

## Korrespondenzanalyse

Theoretische Grundlagen und die Anwendung in SPSS.

Ein Vortrag von Jana Bombik, Erdmute Jahn, Claudia Philipp und Helge Siems an der Hochschule Harz im Rahmen der Vertiefungsrichtung Marktforschung.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Gliederung

- ✘ Helge (Grundgedanke, Einordnung, Analyseschritte)
- ✘ Claudia (Zentroid, Distanzen, Streuung)
- ✘ Jana (Dimensionsreduktion, Geometrie der Spaltenprofile, simultane Darstellung der Spalten- und Zeilenprofile)
- ✘ Erda (Interpretation der Ausgabedateien)

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Helge Siems

Grundgedanke, Einordnung, Analyseschritte

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

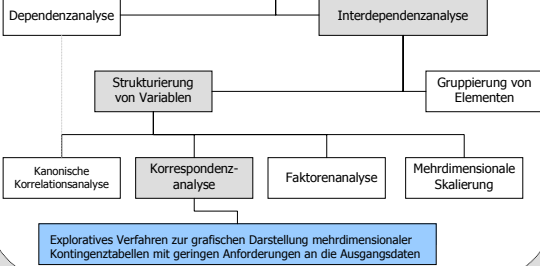
---

---

---

---

# Multivariate Analyseverfahren



Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Ziel der Korrespondenzanalyse

Zeilen und Spalten einer rechteckigen Matrix so analysieren, dass sie als Punkte in einem gering dimensionierten Raum (zwei oder drei Dimensionen) dargestellt werden können.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Basis: Häufigkeitstabelle

| SORTE   | GESCHMACK |        |         |  |
|---------|-----------|--------|---------|--|
|         | herb      | bitter | süßlich |  |
| Sorte 1 | 5         | 7      | 2       |  |
| Sorte 2 | 18        | 46     | 20      |  |
| Sorte 3 | 19        | 29     | 39      |  |
| Sorte 4 | 12        | 40     | 49      |  |
| Sorte 5 | 3         | 7      | 16      |  |

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

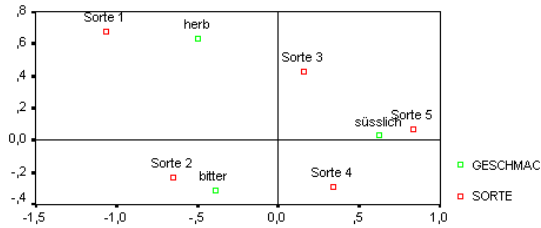
---

---

---

---

## Ziel: Grafische Darstellung



Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Geschichte / Herkunft

- Mathematische Basis: Singular Value Decomposition (SVD)
- Anwendung in der Bio- bzw. Psychometrik
- Reine numerische Darstellung
- Geometrische Form Beginn 1960er (Jean-Paul Benzécri)
- Anwendung in verschiedenen Wissenschaften
- Bedeutung in der Marketingforschung seit Ende 1980er (Hoffman / Franke, Backhaus / Meyer)

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Alternativen zu SPSS

- BASP - Programm für Cluster- und Korrespondenzanalyse (non-Profit Software-Projekt)
- ADE 4 - multivariate analysis and graphical display software (freeware)
- SAS JMP Version 5 - kommerzielles Statistikpaket

Links zu den Produkten auf <http://helge-siems.de/mafosoftware>

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Anforderungen an die Daten und Stärken der Korrespondenzanalyse

### Anforderungen

- Nicht negative Werte
- Rechteckige Matrix

### Stärken / Vorteile

- nicht-metrische Daten möglich (nominal, ordinal)
- Verteilung der Daten unerheblich

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Schritte der Korrespondenzanalyse

1. Erstellung der Ausgangsmatrix
2. Bildung der Spalten- und Zeilenprofile
3. Festlegung der geometrischen Schwerpunkte der Produkte und der Eigenschaften
4. Berechnung der Distanzen zwischen den Produkten und den Eigenschaften
5. Berechnung der Gesamtstreuung in den Daten
6. Reduzierung der Dimensionen

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Aktuelles Beispiel

- Bier-Geschmackstest
- 50 Konsumenten
- drei Produktmerkmale
- Jede Sorte kann zwischen 0 und 3 Merkmalen zugeordnet werden
- Ergibt nominalskalierte (binäre) Daten

