

FRAGESTUNDE

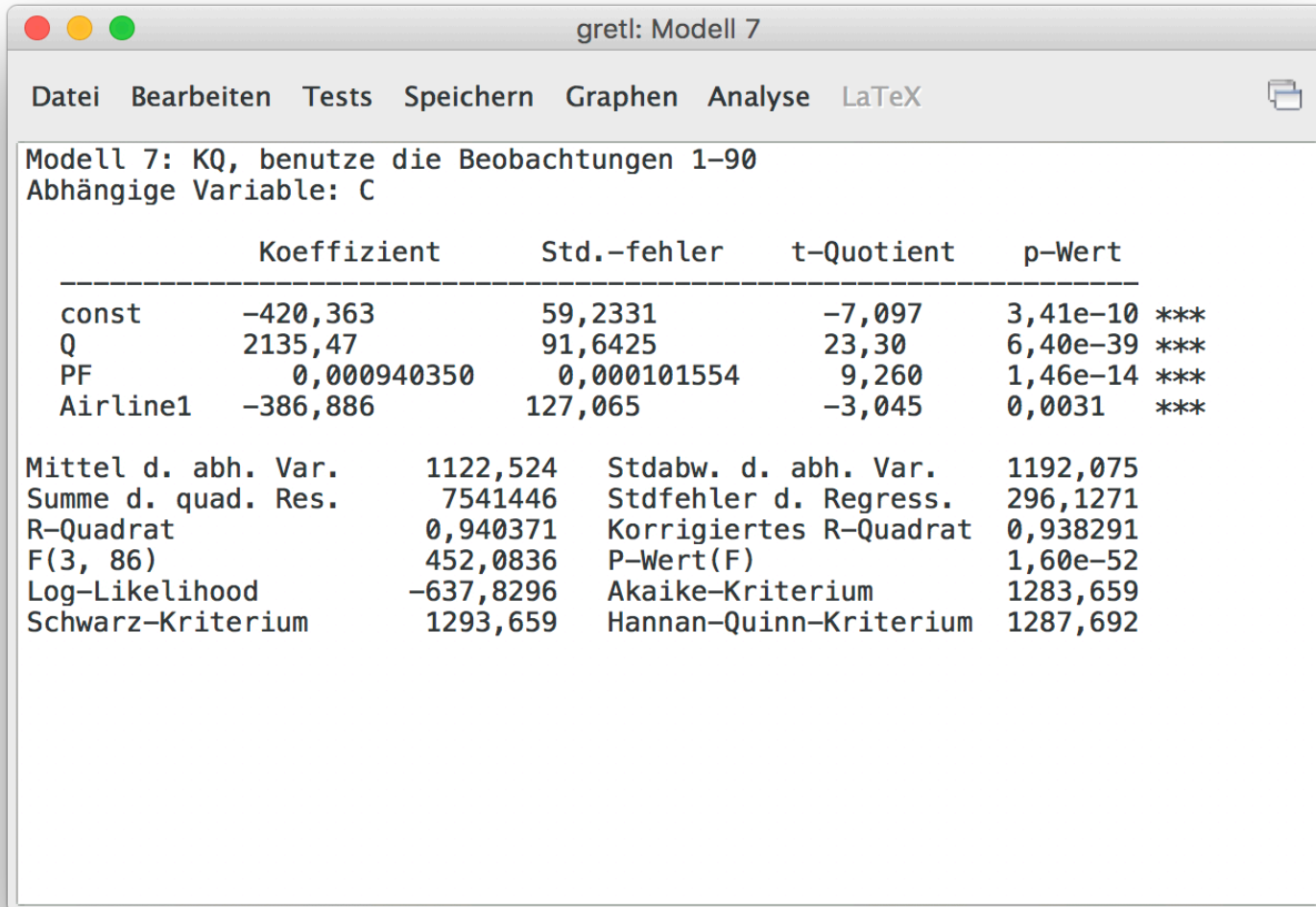
#2 - Allgemein

- **Q:** Müssen wir den Standard Error händisch berechnen können?
- **R:** Nein.

