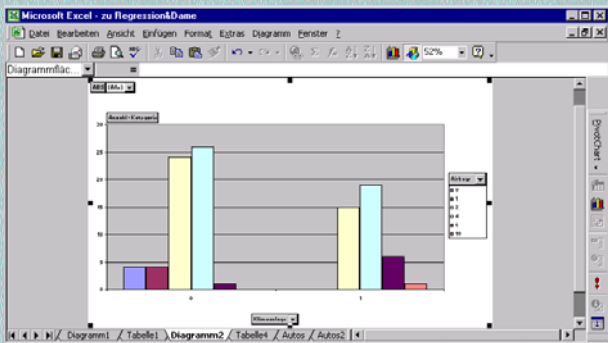
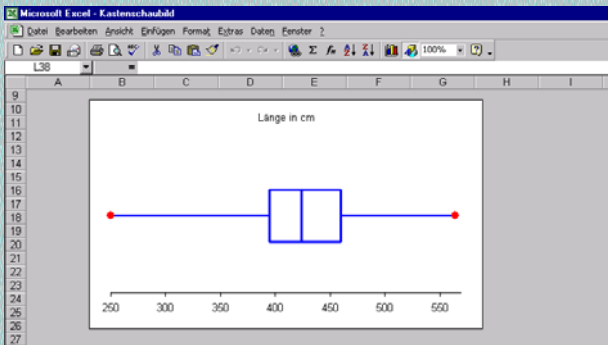


Microsoft Excel - zu Regression4Dname

PKWName	Türen	Platze	Länge	Gewicht	Hultraum	Leistung	Vmax	Bezahl.	Anfangs	ABS	Klimaanlage	PUP	Verbrauch	Preis	Kategorie
1															
2	Alfa Romeo 1.6T Spark	4	5.443	1.230	1.580	88	200	10,5	4	1	0	1	8,2	294.000	M
3	Alfa Romeo 145 1.4 16V	3	5.409	1.140	1.370	76	195	11,2	2	1	0	0	7,9	212.000	C
4	Alfa Romeo 166 2.0 TS	4	5.472	1.420	1.970	114	213	9,6	4	1	1	1	9,7	452.000	L
5	Alfa Romeo GTV 2.0 TS	2	5.428	1.370	1.970	114	211	8,4	2	1	0	1	9,2	391.000	S
6	Audi A3 1.6 Attraction	3	5.415	1.090	1.595	74	188	11	4	1	1	0	7,6	275.640	C
7	Audi A4 1.6	4	5.448	1.200	1.595	74	190	12,4	4	1	1	0	7,9	328.590	M
8	Audi A6 1.8	4	5.480	1.355	1.781	92	203	11,3	4	1	1	1	8,5	441.570	L
9	Audi A8 2.8	4	5.503	1.540	2.771	142	236	8,4	6	1	1	1	9,9	776.700	L
10	Audi TT 1.8T	3	2.404	1.205	1.781	132	228	7,4	4	1	1	1	8	474.700	S
11	Bentley Amage Red Label	4	5.539	2.520	6.750	298	249	6,3	2	1	1	1	19,2	3.503.400	L
12	BMW 316i	4	5.447	1.360	1.895	77	200	12,4	6	1	0	1	7,8	339.000	M
13	BMW 316i compact	3	5.421	1.175	1.895	77	190	11,9	4	1	0	0	7,6	291.000	C
14	BMW 520i	4	5.478	1.470	1.991	110	220	10,2	6	1	1	1	9	496.000	L
15	BMW 728i	4	5.512	1.710	2.793	142	238	8,6	10	1	1	1	10	784.000	L
16	BMW 730i	3	5.495	1.700	1.995	142	238	10,4	4	1	0	0	9,8	700.000	C



DATEN- ANALYSE

mit Excel

Marcus Hudec
Christian Neumann



Vorwort

„Wir leben im Jahrhundert der Statistik!“ lautet der Leitsatz von STAT 4 U , dem Projekt zur Vermittlung statistischer Grundkenntnisse für 10-18 jährige. Und tatsächlich werden neue Erkenntnisse immer mehr aufgrund von systematischen Auswertungen gesammelter Daten gewonnen. Wichtige Entscheidungen werden in den verschiedensten Lebensbereichen aufgrund dieser Erkenntnisse getroffen. Die Flut von statistischen Informationen ist einem ständigen Wachstum unterworfen. Es gehört zunehmen zur Allgemeinbildung, statistische Aussagen in Grundzügen nachvollziehen und einfache Berechnungen selbst durchführen zu können.

Mit zunehmenden Datenmengen und der Verfügbarkeit immer leistungsfähigerer Rechner erfolgt die Analyse von Daten mittlerweile so gut wie ausschließlich mit dem Computer.

Das Arbeitsheft Datenanalyse mit Excel zeigt anhand einer aktuellen Datensammlung, wie Informationen aus einem bestehenden Datensatz mit Hilfe der Software Microsoft Excel gewonnen werden können.

In einem begleitenden Folder für Unterrichtende wird dieses Arbeitsheft in den Kontext mit dem Projekt STAT 4 U gebracht. Außerdem werden Vorschläge zum Einsatz dieses Arbeitheftes präsentiert sowie weitere Anregungen zur Beschaffung und Bearbeitung weiterer Datensätze geboten.

Inhalt

<i>Vorwort</i>	<i>i</i>
<i>Inhalt</i>	<i>ii</i>
1 <i>Einleitung</i>	1
2 <i>Die Datenmatrix - 100 PKWs</i>	2
3 <i>Grundlegende Schritte</i>	4
3.1 <i>Sortieren</i>	4
3.2 <i>Filtern</i>	8
4 <i>Skalenniveaus</i>	16
4.1 <i>Nominalskala</i>	16
4.2 <i>Ordinalskala</i>	17
4.3 <i>Differenzskala</i>	17
4.4 <i>Verhältnisskala</i>	18
4.5 <i>Absolutskala</i>	18
5 <i>Statistische Grafiken</i>	19
5.1 <i>Stabdiagramm</i>	19
5.2 <i>Säulendiagramm</i>	22
5.3 <i>Stängel-Blatt-Diagramm</i>	29
5.4 <i>Histogramm</i>	30
6 <i>Statistische Maßzahlen I: Lagemaße</i>	37
6.1 <i>Minimum und Maximum</i>	37
6.2 <i>Arithmetisches Mittel</i>	39
6.3 <i>Median</i>	42
6.4 <i>Arithmetisches Mittel und Median – ein Vergleich</i>	44
6.5 <i>Teilbereiche</i>	48
6.6 <i>Mittelwert in Gruppen \Leftrightarrow Gesamtmittelwert</i>	51
6.7 <i>Quartile</i>	53
6.8 <i>Mittelwert – Quartile</i>	57
6.9 <i>Kastenschaubild</i>	58
7 <i>Statistische Maßzahlen II: Streuung</i>	61
7.1 <i>Maßzahlen der Streuung</i>	62
7.2 <i>Spannweite</i>	62
7.3 <i>Quartilsabstand</i>	63
7.4 <i>Standardabweichung und Varianz</i>	65
7.5 <i>Varianz</i>	65
7.6 <i>Standardabweichung</i>	68
8 <i>Erweiterungen</i>	70
8.1 <i>Pivot-Tabellen und Pivot-Charts</i>	70
8.2 <i>Histogramm mit unterschiedlicher Klassenbreite</i>	78
9 <i>Lösungsteil</i>	86
<i>Anhang 1: Datenmatrix</i>	87
<i>Anhang 2: Zellbezüge</i>	95

1 Einleitung

Erhobene, gesammelte oder recherchierte Daten liegen zumeist in einer Tabellenform, der sogenannten Datenmatrix, vor. Aufgabe der (explorativen) Datenanalyse ist es, diese Daten mithilfe von statistischen Maßzahlen zu verdichten, das heißt, die in der oft relativ umfangreichen Datenmatrix verborgenen Informationen durch einige wenige Kennzahlen transparent darzustellen.

Neben der Berechnung von Kennzahlen dienen statistische Grafiken zum Erkennen und zur Darstellung von auffälligen Mustern sowie Zusammenhängen in den Daten.

Vom einfachen Mittelwert bis zur computerintensiven Berechnung komplexer multivariater statistischer Verfahren spannt sich der Bogen der Datenanalyse.

In diesem Arbeitsheft soll anhand verschiedener Beispiele gezeigt werden, wie ein gegebener Datensatz durchleuchtet und dargestellt werden kann. Besonderer Wert wird dabei auf die Unterstützung der Arbeit durch MS-Excel gelegt, das eine wertvolle Hilfe und Arbeitserleichterung für die meisten Problemstellungen bietet.

Bietet sich nicht die Möglichkeit, die gezeigten Beispiele an einem Rechner nachzuvollziehen, so liefert dieses Skript trotzdem umfangreiche Hilfestellungen zum Auswerten eines Datensatzes.

2 Die Datenmatrix - 100 PKWs

Unser Beispieldatensatz wurde einer im Handel erhältlichen Autozeitschrift¹ entnommen und zur Durchführung einer Datenanalyse in Excel eingegeben (Abbildung 1). Er enthält Informationen zu 16 verschiedenen Merkmalen von 100 PKWs.

Ausgangspunkt unserer Analyse bildet also eine sogenannte Datenmatrix mit 100 Zeilen und 16 Spalten, wobei jede Zeile dieser Matrix die Daten eines Merkmalsträgers (in unserem Beispiel eines PKW-Modells) und jede Spalte die Daten eines Merkmals (einer Variable, zum Beispiel Länge in cm), welche für jeden Merkmalsträger (Auto) erhoben wurden, enthält.

In Tabelle 1 sind alle 16 Merkmale angeführt.

PKW-Name	Bezeichnung des Modells
Türen	Anzahl der Türen
Plätze	Anzahl der Plätze
Länge	in cm
Gewicht	in kg
Hubraum	in ccm
Leistung	in kW
Vmax	Höchstgeschwindigkeit in km/h
Beschl.	von 0 auf 100 km/h in Sekunden
Airbags	Anzahl der Airbags
ABS	1= ja, 0 = nein
Kilmaanlage	1= ja, 0 = nein
PLIP	elektr. Türöffner 1 = ja, 0 = nein
Verbrauch	in l/100 km
Preis	in ATS
Kategorie	K: Klein, C: Compact, M: Mittelklasse, L: Limousine, S: Sportwagen

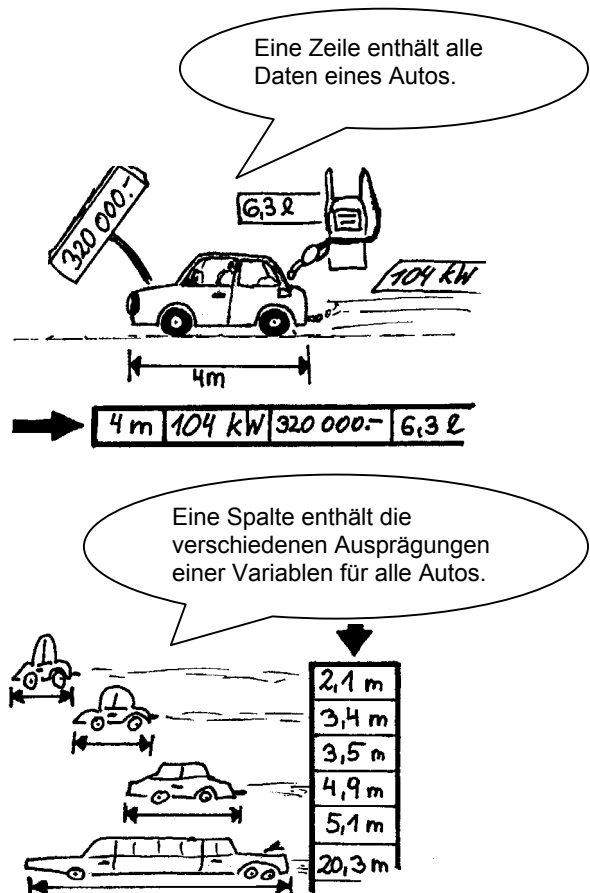


Tabelle 1

Hinweis: Beachte, dass die verschiedenen Merkmale unterschiedliche Eigenschaften haben! So dient der PKW-Name lediglich der Identifikation der Merkmalsträger. Das Merkmal ABS etwa hat nur zwei mögliche Ausprägungen (1 und 0), während die Länge der Autos viele verschiedene Werte annimmt. Eine systematische Diskussion der Eigenschaften der unterschiedlichen Merkmale findet sich in Kapitel 4: Skalenniveaus.

¹ Autorevue: Autokatalog 2000

Microsoft Excel - Datenanalyse mit Excel																	
Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster ?																	
Y80 =																	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2		PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
3		Alfa Romeo 16T Spark	4	5	443	1230	1598	88	200	10,5	4	1	0	1	8,2	294000	M
4		Alfa Romeo 145 1,4 16V	3	5	409	1140	1370	76	185	11,2	2	1	0	0	7,9	212000	C
5		Alfa Romeo 166 2,0 TS	4	5	472	1420	1970	114	213	9,6	4	1	1	1	9,7	452000	L
6		Alfa Romeo GTV 2,0 TS	2	2	429	1370	1970	114	211	8,4	2	1	0	1	9,2	391000	S
7		Audi A3 1,6 Attraction	3	5	415	1090	1595	74	188	11	4	1	1	0	7,6	275640	C
8		Audi A4 1,6	4	5	448	1200	1595	74	190	12,4	4	1	1	0	7,9	328580	M
9		Audi A6 1,8	4	5	480	1355	1781	92	203	11,3	4	1	1	1	8,5	441570	L
10		Audi A8 2,8	4	5	503	1540	2771	142	236	8,4	6	1	1	1	9,9	775700	L
11		Audi TT 1,8T	3	2	404	1205	1781	132	228	7,4	4	1	1	1	8	474700	S
12		Bentley Arnage Red Label	4	5	539	2520	6750	298	249	6,3	2	1	1	1	19,2	3583400	L
13		BMW 316i	4	5	447	1360	1895	77	200	12,4	6	1	0	1	7,8	339000	M
14		BMW 316i compact	3	5	421	1175	1895	77	190	11,9	4	1	0	0	7,6	291000	C
15		BMW 520i	4	5	478	1470	1991	110	220	10,2	6	1	1	1	9	496000	L
16		BMW 728i	4	5	512	1710	2793	142	238	8,6	10	1	1	1	10	784000	L
17		BMW Z3 1,8	2	2	405	1220	1895	87	196	10,4	4	1	0	0	7,8	389000	S
18		BMW Z8	2	2	440	1585	4941	294	250	4,7	4	1	1	1	14,5	1990000	S
19		Chevrolet Corvette	2	2	456	1538	5666	253	274	5,7	2	1	1	1	13,1	997700	S
20		Chrysler Neon 2,0 LE	4	5	439	1245	1996	98	200	10,8	2	1	1	1	7,9	227000	M
21		Chrysler Stratus 2,0 LE	2	5	475	1361	1996	98	205	10,9	2	1	1	1	8,2	328650	L
22		Citroen Saxo 1,1i SX	3	5	372	805	1124	40	162	12,8	2	0	0	1	6,5	147800	K
23		Citroen Xantia 1,8i 16V SX	5	5	452	1290	1761	81	194	10,8	4	1	1	1	8,7	288150	M
24		Citroen XM 2,0i	5	5	471	1462	1998	108	228	8,7	4	1	1	1	10,4	459000	L
25		Citroen Xsara i SX	3	5	417	1020	1360	55	175	12	4	1	0	1	6,9	207480	C
26		Daewoo Lanos 1,4 SE	3	5	407	1015	1349	55	166	15	2	0	1	0	7,9	164900	K
27		Daewoo Nubira 1,6 SX	4	5	450	1178	1598	66	185	11	2	1	1	0	7,8	219900	C
28		Daihatsu Cuore GL	3	4	341	724	989	41	140	12,8	2	0	0	0	5	104900	K
29		Ferrari 456 M GT	2	2	473	1690	5474	325	300	5,2	2	1	1	1	19,8	3175000	S
30		Fiat Barchetta	2	2	392	1085	1747	96	200	8,9	0	0	0	0	8,5	307000	S
31		Fiat Bravo 80 16V SX	3	5	403	1010	1242	60	173	12,5	4	1	0	0	6,8	193000	C
32		Fiat Marea 100 16V SX	4	5	439	1140	1581	76	187	10,7	4	1	0	0	8,3	228000	M
33		Fiat Punto 1,2 ELX	3	5	380	860	1242	44	155	14,3	4	1	0	1	5,7	157900	K
34		Fiat Seicento Young	3	5	332	730	899	29	140	18	2	0	0	0	6,1	99900	K
35		Ford Escort 1,6 Champion	5	5	414	1190	1597	66	177	12,2	2	0	0	1	8,2	189500	C
36		Ford Fiesta 1,3 Flair	3	5	383	912	1299	37	143	19,5	4	0	0	0	6,9	152900	K
37		Ford KA 1,3 50	3	4	362	946	1299	37	147	17,7	2	0	0	0	5,9	129500	K
38		Ford Mondeo 1,6i	4	5	456	1295	1597	70	185	12,7	4	1	1	1	7,5	272000	M
39		Ford Puma 1,4	2	4	398	1050	1388	66	180	11,9	2	0	0	0	7,2	222500	S
40		Honda Accord 1,6i LS	4	5	460	1270	1590	85	190	12,2	4	1	0	1	8,2	279900	M
41		Honda Civic 1,4i Fun	3	5	419	1036	1396	55	165	13,9	2	1	0	0	6,8	189900	C

Abbildung 1

Die vollständige Datenmatrix ist in Anhang 2 aufgelistet.

3 Grundlegende Schritte

Für viele Analysen in der Datenmatrix ist die in Abbildung 1 vorgegebene alphabetische Sortierung nach Autonamen nicht besonders gut geeignet.

Für andere Untersuchungen soll wiederum nur eine bestimmte Auswahl von Autos herangezogen werden (z.B. nur die Sportwägen).

Im Umgang mit einer Datenmatrix sind demnach zu Beginn zwei Instrumente von Bedeutung: Sortieren und Filtern von Daten.

3.1 Sortieren

Einfache Sortierung

Im PKW-Datensatz kann es zum Beispiel von Interesse sein, die Autos dem Preis nach zu sortieren. Mit einer derartigen Sortierung erkennt man auf einen Blick, in welcher Preisklasse ein Auto liegt und welche Autos ungefähr gleich viel kosten. Außerdem ist so auch das billigste und das teuerste Auto leicht zu bestimmen.

EINFACHE SORTIERUNG

Mit Excel kann eine Datenmatrix auf folgende Weise sortiert werden.

- 1) Die Daten, die sortiert werden sollen, werden markiert.
- 2) Der Menüpunkt **DATEN - SORTIEREN** wird angeklickt.

Beachte, dass wirklich alle Daten markiert sind!

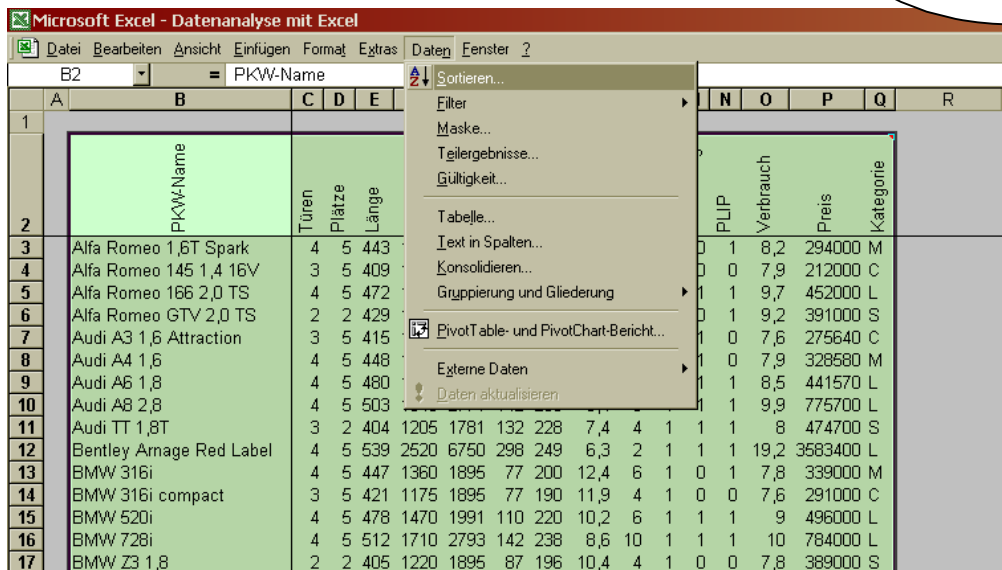


Abbildung 2

3) Es erscheint ein Fenster, in dem man bestimmen kann, wonach die markierte Datenmatrix sortiert werden soll.

Dabei sind bis zu drei Sortierkriterien möglich. Als Sortierreihenfolge kann Aufsteigend (dem Alphabet bzw. dem Zahlenwert nach) oder Absteigend gewählt werden.



Abbildung 3

4) Wir möchten die Liste zunächst nur nach dem Merkmal **PREIS** ordnen. Im ersten Eingabefeld **SORTIEREN NACH** wählen wir also **PREIS**.



Abbildung 4

5) Nachdem die gewünschte Sortierreihenfolge festgelegt ist, genügt ein Klick auf **OK**, worauf die Daten sortiert werden (Abbildung 5).

	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis
3	Fiat Seicento Young	3	5	332	730	899	29	140	18	2	0	0	0	6,1	99900 K
4	Subaru Alto L	3	5	350	730	993	40	150	15,5	0	0	0	0	5,7	99900 K
5	Lada Samara 1,5i	5	5	401	945	1499	57	155	13	0	0	0	0	7,7	99990 C
6	Daihatsu Cuore GL	3	4	341	724	989	41	140	12,8	2	0	0	0	5	104900 K
7	Skoda Felicia 1,3 Viva	5	5	386	985	1289	40	151	15,5	1	0	0	1	6,4	119900 C
8	MCC Smart pure	2	2	250	720	599	33	135	18,9	2	1	0	1	4,8	121500 K
9	Lada 110 1,5i	4	5	427	1020	1499	52	167	13	0	0	0	0	6,6	125900 M
10	Suzuki Swift 1,0 GLS	3	5	375	745	993	39	145	15,9	2	0	1	1	5,5	125900 K
11	Peugeot 106 1,0 ZEN	3	5	368	795	954	37	150	19,4	2	0	0	0	6,2	127500 K
12	Ford KA 1,3 50	3	4	362	946	1299	37	147	17,7	2	0	0	0	5,9	129500 K
13	Seat Arosa 1,0i Stella	3	4	354	864	999	37	151	17,6	2	0	0	0	5,6	132700 K
14	VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590 K
15	Renault Twingo	3	4	343	820	1149	43	151	13,7	4	1	0	0	6	134600 K
16	Skoda Fabia 1,0 Classic	5	5	396	1064	997	37	162	15,4	1	0	0	0	7,1	137000 K
17	Mazda 323 P 1,3i 16V	3	5	404	1000	1324	54	164	13,3	2	0	0	0	7,4	139900 C
18	Suzuki Baleno 1,3GL	3	4	390	910	1298	63	160	13,1	2	0	0	0	6,4	145900 C
19	Citroen Saxo 1,1i SX	3	5	372	805	1124	40	162	12,8	2	0	0	1	6,5	147800 K
20	VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900 K
21	Ford Fiesta 1,3 Flair	3	5	383	912	1299	37	143	19,5	4	0	0	0	6,9	152900 K
22	Fiat Punto 1,2 ELX	3	5	380	860	1242	44	155	14,3	4	1	0	1	5,7	157900 K
23	Mini	2	4	305	770	1273	46	148	13	1	0	0	0	6,6	159000 K
24	Lancia Y 1,2 LS	3	5	372	850	1242	44	160	13,3	1	0	0	1	6,6	159500 K
25	Opel Corsa 1,0 Joy	3	5	373	885	973	40	150	18	2	0	0	0	5,6	159500 K
26	Renault Clio 1,2 Travel	3	5	377	880	1149	43	160	15	4	1	0	1	6,2	160300 K
27	Daewoo Lanos 1,4 SE	3	5	407	1015	1349	55	166	15	2	0	1	0	7,9	164900 K
28	Nissan Micra 1,0 SMouse	3	5	372	915	998	40	150	16,4	4	1	1	1	6,1	169900 K

In dieser Liste sind alle Autos dem Preis nach sortiert, die billigsten Autos stehen ganz oben.

Abbildung 5

Mehrfache Sortierung

Nun wollen wir einen Schritt weiter gehen und die Autos in den einzelnen Kategorien (Kleinwagen, Compact, Mittelklasse, Limousinen und Sportwägen) dem Preis nach vergleichen. Zu diesem Zweck muss die Liste mehrfach sortiert werden: zunächst nach der Kategorie und danach innerhalb der Kategorien nach dem Preis.

MEHRFACHE SORTIERUNG

Mit Excel kann auch diese Aufgabe leicht gelöst werden.

- 1) Die Daten, die sortiert werden sollen, werden markiert.
- 2) Der Menüpunkt **DATEN - SORTIEREN** wird angeklickt.
- 3) Im Fenster wählen wir diesmal zunächst **SORTIEREN NACH: KATEGORIE** und **ANSCHLIEßEND NACH: PREIS** (Abbildung 6).



Abbildung 6

- 4) Die Datenmatrix aus Abbildung 7 ist nun unseren Vorstellungen entsprechend sortiert. Zunächst sind alle Compact-Autos aufgelistet und dem Preis nach sortiert. Daran anschließend folgen Kleinwägen, Limousinen, Mittelklasse-Autos (Markierung in Abbildung 7) und schließlich Sportwägen. Mithilfe der gewählten Sortierung kann ein erster Überblick über die Preise innerhalb der Autokategorien gewonnen werden. So ist etwa der niedrigste oder höchste Wert eines Merkmals leicht abzulesen („Welches Auto einer Kategorie kostet am wenigsten?“). Aus Abbildung 7 erhält man zum Beispiel folgende Informationen auf einen Blick:
 - Der Lada 110 1,5i ist mit einem Preis von 125.900,- das billigste Auto der Mittelklasse.
 - Das teuerste Auto der Mittelklasse kostet 339.000,-. Die Preise der Autos der Mittelklasse bewegen sich also in einem recht großen Bereich von 125.900,- bis 339.000,-.
 - Die meisten Autos der Mittelklasse liegen preislich im Bereich von 200.000,- - 300.000,-.

	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klima	PLUP	Verbrauch	Preis	Kategorie
45	Seat Cordoba 1,6i Signo	2	5	414	1018	1598	55	170	12,7	4	1	0	1	7,2	206800	K
46	Mercedes A 140	5	5	358	1095	1397	60	170	12,9	2	1	0	0	7,1	254146	K
47	Lada 110 1,5i	4	5	427	1020	1498	55	170	12,7	0	0	0	0	6,6	125900	M
48	Skoda Octavia 1,6 LX	5	5	451	1185	1498	55	170	12,7	0	0	0	0	7,6	182000	M
49	VW Bora 1,4	4	5	438	1185	1498	55	170	12,7	0	0	0	0	6,5	223750	M
50	Chrysler Neon 2,0 LE	4	5	438	1185	1498	55	170	12,7	0	0	0	0	7,9	227000	M
51	Fiat Marea 100 16V SX	4	5	438	1185	1498	55	170	12,7	0	0	0	0	7,3	228000	M
52	Seat Toledo 1,6 Stella	4	5	438	1185	1498	55	170	12,7	0	0	0	0	7,7	230900	M
53	Nissan Primera 1,6	4	5	438	1185	1498	55	170	12,7	0	0	0	0	7,2	239900	M
54	Mazda 626 1,8i	4	5	459	1185	1498	55	170	12,7	0	0	0	0	7,6	249900	M
55	Opel Vectra 1,6	4	5	450	1190	1500	55	170	12,7	0	0	0	0	7,4	249900	M
56	Peugeot 406 1,8 SR	2	5	460	1350	1761	66	181	14,8	4	1	1	1	7,4	264900	M
57	Ford Mondeo 1,6i	4	5	456	1295	1597	70	185	12,7	4	1	1	1	7,5	272000	M
58	VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
59	Honda Accord 1,6i LS	4	5	460	1270	1590	85	190	12,2	4	1	0	1	8,2	279900	M
60	Renault Laguna 1,6 16V	5	5	451	1370	1598	79	195	11,5	4	1	1	1	7,5	285000	M
61	Citroen Xantia 1,8i 16V SX	5	5	452	1290	1761	81	194	10,8	4	1	1	1	8,7	288150	M
62	Volvo S40 1,6	4	5	448	1220	1587	80	190	12	4	1	0	0	8	288966	M
63	Alfa Romeo 1,6T Spark	4	5	443	1230	1598	88	200	10,5	4	1	0	1	8,2	294000	M
64	Audi A4 1,6	4	5	448	1200	1595	74	190	12,4	4	1	1	0	7,9	328580	M
65	BMW 316i	4	5	447	1360	1895	77	200	12,4	6	1	0	1	7,8	339000	M
66	Opel Tigra 1,4-16V	3	4	392	980	1389	66	190	11,5	2	1	0	0	7,3	213000	S
67	Ford Puma 1,4	2	4	398	1050	1388	66	180	11,9	2	0	0	0	7,2	222500	S
68	Mazda MX-5 1,6i 16V	2	2	398	1020	1589	81	191	9,7	2	0	0	0	8,1	295000	S
69	Honda Prelude 2,0i	2	4	455	1285	1997	98	201	9,2	2	1	0	1	9,2	299000	S
70	Fiat Barchetta	2	2	392	1085	1747	96	200	8,9	0	0	0	0	8,5	307000	S
71	BMW Z3 1,8	2	2	405	1220	1895	87	196	10,4	4	1	0	0	7,8	389000	S
72	Alfa Romeo GTV 2,0 TS	2	2	429	1370	1970	114	211	8,4	2	1	0	1	9,2	391000	S

Innerhalb der einzelnen Kategorien sind die Autos auch dem Preis nach sortiert.

Abbildung 7



1. *Sortiere die PKW-Datenmatrix nach dem Merkmal Leistung!*
 - a. *Welches Auto hat die geringste Leistung?*
 - b. *Welches Auto hat die größte Leistung?*
 - c. *Wie viele Autos haben eine Leistung von weniger als 45 kW?*

2. *Sortiere die PKWs zunächst nach dem Merkmal Türen und anschließend nach dem Merkmal Preis!*
 - a. *Welcher 5-Türer kostet am wenigsten?*
 - b. *Welches ist das teuerste Auto mit 2 Türen?*
 - c. *Wie viele viertürige Autos sind in der Datenmatrix enthalten?*

3. *Finde mithilfe der Sortierfunktion heraus, welches Auto ohne Airbags am meisten kostet!*

3.2 Filtern

Uns interessieren nur diejenigen PKWs, die kürzer als 4m sind? Es sollen alle Autos, die weniger als 8l auf 100km verbrauchen, untersucht werden? Kein Problem!

FILTERN

Mithilfe von Excel ist es relativ einfach, eine Datenmatrix nach beliebigen Kriterien zu filtern, d.h. eine interessierende Teilmenge aller Zeilen der Datenmatrix zu bilden.

- 1) Zu diesem Zweck muss eine Zelle der Datenmatrix markiert sein (z.B. B3). Dann kann der sogenannte Auto-Filter aktiviert werden. (**DATEN - FILTER - AUTOFILTER**)

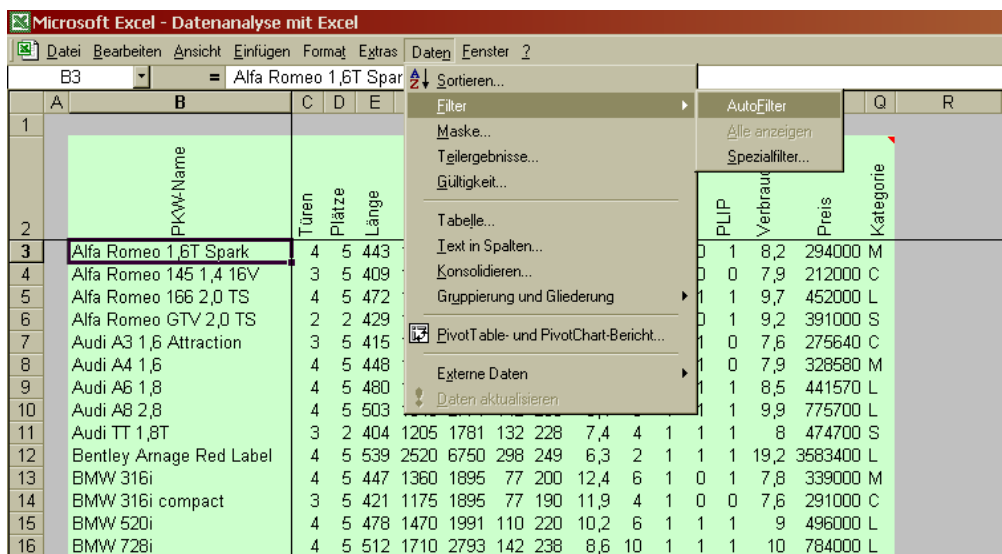


Abbildung 8

- 2) Damit steht uns ein mächtiges Werkzeug zur Verfügung, das vielfältige Optionen eröffnet.

