



## Übung zur Empirischen Wirtschaftsforschung

### VI Einkommensfunktion – Teil 1

*Querschnittsdaten*

*Dummyvariablen*

*Humankapitaltheorie*

*Individualdaten des Sozio-ökonomischen Panel (SOEP)*

*Schätzung von Einkommensfunktionen*

#### Literatur

*Winker, P. (2007), Empirische Wirtschaftsforschung, Kap. 9.*

*Franz, W. (2003), Arbeitsmarktökonomik, 5. Auflage, Kap. 3.*

*Hübler, O. (2005), Einführung in die empirische Wirtschaftsforschung, Kap. 2 & 3.*

*Smolny, W., and Kirbach, M. (2004): Wage differentials between East and West Germany - Is it related to the location or to the people?*

## Arten von Daten

### Querschnittsdaten - Zeitreihendaten

**Zeitreihendaten** (time series) sind Daten, die über die Zeit wiederholt erhoben werden, i.d.R. mit einer regelmäßigen Frequenz. Die Daten werden nach Möglichkeit bei denselben Merkmalsträgern und mit gleicher Methodik und Definition erhoben. Bei Zeitreihendaten liegt zu einer Variable ein Beobachtungswert pro Periode vor, z.B. der:

- aggregierte Konsum pro Quartal von 1960-2009,
- kurzfristige Dreimonatszinssatz pro Monat von 1975-2009,
- DAX-Wert pro Tag von 1993-2009,
- ...

In der Gleichung

$$y_t = \beta_0 + \beta_1 \cdot x_t + \varepsilon_t$$

ist  $t$  der Zeitindex.

**Querschnittsdaten** (cross-section) beziehen sich auf zeitlich ungeordnete Beobachtungen für unterschiedliche Merkmalsträger, die meist zu einem einheitlichen Zeitpunkt erhoben werden. Es liegen somit  $i$  Beobachtungen zu einem Zeitpunkt vor, z.B.:

- die Mietkosten verschiedener Wohnungen in einer Stadt am 1. Juni 2009,
- das Einkommen von Arbeitnehmern im Mai 2008,
- der Exportanteil von Unternehmen im Jahr 2000,
- ...

In der Gleichung

$$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$$

ist  $i$  kein Zeitindex, sondern die Nummer des  $i$ -ten Mieters, Arbeitnehmers bzw. der  $i$ -ten Firma.

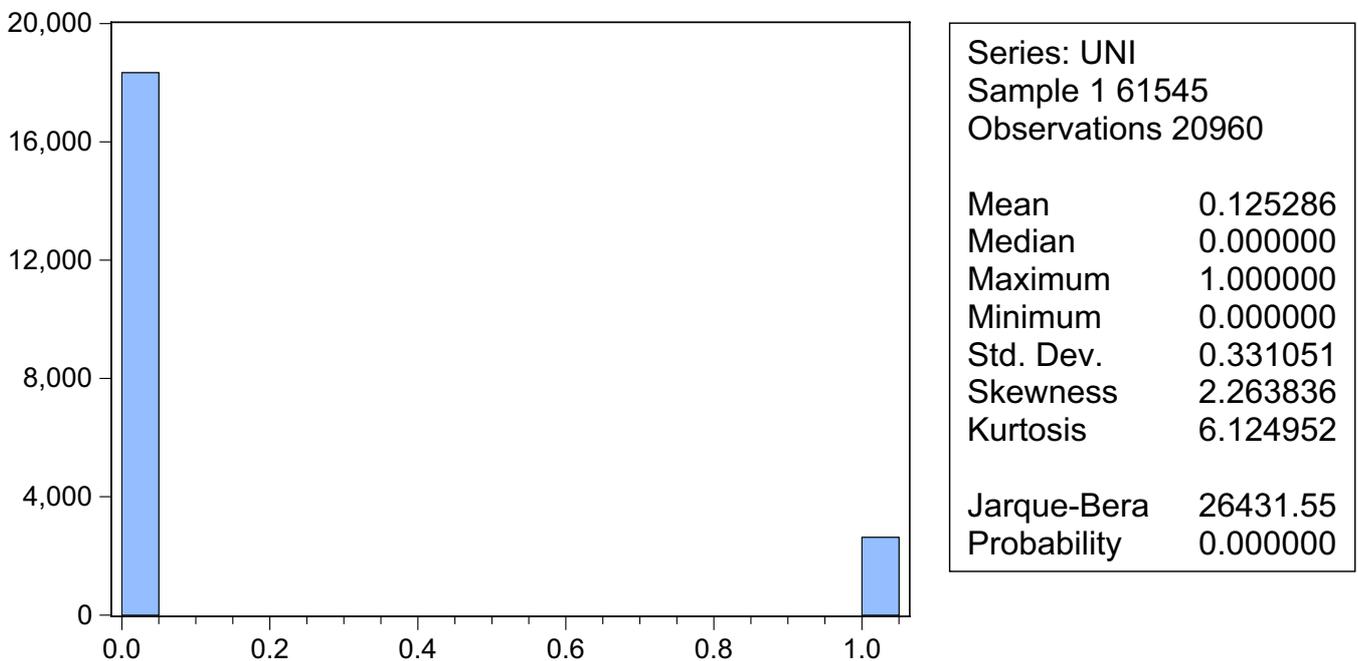
Querschnittsanalysen können ebenso wie Zeitreihenuntersuchungen in EViews durchgeführt werden. Hierzu wird ein Workfile mit *nicht datierten* Daten benötigt. Der Befehl "create u 1 32000" erzeugt hierbei ein Workfile mit 32.000 Beobachtungen.

