

Universität Stuttgart

Fachbereich 10 – Wirtschafts- und Sozialwissenschaften

Institut für Sozialwissenschaften, Abteilung IV

Thomas Krause, M.A.

Statistische Modellbildung 2

Wintersemester 2018/19

## Übungsaufgabe 1

01.11.18

Anna-Maria Enders

Schurwaldstr. 69

70186 Stuttgart

annaenders95@aol.com

Matrikelnummer 3322963

Master of Arts Empirische Politik- und Sozialforschung

3. Fachsemester

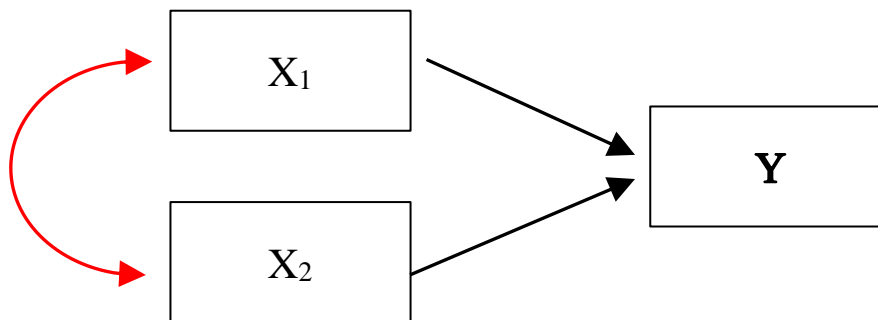
## **Inhaltsverzeichnis**

1.	Aufgabe .....	1
2.	Aufgabe .....	3
2.1	*ALTER .....	3
2.2	*BILDUNG .....	6
2.3	*GESCHLECHT .....	8
2.4	*EINKOMMEN UMBENENNEN .....	10
3.	Aufgabe .....	13
3.1	Modell a: Einkommen auf Alter .....	13
3.2	Modell b: Einkommen auf Bildung .....	16
3.3	Modell c: Einkommen auf Geschlecht .....	19
3.4	Modell ab: Einkommen auf Alter, Bildung und Geschlecht .....	22
3.5	Modell abc: Einkommen auf Alter, Bildung und Geschlecht .....	26
4.	Literaturverzeichnis .....	31

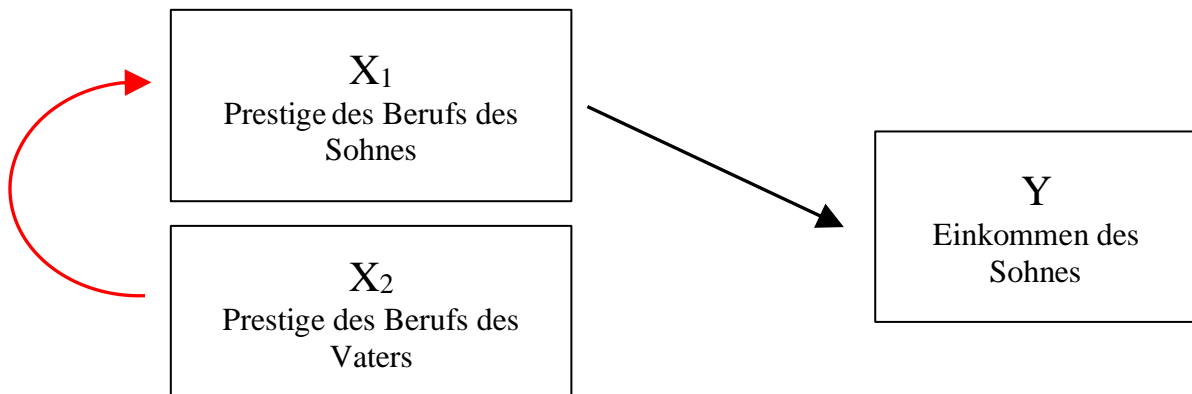
## 1. Aufgabe

1a. Was ist unter Auspartialisierung zu verstehen und wieso ist es aufgrund der beteiligten Mechanismen wichtig, immer mehrere Prädiktorvariablen zu berücksichtigen, auch wenn diese ggf. keinen Einfluss auf die abhängige Variable haben?

