



Workshop  
Auswertung von Statistiken mit Excel

*Hr. Prof. Weißbach, Fr. Strautz*



## Workshop Auswertung von Statistiken mit Excel **Gliederung**

- Excel: Die Basics
- Deskriptive univariate Statistik
  - Lageparameter, Streuungsmaße, Häufigkeitstabellen
- Bivariate Statistik
  - Chi<sup>2</sup>-Test, Kontingenzkoeffizient
- Ausblick:
  - Konfidenzintervalle, T-Test, Korrelation, ANOVA, Lineare Regression

## Die Basics

### Skalenniveaus- Welche Daten werte ich aus?

Skalenniveau	Ausprägung	Operationen	Beispiel
Nominal	Unterschied	Modalwert (Häufigkeit)	Haarfarbe, Augenfarbe
Ordinal	Ränge	Medianwert (Über 50%)	Schulnoten
Metrisch	Werte	Mittelwert, Abstände interpretierbar	Umsatz, Gehalt

2000m Lauf:  
Ergebnis

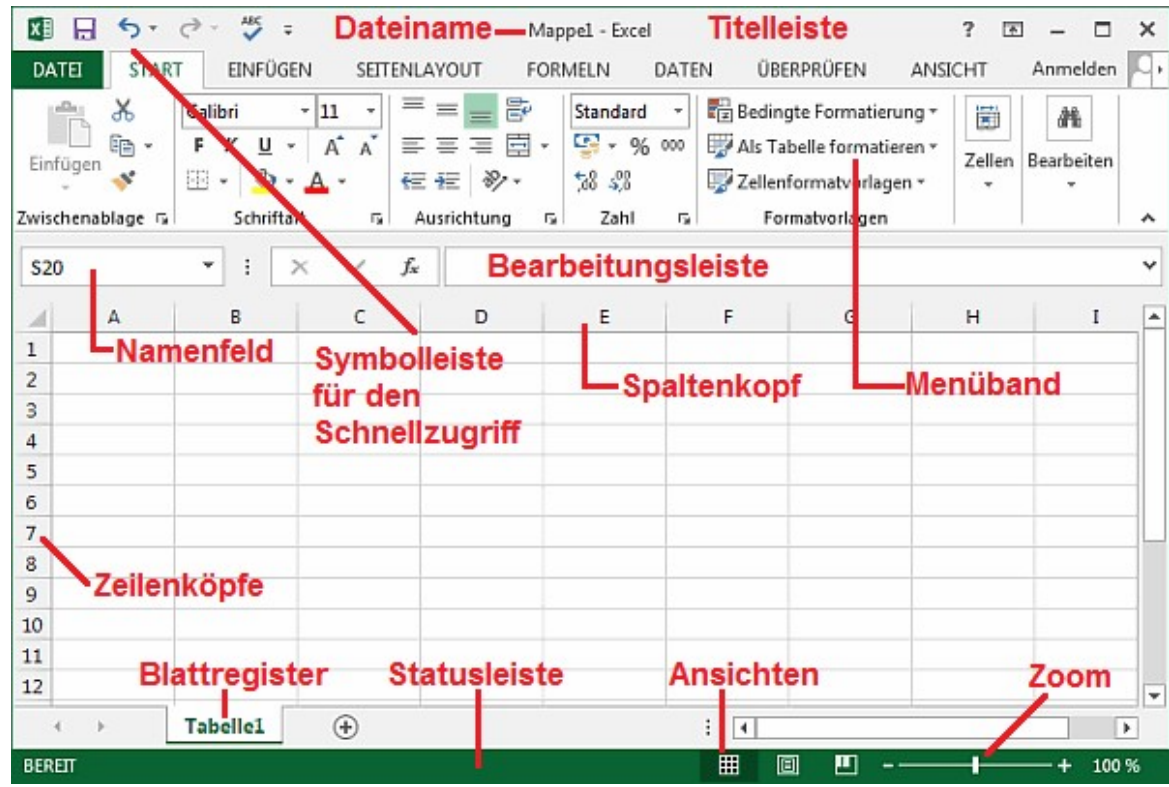
Gut/schlecht

Platz 1/2/3

11:17 Min

# Die Basics

Excel

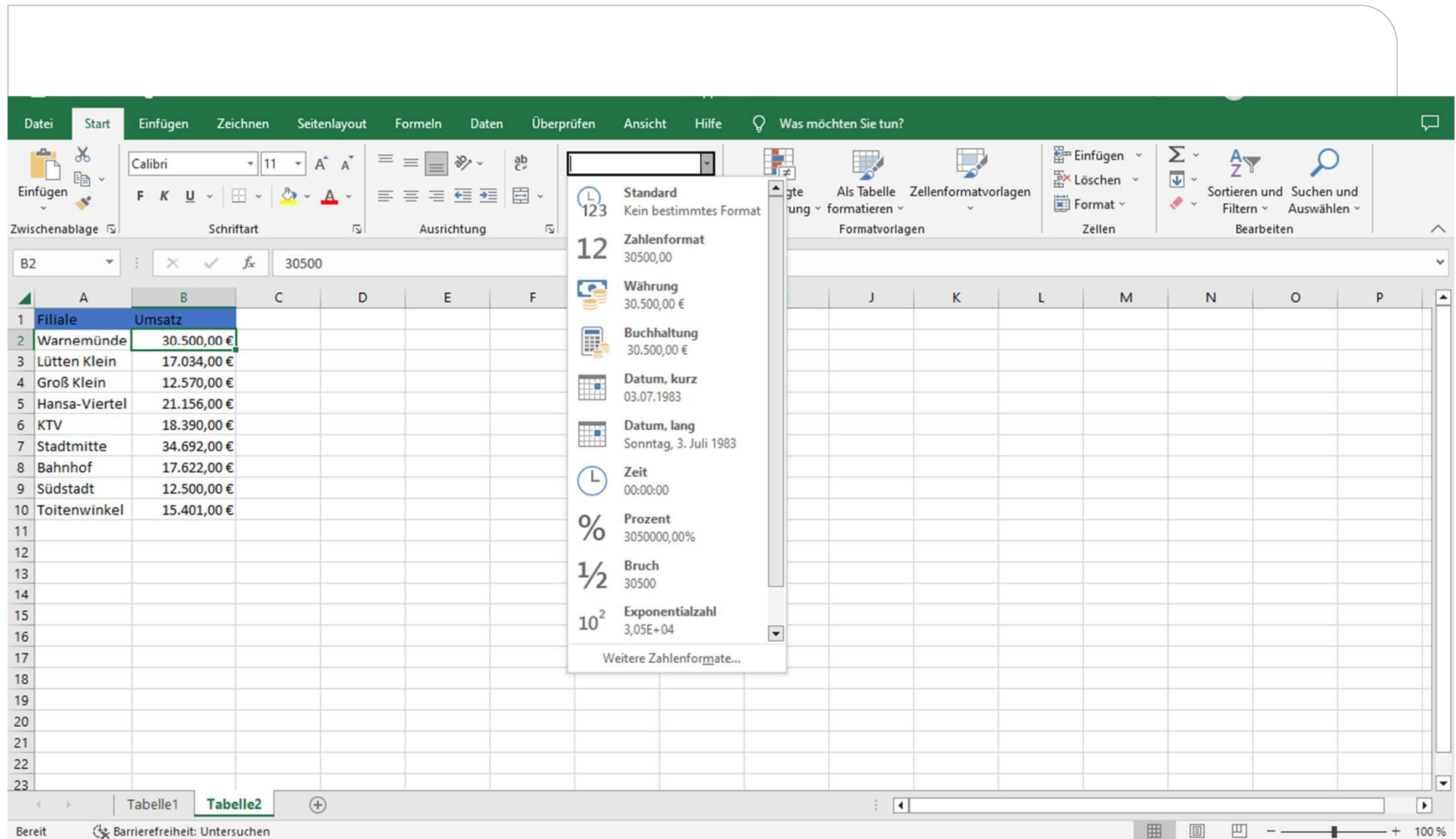


Bildquelle:  
<https://i.pinimg.com/originals/79/ad/c3/79adc3c1624efe201d977fc5c5b94078.png>

# Die Basics

## Zellen formatieren

The screenshot displays the 'Zellen formatieren' (Format Cells) dialog box in Microsoft Excel. The background shows a spreadsheet with columns 'Schüler' and 'Note' and rows 1-21. The 'Zellen formatieren' dialog is open, showing the 'Zahlen' (Number) tab. The 'Kategorie:' (Category) list has 'Zahl' (Number) selected. The 'Beispiel' (Preview) field shows '1,00'. The 'Dezimalstellen:' (Decimal places) spinner is set to '2'. The '1000er-Trennzeichen verwenden (,)' (Use thousands separator) checkbox is unchecked. The 'Negative Zahlen:' (Negative numbers) list has '-1234,10' selected. The dialog has 'OK' and 'Abbrechen' (Cancel) buttons.



The screenshot shows the Microsoft Excel interface. The ribbon is set to 'Start'. The active cell is B2, containing the value '30500'. A dropdown menu for 'Zahlenformat' (Number Format) is open, showing various options. The data table in the background is as follows:

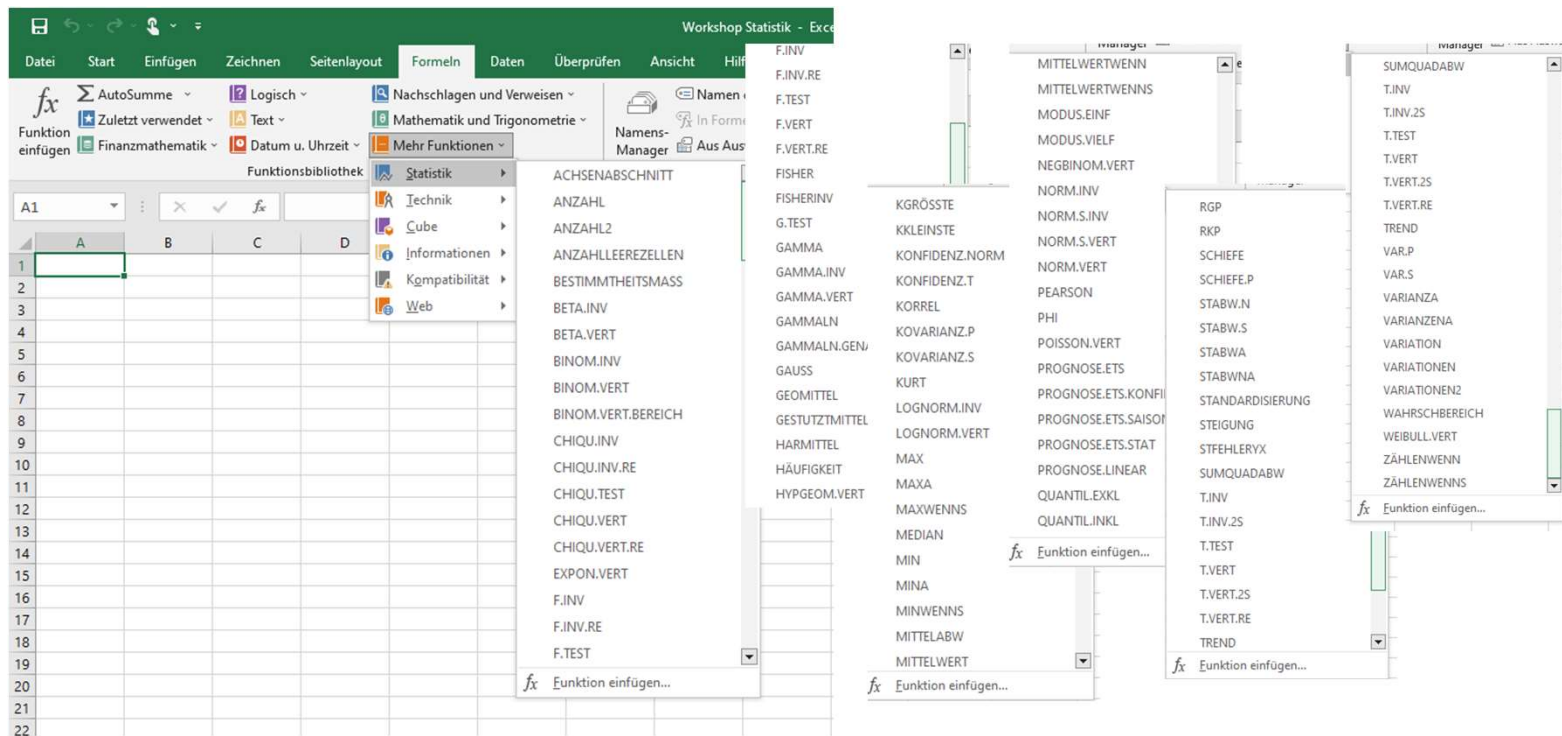
Filiale	Umsatz
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €

The 'Zahlenformat' dropdown menu includes the following options:

- Standard: Kein bestimmtes Format
- Zahlenformat: 30500,00
- Währung: 30.500,00 €
- Buchhaltung: 30.500,00 €
- Datum, kurz: 03.07.1983
- Datum, lang: Sonntag, 3. Juli 1983
- Zeit: 00:00:00
- Prozent: 3050000,00%
- Bruch: 30500
- Exponentialzahl: 3,05E+04

# Die Basics

## Statistik Funktionen in Excel

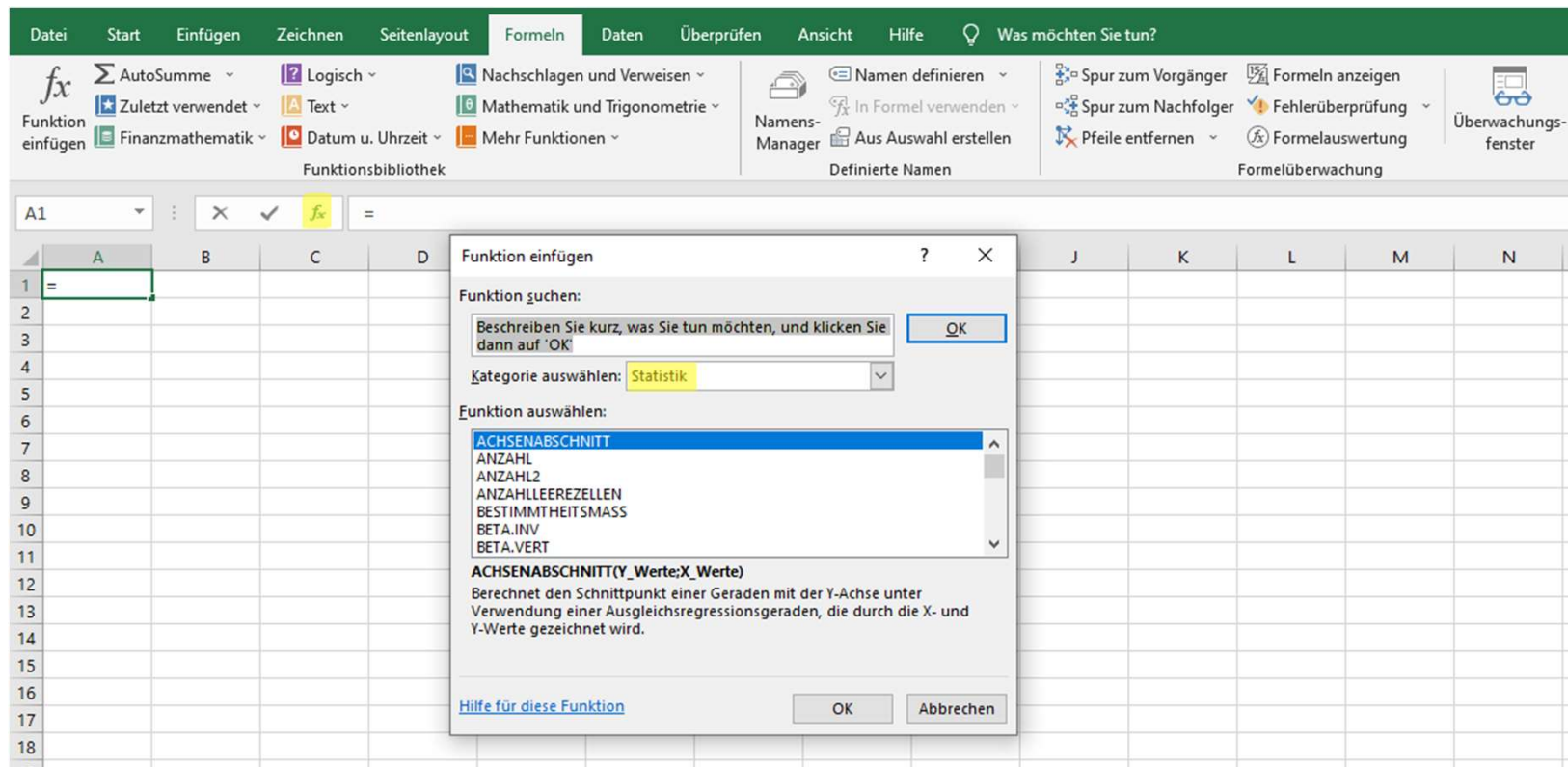


The screenshot shows the Excel interface with the 'Formeln' ribbon active. The 'Statistik' category in the 'Funktionsbibliothek' is expanded, showing a list of statistical functions. Three dropdown menus are open, displaying the following lists of functions:

- Dropdown 1 (left):**
  - ACHSENABSCHNITT
  - ANZAHL
  - ANZAHL2
  - ANZAHLLEEREZELLEN
  - BESTIMMTHEITSMASS
  - BETA.INV
  - BETA.VERT
  - BINOM.INV
  - BINOM.VERT
  - BINOM.VERT.BEREICH
  - CHIQU.INV
  - CHIQU.INV.RE
  - CHIQU.TEST
  - CHIQU.VERT
  - CHIQU.VERT.RE
  - EXPON.VERT
  - F.INV
  - F.INV.RE
  - F.TEST
  - Funktion einfügen...
- Dropdown 2 (middle):**
  - KGRÖSSTE
  - KKLEINSTE
  - KONFIDENZ.NORM
  - KONFIDENZ.T
  - KORREL
  - KOVARIANZ.P
  - KOVARIANZ.S
  - KURT
  - LOGNORM.INV
  - LOGNORM.VERT
  - HARMITTEL
  - HÄUFIGKEIT
  - HYPGEOM.VERT
  - Funktion einfügen...
- Dropdown 3 (right):**
  - MITTELWERTWENN
  - MITTELWERTWENNS
  - MODUS.EINF
  - MODUS.VIELF
  - NEGBINOM.VERT
  - NORM.INV
  - NORM.S.INV
  - NORM.S.VERT
  - NORM.VERT
  - PEARSON
  - PHI
  - POISSON.VERT
  - PROGNOSE.ETS
  - PROGNOSE.ETS.KONFI
  - PROGNOSE.ETS.SAISON
  - PROGNOSE.ETS.STAT
  - PROGNOSE.LINEAR
  - QUANTIL.EXKL
  - QUANTIL.INKL
  - Funktion einfügen...

# Die Basics

## Statistik Funktionen in Excel



The screenshot shows the Microsoft Excel interface with the 'Formeln' (Formulas) ribbon selected. The 'Funktion einfügen' (Insert Function) dialog box is open, displaying the 'Statistik' (Statistical) category. The function 'ACHSENABSCHNITT' is selected in the list. The description for this function is: 'Berechnet den Schnittpunkt einer Geraden mit der Y-Achse unter Verwendung einer Ausgleichsregressionsgeraden, die durch die X- und Y-Werte gezeichnet wird.'



## Die Basics

### Summe berechnen

	A	B	C	D
1	Filiale	Umsatz		
2	Warnemünde	30.500,00 €		
3	Lütten Klein	17.034,00 €		
4	Groß Klein	12.570,00 €		
5	Hansa-Viertel	21.156,00 €		
6	KTV	18.390,00 €		
7	Stadtmitte	34.692,00 €		
8	Bahnhof	17.622,00 €		
9	Südstadt	12.500,00 €		
10	Toitenwinkel	15.401,00 €		
11	Summe	=SUMME(B2:B10)		
12	Mittelwert	SUMME(Zahl1; [Zahl2]; ...)		
13				

Addiert alle markierten Felder

=SUMME(Zelle1:Zelle n)

=SUM(Cell1:Cell n)

## Die Basics

### Anzahl

	A	B	C	D
1	<b>Filiale</b>	<b>Umsatz</b>		
2	Warnemünde	30.500,00 €		
3	Lütten Klein	17.034,00 €		
4	Groß Klein	12.570,00 €		
5	Hansa-Viertel	21.156,00 €		
6	KTV	18.390,00 €		
7	Stadtmitte	34.692,00 €		
8	Bahnhof	17.622,00 €		
9	Südstadt	12.500,00 €		
10	Toitenwinkel	15.401,00 €		
11	<b>Summe</b>	179.865,00 €		
12	<b>Mittelwert</b>	19.985,00 €		
13	<b>Anzahl</b>	=Anzahl(B2:B12)		
14		ANZAHL(Wert1; [Wert2]; ...)		

Zählt markierte Zellen

=ANZAHL(Zelle1: Zelle n)

=COUNT(Cell 1; Cell n)



# Deskriptive univariate Statistik

## Lageparameter

### Modus, Median, Arithmetischer Mittelwert

Skalen-niveau	Operationen	Beispiel
Nominal	Modalwert (Häufigkeit)	Haarfarbe, Augenfarbe
Ordinal	Medianwert (Über 50%)	Schulnoten
Metrisch	Mittelwert	Umsatz, Gehalt

- Modus: Häufigste Ausprägung
- Median: Über 50% der Werte liegen unter diesem Wert, „Wert in der Mitte“
- Arithmetischer Mittelwert: Summe durch Anzahl der Werte

## Lageparameter

### Modus, Median

8	17	2
9	18	2
0	19	3
1	20	3
2	21	1
3		
4	Modus	=MODUS.EINF(B2:B22)

19	18	2
20	19	3
21	20	3
22	21	1
23		
24	Modus	2
25	Median	=MEDIAN(B2:B22)
26		

- =MODUS.EINF(Zelle1:Zelle n)
- =MODE.SNGL(Cell1:Cell n)

- =MEDIAN(Zelle1:Zelle n)
- =MEDIAN(Cell1:Cell n)  Median

1	2	2	2	3	3	4
---	---	---	---	---	---	---

  
 Modus 3x

## Lageparameter

### Mittelwert

	A	B	C	D
1	Filiale	Umsatz		
2	Warnemünde	30.500,00 €		
3	Lütten Klein	17.034,00 €		
4	Groß Klein	12.570,00 €		
5	Hansa-Viertel	21.156,00 €		
6	KTV	18.390,00 €		
7	Stadtmitte	34.692,00 €		
8	Bahnhof	17.622,00 €		
9	Südstadt	12.500,00 €		
10	Toitenwinkel	15.401,00 €		
11	Summe	179.865,00 €		
12	Mittelwert	=MITTELWERT(B2:B10)		
13		MITTELWERT(Zahl1; [Zahl2]; ...)		
14				

$$\mu = \frac{1}{N} \sum_{i=1}^N a_i = \frac{\text{Summe der Werte}}{\text{Anzahl Werte}}$$

=MITTELWERT(Zelle1:Zelle n)

= AVERAGE(Cell1: Cell n)

## Streuungsmaße

### Varianz und Standardabweichung

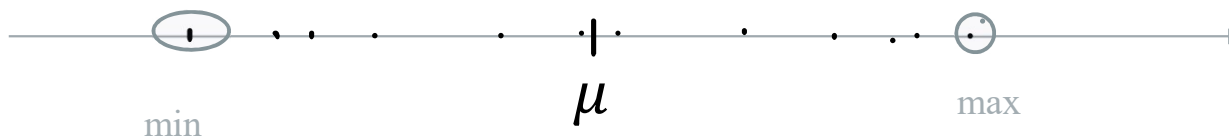
- Varianz: Summe der quadratischen Abweichung der Einzelwerte von ihrem arithmetischen Mittelwert geteilt durch Anzahl

Sigma Quadrat

N= Anzahl Werte

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n (a_i - \mu)^2$$

- Beschreibt Streuung der Werte um den Mittelwert



## Streuungsmaße

### Varianz und Standardabweichung

Varianz in Excel ermitteln

A	B
<b>Filiale</b>	<b>Umsatz</b>
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €
<b>Summe</b>	<b>179.865,00 €</b>
<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>
<b>Varianz</b>	<b>=VAR.P(B2:B10)</b>

- Grundgesamtheit: =VAR.P(Zahl 1:Zahl n)
- =VAR.P(Cell1: Cell n)

Die Funktion VAR.P geht davon aus, dass die ihr übergebenen Argumente einer Grundgesamtheit entsprechen. Wenn die zugehörigen Daten nur eine Stichprobe der Grundgesamtheit darstellen, sollten Sie die Varianz mit der Funktion VAR.S berechnen.

<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>
<b>Varianz</b>	<b>53009890,67</b>



## Streuungsmaße

### Varianz und Standardabweichung

Varianz in Excel ermitteln

A	B
<b>Filiale</b>	<b>Umsatz</b>
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €
<b>Summe</b>	<b>179.865,00 €</b>
<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>
<b>Varianz</b>	<b>=VAR.S(B2:B10)</b>

- Stichprobe: =VAR.S(Zahl 1:Zahl n)
- =VAR.S(Cell1: Cell n)

▪ VAR.S geht davon aus, dass die ihr übergebenen Argumente eine Stichprobe der Grundgesamtheit darstellen. Entsprechen die als Argumente übergebenen Daten dagegen einer Grundgesamtheit, sollte die zugehörige Varianz mithilfe der Funktion VAR.P berechnet werden.