- **Q:** Hat das Monte Carlo Experiment irgendeine Bedeutung für uns im Hinblick auf die Klausur? Wenn ja, würde ich gerne noch einmal die Grundidee dahinter verstehen.
- **R:** Nein.

#1 – Interpretation Dummy

- **Q:** Bedeutet der negative Koeffizient der Dummy - Variablen in Anlage 4 der ersten Probeklausur, dass die „Airline 1“ um 386 günstiger ist?
- **R:** Ja.



gretl: Modell 7

Datei Bearbeiten Tests Speichern Graphen Analyse LaTeX

Modell 7: KQ, benutze die Beobachtungen 1-90
Abhängige Variable: C

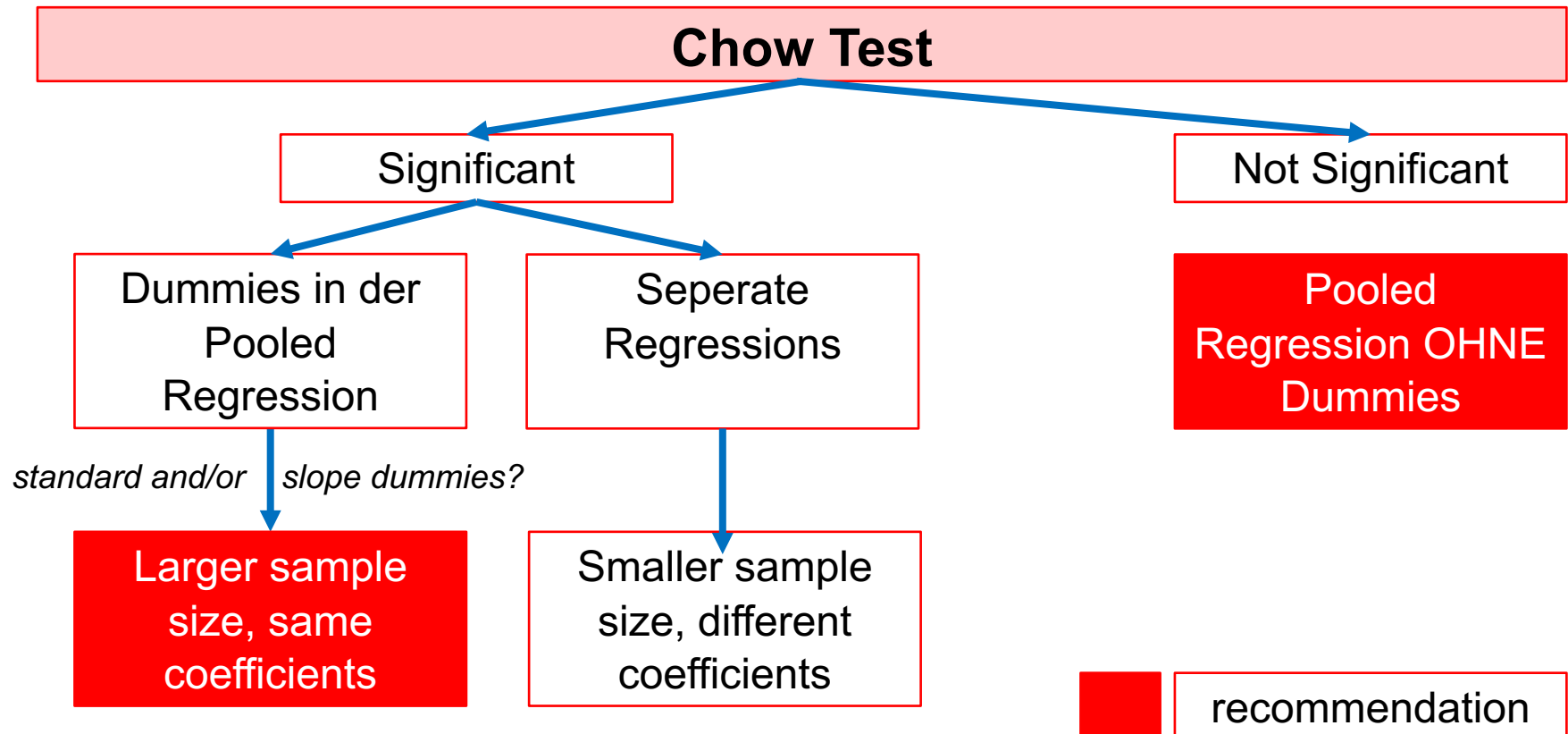
	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	-420,363	59,2331	-7,097	3,41e-10	***
Q	2135,47	91,6425	23,30	6,40e-39	***
PF	0,000940350	0,000101554	9,260	1,46e-14	***
Airline1	-386,886	127,065	-3,045	0,0031	***

