

Über die
Berechnung des Mittelfehlers des
Resultates einer Linientaxierung

Von
J. W. LINDEBERG

HELSINKI 1924

Über die Berechnung des Mittelfehlers des Resultates einer Linientaxierung.

Von

J. W. LINDEBERG.

1. Bei der Linientaxierung eines Gebietes werden bekanntlich die gesuchten Masszahlen so erhalten, dass sie in bezug auf ein repräsentatives, aus äquidistanten Linien von fester Breite bestehendes Teilgebiet durch möglichst genaue direkte Beobachtungen und darauf gegründete Berechnungen bestimmt werden. Wenn auch die gemachten Beobachtungen und Berechnungen exakt wären, ist ein solches Resultat immer deshalb unsicher, weil sich das Liniengebiet in bezug auf die untersuchten Verhältnisse niemals genau so verhält wie das ganze Gebiet. Um eine Auffassung davon zu erhalten, wie grosse Fehler aus diesem Grunde zu erwarten sind, hat man in letzter Zeit allgemein angefangen, auf Wahrscheinlichkeitsbetrachtungen gegründete Berechnungen der Mittelfehler der Resultate vorzunehmen.

Solche Berechnungen dürften zum ersten Mal in der in Schweden vorgenommenen Taxierung des Län Värmland ausgeführt und publiziert worden sein.¹ Später sind ähnliche Berechnungen in Norwegen gemacht worden², und ganz vor kurzem ist der Bericht über eine in Finnland ausgeführte Taxierung erschienen³, wobei die Mittelfehler mittels eines von dem schwedischen Muster etwas abweichenden Verfahrens berechnet sind.

¹ Värmlands läns skogar jämte plan till en taxering af Sveriges samtliga skogar. Stockholm 1914.

² Taxering af Norges skoger utförd af Landsskogtaxeringen:

I Östfold fylke. Hamar 1920.

II Hedmark fylke. Kristiania 1922.

³ YRJÖ ILVESSALO, Untersuchungen über den Zustand der Privatwälder in den mittleren Teilen des Län Tavastehus. Acta Forestalia Fennica 26, Helsinki, 1923.

Das Folgende enthält eine kritische Untersuchung der bisher angewandten Methoden der Fehlerberechnung und einen Versuch, diese zu verbessern.

Eine Linientaxierung kann sehr verschiedene Verhältnisse betreffen, wie die Kubikmasse pro ha, den Waldboden in Prozenten des ganzen Areals, u. s. w. Der Kürze halber wird im folgenden immer nur vom Prozent eines Gebietes gesprochen, ohne dass die Massenerscheinung, wovon die Rede ist, näher bestimmt wird.

2. Bei der schon erwähnten Linientaxierung des Länns Värmland wurde die Berechnung des Resultates und seines Mittelfehlers wie folgt gemacht.

Es sei G das zu taxierende Gebiet, l_1, l_2, \dots, l_n die Längen der Linien, so geordnet, wie diese tatsächlich nacheinander folgten, und p_1, p_2, \dots, p_n die zu den Linien gehörigen Prozentzahlen, welche wir als exakt voraussetzen. Bezeichnet L die Gesamtlänge der Linien, so ist

$$(1) \quad P = \sum_{\mu=1}^n \frac{l_{\mu}}{L} p_{\mu}$$

das Prozent des aus den Linien zusammengesetzten Gebietes. Diese Zahl wurde als Prozent des ganzen Gebietes angenommen.

Im Bericht über die Taxierung ist die Sache zwar nicht ganz in dieser Weise erklärt. Zwecks der nachfolgenden Mittelfehlerberechnung wurden nämlich die Linien in zehn Gruppen so zusammengefasst, dass zu jeder einzelnen Gruppe soweit wie möglich Linien geführt wurden, welche über das ganze Gebiet gleichmässig verteilt waren. Wenn wir die Gesamtlängen der Linien der verschiedenen Gruppen mit L_1, L_2, \dots, L_{10} bezeichnen und P_1, P_2, \dots, P_{10} die zugehörigen Prozente sind, wurde nach dem Bericht P aus der Gleichung

$$P = \sum_{\mu=1}^{10} \frac{L_{\mu}}{L} P_{\mu}$$

berechnet. Dieser Wert ist aber offenbar mit dem Werte (1) identisch.

Aus dem Bericht kann weiter geschlossen werden, dass die Berechnung des Mittelfehlers nach der Formel

$$(2) \quad \varepsilon(P) = \sqrt{\frac{1}{9} \sum_{\mu=1}^{10} \frac{L_{\mu}}{L} (P_{\mu} - P)^2}$$

erfolgte, denn im theoretischen Kapitel desselben sind auf S. 80 als Grund der Berechnungen die beiden Formeln angegeben, die in Par. 188 von Czubers Lehrbuch der Wahrscheinlichkeitsrechnung als Formeln (4) und (10) auftreten.

Bezeichnen wir mit \bar{P} den wahren Wert des zu dem ganzen Gebiete G gehörigen Prozents, so können wir die der Formel (2) zugrunde liegenden Hypothesen wie folgt formulieren:

a) Die Zahlen P_{μ} sind voneinander unabhängige Wahrscheinlichkeitsgrössen, deren Mittelwerte sämtlich gleich \bar{P} sind.

b) Die Streuungen oder Mittelfehler der einzelnen Grössen P_{μ} sind den Quadratwurzeln aus den Längen L_{μ} umgekehrt proportional.

Von diesen Hypothesen scheint wenigstens die zweite wegen ihres speziellen Charakters Bedenken erregen zu müssen. Wenn man an die Verhältnisse in der Fehlertheorie denkt, ist es klar, dass sie erfüllt ist, wenn man sich die Struktur des Gebietes so vorstellen darf, dass das Prozent einer Linienstrecke von gewisser Länge unabhängig davon, in welchem Teile des Gebietes sie gewählt ist, immer eine Wahrscheinlichkeitsgrösse mit derselben festen Streuung ist. Eine solche Annahme scheint aber nicht möglich zu sein. Nun können zwar Verhältnisse, die der Hypothese b) entsprechen, auch in anderer Weise annähernd zustandekommen, es würde aber immer geeignet sein, den Wert des Resultates zu vermindern, wenn man gezwungen wäre, das Berechnungsverfahren auf dieselbe zu stützen. Es kann indes gezeigt werden, dass eine Abänderung der Formel (2) die aus dieser Hypothese herfließende Dunkelheit hebt. Einige theoretische Erörterungen sind zu diesem Zwecke nötig.

