Zur analytischen Statistik.

Von

Willers Jessen.

An die Beobachtungen von Pettenkofer und Buhl hat bekanntlich Ludw. Seidel analytisch-statistische Rechnungen geknüpft, durch welche er die Annahme, dass zwischen Grundwasserstand, Regenmenge und Typhusmorbilität ein causaler Zusammenhang bestehe, mathematisch begründet hat. Zu diesem Behufe hat er sich einer ihm eigenthümlichen Berechnungsweise bedient, welche ohne Zweifel ihre Vorzüge hat, ja unter Umständen vielleicht allein anwendbar ist. Aber im Allgemeinen scheint doch die ältere Methode[1]) vorzüglicher, welche von Poisson herrührt und von Gavarret empfohlen wurde. Es mag daher nicht überflüssig sein, zu zeigen, dass sich die von Seidel gelösten Probleme auch nach dieser Methode behandeln lassen.

Seidel[2]) hat die Typhusmorbilität und den durchschnittlichen Stand des Grundwassers für Jedes der Jahre 1856—64 zusammengestellt. Wenn man unter diesen 9 Jahren die vier nassesten (in denen das Grundwasser am höchsten stand), und die vier trockensten heraussucht, so bleibt noch Eins übrig, welches gerade die Mitte zwischen den nassen und den trockenen Jahren hält. Addirt man nun alle Typhusfälle, welche in den 4 trockenen Jahren vorkamen, und legt man dazu die Hälfte der Fälle, welche in dem mittleren

---

[1]) Poisson, Lehrbuch der Wahrscheinlichkeitsrechnung. Deutsch von Schnuse. Braunschweig 1841. Gavarret, Principes généraux de statistique médicale. Paris 1840. Vgl. auch Fick, Medicinische Physik. Braunschweig 1866. Anhang.

[2]) Diese Zeitschrift Bd. I p. 225. Tab. 4.

Oben wurde für die nassen Jahre der Procentsatz 37 gefunden. Wenn man diese Zahl in der ersten Vertikalreihe der Tafel aufsucht und dann die entsprechende Horizontalreihe durchläuft, so findet man 746 als die Zahl, welche der gegebenen Gesammtzahl der Typhusfälle (882) am nächsten kommt. Geht man von 746 vertikal aufwärts oder abwärts, so findet man in der obersten oder untersten Horizontalreihe 5 % als den wahrscheinlichen Fehler, welcher zu allen Zahlen dieser Vertikalreihe gehört. Wenn man also gerade 746 Krankheitsfälle gehabt hätte, von welchen gerade 37 % in die nassen Jahre gefallen wären, so dürfte man sagen, dass in nassen Jahren wenigstens 37 weniger 5 und höchstens 37 plus 5, also 32 bis 42 % sämmtlicher Krankheitsfälle vorkämen. Annähernd dasselbe gilt aber auch, wie aus der Tabelle ersichtlich, für die wirklich gegebenen Zahlen, obgleich sowohl die Gesammtzahl der Fälle (882) von 746, als auch strenge genommen der Procentsatz etwas von 37 % abweicht. Ueberhaupt wird man mit der Tabelle stets ausreichen[1]), wenn man nur nach Procenten und nicht etwa noch Promille oder noch schärfer rechnen will, was in der Regel überflüssig sein wird. Uebrigens nimmt die Tabelle keine Rücksicht auf weniger als 200 gegebene Krankheitsfälle. Gavarret's Tabelle fängt sogar erst mit der Zahl 300 an; es mag richtig sein, dass man auf kleinere Zahlen eigentlich gar keine statistischen Rechnungen gründen sollte.

Es hat sich also herausgestellt, dass 32—42 % sämmtlicher Krankheitsfälle in die nassen, demnach 58—68 % in die trockenen Jahre fallen. Man kann 212 gegen 1 wetten, dass dieses Resultat richtig sei, dass also eine bedeutende Mehrzahl, wenigstens 58 % der Fälle in trockenen Jahren vorkommen. Aber diese Resultate gelten erstens nur für die Münchener Verhältnisse und zweitens nur für solche Jahre, welche ungefähr denselben Grad von Grund-

[1]) Wenn man mit dem Quadrate eines wahrscheinlichen Fehlers in die Zahlen der zweiten Vertikalreihe dividirt, so erhält man die dazu gehörige Gesammtzahl der Fälle. Sucht man z. B. letztere Zahl für den Procentsatz 37 und den Fehler $4\frac{1}{2}$, so erhält man $\frac{18648}{(4\frac{1}{2})^2} = 921$. Die Tabelle liesse sich also, wenn es nöthig wäre, leicht erweitern.

nässe besitzen, wie die in Rechnung gezogenen. Erst längere und an vielen verschiedenen Orten angestellte Beobachtungen können allgemeiner gültige Procentsätze ergeben.

