»STATISTICAL INQUIRIES« Nr. 12
Issued by »The Statistical Department«, Denmark

An Analysis of the Personal Income Distribution for Wage and Salary Earners in 1955

By Kjeld Bjerke

KØBENHAVN
1964

This report, which is a continuation of *Statistiske Undersøgelser: »Lønmodtagerindkomster, fordeling og sammensætning«*, has been prepared by Kjeld Bjerke, chief of division.

The Statistical Department

In preparing this analysis of variance I have discussed some general problems with Professor, dr. phil. Anders Hald. I have also received excellent assistance from Mr. Karl Vind, lecturer, cand. polit. and Mr. Erik Holm, cand polit., assistant chief of section in this department. I am grateful for the valuable assistance I have received.

Kjeld Bjerke.

An Analysis of the Personal Income Distribution for Wage and Salary Earners in 1955

1. In *Statistiske Undersøgelser:* "Lønmodtagerindkomster, fordeling og sammensætning" an account has been given of the personal income distributions etc. of households and heads of households; this account has been based on a sample survey of the income, consumption and saving of wage and salary earners in 1955 for a few rather rough socio-economic groups.

On the basis of this material and by means of an analysis of variance we have wanted to illustrate the effect which social status, sex, age and geographical area may have on the wages or salary of the individual person. Further, we have studied how the difference between the actual wages or salary and the wages or salary in which only the mentioned factors are considered would be distributed. For if these factors, and perhaps the interaction among them, give a sufficiently exhaustive explanation of the total level of wages and salaries, it might be interesting to study whether this difference between actual and estimated wages and salary (the residual) would be a stochastic variable, which would perhaps be "normally distributed", if only after a transformation.

However, it is probably to be expected that the above-mentioned factors need not exhaust the effect of such factors. Training and education are factors for which no allowance is made direct. Nor is the composition of the household (with children, without children, etc.) included direct as a determinant; this applies also to the conditions of work and employment (including the wage system). But a number of these factors are not independent of the factors included in the calculations, so the residual will perhaps nevertheless show a certain approximation to "normal distribution".

As regards training and education no information is available, and regarding the composition of the household the influence will probably be greater when the household is the sampling unit than in a survey of the incomes of heads of households. In this connection it is also important that it has not been intended to include the total income of the head of household, but solely his income by way of wages or salary (from his main occupation), since the adoption of this procedure would to a greater extent lead us to expect that the difference between actual and estimated wages (salary) will reflect solely the more personal factors.

In the following report we have first examined the assumptions for using an analysis of variance. Thereafter, the parameters for the different factors, social status, sex, age, and geographical area, and their interaction have been estimated. The results have been tested, and finally, the aggregated distributions have been studied, especially the distributions of the residuals.

2. As mentioned, the material dealt with in the following originates from the survey of the income, expenditure and saving in households of wage and salary earners in 1955, comprising approximately 3100 households. The 3100 households were selected by two stages within the capital, provincial towns and rural districts with urban areas, respectively.

In the first stage, municipalities (taxation districts) were drawn at random from groups of uniform municipalities formed in advance, the probability of selection of each municipality (taxation district) being proportional to the number of households in the municipality. Actually, the selection should have been proportional to the number of wage- and salary-earning households, but this number was unknown, and it was estimated that the wage- and salary-earning households constituted a more or less constant share of the total number of households within the mentioned groups of municipalities. In the second stage, wage- and salary-earning households (sampling units) were drawn from each municipality in the sample of municipalities.

It should be mentioned that the definition of the sampling units selected—all participants in the joint consumption—did not quite correspond to the units of selection in the second stage, these having been laid down in the choice of frame for the survey, namely the questionnaires from the population census in 1955. Since, according to the definition adopted in the population census, the household comprises all persons with permanent residence in the joint dwelling with the exception of lodgers providing their own meals, whereas the household concept in the consumer survey comprises only the persons participating in the joint consumption to an extent of at least half the income earned, the household according to the population census will in certain cases be more comprehensive than the unit of the present analysis. This fact involves

certain complications in an exact blowing up of the results of the analysis to cover the whole country and will also be important in connection with an exact calculation of the standard deviation. However, these difficulties were not taken into account as it was assumed that the change which would result from an exact calculation would be insignificant compared to the inaccuracy which had arisen in the course of collecting and processing the questionnaires.

Since the income of the head of household rather than that of the household is the object of this analysis, the selection can no longer be considered a cluster selection. This is of importance to the following calculations, which become less complicated under this assumption.

In the selection of the number of households it was attempted to achieve an error of the total expenditure (incl. saving) per household of the same order as that for the twelve socio-economic groups into which the material was divided, and the result as regards the number of households selected within the individual socio-economic groups was as follows.

Planned and Final Number of Households in the Survey.

	Planned number of households	Number of households in the final material	Error of the average total expenditure ± per cent
Capital with suburbs:			
Higher salaried employees and public servants.......	400	336	3.7
Lower – – – – –	425	469	3.6
Skilled workers in urban industries................	225	206	4.4
Unskilled – – – –	275	251	5.6
Provincial towns with suburbs:			
Higher salaried employees and public servants.......	275	212	5.0
Lower – – – – –	300	341	4.2
Skilled workers in urban industries................	175	154	5.8
Unskilled – – – –	225	213	5.4
Rural districts with urban areas:			
Lower salaried employees and public servants........	375	322	4.0
Skilled workers in urban industries................	200	155	4.5
Unskilled – – – –	275	281	4.1
Agricultural labourers............................	150	160	5.0
Households of wage and salary earners, total........	3300	3100	

The number of households within each of the twelve groups of wage and salary earners is fixed so that there is 95 per cent probability that the deviation between the group's actual—but unknown—average expenditure and the average expenditure found in the survey will be less than approx. 5 per cent.

The material is not *self-weighting* as it has not been proportionally selected, and it is not *orthogonally* distributed.

The study of the error of the *average expenditure* will give a rough impression of the error of the *total average income*, which is probably of the same order as that of the average expenditure.

3. It has already been mentioned that it was the intention to use the income from wages or salary of the head of household for the purpose of the analysis. This income comprises income from main and subsidiary occupations; but for the purpose of the present survey solely the wage-(salary-)income from the main occupation has been used since the connection between social status and this income must be most marked.

Some of the workers were unemployed in 1955. We have not tried to adjust for this fact; however, it has been found most correct not to include workers with long periods of unemployment, i.e. workers receiving more than kr. 500 by way of unemployment benefits. An examination of the material showed that the number of long-term unemployed was as low as 85.

For skilled workers in the building trades, however, the unemployment benefit has been added to the wage-income from the point of view that seasonal unemployment is a normal thing in these trades. For this reason the group of skilled workers has been subdivided into two groups, viz. skilled workers in the building trades and other skilled workers. This subdivision had to be made on the basis of information as to the unemployment insurance fund to which the workers in question belonged; estimates have thus been made separately for skilled bricklayers, carpenters, painters, plumbers and electricians.

In the case of unskilled workers a subdivision has also been made from the belief that a distinction between the members of the Federation of Unskilled Labourers and the other trades would lead to somewhat more homogeneous groups.

Finally, it may be mentioned that in the case of lower salaried employees and public servants a subdivision has been made into men and women. The social status groups distinguished in the survey are accordingly the following: Higher salaried employees and public servants; Lower salaried employees and public servants, broken down into men and women; Skilled workers, divided into building trades and other trades—and finally Unskilled workers divided into members of the Federation of Unskilled Labourers and others. It is obvious that there is correlation between the distribution of the mentioned groups and the income; but for the subdivision the income has not been used as the criterion—except to a small extent in connection with the distinction between higher and lower salaried employees.

As information about higher salaried employees and public servants is available only for the capital and the provincial towns, only these two geographical areas have been included. As we have agricultural labourers only in the rural districts they have not been included in the survey.

While the three factors of social status group, sex and geographical area are non-numerical values, age can be expressed numerically. However, as in the case of social status groups and geographical areas we have preferred to operate with age-groups rather than with the individual statements of age. There is also another reason why age has not been treated as a non-numerical variation, namely that there is no linear correlation between age and income. The following age distribution has been used: below 25 years of age, 25–34, 35–44, 45–59, 60 years of age and over.

It has already been mentioned that it might be an advantage to include more factors, but that the size of the material sets limits to the number of criteria of subdivision which should be used.

4. In the analysis of variance it has been assumed, for instance, that a certain social status exercises a certain influence, and that this influence on the wage (salary) level is the same for each person belonging to the mentioned social status group. If, as in the present survey, the subdivision by social status is fairly rough, the result will be a limitation of the validity of the mentioned assumption, but the more homogeneous the subgroups of the material become, the greater is the possibility that the mentioned assumption will hold good. Similar considerations apply to age and geographical area.

To illustrate the procedure adopted in the analysis, it may be imagined that there are only two subdivision criteria, e.g. age and social status, and that each criterion is divided into three groups, compare the subsequent table.

