



KATHOLISCHE UNIVERSITÄT
EICHSTÄTT-INGOLSTADT

Datenanalyse mit Excel: Eine Einführung

Was sind Daten und wie kann man sie analysieren?

- Daten sind Informationen, die wir messen oder sammeln können, z. B. wie viele Tassen Kaffee jemand pro Tag trinkt oder in welchem Studiengang jemand eingeschrieben ist.
- Solche Daten nennt man empirisch: Das sind Erkenntnisse oder Beobachtungen, die auf Erfahrung, Messung oder Experimenten in der Realität beruhen.
- Datenanalyse bedeutet, diese Informationen systematisch zu untersuchen, um Muster, Zusammenhänge oder Bedeutungen zu erkennen.
- Je nach Fragestellung und Art der Daten kommen dabei unterschiedliche Methoden zum Einsatz

➡ Durch die Analyse von Daten können wir also besser verstehen, wie und warum Dinge zusammenhängen.



Qualitative vs. quantitative Datenanalyse

In der Forschung unterscheidet man zwischen qualitativer und quantitativer Datenanalyse; wesentliche Unterschiede sind in nachfolgender Tabelle zusammengefasst (stark vereinfacht):

	Qualitative Forschung	Quantitative Forschung
Fokus	Ideen erforschen oder Hypothesen/Theorien formulieren	Überprüfung von Hypothesen und Theorien
Datenstruktur	Texte, Bilder	Zahlen, Messwerte
Ziel der Analyse	Tiefes Verständnis einzelner Fälle oder Phänomene	Generalisierbare Aussagen über viele Fälle
Stichprobe	eher klein, gezielt ausgewählt	eher groß, repräsentativ
Ergebnisse ausgedrückt in ...	Wörter	Kennzahlen, Diagramme, Tabellen

Häufig werden qualitative und quantitative Methoden auch kombiniert, um unterschiedliche Forschungsfragen umfassend zu bearbeiten.

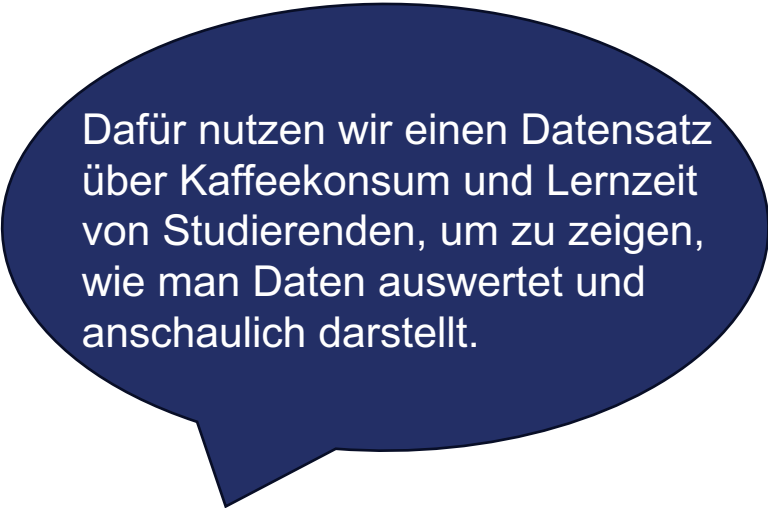
Unser Fokus: Quantitative Datenanalyse mit Excel

In dieser Präsentation konzentrieren wir uns auf die quantitative Analyse von Daten, also auf mess- und zählbare Informationen wie z. B.

- Anzahl der getrunkenen Tassen pro Tag
- Lernzeit in Stunden pro Woche
- Alter oder Semesteranzahl

Diese Daten können wir mit statistischen Methoden in Excel untersuchen, z. B.

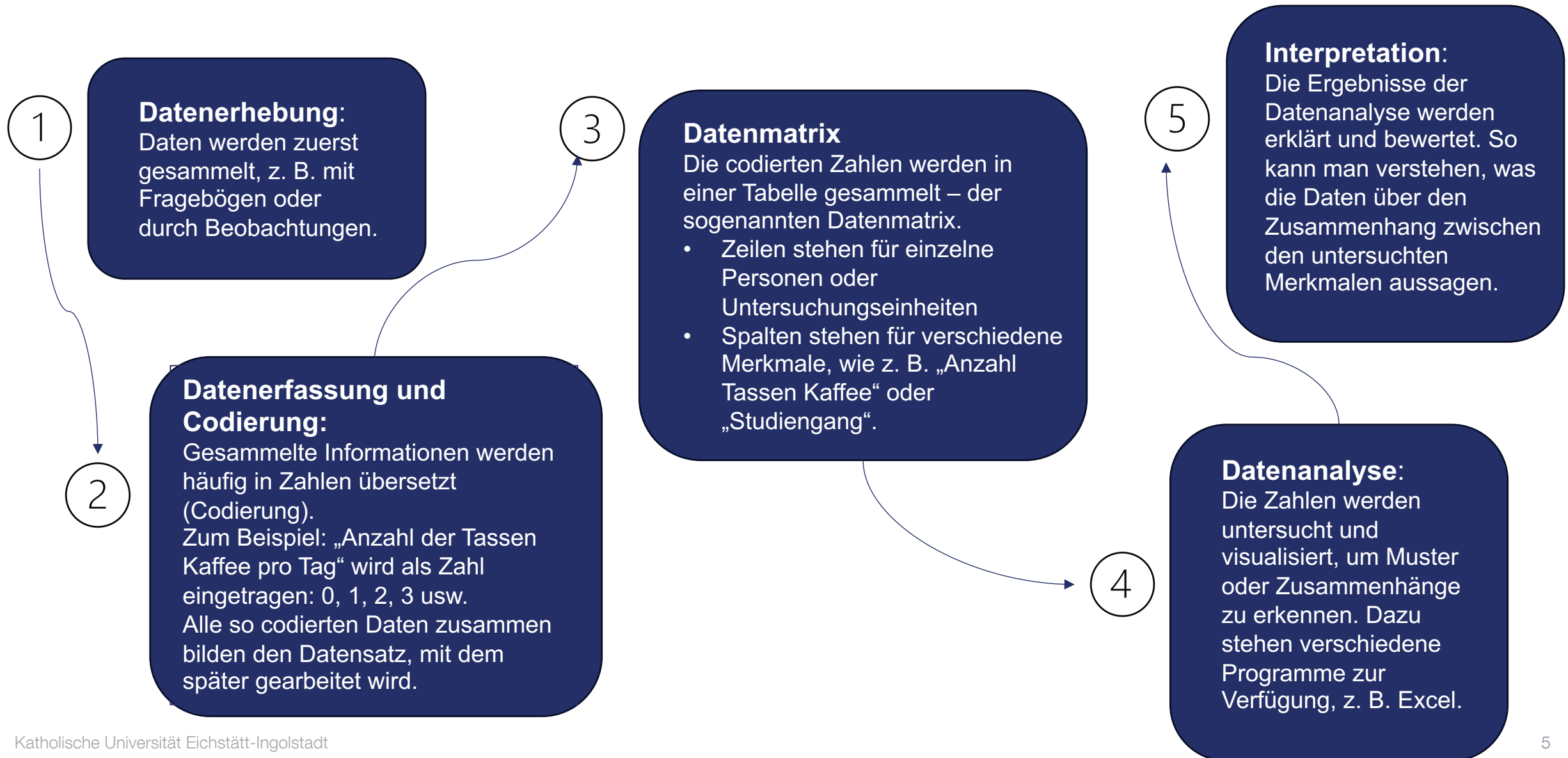
- Mittelwerte berechnen
- Zusammenhänge sichtbar machen
- Ergebnisse in Diagrammen visualisieren



Dafür nutzen wir einen Datensatz über Kaffeekonsum und Lernzeit von Studierenden, um zu zeigen, wie man Daten auswertet und anschaulich darstellt.

