

Paper-ID: VGI\_195913



## Die Ausgleichung vermittelnder Beobachtungen im Rahmen der mathematischen Statistik

W. Eberl <sup>1</sup>

<sup>1</sup> *Wien*

Österreichische Zeitschrift für Vermessungswesen **47** (3), S. 73–84

1959

Bib<sub>T</sub>E<sub>X</sub>:

```
@ARTICLE{Eberl_VGI_195913,  
  Title = {Die Ausgleichung vermittelnder Beobachtungen im Rahmen der  
           mathematischen Statistik},  
  Author = {Eberl, W.},  
  Journal = {{{\0}sterreichische Zeitschrift f{{\"u}r Vermessungswesen}},  
  Pages = {73--84},  
  Number = {3},  
  Year = {1959},  
  Volume = {47}  
}
```



den Zahlenwert der Ausdrücke

$$\frac{(n-i)(i-1)}{\sqrt{3(n-1)}} \quad \text{und} \quad \frac{(n-i)(i-1)}{\sqrt{12(n-1)}}$$

gegenübergestellt.

Für  $n = 10$  erhält man — die Näherungswerte sind unter den strengen Werten angeschrieben — für die Werte von  $i$  zwischen 1 und 10 dafür die Zahlenwerte:

$i =$	1	2	3	4	5	6	7	8	9	10
Formel (12)	0,00	1,59	2,74	3,51	3,90	3,90	3,51	2,74	1,59	0,00
Formel (12a)	0,00	1,54	2,69	3,46	3,85	3,85	3,46	2,69	1,54	0,00
Formel (19)	0,00	0,96	1,62	2,06	2,28	2,28	2,06	1,62	0,96	0,00
Formel (19a)	0,00	0,77	1,35	1,73	1,93	1,93	1,73	1,35	0,77	0,00

Die Näherung (12a) liefert, wie man aus den nur kleinen Abweichungen gegen die Sollwerte erkennt, sehr gute Ergebnisse, während die Näherung (19a) bei  $n = 10$  im Durchschnitt um etwa 15% zu kleine Werte ergibt.

## Die Ausgleichung vermittelnder Beobachtungen im Rahmen der mathematischen Statistik

Von *W. Eberl*

**§ 1. Einleitung.** Trotz der stürmischen Entwicklung der Stochastik<sup>1)</sup> während der letzten Jahrzehnte hat sich die lehrbuchmäßige Darstellung der Ausgleichsrechnung seit den Tagen von C. F. *Gauß* (1777—1855) und F. R. *Helmert* (1843 bis 1917) kaum geändert. Das ist im Hinblick auf beide Disziplinen bedauerlich. Denn einerseits tragen die Methoden der mathematischen Statistik viel weiter als die der traditionellen Ausgleichsrechnung, und andererseits müßte die Beachtung der Tatsache, daß die Ausgleichsrechnung nur ein kleines wenn auch wichtiges Teilgebiet der Regressionstheorie darstellt, zu einer realistischeren Beurteilung der Rolle, die die Stochastik für den Techniker spielt, beitragen.

Der Ausgleichung vermittelnder Beobachtungen kommt eine besondere Bedeutung zu, da einerseits die direkten Beobachtungen als Sonderfälle von vermittelnden angesehen werden können und sich andererseits die Ausgleichung bedingter Beobachtungen meist sehr einfach auf die Ausgleichung vermittelnder Beobachtungen zurückführen läßt.

Die Paragraphen 2 bis 4 enthalten ohne Beweis einige für das Folgende grundlegende Definitionen und Sätze der Stochastik. Ziffer 5 bringt dann einige Sätze der Regressionstheorie samt den zugehörigen meist bekannten Beweisen. Die ausgiebige Verwendung des Summationsübereinkommens auf diesem Gebiet dürfte neu sein und bietet gewisse Vorteile.

<sup>1)</sup> Die Stochastik ist die Lehre vom Zufall und umfaßt Wahrscheinlichkeitstheorie und mathematische Statistik.

§ 2. Einige Grundbegriffe der Wahrscheinlichkeitstheorie.  $X_1, \dots, X_r$  seien  $r$  reelle stetige Merkmale<sup>2)</sup> mit gemeinsamer Verteilung oder  $X = (X_1, \dots, X_r)$  ein  $r$ -dimensionales stetiges Merkmal. Der Wertevorrat  $\mathfrak{W}$  von  $X_1, \dots, X_r$  bzw.  $X$  sei der  $r$ -dimensionale Euklidische Raum  $\mathfrak{E}_r$  oder ein  $r$ -dimensionales Intervall desselben. Die gemeinsame Verteilung von  $X_1, \dots, X_r$  oder die Verteilung von  $X$  ist dann durch eine Dichte  $f(x) = f(x_1, \dots, x_r)$ <sup>3)</sup> bestimmt, die auf höchstens endlich vielen Hyperflächen des  $\mathfrak{E}_r$  unstetig ist. Indem wir außerhalb von  $\mathfrak{W}$   $f(x) \equiv 0$  setzen, können wir  $f(x)$  als eine im ganzen  $\mathfrak{E}_r$  definierte Funktion annehmen.

Die Wahrscheinlichkeit, daß  $X$  einem Bereich  $\mathfrak{B}$  angehört, von dem nur vorausgesetzt wird, daß auf ihm das folgende Integral definiert ist, ist das  $r$ -fache Integral  $W\{X \in \mathfrak{B}\} = \int_{\mathfrak{B}} f(x) dx$  mit  $dx = dx_1 \dots dx_r$ . Natürlich ist  $\int_{\mathfrak{E}_r} f(x) dx = 1$ . Die Erwartung einer Merkmalfunktion  $\varphi(X) = \varphi(X_1, \dots, X_r)$  ist

$$E\varphi(X) = \int_{\mathfrak{E}_r} \varphi(x) f(x) dx, \quad \dots \quad (1)$$

sofern das Integral absolut konvergiert.  $E$  ist ein linearer Operator, so daß für  $n$  Merkmalfunktionen  $\varphi_i(X)$  und  $n$  Konstante  $c_i$  ( $i = 1, \dots, n$ )

$$E \sum_{i=1}^n c_i \varphi_i(X) = \sum_{i=1}^n c_i E\varphi_i(X) \quad \dots \quad (2)$$

gilt.  $(i_1, \dots, i_k)$  sei eine Permutation von  $(1, \dots, r)$  und  $k < r$ . Betrachtet man dann die Verteilung von  $(X_{i_1}, \dots, X_{i_k})$  ganz unabhängig davon, welche Werte  $(X_{i_{k+1}}, \dots, X_{i_r})$  annimmt, so heißt diese Verteilung die Randverteilung von  $(X_{i_1}, \dots, X_{i_k})$ . Die Dichte dieser Randverteilung ist

$$f_{i_1 \dots i_k}(x_{i_1}, \dots, x_{i_k}) = \int_{\mathfrak{E}_{r-k}} f(x) dx_{i_{k+1}} \dots dx_{i_r}. \quad \dots \quad (3)$$