Here μ is an expression of the expected income for the whole population. If it is desired to determine average incomes in which allowance is made for the criteria of subdivision, α_2, for instance, must be added to μ in the case of a person in the second age group. If, moreover, he belongs to the third social group, the parameter β_3 must be added. In this way we arrive at the

Age (i)	Social status (j)				
	Number of observations			Total	Parameters
	1	2	3		
1....................	n_{11}	n_{12}	n_{13}	$N_{1.}$	α_1
2....................	n_{21}	n_{22}	n_{23}	$N_{2.}$	α_2
3....................	n_{31}	n_{32}	n_{33}	$N_{3.}$	α_3
Total...	$N_{.1}$	$N_{.2}$	$N_{.3}$	$N_{..}$	
Parameters...........	β_1	β_2	β_3		μ

expected income in the second age group and third social group. If it is the ith age group and the jth social group, α_i and β_j must consequently be added. The kth observation (income) among the n_{ij} observations we have here can deviate by a certain (possibly) random quantity e_{ijk} from the expected quantity, which is the same for all n_{ij} observations. This quantity, e_{ijk} may also be added. The following model will thus be set up

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$$
$$i = 1, 2, \ldots\ldots r$$
$$j = 1, 2, \ldots\ldots s$$
$$k = 1, 2, \ldots\ldots n_{ij}$$

where μ, the α's and β's are parameters and the e's must be imagined to be normally distributed around 0 and with a standard deviation of the same size in all ij subgroups.

The model set up assumes that there is no interaction among the different subdivision criteria. In the present example with the two variables of age and social status this assumption will not hold good. It is therefore necessary to allow for this fact by introducing into the model a parameter for this interaction.

If we have an arbitrary age group, i, and an arbitrary social status group, j, and we again consider the person (A's) income, the model will take the following form:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$
$$i = 1 \ldots. r; \quad j = 1 \ldots. s; \quad k = 1 \ldots. n_{ij}$$

where the new term $(\alpha\beta)_{ij}$ signifies a parameter which depends on the power of interaction between age and social status. With regard to the determination of estimates of the parameters, the following comments should be made:

The least-squares method is used in determining the parameters and therefore the following quantities are to be minimised with regard to each of the parameters; it should be noted that the number of observations in the individual cells is not the same. Thus the distribution is not orthogonal. In the expression the theoretical values μ, α, and β have been replaced by the estimated values m, a, and b.

$$\sum_i^r \sum_j^s \sum_k^{n_{ij}} (y_{ijk} - m - a_i - b_j - (ab)_{ij})^2$$

Differentiation gives the following result:

$$-2 \sum_i^r \sum_j^s \sum_k^{n_{ij}} (y_{ijk} - m - a_i - b_j - (ab)_{ij}) = 0 \qquad (m)\ (1)$$

$$-2 \sum_j^s \sum_k^{n_{ij}} (y_{ijk} - m - a_i - b_j - (ab)_{ij}) = 0 \qquad (a_i)\ (2)$$

$$-2 \sum_i^r \sum_k^{n_{ij}} (y_{ijk} - m - a_i - b_j - (ab)_{ij}) = 0 \qquad (b_j)\ (3)$$

$$-2 \sum_k^{n_{ij}} (y_{ijk} - m - a_i - b_j - (ab)_{ij}) = 0 \qquad (ab_{ij})\ (4)$$

It will be seen that (4) contains the previous three systems since each of them can be produced from (4) through a suitable summation. (4) can be transcribed as follows.

$$\sum_k^{n_{ij}} y_{ijk} - n_{ij}\, m - n_{ij}\, a_i - n_{ij}\, b_j - n_{ij}\, (ab)_{ij} = 0$$

or by division by n_{ij}

$$m + a_i + b_j + (ab)_{ij} = \sum_{k}^{n_{ij}} \frac{y_{ijk}}{n_{ij}} = \bar{y}_{ij}$$

This, then, is the central condition.

In (4) there are $(r \cdot s)$ equations, but $1 + r + s + r \cdot s$ unknowns. Therefore the system is not determined. This, however, can be achieved by adding the following constraints.

$$\sum_{i}^{r} a_i = 0; \sum_{j}^{s} b_j = 0; \sum_{i}^{r} (ab)_{ij} = 0; \sum_{j}^{s} (ab)_{ij} = 0$$

Here there are $1 + 1 + s + r - 1 = r + s + 1$ constraints, so the system is now determined.

The central condition in (4) will be fulfilled if, since the material is not orthogonal, estimates of the following parameters are formed from the *unweighted* averages:

$$m = y_{..}$$
$$a_i = y_{i.} - y_{..}$$
$$b_j = y_{.j} - y_{..}$$
$$(ab)_{ij} = \bar{y}_{ij} - y_{i.} - y_{.j} + y_{..}$$ [1]

[1] These parameters are described as follows:

$$\bar{y}_{ij} = \frac{1}{n_{ij}} \sum_{k}^{n_{ij}} y_{ijk}$$

$$y_{.j} = \frac{1}{r} \sum_{i}^{r} \bar{y}_{ij}$$

$$y_{i.} = \frac{1}{s} \sum_{j}^{s} \bar{y}_{ij}$$

$$y_{..} = \frac{1}{s} \sum_{j}^{s} y_{.j} = \frac{1}{r} \sum_{i}^{r} y_{i.}$$

These parameters fulfil the constraints set up above.[1])

In the example mentioned here with three age groups and three social groups 9 interaction parameters $(ab)_{ij}$ are possible.

Since income or wages (salary) is the independent variable there may be some reason to assume, at least for more exacting work, that the quantities do not interact additively, but multiplicatively. The point of view is that wages (salary) or income is imagined to be determined by the influence of a great number of mutually independent factors with the property of increasing or reducing income in some proportion. (The law of proportionate effect). By considering the logarithms, the effect will be expressed additively and according to the Central Limit Theorem, we may expect, therefore to get normal distributions for this sum of mutually independent variables.[2]) It follows also that if the logarithm of the mentioned variables is approximately normally distributed, the variable itself will give a distribution with a positive skewness (the long tail to the right). The model will take the following form:

$$y_{ijk} = \mu \cdot \alpha_i \cdot \beta_j \cdot (\alpha\beta)_{ij} \cdot e_{ijk}$$

or if we take the logarithm

$$\log y_{ijk} = \log \mu + \log \alpha_i + \log \beta_j + \log (\alpha\beta)_{ij} + \log e_{ijk}$$

6. In the concrete analysis it was decided to use the assumption of both additive and multiplicative effect.

The volume of calculations will depend on the number of factors to be included and the number of interactions which it may be considered necessary to include.

The solution of the task has been based on the following assumptions: Three factors are taken into account: social status, geographical area, and age. Social status comprised 7 groups, geographical area 2 groups, and age 5 groups. This gives a total of 70 "cells". With regard to the interaction assumptions we have provisionally reckoned with all possible interactions.

Under these assumptions the number of parameters to be determined will come to a total of 144, namely 14 from the main effects, 59 from interactions of the first order, 70 from interactions of the second order, and 1 parameter

[1]) The system of equations of the analysis of variance was set up by Mr. Erik Holm after consultation with Mr. Karl Vind and myself.

[2]) The assumption of independence may be problematical.

Table 1a. *Test for Normality. Additive Model.*

Geographical area	Age	Social group													
		Skewness: $\sqrt{\beta_1}$-test							Kurtosis: *a*-test						
		1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, male	3 Lower sal. empl. and publ. servants, female	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed of Unsk. Lab.	7 Unskilled workers, other	1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, male	3 Lower sal. empl. and publ. servants, female	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed of Unsk. Lab.	7 Unskilled workers, other
The Capital	–24 years	(7)	(13) N	(18) vs	(1)	(12) N	(2)	(11) N	(7)	(13) N	(18) N	(1)	(12) N	(2)	(11) N
	25–34 years	(73) hs	(88) vs	(35) N	(9)	(37) N	(21) N	(25) N	(73) ts	(88) N	(35) N	(9)	(37) N	(21) N	(25) N
	35–44 years	(97) hs	(96) vs	(35) N	(18) vs	(36) hs	(28) N	(29) N	(97) ts	(96) ts	(35) N	(18) N	(36) ts	(28) ts	(29) N
	45–59 years	(125) hs	(73) vs	(60) N	(21) hs	(41) N	(43) N	(50) N	(125) N	(73) ts	(60) N	(21) N	(41) N	(43) N	(50) N
	60 years and over	(34) N	(36) N	(15) N	(8)	(19) N	(10) N	(14) N	(34) N	(36) N	(15) N	(8)	(19) N	(10) N	(14) N
Provincial towns	–24 years	(4)	(17) N	(17) N	(2)	(9)	(4)	(5)	(4)	(17) N	(17) N	(2)	(9)	(4)	(5)
	25–34 years	(60) hs	(75) N	(20) N	(20) N	(36) vs	(18) N	(17) N	(60) ts	(75) ts	(20) N	(20) N	(36) N	(18) N	(17) N
	35–44 years	(63) hs	(80) vs	(17) vs	(9)	(10) N	(37) N	(14) N	(63) ts	(80) ts	(17) ts	(9)	(10) N	(37) N	(14) N
	45–59 years	(73) hs	(69) vs	(22) N	(15) N	(30) N	(31) hs	(29) N	(73) N	(69) ts	(22) N	(15) N	(30) N	(31) ts	(29) N
	60 years and over	(12) N	(20) vs	(3)	(7)	(2)	(7)	(3)	(12) N	(20) N	(3)	(7)	(2)	(7)	(3)

Note: N indicates normal distribution, *hs* positive skewness, *vs* negative skewness, *ts* leptokurtosis, and *tf* platykurtosis.
The number of observations has been shown in brackets.

