Anders Hald

T.N. THIELE'S CONTRIBUTIONS TO STATISTICS

Preprint 1979 No. 6

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

October 1979

H.C.Ø.-tryk
København

Summary.

The background for the work of Thiele (1838-1910) is sketched. He made contributions to the theory of skew distributions, he defined the cumulants and investigated their properties, and he formulated the canonical form of the linear model with normally distributed errors and reduced the general linear model to canonical form by means of an orthogonal transformation thereby providing a new justification for the method of least squares. Furthermore, he formulated the basic principle for the one-way analysis of variance and tested the significance of the variation between groups by means of the difference of the between-group variance and the within-group variance. He derived estimates of the parameters in the two-way classification model without interaction, also for the case with missing observations. Finally he stressed the importance of criticism of the model by means of graphical and numerical analysis of residuals using rational subgroups of observations.

Key words.

History. Thiele. Oppermann. Gram. Transformation of skew distributions. Gram-Charlier Type A series. Cumulants and k-statistics. Canonical form of linear model. Method of least squares. Orthogonal transformations. Analysis of variance. Criticism of model.

The paper was written on the occasion of the 500th anniversary of the University of Copenhagen on the 1st of June 1979.

1. The Background

The works of Laplace and Gauss in the beginning of the 19th century mark a new era in the theory of statistics. The central limit theorem proved by Laplace in 1812 provided the basic tool for the asymptotic theory of estimation and was used in his discussion of the method of least squares. Gauss (1809) gave a probabilistic basis for the method of least squares assuming a uniform distribution of the parameters and defining the best estimate as the one maximizing the posterior density. In his second proof Gauss (1821-1823) gave a non-Bayesian theory of estimation for the linear model defining the best estimate as the linear unbiased estimate with minimum variance.

Within the chosen framework Gauss' theory of estimation was rather complete. Nevertheless, an immense number of papers on the method of least squares was published in the remaining part of the 19th century with the purpose to disseminate knowledge of the method, to give examples of applications and study the distribution of residuals, to discuss the numerical problems in solving the normal equations, to derive the solution for special cases of the linear model and to give other motivations for using the method than those presented by Gauss. Danish astronomers, geodecists and actuaries took part in this developent.

Heinrich Christian Schumacher (1780-1850), Professor of Astronomy at the University of Copenhagen, studied under Gauss in 1808-1809 and was therefore well acquainted with the method of least squares at the time of Gauss' first publication on this subject. When the

Danish Geodetic Institute was founded in 1816 Schumacher became its director and as such he initiated a close cooperation between a geodetic survey of Denmark, starting in 1817, and a geodetic survey of Hannover, which Gauss was commissioned to begin in 1818. A biography (in Danish) of Schumacher has been written by Einar Andersen (1975). Whereas Schumacher accepted the method of least squares on the authority of Gauss the next two generations of statisticians were rather critical towards Gauss' justifications of the method.

Carl Christopher Georg Andræ (1812-1893) succeeded Schumacher as director of the Danish Geodetic Institute. Besides completing the survey and carrying out the analysis of the data he wrote several papers on the metod of least squares. Andræ (1867) pointed out that the best (i.e. minimum variance) estimate of the true value depends on the error distribution. As an example he considered the uniform distribution and proved that the best estimate is the average of the smallest and the largest observation. On these grounds he criticised Gauss' second model and concluded that the method of least squares leads to the best estimate only if the observations are normally distributed. Thus he combined the ideas of the two proofs of Gauss by first deriving the most probable value in the posterior distribution and afterwards proving that this leads to the minimum mean square error. Andræ (1860) also wrote an interesting paper on the estimation of the median and the interquartile distance in a symmetric distribution by means of linear functions of empirical percentage points.

Georg Karl Christian Zachariae (1835-1907), who succeeded Andræ as director, wrote a textbook (1871) on the method of least squares. As a motivation for basing his exposition on the normal distribution he gave a detailed discussion of the hypothesis of elementary errors, i.e. the hypothesis that any observed error may be considered as the sum of a large number of independent elementary errors having different (symmetric) distributions with finite moments, and proved the central limit theorem using characteristic functions following the proof given by Bessel (1838). Based on the ideas of Andræ he gave an excellent exposition of the method of least squares.

About the same time another excellent textbook was written by the German geodecist F.R. Helmert (1872) giving a rather complete survey of the method of least squares as developed by Gauss and his pupils. Helmert's book is essentially based on the second proof by Gauss. With these two books in existence there was no reason for Thiele to write an ordinary textbook on the same subject.

Frederik Moritz Bing (1839-1912), Actuary at the State Life Insurance Company, attacked the usefulness of Bayesian methods in statistics in a paper published in Tidsskrift for Mathematik, 1879. One of the usual, rather heated and confused discussions evolved between Bing and Ludvig Valentin Lorenz (1829-1891), Professor of Mathematics at the Military Academy, who defended the Bayesian view. As a result of this discussion Thiele rejected Bayes' postulate.

Ludvig Henrik Ferdinand Oppermann (1817-1883), Professor of German at the University of Copenhagen and Actuary at the State Life Insurance Company, exerted a profound influence on Thiele. In a paper from 1872 on the justification for the method of least squares he introduces as loss function, L say, and like Laplace and Gauss he defines the best estimator as the one minimizing L, but whereas Laplace and Gauss arbitrarily chose the functional form of L Oppermann wanted to derive the functional form from fundamental principles without specifying the distribution of the observations. We shall sketch his main ideas. He restricted the class of estimators considered to location and scale equivariant functions. Let $L = L(y_1 - m, \ldots, y_n - m)$ where the y's denote n independent observations and m denotes an estimator. First he requires that L for all the observations should equal the sum of the L's for subgroups of observations, which leads to $L = \Sigma F(y_i - m)$, where F is unknown. Next he requires that the best estimator based on all the observations should also be obtainable by finding the best estimators from subgroups of observations and combining them by means of the loss function as if they were direct observations. This leads to a differential equation which has the solution $F(x) = x^2$ and hence the method of least squares.

Oppermann used the distribution which to-day is known as the Gram-Charlier Type A series and estimated the parameters by the method of moments, see Gram (1879, pp. 6-7 and p. 94). Furthermore, he also pointed out the advantages of orthogonal transformations of observations in the linear model, see Gram (1879, p. 7) and Thiele (1903, p. 54).

Jørgen Pedersen Gram (1850-1916), mathematician and actuary, worked closely together with Thiele. They were both actuaries in the Life Insurance Company Hafnia, Thiele from 1872 and Gram from 1875. Gram wrote an important thesis (1879) on series expansions by means of orthogonal functions and the determination of the coefficients by the method of least squares. One of the series discussed is the Gram-Charlier Type A series. He also gave the general formula for variance-stabilizing transformations, $g'(x) = 1/h(x)$, where $g(x)$ denotes the transformation function and $h(x)$ gives the relation between the mean and the standard deviation of the variable in question, see p. 98, 1879. Furthermore, he discussed stratified sampling to estimate the total of a population and derived the rule for optimum allocation of the sample, which was rediscovered by Neyman in 1934, see Gram (1883, p. 181).

2. Thorvald Nicolai Thiele, 1838-1910

T. N. Thiele was born into a well-known Danish family of book-printers, instrument-makers and opticians. His father, Just Mathias Thiele (1795-1874) was a man of many talents, practical as well as artistic. For many years he was private librarian to King Christian VIII, director of the Royal Collection of Prints and secretary of the Royal Academy of Fine Arts. He made a name for himself as dramatist, poet and folklorist. The home in which T. N. Thiele grew up was thus a highly cultured one where many leading intellectuals of the time were frequent guests. He was named after the sculptor Bertel Thorvaldsen who was one of his godfathers. "Little Ida", to whom Hans Christian Andersen first told "Little Ida's Flowers" was his half-sister, Ida Holten Thiele.

T.N. Thiele got his masters degree in astronomy in 1860, his doctors degree in 1866 and worked from 1860 as assistant to Professor Heinrich Louis d'Arrest (1822-1875) at the Copenhagen Observatory. During 1870-1871 he worked on establishing the actuarial basis for the life insurance company Hafnia, which was started in 1872 with Thiele as actuary. In 1875 he became Professor of Astronomy and director of the Copenhagen Observatory. He retired in 1907. In 1895 he became corresponding member of The Institute of Actuaries, London. He took the initiative to found the Danish Society of Actuaries in 1901 and was its President until he died.