$X_{i_1}, \dots, X_{i_k}$  heißen unabhängig voneinander, wenn  $f_{i_1 \dots i_k}(x_{i_1}, \dots, x_{i_k}) = \prod_{j=1}^k f_{i_j}(x_{i_j})$

ist. Ist  $\varphi(X_{i_1}, \dots, X_{i_k}) = \prod_{j=1}^k \varphi_{i_j}(X_{i_j})$  eine Funktion der unabhängigen Merkmale  $X_{i_1}, \dots, X_{i_k}$ , so gilt

$$E \prod_{j=1}^k \varphi_{i_j}(x_{i_j}) = \prod_{i=1}^k E\varphi_{i_j}(x_{i_j}). \quad \dots \quad (4)$$

Für  $p = 1, \dots, r$  heißt  $EX_p = \int_{-\infty}^{+\infty} x_p f_p(x_p) dx_p = \xi_p$  das Mittel und  $VX_p = E(X_p - \xi_p)^2 = \int_{-\infty}^{+\infty} (x_p - \xi_p)^2 f_p(x_p) dx_p = \sigma_p^2$  die Varianz von  $X_p$ .

Es gilt der Verschiebungssatz  $E(X_p - \xi_p)^2 = EX_p^2 - (EX_p)^2$  oder

$$\sigma_p^2 = EX_p^2 - \xi_p^2. \quad \dots \quad (5)$$

Die positive Wurzel  $\sigma_p$  aus der Varianz heißt Streuung<sup>4)</sup> von  $X_p$ .

<sup>2)</sup> Statt Merkmal ist auch Zufallsvariable oder zufällige Variable gebräuchlich.

<sup>3)</sup> Variable werden mit großen oder kleinen Buchstaben bezeichnet, je nachdem für ihren Wertevorrat ein Wahrscheinlichkeitsmaß von Belang ist oder nicht. (Zufällige) Merkmale werden daher durch Großbuchstaben, Variable im Sinne der Analysis durch Kleinbuchstaben ausgedrückt.

<sup>4)</sup> In der Ausgleichsrechnung: Mittlere Abweichung oder mittlerer Fehler.

Sind  $p \neq q$  zwei ganze Zahlen zwischen 1 und  $r$ , so heißt

$$C(X_p, X_q) = E[(X_p - \xi_p)(X_q - \xi_q)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_p - \xi_p)(x_q - \xi_q) f_{pq}(x_p, x_q) dx_p dx_q = \sigma_{pq}$$

die Kovarianz von  $X_p$  und  $X_q$ . Auch für die Kovarianz gilt ein *Verschiebungssatz*:

$$E[(X_p - \xi_p)(X_q - \xi_q)] = E(X_p X_q) - \xi_p \xi_q$$

oder 
$$\sigma_{pq} = E(X_p X_q) - \xi_p \xi_q. \quad \dots (6)$$

Die  $r$ -reihige Matrix  $(\sigma_{pq})$  mit  $\sigma_{pp} = \sigma_p^2$  heißt *Kovarianzmatrix* von  $X$ .

$X_p$  und  $X_q$  heißen *unkorreliert*, wenn  $\sigma_{pq} = 0$  ist. Wegen (4) und (2) sind unabhängige Merkmale immer auch unkorreliert, dagegen müssen unkorrelierte Merkmale nicht unabhängig sein.

Ist  $X$  ein eindimensionales Merkmal mit dem Mittel  $\xi$  und der Varianz  $\sigma^2$ , sind ferner  $a$  und  $b$  zwei beliebige Konstante, so sind Mittel und Varianz des Merkmals  $Y = (X - a)/b$

$$EY = \frac{\xi - a}{b} \quad \text{und} \quad VY = \frac{\sigma^2}{b^2}. \quad \dots (7a, 7b)$$

Sind  $X_1, \dots, X_r$  unkorrelierte Merkmale mit den Varianzen  $\sigma_1^2, \dots, \sigma_r^2$  und sind  $c_1, \dots, c_r$  beliebige Konstante, so ist

$$V \sum_{p=1}^r c_p X_p = \sum_{p=1}^r c_p^2 \sigma_p^2. \quad \dots (8)$$

**§ 3. Zwei Verteilungen.** Die folgenden beiden Verteilungen können als Beispiele für die allgemeineren Definitionen von § 2 dienen.

A. *Die  $r$ -dimensionale Gaußverteilung.* Das Merkmal  $X = (X_1, \dots, X_r)$  heißt ( $r$ -dimensional) *nach Gauß verteilt*, wenn sein Wertevorrat der  $\mathbf{E}_r$  und seine Dichte

$$f(x_1, \dots, x_r) = \frac{\sqrt{\text{Det}(\sigma^{pq})}}{(2\pi)^{\frac{r}{2}}} e^{-\frac{1}{2} \sum_{p=1}^r \sum_{q=1}^r \sigma^{pq} (x_p - \xi_p)(x_q - \xi_q)} \quad \dots (9)$$

ist. Es ist  $EX_p = \xi_p$ ,  $VX_p = \sigma_p^2 = \sigma_{pp}$  und  $C(X_p, X_q) = \sigma_{pq}$ . Die Matrix  $(\sigma^{pq})$  ist symmetrisch, positiv definit und stellt die inverse der Kovarianzmatrix  $(\sigma_{pq})$  dar:

$$(\sigma^{pq}) = (\sigma_{pq})^{-1}.$$

Die  $r$ -dimensionale Gaußverteilung ist durch ihre Parameter  $\xi_p$  und  $\sigma_{pq}$ ;  $p, q = 1, \dots, r$ , vollständig bestimmt.

Man kann zeigen, daß die Randverteilung von  $X_p$  die eindimensionale Gaußverteilung mit dem Mittel  $\xi_p$  und der Varianz  $\sigma_p^2$  ist:

$$f_p(x_p) = \frac{1}{\sqrt{2\pi} \sigma_p} e^{-\frac{(x_p - \xi_p)^2}{2\sigma_p^2}} \quad \dots (10)$$

Statt (10) schreibt man kürzer:  $X_p$  ist nach  $G(\xi_p, \sigma_p^2)$  verteilt. Daraus folgt sofort

*Satz 1:  $r$  gemeinsam nach Gauß verteilte Merkmale  $X_1, \dots, X_r$  sind dann und nur dann voneinander unabhängig, wenn ihre Kovarianzen  $\sigma_{pq}$  ( $p \neq q$ ) verschwinden.*

Denn (9) zerfällt dann und nur dann in  $r$  Faktoren (10). Unkorrelierte Gaußmerkmale sind also auch unabhängig und umgekehrt.

