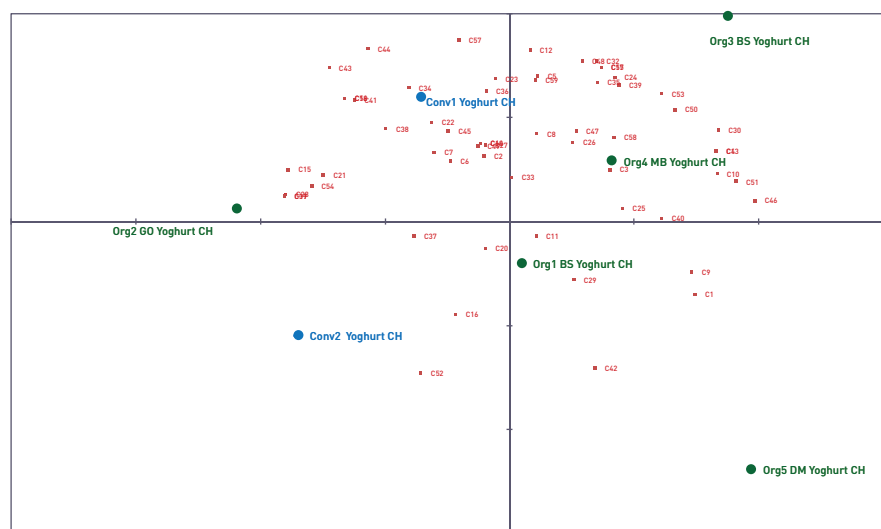


Abb. 7: Internal Preference Mapping (Biplot) für Joghurt (basierend auf einem Verbrauchertest mit organisch und konventionell hergestellten Produkten, durchgeführt in der Schweiz), [Forschungsprojekt ECROPOLIS]



d. h. die Aussagekraft der Ergebnisse erhöht wird, je mehr Produkte ausgewählt werden und je unterschiedlicher sich die Produkte in ihren Eigenschaften sind, welche in die Analyse einbezogen werden.

Sogenannte *Internal Preference Mappings* basieren ausschließlich auf Daten aus einem Verbrauchertest mit Konsumenten. Diese hedonischen Daten (z. B. Einzelwerte der Gesamtbeliebtheit oder Beliebtheit von Geruch, Geschmack, etc.) werden dann mittels PCA (siehe Kapitel 5.4.1) als Akzeptanzlandkarte im 2-dimensionalen Raum visualisiert (Abb. 7).

Produkte, für die Konsumenten ähnliche Vorlieben zeigen, gruppieren sich in diesen Landkarten zusammen. Je grösser die Unterschiede hinsichtlich der Beliebtheit / Akzeptanz, umso grösser die Abstände in der Darstellung. Nicht ersichtlich in einem Internal Preference Mapping sind die sensorischen Eigenschaften, welche die Beliebtheit beim Konsumenten ausmachen. Die Aussagekraft des Internal Preference Mapping ist somit eingeschränkt. Geeignet ist ein Internal Preference Mapping (Abb. 7) v.a. um Konsumentengruppen mit ähnlichen Präferenzen zu identifizieren – auch wenn die Anzahl an Konsumenten im Test nicht für ein „echtes“ Clustering ausreicht.

Für ein *Extended Internal Preference Mapping* werden in die PCA (siehe Kapitel 5.4.1) der Beliebtheitsdaten zusätzlich auch die Daten der deskriptiven sensorischen Analyse (derselben Produkte) als zusätzliche Variablen geplottet.

Hinweis: Biplots, welche mit Hilfe von Preference Mapping Techniken generiert werden, sollten als eine Art Landkarte interpretiert werden, da konkrete Skalenwerte weniger ausschlaggebend für die Interpretation sind, als die Positionierung von Konsumentendaten und Produkten resp. sensorischen Attributen.

5.3.3 Conjoint Analyse

Unter *Conjoint Analyse* versteht man eine multivariate Methode, welche eine Art ganzheitlicher Betrachtung von Beliebtheitsaussagen seitens des Konsumenten ermöglicht. Man geht dabei davon aus, dass Konsumenten einzelne Eigenschaften von Produkten für sich individuell einschätzen

und bewerten und dass sich die Summe dieser Bewertungen dann bei jedem Konsumenten zu einer Gesamtbewertung aufsummiert. Einsatzgebiet der Conjoint Analyse ist neben der Produktentwicklung die Marktsegmentierung. Computergestützt wird simultan die Präferenz beim Konsumenten abgeschätzt und eine Markteinteilung nach Präferenz vorgenommen.

