Datenanalyse in der Außenprüfung

Methodische Fallstricke

IHR REFERENT — CHRISTIAN WESTPHAL

- **Diplom-Volkswirt**
  quantitative Wirtschaftsforschung
  Univ. Kiel, 2006
- **Promotion**
  Statistiklehrstuhl WiWi-Fakultät
  Univ. Marburg, 2013
- **Steuerberater**
  Klein Nordende, 2018

Folien und Programmcode herunterladbar
https://www.westphal.de/stat/
ab 01. September 2020

STATISTISCHE METHODEN IN DER AP

Warum?

- Statistik ist nicht Selbstzweck
- kann helfen bei
  - Gleichmäßigkeits- (85 AO)
  - Wirtschaftlichkeits-grundsatz (88 V S. 2 AO)

Voraussetzungen:

- Korrekte Methodenwahl
- Methodenverständnis
- Berücksichtigung der Schwächen

AGENDA

1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. **Einführung statistisches Testen, Verteilungen**
5. **Zifferntests**
6. **Test auf Lognormalverteilung**
7. **Was sonst noch schiefgehen könnte**
8. Schlussfolgerungen

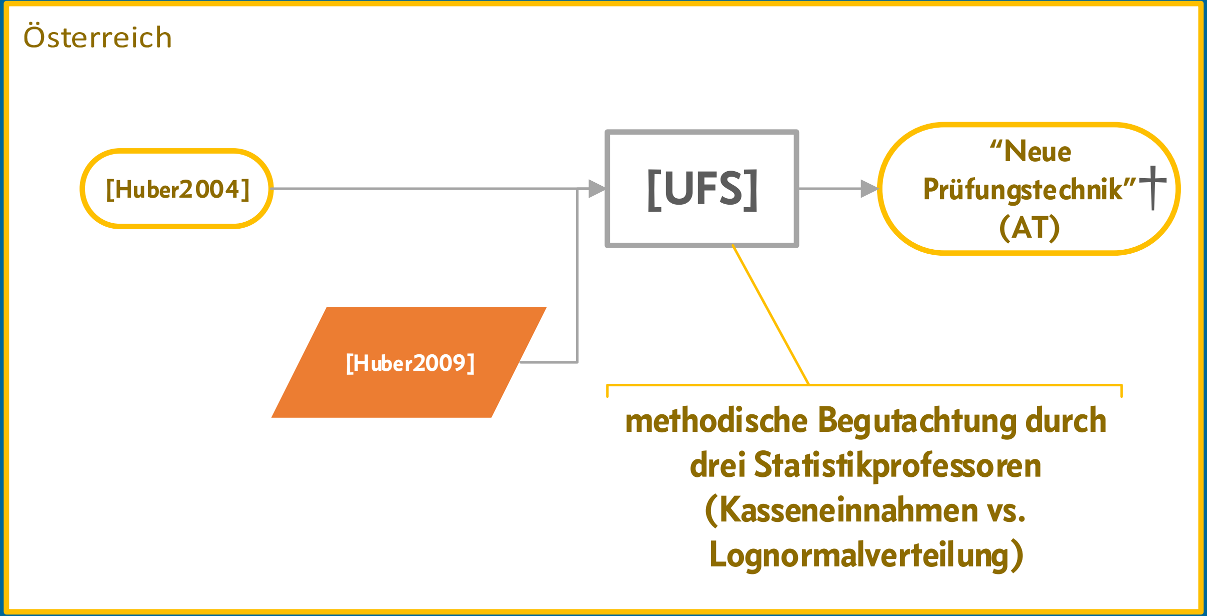
AGENDA

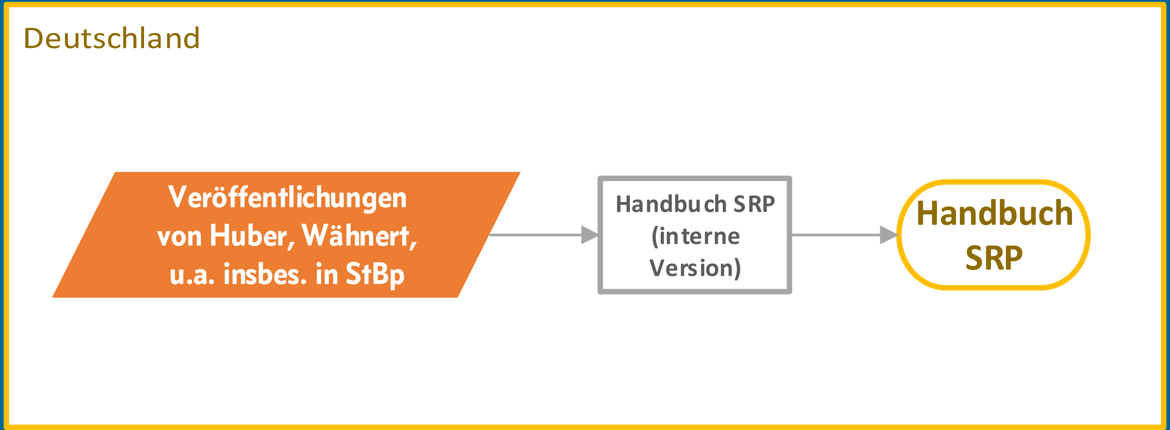
1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

KNOW YOUR ENEMY

Herkunft und Ausmaße des Problems

HERKUNFT


Österreich
[Huber2004]
[UFS]
"Neue Prüfungstechnik"† (AT)
[Huber2009]
methodische Begutachtung durch drei Statistikprofessoren (Kasseneinnahmen vs. Lognormalverteilung)


Deutschland
Veröffentlichungen von Huber, Wähnert, u.a. insbes. in StBp
Handbuch SRP (interne Version)
Handbuch SRP

VORGEHEN [SRP], KURZ: [1]

1. Betriebswirtschaftliches Profil
2. Auswertungen
3. Zeitreihen
4. **Statistische Tests**

ZIELE UND WÜNSCHE

- lückenlose Prüfung auch [insbesondere?] von Kleinstbetrieben [SRP:6]
- automatische, “gerichtsfeste” Hinzuschätzung [SRP:10,36f]
- Schätzung auch bei formell ordnungsmäßiger Buchführung [Huber2004:98f,103,180], [SRP:35f]

ZIELE UND WÜNSCHE

- strafrechtliche Verwertung des SRP-Ergebnisses [SRP:214ff]
- Schaffung eines “Verurteilungsautomatismus”? – auf pseudowissenschaftlicher Grundlage?