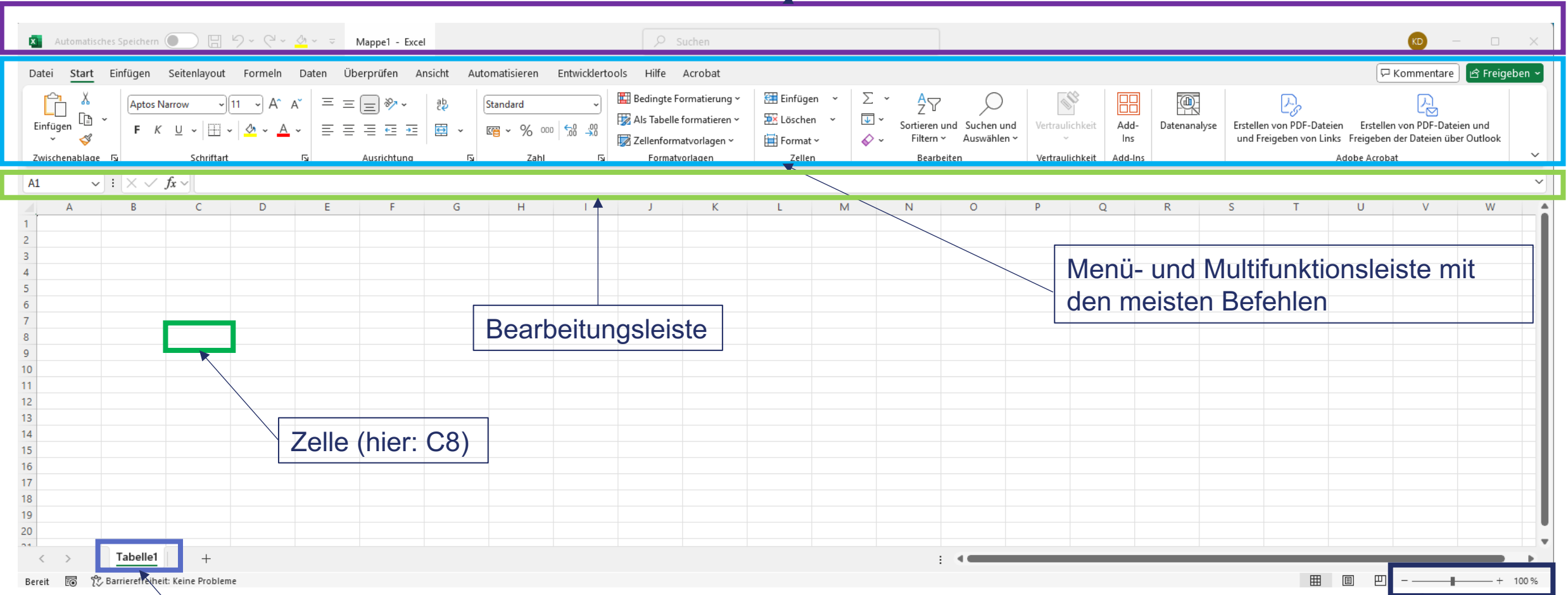
➡ Ziel ist es, grundlegende Techniken zur Auswertung und Darstellung solcher Daten mit Excel kennenzulernen!

Einen Schritt zurück: Von der Erhebung zur Analyse und Interpretation



Excel-Oberfläche (Windows)

Titelleiste



Menü- und Multifunktionsleiste mit den meisten Befehlen

Bearbeitungsleiste

Zelle (hier: C8)

Tabellenblatt

Zoom-Funktion

Beispieldatensatz: Aufbau eines Datensatzes in Excel

ID	Studiengang	Semesterzahl	Tassen pro Tag	Bevorzugter Ort	Kaffeeart	Lernzeit pro Tag	Geschlecht
1	Psychologie	1	4	Café	mit Pflanzenmilch	3 w	w
2	Psychologie	3	1	Zuhause	mit Pflanzenmilch	2 w	w
3	Psychologie	10	0			1 w	w
4	Soziologie	9	2	Campus	Schwarz	3 w	w
5	Anglistik/Amerikanistik	2	5	Campus	Schwarz	6 m	m
6	Soziologie	7	4	Campus	mit Pflanzenmilch	5 m	m
7	Soziologie	6	1	Café	mit Milch	2 w	w
8	Soziologie	4	0			1 m	m
9	Soziale Arbeit	3	1	Campus	mit Pflanzenmilch	2 w	w
10	Anglistik/Amerikanistik	2	0			1 m	m
11	Germanistik	8	1	Zuhause	mit Milch	2 m	m
12	Soziologie	2	2	Café	mit Milch	3 m	m
13	Germanistik	8	5	Campus	Schwarz	6 m	m
14	Anglistik/Amerikanistik	9	5	Zuhause	Espresso	6 w	w
15	Anglistik/Amerikanistik	7	2	Café	Cappuccino	3 w	w
16	Anglistik/Amerikanistik	10	6	Campus	mit Pflanzenmilch	7 w	w
17	Mathematik	9	0			1 w	w
18	Soziale Arbeit	6	4	Café	Cappuccino	5 w	w
19	Anglistik/Amerikanistik	7	3	Zuhause	mit Milch	4 m	m
20	Musik	9	0			1 w	w
21	Germanistik	8	5	Café	mit Milch	6 m	m
22	Musik	1	1	Campus	Cappuccino	2 w	w
23	Mathematik	10	2	Café	Cappuccino	3 m	m
24	Soziologie	7	4	Campus	mit Milch	5 w	w
25	Psychologie	4	3	Campus	mit Pflanzenmilch	4 m	m
26	Anglistik/Amerikanistik	3	0			1 w	w
27	Psychologie	8	1	Campus	mit Milch	2 w	w
28	Mathematik	8	3	Campus	Schwarz	1 w	w
29	Soziologie	3	1	Café	mit Pflanzenmilch	1 m	m
30	Psychologie	8	0			1 m	m

In der obersten Zeile werden die abgefragten Merkmale (Variablen) eingetragen, z. B. ID, Studiengang, Semesterzahl

Jede Zeile = eine befragte Person (hier die Person mit der ID 7)
→ enthält alle abgefragten Informationen zu dieser Person

Jede Spalte = eine Variable
→ z. B. Studiengang, Semesterzahl (hier), getrunkene Kaffeetassen pro Tag

So kann man aus der Tabelle zum Beispiel auf den ersten Blick herauslesen, dass ...

- Person 2 (ID 2) im 3. Semester Psychologie studiert und 1 Tasse Kaffee pro Tag trinkt, bevorzugt zu Hause.
- Person 10 keinen Kaffee trinkt und eine 1 Stunde pro Tag lernt
- Einige Personen im Vergleich viel Kaffee trinken (6 Tassen), andere überhaupt keinen (0 Tassen)
- Insgesamt 30 Personen an der Befragung teilgenommen haben, davon 17 weibliche und 13 männliche.
- ...

Beispieldatensatz: Grundsätzliches

Datei	Start	Einfügen	Seitenlayout	Formeln	Daten	Überprüfen	Ansicht	Automatisieren	Entwicklertools	Hilfe
Einfügen	Schriftart	Ausrichtung	Zahl	Bedingte Formatierung Als Tabelle formatieren Zellenformatvorlagen	Zellen	Bearbeiten	Vertraulichkeit	Add-Ins		
Zwischenablage				Formatvorlagen			Vertraulichkeit	Add-Ins		
L18										
	A	B	C	D	E	F	G	H		
	ID	Studiengang	Semesterzahl	Tassen pro Tag	Bevorzugter Ort	Kaffeeart	Lernzeit pro Tag	Geschlecht		
1	1	Psychologie	1	4	Café	mit Pflanzenmilch	5 w			
2	2	Psychologie	3	1	Zuhause	mit Pflanzenmilch	2 w			
3	3	Psychologie	10	0			1 w			
4	4	Soziologie	9	2	Campus	Schwarz	3 w			
5	5	Anglistik/Amerikanistik	2	5	Campus	Schwarz	6 m			
6	6	Soziologie	7	4	Campus	mit Pflanzenmilch	5 m			
7	7	Soziologie	6	1	Café	mit Milch	2 w			
8	8	Soziologie	4	0			1 m			
9	9	Soziale Arbeit	3	1	Campus	mit Pflanzenmilch	2 w			
10	10	Anglistik/Amerikanistik	2	0			1 m			
11	11	Germanistik	8	1	Zuhause	mit Milch	2 m			
12	12	Soziologie	2	2	Café	mit Milch	3 m			
13	13	Germanistik	8	5	Campus	Schwarz	6 m			
14	14	Anglistik/Amerikanistik	9	5	Zuhause	Espresso	6 w			
15	15	Anglistik/Amerikanistik	7	2	Café	Cappuccino	3 w			
16	16	Anglistik/Amerikanistik	10	6	Campus	mit Pflanzenmilch	7 w			
17	17	Mathematik	9	0			1 w			
18	18	Soziale Arbeit	6	4	Café	Cappuccino	5 w			
19	19	Anglistik/Amerikanistik	7	3	Zuhause	mit Milch	4 m			
20	20	Musik	9	0			1 w			
21	21	Germanistik	8	5	Café	mit Milch	6 m			
22	22	Musik	1	1	Campus	Cappuccino	2 w			
23	23	Mathematik	10	2	Café	Cappuccino	3 m			
24	24	Soziologie	7	4	Campus	mit Milch	5 w			
25	25	Psychologie	4	3	Campus	mit Pflanzenmilch	4 m			
26	26	Anglistik/Amerikanistik	3	0			1 w			
27	27	Psychologie	8	1	Campus	mit Milch	2 w			
28	28	Mathematik	8	3	Campus	Schwarz	1 w			
29	29	Soziologie	3	1	Café	mit Pflanzenmilch	1 m			
30	30	Psychologie	8	0			1 m			

Bei diesem Datensatz handelt es sich um ein fiktives Beispiel.