5.4 B4 - Verknüpfung von Verbraucherdaten mit analytischen Daten

5.4.1 Hauptkomponenten-Analyse (Principal Component Analysis / PCA)

Die Hauptkomponenten-Analyse (PCA) ist ebenfalls eine multivariate statistische Methode. Mittels PCA lassen sich v.a. analytische Daten gut im Hinblick auf Produktunterschiede visualisieren. Diejenigen Produkteigenschaften, welche sich

stark ähneln, werden dafür linear zu Variablen / Hauptkomponenten (principal components) zusammengefasst. Auf diesem Wege wird ein möglichst grosser Anteil der Gesamtvarianz der Daten im 2D-Raum dargestellt. Auch wenn der Gesamtgehalt an Information dabei verringert wird, erreicht man durch die Reduktion auf zwei Dimensionen eine bessere Übersicht und leichtere Interpretation.

Mittels PCA werden Datensätze so aufbereitet, dass objektive Fragestellungen, wie beispielsweise die Frage nach dem Unterschied zwischen Produkten (in Abb. 8, z. B. für Erdbeeryoghurt) und v.a. die Ursachen für diese Unterschiede, d.h. die sensorischen Eigenschaften, visualisiert und interpretiert werden können.

Die Abbildung 8 zeigt einen Biplot einer PCA mit acht Erdbeeryoghurts und zwölf Attributen. Die erste Hauptkomponente (F1) ist auf der x-Achse dargestellt, die y-Achse entspricht der zweiten Hauptkomponente (F2). In diesem Beispiel erklärt die erste Hauptkomponente fast 76 % der Unterschiede zwischen den Produkten, die zweite Hauptkomponente rund 16 %. Die Abstände zwischen den dargestellten Produkten lassen sich interpretieren. Produkte, die nah beieinander liegen, sind tendenziell relativ ähnlich in ihren sensorischen Eigenschaften und umgekehrt. Es bleibt jedoch jeweils zu prüfen, inwiefern dies auch in den sensorischen Dimensionen gültig ist, welche in F1 und F2 wenig Gewicht haben.

5.4.2 External Preference Mapping (Multiple Regressions-Analyse)

Um ein sogenanntes *External Preference Mapping* anzufertigen zu können, benötigt man sowohl analytische Daten, die mittels PCA visualisiert werden, als auch Verbraucherdaten (Einzeldaten oder Cluster), welche mittels Regressionsanalyse in diese Darstellung hinein projiziert werden. Mittels dieser Methode können wichtige Faktoren für die Akzeptanz von Produkten identifiziert werden.

Gleichzeitig lässt sich darstellen, in welcher Abhängigkeit die sensorischen Eigenschaften und ihre Beliebtheit beim Konsumenten stehen. Potenzial für Produktverbesserungen wird bei dieser Analyse ersichtlich.

Autoren:

Annette Bongartz, Dipl. oec. troph., Dozentin und Leiterin der Fachstelle Sensorik am Institut für Lebensmittel und Getränkeinnovation (ILGI)

Urs Mürset, Prof. Dr. sc. nat., Dozent und Leiter der Fachstelle Datenanalyse und Statistik am Institut für Angewandte Simulation (IAS)

ZHAW – Zürcher Hochschule für Angewandte Wissenschaften, Department Life Sciences und Facility Management (LSFM), CH 8820 Wädenswil, Schweiz



DLG e.V., Ausschuss Sensorik
 Eschborner Landstraße 122, 60489 Frankfurt am Main
 Telefon: 069/24788-360, Fax: 069/24788-8360
 E-Mail: B.Schneider@DLG.org; Internet: www.DLG.org/sensorikausschuss.html

Weiterführende Literatur:

