

Prüfung im Fach Mikroökometrie im Wintersemester 2013/14 Aufgaben

Vorbemerkungen:

- Anzahl der Aufgaben:**
- Die Klausur besteht aus 4 Aufgaben.
- Bewertung:**
- Es können maximal 60 Punkte erworben werden. Die Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
- Erlaubte Hilfsmittel:**
- Tabellen der statistischen Verteilungen (sind der Klausur beigelegt)
 - Taschenrechner
 - Fremdwörterbuch
- Wichtige Hinweise:**
- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den exakten Wert der gesuchten Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
 - Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Annahme oder Angabe fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1 (22 Punkte)

Mit Befragungsdaten von 7000 Personen im Alter zwischen 17 und 98 Jahren werden Determinanten der Gesundheitszufriedenheit analysiert. Es liegen folgende Variablen vor:

sat = Zufriedenheit (1 = gering, 2 = mittel, 3 = hoch)
age = Alter
agesq = Alter quadriert
inc = monatliches Einkommen in Tausend Euro
educ = Bildung in Jahren
disease = chronische Krankheit (=1, sonst=0).

Die Schätzergebnisse eines geordneten Probit Modells sind wie folgt:

```
Iteration 0:  log likelihood = -5467.8914

Ordered probit regression           Number of obs   =       7000
                                   LR chi2(5)        =       427.16
                                   Prob > chi2         =       0.0000
Log likelihood = -5265.9881         Pseudo R2       =       ?
```

	sat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age		-.02552	.00560	-4.56	0.000	-.03650	-.01454
agesq		.00035	.00003	11.67	0.000	.00029	.00041
inc		.10637	.00912	11.66	0.000	.08849	.12426
educ		.02188	.00581	3.76	0.000	.01049	.03328
disease		-.62530	.03999	-15.64	0.000	-.70368	-.54692
/cut1		-1.456627	.1312109			-1.713796	-1.199459
/cut2		.0861512	.1298953			-.168439	.3407414

- 1.1 Erklären Sie ausführlich, warum die Schätzung keine Konstante enthält. Beschreiben Sie kurz die notwendige Änderung des Modells, wenn es mit Konstante geschätzt werden soll. (3 Punkte)
- 1.2 Erläutern Sie formal den Zusammenhang zwischen der beobachteten Zufriedenheit y_i und der latenten Zufriedenheit y_i^* . Definieren Sie die verwendeten Größen. (2 Punkte)
- 1.3 Interpretieren Sie den Koeffizienten von *educ* in Bezug auf die beobachtete Zufriedenheit statistisch und inhaltlich. (4 Punkte)
- 1.4 Bestimmen Sie das Alter, in dem die latente Zufriedenheit minimal ist. (2 Punkte)
- 1.5 Berechnen Sie die kompensierende Einkommensvariation für eine chronisch kranke Person. Interpretieren Sie das Ergebnis. (3 Punkte)
- 1.6 Definieren und erläutern Sie das McFadden R^2 . Berechnen Sie das Maß für die vorliegende Schätzung. (4 Punkte)
- 1.7 Sie könnten im vorliegenden Fall auch ein multinomiales Logit Modell schätzen. Gehen Sie auf einen Vorteil und einen Nachteil im Vergleich zum Ordered Probit Modell ein. (4 Punkte)

Aufgabe 2 (19 Punkte)

Die Kinderzahl deutscher Frauen (*kids*) ist die abhängige Variable einer Regressionsanalyse. 150 der 600 in der Stichprobe enthaltenen Frauen hatten bislang keine Kinder. Die erklärenden Variablen sind wie folgt definiert:

- age = Alter in Jahren
- educ = Bildung in Jahren
- rel_chr = 1, falls Zugehörigkeit zur christlichen Religion, 0 sonst (Referenz: konfessionslos)
- rel_oth = 1, falls Zugehörigkeit zu anderer Religion, 0 sonst (Referenz: konfessionslos).

Die Schätzergebnisse eines Tobit-Modells sind wie folgt:

Variable	Koeffizient	Std.fehler
age	0.160 **	(0.078)
educ	-0.550 ***	(0.134)
rel_chr	1.725 ***	(0.405)
rel_oth	2.260 ***	(0.408)
Konstante	9.083 ***	(2.659)
Skalierungsfaktor $\Phi(\mathbf{x}'\boldsymbol{\beta}/\sigma)$	0.234	
Anzahl Beobachtungen (N)	600	

Signifikanzniveau: * <0.1 , ** <0.05 , *** <0.01 .

- 2.1 Erläutern Sie knapp den Unterschied zwischen Stützung und Zensierung. (2 Punkte)
- 2.2 Nennen Sie eine Konsequenz für die Eigenschaften des Schätzers, wenn eine KQ-Schätzung mit gestutzten Daten durchgeführt wird. Berechnen Sie für das vorliegende Beispiel die Anzahl der Beobachtungen, die bei der gestutzten Regression verwendet wird. (2 Punkte)
- 2.3 Nennen Sie drei Arten marginaler Effekte, die auf Basis einer Tobit-Schätzung berechnet werden können. (1.5 Punkte)
- 2.4 Berechnen und interpretieren Sie den (unverzerrten) marginalen Effekt eines weiteren Bildungsjahres auf die Kinderzahl von allen Frauen (d. h. Frauen mit und ohne Kinder). (3 Punkte)
- 2.5 Das Tobit-Modell wurde unter Verwendung der folgenden Log-Likelihood-Funktion geschätzt. Erklären Sie kurz, was die beiden Summen $\sum_{y_i=0}$ bzw. $\sum_{y_i>0}$ jeweils berechnen. Welche Annahme hinsichtlich β_1 und β_2 wird im Tobit-Modell gemacht? (3 Punkte)

$$\ln L = \sum_{y_i=0} \ln \Phi \left(\frac{0 - \mathbf{x}'_i \boldsymbol{\beta}_1}{\sigma} \right) + \sum_{y_i>0} \ln \frac{1}{\sigma} \phi \left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}_2}{\sigma} \right)$$

