Spielerlust und Spielerfrust in 50 Jahren Lotto – ein Beispiel für visuell gesteuerte Datenanalyse

ANDREAS EICHLER, BIELEFELD

Zusammenfassung: *50 Jahre Lotto laden dazu ein, den gewaltigen Datensatz von über 2500 Ausspielungen des Samstagslottos zu erforschen. In diesem Artikel wird ein Teilbereich untersucht: das Tipp-Verhalten der Spieler in Abhängigkeit von dem sich ändernden normativen Spiel-Modells hinsichtlich der Spielregeln und Tipp-Preise. An die Datenanalyse schließt sich eine didaktische Erörterung des Beispiels unter Beachtung der momentanen Tendenzen der Stochastikdidaktik an.*

1 Einführung

Seit fast 50 Jahren bewegen 6 kleine Kugeln die deutschen Spielerherzen. Nach einem eher bescheidenen Anfang am 9. Oktober 1955, als rund eine Million Spieltipps einen Spieleinsatz von 500 000 D-Mark ergaben und die 10 Hauptgewinner mit 5 Richtigen jeweils etwa 13 000 D-Mark einstrichen, entwickelte sich *das* deutsche Volks-Glücksspiel rasant und zieht mittlerweile mit Rekord-Jackpots von vielen Euro-Millionen ganz Deutschland in seinen Bann.

Theoretisch ist das Lotto-Spiel nahezu vollständig ausgeleuchtet. Zum Beispiel, dass der Einzelne so gut wie keine Chance auf den Hauptgewinn hat, es sei denn, er trifft tatsächlich eine von nahezu 140 Millionen Möglichkeiten – was vergleichbar mit dem richtigen Tipp auf eine einzelne Sekunde in rund viereinhalb Jahren ist – ist hinlänglich bekannt. Ebenso sind die Wahrscheinlichkeiten für alle Trefferanzahlen, für Zwillinge, Drillinge oder sonstige Besonderheiten in Einfach- oder Systemtipps ausführlich diskutiert worden.[1]

Wie aber lässt sich abseits der theoretischen Überlegungen zum Lotto das tatsächliche Spielverhalten der Glücksritter beschreiben? 50 Jahre Lotto (das sind über 2500 Samstagsausspielungen) laden gerade dazu ein, sich in dem gewaltigen Datensatz, der im Netz frei zur Verfügung steht[2], auf die Suche nach Mustern und Besonderheiten zu machen. Dazu wird im Folgenden der Spieleinsatz im Samstagslotto im Sinne eines „Datendetektivs“ (Biehler (1997)) untersucht und durch einen Ausblick auf weitere Möglichkeiten ergänzt. Abschließend werden didaktische Anmerkungen zur Behandlung des Lotto-Beispiels in der Schule formuliert.

2 Erforschung der Spieleinsätze im Lotto

Es geht im Folgendenen darum, die Besonderheiten in dem Datensatz zu *sehen* und die nachfolgenden Untersuchungsschritte anhand der grafischen Darstellungen zu konzipieren. Ausgangspunkt ist dabei der Datensatz zum Lottospiel so, wie er im Netz vorhanden ist. Ein erster, noch unbedarfter Schritt besteht in der grafischen Darstellung der Spieleinsätze seit 1955.

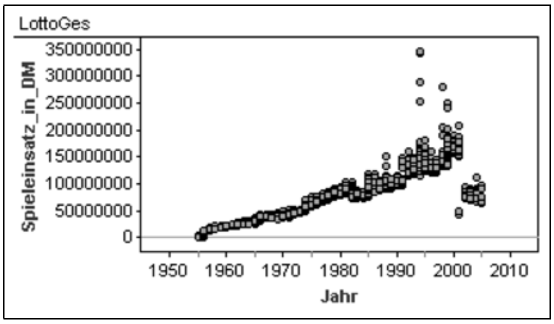

Abb. 1: Entwicklung der Spieleinsätze im Samstagslotto von 1955 bis heute in D-Mark

Auffallend ist hier zunächst zweierlei. Offensichtlich wächst der eingesetzte Betrag kontinuierlich an, nur in den letzten Jahren scheinen die Einsätze einzubrechen. Die letzte Besonderheit ist schnell geklärt, da es sich um die Spieleinsätze nach der Euro-Einführung im Jahr 2002 handelt. Eine Umrechnung aller Beträge in die neue Währung Euro ergibt das folgende Bild.

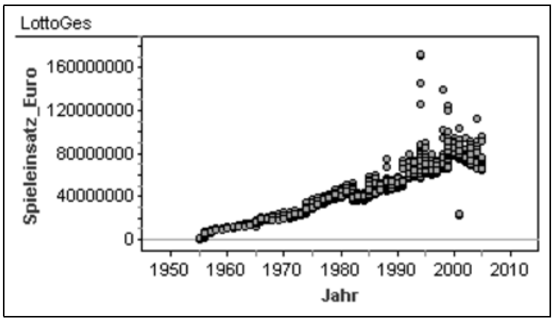

Abb. 2: Entwicklung der Spieleinsätze im Samstagslotto von 1955 bis heute in Euro

Im Sinne einer nicht weiter nachfragenden Statistik wäre man hier fast am Ende der Analyse angelangt.

Die Punktwolke ruft geradezu nach dem Einfügen einer Gerade, die ob nach Augenmaß oder mit der Methode der kleinsten Quadrate schnell eingefügt ist

$$y = 1611800 \cdot x - 3151500000$$

(wenn y den Spieleinsatz und x das Jahr bezeichnet). Die Güte der linearen Abhängigkeit ist hoch ($r \approx 0,96$), die erklärte Varianz (bzw. das Bestimmtheitsmaß) ebenso ($r^2 \approx 0,92$).

Doch es gibt beim genaueren Hinschauen und ebenso beim eingehenderen Einbezug des Sachkontexts noch eine Fülle von Fragen zu stellen und Besonderheiten zu entdecken, die ein weiteres Beschäftigen mit der Punktwolke fruchtbar erscheinen lassen:

1. Gab es in der Lottogeschichte Preiserhöhungen? Wo kann man diese entdecken und wie haben sie sich ausgewirkt?

2. Hat sich die Wiedervereinigung der beiden deutschen Staaten, bei der sich die Zahl der potenziellen Lotto-Spieler schlagartig um ein Viertel erhöhte, in einer deutlichen Mehreinnahme des deutschen Lottoblocks gezeigt?

3. Wie sind die punktuellen Abweichungen von dem Hauptstrang der Lotto-Einnahmen, die so gut durch eine Gerade repräsentiert werden können, zu erklären?

Die drei Fragen, denen im Folgenden nachgegangen werden soll, zielen auf ein tieferes Eintauchen in den Datensatz, die Verbindung des Sachkontexts und der Daten sowie das grafikgesteuerte und detektivische Aufspüren von Besonderheiten in den zunächst leblosen Zahlenkolonnen.