3. Wenn x eine Wahrscheinlichkeitsgrösse bezeichnet, werden wir im folgenden mit $M(x)$ und $\varepsilon(x)$ immer ihren Mittelwert und ihre Streuung bezeichnen.

Es seien x_1, x_2, \dots, x_n voneinander unabhängige Wahrscheinlichkeitsgrössen mit dem gemeinsamen Mittelwert A und den Streuungen $\sigma_1, \sigma_2, \dots, \sigma_n$. Ferner seien h_1, h_2, \dots, h_n positive Konstanten, die der Gleichung

$$(3) \quad \sum_{\mu=1}^n h_{\mu} = 1$$

genügen. Wir betrachten die Grösse

$$x = \sum_{\mu=1}^n h_{\mu} x_{\mu}.$$

Nach einer bekannten Regel ist

$$(4) \quad \varepsilon^2(x) = \sum_{\mu=1}^n h_{\mu}^2 \sigma_{\mu}^2.$$

Da

$$M\{(x_{\mu} - A)^2\} = \sigma_{\mu}^2,$$

ist also

$$\varepsilon^2(x) = M \left\{ \sum_{\mu=1}^n h_{\mu}^2 (x_{\mu} - A)^2 \right\}.$$

Wenn $\xi_1, \xi_2, \dots, \xi_n$ ein System von Beobachtungen der Grössen x_1, x_2, \dots, x_n bezeichnen, ist somit annähernd

$$(5) \quad \varepsilon(x) = \sqrt{\sum_{\mu=1}^n h_{\mu}^2 (\xi_{\mu} - A)^2}.$$

Es seien k_1, k_2, \dots, k_n Konstanten, welche der Gleichung

$$\sum_{\mu=1}^n k_{\mu} \sigma_{\mu}^2 = \sum_{\mu=1}^n h_{\mu}^2 \sigma_{\mu}^2$$

genügen. Dann ist offenbar auch

$$\varepsilon^2(x) = M \left\{ \sum_{\mu=1}^n k_{\mu} (x_{\mu} - A)^2 \right\},$$

und es ist näherungsweise

$$(6) \quad \varepsilon(x) = \sqrt{\sum_{\mu=1}^n k_{\mu} (\xi_{\mu} - A)^2}.$$

Falls die Streuungen σ_{μ} den Quadratwurzeln aus den Konstanten h_{μ} umgekehrt proportional sind, d. h. falls

$$(7) \quad h_1 \sigma_1^2 = h_2 \sigma_2^2 = \dots = h_n \sigma_n^2,$$

so ist wegen (3)

$$\sum_{\mu=1}^n \frac{h_{\mu}}{n} \sigma_{\mu}^2 = \sum_{\mu=1}^n h_{\mu}^2 \sigma_{\mu}^2.$$

Indem wir $k_{\mu} = \frac{h_{\mu}}{n}$ nehmen, erhalten wir aus (6)

$$(8) \quad \varepsilon(x) = \sqrt{\frac{1}{n} \sum_{\mu=1}^n h_{\mu} (\xi_{\mu} - A)^2}.$$

Diese letzte Formel ist aber nicht richtig, wenn die Gleichungen (7) nicht bestehen.

Die Anwendung der Formeln (5) und (8) setzt voraus, dass A bekannt ist. Wir wollen nachsehen, wie sich die obigen Entwicklungen modifizieren, wenn der Näherungswert

$$B = \sum_{\mu=1}^n h_{\mu} \xi_{\mu}$$

anstelle von A angewandt werden muss.

Es ist

$$(9) \quad \begin{aligned} M\{(\xi_{\mu} - B)^2\} &= M\{[-h_1(\xi_1 - A) - h_2(\xi_2 - A) - \dots \\ &\dots - h_{\mu-1}(\xi_{\mu-1} - A) + (1 - h_{\mu})(\xi_{\mu} - A) - h_{\mu+1}(\xi_{\mu+1} - A) \dots \\ &\dots - h_n(\xi_n - A)]^2\} = \\ &= h_1^2 \sigma_1^2 + h_2^2 \sigma_2^2 + \dots + h_{\mu-1}^2 \sigma_{\mu-1}^2 + (1 - h_{\mu})^2 \sigma_{\mu}^2 + \\ &\quad h_{\mu+1}^2 \sigma_{\mu+1}^2 + \dots + h_n^2 \sigma_n^2. \end{aligned}$$

Demnach wird

$$M\left\{\sum_{\mu=1}^n h_{\mu}^2 (\xi_{\mu} - B)^2\right\} = \sum_{\mu=1}^n \alpha_{\mu} h_{\mu}^2 \sigma_{\mu}^2$$

und

$$M\left\{\sum_{\mu=1}^n \frac{h_{\mu}}{n} (\xi_{\mu} - B)^2\right\} = \sum_{\mu=1}^n \beta_{\mu} \frac{h_{\mu}}{n} \sigma_{\mu}^2,$$

wo

$$\alpha_{\mu} = 1 - 2h_{\mu} + \sum_{\mu=1}^n h_{\mu}^2 \quad \text{und} \quad \beta_{\mu} = 1 - h_{\mu}.$$

Nehmen wir als Näherungswert von α_{μ} den Wert $\alpha_{\mu} = \frac{n-1}{n}$, welcher

exakt ist, falls $h_1 = h_2 = \dots = h_n$, so ergibt sich aus der ersten von den obigen Gleichungen mit Rücksicht auf (4)

$$(10) \quad \varepsilon(x) = \sqrt{\frac{n}{n-1} \sum_{\mu=1}^n h_{\mu}^2 (\xi_{\mu} - B)^2}.$$

Sind die Gleichungen (7) erfüllt, so erhalten wir aus der zweiten Gleichung, indem wir $\beta_{\mu} = \frac{n-1}{n}$ nehmen,

$$(11) \quad \varepsilon(x) = \sqrt{\frac{1}{n-1} \sum_{\mu=1}^n h_{\mu} (\xi_{\mu} - B)^2}.$$

Diese letzte Formel ist im wesentlichen die schon erwähnte Formel (10) aus Czubers Lehrbuch, die der Berechnung des Mittelfehlers in der Taxierung von Värmland zugrunde liegt. Wird (10) anstatt (11) benutzt, so ergibt sich als Mittelfehler von P

$$(12) \quad \varepsilon(P) = \sqrt{\frac{10}{9} \sum_{\mu=1}^n \left(\frac{L_{\mu}}{L}\right)^2 (P_{\mu} - P)^2}.$$

Da (10) unabhängig von der Voraussetzung b) ist, haben wir also in (12) eine Formel, die nur der Voraussetzung a) bedarf.

