

## Statistische Auswertung von Versuchsergebnissen

Im Beitrag von LANGENEGGER [1] wurde gezeigt, daß mit verhältnismäßig einfachen meßtechnischen Hilfsmitteln möglich ist, eine Kennlinienschar aufzunehmen und daraus eine entsprechende Güllekennzahl (GK) abzulesen. In diesem Aufsatz gilt das besondere Interesse der Beantwortung von zwei Fragen:

1. Ist die Kennzahl GK eine zweckmäßige und hinreichende Größe zur Charakterisierung des technischen Schwierigkeitsgrades, welches ein Güllegemisch bei Pumpvorgängen verursacht und
2. ist mittels der Kennzahl eine genügend gute Erklärung der Leistung von Dickstolppumpen möglich.

Da zur Klärung beider Fragen Schätzfunktionen errechnet werden müssen, erscheint es zweckmäßig, an dieser Stelle die verwendete Methode kurz zu erläutern.

### 1. Die Formulierung eines Regressionsansatzes

Die multiple Regression wird verwendet, um aus einer Menge von beobachteten (abhängigen und unabhängigen) Variablen eine beste funktionelle Darstellung zu erzeugen. Ist  $y$  die abhängige Variable und sind  $x_1, x_2, \dots, x_m$  die unabhängigen Variablen, so stellt die Form

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m$$

eine multiple lineare Regressionsgleichung mit den Regressionskoeffizienten  $b_0, b_1, \dots, b_m$  dar. Diese Regressionskoeffizienten gilt es zu bestimmen. Anders ausgedrückt besteht die Aufgabe darin, für  $y$  eine Schätzfunktion zu errechnen, wobei mit Hilfe der „Methode der kleinsten Quadrate“ die „besten“ Regressionskoeffizienten ermittelt werden.

Selbstverständlich kann eine Regressionsgleichung auch in nichtlinearer Form angesetzt werden; zum Beispiel:

$$Y = b_0 + b_1 x_1 + b_2 x_1^2 + b_3 x_1^3 + \dots + b_k x_k + b_{k+1} x_k^2 + b_{k+2} x_k^3 + \dots$$

Bei der Formulierung eines Regressionsansatzes ergeben sich nicht selten praktische Schwierigkeiten bei der Auswahl der unabhängigen Variablen, insbesondere dann, wenn eine große Anzahl zur Verfügung steht, und zum anderen bei der Wahl der Funktion selbst, da diese nicht immer auf Grund sachlogischer Überlegungen festgelegt werden kann.

Bei der Entscheidung, welche unabhängigen Veränderlichen in die Regressionsgleichung aufgenommen werden sollen, ist es zweckmäßig, die in Tafel 1 ausgewiesene Korrelationsmatrix zu verwenden, da die Korrelationskoeffizienten eine sehr gute qualitative Aussage über Richtung und Stärke des Zusammenhanges zwischen den unabhängigen Variablen und der abhängigen Variablen einerseits und andererseits zwischen den unabhängigen Variablen untereinander darstellen.

Die Korrelationskoeffizienten werden bestimmt nach

$$Y_{xy} = \frac{\sum (xy) - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

n = Anzahl der Beobachtungen  
- 1 ≤ Y<sub>xy</sub> ≤ + 1

Als Kennziffer für die Güte beziehungsweise für den Anpassungsgrad der ausgewählten Funktion wird das multiple Bestimmtheitsmaß verwendet. Es gibt an, wie groß der Anteil des mittleren Differenzenquadrates auf der Regression an der Gesamtvarianz ist, beziehungsweise welcher Anteil der Gesamtvarianz aus der Abhängigkeit der Veränderlichen  $y$  von den unabhängigen Variablen  $x_1, x_2, \dots, x_m$  erklärt werden kann.

