

Statistik 3

Übung 2

Einleitung und Ziele der Übung

Herzlich Willkommen zur zweiten Übung Statistik 3. Diese Übung behandelt die Partial- und Semipartialkorrelation.

Agenda

- Installieren des Packages ppcor
- Bestimmen von Partialkorrelationen 1. Ordnung
- Bestimmen von Partialkorrelationen höherer Ordnung
- Bestimmen von Semipartialkorrelationen
- (Semi-)Partialkorrelationen basierend auf der Regression
- (Semi-)Partialkorrelationen mit ppcor

1. Installieren des Packages *ppcor*

Zur Erinnerung: R bietet viele Funktionen, die über die Grundlagen hinausgehen, als Package an. Packages müssen einmalig installiert aber bei jedem Start von R neu geladen werden. Man sollte Packages auch nur laden, wenn sie wirklich benötigt werden. Das Package *ppcor* enthält Funktionen, die sehr praktisch sind, wenn viele (Semi-)Partialkorrelationen bestimmt werden sollen. Dafür kann es manche Spezialfälle nicht so gut. Auch gibt es keine eingebaute Funktionen zur Prüfung von Annahmen/Voraussetzungen.

Installieren

- Bei Frage nach Neustart mit nein antworten antworten. Falls nötig nach der Installation selber neustarten.
- Solange ein “Stop” Symbol unten rechts an der Konsole zu sehen ist, arbeitet R noch und wird weitere Befehle erst danach ausführen.

```
1 install.packages("ppcor") # Hier müssen "" genutzt werden
```

Package laden

- Generell möglichst alle notwendigen Packages an einem Ort am Anfang eines Dokuments

```
1 library("ppcor") # Auch ohne "" möglich, da ppcor nach dem Installieren bekannt ist.
```

2. Bestimmen von Partialkorrelationen 1. Ordnung - Einführung 1

In der Vorlesung wurde diese Formel zur Bestimmung der Partialkorrelation 1. Ordnung eingeführt:

$$r_{AB \bullet C} = \frac{r_{AB} - r_{AC} * r_{BC}}{\sqrt{1 - r_{AC}^2} * \sqrt{1 - r_{BC}^2}}$$

Um die Partialkorrelation 1. Ordnung zu bestimmen, werden die Korrelationen benötigt.

2. Bestimmen von Partialkorrelationen 1. Ordnung - Übung 1

In einer Schule wurden die Fähigkeiten von Schüler*innen in drei Fächern erfasst. Es soll untersucht werden, wie stark der Zusammenhang zwischen den Fähigkeiten in zwei Fächern ist, wenn jeweils für das dritte Fach kontrolliert wird (dessen Einfluss herauspartialisiert wird).

	Mathematik	Deutsch
Deutsch	.50	
Physik	.65	.35

Welche drei (verschiedenen) Partialkorrelationen könnten bestimmt werden? Geben Sie die mathematische Bezeichnung in die unteren drei Felder ein. Schreiben Sie einfach alles hintereinander und nutzen Sie die ersten Buchstaben und den normalen Punkt.

z.B. $r_{AB.C}$ (Hinweis: wäre dasselbe wie $r_{BA.C}$. Die Überprüfung erkennt nicht, wenn Sie Lösungen doppelt eingeben)

2. Bestimmen von Partialkorrelationen 1. Ordnung - Übung 2

Formulieren Sie für eine der drei Partialkorrelationen jeweils die statistischen und inhaltlichen Null- und Alternativhypothesen.

Berechnen Sie anschliessend alle drei Partialkorrelationen erster Ordnung!

	Mathematik	Deutsch
Deutsch	.50	
Physik	.65	.35

- $r_{DM.P} =$
- $r_{DP.M} =$
- $r_{MP.D} =$

2. Bestimmen von Partialkorrelationen 1. Ordnung - Einführung 2

Die Partialkorrelationen sind niedriger ausgefallen, als die ursprünglichen Korrelationen. Das heisst, das mindestens ein Teil des Zusammenhangs zwischen den Leistungen in je zwei der Fächer durch die Leistung im dritten Fach erklärt werden kann. Wie bei Korrelationen auch, ist dies allein aber kein Beweis für Kausalität. Dafür bräuchte es Experimentaldaten.

Eine Frage, die hingegen beantwortet werden kann: Sind die Partialkorrelationen noch statistisch signifikant?

Dafür braucht man

- das Signifikanzniveau
- die Stichprobengrösse
- um die Freiheitsgrade zu bestimmen
- dann den empirischen F - oder t -Wert ($t_{emp} = \sqrt{F_{emp}}$) bestimmen
- und dann entweder den R-Befehl `pf()` oder `pt()` für den p -Wert verwenden

2. Bestimmen von Partialkorrelationen 1. Ordnung - Übung 3

Wir gehen von $N = 20$ und $\alpha = .05$ aus.