Im Modell einer Regression von  $Y$  auf die Variablen  $X_1$  und  $X_2$  wird der Einfluss der **X-Variablen untereinander** bereinigt.



Mithilfe eines Fallbeispiels kann dargestellt werden, warum mehrere Prädiktorvariablen berücksichtigt werden müssen, selbst wenn diese keinen Einfluss auf die abhängige Variable haben.



Das Prestige des Berufs des Vaters hat zwar keinen direkten Einfluss auf das Einkommen des Sohnes, es beeinflusst aber das Prestige des Berufs des Sohnes. Damit dieser Effekt bei der Regression von  $Y$ , dem Einkommen des Sohnes, auf  $X$ , das Prestige seines Berufes, ausgeschlossen wird, wird eine Auspartialisierung vorgenommen. Somit kann sichergestellt werden, dass die Varianz von  $Y$  nur auf die Teile von  $X_1$  zurückgeführt werden, die nicht durch  $X_2$  beeinflusst wurden (vgl. Urban & Mayerl, 2011, S. 80-86).

*1b. Wieso können unabhängige Variablen im multiplen Regressionsmodell einen Einfluss auf Y haben, obwohl die bivariate Korrelation zwischen ihnen und Y nicht signifikant ist?*

Unabhängige Variablen können einen Einfluss auf Y haben, indem sie andere unabhängige Variablen beeinflussen, die für das Modell verwendet wurden, jedoch selbst keinen eigenen signifikanten Einfluss auf Y ausüben (siehe oben). Durch diese Beeinflussung üben sie über diese weitere unabhängige Variable einen Einfluss auf Y aus.

## 2. Aufgabe

### 1. Schritt: Variablen auswählen

Alter: v84

Schulabschluss: v86

Geschlecht: v81

Einkommen: einkommen/v420

### 2. Schritt: Recodieren

#### 2.1 \*ALTER

compute alter0= v84-18.

variable labels alter0 "Alter mit 18 als Nullpunkt".

FREQUENCIES VARIABLES=alter0

/STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS  
KURTOSIS.

### Häufigkeiten

Anmerkungen		
Ausgabe erstellt		01-NOV-2018 14:34:39
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition von fehlenden Werten	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.

Verwendete Fälle		Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax		FREQUENCIES VARIABLES=alter0 /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS KURTOSIS.
Ressourcen	Prozessorzeit	00:00:00,05
	Verstrichene Zeit	00:00:00,00

## Statistiken

Alter mit 18 als Nullpunkt

N	Gültig	3468
	Fehlend	3
Mittelwert		31,4403
Median		32,0000
Standardabweichung		17,50623
Schiefe		,062
Kurtosis		-,892
Minimum		,00
Maximum		73,00

## Alter mit 18 als Nullpunkt

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	,00	29	,8	,8	,8
	1,00	54	1,6	1,6	2,4
	2,00	42	1,2	1,2	3,6
	3,00	56	1,6	1,6	5,2
	4,00	41	1,2	1,2	6,4
	5,00	49	1,4	1,4	7,8
	6,00	46	1,3	1,3	9,1
	7,00	58	1,7	1,7	10,8
	8,00	52	1,5	1,5	12,3
	9,00	55	1,6	1,6	13,9
	10,00	49	1,4	1,4	15,3
	11,00	52	1,5	1,5	16,8
	12,00	55	1,6	1,6	18,4

