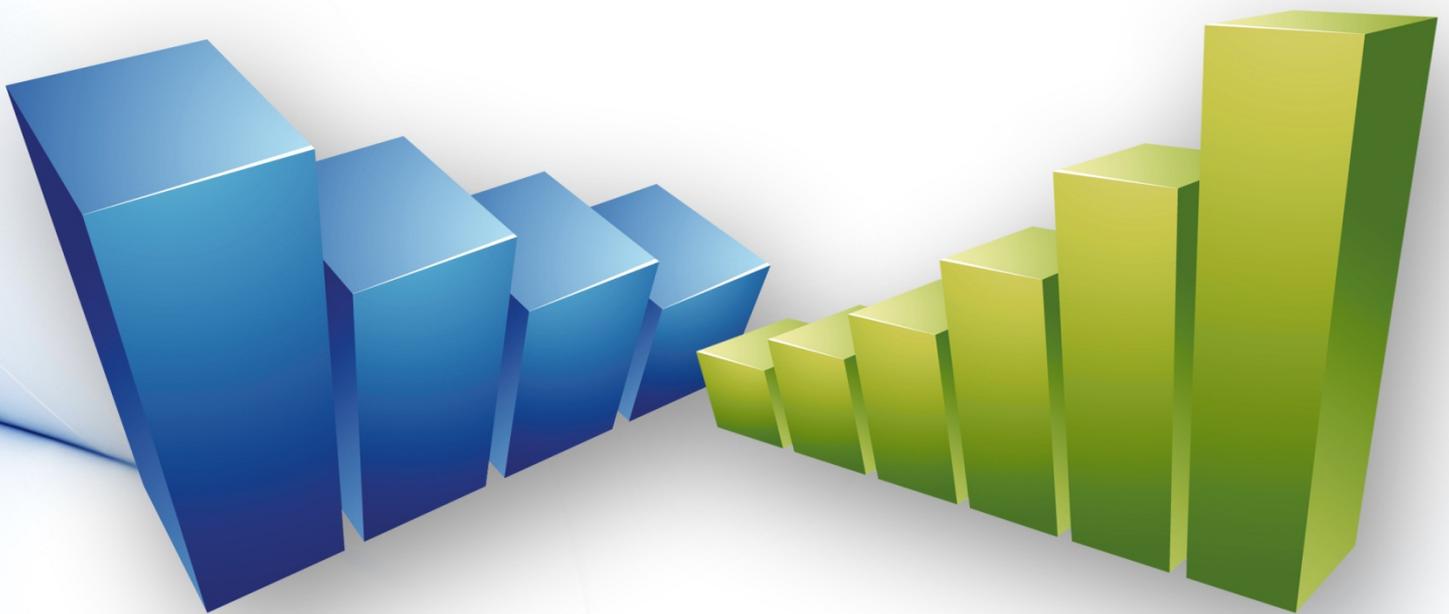


IfaD



G/TURF

Analysen in neuen
Geschwindigkeitsdimensionen



Inhaltsübersicht

1. Einführung
2. Grundlegende Begriffe
3. Aufgabenstellungen für TURF in der Marktforschung
4. Bislang bekannte Algorithmen für TURF-Analysen
5. Der innovative Algorithmus genPLUS
6. Validierung des Algorithmus genPLUS
7. Geschwindigkeit von genPLUS
8. Neue Einsatzmöglichkeiten durch hohe Geschwindigkeiten
9. Das Tool G/TURF
10. Zusammenfassung

1. Einführung

Die optimale Zusammenstellung von Produktportfolios oder Sortimenten gehört seit jeher zu den Aufgaben der Marktforschung. Mit wachsender Produktvielfalt wird diese Aufgabe immer bedeutender, aber auch deutlich anspruchsvoller. Die Anzahl der verfügbaren Produkte für ein Sortiment steigt, Regalplatz, Produktionskapazitäten oder Kosten zwingen aber zur Beschränkung. Ein etabliertes Verfahren zur Ermittlung optimaler Produktkombinationen ist die TURF-Analyse (Total Unduplicated Reach and Frequency). Sie berechnet Nettoreichweiten und Frequenzen von Produktkombinationen. TURF-Analysen stoßen bisher sehr schnell an Grenzen der Rechengeschwindigkeit heute verfügbarer Computer. Bei etwas anspruchsvolleren, aber realistischen Aufgaben benötigen sie astronomische Rechenzeiten für eine optimale Lösung. Steht wenig Zeit zur Verfügung, liefern sie häufig unbrauchbare Ergebnisse. Der innovative TURF-Algorithmus genPLUS, der hier im Weiteren vorgestellt wird, benötigt nur noch Bruchteile der Rechenzeiten ohne relevante Qualitätseinbußen und löst damit dieses Problem. Er ist in das anwenderfreundliche Tool G/TURF integriert und ermöglicht damit jedem die Bearbeitung von Aufgabenstellungen, die bisher unlösbar waren. Im Zusammenspiel mit anderen Verfahren, wie Conjoint-Analysen, ergeben sich zusätzlich neue Einsatzmöglichkeiten.

