Ao. Univ. Prof. Dr. Günther OSSIMITZ
Universität Klagenfurt

MESSEN ALS GRUNDVORAUSSETZUNG VON MODELLBILDUNG UND SIMULATION

(Version Frühjahr 2003)

1. GRUNDLEGENDES ZUM MESSSEN

1.1. Was versteht man unter "Messen"?

"Messen" ist eine der zentralen Aktivitäten zur Ordnung unserer heutigen Welt. Wir messen Längen, Massen, Zeiten, Temperaturen usw. längst nicht nur mehr in physikalischen Versuchen, sondern auch im Alltag. Im Haushalt werden etwa gemessen: die verbrauchte Strommenge, die Wassermenge, die Kochzeit eines Frühstückseies, die Zimmertemperatur, die Wohnfläche, die Wärmedurchlässigkeit von Mauern und Fenstern, Telefongebühreneinheiten usw. Rund um das Auto werden etwa Benzinverbrauch, Fahrzeiten, Geschwindigkeiten, Abgaswerte, Zündzeitpunkt, Verkehrsaufkommen, Unfallrisiken etc. gemessen. Auch im Bereich der Medizin etwa wurden in den letzten Jahrzehnten eine Unzahl neuer Messverfahren zur Verfeinerung der Diagnostik entwickelt.

Im Gegensatz zur wohlentwickelten ***Praxis des Messens***, die seit Beginn der Neuzeit (im "naturwissenschaftlichen Zeitalter", vgl. dazu PIETSCHMANN 1980) einen ungebrochenen Aufschwung erlebt, steckt die ***Theorie des Messens*** heute noch in ihren Kinderschuhen. Eine systematische Auseinandersetzung mit den theoretischen Grundlagen des Messens begann erst in den Dreißigerjahren des 20. Jahrhunderts. Ein wesentlicher Impuls dürfte dabei von der damals einsetzenden Revolution des physikalischen Weltbildes durch die Quantenmechanik und Relativitätstheorie ausgegangen sein. Diese Theorien kamen im wesentlichen zu dem Schluss, dass speziell in den Grenzbereichen physikalischer Messung der Messvorgang selbst und damit indirekt der "Beobachter" das Messergebnis zum Teil ganz erheblich beeinfluss t. Ein einfaches Beispiel dafür ist etwa das Phänomen, dass bei jeder Temperaturmessung das Thermometer auch die Temperatur des zu messenden Gegenstandes beeinfluss t, speziell dann, wenn das zu messende Objekt extrem klein ist. Noch krasser und unmittelbar ersichtlich ist der Einfluss des messenden Organs dort, wo die Messung nicht durch ein Messgerät, sondern durch eine Person (etwa einem Wertungsrichter beim Turnen oder einem Lehrer bei einer Schularbeit) erfolgt.

Entscheidende ***Beiträge zur Messtheorie*** kamen sowohl von Grundlagenforschern (z. B. Wissenschaftsphilosophen, Mathematikern) als auch von Anwendungswissenschaftlern (wie Psychologen oder Ökonomen), die die in ihren Wissenschaften üblichen Messverfahren kritisch reflektieren wollten. Mittlerweile ist die Messtheorie zu einem eigenen Teilgebiet der Angewandten Mathematik geworden. In diesem Artikel möchte ich an einigen Beispielen zeigen, welchen Nutzen man aus theoretischen Reflexionen für das praktischen Messen - speziell im Bereich der Statistik - ziehen kann.

Was versteht man nun unter dem Begriff "messen"? In der Regel wird durch eine Messung ein Sachverhalt quantitativ - also durch Zahlen - dargestellt. Ein Meteorologe etwa kann die gemessene Temperatur eines Ortes zu einem bestimmten Zeitpunkt genauso wie die Luftfeuchtigkeit oder den Luftdruck durch eine Zahl wiedergeben. Bei der Beurteilung der Bewölkungsstruktur wird hingegen eher eine Zuordnung zu bestimmten Wolkentypen (Cumulus, Stratus usw.) und keine zahlenmäßige Bewertung vorgenommen. Die Gesamtbedeckung des Himmels (0/8 = wolkenlos, 8/8 = völlig bedeckt) und die Höheangabe für verschiedene Wolkenschichten erfolgt wieder durch Zahlen. *Messungen sind also in der Regel zahlenmäßige Zuordnungen, fallweise auch Zuordnungen zu bestimmten Typen (Klassifikationen)*. Allerdings ist nicht jede Zuordnung von Zahlen als eine Messung zu betrachten! Die einem Telefonanschluss zugeordnete Rufnummer wird genauso wie die Identifikationsnummer eines Buches in einer Bibliothek kaum als das Ergebnis einer Messung verstanden. Genauso wird die Zuordnung einer Pflanze zu einer bestimmten Gattung innerhalb der Pflanzensystematik kaum als Messung betrachtet. In all diesen Fällen dient die Zuordnung von Zahlen bzw. Kategorien zu Zwecken der Identifikation bzw. Systematisierung.

*Von "**Messen**" sprechen wir eher dann, wenn eine Zuordnung von Zahlen bzw. eine Klassifikation zum Zwecke einer (vergleichenden) Beschreibung bzw. Beurteilung eines Sachverhaltes erfolgt*. Zum Vergleich können dabei andere Messwerte oder auch vorgegebene Sollwerte dienen. Bereits in dieser Definition des Begriffs "Messen" ist eine *Zwecksetzung notwendig impliziert*. Ohne Zweck kann es kein Messen geben, und der Zweck bestimmt ganz nachhaltig, was und wie gemessen wird. Dies wird durch die folgenden Beispiele noch deutlicher werden. Der Zweck liegt auch außerhalb des Messobjektes immer im die Messung durchführenden oder auswertenden "Beobachter". Damit wird bereits auf einer definitorisch-grundsätzlichen Ebene klar, dass es ein *Messen "unabhängig vom Beobachter"* nicht geben kann. In der Regel ist es um der Vergleichbarkeit willen notwendig, durch Angabe genauer Messvorschriften die Freiheit des einzelnen Beobachters beim Messvorgang einzuschränken. Dennoch bleibt ein jeder Messung zugrundeliegender Zweck erhalten. Beispielsweise wäre es für eine Volkszählung fatal, wenn es dem Erhebungsorgan überlassen bliebe, welche Fragen gestellt werden oder auch nur, wie die einzelnen Fragen formuliert werden. Durch die Standardisierung der Erhebung wird eine möglichst einheitliche Erfassung einer großen Zahl von Daten angestrebt. Erst dadurch werden die Zusammenfassung einzelner Daten, der Vergleich zwischen verschiedenen Regionen oder auch zwischen zwei Volkszählungen möglich - darin besteht u.a. auch der Zweck derartiger Erfassungen.

1.2. Reduktion von Komplexität

Jede Messung reduziert Komplexität. Aus der Vielzahl von Attributen eines Sachverhalts wird eine einzelne Zahl oder eine einfache Zugehörigkeit zu einer Kategorie herausabstrahiert. Bei einer Längenmessung sieht man von der Beschaffenheit des Raumes zwischen den beiden Endpunkten der Mess-Strecke ab; bei der Ermittlung des Wasserverbrauchs eines Haushaltes ist es unerheblich, wann innerhalb des betrachteten Zeitraumes, wozu und von wem das Wasser verwendet wurde. Auch die Qualität des Wassers wird von der Wasseruhr nicht erfasst. Bei einer Blutuntersuchung wird die komplizierte chemische Zusammensetzung des Blutes auf einige wenige Kennwerte reduziert. Bei der schulischen Notengebung wird ein höchst komplexes und einmaliges Schülerverhalten über ein ganzes Schuljahr zu einer einzigen Zahl auf einer fünfstufigen Skala reduziert usw. Diese Reduktion bedingt auch ein Umschlagen von ***Qualität***[1] in ***Quantität***. Das Gewicht einer Substanz sagt nichts über die Art der Substanz und ihre Qualität (z.B. im Hinblick auf Verunreinigungen) aus. Das allgemeine gesundheitliche Befinden eines Menschen, die Lebensqualität eines Landstriches oder die Qualität eines Autos lassen sich kaum oder gar nicht durch eine einzelne Messung quantifizieren. Wird der Begriff "Qualität" in einem umfassenderen Sinn verstanden, so ist es nur in Ausnahmefällen möglich, durch Messen Qualität hinreichend zu erfassen. Selbstverständlich können *notwendige Voraussetzungen für eine bestimmte Qualität* gemessen werden. Beispielsweise ist es für die Qualität von Nahrungsmitteln notwendig, dass sie frei von Giftstoffen sind. Allerdings kann durch chemische Analysen kaum eine geschmackliche Überprüfung ersetzt werden. Genauso kann eine Messung verschiedener Kennwerte im menschlichen Blut zwar Hinweise auf verschiedene Krankheiten oder für das Zusammenpassen von Spender- und Empfängerblut liefern. Dennoch wird bei Bluttransfusionen die Eignung einer bestimmten Blutkonserve für einen Patienten erst durch einen empirischen Test ("Kreuzblut") hinreichend sichergestellt. *Diese Beispiele sollen andeuten, dass es prinzipiell ein Problem ist, durch quantitative Analysen umfassende qualitative Aufschlüsse* zu gewinnen.

