

Sommersemester 2009, 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:**
 - a) Für nominale und ordinale Daten (FREQUENCIES)
 - b) metrische Daten (EXAMINE)

1a. Hilfsbefehl **EXECUTE**

1a. Hilfsbefehl 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)

(Wahlstudie 2005)

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

→ 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

```
File Edit View Data Transform Analyze Graphs Utilities Run Window
Help
***Beispiel 1: Alter berechnen, |Wahlstudie 2005 (BTW05).

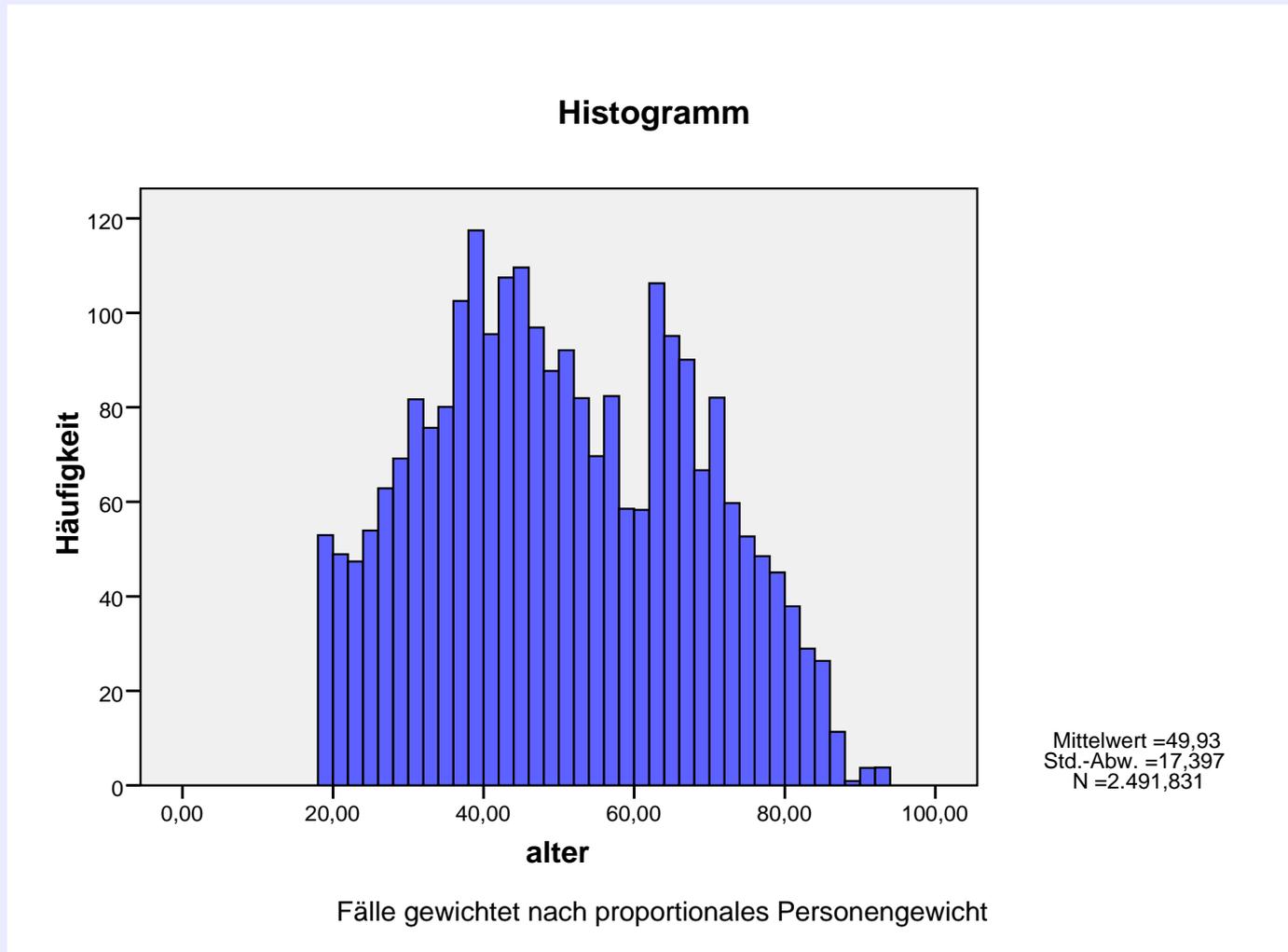
fre s02a.
compute alter = 2005-s02a.
exe.
fre alter/histo.
```