Table 1b. *Test for Normality. Multiplicative Model.*

Geographical area	Age	Social group													
		Skewness: $\sqrt{\beta_1}$-test							Kurtosis: *a*-test						
		1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, male	3 Lower sal. empl. and publ. servants, female	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed of Unsk. Lab.	7 Unskilled workers, other	1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, male	3 Lower sal. empl. and publ. servants, female	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed of Unsk. Lab.	7 Unskilled workers, other
The Capital	–24 years	(7)	(13) vs	(18) vs	(1)	(12) N	(2)	(11) N	(7)	(13) N	(18) ts	(1)	(12) N	(2)	(11) N
	25–34 years	(73) N	(88) vs	(35) N	(9)	(37) N	(21) vs	(25) N	(73) ts	(88) ts	(35) N	(9)	(37) N	(21) N	(25) N
	35–44 years	(97) hs	(96) vs	(35) vs	(18) vs	(36) hs	(28) N	(29) N	(97) N	(96) ts	(35) ts	(18) ts	(36) ts	(28) ts	(29) N
	45–59 years	(125) N	(73) vs	(60) vs	(21) N	(41) vs	(43) vs	(50) vs	(125) N	(73) ts	(60) N	(21) N	(41) N	(43) ts	(50) N
	60 year and over	(34) N	(36) vs	(15) N	(8)	(19) vs	(10) N	(14) vs	(34) N	(36) ts	(15) N	(8)	(19) ts	(10) N	(14) ts
Provincial towns	–24 years	(4)	(17) N	(17) N	(2)	(9)	(4)	(5)	(4)	(17) N	(17) N	(2)	(9)	(4)	(5)
	25–34 years	(60) hs	(75) vs	(20) vs	(20) vs	(36) vs	(18) N	(17) vs	(60) ts	(75) ts	(20) N	(20) N	(36) ts	(18) N	(17) N
	35–44 years	(63) hs	(80) vs	(17) vs	(9)	(10) N	(37) vs.	(14) N	(63) N	(80) ts	(17) ts	(9)	(10) N	(37) N	(14) N
	45–59 years	(73) hs	(69) vs	(22) vs	(15) N	(30) vs	(31) vs	(29) vs	(73) N	(69) ts	(22) N	(15) N	(30) N	(31) ts	(29) N
	60 years and over	(12) N	(20) vs	(3)	(7)	(2)	(7)	(3)	(12) N	(20) ts	(3)	(7)	(2)	(7)	(3)

Note: N indicates normal distribution, *hs* positive skewness, *vs* negative skewness, *ts* leptokurtosis, and *tf* platykurtosis.
The number of observations has been shown in brackets.

for the income level for the whole population. The additive form of the model will accordingly be as follows:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + e_{ijkl}$$

Here μ = an unweighted average of "cell averages".
α = contribution to wages (salary) owing to social status.
β = contribution to wages (salary) owing to geographical area.
γ = contribution to wages (salary) owing to age.

The subsequent terms are interaction parameters and "*e*", finally, is a stochastically varying quantity with the expectation of 0.

7. It has been mentioned that the material comprises only the capital and provincial towns and that certain persons with long periods of unemployment have not been included. A total of 2097 wage (salary) data have been used.

This material has been processed at The Danish Institute of Computing Machinery, and the first problems we have tried to throw light on for the purpose of the analysis of variance are how the frequency distributions within the 70 cells have turned out, and also the size of the standard deviations within the individual cells. The calculations have been made both under additive and multiplicative assumptions.

The assumptions for carrying through the analysis of variance must be that the frequency distributions within the individual cells are approximately normal, and that the standard deviation within the cell does not deviate significantly.

The first assumption has been examined by means of tests for skewness and humpedness (kurtosis).

For skewness the usual $\sqrt{\beta_1}$ test (the moment of the third order) has been used; but for kurtosis a $\sqrt{\beta_2}$ test (the moment of the fourth order) has not been used because this test is not very suitable for less than 200 observations. Instead we have used an "a" test set up by R. C. Geary which can be used for as little as 40 observations.[1]).

[1]) This test may be expressed as follows

$$A = \frac{\sum_{i=1}^{n} (x_i \div \bar{x})}{\sqrt{n \sum (x_i \div \bar{x})^2}}$$

cf. E. S. Pearson and H. C. Hartley: Biometrika Tables for Statisticans Vol. 1, Cambridge 1954.

For the individual social groups broken down by age for the Capital and provincial towns, respectively, tables 1a and 1b show the results of an examination of the extent to which the frequency distributions within the individual cells may be said to be normal under assumptions of an additive and a multiplicative model, respectively. Appendix 1a and b show number of observations, means and standard deviations within cells for the two models. It will be seen that in the case of the additive model there is a greater tendency towards normality than in the case of the multiplicative model; it should be noted that the 99 per cent limit has been used.

The tendency towards normal distribution in the case of the additive model is especially clear in connection with female lower salaried employees, skilled workers, and unskilled workers. On the other hand, the frequency distributions for higher salaried employees, lower male salaried employees, and building workers seem to be skew and on the whole leptokurtic. As was to be expected, in accordance with the above remarks concerning the interaction of abilities, it is interesting to note that the multiplicative model seems to be "better" for higher salaried employees and public servants than the additive one.

For cells with at least 18 observations the connection between significance and non-significance is as follows:

Additive model

a \ $\sqrt{\beta_1}$	*s*	*ns*	
s	10	6	16
ns	2	19	21
	12	25	37

We have gone down to approx. 20 observations although the "a" test is suitable only for at least 40 observations, so statements for between 20 and 40 observations are uncertain.

For 17 groups with between 10 and 18 observations, 16 were non-significant (ns) for "a" test and only 1 deviated significantly. With so few observations this statement is extremely uncertain.

It has been mentioned that for the multiplicative model the agreement is not nearly so god. For cells with at least 18 observations the connection between significance and non-significance is as follows:

Multiplicative model

a \ $\sqrt{\beta_1}$	*s*	*ns*	
s............	15	12	27
ns...........	2	8	10
	17	20	37

For the above-mentioned 17 cells with few observations (between 10 and 18) 5 deviated significantly and 12 non-significantly. As mentioned, the statement is uncertain.

Graphs 1 and 2 show for the 70 cells the connection between means and standard deviations under the two assumptions.

As regards the assumption concerning the standard deviation a Bartlett's test shows that the standard deviations deviate significantly, both in the additive and in the multiplicative model, but in the additive model, cf. graph 1, it is especially the standard deviation for higher salaried employees which contributes to the standard deviations being significantly deviating. Apart from this group there does not on the whole seem to be any marked correlation between mean and standard deviation.

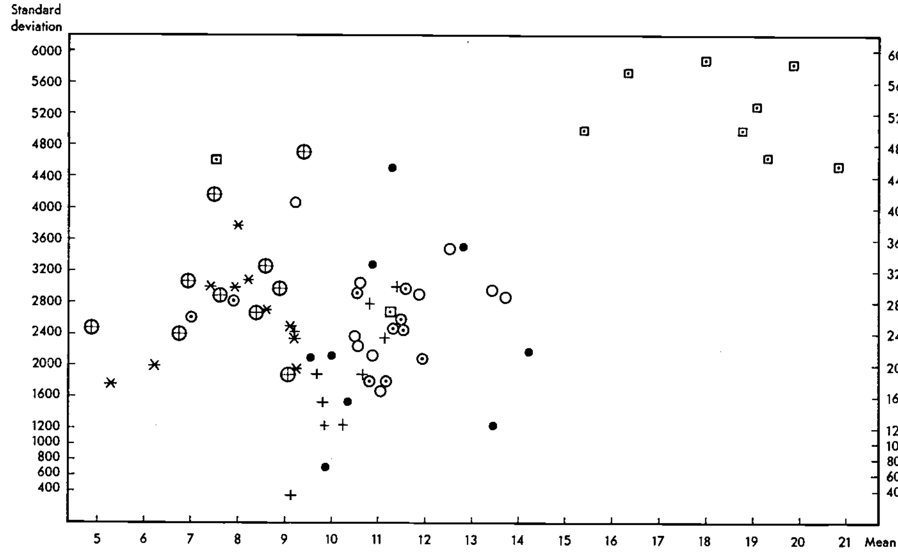
In the case of the connection between mean and standard deviation in the multiplicative model there seems to be a certain tendency towards negative correlation.

A similar impression may be gained by studying the average standard deviation within certain groups of average income distributed by size. The main results will be seen from the following table 2.

For the additive model it will be seen how the level of standard deviations is fairly the same up to kr. 15,000. It is only in the following group that higher employees cause the standard deviations to become appreciably higher.

For the multiplicative model the table shows that the standard deviation for the lowest two income groups is at an appreciably higher level than for the following three groups.

