Zeitschrift: Bulletin de l'Association suisse des électriciens
Herausgeber: Association suisse des électriciens
Band: 39 (1948)
Heft: 6

Artikel: Mathematische Statistik und Tarifwesen II
Autor: Morel, C.
DOI: https://doi.org/10.5169/seals-1057938

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 22.08.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

39e année | N° 6 | Samedi, le 20 Mars 1948

BULLETIN

DE L'ASSOCIATION SUISSE DES ELECTRICIENS

Mathematische Statistik und Tarifwesen II[1])

Von *Ch. Morel*, Feldmeilen

519.24 : 621.317.8

In diesem zweiten Aufsatz werden zunächst die wichtigsten Eigenschaften der Gaußschen Fehlerfunktion gestreift, da diese die Grundlage bildet, auf die sich die praktischen Verfahren zur Auswertung und Prüfung von Statistiken aufbauen. Eine besondere Rolle spielt dabei das Fehlerintegral, dessen graphische Darstellung eine S-förmige Kurve ergibt. Durch eine Maßstabtransformation erhält man das Wahrscheinlichkeitsnetz, in welchem die Fehlerintegralkurve, auch Summenhäufigkeit genannt, als Gerade, und die glockenförmige Fehlerkurve als eine Art Hyperbel erscheint, was die Lösung der meisten sich stellenden Aufgaben wesentlich erleichtert und auch die Analyse der erhaltenen Kurven ohne mathematische Hilfsmittel auf einfachste Weise gestattet. Beispiele praktischer Anwendung dieser Methoden auf das Tarifwesen ergänzen die Darstellungen.

Dans ce deuxième article, l'auteur expose brièvement les principales caractéristiques de la fonction d'erreurs de Gauss, qui est à la base des procédés pratiques d'utilisation et de contrôle des statistiques. Dans ce domaine, l'intégrale des erreurs, représentée graphiquement par une courbe en S, joue un rôle particulier. A l'aide d'une transformation d'échelle, on obtient un réseau de probabilité, dans lequel la courbe intégrale des erreurs (appelée également courbe de fréquences cumulées) devient une droite, tandis que la courbe en cloche se présente sous la forme d'une sorte d'hyperbole, ce qui facilite grandement la solution de la plupart des problèmes posés et permet l'analyse des courbes, sans avoir recours à des procédés mathématiques. Cet exposé est complété par des exemples d'application dans le domaine de la tarification.

5. Die Gaußsche Fehlerfunktion und die Summenhäufigkeiten

a) Geometrische Eigenschaften

Die Grundform der Gaußschen Fehlerfunktion

$$f = \frac{1}{s\sqrt{2\pi}}\, e^{-\frac{(v-\bar{v}^2)}{2s^2}} \qquad (26)$$

kann, wie bereits angedeutet, durch Setzen von

$$y = 2\sqrt{2\pi}\, sf \text{ und } x = \frac{v-\bar{v}}{s} \qquad (27)$$

in die reduzierte oder normale Form

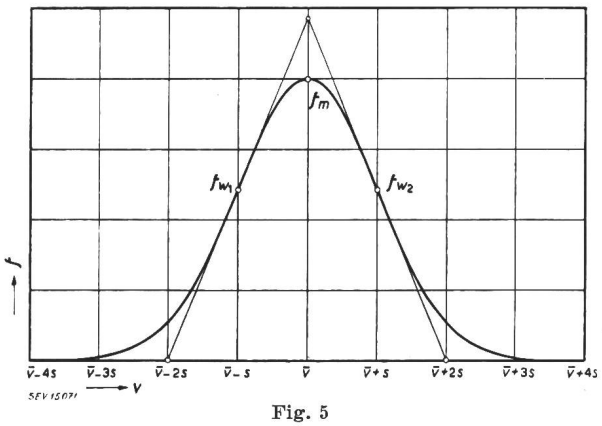

Fig. 5
Grundform der Fehlerfunktion

$$f_m = \frac{1}{s\sqrt{2\pi}} \qquad f_{w1} = f_{w2} = \frac{1}{s\sqrt{2\pi}} \cdot e^{-\frac{1}{2}}$$

$$y = 2\, e^{-\frac{x^2}{2}} \qquad (28)$$

übergeführt werden.

Charakteristisch für den Verlauf der Kurve sind ihre Maxima, Minima und Wendepunkte. Um diese zu finden, setzt man die erste, bzw. die zweite Ableitung der Funktion nach der Variabeln gleich null. Diese Operationen führen zu folgenden Werten:

Tabelle VI

Funktion:	$f = \frac{1}{s\sqrt{2\pi}}\, e^{-\frac{(v-\bar{v})^2}{2s^2}}$	$y = 2e^{-\frac{x^2}{2}}$
Maximum	$v_m = \bar{v} \quad f_m = \frac{1}{s\sqrt{2\pi}}$	$x_m = 0 \; y_m = 2$
Wendepunkte	$v_w = \bar{v} \pm s$ $f_w = \frac{1}{s\sqrt{2\pi}}\, e^{-\frac{1}{2}}$	$x_w = \pm 1$ $y_w = 2\, e^{-\frac{1}{2}}$
Wendetangenten — Schnittpunkte mit Abszissenaxe	$v = \bar{v} \pm 2s$	$x = \pm 2$

Fig. 5 gibt diese Beziehungen für die Grundform, Fig. 6 für die Normal- oder reduzierte Form der Fehlerfunktion wieder.

b) Die einfache Summenhäufigkeit (F-Integral)

In der Auswertung von Statistiken stellt sich oft die Frage, wieviele Einzelwerte der Variabeln einen bestimmten Wert v dieser Variabeln nicht übersteigen, mit andern Worten, wie gross die

[1]) siehe den I. Teil im Bull. SEV Bd. 38(1947), Nr. 6, S. 141...149.

Summe aller Häufigkeiten der Variabeln zwischen ihrem untersten Wert ($-\infty$) und dem bestimmten Wert v ist. Diese Summe, die *einfache Summenhäufigkeit F*, schreibt sich

$$F(v) = \int_{-\infty}^{v} f(v) \cdot dv \qquad (29)$$

Die Summenhäufigkeit stellt zugleich die Wahrscheinlichkeit für das Auftreten irgendeines Einzel-

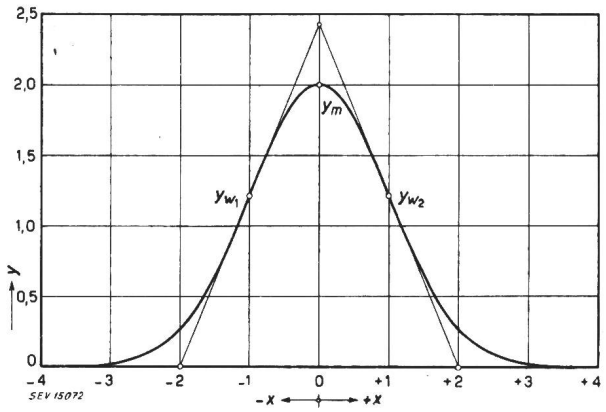

Fig. 6
Reduzierte Form der Fehlerfunktion

$y_m = 2 \qquad y_{w1} = y_{w2} = 2\, e^{-\frac{1}{2}}$

wertes der Statistik zwischen den Grenzen $-\infty$ und v dar. Erstreckt sich die Summierung auf die ganze Statistik, also zwischen den Grenzen $-\infty$ und $+\infty$, so wird die Wahrscheinlichkeit zur Gewissheit, da jeder beliebige Einzelwert innert dieser Grenzen liegen muss. Infolgedessen wird

$$F(\infty) = \int_{-\infty}^{+\infty} f(v)\, dv = 1 \qquad (30)$$

Für die Grundform der Fehlerfunktion gemäss Gleichung (29) lautet das Integral:

$$F(v) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{v} e^{-\frac{(v-\bar{v})^2}{2s^2}}\, dv \qquad (31)$$

Setzt man $u = \dfrac{v-\bar{v}}{\sqrt{2}\, s}$, so wird

$$F(v) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{u} e^{-u^2}\, du$$

oder $$F(v) = \frac{1}{\sqrt{\pi}} \left(\int_{-\infty}^{0} e^{-u^2}\, du + \int_{0}^{u} e^{-u^2}\, du \right).$$

Das bestimmte Integral zwischen $-\infty$ und 0 ist bekannt:

$$\int_{-\infty}^{0} e^{-u^2}\, du = \frac{\sqrt{\pi}}{2}$$

So erhält man

$$F(v) = \frac{1}{2} \left(1 + \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-u^2}\, du \right) \qquad (32)$$

Das verbleibende bestimmte Integral zwischen 0 und u kann nur durch Reihenentwicklung gelöst werden. Es bestehen jedoch Tafeln, wie für die Logarithmen, aus denen die einzelnen Werte von

$$\varphi(u) = \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-u^2}\, du$$

herausgelesen werden können (siehe z. B. Tab. VII).

Tabelle VII

u	$\varphi(u)$	Diff.	u	$\varphi(u)$	Diff.
0,00	0,000	56	0,90	0,797	24
0,05	0,056	56	0,95	0,821	22
0,10	0,112	56	1,00	0,843	37
0,15	0,168	55	1,10	0,880	30
0,20	0,223	53	1,20	0,910	24
0,25	0,276	53	1,30	0,934	18
0,30	0,329	50	1,40	0,952	14
0,35	0,379	49	1,50	0,966	10
0,40	0,428	47	1,60	0,976	8
0,45	0,475	45	1,70	0,984	5
0,50	0,520	43	1,80	0,989	4
0,55	0,563	41	1,90	0,993	2
0,60	0,604	38	2,00	0,995	2
0,65	0,642	36	2,10	0,997	1
0,70	0,678	33	2,20	0,998	1
0,75	0,711	31	2,30	0,999	0
0,80	0,742	29	2,40	0,999	1
0,85	0,771	26	2,50	1,000	
0,90	0,797				

Numerische Werte der Funktion

$$\varphi(u) = \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-u^2}\, du$$

Für die reduzierte Form der Fehlerfunktion erhält man nach Einsetzen von

$$u = \frac{x}{\sqrt{2}}:$$

$$F(x) = \sqrt{2\pi} \left(1 + \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-u^2}\, du \right) \qquad (33)$$

Die Funktion $F(v)$ ist in Fig. 7 graphisch dargestellt.

c) Die Summenhäufigkeit der absoluten Fehler (W-Integral, W-Funktion)

Man kann sich auch fragen, wieviele Einzelwerte um nicht mehr als einen bestimmten absoluten Betrag $z = (v - \bar{v})$ vom Mittelwert abweichen. Diese Anzahl nennt man die *Summenhäufigkeit W der absoluten Fehler*, bzw. der absoluten Abweichungen vom Mittelwert. Die symbolische Formel dafür lautet:

$$W(|v-\bar{v}|) = \int_{\bar{v}-(v-\bar{v})}^{\bar{v}+(v-\bar{v})} f(v)\,dv \qquad (34)$$

Dieser Ausdruck entspricht der Wahrscheinlichkeit des Auftretens eines Einzelwertes der Statistik zwischen $\bar{v}-(v-\bar{v})$ und $\bar{v}+(v-\bar{v})$ bzw. $\bar{v}-z$ und $\bar{v}+z$.

