

Wichtige Begriffe zur Summarischen Risikoprüfung (SRP) und zur Neuen Prüfungstechnik (NiPt)

Dipl.-Finw. Andreas Wähnert, Kiel

Die digitale Form der Außenprüfung hat neue Begriffe in den Betriebsprüfungsalltag hineingetragen. Bereits die GDPdU¹ – bzw. neuerdings die GoBD² – haben neue (technische) Termini etabliert, welche die heutzutage überwiegende elektronische Art der Betriebsverwaltung betreffen.

Durch die digitale Aufzeichnung und Aufbewahrung von Informationen hat sich die Ausgangslage für Außenprüfungen mit wesentlich umfangreicheren Unterlagen (Datenbeständen) und speziellen Fehler- und Manipulationsrisiken grundlegend verändert, so dass auch nur eine angepasste Prüfungsweise dieser gerecht werden kann.³ Im Gegenzug hat die Einführung von IDEA[®] als bundeseinheitliches digitales Prüfungswerkzeug neue Ausdrücke der Datenanalyse mitgebracht, die zum größten Teil aus den Bereichen „Mathematik“ und „Wirtschaftswissenschaften“ stammen und z.B. in der Wirtschaftsprüfung bereits seit langer Zeit bekannt sind.⁴ Die Summarische Risikoprüfung (SRP) hat als systematisches Prüfungsnetz wenige zusätzliche Begriffe und einige notwendige Differenzierungen ergänzt.

Diese Stichwortliste soll eine Hilfestellung für den Betriebsprüfungsalltag darstellen, indem wesentliche Begriffe der Summarischen Risikoprüfung (SRP) bzw. Neuen Prüfungstechnik (NiPt) kurz und (möglichst) allgemeinverständlich erklärt werden. Dabei erhebt sie nicht den Anspruch, mathematische und wirtschaftswissenschaftliche Termini umfassend zu beschreiben, sondern konzentriert sich auf den Kontext der steuerlichen Außenprüfung.

Inhaltsverzeichnis

ABC-Analyse	6
Absoluter Rang (Absolutrang)	6
Abweisungswert	6
Arithmetische Kennzahlen	6
Aufschlagskalkulation	7
Aufschlagsatzzeitreihe	7
Ausbeutekalkulation	7
Beanstandungsanlass (§ 158 AO)	8
Benford-Verteilung (Benford-Gesetz)	8
Betriebsbesichtigung	9
Betriebsprofil (Betriebswirtschaftliches Profil)	9
Beweis	9
Binomialverteilung	9
Biokybernetische Sensitivitätsanalyse	10
Box-Plot (Box-Whisker-Plot)	10
Chancengleichheit	10

¹ „Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen (GDPdU)“, BMF-Schreiben v. 16.7.2001, BStBl I 2001, 415.

² „Grundsätze zur ordnungsgemäßen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD)“, BMF-Schreiben v. 14.11.2014, BStBl I 2014, 1450.

³ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 27.

⁴ IDW-Prüfungshinweis „Einsatz von Datenanalysen im Rahmen der Abschlussprüfung“ (IDW PH 9.330.3).

Chi-Quadrat-Test (χ^2 -Test)	11
Datenanalyse	11
Deskriptive Statistik	11
Dezil	12
Dezimalsprung	12
Digitale Datenanalyse	12
Digitale Ziffernanalyse (digZA)	12
Diskrete Verteilungen	12
Doppelskalierung (Zeitreihenvergleich)	12
Doppelverkürzung	13
Effektiver Rechtsschutz	13
Einerstelle	13
Erstziffer	13
Erwartungsideal	13
Explorative Statistik	13
Feste Zahlenpositionen	13
Formelle Ordnungsmäßigkeit bzw. Mängel	13
Fraktile	14
Freiheitsgrade	14
Fuzzy Logic	14
Geldverkehrsrechnung	14
Gemeinsame Sachaufklärungsverantwortung	15
Gesamtbildmethode	15
Gestutztes Mittel (Gestutztmittel)	15
Gleichverteilung	15
Gleitschlittenzeitreihe	16
Gutachten	17
Häufigkeitsanalyse	17
Hypergeometrische Verteilung	17
Hypothese	17
Idealerwartung	18
IKS	18
Indizien Gesamtheit	18
Induktive Statistik	18
Interaktive Sach(verhalts)aufklärung	18
Internes Kontrollsystem (IKS)	19
Interpolation	19
Iterationsanalyse	19
Kalkulation	20

Klassifizierung.....	20
Koinzidenzanalyse.....	20
Kombinatorik.....	20
Komplettmanipulation.....	21
Konfidenzniveau.....	21
Konjunkturbereinigung (Konjunkturanpassung).....	22
Kongruenz.....	22
Korrelation.....	22
Korrelationskoeffizient.....	22
Kybernetische Sensitivitätsanalyse.....	22
Liquiditätsrechnung.....	22
Logarithmische Normalverteilungen.....	23
Manipulationspsychologie.....	23
Materielle Ordnungsmäßigkeit bzw. Mängel.....	24
Median.....	24
„Mit an Sicherheit grenzender Wahrscheinlichkeit“.....	24
Mittelwert.....	24
Modalklasse (Modalwert).....	24
Monetary Unit Sampling (MUS).....	24
Musterbasierte Datenanalyse.....	25
Nachkalkulation.....	25
Neue interaktive Prüfungstechnik (NiPt).....	25
Newcomb-Benford-Law (NBL).....	26
Nominaler Zeitreihenvergleich.....	26
Normalverteilungen.....	26
Nullhypothese.....	26
Permutation.....	26
Perzentil.....	26
Poisson-Verteilung.....	26
Probabilistische Schlussfolgerungen.....	26
Prozentrang.....	27
Prüfungsnetz.....	27
Punkterwartung.....	27
Quantil.....	27
Quantilsschätzung.....	28
Quartil.....	29
Quintil.....	29
Rangkennzahlen.....	29
Realer Zeitreihenvergleich.....	29

Rechtsschutz	29
Referenz(-Verteilung)	29
Regelbasierte Datenanalyse	29
Relativer Rang (Relativrang)	30
Richtigkeitsvermutung	30
Rollierende Zeitreihe	30
Sachaufklärungsverantwortung	30
Sachverständigengutachten	31
Schichtung	31
Schnittstellenverprobung (SSV)	31
Schwebende Zahlenpositionen	31
Sensitivitätsanalyse	31
Sensitivitätsmodell nach Vester	32
Signifikanz	32
Spitzenkappung	32
SRP (Summarische Risikoprüfung)	33
SRP-Vorlagen	33
Standardabweichung	33
Standardnormalverteilung	34
Statistik	34
Stetige Verteilungen	34
Stichprobenprüfung	34
Stochastik	35
Strukturanalyse	35
Summarische Risikoprüfung (SRP)	35
Variationskoeffizient	35
Verfahrensdokumentation	35
Verifikationsprinzip	36
Vermögenszuwachsrechnung	36
Verteilungsanalyse	36
Waffengleichheit	37
Wahrscheinlichkeitslehre (Stochastik)	37
Wahrscheinlichkeitstest	37
χ^2 -Test	37
z-Test	37
Zeitreihenanalyse	38
Zeitreihenbasierte Schätzung	39
Zeitreihenvergleich	39
Ziffernanalyse	39

Zweitziffer	40
-------------------	----

ABC-Analyse

Die ABC-Analyse ist eine – auch in der Wirtschaftsprüfung verbreitete – Methode zur Einteilung einer größeren Datenmenge in drei verschiedene Klassen ([Klassifizierung](#)) i.d.R. mit folgender effizienzbezogener Ausrichtung:

1. Einzelvorgänge, die aufgrund ihrer Größe alleine unternehmensrelevant sind.
2. Vorgänge, die nicht einzeln aber in ihrer Gesamtheit unternehmensrelevant sind.
3. Vorgänge, die auch in ihrer Summe keine große Bedeutung für das Unternehmen haben.

Bei der ABC-Analyse handelt es sich um eine einfache Art der Datenklassifizierung, die auch ohne Visualisierung der [Häufigkeitsmuster](#) zur Risikoanalyse und Planung geeignet ist. Unter digitalen Analysebedingungen lassen sich weiterführende Häufigkeitsauswertungen, die z.B. in IDEA® als „Schichtung“ bezeichnet werden, vielseitig einsetzen, um durch [Musterverbildlichung](#) Regelmäßigkeiten und Störungen als sinnvolle Prüfungseinstiege zu identifizieren.

Absoluter Rang (Absolutrang)

Absolute Ränge geben die größenbezogene Position eines Werts innerhalb einer Datengesamtheit wieder. Dazu werden die einzelnen Daten nach ihrer Größe sortiert und mit auf- oder absteigenden Rängen durchnummeriert.

In der Besteuerung sind absolute Ränge z.B. aus der Hinzurechnung bei der Gewerbesteuer bekannt, wo nach der alten GewStR 1998 der achtniedrigste⁵ Kontokorrentschuldenstand für die Qualifizierung als Dauerschuldzinsen festgelegt wurde.⁶

Als betriebswirtschaftliche und Prüfungsparameter ([RangKennzahlen](#)) sind absolute Ränge nur dann geeignet, wenn die Datengesamtheiten gleich groß sind. Um diese Problematik grundsätzlich auszuschließen, wird i.d.R. mit relativen Rängen wie [Quantilen](#) (Prozenträngen) gearbeitet.

Abweisungswert

Der Abweisungswert bezeichnet in der [Wahrscheinlichkeitslehre](#) (Stochastik) eine definierbare Aussagesicherheit, ab welcher eine [Hypothese](#) als nicht zutreffend angesehen wird und deshalb verworfen werden muss.

Üblicherweise werden Abweisungsschwellen prozentual angegeben, wie z.B. 95 %⁷ eine in vielen Bereichen eingesetzte Grenze ist.⁸ Mitunter werden im Rahmen von [Wahrscheinlichkeitstests](#) auch die mit einer (Nicht-)Übereinstimmungswahrscheinlichkeit korrespondierenden Testgrößen verwendet: So gibt ein [Chi-Quadrat-Wert](#) $\geq 16,9$ unter 9 [Freiheitsgraden](#) – z.B. als Übereinstimmungsergebnis einer beobachteten Ziffernverteilung auf der Einerstelle von Bareinnahmen mit der [Gleichverteilung](#) – die Aussageklarheit von mindestens 95 % an, dass die beiden [Häufigkeitsverteilungen](#) nicht identisch sind.⁹

Arithmetische Kennzahlen

Arithmetische Kennzahlen sind Planungs- und Prüfungsparameter, die durch Berechnung ermittelt werden. Von einfacher Summenbildung z.B. bei Periodenwerten wie dem Monatsumsatz über Quotienten (z.B. Liquiditätsparameter) bis zu komplexen Kennzahlensystemen sollen sie in der betriebswirtschaftlichen Planung und Beratung sowie in der Prüfung aussagekräftige Beurteilungen ermöglichen, wenn der Umfang an Einzelvorgängen dafür zu groß ist.

In seinem Zeitreihenurteil bezieht sich der X. Senat des BFH auf den Durchschnittswert von Wochenaufschlagsätzen und damit auf eine arithmetische Kennzahl.¹⁰ Der Nachteil derartiger arithmetischer gegenüber [RangKennzahlen](#) ist, dass sie in aller Regel auch durch nicht repräsentative Ausreißer beeinflusst und möglicherweise (einseitig) verzerrt werden.¹¹

⁵ Rang 8 von 365 Tagesständen unter aufsteigender Sortierung.

⁶ GewStR 1998 45 (7).

⁷ 95 % markieren in [Normalverteilungen](#) den Bereich „[Mittelwert](#) \pm doppelte [Standardabweichung](#)“.

⁸ *Becker/Giezek/Weber/Wähnert*, „Der Beanstandungsanlass nach § 158 AO“, DStR 32/2016, 1878, mit Zitaten aus den Naturwissenschaften, der Technik und der Qualitätssicherung.

⁹ Die Umrechnung von Testgrößen in (Nicht-)Übereinstimmungswahrscheinlichkeiten kann mit Hilfe von Tabellenkalkulationsprogrammen wie z.B. MS Excel® objektiv und nachvollziehbar durchgeführt werden.

¹⁰ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 4 und 44 ff.

¹¹ Bspw. gibt das [arithmetische Mittel](#) der Datenmenge, die unter dem Begriff „[Box-Plot](#)“ abgebildet ist, diese mit dem Wert „57“ nicht repräsentativ wieder, während die mittlere [RangKennzahl](#) „[Median](#)“ mit „49“ eindeutig besser den Gehalt der Datenverteilung trifft, weil die beiden hohen Ausreißer letztere nicht verfälschen.

Aufschlagskalkulation

Bei einer Aufschlagskalkulation wird der Sollumsatz durch die Anwendung der produkt- bzw. gruppenbezogenen Aufschlagsätze auf die jeweiligen Waren- oder Leistungseinsätze reproduziert, um den (erklärten) Istumsatz zu verproben.¹²

Eine auffällige (Nach-)Kalkulation erfüllt nach Ansicht des BFH¹³ die Anforderungen an einen Beanstandungsanlass gem. § 158 AO. In der Drei-Stufen-Theorie des BFH werden [Geldverkehrs-](#) und [Vermögenszuwachsrechnungen](#) sowie Kalkulationen den Zeitreihenanalysen als Schätzungsmethode (und wohl auch als Verprobungsmethode) vorgezogen, weil sie den realistischen Besteuerungsgrundlagen näher kämen.¹⁴ Dazu ist anzumerken, dass es sich bei Kalkulationen einerseits um arbeitsaufwendige [Gesamtbildmethoden](#) handelt, auf die nur ordnungsgemäße Aufzeichnungen einen Anspruch haben¹⁵. Andererseits ist die Ergebnisgenauigkeit wesentlich von der Vollständigkeit und Richtigkeit der Ausgangsgrößen abhängig, welche häufig geschätzt werden müssen wie bspw. Eigenverbrauch und Verderb; Fehler der Ausgangsgrößen und bei der wertmäßigen Zuordnung der Aufschlagsätze verursachen über die Multiplikation „Hebelfehler“¹⁶.

Aufschlagsatzzeitreihe

Aufschlagsatzzeitreihen sind Varianten der [Zeitreihenanalyse](#), bei denen der Aufschlagsatz als Faktor aus Wareneinsatz und -umsatz chronologisch nach Perioden wie z.B. Wochen, Monaten oder Quartalen verbildlicht wird, um aus der Übersicht Erkenntnisse für die Planung, Prüfung und Beratung abzuleiten.

Obwohl Aufschlagsatzzeitreihen – zu denen auch die [Gleitschlittenzeitreihen](#) gehören – für eine Schlüssigkeitsprüfung im Sinn des § 158 AO nur mäßig geeignet sind,¹⁷ bildeten sie vor Einführung der [Summarischen Risikoprüfung](#) (SRP) hierfür das Standardinstrument in Wirtschaftsprüfung, Steuerberatung und Betriebsprüfung. Ebenso beschäftigt sich das Zeitreihenurteil des BFH¹⁸ mit einer Aufschlagsatzzeitreihe in Gleitschlittenform, während das FG Hamburg¹⁹ in einem Beschluss den Unterschied zu [doppelskalierten Zeitreihenvergleichen](#) anspricht.

Ausbeutekalkulation

Ausbeutekalkulationen verproben den Sollumsatz durch die (berechnende) Herleitung aus Wareneinsatzmengen und deren Verkaufspreisen.

Eine auffällige (Nach-)Kalkulation erfüllt nach Ansicht des BFH²⁰ die Anforderungen an einen Beanstandungsanlass gem. § 158 AO. In der Drei-Stufen-Theorie des BFH werden [Geldverkehrs-](#) und [Vermögenszuwachsrechnungen](#) sowie Kalkulationen den Zeitreihenanalysen als Schätzungsmethode (und wohl auch als Verprobungsmethode) vorgezogen, weil sie den realistischen Besteuerungsgrundlagen näher kämen.²¹ Dazu ist anzumerken, dass es sich bei Ausbeutekalkulationen einerseits um arbeitsaufwendige [Gesamtbildmethoden](#) handelt, auf die nur ordnungsgemäße Aufzeichnungen einen Anspruch haben²². Andererseits ist die Ergebnisgenauigkeit wesentlich von der Vollständigkeit und Richtigkeit der Ausgangsgrößen abhängig, weil Fehler bei der Ermittlung des Wareneinsatzes (unter Berücksichtigung aller Arten von Verlusten sowie der Gefahr von [Doppelverkürzungen](#)) und/oder der Verkaufspreise (bei mengenrichtiger Zuordnung von Preisänderungen wie Sonderverkäufen) über die Multiplikation „Hebelfehler“²³ verursachen.

¹² Ausführlicher dazu z.B.: *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, Berlin, 4. Auflage 2017, 233 ff.

¹³ BFH v. 18.3.1964 – IV 179/60 U, BStBl III 1964, 381.

¹⁴ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 60 ff.

¹⁵ BFH v. 21.2.1990 – X R 54/87, BFH/NV 1990, 683.

¹⁶ „Hebelwirkungen“ werden vom BFH thematisiert in: Ur. v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 44 ff.

¹⁷ Eine Gegenüberstellung von Faktorzeitreihen – wie Aufschlagsatzzeitreihen –, einseitig und doppelskalierten Zeitreihenvergleichen mit einer Bewertung anhand des [Korrelationskoeffizienten](#) – als statistischem Parameter für die Größenabhängigkeit – findet sich unter „[Zeitreihenanalyse](#)“.

¹⁸ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743; mit einer Klarstellung des Bezugs: RiBFH *Nöcker*, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle – der Registrierkasse Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548.

¹⁹ FG Hamburg v. 31.10.2016 – 2 V 202/16, Rz. 48: „Mit Hilfe der doppelskalierten Monatszeitreihe hat die Betriebsprüfung die Daten des Wareneinkaufs den Umsatzerlösen in zwei Einzelgrafiken gegenüber gestellt. Bei dieser Methode tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander.“

²⁰ BFH v. 18.3.1964 – IV 179/60 U, BStBl III 1964, 381.

²¹ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 60 ff.

²² BFH v. 21.2.1990 – X R 54/87, BFH/NV 1990, 683.

²³ „Hebelwirkungen“ werden vom BFH thematisiert in: Ur. v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 44 ff.

Beanstandungsanlass (§ 158 AO)

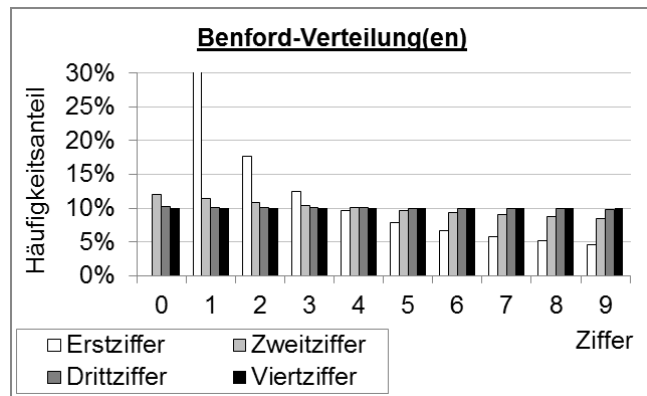
Der Gesetzgeber hat mit dem Beanstandungsanlass – als Teil des Verifikationsprinzips²⁴ – ein besonderes Beweismaß normiert, mit welchem die Gleich- und Gesetzmäßigkeit der Besteuerung sichergestellt werden soll. Neben der Ordnungsmäßigkeit der steuerlichen Aufzeichnungen bildet das Fehlen eines Beanstandungsanlasses die zweite (unabhängige) Voraussetzung des § 158 AO für die „Beweiskraft der Buchführung (und Aufzeichnungen)“ ([Richtigkeitsvermutung](#)).

Der Beanstandungsanlass muss als [probabilistische Schlussfolgerung](#) vom [Beweis](#) nach unten abgegrenzt werden, denn letzterer ist im besten Fall 100 % sicher. Eine zahlenmäßige Klarstellung der erforderlichen Aussagegewissheit fehlt in der Rechtsprechung, lässt sich aber anhand wissenschaftlicher Auseinandersetzungen zu Wahrscheinlichkeitsaussagen²⁵ und der höchstrichterlichen Anerkennung²⁶ bestimmter Verprobungsmethoden zur Vollständigkeitsprüfung von (Bar-)Einnahmen auf den Bereich „90-95 %“ eingrenzen; insofern muss die verbale Interpretation des Beanstandungsanlasses „mit an Sicherheit grenzender Wahrscheinlichkeit“ durch den BFH²⁷ als Fehldeutung angesehen werden, weil dies einen juristischen Beweis umschreibt.

Benford-Verteilung (Benford-Gesetz)

Als „Benford-Gesetz“ oder „Newcomb-Benford-Law (NBL)“ wird die Beobachtung bezeichnet, dass die führenden Zahlenpositionen – besonders in Wirtschaftszahlen – von der nach der [Kombinatorik](#) erwarteten [Gleichverteilung](#) des absoluten Zufalls abweichen – mit eigenen regelmäßigen Mustern.²⁸

Die abgebildeten besonderen Häufigkeitsverteilungen der führenden [schwebenden Zahlenpositionen](#) „Erstziffer“ und „Zweitziffer“ wurden von *Nigrini*²⁹ als Ausgangspunkt für [Musteranalysen](#) in der Wirtschaftsprüfung erkannt und von Ernst & Young³⁰ für Selektionsroutinen übernommen; gleichzeitig bestärken die empirischen Erkenntnisse die theoretische Annahme (annähernd) gleichverteilter Ziffernhäufigkeiten auf den nachfolgenden Zahlenpositionen wie z.B. der Einerstelle in Werten ≥ 100 .



Eine zentrale Anwendung der Musteranalyse auf die Benford-Verteilung stellt die Suche nach menschlichen Eingriffen z.B. in Form der Einnahmeverkürzung dar ([Zifferanalyse](#)), weil sie die Häufigkeitsmuster zerstören³¹; dabei müssen mögliche natürliche Ursachen für Musterabweichungen wie ein [Dezimalsprung](#) – manuell³² oder automatisch³³ – abgeklärt werden. Irritierend sind Bezeichnungen der automatisierten Datenanalyse wie „Benford-Test“, weil eine seriöse Gegenüberstellung einer Ziffernverteilung aus Prüfungsdaten mit der dazugehörigen Benford-Verteilung wegen der Abhängigkeit von der untersuchten Datenmenge³⁴ einen [Wahrscheinlichkeitstest](#) erfordert.

²⁴ BFH v. 16. 5. 2013 – II R 15/12, BStBl II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

²⁵ *Köller/Nissen/Rieß/Sadorf*, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Luchterhand Verlag München, 2004.

²⁶ BFH v. 21.2.1974 – I R 65/72, BStBl II 1974, 591 zur Geldverkehrs- bzw. Vermögenszuwachsrechnung sowie BFH v. 18.3.1964 – IV 179/60 U, BStBl III 1964, 381 zu Kalkulationen; bestätigt durch BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 39.

²⁷ BFH v. 24.6.1997 – VIII R 9/96, BStBl II 1998, 51; so auch: *Seer, Tipke/Kruse*, AO/FGO, § 158 AO Rz. 14.

²⁸ *Benford*, „The law of anomalous numbers“, Proceedings of the American Philosophical Society 1938, 551; die Benford-Verteilungen lassen sich durch eine einzige Formel für alle Zahlenpositionen berechnen.

²⁹ Nachzulesen unter: http://www.nigrini.com/datas_software.htm.

³⁰ Nachzulesen unter: [www.ey.com/Publication/vwLUAssets/EY_-_Forensic_Data_Analytics/\\$FILE/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf](http://www.ey.com/Publication/vwLUAssets/EY_-_Forensic_Data_Analytics/$FILE/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf).

³¹ *Hill*, „Random-number guessing and the first digit phenomenon“, Psychological Reports 62/1999, 967; *Hill*, „The difficulty of faking data“, Chance 12/3/1999, 27.

³² In der [Summarischen Risikoprüfung](#) (SRP) wird die Überprüfung auffälliger Ziffernhäufigkeiten durch nachgelagerte Vertiefungen gewährleistet, welche von den [SRP-Vorlagen](#) zweckmäßig unterstützt werden; vgl.: *Wähner*, „Der Beweiswert der Neuen interaktiven Prüfungstechnik (NiPt) – Chi-Quadrat-Test und vergleichbare Verfahren in der Betriebsprüfung“, BBK 8/2014, 380.

³³ Das Fraunhofer SIT nimmt dafür eine Modellierung des zu untersuchenden Datensatzes vor; nachzulesen unter: <https://www.sit.fraunhofer.de/de/angebote/kompetenzfelder/itforensics/forensische-finanzdatenanalyse>.

³⁴ Die Bedeutung der untersuchten Datenmenge für die (natürliche) Ergebnisstreuung um eine Idealerwartung ([Referenz](#)) herum kann aus der Abb. zum Begriff „[Binomialverteilung](#)“ abgelesen werden.

Betriebsbesichtigung

⇒ [Betriebsprofil](#)

Betriebsprofil (Betriebswirtschaftliches Profil)

Unter dem „Betriebsprofil“ wird im Rahmen einer [Summarischen Risikoprüfung](#) (SRP) die vorgeschaltete Sammlung wichtiger Unternehmensinformationen – mittels Aktenstudium, Betriebsbesichtigung und Befragung der Stpfl. – für eine seriöse Beurteilung der Ergebnisse aus den [regel-](#) und [musterbasierten Datenanalysen](#) verstanden. Der Begriff ist in die AO-Kommentierung übernommen worden.³⁵

Soll z.B. die Schlüssigkeit von Wareneinsatz zu -umsatz mit Hilfe eines [doppelskalierten Zeitreihenvergleichs](#) beurteilt werden, können Informationen aus den Bereichen „Wareneinkauf“, „Warenbezahlung“, „Lagerhaltung“, „Warenverlust“, „Absatzwege“ und „Rechnungslegung“ wichtig sein. Zusätzlich sollte ein Eindruck zur betriebswirtschaftlichen Kenntnis und Erfahrung bestehen. Auch für Musteranalysen wie bspw. die Einerstellenuntersuchung ([Ziffernanalyse](#)) von Bareinnahmen auf die Gleichverteilung – zur Aufdeckung einer möglichen [Komplettmanipulation](#) – sind gewisse betriebliche Bedingungen im Betriebsprofil zu dokumentieren, um Fehldeutungen des Ergebnisses zu vermeiden: Eine [kombinatorisch](#) problematische Preisgestaltung wie z.B. nur auf „0“ und „5“ endende Einzelpreise, sehr niedrige Verkaufsmengen pro Tag sowie pauschal abgerechnete Tagesveranstaltungen, weil diese Konstellationen die Erreichbarkeit aller Ziffern verhindern (können).