- Wir nehmen an, dass eine Befragung mit insgesamt 30 Studierenden durchgeführt wurde
- Diese wurden gebeten, folgende Angaben über sich zu machen:
 - Eingeschriebener Studiengang
 - Aktuelle Semesterzahl
 - Anzahl der pro Tag getrunkenen Tassen Kaffee
 - Bevorzugter Ort, an dem Kaffee getrunken wird
 - Bevorzugte Kaffeeart
 - Lernzeit pro Tag in Stunden
 - Geschlecht

Datenaufbereitung: Daten vorbereiten, bevor man sie analysiert

Bevor die Daten ausgewertet werden können, müssen sie noch einmal überprüft und ggf. vereinheitlicht werden. Das beinhaltet z. B.:

- Einheitliche und richtige Schreibweise prüfen (z. B. bei der Variable *Bevorzugter Ort* die Ausprägung *zu Hause*, nicht *zu Hause* oder *zuhause*)
- Zahlen prüfen: Sind die Werte wirklich sinnvoll? (Z. B. ist zu vermuten, dass niemand 500 Tassen Kaffee pro Tag trinkt) → Möglichkeiten: Abgleich mit dem Fragebogen, ob ein Eingabefehler vorliegt, oder Daten bereinigen (z. B. Wert als fehlend kennzeichnen)
- Formatierungen anpassen: Zahlen als Zahlen, Text als Text, keine Mischformate → das kann man durch einen Rechtsklick in die entsprechenden Zellen unter „Zellen formatieren“ prüfen
- Dubletten löschen: Wenn z. B. dieselbe Person aus Versehen zweimal in der Liste steht, wird eine der Zeilen entfernt, damit die Daten nicht verfälscht werden
- Zeilen mit fehlenden Werten löschen oder als fehlend markieren



Hilfreiche Excel-Befehle:

- **SÄUBERN():** Entfernt nicht druckbare Zeichen aus Texten, z. B. unsichtbare Steuerzeichen, die beim Kopieren aus anderen Programmen mitkommen können.
- **GLÄTTEN():** Entfernt überflüssige Leerzeichen, also z. B. mehrere Leerzeichen zwischen Wörtern oder am Anfang/Ende eines Textes.

Datenanalyse: Berechnen von (statistischen) Kennzahlen

- Kennzahlen sind Zahlen, die wichtige Eigenschaften von Daten zusammenfassen.
- Sie helfen, Muster und Unterschiede in den Daten leichter zu erkennen.
- Beispiele für Kennzahlen:
 - Durchschnitt (Mittelwert): das arithmetische Mittel aller Werte, z. B. die durchschnittliche Anzahl der täglich getrunkenen Tassen Kaffee.
 - Median: der Wert, der in der Mitte liegt, wenn alle Werte der Größe nach geordnet sind; er ist weniger empfindlich gegenüber extrem hohen oder niedrigen Werten (Ausreißern) als der Mittelwert.
 - Modalwert (Modus): der Wert, der am häufigsten vorkommt, z. B. die am häufigsten angegebene Anzahl an Tassen Kaffee im Datensatz.
 - Absolute Häufigkeit: wie oft ein bestimmter Wert im Datensatz (absolut) vorkommt (Anzahl), z. B. wie viele Studierende 0 Tassen Kaffee trinken
 - Relative Häufigkeit: wie viel Prozent der Befragten einen bestimmten Wert haben, z. B. wie viel Prozent der Studierenden 0 Tassen Kaffee trinken
 - Minimum: der kleinste Wert in einem Bereich, z. B. bei „Semesterzahl“ die niedrigste angegebene Semesterzahl
 - Maximum: der größte Wert in einem Bereich, z. B. bei „Semesterzahl“ die höchste angegebene Semesterzahl
- Diese Kennzahlen muss man in Excel nicht zwingend selbst berechnen, sondern kann man sich über Formeln ausgeben lassen.

Datenanalyse: Eingeben von Formeln in Excel

Beispiel: Berechnung des Mittelwerts der Zellen D2 bis D31

Schritt 1: Zelle auswählen

→ Auf die Zelle klicken, in der das Ergebnis der Formel erscheinen soll (meist direkt unter den Daten, hier dann D32)

Schritt 2: Formel eingeben

→ Zuerst ein Gleichheitszeichen (=) eintippen, damit Excel weiß, dass eine Formel eingegeben wird; dann die entsprechende Formel (hier für den Mittelwert) eintippen unter Berücksichtigung des Bereiches, der in die Berechnung eingehen soll (hier D2 bis D31), z. B. =MITTELLWERT(D2:D31)
→ Wichtig: ohne Leerzeichen, mit runden Klammern für den umschließenden Bereich

→ Danach Enter drücken

Schritt 3: Formel überprüfen

→ Wenn man die Zelle auswählt, kann man oben in der Bearbeitungsleiste die eingegebene Formel sehen

D32	=MITTELWERT(D2:D31)						
	A	B	C	D	E	F	G
1	ID	Studiengang	Semesterzahl	Tassen pro Tag	Bevorzugter Ort	Kaffeeart	Lernzeit pro Tag
2	1	Psychologie	1	4	Café	mit Pflanzenmilch	5
3	2	Psychologie	3	1	Zuhause	mit Pflanzenmilch	2
4	3	Psychologie	10	0			1
5	4	Soziologie	9	2	Campus	Schwarz	3
6	5	Anglistik/Amerikanistik	2	5	Campus	Schwarz	6
7	6	Soziologie	7	4	Campus	mit Pflanzenmilch	5
8	7	Soziologie	6	1	Café	mit Milch	2
9	8	Soziologie	4	0			1
10	9	Soziale Arbeit	3	1	Campus	mit Pflanzenmilch	2
11	10	Anglistik/Amerikanistik	2	0			1
12	11	Germanistik	8	1	Zuhause	mit Milch	2
13	12	Soziologie	2	2	Café	mit Milch	3
14	13	Germanistik	8	5	Campus	Schwarz	6
15	14	Anglistik/Amerikanistik	9	5	Zuhause	Espresso	6
16	15	Anglistik/Amerikanistik	7	2	Café	Cappuccino	3
17	16	Anglistik/Amerikanistik	10	6	Campus	mit Pflanzenmilch	7
18	17	Mathematik	9	0			1
19	18	Soziale Arbeit	6	4	Café	Cappuccino	5
20	19	Anglistik/Amerikanistik	7	3	Zuhause	mit Milch	4
21	20	Musik	9	0			1
22	21	Germanistik	8	5	Café	mit Milch	6
23	22	Musik	1	1	Campus	Cappuccino	2
24	23	Mathematik	10	2	Café	Cappuccino	3
25	24	Soziologie	7	4	Campus	mit Milch	5
26	25	Psychologie	4	3	Campus	mit Pflanzenmilch	4
27	26	Anglistik/Amerikanistik	3	0			1
28	27	Psychologie	8	1	Campus	mit Milch	2
29	28	Mathematik	8	3	Campus	Schwarz	1
30	29	Soziologie	3	1	Café	mit Pflanzenmilch	1
31	30	Psychologie	8	0			1
32				2,2			

Datenanalyse: Wichtige Formeln in Excel

Kennzahl	Excel-Formel
Mittelwert	=MITTELWERT(D2:D31)
Median	=MEDIAN(D2:D31)
Modalwert	=MODALWERT(D2:D31)
Summe	=SUMME(D2:D31)
Absolute Häufigkeit	=ZÄHLENWENN(D2:D31;Wert)
Relative Häufigkeit	=ZÄHLENWENN(D2:D31;Wert)/ANZAHL(D2:D31)
Minimum	=MIN(D2:D31)
Maximum	=MAX(D2:D31)



Hinweis:

In den Klammern D2:D31 steht der Zellbereich, auf den sich die Berechnung bezieht. Dieser muss je nach Bereich angepasst werden!

