

Lösung zu Kapitel 10: Beispiel 2

Bewerbungen

Die Daten aus Kendall (1975) beziehen sich auf 48 Bewerbungen um eine Position in einem Unternehmen. Diese Bewerbungen wurden anhand der 15 Variablen im Datenfile **bewerbung.sav** bewertet (je höher der Wert, desto stärker ist die Eigenschaft ausgeprägt).

- Gibt es Gruppen ähnlicher Bewerbungen?

Da der Datensatz ziemlich klein und übersichtlich ist, wurde in diesem Fall das hierarchische Clustern als Zuordnungsalgorithmus gewählt.

PASW

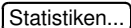
Analysieren

Klassifizieren

Hierarchische Cluster...

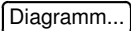
Variablen der Beobachtungen  Variable(n):

(hier alle)

unter 

kann Bereich für Clusteranzahl angegeben werden



unter 

kann Dendrogramm markiert werden



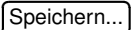
unter 

bei Messniveau gewünschtes Distanzmaß auswählen

bei Cluster-Methode Fusionsmethode auswählen

bei Werte transformieren Variablenstandardisierung wählen



unter 

Speicherung der Clusterzugehörigkeit(en) angeben





Hinweis: Als Messniveau wurde die Pearson-Korrelation (das Ähnlichkeitsmaß) gewählt und die Werte wurden in z-Werte transformiert um die Streuung der Variablen zu vereinheitlichen. Es wurde nach Variablen geclustert.

Auf die Darstellung der Clusterzuordnungsübersicht wird in diesem Fall aus Platzgründen verzichtet. Große Sprünge wären hier allerdings erst im Schritt 45 zu erkennen. Als Ähnlichkeitsmaß wurde die Pearsonkorrelation gewählt und deswegen sind die Koeffizienten auch absteigend geordnet.

Im Eiszapfenplot kann man sehen wann welche Personen zu einem Cluster verschmolzen wurden (► Abbildung 1).

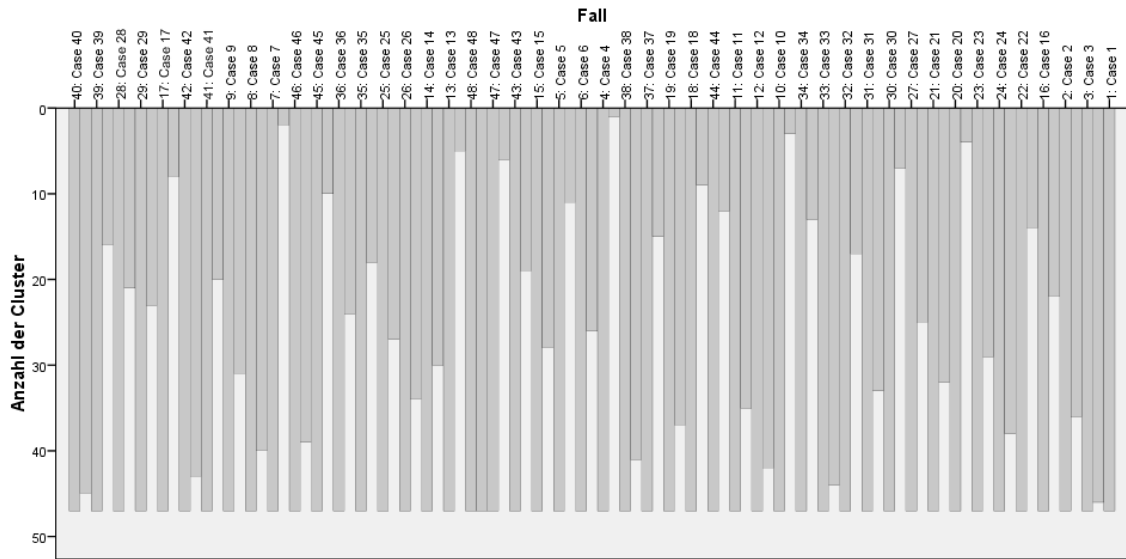


Abbildung 1: Eiszapfenplot des hierarchischen Clusters nach Personen.

Hinweis: Um das Dendrogramm optisch einigermaßen ansprechend zu formatieren wurde nach Doppelklick auf die Ausgabe (des Dendrogramms) der Text markiert und die Schriftart Courier New gewählt und es wurde noch zusätzlich die Schriftgröße verkleinert.

Das Dendrogramm zeigt ein sehr heterogenes Bild (► Abbildung 2). Wenn man nach den Personen clustert können keine Gruppen ähnlicher Bewerbungen identifiziert werden.