In 1867 Thiele married Marie Martine Trolle (1841-1889). They had six children.

Thiele wrote about astronomy, number theory, numerical analysis, actuarial mathematics and mathematical statistics. For Thiele's work as an actuary we refer to the obituary by Gram (1910). We shall give a detailed account of his contributions to mathematical statistics.

Thiele's fundamental contributions to statistics are contained in a book "Almindelig Iagttagelseslære" (General Theory of Observations), 1889, and a paper "Om Iagttagelseslærens Halvinvarianter" (On the Half-Invariants From the Theory of Observations), 1899. They are rather difficult to understand and presumably for that reason he wrote a "popular" version of the book called "Elementær Iagttagelselære" (Elementary Theory of Observations), 1897, covering nearly the same theory but without some of the more difficult proofs and with more examples. A very poor English translation "Theory of Observations" was published in 1903 and reprinted in Annals of Mathematical Statistics, 1931.

In my opinion Thiele's first book is considerably better than the following two versions. At the time of publication it must have been considered highly unconventional as a textbook because it concentrates on skew distributions (instead of the normal), cumulants (instead of central moments) and a new justification and technique for the method of least squares (instead of the Gauss'ian minimum variance, linear unbiased estimation). Since both the book and his 1899 paper are written in Danish they are not widely known and I shall therefore give rather detailed references to help the interested reader. It is easy to find the corresponding results in the English version of his book.

Thiele had the bad habit of not giving precise references or no references at all to other authors even if he used their results freely. He just supposed that his readers were fully acquainted with the literature. (From that point of view it therefore serves him right that he was himself neglected, for instance by K. Pearson and R. A. Fisher.) I have tried to track down the origin of his ideas as far as possible but some of my comments on this matter are pure guesswork.

Translations given in the following of works by Gauss, Thiele and Gram are not literal but intend only to convey the meaning.

In his obituary of Thiele, Gram (1910) writes:

"He thought profoundly and thoroughly on any matter which occupied him and he had a wonderful perseverance and faculty of combination. But he liked more to construct his own methods than to study the methods of other people. Therefore his reading was not very extensive and he often took a one-sided view which had a restrictive influence on the results of his own speculations. Furthermore, he had very great difficulties in finding correct expressions for his thoughts, in writing as well as verbally, even if he occasionally could be rather eloquent. Thiele's importance as a theoretician lies therefore more in the original ideas he started than in his formulations, and his ideas were in many respects not only original but far ahead of his times. Therefore he did not get the recognition he deserved and some time will elapse before his ideas will be brought in such a form that they will be accessible to the great majority, but at that time they will also be fully valued because of their fundamental importance."

Perhaps the time has come now 90 years after the publication of his book.

We shall use modern terminology and notation to make Thiele's work easier to understand. For example, the symmetric functions defined by him and called half-invariants will here be called cumulants. Expectation and variance of a random variable are denoted by E and V, respectively, and the variance will also be denoted by σ^2. Unless the contrary is explicitly stated the random variables (observations) considered are assumed to be independent. Thiele made no contribution to the theory of discrete distributions. We shall therefore assume that the distributions considered are continuous, apart from the considerations on cumulants and estimation methods in Sections 4 and 5 which are of a general nature. The standardixed normal density will be denoted by $\phi(x)$. For brevity we shall use matrix notation in the discussion of the linear model. Matrices are assumed to be conformable and of full rank unless otherwise explicitly stated.

3. Skew distributions

At the time when Thiele became interested in statistics the normal distribution played a predominant role. However, Oppermann, Thiele and Gram working with economic and demographic data as actuaries realized the need for developing a theory of skew distributions.

The Gram-Charlier Type A distribution.

Let $f(x)$ denote a continuous density with finite moments. Thiele (1889, pp. 13-16 and 26-28) writes the density in the form

$$f(x) = k_0 \phi(x) - k_1 \phi'(x) + k_2 \phi''(x)/2! - \dots$$

and determines the unknown coefficients in terms of the cumulants, see Section 4.

To-day this expansion is usually called the Gram-Charlier Type A series. The name is, however, incorrect from a historical point of view. Gnedenko and Kolmogorov (1954, p. 191) point out that the expansion occurs in the work of Tchebychev. Särndal (1971) and Cramér (1972) have written a history of the subject with special regard to Swedish contributions.

As mentioned above Oppermann used the Type A series before Gram. Thiele (1873) wrote a paper about numerical aspects of the series. In his books Thiele did not discuss the convergence of the series, presumably because he only considered the use of a finite part, neither did he comment on the property that the density may become negative.

There are no indications of how they invented the series. I suppose they considered it trivial for the following two reasons. As actuaries they were familiar with the technique of finding approximation formulas by first transforming the given function by subtraction of or division by a suitably chosen simple function and then using a polynomial as approximation to the difference or quotient. Hence, the normal distribution multiplied by a polynomial of low degree was a natural starting point for approximating skew distributions. The second reason was the hypothesis of elementary errors. Using characteristic functions Bessel (1838) had derived the Type A series under the assumption that the distribution of the elementary errors were symmetric and Zachariae (1871) had included a discussion of this problem and of Bessel's proof in his book. Oppermann, Thiele and Gram were therefore fully aware of the probabilistic background of the expansion and presumably they considered an extension to non-symmetric distributions a commonplace.

In his discussion of the distribution of the circumference of trees Gram (1889, p. 114) remarks that a density of the form $f(x) = (a+bx)\exp\{-k(x-c)^2\}$ seems to be adequate. Of course they were also aware of the possibility of using densities of the form

$$f(x) = (k_0 + k_1 x + k_2 x^2 + \dots) e^{-\alpha x}.$$

As a special case Gram (1879, pp. 105-107) fitted a gamma distribution to the distribution of the marriage age for men, using the method of moments to estimate the parameters.

Transformation of skew distributions.

The first paper by Thiele (1878) on statistics has the title " Bemærkninger om skæve Fejlkurver" (Remarks on skew frequency curves). Let the density of x be $\phi(x)$, the standardized normal density. Thiele remarks that the density of $y = f(x)$ then will be $p(y) = \phi(x)/f'(x)$ which gives a general expression for skew distributions when $f(x)$ is non-linear. As a simple example he discusses the transformation $y = \alpha + \beta x + \gamma x^2$ and estimates the three parameters by setting the first three empirical moments about the origin equal to the corresponding theoretical moments.

In 1889, pp. 29-31, he returns to this principle of transformation and points out that if the transformation is not one-to-one $p(y)$ will be the sum of terms of the form $\phi(x)/f'(x)$ corresponding to the number of roots of the equation $f(x) = y$. For $y = \alpha + \beta x + \gamma x^2$ he derives the cumulants of y in terms of the cumulants of x and gives the equations for determination of the parameters. (This section has not been included in the 1897 and 1903 books.)

Thiele had this idea of transformation and used it long before his 1878 paper. In his thesis (1866, pp. 7-8) he pointed out that the geometric, instead of the arithmetic, mean should be used to estimate distances based on certain astronomical measurements. In 1875 he fitted a logarithmic normal distribution to data on the marriage age for females, using $\log(y-\alpha)$ as normally distributed where y denotes the age at marriage, see Gram (1910).

4. Cumulants

Oppermann (1872) distinguishes between the special and the general theory of statistic according as the theory is based on specified distributions or not.

Thiele's (1889) fundamental contribution to the general theory is the introduction of cumulants (half-invariants as he called them) and the development of a corresponding theory, the only assumption being the existence of moments.

Thiele's starting point is the one-to-one correspondence between n observations and a set of symmetric functions of order 1 to n. Beginning with the moments about the origin, m'_r, he next defines the moments about the sample mean, m_r, and uses the mean, the variance and $m_r m_2^{-r/2}$, $r = 3,4,\ldots$, to characterize the distribution. He then writes (1889, p.19): "It is, however, better to use what we shall call the half-invariants defined by the formula

$$m'_{r+1} = \sum_{i=0}^{r} \binom{r}{i} m'_{r-i} h_{i+1} \, , \quad r = 0,1,\ldots \, , \qquad (4.1)$$

where h_i denotes the *i*th empirical half-invariant".

