

## **Erfolgreich testen und analysieren – geht das? Bildung stabiler Testgruppen für den Einzelhandel**

Martin Debus  
Bauer Systems KG  
Burchardstr. 11  
Hamburg  
martin.debus@bauermedia.com

Dr. Sergej Steinberg	Stefan Callsen
Bauer Systems KG	Bauer Systems KG
Burchardstr. 11	Burchardstr. 11
Hamburg	Hamburg
sergej.steinberg@bauermedia.com	stefan.callsen@bauermedia.com

### **Zusammenfassung**

Um neue Märkte zu gewinnen oder aber bestehende Marktanteile zu erhalten bzw. auszubauen besteht eine wesentliche Tätigkeit von Produktions- und Handelsunternehmen darin, neue Produkte einzuführen, Produkte- oder Preise zu variieren, Liefermengen zu optimieren oder schlicht Aufmerksamkeit zu schaffen. Ob nun Produktrelaunch, Anzeigenkampagne oder Preissenkung – alle diese Maßnahmen finden größtenteils entweder zeitlich begrenzt oder regional begrenzt (auf sog. Testmärkten) statt.

Die Auswertung bzw. Bewertung solcher Testmaßnahmen gestalten sich oftmals schwierig, da Absatzzahlen – auch abseits aller Kampagnenbeeinflussungen – vielfach starken Schwankungen unterliegen, etwa aufgrund von Saisonalität, Feiertagen, Wetter, regionalen Vorlieben oder Reaktionen der Wettbewerber. Ebenso begleiten die Handelsstufen (Groß- vs. Einzelhandel) diese Kampagnen unterschiedlich und tragen ggf. differenziert zum Ergebnis bei.

Im Folgenden wird ein in SAS (BASE/STAT) umgesetztes Verfahren vorgestellt, welches die Bewertung der o.g. Testmaßnahmen unterstützen soll. Die Beispiele beziehen sich auf die Bildung von Testgruppen für den Zeitschrifteneinzelhandel, können jedoch auf ähnliche Fragestellungen angewandt werden.

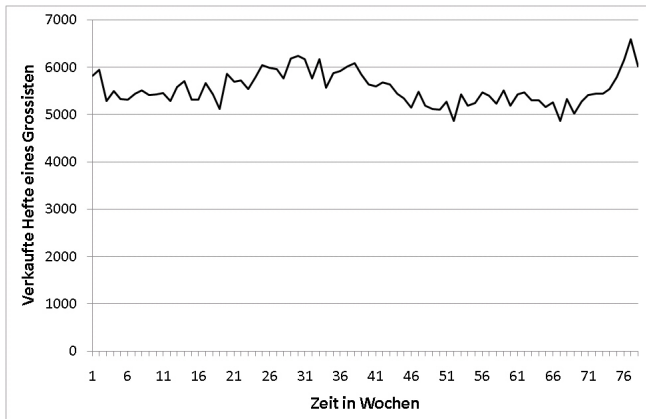
**Schlüsselwörter:** Testen, Verlagswesen, Handel, Zeitreihen

## **1 Problemstellung und Datenlage**

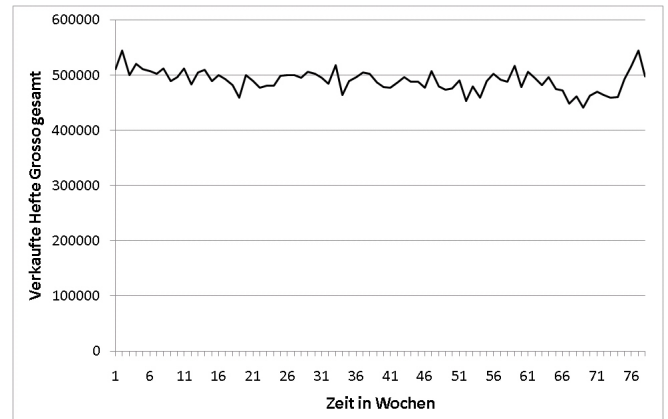
Den Ausgangspunkt zu den hier vorgestellten Überlegungen stellt die Anforderung dar, im Einzelhandel testweise vorgenommene Maßnahmen für einzelne Titel messbar zu machen und damit deren Wirkung auf den Absatz zu beziffern.

