Gruppenanalysen
von Günter Schmidt

Dies ist der elektronische Nachdruck des zweiten Kapitels einer bei Texas Instruments verlegten Broschüre mit dem Titel

Mathematik erleben
Experimentieren
Entdecken
Modellieren
Veranschaulichen

von **Günter Schmidt, Stromberg**

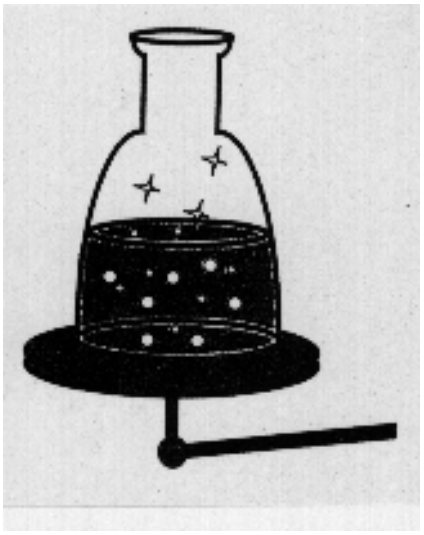
Die anderen Kapitel zu den Themen

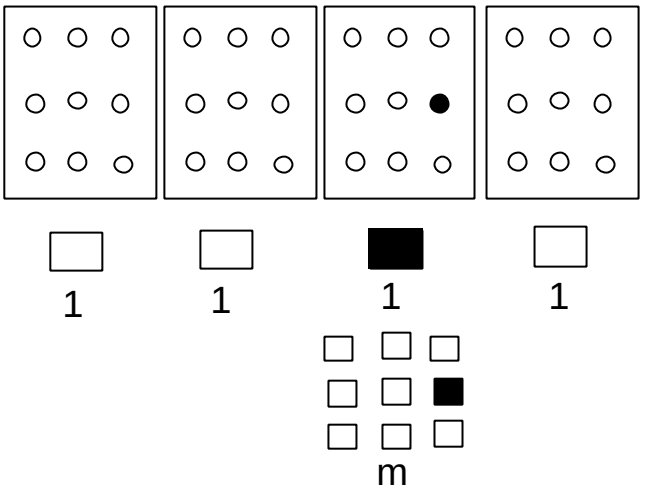
- **Abstandsberechnungen** (Analytische Geometrie)
- **Springbrunnen** (Analysis)
- **Lösen von Gleichungen** (Analysis)
- **Kurvenanpassung** (Statistik/Analysis)

können auf www.ti.com/calc/deutschland/materialien.htm in der entsprechenden Rubrik gefunden werden.

Gruppenanalysen

Themenbereich	
Stochastik, Verbindung zur Analysis (Extremwerte)	
Inhalte	Ziele
• Simulation von Zufallsexperimenten • Erwartungswert • Extremwertbestimmung bei reellen Funktionen	• Problemlösen mit Simulationen und mathematischen Modellen • Anwenden von Verfahren der Stochastik und der Analysis





In der Medizin (Serumdiagnostik) müssen oft große Populationen auf das Vorhandensein eines bestimmten Erregers untersucht werden. Der Aufwand für die Entnahme und Untersuchung einer Blutprobe ist hoch und mit entsprechenden Kosten verbunden. Man ist deshalb an einem Verfahren interessiert, das diesen Aufwand minimiert.

Für den Fall, daß die Wahrscheinlichkeit für eine positive Probe gering ist und daß man außerdem den Erreger mit sehr großer Sicherheit feststellen kann, wurde ein Gruppenverfahren entwickelt: Man faßt mehrere Blutproben zu einer Gruppe zusammen und untersucht zunächst diese Mischprobe. Nur im Fall einer positiven Reaktion einer solchen Gruppenuntersuchung werden dann weitere Einzeluntersuchungen erforderlich.

- Kann man mit einem solchen Gruppenverfahren die Anzahl der Untersuchungen (im Mittel) gegenüber dem Einzeluntersuchungsverfahren deutlich verringern?