Manchmal kann durch eine *Vervielfachung der Quantität* ein genaueres Wissen über die Qualität (zumindest teilweise) wieder zurückgewonnen werden. Beispielsweise wird durch die Messung eines Schalldruckes zwar eine Maßzahl für die Lautstärke eines Geräusches bzw. Klanges gewonnen, über die Art des Signals gibt die Messung jedoch keine Auskunft. Wird hingegen der Schalldruck in sehr kurzen Zeitabständen erfasst (bei der modernen CD-Audiotechnik etwa ca. 45.000 mal pro Sekunde), so kann aus dieser Vielzahl von Einzelmessungen das ursprüngliche Klangbild (für das menschliche Ohr ausreichend) präzise rekonstruiert werden.

2. BEISPIELE VON MESSVORGÄNGEN

2.1. Zählen

Das Zählen ist ein selbstverständlicher Bestandteil unserer Kultur. Tatsächlich verbergen sich hinter jedem Zählvorgang zwei durchaus nichttriviale Prozesse. *Jede Zählung erfordert eine äußere und eine innere Differenzierung*.

Mit ***äußerer Differenzierung*** ist die *Abgrenzung der zu zählenden Objekte von den nicht zu zählenden Objekten* gemeint. Wenn man die Anzahl der Personen in einem

Raum zählt, ist das meist nicht schwierig. Will man hingegen die Anzahl der Personen einer Stadt erfassen, so ist die Abgrenzung zwischen den zu zählenden und nicht zu zählenden Personen schon schwieriger. Zunächst muß räumlich abgegrenzt werden, wo der Stadtbereich genau liegt; weiters muß festgelegt werden, ob Ansässige oder Anwesende gezählt werden sollen; weiters muß der Zeitpunkt der Erhebung festgelegt werden; weiters muß für jeden Menschen festgelegt werden, ob die Zählkriterien auf diese Person zutreffen oder nicht. In der Statistik spricht man in diesem Zusammenhang von der ***räumlichen, zeitlichen und sachlichen Abgrenzung*** *einer* ***statistischen Masse***. Jede Zählung impliziert somit eine Klassifikation aller denkbaren Objekte in solche, die gezählt werden (also "dazu gehören") und solche, die nicht gezählt werden (also "nicht dazugehören"). Nach welchen Regeln festgelegt wird, welche Objekte überhaupt gezählt werden sollen, hat einen wesentlichen Einfluss auf das Zählergebnis! Z.B. macht es bei der Zählung der Zuschauerzahl einer Sportveranstaltung einen Unterschied, ob man die Zahl der verkauften Eintrittskarten oder die Zahl der tatsächlich anwesenden Personen ermittelt: es können Leute mit einer Eintrittskarte zu Hause geblieben sein oder aber auch Leute ohne eine (bezahlte) Eintrittskarte bei der Veranstaltung sein.

Mit ***innerer Differenzierung*** ist gemeint, dass man *die zu zählenden Objekte auch voneinander unterscheiden und identifizieren* können muß; nur so kann sichergestellt werden, dass jedes Objekt genau einmal gezählt wird. So zählt man z.B. einen Haufen Streichhölzer ganz intuitiv, indem man die bereits gezählten Streichhölzer zur Seite auf einen eigenen Haufen "bereits gezählt" legt. Damit wird sicher gestellt, dass jedes Streichholz genau ein mal gezählt wird. Besonders wichtig, aber auch schwierig ist diese innere Differenzierung dann, wenn abstrakte Dinge (wie etwa Geldbeträge) zu zählen sind: etwa bei der Erfassung volkswirtschaftlicher Vorgänge oder bei der Erfassung geldmäßig bewerteter Einheiten (etwa in Bilanzen) ist es von großer Bedeutung, sozusagen nicht denselben Euro mehrmals zu zählen.

Gezählt werden kann nur, was mathematisch gesprochen als *Menge*[2] vorliegt. In dieser Feststellung liegt ein interessanter Gegensatz zwischen dem Zählen als etwas Dynamischen und dem Mengenbegriff bzw. der "Anzahl" als etwas Statischem. Große Zählungen haben es an sich, dass bei Bekanntwerden der Ergebnisse die Zahlen meist nicht mehr aktuell sind, weil sich die gezählten Massen laufend verändern.

Ein besonderes Thema sind ***Mehrfachzählungen***. So hat etwa der Generalsekretär des Österreichischen Fußballbundes anlässlich der Vergabe der Fußball-EM 2008 an Österreich und die Schweiz frohlockt: „Man kann schon ins Schwärmen kommen, wenn man weiß, dass bei so einem Ereignis rund zehn Milliarden Menschen weltweit zusehen“[3]. Zehn Milliarden sind mehr als die gesamte Weltbevölkerung! Eine Posterin hat zu dieser Zählweise karikierend angemerkt, dass man die Zahl der weltweiten Zuseher ja gleich auf 20 Milliarden verdoppeln könne, indem man die Zuseher für jede Halbzeit extra zählt. In gewissen Fällen macht eine Mehrfachzählung eher Sinn: Wenn z. B. die Wiener U-Bahnen das tägliche Fahrgastaufkommen mit 3,5 Millionen Personen angeben, obwohl sich in Wien (inklusive Urlaubern und Pendlern) stets weniger als 2 Millionen Menschen aufhalten.

2.2. Wetterbeobachtungen

Wetterbeobachtungen für meteorologische und klimatologische Zwecke sind ein Musterbeispiel für systematisches naturwissenschaftliches Messen. Aus der ungeheuer komplexen Gesamtheit des weltweiten Wettergeschehens werden nach strengen Richtlinien eine vergleichsweise geringe Zahl quantitativer Daten erhoben. Dennoch werden derart viele Messwerte erfasst, dass zu ihrer Verarbeitung für Wettervorhersagezwecke heute die leistungsfähigsten Computer der Welt eingesetzt werden. Was und wie gemessen wird, hängt einerseits vom Zweck, dem die Messung dienen soll und andererseits von den damit verbundenen Kosten ab. Hinsichtlich der Zwecksetzung unterscheiden sich *Wetterbeobachtungen für synoptische Zwecke* (Wetterkarten für Wetterprognosen) wesentlich von *klimatologischen Wetterbeobachtungen*. Die synoptischen Wetterbeobachtungen erfolgen weltweit stets zum selben Zeitpunkt (synchron) rund um die Uhr im Dreistundenrhythmus um 0h, 3h, 6h usw. Greenwich Mean Time. Klimatologische Wetterbeobachtungen hingegen erfolgen dreimal täglich um 7h, 14h und 21h Ortszeit, die dem Sonnenstand (bzw. der geographischen Länge) entspricht[4]. Für beide Zwecke werden im wesentlichen dieselben Basisdaten erhoben: Luftdruck, Temperatur, Luftfeuchte, Windrichtung und Windgeschwindigkeit, Bewölkungsmenge, Wolkenarten, Sichtweite, Niederschlagsmenge und Wettererscheinungen (Regen, Nebel, Schneefall etc.). Damit diese Daten ohne Probleme weltweit ausgetauscht und analysiert werden können, wurden international einheitliche *Skalen* zur Messung der verschiedenen Wetterelemente vereinbart. Wir werden uns mit einigen dieser Messungen etwas ausführlicher beschäftigen, weil sich die an diesm Beispiel gewonnenen Erkenntnisse verallgemeinern lassen.
Zur ***Temperaturmessung*** ist die *Festlegung eines Nullpunktes und eines Einheitsschrittes* erforderlich: z. B. haben die Celsius- und die (heute ungebräuchliche) Reaumurskala zwar denselben Nullpunkt (beim Gefrierpunkt des Wassers), aber verschiedene Einheitsschritte (100°C = 80°R); während die in den USA noch immer verwendete Fahrenheit-Skala sowohl einen anderen Nullpunkt als auch eine andere Einheitsschrittweite als die Celsiusskala (0°C = 32°F; 100°C = 212°F) hat. *Unabhängig von der verwendeten Skala* sind u.a. folgende Aussagen über die (mit derselben Skala) gemessenen Temperaturen a, b, c und d sinnvoll möglich (a,b,c und d seien z.B. 40°,35°,30° und 20° einer festen, beliebigen Skala):

– a ist größer als b
– b liegt in der Mitte zwischen a und b
– die Differenz a–b ist doppelt so groß wie die Differenz c–d.