- 2.6 Überprüfen Sie mit einem Wald-Test am 5%-Signifikanzniveau die Hypothese, dass die Kinderzahl der Frauen, die einer Religionsgemeinschaft zugehören (d.h. christliche oder andere Religion) von der Kinderzahl der konfessionslosen Frauen verschieden ist. Geben Sie Null- und Alternativhypothesen, Freiheitsgrade, den kritischen Wert und die Testentscheidung an. *Hinweise:* (1) Die Teststatistik des Wald-Tests lautet: $W = \hat{\boldsymbol{\beta}}' \hat{Var}(\hat{\boldsymbol{\beta}})^{-1} \hat{\boldsymbol{\beta}} \sim \chi_k^2$. (2) Verwenden Sie folgende geschätzte inverse Kovarianzmatrix des Schätzvektors $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{rel_chr}, \hat{\beta}_{rel_oth})'$:

$$\hat{Var}(\hat{\boldsymbol{\beta}})^{-1} = \begin{pmatrix} 5 & 0 \\ 0 & 100 \end{pmatrix}$$

(5.5 Punkte)

- 2.7 Nennen Sie zwei Vorteile des Heckman-Selektionsmodells gegenüber dem Tobitmodell. (2 Punkte)

Aufgabe 3 (12 Punkte)

Sie analysieren die Verweildauer der Patienten in einem Krankenhaus.

- 3.1 Definieren Sie den Begriff *duration dependence* am Beispiel des Krankenhausaufenthalts. Unterscheiden Sie dabei zwischen negativer bzw. positiver *duration dependence*. (3 Punkte)
- 3.2 Sie analysieren die Dauer des Krankenhausaufenthalts (gemessen in Tagen) mit einer Weibull-Regression. Als erklärende Variable nehmen Sie das Gewicht des Patienten in Kilogramm (*gewicht*) auf. Interpretieren Sie den Koeffizienten der Variable *gewicht* inhaltlich. (2 Punkte)

```
Weibull regression -- log relative-hazard form
```

No. of subjects =	1445	Number of obs =	1445
No. of failures =	552		
Time at risk =	80013		
Log likelihood =	-853.08318	LR chi2(1) =	165.48
		Prob > chi2 =	0.0000

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
gewicht	.0887867	.0134355	6.61	0.000	.0624535	.1151198
_cons	-3.402094	.3010177	-11.30	0.000	-3.992077	-2.81211
/ln_p	-.2158398	.0389149	-5.55	0.000	-.2921115	-.1395681
p	.8058644	.0313601			.7466852	.8697338
1/p	1.240904	.0482896			1.149777	1.339252

- 3.3 In einer weiteren Weibull-Schätzung wird das Alter des Patienten als zusätzliche erklärende Variable in das Modell aufgenommen. Die Schätzung ergibt einen Log-Likelihood-Wert von -850.61034. Überprüfen Sie, ob sich der Erklärungsgehalt des Modells signifikant verbessert hat. Geben Sie die Teststatistik, Freiheitsgrade und kritischen Wert zum Signifikanzniveau $\alpha = 0.01$ an. Berechnen Sie den empirischen Wert der Teststatistik und treffen Sie eine Testentscheidung. (5 Punkte)
- 3.4 Erklären Sie den Unterschied zwischen dem Weibull-Modell und dem Cox Proportional-Hazard-Modell. (2 Punkte)

Aufgabe 4 (7 Punkte)

Für eine Stichprobe von 11874 Studierenden wird die Arbeitsmarktpartizipation neben dem Studium mit einem linearen Wahrscheinlichkeitsmodell sowie mit Probit- und Logit-Modellen analysiert. Die binäre abhängige Variable ist $y=1$, wenn eine Person arbeitet, und $y=0$ sonst.

- 4.1 Nennen Sie zwei Nachteile eines linearen Wahrscheinlichkeitsmodells (2 Punkte)
- 4.2 Wie ändern sich die Schätzergebnisse für den Vektor β im linearen Wahrscheinlichkeitsmodell wenn die abhängige Variable nicht mit 0 und 1 sondern mit 1 und 2 kodiert ist? (2 Punkte)
- 4.3 Sie berücksichtigen zunächst als erklärende Variable das Alter der Studierenden und schätzen ein Logit-Modell (Logit 1). Danach nehmen Sie zusätzlich das Geschlecht ins Modell auf und schätzen erneut (Logit 2). Die folgende Tabelle vergleicht die Modellgüte beider Modelle. *Hinweis:* AIC bezeichnet das Akaike Informationskriterium und BIC das Schwarz Kriterium.

Model	Obs	df	AIC	BIC
Logit 1	11874	2	9710.496	9725.26
Logit 2	11874	3	9601.815	9623.962

Welches Modell ist zu bevorzugen? Begründen Sie. Erläutern Sie, welche Kriterien bei der Berechnung der beiden Gütemaße berücksichtigt werden. (3 Punkte)