Beweis

Ein Beweis ist ein Beleg dafür, dass eine Aussage richtig ist.

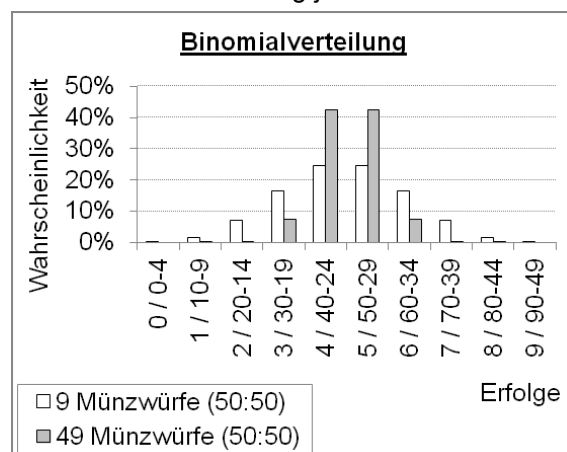
Nach wissenschaftlichem Verständnis wird hierfür auf die eindeutige Kausalität abgestellt: Zwischen dem belegten Vorgang und der zu beweisenden Aussage besteht zweifelsfrei eine Ursache-Wirkungs-Beziehung. Im Sinn [probabilistischer Schlussfolgerungen](#) wird somit die Aussagequalität „mit an Sicherheit grenzender Wahrscheinlichkeit“ erreicht.

Juristische Beweise müssen von einer niedrigeren Aussageklarheit ausgehen, weil die abzuurteilenden Vergangenheitssachverhalte in aller Regel nicht bis zu der genannten Eindeutigkeit aufklärbar sind.³⁶ Der Bundesgerichtshof hat bei einer Wahrscheinlichkeit von 99,5 % eine ausreichende Gewissheit für eine strafrechtliche Verurteilung gesehen.³⁷ Seer – als Autor des AO-Kommentars „Tipke/Kruse“ – sieht darin eine hohe Quantifizierung des Vollbeweises.³⁸ Entsprechend wird auch kriminalistischen Strengbeweisen wie DNA-Spuren, Fingerabdrücken etc. in wissenschaftlichen Studien nur eine Aussagesicherheit oberhalb von 99 % zugebilligt.³⁹

Binomialverteilung

Mit der [diskreten](#) Binomialverteilung wird das Streuverhalten für (dichotome) ja-nein-Zufallsereignisse wiedergegeben wie bspw. den Münzwurf; dabei umfasst die Binomialverteilung jedes Wahrscheinlichkeitsverhältnis der beiden möglichen Ereignisse, z.B. auch die Einzelbetrachtung einer [Ziffernanalyse](#) unter Annahme der [Gleichverteilung](#) mit der Relation 10 % (gewählte Ziffer) zu 90 % (alle anderen Ziffern).

Die nebenstehenden (glockenförmigen) Häufigkeitsverteilungen lassen sich über die [Kombinatorik](#) auszählen oder mit der Funktion der Binomialverteilung (z.B. in Tabellenkalkulationsprogrammen) berechnen. Typisch ist die deutliche Abhängigkeit von der Datenmenge, indem vergleichbar stark abweichende Ergebnisse bei größerem Datenumfang viel unwahrscheinlicher sind. Wegen dieser Eigenschaft



³⁵ Seer, Tipke/Kruse, AO/FGO, § 158 AO Rz. 17.

³⁶ Die Beweisoptionen der Zivilprozessordnung in Titel 6 ff. lassen diesen abgemilderten Anspruch bereits erahnen, indem die genannten Mittel wie „Augenschein“, „Urkunden“, „Parteivernehmung“ etc. üblicherweise eine Restunsicherheit in sich tragen.

³⁷ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

³⁸ Protokoll zum 140. Bochumer Steuerseminar für Praktiker und Doktoranden v. 4.3.2016, „Problemfragen der Schätzung nach § 162 AO“, nachzulesen unter: http://www.kompetenzzentrum-steuerrecht.de/v1/cms/upload/Institut/BochumerSteuerseminar/Protokoll_20100118.pdf.

³⁹ Giezek/Rupprecht/Wähner, „Wie sicher muss ein Verprobungsergebnis sein? – Anleihen aus der Kriminalistik zur sinnvollen Einordnung des Beanstandungsanlasses“, BBK 5/2017, 236, mit diversen Quellen.

der Zufallsstreuung werden [Wahrscheinlichkeitstests](#) wie die Binomialverteilung bei der manipulationsbezogenen [Musteranalyse](#) (z.B. Ziffernanalyse) zur Ergebnisbewertung herangezogen. Die häufig in der Literatur angeführten Bedenken hinsichtlich erforderlicher Mindestdatenmengen sind dann auf die Anwendungsregeln des jeweiligen Wahrscheinlichkeitstests beschränkt.

Zusätzlich griffen bzw. greifen einige Datenanalysewerkzeuge des qualifizierten [Stichprobenverfahrens](#) „[Monetary Unit Sampling](#) (MUS)“ aus Vereinfachungsgründen auf die Binomialverteilung zur beweisfähigen Feststellung des erforderlichen Stichprobenumfangs und zur Hochrechnung gefundener Fehler zurück, obwohl die Einzelwahrscheinlichkeit beim Stichprobenziehen nicht konstant ist, sondern sich durch das Nicht-Zurücklegen verändert (s. [Hypergeometrische Verteilung](#)).

Biokybernetische Sensitivitätsanalyse

⇒ [Sensitivitätsmodell nach Vester](#)

Box-Plot (Box-Whisker-Plot)

Box-Plots (bzw. Box-Whisker-Plots) sind i.d.R. [rangbezogene](#), verbildlichende Einordnungen von Datenmengen als Mittel der [deskriptiven Statistik](#), um aus deren Verteilung den Normbereich und die Ausreißer zu identifizieren.

Die rechtsseitig abgebildete Tabelle zeigt eine Datenmenge aus zwölf Werten, welche für einen besseren Überblick nach dem [absoluten Rang](#) absteigend sortiert ist; durch die Angabe des relativen Prozentrangs ([Quantils](#)) werden auch Vergleiche mit größeren und kleineren Datenumfängen ermöglicht. Innerhalb der Box liegen die „normalen“ Daten der Verteilung, während die „Whisker“ (Fühler) schwache Ausreißer markieren; außerhalb letzterer liegen starke Ausreißer.

Fachübergreifend ist diese Form der Dateneinordnung ein übliches Instrument zur Erlangung eines Überblicks über Datenbestände und auch in der Besteuerung vertreten: z.B. zur Fremdüblichkeitsbewertung von Verrechnungspreisen im Konzernverbund⁴⁰. Im Rahmen der [zeitreihenbasierten Quantilsschätzung](#)

werden die möglichen Schätzungsgrößen wie Monatsaufschlagsätze hinsichtlich ihrer (sachgerechten) Eignung ebenfalls per Box-Plot beurteilt.

Box-Plot			
Rang	Wert	Prozentrang (Quantil)	Box-Whisker-Plot
12	125	100,00%	
11	95	90,90%	
10	55	81,80%	
9	52	72,70%	
8	51	63,60%	
7	49	54,50%	
6	48	45,40%	
5	47	36,30%	
4	46	27,20%	
3	42	18,10%	
2	41	9,00%	
1	38	0,00%	

Chancengleichheit

In seinem Zeitreihenurteil betont der X. Senat des BFH den „Anspruch der Steuerpflichtigen auf einen effektiven Rechtsschutz durch die Gerichte“ und sieht in einer „Excel-Tabelle mit ca. 1 100 Zeilen zu je 10 Spalten, d.h. insgesamt ca. 11 000 Eintragungen“ eine Überforderung des Stpfl. durch die Betriebsprüfung.⁴¹

Dieses Problem der komplexer gewordenen Außenprüfungssituationen betreffend wird in der Literatur jedoch zurecht darauf hingewiesen, dass sich „[d]ie Buchführung ... von einer Durchschreibebuchführung hin zu einer umfassenden Datenverarbeitung weiterentwickelt [hat], die elektronische Form ... nunmehr die übliche [ist]. ... Diese Entwicklung geschah auf Betreiben und im Interesse der Unternehmen ...“⁴² Insofern ist die verstärkte Anwendung der digitalen [Datenanalyse](#) durch die Finanzverwaltung als logische und notwendige Konsequenz auf die schwierigere Prüfungssituation infolge der Digitalisierung der Betriebsverwaltung zu sehen.

Zum Nachvollziehen von digitalen Außenprüfungsergebnissen ist für sachverständige Dritte wie Steuerberater die genaue Benennung der untersuchten Besteuerungsdaten sowie die (übersichtliche) Dokumentation der Ergebnisse erforderlich, denn sämtliche Analyseoperationen der [SRP-Vorlagen](#) aus der [Summarischen Risikoprüfung](#) (SRP) sind z.B. mit Hilfe frei zugänglicher Tabellenkalkulations-

⁴⁰ BMF-Schreiben v. 12. 4. 2005 – IV B 4 – S 1341 – 1/05, BStBl I 2005, 570: „Grundsätze für die Prüfung der Einkunftsabgrenzung zwischen nahestehenden Personen mit grenzüberschreitenden Geschäftsbeziehungen in Bezug auf Ermittlungs- und Mitwirkungspflichten, Berichtigungen sowie auf Verständigungs- und EU-Schiedsverfahren“, Pkt. 3.4.12.5 „Bandbreiten und ihre Einengung“.

⁴¹ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 49-50.

⁴² Graf, „Anforderungen an die Vorlage von Einzelaufzeichnungen – Ende der ‚Schonzeit‘ für Daten aus Vor-Systemen?“, BBK 18/2013, 875, 876.

programme reproduzierbar und in der Wirtschaftsprüfung üblich.⁴³ Generell dürfte im Hinblick auf die Chancengleichheit resp. den Rechtsschutz gelten, dass derartige systematische Datenanalyzesysteme mit mehrperspektivischer Datenaufbereitung, interner Ergebnisabsicherung, standardisierter Methoden-anwendung und Dokumentation einfacher nachzuvollziehen sind als eine unüberschaubare Menge an Einzelprüfschritten und individuell programmierten Makros.

In Literatur⁴⁴ und Rechtsprechung⁴⁵ ist mit Blick auf die neueren Prüfungsansätze auch die Einholung von Sachverständigengutachten vorgeschlagen worden. Dabei ist jedoch darauf zu achten, dass sich auch die wissenschaftliche Expertise an die gesetzlichen Gewissheitsmaßstäbe „juristischer Be-
weis“ und „Beanstandungsanlass“ – in klarer Abgrenzung zum kausalen, wissenschaftlichen Beweis – halten muss.

Chi-Quadrat-Test (χ^2 -Test)

Der Chi-Quadrat-Test⁴⁶ ist ein gängiger⁴⁷ Wahrscheinlichkeitstest, mit dem eine beobachtete Datenmenge – gemäß den empirischen und theoretischen Erkenntnissen der Mathematik – auf die Übereinstimmung mit einer Referenzverteilung überprüft werden kann.

Als vorkonfektionierte Anwendung aus der Stochastik berechnet der Chi-Quadrat-Test einen Gesamtabweichungswert zur Gegenüberstellung der beobachteten Datenmenge und der erwarteten Verteilung. Dieser Chi-Quadrat-Wert korrespondiert – in Abhängigkeit von den Freiheitsgraden – über die Chi-Quadrat-Verteilungen mit einer Übereinstimmungswahrscheinlichkeit.

Die stochastische Absicherung eines Musterabgleichs z.B. bei der Ziffernanalyse der Einerstelle aus Bareinnahmen auf die Gleichverteilung ist deshalb besonders empfehlenswert, weil kleinere Datenmengen eine stärkere Streutendenz aufweisen als größere⁴⁸, was in den Ergebnissen von Wahrscheinlichkeitstests berücksichtigt wird; entsprechend gibt es i.d.R. konkrete Empfehlungen für Mindestdatenmengen zu einzelnen Wahrscheinlichkeitstests.⁴⁹ Mit der stochastischen Hochrechnung eines qualifizierten Stichprobenergebnisses auf die Grundgesamtheit aller relevanten Sachverhalte hat der Bundesgerichtshof⁵⁰ indirekt auch die Übereinstimmungsprüfung einer beobachteten Datenmenge auf eine erwartete Referenzverteilung mittels Wahrscheinlichkeitstest anerkannt, weil beide Ansätze der Statistik auf denselben Grundlagen fußen.

Unkorrekterweise wird in Literatur und Rechtsprechung der Chi-Quadrat-Test häufiger mit der Annahme gleichverteilter Ziffern auf der Einerstelle von Bareinnahmen gleichgesetzt. Diese fehlerhafte Vermischung von (Referenz-)Erwartung „Gleichverteilung“ und stochastischem Übereinstimmungstest verkennt zum einen das breite Einsatzgebiet des Chi-Quadrat-Tests und verkompliziert zum anderen die Besprechung auffälliger Betriebsprüfungsergebnisse unnötig.

Datenanalyse

Die (digitale) Datenanalyse ermöglicht eine umfassendere und schnellere Untersuchung datenintensiver Vorgänge durch Such-, Sortier- und Filterfunktionen und ist somit ein unverzichtbares Instrument für die Beratung und Prüfung im Zeitalter elektronischer Betriebsverwaltungen mit kontinuierlich zunehmenden Datenmengen.

Neben den Datenbankenfunktionen liegt ein wesentliches Verbesserungspotential der digitalen Datenanalyse in umfangreichen Optionen zur Datenverbildlichung und der Interaktivität für aussagekräftigere Gesamtbildmethoden wie im Sinn des § 158 AO (Beanstandungsanlass).

Deskriptive Statistik

⇒ Statistik

⁴³ IDW-Prüfungshinweis „Einsatz von Datenanalysen im Rahmen der Abschlussprüfung“ (IDW PH 9.330.3).

⁴⁴ Krumm, „Rechtsfragen der digitalen Betriebsprüfung (Summarische Risikoprüfung)“, DB 20/2017, 1105, 1107.

⁴⁵ BFH v. 12.7.2017 – X B 16/17, BFH/NV 2017, 1204, Rz. 137.

⁴⁶ Ohne weitere Differenzierung ist mit dem Chi-Quadrat-Test der Verteilungs- bzw. Anpassungstest gemeint, womit Daten auf die Übereinstimmung mit einer bestimmten Verteilung untersucht werden.

⁴⁷ Der Chi-Quadrat-Test gehört beim Abgleich zweier mehrklassiger (≥ 3) Verteilungen zu den gebräuchlichsten Verfahren, was daran zu erkennen ist, dass er in die verbreiteten Tabellenkalkulationsprogramme wie MS Excel® mit einem vollständigen Berechnungssatz integriert ist; bei einer Verteilungsgegenüberstellung mit lediglich zwei Klassen (z.B. Ziffernanalyse aus der Perspektive einer bestimmten Ziffer) ist die 100 % korrekte Binomialverteilung vorzuziehen.

⁴⁸ Diese Besonderheit des natürlichen Streuverhaltens lässt sich an der Kombinatorik der Binomialverteilung nachvollziehen (s. Abb. zu dem Begriff).

⁴⁹ Für die Anwendung eines Chi-Quadrat-Tests sollte jede Ausprägung mindestens einmal erwartet werden und 80 % von ihnen mindestens fünfmal. Auf die Gleichverteilung der zehn Ziffernmöglichkeiten auf der Einerstelle von Bareinnahmen bedeutet das eine Mindestdatenmenge von 50 (10 Ziffern • mindestens 5 erwartete Vorkommen – der erste Teil der Empfehlung entfällt als Folge der Gleichverteilung).

⁵⁰ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

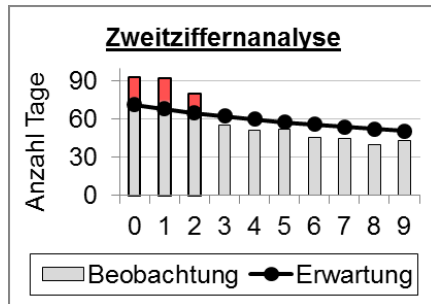
Dezil

⇒ [Quantil](#)

Dezimalsprung

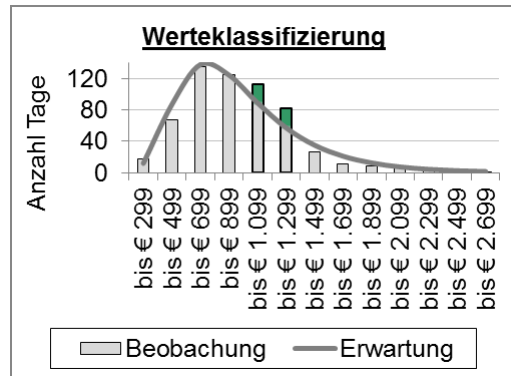
Ein Dezimalsprung im Wertebereich von (Bar-)Einnahmen ist die praxisrelevanteste nicht-manipulative Ursache für eine [signifikante](#) Abweichung von der [Benford-Verteilung](#) der führenden [schwebenden Zahlenpositionen](#) ([Ziffernanalyse](#)).

Obwohl gerade die Dezimalsprünge in z.B. [logarithmisch-normalverteilten](#) Daten die besonderen –



von der [Gleichverteilung](#) abweichenden – Häufigkeitsmuster der Benford-Verteilung erzeugen, bewirken sie bei geringem Wertumfang eine Überpräsenz der kleinen Zweitziiffern auf Kosten der größeren; in der links abgebildeten Gegenüberstellung der Zweitziiffernhäufigkeiten aus Bareinnahmen mit der Benford-Verteilung ist diese regelmäßige Verschiebung rot hervorgehoben. Mit Hilfe einer [Datenklassifizierung](#) unter Fixierung⁵¹ der Zweitziiffer oberhalb des Dezimalsprungs lässt sich schnell und sicher feststellen, ob eine Ballung der Bareinnahmen in

diesem Bereich für die auffällige Häufigkeitsverteilung ursächlich ist: Der rechten Grafik ist ein starkes Vorkommen von Tagessummen im Bereich „> € 899 bis € 1.299“ (Balken mit grüner Spitze) zu entnehmen, wodurch eine Überbetonung der Zweitziiffern „0“, „1“ und „2“ zu Lasten der größeren erklärt wird. Als Ergebnis dieser vertieften [Musteranalyse](#) wird das Indiz der auffälligen Zweitziiffernbeobachtung betrieblich begründet und somit entkräftet.



Zur Abklärung aufgefunderer Zweitziifferauffälligkeiten – und damit zur Absicherung von Indizien – wird ein Dezimalsprung als mögliche Ursache im Rahmen der [Summarischen Risikoprüfung](#) (SRP) selbständig hinterfragt; die [SRP-Vorlagen](#) bieten zur SRP-Prüfungssystematik Werkzeuge, welche diese und andere praxisrelevante Vertiefungen effizient unterstützen.⁵²

Digitale Datenanalyse

⇒ [Datenanalyse](#)

Digitale Ziffernanalyse (digZA)

⇒ [Ziffernanalyse](#)

Diskrete Verteilungen

Diskret verteilte Datenmengen haben ausschließlich ganzzahlige Ergebnisse in ihrer „stufigen“ Struktur, weil eine Münze z.B. nur entweder 3- oder 4-mal auf die beobachtete Seite fallen kann, aber nicht 3,5-mal. Zu den diskreten Funktionen gehören die [Binomial](#)- und die [Hypergeometrische](#), deren Ergebnisverteilungen sich mit Hilfe der [Kombinatorik](#) herleiten lassen.

Doppelskalierung (Zeitreihenvergleich)

Das Mittel der Doppelskalierung dient bei einem [Zeitreihenvergleich](#) dazu, zwei auf Abhängigkeit zu untersuchende Datenreihen in der Größe so aneinander anzupassen, dass die relativen Veränderungen adäquat wiedergegeben werden. Dadurch lässt sich gut erkennen, ob eine Datenreihe die andere sowohl in der Veränderungsrichtung als auch in der -intensität bestimmt ([Korrelation](#)).

Im Gegensatz zur Prüfung zweier Datenreihen auf Übereinstimmung wie insbesondere bei der [Schnittstellenverprobung](#) (SSV) zielen Schlüssigkeitsuntersuchungen auf die nachvollziehbare Handlungslogik zweier als „abhängig“ erwarteter Betriebsfaktoren ab: So kann z.B. ohne vorhandene Ware

⁵¹ Durch eine Klasseneinteilung von Daten in z.B. #-99-Intervalle – ganz korrekt nach Entfernung der Nachkommastellen – wird die Zweitziiffer oberhalb von € 1.000 fixiert: In der Klasse „> € 899 bis € 999“ sind alle Ziffern auf der Zweitposition denkbar wie z.B. bei € 857 die „5“, während im folgenden Intervall „> € 999 bis € 1.099“ lediglich die „0“ vorkommen kann. Aus stochastischer Sicht verzehnfacht sich oberhalb des Dezimalsprungs somit die Wahrscheinlichkeit einer bestimmten Zweitziiffer im Vergleich zum Wertebereich unterhalb.

⁵² Wähnert, „Die Ziffernanalyse als Bestandteil zeitgemäßer Prüfungsansätze“, StBp 9/2012, 241.

kein Absatz erfolgen; unter sachgerechter Periodisierung müssten die beiden Datenreihen „Wareneinkauf“ und „-absatz“ aufzeigen, dass der Unternehmer in Abhängigkeit vom Verkauf Ware nachbezieht. Derartige Zusammenhänge lassen sich aus einem doppelskalierten Zeitreihenvergleich wesentlich besser ablesen als aus [Aufschlagsatzzeitreihen](#) oder einseitig skalierten Zeitreihenvergleichen.⁵³

Tabellenkalkulationsprogramme wie MS Excel[®] bieten die Doppelskalierung von Zeitreihenvergleichen als Formatierungsoption an. Bei solchen Automatikfunktionen wird der Zweck der Doppelskalierung jedoch dann verfehlt, wenn die Hauptbereiche der Grafen nicht aufeinander liegen, was besonders bei Extremausreißern in nur einer Datenreihe vorkommen kann; in dem Fall sollten die beiden Ordinaten (Y-Achsen) individuell angepasst werden.

Doppelverkürzung

Unter „Doppelverkürzung“ wird eine systematische Steuerhinterziehung verstanden, bei der sowohl der Warenabsatz als auch -einkauf in [korrelierendem](#) Umfang weggelassen werden. Der Steuerhinterzieher spart dabei die Umsatz- und die Gewinnbesteuerung auf die Marge.

Systematischer Steuerbetrug wie Doppelverkürzungen soll durch planvolles Vorgehen bekannte Prüfungsmethoden unterwandern, indem keine typischen [Manipulationsspuren](#) entstehen. Einige Warenwirtschaftssysteme unterstützen durch integrierte oder zusätzlich einspielbare Funktionen derartige Steuerhinterziehungen. Der BFH vermisst in seinem Zeitreihenurteil einen „empirische[n] Nachweis der Eignung dieser Verprobungsmethode zur sicheren Aufdeckung von Steuerverkürzungen – insbesondere der mit anderen Methoden nur schwer aufzudeckenden sog. ‚Doppelverkürzungen‘“⁵⁴. Dazu ist anzumerken, dass die Annahme, irgendeine [Gesamtbildmethode](#) könne systematische Einnahmeverkürzungen sicher aufdecken, der Logik systematischer – auf Prüfungsansätze reagierender – Manipulationen widerspricht; demgegenüber ist unzweifelhaft, dass die herkömmlichen [Kalkulationen](#) unterlaufen werden.

Effektiver Rechtsschutz

⇒ [Chancengleichheit](#)

Einerstelle

⇒ [Feste Zahlenpositionen](#)

Erstziffer

⇒ [Schwebende Zahlenpositionen](#) sowie [Benford-Verteilung](#)

Erwartungsideal

⇒ [Referenz](#)

Explorative Statistik

⇒ [Statistik](#)

Feste Zahlenpositionen

Mit festen Zahlenpositionen werden im Rahmen der [Ziffernanalyse](#) die wertbezogen gleichbedeutenden Stellen der Zahlen einer Datenmenge bezeichnet: So hat die Einerstelle in einem kleinen Datensatz von z.B. € 87,50 dieselbe monetäre Bedeutung wie in einem deutlich werthaltigeren Vorgang von bspw. € 3.416,00.

Diese Form der Zahlenbeobachtung entstammt der theoretischen [Referenz](#)-Herleitung, dass die freie Multiplikation von Größen (z.B. Verkaufshäufigkeiten mit den Einzelpreisen) zum absoluten Zufall der [Gleichverteilung](#) führen müsste, und findet z.B. in der Prüfung von (Bar-)Einnahmen auf die Häufigkeiten der Einerstelle (unter Übereinstimmungsbewertung mit dem [Chi-Quadrat-Test](#) oder der [Binomialverteilung](#)) Anwendung. Empirisch wird diese Erwartung nach dem [Benford-Gesetz](#) erst ab der [schwebenden](#) Drittziffer (annähernd) bestätigt.⁵⁵

Formelle Ordnungsmäßigkeit bzw. Mängel

⇒ [Richtigkeitsvermutung](#)

⁵³ Ein Vergleich von Zeitreihenmodellen im Hinblick auf die Abhängigkeit zweier Datenreihen findet sich zum Begriff „[Zeitreihenanalyse](#)“.

⁵⁴ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 43.

⁵⁵ Aus dem Grund der widerstreitenden Referenzen „Benford-Gesetz“ und „Gleichverteilung“ werden bei der Einerstellenuntersuchung der [Summarischen Risikoprüfung](#) (SRP) Datensätze < € 100 automatisch ausgeschlossen, was im Hinblick auf den absoluten Zufall der Gleichverteilung keine Verfälschungen zur Folge hat, aber die besondere Empirie der Wirtschaftsdaten (Benford-Gesetz) beachtet.

Fraktil

Neben [Quantilen](#) sind auch Fraktile relative Ränge, mit denen – im Gegensatz zu [absoluten Rängen](#) – eine vergleichende Bewertung trotz unterschiedlicher Datenmengen möglich ist.