- Wie ist die optimale Gruppengröße m in Abhängigkeit von der Wahrscheinlichkeit p (für einen Befall) zu wählen, damit die mittlere Anzahl der Untersuchungen möglichst klein wird?

Vorüberlegungen

Wir gehen das Problem in mehreren Schritten an.

1. Plausibilitätsbetrachtungen

Wenn die Wahrscheinlichkeit p für den Befall eines Individuums nicht klein ist, so wird man in der Gruppenuntersuchung in der Regel ein positives Ergebnis erwarten. Damit werden dann anschließend die Einzeluntersuchungen nötig. Eine Verringerung der mittleren Zahl der Untersuchungen wird man also nur bei kleinen Wahrscheinlichkeiten p erwarten. Auch die Gruppengröße wird bei kleinen Wahrscheinlichkeiten wohl größer gewählt werden können. Die Höhe der Ersparnis bei dem Gruppenverfahren kann man nur schwer schätzen, ebensowenig die zu einem bestimmten p gehörige optimale Gruppengröße.

2. Simulation

Eine geeignete Simulation des Prozesses kann die Plausibiltätsbetrachtungen unterstützen. Dabei wird man auch einen ersten quantitativen Eindruck erhalten. Hierzu entwickeln wir ein Programm, mit dem wir dann entsprechende Experimente ausführen (simulieren) können. Der Kern dieses Programms wird der Zufallszahlengenerator des TI 92 sein. Der Entwurf des Programms verlangt eine passende Modellierung des Vorgangs, diese kann übersichtlich in einem Struktogramm dargestellt werden.

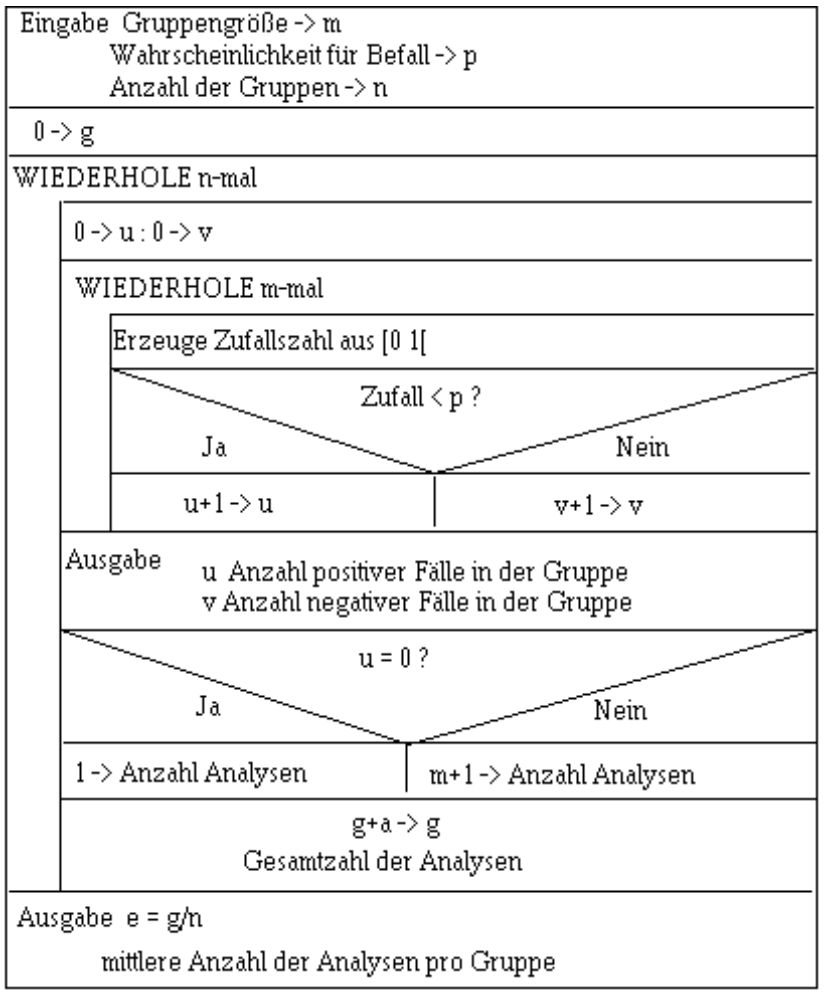
3. Wahrscheinlichkeitstheoretisches Modell

