Vom Wesen und Unwesen der Mathematischen Statistik

Klaus-Dieter Feige

Feige, K.-D. 2006: **On the use and misuse of mathematical statistics.** Ber. Vogelwarte Hiddensee 17: 65-71.
Although mathematical statistics historically do not have the best reputation, there is no doubt that any serious investigation, not only in the field of ornithology, cannot waive it. To encourage volunteer researchers to use statistical methods, the basic principles and assumptions are explained as simple as possible. Some examples serve to show how descriptive and test statistics should be understood and applied. Conclusions are drawn e.g. with regard to an appropriate planning of data sampling in the field. However, even though mathematical statistics can help to understand empirical data common sense remains important to judge the validity of scientific conclusions.

1. Statistik in Zitaten

Die Statistik genießt nicht immer den besten Ruf. Manche glauben, dass die Statistik von Betrügern erfunden wurde – oder zumindest von Politikern. Kaum eine Zeitung, Zeitschrift, Nachrichtensendung oder Dokumentation kommt heute zur Begründung einer Aussage ohne mehr oder weniger plausible Tabellen, Diagramme oder Kurven aus, die sich auf vorgeblich wissenschaftlich erhobene Daten aufbauen. Da diese Aussagen nicht selten von der erlebten Wirklichkeit abweichen, hat nicht nur die (mathematische) Statistik ein Imageproblem.

Das berühmteste Zitat in dieser Hinsicht: „**Ich glaube nur der Statistik, die ich selbst gefälscht habe...**" wird ja gemeinhin WINSTON CHURCHILL (1874-1965) zugesprochen. Dass auch dies wohl eine Legende ist, macht das Imageproblem nicht geringer. Der tatsächliche Urheber dieses Zitats, JOSEPH GOEBBELS (1897-1945), legte dieses WINSTON CHURCHILL mehrfach in den Mund, um diesen und damit auch die Statistik zu diskreditieren.

Selbst der 32. Präsident der USA, FRANKLIN D. ROOSEVELT (1882-1945) meinte: „**Ich stehe Statistiken etwas skeptisch gegenüber. Denn laut Statistik haben ein Millionär und ein armer Schlucker je eine halbe Million.**"

Aber auch in der Gegenwart sieht es dahingehend noch nicht viel besser aus. FRANZ STEINKÜHLER (ehemaliger Gewerkschaftsführer, geb. 1937) äußerte sich zum Thema wie folgt: „**Ich denke bei „Statistik" an den Jäger, der an einem Hasen beim erstenmal knapp links vorbei schoss und beim zweitenmal knapp rechts vorbei. Im statistischen Durchschnitt ergäbe dies einen toten Hasen.**"

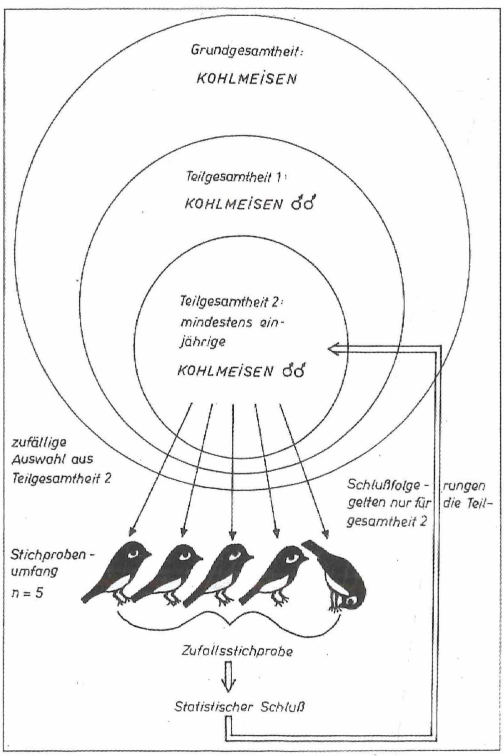
Angesichts eines derart üblen Leumunds - verbreitet durch gewichtige Multiplikatoren - erscheint die Meinung von Bankier HERMANN JOSEF ABS (1901-1994) schon fast wieder sympathisch: „**Die Statistik ist wie eine Laterne im Hafen. Sie dient dem betrunkenen Seemann mehr zum Halt als zur Erleuchtung...**"

