

Aufgabe 1:

Im Datensatz `leprosy` sind die Anzahl von Lepra-Bazillen vor und nach der Behandlung mit Antibiotika enthalten. Es existieren für jeden Patienten folgende Variablen: Medikament (`Drug`), Anzahl der Bazillen (`count`) und Zeit (`time`). Medikament C entspricht einer Placebo-Behandlung. `time 0` entspricht dem Zustand vor der Behandlung, `time 1` dem Zustand nach der Behandlung.

Benutzen Sie den folgenden Code um den Datensatz einzulesen:

```
R>url <- "http://www.statistik.lmu.de/institut/lehrstuhl/semwiso/longitudinal_ss08/"
R>leprosy <- read.table(paste(url, "download/leprosy.txt",
+   sep = ""), header = T)
R>leprosy$Drug <- factor(leprosy$Drug, levels = c("C", "A", "B"))
```

Lösung:

(Aufgabe adaptiert von Michael Hoehle, SS07)

- (a) Bestimmen Sie Mittelwert und Varianz der Anzahlen der Bazillen vor und nach der Behandlung für jede Medikamentengruppe und plotten Sie die Verläufe getrennt nach Gruppen.

Lösung:

```
R>print(xyplot(count ~ time | Drug, groups = id, data = leprosy,
+   type = "b", layout = c(3, 1)))
R>(mean.Pre <- with(subset(leprosy, time == "0"), tapply(count,
+   Drug, mean)))
```

```
      C      A      B
12.9  9.3 10.0
```

```
R>(var.Pre <- with(subset(leprosy, time == "0"), tapply(count,
+   Drug, var)))
```

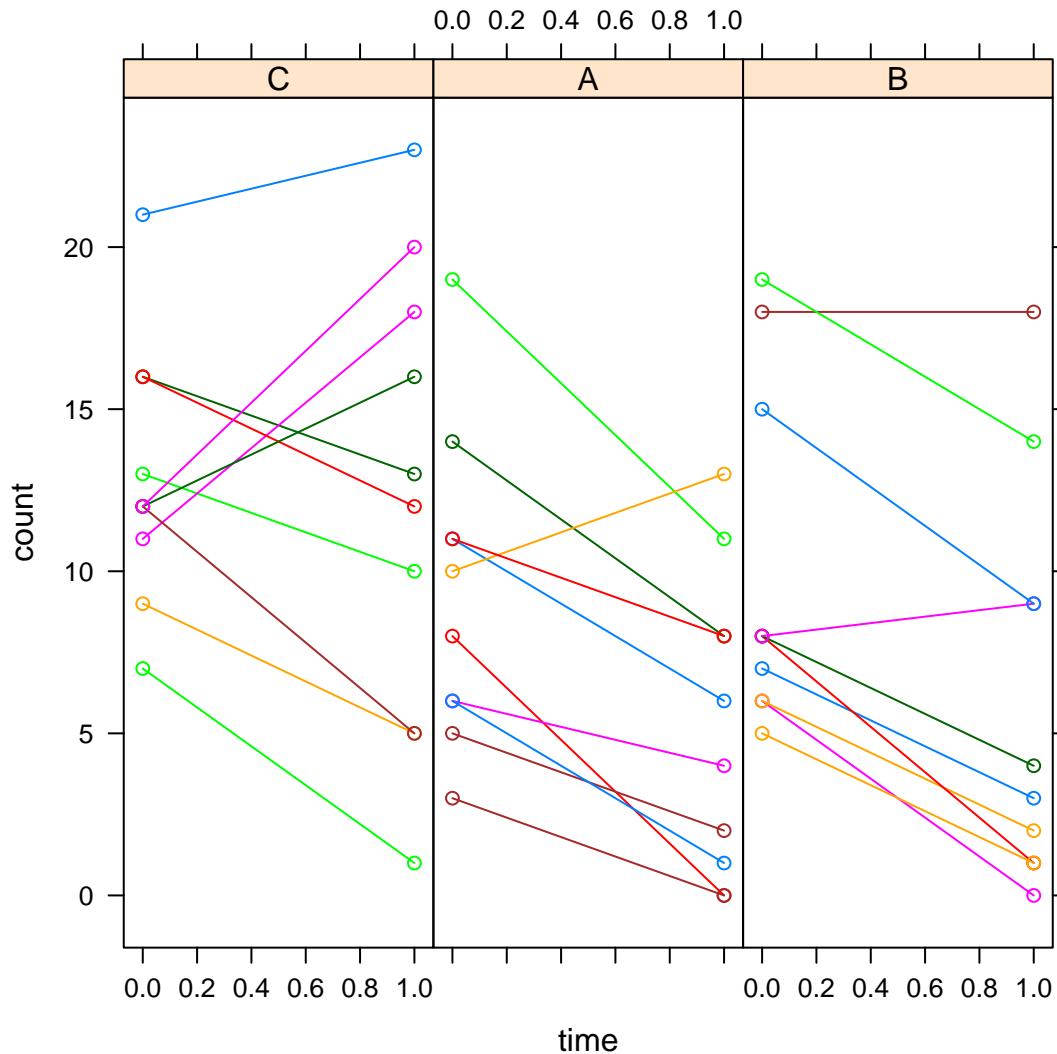
```
      C      A      B
15.65556 22.67778 27.55556
```

```
R>(mean.Post <- with(subset(leprosy, time == "1"), tapply(count,
+   Drug, mean)))
```

```
      C      A      B
12.3  5.3  6.1
```

```
R>(var.Post <- with(subset(leprosy, time == "1"), tapply(count,
+   Drug, var)))
```

```
      C      A      B
51.12222 21.56667 37.87778
```



