Energie-Erzeugung und -Verteilung

Die Seiten des VSE

Die modernen statistischen Methoden im Dienste der Elektrizitätswerke

Von *Ch. Morel*, Zürich

519.24 : 658.8.03 : 621.311

Die in der mathematischen Statistik in den letzten Jahren erzielten Fortschritte rechtfertigen eine immer weitere Anwendung der von dieser noch jungen Wissenschaft zur Verfügung gestellten Mittel, zur Lösung der mannigfaltigen Probleme des Elektrizitätswerkbetriebes. In beschränktem Rahmen dieses Aufsatzes wird versucht, die Grundprinzipien der modernen statistischen Methoden herauszuschälen, und ihre praktische Anwendung anhand einiger einfacher Beispiele zu zeigen. Es wird insbesondere auf die Gefahren einer allzu engen Auslegung der Ergebnisse hingewiesen, die immer mit einem möglichen Fehler behaftet sind, der aber zahlenmässig feststellbar ist und so berücksichtigt werden kann. Die Korrelationsrechnung wird eingehend behandelt, was gleichzeitig erlaubt, zu zeigen, wie die Streuungszerlegung vor sich geht. Schliesslich wird ein kurzer Überblick über die Anwendung der statistischen Verteilungen zur Lösung von Problemen des Betriebes gegeben.

Les progrès réalisés au cours des dernières années dans le domaine de la statistique mathématique justifient l'application toujours plus étendue des méthodes d'investigation que cette science encore jeune met à disposition, à l'étude des problèmes les plus divers de l'exploitation des entreprises de production et de distribution d'énergie électrique. Dans le cadre limité qui lui est imposé, le présent exposé essaie de dégager les principes fondamentaux des méthodes statistiques modernes et de montrer, par quelques exemples simples, leur application pratique. Il insiste en particulier sur les dangers d'une interprétation par trop étroite des résultats qui sont toujours affectés d'une erreur possible, chiffrable il est vrai, mais dont il faut tenir compte. Le problème de la corrélation y est traité en détail, ce qui permet de montrer également comment fonctionne l'analyse de la variance. Un bref aperçu de l'application des distributions statistiques à la solution de problèmes que soulève l'exploitation termine cet exposé.

Vorbemerkung

Die Statistik spielt im Betrieb eine unbestreitbare Rolle, die darin besteht, das Geschehene *zusammenzufassen*, um daraus *Schlüsse* zu ziehen (z. B. Verbesserung des Betriebes) oder um darauf die *Planung* für die Zukunft aufzubauen (z. B. Bau neuer Anlagen). Sie ist nicht Selbstzweck, sondern ein Werkzeug, dessen Verwendung mit gewissen Risiken verbunden ist. Die heutige Konzeption der Statistik und die modernen Prüfverfahren erlauben aber, diese Klippen zu umfahren und, was besonders wichtig ist, die Präzision der Ergebnisse zahlenmässig festzulegen.

Die erste Aufgabe des Statistikers besteht darin, die Angaben zu *sammeln* und sie wenn möglich durch Zahlen auszudrücken. Diese Arbeit mag auf den ersten Blick langweilig erscheinen, sie ist aber nicht ohne Bedeutung, denn von ihr hängt grösstenteils das Endergebnis ab. Alsdann sind die gesammelten Daten zu *beschreiben*, sie sprechen zu lassen, sie lebendig werden zu lassen. Hier gilt es, Summen zu bilden und Mittelwerte auszurechnen, die Variabilität von Zahlenreihen und die Beziehungen zwischen diesen Reihen zu untersuchen. Schliesslich können die Ergebnisse ausgewertet und *ausgelegt* werden, unter Verwendung der verschiedenen im Laufe der letzten Jahrzehnte geschaffenen Prüfverfahren (tests). Letztere Arbeit verlangt insbesondere ein enges Zusammengehen des Statistikers mit dem Praktiker oder dem Spezialisten.

1. Sammeln der Angaben

Die vom Statistiker gesammelten Zahlenreihen stellen keine anonyme Masse dar, sondern Gruppen von Einzelwerten, von denen jeder sein Eigenleben und seine besonderen Eigenschaften besitzt, die bei den Berechnungen zu beachten sind. Die Einzelwerte mögen auf den ersten Blick vielfältig erscheinen. Sie gehören aber zu einer Grundgesamtheit, aus der sie eine Stichprobe bilden. Die Gesetze der Grundgesamtheit gelten aber für jede Stichprobe, allerdings innerhalb gewisser, durch die Wahrscheinlichkeit bestimmter Grenzen. Die statistischen Masszahlen der Stichprobe sind Schätzungen derjenigen der Grundgesamtheit, der sie entstammt, und die Präzision dieser Masszahlen kann beziffert werden.

Diese Auffassung hat einen unmittelbaren, praktischen Wert. Um genaue Auskünfte zu erhalten, bedarf es nicht mehr unendlicher Zahlenreihen. Eine Stichprobe genügt, deren Umfang, neben den statistischen Charakteristiken der betrachteten Gruppe, von der gewünschten Präzision abhängt. Auf die Theorie der Stichprobe kann hier nicht eingegangen werden, da sie aus dem Rahmen dieser Arbeit fallen würde.

In den meisten Fällen liefert der Betrieb im weitesten Sinne des Wortes — also auch der kommerzielle Dienst und die Buchhaltung — die Angaben, die statistisch auszuwerten sind, und zwar in Form von Zahlen. Die Statistik kann aber vor Probleme gestellt werden, für die der Betrieb keine fertigen Unterlagen zu liefern imstande ist. Um zu seinem Ziele zu kommen, muss dann der Statistiker Umfragen vornehmen. Mit diesem Kapitel allein könnten Seiten gefüllt werden. Es seien hier nur einige allgemeine Betrachtungen gegeben. Die Fragebogen sollen möglichst einfach sein und keine suggestiven oder zweideutigen Fragen enthalten. Nötigenfalls soll der Fragebogen vor der endgültigen Redaktion zehn bis zwanzig Personen unterbreitet werden, was erlauben wird, allfällig unklare Fragen besser zu formulieren. Der Aufbau des Fragebogens muss der vorgesehenen Art der Auswertung (manuelle oder maschinelle) Rechnung tragen und Kontrollmöglichkeiten aufweisen.

Die heikelste Frage ist diejenige des Umfanges der Stichprobe, denn die Untersuchung wird kaum auf alle verfügbaren Einzelwerte ausgedehnt werden können. Dies gilt nicht nur für die Umfragen, sondern für alle statistischen Untersuchungen. Es sei diesbezüglich auf die Literaturangabe am Ende dieses Aufsatzes hingewiesen.

Ein anderes Problem, das zu erwähnen ist, ist dasjenige der Abnahme und der Prüfung des Materials. Zum Beispiel die Kontrolle der Lebensdauer von Glühlampen, die Prüfung von Isolatoren, die Kontrolle der Leitfähigkeit von Kupfer, die Prüfung der Isolieröle, die Kontrolle der Qualität der Kohle usw. Da das zu prüfende Material in der Regel bei der Prüfung zerstört wird, stellt sich die Frage, wie gross muss die Stichprobe sein, um bei gegebener Genauigkeit ein Maximum an Information zu gewinnen. Es ist leider nicht möglich, im Rahmen der vorliegenden Arbeit auf diese Frage näher einzutreten. Die Fachliteratur wird hierüber Auskunft geben.

2. Das Beschreiben der statistischen Gesamtheiten

Die gesammelten Daten bilden im allgemeinen eine formlose Masse, die es zu ordnen gilt. Zuerst muss ihre Verteilung untersucht werden. Diese beschreibende Operation erlaubt meistens, das Verteilungsgesetz, dem die Gesamtheit gehorcht, zu erkennen. Die zweite Operation besteht darin, die Charakteristiken oder Parameter der Verteilung zu berechnen. Liegen mehrere Zahlenreihen vor, von denen vermutet wird, dass sie in gewissen Beziehungen zueinander stehen, so werden die statistischen Abhängigkeiten untersucht.

Beschreibung der Angaben

Um die Verteilung erkennen zu können, müssen die Einzelwerte zunächst geordnet werden; dies erfolgt durch Einteilung in Klassen und alsdann innerhalb der Klassen durch Sortierung nach steigendem oder fallendem Wert. Die Gruppierung richtet sich nach dem Charakter der zu untersuchenden Veränderlichen. Die Variationsbreite (Intervall zwischen dem grössten und dem kleinsten Einzelwert) wird in eine Anzahl Klassen gleicher Breite aufgeteilt, derart dass auf jede Klasse eine genügende Anzahl Einzelwerte fällt, dass aber trotzdem der Charakter der Verteilung gewahrt bleibt.

Die zweckmässigste graphische Darstellung ist die Treppenkurve oder Histogramm, denn damit wird gleichzeitig ausgedrückt, dass es sich nicht um eine stetige Funktion, sondern um diskrete Werte handelt.