Für die Freiheitsgrade des t -Werts gilt: $df = N - 2 - k$, wobei k die Ordnung der Partialkorrelation ist.

Beim F -Wert gilt: $df_1 = 1$ und $df_2 = N - 2 - k$

Weiterhin gilt:

$$F_{emp} = \frac{r_{AB \bullet C}^2}{1 - r_{AB \bullet C}^2} * df_2$$

Bestimmen Sie für $r_{DM.P}$ den empirischen F -Wert!

$F_{emp} =$

2. Bestimmen von Partialkorrelationen 1. Ordnung - Übung 4

Nun bleibt noch den zugehörige p -Wert mittels R zu bestimmen.

$$\varrho_{AB \bullet C} \neq 0$$

```
1 pf(F-Wert,df1,df2, lower.tail=FALSE) #oder  
2 2*pt(t-Wert,df, lower.tail=FALSE)
```

$$\varrho_{AB \bullet C} > 0 \quad \text{und} \quad r_{AB \bullet C} > 0 \quad \text{oder} \quad \varrho_{AB \bullet C} < 0 \quad \text{und} \quad r_{AB \bullet C} < 0$$

```
1 0.5 * pf(F-Wert,df1,df2, lower.tail=FALSE) #oder  
2 pt(t-Wert,df, lower.tail=FALSE)
```

$$\varrho_{AB \bullet C} > 0 \quad \text{und} \quad r_{AB \bullet C} < 0 \quad \text{oder} \quad \varrho_{AB \bullet C} < 0 \quad \text{und} \quad r_{AB \bullet C} > 0$$

```
1 1-(0.5 * pf(F-Wert,df1,df2, lower.tail=FALSE)) #oder  
2 pt(t-Wert,df, lower.tail=TRUE)
```

Bestimmen Sie p für $r_{DM.P}!$ $p =$

Damit ist das Ergebnis für $\alpha = .05$.

Bei genügend Teststärke sollte die Alternativhypothese werden.

3. Bestimmen von Partialkorrelationen höherer Ordnung - Einführung

Die Berechnungen für Partialkorrelationen höherer Ordnung sind nahezu identisch zu den Partialkorrelationen 1. Ordnung.

Allerdings benötigt man nicht mehr die Korrelationen als Grundlage, sondern die Partialkorrelationen der Ordnung k-1. Für eine Partialkorrelation der 4. Ordnung wären also die Partialkorrelationen 3. Ordnung notwendig.

$$r_{AB \bullet C_1, \dots, C_k} = \frac{r_{AB \bullet C_1, \dots, C_{k-1}} - r_{AC_k \bullet C_1, \dots, C_{k-1}} * r_{BC_k \bullet C_1, \dots, C_{k-1}}}{\sqrt{1 - r_{AC_k \bullet C_1, \dots, C_{k-1}}^2} * \sqrt{1 - r_{BC_k \bullet C_1, \dots, C_{k-1}}^2}}$$

Da der Rest identisch ist, bestimmen wir nur schnell den Wert für eine Partialkorrelation 2. Ordnung.

3. Bestimmen von Partialkorrelationen höherer Ordnung - Übung

Berechnen Sie $r_{PI.DM}$ Die untenstehende Tabelle enthält die Partialkorrelationen 1. Ordnung, bei denen die Deutschfähigkeiten bereits herausgerechnet wurden.

	Mathematik	Informatik
Informatik	.76	
Physik	.59	.48

• $r_{PI.DM} =$

4. Bestimmen von Semipartialkorrelationen - Einführung

Auch für Semipartialkorrelationen ändert sich nicht viel:

- eine Wurzel entfällt,
- es werden die Korrelationen/Semipartialkorrelationen in den Formeln verwendet.

Hier die Formel, wenn Variablen (ausschliesslich) aus B herauspartialisiert werden. Beachten Sie vor allem die Klammern. Diese zeigen an, aus welcher Variable herauspartialisiert wird.

$$r_{A(B \bullet C_1, \dots, C_k)} = \frac{r_{A(B \bullet C_1, \dots, C_{k-1})} - r_{A(C_k \bullet C_1, \dots, C_{k-1})} * r_{B(C_k \bullet C_1, \dots, C_{k-1})}}{\sqrt{1 - r_{B(C_k \bullet C_1, \dots, C_{k-1})}^2}}$$

Beachten Sie, dass (fast immer) gilt:

$$r_{A(B \bullet C_1, \dots, C_k)} \neq r_{B(A \bullet C_1, \dots, C_k)}$$

Es macht einen Unterschied, ob wir aus A oder aus B herauspartialisieren.