2	<b>Mittelwert</b>	19.985,00 €
3	<b>Anzahl</b>	11
4	<b>Varianz</b>	59636127
5		

## Streuungsmaße

### Varianz und Standardabweichung

- Standardabweichung: positive Quadratwurzel aus der Varianz  
(Wurzel aus der Summe der mittleren quadratischen Abweichung der Einzelwerte von ihrem arithmetischen Mittelwert)

Sigma Quadrat

N= Anzahl Werte

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \sum (a_i - \mu)^2}$$

- Quadratisch gewichtete **durchschnittliche Abweichung vom Mittelwert**

## Streuungsmaße

### Varianz und Standardabweichung

Standardabweichung in Excel ermitteln

Filiale	Umsatz
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €
<b>Summe</b>	<b>179.865,00 €</b>
<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>
<b>Varianz</b>	<b>53009890,67</b>
<b>Stdabweichung</b>	<b>=STABW.N(B2:B10)</b>

- Grundgesamtheit: =STABW.N(Zahl 1:Zahl n)
- = STDEV.P(Cell1: Cell n)

<b>Anzahl</b>	<b>11</b>
<b>Varianz</b>	<b>53009890,67</b>
<b>Stdabweichung</b>	<b>7280,789151</b>

## Streuungsmaße

### Varianz und Standardabweichung

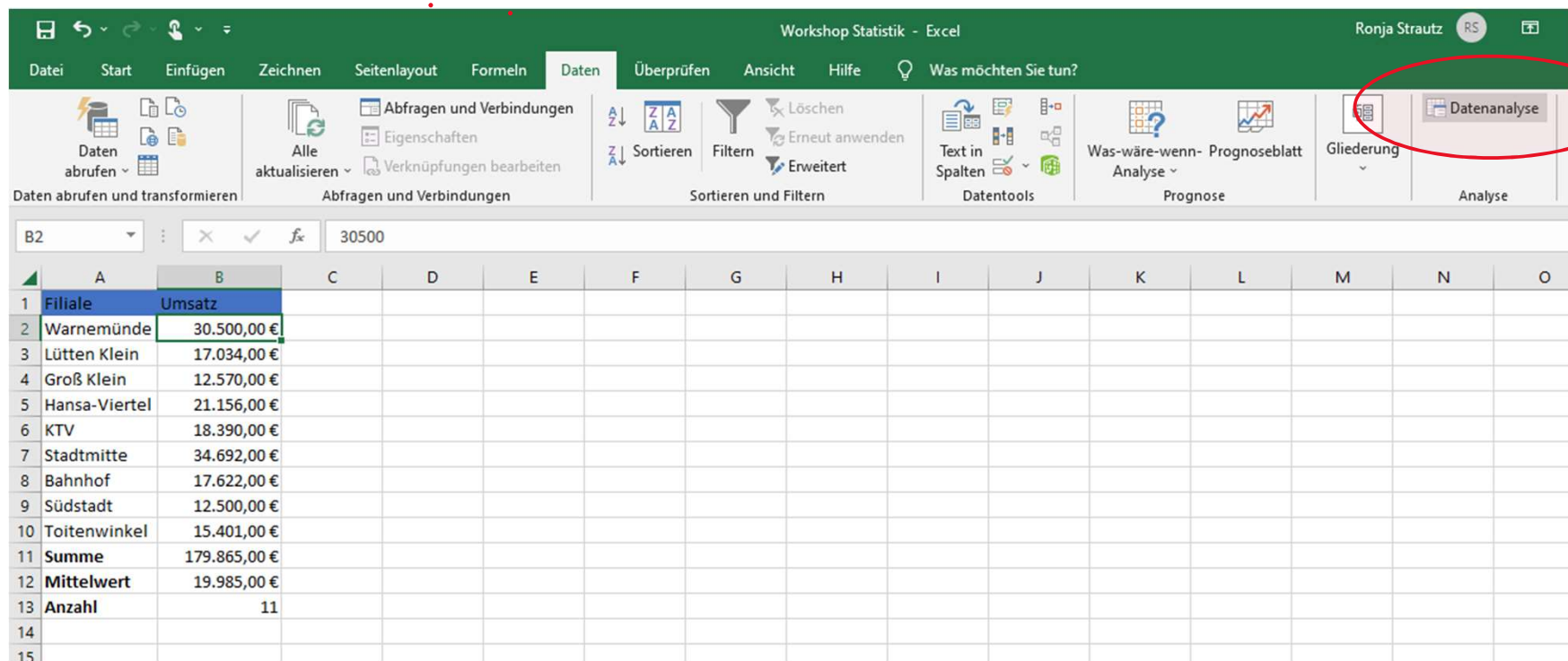
Standardabweichung in Excel ermitteln

Filiale	Umsatz
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €
<b>Summe</b>	<b>179.865,00 €</b>
<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>
<b>Varianz</b>	<b>53009890,67</b>
<b>Stdabweichung</b>	<b>=STABW.S(B2:B10)</b>

- Stichprobe: = STABW.S(Zahl 1:Zahl n)
- = STDEV.S(Cell1: Cell n)

<b>Anzahl</b>	<b>11</b>
<b>Varianz</b>	<b>53009890,67</b>
<b>Stdabweichung</b>	<b>7722,443072</b>

## Deskriptive Datenanalyse- Einfach auf einen Blick

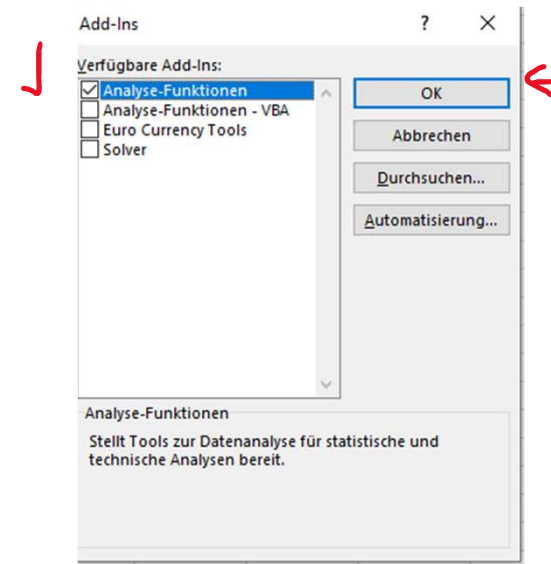
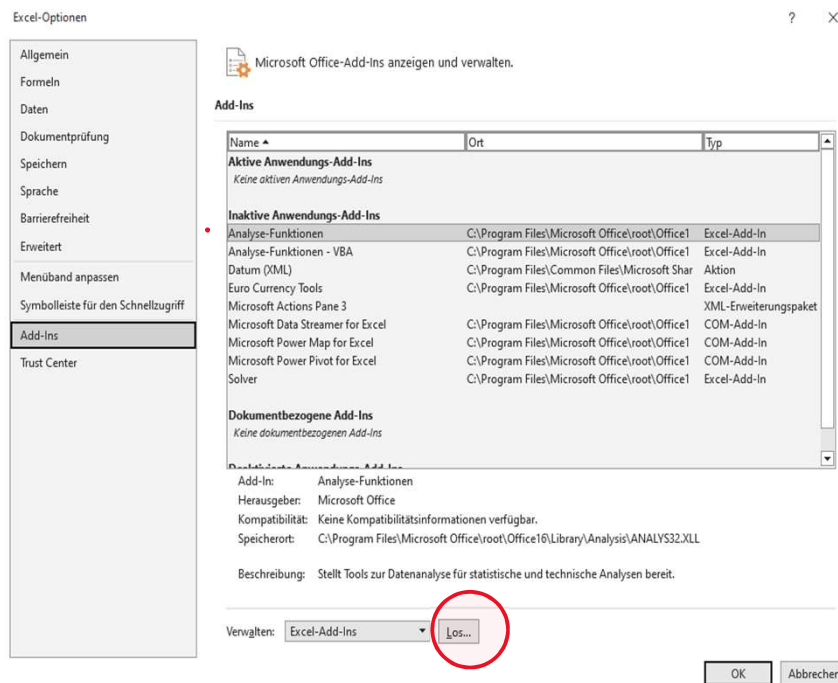


The screenshot shows the Microsoft Excel interface with the 'Daten' (Data) ribbon selected. The 'Datenanalyse' (Data Analysis) button is circled in red. Below the ribbon, a spreadsheet is visible with the following data:

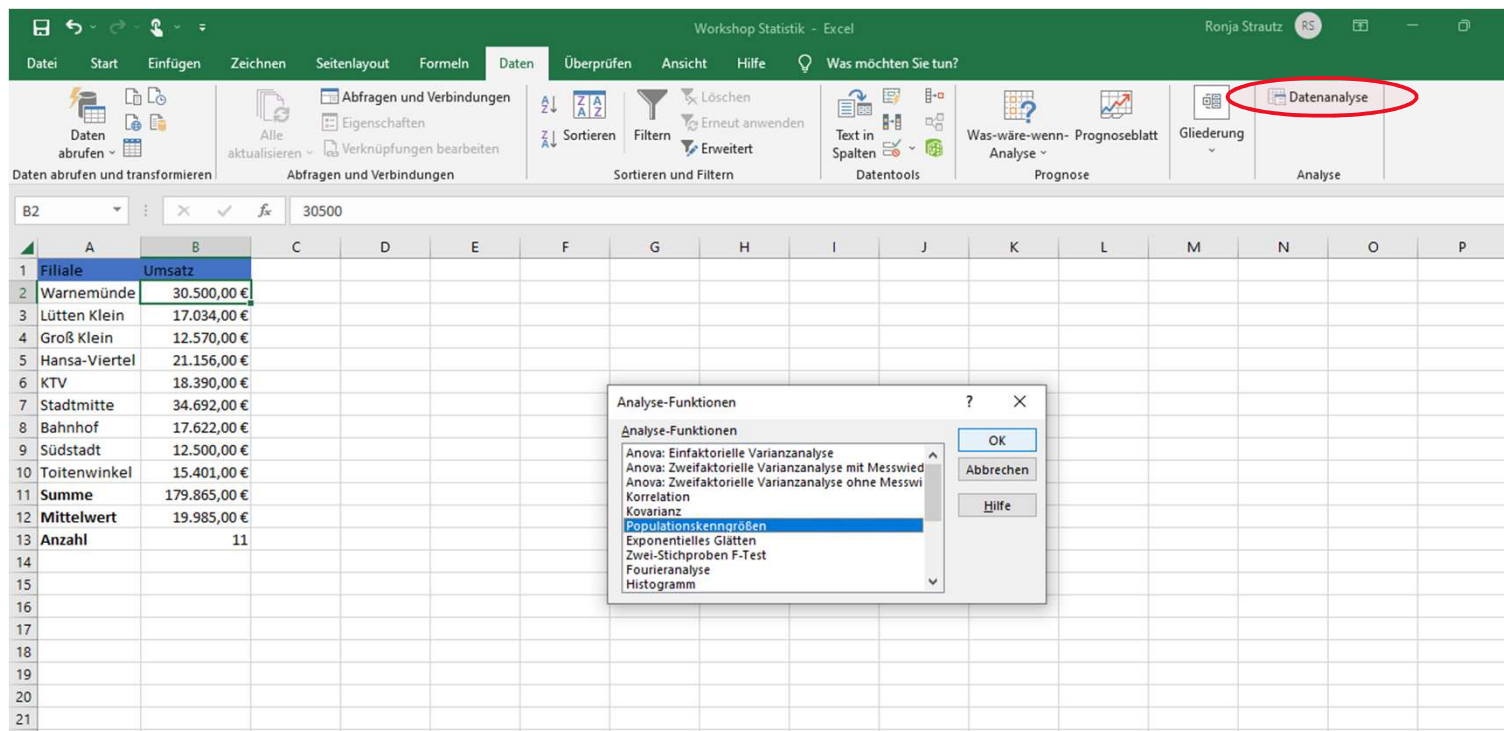
Filiale	Umsatz
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €
<b>Summe</b>	<b>179.865,00 €</b>
<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>

