

**Zeitschrift:** Mitteilungen des Statistischen Bureaus des Kantons Bern  
**Band:** - (1968)  
**Heft:** 54

**Artikel:** Kostenabhängigkeit in den bernischen Bezirksspitalern = Facteurs influants sur les frais dans les hôpitaux des districts bernois  
**Autor:** [s.n.]  
**DOI:** <https://doi.org/10.5169/seals-850384>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. 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. [Siehe Rechtliche Hinweise.](#)

### **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. [Voir Informations légales.](#)

### **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. [See Legal notice.](#)

**Download PDF:** 12.11.2024

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



Mitteilungen des Statistischen Bureaus des Kantons Bern

Neue Folge



**Nr. 54**

**Kostenabhängigkeit  
in den bernischen Bezirksspitalern**

**Facteurs influants sur les frais  
dans les hôpitaux des districts bernois**

Bern 1968

H 2 111 4



Mitteilungen des Statistischen Bureaus des Kantons Bern

Neue Folge

**Nr. 54**

**Kostenabhängigkeit  
in den bernischen Bezirksspitalern**

**Facteurs influants sur les frais  
dans les hôpitaux des districts bernois**

Bern 1968

# Inhaltsverzeichnis

<b>0</b>	<b>Einleitung</b>	5
<b>1</b>	<b>Theorie der stochastischen Abhängigkeit</b>	5
11	Grundsätzliches	5
111	Funktionale und stochastische Abhängigkeit	5
112	Operationsstufen zur Problemlösung	6
113	Systematik	6
12	Einfache lineare Regression und Korrelation	6
121	Das Modell	6
122	Schätzverfahren	7
122.1	Regressionsparameter	7
122.2	Streuung der Einzelwerte	9
122.3	Bestimmtheitsmass. Korrelationskoeffizient	9
123	Prüfen von Hypothesen	9
123.1	Grundsätzliches	9
123.2	Varianzanalyse	11
123.3	Prüfen der Regressionsparameter	11
123.4	Bestimmtheitsmass. Korrelationskoeffizient	12
123.5	Linearitätstest	13
124	Vertrauensgrenzen	15
124.1	Grundsätzliches	15
124.2	Regressionsparameter	15
124.3	Vertrauensgrenzen für die Regressionswerte	16
125	Zwei Regressionsgerade	16
13	Einfache nichtlineare Regression	18
131	Grundsätzliches	18
132	Transformation auf einfache lineare Regression	19
133	Mehrfache lineare Regression	19
134	Regression mit orthogonalen Polynomen	19
14	Mehrfache lineare Regression und Korrelation	20
141	Allgemeines	20
142	Das Modell	20
143	Schätzverfahren	21
143.1	Regressionsparameter	21
143.2	Bestimmtheitsmass	22
143.3	Korrelationskoeffizient	22
144	Prüfen von Hypothesen	23
144.1	Varianzanalyse	23
144.2	Prüfen der Regressionskoeffizienten	23
144.3	Prüfen der Bestimmtheiten	23
145	Vertrauensgrenzen	24
146	Unechte Variable	24
15	Mehrfache nichtlineare Regression und Korrelation	25
151	Allgemeines	25
152	Transformation	25
153	Mehrfache nichtlineare Regression	25
16	Elektronenrechner – Programme	25
<b>2</b>	<b>Die Kostenabhängigkeit in den bernischen Bezirksspitalern</b>	26
21	Problemlage und Zielsetzung	26
22	Theorie der Kostenabhängigkeit	26
221	Grundsätzliches	26
222	Die abhängige Variable	27
223	Die Einflussfaktoren	27
223.1	Die Produktivität	27
223.2	Die Marktpreise	30
223.3	Der Ausnützungsgrad	30
223.4	Die Aufenthaltsdauer der Patienten	32
223.5	Die Betriebsgrösse	33
223.51	Der Kostenverlauf	33
223.52	Massstab der Betriebsgrösse	33



	223.6	Spezialabteilungen. Behandlungsmethode .....	34
	223.7	Symbolik .....	34
23		Die Erhebungsgrundlagen (Krankenanstaltenstatistik) .....	35
24		Kostenabhängigkeit in den Bezirksspitalern .....	35
	241	Analyse der einzelnen Einflussfaktoren .....	35
	241.1	Produktivität .....	35
	241.2	Marktpreise .....	42
	241.3	Ausnützungsgrad .....	44
	241.4	Aufenthaltsdauer .....	44
	241.5	Betriebsgrösse .....	45
	241.6	Zusammenstellung der Ergebnisse .....	46
	242	Interkantonaler Vergleich .....	46
	243	Mehrfache Regression und Korrelation .....	49
	243.1	Die Variablen .....	49
	243.2	Schätzwerte .....	50
	243.3	Prüfen von Hypothesen .....	52
	243.4	Vertrauensgrenzen der Schätzung .....	55
	243.5	Überblick über die Resultate .....	55
25		Die Ergebnisse der Untersuchung .....	56
		Les principaux résultats de l'étude .....	58
<b>3</b>		<b>Synopsis grundlegender Ansätze .....</b>	<b>61</b>
<b>4</b>		<b>Arbeitsablauf bei linearer Regression .....</b>	<b>65</b>
<b>5</b>		<b>Literaturangaben .....</b>	<b>66</b>

# 0 Einleitung

Bei der wissenschaftlichen Arbeit ist sehr oft die Lage gegeben, dass nicht nur eine einzige Veränderliche vorliegt; zwei oder mehrere Variable sind zu untersuchen, wobei sie nicht getrennt, sondern gleichzeitig beobachtet werden. Es ist meistens so, dass eine Variable (die sogenannte abhängige Variable) von einer oder sogar mehreren Variablen (den unabhängigen Variablen) abhängig ist. So kann es sein, dass die Kosten pro Pflage tag (abhängige Variable) von der Produktivität (unabhängige Variable) beeinflusst werden. Die betriebswirtschaftliche Theorie zeigt ferner, dass die Einheitskosten nicht nur von der Produktivität, sondern noch von anderen Einflussfaktoren abhängig sind, wie z.B. den Marktpreisen und dem Beschäftigungsgrad.

Generell kann also argumentiert werden, dass eine Variable (z.B. Kosten) als Funktion von einer oder mehreren anderen Variablen (z.B. Produktivität, Marktpreise) betrachtet werden kann. Welches ist die praktische Bedeutung dieser Erkenntnis? Nun, man wird unter anderem daran interessiert sein, für bekannte Werte der unabhängigen Variablen fehlende Werte der abhängigen Variablen zu berechnen (Interpolation, Extrapolation). Ferner wird es möglich sein, die verwendete Theorie mit Hilfe der Statistik zu überprüfen, zu verifizieren.

## 1 Theorie der stochastischen Abhängigkeit

### 11 Grundsätzliches

#### 111 Funktionale und stochastische Abhängigkeit

Vorerst soll das Wesen der stochastischen Abhängigkeit kurz erläutert werden. Wir gehen aus vom **Begriff der Funktion**. Nach Dirichlet heisst  $y$  eine Funktion von  $x$ , d.h.  $y = f(x)$ , wenn irgendwelchen Werten von  $x$  eindeutig  $y$ -Werte zugeordnet sind. Die Zuordnung kann u.a. folgendermassen erfolgen:

- Mit **Rechenvorschrift**, und zwar explizit, d.h.  $y = f(x)$ , oder implizit, d.h.  $f(x,y) = 0$ .
- Durch eine **Tabelle**.
- Durch eine **grafische Darstellung**.

R. G. D. Allen (Mathematik ..., Berlin 1956, S.29) umschreibt folgendermassen: «Der Begriff der Funktion umschliesst daher die Begriffe der **Beziehung** zwischen den Werten zweier Veränderlicher und der **Abhängigkeit** einer Veränderlichen von der anderen».

Nun hat sich gerade die Betriebswirtschaftslehre sehr oft mit messbaren Grössen zu befassen, wie z.B. mit Kosten, Produktivität, Beschäftigungsgrad, sowie mit den Beziehungen zwischen diesen Grössen. Die Konzeption des funktionalen Zusammenhangs hat daher sehr grosse praktische Bedeutung.

Die funktionale Abhängigkeit ist also durch eine eindeutige Zuordnung charakterisiert, d.h. zu einem gegebenen Wert der unabhängigen Variablen, z.B.  $x_i$ , ist der Wert der abhängigen Variablen  $y_i$  eindeutig gegeben. Im Gegensatz dazu betrachtet man in der Statistik Wertepaare, bei denen zu jedem  $x$ -Wert der Wert  $y_i$  nicht eindeutig gegeben ist; zu jedem  $x_i$  können zwei oder mehrere  $y_i$  gegeben sein. So werden wir folgendem Fall begegnen:

Kosten in Fr.	Produktivität ( $x_k$ )	
	100	105
$y_{ik}$	42.32	39.44
	56.79	46.49

Bei diesem Beispiel gehören zu jedem der beiden  $x$ -Werte je zwei  $y$ -Werte. Man bezeichnet eine derartige Abhängigkeit, also eine mehrdeutige Zuordnung, als **stochastische** (= mutmassliche) **Abhängigkeit**. Das Problem besteht für den Statistiker darin, an Stelle des Punkteschwarms eine mathematische Funktion zu setzen, d.h. die stochastische Abhängigkeit in eine funktionale Abhängigkeit überzuführen.

## 112 Operationsstufen zur Problemlösung

Zur Lösung des gesamten Problemkreises kann es zweckmässig sein, folgende Operationsstufen zu verwenden:

- (1) Festlegung der **Problemlage** bzw. der **Zielsetzung**;
- (2) **Aufstellung des Modells**:
  - Entwicklung der Hypothese, d.h. Bestimmung der Einflussgrössen bzw. der abhängigen Variablen aus der betriebswirtschaftlichen Theorie (oder Empirie);
  - Definition der Einheiten, in denen die verwendeten Variablen gemessen werden;
  - Hypothese über die Form der Beziehung;
  - Formulierung des Modells in Gleichungsform (= Formulierung des Zusammenhangs) → mathematisches Modell in allgemeiner Form.
- (3) **Datenbeschaffung** bzw. Messung der Variablen: Beobachtung, Zählung, Versuch; Primär- oder Sekundärerhebung.
- (4) **Schätzverfahren**:  
Schätzen der Regressionsparameter; Masszahlen zur Messung der **Stärke der stochastischen Abhängigkeit**: Bestimmtheitsmass und Korrelationskoeffizient.
- (5) **Prüfverfahren** (Tests):
  - Ursachen der **Regressionsstreuung** → Streuungszerlegung;
  - Prüfen der Hypothese, ob überhaupt Abhängigkeit zwischen den Variablen besteht;
  - Prüfen der Hypothese über die Form der Abhängigkeit.
- (6) **Regressionsgleichung**:  
Ersetzung der beobachteten Werte von  $y_i$  durch theoretische Werte  $Y_i$ .
- (7) **Vertrauensgrenzen** der Schätzung. Bestimmung des Fehlers, mit welchem die Schätzwerte behaftet sind (Güte der Schätzung).
- (8) **Deutung** der Ergebnisse:  
Analyse bzw. Interpretation der Resultate.

## 113 Systematik

Die nachfolgenden Untersuchungen basieren auf quantitativen Merkmalen. Je nach der **Zahl** der Veränderlichen unterscheidet man:

- Einfache Regression (zwei Variable)
- Mehrfache Regression (drei oder mehr Variable)

Je nach der **Form** der Regressionslinie unterscheidet man lineare oder nichtlineare Regression (Korrelation).

## 12 Einfache lineare Regression und Korrelation

### 121 Das Modell

Gegeben sei das **mathematische Modell** einer Funktionalbeziehung

$$\lambda = \alpha + \beta x \quad (1)$$

mit der abhängigen Variablen  $\lambda$ . Das entsprechende **statistische Modell** hat die Form

$$y = \alpha + \beta x + \varepsilon \quad (2)$$

wobei  $\alpha$  und  $\beta$  unbekannte Parameter sind,  $\varepsilon$  aber eine Zufallsvariable ist. In der Praxis sind  $N$  Wertepaare  $(x_i, y_i)$ ,  $i = 1, 2, \dots, N$ , gegeben. Diese **Beobachtungen** bilden  $N$  Beziehungen:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (3)$$

Nehmen wir nun an, dass für  $\alpha$  und  $\beta$  die Schätzwerte  $a$  und  $b$  vorliegen. Es ist dann

$$y_i = a + b x_i + e_i \quad (4)$$

mit  $e_i$  als Schätzwert für das unbekannte  $\varepsilon_i$ . Als Schätzwert für  $Y_i$  ergibt sich

$$Y_i = a + b x_i \quad (5)$$

In dieser Gleichung (5) erscheinen die Parameter

$a$  = Niveaunkonstante,

$b$  = Regressionskoeffizient.

## 122 Schätzverfahren

### 122.1 Regressionsparameter

Die Methode der kleinsten Quadrate erlaubt uns, Schätzwerte sowohl für die Niveaunkonstante  $a$  wie auch für den Regressionskoeffizienten  $b$  zu finden. Wir stellen folgende Forderung auf:

$$\sum_{i=1}^N e_i^2 = \sum_{i=1}^N (y_i - Y_i)^2 = \sum_{i=1}^N (y_i - a - b x_i)^2 = \text{Minimum}$$

Zur Erfüllung dieser Forderung müssen  $a$  und  $b$  folgendem Gleichungssystem genügen:

$$\left. \begin{aligned} N a + \left\{ \sum_{i=1}^N x_i \right\} b &= \sum_{i=1}^N y_i \\ \left\{ \sum_{i=1}^N x_i \right\} a + \left\{ \sum_{i=1}^N x_i^2 \right\} b &= \sum_{i=1}^N x_i y_i \end{aligned} \right\} \quad (6)$$

Wir erhalten, wenn

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

und

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

für die Niveaunkonstante  $a$ :

$$a = \bar{y} - b_{yx} \bar{x} \quad (7)$$

und für den Regressionskoeffizienten  $b$ :

$$b_{yx} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (8)$$

Es ist nun weiter

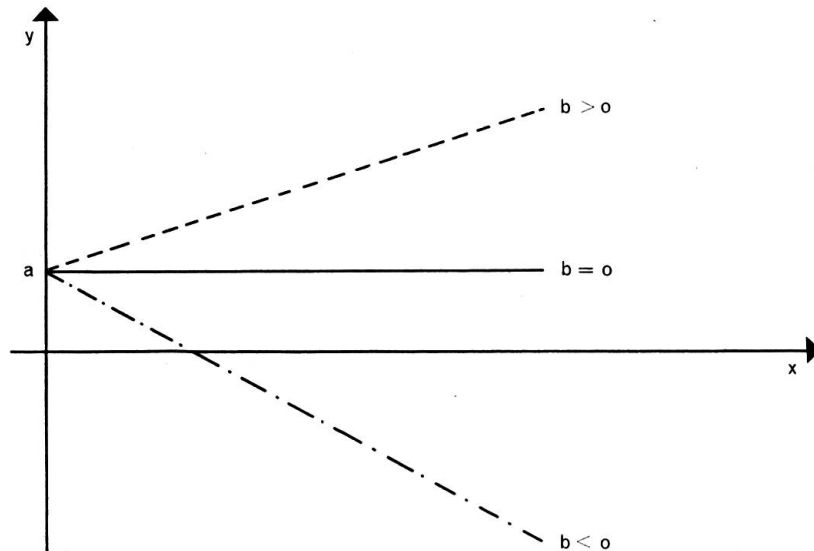
$$\left. \begin{aligned} \sum (x_i - \bar{x})(y_i - \bar{y}) &= \sum x_i y_i - \frac{1}{N} (\sum x_i)(\sum y_i) = S_{xy} \\ \sum (x_i - \bar{x})^2 &= \sum x_i^2 - \frac{1}{N} (\sum x_i)^2 = S_{xx} \\ \sum (y_i - \bar{y})^2 &= \sum y_i^2 - \frac{1}{N} (\sum y_i)^2 = S_{yy} \end{aligned} \right\} \quad (9)$$

Durch Transformation erhalten wir folgende für die numerische Auswertung geeigneten Ansätze:

$$b_{yx} = \frac{\sum x_i y_i - \frac{1}{N} (\sum x_i)(\sum y_i)}{\sum x_i^2 - \frac{1}{N} (\sum x_i)^2} = \frac{S_{xy}}{S_{xx}} \quad (10)$$



Für  $a > 0$  können beim Regressionskoeffizienten  $b$  – grafisch dargestellt – folgende Verhältnisse vorliegen:



Es zeigt somit:

**Vorzeichen von  $b$ :**

- Positiv:       Regressionslinie steigt
- Negativ:      Regressionslinie fällt
- $b = 0$ :       Regressionslinie parallel zur Abszisse

**Betrag von  $b$ :** Winkel der Regressionsgeraden zur Abszisse bzw. Ordinate.

Die **Bedeutung** des Regressionskoeffizienten  $b$ : Der in Ansatz (10) dargestellte Richtungsparameter  $b$  gibt an, um wieviel die abhängige Veränderliche  $y$  variiert (zu- oder abnimmt), wenn die unabhängige Veränderliche  $x$  um eine Einheit zunimmt.

Setzen wir nun in (5) für

$$a = \bar{y} - b\bar{x}$$

so erhalten wir folgende Regressionsgleichung:

$$Y_i = \bar{y} + b_{yx}(x_i - \bar{x}) \tag{11}$$

Die Regressionsgerade ist im Variationsbereich der unabhängigen Veränderlichen  $x$  definiert, d.h. zwischen dem kleinsten und grössten Wert. Sowohl Interpolation wie Extrapolation sind möglich, wobei vor allem bei der Extrapolation sehr oft einige Vorsicht am Platze ist.

**Umkehrung** des Problems:

Es kann sinnvoll sein, von den  $y$ -Werten auf die  $x$ -Werte zu schliessen. Der Ansatz für die Regressionsgerade lautet dann

$$X = \bar{x} + b_{xy}(y - \bar{y}) \tag{12}$$

bzw. auch

$$X = a + b_{xy}y \tag{13}$$

Der Regressionskoeffizient  $b_{xy}$  ist bestimmt aus

$$b_{xy} = \frac{\sum_{i=1}^N x_i y_i - \frac{1}{N} \left( \sum_{i=1}^N x_i \right) \left( \sum_{i=1}^N y_i \right)}{\sum_{i=1}^N y_i^2 - \frac{1}{N} \left( \sum_{i=1}^N y_i \right)^2} = \frac{S_{xy}}{S_{yy}} \tag{14}$$

Es sind somit bei der stochastischen Abhängigkeit zwei Regressionsgeraden möglich

$$Y = \bar{y} + b_{yx}(x - \bar{x})$$

und

$$X = \bar{x} + b_{xy}(y - \bar{y})$$

## 122.2 Streuung der Einzelwerte

Die Streuung der Einzelwerte um die Regressionsgerade ist gemäss folgendem Ansatz (Vgl. E. Weber, Grundriss, 1967, S.332) zu bestimmen:

$$s_{yx}^2 = \frac{1}{N-2} \sum_{i=1}^N (y_i - Y_i)^2 \quad (15)$$

Durch Transformation ergibt sich

$$s_{yx}^2 = \frac{1}{N-2} \left\{ S_{yy} - \frac{S_{yx}^2}{S_{xx}} \right\} \quad (16)$$

Mit (15) bzw. (16) messen wir das Ausmass der Übereinstimmung zwischen den geschätzten und beobachteten Werten der abhängigen Veränderlichen.

## 122.3 Bestimmtheitsmass und Korrelationskoeffizient

Während bei der Regressionsanalyse (vgl. 122.1 oben) die Richtung der stochastischen Abhängigkeit festgestellt wird, gibt uns die **Korrelationsanalyse** Aufschluss über den Grad des Zusammenhangs. Eine geeignete Masszahl ist vor allem das Bestimmtheitsmass B. Es ist folgendermassen definiert:

$$B = \frac{[S(x_i - \bar{x})(y_i - \bar{y})]^2}{[S(x_i - \bar{x})^2][S(y_i - \bar{y})^2]} = \frac{S_{xy}^2}{S_{xx} S_{yy}} \quad (17)$$

Für die numerische Auswertung lässt sich der Ausdruck nach dem ersten Gleichheitszeichen transformieren. Vor allem lässt sich B auch mit Hilfe der beiden Regressionskoeffizienten errechnen. Es ist

$$B = b_{yx} b_{xy} \quad (18)$$

**Deutung** des Bestimmtheitsmasses:

Das Bestimmtheitsmass B gibt den Anteil der Streuung der abhängigen Veränderlichen y an, der sich aus der Variabilität der unabhängigen Veränderlichen x erklären lässt.

Die **Eigenschaften** des Bestimmtheitsmasses:

- (1) B ist positiv, d. h.  $B > 0$ ;
- (2) Variation:  $0 \leq B \leq 1$ ;
- (3) Symmetrie in x und y, d. h. es ist

$$B = \frac{S_{xy}^2}{S_{xx} S_{yy}} = \frac{S_{yx}^2}{S_{yy} S_{xx}}$$

Aus dem Bestimmtheitsmass lässt sich durch einfache Operation der Korrelationskoeffizient gewinnen, der somit in enger Beziehung zu dieser Masszahl steht:

$$r = \sqrt{B} \quad (19)$$

**Eigenschaften** des Korrelationskoeffizienten:

- (1) Vorzeichen:  $r \leq 0$
- (2) Variation:  $-1 \leq r \leq +1$
- (3) Symmetrie in bezug auf die Variablen x und y:

$$r = \frac{S_{xy}}{S_x S_y} = \frac{S_{yx}}{S_y S_x}$$

Für praktische Auswertungen hat das Bestimmtheitsmass grössere Bedeutung; ihm ist vor dem Korrelationskoeffizienten der Vorrang zu geben.

## 123 Prüfen von Hypothesen

### 123.1 Grundsätzliches

Ein **Prüfverfahren** (Test) wird benützt, damit entschieden werden kann, ob eine aufgestellte Hypothese angenommen oder verworfen werden soll, bzw. nach A. Linder (Statistische Methoden ..., S.43): « ..., ob die Angaben einer Stichprobe mit einer Hypothese verträglich sind oder ihr widersprechen. »

Unter einer **Hypothese** versteht man nach Merriam-Webster «a tentative theory or supposition provisionally adopted to explain certain facts and to guide in the investigation of others».

In bezug auf die Arten von Hypothesen unterscheidet man:

- Nullhypothese;
- Alternativhypothese (Gegenhypothese).

Von grosser Bedeutung bei der Regressionsrechnung ist der Test, ob der Richtungsparameter  $b_1$  wesentlich oder nur zufällig von einem zweiten Regressionskoeffizienten  $b_2$  verschieden ist. Wir formulieren folgende Hypothese (einseitiger Test):

- (1) Nullhypothese  $H_0 : b_1 = b_2 = 0.62$
- (2) Alternativhypothese  $H_A : b_1 > b_2$   
[bzw.  $H_A : b_1 < b_2$ ]

Der zweiseitige Test prüft, ob zwischen  $b_1$  und  $b_2$  keine Differenz vorliegt bzw. die Alternativhypothese der gegenseitigen Abweichung, somit z. B.:

- (1) Nullhypothese  $H_0 : b_1 = b_2 = 0.62$
- (2) Alternativhypothese  $H_A : b_1 \neq b_2$

**Fehler bei Prüfverfahren:**

Die Gültigkeit von Hypothesen muss überprüft werden. Die Resultate lassen erkennen, ob die Hypothese angenommen oder aber verworfen werden muss. Dabei können zwei Typen von Fehlern unterlaufen:

- Typ I: Verwerfen einer Hypothese, die richtig ist (= Risiko 1. Art);
- Typ II: Annahme einer Hypothese, die falsch ist (= Risiko 2. Art).

Die folgende Übersicht zeigt die vier Möglichkeiten:

Hypothese	Entscheid	
	Annahme	Ablehnung
Hypothese ist richtig	Richtiger Entscheid	Fehlertyp I
Hypothese ist falsch	Fehlertyp II	Richtiger Entscheid

Unser Entscheid ist also richtig, wenn wir

- (1)  $H_0$  annehmen, wenn  $H_0$  richtig ist;
- (2)  $H_0$  ablehnen, wenn  $H_0$  falsch ist.

Hingegen begehen wir einen Fehler von **Typ I**, wenn wir  $H_0$  ablehnen, wenn diese Hypothese richtig ist, bzw. einen Fehler vom **Typ II**, wenn wir  $H_0$  annehmen, wenn  $H_0$  falsch ist. Schwerwiegender ist meistens Fehlertyp II; man versucht ihn deshalb zu minimalisieren, indem der Stichprobenumfang erhöht wird. Dieses Vorgehen kann mit dem Hinweis darauf begründet werden, dass mehr Informationen bessere Entscheide ermöglichen.

Für die Praxis ist es nützlich, auch bei den Prüfverfahren einen bestimmten **Arbeitsablauf** einzuhalten:

- (1) Problemlage
- (2) Formulierung der Teilhypothese
- (3) Prüfverteilung
- (4) Sicherheitsgrad (Entscheidungswahrscheinlichkeit)
- (5) Sicherheitsschwelle
- (6) Bereitstellung der Prüf-Rechenformel; Berechnung des Prüfwertes
- (7) Entscheid (Vergleich Prüfwert mit Sicherheitsschwelle); → Annahme oder Ablehnung der Hypothese (bzw. Alternativhypothese).
- (8) Schlussfolgerungen.

### 123.2 Varianzanalyse

Die Quadratsumme der Abweichungen der Einzelwerte  $y_i$  von ihrem Mittelwert  $\bar{y}$  kann folgendermaßen zerlegt werden:

$$S(y_i - \bar{y})^2 = S(Y_i - \bar{y})^2 + S(y_i - Y_i)^2 \quad (20)$$

Nun ist aber

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

$$S(Y_i - \bar{y})^2 = b_{yx} \left\{ S x_i y_i - \frac{1}{N} (S x_i) (S y_i) \right\} = b_{yx} S_{xy}$$

$$S(y_i - Y_i)^2 = S_{yy} - b_{yx} S_{xy}$$

Wir erhalten folgendes Schema der Streuungszersetzung:

Streuung	Summe der Quadrate	FG	Durchschnittsquadrate
Auf der Regression	$b_{yx} S_{xy}$	1	$s_1^2 = b_{yx} S_{xy}$
Um die Regression	$S_{yy} - b_{yx} S_{xy}$	$N - 2$	$s_2^2 = \frac{1}{N - 2} \{S_{yy} - b_{yx} S_{xy}\}$
Insgesamt	$S_{yy}$	$N - 1$	.