2.1 Spieleinsatz und Tipp-Preise

Eine e-mail an die offizielle Lotto-Kontaktadresse ergibt folgende Informationen zu den verschiedenen Preiserhöhungen im Lotto: Ein Lotto-Tipp kostete

- 50 Pfennige vom 9.10.1955 bis zum 27.6.1981 (ab dem 1.1.1958 musste der Mindesteinsatz allerdings 1 D-Mark betragen);

- 1 D-Mark vom 4.7.1981 bis zum 30.11.1991;

- 1,25 D-Mark vom 7.12.1991 bis zum 15.5.1999 und

- 1,50 D-Mark bzw. 0,75 Euro (wobei hier vereinfachend der Umrechnungsfaktor 2 von Euro zu D-Mark genommen wird) ab dem 22.5.1999.

Da der Preis für einen Lotto-Tipp nicht konstant ist, bietet das Merkmal Spieleinsatz (in Euro) zwar einen Eindruck über die Einnahmen des deutschen Lotto-Blocks, verschleiert aber die Veränderung des Tipp-Verhaltens. In der grafischen Darstellung der Tipp-Anzahlen seit 1955 wird nämlich deutlich, was schon bei der ersten Betrachtung der Preiserhöhungen zu vermuten ist: Insbesondere die erste Preiserhöhung hat zwar die Einnahmen des deutschen Lotto-Blocks nur wenig beeinflusst, aber offenbar eine Schockwelle ausgelöst, die bis heute fortwirkt. So hat sich die Anzahl der abgegebenen Tipps mit der Preiserhöhung 1981 von rund 170 Millionen auf rund 100 Millionen nahezu halbiert und ist seitdem weitgehend konstant geblieben (wobei die Betonung hierbei auf weitgehend liegt! Vgl. Abb. 3). Auch die (gleichzeitige) Erhöhung der Gewinnobergrenze auf 3 Millionen D-Mark hat offenbar den ungeheuren Einbruch der Tipp-Zahlen nicht verhindern können. Vermutlich ist die Einführung des Mittwochslottos – zunächst als Spiel 7 aus 38 – der (erfolgreiche) Versuch, die Lotto-Flucht auszugleichen.

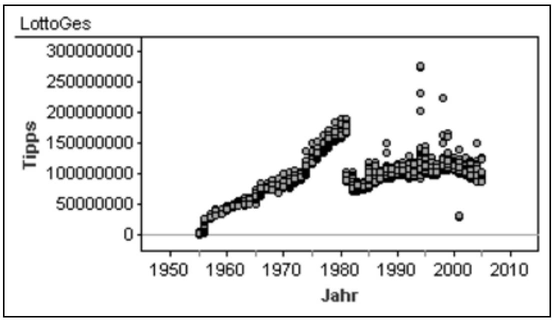

Abb. 3: Entwicklung der Tipp-Anzahl im Samstagslotto von 1955 bis heute

Auch ohne die vertiefte mathematische Beschreibung einzelner Phasen lohnt sich ein Blick auf die weiteren Preiserhöhungen, die offensichtlich bei Weitem nicht die Wirkung der ersten zeigen. Eine Erklärung mag in der weniger drastischen Erhöhung, aber auch in flankierenden Maßnahmen liegen. So ging seit 1981 jede Preiserhöhung mit Erweiterungen des Spiels einher, möglicherweise, um die Preiserhöhung durch ebenfalls erhöhte Attraktivität abzufedern. So gibt es seit dem 7. Dezember 1991 (Preiserhöhung auf 1,25 D-Mark) die Gewinnklassen I (6er mit Superzahl) und VII (3er mit Zusatzzahl).

Die Einführung der (neuen) Gewinnklasse I machte den bereits seit 1985 bestehenden Jackpot ohne Gewinnobergrenze erst interessant:[3] Er wird erst dann fesselnd, wenn er sich füllt. Im Zeitraum von 1985 bis zur Einführung der Superzahl Ende 1991 gab es allerdings in rund 250 Ausspielungen nur sieben, in denen die damalige Gewinnklasse I ('Sechser') nicht belegt war. In keinem Fall folgten zwei Nicht-Belegungen der ersten Gewinnklasse aufeinander, so dass zwar prinzipiell der Aufbau eines Jackpots möglich war, faktisch aber nicht stattfand.

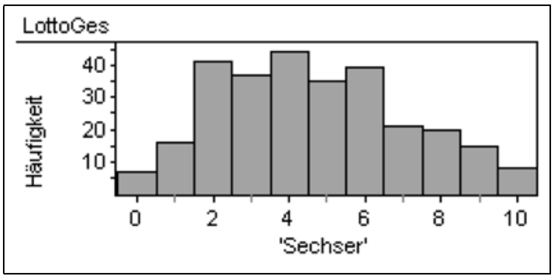

Abb. 4: Belegung der ehemaligen Gewinnklasse I vom 1.6.1985 bis zum 30.11.1991

Erst die Einführung der Superzahl und damit die Verkleinerung der Wahrscheinlichkeit um den Faktor 10, einen Tipp in der (neuen) Gewinnnklasse I ('Sechser mit Superzahl') zu landen, ermöglichte in mehreren Fällen den Aufbau eines Jackpots über mehr als zwei Ausspielungen. Dennoch lässt sich in der Zeitreihe zu den insgesamt abgegebenen Tipps erkennen, dass auch schon die Einführung des Jackpots 1985 den Abwärtstrend stoppen konnte.

Welche weiteren Verbindungen kann man zwischen dem Ringen des deutschen Lotto-Blocks um Spieler und den Daten erkennen? Da gibt es die beiden 'Knicks' oder 'Sprünge' in der offenbar nahezu linear verlaufenden Entwicklung bis 1981 (vgl. Abb. 3). Beide können durch Neuerungen im Lotto-Spiel erklärt werden. Der erste 'Knick' liegt in der Mitte der 60er Jahre und fällt damit mit dem Beginn der Fernsehübertragungen der Lottoziehung am 4. September 1965 zusammen. Es lässt sich vermuten, dass die Werbewirksamkeit der Fernsehübertragung den Anstieg der Tipp-Anzahlen verursacht hat.

Der zweite 'Knick' Mitte der 70er Jahre fällt mit der Anhebung des Höchstgewinns von 500 000 auf 1,5 Millionen D-Mark am 1. Januar 1974 zusammen, die offenbar weitere Spieler mobilisierte bzw. die vermehrte Anzahl von Tipps der bisherigen Spieler bedingte. In der Zeit schließlich nach 1985 scheint sich die Erholung der Tipp-Zahlen bis etwa 2002 zu erstrecken. Danach gehen die Tipp-Zahlen wieder zurück. Ob dies mit der wirtschaftlichen Stimmung in Deutschland, der Einführung des Euro oder auch der Zusammenlegung der Mittwochs- und Samstagsziehung im Lotto zusammenhängt, soll hier nicht weiter thematisiert werden.