Für Linearkombinationen von Gaußmerkmalen gilt

*Satz 2:* Sind  $X_1, \dots, X_r$  gemeinsam nach Gauß verteilt mit den Mitteln  $\xi_p$  und der Kovarianzmatrix  $(\sigma_{pq})$ , sind ferner die  $c_p$  beliebige Konstante, so ist auch  $X =$

$$= \sum_{p=1}^r c_p X_p \text{ nach Gauß verteilt mit dem Mittel } EX = \sum_{p=1}^r c_p \xi_p \text{ und der Varianz } VX =$$

$$= \sum_{p=1}^r \sum_{q=1}^r c_p c_q \sigma_{pq}. \text{ Sind insbesondere die } X_p \text{ unabhängig voneinander, so ist}$$

$$VX = \sum_{p=1}^r c_p^2 \sigma_p^2 \text{ (vgl. (8)).}$$

B. Die *Chi-Quadrat* ( $\chi^2$ -)verteilung. Sind  $X_1, \dots, X_n$  unabhängig voneinander nach  $G(0, 1)$  verteilt, so besitzt  $X = \sum_{i=1}^n X_i^2$  die  $\chi^2$ -Verteilung mit  $n$  Freiheitsgraden (F. R. Helmert, 1876). Der Wertevorrat von  $X$  ist die Halbgerade  $[0, +\infty)$ , die Dichte ist

$$f(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}.$$

Es ist

$$EX = n \text{ und } VX = 2n. \quad \dots (11a, 11b)$$

Mit dieser Definition  $\chi^2$ -verteilter Merkmale hängt eng zusammen

*Satz 3 (Cochran, 1933):*  $X_1, \dots, X_n$  seien unabhängig voneinander nach  $G(0, \sigma^2)$  verteilt. Ferner sei  $\sum_{i=1}^n X_i^2 = Q_1 + \dots + Q_k$ , wo  $Q_j$  eine quadratische Form in den  $X_i$  vom Rang<sup>5)</sup>  $\leq r_j$  ist,  $j = 1, \dots, k$ . Ist dann  $\sum_{j=1}^k r_j = n$ , so sind die  $Q_j/\sigma^2$  unabhängig voneinander nach  $\chi^2$  mit  $r_j$  Freiheitsgraden verteilt.

Einen Beweis findet man in [2].

**§ 4. Das Schätzen von Parametern.** In der Stochastik werden vor allem *Scharen* von Verteilungen betrachtet. Dementsprechend hänge die Dichte  $f(x)$  des Merkmals  $X$  von  $m$  Parametern  $\theta_1, \dots, \theta_m$  ab, die man zu einem  $m$ -dimensionalen Parameter  $\theta = (\theta_1, \dots, \theta_m)$  zusammenfassen kann. Statt  $f(x)$  ist daher genauer  $f(x; \theta)$  zu schreiben.

$X^1, \dots, X^n$  heißen *Beobachtungen* des Merkmals  $X$ , wenn alle  $X^i$  ( $i = 1, \dots, n$ ) unabhängig voneinander wie  $X$  verteilte Merkmale sind, und wenn jedem  $X^i$  ein bestimmter beobachteter Wert  $x^i$  entspricht.

<sup>5)</sup> Eine quadratische Form  $\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$  mit  $a_{ij} = a_{ji}$  hat den Rang  $r$ , wenn die Matrix  $(a_{ij})$  den Rang  $r$  hat.

Die mathematische Statistik hat nun, z. T. in Fortführung von Verfahren der Ausgleichsrechnung, eine Reihe von Methoden entwickelt, durch die sich aus  $n$  Beobachtungen eines Merkmals  $X$  vorteilhafte *Schätzungen*

$$T_a = \hat{\theta}_a^1(X^1, \dots, X^n), \quad a = 1, \dots, m, \quad \dots \quad (12)$$

der unbekanntem Parameter  $\theta_a$  gewinnen lassen. Ersetzt man auf der rechten Seite von (12) die Beobachtungen  $X^i$  durch die beobachteten Werte  $x^i$ , so ergeben sich spezielle *Schätzwerte*  $t_a = \hat{\theta}_a^1(x^1, \dots, x^n)$  für die  $\theta_a$ . Da die  $\hat{\theta}_a^1(X^1, \dots, X^n)$  Merkmalfunktionen sind, sind die  $T_a$  so wie die  $X^i$  Merkmale. Ihre Verteilung ist durch die Verteilung von  $X$  und durch die Funktionen  $\hat{\theta}_a^1$  bestimmt.

*Beispiel:* Sind  $X_1, \dots, X_n$  Beobachtungen eines nach  $G(\xi, \sigma^2)$  verteilten Merkmals  $X$ , so ist der *Durchschnitt*  $\bar{X} = \sum_{i=1}^n X_i/n$  der Beobachtungen eine Schätzung des Mittels  $\xi$  von  $X$ .  $\bar{X}$  ist auf Grund von Satz 2 nach  $G(\xi, \sigma^2/n)$  verteilt. Eine andere Schätzung von  $\xi$  ist  $\tilde{X} = (X_{\min} + X_{\max})/2$ , wo  $X_{\min}$  und  $X_{\max}$  die kleinste bzw. die größte der  $n$  Beobachtungen ist.

Man verwendet die Freiheit, die man in der Wahl der Schätzung (12) hat, um diese Schätzung mit möglichst vielen wünschenswerten Eigenschaften auszustatten. Für die Zwecke der Ausgleichung vermittelnder Beobachtungen sind folgende Vorzüge besonders wichtig:

Eine Schätzung heißt *linear*, wenn sie in den Beobachtungen linear ist, also  $T_a = \sum_{i=1}^n c_{ai} X^i$ , wobei die Koeffizienten  $c_{ai}$  von den Beobachtungen unabhängig sind. Im obigen Beispiel ist  $\bar{X}$  eine lineare Schätzung, da die Koeffizienten unabhängig von den Beobachtungen den Wert  $1/n$  haben. Dagegen ist  $\tilde{X}$  keine lineare Schätzung.  $\tilde{X}$  erscheint zwar zunächst als lineare Form in den  $X^i$ , wobei die Koeffizienten entweder 0 oder  $1/2$  sind, aber es hängt eben von den Beobachtungen selbst ab, ob eine Beobachtung mit 0 oder mit  $1/2$  zu multiplizieren ist. Wenn die Beobachtungen so klein ausfallen, daß ihre Quadrate vernachlässigt werden können, so kann man an Stelle von Schätzungen, die nach ihren Argumenten differenzierbar sind, lineare Schätzungen verwenden.

Von einer guten Schätzung verlangt man, daß die Mitte ihrer Verteilung in den zu schätzenden Parameter zu liegen kommt. Indem man diese Mitte z. B. durch die Erwartung von  $\hat{\theta}_a^1(X^1, \dots, X^n)$  oder durch das Mittel von  $T_a$  definiert, gelangt man zum Begriff der *erwartungstreuen* Schätzung: Für eine solche ist  $ET_a = \theta_a$ . Man wird weiter fordern, daß sich die Verteilung von  $T_a$  möglichst dicht um den zu schätzenden Parameter  $\theta_a$  zusammenballt. Als Maß dieser Zusammenballung kann man z. B. die Varianz verwenden. Im Hinblick auf die Ausgleichung vermittelnder Beobachtungen kann man sich auf lineare und erwartungstreue Schätzungen beschränken und nennt  $T_a$  eine *beste* erwartungstreue lineare Schätzung, wenn für alle derartigen Schätzungen  $T_a^*$

$$E(T_a - \theta_a)^2 \leq E(T_a^* - \theta_a)^2$$

gilt, wenn also  $T_a$  unter allen erwartungstreuen linearen Schätzungen  $T_a^*$  die kleinste Varianz besitzt.

