

Mittelwerte, Zusammenhangsmaße, Hypothesentests in Stata

Statistik II

Wiederholung: Daten & Deskriptive Statistik

Zusammenhangsmaße

- Zwei nominale Variablen

- Zwei ordinale Variablen

- Nominal/intervallskalierte Variablen

- Zwei intervallskalierte Variablen

Inferenzstatistik

- Konfidenzintervalle

- Hypothesentests

Zusammenfassung

Daten verwalten

- ▶ Datenmatrix aus Fällen (Zeilen) und Variablen (Spalten)
- ▶ Neue Variablen erzeugen mit `generate`
- ▶ Vorhandene Variablen verändern mit `replace`
- ▶ Bedingungen für Befehle formulieren mit `if` (z. B. `if v26<8`)
- ▶ Beispiel: Geburtsjahr → Alter in Teilnehmerbefragung inkl. Fehlerkorrektur

Beispiel: Geburtsjahr/Alter

```
. tab geburtsjahr
```

geburtsjahr	Freq.	Percent	Cum.
1983	2	3.23	3.23
1984	1	1.61	4.84
1985	6	9.68	14.52
1986	5	8.06	22.58
1987	22	35.48	58.06
1988	21	33.87	91.94
1989	4	6.45	98.39
19888	1	1.61	100.00
Total	62	100.00	

```
. replace geburtsjahr = 1988 if geburtsjahr == 19888  
(1 real change made)
```

```
. tab geburtsjahr
```

geburtsjahr	Freq.	Percent	Cum.
1983	2	3.23	3.23
1984	1	1.61	4.84
1985	6	9.68	14.52
1986	5	8.06	22.58
1987	22	35.48	58.06
1988	22	35.48	93.55
1989	4	6.45	100.00
Total	62	100.00	

Beispiel: Geburtsjahr/Alter

```
. gen alter = 2009-geburtsjahr
```

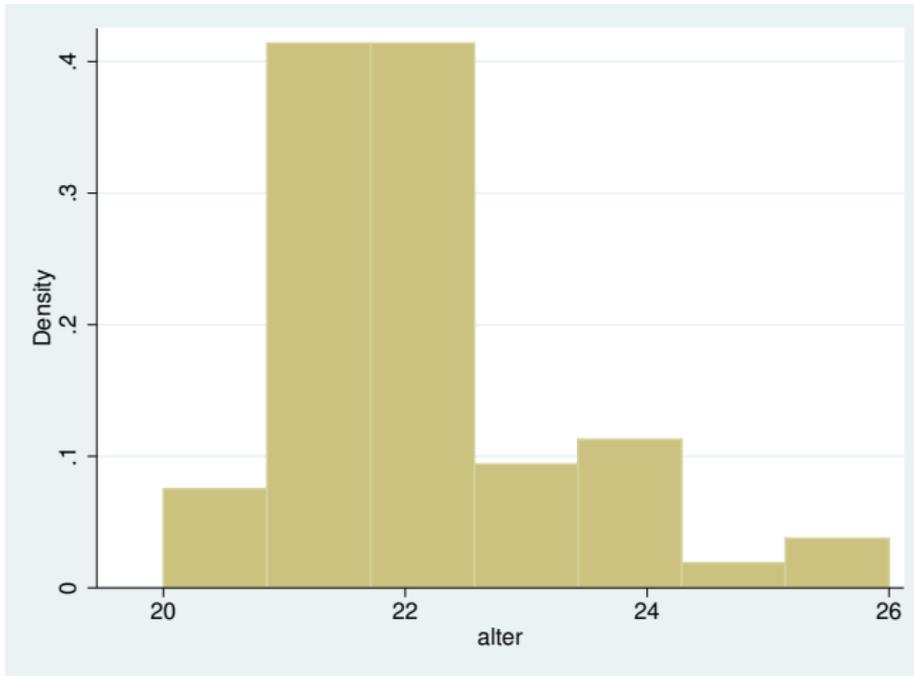
```
(8 missing values generated)
```

```
. tab alter
```

alter	Freq.	Percent	Cum.
20	4	6.45	6.45
21	22	35.48	41.94
22	22	35.48	77.42
23	5	8.06	85.48
24	6	9.68	95.16
25	1	1.61	96.77
26	2	3.23	100.00
Total	62	100.00	

```
. hist alter
```

Histogramm Alter



Verteilung wie beschreiben?

- ▶ Mittelwerte

Verteilung wie beschreiben?

- ▶ Mittelwerte
 - ▶ Modus
 - ▶ Median
 - ▶ (Perzentile)
 - ▶ Arithmetisches Mittel

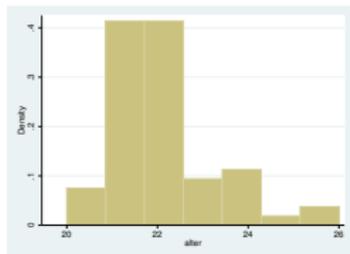
Verteilung wie beschreiben?