Querschnittsdaten haben in der Regel eine höhere Varianz als Zeitreihendaten. Daher ist ein geringerer Wert für das Bestimmtheitsmaß  $R^2$  zu erwarten.

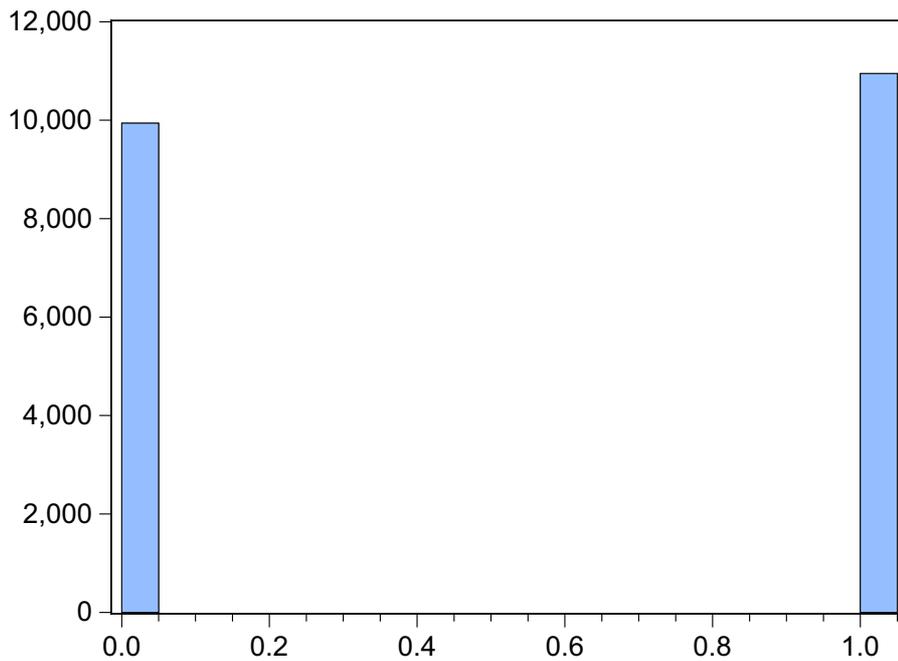
## Der Einsatz von Dummyvariablen

Dummyvariablen sind keine quantitativen, sondern qualitative Variablen. Sie nehmen nur 2 Ausprägungen an, z.B.:

- @seas(1) ist 1 im ersten Quartal, 0 in den anderen drei Quartalen.
- Die Variable UNI hat die Ausprägungen 1 für einen Universitätsabschluss und 0 für keinen Universitätsabschluss.
- Die Variable SEX hat die Ausprägungen 1 für Frau und 0 für Mann.
- ...



- Der Maximalwert ist 1, d.h. ein Universitätsabschluss.
- Der Minimalwert ist 0, d.h. kein Universitätsabschluss.
- Der Median ist 0, d.h. es haben mehr Personen keinen Universitätsabschluss.
- Das arithmetische Mittel beträgt 0.125, d.h. 12,5% der 20.960 Personen haben einen Universitätsabschluss.