Ein deutlich sichtbares Charakteristikum der Zeitreihen sowohl hinsichtlich des Spieleinsatzes als auch der Tipps besteht in den außergewöhnlichen Datenpunkten, die weit ober- bzw. unterhalb der Hauptmasse der Daten liegen. Diese werden später ausführlich betrachtet.

2.2 Spieleinsatz vor und nach der Vereinigung

20 Prozent mehr potenzielle Lotto-Spieler hat die Wiedervereinigung der beiden deutschen Staaten dem Lotto-Block beschert. Das müsste sich doch auch in den Spieleinsätzen bzw. den Anzahlen der abgegebenen Tipps nach 1990 (bzw. nach 1989) niederschlagen. Tut es aber nicht, zumindest weniger, als man es erwarten könnte. Die Boxplots der Tipps in den Jahren um die Wiedervereinigung zeigen einen hinsichtlich des Medians und des ersten und dritten Quartils zwar sichtbaren, jedoch geringen Anstieg von den Jahren 1988 und 1989 zu den Jahren 1990 und 1991.

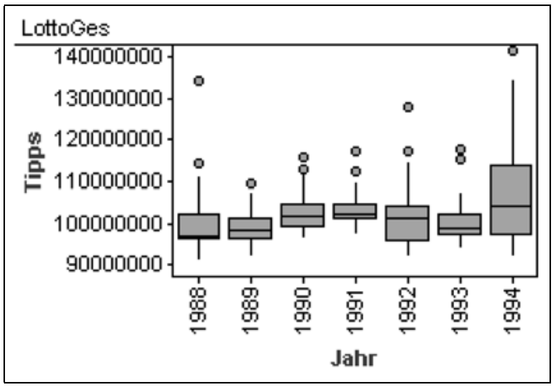

Abb. 5: Entwicklung der Spieleinsätze in der Zeit der Wiedervereinigung

Prozentual beträgt der Anstieg der Tipps von 1989 zu 1990 rund 3,4 Prozent und gibt einen höchstens sehr vagen Hinweis, dass sich die Anzahl der potenziellen Spieler stark erhöht hat. So ist etwa der Anstieg der Tipps vom Jahr 1987 auf das Jahr 1988 deutlich größer (rund 4,7 Prozent). Ein möglicherweise dennoch bestehender Effekt der Wiedervereinigung auf die Anzahl der abgegebenen Tipps überlagert sich spätestens ab Ende 1991 mit einer Spieländerung im Lotto, die sehr viel deutlicher Effekte zeigt – nämlich die Einführung der Superzahl und der damit verbundene temporär enorme Anstieg des Jackpots sowie der Tipp-Anzahlen. Dieser temporäre Anstieg soll abschließend untersucht werden.

2.3 Auswirkungen von Jackpots auf das Tipp-Verhalten

Die Entwicklung der Tipp-Anzahl (vgl. Abb. 3) wie auch der Spieleinsätze (vgl. Abb. 2) zeigt zwei Gruppen oder Cluster von Datenpunkten, die deutlich von den 'normalen' Tipp-Anzahlen bzw. Spieleinsätzen abweichen.

Die Erklärung für die Gruppe der nach unten abweichenden Tipp-Anzahlen scheint banal zu sein (es sind drei nicht-zusammenhängende Lotto-Ausziehungen des Jahres 2001). Mit Blick auf den gesamten Datensatz handelt es sich offenbar um einen Fehler in der Datenerfassung, in dem der Spieleinsatz des Mittwochslottos fälschlicherweise für den Samstag (und umgekehrt) verwendet wurde.

Die Spieleinsätze des Mittwochs-Lottos, das seit dem 2. Dezember 2000 mit den gleichen Regeln wie das Samstags-Lotto gespielt wird und damit schlicht einen weiteren Termin des Samstagslottos darstellt, ist in der unten stehenden Grafik in die bekannte Zeitreihe eingefügt. Eine Beobachtung, die hier nicht weiter verfolgt werden soll, ist die – trotz gleicher Spielregeln und zusammengelegter Jackpots – deutlich niedrigere Tipp-Anzahl des Mittwochs- (eingekreist) gegenüber dem Samstags-Lotto.

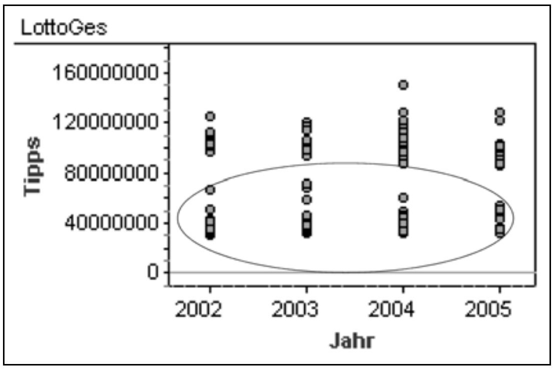

Abb. 6: Tipps im Mittwochs- und Samstagslotto seit Dezember 2000

Eingehender sollen die nach oben abweichenden Tipp-Anzahlen seit 1991 untersucht werden. Eine erste Analyse des Datensatzes bestätigt die plausible Vermutung, dass den sichtbaren 'Ausreißern' eine mehrwöchige Phase des Jackpot-Aufbaus vorangegangen ist.[4] Betrachtet man insgesamt den Zusammenhang zwischen der Länge der 0-Serien, d.h. die Anzahl der Ausspielungen ohne Besetzung der Gewinnklasse I (Sechser mit Superzahl), und der Anzahl der abgegebenen Tipps, so ergibt sich zunächst ein diffuses Bild (vgl. Abb. 7).

Dass sich erst ab einer Länge der 0-Serie von mindestens 3 der Spieleinsatz sichtbar erhöht, ist zwar qualitativ in der Grafik erkennbar, andere Einflüsse, wie etwa die insgesamt bestehende Abnahme an Tipps zu Beginn dieses Jahrhunderts, stören jedoch dieses Bild. Es scheint daher von Nutzen zu sein, einzelne Entwicklungen des Jackpot-Aufbaus zu betrachten.

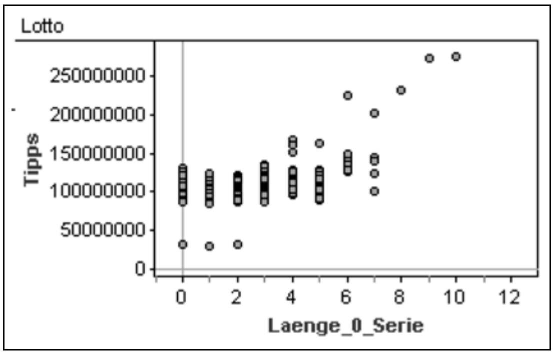

Abb. 7: Tipps in Abhängigkeit von der Länge einer 0-Serie

Dazu bietet sich die längste 0-Serie aus dem Jahr 1994 an. Hier blieb zwischen dem 2. Juli und dem 10. September die Gewinnklasse I zehn Wochen lang unbesetzt.