Solving these equations he finds h_r in terms of the first r moments about the origin and he also derives the simpler relations between the cumulants and the moments about the mean. In particular he finds $h_1 = m_1'$, $h_2 = m_2$, $h_3 = m_3$ and $h_4 = m_4 - 3\,m_2^2$.

Replacing the empirical moments in (4.1) by the theoretical moments, μ_r' , the theoretical cumulants, κ_r, are defined. Thiele derived the cumulants for the binomial, the rectangular, the normal and the Gram-Gharlier Type A distribution and also gave the cumulants of $y = \alpha + \beta x + \gamma\, x^2$ in terms of the cumulants of x. In 1903 he also gave the cumulants for the Poisson and the mixed normal distribution.

Thiele's (1889, pp. 21-22) comments on the interpretation of the cumulants are as follows. The mean, κ_1, depends on both location and scale, the variance, κ_2, depends on the scale but is independent of the location, and the quantities $\gamma_r = \kappa_{r+2}\,\kappa_2^{-(r+2)/2}$, $r = 1,2,\ldots$, are independent of both location and scale and therefore describes the shape of the distribution. The first four cumulants are the most important, the third characterizes the skewness and the fourth the flatness or peakedness of the distribution.

His interpretation of the first four cumulants is presumably influenced by his results for the Type A series. Let the random variable x have the cumulants $\kappa_1, \kappa_2, \ldots$ and consider the standardixed variable $y = (x - \kappa_1)/\sqrt{\kappa_2}$. Thiele proves that the density of y,g(y) say, may be written as

$$g(y) = \phi(y) - \gamma_1\,\phi^{(3)}(y)/3! + \gamma_2\,\phi^{(4)}(y)/4! + \ldots, \qquad (4.2)$$

the following terms being more complicated. Actually he gives the coefficients up to the 8th term (1889, p. 28). Considering the first three terms only it is easy to see how γ_1 and γ_2 influence the shape of the distribution.

For independent random variables, $x_1, \ldots, x_n$, he proves the fundamental (addition) theorem (1889, p. 36)

$$\kappa_r(\Sigma a_i x_i) = \Sigma a_i^r \kappa_r(x_i). \tag{4.3}$$

Finally, he derives the (theoretical) cumulants of the empirical cumulants, i.e. $\kappa_r(h_i)$, (1889, pp. 60-62), for the most important values of r and i. First of all

$$\kappa_r(h_1) = \kappa_r n^{1-r}$$

and

$$\kappa_r(h_1)\{\kappa_2(h_1)\}^{-r/2} = \kappa_r \kappa_2^{-r/2} n^{1-(r/2)},$$

which he uses to prove that the distribution of the mean of independent and identically distributed random variables is asymptotically normal regardless of the distribution of the observations if only the moments exist.

Expressing the higher empirical cumulants in terms of sums of products of observations (auxiliary tables are given) Thiele derives $\kappa_1(h_i)$ for $i = 1, \ldots 6$, $\kappa_2(h_i)$ for $i = 1, \ldots, 4$, $\kappa_3(h_i)$ for $i = 1,2$ and (in 1903) $\kappa_4(h_i)$ for $i = 1,2$. He regrets that he has not succeeded in finding the general formula. Among these results is the important formula (1889, p. 61)

$$\kappa_2(h_2) = \kappa_4(n-1)^2/n^3 + 2\kappa_2^2(n-1)/n^2 .$$

The version used to-day

$$V\{s^2\} = \kappa_4/n + 2\,\kappa_2^2\,/(n-1)\,, \qquad s^2 = h_2 n\,/\,(n-1)\,, \qquad (4.4)$$

is given on p. 39, 1897.

Thiele points out that the variance of the estimates of the cumulants of higher order is relatively large. For the normal distribution he finds (1889, p. 64)

$$\kappa_2(h_r) \simeq (r!)\,\kappa_2^r\,/n \quad \text{for } n \text{ large.}$$

Looking back at the above results one may ask two questions: (1) How did Thiele find the recursion formula (4.1) defining the cumulants? (2) Why did he prefer the cumulants for the central moments? He does not give any answer to the first question. For the second question it is clear from his comments that the first reason was the simple property of the cumulants for the normal distribution and also the simple extension (4.2) to the Type A series. Furthermore, the addition theorem (4.3) and its usefulness in proving asymptotic normality of estimators were decisive for him.

Ten years after the publication of his book Thiele finally succeeded in finding the general definition of the cumulants. In a short paper "Om Iagttagelseslærens Halvinvarianter" from 1899 he defines the cumulants by the equation

$$\exp\{\kappa_1 t + \kappa_2 t^2/2! + \dots\} = 1 + \mu_1' t + \mu_2' t^2/2! + \dots = \int e^{tx} f(x)\,dx, \quad (4.5)$$

where $f(x)$ denotes the density for the random variable with cumulants equal to $\kappa_1, \kappa_2, \dots$.

Equating the coefficients of t^i on both sides of (4.5) he finds

$$\frac{\mu_i'}{i!} = \sum_{r=1}^{i} \Sigma \frac{1}{a!}\left(\frac{\kappa_\alpha}{\alpha!}\right)^a \frac{1}{b!}\left(\frac{\kappa_\beta}{\beta!}\right)^b \cdots \frac{1}{d!}\left(\frac{\kappa_\delta}{\delta!}\right)^d \qquad (4.6)$$

for $i = a\alpha + b\beta + \ldots + d\delta$ and $r = a + b + \ldots + d$.

Taking logarithms on both sides of (4.5) he obtains similarly

$$\frac{\kappa_i}{i!} = \sum_{r=1}^{i} (-1)^{r-1}(r-1)! \, \Sigma \frac{1}{a!}\left(\frac{\mu_\alpha'}{\alpha!}\right)^a \frac{1}{b!}\left(\frac{\mu_\beta'}{\beta!}\right)^b \cdots \frac{1}{d!}\left(\frac{\mu_\delta'}{\delta!}\right)^d \qquad (4.7)$$

for $i = a\alpha + b\beta + \ldots + d\delta$ and $r = a + b + \ldots + d$.

Finally, he obtains the recursive definition (4.1) by differentiating (4.5) and equating coefficients of equal powers of t.

Next he turns to the operational properties of the cumulants.

He remarks that t in (4.5) may be replaced by any suitably chosen operator and that an obvious choice is $t = -D$, where D denotes differentiation. Noting that $e^{-aD} f(x) = f(x-a)$ he proves that

$$e^{bD^2/2} f(x) = \int f(x + u\sqrt{b})\,\phi(u)\,du \, . \qquad (4.8)$$

Setting $f(x) = \phi(x/\sigma)/\sigma$ he finds that the effect of the operator $\exp(bD^2/2)$ on a normal density is to add b to the variance. (It is easy to see from his formula that this is true for any distribution).

Applying the equation (4.5) with $t = -D$ to the normal distribution he finds

$$\exp\{-\kappa_1 D + \kappa_2 D^2/2! - \ldots\}(2\pi\sigma^2)^{-\frac{1}{2}}\exp\{-(x-\xi)^2/2\sigma^2\}$$

$$= \int f(y)\exp\{-y D\}(2\pi\sigma^2)^{-\frac{1}{2}}\exp\{-(x-\xi)^2/2\sigma^2\}\,dy$$

$$= \int f(y)(2\pi\sigma^2)^{-\frac{1}{2}}\exp\{-(x-\xi-y)^2/2\sigma^2\}\,dy\,.$$

Writing $y = (x-\xi) + (y-x+\xi)$ and replacing $f(y)$ in the integral above by its Taylor expansion he obtains

$$f(x) = \exp\{-(\kappa_1-\xi)D + (\kappa_2-\sigma^2)D^2/2! - \kappa_3 D^3/3! + \ldots\} \cdot (2\pi\sigma^2)^{-\frac{1}{2}}\exp\{-(x-\xi)^2/2\sigma^2\}, \qquad (4.9)$$

i.e. a series expansion of a density $f(x)$ with given cumulants in terms of a normal density and its derivatives. A particularly simple result is obtained by choosing $\xi = \kappa_1$ and $\sigma^2 = \kappa_2$.