Obgleich nun also die Formel (2) im allgemeinen durch (12) ersetzt werden muss, ist in unserem Resultat zugleich eine Rechtfertigung der Anwendung der Formel (2) bei der Värmlandtaxierung enthalten. Die Längen L_{μ} waren nämlich hier beinahe gleich, und unter solchen Umständen führen die Formeln (2) und (12) zu Resultaten, die nicht wesentlich verschieden sind.

4. Mit der Anwendung der Formel (2) ist aber ein grösserer Übelstand verbunden, welchem nicht damit abgeholfen wird, dass man (2) durch (12) ersetzt. Die Anwendung der Formeln (10) und (11) für einen so keinen Wert von n wie $n = 10$ scheint nicht zulässig zu sein.

Um die Sache näher zu beleuchten, wollen wir die theoretischen Erörterungen des vorigen Par. wiederaufnehmen, indem wir jetzt noch die folgenden Voraussetzungen machen:

Die Grössen x_{μ} (und also auch die mit ξ_{μ} bezeichneten Zahlen) folgen dem Gauss'schen Gesetz.

Die Zahlen h_{μ} sind alle einander gleich und also gleich $\frac{1}{n}$.

Die Streuungen σ_{μ} sind alle gleich derselben Zahl σ .

Unter diesen Voraussetzungen geben die Formeln (10) und (11) beide die exakte¹ Formel

$$(13) \quad \varepsilon(x) = \sqrt{\frac{1}{n(n-1)} \sum_{\mu=1}^n (\xi_{\mu} - B)^2}.$$

Wir wollen die Grösse

$$\varepsilon \left\{ \sum_{\mu=1}^n (\xi_{\mu} - B)^2 \right\}$$

berechnen.

Aus (9) folgt wegen unserer Voraussetzungen

$$M \left\{ \sum_{\mu=1}^n (\xi_{\mu} - B)^2 \right\} = (n-1) \sigma^2.$$

Ferner ist

$$\sum_{\mu=1}^n (\xi_{\mu} - B)^2 = \frac{n-1}{n} \sum_{\mu=1}^n (\xi_{\mu} - A)^2 - \frac{2}{n} \sum_{\mu, \nu} (\xi_{\mu} - A) (\xi_{\nu} - A),$$

wo die zweite Summe rechts über alle möglichen Wertepaare μ, ν zu erstrecken ist, die die Bedingung $1 \leq \mu < \nu \leq n$ erfüllen. Führen wir die Bezeichnungen

$$a = \frac{n-1}{n} \sum_{\mu=1}^n (\xi_{\mu} - A)^2 - (n-1) \sigma^2$$

$$b = -\frac{2}{n} \sum_{\mu, \nu} (\xi_{\mu} - A) (\xi_{\nu} - A)$$

ein, so ergibt sich deshalb

$$\varepsilon^2 \left\{ \sum_{\mu=1}^n (\xi_{\mu} - B)^2 \right\} = M \{ (a + b)^2 \}.$$

Da die Grössen ξ_{μ} dem Gauss'schen Gesetz folgen, ist

$$M \{ (\xi_{\mu} - A)^4 \} = 3 \sigma^4,$$

und weiter ergibt sich

¹ Hierunter verstehe ich, dass $\varepsilon^2(x)$ exakt gleich dem Mittelwert des Ausdruckes unter der Quadratwurzel ist. Die Formeln (10) und (11) selbst sind in diesem Sinne im allgemeinen nicht exakt.

$$M(a^2) = \frac{2(n-1)^2}{n} \sigma^4,$$

$$M(ab) = 0$$

und

$$M(b^2) = \frac{2(n-1)}{n} \sigma^4.$$

Demnach wird

$$M\{(a+b)^2\} = 2(n-1)\sigma^4$$

und somit

$$\varepsilon \left\{ \sum_{\mu=1}^n (\xi_{\mu} - B)^2 \right\} = \sigma^2 \sqrt{2(n-1)}.$$

Es ist also

$$\frac{\varepsilon \left\{ \sum_{\mu=1}^n (\xi_{\mu} - B)^2 \right\}}{M \left\{ \sum_{\mu=1}^n (\xi_{\mu} - B)^2 \right\}} = \sqrt{\frac{2}{n-1}}.$$

Falls $n = 10$ ist, finden wir für dieses Verhältnis den Wert 0,47. Also ist der Mittelfehler des Ausdruckes unter der Quadratwurzel in (13) für $n = 10$ nahezu die Hälfte des Mittelwertes desselben, was uns zeigt, dass eine Bestimmung von $\varepsilon(x)$ mittels (13) für diesen Wert von n wertlos ist. Hieraus folgt, dass die Anwendung der Formeln (2) und (12) nicht zulässig ist.

Nun ist es ja möglich, dass die bei der Värmlandtaxierung erhaltenen Mittelfehler nichtsdestoweniger aus dem Grunde hinreichend gross sind, weil, wie im Berichte bemerkt worden ist, die Art, in welcher die Linien in Gruppen zusammengeführt wurden, eine Tendenz zu einer zu grossen Streuung der Zahlen P_{μ} bewirkt. Die Berechnungsmethode kann aber nicht hierauf basiert werden, und es ist mir auch nicht in anderer Weise gelungen, die Anwendung der Formeln (2) oder (12) zu rechtfertigen. Damit ein aus (10) erhaltenes Resultat eine selbständige Bedeutung als Mass der Unsicherheit von x habe, muss n doch wenigstens 50 betragen.