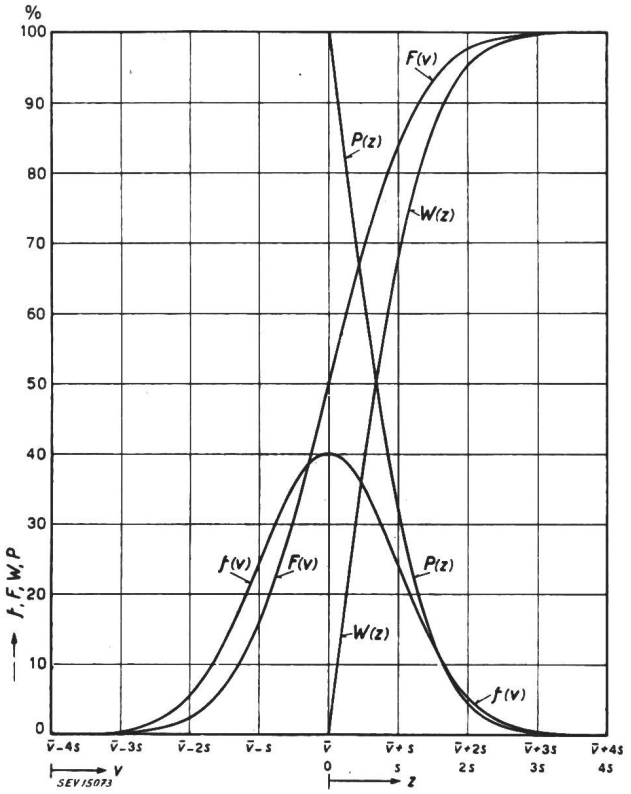

Fig. 7
Grundform der Fehlerfunktion und ihrer Summenhäufigkeiten

Für die Gaußsche Fehlerfunktion (26) kann man schreiben, wenn $(v-\bar{v}) = z$ gesetzt wird:

$$W(z) = \frac{1}{s\sqrt{2\pi}} \int_{-z}^{+z} e^{-\frac{z^2}{2s^2}}\,dz$$

oder, da die Funktion symmetrisch ist:

$$W(z) = \frac{2}{s\sqrt{2\pi}} \int_{0}^{z} e^{-\frac{z^2}{2s^2}}\,dz \qquad (35)$$

Macht man die Substitution $u = \frac{z}{\sqrt{2}\,s}$,

so erhält man

$$W(z) = \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-u^2}\,du \qquad (36)$$

Die numerischen Werte von W können den bereits erwähnten Funktionen-Tafeln entnommen werden.

Für die reduzierte Form der Fehlerfunktion (28) ergibt sich, wenn $u = \frac{x}{\sqrt{2}}$ gesetzt wird

$$W(|x|) = 2\sqrt{2\pi} \cdot \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-u^2}\,du \qquad (37)$$

Die Kurve der W-Funktion ist ebenfalls in Fig. 7 eingezeichnet.

d) Die Summenhäufigkeit der einen bestimmten Wert übersteigenden absoluten Fehler (P-Integral)

Oft begegnet man der Frage, bei wieviel Einzelwerten der Fehler einen bestimmten Betrag $z = (v-\bar{v})$ übersteigt, oder wie gross die Wahrscheinlichkeit ist, dass ein Einzelwert um mehr als diesen Betrag z vom Mittelwert $\bar{v}$ abweicht. Man sieht ohne weiteres, dass diese Funktion, die $P(z)$ geschrieben wird, die Funktion $W(z)$ zu 1 ergänzt:

$$P(z) = 1 - W(z) \qquad (38)$$

Für die normale Verteilung gemäss (26) kann man also setzen, wenn $z = (v-\bar{v})$

$$P(z) = 1 - \frac{2}{s\sqrt{2\pi}} \int_{0}^{z} e^{-\frac{z^2}{2s^2}}\,dz \qquad (39)$$

oder, wenn $u = \frac{z}{\sqrt{2}\,s}$ ist

$$P(z) = 1 - \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-u^2}\,du \qquad (40)$$

Aehnlich gilt für die reduzierte Fehlerfunktion (23)

Tabelle VIII

P	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	P
0,00	∞	1,644 854	1,281 552	1,036 433	0,841 621	0,674 490	0,524 401	0,385 320	0,253 347	0,125 661	0,00
0,01	2,575 829	1,598 193	1,253 565	1,015 222	0,823 894	0,658 838	0,510 073	0,371 856	0,240 426	0,113 039	0,01
0,02	2,326 348	1,554 774	1,226 528	0,994 458	0,806 421	0,643 345	0,495 850	0,358 459	0,227 545	0,100 434	0,02
0,03	2,170 090	1,514 102	1,200 359	0,974 114	0,789 192	0,628 006	0,481 727	0,345 126	0,214 702	0,087 845	0,03
0,04	2,053 749	1,475 791	1,174 987	0,954 165	0,772 193	0,612 813	0,467 699	0,331 853	0,201 893	0,075 270	0,04
0,05	1,959 964	1,439 531	1,150 349	0,934 589	0,755 415	0,597 760	0,453 762	0,318 639	0,189 118	0,062 707	0,05
0,06	1,880 794	1,405 072	1,126 391	0,915 365	0,738 847	0,582 841	0,439 913	0,305 481	0,176 374	0,050 154	0,06
0,07	1,811 911	1,372 204	1,103 063	0,896 473	0,722 479	0,568 051	0,426 148	0,292 375	0,163 458	0,037 608	0,07
0,08	1,750 686	1,340 755	1,080 319	0,877 896	0,706 303	0,553 385	0,412 463	0,279 319	0,150 969	0,025 069	0,08
0,09	1,695 398	1,310 579	1,058 122	0,859 617	0,690 309	0,538 836	0,398 855	0,266 311	0,138 304	0,012 523	0,09

$$P(x) = 2\sqrt{2\pi}\left(1 - \frac{2}{\sqrt{\pi}}\int_0^u e^{-u^2}\,du\right), \text{ wobei}$$

$$u = \frac{x}{\sqrt{2}} \qquad (41)$$

In Fig. 7 ist die der Funktion $P(z)$ entsprechende Kurve eingetragen.

Für die Funktion $P(x)$, die bei der Prüfung von statistischen Masszahlen eine wichtige Rolle spielt, findet man in Tabelle VIII für verschiedene Werte von $P(x)$, bezogen auf $P(0) = 1$, die entsprechenden Werte von x.

e) Die Bedeutung der verschiedenen Summenhäufigkeiten

Liegt eine Statistik vor, so interessiert in erster Linie, ob ihre Verteilung normal ist, d. h. ob sie dem Gaußschen Fehlergesetz gehorcht. Mit Hilfe der abgeleiteten Summenhäufigkeits-Funktionen lassen sich gewisse Normen aufstellen, die eine summarische Ueberprüfung der Verteilung gestatten.

So findet man bei numerischer Auswertung der *W-Funktion* (35) für die entsprechenden Integrationsgrenzen, dass bei einer normalen Verteilung 0,683 oder rund $^2/_3$ aller absoluten Fehler, bzw. Abweichungen vom Mittelwert, kleiner als die mittlere quadratische Abweichung s sein, d. h. innerhalb der Grenzen $\bar{v} \pm s$ liegen müssen. Ferner müssen 0,954 oder rund 95 % aller absoluten Fehler kleiner sein als $2\,s$ und 0,997 oder 99,7 % kleiner als $3\,s$. Praktisch dürfen also nicht mehr als 3 ‰ aller Einzelwerte um mehr als $\pm\,3\,s$ vom Mittelwert abweichen.

Auf diese Weise kann man auch ermitteln, innerhalb welcher Grenzen ein bestimmter Prozentsatz aller Einzelwerte einer Statistik sich bei normaler Verteilung befinden müssen. Die Ausrechnung ergibt z. B., dass

50 % aller Werte innerhalb der Grenzen $\bar{v} \pm 0{,}675\,s$
90 % aller Werte innerhalb der Grenzen $\bar{v} \pm 1{,}645\,s$
95 % aller Werte innerhalb der Grenzen $\bar{v} \pm 1{,}956\,s$
98 % aller Werte innerhalb der Grenzen $\bar{v} \pm 2{,}326\,s$
99 % aller Werte innerhalb der Grenzen $\bar{v} \pm 2{,}576\,s$
liegen müssen.

Kann die Verteilung einer Statistik zum vorneherein als normal angenommen werden, so lässt sich auf einfache Weise die mittlere quadratische Abweichung in erster Annäherung errechnen. Man ermittelt die Grenzen $v + z$ und $v - z$, innerhalb welcher z. B. 95 % aller absoluten Fehler liegen, und dividiert den so erhaltenen Grenzfehler z durch 1,956.

6. Die Prüfung von statistischen Masszahlen

a) Die Sicherheitsschwelle

Jede Statistik muss als eine Stichprobe aus einer Grundgesamtheit betrachtet werden. Entnimmt man dieser Grundgesamtheit eine Anzahl voneinander unabhängiger Stichproben, so werden ihre statistischen Masszahlen nicht in allen Fällen mit denjenigen der Grundgesamtheit übereinstimmen; sie werden streuen mit einer Verteilung, die derjenigen der Grundgesamtheit entspricht. Im Falle einer normalen Verteilung ist somit theoretisch zu erwarten, dass z. B. der Durchschnitt $\bar{v}$ oder die Streuung s^2 der einzelnen Stichproben innerhalb der Grenzen $-\infty$ und $+\infty$ um den Durchschnitt μ oder die Streuung σ^2 der Grundgesamtheit herum streuen werden. Praktisch kann aber nicht jede Abweichung als zufällig hingenommen werden. Es muss eine Grenze definiert werden, innerhalb welcher der Fehler als zufällig gilt und bei deren Ueberschreitung die Abweichung als wesentlich oder gesichert zu betrachten ist. Diese Grenze nennt man Sicherheitsschwelle.