Dabei wird die Datengesamtheit in Bruchteile eingeteilt, so dass z.B. das $\frac{3}{4}$ -Fraktil dem 75 %-Rang (Quantil) entspricht. Sollte sich dabei ein nicht-ganzzahliger Rang ergeben, wird zwischen den beiden angrenzenden Werten [interpoliert](#).

In der beschreibenden (deskriptiven) [Statistik](#) sowie der Betriebswirtschaftslehre werden derartige relative Ränge eingesetzt, um die robusteren⁵⁶ [Rangkennzahlen](#) über variable Datenumfänge hinweg nutzen zu können, z.B. damit vergleichbare Schwellenwerte bei der Beurteilung von Gehältern definiert werden können, selbst wenn zu den Branchen unterschiedlich viele Vergleichswerte vorliegen.

Freiheitsgrade

Die Freiheitsgrade als Anzahl unabhängiger Beobachtungen stellen für bestimmte [statistische](#) Parameter bzw. Ansätze eine notwendige Zusatzinformation zur Beurteilung dar.

Bei der Anwendung des [Chi-Quadrat-Tests](#) als Übereinstimmungstest für beobachtete und erwartete Eigenschaften von Prüfungsdaten (z.B. mit der [Ziffernanalyse](#) der [festen](#) Einerstelle aus Bareinnahmen auf die [Gleichverteilung](#)) geben die Freiheitsgrade die (maximal) „freie“ Klassenanzahl an: Eine willkürliche Ziffernverteilung könnte im Beispiel der Einerstelle höchstens auf neun Ziffern „frei“ verteilen, während für die letzte Ziffer ein fester Rest übrig bleibt; somit ergeben sich neun Freiheitsgrade. Diese Angabe der Klassenanzahl ist für den Chi-Quadrat-Test erforderlich, weil die einzelnen Chi-Quadrat-Klassenwerte vor der Aufsummierung relativ berechnet werden.⁵⁷

Fuzzy Logic

Die Unschärfelogik (Fuzzy Logic) bezeichnet eine Theorie der Mustererkennung, welche z.B. für die Lösung komplexer Probleme eingesetzt wird.⁵⁸

Im Rahmen der [Zeitreihenanalytik](#) bspw. findet sie sich als Suchlogik, Prüffelder über die Abbildungsschärfe von Daten effektiv und effizient aufzudecken, in der Periodenwahl wieder: Um sowohl grundsätzliche Abhängigkeitsstörungen z.B. beim zeitlichen Versatz einer ganzen Datenreihe als auch Einzelprüffelder zu erkennen, sind häufig unterschiedliche Summationsebenen der Vergleichsdaten notwendig. Also muss ein zeitreihenbasiertes Analysesystem von groben Intervallen wie „Quartalen“ zu detaillierteren Perspektiven führen und sollte dazu mehrere Zeitreihenvarianten (z.B. [Gleitschlittenzeitreihe](#)) wegen ihrer unterschiedlichen Ausrichtungen kombinieren.⁵⁹

Geldverkehrsrechnung

Mittels Geldverkehrsrechnungen soll die Vollständigkeit der erklärten Einnahmen verprobt werden, indem die liquiden Geldmittel den Ausgaben gegenübergestellt werden: Reicht das verfügbare Geld nicht für den Verbrauch inkl. getätigter Anschaffungen aus, ergibt sich ein Hinweis auf un versteuerte Erlöse.⁶⁰

Eine auffällige Geldverkehrsrechnung erfüllt nach Ansicht des BFH⁶¹ die Anforderungen an einen Beanstandungsanlass gem. § 158 AO. In der Drei-Stufen-Theorie des BFH werden [Geldverkehrs-](#) und [Vermögenszuwachsrechnungen](#) sowie Kalkulationen den Zeitreihenanalysen als Schätzungsmethode (und wohl auch als Verprobungsmethode) vorgezogen, weil sie den realistischen Besteuerungsgrundlagen näher kämen.⁶² Hierzu ist anzumerken, dass es sich bei Geldverkehrsrechnungen einerseits um arbeitsaufwendige [Gesamtbildmethoden](#) handelt, auf die nur ordnungsgemäße Aufzeichnungen einen Anspruch haben⁶³. Andererseits ist die Ergebnisgenauigkeit wesentlich von der Vollständigkeit und Richtigkeit der Ausgangsgrößen abhängig, wobei private Geldzuflüsse und -abflüsse nach Abschaf-

⁵⁶ Unter der „Robustheit“ von Kennzahlen wird verstanden, dass sie nicht bzw. wenig anfällig gegenüber alltäglichen Sachverhalten sind.

⁵⁷ Für die Beurteilung mehrerer aufsummierter prozentualer Abweichungen wäre z.B. ebenfalls die Anzahl an Prozentwerten notwendig.

⁵⁸ Vester, „Die Kunst vernetzt zu denken – Ideen und Werkzeuge für einen neuen Umgang mit Komplexität“, Deutscher Taschenbuch Verlag, München, 9. Auflage 2012.

⁵⁹ Wähner, „Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz „Summarische Risikoprüfung (SRP)““, DB 45/2016, 2627.

⁶⁰ Ergänzend dazu z.B.: Brinkmann, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, Berlin, 4. Auflage 2017, 238 f.

⁶¹ BFH v. 21.2.1974 – I R 65/72, BStBl II 1974, 591.

⁶² BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 60 ff.

⁶³ BFH v. 21.2.1990 – X R 54/87, BFH/NV 1990, 683.

fung der Vermögenssteuer nur schwierig nachzuverfolgen sind und außerdem das verbreitete⁶⁴ Steuerhinterziehungsmodell, Schwarzlöhne durch nicht erklärte Einnahmen zu bezahlen, im Rahmen von Betriebsprüfungen kaum ausgeschlossen werden kann.

Gemeinsame Sachaufklärungsverantwortung

⇒ [Interaktive Sach\(verhaltens\)aufklärung](#)

Gesamtbildmethode

Im besonderen Bürger-Behörden-Verhältnis des Besteuerungsmassenverfahrens erlaubt die Abgabenordnung der Finanzverwaltung Gesamtbildmethoden zur Übersichtsbeurteilung der Schlüssigkeit von Besteuerungsaufzeichnungen, weil „es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild der Buchführung als ausreichende Grundlage ... genügen kann.“⁶⁵ Nach dem Verifikationsprinzip⁶⁶ führt die Versagung der „Beweiskraft der Buchführung“ aufgrund von (gravierenden) Ordnungswidrigkeiten und/oder einem [Beanstandungsanlass](#) gem. §§ 158 i.V.m. 162 Abs. 2 S. 2 AO zur Amtspflicht der Schätzung.

Mit der Ordnungsmäßigkeit wird die Nachvollziehbarkeit der Aufzeichnungs- und Aufbewahrungsweise an den gesetzlichen Ordnungsvorgaben als formelles Gesamtbild überprüft. Sobald Ordnungsmängel die materielle Prüfung der Besteuerungsaufzeichnungen (in angemessener Zeit) verhindern, hat die Finanzbehörde zu schätzen.⁶⁷

Ein materielles Gesamtbild kann auf Basis von (verbildlichten) Kennzahlen(entwicklungen) oder Nachberechnungen die inhaltliche Logik der Besteuerungsdeklaration bestätigen bzw. mit einem Beanstandungsanlass widerlegen. Der Anwendungserlass zu § 158 AO konkretisiert dazu: „... wenn es nach Verprobung usw. unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt.“ Höchstrichterlich akzeptiert sind [Geldverkehrs-](#) und [Vermögenszuwachsrechnungen](#), [Kalkulationen](#) sowie statistisch qualifizierte Stichprobenprüfungen wie [Monetary Unit Sampling](#) (MUS). Digital unterstützte Prüfungsnetze wie die [Summarische Risikoprüfung](#) (SRP) greifen zusätzlich auf regelbasierte [Zeitreihen-](#) und [musterbasierte Häufigkeitsanalysen](#) zurück, welche in Form eines Prüfungsnetzes viele Informationen aus den Besteuerungsdaten zu einer möglichst aussagekräftigen [Indiziengesamtheit](#)⁶⁸ zusammentragen.

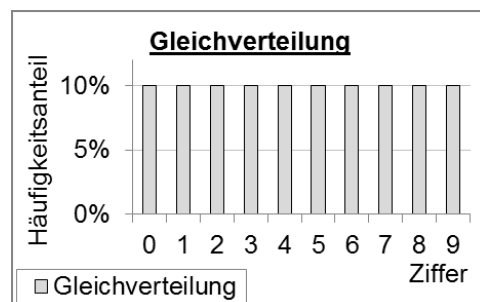
Gestutztes Mittel (Gestutztmittel)

Um den Vorteil der [Rangkennzahlen](#), von (asymmetrischen) Ausreißern nicht verzerrt zu werden, auch bei einem arithmetischen [Durchschnittswert](#) zu erreichen, wird beim Gestutztmittel ein frei bestimmbarer Prozentanteil der kleinsten und größten Werte aus der Berechnung ausgeschlossen.

Gleichverteilung

Als Erwartungsideal ([Referenz](#)) beschreibt die Gleichverteilung den absoluten Zufall mit (annähernd⁶⁹) identischen Häufigkeiten für alle möglichen Einzelergebnisse wie z.B. bei der Annahme, dass auf der [festen](#) Einerstelle von (Bar-)Einnahmen jede Ziffer gleichwahrscheinlich vorkommt.

Im Gegensatz zur empirischen [Benford-Verteilung](#) leitet sich die Punkterwartung „Gleichverteilung“ aus der mathematischen Theorie ab, indem die freie Multiplikation von Größen zum absoluten Zufall führt: Werden z.B. tägliche Zufallshäufigkeiten von Einzelverkäufen mit einem (einheit-



⁶⁴ Diese Erkenntnis hat dazu geführt, dass in der österreichischen Strategie zur Bekämpfung von Kassenmanipulationen der dortigen Finanzkontrolle Schwarzarbeit die Bareinnahmenprüfung via Kassennachschauen zugewiesen wurde; vgl. *Huber*, „Praktische Erfahrungen mit der Kassenführung in Österreich – Umsetzung der Kassennichtlinie 2012 und Kassennachschau durch die Finanzpolizei“, BBK 6/2014, 286.

⁶⁵ *Seer, Tipke/Kruse*, AO/FGO, § 158 AO Rz. 5.

⁶⁶ BFH v. 16. 5. 2013 – II R 15/12, BStBl II 2014, 225, vollständiges Zitat unter dem Begriff „[Beanstandungsanlass](#)“.

⁶⁷ BFH v. 15. 2. 1989 – X R 16/86, BStBl II 1989, 462: „... der ‚Beweisverderber‘ oder ‚Beweisvereiteler‘ aus seinem Verhalten keinen Vorteil ziehen darf ... Zur Vermeidung ... sind auch belastende Unterstellungen ... oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt.“

⁶⁸ BGH v. 27. 4. 2010 – 1 StR 454/09: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern ... Denn die Indizien können in ihrer Gesamtheit ... die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreich[en].“

⁶⁹ In der Spieltheorie wird in dem Zusammenhang von idealen Würfeln, Münzen etc. gesprochen, weil eine 100 %-ig gleichgewichtete Herstellung, die im Ideal eine exakte Gleichverteilung erwarten ließe, technisch kaum möglich ist.

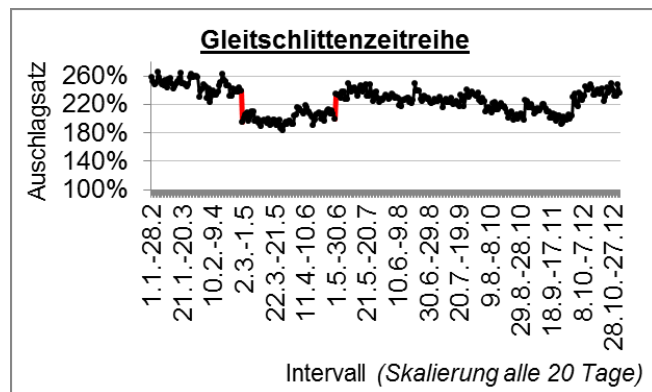
lichen) Stückpreis von € 3 zu Tagessummen aufsummiert, ergeben sich – ab einer Mindestmenge von Vorgängen – alle möglichen Ziffern auf der Einerstelle und erreichen im Ideal jeweils einen Anteil von 10 %. Diese theoretische Hypothese wird empirisch durch das Benford-Gesetz ab der Drittziffer – mit nur noch marginalen Abweichungen⁷⁰ – bestätigt, weshalb eine Abgrenzung zu den [schwebenden](#) Erst- und Zweitziern erfolgen muss.⁷¹

Weil nur wenige praxisferne Absatzkonstellationen im Zusammenhang mit Bareinnahmen die freie Multiplikation (Erreichbarkeit aller Ziffern) stören, dient die Anwendung der Gleichverteilung zur Suche nach menschlichen Ziffernvorlieben (s. [Manipulationspsychologie](#)) in Folge bestimmter Formen von Einnahmeverkürzungen (s. [Zifferanalyse](#)). Wegen der Abhängigkeit derartiger [musterbasierter Methoden](#) von der untersuchten Datenmenge⁷² erfordert die seriöse Ergebnisbeurteilung einen [Wahrscheinlichkeitstest](#), wozu als gängige Verfahren der „[Chi-Quadrat-Test](#)“⁷³ und die „[Binomialverteilung](#)“ in Betracht kommen.

Gleitschlittenzeitreihe

Sog. Gleitschlitten- oder rollierende⁷⁴ Zeitreihen sind Analyse- und Prüfmittel, bei denen die chronologische Wiedergabe von Betriebsdaten bzw. -kennzahlen gleitend erfolgt: z.B. wird bei der im BFH-Zeitreihenurteil⁷⁵ verhandelten Variante der 10-Wochen-Aufschlagsatz um jeweils eine Woche fortgeführt, so dass auf das Intervall „Woche 1 bis 10“ die Perioden „Woche 2 bis 11“ und „Woche 3 bis 12“ folgen.

Ziel dieser besonderen Zeitreihendarstellung ist es, längere, gegenüber alltäglichen Störfaktoren robustere⁷⁶ Zeitreihenperioden mit einer größeren Anzahl an Beobachtungspunkten zu verknüpfen, um einen empfindlicheren Detektor für relevante Abweichungen als Beratungs- bzw. Prüffelder zu erhalten. Folgerichtig wird diese Zeitreihenform in der [Summarischen Risikoprüfung](#) (SRP) (zentral) für das zügige Auffinden von Einzelprüffeldern verwendet.⁷⁷ So sind dem nebenstehenden Diagramm einer tageweise gleitenden 2-Monats-Aufschlagsatzzeitreihe zwei sprunghafte Wertänderungen zu entnehmen: Sowohl beim Ein- (Intervall „1.3. bis 30.4.“) als auch beim Austritt (Intervall „1.5. bis 30.6.“) eines großen Lagereinkaufs am 30.4. wird die Periodenmarge abrupt verändert. Nach Prüfung des Einzelvorgangs kann dieser für die weitere Gesamtbildprüfung in der zeitlichen und/oder Wertzuordnung korrigiert werden ([interaktive Sachaufklärung](#)).



Wie alle Aufschlagsatzzeitreihen sind auch (aufschlagsatzbezogene) [Gleitschlittenzeitreihen](#) für eine Schlüssigkeitprüfung im Sinn des § 158 AO nur mäßig geeignet,⁷⁸ bildeten aber bereits die Grundlage mehrerer finanzgerichtlicher Auseinandersetzungen mit der Zeitreihenanalyse.⁷⁹ Im Zeitreihenur-

⁷⁰ Die Benford-Verteilungen der Dritt- und Viertziffer weisen als maximale Differenzen zur Gleichverteilung 10,18 % (Ziffer „0“ an der Drittposition) bzw. 10,02 % (Ziffer „0“ an der Viertposition) sowie 9,83 % (Ziffer „9“ an der Drittposition) bzw. 9,98 % (Ziffer „9“ an der Viertposition) auf (s. Abb. zum Begriff [„Benford-Verteilung“](#)).

⁷¹ Aus dem Grund der widerstreitenden Referenzen „Benford-Gesetz“ und „Gleichverteilung“ werden bei der Einerstellenuntersuchung der [Summarischen Risikoprüfung](#) (SRP) Datensätze < € 100 automatisch ausgeschlossen, was im Hinblick auf den absoluten Zufall der Gleichverteilung keine Verfallschungen zur Folge hat, aber die besondere Empirie der Wirtschaftsdaten (Benford-Gesetz) beachtet.

⁷² Die Bedeutung der untersuchten Datenmenge für die (natürliche) Ergebnisstreuung um eine Idealerwartung ([Referenz](#)) herum kann aus der Abb. zum Begriff [„Binomialverteilung“](#) abgelesen werden.

⁷³ Die – häufig festzustellende – Gleichsetzung des Chi-Quadrat-Tests mit der Erwartung „Gleichverteilung“ auf der Einerstelle von Bareinnahmen verkennt die Funktion der [Wahrscheinlichkeitstheorie](#), qualifizierte Übereinstimmungstests für bestimmte Vergleichssituationen zu stellen: So eignet sich der Chi-Quadrat-Test für alle Überprüfungen, ob eine mehrklassige (≥ 3) Beobachtung mit einer entsprechend [klassifizierten](#) Erwartung übereinstimmt.

⁷⁴ Harle/Olles, „Die moderne Betriebsprüfung“, NWB Verlag Heme, 3. Auflage 2014, Rz. 1366 ff.

⁷⁵ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743.

⁷⁶ Unter der „Robustheit“ wird verstanden, dass keine bzw. nur eine geringe Anfälligkeit gegenüber alltäglichen Sachverhalten besteht.

⁷⁷ Wähnert, „Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz „Summarische Risikoprüfung (SRP)““, DB 45/2016, 2627.

⁷⁸ Eine Gegenüberstellung von Faktorzeitreihen – wie Aufschlagsatzzeitreihen –, einseitig skalierten und doppelskalierten Zeitreihenvergleichen mit einer Bewertung anhand des [Korrelationskoeffizienten](#) – als statistischem Parameter für die Größenabhängigkeit – findet sich unter [„Zeitreihenanalyse“](#).

⁷⁹ So z.B. FG Münster v. 19.8.2004 – 8 V 3055/04, FG Düsseldorf v. 15.2.2007 – 16 V 4691/06 und FG Köln v. 27.1.2009 – 6 K 3954/07.

teil des BFH wird der Variantenreichtum von Zeitreihen erkannt,⁸⁰ während das FG Hamburg⁸¹ in einem Beschluss den Unterschied zu [doppelskalierten Zeitreihenvergleichen](#) thematisiert.

Gutachten

⇒ [Chancengleichheit](#)

Häufigkeitsanalyse

Häufigkeitsanalysen ordnen eine unübersichtlich große Datenmenge für Prüfungs- und Beratungszwecke anhand der Auszählung nach skalierbaren Eigenschaften wie dem [klassifizierten](#) Wertinhalt, im Aufzeichnungssystem vergebenen Schlüsseln (z.B. Umsatzsteuerschlüssel) etc.

Als einfache Form der Häufigkeitsuntersuchung ist die [ABC-Analyse](#) ein verbreitetes Instrument der Wirtschaftsprüfung. Durch Einführung der digitalen Aufzeichnung und Datenanalyse haben sich die Möglichkeiten sowie die Geschwindigkeit der Datenauswertung wesentlich verbessert, wodurch die Verbildlichung von Häufigkeitsverteilungen mehr Bedeutung gewonnen hat, denn größere Datenmengen neigen dazu, allgemeine oder betriebsbezogene [regelmäßige Muster](#) auszubilden. Diese können zum wirksamen, zügigen Auffinden von Prüf- bzw. Beratungsfeldern sowie zur Aufdeckung von [menschlichen Eingriffen](#) eingesetzt werden.

Hypergeometrische Verteilung

Mit der [diskreten](#) Hypergeometrischen Verteilung wird das Streuverhalten für (dichotome) ja-nein-Zufallsereignisse wiedergegeben, bei denen sich – im Gegensatz zur [Binomialverteilung](#) – die Wahrscheinlichkeit durch „Nicht-Zurücklegen“ fortlaufend verändert.

Als klassisches Erklärungsmodell dient das Urnenexperiment mit zwei Kugelarten: Die erste Kugel der Farbe „A“ wird nach dem Häufigkeitsverhältnis der beiden Kugelarten gezogen.⁸² Weil die erste Kugel nicht wieder in die Urne zurückgelegt wird, ändert sich die Ereigniswahrscheinlichkeit für die zweite Kugel in: $\frac{A \text{ bzw. } (A-1)}{(A+B-1)}$, je nachdem welche Kugelart der erste Versuch ergeben hat. Die re-

sultierenden Häufigkeitsverteilungen lassen sich über die [Kombinatorik](#) auszählen oder mit der Funktion der Hypergeometrischen Verteilung (z.B. in Tabellenkalkulationsprogrammen) berechnen. Wie bei der Binomialverteilung ist die deutliche Abhängigkeit von der Datenmenge typisch, indem vergleichbar stark abweichende Ergebnisse bei größerem Datenumfang viel unwahrscheinlicher sind (s. Abb. zum Begriff „[Binomialverteilung](#)“).

Praktische Prüfungsbedeutung hat die Hypergeometrische Verteilung dadurch, dass das Ziehen und Bewerten von Stichproben ebenfalls der beschriebenen Charakteristik gehorcht. Deshalb liefert die Hypergeometrische Verteilung die korrekten Zusammenhänge zur Planung und Hochrechnung von qualifizierten (mathematisch-statistischen) Stichprobenprüfungen, wie sie als [Gesamtbildmethode](#) anerkannt sind.⁸³ Aus Vereinfachungsgründen arbeiten einige Anwendungstools des in der Wirtschaftsprüfung verbreiteten „[Monetary Unit Sampling](#) (MUS)“ mit der Poisson- bzw. der Binomialverteilung, weil die kumulierte Funktion der Hypergeometrischen Verteilung in Standardsoftware noch nicht so lange üblich ist.

Hypothese

Mit (Null-)Hypothese wird eine Annahme bezeichnet, die durch Überprüfung zu bestätigen oder zu verwerfen ist. Auf die (steuerliche) Prüfung übertragen ist damit die Erwartung gemeint, die an einen Sachverhalt und dessen Dokumentation im Hinblick auf die rechtliche Würdigung geknüpft ist, womit

⁸⁰ BFH v. 25.3.2015 – XR 20/13, BStBl II 2015, 743, Rz. 37; mit einem klareren Bezug des Urteils: RiBFH Nöcker, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse“, NWB 48/2015, 3548.

⁸¹ FG Hamburg v. 31.10.2016 – 2 V 202/16, Rz. 48: „Mit Hilfe der doppelskalierten Monatszeitreihe hat die Betriebsprüfung die Daten des Wareneinkaufs den Umsatzerlösen in zwei Einzelgrafiken gegenüber gestellt. Bei dieser Methode tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander.“

⁸² Wie bei allen Zufallsexperimenten wird im einfachsten Erklärungsmodell davon ausgegangen, dass beide Kugelarten mit derselben Wahrscheinlichkeit „p“ gezogen werden. Sowohl die Binomial- als auch die Hypergeometrische Verteilung können aber auch jede andere Einzelwahrscheinlichkeitsverteilung verarbeiten.

⁸³ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff., § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden“, IDW PS 310 „Repräsentative Auswahlverfahren (Stichproben) in der Abschlussprüfung“, s.: <https://www.idw.de/idw/verlautbarungen/idw-ps-310/43204>, DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/de/norm/din-iso-2859-1/191697831>.

eine Prüfungsmaßnahme dem allgemeinen Duktus „Erwartung ⇒ Beobachtung ⇒ Übereinstimmungsprüfung“ folgt.

Bereits die (Vergangenheits-)Prüfung eines Einzelvorgangs wird trotz sorgfältiger Analyse der dazugehörigen Unterlagen i.d.R. ein gewisses Maß an Unklarheit hinterlassen, weil eine vollständige, objektive Dokumentation nahezu unmöglich ist. Deshalb erfolgt eine Prüfungsbewertung regelmäßig als Wahrscheinlichkeitsaussage ([probabilistische Schlussfolgerung](#)). Bei der Prüfung von komplexen (datenintensiven) und/oder möglicherweise verschleierte Sachverhalten per [Gesamtbildprüfung](#) gem. § 158 AO wird das Maß an Ungewissheit grundsätzlich noch deutlich größer sein, so dass der sachgerechten, nachvollziehbaren Einordnung des indiziellen Prüfungsergebnisses zwischen [Beweis](#), [Beanstandungsanlass](#) und bloßer Möglichkeit eine zentrale Bedeutung zukommt.⁸⁴

Idealerwartung

⇒ [Referenz](#)

IKS

⇒ [Internes Kontrollsystem](#)

Indiziengesamtheit

Der BGH hat festgestellt, dass „die Indizien ... in ihrer Gesamtheit ... die entsprechende Überzeugung vermitteln [können], auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreich[en]“.⁸⁵

Diese Erkenntnis zur erhöhten Beweiskraft von Indiziengesamtheiten geht auf die mathematischen Gesetzmäßigkeiten zurück, dass das gemeinsame Auftreten ([Koinzidenz](#)) von unwahrscheinlichen Einzelereignissen (Indizien) zu einer (deutlich) reduzierten Wahrscheinlichkeit der Gesamtsituation führt. Neben der multiplikativen Verstärkung unabhängiger Indizien, die an der [Kombinatorik](#) bildlich nachvollzogen werden kann, gilt das in verringertem Maß i.d.R. auch für (evtl.) abhängige Beweisanzeichen.⁸⁶

Das Beweisprinzip der Indiziengesamtheiten hat besonders in der – grundsätzlich indiziellen – [Gesamtbildprüfung](#) mit (digitalen) Prüfungsnetzen wie der [Summarischen Risikoprüfung](#) (SRP) an Bedeutung gewonnen, weil die Ablehnung der Beweiskraft der Buchführung (bzw. Aufzeichnungen) nach § 158 AO regelmäßig auf einer Mehrzahl von Indizien fußt.⁸⁷ Mit der gemeinsamen Bewertung von formellen und materiellen Mängeln hat diese generelle Beweisführungsregel tlw. Einzug in die Finanzrechtsprechung gehalten.⁸⁸

Induktive Statistik

⇒ [Statistik](#)

Interaktive Sach(verhalts)aufklärung

Mit der interaktiven Sachaufklärung bietet die digitale Außenprüfung – besonders bei [vorlagenbasierten](#) Prüfungsnetzen wie der [Summarischen Risikoprüfung](#) (SRP) – die Möglichkeit, neue oder korrigierte Erkenntnisse auch bei komplexen [Gesamtbildprüfungen](#) einzupflegen, wonach umgehend die aktualisierte (indizielle) Prüfungssituation zwischen Aufklärung, Möglichkeit und [Beanstandungsanlass](#) vorliegt.