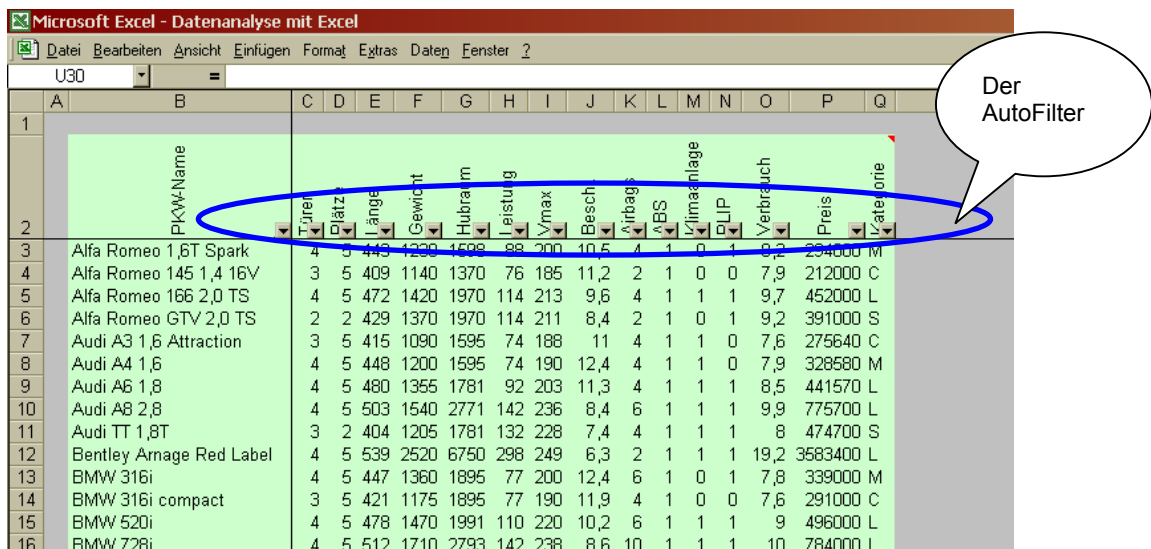


Abbildung 9

FILTERN – BEISPIEL 1: ANALYSE VON AUTOS OHNE ABS

Zu Beginn wollen wir all jene Autos herausfiltern, die kein **ABS** besitzen.

- 1) Dazu ist der Filter in der Spalte **ABS** anzuklicken (siehe Abbildung 9). Es erscheint ein Menü, das mehrere Auswahlmöglichkeiten bietet (Abbildung 10).
- 2) Wir wählen „0“, da dies bedeutet, dass das Auto kein **ABS** besitzt.

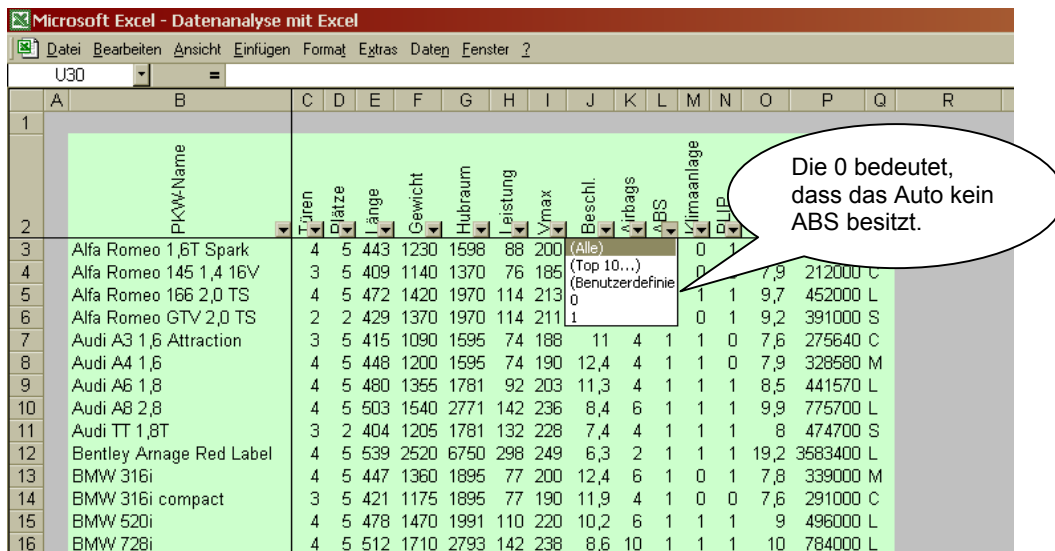


Abbildung 10

- 3) Mit dieser Auswahl werden ausschließlich jene Autos angezeigt, die kein **ABS** besitzen. Das Ergebnis ist in Abbildung 11 dargestellt. Am Bildschirm erkennt man den eingeschalteten Filter daran, dass der Pfeil im entsprechenden Drop-Down-Menü blau ist (im Gegensatz zu den schwarzen Filterpfeilen) und mehrere Zeilen ausgeblendet wurden. Die ausgeblendeten Zeilen beinhalten alle Autos mit **ABS**.

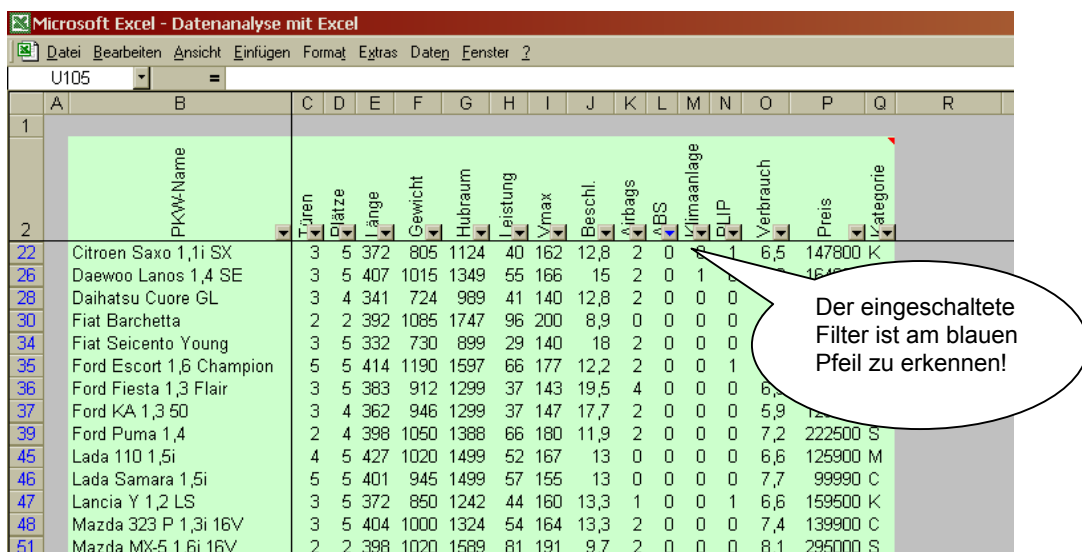


Abbildung 11

Betrachten wir nochmals die Auswahlmöglichkeiten für den Filter des Merkmals **ABS**!

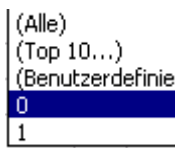


Abbildung 12

(Alle)	Mit diesem Menüpunkt kann ein eingeschalteter Filter wieder deaktiviert werden
(Top Ten)	Ist beim Filtern nach der Variablen ABS unbrauchbar; Für ein Anwendungsbeispiel siehe Filtern - Beispiel 4
(Benutzerdefiniert)	Für ein Anwendungsbeispiel siehe Filtern - Beispiel 5
0	Alle Autos mit ABS = 0 (also ohne ABS) herausfiltern und anzeigen
1	Alle Autos mit ABS = 1 (also mit ABS) herausfiltern und anzeigen

Tabelle 2

Mit dem Filter für die Variable **ABS** kann man also jene Autos herausfiltern, die mit **ABS** ausgestattet bzw. nicht ausgestattet sind. Bei anderen Variablen kann es mehrere Auswahlmöglichkeiten geben. Einige davon werden nachfolgend vorgestellt.

FILTERN – BEISPIEL 2: AUSWAHL NACH DER ANZAHL DER AIRBAGS

- 1) Ist der AutoFilter für **ABS** noch aktiviert, müssen wir diesen zunächst ausschalten (ein Klick auf den blauen Pfeil in der Spalte **ABS** und Anwählen der Option „(Alle)“).
- 2) Nun können wir uns der Variablen **AIRBAGS** zuwenden. Ein Klick auf den schwarzen Pfeil in der Spalte **AIRBAGS** und das entsprechende Filtermenü erscheint (Abbildung 13).

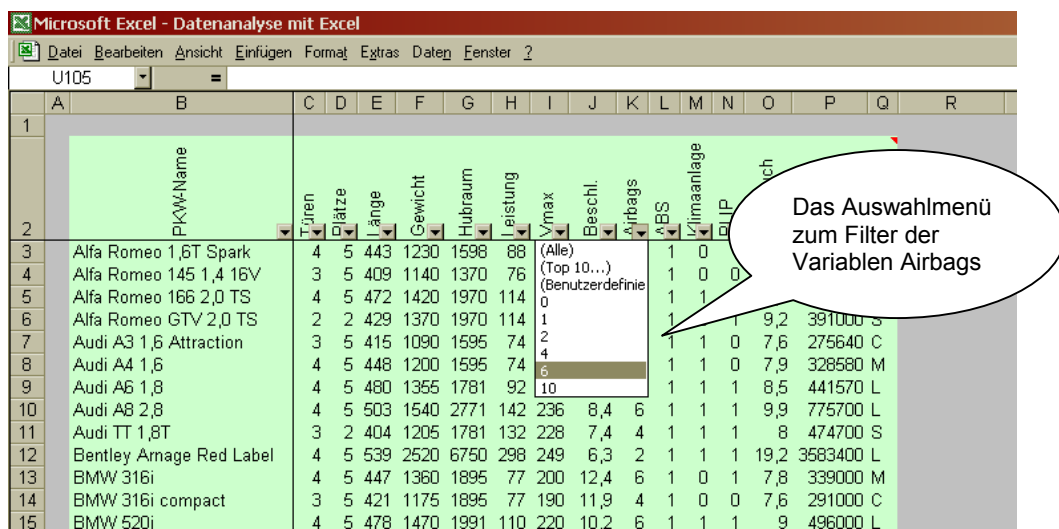


Abbildung 13

- 3) In diesem Auswahlménü gibt es mehr Optionen als in jenem von **ABS**. Während „(Alle)“, „(Top 10 ...)“, „(Benutzerdefiniert)“, „0“ und „1“ gleich bleiben, kommen „2“, „4“, „6“ und „10“ hinzu. Das liegt daran, dass das Merkmal **AIRBAGS** all diese Merkmalsausprägungen besitzt. Während bei **ABS** nur „1“ oder „0“ (ja oder nein) auszuwählen war, ist hier die Anzahl der **AIRBAGS** angegeben. Alle vorkommenden Autos besitzen entweder 0, 1, 2, 4, 6 oder 10 **AIRBAGS**.
- 4) Wir wollen nun jene Autos herausfiltern, die genau 6 **AIRBAGS** besitzen. Dazu wird im Auswahlménü die Zahl 6 angeklickt.

Microsoft Excel - Datenanalyse mit Excel

File Edit View Insert Format Tools Data Window Help

T113 =

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
		PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Preis	Leistung				
10		Audi A8 2,8	4	5	503	1540	2771	142	236	8,4	6	1						
13		BMW 316i	4	5	447	1360	1895	77	200	12,4	6	1	0					
15		BMW 520i	4	5	478	1470	1991	110	220	10,2	6	1		9	45000	L		
56		Mercedes E 200	4	5	480	1510	1998	100	209	9,3	6	1	1	1	9,3	534535	L	
57		Mercedes S 320	4	5	504	1770	3199	165	240	8,2	6	1	1	1	11,5	1106640	L	
74		Peugeot 607 2,2	4	5	488	1455	2230	116	220	9,6	6	1	1	1	7,2	455000	L	
96		Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L	
103																		

Diese Autos haben jeweils 6 Airbags!

Abbildung 14

- 5) In der Tabelle bleiben nur jene Autos eingeblendet, die genau 6 **AIRBAGS** aufweisen.

FILTERN – BEISPIEL 3: MEHRERE FILTER

Es ist durchaus möglich, die Datenmatrix nach mehreren Kriterien gleichzeitig zu filtern. Geschieht dies, so werden nur noch jene Merkmalsträger angezeigt, deren Ausprägungen gleichzeitig alle Filterkriterien erfüllen (logische UND – Verknüpfung).

Wir wollen uns jene PKWs anzeigen lassen, die keine **AIRBAGS** und kein **ABS** besitzen.

- 1) Dazu wird der Filter **AIRBAGS**, der von Beispiel 2 noch auf „6“ steht, angewählt und auf „0“ verändert.
- 2) Im Filter **ABS** wird ebenfalls „0“ ausgewählt.
- 3) Nun sind die Filterpfeile der Variablen **AIRBAGS** und **ABS** blau. In Abbildung 15 ist das Ergebnis dargestellt: diese Autos besitzen weder **AIRBAGS** noch **ABS**.

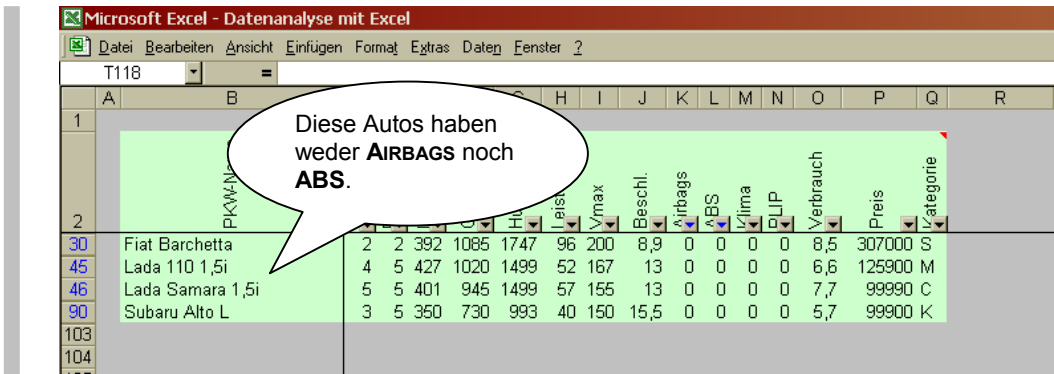


Abbildung 15

FILTERN – BEISPIEL 4: DIE 10 SPARSAMSTEN AUTOS BEIM TREIBSTOFFVERBRAUCH

Dieses Beispiel zum AutoFilter behandelt die Option „(Top 10...)“. Es sollen die zehn Autos mit dem geringsten Treibstoffverbrauch herausgefiltert werden.

1) Im Filtermenü der Variablen **VERBRAUCH** wird „(Top 10...)“ angewählt (Abbildung 16).

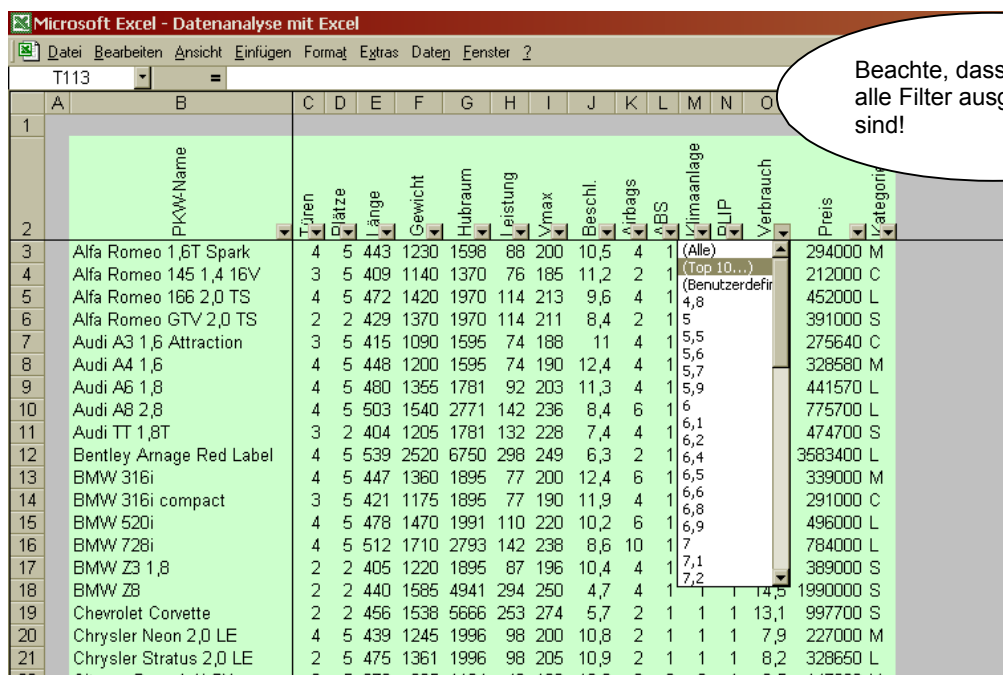


Abbildung 16

2) Es erscheint ein kleines Menü das drei Auswahlmöglichkeiten bietet:

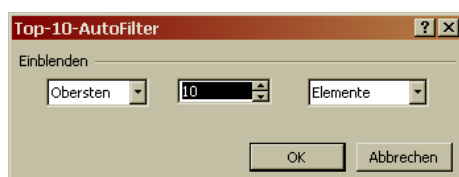


Abbildung 17

- 3) Im ersten Eingabefeld kann man bestimmen, ob man die obersten (größten) oder untersten (kleinsten) Elemente anzeigen lassen will. Da wir uns für die Autos mit dem niedrigsten Verbrauch interessieren, wählen wir „Untersten“ an.
- 4) Im zweiten Eingabefeld kann man die Anzahl der anzuzeigenden Elemente auswählen. Somit lassen sich nicht nur die „Top 10“, sondern auch die „Top 3“ oder „Top 5“ anzeigen. Wir belassen die Einstellung bei „Top 10“.
- 5) Das dritte Eingabefeld ermöglicht neben der Anzeige der obersten oder untersten Elemente auch eine Prozentangabe (etwa die Auswahl der sparsamsten 5% aller Autos). Wir belassen die Einstellung jedoch bei **ELEMENTE**.

	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klima	PJP	Verbrauch	Preis	Kategorie
28	Daihatsu Cuore GL	3	4	341	724	989	41	140	12,8	2	0	0	0	5	104900	K
33	Fiat Punto 1,2 ELX	3	5	380	860	1242	44	155	14,3	4	1	0	1	5,7	157900	K
37	Ford KA 1,3 50	3	4	362	946	1299	37	147	17,7	2	0	0	0	5,9	129500	K
53	MCC Smart pure	2	2	250	720	599	33	135	18,9	2	1	0	1	4,8	121500	K
67	Opel Corsa 1,0 Joy	3	5	373	885	973	40	150	18	2	0	0	0	5,6	159500	K
83	Seat Arosa 1,0i Stella	3	4	354	864	999	37	151	17,6	2	0	0	0	5,6	132700	K
90	Subaru Alto L	3	5	350	730	993	40	150	15,5	0	0	0	0	5,7	99900	K
92	Suzuki Swift 1,0 GLS	3	5	375	745	993	39	145	15,9	2	0	1	1	5,5	125900	K
100	VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
102	VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

Abbildung 18

- 6) Abbildung 18 zeigt jene zehn Autos mit dem geringsten Verbrauch an.

Beachte: Die Autos in Abbildung 18 sind nicht nach dem Treibstoffverbrauch sortiert! Das Auto mit dem geringsten Verbrauch etwa steht an vierter Stelle.

FILTERN – BEISPIEL 5: AUTOS VON 200.000,- BIS 300.000,-

Wenn man einen Datensatz mit Hilfe des Filters bearbeitet, werden die bisher verwendeten Optionen des Auswahlmenüs mitunter nicht allen individuellen Filterwünschen gerecht. Aus diesem Grund lässt sich mit einem Klick auf „(Benutzerdefiniert)“ ein weiteres Menü öffnen, in dem zusätzliche Filtermöglichkeiten zur Verfügung stehen.

Als Beispiel wollen wir alle Autos mit einem Kaufpreis von 200.000,- bis 300.000,- herausfiltern.

- 1) Klickt man den Filter in der Spalte **PREIS** an, erscheint das bereits bekannte Menü (Abbildung 19).

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PUP	Verbrauch	Preis	Kategorie
Alfa Romeo 1,6T Spark	4	5	443	1230	1598	88	200	10,5	4	1	0	1	(Alle)		M
Alfa Romeo 145 1,4 16V	3	5	409	1140	1370	76	185	11,2	2	1	0	0	(Top 10...)		C
Alfa Romeo 166 2,0 TS	4	5	472	1420	1970	114	213	9,6	4	1	1	1	(Benutzerdefiniert)		L
Alfa Romeo GTV 2,0 TS	2	2	429	1370	1970	114	211	8,4	2	1	0	1	99900	99900	S
Audi A3 1,6 Attraction	3	5	415	1090	1595	74	188	11	4	1	1	0	104900	104900	C
Audi A4 1,6	4	5	448	1200	1595	74	190	12,4	4	1	1	0	119900	119900	M
Audi A6 1,8	4	5	480	1355	1781	92	203	11,3	4	1	1	1	121500	121500	L
Audi A8 2,8	4	5	503	1540	2771	142	236	8,4	6	1	1	1	125900	125900	L
Audi TT 1,8T	3	2	404	1205	1781	132	228	7,4	4	1	1	1	129500	129500	L
Bentley Arnage Red Label	4	5	539	2520	6750	298	249	6,3	2	1	1	1	132700	132700	S
BMW 316i	4	5	447	1360	1895	77	200	12,4	6	1	0	1	134590	134590	L
BMW 316i compact	3	5	421	1175	1895	77	190	11,9	4	1	0	0	134600	134600	M
BMW 520i	4	5	478	1470	1991	110	220	10,2	6	1	1	1	137000	137000	C
BMW 728i	4	5	512	1710	2793	142	238	8,6	10	1	1	1	139900	139900	L
BMW Z3 1,8	2	2	405	1220	1895	87	196	10,4	4	1	0	0	145900	145900	L
BMW Z8	2	2	440	1585	4941	294	250	4,7	4	1	1	1	147800	147800	S
Chevrolet Corvette	2	2	456	1538	5666	253	274	5,7	2	1	1	1	149900	149900	S
													152900	152900	S
													174,5	1990000	S
													13,1	997700	S

Abbildung 19

- 2) Nach Klick auf „(Benutzerdefiniert)“ erscheint folgendes Fenster.

Benutzerdefinierter AutoFilter

Zeilen anzeigen:
Preis

entspricht

Und Oder

Verwenden Sie "?" für einzelne Zeichen
Verwenden Sie "*" für Zeichenfolgen

OK Abbrechen

Abbildung 20

- 3) Um das gewünschte Ergebnis zu erhalten, müssen wir das Blatt wie in Abbildung 21 ausfüllen: der Preis soll größer oder gleich 200.000,- **und gleichzeitig** kleiner oder gleich 300.000,- sein.

Benutzerdefinierter AutoFilter

Zeilen anzeigen:
Preis

ist größer oder gleich 200000

Und Oder

ist kleiner oder gleich 300000

Verwenden Sie "?" für einzelne Zeichen
Verwenden Sie "*" für Zeichenfolgen

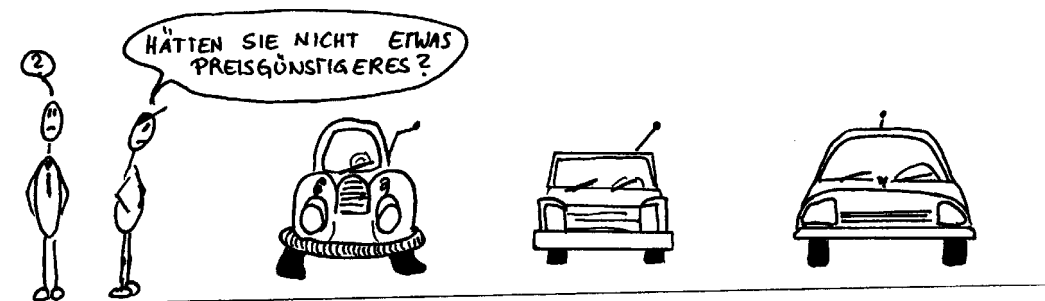
OK Abbrechen

Abbildung 21

- 4) Als Ergebnis werden in der Tabelle nur jene Autos angezeigt, die dem Kriterium aus 3) entsprechen (Abbildung 22).

	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klima	PLP	Verbrauch	Preis	Kategorie
3	Alfa Romeo 1.6T Spark	4	5	443	1230	1598	88	200	10,5	4	1	0	1	8,2	294000	M
4	Alfa Romeo 145 1,4 16V	3	5	409	1140	1370	76	185	11,2	2	1	0	0	7,9	212000	C
7	Audi A3 1,6 Attraction	3	5	415	1090	1595	74	188	11	4	1	1	0	7,6	275640	C
14	BMW 316i compact	3	5	421	1175	1895	77	190	11,9	4	1	0	0	7,6	291000	C
20	Chrysler Neon 2,0 LE	4	5	439	1245	1996	98	200	10,8	2	1	1	1	7,9	227000	M
23	Citroen Xantia 1,8i 16V SX	5	5	452	1290	1761	81	194	10,8	4	1	1	1	8,7	288150	M
25	Citroen Xsara i SX	3	5	417	1020	1360	55	175	12	4	1	0	1	6,9	207480	C
27	Daewoo Nubira 1,6 SX	4	5	450	1178	1598	66	185	11	2	1	1	0	7,8	219900	C
32	Fiat Marea 100 16V SX	4	5	439	1140	1581	76	187	10,7	4	1	0	0	8,3	228000	M
38	Ford Mondeo 1,6i	4	5	456	1295	1597	70	185	12,7	4	1	1	1	7,5	272000	M
39	Ford Puma 1,4	2	4	398	1050	1388	66	180	11,9	2	0	0	0	7,2	222500	S
40	Honda Accord 1,6i LS	4	5	460	1270	1590	85	190	12,2	4	1	0	1	8,2	279900	M
42	Honda Prelude 2,0i	2	4	455	1285	1997	98	201	9,2	2	1	0	1	9,2	299000	S
49	Mazda 626 1,8i	4	5	459	1170	1840	74	182	11,8	4	1	0	0	7,6	249900	M
51	Mazda MX-5 1,6i 16V	2	2	398	1020	1589	81	191	9,7	2	0	0	0	8,1	295000	S
60	Mercedes A 140	5	5	358	1095	1397	60	170	12,9	2	1	0	0	7,1	254146	K
65	Nissan Primera 1,6	4	5	452	1240	1597	73	181	12	4	1	0	0	7,2	239900	M
69	Opel Tigra 1,4-16V	3	4	392	980	1389	66	190	11,5	2	1	0	0	7,3	213000	S
70	Opel Vectra 1,6	4	5	450	1198	1598	55	175	15,5	4	1	0	1	7,4	249900	M
73	Peugeot 406 1,8 SP	2	5	460	1350	1761	55	181	14,6	4	1	1	1	8,4	264900	M

Abbildung 22



4. *Filtere alle Autos mit Klimaanlage aus dem Datensatz!*
5. *Finde mit Hilfe des Filters heraus, welche Mittelklassewagen länger als 450cm sind!*
6. *Welches ist der leistungsstärkste Sportwagen?*
 - a. *Löse die Aufgabe auf zwei Arten (Sortieren/Filtern)!*
 - b. *Welches Werkzeug ist für diese konkrete Fragestellung besser geeignet?*

4 Skalenniveaus

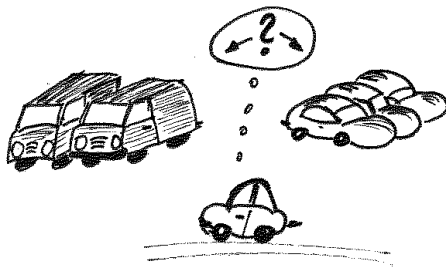
Bereits zu Beginn haben wir gesehen, dass es unterschiedliche Arten von Merkmalen gibt.

Von zentraler Bedeutung bei der Arbeit mit Daten ist es, die Art der Merkmale zu erkennen und bei Analysen zu berücksichtigen. Die Art eines Merkmales wird durch das sogenannte Skalenniveau beschrieben. Vom Skalenniveau eines Merkmals hängt es ab, welche Berechnungen sinnvollerweise angestellt werden können.

Im folgenden werden die verschiedenen Skalenniveaus anhand der Variablen des PKW-Datensatzes vorgestellt.

4.1 Nominalskala

Durch ein nominal skaliertes Merkmal können Merkmalsträger in verschiedene Klassen eingeteilt werden.



Ein Beispiel für ein nominal skaliertes Merkmal in unserem Datensatz ist das Merkmal **KATEGORIE**. Jedes Auto kann eindeutig einer Klasse zugeordnet werden.

Die meisten Rechenoperationen sind für nominal skalierte Merkmale nicht geeignet. (So ist es zum Beispiel nicht sinnvoll, einen „Mittelwert“ zwischen Limousine und Sportwagen berechnen zu wollen.)

Ein Spezialfall der nominalskalierten Merkmale sind die binären Merkmale, die überhaupt nur zwei Ausprägungen (etwa ja/nein oder ein/aus) annehmen können.

Im PKW-Datensatz finden wir noch andere nominal skalierte Variablen: **ABS**, **KLIMAAANLAGE**, **PLIP**.

Transformationsmöglichkeit: Den verschiedenen Kategorien wurden zwar aus Gründen der Anschaulichkeit verschiedene einzelne Buchstaben zugeordnet („K“ für Klein, „C“ für Compact, etc.). Genauso gut hätte man aber auch die Ziffern 1, 2, 3, 4 und 5 für die verschiedenen Kategorien verwenden können.

4.2 Ordinalskala

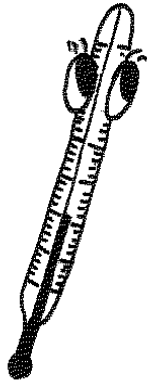
Ordinal skalierte Merkmale können bereits sinnvoll sortiert werden. Gleiche Differenzen haben aber inhaltlich keine oder verschiedene Bedeutung.

Ein Paradebeispiel für ein ordinal skaliertes Merkmal sind die Schulnoten. Hier ordnen wir ganz automatisch einem „Sehr gut“ die 1, einem „Gut“ die 2 usw. zu. Man kann auch bereits sinnvoll sortieren (1 besser als 2). Gleiche Differenzen haben aber inhaltlich noch unterschiedliche Bedeutungen (Bsp.: die Differenz zwischen 2 und 3 hat eine andere inhaltliche Bedeutung als die Differenz zwischen 4 und 5). Im PKW-Datensatz sind keine ordinalskalierten Variablen vorhanden.

4.3 Differenzenskala

Die Differenzenskala erlaubt eine sinnvolle Sortierung und, wie der Name schon sagt, eine sinnvolle Differenzenbildung. Das bedeutet, dass auch die Berechnung von Durchschnitten möglich ist.

Sie besitzt jedoch keinen natürlichen Nullpunkt, so dass die Bildung von Verhältnissen (Division) nicht sinnvoll ist.



Ein anschauliches Beispiel für die Differenzenskala stellt die Temperatur in Grad Celsius dar. Man kann einerseits feststellen, dass es um 10° wärmer ist als zu einem anderen Zeitpunkt (sinnvolle Differenzenbildung) oder dass die Durchschnittstemperatur im Juli 28° betrug (Berechnung von Durchschnitten).

Man kann aber nicht sinnvollerweise feststellen, dass 30°C „doppelt so warm“ wie 15°C sei. Besonders deutlich wird dies, wenn man zwei Temperaturen auf beiden Seiten des Nullpunkts vergleichen will (etwa -10° und $+20^\circ$). Da der Nullpunkt willkürlich und nicht am absoluten Nullpunkt festgelegt wurde, können keine sinnvollen Divisionen angestellt werden.²

Im PKW-Datensatz sind keine Merkmale vorhanden, deren Skalierung der Differenzenskala entspricht.

² Beachte, dass die Division für die Temperatur in $^\circ\text{Kelvin}$ sehr wohl sinnvoll ist. Da 0°Kelvin dem absoluten Nullpunkt entspricht, kann man richtigerweise sagen, dass 100°Kelvin doppelt so warm ist wie 50°Kelvin . Die Temperatur in $^\circ\text{Kelvin}$ ist also ein verhältnisskaliertes Merkmal.

4.4 Verhältnisskala

Die Verhältnisskala erlaubt alle Operationen der Differenzenskala. Zusätzlich sind auch relative Größenvergleiche (also Divisionen) möglich.

Beispiele für verhältnisskalierte Variablen im PKW-Datensatz sind: **LÄNGE**, **GEWICHT**, **HUBRAUM**, **LEISTUNG**, **HÖCHSTGESCHWINDIGKEIT**, **BESCHLEUNIGUNG**, **VERBRAUCH** und **PREIS**.