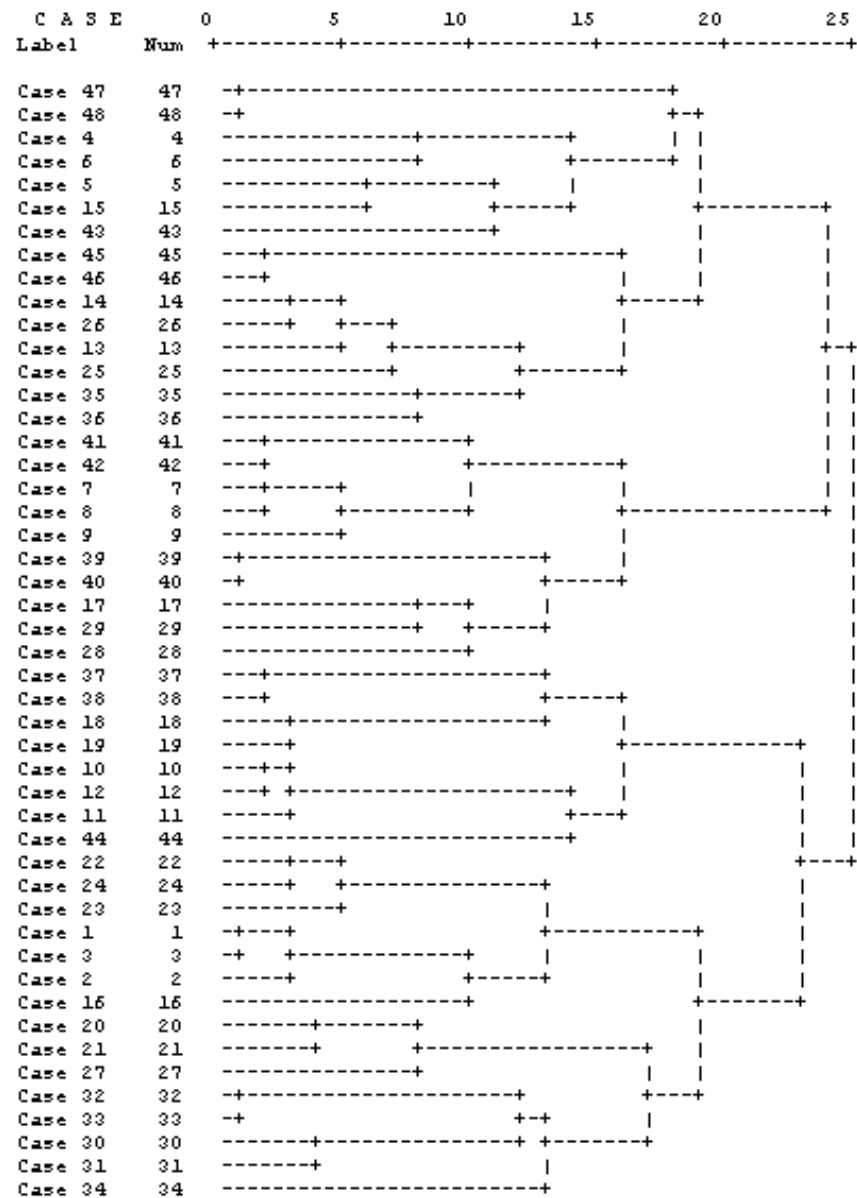


Abbildung 2: Dendrogramm des hierarchischen Clusters nach Personen.

Da das hierarchische Clustern nach Fällen kein besonders eindeutiges Ergebnis lieferte, wurde im Anschluss noch nach Variablen geclustert.

| Zuordnungsübersicht | | | | | | |
|---------------------|--------------------------|-----------|---------------|-------------------------------|-----------|------------------|
| Schritt | Zusammengeführte Cluster | | Koeffizienten | Erstes Vorkommen des Clusters | | Nächster Schritt |
| | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1 | 6 | 12 | ,883 | 0 | 0 | 3 |
| 2 | 8 | 11 | ,860 | 0 | 0 | 4 |
| 3 | 6 | 13 | ,827 | 1 | 0 | 6 |
| 4 | 5 | 8 | ,821 | 0 | 2 | 5 |
| 5 | 5 | 10 | ,766 | 4 | 0 | 6 |
| 6 | 5 | 6 | ,755 | 5 | 3 | 11 |
| 7 | 9 | 15 | ,693 | 0 | 0 | 9 |
| 8 | 4 | 14 | ,685 | 0 | 0 | 10 |
| 9 | 1 | 9 | ,567 | 0 | 7 | 13 |
| 10 | 4 | 7 | ,547 | 8 | 0 | 12 |
| 11 | 2 | 5 | ,456 | 0 | 6 | 12 |
| 12 | 2 | 4 | ,422 | 11 | 10 | 13 |
| 13 | 1 | 2 | ,281 | 9 | 12 | 14 |
| 14 | 1 | 3 | ,070 | 13 | 0 | 0 |

Abbildung 3: Zuordnungsübersicht bei Clustern von Variablen.

Die Koeffizienten der Zuordnungsübersicht haben eine absteigende Ordnung, da als Ähnlichkeitsmaß die Pearson-Korrelation gewählt wurde. Erst in Schritt 12 werden große Sprünge deutlich (► Abbildung 3).

Im Eispapfenplot kann man sehen wann welche Eigenschaften zu einem Cluster verschmolzen wurden (► Abbildung 4).