Graph 1. *Comparison Between Means and Standard Errors. Additive model.*



Graph 2. *Comparison Between Means and Standard Errors. Multiplicative Model.*

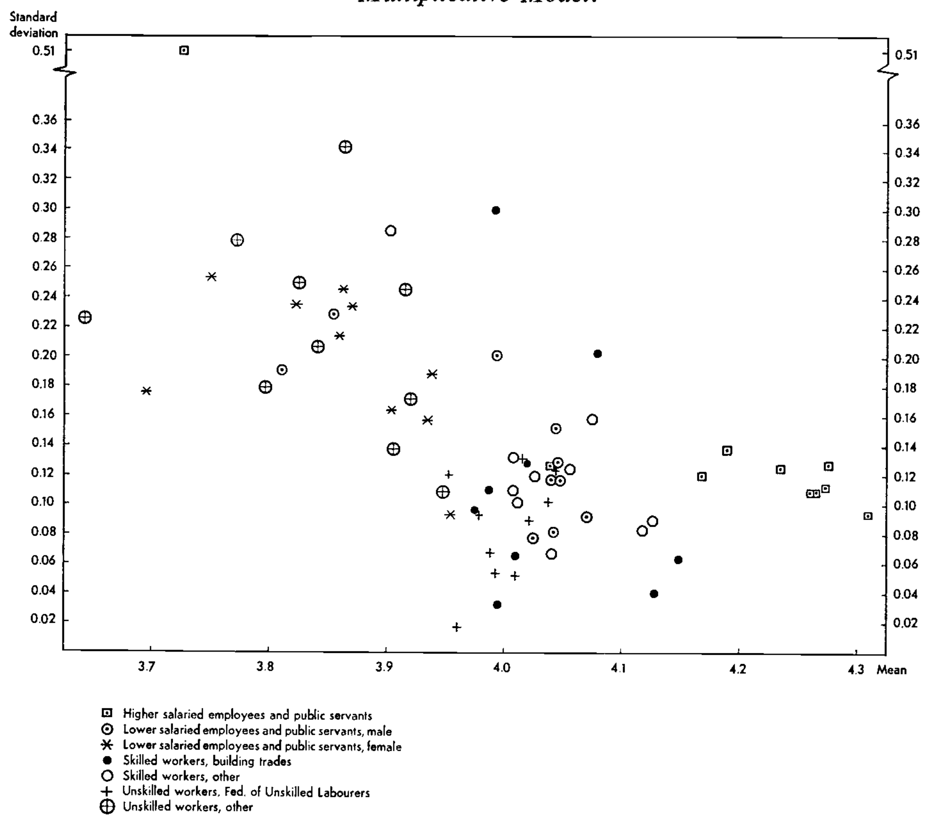


Table 2. *Dependence on Income of Standard Deviation.*

Expected income	Additive model	Multiplikative model
0– 7,999	2.780	0.232
8,000– 9,999	2.680	0.201
10,000–11,999	2.380	0.113
12,000–14,999	2.920	0.109
15,000 and over	5.330	0.120

It has been shown that while the conditions for normal distribution seem to be fulfilled quite well for the additive model, these conditions are poorly fulfilled in the case of the multiplicative model, apart from higher salaried employees and public servants.

With regard to standard deviation this group is significantly deviating, and there seem to be certain characteristic differences in level when this group is distributed by size of income. However, as there seems to be no marked tendencies for the standard deviation to grow disproportionately with income, we have proceeded with the study and tested the results.

8. For the complete additive and multiplicative models[1]), respectively, it has been examined for the purpose of the analysis of variance what interactions among the different factors influence wages (salaries). The result of this examination has been given in table 3.

Table 3. *Complete Model.*

Hypothesis	Degrees of freedom		F-test		Fractiles of the F-distribution							
	f_1	f_2	Additive model	Multiplicative model	50 pct.	70 pct.	90 pct.	95 pct.	97.5 pct.	99 pct.	99.5 pct.	99.95 pct.
RSA = 0	24	2027	0.58	1.31	0.97	.	1.40	1.52	.	.	.	.
RSA = 0 SA = 0	24	2051	5.40	4.44	.	.	.	.	.	.	.	2.30
RSA = 0 RA = 0	4	2051	0.86	1.86	0.84	1.22	1.94	.	.	.	.	.
RSA = 0 RS = 0	6	2051	3.01	2.55	.	.	.	.	2.42	2.81	3.10	.
RSA = 0 RA = 0 RS = 0	10	2051	2.10	2.37	.	.	.	.	2.05	2.32	2.52	.

Note: $Y_{ijkt} = M + R_i + S_j + A_k + (RS)_{ij} + (RA)_{ik} + (SA)_{jk} + (RSA)_{ijk} + e_{ijkt}$
R: geographical area ($i = 1,2$); S: social group ($j = 1,2,3,4,5,6,7$); A: age group ($k = 1,2,3,4,5$).

[1]) I.e. the model where all interaction factors are included.

Table 4. *Parameters in the Additive Model.*

Provincial towns.

Age	1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants males	3 Lower sal. empl. and publ. servants females	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed. of Unsk. Lab.	7 Unskilled workers, other	Total
–24 years	–4563 (4)	–827 (17)	195 (17)	1636 (2)	554 (9)	1564 (4)	1441 (5)	–2224 (58)
25–34 years	–1070 (60)	127 (75)	306 (20)	–526 (20)	431 (36)	–216 (18)	948 (17)	311 (246)
35–44 years	1100 (63)	354 (80)	327 (17)	–631 (9)	100 (10)	–517 (37)	–733 (14)	833 (230)
45–59 years	2184 (73)	351 (69)	–111 (22)	503 (15)	–925 (30)	–765 (31)	–1237 (29)	537 (269)
60 years and over ...	2349 (12)	–5 (20)	–717 (3)	–982 (7)	–160 (2)	–66 (7)	–419 (3)	543 (54)
Total	5690 (212)	–338 (261)	–3070 (79)	619 (53)	691 (87)	–646 (97)	–2946 (68)	10324 (857)

Note: $Y_{ijkt} = M + R_i + S_j + A_k + (SA)_{jk} + e_{ijkt}$.
It should be added that a corresponding table for the Capital would deviate from the table for provincial towns in such a way that 10,324 would be changed to 10,324 + 1,062 = 11,386. Table 4 is based on the information in appendix 2a and agrees with this table and with table 5, apart from rounding errors. The figures in brackets are the numbers.

Table 5. *Expected Income in the Reduced Model.*

Social group — Additive model

Geographical area	Age	1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, males	3 Lower sal. empl. and publ. servants, females	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed. of Unsk. Lab.	7 Unskilled workers, other
The Capital	-24 years..........	10,288	7,996	6,285	11,417	10,406	10,080	7,655
	25-34 years..........	16,316	11,485	8,932	11,790	12,819	10,835	9,698
	35-44 years..........	19,008	12,235	9,475	12,207	13,009	11,056	8,539
	45-59 years..........	19,796	11,936	8,741	13,045	11,689	10,514	7,741
	60 years and over.....	19,967	11,586	8,141	11,567	12,460	11,217	8,563
Provincial towns	-24 years..........	9,226	6,934	5,223	10,355	9,344	9,018	6,593
	25-34 years..........	15,254	10,423	7,870	10,728	11,757	9,773	8,636
	35-44 years..........	17,946	11,173	8,413	11,145	11,947	9,994	7,477
	45-59 years..........	18,734	10,874	7,679	11,983	10,627	9,452	6,679
	60 years and over.....	18,905	10,524	7,079	10,505	11,398	10,155	7,501

Social group — Multiplicative model

Geographical area	Age	1 Higher sal empl. and publ. servants	2 Lower sal. empl. and publ. servants, males	3 Lower sal. empl. and publ. servants, females	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed. of Unsk. Lab.	7 Unskilled workers, other
The Capital	-24 years..........	8,734	7,165	5,542	11,236	9,616	10,000	6,879
	25-34 years..........	15,827	11,345	8,356	11,505	12,529	10,549	8,825
	35-44 years..........	18,634	12,003	8,933	11,612	12,700	10,735	7,850
	45-59 years..........	19,378	11,609	8,080	12,850	11,334	10,219	6,800
	60 years and over.....	19,643	10,998	7,371	10,573	11,899	11,046	7,316
Provincial towns	-24 years..........	7,896	6,477	5,011	10,158	8,694	9,041	6,219
	25-34 years..........	14,309	10,257	7,554	10,402	11,327	9,537	7,978
	35-44 years..........	16,846	10,852	8,076	10,498	11,482	9,705	7,097
	45-59 years..........	17,519	10,495	7,305	11,617	10,247	9,238	6,148
	60 years and over.....	17,758	9,943	6,664	9,559	10,757	9,986	6,614

The interaction among all three factors, social status, geographical area, and age does not seem to be significant, and this holds good of the interaction between geographical area and age. In the case of the interaction between geographical area and social status the statement is lying at Confidence Limit, while in the case of age and social group the interaction is clear. Thus there seems to be reason to take into account only the interaction between age and social group; the reduced model will accordingly be as follows:

Additive model: $Y_{ijkt} = M + R_i + S_j + A_k + (SA)_{jk} + e_{ijkt}$
Multiplicative model: $Y_{ijkt} = M \cdot R_i \cdot S_j \cdot A_k \cdot (SA)_{jk} \cdot e_{ijkt}$

On the basis of this reduced model the parameters used in the analysis of variance have been shown in appendixes 2a and b.

As no allowance has been made for interaction with geographical area, the tables for provincial towns can be converted into figures for the capital by adding a fixed factor, which is kr. 1062 for the additive model[1]); the calculations have, however, been based on the whole material.

[1]) For the multiplicative model the figures must be *multiplied* by a fixed factor.

Owing to this connection between provincial towns and the Capital we shall be concerned in the following only with the provincial towns—and only with the additive model. On the basis of this tabel the expected cell incomes can be determined and an inspection of the parameters of the table will give an impression of the order of the importance of the different factors: social group, age and interaction between social group and age.

Instead of basing the calculations on other unskilled workers and the sixty-year-olds and over, as in the appendix, we have converted the table for the additive model so that the parameters are expressed in such a way that M corresponds to the unweighted average of the cell averages, while S (social status) and A (age) express the effect concerning the marginal distributions for these factors.

This conversion of the table is shown in table 4, which deals with provincial towns. Figures in brackets indicate number of observations.