SITUATION (DE)

- BFH v. 25.03.2015 - X R 20/13: **grds. zulässig** (Rz. 54), aber kritisch
- [SRP:114] „*… kann die derzeitige Zurückhaltung in der Finanzrechtsprechung nur darauf zurückzuführen sein, dass Fehler in der Anwendung befürchtet werden.*"

1. Nein, es gibt auch noch andere Gründe (siehe Abschnitt Fehler 1. und 2. Art)
2. Das kritisierte Bauchgefühl der Jurisdiktion ist durchaus richtig

AGENDA

1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

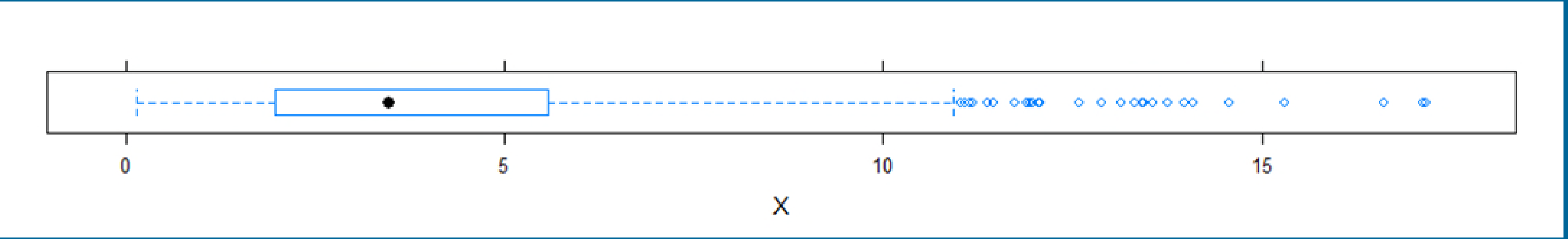
QUANTILE / RICHTSATZSAMMLUNG

Missbrauch deskriptiver Statistiken

DATENBLATT QUANTIL

- Lagemaß
- robust gegen Ausreißer
- Prominentes Beispiel: Median = 50%-Quantil:

0.1 1 1 **4** 8 9 17



VORKOMMEN

Richtsatzsammlung

- „externer Betriebsvergleich"

Summarische Risikoprüfung

- Schätzungsmethode

ZUSTANDEKOMMEN RICHTSATZSAMMLUNG

- beobachte die Zahlen von N Betrieben (sicher ehrlich)
- entferne Ausreißer / unplausible Werte
- Wähle Ober- und Untergrenze
- Resultat: Richtsatzsammlung

- N=?, Auswahlmethode? Geheim!
- was genau passiert hier? Geheim!
- Welche Quantile? Geheim!
- siehe Beyer [2,3,4]

PROBLEME

- Welche Quantile stellen Richtsatzgrenzen dar?
- Legitime Unter-/Überschreitung möglich (irgendwer ist immer Verlierer/Gewinner im Markt) (analog Fehler I. Art, s.u.)
- Geheimes Zahlenwerk, aber Stpfl. verantwortet sich diesem gegenüber – Rechtsstaatlichkeit? BFH kritisch [vgl. 4]

AGENDA

1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

ZEITREIHENVERGLEICHE

VORGEHEN

- Grafische Veranschaulichung von Zeitreihen: (fast) immer eine gute Idee

- „Zoom" bis Einzelzeitpunktbetrachtung, Betrachtung verschiedener Zeiträume: keine gute Idee — „Cherrypicking" – analog dem multiplen Testen (s.u.)

- Bsp.: [SRP:73] „Offensichtliche Regelablösungen [Anm.: Strukturbrüche]… begründen schwerwiegende Zweifel an der inhaltlichen Richtigkeit…" – sehr statischer Begriff von Wettbewerb

MOTTO DES TAGES

- [SRP:54] *„Voraussetzung für die Auswahl geeigneter Methoden zur jeweiligen Prüfungsfrage ist eine* ***ausreichende Methodenkenntnis*** *[…]"*

- [SRP:56] *„Die statistische Kenngröße für gleichgerichtete Abhängigkeiten zweier Datenreihen [gemeint: Zeitreihen] ist der Korrelationskoeffizient."*

METHODENKENNTNIS ZEITREIHENANALYSE

„Die statistische Kenngröße für gleichgerichtete Abhängigkeiten zweier Datenreihen [gemeint: Zeitreihen] ist der Korrelationskoeffizient."

- Korrelationskoeffizient: Zusammenhangsmaß, kein Abhängigkeitsmaß
- Zeitreihen und Korrelationskoeffizient: zumindest gefährlich, [10]

(Scheinkorrelation)

AGENDA

1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

STATISTISCHES TESTEN

Grundlagen und Beispiel

LEBENSSACHVERHALT ↔ STATISTIK

Lebenssachverhalt		Schließende Statistik
Grundannahme („betrügt nicht")	→	Nullhypothese H_0
Forschungshypothese („betrügt")		Alternativhypothese H_1 o. H_A
Buchführungszahlen	→	Stichprobe
		↓
		Teststatistik
		↓ (α-Fehler, Signifikanzniveau)
PRÜFUNGSBERICHT	←	Testentscheidung

LEBENSSACHVERHALT ↔ STATISTIK

Lebenssachverhalt		Schließende Statistik
wir würfeln mit einem fairen Würfel	→	Würfel produziert gleichverteilte Augenzahlen (H_0)
Forschungshypothese („Würfel ist nicht fair”)		Würfel produziert **nicht** gleichverteilte Augenzahlen
600x würfeln	→	{5, 5, 4, 6, 2, 1, 3, 3, 2, ...}
		↓
		Teststatistik (s.u.)
		↓ (α-Fehler, Signifikanzniveau)
WÜRFEL IST FAIR O. UNFAIR	←	H_0 beibehalten o. ablehnen

χ^2-Test (Beispiel: 600x Würfeln)

	1	2	3	4	5	6	Summe
theoret. Häufigkeit (relative)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	600
beobachtete Häufigkeit	87	109	99	88	112	105	600

Beobachtete Häufigkeiten weichen von theoretischen Häufigkeiten ab! Ist die Abweichung so stark, dass wir von einem gezinkten Würfel ausgehen müssen?