Die Verteilungskurve (Häufigkeit je Klasse) wird durch die Summenhäufigkeitskurve (Summe der Häufigkeiten aller Klassen bis zur betrachteten, diese inbegriffen) ergänzt.

Je nach dem vorliegenden Falle werden die absoluten oder die relativen Häufigkeiten gewählt. Als relative Häufigkeit ist das in Prozenten ausgedrückte Verhältnis zwischen der betrachteten Häufigkeit und der Summe aller Häufigkeiten (Umfang der Stichprobe) zu verstehen.

Statistische Verteilungen

Die bekannteste Verteilung ist die Normalverteilung von Gauss-Laplace, auch Fehler- oder Glockenkurve genannt. Es ist eine symmetrische Funktion, deren mathematischer Ausdruck lautet

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Die zwei Parameter dieser Funktion sind μ (Durchschnitt) und σ^2 (Streuung). Durch eine einfache Transformierung kann sie in die Standard-Form übergeführt werden, mit $\mu = 0$ und $\sigma^2 = 1$. Man setzt

$$u = \frac{x-\mu}{\sigma}$$

und erhält

$$\varphi(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}}$$

Der Ausdruck $\varphi(x)\,dx$ stellt die Wahrscheinlichkeit dafür dar, dass eine Abweichung zwischen x und $x + dx$ liegt. Die Summe der Elementarwahrscheinlichkeiten $\int_{-\infty}^{x} \varphi(x)\,dx$ ist die Wahrscheinlichkeit für das Auftreten eines Fehlers, der kleiner oder gleich x ist. Diese Zahl ist immer kleiner als 1; es lässt sich leicht zeigen, dass $\int_{-\infty}^{+\infty} \varphi(x)\,dx = 1$ ist.

Eine andere Verteilung, wovon die Normalverteilung nur ein Grenzfall darstellt, ist die Binomialverteilung, deren Gleichung lautet

$$\varphi(x) = \binom{m}{x} p^x (1-p)^{m-x}$$

wobei

$$\binom{m}{x} = \frac{m!}{x!\,(m-x)!}$$

Es ist dies die Wahrscheinlichkeit dafür, dass ein Ereignis mit Einzelwahrscheinlichkeit $p < 1$ in einer Reihe von m Versuchen $x$mal auftritt (z. B. Wahrscheinlichkeit für die Ziehung von x roten Kugeln bei m Versuchen, wenn die Urne 100 rote Kugeln von insgesamt 1000 Kugeln enthält; $p = 1/10$).

Der Durchschnitt (oder mathematische Hoffnung) wird

$$\mu = m\,p$$

und die Streuung

$$\sigma^2 = m\,p\,(1-p)$$

Strebt der Wert m gegen unendlich zu und bleibt p endlich, so geht die Binomialverteilung in die Normalverteilung über.

Dagegen, wenn m gegen unendlich und p gegen null zustreben, während das Produkt $mp = \lambda$ endlich bleibt, so entsteht eine andere Verteilung: die Poisson'sche Verteilung oder Verteilung der seltenen Ereignisse. Ihr mathematischer Ausdruck lautet:

$$\varphi(x) = \frac{\lambda^x\, e^{-\lambda}}{x!}$$

dabei ist

$$\mu = \lambda \quad \text{und} \quad \sigma^2 = \lambda$$

Berechnung der statistischen Masszahlen

Mit Hilfe der mathematischen Statistik kann eine Gesamtheit numerischer Angaben durch einige Parameter charakterisiert werden. So können Gesamtheiten durch die Gegenüberstellung ihrer Parameter verglichen, oder durch ihre Parameter in vereinfachter Form dargestellt werden.

Das erste zu bestimmende Parameter ist dasjenige der Lage, oder Positionsparameter; in der Regel ist es der Durchschnitt $\bar{x}$. Dieser erfüllt die statistischen Anforderungen am besten, denn er ist passend, wirksam und erschöpfend. Andere, weniger wirksame Positionsparameter sind der Medianwert (Zentralwert, mittelster Wert) und der häufigste Wert; sie können von gewissem Nutzen sein wenn die Verteilung allzusehr von der Normalverteilung abweicht.

Zählt die Stichprobe N Einzelwerte x_i, die in M Klassen mit Mitte x_j und Häufigkeit f_j verteilt sind, so schreibt sich der Durchschnitt

$$\bar{x} = \frac{1}{N} \mathop{\mathrm{S}}_{i=1}^{N} x_i = \frac{1}{N} \mathop{\mathrm{S}}_{j=1}^{M} f_j\, x_j, \quad \text{wobei } N = \mathop{\mathrm{S}}_{j=1}^{M} f_j$$

Das wirksamste Mass der Variabilität ist die Streuung s^2 (oder die Standardabweichung s). Ihre Formel ist

$$s^2 = \frac{1}{N-1} \mathop{\mathrm{S}}_{i=1}^{N} (x_i - \bar{x})^2 = \frac{1}{N-1} \mathop{\mathrm{S}}_{j=1}^{M} f_j\, (x_j - \bar{x})^2$$

Für die praktische Berechnung können folgende Formeln von Nutzen sein:

$$S\,(x_i - \bar{x})^2 = S\,x_i^2 - \frac{1}{N}\,(S\,x_i)^2 =$$

$$= S\,x_i^2 - \bar{x}\,S\,x_i = S\,x_i^2 - N\,\bar{x}^2$$

und

$$S f_j\,(x_j - \bar{x})^2 = S f_j\,x_j^2 - \frac{1}{N}\,(S f_j\,x_j)^2 =$$

$$= S f_j\,x_j^2 - \bar{x}\,S f_j\,x_j = S f_j\,x_j^2 - N\,\bar{x}^2$$

Ein anderes Mass für die Variabilität ist die Variationsbreite $w = x_N - x_1$, die z. B. in der laufenden Fabrikationskontrolle häufig benutzt wird.

Diese Berechnungen haben dann einen präzisen Sinn, wenn die Verteilung normal oder annähernd normal ist. Oft kann eine unsymmetrische Verteilung durch Transformierung in eine Normalverteilung verwandelt werden. In vielen Fällen führt eine Transformierung zum Ziele, bei der $z = \log (a + bx)$ oder einfacher $z = \log x$ gesetzt wird. Es gibt auch Fälle, bei denen $z = x^2$ oder $z = \frac{1}{x}$ gesetzt werden muss.

Ein einfaches Mittel, eine Normalverteilung zu erkennen, besteht in der Aufzeichnung der Summenhäufigkeitskurve auf Wahrscheinlichkeitspapier. Entsteht dabei eine Gerade oder annähernd eine Gerade, so kann auf eine Normalverteilung geschlossen werden.

Statistische Abhängigkeiten

In statistischen Untersuchungen kommen oft mehrere Veränderliche vor, von denen es die gegenseitigen Beziehungen herauszuschälen gilt.

Es scheint angezeigt, hier daran zu erinnern, dass die Beziehungen zwischen statistischen Zahlenreihen in der Regel keine funktionelle sind. Jedem Wert der einen Veränderlichen können mehrere verschiedene Werte der andern Veränderlichen entsprechen. Die mehr oder weniger lockere Abhängigkeit ist eine stochastische. Sie erlaubt, aus den Werten der einen Veränderlichen auf den wahrscheinlichen Wert der andern zu schliessen oder, mit anderen Worten, die Grenzen festzulegen zwischen denen erwartet werden kann, dass die andere Veränderliche variiert. Die Funktion, die die beiden Veränderlichen bindet, ist die Regressionsfunktion. Die Regression kann oft in erster Annäherung als geradlinig (lineare Regression) angenommen werden. Es gibt aber auch Fälle, wo die Regression höherer Ordnung ist. Ein Mass für die Dichte, mit der sich die Einzelwerte um die Regressionsfunktion scharen, liefert das Bestimmtheitsmass, und zwar unabhängig von der Form der Regression.

Die lineare Regression kann einfach:

$$Y = a + b\,(x - \bar{x})$$

oder mehrfach sein:

$$Y = a + b_1\,(x_1 - \bar{x}_1) + b_2\,(x_2 - \bar{x}_2) + \ldots$$

wobei x_1, x_2, ... die unabhängigen Veränderlichen sind.

