

Clusteranalyse Winterreifen

Daniel Stepanovic

Ausgangssituation

Für eine Auswahl von 19 Winterreifen wurden die Ergebnisse eines ÖAMTC-Winterreifentest 2024 erfasst. Bewertet wurden zwei sicherheitsrelevante Kriterien:

- **Fahrsicherheit bei Trockenheit**
- **Fahrsicherheit bei Nässe**

Die Bewertungen erfolgten in Noten von **0.5 (sehr gut)** bis **5.5 (sehr schlecht)**. Zusätzlich wurde die **Reifendimension** (205 oder 215 mm) dokumentiert.

Ziel dieser Analyse ist es, mithilfe eines nicht-hierarchischen Clusteringverfahrens (**k-Means**) Gruppen von Reifen zu identifizieren, die sich anhand ihrer Sicherheitseigenschaften ähneln.

Obwohl es sich bei den beiden Bewertungskategorien streng genommen um **ordinale Variablen** handelt, werden sie in dieser Aufgabe, der Einfachheit halber, **als metrisch behandelt**.

Es sollen **zwei verschiedene Clusterlösungen** mit unterschiedlicher Gruppenzahl gebildet, beschrieben und miteinander verglichen werden.

Datenmanagement

Die Daten werden mit `read.table()` eingelesen.

```
daten = read.table("wi23b095.txt", sep = ";", header= TRUE)
str(daten)
```

```
'data.frame':  19 obs. of  4 variables:
 $ modell          : chr  "Vredestein Wintrac" "Pirelli Cinturato Winter 2" "Viking Wintrac"
 $ dimension       : int   205 215 215 215 215 215 215 205 205 205 215 ...
 $ fahrsicherheit_trocken: num  3.8 2.9 3.6 2 2.5 3.6 2 2.5 3.9 4.2 ...
 $ fahrsicherheit_nass  : num  2.6 2.1 3.1 2 2.5 2.8 1.9 5.5 2.5 3.5 ...
```

```
summary(daten)
```

modell	dimension	fahrsicherheit_trocken	fahrsicherheit_nass
Length:19	Min. :205.0	Min. :1.900	Min. :1.900
Class :character	1st Qu.:205.0	1st Qu.:2.500	1st Qu.:2.300
Mode :character	Median :205.0	Median :2.900	Median :2.500
	Mean :209.7	Mean :3.032	Mean :2.889
	3rd Qu.:215.0	3rd Qu.:3.600	3rd Qu.:3.150
	Max. :215.0	Max. :4.200	Max. :5.500

Standardisierung

Da die beiden Variablen unterschiedliche Mittelwerte und Streuungen aufweisen, werden sie zur Vergleichbarkeit standardisiert:

```
cluster_daten = scale(daten[, 3:4])
```

Clustering mit k-Means

Als Clustermethode verwenden wir **k-Means-Clustering**. Da wir die optimale Clusteranzahl nicht kennen, wird der Algorithmus wiederholt angewendet, wobei jeweils **25 verschiedene Startwerte** gewählt werden, um eine möglichst stabile Lösung zu erhalten.

Als Maß für die Qualität einer Lösung verwenden wir die **Fehlerquadratsumme (WSS)**, welche für jede der Clusterlösungen berechnet wird:

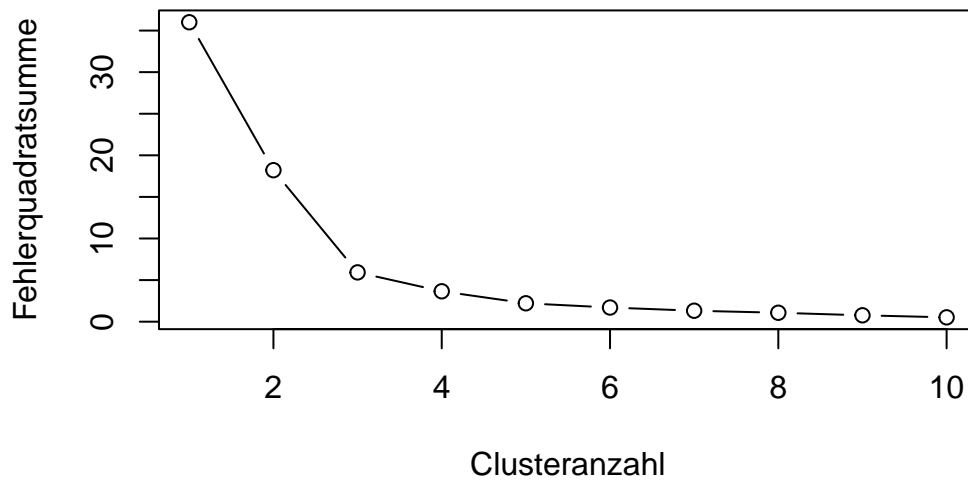
```
set.seed(42)
wss = c()
for (i in 1:10) {
  wss[i] = kmeans(cluster_daten, centers = i, nstart = 25)$tot.withinss
}
wss = data.frame(Clusteranzahl = 1:10, Fehlerquadratsumme = wss)
wss
```

	Clusteranzahl	Fehlerquadratsumme
1	1	36.0000000
2	2	18.2130553
3	3	5.9255573
4	4	3.6533760
5	5	2.2106517
6	6	1.7020621
7	7	1.3175876
8	8	1.0735628
9	9	0.7664732
10	10	0.5199887

Der folgende Elbow-Plot hilft uns bei der Bestimmung jener Lösung, die einen sinnvollen Kompromiss zwischen der Modellkomplexität und der Minimierung der Fehlerquadratsumme darstellt:

```
with(wss, plot(Clusteranzahl, Fehlerquadratsumme, type = "b",
main = "Elbow-Plot zur Bestimmung der Clusteranzahl"))
```

Elbow-Plot zur Bestimmung der Clusteranzahl



Im Elbow-Plot ist ein deutlicher Knick bei drei Clustern zu erkennen. Während sich die Fehlerquadratsumme von einem auf zwei und von zwei auf drei Cluster deutlich verringert, bringt eine Erhöhung auf vier oder mehr Cluster nur noch geringe zusätzliche Verbesserung. Daher werden im Folgenden zwei und drei Clusterlösungen analysiert und verglichen.

Clusteranalyse mit 2 Gruppen

Wir beginnen mit der Analyse einer Lösung mit **zwei Clustern**. Die Berechnung erfolgt mit dem k-Means-Algorithmus und 25 Startwerten:

```
set.seed(42)
km2 = kmeans(cluster_daten, centers = 2, nstart = 25)
daten$cluster2 = km2$cluster
table(km2$cluster)
```

```
1 2
10 9
```

Die 2-Cluster-Lösung teilt die 19 Winterreifen in zwei unterschiedlich große Gruppen auf:

- Cluster 1 enthält 10 Reifenmodelle
- Cluster 2 umfasst die übrigen 9 Reifenmodelle

Damit ergibt sich eine relativ ausgeglichene Verteilung der Beobachtungen auf die beiden Cluster, was für eine stabile und sinnvolle Gruppierung spricht.

Beschreibung der Gruppen (2 Cluster)

```
aggregate(. ~ cluster2, data = daten[, c("fahrsicherheit_trocken",  
"fahrsicherheit_nass", "cluster2")], mean)
```

	cluster2	fahrsicherheit_trocken	fahrsicherheit_nass
1	1	3.550000	3.440000
2	2	2.455556	2.277778

