

Aspekte der angewandten Statistik bei der Aufstellung und Überprüfung von Jahresabschlüssen

Von Dr. W. Gaul, Institut für Angewandte Mathematik der Universität Bonn,
und Dr. K. J. Schmitz, Bundesamt für Finanzen, Bonn-Bad Godesberg

I. Einleitung

In letzter Zeit sind verstärkt mathematisch-statistische Verfahren zur Rationalisierung bei der Inventur und bei Jahresabschlußprüfungen ins Gespräch gekommen. Insbesondere werden Methoden der mathematischen Stichprobenanalyse vorgeschlagen (siehe AWV-Fachinformation (1978), HOVERMANN (1976), NIES (1975), (1976), STROBEL (1973), WYSOCKI (1975)).

Dagegen ist - trotz der bei vorstehenden statistischen Verfahren unverkennbar großen Vorzüge - eine gewisse Zurückhaltung bzgl. ihrer seit dem 1. 1. 1977 gesetzlich zulässigen Anwendung zu beobachten. Gründe dafür scheinen u. a. in dem Einwand zu liegen, daß hier neue Freiheitsgrade bei der Aufstellung des Inventars, der Bilanz und bei der Ermittlung der Besteuerungsgrundlagen eingeräumt werden, die eingehender Diskussion bedürfen.

Mit dieser Thematik befaßt sich die vorliegende Arbeit, wobei im Folgenden insbesondere die Sichtweise derer berücksichtigt wird, die unter Benutzung von Stichprobenverfahren erhaltene Daten zu überprüfen haben.

II. Problemdiskussion

1. Anmerkungen zu Schätzverfahren

Wird z. B. ein Lagerbestand nicht voll erfaßt, sondern nur aufgrund einer oder mehrerer (evtl. geschichteter) Stichproben „hochgerechnet“, so kann unabsichtlich - aber auch absichtlich - der tatsächliche Lagerbestand „falsch geschätzt“ werden. Absichtliche Fehlschätzungen sind zudem schwer nachweisbar, da ja jedes statistische Verfahren, hier also das gewählte Stichprobenverfahren, mit einem zulässigen Restfehler behaftet ist. Somit besteht eine gewisse Unsicherheitssituation, in der fast jeder Geprüfte behaupten könnte - falls bei einer Überprüfung festgestellt wird, daß er seinen Lagerbestand „falsch geschätzt“ hat -, er gehöre zu dem (kleinen) Anteil der-

jenigen, wo (große) Abweichungen zwischen Ist-Bestand und geschätztem Lagerbestand bei Anwendung des entsprechenden statistischen Verfahren eben „zufällig“ eingetreten seien.

Läßt man also statistische Verfahren etwa zur Ermittlung von Besteuerungsgrundlagen zu, so sollte auch ein geeignetes Instrumentarium zur Verfügung stehen, mit dem die Einlassungen des Überprüften gegebenenfalls ohne großen Aufwand bewertet werden können. Wendet man lediglich wiederum ein Schätzverfahren an und vergleicht einen vorgegebenen Schätzwert x mit einem selbst erhobenen Schätzwert y , so wird, da ja $x = y$ fast nie gelten wird, die Bewertung des Absolutbetrages der Differenz $|y - x|$ zum Prüfkriterium. Ist für den vorliegenden Lagerbestand z. B. eine Normalverteilungsannahme naheliegend, so erweist sich diese Situation als problemlos. Wie aber nachfolgend motiviert wird, treffen solche Annahmen i. a. nicht zu. Die Bereitstellung benötigter Information, die eine spezielle Verteilungsannahme rechtfertigen würde, ist in der Regel aufwendig.

2. Verteilungsproblematik

Die der Normalverteilung eigenen Charakteristika sind für Lagerbestände i. a. nicht immer gegeben. Als Beispiel einfachster Art sei der Lagerbestand der 600 Artikel umfassenden Schicht I aus NIES (1975, S. 78) betrachtet. In Abb. 1 (s. nächste Seite linke Spalte oben) wurde dessen tatsächliche Verteilung und diejenige Normalverteilung eingezeichnet, für welche Mittelwert und Standardabweichung mit dem Mittelwert $\mu = 47,15$ und der Standardabweichung $\sigma = 8,62$ des Lagerbestandes übereinstimmt.