- ▶ Mittelwerte
 - ▶ Modus
 - ▶ Median
 - ▶ (Perzentile)
 - ▶ Arithmetisches Mittel
- ▶ Streuungsmaße

Verteilung wie beschreiben?

- ▶ Mittelwerte
 - ▶ Modus
 - ▶ Median
 - ▶ (Perzentile)
 - ▶ Arithmetisches Mittel
- ▶ Streuungsmaße
 - ▶ Spannweite (range)
 - ▶ Varianz
 - ▶ Standardabweichung

Alter deskriptiv

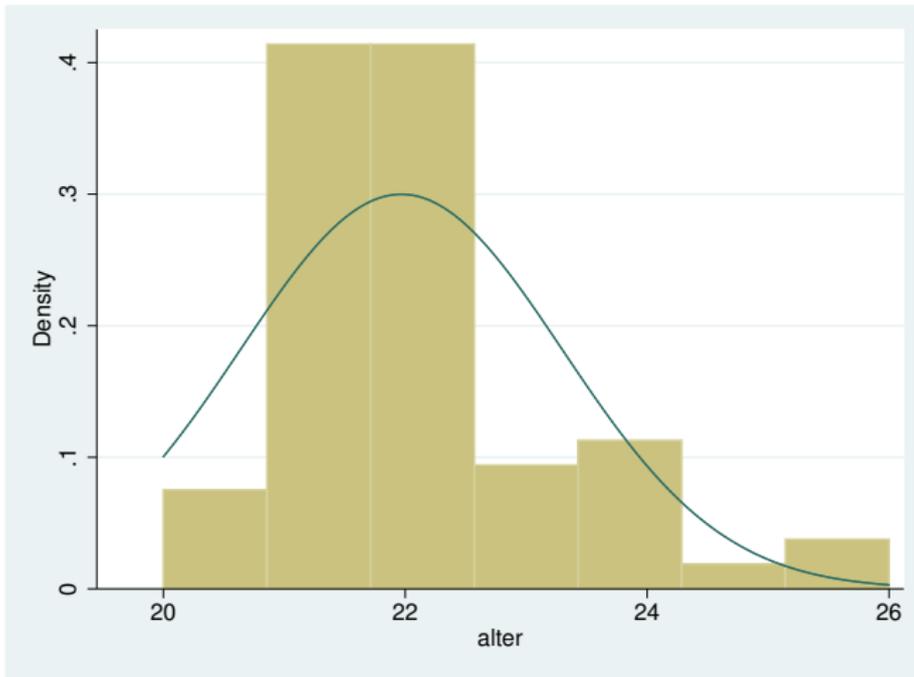


```
. summarize alter,detail
```

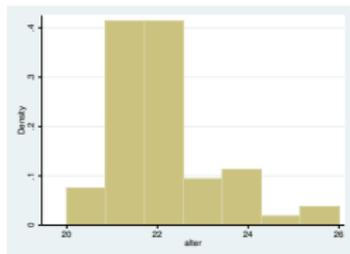
		alter			
Percentiles		Smallest			
1%	20	20		Obs	62
5%	20	20		Sum of Wgt.	62
10%	21	20		Mean	21.96774
25%	21	20		Std. Dev.	1.330201
50%	22	Largest		Variance	1.769434
75%	22	24		Skewness	1.15472
90%	24	25		Kurtosis	4.260386
95%	24	26			
99%	26	26			

- ▶ Verteilung rechtsschief/linkssteil: positive skewness
- ▶ Kurtosis > 0 : schmaler Gipfel (verglichen mit Normalverteilung)
- ▶ Modus?

Alter deskriptiv



Alter deskriptiv



```
. summarize alter,detail
```

		alter			
Percentiles		Smallest			
1%	20	20		Obs	62
5%	20	20		Sum of Wgt.	62
10%	21	20		Mean	21.96774
25%	21	20		Std. Dev.	1.330201
50%	22	Largest		Variance	1.769434
75%	22	24		Skewness	1.15472
90%	24	25		Kurtosis	4.260386
95%	24	26			
99%	26	26			

- ▶ Verteilung rechtsschief/linkssteil: positive skewness
- ▶ Kurtosis > 0 : schmaler Gipfel (verglichen mit Normalverteilung)
- ▶ Modus?

Beispiel: Geburtsjahr/Alter

```
. gen alter = 2009-geburtsjahr  
(8 missing values generated)
```

```
. tab alter
```

alter	Freq.	Percent	Cum.
20	4	6.45	6.45
21	22	35.48	41.94
22	22	35.48	77.42
23	5	8.06	85.48
24	6	9.68	95.16
25	1	1.61	96.77
26	2	3.23	100.00
Total	62	100.00	

```
. hist alter
```

Was ist ein Zusammenhang?