Alle diese Aussagen bleiben in ihrem Wahrheitsgehalt unverändert, wenn man die Temperaturen auf einer anderen Skala messen würde; ihr *Wahrheitsgehalt ist unabhängig von der Wahl der Skala*. Derartige Aussagen werden daher auch als *sinnvolle Aussagen* über Temperaturen angesehen.

Der Wahrheitsgehalt einer Aussage über Quotienten von zwei Temperaturen (etwa der Art "a ist doppelt so groß wie d") ändert sich jedoch, wenn man zu einer Skala mit einem anderen Nullpunkt übergeht. Genauso ändert sich der Wahrheitsgehalt einer Aussage der Art "a ist um 5 größer als b", wenn man zu einer Skala mit anderen Einheitsschritten übergeht. Solche Aussagen werden als *nicht sinnvoll* angesehen, weil sie nichts über die Temperaturen "an sich" aussagen, sondern lediglich etwas über zwei Messwerte auf einer ganz bestimmten Skala. Zu beachten ist, dass "sinnvolle Aussage" keineswegs dasselbe bedeutet wie "wahre Aussage". Auch falsche Aussagen können sinnvoll sein, wenn sie auch nach einer beliebigen zulässigen Skalentransformation falsch bleiben.

Bei der Messung des Luftdrucks sind ebenfalls verschiedene Skalen gebräuchlich. Das von Torricelli um 1640 konzipierte Quecksilberbarometer basiert auf der Idee, den Luftdruck durch die Höhe einer Quecksilbersäule, deren Gewichtsdruck gleich dem Luftdruck ist, zu messen. Dies ergibt einen mittleren Luftdruck von 760 mmHg (heute auch 760 Torr) in Meereshöhe bei $0^{o}C$ Temperatur und 45^{o} geographischer Breite. Heute ist die Druckeinheit Pascal = Netwon/m^2 bzw. 1 bar = 100.000 Pa internationale Norm. Der Nullpunkt liegt (im Gegensatz zu den Temperaturskalen) jedoch für alle Skalen zur Druckmessung gleich. Ein Druck hat genau dann den Wert null, wenn keine Kraft auf die betrachtete Fläche einwirkt. Die verschiedenen Skalen unterscheiden sich lediglich durch die Wahl der Einheit.

Welche Aussagen sind nun für Druckangaben sinnvoll? Eine kurze Überprüfung zeigt, dass alle Aussagen, die wir für Temperaturmessungen als sinnvoll angesehen haben, sinngemäß auch für Druckmessungen gemacht werden können. Darüber hinaus sind auch Aussagen über Quotienten von Druckmessungen, wie etwa

- Druck a ist doppelt so groß wie Druck d
- Druck c ist um 50% größer als Druck d

sinnvolle Aussagen, da sie bei (simultanen) Transformationen der Druckeinheiten ihren Wahrheitsgehalt nicht ändern. Aussagen der Art

- Druck a ist um 10 größer als Druck c

sind hingegen auch für Druckangaben nicht sinnvoll, da ihr Wahrheitsgehalt sich beim Übergang zu einer anderen Druckeinheit ändert.

Zum *sinnvollen Vergleich von Messungen des Luftdruckes* an verschiedenen Orten bedarf es noch einer Reihe von Korrekturen. Bei Verwendung eines Quecksilberbarometers wird der Druck aus der Höhe der Quecksilbersäule ermittelt. Da diese Höhe auch mit der Temperatur schwankt, ist der abgelesene Wert um den Temperatureinfluss zu korrigieren. Da der Luftdruck generell mit zunehmender Meereshöhe abnimmt, wird auch eine "Reduktion auf Meeresniveau" durchgeführt. Bei Orten verschiedener geographischer Breite muß schließlich auch noch der (fliehkraftbedingte) Unterschied in der Erdbeschleunigung berücksichtigt werden. Alle Druckangaben in Wetterkarten oder auch in den meteorologischen Aufzeichnungen der einzelnen Stationen sind bereits um die genannten Einflüsse korrigiert. Dieses Beispiel zeigt, dass oft erst eine *sachgerechte Weiterverarbeitung der Rohdaten* sinnvoll vergleichbare Zahlen liefert. Wir werden in diesem Zusammenhang auch von *direkten* und *abgeleiteten Messungen* sprechen. Während ***direkte*** *Messungen unmittelbar am Messobjekt* vorgenommen werden, (z.B. Ablesung am Barometer) entstehen ***abgeleitete*** *Messungen durch eine zweckorientierte Weiterverarbeitung bereits vorliegender Messergebnisse*. Sämtliche *statistischen Verfahren* können derart als *Verfahren zur Ableitung neuer Messergebnisse* aus bereits vorliegenden Daten verstanden werden.

Ein weiteres Beispiel für ein abgeleitetes Maß ist die *relative Änderung des Luftdrucks*. Dieses Maß kann entweder *quantitativ* durch die Angabe der Druckdifferenz zwischen der letzten und der vorletzten Messung (bei feste Zeitabstand zwischen zwei Messungen) oder *qualitativ* durch eine der Angaben "stark fallend", "fallend", "gleichbleibend", "steigend", "stark steigend" erfolgen. Für die qualitativen Angaben der Druckänderung ist zwar eine Ordnung (Reihenfolge) definiert, es gibt aber keine "Abstände" zwischen den einzelnen Kategorien.

Bei der *Luftfeuchte* gibt es zwei wichtige Maße. Die absolute Luftfeuchte entspricht dem Gewicht des Wasserdampfes je m^3 Luft, die relative Luftfeuchte gibt an, welcher Anteil des für die betreffende Temperatur maximalen Wasserdampfgehaltes (Sättigungsdampfmenge) tatsächlich erreicht wird. Das praktisch relevantere Maß ist die *relative Luftfeuchte*: Bei einer relativen Luftfeuchte von 100% (oder knapp darunter) beginnt diese zu kondensieren (Wolkenbildung). Man kann die relative Luftfeuchte entweder direkt (z.B durch ein Haarhygrometer) oder indirekt messen. Die indirekte Bestimmung basiert im wesentlichen auf der Temperaturmessung mit einem "normalen" und einem ständig feucht gehaltenen Thermometer in einem sogenannten Aspirationspsychrometer. Durch Einsetzen der abgelesenen Werte in Gleichungen oder durch Nachsehen in fertigen Tabellen (Psychrometertafeln) werden absolute bzw. relative Luftfeuchte rechnerisch ermittelt. Dieses Beispiel soll zeigen, dass in besonderen Fällen dieselbe Größe sowohl direkt als auch duch Ableitung aus anderen Messwerten ermittelt werden kann. Eine direkte oder eine abgeleitete Messung zu sein, ist also streng genommen nicht eine Eigenschaft des Messwertes, sondern des Messverfahrens. Sehr häufig können jedoch bestimmte Messwerte nur durch Ableitung aus anderen Messwerten ermittelt werden, sodass man dementsprechend dem Messwert selbst die Eigenschaft “abgeleitete Messung” zuschreibt.