Da auch der *Median*<sup>6)</sup> und der *Modus*<sup>7)</sup> von  $T_a$  zentral gelegene Punkte der Verteilung von  $T_a$  sind und i. a. nur dann mit dem Mittel  $ET_a$  zusammenfallen, wenn  $T_a$  symmetrisch verteilt ist, liegt in der Bestimmung der Mitte der Verteilung von  $T_a$  durch  $ET_a$  meistens eine gewisse Willkür. Die Messung der Zusammenballung einer Verteilung durch die Varianz ist nicht einmal bei Gauß- oder Chiquadratmerkmalen zwangsläufig, da auch jede monotone Funktion der Varianz, z. B. das *Genauigkeitsmaß*  $h^2 = 1/2 \sigma^2$  zu diesem Zweck verwendet werden kann. Die erste Festsetzung rechtfertigt sich jedoch durch ihre Zwangsläufigkeit im Falle symmetrischer Verteilungen und beide erweisen ihre Zweckmäßigkeit durch die Vorzüge der auf ihnen beruhenden Rechenverfahren, auch wenn man sehr allgemeine Verteilungen der  $T_a$  zuläßt.

Die wichtigste Methode zur Gewinnung von guten Schätzungen unbekannter Parameter ist das auf C. F. Gauß zurückgehende *Plausibilitätsprinzip*, das von R. A. Fisher im Jahre 1922 unter dem Namen *Maximum-Likelihood-Principle* weiter ausgebaut wurde.

Hat das Merkmal  $X = (X_1, \dots, X_r)$  die Dichte  $f(x; \theta)$ , wo  $\theta = (\theta_1, \dots, \theta_m)$  ein  $m$ -dimensionaler Parameter ist, sind ferner  $X^1, \dots, X^n$  Beobachtungen von  $X$ , so heißt die Funktion

$$P(x^1, \dots, x^n; \theta) = \prod_{i=1}^n f(x^i; \theta) \quad \dots (13)$$

bei festgehaltenen beobachteten Werten  $x^1, \dots, x^n$  und variablem  $\theta$  die *Plausibilitätsfunktion* des Parameters  $\theta$ . Der einem Parameterwert entsprechende Funktionswert heißt die *Plausibilität* des Parameterwertes. *Ordnet man nun jedem  $n$ -Tupel beobachteter Werte  $(x^1, \dots, x^n)$  den Parameterwert  $t = \hat{\theta}(x^1, \dots, x^n)$  größter Plausibilität*<sup>8)</sup> zu, so heißt  $T = \hat{\theta}(X^1, \dots, X^n)$  die *plausible Schätzung* von  $\theta$ . Natürlich setzt sich die  $m$ -dimensionale Schätzung  $\hat{\theta}$  aus  $m$  eindimensionalen Schätzungen  $\hat{\theta}_1, \dots, \hat{\theta}_m$  zusammen, und es ist  $T_a = \hat{\theta}_a(X^1, \dots, X^n)$  die plausible Schätzung von  $\theta_a$ . Plausible Schätzungen sind durch eine Anzahl vorteilhafter Eigenschaften ausgezeichnet.

6) Der Median von  $T_a$  ist ein (nicht immer eindeutig bestimmter) Wert  $\mu$ , für den  $W\{T_a \leq \mu\} = W\{T_a \geq \mu\} = \frac{1}{2}$  ist.

7) Der Modus von  $T_a$  ist ein (nicht immer eindeutig bestimmter) Wert  $\nu$ , in dem die Dichte von  $T_a$  ihren größten Wert annimmt.

8) Beschränkt man den Wertevorrat des Parameters  $\theta$  auf einen abgeschlossenen Teil des  $\mathcal{E}_m$ , den *Raum der zulässigen Parameter*, und hängt (13) stetig von den  $\theta_a$  ab, so ist die Existenz dieser Größtwerte auch formal gesichert. Der Statistiker kann sich aber ähnlich wie der Physiker solche rein mathematischen Existenzbetrachtungen ersparen, da das Vorhandensein bestimmter Verteilungen usw. in der Erfahrungswelt als gegeben angesehen wird, das heißt als Arbeitshypothese ein für allemal vorausgesetzt werden muß.

Bei diskret verteilten Merkmalen, deren Verteilung von einem unbekanntem Parameter abhängt, hat man die Plausibilitätsfunktion mit Hilfe der Wahrscheinlichkeiten anstatt der Dichten zu definieren. Die Bestimmung plausibler Schätzwerte der unbekanntem Parameter erfolgt dann ganz analog.

**§ 5. Einige Sätze aus der Theorie der linearen Regression.**  $Y_1, \dots, Y_n$  seien  $n$  Beobachtungen.  $EY_i = \sum_{a=1}^m x_{ai} \theta_a$ ,  $VY_i = \sigma^2$ , wo die  $x_{ai}$  für  $a = 1, \dots, m$  und  $i = 1, \dots, n$  bekannte Zahlen und die  $\theta_a$  sowie  $\sigma^2$  unbekanntem Parameter sind. Die Hyperebene  $\eta = \sum_{a=1}^m \theta_a x_a$  heißt *Regressionshyperebene* von  $Y$  in bezug auf die  $x_a$ . Man spricht auch von einer *m-fachen linearen Regression* von  $Y$  in bezug auf die  $x_a$ .

Die weitaus meisten Zusammenhänge, die der Erfahrungswissenschaftler zählend, messend und wägend beobachtet, stellen Überlagerungen von funktionellen Abhängigkeiten und zufälligen Schwankungen dar. Die Regressionsrechnung entwickelt Methoden, mit denen man solche Beziehungen in ihren funktionellen und ihren zufälligen Anteil aufspalten kann. Die Theorie der linearen Regression beschäftigt sich mit dem Fall, wo einer linearen Abhängigkeit  $\eta = \sum_{a=1}^m \theta_a x_a$  einer Beobachtungsgröße  $\eta$  von  $m$  genau bestimmbaren Argumenten  $x_a$  eine Zufallschwankung mit dem Mittel 0 und der Varianz  $\sigma^2$  überlagert ist. Die  $m$  Koeffizienten  $\theta_a$  sind feste unbekanntem Parameter. Bezeichnet man die bekannten Werte, die  $x_1, \dots, x_m$  bei der  $i$ -ten Beobachtung annehmen, der Reihe nach mit  $x_{1i}, \dots, x_{mi}$ , so ist die  $i$ -te Beobachtung ein Merkmal  $Y_i^j$  der eingangs beschriebenen Art. Die Ermittlung des funktionellen Anteiles in der Beziehung der  $x_a$  zu  $Y$  besteht dann, statistisch gesehen, in einer Schätzung der Parameter  $\theta_a$ .