In diesen Gleichungen ist die Konstante a gleich dem Durchschnitt $\bar{y}$ und der Regressionskoeffizient b wird nach der Methode der kleinsten Quadrate ermittelt. So erhält man für die einfache lineare Regression in x:

$$b = \frac{S\,(x - \bar{x})\,(y - \bar{y})}{S\,(x - \bar{x})^2}$$

und für die doppelte lineare Regression

$$b_1 = \frac{S\,(x_1 - \bar{x}_1)\,(x_2 - \bar{x}_2)\,S\,(x_2 - \bar{x}_2)\,(y - \bar{y}) - S\,(x_2 - \bar{x}_2)^2\,S\,(x_1 - \bar{x}_1)\,(y - \bar{y})}{[S\,(x_1 - \bar{x}_1)\,(x_2 - \bar{x}_2)]^2 - S\,(x_1 - \bar{x}_1)^2\,S\,(x_2 - \bar{x}_2)^2}$$

und

$$b_2 = \frac{S\,(x_1 - \bar{x}_1)\,(x_2 - \bar{x}_2)\,S\,(x_1 - \bar{x}_1)\,(y - \bar{y}) - S\,(x_1 - \bar{x}_1)^2\,S\,(x_2 - \bar{x}_2)\,(y - \bar{y})}{[S\,(x_1 - \bar{x}_1)\,(x_2 - \bar{x}_2)]^2 - S\,(x_1 - \bar{x}_1)^2\,S\,(x_2 - \bar{x}_2)^2}$$

Es würde zu weit führen, hier noch die wesentlich komplizierteren Formeln für die Koeffizienten b bei drei- und mehrfacher linearer Regression anzuführen.

Das Bestimmtheitsmass, das nichts anderes ist als das Verhältnis der Streuung der Regressionsfunktion zur Gesamtstreuung der abhängigen Veränderlichen, schreibt sich allgemein

$$B = \frac{S\,(Y - \bar{y})^2}{S\,(y - \bar{y})^2}$$

Für die einfache Regression lautet es:

$$B = \frac{[S\,(x - \bar{x})\,(y - \bar{y})]^2}{S\,(x - \bar{x})^2\,S\,(y - \bar{y})^2} = b\,\frac{S\,(x - \bar{x})\,(y - \bar{y})}{S\,(y - \bar{y})^2}$$

und für die doppelte Regression:

$$B = \frac{b_1\,S\,(x_1 - \bar{x}_1)\,(y - \bar{y}) - b_2\,S\,(x_2 - \bar{x}_2)\,(y - \bar{y})}{S\,(y - \bar{y})^2}$$

Bei der einfachen linearen Regression ist der Korrelationskoeffizient r gleich der Quadratwurzel aus dem Bestimmtheitsmass B:

$$r^2 = B \text{ oder } r = \sqrt{B}$$

Die nichtlineare Regression, z. B.

$$Y = a + b\,x + c\,x^2 + \ldots$$

wird wie folgt behandelt:

Man setzt

$$Y = \bar{y} + b_1\,(x_1 - \bar{x}_1) + b_2\,(x_2 - \bar{x}_2) + \ldots$$

wobei

$$x_1 = x, \quad x_2 = x^2, \text{ usw.}$$

Für die Lösung wird alsdann gleich verfahren wie bei der mehrfachen linearen Regression.

Um die Form der nichtlinearen Regression zu ermitteln, bedient man sich vorteilhaft der von Fisher angegebenen Methode der Orthogonalpolynome. Man wird jedoch auf diese Rechnung verzichten können, wenn der Grad der Regressionsgleichung auf den ersten Blick erkenntlich ist oder sonstwie starke Anhaltspunkte dafür vorliegen.

3. Die Auslegung der Ergebnisse

Die Auslegung erlaubt in erster Linie, aus den Erkenntnissen über die Stichproben auf die Grundgesamtheiten denen sie entstammen, Rückschlüsse zu ziehen, und zwar mit Hilfe der statistischen Prüfverfahren. Sie gestattet ebenfalls, die Variabilität der Erscheinungen mit ihren Ursachen in Beziehung zu bringen: das ist Sache der Streuungszerlegung.

Die statistischen Prüfverfahren

Die hier am häufigsten angewandte Methode besteht darin, über eine oder mehrere Grundgesamtheiten eine Hypothese aufzustellen, und dann mit Hilfe eines geeigneten Prüfverfahrens festzustellen, in welchem Masse die Hypothese durch das Studium einer oder mehrerer Stichproben aus diesen Grundgesamtheiten unzutreffend ist. Wenn dieses Mass stark ist, d. h. wenn die Abweichung wesentlich oder gesichert ist, so muss die Hypothese verworfen werden. Ist das Mass schwach, oder die Abweichung zufällig, so kann nur auf das Fehlen von Gründen zur Verwerfung der Hypothese geschlossen werden. Man wird somit versuchen müssen, die Hypothese so aufzustellen, dass ihre Verwerfung die gewünschte Gewissheit bringt.

Die Sicherheitsschwellen

Vor der Aufzählung der Hauptgesetze, denen die Beziehungen zwischen den Masszahlen einer Grundgesamtheit und denjenigen der dieser entnommenen Stichproben folgen, scheint es nützlich, die gewöhnlich für Vergleiche angewandte Methode kurz zu streifen. Die meisten statistischen Grössen und ihre Abweichungen weisen bekannte Verteilungen auf (normale oder andere Verteilungen), so dass jeder Grösse oder Abweichung eine bestimmte Wahrscheinlichkeit zukommt. Die Erfahrung hat zur Festlegung einer Sicherheitsschwelle geführt. Diese wird definiert als der Wert für den die Wahrscheinlichkeit, dass eine beliebige Abweichung gleich gross oder grösser ist, p % beträgt. Es werden gewöhnlich zwei Sicherheitsschwellen angenommen, die den Wahrscheinlichkeiten $p = 5$ % und 1 % (oder 0,05 und 0,01) entsprechen. Eine Abweichung, deren Wahrscheinlichkeit grösser ist als 5 % wird als zufällig, eine grössere Abweichung, d. h. eine solche, die einer Wahrscheinlichkeit von weniger als 5 % entspricht, als gesichert betrachtet. Beträgt die Wahrscheinlichkeit weniger als 1 %, so gilt die Abweichung als stark gesichert.

In einer Normalverteilung entspricht z. B. die Schwelle 5 % ($u_{0{,}05}$) einer Abweichung von 1,96 σ und die Schwelle 1 % ($u_{0{,}01}$) einer solchen von 2,576 σ. In erster Annäherung kann also gesagt werden, dass jede Abweichung, die mehr als das Doppelte der Standard-Abweichung beträgt ($u > 2\,\sigma$), ge-

Tabelle I

Beobachtete Grösse (Stichprobe)	Genauer Wert (Grundgesamtheit)	Streuung	Verteilung der beobachteten Grösse	
			Kleine Stichprobe	Grosse Stichprobe
Durchschnitt $\bar{x} = \frac{1}{N}\,S\,x$	μ	$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{N}$	normal	normal
Streuung $s^2 = \frac{1}{N-1}\,S\,(x - \bar{x})^2$	σ^2	$\sigma_{s^2}{}^2 \cong \frac{2\,\sigma^4}{N}$	unsymmetrisch	normal
Standard-Abweichung $s = \sqrt{\frac{1}{N-1}\,S\,(x - \bar{x})^2}$	σ	$\sigma_s{}^2 = \frac{\sigma^2}{2\,N}$	unsymmetrisch	normal
Linearer Regressionskoeffizient $b = \frac{S\,(x - \bar{x})\,(y - \bar{y})}{S\,(x - \bar{x})^2}$	β	$\sigma_b{}^2 = \frac{S\,(y - Y)^2}{N - 2}$	unsymmetrisch	normal
Korrelationsfaktor $r = \frac{S\,(x - \bar{x})\,(y - \bar{y})}{\sqrt{S\,(x - \bar{x})^2 \cdot S\,(y - \bar{y})^2}}$	ϱ	$\sigma_r{}^2 = \frac{(1 - \varrho^2)^2}{N - 1}$	unsymmetrisch	annähernd normal

