

Bitte für die Korrektur freilassen!			
Aufgabe	1	2	Σ
Punkte	20	25	45
erzielt			

**Übungs-Klausur-Aufgaben zur Vorlesung
“Generalisierte Regressionsmodelle”
im Wintersemester 2015/2016**

01. Februar 2016

Hinweise zu den Übungsaufgaben:

- (a) Im Folgenden finden Sie zu Übungszwecken zwei ausgewählte Aufgaben aus alten GRM-Klausuren.
- (b) Die Bearbeitungszeit der Bachelor-Klausur beträgt 120 Minuten. Bei der Bachelor-Klausur können 120 Punkte erreicht werden.
- (c) Die Bearbeitungszeit der Klausuren „Grundlagen der generalisierten Regression“ und „Fortgeschrittene generalisierte Regression“ beträgt jeweils 60 Minuten und es können jeweils 60 Punkte erreicht werden.
- (d) Da die Übungsaufgaben einen Umfang von 45 Punkten haben, sollte Sie unter Klausurbedingungen dementsprechend 45 Minuten für die beiden Übungsaufgaben benötigen.

Hinweise zu den Klausurbedingungen:

- (a) Überprüfen Sie bitte, ob Ihre Angabe vollständig ist. Die Angabe sollte (inklusive dieses Deckblatts) 3 Seiten und 2 Aufgaben umfassen.
- (b) Als Hilfsmittel sind ausschließlich ein nicht programmierbarer Taschenrechner sowie die Formelsammlung zugelassen.
- (c) Nicht-Deutsch-Muttersprachler dürfen zusätzlich ein Wörterbuch verwenden der Form: Muttersprache – deutsch / deutsch – Muttersprache.
- (d) Bei Unterschleif gilt die Klausur als nicht bestanden und es erfolgt eine Meldung an das Prüfungsamt.
- (e) Verwenden Sie für Ihre Notizen und Lösungen ausschließlich die Ihnen zur Verfügung gestellten Papierbögen.
- (f) Geben Sie bitte am Ende der Klausur alle von Ihnen zur Korrektur vorgesehenen Blätter sowie diese Angabe ab und kennzeichnen Sie jedes abgegebene Blatt mit Namen und Matrikelnummer.

Aufgabe 1

20 Punkte

Wir betrachten ein binäres Zufallsexperiment („Erfolg“ oder „Misserfolg“) mit Erfolgswahrscheinlichkeit p . Die Wahrscheinlichkeitsfunktion für die Anzahl y an Misserfolgen bis zum r -ten Erfolg ist gegeben durch

$$f(y|p, r) = \binom{y+r-1}{y} p^r (1-p)^y.$$

Hierbei ist r ein fester, bekannter Wert und es gelte $y \in \mathbb{N}_0, r \in \mathbb{N}, p \in (0, 1)$.

- Zeigen Sie, dass die Verteilung von y zu einer einfachen Exponentialfamilie gehört, und bestimmen Sie die Größen $\theta, b(\theta)$, den Dispersionsparameter ϕ sowie Erwartungswert μ und Varianz σ^2 von y .
- Leiten Sie den natürlichen Link $g(\cdot)$ als Funktion des Erwartungswerts μ her. Geben Sie die Responsefunktion $h(\cdot)$, geschrieben als Funktion des linearen Prädiktors η , an.

Aufgabe 2

25 Punkte

Der folgende Datensatz enthält die Ergebnisse einer Wählerbefragung, die vor der Wahl zum deutschen Bundestag 2009 durchgeführt wurden. Aus Gründen der Einfachheit beschränkt sich unser Datensatz auf die $n = 481$ Beobachtungen, in denen eine Wahlabsicht zugunsten der SPD oder der CDU angegeben wurde. Folgende Variablen wurden erhoben:

Variable	Beschreibung
WA	Wahlabsicht (1: SPD, 0: CDU)
Alter_z	Alter in Jahren, zentriert um Mittelwert (50.5 Jahre)
Gewerk	Gewerkschaftsmitglied (1: ja, 0: nein)
Abi	Abitur oder Fachhochschulreife (1: ja, 0: nein)
Religion	3 Kategorien: „evangelisch“, „katholisch“ oder „andere Religion als evangelisch oder katholisch“

Mithilfe der Funktion `glm()` wird eine logistische Regression gerechnet. Es ergibt sich unter anderem folgender Output:

```
glm(formula = WA ~ Alter_z + Gewerk + Abi + as.factor(Religion),
     family = binomial, data = wahl)
```

```
Family: binomial
Link function: logit
```

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.009789   0.172367  -0.057  0.9547
Alter_z      -0.012319   0.005928  -2.078  0.0377
Gewerk        0.870752    [???]    3.122  0.0018
Abi          -0.597899   0.254163  -2.352  0.0187
Religion_kath -0.425527   0.220265  [???]  0.0534
Religion_andere -0.161180  0.244786  -0.658  0.5102
---
```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 659.55  on 480  degrees of freedom
Residual deviance: 634.58  on 475  degrees of freedom
```

```
Number of Fisher Scoring iterations: 4
```

- (a) Geben Sie den linearen Prädiktor des verwendeten Modells an. Wie lauten Struktur- und Verteilungsannahme des Modells?
- (b) Interpretieren Sie die Koeffizienten der Variablen `Alter_z` und `Abi`. Gehen Sie dabei für jede Variable auch auf die Wirkung einer Erhöhung um eine Einheit ein.
- (c) Berechnen Sie basierend auf obigem Modell die Wahrscheinlichkeit dafür, dass ein 40-jähriger katholischer Wähler ohne Abitur, der Gewerkschaftsmitglied ist, beabsichtigt die CDU zu wählen.
- (d) Kann mit den vorliegenden Informationen beurteilt werden, ob die Religion eines Wählers zum Signifikanzniveau $\alpha = 0.10$ einen signifikanten Effekt auf die Wahlabsicht hat? Falls ja: wie? Falls nein: warum nicht?