Mittel d. abh. Var.	1122,524	Stdabw. d. abh. Var.	1192,075
Summe d. quad. Res.	7541446	Stdfehler d. Regress.	296,1271
R-Quadrat	0,940371	Korrigiertes R-Quadrat	0,938291
F(3, 86)	452,0836	P-Wert(F)	1,60e-52
Log-Likelihood	-637,8296	Akaike-Kriterium	1283,659
Schwarz-Kriterium	1293,659	Hannan-Quinn-Kriterium	1287,692

#2 – Chow Test

- **Q:** Was ist, wenn der Chow Test nicht signifikant ist? Dann macht man lieber die pooled Regression? Und wann nutzt man lieber die Dummy Variante oder sind der Chow Test und die Nutzung der Dummies subsidär?

- **R:**



#2 – Chow Test

- **Q:**
- (i) Im Skript 4 Folie 45 Slope Dummy: Wieso sagt der t-Wert hier etwas über die Höhe des Grenzprodukts aus? Sagt nicht immer der Koeffizient etwas über die Höhe aus?
- (ii) Auf der Folie 46 steht noch etwas zur 0-Hypothese: Heißt es im Ergebnis also, dass die Overhead costs von der occupational and regular School gleich sind? Das kann doch eigentlich nicht sein, das hatten wir doch bereits gesehen oder nicht?
- **R:** C.f. next slide.
 - (i) Der t-Wert sagt nie etwas über die ökonomische Signifikanz (hier Grenzprodukt) aus, immer nur über die statistische Signifikanz. Falls ich aber keine statistische Signifikanz habe, sollte ich auch nicht über die ökonomische Signifikanz sprechen.
 - (ii) Anwendung des “Standard” F-Tests, “check for improvement of including new variables”

SLOPE DUMMY VARIABLES

```
. reg COST N OCC NOCC
```

Source	SS	df	MS			
Model	1.0009e+12	3	3.3363e+11	Number of obs =	74	
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) =	49.64	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6803	
				Adj R-squared =	0.6666	
				Root MSE =	81980	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	152.2982	60.01932	2.537	0.013	32.59349	272.003
OCC	-3501.177	41085.46	-0.085	0.932	-85443.55	78441.19
NOCC	284.4786	75.63211	3.761	0.000	133.6351	435.3221
_cons	51475.25	31314.84	1.644	0.105	-10980.24	113930.7

- We can perform t tests as usual.
- The t statistic for the coefficient of *NOCC* is 3.76, so the marginal cost per student in an occupational school is significantly higher than that in a regular school.