He finally states the general formula connecting two densities having the same mean and variance

$$f^*(x) = \exp\{-(\kappa_3^* - \kappa_3)D^3/3! + (\kappa_4^* - \kappa_4)D^4/4! - \ldots\}\, f(x). \qquad (4.10)$$

He adds that one has to check that the series is convergent and gives an example of a divergent series.

In his book from 1903 he uses (4.5) for defining the cumulants and he is therefore able to simplify some of his previous proofs. He does not give the general formulas (4.6) and (4.7) neither does he mention the operational properties. However, he gives a reference to his 1899 paper on p. 49, 1903.

K.Pearson and R.A. Fisher have both commented on Thiele's work.

In his fundamental paper on "Skew Variation in Homogeneous Material" K. Pearson (1895) comments on Thiele's 1889 book but he seems to have overlooked Thiele's most important results since he writes: "They (i.e. the cumulants) are not used, however, to discriminate between various types of generalized curves, nor to calculate the constants of such types".

R.A. Fisher (1928) begins his paper on cumulants and k-statistics by the definition (4.5) without referring to Thiele and furthermore he ascribes (4.4) to Student; it is of course originally due to Gauss (1823, Art. 39). Although it is evident from Fisher's early writings that he was not familiar with the Continental literature it is rather odd that he did not know the English version of Thiele's book, in particular as it is mentioned by Whittaker and Robinson (1924, pp. 171-172). Fisher also derived formulas analogous to (4.6) and (4.7).

The paper by Cornish and Fisher (1937), in which the name cumulant is introduced instead of half-invariant, again starts from (4.5) and goes on to find the operational properties of the cumulants and derives (4.9) which by inversion gives the Cornish-Fisher expansion. They do not mention Thiele.

Fisher gives an (unfair) evaluation of Thiele's work in his "Statistical Methods" (Sixth ed., 1936, p. 22 and pp. 76-78. I have not checked whether these remarks also occur in previous editions.) The interested reader should look for himself.

5. Estimation methods.

Between the times of Gauss and Fisher it was customary to state the result of a statistical analysis in the form of an estimate, t say, and its empirical standard error, $\hat{\sigma}_t$ (or the probable error 0.67449 $\hat{\sigma}_t$), usually as $t \pm \hat{\sigma}_t$. This statement implied that t was approximately normally distributed with standard deviation $\hat{\sigma}_t$. Tests and confidence intervals were then cumputed in the usual manner.

Also Thiele used this method with the additional remark that one should take the third and fourth cumulant into account if necessary.

As mentioned above Thiele used the method of moments in 1878. Let us briefly consider this method before discussing Thiele's principle of estimation. The method of moments uses the first r empirical moments about the origin, $m_1', \ldots, m_r'$, as estimates of the analogously defined theoretical moments $\mu_1', \ldots, \mu_r'$. We may, however, just as well use the empirical central moments, $m_1', m_2, \ldots, m_r$, or the empirical cumulants, $h_1, \ldots, h_r$, as estimates of the corresponding theoretical quantities, $\mu_1', \mu_2, \ldots, \mu_r$ or $\kappa_1, \ldots, \kappa_r$, respectively, because of the one-to-one relations between the three sets of symmetric functions and the analogous relations between the theoretical quantities, i.e. the method of moments has the important property that the same estimates are obtained whatever set of symmetric functions is used as starting point. It should also be noted that the relations referred to are non-linear, at least for symmetric functions of higher order.

Obviously, Thiele wanted to improve the method of moments and in 1889, p. 63, he postulates that the best method of estimation is to set the empirical cumulant equal to its expectation. We shall quote some of his remarks on this matter from 1903, pp. 47-48.

"The struggle for life, however, compels us to consult the oracles. But the modern oracles must be scientific;... It is hardly possible to propose more satisfactory principles than the following: The mean value of all available repetitions can be taken directly, without any change, as an approximation to the presumptive mean. If only one observation without repetition is known, it must itself, consequently, be considered an approximation to the presumptive mean value.... If necessary, we complete our predictions with the mean errors and higher half-invariants,... The ancient oracles did not release the questioner from thinking and from responsibility, nor do the modern ones; yet there is a difference in the manner."

Applying this principle to the empirical cumulants Thiele arrives at the following method of estimation. Let $E\{h_i\} = f_i(\kappa_1,\dots,\kappa_i)$. Considering h_i as the best estimate of $E\{h_i\}$ the estimates $\hat{\kappa}_1,\dots,\hat{\kappa}_r$ of the theoretical cumulants are obtained by solving the r equations $h_i = f_i(\hat{\kappa}_1,\dots,\hat{\kappa}_i)$ $i = 1,\dots r$. Thiele gives the solution for $r = 5$, (1889, p. 63). He does not mention the fact that the estimates are biased, apart from the first three. He does not give formulas for the variance of the estimates but remarks that the main term of the variance may be found by means of the variances of the h's, which he has previously derived.

Thiele did not prove his postulate that h_i is the best estimate of $E\{h_i\}$. To-day we know that he was right in the sense that any symmetric function is the minimum variance estimate among unbiased estimates of its expectation.

As pointed out by Steffensen (1923, 1930) Thiele's method of estimation leads to a dilemma (compared with the method of moments) because other estimates of the cumulants will be obtained if we start from the central moments, say, using m_i as estimate of $E\{m_i\}$. The explanation of this fact is, of course, that the relation between the moments and the cumulants is non-linear so that the conversion formulas are destroyed by introducing unbiasedness of one particular symmetric function.

After Steffensen's (1923) discussion Tschuprow (1924, pp. 468-472) turned the problem around by asking for an unbiased estimate of a given parameter. (This has presumably been done by many other authors, but Tschuprow has a fairly general formulation of the problem.) He did not, however, solve the problem for the cumulants.

Bertelsen (1927, p. 144) found the first four symmetric functions, k_i say, such that $E\{k_i\} = \kappa_i$ for $i = 1,\ldots,4$. Finally, Fisher (1928) by giving the general rules for finding the k-statistics solved the problem and thus completed the work of Thiele. As explained by Cornish and Fisher (1937) Thiele's formula (4.7) is the natural starting point for constructing the k-statistics. Fisher points out that the requirement $E\{k_i\} = \kappa_i$ leads to manageable formulas for the k's which is not the case for the analogous requirement $E\{q_i\} = \mu_i$, say, for the determination of q_i.

Steffensen indicates that Thiele considered his method of estimation as a generalization of the method used by Gauss for estimating the variance. This interpretation assumes that Gauss started from the second empirical moment (or cumulant), found its expectation $E\{h_2\} = \sigma^2(n-1)/n$ and solved the equation $h_2 = \hat{\sigma}^2(n-1)/n$ to find $\hat{\sigma}^2$. However, one may just as well say that Gauss was looking for a symmetric function of order two with expectation σ^2 and this interpretation will lead to Tschuprow's procedure.

Gauss did not formulate a principle, he just solved the problem at hand. Consider the linear model and let Q denote the sum of the n squared residuals. Gauss (1823, Article 38) proved that $E\{Q\} = (n-m)\sigma^2$, where m denotes the number of parameters, and formulated his conclusion nearly as follows: "The value of Q considered as a random variable may be larger or smaller than the expected value but the difference will be of less importance the larger the number of observations so that one may use $\sqrt{Q/(n-m)}$ as an approximate value for σ". Hence, Gauss derived an unbiased estimate of σ^2 but stated his result as a biased estimate of σ. Of course, the standard deviation is the quantity of practical interest.

Gauss required unbiasedness only for the "natural" parameters of the model, i.e. the expectations and the variance, presumably because they combine linearly. There is no doubt that Thiele considered the cumulants as the "natural" symmetric functions to use exactly for this reason, see (4.3). Hence, if any symmetric function is to be chosen for unbiased estimation then it should be the cumulant. Thiele made the mistake to start from the empirical cumulants instead of the theoretical.

Without going into details with the history of the χ^2 distribution it seems reasonable to put on record that Oppermann (1863) in a query states that he has derived the distribution of $s^2 = \Sigma (x_i - \bar{x})^2/(n-1)$ for n normally distributed observations, the proof being by induction. Unfortunately he does not give his result because he does not like the method of proof. Instead he asks the following question: Let f(y) denote the density for the distribution of s^2 so that

$$\int_0^y f(t)dt = (2\pi)^{-n/2} \underset{0 < s^2 < y}{\int \ldots \int} \exp\{- \Sigma x_i^2/2\}\, dx_1 \ldots dx_n \ .$$

How is f(y) to be found by evaluation of this integral? He received no answer.