Im folgenden wird durchwegs das Summationsübereinkommen, allerdings mit verschiedenen Summationsbereichen, verwendet: kommt in einem Produkt ein Zeiger genau zweimal vor, so ist über ihn zu summieren, und zwar im Falle eines Zeigers  $a, b, c, d, e$  oder  $f$  von 1 bis  $m$ , im Falle eines Zeigers  $i, j$  oder  $k$  von 1 bis  $n$ . In dieser Schreibweise ist also statt  $\sum_{a=1}^m x_{ai} \theta_a$  einfach  $x_{ai} \theta_a$  zu schreiben. Die Summen  $\sum_{i=1}^n x_{ai} x_{bi}$  werden kürzer mit  $x_{ai} x_{bi} = S^{ab}$  bezeichnet und stellen die Elemente einer  $m$ -reihigen quadratischen Matrix dar. Im Falle

$$\text{Det}(S^{ab}) \neq 0 \quad \dots (14)$$

ist  $(S_{ab})$  die inverse Matrix von  $(S^{ab})$ :  $(S_{ab}) = (S^{ab})^{-1}$ .

Aus der Determinantentheorie bekannt ist

**Satz 4:** Sind  $(x_{ai})$  und  $(y_{bj})$  zwei Matrizen mit  $m$  Zeilen und  $n (> m)$  Spalten, so erhält man die Determinante der Matrix  $(x_{ai} y_{bj})$ , indem man jede  $m$ -reihige Determinante von  $(x_{ai})$  mit der entsprechenden Determinante von  $(y_{bj})$  multipliziert und alle diese Produkte addiert.

Zwei Beweise dieses Satzes findet man in [5].

Aus Satz 4 ergibt sich sofort eine notwendige und hinreichende Bedingung für (14):

**Satz 5:** (14) gilt dann und nur dann, wenn der Rang von  $(x_{ai})$   $m$  ist.

Denn nach Satz 4 ist  $\text{Det}(S^{ab})$  die Summe der Quadrate aller  $m$ -reihigen Determinanten der Matrix  $(x_{ai})$ .

Die nächsten beiden Sätze geben die Bedeutung des Ranges von  $(x_{ai})$  für die Regressionsaufgabe an.

*Satz 6: Ist der Rang von  $(x_{ai})$  kleiner als  $m$ , so läßt sich die Anzahl der zu schätzenden Parameter  $\theta_a$  verringern.*

Denn wenn der Rang von  $(x_{ai})$  kleiner als  $m$  ist, läßt sich eine Zeile, etwa die  $m$ -te, als Linearkombination der anderen darstellen:  $x_{mi} = k_{a'} x_{a'i}$ , wo  $a'$  von 1 bis  $m-1$  läuft. Dann ist weiter  $\eta_i = x_{ai} \theta_a = x_{a'i} \theta_{a'} + x_{mi} \theta_m = x_{a'i} \theta_{a'} + k_{a'} x_{a'i} \theta_m = x_{a'i} (\theta_{a'} + k_{a'} \theta_m) = x_{a'i} \tilde{\theta}_{a'}$ , wenn man als neue Parameter für  $a' = 1, \dots, m-1$  die  $\tilde{\theta}_{a'} = \theta_{a'} + k_{a'} \theta_m$  einführt.

Es ist nützlich, sich die geometrische Bedeutung einer solchen Verringerung der Parameter an einem einfachen Beispiel klar zu machen. Es sei  $m = 2$ , so daß man die Regressionsaufgabe in der Form  $EZ_i = \theta_1 x_i + \theta_2 y_i$ ,  $VZ_i = \sigma^2$  ansetzen kann. Die Matrix

$$\begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix}$$

habe den Rang 1, so daß also für  $i = 1, \dots, n$ ,  $y_i = kx_i$  ist. Sind  $z_i$  die beobachteten Werte der  $Z_i$ , so liegen die Punkte  $(x_i, y_i, z_i)$  sämtlich in der zur  $xy$ -Ebene senkrechten Ebene  $kx - y = 0$ . Die Regressionsaufgabe, eine Ebene zu finden, die die Punkte  $(x_i, y_i, z_i)$  enthält, wird daher durch  $kx - y = 0$  in trivialer Weise erfüllt. Diese Lösung läßt sich aber nicht in der Gestalt  $z = \theta_1 x + \theta_2 y$  darstellen. Die Regressionsaufgabe  $EZ_i = (\theta_1 + k\theta_2) x_i = \tilde{\theta} x_i$  ist dagegen sinnvoll und bedeutet das Aufsuchen der Regressionsgeraden von  $Z$  in bezug auf  $x$  in der  $xz$ -Ebene. Analog läßt sich die Regressionsgerade von  $Z$  in bezug auf  $y$  in der  $yz$ -Ebene bestimmen. Beide Regressionsgeraden stellen die Projektionen der in der Ebene  $kx - y = 0$  gelegenen Regressionsgeraden von  $Z$  in bezug auf  $\sqrt{x^2 + y^2}$  auf die  $xz$ - bzw.  $yz$ -Ebene dar.

*Satz 7: Besteht zwischen  $\theta_1, \dots, \theta_m$  kein linearer Zusammenhang<sup>9)</sup>, und ist der Rang von  $(x_{ai})$  gleich  $m$ , so lassen sich die  $\eta_i$  nicht als Linearkombination von weniger als  $m$  Parametern darstellen.*

Denn sonst würden sich  $\theta_1, \dots, \theta_m$  als Linearkombination von  $l (< m)$  Größen  $\xi_1, \dots, \xi_l$  darstellen lassen; woraus sich so fort ein linearer Zusammenhang der  $\theta_a$  ergibt.

Der folgende Satz gestattet unter sehr allgemeinen Voraussetzungen eine zufriedenstellende Schätzung der  $\theta_a$  und von  $\sigma^2$ .

*Satz 8: Die Beobachtungen  $Y_i$  mit  $EY_i = x_{ai} \theta_a$  und  $VY_i = \sigma^2$  seien unkorreliert und der Rang von  $(x_{ai})$  sei  $m$ . Dann folgt:*

$$a) \quad T_a = S_{ab} x_{bi} Y_i, \quad a = 1, \dots, m, \quad \dots (15)$$

sind die eindeutig bestimmten besten erwartungstreuen linearen Schätzungen der  $\theta_a$ ,

$$b) \quad S^2 = \sum_{i=1}^n (Y_i - x_{ai} T_a)^2 / (n-m) \quad \dots (16)$$

ist eine erwartungstreue Schätzung von  $\sigma^2$ ,

$$c) \text{ die Kovarianzmatrix von } T = (T_1, \dots, T_m) \text{ ist } (S_{ab} \sigma^2).$$

Beweis: a) Die lineare Schätzung  $z_{ai} Y_i$  von  $\theta_a$  ist erwartungstreu, wenn  $E z_{ai} Y_i = z_{ai} E Y_i = z_{ai} x_{bi} \theta_b = \theta_a = \delta_{ab} \theta_b$ <sup>10)</sup> oder wenn  $(z_{ai} x_{bi} - \delta_{ab}) \theta_b = 0$  ist.