(21)

Es ist ferner

$$s_2^2 = s_{yx}^2 = \frac{1}{N - 2} \{S_{yy} - b_{yx} S_{xy}\}$$

Es kann gezeigt werden, dass

$$s_{yx}^2 = \hat{\sigma}_y^2$$

d.h.  $s_{yx}^2$  ist ein zuverlässiger Schätzwert der unbekanntenen Streuung  $\hat{\sigma}_y^2$ .

**Prüfverfahren:** F-Test

$$F = \frac{\text{DQ (auf der Regression)}}{\text{DQ (um die Regression)}} = \frac{s_1^2}{s_2^2} \quad (22)$$

**Freiheitsgrade:**  $n_1^* = 1$ ;  $n_2^* = N - 2$

**Entscheidungskriterien:**

$$\left. \begin{array}{l} F \geq F_p: \text{Regression gesichert für Irrtumswahrscheinlichkeit von } P\% \\ F < F_p: \text{Weitere Berechnungen nicht sinnvoll} \end{array} \right\} \quad (23)$$

Ist die Regression gesichert, werden von Null verschieden sein:

- Regressionskoeffizient;
- Bestimmtheitsmass (Korrelationskoeffizient).

Folge: Regressionsgerade kann bestimmt werden.

### 123.3 Prüfen der Regressionsparameter

Wir prüfen, ob die Abweichungen der beiden Regressionsparameter  $a$  und  $b$  einer Stichprobe von den entsprechenden Werten der Grundgesamtheit wesentlich oder nur zufällig sind. Als Prüfstreuung dient uns Ansatz (16) oben, d.h.

$$s_{yx}^2 = \frac{1}{N - 2} \left\{ S_{yy} - \frac{S_{yx}^2}{S_{xx}} \right\}$$

Als Prüfverteilung muss die t-Verteilung in Betracht gezogen werden.



### (1) Prüfen der Niveaunkonstanten a:

Wir verwenden folgenden t-Test:

$$t = \frac{a - \alpha}{s_{yx}} \sqrt{N} \quad (24)$$

mit:  $n^* = (N - 2)$  Freiheitsgraden.

Wir setzen nun  $\alpha = 0$ . Es ist dann folgende Hypothese zu prüfen:

$$H_0 : a = 0$$

Der Prüfansatz wird dann zu

$$t = \frac{a}{s_{yx}} \sqrt{N}, \text{ bzw. } t^2 = \frac{a^2 N}{s_{yx}^2}.$$

### (2) Prüfen des Regressionskoeffizienten b:

Auch hier kommt ein t-Test zur Anwendung

$$t = \frac{b_{yx} - \beta}{s_{yx}} \sqrt{S_{xx}} \quad (25)$$

mit:  $n^* = (N - 2)$  Freiheitsgraden.

Wir prüfen folgende Hypothese:

$$H_0 : b_{yx} = \beta$$

Diese Hypothese wird angenommen, wenn

$$\frac{|b_{yx} - \beta|}{s_{yx}} \sqrt{S_{xx}} < t_p$$

andernfalls wird sie abgelehnt.

Nun ist aber  $\beta$  meistens nicht bekannt. Es ist daher sinnvoll, den Richtungsparameter  $b_{yx}$  mit Null zu vergleichen. Die Frage lautet: Ist  $b$  wesentlich oder nur zufällig von Null verschieden? Dies ist bei der Regressionsanalyse die eigentliche Kardinalfrage. Sollte nämlich  $b = 0$  sein, so hat die in Betracht gezogene unabhängige Veränderliche **keinen Einfluss** auf die abhängige Variable.

Der Prüfansatz hat folgende Form:

$$t = \frac{b_{yx}}{s_{yx}} \sqrt{S_{xx}} \quad (26)$$

mit:  $n^* = (N - 2)$  Freiheitsgraden.

Es ist folgende Hypothese zu prüfen:

$$H_0 : b_{yx} = 0$$

Diese Hypothese wird abgelehnt, wenn

$$\frac{|b_{yx}|}{s_{yx}} \sqrt{S_{xx}} > t_p$$

Der Regressionskoeffizient ist in diesem Fall wesentlich von Null verschieden; die Hypothese eines bestehenden Einflusses muss in der Folge akzeptiert, bzw. die Nullhypothese muss abgelehnt werden.

### (3) Prüfen der Differenz zwischen zwei Regressionskoeffizienten

Es ist folgende Hypothese zu prüfen:

$$H_0 : b_1 = b_2$$

d. h. wir prüfen, ob die Differenz zwischen den beiden Regressionskoeffizienten Null ist. Für das Prüfverfahren vgl. E. Weber, Grundriss (6. Auflage 1967), S. 338f.

## 123.4 Prüfen des Bestimmtheitsmasses und des Korrelationskoeffizienten

### (1) Prüfen des Bestimmtheitsmasses

Es ist naheliegend, die Abweichung des Bestimmtheitsmasses von Null zu prüfen; denn für  $B = 0$  muss eine Abhängigkeit zwischen den betrachteten Variablen negiert werden.

Wir stellen also folgende Hypothesen auf:

$$H_0: B = 0$$

$$H_A: B > 0$$

Nach A. Linder (Methoden, S.182) lässt sich das Prüfverfahren aus der Streuungszerlegung gewinnen. Der Ansatz (21) kann zu diesem Zweck auch folgendermassen formuliert werden:

Streuung	Summe der Quadrate	FG	Durchschnittsquadrate
Regression	$BS_{yy}$	1	$s_1^2 = BS_{yy}$
Einzelwerte um Regression	$(1 - B) S_{yy}$	$N - 2$	$s_2^2 = \frac{1}{N - 2} (1 - B) S_{yy}$
Insgesamt	$S_{yy}$	$N - 1$	.

Das Verhältnis der beiden Durchschnittsquadrate ergibt einen F-Test:

$$F = \frac{s_1^2}{s_2^2} = \frac{BS_{yy}(N-2)}{(1-B)S_{yy}} = \frac{B(N-2)}{(1-B)} \quad (27)$$

Mit:  $n_1^* = 1$ ;  $n_2^* = (N-2)$  FG und dem **Entscheidungskriterium**:

$F \geq F_p$ : B ist wesentlich verschieden von Null; Einfluss besteht.

$F < F_p$ : B ist bloss zufällig von Null verschieden.

Das Bestimmtheitsmass B kann auch mit Tafelwerten (Sicherheitspunkte) überprüft werden (vgl. Linder, Methoden, S.183 bzw. S.469-471). Verwendet wird folgender Ansatz:

$$B_p = \frac{F_p}{(N-2) + F_p} \quad (28)$$

Ein Bestimmtheitsmass, das kleiner ist als (28), weicht nur zufällig von Null ab. Es gilt also das Entscheidungskriterium:

$B \geq B_p$ : B ist signifikant von Null verschieden.

## (2) Prüfen des Korrelationskoeffizienten

Auch für das Prüfen des Korrelationskoeffizienten kann ein t-Test angesetzt werden. Der Ansatz lautet

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

mit:  $n^* = (N-2)$  Freiheitsgraden.

Als Entscheidungskriterium gilt:

$t \geq t_p$ : Der Korrelationskoeffizient ist signifikant von Null verschieden.

Es ist selbstverständlich möglich, mit Hilfe von Sicherheitspunkten (Tabellenwerten) das Prüfverfahren durchzuführen.

### 123.5 Linearität der Regression

Wir haben weiter oben (vgl. Ziff.112) dargelegt, dass bei der Aufstellung des Modells eine Hypothese über die **Form der Beziehung** zwischen den betrachteten Variablen zu formulieren ist. Meistens wird man die Einzelwerte in einer Grafik zur Darstellung bringen und auf Grund dieser Zeichnung einen Vorentscheid fällen. Es ist aber angebracht, dieses Vorgehen durch ein Prüfverfahren abzustützen. So

wird man z. B. prüfen, ob die Hypothese der Linearität angenommen werden kann oder aber verworfen werden muss. Geeignet zur Lösung dieses Problems ist die Varianzanalyse bzw. ein F-Test. E. Weber (vgl. Grundriss, 1967, S.339 ff.) bemerkt dazu: «Oftmals ist aus der Anordnung der Wertepaare (x,y) ersichtlich, ob lineare Regression angenommen werden kann. In Zweifelsfällen muss eine Prüfung auf Linearität vorgenommen werden. Diese Prüfung beruht darauf, die Summe der Quadrate der Abweichungen der Mittel der Beobachtungswerte von den entsprechenden Regressionswerten mit der Quadratsumme der Abweichungen der Beobachtungswerte von den jeweiligen Spaltenmitteln zu vergleichen.»

Das führt zu folgender **Anordnung**:

Unabhängige Veränderliche	$x_1$	...	$x_k$	...	$x_p$	Total
Einzelwerte $y_{ik}$	$y_{11}$	...	$y_{1k}$	...	$y_{1p}$	
Allgemein:			$y_{ik}$			
Umfang	$n_1$	...	$n_k$	...	$n_p$	$n = \sum_{k=1}^p n_k$

Nach E. Weber (Grundriss, S.340 ff.) führen wir die Varianzanalyse in zwei Stufen durch:

1. Stufe: Aufteilung der SQ (insgesamt) in SQ (zwischen den Spalten) und SQ (innerhalb der Spalten).
2. Stufe: Zerlegung der SQ (zwischen den Spalten).

Als Ergebnis der zwei Streuungszerlegungen erhalten wir einen F-Test zur Beurteilung, ob Linearität angenommen werden darf oder nicht.

In den nachstehenden Schemata der Streuungszerlegung sind die für numerische Auswertung benötigten Ansätze enthalten. Für die Theorie verweisen wir auf A. Linder (Methoden, S.156 ff.) bzw. E. Weber (Grundriss, S.340/341).

Wir nehmen vorerst eine einfache Streuungszerlegung mit einer Aufteilung der SQ (insgesamt) in SQ (zwischen den Spalten) und SQ (innerhalb der Spalten) vor, d. h. also

$SQ(\text{insgesamt}) = SQ(\text{zwischen Spalten}) + SQ(\text{innerhalb})$ .

Streuungszerlegung 1. Stufe:

Streuung	SQ	FG	DQ
Zwischen den Spalten	$S \frac{1}{n_k} \left( \sum_i y_{ik} \right)^2 - \frac{1}{n} \left( \sum_{k,i} y_{ik} \right)^2 = B$	$p - 1$	.
Innerhalb der Spalten	$A - B = C$	$n - p$	$s_{III}^2 = \frac{C}{n - p}$
Insgesamt	$S_{yy} = A$	$n - 1$	.

(29)

Es ist (vgl. oben)

$$S_{yy} = \sum_{k,i} y_{ik}^2 - \frac{1}{n} \left( \sum_{k,i} y_{ik} \right)^2$$

Das SQ (zwischen den Spalten) lässt sich weiter aufteilen in

$$SQ(\text{zwischen}) = \sum_k [n_k (\bar{y}_{\cdot k} - Y_k)^2] + \sum_k [n_k (Y_k - \bar{y}_{\cdot\cdot})^2]$$

Streuungszerlegung 2. Stufe:

Streuung	SQ	FG	DQ
Auf der Regression	$(S_{xy})^2 / S_{xx} = D$	1	$s_I^2 = D$
Mittelwerte um Regression	$B - D = E$	$p - 2$	$s_{II}^2 = \frac{E}{p - 2}$
Zwischen den Spalten	$S \frac{1}{n_k} \left( \sum_i S_{y_{ik}} \right)^2 - \frac{1}{n} \left( \sum_{k,i} S_{y_{ik}} \right)^2 = B$	$p - 1$	.

(30)

Aus (29) und (30) gewinnen wir folgenden F-Test:

$$F = \frac{s_{II}^2}{s_{III}^2}, \quad \text{mit: } \left. \begin{array}{l} n_1^* = (p-2) \\ n_2^* = (n-p) \end{array} \right\} \text{ Freiheitsgraden}$$

Voraussetzung:  $s_{II}^2 > s_{III}^2$ .

Als **Entscheidungskriterium** verwenden wir:

$F < F_p$ : Die DQ  $s_{II}^2$  und  $s_{III}^2$  sind nur zufällig voneinander verschieden: Lineare Regression ist zulässig. Der Test sagt nur aus, dass lineare Regression zulässig sei. Es ist aber möglich, dass eine andere Form den Daten besser entspricht.

Ist hingegen  $F > F_p$ , so darf keinesfalls lineare Regression angenommen werden.

## 124 Vertrauensgrenzen der Schätzung

### 124.1 Grundsätzliches

Für die Schätzwerte müssen zufällige Schwankungen in Rechnung gestellt werden. Es wird unsere Aufgabe sein, in den folgenden zwei Abschnitten Überlegungen in bezug auf die Unsicherheit der Schätzwerte anzustellen, insbesondere für

- den Regressionskoeffizienten  $b$ , bzw.
- für die Regressionswerte  $Y_i$ .

Dabei wird es darum gehen, eine geeignete Konzeption von Streuungsbereichen zu finden.

### 124.2 Vertrauensgrenzen für die Regressionsparameter

Wie bereits weiter oben angedeutet, nimmt der Regressionskoeffizient  $b$  im Rahmen der Regressionsanalyse eine ausserordentlich wichtige Stellung ein. Es ist daher gerechtfertigt, Überlegungen anzustellen in bezug auf die Unsicherheit des aus einer Stichprobe gewonnenen Schätzwertes.

Wir haben oben in Ziff. 123.3 folgenden t-Test angesetzt

$$t = \frac{b_{yx}}{s_{yx}} \sqrt{S_{xx}}$$

Aus diesem Ansatz können folgende Vertrauensgrenzen bestimmt werden (für die theoretischen Grundlagen vgl. K.W. Smillie, Introduction, S.10/11):

$$b \pm t_p \frac{s_{yx}}{\sqrt{S_{xx}}} \quad (31)$$

Es sind  $n^* = N - 2$  Freiheitsgrade zu berücksichtigen.

Es ist ferner in Ansatz (31):

$$s_{yx}^2 = \frac{1}{N-2} \left\{ S_{yy} - \frac{S_{yx}^2}{S_{xx}} \right\}$$

bzw.

$$S_{xx} = S X_i^2 - \frac{1}{N} (S X_i)^2$$



Selbstverständlich ist es auch möglich, Vertrauensgrenzen für die Niveaunkonstante  $a$  in Rechnung zu stellen:

$$a \pm t_p \frac{s_{yx}}{\sqrt{N}} \quad (32)$$

Sowohl für Ansatz (31) als auch für (32) sind  $n^* = N - 2$  Freiheitsgrade in Rechnung zu stellen.

### 124.3 Vertrauensgrenzen für die Regressionswerte $Y_i$

Die Streuung des zum Wert  $x$  gehörenden Regressionswertes beträgt (vgl. A.Linder, Methoden, S.160; K.W.Smillie, Introduction, S.11), dargestellt in unserer Symbolik,

$$s_Y^2 \sim s_{yx}^2 \left\{ \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right\} \quad (33)$$

Der Ausdruck in der Klammer von (33)

$$\frac{(x_i - \bar{x})^2}{S_{xx}}$$

zeigt, dass  $s_Y^2$  mit zunehmender Entfernung der  $x_i$  von  $\bar{x}$  grösser wird. Mit Hilfe von (33) können wir nun die Vertrauensgrenzen definieren

$$a + bx_i \pm t_p s_Y \quad (34.1)$$

bzw.

$$Y_i \pm t_p s_Y \quad (34.2)$$

Es sind in diesem Ansatz für die Bestimmung des Tafelwertes  $t_p$  insgesamt  $n^* = (N-2)$  Freiheitsgrade zu berücksichtigen.

In der Praxis liegt sehr oft der Fall vor, dass für einen Einzelwert von  $x$  der dazugehörige  $Y$ -Wert prognostiziert werden muss. Für die Streuung gilt dann:

$$s_{Y(E)}^2 \sim s_{yx}^2 \left[ 1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right] \quad (35)$$

Für den Wert  $t_p$  bei der Bestimmung der Vertrauensgrenzen sind auch hier  $n^* = N-2$  Freiheitsgrade vorzumerken.

## 125 Zwei Regressionsgerade

Es ist oft die Sachlage gegeben, dass zwei Regressionsgerade miteinander zu vergleichen sind. Das Vorgehen zur Lösung dieses Problems gestaltet sich folgendermassen:

- (1) Man wird vorerst für beide Regressionsgerade die **Linearität** prüfen;
- (2) Es wird zu prüfen sein, ob die Regressionskoeffizienten  $b_1$  bzw.  $b_2$  überhaupt **von Null verschieden** sind;
- (3) Sind lineare Abhängigkeiten als zulässig erkannt bzw. die Richtungsparameter verschieden von Null, so wird man schliesslich die **Parallelität** prüfen.

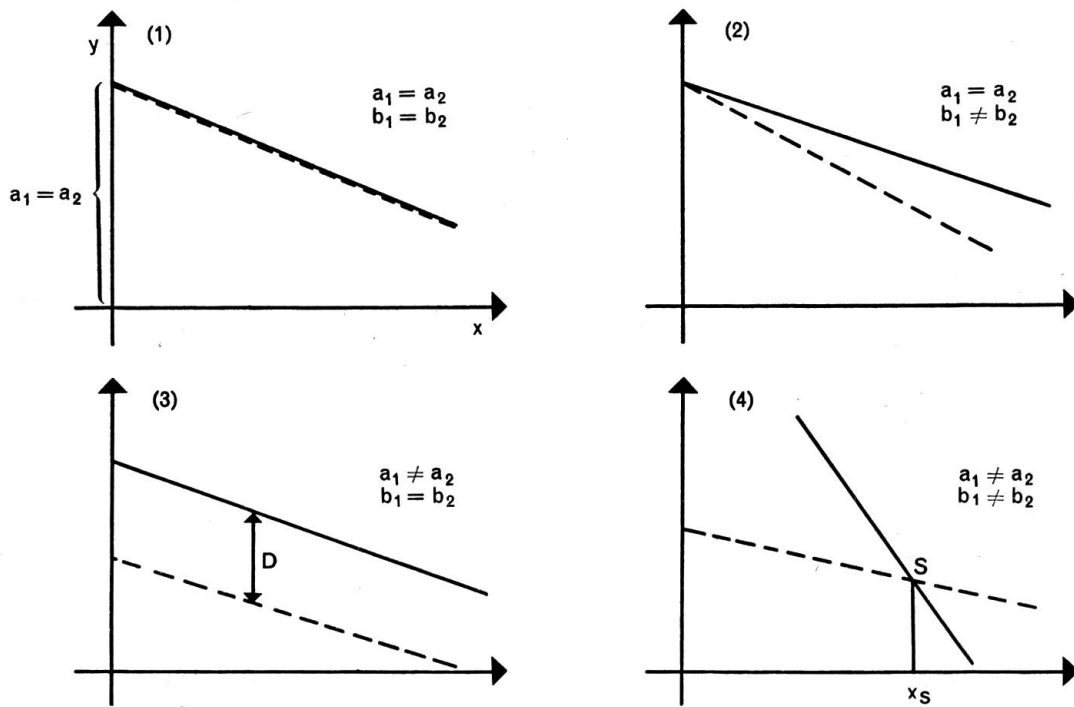
In bezug auf Niveaunkonstanten bzw. Regressionskoeffizienten sind nun folgende Fälle denkmöglich:

Regressions- koeffizienten	Niveaunkonstanten	
	$a_1 = a_2$	$a_1 \neq a_2$
Gleich	$b_1 = b_2$	$b_1 = b_2$
Verschieden	$b_1 \neq b_2$	$b_1 \neq b_2$

Sollte der Fall vorliegen, wo die Niveaunkonstanten verschieden, die Regressionskoeffizienten aber in Vorzeichen und Betrag gleich sind, so kann es nützlich sein, die Distanz  $D$  zwischen beiden Regressionsgeraden zu berechnen und in die Überlegungen einzubeziehen.

Bei Nichtparallelität wird es interessant sein, den Schnittpunkt beider Geraden zu bestimmen.

Aus unserem soeben aufgeführten Schema lassen sich folgende Grafiken gewinnen:



Die beiden Probleme

- Linearität (vgl. Ziff.123.5)
- Verschiedenheit von Null (vgl. 123.3)

haben wir oben behandelt. Für die in der Praxis hauptsächlich vorkommenden Fälle (3) und (4) – verschiedene Niveaunkonstanten – ist es gegeben, Varianzanalyse und Regressionsanalyse zu kombinieren; es resultiert die sogenannte Kovarianzanalyse, auch etwa Mitstreuungszerlegung genannt.

A.Linder (Methoden, S.220) bemerkt dazu: «Die Mitstreuungszerlegung wird vor allem benützt, um festzustellen, wie eine oder mehrere unabhängige Veränderliche sich auf die Unterschiede zwischen den Durchschnitts einer abhängigen Veränderlichen auswirken. Sie kann aber auch dazu dienen, die Unterschiede zwischen mehreren Regressionskoeffizienten zu beurteilen.»

Als Grundlage für die Berechnungen seien N Beobachtungen mit Werteverbindungen (x,y) gegeben. Wir nehmen eine Ausgliederung in p Gruppen (1, 2, ..., k, ..., p) vor. Innerhalb der Teilgesamtheiten sind die entsprechenden Werteverbindungen vorhanden. Aus diesen können Summenwerte und Durchschnitte gebildet werden. Für theoretische Erörterungen verweisen wir u.a. auf die Monographien von A.Linder und E.Weber, bzw. auf das Werk von B.Ostle.

Für die numerische Auswertung gelangen wir zu folgendem Schema der Streuungszerlegung (36):

Streuung	SQ	FG	DQ
1. Gerade .....	$S_{y_1y_1} - b_1 S_{x_1y_1}$		.
k-te Gerade .....	$S_{y_ky_k} - b_k S_{x_ky_k}$		.
p-te Gerade	$S_{y_py_p} - b_p S_{x_py_p}$		.
p Regressionsgerade	$\sum_{k=1}^p S_{y_ky_k} - \sum_{k=1}^p b_k S_{x_ky_k} = E$	$N - 2p$	$s_1^2 = \frac{E}{N - 2p}$
Nichtparallelität	$C - E = D$	$p - 1$	$s_2^2 = \frac{D}{p - 1}$
Innerhalb der Gruppen	$\sum_{k=1}^p S_{y_ky_k} - b^* \left( \sum_{k=1}^p S_{x_ky_k} \right) = C$	$N - p - 1$	$s_3^2 = \frac{C}{N - p - 1}$
Zwischen den Gruppen	$A - C = B$	$p - 1$	$s_4^2 = \frac{B}{p - 1}$
Insgesamt	$S_{yy} - b S_{xy} = A$	$N - 2$	.

Es ist ferner [vgl. SQ (innerhalb Gruppen)]

$$b^* = \frac{S_{xy}^*}{S_{xx}^*} = \frac{\sum_k S_{x_k y_k}}{\sum_k S_{x_k x_k}}$$

Vgl. zu diesem Schema auch: B. Ostle, Statistics in Research, 1964, S. 203.

Aus der Streuungszerlegung (36) resultiert ein F-Test, der Antwort auf die Frage gibt, ob die berechneten Regressionskoeffizienten wesentlich oder nur zufällig voneinander abweichen. Als Prüfstreuung verwenden wir  $s_1^2$ .

Formulierung der **Hypothese**:

$H_0$ :  $b_1 = b_2 = \dots = b_p = b$

$H_A$ : Mindestens zwei Regressionskoeffizienten sind signifikant voneinander verschieden.

Der F-Test lautet

$$I. \quad F = \frac{s_2^2}{s_1^2}; \quad \text{mit: } \left. \begin{array}{l} n_1^* = p - 1 \\ n_2^* = N - 2p \end{array} \right\} \text{ Freiheitsgraden} \quad (37)$$

Die errechnete Verhältniszahl vergleichen wir mit dem Tabellenwert. Bei

$F < F_p$ : Die Regressionskoeffizienten sind nur zufällig voneinander verschieden.

Sind mindestens zwei Regressionskoeffizienten signifikant voneinander verschieden, so kann man prüfen, zwischen welchen Parametern Unterschiede bestehen (t-Test).

Unter der Voraussetzung der Parallelität der Regressionskoeffizienten kann man testen, ob die Abstände  $D_k$  wesentlich oder nur zufällig voneinander verschieden sind. Dazu verwenden wir die untersten drei Zeilen der Streuungszerlegung von (36) und erhalten folgenden F-Test:

$$II. \quad F = \frac{s_4^2}{s_3^2}; \quad \text{mit: } \left. \begin{array}{l} n_1^* = p - 1 \\ n_2^* = N - p - 1 \end{array} \right\} \text{ Freiheitsgraden} \quad (38)$$

Das Entscheidungskriterium:

$F \geq F_p$ : Mindestens zwei der Distanzen sind signifikant voneinander verschieden.

Man kann auch hier prüfen, welche Distanzen voneinander verschieden sind. Ist  $F < F_p$ , so weichen die Distanzen nur zufällig voneinander ab; die Regression durch den Schwerpunkt kann die stochastische Abhängigkeit generell angeben. Wir hätten dann die Verhältnisse, die in Grafik (1) dargestellt sind. Es liegen also vor:

- Linearität;
- Verschiedenheit der beiden b von Null;
- Parallelität;
- Distanzen nur zufällig voneinander verschieden.

In der Praxis dürfte jedoch dieser Spezialfall nur selten gegeben sein.

## 13 Einfache nichtlineare Regression

### 131 Grundsätzliches

Im Abschnitt 12 haben wir das einfachste Modell behandelt; es betrifft die lineare Regression und Korrelation. Als Ergebnis des Linearitätstests muss aber in vielen Fällen die Hypothese einer nichtlinearen stochastischen Abhängigkeit angenommen werden. Sehr oft weisen schon theoretische Untersuchungen auf eine nichtlineare Beziehung hin.

Zur Lösung der vorliegenden Probleme drängen sich folgende Methoden auf:

- Transformation auf den linearen Fall;
- Mehrfache lineare Regression;
- Orthogonale Polynome.

Es wird fallweise zu prüfen sein, welches Verfahren geeignet ist, die Daten auszuwerten.

### 132 Transformation auf einfache lineare Regression

Theoretische oder empirische Untersuchungen lassen es oft als zweckmässig erscheinen, nichtlineare Regression anzunehmen. Der Übergang zu Logarithmen muss sich in vielen Fällen – besonders bei betriebswirtschaftlichen und volkswirtschaftlichen Problemen – als naheliegendste und einfachste Lösungsmöglichkeit aufdrängen. Folgende Ansätze können einer derartigen Transformation unterzogen werden:

$$Y = ab^x \quad (39)$$

Die Transformation ergibt

$$\log Y = \log a + x(\log b)$$

Die Bestimmung der Parameter erfolgt in der oben dargestellten Art und Weise (vgl. Ziff.12).

$$Y = ax^b \quad (40)$$

Das Ergebnis der Transformation

$$\log Y = \log a + b(\log x)$$

Auch hier stellen sich keine neuen Probleme. Die Erkenntnisse aus der linearen Regression finden analoge Anwendung.