- ▶ Allgemein: gemeinsames „Muster“ in der Verteilung zweier Variablen (kausal?)
- ▶ Skalenniveaus – Zusammenhangsmaße
- ▶ Gerichtete vs. ungerichtete Zusammenhänge

Maße auf der Basis von χ^2

- ▶ Vergleich empirische Tabelle – Indifferenztablelle
- ▶ Für jede Zelle Differenz zwischen beobachteten/erwarteten Werten ermitteln
- ▶ Abweichungen quadrieren
- ▶ Quadrierte Abweichungen durch erwartete Werte teilen
- ▶ Summe der Beiträge: χ^2
 - ▶ Wert zwischen 0 und $+\infty$
 - ▶ Abhängig von Fallzahl
 - ▶ Stärke des Zusammenhangs
 - ▶ Kategorienzahl
- ▶ Cramer's V, ϕ , $C = \sqrt{\frac{\chi^2}{n \times (R-1)}}$
- ▶ λ ?

Zusammenhang Erst- und Zweitstimme?

```
/*Kreuztabelle Erst- und Zweitstimme mit Indifferenztafel*/  
tab erststimme zweitstimme,exp
```

erststimme	zweitstimme						Total
	FDP	Grüne	SPD	Piraten	CDU/CSU	Linke	
FDP	4 1.0	0 1.8	0 1.6	0 0.7	2 0.8	0 0.1	6 6.0
Grüne	2 2.9	11 5.5	2 4.9	3 2.0	0 2.3	0 0.3	18 18.0
SPD	1 4.4	7 8.3	14 7.4	3 3.0	1 3.5	1 0.4	27 27.0
Linke	0 0.2	0 0.3	0 0.3	1 0.1	0 0.1	0 0.0	1 1.0
CDU/CSU	2 1.5	1 2.8	1 2.5	0 1.0	5 1.2	0 0.1	9 9.0
Sonstige Parteien	1 0.2	0 0.3	0 0.3	0 0.1	0 0.1	0 0.0	1 1.0
Total	10 10.0	19 19.0	17 17.0	7 7.0	8 8.0	1 1.0	62 62.0

```
/* stunde2.do ends here */  
end of do-file
```

Zusammenhang Erst- und Zweitstimme?

Key								
Frequency								
column percentage								
erststimme		zweitstimme					Total	
	FDP	Grüne	SPD	Piraten	CDU/CSU	Linke		
FDP	4 40.00	0 0.00	0 0.00	0 0.00	2 25.00	0 0.00	6 9.68	
Grüne	2 20.00	11 57.09	2 11.76	3 42.86	0 0.00	0 0.00	18 29.03	
SPD	1 10.00	7 36.84	14 82.35	3 42.86	1 12.50	1 100.00	27 43.55	
Linke	0 0.00	0 0.00	0 0.00	1 14.29	0 0.00	0 0.00	1 1.61	
CDU/CSU	2 20.00	1 5.26	1 5.88	0 0.00	5 62.50	0 0.00	9 14.52	
Sonstige Parteien	1 10.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1 1.61	
Total	10 100.00	19 100.00	17 100.00	7 100.00	8 100.00	1 100.00	62 100.00	

Pearson chi2(25) = 66.5231 Pr = 0.000
 Cramér's V = 0.4632

Ordinale Zusammenhänge

- ▶ Zwei ordinale Variablen → Richtung
 - ▶ Mehr x , mehr y ; weniger x , weniger y → positiver Zusammenhang
 - ▶ Mehr x , weniger y ; weniger x , mehr y → negativer Zusammenhang

Ordinale Zusammenhänge

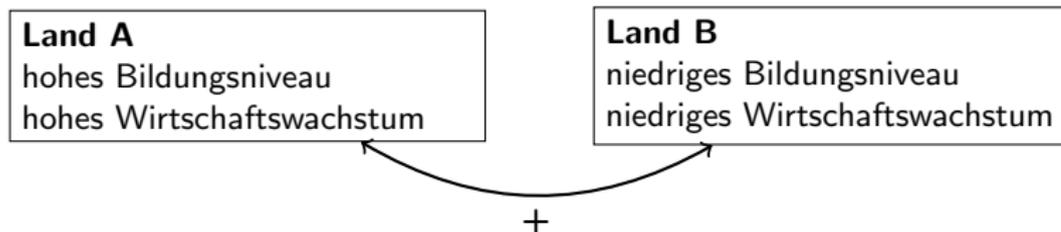
- ▶ Zwei ordinale Variablen → Richtung
 - ▶ Mehr x , mehr y ; weniger x , weniger y → positiver Zusammenhang
 - ▶ Mehr x , weniger y ; weniger x , mehr y → negativer Zusammenhang
- ▶ Wie mißt man das?