5. Ich gehe jetzt zu dem Verfahren über, das Herr Y. LVESSALO bei seinen oben erwähnten Mittelfehlerberechnungen angewandt hat und dessen Grundgedanke von W. CAJANUS herrührt.

¹ Dieses Resultat ist keineswegs neu. Es findet sich in viel allgemeinerer Form schon bei GAUSS, Theoria combinationis observationum erroribus minimis obnoxiae. In den üblichen Lehrbüchern der Wahrscheinlichkeitsrechnung werden aber die Konsequenzen desselben in merkwürdiger Weise vernachlässigt.

Der Verlauf der Zahlen p_{μ} wurde zuerst so graphisch dargestellt, dass jeder Zahl p_{μ} der Punkt zugeordnet wurde, dessen Abszisse μ und dessen Ordinate p_{μ} waren. Die erhaltene Punktfolge wurde sodann mittels einer mit der Hand gezeichneten Kurve K ausgeglichen, wobei versucht wurde, nur das im Verlauf der Zahlen p_{μ} unzweifelhaft Systematische im Gang der Kurve K zum Ausdruck kommen zu lassen und alle übrigen Variationen der p_{μ} mittels derselben auszugleichen. Zu jedem der anfangs gezeichneten Punkte gehört ein Punkt von K mit derselben Abszisse und der Ordinate p'_{μ} . Die Berechnung des Mittelfehlers des gemäss (1) berechneten Prozents P erfolgte nach der Formel

$$(14) \quad \varepsilon(P) = \sqrt{\frac{1}{n} \sum_{\mu=1}^n \frac{l_{\mu}}{L} (p_{\mu} - p'_{\mu})^2}.$$

Man gelangt zu dieser Formel, wenn man die folgenden Voraussetzungen macht:

a) Die Zahlen p_{μ} sind voneinander unabhängige Wahrscheinlichkeitsgrössen, deren Mittelwerte p'_{μ} sind.

b) Das wahre Prozent P des Gebietes ist gleich

$$\sum_{\mu=1}^n \frac{l_{\mu}}{L} p'_{\mu}.$$

c) Die Streuungen der Grössen p_{μ} , die wir mit $\sigma_1, \sigma_2, \dots, \sigma_n$ bezeichnen wollen, sind den Quadratwurzeln aus den Längen l_1, l_2, \dots, l_n umgekehrt proportional.

Dann ist nämlich einerseits

$$\varepsilon^2(P) = M \left\{ \left[\sum_{\mu=1}^n \frac{l_{\mu}}{L} p_{\mu} - \bar{P} \right]^2 \right\} = M \left\{ \left[\sum_{\mu=1}^n \frac{l_{\mu}}{L} (p_{\mu} - p'_{\mu}) \right]^2 \right\} = \sum_{\mu=1}^n \left(\frac{l_{\mu}}{L} \right)^2 \sigma_{\mu}^2,$$

und andererseits

$$M \left\{ \frac{1}{n} \sum_{\mu=1}^n \frac{l_{\mu}}{L} (p_{\mu} - p'_{\mu})^2 \right\} = \frac{1}{n} \sum_{\mu=1}^n \frac{l_{\mu}}{L} \sigma_{\mu}^2 = \sum_{\mu=1}^n \left(\frac{l_{\mu}}{L} \right)^2 \sigma_{\mu}^2,$$

was unmittelbar zu (14) führt.

Hinsichtlich dieser Formel gilt offenbar in Analogie mit dem, was wir in Par. 3 vorgeführt haben, dass, falls wir anstatt ihrer die Formel

$$(15) \quad \varepsilon(P) = \sqrt{\sum_{\mu=1}^n \left(\frac{l_{\mu}}{L}\right)^2 (p_{\mu} - p'_{\mu})^2}$$

benutzen, die Voraussetzung c) nicht mehr nötig ist. Es scheint aber jetzt sogar notwendig, die Formel (14) durch (15) zu ersetzen, denn da die Längen l_{μ} absehbar variieren können, kann die Anwendung der Formel (14) bei ungünstigen Umständen zu bedeutenden Fehlern im Resultate führen.

Bei der Anwendung der jetzt in Frage stehenden Methode tritt die richtige Zeichnung der Kurve K als neue wesentliche Schwierigkeit auf. Da man unter allen Umständen gezwungen ist, eine gewisse Homogenität des zu taxierenden Gebietes vorauszusetzen, sollte es keine Bedenken erregen, die Annahme zu machen, dass der Verlauf der Streuungen eine gewisse Stetigkeit aufweist. Es ist mir aber nicht gelungen, hieraus befriedigende Regeln für die Zeichnung von K abzuleiten. Die Vorschriften werden künstlich und in hohem Grade willkürlich.

Es ist kein Zweifel, dass ein geschickter und erfahrener Forscher mittels dieses Verfahrens zuverlässige Resultate erhalten kann, und es mag als ein Verdienst desselben hervorgehoben werden, dass eine eingehende Kenntnis der Struktur des Gebietes bei der Zeichnung der Kurve K verwertet werden kann, um ein sicheres Resultat zu erhalten. Es ist aber in methodischer Hinsicht sehr unbefriedigend, das Urteil darüber, wie die Kurve K gezeichnet werden soll, dem Handhaber der Berechnungen zu überlassen, und es scheint deshalb wünschenswert, die Zeichnung der Kurve durch irgendeinen anderen Vorgang ersetzen zu können.

Im folgenden soll aus der Formel (15) eine neue abgeleitet werden, worin die p'_{μ} sozusagen eliminiert sind und welche deshalb eine Berechnungsmethode gibt, der keine Willkürlichkeit mehr anhaftet. Natürlich kann dies nur in der Weise geschehen, dass gewisse Hypothesen eingeführt werden. Diese sind aber solcher Art, dass, wie mir scheint, keine grösseren Bedenken gegen sie erhoben werden können.