Ist hier die Angabe der tatsächlich vorliegenden Verteilungsfunktion - die Treppenfunktion der Abb. 1 - wegen der Einfachheit des gewählten Beispiels noch unproblematisch, können entsprechende Untersuchungen bei umfangreicher gelagerter Fakten - man siehe die AWV-Fachinformation (1978, S. 13 u. Anhang) für die Über-

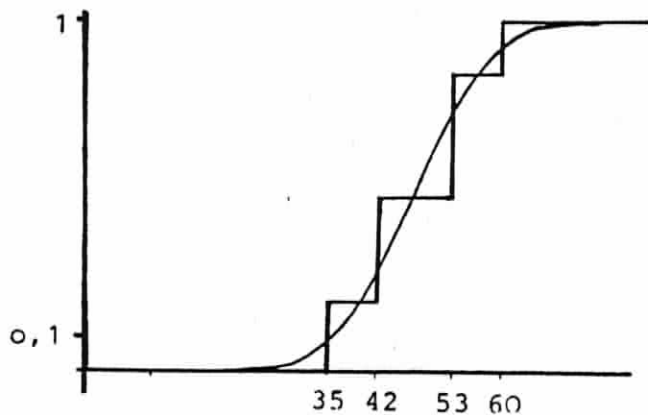


Abb. 1

prüfung der Annahme, ob im dortigen Fall die Unterstellung einer logarithmischen Normalverteilung gerechtfertigt ist – rasch zu erheblichem Arbeitsaufwand führen. Bei einer Routine-Überprüfung, der nur begrenzt Zeit zur Verfügung steht, kann dies aber dem Prüfer nicht zugemutet werden. Dem zu Überprüfenden steht es natürlich frei, durch Bereitstellung umfassenderer Informationen eine auf seine spezielle Situation zugeschnittene Prüfungsmethode zu ermöglichen. Solche zusätzliche Informationsbeschaffung schmälert jedoch den eingangs erwähnten Rationalisierungseffekt.

Somit ist nach vorstehenden Ausführungen der Einsatz statistischer Verfahren begründet, die nicht von einer bestimmten Verteilungsannahme abhängen. Ist in einzelnen Fällen speziellere Information vorhanden, kann diese ja durch entsprechende Verfahren berücksichtigt werden.

III. Verteilungsunabhängiges Testverfahren

Im Folgenden wird die Überprüfung von auf Stichprobenbasis erhobenen Daten durch eine Methode beschrieben, die die Problematik der nicht bekannten Verteilungsannahme umgeht. Zur korrekten Darstellung sind einige mathematische Bezeichnungen erforderlich, deren Bedeutung in den nachfolgenden Beispielen anschaulich wird.

1. Empirische Verteilungen

Zu Beginn liegt nur die Stichprobe des zu Überprüfenden vor. Sei $\{x_1, \dots, x_m\}$ die Menge der entsprechenden Stichprobendaten. Zur Überprüfung wird vom Prüfer ebenfalls eine Menge von Stichprobendaten $\{y_1, \dots, y_n\}$ ausgewählt. m, n geben die Anzahlen der jeweils gewählten Stichprobenelemente an und können verschieden sein. Die Kenntnis der Stichprobendaten ist äquivalent zur Kenntnis der jeweiligen empirischen Verteilungsfunktionen, die sich folgendermaßen bestimmen lassen (siehe auch die zu den späteren Beispielen gehörigen Abbildungen).

Für jede Zahl r kann man die Anzahl der Stichprobendaten betrachten, welche einen Wert kleiner oder gleich r haben. Diese Anzahl dividiert durch die Gesamtzahl der Stichprobendaten ergibt den Wert der empirischen Verteilungsfunktion an der Stelle r . Mathematisch ausgedrückt:

Sei r eine reelle Zahl, $(-\infty, r]$ die Menge der reellen Zahlen, die kleiner/gleich r sind. Bezeichne
 $\#(\{x_1, \dots, x_m\} \cap (-\infty, r])$

die Anzahl der Stichprobendaten aus $\{x_1, \dots, x_m\}$, die kleiner/gleich r sind, so heißt

$$F_m(r) = \frac{1}{m} \cdot \#(\{x_1, \dots, x_m\} \cap (-\infty, r])$$

empirische Verteilungsfunktion bzgl. $\{x_1, \dots, x_m\}$.