2. Grundlegende Begriffe

Zunächst ist es hilfreich, die beiden grundlegenden Begriffe aus der Mediaforschung für andere Anwendungen neu zu definieren, weil eine direkte Übernahme zu Unklarheiten oder sogar Fehlinterpretationen führen kann. Im Falle der Planung einer Produktlinie bedeuten:

- *Nettoreichweite*: Anzahl der Personen in der Stichprobe, für die zumindest eine der angebotenen Produktvarianten einer Produktlinie zum Kauf in Frage käme. In die Nettoreichweite geht jede Person also nur einmal ein, egal wie viele Produktvarianten sie kaufen würde.
- *Frequenz*: Anzahl der Produkte einer Produktlinie, die für eine Person zum Kauf in Frage kämen, summiert über alle Personen der Stichprobe. Es werden also Doppel- und Mehrfachauswahlen einer Person berücksichtigt.

3. Aufgabenstellungen für TURF in der Marktforschung

Ein Beispiel soll die Einsatzmöglichkeiten von TURF-Analysen in der Marktforschung und die Anforderungen an die Rechengeschwindigkeit zeigen.

Bestückung eines Snackautomaten

Für einen Snackautomaten auf dem Bahnhof kann der Betreiber aus 100 Produkten auswählen. Der Automat hat Platz für 30 verschiedene Snacks. Die Aufgabe besteht nun darin, diejenigen Produkte auszuwählen, die möglichst viele verschiedene Personen ansprechen und damit den höchstmöglichen Absatz bringen (Nettoreichweite). Basis für die Analyse ist eine Befragung potentieller Kunden mit einer skalierten Abfrage zur Kaufwahrscheinlichkeit der 100 verfügbaren Produkte. Nun reicht es nicht, den Automaten mit den 30 Produkten mit der höchsten durchschnittlichen Kaufwahrscheinlichkeit zu bestücken. Damit würden unter Umständen viele Personen mehrere kaufbare Produkte finden, aber nur eines kaufen wollen. Andere würden dagegen gar kein kaufbares Produkt finden. Es gilt also diejenige Produktkombination zu bestimmen, die für möglichst viele Personen ein kaufbares Produkt bereithält. Für den Snackautomaten in diesem Beispiel gibt es ca. 29.000.000.000.000.000.000.000 (29 Quadrillionen) Alternativen, ein Set mit 30 Produkten zusammenzustellen. Daraus die besten Sets zu finden, war bisher eine unlösbare Aufgabe. Die Rechenzeit für die Ermittlung der 100 Sets mit der besten Nettoreichweite beträgt mit herkömmlichen Algorithmen (vollständige Suche) geschätzte 330 Milliarden Jahre, mit dem neuen Algorithmus ca. 16,5 Stunden.

4. Bislang bekannte Algorithmen für TURF-Analysen

Die bislang bekannten und in der Praxis verwendeten Algorithmen zur TURF-Analyse lassen sich in drei verschiedene Klassen einteilen:

- *Vollständig*: Es werden Nettoreichweiten oder Frequenzen für alle Kombinationsmöglichkeiten für die gewünschte Set-Größe von Produkten berechnet. Die Vorgehensweise ist einfach und es wird eine zu 100% richtige Lösung gefunden. Mit der Größe der Aufgabenstellung steigt die Rechenzeit allerdings sehr stark überproportional an.
- *Heuristisch*: Es werden nicht alle Kombinationsmöglichkeiten durchgerechnet, sondern der Rechenprozess unter verschiedenen Annahmen abgekürzt. Diese Verfahren finden häufig viele gute Kombinationen in kurzer Zeit. Gerade bei größeren Aufgabenstellungen, bei denen die Zeitersparnis relevant wird, sind die Ergebnisse aber nicht verlässlich.
- *Direkt*: Ebenso wie beim heuristischen Ansatz werden nicht alle Kombinationen durchlaufen. Allerdings verwerfen diese Ansätze nur Kombinationen, die aufgrund eindeutiger Kriterien nicht zu den besten Kombinationen gehören können. Dadurch sparen auch sie erheblichen Zeitaufwand ein, bei gleicher Zuverlässigkeit wie die vollständigen Suchläufe. Bei größeren Aufgabenstellungen steigt aber auch bei diesen Verfahren die benötigte Rechenzeit – abhängig von der Aufgabenstellung – häufig in praxisferne Regionen.

5. Der innovative Algorithmus genPLUS

Der neue genPLUS Algorithmus verknüpft ein Verfahren aus der genetischen Optimierung mit der heuristischen Suche. Die heuristische Suche zeichnet sich durch sehr hohe Geschwindigkeit aus. Allerdings werden im Allgemeinen nicht alle besten Kombinationen einer vorgegebenen Set-Größe gefunden. Zudem findet dieser Ansatz häufig nicht die geforderte Anzahl an Kombinationen. Genetische Verfahren liefern im Gegensatz zu der heuristischen Suche bei weitem bessere Ergebnisse und auch eine ausreichende Zahl von Kombinationen, brauchen dafür aber eine längere Rechenzeit. Durch die Integration der beiden Verfahren werden ihre Stärken gebündelt und die Nachteile der einzelnen Ansätze vermieden. In extrem kurzer Zeit können damit nahezu 100% richtige Lösungen gefunden werden. Geringfügige Abweichungen in den gefundenen Kombinationen gegenüber der vollständigen Suche ergeben sich eher im hinteren Bereich von Kombinationen mit schlechteren Reichweiten oder Frequenzen und sind damit für den praktischen Einsatz nicht relevant.

6. Validierung des Algorithmus genPLUS

Eine verlässliche Validierung des neuen genPLUS Algorithmus ist naturgemäß nur bei Aufgabenstellungen möglich, bei denen durch die Anwendung des vollständigen oder direkten Algorithmus eine sichere Analyse aller Kombinationsmöglichkeiten in einem vertretbaren zeitlichen Rahmen zu realisieren ist. Für alle Testaufgaben, die unter diesen Bedingungen durchgeführt wurden, konnte der neue Algorithmus die optimalen Produktkombinationen zu nahezu 100% ermitteln. Produktkombinationen mit der besten Reichweite wurden immer gefunden. Dabei wurden direkte Suchläufe mit über 60 Tagen Laufzeit in die Validierung einbezogen.¹ Die Stabilität der Ergebnisse aller Validierungsstudien und die Struktur von genPLUS lassen darüber hinaus die Annahme zu, dass die optimalen Produktkombinationen auch jenseits dieser Grenzen verlässlich gefunden werden.

7. Geschwindigkeit von genPLUS

Der Algorithmus genPLUS ermöglicht die Lösung von Aufgaben, von denen man vorher Jahre entfernt war. Die nachfolgende Tabelle zeigt einige Beispiele für die Zeitersparnis. Die Rechenzeiten für den vollständigen Algorithmus im Bereich mehrerer Jahre sind auf Basis kleinerer Aufgabenstellungen im Bereich von bis zu 5 Tagen hochgerechnet. Gefordert sind jeweils die 100 Kombinationen mit den höchsten Reichweiten.

Aufgabe	Set-Größe (Anzahl Produkte in Kombination)	Anzahl alter- nativer Produkte	Anzahl Kombinatio- nen (ca.)	Rechenzeit vollständiger Algorithmus (P-STAT)	Rechen- zeit genPLUS
Testaufgabe	12	49	92 Milliarden	9 Stunden	2,5 Minuten
Ermittlung eines optimalen Pro- duktportfolios	10	76	1 Billiarde	4,5 Tage	4,5 Minuten
Produktlinien- Optimierung mittels Conjoint	5	5.760	53 Milliarden	600 Jahre ²	3,5 Stunden
Bestückung eines Automaten	30	100	29 Quadrillionen	330 Milliarden Jahre ³	16,5 Stunden