SPSS Processor is ready

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		73,00
	21,00	19	,7		76,00
	22,00	25	1,0		77,00
	23,00	23	,9		78,00
	24,00	24	1,0		79,00
	25,00	30	1,2		80,00
	26,00	34	1,3		81,00
	27,00	29	1,2		82,00
	28,00	31	1,2		83,00
	29,00	38	1,5		84,00
	30,00	38	1,5		85,00
	31,00	44	1,7		86,00
	32,00	39	1,5		87,00
	33,00	36	1,4		88,00
	34,00	35	1,4		89,00
	35,00	46	1,8		90,00
	36,00	49	1,9		91,00
	37,00	54	2,1		92,00
					93,00
					94,00
					95,00
					96,00
					97,00
					98,00
					99,00
					100,00
					101,00
					102,00
					103,00
					104,00
					105,00
					106,00
					107,00
					108,00
					109,00
					110,00
					111,00
					112,00
					113,00
					114,00
					115,00
					116,00
					117,00
					118,00
					119,00
					120,00
					121,00
					122,00
					123,00
					124,00
					125,00
					126,00
					127,00
					128,00
					129,00
					130,00
					131,00
					132,00
					133,00
					134,00
					135,00
					136,00
					137,00
					138,00
					139,00
					140,00
					141,00
					142,00
					143,00
					144,00
					145,00
					146,00
					147,00
					148,00
					149,00
					150,00
					151,00
					152,00
					153,00
					154,00
					155,00
					156,00
					157,00
					158,00
					159,00
					160,00
					161,00
					162,00
					163,00
					164,00
					165,00
					166,00
					167,00
					168,00
					169,00
					170,00
					171,00
					172,00
					173,00
					174,00
					175,00
					176,00
					177,00
					178,00
					179,00
					180,00
					181,00
					182,00
					183,00
					184,00
					185,00
					186,00
					187,00
					188,00
					189,00
					190,00
					191,00
					192,00
					193,00
					194,00
					195,00
					196,00
					197,00
					198,00
					199,00
					200,00
					201,00
					202,00
					203,00
					204,00
					205,00
					206,00
					207,00
					208,00
					209,00
					210,00
					211,00
					212,00
					213,00
					214,00
					215,00
					216,00
					217,00
					218,00
					219,00
					220,00
					221,00
					222,00
					223,00
					224,00
					225,00
					226,00
					227,00
					228,00
					229,00
					230,00
					231,00
					232,00
					233,00
					234,00
					235,00
					236,00
					237,00
					238,00
					239,00
					240,00
					241,00
					242,00
					243,00
					244,00
					245,00
					246,00
					247,00
					248,00
					249,00
					250,00
					251,00
					252,00
					253,00
					254,00
					255,00
					256,00
					257,00
					258,00
					259,00
					260,00
					261,00
					262,00
					263,00
					264,00
					265,00
					266,00
					267,00
					268,00
					269,00
					270,00
					271,00
					272,00
					273,00
					274,00
					275,00
					276,00
					277,00
					278,00
					279,00
					280,00
					281,00
					282,00
					283,00
					284,00
					285,00
					286,00
					287,00
					288,00
					289,00
					290,00
					291,00
					292,00
					293,00
					294,00
					295,00
					296,00
					297,00
					298,00
					299,00
					300,00
					301,00
					302,00
					303,00
					304,00
					305,00
					306,00
					307,00
					308,00
					309,00
					310,00
					311,00
					312,00
					313,00
					314,00
					315,00
					316,00
					317,00
					318,00
					319,00
					320,00
					321,00
					322,00
					323,00
					324,00
					325,00
					326,00
					327,00
					328,00
					329,00
					330,00
					331,00
					332,00
					333,00
					334,00
					335,00
					336,00
					337,00
					338,00
					339,00
					340,00
					341,00
					342,00