SLOPE DUMMY VARIABLES

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	
Model	1.0009e+12	3	3.3363e+11	Number of obs = 74
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) = 49.64
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.6803
				Adj R-squared = 0.6666
				Root MSE = 81980

```
. reg COST N
```

Source	SS	df	MS	
Model	5.7974e+11	1	5.7974e+11	Number of obs = 74
Residual	8.9160e+11	72	1.2383e+10	F(1, 72) = 46.82
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.3940
				Adj R-squared = 0.3856
				Root MSE = 1.1e+05

- We can also perform an **F test** of the joint explanatory power of the dummy variables, comparing *RSS* when the dummy variables are included with *RSS* when they are not.
- The null hypothesis is that the coefficients of *OCC* and *NOCC* are both equal to 0. The alternative hypothesis is that one or both are nonzero.
- The improvement in the fit on adding the dummy variables is the reduction in *RSS*.

SLOPE DUMMY VARIABLES

```
. reg COST N OCC NOCC
```

Source	SS	df	MS
Model	1.0009e+12	3	3.3363e+11
Residual	4.7045e+11	70	6.7207e+09
Total	1.4713e+12	73	2.0155e+10

```
Number of obs =      74
F( 3,      70) =    49.64
Prob > F       =    0.0000
R-squared      =    0.6803
Adj R-squared  =    0.6666
```

$$F(2,70) = \frac{(8.92 \times 10^{11} - 4.70 \times 10^{11}) / 2}{4.70 \times 10^{11} / 70} = 31.4$$

```
. reg COST N
```

Source	SS	df	MS
Model	5.7974e+11	1	5.7974e+11
Residual	8.9160e+11	72	1.2383e+10
Total	1.4713e+12	73	2.0155e+10

```
Number of obs =      74
F( 1,      72) =    46.82
Prob > F       =    0.0000
R-squared      =    0.3940
Adj R-squared  =    0.3856
Root MSE      =    1.1e+05
```

- The improvement in the fit on adding the dummy variables is the reduction in *RSS*.
- The cost is 2 because 2 extra parameters, the coefficients of the dummy variables, have been estimated, and as a consequence the number of degrees of freedom remaining has been reduced from 72 to 70. This is 70 because there are 74 observations and 4 parameters have been estimated.