The table gives the results of the analysis of variance and shows that not only social status and age are very important to the level of wages (salary), but also the interaction between the two factors is of great importance. Where there are few observations, the results will, of course, be uncertain.

Table 6. *Analysis of Variance.*

	Sums of squares (in mill. kr.)	Degrees of freedom	Variance (in 1000)	F-test
	Additive model.			
Between groups...........	31,936	69	462,800	36.8
Within groups (Residual distribution)...........	25,502.4	2,027	12,580	–
Total....................	57,438.4	2,096	27,400	–
	Multiplicative model.			
Between groups...........	41,474.7	69	0.60110	30.1
Within groups (Residual distribution)...........	40,526.4	2,027	0.01999	–
Total....................	82,001.1	2,096	0.03912	–

If we look, first, at the marginal distributions, the table shows that by social groups there are very wide differences in income level, while the effect of the (marginal) age shows that, to begin with, income increases fairly rapidly with age and thereafter it sags.

Concerning the individual social groups it should be noted that the income level for higher salaried employees and public servants is considerably above the average, but that interaction between age and social group leads to an appreciable acceleration in income by age.

By way of contrast, the income level for lower female salaried employees and public servants is quite considerably below the average, and here interaction between age and social group seems to be reflected first in relatively rising income by age and thereafter a decline. In the case of other unskilled workers, on the other hand, the income is generally falling by age.

For skilled workers and members of the Federation of Unskilled Labourers and lower male salaried employees the income level does not seem to deviate so much from the average; it is somewhat higher for skilled workers and somewhat lower for Federation members. For the groups of skilled workers in building trades and Federation members there is a general tendency for the interaction between age and social group to be reflected in relatively declining income.

By means of the parameters from appendixes 2a and b the expected incomes have been calculated in table 5 on the basis of the reduced model and under additive as well as multiplicative assumptions.

The expected incomes according to the multiplicative model are everywhere lower than the expected incomes according to the additive model, the former expected incomes being geometric means and the latter arithmetic means. The estimated expected incomes deviate on the whole very little from the means, as will be seen from appendix 1. This is roughly what is to be expected in view of the results of the analysis of variance.

9. It has been mentioned above that the factors which are taken into account (social group, age and geographical area) are only some of the factors which it might have been interesting to draw into the survey.

It has been studied in the analysis of variance how great a part of the wages (salary) earned can be ascribed to the agency of these three factors and how much remains unexplained. These calculations have been given in table 6.

The F-test deviates significantly and, as will be seen from the table, slightly more than half of the sums-of-squares values seem to be attributable to the agency of the three factors.

The F-test gives a very high value so there is no doubt that even though the theoretical conditions for an analysis of variance are not quite fulfilled, it is permissible, on the basis of the available material, to draw the conclusion referred to above that around half the sums-of-squares values seem to be attributable to the agency of the mentioned factors.

Similar results, although including some more factors, have been arrived at in a British survey by Hill[1]) and in an American survey undertaken by Adams.[2])

It may be mentioned that in Hill's survey the analysis of variance gave the following results on page 26 for the additive and the multiplicative model.

10. With the point of departure in the information about the expected incomes within the individual cells, we have calculated, for each of the wage and salary earners included in the survey, the difference between his actual wages (salary) and the expected wages (salary) according to the reduced model. In connection with these calculations it must be borne in mind that the difference

[1]) T. P. Hill: "An Analysis of the Distribution of Wages and Salaries in Great Britain," Econometrics, vol. 27,3 – July 1959.

[2]) Gerard Adams: »The Size of Individual Incomes: Socio-Economic Variables and Chance Variation«, The Review of Economics and Statistics, Vol. XL. Cambridge, Mass. 1958.

	Sums of squares (mill. £)	Degrees of freedom	F-test
Additive model.			
Between groups	50.3	47	30.1
Within groups (Residual distrib.)	48.5	2366	–
Multiplicative model.			
Between groups	19.8	47	28.1
Within groups (Residual distrib.)	20.5	1366	–

is dependent on the criteria for subdivision on which the calculation of the expected size of "cell averages" has been based.

In the comparison of the actual distribution and the residual distribution for the whole population the fact has been taken into account that the different social groups are not self-weighting. The distributions for the whole population, therefore, should illustrate the conditions of the population of wage and salary earners in 1955.

The first thing we have tried to throw light on is the actual distribution for the population compared with the distributions we get when the incomes are estimated by means of the reduced model, the additive as well as the multiplicative one.

The results have been shown in table 7. It will be seen how the actual distribution has a marked positive skewness—and an »a«-test shows that the distribution is also leptokurtic. By undertaking the estimation by means of the reduced model the effect of the three factors is isolated—and therefore it is not surprising that the distributions, both under additive and multiplicative assumptions, give a far greater accumulation around the "average income" than the original distribution, since that part of standard deviation which is due to more individual circumstances is not included in these distributions. It will be seen that the distribution according to the multiplicative model has a somewhat smaller accumulation around the average than the distribution according to the additive model.

While these distributions illustrate the importance of the factors which it has been possible to allow for in the calculations, the residual distributions will illustrate the importance of the other factors which influence wages (salaries).

Table 7. *Comparison between Original Distribution and Distribution Estimated on the Basis of the Reduced Additive and Multiplicative Models.*

	Original distribution	Distribution based on additive model	Distribution based on multiplicative model
– 2,499	1.7	–	–
2,500– 4,999	4.9	–	–
5,000– 7,499	10.6	8.1	18.5
7,500– 9,999	26.1	31.5	29.5
10,000–12,499	29.4	43.8	35.4
12,500–14,999	15.2	4.6	5.8
15,000–17,499	6.0	3.0	3.0
17,500–19,999	2.1	9.0	7.8
20,000–22,499	1.4	–	–
22,500–24,999	1.1	–	–
25,000–27,499	0.8	–	–
27,500–29,999	0.4	–	–
30,000–32,499	0.1	–	–
32,500–34,999	0.1	–	–
35,000–37,499	0.1	–	–
37,500 and over	–	–	–
	100.0	100.0	100.0

The residual distributions for the reduced model estimated both under additive and multiplicative assumptions will be seen in table 8, which shows also the original distribution.

The tendency of the distributions to approach normal (log-normal) distributions has been tested, compare the subsequent table.

As mentioned, the original distribution has a marked positive skewness and is leptokurtic, and χ^2-test shows, in fact, a significant deviation from a normal distribution. Both the residual distributions, too, are leptokurtic, but the distribution under the additive assumption is positively skew, whereas the distribution under the multiplicative assumption is negatively skew.

While the distribution under the multiplicative assumption also for the χ^2-test deviates significantly from the normal distribution, this is not the case with the distribution under the additive assumption. A further illustration of the distributions will be found in graph 3. This shows, also, that there seems to be best agreement with the normal distribution in the case of the residual

Table 8. *Comparison Between Actual Distributions and Normal Distribution with the Same Means and Standard Deviations.*

	Original distribution		Additive reduced model			Multiplicative reduced model		
	Actual aggregated distribution	Normal distribution		Actual distribution	Normal distribution		Actual distribution	Normal distribution
- 4,999...	6	9	--7,001...	1	1	--0.1501...	9	16
5,000- 7,499...	10	12	-7,000--3,001...	13	16	-0.1500--0.1001...	6	10
7,500- 9,999...	24	18	-3,000--1,001...	21	21	-0.1000--0.0501...	12	12
10,000-12,499...	29	21	-1,000- -1...	16	13	-0.0500--0.0001...	18	13
12,500-14,999...	17	19	0- 999...	17	12	0.0000- 0.0499...	21	13
15,000-17,499...	7	12	1,000- 2,999...	21	20	0.0500- 0.0999...	16	12
17,500-22,499...	4	8	3,000- 6,999...	9	16	0.1000- 0.1499...	9	9
22,500-........	3	1	7,000-........	2	1	0.1500-..........	9	15
	100	100		100	100		100	100

distribution under the additive assumptions. Also the fractile diagrams in appendix 3 show the same thing.

It is no wonder that the additive model should give better agreement with normality than the logarithmic model; this is only natural if it is borne in mind that the cell distributions in table 1 are on the whole more normal under the additive assumption than under the multiplicative assumption. As the

	Original distribution	Residual income: additive reduced model	Residual income: multiplicative reduced model
Number of observations...........	2097	2097	2097
Means...........................	11349	—28	—0.0014
Standard Deviation..............	4627	3168	0.1463
χ^2-test..........................	18.8 (3) s	13.7 (8) ns	29.0 (6) s
Measure of skewness ($\sqrt{\beta_1}$-test).....	1.47 (hs)	1.08 (hs)	—1.92 (vs)
Kurtosis (a-test)..................	0.6935 (ts)	0.7034 (ts)	0.6658 (ts)

hs = positively skew, vs = negatively skew, ts = leptokurtic, s = significantly deviating, ns = not significantly deviating. Figures in brackets indicate degress of freedom.