Abbildung 1: Übersicht der Leistungsfähigkeit hinsichtlich Geschwindigkeit des neuen Tools anhand ausgewählter Beispiele

¹ Die umfangreichste Validierung beruhte auf einer Suche nach den hinsichtlich Reichweite besten Sets aus 10 Items aus einer Gesamtheit von 200 Items. Daraus resultieren über 22 Milliarden mögliche Kombinationen. Die 1.000 besten davon waren zu ermitteln. Ein vollständiger Suchlauf würde dafür bei heute verfügbaren Rechnern mehrere tausend Jahre dauern. Der Datensatz umfasste 1.000 Fälle.

² Hochgerechnet für den Fall, dass es P-STAT erlauben würde, mehr als 210 alternative Produkte zu berücksichtigen

³ Hochgerechnet

8. Neue Einsatzmöglichkeiten durch hohe Geschwindigkeiten

Durch die außerordentliche Geschwindigkeit des neuen Algorithmus eröffnen sich viele neue Einsatzmöglichkeiten für TURF-Analysen. Eine davon stellen wir im Folgenden vor.

Entwicklung eines Produktportfolios auf Basis einer Conjoint-Analyse

Neue Anwendungsmöglichkeiten gibt es für die praktische Umsetzung von Ergebnissen aus Conjoint-Analysen zur Entwicklung eines optimalen Produktportfolios. Aus Conjoint-Analysen ergeben sich schon aus wenigen Produktmerkmalen und Ausprägungen Beurteilungen für eine sehr große Anzahl unterschiedlicher Produktvarianten. Die Kaufbereitschaft kann zusätzlich im Anschluss an das Conjoint für Produkte mit unterschiedlicher Wertigkeit abgefragt werden. Daraus kann die Kaufbereitschaft für alle Produkte anhand ihrer Conjoint-Nutzenwerte errechnet werden. Für ein einfaches Conjoint mit 7 Merkmalen und insgesamt 25 Ausprägungen ergeben sich schon 5.760 mögliche Produkte mit ihren Kaufwahrscheinlichkeiten. Daraus ein Portfolio mit 5 Produkten zu ermitteln, das möglichst viele verschiedene Käufer anspricht, erfordert die Prüfung von 53 Milliarden Kombinationsmöglichkeiten. Eine Lösung mit herkömmlichen Methoden benötigt ca. 600 Jahre. Mit dem neuen Algorithmus ist die Aufgabe in 3,5 Stunden gelöst.

9. Das Tool G/TURF

Die umfassenden Möglichkeiten der TURF-Analyse mit dem genPLUS Algorithmus sind in das anwenderfreundliche Tool G/TURF integriert. Die Nutzer kommen mit durchdachten Voreinstellungen und automatischen Anpassungen immer sicher zu verlässlichen Ergebnissen. Anhand der Aufgabenstellung (Anzahl Produkte, auszugebende Anzahl bester Ergebnisse etc.) werden automatisch die besten Verfahrenseinstellungen bestimmt, so dass ein Anwender keine speziellen Kenntnisse zu den einzelnen Verfahren besitzen muss. Zusätzlich sind neben dem genPLUS Algorithmus auch ein vollständiger und ein direkter Algorithmus implementiert.

Auf das Online-Tool hat der Anwender von überall ohne vorherige Installation Zugriff. Dadurch entfallen Barrieren, es in Unternehmen einzusetzen. Die Verbindung wird auf einer sicheren Leitung hergestellt und ist passwortgeschützt. Der Anwender greift automatisch immer auf die aktuellste Version zu. Eigene Analysen kann der Anwender über die gezielte Weitergabe von Rechten mit Kollegen oder Kunden teilen.

Die wichtigsten Features von G/TURF im Überblick:

- Wahl zwischen vollständigem, direktem und genPLUS Algorithmus
- Wahl zwischen Reichweiten- und Frequenz-Optimierung
- Sortierung nach Reichweite und Frequenz
- Ausgabe von bis zu 10.000 Kombinationen je Set-Größe
- Definition von "Must haves" und "No goes"
- Gewichtung der Fälle und/oder Produkte
- Signifikanztests
- Ausgabe von Shapley values

Kurzdarstellung von G/TURF anhand eines Fallbeispiels

Ein Automatenbetreiber überlegt, ob er Automaten größerer Kapazität aufstellen und wie er diese dann bestücken soll. Zurzeit besitzt er Automaten mit einer Kapazität von 15 Produkten. Zur Unterstützung seiner Entscheidung steht eine Datei zur Verfügung, die von 500 Befragten die Angaben zu 100 Produkten enthält, ob sie diese an einem Automaten kaufen würden (1) oder nicht (0).