Durch den absoluten Nullpunkt sind Aussagen wie etwa „Auto A ist doppelt so teuer wie Auto B“ oder „Auto C ist um 20 Prozent stärker als Auto D“ sinnvoll.

4.5 Absolutskala

Die Absolutskala wird dadurch charakterisiert, dass es keine Umrechnungsmöglichkeiten gibt (etwa in unterschiedliche Maßeinheiten), da es sich bei den Merkmalsausprägungen um absolute Zahlen handelt.

Im PKW-Datensatz sind drei absolutskalierte Variablen vorhanden: **TÜREN**, **PLÄTZE** und **AIRBAGS**.



7. Welche der nominalskalierten Merkmale im PKW-Datensatz sind binär?
8. Finde weitere Beispiele für nominal skalierte Merkmale außerhalb des PKW-Datensatzes!
9. Welchem Skalenniveau können folgende Merkmale zugeordnet werden?

N ... Nominalskala

V ... Verhältnisskala

O ... Ordinalskala

A ... Absolutskala

D ... Differenzenskala

Anzahl der Unfälle auf der A1 im Jahr 1999	
Güteklassen von Eiern	
Beruf	
Körpergröße	
Platzierungen im Abfahrtslauf	
Prüfungsdauer	
Geschlecht	
Jahreszahlen	
Seitenanzahl eines Buches	

5 Statistische Grafiken

Die Darstellung der Daten in Diagrammen oder statistischen Grafiken ermöglicht zumeist eine einfache Erfassung unterschiedlicher Eigenschaften der Daten. Je nach Diagramm und Merkmal können etwa Aufschlüsse über Häufigkeitsverteilungen, Verläufe oder Anteile gewonnen werden. Für unterschiedliche Merkmale sind auch verschiedene Diagramme gebräuchlich. Eine kleine Auswahl wird im folgenden vorgestellt.

5.1 Stabdiagramm

Ein Stabdiagramm dient zur Darstellung der Aufteilung unterschiedlicher Merkmalsausprägungen eines Merkmals auf die verschiedenen Merkmalsträger.

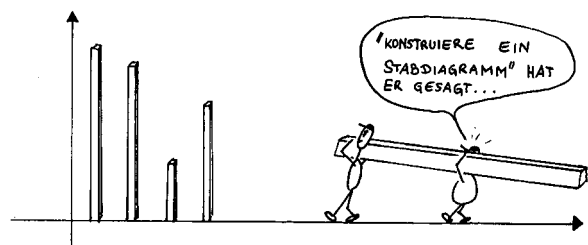
Anhand des Merkmals Kategorie soll die Konstruktion eines Stabdiagrammes demonstriert werden.

- 1) Die verschiedenen Merkmalsausprägungen des Merkmals Kategorie sind „C“, „K“, „L“, „M“ und „S“.
- 2) Bevor wir ein Stabdiagramm konstruieren können, müssen wir herausfinden, wie viele Autos der verschiedenen Ausprägungen in der Datenmatrix vorhanden sind.

Zu diesem Zweck kann man die Autos einzeln abzählen und eine Strichliste anfertigen:

Ausprägung	Strichliste	Häufigkeit
„C“	 	20
„K“	 	24
„L“	 	19
„M“	 	19
„S“	 	18

Tabelle 3



Hinweis:

Die Anzahl der Autos in jeder Kategorie kann man am besten abzählen, wenn man die Datenmatrix zuvor nach der Kategorie sortiert! Man kann sich die Anzahl der Autos in den einzelnen Kategorien aber auch durch eine Formel berechnen lassen. Die Formel wird im folgenden vorgestellt.

HÄUFIGKEIT

Für die Berechnung der Häufigkeit des Vorkommens von bestimmten Werten in einer Datenmatrix bietet Excel den Befehl

=ZÄHLENWENN(DATENBEREICH;SUCHKRITERIEN) an.

Der Datenbereich gibt jene Zellen an, in denen die Formel nach den Werten sucht, die in den Suchkriterien angegeben werden. Ausgegeben wird die Häufigkeit, mit der die gesuchten Werte im Datenbereich vorkommen. Wie dies im Beispiel der **KATEGORIEN** funktioniert, ist in Abbildung 23 zu sehen³.

Microsoft Excel - Datenanalyse mit Excel

	A	B	M	N	O	P	Q	R	S	T
1										
2		PKW-Name	Klimaanlage	PLJP	Verbrauch	Preis	Kategorie			
94		Volvo S40 1,6	0	0	8	288966	M			
95		Volvo S70 2,4	1	1	8,6	422441	L			
96		Volvo S80 2,4	1	1	8,5	455466	L			
97		VW Beetle 1,6	0	0	8,1	232050	C			
98		VW Bora 1,4	0	0	6,5	223750	M			
99		VW Golf 1,4	0	0	6,4	203990	C			
100		VW Lupo 1,0	0	0	5,6	134590	K			
101		VW Passat 1,6	0	0	8,1	276450	M			
102		VW Polo 1,0	0	0	5,7	149900	K			
103										
104										
105						C	=ZÄHLENWENN(\$Q\$3:\$Q\$102;P105)			
106						K				
107						L				
108						M				
109						S				

Abbildung 23

Microsoft Excel - Datenanalyse mit Excel

	A	B	M	N	O	P	Q	R	S	T
1										
2		PKW-Name	Klimaanlage	PLJP	Verbrauch	Preis	Kategorie			
94		Volvo S40 1,6	0	0	8	288966	M			
95		Volvo S70 2,4	1	1	8,6	422441	L			
96		Volvo S80 2,4	1	1	8,5	455466	L			
97		VW Beetle 1,6	0	0	8,1	232050	C			
98		VW Bora 1,4	0	0	6,5	223750	M			
99		VW Golf 1,4	0	0	6,4	203990	C			
100		VW Lupo 1,0	0	0	5,6	134590	K			
101		VW Passat 1,6	0	0	8,1	276450	M			
102		VW Polo 1,0	0	0	5,7	149900	K			
103										
104										
105						C	20			
106						K	24			
107						L	19			
108						M	19			
109						S	18			

Abbildung 24

³ Zur Vereinfachung werden in der Formel relative und absolute Zellbezüge verwendet. Eine Erklärung der Funktionsweise dieser Zellbezüge findet sich in Anhang 2.

- 3) Aus der oben vorgestellten Strichliste bzw. mithilfe der Excel-Funktion wird eine Häufigkeitstabelle erstellt (Abbildung 25).

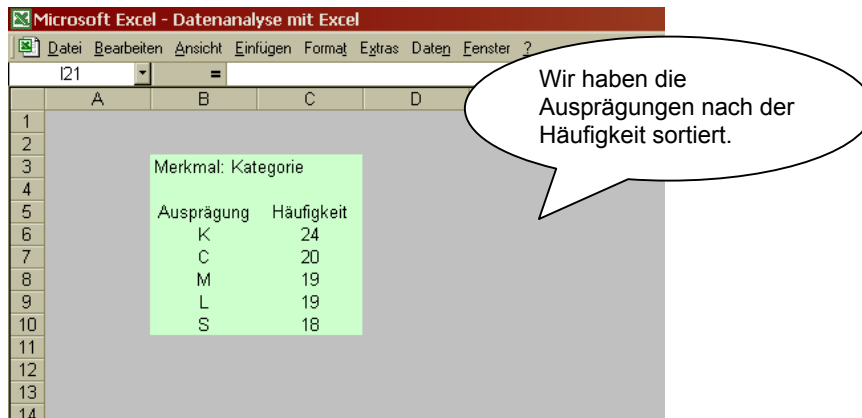


Abbildung 25

- 4) Nun können wir mit der Konstruktion eines Stabdiagrammes beginnen. Zunächst benötigen wir eine x-Achse, auf der die verschiedenen Merkmalsausprägungen („K“, „C“, „M“, „L“, „S“) aufgetragen werden. Wir beginnen mit der Ausprägung, die am öftesten vorkommt („K“).

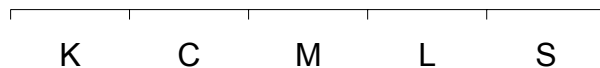


Abbildung 26

- 5) Für jede Kategorie wird ein Stab ins Diagramm gezeichnet. Die Länge des Stabes entspricht dabei genau der Anzahl der Merkmalsträger (PKW), die diese Merkmalsausprägung besitzen. Die Anzahl wird auf der y-Achse eingetragen. Ein fertiges Stabdiagramm könnte dann wie jenes in Abbildung 27 aussehen.

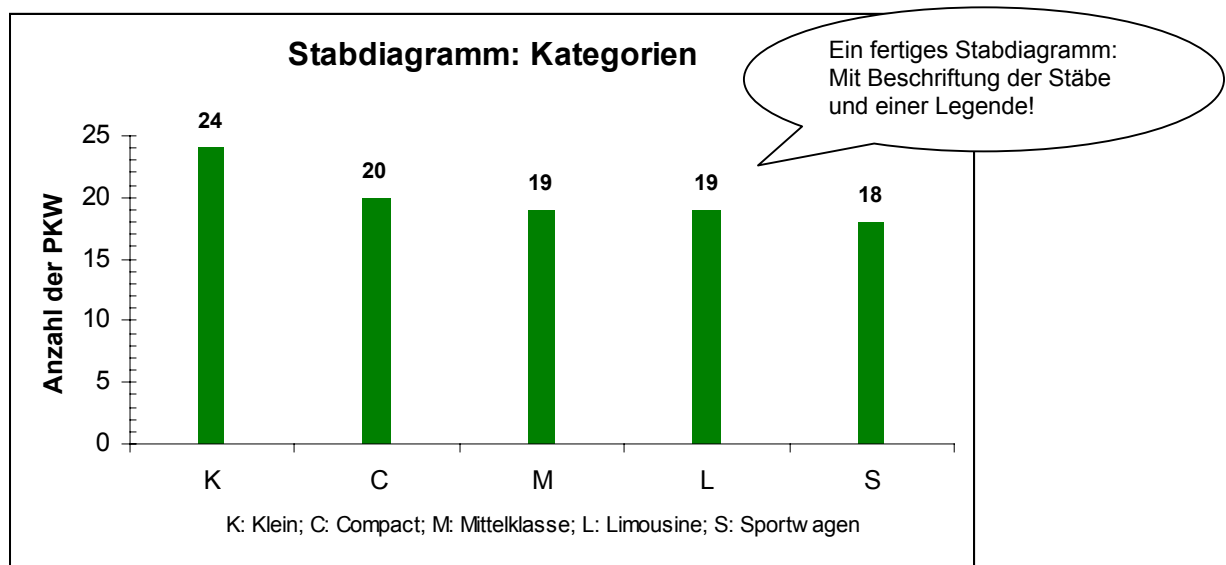


Abbildung 27

- 6) Aus dem Diagramm kann man ablesen, dass die Kleinwägen mit 24 Stück in der Datenmatrix am stärksten vertreten sind. Weiters wurden 20 Compact-Autos, jeweils 19 Mittelklassewägen und Limousinen sowie 18 Sportwägen in die Untersuchung aufgenommen.

STABDIAGRAMM

Während ein Stabdiagramm händisch sehr leicht anzufertigen und für mehrere Darstellungen nützlich ist, bietet Excel diesen Diagrammtyp nicht standardmäßig an⁴. Statt den dünnen Stäben werden ausgefüllte Säulen in einem sogenannten Säulendiagramm angeboten. Aus diesem Grund wird im folgenden Abschnitt die Konstruktion eines Säulendiagramms mit Excel ausführlicher präsentiert.

5.2 Säulendiagramm

Eine Alternative zum Stabdiagramm stellt das Säulendiagramm dar. Vom Prinzip her sind Stab- und Säulendiagramm auch völlig gleich aufgebaut, nur werden hier Säulen anstatt der Stäbe verwendet. Die Höhe der Säulen repräsentiert wieder die Anzahl der verschiedenen Merkmalsausprägungen.

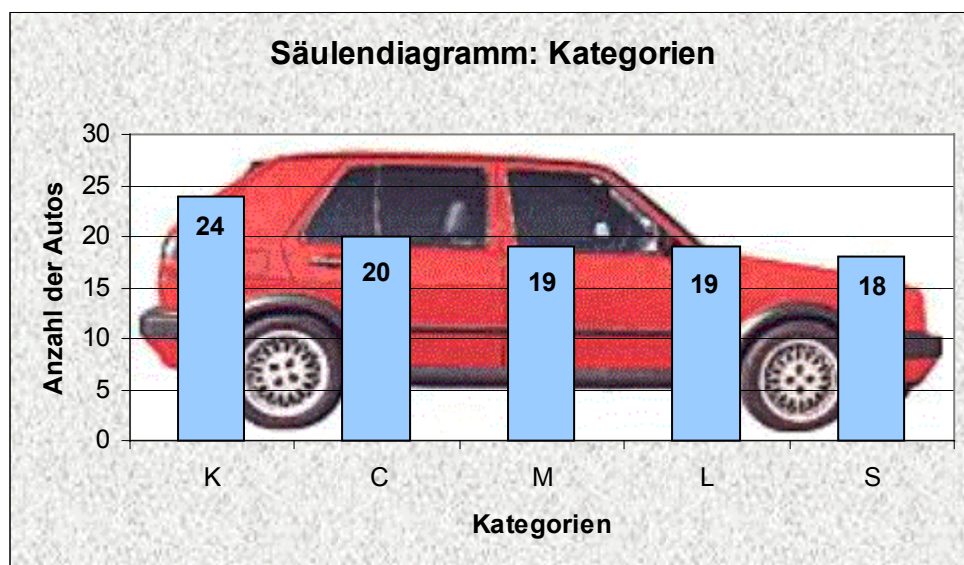


Abbildung 28

Ein Vergleich zwischen Abbildung 22 und Abbildung 28 zeigt, dass sich nur die Darstellung, nicht aber die enthaltene Information geändert hat. Letztlich ist die Auswahl zwischen Stab- und Säulendiagramm nur eine Geschmacksfrage.

⁴ Auch das „Stabdiagramm“ aus Abbildung 27 ist in Wirklichkeit ein Excel-Säulendiagramm. Die Säulen wurden einfach möglichst schmal gewählt. Grundsätzlich ist die Wahl zwischen Stab- und Säulendiagramm beliebig.

SÄULENDIAGRAMM

Wie kommt man zum Säulendiagramm aus Abbildung 28? Die Konstruktion wird im folgenden Schritt für Schritt dargestellt.

- 1) Zunächst benötigt man eine Tabelle, in der die darzustellenden Daten aufgelistet sind. Wie beim Stabdiagramm benötigen wir also wieder die Anzahl der Autos mit den Kategorien „C“, „K“, „L“, „M“ und „S“. Wie man zu der Anzahl gelangt, wurde bereits im Abschnitt 5.1. gezeigt.
- 2) Wir markieren die Daten (Abbildung 29).

The screenshot shows an Excel spreadsheet with the following data:

Ausprägung	Häufigkeit
K	24
C	20
M	19
L	19
S	18

Abbildung 29

- 3) Mit **EINFÜGEN - DIAGRAMM** wird der **DIAGRAMM-ASSISTENT** aufgerufen

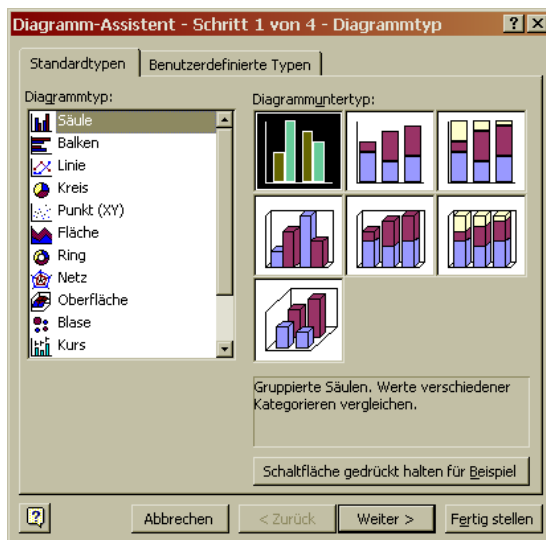


Abbildung 30

- 4) In Schritt 1 bietet der **DIAGRAMM-ASSISTENT** eine Vielzahl von Diagrammtypen und Untertypen an. Der von uns gewünschte Typ (Säule) ist bereits markiert, so dass wir sofort auf **WEITER** klicken können.

- 5) Da wir unter Punkt 2 bereits die relevanten Daten markiert haben, brauchen auch in Schritt 2 keine Änderungen vorgenommen werden. Klick auf **WEITER**.

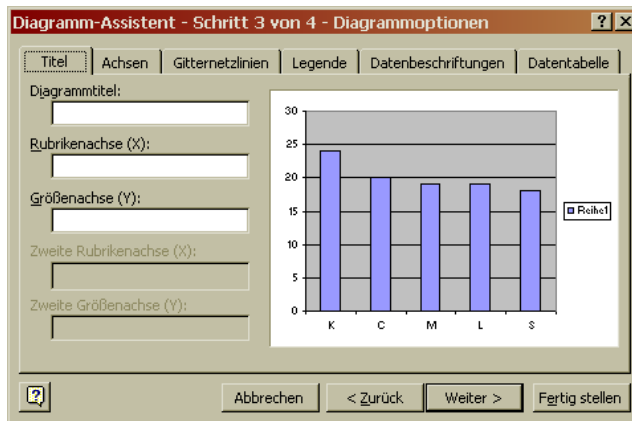


Abbildung 31

- 6) In Schritt 3 wollen wir einige der zahlreichen angebotenen Diagrammoptionen ändern. Das Register mit dem Namen **TITEL** wird entsprechend der Abbildung 32 ausgefüllt.

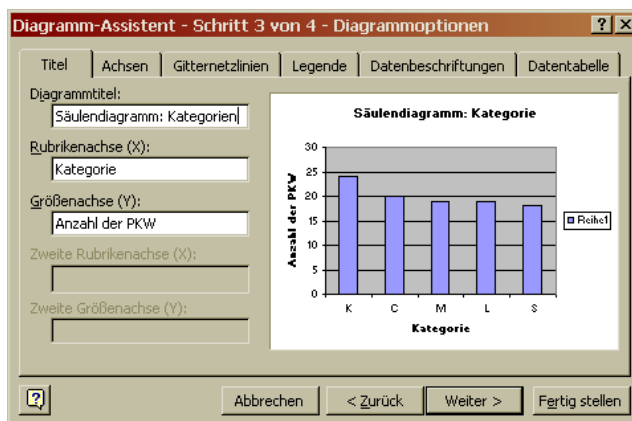


Abbildung 32

Weiters wollen wir die Legende nicht anzeigen lassen und schalten diese Option daher im Register **LEGENDE** aus (Abbildung 33).

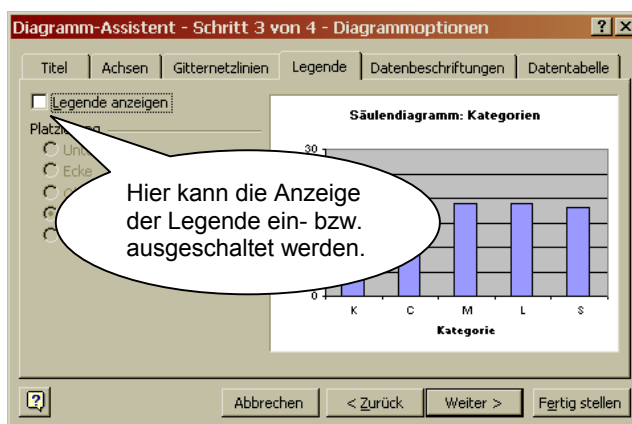


Abbildung 33

Schließlich soll auch noch die Anzahl der PKWs jeder **KATEGORIE** an der entsprechenden Säule angezeigt werden. Am Blatt **DATENBESCHRIFTUNGEN** wird **WERT ANZEIGEN** aktiviert.

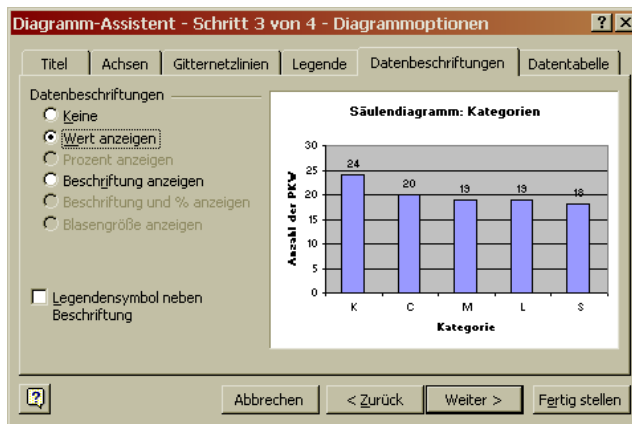


Abbildung 34

7) Da wir in Schritt 4 nichts mehr ändern wollen, klicken wir auf **FERTIG STELLEN**. Das Diagramm aus Abbildung 35 wird in unserem Arbeitsblatt eingefügt.

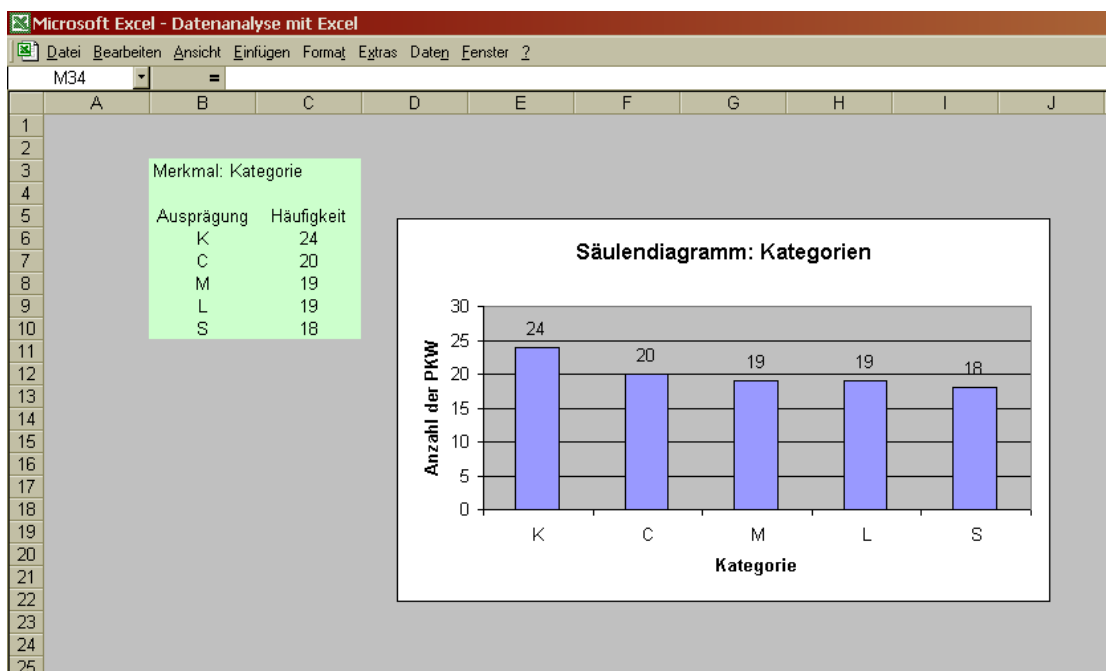


Abbildung 35

Vom Informationsgehalt her entspricht das Diagramm aus Abbildung 35 dem Diagramm aus Abbildung 28. Es soll jedoch einmal demonstriert werden, wie verschiedene grafische Aspekte verändert werden können. Diese Veränderungen sind meist nicht notwendig, können aber die Attraktivität einer Grafik erhöhen und damit zu einer besseren und anschaulicheren Informationsvermittlung beitragen.

- 1) Der Hintergrund soll verändert werden und eine ansprechendere Gestaltung erhalten. Dazu klicken wir zunächst mit der rechten Maustaste in den grauen Bereich des Diagramms und wählen im erscheinenden Kontextmenü die Option **ZEICHNUNGSFLÄCHE FORMATIEREN**.

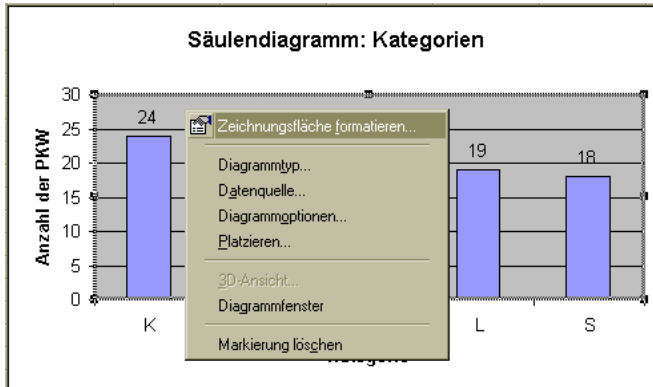


Abbildung 36

- 2) Im Fenster **ZEICHNUNGSFLÄCHE FORMATIEREN** kann man diverse Einstellungen für Rahmen- und Flächengestaltung der Zeichnungsfläche vornehmen. Wir möchten für die Fläche einen Fülleffekt und klicken aus diesem Grund auf die Schaltfläche **FÜLLEFFEKTE**.

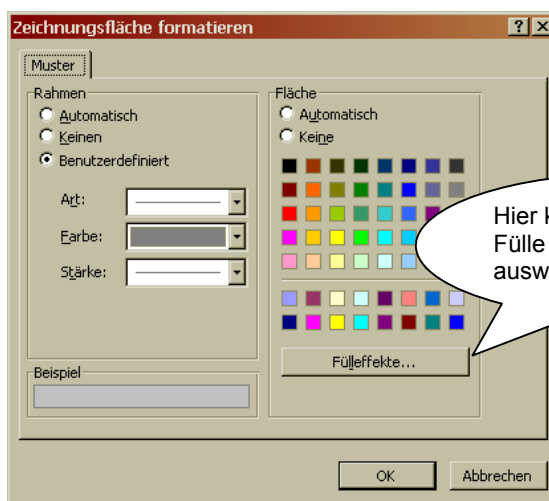


Abbildung 37

- 3) Es erscheint ein weiteres Fenster (**FÜLLEFFEKTE**), das eine große Auswahl von Effekten anbietet. Wir wählen die Option **GRAFIK AUSWÄHLEN** vom Register **GRAFIK** (Abbildung 38). Im folgenden Fenster **BILD AUSWÄHLEN** kann eine beliebige Grafik angegeben werden, die als Hintergrund im Diagramm erscheinen soll.

Ein Klick auf **OK** im Fenster **FÜLLEFFEKTE** und ein **OK** im Fenster **ZEICHNUNGSFLÄCHE FORMATIEREN** führt zum gewünschten Hintergrund in unserem Diagramm (Abbildung 40).

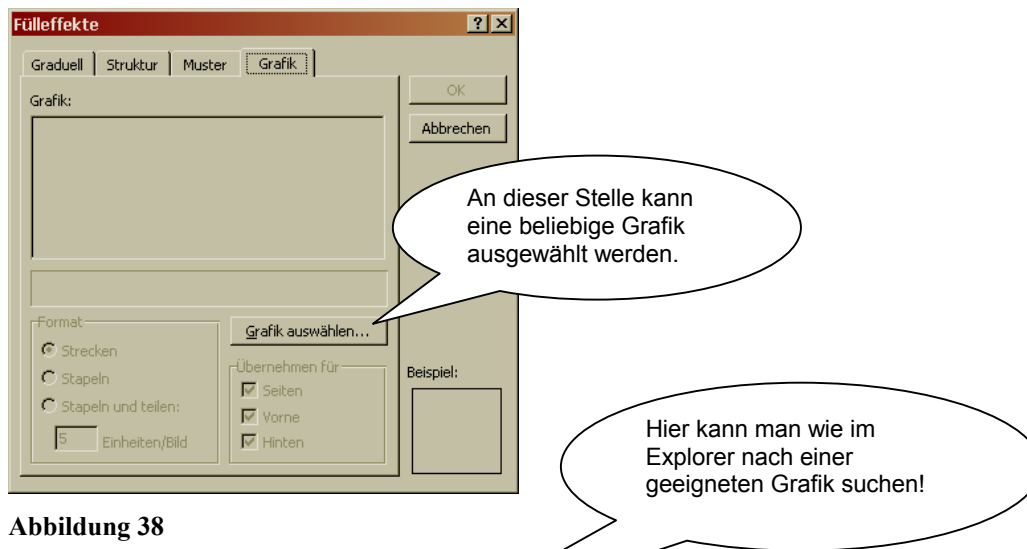


Abbildung 38

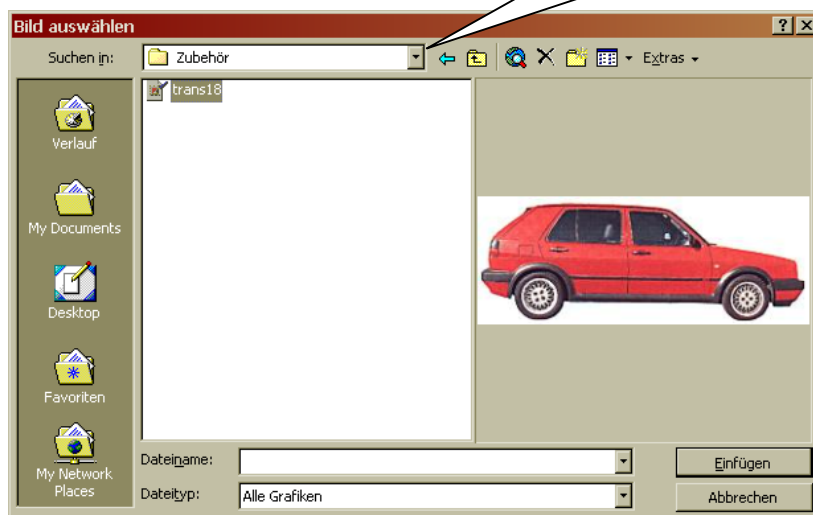


Abbildung 39

Hinweis: Damit eine Grafik angewählt werden kann, muss sie natürlich auf einem Speichermedium (Festplatte, Diskette, CD, etc.) verfügbar sein. Man kann den genauen Pfad, unter dem die Grafik gespeichert ist, suchen – siehe Sprechblase bei Abbildung 39!

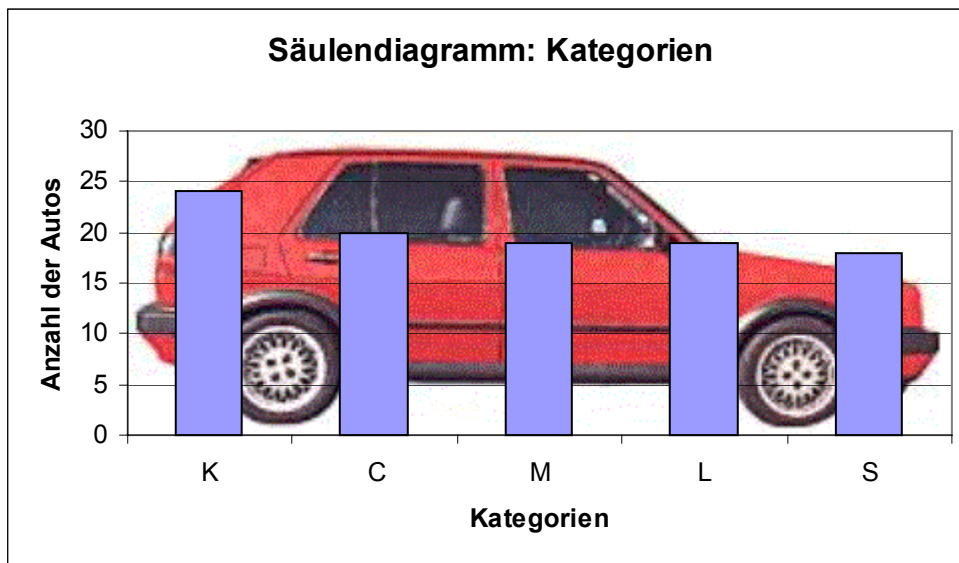


Abbildung 40

- 4) Nun soll das Diagramm noch durch einen weiteren Hintergrund, dem Zeitungspapier-Muster umrahmt werden. Durch einen Rechtsklick irgendwo am Rand des Diagrammes erscheint ein Menü, in dem **DIAGRAMMFLÄCHE FORMATIEREN** angewählt werden kann. Die weitere Vorgangsweise ist ähnlich der in Schritt 3: Im Fenster **FÜLLEFFEKTE** wird nun allerdings das Register **STRUKTUREN** angewählt. Die erste Struktur ist bereits das gewünschte Zeitungspapier-Muster. Dieses wird markiert und **OK** angewählt. Ein weiteres **OK** im Fenster **DIAGRAMMFLÄCHE FORMATIEREN** bringt eine optische Verbesserung des Diagramms mit sich.
- 5) Zuletzt wollen wir noch die Säulen selbst etwas verändern. Durch einen Rechtsklick auf eine beliebige Säule erscheint ein Menü, in dem **DATENREIHEN FORMATIEREN** ausgewählt wird.

Im folgenden Fenster können wir im Register **MUSTER** eine Farbe für die Fläche auswählen, z.B. blau.

Schließlich wollen wir noch im Register **DATENBESCHRIFTUNG** das Kästchen **WERT ANZEIGEN** aktivieren.