Abbildung 1 zeigt die Summe der verkauften Hefte eines Titels, die von den Einzelhändlern eines zufällig ausgewählten Grossisten innerhalb von eineinhalb Jahren abge-

setzt wurden (wöchentliches Erscheinungsintervall). Durch verkaufsfördernde Maßnahmen ist für diese Gruppe ein Zuwachs des Verkaufs um maximal 1% anzunehmen. Dies würde für die Grossisten aus der Beispielreihe einen Anstieg von durchschnittlich 6700 auf 6767 verkaufte Hefte bedeuten. Dieser erwartete Anstieg ist aufgrund mehrerer Faktoren schwer messbar: Die in den Markt gegebene Auflage und die Gesamtzahl der verkauften Hefte (Abbildung 2) schwankt.



**Abbildung 1:** Zeitreihe verkaufter Hefte eines Grossisten



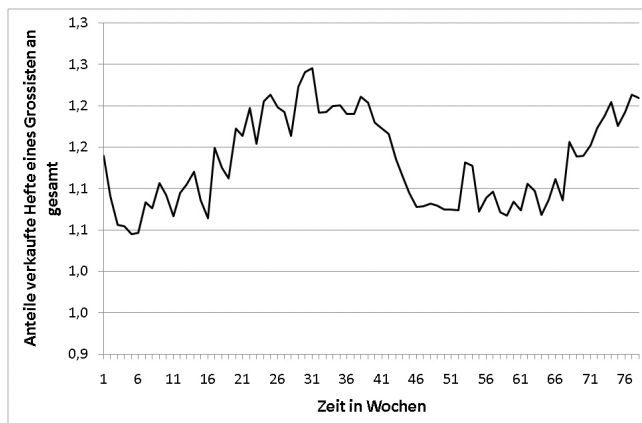
**Abbildung 2:** Zeitreihe verkaufter Hefte aller Grossisten

Als Lösung bietet es sich an, die Verkaufspersormance der einzelnen Grossisten ins Verhältnis zu den insgesamt verkauften Heften zu setzen (Abbildung 3):  $a(i)_t = \frac{v(i)_t}{V_t}$ , wobei  $a(i)_t$  der Anteil der verkauften Hefte bei einem Grossisten  $i$  zum Zeitpunkt  $t$  ist,  $v(i)_t$  die verkauften Hefte des Grossisten und  $V_t$  die Anzahl der verkauften Hefte insgesamt zum Zeitpunkt  $t$ .

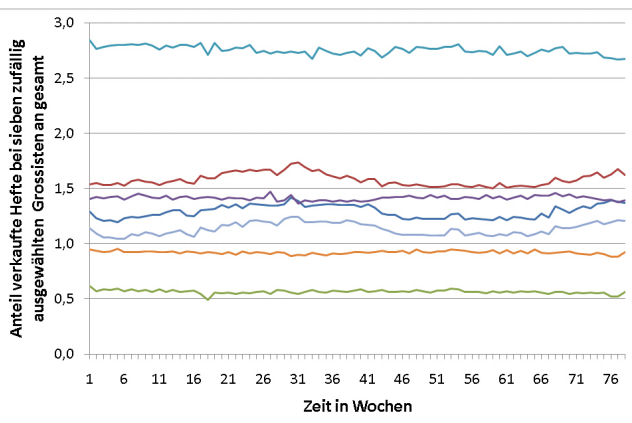
Nun werden  $k$  Grossisten (im Beispiel sieben Grossisten) nach dem Zufallsprinzip für den Test selektiert (Abbildung 4) und deren Anteile zur neuen Zeitreihe  $a(j)_t$  aufsummeriert:  $a(j)_t = \sum_1^k a(i_k)_t$ .

Die Zeitreihe (Abbildung 5) wird nun normiert, in dem lediglich die Abweichung um den Mittelwert erhalten bleibt:  $n(j)_t = a(j)_t - \overline{a(j)}$ .