## Deskriptive Datenanalyse- Einfach auf einen Blick

Datei -> Optionen-> Add-ins-> Analyse Funktionen->Los



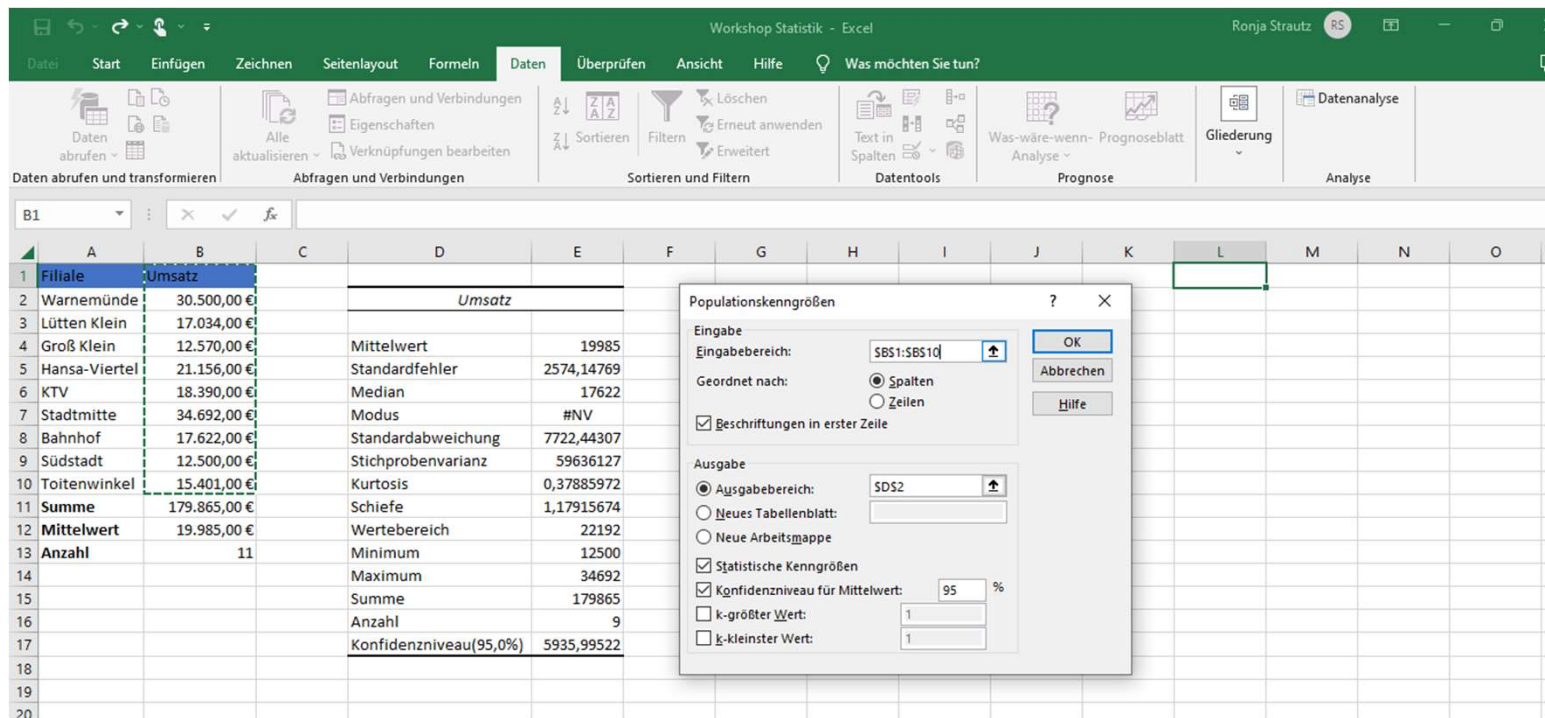
## Deskriptive Datenanalyse- Einfach auf einen Blick



The screenshot shows the Microsoft Excel interface with the 'Daten' (Data) ribbon active. The 'Datenanalyse' (Data Analysis) icon is circled in red. The 'Analyze-Tools Task Pane' is open, showing a list of statistical functions. The 'Populationskenngrößen' (Population Statistics) option is selected.

Filiale	Umsatz
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €
<b>Summe</b>	<b>179.865,00 €</b>
<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>

## Deskriptive Datenanalyse- Einfach auf einen Blick



The screenshot shows an Excel spreadsheet with the following data:

Filiale	Umsatz
Warnemünde	30.500,00 €
Lütten Klein	17.034,00 €
Groß Klein	12.570,00 €
Hansa-Viertel	21.156,00 €
KTV	18.390,00 €
Stadtmitte	34.692,00 €
Bahnhof	17.622,00 €
Südstadt	12.500,00 €
Toitenwinkel	15.401,00 €
<b>Summe</b>	<b>179.865,00 €</b>
<b>Mittelwert</b>	<b>19.985,00 €</b>
<b>Anzahl</b>	<b>11</b>

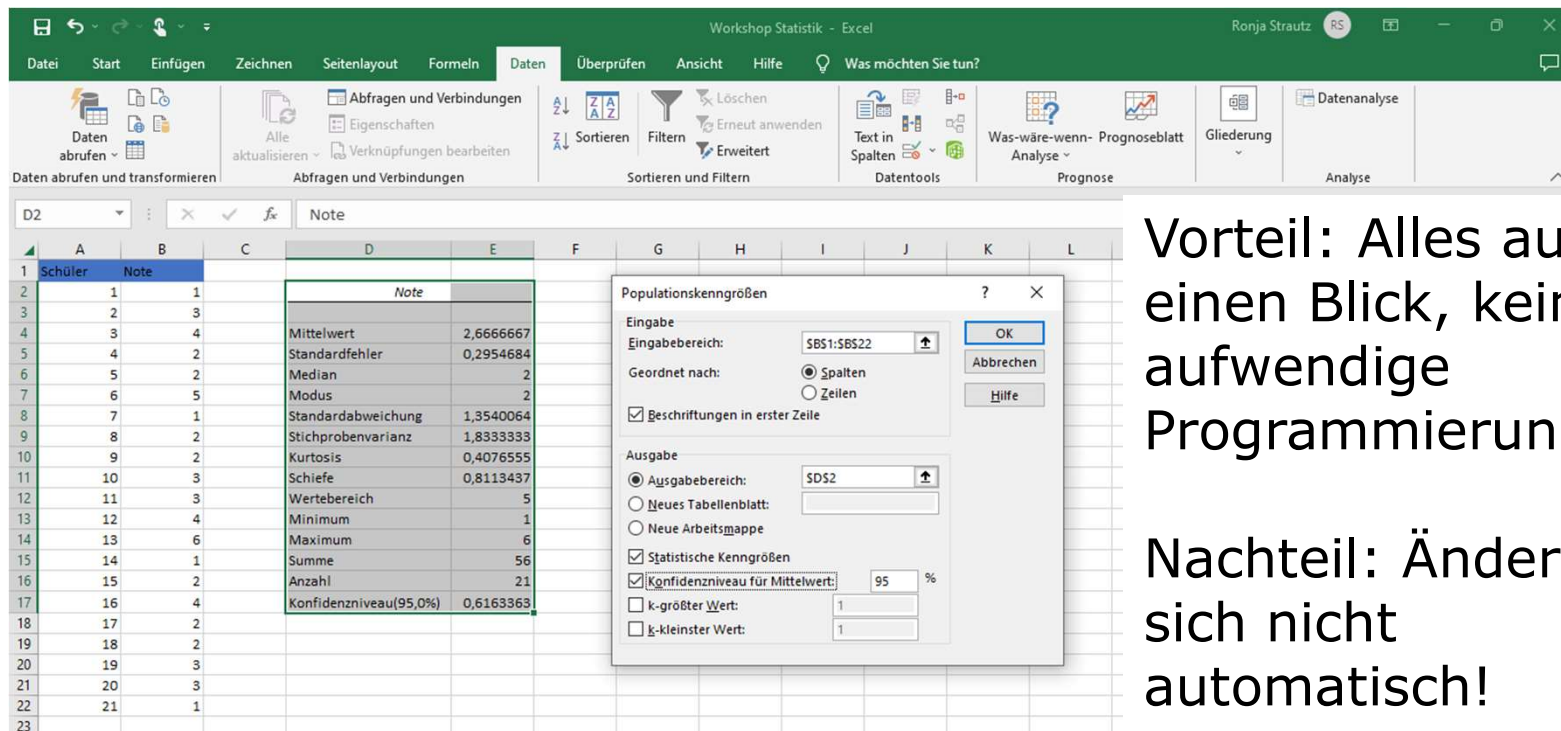
  

The 'Populationskenngrößen' dialog box is open, showing the following settings:

- Eingabe:** SBS1:SBS1Q
- Geordnet nach:** Spalten
- Beschriftungen in erster Zeile
- Ausgabe:**
  - Ausgabebereich: SDS2
  - Neues Tabellenblatt
  - Neue Arbeitsmappe
- Statistische Kenngrößen
- Konfidenzniveau für Mittelwert: 95 %
- k-größter Wert: 1
- k-kleinsten Wert: 1



## Deskriptive Datenanalyse- Einfach auf einen Blick



Schüler	Note
1	1
2	3
3	4
4	2
5	2
6	5
7	1
8	2
9	2
10	3
11	3
12	4
13	6
14	1
15	2
16	4
17	2
18	2
19	3
20	3
21	1
22	

Note	
Mittelwert	2,666667
Standardfehler	0,2954684
Median	2
Modus	2
Standardabweichung	1,3540064
Stichprobenvarianz	1,8333333
Kurtosis	0,4076555
Schiefe	0,8113437
Wertebereich	5
Minimum	1
Maximum	6
Summe	56
Anzahl	21
Konfidenzniveau(95,0%)	0,6163363

**Populationskenngrößen**

Eingabe  
 Eingabebereich: SB\$1:SB\$22  
 Geordnet nach:  Spalten  Zeilen  
 Beschriftungen in erster Zeile

Ausgabe  
 Ausgabebereich: SD\$2  
 Neues Tabellenblatt  
 Neue Arbeitsmappe  
 Statistische Kenngrößen  
 Konfidenzniveau für Mittelwert: 95 %  
 k-größter Wert: 1  
 k-kleinsten Wert: 1

Vorteil: Alles auf einen Blick, keine aufwendige Programmierung

Nachteil: Ändert sich nicht automatisch!

# Deskriptive Statistik

## Häufigkeitstabellen

Schüler	Note
1	1
2	2
3	3
4	4
5	2
6	5
7	6
8	7
9	8
10	9
11	10
12	11
13	12
14	13
15	14
16	15
17	16
18	17
19	18
20	19
21	20
22	21
23	1

Note	
Mittelwert	2,6666667
Standardfehler	0,2954684
Median	2
Modus	2
Standardabweichung	1,3540064
Stichprobenvarianz	1,8333333
Kurtosis	0,4076555
Schiefte	0,8113437
Wertebereich	5
Minimum	1
Maximum	6
Summe	56
Anzahl	21
Konfidenzniveau(95,0%)	0,6163363

=ZÄHLEWENN(Zelle1:  
Zelle n; Wert)

=COUNTIF( Cell 1:Cell n;  
Value)

Note	Häufigkeit	%	kum%
1	=ZÄHLENWENN(B2:B22;G3)		
2	ZÄHLENWENN(Bereich; Suchkriterien)		
3			
4			
5			
6			

=ZÄHLENWENN(\$B\$2:\$B\$22;G4)

Note	Häufigkeit	%
1	4	
2	5	
3	3	
4	2	
5	1	
6	1	

# Deskriptive Statistik

## Häufigkeitstabellen

MITTELWE... : X ✓ fx =H3/\$H\$9

Schüler	Note
1	1
2	2
3	3
4	4
5	2
6	2
7	5
8	1
9	2
10	2
11	3
12	3
13	4
14	6
15	1
16	2
17	4
18	2
19	2
20	3
21	3
22	1
23	

Note	Häufigkeit	%	kum%
1	4	=H3/\$H\$9	
2	7	33,33%	
3	5	23,81%	
4	3	14,29%	
5	1	4,76%	
6	1	4,76%	
Summe	21		

Prozent

% 000 ←,0 ,00 →,0

Zahl

Bedingte Formatierung

Note	Häufigkeit	%
1	4	19,05%
2	7	33,33%
3	5	23,81%
4	3	14,29%
5	1	4,76%
6	1	4,76%
Summe	21	

Absolute Häufigkeit  
/Summe

=relative Häufigkeit

# Deskriptive Statistik

## Häufigkeitstabelle

Schüler	Note					Häufigkeitstabelle			
			Note			Note	Häufigkeit	%	kum%
1	1	1				1	4	19,05%	19,05%
2	2	3				2	7	33,33%	52,38%
3	3	4	Mittelwert	2,666667		3	5	23,81%	=J4+I5
4	4	2	Standardfehler	0,2954684		4	3	14,29%	90,48%
5	5	2	Median			5	1	4,76%	95,24%
6	6	5	Modus	2		6	1	4,76%	100,00%
7	7	1	Standardabweichung	1,3540064		Summe	21		
8	8	2	Stichprobenvarianz	1,8333333					
9	9	2	Kurtosis	0,4076555					
10	10	3	Schiefe	0,8113437					
11	11	3	Wertebereich	5					
12	12	4	Minimum	1					
13	13	6	Maximum	6					
14	14	1	Summe	56					
15	15	2	Anzahl	21					
16	16	4	Konfidenzniveau(95,0%)	0,6163363					
17	17	2							
18	18	2							
19	19	3							
20	20	3							
21	21	1							

Kummulierte Häufigkeiten: Aufsummieren der relativen Häufigkeiten auf einhundert Prozent.