Series: SEX	
Sample 1 61545	
Observations 20886	
Mean	0.524083
Median	1.000000
Maximum	1.000000
Minimum	0.000000
Std. Dev.	0.499432
Skewness	-0.096444
Kurtosis	1.009302
Jarque-Bera	3481.075
Probability	0.000000

- Der Maximalwert ist 1, d.h. die Person ist weiblich.
- Der Minimalwert ist 0, d.h. die Person ist männlich.
- Der Median ist 1, d.h. unter den beobachteten Personen befinden sich mehr Frauen.
- Das arithmetische Mittel ist 0.524, d.h. 52,4% der 20.886 Personen sind Frauen.

# Ökonomische Theorie

## Der Einfluß des Humankapitals auf das Einkommen

Nach der neoklassischen Theorie entspricht der Lohn im Gleichgewicht dem Grenzprodukt der Arbeit. Im Rahmen der Humankapitaltheorie (als Erweiterung des neoklassischen Modells) richtet sich der Lohn nach der Produktivität und dem Humankapitalbestand, d.h. Individuen investieren in ihr Humankapital, um durch eine höhere Produktivität einen höheren Reallohn zu erhalten.

Die Akkumulation des Humankapitals eines Erwerbstätigen bestimmt sich analog zu Sachkapitalinvestitionen. Formal lässt sich das mittels folgender Gleichung darstellen:

$$HK_t = HK_{t-1} + I_{HK} - \delta \cdot HK .$$

Der Humankapitalbestand einer Periode  $t$  ergibt sich aus dem Humankapitalbestand der Vorperiode ergänzt um Investitionen in das Humankapital  $I_{HK}$  und korrigiert um Abschreibungen  $\delta$  auf Humankapital. Investitionen werden durchgeführt, falls der Ertrag höher ist als die anfallenden Kosten:

$$\sum_{t=1}^T = \frac{w_t^q - w_t}{(1+r)^t} - C_0 > 0 .$$

Der Ertrag der Investition entspricht der Summe der Differenzen des Reallohns  $w^q$  abzüglich des Reallohns  $w$ , den das Individuum ohne die Investition in Humankapital realisiert hätte bis zur Endperiode  $T$ , in der ein Erwerbstätiger aus dem Erwerbsleben ausscheidet. Der Ertrag jeder Periode wird um einen Abzinsungsfaktor  $r$  korrigiert. In der Periode, in der die Fort- und Weiterbildung durchgeführt wird, entstehen Kosten  $C_0$ .

Schul-, Aus-, Fort- und Weiterbildung sind Investitionen in das Humankapital, von denen Wirtschaftssubjekte zukünftige Erträge erwarten. Wenn die Summe der zukünftigen Erträge höher ist als die anfallenden Kosten, werden diese Investitionen durchgeführt, andernfalls wird, zumindest aus finanzieller Sicht, auf sie verzichtet. Die Verteilung der Arbeitseinkommen richtet sich demzufolge nach der Investition in (Aus-) Bildung, dem Humankapitalbestand. Personen mit einem höheren Schul- und Berufsabschluß sollten somit ein größeres Einkommen haben als Personen mit einer niedrigen Schul- bzw. Berufsausbildung. Einkommensunterschiede werden demzufolge durch die Dauer und die Kosten der Ausbildung bestimmt.

Im Folgenden soll anhand eines empirischen Modells untersucht werden, inwiefern Investitionen in Humankapital, z.B. in Form unterschiedlicher Schul- und Berufsabschlüsse zu einem höheren Einkommen führen. Als Datengrundlage dienen Zahlen aus dem Sozio-ökonomischen Panel (SOEP).