Datenanalyse: Kennzahlen im Beispieldatensatz

	A	B	C	D	E	F	G
1	ID	Studiengang	Semesterzahl	Tassen pro Tag	Bevorzugter Ort	Kaffeeart	Lernzeit pro Tag
2	1	Psychologie	1	4	Café	mit Pflanzenmilch	5
3	2	Psychologie	3	1	Zuhause	mit Pflanzenmilch	2
4	3	Psychologie	10	0			1
5	4	Soziologie	9	2	Campus	Schwarz	3
6	5	Anglistik/Amerikanistik	2	5	Campus	Schwarz	6
7	6	Soziologie	7	4	Campus	mit Pflanzenmilch	5
8	7	Soziologie	6	1	Café	mit Milch	2
9	8	Soziologie	4	0			1
10	9	Soziale Arbeit	3	1	Campus	mit Pflanzenmilch	2
11	10	Anglistik/Amerikanistik	2	0			1
12	11	Germanistik	8	1	Zuhause	mit Milch	2
13	12	Soziologie	2	2	Café	mit Milch	3
14	13	Germanistik	8	5	Campus	Schwarz	6
15	14	Anglistik/Amerikanistik	9	5	Zuhause	Espresso	6
16	15	Anglistik/Amerikanistik	7	2	Café	Cappuccino	3
17	16	Anglistik/Amerikanistik	10	6	Campus	mit Pflanzenmilch	7
18	17	Mathematik	9	0			1
19	18	Soziale Arbeit	6	4	Café	Cappuccino	5
20	19	Anglistik/Amerikanistik	7	3	Zuhause	mit Milch	4
21	20	Musik	9	0			1
22	21	Germanistik	8	5	Café	mit Milch	6
23	22	Musik	1	1	Campus	Cappuccino	2
24	23	Mathematik	10	2	Café	Cappuccino	3
25	24	Soziologie	7	4	Campus	mit Milch	5
26	25	Psychologie	4	3	Campus	mit Pflanzenmilch	4
27	26	Anglistik/Amerikanistik	3	0			1
28	27	Psychologie	8	1	Campus	mit Milch	2
29	28	Mathematik	8	3	Campus	Schwarz	1
30	29	Soziologie	3	1	Café	mit Pflanzenmilch	1
31	30	Psychologie	8	0			1
32				2,2			



Hinweise zum sog. Skalenniveau:
Nicht alle Kennzahlen passen zu jeder Art von Daten:

- Mittelwert und Median sind nur sinnvoll bei Zahlen (z. B. Tassen Kaffee, Semesterzahl).
- Der Modalwert kann auch bei kategorischen Daten verwendet werden (z. B. Studiengang, bevorzugter Ort).
- Das heißt: Man muss vorher wissen, welche Art von Daten man hat, um die richtige Kennzahl zu wählen.

Durchschnittlich 2,2 Tassen pro Tag werden von den Befragten getrunken.

Datenanalyse: Kennzahlen im Beispieldatensatz

	A	B	C	D	E	F	G
1	ID	Studiengang	Semesterzahl	Tassen pro Tag	Bevorzugter Ort	Kaffeeart	Lernzeit pro Tag
2	1	Psychologie	1	4	Café	mit Pflanzenmilch	5
3	2	Psychologie	3	1	Zuhause	mit Pflanzenmilch	2
4	3	Psychologie	10	0			1
5	4	Soziologie	9	2	Campus	Schwarz	3
6	5	Anglistik/Amerikanistik	2	5	Campus	Schwarz	6
7	6	Soziologie	7	4	Campus	mit Pflanzenmilch	5
8	7	Soziologie	6	1	Café	mit Milch	2
9	8	Soziologie	4	0			1
10	9	Soziale Arbeit	3	1	Campus	mit Pflanzenmilch	2
11	10	Anglistik/Amerikanistik	2	0			1
12	11	Germanistik	8	1	Zuhause	mit Milch	2
13	12	Soziologie	2	2	Café	mit Milch	3
14	13	Germanistik	8	5	Campus	Schwarz	6
15	14	Anglistik/Amerikanistik	9	5	Zuhause	Espresso	6
16	15	Anglistik/Amerikanistik	7	2	Café	Cappuccino	3
17	16	Anglistik/Amerikanistik	10	6	Campus	mit Pflanzenmilch	7
18	17	Mathematik	9	0			1
19	18	Soziale Arbeit	6	4	Café	Cappuccino	5
20	19	Anglistik/Amerikanistik	7	3	Zuhause	mit Milch	4
21	20	Musik	9	0			1
22	21	Germanistik	8	5	Café	mit Milch	6
23	22	Musik	1	1	Campus	Cappuccino	2
24	23	Mathematik	10	2	Café	Cappuccino	3
25	24	Soziologie	7	4	Campus	mit Milch	5
26	25	Psychologie	4	3	Campus	mit Pflanzenmilch	4
27	26	Anglistik/Amerikanistik	3	0			1
28	27	Psychologie	8	1	Campus	mit Milch	2
29	28	Mathematik	8	3	Campus	Schwarz	1
30	29	Soziologie	3	1	Café	mit Pflanzenmilch	1
31	30	Psychologie	8	0			1
32			8				

Der häufigste Wert in der Datenreihe ist 8, die meisten Studierenden sind also im 8. Semester.

Datenanalyse: Kennzahlen im Beispieldatensatz

G33		:		=ZÄHLENWENN(G2:G31;6)/ANZAHL(G2:G31)			
	A	B	C	D	E	F	G
1	ID	Studiengang	Semesterzahl	Tassen pro Tag	Bevorzugter Ort	Kaffeeart	Lernzeit pro Tag
2	1	Psychologie	1	4	Café	mit Pflanzenmilch	5
3	2	Psychologie	3	1	Zuhause	mit Pflanzenmilch	2
4	3	Psychologie	10	0			1
5	4	Soziologie	9	2	Campus	Schwarz	3
6	5	Anglistik/Amerikanistik	2	5	Campus	Schwarz	6
7	6	Soziologie	7	4	Campus	mit Pflanzenmilch	5
8	7	Soziologie	6	1	Café	mit Milch	2
9	8	Soziologie	4	0			1
10	9	Soziale Arbeit	3	1	Campus	mit Pflanzenmilch	2
11	10	Anglistik/Amerikanistik	2	0			1
12	11	Germanistik	8	1	Zuhause	mit Milch	2
13	12	Soziologie	2	2	Café	mit Milch	3
14	13	Germanistik	8	5	Campus	Schwarz	6
15	14	Anglistik/Amerikanistik	9	5	Zuhause	Espresso	6
16	15	Anglistik/Amerikanistik	7	2	Café	Cappuccino	3
17	16	Anglistik/Amerikanistik	10	6	Campus	mit Pflanzenmilch	7
18	17	Mathematik	9	0			1
19	18	Soziale Arbeit	6	4	Café	Cappuccino	5
20	19	Anglistik/Amerikanistik	7	3	Zuhause	mit Milch	4
21	20	Musik	9	0			1
22	21	Germanistik	8	5	Café	mit Milch	6
23	22	Musik	1	1	Campus	Cappuccino	2
24	23	Mathematik	10	2	Café	Cappuccino	3
25	24	Soziologie	7	4	Campus	mit Milch	5
26	25	Psychologie	4	3	Campus	mit Pflanzenmilch	4
27	26	Anglistik/Amerikanistik	3	0			1
28	27	Psychologie	8	1	Campus	mit Milch	2
29	28	Mathematik	8	3	Campus	Schwarz	1
30	29	Soziologie	3	1	Café	mit Pflanzenmilch	1
31	30	Psychologie	8	0			1
32							4
33							0,13333333

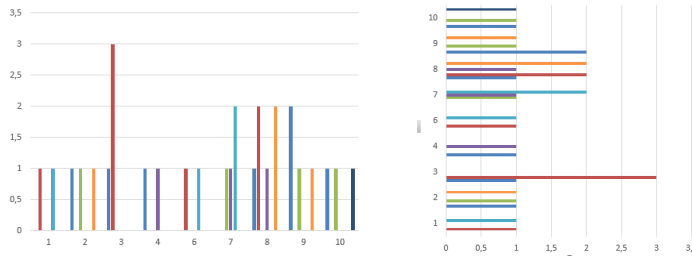
Absolute Häufigkeit:
Insgesamt 4 Studierende geben an, 6 Stunden pro Tag zu lernen.

Relative Häufigkeit:
Das sind 13,33 Prozent (diese Formel ist in der Bearbeitungszeile oben zu sehen)

Visualisierung der Daten: Diagrammarten

Je nachdem, welche Daten oder Ergebnisse man zeigen möchte, eignen sich verschiedene Diagrammtypen:

- Säulen-/ und Balkendiagramm:



→ Diese Diagramme eignen sich zum Vergleich verschiedener Gruppen oder Kategorien. Zum Beispiel kann im Beispieldatensatz gezeigt werden, wie viele Studierende in welchem Semester sind. Jede Säule/jeder Balken steht für ein Semester und ihre Höhe/Breite zeigt die Anzahl der Studierenden in diesem Semester an.

- Kreisdiagramm:



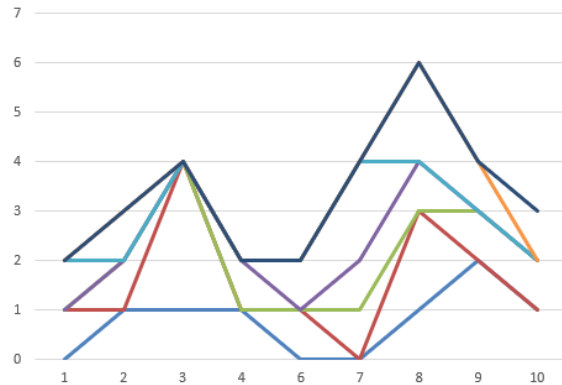
→ Mit einem Kreisdiagramm werden Anteile an einem Ganzen dargestellt. Zum Beispiel kann gezeigt werden, wie sich die bevorzugten Kaffeearten unter den Studierenden verteilen. Der Kreis wird in Teile aufgeteilt, die zeigen, wie groß der Anteil der einzelnen Kaffeearten im Verhältnis zur Gesamtzahl der Befragten ist.