13,00	54	1,6	1,6	20,0
14,00	52	1,5	1,5	21,5
15,00	43	1,2	1,2	22,7
16,00	53	1,5	1,5	24,2
17,00	62	1,8	1,8	26,0
18,00	48	1,4	1,4	27,4
19,00	33	1,0	1,0	28,3
20,00	50	1,4	1,4	29,8
21,00	46	1,3	1,3	31,1
22,00	38	1,1	1,1	32,2
23,00	50	1,4	1,4	33,7
24,00	43	1,2	1,2	34,9
25,00	68	2,0	2,0	36,9
26,00	57	1,6	1,6	38,5
27,00	74	2,1	2,1	40,6
28,00	75	2,2	2,2	42,8
29,00	69	2,0	2,0	44,8
30,00	92	2,7	2,7	47,4
31,00	82	2,4	2,4	49,8
32,00	70	2,0	2,0	51,8
33,00	85	2,4	2,5	54,3
34,00	81	2,3	2,3	56,6
35,00	74	2,1	2,1	58,7
36,00	83	2,4	2,4	61,1
37,00	68	2,0	2,0	63,1
38,00	61	1,8	1,8	64,9
39,00	56	1,6	1,6	66,5
40,00	60	1,7	1,7	68,2
41,00	53	1,5	1,5	69,7
42,00	64	1,8	1,8	71,6
43,00	66	1,9	1,9	73,5
44,00	62	1,8	1,8	75,3
45,00	40	1,2	1,2	76,4
46,00	61	1,8	1,8	78,2
47,00	49	1,4	1,4	79,6
48,00	46	1,3	1,3	80,9
49,00	40	1,2	1,2	82,1
50,00	32	,9	,9	83,0
51,00	40	1,2	1,2	84,1
52,00	48	1,4	1,4	85,5
53,00	51	1,5	1,5	87,0

	54,00	43	1,2	1,2	88,2
	55,00	60	1,7	1,7	90,0
	56,00	47	1,4	1,4	91,3
	57,00	42	1,2	1,2	92,5
	58,00	32	,9	,9	93,5
	59,00	34	1,0	1,0	94,4
	60,00	26	,7	,7	95,2
	61,00	35	1,0	1,0	96,2
	62,00	21	,6	,6	96,8
	63,00	22	,6	,6	97,4
	64,00	19	,5	,5	98,0
	65,00	16	,5	,5	98,4
	66,00	11	,3	,3	98,8
	67,00	13	,4	,4	99,1
	68,00	9	,3	,3	99,4
	69,00	10	,3	,3	99,7
	70,00	4	,1	,1	99,8
	71,00	2	,1	,1	99,9
	72,00	3	,1	,1	99,9
	73,00	2	,1	,1	100,0
	Gesamt	3468	99,9	100,0	
Fehlend	System	3	,1		
Gesamt		3471	100,0		

## 2.2 \*BILDUNG

Recode v86 (1=0) (2=1) (3=2) (4=3) (5=4) (else=SYSMIS) into bildung\_rec.  
variable labels bildung\_rec "Abschluss umkodiert".  
value labels bildung\_rec 0"Kein Schulabschluss" 1"Hauptschule" 2"Realschule"  
3"Fachhochschulreife" 4"Abi".

FREQUENCIES VARIABLES=bildung\_rec  
/STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS  
KURTOSIS.



# Häufigkeiten

## Anmerkungen

Ausgabe erstellt		01-NOV-2018 14:34:39
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition von fehlenden Werten	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax		FREQUENCIES VARIABLES=bildung_rec /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS KURTOSIS.
Ressourcen	Prozessorzeit	00:00:00,06
	Verstrichene Zeit	00:00:00,00

## Statistiken

Abschluss umkodiert

N	Gültig	3427
	Fehlend	44
Mittelwert		2,3262
Median		2,0000
Standardabweichung		1,21232

Schiefe	,252
Kurtosis	-1,296
Minimum	,00
Maximum	4,00

### Abschluss umkodiert

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Kein Schulabschluss	64	1,8	1,9	1,9
	Hauptschule	974	28,1	28,4	30,3
	Realschule	1144	33,0	33,4	63,7
	Fachhochschulreife	270	7,8	7,9	71,5
	Abi	975	28,1	28,5	100,0
	Gesamt	3427	98,7	100,0	
Fehlend	System	44	1,3		
Gesamt		3471	100,0		

## 2.3 \*GESCHLECHT

Recode v81 (2=0) (1=1) (else=copy) into geschl\_rec.  
variable labels geschl\_rec "Geschlecht mit w=0".  
value labels geschl\_rec 0"weiblich" 1"männlich".

```
FREQUENCIES VARIABLES=geschl_rec
/STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS
KURTOSIS.
```