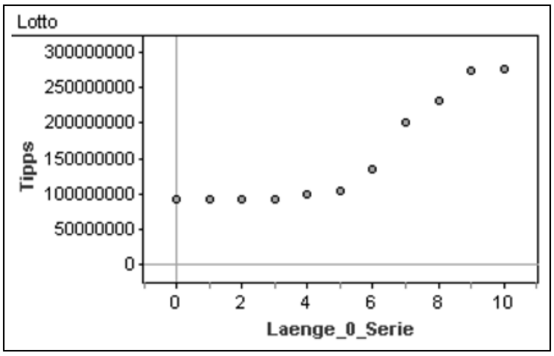

Abb. 8: Tipps in Abhängigkeit von der Länge einer 0-Serie zwischen dem 2. Juli und 10. September 1994

Rein qualitativ erkennt man drei Phasen: die Phase, in der wenig oder nichts passiert, die Phase eines sprunghaften Anstiegs der abgegebenen Tipps, und am Ende dieser kurzen Zeitreihe deutet sich möglicherweise eine Phase der Ermüdung an. Mathematisch sollen im Folgenden zwei Modelle zur Beschreibung dieses Phänomens verwendet werden, eine elementare, stückweise lineare Funktion sowie die logistische Funktion zur Beschreibung von Wachstumsprozessen:

$$f(x) := \begin{cases} g_1 & x \leq 5 \\ \frac{g_2 - g_1}{4} \cdot (x-5) + g_1 \quad \text{für} & 5 < x \leq 9 \\ g_2 & x > 9 \end{cases}$$

$$g(x) := \frac{l_1}{1 + l_2 \cdot e^{-l_3 \cdot x}} + l_4$$

Beiden mathematischen Modellen ist gemein, dass sie ein Supremum besitzen, nämlich g_1 im linearen und $l_1 + l_4$ im logistischen Modell.

Elementar lassen sich die drei genannten Phasen zunächst mit dem linearen Modell bestimmen (vgl. Abb. 9). Mathematisch sind nur die konstanten Anteile der stückweise linearen Funktion zu schätzen bzw. in die Punktwolke einzupassen. Die lineare Funktion mit positiver Steigung ergibt sich durch die Zwei-Punkt-Form der Gerade (unter der vernünftigen Annahme, dass die Funktion an den Enden des durch die zweite Phase gegebenen Intervalls keine Sprungstellen hat).[5]

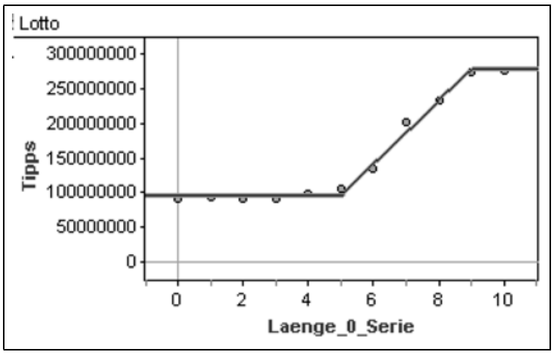

Abb. 9: Tipps in Abhängigkeit von der Länge einer 0-Serie zwischen dem 2. Juli und 10. September 1994

An dieser Stelle wird keine Optimierung der Anpassung für notwendig erachtet, auch wenn sie über die abschnittsweise Berechnung der Regressionsgerade mit Hilfe der Minimierung der Summe der Residuenquadrate (bzw. der Abweichungsquadrate) möglich wäre.

Wichtigeres Ergebnis ist hier, dass per Augenmaß eine die Punktwolke sinnvoll repräsentierende Funktion eingepasst werden kann. Die Gleichung der linearen Funktion hat hier etwa folgende Gestalt:

$$f(x) := \begin{cases} 95 \cdot 10^6 & x \leq 5 \\ 182 \cdot 10^6 \cdot x - 816 \cdot 10^6 \quad \text{für} & 5 < x \leq 9 \\ 277 \cdot 10^6 & x > 9 \end{cases}$$

Dieses elementar zu erstellende Modell hat im Sachkontext den Nachteil, dass man davon ausgeht, dass der Tipp-Markt quasi abrupt gesättigt ist und keine weitere Steigerung der Tippzahlen zulässt.

Ein anderes, aber mathematisch komplexeres Modell bietet die logistische Funktion, die die allmähliche Sättigung der Tippanzahlen beschreibt und hier etwa folgende Gestalt hat:[6]

$$g(x) := \frac{228 \cdot 10^6}{1 + 293 \cdot e^{-0{,}7765 \cdot x}} + 84 \cdot 10^6$$

Obwohl das Modell plausibler zu sein scheint, ist hier das einfachere lineare Modell hinsichtlich der Summe der Residuenquadrate überlegen. So ist diese Summe rund halb so groß wie im zweiten Modell.

Die Sättigung ist im ersten Modell bei einem Wert von etwa $280 \cdot 10^6$ erreicht, im zweiten Modell bei einer Tipp-Anzahl von etwa $300 \cdot 10^6$.

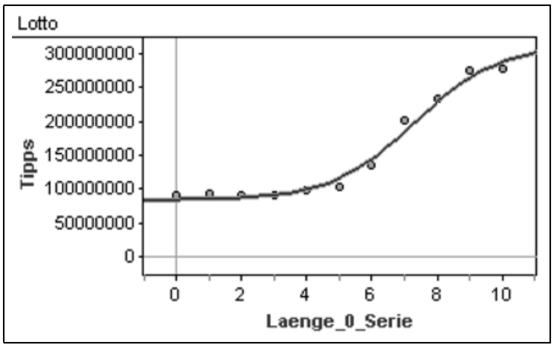

Abb. 10: Tipps in Abhängigkeit von der Länge einer 0-Serie zwischen dem 2. Juli und dem 10. September 1994

Ohne auf eine weitere Residuen- oder Regressionsanalyse, die den Schulrahmen sprengen würde, einzugehen, soll abschließend vielmehr der Frage nachgegangen werden, ob beide oder zumindest eines der beiden Modelle geeignet ist, auch die anderen Anstiege von Tipp-Anzahlen bei längeren 0-Serien zu repräsentieren. Eine solche Modell-Überprüfung ist nicht zuletzt dadurch unbedingt erforderlich, dass es auf einer einzigen 0-Serie und die Modell-Annahme einer Sättigungsphase allein auf einem Datum basiert (obwohl die Annahme einer wie auch immer gearteten Sättigung bzw. Beschränkung der Tipp-Anzahlen selbstverständlich plausibel ist).