Beispiel 1: Berechnung Alter in Jahren



Beispiel 1: Berechnung Alter in Jahren

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

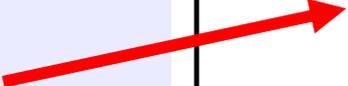
SPSS Prozessor ist bereit

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	1,5	24,5

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 z
	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

```
File Edit View Data Transform Analyze Graphs Utilities Run Window Help
***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.
```

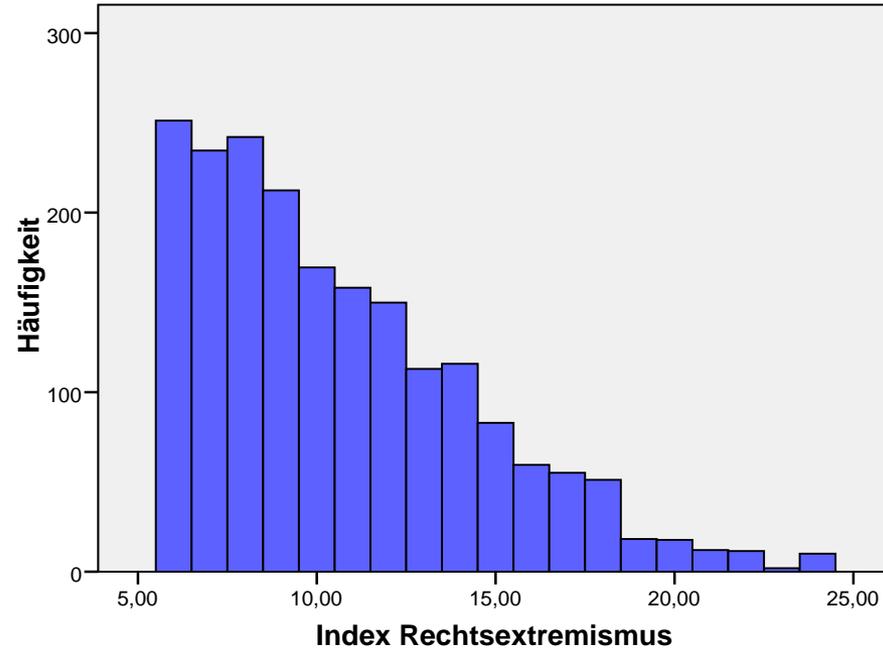
SPSS Processor is ready

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			
Fehlend	System			
Gesamt		2540	100,0	

Histogramm



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

Fälle gewichtet nach proportionales Personengewicht

Übungsaufgaben 1 – 3 (Allbus 2004)

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?