Visualisierung der Daten: Diagrammarten

Je nachdem, welche Daten oder Ergebnisse man zeigen möchte, eignen sich verschiedene Diagrammtypen:

- Liniendiagramm:

→ Liniendiagramme eignen sich zur Darstellung von Veränderungen über eine Zeitspanne oder geordnete Werte. Beispielsweise könnte man die durchschnittliche Anzahl der Tassen Kaffee pro Tag, die ein bestimmter Studierender trinkt, über die Semester hinweg visualisieren (dies ist in diesem Datensatz aber nicht möglich, da dies nicht erhoben wurde).



- ...
- Für weiterführende Analysen gibt es auch Diagrammtypen wie den Boxplot, der Verteilungen und Streuungen von Daten anschaulich macht.

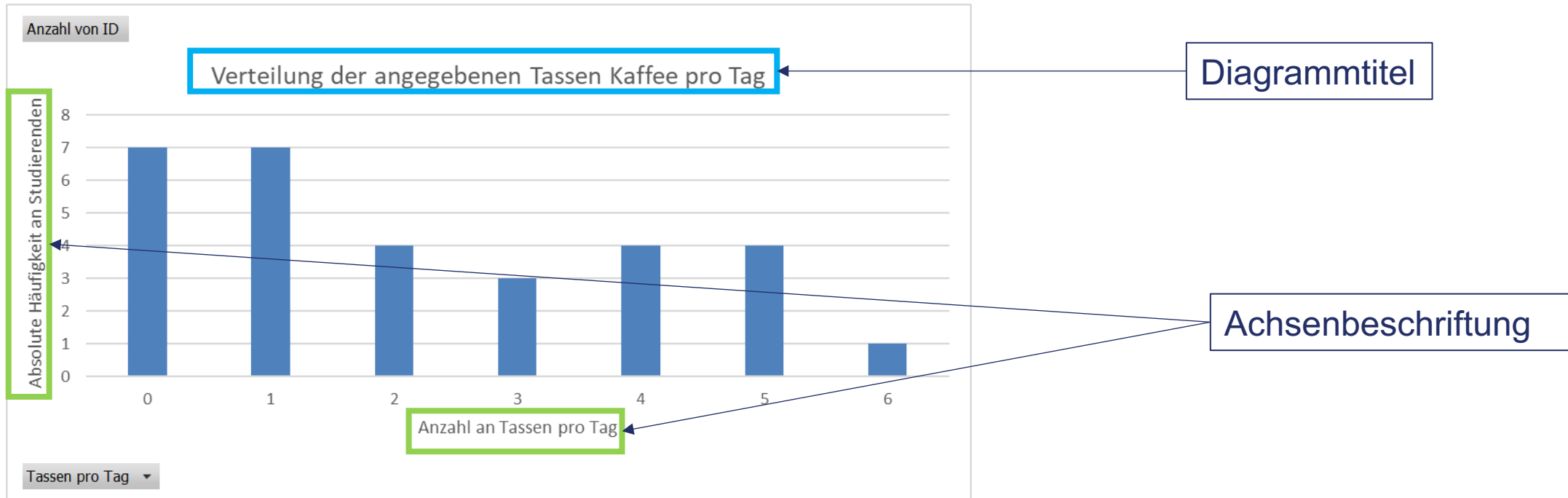
Visualisierung der Daten: Diagrammgestaltung

Damit Diagramme gut verständlich und aussagekräftig sind, sollten einige Gestaltungsregeln beachtet werden:

- Titel
 - Der Titel eines Diagramms sollte kurz und präzise angeben, was dargestellt wird. Ein klarer Titel ermöglicht es, den Inhalt schnell zu erfassen. Beispiel: „Anzahl der Studierenden pro Semester“ oder „Verteilung der bevorzugten Kaffeearten“.
- Achsenbeschriftungen (falls Achsen vorhanden sind)
 - Beide Achsen eines Diagramms müssen deutlich beschriftet sein, um die dargestellten Werte verständlich zu machen. Die x-Achse (waagrecht) zeigt dabei Kategorien oder Zeitpunkte, wie etwa „Semesterzahl“ oder „Kaffeeart“. Die y-Achse (senkrecht) gibt die Werte oder Häufigkeiten an, beispielsweise „Anzahl der Studierenden“ oder „Anteil in Prozent“. Eine fehlende oder unklare Beschriftung erschwert das Verständnis.
- Farben
 - Farben dienen dazu, verschiedene Gruppen oder Kategorien voneinander abzuheben. Dabei sollten gut unterscheidbare Farben gewählt werden, die nicht zu grell sind. Wichtig ist, dass für gleiche Kategorien in verschiedenen Diagrammen immer dieselben Farben verwendet werden. So können Betrachter die Kategorien schneller erkennen und besser vergleichen. Beispiel: Wenn „mit Milch“ in allen Diagrammen blau und „mit Pflanzenmilch“ immer grün dargestellt wird, erleichtert dies das Verständnis und die Orientierung.

Visualisierung der Daten: Diagrammgestaltung

Damit Diagramme gut verständlich und aussagekräftig sind, sollten einige Gestaltungsregeln beachtet werden:



Visualisierung der Daten: Pivot-Tabellen

Wenn Daten nicht wie im Beispieldatensatz vorgruppiert sind, sondern pro Person oder Ereignis mehrere Datenzeilen vorliegen, lassen sich keine sinnvollen Diagramme direkt daraus erstellen.

→ Lösung: Mit einer Pivot-Tabelle lassen sich solche Daten schnell gruppieren, zählen oder mitteln, ohne Formeln verwenden zu müssen. Die Ergebnisse können anschließend direkt in ein Pivot-Diagramm übernommen werden.



Was ist eine Pivot-Tabelle?

= eine spezielle Auswertungstabelle, mit der sich große Datenmengen übersichtlich zusammenfassen lassen.

Man spricht auch von einer Kreuztabelle, weil sie Daten nach Zeilen- und Spaltenkategorien gruppiert.

Beispielhafte Fragen, die sich damit beantworten lassen:

Wie viele Personen bevorzugen welche Kaffeeart?
In welchem Semester sind wie viele Studierende?

Typische Funktionen:

Zählen (z. B. Häufigkeit pro Kategorie)

Mittelwert berechnen

Werte nach Gruppen vergleichen

Visualisierung der Daten: Erstellen von Pivot-Tabellen

Beispiel: Anteile der bevorzugten Kaffeeart unter den Befragten

- Schritt 1: Datensatz markieren
→ Den gesamten Datensatz (inklusive Überschriften) markieren.
- Schritt 2: Pivot-Sheet erstellen
→ Im Reiter „Einfügen“ den Befehl „PivotTable“ wählen, dann „Neues Arbeitsblatt“ für die Position der Tabelle auswählen.
→ Es öffnet sich ein neues Sheet (dieses kann durch einen Doppelklick auf den Namen noch umbenannt werden).

The screenshot shows the Microsoft Excel interface. The 'Einfügen' (Insert) ribbon is active, and the 'PivotTable' button is highlighted. A data table is visible in the background, and a 'PivotTable aus Tabelle oder Bereich' dialog box is open, showing options to create a new pivot table on a new worksheet.

ID	Studiengang	Semesterzahl	Tassen pro Tag	Bevorzugter Ort	Kaffeeart	Lernzeit pro Tag	Geschlecht
1	Psychologie	1	4	Café	mit Pflanzenmilch	5 w	
2	Psychologie	3	1	Zuhause	mit Pflanzenmilch	2 w	
3	Psychologie	10	0			1 w	
4	Soziologie	9	2	Campus	Schwarz	3 w	
5	Anglistik/Amerikanistik	2	5	Campus	Schwarz	6 m	
6	Soziologie	7	4	Campus	mit Pflanzenmilch	5 m	
7	Soziologie	6	1	Café	mit Milch	2 w	
8	Soziologie	4	0			1 m	
9	Soziale Arbeit	3	1	Campus	mit Pflanzenmilch	2 w	
10	Anglistik/Amerikanistik	2	0			1 m	
11	Germanistik	8	1	Zuhause	mit Milch	2 m	
12	Soziologie	2	2	Café	mit Milch	3 m	
13	Germanistik	8	5				
14	Anglistik/Amerikanistik	9	5				
15	Anglistik/Amerikanistik	7	2				
16	Anglistik/Amerikanistik	10	6				
17	Mathematik	9	0				
18	Soziale Arbeit	6	4				
19	Anglistik/Amerikanistik	7	3				
20	Musik	9	0				
21	Germanistik	8	5				
22	Musik	1	1				
23	Mathematik	10	2				
24	Soziologie	7	4				
25	Psychologie	4	3				
26	Anglistik/Amerikanistik	3	0			1 w	
27	Psychologie	8	1	Campus	mit Milch	2 w	
28	Mathematik	8	3	Campus	Schwarz	1 w	
29	Soziologie	3	1	Café	mit Pflanzenmilch	1 m	
30	Psychologie	8	0			1 m	