Ein Klick auf **OK** und das Diagramm aus Abbildung 28 ist fertig!

Beachte: Wir konnten die Anordnung der Säulen beliebig wählen. Wir haben als Kriterium die Höhe der Säule gewählt, genauso gut hätte man die Kategorien dem Alphabet nach ordnen können. Dies ist allerdings nur bei nominal skalierten Variablen erlaubt! Bei allen anderen Variablen gibt es ein logisches Kriterium, nach dem die Säulen sortiert werden müssen. Beispiel: Bei der Anzahl der Türen (oder der Sitzplätze, oder der Anzahl der Airbags) muss eben nach dieser Anzahl sortiert werden!

5.4 Histogramm

Klassenbildung

Beim Säulendiagramm wird für jede vorkommende Merkmalsausprägung eines Merkmals eine Säule gezeichnet. Daraus kann man schließen, dass dieser Diagrammtyp nicht für die Darstellung aller Merkmale geeignet sein kann, besonders wenn diese viele Ausprägungen besitzen.

Betrachtet man zum Beispiel die Variable **LEISTUNG**, so sieht man, dass beinahe jedes Auto eine unterschiedliche kW-Zahl besitzt. Bei der Untersuchung derartiger Variablen ist eine Klassenbildung oft sinnvoll. Mehrere Werte werden zu einer Klasse zusammengefaßt und gemeinsam dargestellt.

Nun könnte man argumentieren, dass man in diesem Fall eben für jede Klasse eine Säule zeichnet und die Höhe der Säule eben die Anzahl der Merkmalsträger mit Merkmalsausprägungen in einer bestimmten Klasse darstellt. Dagegen spricht zunächst, dass ein Abstand zwischen den Säulen nicht dem darzustellenden Sachverhalt entspricht, da zwischen den Klassen keine Zwischenräume vorhanden sein sollen.

Das korrekte Diagramm für die Darstellung der Anzahl der Merkmalsausprägungen in Klassen heißt Histogramm.

Bei der Klasseneinteilung in gleich breite Klassen sind die Unterschiede zwischen Histogramm und Säulendiagramm auch gering: in diesem Fall entspricht ein Histogramm einem Säulendiagramm ohne Abstand zwischen den Säulen. Die grundlegenden Unterschiede offenbaren sich erst bei der Wahl unterschiedlicher Klassenbreiten. Das entsprechende Histogramm wird in Kapitel 8 Erweiterungen gezeigt.

Wie soll man Klassen bilden?

Für die Bildung von Klassen gibt es keine universelle Patentlösung. Am ehesten läßt sich argumentieren, dass die Klassenbildung nach sachlogischen Argumenten erfolgen soll. Das heißt, dass die Klassen sinnvoll mit der gewünschten Untersuchung zusammenhängen sollen. Weiters müssen die Klassen nahtlos aneinander anschließen, dürfen einander aber nicht überlappen (Beispiel: Eine Einteilung in die Klassen 0-60kW und 40-80kW ist nicht zulässig, da ein Auto mit 50kW nicht eindeutig zugeordnet werden könnte). Schließlich ist es sehr empfehlenswert, alle Klassen gleich breit zu wählen⁵ (Beispiel: jeweils 40kW Bandbreite).

⁵ Unterschiedliche Klassenbreiten stellen sowohl in der Darstellung als auch in der Anschaulichkeit ein Problem dar und sollten daher, wenn möglich, vermieden werden. Wie man damit umgehen kann, ist in Abschnitt 8.2. Histogramm mit unterschiedlicher Klassenbreite festgehalten.

Darstellung in einem Histogramm

1) In unserem Beispiel wollen wir überprüfen, ob sich im Datensatz eher schwächer oder stärker motorisierte Autos befinden. Hier bietet sich eine Einteilung z.B. in 40er-Schritten (0kW, 40kW], (40kW, 80kW], (80kW, 120kW] an, wobei wir Autos, die genau an der Grenze liegen (40kW, 80kW, 120kW,...) jeweils der niedrigeren Klasse zuordnen wollen. Außerdem müssen wir überlegen, wie viele Klassen benötigt werden. Da das Maximum der Leistung 325kW beträgt, ist die Klasseneinteilung bis zu [320kW, 359kW) ausreichend.

2) Nun muss die Anzahl der Autos in den von uns definierten Klassen bestimmt werden. Dazu kann man wieder eine Strichliste anfertigen.

Leistung in kW	Strichliste	Anzahl
(0-40]		15
(40-80]		46
(80-120]		25
(120-160]		4
(160-200]		3
(200-240]		3
(240-280]		1
(280-320]		2
(320-360]		1

Tabelle 4

Zur Berechnung der Klassenhäufigkeiten bietet Excel folgende Funktion an.

HÄUFIGKEIT IN KLASSEN

Für die Berechnung der Anzahl der Merkmalsträger mit einer Merkmalsausprägung in einem bestimmten Bereich ist der Befehl

= **HÄUFIGKEIT (DATENBEREICH;KLASSEN)** geeignet.

Der Datenbereich gibt dabei jene Zellen an, in denen nach den Werten gesucht wird. In unserem Beispiel ist dies die Spalte Leistung (H3:H102)

Die Klassen stellen jeweils die oberen Klassengrenzen dar. Die Eingabe „40“ bedeutet also, dass alle Elemente des Datenbereichs gezählt werden, deren Werte im Intervall von 0 bis 40 liegen.

1) Zunächst werden alle oberen Klassengrenzen der gewünschten Klassen eingegeben (Abbildung 42).

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.
VW Passat 1,6	4	5	468	1250	1595	74	192	
VW Polo 1,0	3	5	374	982	999	37	151	

Leistung in kW
40
80
120
160
200
240
280
320
360

Diese Zahlen stellen jeweils die Obergrenzen unserer gewünschten Klassen dar! Wir wollen also die Anzahl der Autos mit 0-40kW, 40-80kW, 80-120 kW usw. Autos mit genau 40kW, 80kW, 120kW, ... werden jeweils der unteren Klasse zugeordnet.

Abbildung 42

2) Danach werden die Felder neben der Eingabe markiert (Abbildung 43).

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLJP	Verbrauch
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7

Leistung in kW
40
80
120
160
200
240
280
320
360

Abbildung 43

3) Während alle Felder markiert bleiben (!), wird die Formel eingegeben - allerdings ohne Return zu drücken! (Abbildung 44). Beachte, dass in der einen Formel sowohl der gesamte Datenbereich (H3:H102) als auch der gesamte Klassenbereich (H106:H114) angegeben wird! Eine solche Formel kann auch als Matrixformel eingegeben werden. Die Besonderheit einer Matrixformel besteht darin, dass sie mehrere Ergebnisse gleichzeitig berechnet und ausgibt. In unserem Fall werden die Häufigkeiten aller Klassen gleichzeitig berechnet.

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

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7

Below the table, a chart titled 'Leistung in kW' is shown. The y-axis represents power in kW with values: 40, 80, 120, 160, 200, 240, 280, 320, 360. The x-axis represents frequency. A data series is plotted with the formula `=häufigkeit(H3:H102;H106:H114)`.

Abbildung 44

- 4) Damit Excel erkennt, dass wir eine Matrixformel eingegeben haben, drücken wir nicht Return, sondern gleichzeitig STRG-UMSCHALT-RETURN. Die Formel wird in geschwungene Klammern gesetzt und für den gesamten markierten Bereich berechnet (Abbildung 45).

The screenshot shows the same Excel spreadsheet as in Abbildung 44, but now the frequency distribution is completed. A speech bubble points to the data:

15 Autos haben 0-40kW, 46 Autos 40-80 kW, 25 Autos 80-120 kW und eines sogar 320-360 kW. Und für die gesamte Berechnung genügt eine einzige Formel!

The completed data table is as follows:

Leistung in kW	Häufigkeit
40	15
80	46
120	25
160	4
200	3
240	3
280	1
320	2
360	1

Abbildung 45

3) Das fertige Histogramm soll schließlich folgendermaßen aussehen:

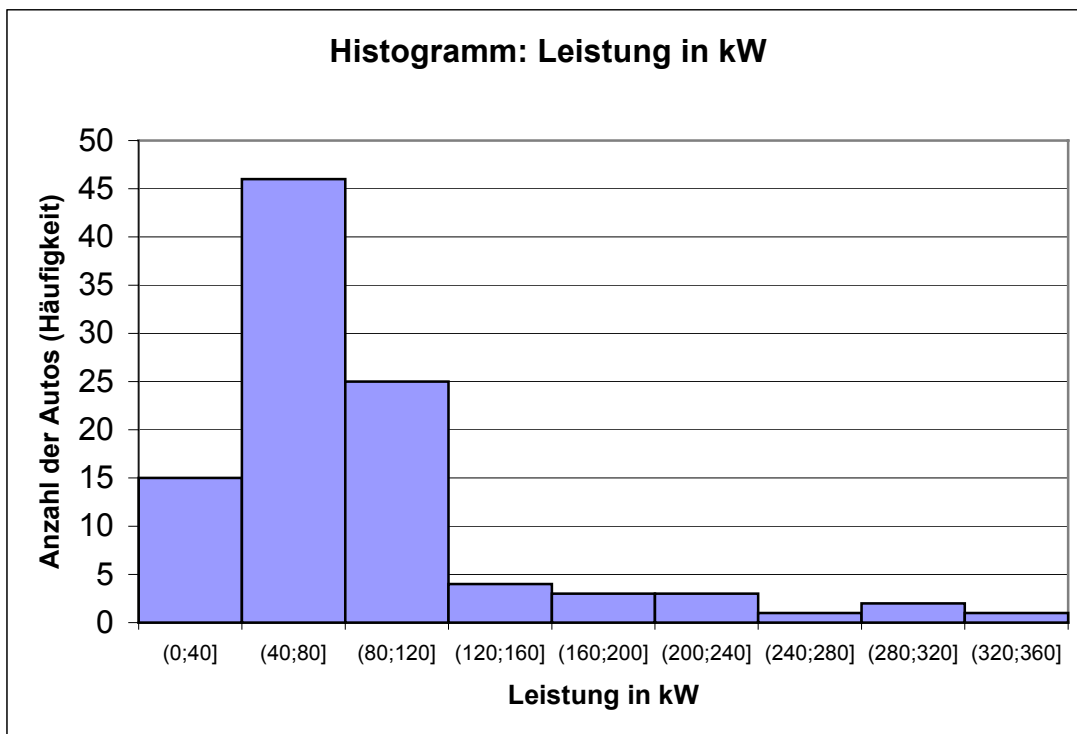


Abbildung 46

HISTOGRAMM

Nur wenn **gleiche Klassenbreiten** vorliegen, kann man ein Histogramm einfach mit einem Säulendiagramm erzeugen. Wie man ein Säulendiagramm mithilfe von Excel erstellt, wurde bereits in Abschnitt 5.2. Säulendiagramm gezeigt.

1) Zunächst fertigt man eine Liste an, in der die Anzahl der Autos in den verschiedenen Leistungsklassen eingetragen wird.

Leistung in kW	Anzahl der Autos (Häufigkeit)
(0;40]	15
(40;80]	46
(80;120]	25
(120;160]	4
(160;200]	3
(200;240]	3
(240;280]	1
(280;320]	2
(320;360]	1

Abbildung 47

2) Nach Anfertigung eines Säulendiagramms sollte die Grafik wie in Abbildung 48 aussehen.

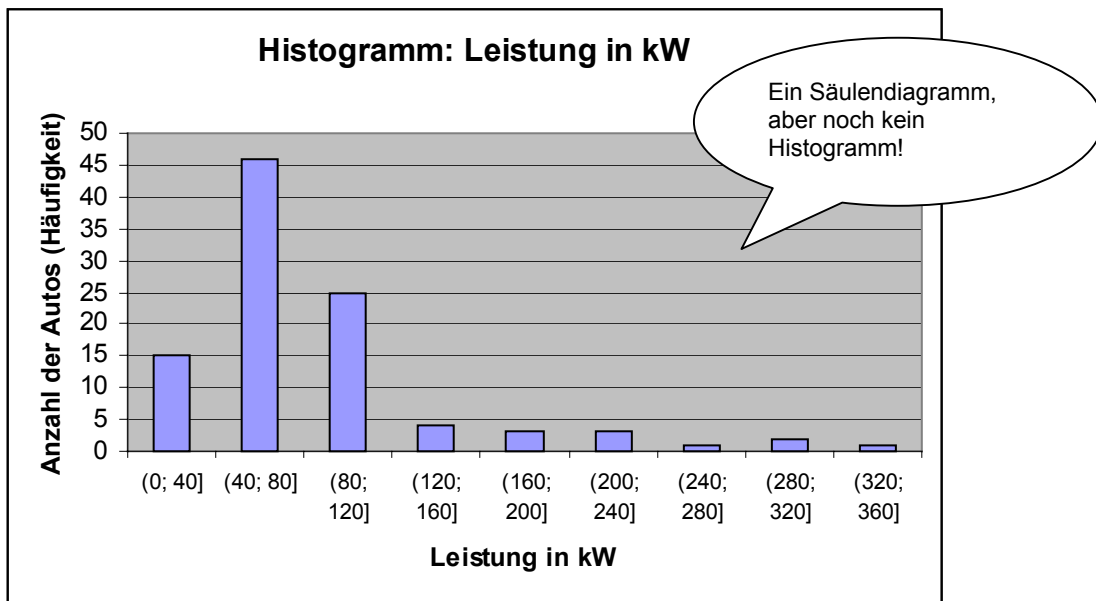


Abbildung 48

Beachte, dass dies noch kein Histogramm, sondern „nur“ ein Säulendiagramm ist. In einem Histogramm dürfen sich zwischen den Flächen keine Abstände befinden. Wir müssen also noch die Zwischenräume entfernen.

- 3) Durch einen Rechtsklick auf eine der Säulen erscheint ein Menü, in dem man **DATENREIHEN FORMATIEREN** anwählt.

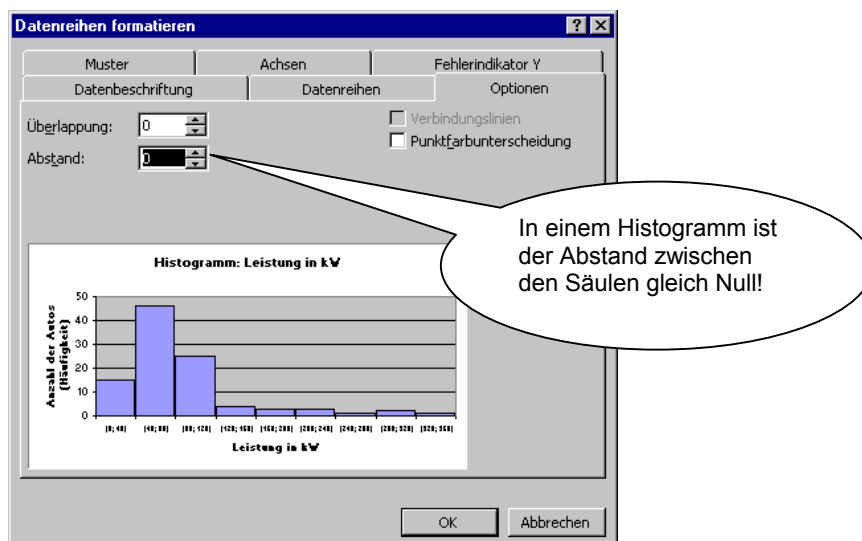


Abbildung 49

- 4) In diesem Fenster wird das Register **OPTIONEN** angewählt und dort der **ABSTAND** auf 0 verringert (Abbildung 49). Durch etwas Formatierung des Diagramms erhält man das Histogramm, wie es in Abbildung 46 gezeigt wird.



10. Welche statistischen Grafiken (Säulendiagramm, Stängel-Blatt-Diagramm, Histogramm) können für die Darstellung folgender Merkmale verwendet werden? Nenne jeweils einen Grund, warum bzw. warum nicht!

a. Merkmal: Anzahl der Plätze

Säulendiagramm: ja/nein Begründung:

Stängel-Blatt-Diagramm: ja/nein Begründung:

Histogramm: ja/nein Begründung:

b. Merkmal: Länge

Säulendiagramm: ja/nein Begründung:

Stängel-Blatt-Diagramm: ja/nein Begründung:

Histogramm: ja/nein Begründung:

c. Merkmal: Hubraum

Säulendiagramm: ja/nein Begründung:

Stängel-Blatt-Diagramm: ja/nein Begründung:

Histogramm: ja/nein Begründung:

11. Konstruiere

a. ein Säulendiagramm ...

b. ein Stängel-Blatt-Diagramm ...

c. ein Histogramm ... *... mit einem passenden Merkmal Deiner Wahl!*

6 Statistische Maßzahlen I: Lagemaße

6.1 Minimum und Maximum

Minimum

Das Minimum ist der kleinste vorkommende Wert einer Variablen. Wir haben das Minimum bereits im Abschnitt 3.1 Sortieren kennen gelernt. Da die Sortierfunktion die Werte der Größe nach ordnet, kann man diese zur Bestimmung des Minimums verwenden. Mit Excel ist es aber auch möglich, sich nur den Wert des Minimums anzeigen zu lassen. Als Beispiel berechnen wir die minimale Leistung.



MINIMUM

= min (DATENBEREICH)

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLJP	Verbrauch	Preis	Kategorie
Volvo S70 2,4	4	5	472	1391	2435	103	200	9,9	4	1	1	1	8,6	422441	L
Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L
VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

Abbildung 50

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLJP	Verbrauch	Preis	Kategorie
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

Abbildung 51

Die minimale **LEISTUNG** unter den Autos im Datensatz beträgt also 29 kW.

Maximum

Das Maximum ist der größte vorkommende Wert einer Variablen.

MAXIMUM

= max (DATENBEREICH)

The screenshot shows an Excel spreadsheet with the following data:

	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
96	Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L
97	VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
98	VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
99	VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
100	VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
101	VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
102	VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K
105		=max(H3:H102)														

Abbildung 52

Die größte **LEISTUNG** beträgt 325 kW.



12. Berechne das geringste Gewicht aller Autos im Datensatz mithilfe der Funktion für die Berechnung des Minimums!

13. Wie schnell fährt das schnellste Auto? (Verwende die Funktion max!)

6.2 Arithmetisches Mittel⁶

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Arithm. Mittel = Summe der Beobachtungen/Anzahl der Beobachtungen

Die Berechnung des arithmetischen Mittels ist formal denkbar einfach: Alle Beobachtungen werden addiert und durch die Anzahl der Beobachtungen geteilt. Das arithmetische Mittel verschafft einen ersten Überblick über die Lage der Daten.

Als Beispiel wollen wir das arithmetische Mittel der **LÄNGEN** der untersuchten Autos berechnen.

Ohne Computer ist dies zwar problemlos möglich, aber bereits eine etwas langwierige Angelegenheit. Zunächst müssen alle **LÄNGEN** addiert werden:

$$372\text{cm} + 407\text{cm} + 341\text{cm} + \dots + 472\text{cm} = 42617\text{cm.}$$

Nun wird die Gesamtlänge durch die Anzahl der Autos dividiert:

$$42617/100 = \underline{426,17}$$

Das arithmetische Mittel der **LÄNGEN** aller untersuchten Autos beträgt also 426,17cm.

ARITHMETISCHES MITTEL

Mit Excel kann das arithmetische Mittel auf diese Weise berechnet werden:

Alle Längen werden mit Hilfe der Formel **=SUMME(DATENBEREICH)** berechnet.

	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
96	Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L
97	VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
98	VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
99	VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
100	VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
101	VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
102	VW Polo 1,0	3	5	374	962	999	37	151	18,4	2	1	0	0	5,7	149900	K
105		=summe(E3:E102)														

Abbildung 53

⁶ Das arithmetische Mittel wird üblicherweise als „der Mittelwert“ bezeichnet. Da es aber auch noch andere Mittelwerte gibt, wollen wir hier den exakten Ausdruck verwenden.

Die Gesamtlänge wird durch die Anzahl der Autos dividiert.

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L
VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

426,17
=E105/100

Abbildung 54

Das arithmetische Mittel der Variablen Länge beträgt 426,17 cm.

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L
VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

426,17
426,17

Das arithmetische Mittel der Variablen Länge beträgt 426,17 cm.

Abbildung 55

Es geht aber noch einfacher! Excel bietet eine Formel an, die den Mittelwert direkt berechnet.

=MITTELWERT(DATENBEREICH)

- 1) Als Datenbereich wird jener Bereich angegeben, in dem sich die Daten befinden, deren Mittelwert berechnet werden soll. Nach Bestätigung der Eingabe wird das arithmetische Mittel für die markierten Daten automatisch berechnet (Abbildung 56, Abbildung 57).

Microsoft Excel - Datenanalyse mit Excel

REST | X ✓ = =mittelwert(E3:E102)

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L
VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

105 =mittelwert(E3:E102)

Abbildung 56

Microsoft Excel - Datenanalyse mit Excel

T120 | =

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L
VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

105 426,17

Das Ergebnis ist natürlich das Gleiche wie in Abbildung 55!

Abbildung 57



14. Berechne das arithmetische Mittel der Variablen Gewicht

- a. mittels Addition aller Werte und Division durch die Anzahl!
- b. mithilfe der vorgegebenen Formel!

6.3 Median

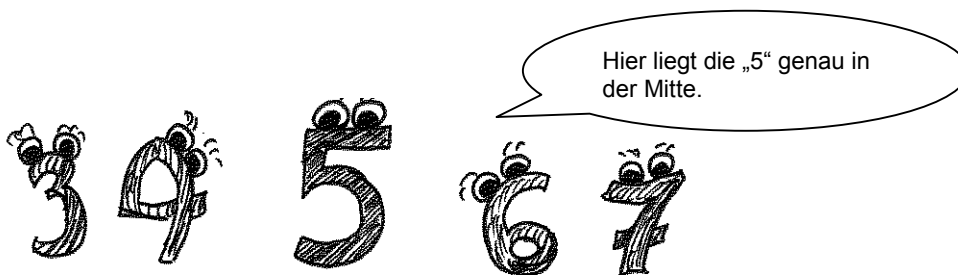
$$\tilde{x}_{0,5} = \begin{cases} x_{((n+1)/2)} & , \text{ falls } n \text{ ungerade} \\ \frac{1}{2} (x_{(n/2)} + x_{((n+2)/2)}) & , \text{ falls } n \text{ gerade} \end{cases}$$

Der Median befindet sich in einer sortierten Liste von Daten genau in der Mitte: Er ist dadurch charakterisiert, dass jeweils mindestens die Hälfte aller Beobachtungen einen Wert größer oder gleich bzw. kleiner oder gleich dem Median annehmen.

Der Median ist wie der Mittelwert eine Lagemaßzahl, also eine einzelne Kennziffer, die die „Lage“ vieler Daten beschreiben soll.

Die Bestimmung des Medians funktioniert wie folgt: die Daten werden der Größe nach sortiert. Jener Wert, der genau in der Mitte liegt, entspricht dem Median.

Dies ist allerdings nur bei einer *ungeraden Anzahl von Werten* ohne weiteres möglich:



Bei einer *geraden Anzahl von Werten* befindet sich kein einzelner Wert in der Mitte. Hier nimmt man die beiden mittleren Werte und berechnet deren arithmetisches Mittel. Dieser Wert wird dann als der Median bezeichnet.



Als Beispiel wollen wir den Median des **HUBRAUMS** im PKW-Datensatz bestimmen. Ohne Computer wird die Angelegenheit ziemlich mühsam, da zunächst alle Werte sortiert werden müssen.

Wert #	1	2	...	99	100
Hubraum in ccm	599	899	...	5.666	6.750

Tabelle 5

Sobald diese Hürde genommen wurde, wird es jedoch ziemlich einfach. Man muss lediglich die beiden mittleren Werte nehmen (es liegt eine gerade Zahl von Werten vor!) und deren arithmetisches Mittel berechnen. Das Ergebnis ist der Median.

Wert #	...	49	50	51	52	...
Hubraum in ccm	...	1.595	1.595	1.595	1.597	...

Tabelle 6

Die beiden mittleren Werte betragen jeweils 1.595 ccm.

Der Mittelwert ergibt sich dann aus $(1595+1595)/2 = 1595$ ccm.

Der Median des **HUBRAUMS** beträgt also 1595 ccm.

MEDIAN

Mit Excel beschränkt sich die Berechnung wiederum auf die Eingabe einer einzigen Formel, wobei die Daten vorher nicht einmal sortiert werden müssen.

= MEDIAN(DATENBEREICH)

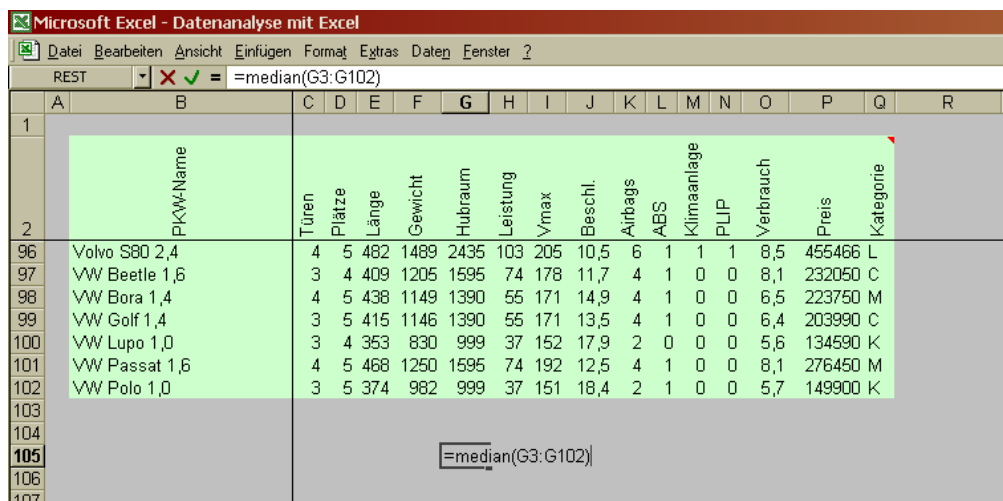


Abbildung 58

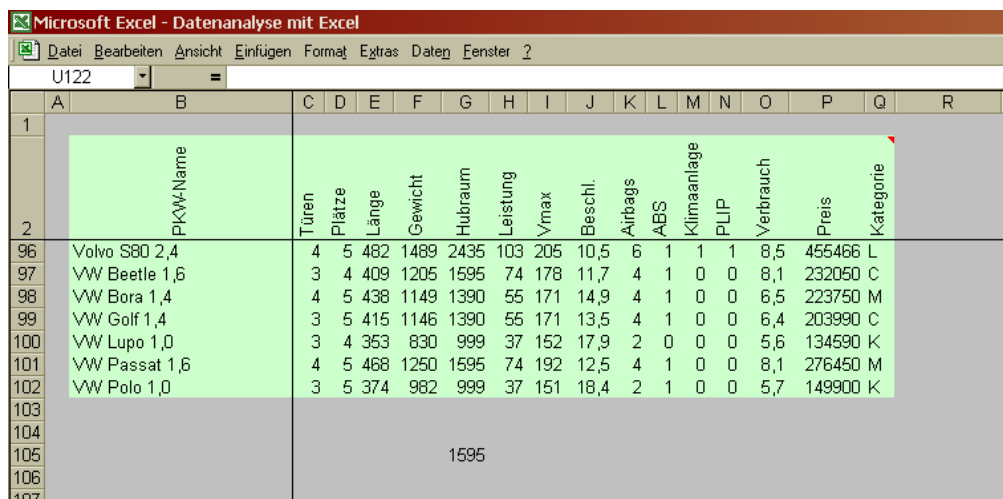


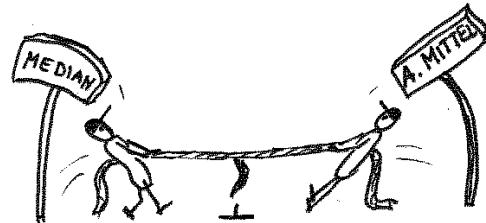
Abbildung 59



15. Berechne den Median der Variablen Gewicht auf zwei Arten!

6.4 Arithmetisches Mittel und Median – ein Vergleich

Besonders jetzt, da es uns so leicht fällt das arithmetische Mittel oder den Median aller Variablen zu berechnen, stellt sich die Frage, welche Aussage diese beiden Werte haben.



Am Beispiel des PKW-Datensatzes lassen sich die unterschiedlichen Eigenschaften dieser beiden Werte gut darstellen.

Zum Vergleich betrachten wir das arithmetische Mittel und den Median von zwei verschiedenen Variablen - der **LÄNGE** in Metern und des **PREISES** in Schilling.

	arithm. Mittel	Median
Länge in m	426,2	424
Preis in S	439.128	244.900

Tabelle 7

Aus Tabelle 7 wird eine Tatsache sofort sehr deutlich: Bei der Länge unterscheiden sich arithmetisches Mittel und Median kaum, beim Preis ist ein großer Unterschied zu erkennen. Wie kann so etwas zustande kommen? Wir wissen bereits, wie arithmetisches Mittel bzw. Median berechnet werden. In die Berechnung des Mittelwerts fließen alle Werte mit ein, in die Berechnung des Medians der Wert (die beiden Werte), der (die) sich in der Mitte befindet (befinden).

Exkurs: Ein einfaches Beispiel

Ein einfaches Beispiel zeigt, welche großen Unterschiede diese Berechnungsarten verursachen können. Für dieses Beispiel greifen wir aus dem PKW - Datensatz zehn Modelle heraus: vier Kleinwägen, fünf Mittelklasseautos und einen starken Sportwagen. Diese zehn Autos teilen wir in zwei Gruppen zu je fünf und betrachten deren Leistung.

Gruppe 1	Leistung in kW		Gruppe 2	Leistung in kW
Mazda 626 1,8i	74		Peugeot 106 1,0 ZEN	37
Seat Toledo 1,6 Stella	74		Seat Arosa 1,0i Stella	37
VW Passat 1,6	74		VW Lupo 1,0	37
Fiat Marea 100 16V SX	76		Suzuki Swift 1,0 GLS	39
BMW 316i	77		Mercedes CL 500	225
Arithmetisches Mittel	75		Arithmetisches Mittel	75

Tabelle 8

Betrachten wir nun das arithmetische Mittel der Leistung in den beiden Gruppen. Sowohl in Gruppe A $(74+74+74+76+77)/5 = 75$ als auch in Gruppe B $(37+37+37+39+225)/5 = 75$ beträgt das arithmetische Mittel 75 kW.

Wir sehen also: die durchschnittliche Leistung der Autos ist in beiden Gruppen genau gleich groß.



Wir sehen aber auch: das arithmetische Mittel kann offensichtlich keine Aussagen über die *Aufteilung* der Werte machen. Während sich in der ersten Gruppe nur annähernd gleich starke Mittelklasse-Autos befinden, wird das Leistungsdefizit der Kleinwägen in Gruppe 2 für das arithmetische Mittel durch einen einzigen starken Sportwagen wettgemacht.

Sehen wir uns nun den Median an: sortiert man die kW-Werte der Größe nach, steht in Gruppe 1 der Wert 74 in der Mitte, in Gruppe 2 der Wert 37. Tabelle 9 faßt die Ergebnisse nochmals zusammen.

	arithm. Mittel	Median
Gruppe 1	75	74
Gruppe 2	75	37

Tabelle 9

In **Gruppe 1** sind arithmetisches Mittel und Median beinahe gleich groß. Dies zeigt an, dass die *Abweichungen von der Mitte* nach beiden Richtungen ungefähr gleich groß sind.

In **Gruppe 2** hingegen ist das arithmetische Mittel deutlich höher als der Median. Daraus kann man erkennen, dass es einen oder mehrere *Ausreißer nach oben* geben muß. Diese ziehen das arithmetische Mittel nach oben, lassen den Median aber unbehelligt. In

Gruppe 2 ist es der Mercedes CL 500, der mit seiner hohen Leistung das arithmetische Mittel auf 75kW bringt, während die restlichen Autos in dieser Gruppe eher schwächer motorisiert sind.

Umgekehrt würde ein arithmetisches Mittel, das kleiner ist als der Median, darauf hindeuten, dass es einen oder mehrere Ausreißer nach unten geben muß.

Exkurs Ende

Kehren wir nun zurück zur Betrachtung der **LÄNGE** und des **PREISES**.

Bei der Variablen **LÄNGE** sind arithmetisches Mittel und Median beinahe gleich. Es liegt also die Vermutung nahe, dass die Abweichungen nach unten und nach oben von diesen zentralen Werten etwa gleich sein werden. Zur Veranschaulichung dieser Vermutung eignet sich ein Histogramm, das in Abbildung 60 gezeigt wird.