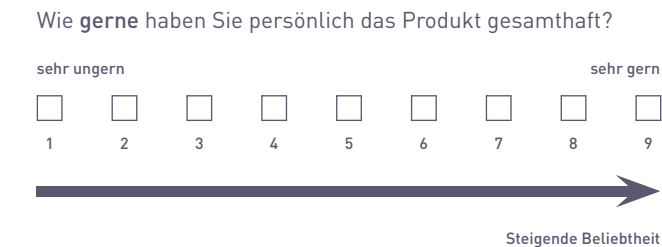
- Arditti, S.: Preference Mapping: a case study. Food Quality and Preference 8, 323-327, 1997.
- Backhaus, K., Erichson, B., Plinke, W. und Weiber R.: Multivariate Analysemethoden. Springer-Verlag, Berlin, Heidelberg, New York, 13. Auflage. 2010.
- Busch-Stockfisch, M. (Hrsg.): Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung. Loseblattsammlung. 21. Aktualisierungslieferung 05/09. Behr's Verlag GmbH & Co.KG. Hamburg, 2002.
- Derndorfer, E.: Lebensmittelsensorik. Facultas Universitätsverlag, Wien, 2006.
- Ellert, J.: Verbraucherakzeptanz und Preference Driving Factors bei Zuchtlachsen (Salmo Salar L.) unterschiedlicher Herkunft. Diplomarbeit. Hochschule für Angewandte Wissenschaften Hamburg, 2006.
- Faber, N. M., Mojte, J., Poelman, A. A. M.: Simple improvement of consumer fit in external preference mapping. Food Quality and Preference 14, 455-461, 2003.
- Greenhoff, K. and MacFie, H. J. H.: Preference Mapping in practice. In: H.J.H. MacFie und D.M.H. Thomson, Measurement of Food Preferences (S.137-147), Glasgow: Black Academic and Professional, 1994.
- Kemp, S.E., Hollowood, T. and Hort, J.: Sensory Evaluation: A practical handbook. Wiley-Blackwell, 2009.
- Lawless, H.T. and Heymann, H.: Sensory Evaluation of Food, principles and Practices. Maryland (Aspen Publishers), 1999.
- McEwan, J. A.: Preference Mapping for Product Optimization. In: Multivariate Analysis of Data in Sensory Science. Edited by T. Naes and E. Risvik, Elsevier Science, 1996.
- Moskowitz, H.R., Muñoz, A.M. and Gacula, M.C.: Viewpoints and Controversies in Sensors Science and Consumer Product testing. Trumbull, Connecticut (Food & Nutrition Press, Inc.), 2003.
- Meullenet, J.F.; Xiong, R and Findlay, C.J.: Multivariate and Probabilistic Analyses of Sensory Science Problems. First Edition, 2007, IFT Press, Blackwell Publishing, p. 207 ff
- Meyners, M., Kunert, J., Busch-Stockfisch, M. (Hrsg.): Multivariate Verfahren. In Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung. Behr's Verlag GmbH & Co.KG. Hamburg, 2002.
- O'Mahony, M.: Sensory Evaluation of Food: Statistical Methods and Procedures. Food Science and Technology; 16, 1986.
- Popper, R. and Gibes, K.: Workshop summary: Data analysis workshop: getting the most out of just-about-right data - Abstracts. Food Quality and Preference 15 (7-8): 891-899, 2004.

STATISTISCHE METHODEN IN DER SENSORIK (TEIL 2): „VERBRAUCHERTESTS“

Der Einsatz sensorischer Prüfverfahren und die damit verbundene Erhebung sowohl objektiver wie subjektiver Datenmengen ist nur dann sinnvoll und erfolgsbringend, wenn man diese Daten auch adäquat auswertet, d. h. Gesetzmäßigkeiten und Abhängigkeiten mittels geeigneter statistischer Methoden berechnet und auf ihre Signifikanz hin überprüft. *Analytische Prüfmethoden*, ausgeführt von sogenannten Fachpanels resp. trainierten Fachpersonen (Prüfern) und deren statistische Auswertemöglichkeiten standen bereits im Fokus des Arbeitsblattes 01/2011 des DLG Lebensmittel Magazins. Das vorliegende Arbeitsblatt befasst sich daher ausschließlich mit gängigen *Methoden der Verbraucherforschung*, d. h. den Verfahren zur Erhebung subjektiver Konsumentendaten, sowie v.a. den zugehörigen adäquaten statistischen Tests und Auswerteverfahren.

Ziel ist es dem Leser zunächst einen Überblick (Abb. 1) zu verschaffen und nachfolgend sowohl ein Grundverständnis als auch eine gewisse Handlungskompetenz auf diesem Gebiet zu vermitteln. Für diejenigen, die einen grösseren Bedarf an detaillierten Informationen und genauen Handlungsvorschriften für die Durchführung statistischer Analysen haben, existiert spezielle Fachliteratur. Interessant in diesem Zusammenhang sind sicherlich auch die weiterführenden Literaturquellen am Ende

Abb. 2: Hedonische Skala (9-teilig)



des Arbeitsblattes. In der Praxis kann es zudem oft hilfreich und effizient sein, professionelle Unterstützung und Beratung durch spezialisierte Institutionen in Anspruch zu nehmen.

1. Übersicht der Methoden

Im Rahmen der Verbraucherforschung werden sowohl *qualitative* als auch *quantitative Datenerhebungsverfahren* eingesetzt. Informationen, welche mittels qualitativer Methoden erhoben werden (Abb. 1, A), können ausschließlich durch Transkription, Gruppierung und/oder Zusammenfassung ausgewertet, dargestellt und interpretiert werden. Daten quantitativer Art (Abb. 1, B) können dagegen mit verschiedenen statistischen Auswerteverfahren analysiert werden. Abhängig vom Datenerhebungsverfahren unterscheidet man neben den rein deskriptiven Methoden zwischen parametrischen und nicht-parametrischen Verfahren. Es folgen jedoch zunächst ein paar grundsätzliche Informationen und Überlegungen, welche bei der Anwendung statistischer Methoden in der Verbraucherforschung hilfreich sind:

2. Gängige Skalentypen

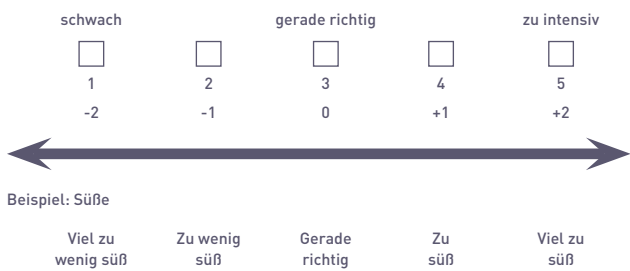
Häufigster Skalentyp, welcher im Rahmen von Verbrauchertests angewendet wird, ist die *9-teilige hedonische Skala* (Abb. 2), welche die Beliebtheit beim Konsumenten sinngemäss

Abb. 1: Methoden der Verbraucherforschung und zugehörige adäquate statistische Tests und Auswerteverfahren in der Übersicht

Methoden der Verbraucherforschung	Statistische Tests und Auswerteverfahren
A. Qualitativ	
1. Persönliches Interview (face-to-face), qualitativ, ohne festgelegte Struktur	Kein Einsatz statistischer Test-Methoden Auswertung resp. Wiedergabe und Interpretation der erhobenen Informationen durch - Transkription - Gruppierung - Zusammenfassung
2. Qualitative Befragung via Papier- oder Online-Fragebogen, mit festgelegter Struktur	
3. Fokusgruppe (moderierte Gruppendiskussion)	
4. Ethnografische Untersuchung (Beobachtung)	
B. Quantitativ	
1. Persönliches Interview (face-to-face) & Befragung via Papier- oder Online-Fragebogen, quantitativ, mit festgelegter Struktur	Auszählen der erhobenen Informationen Darstellen von Häufigkeitswahrscheinlichkeiten
2. Konzept-Tests (ohne Produkt) oder Produkt-Test (mit Produkt), z. B.: Akzeptanztests, Präferenztests (Rankings)	
a. 9-teilige Hedonische Skala	
b. JAR-Skala (just about right)	Deskriptive Statistische Methoden Mittelwerte, Standardabweichungen, Häufigkeitsverteilungen Induktive Statistische Methoden: 1. parametrisch - t-Test (Vergleich von 2 Produkten) - ANOVA zzgl. post hoc Test (Vergleich von > 2 Produkten) 2. nicht-parametrisch - Wilcoxon-Test (Vergleich von 2 Produkten) - Friedman-Test zzgl. post hoc Test (Vergleich von > 2 Produkten)
c. Rangierungsskala	
3. Segmentierung	Cluster-Analyse Preference Mapping (internal) Conjoint Analyse
4. Verknüpfung von Verbraucherdaten mit analytischen Daten	Hauptkomponentenanalyse (PCA) Preference Mapping (external)

Abb. 3: JAR Skala (just about right)

Wie empfinden Sie persönlich die Eigenschaft des Produkts?



von „habe ich überhaupt nicht gern“ über „weder noch“ bis „habe ich außerordentlich gern“ abfragt. Alternativ können auch 7- oder 5-teilige hedonische Skalen eingesetzt werden, um die Beurteilung für den Konsumenten etwas einfacher zu gestalten. Eine Reduktion der Skalenunterteilung um 2 oder gar 4 Kategorien, geht jedoch zu Lasten der Aussagekraft eines Tests, da - analog zu diesem gröberen Raster – die Verteilung der Akzeptanz dann auch nur mit geringerem Detaillierungsgrad aufgezeigt werden kann.

Neben den beschriebenen hedonischen Skalen sind häufig auch Rangierungsskalen zur Erhebung von Präferenzdaten und auch sogenannte *JAR (just-about-right)* Skalen (Abb. 3) im Einsatz. In der Regel sind JAR-Skalen 5-teilig und eignen sich für die subjektive Intensitätseinschätzung verschiedener Produkteigenschaften – von „viel zu wenig (viel zu wenig intensiv)“ über „zu wenig (zu wenig intensiv)“, „gerade richtig“, „zu viel (zu intensiv)“ bis hin zu „viel zu viel (viel zu intensiv)“ – beim Konsumenten abzufragen.

3. Parametrische vs. nicht-parametrische statistische Auswerteverfahren

Bei enger Auslegung des Sachverhalts können im Zusammenhang mit der Erhebung der Gesamtbefindlichkeit via hedonischer Skala keine *parametrischen Testverfahren* angewendet werden, da der eingesetzte Skalentyp keine metrischen (Intervall- oder Verhältnis-skalierten) Messwerte erhebt. Anders als bei trainierten Prüfern, die z. B. eine Profilierungsskala als echtes Messinstrument nutzen, geben Konsumenten auf der hedonischen Skala nämlich eher qualitative Einschätzungen ab. Trotzdem werden bei der Auswertung von Verbraucherstudien verbreitet parametrische Verfahren (z. B. t-Test und ANOVA) angewendet, sofern die Daten zumindest normalverteilt erscheinen. Häufiger ist aber der Fall, dass Datensätze aus Verbraucherstudien tatsächlich nicht normalverteilt sind und dann auch definitiv *nicht-parametrisch* ausgewertet werden sollten (sondern alternativ mittels Wilcoxon-Test oder Friedman-Test).