6. Wenn $\mu \neq \nu$, ist offenbar

$$(16) \quad M \left\{ \frac{l_{\mu}^2}{L^2} (p_{\mu} - p'_{\mu})^2 + \frac{l_{\nu}^2}{L^2} (p_{\nu} - p'_{\nu})^2 \right\} = M \left\{ \left[\frac{l_{\mu}}{L} (p_{\mu} - p'_{\mu}) - \frac{l_{\nu}}{L} (p_{\nu} - p'_{\nu}) \right]^2 \right\}.$$

Setzen wir

$$\frac{l_{\mu} + l_{\nu}}{2L} (p_{\mu} - p_{\nu}) = a$$

$$\frac{l_{\mu} - l_{\nu}}{2L} (p_{\mu} + p_{\nu} - p'_{\mu} - p'_{\nu}) = b$$

$$\frac{l_{\mu} + l_{\nu}}{2L} (p'_{\nu} - p'_{\mu}) = c,$$

so ist

$$\frac{l_{\mu}}{L} (p_{\mu} - p'_{\mu}) - \frac{l_{\nu}}{L} (p_{\nu} - p'_{\nu}) = a + b + c,$$

und also hat man

$$(17) \quad M \left\{ \left[\frac{l_{\mu}}{L} (p_{\mu} - p'_{\mu}) - \frac{l_{\nu}}{L} (p_{\nu} - p'_{\nu}) \right]^2 \right\} =$$

$$M(a^2) + M(b^2) + M(c^2) + 2M(ab) + 2M(ac) + 2M(bc).$$

Weil $M(b) = 0$ und c eine Konstante ist, findet man

$$M(b^2) = \left(\frac{l_{\mu} - l_{\nu}}{2L}\right)^2 (\sigma_{\mu}^2 + \sigma_{\nu}^2),$$

$$M(c^2) = \left(\frac{l_{\mu} + l_{\nu}}{2L}\right)^2 (p'_{\mu} - p'_{\nu})^2,$$

$$M(ac) = -\left(\frac{l_{\mu} + l_{\nu}}{2L}\right)^2 (p'_{\mu} - p'_{\nu})^2,$$

$$M(bc) = 0.$$

Um $M(ab)$ zu bestimmen, schreiben wir

$$ab = \frac{l_{\mu}^2 - l_{\nu}^2}{4L^2} \left\{ [(p_{\mu} - p'_{\mu})^2 - (p_{\nu} - p'_{\nu})^2] + (p'_{\mu} - p'_{\nu})(p_{\mu} + p_{\nu} - p'_{\mu} - p'_{\nu}) \right\}.$$

Hieraus sieht man, dass

$$M(ab) = \frac{l_{\mu}^2 - l_{\nu}^2}{4L^2} (\sigma_{\mu}^2 - \sigma_{\nu}^2).$$

Bemerken wir noch die Identität

$$\left(\frac{l_{\mu} - l_{\nu}}{2L}\right)^2 (\sigma_{\mu}^2 + \sigma_{\nu}^2) + \frac{l_{\mu}^2 - l_{\nu}^2}{4L^2} (\sigma_{\mu}^2 - \sigma_{\nu}^2) = \frac{l_{\mu} - l_{\nu}}{2L^2} (l_{\mu} \sigma_{\mu}^2 - l_{\nu} \sigma_{\nu}^2),$$

so ergibt sich aus (16) und (17)

$$(18) \quad M \left\{ \frac{l_\mu^2}{L^2} (p_\mu - p'_\mu)^2 + \frac{l_\nu^2}{L^2} (p_\nu - p'_\nu)^2 \right\} = \\ M \left\{ \left(\frac{l_\mu + l_\nu}{2L} \right)^2 (p_\mu - p_\nu)^2 + \frac{l_\mu - l_\nu}{2L^2} (l_\mu \sigma_\mu^2 - l_\nu \sigma_\nu^2) - \left(\frac{l_\mu + l_\nu}{2L} \right)^2 (p'_\mu - p'_\nu)^2 \right\}.$$

Diese Formel wollen wir dazu benutzen, den Ausdruck unter der Quadratwurzel in (15) umzuformen. Es ist offenbar

$$(19) \quad \sum_{\mu=1}^n \frac{l_\mu^2}{L^2} (p_\mu - p'_\mu)^2 = \frac{1}{2} \sum_{\mu=1}^{n-1} \left\{ \frac{l_\mu^2}{L^2} (p_\mu - p'_\mu)^2 + \frac{l_{\mu+1}^2}{L^2} (p_{\mu+1} - p'_{\mu+1})^2 \right\} + \\ \frac{1}{2} \frac{l_1^2}{L^2} (p_1 - p'_1)^2 + \frac{1}{2} \frac{l_n^2}{L^2} (p_n - p'_n)^2,$$

und hieraus ergibt sich mit Benutzung von (18)

$$(20) \quad M \left\{ \sum_{\mu=1}^n \frac{l_\mu^2}{L^2} (p_\mu - p'_\mu)^2 \right\} = M \left\{ \frac{1}{2} \sum_{\mu=1}^{n-1} \left(\frac{l_\mu + l_{\mu+1}}{2L} \right)^2 (p_\mu - p_{\mu+1})^2 \right\} + R,$$

wo

$$R = R_1 + R_2 + R_3$$

und

$$R_1 = \frac{1}{2} \frac{l_1^2}{L^2} \sigma_1^2 + \frac{1}{2} \frac{l_n^2}{L^2} \sigma_n^2,$$

$$R_2 = \frac{1}{2} \sum_{\mu=1}^{n-1} \frac{l_\mu - l_{\mu+1}}{2L^2} (l_\mu \sigma_\mu^2 - l_{\mu+1} \sigma_{\mu+1}^2),$$

$$R_3 = -\frac{1}{2} \sum_{\mu=1}^{n-1} \left(\frac{l_\mu + l_{\mu+1}}{2L} \right)^2 (p'_\mu - p'_{\mu+1})^2.$$

Wir führen die Voraussetzung ein, dass sämtliche Grössen R_1 , $|R_2|$ und $|R_3|$, und also auch $|R|$, so klein sind, dass sie im Vergleich mit dem ersten Gliede rechts von (20) vernachlässigt werden können. Dann ergibt sich, indem wir den Ausdruck in der grossen Parenthese rechts in (20) unter die Quadratwurzel in (15) einführen,

$$(21) \quad \varepsilon(P) = \sqrt{\frac{1}{2} \sum_{\mu=1}^{n-1} \left(\frac{l_\mu + l_{\mu+1}}{2L} \right)^2 (p_\mu - p_{\mu+1})^2}.$$

Dies ist die im vorigen Par. in Aussicht gestellte Formel, worin die p'_μ nicht mehr auftreten.