Etwas derber formuliert dies EDOUARD HERRIOT (1872-1957), französischer Politiker, Radikalsozialist, 1924/25 u. 1932 Ministerpräsident, 1947-1954 Präsident der französischen Nationalversammlung): „**Die Statistik ist eine sehr gefällige Dame. Nähert man sich ihr mit entsprechender Höflichkeit, dann verweigert sie einem fast nie etwas.**"

WALTER KRÄMER, Professor für Wirtschafts- und Sozialstatistik in Dortmund, hat die Fehlermöglichkeiten der Statistik in einem Buch aufgegriffen: „So lügt man mit Statistik", Campus-Verlag, 8. Auflage 1998. Die mathematische Statistik hat diesen Ruf aber kaum verdient. Man macht ja schließlich nicht das Auto dafür verantwortlich, wenn der Fahrer einschläft und gegen einen Baum fährt. Ähnlich ist es mit den Aussagen der mathematischen Statistik. So kann man ELISABETH NOELLE-NEUMANN (*1916), deutsche Marktforscherin, folgen, die da meint: „**Für mich das Informationsmittel der Mündigen. Wer mit ihr umgehen kann, kann weniger leicht manipuliert werden. Der Satz „Mit Statistik kann man alles beweisen" gilt nur für die Bequemen, die keine Lust haben, genau hinzusehen.**"

Oder auch MARTIN KRUSE (*1929), deutscher evangelischer Theologe, 1985-1991 Vorsitzender des Rates der EKD: „**Die Statistik ist eine Wanderkarte: Wenn man sie zu sehen bekommt, ist sie von der Realität schon etwas überholt. Dennoch gibt sie Orientierung. Man muss sie mit Verstand lesen können, sonst geht man in die Irre.**"

Und mit diesem versöhnlichen Abschlusszitat können wir uns getrost dem eigentlichen Anliegen der mathematischen Statistik zuwenden. Über kurz oder lang sind Vorurteile sowieso nicht zu beheben. Und auch in der Branche der Statistiker wird es immer schwarze Schafe geben. Die Statistik ist nicht um ihrer selbst Willen da, sondern ein geeignetes Werkzeug in der Hand des kundigen Wissenschaftlers.



2. Das Grundprinzip der mathematischen Statistik

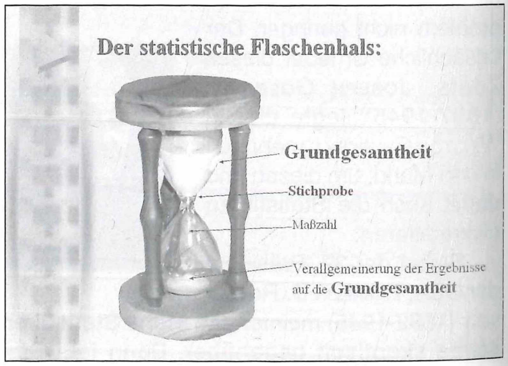
2.1 Grundgesamtheit und Stichprobe

Voraussetzung für die Anwendung mathematisch-statistischer Verfahren ist die mehr oder weniger zufällige Variabilität eines messbaren Merkmals in einer definierten Gruppe von Objekten. Voraussetzung ist weiterhin, dass wir nur einen Teil der Gesamtheit (Stichprobe) aller Objekte dieser Gruppe (Grundgesamtheit, Population) messen können oder wollen.

Das Ziel der Statistik besteht nun darin, anhand der Messungen des betrachteten Merkmals in der Teilpopulation eine hinreichend gesichert Verallgemeinerung zum Merkmal auf die Gesamt-Population erhalten zu können. Die Grundgesamtheit muss quasi durch den Flaschenhals der Stichprobe hindurchlaufen um dann aus diesen zufälligen Daten wieder auf die Gesamtheit schließen zu können.