SPSS Prozeduren zur univariaten Statistik

1. Häufigkeiten (**FREQUENCIES**):

- ▶ Analysieren ▶ Deskriptive Statistiken
 - ▶ Häufigkeiten

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

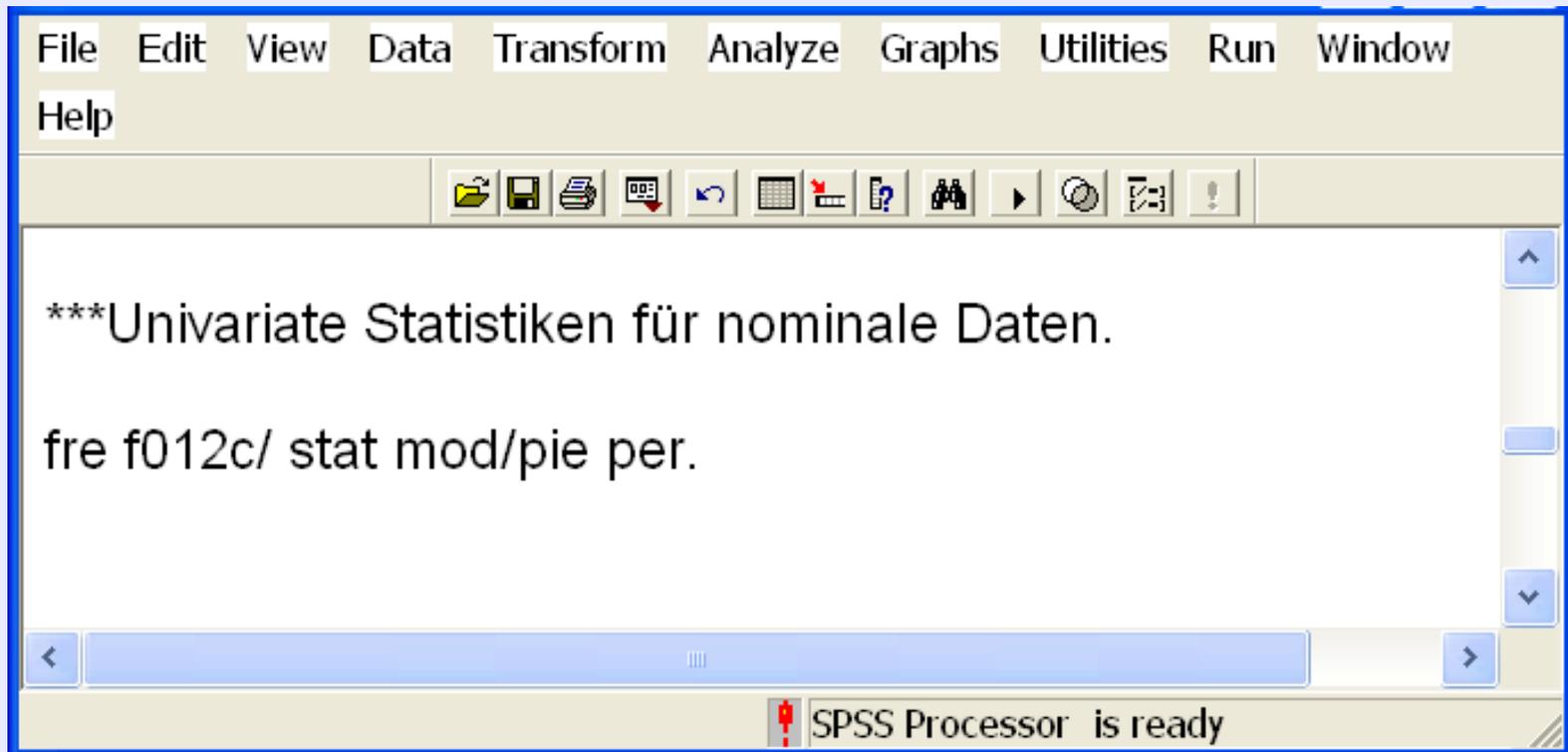
- ▶ Analysieren ▶ Deskriptive Statistiken
 - ▶ Explorative Datenanalyse

1. 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}]
```

1. FREQUENCIES (Nominale Daten)

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