<sup>9)</sup> Wenn man sie als Variable im Raum der zulässigen Parameter auffaßt, vgl. Fußnote <sup>8)</sup>.

<sup>10)</sup>  $\delta_{ab}$  ist ebenso wie  $\delta_{ij}$  das Kroneckersche Delta:  $(\delta_{ab})$  und  $(\delta_{ij})$  sind die  $m$ - bzw.  $n$ -zeilige Einheitsmatrix.

Wegen der Unabhängigkeit der  $z_{ai}$  von den  $\theta_b$  muß dann für  $b = 1, \dots, m$

$$z_{ai} x_{bi} - \delta_{ab} = 0 \quad . . . (17)$$

sein.

Wegen (8) ist  $V_{z_{ai}} Y_i = \sigma^2 \sum_{i=1}^n z_{ai}^2$ . Diese Größe soll unter der Nebenbedingung (17) ein Minimum werden. Dazu hat man das freie Minimum von  $F(z_{a1}, \dots, z_{am}; \lambda_1, \dots, \lambda_m) = \sum_{i=1}^n z_{ai}^2 - 2 \sum_b \lambda_b (x_{bi} z_{ai} - \delta_{ab})$  zu ermitteln. Partielle Ableitung nach den Variablen gibt

$$z_{ai} - \lambda_b x_{bi} = 0 \quad . . . (18)$$

und (17). Einsetzen von  $z_{ai}$  aus (18) in (17) gibt  $\lambda_b x_{bi} x_{ci} = \delta_{ac}$  oder  $\lambda_b S^{bc} = \delta_{ac}$ . Wegen Satz 5 und der Symmetrie der Matrix  $(S^{bc})$  ergibt sich zunächst  $\lambda_b = S_{bc} \delta_{ac} = S_{ab}$  und schließlich  $z_{ai} = S_{ab} x_{bi}$ . Man zeigt leicht, daß die mit diesen  $z_{ai}$  gebildeten  $T_a$  auch wirklich die in a) angegebenen Eigenschaften haben.

c) Wegen (5) und (6) und wegen der Unkorreliertheit der  $Y_i$  ist

$$E(Y_i Y_j) = \delta_{ij} \sigma^2 + E Y_i \cdot E Y_j \quad . . . (19)$$

Weiter ist  $E(T_a T_b) = E(S_{ac} x_{ci} Y_i S_{bd} x_{dj} Y_j) = x_{ci} x_{dj} S_{ac} S_{bd} E(Y_i Y_j)$ . Einsetzen von (19) führt auf  $x_{ci} x_{dj} S_{ac} S_{bd} (\delta_{ij} \sigma^2 + E Y_i \cdot E Y_j) = S^{cd} S_{ac} S_{bd} \sigma^2 + x_{ci} x_{dj} S_{ac} S_{bd} x_{ei} \theta_e x_{jf} \theta_f = \delta_{ad} S_{bd} \sigma^2 + S^{ce} S^{df} S_{ac} S_{bd} \theta_e \theta_f$ , also

$$E(T_a T_b) = S_{ab} \sigma^2 + \theta_a \theta_b \quad . . . (20)$$

Wegen (6) ist dann

$$C(T_a, T_b) = S_{ab} \sigma^2,$$

womit c) bewiesen ist.

b)  $E[(n-m)S^2] = E[(Y_i - x_{ai} T_a)(Y_i - x_{bi} T_b)] = E(Y_i Y_i) - 2 x_{ai} x_{bj} S_{ab} E(Y_i Y_j) + x_{ai} x_{bi} E(T_a T_b)$ . Einsetzen von (19) und (20) liefert.  $\delta_{ii} \sigma^2 + E Y_i \cdot E Y_i - 2 x_{ai} x_{bj} S_{ab} (\delta_{ij} \sigma^2 + E Y_i \cdot E Y_j) + S^{ab} (S_{ab} \sigma^2 + \theta_a \theta_b) = n \cdot \sigma^2 + S^{ab} \theta_a \theta_b - 2 S^{ab} S_{ab} \sigma^2 - 2 x_{ai} x_{bj} S_{ab} x_{ci} \theta_c x_{dj} \theta_d + S^{ab} S_{ab} \sigma^2 + S^{ab} \theta_a \theta_b$ . Der vierte Summand ist  $-2 S^{ac} S^{bd} S_{ab} \theta_c \theta_d = -2 \delta_{bc} S^{bd} \theta_c \theta_d = -2 S^{cd} \theta_c \theta_d$  und hebt sich daher mit dem zweiten und dem letzten auf. Der dritte gibt zusammen mit dem fünften  $-\delta_{aa} \sigma^2 = -m \sigma^2$ . Insgesamt wird also  $E[(n-m)S^2] = (n-m)\sigma^2$ , womit b) bewiesen ist.

Eine Aussage über die Varianz von  $S^2$  fehlt, da keine Voraussetzung über die vierten Momente der  $Y_i$  gemacht werden.

Ergänzt man die Voraussetzungen von Satz 8 durch die Annahme, daß die  $Y_i$  nach Gauß verteilt sind, so werden die  $Y_i$  nach Satz 1 unabhängig und man erhält

*Satz 9: Die Beobachtungen  $Y_1, \dots, Y_n$  seien unabhängig voneinander nach  $G(x_{ai}, \theta_a, \sigma^2)$  verteilt und  $(x_{ai})$  habe den Rang  $m$ .*

a) *Dann sind die eindeutig bestimmten besten erwartungstreuen linearen Schätzungen (15) auch die plausiblen Schätzungen der  $\theta_a$ .  $T = (T_1, \dots, T_m)$  ist nach Gauß verteilt mit den Mitteln  $\theta_a$  und der Kovarianzmatrix  $(S_{ab} \sigma^2)$ .  $S_{ab} \sigma^2$  wird durch  $S_{ab} S^2$  erwartungstreu geschätzt.*

b)  $(n - m) S^2/\sigma^2$  ist unabhängig von  $S^{ab} (T_a - \theta_a) (T_b - \theta_b)$  wie  $\chi^2$  mit  $(n - m)$  F. g. verteilt. Es ist  $ES^2 = \sigma^2$  und  $VS^2 = 2 \sigma^4/(n-m)$ .  $VS^2$  wird durch  $2 S^4/(n - m + 2)$  erwartungstreu geschätzt.