Abbildung 2 zeigt die Einstellmöglichkeiten für die Berechnung in G/TURF. Die im Hinblick auf die Beantwortung der Fragestellung wichtigsten werden im Folgenden erläutert.

Eine vollständige Suche würde bei dem gegebenen Umfang der Aufgabenstellung viele Jahre in Anspruch nehmen und ist somit nicht praktikabel. Insofern wird das Verfahren genPLUS gewählt.

The screenshot displays the G/TURF web application interface. The top navigation bar includes a 'LOGOUT' button, the user's role 'Workshop (workshop)', and the application name 'G/TURF'. A 'Zurück zur Übersicht' button is also present. The main content area is divided into several sections:

- Projekt:** A sidebar menu with options: anlegen, kopieren, freigeben, löschen, auswählen.
- Snackautomat:** A menu with options: bearbeiten, Datenimport, Variablen, Selektionen.
- Modell:** A menu with options: anlegen, kopieren, löschen, auswählen.
- R&R:** A menu with options: Bezeichnung, Analyse-Sets, Bedingungen, **Optionen** (highlighted), Berechnung.

The main configuration area is titled 'Modelloptionen' and contains the following settings:

- Verfahren:** genPLUS (selected in a dropdown)
- Selektion:** Es wurden noch keine Selektionen definiert.
- Fall-Gewichtung:** Es wurden noch keine Variablen mit den Fall-Gewichtungen definiert.
- Item-Gewichtung verwenden:**
- Minimale Set-Größe:** 15
- Maximale Set-Größe:** 20
- Sortierung:** Reichweite (selected in a dropdown)
- Frequenz:** Summe (selected in a dropdown)
- Untere Grenze (inklusive):**
- Shapley values:**

The 'Signifikanztests' section includes:

- Vergleich der besten Kombination jeder Set-Größe:**
- Vergleich aller Kombinationen innerhalb einer Set-Größe:**
- Irrtumswahrscheinlichkeit:** 1 %

The 'Ausgabe' section includes:

- Sprache:** Deutsch (selected in a dropdown)
- Datumsformat:** 04.04.2012 09:52 (selected in a dropdown)
- Nachkommastellen:** 3
- Dezimaltrennzeichen:** , (selected in a dropdown)
- Tausendertrennzeichen:** . (selected in a dropdown)
- Anzahl beste Ergebnisse:** 10
- Tabelle Alternativen:**
- Individuelle Ergebnisse:**
- Kombinationen:**

A 'speichern' button is located at the bottom right of the configuration area.

Abbildung 2: Ein Blick auf die praxisrelevanten Einstellmöglichkeiten von G/TURF

Da die Automaten derzeit maximal 15 Produkte fassen können, wird als minimale Set-Größe 15 gewählt. Der Automatenbetreiber möchte aber insbesondere wissen, inwieweit sich größere Automaten lohnen. Insofern wird als maximale Set-Größe eine Zahl größer als 15, z.B. 20 vorgegeben.

Bei der Sortierung ist das Optimierungskriterium zu bestimmen, d.h. ob die Kombinationen mit höchster Reichweite oder höchster Frequenz identifiziert werden sollen. Entscheidungsgrundlage für den Automatenbetreiber ist in diesem Fall die Reichweite. Entsprechend werden Kombinationen zunächst nach der Reichweite und Kombinationen gleicher Reichweite dann nach der Frequenz sortiert.

Neben den Algorithmen ist eine Besonderheit des Tools, Signifikanztests durchführen zu können. Zum einen besteht die Möglichkeit zu untersuchen, ob sich die maximalen Reichweiten bzw. Frequenzen zwischen den unterschiedlichen Set-Größen unterscheiden. Zum anderen können Signifikanztests innerhalb einer Set-Größe durchgeführt werden. Interessant ist diese Option, falls die beste Kombination nicht infrage kommt, man aber eine Kombination gleicher Größe mit zumindest nicht signifikant geringerer Reichweite bzw. Frequenz (entsprechend dem gewählten Optimierungskriterium) bestimmen möchte. Für die beschriebene Fragestellung ist der erste Signifikanztest relevant. Gewünscht wird, für jede Set-Größe die 10 besten Kombinationen zu bestimmen.