χ^2-Test (Beispiel: 600x Würfeln)

	1	2	3	4	5	6	Summe
theoret. Häufigkeit (relative)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	600
beobachtete Häufigkeit	87	109	99	88	112	105	600
Abweichung	-13	+9	-1	-12	+12	+5	0
Abweichung relativ	$\frac{-13}{600} = -0.0217$	0.0150	-0.0017	-0.0200	0.0200	0.0083	0

χ^2-Test (Beispiel: 600x würfeln)

	1	2
…	…	…
…	…	…
Abweichung	-13	…
Abweichung relativ	$\frac{-13}{600} = -0.0217$	…
χ^2	$600 \times \frac{(-0.0217)^2}{1/6} = 1.6952$	…

χ^2-Test (Beispiel: 600x Würfeln)

	1	2	3	4	5	6	Summe
theoret. Häufigkeit (relative)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	100 (1/6)	600
beobachtete Häufigkeit	87	109	99	88	112	105	600
Abweichung	-13	+9	-1	-12	+12	+5	0
Abweichung relativ	$\frac{-13}{600} = -0.0217$	0.0150	-0.0017	-0.0200	0.0200	0.0083	0
χ^2	$600 \times \frac{(-0.0217)^2}{1/6} = 1.69$	0.81	0.01	1.44	1.44	0.25	**5.64**

χ^2-TEST: TESTENTSCHEIDUNG

- Kritischer Wert für Teststatistik:
$$\chi^2(0.95, \mathrm{df} = 5) = 11.0705$$
- Teststatistik: $5.6452 < 11.0705$
- Testentscheidung: H_0 beibehalten, Fairness des Würfels wird weiterhin angenommen

LITERATUREMPFEHLUNG

[Fahrmeir2016]: als gut verständlich bekannt

[Schlittgen2012]: meine Wahl

AGENDA

1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

NEWCOMB-BENFORD-LAW (NBL)

„SIGNIFICANT DIGIT LAW"

Eine magische Eigenschaft von Wirtschaftsdaten?

BENFORDVERTEILUNG (NBL)

exponentieller Prozess (z.B. prozentuales Wachstum):

$$X_t = (1 + r)^t, \qquad t: 1 \to \infty, \qquad \mathrm{r} > 0$$

z.B.: $X_1 = (1 + 0.0001)^1$ $\quad X_2 = (1 + 0.0001)^2$ $\quad X_3 = (1 + 0.0001)^3$ $\quad \ldots$

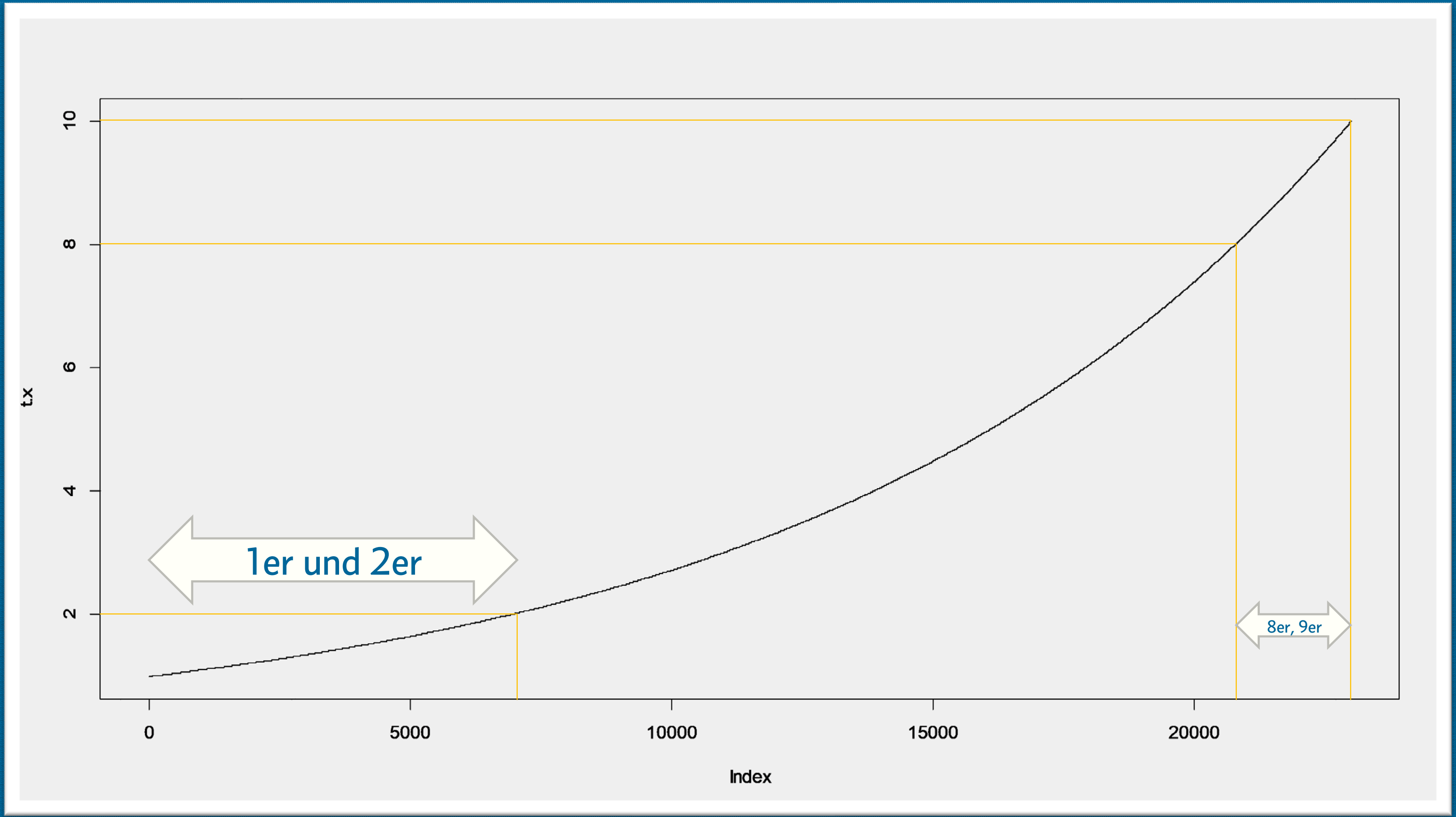
Beispiel (in R [S1]):

```
t.x <- (1 + 0.0001)^( c(1:23027))
t.breaks <- seq(1:10)
round( table( cut(t.x, t.breaks, right=FALSE))/length(t.x), 3)
```

NBL für 1. Ziffer (D1)

[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)
0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

GRAFISCHE VERANSCHAULICHUNG



BENFORDVERTEILUNG (NBL)

Weitere Situationen bekannt, in denen NBL zu gelten scheint oder gar bewiesen ist [vgl. 14]

Kein bewiesenes, **stets** geltendes mathematisches Gesetz!
[vgl. 5,6,7,13,14]
auch nicht in Finanzdaten [vgl. 15:9]

Alternative Erklärungen in [8, 15]
ohne Mathematikkenntnisse lesbare Zusammenfassung in [14]

IN DER SRP

- [SRP:127]: Anwendung auf Tageskasseneinnahmen, Verteilungserwartung der **Zweit**ziffer sei NBL

  ?

- [SRP:135]: Zweitziffer (D2) (123.45) wird mit NBL verglichen.

