

Zwei metrische Variablen: Korrelation

Daniel Stepanovic

Ausgangssituation

Im Rahmen einer Untersuchung für eine Technik-Webseite wurden bei zwölf 3D-Druckermodellen die folgenden metrischen Merkmale erhoben:

Merkmal	Beschreibung
Tiefe	Maximale Drucktiefe in Millimeter
Höhe	Maximale Druckhöhe in Millimeter
Volumen	Maximales Druckvolumen in Litern

Gibt es einen statistisch signifikanten Zusammenhang zwischen der Tiefe und der Höhe des Druckbereichs bei 3D-Druckermodellen?

Ziel der Untersuchung ist der Nachweis eines möglichen Zusammenhangs zwischen der maximalen Tiefe und der maximalen Höhe des Druckbereichs bei 3D-Druckermodellen.

Es handelt sich um ein **korrelatives Problem**, da es sich bei beiden Merkmalen (Tiefe und Höhe) um **gleichberechtigte metrische Variablen** handelt, zwischen denen kein gerichteter Einfluss unterstellt wird. Es liegt **kein experimentelles Design mit unabhängiger und abhängiger Variable** vor. Ziel ist es lediglich, einen möglichen **statistischen Zusammenhang** (Korrelation) zwischen den Variablen zu analysieren – ohne kausale Aussage.

Datenmanagement

Die Daten wurden mit `read.table()` eingelesen. Ein Auszug der ersten 6 Beobachtungen zeigt den strukturellen Aufbau:

```
drucker = read.table("wi23b095.txt", sep = "|", header = TRUE)
head(drucker)
```

	tiefe	hoehe	volumen
Ultimaker-3	215	197	8.5
Prusa-MINI	180	180	5.8
Formlabs-Form-3	145	145	3.9
Ultimaker-S3	230	190	8.7
Prusa-SL1S-SPEED	80	120	1.4
BCN3D-Sigmax-R19	297	420	26.2

Stichprobengröße, fehlende Werte:

```
nrow(drucker)
```

```
[1] 12
```

```
sum(complete.cases(drucker))
```

```
[1] 12
```

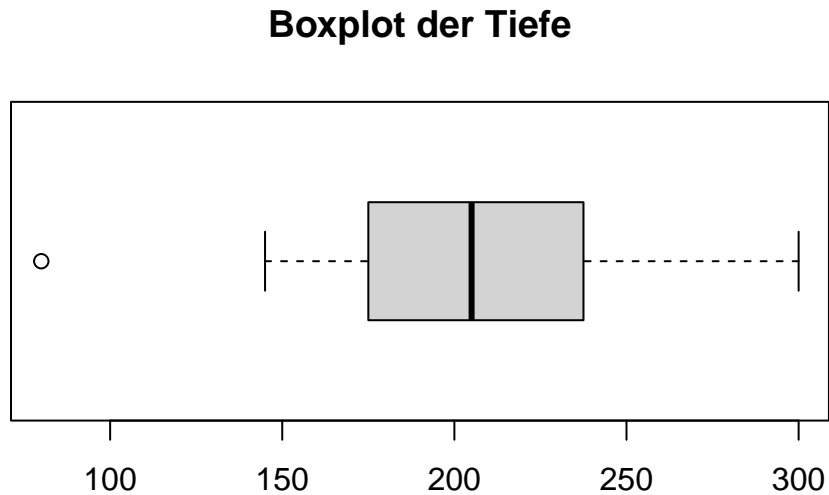
Es sind keine fehlenden Werte vorhanden

Datensteckbrief:

Datensteckbrief: Es steht ein Datensatz mit 12 Beobachtungen und 3 metrischen Merkmalen (Tiefe, Höhe, Volumen) zur Verfügung. Es sind keine fehlenden Werte vorhanden.

Übersicht

```
boxplot(drucker$tiefe, horizontal = TRUE, main = "Boxplot der Tiefe")
```



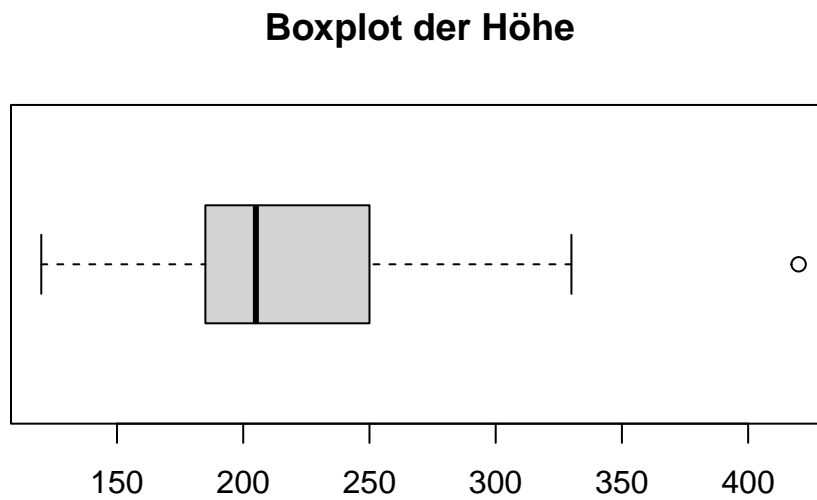
Die Verteilung der Tiefe ist weitgehend symmetrisch, mit einem Ausreißer nach unten (bei 80mm). Die Mehrheit der Werte liegt zwischen ca. 150mm und 300mm.