Graph 3. *Comparison of Actual Distribution With Expected Distribution.*

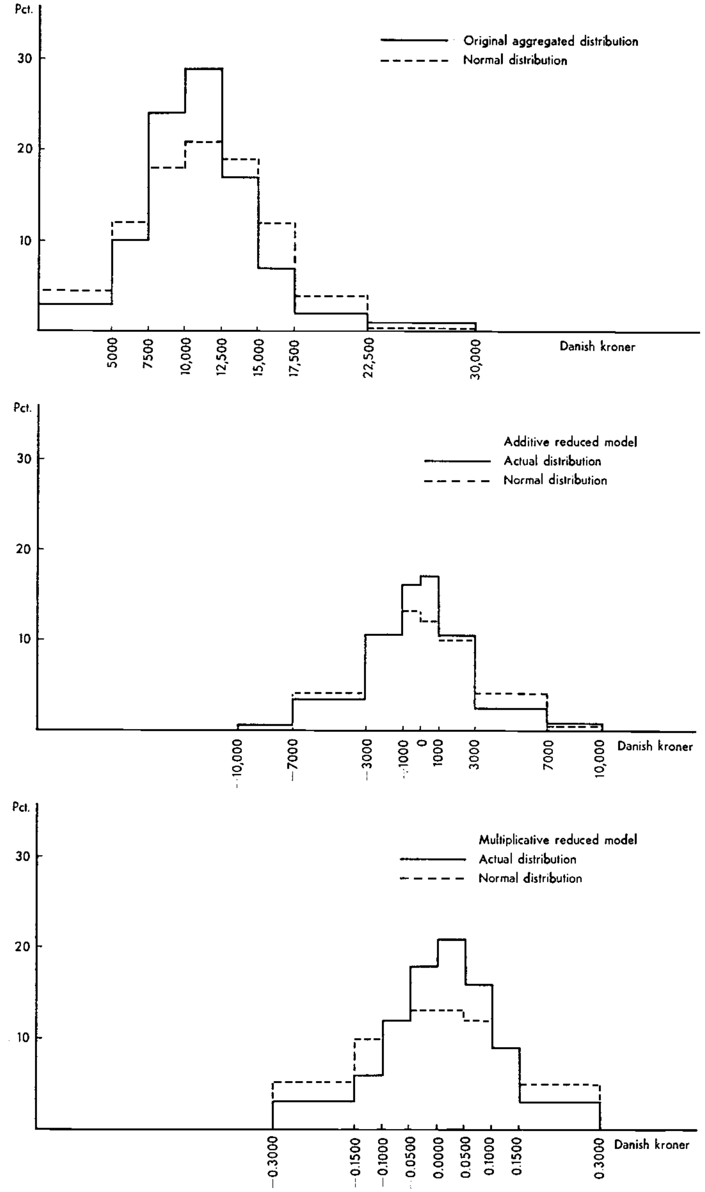


Table 9. *Number of Wage and Salary Earners by Size of Wages and Salaries.*

Income group kr.	The Capital							Provincial towns						
	Higher sal. empl. and publ. servants	Lower sal. empl. and publ. servants, males	Lower sal. empl. and publ. servants, females	Skilled workers, building trades	Skilled workers, other	Unskilled workers, Fed. of Unsk. Lab.	Unskilled workers, other	Higher sal. empl. and publ. servants	Lower sal. empl. and publ. servants, males	Lower sal. empl. and publ. servants, females	Skilled workers, building trades	Skilled workers, other	Unskilled workers, Fed. of Unsk. Lab.	Unskilled workers, other
– 2,499.	»	2	3	2	»	»	6	1	1	5	»	1	»	4
2,500– 4,999.	1	5	10	»	1	3	13	»	11	10	1	2	2	11
5,000– 7,499.	3	11	44	»	8	4	36	»	16	22	3	5	6	16
7,500– 9,999.	5	40	56	1	13	27	33	9	38	25	22	23	56	27
10,000–12,499.	17	145	41	14	45	48	32	14	127	14	23	37	28	10
12,500–14,999.	53	94	9	25	56	16	6	53	59	3	3	19	4	»
15,000–17,499.	101	8	»	14	14	5	1	57	7	»	1	»	»	»
17,500–19,999.	47	»	»	»	4	1	1	32	2	»	»	»	1	»
20,000–22,499.	39	1	»	1	3	»	»	12	»	»	»	»	»	»
22,500–24,999.	28	»	»	»	»	»	1	14	»	»	»	»	»	»
25,000–27,499.	17	»	»	»	1	»	»	7	»	»	»	»	»	»
27,500–29,999.	9	»	»	»	»	»	»	6	»	»	»	»	»	»
30,000–32,499.	7	»	»	»	»	»	»	2	»	»	»	»	»	»
32,500–34,999.	4	»	»	»	»	»	»	2	»	»	»	»	»	»
35,000–37,499.	3	»	»	»	»	»	»	1	»	»	»	»	»	»
37,500–39,999.	»	»	»	»	»	»	»	1	»	»	»	»	»	»
40,000–42,499.	»	»	»	»	»	»	»	1	»	»	»	»	»	»
42,500–44,999.	1	»	»	»	»	»	»	»	»	»	»	»	»	»
45,000–47,499.	1	»	»	»	»	»	»	»	»	»	»	»	»	»
47,500–49,999.	»	»	»	»	»	»	»	»	»	»	»	»	»	»
50,000–......	»	»	»	»	»	»	»	»	»	»	»	»	»	»
Total........	336	306	163	57	145	104	129	212	261	79	53	87	97	68

standard deviations deviate significantly from each other, and since several distributions are leptokurtic, we cannot expect normal (aggregated) residual distributions, but leptokurtic distributions. Incidentally, Hill has emphasized that even aggregation of normal distributions will give leptokurtic distributions if the standard deviations are different.

11. Table 1 showed how some of the cell distributions were normally distributed, while the other distributions were leptokurtic and either positively or negatively skew.

As it is interesting to ascertain how an aggregation influences the distributions, we shall show, in the following, how aggregation by age influences the 14 social groups. In table 9 we have therefore shown the frequency distributions for the 14 social groups, 7 for the capital and 7 for the provincial towns.

To get a clear idea of the appearance of these distributions the frequency distributions have moreover been plotted in graph 4, and in table 10 three tests show the extent to which the distributions seem to be normal or log-normal; the 99 per cent significance level has been used. Besides the two tests already mentioned, $\sqrt{\beta_1}$-test and a-test, a χ^2-test has been used.

As was to be expected from the results in table 1 for the individual cell distributions, the aggregated distributions show a greater tendency towards normality than log-normality; I have therefore only considered this assumption.

As probably was to be expected from the results in table 1, only the distribution for lower female salaried employees and public servants fulfils the conditions for normality in all three tests. Further, the different groups of workers show some tendency towards normality in certain of the tests, whereas the other two groups of salaried employees and public servants are, on the whole, significantly deviating from normal distributions. Also this was to be expected from the results in table 1. The distributions are leptokurtic and, in the case of higher salaried employees and public servants, positively skew; for lower male salaried employees and public servants the distributions are negatively skew. This will be seen clearly from the graphs.

As regards the »a«-test, it proves that in all the cases where the test shows a significant deviation, this is because the distributions are leptokurtic, which is not surprising since even those cell distributions which were not normal were all leptokurtic.

More generally, it may probably be said that in the case of the test a significant skewness may appear if, for more specified social groups, an aggregation is undertaken of normal distributions with different averages or in the case of "naturally" skew distributions.

Now, the significant deviations for the $\sqrt{\beta_1}$-test and the »a«-test generally go together, which is probably a reflection of the fact that we are operating with aggregated distributions which even for more specified social groups have been skew distributions with different standard deviations. By age, this has been the case for some social groups, and this may indicate that something similar will also hold good of more specified social groups. Whether we shall be able to arrive at normal distributions through very detailed breakdowns by social groups, is difficult to say, but it may probably be expected that even in these cases the distributions will have significantly different standard deviations. In support of these reflections reference is made also to the results for the distributions within the individual "cells".