Wenn man mit den Monaten[1]) ähnlich verfährt, wie mit den Jahren, so erhält man $4^1/_2$ (relativ) nasse und $4^1/_2$ trockene Januare, Februare u. s. f. Addirt man die Krankheitsfälle einerseits der nassen, andrerseits der trockenen gleichnamigen Monate, so ergeben sich die in Tab. 2 zusammengestellten Summen.

Tab. 2.

	Nass	Trocken
Januar	49	74
Februar . . .	33	76
März	26	36
April	13,5	33,5
Mai	15,5	31,5
Juni	25	29
Juli	20	23
August	25,5	35,5
September . .	13	25
October	21	47
November. . .	48,5	60,5
December. . .	37,5	85,5
Summe	327,5	556,5

Durch dieses Verfahren wird ebensowohl, wie durch das von Seidel geübte, der etwaige Einfluss der Jahreszeiten, aber nicht zugleich die Zahl der Krankheitsfälle eliminirt. Man sieht daher hier unmittelbar, dass trockene Monate stets mehr, ja meistens viel mehr Typhus liefern, als ebensoviele, gleichnamige nasse. Die Gesammtzahl aller Fälle ist hier 884, wovon $63\,^0/_0$ in trockene Monate fielen. Das Resultat ist also dasselbe wie bei den Jahren.

Wenn man in ähnlicher Weise je drei Monate zusammenfasst, so erhält man in Summa für die nassen Monate 195, für die mittleren 284, für die trockenen 405 Krankheitsfälle. Diese Zahlen lassen sich in verschiedener Weise mit einander vergleichen; das

[1]) Seidel l. c. Bd. II. p. 222. Tab. 1 u. 2.

bemerkenswertheste Resultat dürfte folgende Rechnung geben. Die mittlere und trockene Gruppe enthalten zusammen 72 Monate mit 689 Fällen, also auf 36 Monate 344 Fälle. Auf ebenso viele Monate fallen in der nassen Gruppe 195 Fälle. Von der Summe ($344 + 195 = 539$) kommen demnach 36% auf die letztere Gruppe. Auf ähnliche Weise erhält man im Vergleich mit den beiden übrigen Gruppen für die mittleren Monate 49, für die trockenen 63%, alle drei Procentsätze mit dem wahrscheinlichen Fehler 6. Die mittlere Gruppe hat also nahezu die Hälfte aller Fälle, welche in 72 Monaten oder überhaupt einer gewissen Zeit vorkamen, mit andern Worten die Durchschnittszahl geliefert; in der That ist $\frac{884}{3} = 295$. Diese Gruppe ist somit eine mittlere in Beziehung nicht allein auf Grundnässe, sondern auch auf Morbilität. Den Ausschlag dagegen haben allein die beiden andern Gruppen gegeben und zwar fast genau in demselben Verhältnisse (36 zu 63), wie vorhin (37 zu 63) gefunden wurde.

Untersucht man auf gleiche Weise den Zusammenhang zwischen Typhusmorbilität und Regenmenge, so erhält man[1] $4\frac{1}{2}$ regenarme Jahre mit 522 und eben so viele regenreiche mit 304 Krankheitsfällen. Die Gesammtzahl ist 826, von denen wiederum 63% in die trockenen Jahre fielen. Das Resultat ist abermals dasselbe, auch ist die Uebereinstimmung leicht erklärlich. Unter den 8 Jahren 1856—63 waren nämlich die vier Jahre 56, 57, 58, 63 zugleich grundtrockener, regenärmer und typhusreicher als die vier übrigen. Alle Zahlen würden daher identisch geworden sein, wenn nicht beim Grundwasser das Jahr 64, beim Regen das Jahr 55 noch hinzugenommen wäre.

Zieht man die Monate in gleicher Weise, wie vorhin, in Betracht, so erhält man die in Tab. 3 zusammengestellten Summen.

---

[1]) Seidel, l. c. Bd. II. p. 147. Tab. 2.

Tab. 3.