Visualisierung der Daten: Erstellen von Pivot-Tabellen

Beispiel: Anteile der bevorzugten Kaffeeart unter den Befragten

- Schritt 3: Pivot-Tabelle erstellen

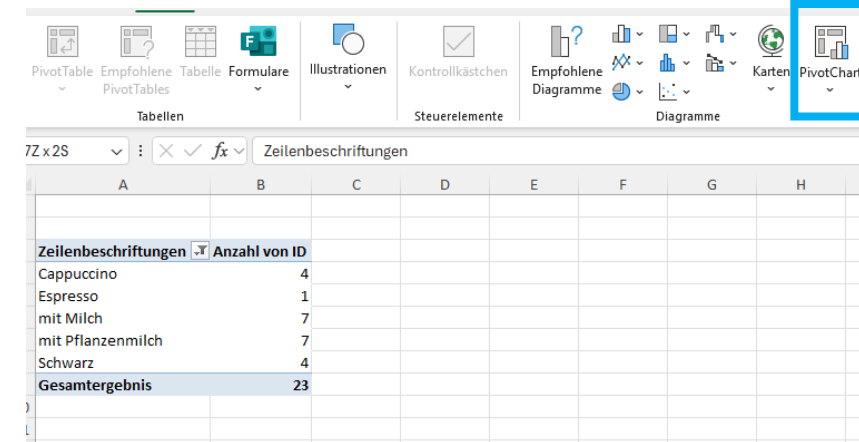
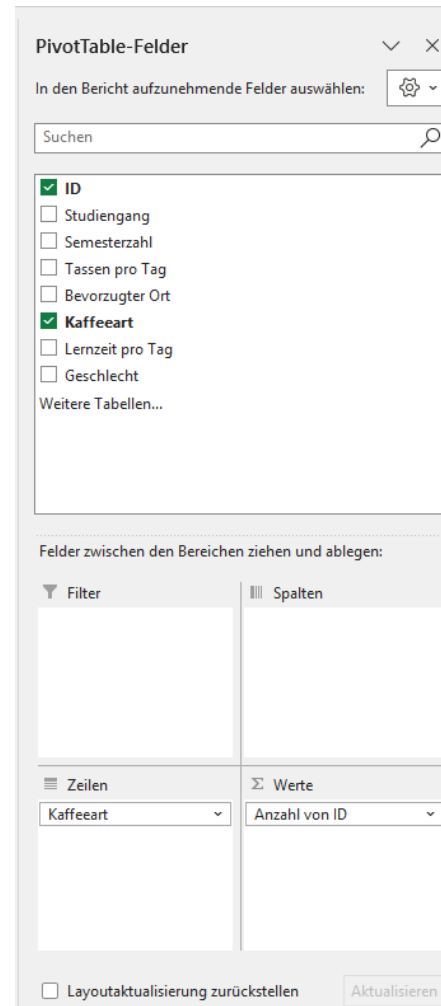
→ Im rechten Feld die Variable „Kaffeeart“ in den Bereich „Zeilen“ ziehen, die Variable „ID“ in den Bereich „Werte“ (Standard: „Anzahl von ...“ → Rechtsklick auf „Wertefeldeinstellung“ → „Anzahl“).

→ Es entsteht eine Tabelle im Sheet.

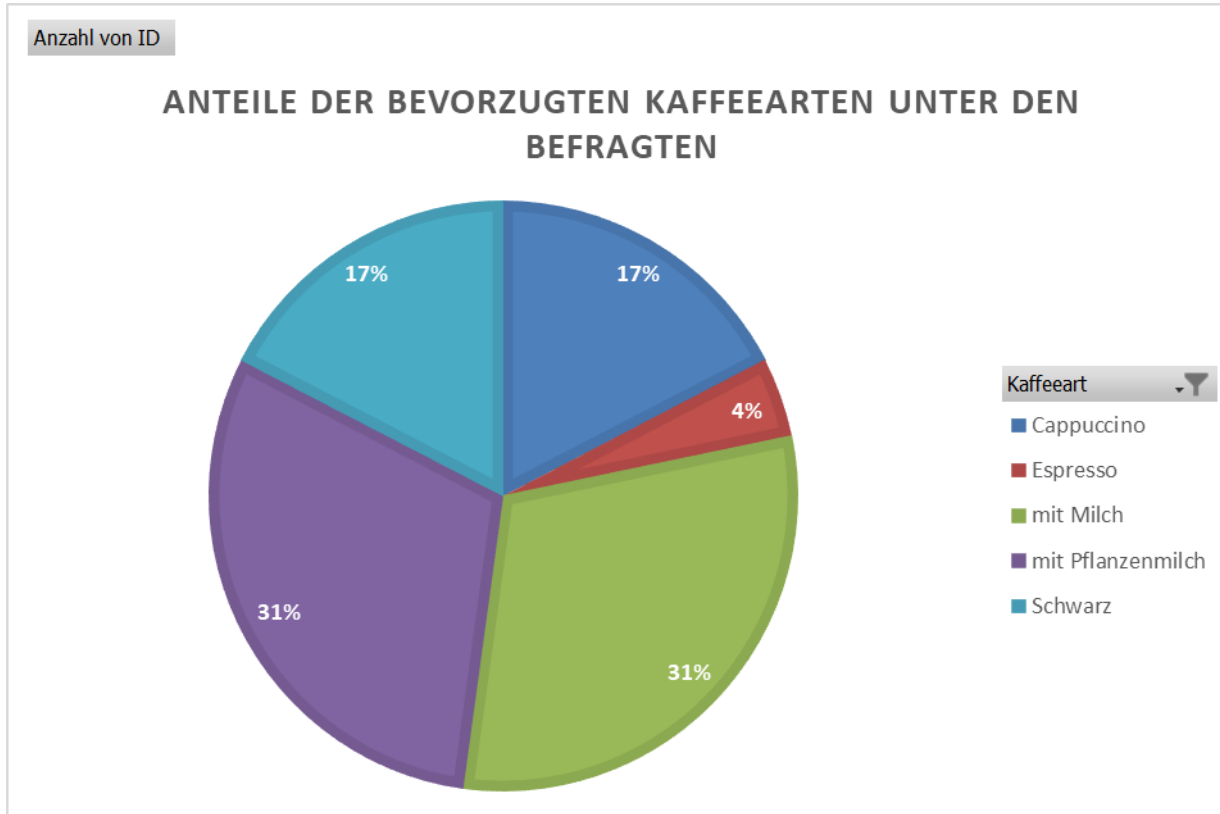
- Schritt 4: Diagramm erstellen

→ Die entstandene Tabelle markieren, mit einem Klick auf „PivotChart“ ein passendes Diagramm, z. B. ein Kreisdiagramm, erstellen.

→ Ggf. die Titel, die Formatvorlage und die Farbe noch anpassen (über den Pinsel rechts oben im Diagramm bzw. über Doppelklick auf Achse).

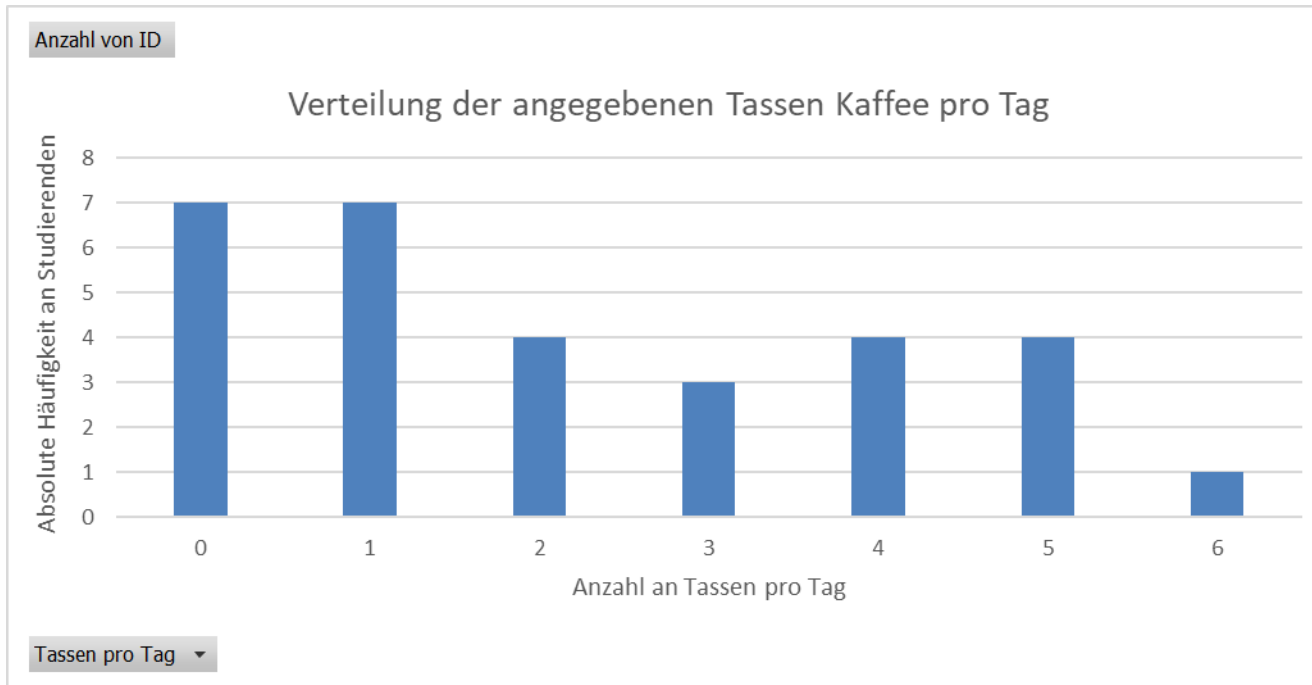


Visualisierung der Daten: Diagramme im Beispieldatensatz

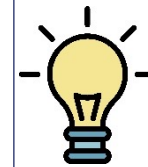


Das Kreisdiagramm zeigt die prozentualen Anteile der bevorzugten Kaffeearten unter den Befragten.
→ Die Darstellung macht auf einen Blick sichtbar, welche Kaffeearten besonders häufig gewählt wurden und wie groß deren jeweiliger Anteil an der Gesamtgruppe ist.

