

# Sommersemester 2008, Statistik mit SPSS



## Überblick:

1. Hilfsbefehl und Datentransformationsbefehl (EXECUTE und COMPUTE)
2. Möglichkeiten und Aufgabe der deskriptiven (univariaten) Statistik
3. SPSS- Prozeduren zur univariaten Statistik
4. Statistische Kennwerte und Grafiken für ...
  - a) nominale Daten
  - b) ordinale Daten
  - c) metrische Daten

## 1a. Hilfsbefehl EXECUTE

## EXECUTE

- SPSS-Befehle zur Datenbeschreibung, Transformationen und Auswahl von Daten werden vom Programm nicht sofort ausgeführt, sondern zunächst nur auf Syntaxfehler hin überprüft.
- Erst bei der nächsten Statistikprozedur (z.B. fre) wird der Befehl ausgeführt und das Ergebnis (z.B. die neue Variable) in der Arbeitsdatei gespeichert.
- Soll ein Befehl (z.B. recode, compute) sofort umgesetzt werden, dann muss nach dem betreffenden Befehl die SPSS-Anweisung „EXECUTE“ folgen.

1b. Datentransformationsbefehl **COMPUTE** (Legt neue Variablen an, berechnet neue Variablen)

## Variablen berechnen: COMPUTE

### Anwendungsbeispiele:

- Viele benötigte Variablen existieren im Datensatz nicht, z.B.
  - Alter (berechnet aus Geburtsjahr und Befragungsjahr)
  - Additiver Index ‚Rechtsextremismus‘ (Addition versch. Items)
  - Berechnung Fernsehdauer in Stunden pro Tag
  - BMI (berechnet aus Körpergröße und Gewicht)

## Variablen berechnen: COMPUTE

**Beispiel 1:** Aus der Variable vjahr (Geburtsjahr) soll das Alter der Befragten errechnet werden (BTW05-ZA-Datensatz)

→ COMPUTE zielvariable = arithmetischer Ausdruck.

compute alter = ????

## Beispiel 1: Berechnung Alter in Jahren

Geburtsjahr	vjahr
Befragungszeitpunkt	2005

Formel für die Berechnung des Alters:

$$2005 - \text{vjahr}$$

**Befehl zur Erstellung einer neuen Variable**

Vollständiger SPSS-Befehl:

```
COMPUTE alter = 2005 - vjahr.
```

**Ausdruck**

**Zielvariable**

# Beispiel 1: Berechnung Alter in Jahren

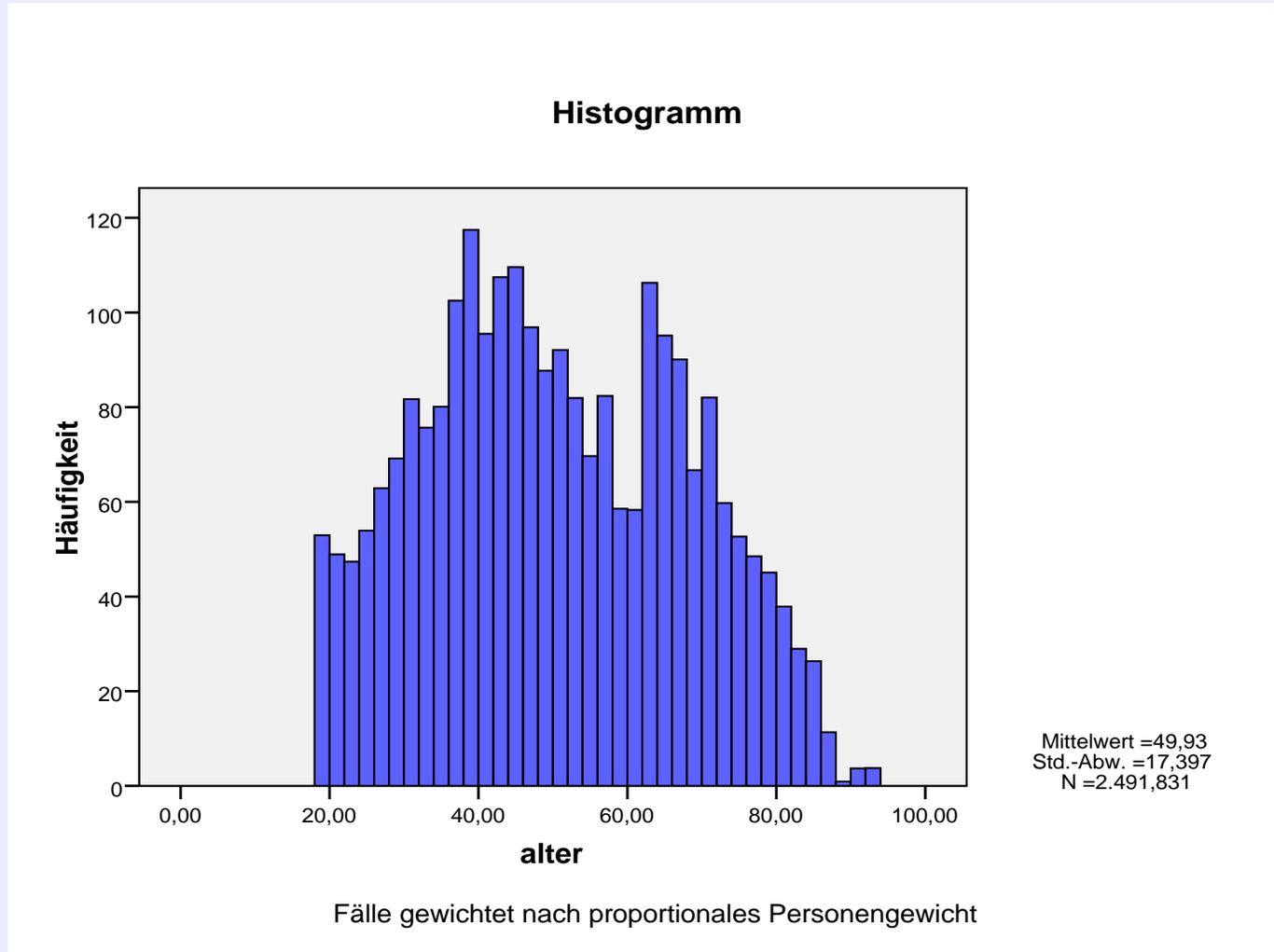
		alter			
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	18,00	32	1,2	1,3	1,3
	19,00	21	,8	,9	2,1
	20,00	30	1,2		76,00
	21,00	19	,7		77,00
	22,00	25	1,0		78,00
	23,00	23	,9		79,00
	24,00	24	1,0		80,00
	25,00	30	1,2		81,00
	26,00	34	1,3		82,00
	27,00	29	1,2		83,00
	28,00	31	1,2		84,00
	29,00	38	1,5		85,00
	30,00	38	1,5		86,00
	31,00	44	1,7		87,00
	32,00	39	1,5		89,00
	33,00	36	1,4		90,00
	34,00	35	1,4		91,00
	35,00	46	1,8		92,00
	36,00	49	1,9		94,00
	37,00	54	2,1		
				Gesamt	2492
				Fehlend System	48
				Gesamt	2540

20	,8	,8	91,7
20	,8	,8	92,5
29	1,1	1,2	93,7
25	1,0	1,0	94,6
20	,8	,8	95,5
27	1,1	1,1	96,5
11	,4	,4	97,0
13	,5	,5	97,5
16	,6	,7	98,1
13	,5	,5	98,7
13	,5	,5	99,2
7	,3	,3	99,5
4	,2	,2	99,7
1	,0	,0	99,7
2	,1	,1	99,8
2	,1	,1	99,8
2	,1	,1	99,9
1	,0	,0	100,0
		Gesamt	100,0

Überprüfen ob Missings gesetzt sind!

# Beispiel 1: Berechnung Alter in Jahren



## Beispiel 1: Berechnung Alter in Jahren

The screenshot shows the SPSS Syntax Editor window titled "Beispiele 07-05-08 - SPSS Syntax-Editor". The menu bar includes "Datei", "Bearbeiten", "Ansicht", "Daten", "Transformieren", "Analysieren", "Grafiken", "Extras", "Ausführen", "Fenster", and "Hilfe". The toolbar contains various icons for file operations and execution. The main text area contains the following syntax script:

```
*** Beispiel 1: Alter berechnen, Wahlstudie 2005  
(BTW05).  
  
fre s02a.  
compute alter = 2005-s02a.  
exe.  
fre alter/histo.
```

The status bar at the bottom indicates "SPSS Prozessor ist bereit".