Weiters ist die relative Luftfeuchte als *Quotient zweier gleichartiger Zahlen* (tatsächliche Dampfmenge/Sättigungsdampfmenge) eine *dimensionslose Zahl* zwischen 0 und 1 (bzw. 0% und 100%), ein sogenanntes *normiertes Maß*[5]. Die absolute Luftfeuchte hat ähnlich wie der Luftdruck einen fixen Nullpunkt und eine beliebig wählbare Einheit[6].

Die aktuellen *Wettererscheinungen* (Regen, Schneefall, Nebel, Gewitter usw.) werden durch direkte Beobachtung ermittelt und durch in standardisierten Kategorien festgehalten. Das Ergebnis der Messung ist hier nicht eine bestimmte Zahl, sondern die Zugehörigkeit zu einem bestimmten Typus (z.B. könnte bei einer bestimmten Wetterbeobachtung eine Wettererscheinung vom Typ "Nebel" festgestellt werden). Da hier mehr eine *qualitative Unterscheidung* als eine quantitative Erfassung zugrundeliegt, spricht man auch von *qualitativem Messen*. Auch die Bewölkungsart wird im wesentlichen in Form einer qualitativen Unterscheidung beurteilt.

2.3. Erfassung von Zeiträumen

Während sich die bisher diskutierten Messungen (mit Ausnahme der Druckänderung) alle auf einen *Zeitpunkt* bezogen haben, beziehen sich die Messungen von Sonnenscheindauer und Niederschlagsmenge jeweils auf einen *Zeitraum* (z.B. ein Tag oder ein Jahr). Für einen bestimmten Zeitpunkt läßt sich lediglich feststellen, ob die Sonne scheint oder nicht. Dieser Sachverhalt läßt sich mathematisch durch eine Funktion der Art

$f(t) = 0$: zum Zeitpunkt t scheint keine Sonne

$f(t) = 1$: zum Zeitpunkt t scheint die Sonne

darstellen[7]. Die Gesamtsonnenscheindauer GS im Zeitintervall von t_0 bis t_1 kann dann durch folgendes Integral dargestellt werden:

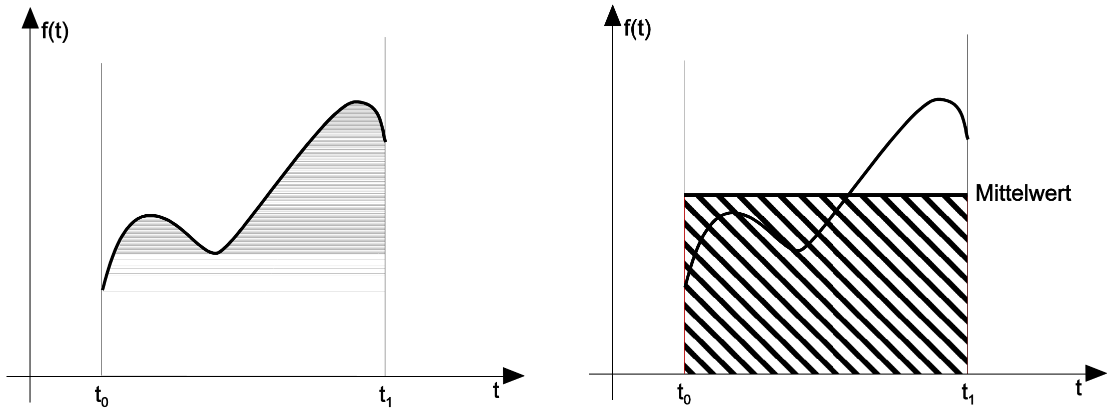
$$GS(t_0, t_1) = \int_{t_0}^{t_1} f(t)\, dt$$

Analog könnte die Gesamtniederschlagsmenge als Integral von Niederschlagsintensitäten, die zu jeden Zeitpunkt erfasst werden (gemessen z.B. in mm/h) aufgefasst werden. Die praktische Bestimmung der Niederschlagsmenge erfolgt jedoch meist so, dass der anfallende Niederschlag in einem geeigneten Gefäß über den gesamten Beobachtungszeitraum gesammelt und anschließend erst gemessen wird[8]. Die Sonnenscheindauer hingegen wird gemessen, indem die tatsächlichen Zeiten des Sonnenscheins mit einem Brennglas auf einem Papierstreifen eingebrannt werden. Die Gesamtlänge der geschwärzten Spur ergibt die Sonnenscheindauer des betreffenden Tages, was einer graphischen Auswertung des oben angeführten Integrals entspricht.

Wie lassen sich auch Temperatur, Luftdruck, Luftfeuchte auf ganze Zeiträume beziehen? Dazu gibt es zwei Möglichkeiten:

a) Permanente Aufzeichnung der betreffenden Messgröße auf ein Registriergerät (Thermograph, Barograph etc.).
b) Wiederholte Einzelmessungen.

Die erste Methode liefert zwar (im Rahmen der Aufzeichnungsgenauigkeit) ein vollständigeres Bild der zeitlichen Entwicklung; die zweite jedoch vielfach leichter weiterzuverarbeitende (weil nummerische) Daten. Allgemein spricht man bei Daten, die über einen ganzen Zeitraum erfasst werden, von (kontinuierlichen bzw. diskreten) *Zeitreihen*. Die graphische Aufzeichnung liefert eine Art Funktionsgraphen einer Funktion f, wobei der jeder Funktionswert f(t) einfach der Messwert zum Zeitpunkt t ist. Aus dem Graphen können durch weiterführende Analysen eine Vielzahl weiterer Kenngrößen wie Maximum, Minimum oder Mittelwert für den betrachteten Zeitraum ermittelt werden. Die Ermittlung des Mittelwertes basiert im wesentlichen auf der Idee, die Fläche unter der Wertekurve durch ein Rechteck gleicher Fläche zu ersetzen.



Die Höhe des Rechteckes entspricht dann dem Mittelwert MW, der definiert ist als

$$MW = \frac{\int_{t_0}^{t_1} f(t)\,dt}{t_1 - t_0}$$

Das Integral im Zähler entspricht dem punktierten Flächeninhalt unter der Kurve von t_0 bis t_1 (linke Abbildung oben). Für die praktische nummerische Auswertung eines derartigen Integrals wird man jedoch in der Regel auf eine Reihe von Einzelwerten zurückgreifen müssen und damit das Integral durch einen (gewichteten) Mittelwert näherungsweise bestimmen.

2.4. Mittelwerte

Mittelwerte sind ein einfaches und gängiges Beispiel für ein abgeleitetes Maß. Sie haben den Zweck, eine Reihe von gleichartigen Messwerten (z.B. Temperaturwerten eines Tages) durch eine einzige "typische" Kennzahl zu ersetzen. Wie bei vielen anderen abgeleiteten Maßen auch geht es um die Verdichtung von Information: anstelle einer Fülle verschiedener Einzelwerte möchte man einen "handfesten" Durchschnittswert. Aus der Fülle möglicher Zentralmaße besprechen wir hier nur das gewöhnliche sowie das gewogene arithmetische Mittel.

Der gewöhnliche Mittelwert $\bar{x}$ wird berechnet, indem man die betrachteten Messwerte x_i; i=1 ...n summiert und die Summe durch die Zahl der Werte (n) dividiert:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$

Dahinter kann man die Idee sehen, die Summe der einzelnen Messwerte auf die n Beobachtungspunkte gleichmäßig aufzuteilen. Auf diese Weise werden etwa in der Meteorologie Mittelwerte verschiedener Messdaten ermittelt. Bei der klimatologischen Temperaturmessung wird das Tagesmittel jedoch folgend berechnet:

$$t_{mittel} = \frac{t_1 + t_2 + 2t_3}{4}$$

t_1: Temperatur um 7 Uhr Ortszeit
t_2: Temperatur um 14 Uhr Ortszeit
t_3: Temperatur um 21 Uhr Ortszeit

Die Temperatur um 21 Uhr wird quasi zweimal erfasst, sodass im Nenner auch 4 statt 3 steht. Hinter dieser Vorgangsweise steht die Idee, mit dem 21 Uhr-Wert auch einen möglichst guten Ersatzwert für die fehlende Nachtmessung in die Mittelbildung einzubeziehen. Allgemein spricht man in solchen Fällen, bei denen die einzelnen Messwerte unterschiedliche "Gewichte" erhalten (t_1 und t_2 haben das Gewicht 1, t_3 das Gewicht 2) von einem *gewichteten Mittelwert*. Beim gewöhnlichen arithmetischen Mittel sind alle Gewichte gleich 1.