Es gibt in der Geschichte des Jackpots nach 1991 fünf weitere Serien, in denen sieben Wochen lang die Gewinnklasse I unbesetzt blieb. Diese sollen im Folgenden betrachtet werden. Die weiteren acht Serien mit einer Länge von sechs Wochen sowie die kürzeren Serien werden hier nicht beachtet, da sich der sprunghafte Anstieg der Tipp-Anzahlen nach den bisherigen Beobachtungen erst bei längeren Serien zeigt.

Es zeigt sich hinsichtlich des gewählten Modells, dass die Jackpotserien in drei Gruppen zerfallen. Zunächst gibt es drei Jackpotserien (Klasse 1), zu denen das bzw. die gewählten linearen bzw. logitischen Modelle passen (auch wenn in diesen Serien keine Sättigung erkennbar ist).

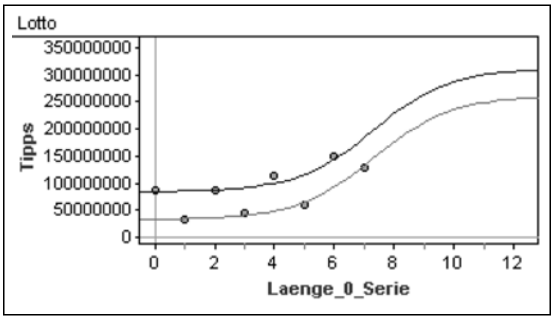

Abb. 11: Klasse 1 der Jackpotserien

Der in der Grafik zu sehende 'Doppelgraph' erklärt sich durch die Gleichschaltung des Samstag- und des Mittwochslottos ab dem 2. Dezember 2000. Ab diesem Zeitpunkt wird der Jackpot vom Samstag auf den Mittwoch und umgekehrt übernommen.

In der Grafik ist sowohl für die Samstagsausspielungen als auch die Mittwochsausspielungen allein das logistische Modell verwendet worden. Das lineare Modell lässt sich analog verwenden. Es zeigt sich, dass einerseits die Tipps am Mittwoch deutlich geringer sind als am Samstag, und dass in diesen beiden Jackpotserien der absolute Zuwachs nahezu identisch ist. Das heißt, die gewählten Modellen sind bis auf eine vertikale Translation (also ein Startniveau bezüglich der Tipps) identisch.

Daneben gibt es zwei weitere Jackpotserien (Klasse 2), die einen deutlich geringeren Anstieg der Tipp-Anzahlen offenbaren.

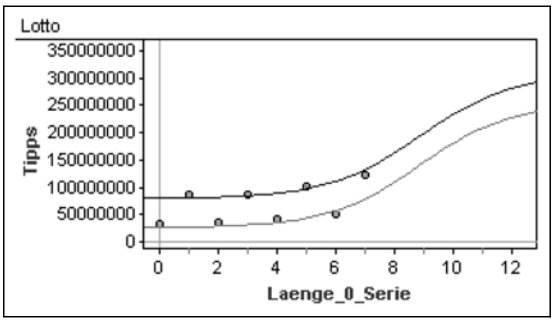

Abb. 12: Klasse 2 der Jackpotserien

Deskriptiv müsste man bei der bestehenden Modellannahme die die Steigung der Funktionen beinflussenden Parameter ändern, etwa in:

$$f(x) := \begin{cases} 88 \cdot 10^6 & & x \leq 5 \\ 26 \cdot 10^6 \cdot x - 42 \cdot 10^6 & \text{für} & 5 < x \leq 12 \\ 270 \cdot 10^6 & & x > 12 \end{cases}$$

$$g(x) := \frac{228 \cdot 10^6}{1 + 293 \cdot e^{-0{,}641 \cdot x}} + 80 \cdot 10^6$$

Die geänderte Steigung könnte als allmählicher Anstieg und als Erreichen der Sättigungsphase erst ab einer Länge der 0-Serie von 12 Wochen interpretiert werden. Für das lineare Modell sind hier die Parameter unter der Annahme der gleichen absoluten Tipp-Zunahme wie in der ersten Klasse von Jackpotserien konstruiert worden.

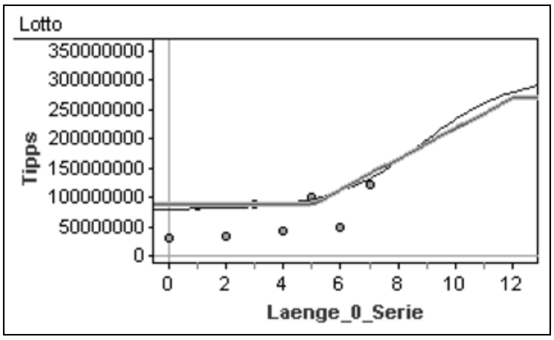

Abb. 13: Klasse 2 der Jackpotserien, lineares und logistisches Modell

Ohne weiteren Zusatz ist so eine Modellvariierung natürlich unbefriedigend. Interessant wird diese, wenn man die Zeitpunkte der unterschiedlichen Jackpotserien mit einbezieht: Die zunächst betrachteten drei Jackpotserien könnte man (wenn man so will) als Ausdruck der aufgestauten Jackpotlust der Spieler bezeichnen. Die eine ist die erste Jackpotserie überhaupt. Vor den beiden anderen hatte dagegen über fünf respektive über zweieinhalb Jahre keine Jackpotserie mehr als sechs Wochen bestanden. Die beiden Jackpotserien mit einem geringeren Anstieg der Tippanzahlen folgten einer vorangehenden Jackpotserie nach weniger als einem dreiviertel Jahr. Der weitaus geringere Anstieg der Tippanzahlen könnte also durch eine gewisse Jackpotmüdigkeit begründet sein.

Schließlich gibt es noch eine Jackpotserie, bei der überhaupt kein Anstieg der Tipp-Anzahlen erkennbar ist (vgl. Abb. 12).

Ist hier möglicherweise die Neutralisierung der verringerten Jackpotlust (der letzte Jackpot liegt knapp neun Monate zurück) und die Unsicherheit mit der Einführung des Euro zu beobachten?