⇒ Große Varianzen deuten auf Überdispersion hin.

- (b) Passen Sie mit Hilfe verallgemeinerter Schätzgleichungen (GEE) ein Modell an die Daten an, welches als Einflussgrößen die Variablen Zeit und eine Interaktion von Behandlung mit Zeit hat. Gehen Sie von unabhängigen Beobachtungen aus. Was verbirgt sich hinter den 'robusten' und den 'naiven' Schätzungen der Varianzen der Parameter? Beurteilen Sie wie sinnvoll die getroffene Annahme für die Korrelationsstruktur der Daten ist.

Hinweis: Benutzen Sie das `gee`-Paket.

Lösung:

Kein Haupteffekt für Drug weil randomisierte Studie (→ Annahme dass Verteilung VOR Behandlung in allen 3 Gruppen gleich ist.)

```
R>library(gee)
R>leprosy.gee.1 <- gee(count ~ time + time:Drug, data = leprosy,
+   family = poisson(link = "log"), id = id, corstr = "independence")
```

```
(Intercept)      time  time:DrugA  time:DrugB
  2.3733542    0.1362451  -0.8418924  -0.7013105
```

```
R>summary(leprosy.gee.1)
```

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:

Link: Logarithm
 Variance to Mean Relation: Poisson
 Correlation Structure: Independent

Call:

```
gee(formula = count ~ time + time:Drug, id = id, data = leprosy,
    family = poisson(link = "log"), corstr = "independence")
```

Summary of Residuals:

	Min	1Q	Median	3Q	Max
	-11.3000000	-4.4083333	-0.5166667	3.3750000	11.9000000

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	2.3733542	0.1031749	23.0032164	0.08013786	29.6158898
time	0.1362451	0.1962463	0.6942557	0.16833050	0.8093905
time:DrugA	-0.8418924	0.3042060	-2.7675078	0.31545382	-2.6688294
time:DrugB	-0.7013105	0.2899301	-2.4188948	0.34932356	-2.0076244

Estimated Scale Parameter: 3.427708

Number of Iterations: 1

Working Correlation

	[,1]	[,2]
[1,]	1	0
[2,]	0	1

'Naive' Varianzschätzung aus Modell-basierter Schätzung für $\text{Cov}(\beta)$. 'Robuste' Varianzschätzung aus Sandwich-Estimator (s. Folien 11.19 f.). Hier teilweise deutliche Abweichung, was darauf hindeutet dass die Annahme unkorrelierter Beobachtungen (`corstr = 'independence'`) eher ungünstig ist. Wenn Sandwich-Estimator und Modell-basierter Schätzer sehr ähnlich sind bedeutet dass ja eben dass $\mathbf{M} \approx \mathbf{B}$ bzw. $\widehat{\text{Cov}}(\mathbf{Y}_i) \approx \hat{\mathbf{V}}_i$, dass also die working correlation nah an der beobachteten liegt.