Entsprechend erhält man mit

$$G_n(r) = \frac{1}{n} \cdot \#(\{y_1, \dots, y_n\} \cap (-\infty, r])$$

die empirische Verteilungsfunktion bzgl. $\{y_1, \dots, y_n\}$.

Für die schon zu Beispielszwecken benutzte Schicht I aus NIES (1975, S. 78 ff.) hat die zur 23-elementigen Stichprobenmenge

$\{y_1, \dots, y_{23}\} = \{53, 42, 42, 35, 42, 53, 53, 35, 53, 53, 42, 35, 53, 42, 42, 53, 53, 53, 53, 42, 60, 53\}$

gehörige empirische Verteilungsfunktion G_{23} die folgende Gestalt

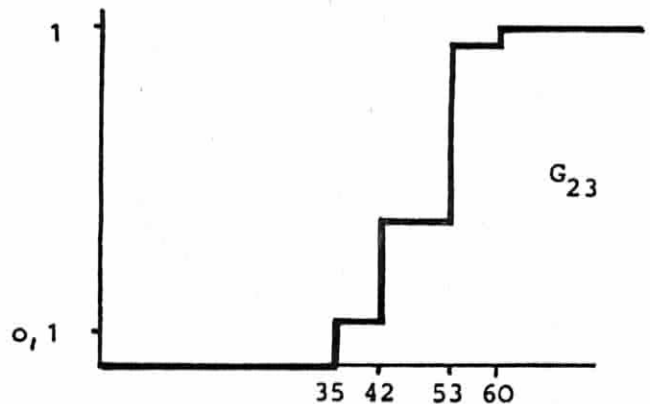


Abb. 2(a)

Da in der obigen Stichprobenmenge keine Daten kleiner als 35 enthalten sind, hat die empirische Verteilungsfunktion für alle Zahlen r kleiner als 35 den Wert 0. An der Stelle 35 springt die empirische Verteilungsfunktion auf den Wert $\frac{3}{23}$, da die 35 dreimal in obiger Stichprobe auftritt, bleibt dann für alle Zahlen r kleiner als 42 unverändert, um an der Stelle 42 einen weiteren Sprung der Höhe $\frac{7}{23}$ zu machen u.s.w.

Abbildung 2(b) vergleicht die so gefundene empirische Verteilungsfunktion mit der in Abbildung 1 gezeigten tatsächlich vorliegenden Verteilung der Schicht I.



Abb. 2(b)

Erläuterung zu Abb. 2(b): Die dick ausgezogene Linie entspricht der empirischen Verteilungsfunktion G_{23} der Stichprobe aus NIES (1975, S. 78 ff.). Die dünne Linie entspricht der tatsächlich vorliegenden Verteilung dieser Schicht. Man erkennt, daß die empirische Verteilungsfunktion von der tatsächlichen zwar etwas abweicht, aber die Übereinstimmung ist noch recht gut.

Auf der Grundlage der bzgl. seiner Prüfstichprobe $\{y_1, \dots, y_n\}$ erhaltenen empirischen Verteilungsfunktion $G_n(r)$ führt der Prüfer jetzt einen Vergleich mit der entsprechend einfach zu bestimmenden empirischen Verteilung $F_m(r)$ bzgl. der Stichprobe $\{x_1, \dots, x_m\}$ des zu Überprüfenden durch. Sind beide Stichproben repräsentativ, werden sich ihre zugehörigen empirischen Verteilungsfunktionen – und die tatsächlich vorliegende Verteilungsfunktion – kaum unterscheiden. Sind gewisse Stichprobenelemente unterrepräsentiert – z. B. solche mit hohen Bewertungen – oder überrepräsentiert, wird dies durch die Differenz der Funktionswerte deutlich, wie folgende Beispiele zeigen.

Würde der Überprüfte z. B. die Stichprobe

$\{x_1, \dots, x_{20}\} = \{35, 35, 42, 35, 53, 35, 35, 35, 35, 35, 35, 35, 35, 60, 42, 35, 35, 42, 35, 35, 35\}$

vorlegen, so ergibt sich die empirische Verteilungsfunktion F_{20}^I von Abb. 3(a).