Es errechnet sich nach

$$B_{y/1,2,\dots,m} = \frac{\sum (Y-\bar{y})^2}{\sum (y-\bar{y})^2}$$

**Tafel 1: Korrelationskoeffizienten  $r$ , für alle 13 Kombinationen der Variablen  $x_i, i = 1, 2, \dots, 12$  und  $y$**

Die Korrelationskoeffizienten  $r$  0,34 sind signifikant mit einer Fehlerwahrscheinlichkeit von max. 5 %

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$Y$
$x_1$		-0,15	-0,57	-0,09	0,98	-0,17	-0,53	0,01	0,18	0,37	0,42	0,36	0,60
$x_2$			0,79	-0,31	-0,09	0,99	0,73	-0,28	0,38	0,67	0,68	0,56	0,09
$x_3$				0,19	0,50	0,75	0,98	0,19	0,47	0,54	0,49	0,50	-0,00
$x_4$					-0,15	-0,28	0,14	0,97	0,31	0,22	0,04	0,20	0,08
$x_5$						-0,11	-0,46	-0,02	0,24	0,40	0,49	0,43	0,67
$x_6$							0,76	-0,27	0,40	0,67	0,69	0,57	0,11
$x_7$								0,16	0,47	0,55	0,53	0,53	0,04
$x_8$									0,41	0,31	0,17	0,32	0,22
$x_9$										0,71	0,70	0,90	0,61
$x_{10}$											0,96	0,91	0,56
$x_{11}$												0,94	0,64
$x_{12}$													0,69
$Y$													

$m$  = Anzahl der unabhängigen Variablen  
 $y$  = Beobachtungswerte  
 $Y$  = Funktionswerte  
 $\bar{y}$  = Mittelwert  
 $0 \leq B_{j1,2,\dots,m} \leq 1$

Für eine große Anzahl von Variablen und oder für eine große Beobachtungsmenge verlangt die Regressionsanalyse ebenfalls eine große Anzahl von arithmetischen Operationen. Die Auswertung der Unterlagen erfolgte deshalb auf einer elektronischen Datenverarbeitungsanlage.

Bei dem verwendeten Rechnerprogramm werden stufenweise Zwischenresultate bestimmt, die wertvolle statistische Information liefern. Diese als Zwischenergebnisse anfallenden Regressionsgleichungen erhält man, indem schrittweise jeweils eine unabhängige Variable mehr zum Aufbau der Regressionsgleichung benutzt wird, bis schließlich am Ende alle verbraucht sind. Beispielsweise hat somit die Lösungsfolge für einen linearen, multiplen Ansatz mit 3 unabhängigen Veränderlichen folgendes Aussehen:

$$\begin{aligned}
 Y &= b_0 + b_1 x_1 \\
 Y &= b'_0 + b'_1 x_1 + b_2 x_2 \\
 Y &= b''_0 + b''_1 x_1 + b''_2 x_2 + b_3 x_3
 \end{aligned}$$

Dabei kann mittels des gleichzeitig errechneten Bestimmtheitsmaßes die Verbesserung in der Varianzerklärung durch Hinzunahme der zweiten und dritten Variablen direkt abgelesen werden.

## 2. Erklärung der Güllekennzahl GK mittels der einzelnen Güllekomponenten

Zur Beantwortung der Frage, welche unabhängigen Variablen zur Erklärung der Güllekennzahl am besten geeignet sind, dienen die in Tafel 1 errechneten einfachen Korrelationskoeffizienten. Man sieht sofort, daß die Trockenmassen der Rückstände aus den vier Sieben ( $x_5, x_6, x_7, x_8$ ) durchwegs höher mit  $y$  korrelieren als die korrespondierenden feuchten Rückstände ( $x_1, x_2, x_3, x_4$ ) und somit besser zur Schätzung der Flüssigmist-Kennzahl geeignet sind. Dieser Tatbestand wurde durch die Rechenergebnisse auch bestätigt. Von besonders starken Einfluß sind dabei die groben Anteile aus Sieb 1, also die Variablen  $x_1$  beziehungsweise  $x_5$ . Ferner zeigt die Korrelationsmatrix, daß es aufgrund der hohen Interkorrelationen wenig sinnvoll ist, neben den Variablen  $x_1, x_2, x_3, x_4$  gleichzeitig die entsprechenden Trockenmassen  $x_5, x_6, x_7, x_8$  im Regressionsansatz zu verwenden.

