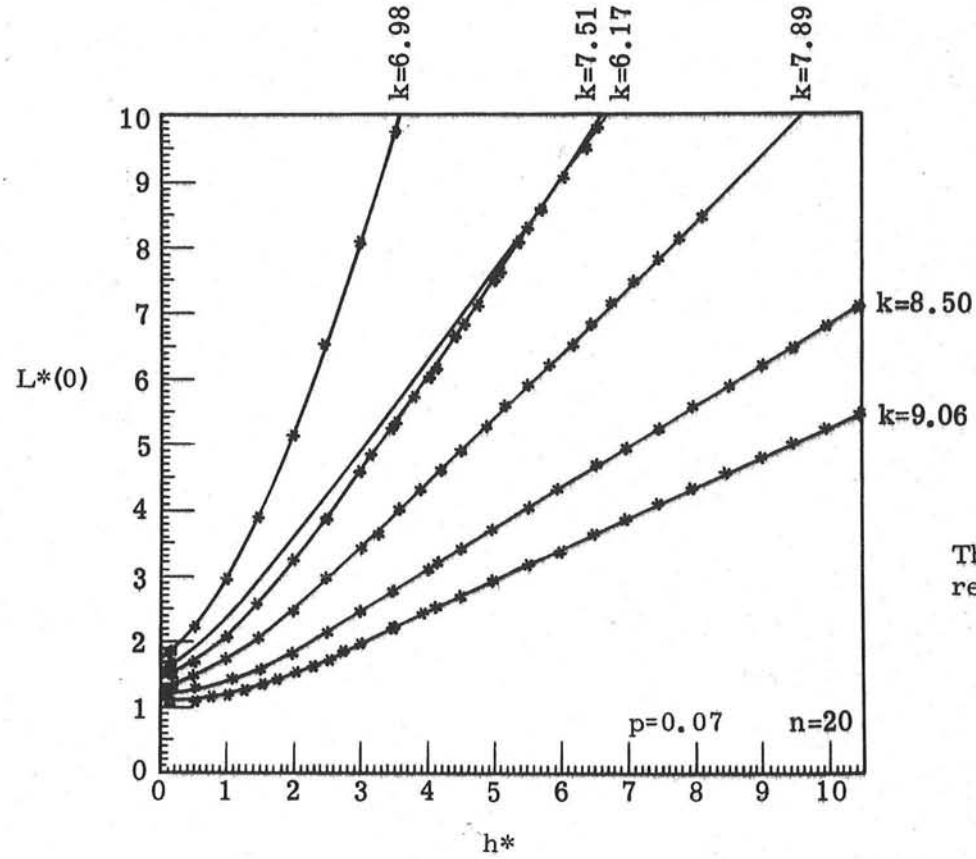


The starred curves relate L(0) and h.

Figure 19.