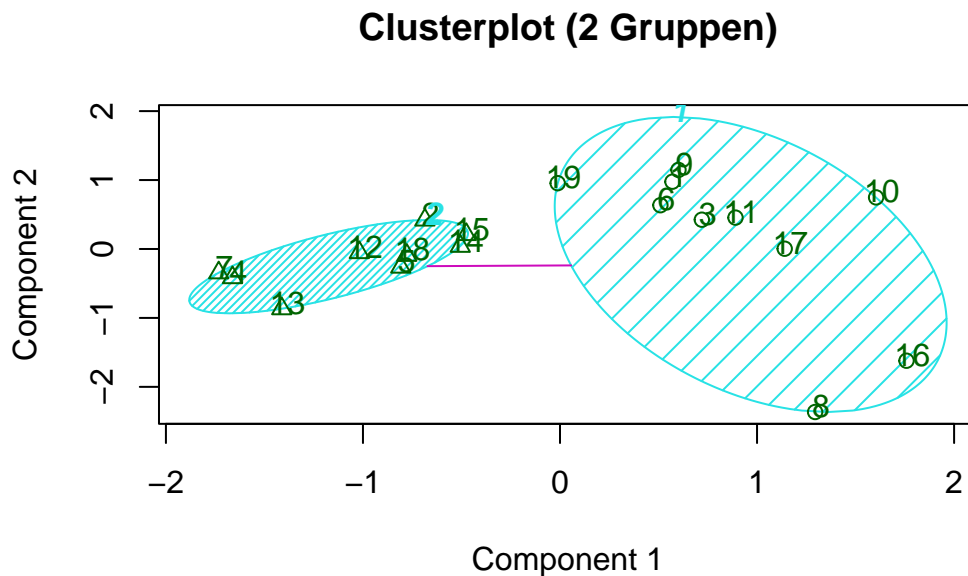
Die Clusterlösung mit zwei Gruppen zeigt eine klare Trennung zwischen Reifen mit besseren und schlechteren Sicherheitswerten:

- Cluster 1 (Durchschnittswerte: trocken = 3.55, nass = 3.44) umfasst Reifen mit insgesamt schlechterer Fahrsicherheit, sowohl auf trockener als auch auf nasser Fahrbahn.

-Cluster 2 (Durchschnittswerte: trocken = 2.46, nass = 2.28) enthält Reifen mit deutlich besseren Bewertungen in beiden Kategorien.

Visualisierung

```
library(cluster)
clusplot(cluster_daten, km2$cluster, labels = 2,
shade = TRUE, main = "Clusterplot (2 Gruppen)")
```



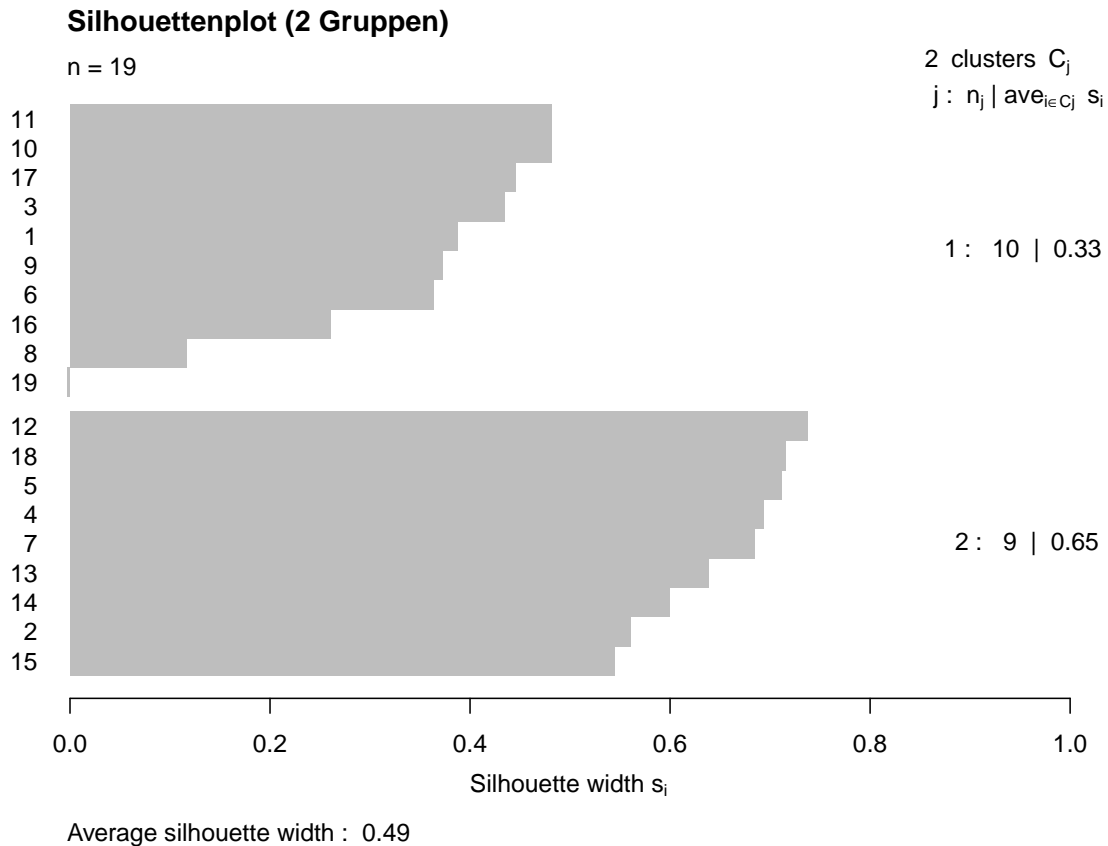
These two components explain 100 % of the point variability.