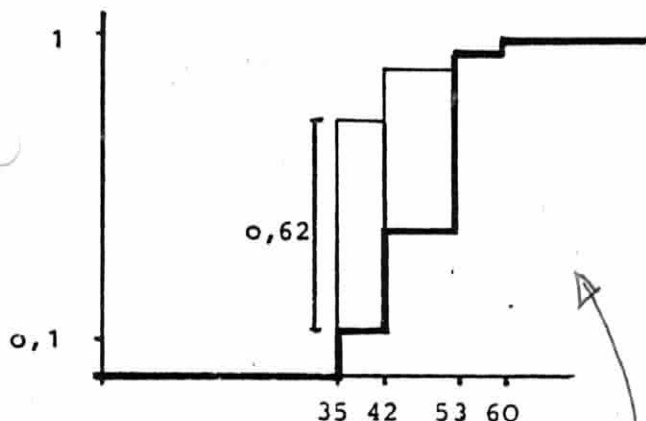


Abb. 3(a)

Abb. 3(b) vergleicht diese mit der empirischen Verteilungsfunktion G_{23} des Prüfers.

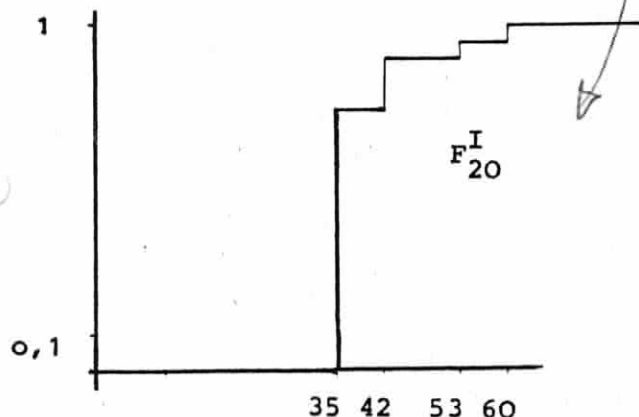


Abb. 3(b)

Erläuterung zu Abb. 3(b): Die dünn ausgezogene Linie entspricht der empirischen Verteilungsfunktion des Überprüften F_{20}^I . Die dick ausgezogene Linie entspricht der empirischen Verteilungsfunktion des Prüfers G_{23} .

Schon anschaulich erkennt man eine große Diskrepanz zwischen den beiden empirischen Verteilungsfunktionen.

Würde der Überprüfte hingegen folgende Stichprobe

$\{x_1, \dots, x_{20}\} = \{35, 42, 42, 60, 53, 53, 53, 42, 35, 53, 53, 42, 42, 42, 60, 53, 53, 53, 42, 42\}$

vorlegen, so ergibt sich die empirische Verteilungsfunktion F_{20}^{II} von Abb. 4(a).



Abb. 4(a)

Abb. 4(b) vergleicht diese wiederum mit der empirischen Verteilungsfunktion des Prüfers G_{23} .

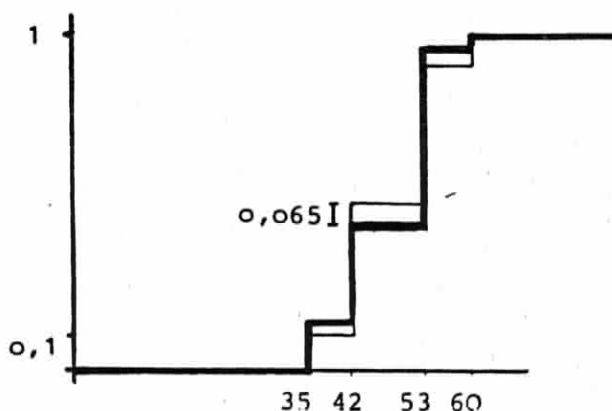


Abb. 4(b)

Erläuterung zu Abb. 4(b): Die dünn ausgezogene Linie entspricht der empirischen Verteilungsfunktion des Überprüften F_{20}^{II} . Die dick ausgezogene Linie entspricht der empirischen Verteilungsfunktion des Prüfers G_{23} .