2.2 Verteilungsvoraussetzungen

Für die korrekte Bewertung einer Stichprobe aus einer Grundgesamtheit müssen wir genaugenommen die tatsächliche Verteilung des Merkmals in der Population kennen! Da wir diese in der Regel aber nicht kennen, nutzen wir Näherungen oder versuchen uns an „verteilungsfreien" Verfahren der Auswertung.



Typische idealisierte Verteilungen sind die in jedem Lehrbuch detailliert beschriebenen Normalverteilungen, Poissonverteilungen oder die Binomialverteilungen. Die Menge weiterer denkbarer und plausibel hergeleiteter Verteilungen ist schier unbegrenzt, was den Nicht-Mathematiker jedoch eher irritiert, denn hilft. In den meisten Fällen genügen aber die Näherungen aus den „griffigen" Verteilungen – wenngleich mit einem Risiko der Gesamtgenauigkeit für unsere Interpretationen auf die Grundgesamtheit oder Population.

2.3 Risiken der Verallgemeinerung

Die Risiken der statistischen Verallgemeinerungen lassen sich jedoch minimieren mit

1. der genauen Definition der Grundgesamtheit,
2. der exakten Zufallsstichproben-Entnahme,
3. der genauen Messung der Zieldaten,
4. der hinreichenden Kenntnis der Verteilungsvoraussetzungen des gemessenen Merkmals,
5. der genauen Berechnung der Maßzahlen,
6. dem gesunden Menschenverstand bei der Interpretation der Ergebnisse.

Man sieht sofort, dass selbst bei Beachtung der Positionen 1-5 der Imageverlust der Statistik durch den bewussten oder unbewussten Missbrauch bei der Interpretation der Ergebnisse entsteht. Diverse Beispiele aus der vieljährigen Beratungspraxis können dies teilweise auf kuriose Weise belegen.

Die Risiken der Verallgemeinerungen von der Stichprobe auf die Grundgesamtheit unterscheiden sich auf zumindest zweierlei Art. Risiken sind in diesem Sinne Wahrscheinlichkeiten dafür, dass eine mit einem statistischen Test gefundene verallgemeinerte Aussage falsch ist! Das **Risiko 1. Art** definiert die Wahrscheinlichkeit dafür, dass ein gefundener Unterschied nicht existiert, die Testaussage also falsch ist. Das **Risiko 2. Art** kennzeichnet die Wahrscheinlichkeit, dass die Alternativaussage: „Es besteht kein Unterschied" – nicht wahr ist.

Beispiel: Bei einem statistischen Test wurde festgestellt, dass die mittleren Flügellängen von Kohlmeisenmännchen und –weibchen einer bestimmten Teilpopulation verschieden sind. Die Aussage gilt mit einem Risiko 1. Art á = 0,05, wobei der zu erreichende Risikowert vor der Testausführung festgelegt wird. Bei einem anderen vergleichenden Test konnte für Grünfinken hinsichtlich des gleichen Merkmals kein Unterschied zwischen den Geschlechtern nachgewiesen werden. Diese Aussage gilt mit einem Risiko 2. Art â = 0,10, das ebenfalls vor der Testdurchführung definiert wurde. Die Risiken 1. und 2. Art sind verfahrensgerecht vor einem Test festzulegen.

Die Praxis sieht insbesondere beim Risiko 1. Art jedoch anders aus. Anhand der berechneten Testmaßzahl wird das jeweilige Risiko einer Falschaussage anhand des Testergebnisses nachträglich bestimmt. Damit begeht man jedoch einen schwerwiegenden methodischen Fehler.

Wenn man aus einer Grundgesamtheit mehrere unabhängige Zufallsstichproben zieht – das selbe Experiment also mehrfach in derselben Grundgesamtheit wiederholt - kann man feststellen, dass auch die Testmaßzahlen einer Variabilität unterliegen und damit eine Verteilung haben. Das nachträgliche Definieren der Sicherheit der Testaussage wäre dann so, als würde man versuchen, sich an den eigenen Haaren aus dem Statistiksumpf zu ziehen. Auch diese tausendfach in der Literatur vollzogene Auswertungspraxis trägt so zum Imageverlust der Statistik bei, da sie im Durchschnitt höhere Genauigkeiten der Untersuchungen vortäuscht, als tatsächlich vorhanden sind. Durch das Nachführen der Risikoangabe begeht man also einen **Fehler 3. Art**!