4. Complete vs. Incomplete Block Design

Bei Verbraucherstudien empfiehlt sich die Anwendung sogenannter *complete block designs*, d. h. jeder der befragten Konsumenten evaluiert jedes Produkt. Nur wenn dies berücksichtigt wird, ist es möglich sogenannte gekoppelte Auswertungen, wie z. B. eine 2-faktorielle ANOVA, durchzuführen, welche mit dem 1. Faktor das Produkt und dem 2. Faktor die Prüfpersonen analysiert. Im Falle von sogenannten *incomplete*

block designs ist keine gekoppelte Analyse möglich und der 1-faktorielle Ansatz ist angeraten.

5. Statistische Tests und Auswerteverfahren – im Detail

Es folgt eine detailliertere Betrachtung gängiger statistischer Auswerteverfahren, v.a. im Bereich „Quantitative Verbraucherforschung“ analog zu Abb. 1 für die Themen: Konzept-/Produkt-Tests (B2), Segmentierung (B3) und Verknüpfung von Verbraucherdaten mit analytischen Daten (B4).

5.1 B2 – Deskriptive statistische Testverfahren (a)

Als einer der ersten Schritte in der Analyse von Daten sollte immer die grundlegende Ermittlung *deskriptiver statistischer Parameter* verfolgt werden - z. B. die Berechnung von Mittelwerten und Standardabweichungen. Im Falle von Akzeptanzdaten können mit Hilfe des Mittelwertes bereits erste Rückschlüsse auf die Beliebtheit eines Produktes resp. einer Produkteigenschaft gemacht werden. Die Streuung der Daten gibt zudem Aufschluss über die Heterogenität der Daten und eine Darstellung der Daten in Form von Häufigkeitsdiagrammen ermöglicht die Erfassung und Interpretation der eigentlichen Datenlage einer Stichprobe.

Sofern gleichzeitig berechnet wird, wie viele Prozent der Stichprobe einen bestimmten Skalenbereich abdecken, kann man sogenannte *Top-Boxes* ausweisen. Eine Top-2-Box gibt z. B. an, wie viele Prozent einer Stichprobe im Falle eines Akzeptanztests mit 9-teiliger hedonischer Skala die Skalenbewertung 9 und 8 vorgenommen haben. Sofern im Rahmen der Datenerhebung bestimmte Zielgruppen in unterschiedlicher demografischer Zusammensetzung (z. B. nach Alter, Geschlecht, etc.) angesprochen werden – wie dies bei Verbrauchertests sehr häufig der Fall ist – können die genannten statistischen Parameter auch separat für diese Untergruppen ausgewertet und miteinander verglichen werden. Dies nennt man *Split-Analysen*. Um sinnvolle Split-Analysen machen zu können, ist jedoch unbedingt zu beachten, dass die Gesamtstichprobe ausreichend groß gewählt wird.

Daten sollten immer auch auf Normalverteilung geprüft werden. In Abhängigkeit des Ergebnisses sind die geeigneten statistischen Tests auszuwählen. Bei normalverteilten Daten können parametrische Verfahren eingesetzt werden; sind die Daten hingegen nicht normalverteilt, werden nicht-parametrische Verfahren genutzt. Einerseits lässt sich eine Normalverteilung visuell (grob) abschätzen, indem man die Datenmenge als Verteilungskurve darstellt und mit dem Verlauf einer Normalverteilungskurve vergleicht. Andererseits kann man den Sachverhalt auch mittels statistischer Tests (z. B. Shapiro Wilk-Test) überprüfen. Der resultierende p-Wert gibt dann Aufschluss über die Art der Verteilung. Da es bei diesen Tests nicht auf den Alpha-Fehler ankommt, sondern auf den (um einiges grösseren) Beta-Fehler, empfiehlt sich die Entscheidungsgrenze eher bei $p = 0.01$ als bei $p = 0.05$.

5.2 B2 - Induktive statistische Testverfahren

5.2.1 a – parametrisch (1.)

t-Test

Unter der Voraussetzung, dass mittels 9-teiliger hedonischer Skala erhobene Daten beim Vergleich von maximal 2 Produkten

im Hinblick auf eine ausgewählte Produkteigenschaft, normalverteilt sind, ist es möglich – analog zur Auswertung analytischer Daten – einen *t-Test* zur Prüfung der Mittelwertunterschiede anzuwenden.