Beweis: Da mit den Voraussetzungen dieses Satzes auch die des vorhergehenden erfüllt sind, gelten die Aussagen a) bis c) von Satz 8 auch hier.

a) Daß man unter der Voraussetzung der Gaußverteilung (15) als plausible Schätzungen ableiten kann, gehört seit Gauß zum Bestand der Ausgleichsrechnung und braucht hier nicht vorgeführt zu werden. Als lineare Funktionen von Gaußmerkmalen sind die  $T_a$  nach Satz 2 selbst nach Gauß verteilt. Diese Verteilung ist nach § 3 durch die Mittel  $ET_a = \theta_a$  und die Kovarianzen  $C(T_a, T_b) = S_{ab} \sigma^2$  vollständig bestimmt.

b) Zur Ermittlung der Verteilung von (16) führt man folgende Zerlegung durch:

$$Q = \sum_{i=1}^n (Y_i - x_{ai} \theta_a)^2 = \sum_{i=1}^n [(Y_i - x_{ai} T_a) + x_{ai} (T_a - \theta_a)]^2 = (Y_i - x_{ai} T_a) (Y_i - x_{bi} T_b) + 2 x_{bi} (Y_i - x_{ai} T_a) (T_b - \theta_b) + x_{ai} x_{bi} (T_a - \theta_a) (T_b - \theta_b) = Q_1 + Q_2 + Q_3.$$

$$Q_2 \text{ verschwindet wegen } x_{bi} (Y_i - x_{ai} T_a) = x_{bi} Y_i - x_{bi} x_{ai} S_{ac} x_{cj} Y_j = x_{bi} Y_i - S^{ab} S_{ac} x_{cj} Y_j = x_{bi} Y_i - \delta_{bc} x_{cj} Y_j = x_{bi} Y_i - x_{bj} Y_j = 0.$$

$Q$ ,  $Q_1$  und  $Q_3$  werden umgeformt, indem man in ihnen statt der Merkmale  $Y_i$  die Merkmale  $Z_i = Y_i - EY_i = Y_i - x_{ai} \theta_a$  einführt, deren Mittel 0 sind.

Zunächst ist  $Q = Z_i Z_i$ . Ferner ist  $T_a - \theta_a = S_{ab} x_{bi} Y_i - \theta_a = S_{ab} x_{bi} (Z_i + x_{ci} \theta_c) - \theta_a = S_{ab} x_{bi} Z_i + S_{ab} S^{bc} \theta_c - \theta_a = S_{ab} x_{bi} Z_i + \delta_{ac} \theta_c - \theta_a$ , also

$$T_a - \theta_a = S_{ab} x_{bi} Z_i. \quad \dots (21)$$

$$\begin{aligned} \text{Aus } Q_1 &= [(Y_i - EY_i) - x_{ai} (T_a - \theta_a)] [(Y_i - EY_i) - x_{bi} (T_b - \theta_b)] = \\ &= (Z_i - x_{ai} x_{cj} S_{ac} Z_j) (Z_i - x_{bi} x_{dk} S_{bd} Z_k) = (\delta_{ij} - x_{ai} x_{cj} S_{ac}) Z_j \\ &(\delta_{ik} - x_{bi} x_{dk} S_{bd}) Z_k \text{ erhält man nach kurzer Vereinfachung } Q_1 = (\delta_{ij} - x_{ai} x_{bj} \\ &S_{ab}) Z_i Z_j. \end{aligned}$$

Mit Hilfe von (21) bekommt man sofort  $Q_3 = x_{ai} x_{bj} S_{ab} Z_i Z_j$ .

Insgesamt ergibt sich also die Zerlegung  $Q = Q_1 + Q_3$  oder

$$Z_i Z_i = (\delta_{ij} - x_{ai} x_{bj} S_{ab}) Z_i Z_j + x_{ai} x_{bj} S_{ab} Z_i Z_j. \quad \dots (22)$$

Für  $c = 1, \dots, m$  ist  $(\delta_{ij} - x_{ai} x_{bj} S_{ab}) x_{ci} = x_{cj} - S^{ac} S_{ab} x_{bj} = x_{cj} - x_{cj} = 0$ . Der Rang von  $(x_{ci})$  ist  $m$ , daher bestehen zwischen den Zeilen der Matrix  $(\delta_{ij} - x_{ai} x_{bj} S_{ab})$  mindestens  $m$  linear unabhängige Beziehungen, so daß  $Q_1$  höchstens den Rang  $n - m$  hat.

$Q_3$  läßt sich durch eine reguläre lineare Transformation, deren erste  $m$  Zeilen  $V_a = x_{ai} Z_i$  lauten, in eine Form der  $m$  Merkmale  $V_a$  überführen und hat daher höchstens den Rang  $m$ .

Nach Satz 3 sind daher  $Q_1/\sigma^2 = (n - m) S^2/\sigma^2$  und  $Q_3/\sigma^2 = (T_a - \theta_a) (T_b - \theta_b) S^{ab}/\sigma^2$  unabhängig voneinander nach  $\chi^2$  mit  $n - m$  bzw.  $m$  F. g. verteilt. Aus (11 a) und (7b) folgt dann wieder  $E(Q_1/\sigma^2) = n - m$  oder  $ES^2 = \sigma^2$ . Aus (11 b) und (7b) ergibt sich die Varianz von  $S^2$ :  $V(Q_1/\sigma^2) = 2 (n - m)$  oder  $VS^2 = 2 \sigma^4/(n - m)$ .

Wegen (5) ist  $ES^4 = VS^2 + (ES^2)^2$ . Daher ist  $ES^4 = 2\sigma^4/(n-m) + \sigma^4 = (n-m+2)\sigma^4/(n-m)$ , so daß  $2S^4/(n-m+2)$  eine erwartungstreue Schätzung von  $VS^2$  ist. Damit ist Satz 9 zur Gänze bewiesen.

Wenn die Beobachtungen  $Y_1, \dots, Y_n$  verschiedene Varianzen  $\sigma_i^2 = \frac{\sigma^2}{p_i}$  haben, wo die Gewichte  $p_i$  bekannt sind und  $\sigma^2$  unbekannt ist, so betrachtet man anstelle der Regressionsaufgabe für die  $Y_1, \dots, Y_n$  die Regressionsaufgabe für die Beobachtungen  $\tilde{Y}_1 = \sqrt{p_1} Y_1, \dots, \tilde{Y}_n = \sqrt{p_n} Y_n$  (nicht summieren,  $n$  ist eine feste Zahl!). Wegen (7a) ist  $E\tilde{Y}_i = \tilde{x}_{ai}\theta_a$ , wo  $\tilde{x}_{a1} = \sqrt{p_1} x_{a1}, \dots, \tilde{x}_{an} = \sqrt{p_n} x_{an}$ . Wegen (7b) ist  $V\tilde{Y}_i = \sigma^2$ . Die Lösung der Regressionsaufgabe für die Beobachtungen  $\tilde{Y}_i$  kann also mit Hilfe der Sätze 8 und 9 durchgeführt werden. Aus der Schätzung  $S^2$  von  $\sigma^2$  ergeben sich dann Schätzungen  $S_i^2 = S^2/p_i$  der  $\sigma_i^2$ .