Ordinale Zusammenhänge

- ▶ Zwei ordinale Variablen → Richtung
 - ▶ Mehr x , mehr y ; weniger x , weniger y → positiver Zusammenhang
 - ▶ Mehr x , weniger y ; weniger x , mehr y → negativer Zusammenhang
- ▶ Wie mißt man das?
- ▶ Vergleich von Paaren von Beobachtungen

Ordinale Zusammenhänge

- ▶ Zwei ordinale Variablen → Richtung
 - ▶ Mehr x , mehr y ; weniger x , weniger y → positiver Zusammenhang
 - ▶ Mehr x , weniger y ; weniger x , mehr y → negativer Zusammenhang
- ▶ Wie mißt man das?
- ▶ Vergleich von Paaren von Beobachtungen



Paarvergleich und Berechnung von γ

- ▶ *Konkordantes* Paar $A \Leftrightarrow B$: B hat mehr von x (z. B. Bildung) und mehr von y (z. B. politisches Interesse) als A
- ▶ *Diskonkordantes* Paar $A \Leftrightarrow B$: B hat mehr von x (z. B. Bildung) als A , aber *weniger* von y (z. B. politisches Interesse) als A
- ▶ γ : Verhältnis konkordante – diskonkordante Paare
 - ▶ Konkordante Paare überwiegen: positiver Zusammenhang
 - ▶ Diskonkordante Paare überwiegen: negativer Zusammenhang
- ▶ Paare mit identischen Werten für eine oder beide Variablen: „ties“
- ▶ Werden bei γ ignoriert
- ▶ $\tau_b = \frac{N_C - N_D}{\sqrt{(N_C + N_D + T_x) \times (N_C + N_D + T_Y)}}$ (konkordante Paare im Nenner berücksichtigt)

Wie konsistent sind wirtschaftsliberale Einstellungen?

- ▶ „Politik sollte sich aus Wirtschaft heraushalten“
- ▶ „Weitere Öffnung der Weltmärkte dient Wohl aller“
- ▶ Problem: Codierungen

```
. gen politikraus= themenidiepolitiksolntesichausde
(3 missing values generated)
. gen globalisierunggut = themenidieweiteroeffnungderweltr
(3 missing values generated)
. tab themenidieweiteroeffnungderweltr
themen1 [Die weitere Öffnung
der Weltmärkte dient dem
Wohl aller.]
```

	Freq.	Percent	Cum.
1. Stimme eher nicht zu	21	31.34	31.34
2. Stimme eher zu	21	31.34	62.69
3. Stimme voll und ganz zu	6	8.96	71.64
4. Weder noch	14	20.90	92.54
5. Stimme ueberhaupt nicht zu	5	7.46	100.00
Total	67	100.00	

```
. tab themenidiepolitiksolntesichausde
themen1 [Die Politik sollte
sich aus der Wirtschaft
heraushalten.]
```

	Freq.	Percent	Cum.
1. Stimme ueberhaupt nicht zu	23	34.33	34.33
2. Stimme eher zu	15	22.39	56.72
3. Stimme eher nicht zu	18	26.87	83.58
4. Weder noch	10	14.93	98.51
5. Stimme voll und ganz zu	1	1.49	100.00
Total	67	100.00	

```
. recode globalisierunggut (1=2) (2=4) (3=5) (4=3) (5=1)
(globalisierunggut: 67 changes made)
. recode politikraus (1=1) (3=2) (4=3) (2=4) (5=5)
(politikraus: 43 changes made)
```

Wie konsistent sind wirtschaftsliberale Einstellungen?

```
. tab politikraus globalisierunggut, gamma taub
```

politikraus	globalisierunggut					Total
s	1	2	3	4	5	
1	1	13	1	6	2	23
2	3	3	3	7	2	18
3	0	1	5	3	1	10
4	1	3	5	5	1	15
5	0	1	0	0	0	1
Total	5	21	14	21	6	67

gamma = 0.1297 ASE = 0.131
Kendall's tau-b = 0.0997 ASE = 0.101

- ▶ Zusammenhang in der richtigen Richtung
- ▶ Nicht besonders stark
- ▶ Konkordante Paare und ties

Warum?

- ▶ Vergleich eines intervallskalierten Merkmals
- ▶ Über zwei oder mehr Gruppen (nominalskalierte Variable)
- ▶ Sind weibliche Teilnehmer jünger (mangels Wehrpflicht)?
- ▶ $\eta^2 = \frac{SAQ_{gesamt} - SAQ_{Kategorien}}{SAQ_{gesamt}}$

```
. tabstat alter ,by(geschlecht) stat (mean n)
```

Summary for variables: alter
 by categories of: geschlecht

geschlecht	mean	N
Nicht zutreffend	22	2
maennlich	22.26316	38
weiblich	21.45455	22
Total	21.96774	62

```
. anova alter geschlecht
```

Source	Partial SS	df	MS	F	Prob > F
Model	9.11251736	2	4.55625868	2.72	0.0741
geschlecht	9.11251736	2	4.55625868	2.72	0.0741
Residual	98.8229665	59	1.67496553		
Total	107.935484	61	1.76943416		

Kovarianz und Korrelation?