Remarks

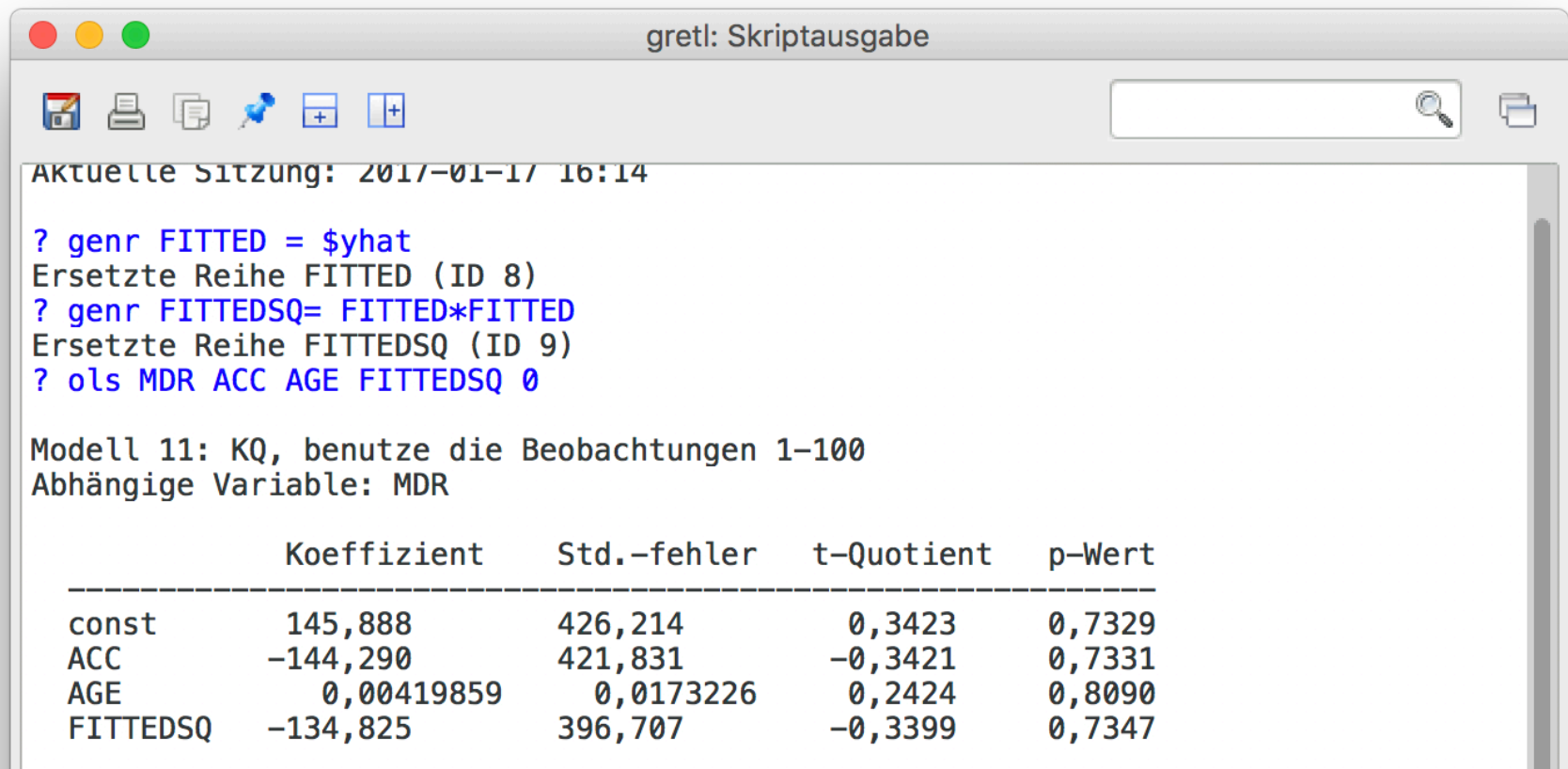
- Zähler
 - *Benefit*: RSS vorher versus neuer
 - *Costs*: decrease in DF: two additional variables, i.e. the two dummies
- Nenner: “Adjustment“: new RSS / new DF (=74-4)
- Even at 0.1 percent level significant (critical value = 7.6)

#2 – Chow Test

- **Q:** Wieso spricht man beim Chow Test bei "k" von Kosten? Vielleicht für mich noch einmal eine Erklärung der Bedeutung!
- **R:**
- Jede zusätzliche Variable ("k") reduziert die dfs und damit "Information" bzw. macht es schwieriger meine Regression durch die Datenpunkte zu legen.
- Bei den separated Regressions muss ich separat meine Betas bestimmen, daher brauche ich doppelt so viele "k"s. Somit "verliere" ich "k" Informationen.

#1 - Ramsey

- **Q:** Ramsey Test: Wird bei der Betrachtung des Signifikanzniveaus nur β oder $\beta \cdot \text{RSS}$ angesehen?
- **R:** Nur "FITTED^2" und die Signifikanz des Koeffizienten beachten!



```
gret!: Skriptausgabe
AKTUELLE SITZUNG: 2017-01-17 16:14
? genr FITTED = $yhat
Ersetzte Reihe FITTED (ID 8)
? genr FITTEDSQ= FITTED*FITTED
Ersetzte Reihe FITTEDSQ (ID 9)
? ols MDR ACC AGE FITTEDSQ 0

Modell 11: KQ, benutze die Beobachtungen 1-100
Abhängige Variable: MDR
```

	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	145,888	426,214	0,3423	0,7329
ACC	-144,290	421,831	-0,3421	0,7331
AGE	0,00419859	0,0173226	0,2424	0,8090
FITTEDSQ	-134,825	396,707	-0,3399	0,7347

#1 - Logit

- **Q:** Können wir bitte noch einmal auf das Logit Modell der zweiten Probeklausur eingehen? Insb. auf $P=F(z)$. Wie werden die 35,5% berechnet?
- **R:** C.f. next slide.