Varianz-Analyse (Analysis of Variance → ANOVA)

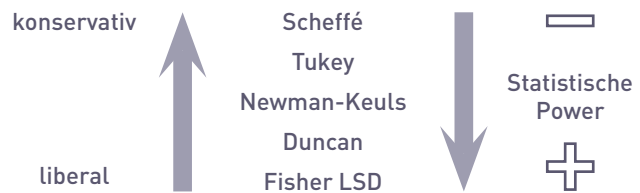
Werden mehr als 2 Produkte im Hinblick auf eine ausgewählte Produkteigenschaften miteinander verglichen, so ist es im Falle normalverteilter Daten, welche mittels 9-teiliger hedonischer Skala erhoben wurden – wiederum analog zur Auswertung analytischer Daten – üblich eine *Varianzanalyse* (ANOVA = „analysis of variance“) zur Prüfung der Mittelwertunterschiede durchzuführen.

Bei der Auswertung von Verbrauchertests sieht man es in der Praxis häufig, dass Daten - auch wenn sie mit einem „complete block design“ erhoben wurden - in der Regel nicht mit einem 2-faktoriellen, sondern lediglich einem 1-faktorieller Ansatz der ANOVA ausgewertet werden. Natürlich ist es bei Verbrauchertests so, dass man mit einer hohen Heterogenität unter den Befragten rechnen muss. Die Betrachtung des 2. Faktors (Prüfer) erscheint somit irrelevant. Um jedoch die Datenlage eines „complete block designs“ optimal auszunützen, ist es notwendig Daten gekoppelt in einem 2-faktoriellen Design, welches sowohl Produkt wie Prüfer berücksichtigt, zu analysieren.

Sofern mittels ANOVA dann signifikante Produkteffekte gezeigt werden können, muss mittels post-hoc-Tests lokalisiert werden, welche Produkte sich effektiv signifikant unterscheiden. Im Gegensatz zu klassischen paarweisen Tests, betrachten die post-hoc-Tests die Proben in ihrer kompletten Matrix. Dies geschieht mittels spezieller Teststatistiken für multiple Vergleiche.

Konfrontiert mit einer riesigen Auswahl an verschiedenen post-hoc-Tests, ist man als praktischer Anwender schnell überfordert. Man findet in der Literatur jedoch Empfehlungen und reduzierte Auswahllisten. Gängig sind in der Verbraucherforschung z. B. Tests wie: *Duncan, Fisher LSD, Newman-Keuls, Scheffé und Tukey*. Diese Tests unterscheiden sich voneinander in der Konservativität im Hinblick auf falsch-positive Signifikanzaussagen sowie in ihrer statistischen Power (Abb. 4).

Abb. 4: Einteilung der post-hoc-tests nach ihrer Konservativität und statistischen Power [Abbildung nach <http://www.reiter1.com/> [29.08.2011]]



Wählt man einen eher konservativen post-hoc-Test, muss man eine niedrige statistische Power (= höhere Chance auf einen Beta-Fehler) in Kauf nehmen, hat aber ein geringeres Risiko auf einen Alpha-Fehler und umgekehrt.

5.2.2 a – nicht-parametrisch (2.)

Wilcoxon-Test

Unter der Voraussetzung, dass mittels 9-teiliger hedonischer Skala erhobene Daten bei einem Vergleich von maximal 2 Pro-

dukten im Hinblick auf eine ausgewählte Produkteigenschaft, normalverteilt sind, kann man zur Prüfung von Mittelwertunterschieden einen *Wilcoxon-Test* einsetzen.

Friedman-Test

Werden mehr als 2 Produkte im Hinblick auf eine ausgewählte Produkteigenschaften miteinander verglichen, so eignet sich im Fall nicht normalverteilter Daten der *Friedman-Test* zur Prüfung der Mittelwertunterschiede. Voraussetzung für die Durchführung des Friedman-Tests ist, dass jeder Prüfer jede Probe genau einmal getestet hat. Es handelt sich um gekoppelte Stichproben. Die erhobenen Daten innerhalb einer Stichprobe müssen aber unabhängig voneinander sein. Beim Friedman-Test wird die Beurteilung der Proben jedes einzelnen Prüfers dann in eine Beliebtheitsrangfolge gebracht und es ergeben sich über alle Prüfer für jede Probe entsprechende Rangsummen. Aus diesen Rangsummen wird nachfolgend das Rangmittel gebildet, welches hier eine vergleichbare Rolle spielt wie die Mittelwerte bei einer Varianzanalyse.

Sofern der Friedman-Test signifikante Produkteffekte nachgewiesen hat, kann man mittels post-hoc-Tests feststellen, welche Produkte sich effektiv signifikant unterscheiden. Mögliche post-hoc-Tests sind *Fisher LSD* und *Nemenyi-Test*.

Hinweis: Der Friedman-Test wird auch bei der Auswertung ordinaler Daten (z. B. bei Präferenz-Tests / Rankings) eingesetzt, z. B. wenn 3 bis 6 Produkte einer Rangierung nach Beliebtheit unterzogen werden (Variante B2c).