## Häufigkeiten

### Anmerkungen

Ausgabe erstellt

01-NOV-2018 14:34:39

Kommentare

Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition von fehlenden Werten	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax		FREQUENCIES VARIABLES=geschl_rec /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS KURTOSIS.
Ressourcen	Prozessorzeit	00:00:00,05
	Verstrichene Zeit	00:00:00,00

## Statistiken

Geschlecht mit w=0

N	Gültig	3471
	Fehlend	0
Mittelwert		,5076
Median		1,0000
Standardabweichung		,50001
Schiefe		-,031
Kurtosis		-2,000
Minimum		,00
Maximum		1,00

## Geschlecht mit w=0

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	weiblich	1709	49,2	49,2	49,2
	männlich	1762	50,8	50,8	100,0

Gesamt	3471	100,0	100,0
--------	------	-------	-------

## 2.4 \*EINKOMMEN UMBENENNEN

COMPUTE einkommen = v420.

FREQUENCIES VARIABLES=einkommen  
 /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS  
 KURTOSIS.

## Häufigkeiten

Anmerkungen		
Ausgabe erstellt		01-NOV-2018 14:34:39
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition von fehlenden Werten	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax		FREQUENCIES VARIABLES=einkommen /STATISTICS=STDDEV MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS KURTOSIS.

Ressourcen	Prozessorzeit	00:00:00,11
	Verstrichene Zeit	00:00:00,00

## Statistiken

einkommen

N	Gültig	3065
	Fehlend	406
Mittelwert		11,1491
Median		11,0000
Standardabweichung		4,96229
Schiefe		,034
Kurtosis		-,856
Minimum		1,00
Maximum		22,00

## einkommen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1,00	37	1,1	1,2	1,2
	2,00	39	1,1	1,3	2,5
	3,00	103	3,0	3,4	5,8
	4,00	136	3,9	4,4	10,3
	5,00	157	4,5	5,1	15,4
	6,00	166	4,8	5,4	20,8
	7,00	199	5,7	6,5	27,3
	8,00	147	4,2	4,8	32,1
	9,00	252	7,3	8,2	40,3
	10,00	189	5,4	6,2	46,5
	11,00	145	4,2	4,7	51,2
	12,00	129	3,7	4,2	55,4
	13,00	342	9,9	11,2	66,6
	14,00	187	5,4	6,1	72,7
	15,00	242	7,0	7,9	80,6
	16,00	104	3,0	3,4	84,0
	17,00	129	3,7	4,2	88,2
	18,00	72	2,1	2,3	90,5
	19,00	171	4,9	5,6	96,1
	20,00	55	1,6	1,8	97,9

	21,00	51	1,5	1,7	99,6
	22,00	13	,4	,4	100,0
	Gesamt	3065	88,3	100,0	
Fehlend	System	406	11,7		
Gesamt		3471	100,0		

### 3. Aufgabe

#### 3.1 Modell a: Einkommen auf Alter

\*Model a.

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA CHANGE

/NOORIGIN

/DEPENDENT einkommen

/METHOD=ENTER alter0.

## Regression

Anmerkungen		
Ausgabe erstellt		01-NOV-2018 14:34:39
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition für "fehlend"	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Die Statistiken basieren auf Fällen, bei denen für keine verwendete Variable Werte fehlen.

Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /NOORIGIN /DEPENDENT einkommen /METHOD=ENTER alter0.
Ressourcen	Prozessorzeit	00:00:00,05
	Verstrichene Zeit	00:00:00,00
	Erforderlicher Speicherplatz	37088 Byte
	Zusätzlich erforderlicher Speicherplatz für Diagramme von Residuen	0 Byte

### Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
einkommen	11,1465	4,96107	3064
Alter mit 18 als Nullpunkt	31,8482	17,34265	3064

### Korrelationen

		einkommen	Alter mit 18 als Nullpunkt
Korrelation nach Pearson	einkommen	1,000	,068
	Alter mit 18 als Nullpunkt	,068	1,000
Sig. (1-seitig)	einkommen	.	,000
	Alter mit 18 als Nullpunkt	,000	.
N	einkommen	3064	3064
	Alter mit 18 als Nullpunkt	3064	3064

### Aufgenommene/Entfernte Variablen<sup>a</sup>

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Alter mit 18 als Nullpunkt <sup>b</sup>	.	Einschluß



- a. Abhängige Variable: einkommen  
b. Alle gewünschten Variablen wurden eingegeben.