Äquivalentes Modell: Naives quasi-poisson glm:

```
R>summary(glm(formula = count ~ time + time:Drug, family = quasipoisson(link = "log"),
+ data = leprosy))
```

Call:

```
glm(formula = count ~ time + time:Drug, family = quasipoisson(link = "log"),
    data = leprosy)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-4.1929	-1.5772	-0.1562	1.0958	3.8929

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.3734	0.1032	23.003	< 2e-16 ***

```

time          0.1362      0.1962   0.694  0.49039
time:DrugA    -0.8419      0.3042  -2.768  0.00764 **
time:DrugB    -0.7013      0.2899  -2.419  0.01884 *

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 3.427716)

```

Null deviance: 255.20 on 59 degrees of freedom
Residual deviance: 207.17 on 56 degrees of freedom
AIC: NA

```

Number of Fisher Scoring iterations: 5

⇒ Schätzer gleich, Varianzen der Parameter entsprechen naiver Schätzung aus gee.

- (c) Schätzen Sie das Modell erneut unter der Annahme dass Beobachtungen am selben Patienten nicht unabhängig sind.

Lösung:

```

R>leprosy.gee.2 <- gee(count ~ time + time:Drug, data = leprosy,
+   family = poisson(link = "log"), id = id, corstr = "unstructured")

```

```

(Intercept)          time  time:DrugA  time:DrugB
  2.3733542    0.1362451  -0.8418924  -0.7013105

```

```

R>summary(leprosy.gee.2)

```

```

GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

```

```

Model:
Link:                Logarithm
Variance to Mean Relation: Poisson
Correlation Structure: Unstructured

```

```

Call:
gee(formula = count ~ time + time:Drug, id = id, data = leprosy,
    family = poisson(link = "log"), corstr = "unstructured")

```

```

Summary of Residuals:
      Min       1Q   Median       3Q      Max
-9.7025096 -4.7333333 -0.4000985  3.5166667 12.2974904

```

```

Coefficients:
              Estimate Naive S.E.      Naive z Robust S.E.      Robust z
(Intercept)  2.373354164  0.1034103 22.95084777  0.08013786 29.61588978
time         -0.002875909  0.1239320 -0.02320554  0.15700486 -0.01831732
time:DrugA   -0.562568679  0.2023204 -2.78058272  0.22198194 -2.53429933
time:DrugB   -0.495284414  0.1977550 -2.50453486  0.23420052 -2.11478786

```

Estimated Scale Parameter: 3.443369
 Number of Iterations: 5

Working Correlation
 [,1] [,2]
 [1,] 1.0000000 0.7383648
 [2,] 0.7383648 1.0000000

⇒ starke geschätzte Korrelation der Beobachtungen, deutlich bessere Übereinstimmung zwischen naiver und robuster Varianzschätzung (s.o.).

Vergleich Parameterschätzungen:

R>summary(leprosy.gee.1)\$coefficients

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	2.3733542	0.1031749	23.0032164	0.08013786	29.6158898
time	0.1362451	0.1962463	0.6942557	0.16833050	0.8093905
time:DrugA	-0.8418924	0.3042060	-2.7675078	0.31545382	-2.6688294
time:DrugB	-0.7013105	0.2899301	-2.4188948	0.34932356	-2.0076244

R>summary(leprosy.gee.2)\$coefficients

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	2.373354164	0.1034103	22.95084777	0.08013786	29.61588978
time	-0.002875909	0.1239320	-0.02320554	0.15700486	-0.01831732
time:DrugA	-0.562568679	0.2023204	-2.78058272	0.22198194	-2.53429933
time:DrugB	-0.495284414	0.1977550	-2.50453486	0.23420052	-2.11478786

Behandlungseffekte werden kleiner geschätzt.

- (d) Überprüfen Sie mit Hilfe der (multivariaten) Wald-Statistik, ob Behandlungseffekte der beiden Antibiotika vorliegen (Nullhypothese: $\beta_3 = \beta_4 = 0$). Interpretieren Sie die Parameterschätzer.