It is well-known that Abbe (1863) in the same year solved the simpler problem of finding the distribution of $\Sigma x_i^2/n$, assuming that $E\{x_i\} = 0$, by evaluating the corresponding integral and that Helmert (1876 a,b) found the distribution of m_2. It is a peculiar fact that Helmert did not present this distribution in his book (1907).

Thiele (1903, p. 46) derived the moment-generating function of h_2 ($= m_2$) but did not go any further.

6. The linear model with normally distributed errors

Thiele points out that the observations, apart from the errors, usually will be unknown functions of certain parameters. It is therefore necessary to consider the following three problems: (1) to formulate a hypothesis about the means and the error distribution, (2) to estimate the parameters in this model, and (3) to criticise the model by means of an analysis of the residuals.

After some very general remarks he states that the only problem which so far has been solved satisfactorily is the one where the means are linear functions of the parameters (or the original hypothesis may be linearized) and the errors are independent and normally distributed. He also adds that the problem of estimation is a technical matter, whereas the specification and the criticism are the important problems.

Thiele could not accept the proofs of Gauss (1809) and Andræ (1867) of the method of least squares because they were based on a uniform prior distribution and Thiele was a non-Bayesian, see (1889, p. 77). He was, however, of the same opinion as Andræ that the method of least squares should only be used for normally distributed observations. Nevertheless, it is rather strange that he does not comment on the second proof by Gauss (1821) based on the principle of minimum variance.

Contrary to Laplace, Gauss and Oppermann, who had used minimization of a loss function to find the best estimate Thiele wanted to derive estimates of the parameters in the linear model with normally distributed errors from "self-evident" principles, i.e. without the use of a loss function. His great invention was what we to-day call the canonical form of the linear hypothesis. On p.68, 1889 he writes as follows:"We shall first discuss a special case, which does not occur in practice, namely the case where the observations fall into two groups so that in the one group every observation determines its unknown mean whereas in the other group the expected value of each observation is a specified number, given by theory and independent of the means in the first group". (Similar considerations may be found on pp. 66-67, 1903).

Hence, in modern terminology Thiele's special case may be formulated as follows. Let $y_1,\ldots,y_n$ be independent and normally distributed with means $E\{y_i\} = \eta_i$ for $i = 1,\ldots,m$, the η's being unknown, $E\{y_i\} = \eta_{i0}$ for $i = m+1,\ldots,n$, the η_0's being given numbers, and unknown variance σ^2.

He then states as evident that y_i should be used as estimate of η_i for $i = 1,\ldots,m$, and since $E\{(y_i - \eta_{i0})^2\} = \sigma^2$ for $i = m+1,\ldots,n$ that

$$s^2 = \sum_{i=m+1}^{n} (y_i - \eta_{i0})^2/(n-m)$$

should be used as estimate of σ^2. Furthermore, the criticism of the model should be based on the n-m errors.

After having solved the estimation problem for the canonical form Thiele goes on to show how the general linear model may be transformed to the canonical form. He points out that the adequate mathematical tool is the theory of orthogonal transformations and gives an exposition of this theory adjusted to the needs of statistics.

We shall give an account of the most important of Thiele's results, keeping to his ideas in the proofs but using matrix notation for brevity. Let $(y_1,\ldots,y_n) = Y'$ denote n independent and normally distributed random variables with mean $(\eta_1,\ldots,\eta_n) = \eta'$ and common variance σ^2. In Thiele's exposition the variances were supposed to be proportional to σ^2 with known proportionality constants. We shall, however, keep to the simpler model since the generalization is trivial.

Thiele called independent linear functions of the observations for "free functions". We shall summarize his most important results on free functions without giving the proofs.

Let $A = (A_1,\ldots,A_m)$ denote m given vectors of dimension n, and let similarly $B = (B_1,\ldots,B_r)$.

Two functions $A_1'Y$ and $B_1'Y$ are said to be independent (free) if and only if $A_1'B_1 = 0$. The definition is based on the requirement that $V(A_1'Y + B_1'Y) = V(A_1'Y) + V(B_1'Y)$.

Two systems of functions $A'Y$ and $B'Y$ are said to be independent if any function in the one system is independent of all the functions in the other system, i.e. if $A'B = 0$.

Any given function $B_1'Y$ may uniquely be written as a sum $B_1'Y = B_0'Y + L'A'Y$, say, where the first component $B_0'Y$ is independent of the m given functions $A'Y$, which means that $L = (A'A)^{-1}A'B_1$ and $B_0 = B_1 - AL$. In particular, for two functions $A_1'Y$ and $B_1'Y$ we have that $B_0'Y$ is independent of $A_1'Y$ where

$$B_0'Y = B_1'Y - (A_1'B_1)A_1'Y/(A_1'A_1).$$

From a system of m functions given by the coefficients $A_1,\ldots,A_m$ we may construct a system of m independent functions using the method above. Using $A_1'Y$ as the first we find the coefficients of m-1 independent functions as

$$A_{2i} = A_i - A_1(A_1'A_i)/(A_1'A_1), \text{ since } A_1'A_{2i} = 0 \text{ for } i = 2,\ldots,m.$$

Similarly we construct m-2 functions independent of $A_1'Y$ and $A_{22}'Y$ by means of the coefficients

$$A_{3i} = A_{2i} - A_{22}(A_{22}'A_{2i})/(A_{22}'A_{22}), \text{ since } A_{22}'A_{3i} = 0 \text{ for } i = 3,\ldots,m,$$

and so on. This procedure is identical with Gauss' algorithm for solving the normal equations.

A system of n independent functions is called a complete system. $Z = Q'Y$ where $Q'Q = I$ is a complete system. Let Q be partioned into two matrices A and B so that

$$Z = \begin{pmatrix} A'Y \\ B'Y \end{pmatrix} \qquad \begin{matrix} m \text{ functions} \\ r = n-m \text{ functions.} \end{matrix}$$

It follows from the results above that any linear function of Y may be written as a sum of two independent functions, the first being independent of A'Y and the second independent of B'Y. In particular, this is true for each of the elements of Y.

Thiele naturally also mentions that distances are invariant under orthogonal transformations.

Thiele writes that Oppermann and Helmert have pointed out the usefulness of working with independent functions. Helmert (1872, p. 164) introduced the concept of an equivalent system of linear functions defined by the requirement that it should give the same estimates as the original observations and as the most important example he used the reduced normal equations whose right-hand sides are independent functions of the observations. It was, however, Thiele who developed a general theory of free functions.

As usual at that time Thiele considers two formulations of the linear model. In the first case, which according to Gauss is called "Adjustment by Correlates" the problem is to estimate η and σ^2 under the restriction $A_1'\eta = \zeta_{10}$, where A'_1 denotes a given (r x n) matrix of rank $r = n-m$, $0 < r < n$, and ζ_{10} is a given vector of dimension r. In the second case, called "Adjustment by Elements", it is assumed that $\eta = X\beta$, where X denotes a given (n x m) matrix of rank m and β is an unknown vector of dimension m.

Adjustment by Correlates

In the following we shall introduce partitioned vectors and matrices of dimensions r,m,(r x r),(r x m) etc. without each time defining the symbol and specifying its dimension because it follows from the context.