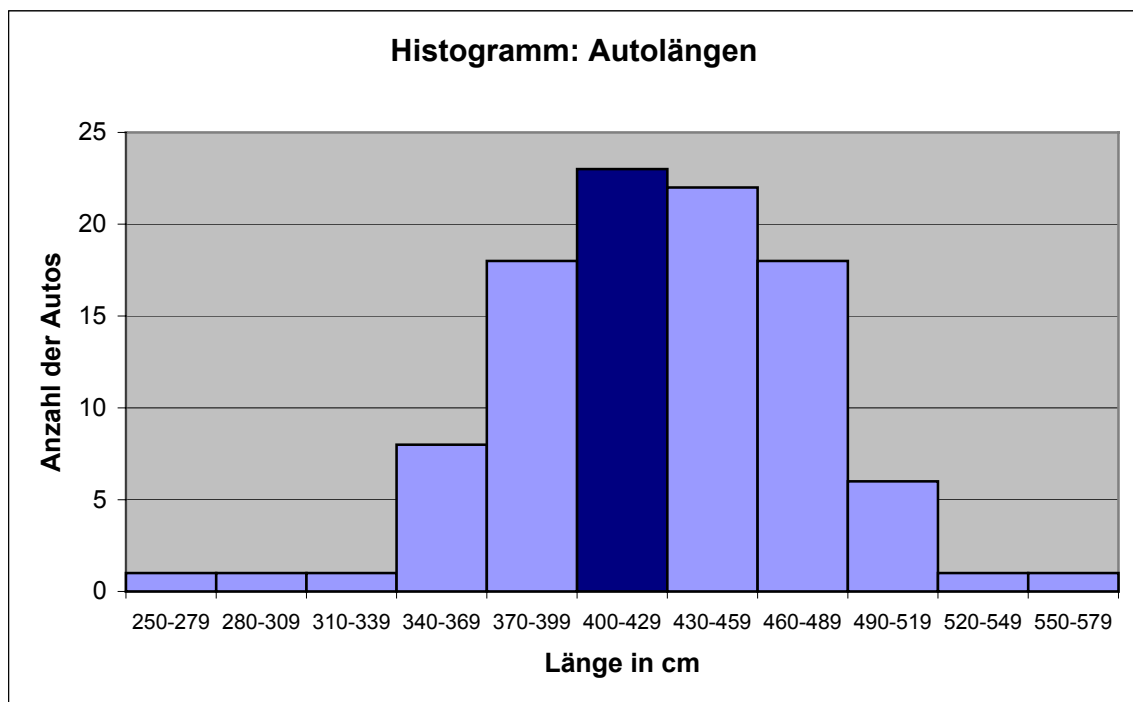


Abbildung 60

Die Abbildung zeigt, dass wir mit unserer Vermutung durchaus richtig liegen. Die Werte liegen gleichermaßen links wie rechts vom Mittelwert und Median (sowohl Mittelwert als auch Median liegen in dem dunkel markierten Bereich). Es gibt einige sehr lange, dafür auch einige besonders kurze Autos. Alles in allem sind die Autolängen aber recht gleichmäßig um die Mitte angeordnet.

Ein anderes Bild ergibt sich, wenn man den Preis der Autos betrachtet.

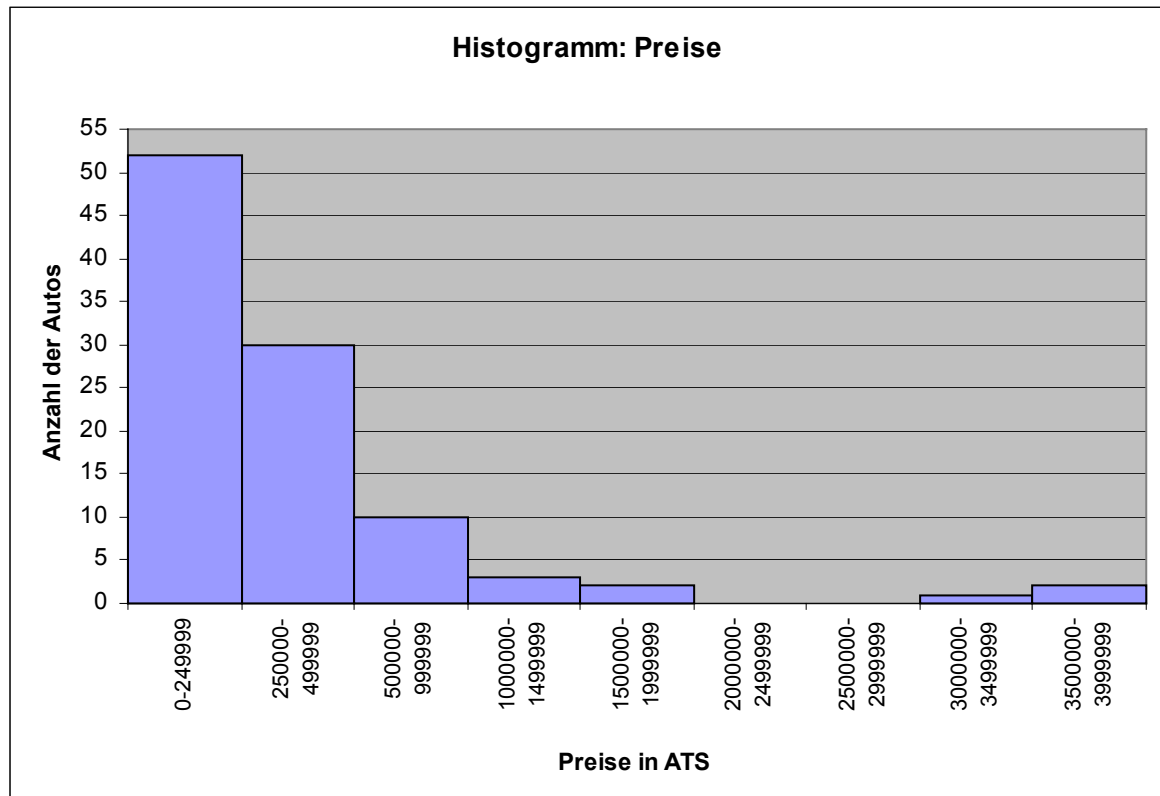


Abbildung 61

Hier zeigt sich ganz deutlich, dass die allermeisten Autos (~82%) in der Preisklasse zwischen 0 und 500.000 S liegen. Es gibt aber auch einige Autos, deren Preis sogar jenseits der 3 Mio S - Marke liegt. Diese Ausreißer nach oben beeinflussen das arithmetische Mittel recht stark, lassen aber den Median völlig unberührt. Aus diesem Grund ist das arithmetische Mittel in diesem Fall so deutlich höher als der Median.

Mittelwert oder Median? Es zeigt sich also, dass sowohl der Mittelwert als auch der Median allein unzureichende Maßzahlen sind, um die Verteilung der Daten zu beschreiben. Gemeinsam gewinnen sie aber einiges an Aussage, man kann nicht nur einen zentralen Wert angeben, sondern auch ungefähre Angaben über die Verteilung der Werte und Ausreißer machen.



16. Berechne Mittelwert und Median der Variablen Gewicht, Hubraum, Leistung, Höchstgeschwindigkeit, Beschleunigung und Verbrauch! Welche Vermutungen über Aufteilung der Werte und Ausreißer kann man aufgrund der Ergebnisse anstellen?

6.5 Teilbereiche

Bisher haben wir folgende Berechnungen angestellt:

- Einerseits wurde der **gesamte Datensatz** untersucht und etwa Minimum, Maximum, Mittelwert oder Median berechnet.
- Andererseits haben wir gesehen, wie man **einen Teil** herausfiltert (z.B. alle Sportwagen) und Berechnungen für diese Teilmenge anstellt.

Mitunter ist es aber auch von Interesse, Berechnungen für mehrere Teilbereiche gleichzeitig anzustellen.

Beispiel: Inwiefern unterscheiden sich die verschiedenen PKW - **KATEGORIEN** hinsichtlich des **PREISES**? Um wie viel ist etwa ein Mittelklasseauto durchschnittlich teurer als ein Wagen der Compact-Klasse? Und in welchen Preisbereich passen die Sportwagen?

TEILBEREICHE

Fragen wie diese können mit Excel am einfachsten mit dem Menüpunkt **DATEN - TEILERGEBNISSE** beantwortet werden. Wir wollen also zum Beispiel das arithmetische Mittel der **PREISE** für jede **KATEGORIE** berechnen.

- 1) Zunächst wird der gesamte Datenbereich markiert.
- 2) Danach ist es notwendig, den Datensatz nach jener Variablen zu sortieren, für die man die Teilergebnisse berechnen will. In unserem Fall ist das die **KATEGORIE**. Wir sortieren die Daten also nach der **KATEGORIE**.
- 3) Mit dem Befehl **DATEN - TEILERGEBNISSE** gelangt man in ein Dialogfenster, das die gewünschten Optionen anbietet (Abbildung 62, Abbildung 63).

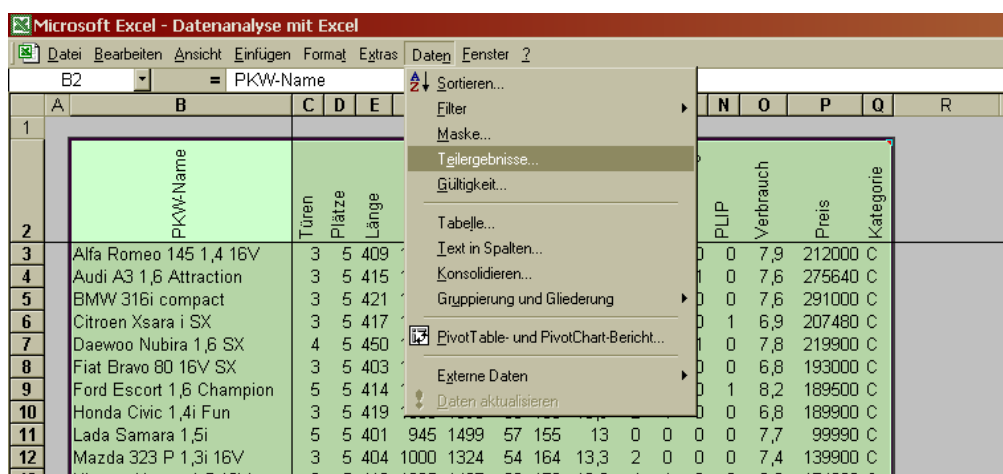


Abbildung 62

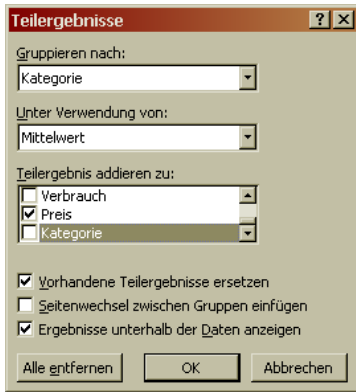


Abbildung 63

4) Im Dialogfenster von Abbildung 63 sind bereits die gewünschten Optionen eingestellt: Wir wollen für die verschiedenen **KATEGORIEN** den Mittelwert der **PREISE** berechnen.

Das bedeutet, dass zunächst die Variable angegeben werden muss, für die etwas berechnet werden soll: **GRUPPIEREN NACH: KATEGORIE**

Danach wählen wir aus, was wir für die angegebene Variable berechnen wollen: **UNTER VERWENDUNG VON: MITTELWERT**

Schließlich müssen wir noch angeben, für welche Variable wir den Mittelwert in jeder Kategorie berechnet haben wollen: **TEILERGEBNIS ADDIEREN ZU: PREIS**

Nach einem Klick auf OK erscheint folgendes Ergebnis.

PKW-Name	Türen	Preis	Kategorie
Subaru Alto L	3	5 360	K
Suzuki Swift 1,0 GLS	3	5 375	K
VW Lupo 1,0	3	4 353 830 999	K
VW Polo 1,0	3	5 374 982 999	K
148939 K Mittelwert			
Alfa Romeo 166 2,0 TS	4	5 472 1420 1970	L
Mazda Xedos 9 2,0 V6	4	5 483 1455 1995	L
Mercedes E 200	4	5 480 1510 1998	L
Mercedes S 320	4	5 504 1770 3199	L
Nissan Maxima 2,0 V6 24V	4	5 477 1465 1995	L
Opel Omega 2,2 16V	4	5 490 1605 2198	L
Peugeot 607 2,2	4	5 488 1455 2230	L
Renault Safrane 2,5 RXE	5	5 477 1630 2435	L
Rolls Royce Silver Seraph	4	5 564 2450 5379	L
Volvo S70 2,4	4	5 472 1391 2435	L
Volvo S80 2,4	4	5 482 1489 2435	L
880000 L Mittelwert			
Alfa Romeo 1,6T Spark	4	5 443 1230 1598	M

Abbildung 64

- 5) Für jede einzelne **KATEGORIE** wurde nun am Ende der **KATEGORIE** der entsprechende **MITTELWERT** des **PREISES** eingefügt. Weiters wurde am linken Rand eine Gliederungsleiste eingefügt, die es ermöglicht, die eigentlichen Daten mit einem Klick auszublenden, um nur noch die Teilergebnisse anzuzeigen.

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
194638 C Mittelwert															
148939 K Mittelwert															
880000 L Mittelwert															
256537 M Mittelwert															
825071 S Mittelwert															
439128 Gesamtmittelwert															

Abbildung 65

- 6) In Abbildung 65 sind nur mehr die Mittelwerte der **PREISE** der einzelnen **KATEGORIEN** und der Gesamtmittelwert sichtbar, die eigentlichen Daten wurden ausgeblendet. Dies ermöglicht einen schnellen und übersichtlichen Vergleich der Berechnungen für die Teilbereiche. Durch klicken auf die -Symbole am linken Rand (Abbildung 65) können die Daten wieder eingeblendet werden.
- 7) Ausschalten der Teilergebnisse: Mithilfe von **DATEN - TEILERGEBNISSE - ALLE ENTFERNEN** werden die angezeigten Teilergebnisse gelöscht und der Datensatz wieder in den vorherigen Zustand zurückversetzt.



17. Verwende den Befehl **Teilergebnisse**, um folgende Fragen zu beantworten!

- Sind die Autos mit Klimaanlage im Mittel teurer als jene ohne Klimaanlage?
- In welcher Kategorie beträgt das Minimum des Verbrauchs 6,5l?
- Welches Auto mit 2 Plätzen hat die größte Länge?
Wie hätte man diese Aufgabe anders (einfacher?) lösen können?
- Wie viele Wagen der verschiedenen Kategorien haben ABS?

6.6 Mittelwert in Gruppen ⇔ Gesamtmittelwert

Im vorigen Abschnitt wurde gezeigt, wie man Berechnungen für mehrere Teilmengen gleichzeitig anstellt. So wurde etwa das arithmetische Mittel des **PREISES** für jede **KATEGORIE** berechnet. Es stellt sich nun die Frage, ob man aus den Gruppenmittelwerten auch den Gesamtmittelwert berechnen kann. Und wenn ja, wie?

Exkurs: So nicht!

Bevor die richtige Vorgangsweise demonstriert wird, soll hier vor einem häufig gemachten Fehler gewarnt werden: der Gesamtmittelwert ist nicht einfach der Mittelwert der Gruppenmittelwerte!

Erinnern wir uns an die Berechnung des Mittelwertes: Alle einzelnen Daten werden addiert und durch ihre Anzahl dividiert.

Im PKW-Datensatz ergibt dies für den Preis einen Gesamtmittelwert von **439.128,-**

Berechnet man den Mittelwert der Gruppenmittelwerte für die einzelnen Kategorien (194.638,- + 148.939,- + 256.537,- + 825.071,- + 880.000,-)/5 = **461.036,7**

gelangt man offensichtlich zu einem anderen Ergebnis.

Warum dies so ist, sei an einem kleinen Beispiel erklärt: Wir nehmen wieder fünf Autos aus dem Datensatz, teilen diese in zwei Gruppen (A und B) und betrachten deren **LEISTUNG**.

Gruppe A	Leistung in kW	Gruppe B	Leistung in kW
Citroen Saxo 1,1i SX	40	VW Beetle 1,6	74
Opel Corsa 1,0 Joy	40	Mazda 626 1,8i	74
		Audi A3 1,6 Attraction	74
Mittelwert	40	Mittelwert	74

Tabelle 10

Beide Autos in Gruppe A haben eine **LEISTUNG** von 40 kW, der Mittelwert in Gruppe A beträgt also ebenfalls 40 kW. In Gruppe B haben alle drei Autos 74 kW, der Mittelwert in Gruppe B beträgt also ebenfalls 74 kW.

Berechnet man den Mittelwert der Gruppenmittelwerte, erhält man $(40+74)/2=57$.

Berechnet man jedoch den Gesamtmittelwert, ist das Ergebnis $(40+40+74+74+74)/5 = 60,4$.

Der Unterschied kommt dadurch zustande, dass der Gesamtmittelwert jeden einzelnen Wert gleich stark berücksichtigt. 40 kW kommen nur 2x vor, 74 kW jedoch 3x. Mit der Berechnung der Gruppenmittelwerte verliert man aber die Information über die Anzahl

der Elemente innerhalb der Gruppe, wodurch 2x40kW plötzlich „gleich stark“ wie 3x74kW gewichtet werden.

Aus diesem Grund ist üblicherweise der **Gesamtmittelwert ungleich dem Mittelwert der Gruppenmittelwerte**.

Exkurs Ende

So geht´s!

Will man den Gesamtmittelwert mithilfe der Mittelwerte der Gruppen berechnen, muss man also auch noch die Anzahl der Gruppenelemente berücksichtigen.


$$\text{Gesamtmittelwert} = \frac{\text{Gr.mittelwert}_1 * \text{Zahl Elemente in Gr.}_1 + \text{Gr.mittelwert}_2 * \text{Zahl Elemente in Gr.}_2 + \dots}{\text{Gesamtzahl der Elemente}}$$

Im Beispiel der Autopreise würde die Berechnung des Gesamtmittelwerts aus den Gruppenmittelwerten folgendermaßen funktionieren.

Gruppe	Mittelwert	Anzahl der Elemente	Mittelwert*Anzahl
C	194.638	20	3.892.760
K	148.939	24	3.574.536
L	880.000	19	16.720.000
M	256.537	19	4.874.203
S	825.071	18	14.851.278
Summe	2.305.185	100	43.912.777
Mittelwert der Gruppenmittelwerte: 2.305.185/5 =			461.037
Gesamtmittelwert			439.128

Abbildung 66

Die einzelnen Gruppenmittelwerte (C, K, M, S, X) werden jeweils mit der Anzahl der Autos in den entsprechenden Gruppen multipliziert. Die Ergebnisse werden addiert und die Summe durch die Gesamtanzahl der Autos dividiert. Das Ergebnis ist der Gesamtmittelwert.



18. Wann ist der Gesamtmittelwert gleich dem Mittelwert der Gruppenmittelwerte?

19. Teile die Autos in zwei Gruppen (mit/ohne) ABS und berechne für jede Gruppe den Mittelwert des Preises! Berechne danach den Mittelwert der beiden Gruppenmittelwerte! Ist dieser gleich dem Gesamtmittelwert?

6.7 Quartile

Wie wir bereits in der Diskussion um Mittelwert und Median gesehen haben, ist eine einzelne Maßzahl oft nicht ausreichend, um Strukturen in den Daten zufriedenstellend zu beschreiben.

Aus diesem Grund entstand die Idee, dass man die Daten nicht nur in zwei Hälften, sondern auch in vier Viertel teilen könnte.



Dazu werden die Daten wie bei der Berechnung des Medians der Größe nach sortiert. Anschließend wird untersucht, in welchem Bereich jeweils ein Viertel der Daten liegt. Die Grenzen dieser Bereiche werden Quartile⁷ genannt (lat. Quartus: der Vierte)

Am einfachsten sind das 0., 2. und 4. Quartil zu bestimmen. Das 0. Quartil ist nichts anderes als das Minimum, das 2. Quartil ist der Median und das 4. Quartil ist das Maximum.

Quartil	Bezeichnung	Eigenschaft
0.	Minimum	Größer 0% der Daten
1.		Größer gleich 25% der Daten
2.	Median	Größer gleich 50% der Daten
3.		Größer gleich 75% der Daten
4.	Maximum	Größer gleich 100% der Daten

Tabelle 11

Beispiel: Wir wollen die 5 Quartile der Autopreise bestimmen.

Das 0.Quartil und das 4. Quartil sind, wie gesagt, nichts anderes als Minimum und Maximum.

0. Quartil	99.900,-
4. Quartil	3.829.000,-

Tabelle 12

⁷ Die Unterteilung erfolgt mitunter auch noch genauer z.B. in Dezile (Zehntel), für unsere Zwecke sind Quartile aber ausreichend. Bei der Unterteilung in n Teile wird ein Teil allgemein als Quantil bezeichnet.

Nun ist das 2. Quartil (der Median) an der Reihe. Auch diesen haben wir bereits berechnet, er beträgt 244.900,-

0. Quartil	99.900,-
2. Quartil	244.900,-
4. Quartil	3.829.000,-

Tabelle 13

Jetzt fehlen noch das erste und das dritte Quartil.

Die Berechnung des ersten Quartils erfolgt ganz ähnlich der Berechnung des Medians bei einer geraden Anzahl von Werten. Erinnern wir uns: bei einer geraden Anzahl von Werten wurden die beiden mittleren Werte herangezogen. Deren arithmetisches Mittel entsprach dem Median.

Für das erste Quartil werden nun ebenfalls zwei Werte herangezogen - der letzte Wert im ersten Viertel und der erste Wert im zweiten Viertel. In unserem Fall sind dies die Werte an der 25. und 26. Stelle (164.900,- und 169.900,-).

Wert #	...	24	25	26	27	...
Preis	...	160.300	164.900	169.900	174.900	...

Hier endet das 1. Viertel ...

Tabelle 14

Von diesen beiden Werten wird nun allerdings nicht das arithmetische Mittel berechnet. Um dem Umstand Rechnung zu tragen, dass links vom 1. Quartil nur ein Viertel der Werte liegt, rechts vom 1. Quartil jedoch drei Viertel, wird auch für das 1. Quartil ein Viertel vom 25. Wert und drei Viertel vom 26. Wert genommen und diese beiden Werte addiert.⁸

Das 1. Quartil ergibt sich also aus $\frac{1}{4} \cdot 164.900 + \frac{3}{4} \cdot 169.900 = 168.650,-$.

Für das 3. Quartil wird spiegelverkehrt zum 1. Quartil der letzte Wert der ersten drei Viertel und der erste Wert des letzten Viertels herangezogen. In unserem Fall sind dies der 75. und der 76. Wert.

... und hier beginnt das letzte Viertel!

Wert #	...	74	75	76	77	...
Preis	...	410.000	422.441	441.570	452.000	...

Tabelle 15

Das 3. Quartil ergibt sich nun aus $\frac{3}{4} \cdot 422.441 + \frac{1}{4} \cdot 441.570 = 427.223,25$.

⁸ Für die Berechnung des 1. bzw. 3. Quartils verwendet Excel diese relativ komplizierte Formel. In der Literatur finden sich jedoch auch einfachere Berechnungsmethoden.

Zusammengefaßt ergeben sich für die Variable **PREIS** folgende Quartile:

Quartil	Preis in ATS
0. Quartil	99.900,-
1. Quartil	168.650,-
2. Quartil	244.900,-
3. Quartil	427.223,25
4. Quartil	3.829.000,-

Tabelle 16

QUARTILE

Mit Excel lassen sich alle Quartile recht einfach berechnen, so dass man sich dabei sogar die eigene Berechnung von Minimum, Maximum und Median ersparen kann.

Die Formel für die Berechnung lautet:

= QUARTILE(DATENBEREICH; QUARTIL)

Der Datenbereich gibt wieder jenen Bereich an, in dem die Daten stehen, deren Quartile berechnet werden sollen. Nach dem Strichpunkt muß das gewünschte Quartil angegeben werden. Es empfiehlt sich natürlich, gleich alle fünf Quartile berechnen zu lassen.

Als Beispiel wollen wir nochmals die Quartile der Variablen **PREIS** berechnen.

1) Wir fertigen unterhalb der Daten eine kleine Tabelle an.

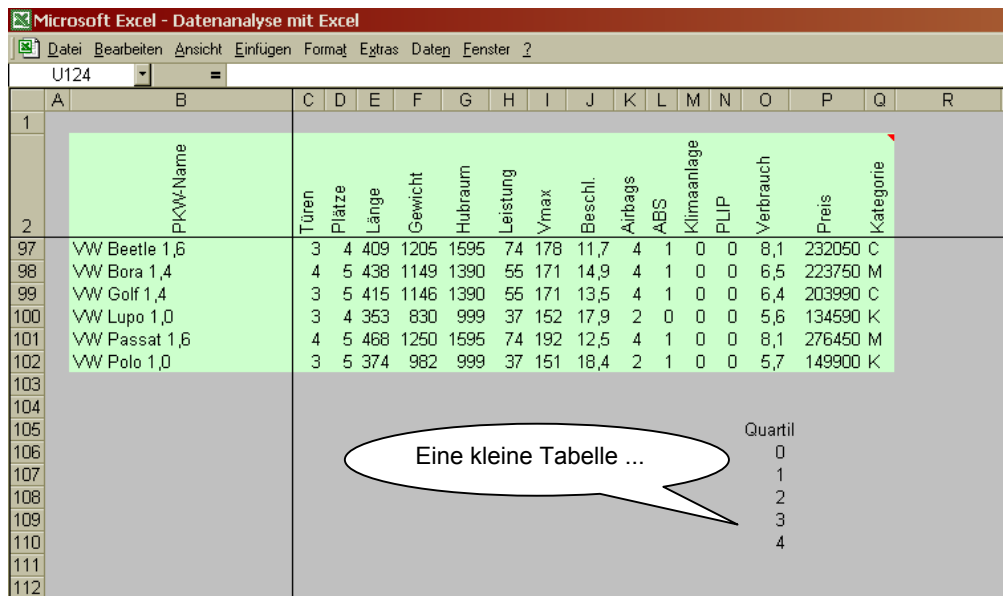


Abbildung 67

2) Nun wird die Formel für die Berechnung der Quartile eingegeben.

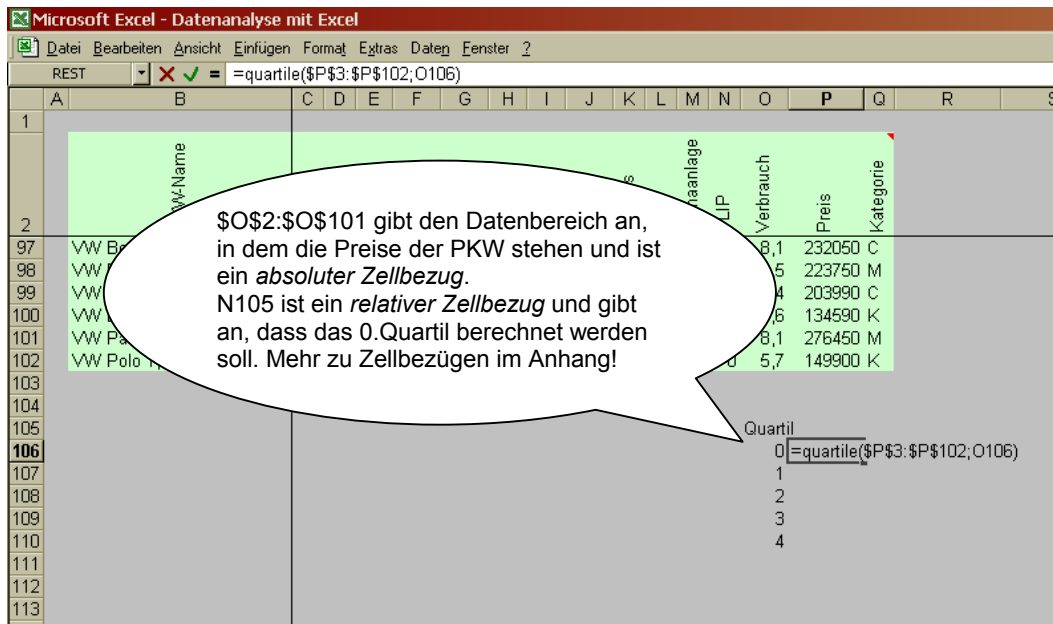


Abbildung 68

3) Nach Bestätigung der Eingabe wird das 0. Quartil (Minimum) angegeben. Der kleinste **PREIS** beträgt also 99.000,-.

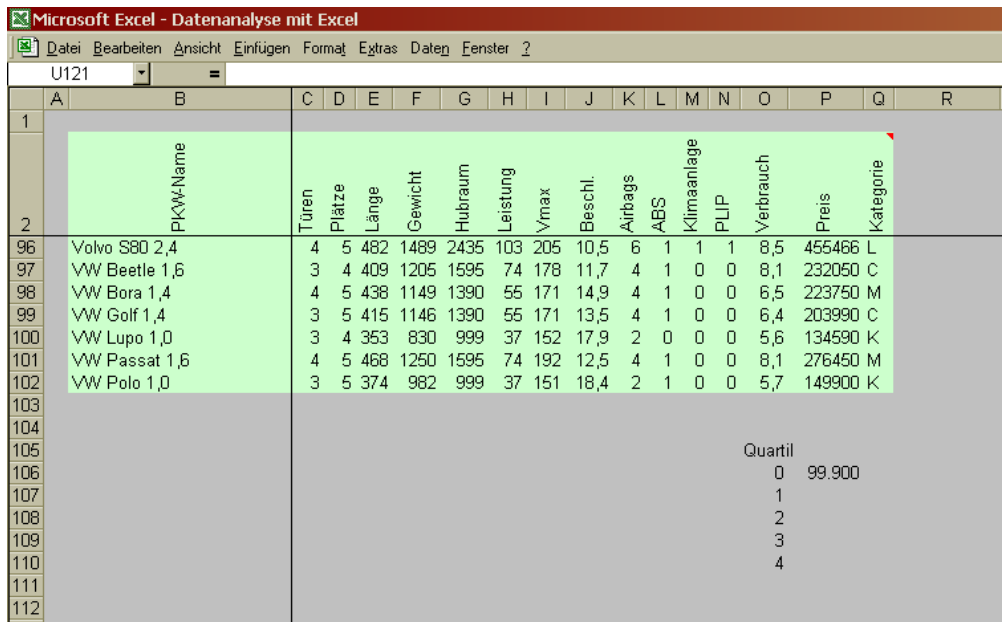


Abbildung 69

4) Nachdem die Formel entsprechend auf alle Quartile angewendet wurde⁹, erhält man folgendes Ergebnis (Abbildung 70).

⁹ Nachdem absolute bzw. relative Zellbezüge verwendet wurden, kann die Formel einfach nach unten gezogen werden. Wie Zellbezüge funktionieren, wird im Anhang 2 erklärt.

Microsoft Excel - Datenanalyse mit Excel

U122

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1																		
2		PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie	
96		Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	L	
97		VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C	
98		VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M	
99		VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C	
100		VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K	
101		VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M	
102		VW Polo 1,0	3	5	374	982	999						0	0	5,7	149900	K	
103																		
104																		
105																		
106																		
107																		
108																		
109																		
110																		
111																		
112																		

Alle Quartile auf einen Blick!

Quartil	Preis
0	99.900
1	168.650
2	244.900
3	427.223,25
4	3.829.000

Abbildung 70

6.8 Mittelwert – Quartile

Die Betrachtung der Quartile erlaubt nun genauere Aussagen über die Aufteilung der Werte als der Median alleine.

Am Beispiel der Autopreise können wir anhand der Quartile ablesen, dass ein Viertel der untersuchten Autos in der Preisklasse von 99.900,- bis 168.650,- liegt.

Die Hälfte der Autos kostet nicht mehr als 244.900,- und drei Viertel sind nicht teurer als 427.223,25.

Das letzte Viertel der Autos liegt preislich in der sehr großen Klasse zwischen 427.223,25 und 3.829.000,-.

Noch eine Bemerkung zum Vergleich des Mittelwerts mit den Quartilen. In der Diskussion um Mittelwert und Median haben wir festgestellt, dass eine große Differenz zwischen Mittelwert und Median auf eine ungleichmäßige Verteilung der Werte um die Mitte bzw. auf Ausreißer hindeutet.

Umso mehr gilt diese Feststellung für die Quartile im Vergleich zum Mittelwert. Wenn der Mittelwert wie beim **PREIS** sogar größer ist als das dritte Quartil, deutet das auf extreme Ausreißer und eine sehr ungleichmäßige Verteilung um den Mittelwert hin. Das dies beim **PREIS** der Fall ist, haben wir ja bereits in einigen Tabellen und Diagrammen feststellen können.

6.9 Kastenschaubild

Für die gemeinsame Darstellung aller Quartile wurde das sogenannte Kastenschaubild entwickelt. Dieses zeigt auf einen Blick, wie die fünf Quartile auf den gesamten Datenbereich verteilt sind.

Betrachten wir ein solches Kastenschaubild für das Merkmal **LÄNGE** im Detail.