Die Koeffizienten liegen hier durchwegs zwischen 0,97 und 0,99.

Für die jetzt im einzelnen angegebenen Schätzfunktionen für die Güllekennzahl konnte durch die Hinzunahme von quadratischen Termen fast immer eine Verbesserung im Anpassungsgrad gegenüber Funktionen mit nur linearen Termen erzielt werden. Als Charakteristikum für die Güte der Anpassung ist das oben definierte multiple Bestimmtheitsmaß mit ausgewiesen.

### 1. Schätzung von $y$ mit unabhängigen Variablen $x_1, x_2, x_3, x_4$ (feuchte Rückstände aus den Sieben 1 bis 4)

$$Y = f(x_1, x_1^2, x_2, x_2^2, x_3, x_3^2, x_4, x_4^2)$$

	$B_{y1 \dots 4}$
(1) $Y = -1,15 + 0,096 \cdot x_1$	0,359
(2) $Y = 10,84 - 0,18 \cdot x_1 + 0,0010 \cdot x_1^2$	0,591
(3) $Y = 6,23 - 0,191 \cdot x_1 + 0,0011 \cdot x_1^2 + 0,068 \cdot x_2$	0,651
(4) $Y = 6,94 - 0,192 \cdot x_1 + 0,011 \cdot x_1^2 + 0,047 \cdot x_2 + 0,00011 \cdot x_2^2$	0,651
(5) $Y = -5,49 - 0,055 \cdot x_1 + 0,00083 \cdot x_1^2 - 0,28 \cdot x_2 + 0,0011 \cdot x_2^2 + 0,15 \cdot x_3$	0,733
(6) $Y = -8,97 - 0,044 \cdot x_1 + 0,00080 \cdot x_1^2 - 0,32 \cdot x_2 + 0,0012 \cdot x_2^2 + 0,23 \cdot x_3 - 0,00024 \cdot x_3^2$	0,737
(7) $Y = -8,77 - 0,028 \cdot x_1 + 0,00077 \cdot x_1^2 - 0,35 \cdot x_2 + 0,0012 \cdot x_2^2 + 0,26 \cdot x_3 - 0,00027 \cdot x_3^2 - 0,032 \cdot x_4$	0,739

$$\begin{aligned}
 (8) \quad Y &= -28,07 - 0,033 \cdot x_1 + 0,00077 \cdot x_1^2 - \\
 &\quad - 0,30 \cdot x_2 + 0,0011 \cdot x_2^2 + 0,16 \cdot x_3 + \\
 &\quad + 0,00012 \cdot x_3^2 + 0,49 \cdot x_4 - 0,0027 \cdot x_4^2 \quad 0,752
 \end{aligned}$$

### 2. Schätzung von $y$ mit den unabhängigen Variablen $x_5, x_6, x_7, x_8$ (Trockenmasse der Rückstände aus den Sieben 1 bis 4, sowie der ausgewaschenen Substanz).