# Beispiel 1: Berechnung Alter in Jahren

**alter**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	18,00	32	1,2	1,3	1,3
	19,00	21	,8	,9	2,1
	20,00	30	1,2	1,2	3,3
	21,00	19	,7	,7	4,1
	22,00	25	1,0	1,0	5,1
	23,00	23	,9	,9	6,0
	24,00	24	1,0	1,0	7,0
	25,00	30	1,2	1,2	8,2
	26,00	34	1,3	1,3	9,5
	27,00	29	1,2	1,2	10,7
	28,00	31	1,2	1,3	11,9
	29,00	38	1,5	1,5	13,5
	30,00	38	1,5	1,5	15,0
	31,00	44	1,7	1,8	16,7
	32,00	39	1,5	1,6	18,3
	33,00	36	1,4	1,5	19,8
	34,00	35	1,4	1,4	21,2
	35,00	46	1,8	1,8	23,0
	36,00	40	1,5	2,0	24,0

Aus dem Geburtsjahr  
wird das Alter

## Zusatz: Struktur des Compute-Befehls

➔ COMPUTE zielvariable = arithmetischer Ausdruck.

Arithmetischer Ausdruck (mathematische Formel):

- Bereits definierte Variablen
- Zahlen
- arithmetische Operationen:
  - Potenzfunktion (\*\*),
  - Multiplikation (\*),
  - Division (/),
  - Addition (+),
  - Subtraktion (-)
- Klammern
- Funktionen

## Zusatz: Struktur des Compute-Befehls

→ COMPUTE zielvariable = arithmetischer Ausdruck.

Abarbeitungsreihenfolge beim arithmetischen Ausdruck:

1. Klammern
2. Funktionen
3. \*\*
4. \*, /
5. +, -

## Beispiel 2: Indexerstellung

Index: Zusammenfassung mehrerer Items zu einer neuen Variable. Neue Variable beinhaltet verschiedene Aspekte eines theoretischen Konstruktes.

Aufgabe: Aus den sechs Variablen f062\_1 bis f062\_6 (Wahlstudie 2005) soll ein additiver Index ‚Rechtsextremismus‘ erstellt werden.

**f062\_1 Rechtsextr: herausragende Leistung deutschen Volkes**

Gültig	1 lehne voellig
	2 lehne eher a
	3 stimme eher
	4 stimme voel
	Gesamt
Fehlend	8 weiss nicht
	9 keine Angat
	Gesamt
Gesamt	

**f062\_2 Rechtsextr: unwertes Leben**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 lehne voellig ab	1647	64,8	68,8	68,8
	2 lehne eher ab	330	13,0	13,8	82,5
	3 stimme eher zu	250	9,8	10,4	93,0
	4 stimme voellig zu	168	6,6	7,0	100,0
	Gesamt	2395	94,3	100,0	
Fehlend	8 weiss nicht	114	4,5		
	9 keine Angabe	30	1,2		
	Gesamt	145	5,7		
Gesamt		2540	100,0		

## Beispiel 2: Indexerstellung

### Beachte:

- vor dem Compute-Befehl: Sind missing values definiert?
- alle Items müssen den gleichen Wertebereich aufweisen (hier: 1 bis 4)
- alle Items müssen die gleiche Polung aufweisen (hier: je höher der Wert, desto rechtsextremer die Einstellung)
- Wertebereich der neuen Variable: 6 bis 24, d.h. Wert 6 = keine rechtsextreme Einstellung, Wert 24 = stärkste rechtsextreme Einstellung

## Beispiel 2: Indexerstellung

```
***Beispiel 2: Additiver Index Rechtsextremismus (BTW05).  
  
fre f062_1 to f062_6.  
Compute rex_idx=f062_1+f062_2+f062_3+f062_4+f062_5+f062_6.  
  
var lab rex_idx 'Index Rechtsextremismus'.  
val lab rex_idx 6 'keine rechtsextreme Einstellung'  
24 'starke rechtsextreme Einstellung' .  
  
exe.  
fre rex_idx/histo.
```