Im Sinn der gemeinsamen Sachaufklärungsverantwortung⁸⁹ können dadurch (nahezu) alle konkreten Informationen in einen laufenden Prüfungsprozess eingebunden werden: Vom Ausschluss fehlerhaft einbezogener Vorgänge über die nachträgliche Berücksichtigung sowie (Teil-)Korrekturen von Werten und Datumsbezügen bis hin zur Anpassung kompletter Datenreihen.

⁸⁴ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Luchterhand Verlag München, 2004.

⁸⁵ BGH v. 27.4.2010 – 1 StR 454/09, NStZ 2011, 108, Rz. 18 ff.

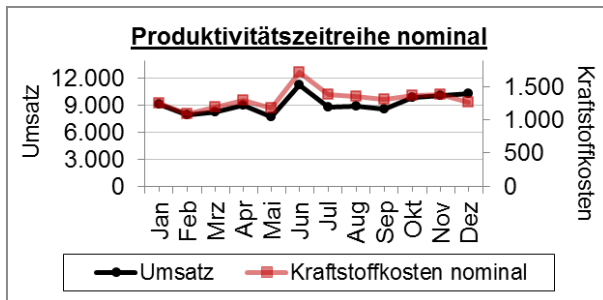
⁸⁶ Lediglich für die Situation, dass zwei Indizien vollständig abhängig sind, gilt, dass die Wahrscheinlichkeit des gemeinsamen Auftretens nicht geringer ist als die für eines der beiden Einzelbeweisanzeichen.

⁸⁷ Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Strafverfahren“, NZWiSt 1/2014, 24, Webel/Wähnert, „Der Verdacht einer Steuerstraftat während der Außenprüfung – § 10 BpO im Kontext der Digitalisierung“, NWB 44/2014, 3324 sowie Webel/Wähnert, „Summarische Risikoprüfung: Einleitung von Strafverfahren in der digitalen Welt“, PStR 3/2015, 70.

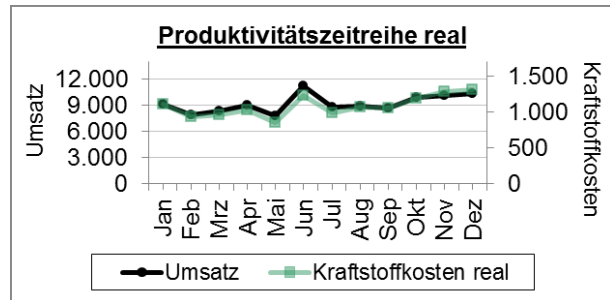
⁸⁸ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 60 ff. und RIBFH Nöcker, „Update Betriebsprüfung – Neues vom BFH und die Auswirkungen für die Praxis“, NWB 42/2016, 3157 zum Drei-Stufen-Modell für die sachgerechte Auswahl einer Schätzungsmethode, wobei das Zusammenfallen von formellen und materiellen Mängeln die Anwendung größerer Schätzungen rechtfertigt.

⁸⁹ Seer, Tipke/Kruse, AO/FGO, § 158 AO Rz. 14.

Der auf der linken Seite abgebildete [doppelskalierte Zeitreihenvergleich](#) aus Erlösen eines Taxiunternehmens und den Kraftstoffkosten (als wertrelevanter, variabler Kostenfaktor) leidet – in Bezug auf die [korrelierende](#) Handlungslogik gem. § 158 AO – unter der Beförderungspreisbindung (per Verordnung) bei gleichzeitig starken Preisschwankungen am Kraftstoffmarkt. Durch die – in der Volkswirtschaftslehre übliche – Herausrechnung der konjunkturellen Preiseinflüsse mittels Preisindizes (auf ein⁹⁰ Basisniveau) werden die Marktschwankungen eliminiert, so dass im



Zeitreihendiagramm auf der rechten Seite (quasi) eine Verprobung der Mengenabhängigkeit wiedergegeben wird; sofern die Beförderungspreise im betreffenden Zeitraum ebenfalls neu festgesetzt wurden, ist hier eine entsprechende Korrektur notwendig. Im Vergleich zur nominalen Korrelationsverprobung ist der realen eine etwas bessere [Kongruenz](#) beider Grafen zu entnehmen, womit eine (noch) klarere Handlungslogik von Mitteleinsatz und Leistung resultiert. Als wesentlicher Unterschied ist das nominale Absinken der Kraftstoffkosten im Dezember offensichtlich auf einen Preisrückgang zurückzuführen.



den die Marktschwankungen eliminiert, so dass im

Internes Kontrollsystem (IKS)

Bereits in den „Grundsätze[n] ordnungsmäßiger EDV-gestützter Buchführungssysteme (GoBS)“⁹¹ wird ein internes Kontrollsystem als ein Kriterium für die Einhaltung der (formellen) Ordnungsmäßigkeit (s. [Richtigkeitsvermutung](#)) digitaler Besteuerungsaufzeichnungen aufgeführt, weil die (weitreichende) elektronische Automatisierung der Aufzeichnungs- und Aufbewahrungsprozesse überwacht werden muss.

Wesentliche IKS-Bestandteile werden in den GoBD⁹² beispielhaft mit „Zugangs- und Zugriffsberechtigungskontrollen, auf Basis entsprechender Zugangs- und Zugriffsberechtigungskonzepte ..., Funktionstrennungen, Erfassungskontrollen (Fehlerhinweise, Plausibilitätsprüfungen), Abstimmungskontrollen bei der Dateneingabe, Verarbeitungskontrollen, Schutzmaßnahmen gegen die beabsichtigte und unbeabsichtigte Verfälschung von Programmen, Daten und Dokumenten“ aufgelistet, um fehlerhafte Systembedienungen oder -funktionen zu verhindern (s.a. [Schnittstellenverprobung](#)).

Interpolation

Als Ermittlung von Zwischenwerten zu [diskreten](#) Daten ist eine Interpolation bei relativen Rängen wie [Quantilen](#) immer dann notwendig, wenn der berechnete Rang nicht ganzzahlig ist.

Obwohl es mehrere Interpolationsverfahren zur [stetigen](#) Annäherung diskreter Datenreihen gibt, wird bei den Standardverfahren zur Berechnung von „Zwischenrängen“ wie in Tabellenkalkulationsprogrammen i.d.R. eine lineare Form eingesetzt. Die [Summarische Risikoprüfung](#) (SRP) greift im Rahmen ihrer Schätzungsoption „[Quantilsschätzung](#)“ ebenfalls auf diese verbreitete Interpolationsweise mittels Quantilsfunktionen zurück.

Iterationsanalyse

Ähnlich der [Ziffern-](#), Struktur- und [Verteilungsanalyse](#) gehört die Iterationsanalyse zu den [musterbasierten Verfahren](#), indem das serielle Wiederholungsverhalten in Besteuerungsdaten auf dessen Übereinstimmung mit einer allgemein oder betriebsspezifisch bekannten Verteilung überprüft wird.

⁹⁰ Werden bei einem auf Abhängigkeit (Korrelation) angelegten (doppelskalierten) Zeitreihenvergleich die Preisveränderungen mit Hilfe von Preisindizes eliminiert, ist der Bezugszeitpunkt unerheblich: Der davon beeinflusste Verhältnisfaktor wird über die größenanpassende Doppelskalierung ohnehin aus der Gegenüberstellung entfernt, um die erwartete Mengenschlüssigkeit erkennen zu können.

⁹¹ BFM-Schreiben v. 7.11.1995, BStBl 1995 I, 738, Pkt. IV.

⁹² „Grundsätze zur ordnungsgemäßen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD)“, BMF-Schreiben v. 14.11.2014, BStBl I 2014, 1450, Pkt. 6.

Bezogen z.B. auf das Ziffernvorkommen an der [festen](#) Einerstelle manipulationsverdächtiger Bar-einnahmen werden hierbei nicht die Häufigkeiten der möglichen Ziffern sondern die chronologischen Datensatzabstände zu denselben Ziffern ausgezählt, also zu jeder Ziffer die jeweilige Tagesanzahl bis sie erneut vorkommt.

Diese Methode, nach bestimmten, aus der Praxis bekannten⁹³ Steuerhinterziehungsszenarien gezielt zu suchen, nutzt zum einen die Kenntnis, dass zufällige serielle Datenfolgen ein anderes Wiederholungsverhalten aufweisen, als Zufallsgeneratoren bei softwaregestützter Manipulation erzeugen. Zum anderen folgt sie der Effizienzmaßgabe aus § 7 BpO die „Dauer [einer Betriebsprüfung] ... auf das notwendige Maß zu beschränken“.

Kalkulation

⇒ [Aufschlagskalkulation](#) sowie [Ausbeutekalkulation](#)

Klassifizierung

Grundlage einer [Häufigkeitsanalyse](#) größerer Datenmengen z.B. zur [Mustererkennung](#) ist die Einteilung in Klassen (Klassifizierung).

Bei [diskreten](#) Datenmerkmalen können diese direkt als Klassen herangezogen werden, wenn bspw. bei der [Ziffernanalyse](#) nach jeder möglichen Ziffer ausgezählt wird. Besonders für gezielte Untersuchungsrichtungen wie bei einer Ziffernanalyse aus der Perspektive einer bestimmten Ziffer können Datenmerkmale auch in Gruppen zusammengefasst werden.⁹⁴

[Stetige](#) Datenmerkmale wie z.B. der Wertinhalt erfordern eine Festlegung von Klassengrenzen (Schichtober- und Schichtuntergrenzen). Im Rahmen musterbezogener Analysen werden die Differenzierungsklassen i.d.R. entweder linear oder [logarithmisch](#) gleichmäßig definiert. Die [Summarische Risikoprüfung](#) (SRP) greift hierfür auf die – aus der deskriptiven [Statistik](#) bekannte und in der Betriebsprüfungspraxis bewährte – Wurzelmethode zurück: Der zu klassifizierende lineare oder logarithmische Zahlenbereich wird in $\sqrt{\text{Datenanzahl}}$ aber – aus Übersichtsgründen – max. 30 Intervalle unterteilt.

Koinzidenzanalyse

Mit Hilfe von Koinzidenzanalysen sollen Chancen und Risiken aus den Erfahrungen der Vergangenheit zu unterschiedlichen Faktorvariationen eingeschätzt werden.

Dazu wird eine mehrdimensionale Matrix aus allen relevanten Vorgangsmerkmalen (z.B. Umsatzsteuerschlüssel) mit deren möglichen Ausprägungen gebildet. Die zur Verfügung stehenden Daten werden (elektronisch) im Sinn der Matrix ausgewertet, um ein besonders erfolgversprechendes Zusammentreffen (Koinzidenz) von Faktoren anzustreben bzw. vor einem kritischen gewarnt zu werden.

Einige professionelle Datenanalysesysteme basieren auf dieser Herangehensweise – vermutlich auch, weil hiermit eine fast vollständige Automatisierung von Selektionsprozessen möglich wird. Die Grenzen liegen zum einen darin, dass wesentliche Erfolgs- bzw. Risikofaktoren beim Erstellen der Matrix nicht bekannt sind, und zum anderen in der fehlenden Auseinandersetzung des Prüfungspersonals mit der Datenqualität: Nach den Erkenntnissen der [Fuzzy Logic](#) ist die Summationsebene von per [Klassifizierung](#) skalierten ([stetigen](#)) Informationen entscheidend dafür, ob [Regeln](#) und [Muster](#) bzw. deren Störungen erkannt werden können.

Kombinatorik

Das mathematische Teilgebiet der Kombinatorik beschäftigt sich u.a. mit den möglichen Ergebnisvariationen (Permutationen) in [diskreten](#) Datenverteilungen so wie bspw. mit den Wahrscheinlichkeitsverhältnissen beim Würfeln.

In nebenstehender Grafik werden die möglichen Kombinationen dreier aufeinander folgender Tagesbareinnahmen aus der Sicht einer bestimmten Ziffer ([Ziffernanalyse](#) zur Referenz „[Gleichverteilung](#)“) mit den jeweiligen Wahrscheinlichkeiten dargestellt: Die Trefferwahrscheinlichkeit für die eine Ziffer (z.B. „4“) beträgt bei der Gleichverteilung 10 %, so dass die Nicht-Treffer-Wahrscheinlichkeit für

⁹³ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 27: „Elektronische Kassensysteme sind durch Umprogrammierung in nahezu beliebiger Weise manipulierbar; von derartigen Manipulationsmöglichkeiten machen Teile der betrieblichen Praxis nach dem Erkenntnisstand des Senats durchaus Gebrauch ...“.

⁹⁴ Bei der Ziffernanalyse aus der Sicht einer einzelnen (unterbewusst bevorzugten bzw. gemiedenen) Ziffer bildet die fokussierte eine Klasse, wohingegen alle anderen möglichen Ziffern zu einer gemeinsamen Klasse gruppiert werden.

die zusammengefassten anderen Ziffern 90 % ergibt. Über den Ergebnisbaum werden die $2 \cdot 2 \cdot 2 = 2^3 = 8$ möglichen Konstellationen verbildlicht. Die Eintrittswahrscheinlichkeit jeder einzelnen Permutation errechnet sich nach Laplace als Produkt aus den Einzelwahrscheinlichkeiten⁹⁵, also für drei Treffer mit: $10 \% \cdot 10 \% \cdot 10 \% = 0,1 \%$; im Umkehrschluss ist es unter Annahme der Gleichverteilung zu 99,9 % unwahrscheinlich, dass die fokussierte Ziffer in allen drei Bareinnahmewerten vorkommt. Soll die Wahrscheinlichkeit für eine Kombination mit mehreren möglichen Ausprägungen wie z.B. bei zwei Treffern ermittelt werden, sind die Einzelwahrscheinlichkeiten wie in der rechten Spalte aufzusummieren. Analog lassen sich Wahrscheinlichkeitsbereiche wie bspw. „mindestens ein Treffer“ per Addition auswerten.

Kombinatorik			P _(x)	P _(2 Treffer)
1. Wert	2. Wert	3. Wert		
		Treffer (10 %)	0,1%	nein
	Treffer (10 %)	Fehler (90 %)	0,9%	0,9%
Treffer (10 %)		Treffer (10 %)	0,9%	0,9%
	Fehler (90 %)	Fehler (90 %)	8,1%	nein
		Treffer (10 %)	0,9%	0,9%
	Treffer (10 %)	Fehler (90 %)	8,1%	nein
Fehler (90 %)		Treffer (10 %)	8,1%	nein
	Fehler (90 %)	Fehler (90 %)	72,9%	nein
			100,0%	2,7%
<i>Binomialverteilung</i>				
Werte:			3	
Treffer:			2	
p _{Treffer} :			10%	
kumuliert:			nein	2,7%

[Wahrscheinlichkeitsverteilungen](#) ermöglichen mit Hilfe einer einzigen Funktion die Berechnung auch deutlich komplexerer Fragestellungen, wie sie z.B. in der Ziffernanalyse für mehrere hundert Besteuerungsdaten auftreten. In der Beispielsabbildung ergänzt der graue Bereich die Wahrscheinlichkeitsergebnisse der einschlägigen [Binomialverteilung](#) für „exakt zwei Treffer“, was über die Zusatzinformation „(nicht) kumuliert“ angesteuert wird.

Komplettmanipulation

„Komplettmanipulation“ meint das freie Erfinden von Besteuerungsgrundlagen.

Aus der prädigitalen Zeit ist vor allem die (ordnungswidrige) Aufzeichnungsweise des händischen Eintragens von täglichen Bareinnahmen in ein Kassenbuch bekannt, wobei anstelle der tatsächlichen Werte frei erfundene Zahlen für eine niedrigere Besteuerung angesetzt werden. Diese verbreitete Steuerhinterziehungsweise findet sich auch in der elektronischen Betriebsverwaltung z.B. mit dem Double-Till-Szenario: Eine Kasse wird zur korrekten Aufzeichnung für die Personalkontrolle verwendet, während in eine typgleiche zweite Kasse im „Backoffice“ Wunschbeträge eingegeben und die begrenzten Informationen durch reduzierte Berichtsausgaben verschleiert werden.⁹⁶

Aus [manipulationspsychologischer](#) Sicht ist das freie Erfinden von Besteuerungsdaten ein geeigneter Anknüpfungspunkt, weil dabei unbewusste Ziffernvorlieben und -abneigungen⁹⁷ derart deutlich hervortreten können, dass sich mit einer [sehr geringen Restunsicherheit](#) der menschliche Eingriff aufdecken lässt. Ähnliche Steuerhinterziehungsarten wie z.B. das vorherige Bezahlen von (unversteuerten) „Schwarzlöhnen“ mit der anschließenden Eintragung der verbliebenen Bareinnahmen sind davon abzugrenzen, weil hierbei keine Ziffernspuren entstehen können.

Konfidenzniveau

Mit dem Konfidenzniveau wird die (festgelegte) Gewissheit bezeichnet, mit der die Bestätigung oder Ablehnung ([Abweisungswert](#)) einer Aussage ([Hypothese](#)) erfolgt bzw. erfolgen soll. Im Zuge einer Qualitätsaussage soll z.B. mit der Sicherheit von x % eine Produkteigenschaft bestätigt werden.⁹⁸ Bei der ablehnenden Form ist bspw. eine Klarheit von x % notwendig, um mit einem Manipulationstest wie der [Ziffernanalyse](#) einen [Beanstandungsanlass](#) nach § 158 AO zu begründen.

Konfidenzmaße werden im Rahmen der [Wahrscheinlichkeitslehre](#) verwendet, um – auf Basis der empirischen und theoretischen Erfahrungen der Mathematik – [statistisch](#) ausreichend eindeutige Aus-

⁹⁵ Die Voraussetzung für eine Gesamtwahrscheinlichkeitsberechnung nach Laplace, dass es sich um unabhängige Ereignisse handelt, ist bei den Tagessummen von Bareinnahmen, welche aus einer Vielzahl von Einzelvorgängen zusammengesetzt sind, erfüllt. Ein möglicherweise begrenzter Ergebnisraum z.B. durch Einzelpreise, die ausschließlich auf „€ #.##5“ enden, lässt sich in der Kombinatorik berücksichtigen.

⁹⁶ Teutemacher, „Handbuch zur Kassenführung“, NWB Verlag Herne, 2015, 137 ff. mit einer Offenlegung einiger verbreiteter Manipulationsmethoden.

⁹⁷ Hill, „Random-number guessing and the first digit phenomenon“, Psychological Reports 62/1999, 967; Hill, „The difficulty of faking data“, Chance 12/3/1999, 27.

⁹⁸ DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/de/norm/din-iso-2859-1/191697831>.

sagequalitäten für [Übereinstimmungstests](#) sowie [Stichprobenverfahren](#) zu definieren: Die beiden verbreitetsten Konfidenzniveaus für die indizielle und einzelbeweisfähige Gewissheit sind 95 %⁹⁹ bzw. 99,5 %¹⁰⁰.

Konjunkturbereinigung (Konjunkturanpassung)

⇒ [Interaktive Sach\(verhaltens\)aufklärung](#)

Kongruenz

Kongruente Zeitreihengrafen liegen deckungsgleich übereinander.

Während bei der Übereinstimmungsprüfung zweier Datenreihen per Zeitreihenvergleich (z.B. zur Überprüfung von [Schnittstellen](#) innerhalb einer Betriebsverwaltung) beide Betriebsfaktoren dafür derselben Größenskalierung (Y-Achse) folgen müssen (einseitig skaliertes [Zeitreihenvergleich](#)), erfordert eine Abhängigkeitsprüfung ([Korrelation](#)) die Anpassung auf dasselbe Wertniveau bspw. mit Hilfe der [Doppelskalierung](#). Der Variantenreichtum in der Ausgestaltung eines Zeitreihenvergleichs¹⁰¹ führt mit hin dazu, dass Beobachtungen wie „Kongruenz“ oder „Parallelität“¹⁰² ganz unterschiedliche Analyseergebnisse bedeuten können.¹⁰³

Korrelation

Unter „Korrelation“ wird in der [Statistik](#) die Abhängigkeit zweier Daten(reihen) verstanden.

Bezogen auf den [Zeitreihenvergleich](#)¹⁰⁴ steht die Korrelation für die Güte der erwarteten¹⁰⁵ Abhängigkeit zweier Größenentwicklungen im Sinn der Handlungslogik: Verändert sich ein relevanter Betriebsfaktor wie z.B. der Absatz, werden die Handelnden im Unternehmen mit einer gleichgerichteten, möglichst adäquaten Anpassung variabler Faktoren gegensteuern, um Erfolgsmaximierung anzustreben. Bei einigen Betriebsgrößen wie insbesondere Einsatz und Verkauf von Waren bzw. Leistungen stellt die Wechselwirkung eine operative Zwangsläufigkeit dar, während sie bspw. bei variablen Kosten eine Schlüssigkeitsannahme (an die vernünftige Unternehmensführung) ist, und entdeckte Implausibilitäten deshalb einen [Beanstandungsanlass](#) gem. § 158 AO darstellen können.

Korrelationskoeffizient

Der statistische Indikator für die Abhängigkeit ([Korrelation](#)) zweier Datenreihen wie z.B. beim [Zeitreihenvergleich](#) ist der Korrelationskoeffizient, der u.a. in Tabellenkalkulationsprogrammen allgemein zur Verfügung steht.

Als lineares Zusammenhangsmaß nimmt der Korrelationskoeffizient Größen von „-1“ für die 100 %-ige Gegenläufigkeit der Grafen über „0“ bei nicht vorhandener Abhängigkeit bis zu „+1“ als Anzeichen für eine vollständige positive Beeinflussung ein. Besonders mit einer [vorlagengesteuerten](#) zeitlichen Eingrenzbarkeit wie bei der [Summarischen Risikoprüfung](#) (SRP) ist er eine geeignete Ergänzung für die verbildlichte Schlüssigkeitsprüfung zweier als abhängig erwarteter Betriebsfaktoren mittels [doppelskaliertem](#) Zeitreihenvergleich.¹⁰⁶

Kybernetische Sensitivitätsanalyse

⇒ [Sensitivitätsmodell nach Vester](#)

Liquiditätsrechnung

⇒ [Geldverkehrsrechnung](#) sowie [Vermögenszuwachsrechnung](#)

⁹⁹ Das Konfidenzmaß „95 %“ leitet sich aus den Rang-Wert-Verhältnissen der [Normalverteilungen](#) ab: Es gibt den Normalbereich zuzüglich schwacher Ausreißer mit „Mittelwert ± doppelte Standardabweichung“ wieder.

¹⁰⁰ So auch: BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff. Nach wissenschaftlichen Einschätzungen entspricht das Maß „99-99,5 %“ der Eindeutigkeit kriminalistischer Beweise: *Giezek/Rupprecht/Wähnert*, „Wie sicher muss ein Verprobungsergebnis sein? – Anleihen aus der Kriminalistik zur sinnvollen Einordnung des Beanstandungsanlasses“, BBK 5/2017, 236, mit diversen Quellen.

¹⁰¹ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 37.

¹⁰² Der parallele Verlauf zweier Zeitreihengrafen abhängiger Betriebsfaktoren in einem einseitig skalierten Zeitreihendiagramm entspricht nicht der [Ideal-erwartung](#), denn die [Hypothese](#) der Korrelation ist – im Gegensatz zur Datenreihenübereinstimmung – ein (nahezu) gleichmäßiges Verhältnis: Erhöht sich bspw. der Wareneinsatz um € 10.000 soll der Umsatz um diesen Wert zuzüglich des Aufschlagsatzes ansteigen.

¹⁰³ Eine Gegenüberstellung dreier weit verbreiteter Arten von Zeitreihenanalysen und ihrer Ergebnisse zu demselben Datenbestand findet sich unter dem Begriff „[Zeitreihenanalyse](#)“.

¹⁰⁴ Mit Beispieldiagrammen unter dem Stichwort „[Zeitreihenanalyse](#)“.

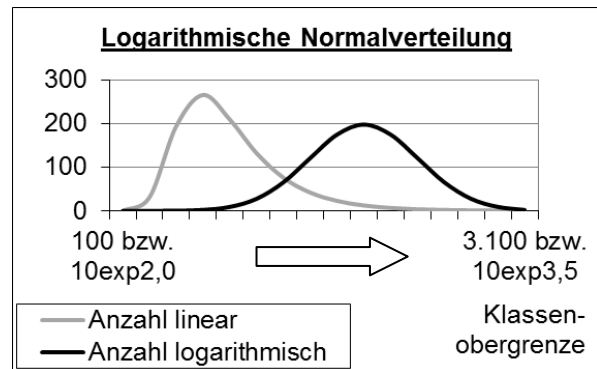
¹⁰⁵ Die statistisch oder grafisch bestätigte Korrelation zweier Daten(reihen) belegt nicht deren kausalen Zusammenhang. Die Ursache-Wirkungs-Beziehung von zwei untersuchten Betriebsfaktoren ergibt sich aus deren schlüssiger Auswahl für den Zeitreihenvergleich.

¹⁰⁶ Eine Gegenüberstellung unterschiedlicher Zeitreihenvarianten in Kombination mit dem Korrelationskoeffizienten zu demselben Datenbestand wird unter dem Begriff „[Zeitreihenanalyse](#)“ erläutert.

Logarithmische Normalverteilungen

Durch die [stetigen](#) Logarithmischen Normalverteilungen wird ein in Biologie, Physik und Wirtschaft verbreitetes¹⁰⁷ – von den [Normalverteilungen](#) abweichendes – Streuungsmuster für Fälle beschrieben, in denen der mögliche Wertebereich im unteren Bereich begrenzt ist wie z.B. beim Kraftstoffverbrauch von Automotoren wegen technischer Gegebenheiten.

