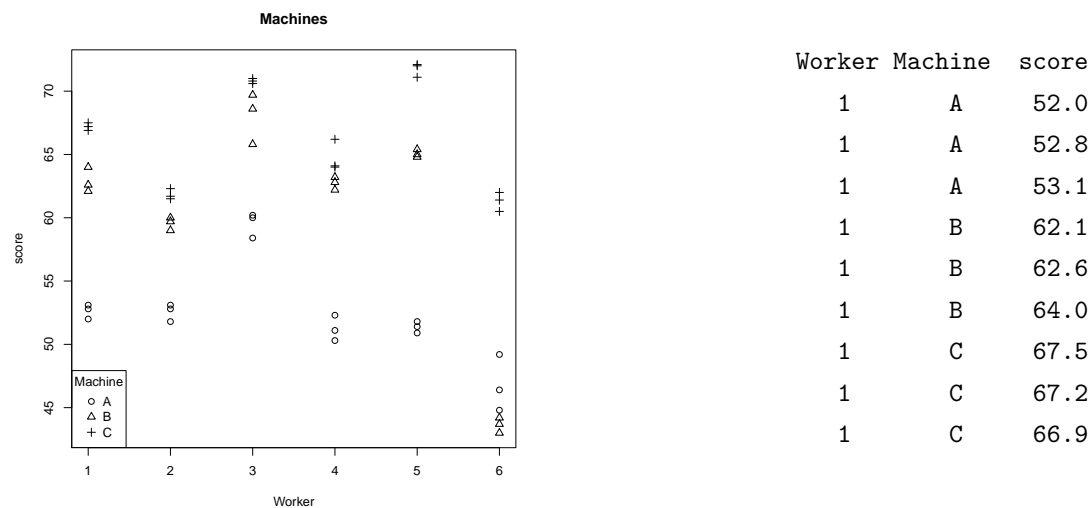


**Aufgabe 1**

Der Datensatz *Machines* aus dem R-Paket nlme wird in der Hilfe folgendermaßen beschrieben:

Data on an experiment to compare three brands of machines used in an industrial process are presented in Milliken and Johnson (p. 285, 1992). Six workers were chosen randomly among the employees of a factory to operate each machine three times. The response is an overall productivity score taking into account the number and quality of components produced.

Es handelt sich um ein randomisiertes geblocktes Design mit wiederholten Messungen. Abbildung 1 zeigt eine mögliche graphische Darstellung der Scores geordnet nach Arbeitern und Maschinentyp, sowie die Daten für Arbeiter 1.



**Abbildung 1:** Produktivitätsscores auf drei Maschinen getrennt nach Arbeiter (links) und Daten für Arbeiter 1 (rechts).

- (a) Handelt es sich um ein gekreuztes oder ein genestetes Design? Begründen Sie Ihre Antwort.
- (b) Sie möchten nun den Produktivitätsscore in Abhängigkeit der Maschine und der Arbeiter modellieren. Warum sollte man dazu feste und zufällige Effekte verwenden?
- (c) Stellen Sie ein Modell mit festen Effekten für die Maschine und zufälligen Effekten für die Arbeiter auf. Geben Sie die Modellgleichung für eine Beobachtung  $y_{ijk}$  an. Dabei ist  $y_{ijk}$  der Score für Arbeiter  $i$  an Maschine  $j$  in der  $k$ -ten Wiederholung. Vergessen Sie dabei nicht Verteilungsannahmen anzugeben.
- (d) Schreiben Sie das Modell aus (c) für den Arbeiter  $i$  auf, also die Modellgleichung für den Vektor  $y_i = (y_{i11}, \dots, y_{i33})'$ .
- (e) Schreiben Sie die Modellgleichung aus (c) und (d) in Matrixnotation für den ganzen Datensatz auf. Geben Sie auch die Kovarianzmatrizen für die zufälligen Effekte und für die Residuen an. Leiten Sie daraus die Kovarianzmatrix der Zielvariable ab.
- (f) Wenn man das Modell aus (c)/(d)/(e) in R schätzt, erhält man folgenden Output:

```
Fixed: score ~ -1 + Machine
      MachineA MachineB MachineC
      52.35556 60.32222 66.27222
Random effects:
Formula: ~1 | Worker
      (Intercept) Residual
StdDev:    5.146552 3.161647
```

Die Prognosen für die zufälligen Effekte sind:

```
(Intercept)
6 -8.70711058
2 -1.59425968
4 -0.06931564
1  1.21035769
3  6.21174760
5  2.94858062
```

Interpretieren Sie die Parameter aus dem Modell sowie die zufälligen Effekte für Arbeiter 6 und für Arbeiter 3. Welche Leistung erwartet man laut Modell für Arbeiter 6 an Maschine B?