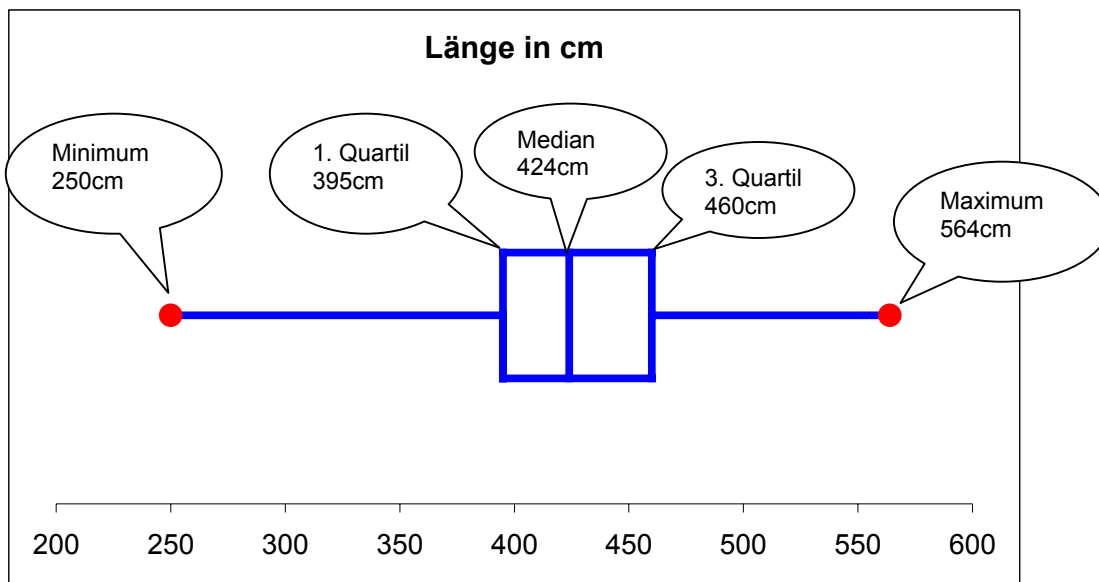


Abbildung 71

- Der rote Punkt links gibt das Minimum an. Zwischen dem roten Punkt und dem Beginn des Kastens (1. Quartil) liegen 25% der Daten.
- Die mittleren 50% der Daten werden durch den Kasten markiert. Der Kasten selbst wird durch die Median-Linie wieder in zwei 25%-Blöcke getrennt.
- Vom Kasten bis zum Maximum liegen schließlich wieder 25% der Daten.

Wir sehen also: ein Kastenschaubild stellt alle Quartile übersichtlich dar und ermöglicht so auf einen Blick ungefähre Aussagen über die Verteilung der Daten.

Betrachten wir noch ein zweites Beispiel: das Kastenschaubild für das Merkmal **PREIS**.

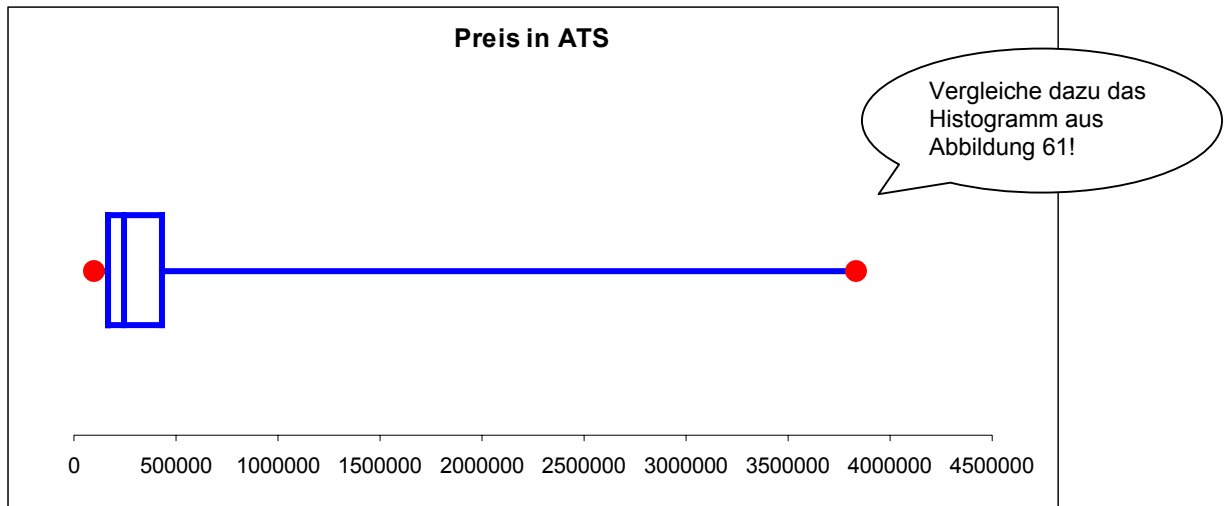


Abbildung 72

Auch so kann ein Kastenschaubild aussehen. Wir erkennen, dass 75% der Autos weniger als 500.000 S kosten, die Hälfte sogar weniger als 250.000 S.

KASTENSCHAUBILD

Obwohl das Kastenschaubild nicht standardmäßig als Diagrammtyp angeboten wird, kann man am Tabellenblatt Kastenschaubild der Datei „Datenanalyse mit Excel.xls“ ein solches Diagramm anfertigen. Im gelben Bereich muss man nur den Variablennamen und die vier Quartile eingeben und Excel zeichnet automatisch ein passendes Diagramm.

Variable	Preis in ATS
Minimum	99900
1. Quartil	168650
Median	244900
3. Quartil	427223,25
Maximum	3829000

Abbildung 73



20. Zeichne das Kastenschaubild der Variablen Höchstgeschwindigkeit! (mithilfe der Datei Kastenschaubild oder der Vorlage aus Abbildung 74) Was kannst du über die Verteilung der Daten aussagen?

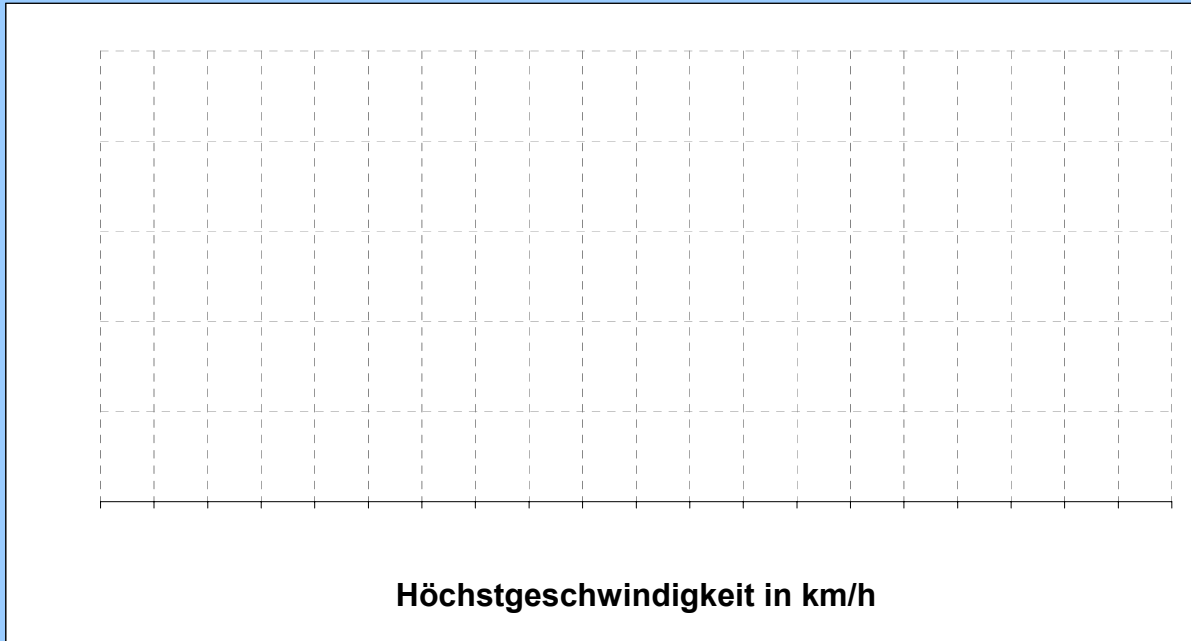


Abbildung 74

21. Zeichne das Kastenschaubild der Variablen Hubraum! Was kannst du über die Verteilung der Daten aussagen?

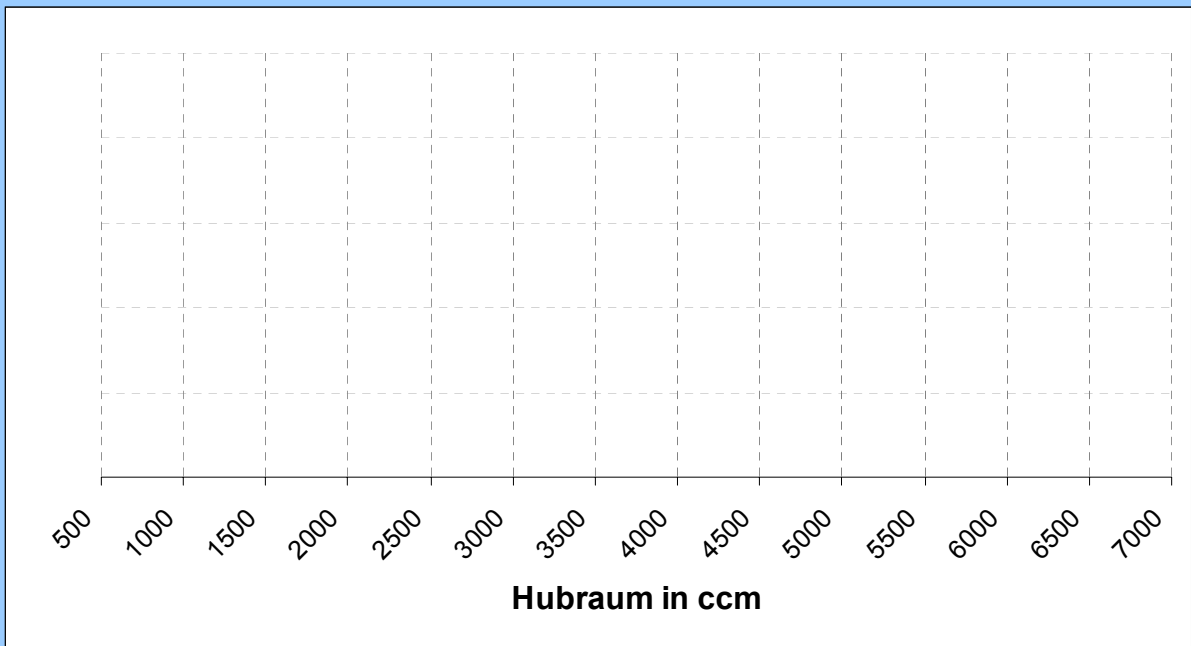


Abbildung 75

7 Statistische Maßzahlen II: Streuung

Bereits in der Diskussion um Mittelwert und Median haben wir gesehen, dass eine Zahl allein relativ wenig über die Gesamtheit der Daten aussagen kann. Es kommt nicht nur darauf an, um welchen Wert sich die Daten gruppieren, sondern z.B. auch darauf, wie nah sich die Daten um diesen zentralen Wert scharen.

Dazu ein kleines Beispiel: Nehmen wir wieder zwei Gruppen von jeweils 5 Autos.

Gruppe A	Leistung in kW	Gruppe B	Leistung in kW
Ford Puma 1,4	66	MCC Smart pure	33
Ford Mondeo 1,6i	70	VW Lupo 1,0	37
VW Passat 1,6	74	Seat Toledo 1,6 Stella	74
Fiat Marea 100 16V SX	76	Mazda Xedos 9 2,0 V6	105
BMW 316i	77	Alfa Romeo 166 2,0 TS	114

Tabelle 17

Vergleichen wir nun Mittelwert und Median der Leistung in beiden Gruppen:

	Gruppe A	Gruppe B
Mittelwert	75	75
Median	74	74

Tabelle 18

Wir erkennen sofort, dass in beiden Gruppen sowohl der Mittelwert als auch der Median der **LEISTUNG** gleich groß sind und sich Mittelwert und Median auch kaum voneinander unterscheiden. Trotzdem ist die Zusammensetzung der beiden Gruppen keineswegs gleich, was die **LEISTUNG** betrifft: In Gruppe A befinden sich fünf annähernd gleich starke Mittelklasseautos, während Gruppe B aus zwei kleineren, einem Mittelklasse- und zwei größeren Autos besteht.

Der Unterschied liegt hier in der *Streuung* um den Mittelwert. Die schwächeren Autos in Gruppe B werden im Mittel durch die stärkeren wieder ausgeglichen, an der Streuung erkennt man jedoch den Unterschied (Abbildung 76).

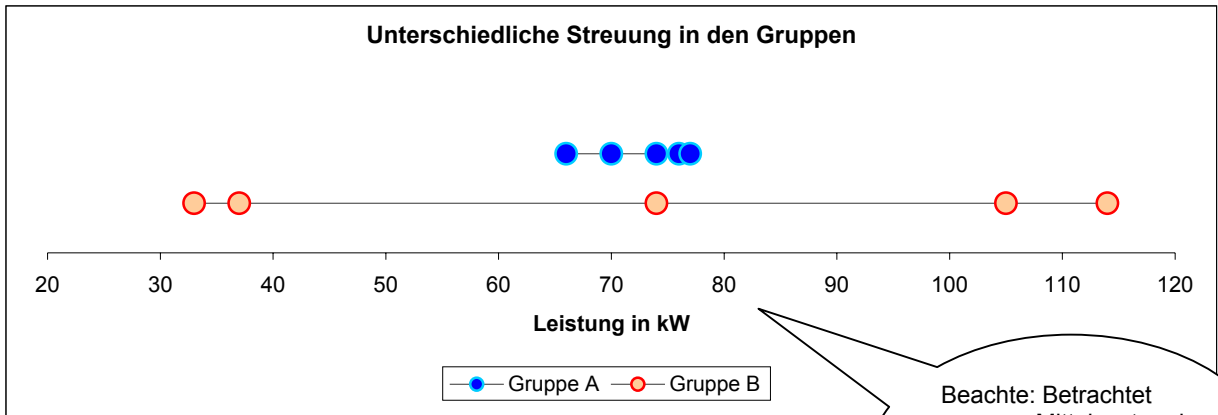


Abbildung 76

Beachte: Betrachtet man nur Mittelwert und Median, erkennt man zwischen den beiden Gruppen keinen Unterschied!

7.1 Maßzahlen der Streuung

Um die Streuung zu messen, gibt es mehrere gebräuchliche Kennzahlen. Vier häufig verwendete werden hier vorgestellt.

7.2 Spannweite

R von engl.: range

$R = x_{(n)} - x_{(1)}$

Die Spannweite ist die Differenz zwischen Maximum und Minimum einer Beobachtungsreihe.

Die Spannweite ist das einfachste Maß der Streuung. Sie berechnet sich lediglich aus der Differenz zwischen Maximum und Minimum und ist demnach nicht immer besonders aussagekräftig.

In unserem Beispiel genügt sie jedoch bereits, um auf den Unterschied aufmerksam zu machen.

	Gruppe A	Gruppe B
Minimum	66	114
Maximum	77	33
Spannweite	11	81

Tabelle 19

Eine größere Spannweite bei gleichem Mittelwert und Median in den Gruppen bedeutet eine größere Streuung.

Zum gesamten PKW-Datensatz zurückgekehrt, wollen wir nun die Spannweite des **HUBRAUMS** berechnen. Dazu benötigen wir das Minimum und das Maximum. Die Differenz ergibt dann die Spannweite.

The screenshot shows an Excel spreadsheet with the following data:

	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie
96	Volvo S80 2,4	4	5	482	1.489	2.435	103	205	11	6	1	1	1	9	455.466	L
97	VW Beetle 1,6	3	4	409	1.205	1.595	74	178	12	4	1	0	0	8	232.050	C
98	VW Bora 1,4	4	5	438	1.149	1.390	55	171	15	4	1	0	0	7	223.750	M
99	VW Golf 1,4	3	5	415	1.146	1.390	55	171	14	4	1	0	0	6	203.990	C
100	VW Lupo 1,0	3	4	353	830	999	37	152	18	2	0	0	0	6	134.590	K
101	VW Passat 1,6	4	5	468	1.250	1.595	74	192	13	4	1	0	0	8	276.450	M
102	VW Polo 1,0	3	5	374	982	999	37	151	18	2	1	0	0	6	149.900	K
103																
104																
105						Minimum	599									
106						Maximum	6.750									
107																
108						Spannweite	6.151									
109																
110																
111																
112																

Abbildung 77

7.3 Quartilsabstand

$$\tilde{x}_{0,75} - \tilde{x}_{0,25}$$

Der Quartilsabstand ist die Differenz zwischen drittem und erstem Quartil.

Die Spannweite hängt bloß von zwei Werten ab, dem kleinsten und dem größten. Falls nun einer dieser beiden Werte ein Ausreißer ist, wird der Wert der Spannweite durch diesen Ausreißer stark beeinflusst.

Aus diesem Grund wird gerne ein ähnliches Streuungsmaß verwendet, das nicht so stark von einem einzelnen Wert abhängig ist - der Quartilsabstand. Dieser Abstand gibt die Differenz zwischen 1. und 3. Quartil an. Innerhalb dieses Quartilsabstandes liegen 50% der Werte (ein Viertel liegt unterhalb des 1. Quartils, ein Viertel oberhalb des 3. Quartils, bleiben zwei Viertel dazwischen).

Wir bleiben beim Hubraum und wollen dessen Quartilsabstand berechnen. Dazu benötigen wir zunächst das 1. und 3. Quartil. Wie diese zu berechnen sind, wissen wir bereits. Die Differenz zwischen 1. und 3. Quartil ergibt sich dann zum Quartilsabstand.

Microsoft Excel - Datenanalyse mit Excel																	
U120																	
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLP	Verbrauch	Preis	Kategorie	
96	Volvo S80 2,4	4	5	482	1.489	2.435	103	205	11	6	1	1	1	9	455.466	L	
97	VW Beetle 1,6	3	4	409	1.205	1.595	74	178	12	4	1	0	0	8	232.050	C	
98	VW Bora 1,4	4	5	438	1.149	1.390	55	171	15	4	1	0	0	7	223.750	M	
99	VW Golf 1,4	3	5	415	1.146	1.390	55	171	14	4	1	0	0	6	203.990	C	
100	VW Lupo 1,0	3	4	353	830	999	37	152	18	2	0	0	0	6	134.590	K	
101	VW Passat 1,6	4	5	468	1.250	1.595	74	192	13	4	1	0	0	8	276.450	M	
102	VW Polo 1,0	3	5	374	982	999	37	151	18	2	1	0	0	6	149.900	K	
103																	
104																	
105							1.Quartil	1.324									
106							3.Quartil	1.995									
107																	
108							Quartilsabstand	671									
109																	
110																	
111																	
112																	

Abbildung 78

Quartilsabstand ⇔ Spannweite

Der Vergleich des Quartilsabstandes mit der Spannweite läßt einige interessante Aufschlüsse über die Verteilung der Daten zu. Als Beispiel betrachten wir wieder den **HUBRAUM** der Autos. Innerhalb des Bereichs der Spannweite 6.151 ccm liegen alle Autos, innerhalb des Quartilsabstandes von 671,25 ccm die Hälfte. Das heißt aber auch, dass sich die andere Hälfte auf einen Bereich von rund 5.500 ccm verteilt. Es muss also einige Werte geben, die weit entfernt vom Median liegen.

Die beste Darstellung von Spannweite und Quartilsabstand ist übrigens ebenfalls durch ein Kastenschaubild (Abbildung 79) möglich!

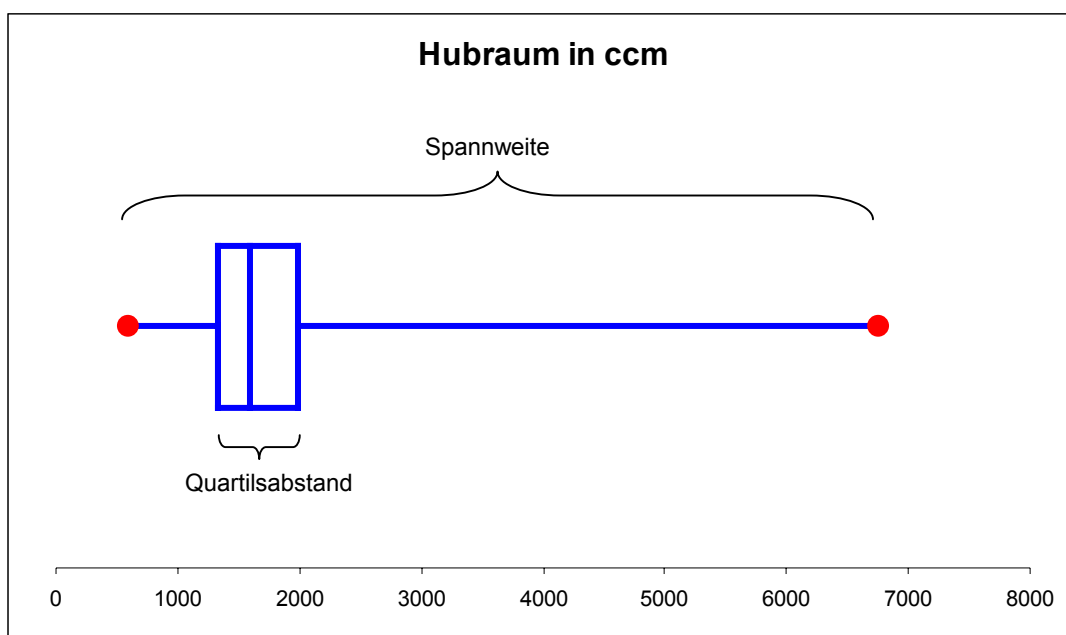


Abbildung 79



22. Berechne die Spannweite der Variablen Verbrauch!

23. Berechne Spannweite und Quartilsabstand der Variablen Preis und Leistung! Welche Aussagen kannst du mit diesen Maßzahlen über die Verteilung des Preises bzw. der Leistung machen? Zeichne ein Kastenschaubild!

7.4 Standardabweichung und Varianz

Die bei weitem am häufigsten gebräuchlichen Maßzahlen für die Streuung sind die Standardabweichung und die Varianz.

7.5 Varianz

$$s_{\bar{x}}^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

Die Varianz ist die mittlere quadratische Abweichung der einzelnen Beobachtungen vom arithmetischen Mittel.

Wie funktioniert die Berechnung der Varianz im Detail? Betrachten wir die Formel einmal näher:

- 1) $x_i - \bar{x}$: Von einer Beobachtung wird der Mittelwert abgezogen. Da der Mittelwert in der Mitte aller Daten liegt, ist dieser natürlich ein gutes Maß für den Abstand der einzelnen Beobachtung von der Mitte.
- 2) $(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots$: Dieser Abstand wird für alle Beobachtungen berechnet. Bevor die Abstände der einzelnen Beobachtungen vom Mittelwert addiert werden, werden sie noch quadriert. Warum? Nun, der Mittelwert ist gerade so definiert, dass er genau in der Mitte aller Daten liegt. Berechnet man bloß die Abstände aller Werte vom Mittelwert, heben sich alle positiven und negativen Abstände auf und man würde für die Summe aller Abstände immer Null erhalten.
Um diesem Problem zu entgehen, quadriert man zunächst alle Abstände. Nun hat man ausschließlich positive (quadratische) Abstände, die man addieren kann, ohne dass sie einander gegenseitig aufheben.
- 3) Die Summe aller quadrierten Abstände werden nun noch durch n , also die Anzahl der Beobachtungen dividiert. So erhält man den mittleren quadratischen Abstand, die Varianz.

VARIANZ

Will man die Varianz mit Hilfe von Excel berechnen, kann man sich grundsätzlich an obige Formel halten und diese entsprechend herleiten.

Als Beispiel soll die Berechnung der Varianz anhand der Variablen **HUBRAUM** des PKW-Datensatzes demonstriert werden.

	AR	AS	AT	AU
1	x_i (Hubraum)			
91	BMW 728i	2793		
92	Mercedes SL 280	2799		
93	Mercedes S 320	3199		
94	Jaguar XJ8 3,2 Exec.	3248		
95	Jaguar XK8 Coupé	3996		
96	BMW Z8	4941		
97	Mercedes CL 500	4966		
98	Rolls Royce Silver Seraph	5379		
99	Ferrari 456 M GT	5474		
100	Chevrolet Corvette	5666		
101	Bentley Arnage Red Label	6750		
102				
103	Arithmetisches Mittel	1.886,50		
104				
105				
106				

Abbildung 80

- 1) Zu Beginn wird das arithmetische Mittel des **HUBRAUMS** berechnet (Abbildung 80). Dieses beträgt 1.886,5 ccm.
- 2) Von jedem einzelnen Wert muss nun das arithmetische Mittel abgezogen werden.

Hierbei erspart das Verwenden von relativen bzw. absoluten Zellbezügen eine Menge Arbeit. Diese Zellbezüge werden im Anhang 2 näher erläutert.

	AR	AS	AT	AU
1	x_i (Hubraum)		$x_i - \bar{x}$	
91	BMW 728i	2793		
92	Mercedes SL 280	2799		
93	Mercedes S 320	3199		
94	Jaguar XJ8 3,2 Exec.	3248		
95	Jaguar XK8 Coupé	3996		
96	BMW Z8	4941		
97	Mercedes CL 500	4966		
98	Rolls Royce Silver Seraph	5379		
99	Ferrari 456 M GT	5474		
100	Chevrolet Corvette	5666		
101	Bentley Arnage Red Label	6750	=AS101-\$AS\$103	
102				
103	Arithmetisches Mittel	1.886,50		
104				
105				
106				

Wenn man die Formel so eingibt, kann man sie nachher leicht vervielfältigen!

Abbildung 81

	AR	AS	AT	AU
1		x_i (Hubraum)	$x_i - \bar{x}$	
91	BMW 728i	2793	906,5	
92	Mercedes SL 280	2799	912,5	
93	Mercedes S 320	3199	1.312,5	
94	Jaguar XJ8 3,2 Exec.	3248	1.361,5	
95	Jaguar XK8 Coupé	3996	2.109,5	
96	BMW Z8	4941	3.054,5	
97	Mercedes CL 500	4966	3.079,5	
98	Rolls Royce Silver Seraph	5379	3.492,5	
99	Ferrari 456 M GT	5474	3.587,5	
100	Chevrolet Corvette	5666	3.779,5	
101	Bentley Arnage Red Label	6750	4.863,5	
102				
103	Arithmetisches Mittel	1.886,50		
104				
105				
106				

Abbildung 82

3) Als nächster Schritt wird jede einzelne dieser Differenzen quadriert (wir arbeiten wieder mit relativen Zellbezügen; siehe Abbildung 83 und Abbildung 84).

	AR	AS	AT	AU
1		x_i (Hubraum)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
91	BMW 728i	2793	906,5	
92	Mercedes SL 280	2799	912,5	
93	Mercedes S 320	3199	1.312,5	
94	Jaguar XJ8 3,2 Exec.	3248	1.361,5	
95	Jaguar XK8 Coupé	3996	2.109,5	
96	BMW Z8	4941	3.054,5	
97	Mercedes CL 500	4966	3.079,5	
98	Rolls Royce Silver Seraph	5379	3.492,5	
99	Ferrari 456 M GT	5474	3.587,5	
100	Chevrolet Corvette	5666	3.779,5	
101	Bentley Arnage Red Label	6750	4.863,5	=AT101^2
102				
103	Arithmetisches Mittel	1.886,50		
104				
105				
106				

Abbildung 83

	AR	AS	AT	AU
1		x_i (Hubraum)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
91	BMW 728i	2793	906,5	821.742,3
92	Mercedes SL 280	2799	912,5	832.656,3
93	Mercedes S 320	3199	1.312,5	1.722.656,3
94	Jaguar XJ8 3,2 Exec.	3248	1.361,5	1.853.682,3
95	Jaguar XK8 Coupé	3996	2.109,5	4.449.990,3
96	BMW Z8	4941	3.054,5	9.329.970,3
97	Mercedes CL 500	4966	3.079,5	9.483.320,3
98	Rolls Royce Silver Seraph	5379	3.492,5	12.197.556,3
99	Ferrari 456 M GT	5474	3.587,5	12.870.156,3
100	Chevrolet Corvette	5666	3.779,5	14.284.620,3
101	Bentley Arnage Red Label	6750	4.863,5	23.653.632,3
102				
103	Arithmetisches Mittel	1.886,50		
104				
105			Summe	116.033.743,0
106				
107				

Abbildung 84

- 4) Die Summe der quadrierten Abweichungen müssen wir noch durch die Anzahl der Autos (=100) dividieren. Für die Varianz erhält man dann

$$s_{\bar{x}}^2 = \frac{116.033.743}{100} = 1.160.337,43$$

Excel macht es uns aber noch einfacher: Man kann die Varianz auch mit einer Formel berechnen!

=VARIANZEN(DATENBEREICH)

Beachte, dass die Formel *Varianzen* heißt. Es gibt nämlich auch die Formel *Varianz*, die zu einem etwas anderen Ergebnis führt.

Microsoft Excel - zu Regression&Dame

	AR	AS	AT
1		x_i (Hubraum)	
91	BMW 728i	2793	
92	Mercedes SL 280	2799	
93	Mercedes S 320	3199	
94	Jaguar XJ6 3,2 Exec.	3248	
95	Jaguar XK8 Coupé	3996	
96	BMW Z8	4941	
97	Mercedes CL 500	4966	
98	Rolls Royce Silver Seraph	5379	
99	Ferrari 456 M GT	5474	
100	Chevrolet Corvette	5666	
101	Bentley Amage Red Label	6750	
102			
103		=varianzen(AS2:AS101)	

Abbildung 85

Microsoft Excel - zu Regression&Dame

	AR	AS	AT
1		x_i (Hubraum)	
91	BMW 728i	2793	
92	Mercedes SL 280	2799	
93	Mercedes S 320	3199	
94	Jaguar XJ6 3,2 Exec.	3248	
95	Jaguar XK8 Coupé	3996	
96	BMW Z8	4941	
97	Mercedes CL 500	4966	
98	Rolls Royce Silver Seraph	5379	
99	Ferrari 456 M GT	5474	
100	Chevrolet Corvette	5666	
101	Bentley Amage Red Label	6750	
102			
103	Varianz	1.160.337,43	

Abbildung 86

7.6 Standardabweichung

$$s_{\bar{x}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Die Standardabweichung ist die Wurzel der Varianz.

Wie wir gesehen haben, ist die Varianz ein logisch hergeleitetes Maß für die Streuung. Ein kleines Problem ergibt sich allerdings mit der Interpretation. Nehmen wir dazu das obige Beispiel: die Varianz des **HUBRAUMS** ist gleich 1.160.337,43 ccm². Nun sind quadrierte Kubikzentimeter ebenso schwer vorstellbar wie etwa Sekunden oder manch andere Einheiten zum Quadrat. Aus diesem Grund hat sich eine zweite Maßzahl neben der Varianz etabliert, die dieser sehr ähnlich ist: die Standardabweichung.

Die Standardabweichung ist nämlich nichts anderes als die Wurzel der Varianz. Sie besitzt den Vorteil, dass sie in derselben Dimension wie die Beobachtungen gemessen wird und somit oft anschaulicher ist.

Die Standardabweichung des **HUBRAUMS** beträgt ungefähr 1.077 ccm.

	AR	AS	AT
1	x_i (Hubraum)		
91	BMW 728i	2793	
92	Mercedes SL 280	2799	
93	Mercedes S 320	3199	
94	Jaguar XJ8 3,2 Exec.	3248	
95	Jaguar XK8 Coupé	3996	
96	BMW Z8	4941	
97	Mercedes CL 500	4966	
98	Rolls Royce Silver Seraph	5379	
99	Ferrari 456 M GT	5474	
100	Chevrolet Corvette	5666	
101	Bentley Arnage Red Label	6750	
102			
103	Varianz	1.160.337,43	
104	Standardabweichung	1.077,19	
105			
106			

Abbildung 87

STANDARDABWEICHUNG

Die Formel für die Berechnung der Standardabweichung in Excel lautet

=STABWN(DATENBEREICH)

Beachte wie bei der Varianz die genaue Eingabe! Es gibt auch hier Formeln, die sehr ähnlich lauten!

	AR	AS	AT
1	x_i (Hubraum)		
91	BMW 728i	2793	
92	Mercedes SL 280	2799	
93	Mercedes S 320	3199	
94	Jaguar XJ8 3,2 Exec.	3248	
95	Jaguar XK8 Coupé	3996	
96	BMW Z8	4941	
97	Mercedes CL 500	4966	
98	Rolls Royce Silver Seraph	5379	
99	Ferrari 456 M GT	5474	
100	Chevrolet Corvette	5666	
101	Bentley Arnage Red Label	6750	
102			
103		=stabwn(AS2:AS101)	
104			
105			
106			

Abbildung 88

Abbildung 89



24. Berechne

- Varianz und Standardabweichung der Variablen Verbrauch!
- Varianz und Standardabweichung der Variablen Leistung auf zwei verschiedene Arten!

8 Erweiterungen

8.1 Pivot-Tabellen und Pivot-Charts

Für etwas fortgeschrittenere Anwender bietet Excel ein interaktives Werkzeug, das zum schnellen Zusammenfassen großer Datenmengen besonders gut geeignet ist - die Pivot-Tabelle.

Eine Pivottable wird verwendet, wenn man zueinander in Beziehung stehende Summen vergleichen möchte, insbesondere wenn lange Zahlenreihen zusammengefasst und verschiedene Fakten zu jedem Wert verglichen werden sollen. Da Pivottabellen interaktiv sind, kann die Ansicht der Daten so geändert werden, dass mehrere Details angezeigt oder verschiedene Zusammenfassungen berechnet werden.