Das Risiko 2. Art wird in den Veröffentlichungen meistens nicht einmal erwähnt, wenn es denn den Autoren überhaupt bekannt ist. Es kann aber durchaus schon einmal größer 0,5 sein! Und das bedeutet, dass trotz der behaupteten Gleichheit aufgrund einer z.B. zu kleinen Stichprobe, dennoch die Wahrscheinlichkeit, dass ein Unterschied zwischen den zu vergleichenden Grundgesamtheiten besteht, größer ist als die Testaussage. Es ist fatal, etwas ruhigen Gewissens zu behaupten, aber trotzdem die Alternativbehauptung wahrscheinlicher ist. Wir sollten uns auch als Biologen im Umgang mit der mathematischen Statistik keine derartigen Patzer mehr leisten.

3. Merkmalsskalierungen und Variablen

Die in der Biologie zu erfassenden Merkmale lassen sich wie folgt systematisieren:
Qualitative Merkmale:
Dichotome oder duale Skalierung
(z.B. männlich, weiblich)
Nominale Skalierung
(z.B. Müller, Meier, Lehmann)
Ordinale Skalierung
(z.B. Stadium I, Stadium II, Stadium III)
Quantitative Merkmale:
Diskrete Skalierung
(z.B. Gelegegröße oder Federzahl...)
Metrische Skalierung
(z.B. Flügellänge oder Gewicht...)

Dabei wächst der **Informationsgehalt** der Merkmale in der obigen Auflistung von oben nach unten – und damit fallen die Stichprobenumfänge, die notwendig sind, um gleichwertige Aussagen aus Stichproben zu erhalten.

4. Deskriptive Statistik

Die beschreibende oder deskriptive Statistik umfasst ursprünglich die grafische Darstellung der gemessenen Daten. Dafür bieten heute viele Datenbankprogramme variable Möglichkeiten, wie z.B. die Histogramme (siehe unten). Bereits diese einfachen Datenaufbereitungen lassen viele Möglichkeiten der Interpretation zu und sind vielmehr noch Grundlage der Thesenbildung. Sie erleichtern dann auch die Interpretation der statistischen Maßzahlen.

4.1 Mittelwert, Median, geometrisches Mittel, gestutzte Verteilungen

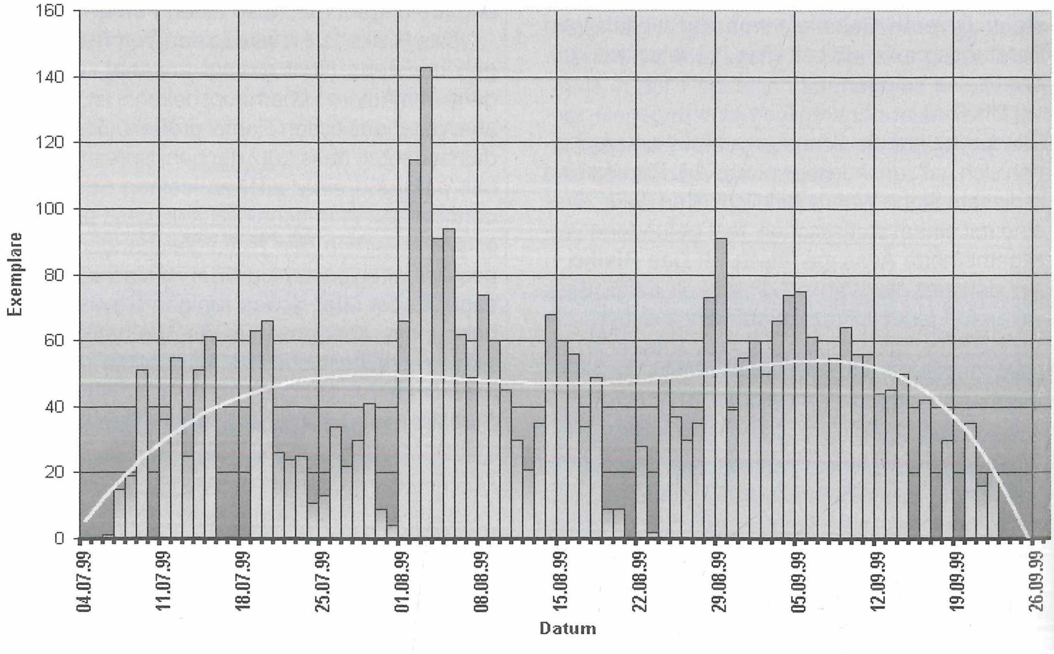
Auch wenn die Maßzahl eines irgendwie typischen Mittelwertes nicht immer geeignet erscheint (z.B. bei mehrgipfligen Verteilungen), sind sie dennoch die häufigsten Maßzahlen in der ornithologischen Literatur – und nicht nur dieser.