Nach der Normierung der entstandenen Zeitreihe auf die Abweichung um den Mittelwert ergibt sich das in Abbildung 6 gezeigte Bild.

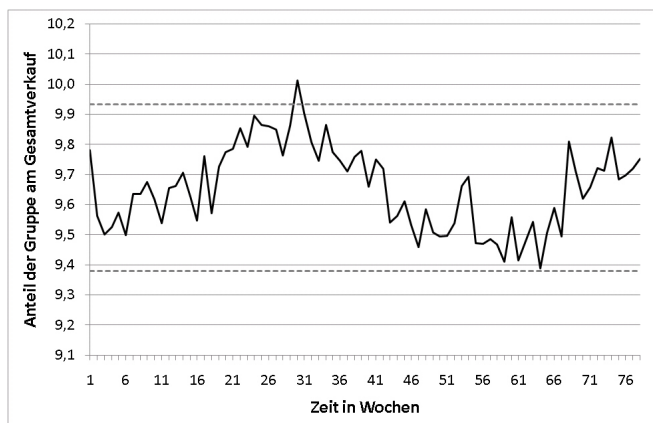


**Abbildung 3:** Zeitreihe der Anteile der verkauften Hefte eines Grossisten am Gesamtverkauf

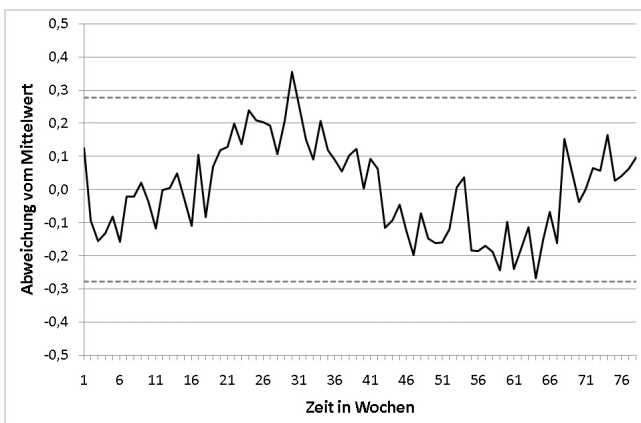


**Abbildung 4:** Zeitreihen eine Gruppe von zufällig ausgewählten Grossisten

Nach dieser Bereinigung verbleibt das Problem wie Steigerungen von weniger als 1% (für dieses Beispiel ein Anstieg des Anteils um mehr als 0,1 Prozentpunkte) sicher gemessen werden können. Das Rauschen dieser Reihe ist über das Konfidenzintervall um den Mittelwert von 0 mit einem Fehler von 5% ( $0 \pm 2\sigma$ ) dargestellt. Die Standardabweichung beträgt in diesem Fall bereits 0,27. Die Wahrscheinlichkeit ist damit groß, dass die Effekte der verkaufsfördernden Maßnahme im weißen Rauschen untergehen.



**Abbildung 5:** Zeitreihe der Anteile der verkauften Hefte einer Gruppe von Grossisten am Gesamtverkauf



**Abbildung 6:** Normierte Zeitreihe der Anteile der verkauften Hefte einer Gruppe von Grossisten am Gesamtverkauf

Die Frage lautet also: Wie muss die Gruppe von Grossisten ausgewählt werden, damit die Veränderungen in den Verkaufszahlen genauer gemessen werden können?

## 2 Verfahren zur Bildung stabiler Testgruppen

### 2.1 Algorithmus

Die Idee für den Algorithmus besteht darin, die Zeitreihen so miteinander zu kombinieren, dass sich das Rauschen ausgleicht und somit die Streuung (Standardabweichung) innerhalb der Gruppen minimiert wird. Bildhaft ausgedrückt ist die Frage also, welche Zeitreihen aufeinandergestapelt werden müssen, um eine möglichst gerade Linie zu erhalten.