gretl: Skriptausgabe

gretl-Version 2016c
 Aktuelle Sitzung: 2017-01-18 15:59

? logit OWNRENT INCOME 0

Modell 4: Logit, benutze die Beobachtungen 1-100
 Abhängige Variable: OWNRENT
 Standardfehler basieren auf Hessematrix

	Koeffizient	Std.-fehler	z	Steigung
const	-2,30079	0,598817	-3,842	
INCOME	0,505045	0,164518	3,070	0,115575

Mittel d. abh. Var.	0,360000	Stdabw. d. abh. Var.	0,482418
McFadden-R-Quadrat	0,095499	Korrigiertes R-Quadrat	0,064890
Log-Likelihood	-59,10177	Akaike-Kriterium	122,2035
Schwarz-Kriterium	127,4139	Hannan-Quinn-Kriterium	124,3123

Zahl der 'korrekt vorhergesagten' Fälle = 72 (72,0%)
 f(beta'x) bei Mitteln der unabhängigen Var's = 0,229
 Likelihood-Quotienten-Test: Chi-Quadrat(1) = 12,4801 [0,0004]

		Prognostiziert	
		0	1
Tatsächlich	0	60	4
	1	24	12

Probit Model Approach	
Estimation parameters	
b0	-2,3008
b1	0,5050
X	3,3693
Z	-0,5991
exp(-Z)	1,8206
f	0,2288
p=F(Z)	35,5%
marginal Änderung	11,6%

#1 - Logit

- **R:**

$$Z = \beta_1 + \beta_2 X$$

- $Z = \beta_0 + \beta_1 * X = -2,3 + 0,5 * 3,4 = -0,6$
- -> "Schätzwert wie bei normaler Regression"

$$p = F(Z) = \frac{1}{1 + e^{-Z}}$$

- $\exp(-Z) = 1,8$
- $p = 35,5\%$
- Erwartungswert, dass ich ein Haus besitze ist damit 35,5%, gegeben Mittelwert des Einkommens (=USD 34,000 Jahreseinkommen).

#1 - Logit

- Nehmen wir an, mein Einkommen steigt um 1 Punkt, also USD 10,000. Dann ist der marginale Effekt + 11,6%. Dies ist $\beta_2 * f(X) = 0,51 * 0,23 = 11,6\%$.

```
gretl: Skriptausgabe

gretl-Version 2016c
Aktuelle Sitzung: 2017-01-18 15:59

? logit OWNRENT INCOME 0

Modell 4: Logit, benutze die Beobachtungen 1-100
Abhängige Variable: OWNRENT
Standardfehler basieren auf Hessematrix
```

	Koeffizient	Std.-fehler	z	Steigung
const	-2,30079	0,598817	-3,842	
INCOME	0,505045	0,164518	3,070	0,115575

```
Mittel d. abh. Var.      0,360000  Stdabw. d. abh. Var.      0,482418
McFadden-R-Quadrat      0,095499  Korrigiertes R-Quadrat    0,064890
Log-Likelihood           -59,10177 Akaike-Kriterium          122,2035
Schwarz-Kriterium        127,4139  Hannan-Quinn-Kriterium    124,3123

Zahl der 'korrekt vorhergesagten' Fälle = 72 (72,0%)
f(beta'x) bei Mitteln der unabhängigen Var's = 0,229
```

#1 - Logit

- Über den marginalen Effekt gerechnet wäre mein Wahrscheinlichkeit ein Haus zu besitzen: $35,5\% + 11,6\% = 47,1\%$.
- Alternativ kann ich aber auch wieder Z ausrechnen...

	A	B	C	D	E	F
1	Logit Model Approach					
2	Estimation parameters					
3	b0	-2,3008				
4	b1	0,5050				
5	X	4,4000				
6	Z	-0,0786				
7						
8	exp(-Z)	1,0818				
9	f	0,2496				
10	p=F(Z)	48,0%				
11	marginal Änderung	12,6%				
12						