### Modellzusammenfassung

Modell	R	R- Quadrat	Korrigierte s R- Quadrat	Standardf ehler des Schätzers	Statistikwerte ändern			
					Änderung in R- Quadrat	Änderung in F	df1	df2
1	,068 <sup>a</sup>	,005	,004	4,95047	,005	14,137	1	3062

### Modellzusammenfassung

Modell	Sig. Änderung in F
1	,000

- a. Einflußvariablen : (Konstante), Alter mit 18 als Nullpunkt

### ANOVA<sup>a</sup>

Modell		Quadratsum me	df	Mittel der Quadrate	F	Sig.
1	Regression	346,446	1	346,446	14,137	,000 <sup>b</sup>
	Nicht standardisierte Residuen	75040,758	3062	24,507		
	Gesamt	75387,203	3063			

- a. Abhängige Variable: einkommen  
b. Einflußvariablen : (Konstante), Alter mit 18 als Nullpunkt

## Koeffizienten<sup>a</sup>

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	10,529	,187		56,295	,000
	Alter mit 18 als Nullpunkt	,019	,005	,068	3,760	,000

a. Abhängige Variable: einkommen

### 3.2 Modell b: Einkommen auf Bildung

\*Model b.

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA CHANGE

/NOORIGIN

/DEPENDENT einkommen

/METHOD=ENTER bildung\_rec.

## Regression

### Anmerkungen

Ausgabe erstellt		01-NOV-2018 14:34:39
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>

	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition für "fehlend"	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Die Statistiken basieren auf Fällen, bei denen für keine verwendete Variable Werte fehlen.
Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /NOORIGIN /DEPENDENT einkommen /METHOD=ENTER bildung_rec.
Ressourcen	Prozessorzeit	00:00:00,05
	Verstrichene Zeit	00:00:00,00
	Erforderlicher Speicherplatz	37088 Byte
	Zusätzlich erforderlicher Speicherplatz für Diagramme von Residuen	0 Byte

## Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
einkommen	11,1845	4,94769	3040
Abschluss umkodiert	2,3178	1,20122	3040

## Korrelationen

		einkommen	Abschluss umkodiert
Korrelation nach Pearson	einkommen	1,000	,255
	Abschluss umkodiert	,255	1,000
Sig. (1-seitig)	einkommen	.	,000
	Abschluss umkodiert	,000	.
N	einkommen	3040	3040
	Abschluss umkodiert	3040	3040

## Aufgenommene/Entfernte Variablen<sup>a</sup>

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Abschluss umkodiert <sup>b</sup>	.	Einschluß

a. Abhängige Variable: einkommen

b. Alle gewünschten Variablen wurden eingegeben.

## Modellzusammenfassung

Modell	R	R- Quadrat	Korrigierte s R- Quadrat	Standardf ehler des Schätzers	Statistikwerte ändern			
					Änderung in R- Quadrat	Änderung in F	df1	df2
1	,255 <sup>a</sup>	,065	,065	4,78447	,065	211,879	1	3038

## Modellzusammenfassung

Modell	Statistikwerte ändern	
	Sig. Änderung in F	
1		,000

a. Einflußvariablen : (Konstante), Abschluss umkodiert

### ANOVA<sup>a</sup>

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	4850,148	1	4850,148	211,879	,000 <sup>b</sup>
	Nicht standardisierte Residuen	69543,325	3038	22,891		
	Gesamt	74393,473	3039			

a. Abhängige Variable: einkommen

b. Einflußvariablen : (Konstante), Abschluss umkodiert

### Koeffizienten<sup>a</sup>

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	8,747	,189		46,376	,000
	Abschluss umkodiert	1,052	,072	,255	14,556	,000

a. Abhängige Variable: einkommen

### 3.3 Modell c: Einkommen auf Geschlecht

#### REGRESSION

```

/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/NOORIGIN
/DEPENDENT einkommen
/METHOD=ENTER geschl_rec.