  Was denn jetzt?

- [SRP:136]: *"Die Zweitziffer nähert sich bereits deutlich der Gleichverteilung an"*

- **[SRP] *meint* verallgemeinertes NBL, [vgl. 5], verkennt: NBL Law beschreibt gemeinsame Verteilung der ersten n Ziffern – isolierte Betrachtung von Randverteilungen von Ziffern ist nicht NBL-begründet!**

GEMEINSAME VERTEILUNG I. U. II.-ZIFFER

	0	1	2	3	4	5	6	7	8	9	2. Ziffer
1	10	11	12	13	14	15	16	17	18	19	
2	20	21	22	23	24	25	26	27	28	29	
3	30	31	32	33	34	35	36	37	38	39	
4	40	41	42	43	44	45	46	47	48	49	
5	50	51	52	53	54	55	56	57	58	59	
6	60	61	62	63	64	65	66	67	68	69	
7	70	71	72	73	74	75	76	77	78	79	
8	80	81	82	83	84	85	86	87	88	89	
9	90	91	92	93	94	95	96	97	98	99	
1. Ziffer											

GEMEINSAME VERTEILUNG I. U. II.-ZIFFER

Unter dem NBL hat die gesamte Tabelle (gemeinsame Verteilung) so auszusehen:

pbenf(2)	0	1	2	3	4	5	6	7	8	9	Randverteilung 1.-Ziffer
1	.0414	.0378	.0348	.0322	.0300	.0280	.0263	.0248	.0235	.0223	**.3010**
2	.0212	.0202	.0193	.0185	.0177	.0170	.0164	.0158	.0152	.0147	**.1761**
3	.0142	.0138	.0134	.0130	.0126	.0122	.0119	.0116	.0113	.0110	**.1249**
4	.0107	.0105	.0102	.0100	.0098	.0095	.0093	.0091	.0090	.0088	**.0969**
5	.0086	.0084	.0083	.0081	.0080	.0078	.0077	.0076	.0074	.0073	**.0792**
6	.0072	.0071	.0069	.0068	.0067	.0066	.0065	.0064	.0063	.0062	**.0669**
7	.0062	.0061	.0060	.0059	.0058	.0058	.0057	.0056	.0055	.0055	**.0580**
8	.0054	.0053	.0053	.0052	.0051	.0051	.0050	.0050	.0049	.0049	**.0512**
9	.0048	.0047	.0047	.0046	.0046	.0045	.0045	.0045	.0044	.0044	**.0458**
Randverteilung 2.-Ziffer	**.1197**	**.1139**	**.1088**	**.1043**	**.1003**	**.0967**	**.0934**	**.0904**	**.0876**	**.0850**	1.00

VON [SRP] IGNORIERT

pbenf(2)	0	1	2	3	4	5	6	7	8	9	Randverteilung 1.-Ziffer
1	.0414	.0378	.0348	.0322	.0300	.0280	.0263	.0248	.0235	.0223	**.3010**
2	.0212	.0202	.0193	.0185	.0177	.0170	.0164	.0158	.0152	.0147	**.1761**
3	.0142	.0138	.0134	.0130	.0126	.0122	.0119	.0116	.0113	.0110	**.1249**
4	.0107	.0105	.0102	.0100	.0098	.0095	.0093	.0091	.0090	.0088	**.0969**
5	.0086	.0084	.0083	.0081	.0080	.0078	.0077	.0076	.0074	.0073	**.0792**
6	.0072	.0071	.0069	.0068	.0067	.0066	.0065	.0064	.0063	.0062	**.0669**
7	.0062	.0061	.0060	.0059	.0058	.0058	.0057	.0056	.0055	.0055	**.0580**
8	.0054	.0053	.0053	.0052	.0051	.0051	.0050	.0050	.0049	.0049	**.0512**
9	.0048	.0047	.0047	.0046	.0046	.0045	.0045	.0045	.0044	.0044	**.0458**
Randverteilung 2.-Ziffer	.1197	.1139	.1088	.1043	.1003	.0967	.0934	.0904	.0876	.0850	1.00

BENFORDVERTEILUNG - SIMULATION

Verkaufsstand für **Bratwurst** und **Bier**:

- Wer Bratwurst kauft, kauft in 60% der Fälle auch Bier
- 70% der Kunden kaufen Bratwurst
- tägliche Kundenzahl X_t ist poissonverteilt
  $X_t \sim Poi(\lambda=250)$
- Bratwurst kostet €3.10, Bier kostet €2.50
- Simulation: 750 Tage (~3 Jahre) (Software: R [S1])

BENFORDVERTEILUNG - SIMULATION

Bratwurst/Bier	ja	nein	Σ
ja	0.42	0.28	0.7
nein	0.30	0	0.3
Σ	0.72	0.28	1.0

```
set.seed(42)
t.tagesumsatz <- rep(0,750); t.D2 <- rep(0,750); t.ones <- rep(0,750)
for (i in 1:750){
   t.tagesumsatz[i] <- sum(sample(c(3.1,2.5,5.6),size=rpois(1,lambda=250),replace=T,prob=c(0.28
                                 ,0.3,0.42)))
   t.D2[i]       <-       as.numeric(strsplit(as.character(t.tagesumsatz[i]),'')[[1]][2])
   t.ones[i]     <-       as.numeric(strsplit(as.character(t.tagesumsatz[i]%%10),'')[[1]][1])
}
```

BENFORDVERTEILUNG - SIMULATION

```
table(t.D2)/750
table(t.ones)/750

prob.D2 <- c(0.120,0.114,0.109,0.104,0.100,0.097,0.093,0.090,0.088,0.085)

# Tests auf Gleichverteilung
chisq.test(table(t.D2))
chisq.test(table(t.ones))

# Tests auf Randverteilung NBL 2. Ziffer
chisq.test(table(t.D2),p=prob.D2)
```

SIMULATION

simulation_benford.R

	0	1	2	3	4	5	6	7	8	9	Gleich-verteilung?	NBL?
D1	0	0.444	0	0	0	0	0	0	0.084	0.472		
D2	0.448	0.044	0.024	0.056	0.048	0.068	0.072	0.096	0.056	0.088	100%	100%
Einer	0.108	0.148	0.112	0.116	0.104	0.084	0.088	0.076	0.076	0.088	~5%	
Gleich vert.	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	%-Zahlen der Ablehnung bei 10k-facher Simulation, p=0.05 simulation_benford.R	
NBL D1		0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046		
NBL D2	0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.088	0.085		

→ **Gegenbeispiel liegt vor**
→ **Behauptung "2. Ziffer muss Benfordverteilt sein" widerlegt**

RESULTATE I

- Theorie: **Gemeinsame** Ziffernverteilung folgen unter bestimmten Bedingungen dem NBL
- [SRP:136]: erkennt, dass Erstziffer bei Kasseneinnahmen i.d.R. offensichtlich nicht NBL (Grund evtl. fehlender exponentieller Prozess, s.o.)
- [SRP]: wendet daher NBL isoliert auf Zweitziffer an – Fundierung?

