

Statistik 3

Übung 1

Einleitung und Ziele der Übung

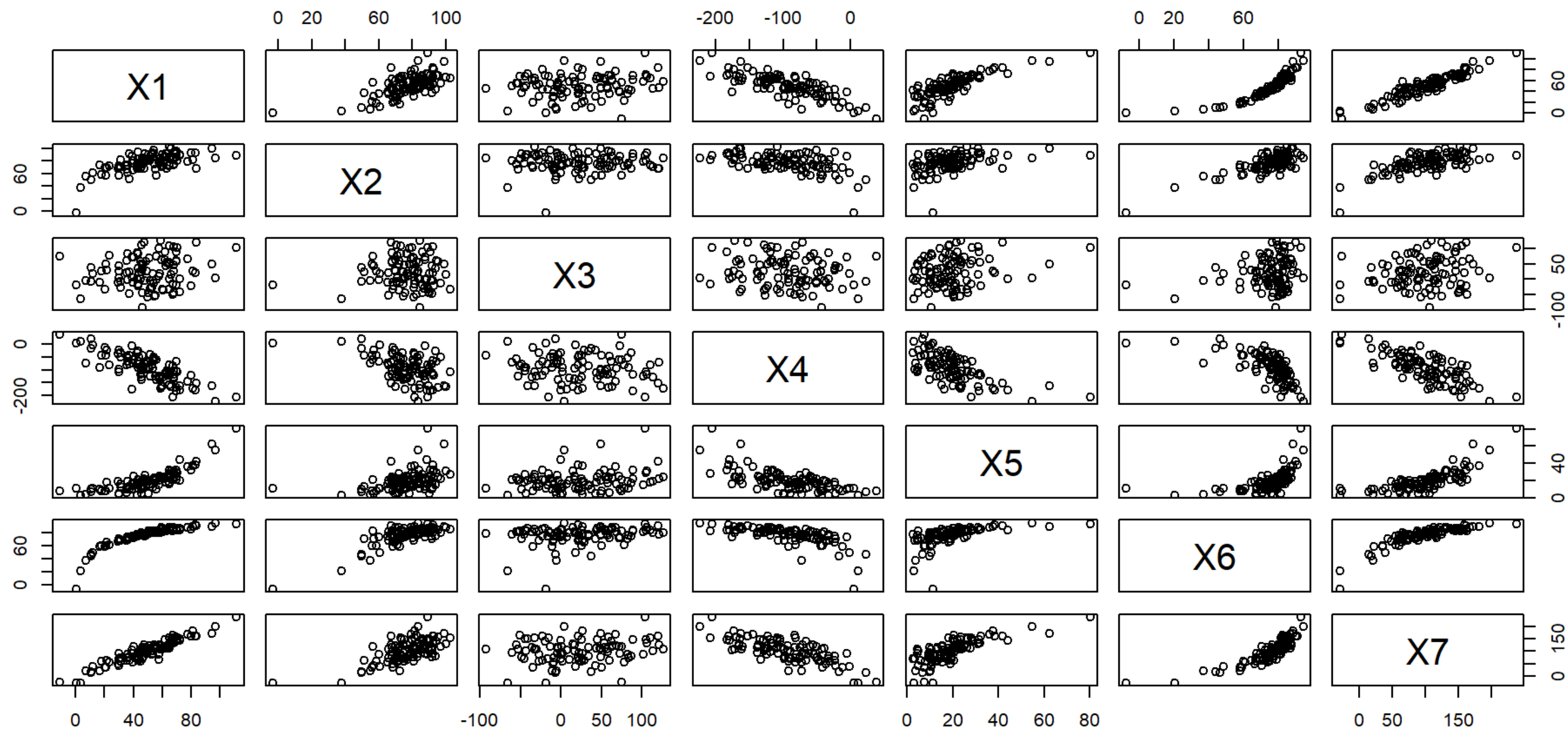
Herzlich Willkommen zur Übung Statistik 3. Diese erste Übung dient der Wiederholung der einfachen linearen Regression. Vieles wird Ihnen sicherlich vertraut vorkommen. Häufig sind es auch Sachen, die Studierende in der Prüfung bzw. unter Stress gerne falsch machen.

Agenda

- Kurze Vorstellung Ihres Übungsleiter*in und ihrer/-s Tutor*in
- Visualisieren von Zusammenhängen und deren Interpretation
- Bestimmen von Regressionskoeffizienten schriftlich (mit Taschenrechner)
- Regressionen in R durchführen und auf statistische Signifikanz prüfen
- Überprüfung von Voraussetzungen/Annahmen und Interpretation des R-Outputs

1. Visualisieren von Zusammenhängen und deren Interpretation 1

Betrachten Sie folgende Graphen. Entscheiden Sie jeweils ob ein linearer Zusammenhang bestehen könnte. Was spricht dafür, was spricht dagegen. Überlegen Sie erst allein und tauschen Sie sich anschliessend mit Ihren Sitznachbar*innen aus.



1. Visualisieren von Zusammenhängen und deren Interpretation 2

Wie Sie sicherlich gesehen haben, gab es einige Zusammenhänge die eindeutig linear aussahen, z.B. X1 und X4, X1 und X7, X4 und X7.

Genauso gab es einige Zusammenhänge die eindeutig nicht linear aussahen, z.B. X1 und X6, X6 und X7, X3 und X6.