- ▶ Varianz: Abweichung einer Variablen von ihrem Mittelwert
- ▶ Kovarianz: *gemeinsame* Abweichung zweier Variablen von ihren Mittelwerten
- ▶ Linearer Zusammenhang
 - ▶ Positiver Zusammenhang: überdurchschnittliche Werte von x , überdurchschnittliche Werte von y und umgekehrt
 - ▶ Negativer Zusammenhang: überdurchschnittliche Werte von x , *unter*durchschnittliche Werte von y und umgekehrt
- ▶ Abweichungsprodukte \rightarrow Kovarianz zwischen $-\infty$ und $+\infty$
- ▶ Hängt ab von Stärke des Zusammenhangs und Skalierung
- ▶ Teilen durch Produkt der Standardabweichung \rightarrow Korrelationskoeffizient r

Zusammenhang zwischen Bearbeitungsdauer und politischem Wissen?

- ▶ Besonders informierte Studierende besonders schnell?
- ▶ Oder besonders langsam? → Zeitdauer + Wissenindex

Bearbeitungsdauer berechnen

```
.  
. keep if abgeschlossen =="Y"  
(7 observations deleted)  
  
.   
. /*Zeitangaben aus Datensatz in internes Format bringen und formatieren*/  
.   
. gen beginn = clock(datumgestartet , "DM20Yhm")  
. gen ende = clock(datumletzteaktivitt , "DM20Yhm")  
. format %tc beginn  
. format %tc ende  
  
.   
. /*Bearbeitungszeit (in Millisekunden) errechnen*/  
.   
. gen dauer=ende-beginn  
. /*Umrechnen in Minuten*/  
. gen minuten = dauer/60000  
  
. 
```

Bearbeitungsdauer berechnen

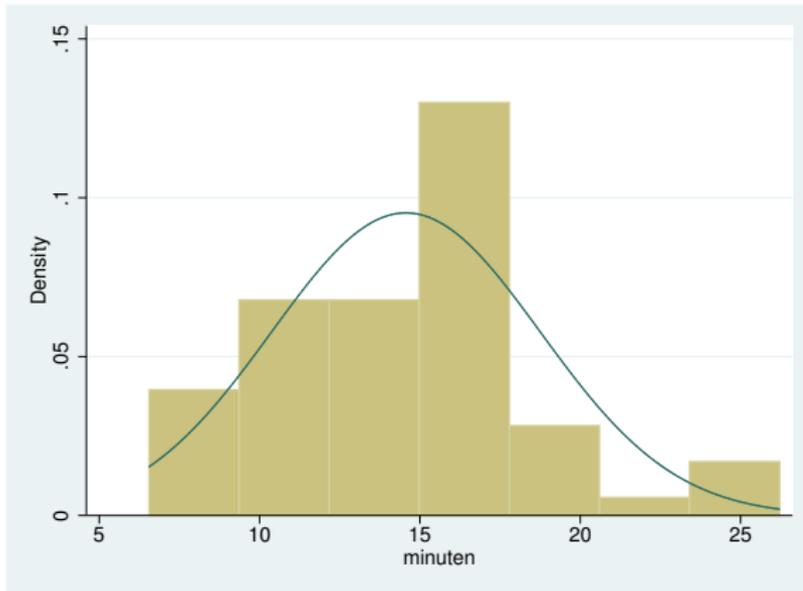
```
. list beginn ende dauer minuten in 1/10
```

```
1
```

	beginn	ende	dauer	minuten
1.	27oct2009 22:48:00	27oct2009 22:58:56	655360	10.92267
2.	27oct2009 22:52:22	27oct2009 23:01:07	524288	8.738133
3.	27oct2009 23:07:40	27oct2009 23:14:13	393216	6.5536
4.	27oct2009 23:44:48	28oct2009 00:02:17	1048576	17.47627
5.	28oct2009 00:00:06	28oct2009 00:13:12	786432	13.1072
6.	28oct2009 00:52:31	28oct2009 01:10:00	1048576	17.47627
7.	28oct2009 08:20:21	28oct2009 08:37:50	1048576	17.47627
8.	28oct2009 08:33:28	28oct2009 08:50:56	1048576	17.47627
9.	28oct2009 08:44:23	28oct2009 08:59:40	917504	15.29173
10.	28oct2009 08:53:07	28oct2009 09:08:25	917504	15.29173

```
1
```