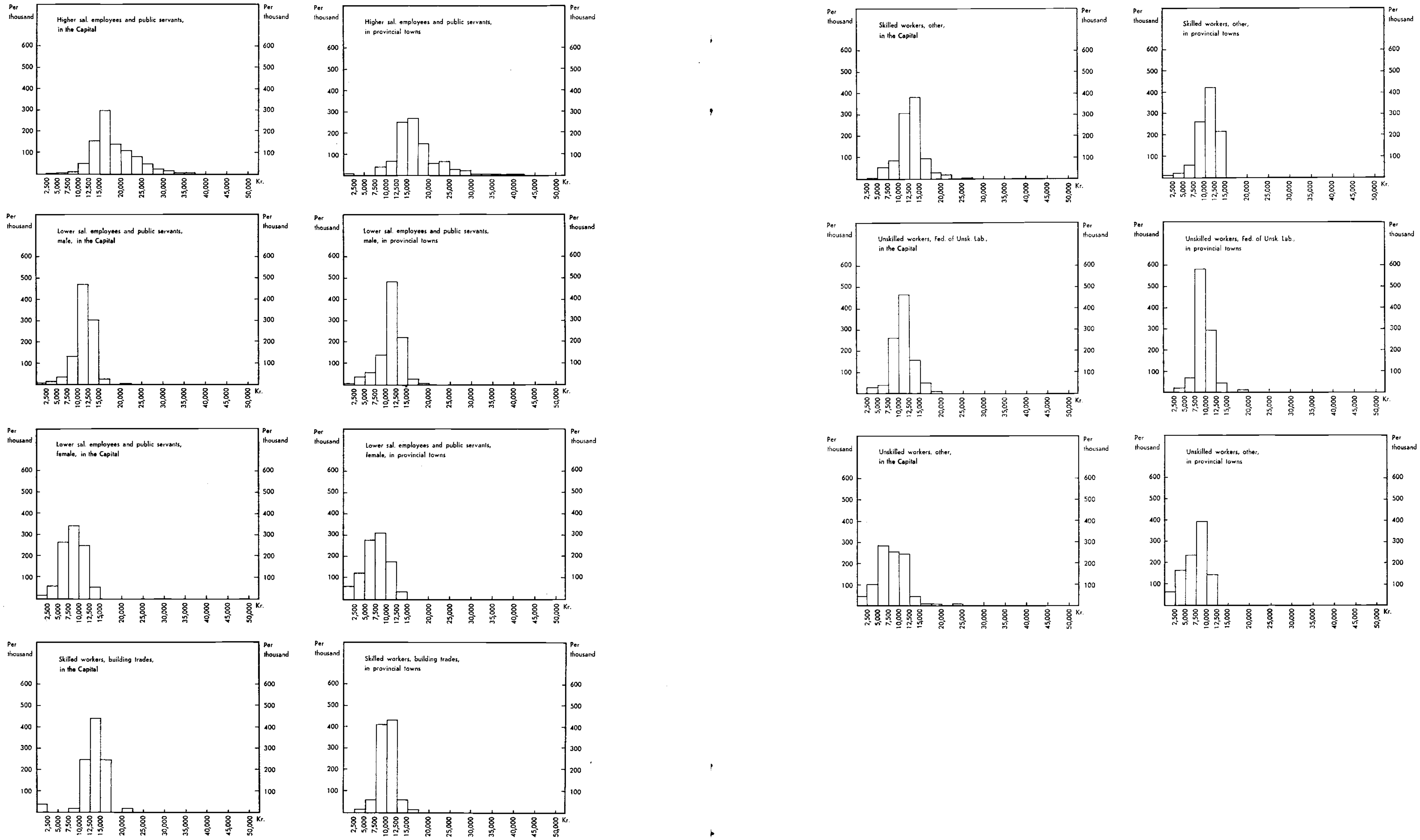
It will, in fact, be seen from table 1 that some of the distributions within "cells" are normal, whereas the aggregated distributions where age has not been used as the subdivision criterion, are not. Further, the analysis of the standard deviations within "cells" has shown that these standard deviations are significantly deviating, a fact which will also influence the result of an aggregation.

Table 10. *Actual Distributions—Level of Significance corresponding to 99 per cent.*

Geographical area	Test	Social group: Normal Distribution 1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, male	3 Lower sal. empl. and publ. servants, female	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed of Unsk. Lab.	7 Unskilled workers, other	Log-normal Distribution 1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, male	3 Lower sal. empl. and publ. servants, female	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed of Unsk. Lab.	7 Unskilled workers, other
The Capital	χ^2-test	s	s	N	N	s	N	N	s	s	s	s	s	N	s
	$\sqrt{\beta_1}$-test	hs	vs	N	vs	N	N	hs	N	vs	vs	vs	vs	vs	vs
	a-test	N	ts	N	ts	ts	ts	N	N	ts	ts	ts	ts	ts	ts
Provincial towns	χ^2-test	s	s	N	¹)	N	s	N	s	s	s	¹)	s	s	s
	$\sqrt{\beta_1}$-test	hs	vs	N	N	vs	hs	N	vs	vs	vs	vs	vs	vs	vs
	a-test	ts	ts	N	N	N	ts	tf	ts	ts	N	ts	ts	ts	N

Note: *N* means normally distributed, *S*: significantly deviating; *hs* means positively skew, *vs*: negatively skew; *ts* means leptokurtic and *tf*: platykurtic.
¹) Cannot be estimated as there are too few degrees of freedom.

Graph 4. *Number of Wage and Salary Earners by Size of Income from Wages and Salaries for Individual Social Groups.*



Appendix 1a: *Additive Model: Number of Observations, Means, Standard Deviations and Tests.*

Geographical area	Age	Social group: Higher sal. empl. and publ. servants		Lower sal. empl. and publ. servants, male		Lover sal. empl. and publ. servants, female		Skilled workers, building trades		Skilled workers, other		Unskilled workers, Fed. of Unsk. Lab.		Unskilled workers, other	
The Capital	–24 years	7 11280	−1.134	13 7903	−0.3775	18 6227	−2.084	1 13000	–	12 10520	−0,1962	2 9860	−0,0000	11 8430	0,1394
		2680	0.7637	2860	0.7967	2007	0.6900	–	–	2273	0,8482	1217	1,000	2683	0,9172
	25–34 years	73 16290	6.463	88 11190	−0.7645	35 9187	0.0917	9 13480	−0.0012	37 13690	0,4414	21 10810	−0,1295	25 9384	0,9553
		5731	0.6292	1794	0.7770	1957	0.8675	1233	0.7895	2869	0,7565	2788	0,7958	4704	0,7236
	35–44 years	97 19020	2.949	96 11970	−0.5677	35 9096	−0.4900	18 12830	−1.913	36 13380	5,124	28 11430	0,2132	29 8898	−0,0800
		5319	0.7409	2096	0.7443	2510	0.7549	3553	0.7433	2958	0,6020	3001	0,6690	2963	0,8704
	45–59 years	125 19790	0.6722	73 11510	−2.001	60 8537	−0.0326	21 14210	1.479	41 11800	−0,3025	43 10660	−0,7396	50 7616	−0,0419
		5849	0.7851	2513	0.7219	2705	0.8304	2186	0.7523	2921	0,9360	1877	0,7457	2883	0,8114
	60 years and over	34 19290	−0.00003	36 11570	−0.0231	15 7954	−0.6155	8 11270	−0.9042	19 12490	−0,4208	10 11140	−0,1424	14 8571	−0,6178
		4651	0.8663	2974	0.7170	2965	0.8467	4509	0.7316	3480	0,7533	2365	0,7800	3265	0,7901
Provincial towns	–24 years	4 7484	−0.6293	17 7005	−0.0511	17 5285	−0.2666	2 9563	−0.0000	9 9196	−0,4222	4 9128	1.306	5 4887	0,1328
		4618	0.8210	2625	0.8759	1760	0.8677	2102	1.0000	4072	0,8496	364,9	0,8655	2485	0,9249
	25–34 years	60 15290	5.269	75 10770	0.0078	20 7423	−0.0074	20 9969	−0.6527	36 10860	−2,217	18 9804	0,2689	17 9098	−1,027
		4975	0.6461	1805	0.7276	2999	0.7815	2133	0.7726	2124	0,7421	1542	0,8257	1882	0,7265
	35–44 years	63 17930	3.105	80 11490	−0.4233	17 9195	−2.967	9 9903	0.1661	10 10610	−0,0404	37 9708	−0,2248	14 6735	−0,0260
		5877	0.7247	2571	0.7346	2358	0.6458	703,5	0.8735	2986	0,8912	1894	0,7914	2397	0,8875
	45–59 years	73 18740	1.314	69 11330	−0.6859	22 8237	−0.2865	15 10350	0.0833	30 10470	−0,0774	31 9252	2,512	29 6895	−0,1535
		4995	0.7427	2479	0.7369	3091	0.8297	1555	0.7830	2352	0,7623	2434	0,5845	3058	0,8666
	60 years and over	12 20830	0.0963	20 10550	−2.273	3 8012	−0.1937	7 10850	0.2670	2 11070	−0,0000	7 10260	0,2443	3 7465	0,1343
		4551	0.9195	2916	0.7029	3776	0.9008	3274	0.8143	1712	1,0000	1249	0,8653	4169	0,8882

Note:

No. of observations Means	Measure of Skewness $\sqrt{\beta_1}$-test
Standard Deviation	Geary's *a*-test

Appendix 1b: *Multiplicative Model: Number of Observations, Means, Standard Deviation and Tests.*