Es gibt aber einen grossen Graubereich, hier muss dann entweder inhaltlich entschieden werden oder es müssen genauere Analysen erfolgen.

Diese Daten werden später nochmals im Rahmen der Analyse mit R genauer angeschaut.

2. Bestimmen von Regressionskoeffizienten mit dem Taschenrechner 1

Jetzt sollen Sie als nächstes nochmal Regressionskoeffizienten mit Taschenrechner und per Hand bestimmen. Dafür sollen nur die Daten von 4 Personen verwendet werden und Merkmal 1 durch Merkmal 2 vorhergesagt werden.

Person	Merkmal 1	Merkmal 2
A	5	11
B	3	7
C	5	10
D	7	8

Überlegen Sie zunächst selbstständig, welche Werte Sie berechnen müssen (und welche Werte Sie dafür benötigen). Nutzen Sie dann die Formelsammlung für Statistik 3 um die Koeffizienten zu bestimmen. Auf der nächsten Seite finden Sie eine Eingabemaske, mit der Sie Ihre Zwischenergebnisse überprüfen können.

2. Bestimmen von Regressionskoeffizienten mit dem Taschenrechner 2

Es wird alles auf 2 Nachkommastellen gerundet und mit diesen Werten weitergerechnet. Bitte geben Sie die Nachkommastellen auch an, selbst wenn diese 0 sind.

Z.B. 10.00 statt 10

Wert		Ergebnis
$\overline{x}$	=	
$\overline{y}$	=	
$\hat{\sigma}_X^2$	=	
$\hat{\sigma}_{XY}$	=	
b_1	=	
b_0	=	

Schreiben Sie anschliessend die Regressionsgleichung mit den eingesetzten Koeffizienten auf! (Hinweis: Grossbuchstaben; den Fehler und die Nachkommastellen nicht vergessen!)

2. Bestimmen von Regressionskoeffizienten mit dem Taschenrechner 3

Was sagen b_0 und b_1 jeweils inhaltlich aus? Wie stehen b_1 und r inhaltlich zueinander (in der einf. lin. Reg.)?
Überlegen Sie sich jeweils kurz eine Antwort auf die 3 Aspekte.

3. Regressionen in R durchführen und auf statistische Signifikanz prüfen 1

Jetzt gehen wir zurück auf die Zusammenhänge vom Anfang. Laden Sie zunächst den Datensatz **Uebung_Statistik3_1A.csv** herunter und importieren Sie ihn nach R mit dem `df`. Er sollte aus 100 Zeilen und 7 Variablen bestehen.

Adaptieren Sie anschliessend den untenstehenden Code um sich das Ergebnis für die Regression von folgenden drei Varianten anzuschauen.

- X3 vorhergesagt durch X1
- X4 vorhergesagt durch X1
- X6 vorhergesagt durch X1

```
1 RegressionAVUV <- lm(AV ~ UV, DATENSATZ)
2 summary(RegressionAVUV)
```

Sobald Sie die drei Summaries haben, sind sie bereit für die nächste Aufgabe.

3. Regressionen in R durchführen und auf statistische Signifikanz prüfen 2

Bitte auf 2 Nachkommastellen runden, ausser bei p -Werten, dort 3 Nachkommastellen.
Bitte geben Sie die Nachkommastellen auch an, selbst wenn diese 0 sind.
Bei Werten, deren Minimum nur 0 oder -1 betragen kann und ihr Maximum 1, ist die führende 0 wegzulassen.

AV	Wert	Ergebnis
X3	b_0	=
X4	b_1	=
X3	R^2	=
X6	df_{b_0}	=
X4	df_{b_0}	=
X3	p_{b_1}	=

Auf der nächsten Seite geht es weiter.

3. Regressionen in R durchführen und auf statistische Signifikanz prüfen 3

Warum unterscheiden sich die Freiheitsgrade der Regressionen?

Welche Koeffizienten sind signifikant für $\alpha = .05$, welche für $\alpha = .10$?

Warum können Sie die Regressionskoeffizienten zwischen den Modellen nicht direkt vergleichen, um zu unterscheiden, welche Variable am besten durch X1 vorhergesagt wird? Was bräuchten wir um dies zu tun?

Wie können Sie sich die Residuen ausgeben lassen, wie die geschätzten Werte?

Wie gross ist der geschätzte Wert für X4, wenn X1 den Wert 5 hat?

4. Überprüfung von Voraussetzungen/Annahmen und Interpretation des R-Outputs - 1

Schauen Sie zunächst nochmals nach, wie Streudiagramme für unsere drei Regressionen aussahen. Im Idealfall sollten wir bei der Prüfung der Annahmen, Verletzungen entdecken, wenn das Streudiagramm auch schon nicht linear aussah.

Lassen Sie sich die Diagramme zur Überprüfung der Annahmen für jeder der Regressionen nach folgendem Schema ausgeben.

```
1 library(performance)
2 check_model(RegressionAVUV)
```

Vergleichen Sie anschliessend die Ergebnisse für die verschiedenen Regressionen. Welches Regression erfüllt die Annahmen am besten, welches eigentlich gar nicht und welches ist noch in Ordnung aus ihrer Sicht? Begründen Sie. Was bedeutet dies für die Durchführbarkeit der Regressionen und die erzielten Ergebnisse?