Visualisierung der Daten: Diagramme im Beispieldatensatz



Das Diagramm stellt dar, wie sich die Anzahl der täglich konsumierten Tassen Kaffee unter den befragten Studierenden verteilt. Die x-Achse zeigt die Anzahl der Tassen pro Tag, die y-Achse die Anzahl der Personen mit diesem Konsumverhalten.



Hinweis:
Ist man nicht an absoluten Häufigkeiten, sondern relativ interessiert, kann man sich in der Pivot-Tabelle über Rechtsklick → „Werte anzeigen als ...“ auch Prozente ausgeben lassen.

Visualisierung der Daten: Pivot-Tabellen

Bisher wurden nur einzelne Variablen betrachtet, also jeweils eine Kategorie in den Spalten (z. B. Kaffeeart).

→ Möchte man jedoch den Zusammenhang zwischen zwei Variablen untersuchen, beispielsweise wie sich das Geschlecht je nach Semesterzahl verteilt, müssen zwei Merkmale gleichzeitig einbezogen werden (hier Geschlecht, Semesterzahl)

→ Hierfür sind Kreuztabellen (Pivot-Tabellen) besonders geeignet, da sie Daten nach zwei oder mehr Kategorien gleichzeitig gruppieren und auswerten können.

Dazu muss man „Schritt 3: Pivot-Tabelle erstellen“ (Folie 21) ein wenig anpassen:

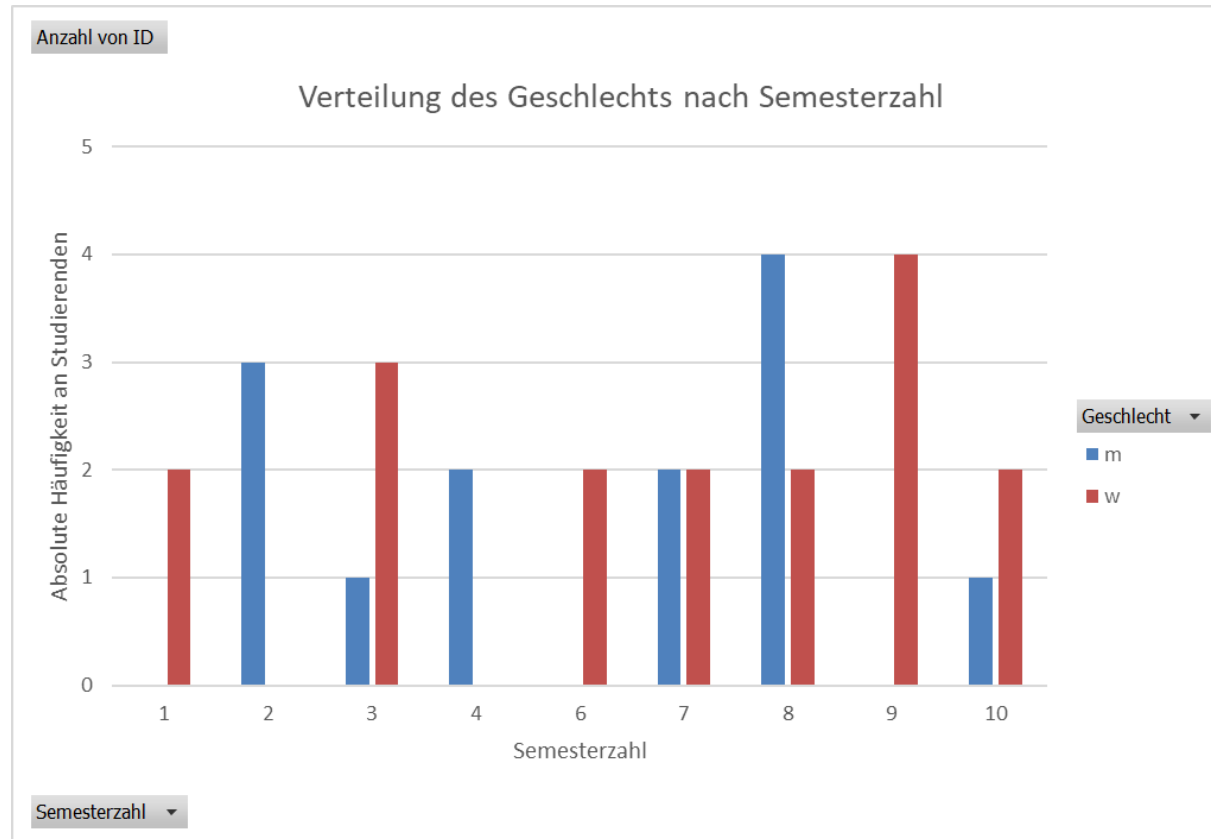
→ Ziehe im rechten Pivot-Tabellen-Feld die Variable „Semesterzahl“ in den Bereich „Zeilen“

→ Um eine Kreuzung mit einer zweiten Variable darzustellen, ziehe diese (z. B. „Geschlecht“) in den Bereich „Spalten“

→ Ziehe die Variable „ID“ in den Bereich „Werte“ (Standard: „Anzahl von ...“)

The screenshot shows the 'PivotTable-Felder' (PivotTable Fields) task pane. At the top, it says 'In den Bericht aufzunehmende Felder auswählen:' (Select fields to add to the report). Below this is a search bar labeled 'Suchen'. A list of fields is shown with checkboxes: ☒ ID, ☐ Studiengang, ☒ Semesterzahl, ☐ Tassen pro Tag, ☐ Bevorzugter Ort, ☐ Kaffeeart, ☐ Lernzeit pro Tag, and ☒ Geschlecht. Below the list is a link 'Weitere Tabellen...'. A horizontal dashed line separates this from the 'Felder zwischen den Bereichen ziehen und ablegen:' (Move fields between areas) section. This section has four boxes: 'Filter' (empty), 'Spalten' (containing 'Geschlecht'), 'Zeilen' (containing 'Semesterzahl'), and 'Werte' (containing 'Anzahl von ID'). At the bottom, there is a checkbox 'Layoutaktualisierung zurückstellen' (Reset layout update) and an 'Aktualisieren' (Refresh) button.

Visualisierung der Daten: Diagramme im Beispieldatensatz



Das Diagramm zeigt, wie das Geschlecht der Studierenden nach Semesterzahl verteilt ist. Es lässt sich zum Beispiel ablesen, dass

- nur männliche Studierende aus dem 6. Semester an der Befragung teilgenommen haben
- 1 männlicher und 3 weibliche Studierende aus dem dritten Semester vertreten sind
- ...

Grenzen von Excel

- Für große Datenmengen ungeeignet: Excel ist vor allem für kleinere bis mittlere Datenmengen gedacht. Bei sehr vielen Zeilen wird es langsam und unübersichtlich.
- Keine komplexen statistischen Verfahren ohne Zusatztools: Regressionsanalysen oder fortgeschrittene statistische Modelle sind nur mit Add-Ins oder in anderen Programmen möglich (z. B. R, SPSS).
- Zusammenarbeit in Teams nur eingeschränkt möglich: Bei gleichzeitiger Bearbeitung einer Datei kann es zu Problemen kommen. Für größere Teamprojekte sind spezialisierte Tools besser geeignet (z. B. Datenbanken oder Online-Plattformen).