Das abgebildete Beispiel für eine Logarithmische Normalverteilung zeigt die in linearer Auftragung¹⁰⁸ rechtsschiefe Struktur. Werden anstatt der Werte deren Logarithmen¹⁰⁹ linear nach Klassen ausgezählt bzw. dies über eine logarithmische X-Achsenkalierung¹¹⁰ ([Klassifizierung](#)) erreicht, wandelt sich der asymmetrische Typ in die Glockenform einer [Normalverteilung](#) – unabhängig von der verwendeten Basis. Logarithmische Normalverteilungen sind mit Hilfe ihrer Kenngrößen „[Mittelwert](#)“ und „[Standardabweichung](#)“ vollständig definiert, so dass jede Klassenhäufigkeit einzeln oder kumuliert berechnet werden kann. Deshalb eignen sie sich gut für [Musteranalysen](#).



Viele Wirtschaftsdaten wie bspw. Betriebsgrößen, Produktionsumfänge, Kapital, Bankeinlagen, Fabrikationskosten, Einkommen, Lagerpositionen, Umsätze etc.¹¹¹ sind auch zu den niedrigen Werten hin begrenzt und weisen regelmäßig logarithmisch-normalverteilte Häufigkeitsmuster auf. Schon vor Jahrzehnten wurden diese statistischen Kenntnisse für [Gesamtbildmethoden](#) in der Betriebsprüfung vereinzelt eingesetzt.¹¹² Besonders aus [manipulationspsychologischer](#) Sicht sind derartige Prüfungsansätze sinnvoll, weil die wahrscheinlichste Spur der Steuerhinterziehung „[Spitzenkappung](#)“ – auch unter Einsatz systematischer Manipulationstools – offengelegt werden kann¹¹³; die Abgrenzung von Auffälligkeiten gegenüber nicht-manipulativen menschlichen Eingriffen kann durch eine Prüfsystematik wie bei der [Verteilungsanalyse](#) der [Summarischen Risikoprüfung](#) (SRP) gewährleistet werden.

Manipulationspsychologie

In der „Psychologie Heute“ wird gefragt, ob sich aus „Studienergebnisse[n zur Steuerhinterziehung] eine Empfehlung für die Finanzämter ableiten“ ließe, denn „Psychologen sind dem Phänomen der Steuermoral auf der Spur“.¹¹⁴ Wie bei jeder Kriminalitätsbekämpfung sind auch zur Aufdeckung von Steuerhinterziehungen Kenntnisse zu verbreiteten Verbrechensabläufen notwendig. Aus psychologischen Studien und Theorien sind mindestens die folgenden drei Ansatzpunkte bekannt:

1. Das freie Erfinden von Besteuerungsgrundlagen hinterlässt häufig so starke Vorlieben oder Abneigungen für einzelne Ziffern in den Zahlen¹¹⁵, dass diese per [Ziffernanalyse](#) als hochwahrscheinliche Spur eines menschlichen Eingriffs identifiziert werden können.
2. Durch das software-unterstützte Umgehen von Ziffernauffälligkeiten per Zufallsgenerator wird eine andere (chronologische) Abfolge an Wiederholungen generiert, als sie aus der [Statistik](#) bekannt ist, weshalb die [Iterationsanalyse](#) ein erfolversprechender Prüfungsansatz ist.
3. Grundsätzlich führen Wiederholungen von Gesetzesbrüchen zur Herabsetzung der Hemmschwelle, was in Verbindung mit dem – auf Veränderungen reagierenden – Belohnungssystem im Gehirn i.d.R. zu (sprunghaften) Erhöhungen der Häufigkeit und Intensität führt.¹¹⁶ Bei der

¹⁰⁷ Limpert/Stahel/Abbt, "Lognormal distributions across the sciences: keys and clues", BioScience May 2001, Vol. 51, № 5, 341.

¹⁰⁸ Eine lineare Unterteilung der X-Achsen zur Beispielsgrafik könnte z.B. sein: „100, 300, 500, 700, ..., 3.100“.

¹⁰⁹ Der Logarithmus sucht den Exponenten zu einer bestimmten Basis, um den Bezugswert auszudrücken wie bspw. $10^x = 1.000 \Rightarrow x = 3$.

¹¹⁰ Eine logarithmische Unterteilung der X-Achse zur Beispielsgrafik könnte z.B. sein: „ $10^{2,0}$, $10^{2,1}$, $10^{2,2}$, $10^{2,3}$... $10^{3,5}$ “; mit den dazugehörigen linearen Klassenobergrenzen „100, ca. 126, ca. 158, ca. 200, ..., ca. 3.162“ nimmt die Intervallspanne kontinuierlich zu.

¹¹¹ Rau, „Statistisch-mathematische Methoden der steuerlichen Betriebsprüfung und die Strukturanalyse als ergänzende Alternative“, Reihe: „Steuern, Wirtschaft und Recht“, Bd. 316, 2012, 43 mit weiteren Zitaten.

¹¹² Schmitz/Gebbers/Huber, „Nachruf auf Wilhelm Baselt – ‚Vater‘ der modernen steuerlichen Prüfungstechnik“, StBp 4/2011, 93.

¹¹³ Mit einem Überblick über den Einsatz der Logarithmischen Normalverteilungen im Rahmen von Betriebsprüfungen: Wähner, „Logarithmische Normalverteilung als Prüfungsmittel für Wirtschaftsdaten – Die Logik eines leistungsfähigen Verteilungsmusters“, StBp 1/2016, 1.

¹¹⁴ Tenzer, „Warum Menschen Steuern zahlen – oder eben nicht“, Psychologie Heute 9/2013, 30.

¹¹⁵ Hill, „Random-number guessing and the first digit phenomenon“, Psychological Reports 62/1999, 967; Hill, „The difficulty of faking data“, Chance 12/3/1999, 27.

¹¹⁶ Ausführlicher zu den physiologischen Zusammenhängen bei der Steuerhinterziehung: Wähner, „Manipulationspsychologie – Kann Manipulation dauerhaft unsichtbar bleiben?“, StBp 4/2013, 102.

(Bar-) Einnahmenhinterziehung lässt sich diese Dynamik über die [Musterabweichung](#) von der erwarteten [Logarithmischen Normalverteilung](#) gezielt suchen – im Rahmen der [Summarischen Risikoprüfung](#) (SRP) geschieht dies mittels [Verteilungsanalyse](#).

Materielle Ordnungsmäßigkeit bzw. Mängel

⇒ [Richtigkeitsvermutung](#)

Median

Als 50 %-[Quantil](#) unterteilt der Median eine aufsteigend sortierte Datenmenge in die eine Hälfte mit kleineren Werten und die andere mit größeren; sollte sich dabei ein nicht-ganzzahliger Rang ergeben, wird zwischen den beiden angrenzenden Werten [interpoliert](#).

In der beschreibenden (deskriptiven) [Statistik](#) von Wirtschaftsdaten liefert der Median häufig einen aussagekräftigeren Stellvertreterwert als das [arithmetische Mittel](#), weil asymmetrische Häufigkeitsstrukturen wie die [Logarithmischen Normalverteilungen](#) besser repräsentiert werden.

„Mit an Sicherheit grenzender Wahrscheinlichkeit“

⇒ [Beweis](#) sowie [Beanstandungsanlass](#)

Mittelwert

Das arithmetische Mittel ist der verbreitetste [statistische](#) Stellvertreter für eine größere Datenmenge und berechnet sich als Quotient aus der Summe aller Einzelwerte geteilt durch die Anzahl. Wichtige (Streu-)Verteilungen wie z.B. die [Normalverteilungen](#) sind zudem durch den Mittelwert und die [Standardabweichung](#) der Daten ausreichend bestimmt.

Sofern die einzelnen Daten einer Mittelwertberechnung eine unterschiedliche Gewichtung haben, muss dies beachtet werden: So hat das FG München darauf hingewiesen, dass die Margen einzelner Produktgruppen mit ihrem Wareneinsatzanteil (gewogen) in den Gesamtaufschlagsatz einfließen müssen.¹¹⁷

Zur Beschreibung von – häufig asymmetrisch verteilten ([Logarithmische Normalverteilungen](#)) – Wirtschaftsdaten eignet sich der durchschnittliche Rang „[Median](#)“ deutlich besser, weil er nicht einseitig verzerrt wird. Weist eine nicht-symmetrische Datenmenge nur wenige einseitige Ausreißer mit Verfälschungspotential auf, kann auch das [gestutzte Mittel](#) zu einem geeigneten Repräsentativ führen. Darüber hinaus bietet die beschreibende (deskriptive) Statistik mit dem geometrischen und harmonischen weitere Mittelwerte für spezielle Anwendungen an.

Modalklasse (Modalwert)

In der beschreibenden (deskriptiven) [Statistik](#) bezeichnen die Modalklasse bzw. der Modalwert den Bereich einer [stetigen](#) resp. [diskreten Häufigkeitsverteilung](#) mit der größten Anzahl an Datensätzen.

Monetary Unit Sampling (MUS)

Monetary Unit Sampling (MUS)¹¹⁸ ist eine mathematisch-statistische [Stichprobenprüfung](#), wie sie durch Gesetz¹¹⁹, höchstrichterliche Rechtsprechung¹²⁰ und die IDW-Prüfungsstandards¹²¹ anerkannt sind und z.B. in der Medizin¹²² oder Qualitätssicherung¹²³ Anwendung finden.

Bei der Bestimmung eines ausreichend großen Stichprobenumfangs und für die Hochrechnung des Stichprobenergebnisses auf die Grundgesamtheit greift Monetary Unit Sampling (MUS) auf die [Wahrscheinlichkeitslehre](#) als mathematisches Spezialgebiet der Datenstreuung zurück.¹²⁴ Die Besonderheit des Monetary Unit Sampling (MUS) gegenüber anderen mathematisch-statistischen Stichprobenprüfungen liegt in der wertproportionalen Ziehungsweise: Dabei werden die Geldeinheiten der Grundgesamtheit in ihrer seriellen Ordnung durch die ermittelte Stichprobenanzahl in Intervalle unter-

¹¹⁷ FG München v. 30.8.2011 – 10 V 735/11.

¹¹⁸ Giezek, „Monetary Unit Sampling – Der Einsatz statistischer Verfahren im Rahmen der Jahresabschlussprüfung“, Gabler Verlag Wiesbaden, 2011.

¹¹⁹ § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden.“

¹²⁰ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

¹²¹ IDW PS 310 „Repräsentative Auswahlverfahren (Stichproben) in der Abschlussprüfung“, s.: <https://www.idw.de/idw/verlautbarungen/idw-ps-310/43204>.

¹²² Bortz/Schuster, „Statistik für Human- und Sozialwissenschaftler“, Springer-Verlag Berlin Heidelberg, 2010, 11.

¹²³ DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/de/norm/din-iso-2859-1/191697831>.

¹²⁴ Die einschlägige Streuungscharakteristik zur Stichprobenprüfung ist die [Hypergeometrische Verteilung](#). Ältere Analysesets zum Monetary Unit Sampling (MUS) verwenden z.T. die Poisson- oder [Binomialverteilung](#), weil deren integrale Flächenfunktionen zum Berechnen von Wahrscheinlichkeitsbereichen weniger kompliziert sind.

teilt und aus jedem Intervall ein Zufallsobjekt¹²⁵ gezogen. Durch den Bezug der Geldeinheit zu einem Vorgang fällt dieser in die Stichprobe. Diese Wertorientiertheit der Stichprobenziehung führt dazu, dass die Ziehungswahrscheinlichkeit wertabhängig ist und Sachverhalte mit einem Wert $\geq 2 \cdot \text{Intervall} - 1$ sicher in die Stichprobe gelangen. Aus diesem Grund ist eine direkte¹²⁶ Anwendung des Monetary Unit Sampling (MUS) nur zum Auffinden von Überbewertungen geeignet, denn niederwertige Vorgänge fallen wahrscheinlicher durch das Stichprobensieb.

Wegen der wechselseitigen Beziehung der Wahrscheinlichkeiten zum Stichprobenumfang¹²⁷ müssen für die MUS-Anwendung bestimmte Parameter festgelegt werden, um eine einheitliche Beurteilungsgrundlage zu gewährleisten: Aus der Wirtschaftsprüfung stammt die Konkretisierung mit einer Aussagegewissheit (Konfidenzniveau) von mindestens 95 %¹²⁸, dass bei einer fehlerfreien Stichprobe der hochgerechnete Fehler in der Grundgesamtheit die Wesentlichkeitsgrenze (Materiality) von 3 % nicht übersteigt.¹²⁹ Daraus ergibt sich nach der Hypergeometrischen Verteilung regelmäßig ein Stichprobenumfang von 99 Geldeinheiten, wobei Doppelziehungen von Vorgängen, die über zwei oder mehrere Intervalle reichen, zu einer geringeren Anzahl an Sachverhalten führen können. Sind alle gezogenen Vorgänge fehlerfrei, kann mit hinreichender Sicherheit geschlossen werden, dass der Fehler der unprüfbar großen Grundgesamtheit nicht über 3 % liegt.

Musterbasierte Datenanalyse

Häufigkeitsauswertungen (als Gesamtbildmethode) mit mehr als drei Klassen (ABC-Analyse) sind aufgrund der menschlichen Wahrnehmung für eine tabellarische Analyse weniger geeignet als für eine verbildlichte.¹³⁰ Da zudem aus der Statistik und Volkswirtschaftslehre bekannt ist, dass natürliche Datenstreuungen regelmäßigen Mustern wie den Normalverteilungen oder Logarithmischen Normalverteilungen folgen, ist die Gegenüberstellung von untersuchten Daten mit diesen allgemeinen oder betriebsinternen Referenzen ein wirksamer, effizienter Ansatz, um Prüfungs- bzw. Beratungsfelder zu erkennen oder die Schlüssigkeit gem. § 158 AO zu beurteilen (Beanstandungsanlass, Ziffernanalyse und Verteilungsanalyse).

„Unser Dilemma im Umgang mit der Komplexität unserer Welt lässt sich darauf zurückführen, dass wir wohl darin ausgebildet wurden einfache logische Schlüsse zu ziehen und nahe liegende Ursache-Wirkungs-Beziehungen zu definieren. ... Simple Ursache-Wirkungs-Beziehungen gibt es nur in der Theorie, nicht in der Wirklichkeit.“¹³¹ Infolge dieser Erkenntnis gründet sich mit dem Sensitivitätsmodell nach Vester auch eins der erfolgreichsten Lösungssysteme für komplexe Probleme auf die Identifikation und Bewertung von (Beziehungs-)Mustern.

Nachkalkulation

⇒ Aufschlagskalkulation sowie Ausbeutekalkulation

Neue interaktive Prüfungstechnik (NiPt)

Unter dem Namen „Neue interaktive Prüfungstechnik (NiPt)“ findet seit 2009 ein Austausch zur digitalen Außenprüfung zwischen Experten mehrerer Länder auf Initiative aus Österreich und Deutschland statt.¹³²

Als Reaktion auf die veränderten Steuerausfallrisiken durch die Digitalisierung der Betriebsverwaltung¹³³ werden in diesem Arbeitskreis Erfahrungen und Ansätze zusammengeführt, um ein besseres

¹²⁵ Der Zufallsmodus der Stichprobenziehung kann durch die einmalige Zufallsauswahl innerhalb des ersten Intervalls mit jeweiliger Addition des Intervallumfangs oder durch Einsatz eines Zufallsgenerators für jedes Intervall erreicht werden. MUS-Analysesets bieten häufig beide Ziehungsmodi an.

¹²⁶ Durch eine Wertumkehr der einzelnen Vorgänge am Maximum (Stichprobenwert eines Vorgangs = „Maximum in der Grundgesamtheit – Wert des Vorgangs“) ist eine Übertragung des Monetary Unit Sampling (MUS) auch auf Unterbewertungen denkbar.

¹²⁷ Aus der Grafik zum Begriff „Binomialverteilung“ ist dieser Zusammenhang ablesbar.

¹²⁸ Das Konfidenzmaß „95 %“ leitet sich aus den Rang-Wert-Verhältnissen der Normalverteilungen ab: Es gibt den Normalbereich zuzüglich schwacher Ausreißer mit „Mittelwert ± doppelte Standardabweichung“ wieder.

¹²⁹ Mit einem Berechnungsbeispiel: *Giezek/Wähner/Becker*, „Anwendung des Monetary Unit Sampling (MUS) in der steuerlichen Betriebsprüfung – Qualifizierte Untersuchung und Bewertung von Stichproben“, StBp 12/2016, 347.

¹³⁰ *Barnard* in einer US-amerikanischen Zeitschrift der Werbebranche 1921: „One look is worth a thousand words“, nachzulesen auf der Internetseite des Ostasieninstituts der Hochschule Ludwigshafen am Rhein, abrufbar unter: <http://www.oai.de/de/component/content/article/45-publikationen/spruchwort/811-ein-bild-sagt-mehr-als-tausend-worte.html>.

¹³¹ *Vester*, „Die Kunst vernetzt zu denken – Ideen und Werkzeuge für einen neuen Umgang mit Komplexität“, Deutscher Taschenbuch Verlag München, 9. Auflage 2012, 15.

¹³² *Huber/Wähner*, „Der systematische Einsatz neuer Prüfverfahren in der Finanzverwaltung – Neue interaktive Prüfungstechnik (NiPt)“, NWB 36/2009, 2814; *Huber/Wähner*, „Das Kölner Zeitreihenurteil und das Projekt ‚NiPt‘“, StBp 7/2009, 207.

¹³³ *Teutemacher*, „Handbuch zur Kassenführung“, NWB Verlag Herne, 2015, 137 ff. mit einer Offenlegung einiger verbreiteter Manipulationsmethoden. BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 27: „Elektronische Kassensysteme sind durch Umprogrammierung in nahezu beliebiger Weise

Grundlagenverständnis und eine bedienerführende Systematik in der Methodenanwendung zu etablieren. Die [Summarische Risikoprüfung](#) (SRP) mit ihren [SRP-Vorlagen](#) fußt zu einem wesentlichen Anteil auf diesem länderübergreifenden Austausch.

Newcomb-Benford-Law (NBL)

⇒ [Benford-Verteilung](#)

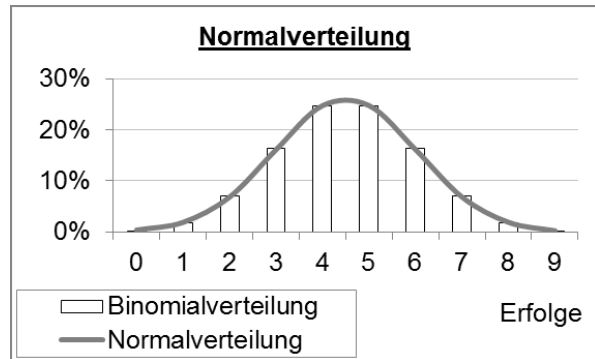
Nominaler Zeitreihenvergleich

⇒ [Interaktive Sach\(verhalts\)aufklärung](#)

Normalverteilungen

Die glockenförmigen Normalverteilungen beschreiben das bekannteste Streuverhalten in (natürlich entstandenen) Datenmengen.

Normalverteilungen sind durch ihre Kenngrößen „[Mittelwert](#)“ und „[Standardabweichung](#)“ vollständig definiert, so dass jede Klassenhäufigkeit einzeln oder kumuliert berechnet werden kann. Wie der Grafik zu entnehmen ist, lassen sich durch Übertragung dieser (spezifischen) Parameter aus der [Binomialverteilung](#) deren [kombinatorische](#) Zufallsverhältnisse auch auf [stetige](#) Beobachtungen ausweiten. Zudem sind die Rangwert-Relationen in allen Normalverteilungen identisch, so dass sie aus der [Standardnormalverteilung](#) grundsätzlich abgeleitet werden können.



Aufgrund der empirischen Erfahrungen und der theoretischen Untermauerung gelten die Normalverteilungen als beste dem Menschen bekannte allgemeine Wiedergabe des Streuverhaltens und finden entsprechend Anwendung in der statistischen Kommunikation von Beobachtungen¹³⁴ sowie in Form von [Wahrscheinlichkeitstests](#) zur Bestätigung von Studienergebnissen und Qualifizierungsarbeiten, in der Qualitätssicherung¹³⁵ und Beweisführung¹³⁶. Im Rahmen der [Summarischen Risikoprüfung](#) (SRP) werden alle [Musteranalysen](#) zur indiziellen Aufdeckung [menschlicher Eingriffe](#) mit Hilfe einschlägiger Wahrscheinlichkeitstests beurteilt.

Nullhypothese

⇒ [Hypothese](#)

Permutation

⇒ [Kombinatorik](#)

Perzentil

⇒ [Quantil](#)

Poisson-Verteilung

⇒ [Hypergeometrische Verteilung](#)

Probabilistische Schlussfolgerungen

„Probabilistische Schlussfolgerung“ ist ein Fachterminus für eine Wahrscheinlichkeitsaussage, wie sie z.B. im Rahmen von Prüfungsmaßnahmen und Gutachten die Regel ist, weil sich die Vergangenheit nicht vollständig, objektiv rekonstruieren lässt und somit eine Restunsicherheit bei der Beschreibung und Beurteilung von Vorgängen verbleiben muss.

manipulierbar; von derartigen Manipulationsmöglichkeiten machen Teile der betrieblichen Praxis nach dem Erkenntnisstand des Senats durchaus Gebrauch ...“. Darüber hinaus zu den Risiken der schnellen und massenhaften Datenverarbeitung z.B.: www.elektronische-steuerprüfung.de, Newsletter 4-2016 v. 22.4.2016: „(IT-)Management: Ein Drittel aller Unternehmen verlieren Daten bei der Migration: Ein Drittel (32 %) der Organisationen verlieren Daten, während sie Daten zwischen Geräten migrieren oder Betriebssysteme aktualisieren. Das ist das Ergebnis einer weltweiten Befragung unter fast 600 IT-Administratoren ...“.

¹³⁴ Als ein typisches Beispiel zur Diskussion ethnischer Unterschiede in der Intelligenzverteilung: *Herrnstein/Murray*, „The bell curve – Intelligence and class structure in American life“, The Free Press New York, II. Title 1994.

¹³⁵ DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/de/norm/din-iso-2859-1/191697831>.

¹³⁶ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

Erfahrene Gutachter deutscher Kriminalämter empfehlen eine Verallgemeinerung von Wahrscheinlichkeitsaussagen durch (geschätzte) numerische Einordnungen, denn „[d]ie numerische Schätzung bewirkt insbesondere eine Vermeidung von Missverständnissen oder möglichen Unterschieden in den scheinbar gleichen Größenordnungen verschiedener Individuen bei identischen verbalen Ausdrücken. Soweit aus dieser Schätzung klar hervorgeht, dass sie ein numerischer Schätzwert ist, wird auch keine erhöhte Exaktheit behauptet. Sie erfordert vom Sachverständigen schlicht eine präzisere Einschätzung. Sie erhöht die intersubjektive Beurteilbarkeit dessen, was gemeint ist.“¹³⁷ Die vorgeschlagene Konkretisierung verbaler Wahrscheinlichkeiten¹³⁸ durch (geschätzte) numerische versachlicht auch die Rechtsprechung, indem z.B. die gesetzlich geforderte Klarheit des [Beanstandungsanlasses](#) nach § 158 AO eingegrenzt wird: Der juristische Beweis ordnet sich laut BGH-Rechtsprechung¹³⁹ und Kriminalistik¹⁴⁰ im Bereich der beiden höchsten Qualitäten ein. Da „eine Stufe an Unterschiedlichkeit in den Wahrscheinlichkeitsgraden ... tolerabel [sei], größere Differenzen jedoch keinesfalls“¹⁴¹, muss der gesetzliche Beanstandungsanlass zwangsläufig in dem Wahrscheinlichkeitsrahmen „90-95 %“ belegen sein; damit kann die verbale Interpretation der Verfahrensrechtsnorm durch den BFH „mit an Sicherheit grenzender Wahrscheinlichkeit“ nicht korrekt sein.¹⁴²

Konkretisierung von Wahrscheinlichkeitsaussagen	
Verbale Wahrscheinlichkeit	Numerische Zuordnung
"non liquet"	50%
"mit leicht überwiegender Wahrscheinlichkeit"	75%
"mit überwiegender Wahrscheinlichkeit"	90%
"mit hoher Wahrscheinlichkeit"	95%
"mit sehr hoher Wahrscheinlichkeit"	99%
"mit an Sicherheit grenzender Wahrscheinlichkeit"	99,99%

Prozentrang

⇒ [Quantil](#)

Prüfungsnetz

⇒ [Summarische Risikoprüfung](#) (SRP)

Punkterwartung

⇒ [Referenz](#)

Quantil

Quantile sind relative Ränge, mit denen – im Gegensatz zu [absoluten Rängen](#) – eine vergleichende Bewertung trotz unterschiedlicher Datenmengen möglich ist.

Dabei wird die Datengesamtheit in Prozentanteile unterteilt, so dass z.B. das 75 %-Quantil den – aufsteigend sortiert – 75 % größten Wert abzählt; sollte sich dabei ein nicht-ganzzahliger Rang ergeben, wird zwischen den beiden angrenzenden Werten [interpoliert](#). Wegen der verbreiteten Nutzung von Prozenträngen haben sich spezielle Unterteilungen etabliert, welche – aus dem Lateinischen abgeleitet – die Anzahl der entstehenden Bereiche angeben: z.B. Perzentile (100), Dezile (10), Quintile (5) und Quartile (4); „Perzentil“ steht im Englischen anstelle von „Quantile“ auch für die allgemeine Bezeichnung der Prozentränge.

In der beschreibenden (deskriptiven) [Statistik](#) sowie der Betriebswirtschaftslehre werden derartige relative Ränge eingesetzt, um die robusteren¹⁴³ [Rangkennzahlen](#) über variable Datenumfänge hinweg nutzen zu können, z.B. damit vergleichbare Schwellenwerte bei der Beurteilung von Gehältern definiert werden können, selbst wenn zu verschiedenen Branchen unterschiedlich viele Vergleichswerte

¹³⁷ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Luchterhand Verlag München, 2004, 13.

¹³⁸ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Luchterhand Verlag München, 2004, 33.

¹³⁹ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 26.

¹⁴⁰ Giezek/Rupprecht/Wähner, „Wie sicher muss ein Verprobungsergebnis sein? – Anleihen aus der Kriminalistik zur sinnvollen Einordnung des Beanstandungsanlasses“, BBK 5/2017, 236, mit diversen Quellen.