Tabelle II

Grosse Stichprobe (N > 100)		Kleine Stichprobe (N < 100)	
Abweichung zwischen beobachtetem Wert und genauem Wert	Abweichung zwischen zwei beobachteten Werten	Abweichung zwischen beobachtetem Wert und genauem Wert	Abweichung zwischen zwei beobachteten Werten
1. Durchschnitte			
$u = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ $\sigma^2_{\bar{x}} = \frac{s^2}{N}$	$u = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_d}$ $\sigma_d^2 = \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}$	$t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ $\sigma^2_{\bar{x}} = \frac{s^2}{N}$ $n = N - 1$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_d}$ $\sigma_d^2 = \frac{S(x_1 - \bar{x}_1)^2 + S(x_2 - \bar{x}_2)^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2}\right)$ $n = N_1 + N_2 - 2$
2. Standard-Abweichungen			
$u = \frac{s - \sigma}{\sigma_s}$ $\sigma_s^2 = \frac{s^2}{2\,N}$	$u = \frac{s_1 - s_2}{\sigma_d}$ $\sigma_d^2 = \frac{s_1^2}{2\,N_1} + \frac{s_2^2}{2\,N_2}$	$F = \frac{s^2}{\sigma^2}$ $n_1 = N - 1 \quad n_2 = \infty$	$F = \frac{s_1^2}{s_2^2}$ $n_1 = N_1 - 1 \quad n_2 = N_2 - 1$
3. Regressionskoeffizienten			
$u = \frac{b - \beta}{\sigma_b} \sqrt{S(x - \bar{x})^2}$ $\sigma_b^2 = \frac{S(y - Y)^2}{N - 2}$	$u = \frac{b_1 - b_2}{\sigma_d}$ $\sigma_d^2 = \frac{S(y_1 - Y_1)^2 + S(y_2 - Y_2)^2}{N_1 + N_2 - 4} \cdot \left(\frac{1}{S(x_1 - \bar{x}_1)^2} + \frac{1}{S(x_2 - \bar{x}_2)^2}\right)$	$t = \frac{b - \beta}{\sigma_b} \sqrt{S(x - \bar{x})^2}$ $\sigma_b^2 = \frac{S(y - Y)^2}{N - 2}$ $n = N - 2$	$t = \frac{b_1 - b_2}{\sigma_d}$ $\sigma_d^2 = \frac{S(y_1 - Y_1)^2 + S(y_2 - Y_2)^2}{N_1 + N_2 - 4} \cdot \left(\frac{1}{S(x_1 - \bar{x}_1)^2} + \frac{1}{S(x_2 - \bar{x}_2)^2}\right)$ $n = N_1 + N_2 - 4$
4. Korrelationskoeffizienten			
$u = \frac{r - \varrho}{\sigma_\varrho}$ $\sigma_\varrho^2 = \frac{(1 - \varrho^2)^2}{N - 1}$	$u = \frac{r_1 - r_2}{\sigma_d}$ $\sigma_d^2 = s^2_{r1} + s^2_{r2}$ $s_r^2 = \frac{(1 - r^2)^2}{N - 1}$	$t = \frac{r}{\sqrt{1 - r^2}} \sqrt{N - 2}$ $n = N - 2$	$u = \frac{z_1 - z_2}{\sigma_d}$ $z = \frac{1}{2} [l\,n\,(1 + r) - l\,n\,(1 - r)]$ $\sigma^2_d = s^2_{z1} + s^2_{z2}$ $s_z = \frac{1}{N - 3}$

In dieser Tabelle bedeuten:

u = Normalverteilung
t = Verteilung von Student-Fisher; n = Freiheitsgrad
F = Verteilung von Fisher-Snedecor; n_1 und n_2 = Freiheitsgrade.

Bei den praktischen Berechnungen können σ und ϱ durch die aus den Stichproben geschätzten Werte s und r ersetzt werden.

sichert ist. Die praktische Anwendung der Sicherheitsschwellen soll weiter unten näher dargelegt werden.

Anwendung der Prüfverfahren

Die Fälle, in denen statistische Prüfverfahren angewendet werden, sind meistens Vergleiche zwischen einer Stichprobe und der Grundgesamtheit, der sie entstammt (Vergleich der nach Beobachtung geschätzten Parameter mit den genauen Werten der Grundgesamtheit) oder Vergleiche zwischen zwei oder mehreren Stichproben (Vergleich der geschätzten Parameter ohne Kenntnis der genauen Werte.)

In beiden Fällen wird nach der Null-Hypothese geprüft, d. h. es wird a priori angenommen, dass die Stichprobe oder die Stichproben der gleichen Grundgesamtheit entstammen. Fällt die Prüfung negativ aus, so kann daraus geschlossen werden, dass die Stichprobe nicht der betrachteten Grundgesamtheit entstammt, oder das die einander gegenübergestellten Stichproben verschiedenen Grundgesamtheiten entnommen wurden.

Vergleich von statistischen Masszahlen (Parameter)

Die Stichprobenfehler der hauptsächlichsten statistischen Grössen, die in den verschiedenen Prüfungen oder Testen vorkommen, sind in der Tabelle I zusammengefasst.

Die Verteilung der Abweichungen zwischen beobachteten Masszahlen, oder zwischen einer beobachteten Masszahl und ihrem genauen Wert ist je nach dem Umfang der Stichprobe verschieden.

Die bei solchen Vergleichen anzuwendenden Formeln sind in der Tabelle II enthalten.

Streuungszerlegung

Die Streuung einer statistischen Gesamtheit ist einer Vielheit von Ursachen zuzuschreiben, von denen jede eine kleine Wirkung ausübt. Unter diesen Ursachen befinden sich aber immer einige, deren Wirkung ausgesprochen ist. Es ist die Aufgabe der Streuungszerlegung, diese Ursachen, die kontrollierten Ursachen, herauszuschälen und ihre Wirkung an der Wirkung der übrigen, unkontrollierbaren Ursachen, d. h. an der Reststreuung zu messen.

Diese Methode verlangt eine Gruppierung der beobachteten Grössen oder Einzelwerte in homogene Gruppen und Untergruppen, die die Wirkung der Ursachen, deren Einfluss zu untersuchen ist, hervortreten lassen. Oft ist es nötig, ausgedehnte Versuchspläne aufzustellen, die sehr gründliche Kenntnisse der Materie erfordern. Im nächsten Kapitel sollen einige praktische Beispiele der Streuungszerlegung gegeben werden.

4. Praktische Beispiele

A. Vergleich von Durchschnitten und Streuungen

Im Tarifwesen, wie auch im Betriebe und in der Buchhaltung bedient man sich oft des Durchschnittes. Man spricht von mittlerem Verbrauch, von mittlerem Preis, ja sogar von Durchschnittsbezügern. Man vergleicht untereinander die Mittelwerte aus mehreren Jahren oder Zeitabschnitten; man zieht aus diesen Vergleichen Schlussfolgerungen, die oft zur Begründung von Anträgen oder von Massnahmen dienen. Solche Vergleiche können unter Umständen irrig, ja irreführend sein, denn sie berücksichtigen eine Erscheinung nicht, die doch massgebend ist: die Streuung.

Als Beispiel diene eine Gruppe von Haushaltbezügern; es sei die Auswirkung einer Tarifänderung auf ihre Energierechnungen zu untersuchen. Um die Berechnungen zu vereinfachen, sollen aus der Abonnentenkartei deren 20 rein zufällig herausgegriffen werden. Das Prinzip bleibt sich gleich für eine Stichprobe grösseren Umfanges. Allein die dem Prüfverfahren zu Grunde zu legende Verteilung ändert sich je nachdem die Stichprobe nur einige Individuen zählt oder deren Tausende umfasst.

Die für die gewählte Abonnentengruppe gültigen Angaben gehen aus Tabelle III hervor, die auch den Rechnungsgang zu verfolgen gestattet. Die Rechnungsbeträge in Schweizerfranken sind aufgerundet. Sie beziehen sich jeweils auf den effektiven Verbrauch im Vergleichsjahr.

Tabelle III

Abonnent Nr.	Rechnung nach altem Tarif		Rechnung nach neuem Tarif		Differenz neuer-alter Tarif	
	x'	x'^2	x''	x''^2	$x_i = x'' - x'$	x_i^2
1	91	8 281	64	4 096	− 27	729
2	63	3 969	43	1 849	− 20	400
3	243	59 049	207	42 849	− 36	1 296
4	185	34 225	171	29 241	− 14	196
5	164	26 896	154	23 716	− 10	100
6	205	42 025	199	39 600	− 6	36
7	246	60 516	226	51 076	− 20	400
8	120	14 400	126	15 876	+ 6	36
9	187	34 969	190	36 100	+ 3	9
10	141	19 881	146	21 316	+ 5	25
11	112	44 944	121	14 641	+ 9	81
12	51	2 601	57	3 249	+ 6	36
13	24	576	22	484	− 2	4
14	30	900	37	1 369	+ 7	49
15	35	1 225	38	1 444	+ 3	9
16	36	1 296	38	1 444	+ 2	4
17	38	1 444	55	3 025	+ 17	289
18	27	729	45	2 025	+ 18	324
19	16	256	27	729	+ 11	121
20	11	121	26	676	+ 15	225
Summen	2 025	358 303	1 992	294 803	− 33	4 369
Durchschnitte	101,25	—	99,6	—	− 1,65	—

Damit der Unterschied in der Auswirkung beider Tarife gesichert erscheint, muss der Durchschnitt aus den einzelnen Abweichungen zwischen beiden Tarifen entscheidend von Null verschieden sein. Wenn man aus der Grundgesamtheit, aus der die untersuchte Gruppe auch eine Stichprobe darstellt, eine beliebige Anzahl Stichproben herausziehen würde, jede vom Umfang N, so würde sich herausstellen, dass die für jede Stichprobe berechneten Werte des Ausdruckes $t = \frac{\bar{x}}{s}\sqrt{N}$ nach einem bestimmten Gesetz verteilt sind, nach der t-Funktion, die den Vorteil aufweist, von der Streuung σ der Grundgesamtheit unabhängig zu sein. Der Verlauf dieser Funktion ist verschieden je nach dem Freiheitsgrade n, der hier gleich $N - 1 = 19$ ist. Je grösser der Freiheitsgrad ist, um so stärker nähert sich die t-Verteilung der normalen oder Gauss'schen

Verteilung, die einen Sonderfall davon für $N = \infty$ darstellt. In speziellen Tafeln findet man für jeden Freiheitsgrad die t-Werte, die einem $p = 0{,}05$ und $p = 0{,}01$ entsprechen.