Man kann die Gewichte auch **normieren,** indem man jedes Gewicht durch die Summe der Gewichte dividiert. Beim obigen Temperaturmittel sind die normierten Gewichte 1/4, 1/4 und 1/2, bei ungewichteten Mittelwertbildungen von n Größen wäre jedes normierte Gewicht 1/n. Es gilt stets für die *normierten* Gewichte g_i:

$$\sum_{i=1}^{n} g_i = 1 \quad \text{und} \quad \bar{x} = \sum_{i=1}^{n} g_i x_i$$, wobei $\bar{x}$ ein gewogenes arithmetisches Mittel ist.

Beispiel: Die Temperatur betrage an einem Tag um 7h: 12°, 14h: 24°, 21h: 16°. Das klimatologische Tagesmittel beträgt (12+24+2*16)/4 = 17° bzw. nach der obigen Formel 12°/4 + 24°/4 + 16°/2 = 17°. Man multipliziert jeden zu mittelnden Wert mit seinem normierten Gewicht und summiert diese Produkte zum gewichteteten Mittelwert. In unserem Beispiel beträgt das gewichtete Temperaturmittel 17°, das ungewichtete Mittel beträgt (12+24+16)/3 = 17,33°.
Anschaulich wird das normierte Gesamtgewicht 1 in n gleiche bzw. verschieden große Teile (beim gewichteten Mittel) zerlegt und jedes Gewicht einem bestimmten Messwert zugewiesen. Der Mittelwert kann dann mathematisch als das Skalarprodukt des Gewichtsvektors **g** [im obigen Beispiel ist **g** = (0,25; 0,25; 0,5)] mit dem Wertevektor **x** [im Beispiel ist **x** = (12°, 24°, 16°)] aufgefasst werden. Damit ist klar, dass man für eine gegebene Zahl von Messwerten prinzipiell beliebig viele verschiedene gewogene Mittelwerte konstruieren kann: Jeder denkbaren Aufteilung der Größe 1 in n Teile entspricht ein gewogenes Mittel der n Messwerte. In der Praxis muß daher die Gewichtung der einzelnen Werte nach sachlichen Gesichtspunkten geschehen. Hinter dem gewöhnlichen arithmetischen Mittel steht etwa die Grundannahme, dass alle Messwerte als gleich (ge)wichtig in die Durchschnittsbildung eingehen sollen. Falls die zu mittelnden Werte ihrerseits Mittelwerte unterschiedlich großer Massen sind, dann sind die Mächtigkeit (Anzahl der Elemente) der betrachteten Massen geeignete Gewichte.
Beispiel: Familie Maier hat vier Kinder im Alter von 8, 9, 11 und 12 Jahren; Familie Müller zwei Kinder im Alter von 2 und 6 Jahren. Der Gesamtmittelwert aller sechs Kinder beträgt somit (2+6+8+9+11+12) / 6 = 8 Jahre. Wir wollen nun annehmen, dass wir von beiden Familien nur die Kinderzahl und deren Durschnittsalter innerhalb der

Familie wissen: die vier Kinder der Familie Maier sind im Schnitt zehn Jahre und die zwei Kinder der Familie Müller im Schnitt vier Jahre alt. Dann erhält man den korrekten Gesamtmittelwert, indem man jeden der beiden Durchschnittswerte mit der Anzahl der Kinder gewichtet: $(4 \cdot 10 + 2 \cdot 4)/6 = 8$ Jahre. Der einfache Mittelwert der beiden Mittelwerte in den Familien $(4 + 10) / 2 = 7$ Jahre liefert hier nicht das korrekte Ergebnis!

Das folgende Beispiel "Nutzwertanalyse" zeigt, wie in einer komplexen Entscheidungssituation gewichtete Bewertungen zu einem Gesamturteil verknüpft werden können.

2.5. Anwendung von gewogenen Mittelwerten: Nutzwertanalyse

Unter *Nutzwertanalyse* versteht man eine Gruppe von Verfahren, mit denen in einer komplexen Entscheidungssituation unter mehreren Alternativen (Möglichkeiten) die beste, d.h. die mit dem größten Nutzwert, ermittelt werden kann. Beispiele für derartige Entscheidungssituationen sind: Auswahl einer EDV-Anlage unter mehreren Angeboten, Wahl eines Geschäftspartners unter mehreren möglichen Kandidaten, Auswahl eines Bewerbers von mehreren Kandidaten bei einer Stellenausschreibung o.ä. Die Auswahloptionen (Kandidaten, Angebote), aus denen eine auszuwählen ist, nennt man Alternativen. Die *Grundidee der Nutzwertanalyse besteht darin, die Gesamtbeurteilung jeder Alternative aus einer Reihe von Einzelbewertungen zusammen zu setzen*. Methodisch wird folgend vorgegangen:

1) Erstellen eines *Anforderungsprofiles* (Gesamtheit aller Bewertungskriterien, nach denen jede Alternative bewertet werden soll).
2) Ermittlung eines Nutzwertes (Teilnutzens) für jede Alternative bezüglich jedes Kriteriums des Anforderungsprofiles. Dies ergibt für jede Alternative einen Bewertungsvektor, dessen Komponenten genau die Bewertungen gemäß den einzelnen Kriterien des Anforderungsprofiles darstellen.
3) Gewichtung jedes Kriteriums des Anforderungsprofiles. Die Summe der Gewichte muss 1 (=100%) betragen.
4) Für jede Alternative wird der gewichtete Mittelwert aller Bewertungen errechnet Auswahl der Alternative mit dem höchsten gewichteten Mittelwert.

Die einzelnen Varianten der Nutzwertanalyse unterscheiden sich primär in der konkreten Durchführung von Schritt 3). Eine einfache Möglichkeit besteht darin, einen *Gesamtnutzen* als *gewichtetes arithmetisches Mittel der einzelnen Teilnutzen* zu ermitteln. Die Gewichte repräsentieren dabei die Wichtigkeit der einzelnen Kriterien im Anforderungsprofil. Diese Vorgangsweise wird im folgenden Beispiel beschrieben und kritisch diskutiert.

Beispiel: Für den Kauf einer EDV-Anlage liegen die Angebote von acht Bietern A - H vor. Mittels einer Nutzwertanalyse soll die Entscheidung, welcher Bieter den Zuschlag erhält, unterstützt werden. Folgende Vorgangsweise wurde gewählt:

1) Es wird ein Anforderungsprofil, bestehend aus einer Reihe von Bewertungskriterien, aufgestellt. Dazu wird oft stufenweise vorgegangen: Zunächst werden Hauptkriterien (Kriteriengruppen) festgelegt. Danach werden in jeder Gruppe eine Reihe von relevanten Einzelkriterien, nach denen tatsächlich beurteilt werden kann, gebildet. Durch diese Vorgangsweise wird vor allem das Gewichten der einzelnen Kriterien (siehe nächster Punkt) erleichtert.
2) Die einzelnen Kriterien werden entsprechend ihrem Anteil am Gesamtnutzen innerhalb des Anforderungsprofiles derart gewichtet, dass die Summe der Gewichte 1 ergibt. Zweckmäßigerweise teilt man dazu das verfügbare Gesamtgewicht zunächst nur auf die Hauptkriterien auf (Level 0). Anschließend gewichtet man jedes Einzelkriterium innerhalb der Gruppe derart, dass die Summe dieser Gewichte wiederum 1 beträgt (Level 1). Das Gewicht jedes Einzelkriteriums innerhalb des gesamten Anforderungsprofiles ist dann das Produkt aus dem Gruppengewicht (Level 0) mit dem Gewicht innerhalb der Gruppe (Level 1). Die Summe der so erhaltenen Gewichte über alle Einzelkriterien ist damit wiederum 1. Insgesamt ergibt sich eine *Gewichtungstabelle* für das gesamte Anforderungsprofil (siehe nächste Seite).
3) Nun wird jede Alternative bezüglich jedes Kriteriums auf einer Punkteskala (hier von 0 - 10) bewertet. Dies erfordert in vielen Fällen eine subjektive Einschätzung, bei mehreren an der Entscheidung Beteiligten oft auch eine intensive Diskussion. Ein Ausschnitt aus der Bewertungstabelle ist auf Seite 17 abgebildet.
4) Für jede Alternative wird das gewichtete arithmetische Mittel der Punktewerte entlang des gesamten Leistungsprofiles als Gesamtnutzwert errechnet; die Präferenzordnung unter den Alternativen ergibt sich aus der Rangordnung der Gesamtnutzwerte.