	Nass	Trocken
Januar	49,5	72,5
Februar. . . .	49,5	58,5
März	25,5	36,5
April	25,5	21,5
Mai	20,5	26,5
Juni.	24	30
Juli	13	21
August	23	33
September . .	13	22
October. . . .	27	38
November. . .	38	55
December. . .	32	79
Summe	340,5	493,5

Der Zusammenhang zwischen Regenmenge und Typhusmorbilität tritt hier ebenfalls deutlich hervor, jedoch, wie auch Seidel fand, weniger entschieden als beim Grundwasser. Durchgehends ergaben zwar die trockenen Monate mehr Krankheitsfälle als die nassen, aber die Differenz ist öfters nur unbedeutend, einmal sogar negativ. In den Summen dagegen tritt das Gesetz wieder scharf hervor. Die Gesammtzahl der Krankheitsfälle ist 834, wovon 41% in die nassen, 59% in die trockenen Monate fielen (w. F. = 5). Letzterer Procentsatz ist allerdings etwas kleiner, als der früher durchweg gefundene (63), aber die Differenz ist zu unbedeutend, um weitere Schlüsse zu gestatten. Diese Behauptung lässt sich durch eine einfache Rechnung beweisen. Die Differenz der beiden Procentsätze (63—59) ist 4, jeder derselben ist mit dem wahrscheinlichen Fehler 5 behaftet. Wenn man nun die beiden Fehler quadrirt, die Quadrate addirt und aus der Summe die Wurzel zieht ($\sqrt{5^2+5^2}$), so erhält man den wahrscheinlichen Fehler der Differenz. Dieser ist hier nahezu 7; die Differenz liegt also zwischen 4—7 und 4+7 oder zwischen — 3 und 11, kann daher möglicher Weise auch Null sein. Wenn dagegen die Differenz grösser wäre, als ihr wahrscheinlicher Fehler, so könnte sie keinenfalls Null sein; dann würde man berechtigt sein, die beiden Procentsätze für wirklich different zu erklären.

Wenn man den Einfluss der Jahreszeiten untersuchen will, so muss man die Monate gleich lang, etwa zu 30 Tagen, annehmen, und die Morbilitätsziffern zuvörderst auf diese Zeit reduciren. In Tab. 4 sind die Monate, nach ihren durchschnittlichen Grundwasserständen geordnet, nebst ihren durchschnittlichen Regenmengen und nebst ihren aus Tab. 2 reducirten Morbilitätsziffern aufgeführt.

Tab. 4.

	Mittlerer Grund-wasser-stand.	Mittlere Regen-menge.	Reducirte Morbilität.
Juli	13,56	46,65	42
Juni	13,69	57,32	54
August	13,73	40,39	59
Mai	14,04	40,32	45
April	14,14	23,73	47
September	14,15	37,17	28
März	14,17	19,19	60
Februar	14,28	10,37	116
October	14,33	18,79	66
November	14,47	24,88	109
Januar	14,54	19,54	119
December	14,63	16,67	119
Summe	—	—	864

Wie man leicht findet, haben die Sommermonate (April bis September) 275, die Wintermonate dagegen 589 Krankheitsfälle, resp. 32 und 68 % der Gesammtzahl geliefert; zugleich sind die letztern [1]) auch grundtrockener und regenärmer als die ersteren. Es lässt sich desshalb nicht ermitteln, ob der Einfluss der Jahreszeiten etwa noch auf einem andern Umstande, als Nässe und Trockenheit, beruht. Der etwas grössere Procentsatz für die trockenen Monate (68 statt der früher durchgängig gefundenen 63) scheint darauf zu deuten, indessen ist der Unterschied zu unbedeutend, um

[1]) Wenn man die abnorme Regenmenge des Nov. 56 aus der Rechnung lässt, so stellt sich die durchschnittliche Regenmenge des November auf 20,29.

beweisend zu sein. Alle vier Jahreszeiten mit einander zu vergleichen, erlaubt die Kleinheit der Morbilitätsziffern nicht.

Die Resultate, welche hier und welche von Seidel erhalten wurden, stimmen durchaus überein, ein praktischer Beweis, dass beide Methoden brauchbar sind. Beide beruhen auch wesentlich auf denselben Grundlagen, doch wird sich Gavarret's Methode oder wenigstens ihre Anwendungsweise am leichtesten populär·darstellen lassen. Gavarret selbst und nach ihm Fick haben dies bereits mit Erfolg versucht; vielleicht ist es gerade jetzt an der Zeit, der analytischen Statistik allgemeineren Eingang zu verschaffen. Mit Recht behauptete Wittstein in der Naturforscher-Versammlung zu Hannover, dass diese Wissenschaft, obwohl sie jetzt kaum existire, in Zukunft Probleme lösen werde, von welchen wir jetzt noch gar keine Ahnung hätten. Die Münchener Untersuchungen sind ein Beleg für die Wahrheit dieser Worte. Wer hätte vor wenigen Jahren geglaubt, dass der Zusammenhang einer Krankheit mit meteorologischen Verhältnissen sich mathematisch würde beweisen lassen? Und doch ist dieser Beweis durch die gemeinsame Thätigkeit der Münchener Forscher jetzt wirklich und unzweifelhaft geliefert worden.