$$Y = f(x_5, x_5^2, x_6, x_6^2, x_7, x_7^2, x_8, x_8^2, x_9, x_9^2)$$

	$B_{y5 \dots 9}$
(9) $Y = -1,73 + 1,02 \cdot x_5$	0,451
(10) $Y = 7,73 - 0,94 \cdot x_5 + 0,067 \cdot x_5^2$	0,646
(11) $Y = 3,78 - 0,95 \cdot x_5 + 0,069 \cdot x_5^2 + 0,60 \cdot x_6$	0,692
(12) $Y = 3,43 - 0,95 \cdot x_5 + 0,069 \cdot x_5^2 + 0,72 \cdot x_6 - 0,0063 \cdot x_6^2$	0,692
(13) $Y = -5,98 + 0,11 \cdot x_5 + 0,049 \cdot x_5^2 - 2,73 \cdot x_6 + 0,099 \cdot x_6^2 + 1,53 \cdot x_7$	0,768
(14) $Y = -9,66 + 0,21 \cdot x_5 + 0,046 \cdot x_5^2 - 3,24 \cdot x_6 + 0,12 \cdot x_6^2 + 2,45 \cdot x_7 - 0,034 \cdot x_7^2$	0,774
(15) $Y = -12,28 - 0,0085 \cdot x_5 + 0,051 \cdot x_5^2 - 2,90 \cdot x_6 + 0,13 \cdot x_6^2 + 1,83 \cdot x_7 - 0,028 \cdot x_7^2 + 0,94 \cdot x_8$	0,789
(16) $Y = -13,60 - 0,012 \cdot x_5 + 0,051 \cdot x_5^2 - 2,86 \cdot x_6 + 0,13 \cdot x_6^2 + 1,76 \cdot x_7 - 0,025 \cdot x_7^2 + 1,34 \cdot x_8 - 0,024 \cdot x_8^2$	0,789
(17) $Y = -33,59 - 0,68 \cdot x_5 + 0,059 \cdot x_5^2 - 1,95 \cdot x_6 + 0,091 \cdot x_6^2 + 0,28 \cdot x_7 + 0,0078 \cdot x_7^2 + 3,44 \cdot x_8 - 0,20 \cdot x_8^2 + 0,65 \cdot x_9$	0,862
(18) $Y = 36,17 - 0,28 \cdot x_5 + 0,051 \cdot x_5^2 - 2,73 \cdot x_6 + 0,11 \cdot x_6^2 + 1,67 \cdot x_7 - 0,034 \cdot x_7^2 + 1,51 \cdot x_8 - 0,12 \cdot x_8^2 - 2,40 \cdot x_9 + 0,033 \cdot x_9^2$	0,892

Das Endtableau — Schätzfunktion (18) — zeigt, daß durch die simultane Berücksichtigung von fünf einigermaßen unabhängigen Variablen (mit geringen Interkorrelationen) ein multiples Bestimmtheitsmaß von 0,892, entsprechend einer Varianzerklärung von 89,2 %, erreicht wird. Somit können 89 % der Unterschiede der Güllekennzahl durch die fünf wesentlichen Flüssigmist-Komponenten erklärt werden. Die groben Feststoffanteile — Trockenmasse aus Sieb 1 — brachten mit 64,6 % Varianzerklärung den höchsten Anteil. Durch die Hinzunahme der Trockenmassen aus Sieb 2, 3 und 4 konnte eine Verbesserung um insgesamt 14,3 % auf 78,9 % erzielt werden, während der letzte Faktor (Trockenmasse der ausgewaschenen Substanz) nochmals einen Anstieg in der Varianzerklärung um 10,3 % auf insgesamt 89,2 % brachte. Die relativ geringe nicht erklärte Restvarianz von knapp 11 % muß nicht erfaßten Einflußgrößen zugeschrieben werden; es ist jedoch durchaus denkbar, daß eine stärkere Fraktionierung in der Absiebung das Ergebnis weiter verbessern könnte. Damit kann, zumindest im Rahmen des vorliegenden Versuchsmaterials, die einfach zu ermittelnde Güllekennzahl als durchaus geeignete Kenngröße für die Zusammensetzung eines Güllegemisches und dem daraus resultierenden Widerstand angesehen werden.

Da eine grafische Darstellung einer Hyperfläche 2. Grades nicht möglich ist, sind die Ergebnisse der Funktion (18) in Tafel 2 wiedergegeben. Dabei sind die Beobachtungswerte für die Güllekennzahl ( $y$ ) den zugehörigen errechneten Schätzwerten ( $Y$ ) gegenübergestellt sowie die absoluten und relativen Differenzen angegeben. Die verhältnismäßig großen Abweichungen im unteren Bereich, der etwa mit  $0 \leq GK \leq 5$  abgegrenzt werden kann, dürften auf der in diesem Bereich schnelleren Entmischung der Gemische während der Messung zurückzuführen sein.