Das Prinzip der Nutzwertanalyse besteht somit darin, eine komplexe Gesamtbewertung in eine Reihe kleinerer Einzelbewertungen zu zerlegen und aus diesen die Gesamtbewertung nach einer bestimmten Rechenregel abzuleiten. Diese Vorgangsweise ist bei der Erfassung komplexer Sachverhalte durchaus üblich. Dabei bleibt der Gesamturteilsaufwand hoch, da eine Vielzahl von Einzelbewertungen durchzuführen ist. Lediglich die Komplexität jedes einzelnen Urteils ist wesentlich geringer als im Falle einer unmittelbaren Gesamtbeurteilung.

Hinter der Vorgangsweise des Aufspaltens in Einzelurteile und anschließender formaler Aggregation steht das Prinzip "Das Ganze ist die Summe seiner Teile". Konkreter sind dies folgende Voraussetzungen bzw. Modellannahmen :
1) Die einzelnen *Bewertungskriterien sind unabhängig voneinander erfassbar und messbar*.
2) Die einzelnen Kriterien *ergänzen einander zu einer vollständigen Beschreibung des Gesamtsystems*.

Die *Unabhängigkeitsannahme* ist eng mit der Ermittlung der Teilnutzen und ihrer Weiterverarbeitung durch die mathematischen Operation des Addierens zu einem Gesamtnutzen verbunden. Jeder Teilnutzen wird unabhängig von den übrigen Teilnutzen erfasst; weiters werden Zusammenhänge, die es zwischen den einzelnen Kriterien gibt, nicht erfasst. Auch bei der Addition der gewichteten Teilnutzen zu einem Gesamtnutzen fließen keinerlei Zusammenhänge zwischen den einzelnen Bewertungen ein. Bei einem additiv ermittelten Gesamtnutzen wäre es beispielsweise denkbar, dass eine EDV-Anlage mit erstklassiger Ausstattung den höchsten Gesamtnutzen aufweist, obwohl sie gar keinen Netzanschluss besitzt. Oder es könnten für eine Anlage eine Reihe von Erweiterungsmöglichkeiten positiv in die Ermittlung des Gesamtnutzens eingehen, die jedoch nicht alle zugleich realisiert werden können (etwa wegen Platzbeschränkungen in der Hardware).

Die *Vollständigkeitsannahme* steht hinter der Vorgangsweise, die *Summe der gewichteten Teilnutzen bereits als Gesamtnutzen* zu betrachten. Schon das extreme Beispiel des vergessenen Netzschalters zeigt, dass eine *vollständige Erfasssung sämtlicher relevanter Punkte praktisch kaum realisierbar* und auch wenig zielführend ist, weil eine additive Verknüpfung sehr vieler kleiner Teilaspekte erst recht dazu führen kann, dass eine wenig akzeptable Alternative den höchsten Gesamtnutzen aufweist. Eine Nutzwertanalyse kann niemals gewährleisten, dass völlig untaugliche Alternativen auch sicher ausgeschieden werden. Daher werden in der Praxis nur mehr diejenigen Alternativen in die Analyse mit einbezogen, die tatsächlich als auszuwählende Alternative in Frage kommen.

Darüber hinaus basiert jede Nutzwertanalyse bereits von ihrem Ansatz her auf folgenden Grundannahmen:

a) **Konstanz der Alternativen**: Es muß bereits vor der Analyse feststehen, zwischen welchen Alternativen zu entscheiden ist. Kompromisse o.ä. oder eine Erweiterung der Wahlmöglichkeiten sind nicht vorgesehen.
b) **Konstanz der Kriterien**: Es müssen sämtliche Kriterien für alle Alternativen in gleicher Gewichtung angewendet werden (können). Dies ist in der Praxis oft sehr schwierig oder nur willkürlich möglich, da es oft von der Gesamtkonstellation einer Alternative abhängt, wie stark ein bestimmtes Kriterium zu werten ist oder welche Kriterien überhaupt wichtig sind.

Worin liegt nun der *Vorteil einer Nutzwertanalyse*? Ich sehe den entscheidenden Vorteil weniger darin, dass durch ein derartiges Verfahren Entscheidungen automatisiert werden können, sondern im Gegenteil eher darin, dass die kompetente Anwendung eines derartigen Verfahrens zu *Reflexion und Diskussion* zwingt. Bei der Erstellung der Kriterien und deren Gewichtung muß reflektiert (und gegebenenfalls auch diskutiert) werden, was man eigentlich will. Diskussionen bei der Ermittlung der einzelnen Teilnutzen tragen zu einer kritischen Auseinandersetzung mit den Charakteristika der einzelnen Alternativen wesentlich bei.

Ein weiterer Vorteil besteht in der *Transparenz des Verfahrens*. Der Einfluss jeder einzelnen Gewichtung und Bewertung auf den Gesamtnutzen ist leicht erfassbar. Dies hilft auch, die in dem Verfahren implizierte Willkür (zumindest in Teilbereichen) transparent zu machen.

3. SKALENNIVEAUS

Dieses Kapitel soll eine Übersicht über die wichtigsten beim Messen und in der Statistik vorkommenden **Skalenniveaus** bieten. Dabei umfasst ein Skalenniveau im wesentlichen die Menge aller Skalierungen, die durch eine bestimmte Klasse zulässiger Skalentransformationen ineinander übergeführt werden können. Damit ist die *Angabe der zulässigen Skalentransformationen das entscheidende Charakteristikum bei der Definition bzw. Festlegung eines Skalenniveaus*. In der Praxis ist bei der Ermittlung des Skalenniveaus zu überlegen, welche Skalentransformationen zulässig sind. Das Skalenniveau ist dabei keine Eigenschaft des zu messenden Merkmals, sondern eine des Messverfahrens. Beispielsweise kann der Luftdruck sowohl quantitativ als auch qualitativ (etwa in der Art "hoch", "normal", "tief") gemessen werden.

Beispiele für verschiedene Messungen und ihnen zugrundeliegende Skalenniveaus bzw. Skalentransformationen sind:

– Bei der Temperaturmessung ist eine Transformation des Nullpunktes und der Einheitsschrittweite zulässig. Dies entspricht einer Skalentransformation des Typs
  $f(x) = rx + d$ mit
  x : Wert auf der ursprünglichen Skala
  $f(x)$: Wert auf der transformierten Skala
  $r > 0$: Transformationsfaktor der Einheit
  d : Verschiebung des Nullpunktes um den Wert d.

– Bei der Messung von Drücken ist nur eine Transformation der Einheit zulässig. Alle zulässigen Skalentransformationen sind vom Typ $f(x) = rx$ mit x, $f(x)$ und r wie oben.

- Bei der Angabe der Anzahl der Gewitter oder der Anzahl der Tage mit Temperaturen unter 0° (Frosttage) innerhalb eines Jahres für einen bestimmten Ort ist keine Datentransformation zulässig. Jede Anzahl ist eine feste, nicht transformierbare Zahl.

– Bei der Angabe von Wettererscheinungen (Nebel, Hagel etc.) ist eine beliebige "Skalentransformation" zulässig. Jede Wettererscheinung kann durch irgendeine beliebige Zahl codiert werden, solange verschiedene Erscheinungen verschieden verschlüsselt werden.

– Bei der Angabe von Windgeschwindigkeiten oder Erdbebenstärken gibt es jeweils mehrere gebräuchliche Skalen. Ihre Gemeinsamkeit besteht darin, dass beim Übergang von einer Skala zu einer anderen die **(AN-)Ordnung** der Daten erhalten bleibt: wenn für zwei beliebige Werte auf einer Skala $a < b$ gilt, so muß auch für jede transformierte Skala $f(a) < f(b)$ gelten.