Anschließend entwickeln wir mit Hilfe der Kenntnisse aus der Wahrscheinlichkeitsrechnung ein mathematisches Modell. Hierzu berechnen wir den Erwartungswert der Anzahl Z der notwendigen Untersuchungen in Abhängigkeit der vorgegebenen Wahrscheinlichkeit p und der willkürlich zu wählenden Gruppengröße m. Dies ermöglicht dann, die eventuelle Ersparnis des Gruppenverfahrens gegenüber dem Einzelverfahren zu berechnen und diese zu optimieren, d.h. die Gruppengröße m_0 für die größte Ersparnis zu finden. Für verschiedene Wahrscheinlichkeiten p können wir die jeweilige optimale Gruppengröße und die gewonnene Ersparnis tabellieren.

Lösungsskizze zur Simulation

Über die Taste APPS (Applications) rufen wir den Programm-Editor auf und wählen hier New zum Erstellen eines neuen Programms mit dem Namen "probe". Beim späteren Aufruf des Programms müssen wir "probe()" eingeben.

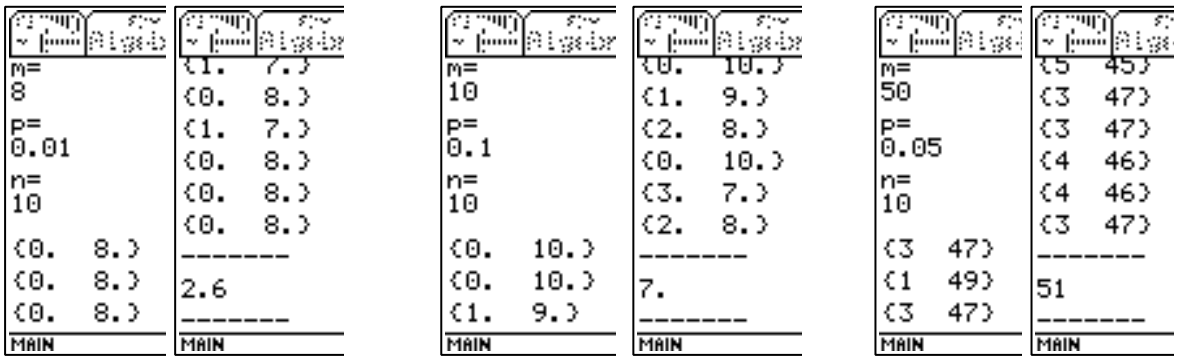
Struktogramm



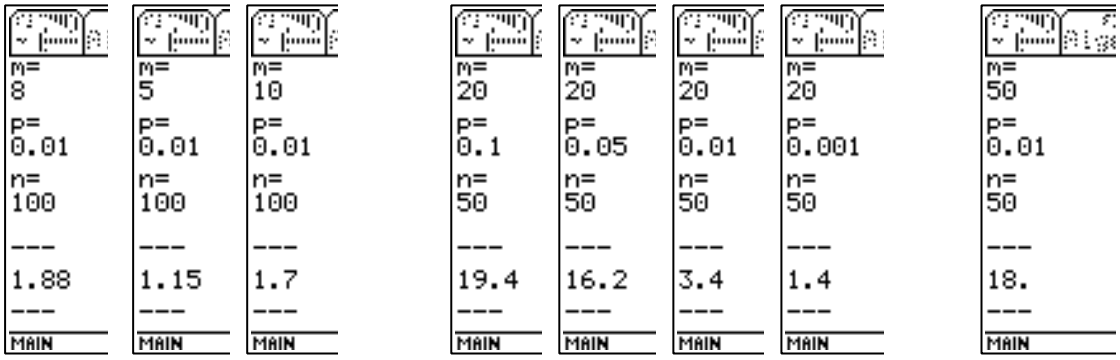
Programm

```
: probe()
: Prgm
: ClrIO
: Input " m=", m
: Input " p=", p
: Input " n=", n
: 0->g
: For i,1,n,1
: 0->u : 0->v
: For j,1,m,1
: rand() ->zufall
: If  zufall<p Then
: u+1->u
: Else
: v+1->v
: EndIf
: EndFor
: Disp {u,v}
: Pause
: If u=0 Then
: 1->a
: Else
: m+1->a
: EndIf
: g+a->g
: EndFor
: g/n->e
: Disp "------"
: Disp e
: Disp "------"
: EndPrg
```