In der mathematischen Statistik ist es nun üblich zu sagen, ein Fehler oder eine Abweichung vom Mittelwert sei *wesentlich*, wenn der entsprechende P-Wert kleiner als 0,01 (1 %) und *zufällig*, wenn der entsprechende P-Wert grösser als 0,05 (5 %) ist. Mit anderen Worten, eine Abweichung gilt als gesichert, wenn sie grösser als 2,576 s und als zufällig, wenn sie kleiner als 1,956 s ist. Für Werte, die dazwischen liegen, muss erst eine besondere Untersuchung zeigen, ob die Abweichung gesichert ist oder nicht. Für praktische Berechnungen und um sich diese besondere Untersuchung zu ersparen, kann man mit der einen Grenze von $P = 0{,}05$ (5 %) auskommen. Damit nimmt man in Kauf, dass in 5 % von allen möglichen Fällen eine Abweichung als wesentlich betrachtet wird, die an sich rein zufällig ist. Mit dieser Grenze wird bei der Prüfung von Abweichungen ein strengerer Massstab angelegt, als mit der Grenze $P = 0{,}01$.

b) Die Prüfung von Durchschnitten

Die am häufigsten verwendete statistische Masszahl ist der Durchschnitt. Hat man den Durchschnitt $\bar{v}$ einer Stichprobe zu N Einzelwerten ermittelt, so kann man sich fragen, wieweit dieser Durchschnitt mit dem für die Grundgesamtheit zu erwartenden Durchschnitt μ übereinstimmt. Allein dieser letzte Wert ist in der Regel unbekannt. Trotzdem weiss man von ihm, dass er gleich ist dem Durchschnitt $\mu_{\bar{v}}$ aus allen Durchschnitten $\bar{v}$ der möglichen einzelnen Stichproben. Ist nun σ^2 die Streuung der Grundgesamtheit, so wird die Verteilung der Durchschnitte der Stichproben eine Streuung $\sigma_{\bar{v}}^2 = \frac{\sigma^2}{N}$ aufweisen, da jede Stichprobe N Einzelwerte umfasst. Die Gleichung für die Verteilung der Durchschnitte wird also lauten

$$f_{(\bar{v})} = \frac{1}{\sigma_{\bar{v}}\sqrt{2\pi}}\, e^{-\frac{(\bar{v}-\mu)^2}{2\sigma_{\bar{v}}^2}} \qquad (42)$$

wobei die Variable nun $\bar{v}$ ist.

Geht man von der Sicherheitsschwelle $P = 0{,}05$ aus, so muss nach Abschnitt 5, lit. e, der Durchschnitt

$\bar{v}$ innerhalb der Grenzen $\mu \pm 1{,}956\ \sigma_{\bar{v}}$ liegen, wenn er nur zufällig von μ abweicht. Im Grenzfall ist also

$$\bar{v} = \mu \pm 1{,}956\ \sigma_{\bar{v}}$$

$$\text{oder da } \sigma_{\bar{v}} = \frac{\sigma}{\sqrt{N}}$$

$$\bar{v} = \mu \pm 1{,}956\ \frac{\sigma}{\sqrt{N}}$$

Ist für die gewählte Stichprobe N gross genug, so kann man, ohne einen grossen Fehler zu begehen, für die Streuung σ^2 der Grundgesamtheit, die Streuung s^2 der Stichprobe in die Gleichung einsetzen, die dann lautet:

$$\bar{v} = \mu \pm 1{,}956\ \frac{s}{\sqrt{N}} \qquad (43)$$

Da von einem $P = 0{,}05$ ($W = 0{,}95$) ausgegangen wurde, besagt die Gleichung (43), dass in 95 % aller Stichproben ein Durchschnitt zu erwarten ist, der um höchstens $\pm\ 1{,}956\ \frac{s}{\sqrt{N}}$ von dem für die Grundgesamtheit theoretisch zu erwartenden Urdurchschnitt abweicht. Mit andern Worten, der errechnete Durchschnitt wird mit 95 % Wahrscheinlichkeit um nicht mehr als $\pm\ 1{,}956\ \frac{s}{\sqrt{N}}$ vom Urdurchschnitt abweichen. Wird diese Wahrscheinlichkeit auf 99 % erhöht ($P = 0{,}01$), so vergrössert sich die Grenze der möglichen Abweichung auf $\pm\ 2{,}576\ \frac{s}{\sqrt{N}}$.

Für das im ersten Aufsatz erwähnte Zahlenbeispiel aus dem Tarifwesen, mit $N = 225$ Einzelwerten, war der mittlere spezifische Energieverbrauch $\bar{q}$ zu 36,5 kWh pro Parametereinheit (Grundeinheit = GE) und die mittlere quadratische Abweichung s zu 13,03 ($s^2 = 169{,}99$) ermittelt worden.

Für $P = 0{,}05$ gilt in diesem Falle die Relation

$$\bar{q} = 36{,}5 = \mu \pm \frac{1{,}956 \cdot 13{,}03}{\sqrt{225}} = \mu \pm 1{,}7$$

Mit 95 %iger Wahrscheinlichkeit kann der errechnete Durchschnitt von 36,5 um höchstens $\pm$ 1,7, d. h. um höchstens $\pm$ 4,65 % vom theoretischen Urdurchschnitt abweichen.

Setzt man die höchstzulässige Abweichung zum vorneherein fest, so kann man ohne weiteres aus der ebenfalls als bekannt angenommenen Streuung den erforderlichen Umfang der Stichprobe ermitteln.

Dabei ist leicht festzustellen, dass, um eine zulässige Verringerung der Abweichungsgrenze auf die Hälfte zu erreichen, der Umfang der Stichprobe auf das Vierfache erhöht werden muss.

Ist der Durchschnitt der Grundgesamtheit bekannt, so kann ohne Schwierigkeit ermittelt werden, ob der errechnete Durchschnitt der Stichprobe von jenem wesentlich oder nur zufällig abweicht. Aus Formel (43) geht hervor, dass der Ausdruck $(v-\mu)\ \frac{\sqrt{N}}{s}$ höchstens $\pm$ 1,956 (bei $P = 0{,}05$) betragen darf. Ergibt die Ausrechnung einen höheren Wert, so gilt der Unterschied als gesichert und die Stichprobe entspricht mit 95 %iger Wahrscheinlichkeit nicht der Grundgesamtheit.

c) Die Prüfung von Streuungen

Ist σ die Streuung der Grundgesamtheit und s die Streuung von Stichproben grösseren Umfanges, so ist die Verteilung der Streuungen s normal, mit einer Streuung $\sigma_s^2 = \frac{\sigma^2}{2N}$.

Geht man wiederum von einem $P = 0{,}05$ aus, so kann man, wie für den Durchschnitt im vorangehenden Abschnitt, setzen:

$$s = \sigma \pm 1{,}956\ \sigma_s$$

$$\text{oder } s = \sigma \left(1 \pm \frac{1{,}956}{\sqrt{2N}}\right) \qquad (44)$$

Aus dieser Gleichung geht hervor, dass die aus der Stichprobe errechnete Streuung mit 95 %iger Wahrscheinlichkeit um höchstens $\pm\ \frac{1{,}956}{\sqrt{2N}}$ von der Streuung der Grundgesamtheit abweichen darf, wenn sie innerhalb der Zulässigkeitsgrenze bleiben soll. Mit andern Worten: Man nimmt in Kauf, dass in 5 % aller möglichen Fälle eine Streuung s zu erwarten ist, die bei

$N =$	100	um	13,8 %
$N =$	1 000	um	4,4 %
$N =$	5 000	um	2,0 %
$N =$	10 000	um	1,4 %

oder mehr von der Streuung σ der Grundgesamtheit abweicht. Für das erwähnte Zahlenbeispiel lautet die Relation (44):

$$s = 13{,}03 = \sigma \left(1 \pm \frac{1{,}956}{\sqrt{2 \cdot 225}}\right) = \sigma\ (1 \pm 0{,}092)$$

Mit 95 %iger Wahrscheinlichkeit wird für das gewählte Beispiel die Streuung s um nicht mehr als 9,2 % von der Urstreuung σ abweichen.

Nun kann der kleine Fehler korrigiert werden, der in der Gleichung (43) dadurch entstand, dass statt σ nur s gesetzt wurde. Die neue Gleichung sollte demnach korrekt lauten

$$\bar{v} = \mu \pm 1{,}956\ \frac{\sigma}{\sqrt{N}} \left(1 \pm \frac{1{,}956}{\sqrt{2N}}\right) \qquad (45)$$

Für das Zahlenbeispiel ergibt sich:

$$\bar{v} = \mu \pm 1{,}7 \cdot (1 + 0{,}092)$$
$$\bar{v} = \mu \pm 1{,}86.$$

Die mit 95 %iger Wahrscheinlichkeit zu erwartende höchste Abweichung des aus der Stichprobe

errechneten Durchschnittes vom Urdurchschnitt hat sich durch Berücksichtigung der Abweichungsmöglichkeit der Streuung von 1,7 auf 1,86, d. h. um rund 9 % erhöht. Auf den errechneten Durchschnitt von 36,5 bezogen bedeutet dies eine Erweiterung der Grenze von ± 4,65 auf ± 5,10 %, was tatsächlich unbedeutend ist und für eine erste Annäherung die einfachere Formel (43) durchaus rechtfertigt, besonders wenn der Umfang der Stichprobe (N) gross ist.

7. Eine graphische Lösung: Das Wahrscheinlichkeitsnetz

a) Die Entstehung des Wahrscheinlichkeitsnetzes

Die Transformation einer praktisch erhaltenen Häufigkeitskurve (Treppenkurve) auf den Massstab der normalen Häufigkeitskurve ist bereits erläutert worden. Diese, eine Beurteilung der Verteilung einer Statistik oder Stichprobe ermöglichende Operation ist jedoch umständlich, da sie eine grosse Rechenarbeit verlangt. Es wurde deshalb nach einer einfachen graphischen Methode gesucht, welche die Verteilung einer Statistik und die mit dieser Verteilung eng zusammenhängenden Summenhäufigkeiten zu kontrollieren gestattet. Diese Methode beruht auf der Anwendung des *Wahrscheinlichkeitsnetzes*, in welchem — normale Verteilung vorausgesetzt — die Summenhäufigkeitskurven als Geraden und die Häufigkeits- oder Fehlerkurve als eine Art Hyperbel erscheinen.