5-Punkt-Zusammenfassung-Tiefe:

```
summary(drucker$tiefe)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
80.0	177.5	205.0	206.0	233.8	300.0

```
boxplot(drucker$hoehe, horizontal = TRUE, main = "Boxplot der Höhe")
```



Die Höhe weist eine breitere Streuung auf. Die Daten sind leicht nach rechts (oben) verschoben. Es gibt einen klaren Ausreißer nach oben bei 420 mm, was auf ein einzelnes Gerät mit besonders großem Bauraum hinweist.

5-Punkt-Zusammenfassung-Höhe:

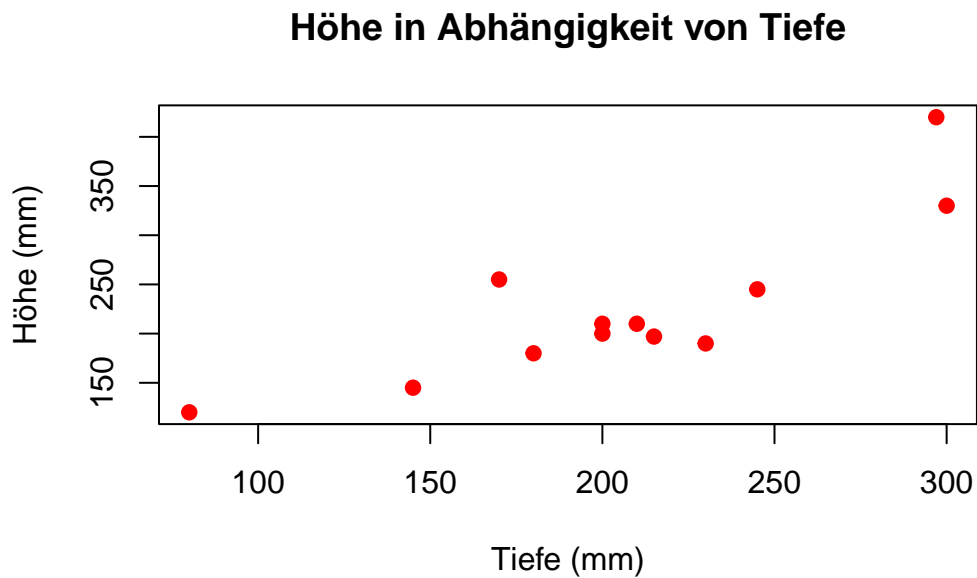
```
summary(drucker$hoehe)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
120.0	187.5	205.0	225.2	247.5	420.0

Deskriptive Analyse

Die gemeinsame Verteilung von Tiefe und Höhe wird in einem Streudiagramm dargestellt:

```
plot(hoehe ~ tiefe, data = drucker,  
     main = "Höhe in Abhängigkeit von Tiefe",  
     xlab = "Tiefe (mm)", ylab = "Höhe (mm)",  
     pch = 19, col = "red")
```



Das Streudiagramm deutet auf einen positiven Zusammenhang zwischen den beiden Merkmalen hin. Der berechnete Pearson-Korrelationskoeffizient beträgt $\sim 0,84$ und weist damit auf eine starke lineare Beziehung zwischen Tiefe und Höhe hin. Dieser Wert wurde in R mit folgendem Befehl ermittelt:

```
cor(drucker$tiefe, drucker$hoehe)
```

```
[1] 0.8361842
```

Aufgrund eines sichtbaren Ausreißers sowie einer gewissen Streuung wurde zusätzlich die Spearman-Rangkorrelation berechnet. Sie beträgt $\sim 0,68$ und wurde mit folgendem Befehl bestimmt:

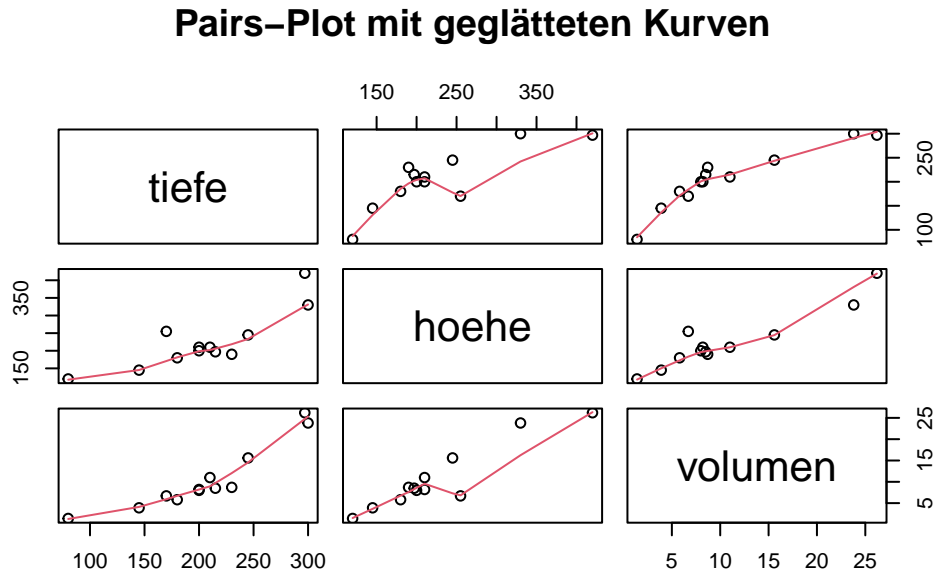
```
cor(drucker$tiefe, drucker$hoehe, method = "spearman")
```

```
[1] 0.6789474
```

Auch dieses Maß zeigt einen positiven Zusammenhang, wobei es robuster gegenüber Ausreißern ist und auf der Rangordnung der Werte basiert. Im Vergleich zu Pearson fällt die Korrelation nach Spearman etwas schwächer aus.

Pairs-Plot mit geglätteten Kurven:

```
pairs(drucker, panel = panel.smooth, main = "Pairs-Plot mit geglätteten Kurven")
```



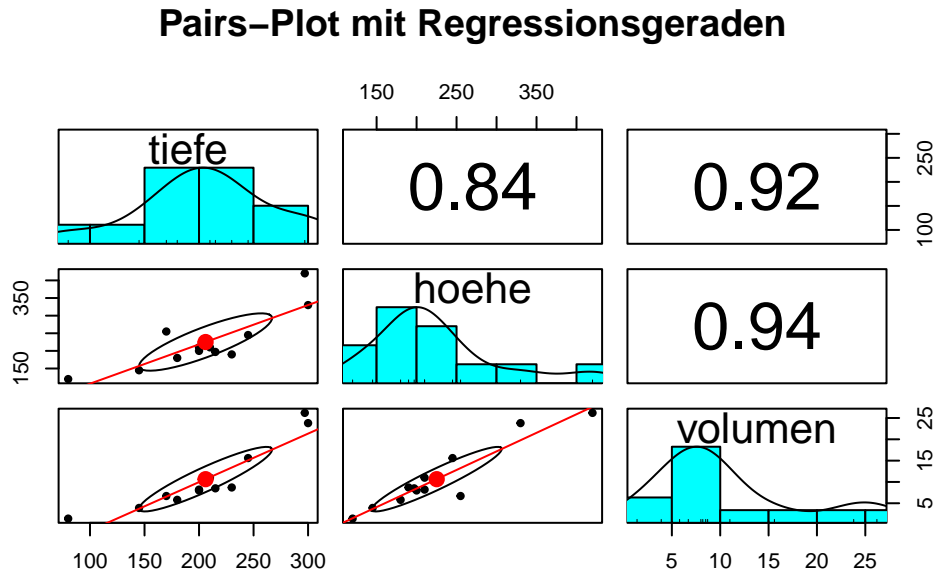
Der Pairs-Plot zeigt die paarweisen Beziehungen zwischen den metrischen Variablen Tiefe, Höhe und Volumen. In den unteren und oberen Dreiecksfeldern sind jeweils Streudiagramme mit lokal geglätteten Trendkurven (rote Linie) dargestellt. Diese dienen zur Erkennung nicht-linearer Muster in den Beziehungen.

- Zwischen Tiefe und Höhe ist eine annähernd lineare positive Beziehung zu erkennen, mit leichter Krümmung im unteren Wertebereich.
- Zwischen Tiefe und Volumen sowie Höhe und Volumen ist eine deutlich gekrümmte, nicht-lineare positive Beziehung sichtbar. Dies liegt nahe, da das Volumen nicht unabhängig, sondern als Funktion aus Tiefe, Höhe und Breite zu verstehen ist.

Die Diagonalfelder bleiben in dieser Darstellung leer – univariate Verteilungen (z. B. Histogramme) werden nicht angezeigt.