```
File Edit View Data Transform Analyze Graphs Utilities Run Window
Help
***Univariate Statistiken für nominale Daten.
fre f012c/ stat mod/pie per.
SPSS Processor is ready
```

Statistiken

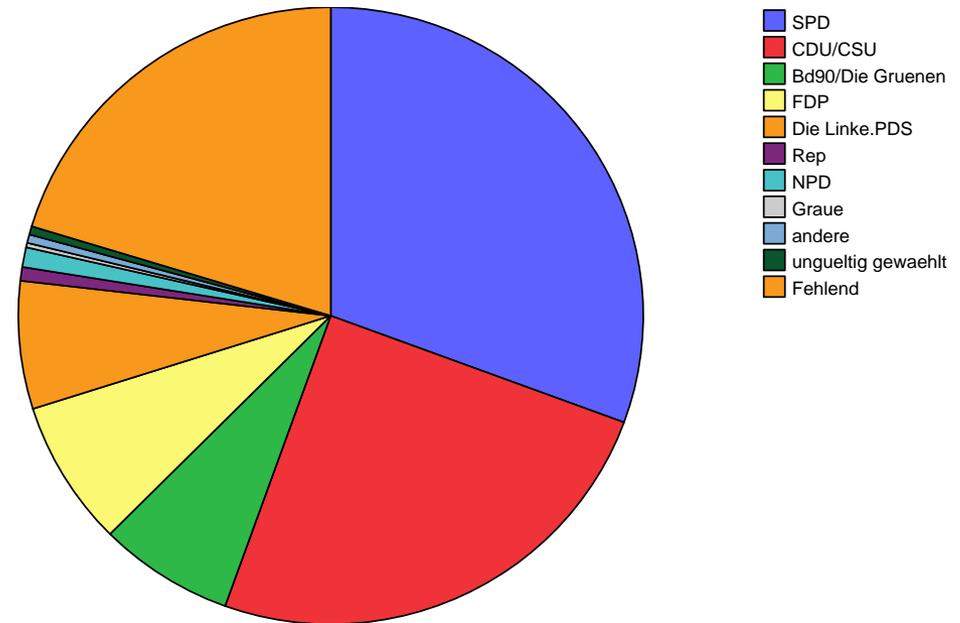
f012c BTW 05 Nachwahlbefragung Zweitstimme

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

1. FREQUENCIES (Ordinale Daten)

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

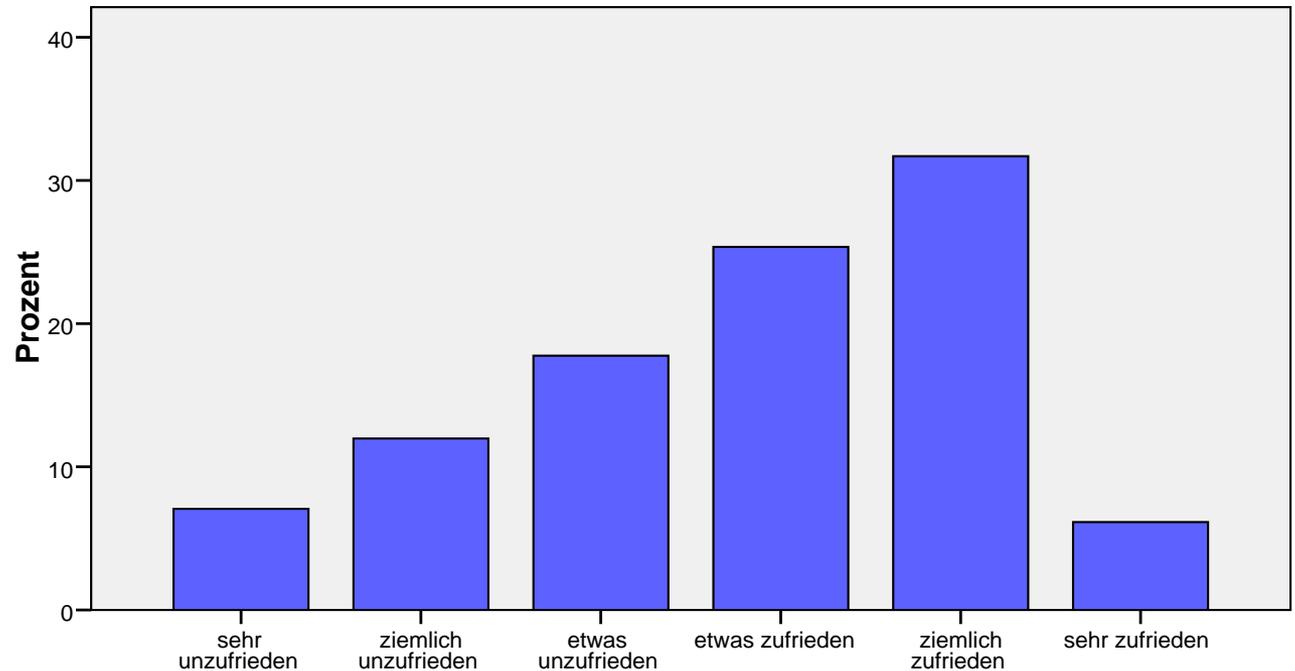
```
File Edit View Data Transform Analyze Graphs Utilities Run Window Help  
***Univariate Statistiken für ordinale Daten.  
fre f003/format notable/nti 4/stat min max med mod/bar per.  
SPSS Processor is ready
```

Statistiken

f003 Demokratie
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

2. 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.
    
```