- (g) Stellen Sie ein Modell mit einem festen Effekt für den Maschinentyp und zufälligen Effekten für die Interaktionen aus Arbeitern und Maschinen auf (ohne zufällige Effekte für die Arbeiter). Nehmen Sie an, dass die zufälligen Interaktionseffekte innerhalb jedes Arbeiters korreliert sein können. Für die Residuen wird Unabhängigkeit angenommen. Geben Sie die Modellgleichung für Arbeiter  $i$  an, also das Modell für den Vektor  $y_i$ .
- (h) Wenn man das Modell aus (g) in R schätzt, erhält man folgenden Output:

```
Fixed: score ~ -1 + Machine
      MachineA MachineB MachineC
      52.35556 60.32222 66.27222
Random effects:
Formula: ~Machine - 1 | Worker
Structure: General positive-definite, Log-Cholesky parametrization
      StdDev      Corr
MachineA 4.0792807 MachnA MachnB
MachineB 8.6252908  0.803
MachineC 4.3894795  0.623  0.771
Residual 0.9615766
```

Für die zufälligen Effekte erhält man:

```
MachineA MachineB MachineC
6 -5.5916013 -16.5838056 -5.0300737
2  0.1838844 -0.8033247 -4.2822823
4 -1.0238706  2.3284558 -1.4146283
1  0.3119890  2.5532237  0.9302963
3  6.9692218  7.7793500  4.4735105
5 -0.8496233  4.7261008  5.3231775
```

Interpretieren Sie die Korrelation zwischen den zufälligen Effekten. Welche Leistung erwartet man nach Modell (h) für Arbeiter 6 an Maschine B?

## Aufgabe 2

In dieser Aufgabe soll der Datensatz *Oxid* aus dem R-Paket *nlme* ausgewertet werden. Die Zielgröße ist die Dicke der Oxidschicht (*Thickness*) auf Siliziumscheiben (*Wafer*). Dazu wurden zufällig acht Parzellen (*Lot*) ausgewählt. Aus jeder Parzelle wurden drei Siliziumscheiben gezogen, bei denen je an drei unterschiedlichen Stellen (*Site*) die Dicke der Oxidschicht gemessen wurde. Das Ziel der Studie ist es, die Varianzkomponenten zu schätzen, die zu den verschiedenen Ebenen (Parzellen, Scheiben) gehören, um die Ursachen für die Variabilität in der Ablagerung von Oxid zu evaluieren (Pinheiro, J. C.; Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York). Im folgenden sehen Sie die Beobachtungen für die erste Parzelle.

Lot	Wafer	Site	Thickness
1	1	1	2006
1	1	2	1999
1	1	3	2007
1	2	1	1980
1	2	2	1988
1	2	3	1982
1	3	1	2000
1	3	2	1998
1	3	3	2007

- Handelt es sich um ein gekreuztes oder ein genestetes Design? Begründen Sie Ihre Antwort.
- Sie möchten nun die Dicke der Oxidschicht in Abhängigkeit der Parzellen und Scheiben modellieren. Warum sollte man zufällige Effekte verwenden?
- Stellen Sie ein Modell für die Dicke der Oxidschicht  $y_{ijk}$  gemessen an der  $k$ -ten Stelle an der  $j$ -ten Scheibe in der  $i$ -ten Parzelle auf. Vergessen Sie nicht die Modellannahmen anzugeben.
- Wenn man das Modell in R schätzt, erhält man u.a. folgenden Output:

```
Fixed: Thickness ~ 1
      (Intercept)
      2000.153
Random effects:
Formula: ~1 | Lot
      (Intercept)
      StdDev:    11.39768
Formula: ~1 | Wafer %in% Lot
      (Intercept) Residual
      StdDev:    5.988802 3.545341
```

Die Prognosen für die zufälligen Effekte sind:

```
Level: Lot          Level: Wafer %in% Lot
(Intercept)        (Intercept)
1 -3.4634693        1/1  6.54599261
2 -11.2216405       1/2 -11.95893915
3  0.8690159        1/3  4.45672612
4 -4.4710240        2/1  0.65859292
5 13.4634497        2/2 -0.83374029
```

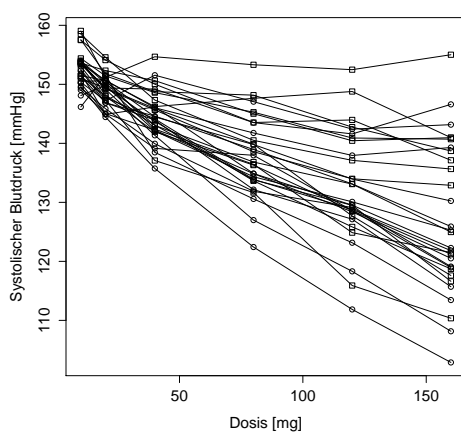
6	19.4080224	2/3	-2.92300678
7	-8.1989764	[...]	
8	-6.3853779	8/1	-0.09019651
		8/2	1.40213670
		8/3	-3.07486292

Vergleichen Sie die geschätzten Varianzkomponenten für die Ebene der Parzellen und der Scheiben. Auf welcher Ebene gibt es mehr Variabilität? Welche Dicke der Oxidschicht erwartet man laut Modell für Parzelle 8? Welche Dicke der Oxidschicht erwartet man laut Modell für Scheibe 2 aus Parzelle 8?