Die folgende Tabelle bietet einen Überblick über die wichtigsten Skalenniveaus und die ihnen entsprechenden zulässigen Skalentransformationen:

Bezeichnung des Skalenniveaus		Zulässige Skalentransformat.	Beispiele	mögliche sinnvolle Aussagen
quantitative (metrische, kardinale) Skalen	Absolutskala	keine	Anzahlen	$a = 10$ (jede denkbare Aussage ist sinnvoll)
	Verhältnisskala (Rationalskala)	$f(x) = r \cdot x \quad r > 0$ (jede homogen-lineare Transf.)	Länge, Masse, Druck, Zeitdauer	$a = 2b$ (beliebige Quotienten von zwei Messwerten sind sinnvoll)
	Intervallskala	$f(x) = r \cdot x + d;\ r > 0$ (affin-lineare Transformationen)	Temperatur, kalendarische Zeitmessung	$a - b = 2(c - d)$ (Differenzen, arithm. Mittelwerte, Streuungen sinnvoll)
qualitative Skalen	Ordinalskala (Rangskala)	$a > b \rightarrow f(a) > f(b)$ (jede monotone Transformation)	Härte, Helligkeit, Erdbebenstärken, Beliebtheitsskalen	$a > b$, Vergleich von Medianen und Quartilsabständen
	Nominalskala	$a = b \leftrightarrow f(a) = f(b)$ jede bijektive Transformation	Schultyp, Geschlecht, Sprachkenntnisse, Automarke	$a = b$, Modus als einziges Zentralmass sinnvoll

Die hier aufgeführten Skalenniveaus sind *hierarchisch geordnet*: jedes Skalenniveau besitzt auch alle Eigenschaften der darunterliegenden "niedrigeren" Niveaus. So ist jede Intervallskala auch eine Ordinal- und eine Nominalskala, weil jede affin lineare Transformation auch monoton und bijektiv ist. Die Mengen der zulässigen Skalentransformationen sind also jeweils echte Teilmengen der zulässigen Skalentransformationen der darunterliegenden Niveaus.

Neben den hier angeführten Skalenniveaus gibt es noch weitere, die durch andere Klassen zulässiger Transformationen charakterisiert sind. Diese lassen sich jedoch nur mehr zum Teil in die oben skizzierte Hierarchie einordnen.

4. SINNVOLLE UND SINNLOSE AUSSAGEN ÜBER MESSGRÖSSEN

In der Messtheorie unterscheidet man zwischen "sinnvollen" und "sinnlosen" Aussagen über Messungen in folgender Weise: eine Aussage ist **sinnvoll**, wenn sie unter allen zulässigen Skalentransformationen ihren Wahrheitsgehalt nicht ändert. Eine falsche Aussage, die unter allen zulässigen Transformationen falsch bleibt, ist genauso sinnvoll wie eine wahre Aussage, die ebenfalls für alle zulässig transformierten Skalen wahr bleibt. Durch dieses Konzept werden Aussagen, die nur aufgrund einer speziellen Skalenwahl wahr sind und bei einer zulässigen Skalentransformation zu falschen Aussagen werden, als **nicht sinnvoll** ausgeschieden. Voraussetzung dafür ist allerdings, dass man die Menge der zulässigen Skalentransformationen - und damit das

Skalienniveau - kennt. Dies ist aber in der Praxis oft nicht einfach. Welche Skalentransformationen sind z.B. für die schulische Notenskala "zulässig"? Man kann hier sehr unterschiedlicher Auffassung sein. Wer in Noten lediglich Bezeichnungen für eine geordnete Abfolge von Leistungsstufen sieht, wird jede monotone (ordnungserhaltende = ordinale) Skalentransformation als zulässig betrachte, weil es ihm ja prinzipiell egal ist, welche Zahlenwerte die einzelnen Notenstufen erhalten, solange nicht die Reihenfolge durcheinander gebracht wird. Jemand anders könnte sich wiederum darauf berufen, das die Notenstufen 1 - 5 gesetzlich festgelegt sind und daher prinzipiell keine Skalentransformation zulässig ist, was dann heißen würde, dass Schulnoten absolut skaliert sind! Auch dazwischenliegende Auffassungen wären denkbar. Es hat sehr gravierende Konsequenzen, ob man nun Schulnoten als ordinal oder als absolutskaliert (oder auch nur als intervallskaliert) ansieht. Beispielsweise ist die Bildung von arithmetischen Mittelwerten auf Ordinalskalenniveau sinnlos, auf Absolutskalenniveau keineswegs. Das mess theoretische Konzept sinnhafter und sinnloser Aussagen basiert somit ganz wesentlich darauf, dass darüber Einigkeit herrscht, welche Skalentransformationen zulässig sind. Bei vielen Messungen ist es aber schwierig, überhaupt ein Skalenniveau anzugeben. Welches Skalenniveau besitzt etwa die Dioptrienzahl, mit der Fehlsichtigkeit gemessen wird oder ein Aktienindex oder der Intelligenzquotient?

Neben der mess theoretischen Konzeption gibt es noch eine andere Möglichkeit, sinnvolle und sinnlose Aussagen zu charakterisieren. Man kann eine *Aussage über Messungen auch dann als sinnvoll ansehen, wenn sie auf* ***plausiblen Operationen*** basiert. Eine Aussage der Art "a ist kleiner als b" ist etwa dann sinnvoll, wenn es plausibel ist, a und b der Größe nach zu vergleichen. Eine Aussage über Mittelwerte kann nur dann sinnvoll sein, wenn auch die Bildung der betreffenden Mittelwerte sachlich plausibel ist. Eine solche plausible Deutung des Mittelwertes könnte etwa darin bestehen, dass man sich die einzelnen Werte additiv zusammengefass t und anschließend gleichmäßig aufgeteilt denken kann. So könnte etwa die Mittelung mehrerer Einkommen (z.B. aller Mitarbeiter einer Abteilung) plausibel gemacht werden; die Mittelung von Temperaturen hingegen schon weniger. Bei der Temperaturmittelung könnte wiederum der Wunsch nach einer Minimierung der Abweichungsquadrate vom gewünschten Zentralwert ein plausibles Argument für das Bilden des arithmetischen Mittels sein. Diese Beispiele zeigen, dass eine formale Operation (wie das Vergleichen, das Mitteln) dann plausibel ist, wenn Strukturen, die der Operation eigen sind (wie etwa die Ausgleichseigenschaft oder die Minimalitätseigenschaft des Mittels) auch im Sachkontext (zumindest teilweise) vorhanden sind. Plausibel ist eine formale Operation dann, wenn ihre Struktureigenschaften mit Struktureigenschaften des Sachproblems zusammenhängt. Bei der Mittelbildung von Schulnoten ist es sehr schwierig, die formale Operation auf diese Weise plausibel zu machen.

Diese auf Plausibilität fußende Auffassung davon, was sinnvolle Aussagen sind, schließt die erstere mit ein. Falls eine Aussage ihren Wahrheitsgehalt unter zulässigen Skalentransformationen ändert, so ist dies auch ein gravierendes plausibles Gegenargument gegen ihre Sinnhaftigkeit. Umgekehrt aber ist die Invarianz des Wahrheitsgehaltes unter allen zulässigen Skalentransformationen nicht das einzige Kriterium für Sinnhaftigkeit im Sinne von Plausibilität.

5. GENAUIGKEIT VON MESSUNGEN

Dass Messwerte ungenau sein können, ist eine bekannte Tatsache. Dabei unterscheidet sich die Art der möglichen Messfehler bei direkten bzw. indirekten Messungen deutlich.

5.1 Genauigkeit von direkten Messungen

Bei direkten Messungen kann es Mängel im Messverfahren selbst oder auch bei der Durchführung eines (an sich exakten) Messverfahrens geben. Ein Beispiel für ein mangelhaftes Messverfahren ist ein Fragebogen mit zweideutig formulierten Fragen; ein Beispiel für eine mangelhafte Durchführung wäre, wenn man mit einem Fragebogen z.T. andere Personen befragt als die, über die man eigentlich Auskunft möchte.[9] Auch bei einer simplen Längenmessung mittels Maßband kann entweder das Maßband selbst ungenau skaliert sein oder es können auch Fehler beim Anwenden auftreten, indem z. B. das Maßband ungenau angelegt oder schlecht abgelesen wurde.