2. EXAMINE (Metrische Daten)

```
***Univariate Statistiken für metrische Daten.  
exa f057_c/plot hist/stat des ext.
```

SPSS Processor is ready

Deskriptive Statistik

			Statistik	Standardfehler	
f0 gl	f057_c Merkel: glaubwuerdig	Mittelwert	3,17	,032	nt
		95% Konfidenzintervall des Mittelwerts	3,11		0%
		Untergrenze			
		Obergrenze	3,24		
		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

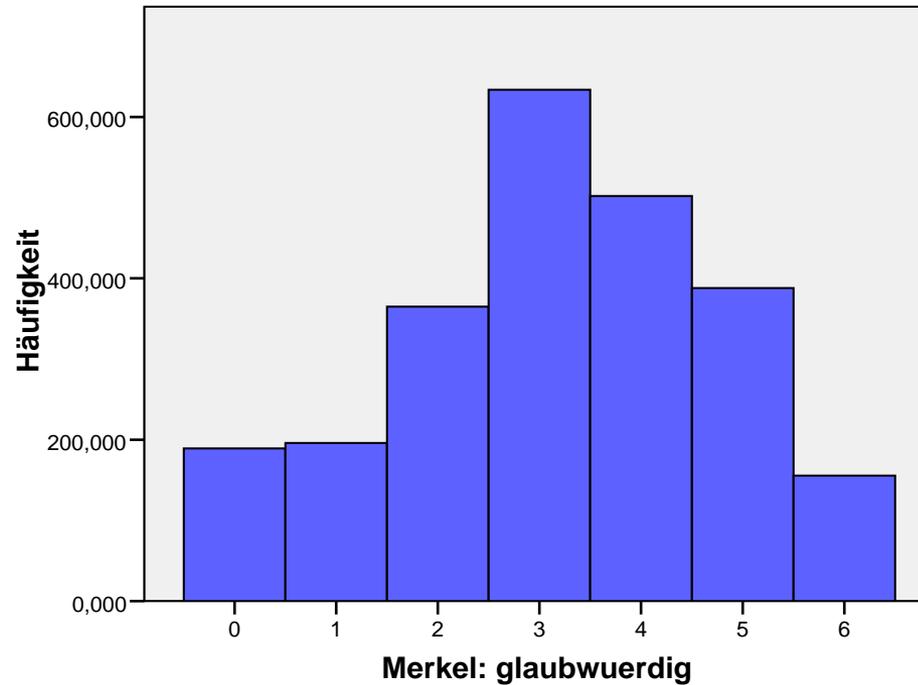
	Fallnummer	Wert
--	------------	------

f057_c Merkel: Größte Wert
glaubwuerdig

Kleinste Wert

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

Histogramm



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

Fälle gewichtet nach proportionales Personengewicht

Übungsaufgaben 4 und 5 (Allbus 2004)