Let us supplement the r linear restrictions $A_1'\eta = \zeta_{10}$ by m linear functions $A_2'\eta = \zeta_2$ which we want to estimate, i.e. A_2' is known whereas ζ_2 is unknown. The linear model

$$Z = A'Y = \begin{pmatrix} A_1'Y \\ A_2'Y \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \quad \text{and } E\{Z\} = \begin{pmatrix} \zeta_{10} \\ \zeta_2 \end{pmatrix}$$

will in general not be in canonical form and we therefore transform to

$$T = BZ = BA'Y,$$

and, to get independence, we require that

$$BA' = Q' \quad \text{and} \quad Q'Q = I.$$

It is known from Gauss' algorithm for solving the normal equations that there exists a lower-triangular matrix B satisfying the equation $BA'AB' = I$. Writing

$$B = \begin{pmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{pmatrix}$$

we find

$$T = BZ = \begin{pmatrix} B_{11}Z_1 \\ B_{21}Z_1 + B_{22}Z_2 \end{pmatrix} = \begin{pmatrix} Q_1'Y \\ Q_2'Y \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

and

$$\theta = E\{T\} = B\,\zeta = \begin{pmatrix} B_{10}\ \zeta_{10} \\ B_{21}\ \zeta_{10} + B_{22}\zeta_2 \end{pmatrix} = \begin{pmatrix} Q_1'\eta \\ Q_2'\eta \end{pmatrix} = \begin{pmatrix} \theta_{10} \\ \theta_2 \end{pmatrix}$$

where the elements of T are independent and normally distributed random variables, the first r having known means and the last m having unknown means.

Having found the canonical form any estimation problem may be solved by expressing the function to be estimated, $L'\eta$ say, in terms of θ_{10} and θ_2 and replacing θ_2 by T_2. Equivalently we may start from $L'Y$, transform to a linear combination of T_1 and T_2 and replace T_1 by its true value θ_{10}.

Let us first investigate the estimation of ζ_2 using the inverse transformation of the one above. Setting

$$B^{-1} = \begin{pmatrix} B^{11} & 0 \\ B^{21} & B^{22} \end{pmatrix} = \begin{pmatrix} B_{11}^{-1} & 0 \\ -B_{22}^{-1}B_{21}B_{11}^{-1} & B_{22}^{-1} \end{pmatrix}$$

we find

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} B^{11}T_1 \\ B^{21}T_1 + B^{22}T_2 \end{pmatrix}$$

which gives

$$\hat{\zeta}_2 = B^{21}\theta_{10} + B^{22}T_2 = Z_2 + B_{22}^{-1}B_{21}(Z_1 - \zeta_{10})$$

as estimate of ζ_2. Using that

$$Z_2 - \hat{\zeta}_2 = B^{21}(T_1 - \theta_{10})$$

Thiele proves that

$$V\{Z_{2i} - \hat{\zeta}_{2i}\} = V\{Z_{2i}\} - V\{\hat{\zeta}_{2i}\} ,$$

where Z_{2i} denotes the ith element of Z_2. Hence, the variance of $\hat{\zeta}_{2i}$ is smaller than the variance of Z_{2i}.

The estimate of η is obtained from

$$Y = Q_1T_1 + Q_2T_2$$

which leads to

$$\hat{\eta} = Q_1\theta_{10} + Q_2T_2 = Y - Q_1(T_1 - \theta_{10}) .$$

Hence, to compute $\hat{\eta}$ we need only find T_1 and θ_{10}. Note that $A_1'\hat{\eta} = \zeta_{10}$.

From

$$Y - \hat{\eta} = Q_1(T_1 - \theta_{10})$$

it follows that

$$V\{y_i - \hat{\eta}_i\} = \sigma^2 \sum_{j=1}^{r} q_{ij}^2 .$$

Since

$$V\{y_i\} = \sigma^2 \quad \text{and} \quad V\{\hat{\eta}_i\} = \sigma^2 \sum_{j=r+1}^{n} q_{ij}^2$$

we have

$$V\{y_i - \hat{\eta}_i\} = V\{y_i\} - V\{\hat{\eta}_i\} \quad ,$$

so that $V\{\hat{\eta}_i\} \leqq V\{y_i\}$.

Considering the sum of squared residuals we get

$$(Y - \hat{\eta})'(Y - \hat{\eta}) = (T_1 - \theta_{10})'(T_1 - \theta_{10})$$

and since the expectation of the right-hand side equals $r\,\sigma^2$ the estimate of σ^2 becomes

$$s^2 = (Y - \hat{\eta})'(Y - \hat{\eta})/(n-m) .$$

For the sum of squared errors under the restriction $A_1'\eta = \zeta_{10}$ we have

$$(Y - \eta)'(Y - \eta) = (T - \theta)'(T - \theta) = (T_1 - \theta_{10})'(T_1 - \theta_{10}) + (T_2 - \theta_2)'(T_2 - \theta_2) .$$

Hence, the minimum with respect to η is obtained for $\theta_2 = T_2$, i.e.

$$\min_\eta (Y - \eta)'(Y - \eta) = (T_1 - \theta_{10})'(T_1 - \theta_{10})$$

which according to the result above gives $\eta = \hat{\eta}$. This property has led to the name "the method of least squares".

To test a hypothetical value of σ^2 the quantity $(n-m)s^2/\sigma^2$ is compared with its expectation $(n-m)$ taking the standard deviation $\sqrt{2(n-m)}$ into account.

Thiele continually stresses the importance of the criticism of the model. He recommends three procedures: (1) graphical analysis of the residuals, (2) analysis of the variation of signs of residuals, and (3) comparison of residuals with their standard deviation. For all three procedures he points out the importance of using rational subgroups of observations to detect systematic deviations from the assumptions.

Of course, if T_1 and θ_{10} have been computed then the n-m differences should be investigated. However, usually the criticism is based on $y_i - \hat{\eta}_i$ and the corresponding $V_e\{y_i - \hat{\eta}_i\}$, say, where σ^2 has been replaced by s^2. For subgroups Thiele uses the approximate test procedure to compare $\Sigma(y_i - \hat{\eta}_i)^2$ with its estimated mean $\Sigma\, V_e\{y_i - \hat{\eta}_i\} = M$, say, using $\sqrt{2M}$ as standard deviation. He used this form for criticism also in cases where the parameters had been estimated by other means than the method of least squares, see Thiele (1871) on the graduation of mortality data.

Finally we note the following special case of the general theory: Let $Z_1 = A_1'Y$ and $Z_2 = A_2'Y$ be independent and let $E\{Z_1\} = \zeta_{10}$ be known. The least squares estimate of $\zeta_2 = A_2'\eta$ is then Z_2. This result follows from the assumption that $A_1'A_2 = 0$ which leads to $B_{21} = 0$ so that $Z_1 = B_{11}^{-1}\, T_1$ and $Z_2 = B_{22}^{-1}\, T_2$.

Adjustment by elements

Let us partition the n equations $\eta = X\beta$ into the first n-m and the last m as indicated by

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} X_1\beta \\ X_2\beta \end{pmatrix}$$

assuming that the rank of the (m x m) matrix X_2 equals m. Eliminating β we get $\beta = X_2^{-1}\eta_2$ and $\eta_1 = X_1X_2^{-1}\eta_2$. Setting $A_1' = (I, -X_1X_2^{-1})$ and

$$Z_1 = A_1'Y = Y_1 - X_1X_2^{-1}Y_2 ,$$

where Y has been partitioned analogously to η, we find $E\{Z_1\} = A_1'\eta = 0$. Hence, the estimation problem may be solved by means of the methods given in the previous section putting $\zeta_{10} = 0$.

Noting that $A_1'X = 0$ it follows that the simplest choice of Z_2 is $Z_2 = X'Y$ which makes Z_1 and Z_2 independent. The least squares estimate of $\zeta_2 = E\{Z_2\} = X'X\beta$ is then Z_2 so that the equation for $\hat{\beta}$ becomes

$$X'X\hat{\beta} = X'Y ,$$

i.e. the normal equation.

Using Gauss' algorithm we multiply the equation by a lower-triangular matrix, G say, which leads to

$$GX'X\hat{\beta} = GX'Y, \quad \text{where} \quad GX'XG' = D = \text{diag}(d_1,\ldots,d_m).$$

(A detailed discussion of the Gaussian algorithm in matrix notation has been given by Henry Jensen (1944)). Thiele observes that $GX'Y = T$, say, represent m independent functions and that $V\{t_i\} = \sigma^2 d_i$.

Introducing λ as new parameter by the transformation $\beta = G'\lambda$ the reduced normal equation becomes $D\,\hat{\lambda} = T$ so that the elements of $\hat{\lambda}$ are independent and $V\{\hat{\lambda}_i\} = \sigma^2/d_i$. Transforming backwards Thiele finds the properties of $\hat{\eta} = X\hat{\beta} = XG'\,\hat{\lambda}$ and of $P'\hat{\beta} = P'G'\hat{\lambda}$ by means of the properties of $\hat{\lambda}$.