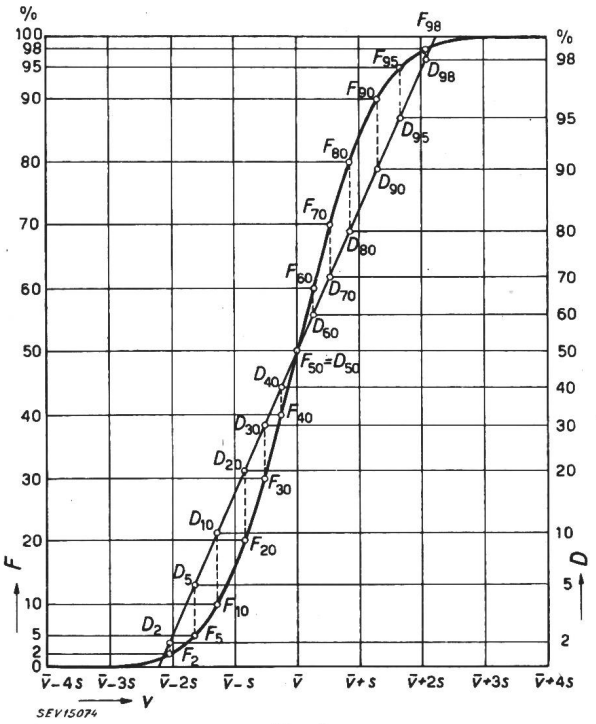

Fig. 8
Konstruktion des Wahrscheinlichkeitsnetzes
links: normale Teilung
rechts: Wahrscheinlichkeitsteilung

Nimmt man eine normale Summenhäufigkeitskurve (z. B. diejenige von Fig. 5) und zieht eine beliebige, schräge Gerade d durch die Mitte der Kurve (entsprechend der Abszisse $\bar{v}$), so ist jedem Punkt F der Kurve mit Abszisse v ein Punkt D der Geraden d mit der gleichen Abszisse zugeordnet. Die Punkte D ergeben eine neue Teilung der Ordinatenachse, die von der Mitte aus nach oben und nach unten immer gestreckter wird, etwa wie eine umgekehrte logarithmische Teilung. In Fig. 8 sind die entsprechenden Punkte F der Kurve (normale Teilung) und der Geraden D (Wahrscheinlichkeitsteilung) zur Verdeutlichung mit gleichlautenden Indices versehen. Zudem ist im linken Teil der Figur die alte und im rechten Teil die neue Teilung eingezeichnet. Das auf diese Weise erhaltene Netz, das sich theoretisch nach oben und nach unten bis ins Unendliche erstreckt, praktisch aber nur etwa zwischen 0,1 und 99,9 verwendet wird, nennt man das Wahrscheinlichkeitsnetz. Aus praktischen Gründen werden die Einzelwerte von F in Prozenten der Endsumme angegeben.

In diesem Netz erscheint also die F-Kurve (einfache Summenhäufigkeit) als eine Gerade (bei normaler Verteilung), deren Neigung vom Massstab der Variabeln v und von der Streuung s^2 abhängt. Bei gleichbleibendem v-Maßstab wird die Gerade um so steiler, je kleiner die Streuung s^2 ist.

b) Die Häufigkeitskurve

Zeichnet man die normale Häufigkeitskurve in die untere Hälfte des Wahrscheinlichkeitsnetzes ein, so erhält man nicht mehr die bekannte, glockenförmige, sondern eine hyperbelartige Kurve, deren

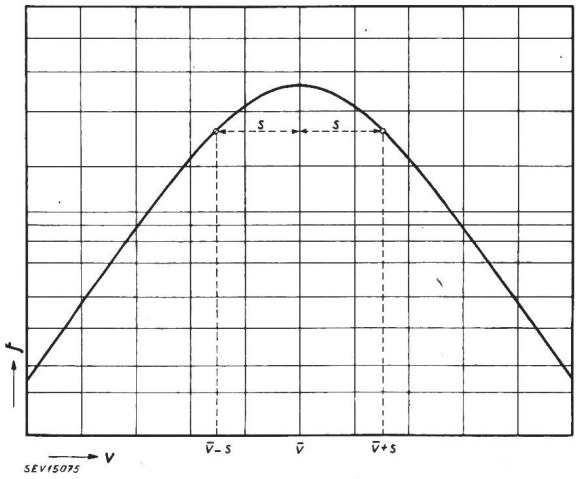

Fig. 9
Darstellung der Fehlerfunktion im Wahrscheinlichkeitsnetz (einfache Häufigkeit)

äussere Äste infolge der Dehnung der Teilung angenähert zu Geraden gestreckt sind (Fig. 9). Die neue Kurve ist leichter zu zeichnen und gibt, vor allem für die extremen Werte der Variabeln, eine bessere Uebersicht der Häufigkeitsverteilung.

c) Die P- und W-Kurven

Bei zweckmässiger Anpassung des Maßstabes der Ordinatenteilung [W % $= 2\,(F - 50)$ %] und Verwendung der oberen Netzhälfte wird auch die W-Kurve (Summenhäufigkeit der absoluten Fehler) zu einer durch den Ursprung gehenden geneigten Geraden (Fig. 10).

Auch hier ist ohne weiteres ersichtlich, dass für einen bestimmten Maßstab der Variabeln die W-Gerade um so steiler wird, je kleiner die Streuung ist.

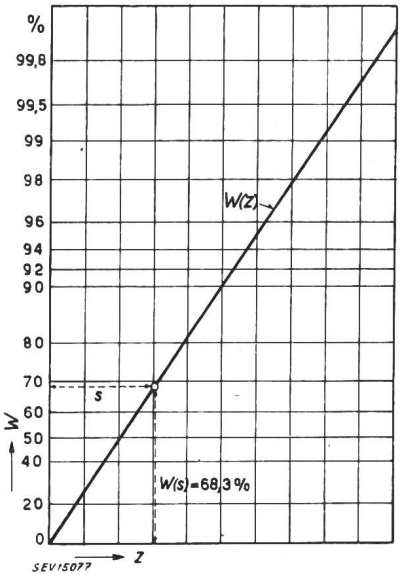

Fig. 10
Summenhäufigkeit der absoluten Fehler

In der untern Hälfte des Wahrscheinlichkeitsnetzes wird schliesslich nach Anpassung des Masstabes (P % = 2 F %) auch die P-Kurve (Summenhäufigkeit der einen bestimmten Wert übersteigenden absoluten Fehler) zu einer Geraden (Fig. 11).

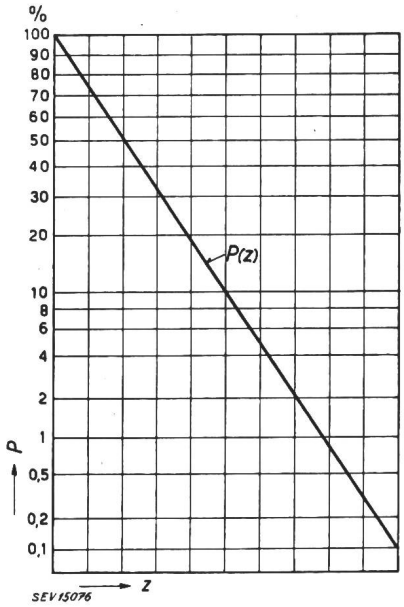

Fig. 11
Summenhäufigkeit der einen bestimmten Wert z übersteigenden absoluten Fehler

d) Die Wahl des Maßstabes für die Variable im Wahrscheinlichkeitsnetz

In allen diesen Fällen ist die Wahl des Maßstabes für die Variable (v bei den f- und F-Kurven, und z bei den W- und P-Kurven) theoretisch nur durch die Breite des verwendeten Wahrscheinlichkeitsnetzes beschränkt. Sollen jedoch mehrere Statistiken einander gegenübergestellt werden, so ist es zweckmässig, einen gemeinsamen Maßstab zu nehmen. Eine für überschlägige Untersuchungen geeignete Methode besteht darin, die einzelnen Werte, bzw. Klassenmitten und Klassengrenzen der Variabeln, in Prozenten des Durchschnittes auszudrücken. So erhält man leicht vergleichbare Kurven, aus deren Verlauf und vor allem aus deren Neigung qualitative Schlüsse über die Streuung gezogen werden können. Sind die Streuungen der einzelnen Statistiken bekannt (was aber selten der Fall ist), so kann für jede Statistik die Variable als ein Vielfaches der mittleren quadratischen Abweichung s ausgedrückt werden; die einzelnen Kurven müssen dann untereinander und mit einer zum voraus leicht zu zeichnenden Standard-Kurve identisch sein, wenn bei allen Statistiken die Verteilung normal ist, d. h. wenn diese Statistiken den Gesetzen der Wahrscheinlichkeitsrechnung genügen. Eine andere praktisch bewährte Vergleichs- und Untersuchungsmethode soll weiter unten näher erläutert werden.

8. Praktische Anwendung des Wahrscheinlichkeitsnetzes

a) Allgemeines

Im ersten Teil dieser Studie ist bereits auf die wesentliche Vereinfachung hingewiesen worden, welche die Gruppierung der N Einzelwerte einer Statistik oder Stichprobe in M Klassen von der Breite w mit Klassenmitte v_j und Klassenhäufigkeit n_j für die Berechnung der statistischen Masszahlen mit sich bringt. Für die praktische Prüfung der Verteilung und für die Ermittlung der Summenhäufigkeiten ist diese Gruppierung geradezu unentbehrlich.

Es ist bereits auf den Vorteil hingewiesen worden, die Klassenbreite mit der mittleren quadratischen Abweichung s in Beziehung zu bringen, um Vergleiche zu erleichtern. Da aber die Berechnung der Streuung s^2 umständlich ist, zieht man in der Praxis einen Wert vor, der in einem festen Verhältnis zu ihr steht, aber leichter zu ermitteln ist: die Grundspanne.

Ist g_j die obere Grenze der Klasse j mit Mitte v_j, so entspricht dieser Grenze g_j die einfache Summenhäufigkeit F_j, in Analogie zum F-Integral der Fehlerfunktion. Ist weiter n_j die Häufigkeit der Klasse j, so wird

$$F_j = \sum_{j=1}^{j} n_j$$

Im Grenzfall, d. i. für $j = M$, erhält man

$$F_M = \sum_{j=1}^{M} n_j = N$$

Zweckmässig wird F_j, wie auch n_j in Prozenten von N ausgedrückt.