### 133 Mehrfache lineare Regression

In bestimmten Fällen führt die Methode der mehrfachen linearen Regression zum verfolgten Ziel, eine nichtlineare Beziehung statistischen Verfahren zu unterwerfen (vgl. dazu A.Linder, Methoden, S.207 ff.).

Eine Regressionsgleichung von der Form

$$Y = a + bx + cx^2$$

kann in die Regressionsgleichung

$$Y = a + b_1x_1 + b_2x_2$$

oder auch

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

übergeführt werden, mit  $x_1 = x$  und  $x_2 = x^2$ . Für die Berechnung der Regressionskoeffizienten  $b_1$  und  $b_2$  sind die entsprechenden Ansätze in Abschnitt 14 (mehrfache lineare Regression) massgebend.

Man wird im Anschluss an die Berechnungen zur Auffindung der Regressionsparameter verschiedene Hypothesen prüfen, so z. B. ob die Regressionsgleichung mit den beobachteten Werten gut übereinstimmt. Ferner könnte man die Frage stellen, ob nicht das lineare Glied weggelassen werden könnte, d. h. man wird für  $b_1$  die Hypothese prüfen, ob

$$H_0: \beta_1 = 0$$

Man wird eine Streuungszerlegung mit einem F-Test ansetzen, um zu prüfen, ob sich der Wert von  $b_1$  mit der Hypothese verträgt, dass  $\beta_1 = 0$ . Ist dies der Fall, wird man die eine der unabhängigen Variablen weglassen können.

### 134 Regression mit orthogonalen Polynomen

Bisher haben wir bei einfacher linearer Regression folgenden Ansatz benützt

$$Y = a_0 + a_1x$$

Durch Erweiterung erhalten wir ein allgemeines Polynom p-ten Grades

$$Y = a_0 + a_1x + a_2x^2 + \dots + a_px^p \quad (41)$$

Wir stellen die Forderung F auf

$$F = S(y_i - Y_i)^2 = \text{Minimum} .$$

Die Bestimmungsgleichungen lassen sich mit Hilfe der Determinantenmethode lösen.



Die Berechnungen können vereinfacht werden, wenn wir an Stelle von (41) folgenden Ansatz verwenden:

$$Y = A_0 + A_1\varphi_1 + \dots + A_p\varphi_p \quad (42)$$

In dieser Formel sind

$\varphi_k$  ( $k = 1, 2, \dots, p$ ) = Orthogonale Polynome;

$A_k$  ( $k = 0, 1, \dots, p$ ) = Konstanten, definiert durch folgende Ansätze:

$$A_0 = SY/N, \text{ bzw.} \quad (43)$$

$$A_k = \frac{SY\varphi_k}{S(\varphi_k)^2}, \quad k = 1, 2, \dots, p \quad (44)$$

Für eingehendere Darstellungen vgl. u. a.:

**Draper, N. R.:** Applied Regression Analysis, 1966, S.150 ff.

**Fisher, R. A.:** Statistical Methods, 13th ed. 1963, insbes. Kap.27 und 28 (S.147–156).

**Grossen, H.:** Regression mit orthogonalen Polynomen, Bern 1948.

**Linder, A.:** Statistische Methoden, 1964, S.210 ff.

**Mather, K.:** Statistical Analysis in Biology, London 1964 (S.129 ff.).

**Weber, E.:** Grundriss, 6. Aufl. 1967, S.352 ff.

Für Tafelwerte stehen zur Verfügung:

**Fisher/Yates:** Statistical Tables, 6th ed. 1963, S.33 ff. bzw. 98 ff.

Die Methode der orthogonalen Polynome ist vor allem gut geeignet zur Behandlung von Zeitreihen (äquidistante Werte von  $x$ ).

## 14 Mehrfache lineare Regression und Korrelation

### 141 Allgemeines

In den vorhergehenden Abschnitten haben wir zwei Veränderliche in ihrer Abhängigkeit untersucht. Es gibt aber Problemstellungen, wo drei und mehr Variable gleichzeitig analysiert werden müssen. So kann es vorkommen, dass die Kosten pro Pflage tag nicht nur von der Produktivität abhängig sind, sondern auch von den Marktpreisen der Produktionsfaktoren. Eine abhängige Variable steht dann mehreren Einflussfaktoren, mehreren unabhängigen Variablen gegenüber. Solche Probleme lassen sich nur mit Hilfe der mehrfachen linearen Regression und Korrelation lösen. Dabei bildet die Theorie der einfachen linearen Regression und Korrelation die Basis, auf dem die mehrfache Regression aufgebaut werden kann.

### 142 Das Modell

Wie in Abschnitt 121 oben nehmen wir eine lineare Beziehung zwischen einer abhängigen und  $p$  unabhängigen Variablen an:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (45)$$

Es sollen ferner  $n$  Werteverbindungen vorhanden sein:

$$(x_{1i}, x_{2i}, \dots, x_{pi}, y_i), \quad i = 1, 2, \dots, n,$$

für  $p$  unabhängige Variable  $x_1, x_2, \dots, x_p$  und einer abhängigen Veränderlichen  $y$ . Die Beobachtungen genügen, analog zu Ziff.121, folgenden  $n$  Beziehungen:

$$y_i = \alpha + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (46)$$

Für  $\alpha, \beta_1, \beta_2, \dots, \beta_p$  sollen nun Schätzwerte vorliegen, so dass (46) wird:

$$y_i = a + b_1 x_{1i} + \dots + b_p x_{pi} + \varepsilon_i \quad (47)$$

Als Schätzwert für  $Y_i$  ergibt sich damit für zwei unabhängige Variable:

$$Y_i = a + b_1 x_{1i} + b_2 x_{2i} \quad (48)$$

Der Ansatz (48) enthält die Niveaunkonstante  $a$  und die beiden Regressionskoeffizienten  $b_1$  und  $b_2$ , die zu bestimmen sind.

Für die Bedingungen und Annahmen in bezug auf das Modell vgl. K.W. Smillie, Introduction, S.41 oben. Erwähnt sei lediglich die Beschränkung (vgl. Smillie, S.41, Ziff.3.2.5): «The number of sets of observations is greater than the number of regression coefficients to be estimated, i.e.  $n > p + 1$ , and no one independent variable is a linear combination of any of the remaining independent variables». Wie bei der einfachen linearen Regression sind Schätzwerte aufzusuchen, Hypothesen zu prüfen und Vertrauensgrenzen zu bestimmen.

### 143 Schätzverfahren

#### 143.1 Regressionsparameter

Vorerst sind die mehrfachen (partiellen) Regressionskoeffizienten zu bestimmen. Gegeben seien zwei unabhängige Veränderliche und eine abhängige Variable:

##### (1) Unabhängige Veränderliche

$$X_1: x_{11}, x_{12}, \dots, x_{1N}$$

$$X_2: x_{21}, x_{22}, \dots, x_{2N}$$

##### (2) Abhängige Veränderliche

$$Y: y_1, y_2, \dots, y_N$$

Es bestehen also folgende Werteverbindungen:

$$(x_{1i}, x_{2i}; y_i); \quad i = 1, 2, \dots, N$$

Es gilt die Forderung

$$S(y_i - a - b_1 x_{1i} - b_2 x_{2i})^2 = \text{Minimum.}$$

Wir verzichten auf Beweise und geben als Resultat:

##### Niveaunkonstante a:

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 \quad (49)$$

oder auch

$$a = \frac{S y_i - b_1 S x_{1i} - b_2 S x_{2i}}{N} \quad (50)$$

##### Regressionskoeffizienten $b_1$ und $b_2$ :

$$b_1 = \frac{1}{\Delta} \begin{vmatrix} S_{x_1 y} & S_{x_1 x_2} \\ S_{x_2 y} & S_{x_2 x_2} \end{vmatrix} \quad (51)$$

$$b_2 = \frac{1}{\Delta} \begin{vmatrix} S_{x_1 x_1} & S_{x_1 y} \\ S_{x_1 x_2} & S_{x_2 y} \end{vmatrix}$$

wobei für den Ausdruck  $\Delta$  gilt

$$\Delta = \begin{vmatrix} S_{x_1 x_1} & S_{x_1 x_2} \\ S_{x_1 x_2} & S_{x_2 x_2} \end{vmatrix}$$

Die mehrfachen (partiellen) Regressionskoeffizienten messen die stochastische Abhängigkeit je zweier Veränderlicher unter Konstantsetzung der dritten Variablen, deren Einfluss dadurch bei der Analyse eliminiert wird. Man könnte deshalb auch setzen:

$$b_1 = b_{y x_1 \cdot x_2}$$

bzw.

$$b_2 = b_{y x_2 \cdot x_1}$$

##### Die Methode der Multiplikatoren:

(vgl. dazu E. Weber, Grundriss, S.345 ff.)

Wir bezeichnen die Multiplikatoren mit dem Buchstaben c und verwenden – bei drei unabhängigen Variablen – folgendes Gleichungssystem:

$$\begin{aligned}
c_{\cdot 1} S_{x_1 x_1} + c_{\cdot 2} S_{x_1 x_2} + c_{\cdot 3} S_{x_1 x_3} &= \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \\
c_{\cdot 1} S_{x_2 x_1} + c_{\cdot 2} S_{x_2 x_2} + c_{\cdot 3} S_{x_2 x_3} &= \\
c_{\cdot 1} S_{x_3 x_1} + c_{\cdot 2} S_{x_3 x_2} + c_{\cdot 3} S_{x_3 x_3} &=
\end{aligned} \tag{52}$$

Die Auflösung des Systems liefert die Matrix der Multiplikatoren, d. h. man erhält die Regressionskoeffizienten aus den Gleichungen

$$\begin{aligned}
b_1 &= c_{11} S_{x_1 y} + c_{12} S_{x_2 y} + c_{13} S_{x_3 y} \\
b_2 &= c_{12} S_{x_1 y} + c_{22} S_{x_2 y} + c_{23} S_{x_3 y} \\
b_3 &= c_{13} S_{x_1 y} + c_{23} S_{x_2 y} + c_{33} S_{x_3 y}
\end{aligned} \tag{53}$$

Für die Auswertung vgl. E.Weber, Grundriss, S.346f.: Schema für die Berechnung der partiellen Regressionskoeffizienten.

Die Methode der Multiplikatoren ist sehr gut geeignet, um zu einem bestimmten System (z. B. zwei unabhängige und eine abhängige Variable) **eine oder mehrere Variablen hinzuzufügen** (vgl. dazu die Arbeit von W.G.Cochran: The omission or addition of an independent variable in multiple linear regression; in Suppl. to the Journal of the Royal Statistical Society, Vol. V, 1938, S.171 ff.). Vor allem sind die Ansätze von Cochran geeignet für die Erstellung eines Elektronenrechner-Programmes.

### 143.2 Die Bestimmtheit

#### (1) Totales Bestimmtheitsmass

Für zwei unabhängige und eine abhängige Variable hat der Ansatz für das Bestimmtheitsmass (totale Bestimmtheit) folgende Form:

$$B_T = \frac{1}{S_{yy}} \{b_1 S_{x_1 y} + b_2 S_{x_2 y}\} \tag{54}$$

Die Deutung des Bestimmtheitsmasses kann selbstverständlich auch aus der Streuungszerlegung heraus erfolgen.

#### (2) Partielle Bestimmtheitsmasse

Setzen wir für

$$B_1 = B_{y x_1 \cdot x_2}$$

und

$$B_2 = B_{y x_2 \cdot x_1}$$

dann erhalten wir für die partiellen Bestimmtheitsmasse erster Ordnung:

$$\left. \begin{aligned}
B_1 &= \frac{B_{y x_1} - 2\sqrt{B_{y x_1} B_{y x_2} B_{x_1 x_2}} + B_{y x_2} B_{x_1 x_2}}{(1 - B_{y x_2})(1 - B_{x_1 x_2})} \\
B_2 &= \frac{B_{y x_2} - 2\sqrt{B_{y x_2} B_{y x_1} B_{x_1 x_2}} + B_{y x_1} B_{x_1 x_2}}{(1 - B_{y x_1})(1 - B_{x_1 x_2})}
\end{aligned} \right\} \tag{55}$$

Diese Bestimmtheitsmasse nach Ansatz (55) lassen sich auch darstellen in den Bestimmtheitsmassen nullter Ordnung.

### 143.3 Korrelationskoeffizient

Sowohl der totale wie der partielle Korrelationskoeffizient lassen sich aus der allgemeinen Formel

$$r = \sqrt{B}$$

herleiten. Auch hier lassen sich die partiellen Korrelationskoeffizienten erster Ordnung auf Koeffizienten nullter Ordnung zurückführen.

## 144 Prüfen von Hypothesen

### 144.1 Streuungszerlegung

Es wird vorerst generell zu entscheiden sein, ob eine mehrfache lineare Regression statthaft bzw. sinnvoll ist. Wir zerlegen zu diesem Zweck wiederum die Quadratsummen und gelangen zu folgendem Schema der Streuungszerlegung: (zwei unabhängige, eine abhängige Variable)

Streuung	SQ	FG	DQ
Auf Regression	$b_1 S_{x_1 y} + b_2 S_{x_2 y} = C$	2	$s_1^2 = \frac{1}{2} C$
Um Regression	$A - C = B$	$N - 3$	$s_2^2 = \frac{B}{N - 3}$
Insgesamt	$S_{yy} = \sum y_i^2 - \frac{1}{N} \left( \sum y_i \right)^2 = A$	$N - 1$	.

Der Testansatz lautet

$$F = \frac{\text{DQ (auf Regression)}}{\text{DQ (um Regression)}} = \frac{s_1^2}{s_2^2} \quad (57)$$

mit:  $n_1^* = 2$ ;  $n_2^* = (N-3)$  FG

Geprüft wird die Hypothese

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Entscheidungskriterium:

$F \geq F_p$ : Mehrfache lineare Regression ist zulässig.

### 144.2 Prüfen der partiellen Regressionskoeffizienten

Es bestehen dafür zwei grundsätzliche Möglichkeiten:

- (1) Prüfen der Abweichung des Regressionskoeffizienten  $b_1$  von einem Wert  $\beta_1$ ;
- (2) Prüfen, ob ein partieller Regressionskoeffizient wesentlich oder nur zufällig von Null abweicht.

In beiden Fällen wird ein t-Test als Prüfverfahren dienen, wobei die Multiplikatoren im Prüfansatz Verwendung finden.

Bei K.W. Smillie hat dieser Ansatz folgende Form (vgl. Introduction, S.48, Ansatz 3.4.3):

$$t = \frac{b_j - \beta_j}{s \sqrt{c_{jj}}} = \frac{b_j - \beta_j}{s_{b_j}}$$

Die Zahl der Freiheitsgrade beträgt  $n^* = N - p - 1$ .

Der Vergleich von t mit dem diesbezüglichen Tabellenwert zeigt, ob die formulierte Hypothese anzunehmen oder zu verwerfen ist.

### 144.3 Prüfen der Bestimmtheiten

#### (1) Mehrfache (totale) Bestimmtheit

Aus Ansatz (54) ergibt sich folgende Verallgemeinerung für das Bestimmtheitsmass:

$$B_T = \frac{1}{S_{yy}} \{ b_1 S_{x_1 y} + b_2 S_{x_2 y} + \dots + b_p S_{x_p y} \}$$

Als Prüfverfahren dient ein F-Test, der aus der Streuungszerlegung abgeleitet wird (vgl. A.Linder, Methoden)

$$F = \frac{B(N-p-1)}{(1-B)p} \quad (58)$$

mit:  $n_1^* = p$  und  $n_2^* = (N-p-1)$  FG

Der F-Wert wird mit dem benötigten Tafelwert verglichen. Wenn

$F > F_p$ : B ist wesentlich von Null verschieden.

Andernfalls ist B nur zufällig von Null verschieden.

## (2) Partielle Bestimmtheiten

Für partielle Bestimmtheiten (p-1)ter Ordnung ist für das Prüfen der Nullhypothese ebenfalls ein F-Test anzusetzen:

$$F = \frac{B(N-2p)}{(1-B)p} \quad (59)$$

mit:  $n_1^* = p$  und  $n_2^* = (N-2p)$  FG.

Aus Tabellen mit F-Verteilung ist der Tabellenwert zu entnehmen und mit dem errechneten F-Wert zu vergleichen.

## 145 Vertrauensgrenzen

Für die berechneten Regressionswerte kann man Vertrauensgrenzen bestimmen (vgl. dazu A. Linder, Methoden, S.196). Analog zu Ansatz (34.2) oben gilt auch hier:

$$Y \pm t_p s_Y$$

Die Streuung  $s_Y^2$  des Regressionswertes Y kann gemäss folgendem Ansatz berechnet werden:

$$s_Y^2 \sim s_2^2 \left\{ \frac{1}{N} + c_{11}(x_1 - \bar{x}_1)^2 + 2c_{12}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + \dots + 2c_{1p}(x_1 - \bar{x}_1)(x_p - \bar{x}_p) \right. \\ \left. + c_{22}(x_2 - \bar{x}_2)^2 + \dots + 2c_{2p}(x_2 - \bar{x}_2)(x_p - \bar{x}_p) \right. \\ \left. + \dots + 2c_{3p}(x_3 - \bar{x}_3)(x_p - \bar{x}_p) \right. \\ \left. + \dots + c_{pp}(x_p - \bar{x}_p)^2 \right\} \quad (60)$$

Vgl. zu (60) A. Linder, Statistische Methoden, 1964, S.196.

Wie in Ansatz (35) kann man an Stelle der Vertrauensgrenzen für Y diejenigen für einen Einzelwert ausrechnen; in der Klammer des Ausdruckes (60) ist in diesem Fall der Wert 1 hinzuzufügen.

mit:  $n^* = (N-p-1)$  Freiheitsgraden

$c_{ik}$  = Multiplikatoren

$s_2^2$  = DQ (um Regression)

Wie bereits weiter oben erläutert, wird auch hier  $s_Y^2$  und damit die Ungenauigkeit um so grösser, je mehr wir uns von den Durchschnitten der unabhängigen Variablen ( $\bar{x}_1, \bar{x}_2, \dots$ ) entfernen.

## 146 Unechte Variable

Es kann in der Praxis erforderlich sein, qualitative Variable in die Untersuchungen einzubeziehen. So ist bei betriebswirtschaftlichen Spitalanalysen die Sachlage gegeben, den Einfluss von Spezialabteilungen auf die Kosten pro Pflage tag abzuklären. Wir haben dann z. B. drei Einflussgrössen:

$x_1$  = Produktivität;

$x_2$  = Durchschnittliche Personalkosten;

$x_3$  = Spezialabteilungen mit der Setzung:

0 Keine Spezialabteilung

1 Es bestehen Spezialabteilungen.

Die statistische Basistabelle sieht dann folgendermassen aus:

Spital	Produktivität $x_1$	Personalkosten (Fr.) $x_2$	Spezialabteilung $x_3$	Betriebskosten pro Pflage-tag (Fr.) $y$
A	129	9438	0	32.67
B	83	10621	1	59.95
C	111	9995	0	46.77
.				
.				
.				

Für Einzelheiten vgl. die Monographie von K.W. Smillie, S.69 ff. (dummy variables). Diese Arbeit enthält auch ein durchgerechnetes Beispiel.

## 15 Mehrfache nichtlineare Regression und Korrelation

### 151 Allgemeines

Im Abschnitt 14 haben wir bei den Ausführungen über die Mehrfachkorrelation immer lineare Abhängigkeit vorausgesetzt. Wie dies die Autoren Ezekiel-Fox (vgl. Methods, 3rd ed. 1959, S.204 ff.) andeuten, kann die Annahme linearer Beziehungen bei verschiedenen Untersuchungen zu begrenzten Ergebnissen führen. Es ist daher gegeben, auch bei mehrfacher Regression mit nichtlinearen Ansätzen zu arbeiten.

### 152 Transformation

In Abschnitt 132 haben wir gezeigt, dass durch geeignete Transformationen die nichtlineare in lineare Regression übergeführt werden kann. Selbstverständlich ist dieses Verfahren auch bei mehrfacher nichtlinearer Regression anwendbar. Man wird also für eine bestimmte Variable den Übergang zu Logarithmen vornehmen.

### 153 Mehrfache nichtlineare Regression

Analog zu Abschnitt 133 können wir mit Hilfe einer Substitution das Problem zu lösen versuchen. Es sei z. B.

$$Y = \bar{y} + b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + b_3(x_2 - \bar{x}_2)^2$$

Die Abhängigkeit in bezug auf die zweite Variable sei also nichtlinear. Wir setzen nun

$$x_2^2 = x_3$$

Das führt uns zu folgender Gleichung:

$$Y = \bar{y} + b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + b_3(x_3 - \bar{x}_3)$$

Die Verwendung von Elektronenrechnern erspart die mühselige Rechenarbeit auf Tischrechnern.

## 16 Elektronenrechner - Programme

Heute stehen zur Berechnung der notwendigen Hilfszahlen, Parameter, Regressionswerte usw. Elektronenrechner zur Verfügung, die in einem Bruchteil des Zeitaufwandes, der früher für «Handarbeit» benötigt wurde, alle Ergebnisse bereitstellen. Es gibt in der Tat eine Reihe vorzüglicher Programme, die vor allem bei mehrfacher Regression gute Dienste leisten. Ein grosser Teil der untenstehenden Resultate wurde jedoch zu Kontrollzwecken zweifach errechnet, einmal auf dem Tischrechner, zum andern Mal auf dem Computer. Wir wollen uns hier nicht weiter mit Programmierung befassen; die diesbezüglichen Probleme werden in der Spezialliteratur eingehend dargestellt.



## 2 Die Kostenabhängigkeit in den bernischen Bezirksspitalern

### 21 Problemlage und Zielsetzung

Die Spitäler üben im Rahmen der Wirtschafts- und Gesellschaftsordnung eine ausserordentlich wichtige Funktion aus. Ihre Finanzgebarung und insbesondere die Kosten ihrer Dienstleistungen sind deshalb von grösster Bedeutung, da es ganz offensichtlich nicht gleichgültig ist, ob der Patient eine einigermassen einkommensadäquate Rechnung erhält oder ob die Kosten eines Krankheitsfalles für die Mehrheit der «Konsumenten von Spitalleistungen» untragbar sind. Aus diesem Grunde ist es sicher kein Luxus, wenn die Betriebsrechnungen der Spitäler von Zeit zu Zeit untersucht werden, wobei vor allem der Kostenseite grosse Aufmerksamkeit zu schenken ist. Im Vordergrund steht die **Frage, durch welche Einflussfaktoren die Höhe der Kosten bedingt ist**. Mit anderen Worten: Es ist die **Abhängigkeit der Kosten** von näher zu bestimmenden Faktoren zu analysieren. Die Problemlage für die vorliegende Untersuchung kann daher folgendermassen formuliert werden:

Durch **welche Faktoren** werden die Kosten der Bezirksspitäler beeinflusst, und welches ist die **Stärke des Einflusses** dieser Faktoren?

Das **Ziel der Arbeit** ist die Erfassung der Einflussfaktoren der Kosten der Spitalleistungen sowie die Berechnung der Stärke dieses Einflusses.

In bezug auf die **Methodik** ist zu erwähnen, dass wir uns des oben dargestellten Verfahrens der Regressions- und Korrelationsanalyse bedienen.

### 22 Theorie der Kostenabhängigkeit

#### 221 Grundsätzliches

Die Feststellung unterschiedlicher Kostenhöhe im Zeitablauf bzw. zwischen mehreren Bezirksspitalern ist wenig aufschlussreich; diese Information sagt uns nicht sehr viel. Weitaus wichtiger ist die Kenntnis der **Ursachen** von Kostenveränderungen im Laufe der Monate oder Jahre sowie verschiedener Kosten-niveaus der betrachteten Spitalbetriebe. Vor allem ist eine zweckmässige Budgetierung, Kostenplanung und Preispolitik bzw. jede **Einflussnahme auf die Kosten** erst dann möglich, wenn wir die Einflussfaktoren der Kosten bzw. die Richtung und Stärke ihrer Einwirkung kennen. Es ist daher die vordringlichste **Aufgabe einer Kostenanalyse, nach Funktionalzusammenhängen zwischen den Relativkosten und anderen betriebswirtschaftlichen Grössen zu suchen**.

Die Quantifizierung der Zusammenhänge kann auf folgende Art und Weise geschehen:

- (1) Für verschiedene Variable liegen die Daten in Form von **Zeitreihen** vor; die Parameter werden aus Zeitreihen bestimmt (Zeitvergleich).
- (2) Gegeben sind Einflussgrössen und abhängige Variable für eine Mehrzahl von Spitalbetrieben, wobei nur die Ergebnisse eines einzigen Zeitpunktes bzw. einer einzigen Zeitperiode (z. B. ein Jahr) vorliegen; die Parameter sind also das Resultat eines **Betriebsvergleiches**.
- (3) Es kann eine Kombination beider Verfahren in Erwägung gezogen werden.

Die vorliegende Untersuchung basiert auf einem Betriebsvergleich; die Daten beziehen sich auf die Betriebsrechnungen 1966 der oben genannten Spitäler.

Welches sind nun die Variablen, welche in die Berechnungen eingeschlossen werden sollen?

#### **Abhängige Variable:**

$y =$  Betriebskosten (1966) pro Pflage-tag in Franken  $= K_B/E$

#### **Einflussfaktoren** (unabhängige Variable):

In der Literatur spielt vor allem die Abhängigkeit vom Beschäftigungsgrad eine grosse Rolle. Wir betrachten mehrere Faktoren, wobei sich folgende Grössen als bedeutsam erwiesen haben:

- Produktivität (Arbeitsproduktivität);
- Marktpreise (Löhne, Nahrungsmittel, medizinisches Material usw.);
- Ausnützungsgrad;
- Betriebsgrösse;
- Durchschnittliche Aufenthaltsdauer der Patienten;
- Spezialabteilungen im Spital.

Es wird nun so vorzugehen sein, dass zuerst abhängige und unabhängige Variable einzeln in einfacher Regression verbunden werden. Anschliessend soll vermitteltst mehrfacher Regression und Korrelation der Einfluss mehrerer Grössen auf die relativen Betriebskosten untersucht werden. Es ist also die Hypothese der Abhängigkeit der Betriebskosten pro Pflgetag von den verschiedenen oben genannten Einflussfaktoren zu prüfen.