Median erkennbar: erste Spalte mit über 50%







# Deskriptive Statistik

## Einfaches Diagramm zu Häufigkeiten

Möglichkeit 1

	F	G	H	I	J	K	L
		Häufigkeitstabelle					
		Note	Häufigkeit	%	kum%		
		1	4	19,05%	19,05%		
6667		2	7	33,33%	52,38%		
4684		3	5	23,81%	76,19%		
2		4	3	14,29%	90,48%		
2		5	1	4,76%	95,24%		
0064		6	1	4,76%	100,00%		
3333		Summe	21				

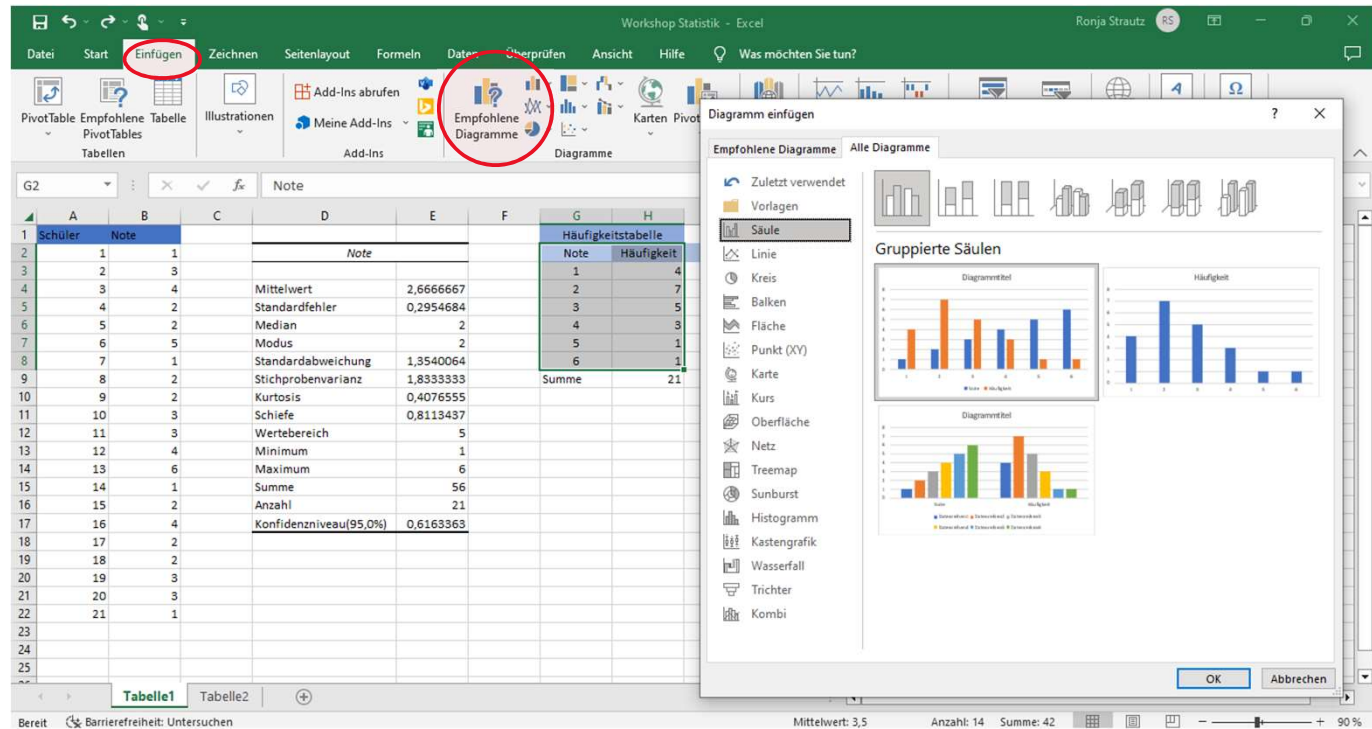
Formatierung	<u>D</u> igramme	Ergebnisse	T	S	
					
Gruppier...	Gruppier...	Punkt (XY)	Linie	Kreis	Weitere...

Empfohlene Diagramme helfen Ihnen, Daten zu visualisieren.

# Deskriptive Statistik

## Einfaches Diagramm zu Häufigkeiten

Möglichkeit 2



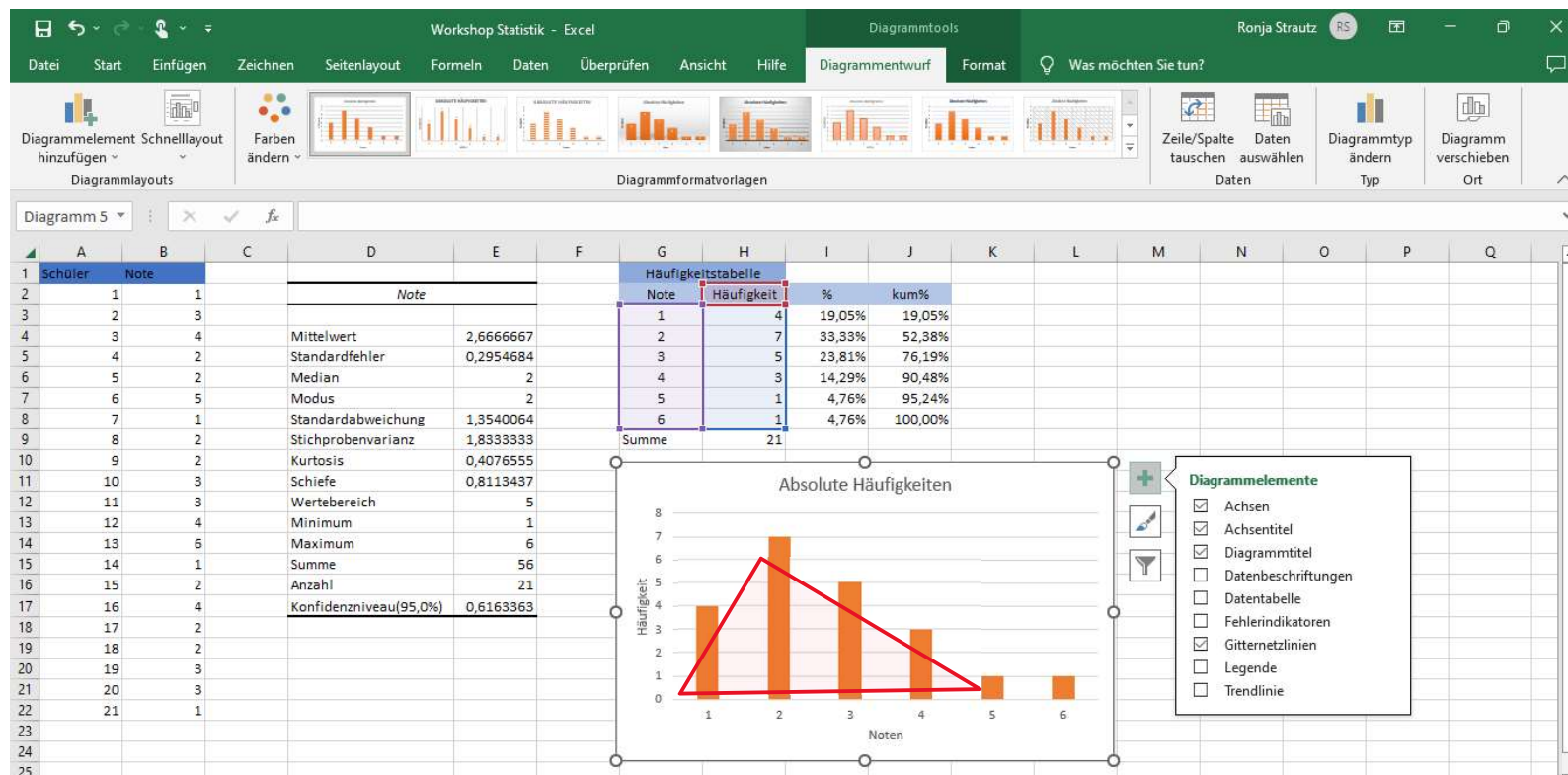
The screenshot shows an Excel spreadsheet with the following data table:

Schüler	Note				Häufigkeitstabelle
1	1				Note
2	2				1 4
3	3				2 7
4	4	Mittelwert	2,666667		3 5
5	2	Standardfehler	0,2954684		4 3
6	2	Median	2		5 1
7	5	Modus	2		6 1
8	1	Standardabweichung	1,3540064		Summe
9	2	Stichprobenvarianz	1,8333333		21
10	2	Kurtosis	0,4076555		
11	3	Schiefte	0,8113437		
12	3	Wertebereich	5		
13	4	Minimum	1		
14	6	Maximum	6		
15	1	Summe	56		
16	2	Anzahl	21		
17	4	Konfidenzniveau(95,0%)	0,6163363		

The 'Empfohlene Diagramme' task pane is open, showing various chart options. The 'Säule' (Column) chart type is selected, and the 'Gruppierter Säulen' (Grouped Column) chart is highlighted as a recommended option for the data.

# Deskriptive Statistik

## Einfaches Diagramm zu Häufigkeiten





# Bivariate Statistik

## Zusammenhangsmaße



## Zusammenhangsmaße

### Chi<sup>2</sup>, Kontingenzkoeffizient

- **Chi<sup>2</sup> Unabhängigkeitstest**
- ab Nominalskala
- 2 Merkmale erhoben (Bivariat)
- Einfachste Fall: 4-Felder-Kreuztabelle
- Basiert aus Vergleich beobachtete und erwartete Häufigkeiten

	A	B	C	D	E
1	Haben Sie im vergangenen Monat Alkohol getrunken?				
2	100 Personen				
3			ja	nein	Zeilensumme
4		Mann	50	10	60
5		Frau	20	20	40
6		Spaltensumme	70	30	100
7					
8					

=SUMME(C4:D4)

=SUMME(C4:C5)

## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

- Erwartung/Annahme des Chi<sup>2</sup>-Tests: Unabhängigkeit  
d.h. Ausprägung einer Variable ist unabhängig von der Ausprägung der zweiten Variable
- Erwartete Häufigkeiten für die jeweilige Zelle  
= Das Produkt der Randhäufigkeiten/ Anzahl Beobachtungen

	A	B	C	D	E
1	Haben Sie im vergangenen Monat Alkohol getrunken?				
2	100 Personen				
3			ja	nein	Zeilensumme
4		Mann	50	10	60
5		Frau	20	20	40
6		Spaltensumme	70	30	100
7					
8					

Beobachtet(Mann&Ja); 50  
 Erwartet:  $60 \cdot 70 / 100 = 35$

# Zusammenhangsmaße

## Chi<sup>2</sup> Unabhängigkeitstest

- Berechnung in Excel

MITTELWE...   X ✓ fx   =E5*C6/E6													
	A	B	C	D	E	F	G	H	I	J	K		
1	Haben Sie im vergangenen Monat Alkohol getrunken?												
2	100 Personen												
3			ja	nein	Zeilensumme			ja	nein	Zeilensumme			
4		Mann		50	10			60		Mann		60	
5		Frau		20	20			40		Frau	=E5*C6/E6	40	
6		Spaltensumme		70	30			100		Spaltensumme	70	30	100
7													
8													

## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

#### Beobachtet

	A	B	C	D	E
1	Haben Sie im vergangenen Monat Alkohol getrunken?				
2	100 Personen				
3			ja	nein	Zeilensumme
4		Mann	50	10	60
5		Frau	20	20	40
6		Spaltensumme	70	30	100
7					
8					

#### Erwartet

		ja	nein	Zeilensumme
Mann		42	18	60
Frau		28	12	40
Spaltensumme		70	30	100

Die Variablen X und Y sind statistisch unabhängig, wenn die erwarteten gleich den beobachteten Häufigkeiten sind.

Weichen die Häufigkeiten voneinander ab? Sind die Unterschiede zufällig?

## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

- Formel

$$\chi^2 = \sum \frac{(\text{beobachteter} - \text{erwarteter Wert})^2}{\text{erwarteter Wert}}$$

- Was prüft der Test?
- Unabhängigkeit von X und Y, Unterschiede der beobachteten und erwarteten Häufigkeiten
- Hypothese  $H_0$  = X und Y sind statistisch unabhängig, d.h. die Ausprägung von X hat keinen Einfluss auf die Ausprägung von Y oder umgekehrt
- In unserem Fall: es besteht kein Zusammenhang zwischen Geschlecht und Alkoholkonsum

## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

- Was prüft der Test?
- Unabhängigkeit von X und Y, Unterschiede der beobachteten und erwarteten Häufigkeiten
- Hypothese  $H_0$ : X und Y sind statistisch unabhängig, d.h. die Ausprägung von Y hat keinen Einfluss auf die Ausprägung von X oder umgekehrt
- In unserem Fall: es besteht kein Zusammenhang zwischen Geschlecht und Alkoholkonsum
- Irrtumswahrscheinlichkeit 5%: Mit 5% Wahrscheinlichkeit treffen wir die falsche Entscheidung bei der Ablehnung/Nichtablehnung der Nullhypothese
- Vergleich des errechneten Chi<sup>2</sup>-Wertes ( $\chi^2$ ) mit einem Wert der Chi<sup>2</sup> Verteilung

## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

- Hypothese  $H_0$ : X und Y sind statistisch unabhängig, d.h. die Ausprägung von Y hat keinen Einfluss auf die Ausprägung von X oder umgekehrt
- In unserem Fall: es besteht kein Zusammenhang zwischen Geschlecht und Alkoholkonsum
- Irrtumswahrscheinlichkeit 5%: Mit 5% Wahrscheinlichkeit treffen wir die falsche Entscheidung bei der Ablehnung/Nichtablehnung der Nullhypothese
- Vergleich des errechneten Chi<sup>2</sup>-Wertes ( $\chi^2$ ) mit einem Wert der Chi<sup>2</sup> Verteilung
- $\chi_{emp}^2 > \chi^2$   $H_0$  ablehnen
- $\chi_{emp}^2 < \chi^2$   $H_0$  nicht ablehnen

## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

- Irrtumswahrscheinlichkeit 5%: Mit 5% Wahrscheinlichkeit treffen wir die falsche Entscheidung bei der Ablehnung/Nichtablehnung der Nullhypothese
- In Excel: Chi<sup>2</sup> Teststatistik

Haben Sie im vergangenen Monat Alkohol getrunken?							
100 Personen							
	ja	nein	Zeilensumme		ja	nein	Zeilensumme
Mann	50	10	60	Mann	42	18	60
Frau	20	20	40	Frau	28	12	40
Spaltensumme	70	30	100	Spaltensumme	70	30	100

=CHIQU.TEST(C4:D5;H4:I5)	=CHISQ.TEST(observed frequencies;expected frequencies)
CHIQU.TEST(Beob_Messwerte; Erwart_Werte)	

Gibt den Signifikanzwert wieder!