Der Clusterplot zeigt eine klare räumliche Trennung der beiden Gruppen. Die beiden Cluster liegen deutlich auseinander, was auf eine gute Trennschärfe hinweist:

- Cluster 1 (links) ist kompakt und homogen, die Beobachtungen liegen eng beieinander. Dies spricht für eine Gruppe von Reifen mit ähnlichen Sicherheitseigenschaften.
- Cluster 2 (rechts) ist größer und etwas verstreuter, aber dennoch klar abgegrenzt. Diese Gruppe enthält Reifen mit unterschiedlicherer, aber tendenziell besserer Bewertung.

Silhouettenanalyse

```
sil2 = silhouette(km2$cluster, dist(cluster_daten))  
plot(sil2, main = "Silhouettenplot (2 Gruppen)")
```



- Die durchschnittliche Silhouettenbreite beträgt 0.49. Dieser Wert liegt an der Grenze zwischen schwacher und akzeptabler Struktur
- Cluster 2 ist mit einem Durchschnittswert von 0.65 klarer abgegrenzt und weist eine hohe Homogenität auf.
- Cluster 1 ist mit einer Silhouettenbreite von 0.33 weniger trennscharf. Einige Beobachtungen (z.B.: Nr. 19 oder 16) haben relativ niedrige Werte, was auf eine schwächere Zuordnung hindeutet.

Positiv ist, dass keine Silhouettenbreite negativ ist.

Auffällig ist Beobachtung Nr. 19, deren Silhouettenwert fast 0 beträgt und daher zwischen den Clustern liegt.

Insgesamt kann die 2-Gruppen-Lösung als strukturell brauchbar beurteilt werden, wobei eine feinere Lösung (z.B.: mit 3 Clustern) eventuell bessere Gruppierung erlaubt.

Clusteranalyse mit 3 Gruppen

```
set.seed(42)
km3 = kmeans(cluster_daten, centers = 3, nstart = 25)
daten$cluster3 = km3$cluster
table(km3$cluster)
```

```
1 2 3
2 9 8
```

Die 3-Cluster-Lösung teilt die 19 Winterreifenmodelle wie folgt auf:

- Cluster 1: 2 Beobachtungen
- Cluster 2: 9 Beobachtungen
- Cluster 3: 8 Beobachtungen

Damit ergibt sich eine leicht ungleichmäßige Verteilung: Cluster 2 ist am größten, während Cluster 1 nur zwei Beobachtungen umfasst. Der sehr kleine Cluster 1 könnte auf eine extrem abweichende Gruppe hinweisen z.B.: Reifen mit besonders schlechten oder besonders guten Sicherheitswerten.