## 222 Die abhängige Variable

Wie bereits im vorhergehenden Abschnitt dargelegt, verwenden wir als abhängige Variable die « Betriebskosten pro Pflgetag ».

Die Betriebskosten setzen wir aus folgenden **Kostenarten** zusammen (vgl. Ziff. II der Betriebsrechnung):

- Personalkosten
- Allgemeine Verwaltungskosten
- Nahrungsmittel und Getränke
- Medizinisches Material
- Elektrizität, Wärme, Wasser
- Inventaranschaffungen und -unterhalt
- Unterhalt der Gebäude und des Umgeländes
- Übrige Betriebskosten

---

$$= \text{Total Betriebskosten (pro Periode)} = K_B$$

Im Spital gilt als **Leistungseinheit bzw. Kostenträger der Pflgetag**, d. h. das vom Patienten während eines Tages belegte Krankenbett, wobei wir uns bewusst sind, dass der Pflgetag keine homogene Leistung darstellt.

Es ist nun gegeben, die totalen Betriebskosten einer Periode zur Leistungseinheit, dem Pflgetag, in Beziehung zu setzen, d. h. es wird ermittelt, welcher Betrag von den in einer Abrechnungsperiode entstandenen Kosten auf eine Leistungseinheit entfällt. Diese Kosten pro Pflgetag können zwischen den einzelnen Spitalbetrieben sehr stark voneinander abweichen. Es gilt nun abzuklären, auf welche Einflussfaktoren die unterschiedliche Höhe der Leistungseinheitskosten zurückzuführen ist.

## 223 Die Einflussfaktoren

### 223.1 Die Produktivität

Die Autoren H. Ulrich und W. Hill sind der folgenden Meinung: « Für die Wirtschaftlichkeit der Betriebsleistung ist nicht die absolute Höhe der Gesamtkosten, sondern ihr Verhältnis zur Produktion, d. h. zur Betriebsleistung massgebend. Wie gross diese Leistung ist, hängt von der Organisation und der technischen Einrichtung des Betriebes, ganz besonders aber vom Einsatz jedes einzelnen Mitarbeiters ab... Dieses Verhältnis von Arbeitseinsatz und Produktion bezeichnet man als Beschäftigungsintensität oder **Produktivität** » (vgl. Ulrich/Hill, 1964, S. 68).

Die Produktivität ist eine ausserordentlich wichtige betriebswirtschaftliche Kennziffer; sie ist von zentraler Bedeutung. Gerade bei einem Dienstleistungsbetrieb kann es nicht gleichgültig sein, mit wieviel Arbeitskräften eine bestimmte Zahl von Leistungseinheiten – also Pflgetage – erstellt werden. Die Berechnung dieser Masszahl bietet nun aber einige Schwierigkeiten. Es würde jedoch zu weit führen, hier die ganze Problematik der Produktivitätsmessung darzulegen.

Die Produktivität, auch etwa Arbeitsergiebigkeit genannt, wird als Verhältnis zwischen Ergebnis der Produktion und Faktoreinsatz definiert. Man kann also im Zähler des Dezimalbruches die Menge der erzeugten Güter oder Dienstleistungen einer Zeitperiode aufführen, im Nenner dagegen die Quantität der Faktoren, die zur Erzeugung dieser Produktion notwendig waren. Sehr oft wird die Verhältniszahl Produktionsergebnis : Arbeitseinsatz gebildet; wir erhalten dann die sogenannte Arbeitsproduktivität. Es sei also:

$$\begin{aligned} P_A &= \text{Arbeitsproduktivität} \\ E &= \text{Produktionsergebnis} \\ A &= \text{Arbeitseinsatz} \end{aligned}$$

Der folgende Ansatz definiert die Arbeitsproduktivität:

$$P_A = \frac{E}{A} \quad (61)$$

Fourastié (La productivité, 1965, S.55) umschreibt die Produktivität als Teilungsergebnis einer Produktion durch einen der Produktionsfaktoren («La productivité est le quotient d'une production par l'un des facteurs de la production»).

Ausgangspunkt für die Ermittlung der Produktivität ist die Produktionsfunktion. Bei Verwendung der drei Produktionsfaktoren:

B = Boden  
A = Arbeit  
K = Kapital

kann das Ergebnis der Produktion als Funktion dieser drei Faktoren bezeichnet werden. Es ist also

$$E = f(B, A, K)$$

Welche Form hat nun diese Produktionsfunktion? In der Literatur findet man verschiedene Ansätze unterschiedlicher Güte. Eine interessante Lösung haben die beiden Amerikaner **C. W. Cobb** und **P. H. Douglas** gefunden (vgl. dazu: Cobb/Douglas, A Theory of Production, American Economic Review, 1928, Suppl., S.139–165). Die dort verwendete Produktionsfunktion enthält, abgesehen von der abhängigen Variablen P (= Produktion in physischen Einheiten) nur zwei Einflussgrößen:

L = Arbeit (= labor)  
C = Kapital (= capital)

Die Produktionsfunktion (gesamte Volkswirtschaft) hat nach den Ausführungen der zwei genannten Autoren folgende Form:

$$P = \beta L^k C^m \quad (62)$$

Cobb und Douglas haben ferner eine Vereinfachung vorgesehen, dergestalt dass

$$k + m = 1$$

Ansatz (62) erhält somit nachstehende Struktur:

$$P = \beta L^k C^{1-k}; \quad 0 < k < 1 \quad (63)$$

(vgl. dazu Cobb/Douglas, a. a. O., S.156).

Es ist sinnvoll, die Arbeits- bzw. Kapitalelastizität zu berechnen. Wir gehen aus von Ansatz (63).

(1) Arbeitselastizität  $e_L$ :

Die Arbeitselastizität kann folgendermassen definiert werden:

$$e_L = \frac{\delta P}{\delta L} \cdot \frac{L}{P}$$

Die partielle Ableitung ergibt nun

$$\frac{\delta P}{\delta L} = \beta k L^{k-1} C^{1-k}$$

Damit erhalten wir als Elastizität

$$e_L = \beta k L^{k-1} C^{1-k} \frac{L}{\beta L^k C^{1-k}} = k \quad (64)$$

(2) Kapitalelastizität  $e_C$ :

Analog zu Ziff.1 ist

$$e_C = \frac{\delta P}{\delta C} \cdot \frac{C}{P}$$

Wir verwenden ferner den Ausdruck

$$\frac{\delta P}{\delta C} = \beta L^k (1-k) C^{-k}$$

Daraus resultiert schliesslich für die Elastizität

$$e_c = \beta L^k (1-k) C^{-k} \frac{C}{\beta L^k C^{1-k}} = \underline{1-k} \quad (65)$$

Im Ansatz (63) stellen somit die beiden Exponenten  $k$  und  $1-k$  die **Produktionselastizitäten** in bezug auf die Faktoren Arbeit und Kapital dar.

Wir definieren ferner die **durchschnittliche Produktivität**:

$$\text{Arbeit: } D_L = \frac{P}{L} \quad (66)$$

$$\text{Kapital: } D_C = \frac{P}{C} \quad (67)$$

Es ist leicht einzusehen, dass dann die **Grenzproduktivitäten**  $G_L$  und  $G_C$  folgende Form haben:

$$\text{Arbeit: } G_L = \frac{\delta P}{\delta L} = k \frac{P}{L} \quad (68)$$

$$\text{Kapital: } G_C = \frac{\delta P}{\delta C} = (1-k) \frac{P}{C} \quad (69)$$

Damit wären einige wichtige Beziehungen aus der Produktionstheorie hergeleitet. Auf die Darstellung der sogenannten Wertproduktivität verzichten wir, da ihre Verwendung in der vorliegenden Untersuchung keine interessanten Ergebnisse liefern dürfte. Von praktischer Bedeutung sind die Ansätze (66) und (67), welche die Durchschnittsproduktivitäten definieren; vor allem Ansatz (66) werden wir in der Folge in die Berechnungen einschliessen.

**Die Messung** des Produktionsergebnisses sowie des Faktoreinsatzes, die im Spitalbetrieb einige Schwierigkeiten bereitet, kann auf verschiedene Art und Weise erfolgen. Wir verwenden wiederum Ansatz (61) mit der dort benützten Symbolik.

**Das Produktionsergebnis:** Es hat sich auf der Basis der zur Verfügung stehenden Unterlagen (Krankenanstaltenstatistik) gezeigt, dass der Pfl egetag geeignet ist, die Produktion zu messen (Damrau, 1957, S. 15, erklärt: «... the unit of performance or achievement is a patient day of care»). Man wird also im Zähler des Ansatzes (61) die Zahl der Patientenpfl egetage einer bestimmten Periode einsetzen. Es ist somit

$$\text{Produktionsergebnis: } E = \text{Zahl der Pfl egetage.}$$

Eine andere Möglichkeit besteht darin, die Zahl der ausgenützten Betten als Produktionsergebnis zu verwenden:

$$\text{Produktionsergebnis: } V = \text{ausgenützte Betten.}$$

**Der Arbeitseinsatz:** Hier stehen zwei Möglichkeiten im Vordergrund. Zur Berechnung der Arbeitsproduktivität können im Nenner verwendet werden:

- Die Zahl der Arbeitsstunden;
- Die Zahl der Arbeitskräfte (Beschäftigte).

Es ist klar, dass bei Zugrundelegung der Zahl der Arbeitskräfte die Arbeitszeit einen Einfluss auf das Ergebnis hat; dies vor allem bei Zeitvergleichen. Die Krankenanstaltenstatistik enthält keine Arbeitsstundenstatistik des Personals. Der Nenner von Ansatz (61) schliesst daher den Bestand der Arbeitskräfte auf den 31. Dezember des Jahres ein. Man hätte auch den Mittelwert zweier Bestände verwenden können; die Resultate wären aber unseres Erachtens nicht besser ausgefallen. Es ist also:

$$A = \text{Arbeitseinsatz} = \text{Zahl der Arbeitskräfte}$$

Die Arbeitsproduktivität kann folgendermassen berechnet werden:

$$P_E = \frac{E}{A} = \frac{\text{Zahl der Pflegetage}}{\text{Zahl der Arbeitskräfte}} \quad (70)$$

Setzen wir im Zähler die Zahl der ausgenützten Betten ein, so erhalten wir als Ausdruck der Produktivität

$$P_V = \frac{V}{A} = \frac{\text{Ausgenützte Betten}}{\text{Zahl der Arbeitskräfte}} \quad (70a)$$

Wir haben für unsere Analyse die Produktivität sowohl nach (70) wie auch nach (70a) berechnet.

**Der Kapitaleinsatz:** Die Messung des Kapitaleinsatzes ist im Spitalbetrieb ziemlich schwierig. Die Zahlen aus dem Vermögensausweis der Krankenanstaltenstatistik sind für den Betriebsvergleich nicht geeignet, so dass wir darauf verzichtet haben, die Kapitalproduktivität in unsere Berechnungen einzubeziehen.

Zu erwähnen ist abschliessend, dass eine faktorbezogene Produktivität nicht mehr aussagt, als sie ist, nämlich eine statistische Verhältniszahl; sie ist also keine Zurechnungsgrösse.

Es wäre ferner gegeben, in die Überlegungen folgende Grössen einzubeziehen:

Kapitalintensität:  $I = \frac{C}{A}$ , bzw.

Technischer Fortschritt: F

Mangels Unterlagen kann weder die eine noch die andere Grösse in den Berechnungen berücksichtigt werden.

### 223.2 Die Marktpreise

Es gilt grundsätzlich folgende Relation:

$$\text{Kosten} = \text{Menge} \times \text{Preis}$$

Auf dem Beschaffungsmarkt müssen für die Produktionsfaktoren Preise bezahlt werden. Im Zeitablauf können sich die Preise – und auch die bei der Leistungserstellung verbrauchten Mengen – ändern, so dass die Kosten variieren. Gleichfalls ist es möglich, dass die verschiedenen Spitalbetriebe für gleiche Produktionsfaktoren unterschiedliche Preise bezahlen. So wird kaum ein Betrieb genau gleiche Löhne ausrichten wie der andere. Auch wenn Abmachungen bestehen, bleibt ein gewisser Spielraum für Lohndifferenzen.

Als wichtigste Kosten figurieren in der Betriebsrechnung die Personalkosten, d. h. Löhne und Gehälter. Von erheblicher Bedeutung sind ebenfalls Nahrungsmittelkosten und Kosten für medizinisches Material. Leider fehlt eine detaillierte Lohn- und Preisstatistik. Wir haben deshalb für die einzelnen Betriebe als ziemlich groben Ersatz die durchschnittlichen Personalkosten pro Angestellten verwendet. Bei den Nahrungsmitteln und beim medizinischen Material wurde ein Index (Kantonsmittel = 100) errechnet, ebenfalls ein Vorgehen, das nicht gerade unseren Wünschen entspricht. Die übrigen Positionen (Verwaltungskosten, Elektrizität, Wasser usw.) haben wir vernachlässigt, da Löhne, Preise für Nahrungsmittel und medizinisches Material von ausschlaggebender Bedeutung sind.

### 223.3 Der Ausnützungsgrad

In der betriebswirtschaftlichen Terminologie spielt der sogenannte Beschäftigungsgrad eine grosse Rolle. Ulrich/Hill, 1964, geben folgende Begriffsumschreibung: «Der Beschäftigungsgrad drückt aus, in welchem Grade die bestehende Leistungskapazität des Betriebes ausgenutzt wird». Im Spitalbetrieb wird diese Grösse als Bettenausnützung oder kurz Ausnützung bezeichnet. Der Ansatz für die Berechnung dieser Relativzahl lautet

$$G_B(\%) = \frac{V}{Q} \cdot 100 \quad (71)$$

mit:  $V = E/365$  bzw.  $V = E/30$ .

Die Symbole haben folgende Bedeutung:

$G_B$  = Ausnützungsgrad der Betten (im Mittel der betrachteten Periode)

$V$  = Belegte (ausgenützte) Betten

$Q$  = Bettenkapazität (normaler Bettenbestand)

$E$  = Krankenpflegetage einer Periode (z. B. Jahr oder Monat).

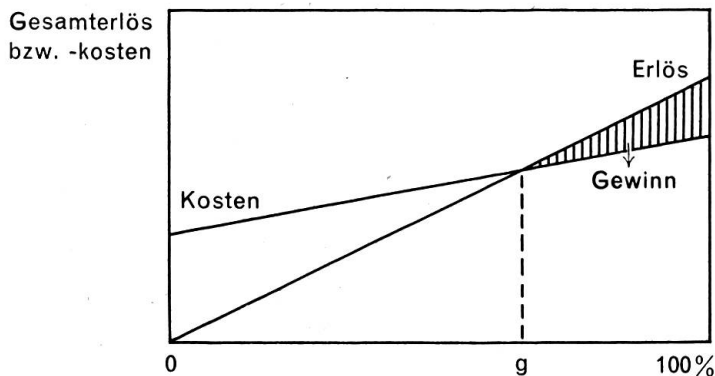
Wie verhalten sich nun die Kosten bei veränderlichem Ausnützungsgrad? Wir stellen also die Frage nach dem Einfluss des Ausnützungsgrades auf die Gesamtkosten bzw. Relativkosten, d. h. die Kosten pro Pflgetag. Zur Lösung dieses Problems müssen die Kosten in variable und fixe Kosten unterteilt werden. Es gibt also:

- Gesamtkosten, die sich bei unterschiedlichem Ausnützungsgrad verändern (aber pro Leistungseinheit, d. h. pro Pflgetag gleich bleiben); es sind dies die **variablen** (proportionalen) **Kosten**.
- Gesamtkosten, die bei verschiedenem Ausnützungsgrad gleich bleiben (aber pro Leistungseinheit, d. h. pro Pflgetag veränderlich sind); sie werden als **fixe Kosten** bezeichnet.

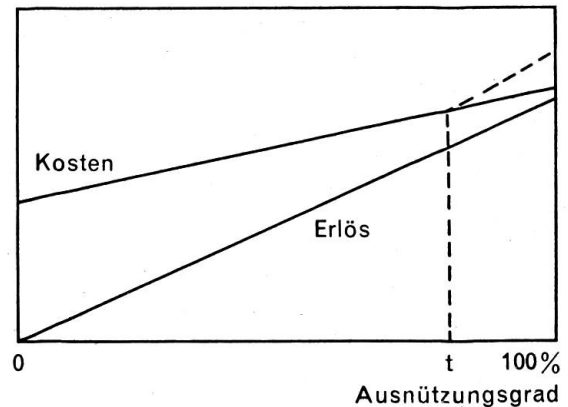
Es stellt sich nun sofort die Frage nach dem Verlauf bzw. der Form der Gesamtkostenkurve. Die neuere Kostentheorie ist der Auffassung, dass ein linearer Kostenverlauf angenommen werden kann. Möglich sei ein progressives Verhalten der Kosten an der obersten Grenze der Leistungskapazität. Beziehen wir die Erlöskurve in unsere Überlegungen ein, so können wir die in der Praxis vorliegenden Verhältnisse in folgenden Grafiken zur Darstellung bringen:

### I. Gesamtkosten und -erlös

a. Gewinnmöglicher Betrieb



b. Verlustbetrieb



$g$  = Gewinnschwelle (Breakeven-point)

Im Fall Ia) erzielt der Betrieb bis zur Erreichung der Gewinnschwelle  $g$  einen Verlust, bei höherem Ausnützungsgrad aber einen Gewinn. Es gilt also

$G_B < g$ : Verlust

$G_B > g$ : Gewinn

Bei den Bezirksspitalern treffen eher die in Figur Ib) dargestellten Verhältnisse zu. Steigt die Erlöskurve stärker als die Kostenkurve, so hat auch hier der Ausnützungsgrad einen günstigen Einfluss auf das Betriebsergebnis, wird doch das Defizit mit steigender Bettenbelegung normalerweise kleiner werden. Würde jedoch bei

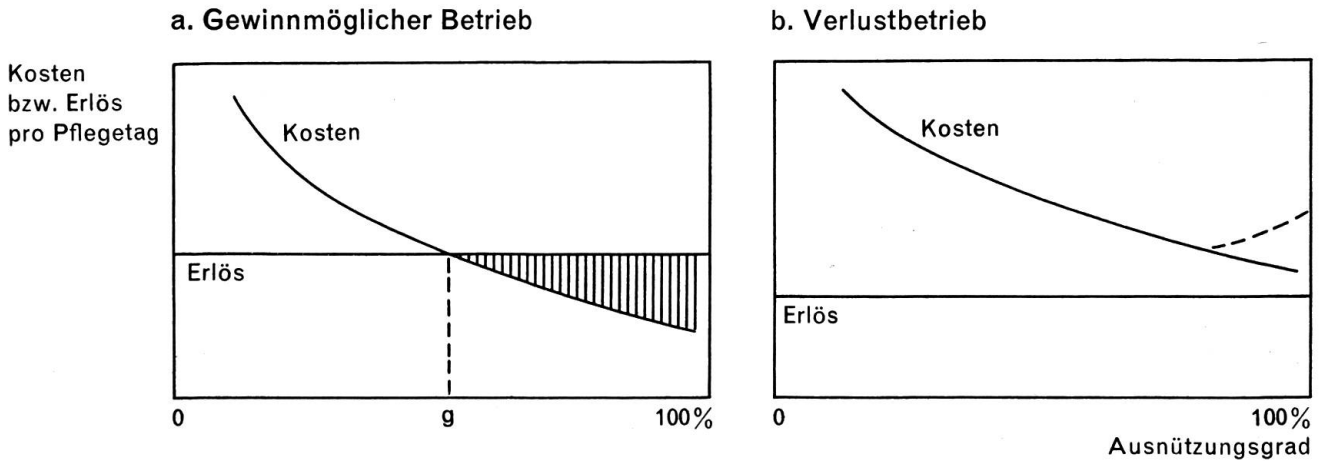
Ausnützung  $> t$

eine starke Progression einsetzen, müsste eine nahezu 100%ige Belegung wieder grössere Verluste bringen, als dies bei etwas geringerem Belegungsgrad der Fall gewesen wäre.

Für die Pflgetagkosten (bzw. Erlös pro Pflgetag) ergibt sich aus Fig. Ia und Ib folgender Verlauf für veränderlichen Ausnützungsgrad:



## II. Relativkosten/relativer Erlös



Es geht aus Grafik (a) bzw. (b) hervor, dass auch der gewinnmögliche Betrieb eine Verlustzone (Ausnutzung < g) hat. Gewinnmögliche Spitäler werden jedoch darnach trachten, die Bettenbelegung möglichst hoch zu halten. Auch für Verlustbetriebe wirkt sich ein Ausnutzungsgrad von 80–90% günstig aus.

### 223.4 Die Aufenthaltsdauer der Patienten

Es kann keine Zweifel daran geben, dass die Aufenthaltsdauer (= Verweildauer) auf die Gesamtkosten je Fall bzw. auf die Kosten pro Pflage-tag einen Einfluss hat. Auch hier wird man zwischen fixen und variablen Kosten unterscheiden müssen.

#### I. Fixkosten

##### a. Absolut fixe Kosten ( $K_{af}$ )

Verschiedene Fallkosten sind von der Aufenthaltsdauer des Patienten unabhängig, so z. B. Verwaltungsarbeiten bei Aufnahme und Entlassung, einmalig anfallende Operationskosten, Röntgenaufnahmen und EKG.

##### b. Intervall-fixe Kosten ( $K_{if}$ )

Zu dieser Kategorie gehören u. a. die Wäschekosten (z. B. wenn der Wechsel der Wäsche wöchentlich erfolgt).

#### II. Variable Kosten ( $K_v$ )

Es sind dies vor allem die Kosten der Pflege, der Verköstigung und der ärztlichen Betreuung der Kranken.

Generell gilt, dass mit steigender Aufenthaltsdauer die Gesamtkosten pro Pflage-tag sinken und umgekehrt, da sich die fixen bzw. intervall-fixen Kosten auf eine zunehmende Anzahl von Pflage-tag verteilen. Die in Abschnitt 223.3 dargestellten Grafiken (IIa) und (IIb) lassen sich analog auf das vorliegende Problem übertragen. Es ist nun noch die Frage der Messung der Patienten-Aufenthaltsdauer abzuklären. F. Gonzenbach gibt in seinen «Unterlagen», Aarau 1960, S. 112, folgende Begriffsumschreibung: «Die **durchschnittliche Aufenthaltsdauer der Patienten** ergibt sich aus dem Total der Krankenpflage-tag im Berichtsjahr geteilt durch die Gesamtzahl der in der gleichen Zeitspanne verpflegten Kranken».

Wir verwenden folgende Symbolik:

- D = Aufenthaltsdauer (durchschnittliche)
- E = Anzahl Krankenpflage-tag pro Periode
- R = Anzahl pro Periode verpflegte Kranke

Es ist also

$$D = \frac{E}{R} \quad (72)$$

In der Praxis stellt sich die Frage nach der Berücksichtigung der Mehrkosten bei kurzer und sehr kurzer Behandlung. Die Regressionsanalyse könnte, sofern vollständige Unterlagen zur Verfügung stehen, Aufschluss geben über die prozentuale Mehrbelastung (Kostenelastizität).

## 223.5 Die Betriebsgrösse

### 223.51 Kostenverlauf bei unterschiedlicher Betriebsgrösse

Die Problemlage lautet: Hat die Betriebsgrösse einen Einfluss auf das Kostenniveau, d. h. müssen vor allem je nach Grösse des Spitalbetriebes unterschiedliche Relativkosten in Rechnung gestellt werden? Es ist ziemlich schwierig, diese Frage zu beantworten; es kann daher keine allgemeingültige Antwort gegeben werden. Die betriebswirtschaftliche Theorie führt mit Berechtigung Gründe an, dass für zunehmende Betriebsgrösse Kostenersparnisse budgetiert werden können. Das Gegenteil ist aber auch möglich; grösserer Umfang des Betriebes kann unter bestimmten Umständen zu Kostenerhöhungen führen.

**Kostenersparnisse:** Sie können u. a. dann eintreten, wenn bei wachsender Betriebsgrösse die Bestellmengen anwachsen, auf denen dann Mengenrabatte gewährt werden. Weiter kann die Spezialisierung zu Kostenreduktionen führen; dies vor allem dann, wenn die verwendeten Produktionsmittel in hohem Masse ausgenützt sind.

**Kostensteigerungen:** Nach H. Koch (1959) führt eine Zunahme des Betriebsvolumens zu Kostensteigerungen auf Grund folgender Tatbestände:

- Disproportionierung zwischen dem Umfang der Unternehmungsleitung und dem Volumen des Produktionsmitteleinsatzes;
- Schwierigkeiten der Koordinierung;
- Abnahme der qualitativen Elastizität des Faktors menschliche Arbeit;
- Schwerfälligkeit des Leitungsapparates bei der Grossunternehmung.

Ein anderer Autor vertritt in bezug auf den Einflussfaktor «Betriebsgrösse» allerdings folgende Meinung: «Eine grössere Bettenzahl erweist sich nicht als Faktor, der die Kosten zum vornherein mit gewisser Gesetzmässigkeit erhöht und gleichsam als Kriterium bei einem Betriebskostenvergleich dienen könnte» (J. Bischofberger, 1965).

Interessant ist der Aufsatz von Dr. med. Hans Büchel, Kantonsarzt in Zürich (vgl. VESKA, Zeitschrift für das gesamte Krankenhauswesen, Nr. 11, 1957, S. 648 ff.: «Die rationelle Betriebsgrösse des Krankenhauses»). Er erwähnt darin Angaben der American Hospital Association aus dem Jahre 1956, basierend auf umfangreichem Material. Die von ihm wiedergegebene Tabelle lässt gewisse Tendenzen im Kostenverlauf nach Betriebsgrösse erkennen (vgl. H. Büchel, a. a. O. S. 653):

Betriebsgrösse	Ausgaben (in S) pro Pfl egetag
– 24 Betten	22.36
25– 49 Betten	19.76
50– 99 Betten	21.45
100–199 Betten	23.89
200–299 Betten	26.27
300–499 Betten	25.98
500 u. m. Betten	29.59

H. Büchel gibt dazu folgende Erläuterung (S. 653): «Die Betriebskosten pro Pfl egetag steigen mit der Bettenzahl, ausgenommen bei den Miniaturbetrieben, die ebenfalls teurer sind. Die Begründung findet sich fast ausschliesslich im Personalsektor.» Für den Kanton Zürich notiert er im Betriebsjahr 1956 ähnliches Verhalten der Kosten.

### 223.52 Massstab der Betriebsgrösse

Welche Messgrössen sind bei den Spezialbetrieben zur Quantifizierung des Betriebsvolumens zweckdienlich?