Based on the equation

$$(Y - \hat{\eta})'(Y - \hat{\eta}) = Y'Y - T'D^{-1}T$$

Thiele remarks that if η is a series with an unknown number of terms then an orthogonal transformation makes it possible to judge the importance of each term (i.e. test the significance of each new coefficient included) by comparing $t_i^2/(\sigma^2 d_i)$ with its expectation which equals 1. He illustrates this with an example using orthogonal polynomials.

The singular case. The method of fictitious observations.
Let the rank of X be less than m so that there is no unique solution of the normal equations. Thiele points out that a practical method for handling this case is to introduce fictitious observations so as to make the solution determinate.

Let the rank of X be m-1. Thiele introduces the enlarged model

$$Y = X\beta + \varepsilon$$

$$z = c'\beta + \delta,$$

where z denotes a fictitious observation with mean $c'\beta$ and variance σ^2, the row vector c' being linearly independent of the rows of X. The normal equation for the enlarged model is

$$(X'X + cc')\;\hat{\beta}_0 = X'Y + cz\ .$$

We may then ask under what conditions the unique solution $\hat{\beta}_0$ will satisfy the original normal equation $X'X\hat{\beta} = X'Y$. Eliminating $X'Y$ we find

$$X'X(\hat{\beta} - \hat{\beta}_0) = c(c'\hat{\beta}_0 - z)$$

so that $X'X\hat{\beta}_0 = X'X\hat{\beta}$ if and only if $c'\hat{\beta}_0 = z$, which is Thiele's condition for using fictitious observations, see p. 94, 1889. The estimate $\hat{\beta}_0$ obviously depends on the fictitious observation z. Hence, we are able to estimate only linear functions $P'\beta$ for which $P'\hat{\beta}_0$ does not depend on z, i.e. what to-day is called estimable functions. Thiele uses this method in the two-way analysis of variance.

The method used to-day is to introduce identifiability constraints, which may be considered as a special case of Thiele's method.

An extension of the linear model

Consider a stochastic process (Brownian motion) which at times $t_0, t_1, \ldots, t_n$ takes on the values $z_0, z_1, \ldots, z_n$ and assume that $z_{i+1} - z_i$, $i = 0,1,\ldots,n-1$, is normally distributed with mean 0 and variance $\sigma^2(t_{i+1} - t_i) = \sigma^2 k_i^2$, say. Suppose that the observation y_i for given value of z_i is normally distributed with mean z_i and variance ω^2. The problem is to estimate the z's, σ^2 and ω^2. This is Thiele's (1880) formul·tion of a model which he used to describe the "quasi-systematic" variations of an instrument-constant.

Starting from preliminary values of σ^2 and ω^2 Thiele solves the problem of estimating the z's by minimizing

$$\sum_{i=0}^{n}\left(\frac{y_i - z_i}{\omega}\right)^2 + \sum_{i=0}^{n-1}\left(\frac{z_{i+1} - z_i}{\sigma k_i}\right)^2 .$$

This leads to the normal equations for the determination of the estimates, $\hat{z}_i$. Also $V\{\hat{z}_i\}$ and $V\{\hat{z}_{i+1} - \hat{z}_i\}$ are found.

Noting that the estimate of σ^2 in the linear model equals

$$s^2 = \Sigma\ (y_i - \hat{\eta}_i)^2/(n-m) = \Sigma\ (y_i - \hat{\eta}_i)^2/(n - \Sigma\ V\{\hat{\eta}_i\}\sigma^{-2})$$

Thiele proposes to use

$$\Sigma\ (y_i - \hat{z}_i)^2/(n+1- \Sigma\, V\{\hat{z}_i\}\omega^{-2})$$

as estimate of ω^2 and

$$\Sigma\ (\hat{z}_{i+1} - \hat{z}_i)^2 k_i^{-2} / (n - \Sigma\ V\{\hat{z}_{i+1} - \hat{z}_i\}(\sigma\, k_i)^{-2})$$

as estimate of σ^2. The problem is then solved by iteration.

Finally Thiele extends the model by letting the mean of y_i for given z_i be a linear function of $x_1,\ldots,x_m$.

7. Analysis of variance

One-way classification

This problem is discussed only in the 1897 edition, see pp. 41-44. Thiele formulates very clearly the basic ideas of the analysis of variance and carries out such an analysis for an example with 20 groups and 25 observations per group.

He first finds the means and variances within groups, $\bar{x}_i$ and s_i^2 , $i = 1,\ldots,20$, and the average variance within groups, s_w^2 say. As a preliminary investigation he computes the 20 standardized deviations $(\bar{x}_i - \bar{x})\sqrt{25}/s_w$ and remarks that the variation is not significant because all these values lie between - 2.1 and + 2.1. He then adds the following important remark: "The most efficient test for the hypothesis is obtained by comparison of the variance between groups and the variance within groups since any systematic variation in the true means will increase the variance between groups". He computes $d = s_b^2 - s_w^2$, where s_b^2 denotes the variance between groups and concludes that this difference is insignificant because it is smaller than the standard deviation of s_b^2, which he estimates by means of (4.4) setting $\kappa_4 = 0$. Of course, he should have estimated the standard deviation of d but he has presumably not found it worth while to do so.

After having carried out an analysis of the variation of the means he gives an analogous analysis of the variances within groups to test the hypothesis that the true variances are equal.

Two-vay classification without interaction

Thiele treats this model as a special case of the general linear model and derives estimates of the parameters in the form used to-day. His starting point (1889, pp. 96-99) is the following problem. Consider the observation of the passage times for k stars over m parallel threads. The true values of the observations may then be written as

$$E\{y_{ij}\} = \eta_{ij} = \alpha_i + \beta_j/h_i \ , \ i = 1,\ldots,k \text{ and } j = 1,\ldots,m,$$

where h_i denotes a known velocity, and the y's are supposed to be independent and normally distributed with $V\{y_{ij}\} = \sigma^2 q_i$, where q_i is known. For simplicity we shall set $q_i = 1$ in the following.

Corresponding to the special structure of the design matrix Thiele develops a computational technique for the construction of the normal equations, called the method of partial eliminations, by which he gets the unknowns separated. Furthermore, he adds a fictitious observation z to make the solution determinate. He then finds the estimates

$$\hat{\alpha}_i = \bar{y}_{i.} - z/mh_i$$

and

$$\hat{\beta}_j = w^{-1} \Sigma_i h_i^{-1}(y_{ij} - \bar{y}_{i.}) + z/m, \quad \text{for} \quad w = \Sigma_i h_i^{-2}.$$

Thiele remarks that α_i and β_j are not estimable, as the estimates depends on the fictitious observation z, but that differences such as $h_i\alpha_i - h_\nu\alpha_\nu$ and $\beta_j - \beta_\mu$ are estimable. Generalizing, he adds that contrasts (as they are called to-day) are estimable and gives the following estimates:

$$\Sigma_i a_i h_i \hat{\alpha}_i = \Sigma_i a_i h_i \bar{y}_{i.} \quad \text{for } \Sigma_i a_i = 0, \quad \text{and } V = \sigma^2 m^{-1} \Sigma_i (a_i h_i)^2$$

and

$$\Sigma_j b_j \hat{\beta}_j = w^{-1} \Sigma_i \Sigma_j h_i^{-1} b_j (y_{ij} - \bar{y}_{i.}) \text{ for } \Sigma_j b_j = 0, \text{ and } V = \sigma^2 w^{-1} \Sigma_j b_j^2,$$

where V denotes the variance of the estimate.

He also gives the estimate $\hat{\eta}_{ij}$ and its variance, finds the variance of $y_{ij} - \hat{\eta}_{ij}$ and proves that

$$E\{\Sigma_i \Sigma_j (y_{ij} - \hat{\eta}_{ij})^2\} = \sigma^2 (m-1)(k-1)$$

which leads to the usual estimate of σ^2.

For $h_i = 1$ the formulas simplify to the standard ones used to-day.

In 1897 and 1903 he only discusses the simple model with $\eta_{ij} = \alpha_i + \beta_j$ and the same variance for all y_{ij}, see pp. 100-103, 1903.