¹⁴¹ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Luchterhand Verlag München, 2004, 33.

¹⁴² BFH v. 24.6.1997 – VIII R 9/96, BStBl II 1998, 51; so auch: Seer, Tipke/Kruse, AO/FGO, § 158 AO Rz. 14.

¹⁴³ Unter der „Robustheit“ von Kennzahlen wird verstanden, dass sie nicht bzw. wenig anfällig gegenüber alltäglichen Sachverhalten sind.

vorliegen. Entsprechend lassen sich z.B. auch Monatsaufschlagsätze trotz unterschiedlich langer Betrachtungszeiträume vergleichen wie bei der [Quantilsschätzung](#).

Quantilsschätzung

Im Fall der Schätzungsnotwendigkeit gem. § 162 Abs. 2 S. 2 AO stellt die Quantilsschätzung eine Schätzungsoption¹⁴⁴ dar, um die systematisch erarbeiteten Erkenntnisse der [Summarischen Risikoprüfung](#) (SRP) für die Feststellung eines sachgerechten „Schätzungsrahmens“¹⁴⁵ zu nutzen.

Als zeitreihenbasierte Schätzungsmethode werden hierfür Monatskennzahlen wie i.d.R. der Aufschlagsatz zu Grunde gelegt, um in einem betriebsspezifisch festgelegten Schätzungsrahmen an die oberste Grenze zu gehen¹⁴⁶, damit dem Geschätzten kein Anreiz gegeben wird, sich weiterhin schätzen zu lassen.¹⁴⁷ Geeignet ist die Quantilsschätzung mit ihrer Zielgröße „objektiviertes Leistungsfähigkeitsmaß“¹⁴⁸ nach der Drei-Stufen-Theorie des BFH¹⁴⁹ für Schätzungsfälle mit gravierenden Ordnungsmängeln, in denen zeitaufwendige¹⁵⁰ Einzelermittlungen wie [Geldflussrechnungen](#) und [Kalkulationen](#) keine Gewähr für ein genaueres Ergebnis bieten können.¹⁵¹

In der Durchführung folgt die Quantilsschätzung folgenden Überlegungen und Arbeitsschritten, wie sie zur Ableitung eines Schätzungsergebnisses aus einer Zeitreihenanalyse erforderlich sind – mit Mitteln, die in den Wissenschaften allgemein Anerkennung genießen:¹⁵²

1. Sofern Besteuerungsaufzeichnungen unvollständig sind oder der Verdacht einer Manipulation vorliegt (z.B. wegen Ordnungsmängeln, welche die Nachvollziehbarkeit der Entstehung verhindern), geben Kennzahlen mit ihrer „Robustheit“¹⁵³ den Kern der (Betriebs-)Verhältnisse häufig besser wieder als Hochrechnungen auf Basis der nicht vollständigen oder falschen Ursprungsdaten.¹⁵⁴
2. Monatskennzahlen sind in aller Regel robust genug, weshalb sie in den BWAs der Beratungspraxis für existenzrelevante Entscheidungen wie z.B. Investitionen herangezogen werden, und weisen gegenüber den – gem. Schätzungsnotwendigkeit – falschen Jahresgrößen eine ausreichende Streuung auf, um einen Schätzungsrahmen feststellen zu können.
3. Um bspw. die einzelnen Monatsaufschlagsätze hinsichtlich ihrer Schätzungseignung zu bewerten, ist eine statistische Einordnung notwendig. Dabei sind [Rangordnungen arithmetischen](#) vorzuziehen, weil ansonsten die als ungeeignet erkannten Ausreißer das Ergebnis dennoch beeinflussen würden. Aufgrund unterschiedlich langer Prüfungszeiträume sollten die Ränge jedoch wie bei [Quantilen](#) relativ gewählt werden, damit der Ansatz verallgemeinerbar ist.
4. Das übliche Verbildlichungsmittel zur Rangeinordnung in der deskriptiven (beschreibenden) [Statistik](#) ist der [Box-Plot](#), der den Normbereich der Daten von den schwachen und starken Ausreißern trennt.

¹⁴⁴ Die Quantilsschätzung ist weder gleichbedeutend mit der Summarischen Risikoprüfung (SRP) noch dogmatisch mit dieser verknüpft, sondern stellt eine Schätzungsoption unter mehreren dar; so auch bereits die erste Vorstellung: *Wähnert*, „Die Quantilsschätzung – Schätzungsmöglichkeiten mit summarischen Methoden“, BBK 20/2008, 1083, und alle folgenden Beschreibungen wie z. B.: *Schumann/Wähnert*, „Die Quantilsschätzung – Schätzung anhand eines objektivierte Leistungsfähigkeitsmaßes“, Stbg 2012, 535.

¹⁴⁵ *Seer, Tipke/Kruse, AO/FGO*, § 162 AO Rz. 44: „Naturgemäß besteht bei der Schätzung nach Wahrscheinlichkeitsgrundsätzen eine Bandbreite möglicher Wertansätze (sog. Schätzungsrahmen). Soweit sich die Schätzung innerhalb des Rahmens bewegt, den der durch die FinBeh. unter Mitwirkung des Stpfl. festgestellte Sachverhalt vorgibt, ist sie nicht zu beanstanden ... Der Schätzungsrahmen ist umso größer, je ungesicherter das Tatsachenmaterial ist, auf dem die Schätzung basiert.“

¹⁴⁶ BFH v. 9.3.1967 – IV 184/63, BStBl III 1967, 349: „Müssen wegen grober Verstöße gegen die steuerlichen Pflichten Umsätze und Gewinne im Wege der Schätzung gegenüber den Erklärungen erheblich erhöht werden, so ist das FA im allgemeinen nicht nur berechtigt, sondern verpflichtet, in dem gegebenen Schätzungsrahmen an die oberste Grenze zu gehen.“; auch gemäß (RiBFH) *Nöcker* „zulässig“, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 2015, 3548, 3549.

¹⁴⁷ FG Berlin-Brandenburg v. 30.6.2017 – 10 V 10219/16: „Andererseits darf es keinen Anreiz geben, sich weiterhin schätzen zu lassen; deshalb ist die Behörde nicht nur berechtigt, sondern sogar verpflichtet, in dem gegebenen Schätzungsrahmen an die obere Grenze zu gehen ...“.

¹⁴⁸ *Schumann/Wähnert*, „Die Quantilsschätzung – Schätzung anhand eines objektivierte Leistungsfähigkeitsmaßes“, Stbg 2012, 535.

¹⁴⁹ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 60 ff.

¹⁵⁰ BFH v. 21. 2. 1990 – X R 54/87, BFH/NV 1990, 683: „Die Geldverkehrsrechnung ist eine arbeitsaufwendige Schätzungsmethode. Die vom Kläger erklärten Buchführungsergebnisse beruhen auf einer formell ordnungswidrigen Buchführung. Daher ist die Anwendung weniger arbeitsaufwendiger Schätzungsmethoden statthaft.“

¹⁵¹ *Giezek/Wähnert*, „Schätzungsmethoden der Betriebsprüfung im Vergleich – Einordnung der wichtigsten Schätzungsverfahren durch Praxis und Wissenschaft, BBK 21/2017, 998; *Pump/Wähnert*, „Das BFH-Urteil zum Zeitreihenvergleich als Verprobungs- und Schätzungsmethode Was bedeuten die Ausführungen für die Praxis?“, NWB 39/2015, 2869.

¹⁵² *Wähnert*, „Schätzungen mit Hilfe von Zeitreihenanalysen – Schrittweise Beschreibung der Quantilsschätzung“, StBp 11/2017, 323.

¹⁵³ Unter der „Robustheit“ von Kennzahlen wird verstanden, dass sie nicht bzw. wenig anfällig gegenüber alltäglichen Sachverhalten sind.

¹⁵⁴ Insofern zielt die Quantilsschätzung auf entsprechende Schätzungsfälle mit ordnungswidrigen Unterlagen; *Schumann/Wähnert*, „Die Quantilsschätzung – Schätzung anhand eines objektivierte Leistungsfähigkeitsmaßes“, Stbg 2012, 535, 538 bereits mit dem Hinweis auf den Zusammenhang zwischen Quantilsschätzung und Ordnungswidrigkeit der Aufzeichnungen.

5. Für die Schwellenwerte des Schätzungs-Box-Plots werden bei der Quantilsschätzung die Rang-Wert-Verhältnisse¹⁵⁵ der [Normalverteilungen](#) übernommen, weil diese im mathematischen Spezialgebiet für Datenstreuungen – der „[Wahrscheinlichkeitslehre](#)“ – das verbreitetste Verhalten wiedergeben.
6. Im sich daraus ergebenden Schätzungsrahmen wird mit dem 80 %-Quantil (defensiv¹⁵⁶) am oberen Rand geschätzt.

Wie bei jedem Standardverfahren zur Schätzung soll auch im Rahmen der Quantilsschätzung abschließend überschlägig hinterfragt werden, ob die ermittelte Hinzuschätzung auf Grundlage der Betriebsverhältnisse „wirtschaftlich vernünftig und möglich“¹⁵⁷ erscheint.¹⁵⁸

Quartil

⇒ [Quantil](#)

Quintil

⇒ [Quantil](#)

Rangkennzahlen

Neben [arithmetischen Kennzahlen](#) wie Periodensummen oder Quotienten (z.B. Liquiditätsparameter) setzen Planung und Prüfung zur aussagekräftigen Beurteilung umfangreicher Datenmengen bzw. komplexer Vorgänge gerne auch auf Rangkennzahlen, die sich aus der wertbezogenen Sortierung mit der resultierenden Rangzuordnung ergeben.

Je nachdem ob die zum Vergleich herangezogenen Datenmengen vergleichbar groß sind oder deutlich abweichen können, kommen dabei [absolute](#) oder relative Ränge ([Fraktile](#) bzw. [Quantile](#)) zum Einsatz. Letztere ermöglichen eine Beurteilung über variable Datenumfänge hinweg, z.B. um einheitliche Schwellenwerte bei der Beurteilung von Gehältern zu definieren, selbst wenn zu verschiedenen Branchen unterschiedlich viele Vergleichswerte vorliegen.

Der Vorteil von Rangkennzahlen gegenüber arithmetischen ist, dass die Zielgröße nicht durch unrepräsentative Ausreißer beeinflusst und möglicherweise (einseitig) verzerrt wird.¹⁵⁹

Realer Zeitreihenvergleich

⇒ [Interaktive Sach\(verhaltens\)aufklärung](#)

Rechtsschutz

⇒ [Chancengleichheit](#)

Referenz(-Verteilung)

Referenzen sind Vergleichswerte oder -datenmengen, die im Rahmen der Planung, Beratung und Prüfung zur Beurteilung herangezogen werden. Sie können aus internen Erfahrungen (der Vergangenheit) oder aus externen (Forschungs-)Erkenntnissen wie z.B. bei der [Benford-](#) oder [Logarithmischen Normalverteilung](#) stammen.

Für einen aussagekräftigen Abgleich der [Häufigkeitsverteilungen](#) von untersuchten Daten mit einer sog. Punkterwartung (Referenz) ist zu beachten, dass diese sich nach empirischen Beobachtungen und den theoretischen Beschreibungen der [Statistik](#) mit zunehmendem Umfang dem Ideal annähern.¹⁶⁰ Um diese bekannte Streuungscharakteristik bei der Übereinstimmungsbewertung zu beachten, erfolgen derartige Gegenüberstellungen häufig unter Einsatz von [Wahrscheinlichkeitstests](#) wie dem [Chi-Quadrat-Test](#), deren Ergebnisse die untersuchte Datenmenge berücksichtigen.

Regelbasierte Datenanalyse

Im Gegensatz zur [musterbasierten Datenanalyse](#) sucht die regelbasierte in untersuchten Daten nach Entwicklungs- oder Abhängigkeitsregeln.

¹⁵⁵ Zum Begriff „[Standardnormalverteilung](#)“ sind die Rang-Wert-Verhältnisse aller Normalverteilungen als Grafik abgebildet.

¹⁵⁶ Der obere Rand des Normbereichs „[Mittelwert](#) ± [Standardabweichung](#)“ ergibt sich in allen Normalverteilungen mit dem ca. 84 %-Quantil (ca. 16 % untere Ausreißer zuzüglich ca. 68 % Normwerte).

¹⁵⁷ BFH v. 18.12.1984 – VIII R 195/82, BStBl II 1986, 226.

¹⁵⁸ *Schumann/Wähner*, „Die Quantilsschätzung – Schätzung anhand eines objektivierten Leistungsfähigkeitsmaßes“, Stbg 2012, 535, 538 z.B. mit dem konkreten Hinweis.

¹⁵⁹ So beschreibt bspw. die mittlere [Rangkennzahl](#) „Median“ mit „49“ die unter dem Begriff „[Box-Plot](#)“ abgebildete Datenmenge eindeutig besser als das [arithmetische Mittel](#) mit dem Wert „57“, weil die beiden hohen Ausreißer letzteren verfälschen.

¹⁶⁰ Aus der Grafik zum Begriff „[Binomialverteilung](#)“ ist dieser Zusammenhang ablesbar.

[Zeitreihenanalysen](#) bilden z.B. einen variantenreichen¹⁶¹ Block an Instrumenten, um die (chronologische) Entwicklung eines Betriebsfaktors oder den Zusammenhang zweier an einer erwarteten Regel ([Hypothese](#)) nachzuverfolgen. Wie bei den musterbezogenen [Häufigkeitsauswertungen](#) wird hierbei auf den wahrnehmungsphysiologischen Vorteil der Verbildlichung größerer Datenmengen gesetzt.¹⁶² Je nach Analysezweck kann über die Zeitreihenart und die Diagrammgestaltung der Fokus sehr unterschiedlich gewählt¹⁶³ und somit die Aussagekraft entscheidend beeinflusst werden.

Entgegen finanzgerichtlichen Einschätzungen¹⁶⁴ finden regelbasierte Zeitreihenanalysen in Beratung und Prüfung bei vielen komplexen Fragestellungen Anwendung: Zur Erlangung eines Überblicks, zur Identifikation von betriebsspezifischen Beratungs- resp. Prüffeldern sowie zur Transparenzsteigerung und Schlüssigkeitsbewertung von fragwürdigen Gestaltungen¹⁶⁵, denn viele alltäglichen Probleme der Beratungs- und Prüfungspraxis lassen sich aufgrund ihrer Komplexität über einfache Ursache-Wirkungs-Beziehungen nicht lösen.¹⁶⁶

Relativer Rang (Relativrang)

⇒ [Quantil](#) sowie [Fraktile](#)

Richtigkeitsvermutung

Das besondere Behörden-Bürger-Verhältnis im Besteuerungsverfahren (Verifikationsprinzip¹⁶⁷) sieht gem. § 158 AO zwei (unabhängige) Voraussetzungen für die „gesetzliche Vermutung“¹⁶⁸ der sachlichen Richtigkeit erklärter Besteuerungsgrundlagen vor.¹⁶⁹

Durch die Einhaltung der gesetzlich vorgeschriebenen Ordnungsvorgaben (formelle Ordnungsmäßigkeit) soll sichergestellt werden, dass sich ein „sachverständige[r] Dritte[r] innerhalb angemessener Zeit einen Überblick über die Geschäftsvorfälle und über die Lage des Unternehmens“¹⁷⁰ verschaffen kann.¹⁷¹

Als materielle Ordnungsmäßigkeit ist in § 158 AO mit den Worten „soweit nach den Umständen des Einzelfalls kein Anlass ist, ihre sachliche Richtigkeit zu beanstanden“ (Fehlen eines [Beanstandungsanlasses](#)) die inhaltliche Plausibilität normiert. Jeder sachliche Einzelfehler sowie nicht oder nicht in angemessener Zeit aufklärbare ([interaktive Sachaufklärung](#)) Unschlüssigkeiten infolge einer Gesamtbildprüfung wie z.B. [Geldverkehrsrechnung](#), [Kalkulation](#), [Zeitreihenanalyse](#), [Häufigkeitsuntersuchung](#), [Manipulationstest](#) oder [Stichprobenprüfung](#) stellt ein Indiz gegen die inhaltliche Richtigkeit von Besteuerungsaufzeichnungen dar.

Sofern eine von beiden Voraussetzungen unheilbar nicht erfüllt ist, tritt nach § 162 Abs. 2 S. 2 AO die Amtspflicht der Schätzung ein.

Rollierende Zeitreihe

⇒ [Gleitschlitzenzeitreihe](#)

Sachaufklärungsverantwortung

⇒ [Interaktive Sach\(verhaltens\)aufklärung](#)

¹⁶¹ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 37.

¹⁶² *Barnard* in einer US-amerikanischen Zeitschrift der Werbebranche 1921: „One look is worth a thousand words“, nachzulesen auf der Internetseite des Ostasieninstituts der Hochschule Ludwigshafen am Rhein, abrufbar unter: <http://www.oai.de/de/component/content/article/45-publikationen/spruchwort/811-ein-bild-sagt-mehr-als-tausend-woerter.html>.

¹⁶³ Eine Gegenüberstellung von Zeitreihenvarianten findet sich unter „[Zeitreihenanalyse](#)“.

¹⁶⁴ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 56 ff. mit der Einschränkung auf den Warenfluss leichtverderblicher Produkte mit konstanten Betriebsverhältnissen.

¹⁶⁵ *Wolenski*, „Das Zeitreihenurteil des X. Senats des BFH vom 25.3.2015“, StbG 6/2016, 268 mit einem Praxisbeispiel auf Seite 271, welches die Unschlüssigkeit der Besteuerungsaufzeichnungen für einen Betrieb unzweifelhaft offenlegt, obwohl dieser den vom BFH festgelegten Voraussetzungen nicht entspricht.

¹⁶⁶ *Vester*, „Die Kunst vernetzt zu denken – Ideen und Werkzeuge für einen neuen Umgang mit Komplexität“, Deutscher Taschenbuch Verlag München, 9. Auflage 2012, 15.

¹⁶⁷ BFH v. 16. 5. 2013 – II R 15/12, BStBl II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

¹⁶⁸ AEO zu § 158 AO.

¹⁶⁹ Ausführlich hierzu: *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, Berlin, 4. Auflage 2017, 103 ff.

¹⁷⁰ § 145 Abs. 1 S. 1 AO.

¹⁷¹ Hierzu ausführlich: „Grundsätze zur ordnungsgemäßen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD)“, BMF-Schreiben v. 14.11.2014, BStBl I 2014, 1450.

Sachverständigengutachten

⇒ [Chancengleichheit](#)

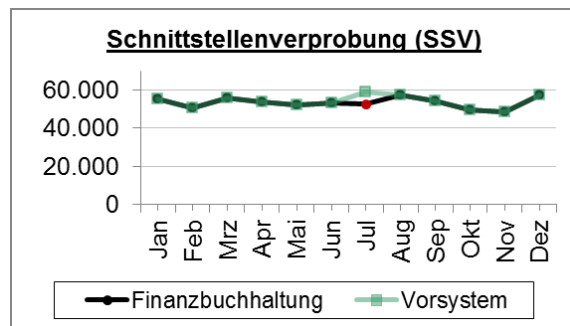
Schichtung

⇒ [Klassifizierung](#)

Schnittstellenverprobung (SSV)

Infolge der Digitalisierung der Betriebsverwaltung haben sich der Automatisierungsgrad und die Anzahl an Informationsschnittstellen wesentlich erhöht. Wegen der unter Spezialisten bekannten¹⁷² Risiken, die damit u.a. für die Steuererklärung verbunden sind, sollten sowohl die externen als auch systeminterne Datenschnittstellen durch [interne Kontrollsysteme](#) (IKS) überwacht werden.

In der Betriebsprüfung hat sich mit der Schnittstellenverprobung (SSV) ein eigenes Prüffeld entwickelt¹⁷³, welches häufig mit Hilfe der [SRP-Vorlagen](#) aus der [Summarischen Risikoprüfung](#) (SRP) bearbeitet wird. So ist in der nebenstehenden Grafik eine monatsweise Kontrolle der externen Schnittstelle aus Vorsystem (z.B. Kasse oder Warenwirtschaftssystem) und Finanzbuchhaltung zu sehen; aufgrund der Prüfung auf Übereinstimmung – im Gegensatz zur Abhängigkeitsprüfung ([Korrelation](#)) – wird ein einseitig skaliertes [Zeitreihenvergleich](#) gewählt, der die chronologische Entwicklung beider Datenreihen am selben Größenmaßstab wiedergibt. Neben der überwiegenden [Kongruenz](#) der Grafen lässt sich mit der (scheinbar) unvollständigen Übernahme im Juli zügig eine Regelstörung identifizieren. Durch eine sinnvolle Kombination aus [regel-](#) und [musterbasierten Datenanalysen](#) wie z.B. bei der Summarischen Risikoprüfung (SRP)¹⁷⁴ können ursächliche Einzelsachverhalte effektiv und schnell aufgefunden werden.



Schwebende Zahlenpositionen

Bei der [ziffernanalytischen](#) Betrachtung von Zahlen aus der Frontsicht ergeben sich die schwebenden Positionen „Erstziffer“, „Zweitziffer“, „Drittziffer“ etc. Dabei nimmt bspw. die Zweitziffer in einem Datensatz von geringerem Wert wie z.B. € 87,50 eine niedrigere monetäre Bedeutung ein wie in einem deutlich werthaltigeren Vorgang von bspw. € 3.416,00.

Diese Form der Zahlenbeschreibung entstammt der empirischen Datenforschung, aus der die regelmäßige Beobachtung der [Benford-Verteilung](#) hervorgegangen ist. In Datenanalyseprogrammen finden vorkonfektionierte Tools zur Gegenüberstellung der schwebenden Ziffernverteilungen mit der Benford-[Referenz](#) für die Risikoabschätzung¹⁷⁵ und bei der Suche nach Manipulationsspuren¹⁷⁶ Anwendung.

Im Verhältnis zu theoretischen Annahmen zu [festen Zahlenpositionen](#) muss bei der Untersuchung beachtet werden, dass empirische Erkenntnisse grundsätzlich vorrangig sind, weshalb in den [SRP-Vorlagen](#) der [Summarischen Risikoprüfung](#) (SRP) eine automatische Vermeidung der Überschneidung erfolgt.¹⁷⁷

Sensitivitätsanalyse

In zwei Entscheidungen fordert der BFH für die „dort streitgegenständliche Variante des Zeitreihenvergleichs“ als Schätzungsmethode (und wohl auch als Verprobungsmethode) von Amts wegen eine

¹⁷² Nachzulesen z.B. unter: www.elektronische-steuerprüfung.de, Newsletter 4-2016 v. 22.4.2016: „(IT-)Management: Ein Drittel aller Unternehmen verlieren Daten bei der Migration: Ein Drittel (32 %) der Organisationen verlieren Daten, während sie Daten zwischen Geräten migrieren oder Betriebssysteme aktualisieren. Das ist das Ergebnis einer weltweiten Befragung unter fast 600 IT-Administratoren ...“

¹⁷³ *Webel/Danielmeyer*: „Schnittstellenverprobung elektronischer Betriebsverwaltungen – das kommende Standardprüffeld?“, StBp 12/2015, 353; *Becker/Danielmeyer/Neubert/Unger*: „Digitale Offensive“ der Finanzverwaltung: Die Schnittstellen-Verprobung (SSV)“, DStR 51/2016, 2983.

¹⁷⁴ *Wähnert*, „Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz „Summarische Risikoprüfung (SRP)““, DB 45/2016, 2627.

¹⁷⁵ Nachzulesen unter: [www.ey.com/Publication/vwLUAssets/EY_-_Forensic_Data_Analytics/\\$FILE/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf](http://www.ey.com/Publication/vwLUAssets/EY_-_Forensic_Data_Analytics/$FILE/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf).

¹⁷⁶ Das Fraunhofer SIT benennt als „flexiblere und viel wirkungsvollere Alternative [zur Betrugserkennung in Finanzdaten] ... die modellbasierte ... Benford-Analyse“; nachzulesen unter: <https://www.sit.fraunhofer.de/de/angebote/kompetenzfelder/itforensics/forensische-finanzdatenanalyse>.

¹⁷⁷ Für die Ziffernanalyse der Einerstelle in (Bar-)Einnahmen auf die Annahme des absoluten Zufalls ([Gleichverteilung](#)) werden in den SRP-Vorlagen der Summarischen Risikoprüfung (SRP) alle Werte unter € 100 vorlagengesteuert ausgeschlossen, weil bei diesen die abweichenden Erfahrungen des Benford-Gesetzes im Widerspruch zur [Hypothese](#) stünden.

Sensitivitätsanalyse zur Absicherung der Ergebnisse.¹⁷⁸ Damit sollen die „nicht behebbaren Unsicherheiten bei einzelnen Parametern – vor allem solche Unsicherheiten, die darauf beruhen, dass der Zeitreihenvergleich auf einem Vergleich der wöchentlichen Wareneinsätze beruht“¹⁷⁹ auf mögliche Auswirkungen auf das Zeitreihenresultat hinterfragt werden. Insofern ist diese Sensitivitätsanalyse von dem verbreiteten [Sensitivitätsmodell nach Vester](#) zu unterscheiden.

Bzgl. der vom BFH „Hebelwirkungen“¹⁸⁰ genannten und somit als gravierend angesehenen Verzerrungen durch Zuordnungsunschärfen in Zeitreihenanalysen ist darauf hinzuweisen, dass – sowohl in den Berechnungen als auch Ausführungen – von einer wöchentlichen Periodisierung durch das Finanzamt ausgegangen wird. Tatsächlich liegen der – tlw. neben der [Summarischen Risikoprüfung](#) (SRP) – eingesetzten 10-Wochen-[Gleitschlittenzitreihe](#) 10-Wochen-Summen¹⁸¹ für Wareneinsatz und -umsatz zu Grunde.¹⁸² Das gewählte Intervall ist entscheidend für die Verzerrungsschwere von Zeitreihenungenauigkeiten, weil bspw. die größeren Monats- gegenüber Wochensummen als Bezugswerte von Zuordnungsfehlern wesentlich kleinere Auswirkungen auf den berechneten Faktor „Rohgewinnaufschlagsatz“ zulassen.¹⁸³

Sensitivitätsmodell nach Vester

Der Biochemiker und Systemforscher Vester hat mit seinem Sensitivitätsmodell einen [musterbasierten](#) Lösungsansatz für komplexe Probleme und Fragestellungen entwickelt, welcher weltweit in der Unternehmensplanung, Entwicklungshilfe, Städteplanung, Verkehrsplanung, Sicherheitspolitik wie die Terrorbekämpfung etc. im Einsatz ist.