=CHIQU.TEST(beobachtete  
Zellenhäufigkeiten; erwartete  
Zellenhäufigkeiten)  
=CHISQ.TEST(observed  
frequencies;expected frequencies)



## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

Haben Sie im vergangenen Monat Alkohol getrunken?  
100 Personen

	ja	nein	Zeilensumme		ja	nein	Zeilensumme
Mann	50	10	60	Mann	42	18	60
Frau	20	20	40	Frau	28	12	40
Spaltensumme	70	30	100	Spaltensumme	70	30	100

`=CHIU.TEST(C4:D5;H4:I5)`  
 CHIQU.TEST(Beob\_Messwerte; Erwart\_Werte)

Signifikanzw.	0,000365966
---------------	-------------

Gibt den Signifikanzwert, auch p-Wert genannt, wieder!

- Mit 5% Wahrscheinlichkeit treffen wir die falsche Entscheidung bei der Ablehnung der Nullhypothese
- Wenn Wert  $\leq 0,05$  ist, können wir mit 95%iger Sicherheit die Nullhypothese ablehnen, wenn der Wert  $>0,05$  ist kann die Nullhypothese nicht abgelehnt werden.

## Zusammenhangsmaße

### Chi<sup>2</sup> Unabhängigkeitstest

- Irrtumswahrscheinlichkeit 5%: Mit 5% Wahrscheinlichkeit treffen wir die falsche Entscheidung bei der Ablehnung/Nichtablehnung der Nullhypothese
- Wenn Wert  $\leq 0,05$  ist, können wir mit 95%iger Sicherheit die Nullhypothese ablehnen
  
- $0.00036 < 0,05$
- $H_0$  („ es besteht kein Zusammenhang zwischen Geschlecht und Alkoholkonsum“) **verwerfen**
- →Es gibt einen signifikanten Zusammenhang zwischen dem Alkoholkonsum und dem Geschlecht, die Nullhypothese kann zum 5% Signifikanzniveau verworfen werden.
  
- **Chi<sup>2</sup> Test trifft keine Aussage über Stärke des Zusammenhanges oder die Richtung, oder welche Variable welche beeinflusst!**

## Zusammenhangsmaße

### Kontingenzkoeffizient

- $\chi^2$  ist abhängig von der Menge der erhobenen Daten
- Normierung von  $\chi^2$  -> Kontingenzkoeffizient C

Hierfür wird der empirische  $\chi^2$  Wert benötigt

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

- Korrigierter Kontingenzkoeffizient  $C_{\text{korr}}$ 
  - Wertebereich 0-1, Interpretierbar
  - 0 Abwesenheit eines Zusammenhanges, 1 perfekter Zusammenhang

$$C_{\text{korr}} = C * \sqrt{\frac{\min(r,s)}{\min(r,s)-1}}$$

Kleinste Anzahl der Zeilen oder Spalten in Tabelle

# Zusammenhangsmaße

## Kontingenzkoeffizient

$$\chi^2 = \sum \frac{(\text{beobachteter} - \text{erwarteter Wert})^2}{\text{erwarteter Wert}}$$

haben sie im vergangenen Monat Alkohol getrunken?

10 Personen

	ja	nein	Zeilensumme		ja	nein	Zeilensumme
Mann	50	10	60	Mann	42	18	60
Frau	20	20	40	Frau	28	12	40
Spaltensumme	70	30	100	Spaltensumme	70	30	100

$\chi^2$			$=((C4-H4)^2)/H4$	
Signifikanzw	0,000365966		2,28571	5,33333

Chi <sup>2</sup>	$=\text{SUMME}(E8:F9)$	1,52381	3,55556
Signifikanzw	0,000365966	2,28571	5,33333

Chi <sup>2</sup>	12,69841
Signifikanzw	0,000365966

## Zusammenhangsmaße

### Kontingenzkoeffizient

$$C = \sqrt{\frac{x^2}{N+x^2}}$$

	ja	nein	Zeilensumme	
Mann	50	10	60	
Frau	20	20	40	
Spaltensumme	70	30	100	
hi <sup>2</sup>	12,69841		1,52381	3,55556
gnifikanzw	0,000365966		2,28571	5,33333
C	=WURZEL(B8/(E6+B8))			

Achtung  
 Klammern  
 setzen, sonst  
 Punkt vor  
 Strichrechnung

$$C \text{ korr} = C * \sqrt{\frac{\min(r,s)}{\min(r,s)-1}}$$

C	0,335672543
C korr	=B10*WURZEL(2/(2-1))

-> mittelstarker Zusammenhang

Chi<sup>2</sup> Test und Kontingenzkoeffizient auch für mehr als 2 Ausprägungen, für ordinale und metrische Variablen nur eingeschränkt möglich

## Literatur

- J. Bley Müller, R. Weißbach, A. Dörre (2020): Statistik für Wirtschaftswissenschaftler, Vahlen, München, 18. Auflage.
- (2021): Statistische Formeln und Tabellen, Vahlen, München, 14. Auflage.
- (2018): Übungen zu Statistik für Wirtschaftswissenschaftler, Vahlen, München.

## Ausblick

- Konfidenzintervalle, T-Test
- Korrelation
- ANOVA
- Lineare Regression



## Univariate Datenanalyse

### **Konfidenzintervalle, T-Test**

Bildet die Stichprobe gut die Grundgesamtheit ab?

Können wir zulässige Schlüsse aus der Stichprobe ziehen? (Repräsentationsschluss)

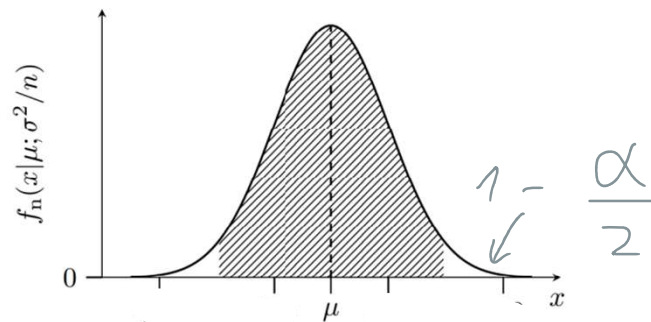
Wir versuchen also möglichst gut den wahren Wert der Grundgesamtheit zu schätzen und geschätzte Parameter auf ihre Signifikanz zu testen

# Univariate Datenanalyse

## Konfidenzintervalle

Mit Hilfe des Geschätzten Wertes wird ein Intervall ermittelt, dass mit einer bestimmten Wahrscheinlichkeit, dem Signifikanzniveau  $\alpha$ , den wahren Wert enthält.

**Idee:** bestimme Intervall um  $\mu$  mit Fläche  $(1 - \alpha)$  (z.B. 95%):



$$\alpha = 0,05$$

$\square W(L \leq \mu \leq R) = 0,95$

$$\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



## Univariate Statistik

### Konfidenzintervall

$$\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$\bar{x}$  Mittelwert: =MITTELWERT(Zelle1:Zelle n)

$z_{1-\alpha/2}$  Quantil der Standardnormalverteilung; wenn  $\alpha = 0,05$   $z = 1,96$ , wenn  $\alpha = 0,01$   
 $z = 2,57$

$\frac{\sigma}{\sqrt{n}}$  Standardfehler des Mittelwertes, Standardabweichung/Wurzel der Anzahl

= STABW.S(Zelle 1:Zelle n)/ WURZEL(=ANZAHL(Zelle 1:Zelle n))

= AVERAGE(Cell1: Cell n)

if  $\alpha = 0,05$   $z = 1,96$ , if  $\alpha = 0,01$   $z = 2,57$

= STDEV.S(Cell1: Cell n)/SQRT(=COUNT(Cell 1; Cell n))

Oder einfach  
mithilfe der  
Datenanalyse



# Univariate Statistik

## Konfidenzintervall

MITTELWE...				=J4+I5		
	A	B	C	D	E	F
1	Schüler	Note				
2	1	1		Note		
3	2	3				
4	3	4		Mittelwert	2,6666667	
5	4	2		Standardfehler	0,2954684	
6	5	2		Median	2	
7	6	5		Modus	2	
8	7	1		Standardabweichung	1,3540064	
9	8	2		Stichprobenvarianz	1,8333333	
10	9	2		Kurtosis	0,4076555	
11	10	3		Schiefe	0,8113437	
12	11	3		Wertebereich	5	
13	12	4		Minimum	1	
14	13	6		Maximum	6	
15	14	1		Summe	56	
16	15	2		Anzahl	21	
17	16	4		Konfidenzniveau(95,0%)	0,6163363	
18	17	2				
19	18	2				
20	19	3				
21	20	3				
22	21	1				

## Univariate Statistik

### Parametertest: T-Test

Unterscheidet sich der Mittelwert (oder zB. Anteilswert) der Stichprobe *signifikant* von einem vorgegebenen Wert. Zum Beispiel. Die Durchschnittsnote der Klassenarbeit ist 2.

Hypothese  $H_0$ : Der Mittelwert der Grundgesamtheit = 2

Hypothese  $H_A$ : Der Mittelwert der Grundgesamtheit  $\neq$  2

Nach dem Test:  $H_0$  ablehnen oder nicht ablehnen, zu bestimmten Signifikanzniveau und kritischem Wert

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} < t_{n-1, \frac{1-\alpha}{2}} \quad H_0 \text{ nicht ablehnen}$$

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} > t_{n-1, \frac{1-\alpha}{2}} \quad H_0 \text{ ablehnen}$$

## Univariate Statistik

### Parametertest: T-Test

Nach dem Test:  $H_0$  ablehnen oder nicht ablehnen, zu bestimmten Signifikanzniveau und kritischem Wert

$$T_{emp} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$  Mittelwert: =MITTELWERT(Zelle1:Zelle n)

$\mu$  Vorgegebener Wert

$\frac{\sigma}{\sqrt{n}}$  Standardfehler des Mittelwertes, Standardabweichung/Wurzel der Anzahl, =  
STABW.S(Zelle 1:Zelle n)/ WURZEL(=ANZAHL(Zelle 1:Zelle n))

## Univariate Statistik

### Parametertest: T-Test

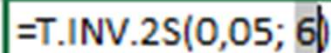
Nach dem Test:  $H_0$  ablehnen oder nicht ablehnen, zu bestimmten Signifikanzniveau und **kritischem Wert**

**$t_{n-1, 1-\alpha/2}$**  Quantil der Studentverteilung


n-1 sind die sogenannten Freiheitsgrade (degrees of freedom (df))

n = ANZAHL(Zelle 1: Zelle n)

Signifikanzniveau 0,05



=T.INV.2S(0,05; 6)



T.INV.2S(Wahrsch; Freiheitsgrade)

2-seitiger T-Test

## Univariate Statistik

### Parametertest: T-Test

P-Wert

n-1 sind die sogenannten Freiheitsgrade (degrees of freedom (df))

n = ANZAHL(Zelle 1: Zelle n)

Signifikanzniveau  $\leftrightarrow 0,05$ ?

df: n-1	6
T emp	2,703
T krit	2,364624252
p wert	=T.VERT.2S(ABS(D27);6)
	T.VERT.2S(x; Freiheitsgrade)

Absolutwerte ABS, da T emp auch negativ sein kann

## Univariate Statistik

### Parametertest: T-Test

Nach dem Test:  $H_0$  ablehnen oder nicht ablehnen, zu bestimmten Signifikanzniveau und kritischem Wert

$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} < t_{n-1, 1-\alpha/2}$   $H_0$  nicht ablehnen kein signifikanter Unterschied zwischen gegebenem Wert und Stichprobenmittelwert

$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} > t_{n-1, 1-\alpha/2}$   $H_0$  ablehnen, Stichprobenmittelwert und gegebener Wert sind signifikant verschieden

Oder Über p-Wert

Ist p-Wert  $> 0,05$   $H_0$  nicht ablehnen

Ist p-Wert  $\leq 0,05$   $H_0$  ablehnen, der Unterschied ist signifikant

Wie lautet Entscheidung?