«Nur solche Begriffsinhalte dürfen als Massstäbe benützt werden, deren quantitative Änderung mit einer Variation der Einsatzmengen **sämtlicher Produktionsmittel** und nicht nur eines Teiles der Faktoren verbunden ist» (H. Koch).

Im Vordergrund stehen daher folgende zwei Grössen:

- Die **Bettenzahl** (angebotene Betten);
- Die Höhe des im Anlagevermögen investierten **Kapitals**.

Wir verfügen nur über Unterlagen in bezug auf die offerierten Betten (= normaler Bettenbestand gemäss Ziff. E der Krankenanstaltenstatistik, gezählt jeweils am 31. Dezember des betreffenden Jahres). Es wäre nützlich gewesen, die Höhe des Anlagevermögens zu kennen. Leider sind aber die Zahlen der untersuchten Betriebe nicht vergleichbar.

### 223.6 Spezialabteilungen, Behandlungsmethode

Wir haben den Versuch unternommen, die Spezialabteilungen der Bezirksspitäler als unechte Variable in die Regressionsanalyse einzubeziehen (vgl. Abschnitt 146 oben).

Unsere Überlegungen beruhen auf der Annahme, dass das Bestehen mehrerer Spezialabteilungen eine Kostensteigerung zur Folge habe. Entsprechend dieser Hypothese haben wir die unabhängige Veränderliche «Spezialabteilung» als unechte Variable gemäss nachstehender Setzung in unser System der Kostenabhängigkeit einbezogen:

0 keine/eine Spezialabteilung

1 zwei oder mehrere Spezialabteilungen

Sehr schwierig quantifizierbar ist die Behandlungsmethode bzw. die ärztliche Behandlungs- und Verordnungsweise. Auf Grund des zur Verfügung stehenden Materials ist es nicht möglich, diesen Einflussfaktor einzubeziehen. Uns scheint allerdings, dass S. Eichhorn (1957) richtig sieht, wenn er bemerkt: «Im allgemeinen kann man feststellen, dass neue und bessere Behandlungsmethoden eine Zunahme von Behandlungs- und Pflegeintensität bewirken. Das aber bedeutet: erhöhter Personaleinsatz, Anschaffung neuer medizinischer Apparate und Geräte, verstärkter, meist verteuerter Medikamentenverbrauch. In aller Regel ist also diese Art der qualitativen Änderung der Einsatzfaktoren – Verbesserungen in den Methoden der stationären Behandlung und Pflege von Kranken – mit einer kostenerhöhenden Tendenz verbunden.»

Um auf die Quantifizierung der Behandlungsmethode zurückzukommen: Es wäre auf Grund einer Statistik der Krankheitsarten die Hypothese zu prüfen, ob eine Abhängigkeit zwischen diesem Einflussfaktor und der Höhe der Betriebskosten pro Pfl egetag besteht.

### 223.7 Symbolik

Wir wollen in der Folge die verwendeten Symbole und ihre Bedeutung in einer Übersicht wiedergeben:

---

$K_B$	= Totale Betriebskosten einer Periode	
$E$	= Anzahl Krankenpfl egetage einer Periode (= Produktionsergebnis)	
$K_B/E$	= Betriebskosten pro Pfl egetag = T	

---

$E$	= Anzahl Krankenpfl egetage einer Periode	
$A$	= Zahl der Arbeitskräfte (Bestand)	
$V$	= Im Mittel belegte (ausgenützte) Betten	
$P_E$	= Pfl egetag-Produktivität der Arbeit (= $E/A$ )	(70)
$P_V$	= Betten-Produktivität der Arbeit (= $V/A$ )	(70a)

---

$L$	= Index der durchschnittlichen Personalkosten	
$N$	= Index der Nahrungsmittelkosten	
$M$	= Index der Kosten für medizinisches Material	

---

$Q$	= Bettenkapazität (normaler Bettenbestand)	
$G_B$	= Ausnützungsgrad {= $(V/Q) \cdot 100$ }	(71)

---

$R$	= Anzahl pro Periode verpflegte Kranke	
$D$	= Durchschnittliche Aufenthaltsdauer (= $E/R$ )	(72)

---

$B$	= Betriebsgrösse (Bettenzahl)	
$S$	= Spezialabteilungen	

---

Aus dieser Aufstellung geht hervor, dass für die Berechnung der Kostenabhängigkeit nachstehende Variable zu berücksichtigen sind:

Variable	Nr.	Symbol	Bezeichnung
(1) Abhängige		K <sub>B</sub> /E	Betriebskosten pro Pflgetag
(2) Einflussfaktor	1	P <sub>A</sub>	Produktivität der Arbeit
	2	G <sub>B</sub>	Ausnützungsgrad in %
	3	B	Betriebsgrösse
	4	D	Aufenthaltsdauer (Patienten)
	5	L	Durchschnittliche Personalkosten
	6	N	Nahrungsmittelpreise
	7	M	Medizinisches Material
	8	S	Spezialabteilungen

Für die Grössen 6 und 7 bestehen nur Approximativ-Zahlen. Die unabhängige Variable Nr.8 ist nur bedingt verwendungsfähig; die Alternative 0/1 hat substitutiven Charakter. Da sämtliche Spitäler fast ohne Ausnahme einen sehr hohen Ausnützungsgrad haben, wird auch dieser Einflussfaktor zu keinen interessanten Resultaten führen.

## 23 Die Erhebungsgrundlagen

Als Erhebungsgrundlage für unsere Berechnungen verwenden wir die Krankenanstaltenstatistik und Jahresrechnung der Bezirksspitäler für das Jahr 1966. Diese Formulare enthalten u. a. folgende Informationen:

- Personalbestand, Dienstage;
- Zimmer- und Bettenbestand;
- Zahl der Kranken und Pflgetage;
- Aufenthaltsdauer der Patienten;
- Bettenbesetzung;
- Betriebsrechnung (Ertrag und Aufwand);
- Vermögensausweis.

In die nachstehende Untersuchung werden nur die Betriebskosten einbezogen, da vor allem Zins- und Abschreibungskosten für den Betriebsvergleich ungeeignet sind. In bezug auf die erfassten Betriebe ist zu erwähnen, dass es sich um die 31 Bezirksspitäler des Kantons Bern handelt (ohne Tiefenau- und Zieglerspital Bern).

## 24 Kostenabhängigkeit in den Bezirksspitalern

### 241 Die einzelnen Einflussfaktoren

#### 241.1 Produktivität

Wir nehmen an, dass die Betriebskosten pro Pflgetag abhängig seien von der Arbeitsproduktivität, d. h. es sei

$$T (= K_B/E) = f(P)$$

Die Produktivität der Arbeit wurde sowohl nach Ansatz (70) wie auch nach (70a) berechnet. Es liegen also für 31 Bezirksspitäler die Zahlenwerte des Rechnungsjahres 1966 vor (für T, P<sub>E</sub> und P<sub>V</sub>). Wir wollen zuerst die Abhängigkeit der Betriebskosten von der Bettenproduktivität der Arbeit untersuchen:

Es ist also:

$$T = f(P_V)$$

Abhängige Variable y: Betriebskosten pro Pflgetag (T)

Einflussfaktor x: Produktivität der Arbeit (P<sub>V</sub>)

Gegeben sei folgende Tabelle:

Spital Nr.	Produktivität $x_i$	Betriebskosten in Fr. $y_i$
1	129	30.50
2	83	54.65
⋮	⋮	⋮
⋮	⋮	⋮

Wir bilden Spitalgruppen gleicher (ähnlicher) Produktivität und erhalten folgende Aufstellung:

Produktivitätsklasse	Häufigkeit $N_j$	Mittelwert	
		$\bar{x}_j$	$\bar{y}_j$
0 bis 75	1	66	69.15
76 bis 100	12	90	50.45
101 bis 125	12	115	40.18
126 bis 150	3	138	31.15
151 +	3	157	26.98

Wir erkennen, dass mit steigender Produktivität (Mittelwert) die durchschnittlichen Betriebskosten sinken. Wir stellen also die Hypothese auf, dass mit steigender Produktivität die Betriebskosten pro Pflegetag kleiner werden. Ferner nehmen wir linearen Zusammenhang zwischen den beiden Variablen an. Somit ist

$$Y = a - b_{yx}x$$

d. h. wir verwenden ein lineares Modell, wobei der Regressionsparameter  $b$  mit einem negativen Vorzeichen zu versehen sei.

Aus den Berechnungen resultieren folgende Hilfswerte:

$$\bar{x} = 110,00$$

$$\bar{y} = 42,9397$$

Für den Regressionskoeffizienten  $b$  erhalten wir

$$b_{yx} = -0,382582$$

Für die Niveaunkonstante  $a$  ergibt sich

$$a = 85,023742$$

Die **Regressionsgleichung** lautet somit

$$Y = 85,024 - 0,383x.$$

Sie ist definiert im Bereich

$$x_k = 66 \text{ und } x_g = 168$$

Die Niveaunkonstante hat hier den Charakter einer reinen Rechengrösse; denn es lässt sich kaum der Fall denken, dass die Produktivität gleich Null wäre.

Die Interpretation des Regressionskoeffizienten  $b_{yx} = -0,38$ : Steigt die Produktivität (ausgenützte Betten pro Angestellten) um eine Einheit, so sinken die Betriebskosten pro Pflegetag um 38 Rappen. Sinkt hingegen die Produktivität um eine Einheit, so steigen die Kosten um den gleichen Betrag.

Es ist sinnvoll, aus der Regressionsgleichung

$$Y = 85,024 - 0,383x$$

einige **Regressionswerte**  $Y_i$  zu bestimmen (z.B. für den kleinsten und grössten Wert sowie für den Mittelwert). Die Resultate dieser Berechnungen sind in der folgenden Tabelle eingetragen:

$x_i$	a	$b_{yx}x$	$Y_i$
$x_k = 66$	85,024	25,250	59,77
$\bar{x} = 110$		42,084	42,94
$x_g = 168$		64,274	20,75

Bestimmen wir für sämtliche Spitaler die Regressionswerte  $Y_i$ , so konnen wir anschliessend die Abweichung der  $(y_i - Y_i)$  bilden. Das ergibt, wenn wir diese Berechnungen fur einige Spitaler durchfuhren, folgendes Bild:

Spital	Beobachteter Wert $y_i$	Errechneter Wert $Y_i$	$y_i - Y_i$
Meiringen	30.50	35.67	-5.17
Interlaken	54.65	53.27	1.38
Frutigen	40.54	42.56	-2.02
Erlenbach	46.49	44.85	1.64
.			
.			
.			

Wir sehen sofort, dass

- positive und negative Abweichungen vorkommen;
- die Abweichungen unterschiedlich gross sind.

Diese beiden Tatsachen lassen es als zweckmassig erscheinen, einen **Streubereich der Einzelwerte** zu berechnen. Nach Ansatz (16) oben ist

$$s_{yx}^2 = \frac{1}{N-2} \left\{ S_{yy} - \frac{S_{yx}^2}{S_{xx}} \right\}$$

Als Ergebnis der Berechnungen erhalten wir

$$s_{yx}^2 = 28,870, \quad \text{bzw.}$$

$$s_{yx} = 5,373.$$

Mit diesem Wert  $s_{yx}$  messen wir das Ausmass der Ubereinstimmung zwischen den errechneten ( $Y_i$ ) und den beobachteten ( $y_i$ ) Werten der abhangigen Veranderlichen.

Ansatz (17) oben definiert das **Bestimmtheitsmass**

$$B = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

In unserem behandelten Fall betragt

$$B = 0,758$$

d.h. rund 76% der Streuung der abhangigen Veranderlichen (Betriebskosten pro Pflegetag) lasst sich aus der Variabilitat des Einflussfaktors Produktivitat erklaren. Der Korrelationskoeffizient betragt:

$$r = \sqrt{B} = -0,87$$



Auf Grund dieser beiden letzten Zahlen wird man folgern, dass die Betriebskosten in hohem Masse durch die Höhe der Produktivität bestimmt sind.

Aus der Regressionsgleichung lässt sich die **Elastizität der Kosten** errechnen. Wir bestimmen für einige Werte die sogenannte Punktelastizität. Sie ist gemäss folgendem Ansatz definiert:

$$P_v \epsilon_T = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{dT}{dP} \cdot \frac{P}{T} \quad (73)$$

Der Elastizitätskoeffizient  $\epsilon$  gibt an, um wieviel Prozent die Betriebskosten pro Pflgetag variieren, wenn die Produktivität eine Änderung von 1% erfährt.

Es ist nun offensichtlich

$$\frac{dT}{dP} = b_{yx} = -0,382\ 582$$

Wir errechnen für einige Werte der unabhängigen Variablen die Kostenelastizität.

Die Ergebnisse:

x	Y	$\frac{x}{Y}$	$P_v \epsilon_T$
66	59.77	1,104	-0,42
$\bar{x} = 110$	42.94	2,562	-0,98
168	20.75	8,096	-3,10

Die mittlere Elastizität liegt sehr nahe bei der Einheit, d. h. bei Variation der Produktivität um 1% verändern sich auch die Betriebskosten um 1%. Bei sehr guter Produktivität – dies geht aus der Tabelle hervor – steigen die Betriebskosten um einen ziemlich hohen Prozentsatz, sofern sich die Relation des Produktionsergebnisses zu den Arbeitskräften in ungünstiger Richtung verändert.

Wie wir in Abschnitt 122.1 gezeigt haben, kann es unter Umständen von Bedeutung sein, von den y-Werten auf die x-Werte, d. h. von den Betriebskosten auf die Produktivität zu schliessen. Die Regressionsgleichung lautet für diesen Fall:

$$X = 195,125 - 1,982 y.$$

Wir gehen nun über zu den **Prüfverfahren**, d. h. wir wollen verschiedene Hypothesen prüfen.

Vorerst soll vermittelst der Varianzanalyse geprüft werden, ob die Regression gesichert ist. Ausgangspunkt bildet Ansatz (21). Die numerische Auswertung führt zu folgenden Resultaten:

Streuung	SQ	FG	DQ
Auf Regression	2 628,7926	1	2 628,7926
Um Regression	837,2267	29	28,8699
Insgesamt	3 466,0193	30	.

Das führt zu folgendem F-Test:

$$F = \frac{2628,7926}{28,8699} = 91,056$$

mit:  $n_1^* = 1$ ,  $n_2^* = 29$  FG

Es ist daher:  $F_{0,001} = 13,391$

Daraus folgt:  $F \gg F_{0,001}$ .

Die Regression ist somit für eine Irrtumswahrscheinlichkeit von 1‰ gesichert.

Weiter prüfen wir den Regressionskoeffizienten, d. h. dessen Verschiedenheit von Null. Als Prüfansatz dient uns (26)

$$t = \frac{b_{yx}}{s_{yx}} \sqrt{S_{xx}}$$

mit:  $n^* = (N-2)$  Freiheitsgraden.

Wir prüfen also die Hypothese, ob

$$H_0: b_{yx} = 0$$

Diese Hypothese muss abgelehnt werden, sofern

$$\frac{|b_{yx}|}{s_{yx}} \sqrt{S_{xx}} > t_p$$

Die Hilfswerte betragen

$$b_{yx} = -0,382582$$

$$s_{yx} = 5,373071$$

$$S_{xx} = 17960$$

Es ist somit

$$t = \frac{0,382582}{5,373071} \cdot 134,015 = 9,542$$

Es ist weiter:  $t_{0,001;29} = 3,659$

Daher:  $t > t_{0,001}$

Die Folgerung lautet:

Der Regressionskoeffizient ( $b_{yx} = -0,383$ ) ist wesentlich von Null verschieden; die Hypothese des Einflusses der Produktivität auf die Betriebskosten pro Pflgetag muss akzeptiert werden.

Wir prüfen ferner, ob das Bestimmtheitsmass B wesentlich oder nur zufällig von Null verschieden ist. Es seien also folgende Hypothesen zu prüfen:

$$H_0: B = 0$$

$$H_A: B > 0.$$

Als Prüfansatz dient uns (27) oben

$$F = \frac{B(N-2)}{1-B} = \frac{0,758447 \cdot 29}{0,241553}$$

$$F = 91,056; \quad F_{0,001} = 13,391$$

Es ist daher  $F > F_{0,001}$

Folgerung: Das Bestimmtheitsmass ist wesentlich von Null verschieden;  $H_0$  ist zu verwerfen.

Das letzte Prüfverfahren bezieht sich auf die **Form der Beziehung** zwischen den beiden Variablen Produktivität und Betriebskosten, d. h. wir prüfen, ob lineare Regression zulässig sei. Zur Testung der Linearitäts-Hypothese verwenden wir die Ansätze (29) und (30) oben.

Wir führen die Streuungserlegung in zwei Stufen durch.

### 1. Stufe der Streuungserlegung:

Streuung	SQ	FG	DQ
Zwischen Spalten	3050,5184	24	$s_{III}^2 = 69,2502$
Innerhalb Spalten	415,5009	6	
Insgesamt	3466,0193	30	.

In den F-Test wird die Streuung eingehen

$$s_{III}^2 = 69,2502$$

Es sind nun weiter die SQ (zwischen den Spalten) aufzuteilen. Dies führen wir in der zweiten Stufe der Streuungszerlegung durch.

## 2. Stufe der Streuungszerlegung:

Streuung	SQ	FG	DQ
Auf Regression	2628,7926	1	$s_I^2 = 2628,7926$
Um Regression	421,7258	23	$s_{II}^2 = 18,3359$
Zwischen Spalten	3050,5184	24	

Wir vereinigen  $s_{II}^2$  und  $s_{III}^2$  im F-Test:

$$F = \frac{s_{II}^2}{s_{III}^2} = \frac{18,3359}{69,2502}; \quad F < 1$$

Verwenden wir den Reziprokwert von F mit den entsprechenden Freiheitsgraden, so darf für eine Irrtumswahrscheinlichkeit von 1% Linearität angenommen werden. Es heisst dies nicht, dass unbedingt lineare Regression als beste Anpassung zu gelten hat.

Wir haben nachträglich doch noch nichtlineare Regression angesetzt von der Form

$$Y = a - b \log x$$

und in der Tat gefunden, dass dieser Ansatz den Verhältnissen besser entspricht. Das neu errechnete Bestimmtheitsmass beträgt 0,772, ist also nur um ein wenig grösser als dasjenige aus linearer Regression. Da die Verbesserung nur 1,4 Prozentpunkte beträgt, werden wir bei der mehrfachen Regression mit linearen Ansätzen arbeiten.

Bereits weiter oben wurde erwähnt (vgl. Abschnitt 124), dass für die Schätzwerte zufällige Schwankungen berücksichtigt werden müssen, d.h. es sind die **Vertrauensgrenzen der Schätzung** zu bestimmen. Wir berechnen diese Grenzen vorerst für den Regressionskoeffizienten  $b_{yx}$ ; dies geschieht unter Verwendung des Ansatzes (31):

$$b_{yx} \pm t_p \frac{s_{yx}}{\sqrt{S_{xx}}}; \quad n^* = N - 2 \text{ FG}$$

Die Hilfwerte betragen

$$s_{yx} = 5,373 \text{ 07}; \quad t_{0,05;29} = 2,045$$

$$S_{xx} = 17960; \quad b_{yx} = -0,3826$$

$$t_{0,05;29} \frac{s_{yx}}{\sqrt{S_{xx}}} = 2,045 \cdot 0,040 \text{ 093} = 0,0820$$

Es ist daher:  $-0,3826 \pm 0,0820$ .

$$\underline{b_u = -0,3006}; \quad \underline{b_o = -0,4646}$$

Vertrauensgrenzen für die **Regressionswerte**  $Y_i$ :

Ausgangspunkt bildet Ansatz (33):

$$s_Y^2 \sim s_{yx}^2 \left\{ \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right\}$$

Die Hilfwerte:

$$N = 31; \quad 1/N = 0,032 \text{ 258}$$

$$\bar{x} = 110; \quad s_{yx}^2 = 28,869 \text{ 887}$$

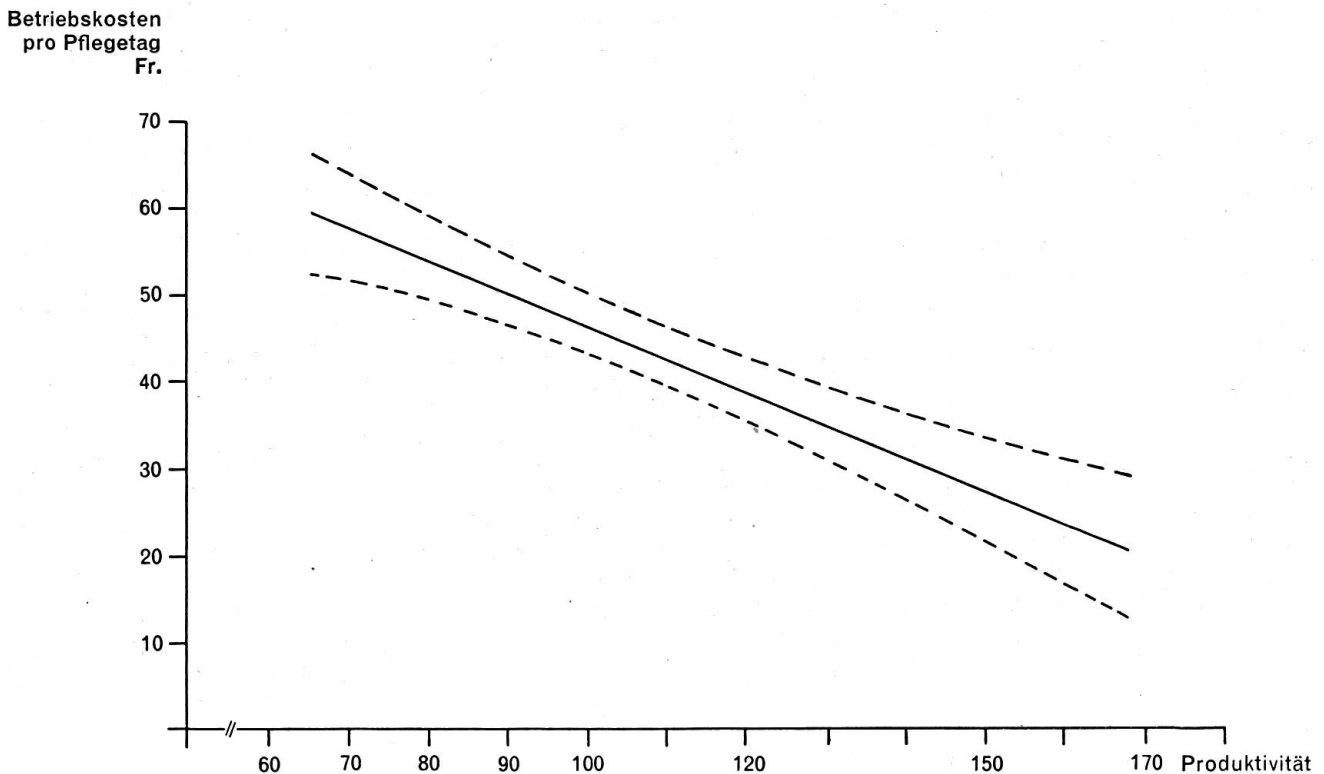
$$S_{xx} = 17960$$

Die Berechnungen wurden mit Hilfe eines FORTRAN-II-Programms auf dem Elektronenrechner der Universität Bern durchgeführt.

Die Ergebnisse unter Berücksichtigung einer Irrtumswahrscheinlichkeit von 1‰ lauten

$x_i$	$s_Y$	$s_Y t_{0,001}$	$Y_i$	Vertrauensgrenzen	
				untere	obere
(0)	(1)	(2)	(3)	(4)	(5)
66	2,011	7,36	59,77	52,41	67,13
70	1,872	6,85	58,24	51,39	65,09
80	1,542	5,64	54,42	48,78	60,06
90	1,255	4,59	50,59	46,00	55,18
100	1,045	3,82	46,77	42,95	50,59
110	0,965	3,53	42,94	39,41	46,47
120	1,045	3,82	39,11	35,29	42,93
130	1,255	4,53	35,29	30,70	39,88
140	1,542	5,64	31,46	25,82	37,10
150	1,872	6,85	27,64	20,79	34,49
160	2,225	8,14	23,81	15,67	31,95
168	2,518	9,21	20,75	11,54	29,96

Tragen wir die Zahlenwerte aus den Spalten (3), (4) und (5) in einer Grafik auf, so ergibt sich das Bild der stochastischen Abhängigkeit der Betriebskosten von der Produktivität:



Fassen wir die Ergebnisse unserer bisherigen Berechnungen kurz zusammen:

Gegeben waren die beiden Variablen:

$y$  = Betriebskosten pro Pflage tag (in Fr.) 1966

$x$  = Produktivität der Arbeit.

Wir haben einen linearen Ansatz zur Anwendung gebracht, wobei die Auswertung folgende Regressionsgleichung lieferte:

$$Y = 85,024 - 0,383 x$$

Aus dieser Gleichung lassen sich Regressionswerte bzw. Abweichungen der beobachteten von den errechneten Werten gewinnen. Das Bestimmtheitsmass B als Masszahl der Stärke des Zusammenhangs beträgt 0,76, ein Wert, der stark gesichert ist.

Als weiteres Resultat der Berechnungen erhielten wir die Kostenelastizität, die im Mittel  $-1,0$  beträgt. Für kleine Werte der unabhängigen Variablen ist sie kleiner, für grosse Werte erheblich grösser. Es lassen sich aus dieser Tatsache leicht Schlüsse für die einzelnen Spitalbetriebe ziehen.

Die **Prüfverfahren** ergeben:

Varianzanalyse: Regression stark gesichert.

Regressionskoeffizient: Verschieden von Null.

Bestimmtheitsmass: Verschieden von Null.

Form der Beziehung: Linearität erlaubt.

Aus den bisher aufgeführten Resultaten sind für die Anwendung in der Praxis von Bedeutung:

- Regressionskoeffizient (Grenzneigung);
- Elastizität der Kosten.

Die Regressionsgleichung lässt sowohl **Interpolation** als auch – unter entsprechenden Vorsichtsmassnahmen – **Extrapolation** zu. Vor allem die Auswirkungen der Abnahme der Produktivität auf die Betriebskosten können beurteilt werden. Es sei in diesem Zusammenhang noch einmal auf die zentrale Stellung der Produktivität bei Dienstleistungsbetrieben hingewiesen.

So weit die Analyse der Ergebnisse für stochastische Abhängigkeit zwischen Betriebskosten pro Pflege-tag (T) und Produktivität ( $P_v$ ).