Bearbeitungsdauer berechnen



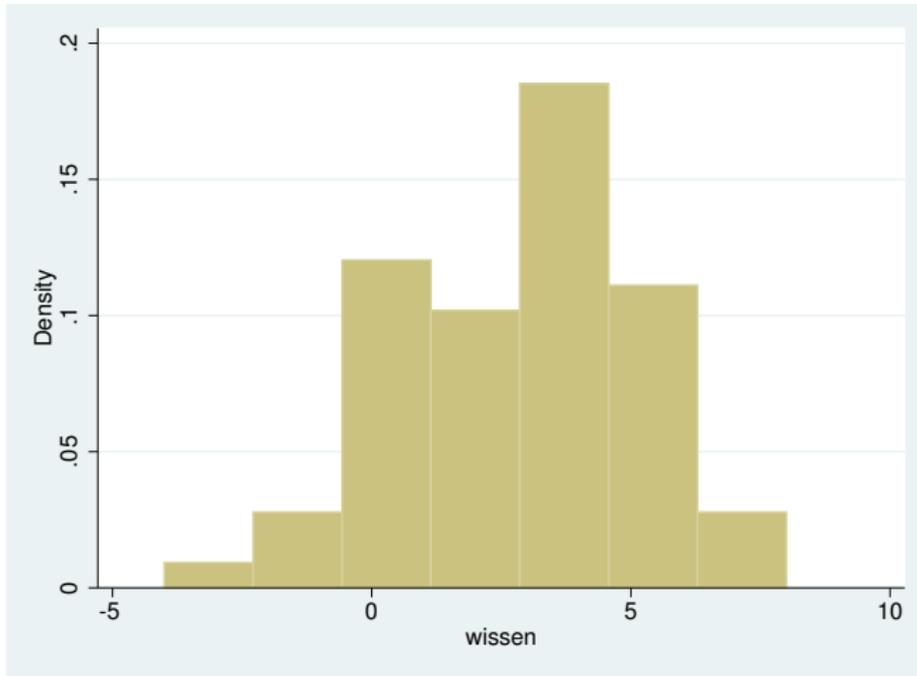
Index für Wissen EU-Länder berechnen

- ▶ Richtige Antwort +1 Punkt
- ▶ Falsche Antwort -1 Punkt
- ▶ „Unsicher“ oder keine Antwort (missing) 0 Punkt

```
. gen wissen=0
.
. /*Schleife fuer Laender, die EU-Mitglieder sind*/
. /*1 =ja, 2= nein, 3= unsicher*/
.
.   foreach land of varlist eucountriesungarn eucountriesirland eucountrieslit
> auen eucountriesmalta eucountrieszypern eucountrieschweden {
2.   replace wissen = wissen + 1 if `land' == 1
3.   replace wissen = wissen - 1 if `land' == 2
4.   }
(53 real changes made)
(4 real changes made)
(60 real changes made)
(2 real changes made)
(53 real changes made)
(6 real changes made)
(54 real changes made)
(7 real changes made)
(49 real changes made)
(9 real changes made)
(7 real changes made)
(53 real changes made)

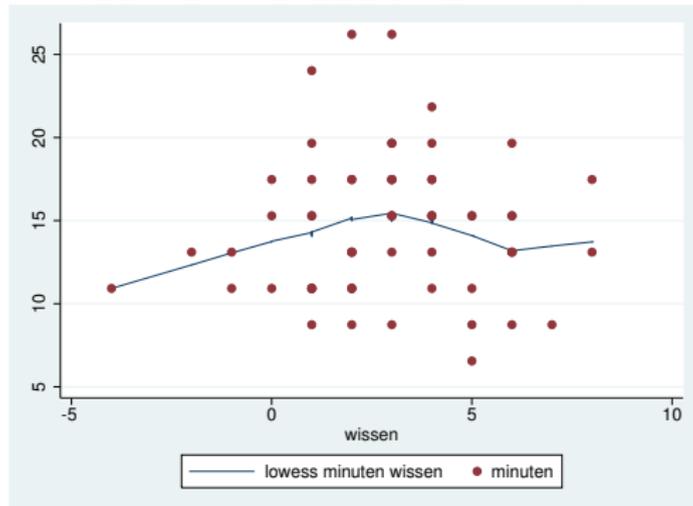
.
. /*Schleife fuer Laender, die keine EU-Mitglieder sind*/
.   foreach land of varlist eucountriesrkei eucountriesnorwegen eucountrieskro
> atien eucountriesukraine eucountriesgeorgien {
2.   replace wissen = wissen + 1 if `land' == 2
3.   replace wissen = wissen - 1 if `land' == 1
4.   }
(0 real changes made)
(62 real changes made)
(10 real changes made)
```