$H_0$  ablehnen, Der Stichprobenmittelwert und der gegebene Mittelwert sind signifikant verschieden, mit einer Irrtumswahrscheinlichkeit von 0,05.

df: n-1	6
T emp	2,703
T krit	2,364624252
p wert	0,035434324

## Bivariate Datenanalyse

### Korrelation

Wie ist das Verhältnis von zwei metrischen Variablen. Korrelieren 2 Variablen miteinander?  
 Wie verhält sich Variable 2, wenn Variable 1 Steigt oder sinkt?

Pearsons Korrelationskoeffizient

$$Correl(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

In Excel:

Körpergröße	Schuhgröße		
134	34		
167	38		
170	39		
156	39		
180	42	KORREL	=KORREL(I33:I41;J33:J41)
195	46		KORREL(Matrix1; Matrix2)
165	40		
183	44		
166	39		



## Bivariate Datenanalyse

### Korrelation

Wie ist das Verhältnis von zwei metrischen Variablen. Korrelieren 2 Variablen miteinander?  
 Wie verhält sich Variable 2, wenn Variable 1 Steigt oder sinkt?

Pearsons Korrelationskoeffizient

$$Correl(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

In Excel:

ODER

Körpergröße	Schuhgröße		
134	34		
167	38		
170	39		
156	39		
180	42	KORREL	0,948723
195	46	PEARSON	=PEARSON(I33:I41;J33:J41)
165	40		
183	44		
166	39		

## Bivariate Datenanalyse

### Korrelation

Wie ist das Verhältnis von zwei metrischen Variablen. Korrelieren 2 Variablen miteinander?

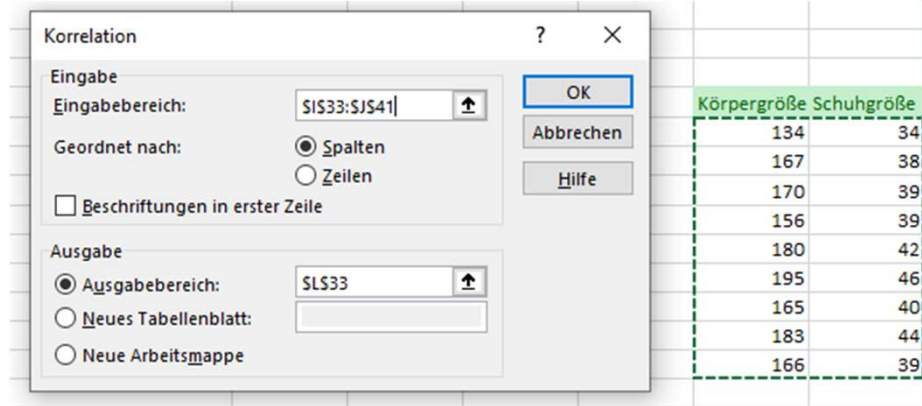
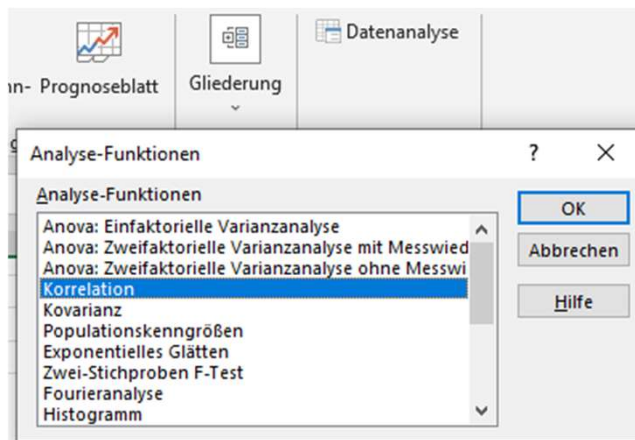
Wie verhält sich Variable 2, wenn Variable 1 Steigt oder sinkt?

Pearsons Korrelationskoeffizient

$$\text{Correl}(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

In Excel:

ODER



## Bivariate Datenanalyse

### Korrelation-Interpretation

Solange sich der Korrelationskoeffizient näher an +1 oder -1 befindet, weist er auf eine positive (+1) oder negative (-1) Korrelation zwischen den Matrizen hin. Positive Korrelation bedeutet, dass, wenn die Werte in einer Variablen steigen, auch die Werte in der anderen Variablen größer werden. Ein Korrelationskoeffizient, der näher an 0 liegt, gibt eine keine oder schwache Korrelation an.

Korrelation bedeutet nicht zwingend Kausalität!, Pearsons Korrelationskoeffizient macht keine Unterschiede zwischen abhängiger und unabhängiger Variable.

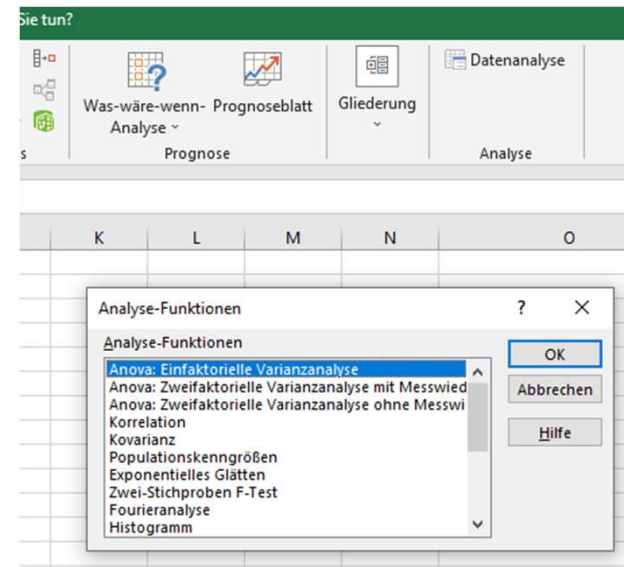
Körpergröße	Schuhgröße		Spalte 1	Spalte 2
134	34			
167	38	Spalte 1	1	
170	39	Spalte 2	0,948723	1
156	39			
180	42	KORREL	0,948723	
195	46	PEARSON	0,948723	
165	40			
183	44			
166	39			

## Bi- und Multivariate Datenanalyse

### ANOVA- Analysis of Variance

Gruppen vergleichen, Hypothesen Testen, ob es unterschiede zwischen den Gruppen gibt.  
 Einfaktoriell: Ein Faktor unterscheidet die Gruppen; hier zum Beispiel die Klasse.  
 Gemessen wurde die Punktzahl beim gleichen Test

	Klasse A	Klasse B
	34	100
	56	53
	79	36
	80	67
	94	51
	87	40
	97	79
	100	83
	25	61
	40	18
	13	68



## Bi- und Multivariate Datenanalyse

### ANOVA

Gruppen vergleichen, Hypothesen Testen, ob es unterschiede zwischen den Gruppen gibt.

Einfaktoriell: Ein Faktor unterscheidet die Gruppen; hier zum Beispiel die Klasse.

Gemessen wurde die Punktzahl beim gleichen Test

Klasse A	Klasse B
34	100
56	53
79	36
80	67
94	51
87	40
97	79
100	83
25	61
40	18
13	68

Anova: Einfaktorielle Varianzanalyse

Eingabe  
Eingabebereich: SBS32:SC\$43

Geordnet nach:  Spalten  Zeilen

Beschriftungen in erster Zeile

Alpha: 0,05

Ausgabe  
 Ausgabebereich: ES34  
 Neues Tabellenblatt:  
 Neue Arbeitsmappe

OK  
Abbrechen  
Hilfe

## Bi- und Multivariate Datenanalyse

### ANOVA

Gruppen vergleichen, Hypothesen testen, ob es Unterschiede zwischen den Gruppen gibt.

$H_0$ : Es gibt keinen Unterschied zwischen den Gruppen, Signifikanzniveau 0,05

Anova: Einfaktorielle Varianzanalyse						
ZUSAMMENFASSUNG						
Gruppen	Anzahl	Summe	Mittelwert	Varianz		
Klasse A	11	705	64,09090909	995,6909091		
Klasse B	11	656	59,63636364	547,2545455		
ANOVA						
Streuungsursache	Quadratsummen (SS)	Freiheitsgrade (df)	Mittlere Quadratsumme (MS)	Prüfgröße (F)	P-Wert	kritischer F-Wert
Unterschiede zwischen den Gruppen	109,1363636	1	109,1363636	0,141464967	0,71078914	4,3512435
Innerhalb der Gruppen	15429,45455	20	771,4727273			
Gesamt	15538,59091	21				

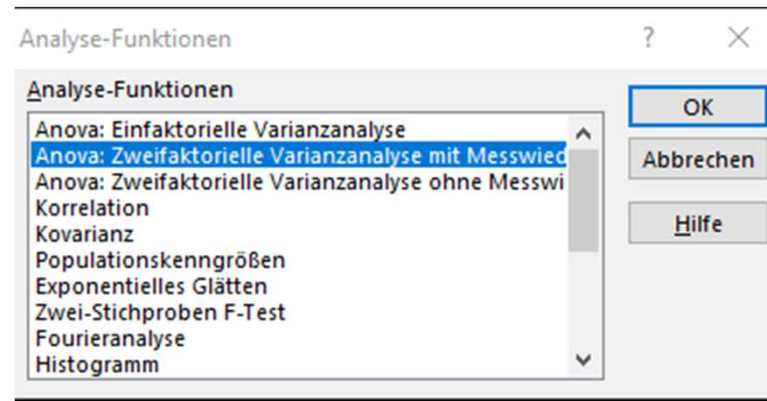
Prüfgröße < Krit. Wert, P-Wert > 0,05: Es gibt keinen signifikanten Unterschied zwischen den Gruppen,  $H_0$  kann nicht abgelehnt werden

## Bi- und Multivariate Datenanalyse

### ANOVA- Analysis of Variance

Zweifaktoriell: Zwei Faktoren unterscheiden die Gruppen; hier zum Beispiel die Klasse und Test.

	Klasse A	Klasse B
Test 1	50	100
	100	53
	79	36
	80	67
	94	51
Test 2	87	40
	30	79
	100	83
	25	61
	20	18



# Bi- und Multivariate Datenanalyse

## ANOVA

	Klasse A	Klasse B
Test 1	50	100
	100	53
	79	36
	80	67
	94	51
Test 2	87	40
	30	79
	100	83
	25	61
	20	18

Anova: Zweifaktorielle Varianzanalyse mit Messwied... ? X

**Eingabe**

Eingabebereich: SAS51:SC\$51 ↑

Zeilen je Stichprobe: 5

Alpha: 0,05

**Ausgabe**

Ausgabebereich: SE\$53 ↑

Neues Tabellenblatt:

Neue Arbeitsmappe

OK  
Abbrechen  
Hilfe



## Bi- und Multivariate Datenanalyse

### ANOVA

Anova: Zweifaktorielle Varianzanalyse mit Messwiederholung						
ZUSAMMENFASSUNG	Klasse A	Klasse B	Gesamt			
<b>Test 1</b>						
Anzahl	5	5	10			
Summe	403	307	710			
Mittelwert	80,6	61,4	71			
Varianz	373,8	586,3	529,1111111			
<b>Test 2</b>						
Anzahl	5	5	10			
Summe	262	281	543			
Mittelwert	52,4	56,2	54,3			
Varianz	1441,3	745,7	976,0111111			
<b>Gesamt</b>						
Anzahl	10	10				
Summe	665	588				
Mittelwert	66,5	58,8				
Varianz	1027,611111	599,5111111				
<b>ANOVA</b>						
Streuungsursache	Quadratsummen (SS)	Freiheitsgrade (df)	Mittlere Quadratsumme (MS)	Prüfgröße (F)	P-Wert	kritischer F-Wert
Stichprobe	1394,45	1	1394,45	1,772361857	0,20174879	4,49399848
Spalten	296,45	1	296,45	0,376791332	0,547954297	4,49399848
Wechselwirkung	661,25	1	661,25	0,840456293	0,372878503	4,49399848
Fehler	12588,4	16	786,775			
Gesamt	14940,55	19				