RESULTATE II

- Simulation: Zumindest Gegenbeispiel einfach möglich

- Bringschuld [SRP]:

    - Demonstrieren, dass es **Beispiele** gibt, in denen die Verteilungsannahmen zutreffen
    - mathematische Beweisführung für Anwendbarkeit der Methoden

    - Unter welchen **Parametern** die Verteilungsannahmen zutreffen (wenn überhaupt)

    - Wie empfindlich die Annahmen auf **Änderungen** der Parameter reagieren

RESULTATE III

- Gleichverteilung der Einerstelle — passt hier (ist auch theoretisch naheliegend [9], aber auch hier Vorsicht)

RECHTSPRECHUNG

- Finanzgericht Hamburg Beschluss v. 31.10.2016 - 2 V 202/16: stützt sich u.a. auf Abweichung Zweitziffer von NBL

AGENDA

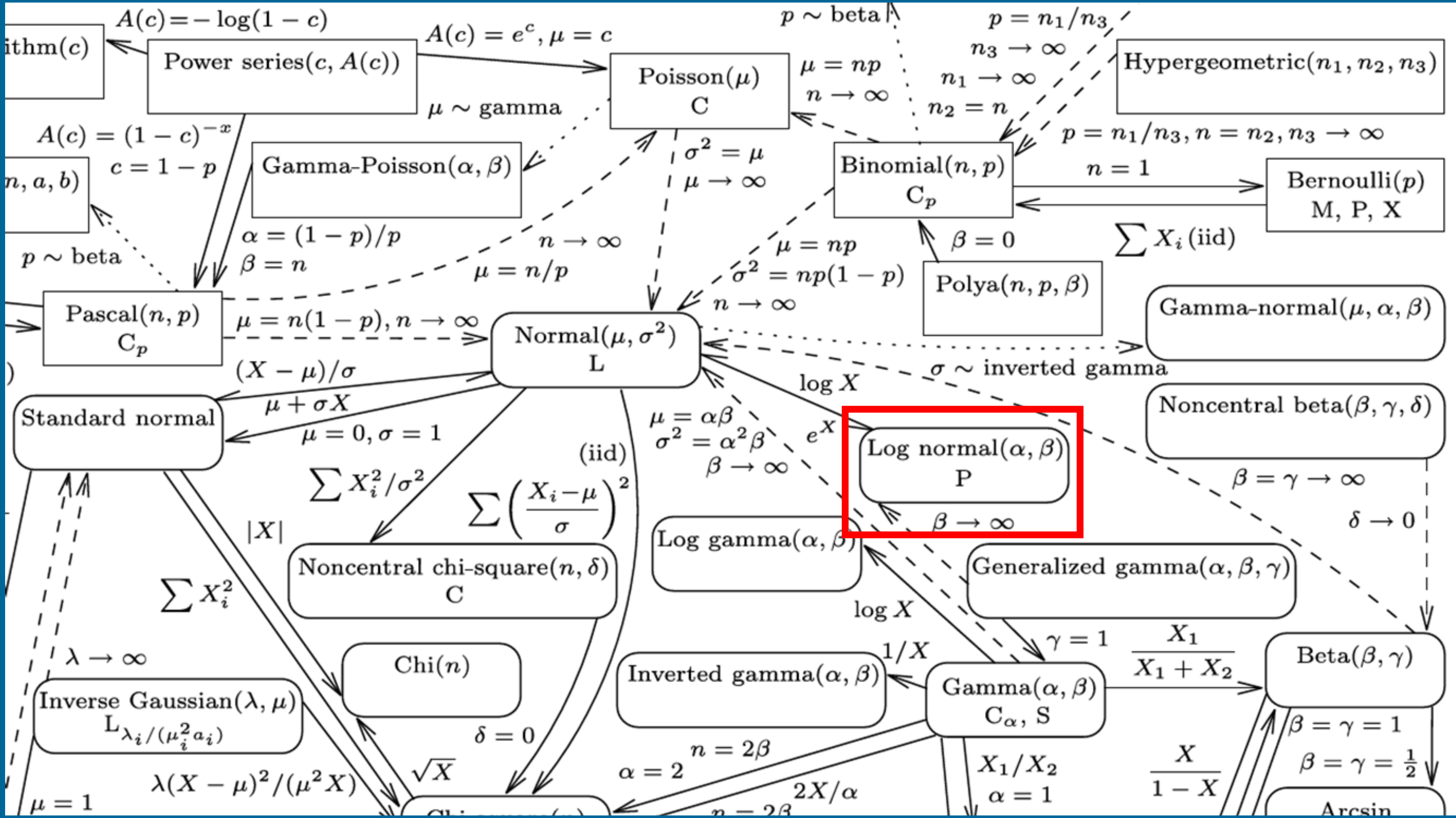
1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

KASSE VS. LOGNORMALVERTEILUNG

Ein Fehlzitat mit schweren Folgen und der Dunning-Kruger-Effekt

ALLES LOGNORMALVERTEILT?

Abbildung aus [11]



VERTEILUNGSTRANSFORMATION (MGF-TECHNIK) (RICHTIG)

$$X \sim N(0,1) \qquad\qquad Y = X^2$$

Welcher Verteilung folgt Y?

$$m_y(t) = \mathrm{E}[e^{tY}] = \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \underbrace{\left(\frac{1/2}{1/2 - t}\right)^{\frac{1}{2}}}$$

MGF einer Chi-Quadrat-verteilten ZV → $Y \sim \chi^2(1)$

VERTEILUNGSTRANSFORMATION (HUBER/WÄHNERT-TECHNIK) (FALSCH)

- Kasseneinahmen → Tageslosung
- Tageslosung: Wirtschaftsdaten
- Wirtschaftsdaten sind lognormalverteilt (belegt durch Fehlzitat, s.u.) → Kasseneinnahmen sind lognormalverteilt ☐

LOGNORMALVERTEILUNG

$$Y = \prod_{t=1}^{T} (1 + r_t)$$

Bsp: Jahresrendite einer Aktie, $T = 250$ Handelstage

$$= (\mathbf{1} + \boldsymbol{r}_1) \times (\mathbf{1} + \boldsymbol{r}_2) \times \cdots \times (\mathbf{1} + \boldsymbol{r}_T)$$

häufig schon fraglich

$$r_t \overset{\text{iid}}{\sim} (\mu, \sigma)$$

$$\rightarrow Y \sim \text{Lognormal (via clt)}$$

SIMULATION

simulation_lognormal.R

- Tageskasseneinnahmen aus vorherigem Beispiel wiederverwendet

- 10.000x Test auf Lognormalverteilung, $\propto = 0.01 = 1\%$
- Bei Lognormalverteilung: 100 Ablehnungen erwartet

- Ergebnis: ~10% Ablehnungen (Shapiro-Wilk-Test) – hätten nur ~1% sein dürfen (KS-Test: ~4%, χ^2-Test: ~1.8%)
- warum nicht mehr Ablehnungen, siehe „Güte"

RICHTIG: WIE IST Y_t VERTEILT?