Index für Wissen EU-Länder berechnen



Streudiagramm Wissen – Bearbeitungsdauer

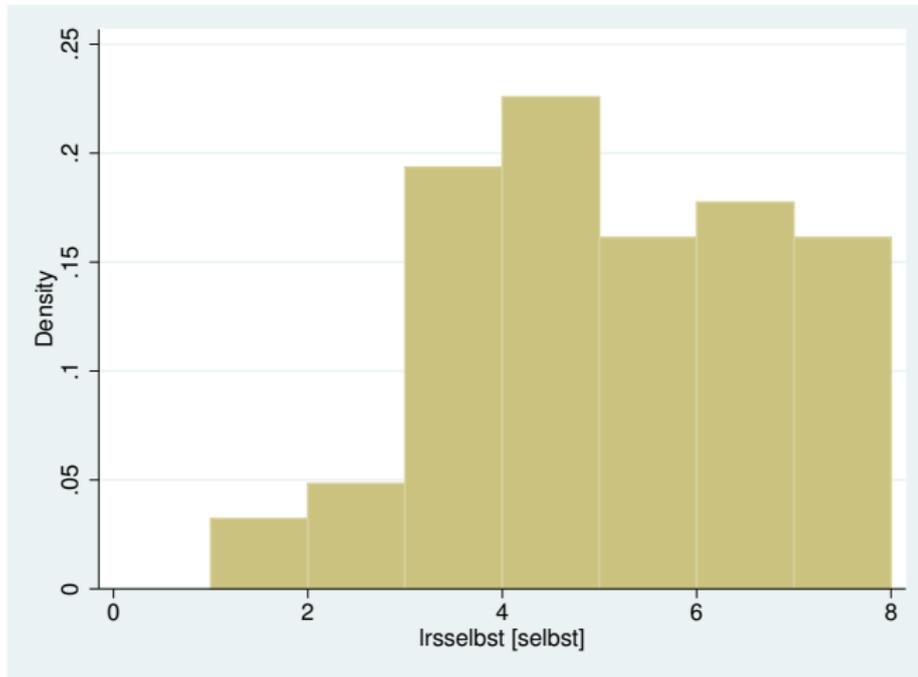
```
. graph twoway (lowess minuten wissen) (scatter  
minuten wissen)  
. corr wissen minuten →0.03
```



Wie berechnet man Konfidenzintervalle?

- ▶ Voraussetzung: Zufallsstichprobe (hier *nicht wirklich* erfüllt)
- ▶ Wenn Zufallsstichprobe, wird sich Stichprobenwert (z. B. Mittelwert) über unendlich viele Stichproben mit Umfang n in regelmäßiger Weise verteilen → theoretische Verteilung, Standardfehler
- ▶ Konfidenzintervall:
 - ▶ Ausgangspunkt: eine tatsächliche vorhandene Stichprobe
 - ▶ α festlegen
 - ▶ Für z. B. 95% aller Stichproben schließt Intervall wahren Mittelwert ein
 - ▶ Für alle Stichprobenkennwerte berechenbar (wenn Standardfehler bekannt)
 - ▶ (Meistens) symmetrisch

LRS Umfrage



LRS Umfrage

```
. summ lrsselbstselbst ,det
```

```
lrsselbst [selbst]
```

Percentiles		Smallest		
1%	1	1		
5%	2	1		
10%	3	2	Obs	62
25%	3	2	Sum of Wgt.	62
50%	4.5		Mean	4.677419
		Largest	Std. Dev.	1.744069
75%	6	8		
90%	7	8	Variance	3.041777
95%	8	8	Skewness	.0923525
99%	8	8	Kurtosis	2.365833

```
. ci lrsselbstselbst
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
lrsselbsts-t	62	4.677419	.221497	4.234509	5.12033

Logik des Hypothesentests

- ▶ Nullhypothese vs. Alternativhypothese
- ▶ Voraussetzung: Zufallsstichprobe (hier *nicht wirklich* erfüllt)
- ▶ Wenn Zufallsstichprobe, wird sich Stichprobenwert (z. B. Mittelwert) über unendlich viele Stichproben mit Umfang n in regelmäßiger Weise verteilen → theoretische Verteilung, Standardfehler
- ▶ Vergleich Testergebnis mit theoretischer Verteilung (Modell für Stichprobenziehung unter H_0)
- ▶ Wie wahrscheinlich ist Testergebnis wenn H_0 gilt bzw. Testergebnis unwahrscheinlicher als α ?

Hypothesentests aus Statistik I

1. χ^2 -Test (zwei nominalskalierte Variablen)
2. z-Test (Vergleich Stichprobenmittelwert mit bekannter Grundgesamtheit)
3. t-Test (Mittelwert zweier Gruppen in Grundgesamtheit identisch?)
 - 3.1 Für abhängige Stichproben (Wiederholungsmessung)
 - 3.2 Für unabhängige Stichproben (unterschiedliche Untersuchungsobjekte)

Logik χ^2 -Test

- ▶ Abweichungen zwischen empirischer Kreuztabelle und Indifferenztablelle \rightarrow empirischer χ^2 -Wert
- ▶ Wenn in GG Nullhypothese gilt: Abweichungen aufgrund von Stichprobenfehlern \rightarrow empirischer χ^2 -Wert $\neq 0$
- ▶ Über viele Stichproben hinweg entspricht Verteilung empirischer χ^2 -Werte theoretischer χ^2 -Verteilung mit $(R - 1) \times (C - 1)$ Freiheitsgraden

Logik χ^2 -Test

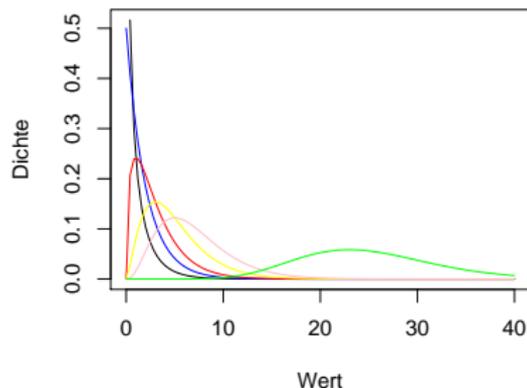
Key								
Frequency								
column percentage								
erststimme	zweitstimme							Total
	FDP	Grüne	SPD	Piraten	CDU/CSU	Linke		
FDP	4 40.00	0 0.00	0 0.00	0 0.00	2 25.00	0 0.00	6 9.68	
Grüne	2 20.00	11 57.09	2 11.76	3 42.86	0 0.00	0 0.00	18 29.03	
SPD	1 10.00	7 36.84	14 82.35	3 42.86	1 12.50	1 100.00	27 43.55	
Linke	0 0.00	0 0.00	0 0.00	1 14.29	0 0.00	0 0.00	1 1.61	
CDU/CSU	2 20.00	1 5.26	1 5.88	0 0.00	5 62.50	0 0.00	9 14.52	
Sonstige Parteien	1 10.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1 1.61	
Total	10 100.00	19 100.00	17 100.00	7 100.00	8 100.00	1 100.00	62	

Pearson chi2(25) = 66.5231 P< = 0.000
 Cramer's V = 0.4632

Wie hoch ist der kritische Wert?

► $\alpha = 0.05$, $df = 25$