## Datengrundlage

### Das SOEP

Das sozio-ökonomische Panel (SOEP) ist eine repräsentative Wiederholungsbefragung privater Haushalte in Deutschland, die seit 1983 im jährlichen Rythmus bei den selben Personen und Familien vom DIW Berlin und Infratest durchgeführt wird. Sie dient der Untersuchung des Wandels objektiver Lebensbedingungen und der subjektiv wahrgenommenen Lebensqualität und liefert dabei u.a. Informationen über:

- Erwerbs- und Familienbiographien
- Erwerbsbeteiligung und berufliche Mobilität
- Einkommensverläufe
- Gesundheit
- Lebenszufriedenheit
- gesellschaftliche Partizipation und Zeitverwendung
- Persönlichkeitsmerkmale
- Haushaltszusammensetzung, Wohnsituation

Dabei enthält es unterschiedliche Stichproben:

- A Deutsche (West)
- B Ausländer (West)
- C Deutsche (Ost)
- D Zuwanderer 1984-93
- E Ergänzung 1998
- F Ergänzung 2000
- G Hocheinkommensbezieher

Auf der Website <http://panel.gsoep.de/soepinfo> kann man z. B. die entsprechenden Fragebögen einsehen. Im Workfile `hk07.wf1` sind ausgewählte Variablen aus der aktuellen Welle von 2007 enthalten.

Daten im workfile HK07.wf1

*EINK* Bruttomonatseinkommen in €

*ALTER* Alter in Jahren

*SEX* Geschlecht, Dummyvariable, 1 für Frauen, 0 für Männer

*OST* Ostdeutschland

*WEST* Westdeutschland

*OHNE* kein Schulabschluss

*HAUPT* Hauptschulabschluss

*REAL* Realschulabschluss

*FACH* Fachschule

*ABI* Abitur

*KEIN* ohne Berufsabschluss

*LEHRE* Lehre

*MEIST* Meister

*FH* Fachhochschulabschluss

*UNI* Universitätsabschluss

*TENUB* Betriebszugehörigkeit in Jahren

*EX* Berufserfahrung

*DAUER* Dauer der Ausbildung in Jahren

*STUND* durchschnittliche wöchentliche Arbeitszeit

*PSAMPLE* Stichprobenart (1: Deutsche (West), 2: Ausländer (West),  
3: Deutsche (Ost), 4: Zuwanderer 1984-93, 5: Ergänzung 1998,  
6: Ergänzung 2000, 7: Hocheinkommensbezieher 2002)

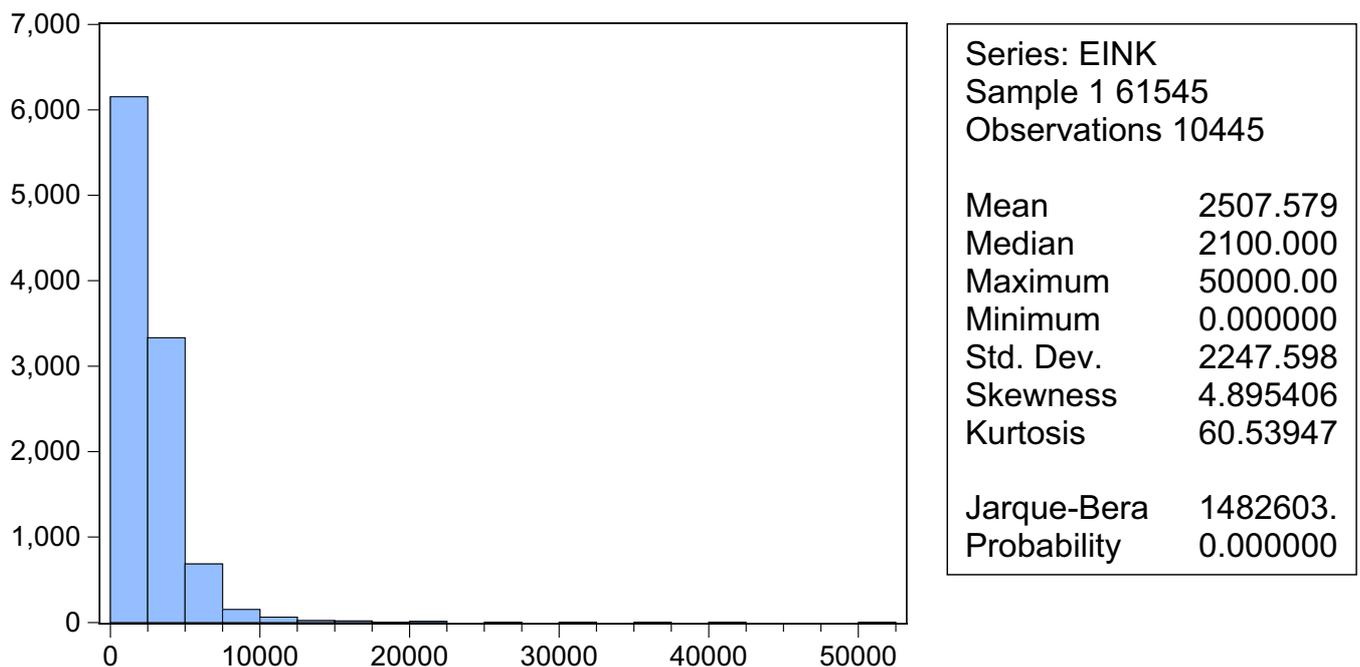
Quelle: Ausgewählte Daten des SOEP für das Jahr 2007.