Neben dem klassischen arithmetischen Mittelwert sind diverse Variationen entwickelt worden, die die Maßzahl robust gegen Ausreißer, Verteilungsabweichungen u.a. machen sollen.

In Rasch u.a. (1996): „Verfahrensbibliothek Versuchsplanung und –auswertung", München, Wien sind die wichtigsten mittelnden Maßzahlen, aber auch deren Verwendung begründet dargestellt worden (siehe Übersicht 1 und 2).

4.2 Maßzahlen der Datenstreuung

Ähnlich mannigfaltig wie die Mittelwert-Maße sind die Maßzahlen der Datenstreuung. Hierzu zählen mehr oder weniger kompliziert zu berechnende Kenngrößen wie Spannweite, Varianz, Stan-



Übersicht 1: Hinweise zur Auswahl geeigneter Maßzahlen der Lage

Skalierung der Beobachtungswerte		Form der Häufigkeitsverteilung		
		Eingipflig		Mehrgipflig
		Symmetrisch	Asymmetrisch	(zunächst prüfen, ob Zerlegung möglich und zweckmäßig ist)
ordinal-skaliert		Zentralwert	Zentralwert Modalwert	Zentralwert Modalwerte
metrisch	intervall-skaliert	arithmetisches Mittel Zentralwert Modalwert Quantile	Modalwert Zentralwert arithmetisches Mittel Quantile	Modalwerte Zentralwert arithmetisches Mittel
	verhältnis-skaliert	arithmetisches Mittel Zentralwert (falls n sehr klein) Modalwert Quantile geometrisches Mittel harmonisches Mittel	Modalwert Quantile Zentralwert arithmetisches Mittel geometrisches Mittel	Modalwerte Zentralwert