- (e) Berechnen Sie die Kovarianz zweier Messungen an der gleichen Scheibe, sowie die Kovarianz zweier Messungen in der gleichen Parzelle aber für unterschiedliche Scheiben je bei marginaler und konditionaler Sichtweise.

### Aufgabe 3

In einer klinischen Studie zur Dosisfindung für ein blutdrucksenkendes Medikament werden 30 Patienten (Variable *ID*) 6 verschiedene Dosen (*dosis*, 10, 20, 40, 80, 120, 160 mg) verabreicht und es wird jeweils der systolische Blutdruck (*SBD*) gemessen. Die Messungen sind in Abbildung 2 dargestellt, getrennt für Männer (*gender* = 1) und Frauen (*gender* = 0). (Simulierte Daten)



	ID	gender	dosis	SBD
	1	1	0	153.7581
	2	1	0	149.2582
	3	1	0	145.6593
	4	1	0	138.5844
	5	1	0	133.9929
	6	1	0	132.8894

**Abbildung 2:** Systolischer Blutdruck nach Gabe verschiedener Dosen eines blutdrucksenkenden Medikaments für Frauen (Quadrate) und Männer (Kreise), sowie die ersten Zeilen des Datensatzes.

- (a) Zunächst wird ein lineares Modell angepasst, in dem die Beobachtungen als unabhängige Beobachtungen behandelt werden:

$$SBD_i = \beta_0 + \beta_1 \text{gender}_i + \beta_2 * (\text{dosis}_i) + \beta_3 \text{gender}_i \text{dosis}_i + \varepsilon_i, \quad i = 1, \dots, 180, \quad \varepsilon_i \text{ unabh.}$$

Warum ist diese Modellierung problematisch?

- (b) Nun wird ein Modell mit festen Effekten für Alter, Geschlecht sowie der Interaktion von Alter mit Geschlecht und zufälligen Intercepts aufgestellt. Warum ist diese Modellierung besser für die Daten geeignet?

$$SBD_{ij} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{dosis}_{ij} + \beta_3 \text{gender}_i \text{dosis}_{ij} + b_i + \varepsilon_{ij}, \quad i = 1, \dots, 30, \quad j = 1, \dots, 6$$

$$b_i \stackrel{\text{iid}}{\sim} N(0, \tau^2), \quad \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \quad b_i, \varepsilon_{ij} \text{ unabhängig}$$

- (c) Vergleichen Sie die geschätzten Standardfehler der festen Parameter des Modells aus (a), das die Abhängigkeit in den Daten ignoriert und des gemischten Modells aus (b), das die Abhängigkeit modelliert. Die Parameter und ihre Standardabweichungen sind in Tabelle 1 zusammengefasst.

**Tabelle 1:** Schätzer der festen Effekte mit geschätzten Standardabweichungen in Klammern für das einfache lineare Modell (LM) und das Modell mit zufälligen Effekten (LMM).

	Intercept $\hat{\beta}_0(\hat{s}d)$	gender $\hat{\beta}_1(\hat{s}d)$	dosis $\hat{\beta}_2(\hat{s}d)$	dosis:gender $\hat{\beta}_3(\hat{s}d)$
LM	153.81 (1.22)	-2.58 (1.73)	-0.15 (0.013)	-0.025 (0.020)
LMM	153.81 (1.59)	-2.58 (2.24)	-0.15 (0.009)	-0.025 (0.013)

- (d) Betrachten Sie ein Modell mit zufälligen Intercepts und zufälligen Steigungen über die Dosis:

$$SBD_{ij} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{dosis}_{ij} + \beta_3 \text{gender}_i \text{dosis}_{ij} + b_{0i} + b_{1i} \text{dosis}_{ij} + \varepsilon_{ij}, \quad i = 1, \dots, 30, j = 1, \dots, 6$$

$$\mathbf{b}_i = (b_{0i}, b_{1i})' \stackrel{\text{iid}}{\sim} N(0, \mathbf{D}), \quad \mathbf{D} = \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{pmatrix} \text{ positiv semi-definit}$$

$$\varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2),$$

$\mathbf{b}_i, \varepsilon_{ij}$  unabhängig

Man erhält folgenden Output in R für die Schätzungen:

```
Fixed: SBD ~ gender + dosis + gender * dosis
(Intercept)      gender      dosis gender:dosis
153.81288362    -2.58605855    -0.15363045    -0.02535405
```

Random effects:

```
Formula: ~1 + dosis | ID
Structure: General positive-definite, Log-Cholesky parametrization
StdDev      Corr
(Intercept) 0.17587701 (Intr)
dosis       0.07395978 0.939
Residual    2.35185642
```

Interpretieren Sie die geschätzte Korrelation von 0.939 der zufälligen Effekte.

- (e) Berechnen Sie die Kovarianz zweier Blutdruckmessungen an der gleichen Person bei `dosis = 10` und `dosis = 20` (marginale und konditionale Sichtweise).