Y_t: Kasseneinnahmen des Tages t

N_t: Anzahl der Kunden des Tages t

I_t: Anzahl der angebotenen Artikel an Tag t

$q_{i,n,t}$: Menge des Artikels i, die Kunde n an Tag t kauft

$p_{i,t}$: Preis des Artikels i am Tag t

Nur noch Fleißarbeit:

Wie ist Y_t wirklich verteilt? Unter welchen Annahmen?

→valide H_0-Verteilung

$$Y_t = \sum_{n=1}^{N_t} \overbrace{\sum_{i=1}^{I_t} (q_{i,n,t} \times p_{i,t})}^{\text{Einkauf Kunde n}}$$

Summe, kein Produkt! Passt nicht zur Lognormalverteilung

WOHER KOMMT DER UNFUG?

[Huber2004:154]

“Tageseinnahmen .. entstehen üblicherweise zufällig durch Kombination von Preisen mit Verkaufsvorgängen. Wenn viele Zufallsgrößen bei der Entstehung einer rechnerischen Größe multiplikativ zusammenwirken, die Wirkung der Zufallsänderung also jeweils der zuvor bestehenden Größe proportional ist, ergibt sich strukturell eine bestimmte Verteilung. Bei wirtschaftlichen Vorgängen ist das die logarithmische Normalverteilung (in der Folge LogNV), die nach einheitlicher Meinung der Lehre bei wirtschaftlichen Vorgängen vorherrscht.”

WOHER KOMMT DER UNFUG?

[Huber2004:154]

"Tageseinnahmen .. entstehen üblicherweise zufällig durch Kombination von Preisen mit Verkaufsvorgängen. Wenn viele Zufallsgrößen bei der Entstehung einer rechnerischen Größe multiplikativ zusammenwirken, die Wirkung der Zufallsänderung also jeweils der zuvor bestehenden Größe proportional ist, ergibt sich strukturell eine bestimmte Verteilung. Bei wirtschaftlichen Vorgängen ist das die logarithmische Normalverteilung (in der Folge LogNV), die nach einheitlicher Meinung der Lehre bei wirtschaftlichen Vorgängen vorherrscht."

WOHER KOMMT DER UNFUG?

Zur Erinnerung

[Huber2004:154]

"Tageseinnahmen .. entstehen üblicherweise zufällig durch Kombination von Preisen mit Verkaufsvorgängen. Wenn viele Zufallsgrößen bei der Entstehung einer rechnerischen Größe multiplikativ zusammenwirken, die Wirkung der Zufallsänderung also jeweils der zuvor bestehenden Größe proportional ist, ergibt sich strukturell eine bestimmte Verteilung. Bei wirtschaftlichen Vorgängen ist das die logarithmische Normalverteilung (LogNV), die nach einheitlicher Meinung der Lehre bei wirtschaftlichen Vorgängen vorherrscht."

Jahresrendite einer Aktie:

$$Y = (\mathbf{1} + \boldsymbol{r_1}) \times (\mathbf{1} + \boldsymbol{r_2}) \times \cdots \times (\mathbf{1} + \boldsymbol{r_T})$$

Tageseinnahmen einer Kasse:

$$Y_t = \sum_{n=1}^{N_t} \overbrace{\sum_{i=1}^{I_t} (q_{i,n,t} \times p_{i,t})}^{\text{Einkauf Kunde n}}$$

WOHER KOMMT DER UNFUG?

[Huber2004:154]

"Tageseinnahmen .. entstehen üblicherweise zufällig durch Kombination von Preisen mit Verkaufsvorgängen. Wenn viele Zufallsgrößen bei der Entstehung einer rechnerischen Größe multiplikativ zusammenwirken, die Wirkung der Zufallsänderung also jeweils der zuvor bestehenden Größe proportional ist, ergibt sich strukturell eine bestimmte Verteilung. Bei wirtschaftlichen Vorgängen ist das die logarithmische Normalverteilung (in der Folge LogNV), die nach einheitlicher Meinung der Lehre bei wirtschaftlichen Vorgängen vorherrscht."

[Fehlzitat z.B. Sachs2002:173]

WOHER KOMMT DER UNFUG?

[SRP:156] weiß es dann besser als [Sachs2002:173]:

"Als mathematisch-theoretische Erklärung für das asymmetrische Streuungsverhalten der Logarithmischen Normalverteilung wird ein multiplikatives Zusammenwirken von Zufallsgrößen angeführt. M.E. dürfte jedoch …"