**§ 6. Die Ausgleichung vermittelnder Beobachtungen.** Die Ausgleichung vermittelnder Beobachtungen ordnet sich nun mühelos in den kleinen Ausschnitt der linearen Regressionstheorie ein, der im letzten Paragraphen in Form einiger markanter Definitionen und Sätze dargestellt wurde.

Jeder Ausgleichung vermittelnder Beobachtungen liegt ein funktionaler Zusammenhang

$$y = f(x_1, \dots, x_m; \theta_1, \dots, \theta_m) \quad \dots (23)$$

zugrunde. Dabei sind die  $x_1, \dots, x_m$  oft verfügbare und stets bekannte Argumente, dagegen  $\theta_1, \dots, \theta_m$  unbekannte Parameter, die zu bestimmen sind. Die Funktionswerte  $y_i$  werden für  $n$  1-Tupel  $x = (x_{1i}, \dots, x_{mi})$ ,  $i = 1, \dots, n$ , beobachtet. Da diese Beobachtungen aber mit unvermeidlichen Fehlern behaftet sind, ist die  $i$ -te Beobachtung  $Y_i$  als Beobachtung im Sinne der Stochastik, daß heißt als (zufälliges) Merkmal aufzufassen. Die Ausgleichung vermittelnder Beobachtungen besteht also statistisch gesehen im Schätzen der Parameter  $\theta_1, \dots, \theta_m$  auf Grund von beobachteten Werten  $y_1, \dots, y_n$ . Dabei darf vorausgesetzt werden, daß die  $\theta_1, \dots, \theta_m$ , solange man sie als Variable betrachtet, unabhängig sind und daß die Anzahl  $n$  der Beobachtungen größer ist als die Anzahl  $m$  der unbekannt Parameter.

Wenn nun, wie in der Geodäsie, Physik und Chemie, die Beobachtungen sehr genau, also die Varianzen der  $Y_i$  sehr klein sind, lassen sich im allgemeinen aus den ersten  $m$  beobachteten Werten  $y_1, \dots, y_m$  sehr genaue Näherungswerte  $\overset{\circ}{\theta}_1, \dots, \overset{\circ}{\theta}_m$  der unbekannt Parameter berechnen. Die Differenzen  $\theta_a - \overset{\circ}{\theta}_a$  zwischen den tatsächlichen und den Näherungswerten der Parameter werden dann auf jeden Fall klein sein, so daß man unter den entsprechenden Differenzierbarkeitsvoraussetzungen über (23) die Taylorentwicklung dieser Funktion nach den Gliedern erster Ordnung abbrechen darf:

$$y - f(x; \overset{\circ}{\theta}) = \sum_{a=1}^m \frac{\partial f}{\partial \theta_a}(x; \overset{\circ}{\theta})(\theta_a - \overset{\circ}{\theta}_a). \quad \dots (24)$$

Anstatt der ursprünglichen Beobachtungen  $Y_i$  und Parameter  $\theta_a$  betrachtet man die neuen Beobachtungen  $Y_i - f(x; \overset{\circ}{\theta})$  bzw. neuen Parameter  $\theta_a - \overset{\circ}{\theta}_a$ .

Die Beschränkung 'auf die Ausgleichung vermittelnder Beobachtungen, die in den unbekanntem Parametern linear sind, findet ihre Begründung in der Kleinheit der  $\theta_a - \hat{\theta}_a$ . Die den Sätzen (8) und (9) innewohnende Beschränkung auf Parameterschätzungen, die in den Beobachtungen linear sind, rechtfertigt sich durch die Kleinheit der absoluten Beträge der  $Y_i - f(x; \hat{\theta})$ .

Schreibt man nun statt  $Y_i - f(x; \hat{\theta})$  und  $\theta_a - \hat{\theta}_a$  wieder  $Y^i$  bzw.  $\theta_a$ , so zeigt sich, daß die Aufgabe der Ausgleichung vermittelnder Beobachtungen zusammenfällt mit der Aufgabe der linearen Regression, wie sie zu Beginn von § 5 formuliert wird. Alle Ergebnisse der traditionellen Ausgleichsrechnung und noch einiges mehr ergeben sich dann aus den entsprechenden Sätzen der Regressionstheorie, deren wichtigste die in § 5 vorgeführten Sätze 8 und 9 sind.

**§ 7. Zusammenfassung und Schluß.** In dieser Arbeit wird also gezeigt, daß die Ausgleichung vermittelnder Beobachtungen einen Teil der linearen Regressionstheorie bildet, die seit den Tagen von C. F. Gauß und F. R. Helmert eine erhebliche Entwicklung erfahren hat. Die Erkenntnis dieser Tatsache ist keineswegs neu, gewinnt aber heute aus zwei Gründen eine ständig wachsende Bedeutung: einerseits sind die fortgeschrittensten theoretischen Methoden gerade gut genug, um aus dem sich ständig erweiternden Feld experimenteller Erfahrungen das Äußerste an Auskunft herauszuholen, was diese Daten zu geben imstande sind. Andererseits zwingt die zunehmende Verfeinerung, Verästelung und Vertiefung moderner mathematischer Methoden alle an diesem Fortschritt Beteiligten und Interessierten, den heutigen Stand der Dinge aus den ökonomischen Gründen in seiner rationellsten Form darzustellen.

Eine weitere Arbeit wird der Ausgleichung bedingter Beobachtungen im Rahmen der mathematischen Statistik gewidmet sein.

### § 8. Literatur.

[1]: W. G. Cochran: The Distribution of Quadratic Forms in a Normal System. Proceedings of the Cambridge Philosophical Society 30 (1933).

[2]: H. Cramer: Mathematical Methods of Statistics. University Press, Princeton, 1946.

[3]: R. A. Fisher: On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society, London, 222 (1922).

[4]: F. R. Helmert: Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und einige damit im Zusammenhang stehende Fragen, Zeitschrift für Mathematik und Physik 21 (1876).

[5]: G. Kowalewski: Einführung in die Determinantentheorie, Veit u. Co., Leipzig, 1909.

[6]: J. W. Linnik: Die Methode der kleinsten Quadrate und Grundlagen einer Theorie der Versuchsauswertungen. Staatl. Verlag für physikalisch-mathematische Literatur, Moskau, 1958 (russisch).

[7]: C. R. Rao: Advanced Statistical Methods in Biometric Research, Wiley, New York, 1952.

[8]: B. L. van der Waerden: Mathematische Statistik. Springer, Berlin, 1957.

Weitere Literaturhinweise auf Arbeiten über lineare Regressionstheorie finden sich vor allem in [7]. [6] konnte der Verfasser bei der Abfassung dieser Arbeit aus sprachlichen Gründen leider nicht verwerten. Der Deutsche Verlag der Wissenschaften zu Berlin bereitet eine deutsche Übersetzung dieses Buches vor. Den Hinweis auf [6] verdankt der Verfasser Herrn W. Richter in Dresden.