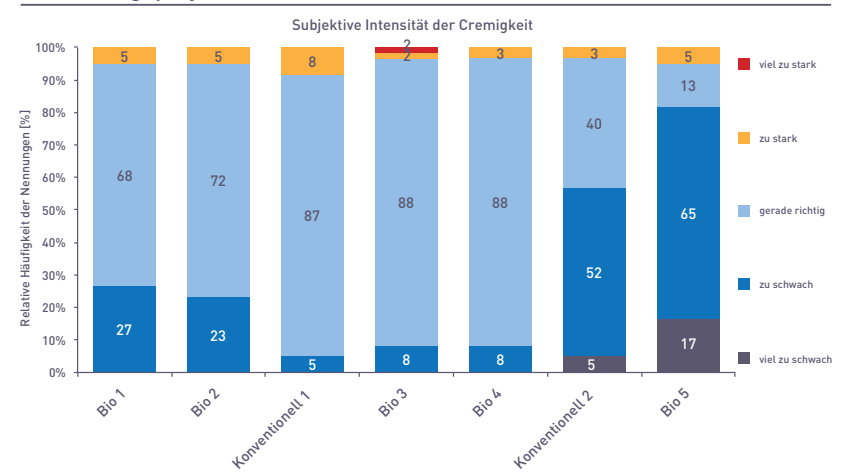
5.2.3 b - Penalty-Analyse

Im Rahmen von Verbrauchertests kommen neben den klassischen hedonischen Skalen (in der Regel 9-teilig) auch sogenannte „*just about right*“- oder *JAR-Skalen*, zum Einsatz. JAR-Skalen sind 5-teilige Skalen, die es ermöglichen die individuelle (subjektive) Einschätzung der Intensität verschiedener Produkteigenschaften vom Konsumenten zu erfragen.

Eine Darstellung der relativen Häufigkeiten der Skalenbewertungen (Abb. 5) gibt zunächst Aufschluss über die Datenlage und somit über die Einschätzung der Intensität der betrachteten Produkteigenschaft durch die Konsumenten. Aus einer derartigen Darstellung (Abb. 5) kann jedoch nicht abgeschätzt werden, wie gross der Einfluss ist, den ein Produkt mit einer Bewertung „viel zu stark (viel zu intensiv)“ oder „viel zu schwach (viel zu wenig intensiv)“ auf die Gesamtbefindlichkeit (erhoben via 9-teiliger hedonischer Skala) haben könnte. Hierzu muss eine sogenannte Penalty-Analyse durchgeführt werden. Beliebtheits- resp. Akzeptanzwerte (i. d. R. die Gesamtbefindlichkeit) werden dafür mit den JAR Daten kombiniert. Ziel ist es dabei aufzudecken, wo die untersuchten Produkte Verbesserungspotenzial besitzen.

Für die Berechnung der *Penalties* wird die 5-teilige JAR Skala zunächst zu einer 3-teiligen Skala modifiziert. Die Kategorien „zu stark“ und „viel zu stark“ resp. „zu schwach“ und „viel zu schwach“ werden jeweils zusammengefasst. Es resultieren die beiden Kategorien „zu stark“ und „zu schwach“. Die Akzeptanzmittelwerte derjenigen Konsumenten, welche das Produkt als „gerade richtig“ eingestuft haben, werden dann mit den Akzeptanzmittelwerten der Konsumenten verglichen, welche das Produkt als „zu schwach“ oder „zu stark“ eingestuft haben. Anschliessend werden die sogenannten Mittelwerts-Abfälle oder *mean drops* berechnet. D. h. es wird jeweils die Differenz zwischen dem Mittelwert der JAR-Kategorie („gerade richtig“)

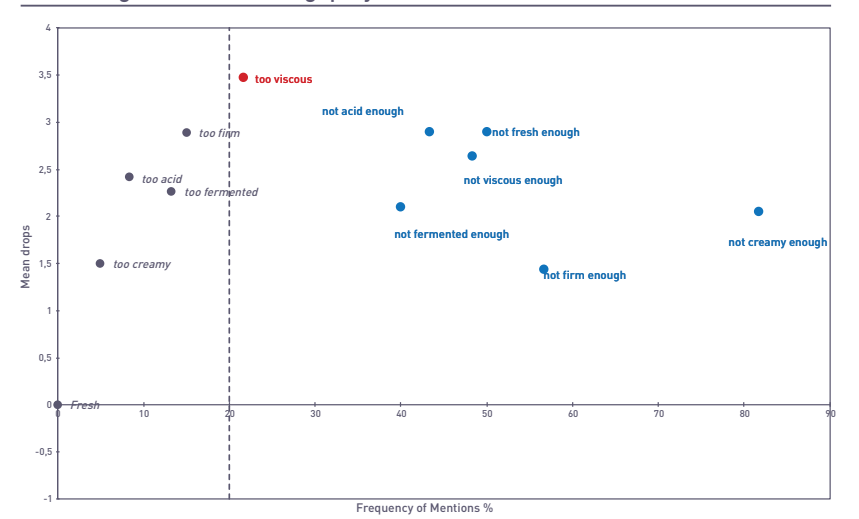
Abb. 5: Häufigkeitsverteilungen für den Deskriptor „Cremigkeit“ bei 7 verschiedenen Natur-Joghurts, ermittelt mit JAR-Skalen [Forschungsprojekt ECROPOLIS]