He gives an example with 11 measurements of the abscissas of 3 points on a line, the position of the scale being different for each of the 11 measurements. Moreover, in 6 of the 11 cases one of the measurements is lacking. He derives the normal equations for this two-way analysis of variance with missing observations and expresses the solution in terms of a fictitious observation as above. Finally, he computes $\hat{\eta}_{ij}$ and s^2 and uses the ratios $(y_{ij} - \hat{\eta}_{ij})^2/V_e\{y_{ij} - \hat{\eta}_{ij}\}$, where V_e denotes the estimated variance, and sums of these over rational subgroups for the criticism of the model.

Acknowledgements.

My thanks are due to Sven Danø for information on the Thiele family and to Søren Johansen for discussions on the manuscript.

Abbe, E. (1863): Über die Gesetzmässigkeit in der Verteilung der Fehler bei Beobachtungsreihen. Dissertation, Jena 1863. See M.G. Kendall: The Work of Ernst Abbe, Biometrika (1971), 58, 369-373.

Andersen, E. (1975): Heinrich Christian Schumacher. Geodætisk Instituts Forlag, København, 209 pp.

Andræ, C.G. (1860): Udvidelse af en af Laplace i Mécanique céleste angivet Methode for Bestemmelsen af en ubekjendt Størrelse ved givne umiddelbare Iagttagelser. Oversigt Vidensk. Selsk. Forhandlinger, 1860, 1-38.

Andræ, C.G. (1867): Om den rette Begrundelse af de mindste Quadraters Methode. Den danske Gradmaaling, Vol. 1, 556-561, Kjøbenhavn.

Bertelsen, N.P (1927): On the Compatibility of Frequency Constants, and on Presumptive Laws of Error. Skandinavisk Aktuarietidskrift, 10, 129-156.

Bessel, F.W. (1838): Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler. Astronomische Nachrichten, 15, 369-404.

Cornish. E.A. and Fisher, R.A. (1937): Moments and Cumulants in the Specification of Distributions. Revue de l'Institut Intern. Statist., 4, 1-14.

Cramér, H. (1972): On the History of Certain Expansions Used in Mathematical Statistics. Biometrika, 59, 205-207.

Fisher, R.A. (1928): Moments and Product Moments of Sampling Distributions. Proc. London Math. Soc., Series 2, 30, 199-238.

Fisher, R.A. (1936): Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh, 6th Ed., 339 pp.

Gauss, C.F. (1809, 1821, 1823): Abhandlungen zur Methode der kleinsten Quadrate von Carl Friedrich Gauss. In deutscher Sprache herausgegeben von A. Börsch und P. Simon. Berlin, 1887, 208 pp.

Gnedenko, B.V. and Kolmogorov, A. N. (1954): Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Cambridge, Mass., 254 pp.

Gram, J.P. (1879): Om Rækkeudviklinger, bestemte ved Hjælp af de mindste Kvadraters Methode. Dissertation, Høst og Søn, København , 124 pp. See also Über die Entwickelung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate. J. Reine Angew. Math. 1883, 94, 41-73.

Gram, J.P. (1883): Om Beregning af en Bevoxnings Masse ved Hjælp af Prøvetræer. Tidsskrift for Skovbrug, 6, 137-198.

Gram, J.P. (1889): Om Konstruktion af Normal-Tilvæxtoversigter, med særligt Hensyn til Iagttagelserne fra Odsherred. Tidsskrift for Skovbrug, 11, 97-151.

Gram, J.P. (1910): Professor Thiele som Aktuar. Dansk Forsikringsårbog, 1910, 26-37.

Helmert, F.R. (1872): Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. Teubner, Leipzig, 348 pp. 2nd. Ed. 1907, 578 pp.

Helmert, F.R. (1876a): Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit im Zusammenhange stehende Fragen. Zeitschrift für Mathematik und Physik, 21, 192-218.

Helmert, F.R. (1876b): Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers direchter

Beobachtungen gleicher Genauigkeit. Astronomische Nachrichten, 88, 113-132.

Jensen, H. (1944): An Attempt at a Systematic Classification of Some Methods for the Solution of Normal Equations. Geodætisk Institut, Meddelelse No. 18, København, 45 pp.

Oppermann, L. (1863): En Forespørgsel (Af en Dilettant). Tidsskrift for Mathematik. Første Rk., 5, 16.

Oppermann, L. (1872): Zur Begründung der algemeinen Fehlertheorie. Methode der kleinsten Quadrate. Hamburg. 20 pp.

Pearson, K. (1895): Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. Phil. Trans. Roy. Soc. London, Series A, 186, 343-414.

Steffensen, J.F. (1923): Et Dilemma fra Iagttagelseslæren. Matematisk Tidsskrift, 1923, 72-76.

Steffensen, J.F. (1930): Some Recent Researches in the Theory of Statistics and Actuarial Science. Cambridge Univ. Press, 52 pp.

Särndal, C.E. (1971): The Hypothesis of Elementary Errors and the Scandinavian School in Statistical Theory. Biometrika, 58, 375-391.

Thiele, T.N. (1866): Undersøgelse af Omløbsbevægelsen i Dobbeltstjærnesystemet Gamma Virginis, udført tildels efter ny Methoder. Dissertation. Kjøbenhavn, 104 pp.

Thiele, T.N. (1871): En mathematisk Formel for Dødeligheden. Reitzel, Kjøbenhavn , 19 pp.

Thiele, T.N. (1873): Om en tilnærmelsesformel. Tidsskrift for Mathematik. Tredie Rk., 3,22-31.

Thiele, T.N. (1878): Bemærkninger om skæve Fejlkurver. Tidsskrift for Matematik. Fjerde Rk., 2, 54-57.

Thiele, T.N. (1880): Om Anvendelse af mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en "systematisk" Karakter. Vidensk. Selsk. Skr., 5. Rk., naturvid. og mat. Afd., 12, 5, 381-408. See also Sur la Compensation de quelques Erreurs quasi-systématiques par la Méthode des moindres Carrés. Reitzel, København, 1880.

Thiele, T.N. (1889): Forelæsninger over Almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Metode. Reitzel, København, 121 pp.

Thiele, T.N. (1897): Elementær Iagttagelseslære. Gyldendal, København, 129 pp.

Thiele, T.N. (1899): Om Iagttagelseslærens Halvinvarianter. Oversigt Vidensk. Selsk. Forhandlinger, 1899, 135-141.

Thiele, T.N. (1903): Theory of Observations. Layton, London, 144 pp. Reprinted in The Annals of Mathematical Statistics, 1931, 2, 165-308.

Tschuprow, A.A. (1924): Ziele und Wege der stochastischen Grundlegung der statistischen Theorie. Nordisk Statistisk Tidskrift, 3, 433-493.

Whittaker, E.T. and Robinson, G. (1924): The Calculus of Observations. Blackie, London 395 pp.

Zachariae, G. (1871): De mindste Kvadraters Methode. Nyborg, 234 pp. 2nd Ed., 1887, Kjøbenhavn, 270 pp.

PREPRINTS 1978

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Tjur, Tue: Statistical Inference under the Likelihood Principle.

No. 2 Hering, Heinrich: The Non-Degenerate Limit for Supercritical Branching Diffusions.

No. 3 Henningsen, Inge: Estimation in M/G/1-Queues.

No. 4 Braun, Henry: Stochastic Stable Population Theory in Continuous Time.

No. 5 Asmussen, Søren: On some two-sex population models.

No. 6 Andersen, Per Kragh: Filtered Renewal Processes with a Two-Sided Impact Function.

No. 7 Johansen, Søren & Ramsey, Fred L.: A Bang-Bang Representation for 3x3 Embeddable Stochastic Matrix.

No. 8 Braun, Henry: A Simple Method for Testing Goodness of Fit in the Presence of Nuisance Parameters.

No. 9 Lauritzen, S.L. & Speed, T.P. & Vijayan, K.: Decomposable Graphs and Hypergraphs.

No. 10 Hald, Anders: On the Statistical Theory of Sampling Inspection by Attributes.

No. 11 Darroch, J.N. & Lauritzen, S.L. & Speed, T. P.: Markov Fields and Log-linear Interaction Models for Contingency Tables.

No. 12 Speed, T.P.: A Note on Nearest-Neighbour Gibbs and Markov Probabilities.

No. 13 Andersson, Steen A.: Canonical Correlations with Respect to a Complex Structure.