Wir haben weiter oben gezeigt, dass die Produktivität auch durch Inbeziehungsetzung der Zahl der Pflegetage mit der Zahl der Arbeitskräfte errechnet werden kann. Daraus resultierte Ansatz (70). Auf dem Computer wurde die Regression ausgewertet zwischen

$y$  = Betriebskosten pro Pflegetag

$x$  = Pflegetag-Produktivität (=  $E/A$ )

Für diesen Fall wurde ein nichtlinearer Ansatz gewählt:

$$Y = a - b \log x$$

Die Auswertung liefert folgende Werte:

$a = 294,424$

$b = -97,010$

Die Regressionsgleichung lautet somit

$$Y = 294,424 - 97,010 \log x$$

Das Bestimmtheitsmass beträgt

$$B = 0,765$$

Auch die Verwendung der Pflegetage als Ausdruck des Produktionsergebnisses führt zu positiven Resultaten in bezug auf die Abhängigkeit der Betriebskosten pro Pflegetag von der Produktivität. Es kann also wahlweise die Produktivität auf Grund der Pflegetage bzw. der ausgenützten Betten berechnet werden.

## 241.2 Die Marktpreise

Im Spitalbetrieb haben drei Preiskategorien grosses Gewicht:

- Löhne des Personals,
- Preise für Nahrungsmittel,
- Preise für medizinisches Material.

### (1) Die Löhne

Wie bereits erwähnt, fehlt eine vergleichbare, detaillierte Lohnstatistik. Für alle 31 Bezirksspitäler wurden daher Mittelwerte errechnet, indem das Total der Personalkosten durch den Angestelltenbestand dividiert wurde. Anschliessend haben wir eine lineare Regression angesetzt mit:

$y$  = Betriebskosten pro Pflergetag (in Fr.)  
 $x$  = Personalkosten 1966 pro Angestellten.

Die Regressionsgleichung:

$$Y = 5,536 + 0,0037 x$$

Steigen also die Löhne des Personals um 1000 Franken pro Jahr, so erhöhen sich die Betriebskosten pro Pflergetag um Fr.3.70.

Die Regression ist gesichert für eine Irrtumswahrscheinlichkeit von 1%. Das Bestimmtheitsmass hat den Wert von

$$B = 0,24$$

d.h. nur rund 24% der Streuung der Betriebskosten lässt sich aus der Variabilität der durchschnittlichen Personalkosten erklären.

### (2) Preise für Nahrungsmittel

Auch hier liegt für die einzelnen Spitalbetriebe keine Statistik der Preise bzw. der verbrauchten Mengen vor. Wir haben als Ersatz einen Index (Kantonsmittel = 100) errechnet und diesen zu den Betriebskosten in Beziehung gebracht.

$y$  = Betriebskosten pro Pflergetag (in Fr.) 1966;  
 $x$  = Index der Nahrungsmittel.

Die Regressionsgleichung lautet

$$Y = 5,474 + 0,375 x$$

Wie zu erwarten war, steigen bei zunehmender Belastung durch die Nahrungsmittel auch die Betriebskosten.

Das Bestimmtheitsmass beträgt:  $B = 0,44$ . Die Regression ist gesichert für eine Wahrscheinlichkeit von 1‰, immerhin ein Resultat, das sich sehen lässt.

### (3) Preise für medizinisches Material

Auch hier besteht die gleiche Sachlage wie bei den Nahrungsmitteln; detaillierte Unterlagen stehen nicht zur Verfügung. Stellvertretend für die wünschbaren Daten wurde wiederum ein Index berechnet (Mittel von 31 Spitälern gleich 100 gesetzt).

Die verwendeten Variablen:

$y$  = Betriebskosten pro Pflergetag (in Fr.) 1966;  
 $x$  = Index des medizinischen Materials.

Die Regressionsgleichung lautet

$$Y = 21,857 + 0,211 x$$

d.h. wir haben einen linearen Ansatz benützt.

Das Bestimmtheitsmass:  $B = 0,690$ .

Wir haben noch mit Hilfe der Varianzanalyse geprüft, ob die Regression gesichert ist. Unter Verwendung des Ansatzes (21) erhalten wir für die numerische Auswertung:

Streuung	SQ	FG	DQ
Auf Regression	2 390,403 390	1	$s_1^2 = 2 390,403$
Um Regression	1 075,615 907	29	$s_2^2 = 37,090$
Insgesamt	3 466,019 297	30	.

Der F-Test ergibt

$$F = \frac{s_1^2}{s_2^2} = \frac{2390,403}{37,090} = 64,448$$

Es ist  $F_{0,001} = 13,391 < F$ .

**Folgerung:** Die Regression ist für eine Wahrscheinlichkeit von 1‰ gesichert.

Damit hätten wir die Untersuchungen in bezug auf die wichtigsten Beschaffungspreise für Spitalbetriebe abgeschlossen. Die Situation kann bei den Löhnen einigermaßen befriedigen, bei den Preisen für Nahrungsmittel und medizinisches Material jedoch kaum. Für künftige Untersuchungen wird es notwendig sein, die erforderlichen Lohn- und Preisstatistiken zu erstellen.

### 241.3 Der Ausnützungsgrad

In Ansatz (71) wurde der Ausnützungsgrad definiert:

$$G_B(\%) = \frac{V}{Q} \cdot 100$$

Wir suchen nach einer Abhängigkeit zwischen dieser Grösse und den Betriebskosten pro Pflgetag:

y = Betriebskosten pro Pflgetag (in Fr.) 1966

x = Ausnützungsgrad in Prozent.

Im Jahr 1966 ergeben die Berechnungen, dass von den 31 Bezirksspitalern nicht weniger als 30 eine Bettenausnützung ausweisen, die über 50% liegt, wobei 25 Betriebe einen Beschäftigungsgrad von mehr als 60% hatten. Es ist daher gemäss den Darlegungen in Abschnitt 223.3 zu vermuten, dass auf Grund des vorliegenden Materials eine Regression und Korrelation keine verwertbaren Ergebnisse liefern wird. Dies ist in der Tat der Fall. Die Regressionsgleichung:

$$Y = 43,996 - 0,015 x$$

Mit steigendem Ausnützungsgrad sinken also die Betriebskosten pro Pflgetag, ein Resultat, das die theoretischen Aussagen erhärten könnte. Allerdings zeigen die Prüfverfahren (Varianzanalyse), dass die Regression nicht zulässig ist. Das Bestimmtheitsmass B hat nur den Wert  $B = 0,00025$ . Es kann gefolgert werden, dass die Einflussgrösse « Ausnützungsgrad » bei der vorliegenden Analyse nicht eingeschlossen werden darf.

### 241.4 Die Aufenthaltsdauer der Patienten

Es war oben im Ansatz (72) die Relation gegeben

$$D = \frac{E}{R}$$

mit:

E = Anzahl Krankenpflgetage

R = Anzahl verpflegte Kranke.

Vernachlässigen wir die intervallfixen Kosten, so können wir nach einer linearen, evtl. nichtlinearen Regression suchen zwischen:

y = Betriebskosten pro Pflgetag (in Fr.) 1966

x = Durchschnittliche Aufenthaltsdauer der Patienten.

Diese Abhängigkeit könnte unter Umständen für die zu verfolgende Preispolitik von Bedeutung sein, indem Patienten mit kurzer Aufenthaltsdauer möglicherweise mit etwas höheren Tarifen zu belasten wären als Spitalbenützer, die längere Zeit behandelt werden.



Wir nehmen vorerst lineare Abhängigkeit an und erhalten folgende Regressionsgleichung:

$$Y = 66,368 - 1,328 x$$

Unsere in Abschnitt 223.4 vertretene Theorie, die Kosten pro Pflgetag seien abhängig von der Pflegedauer, wobei steigende Behandlungszeit sich in einer Reduktion der Kosten geltend mache, scheint auf Grund der analysierten Daten der Bezirksspitäler aus dem Jahr 1966 richtig zu sein.

Das Bestimmtheitsmass:  $B = 0,444$ , d. h. rund 44% der Streuung der Betriebskosten lässt sich aus der Variabilität des Einflussfaktors Aufenthaltsdauer (bei einfacher linearer Regression) erklären.

Die Varianzanalyse gemäss (21) ergibt

Streuung	SQ	FG	DQ
Auf Regression	1 539,197 977	1	1 539,197 977
Um Regression	1 926,821 320	29	66,442 114
Insgesamt	3 466,019 297	30	.

Wir schreiten zum F-Test

$$F = \frac{1539,198}{66,442} = 23,166$$

$$F_{0,001} = 13,391 < F$$

**Folgerung:** Die Regression (Betriebskosten/Aufenthaltsdauer) ist für eine Irrtumswahrscheinlichkeit von 1‰ gesichert.

#### 241.5 Die Betriebsgrösse

In Abschnitt 223.51 haben wir gezeigt, dass die Betriebsgrösse sich in unterschiedlichem Ausmass auf die Höhe der Kosten auswirken kann. Bei veränderlichem – zunehmendem – Betriebsvolumen sind sowohl Kostenersparnisse wie Kostensteigerungen möglich. Im Aufsatz von H. Büchel (1957) wird die Auffassung vertreten, dass die Betriebskosten pro Pflgetag mit der Bettenzahl steigen, wobei Kleinbetriebe ebenfalls teurere Dienstleistungen haben.

Es wird daher gegeben sein, vorerst mit einem **nichtlinearen Ansatz** von der Form

$$Y = a - bx + cx^2$$

zu arbeiten. Zur Lösung des Problems dient die Methode der mehrfachen linearen Regression (vgl. Abschnitt 133 oben). Aus den Berechnungen geht die folgende Regressionsgleichung hervor:

$$Y = 39,192697 - 0,000006147 x_1 + 0,00000002624 x_2$$

Die Streuungszersetzung zeigt, dass die berechnete Regressionskurve einigermaßen mit den beobachteten Werten verträglich ist; hingegen wirkt dieser nichtlineare Ansatz nicht sehr überzeugend.

Möglicherweise kann die lineare Approximation bessere Ergebnisse zeitigen. Wir gehen daher zu einem linearen Ansatz über, wobei ebenfalls:

$y$  = Betriebskosten pro Pflgetag (in Fr.) 1966;

$x$  = Betriebsgrösse (Zahl der Betten).

Der Computer liefert folgende Resultate (Regressionsgleichung):

$$Y = 35,693 + 0,059 x$$

Das Bestimmtheitsmass beträgt:  $B = 0,253$ . Die Varianzanalyse gibt den Hinweis, dass die Regression mit einer Irrtumswahrscheinlichkeit von 1% gesichert ist.

Die Regressionsgleichung lässt sich derart interpretieren, dass mit steigender Bettenzahl die Betriebskosten pro Pflage tag steigen. Die Kostenwirkung ist nicht sehr kräftig, steigen doch die Relativkosten bei Erhöhung des Betriebsvolumens um 10 Betten nur um durchschnittlich 60 Rappen. Immerhin muss die Hypothese steigender Einheitskosten für zunehmende Bettenzahl akzeptiert werden.

### 241.6 Zusammenstellung der Ergebnisse

In der nachstehenden Tabelle sollen die Ergebnisse der bisherigen Untersuchung zusammengestellt werden (nur lineare Regression und Korrelation).

Als abhängige Variable figurieren stets die Betriebskosten pro Pflage tag für das Jahr 1966. Es ist also

$y$  = Betriebskosten pro Pflage tag (in Fr.).

Einfache lineare Regression und Korrelation:  
(Bezirksspitäler Kt. Bern;  $N = 31$ )

Unabhängige Variable		Parameter		B	Varianzanalyse	
Nr.	Bezeichnung	a	b		F-Wert	Wertung
x <sub>1</sub>	Produktivität	8,5024	-0,383	0,76	91,1	***
x <sub>4</sub>	Medizin. Material	21,857	0,211	0,69	64,4	***
x <sub>3</sub>	Nahrungsmittel	5,474	0,375	0,47	26,0	***
x <sub>6</sub>	Aufenthaltsdauer	66,368	-1,328	0,44	23,2	***
x <sub>7</sub>	Betriebsgrösse	35,693	0,059	0,25	9,8	**
x <sub>2</sub>	Personalkosten	5,536	0,0037	0,24	9,2	**
x <sub>5</sub>	Ausnutzungsgrad	43,996	-0,0149	0,00025	F < 1	(-)

In der Spalte «Wertung» bedeuten die Symbole

- \*\*\* gesichert für Wahrscheinlichkeit von 1<sup>0</sup>/<sub>100</sub>;
- \*\* gesichert für Wahrscheinlichkeit von 1 %;
- (-) Regression nicht gesichert.

Auf Grund dieser Resultate müsste man für sämtliche Einflussfaktoren mit Ausnahme des Ausnutzungsgrades einfache lineare Regression zwischen den sechs aufgeführten unabhängigen Variablen und den Betriebskosten pro Pflage tag annehmen. Vor allem steht die Produktivität (x<sub>1</sub>) mit einem Bestimmtheitsmass von  $B = 0,76$  im Vordergrund. Die durchschnittlichen Personalkosten scheinen bei den betrachteten Betrieben einen nicht unwesentlichen Einfluss auszuüben; allerdings sind nur 24% des Variabilitätsanteils durch diese Grösse erklärt. Soweit die Resultate aus der Untersuchung der einzelnen Abhängigkeiten mit linearer Regression und Korrelation.

Es könnte nun vermutet werden, dass zwischen einzelnen Einflussfaktoren eine Abhängigkeit besteht. Wir werden bei der Behandlung der mehrfachen Regression und Korrelation auf dieses Problem zurückkommen.

### 242 Interkantonaler Vergleich

Es ist nicht uninteressant, die Zahlen der bernischen Bezirksspitäler mit denjenigen anderer Kantone zu vergleichen. Vor allem ist es sinnvoll, die Regressionsparameter für eine stochastische Kostenabhängigkeit mit folgenden Variablen zu berechnen und nebeneinanderzustellen:

$y$  = Totalkosten pro Pflage tag (in Fr.);  
 $x$  = Arbeitsproduktivität.

Es wird nützlich sein, Regressionsrechnung und Streuungszerlegung zu verbinden und vermittelst der **Kovarianzanalyse** Regressionskoeffizienten und Durchschnitte zu vergleichen.

Die Vergleichswerte aus der einfachen linearen Regression betragen:

Grösse	Kt. Bern	Ostschweiz	Vereinigt
N	31	16	47
a	91,176	116,951	103,392
b	-0,407	-0,609	-0,501
B	0,75	0,83	0,78
F-Wert	88,056	68,233	154,575
F <sub>0,001</sub>	13,4	17,1	12,3
Wertung	***	***	***

Die Regressionsgleichungen:

Kanton Bern:  $Y = 91,2 - 0,407 x$

Ostschweiz:  $Y = 116,9 - 0,609 x$

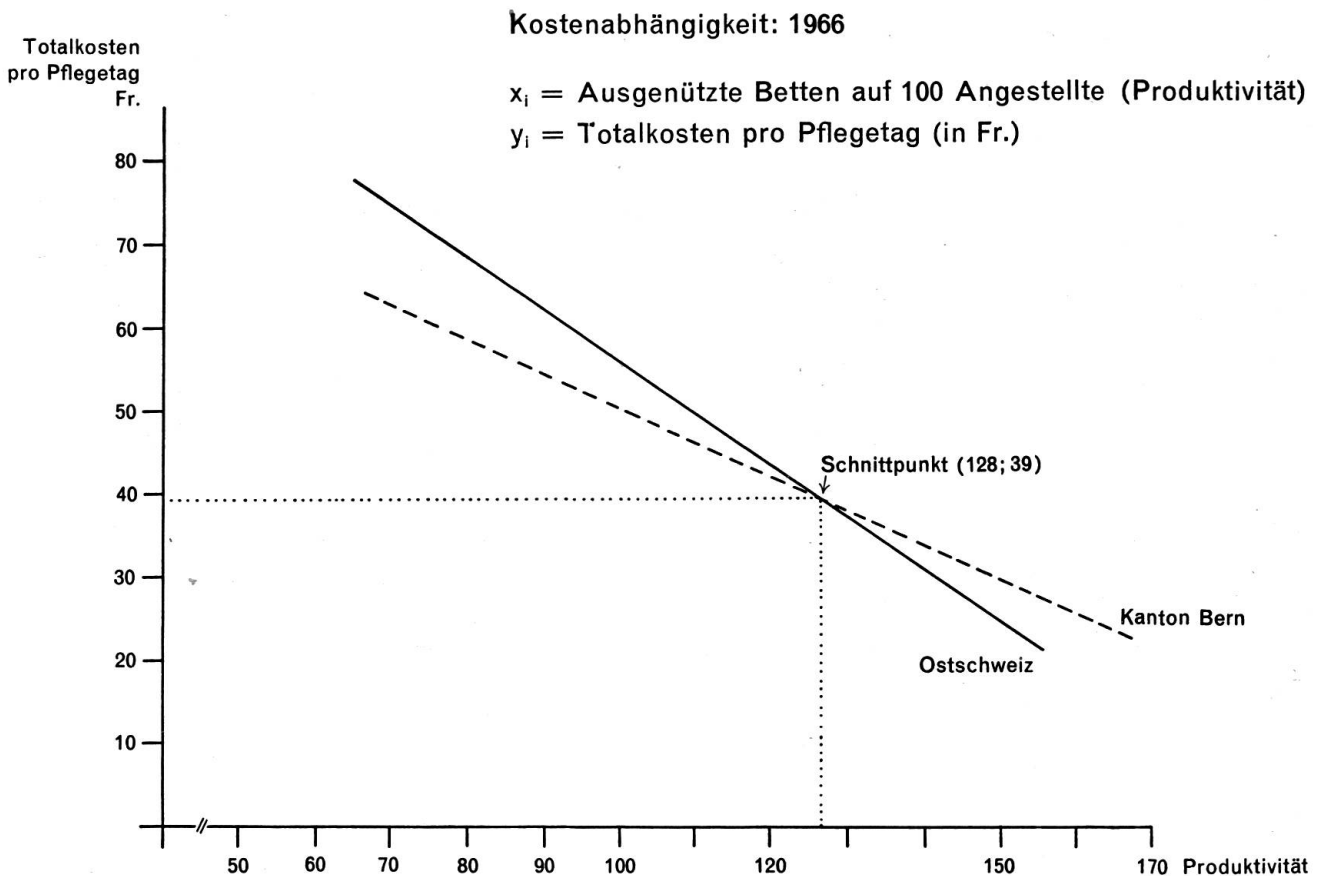
Die Berechnungen der Regressionsgleichung «Ostschweiz» basieren zur Hauptsache auf den Ergebnissen einiger Bezirksspitäler aus der Nord- und Ostschweiz.

Aus den beiden Regressionsgleichungen lässt sich – provisorisch – ein Schnittpunkt bestimmen, da die beiden Regressionskoeffizienten auf den ersten Blick verschieden voneinander sind (diese Hypothese wird später zu prüfen sein).

Schnittpunkt:  $x = 127,82 \sim 128$

$y = 39,13 \sim 39$

Die folgende Grafik gibt die beiden Regressionsgleichungen mit ihrem Schnittpunkt:



Aus den Regressionsgleichungen können die Regressionswerte  $Y_i$  bestimmt werden:

Produktivität $x_i$	Kanton Bern		Ostschweiz	
	$x_i$	$Y_i$	$x_i$	$Y_i$
Kleinster Wert	66	64.30	65	77.38
100	100	50.46	100	56.07
Grösster Wert	168	22.77	157	21.36

Sofern die Hypothese verschiedener Regressionskoeffizienten angenommen werden muss, lässt sich aus der Grafik entnehmen, dass im Kanton Bern bei den Spitälern mit tiefer Produktivität die Totalkosten kleiner sind als in der Ostschweiz; umgekehrt liegen die bernischen Kosten pro Pflegetag für Spitäler mit hoher Produktivität über denjenigen ostschweizerischer Spitäler. Für bernische Betriebe hat eine Produktivitätsverschlechterung einen geringeren Anstieg der Relativkosten zur Folge als bei den Vergleichsbetrieben der Ostschweiz.

Wie in Abschnitt 125 dargelegt, sind nun verschiedene Hypothesen zu prüfen.

(1) Vorerst ist abzuklären, ob für beide Regressionsgeraden **Linearität** angenommen werden kann.

Wir verwenden die Prüfansätze (29) und (30). Die Resultate zeigen, dass sowohl für den «Kanton Bern» wie auch bei der «Ostschweiz» **lineare Regression zulässig** ist.

(2) Weiterhin ist zu prüfen, ob die Regressionskoeffizienten von Null verschieden sind.

Regressionskoeffizient Kanton Bern:  $b_{BE}$

Regressionskoeffizient Ostschweiz:  $b_{OS}$

Es ist also zu prüfen, ob

$H_0: b_{BE} = 0$ , bzw.

$H_0: b_{OS} = 0$

Verwendung findet der Testansatz (26). Die Resultate lassen erkennen, dass die Nullhypothese verworfen werden muss; **beide Regressionskoeffizienten sind wesentlich von Null verschieden**.

(3) Für beide Spitalgruppen wurde lineare Regression als zulässig erkannt. Ein Prüfverfahren hat weiter gezeigt, dass die beiden Regressionskoeffizienten verschieden von Null sind. Es ist somit gegeben, die Parallelität der Geraden zu prüfen. Als Prüfansatz kommt (36) zur Anwendung.

Die Berechnungen (zwei Regressionsgerade) ergeben:

Streuung	SQ	FG	DQ
Um 1. Regressionsgerade (BE)	980,750	29	.
Um 2. Regressionsgerade (OS)	487,978	14	.
Beide Regressionsgeraden	1 468,728	43	$s_1^2 = 34,156$
Nichtparallelität	192,222	1	$s_2^2 = 192,222$
Innerhalb Gruppen	1 660,950	44	$s_3^2 = 37,749$
Zwischen Gruppen (Abstand)	355,864	1	$s_4^2 = 355,864$
Insgesamt	2 016,814	45	.

Prüfverfahren: Nichtparallelität

$$F = \frac{s_2^2}{s_1^2} = \frac{192,222}{34,156} = 5,628$$

$$F_{0,05} \sim 4,1; \quad F > F_{0,05; 1,43}$$

**Folgerung:** Die Regressionskoeffizienten sind wesentlich voneinander verschieden. Die beiden Regressionsgleichungen für den Kanton Bern und die Ostschweiz können daher nicht zu einer einzigen Gleichung mit einem gemeinsamen Regressionskoeffizienten vereinigt werden. Für beide Gebiete liegen unterschiedliche Verhältnisse vor, d. h. Produktivitätsveränderungen haben verschiedene Auswirkungen auf die Relativkosten, je nachdem es sich um bernische oder ostschweizerische (vor allem zürcherische) Spitalbetriebe handelt.

Der Prüfansatz (36) könnte vor allem dann Verwendung finden, wenn Produktivitätsuntersuchungen für mehrere Gebiete vorzunehmen sind. Der generelle Test erlaubt die Prüfung der Gesamtheit der Regressionskoeffizienten. Vermittelt eines t-Testes kann anschliessend geprüft werden – sofern der F-Test anspricht –, zwischen welchen Regressionskoeffizienten signifikante Unterschiede bestehen:

$$t^2 = \frac{(b_j - b_k)^2}{s_1^2} \cdot \frac{S_{x_j x_j} S_{x_k x_k}}{S_{x_j x_j} + S_{x_k x_k}} \quad (74)$$

mit:  $n^* = (N - 2p)$  Freiheitsgraden.

Liegt Parallelität vor, d. h. sind die Regressionskoeffizienten nur zufällig voneinander verschieden, so kann es sinnvoll sein, den **Abstand** zwischen den beiden Regressionsgeraden zu bestimmen. In unserem Fall der Nichtparallelität wird selbstverständlich ein ganz bestimmter Abstand interessieren. So könnte man sich z. B. fragen, wie gross der Unterschied der Relativkosten bei einer Produktivität von 80 ist.

Wir schliessen diesen Abschnitt über den interkantonalen Vergleich ab. Die Zahlen geben einige interessante Aufschlüsse. Sie lassen erkennen, dass offensichtlich regionale Kostenunterschiede bestehen, was Hinweise auf bestimmte Kostensenkungsmöglichkeiten gibt. Es könnte nicht unzweckmässig sein, diesen regionalen Vergleich auszubauen und die wichtigsten Zahlen einer eingehenden Analyse zu unterziehen.

## 243 Mehrfache Regression und Korrelation

### 243.1 Die Variablen

In Abschnitt 241 haben wir jeweils mit einfacher linearer – oder nichtlinearer – Regression untersucht, welchen Einfluss die analysierte unabhängige Variable auf die Betriebskosten pro Pflgetag der bernischen Bezirksspitäler ausübt. Es ist nun gegeben, die verschiedenen Variablen gleichzeitig zu untersuchen, d. h. ein Modell von der Art des Ansatzes (45) bzw. (46) oben zu verwenden.

Als Variable gehen in diesen Ansatz ein:

#### (1) Abhängige Variable (y):

$y = T = K_B/E =$  Betriebskosten pro Pflgetag (in Fr.) 1966.

#### (2) Einflussfaktoren (x):

$x_1 =$  Arbeitsproduktivität (ausgenützte Betten pro Angestellten;  $P_A$ );

$x_2 =$  Ausnützungsgrad in % ( $G_B$ );

$x_3 =$  Betriebsgrösse (B);

$x_4 =$  Durchschnittliche Aufenthaltsdauer der Patienten (D);

$x_5 =$  Durchschnittliche Personalkosten (L);

$x_6 =$  Nahrungsmittel-«Preise» (N);

$x_7 =$  «Preise» für medizinisches Material (M);

$x_8 =$  Spezialabteilungen (S): Unechte Variable.

Für Einzelheiten verweisen wir auf Abschnitt 241 bzw. die in 223 gegebene Theorie.

Es ist somit

$$T = f(P_A, G_B, B, D, L, N, M, S)$$

oder, bei Verwendung eines linearen Ansatzes,

$$T = a + b_1 P_A + b_2 G_B + b_3 B + b_4 D + b_5 L + b_6 N + b_7 M + b_8 S$$

Unsere Aufgabe ist es, die Regressionsparameter aufzusuchen, Bestimmtheitsmasse zu errechnen, Hypothesen zu prüfen und Vertrauensgrenzen zu bestimmen. Die Berechnungen wurden auf dem Elektronenrechner der Universität Bern durchgeführt.

Zu erwähnen ist bereits hier, dass die einzelnen Einflussfaktoren in der Reihenfolge ihrer Bedeutung eingeschlossen werden, wobei sich in unserem Fall nachstehende Rangfolge der Wichtigkeit ergibt:

- $x_1$  = Produktivität;
- $x_5$  = Personalkosten pro Angestellten;
- $x_7$  = Medizinisches Material («Preise»);
- $x_6$  = Nahrungsmittelpreise;
- $x_4$  = Durchschnittliche Aufenthaltsdauer;
- $x_3$  = Betriebsgrösse;
- $x_8$  = Spezialabteilungen;
- $x_2$  = Ausnützungsgrad.