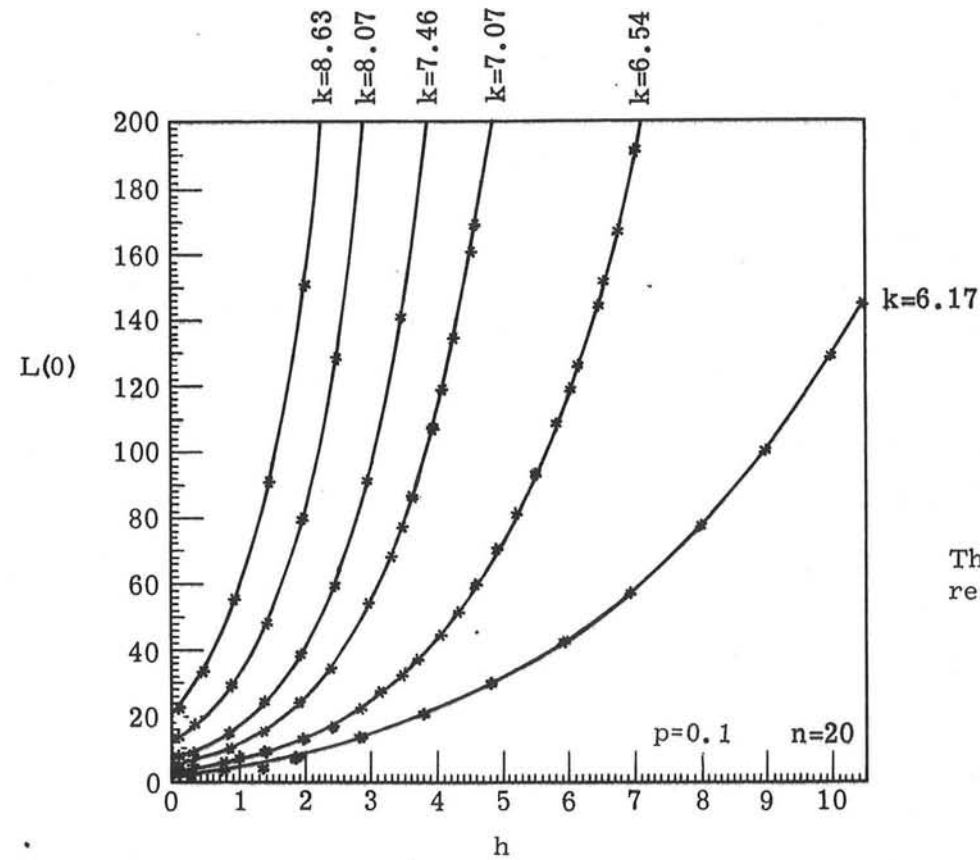


The starred curves relate L*(0) and h*.

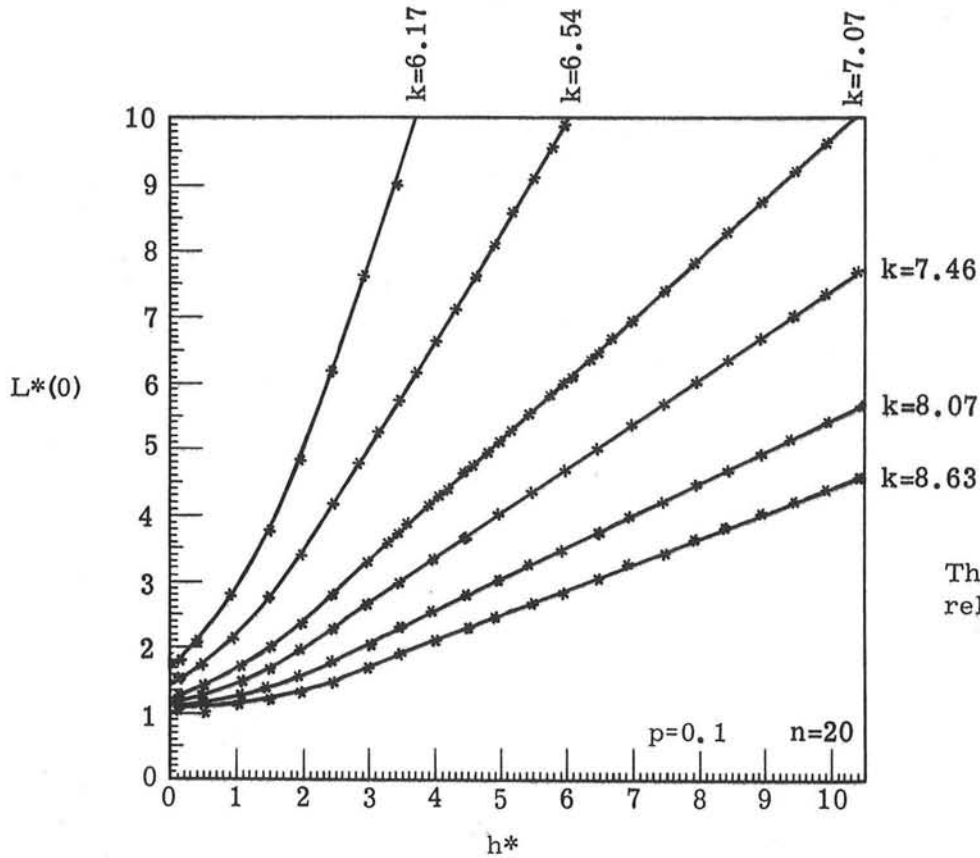


The starred curves relate L(0) and h.

Figure 20.



The starred curves relate to L*(0) and h*.



The starred curves relate to L(0) and h.

Figure 21.

DISTRIBUTION:
H. E. Lenander, 1600
J. M. Wiesen, 1640
H. E. Walker, 1641
R. E. Butler, 1642
M. C. Carter, 1643
C. R. Clark, 1643
R. G. Easterling, 1643
D. A. Edelman, 1643
I. J. Hall, 1643
G. J. Lombardi, 1643
N. R. Mohberg, 1643
W. H. Robertson, 1643
D. D. Sheldon, 1643
R. R. Prairie, 1643 (10)
T. M. Burford, 1700
J. W. Weihe, 1710
M. J. Norris, 1720
H. H. Wicke, 1721
R. J. Thompson, 1722
J. J. Deely, 1723
G. P. Steck, 1723
W. J. Zimmer, 1723 (10)
R. G. Clem, 1730
H. D. Sivinski, 1740
R. W. Devore, 7410
R. H. Schultz, 7420
A. P. Gruer, 7430
W. E. Boyes, 7450
O. H. Schreiber, 8139
B. F. Hefley, 8232 (5)
B. R. Allen, 3421
R. S. Gillespie, 3411
W. J. Wagoner, 3413 (DTIE) (4)
C. H. Sproul, 3428-2 (10)

uqo