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Ausgangsdaten

Korrespondenztabelle

| SORTE        | GESCHMAC |        |         | Aktiver Rand |
|--------------|----------|--------|---------|--------------|
|              | herb     | bitter | süßlich |              |
| Sorte 1      | 5        | 7      | 2       | 14           |
| Sorte 2      | 18       | 46     | 20      | 84           |
| Sorte 3      | 19       | 29     | 39      | 87           |
| Sorte 4      | 12       | 40     | 49      | 101          |
| Sorte 5      | 3        | 7      | 16      | 26           |
| Aktiver Rand | 57       | 129    | 126     | 312          |

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

Daten sind wegen unterschiedlicher  
Randhäufigkeiten erst nach  
Normierung vergleichbar

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

## Zeilen- bzw. Markenprofile

Zeilenprofile

| SORTE   | GESCHMACK |        |         | Aktiver Rand |
|---------|-----------|--------|---------|--------------|
|         | herb      | bitter | süßlich |              |
| Sorte 1 | ,357      | ,500   | ,143    | 1,000        |
| Sorte 2 | ,214      | ,548   | ,238    | 1,000        |
| Sorte 3 | ,218      | ,333   | ,448    | 1,000        |
| Sorte 4 | ,119      | ,396   | ,485    | 1,000        |
| Sorte 5 | ,115      | ,269   | ,615    | 1,000        |
| Masse   | ,183      | ,413   | ,404    |              |

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

## Spalten- bzw. Merkmalsprofile

### Spaltenprofile

| SORTE        | GESCHMACK |        |         |       |
|--------------|-----------|--------|---------|-------|
|              | herb      | bitter | süßlich | Masse |
| Sorte 1      | ,088      | ,054   | ,016    | ,045  |
| Sorte 2      | ,316      | ,357   | ,159    | ,269  |
| Sorte 3      | ,333      | ,225   | ,310    | ,279  |
| Sorte 4      | ,211      | ,310   | ,389    | ,324  |
| Sorte 5      | ,053      | ,054   | ,127    | ,083  |
| Aktiver Rand | 1,000     | 1,000  | 1,000   |       |

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

## Claudia Philipp

Zentroid, Distanzen, Streuung

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

## Festlegung des geometrischen Schwerpunktes

- Messung der in den Daten enthaltenen Streuungsinformation, indem die Abweichungen der Marken- bzw. Merkmalsprofile von ihrem geometrischen Schwerpunkt berechnet werden.
- Schwerpunkt stellt das Durchschnittsprofil der Punktwolke dar (Zentroid)

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

## Festlegung des geometrischen Schwerpunktes

- Datenvektor des Schwerpunktes  $q$  der Biersortenprofile lässt sich berechnen, indem man die Zeilenprofile  $r_i$  mit ihrer Masse  $w_i$  gewichtet und anschließend aufsummiert.

Formel:

$$q = \sum_i w_i r_i = w_1 r_1 + w_2 r_2 + w_3 r_3 + w_4 r_4 + w_5 r_5$$

### Besonderheit:

Koordination des geometrischen Schwerpunktes der Reihenprofile  $q$  entspricht exakt den Massen des Spaltenprofile  $q_j$

---

---

---

---

---

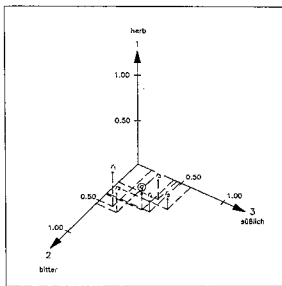
---

---

---

220

Abb. 50: Graphische Darstellung der Biersortenprofile und ihres geometrischen Schwerpunktes  $q$  im dreidimensionalen Raum



---

---

---

---

---

---

---

---

## Berechnung der Distanzen zwischen den Punkten

- Abstände der Punkte zueinander bzw. zum Schwerpunkt geben Aufschluss über die Ähnlichkeiten bzw. Unterschiede ihrer Profile
- **Punkte dicht beieinander:**  
große Gemeinsamkeiten hinsichtlich der Merkmalszuordnungen
- **Punkte weiter entfernt:**  
kaum Gemeinsamkeiten hinsichtlich der Merkmalszuordnungen
- **Punkte nah am Zentroiden:**  
Alternativen die dem Durchschnittsprofil ähneln

---

---

---

---

---

---

---

---

## Berechnung der Distanzen zwischen den Punkten

- Berechnung der räumlichen Entfernung zwischen zwei Punkten normalerweise mit Hilfe des euklidischen Distanzmaßes

➤ **Aber hier ungünstig da:**

- sie nicht berücksichtigt, dass die Dimensionen in Abhängigkeit von den Spaltenhäufigkeiten unterschiedliche Skalengrößen aufweisen
- Entfernungen zwischen den Produktpunkten würden durch die unterschiedlichen Massen verfälscht werden.
  - große Massen → starke Beeinflussung
  - kleine Masse → schwache Beeinflussung