Da die Rechnung damit eher einfacher als schwerer geworden ist, sparen wir uns ein Rechenbeispiel und wechseln direkt zum nächsten Thema.

5. (Semi-)Partialkorrelationen basierend auf der Regression - Einführung

In der Vorlesung wurde besprochen, dass eine Partialkorrelation zwischen A und B nichts anderes ist, als die Korrelation der Residuen, wenn A und B jeweils durch die Drittvariablen vorhergesagt werden.

```
1 RegA <- lm(A~C1+...+Ck,DATENSATZ)
2 RegB <- lm(B~C1+...+Ck,DATENSATZ)
3 cor.test(residuals(RegA),residuals(RegB))
4 #Beachte: da cor.test() von einer Korrelation ausgeht,
5 #stimmen die Freiheitsgrade, t-Wert und p-Wert nicht
```

Die Semipartialkorrelation funktioniert ähnlich, ausser das hier nur die Residuen aus einer Variable verwendet werden:

```
1 # Die Drittvariablen werden hier nur aus B herauspartialisiert
2 RegB <- lm(B~C1+...+Ck,DATENSATZ)
3 cor.test(DATENSATZ$A,residuals(RegB))
4 #Beachte: da cor.test() von einer Korrelation ausgeht,
5 #stimmen die Freiheitsgrade, t-Wert und p-Wert nicht
```

Die p -Werte müssen wie zuvor händisch berechnet werden!

5. (Semi-)Partialkorrelationen basierend auf der Regression - Übung

Nehmen Sie an, dass die Studie im nächsten Jahr wiederholt wurde. Laden Sie dazu den Datensatz **Uebung_Statistik3_2A.csv** herunter und importieren Sie ihn in R.

Der Datensatz hat Spalten und Zeilen.

Bestimmen Sie folgende (Semi-)Partialkorrelationen mittels R:

- $r_{DM \bullet I} =$
- $r_{D(M \bullet I)} =$
- $r_{M(D \bullet I)} =$
- $r_{MD \bullet IP} =$
- $r_{M(D \bullet IP)} =$

Überlegen Sie, was der geringe Unterschied zwischen $r_{DM \bullet I}$ und $r_{MD \bullet IP}$ verrät.

6. (Semi-)Partialkorrelationen mit ppcor - Einführung 1

Falls Sie das Package *ppcor* noch nicht in R geladen haben, sollten Sie es spätestens jetzt tun.

Das Package bietet zwei Funktionen, die relevant sind: `pcor()` und `spcor()`. Beide funktionieren sehr ähnlich:

- Sie erhalten als Input einen Datensatz, der nur die relevanten Spalten beinhaltet
- `pcor()` ermittelt Partialkorrelationen, `spcor()` Semipartialkorrelationen
- Es werden immer die (Semi-)Partialkorrelationen von zwei Variablen unter Herauspartialisierung aller anderen Variablen bestimmt (= höchste mögliche Ordnung)
- der Output enthält jeweils drei Tabellen:
 - die Koeffizienten
 - die p -Werte
 - die t -Werte
- Spalte und Zeile benennen die Variablen, deren Zusammenhang analysiert wurde
- bei `spcor()` verrät die Spalte aus welcher Variable herauspartialisiert wurde
- die p -Werte und t -Werte sind korrekt

6. (Semi-)Partialkorrelationen mit ppcor - Einführung 2

Nehmen wir an, wir haben einen Datensatz data mit fünf Spalten (VP_Code,A,B,C,D)

```
1 library(ppcor)
2 # Partialkorrelationen zweiter Ordnung
3 pcor(data[,2:5]) #2. bis 5. Spalte ist relevant oder
4 pcor(data[,c("A","B","C","D")]) #oder
5
6 # Semipartialkorrelationen zweiter Ordnung
7 spcor(data[,2:5]) #2. bis 5. Spalte ist relevant oder
8 spcor(data[,c("A","B","C","D")]) #oder
```

```
1 # Partialkorrelationen erster Ordnung
2 pcor(data[,c("A","B","C",)])
3 pcor(data[,c("A","B","D",)])
4 pcor(data[,c("A","C","D",)])
5 pcor(data[,c("B","C","D",)]) #oder
6
7 pcor(data[,2:4])
8 pcor(data[,c(2,3,5)])
9 pcor(data[,c(2,4,5)])
10 pcor(data[,3:5])
```

6. (Semi-)Partialkorrelationen mit ppcor - Übung

Bestimmen Sie folgende (Semi-)Partialkorrelationen mittels pcor() und spcor(). Sie sollten auf nahezu identische Werte kommen wie zuvor.

- $r_{DM \bullet I} =$
- $r_{D(M \bullet I)} =$
- $r_{M(D \bullet I)} =$
- $r_{MD \bullet IP} =$
- $r_{M(D \bullet IP)} =$