Wir wollen aber die neu eingeführten Voraussetzungen näher in Betracht ziehen.

7. Wenn weder die Längen l_μ noch die Streuungen σ_μ sehr stark variieren, wird R_1 ungefähr den n -ten Teil der linken Seite von (20) ausmachen. Ein solcher Fehler kann, wenn n nicht sehr klein ist, ohne Bedenken vernachlässigt werden.

Hinsichtlich der Grösse R_2 ist zunächst zu bemerken, dass sie, sowohl wenn die Längen l_μ alle einander gleich sind, als im Falle, wo $l_1 \sigma_1^2 = l_2 \sigma_2^2 = \dots = l_n \sigma_n^2$, gleich Null ist. Damit R_2 einen absehbaren Wert erhalte, müssen also die Längen l_μ sowie die Produkte $l_\mu \sigma_\mu^2$ bedeutend variieren, und dies muss noch so geschehen, dass die verschiedenen Glieder von R_2 einander nicht aufheben. Ein solcher Sachverhalt würde eine ganz besondere Struktur des Gebietes voraussetzen, und es scheint deshalb berechtigt, das Glied R_2 zu vernachlässigen.

Auch die Grösse $|R_3|$ wird im allgemeinen klein sein, denn die Variationen der Zahlen p'_μ sollen ja den systematischen Veränderungen der Prozente p_μ entsprechen, und diese haben meistens einen allmählichen stetigen Charakter. Die Vernachlässigung des Gliedes R in (20) ist somit wohl begründet.

Derjenige Teil von R , von welchem am öftesten zu befürchten sein dürfte, dass er einen nicht zu vernachlässigenden Wert erhält, ist ohne Zweifel R_3 . Es sei deshalb noch darauf aufmerksam gemacht, dass dieses Glied immer negativ ist und dass seine Vernachlässigung aus diesem Grunde nur bewirkt, dass der Mittelfehler zu gross geschätzt wird. Die Vergrößerung dürfte aber im allgemeinen nicht sehr bedeutend sein, und sie kann deshalb als ein nützlicher Sicherheitszuschlag betrachtet werden.

8. Die Formel (21) ist nicht die einzige, die auf Grund der Entwicklungen des Par. 6 zur Berechnung des Mittelfehlers in Frage kommen kann. Es ist auch, falls n eine gerade Zahl ist,

$$\sum_{\mu=1}^n \frac{l_\mu^2}{L} (p_\mu - p'_\mu)^2 = \sum_{\nu=1}^{\frac{n}{2}} \left\{ \frac{l_{2\nu-1}^2}{L^2} (p_{2\nu-1} - p'_{2\nu-1})^2 + \frac{l_{2\nu}^2}{L^2} (p_{2\nu} - p'_{2\nu})^2 \right\},$$

und wenn man diese Gleichung anstatt (19) nimmt, wird man zu der Formel

$$(22) \quad \varepsilon(P) = \sqrt{\sum_{\nu=1}^n \left(\frac{l_{2\nu-1} + l_{2\nu}}{2L} \right)^2 (p_{2\nu-1} - p_{2\nu})^2}$$

geführt. Ist n ungerade, kann man noch dieselbe Formel anwenden, wenn man eine von den Zahlen p_μ weglässt.

Wenn eine grosse Anzahl von Linien, oder also von Zahlen p_μ , vorliegt, empfiehlt sich sogar die Anwendung der Formel (22) vor der Formel (21), weil die Rechnung mit jener weniger mühsam ist. Ist aber die Anzahl der Zahlen p_μ nicht gross, so muss die Formel (21) zur Anwendung kommen, weil der Mittelfehler des Ausdrucks unter der Quadratwurzel hier kleiner ist. Um eine Vorstellung von der relativen Zuverlässigkeit der beiden Formeln zu gewinnen, wollen wir die Mittelfehler der Ausdrücke unter den Quadratwurzeln berechnen. Hierbei machen wir die folgenden Voraussetzungen:

Die Grössen p_μ folgen alle dem Gauss'schen Gesetz.

Die Mittelwerte dieser Grössen sind alle gleich p , und ihre Streuungen haben sämtlich denselben Wert σ .

Die Längen l_μ sind alle einander gleich.

Wir beginnen mit der Formel (21) und haben also den Mittelfehler des Ausdrucks

$$(23) \quad \sum_{\mu=1}^{n-1} (p_\mu - p_{\mu+1})^2$$

zu berechnen. Indem wir zur Abkürzung $p_\mu - p_{\mu+1} = x_\mu$ setzen, erhalten wir durch eine leichte Rechnung

$$M(x_\mu^2) = 2\sigma^2, \quad M(x_\mu^4) = 12\sigma^4$$

und

$$M(x_\mu^2 x_\nu^2) = 4\sigma^4, \quad \text{falls } |\mu - \nu| > 1$$

$$M(x_\mu^2 x_\nu^2) = 6\sigma^4 \quad \text{»} \quad |\mu - \nu| = 1.$$

Da

$$M\left\{ \sum_{\mu=1}^{n-1} x_\mu^2 \right\} = 2(n-1)\sigma^2,$$

ist das Quadrat des Mittelfehlers von (23) gleich

$$M\left\{ \left[\sum_{\mu=1}^{n-1} x_\mu^2 - 2(n-1)\sigma^2 \right]^2 \right\},$$

was im Hinblick auf die oben angegebenen Resultate den Wert

$$(12n - 16)\sigma^4$$

gibt. Also wird

$$(24) \quad \frac{\varepsilon \left\{ \sum_{\mu=1}^{n-1} (p_\mu - p_{\mu+1})^2 \right\}}{M \left\{ \sum_{\mu=1}^{n-1} (p_\mu - p_{\mu+1})^2 \right\}} = \sqrt{\frac{3n-4}{(n-1)^2}}.$$

Durch analoge Rechnungen findet man

$$(25) \quad \frac{\varepsilon \left\{ \sum_{\nu=1}^{\frac{n}{2}} (p_{2\nu-1} - p_{2\nu})^2 \right\}}{M \left\{ \sum_{\nu=1}^{\frac{n}{2}} (p_{2\nu-1} - p_{2\nu})^2 \right\}} = \sqrt{\frac{4}{n}}.$$

Für alle Werte $n > 4$ ist (24) kleiner als (25), und deshalb ist die Formel (21) im allgemeinen vorteilhafter als (22). Wenn n nicht mehr als 40 bis 50 beträgt, ist leider auch noch der Wert (24) sehr gross. Für $n = 40$ ergibt sich der Wert 0,27, was noch eine viel zu grosse Unsicherheit im Resultate zeigt.