Die Rechnung ergibt

$$s^2 = \frac{1}{19}\left(4369 - \frac{33^2}{20}\right) = 227{,}0 \text{ und}$$

$$s = \sqrt{227} = 15{,}07$$

so dass

$$t = \frac{1{,}65\sqrt{20}}{15{,}07} = 0{,}49$$

In der Tafel der t-Verteilung findet man, unter $n = 19$

für $p = 0{,}05$ $t_{0{,}05} = 2{,}093$

und für $p = 0{,}01$ $t_{0{,}01} = 2{,}861$

Der für t errechnete Wert ist wesentlich kleiner als $t_{0{,}05}$. Der Unterschied zwischen beiden Tarifen ist somit rein zufällig. Mit andern Worten: die durch den neuen Tarif verursachte Änderung der einzelnen Rechnungen bleibt innerhalb der natürlichen Streuung der betrachteten Abonnentengruppe. Der neue Tarif wird somit für die Unternehmung keine wesentliche Auswirkung haben, unter der Bedingung, dass der Energieverbrauch gleich bleibt, und dass die Abonnenten ihre Gewohnheiten nach der Einführung des neuen Tarifs beibehalten. Die vorliegende Berechnung setzt ferner voraus, dass der neue Tarif für alle Bezüger obligatorisch ist.

Dieses Beispiel bestätigt das bezüglich der Durchschnitte weiter oben Gesagte. Wenn man lediglich auf die Summen, bzw. Durchschnitte abgestellt hätte, um die Auswirkung des neuen Tarifes zu beurteilen, so hätte man einfach feststellen können, dass dieser Tarif einen Einnahmenausfall von der Grössenordnung 1,5 % mit sich bringt. Man wäre dabei versucht gewesen anzunehmen, dass dieser Ausfall sich bei jedem Abonnent eingestellt hätte. Ein Blick auf Tabelle I genügt aber, um zu erkennen, dass die Abonnenten sehr verschieden getroffen werden. Die einen erleiden eine starke Erhöhung ihrer Rechnung, während die andern mit einer namhaften Verbilligung ihres Energiebezuges wegkommen. Die starke Turbulenz innerhalb der betrachteten Abonnentengruppe wird durch den hohen Wert der Streuung bestätigt. Die mittlere quadratische Abweichung beträgt tatsächlich etwa das Zehnfache des Durchschnittes aus den Einzeldifferenzen.

Es können auch die Streuungen der beiden Reihen «Rechnung nach altem Tarif» und «Rechnung nach neuem Tarif» einander gegenübergestellt werden. Man rechnet für jede Serie s^2 und bildet das Quotient aus den beiden Werten. Für diesen Vergleich bedient man sich der Funktion F, die nur vom Freiheitsgrad der beiden Streuungen abhängt. Für die Werte F, die $p = 0{,}05$ und $p = 0{,}01$ entsprechen, gibt es besondere Tafeln. Man setzt $F = \frac{s'^2}{s''^2}$.

Für die Reihen von Tabelle I ist der Quotient der beiden Streuungen s'^2 und s''^2 identisch mit dem Quotient der Quadratsummen, da beide Reihen den gleichen Umfang haben. Als Quotient erhält man 1,197. In den Tafeln findet man für $p = 0{,}05$

$$F_{0{,}05} = 2{,}308 \text{ für } n_1 = 12 \text{ und } n_2 = 19$$
$$F_{0{,}05} = 2{,}114 \text{ für } n_1 = 24 \text{ und } n_2 = 19$$

Der genaue Wert für $n_1 = 19$ und $n_2 = 19$ kann interpoliert werden. Diese Rechnung ist jedoch überflüssig, da das errechnete F (1,917) wesentlich kleiner ist, als die Werte, die den genauen Wert einrahmen. Es ergibt sich daraus, dass der Unterschied zwischen den beiden Streuungen nur zufällig ist. Mit andern Worten wird die Einführung des neuen Tarifes keine Umwälzung in den Stromrechnungen hervorrufen, die über das Mass der Streuung unter dem alten Tarif hinausginge.

B. Vergleich von einfachen linearen Regressionen

In diesem Beispiel gilt es, den Einfluss der Einführung eines neuen Tarifes auf die Entwicklung des Energieverbrauches der Abonnenten zu untersuchen. Sind zum Beispiel für die betrachteten Abonnenten die jährlichen Verbrauchszahlen für einige Jahre vor und nach dem Tarifwechsel bekannt, so können

Tabelle IV

Abonnent Nr.	Energieverbrauch (y_i') in kWh, vor dem Tarifwechsel					Energieverbrauch (y_i'') in kWh, nach dem Tarifwechsel				
	Jahr 1	Jahr 2	Jahr 3	Jahr 4	Total	Jahr 1	Jahr 2	Jahr 3	Jahr 4	Total
1	170	218	296	278	962	483	423	571	516	1 993
2	88	95	101	100	384	134	118	125	146	523
3	117	132	124	109	482	134	121	152	166	573
4	104	113	116	120	453	170	171	138	176	655
5	216	226	247	265	954	245	242	261	228	976
6	151	167	167	144	629	179	212	208	247	846
7	72	82	75	84	313	135	134	163	161	593
8	134	124	146	162	566	175	209	205	225	814
9	81	81	81	78	321	89	87	103	115	394
10	140	104	96	93	433	156	211	299	262	928
11	119	105	101	144	469	166	160	153	124	603
12	161	177	185	201	724	197	199	194	184	774
13	164	172	175	170	681	244	236	259	291	1 030
14	153	159	155	156	623	221	220	224	167	832
15	71	92	101	101	365	130	146	182	238	696
Total	1 941	2 047	2 166	2 205	8 359	2 858	2 889	3 227	3 246	12 230
Durchschnitt	129	136	144	147	139	191	193	216	216	204

die Regressionen der beiden Perioden miteinander verglichen werden. Die Ausgangszahlen sind in Tabelle IV enthalten.

a) Berechnung der Regression «vor Tarifwechsel»

$$Y' = \bar{y}' + b'(x' - \bar{x}')$$

$$\bar{y}' = 8359 : 15 = 139$$

$$\bar{x}' = (1 + 2 + 3 + 4) : 4 = 2{,}5$$

$$b' = (1941 + 2 \cdot 2047 + 3 \cdot 2166 + 4 \cdot 2205 - 2{,}5 \cdot 8539) : (450 - 150^2 : 60) = 6{,}07$$

$$Y' = 139 + 6{,}07\,(x' - 2{,}5) = 6{,}07\,x' + 124$$

b) Berechnung der Regression «nach Tarifwechsel»

$$Y'' = \bar{y}'' + b''(x'' - \bar{x}'')$$

$$\bar{y}'' = 12\,230 : 15 = 204$$

$$\bar{x}'' = (1 + 2 + 3 + 4) : 4 = 2{,}5$$

$$b'' = (2858 + 2 \cdot 2889 + 3 \cdot 3227 + 4 \cdot 3246 - 2{,}5 \cdot 12\,230) : (450 - 150^2 : 60) = 10{,}08$$

$$Y'' = 204 + 10{,}08\,(x'' - 2{,}5) = 10{,}08\,x'' + 179$$

c) Vergleich der beiden Regressionskoeffizienten

$$t = \frac{b'' - b'}{\sigma_d}$$

$$\sigma_d^2 = \frac{S(y' - Y')^2 + S(y'' - Y'')^2}{N' + N'' - 4} \cdot \left(\frac{1}{S(x' - \bar{x}')^2} + \frac{1}{S(x'' - \bar{x}'')^2}\right)$$

$$S(y' - Y')^2 = S\,y'^2 - \frac{1}{N}(S\,y')^2 - \frac{\left[S\,x'y' - \frac{1}{N}S\,x'\,S\,y'\right]^2}{S\,x'^2 - \frac{1}{N}(S\,x')^2}$$

$$S(y' - Y')^2 = 1\,330\,877 - \frac{8359^2}{60} - \frac{[21\,353 - 150 \cdot 8359 : 60]^2}{450 - 150^2 : 60} = 163\,563$$

$$S(y'' - Y'')^2 = 3\,277\,086 - \frac{12\,230^2}{60} - \frac{[31\,331 - 150 \cdot 12\,230 : 60]^2}{450 - 150^2 : 60} = 776\,591$$

$$\sigma_d^2 = \frac{163\,563 + 776\,591}{60 + 60 - 4} \cdot \left(\frac{1}{75} + \frac{1}{75}\right) = 216{,}13$$

$$\sigma_d = \sqrt{216{,}13} = 14{,}701$$

$$t = \frac{10{,}08 - 6{,}07}{14{,}701} = 0{,}273$$

In der Tafel findet man, für $n = 116$

$$t_{0,05} = 1{,}98 \qquad t_{0,01} = 2{,}617$$

Der Unterschied zwischen den beiden Regressionskoeffizienten ist nur zufällig (Wahrscheinlichkeit grösser als 0,05). Der neue Tarif hat also den jährlichen Zuwachs des Energieverbrauches nicht wesentlich beeinflusst. Dagegen ist der Unterschied zwischen den beiden Durchschnitten $\bar{y}'$ und $\bar{y}''$ stark gesichert, wie die nachfolgende Prüfung zeigt.