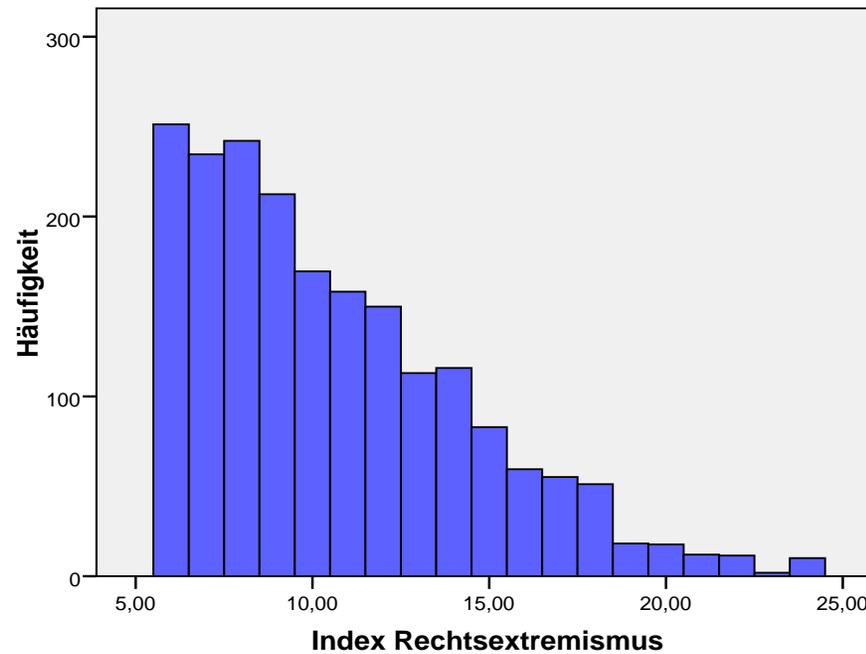
SPSS Prozessor ist bereit

# Beispiel 2: Indexerstellung

rex\_idx Index Rechtsextremismus

		Häufi	Gültige	Kumulierte
Gültig	6,00 keine rechtsextreme Einstellung			
	7,00			
	8,00			
	9,00			
	10,00			
	11,00			
	12,00			
	13,00			
	14,00			
	15,00			
	16,00			
	17,00			
	18,00			
	19,00			
	20,00			
	21,00			
	22,00			
	23,00			
	24,00 starke rechtsextreme Einstellung			
Gesamt				
Fehlend	System			
Gesamt		2540	100,0	

Histogramm



Mittelwert =10,61  
 Std.-Abw. =3,845  
 N =1.966,8

## Weitere Beispiele (Allbus 2004)

- Berechnung Fernsehdauer in Stunden pro Tag
- BMI (berechnet aus Körpergröße und Gewicht)

## 2. Möglichkeiten und Aufgabe der deskriptiven (univariaten) Statistik

# Möglichkeiten und Aufgabe der univariaten Statistik

- Überblick über univariate Verteilungen: Häufigkeitstabellen und graphische Darstellung
- Zusammenfassung der Eigenschaften einer Verteilung in einem Wert (bzw. wenigen Werten): Kennwerte univariate Verteilungen wie z.B. Lagemaße, Streuungsmaße, Schiefe, Kurtosis
- Insbesondere Verteilung, Lage, Streuung und Schiefe der sollten untersucht werden:
  - Fehler im Datensatz (z.B. Alter)
  - sind fehlende Werte definiert?
  - zu schiefe Verteilungen häufig problematisch
  - Ausreißeridentifikation
  - Anwendungsvoraussetzungen erfüllt?

# Übersicht über die SPSS Prozeduren

## 1. Häufigkeiten (**FREQUENCIES**):

- ▶ Analysieren ▶ Deskriptive Statistiken
- ▶ Häufigkeiten

## 2. Explorative Datenanalyse (**EXAMINE**):

- ▶ Analysieren ▶ Deskriptive Statistiken
- ▶ Explorative Datenanalyse

## (3. Deskriptive Statistiken (DESCRIPTIVES):

- ▶ Analysieren ▶ Deskriptive Statistiken
- ▶ Deskriptive Statistiken )

## (4. Fälle zusammenfassen (SUMMARIZE):

- ▶ Analysieren ▶ Berichte ▶ Fälle zusammenfassen)