Unterschied zwischen Test 1 und Test 2? Nein, nicht signifikant

Unterschied zwischen Klasse A und Klasse B? Nein, nicht signifikant

Beeinflussen sich die Faktoren signifikant untereinander? Nein

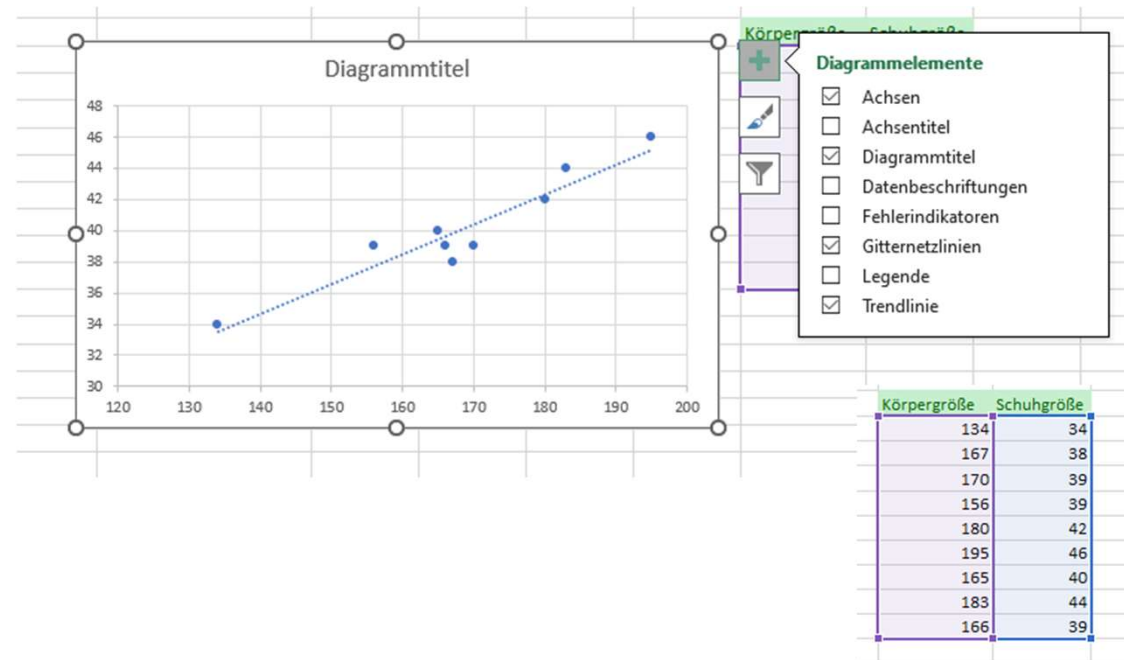
# Multivariate Datenanalyse

## Lineare Regression

Welchen Einfluss haben eine oder mehrere unabhängige Variablen auf eine Abhängige Variable?

Scatterplot-  
Streudiagramm

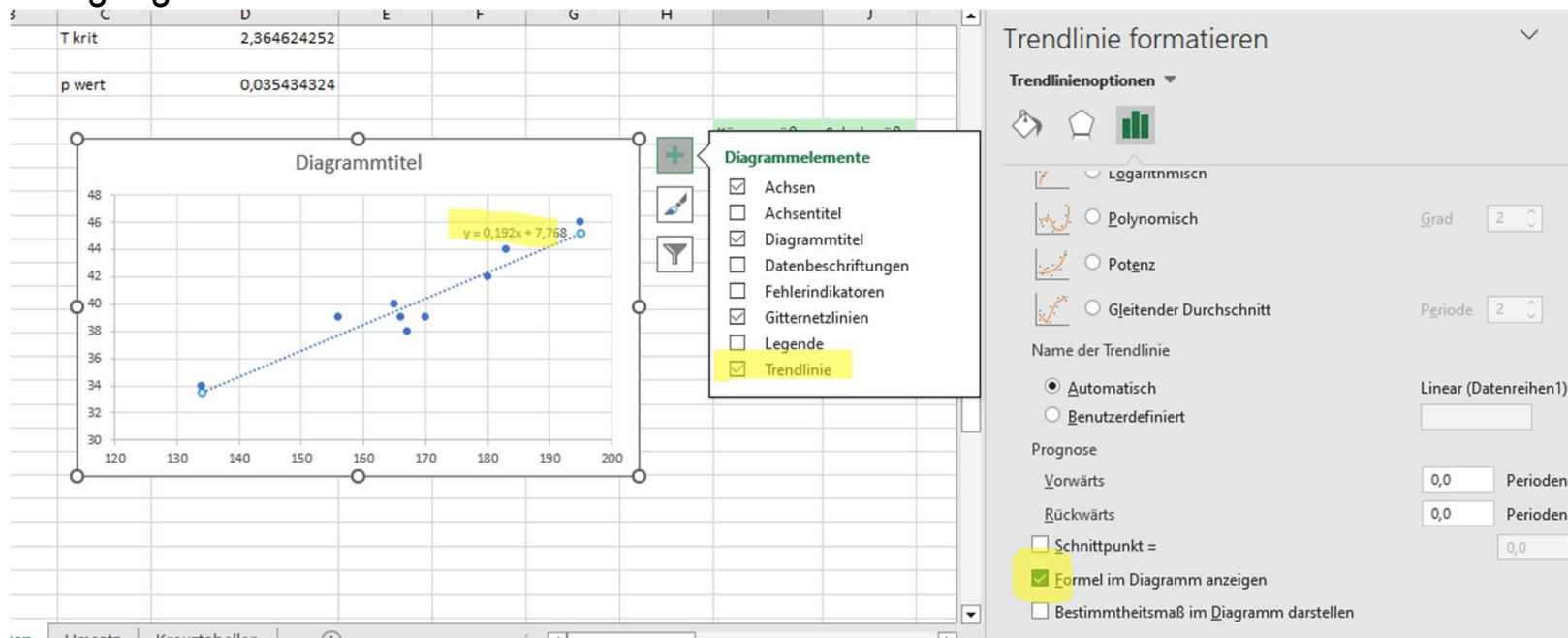
Trend erkennbar



# Multivariate Datenanalyse

## Lineare Regression

Regressionsgerade und Regressionskoeffizienten  $y=b+mx$ ,  $b$  ist Achsenabschnitt,  $m$  ist die Steigung



# Multivariate Datenanalyse

## Lineare Regression

Regressionsgerade und Regressionskoeffizienten  $y=b+mx$ ,

$b$  ist Achsenabschnitt,

$$b = \bar{y} - m \cdot \bar{x}$$

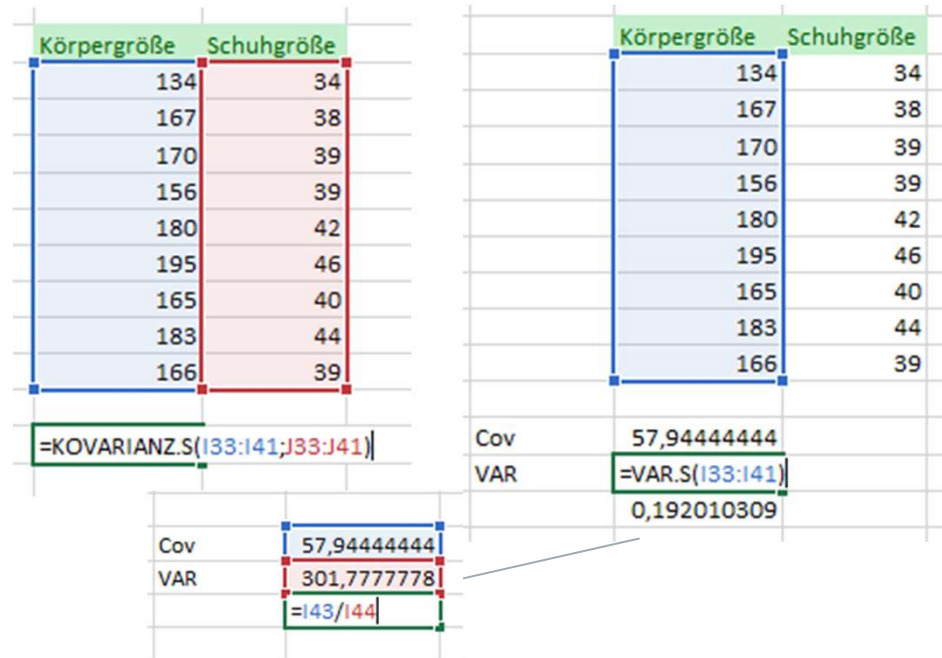
$m$  ist die Steigung

$$m = \frac{\text{Cov}[x, y]}{\text{Var}[x]}$$

Sinkt/Steigt  $x$  um eine Einheit, sinkt oder steigt  $y$  um  $m$  Einheiten

Negatives  $V_z$  -> sinkt

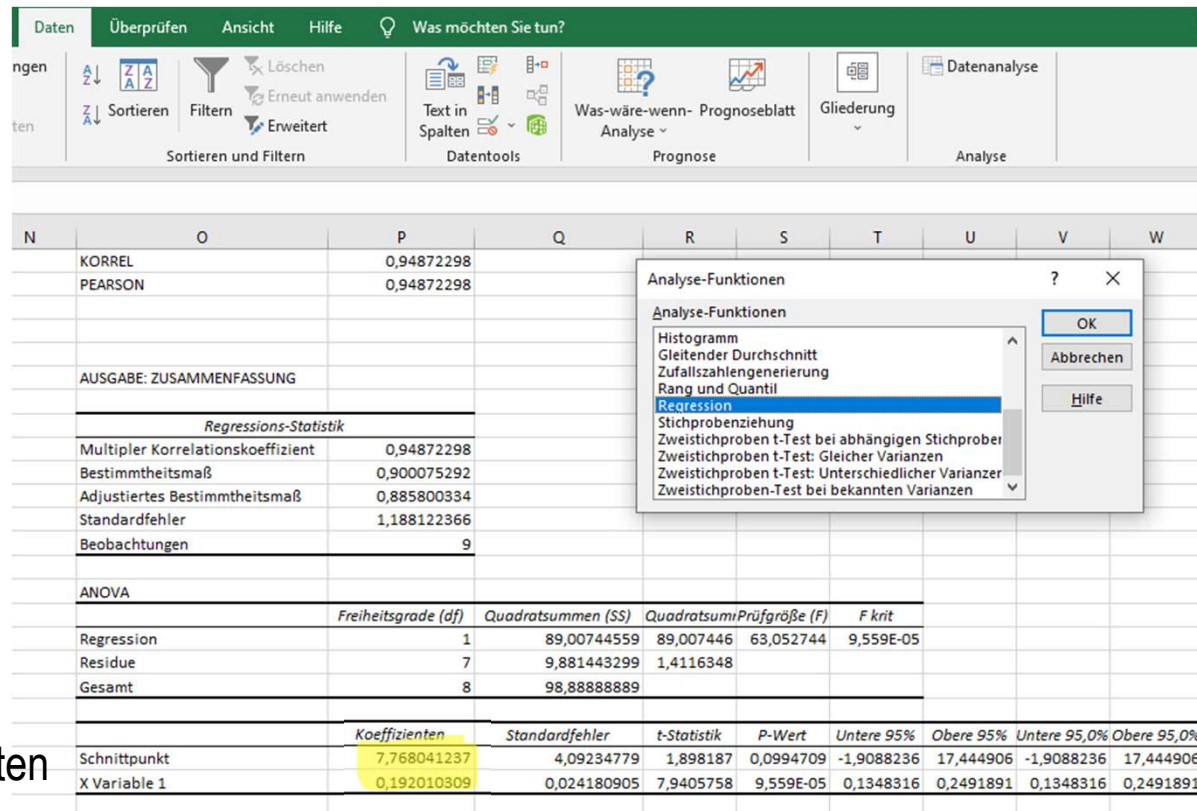
Positives  $V_z$  -> steigt



# Multivariate Datenanalyse

## Lineare Regression

Erweiterte Statistik über die Datenanalysefunktion



The screenshot displays the Microsoft Excel interface with the 'Datenanalyse' (Data Analysis) task pane open. The 'Regression' option is selected in the list of analysis tools. The background spreadsheet shows the following data:

N	O	P	Q	R	S	T	U	V	W
	KORREL		0,94872298						
	PEARSON		0,94872298						
AUSGABE: ZUSAMMENFASSUNG									
<b>Regressions-Statistik</b>									
	Multipler Korrelationskoeffizient		0,94872298						
	Bestimmtheitsmaß		0,900075292						
	Adjustiertes Bestimmtheitsmaß		0,885800334						
	Standardfehler		1,188122366						
	Beobachtungen		9						
ANOVA									
		Freiheitsgrade (df)	Quadratsummen (SS)	Quadratsumme/Prüfgröße (F)	F krit				
	Regression	1	89,00744559	89,007446	63,052744	9,559E-05			
	Residue	7	9,881443299	1,4116348					
	Gesamt	8	98,88888889						
		Koeffizienten	Standardfehler	t-Statistik	P-Wert	Untere 95%	Obere 95%	Untere 95,0%	Obere 95,0%
	Schnittpunkt	7,768041237	4,09234779	1,898187	0,0994709	-1,9088236	17,444906	-1,9088236	17,444906
	X Variable 1	0,192010309	0,024180905	7,9405758	9,559E-05	0,1348316	0,2491891	0,1348316	0,2491891

Regressionskoeffizienten

# Multivariate Datenanalyse

## Lineare Regression

Bestimmtheitsmaß erklärt, welcher Anteil der Streuung von dem Regressionsmodell erklärt wird, wie gut also die lineare Regression zu den Daten passt.  $R^2$  liegt zwischen 0 und 1 (0=erklärt keinen Teil der Streuung, 1=erklärt perfekt die Streuung der Daten)