Konkret sind dazu folgende Schritte nötig:

1. Berechnen der aktuellen Standardabweichung für alle normierten Zeitreihen bzw.

$$\text{Zeitreihengruppierungen } n(j) : \sigma(n(j)) = \sqrt{\frac{\sum_{t=1}^T ((\overline{n(j)}) - n(j)_t)^2}{T}}$$

2. Bilden von allen möglichen Paarkombinationen für diese Zeitreihen.
3. Addieren der Zeitreihenwerte der Paarkombinationen in der Form  $n(k) = n(j_1) + n(j_2)$ .
4. Berechnen der neuen Standardabweichungen für alle Paarkombinationen.
5. Das Paar mit der geringsten Standardabweichung wird zu einer neuen Zeitreihe gruppiert.
6. Überprüfung, ob das Entfernen eines der Unterobjekte aus der neuen Zeitreihe die Standardabweichung der neuen Zeitreihe minimiert.
7. Wenn ja, dann wird das Unterobjekt aus der Gruppierung entfernt.
8. Wiederhole die Schritte 1 bis 7, bis ein Abbruchkriterium erreicht ist.

Der Abbruch kann über die Begrenzung der Anzahl der Zeitreihen erfolgen, die pro Gruppe zusammengelegt werden dürfen. Weitere Abbruchkriterien sind die Limitierung für den Mittelwert der Zeitreihengruppen, die nicht überschritten werden darf, ein Minimalwert für die Anzahl der Gruppen, die entstehen sollen und eine Einschränkung wie oft die oben beschriebenen Schritte maximal ausgeführt werden sollen.

Eine solche gefundene Gruppe kann dann für einen Test verwendet werden, wodurch die Messgenauigkeit gegenüber einer zufälligen Auswahl der Gruppenteilnehmer deutlich erhöht wird.

### 2.2 Implementierung in SAS

Die Implementierung in SAS erfolgte über zwei parametergesteuerte Makros und ist allgemeingültig für sämtliche Problemstellungen, die sich über komplementäre Zeitreihen beschreiben lassen.

```

%macro create_groups(input1,
                    input2,
                    steps,
                    max_elements,
                    max_mean,
                    min_groups
                    );
%do i=1 %to &steps.;
  %if &break. = 0 %then %do;
    %group_items(input1=&input1.,
                input2=&input2.,
                step=&i.,
                max_elements=&max_elements.,
                max_mean=&max_mean.,
                min_groups=&min_groups.
                );
    proc append data=history_tmp01 base=history;
    run;
    proc append data=history_tmp02 base=history;
    run;
  %end;
%end;
%mend;

%let break=0;
%create_groups(input1=list1,
              input2=list2,
              steps=150,
              max_elements=12,
              max_mean=15,
              min_groups=8
              );

```

Das Makro `create_groups` hat sechs Parameter. Es werden eine Zeitreihe mit Werten pro zu gruppierenden Objekt übergeben, sowie eine Liste mit den Zuordnungen der Objekte zu Gruppen. Die Objekte können dabei schon vorgruppiert sein (d.h. man erzwingt, dass gewisse Objekte zu einer Gruppe gehören) oder sie sind nicht zugeordnet (d.h. jedes Objekt ist in einer eigenen Gruppe). Es werden darüber hinaus Parameter übergeben, die die Abbruchbedingungen für den Algorithmus bestimmen. Das sind die Anzahl der maximal auszuführenden Gruppierungsschritte, die Maximalanzahl der Elemente, die in einer Gruppe vorkommen dürfen, der maximale Mittelwert der Zeitreihengruppe und die Mindestanzahl von Gruppen, die nicht unterschritten werden soll, um zu verhindern, dass schließlich alle Zeitreihen zu einer Gruppe zusammengelegt werden.

Innerhalb dieses Makros wird das Makro `group_items` aufgerufen, in dem der eigentliche in Abschnitt 2.1 skizzierte Algorithmus implementiert ist. Im Dataset `history` wird pro Iteration festgehalten welche Elemente gruppiert wurden und ob ggf. Elemente wieder aus einer Gruppe entfernt wurden.