# Die Prozedur FREQUENCIES

```

FREQUENCIES [VARIABLES=]varlist
    [ /FORMAT=[ {DVALUE} ] [ {NOTABLE } ] ]
                {AFREQ }      {LIMIT(n)}
                {DFREQ }
    [ /MISSING=INCLUDE ]
    [ /BARCHART=[MIN(n) ] [MAX(n) ] [ {FREQ(n) } ] ]
                                    {PERCENT(n)}
    [ /PIECHART=[MIN(n) ] [MAX(n) ] [ {FREQ } ] ]
                                    {PERCENT}
                                [ {MISSING } ] ]
                                {NOMISSING}
    [ /HISTOGRAM=[MIN(n) ] [MAX(n) ] [ {FREQ(n) } ] ]
                                    {PERCENT(n)}
                                [ {NONORMAL} ] [ INCREMENT(n) ] ]
                                {NORMAL }
    [ /NTILES=n ]
    [ /PERCENTILES=value list ]
    [ /STATISTICS=[DEFAULT] [MEAN] [STDDEV] [SUM]
                    [MINIMUM] [MAXIMUM] [RANGE]
                    [SEMEAN] [VARIANCE] [SKEWNESS] [SESKEW]
                    [MODE] [KURTOSIS] [SEKURT] [MEDIAN]
                    [ALL] [NONE] ]
    [ /GROUPED=varlist [ { (width) } ] ]
                                { (boundary list) }
    [ /ORDER = [ {ANALYSIS} ]
                [ {VARIABLE} ] ]
    
```

## a) Nominale Daten

→ Für die Variable f012c (Wahlentscheidung, Nachwahlbefragung) soll der Modus und ein Tortendiagramm ausgegeben werden.

The screenshot shows the SPSS Syntax-Editor window titled "Beispiele 07-05-08 - SPSS Syntax-Editor". The menu bar includes "Datei", "Bearbeiten", "Ansicht", "Daten", "Transformieren", "Analysieren", "Grafiken", "Extras", "Ausführen", "Fenster", and "Hilfe". The toolbar contains various icons for file operations and analysis. The main text area contains the following SPSS syntax command:

```
***Univariate Statistiken für nominale Daten.  
  
fre f012c/ stat mod/pie per.
```

The status bar at the bottom indicates "SPSS Prozessor ist bereit".

### Statistiken

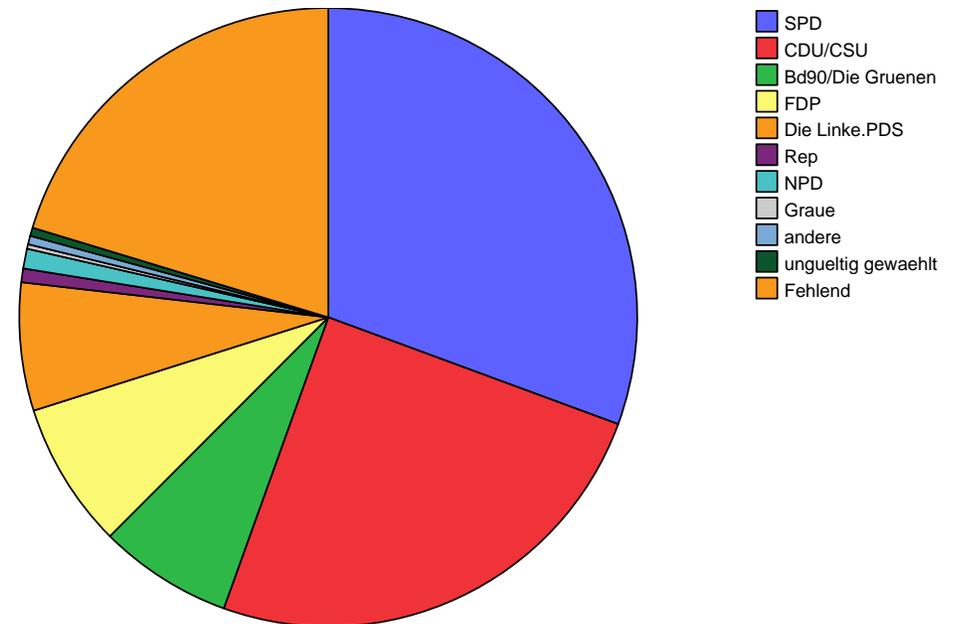
f012c BTW 05 Nachwahlbefragung Zweitstimme