Und jetzt sind Sie dran: Übung anhand des Beispieldatensatzes

1. Berechnen Sie für die Variable „Lernzeit pro Tag“ folgende Kennzahlen: Minimum, Maximum, Mittelwert und Median. Erklären Sie, was jede Kennzahl bedeutet, und was die Werte über das Lernverhalten der Studierenden aussagen.
2. Erstellen Sie ein Säulendiagramm, das zeigt, wie viele Studierende in den einzelnen Semestern sind. Fügen Sie anschließend Achsen- und einen Diagrammtitel hinzu und formulieren Sie diese möglichst präzise.
3. Untersuchen Sie den Zusammenhang zwischen der Lernzeit pro Tag und dem Geschlecht. Visualisieren Sie die Ergebnisse in einer Tabelle und im Anschluss in einem geeigneten Diagramm.



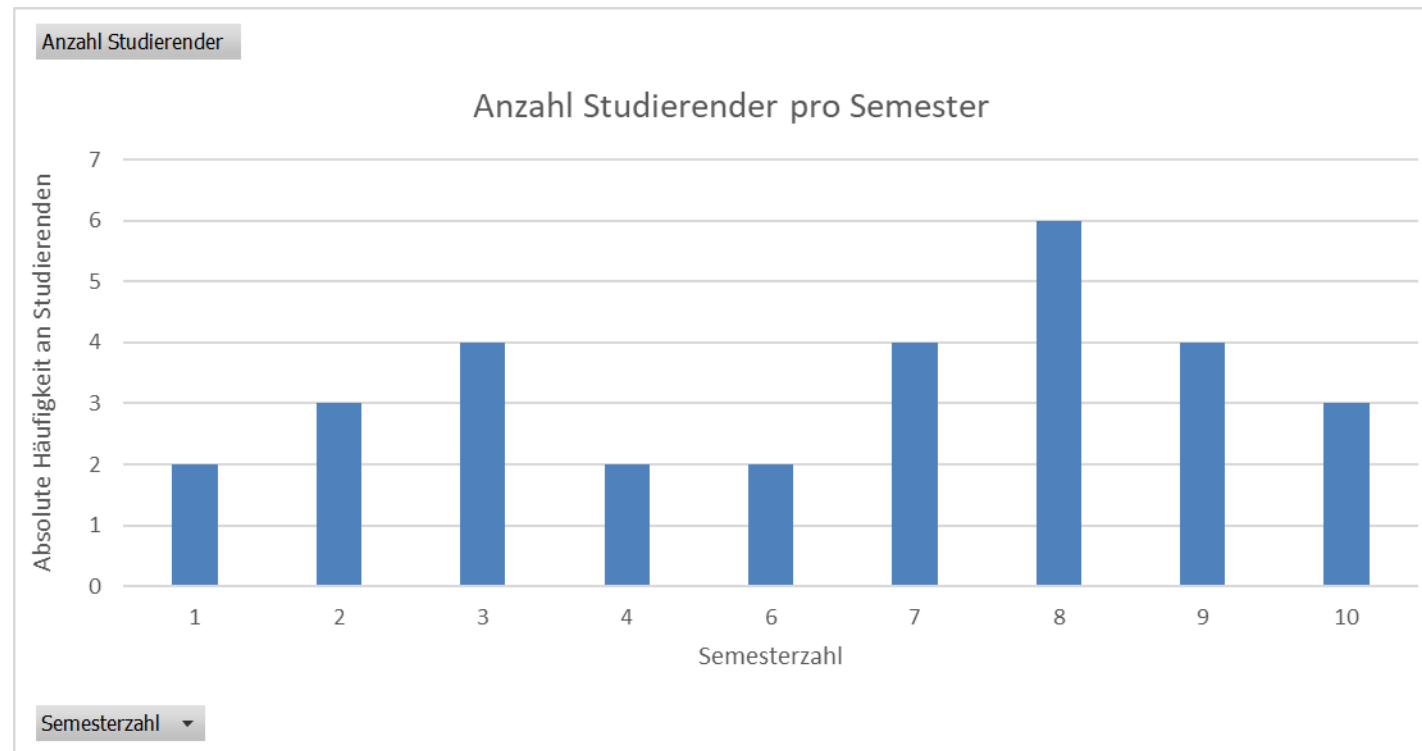
Und jetzt sind Sie dran: Lösung

1. Berechnen Sie für die Variable „Lernzeit pro Tag“ folgende Kennzahlen: Minimum, Maximum, Mittelwert und Median. Erklären Sie, was jede Kennzahl bedeutet, und was die Werte über das Lernverhalten der Studierenden aussagen.
 - Minimum = 1 (Befehl: =MIN(G2:G31)); die kürzeste Lernzeit, die ein Studierender pro Tag angegeben hat.
 - Maximum = 7 (Befehl: =MAX(G2:G31)); die längste Lernzeit pro Tag, die angegeben wurde.
 - Mittelwert = 3,07 (Befehl: =MITTELWERT(G2:G31)); der Durchschnittswert (gerundet auf zwei Dezimalstellen) aller Lernzeiten, also die Summe aller Lernzeiten geteilt durch die Anzahl der befragten Studierenden.
 - Median = 2,5 (Befehl: =MEDIAN(G2:G31)); die Lernzeit, die genau in der Mitte liegt, wenn alle Lernzeiten der Größe nach sortiert werden.

Und jetzt sind Sie dran: Lösung

- Erstellen Sie ein Säulendiagramm, das zeigt, wie viele Studierende in den einzelnen Semestern sind. Fügen Sie anschließend Achsen- und einen Diagrammtitel hinzu und formulieren Sie diese möglichst präzise.

Zum Beispiel:

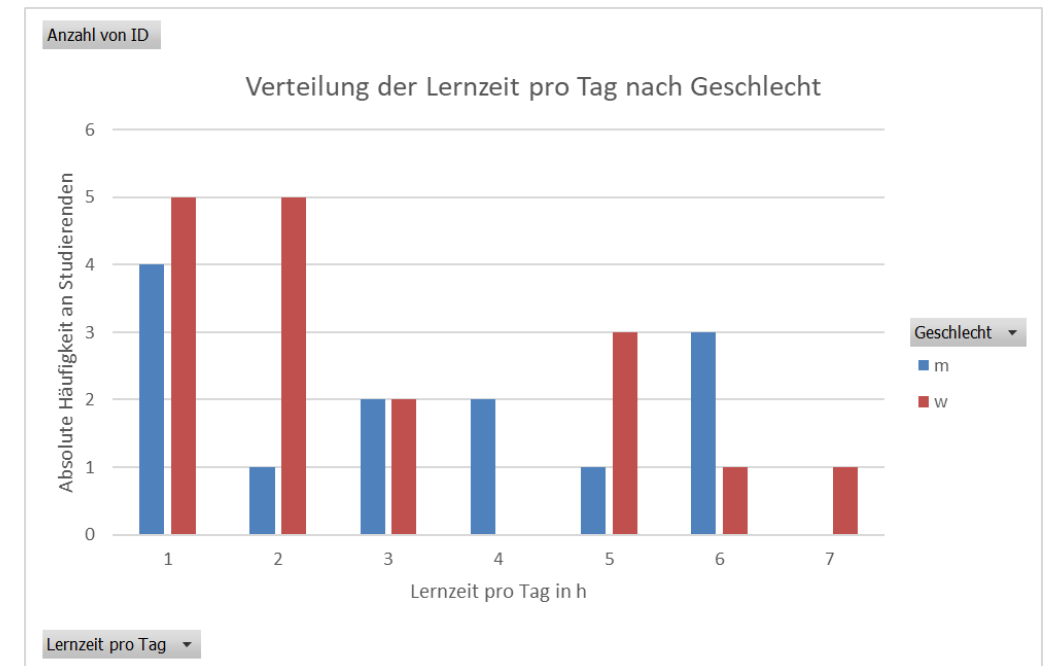


Und jetzt sind Sie dran: Lösung

3. Untersuchen Sie den Zusammenhang zwischen der Lernzeit pro Tag und dem Geschlecht. Visualisieren Sie die Ergebnisse in einer Tabelle und im Anschluss in einem geeigneten Diagramm.

Zum Beispiel:

Anzahl von ID	Geschlecht		
Lernzeit pro Tag in h	m	w	Gesamtergebnis
1	4	5	9
2	1	5	6
3	2	2	4
4	2		2
5	1	3	4
6	3	1	4
7		1	1
Gesamtergebnis	13	17	30



Literatur und hilfreiche Videos

Zur Einführung:

Schels, Ignatz (2023): *Excel 2021. Das Kompendium*. Burgthann: Markt + Technik Verlag.

Weiß, Martin (2021): *Excel Pivot-Tabellen für dummies*. Weinheim: Wiley-VCH.

Digitale Profis: Microsoft Excel Grundlagen - Komplettes Tutorial für Anfänger:innen,
<https://www.youtube.com/watch?v=SVMJXAfQRKk> (zuletzt abgerufen am 03.08.2025).

Für Fortgeschrittene:

Duller, Christine (2025): *Einführung in die Statistik mit EXCEL und SPSS. Ein anwendungsorientiertes Lehr- und Arbeitsbuch*. Heidelberg: Springer Gabler Berlin.



KATHOLISCHE UNIVERSITÄT
EICHSTÄTT-INGOLSTADT

Viel Spaß beim Forschen!