## Schätzung von Einkommensfunktionen

### Das additive Modell

Im einfachen additiven empirischen Modell wird die endogene Variable *Einkommen* durch verschiedene exogene Variablen erklärt:

$$Eink = \beta_0 + \beta_1 \cdot Schule + \beta_2 \cdot Beruf + \beta_3 \cdot Berufserfahrung + \beta_4 \cdot Geschlecht + \varepsilon .$$



- Im Beispiel liegen 10.445 Beobachtungen vor.
- Der Wert des arithmetischen Mittels ist 2.507,58 €.
- Die Verteilung ist links-steil bzw. rechts-schief, der Median ist kleiner als das arithmetische Mittel.
- Die Standardabweichung beträgt 2.247,60 €.

In einer ersten Schätzung wird das Einkommen durch den Schulabschluss erklärt:

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 12:20
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4502
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1939.922	102.4056	18.94352	0.0000
HAUPT	70.75306	113.4231	0.623798	0.5328
REAL	98.33049	109.4647	0.898285	0.3691
FACH	1172.601	144.5424	8.112509	0.0000
ABI	1133.959	114.0617	9.941630	0.0000

```

=====
R-squared          0.080265      Mean dependent var 2332.069
Adjusted R-squared 0.079447      S.D. dependent var 1707.730
S.E. of regression 1638.489      Akaike info criteri17.64205
Sum squared resid  1.21E+10     Schwarz criterion  17.64917
Log likelihood     -39707.25      Hannan-Quinn criter17.64456
F-statistic        98.11337      Durbin-Watson stat 1.692184
Prob(F-statistic)  0.000000
=====

```

- HAUPT, REAL, FACH und ABI sind Dummyvariablen für einen Hauptschulabschluss, Realschulabschluss, Fachabitur bzw. Abitur.
- Personen ohne Schulabschluss werden nicht als Dummyvariable aufgenommen, da nur  $i-1$  Dummyvariablen bei  $i$  Ausprägungen modelliert werden dürfen. Personen ohne Schulabschluss haben ein Einkommen von 1.939,92 € (Koeffizient der Konstanten C).
- Personen mit Abitur haben ein Einkommen von  $1.939,92 + 1.133,96 = 3.073,88$  €.
- Die Variablen für einen Haupt- und Realschulabschluss sind nicht signifikant.
- Der Erklärungsgehalt des Modells ist mit ca. 8% gering.

**Prüfung der Schätzergebnisse:**

- erwartetes Vorzeichen?
- Größenordnung?
- ökonomische Plausibilität?

Die Erklärung des Einkommens durch den Berufsabschluss liefert folgendes Ergebnis:

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 12:25
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4644
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1768.626	39.12975	45.19901	0.0000
LEHRE	308.2295	50.30387	6.127351	0.0000
MEIST	751.8262	91.45912	8.220352	0.0000
FH	984.7375	81.41067	12.09593	0.0000
UNI	1975.190	74.03704	26.67840	0.0000

```

=====
R-squared          0.148540      Mean dependent var 2298.771
Adjusted R-squared 0.147806      S.D. dependent var 1701.053
S.E. of regression 1570.316      Akaike info criteri17.55702
Sum squared resid  1.14E+10      Schwarz criterion  17.56395
Log likelihood     -40762.39      Hannan-Quinn criter17.55946
F-statistic        202.3222      Durbin-Watson stat 1.684968
Prob(F-statistic)  0.000000
=====

```

- LEHRE, MEIST, FH und UNI sind Dummyvariablen für Lehre, Meister, Fachhochschul- bzw. Universitätsabschluss.
- Personen ohne Berufsabschluss haben ein Einkommen von 1.768,63 €.
- Personen mit Universitätsabschluss erhalten ein Einkommen in Höhe von  $1.768,63 + 1.975,19 = 3.743,82$  €.
- Alle Variablen sind signifikant.
- Der Erklärungsgehalt des Modells ist deutlich gestiegen und beträgt nun ca. 14,9%.