### 3. Erklärung der Pumpenleistung mittels der Güllekennzahl GK

Der Zusammenhang zwischen Pumpenleistung und Kennzahl GK wurde an zwei verschiedenen Dickstoffpumpen untersucht. Dabei wurden einfache lineare Regressionsgleichun-

**Tafel 2: Beobachtungswerte und zugehörige errechnete Schätzwerte für die Güllekennzahl GK**

	Beobachtungswert y	Schätzwert (Funktionswert) Y	Differenz absolut y - Y	Differenz relativ mit y = 100 %
1	2,00	6,17	-4,17	-208,7
2	21,00	22,37	-1,37	-6,5
3	9,00	8,30	0,70	7,8
4	1,50	4,17	-2,67	-178,2
5	0,50	3,85	-3,25	-669,2
6	0,25	2,15	-1,90	-761,3
7	4,0	4,05	-0,05	-1,1
8	11,0	11,45	-0,45	-4,1
9	19,0	21,84	-2,84	-14,9
10	5,0	1,76	3,24	64,8
11	2,0	-1,88 <sup>2)</sup>	3,88	194,2
12	16,0 <sup>1)</sup>	19,55	-3,55	-22,2
13	16,0 <sup>1)</sup>	24,93	-8,93	-55,8
14	13,0	11,25	1,75	13,5
15	1,0	-1,00	2,00	200,0
16	1,20	6,96	-5,76	-480,0
17	11,0	5,35	5,65	51,3
18	3,00	2,37	0,63	21,0
19	2,0	0,66	1,34	67,2
20	2,0	-0,11	2,11	105,5
21	12,0	10,95	1,05	8,8
22	9,0	9,0	0,00	0,0
23	6,0	0,85	5,15	85,8
24	9,0	3,89	5,11	56,7
25	12,0	15,76	-3,76	-31,4
26	9,0	12,69	-3,69	-41,1
27	6,0	8,8	-2,8	-46,7
28	3,0	3,15	-0,15	-5,2
29	56,0	55,95	0,05	0,1
30	52,0	37,32	14,68	28,2
31	21,0	20,01	0,99	4,7
32	14,0	16,32	-2,32	-16,5
33	12,0	12,56	-0,56	-4,6

**Anmerkungen:**

- 1) Für gleiche Beobachtungswerte können sich aufgrund der zugehörigen unterschiedlichen x-Werte voneinander abweichende Funktionswerte ergeben;
- 2) Negative Werte für die Kennzahl GK sind nicht definiert, können jedoch als Funktionswert wegen der größeren Ungenauigkeit im Bereich kleinerer GK-Werte rechenstechnisch entstehen

gen verwendet. Für die Pumpe I mit 18 vorliegenden Beobachtungen ergab sich folgende Schätzfunktion

$$Y_1 = 223,626 - 3,799 \cdot x_1 \text{ [m}^3/\text{h]} \quad (1)$$

mit einem Bestimmtheitsmaß von 0,979; das besagt, daß in diesem Fall die Unterschiede in den gemessenen Leistungen durch die Güllekennzahl GK zu etwa 98 % erklärt werden. (Die Ergebnisse sind im Beitrag LANGENEGGER, Tafel 2, Bild 6 wiedergegeben.)

Da parallel zu den Kennzahlen GK auch die Trockenmassegehalte der Güllegemische festgestellt wurden, konnte gleichzeitig versucht werden die Pumpenleistung durch die Trockenmasse insgesamt zu erklären. Ein linearer Ansatz mit Prozent Trockenmasse insgesamt als unabhängige Variable ergab die Funktion

$$Y_1 = 309,931 - 19,297 \cdot x_2 \text{ [m}^3/\text{h]} \quad (2)$$

mit nur 51,78 % Varianzerklärung.

Durch Hinzufügen eines quadratischen Gliedes konnte mit der Schätzfunktion

$$Y_1 = 226,089 + 20,901 \cdot x_2 - 3,237 \cdot x_2^2 \text{ [m}^3/\text{h]}$$

die Varianzerklärung auf 76,87 % verbessert werden. Sie bleibt jedoch immer noch um etwa 21 % gegenüber Funktion (1) zurück. Damit ist die Kennzahl GK weitaus besser zur Leistungsbestimmung von Dickstoffpumpen geeignet als der bisher vielfach verwendete Trockenmassegehalt.