N	Gültig	2024
	Fehlend	516
Modus		10

f012c BTW 05 Nachwahlbefra

		Häufigkeit
Gültig	10 SPD	778
	11 CDU/CSU	632
	12 Bd90/Die Gruenen	178
	13 FDP	194
	14 Die Linke.PDS	170
	15 Rep	18
	16 NPD	26
	17 Graue	6
	18 andere	11
	19 ungueltig gewaehlt	11
Gesamt		2024
Fehlend	-1 trifft nicht zu	318
	98 kann mich nicht erinnern	10
	99 verweigert	188
	Gesamt	516
Gesamt		2540

BTW 05 Nachwahlbefragung Zweitstimme



Fälle gewichtet nach proportionales Personengewicht

100,0

## b) Ordinale Daten

→ Für die Variable f003 (Demokratiezufriedenheit) sollen Modus, Median und Quartile und zusätzlich ein Balkendiagramm ausgegeben werden.

The screenshot shows the SPSS Syntax Editor window titled "Beispiele 07-05-08 - SPSS Syntax-Editor". The menu bar includes "Datei", "Bearbeiten", "Ansicht", "Daten", "Transformieren", "Analysieren", "Grafiken", "Extras", "Ausführen", "Fenster", and "Hilfe". The toolbar contains various icons for file operations and analysis. The main text area contains the following SPSS syntax command:

```
***Univariate Statistiken für ordinale Daten.  
  
fre f003/format notable/nti 4/stat min max med mod/bar per.
```

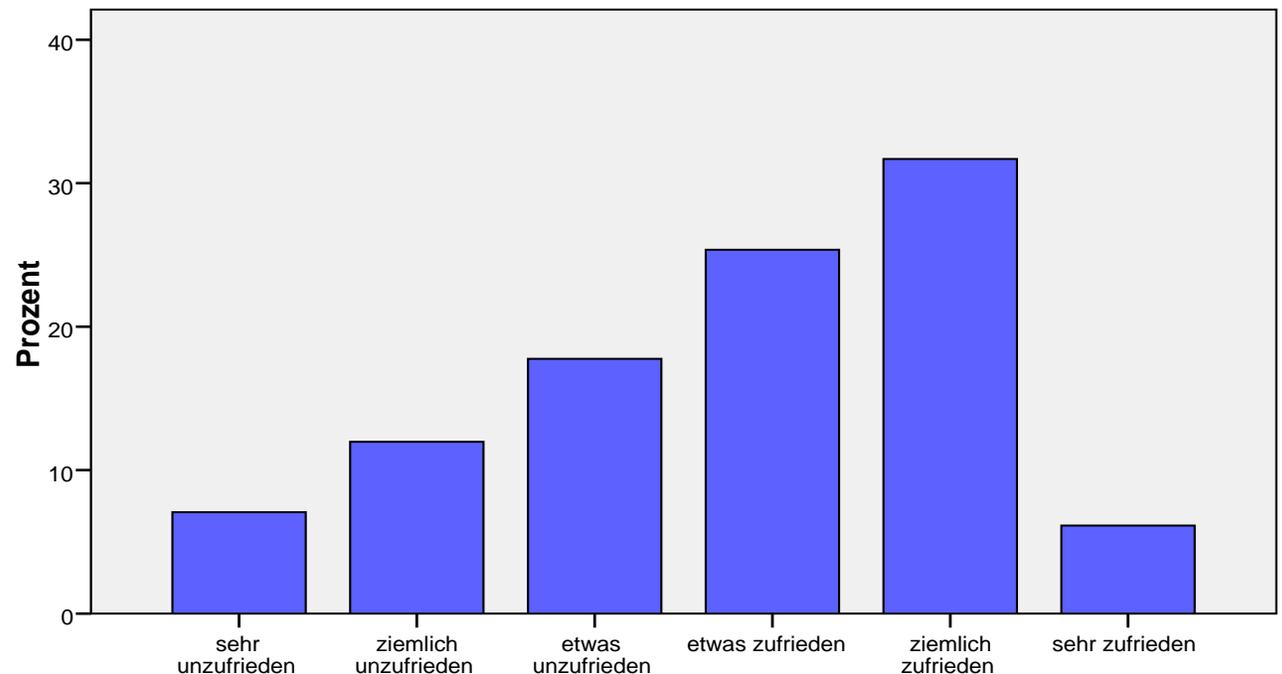
The status bar at the bottom indicates "SPSS Prozessor ist bereit".