Bereits dieser Aufstellung ist zu entnehmen, dass Produktivität und durchschnittliche Personalkosten an oberster Stelle rangieren, beide Einflussfaktoren also von zentraler Bedeutung sind.

Es wird später darzulegen sein, in welchem Ausmass die einzelnen Variablen zur Erklärung der Variabilität beitragen.

### 243.2 Schätzwerte

Wir nehmen linearen Zusammenhang an und geben vorerst die Resultate der Berechnungen in bezug auf die Regressionsparameter, wobei die Variablen wiederum in der Reihenfolge ihrer Bedeutung aufgeführt sind:

Einflussfaktor	Niveau-konstante a	Partielle Regressionskoeffizienten bei Einschluss von:								
		$b_j$	$x_1$	$x_5$	$x_7$	$x_6$	$x_4$	$x_3$	$x_8$	$x_2$
Produktivität	85,0	$b_1$	-0,383	-0,361	-0,281	-0,258	-0,289	-0,270	-0,269	-0,264
Personalkosten	53,3	$b_5$		0,003	0,002	0,002	0,002	0,002	0,002	0,002
Medizin. Material	44,6	$b_7$			0,079	0,062	0,070	0,067	0,069	0,071
Nahrungsmittel	34,4	$b_6$				0,083	0,086	0,092	0,088	0,087
Aufenthaltsdauer	28,8	$b_4$					0,291	0,270	0,245	0,238
Betriebsgrösse	26,8	$b_3$						0,011	0,018	0,018
Spezialabteilung	27,4	$b_8$							-1,903	-1,835
Ausnützungsgrad	27,9	$b_2$								-0,014

Die Aufstellung enthält die sogenannten partiellen Regressionskoeffizienten, welche den Einfluss jeder Variablen in ursprünglichen Einheiten wiedergeben. So nehmen z. B. die Betriebskosten pro Pflage tag um 3,61 Franken ab, wenn – unter Konstanz der Löhne – die Produktivität um 10 Einheiten zunimmt (oder umgekehrt).

Aus der Tabelle sind die Veränderungen der Niveaunkonstante bzw. der Regressionskoeffizienten bei Einbezug eines zusätzlichen Einflussfaktors sehr gut zu erkennen. Die Regressionsgleichungen können dieser Aufstellung entnommen werden. Der alleinige Einbezug der unabhängigen Variablen «Produktivität» liefert die bereits früher erörterte Regressionsgleichung

$$Y = 85,0 - 0,383 x$$

Das totale Bestimmtheitsmass gibt eine Information über die Stärke der stochastischen Verbundenheit und wird nach Ansatz (54) berechnet. Wir werden nun die Auswirkungen eines Einbezugs einer zusätzlichen Variablen auf die totale Bestimmtheit betrachten.



## Die Resultate:

Einflussfaktor	B <sub>T</sub>	F-Wert (aus F-Test); Einbezug von:							
		X <sub>1</sub>	X <sub>5</sub>	X <sub>7</sub>	X <sub>6</sub>	X <sub>4</sub>	X <sub>3</sub>	X <sub>8</sub>	X <sub>2</sub>
Produktivität	0,76	91,1	195,1	132,3	131,0	169,8	178,8	191,1	134,8
Personalkosten	0,90		43,0	35,0	49,9	74,3	85,4	88,4	81,9
Medizin. Material	0,95			26,3	18,4	29,2	36,1	40,3	33,9
Nahrungsmittel	0,96				9,3	13,2	19,9	19,4	18,4
Aufenthaltsdauer	0,97					9,1	10,3	8,9	7,8
Betriebsgrösse	0,98						9,4	11,3	11,1
Spezialabteilung	0,98							2,8	2,5
Ausnützungsgrad	0,98								0,2

Aus dieser Tabelle ist ersichtlich, dass schon die unabhängige Variable Nr. 7 (medizinisches Material) nicht mehr sehr viel zur Erhöhung des totalen Bestimmtheitsmasses beiträgt; massgebliche Erklärungsfaktoren sind also Produktivität und durchschnittliche Personalkosten (Löhne). Beziehen wir diese beiden Grössen in die mehrfache Regression ein, so erreicht das totale Bestimmtheitsmass einen Wert von  $B = 0,90$ . Ohne weitere Untersuchungen ist man geneigt, die drei unabhängigen Variablen Nr. 1, 5 und Nr. 7 in die mehrfache Regression und Korrelation einzuschliessen.

Es taucht nun die Frage auf, ob **Abhängigkeiten zwischen den unabhängigen Variablen** bestehen. Aufschlussreich ist in dieser Beziehung eine Tabelle der einfachen Korrelationskoeffizienten:

Einflussfaktor		X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	y
X <sub>1</sub>	Produktivität	0,18	-0,41	0,66	-0,13	-0,60	0,63	-0,38	-0,87
X <sub>2</sub>	Ausnützungsgrad		0,13	-0,11	0,13	-0,01	0,26	0,20	-0,02
X <sub>3</sub>	Betriebsgrösse			-0,28	0,24	0,20	0,33	0,79	0,50
X <sub>4</sub>	Aufenthaltsdauer				-0,39	-0,48	-0,63	-0,34	-0,67
X <sub>5</sub>	Personalkosten					0,13	0,42	0,20	0,49
X <sub>6</sub>	Nahrungsmittel						0,60	0,16	0,69
X <sub>7</sub>	Medizin. Material							0,34	0,83
X <sub>8</sub>	Spezialabteilung								0,42

Die Aufstellung zeigt deutlich, dass zum Teil enge Korrelation zwischen einzelnen unabhängigen Variablen besteht, so z. B. zwischen Produktivität und Aufenthaltsdauer, Betriebsgrösse und Spezialabteilungen, Aufenthaltsdauer und Nahrungsmitteln, Aufenthaltsdauer und medizinischem Material.

Auf Grund dieser Tatsache sowie der Tabelle der totalen Bestimmtheiten ist es gegeben, nur die beiden unabhängigen Variablen **Produktivität** und **durchschnittliche Personalkosten** in die Überlegungen einzubeziehen. Dies aus den beiden dargelegten Gründen:

1. Einen wesentlichen Beitrag zur Erklärung der Variabilität liefern nur Produktivität und Löhne;
2. Zwischen Produktivität einerseits und Betriebsgrösse bzw. Aufenthaltsdauer, Nahrungsmittel, medizinischem Material andererseits bestehen ziemlich enge Korrelationen.

### Schlussfolgerung:

Gestützt auf das zur Verfügung stehende Material sowie die soeben erläuterten Auswertungen wird man folgende Variablen für die mehrfache Regression berücksichtigen:

$y = T =$  Betriebskosten pro Pflgetag (in Fr.) 1966

$x_1 = P_A =$  Arbeitsproduktivität

$x_5 = L =$  Durchschnittliche Personalkosten.



Die Ergebnisse aus mehrfacher linearer Regression und Korrelation:

Einflussfaktor	Niveau-konstante a	Regressionskoeffizient bei Einschluss von			Bestimmtheit B <sub>T</sub>	F-Test		
		b <sub>j</sub>	x <sub>1</sub>	x <sub>5</sub>		x <sub>1</sub>	x <sub>5</sub>	Wertung
x <sub>1</sub> : Produktivität	85,0	b <sub>1</sub>	-0,383	-0,3609	0,76	91,1	195,1	(***)
x <sub>5</sub> : Personalkosten	53,3	b <sub>5</sub>		0,0029	0,90		43,5	(***)

(\*\*\*) bedeutet: Gesichert für P = 1<sup>o</sup>/<sub>100</sub>.

Wir haben somit sechs von acht Variablen weggelassen und nur deren zwei als gesicherte Einflussfaktoren einbezogen. Man könnte der Auffassung sein, dass dies ein recht dürftiges Ergebnis sei. Dem ist jedoch nicht so; Spitäler sind Dienstleistungsbetriebe reinsten Wassers. Die Relation Pflage : Personal sowie die Lohnhöhe sind von ausschlaggebender Bedeutung für die Höhe der Betriebskosten pro Pflage. Unsere Regressionsgleichung

$$Y = 53,3 - 0,361 x_1 + 0,0029 x_5$$

muss unseres Erachtens als sinnvolles Resultat der Regressionsanalyse gewertet werden. **Mit steigender Produktivität sinken ceteris paribus die Kosten, mit wachsenden Löhnen nehmen sie zu.** Die partiellen Regressionskoeffizienten geben Auskunft über Richtung und Ausmass der Kostenveränderungen bei Variation der Einflussfaktoren Produktivität und durchschnittlicher Lohnhöhe bzw. durchschnittlichen Personalkosten.

243.3 Prüfen von Hypothesen

(1) Streuungszzerlegung

Es soll vorerst abgeklärt werden, ob es statthaft ist, mehrfache lineare Regression anzusetzen. Wir verwenden zur Abklärung dieser Frage folgende Streuungszzerlegung:

Streuung	SQ	FG	DQ
Auf Regression	$S(Y_i - \bar{y})^2 = A$	2	$s_1^2 = A/2$
Um Regression	$S(y_i - Y_i)^2 = B$	N - 3	$s_2^2 = B/(N - 3)$
Insgesamt	$S(y_i - \bar{y})^2 = C$	N - 1	.

Als Prüfverfahren ergibt sich ein F-Test:

$$F = \frac{DQ \text{ (auf Regression)}}{DQ \text{ (um Regression)}} = \frac{s_1^2}{s_2^2}$$

mit: n<sub>1</sub><sup>\*</sup> = 2 und n<sub>2</sub><sup>\*</sup> = (N-3) FG

**Entscheid:** Bei F ≥ F<sub>p</sub> sind die Streuungen wesentlich voneinander verschieden, d. h. mehrfache lineare Regression ist zulässig (damit ist auch B > 0).

In unserem Fall lautet die Problemstellung: Ist zweifache lineare Regression zulässig, mit

y = Betriebskosten pro Pflage;

x<sub>1</sub> = Produktivität;

x<sub>5</sub> = Durchschnittliche Personalkosten.

In die Varianzanalyse sind also zwei unabhängige und eine abhängige Variable einzubeziehen.

Das Ergebnis der Berechnungen:

Streuung	SQ	FG	DQ
Auf Regression ( $x_1$ und $x_5$ )	3 135,961	2	$s_1^2 = 1 567,981$
Um Regression	330,058	28	$s_2^2 = 11,788$
Insgesamt	3 466,019	30	.

**F-Test:**

$$F = \frac{s_1^2}{s_2^2} = \frac{1 567,981}{11,788} = 133,015 (***)$$

$$\left. \begin{array}{l} n_1^* = 2 \\ n_2^* = 28 \end{array} \right\} F_{0,001; 2,28} = 13,498$$

Somit:  $F \gg F_{0,001}$ .

**Folgerung:** Zweifache lineare Regression ist für eine Irrtumswahrscheinlichkeit von  $P = 1/\infty$  gesichert und zulässig.

Wir prüfen nun weiter die Hypothese, ob die zusätzlich eingeschlossene (allgemein:  $p-k$ ) Variable «Löhne» einen wesentlichen Beitrag zum Summenquadrat beider (allgemein:  $p$ ) Variablen leistet (vgl. dazu: K.W. Smillie, Introduction, S.50ff.).

Die Streuungszerlegung:

Streuung	SQ	FG	DQ
Bezüglich $x_1$	2 628,792	1	2 628,792
Zusätzlich bezüglich $x_5$	507,169	1	507,169
Auf Regression (bezüglich $x_1$ und $x_5$ )	3 135,961	2	.
Um Regression	330,058	28	$s^2 = 11,788$
Insgesamt	3 466,019	30	.

Der F-Test ergibt

$$F = \frac{507,169}{11,788} = 43,077 > F_{0,001} (= 13,498)$$

**Folgerung:** Aus der Streuungszerlegung geht hervor, dass die zusätzlich eingeführte Variable  $x_5$  (= durchschnittliche Personalkosten) einen wesentlichen Beitrag zum Summenquadrat liefert; an Stelle der Regressionsgleichung

$$Y = 85,0 - 0,383 x$$

wird man folgende Gleichung verwenden:

$$Y = 53,3 - 0,361 x_1 + 0,0029 x_5$$

Es ist nun noch der partielle Regressionskoeffizient  $b_1$  zu prüfen:

Streuung	SQ	FG	DQ
Bezüglich $x_5$	835,705	1	835,705
Zusätzlich bezüglich $x_1$	2 300,256	1	2 300,256
Auf Regression (bezüglich $x_1$ und $x_5$ )	3 135,961	2	.
Um Regression	330,058	28	$s^2 = 11,788$
Insgesamt	3 466,019	30	.

Der F-Test liefert

$$F = \frac{2\,300,256}{11,788} = 195,139 (***)$$

Der Einbezug der Produktivität erhöht das Summenquadrat für die Regressionswerte wesentlich.

## (2) Partielle Regressionskoeffizienten

Für den Fall der Nullhypothese prüfen wir, ob  $b_j$  zufällig oder wesentlich von Null verschieden ist. Als Prüfansatz gilt

$$t = \frac{b_j}{s\sqrt{c_{jj}}} = \frac{b_j}{s_{b_j}}$$

mit:  $n^* = N - p - 1$  Freiheitsgraden;

$s^2 = DQ$  (um Regression) in (56);

$c_{jj}$  = Multiplikator.

Die Auswertung ergibt:

Partielle Regressionskoeffizienten	b	t-Wert	$t_{0,001}$	Wertung
$b_1$	-0,3609	13,969	3,674	(***)
$b_5$	0,002876	6,559	3,674	(***)

**Folgerung:** Beide Regressionskoeffizienten  $b_1$  und  $b_5$  (Produktivität und Löhne) sind wesentlich von Null verschieden.

## (3) Mehrfache (totale) Bestimmtheit

Als Prüfansatz verwenden wir (58) oben; er hat die Form

$$F = \frac{B(N - p - 1)}{(1 - B)p} = \frac{25,333\,644}{0,190\,454}$$

Das Resultat der Division ergibt

$$F = 133,017 \gg F_{0,001} (= 13,498)$$

Das totale Bestimmtheitsmass ist somit wesentlich von Null verschieden; wir gelangen hier zu den gleichen Schlussfolgerungen wie bei der Varianzanalyse weiter oben.

## 243.4 Vertrauensgrenzen

### (1) Regressionskoeffizienten

Die Vertrauensgrenzen für die partiellen Regressionskoeffizienten sind gegeben durch

$$\left. \begin{aligned} b_j \pm t_p s / \sqrt{c_{jj}}, & \quad \text{bzw.} \\ b_j \pm t_p s_{b_j} \end{aligned} \right\} \quad (75)$$

mit:  $n^* = N - p - 1$  Freiheitsgraden.

Für unsere Kostenabhängigkeit ergibt sich:

Variable	s <sub>b<sub>j</sub></sub>	t <sub>0,01</sub> s <sub>b<sub>j</sub></sub>
Produktivität	0,025 832 3	0,071 375
Löhne	0,000 438 5	0,001 212

mit:  $t_{0,01} = 2,763$

Daraus erhalten wir bei einer vorgegebenen Wahrscheinlichkeit von 1% folgende Werte:

Variable	t <sub>0,01</sub> s <sub>b<sub>j</sub></sub>	b <sub>j</sub>	u <sub>b<sub>j</sub></sub>	o <sub>b<sub>j</sub></sub>
(0)	(1)	(2)	(3)	(4)
Produktivität	0,071 375	-0,360 857	-0,290	-0,432
Löhne	0,001 212	0,002 876	0,001 66	0,004 09

Die Spalte (2) enthält die partiellen Regressionskoeffizienten, Spalten (3) und (4) die Werte  $b_j \pm t_p s_{b_j}$ , d. h. im Intervall von -0,290 und -0,432 ist z. B. der Regressionskoeffizient  $b_1$  enthalten.

### (2) Regressionswerte Y

Wie bereits in Abschnitt 145 dargelegt, sind die aus der Regressionsgleichung errechneten Schätzwerte Y zufälligen Schwankungen unterworfen; man wird deshalb eine Streuung  $s_y^2$  bestimmen. Dies geschieht unter Verwendung des Ansatzes (60) oben. Nach Smillie (vgl. Introduction, S. 49) lassen sich die Vertrauensgrenzen gemäss folgendem, in der Symbolik etwas geänderten Ausdruck berechnen:

$$Y \pm t_p s_2 \sqrt{\left( \sum_{j, k=0}^p c_{jk} X_j X_k \right)}$$

Für einen Einzelwert ist in der Klammer der Wert 1 hinzuzufügen. Unsere Hauptaufgabe war es, die Kosten-Einflussfaktoren aufzufinden; wir verzichten daher auf die Berechnung der Vertrauensgrenzen der Regressionswerte.

## 243.5 Überblick über die Resultate aus mehrfacher Regression

Wir haben untersucht, ob die in Abschnitt 243.1 aufgeführten unabhängigen Variablen als Einflussfaktoren der Betriebskosten zu betrachten seien. Die jeweilige Veränderung des totalen Bestimmtheitsmasses bei Einbezug einer zusätzlichen Variablen sowie bestehende Korrelation zwischen verschiedenen unabhängigen Variablen lassen es als richtig erscheinen, nur die beiden Faktoren «Produktivität» und «durchschnittliche Personalkosten» in die mehrfache Regression einzubeziehen. Es gilt also der Ansatz

$$T = f(P_A, L)$$

bzw., bei Verwendung eines linearen Ansatzes

$$T = a + b_1 P_A + b_5 L$$

Aus den Berechnungen erhalten wir folgende Regressionsgleichung:

$$T = 53,3 - 0,361 P_A + 0,0029 L.$$

Diese Gleichung besagt, dass mit steigender Produktivität die Betriebskosten pro Pflage tag sinken, bei Lohnerhöhungen jedoch auch die Relativkosten ansteigen. Es ist dies unseres Erachtens ein vernünftiges Resultat. Je weniger Personal (in Relation zu den Pflage tagen) man einsetzen wird, desto tiefer sind die Kosten pro Pflage tag; je höher die Löhne, um so höher auch die Relativkosten. Der ziemlich hohe Wert für die totale Bestimmtheit ( $B = 0,90$ ) lässt die Folgerung zu, dass die Betriebskosten pro Pflage tag durch die beiden erwähnten Einflussgrössen gut bestimmt sind. Die Prüfverfahren, vor allem die Varianzanalyse, haben gezeigt, dass mehrfache Regression zulässig ist bzw. beide Faktoren einen wesentlichen Beitrag zum Summenquadrat liefern. Der t-Test bestätigt die Vermutung, dass die beiden partiellen Regressionskoeffizienten wesentlich von Null verschieden sind.

## 25 Die Hauptergebnisse der Untersuchung

(1) Die Spitalkosten sind in den letzten 15 bis 20 Jahren ausserordentlich stark angestiegen, mit entsprechenden finanziellen Wirkungen auf Spitalbenützer, Krankenkassen und öffentliche Hand bzw. Steuerzahler. Es ist daher gegeben, die **Frage nach den Kosten-Einflussfaktoren** zu stellen.

Das **Hauptproblem** kann folgendermassen formuliert werden: Von welchen Faktoren werden die Kosten, insbesondere die Betriebskosten pro Pflage tag, beeinflusst? Es war daher das Ziel der vorliegenden Untersuchung, die Einflüsse auf die Spitalleistungskosten zu erfassen und die Stärke dieser Einflüsse zu berechnen.

(2) Die betriebswirtschaftliche **Theorie der Kostenabhängigkeit** erklärt grundsätzlich die Kosten als von folgenden Faktoren abhängig:

- Produktivität;
- Marktpreise (bzw. Verbrauchsmengen);
- Ausnützungsgrad (auch Beschäftigungsgrad genannt);
- Betriebsgrösse.

Für den Spitalbetrieb kommen ferner in Betracht:

- Aufenthaltsdauer der Patienten;
- Art und Anzahl von Spezialabteilungen;
- Behandlungsmethode.

Diese Aufzählung ist nicht vollständig; sie schliesst aber unseres Erachtens die wichtigsten Faktoren ein.

Als bedeutendster Einflussfaktor erscheint an erster Stelle die **Produktivität**. Sie ist im Spitalbetrieb nicht leicht zu bestimmen. Die von uns verwendete statistische Beziehungsgrösse ist eine faktorbezogene Produktivität und lässt keinen Schluss auf einen Kausalzusammenhang zu; insbesondere kann eine «Verschlechterung» dieser Masszahl u. a. auf den medizinischen Fortschritt zurückgeführt werden, der eine Vermehrung des Pflege- und Hilfspersonals zur Folge hat. Die in unserer Untersuchung benützte Arbeitsproduktivität berechnet sich gemäss Ansatz (70) bzw. (70a), wobei diese Formeln auf der generellen Beziehung beruhen

$$P = \frac{\text{Output}}{\text{Input}} = \frac{\text{Produktionsergebnis}}{\text{Arbeitseinsatz}}$$

Es zeigen sich sofort die Schwächen dieser so bestimmten Arbeitsproduktivität. Das Produktionsergebnis, der Pflage tag, ist eine ziemlich komplexe Grösse. Vergleiche zwischen mehreren Spitälern liefern nicht unbedingt eindeutige Schlüsse. Ein Weg zur Verbesserung der Messung des Outputs wäre möglicherweise der, die Pflage tage mit den Ergebnissen der Krankenstatistik zu gewichten, wobei die Personalbeanspruchung entsprechend zu berücksichtigen wäre. Auch beim Input, dem Arbeitseinsatz, sind Vorbehalte anzubringen. Die Verwendung des Personalbestandes im Nenner des Bruches hat gewisse Nachteile. Besser wäre es, Personal-Arbeitsstunden zu verwenden; dadurch erfahren Spitäler mit halbtagsweise beschäftigtem Personal eine gerechtere Beurteilung. Weiter wäre beim Input

das Problem der Pflegeschulen zu lösen. Auch hier könnte eine zweckentsprechend geführte Arbeitsstunden-Statistik zu besseren Ergebnissen führen. Allgemein kann gesagt werden, dass der Kosten-Einflussfaktor «Arbeitsproduktivität» in unserer Analyse eine eher approximative Aussagefähigkeit besitzt. Es wäre vielleicht besser, direkt von einer Pflegetage-Personal-Relation zu sprechen.

In bezug auf die **Marktpreise** bzw. **Verbrauchsmengen** wären bessere statistische Unterlagen erwünscht; die zur Verfügung stehenden Unterlagen sind ziemlich mangelhaft.

Die restlichen oben aufgeführten Einflussfaktoren sind u.E. statistisch richtig erfasst, und ihrer Einführung in eine mehrfache Regression würde somit nichts im Wege stehen. Die quantitative Erfassung der Behandlungsmethode ist allerdings nur unter grossen Schwierigkeiten möglich; da jegliche Unterlagen fehlten, konnte diese Variable nicht berücksichtigt werden. Hingegen haben die Faktoren Ausnutzungsgrad, Betriebsgrösse, Aufenthaltsdauer und Spezialabteilung Eingang in unser Modell der Kostenabhängigkeit gefunden.

(3) In bezug auf die **Methode** ist zu erwähnen, dass wir zwei Verfahren verwendeten:

- Statistik: Vermittelt der Regressions- und Korrelationsrechnung musste abgeklärt werden, welche der Einflussfaktoren (von der betriebswirtschaftlichen Theorie vorgegeben) effektiv geeignet sind, die Höhe der Betriebskosten pro Pflegetag zu erklären;
- Betriebswirtschaft: Die benötigten Parameter können aus Betriebsvergleich oder Zeitreihen hervorgehen. In unserer Untersuchung haben wir die Daten von 31 Bezirksspitalern des Kantons Bern für das Jahr 1966 verwendet; die Regressionsgleichungen sind somit das Resultat eines Betriebsvergleichs.

In einem späteren Zeitpunkt soll eine vergleichende Analyse von Zeitreihen vorgenommen werden, sobald die Unterlagen vollständig sind.

(4) Als **Erhebungsgrundlage** für die Berechnungen verwenden wir die Krankenanstaltenstatistik und die Jahresrechnung der Bezirksspitäler für das Jahr 1966. Die Statistikbogen liefern wertvolle Informationen; trotzdem wäre ein Ausbau des Basismaterials bzw. der Führungsunterlagen der Spitäler erwünscht. So könnte ein Betriebsabrechnungsbogen und eine Leistungsergebnisrechnung nützliche Dienste leisten. Weiter wäre es der Sache förderlich, wenn für bestimmte Erhebungsmerkmale Monatszahlen vorlägen, so u. a. für Patientenzahlen, Ausnutzungsgrad usw.

(5) Vermittelt **einfacher linearer und nichtlinearer Regression** versuchen wir vorerst abzuklären, ob für die einzelnen Faktoren die Hypothese des Kosteneinflusses angenommen oder verworfen werden muss. In bezug auf die Produktivität zeigten die Berechnungen, dass die Hypothese einer linearen Abhängigkeit zwischen dieser Variablen und den Betriebskosten pro Pflegetag angenommen werden muss; mit sinkender Produktivität steigen die Relativkosten. Rund 76% der Streuung der abhängigen Veränderlichen (Betriebskosten) lässt sich aus der Variabilität des Einflussfaktors Arbeitsproduktivität erklären. Die gleichen Schlussfolgerungen müssen gezogen werden in bezug auf die durchschnittlichen Personalkosten ( $B = 0,24$ ), die Nahrungsmittel ( $B = 0,44$ ) und das medizinische Material ( $B = 0,69$ ). Die einfache lineare Regression ergibt weiter eine gesicherte Abhängigkeit mit negativem Vorzeichen für Aufenthaltsdauer der Patienten und Relativkosten. Für Betriebsgrösse und Kosten ist lineare Abhängigkeit erlaubt; ein nichtlinearer Ansatz gibt bessere, jedoch nicht sehr überzeugende Resultate. Schliesslich ist auch zwischen Löhnen und Betriebskosten einfache lineare Regression zulässig; Pflegetag-Kostensteigerung als Folge der Lohnerhöhung ist ja eine für die Betriebsleiter nicht unbekannt betriebswirtschaftliche Erscheinung. Kein positives Ergebnis zeitigte die Verbindung von Ausnutzungsgrad und Relativkosten. Dies verwundert uns nicht, weisen doch fast alle Spitäler einen hohen bis sehr hohen Ausnutzungsgrad auf.