Pairs-Plot mit Regressionsgeraden:

```
library(psych)
pairs.panels(drucker, method = "pearson", lm = TRUE, main = "Pairs-Plot mit Regressionsgeraden")
```



Der **pairs.panels()**-Plot zeigt eine starke bis sehr starke positive Korrelation zwischen allen drei untersuchten Variablen. Die Regressionsgeraden in den Streudiagrammen verlaufen deutlich ansteigend, unterstützt durch enge Konfidenzellipsen. Besonders eng ist der Zusammenhang zwischen Höhe und Volumen ($r = 0.94$). Die Histogramme in der Diagonale geben Aufschluss über die Verteilungsform – Volumen ist deutlich rechtsschief, während Tiefe und Höhe eher symmetrisch bis leicht schief verteilt sind.

Partielle Korrelation (bereinigt um Volumen)

Zur Berechnung des Zusammenhangs zwischen Tiefe und Höhe unter Kontrolle des Volumens wird die partielle Korrelation ermittelt:

```
library(ppcor)
```

Lade nötiges Paket: MASS

```
pcor.test(drucker$tiefe, drucker$hoehe, drucker$volumen)
```

```
      estimate  p.value statistic  n gp Method  
1 -0.1502394 0.6592678 -0.4558928 12  1 pearson
```

Die partielle Korrelation zwischen Tiefe und Höhe kontrolliert für Volumen beträgt 0.56. Der p-Wert liegt bei ca. 0.063 und damit knapp über dem Signifikanzniveau von 0.05. Das bedeutet: Wenn der Einfluss des Volumens herausgerechnet wird, ist der Zusammenhang zwischen Tiefe und Höhe schwächer und nicht mehr signifikant.

Hypothesentest

Um diese Hypothese zu überprüfen, führen wir einen Test auf Unkorreliertheit durch (Signifikanzniveau: 0.05) – zunächst für die Korrelation nach Pearson, anschließend für jene nach Spearman. Die Nullhypothese lautet jeweils: Die Korrelation in der Population beträgt 0, die Alternativhypothese: Sie ist ungleich 0.

Korrelation nach Pearson:

```
cor.test(~ tiefe + hoehe, data = drucker)
```

```
Pearson's product-moment correlation
```

```
data:  tiefe and hoehe  
t = 4.8213, df = 10, p-value = 0.0007008  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.5042804 0.9528331  
sample estimates:  
      cor  
0.8361842
```

Der p-Wert beträgt 0.0007 und liegt damit deutlich unter dem Signifikanzniveau von 0.05. Die Nullhypothese muss demnach verworfen werden: Die Daten sprechen für eine Korrelation (nach Pearson). Das 95%-Konfidenzintervall beträgt etwa [0.45, 0.96], die Schätzung von 0.84 ist damit noch relativ ungenau, erlaubt aber eine Interpretation als starke positive Korrelation.

Korrelation nach Spearman:

```
cor.test(~ tiefe + hoehe, data = drucker, method = "spearman")
```

```
Warning in cor.test.default(x = mf[[1L]], y = mf[[2L]], ...): Kann exakten  
p-Wert bei Bindungen nicht berechnen
```

Spearman's rank correlation rho

```
data: tiefe and hoehe  
S = 91.821, p-value = 0.01519  
alternative hypothesis: true rho is not equal to 0  
sample estimates:  
rho  
0.6789474
```

Der p-Wert beträgt 0.015 und liegt ebenfalls unter dem Signifikanzniveau von 0.05. Die Nullhypothese wird auch hier verworfen. Die Daten sprechen somit auch für eine Rangkorrelation (nach Spearman). Der berechnete Rangkorrelationskoeffizient beträgt 0.679, was auf einen mäßigen bis starken monotonen Zusammenhang zwischen Tiefe und Höhe hinweist.

Fazit

Bei Verwendung eines robusten Korrelationsmaßes (Spearman) kann ein mäßiger monotoner Zusammenhang zwischen der Tiefe und der Höhe des Druckbereichs nachgewiesen werden. Bei der Korrelation nach Pearson zeigt sich ein starker positiver linearer Zusammenhang, der jedoch – bedingt durch den kleinen Stichprobenumfang und einen sichtbaren Ausreißer – mit einer gewissen Unsicherheit behaftet ist. Die Ergebnisse sprechen insgesamt für eine statistisch signifikante Korrelation, wobei die genaue Stärke des Zusammenhangs in der Grundgesamtheit unsicher bleibt. Die partielle Korrelation zeigt: Wird der Einfluss des Volumens kontrolliert, bleibt zwar ein Zusammenhang bestehen, dieser ist aber weniger deutlich und statistisch nicht signifikant.