## Statistiken

f003 Demokrat  
tatsaechlicher I

N	Gr
	Fe
Median	
Modus	
Minimum	
Maximum	
Perzentile	25
	50
	75

Demokratie: Zufriedenheit mit tatsaechlicher Demokratie in BRD



Demokratie: Zufriedenheit mit tatsaechlicher Demokratie in BRD

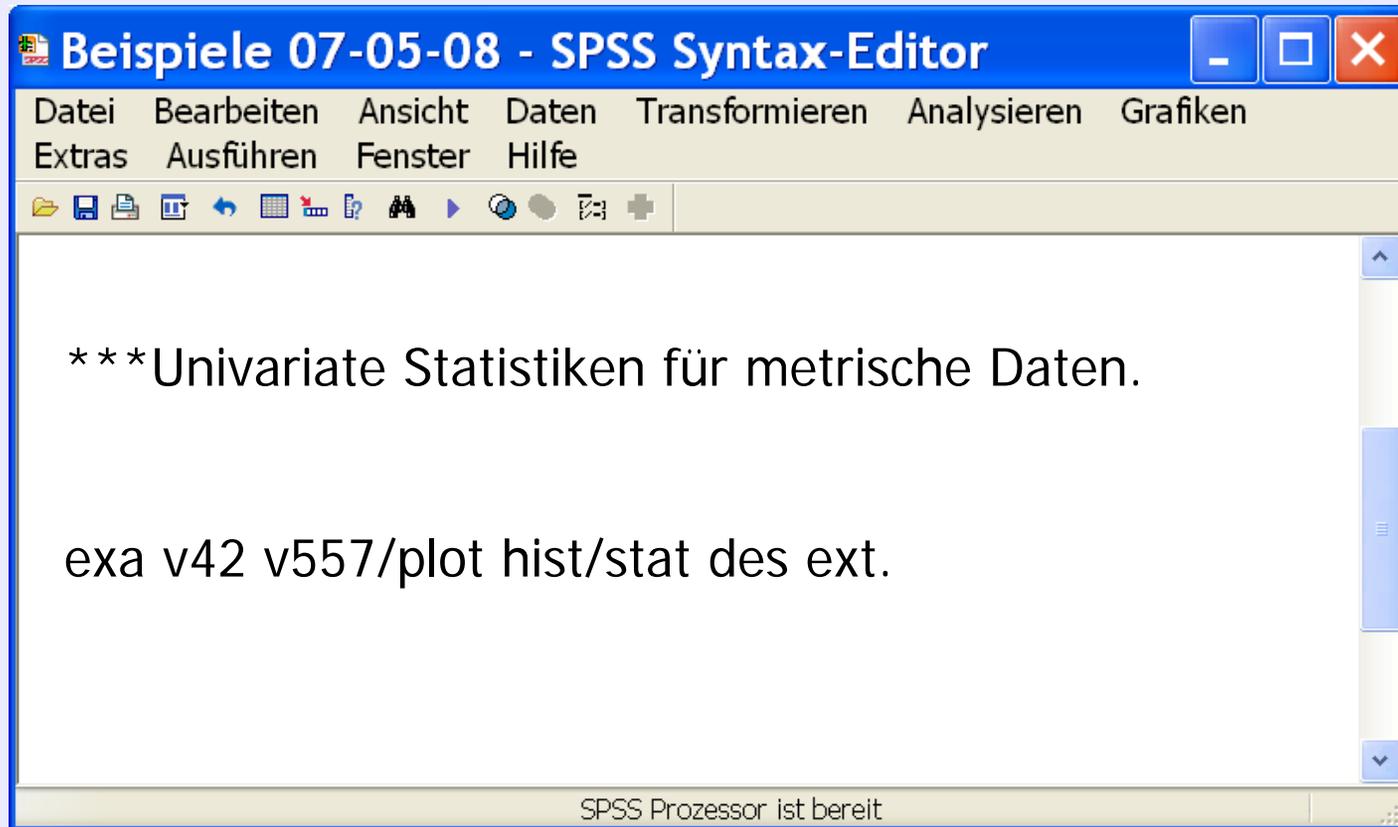
Fälle gewichtet nach proportionales Personengewicht