Wenn $n = 200$, gibt (25) den Wert 0,14, was schon als eine hinreichende Genauigkeit bezeichnet werden kann. Dann empfiehlt sich die Anwendung der Formel (22).

9. Wenn es sich um grössere Gebiete handelt, kann es vorkommen, dass eine vollständige, auf Messungen gegründete exakte Taxierung der Linien nicht möglich ist. Man kann gezwungen sein, sich damit zu begnügen, äquidistante Probestücke auf den Linien genau zu taxieren und die dazwischenliegende Teile mittels Okulartaxierung zu schätzen oder vollständig ausser Acht zu lassen. Dann sind die zur Verfügung stehenden Werte p_μ nur Näherungswerte der zu den Linien gehörigen Prozente. Falls diese Zahlen als von systematischen Fehlern frei angesehen werden können, gelten nichtsdestoweniger alle im vorigen betreffs der als exakt vorausgesetzten Prozente gemachten Überlegungen. Also, die eine oder die andere der Formeln (21) und (22) muss, je nach der Anzahl der Linien, weiterhin zur Anwendung kommen.

10. Wenn die Anzahl der Linien klein ist, womit ich eine Anzahl kleiner als 100 verstehe, kann die aus dieser kleinen Anzahl herrührende

Unsicherheit in unseren Formeln dadurch gehoben werden, dass das Gebiet so in Teilgebiete zerlegt wird, dass die neu eingeführten Grenzen die Taxierungslinien schneiden. Ich will dies an einem einfachen Beispiel zeigen.

Es sei angenommen, dass die Längen l_μ alle gleich sind, die p_μ alle dieselben Mittelwerte und Streuungen haben und dem Gauss'schen Gesetz folgen. Indem wir das Gebiet G in zwei Teilgebiete G_1 und G_2 durch eine Grenzlinie, die die der Taxierung zu Grunde liegenden Linien halbiert, zerlegen, nehmen wir weiter an, dass die Prozente $p_\mu^{(1)}$ und $p_\mu^{(2)}$ der Linien der Teilgebiete ebenfalls sämtlich gleiche Mittelwerte und Streuungen haben und dem Gauss'schen Gesetz folgen. Das für G erhaltene Prozent sei P , und die für die Teilgebiete erhaltenen Prozente mögen mit P_1 und P_2 bezeichnet werden, wobei offenbar

$$(26) \quad P = \frac{1}{2} (P_1 + P_2).$$

Erfolgt nun die Berechnung des Mittelfehlers von P direkt für das ganze Gebiet G mittels der Formel (21), so ist nach (24)

$$(27) \quad \frac{\varepsilon \langle \varepsilon^2(P) \rangle}{M \langle \varepsilon^2(P) \rangle} = \sqrt{\frac{3n-4}{(n-1)^2}}.$$

Andererseits denken wir uns, dass wir zuerst die Mittelfehler von P_1 und P_2 mittels (21) und sodann den Mittelfehler von P mittels der sich aus (26) ergebenden Gleichung

$$\varepsilon^2(P) = \frac{1}{4} \{ \varepsilon^2(P_1) + \varepsilon^2(P_2) \}$$

berechnen. Da wegen unserer Voraussetzungen

$$M \langle \varepsilon^2(P_1) \rangle = M \langle \varepsilon^2(P_2) \rangle$$

und

$$\varepsilon \langle \varepsilon^2(P_1) \rangle = \varepsilon \langle \varepsilon^2(P_2) \rangle,$$

so ist dann

$$M \langle \varepsilon^2(P) \rangle = \frac{1}{2} M \langle \varepsilon^2(P_1) \rangle$$

und

$$\varepsilon \langle \varepsilon^2(P) \rangle = \frac{1}{4} \sqrt{\varepsilon^2 \langle \varepsilon^2(P_1) \rangle + \varepsilon^2 \langle \varepsilon^2(P_2) \rangle} = \frac{\varepsilon \langle \varepsilon^2(P_1) \rangle}{2\sqrt{2}}$$

und mithin

$$\frac{\varepsilon \langle \varepsilon^2(P) \rangle}{M \langle \varepsilon^2(P) \rangle} = \frac{1}{\sqrt{2}} \frac{\varepsilon \langle \varepsilon^2(P_1) \rangle}{M \langle \varepsilon^2(P_1) \rangle} = \frac{1}{\sqrt{2}} \sqrt{\frac{3n-4}{(n-1)^2}}.$$

Wenn wir dieses Resultat mit (27) vergleichen, sehen wir, dass die Zerlegung unseres Gebietes in zwei Teilgebiete zu einer wesentlich sicheren Bestimmung des Mittelfehlers führt.

Aus diesem Resultate folgt, dass, obgleich man gezwungen sein kann, Mittelfehler auf Grund einer Anzahl von Linienprozenten zu berechnen, die zu klein ist, damit das Resultat eine selbständige Bedeutung habe, das Resultat doch nicht wertlos zu sein braucht. In Verbindung mit anderen Resultaten derselben Art kann dasselbe zu einem hinreichend sicheren Mittelfehler eines grösseren Gebietes führen. Wir wollen deshalb noch besonders nachsehen, wie die Rechnung gemacht werden muss, falls die Anzahl der Linien sehr klein ist, worunter ich, um zu präzisieren, eine Anzahl kleiner als 20 verstehen will.

Vor allem müssen wir dafür Sorge tragen, dass die erhaltenen Mittelfehler nicht systematisch zu klein werden. Die Vernachlässigung des Gliedes R_1 in dem Ausdruck (20) kann eine solche Wirkung haben. Im Par. 7 haben wir schon erwähnt, dass die Grösse dieses Gliedes im allgemeinen ungefähr ein n :ter Teil des Mittelwertes des Ausdrucks unter der Quadratwurzel in (15) beträgt. Multiplizieren wir den Ausdruck unter der Quadratwurzel in (21) mit $\frac{n}{n-1}$, so wird daher die Wirkung dieser Fehlerquelle aufgehoben. Falls n sehr klein ist, muss man somit die Formel (21) durch die Formel

$$\varepsilon(P) = \sqrt{\frac{n}{2(n-1)} \sum_{\mu=1}^{n-1} \left(\frac{l_\mu + l_{\mu+1}}{2L} \right)^2 (p_\mu - p_{\mu+1})^2}$$

ersetzen.