Als Spanne G bezeichnet man die absolute Differenz zwischen zwei in bezug auf $\bar{v}$ symmetrischen Grenzwerten g, deren entsprechende F-Werte (in Prozenten von N ausgedrückt) zusammen 100 ergeben. Die Spanne zwischen g_5, entsprechend F = 5 % und g_{95}, entsprechend F = 95 % heisst die

Grundspanne G_{90}. Sie umfasst 90 % aller Einzelwerte der Statistik und entspricht bei normaler Verteilung dem W-Integral für den absoluten Fehler $z = |\bar{v} - g_5| = |\bar{v} - g_{95}|$ [2]).

Der Klassengrenzwert g_{50} ist identisch mit dem Medianwert m und bei normaler Verteilung auch mit dem Durchschnitt $\bar{v}$. Diesem Grenzwert entspricht theoretisch die Spanne G_0.

Die Spanne G_{50} gibt bei normaler Verteilung die obere und untere Grenze des wahrscheinlichen Fehlers an, d. h. des Fehlers, für welchen die Summenhäufigkeit W gleich 0,5 oder 50 % ist, während die mittlere quadratische Abweichung s, bzw. die Grenzen $\bar{v} \pm s$ der Spanne G_{68} entsprechen.

chungen vom Mittelwert als zufällig betrachtet, während ausserhalb dieser Spanne fallende Abweichungen als gesichert gelten. Es wäre somit naheliegend, die Klassenbreite w mit dieser Spanne G_{95} in eine einfache Beziehung zu bringen. Die Erfahrung hat aber gezeigt, dass es zweckmässiger ist, von der zuverlässiger ermittelbaren Grundspanne G_{90} auszugehen, und diese je nach Feinheit der vorzunehmenden Untersuchung in 5, 10 oder 20 gleich grosse Teile zu teilen, so dass sich zum Beispiel bei der 10er Teilung eine Klassenbreite

$$w = \frac{G_{90}}{10} = \frac{g_{95} - g_5}{10}$$

Tabelle IX

Klassenmitte v_j			obere Klassengrenze g_j	Häufigkeit n_j pro Klasse in % von N			Summenhäufigkeit F_j in % von N		
5er Teilung	10er Teilung	20er Teilung		5er Teilung	10er Teilung	20er Teilung	5er Teilung	10er Teilung	20er Teilung
10	10	10	10	0,15	0,15	0,15	0,15	0,15	0,15
		12,5	15			0,10			0,25
	15	17,5	20		0,27	0,17		0,42	0,42
		22,5	25			0,26			0,68
20	25	27,5	30	0,91	0,64	0,38	1,06	1,06	1,06
		32,5	35			0,56			1,62
	35	37,5	40		1,35	0,79		2,41	2,41
		42,5	45			1,11			3,52
40	45	48,5	50	3,94	2,59	1,47	5,00	5,00	5,00
		52,5	55			1,93			6,93
	55	57,5	60		4,41	2,48		9,41	9,41
		62,5	65			3,07			12,48
60	65	67,5	70	11,18	6,77	3,70	16,18	16,18	16,18
		72,5	75			4,37			20,55
	75	77,5	80		9,36	4,99		25,54	25,54
		82,5	85			5,53			31,07
80	85	87,5	90	20,92	11,56	6,03	37,10	37,10	37,10
		92,5	95			6,36			43,46
	95	97,5	100		12,90	6,54		50,00	50,00
		102,5	105			6,54			56,54
100	105	107,5	110	25,80	12,90	6,36	62,90	62,90	62,90
		112,5	115			6,03			68,93
	115	117,5	120		11,56	5,53		74,46	74,46
		122,5	125			4,99			79,45
120	125	127,5	130	20,92	9,36	4,37	83,82	83,82	83,82
		132,5	135			3,70			87,52
	135	137,5	140		6,77	3,07		90,59	90,59
		142,5	145			2,48			93,07
140	145	147,5	150	11,18	4,41	1,93	95,00	95,00	95,00
		152,5	155			1,48			96,48
	155	157,5	160		2,59	1,11		97,59	97,59
		162,5	165			0,79			98,38
160	165	167,5	170	3,94	1,35	0,56	98,94	98,94	98,94
		172,5	175			0,38			99,32
	175	177,5	180		0,64	0,26		99,58	99,58
		182,5	185			0,17			99,75
180	185	187,5	190	0,91	0,27	0,10	99,85	99,85	99,85
190	190	190	190	0,15	0,15	0,15	100,00	100,00	100,00

b) Wahl der Klassenbreite

Bei der Prüfung von statistischen Masszahlen werden alle innerhalb der Spanne G_{95} (entsprechend $W = 0{,}95$ oder $P = 0{,}05$) liegenden Abwei-

ergibt. Diese Klassenbreite gilt selbstverständlich auch für die Bereiche ausserhalb der Grundspanne.

c) Die Häufigkeitsteilung

In Tabelle IX sind die bei normaler Verteilung zu erwartenden Häufigkeiten und Summenhäufigkeiten für eine Statistik mit Durchschnitt $\bar{v} = 100$ und Streuung $s^2 = 923$ ($s = 30{,}4$) bei 5er, 10er und 20er Teilung der Grundspanne G_{90} zusammengestellt.

[2]) Der Grund, warum für die Spanne das Symbol G und nicht W gewählt wurde, liegt darin, dass W seinem Wesen nach das Integral der Fehlerhäufigkeiten zwischen zwei in bezug auf $\bar{v}$ symmetrischen Abszissenwerten ist, während die Spanne G die Strecke der Abszissenachse zwischen diesen zwei Werten misst.

In der Praxis werden sich je nach der Streuung der betrachteten Statistik andere Klassenbreiten w, Klassenmitten v_j und Klassengrenzen g_j ergeben. Die relativen Werte der Häufigkeit und der Summenhäufigkeit (in Prozenten des Statistikumfanges N) werden aber bei normaler Verteilung immer den Tabellenwerten entsprechen.

Hieraus ergibt sich nebenbei eine einfache Methode zur angenäherten Berechnung der Streuung. Für die Statistik von Tabelle IX liegen die Grenzen g_5 und g_{95} bei 50 und 150. Die Grundspanne G_{90} beträgt 100 und die entsprechende absolute Abweichung 50. Nach Abschnitt 1e) müssen bei normaler Verteilung 90 % aller Werte innerhalb der Grenze $\bar{v} \pm 1{,}645\ s$ liegen. Kann man annehmen, dass die Verteilung der zu untersuchenden Statistik normal ist, so entspricht die absolute Abweichung 50 der Grenze $\pm$ 1,645 s. Somit erhält man

$$s = \frac{50}{1{,}645} = 30{,}4 \text{ und } s^2 = 923$$

Auch wenn die Verteilung nur angenähert normal ist, was praktisch meistens der Fall ist, liefert diese einfache Rechnung für die Streuung einen für überschlägige Betrachtungen hinreichend genauen Wert.

Bei der graphischen Darstellung pflegt man die Klassenhäufigkeit n_j der Klassenmitte v_j und die Summenhäufigkeit F_j der oberen Klassengrenze g_j zuzuordnen (Fig. 12).

d) Numerisches Beispiel

Zur Illustrierung der geschilderten graphischen Methode sei das frühere Beispiel herangezogen (Tabelle IV). Die 225 Werte umfassende Statistik ($N = 225$) besass nach der «klassischen» Berechnungsweise folgende Charakteristiken:

Umfang $N = 225$

Durchschnitt $\bar{q} = 36{,}5$

Streuung $s^2 = 169{,}99$

mittlere quadratische Abweichung $s = 13{,}03$.

Das Ordnen der Statistik nach steigendem Wert von q zeigt, dass 5 % aller Werte die Grenze

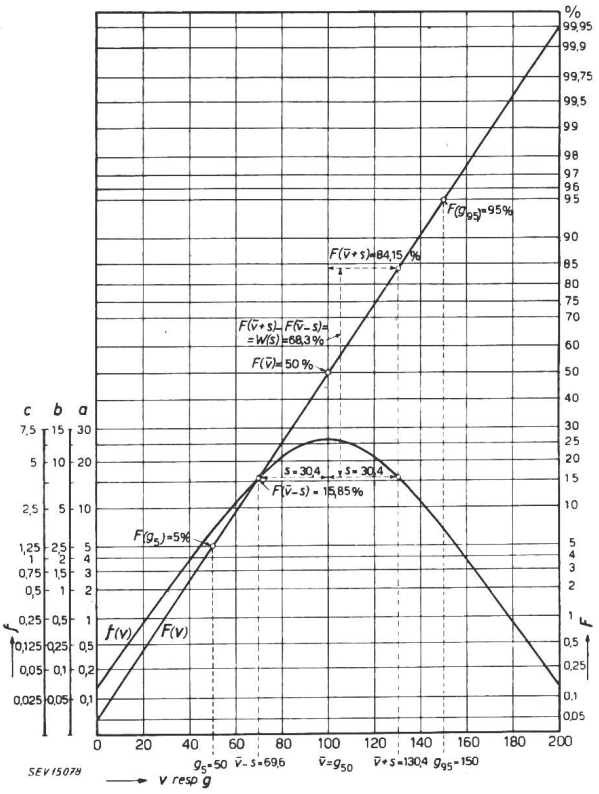

Fig. 12
Darstellung der Normalverteilung nach Tabelle IX

Tabelle X

Teilungsklassen	Obere Klassengrenze g_j		Klassenmitte q_j		Klassenhäufigkeit n_j		Summenhäufigkeit F_j		Normalverteilung	
	absolut	% von $\bar{q}$	absolut	% von $\bar{q}$	absolut	% von N	absolut	% von N	n_j % von N	F_j % von N
	0	0							0,2	0,2
			3,1	5,8	0	0			0,9	
äussere	6,4	17,5					0	0		1,1
			10,7	29,3	11	4,9			3,9	
g_5	15,0	41,1					11	4,9		5,0
			19,3	52,9	19	8,4			11,2	
	23,6	64,7					30	13,3		16,2
			27,9	76,4	59	26,3			20,9	
	32,2	88,2					89	39,6		37,1
			36,5	100,0	66	29,3			25,8	
innere	40,8	111,8					155	68,9		62,9
			45,1	123,6	34	15,1			20,9	
	49,4	135,3					189	84,0		83,8
			53,7	147,1	25	11,1			11,2	
g_{95}	58,0	158,9					214	95,1		95,0
			62,3	170,7	7	3,1			3,9	
	66,6	182,5					221	98,2		98,9
			70,9	194,2	2	1,0			0,9	
	75,2	206,0					223	99,2		99,8
äussere			79,5	217,8	1	0,4			0,2	
	83,8	229,6					224	99,6		100,0
			88,1	241,4	1	0,4				
	92,4	253,1					225	100,0		

$g_5 = 15$ und 95 % die Grenze $g_{95} = 58$ nicht übersteigen. Die Grundspanne errechnet sich somit zu

$$G_{90} = 58 - 15 = 43$$

Für eine überschlägige Untersuchung genügt eine 5er Teilung der Grundspanne. Somit wird die Klassenbreite

$$w = \frac{G_{90}}{5} = \frac{43}{5} = 8{,}6$$

Die sich daraus ergebenden oberen Klassengrenzen g_j und Klassenmitten q_j, sowie die aus der Auszählung resultierenden Klassenhäufigkeiten n_j und Summenhäufigkeiten F_j sind in Tabelle X zusammengestellt. Um die graphische Darstellung zu erleichtern, figurieren daneben die relativen Werte, bei Klassenmitten und Klassengrenzen in Prozenten des Durchschnittes, bei den Häufigkeiten in Prozenten des Umfanges der Stichprobe. Zur Kontrolle sind die bei normaler Verteilung zu erwartenden Häufigkeitswerte in Prozenten von N angeführt.