# Die Prozedur EXAMINE

```

EXAMINE VARIABLES=varlist [[BY varlist] [varname BY varname]]
    [/COMPARE={GROUP** }]
                {VARIABLE}
    [/ID={ $CASENUM**}]
                {varname }
    [/PERCENTILES[(value list)={HAVERAGE }] [NONE]]
                                {WAVERAGE }
                                {ROUND }
                                {AEMPIRICAL}
                                {EMPIRICAL }
    [/PLOT=[STEMLEAF**] [BOXPLOT**] [NPLOT]]
                [SPREADLEVEL(value)] [HISTOGRAM]
                [{ALL }]
                {NONE}
    [/STATISTICS=[DESCRIPTIVES**] [EXTREME({5})]]
                                                {n}
                [{ALL }]
                {NONE}
    [{/TOTAL }]
    {/NOTOTAL}
    .....
    [/MISSING={LISTWISE**} [INCLUDE]]
                {REPORT }
                {PAIRWISE }
    **Default if the subcommand is omitted.
    
```

## c) Metrische Daten



## Deskriptive Statistik

			Statistik	Standardfehler	
f0 gl	f057_c Merkel: glaubwuerdig	Mittelwert	3,17	,032	ent
		95% Konfidenzintervall des Mittelwerts	Untergrenze Obergrenze	3,11 3,24	0%
		5% getrimmtes Mittel	3,19		
		Median	3,00		
		Varianz	2,566		
		Standardabweichung	1,602		
		Minimum	0		
		Maximum	6		
		Spannweite	6		
		Interquartilbereich	2		
		Schiefe	-,249	,050	
		Kurtosis	-,577	,099	

Schiefe:

- Wert 0 = symmetrische Verteilung
- Wert > 0 = rechtsschiefe Verteilung
- Wert < 0 = linksschiefe Verteilung

Kurtosis:

- Wert 0 = symmetrische Verteilung
- Wert > 0 = schmale (steile) Verteilung
- Wert < 0 = breite (flache) Verteilung

**Extremwerte**

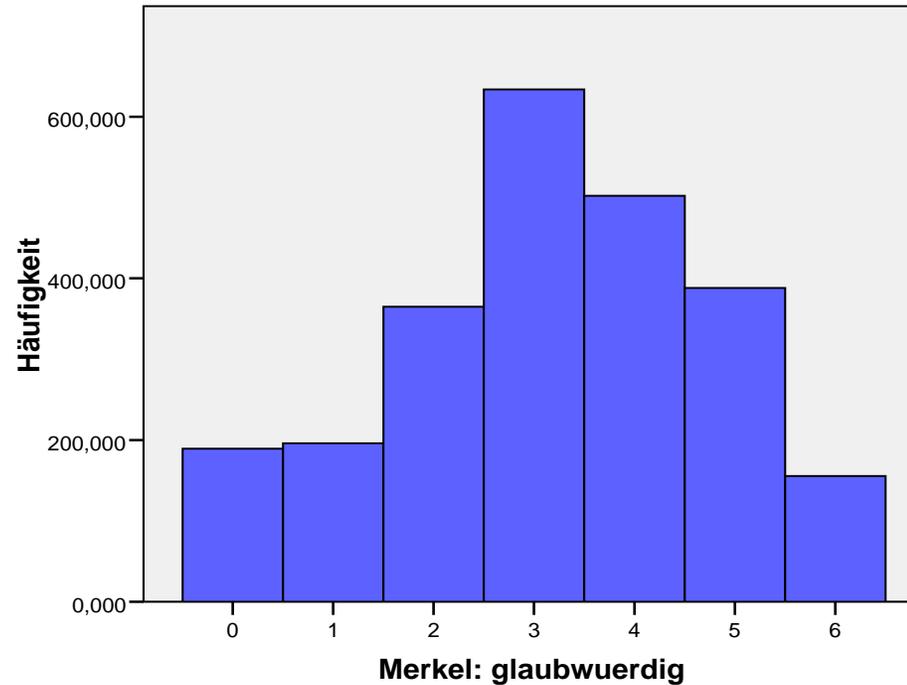
	Fallnummer	Wert
--	------------	------

f057\_c Merkel: Größte Werte  
glaubwuerdig

Kleinste Werte

- a. Nur eine partielle Liste von F  
Tabelle der oberen Extremw
- b. Nur eine partielle Liste von F  
Tabelle der unteren Extremw

**Histogramm**



Mittelwert =3,17  
Std.-Abw. =1,602  
N =2.429,738

Fälle gewichtet nach proportionales Personengewicht