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Berechnung der Distanzen zwischen den Punkten

- Berechnung mit **gewichteter euklidischer Distanz**  
- zusätzliche Gewichtung der Distanz mit der Masse

➤ Formel:

$$d^2_{(i,k)} = \sum_j \frac{(n_{ij} / n_i - n_{kj} / n_k)^2}{n_j / n} = \frac{(r_{ij} - r_{kj})^2}{q_j}$$

- Gewichtete euklidische Distanz ist proportional zur Chi<sup>2</sup>-Statistik, deshalb auch Chi<sup>2</sup>-Distanz genannt.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Vorteil der Chi<sup>2</sup>-Distanz

- Betrachter kann die Entfernung zwischen zwei Punkten in der räumlichen Darstellung als euklidische Distanz interpretieren.
- Durch die Gewichtung der Punktekoordinaten mit den Massen entspricht sie der gewünschten Chi<sup>2</sup>-Distanz, so dass man die Verzerrungseffekte, die durch die unterschiedliche Spaltenhäufigkeiten hervorgerufen werden, eliminiert.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---



## Berechnung der Gesamtstreuung in den Daten

- Berechnung der räumlichen Abstände eines Punktes vom geometrischen Schwerpunkt i. d. R. mit Chi<sup>2</sup> - Distanz

$$d^2_{(i,k)} = \sum_j \frac{(n_{ij}/n_i - n_{kj}/n_k)^2}{n_j/n} = \frac{(r_{ij} - q_j)^2}{q_j}$$

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Berechnung der Gesamtstreuung in den Daten

- Berechnung der Gesamtstreuung „total inertia“ um zu verdeutlichen, wie stark die einzelnen gewichteten Markenprofile um den Schwerpunkt streuen

$$\sum_i w_i d^2_i = \sum_{ij} w_i \frac{(r_{ij} - q_j)^2}{q_j} = Chi^2 / n = in(I)$$

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Jana Bombik

Dimensionsreduktion, Spaltenprofile, simultane Darstellung der Spalten- und Zeilenprofile

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Dimensionsreduktion

### Ziel:

- Bestimmung der optimalen Unterräume bei möglichst geringem Informationsverlust
- Graphische Deutung ist nur bei **zwei** höchstens drei Dimensionen möglich

$$K = \min(r,c)-1 \quad K = \text{max. Dimensionalität}$$

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

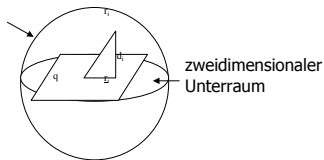
---

---

---

## Dimensionsreduktion

Dreidimensionaler Ursprungsraum



$r_i$  = beliebiger Profilpunkt

$l_i$  = Punkt im Unterraum mit kürzester Entfernung zu  $r_i$

$d_j$  = Entfernung zwischen  $r_i$  und  $l_i$

$q_j$  = Masse der Spaltenprofile

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Dimensionsreduktion

### Vorgehensweise:

- Bildung des Durchschnitts über die kürzeste Entfernung aller Punkte der betrachteten mehrdimensionalen Punktwolke zum Unterraum
- Methode: „Singular Value Decomposition“ (SVD)

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

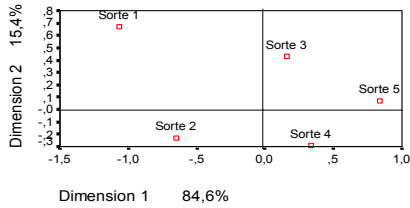
---

---

---

---

### Graphische Darstellung der Biersortenprofile im zweidimensionalen Raum



Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

### Geometrie der Spaltenprofile

- Spaltenprofile werden in einem fünfdimensionalen Raum positioniert, den die fünf Biersorten aufspannen
- Zentroid der Spaltenprofile= Summe der mit ihrer Masse gewichteten Zeilenprofile
- Die Koordinaten des Durchschnittsprofils der Merkmale entsprechen exakt den Massen der Zeilenprofile der fünf Biere

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

### Geometrie der Spaltenprofile

- Symmetrie äußert sich darin, dass der Zentroid der Spaltenprofile exakt den Massen der Zeilenprofile entsprechen und umgekehrt.
- Berechnung der Distanz zwischen einem Spaltenpunkt und dem Zentroiden:

$$d^2_{(j,k)} = \sum_i (c_{ij} - w_i)^2 / w_i$$

- Die Gesamtstreuung in den Daten und die Erklärungsbeiträge der Hauptachsen für die Analyse der Zeilen- und Spaltenprofile sind identisch.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