Im Wahrscheinlichkeitsnetz von Fig. 13 sind die Werte der Tabelle X graphisch aufgetragen.

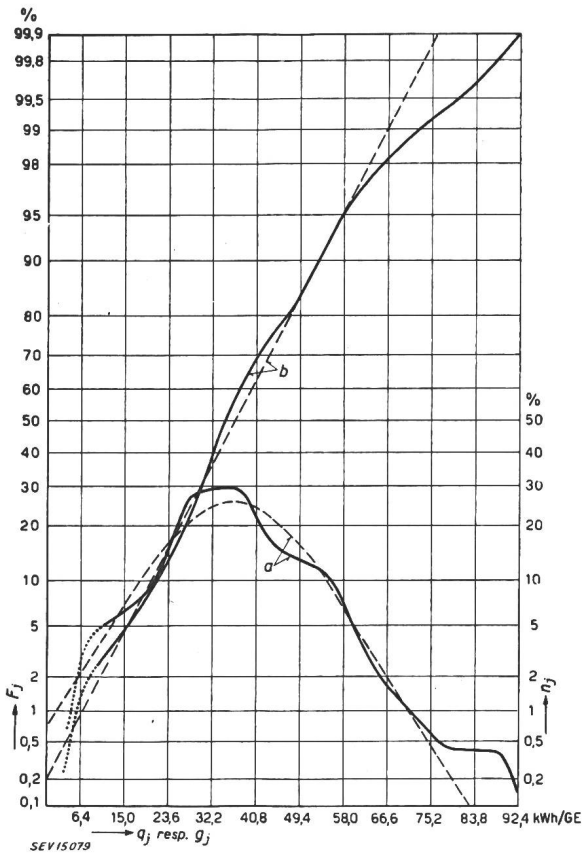

Fig. 13
Darstellung der Mischverteilung nach Tabelle X
——— Mischverteilung
– – – – Normalverteilung gleicher Grundspanne (bzw. Streuung).
a einfache Häufigkeit *b* Summenhäufigkeit

Im Bereiche der Grundspanne G_{90}, d. h. zwischen den Grenzen g_5 und g_{95} weichen die tatsächlichen Kurvenwerte (volle Kurve) nicht sehr stark von den für eine normale Verteilung zu erwartenden Werten ab (gestrichelte Kurve). Wenn der Umfang der Stichprobe wächst, tritt auch in den äusseren Teilungsklassen eine weitgehende Angleichung an die Normalwerte ein.

Aus der Kenntnis der Grundspanne G_{90} lässt sich die mittlere quadratische Abweichung s kontrollieren. Die Grundspanne G_{90} entspricht dem Bereich $\bar{q} \pm 1{,}645\ s$, innerhalb welchem bei normaler Verteilung 90 % aller Werte liegen müssen. Somit ist

$$g_5 = \bar{q} - 1{,}645\ s$$
$$g_{95} = \bar{q} + 1{,}645\ s$$

oder in Zahlen

$$g_5 = 15{,}0 = 36{,}5 - 1{,}645\ s$$
$$g_{95} = 58{,}0 = 36{,}5 + 1{,}645\ s$$

woraus sich ergibt

$$s = 13{,}08$$

und $$s^2 = 171$$

Gegenüber den «klassisch» errechneten Werten beträgt der Unterschied 0,05 oder 0,4 % für s und rund 1,0 oder 0,6 % für s^2. Diese Unterschiede sind also vernachlässigbar klein.

Aber auch rein graphisch kann die Streuung direkt ermittelt werden. Bei normaler Verteilung müssen 68,3 % aller Werte innerhalb der Grenzen $\bar{q} \pm s$ liegen. Auf der Summenhäufigkeitskurve (bzw. -geraden im Wahrscheinlichkeitsnetz) entspricht also $\bar{q} + s$ einem $F_{\bar{q}+s} = 84{,}15$ % und $\bar{q} - s$ einem $F_{\bar{q}-s} = 15{,}85$ %. Die Spanne zwischen den Abszissen dieser beiden F-Punkte ist somit gleich 2 s. In Fig. 13 ist diese Spanne rund 70 % von $\bar{q}$ oder 25,6, womit $s = 12{,}8$ wird. Bei grösserem Statistikumfang würde auch dieser rein graphisch ermittelte Wert sich dem theoretischen Wert besser nähern.

Der Vorteil der geschilderten graphischen Methode liegt in ihrer Einfachheit, aber auch darin, dass es genügt, die leicht zu erhaltenden Häufigkeits- und Summenhäufigkeitskurven im Wahrscheinlichkeitsnetz zu zeichnen, um auf den ersten Blick zu erkennen, ob die Verteilung normal oder annähernd normal ist, oder ob sie Unregelmässigkeiten aufweist, die auf eine Mischverteilung hindeuten. Die Zerlegung solcher Mischverteilungen in eine Anzahl normaler Verteilungen ist erst durch die Anwendung des Wahrscheinlichkeitsnetzes auf eine einfache und elegante Art möglich gemacht worden. Diese Zerlegung erlaubt, wertvolle Schlüsse auf die Auswirkung von Faktoren, die bei der Untersuchung vorerst nicht berücksichtigt wurden, zu ziehen.

9. Analyse von Häufigkeitsverteilungen mit Hilfe des Wahrscheinlichkeitsnetzes

a) Allgemeines

Das Beispiel von Fig. 13 zeigt, dass, zumal beim Energieverbrauch, wo stets mehr als ein Faktor im

Spiele steht, die die Verteilung der spezifischen Verbrauchswerte wiedergebende Häufigkeitskurve stets mehr oder weniger von der normalen Verteilung nach der Gaußschen Formel abweicht. Die Erfahrung zeigt ferner, dass die «Anomalien» der Kurven in der Regel auch bei starker Vermehrung des Statistikumfanges weiter bestehen bleiben. Wohl verschwinden mit zunehmendem Umfang infolge zuverlässigerer Mittelwertbildung die kleinen Unregelmässigkeiten, vor allem bei den untersten und obersten, meist schwach besetzten Klassen, aber die grossen, typischen Abweichungen vom normalen Kurvenverlauf behaupten sich.

Diese Erscheinung darf keineswegs als ein Versagen der Gaußschen Fehlertheorie gedeutet werden. Läge z. B. für den Beleuchtungsenergieverbrauch nur ein beeinflussender Faktor, z. B. die Raumzahl, vor, so müsste die Häufigkeitsverteilung unbedingt normal ausfallen. Sobald sich aber weitere Faktoren hinzugesellen, z. B. die Raumgrösse, die Ausdehnung der Installation, die Lebensgewohnheiten der Abnehmer usw., so müssen diese eine Veränderung der Verteilung hervorrufen. Diese Einflüsse weitgehend herauszuschälen und zu deuten gestattet die nachfolgend beschriebene analytische Methode.

Wie jede Schwingungskurve nach der Fourierschen Analyse als eine Summe reiner Sinuskurven verschiedener Phasen, Frequenzen und Amplituden betrachtet werden kann, so lässt sich auch jede beliebige Häufigkeitskurve (als Bild eines beliebigen Urkollektivs) in einzelne normale Häufigkeitskurven (als Bilder normaler Teilkollektive) mit verschiedenen Durchschnitten, Streuungen (bzw. Grundspannen) und Scheitelwerten (bzw. Umfängen) zerlegen. Die Deutung der charakteristischen Grössen der Teilkollektive erlaubt dann, auf die Eigenschaften des Urkollektivs Rückschlüsse zu ziehen.

Nebst statistisch-mathematischem, also rein handwerksmässigem Geschick verlangt diese graphisch-rechnerische Analyse, und noch viel mehr die Deutung der Ergebnisse, eine gründliche Beherrschung der untersuchten Materie und viel Verständnis für ihre inneren Zusammenhänge.

b) Das Werkzeug

Von wesentlicher Bedeutung für die Kennzeichnung einer normalen Häufigkeitskurve sind die Grundspanne G_{90}, die ein praktisches Mass für die Streuung bildet, und der Umfang des durch die Kurve dargestellten Kollektivs.

Zu den beiden Endpunkten der Grundspanne G_{90} einer normalen Häufigkeitskurve, als Abszissen genommen, gehören zwei gleich grosse Ordinaten, die in einem bestimmten Verhältnis zu dem dem Durchschnitt $\bar{q}$ zugeordneten Scheitelwert n_m der Kurve stehen.

Nach Tabelle VI hat der Scheitelpunkt der Kurve die Ordinate

$$f_m = \frac{1}{s\sqrt{2\pi}}$$

Die Endpunkte g_5 und g_{95} der Grundspanne G_{90} entsprechen den Abszissenwerten $\bar{q} - 1{,}645\ s$ und $\bar{q} + 1{,}645\ s$. Setzt man diese Werte in Formel (26) ein, so erhält man

$$f_{g5} = f_{g95} = \frac{1}{s\sqrt{2\pi}} \cdot e^{-\frac{1{,}645^2}{2}}$$

und das gesuchte Verhältnis ergibt sich zu

$$\gamma_{90} = \frac{f_{g5}}{f_m} = e^{-\frac{1{,}645^2}{2}} = 0{,}258 \text{ oder rund } 0{,}26 \qquad (46)$$

Das Wertvolle an dieser Beziehung ist, dass der Faktor γ_{90} konstant und von der Streuung unabhängig ist.