Hier erkennt man anschaulich nur eine geringe Diskrepanz zwischen den beiden empirischen Verteilungsfunktionen.

2. Testverfahren

Die Beispiele zeigen, daß der Ausdruck

$$D_{m,n} = \max_r |F_m(r) - G_n(r)|$$

welcher den maximalen Absolutbetrag der Differenz zwischen der empirischen Verteilungsfunktion des zu Überprüfenden $F_m(r)$ und der empirischen Verteilungsfunktion des Prüfers $G_n(r)$ mißt, geeignet zur Bewertung der Stichprobe des Geprüften ist. In Abb. 3(b) und Abb. 4(b) ist $D_{m,n}$ mit 0,62 bzw. 0,065 eingezeichnet.

Die Frage, welche Größe dieser Ausdruck haben muß, damit man die Stichprobe des Überprüften als ungeeignet verwirft, ist mit Hilfe des unter dem Namen Kolmogorov-Smirnov-Zwei-Stichproben-Test bekannten Verfahrens zu beantworten, das z. B. aus folgenden Gründen für die vorliegende Situation geeignet ist.

Es reagiert nicht nur auf verschiedene Mittelwerte, sondern auch auf Unregelmäßigkeiten zwischen den Stichproben. Es ist einfach zu handhaben, da nur die Kenntnis

der beiden Stichprobenmengen benötigt wird. Zudem sind Vergleichswerte für den Ausdruck $D_{m,n}$ vertafelt (HARTER/OWEN (1970), PEARSON/HARTLEY (1972)). Die in einer solchen Tafel für diverse m und n und verschiedene Signifikanzniveaus α aufgelisteten Vergleichswerte $c_{m,n,\alpha}$ – auch kritische Werte genannt – erlauben nämlich, wenn für das aus den Stichproben errechnete $D_{m,n}$ die Ungleichung

$$m \cdot n \cdot D_{m,n} \geq c_{m,n,\alpha}$$

gilt, die Stichprobe des Überprüften als ungeeignet zu verwerfen. Da ja, wie schon erwähnt, jedes statistische Verfahren mit einem Restfehler behaftet ist, gibt hier das Signifikanzniveau α die Wahrscheinlichkeit dafür an, daß die Stichprobe des Überprüften fälschlicherweise verworfen wird. Dieses Restrisiko ist der Preis dafür, daß beim Überprüfen durch den Einsatz statistischer Methoden erhebliche Rationalisierungseffekte auftreten. Schließlich trägt ja auch der Prüfer ein Restrisiko, wenn er bei der Aufstellung von Jahresabschlüssen statistische Verfahren zuläßt.

Man beachte dabei, daß in der hier vorgenommenen praktischen Anwendung nicht genau das exakte gewählte Signifikanzniveau α eingehalten wird, sondern daß das tatsächliche Signifikanzniveau in der Regel kleiner als α sein wird, da der Test bei dieser Anwendung auf nicht-stetige Verteilungen konservativ ist (NOETHER (1967), S. 14ff.).

Mit anderen Worten, hat man $\alpha = 0,05$ gewählt, d. h. rechnet man damit, daß in 5% der Fälle der Test die Stichprobe des Geprüften verwirft, obwohl dieser seine Stichprobe korrekt erhoben hat, so wird in der Anwendung tatsächlich in weniger als 5% der Fälle eine solche Fehlentscheidung getroffen werden.

Betrachtet man nun Beispiel 3(b), so hat man dort

$$D_{20,23} = 0,62$$

Man entnimmt einer der genannten Tafeln (hier PEARSON/HARTLEY (1972), daß der errechnete Wert

$$m \cdot n \cdot D_{m,n} = 20 \cdot 23 \cdot 0,62 = 285,2$$

größer ist als der Wert

$$c_{20,23,0,1} = 164$$

ja sogar noch größer als der Wert

$$c_{20,23,0,001} = 262$$

Das heißt: Hat der Geprüfte (und natürlich auch der Prüfer) ordnungsgemäß die Stichprobe erhoben, so würde eine Zahl größer/gleich als 164 bzw. 262 nur in 10% bzw. 1% der Fälle auftreten.

In einer solchen Situation wird man natürlich dazu neigen, an der Repräsentativität der Stichprobe zu zweifeln und weitere Überprüfungen anstellen.