Geographical area	Age	Social group													
		Higher sal. empl. and publ. servants		Lower sal. empl. and publ. servants, male		Lover sal. empl. and publ. servants, female		Skilled workers, building trades		Skilled workers, other		Unskilled workers, Fed. of Unsk. Lab.		Unskilled workers, other	
The Capital	– 24 years	7 4.039	−2.136	13 3.856	−3.495	18 3.749	−7.165	1 4.114	–	12 4.011	−0.4519	2 3.992	−0.0001	11 3.906	0.0254
		0.1246	0.7508	0.2280	0.7076	0.2536	0.6326	–	–	0.1015	0.8385	0.0537	1.0000	0.1373	0.8900
	25–34 years	73 4.190	−0.0263	88 4.042	−4.453	35 3.954	0.0235	9 4.128	−0.0811	37 4.127	0.0185	21 4.017	−1.635	25 3.914	−0.7965
		0.1367	0.6894	0.0810	0.7091	0.0927	0.8733	0.0402	0.8038	0.0891	0.7740	0.1305	0.7354	0.2462	0.7391
	35–44 years	97 4.265	0.4627	96 4.070	−7.165	35 3.936	−3.699	18 4.078	−9.028	36 4.118	1.360	28 4.043	−0.9149	29 3.921	−0.8704
		0.1086	0.7847	0.0924	0.6554	0.1471	0.6908	0.2034	0.5660	0.0833	0.6453	0.1228	0.6385	0.1705	0.8250
	45–59 years	125 4.278	−0.0561	73 4.046	−5.999	60 3.905	−1.561	21 4.148	0.4319	41 4.056	−1.140	43 4.020	−3.580	50 3.841	−1.251
		0.1273	0.7792	0.1281	0.6400	0.1638	0.7740	0.0631	0.07760	0.1237	0.8095	0,0893	0.6850	0.2061	0.7690
	60 years and over	34 4.272	−0.3967	36 4.044	−8.877	15 3.860	−1.202	8 3.991	−3.728	19 4.074	−3.743	10 4.037	−1.035	14 3.863	−7.515
		0.1117	0.8322	0.1524	0.5848	0.2142	0.8264	0.2989	0.6362	0.1584	0.6657	0.1017	0.7258	0.3414	0.5529
Provincial towns	– 24 years	4 3.726	−1.213	17 3.810	−0.4660	17 3.694	−1.081	2 3.975	−0.0000	9 3.901	−1.849	4 3.960	−0.1208	5 3.643	0.0175
		0.5132	0.8594	0.1902	0.8716	0.1760	0.8464	0.0962	1.000	0.2846	0.7859	0.0172	0.8552	0.2252	0.9045
	25–34 years	60 4.167	0.7249	75 4.026	−0.9418	20 3.823	−2.022	20 3.987	−2.405	36 4.024	−11.23	18 3.987	0.0465	17 3.948	−2.933
		0.1199	0.7230	0.0773	0.7046	0.2347	0.7975	0.1097	0.7198	0.1194	0.6070	0.0670	0.8392	0.1079	0.6833
	34–44 years	63 4.235	0.6125	80 4.047	−2.637	17 3.937	−9.636	9 3.995	0.0816	10 4.008	−0.2577	37 3.978	−1.317	14 3.798	−0.6927
		0.1240	0.7595	0.1162	0.6828	0.1890	0.5252	0.0306	0.8756	0.1324	0.8620	0.0942	0.7586	0.1796	0.8279
	45–59 years	73 4.259	0.0898	69 4.041	−3.602	22 3.870	−3.132	15 4.010	−0.0025	30 4.008	−1.104	31 3.952	−2.014	29 3.773	−1.810
		0.1092	0.7635	0.1170	0.6546	0.2341	0.7304	0.0653	0.8026	0.1086	0.7440	0.1202	0.5731	0.2782	0.7485
	60 years and over	12 4.309	0.0344	20 3.993	−8.608	3 3.863	−0.3424	7 4.018	0.0057	2 4.041	−0.0007	7 4.009	0.1402	3 3.826	0.0026
		0.0940	0.9324	0.2014	0.5902	0.2453	0.9236	0.1301	0.8575	0.0672	1.000	0.0518	0.8661	0.2496	0.8277

Note:

No. of observations Means	Measure of skewness $\sqrt{\beta_1}$-test
Standard Deviation	Geary's a-test

Appendix 2a. *Parameters in the Additive Model. Provincial Towns.*

Age group	Social group							
	1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, males	3 Lower sal. empl. and publ. servants, females	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed. of Unsk. Lab.	7 Unskilled workers, other	Age group component A_k
-24 years...	−8,771	−2,682	−948	758	−1,146	−229	»	−908
25–34 years...	−4,786	−1,236	−344	−912	−776	−1,517	»	1,135
35–44 years...	−935	673	1,358	664	573	−137	»	−24
45–59 years...	651	1,172	1,422	2,300	51	119	»	−822
60 years and over....	»	»	»	»	»	»	»	»
Social group component S_j	11,404	3,023	−422	3,004	3,897	2,654	»	7,501

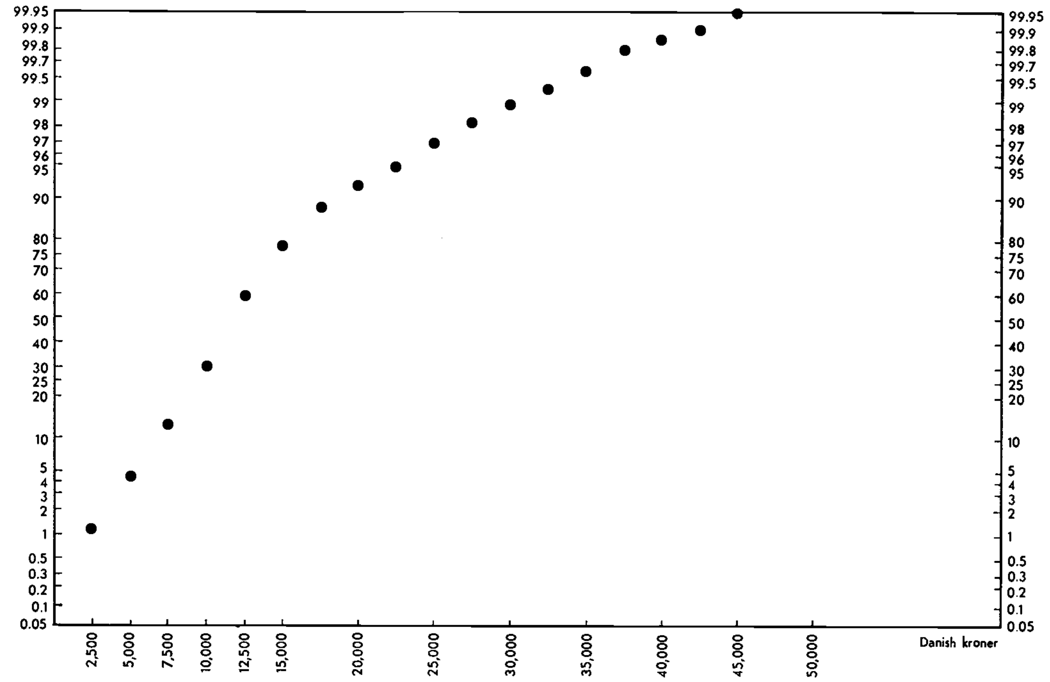
Note: Model $S_{ijkt} = M + R_i + S_j + A_k + (SA)_{jk} + e_{ijkt}$. Interactions of the second order and interactions between geographical area and social groups and between geographical area and age are considered non-significant.
For the Capital the factor $R_i = 1062$ is to be added.

Appendix 2b. *Parameters in the Multiplicative Model Multiplied by 10⁴. Provincial Towns.*

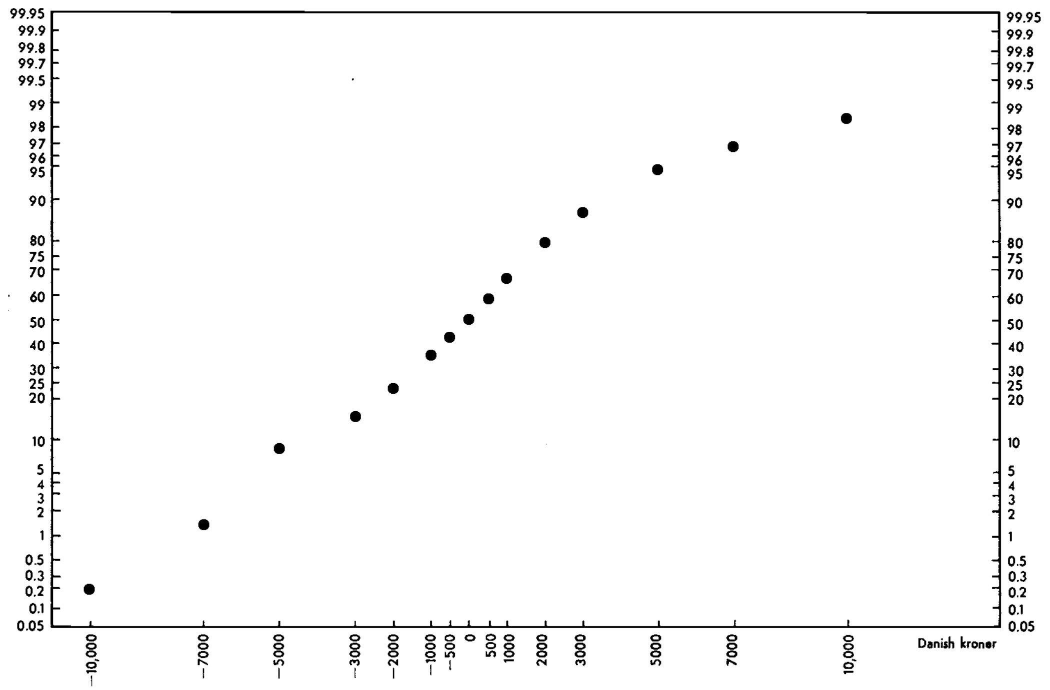
Age group	Social group							
	1 Higher sal. empl. and publ. servants	2 Lower sal. empl. and publ. servants, males	3 Lower sal. empl. and publ. servants, females	4 Skilled workers, building trades	5 Skilled workers, other	6 Unskilled workers, Fed. of Unsk. Lab.	7 Unskilled workers, other	Age group component A_k
-24 years...	−3,252	−1,593	−979	523	−657	−164	»	−268
25–34 years...	−1,752	−679	−269	−447	−590	−1,014	»	814
35–44 years...	−535	74	529	101	−23	−430	»	306
45–59 years...	259	553	717	1.165	107	−20	»	−318
60 years and over....	»	»	»	»	»	»	»	»
Social group component S_j	4,289	1,770	32	1,599	2,112	1,789	»	38,295

Note: Model $S_{ijkt} = M \cdot R_i \cdot S_j \cdot A_k \cdot (SA)_{jk} \cdot e_{ijkt}$. Interactions of the second order and interactions between geographical area and social groups and between geographical area and age are considered non-significant.
For the Capital the factor $R_i = 438$ (10^4) is to be added.

Appendix 3,1. *Aggregated Actual Distribution.*



Appendix 3,2. *Residual Distribution. Additive Model.*



Appendix 3,3. *Residual Distribution. Multiplicative Model.*