5.2 Genauigkeit von abgeleiteten Messungen

Bei abgeleiteten Messungen ist das Messverfahren im Allgemeinen durch eine exakte Berechnungsvorschrift gegeben: z. B. der Durchschnittverbrauch pro 100 km eines Autos ergibt sich aus dem Treibstoffverbrauch dividiert durch die gefahrenen Kilometer mal 100. Ungenauigkeiten ergeben sich bei abgeleiteten Messungen aus Ungenauigkeiten der Ausgangsdaten. Dabei kann man zumeist relativ einfach abschätzen, wie sich ein Fehler einer Eingabegröße auf die Ergebnisgröße auswirkt: man nimmt von der fehlerbehafteten Größe einen oberen und unteren Grenzwert an und ermittelt dann (unter der Annahme, alle anderen Eingabewerte wären exakt gegeben) die Variabilität der Ergebnisgröße. Für die Fehlerfortpflanzung besonders kritisch sind dabei Fälle, bei denen man durch eine nahe bei Null liegende fehlerbehaftete Zahl dividieren muß: Wenn z.B. (bei Zähler = 1) die Zahl im Nenner an der dritten Stelle hinter dem Komma falsch ist und 0,011 statt richtig 0,01 lautet, dann ergibt sich als Ergebnis 1/0,011 die Zahl 90,90... statt der Zahl 1/0,01 = 100. Der Fehler liegt im Ergebnis an der zweiten Stelle vor dem Komma, obwohl er bei den Ausgangsdaten nur in der Größenordnung 1/1000 war!

5.3 Messfehler bei technischen Messungen

Bei technischen Messungen ist die Analyse von Messfehlern durchaus üblich. Einerseits versucht man, die Größenordnung der möglichen Fehler zu ermitteln, andererseits wird durch Fehlerausgleichverfahren versucht, rechnerisch die

Gesamtgenauigkeit der Messung zu erhöhen, indem man etwa dieselbe Messung mehrfach durchführt und die verschiedenen Messwerte mittelt. Vielfach sind auch auf sachlogischen Zusammenhängen beruhende Fehlerausgleichverfahren üblich: Wenn z.B. in der Geodäsie ein geschlossener Polygonzug ermittelt wird, dann muß man Ende des Polygonzuges rechnerisch wieder exakt am Ausgangspunkt angekommen sein. Auf diese Weise erhält man auch eine präzise Auskunft über den Gesamtfehler, der sich im Zuge des Verfahrens ergeben hat.
Begründer der Fehlerrechnung ist Carl Friedrich Gauß, der im 19. Jahrhundert u.a. die Grundlagen für die moderne Geodäsie schuf. Im 20. Jahrhundert wurden für eine Reihe physikalischer Grundgrößen (Länge, Zeit, Lichtgeschwindigkeit) unglaublich genaue Messverfahren entwickelt, die eine Größe mit bis zu acht (oder gar noch mehr) gültigen Stellen liefern. Mit üblichen direkten Messverfahren erreicht man etwa eine Genauigkeit von drei bis vier, höchstens fünf gültigen Stellen: Wenn man z.B. eine Masse von 1 kg auf Gramm oder eine Länge von einem Meter auf mm genau abmisst, dann sind das drei gültige Stellen. Eine Waage, die auf fünf Stellen genau wägt, ist technisch kaum konstruierbar: Eine solche Waage müsste etwa einen PKW so genau abwägen, dass Hineinlegen eines einfachen Briefes in das Auto die Anzeige verändert. Am genauesten sind heute die Möglichkeiten der Zeitmessung mit Hilfe von höchstfrequent schwingenden Quarzen oder gar von Atomuhren.

5.4 Messfehler bei wirtschaftlichen Messungen

Die Genauigkeit wirtschaftlicher Messungen ist in der Regel viel geringer als die von technischen Messungen. Einen ausführlichen Beleg für diesen Befund liefert der Nobelpreisträgers Oskar MORGENSTERN in seinem Klassiker "Über die Genauigkeit von wirtschaftlichen Messungen." MORGENSTERN ist wagte es, die Genauigkeit etwa von Wirtschaftsstatistiken zu hinterfragen, indem er z.B. die Kohleexporte von England nach Frankreich mit den Kohlenimporten von Frankreich aus Englang (was theoretisch in einem Jahr dieselbe Größe sein sollte!) zu vergleichen. Die Ergebnisse waren ernüchternd: Abweichungen in der Größenordnung von ±10% stufte MORGENSTERN als "sehr gut", in der Größenordnung von 10-25% als "durchschnittlich" und erst Abweichungen von mehr als einem Viertel als "hoch" ein!

6. LITERATUR

Zur Messtheorie gibt es derzeit noch kaum deutschsprachige Literatur. Daher habe ich die Literaturliste eher kurz gehalten.

BRAUN, Günther (1982): Der Beitrag der Nutzwertanalyse zur Handhabung eines multidimensionalen Zielsystems. In: WiSt Wirtschaftswissenschaftliches Studium, Jg. 11, Heft 2, S. 49 - 54.

GROB, Heinz Lothar (1985): Fallstudie zur Nutzwertanalyse. In: WiSt Wirtschaftswissenschaftliches Studium, Jg. 14, Heft 3, S.150-153.

MORGENSTERN, Oskar (1965): Über die Genauigkeit von wirtschaftlichen Messungen. Physica Verlag, Wien.

PIETSCHMANN, Herbert (1980): Das Ende des naturwissenschaftlichen Zeitalters. Ullstein TB, Stuttgart.

PIRSIG, Robert (1976): Zen und die Kunst, ein Motorrad zu warten. Fischer TB, Frankfurt/Main.

ROBERTS, Fred (1979): Measurement Theory. Addison Wesley, New York.

ZANGEMEISTER, C. (1976): Nutzwertanalyse in der Systemtechnik, München 1976.

ANMERKUNGEN

1. "Qualität" ist hier in einem umfassenden Sinn gemeint, wie er etwa von PIRSIG (1980) verwendet wird.

2. im Sinne der naiven Definition CANTOR's: "Eine Menge ist eine Zusammenfassung bestimmter, wohlunterschiedener Objekte unserer Anschauung oder unseres Denkens zu einem Ganzen. Diese Objekte heißen auch Elemente der Menge." Die äußere Differenzierung wird hier mit diem Beifügung "bestimmter", die innere Differenzierung durch das Adjektiv "wohlunterschiedener" angesprochen.

3. Quelle: http://derstandard.at/?id=1160891

4. 14 h Ortszeit z.B. bedeutet genau 2 Stunden nach dem (mittleren) Sonnenhöchststand des betreffenden geographischen Längenkreises.

5. Weitere Beispiele für normierte Maße sind etwa die Maßzahlen zur Himmelsbedeckung oder Wahrscheinlichkeiten zur Messung von Unsicherheit.

6. International üblich ist die Einheit kg/m^3.

7.Wir gehen hier der Einfachheit halber von einer “diskreten” Sonnenscheinfunktion aus. Man könnte auch eine kontinuierliche Funktion verwenden, indem man etwa die Sonneneinstrahlung pro m2 und Sekunde in Watt misst.

8. Es gibt aber auch Geräte, mit denen die laufende Niederschlagstätigkeit aufgezeichnet wird, indem z.B. das Ansteigen des Niederschlags im Gefäß registriert wird. Solche Niederschlagsmesser sind jedoch recht kompliziert, da Schnee und Hagel erst verfüssigt werden müssen und gleichzeitig keine hohen Verdunstungsverluste entstehen dürfen. - Da der Niederschlag zu den wichtigsten Wetterfaktoren zählt, (Landwirtschaft, Regulierung von Stauseen, Überschwemmungskatastrophen) gibt es neben den Beobachtungen in Klimastationen noch eine dichtes Netz automatischer Niederschlagsmess geräte.

9. Dies kommt in der Praxis häufiger vor, als man zunächst meinen möchte. So will z.B. ein Psychologe etwas über die Intelligenz von Menschen (ganz allgemein) erfahren und führt dazu eine Untersuchung unter Studierenden der Psychologie durch. Dies liefert vielleicht Ergebnisse über die Intelligenz von Psychologiestudenten, aber nicht von Menschen im Allgemeinen. (Psychologiestudenten sind vermutlichkeine besonders typische Gruppe von Menschen, was Intelligenz angeht.)