und der beiden nicht-JAR-Kategorien („zu schwach“ resp. „zu stark“) ermittelt. Mithilfe eines multiplen Vergleichstest zwischen den Mittelwerten der JAR- und der beiden nicht-JAR-Kategorien wird ein p-Wert berechnet, welcher eine Aussage bzgl. der Signifikanz der Unterschiede zwischen den Gruppen zulässt. Sofern ein signifikanter Unterschied gefunden wird, ist dies ein Hinweis darauf, dass die subjektive Intensitätseinschätzung auf der JAR-Skala die Ergebnisse der hedonischen 9-Punkteskala (Gesamtbefindlichkeit) beeinflusst. Je höher die Penaltät (Penalty), umso mehr Einfluss hat die betrachtete Eigenschaft auf die Gesamtbefindlichkeit (Akzeptanz). Die Penaltäten-Analyse ermittelt also, wie stark die Gesamtbefindlichkeit eines Produktes abnimmt, wenn eine Produkteigenschaft nicht als „gerade richtig“ eingeschätzt wird. Zur Visualisierung des Verbesserungspotentials einzelner Produkte im Hinblick auf bestimmte Merkmale ist es möglich die mean drops in Relation zur Häufigkeit der Nennungen (Abb. 6) darzustellen.

Achtung: Dieses Verfahren sollte nur angewendet werden, wenn mehrere (mindestens drei) Produkteigenschaften pro

Abb. 6: Darstellung der „mean drops“ eines ausgewählten Natur-Joghurts [Forschungsprojekt ECROPOLIS]



Produkt mittels JAR-Skala untersucht wurden. Außerdem ist es für die Relevanz resp. Aussagekraft der Penalty Analyse wichtig darauf zu achten, dass nur solche nicht-JAR-Kategorien in die Berechnung einbezogen werden, auf welche sich eine zuvor festzulegende Mindestanzahl an Prüfern vereint (Schwellenwert, meist in % angegeben). Je kleiner die Anzahl an Prüfern im Test, umso höher der geeignete Schwellenwert. Und nicht zuletzt ist es auch entscheidend für die Interpretation von JAR-Daten, im Hinterkopf zu behalten, dass es sich um Daten handelt, welche subjektive Intensitätseinschätzungen widerspiegeln. Es handelt sich zwar um Intensitätsmessungen, dennoch sagen diese etwas ganz anderes aus, als objektive

Intensitätsmessungen resp. deskriptive Daten, welche mit Hilfe von trainierten Prüfern erhoben wurden.

5.3 B3 - Segmentierung

5.3.1 Cluster-Analyse

Zur Cluster-Analyse gehört einerseits die *Hierarchische Cluster-Analyse*, welche v.a. für deskriptive Daten eingesetzt wird und dabei Produkte, die in ihren sensorischen Eigenschaften ähnlich sind, in Clustern zusammenfasst. Andererseits kennt man die *k-means-Cluster-Analyse*, welche v.a. im Zusammenhang mit Verbrauchertests zum Einsatz kommt und die Datenmenge im Hinblick auf die grössten Unterschiede im Akzeptanzverhalten hin unterteilt. Sofern für die Datenerhebung ein „complete-block-Design“ gewählt wurde, ist die k-means-Cluster-Analyse daher geeignet das Gruppenverhalten von Konsumenten zu analysieren, d.h. festzustellen, ob Untergruppen existieren, welche sich in ihrem Akzeptanzverhalten unterscheiden. In Abhängigkeit der Stichprobengröße des Verbrauchertests wird im Vorfeld der Analyse die Anzahl Cluster definiert, die gebildet werden soll.

5.3.2 Internal Preference Mapping (= Multiple Regressions-Analyse)

Unter dem Begriff *Preference Mapping* versteht man grundsätzlich verschiedene multivariate statistische Verfahren. Hauptsächlich lassen sich drei verschiedene Varianten unterscheiden:

- Internal Preference Mapping
- Extended Preference Mapping
- External Preference Mapping

Allen Preference Mapping Techniken gemein ist, dass nie weniger als 6 Proben gleichzeitig untersucht werden sollten. Generell kann man davon ausgehen, dass der resultierende Versuchsraum umso grösser wird,