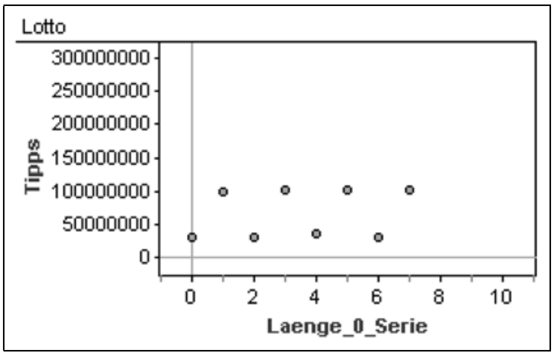

Abb. 14: Jackpotserie ohne Anstieg der Tipp-Anzahlen

Wenn auch Vieles auf Grund der schmalen Datenbasis unsicher bleibt, die gewählten Modelle ermöglichen dennoch eine Beschreibung des Spielerverhaltens in längeren Jackpotserien, die hinsichtlich des Sachkontexts plausibel sind. Wie jede deskriptive Musteranpassung an einen Datensatz (hier die lineare bzw. logistische Funktion) enthält auch diese das Potenzial für eine Prognose für zukünftige Jackpotserien. Diese umfasst folgende Vermutungen:

- In längeren Jackpot- bzw. 0-Serien zeigt sich nach etwa fünf Wochen gleichbleibender Tipp-Anzahlen eine Phase des sprunghaften Anstiegs, die ab zehn bzw. mehr als zehn Wochen in eine Phase der Ermüdung übergeht.

- Die Phase des Anstiegs in einer 0-Serie ist durch den Zeitabstand zur vorangegangenen 0-Serie bestimmt. 0-Serien mit einem Abstand zur vorangegangenen, der größer als zweieinhalb Jahre ist, zeigen einen deutlich stärkeren Anstieg als solche mit einem Abstand von weniger als einem Jahr zur vorangegangenen 0-Serie (die mit einem Abstand zwischen einem dreiviertel Jahr und zweieinhalb Jahren gibt es zurzeit nicht).

- Der absolute Anstieg der Tippanzahlen bei Jackpotserien ist im Mittwochs- und Samstagslotto nahezu identisch (der prozentuale ist damit im Mittwochslotto deutlich größer).

Inwieweit das bzw. die gewählten Modelle tatsächlich tragfähig für die Beschreibung von 0-Serien sind oder ob andere Modelle vorzuziehen sind, kann endgültig erst die Zukunft des Lotto-Spiels zeigen.

3 Anmerkungen zum didaktischen Nutzen

Im Folgenden sollen zunächst zwei Anmerkungen zu einer Verwendung des Lotto-Beispiels in der Schule gemacht werden. Anschließend werden wenige Nach- und einige Vorteile bzw. Möglichkeiten dieses Beispiels diskutiert.

3.1 Das Lotto-Beispiel in der Schule

Bei der Verwendung des Lotto-Datensatzes in der Schule sind natürlich Herangehensweisen möglich und denkbar, die von dem hier vorgestellten Modellierungsgang abweichen. Etwa könnte man zunächst mit der Punktwolke zu den Tipp-Anzahlen oder mit einem weiter veränderten Datensatz beginnen. Hier sollte dagegen allein der nicht vorkonstruierte Gang beschrieben werden, der mit dem Datensatz so beginnt, wie er im Netz verfügbar ist (dieser enthält z.B. keine Angaben über die Tipp-Anzahlen).

Es wurden zwei Modelle vorgestellt, ein lineares und ein logistisches. Das lineare ist sicher bereits in der Sekundarstufe I verfügbar. Mit diesem Modell lassen sich die wesentlichen Schritte der hier vorgestellten Datenanalyse durchführen. Das logistische Modell ist dagegen allein für die Sekundarstufe II geeignet, etwa bei der übergreifenden Behandlung von Wachstumsprozessen.

3.2 Grenzen des Lotto-Beispiels

Im Sinne der in der Datenanalyse bzw. des anwendungsorientierten Mathematikunterrichts nicht aufzuhebenden Trennung der Mathematik mit dem Sachkontext kann die Erforschung des Lotto-Datensatzes erst dann ansetzen, wenn es für Schülerinnen und Schüler sinnvoll ist, sich mit diesem Glücksspiel zu beschäftigen. Damit ist eine Behandlung dieses Beispiels erst am Ende der Sekundarstufe I angemessen. Denn das verständige Entdecken setzt einen gewissen Umgang mit Glücksspielen, mit Geld und mit den alltäglichen Medien, die über Lotto berichten, voraus. Auch im Sinne einer Verbindung von Explorativer Datenanalyse und der Wahrscheinlichkeitsrechnung sowie der Kombinatorik als ihr Hilfsmittel ist diese Einschränkung sinnvoll.[7]

Ein Erforschen der Lotto-Daten ohne den Rechner ist sinnlos. Obwohl dies im Sinne der allgemeinen didaktischen Diskussion keine Grenze, sondern eher eine Möglichkeit ist, scheint es in der offenbar (noch?) überwiegend rechnerfreien Schul-Stochastik zumindest momentan eine Grenze darzustellen.[8]

3.3 Möglichkeiten des Lotto-Beispiels

Die Möglichkeiten, die die Erforschung des Lotto-Datensatzes umfasst, sollen anhand dreier Aspekte skizziert werden. Diese Aspekte betreffen

- die Formulierung von Standards des Mathematik- und Stochastikunterrichts, etwa in der *Leitidee Daten und Zufall* durch die Kultusministerkonferenz (KMK (2003)),
- die *Prinzipien der Explorativen Datenanalyse*, wie sie von Tukey et al. (1982) formuliert und seitdem in vielen didaktischen Arbeiten zitiert und ausgestaltet wurden, sowie
- die Phasen des *statistischen Denkens*, das als daten- und realitätsorientiertes Pendant des stochastischen Denkens seit einiger Zeit in der Stochastikdidaktik diskutiert wird (vgl. Pfannkuch/Wild (1999)).

Diese drei Aspekte sollen im Folgenden kurz präzisiert und auf das Lotto-Beispiel bezogen werden.

Standards Die Kultusministerkonferenz veröffentlichte 2003 allgemeinverbindliche Standards für den Stochastikunterricht in der Sekundarstufe I, die in neu gestaltete Lehrpläne der einzelnen Bundesländer eingegangen sind bzw. eingehen werden. Zu diesen Standards, die eine wesentliche Verringerung der Wahrscheinlichkeitsrechnung zugunsten der Datenanalyse umfassen, gehören:

- die Planung, Durchführung und Auswertung von statistischen Erhebungen,
- die systematische Datensammlung und -aufbereitung insbesondere mit grafischen Verfahren unter Verwendung des Rechners,
- die Interpretation der Daten mit Hilfe von Kenngrößen,
- die Verbindung von Realität und Mathematik durch die Interpretation von Argumenten, die auf der Datenanalyse beruhen und
- die Beschreibung von Zufallsphänomenen in alltäglichen Situationen und deren mathematische Modellierung mit Wahrscheinlichkeiten.

Während die Planung der Datenerhebung in diesem Beispiel keine Rolle spielt, werden die übrigen Standards offenbar bedient, wobei der Analyseprozess stets durch die in den Standards zentral genannten grafischen Verfahren gesteuert wird. Auch der letzte Punkt wird angesprochen, da das Ersetzen einer Punktwolke bzw. einer Zeitreihe als spezieller Punktwolke durch eine Funktion stets eine Reduzierung der Daten um den Zufall umfasst.