Der Befehl `Pause` im Programm ermöglicht das schrittweise Abrufen (mit `Enter`) der einzelnen Simulationsergebnisse. Durch Experimentieren mit diesem Simulationsprogramm können wir nun einen ersten Überblick gewinnen. Zur Illustration hier drei Versuche:



Wenn man auf die Ausgabe der Einzelergebnisse verzichtet (Löschen von `Disp {u,v}` und `Pause`), kann man sich die mittleren Anzahlen bei längeren Simulationsserien ausgeben lassen.



Lösungsskizze zum Theoretischen Modell

1. Erwartungswert

Die Zufallsgröße Z beschreibe die Anzahl der notwendigen Untersuchungen für eine Gruppe vom Umfang m. Z kann die Werte 1 oder m+1 annehmen. Wenn p die Wahrscheinlichkeit für eine positive Einzelprobe ist, so berechnen wir die Wahrscheinlichkeiten

$$P(Z=1) = (1-p)^m \quad \text{und} \quad P(Z=m+1) = 1 - (1-p)^m$$

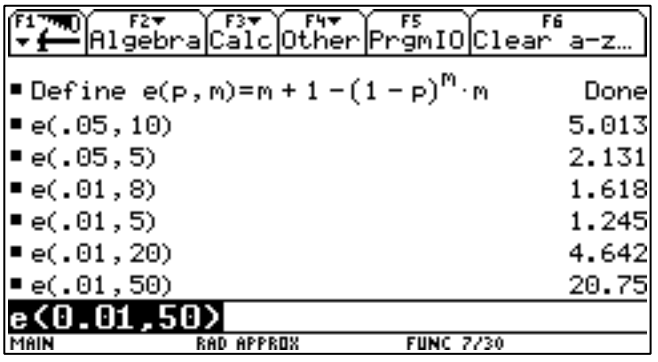
Für den Erwartungswert der Zufallsgröße Z gilt dann:

$$E(Z) = (1-p)^m * 1 + (1 - (1-p)^m) * (m+1)$$

$$\boxed{E(Z) = (m+1) - (1-p)^m * m}$$

Wir berechnen einige Erwartungswerte für verschiedene p und m und vergleichen mit den Simulations-ergebnissen.

Indem wir e als Funktion von p und m definieren
Define im Menu F4
können wir durch Aufrufen dieser Funktion beliebige Werte berechnen.



2. Ersparnis

Bei der Gruppengröße m ist die erwartete Ersparnis SP gegenüber den m Einzeluntersuchungen durch die Differenz m - E(Z) gegeben. Es gilt also

$$SP = m - ((m+1)-(1-p)^m * m) = m*(1-p)^m - 1$$

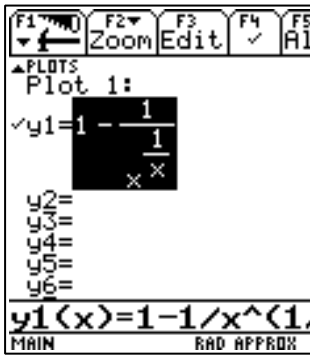
Zum besseren Vergleich benutzen wir die ”relative Ersparnis” c bezogen auf ein Tier:

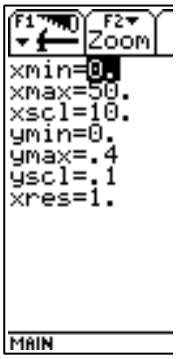
$$c = \frac{SP}{m} = (1-p)^m - \frac{1}{m}$$