Lösung:

Die Wald-Statistik für $H_0 : \mathbf{L}\boldsymbol{\beta} = 0$ ist $W^2 = (\mathbf{L}\hat{\boldsymbol{\beta}})' \left(\mathbf{L}\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}})\mathbf{L}' \right)^{-1} (\mathbf{L}\hat{\boldsymbol{\beta}})$. Die Kontrastmatrix in diesem Fall ist $L = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

```
R>beta <- as.matrix(coef(leprosy.gee.2))
R>L1 <- matrix(c(0, 0, 1, 0, 0, 0, 0, 1), ncol = 4, byrow = TRUE)
R>Sigma.star <- leprosy.gee.2$robust.variance
R>Sigma1 <- L1 %*% Sigma.star %*% t(L1)
R>(W1 <- t(L1 %*% beta) %*% solve(Sigma1) %*% (L1 %*% beta))
```

```
      [,1]
[1,] 7.339606
```

```
R>1 - pchisq(W1, 2)
```

```
      [,1]
[1,] 0.02548149
```

⇒ Deutlich signifikanter Effekt der Behandlungen gegenüber Placebo.

Interpretation der Parameterschätzer: $\exp(\beta_3)$ gibt die multiplikative Veränderung des Zeiteffektes in Behandlungsgruppe 'A' gegenüber der Placebogruppe an. D.h. in diesem Fall: Nach der Behandlung ist die erwartete Bazillenbelastung in Gruppe 'A' etwa das 0.57-fache der Belastung in der Placebo-Gruppe ($\exp(\beta_3) \approx 0.5695$).

- (e) Überprüfen Sie, ob die Behandlungseffekte der beiden Antibiotika gleich sind (Nullhypothese: $\beta_3 = \beta_4$).

Lösung:

Die Kontrastmatrix in diesem Fall ist $L = (0 \ 0 \ 1 \ -1)$

```
R>L2 <- matrix(c(0, 0, 1, -1), ncol = 4, byrow = TRUE)
R>Sigma2 <- L2 %*% Sigma.star %*% t(L2)
R>(W2 <- t(L2 %*% beta) %*% solve(Sigma2) %*% (L2 %*% beta))
```

```
      [,1]
[1,] 0.08847197
```

```
R>1 - pchisq(W2, 1)
```

```
      [,1]
[1,] 0.7661288
```

Die Nullhypothese, dass die beiden Behandlungen gleich effektiv sind, kann also nicht abgelehnt werden.

- (f) Fassen Sie beide Antibiotikabehandlungen zusammen und passen Sie das Modell entsprechend an. Geben sie ein Konfidenzintervall für den Behandlungseffekt an.

Lösung:

```
R>leprosy$trt <- factor(iffelse(leprosy$Drug == "C", "Placebo",
+ "Treatment"))
R>leprosy.gee.3 <- gee(count ~ time + time:trt, data = leprosy,
+ family = poisson(link = "log"), id = id, corstr = "unstructured")
```

```
(Intercept)          time time:trtTreatment
      2.3733542          0.1362451          -0.7691331
```

```
R>summary(leprosy.gee.3)
```

```
GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)
```

Model:

```
Link:                      Logarithm
Variance to Mean Relation: Poisson
Correlation Structure:     Unstructured
```

Call:

```
gee(formula = count ~ time + time:trt, id = id, data = leprosy,
```

```
family = poisson(link = "log"), corstr = "unstructured")
```

Summary of Residuals:

	Min	1Q	Median	3Q	Max
	-9.7027348	-4.7333333	-0.5080582	3.5166667	12.2972652

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	2.373354164	0.1027701	23.09381541	0.08013786	29.61588978
time	-0.002854870	0.1231875	-0.02317500	0.15700442	-0.01818337
time:trtTreatment	-0.527827856	0.1650404	-3.19817307	0.19882649	-2.65471601

Estimated Scale Parameter: 3.400867

Number of Iterations: 5

Working Correlation

	[,1]	[,2]
[1,]	1.000000	0.738253
[2,]	0.738253	1.000000

Konfidenzintervall für multiplikativen Effekt $\exp(\beta)$

```
R>exp(coef(leprosy.gee.3)[3] + c(0, -1, 1) * qnorm(0.975) * sqrt(leprosy.gee.3$robust.varia  
+ 3))
```

```
[1] 0.5898849 0.3995084 0.8709810
```

⇒ Durch Antibiotikabehandlung etwa 40% weniger Bazillen KI: (13% bis 61% weniger)