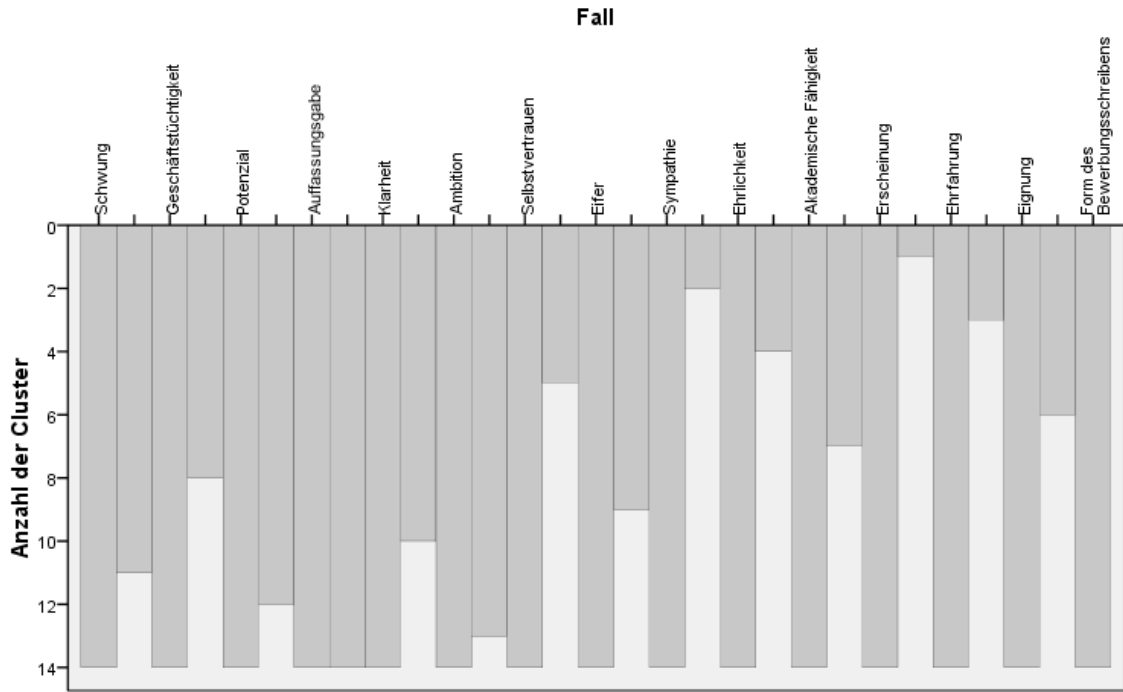


Abbildung 4: Eispapfenplot des hierarchischen Clusters nach Variablen (Eigenschaften).

*****HIERARCHICAL CLUSTER ANALYSIS*****

Dendrogram using Average Linkage (Between Groups)

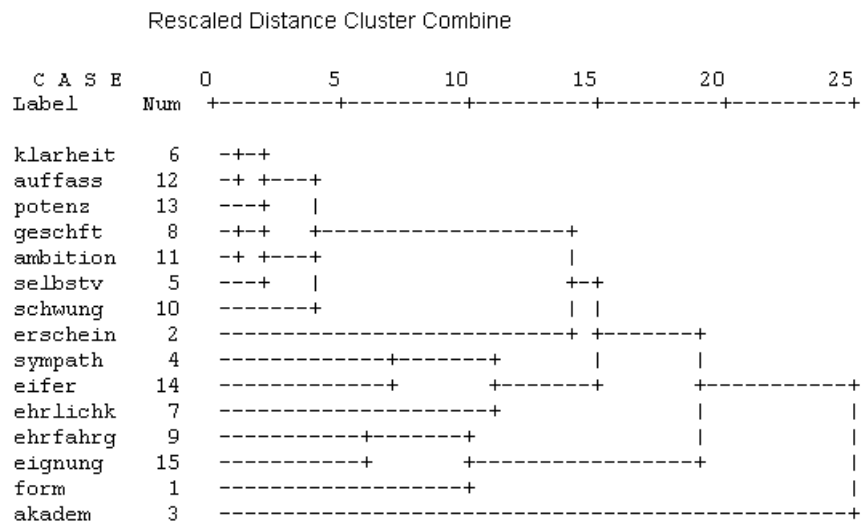


Abbildung 5: Dendrogramm des hierarchischen Clusters nach Variablen (Eigenschaften).

Hinweis: Um das Dendrogramm optisch einigermaßen ansprechend zu formatieren wurde nach Doppelklick auf die Ausgabe (des Dendrogramms) der Text markiert und die Schriftart Courier New gewählt.

Im Dendrogramm kann man die beiden Cluster gut erkennen. Die Variablen `klarheit` (Klarheit), `auffass` (Auffassungsgabe), `potenz` (Potenzial), `geschft` (Geschäftspotential), `ambition` (Ambition) und `selbstv` (Selbstvertrauen) bilden ein sehr homogenes Cluster.