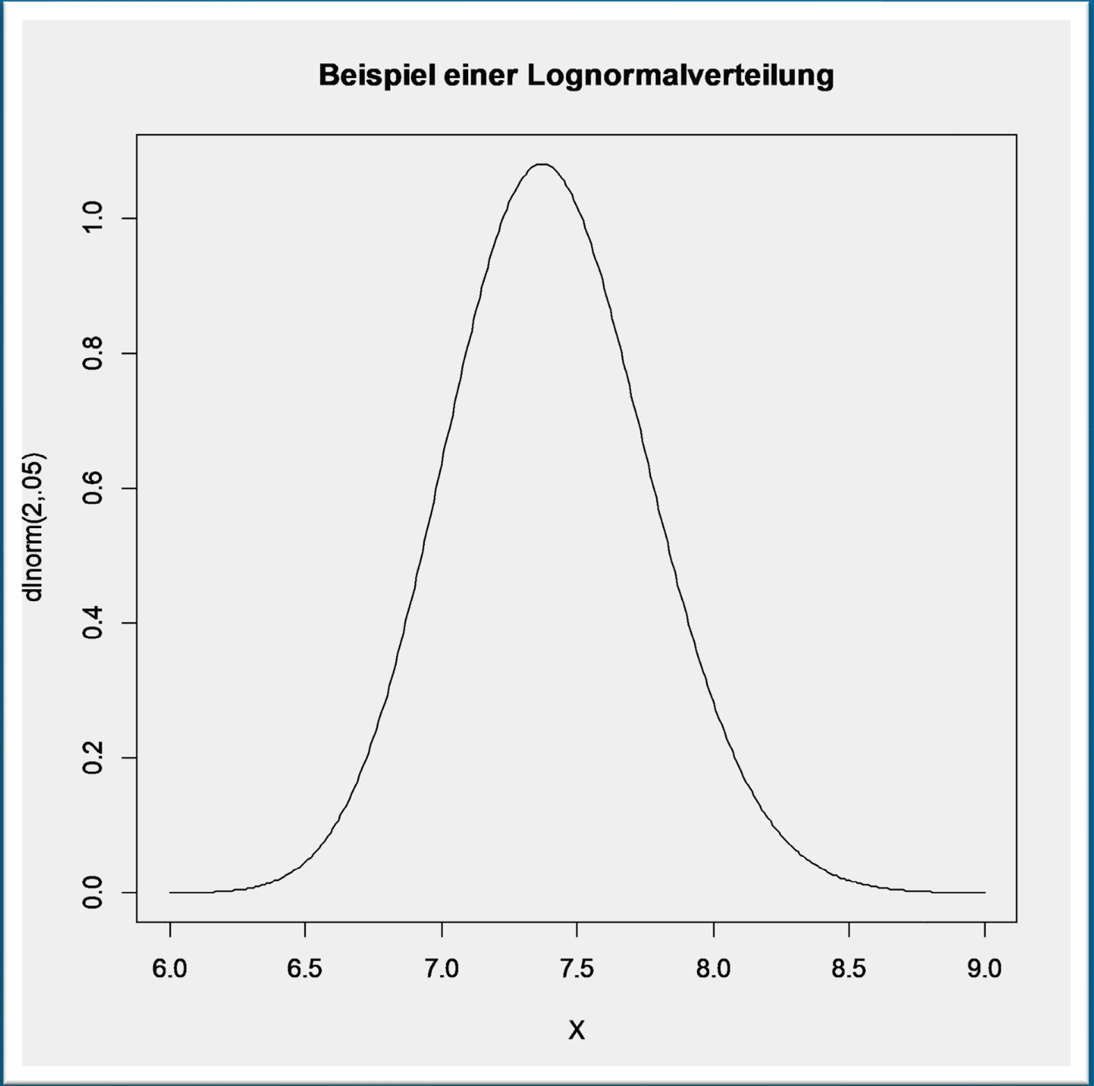
WOHER KOMMT DER UNFUG?

[SRP:156] weiß es dann besser als [Sachs2002:173]:

"Als mathematisch-theoretische Erklärung für das asymmetrische Streuungsverhalten der Logarithmischen Normalverteilung wird ein multiplikatives Zusammenwirken von Zufallsgrößen angeführt. **M.E. dürfte jedoch ...**"

WOHER KOMMT DER UNFUG?

[SRP:156] …Zufallsgrößen angeführt. **M.E.** dürfte jedoch der entscheidende Unterschied zur glockenförmig symmetrischen Zufallsstreuung in einer linksseitigen Begrenzung der möglichen Datenausprägungen liegen…"



RECHTSPRECHUNG

- [UFS] mit qualifizierten gutachterlichen Stellungnahmen [UFS:36-40 u. 128-136].

AGENDA

1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

WARUM FÄLLT DAS NICHT AUF?

Plausibilität durch geringe Güte

GÜTE / POWER

`simulation_chisq.R`

- Wie gut erkennt ein Test eine Abweichung „vom Verteilungsideal” / von der Nullhypothese?

- χ^2-Test: einfach und vielseitig, aber u.U. niedrige Güte

- Beispiel: 10.000 Stichproben je 750 Beobachtungen $X \sim N(20, 1.65)$, Test auf Lognormalverteilung, $\alpha = 0.05$

SIMULATION GÜTE: ERGEBNIS

- **Keine** Stichprobe stammt aus einer Lognormalverteilung

- Die Nullhypothese: “Stichprobe stammt aus Lognormalverteilung” wird **nur in ~16% der Fälle abgelehnt** (Richtig: 100% Ablehung).

→ Obwohl Kasseneinnahmen nie aus Lognormalverteilungen stammen dürften, ist der Test so schlecht, dass er trotzdem häufig nicht ablehnt

FEHLER I. UND II. ART

Fehlerrate x Testhäufigkeit = erwartete falsche Testresultate

FEHLER I. UND II. ART

Testergebnis / Nullhypothese	“H_0 ist wahr”	“H_0 ist unwahr”
ist wahr	✔	**Fehler I. Art** **α-Fehler** **(Ehrlicher Stpfl. wird als Hinterzieher erkannt)**
ist unwahr	Fehler II. Art β-Fehler (Steuerhinterzieher wird nicht erkannt)	✔

FEHLER I. ART

- wird vor Durchführung des Tests festgelegt, Wertebereich zwischen 0 und 1 (0% bis 100%)
- 0%: gängige Tests werden H_0 nie ablehnen
- 100%: gängige Tests werden H_0 immer ablehnen
- üblicher Werte: 0.05 (5%), 0.01 (1%), 0.001 (0.1%)

FEHLER I. ART

- Bedeutung von 5% Fehler erster Art:
- Trifft H_0 ("Stpfl. ist steuerehrlich") **immer** zu,
- so wird vom Test trotzdem in 5% der Fälle "Stpfl. ist steuerehrlich" verworfen

FEHLER I. ART

Trifft H_0 (“Stpfl. ist steuerehrlich”) **immer** zu, so wird vom Test trotzdem in 5% der Fälle “Stpfl. ist steuerehrlich” verworfen:

2018: 188.973 Betriebsprüfungen [16]

falls alle steuerehrlich: Erwartung

9448x Ablehnung der Nullhypothese bei $\alpha = 0.05$

1889X Ablehnung der Nullhypothese bei $\alpha = 0.01$

MULTIPLES TESTEN

Daten quälen, bis sie endlich gestehen

FEHLER I. ART (FORTSETZUNG)

- steuerehrlicher Stpfl. hat bei 5% Fehler I. Art**[!]** 5%-Chance (fälschlich) als steuerunehrlich erkannt zu werden
- analog 20-seitiger Würfel: Wahrscheinlichkeit für eine 20
- SRP sieht aber nicht nur einen statistischen Test vor (Benford 2. Ziffer, Gleichvert. Einerstelle, Lognormal)

FEHLER I. ART (FORTSETZUNG)

- SRP sieht aber nicht nur einen statistischen Test vor (Benford 2. Ziffer, Gleichvert. Einerstelle, Lognormal)

- Bei Unabhängigkeit

$$Pr(\text{"}kein\ Test\ lehnt\ ab\text{"}) = 0.95^3 = 0.86$$

→ 14% Chance, dass einer von drei Tests anschlägt (bei steuerehrlichem Stpfl.)