11. Es wäre von grossem Interesse gewesen, die in der Taxierung von Värmland erhaltenen Mittelfehler mit denjenigen zu vergleichen, welche die Formel (21) liefert. Da in dem Bericht über diese Taxierung wie in den Berichten über die norwegischen Taxierungen nur die die Liniengruppen betreffenden Prozentzahlen mitgeteilt sind, ist mir eine solche Vergleichung nicht möglich gewesen. Dagegen hat Herr Y. LIVESALO gütigst eine Umrechnung einiger von ihm in der oben zitierten Arbeit erhaltenen Mittelfehler mittels der Formel (21) besorgt. Ich teile hier sowohl die mit seinem früheren Verfahren als die mit der neuen Formel erhaltenen Resultate mit. Die den Waldboden in Prozenten der Landfläche betreffenden Resultate sind folgende:

Gebiet	Mittelfehler nach dem alten Verfahren	Mittelfehler nach Formel (21)	Gewicht des Gebietes
Sahalahti und Pento	0,85	0,92	0,629
Ilvesvuori	1,66	2,16	0,095
Kotala	1,05	1,87	0,086
Vehkajärvi und Pajulahti .	1,74	1,75	0,190
Alle Gebiete zusammen ..	0,66	0,72	1

Als Mittelfehler des durchschnittlichen Kubikinhalts pro ha Waldboden wurden die folgenden Zahlen erhalten:

Gebiet	Mittelfehler nach dem alten Verfahren	Mittelfehler nach Formel (21)	Gewicht des Gebietes
Sahalahti und Pento	1,40	1,26	0,609
Ilvesvuori	2,93	3,50	0,095
Kotala	1,19	2,63	0,085
Vehkajärvi und Pajulahti .	2,36	1,66	0,211
Alle Gebiete zusammen ..	1,03	0,93	1

Die Übereinstimmung in den definitiven, das ganze Gebiet berührenden Zahlen kann als ziemlich gut bezeichnet werden. Da ich erwartete, dass die von der Formel (21) gelieferten Zahlen im allgemeinen etwas grösser ausfallen würden als die früher erhaltenen und es sich doch in der zweiten Tabelle betreffs der zwei grösseren Gebiete umgekehrt verhält, habe ich die originalen Diagramme, in denen die Ausgleichungen der Punktfolgen vorgenommen wurden, untersucht. Es zeigte sich hierbei, dass man in dem wichtigsten, das Gebiet Sahalahti und Pento betreffenden Diagramm eine plausible Mittelwertkurve zeichnen konnte, die zu einem wesentlich kleineren Mittelfehler geführt hätte. In dem demnächst wichtigsten Diagramm des Gebietes Vehkajärvi und Pajulahti trat der systematische Verlauf der Punkte nicht ganz deutlich

hervor. Um keinen zu kleinen Mittelfehler zu erhalten, war deshalb eine horizontale Gerade als ausgleichende Kurve genommen. Diese Tatsachen scheinen mir den oben erwähnten Umstand gut zu erklären. Ich glaube deshalb, dass die Formel (21) ein gutes Instrument ist, eine solche Bestimmung des Mittelfehlers mechanisch zu erhalten, die durch die von Y. ILVESSALO früher angewandte Methode erzielt ist.

12. Es sei schliesslich bemerkt, dass die Formel (21) noch anwendbar ist, wenn es sich nicht um eine eigentliche Linientaxierung handelt, sondern um eine Taxierung auf Grund von kleineren Probeflächen, die über das ganze Gebiet gleichmässig verteilt sind. Wir wollen annehmen, dass die Probeflächen z. B. aus gleich grossen quadratischen Flächenstücken bestehen, deren Mittelpunkte sich in den Kreuzungspunkten eines rektangulären Netzes befinden. Dann können die verschiedenen Probeflächen mittels zweier Indices fixiert werden; wir bezeichnen mit $p_{\mu, \nu}$ die zu denselben gehörigen Prozentzahlen. Der Mittelfehler des Mittelwertes P aus diesen Prozentsen kann vermittels der Formel (21) nachgebildeten Formel

$$(28) \quad \varepsilon(P) = \sqrt{\frac{1}{4n^2} \sum (p_{\mu, \nu} - p_{\mu', \nu'})^2}$$

berechnet werden, wo n die Anzahl der Probeflächen bedeutet und die Summe über alle möglichen Wertsysteme μ, ν, μ', ν' zu erstrecken ist, welche den Bedingungen

$$\mu' < \mu, \quad \nu' < \nu, \quad \mu + \nu - \mu' - \nu' = 1$$

genügen.

Eine einzelne Zahl $p_{\mu, \nu}$ tritt im allgemeinen in vier Gliedern der Summe auf. Die zu den an der Grenze des Gebietes gelegenen Probeflächen gehörigen Prozente sind aber meistens nur in drei oder zwei Gliedern enthalten. Es bezeichne α die Anzahl der Prozente erster Art (die also in drei Gliedern auftreten) und β die Anzahl der Flächen zweiter Art. Falls $\alpha + 2\beta$ bedeutend ist oder, um zu präzisieren, falls $\alpha + 2\beta$ mehr als ein Fünftel von n beträgt, empfiehlt sich die Korrektur der Formel (28), die der im Par. 10 erwähnten entspricht. Man findet für diesen Fall

$$\varepsilon(P) = \sqrt{\frac{1}{n(4n - \alpha - 2\beta)} \sum (p_{\mu, \nu} - p_{\mu', \nu'})^2}$$

Zum Schluss sei bemerkt, dass, wenn die Anzahl der Probeflächen gross ist, man, um die Rechnungen weniger mühsam zu machen, eine solche Abänderung der Formel (28) vornehmen kann, die der in Par. 8 erwähnten Reduktion der Formel (21) entspricht.