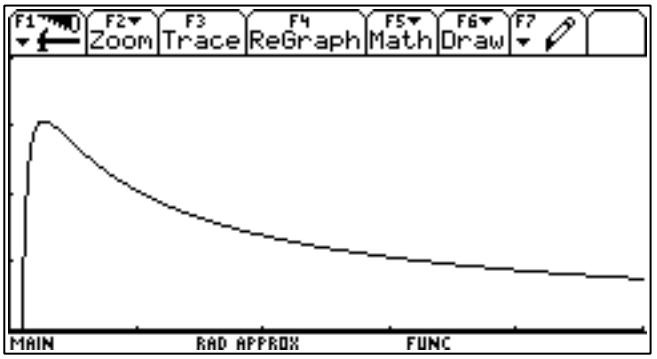
Günstig ist das Gruppenverfahren nur dann, wenn c>0. Dies ist der Fall, wenn

$$(1-p)^m - \frac{1}{m} > 0 \text{ und damit } \quad p < 1 - \frac{1}{\sqrt[m]{m}}$$

Wir geben die Funktion $x -> 1 - \frac{1}{\sqrt[x]{x}}$ ein und betrachten den Graph





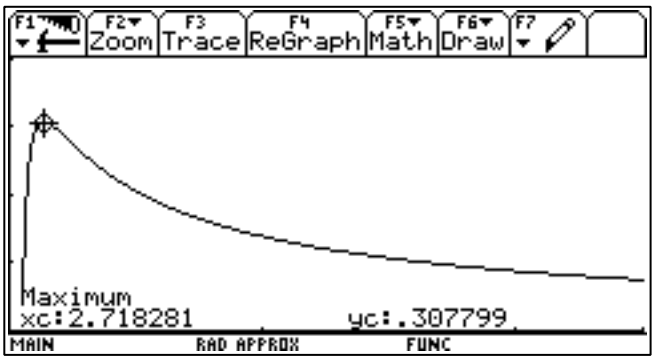


Interpretation des Graphen für unser Problem:

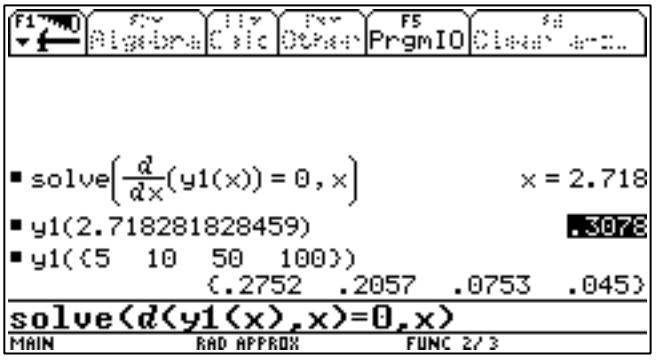
Für große Gruppengrößen muß p sehr klein sein, damit sich überhaupt eine Ersparnis ergibt. Für Werte von p, die größer als das Maximum der Funktion sind, bringt das Gruppenverfahren überhaupt keine Vorteile mehr, gleichgültig, welche Gruppen-größe m man wählt. Dieser Maximalwert liegt etwas über 0.3.

Wir wollen diesen Maximalwert genau bestimmen.

Hierzu benutzen wir in der Graphikdarstellung aus dem Menu F5 die Option Maximum: Nach Wahl einer unteren Grenze (lower bound? <- 0.1) und einer oberen Grenze (upper bound? <- 10) wird das Maximum berechnet und in der Graphik angezeigt.



Wir können das Maximum auch mit Hilfe der Nullstelle der 1. Ableitung berechnen.
Durch Einsetzen in y1(x) erhalten wir den Wert für die maximale Wahrscheinlichkeit.
Entsprechend können wir auch für bestimmte Gruppengrößen die Wahrscheinlichkeiten berechnen, oberhalb der eine Gruppenuntersuchung auf keinen Fall mehr eine Ersparnis bringen kann.