Beschreibung der Gruppen (3 Cluster)

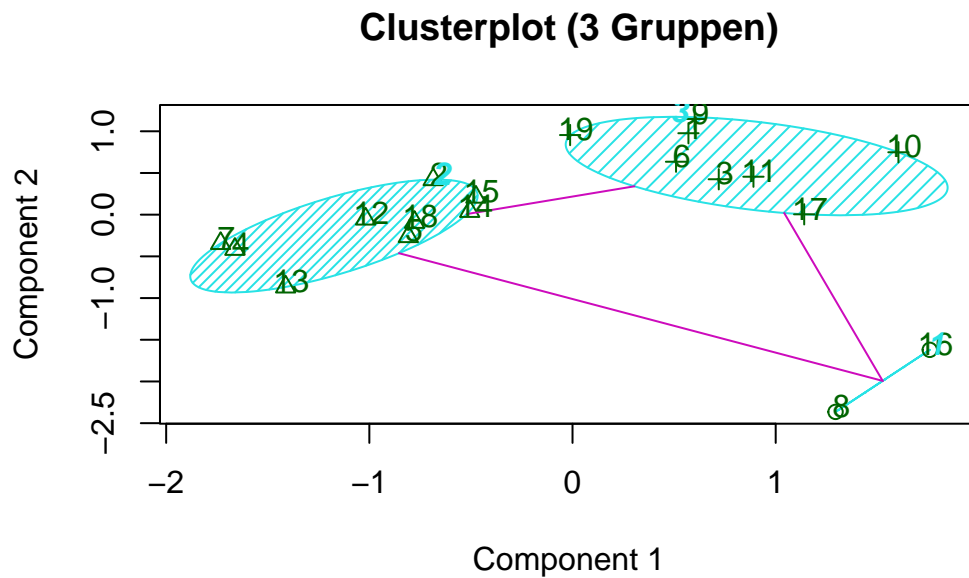
```
aggregate(. ~ cluster3, data = daten[, c("fahrsicherheit_trocken",
"fahrsicherheit_nass", "cluster3")], mean)
```

	cluster3	fahrsicherheit_trocken	fahrsicherheit_nass
1	1	2.800000	5.400000
2	2	2.455556	2.277778
3	3	3.737500	2.950000

- Cluster 1 für Reifen mit gravierenden Schwächen bei Nässe
- Cluster 2 steht für gute Allrounder
- Cluster 3 für Modelle mit klarer Schwäche bei Trockenheit

Visualisierung

```
clusplot(cluster_daten, km3$cluster, labels = 2,  
         shade = TRUE, main = "Clusterplot (3 Gruppen)")
```



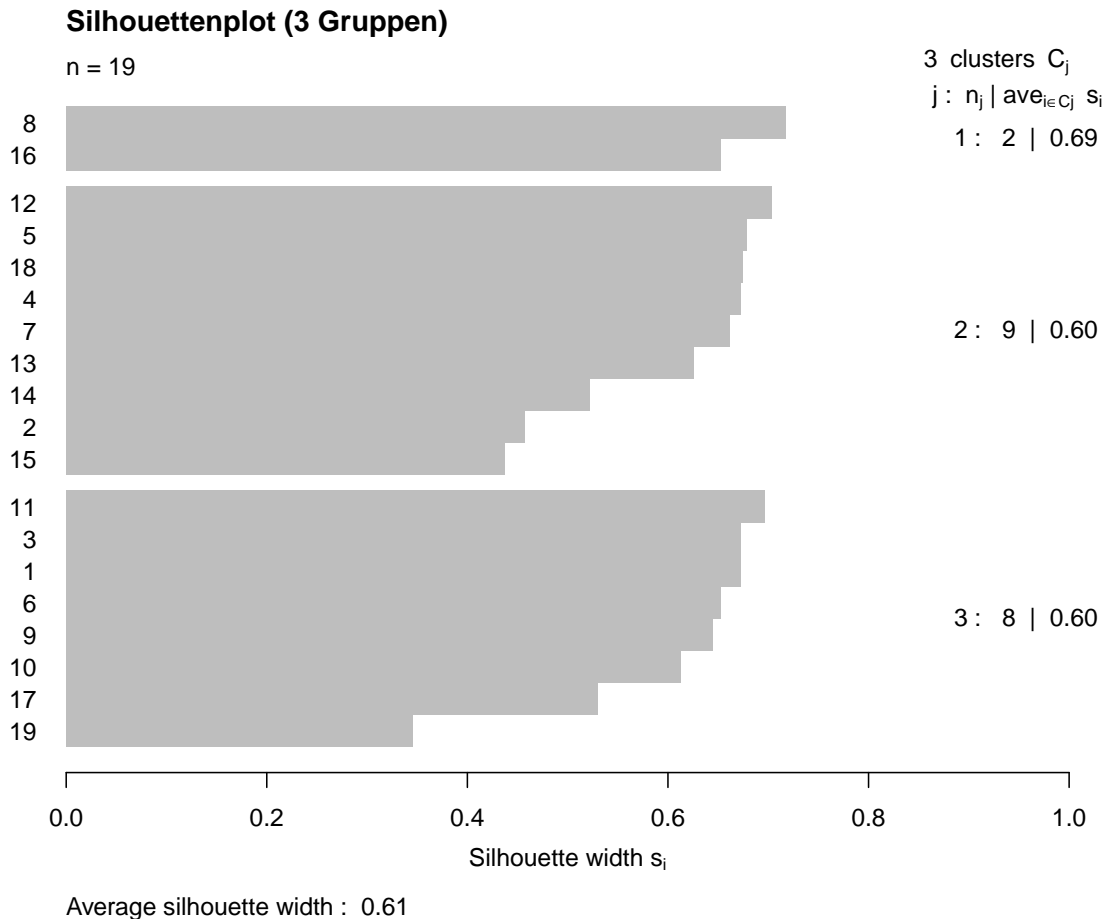
These two components explain 100 % of the point variability.