Weil einfache (lineare) Ursache-Wirkungs-Prinzipien in komplexen – datenintensiven oder möglicherweise verschleierte – Situationen nicht existieren¹⁸⁴, bildet er eine Matrix aus den vordefinierten systemrelevanten Variablen und ordnet – vereinfacht skaliert – die gegenseitigen Beeinflussungen ([Korrelationen](#)) zu. Aus den sich ergebenden Mustern lassen sich die komplexen Zusammenhänge wie Rückkopplungen erkennen. Zusätzlich trägt die Möglichkeit der [interaktiven](#) Simulation durch Veränderung einzelner Faktoren (bzw. Beziehungen) zur korrekten Entscheidungsfindung bei. Als wesentlichen Entwicklungsbezug nennt Vester die neuro-hormonelle Selbstregulation (Kybernetik) in der Biologie.

Signifikanz

Unter „signifikanter Abweichung“ wird in der [Statistik](#) ein ([stochastisches](#)) Übereinstimmungsergebnis verstanden, welches durch Überschreiten des [Konfidenzniveaus](#) zu einer eindeutigen Ablehnung der Annahme ([Hypothese](#)) führt (z.B. bei der [musterbasierten Ziffern-](#) oder [Verteilungsanalyse](#)).

Mit vorher festgelegten Signifikanzgrenzen wird im Bereich von Wahrscheinlichkeitsaussagen ([probabilistischen Schlussfolgerungen](#)) gearbeitet, wie sie die Vergangenheitsrekonstruktion bei Prüfungen und Gutachten erfordert. Dabei sind Ablehnungsklarheiten von 95 %, 99 % oder 99,9 % üblich, die in Steuerungs- bzw. Strafverfahren an die Anforderungen an (juristische) [Beweise](#) und [Beanstandungsanlässe](#) angepasst werden müssen.

Spitzenkappung

Die wahrscheinlichste Spur systematischer, softwareunterstützter¹⁸⁵ Steuerhinterziehung ist der verstärkte Eingriff in höhere (Bar-)Einnahmen mit entsprechenden [Musterverzerrungen](#)¹⁸⁶, weil sie sich aus der psychologischen Erkenntnis einer (stetigen) Steigerung wiederholter krimineller Handlungen ([Manipulationspsychologie](#)) ergibt.

¹⁷⁸ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 69, sowie BFH v. 12.7.2017 – X B 16/17, BFH NV 2017, 1204, Rz. 109.

¹⁷⁹ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 69.

¹⁸⁰ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 44.

¹⁸¹ Entgegen der Beschreibung durch den BFH (BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 4) wird hierbei auch nicht der [Mittelwert](#) der Wochenaufschlagsätze gebildet sondern der Gesamtaufschlagsatz für die betreffenden 10-Wochen-Zeiträume.

¹⁸² Wolenski, „Das Zeitreihenurteil des X. Senats des BFH vom 25.3.2015“, Stbg 6/2016, 268.

¹⁸³ Mit einem Beispiel in Fußnote 5: Wolenski, „Das Zeitreihenurteil des X. Senats des BFH vom 25.3.2015“, Stbg 6/2016, 268.

¹⁸⁴ Vester, „Die Kunst vernetzt zu denken – Ideen und Werkzeuge für einen neuen Umgang mit Komplexität“, Deutscher Taschenbuch Verlag, München, 9. Auflage 2012, 15: „Unser Dilemma im Umgang mit der Komplexität unserer Welt lässt sich darauf zurückführen, dass wir wohl darin ausgebildet wurden einfache logische Schlüsse zu ziehen und nahe liegende Ursache-Wirkungs-Beziehungen zu definieren. ... Simple Ursache-Wirkungs-Beziehungen gibt es nur in der Theorie, nicht in der Wirklichkeit“.

¹⁸⁵ Teutemacher, „Handbuch zur Kassenführung“, NWB Verlag Heme, 2015, 137 ff. mit einer Offenlegung einiger verbreiteter Manipulationsmethoden.

¹⁸⁶ Eine entsprechende Grafik findet sich zum Begriff [„Verteilungsanalyse“](#).

Im Rahmen der [Summarischen Risikoprüfung](#) (SRP) wird diese Form der Einnahmenverkürzung „Spitzenkappung“ genannt und mit der [Verteilungsanalyse](#) gezielt nachverfolgt. Somit wird – wie bei der Ziffernanalyse zur Entdeckung einer [Komplettmanipulation](#) – ein bestimmtes [Häufigkeitsbild](#) zur Aufdeckung einiger heutzutage verbreiteter digitaler Hinterziehungsszenarien genutzt. Durch den Einsatz von [Wahrscheinlichkeitstests](#) für die Übereinstimmungsprüfung lässt sich bei vorgefundener Auffälligkeit feststellen, ob die Spuren derart deutlich sind, dass sich unter [sehr geringer Restunsicherheit](#) ein menschlicher Eingriff annehmen lässt.

SRP (Summarische Risikoprüfung)

⇒ [Summarische Risikoprüfung](#)

SRP-Vorlagen

Die Anwendung der [Summarischen Risikoprüfung](#) (SRP) wird in ihrer Basisversion durch zwei Vorlagen in MS Excel® bzw. OpenOffice® erleichtert, indem die wichtigsten Auswertungen entlang der SRP-Prüfungssystematik¹⁸⁷ vorprogrammiert sind. Dadurch bestimmen die Prüfer weiterhin den Prüfungsprozess, benötigen jedoch deutlich weniger Softwarekenntnisse und bekommen – im Gegensatz zu einer Auswahl an Datenanalysemakros für bestimmte Prüfschritte – einen sinnvollen Ablauf diverser Einzelanalysen vorgeschlagen.

Die auf [Regeln](#) und [Muster](#) zu untersuchenden Daten werden aus einer Datenquelle – im Rahmen der Außenprüfung üblicherweise IDEA® – in die Vorlagen kopiert, wonach die Daten und Analysen aufbereitet sind. Alle Ergebnisse sind dokumentationsvorbereitet, d.h. sie enthalten alle notwendigen Informationen, um unter Angabe der untersuchten Daten mit Standardsoftware die Auswertungen nachzuvollziehen. Darüber hinaus bieten beide SRP-Vorlagen interaktive Optionen, um neue Erkenntnisse bzw. Berichtigungen zu den untersuchten Daten nachvollziehbar zu berücksichtigen, wonach alle Auswertungen automatisch angepasst sind [[interaktive Sach\(verhaltens\)aufklärung](#)].

Die SRP-Vorlage BWA führt unterschiedliche regelbasierte [Zeitreihenanalysen](#) vom groben Überblick zu detaillierteren Perspektiven durch. Auf diese Weise wird das Erkennen von Regelstörungen in veränderlichen und abhängigen Betriebsgrößen als sinnvolle Prüffelder sowie deren Eingrenzung bis zu ursächlichen Sachverhalten unterstützt.

Mit der SRP-Vorlage Stochastik werden zu den eingepflegten Daten [Häufigkeitsmuster](#) wiedergegeben. Abweichungen von gleichmäßigen Mustern können effizient Prüffelder aufzeigen, welche ebenfalls mit Hilfe der SRP-Vorlagen weiter konkretisiert werden können. Im Bereich der (Bar-)Einnahmenprüfung können Musterstörungen im Vergleich zur [Gleich-](#), [Benford-](#) bzw. [Logarithmischen Normalverteilung typische Eingriffsspuren](#) offenlegen; zur seriösen Beurteilung von festgestellten Musterabweichungen greift die SRP-Vorlage auf die stochastischen Übereinstimmungstests „[Chi-Quadrat-Test](#)“ und „[Binomialverteilung](#)“ zurück, weil die Anzahl der untersuchten Daten hierbei von großer Bedeutung ist. Außerdem erleichtern die Analysen der SRP-Vorlagen eine gezielte Abklärung, ob entdeckte Auffälligkeiten auf besondere Einzelfallumstände wie z.B. einen [Dezimalsprung](#) zurückzuführen sind.

Standardabweichung

Als absolute¹⁸⁸ Form der vorzeichenneutralen, durchschnittlichen Differenz jedes einzelnen Datensatzes zum [Mittelwert](#) soll die Standardabweichung in der beschreibenden (deskriptiven) [Statistik](#) die Streuungsstärke innerhalb einer Datenmenge wiedergeben. Wichtige (Streu-)Verteilungen wie z.B. die [Normalverteilungen](#) sind zudem durch den Mittelwert und die Standardabweichung der Daten ausreichend bestimmt.

Berechnet wird die Standardabweichung als Summe der Betragsabweichungen aller Einzeldaten zum Mittelwert, welche (anschließend) durch die Datenanzahl geteilt wird.

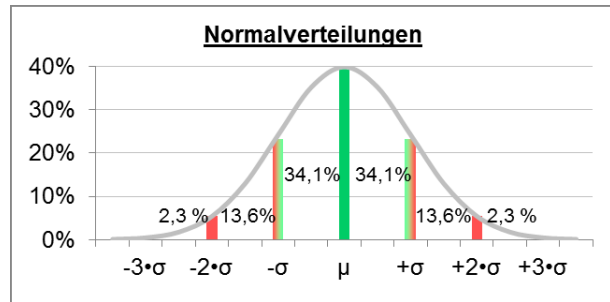
Bei der Beschreibung von – häufig asymmetrisch verteilten ([Logarithmische Normalverteilungen](#)) – Wirtschaftsdaten ist zu beachten, dass die symmetrische Standardabweichung das Streuverhalten nicht korrekt beschreiben kann.

¹⁸⁷ Wähnert, „Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz „Summarische Risikoprüfung (SRP)“, DB 45/2016, 2627.

¹⁸⁸ Die relative Form der Standardabweichung nennt sich [Variationskoeffizient](#).

Standardnormalverteilung

Die Standardnormalverteilung mit den definierten Kenngrößen [Mittelwert](#) = 0 und [Standardabweichung](#) = 1 dient als „Schablone“ für die Herleitung jeder beliebigen [Normalverteilung](#)¹⁸⁹, weil deren Rang-Wert-Verhältnisse identisch sind: Der Datennormbereich „Mittelwert (μ) \pm Standardabweichung (σ)“ umfasst ca. 68,2 % der Daten, während jeweils ca. 13,6 % „schwache“ Ausreißer bis zur doppelten Durchschnittsabweichung reichen und damit auf beiden Seiten ca. 2,3 % der Datenmenge als starke, unrepräsentative Werte verbleiben.



Wegen dieser starren Rang-Wert-Verhältnisse des bekanntesten (zufälligen) Streuverhaltens in Datenmengen werden (wissenschaftliche) Studienergebnisse häufig auch über die Mittelwert- bzw. Flächenrelationen zur Standardabweichung der [Referenz](#) kommuniziert wie z.B. bei der Verteilung des Intelligenzquotienten bestimmter Ethnien zur Vergleichsgruppe.¹⁹⁰

Statistik

Als Teilgebiet der Mathematik fasst die Statistik die theoretischen Ansätze und empirischen Erfahrungen zu größeren Datenmengen zusammen. Damit leistet sie eine wesentliche Qualifizierung für die Natur- und Wirtschaftswissenschaften, indem sie komplexe – nicht über singuläre Ursache-Wirkungs-Beziehungen erklärbar – Vorgänge vergleich- und reproduzierbar beschreibt.

Je nach Anwendungsgebiet wird zwischen deskriptiv für die Wiedergabe der charakteristischen Merkmale einer Datenmenge, explorativ (erforschend) und induktiv (ableitend) unterschieden. Die induktive Statistik (auch Inferenzstatistik) schlussfolgert entsprechend der Erkenntnisse aus der [Wahrscheinlichkeitstheorie](#) aus einer Stichprobe auf die Grundgesamtheit und ist somit zentraler Bestandteil der [Stichprobenprüfung](#).

Stetige Verteilungen

Stetig verteilte Datenmengen zeigen – im Gegensatz zu [diskreten](#) – einen kontinuierlichen Verlauf, indem auch nicht-ganzzahlige Ergebnisse möglich sind wie z.B. bei der statistischen Auswertung von Körpergrößen. Zu den stetigen Funktionen gehören die [Normalverteilungen](#) und die [Logarithmischen Normalverteilungen](#); erstere können zur kontinuierlichen Übertragung der [Binomialverteilung](#) genutzt werden.

Stichprobenprüfung

Sobald eine Vollprüfung nicht möglich oder aus anderen Gründen nicht vertretbar¹⁹¹ ist, kann eine Einschätzung der Vollständigkeit und Richtigkeit nur auf der Basis eines [Gesamteindrucks](#) oder eines Auszugs erfolgen. Die Stichprobenprüfung gründet ihr Urteil dabei auf die bewusste oder zufällige Auswahl eines Teils der Grundgesamtheit an Vorgängen.

Der bewussten Stichprobenauswahl liegen i.d.R. Kriterien wie „Erfahrung“, „Wert der Position“ etc. zu Grunde. Ein belastbarer Rückschluss auf die Grundgesamtheit ist auf diese Weise jedoch nicht möglich¹⁹², weshalb derartige Stichprobenanalysen Teilprüfungen bedeuten.

Demgegenüber beachten mathematisch-statistische Stichprobenprüfungen wie [Monetary Unit Sampling](#) (MUS) die Erkenntnisse der [Wahrscheinlichkeitslehre](#) zum mengenabhängigen¹⁹³ Streuverhalten und lassen deshalb eine aussagekräftige Hochrechnung vorgefundener Fehler auf die zu er-

¹⁸⁹ Auch [Logarithmische Normalverteilungen](#) lassen sich über die Logarithmen der Datenwerte auf diesem Weg beschreiben.

¹⁹⁰ *Hernstein/Murray*, „The bell curve – Intelligence and class structure in American life“, The Free Press New York, II. Title 1994.

¹⁹¹ Nach dem Verifikationsprinzip darf im Besteuerungsverfahren die Dauer von Einzelprüfungen nicht so niedrige Prüfungsquoten verursachen, dass die Gleichmäßigkeit der Besteuerung nicht mehr gewährleistet werden kann (BFH v. 16. 5. 2013 – II R 15/12, BStBl II 2014, 225 mit dem wörtlichen Zitat unter dem Begriff „[Beanstandungsanlass](#)“).

¹⁹² *Giezek*, „Monetary Unit Sampling – Der Einsatz statistischer Verfahren im Rahmen der Jahresabschlussprüfung“, Gabler Verlag Wiesbaden, 2011, 25 ff.

¹⁹³ Die Bedeutung der untersuchten Datenmenge für die (natürliche) Ergebnisstreuung um eine Idealerwartung ([Referenz](#)) herum kann aus der Abb. zum Begriff „[Binomialverteilung](#)“ abgelesen werden.

wartende Fehlerspanne der Grundgesamtheit zu. Aus diesem Grund sind mathematisch-statistische Stichprobenprüfungen im Handelsrecht¹⁹⁴ und in der Wirtschaftsprüfung¹⁹⁵ zugelassene Prüfverfahren.

Stochastik

⇒ [Wahrscheinlichkeitslehre](#)

Strukturanalyse

⇒ [Verteilungsanalyse](#)

Summarische Risikoprüfung (SRP)

Die Summarische Risikoprüfung (SRP)¹⁹⁶ ist ein systematisches Prüfungsnetz, welches diverse verbildlichende Datenanalyseansätze so kombiniert, dass innerhalb kurzer Zeit ein guter Überblick über eine große bzw. möglicherweise verschleierte Datenmenge erzielt werden kann.

Damit ist sie besonders für den Prüfungseinstieg geeignet, um die Vertiefungsnotwendigkeit festzustellen sowie betriebsspezifische Prüffelder aufzudecken und bis zum Einzelsachverhalt nachzuvollziehen.¹⁹⁷ Ebenso können komplexe Vorgänge transparenter aufbereitet werden, um eine qualifizierte rechtliche Beurteilung zu unterstützen wie z.B. bei Verrechnungspreisfragen.¹⁹⁸ In der Anwendung auf die Vollständigkeit von (Bar-)Einnahmen soll ein Ergebnisnetz erarbeitet werden, das – im Sinn einer sich gegenseitig absichernden [Indiziengesamtheit](#)¹⁹⁹ – eine möglichst sichere Aussage zur materiellen Richtigkeit gem. § 158 AO gewährleistet; zusätzlich ermöglichen die interaktiven Optionen der SRP-Vorlagen das zügige Einarbeiten bzw. Korrigieren der untersuchten Daten infolge weiterer Sachverhaltsaufklärung [[interaktive Sach\(verhaltens\)aufklärung](#)].

Über die [SRP-Vorlagen](#) ist die Summarische Risikoprüfung in einen [regel-](#) und einen [musterbasierten](#) Bereich unterteilt. Im regelbasierten Teil werden unterschiedliche Zeitreihenvarianten so verknüpft, dass prüfungsrelevante Größenänderungen und -unabhängigkeiten (Regelstörungen) möglichst effektiv erkannt und effizient eingegrenzt werden können. Mit Hilfe von zielgerichtet vernetzten Musteranalysen können zudem auffällige Wert- bzw. Betragshäufigkeiten als Prüffelder festgestellt werden. In Vollständigkeitsprüfungen von (Bar-)Einnahmen werden die Musteranalysen speziell dafür eingesetzt, um [typische Eingriffsspuren](#) von Einnahmeverkürzungen zu suchen. Bei der systematischen Prüffeldsuche werden die beiden SRP-Teile auch gezielt kombiniert, um Einzelsachverhalte aufzufinden.²⁰⁰ Die Aufarbeitung komplexer Prüfungsfragen setzt häufig eine über die SRP-Vorlagen hinausgehende freie Datenvisualisierung voraus.

Variationskoeffizient

Als relative [Standardabweichung](#) soll der Variationskoeffizient auch einen Vergleich der Streuungsintensität zwischen Datenmengen mit unterschieden Wertenniveaus ermöglichen.

Dazu wird die Standardabweichung durch den Mittelwert geteilt und als prozentuale Streuungsstärke wiedergegeben.

Verfahrensdokumentation

Zur Überprüfung der formellen Ordnungsmäßigkeit von Bestimmungsaufzeichnungen (als Voraussetzung der [Richtigkeitsvermutung](#) erklärter Besteuerungsgrundlagen) ist das Nachvollziehen der Aufzeichnungs- und Aufbewahrungsprozesse durch die Betriebsprüfung erforderlich. Um dies auch für die immer umfangreicher digitalisierten Abläufe sicherzustellen, wird bereits in den „Grundsätze[n] ordnungsmäßiger EDV-gestützter Buchführungssysteme (GoBS)“ eine Verfahrensdokumentation als Pflichtbestandteil angeführt.²⁰¹

In den GoBD wird die Funktion der Verfahrensdokumentation als Beschreibung aller personellen und programmtechnischen Aufzeichnungs- und Aufbewahrungsprozesse bestehend aus „einer allge-

¹⁹⁴ § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden.“

¹⁹⁵ IDW PS 310 „Repräsentative Auswahlverfahren (Stichproben) in der Abschlussprüfung“, s.: <https://www.idw.de/idw/verlautbarungen/idw-ps-310/43204>.

¹⁹⁶ Eine umfangreichere, allgemein zugängliche Beschreibung der Summarischen Risikoprüfung (SRP) mit Hinweisen auf weiterführende Zeitschriftenbeiträge ist zu finden unter: https://de.wikipedia.org/wiki/Summarische_Risikopruefung.

¹⁹⁷ Wähnert, „Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz „Summarische Risikoprüfung (SRP)““, DB 45/2016, 2627.

¹⁹⁸ Werner/Wähnert, „Verbildlichende Prüfung am Beispiel der Gewinnverlagerung – Die Überlegenheit der Visualisierung von Datenregeln und -verteilungen gegenüber Kennzahlen“, IStR 19/2016, 811.

¹⁹⁹ BGH v. 27.4.2010 – 1 StR 454/09, NSiZ 2011, 108.

²⁰⁰ Wähnert, „Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz „Summarische Risikoprüfung (SRP)““, DB 45/2016, 2627.

²⁰¹ BFM-Schreiben v. 7.11.1995, BStBl 1995 I, 738, Pkt. VI.

meinen Beschreibung, der Anwenderdokumentation, der technischen Systemdokumentation und einer Betriebsdokumentation“ näher definiert, wenngleich alleinige Schwächen der Verfahrensdokumentation „kein[en] formelle[n] Mangel mit sachlichem Gewicht“ darstellen.²⁰²

Verifikationsprinzip

⇒ [Richtigkeitsvermutung](#)

Vermögenszuwachsrechnung

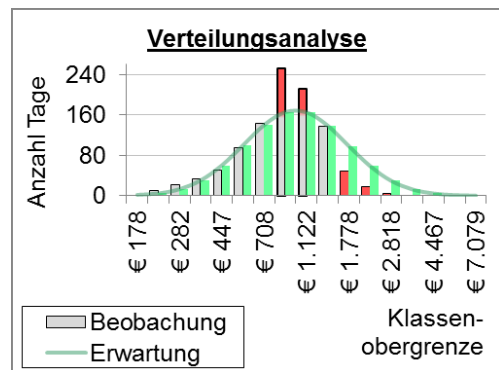
Bei der Vermögenszuwachsrechnung als Form der [Geldverkehrsanalyse](#) liegt der Fokus nicht auf der Logik, den laufenden Geldbedarf dem verfügbaren -zufluss gegenüberzustellen, sondern die Vermögensmehrungen auf die Mittelherkunft zu hinterfragen. Auf diese Weise soll ebenso die Vollständigkeit der erklärten Einnahmen verprobt werden.²⁰³

Auffällige Geldverkehrsrechnungen erfüllen nach Ansicht des BFH²⁰⁴ die Anforderungen an einen Beanstandungsanlass gem. § 158 AO. In der Drei-Stufen-Theorie des BFH werden [Geldverkehrs-](#) und [Vermögenszuwachsrechnungen](#) sowie Kalkulationen den Zeitreihenanalysen als Schätzungsmethode (und wohl auch als Verprobungsmethode) vorgezogen, weil sie den realistischen Besteuerungsgrundlagen näher kämen.²⁰⁵ Hierzu ist anzumerken, dass es sich bei Vermögenszuwachsrechnungen einerseits um arbeitsaufwendige [Gesamtbildmethoden](#) handelt, auf die nur ordnungsgemäße Aufzeichnungen einen Anspruch haben²⁰⁶. Andererseits ist die Ergebnisgenauigkeit wesentlich von der Vollständigkeit und Richtigkeit der Ausgangsgrößen abhängig, wobei private Geldzu- und -abflüsse sowie Vermögensmehrungen nach Abschaffung der Vermögenssteuer nur schwierig nachzuverfolgen sind und zudem das verbreitete²⁰⁷ Steuerhinterziehungsmodell, Schwarzlöhne durch nicht erklärte Einnahmen zu bezahlen, im Rahmen von Betriebsprüfungen kaum ausgeschlossen werden kann.

Verteilungsanalyse

Mit der Verteilungsanalyse [als Teil der [Summarischen Risikoprüfung](#) (SRP)] wird die [Häufigkeitsverteilung](#) von nach Größenklassen ausgezählten ([klassifizierten](#)) Besteuerungsdaten [musterbezogen](#) einer betriebsnah rekonstruierten²⁰⁸ [Logarithmischen Normalverteilung](#) gegenübergestellt, um einerseits Einzelprüffelder effizient zu erkennen und andererseits das typische Manipulationsbild der [Spitzenkappung](#) indiziell ([Beanstandungsanlass](#)) aufzudecken; sie ähnelt der Strukturanalyse.²⁰⁹

Dem nebenstehenden Diagramm sind die beobachteten und erwarteten Häufigkeiten von Tagen im jeweiligen Bareinnahmenbereich zu entnehmen. Infolge der logarithmischen²¹⁰ Klassifizierung bildet sich als [Referenz](#) eine symmetrische [Normalverteilung](#) aus, die an den wertärmeren²¹¹ (linksseitigen) Balken der ausgezählten Tagesbareinnahmen fortentwickelt wurde. Im Analyseergebnis ist deutlich zu sehen, dass die kurzen Balken im Hocheinnahmebereich (rechte Seite) eine Unterbesetzung zu



²⁰² „Grundsätze zur ordnungsgemäßen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD)“, BMF-Schreiben v. 14.11.2014, BStBl I 2014, 1450, Pkt. 10.1, Rz. 153 und 155.

²⁰³ Kritisch hierzu: *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, Berlin, 4. Auflage 2017, 238.

²⁰⁴ BFH v. 21.2.1974 – I R 65/72, BStBl II 1974, 591.

²⁰⁵ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 60 ff.

²⁰⁶ BFH v. 21.2.1990 – X R 54/87, BFH/NV 1990, 683.

²⁰⁷ Diese Erkenntnis hat dazu geführt, dass in der österreichischen Strategie zur Bekämpfung von Kassenmanipulationen die Bareinnahmenprüfung der dortigen Finanzkontrolle Schwarzarbeit via Kassennachschauen zugewiesen wurde; vgl. *Huber*, „Praktische Erfahrungen mit der Kassenführung in Österreich – Umsetzung der Kassenrichtlinie 2012 und Kassennachschau durch die Finanzpolizei“, BBK 6/2014, 286.

²⁰⁸ Die spezifische Logarithmische Normalverteilung wird aus den Kenngrößen „Mittelwert“ und „Standardabweichung“ der untersuchten Daten gebildet.

²⁰⁹ Im Gegensatz zur Strukturanalyse, die mit der – auf 100 % zulaufenden – kumulativen Verteilungsfunktion arbeitet, stellt die Verteilungsanalyse die klassenbezogenen Häufigkeiten der untersuchten Daten denen der abgeleiteten Logarithmischen Normalverteilung gegenüber, um Vermischungseffekte zu vermeiden. Die Strukturanalyse geht auf den ehemaligen Betriebsprüfer und Dozenten der Bundesfinanzakademie Wilhelm Baselt zurück (*Schmitz/Gebbers/Huber*, „Nachruf auf Wilhelm Baselt – ‚Vater‘ der modernen steuerlichen Prüfungstechnik“, StBp 4/2011, 93) und wurde von *Gebbers* ausführlich vorgestellt: „Die Strukturanalyse – Teil I bis III“, StBp 8-10/2007, 225, 257 und 289.