$$t = \frac{\bar{y}'' - \bar{y}'}{\sigma_d}\sqrt{\frac{N'\,N''}{N' + N''}}$$

$$\sigma_d^2 = \frac{1}{N' + N'' - 2}\left[S(y' - \bar{y}')^2 + S(y'' - \bar{y}'')^2\right]$$

$$\sigma_d^2 = \frac{1}{60 + 60 - 2}\,[166\,642 + 784\,204] = 8058{,}2$$

$$\sigma_d = \sqrt{8058{,}2} = 89{,}767$$

$$t = \frac{204 - 139}{89{,}767} \cdot \sqrt{\frac{60^2}{60 + 60}} = 3{,}966$$

für

$n = 118$, $t_{0{,}05} = 1{,}98$; $t_{0{,}01} = 2{,}617$; $t_{0{,}001} = 3{,}373$

Mit andern Worten: die Wahrscheinlichkeit für einen solchen Unterschied ist so klein ($< 0{,}001$), dass kein Zufall vorliegen kann.

Der Tarifwechsel hatte demnach eine sinnfällige, plötzliche Zunahme des Energieverbrauches zur Folge.

C. Das Prüfen der einfachen Regression

Im Verlaufe einer Tarifstudie stellte sich die Aufgabe, die Beziehung zwischen dem Einnahmenausfall und der reduzierten Bodenfläche der gewerblich benützten Räume bei einer Anzahl Gewerbetreibenden zu prüfen. Unter Einnahmenausfall ist hier der Unterschied zwischen der Rechnung für den Lichtkonsum nach dem alten kWh-Tarif und dem um den Grundpreis für die Wohnräume erhöhten Arbeitspreis nach dem neuen Grundpreistarif zu verstehen.

Die reduzierte Bodenfläche der gewerblichen Räume ist gleich ihrer wirklichen Bodenfläche, multipliziert mit einem Faktor, der von der Beleuchtungsstärke abhängt, die für die Arbeit in diesen Räumen erforderlich ist. Die Räume sind in 4 Kategorien eingeteilt, für die die Faktoren gleich 1, ½, ¼ oder ⅛ sind.

Die Stichprobe umfasst 86 Abonnenten.

Wir setzen

y = Einnahmenausfall in Fr.
x = gesamte reduzierte Fläche der Räume eines Abonnenten, in m^2
$N = 86$.

Da die x-Werte zwischen 1 und 72 variieren, werden Klassen von der Breite 3 gemacht. Die Klassenmitten x_j sind dann 2, 5, 8 ...

Zuerst werden die y-Werte in eine Tabelle (V) gruppiert, die zugleich die Quadratwerte zu berechnen erlaubt.

Tabelle V

$x_j =$	2		5		8		
	y_j	y_j^2	y_j	y_j^2	y_j	y_j^2	
	39	1521	22	484	60	3 600	
	19	361	75	5 625	48	2 304	
	.	.	.	.	.	.	
	.	.	.	.	.	.	
	.	.	.	.	.	.	
$S =$	53	2259	411	24 607	1139	114 821	

Von dieser Tabelle gehen wir zur Tabelle VI über, deren Kolonnen in der nachfolgenden Rechnung wieder vorkommen werden.

b) *Bestimmtheitsmass*

$$B = \frac{[S(x-\bar{x})(y-\bar{y})]^2}{S(x-\bar{x})^2\, S(y-\bar{y})^2} = \frac{b\, S(x-\bar{x})(y-\bar{y})}{S(y-\bar{y})^2}$$

$$S(y-\bar{y})^2 = S y_j^2 - \frac{1}{N}(S y_j)^2 =$$

$$= 3461123 - \frac{1}{86}\, 12077^2 = 1765147$$

$$B = \frac{9{,}83 \cdot 139723}{1765147} = 0{,}778 = 77{,}8\,\%$$

c) *Prüfung des Regressionskoeffizienten*

Frage: ist b stark von Null verschieden?

Tabelle VI

x_j	f_j	$f_j x_j$	x_j^2	$f_j x_j^2$	$\overset{j}{S} y_j$	$(\overset{j}{S} y_j)^2$	$\overset{j}{S} y_j^2$	$\frac{1}{f_j} (\overset{j}{S} y_j)^2$	$x_j \overset{j}{S} y_j$
2	5	10	4	20	53	2 809	2 259	561,80	106
5	12	60	25	300	411	168 921	24 607	14 076,75	2 055
8	18	144	64	1 152	1 139	1 297 321	114 821	72 073,39	9 112
11	10	110	121	1 210	1 154	1 331 716	185 190	133 171,60	12 694
14	7	98	196	1 372	622	386 884	65 336	55 269,14	8 708
17	6	102	289	1 734	924	853 776	169 052	142 296,00	16 708
20	8	160	400	3 200	1 302	1 695 204	229 012	211 900,50	26 040
23	6	138	529	3 174	1 174	1 378 276	259 844	229 712,67	27 002
26	3	78	679	2 037	682	465 124	167 426	155 041,33	17 732
29	1	29	841	841	396	156 816	156 816	156 816,00	11 484
32	3	96	1 024	3 072	1 006	1 012 036	341 860	337 345,33	32 192
35	2	70	1 225	2 450	823	677 329	396 125	338 664,50	28 805
38	2	76	1 444	2 888	633	400 689	200 889	200 344,50	24 050
59	1	59	3 481	3 481	618	381 924	381 924	381 924,00	36 462
62	1	62	3 844	3 844	329	108 241	208 241	108 241,00	20 398
71	1	71	5 041	5 041	811	657 721	657 721	657 721,00	57 581
—	86	1 363	—	35 816	12 077	—	3 461 123	3 225 159,51	331 129
	$S f_j = N$	$S f_j x_j = S x$		$S f_j x_j^2 = S x^2$	$\overset{M}{S} \overset{j}{S} y_j = S y$		$\overset{M}{S} \overset{j}{S} y_j^2 = S y^2$	$\overset{M}{S} \frac{1}{f_j} (\overset{j}{S} y_j)^2$	$\overset{M}{S} (x_j \overset{j}{S} y_j) = S x y$

a) *Regressionsgleichung*

$$Y = \bar{y} + b\,(x-\bar{x})$$

$$\bar{y} = \frac{1}{N} S f_j y_j = \frac{1}{86} \cdot 12077 = 140{,}43$$

$$\bar{x} = \frac{1}{N} S f_j x_j = \frac{1}{86} \cdot 1363 = 15{,}85$$

$$b = \frac{S(x-\bar{x})(y-\bar{y})}{S(x-\bar{x})^2}$$

$$S(x-\bar{x})(y-\bar{y}) = S x_j y_j - \frac{1}{N}(S f_j x_j)(S y_j) =$$

$$= 331129 - \frac{1}{86}\, 1363 \cdot 12077 = 139723$$

$$S(x-\bar{x})^2 = S f_j x_j^2 - \frac{1}{N}(S f_j x_j)^2 =$$

$$= 35816 - \frac{1}{86}\, 1363^2 = 14214$$

$$b = \frac{139623}{14214} = 9{,}83$$

$$Y = 140{,}43 + 9{,}83\,(x - 15{,}85) \qquad Y = 9{,}83\, x - 15{,}37$$

Wir setzen

$$t = \frac{b}{s_b} \sqrt{S(x-\bar{x})^2} \text{ wobei } s_b^2 = \frac{1}{N-2}\, S(y-Y)^2$$

$$S(y-Y)^2 = S(y-\bar{y})^2 - S(Y-\bar{y})^2 =$$

$$= S y_j^2 - \frac{1}{N}(S y_j)^2 - \frac{\left[S x_j y_j - \frac{1}{N} S x_j\, S y_j\right]^2}{S f_j x_j^2 - \frac{1}{N}(S f_j x_j)^2} =$$

$$= 1765147 - \frac{139723^2}{14214} =$$

$$= 1765147 - 1373471 = 391676$$

$$s_b^2 = \frac{1}{84} \cdot 391676 = 4662{,}8 \qquad s_b = 68{,}285$$

$$t = \frac{9{,}83}{68{,}285} \cdot \sqrt{14214} = 17{,}163^{**}$$

Für den Freiheitsgrad

$$n = N - 2 = 84$$

ist $\qquad t_{0,05} \cong 1{,}99 \quad$ und $\quad t_{0,01} \cong 2{,}64$

Da der für t gefundene Wert wesentlich grösser als $t_{0,01}$ ist, kann die Regression als stark gesichert angesehen werden.

d) *Streuungszerlegung*

Die Streuungszerlegung beruht auf dem Umstand, dass die Summe der Quadrate der Abweichungen gleich ist der Summe der Quadrate der durch die kontrollierbaren Ursachen bewirkten Abweichungen, vermehrt um die Summe der Quadrate der den zufälligen Ursachen zuzuschreibenden Abweichungen (Reststreuung). Jede dieser Summen ist mit einem bestimmten Freiheitsgrad behaftet. Der Vergleich der Durchschnittsquadrate (Quotient aus der Quadratsumme und dem Freiheitsgrade) mit der Reststreuung zeigt, ob die Wirkung der betrachteten Ursache gesichert oder zufällig ist.