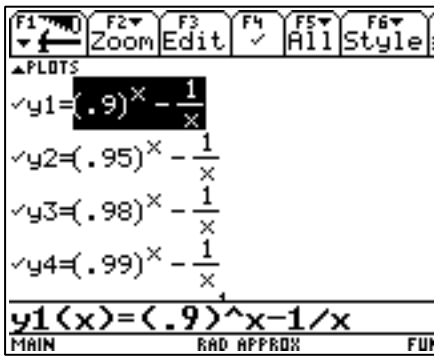
Mit unserem Rechner können wir dies in einem Schritt tun, indem wir in y1(x) für x eine Liste { } von Werten einsetzen.

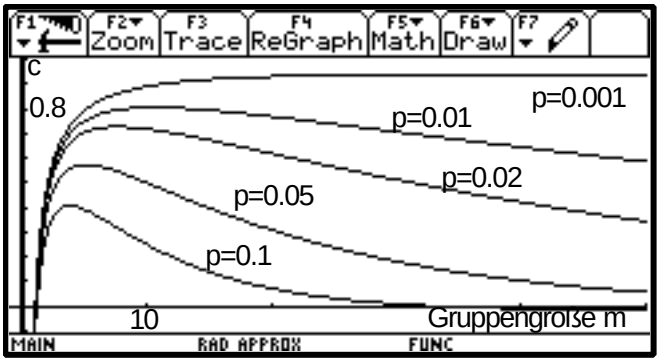
3. Optimale Gruppengröße

Bei festem (geschätztem) p können wir die relative Ersparnis in Abhängigkeit der gewählten Gruppengröße m untersuchen.

$$c_p = c_p(m) = (1-p)^m - \frac{1}{m}$$

Wir plotten Graphen von c_p für für verschiedene p:



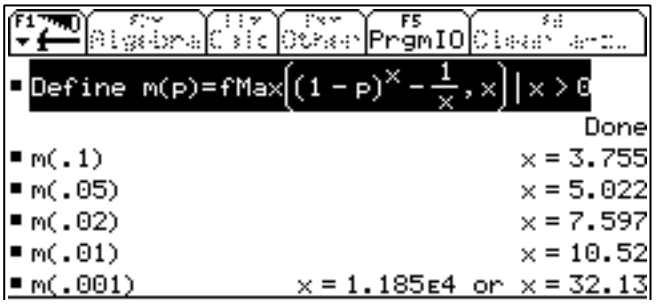


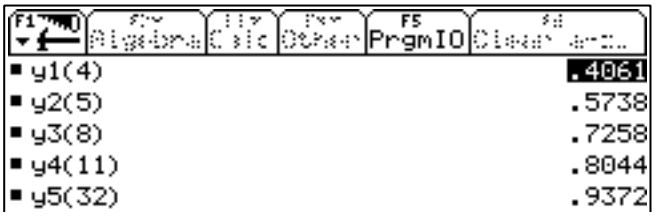
Interpretation der Graphen:

Für jede Wahrscheinlichkeit p gibt es offenbar eine Gruppengröße m, für die die relative Ersparnis c ein Maximum annimmt. Bei relativ großen p ist dieses Maximum recht scharf ausgeprägt, bei kleinerem p gibt es einen größeren Bereich für die Wahl einer günstigen Gruppengröße. Die relative Ersparnis kommt für kleine p nahe an 1.

Zur genaueren Bestimmung des jeweiligen Maximums können wir die `Trace`-Option `F3` oder die oben bereits benutzte Maximum-Option aus dem Menu `F5` benutzen. Wir wollen hier noch eine zusätzliche Möglichkeit der Berechnung zeigen. Dazu benutzen wir die Funktion `fMax` aus dem `Calc`-Menu und definieren damit m(p). Das Aufrufen dieser Funktion für verschiedene p liefert uns

die zugehörigen optimalen Werte für m. Damit können wir dann die optimale relative Ersparnis zu dem jeweilig passend gerundeten Wert errechnen.





Der für p=0.001 zusätzlich ausgegebene Wert von 11850 kann durch graphische und rechnerische Überprüfung ausgeschlossen werden. (Wieso liefert der Rechner diesen Wert?)