```
. displ invchi2(1,0.95)  
3.8414588  
. displ invchi2(25,0.95)  
37.652484
```



z-Test

- Ist der Mittelwert von LRS tatsächlich < 6 ?

```
. ttest lrsselbstselbst =6
```

```
One-sample t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
lrssel-t	62	4.677419	.221497	1.744069	4.234509	5.12033

```

mean = mean(lrsselbstselbst)
Ho: mean = 6
Ha: mean < 6
Pr(T < t) = 0.0000

t = -5.9711
degrees of freedom = 61
Ha: mean != 6
Pr(|T| > |t|) = 0.0000
Ha: mean > 6
Pr(T > t) = 1.0000
    
```

LRS Umfrage

```
. summ lrsselbstselbst ,det
```

```
lrsselbst [selbst]
```

Percentiles		Smallest		
1%	1	1		
5%	2	1		
10%	3	2	Obs	62
25%	3	2	Sum of Wgt.	62
50%	4.5		Mean	4.677419
		Largest	Std. Dev.	1.744069
75%	6	8		
90%	7	8	Variance	3.041777
95%	8	8	Skewness	.0923525
99%	8	8	Kurtosis	2.365833

```
. ci lrsselbstselbst
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
lrsselbsts-t	62	4.677419	.221497	4.234509	5.12033

t-Test

- ▶ Unterscheiden sich Gruppen bzw. sind Gruppenmittelwerte identisch?
- ▶ Frauen weniger radikal (weniger links) als Männer?

```
. tab geschlecht
```

geschlecht	Freq.	Percent	Cum.
Nicht zutreffend	2	3.17	3.17
maennlich	39	61.90	65.08
weiblich	22	34.92	100.00
Total	63	100.00	

```
. gen frau=.
(63 missing values generated)
. replace frau=0 if geschlecht ==2
(39 real changes made)
. replace frau=1 if geschlecht ==3
(22 real changes made)
```

```
. tab frau
```

frau	Freq.	Percent	Cum.
0	39	63.93	63.93
1	22	36.07	100.00
Total	61	100.00	

t-Test

- ▶ Unterscheiden sich Gruppen bzw. sind Gruppenmittelwerte identisch?
- ▶ Frauen weniger radikal (weniger links) als Männer?

```
. ttest lrsselbstselbst ,by(frau)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	38	4.894737	.308231	1.900064	4.270201	5.519272
1	22	4.454545	.2995471	1.405	3.831603	5.077488
combined	60	4.733333	.2240695	1.735635	4.284971	5.181695
diff		.4401914	.4653922		-.4913922	1.371775

```
diff = mean(0) - mean(1)                                t = 0.9459
Ho: diff = 0                                             degrees of freedom = 58
Ha: diff < 0                                             Ha: diff != 0
Pr(T < t) = 0.8259                                       Pr(|T| > |t|) = 0.3481
                                                           Ha: diff > 0
                                                           Pr(T > t) = 0.1741
```

Literatur für nächste Woche (Regression)

- ▶ Berk (2004, S. 13-17, 39-56) und
- ▶ Fox (1997, S. 86-88, 101, 204-205, 212-213)
- ▶ (beides im ReaderPlus)

Zusammenfassung

- ▶ Fast alle Berechnungen aus Statistik I mit ein bis zwei Befehlen umsetzbar
- ▶ Wichtig:
 - ▶ Verstehen was man tut
 - ▶ Daten kontrollieren und ggf. umkodieren