Übersicht 2: Maßzahlen der Lage und ihre Eigenschaften

Maßzahl	Vereinfachte Berechnungsmethode	Eigenschaften, Bemerkungen
Arithmetisches Mittel $y = \frac{1}{n}\sum_{i=1}^{n} y_i$	Nach Klasseneinteilung (siehe Verfahren **2/11/0003**): $\bar{y} = A + \frac{h}{n}\sum_{i=1}^{k} f_i z_i$	Die Summe der Abweichungen aller Beobachtungswerte y_i von ihrem arithmetischen Mittel $\bar{y}$ ist Null. Die Summe der Quadrate der Abweichungen aller Beobachtungswerte von ihrem arithmetischen Mittel ist eine Minimum: $SQ_y = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \min_b \sum_{i=1}^{n} (y_i - b)^2$
Gewogenes arithmetisches Mittel $\bar{y}_g = \frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i}$	Nach Klasseneinteilung: $\bar{y}_g = A + h \frac{\sum_{i=1}^{k} f_i w_i z_i}{\sum_{i=1}^{k} f_i w_i}$	Als Gewichte w_i können auftreten: - Stichprobenumfänge, falls die y_i Mittelwerte von Stichproben verschiedenen Umfanges sind - Wichtung auf Grund sachlogischer Überlegungen
Zentralwert (Median) $Z = \begin{cases} y_{\left(\frac{n+1}{2}\right)} & \text{falls n ungerade} \\ 0{,}5\left(y_{\left(\frac{n}{2}\right)} + y_{\left(\frac{n}{2}+1\right)}\right), & \text{falls n gerade} \end{cases}$	Nach Klasseneinteilung: $Z = y''_{j-1} + \frac{h}{2}\left(\frac{n - 2\sum_{i=1}^{j-1} f_i}{f_j}\right)$, falls Z in die Klasse j fällt	Z halbiert die nach ihrer Größe geordnete Rangreihe der Beobachtungswerte. Die Summe der absoluten Abweichungen aller Beobachtungswerte y_i von Z ist ein Minimum: $\sum_{i=1}^{n} \lvert y_i - Z \rvert = \min_b \sum_{i=1}^{n} \lvert y_i - b \rvert$
Modalwert $Mo = y_{[\max(f_j)]}$ Abszissenwert eines Gipfels des Häufigkeitspolygons		Vorteile: - Kann leicht aus Häufigkeitstabelle oder graphischer Darstellung abgelesen werden. - Ist zur Charakterisierung mehrgipfliger Verteilungen geeignet. Nachteile: - Streut stärker als $\bar{y}$. - Oft gibt es mehrere Modalwerte.
Geometrisches Mittel $G = \sqrt[n]{\prod_{i=1}^{n} y_i}$	$\lg G = \frac{1}{n}\sum_{i=1}^{n} \lg y_i$	Anwendung bei Verhältniszahlen und Zeitreihen Voraussetzung: $y_i > 0 \; (i = 1, \ldots, n)$
Harmonisches Mittel $H = \frac{n}{\sum_{i=1}^{n} \frac{1}{y_i}}$		Anwendung als Mittel von Leistungslimits bei freiem Zeitverbrauch Voraussetzung: $y_i > 0 \; (i = 1, \ldots, n)$

dardabweichung, Variationskoeffizient, Quantilabstände, mittlere absolute Abweichung. Eine ähnliche Übersicht wie bei den Maßzahlen der Lage findet sich ebenfalls bei Rasch u.a. (1996) und in vielen biomathematischen Lehrbüchern.

5. Statistische Tests

Neben den beschreibenden Verfahren bilden die statistischen Tests das Herz der mathematischen Statistik. Hierbei wird jeweils eine These gegen die zugehörige Antithese geprüft. Dabei besteht in der Regel das Ziel darin, eine Differenz nachzuweisen, also die Ungleichheit zwischen zwei Populationen nachzuweisen:

Ziel H_A	**Annahme** H_0	
ungleich ($\neq$)	gleich ($=$)	Zweiseitiger Test
größer ($>$)	kleiner oder gleich ($\leq$)	Einseitige Tests
kleiner ($<$)	größer oder gleich ($\geq$)	

Beispielsweise könnte die Nullhypothese lauten: *Zwischen der mittleren Gelegegröße von spanischen und polnischen Kohlmeisen besteht kein Unterschied* gegenüber der Alternativhypothese: *Es gibt einen Unterschied (egal ob größer oder kleiner).*

Im Fall des einseitigen Testes sollte schon eine begründete Hypothese bestehen, die dann zwischen den Thesen: *Die mittlere Gelegegröße spanischer Kohlmeisen ist größer oder zumindest gleich der von polnischen Kohlmeisen* gegen *Die mittlere Gelegegröße spanischer Kohlmeisen ist kleiner als die von polnischen Kohlmeisen.*

Am Beispiel des Tests der Fangwahrscheinlichkeit der beiden Geschlechter von verschiedenen Meisenarten in einem Kontrollgebiet, soll die These: *Bei allen Arten ist die Fangwahrscheinlichkeit gleich groß* gegen *Die Wahrscheinlichkeiten sind verschieden* geprüft werden. Die folgende Abbildung demonstriert das Testverfahren und belegt hier durch die signifikante Testwertprüfung die Ungleichhypothese.