Ebenso kann nachgewiesen werden, dass bei normaler Verteilung der Umfang N in einer bestimmten Beziehung zum Scheitelwert f_m steht, die aber von der Streuung abhängt. Die Beziehung lautet

$$N = f_m\ s\sqrt{2\pi}$$

Für praktische Berechnungen, bei denen die Variabeln in Klassen von der Breite w eingeteilt sind, wobei die grösste Klassenhäufigkeit (Scheitelwert) mit n_m bezeichnet wird, schreibt sich die Formel

$$N = \frac{n_m}{w}\ s\sqrt{2\pi}$$

Da aber $G_{90} = 2 \cdot 1{,}645\ s$ und $\sqrt{2\pi} =$ rund 2,5 ist, so erhält man

$$N = 0{,}76\ G_{90}\ \frac{n_m}{w} \qquad (47)$$

Für die rasche Skizzierung einer normalen Häufigkeitskurve im Wahrscheinlichkeitsnetz ist es noch nützlich zu wissen, dass ausserhalb der Grundspanne die beiden Kurvenäste praktisch gradlinig verlaufen.

c) Die Darstellung

Bei der graphischen Darstellung ist es hier nicht nötig, von der Grundspanne des Urkollektivs auszugehen, und diese in 5, 10 oder 20 Teile zu teilen, um die *Klassenbreite* zu erhalten. Es ist sogar vorteilhafter, hiefür eine runde ganze Zahl zu wählen, 1, 2, 5 oder 10, je nach Grösse der Variabeln.

Wichtig ist nur, die Breite der Klassen nicht zu klein zu wählen, da sonst auf jede Klasse, besonders auf die äussersten, zu wenig Einzelwerte entfallen, so dass die Kurve unter Umständen zackig wird. Anderseits darf die Unterteilung nicht zu grob sein, sonst verschwinden die charakteristischen Unregelmässigkeiten der Kurve. Mit einiger Uebung ist es nicht schwer, die Klassenbreite dem gewünschten Feinheitsgrad der Untersuchung anzupassen.

Für die rasche und genaue Aufzeichnung der Teilkollektive mit normaler Verteilung kann die 5er oder 10er Teilung der Grundspanne gute Dienste leisten, da die einzelnen Punkte aus der Tabelle IX herausgelesen werden können. Dabei muss aber der abgelesene Tabellenwert entspre-

chend dem Verhältnis der Klassenbreite des darzustellenden Teilkollektivs zur Klassenbreite des vorhandenen Urkollektivs und vom Umfang des Teilkollektivs zum Umfang des Urkollektivs, sowie dem für die Häufigkeit gewählten Maßstabe, umgerechnet werden.

Da für die Darstellung der Kurven die untere Hälfte des Wahrscheinlichkeitsnetzes (0...50 %) verwendet wird, wählt man als Maßstab *für die Klassenhäufigkeit*, die zweckmässig in absoluten Zahlen ausgedrückt wird, ein ganzes Vielfaches des vorhandenen Prozentmaßstabes, dermassen, dass der Kurvenscheitel nicht über die 50 %-Linie hinausragt, aber möglichst nahe an diese herankommt (bei $n_m = 120$ setzt man z. B. 125 = 50 % und bei $n_m = 160$, 200 oder 250 = 50 %).

In den meisten Fällen ist es angezeigt, die Häufigkeitskurve des Urkollektivs *auszugleichen*, vor allem wenn sie wegen zu kleinen Umfangs etwas zackig ausfällt. Sehr oft genügt eine Verschiebung der Klassengrenzen um eine halbe Klassenbreite, um nach erneuter Auszählung eine weitere Punktreihe zu erhalten, die es erlaubt, eine stetige Kurve zu zeichnen. Genügt dies nicht, so können noch die Klassen je zu zweien zusammengenommen werden, was im entsprechenden Maßstab eine weitere Punktreihe ergibt. Mitunter kann es auch nötig sein, die Summenhäufigkeitskurve zur Ausgleichung heranzuziehen, indem man die auffällig daneben geratenen Punkte berichtigt und hierauf die einfache Häufigkeitskurve zurückkonstruiert.

d) Das Vorgehen bei der Analyse

In der Regel verrät das Aussehen der die Mischverteilung darstellenden, ausgeglichenen Urkurve, wo das wichtigste, oder vielleicht auch schon, wo die anderen Teilkollektive liegen können. Anhaltspunkte dafür liefern die verschiedenen Buckel und der Verlauf der äussersten Kurvenäste.

Ist nur *ein* Teilkollektiv ohne weiteres erkennbar, so wird dieses zunächst von Hand einskizziert, mit Scheitel und ungefährer Oeffnung. Die unter b) abgeleiteten Beziehungen erlauben, die Grundspanne und dann den Umfang zu bestimmen, worauf die Kurve genau gezeichnet werden kann, sei es durch Heranziehung der 5er oder 10er Teilung der Grundspanne, oder über den Umweg der Summenhäufigkeitsgeraden. Mit der durch Subtraktion erhaltenen Restkurve wird gleich verfahren, bis alle Teilkollektive ermittelt sind. Wenn nötig, werden an den einzelnen Teilkollektiven noch Korrekturen angebracht, bis die Summenkurve mit der Urkurve praktisch übereinstimmt.

Wenn sich die Teilkollektive alle sofort erkennen lassen, so werden sie miteinander einskizziert. Aus Scheitelwert und Grundspanne wird der Umfang jedes Teilkollektivs ermittelt; die Summe muss den Umfang des Urkollektivs ergeben. Diese erste Kontrolle vor dem genauen Zeichnen der Teilkurven erspart in der Regel spätere Korrekturen.

e) Zahlenbeispiel

Als Beispiel diene wiederum die schon mehrfach verwendete Statistik (Tabelle IV) mit folgenden Charakteristiken:

Umfang N = 225
Durchschnitt $\bar{q}$ = 36,5
Streuung s^2 = 169,99.

Wenn die Klassenbreite w zu 20 % des gleich 100 % gesetzten Durchschnittes $\bar{q}$ gewählt wird, so ergibt sich folgende Urverteilung:

Tabelle XI

Obere Klassengrenze g_i	30	50	70	90	110	130	150	170	190	210	230	250
Klassenmitte q_i	20	40	60	80	100	120	140	160	180	200	220	240
Klassenhäufigkeit n_i	6	12	20	53	57	30	27	13	4	1	1	1

Nach dem Ausgleichen durch Verschieben der Klassengrenzen und durch Zusammenlegen je zweier Nachbarklassen ergibt sich die in Fig. 14 gestrichelt gezeichnete Kurve. Entsprechend dem grössten Wert von $n_i = 57$ wurde für den Häufigkeitsmaßstab 50 % = 100 gesetzt.

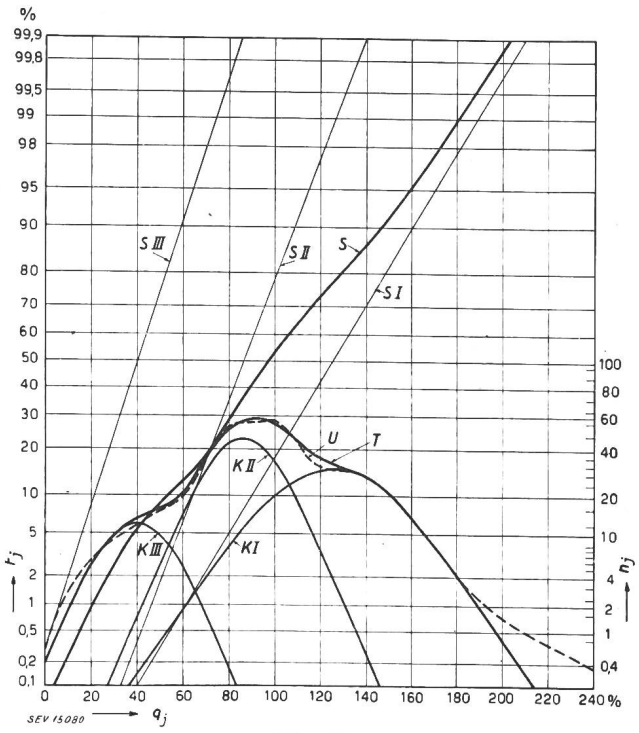

Fig. 14
Zerlegung einer Mischverteilung *U* in drei Normalverteilungen *I*, *II* und *III*
K einfache Häufigkeit *S* Summenhäufigkeit

Die gestrichelte Häufigkeitskurve des spezifischen Energieverbrauches pro Raum ist unregelmässig. Es liegt also eine Mischverteilung vor. Der Kurvenverlauf lässt vermuten, dass es sich hier um mindestens drei Teilkollektive handelt. Die Auswertung nach dem geschilderten Verfahren bestätigt in erster Annäherung diese Annahme und führt zu folgendem Ergebnis:

	Kurve *I*	Kurve *II*	Kurve *III*
Scheitelwert n_m	30	47	12
Abszisse q_m des Scheitelwertes (Durchschnitt)	125	86	40
Grundspanne G_{90}	90	56	48
Umfang N	103	100	22

Die drei Teilkurven *K I*, *K II* und *K III* sind in Fig. 14 voll ausgezogen, ebenso die zugehörigen Summenhäufigkeitsgeraden *S I*, *S II* und *S III*. Der Umfang des durch Summierung der drei Teilkurven erhaltenen neuen Kollektivs *T* (Fig. 14) errechnet sich zu 225. Aus der zugehörigen Summenhäufigkeitskurve *S* lässt sich die Grundspanne des neuen Summenkollektivs *T* zu G_{90} = 118 % oder 43,1 kWh/GE ermitteln.

Das Ergebnis der rechnerischen Kontrolle zeigt Tabelle XII.

Tabelle XII

Obere Klassengrenze g_j	30	50	70	90	110	130	150	170	190	210	230	250
Klassenmitte q_j	20	40	60	80	100	120	140	160	180	200	220	240
Klassenhäufigkeiten n_j (absolut) ausgeglichene Urkurve *U*	5,8	11,8	19,8	53,0	57,0	30,8	26,8	13,4	4,2	1,4	0,7	0,3
Teilkurve *K I*	—	0,3	2,0	8,0	19,9	28,8	25,3	13,5	4,3	0,8	0,1	—
Teilkurve *K II*	—	1,8	15,5	41,7	32,8	7,7	0,5	—	—	—	—	
Teilkurve *K III*	5,5	11,0	5,1	0,4	—	—	—	—	—	—	—	
Summenkurve *T*	5,5	13,1	22,6	50,1	52,7	36,5	25,8	13,5	4,3	0,8	0,1	

Aus der Grundspanne G_{90} = 43,1 des Summenkollektivs *T* errechnet sich die mittlere quadratische Abweichung zu

$$s = \frac{43{,}1}{2 \cdot 1{,}645} = 13{,}08$$

was mit dem früher ermittelten Werte (13,03) sehr gut übereinstimmt.