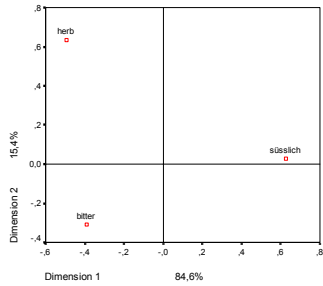
---

---

---

---

### Graphische Darstellung der Merkmalsprofile im zweidimensionalen Raum



Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

### Simultane graphische Darstellung der Zeilen- und Spaltenprofile

- Spalten- und Zeilenprofile können so übereinander gelegt werden, dass sich ihre Schwerpunkte und Hauptachsen decken
- Symmetrie hängt damit zusammen, dass die „singular values“ und damit auch deren Quadrate („principal inertias“) in beiden Analysen identisch sind.
- Symmetrie zwischen den Koordinaten der Zeilen- und Spaltenprofile ist für die Interpretation der Korrespondenz zwischen den beiden Punktwolken ist in der graphischen Darstellung von großer Bedeutung.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

### Deutung des Zusammenhangs:

Die räumliche Positionierung einer Biermarke liegt tendenziell immer dann in der Nähe des Merkmals, wenn die Alternative bezüglich dieses Merkmals eine hohe bedingte relative Häufigkeit aufweist.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

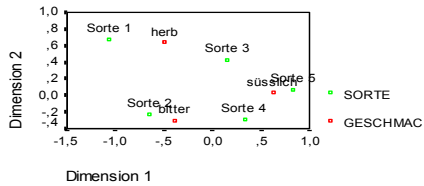
---

---

---

---

## Simultane Darstellung



Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

## Zeilenprofile

| SORTE   | GESCHMAC |        |         | Aktiver Rand |
|---------|----------|--------|---------|--------------|
|         | herb     | bitter | süßlich |              |
| Sorte 1 | ,357     | ,500   | ,143    | 1,000        |
| Sorte 2 | ,214     | ,548   | ,238    | 1,000        |
| Sorte 3 | ,218     | ,333   | ,448    | 1,000        |
| Sorte 4 | ,119     | ,396   | ,485    | 1,000        |
| Sorte 5 | ,115     | ,269   | ,615    | 1,000        |
| Masse   | ,183     | ,413   | ,404    |              |

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

---

---

## Interpretation der Darstellung

Ein Merkmal orientiert sich so entlang der Hauptachsen, dass es tendenziell in der Nähe solcher Alternativen liegt, die bezüglich dieses Merkmals im Vergleich zu anderen Konkurrenzprodukten hohe bedingte relative Häufigkeiten aufweisen.

**Es werden auch Beziehungen zwischen den beiden Wolken visualisiert !**

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

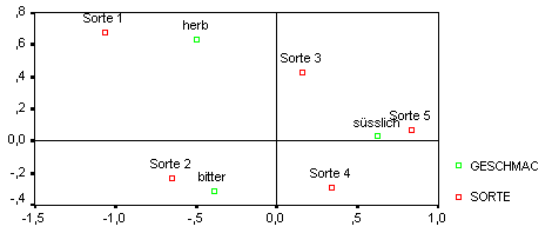
---

---

---

---

## Simultane Darstellung



Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## WICHTIG!!!!

Die räumliche Entfernung zwischen einem Punkt und einem Merkmal kann nicht direkt interpretiert werden.

Warum???

Aufgrund der mathematischen Vorgehensweise ist eine Distanz zwischen Punkten unterschiedlicher Wolken nicht definiert wurde.

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Erdmute Jahn

Interpretation der Ausgabedateien

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Ablauf in SPSS

- /Analysieren/Dimensionsreduktion/Korrespondenzanalyse
- Zeile → Sorte (Bereich definieren: Anzahl der Sorten)
- Spalte → Merkmale (Bereich definieren: Anzahl der Merkmale)
- Modelle → Distanzmaß:  $\chi^2$  → Normalisierungsmethode: symmetrisch
- Statistik → Korrespondenztabelle → Übersicht Zeilenpunkte und Spaltenpunkte → Zeilen- und Spaltenprofile
- Diagramme → Biplot → Zeilenpunkte und Spaltenpunkte

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---

## Vielen Dank für die Aufmerksamkeit

- Folien, Handout und Beispiele verfügbar auf <http://www2.hs-harz.de/~u16038/mafo/>

Hochschule Harz - Jana Bombik, Erdmute Jahn, Claudia Philipp, Helge Siems – 4. Juli 2002

---

---

---

---

---

---

---

---