Der Quellcode zeigt einen Beispielaufwurf des Algorithmus mit den Listen list1 und list2 sowie einer Maximalanzahl von 150 Schritten, mit maximal 12 Elementen pro Zeitreihengruppe, deren Mittelwert 15 nicht überschreiten soll. Desweiteren sollen nicht weniger als 8 Gruppen gebildet werden.

**Tabelle 1:** Auszug aus der Tabelle History mit den einzelnen Gruppierungsschritten

Step	Std. A	Std. B	Std. AB	Text
1	0,0267	0,1068	0,0164	Neues Cluster 68 besteht aus: 45 und 56
2	0,0781	0,0466	0,0348	Neues Cluster 69 besteht aus: 22 und 53
3	0,0259	0,0678	0,0147	Neues Cluster 70 besteht aus: 17 und 52
4	0,0519	0,0515	0,0373	Neues Cluster 71 besteht aus: 48 und 57
5	0,0241	0,0572	0,0160	Neues Cluster 72 besteht aus: 20 und 62
6	0,0482	0,0537	0,0396	Neues Cluster 73 besteht aus: 1 und 59
7	0,0188	0,0506	0,0100	Neues Cluster 74 besteht aus: 4 und 54
8	0,0207	0,0497	0,0121	Neues Cluster 75 besteht aus: 36 und 67
9	0,0440	0,0491	0,0351	Neues Cluster 76 besteht aus: 24 und 61
10	0,0218	0,0475	0,0140	Neues Cluster 77 besteht aus: 5 und 63
11	0,0371	0,0434	0,0255	Neues Cluster 78 besteht aus: 3 und 60
12	0,0279	0,0439	0,0198	Neues Cluster 79 besteht aus: 6 und 64
13	0,0298	0,0434	0,0224	Neues Cluster 80 besteht aus: 40 und 58
14	0,0224	0,0396	0,0131	Neues Cluster 81 besteht aus: 41 und 73
14			0,0128	Element 59 aus Cluster 81 entfernt.
15	0,0211	0,0537	0,0126	Neues Cluster 82 besteht aus: 15 und 59
16	0,0361	0,0224	0,0108	Neues Cluster 83 besteht aus: 8 und 51

Tabelle 1 zeigt die ersten 16 Schritte des Algorithmus angewendet auf die beschriebenen Daten. In Schritt 14 wird ein drittes Element zu einer bereits bestehenden Gruppe von zwei Elementen hinzugefügt. Es stellt sich heraus, dass das Entfernen eines der alten Elemente die Standardabweichung reduzieren würde, also wird Element 59 aus der Gruppe entfernt.

### 3 Validierung des Verfahrens

Zur Validierung des Verfahrens wurden einerseits alle Grossisten einmal zufällig in 10 Gruppen aufgeteilt und einmal mittels des Algorithmus (Tabelle 2). Anschließend wurden die entstehenden Mittelwerte und Standardabweichungen der Verkaufsanteile der Gruppen berechnet. Bei der Zufallsgruppierung weisen die Gruppen eine durchschnittliche Standardabweichung von 0,096 auf, wobei die Werte hier in den einzelnen Gruppen stark schwanken. Bei der Gruppierung über den Algorithmus wird die durchschnittliche Standardabweichung in den Gruppen auf 0,042 gesenkt. Das bedeutet, dass sich das Rauschen innerhalb der Gruppe ausgleicht und somit eventuelle von außen induzierte Ausschläge besser gemessen werden können.