Die Deutung des Ergebnisses ist hier nicht schwer. Das Urkollektiv kann als die Summe von drei Teilkollektiven normaler Verteilung angesehen werden, die sich um die spezifischen Verbrauchswerte 40 % = 14,6 kWh/GE, 86 % = 31,4 kWh/GE und 125 % = 45,7 kWh/GE scharen (1 GE = 1 Hauptraum). Das Kollektiv mit dem kleinsten spezifischen Verbrauch (*K III*) ist schwach besetzt, die beiden anderen ungefähr gleich stark. Die Kollektive *K III* und *K II* weisen eine ziemlich starke Konzentration auf, während das Kollektiv *K I* wesentlich stärker streut.

Da es sich bei dieser Untersuchung um den am Licht-Doppeltarifzähler gemessenen Energieverbrauch im Haushalt handelt, unter Ausschluss von Küche und Warmwasserbereitung, können die drei ermittelten Teilkollektive nur Abonnentengruppen mit verschiedenen Verbrauchseigenschaften darstellen. Das Kollektiv *K III*, mit dem kleinsten Umfang (rund 10 % des Gesamtumfanges) dürfte diejenigen Abonnenten umfassen, die neben dem eigentlichen Beleuchtungsverbrauch keinen nennenswerten oder gar keinen Konsum für Kleinapparate aufweisen. Dies geht auch aus dem relativ schwachen spezifischen Energieverbrauch dieser Gruppe pro Hauptraum hervor. Das mittlere Kollektiv *K II*, das rund 44 % aller Abonnenten umfasst, stellt mit grosser Wahrscheinlichkeit die Abnehmer dar, die neben der Beleuchtung einen regen Gebrauch von ihren Kleinapparaten, z. B. von Bügeleisen, Staubsauger und Radio, machen. Der mittlere spezifische Verbrauch dieser Gruppe stellt sich auf 31,4 kWh pro Hauptraum, was für eine Wohnung mit 5 Haupträumen (4 Zimmer mit Küche, ohne Bad oder 3 Zimmer mit Küche und Bad) einem mittleren Jahresverbrauch von 157 kWh entspricht. Das Kollektiv *K III* schliesslich dürfte diejenigen Abonnenten gruppieren, die neben der Beleuchtung und den Kleinapparaten auch noch Raumheizung in mässigen Grenzen betreiben; sie machen rund 46 % aller Abnehmer aus. Für eine Wohnung mit 5 Haupträumen beträgt in diesem Falle der mittlere Jahresverbrauch 228,5 kWh.

Nimmt man mangels genauerer Angaben an, die Verteilung der verschiedenen Wohnungskategorien sei in allen drei Teilkollektiven identisch und gleich der Verteilung im Gesamtkollektiv, so ergeben sich die in Tabelle XIII zusammengestellten Werte.

Tabelle XIII

Kollektiv	Anteil am Gesamtkollektiv %	Anzahl der Haupträume	Mittl. spez. Verbrauch pro Hauptraum kWh	Jahresverbrauch pro Kollektiv kWh
III	10	103	14,6	1504
II	44	454	31.4	14256
I	46	475	45,7	21707
Total	100	1032	36.3	37467

Diese Werte stimmen mit den früher auf andere Arten ermittelten Werten ziemlich gut überein (mittl. spez. Verbrauch = 36,5 kWh/GE und Gesamtverbrauch = 37 293 kWh).

Die durchgeführte Analyse erlaubt weiter, den mutmasslichen mittleren Energieverbrauch pro Hauptraum, je für Beleuchtung, Kleinapparate und Zusatzheizung in der Uebergangszeit (denn 1938 wurde noch nicht so intensiv elektrisch geheizt, wie während und nach dem zweiten Weltkrieg) zu ermitteln. Die Differenzbildung zwischen den drei Teilkollektiven ergibt für

Beleuchtung:	14,6 kWh/Hauptraum
Kleinapparate:	16,8 kWh/Hauptraum
Zusatzheizung:	14,3 kWh/Hauptraum

Mit diesen Zahlen fällt es nicht schwer, für jede Wohnungsgrösse die entsprechenden Werte zu berechnen.

f) Schlussbemerkung

Eine solche mathematisch-statistische Untersuchung des Haushaltenergieverbrauches ist unseres Wissens noch nie durchgeführt worden. Es dürfte sicher von Interesse sein, ob die so ermittelten Zahlen mit den praktischen Erfahrungen übereinstimmen. Sollte dies zutreffen, so wäre mit der geschilderten Methode ein wirkungsvolles Instrument geschaffen, um die Struktur des Energieverbrauches ohne umfangreiche und langwierige statistische Erhebungen (im landläufigen Sinne des Wortes) zu untersuchen.

Literatur

[1] *Linder, Arthur:* Statistische Methoden für Naturwissenschafter, Mediziner und Ingenieure. 150 S. — Basel, 1945.

[2] *Fueter, Rudolf:* Das Mathematische Werkzeug des Chemikers, Biologen, Statistikers und Soziologen; Vorlesungen über die höheren mathematischen Begriffe in Verbindung mit ihren Anwendungen. 303 S. — Zürich u. Leipzig, 2.*ed. 1930.

[3] *Klezl, Felix:* Allgemeine Methodenlehre der Statistik, ein Lehrbuch für alle wissenschaftlichen Hochschulen. 295 S. — Wien, 2. ed. 1946.

[4] *Hack, Franz:* Wahrscheinlichkeitsrechnung. 123 S. — Sammlung Göschen, Nr. 508. Berlin u. Leipzig, 1914.

[5] *Daeves, Karl,* u. *August Beckel:* Auswertung von Betriebszahlen und Betriebsversuchen durch Grosszahl-Forschung; eine Methodik zur Vermeidung von Ausschussware und zur Verbesserung der Qualität. 43 S. — Berlin, 2. ed. 1943. [Auch in: Chem. Fabrik Bd. 14 (1941), S. 131 ff.]

[6] *Daeves, Karl:* Praktische Grosszahl-Forschung, Methoden zur Betriebs-Überwachung und Fehlerbeseitigung. VI + 132 S. — Berlin, 1933.

[7] *Meyer de Stadelhofen, Jean:* Statische Erhebungen über die Radiohörer und den für den Radiobetrieb konsumierten elektrischen Strom. Techn. Mitt". schweiz. Telegr.- u. Teleph.-Verw. Bd. 24(1946), Nr. 6, S. 258...265.

[8] *Beckel, August,* u. *Karl Daeves:* Ein neues Hilfsmittel der Grosszahlforschung. Stahl u. Eisen Bd. 54(1934), Nr. , S. 1305... .

[9] *Beckel, August:* Anwendung der Grosszahl-Forschung in der Lebensmittelchemie. Z. Unters. Lebensm. Bd. 66 (1933), Nr. 1/2, S. 158...177.

[10] *Beckel, August:* Grosszahlforschung in der Lebensmittelchemie: Buttersäurezahl. Z. Unters. Lebensm. Bd. 79 (1940), Nr. 1/2, S. 128...137.

Adresse des Autors:
Ch. Morel, dipl. Ingenieur, Deyenstrasse, Feldmeilen (ZH).

Etude sur la marche en parallèle des centrales des Forces Motrices Bernoises S. A.

Par *Ch. Jean-Richard,* Muri (BE)

621.3.016.32:621.311.21(494.24)

La marche en parallèle des centrales des Forces Motrices Bernoises s'effectue au moyen d'un réseau à 45 kV, formé de plusieurs mailles, mises en parallèle. A trois endroits de ce réseau il est opportun de maintenir la tension constante pour une valeur moyenne du facteur de puissance. Aux autres endroits il y a lieu de régler la puissance réactive en fonction de la puissance active dans le but d'orienter le flux de la puissance réactive vers les consommateurs.

L'emploi systématique de régulateurs automatiques, soit de tension soit de puissance réactive et de relais «Deltavar» pour le choix automatique du mode de réglage, présentera l'avantage de rendre l'exploitation des réseaux à la fois plus simple, plus économique et plus stable.

Der Parallelbetrieb der einzelnen Kraftwerke der BKW (Bernischen Kraftwerke A.-G.) erfolgt über ein 45-kV-Netz, das aus mehreren parallel geschalteten Maschen besteht. An drei Stellen dieses Netzes ist es angezeigt, die Spannung bei einem mittleren Leistungsfaktor konstant zu halten. An den übrigen Stellen sollte die Blindleistung in Abhängigkeit der Wirkleistung so reguliert werden, dass die Blindleistung in Richtung der Verbraucher fliesst.

Die systematische Verwendung automatischer Regler, seien es Spannungs- oder Blindleistungsregler, und von «Deltavar»-Relais für die automatische Wahl der Regulierart weist den Vorteil auf, den Betrieb der Netze gleichzeitig einfacher, wirtschaftlicher und ausgeglichener zu gestalten.

Huit centrales sont exploitées par les FMB (Forces Motrices Bernoises). Deux d'entre elles se trouvent à l'une des extrémités d'un réseau à 150 kV de forme triangulaire. Leur puissance disponible est de 260 MW et l'énergie de 800 000 MWh par an. L'étendue du réseau est de 628 km de lignes simples, montées sur pylones, principalement à deux ternes. La puissance réactive à vide est de 41 MVar à 150 kV.

Les six autres centrales sont échelonnées le long des cours d'eau qui sillonnent le canton de Berne. Leur puissance est de 79 MW et l'énergie de 460 000 MWh par an.

La pointe maximale de la charge des réseaux des FMB au mois d'août 1947 a été de 329 MW.

Les six centrales sont reliées par un réseau à 45 kV, formé de plusieurs mailles. En outre, elles alimentent chacune une partie d'un réseau à 16 kV, de même que des sous-stations 45/16 kV.

Aux barres omnibus à 16 kV de ces centrales la tension est prescrite par un diagramme hebdomadaire ajusté à la main et maintenue par régulateurs automatiques de tension. Cette tactique peut être désignée à juste titre par classique. Elle remonte aux débuts du service électrique, les besoins étant déterminés par une clientèle dont les occupations se règlent d'après les heures du jour.

Logiquement, l'emploi de régulateurs automatiques de tension aux barres omnibus à 16 kV des sous-stations donne de bons résultats.

Les deux centrales branchées au réseau à 150 kV sont assez puissantes pour imposer la tension à tout le réseau de manière à ce qu'elle soit constante en un point donné de ce réseau.

Le réseau à 45 kV n'est pas aussi facile à exploiter. Il assure la marche en parallèle des centrales entre elles tout en étant placé entre le réseau à