Die Erklärung des Einkommens durch den Schul- und Berufsabschluss ergibt:

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 12:28
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4492
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1672.716	100.5809	16.63054	0.0000
HAUPT	119.8877	110.2409	1.087506	0.2769
REAL	36.59987	106.7441	0.342875	0.7317
FACH	879.4480	141.6157	6.210103	0.0000
ABI	337.6250	117.2384	2.879817	0.0040
LEHRE	301.7237	52.39654	5.758466	0.0000
MEIST	713.0575	92.97069	7.669702	0.0000
FH	770.5994	87.49305	8.807550	0.0000
UNI	1753.281	92.74539	18.90423	0.0000

```

=====
R-squared           0.156608   Mean dependent var 2334.580
Adjusted R-squared  0.155103   S.D. dependent var 1708.385
S.E. of regression  1570.318   Akaike info criteri17.55795
Sum squared resid   1.11E+10   Schwarz criterion  17.57079
Log likelihood      -39426.14   Hannan-Quinn criter17.56247
F-statistic         104.0552   Durbin-Watson stat 1.684046
Prob(F-statistic)   0.000000
=====

```

- Personen ohne Schul- oder Berufsabschluss haben ein Einkommen von 1.672,72 € (der Koeffizient der Konstanten C).
- Personen mit Abitur und Universitätsabschluss haben ein Einkommen von  $1.672,72 + 337,63 + 1.753,28 = 3.763,63$  €.
- Die Variablen FACH, ABI, LEHRE, MEIST, FH und UNI sind signifikant, während die Koeffizienten HAUPT und REAL insignifikant sind.
- Der Erklärungsgehalt des Modells steigt im Vergleich zum vorherigen Modell nur leicht auf ca. 15,7%.

Zur Erklärung des Einkommens kommen neben Schul- und Berufsabschluss weitere Variablen in Betracht:

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 12:35
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4484
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1617.526	111.0702	14.56310	0.0000
HAUPT	90.52650	102.3509	0.884472	0.3765
REAL	276.8814	99.99983	2.768819	0.0056
FACH	1008.123	132.3892	7.614839	0.0000
ABI	714.8310	112.2193	6.369949	0.0000
LEHRE	301.7840	48.07352	6.277551	0.0000
MEIST	459.4694	85.76351	5.357400	0.0000
FH	719.0184	80.91202	8.886422	0.0000
UNI	1548.399	85.81840	18.04274	0.0000
SEX	-1191.428	43.63574	-27.30396	0.0000
EX	20.70335	2.013341	10.28308	0.0000

```

=====
R-squared          0.290965      Mean dependent var 2338.168
Adjusted R-squared 0.289380      S.D. dependent var 1707.782
S.E. of regression 1439.631      Akaike info criteri17.38461
Sum squared resid  9.27E+09      Schwarz criterion  17.40033
Log likelihood     -38965.30      Hannan-Quinn criter17.39015
F-statistic        183.5572      Durbin-Watson stat 1.551809
Prob(F-statistic) 0.000000
=====

```

- SEX ist eine Dummyvariable, die den Wert 1 für Frauen und den Wert 0 bei Männern annimmt. Frauen verdienen bei sonst gleichen Voraussetzungen 1.191,43 € weniger als Männer. Eine Aussage darüber, ob geringeres Humankapital (Babypause) oder Diskriminierung der Grund für den geringeren Verdienst ist, kann aufgrund dieser Schätzung nicht gemacht werden.
- EX ist eine Variable für die Berufserfahrung: Ein Jahr mehr Berufserfahrung erhöht den Verdienst um 20,70 €.
- Alle Variablen bis auf die Variable Hauptschulabschluss sind signifikant. Der Erklärungsgehalt des Modells steigt im Vergleich zum vorherigen Modell auf ca. 29%.