Zur abschließenden Dokumentation der Ergebnisse eignet sich eine Tabelle, in der wir zu verschiedenen Wahrscheinlichkeiten p die jeweils optimale Gruppengröße , die maximale relative Ersparnis c und die minimale relative Anzahl (1-c) der notwendigen Untersuchungen pro Individium angeben.

Wahrscheinlichkeit p	optimale Gruppen-größe m	maximale relative Ersparnis c	minimale relative Untersuchungszahl 1-c
0.001	32	0.94	0.06
0.01	11	0.80	0.20
0.02	8	0.73	0.27
0.05	5	0.57	0.43
0.1	4	0.40	0.60

Zusatzaufgaben und Erweiterungen

- graphische Darstellung der Abhängigkeit der minimalen relativen Untersuchungsanzahl in Abhängigkeit von der Gruppengröße

- Berechnung der optimalen Gruppengröße und Ersparnis, wenn die Gruppenanalyse gegenüber der Einzelanalyse mehr Kosten verursacht, z.B. das 5-fache (k-fache).

Literatur

[1] R.Dorfman, The detection of defective members of large populations, Ann.Math.Stat., Vol 14 (1943), pp 436-440

[2] J.B.Pomeranz, On Pooled Testing, Math. Scientist, 1976, 1, Seiten 99-105

[3] Thomas H. Kick, Das Problem der unechten Münzen, MNU 35 (1982), Heft 1, Seiten 23-26

[4] Kultusministerium Rheinland-Pfalz, Handreichung zum Lehrplan Mathematik , Grund- und Leistungsfach in der Oberstufe des Gymnasiums (Mainzer Studienstufe) - Der Computer als Werkzeug im Mathematikunterricht Teil 2 (1988), Seite 199-205, Worms 1988

[5] Ministerium für Bildung und Kultur Rheinland-Pfalz, Handreichung zum Lehrplan Mathematik , Grund- und Leistungsfach in der Oberstufe des Gymnasiums (Mainzer Studienstufe) - Der Computer als Werkzeug im Mathematikunterricht Teil 3 (1992), Seite 53-55, Worms 1992

[6] Glaser/Scheid/Wellstein (Hrsg), SIGMA Grundkurs Stochastik, Seite 80, Stuttgart 1982

Kurzkommentar zur Literatur:

Der Artikel von Dorfman [1] ist grundlegend, er bezieht sich auf eine zu der Zeit (1943) aktuelle Anwendung bei der Armee: " The method will be described by showing its application to a large scale-project on which the United States Public Health Service and the Selective Service System are now engaged. The object of the program is to weed out all syphilitic men called for induction. Under this program each prospective inductee is subjected to a ´Wassermann-type` blood test...." . (Die umstrittene Anwendung von Pool-Tests spielt auch in der gegenwärtigen Aufarbeitung des "Aids-Skandals" (Blutkonserven) in der Bundesrepublik eine Rolle).

Der Dorfmann-Artikel wurde in den Folgejahren in vielen Abhandlungen wissenschaftlicher Zeitschriften aufgegriffen, in denen das von Dorfmann geschilderte Verfahren weiter untersucht wird. Ein solcher Artikel ist der von Pomeranz [2.2], hier findet man auch weitere Literaturhinweise.

In dem Artikel von Kick [3] wird das gleiche Verfahren in interessantem Zusammenhang mit Münzwägeproblemen in einer für die Schule relevanten Weise bearbeitet.

In dem in [4] dargestellten Unterrichtsbeispiel wird das Verfahren in einem Anwendungszusammenhang zur Lebensmittelüberwachung (Untersuchung von Hausschweinen auf Trichinen) behandelt. Der Computereinsatz bezieht sich dabei auf die Verwendung "kleiner" BASIC-Programme und eines Funktionenplotters. In [5] wird ein Vorschlag gemacht, wie eine Abituraufgabe zu diesem Thema aussehen könnte. Eine entsprechende Aufgabe für den Unterricht ist in dem Schulbuch [6] aufgeführt, hier ist der Einsatz des Computers noch nicht intendiert.