Der Plot zeigt eine klare Trennung der drei Cluster ohne Überlappung.

- Cluster 1 (unten rechts) enthält wenige, aber deutlich abweichende Reifen mit schlechter Nässebewertung
- Cluster 2 (oben rechts) umfasst sicherere Modelle mit guten Bewertungen
- Cluster 3 (links) sammelt Reifen mit Schwächen auf trockenerer Fahrbahn.

Silhouettenanalyse

```
sil3 = silhouette(km3$cluster, dist(cluster_daten))  
plot(sil3, main = "Silhouettenplot (3 Gruppen)")
```



Die durchschnittliche Silhouettenbreite beträgt 0.61 und weist auf eine gute Clusterstruktur hin. Alle drei Cluster sind ähnlich gut strukturiert (zwischen 0.60–0.69), keine negativen Werte treten auf. Besonders Cluster 1 zeigt mit 0.69 eine sehr klare Abgrenzung, aber auch Cluster 2 und 3 sind stabil. Die 3er-Lösung liefert somit eine homogene und gut trennbare Gruppierung.

Vergleich der Clusterlösungen

Silhouettenkoeffizienten

```
cat("Durchschnittliche Silhouettenbreite (2 Gruppen):", mean(sil2[, 3]), "\n")
```

Durchschnittliche Silhouettenbreite (2 Gruppen): 0.485526

```
cat("Durchschnittliche Silhouettenbreite (3 Gruppen):", mean(sil3[, 3]), "\n")
```

Durchschnittliche Silhouettenbreite (3 Gruppen): 0.612128

- Die 2-Gruppen-Lösung zeigt eine stärkere Trennung mit höherem Silhouettenwert.
- Die 3-Gruppen-Lösung bietet eine differenziertere Betrachtung, z.B. ein zusätzlicher Cluster mit guter Trocken-, aber schwacher Nassleistung.

Clusterinhalt (qualitativ)

- In der 2er-Lösung ergibt sich eine klare Trennung in sicherere vs. weniger sichere Reifen.
- In der 3er-Lösung wird zusätzlich ein differenziertes Profil sichtbar, etwa Reifen mit durchschnittlicher Trockensicherheit, aber starker Nässe-Schwäche.

Fazit

Die Winterreifen lassen sich anhand der beiden sicherheitsrelevanten Merkmale sinnvoll clustern. Die 2-Cluster-Lösung zeigt eine klarere Trennung, während die 3-Cluster-Lösung eine differenziertere Interpretation ermöglicht.

Die Wahl der besseren Lösung hängt vom Verwendungszweck ab:

- Für einfache Kundenempfehlungen könnte die 2-Gruppen-Lösung genügen.
- Für eine feinere Marktanalyse bietet die 3-Gruppen-Lösung mehr Informationsgehalt.

Beide Lösungen sind plausibel und aussagekräftig für die Bewertung der Reifenleistung.