Mithilfe der Pivottable können nicht nur zweidimensionale Tabellen, sondern auch drei-, vier oder höherdimensionale Datenmatrizen erzeugt werden.

Als Beispiel möchten wir mithilfe der Pivot-Tabelle etwa die Autos mit bzw. ohne **AIRBAGS** bzw. **ABS** bezüglich mehrerer Merkmale (etwa **VERBRAUCH** und **PREIS**) untersuchen.

ANLEGEN EINER PIVOT-TABELLE

- 1) Zunächst muss eine Zelle des zu untersuchenden Datenbereiches markiert sein (z.B. A2).
- 2) Der Befehl zum Erstellen einer Pivot-Tabelle lautet **DATEN – PIVOTTABLE- UND PIVOTCHART-BERICHT**.

The screenshot shows the Microsoft Excel interface with a data table of cars. The 'Daten' menu is open, and the 'PivotTable- und PivotChart-Bericht...' option is highlighted. The data table is as follows:

	A	B	C	D	E		M	N	O	P	Q						
	PKW-Name	Türen	Plätze	Länge	Gewicht	Summenmenge	PLUP	Verbrauch	Preis	Kategorie							
2	Alfa Romeo 1.6T Spark	4	5	443	1.2	0	1	8,2	294.000	M							
3	Alfa Romeo 145 1,4 16V	3	5	409	1.1	0	0	7,9	212.000	C							
4	Alfa Romeo 166 2,0 TS	4	5	472	1.4	1	1	9,7	452.000	L							
5	Alfa Romeo GTV 2,0 TS	2	2	429	1.3	0	1	9,2	391.000	S							
6	Audi A3 1,6 Attraction	3	5	415	1.0	1	0	7,6	275.640	C							
7	Audi A4 1,6	4	5	448	1.2	1	0	7,9	328.580	M							
8	Audi A6 1,8	4	5	480	1.3	1	1	8,5	441.570	L							
9	Audi A8 2,8	4	5	503	1.5	1	1	9,9	775.700	L							
10	Audi TT 1,8T	3	2	404	1.2	1	1	8	474.700	S							
11	Bentley Arnage Red Label	4	5	539	2.520	6.750	298	249	6,3	2	1	1	1	19,2	3.583.400	L	
12	BMW 316i	4	5	447	1.360	1.895	77	200	12,4	6	1	0	1	7,8	339.000	M	
13	BMW 316i compact	3	5	421	1.175	1.895	77	190	11,9	4	1	0	0	7,6	291.000	C	
14	BMW 520i	4	5	478	1.470	1.991	110	220	10,2	6	1	1	1	9	496.000	L	
15	BMW 728i	4	5	512	1.710	2.793	142	238	8,6	10	1	1	1	10	784.000	L	
16	BMW Z3 1,8	2	2	405	1.220	1.895	87	196	10,4	4	1	0	0	7,8	389.000	S	

Abbildung 90

- 3) Es erscheint ein Assistent, der in Schritt 1 bereits die gewünschten Einstellungen vorweist. Wir können also gleich auf **WEITER** klicken.

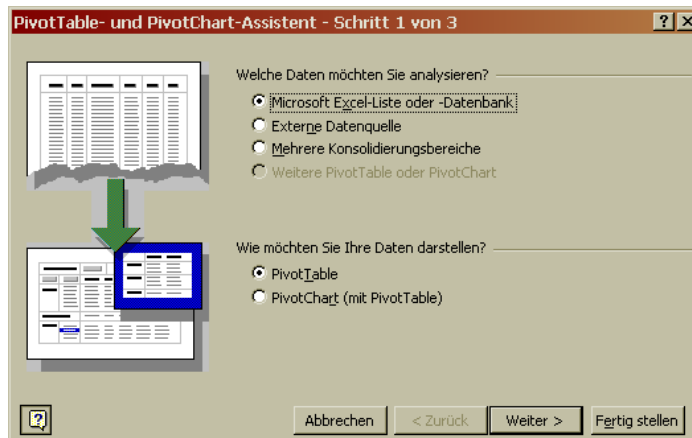


Abbildung 91

- 4) Auch in Schritt 2 und 3 sind bereits die gewünschten Einstellungen vorhanden, sodass wir auf **WEITER** bzw. **FERTIG STELLEN** klicken.

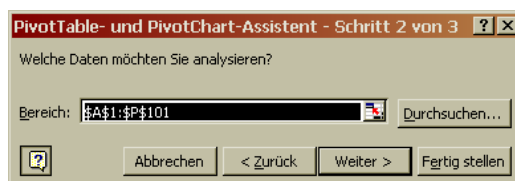


Abbildung 92

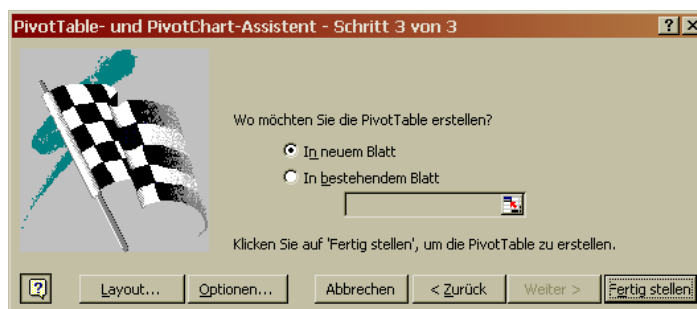


Abbildung 93

- 5) In einem neuen Blatt wird nun eine Pivot-Tabelle erstellt. (Abbildung 94)

Achtung: Entfernt man die Markierung von ihrer Voreinstellung (Zelle A3), verschwinden auch die Optionen aus der Menüleiste. In diesem Fall muss wieder die Zelle (A3) markiert werden.

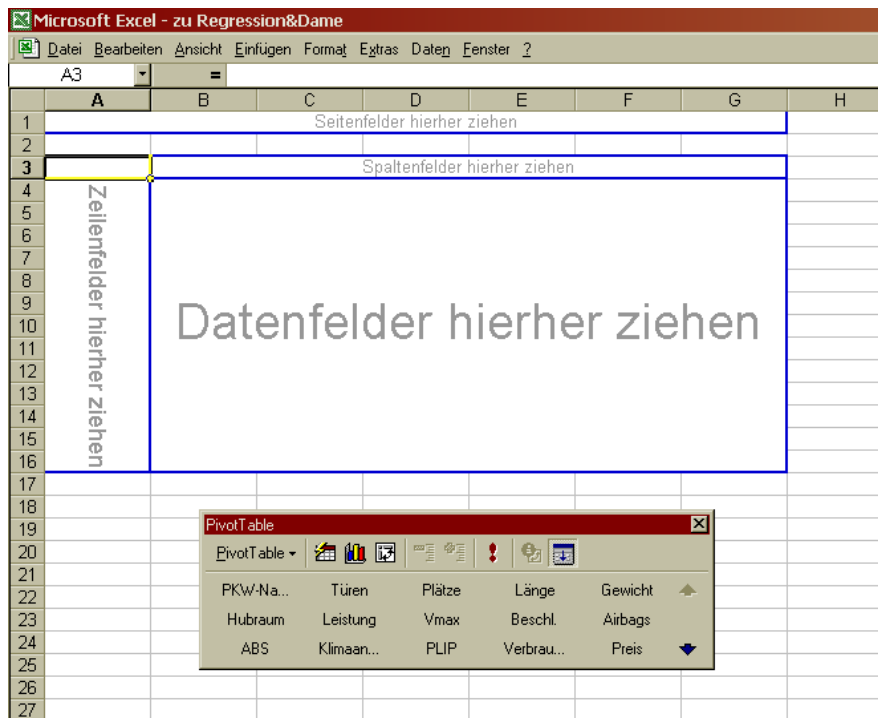


Abbildung 94

ERSTE AUSWERTUNGEN MIT DER PIVOT-TABELLE

Nachdem wir nun eine Pivot-Tabelle erfolgreich angelegt haben, können wir recht einfach eine erste Auswertung vornehmen.

- 1) Zunächst klicken wir auf **AIRBAGS** und ziehen das Kästchen zu den Zeilenfeldern in Spalte A. Danach klicken wir auf **ABS** und ziehen das Kästchen zu den Spaltenfeldern in Zeile 3. Die Pivot-Tabelle sieht nun folgendermaßen aus.

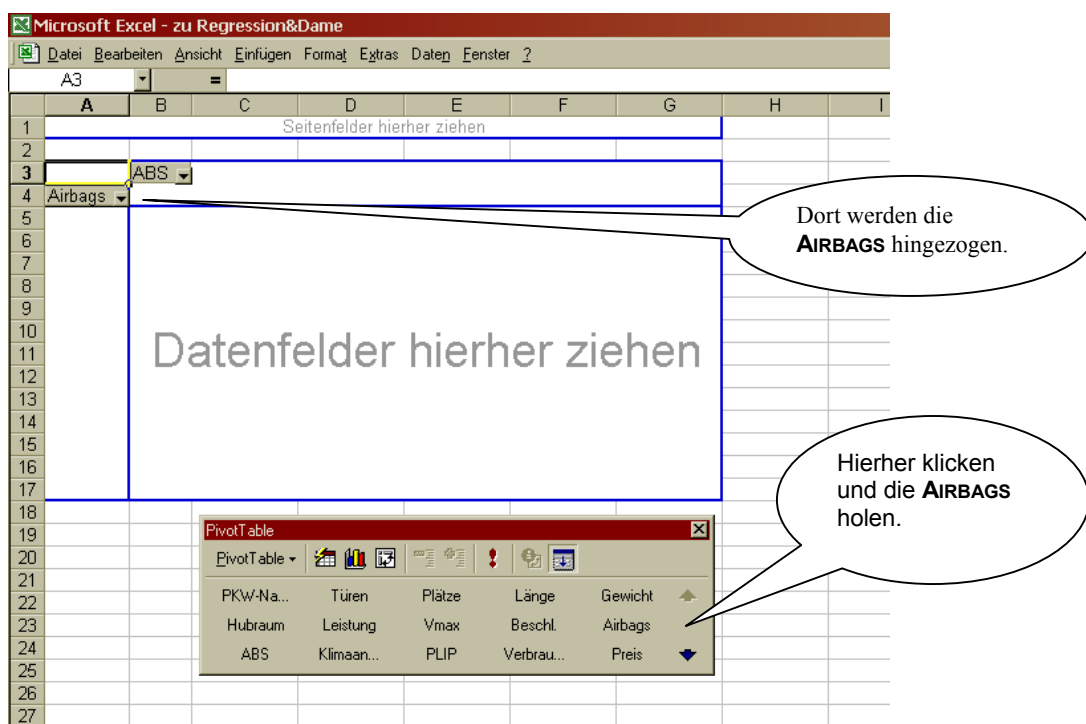


Abbildung 95

- 2) Jetzt fehlt noch das Merkmal, welches eigentlich untersucht werden soll. Wir wollen mit dem **VERBRAUCH** beginnen, klicken also **VERBRAUCH** an und schieben das Kästchen in den großen Kasten mit der Beschriftung „Datenfelder hierher ziehen“. Das Ergebnis ist in Abbildung 96 abgebildet.

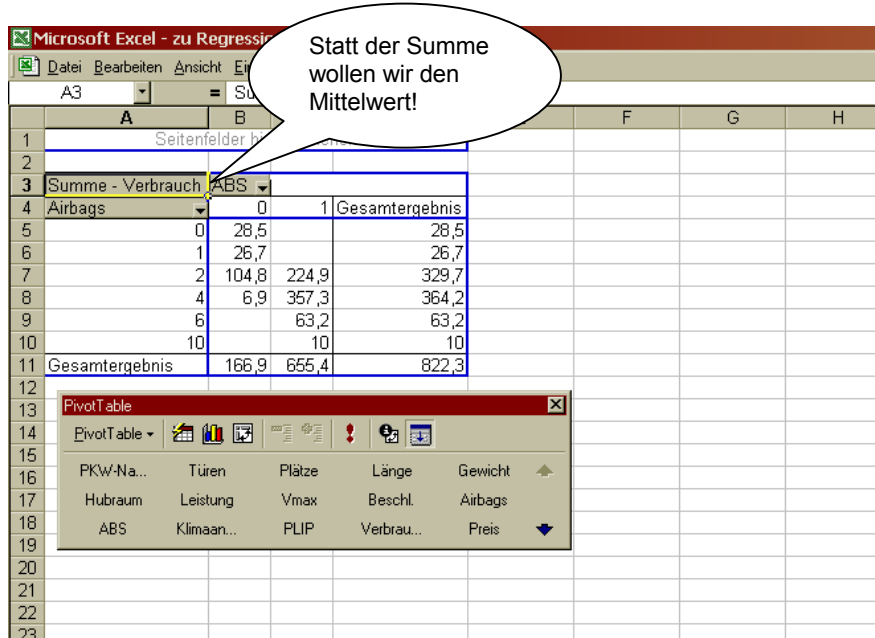


Abbildung 96

- 3) Zufriedenstellend ist das Ergebnis noch nicht ganz! Statt der Summe des **VERBRAUCHS** wäre zum Beispiel das arithmetische Mittel interessanter, um etwa zu vergleichen, ob Autos mit oder ohne **ABS** durchschnittlich mehr Treibstoff verbrauchen. Deshalb machen wir einen Rechtsklick auf die Zelle **SUMME-VERBRAUCH** (A3) und wählen im Menü den Punkt **FELDEIGENSCHAFTEN** (Abbildung 97). Es erscheint ein Fenster, das sogenannte **PIVOTTABLE – FELD** (Abbildung 98).

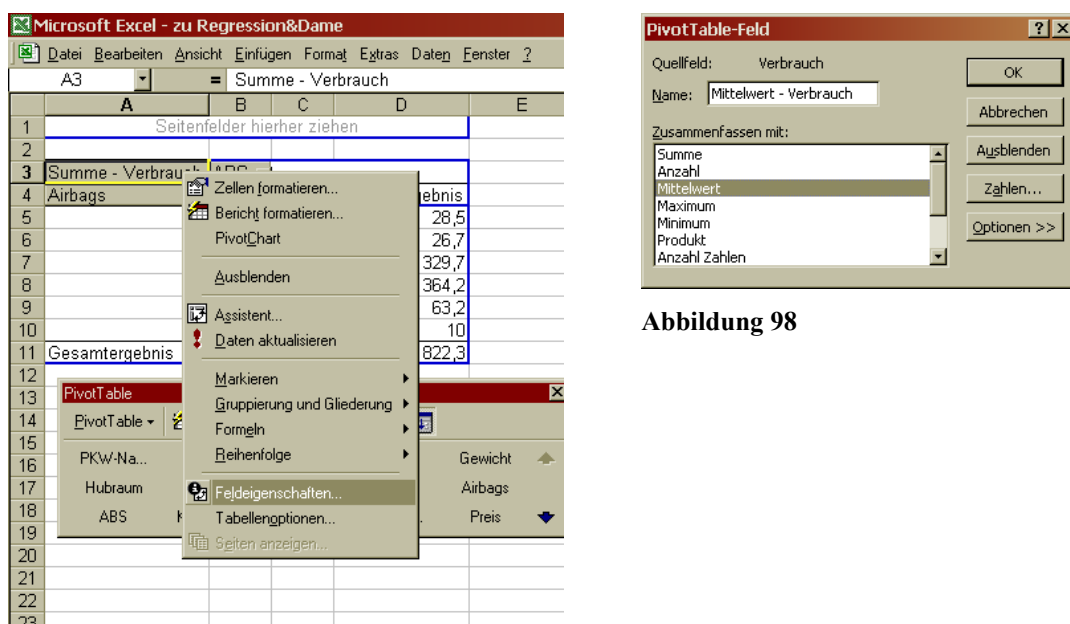


Abbildung 98

Abbildung 97

- 4) Im **PIVOTTABLE-FELD** wählen wir **MITTELWERT** und klicken auf **OK**. Die resultierende Pivot-Tabelle sieht bereits viel informativer aus!

The screenshot shows an Excel PivotTable with the following data:

Airbags	ABS	Gesamtergebnis
0	7,13	7,13
1	6,68	6,68
2	6,55	9,78
4	6,90	8,12
6	9,03	9,03
10	10,00	10,00
Gesamtergebnis	6,68	8,74

The PivotTable also shows a grand total of 8,22 for the overall average consumption.

Abbildung 99

Aus der Pivot-Tabelle in Abbildung 99 kann man verschiedenste interessante Tatsachen herauslesen, zum Beispiel:

- Der durchschnittliche Verbrauch der Autos die **ABS** besitzen beträgt 8,74l/100 km.
- Der durchschnittliche Verbrauch der Autos ohne **ABS** beträgt 6,68l/100 km.
- Die Autos ohne **ABS**, aber mit 2 Airbags sind im Durchschnitt am Sparsamsten. (6,55l/100 km)
- In der Datenmatrix gibt es kein Auto mit keinem oder einem **AIRBAG** und **ABS**. Außerdem haben alle Autos mit 6 oder 10 **AIRBAGS** auch **ABS**.
- Der Mittelwert des **VERBRAUCHS** aller Autos beträgt 8,22l/100 km.

Weitere Auswertungen sind nun besonders einfach. Einige Beispiele werden im folgenden Abschnitt demonstriert.

WEITERFÜHRENDE AUSWERTUNGEN MIT PIVOT-TABELLEN

Wir wollen wissen, wie viele Autos eigentlich für den Mittelwert in den verschiedenen Gruppen verantwortlich sind? Kein Problem!

- 1) Dazu ziehen wir die Variable **VERBRAUCH** nochmals in den großen Kasten (zwischen Zelle B5 und D11). Nun berechnet Excel für jede Gruppe den Mittelwert und die Summe des **VERBRAUCHS**.

- 2) Da die Summe nicht gewünscht ist, klicken wir mit der rechten Maustaste auf **SUMME – VERBRAUCH**. Im folgenden Menü wählen wir wieder Feldeigenschaften.
- 3) Im erscheinenden Fenster **PIVOTTABLE–FELD** wählen wir diesmal **ANZAHL**. Neben dem Mittelwert des **VERBRAUCHS** in den einzelnen Gruppen wird nun auch die Anzahl der Autos angezeigt, die für diesen Mittelwert verantwortlich ist.

The screenshot shows an Excel PivotTable with the following data:

Airbags	Mittelwert - Verbrauch	Anzahl - Verbrauch	Gesamtergebnis
0	7,13	4,00	7,13
1	6,68	4,00	6,68
2	6,55	9,78	8,45
4	6,90	8,12	8,09
6	9,03	7,00	9,03
10	10,00	1,00	10,00
Gesamt: Mittelwert - Verbrauch	6,68	8,74	8,22
Gesamt: Anzahl - Verbrauch	25,00	75,00	100,00

The PivotTable Field List shows: PKW-Na..., Türen, Plätze, Länge, Gewicht.

Abbildung 100

- 4) Jetzt wollen wir noch einen Schritt weiter gehen und wissen, welche Autos mit sechs **AIRBAGS** und **ABS** ausgestattet sind! Ein Doppelklick in Zelle D13 oder D14 (dort befinden sich die Autos mit 6 **AIRBAGS** und **ABS**) genügt: In einem neuen Tabellenblatt werden alle 7 Autos mit 6 **AIRBAGS** und **ABS** ausgegeben!

The screenshot shows a filtered data table with the following columns: PKW-Name, Türen, Plätze, Länge, Gewicht, Hubraum, Leistung, Vmax, Beschl., Airbag, and Ka. A callout bubble points to the filtered data with the text: "Eine besonders einfache Möglichkeit zu filtern!"

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbag	Ka	
Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	8,5	455466 L
Peugeot 607 2.2	4	5	488	1455	2230	116	220	9,6	6	7,2	455000 L
Mercedes S 320	4	5	504	1770	3199	165	240	8,2	6	11,5	1106640 L
Mercedes E 200	4	5	480	1510	1998	100	209	9,3	6	9,3	534535 L
BMW 520i	4	5	478	1470	1991	110	220	10,2	6	9	496000 L
BMW 316i	4	5	447	1360	1895	77	200	12,4	6	7,8	339000 M
Audi A8 2,8	4	5	503	1540	2771	142	236	8,4	6	9,9	775700 L

Abbildung 101

PIVOT-CHARTS

Mit Hilfe eines Pivot-Charts kann man eine Pivot-Tabelle auch grafisch darstellen. Wir wollen zum Beispiel herausfinden, ob Autos mit **KLIMAANLAGE** durchschnittlich teurer sind als Autos ohne **KLIMAANLAGE**. Damit das Ergebnis nicht verzerrt wird, wollen wir

alle **KATEGORIEN** getrennt voneinander untersuchen¹⁰. Das Ergebnis soll grafisch dargestellt werden.

- 1) Die Fragestellung ist natürlich am komfortabelsten mithilfe der Pivot-Tabelle zu lösen. Wir benötigen also **KLIMAANLAGE** als Spaltenfeld, **KATEGORIE** als Zeilenfeld und den **PREIS** im Datenfeld. Die fertige Pivot-Tabelle zeigt Abbildung 102.

Kategorie	Mittelwert - Preis	Gesamtergebnis
C	186744,38	194637,50
K	148277,43	148938,58
L	880000,11	880000,11
M	246812,77	256536,63
S	368570,67	825070,89
Gesamtergebnis	214024,37	439127,50

Abbildung 102

- 2) Zur grafischen Darstellung klickt man mit der rechten Maustaste in die Tabelle und wählt den Menüpunkt **PIVOTCHART** aus.

Kategorie	Mittelwert - Preis	Gesamtergebnis
C	186744,38	194637,50
K	148277,43	148938,58
L	880000,11	880000,11
M	246812,77	256536,63
S	368570,67	825070,89
Gesamtergebnis	214024,37	439127,50

Abbildung 103

¹⁰Es ist klar, dass Limousinen teurer sein werden als Kleinwägen. Das liegt aber nicht nur daran, dass der Anteil an Klimaanlage bei Limousinen größer ist als bei Kleinwägen. Aus diesem Grund betrachten wir alle Kategorien einzeln: Sind z.B. die Kleinwägen mit Klimaanlage teurer als die Kleinwägen ohne Klimaanlage?

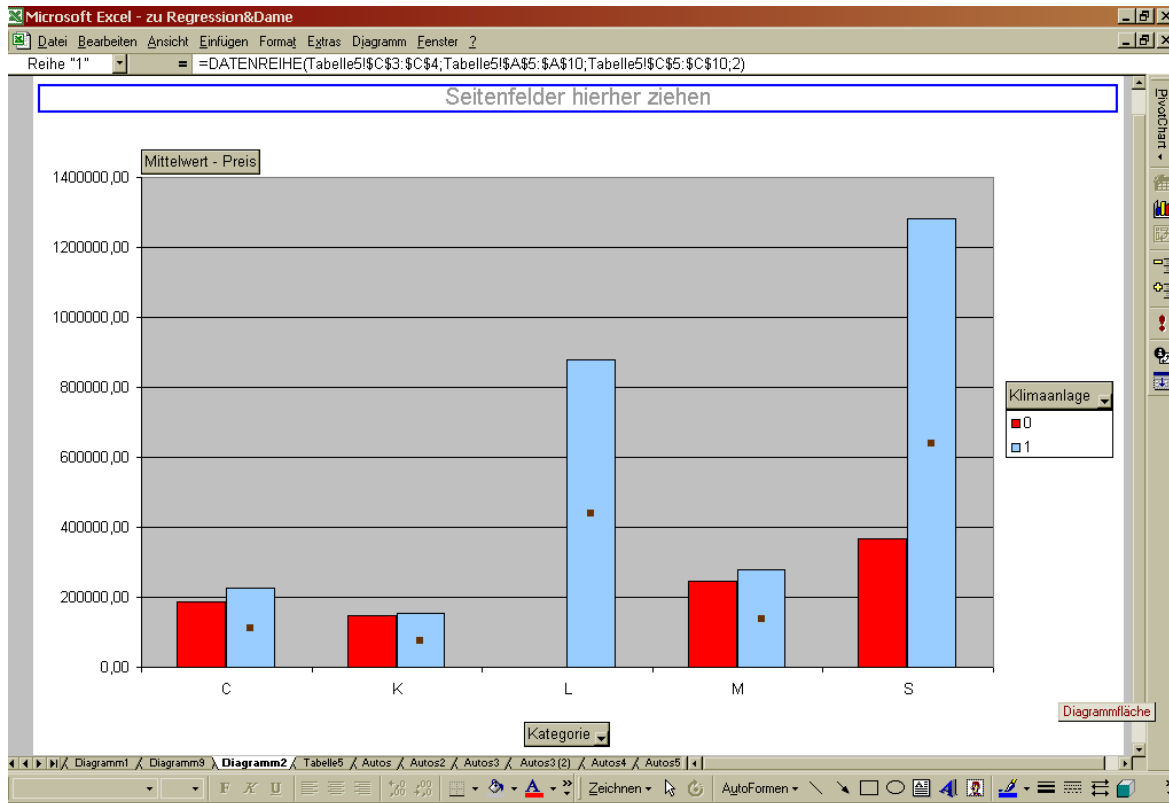


Abbildung 104

3) Nach Wahl des geeigneten Diagrammtyps (Säulen nebeneinander) erscheint eine Grafik, die unsere Frage übersichtlich beantwortet

- In allen **KATEGORIEN** ist der Mittelwert des **PREISES** der Autos mit **KLIMAAANLAGE** höher als jener ohne.
- Ein besonders großer Preisunterschied ist in der **KATEGORIE „S“** zu finden.
(Anmerkung: In dieser **KATEGORIE** wird der Mittelwert von einigen sehr teuren Autos stark beeinflusst.)
- In der Klasse „L“ gibt es keine Autos ohne **KLIMAAANLAGEN**.
- In den Klassen „C“, „K“ und „M“ ist der Preisunterschied geringer.

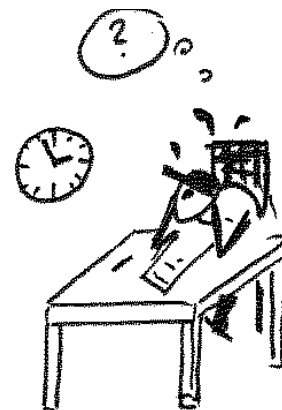
Wir sehen also: stellt man Preisvergleiche zwischen PKWs mit und ohne **KLIMAAANLAGE** an, ist es sinnvoll gleichzeitig zwischen den **KATEGORIEN** zu unterscheiden. Mithilfe der Pivot-Tabelle und des Pivot-Charts funktioniert dies besonders einfach.

8.2 Histogramm mit unterschiedlicher Klassenbreite

Wie bereits in Abschnitt 5.4. diskutiert, dient ein Histogramm zur Darstellung von Daten, die in Klassen zusammengefaßt sind. Es wurde empfohlen, alle Klassen möglichst gleich breit zu wählen. In diesem Fall ist ein Histogramm einem Säulendiagramm sehr ähnlich, anschaulich und einfach herzustellen.

Es gibt aber auch Fälle, in denen die Klassen aus praktischen Gründen nicht alle gleich breit gewählt werden können. In diesem Fall muss bei der Konstruktion eines Histogramms eine spezielle Besonderheit berücksichtigt werden, die im folgenden an einem Beispiel gezeigt wird.

Beispiel: Bei einem Eignungstest kann man maximal 100 Punkte erreichen. Nehmen wir folgendes an: Ein Ergebnis von 91 bis 100 Punkten bedeutet einen ausgezeichneten Erfolg, mit 71-90 Punkten hat man den Test bestanden, zwischen 41 und 70 Punkten darf man den Test nochmals wiederholen und bei 0-40 Punkten hat man den Test nicht bestanden.



In Tabelle 20 ist das Ergebnis des letzten Eignungstests aufgelistet. Wie kann man dieses Ergebnis mit einem Histogramm geeignet darstellen?

	Anzahl der Teilnehmer (= absolute Häufigkeit)	relative Häufigkeit
Nicht bestanden (0-40)	20	0,250
Wiederholen (41-70)	40	0,500
Bestanden (71-90)	10	0,125
Ausgezeichneter Erfolg (91-100)	10	0,125

Tabelle 20

Da gruppierte Daten vorliegen, ist für die Darstellung der Häufigkeiten in den Klassen ein Histogramm am besten geeignet. Gleiche Klassenbreiten sind in diesem Fall allerdings nicht sinnvoll, da die Breiten aus sachlogischen Gründen bereits vorgegeben sind. Aus diesem Grund muß die Darstellung in einem Histogramm mit unterschiedlichen Klassenbreiten erfolgen.

Bei der Konstruktion eines Histogramms mit unterschiedlichen Klassenbreiten sind folgende Punkte zu berücksichtigen:

- Wie bei jedem Histogramm sollten die Klassen aneinander anstossen, aber einander nicht überschneiden.
- Die **Häufigkeiten** werden durch die **Flächen der Rechtecke** beschrieben. Die Fläche der Rechtecke ist also proportional zur Klassenhäufigkeit, nicht die Höhe der Rechtecke.
- Auch die erste oder letzte Klasse darf nicht offen sein.

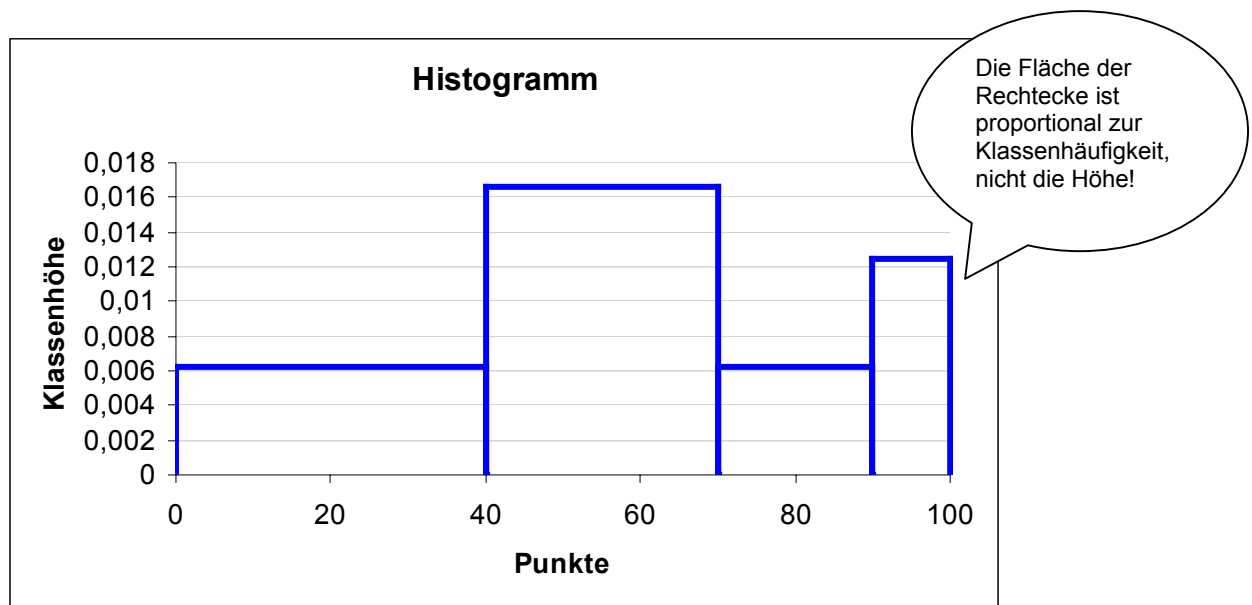


Abbildung 105

Abbildung 105 zeigt ein Histogramm für die Daten aus Tabelle 20. Es verdeutlicht sehr anschaulich die Proportionalität der Flächen zu den Klassenhäufigkeiten:

In der ersten Klasse (0-40 Punkte) befinden sich doppelt so viele Personen wie in der dritten (71-90 Punkte). Das Rechteck der ersten Klasse ist auch genau doppelt so groß wie das der dritten Klasse. Da es sich aber auch über eine doppelt so große Länge erstreckt (40 Punkte gegenüber 20 Punkten) ist es nicht höher als das erste Rechteck!

Sowohl in der dritten (71-90 Punkte) als auch in der vierten Klasse (91-100 Punkte) befinden sich 10 Personen. Die Fläche der beiden Rechtecke ist auch gleich groß. Da sich das Rechteck der vierten Klasse gegenüber dem der dritten nur über die halbe Punktelänge erstreckt, ist es aber doppelt so hoch.

Wir sehen also: bei einem Histogramm mit unterschiedlichen Klassenbreiten kommt es auf die Flächen an!

Konstruktion eines Histogramms mit unterschiedlicher Klassenbreite

1) Will man selbst ein solches Histogramm anfertigen, benötigt man zumindest folgende Bestandteile:

- Klassenbeginn und Klassenende der einzelnen Klassen
- absolute Häufigkeiten der Klassen

Bezeichnung	Klassenbeginn ¹¹	Klassenende	Anzahl der Teilnehmer (= absolute Häufigkeit)
Nicht bestanden	0	40	20
Wiederholen	40	70	40
Bestanden	70	90	10
Ausgezeichneter Erfolg	90	100	10

Tabelle 21

2) Aus den absoluten Klassenhäufigkeiten werden die relativen Klassenhäufigkeiten berechnet. Dazu dividiert man die Anzahl der Elemente in den Klassen durch die Gesamtanzahl der Elemente. In unserem Beispiel gibt es insgesamt 80 Teilnehmer. Die relative Häufigkeit in der ersten Klasse beträgt $20/80 = 0,250$.