Betrachtet man nun Beispiel 4(b), so hat man jetzt

$$D_{20,23} = 0,065$$

also

$$m \cdot n \cdot D_{m,n} = 20 \cdot 23 \cdot 0,065 = 29,9$$

Man entnimmt der genannten Tafel, daß selbst zu $\alpha = 0,1$ der bekannte Wert $c_{20,23,0,1} = 164$ gehört.

Selbst wenn man also zu Ungunsten des Geprüften 10% Fehler im Verfahren zulassen würde, müßte sich eine Zahl größer/gleich 164 ergeben, damit an seiner Stichprobe

Zweifel angemeldet werden könnten. Das heißt, hier scheint alles in Ordnung zu sein.

Für die praktische Anwendung des Verfahrens beachte man, daß fertige Computerprogramme vorliegen (siehe BEUTEL u. a. (1978)).

Liegt zwischen der Entnahme der beiden Stichproben ein längerer Zeitraum, innerhalb dessen sich die Zusammensetzung des Lagerbestandes – allerdings nicht grundlegend – geändert hat, so müssen beide Stichproben durch Zu- bzw. Abrechnung der in der Zwischenzeit erfolgten Ab- und Zugänge auf einen gemeinsamen Zeitpunkt bezogen werden, damit das Verfahren angewendet werden kann.

IV. Zusammenfassung

1. Seit einiger Zeit werden aus Rationalisierungsgründen Methoden der mathematischen Stichprobenanalyse bei der Inventur und bei der Aufstellung von Jahresabschlüssen zugelassen.
2. Der Aussagegehalt von auf Stichprobenbasis erhobenen Daten kann mit entsprechenden statistischen Verfahren überprüft werden.
3. Zur Anwendung eines auf spezielle Verteilungsannahmen gestützten statistischen Verfahrens steht in der Praxis i. d. R. nicht genug Information zur Verfügung, daher wird ein verteilungsunabhängiges Verfahren vorgestellt.
4. Das vorgeschlagene Testverfahren wurde gewählt, weil
 - es nicht nur auf verschiedene Mittelwerte, sondern auch auf Unregelmäßigkeiten zwischen den Stichproben reagiert,
 - es einfach zu handhaben ist,
 - die benötigten Vergleichswerte vertafelt sind,
 - fertige Computer-Programme vorliegen.
5. Beispiele sind zur Verdeutlichung beigelegt.

Literatur

- BEUTEL, P./H. KUFFNER/E. ROCK/W. SCHUBO (1978): SPSS 7. C. Fischer Verlag, Stuttgart.
- HARTER, H. L./D. B. OWEN (1970): Selected Tables in Mathematical Statistics. Markham Publishing Co., Chicago.
- HOVERMANN, K. (1976): Phasen der Gewinnung stichprobenfundierter Prüfurteile bei Jahresabschlußprüfungen. Die Wirtschaftsprüfung 29, Heft 10/1976.
- NIES, H. (1975): Rationalisierung der körperlichen Bestandsaufnahme des Vorratsvermögens. Die Steuerliche Betriebsprüfung 15, Heft 4/1975.
- NIES, H. (1976): Mathematisch-statistische Methoden als Hilfsmittel zur Aufstellung des Inventars. Die Steuerliche Betriebsprüfung 16, Heft 11/1976.
- NOETHER, G. E. (1967): Elements of Nonparametric Statistics. John Wiley & Sons.
- PEARSON, E. S./H. O. HARTLEY (1972): Biometrika Tables for Statisticians, Vol. II. University Press, Cambridge.
- STROBEL, W. (1973): Die Bedeutung mathematischer Stichprobenanalysen für die Praxis von Jahresabschlußprüfungen. Die Wirtschaftsprüfung 26, Heft 18/1973.
- v. WYSOCKI, K. (1975): Zur Verwendung von Nomogrammen bei der Auswertung von Zufallsstichproben. Die Wirtschaftsprüfung 28, Heft 18/1975.
- v. WYSOCKI, K. (1976): Die Kapitalflußrechnung als Ergänzung des Jahresabschlusses. Die Wirtschaftsprüfung 29, Heft 1/2/1976.