In einem zweiten Versuch wurde die Pumpe II mit geringeren Förderleistungen eingesetzt. Mit 13 Meßwerten wurde die Regressionsfunktion

$$Y_2 = 63,749 - 1,146 \cdot x_1 \text{ [m}^3/\text{h]}$$

errechnet, wobei die Varianzerklärung mit 95,81 % fast die gleiche Güte erreicht. (Die Ergebnisse sind im Beitrag LANGENEGGER, Tafel 3, Bild 7, wiedergegeben.)

**Schrifttum**

- [1] LANGENEGGER, G.: Eine Meßmethode zur Bewertung der Pumpfähigkeit von Gülle und Flüssigmist. Landtechnische Forschung 18 (1970), S. 55-61

E. Ortmaier

**Informieren und Dokumentieren mit neuen Systemen?**

Die Flut der Informationen ist mit den hergebrachten Systemen nicht mehr zu bewältigen. Immer schwieriger wird es, an der weltweiten Wissenschaftsexpansion zu partizipieren. Dies aber ist notwendig, da der Grad der Informiertheit mit über den technologischen, wirtschaftlichen und sozialen Fortschritt entscheidet. Seit Jahren wird daher nicht nur in der Bundesrepublik Deutschland der Aufbau eines den heutigen und morgigen Erfordernissen gerecht werdenden Informations- und Dokumentationswesens gefordert. Erst unlängst hat der Deutsche Verband technisch-wissenschaftlicher Vereine in einem Memorandum noch einmal nachdrücklich auf die Gesamtbedeutung dieser Aufgabe hingewiesen und zu entsprechenden Aktivitäten gedrängt. Wie nun aus dem Bundesministerium für Bildung und Wissenschaft zu erfahren ist, soll noch in diesem Jahr ein Gesamtprogramm für das technisch-wissenschaftliche Dokumentationswesen entworfen werden. Hierzu soll in Kürze eine Kommission berufen werden.

**Patent- und Lizenzbilanz blieb negativ**

In der deutschen Patent- und Lizenzbilanz hat sich die Schere zwischen Einnahmen und Ausgaben weiter geöffnet. Sie schloß im vergangenen Jahr mit einem Saldo von 620 Mio. DM ab. Im Jahr zuvor betrug die Differenz nur 394 Mio. DM. Größte Lizenz- und Patentnehmer Deutschlands waren wieder die europäischen Länder, die Vereinigten Staaten, Japan und Brasilien. Sie sind mit 85 % an den deutschen Einnahmen beteiligt. Die höchsten Lizenz- und Patentgebühren zahlte Deutschland an die Vereinigten Staaten (nahezu 50 %) und an die Schweiz (24 %). Die größten Einnahmen erzielte die Chemische Industrie. Auf sie entfielen 41 % der Gesamteinnahmen. Mit 31 % der Gesamtausgaben war sie aber auch größter Patent- und Lizenzkunde des Auslandes. Die elektrotechnische Industrie war an den Gesamteinnahmen mit 15,1 % und an den Gesamtausgaben mit 29,9 % beteiligt. Sie hatte damit erneut das größte Defizit. Es betrug im Jahre 1969 rund 242 Mio. DM gegenüber 147 Mio. DM bei der Chemischen Industrie. Im Zuge der internationalen Verflechtung kann aus der Patent- und Lizenzbilanz jedoch nicht mehr auf Rückstand oder Fortschritt der Forschung in einem Lande geschlossen werden. Immerhin zahlten die Unternehmen mit maßgeblich ausländischer Beteiligung 54 % der gesamten deutschen Gebühren.

\*

Das Handwerk erhielt zur Förderung von Forschung und Entwicklung, zur Einrichtung von Ausbildungsstätten und zur überbetrieblichen Lehrlingsunterweisung vom Bundesministerium für Wirtschaft im vergangenen Jahr Zuwendungen von 12 Millionen DM. In diesem Jahr sind für die gleichen Zwecke 14,25 Millionen DM bereitgestellt worden.