Wichtig ist ebenso die Verbindung zu anderen Leit-Ideen des Mathematikunterrichts, insbesondere zur Leitidee des funktionalen Zusammenhangs. So ist ein wesentlicher Aspekt der Analyse der Lotto-Daten die Verwendung einer oder mehrerer Funktionen „*zur Beschreibung quantitativer Zusammenhänge*“ (KMK (2003, S. 15). Die Funktion ist hier ein ideales Muster zur Beschreibung der Realität.

An dieser Stelle soll dennoch betont werden, dass der Lotto-Datensatz kein didaktisch konstruiertes, besonders wertvolles Beispiel darstellt, durch das der Gedanke der Funktion als modellhafte Beschreibung von Daten herausragend berührt wird. Vielmehr handelt es sich bei dem Lotto-Datensatz um ein reichhaltiges Entdeckungsfeld, das eben *auch* die Mustererkennung in Daten bzw. eine Funktionsanpassung in einer Zeitreihe ermöglicht. Der Gedanke einer Prognose ist hier als Zusatz zu verstehen, ermöglicht aber den Einstieg in eine grundlegendere Behandlung von Prognosen auf Grund statistischer Daten.

Im Sinne eines Spiralcurriculums tauchen die in den Standards formulierten Ideen auch in Überlegungen zum Mathematikunterricht in der Sekundarstufe II auf (vgl. etwa Borneleit et al. (2000)) und sind damit auf der Basis komplexerer mathematischer Betrachtungsweisen übertragbar.

Prinzipien der Explorativen Datenanalyse Prinzipiell kann die Explorative Datenanalyse im Gegensatz zur klassischen Beschreibenden Statistik als Umkehrung eines Bearbeitungsablaufs *Modell–Datenerhebung und -auswertung* verstanden werden (vgl. Borovcnik/Ossimitz (1984)). So geht es um das detektivische Durchforsten eines Datensatzes mit einer möglichst flexiblen Verwendung der statistischen Methoden, um Besonderheiten und Muster zu entdecken, die dann Grundlage für das Aufstellen eines Modells sein können. Grundlage dieses induktiven Vorgehens sind (vgl. Tukey et al. (1982))

- die zentrale Verwendung grafischer Methoden (*eine Grafik sagt mehr als 1000 Worte*),
- die Verwendung robuster Methoden, etwa der Einsatz von Median (bzw. Quantilen) sowie zugehöriger Streuparameter und
- die Clusterung von Daten, d.h. die Einschränkung der Datenanalyse aus sachlichen Gründen auf einen Teil der Daten.

Die Steuerung des Analyseprozesses der Lotto-Daten geschieht bei dem hier beschriebenen Vorgehen offensichtlich auf Grund der grafischen Darstellungen. Man untersucht, was man *sieht* und versucht, das Gesehene mathematisch bzw. statistisch zu fassen und mit dem Sachkontext zu verbinden.

Die Verwendung robuster Methoden als Prinzip der Explorativen Datenanalyse kann und sollte nicht mehr rigider Bestandteil der schulischen Datenanalyse sein. So geht es – insbesondere mit Hilfe des Rechners – vielmehr um ein verständiges Einsetzen aller statistischer Methoden. Dabei meint verständig, dass man unterschiedliche Methoden wie etwa das arithmetische Mittel oder den Median im Bewusstsein der daraus möglicherweise unterschiedlichen Dateninterpretationen einsetzen kann. In der Analyse der Lotto-Daten sind robuste und nicht-robuste Methoden vermischt. Diese Mischung wird etwa bei der Anpassung eines Funktionsgraphen an die Punktwolken sichtbar, z.B. bei der Frage, ob man versucht, die Summe der Abweichungsquadrate zu minimieren, oder ob man sich von anderen (und damit zumeist robusteren) mathematischen Methoden leiten lässt.

Die Clusterung ist in der Analyse der Lotto-Daten ein durchgängiges Prinzip. So werden die Cluster von Spieleinsätzen, die durch unterschiedliche Währungen entstehen, angeglichen. Gleiches geschieht mit den Clustern der Spieleinsätze bzw. der Tipps, die sich hinsichtlich der unterschiedlichen Preisstufen in der Lottogeschichte ergeben. Schließlich werden ganz bestimmte Cluster herausgegriffen und gesondert untersucht, um ein Modell – hier die Tipp-Anzahl beim Ansteigen des Jackpots – zu erstellen.

Statistisches Denken Das statistsche Denken, das bei Pfannkuch/Wild (1999) in ein umfassenderes Modell des Problemlösens eingebunden ist, enthält folgende Phasen oder Katgorien:

- Erkennen der Notwendigkeit realer Daten (recognition of the need for data),
- Flexible Repräsentation der relevanten Daten und deren Diskussion (transnumeration),
- Einsicht in die Variabilität von und Muster in den Daten (consideration of variation),
- Einbinden der Daten in statistische Modelle (reasoning with statistical models) und
- Verbinden von Kontext und Statistik (integrating the statistical and contextual).

Das Erkennen der Notwendigkeit statistischer Daten zur Beschreibung und Beurteilung eines Sachkontexts ist hier vorweggenommen und prinzipiell eher eine Anforderung an allgemeine Problemlöseprozesse, bei der die Bezugnahme auf statistische Daten nicht offensichtlich ist.

Die flexible Repräsentation der Daten ist ein wesentliches Moment der Analyse der Lotto-Daten. Dabei ist die Analyse geleitet durch die im vorangegangenen Abschnitt diskutierten Prinzipien der Explorativen Datenanalyse. Sie umfasst die unterschiedlichen grafischen Darstellungen, die Reduktion auf statistische Maßzahlen und schließlich die Darstellungen ein und desselben Sachverhalts in verschiedener Form auf Grund von Datentransformationen.

Die Variabilität der Lotto-Daten zum Spieleinsatz bzw. zur Zahl der abgegebenen Tipps ist in jeder Phase des Analyseprozesses sichtbar. Im Zusammenhang mit dem Einbinden der Daten in Modelle geht es um den zentralen Aspekt des statistischen Denkens: Die Aufteilung der statistischen Daten in ein Muster (bzw. ein mathematisches Modell) und Residuen (bzw. Zufall). Die mathematischen Muster bestehen etwa in der Reduktion der Datenvielfalt auf eine Funktion, die prinzipiell Trends der Spieleinsätze oder der Tipp-Anzahl beschreiben und zum Teil erklären helfen und damit auch als Gundlage für (vorsichtige) Prognosen dienen kann, die in der Zukunft zu falsifizieren oder auch zu verifizieren sind. Diese mathematischen Modelle heben allerdings die Variabilität der Daten nicht auf, sondern stellen ein Muster dar, um die die Daten streuen.