In unserem Falle kann die Gesamtstreuung zerlegt werden in eine Streuung zwischen den Klassen und eine Streuung innerhalb der Klassen (Reststreuung). Die Streuung zwischen den Klassen besteht ihrerseits aus zwei Streuungen: die Streuung aus der Regression (Streuung der Punkte der Regressionsgeraden um den Durchschnitt) und die Streuung der Klassendurchschnitte gegenüber der Regressionsgeraden. So ergibt sich die Tabelle VII, wo M gleich der Anzahl Klassen der Veränderlichen x ist (also 24).

Tabelle VII

Ursache der Streuung	Summe der Quadrate	Freiheitsgrade
Lineare Regression	$Sf_j(Y_j - \bar{y})^2 = SC_r$	1
Abweichung von der Regressionsgeraden	$Sf_j(\bar{y}_j - Y_j)^2 = SC_d$	$M - 2$
Total zwischen Klassen	$Sf_j(\bar{y}_j - \bar{y})^2 = SC_c$	$M - 1$
Rest (innerhalb der Klassen)	$S(y - \bar{y}_j)^2 = SC_e$	$N - M$
Insgesamt	$S(y - \bar{y})^2 = SC_t$	$N - 1$

Die Quadratsummen SC_t und SC_r sind bereits bekannt. Es genügt nun, eine der beiden Summen SC_e oder SC_d zu berechnen, um die andere zu ermitteln, denn $SC_r + SC_d = SC_c$ und $SC_c + SC_e = SC_t$. Die am einfachsten zu ermittelnde ist SC_e, denn

$$S(y - \bar{y}_j)^2 = S\,y^2 - S\frac{1}{f_j}(S\,y_j)^2 =$$
$$= 3461123 - 3225160 = 235963$$

Die Streuungszerlegung erhält folgende Form (Tabelle VIII):

Die Prüfung von $\frac{CM_c}{CM_e}$ zeigt, dass die Korrelation zwischen y und x stark gesichert ist, unabhängig von jeder Hypothese über den Charakter dieser Abhängigkeit.

Nach der Prüfung von $\frac{CM_d}{CM_e}$ sind die Abweichungen von der Regressionsgeraden schwach gesichert, so dass die Regression nur in erster Annäherung als linear angenommen werden kann.

e) *Auslegung*

Die Regressionsformel besagt, dass der Einnahmenausfall im Durchschnitt Fr. 9.83 pro m^2 reduzierter Bodenfläche beträgt, dass aber von diesem Ergebnis ein fester Betrag von Fr. 15.37 abgezogen werden muss.

Das Bestimmtheitsmass lässt erkennen, dass die Korrelation verhältnismässig eng ist, denn rund 78 % der Variabilität des Einnahmenausfalls lässt sich durch die Variabilität der reduzierten Bodenfläche erklären.

Die Prüfung des Regressionskoeffizienten zeigt, dass dieser stark verschieden von Null ist. Die Regression ist somit stark gesichert.

Den gleichen Schluss ergibt die Streuungszerlegung, aus der zudem hervorgeht, dass die Variabilität der Abweichungen gegenüber der Regressionsgeraden nur schwach gesichert ist. Es kann also in erster Annäherung angenommen werden, dass die Regression linear ist. Wollte man noch weiter gehen, so müsste nun der Grad der Regressionsgleichung ermittelt werden.

D. *Die mehrfache Regression*

Im vorhergehenden Beispiel wurde vorläufig angenommen, dass der Grundpreis für die Wohnräume gegeben sei, und dass der Einnahmenausfall in keinerlei Beziehung zu ihm stehe. Dies ist aber nicht der Fall, so dass nun untersucht werden muss, wie der Einnahmenausfall sowohl von den Wohnräumen (Anzahl) als auch von den gewerblichen Räumen (reduzierte Bodenfläche) abhängt. Die Gleichung lautet:

$$Y = a + b_1 x_1 + b_2 x_2$$

wobei

Y = Einnahmenaufsall in Fr.
x_1 = Anzahl der Wohnräume
x_2 = reduzierte Bodenfläche in m^2

Tabelle VIII

Streuung	Freiheitsgrade n	Summe der Quadrate SC	Durchschnittsquadr. CM	$F = \frac{CM}{CM_e}$
Regression	1	1 373 471	729 634 (CM_r)	—
Abweichung von der Regression	22	155 713	7 078 (CM_d)	1,8597*
Total zwischen Klassen	23	1 529 184[1])	69 095 (CM_c)	18,154**
Rest	62	235 963	3 806 (CM_e)	—
Insgesamt	85	1 765 147	—	—

Den Tafeln entnimmt man für

$$n_1 = 22 \quad \text{und} \quad n_2 = 60$$
$$F_{0,05} = 1,70 \quad \text{und} \quad F_{0,01} = 2,12$$

[1]) Diese Zahl kann mit der Formel für SC_c kontrolliert werden.

$$S\,f_j(\bar{y}_j - \bar{y})^2 = S\,f_j\,\bar{y}^2 - \frac{1}{N}(S\,f_j\,\bar{y}_j)^2 = \overset{M}{S}\frac{1}{f_j}(\overset{j}{S}\,y_j)^2 - \frac{1}{N}(\overset{j}{S}\,y_j)^2$$

Die Ausrechnung ergibt $SC_c = 3\,225\,160 - \frac{1}{86} \cdot 12\,077^2 = 1\,529\,184$.

a, b_1 und b_2 = zu bestimmende Konstanten
b_1 und b_2 = Koeffizienten der mehrfachen Regression.

Die Gleichung kann geschrieben werden:

$$Y - \bar{y} = b_1 (x_1 - \bar{x}_1) + b_2 (x_2 - \bar{x}_2)$$

Für die Berechnungen wird vorteilhaft eine Tabelle aufgestellt, die folgende Werte enthält:

a) die Veränderlichen x_1, x_2, y und ihre Summe T,
b) die Quadrate der Veränderlichen und ihre Summen,
c) die Produkte der Veränderlichen je zu zwei,
d) die Summen jeder Kolonne.

Mit diesen Werten können folgende Quadratsummen berechnet werden

$$S_1 = S (x_1 - \bar{x}_1)^2 = S x_1^2 - \frac{1}{N} (S x_1)^2$$

$$S_2 = S (x_2 - \bar{x}_2)^2 = S x_2^2 - \frac{1}{N} (S x_2)^2$$

$$S_3 = S (y - \bar{y})^2 = S y^2 - \frac{1}{N} (S y)^2$$

$$S_4 = S (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) = S x_1 x_2 - \frac{1}{N} (S x_1)(S x_2)$$

$$S_5 = S (x_1 - \bar{x}_1)(y - \bar{y}) = S x_1 y - \frac{1}{N} (S x_1)(S y)$$

$$S_6 = S (x_2 - \bar{x}_2)(y - \bar{y}) = S x_2 y - \frac{1}{N} (S x_2)(S y)$$

$$S_7 = S (T - \bar{T})^2 = S T^2 - \frac{1}{N} (S T)^2$$

Die letzte Summe dient der Kontrolle, denn es lässt sich leicht zeigen, dass

$$S_1 + S_2 + S_3 + 2 (S_4 + S_5 + S_6) = S_7$$

Für b_1, b_2 und B ergibt sich

$$b_1 = \frac{S_6 \cdot S_4 - S_5 \cdot S_2}{S_4^2 - S_1 \cdot S_2}$$

$$b_2 = \frac{S_5 \cdot S_4 - S_6 \cdot S_4}{S_4^2 - S_1 \cdot S_2}$$

$$B = \frac{b_1 \cdot S_5 + b_2 \cdot S_6}{S_3}$$

(Fortsetzung folgt)

Zähler mit Maximumanzeiger mit zwei Messwerken

Von *A. Berner, H. Feuz* und *Service de l'électricité*, Neuenburg

621.317.785

Die Verfasser beschreiben einen von ihnen entwickelten Zähler mit zwei Messwerken, die abwechslungsweise auf Ende einer Abrechnungsperiode in Betrieb gesetzt werden können.