(6) Ein **interkantonaler Vergleich** zwischen bernischen und ostschweizerischen Betrieben ergibt unterschiedliche Resultate in bezug auf die Kostenabhängigkeit bei Benützung der Arbeitsproduktivität als Einflussgrösse. Es können somit regionale Abweichungen vermutet werden, die in Zukunft näher analysiert werden sollten. Die Möglichkeit von realisierbaren Kosteneinflussmassnahmen könnte besser abgeklärt werden.

(7) Auf Grund der betriebswirtschaftlichen Theorie und der Ergebnisse der einfachen linearen und nichtlinearen Regression und Korrelation wurden abhängige (Betriebskosten pro Pflegetag) und unabhängige Variable in einer **mehrfachen Regression und Korrelation** vereinigt. Verschiedene



Einflussfaktoren konnten in der endgültigen Regressionsgleichung keinen Platz finden, da zwischen mehreren unabhängigen Variablen hohe Korrelation bestand, bzw. einige Grössen nur unwesentlich zur Erhöhung des totalen Bestimmtheitsmasses beitrugen. Es bleibt schliesslich folgende Regressionsgleichung:

$$T = 53,3 - 0,361 P_A + 0,0029 L,$$

wobei:

T = Betriebskosten pro Pflage tag (in Fr.) 1966;

P<sub>A</sub> = Arbeitsproduktivität;

L = Durchschnittliche Personalkosten (in Fr.) 1966.

Für sinkende Produktivität und steigende Löhne müssen wachsende Betriebskosten pro Pflage tag budgetiert werden. Dieses Ergebnis ist insofern bescheiden, als die Mehrzahl der betrachteten Einflussfaktoren aus mehrfacher Regression nicht berücksichtigt werden konnte. Es scheint jedoch, dass Ausnützungsgrad und Betriebsgrösse, Preise und Verbrauchsmengen Chancen haben, bei einer Zeitreihenanalyse wiederum eine ins Gewicht fallende Rolle zu spielen.

(8) Es könnte sinnvoll sein, die vorliegende Kostenuntersuchung aus einem Betriebsvergleich mit einer Zeitreihenanalyse zu ergänzen. Weiter wäre es erwünscht, auf Grund verbesserter und umfangreicherer Unterlagen die Kostenabhängigkeit in Krankenanstalten erneut zu beleuchten.

Spitäler üben eine nicht zu unterschätzende Funktion aus. Man wird daher, wenn es um die Gesundheit und das Leben von Menschen geht, nicht allzu spitz rechnen dürfen. Eine gewisse Kostenkontrolle und aufmerksames Kostenbewusstsein sind aber auch im Spitalsektor notwendig.

Bern, im März 1968

## 25 Les principaux résultats de l'étude

(1) La hausse considérable des frais d'hôpital, au cours des 15 à 20 dernières années, a de lourdes conséquences financières pour les patients, les caisses-maladie, les pouvoirs publics et, partant, pour le contribuable. Il convient donc d'étudier les **facteurs d'influence** entrant en ligne de compte.

Le problème essentiel peut se formuler ainsi: Quels sont les facteurs qui influent sur les frais, en particulier sur les frais d'exploitation par journée de malade? Notre analyse a pour objectif de déterminer les facteurs qui agissent sur le coût des prestations hospitalières et d'évaluer le poids de leur influence.

(2) Selon la **théorie de l'économie d'entreprise**, les frais dépendent en principe des facteurs suivants:

- productivité;
- prix du marché (resp. volume de la consommation);
- degré de mise à profit (ou d'occupation);
- grandeur de l'exploitation.

Dans le cas d'un hôpital, il faut prendre en outre en considération:

- la durée de séjour des patients;
- la nature et le nombre des services spécialisés;
- le mode de traitement.

Cette énumération n'est certes pas exhaustive, mais elle indique à notre avis les facteurs essentiels.

La **productivité** apparaît au premier rang des facteurs d'influence. Il n'est cependant pas facile de la déterminer dans un hôpital. Nous utilisons une proportion statistique de facteurs, dont on ne peut déduire un rapport de causalité. Cette relation peut être détériorée, par exemple, par les progrès de la médecine, qui exigent plus de personnel soignant et de personnel auxiliaire. La productivité du travail, utilisée dans notre étude, se calcule selon les formules (70 ou 70a) fondées sur la relation générale:

$$P = \frac{\text{Output}}{\text{Input}} = \frac{\text{Production obtenue}}{\text{Travail fourni}}$$



On remarque aussitôt les points faibles de la productivité du travail calculée de cette manière. Le résultat de la production, la journée de malade, est une grandeur assez complexe. Des comparaisons entre plusieurs hôpitaux ne permettent pas nécessairement des conclusions précises. Un moyen d'améliorer la mesure des « sorties » consisterait peut-être à pondérer les journées de malade avec la statistique des malades, en tenant compte de la mise à contribution plus ou moins forte du personnel. Des réserves s'imposent également pour le travail fourni. Le fait de placer l'effectif du personnel au dénominateur de la fraction a certains inconvénients. Il vaudrait mieux prendre les heures de travail; cela donnerait une image plus proche de la vérité pour les hôpitaux qui emploient du personnel à la demi-journée. Aux entrées (input), il faudrait encore résoudre le problème des écoles d'infirmières. D'une manière générale, nous devons reconnaître que, dans notre analyse, la valeur documentaire du facteur d'influence « productivité du travail » reste assez approximative. Il serait peut-être préférable de parler simplement d'une relation entre journées de malade et personnel.

Concernant les **prix du marché**, resp. le **volume de la consommation**, il serait souhaitable que l'on ait de meilleures données statistiques; la documentation disponible laisse passablement à désirer.

Le relevé statistique des autres facteurs d'influence mentionnés est juste à notre avis; et rien ne s'oppose à ce qu'on les introduise dans une régression multiple. Une appréciation quantitative du mode de traitement est toutefois très difficile. Faute de données, il n'a pas été possible de prendre cette variable en considération. Nous avons en revanche fait entrer dans notre modèle sur la dépendance des frais les facteurs : degré de mise à profit, grandeur de l'exploitation, durée des séjours et services spécialisés.

(3) Concernant la méthode, signalons que nous avons utilisé deux procédés:

- Statistique. Par des calculs de régression et de corrélation, il a fallu déterminer les facteurs d'influence (cités par la théorie de l'économie d'entreprise) effectivement propres à expliquer le niveau des frais d'exploitation par journée de malade.
- Economie d'entreprise. Les paramètres nécessaires peuvent être tirés de comparaisons entre établissements ou de séries chronologiques. Notre étude se fonde sur les données des 31 hôpitaux de district du canton de Berne, pour 1966. Les équations de régression sont donc le résultat d'une comparaison entre établissements.

Une analyse comparative de séries chronologiques se fera plus tard (dès que la documentation sera complète).

(4) Comme **données de base** pour les calculs, nous utilisons la statistique des établissements pour malades et le compte annuel des hôpitaux de district, pour l'exercice 1966. Les bulletins statistiques donnent certes de précieuses informations, mais il serait souhaitable que l'on développe encore la documentation fondamentale, c'est-à-dire les renseignements sur la gestion des hôpitaux. Un compte d'exploitation et un compte sur le résultat des prestations, par exemple, seraient très utiles. Il serait aussi très avantageux d'avoir des chiffres mensuels sur certains éléments (nombre de patients, degré d'occupation, etc.).

(5) Au moyen d'une **régression simple, linéaire ou non linéaire**, nous essayons d'abord d'élucider, pour les différents facteurs, si l'on doit admettre ou rejeter l'hypothèse d'une influence sur les frais. Concernant la productivité, les calculs ont démontré que l'on doit admettre l'hypothèse d'une relation linéaire entre ces variables et les frais d'exploitation par journée de malade; les frais relatifs augmentent lorsque la productivité diminue. Les fluctuations de la productivité expliquent environ 76% de la dispersion des variables qui en dépendent (frais d'exploitation). On peut tirer les mêmes conclusions quant aux frais moyens de personnel ( $B = 0,24$ ), les produits alimentaires ( $B = 0,44$ ) et le matériel médical ( $B = 0,69$ ). La régression linéaire simple indique en outre une dépendance certaine (précédée du signe négatif) entre la durée de séjour des patients et les frais relatifs. On peut accepter une relation linéaire entre la grandeur de l'établissement et les frais; mais une relation non linéaire donne de meilleurs résultats, quoique pas très convaincants. Enfin on peut aussi établir une régression linéaire simple entre les salaires et les frais d'exploitation. Une hausse des salaires se traduit naturellement par une augmentation du coût de la journée de malade; c'est un principe bien connu en économie d'entreprise. La comparaison entre le degré d'occupation et les frais relatifs n'a pas donné de résultat positif. Cela n'a rien de surprenant, puisque presque tous les hôpitaux sont fortement ou même très fortement occupés.

(6) Une **comparaison intercantonale** entre les établissements bernois et ceux de la Suisse orientale fait apparaître des résultats inégaux concernant l'influence de la productivité du travail sur les frais. Il

semble donc exister des différences régionales qu'il conviendrait d'analyser de plus près à l'avenir. On pourrait mieux rechercher la possibilité de mettre en pratique des mesures propres à restreindre les frais.

(7) Partant de la théorie de l'économie d'entreprise et des résultats des régressions et corrélations simples, linéaires ou non linéaires, on a réuni des variables dépendantes (frais d'exploitation par journée de malade) et indépendantes dans une **régression** et une **corrélation multiples**. Plusieurs facteurs d'influence n'ont pu trouver place dans l'équation définitive, parce qu'il existait une étroite corrélation entre plusieurs variables indépendantes ou que certaines grandeurs contribuaient de façon négligeable à l'augmentation du total déterminé. En fin de compte, il reste la régression suivante:

$$T = 53,3 - 0,361 P_A + 0,0029 L$$

dans laquelle

T = frais d'exploitation (en fr.) par journée de malade en 1966

P<sub>A</sub> = productivité du travail

L = frais moyens de personnel en 1966 (en fr.).

Les frais d'exploitation par journée de malade augmentent en cas de baisse de la productivité et de hausse des salaires. Ce résultat est modeste du fait que la plupart des facteurs d'influence n'ont pu être considérés dans la régression multiple. Il semble toutefois que le degré de mise à profit, la grandeur de l'établissement, les prix et le volume de la consommation aient des chances de jouer un rôle notable dans une analyse de séries chronologiques.

(8) Il pourrait être judicieux de compléter cette étude, fondée sur une comparaison entre établissements, par une analyse de séries chronologiques. Il serait en outre souhaitable que l'on puisse réunir une documentation meilleure et plus vaste, pour examiner à nouveau les facteurs d'influence qui déterminent les frais dans les établissements pour malades.

Les hôpitaux remplissent une fonction des plus importantes, et l'on ne doit pas compter de trop près quand il s'agit de la santé et de la vie d'êtres humains. Cela ne nuirait pourtant pas que les frais soient soumis à un certain contrôle et qu'on y voue une plus grande attention même dans les hôpitaux.

Berne, mars 1968

### 3 Synopsis grundlegender Ansätze

Ansatz	Nr.
<b>Einfache lineare Regression und Korrelation</b>	
<b>(1) Das Modell</b>	
$y = \alpha + \beta x + \varepsilon, \text{ bzw.}$	(2)
$y_i = \alpha + \beta x_i + \varepsilon_i.$	(3)
<b>(2) Schätzverfahren</b>	
Niveaunkonstante a:	
$a = \bar{y} - b_{yx} \bar{x}$	(7)
Regressionskoeffizient b:	
$b_{yx} = \frac{S(x_i - \bar{x})(y_i - \bar{y})}{S(x_i - \bar{x})^2} = \frac{Sx_i y_i - \frac{1}{N}(Sx_i)(Sy_i)}{Sx_i^2 - \frac{1}{N}(Sx_i)^2} = \frac{S_{xy}}{S_{xx}}$	(8), (10)
Regressionsgleichung:	
$Y_i = \bar{y} + b_{yx}(x_i - \bar{x}), \text{ bzw.}$	(11)
$Y_i = a + b_{yx} x_i$	
Streuung der Einzelwerte:	
$s_{yx}^2 = \frac{1}{N-2} \left\{ S_{yy} - \frac{S_{yx}^2}{S_{xx}} \right\} = \frac{1}{N-2} \left\{ S_{yy} - b_{yx} S_{yx} \right\}$	(16)
Bestimmtheitsmass; Korrelationskoeffizient:	
$B = \frac{S_{xy}^2}{S_{xx} S_{yy}} = b_{yx} b_{xy}$	(17),
$r = \sqrt{B}$	(18)

**(3) Prüfen von Hypothesen**

Varianzanalyse:

Streuung	SQ	FG	DQ
Auf Regression	$b_{yx} S_{xy}$	1	$s_1^2 = b_{yx} S_{xy}$
Um Regression	$S_{yy} - b_{yx} S_{xy}$	$N - 2$	$s_2^2 = s_{yx}^2$
Insgesamt	$S_{yy}$	$N - 1$	.

(21)

Niveaunkonstante a:

$$t = \frac{a - \alpha}{s_{yx}} \sqrt{N}, \quad \text{bzw. } t = \frac{a}{s_{yx}} \sqrt{N} \quad (24)$$

Regressionskoeffizient b:

$$t = \frac{b_{yx} - \beta}{s_{yx}} \sqrt{S_{xx}}, \quad \text{bzw. } t = \frac{b_{yx}}{s_{yx}} \sqrt{S_{xx}} \quad (25), (26)$$

Bestimmtheitsmass:

$$F = \frac{B S_{yy} (N - 2)}{S_{yy} (1 - B)} = \frac{B (N - 2)}{(1 - B)} \quad (27)$$

mit:  $n_1^* = 1$  und  $n_2^* = (N - 2) FG$

**(4) Vertrauensgrenzen der Schätzung**

Regressionsparameter:

$$a \pm t_P s_{yx} / \sqrt{N} \quad (32)$$

$$b \pm t_P s_{yx} / \sqrt{S_{xx}} \quad (31)$$

Regressionswerte  $Y_i$ :

$$s_Y^2 \sim s_{yx}^2 \left\{ \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right\}, \quad \text{bzw.} \quad (33)$$

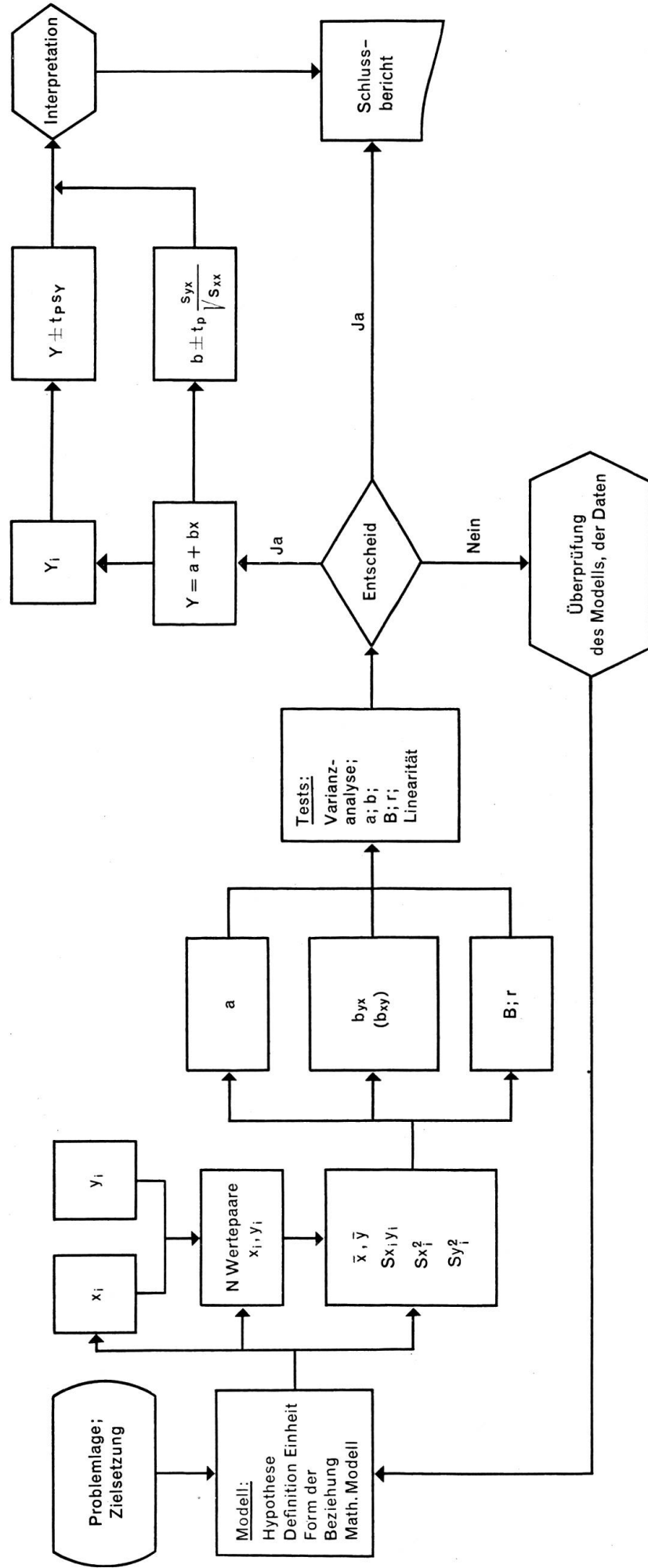
$$Y_i \pm t_P s_Y \quad (34.2)$$

Ansatz	Nr.
<p style="text-align: center;"><b>Einfache nichtlineare Regression</b></p> <p><b>(1) Transformation auf den linearen Fall</b></p> <p style="margin-left: 40px;"><math>Y = ab^x</math>, bzw. <math>\log Y = \log a + x (\log b)</math></p> <p style="margin-left: 40px;"><math>Y = ax^b</math>, bzw. <math>\log Y = \log a + b (\log x)</math></p> <p><b>(2) Mehrfache lineare Regression</b></p> <p style="margin-left: 40px;"><math>Y = a + b_1x_1 + b_2x_2</math>, mit: <math>x_1 = x</math>; <math>x_2 = x^2</math></p> <p><b>(3) Orthogonale Polynome</b></p> <p style="margin-left: 40px;"><math>Y = a_0 + a_1x + a_2x^2 + \dots + a_px^p</math>, bzw.</p> <p style="margin-left: 40px;"><math>Y = A_0 + A_1\varphi_1 + \dots + a_p\varphi_p</math></p>	<p>(39)</p> <p>(40)</p> <p>(41)</p> <p>(42)</p>
<p style="text-align: center;"><b>Mehrfache lineare Regression</b></p> <p><b>(1) Das Modell</b></p> <p style="margin-left: 40px;"><math>y = \alpha + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \varepsilon</math></p> <p><b>(2) Schätzverfahren</b></p> <p>Niveaunkonstante a (3 Variable):</p> $a = \frac{Sy_i - b_1Sx_{1i} - b_2Sx_{2i}}{N}$ <p>Regressionskoeffizienten:</p> $b_1 = \frac{1}{\Delta} \begin{vmatrix} S_{x_1y} & S_{x_1x_2} \\ S_{x_2y} & S_{x_2x_2} \end{vmatrix}; \quad \text{bzw.} \quad b_2 = \frac{1}{\Delta} \begin{vmatrix} S_{x_1x_1} & S_{x_1y} \\ S_{x_1x_2} & S_{x_2y} \end{vmatrix}$ <p>mit: <math>\Delta = \begin{vmatrix} S_{x_1x_1} &amp; S_{x_1x_2} \\ S_{x_1x_2} &amp; S_{x_2x_2} \end{vmatrix}</math></p>	<p>(45)</p> <p>(50)</p> <p>(51)</p>

Ansatz	Nr.
<p>Bestimmtheitsmass (totales):</p> $B_T = \frac{1}{S_{yy}} \{b_1 S_{x_1y} + b_2 S_{x_2y}\}$	(54)
<p><b>(3) Prüfen von Hypothesen</b></p>	
<p>Varianzanalyse:</p> $F = \frac{\text{DQ (auf Regression)}}{\text{DQ (um Regression)}} = \frac{s_1^2}{s_2^2}$	(57)
<p>Partielle Regressionskoeffizienten:</p> $t = \frac{b_j - \beta_j}{s/\sqrt{c_{jj}}} = \frac{b_j - \beta_j}{s_{b_j}}$ <p><math>n^* = (N - p - 1)</math> FG</p>	
<p>Totale Bestimmtheit:</p> $F = \frac{B(N - p - 1)}{(1 - B)p}$ <p>mit: <math>n_1^* = p</math> und <math>n_2^* = (N - p - 1)</math> FG</p>	(58)
<p>Partielle Bestimmtheit:</p> $F = \frac{B(N - 2p)}{(1 - B)p}$ <p>mit: <math>n_1^* = p</math> und <math>n_2^* = (N - 2p)</math> FG</p>	(59)
<p><b>(4) Vertrauensgrenzen der Schätzung</b></p>	
<p>Partielle Regressionskoeffizienten:</p> $b_j \pm t_p s/\sqrt{c_{jj}}, \quad \text{bzw. } b_j \pm t_p s_{b_j}$	(75)
<p>Regressionswerte:</p> $Y \pm t_p s_Y; \quad s_Y^2 \text{ nach (60).}$	

# 4 Rechnungsgang Regressions- und Korrelationsanalyse (einfach, linear)

Problemlage; Zielsetzung; Modell      Daten; Hilfswerte      Parameter; Abhängigkeitsmasse      Tests      Regressionsgleichung      Vertrauensgrenzen      Interpretation; Schlussbericht





# 5 Literaturverzeichnis

## 51 Theorie der stochastischen Abhängigkeit

### 511 Lehrbücher

**Anderson, R.L. and Bancroft, T.A.**

Statistical theory in research, New York 1952

**Bennet, C.A. and Franklin, N.L.**

Statistical analysis in chemistry and the chemical Industry, New York 1961

**Bowker, A.H. and Lieberman, G.J.**

Engineering statistics, Englewood Cliffs, 1965

**Brownlee, K.A.**

Statistical theory and methodology in science and engineering, New York 1965, 3rd ed.

**Bryant, E.C.**

Statistical analysis, New York 1966, 2nd ed.

**Davies, O.L.**

Statistical methods in research and production, London 1961, 3rd ed.

**Fisher, R.A.**

Statistical methods for research workers, Edinburgh-London 1963, 13th ed.

**Goulden, C.H.**

Methods of statistical analysis, New York 1956, 2nd ed.

**Graybill, F.A.**

An introduction to linear statistical models, Vol. I, New York 1961

**Guenther, W.C.**

Concepts of statistical inference, New York 1965

**Johnson, N.L. and Leone, F.C.**

Statistics and experimental design: In engineering and the physical sciences, Vol. I, New York 1964

**Kendall, M.G. and Stuart, B.**

The advanced theory of statistics, Vol. 2: Inference and relationship, London 1961

**Linder, A.**

Statistische Methoden für Naturwissenschaftler, Mediziner und Ingenieure, Basel-Stuttgart 1964, 4. Aufl.

**Ostle, B.**

Statistics in research. Basic concepts and techniques for research workers, Ames, Iowa 1964, 2nd ed.

**Snedecor, G.W.**

Statistical methods, Ames, Iowa 1959, 5th ed.

**Steel, R.G.D. and Torrie, J.H.**

Principles and procedures of statistics, New York 1960

**Weber, E.**

Grundriss der biologischen Statistik, Jena 1967, 6. Aufl.

## 512 Spezialwerke

### **Acton, F.S.**

Analysis of straight-line data, New York 1959

### **Deming, W.E.**

Statistical adjustment of data, New York 1964

### **Draper, N.R. and Smith, H.**

Applied regression analysis, New York 1966

### **Dwyer, P.S.**

Linear Computations, New York 1951

### **Ezekiel, M. and Fox, K.A.**

Methods of correlation and regression analysis. Linear and curvilinear, New York 1959, 3rd ed.

### **Förster, E. und Egermayer, F.**

Korrelations- und Regressionsanalyse. Ein Leitfaden für Oekonomen, Berlin 1966

### **Hellwig, Z.**

Linear regression and its application to economics, Oxford 1963

### **Linnik, J.W.**

Die Methode der kleinsten Quadrate in moderner Darstellung, Berlin 1961

### **Smillie, K.W.**

An introduction to regression and correlation, New York 1966

### **Williams, E.J.**

Regression analysis, New York 1959

## 513 Aufsätze

### **Cochran, W.G.**

The omission or addition of an independent variate in multiple linear regression, in: Suppl. to the Journal of the Royal Statistical Society, 1938, Vol. V, S.171ff.

### **Hader, R.J. and Grandage, A.H.E.**

Simple and multiple regression analysis, in: Chew, V.: Experimental designs in industry, New York 1958, S.108-137

## 52 Tabellen/Formelsammlungen

### **Fisher, R.A. and Yates, F.**

Statistical tables for biological, agricultural and medical research, Edinburgh 1963, 6th ed.

### **Graf, U., Henning, H.-J. und Stange, K.**

Formeln und Tabellen der mathematischen Statistik, Berlin 1966

### **Linder, A.**

Handliche Sammlung mathematisch-statistischer Tafeln, Basel 1961

### **Smirnov, N.V.**

Tables for the distribution and density functions of t-distribution («Student's» distribution), Oxford 1961

## 53 Betriebswirtschaftslehre

**Anthony, R.N.**

Management accounting principles, Homewood, Ill., 1965

**Bureau International du Travail (BIT)**

Le coût des soins médicaux, Etudes et documents, Nouvelle Série no 51, Genève 1959

**Cobb, C.W. and Douglas, P.H.**

A theory of production, A. E. R., 1928, Suppl.

**Damrau, W.F.**

The rise of municipal hospital expenditures in New York City, 1914–1954, Diss. New York 1957

**Fourastié, J.**

La productivité, Paris 1965

**Gonzenbach, F.**

Unterlagen für Buchhaltung und Statistik der schweizerischen Krankenanstalten. Normalkontenplan, Buchungskatalog, Anleitung für Krankenanstaltsstatistik, Aarau 1960, 2. Aufl.

**Graf, A., Hunziker, A. und Scheerer, F.**

Betriebsstatistik und Betriebsüberwachung, Zürich 1961, 2. Aufl.

**Lembcke, P.A.**

Hospital efficiency – a lesson from Sweden in: Hospitals J.A.H.A., April 1, 1959, Vol. 33

**Montacute, Ch.**

Costing and efficiency in hospitals, London 1962

**Schnettler, A.**

Betriebsanalyse, Stuttgart 1958

**Schnettler, A.**

Betriebsvergleich, Stuttgart 1961, 3. Aufl.

**Taylor, Ph.J. and Nelson, B.O.**

Management accounting for hospitals, Philadelphia 1964

**Viel, J.**

Betriebs- und Unternehmungsanalyse, Zürich 1958, 2. Aufl.