```

# Regression

## Anmerkungen

Ausgabe erstellt		01-NOV-2018 14:34:39
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition für "fehlend"	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Die Statistiken basieren auf Fällen, bei denen für keine verwendete Variable Werte fehlen.
Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /NOORIGIN /DEPENDENT einkommen /METHOD=ENTER geschl_rec.
Ressourcen	Prozessorzeit	00:00:00,05
	Verstrichene Zeit	00:00:01,00
	Erforderlicher Speicherplatz	37088 Byte
	Zusätzlich erforderlicher Speicherplatz für Diagramme von Residuen	0 Byte

## Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
einkommen	11,1491	4,96229	3065
Geschlecht mit w=0	,5250	,49946	3065

## Korrelationen

		einkommen	Geschlecht mit w=0
Korrelation nach Pearson	einkommen	1,000	,350
	Geschlecht mit w=0	,350	1,000
Sig. (1-seitig)	einkommen	.	,000
	Geschlecht mit w=0	,000	.
N	einkommen	3065	3065
	Geschlecht mit w=0	3065	3065

## Aufgenommene/Entfernte Variablen<sup>a</sup>

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Geschlecht mit w=0 <sup>b</sup>	.	Einschluß

a. Abhängige Variable: einkommen

b. Alle gewünschten Variablen wurden eingegeben.

## Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Statistikwerte ändern			
					Änderung in R-Quadrat	Änderung in F	df1	df2
1	,350 <sup>a</sup>	,122	,122	4,64986	,122	426,572	1	3063

## Modellzusammenfassung

Modell	Sig. Änderung in F
1	,000

a. Einflußvariablen : (Konstante), Geschlecht mit w=0

ANOVA <sup>a</sup>						
Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	9223,011	1	9223,011	426,572	,000 <sup>b</sup>
	Nicht standardisierte Residuen	66225,849	3063	21,621		
	Gesamt	75448,860	3064			

a. Abhängige Variable: einkommen

b. Einflußvariablen : (Konstante), Geschlecht mit w=0

Koeffizienten <sup>a</sup>						
		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
Modell		Regressionskoeffizient B	Standardfehler	Beta	T	Sig.
1	(Konstante)	9,326	,122		76,527	,000
	Geschlecht mit w=0	3,474	,168	,350	20,654	,000

a. Abhängige Variable: einkommen

### 3.4 Modell ab: Einkommen auf Alter, Bildung und Geschlecht

REGRESSION

/DESCRIPTIVES MEAN STDDEV CORR SIG N

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA CHANGE

/NOORIGIN

/DEPENDENT einkommen

/METHOD=ENTER alter0 bildung\_rec.



# Regression

Anmerkungen		
Ausgabe erstellt		01-NOV-2018 14:34:40
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition für "fehlend"	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Die Statistiken basieren auf Fällen, bei denen für keine verwendete Variable Werte fehlen.
Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /NOORIGIN /DEPENDENT einkommen /METHOD=ENTER alter0 bildung_rec.
Ressourcen	Prozessorzeit	00:00:00,05
	Verstrichene Zeit	00:00:00,00
	Erforderlicher Speicherplatz	37536 Byte

Zusätzlich erforderlicher Speicherplatz für Diagramme von Residuen	0 Byte
--	--------

## Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
einkommen	11,1820	4,94647	3039
Alter mit 18 als Nullpunkt	31,9885	17,27385	3039
Abschluss umkodiert	2,3172	1,20103	3039

## Korrelationen

		einkommen	Alter mit 18 als Nullpunkt	Abschluss umkodiert
Korrelation nach Pearson	einkommen	1,000	,057	,255
	Alter mit 18 als Nullpunkt	,057	1,000	-,268
	Abschluss umkodiert	,255	-,268	1,000
Sig. (1-seitig)	einkommen	.	,001	,000
	Alter mit 18 als Nullpunkt	,001	.	,000
	Abschluss umkodiert	,000	,000	.
N	einkommen	3039	3039	3039
	Alter mit 18 als Nullpunkt	3039	3039	3039
	Abschluss umkodiert	3039	3039	3039

## Aufgenommene/Entfernte Variablen<sup>a</sup>

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Abschluss umkodiert, Alter mit 18 als Nullpunkt <sup>b</sup>	.	Einschluß

a. Abhängige Variable: einkommen

b. Alle gewünschten Variablen wurden eingegeben.

### Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Statistikwerte ändern			
					Änderung in R-Quadrat	Änderung in F	df1	df2
1	,286 <sup>a</sup>	,082	,081	4,74111	,082	135,439	2	3036

### Modellzusammenfassung

Modell	Statistikwerte ändern	
	Sig. Änderung in F	
1	,000	

a. Einflußvariablen : (Konstante), Abschluss umkodiert, Alter mit 18 als Nullpunkt

### ANOVA<sup>a</sup>

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	6088,806	2	3044,403	135,439	,000 <sup>b</sup>
	Nicht standardisierte Residuen	68243,566	3036	22,478		
	Gesamt	74332,372	3038			

a. Abhängige Variable: einkommen

b. Einflußvariablen : (Konstante), Abschluss umkodiert, Alter mit 18 als Nullpunkt

## Koeffizienten<sup>a</sup>

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regression skoeffizient B	Standardfehler	Beta		
1	(Konstante)	7,165	,282		25,390	,000
	Alter mit 18 als Nullpunkt	,039	,005	,135	7,496	,000
	Abschluss umkodiert	1,199	,074	,291	16,124	,000

a. Abhängige Variable: einkommen

### 3.5 Modell abc: Einkommen auf Alter, Bildung und Geschlecht

#### REGRESSION

```

/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/NOORIGIN
/DEPENDENT einkommen
/METHOD=ENTER alter0 bildung_rec geschl_rec.

```

## Regression

### Anmerkungen

Ausgabe erstellt		01-NOV-2018 14:34:40
Kommentare		
Eingabe	Daten	/Users/annaenders/Downloads/allbus2014.sav
	Aktiver Datensatz	DataSet1

	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	3471
Behandlung fehlender Werte	Definition für "fehlend"	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Die Statistiken basieren auf Fällen, bei denen für keine verwendete Variable Werte fehlen.
Syntax	REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /NOORIGIN /DEPENDENT einkommen /METHOD=ENTER alter0 bildung_rec geschl_rec.	
Ressourcen	Prozessorzeit	00:00:00,05
	Verstrichene Zeit	00:00:00,00
	Erforderlicher Speicherplatz	38064 Byte
	Zusätzlich erforderlicher Speicherplatz für Diagramme von Residuen	0 Byte

## Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
einkommen	11,1820	4,94647	3039
Alter mit 18 als Nullpunkt	31,9885	17,27385	3039
Abschluss umkodiert	2,3172	1,20103	3039
Geschlecht mit w=0	,5255	,49943	3039

## Korrelationen

		einkommen	Alter mit 18 als Nullpunkt	Abschluss umkodiert	Geschlecht mit w=0
Korrelation nach Pearson	einkommen	1,000	,057	,255	,350
	Alter mit 18 als Nullpunkt	,057	1,000	-,268	-,004
	Abschluss umkodiert	,255	-,268	1,000	-,027
	Geschlecht mit w=0	,350	-,004	-,027	1,000
Sig. (1-seitig)	einkommen	.	,001	,000	,000
	Alter mit 18 als Nullpunkt	,001	.	,000	,412
	Abschluss umkodiert	,000	,000	.	,067
	Geschlecht mit w=0	,000	,412	,067	.
N	einkommen	3039	3039	3039	3039
	Alter mit 18 als Nullpunkt	3039	3039	3039	3039
	Abschluss umkodiert	3039	3039	3039	3039
	Geschlecht mit w=0	3039	3039	3039	3039

### Aufgenommene/Entfernte Variablen<sup>a</sup>

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Geschlecht mit w=0, Alter mit 18 als Nullpunkt, Abschluss umkodiert <sup>b</sup>	.	Einschluß

a. Abhängige Variable: einkommen

b. Alle gewünschten Variablen wurden eingegeben.

## Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Statistikwerte ändern			
					Änderung in R-Quadrat	Änderung in F	df1	df2
1	,459 <sup>a</sup>	,211	,210	4,39627	,211	270,330	3	3035

## Modellzusammenfassung

Modell	Statistikwerte ändern	
	Sig. Änderung in F	
1	,000	

a. Einflußvariablen : (Konstante), Geschlecht mit w=0, Alter mit 18 als Nullpunkt, Abschluss umkodiert

## ANOVA<sup>a</sup>

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	15674,220	3	5224,740	270,330	,000 <sup>b</sup>
	Nicht standardisierte Residuen	58658,152	3035	19,327		
	Gesamt	74332,372	3038			

a. Abhängige Variable: einkommen

b. Einflußvariablen : (Konstante), Geschlecht mit w=0, Alter mit 18 als Nullpunkt, Abschluss umkodiert

## Koeffizienten<sup>a</sup>

		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
Modell		Regressionskoeffizient B	Standardfehler	Beta	T	Sig.
1	(Konstante)	5,151	,277		18,603	,000
	Alter mit 18 als Nullpunkt	,040	,005	,140	8,345	,000
	Abschluss umkodiert	1,244	,069	,302	18,035	,000
	Geschlecht mit w=0	3,558	,160	,359	22,270	,000

a. Abhängige Variable: einkommen

3a Vergleichen Sie die Regressionskoeffizienten über die Modelle und erläutern Sie, was hier festzustellen ist!

Modellbezeichnung	Enthaltene Variablen	Regressionskoeffizient B	Beta
Modell a	Einkommen auf Alter	,019	,068
Modell b	Einkommen auf Bildung	1,052	,255
Modell c	Einkommen auf Geschlecht	3,474	,168
Modell ab	Einkommen auf Alter und Bildung	Alter: ,039 Bildung: 1,199	Alter: ,135 Bildung: ,291
Modell abc	Einkommen auf Alter, Bildung und Geschlecht	Alter: ,040 Bildung: 1,244 Geschlecht: 3,558	Alter: ,140 Bildung: ,302 Geschlecht: ,359

Tabelle 1: Regressionskoeffizient B für die berechneten Modelle. Eigene Darstellung.

Die standardisierten Beta-Werte für die einzelnen Variablen erhöhen sich, wenn weitere Variablen in das Modell miteinbezogen werden. Der Einfluss der einzelnen Variablen wird also erhöht, wenn Effekte zwischen den Variablen auspartialisiert wurden.



3b Vergleichen Sie  $R^2$  über die Modelle und erläutern Sie, was hier festzustellen ist!

Modellbezeichnung	Enthaltene Variablen	$R^2$
Modell a	Einkommen auf Alter	,005
Modell b	Einkommen auf Bildung	,065
Modell c	Einkommen auf Geschlecht	,122
Modell ab	Einkommen auf Alter und Bildung	,082
Modell abc	Einkommen auf Alter, Bildung und Geschlecht	,211

Tabelle 2:  $R^2$  für die berechneten Modelle. Eigene Darstellung.

In Vergleich der bivariaten Regressionsmodelle hat das Geschlecht auf das Einkommen die höchste Erklärungskraft von 12 %.

Im Vergleich der multivariaten Regressionsmodelle zeigt sich, dass der Einfluss der beiden Variablen Alter und Bildung auf das Einkommen verglichen mit den jeweiligen bivariaten Modellen nun auf 8 % steigt. Die höchste Erklärungskraft hat das multivariate Modell, in denen alle drei Kontrollvariablen, Alter, Bildung und Geschlecht berücksichtigt werden. Hier hat das Modell 21 % Erklärungskraft.

#### 4. Literaturverzeichnis

Urban, D., & Mayerl, J. (2011). *Regressionsanalyse: Theorie, Technik und Anwendung*. Wiesbaden: VS-Verlag.