Die Vielzahl der Tests darstellen zu wollen, wäre hier schlichtweg unmöglich. RASCH u.a. (1996) versuchen eine Systematisierung der wichtigsten tausend Tests. Das machen aber auch viele andere Lehrbücher der Biometrie und

Chi²-Test

$n_{ges}=77$	Kohlmeise	Blaumeise	Tannenmeise	
männlich	13	3	5	21
weiblich	15	22	19	56
	28	25	24	

$$\chi^2_{df=(k-1)\times(l-1)} = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(f_{b(i,j)}-f_{e(i,j)})^2}{f_{e(i,j)}} = 8{,}61$$

$$\chi^2_{krit,df=2,\alpha=0.05} = 5{,}99 \Rightarrow \text{signifikant}$$

Statistik. Man muss nur den Mut haben, da mal hinein zu sehen.

6. Weitere statistische Schätzverfahren

Der größte Bereich der mathematischen Analyse umfasst die Bewertung von Wechselwirkungen zwischen verschiedenen Merkmalen (Faktoren) in einer Population oder zwischen verschiedenen Populationen. Ein typisches Maß für den linearen Grad der Wechselbeziehungen zwischen zwei gleichzeitig an den Objekten einer Population zufällig auftretenden Merkmalen ist der **Korrelationskoeffizient**. Nur auch von diesen gibt es je nach Definition des Zusammenhangs verschiedene Berechnungsvarianten. Die lineare Beziehung zwischen den Merkmalen kann zudem mittels **linearer Regression** als mathematische Funktion dargestellt werden. Und betrachtet man gar ein Zielmerkmal als Funktion von mehreren Einflussgrößen, so spricht man von multipler Korrelation und je nach Regressionsmodell von **mehrfach quasilinearer Regression, polynominaler Regression** oder bei besonders schwierig zu determinierenden Modellen von **eigentlich nichtlinearer Regression**.

Die Palette der in der Biomathematik genutzten Algorithmen und Verfahren lässt sich noch weiter ausbreiten. Das kann aber hier nicht das Ziel einer Einführung in die Mannigfaltigkeit der mathematischen Statistik sein. Trotz der scheinbar unübersichtlichen Vielfalt möchte ich Sie ermuntern sich von allen Vorurteilen nicht abschrecken zu lassen (und sei es mit Hilfe eines Statistikers), ihre oft mühsam erhobenen Daten einer genaueren Analyse zu unterziehen. Der Wissenszuwachs und die Sicherheit der eigenen Aussagen gewinnen.

7. Statistische Versuchsplanung – warum und wie?

Um Kosten zu sparen und um auch reproduzierbare Ergebnisse mit der gewünschten Genauigkeit zu erhalten ist eine statistische Versuchsplanung notwendig.

Mit ihr werden

- die Versuchsfragen präzisiert,
- das statistische Modell gewählt,
- die Versuchsanlage konstruiert und gewählt,
- der Stichprobenumfang berechnet und
- die Versuchsauswertung geplant.

Versuche mit falschem Aufbau oder zu geringem, aber auch zu großem Datenumfang kosten wegen der mangelnden Ergebnisgenauigkeit oder der höher als notwendig ermittelten Präzision die Wirtschaft oft Milliarden an Euro, meist viel Zeit und Arbeit, und sie sind damit für die Arbeit der Statistiker imageprägend.

8. Tücken der Statistik

Bei allen **Problemen**, die sich bereits im Vorfeld der Datenerhebung und Modellbildung ergeben: Bleiben Sie gegenüber den nackten ausgerechneten Zahlen kritisch! Holen Sie diese mit dem **gesunden Menschenverstand** (also ihrem Fachwissen) wieder ins Leben zurück! Achten Sie auf die Risiken, die sie durch die Stichprobenentnahmen eingegangen sind, und sie stehen nicht als potentieller Betrüger da.

Die Mathematische Statistik ist nur ein **Werkzeug** für den Motor Ihres Autos, mit dem sie durch sumpfiges Gelände müssen. Nicht mehr, aber auch nicht weniger...

Anschrift des Autors:

Dr. Klaus Dieter Feige
Lewitzweg 23
19372 Matzlow