Die Tabellen 1 und 2 zeigen Ausschnitte aus der csv-Datei, die alle angeforderten Ergebnisse enthält und nach Abschluss der Berechnungen zum Download bereitgestellt wird.

Analyse-Set	Set-Größe	Rang	Kombination	Reichweite (abs)	Reichweite (%)	Frequenz (abs)
1	20	1	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 93 96 97	440	88,0	2180
1	20	2	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 84 91 97 100	440	88,0	2176
1	20	3	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 96 97 100	440	88,0	2175
1	20	4	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 77 81 91 97 100	440	88,0	2172
1	20	5	3 12 15 26 28 29 33 34 35 37 38 42 46 70 71 72 81 91 97 100	440	88,0	2170
1	20	6	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 93 96 98	440	88,0	2167
1	20	7	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 85 91 97 99	440	88,0	2165
1	20	8	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 84 87 91 97	440	88,0	2160
1	20	9	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 87 91 96 97	440	88,0	2159
1	20	10	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 85 91 97 100	440	88,0	2156
1	20	10	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 77 81 87 91 97	440	88,0	2156

Tabelle 1: Die Kombinationen mit 20 Produkten mit der höchsten Nettoreichweite

In Tabelle 1 sind die besten Kombinationen mit 20 Produkten mit der höchsten Nettoreichweite dargestellt. Die auf dem ersten Rang enthält u.a. die Produkte mit den Kodierungen 3, 12, 15, etc. Mit dieser wird die maximale Reichweite von 440 erzielt. D.h. für 440 der 500, also 88% der Befragten ist unter diesen 20 Produkten mindestens eins dabei, das sie kaufen würden. Für diese Kombination ergibt sich eine Frequenz von 2180. Die folgenden 9 Kombinationen unterscheiden sich nicht in der Reichweite, aber in der Frequenz. Die letzten beiden sind bezüglich beider Werte identisch, so dass sie auf demselben Rang liegen und insofern nicht nur die 10 besten, sondern die 11 besten Kombinationen ausgegeben werden.

In Tabelle 2 bedeutet ein "x", dass die Reichweite der besten Kombination einer Set-Größe signifikant höher ist als die Reichweite der besten Kombination einer anderen Set-Größe. In diesem Beispiel ist demnach bei richtiger Bestückung die maximale Reichweite eines Automaten mit einer Kapazität von 20 Produkten signifikant ($p < 0.01$) höher als die eines Automaten mit 15 Produkten, nicht aber signifikant höher als die eines Automaten mit 16 oder mehr Produkten. D.h. im Sinne einer signifikanten Verbesserung der Nettoreichweite würde sich ein Wechsel von Automaten mit einer Kapazität von 15 Produkten auf 20 Produkte lohnen.

Analyse-Set	Set-Größe	Rang	Kombination	Reichweite signifikant höher zur besten Kombination der Set-Größe ...					
				15	16	17	18	19	20
1	20	1	3 12 15 26 28 29 33 34 35 37 38 42 46 70 72 81 91 93 96 97	x	-	-	-	-	

Tabelle 2: Ergebnisse der Signifikanztests

10. Zusammenfassung

Das Tool G/TURF mit dem innovativen genPLUS Algorithmus, das für den Marktforscher entwickelt wurde, stellt eine für die Marktforschungs-Praxis bedeutende Optimierung der TURF-Analyse dar. Der neue Algorithmus macht viele TURF-Analysen überhaupt erst möglich und eröffnet damit Lösungsmöglichkeiten für bisher unlösbare Aufgabenstellungen.



Uhlandstraße 68
D-22087 Hamburg

Tel.: +49 (0)40 25 17 13 - 0
Fax: +49 (0)40 25 17 13 - 33
www.ifad.de

Geschäftsführer: Martin Cyrus

Multivariate Analysen, Statistik:

Johannes Lüken
Tel.: +49 (0)40 25 17 13 - 22
E-Mail: jlueken@ifad.de