AGENDA

1. Summarische Risikoprüfung
2. Quantile
3. Zeitreihen
4. Einführung statistisches Testen, Verteilungen
5. Zifferntests
6. Test auf Lognormalverteilung
7. Was sonst noch schiefgehen könnte
8. Schlussfolgerungen

SCHLUSSFOLGERUNGEN

Unschön für die SRP

ANALOGIE STB-EXAMEN

Methodenwahl [SRP]

"Ermitteln Sie stpfl. Gewinn des Einzelkaufmanns."

1. stpfl. Gewinn ermittelt sich nach UStG (Unfug)
2. EUR 2400 aus VHS-Kursen sind steuerfrei
   gem. § 4 Nr. 21 UStG
   (Übungsleiterpauschale zwar irgendwie erkannt, rettet aber auch nichts mehr)

"Schichtung" etc. hilft dann auch nichts mehr

ANWENDUNGSPROBLEME

- Vertrauen auf (richtige) statistische Methoden führt zu
  - Veranlagungsmaßnahmen (Hinzuschätzungen)
  - erheblicher Überschätzung des Aussagewertes der durchgeführten Analysen
  - „automatisiertem" Strafrecht
- Aber: Fehler 1. Art → "By Design" werden zwangsläufig massenhaft Unschuldige getroffen.

METHODENPROBLEME

- Einsatz statistischer Methoden ist empfindlich für
  - Auswahl des richtigen Modells
  - Verletzung von Annahmen
  - Unkenntnis der Bedeutung der Rahmenbedingungen
- Erfordert mathematische Begründung der Modellwahl

ABWEHR UNBERECHTIGTER MAßNAHMEN / ANSPRÜCHE

- penibles Einhalten der Ordnungsvorschriften → zurecht skeptische Rechtsprechung gegenüber SRP
- Anwendung SRP: nicht verhinderbar
- Maßnahmen nur aus SRP-Ergebnis? → vgl. UFS Wien, Methodik muss von qualifizierten Gutachtern geprüft werden (wird nicht standhalten)
- Mischproblem: Ordnungsmäßigkeit der Buchführung verworfen und SRP-Ergebnis? → divide & conquer

POSITIVES

- Möglichkeit der statistischen Methodik wirkt nicht nur in eine Richtung:
  - Beispiel: Unsicherheitszuschlag ist der Höhe nach zu begründen [11] und [BFH VIII R 5/14] – Statistische Methodik u.U. geeignet
  - Beispiel: Entlastungsbeweis

AUSBLICK

- Qualifikationserfordernis für Entwicklung und Anwendung statistischer Methoden?
- Widerstand erwarten: Lebenswerk von Huber/Währnert

QUELLEN

QUELLEN [NUMERISCH]

[1] Beyer. “Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter.” NWB. 2019. S. 2931-2939.

[2] Beyer. “Ist die Richtsatzsammlung eine repräsentative Datenbasis?” NWB. 2018. S. 2921f

[3] Beyer. “Kritische Fragen zur Richtsatzsammlung als Schätzungsbasis.” NWB. 2018. S. 3232-3237.

[4] Beyer. “Aktuelles zur Schätzung mit BMF-Richtsatzsammlung … .” NWB. 2019. S. 3479-3484.

[5] Hill. “A Statistical Derivation of the Significant-Digit Law.” Statistical Science. 1995. Vol. 10, S. 354-363.

[6] Hill. “The First Digit Phenomenon: … .” American Scientist. 1998. Vol. 86, S. 358-363.

[7] Berger und Hill. “Benford’s Law Strikes Back: … .” Mathematical Intelligencer. 2011. Vol. 33, S. 85-91.

[8] Fewster. “A Simple Explanation of Benford’s Law.” American Statistician . 2009. Vol. 63, S. 26-32.

[9] Dlugosz und Müller-Funk. “The value of the last digit: statistical fraud detection with digit analysis.” Advances in Data Analysis and Classification. 2009. Vol. 3, S. 281-290.

[10] Yule. “Why do we Sometimes get Nonsense-Correlations between Time-Series? … .” Journal of the Royal Statistical Society. 1926. Vol. 89, S. 1-63.

[11] Leemis und McQueston. :Univariate Distribution Relationships.” American Statistician. 2008. Vol. 62, S. 45-53.

[12] Nöcker. “Die „neuen“ BFH-Urteile zur Schätzung bei einem Einnahmenüberschussrechner.“ NWB. 2018. S. 2850-2855

[13] Goodman. “The promises and pitfalls of Benford‘s law.“ Significance. 2016. S. 38-41.

[14] Torres, Gamero und Sola. “How do numbers begin? (The first digit law).“ European Journal of Physics. 2007. Vol. 28. S. L17-L25.

[15] Scott und Fasli. “Benford‘s Law: An Empirical Investigation and a Novel Explanation.“ CSM Technical Report. 2001. http://repository.essex.ac.uk/8664/

[16] BMF. „Ergebnisse der steuerlichen Betriebsprüfung 2018.“https://www.bundesfinanzministerium.de/Monatsberichte/2019/10/Inhalte/Kapitel-3-Analysen/3-4-ergebnisse-steuerliche-betriebspruefung-2018.html

QUELLEN [NAMENTLICH]

[Fahrmeir2016]: Fahrmeir, Heumann, Künstler, Pigeot, Tutz. “Statistik – Der Weg zur Datenanalyse.” Springer. 2016. 8. Auflage.

[Huber2004]: Huber. “Die neue Prüfungstechnik in der Betriebsprüfung.” LexisNexis. 2004.0

[Sachs2002]: Sachs. “Angewandte Statistik.” Springer. 2002. 10. Auflage.

[Schlittgen2012]: Schlittgen. “Einführung in die Statistik.” Oldenbourg. 2012. 12. Auflage.

[SRP]: Wähnert. “Handbuch für die Summarische Risikoprüfung.” FinMin SH. 2020. 2. öffentliche Auflage.

[UFS]: UFS Wien. Entscheidung vom 12.3.2009. RV/0328-W/06.

NÜTZLICHE SOFTWARE

[S1] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[S2] G. Grothendieck (2017). sqldf: Manipulate R Data Frames Using SQL. R package version 0.4-11. https://CRAN.R-project.org/package=sqldf

[S3] Dieter William Joenssen (2015). BenfordTests: Statistical Tests for Evaluating Conformity to Benford's Law. R package version 1.2.0. https://CRAN.R-project.org/package=BenfordTests

[S4] Juergen Gross and Uwe Ligges (2015). nortest: Tests for Normality. R package version 1.0-4. https://CRAN.R-project.org/package=nortest1

DANKE FÜR IHRE AUFMERKSAMKEIT