Der Einbezug des Sachkontexts ist bei der Analyse des Lotto-Datensatzes unabdingbar. Deutlich wird hier, wie die Veränderung eines *normativen Modells* die Daten hinsichtlich der betrachteten Merkmale Spieleinsatz und Anzahl der abgegebenen Tipps beeinflusst. D.h. die sich ändernden Spielregeln und Tipp-Preise steuern ebenso wie die als zufällig zu betrachtenden 0-Serien das Verhalten der Spieler. Der Sachkontext leitet schließlich die Konstruktion der Modelle, in dem etwa trotz der bisher denkbar schlechten Datenlage allein aus Plausibilitätsgründen von einer Sättigungsphase der Tippanzahlen ausgegangen wird.

4 Schlussbemerkung

Die Analyse der Lotto-Daten birgt eine Möglichkeit, im Stochastikunterricht reale Daten im Sinne der für die Schule geforderten Standards der Leitidee Daten und Zufall zu behandeln. Wesentliche auf statistischen Grundbegriffen aufbauende Methoden können zur Analyse verwendet werden.

Dabei ist es möglich, die Methoden je nach Klassenstufe weitgehend elementar zu halten. Es ist aber ebenso möglich, mit sehr viel komplexeren mathematischen bzw. statistischen Methoden zu arbeiten, wie dies nur im Vergleich der Modelle, die auf einer linearen bzw. logistischen Funktion basieren, aufgezeigt wurde. Insbesondere die Optimierung der Funktionsanpassung birgt Möglichkeiten oder aber auch Probleme, die über die Schule hinausweisen. Darüber hinaus umfasst der Lotto-Datensatz die Möglichkeit, quasi nebenbei Elemente der Zeitreihenanalyse zu behandeln und schließlich einen Einstieg in das Themenfeld Prognose zu leisten.

Von den vielen Untersuchungsmöglichkeiten, die der Lotto-Datensatz bietet, ist in dieser Arbeit nur ein kleiner Aspekt ausgewählt worden, der Appetit auf mehr machen soll. Die Untersuchung von Quoten, den sogenannten *Narrenzahlen*, kleinen und großen Katastrophen und nicht zuletzt die Verbindung des wahrscheinlichkeitstheoretischen Wissens zum Lotto mit den empirischen Daten sind zukünftige Betätigungmöglichkeiten für Datendetektive in der 50jährigen Geschichte eines Glücksspiels, das gleichermaßen die Öffentlichkeit wie auch die kleinen und großen Mathematiker in seinen Bann ziehen kann.

Anmerkungen

[1] Stellvertetend für alle Lotto-Artikel kann hier die Fülle der Arbeiten von Strick (z.B. (2003)) gelten, der sich seit Jahrzehnten mit dem Lotto beschäftigt.

[2] Diese sind als txt-Datei auf der Webpage von Lotto-Rheinland-Pfalz zu finden (vgl. http://www.lotto-rlp.de). In diesem Datensatz sind leider manche Inkonsistenzen enthalten. Einen bereinigten Datensatz im Excel oder fathom-Format erhält man auf der persönlichen Seite des Autors innerhalb folgender Homepage: http://www.tu-braunschweig.de/idm.

[3] Beim Jackpot wird die prozentual festgelegte Gewinnausschüttung der Gewinnklasse I (bis zum 7.12.1991 der 'Sechser', danach der 'Sechser' mit Superzahl) zu der Gewinnausschüttung der nächsten Ziehung hinzugefügt, wenn es keinen (richtigen) Tipp in der Gewinnklasse I gab.

[4] Die Software *fathom* unterstützt diesen ersten Analyseprozess sehr gut. So können einzelne Datenpunkte markiert und zu diesen alle erhobenen Merkmale in einer Infobox angezeigt werden. Letztere erlaubt wiederum das leicht handhabbare Betrachten der Vorgänger- oder Nachfolge-Ziehungen.

[5] Die Verwendung von *fathom* ermöglicht eine sehr einfach zu handhabende Eingabe einer Funktion in ein bestehendes Diagramm. Die zu schätzenden Parameter lassen sich durch Schieberegler wiederum sehr leicht anpassen. Zudem gibt es die Möglichkeit, sich die Summe der Residuenquadrate zu einer eingepassten Funktion anzeigen zu lassen.

[6] Die Parameter der logistischen Funktion sind hier mit fathom angepasst worden. Vgl. Anmerkung 5.

[7] Beispielsweise ist in Niedersachsen die elementare Behandlung des Lotto-Spiels auf die Klasse 10 beschränkt (vgl. RRL (2003) oder Griesel/Postel (1996)). Unglücklich ist hier die Trennung von Datenanalyse und Wahrscheinlichkeitsrechnung, wobei die Datenanalyse quasi als kleine Schwester der Wahrscheinlichkeitsrechnung in den unteren Klassen der Sekundarstufe I behandelt wird.

[8] Auch wenn es sicherlich viele Ausnahmen gibt, scheint die genannte These auf Grund von Erfahrungen in der Lehrerfortbildung im Rahmen von SINUS sowie der Erforschung von Lehrervorstellungen zum Stochastikunterricht (vgl. Eichler (2005)) angemessen zu sein.

Literatur

Borneleit, P., Danckwerts, R., Henn, H.-W., Weigand, H.-G. (2000): Expertise zum Mathematikunterricht in der gymnasialen Oberstufe. www.math.uni-siegen.de/didaktik/downl/expertise.pdf.

Biehler, R. (1997): Auf Entdeckungsreise in Daten. In: mathematiklehren 97, 1997, 4-5.

Eichler, A. (2005): Individuelle Stochastikcurricula von Lehrerinnen und Lehrern. Hildesheim: Franzbecker.

Griesel, H., Postel, H. (1996, Hrsg): Elemente der Mathematik 10, Niedersachsen. Hannover: Schroedel.

Pfannkuch, M., Wild, C. (1999): Statistical Thinking in Empirical Enquiry. In: International Statistical Review 67(3) 1999, 223-248.

Niedersächsisches Kultusministerium (2003, Hrsg.): Rahmenrichtlinien für das Gymnasium, Klasse 7 – 10 Mathematik. Hannover: Schroedel.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2003): Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Bonn.

Tukey, P.A., Gnanadesikan, R., Kettering, J.R., Siegel, A.F. (1982): Themen aus der Datenanalyse: Begriffe, Methoden, Beispiele und Pädagogik. In: Der Mathematikunterricht 28(1) 1982, 28-56.

Anschrift des Verfassers
Andreas Eichler
Institut für Didaktik der Mathematik
Universität Bielefeld
Postfach 10 01 31
33501 Bielefeld
andreas.eichler@uni-bielefeld.de