²¹⁰ Eine logarithmische Klassifizierung wie auf der X-Achse der Beispielsgrafik teilt in exponentiell lineare Intervalle ein: „ $10^{2,25}$, $10^{2,35}$, $10^{2,45}$, $10^{2,55}$... $10^{3,85}$ “; die dazugehörigen linearen Klassenobergrenzen „178, 224, 282, 355, ..., 7.079“ (Skalierung nur alle zwei Klassen) bedeuten kontinuierlich zunehmende Klassenspannen.

²¹¹ Dieser Teil der Einnahmenstruktur gilt aus [psychologischer](#) Sicht sowie aufgrund praktischer Zwänge als weniger manipulationsgefährdet.

Gunsten der mittleren Klassen aufweisen. Um die von der Datenanzahl abhängige Referenztreue der Streuung²¹² zu berücksichtigen, wird das Vergleichsergebnis im Rahmen der Summarischen Risikoprüfung (SRP) mit einem Wahrscheinlichkeitstest gegen unkritische Abweichungen (s. [Signifikanz](#)) abgesichert.²¹³

Sofern als betriebliche Besonderheit – und Ursache für die Nichtübereinstimmung – eine häufige Vollausslastung (weitgehend) ausgeschlossen werden kann²¹⁴, deutet das Analyseresultat aus Schlüssigkeitssicht auf die kriminalpsychologisch wahrscheinlichste Spur einer (digital unterstützten) Einnahmenverkürzung hin: Ein stärkerer Eingriff in die höheren Bareinnahmen (Spitzenkappung).

Waffengleichheit

⇒ [Chancengleichheit](#)

Wahrscheinlichkeitslehre (Stochastik)

Die Wahrscheinlichkeitslehre – als Teilgebiet der Stochastik – fasst die empirischen und theoretischen Erkenntnisse zu unsicheren Ereignissen (s.a. [Kombinatorik](#)) zusammen. Auf dieser Grundlage können Wahrscheinlichkeiten zum Auftreten bestimmter Beobachtungen zwischen „unmöglich“ und „sicher“ berechnet und somit [probabilistische Schlussfolgerungen](#) mit Aussagegewissheiten abgeleitet werden.

Als bestmögliche Beschreibung der nicht vollständig und objektiv rekonstruierbaren Vergangenheit, komplexer Vorgänge sowie zukünftiger Ereignisse findet die Wahrscheinlichkeitslehre – i.d.R. in Form von [Wahrscheinlichkeitstests](#) – in fast allen Lebensbereichen Anwendung wie bspw. der Absicherung von Studienergebnissen²¹⁵, im Rahmen der industriellen Qualitätssicherung²¹⁶, in der beweisfähigen Stichprobenprüfung²¹⁷ und der automatischen Prozesssteuerung²¹⁸.

Wahrscheinlichkeitstest

Wahrscheinlichkeitstests wie der [Chi-Quadrat-](#) und der [z-Test](#) sind vorkonfektionierte Hilfsmittel der [Wahrscheinlichkeitslehre](#), um deren (empirische und theoretische) Erkenntnisse – leicht anwendbar – auf eine bestimmte [Hypothese](#) zu übertragen wie bspw. bei der Annahme, dass die Ziffernverteilungen der [schwebenden Anfangspositionen](#) in (nachfragebedingt) zufällig entstandenen Bareinnahmen der [Benford-Verteilung](#) folgen ([Ziffernanalyse](#)). Als Ergebnis wird eine Übereinstimmungswahrscheinlichkeit ausgegeben, welche eine qualifizierte Beurteilung (s. [probabilistische Schlussfolgerung](#)) ermöglicht.

Weil Wahrscheinlichkeitstests neben der Form des natürlichen Streuverhaltens auch die Abhängigkeit von der Anzahl an Vorgängen²¹⁹ berücksichtigen, gelten stochastische Bewertungen unter Unsicherheit, wie sie im Prüfungs- und Gutachtenbereich üblich sind, als „qualifiziert“. Sie sind objektiv nachvollziehbar und generell akzeptiert.²²⁰

χ^2 -Test

⇒ [Chi-Quadrat-Test](#)

z-Test

Der z-Test – auch Gauß-Test – ist ein [Wahrscheinlichkeitstest](#), um eine Datenstichprobe auf die Zugehörigkeit zu einer bekannten [normalverteilten](#) Grundgesamtheit zu testen. So können bspw. Umsatzdaten, die mittels Nachschau als Stichprobe gewonnen wurden, anhand ihrer Logarithmen da-

²¹² Aus der Grafik zum Begriff „[Binomialverteilung](#)“ ist dieser Zusammenhang ablesbar.

²¹³ Die Summarische Risikoprüfung (SRP) arbeitet hierbei mit dem [Chi-Quadrat-Test](#) (nur für den Hauptbereich mit Klassenerwartungen ≥ 5). Hierdurch wird eine vorsichtiger Beurteilung als bei üblichen Übereinstimmungstests zur Normalverteilung wie bspw. dem Anderson-Darling-Test gewährleistet.

²¹⁴ Im abgebildeten Fall kommen Tage mit betriebsbezogen hohen Bareinnahmen (rechte rote Balken) vor, sind nur deutlich seltener als aus der Nachfragekonstruktion zu erwarten ist. Sollte es für diese – problemlos identifizierbaren – Tage keine strukturellen Unterschiede geben, die das ausnahmsweise Erzielen der Einnahmehöhe begründen können, ist eine Kapazitätsgrenze (häufige Vollausslastung) als Ursache für die Musterabweichung nicht nachvollziehbar.

²¹⁵ Vgl. z.B. *Bortz/Schuster*, „Statistik für Human- und Sozialwissenschaftler“, Springer-Verlag Berlin Heidelberg, 2010, 11.

²¹⁶ DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/norm/din-iso-2859-1/191697831>.

²¹⁷ § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden.“ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

²¹⁸ Vgl. z.B. *Richter*, „Prinzipien der Künstlichen Intelligenz“, Teubner-Verlag Stuttgart, 1992.

²¹⁹ Deshalb sind bei [musterbasierten Datenanalysen](#), die mit einem Wahrscheinlichkeitstest beurteilt werden, keine allzu großen Mindestdatenmengen erforderlich und lassen sich häufig aus den Anwendungsempfehlungen berechnen wie beim [Chi-Quadrat-Test](#). Aus der Grafik zum Begriff „[Binomialverteilung](#)“ ist der Zusammenhang zwischen Streuungsintensität und Datenanzahl ablesbar.

²²⁰ Eine beispielhafte Auflistung der Anwendungsbereiche findet sich unter dem Begriff „[Wahrscheinlichkeitslehre](#)“.

raufhin geprüft werden, ob sie zur umfangreicheren Werteverteilung ([Logarithmische Normalverteilung](#)) aus einer Betriebsprüfung passen.

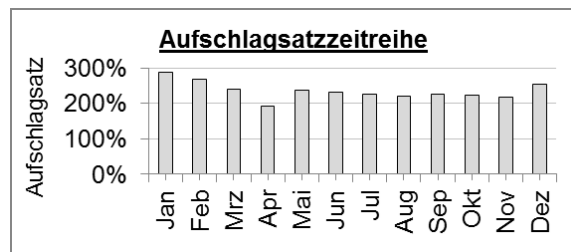
Als vorkonfektionierte Anwendung aus der [Stochastik](#) berechnet der z-Test ein Abweichungsmaß, das die Übereinstimmung im Wertespektrum der [Standardnormalverteilung](#) wiedergibt²²¹ und eine qualifizierte Beurteilung (s. [probabilistische Schlussfolgerung](#)) nach den Erkenntnissen der Wahrscheinlichkeitslehre ermöglicht. Die bekanntermaßen stärkere Streutendenz kleinerer Datenmengen²²² fließt dabei in die Berechnung des z-Werts bereits ein.

Mit der stochastischen Hochrechnung eines qualifizierten Stichprobenergebnisses auf die Grundgesamtheit der relevanten Sachverhalte hat der Bundesgerichtshof²²³ indirekt auch die Übereinstimmungsprüfung durch einen z-Test anerkannt, weil beide Ansätze der [Statistik](#) auf denselben Grundlagen fußen.

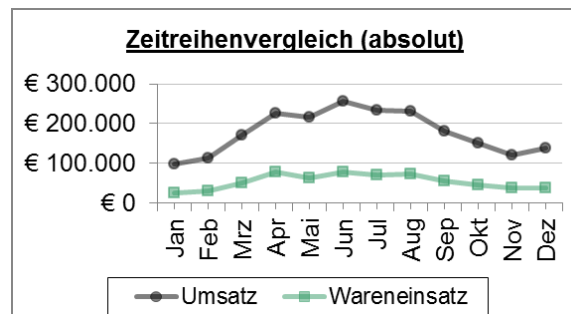
Zeitreihenanalyse

Das chronologische Nachverfolgen der Entwicklung einer Betriebsgröße bzw. der Abhängigkeit mehrerer ([Korrelation](#)) ist ein beliebtes Mittel u.a. in der Wirtschaftsprüfung²²⁴, um wirksam und effizient zu prüfen resp. planen. Im Hinblick auf die Aussage kommt es bei dem variantenreichen²²⁵ Instrument wesentlich auf den Zeitreihentyp und die Ausgestaltung an, wie die Gegenüberstellung dreier Zeitreihenanalysen zum selben Datenbestand demonstriert:

Eine Aufschlagsatzzeitreihe²²⁶ gibt die Veränderung des Faktors aus Wareneinsatz und -umsatz (Aufschlagsatz) wieder. Diese Art Zeitreihenanalyse war vor Einführung der [Summarischen Risikoprüfung](#) (SRP) die verbreitetste im Revisionsbereich, um die Schlüssigkeit von Warenflüssen gem. § 158 AO zu beurteilen ([Beanstandungsanlass](#)). In Rechtsprechung und Literatur wurden jedoch Bedenken wegen der Verzerrungsanfälligkeit durch Sondersachverhalte wie z.B. Bestandseinkäufe geäußert.²²⁷



Mit der digitalen Datenanalyse kommen vermehrt auch vergleichende Gegenüberstellungen beider Einzelgrößen zum Einsatz. Hierbei muss streng das Prüfungsziel differenziert werden: Sollen zwei Datenreihen auf Identität hin untersucht werden wie bspw. bei der Überprüfung von Aufzeichnungsschnittstellen ([Schnittstellenverprobung](#)), ist Deckungsgleichheit ([Kongruenz](#)) der Grafen zu erwarten. Wird allerdings die Abhängigkeit zweier Betriebsfaktoren im Sinn kaufmännischer Handlungslogik angenommen, kann eine (absolute²²⁸) Gegenüberstellung die (relative) Korrelation – trotz des grundsätzlichen Vorteils der Datenvisualisierung²²⁹ – nicht gut nachvollziehbar aufzeigen.



Erst die Größenanpassung beider Betriebsfaktoren auf ein vergleichbares Niveau z.B. mittels [Doppeleskalierung](#) ermöglicht eine bildhafte Analyse, ob der unternehmerisch beeinflussbare Wareneinsatz gleichgerichtet und adäquat auf den vom Markt aufkotroyierten Umsatz reagiert. Der statistische Indikator für die Abhängigkeit zweier Datenreihen „[Korrelationskoeffizient](#)“ bestätigt mit dem fast perfekten Grad von 98,7 % für die Beispieldaten, dass diese echte Korrelationsprüfung eine erheblich qualifi-

²²¹ Die Variable der Standardnormalverteilung wird – in Abweichung vom üblichen „x“ – mit „z“ benannt, um Eindeutigkeit zu erreichen.

²²² Diese Besonderheit des natürlichen Streuverhaltens lässt sich an der [Kombinatorik](#) der [Binomialverteilung](#) nachvollziehen (s. Abb. zu dem Begriff).

²²³ BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

²²⁴ IDW Prüfungshinweis „Einsatz von Datenanalysen im Rahmen der Abschlussprüfung“ (IDW PH 9.330.3).

²²⁵ BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 37.

²²⁶ Auch die [Gleitschlitzenzeitreihe](#), auf die sich das BFH-Urteil v. 25.3.2015 – X R 20/13 (BStBl II 2015, 743) bezieht (RiBFH Nöcker, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse“, NWB 48/2015, 3548) gehört zu den Aufschlagsatzzeitreihen.

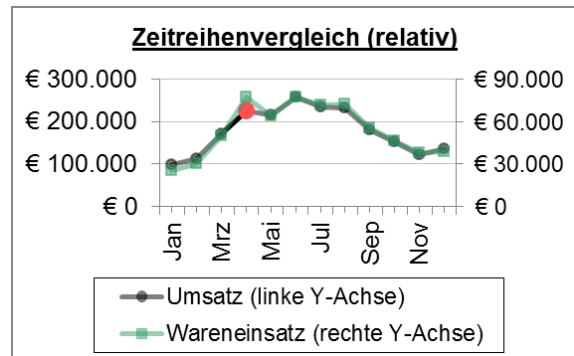
²²⁷ Wigger, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 6/2008, 168.

²²⁸ Für beide Datenreihen (mit absoluten Werten) wird dieselbe Y-Achse verwendet.

²²⁹ Barnard in einer US-amerikanischen Zeitschrift der Werbebranche 1921: „One look is worth a thousand words“, nachzulesen auf der Internetseite des Ostasieninstituts der Hochschule Ludwigshafen am Rhein, abrufbar unter: <http://www.oai.de/de/component/content/article/45-publikationen/sprichwort/811-ein-bild-sagt-mehr-als-tausend-woerter.html>.

ziertere Grundlage zur Beurteilung der Schlüssigkeit gem. § 158 AO bildet.²³⁰ Zusätzlich lassen sich einzelfallbezogene Prüffelder mit dem Beispiel „niedriger Umsatz im April“ effektiv und effizient aus dem Diagramm ablesen – im Sinn des § 7 BpO.

Durch [vorlagenunterstützte](#) Prüfungsnetze wie die Summarische Risikoprüfung (SRP) eröffnet die digitale Datenanalyse darüber hinaus die Option, neue Erkenntnisse – bspw. auch infolge konkreter Einlassungen – in eine Plausibilitätsprüfung jederzeit [interaktiv](#) einzupflegen und damit die gemeinsame Verantwortung zur Sachverhaltsaufklärung²³¹ umzusetzen.



Zeitreihenbasierte Schätzung

Zeitreihenbasierte Schätzungsmethoden wie die [Quantilsschätzung](#) verwenden für die Feststellung einer (sachgerechten) Schätzungshöhe – nach eingetretener Amtspflicht der Schätzung – die Erkenntnisse aus einer (verprobenden) [Gesamtbildprüfung](#) mit Hilfe von [Zeitreihenanalyse\(n\)](#), um den Vorteil der „Robustheit“²³² von Kennzahlen gegenüber Hochrechnungen auf der Basis unvollständiger oder evtl. falscher Ursprungsdaten zu nutzen; damit beziehen sich die derzeitigen zeitreihenbasierten Schätzungsansätze auf ordnungswidrige Aufzeichnungen, die den Verdacht materieller Fehler erwecken.²³³

Als interne Schätzungsmethoden sind zeitreihenbasierte nach ständiger Finanzrechtsprechung externen Ansätzen grundsätzlich überlegen.²³⁴ Damit aus der hervorgehenden Mehrzahl von möglichen Schätzungsparametern ein sinnvoller „Schätzungsrahmen“²³⁵ abgeleitet werden kann, sind jedoch einige Überlegungen und Arbeitsschritte erforderlich: Zur rang- bzw. wertmäßigen Einordnung der betriebsinternen Daten sowie zur anschließenden Bewertung hinsichtlich ihrer Schätzungseignung bieten sich hierfür die üblichen Methoden des mathematischen Spezialgebiets der „[Statistik](#)“ an.²³⁶

Zeitreihenvergleich

⇒ [Zeitreihenanalyse](#)

Ziffernanalyse

Die Ziffernanalyse ist ein Prüfinstrument, welches zunächst in der Wirtschaftsprüfung²³⁷ zur Aufdeckung menschlicher Eingriffe in Unternehmenszahlen wie insbesondere Unterschlagungen eingesetzt wurde. Dabei gründet sich die Wirksamkeit der Methode auf die [manipulationspsychologische](#) Erkenntnis, dass sich beim freien Erfinden von Zahlen Ziffernvorlieben und -abneigungen unterbewusst bis zur mathematischen [Signifikanz](#) einschleichen können.²³⁸

Die Durchführung einer Ziffernanalyse besteht aus der beobachtenden Auszählung der Daten nach einer [schwebenden Front-](#) oder einer [festen Endposition](#), der Gegenüberstellung einer [Referenz](#) sowie

²³⁰ Wolenski zeigt in „Das Zeitreihenurteil des X. Senats des BFH vom 25.3.2015“ (Stbg 6/2016, 268, 271) eindrücklich, dass aussagekräftige Schlüssigkeiturteile auf diesem Weg auch für Betriebe möglich sind, die nicht in das bisherige Muster der Finanzrechtsprechung fallen (BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 55), weil sie z.B. mit haltbaren Waren handeln.

²³¹ Seer, Tipke/Kruse, AO/FGO, § 158 AO Rz. 14.

²³² Unter der „Robustheit“ von Kennzahlen wird verstanden, dass sie nicht bzw. wenig anfällig gegenüber alltäglichen Sachverhalten sind; Giezek/Wähner, „Schätzungsmethoden der Betriebsprüfung im Vergleich – Einordnung der wichtigsten Schätzungsverfahren durch Praxis und Wissenschaft, BBK 21/2017, 998.

²³³ Dieser Anwendungsbereich geht konform mit der Drei-Stufen-Theorie (BFH v. 25.3.2015 – X R 20/13, BStBl II 2015, 743, Rz. 60 ff. und RiBFH Nöcker, „Update Betriebsprüfung – Neues vom BFH und die Auswirkungen für die Praxis“, NWB 42/2016, 3157) und wurde z.B. in Schumann/Wähner, „Die Quantilsschätzung – Schätzung anhand eines objektivierte Leistungsmaßes“ (Stbg 2012, 535, 538) so vorgestellt.

²³⁴ RiBFH Nöcker, „Update Betriebsprüfung – Neues vom BFH und die Auswirkungen für die Praxis“, NWB 42/2016, 3157, 3165, sowie vergleichbar: Seer, Tipke/Kruse, AO/FGO, § 158 AO Rz. 20.

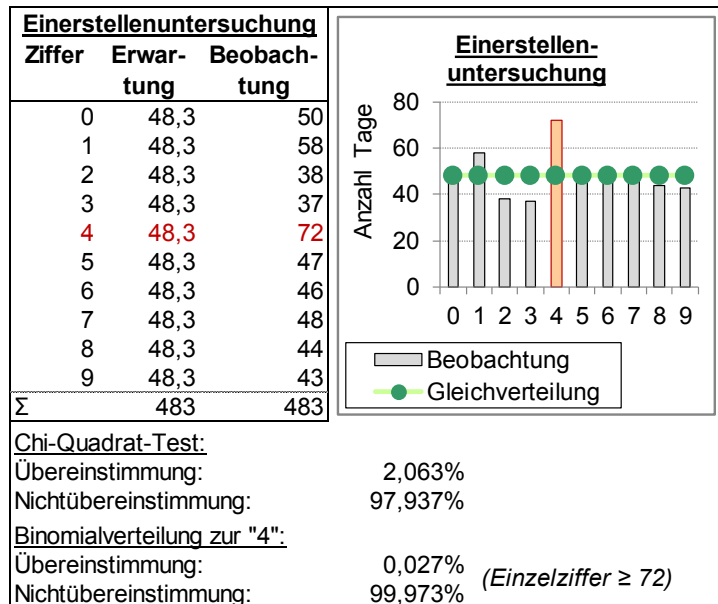
²³⁵ Seer, Tipke/Kruse, AO/FGO, § 162 AO Rz. 44: „Naturgemäß besteht bei der Schätzung nach Wahrscheinlichkeitsgrundsätzen eine Bandbreite möglicher Wertansätze (sog. Schätzungsrahmen). Soweit sich die Schätzung innerhalb des Rahmens bewegt, den der durch die FinBeh. unter Mitwirkung des Stpfl. festgestellte Sachverhalt vorgibt, ist sie nicht zu beanstanden ... Der Schätzungsrahmen ist umso größer, je ungesicherter das Tatsachenmaterial ist, auf dem die Schätzung basiert.“

²³⁶ Wähner, „Schätzungen mit Hilfe von Zeitreihenanalysen – Schrittweise Beschreibung der Quantilsschätzung“, StBp 11/2017, 323.

²³⁷ Diese Einsatzmöglichkeit der Ziffernanalyse wurde frühzeitig von Nigrini beschrieben (nachzulesen unter: http://www.nigrini.com/datas_software.htm) und bspw. von Ernst & Young in entsprechende Prüfungsschemata übernommen (nachzulesen unter: [www.ey.com/Publication/vwLUAssets/EY_-_Forensic_Data_Analytics/\\$FILE/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf](http://www.ey.com/Publication/vwLUAssets/EY_-_Forensic_Data_Analytics/$FILE/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf)).

²³⁸ Hill, „Random-number guessing and the first digit phenomenon“, Psychological Reports 62/1999, 967; Hill, „The difficulty of faking data“, Chance 12/3/1999, 27.

der Übereinstimmungsbewertung. Im abgebildeten Beispiel wurde die (feste) Einerstelle von Bareinnahmen nach den zehn möglichen Ziffern ausgezählt (Beobachtung). Als ideale Punkterwartung (Referenz) steht dem die [Gleichverteilung](#) aufgrund der theoretischen Erwägungen der [Kombinatorik](#)²³⁹ sowie der empirischen Erkenntnisse der [Benford-Verteilung](#) für Zahlen ≥ 100 gegenüber. Um bei der Übereinstimmungsbewertung die von der Datenanzahl abhängige Abweichungstendenz²⁴⁰ zu beachten, wird mit einem [Wahrscheinlichkeitstest](#) auf die mathematische Spezialdisziplin zurückgegriffen: Für die Gesamtbetrachtung aller Ziffern der [Chi-Quadrat-Test](#) resp.



für die Einzelperspektive einer Ziffer die [Binomialverteilung](#)²⁴¹. Als Resultat der qualifizierten Ziffernanalyse wird mit einer Aussagegewissheit von 97,9 % festgestellt, dass die gesamte Ziffernverteilung nicht rein zufällig entstanden ist und dass die beobachtete Einzelziffernhäufung „4“ zu 99,9 % nicht auf den Zufall zurückgehen kann. Als mögliche Ursachen unzufälliger Entstehung kommen Datenzuordnungsfehler, Gestaltungen²⁴² oder Manipulationen in Form des freien Erfindens in Betracht. Sofern die ersten beiden Begründungen nicht vorliegen, resultiert ein Indiz der berechneten Beweiskraft (s. [probabilistische Schlussfolgerung](#)) auf manipulierte Besteuerungsdaten, welches im Kontext des § 158 AO ([Beanstandungsanlass](#)) einzuordnen ist.

[Vorlagengestützte](#) Prüfungsnetze wie die [Summarische Risikoprüfung](#) (SRP) können (weitgehend) sicherstellen, dass derartige [musterbasierte](#) Manipulationstests grundsätzlich mit einem Wahrscheinlichkeitstest abgesichert und die alternativen menschlichen Eingriffsformen untersucht werden. So gehört zur Einerstellenprüfung der SRP die Abklärung, ob im Beispielsfall dominierende Einzelpreise ein verstärktes Vorkommen der „4“ bewirkt haben können und ob Veranstaltungstage mit auf „4“ endenden Pauschalpreisen vorliegen. Auch bei der weiteren Ziffernanalyse der SRP²⁴³ zur Zweitziffer wird der praxisrelevante betriebliche Faktor „begrenzter Zahlenraum“ mit dem Schwerpunkt [„Dezimalsprung“](#) gezielt hinterfragt.

Zweitziffer

⇒ [Schwebende Zahlenpositionen](#) sowie [Benford-Verteilung](#)

²³⁹ Die tägliche Multiplikation und Addition einer Vielzahl von Verkäufen zu deren Einzelpreisen führt in den meisten Fällen der Preisgestaltung zu einer rein zufälligen Entstehung der Einerstelle und damit zur [Gleichverteilung](#).

²⁴⁰ Die Bedeutung der untersuchten Datenmenge für die (natürliche) Ergebnisstreuung um eine Idealerwartung ([Referenz](#)) herum kann aus der Abb. zum Begriff [„Binomialverteilung“](#) abgelesen werden.

²⁴¹ Zur Beurteilung der zu starken Präsenz der Ziffer „4“ wird die [Binomialverteilung](#) als kumulierte Flächenfunktion für den Bereich „72-mal bis alle Datensätze (483)“ angewendet.

²⁴² Unter betriebliche Gestaltungen, welche die freie Entwicklung der Tagesbareinnahmen behindern, sind bzgl. der Einerstelle zwei Faktoren zu nennen: Zum einen können einige Preisgestaltungen wie besonders (fast) ausschließlich auf „0“ und/oder „5“ endende Einzelpreise ohne Nachkommastellen die Erreichbarkeit der Ziffern in den Tagessummen beschränken. Zum anderen können Tage mit sehr wenigen Vorgängen zu einem begrenzten Ziffernspektrum führen wie insbesondere bei Tagesveranstaltungen im Gastronomiebereich, wo der Pauschalpreis die Kombinatorik ersetzt.

²⁴³ Kritik an der Summarischen Risikoprüfung (SRP) aus der Literatur, die sich auf die Erstziffernuntersuchung bezieht [z.B. *Petersen*, „Benford's Law und die Betriebsprüfung – Eine Erwiderung auf Wähner's „Interaktive Außenprüfung – Möglichkeiten der gemeinsamen Sachverhaltsaufklärung im digitalen Zeitalter“ (Stbg 2014 S. 20 ff.) aus wissenschaftshistorischer Perspektive“, Stbg 12/2015, 506] ist somit gegenstandslos, weil diese von Beginn an nie Bestandteil der SRP gewesen ist (klargestellt u.a. in: *Wähner*, „Eine Antwort auf Petersens ‚wissenschaftshistorische‘ Betrachtung des Benford's Law: Sachgerechte Maßstäbe für die Neue Prüfungstechnik“, Stbg 12/2015, 511).