Les auteurs décrivent un compteur de leur conception comportant deux cadrans entrant alternativement en fonction à la fin d'une période de facturation.

Der wesentlichste Nachteil der heute üblichen Zähler mit Maximumanzeiger besteht in der Notwendigkeit, nach jeder Ablesung den Zeiger auf Null zurückzustellen. Wird die Rückstellung anlässlich der Ablesung der Zähler vorgenommen, so verbleibt dem Werk, im Falle von Reklamationen, kein Beweismittel. Zwar könnte die Ablesung in Anwesenheit des Abonnenten vorgenommen werden, doch würde dadurch für den Zählerableser viel Zeit verlorengehen; wären viele derartige Zähler abzulesen, so würde sich dies auf die Arbeit des Zählerablesers nachteilig auswirken. Da zudem in vielen Energielieferungsverträgen, in denen solche Zähler vorgesehen sind (namentlich bei Zweigliedtarifen), monatliche Abrechnung vereinbart ist, müssten bei allen diesen Abonnenten die Zähler am letzten Tag des Monates abgelesen werden, was praktisch unmöglich ist. Wohl gibt es Zähler, die die Maximalleistung graphisch aufzeichnen, wie z. B. der «Maxigraph» und der «Printo-Maxigraph», doch sind diese vielfach im Verhältnis zur Bedeutung des Energielieferungsvertrages zu teuer.

Der Zähler mit doppeltem Messwerk, den Fig. 1 zeigt (Schweizer Patent Nr. 294 742), hilft diesen Nachteilen ab. Er besitzt zwei Messwerke, die abwechslungsweise in Betrieb gesetzt werden können; beide umfassen:

— einen Maximalleistungsanzeiger (kW),
— zwei Zählwerke, je für Hoch- und Niedertarif,
— ein Schauzeichen, das angibt, ob das Messwerk in Betrieb ist oder nicht.

Nehmen wir an, dass es sich um monatliche Ablesungen handelt, dass wir uns Mitte Juni befinden und dass die Kilowattstunden und Kilowatt auf dem oberen Messwerk «S» registriert werden. Das untere Messwerk «I» ist seit dem 31. Mai ausser Betrieb, was durch eine Signalplatte «Halt» angezeigt wird. Das nicht in Betrieb stehende Messwerk «I» gibt die kWh- und kW-Stände am 31. Mai wieder.

Am 30. Juni, um Mitternacht, wird die Registrierung mittels Fernsteuerung oder durch eine Umschaltuhr vom Messwerk «S» auf das Messwerk «I» umgeschaltet. Im Augenblick der Umschaltung wird der Stand des Maximumanzeigers «I» automatisch auf Null gestellt, die Signalplatte «Halt» des Messwerkes «I» verschwindet und erscheint dafür beim Messwerk «S», das nunmehr stillsteht. Vom 1. Juli um 00.00 Uhr bis zum 31. Juli um Mitternacht erfolgt die Registrierung durch das Werk «I».

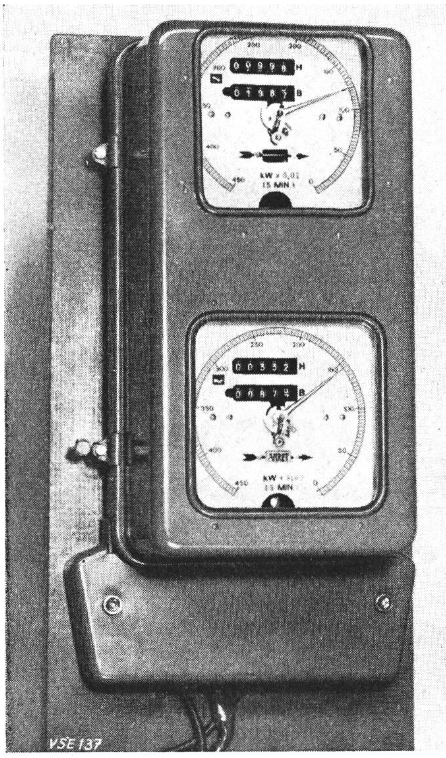

Fig. 1
Zähler mit Maximumanzeiger mit zwei Messwerken

Auf diese Weise bleiben die kW- sowie die kWh-stände für Hoch- und Niedertarif, die sich am 30. Juni um Mitternacht ergaben, auf dem Messwerk «S» während des ganzen Monates Juli stehen. Der Zählerableser hat also einen ganzen Monat Zeit, um die Zählerstände abzulesen; auch der Abonnent hat so lange Gelegenheit, um seinen Zähler zu kontrollieren.

Die Möglichkeit, die kWh-Stände an einem bestimmten Tag und zu einer bestimmten Stunde blockieren zu können, is besonders wichtig bei den tarifmässig bedingten Saisonumschaltungen, namentlich bei Grossabonnenten.

Ist im Energielieferungsvertrag vereinbart, dass die Maximalleistung nur während bestimmter Stunden, z. B. während der Spitzenzeiten, registriert werden soll, kann die Aufzeichnung der Maximalleistung auf dem registrierenden Messwerk während der übrigen Zeit ausser Betrieb gesetzt werden.

Sämtliche Befehle, wie

— Umschaltung der Messwerke,
— Übergang vom Hoch- zum Niedertarif und umgekehrt,
— In- und Ausserbetriebsetzung der Registrierung der Maximalbelastung

können ohne Schwierigkeiten durch Fernsteuerung oder durch Schaltuhren ausgeführt werden; bestehende Schaltuhren können ohne weiteres mit einer zusätzlichen Vorrichtung für die monatliche Umschaltung ausgerüstet werden.

Wirtschaftliche Mitteilungen

Unsere Kraft — Die Elektrizität

Unter diesem Titel ist im Orell-Füssli-Verlag Zürich eine Broschüre von Josef Jaeger über die schweizerische Elektrizitätswirtschaft erschienen. Das Büchlein richtet sich in erster Linie an die breite Oeffentlichkeit. Der Verfasser hat sich zum Ziele gesetzt, die hauptsächlichsten Probleme unserer Elektrizitätswirtschaft auf allgemein verständliche Weise darzulegen und damit einen anschaulichen Einblick in das Wesen unserer Elektrizitätswirtschaft zu vermitteln. Von aktueller Bedeutung dürften besonders die Ausführungen über die Frage des Stromexportes sowie über die Beziehungen zwischen Elektrizitätswirtschaft und Heimatschutz sein. Die Broschüre ist von W. Bär reich illustriert worden.

Verbandsmitteilungen

Eidgenössische Mass- und Gewichtskommission

Im Jahresbericht des VSE (Bull. SEV 1954, Nr. 11, Seite 427/15) wurde irrtümlich mitgeteilt, dass der inzwischen verstorbene Herr Direktor Thorens zum Präsidenten der eidgenössischen Mass- und Gewichtskommission gewählt worden sei. Wir bitten unsere Leser, diesen Irrtum zu entschuldigen.

Nun hat seither der Bundesrat zum neuen Präsidenten dieser Kommission, an Stelle von Herrn Prof. Dr. Paul Joye, Fribourg, der auf Ende 1953 zurückgetreten ist, Herrn Dir. Dr. h. c. *Karl Bretscher*, Delegierter des Verwaltungsrates der Winkler, Fallert A.-G., Bern, gewählt.

Jubiläumsfonds ETH 1955
Spende der Elektrizitätswerke

Die Sammlung, die der VSE bei seinen Mitgliedern zugunsten des Jubiläumsfonds ETH 1955 durchführt, hat bis heute ein sehr erfreuliches Ergebnis gezeitigt. Es sind rund 200 Spenden im Gesamtbetrage von rund Fr. 390 000.— eingegangen. Darunter sind 3 Spenden zu je Fr. 50 000.—, 10 Spenden zwischen Fr. 10 000.— und Fr. 30 000.— und 29 Spenden zwischen Fr. 1000.— und Fr. 7000.— zu erwähnen. Allen Spendern sei bestens gedankt.

Die Sammlung geht weiter. Einzahlungen sind auf Postcheckkonto VIII 4417 Zürich erbeten.

Redaktion der «Seiten des VSE»: Sekretariat des Verbandes Schweizerischer Elektrizitätswerke, Seefeldstrasse 301, Zürich 8, Telephon (051) 34 12 12, Postcheckkonto VIII 4355, Telegrammadresse: Electrunion, Zürich.

Redaktor: *Ch. Morel,* Ingenieur.

Sonderabdrücke dieser Seiten können beim Sekretariat des VSE einzeln und im Abonnement bezogen werden.