**Tabelle 2:** Gruppierung zufällig und mit Minimierung der Standardabweichung

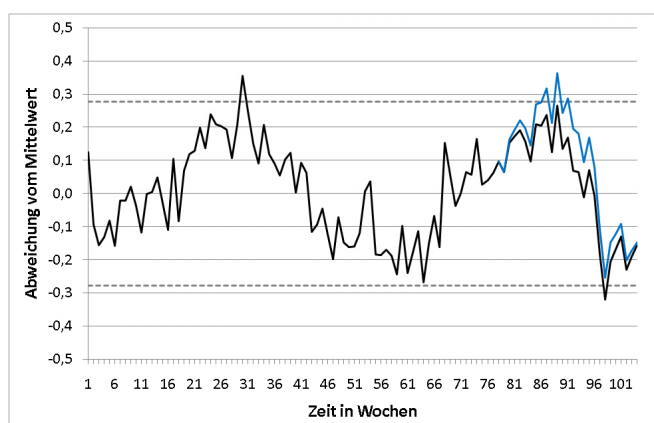
Zufallsgruppierung			Gruppierung über Algorithmus		
Gruppen-ID	Mittelwert	Std	Gruppen-ID	Mittelwert	Std
1	13,097	0,057	112	8,358	0,042
2	9,656	0,139	113	9,620	0,046
3	6,327	0,078	117	11,837	0,048
4	14,563	0,179	119	8,923	0,040
5	5,530	0,086	123	11,428	0,042
6	13,122	0,102	125	7,033	0,031
7	13,700	0,076	126	11,067	0,046
8	9,136	0,060	127	9,665	0,041
9	10,281	0,114	128	10,424	0,044
10	4,587	0,074	129	11,644	0,037
Mittelwert	10,000	<b>0,096</b>	Mittelwert	10,000	<b>0,042</b>

Eine weitere Validierung wurde durchgeführt, indem synthetische Testdaten erstellt wurden, die die erwarteten Effekte in den Verkaufsanteilen widerspiegeln sollen.

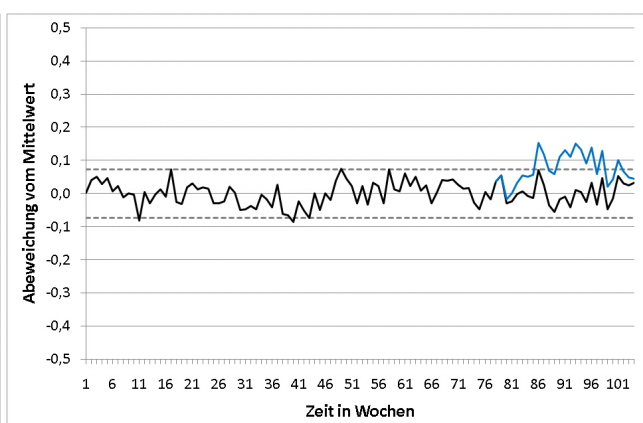
**Tabelle 3:** Faktoren für die Erzeugung der synthetischen Testdaten

<b>T</b>	78	79	80	81	82	83	84	85	86
<b>Faktor</b>	1,000	1,000	1,001	1,002	1,003	1,004	1,005	1,006	1,007
<b>T</b>	87	88	89	90	91	92	93	94	95
<b>Faktor</b>	1,008	1,009	1,010	1,012	1,014	1,015	1,014	1,012	1,010
<b>T</b>	96	97	98	99	100	101	102	103	104
<b>Faktor</b>	1,009	1,008	1,007	1,006	1,005	1,004	1,003	1,002	1,001

Hierbei wurden die Zeitreihen so modifiziert, dass sie einen um bis zu 1,5% höheren Verkauf simulieren. So wurde auf Basis der in Abschnitt 2 gebildeten Daten eine bestehende Zeitreihe von 104 Wochen ab Woche 78 mit einem Faktor belegt, der die Anzahl der verkauften Hefte in den gruppierten Elementen um bis zu 1,5% nach dem in Tabelle 3 dargestellten Schlüssel steigert.



**Abbildung 7:** Gruppierung über zufällige Auswahl



**Abbildung 8:** Gruppierung mit Minimierung der Standardabweichung

Verglichen wurde nun, inwieweit sich der daraus entstehenden Effekt in einer zufällig zusammengestellten Gruppe von Testgrossisten gegenüber einer mit dem Algorithmus zusammengestellten Gruppe besser messen lässt. Die Ergebnisse sind in den Abbildungen 7 und 8 dargestellt. Es ist sichtbar, dass das Konfidenzintervall bei der zufälligen Gruppierung deutlich größer ausfällt als bei Anwendung des Gruppierungsalgorithmus und die 2-Sigma-Grenzen bei der Gruppierung über den Algorithmus eindeutig durchstoßen werden.