Bezeichnung	Klassenbeginn	Klassenende	Anzahl der Teilnehmer (= absolute Häufigkeit)	Relative Häufigkeit
Nicht bestanden	0	40	20	0,250
Wiederholen	40	70	40	0,500
Bestanden	70	90	10	0,125
Ausgezeichneter Erfolg	90	100	10	0,125

Tabelle 22

3) Für die Berechnung der Höhe der Rechtecke muss die relative Klassenhäufigkeit noch auf die Breite der Klasse bezogen werden. Wir dividieren also die relative Klassenhäufigkeit durch die Klassenbreite. In der ersten Klasse ergibt dies eine Höhe von $0,250/40 = 0,0063$.

Bezeichnung	Klassenbeginn	Klassenende	Anzahl der Teilnehmer (= absolute Häufigkeit)	Relative Häufigkeit	Rechteckshöhe
Nicht bestanden	0	40	20	0,250	0,0063
Wiederholen	40	70	40	0,500	0,0167
Bestanden	70	90	10	0,125	0,0063
Ausgezeichneter Erfolg	90	100	10	0,125	0,0125

Tabelle 23

¹¹ Damit die Klassen nahtlos aneinanderstoßen, werden die Intervalle folgendermaßen gewählt: [0; 40], (40; 70], (70; 90], (91, 100] Da es nur ganze Punktzahlen gibt, entspricht dies genau der vorgegebenen Punkteinteilung (0-40; 41-70; 71-90; 91-100)

- 4) Nun hat man alle benötigten Bestandteile berechnet und kann ein Histogramm wie in Abbildung 105 konstruieren.

HISTOGRAMM MIT UNTERSCHIEDLICHER KLASSENBREITE

Excel bietet die oben beschriebene Diagrammerstellung nicht standardmäßig an. Wie bereits demonstriert, kann man sich bei einem Histogramm mit gleichen Klassenbreiten einfach mit einem Säulendiagramm behelfen.

Für unterschiedliche Klassenbreiten wird in der beiliegenden Excel-Datei *Datenanalyse mit Excel.xls* im Tabellenblatt Histogramm eine Möglichkeit angeboten, Histogramme auch mit unterschiedlichen Klassenbreiten zu konstruieren.

Als Eingaben sind der Beginn der ersten Klasse und das Ende aller Klassen sowie die absoluten Klassenhäufigkeiten erforderlich. Mit einem Klick auf Skalieren wird ein entsprechendes Histogramm ausgegeben. Dieses Programm kann selbstverständlich auch Histogramme mit gleicher Klassenbreite ausgeben!

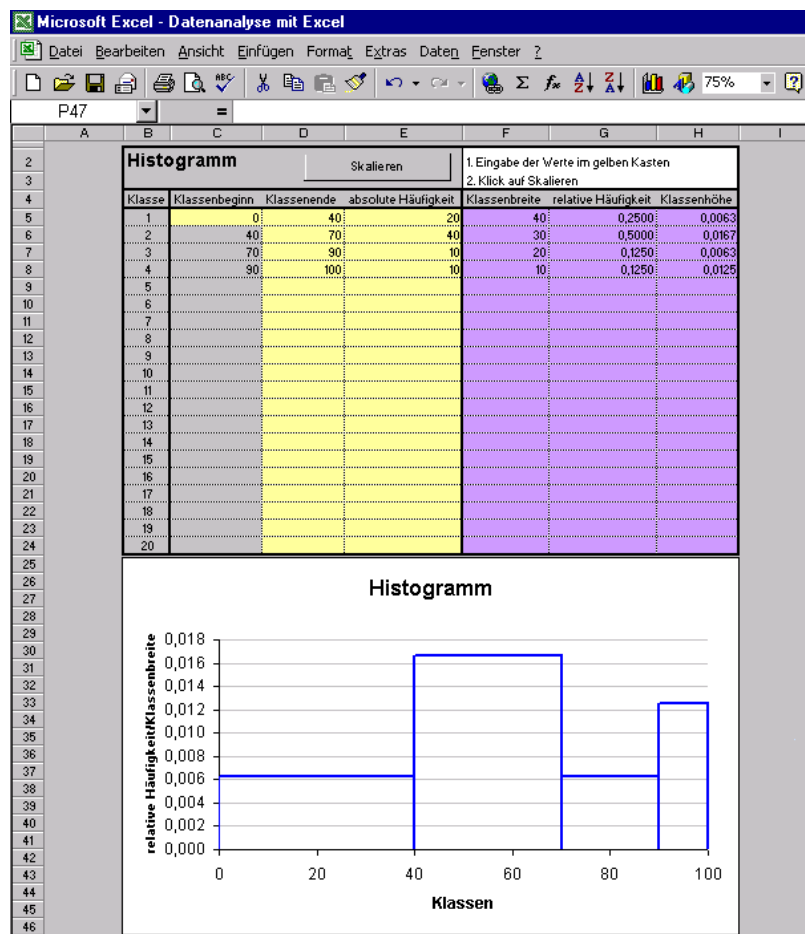


Abbildung 106

WIE EIN EXCEL-HISTOGRAMM ENTSTEHT

Im folgenden wird gezeigt, welche Ideen hinter dem Histogramm aus Abbildung 106 stehen und wie man sich ein solches Diagramm selbst anfertigen kann. Dazu wollen wir wieder die Noten des Testergebnisses heranziehen und folgendes Histogramm erstellen.

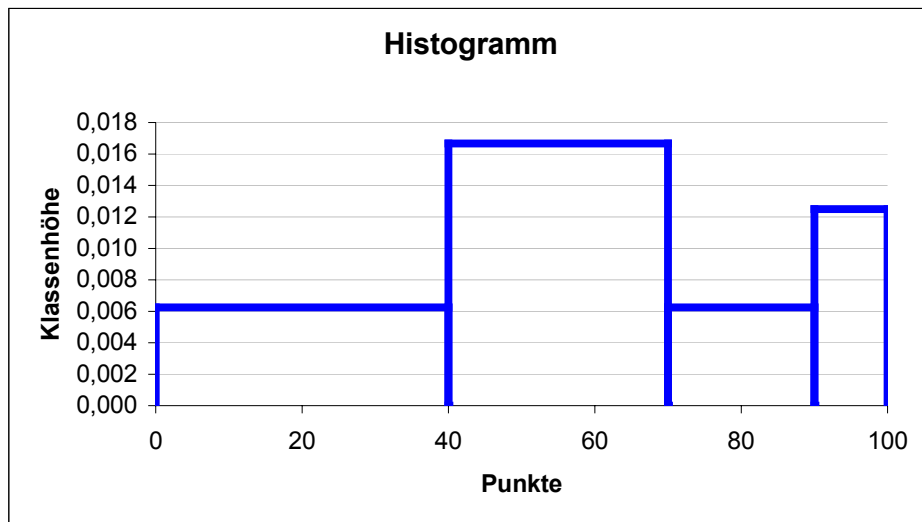


Abbildung 107

- 1) Die erste Idee zur Konstruktion des Histogramms betrifft die Wahl des Diagrammtyps. Mithilfe eines Punkt (XY) – Diagramms kann man einen beliebigen Linienzug (auch ohne Datenpunkte zeichnen).

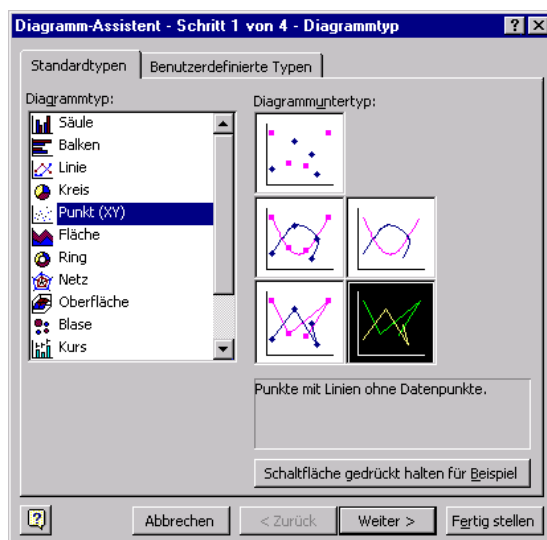


Abbildung 108

- 2) Wenn man die Eckpunkte aller Rechtecke des Histogramms in der richtigen Reihenfolge eingibt, kann man mit diesem Linienzug ein ganzes Histogramm anfertigen. Dazu bezeichnen wir jede Klasse mit einem Buchstaben (a, b, c und d) und jeden Eckpunkt mit einer Ziffer (1-4). Wir benötigen also folgenden Linienzug: 1a,1b, 1c, 1d, 2a, 2b, 2c, 2d, 3a, 3b, 3c, 3d, 4a, 4b, 4c, 4d (Abbildung 109).

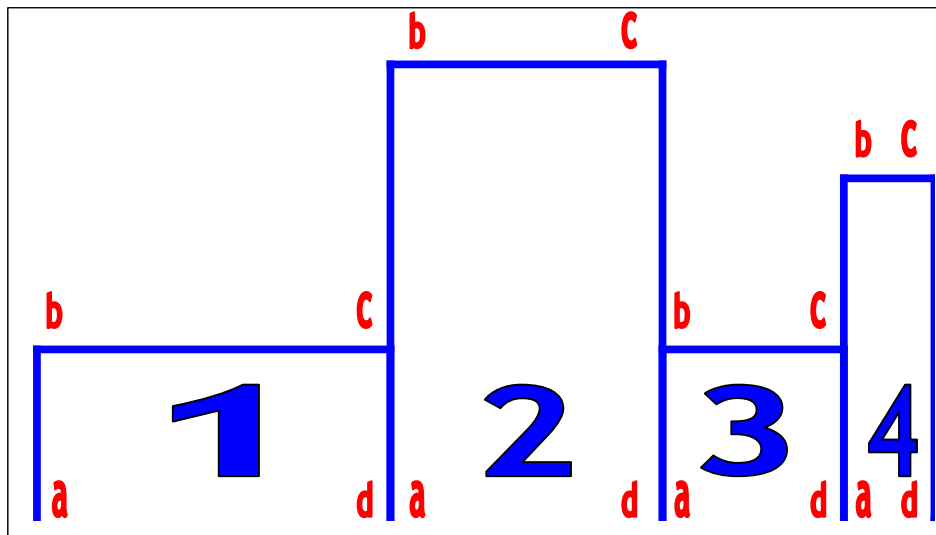


Abbildung 109

3) Die Punkte mit a (also 1a, 2a, 3a und 4a) haben folgende Koordinaten:

X: Klassenbeginn der jeweiligen Klasse

Y: 0

Die Punkte mit b haben folgende Koordinaten:

X: Klassenbeginn der jeweiligen Klasse

Y: Klassenhöhe der jeweiligen Klasse

Die Punkte mit c haben die Koordinaten:

X: Klassenende der jeweiligen Klasse

Y: Klassenhöhe der jeweiligen Klasse

Die Punkte d haben schließlich die Koordinaten:

X: Klassenende der jeweiligen Klasse

Y: 0

4) Aus der in Abbildung 110 dargestellten Eingabemaske wird eine ebenfalls in der Abbildung dargestellte Liste konstruiert. Diese Liste enthält bereits alle benötigten Punkte des Linienzuges (von 1a bis 4d) mit den korrekten Koordinaten. Der einzige Mangel an dieser Liste ist die Sortierung der Punkte (1a, 2a, 3a, 4a, 1b, 2b, etc). Wir benötigen eine Sortierung zunächst nach der Klasse und erst danach nach den Punkten. Dazu werden alle Klassen und Punkte inklusive deren Koordinaten markiert Abbildung 111 und zuerst nach Klasse und anschließend nach Punkt sortiert (Abbildung 112). Als Ergebnis erhält man die in 2) angestrebte Sortierung: 1a, 1b, 1c, 1d, 2a, ... (Abbildung 113).

	J	K	L	M	N	O	P	Q
2								
3								
4		Klasse	Klassenbeginn	Klassenende	absolute Häufigkeit	Klassenbreite	relative Häufigkeit	Klassenhöhe
5		1	0	40	20	40	0,2500	0,0063
6		2	40	70	40	30	0,5000	0,0167
7		3	70	90	10	20	0,1250	0,0063
8		4	90	100	10	10	0,1250	0,0125
9								
10								
11		Klasse	Punkt	X-Koordinate	Y-Koordinate			
12		1	a	0	0			
13		2	a	40	0			
14		3	a	70	0			
15		4	a	90	0			
16		1	b	0	0,0063			
17		2	b	40	0,0167			
18		3	b	70	0,0063			
19		4	b	90	0,0125			
20		1	c	40	0,0063			
21		2	c	70	0,0167			
22		3	c	90	0,0063			
23		4	c	100	0,0125			
24		1	d	40	0			
25		2	d	70	0			
26		3	d	90	0			
27		4	d	100	0			
28								

Abbildung 110

	Klasse	Klassenbeginn	Klassenende	absolute Häufigkeit	Klassenbreite	relative Häufigkeit	Klassenhöhe
4							
5	1	0	40	20	40	0,2500	0,0063
6	2	40	70	40	30	0,5000	0,0167
7	3	70	90	10	20	0,1250	0,0063
8	4	90	100	10	10	0,1250	0,0125
9							
10							
11	Klasse	Punkt	X-Koordinate	Y-Koordinate			
12	1	a	0	0			
13	2	a	40	0			
14	3	a	70	0			
15	4	a	90	0			
16	1	b	0	0,0063			
17	2	b	40	0,0167			
18	3	b	70	0,0063			
19	4	b	90	0,0125			
20	1	c	40	0,0063			
21	2	c	70	0,0167			
22	3	c	90	0,0063			
23	4	c	100	0,0125			
24	1	d	40	0			
25	2	d	70	0			
26	3	d	90	0			
27	4	d	100	0			
28							

Abbildung 111

Sortieren

Sortieren nach
 Klasse Aufsteigend Absteigend

Anschließend nach
 Punkt Aufsteigend Absteigend

Zuletzt nach
 Aufsteigend Absteigend

Liste enthält
 Überschrift keine Überschrift

Optionen... OK Abbrechen

Abbildung 112

	J	K	L	M	N	O	P	Q
2								
3								
4		Klasse	Klassenbeginn	Klassenende	absolute Häufigkeit	Klassenbreite	relative Häufigkeit	Klassenhöhe
5		1	0	40	20	40	0,2500	0,0063
6		2	40	70	40	30	0,5000	0,0167
7		3	70	90	10	20	0,1250	0,0063
8		4	90	100	10	10	0,1250	0,0125
9								
10								
11		Klasse	Punkt	X-Koordinate	Y-Koordinate			
12		1	a	0	0			
13		1	b	0	0,0063			
14		1	c	40	0,0063			
15		1	d	40	0			
16		2	a	40	0			
17		2	b	40	0,0167			
18		2	c	70	0,0167			
19		2	d	70	0			
20		3	a	70	0			
21		3	b	70	0,0063			
22		3	c	90	0,0063			
23		3	d	90	0			
24		4	a	90	0			
25		4	b	90	0,0125			
26		4	c	100	0,0125			
27		4	d	100	0			
28								

Abbildung 113

5) Nun können die X- und Y-Koordinaten markiert und in einem Diagramm (Abbildung 108) dargestellt werden. Mit etwas grafischer Umgestaltung erhält man als Ergebnis das Diagramm aus Abbildung 107.

Anmerkung: Hinter dem Button „Skalieren“ in der Eingabemaske verbirgt sich ein Makro, das lediglich dafür sorgt, dass das Histogramm in die Mitte des Diagramms gerückt wird. Zur eigentlichen Konstruktion des Histogramms, wie sie hier dargestellt wurde, ist kein Makro notwendig!

9 Lösungsteil

1) a)

- *Datenbereich markieren*
- *Anwählen von: Daten – Sortieren*
- *Sortieren nach: Leistung*
- *OK → Daten werden nach Leistung sortiert.*
- *Das erste angezeigte Auto hat die geringste Leistung (Fiat Seicento Young, 29 kW).*

b)

- *wie a)*
- *Das letzte angezeigte Auto, also der Ferrari 456 M GT mit 325 kW.*

c)

- *wie a)*
- *Da die PKWs nach dem Merkmal Leistung sortiert sind, muss man alle Autos zählen bis man auf das erste mit einer Leistung von 45 kW oder mehr stößt. 20 Autos haben eine Leistung von weniger als 45 kW.*

2) a)

- *Datenbereich markieren*
- *Anwählen von: Daten – Sortieren*
- *Sortieren nach: Türen; anschließend nach: Preis*
- *OK → Daten werden zunächst nach Anzahl der Türen und innerhalb der Gruppen nach dem Preis sortiert.*
- *Unter allen PKW mit 5 Türen ist der Lada Samara 1,5i mit einem Preis von 99.990,- am billigsten.*

b)

- *wie a)*
- *Unter allen PKW mit 2 Türen ist der Ferrari 456 M GT mit einem Preis von 3.175.000,- am teuersten.*

c)

- *Datenbereich markieren*
- *Anwählen von: Daten – Sortieren*
- *Sortieren nach: Türen*

- *In der Datenmatrix gibt es 32 Viertürer.*
- 3)
- *Datenbereich markieren*
 - *Anwählen von: Daten – Sortieren*
 - *Sortieren nach: Airbags; anschließend nach: Preis*
 - *Das teuerste Auto ohne Airbags ist der Fiat Barchetta.*
- 4)
- *Markierung muss sich irgendwo in der Datenmatrix befinden*
 - *Anwählen von: Daten – Filter – AutoFilter*
 - *Anklicken des Filterkästchens Klimaanlage*
 - *Auswählen von „1“ → Es werden nur mehr die Autos mit Klimaanlage angezeigt.*
- 5)
- *Markierung muss sich irgendwo in der Datenmatrix befinden*
 - *Anwählen von: Daten – Filter – AutoFilter*
 - *Anklicken des Filterkästchens Kategorie*
 - *Auswählen von „M“*
 - *Anklicken des Filterkästchens Länge*
 - *Auswählen von „(Benutzerdefiniert)“*
 - *Eingabe von Länge „ist größer als“ „450“.*
 - *OK → Beide Filter sind eingeschaltet. Es werden nur mehr die Mittelklassewagen mit einer Länge über 450 cm angezeigt.*
- 6) a)
- **Sortieren:**
 - *Datenbereich markieren*
 - *Anwählen von: Daten – Sortieren*
 - *Sortieren nach: Kategorie, Anschließend nach: Leistung*
 - *Der leistungsstärkste Sportwagen ist der Ferrari 456 M GT mit 325 kW.*
 - **Filtern:**
 - *Markierung muss sich irgendwo in der Datenmatrix befinden*
 - *Anwählen von: Daten – Filter – AutoFilter*

- *Anklicken des Filterkästchens Kategorie*
- *Auswählen von „S“*
- *Anklicken des Filterkästchens Leistung*
- *Auswählen von „(Top Ten)“*
- *Auswählen von „Obersten“ „1“ „Elemente“*
- *OK → Nur mehr der Ferrari 456 M GT wird angezeigt.*

b)

- *Im konkreten Fall erscheint das Sortieren etwas einfacher.*

7)

- *ABS, Klimaanlage, PLIP*

8)

- *Beruf, Geschlecht, Rasse, Haarfarbe*

9)

- *A, O, N, V, O, V, N, D, A*

10) a)

- *Säulendiagramm: ja; einzelne Merkmalsausprägungen*
- *Stängel-Blatt-Diagramm: nein; keine geeignete Klassenbildung möglich*
- *Histogramm: nein; keine Klassenbildung*

b)

- *Säulendiagramm: nein; zu viele verschiedene Merkmalsausprägungen*
- *Stängel-Blatt-Diagramm: (eher) ja; Zehner- und Hunderterstellen als Stängel, Einerstellen als Blätter*
- *Histogramm: ja; Klassenbildung*

c)

- *Säulendiagramm: nein; zu viele verschiedene Merkmalsausprägungen*
- *Stängel-Blatt-Diagramm: (eher) ja; z.B. Hunderter- und Tausenderstellen als Stängel*
- *Histogramm: ja; Klassenbildung*

11) a)

- *Säulendiagramm: Türen, Plätze, Airbags, ABS, Klimaanlage, PLIP, Kategorie*

b)

- *Stängel-Blatt-Diagramm: siehe c); teilweise keine günstige Möglichkeit in Stängel und Blätter zu teilen*

c)

- *Histogramm: Länge, Gewicht, Hubraum, Leistung, Höchstgeschwindigkeit, Beschleunigung, Verbrauch, Preis*

12)

- *=min(Datenbereich) → Das geringste Gewicht beträgt 720 kg.*

13)

- *=max(Datenbereich) → Das schnellste Auto fährt 300 km/h.*

14) a)

- *Addition des Gewichts aller Autos*
- *Division durch die Anzahl → Das mittlere Gewicht beträgt ungefähr 1.212 kg.*

b)

- *=mittelwert(Datenbereich) → Das mittlere Gewicht beträgt ungefähr 1.212 kg.*

15)

- **1.Art:**
- *Sortieren nach Gewicht*
- *Berechnen des arithmetischen Mittels von Wert #50 und Wert #51 → Der Median des Gewichts beträgt 1.184 kg.*
- **2.Art:**
- *=median(Datenbereich) → Der Median des Gewichts beträgt 1.184 kg.*

16)

Variable	Median	Mittelwert
Gewicht	1.184,00	1.211,86
Hubraum	1.595,00	1.886,50
Leistung	74,00	88,33
Vmax	185,00	188,30
Beschleunigung	12,00	12,05
Verbrauch	7,70	8,22

Tabelle 24

17) a)

- *Sortieren nach: Klimaanlage*
- *Anwählen von Daten – Teilergebnisse*
- *Gruppieren nach: Klimaanlage; unter Verwendung von: Mittelwert; Teilergebnisse addieren zu: Preis*
- *OK → Die Autos mit Klimaanlage sind im Mittel teurer als jene ohne Klimaanlage.*

b)

- *Sortieren nach: Kategorie*
- *Anwählen von Daten – Teilergebnisse*
- *Gruppieren nach: Kategorie; unter Verwendung von: Minimum; Teilergebnisse addieren zu: Verbrauch*
- *OK → 6,5 l ist das Minimum des Verbrauchs in der Kategorie Mittelklasseautos.*

c)

- *Sortieren nach: Plätze*
- *Anwählen von Daten – Teilergebnisse*
- *Gruppieren nach: Plätze; unter Verwendung von: Maximum; Teilergebnisse addieren zu: Länge*
- *OK → Die größte Länge bei Zweisitzern hat das Jaguar XK 8 Coupé mit 476 cm.*
- **Einfacher:** *etwa Sortieren nach: Plätze; anschließend nach: Länge*

d)

- *Sortieren nach: Kategorie*
- *Anwählen von Daten – Teilergebnisse*
- *Gruppieren nach: Kategorie; unter Verwendung von: Summe; Teilergebnisse addieren zu: ABS*
- *OK → C: 15, K: 9, L: 19, M: 17, S: 15*

18)

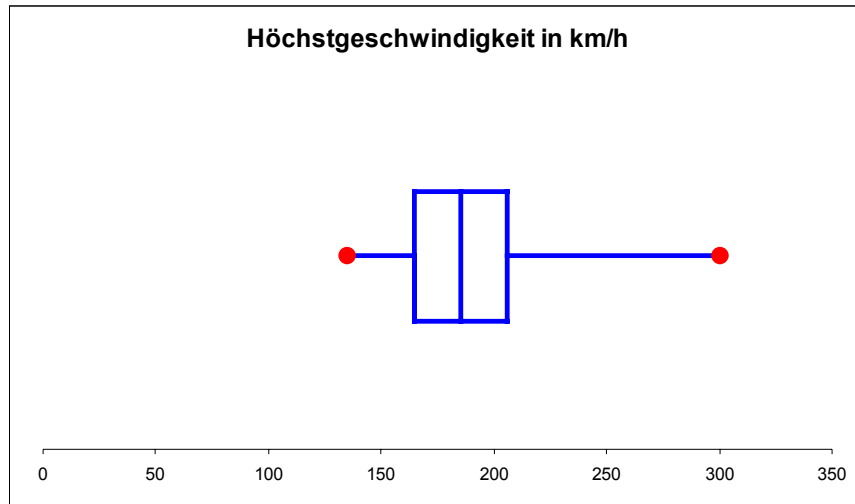
- *Wenn alle Gruppen die gleiche Anzahl von Elementen haben.*

19)

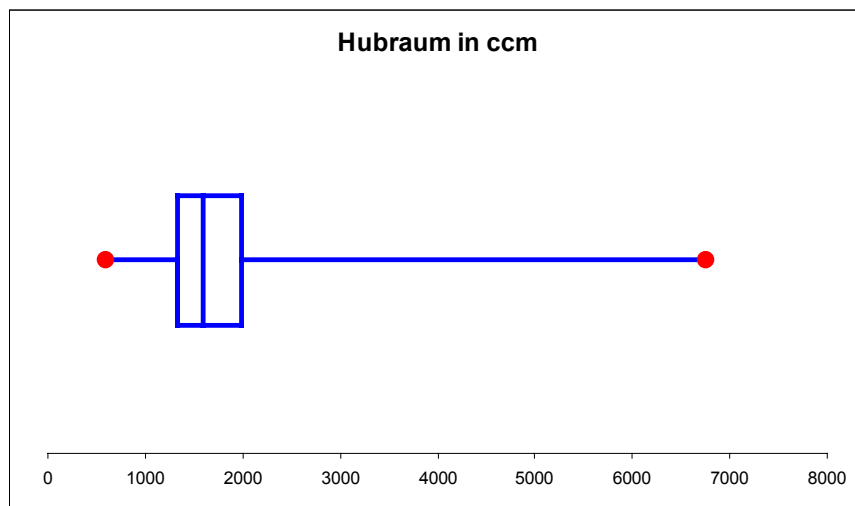
- *Mittelwert des Preises der Autos ohne ABS: 154.523,-*
- *Mittelwert des Preises der Autos mit ABS: 533.997,-*
- *Mittelwert der beiden Gruppen: 344.260,-*

- Gesamtmittelwert: 439.128,-

20)



21) a)



22) a)

- =varianzen(Datenbereich)
- =stabwn(Datenbereich)

b)

- Varianz
- von jeder Beobachtung Mittelwert abziehen
- jedes Einzelergebnis quadrieren
- Einzelergebnisse addieren
- Summe durch Anzahl dividieren
- Standardabweichung
- Wurzel der Varianz

Anhang 1: Datenmatrix

PKW-Name	Türen	Plätze	Länge	Gewicht	Hubraum	Leistung	Vmax	Beschl.	Airbags	ABS	Klimaanlage	PLIP	Verbrauch	Preis	Kategorie
Alfa Romeo 1,6T Spark	4	5	443	1230	1598	88	200	10,5	4	1	0	1	8,2	294000	M
Alfa Romeo 145 1,4 16V	3	5	409	1140	1370	76	185	11,2	2	1	0	0	7,9	212000	C
Alfa Romeo 166 2,0 TS	4	5	472	1420	1970	114	213	9,6	4	1	1	1	9,7	452000	X
Alfa Romeo GTV 2,0 TS	2	2	429	1370	1970	114	211	8,4	2	1	0	1	9,2	391000	S
Audi A3 1,6 Attraction	3	5	415	1090	1595	74	188	11	4	1	1	0	7,6	275640	C
Audi A4 1,6	4	5	448	1200	1595	74	190	12,4	4	1	1	0	7,9	328580	M
Audi A6 1,8	4	5	480	1355	1781	92	203	11,3	4	1	1	1	8,5	441570	X
Audi A8 2,8	4	5	503	1540	2771	142	236	8,4	6	1	1	1	9,9	775700	X
Audi TT 1,8T	3	2	404	1205	1781	132	228	7,4	4	1	1	1	8	474700	S
Bentley Arnage Red Label	4	5	539	2520	6750	298	249	6,3	2	1	1	1	19,2	3583400	X
BMW 316i	4	5	447	1360	1895	77	200	12,4	6	1	0	1	7,8	339000	M
BMW 316i compact	3	5	421	1175	1895	77	190	11,9	4	1	0	0	7,6	291000	C
BMW 520i	4	5	478	1470	1991	110	220	10,2	6	1	1	1	9	496000	X
BMW 728i	4	5	512	1710	2793	142	238	8,6	10	1	1	1	10	784000	X
BMW Z3 1,8	2	2	405	1220	1895	87	196	10,4	4	1	0	0	7,8	389000	S
BMW Z8	2	2	440	1585	4941	294	250	4,7	4	1	1	1	14,5	1990000	S
Chevrolet Corvette	2	2	456	1538	5666	253	274	5,7	2	1	1	1	13,1	997700	S
Chrysler Neon 2,0 LE	4	5	439	1245	1996	98	200	10,8	2	1	1	1	7,9	227000	M
Chrysler Stratus 2,0 LE	2	5	475	1361	1996	98	205	10,9	2	1	1	1	8,2	328650	X
Citroen Saxo 1,1i SX	3	5	372	805	1124	40	162	12,8	2	0	0	1	6,5	147800	K
Citroen Xantia 1,8i 16V SX	5	5	452	1290	1761	81	194	10,8	4	1	1	1	8,7	288150	M
Citroen XM 2,0i	5	5	471	1462	1998	108	228	8,7	4	1	1	1	10,4	459000	X
Citroen Xsara i SX	3	5	417	1020	1360	55	175	12	4	1	0	1	6,9	207480	C
Daewoo Lanos 1,4 SE	3	5	407	1015	1349	55	166	15	2	0	1	0	7,9	164900	K
Daewoo Nubira 1,6 SX	4	5	450	1178	1598	66	185	11	2	1	1	0	7,8	219900	C
Daihatsu Cuore GL	3	4	341	724	989	41	140	12,8	2	0	0	0	5	104900	K
Ferrari 456 M GT	2	2	473	1690	5474	325	300	5,2	2	1	1	1	19,8	3175000	S
Fiat Barchetta	2	2	392	1085	1747	96	200	8,9	0	0	0	0	8,5	307000	S
Fiat Bravo 80 16V SX	3	5	403	1010	1242	60	173	12,5	4	1	0	0	6,8	193000	C
Fiat Marea 100 16V SX	4	5	439	1140	1581	76	187	10,7	4	1	0	0	8,3	228000	M
Fiat Punto 1,2 ELX	3	5	380	860	1242	44	155	14,3	4	1	0	1	5,7	157900	K
Fiat Seicento Young	3	5	332	730	899	29	140	18	2	0	0	0	6,1	99900	K
Ford Escort 1,6 Champion	5	5	414	1190	1597	66	177	12,2	2	0	0	1	8,2	189500	C

Ford Fiesta 1,3 Flair	3	5	383	912	1299	37	143	19,5	4	0	0	0	6,9	152900	K
Ford KA 1,3 50	3	4	362	946	1299	37	147	17,7	2	0	0	0	5,9	129500	K
Ford Mondeo 1,6i	4	5	456	1295	1597	70	185	12,7	4	1	1	1	7,5	272000	M
Ford Puma 1,4	2	4	398	1050	1388	66	180	11,9	2	0	0	0	7,2	222500	S
Honda Accord 1,6i LS	4	5	460	1270	1590	85	190	12,2	4	1	0	1	8,2	279900	M
Honda Civic 1,4i Fun	3	5	419	1036	1396	55	165	13,9	2	1	0	0	6,8	189900	C
Honda Prelude 2,0i	2	4	455	1285	1997	98	201	9,2	2	1	0	1	9,2	299000	S
Jaguar XJ8 3,2 Exec.	4	5	502	1710	3248	174	225	8,5	2	1	1	1	12	890700	X
Jaguar XK8 Coupé	2	2	476	1615	3996	209	250	6,7	2	1	1	1	11,9	1102500	S
Lada 110 1,5i	4	5	427	1020	1499	52	167	13	0	0	0	0	6,6	125900	M
Lada Samara 1,5i	5	5	401	945	1499	57	155	13	0	0	0	0	7,7	99990	C
Lancia Y 1,2 LS	3	5	372	850	1242	44	160	13,3	1	0	0	1	6,6	159500	K
Mazda 323 P 1,3i 16V	3	5	404	1000	1324	54	164	13,3	2	0	0	0	7,4	139900	C
Mazda 626 1,8i	4	5	459	1170	1840	74	182	11,8	4	1	0	0	7,6	249900	M
Mazda Demio 1,3 SE	5	5	381	960	1324	46	151	14	2	1	0	0	7	183990	K
Mazda MX-5 1,6i 16V	2	2	398	1020	1589	81	191	9,7	2	0	0	0	8,1	295000	S
Mazda Xedos 9 2,0 V6	4	5	483	1455	1995	105	202	10,7	2	1	1	1	9	407000	X
MCC Smart pure	2	2	250	720	599	33	135	18,9	2	1	0	1	4,8	121500	K
Mercedes CL 500	2	5	499	1865	4966	225	250	6,5	4	1	1	1	12,5	1556674	S
Mercedes CLK 200	2	5	457	1375	1998	100	208	11	2	1	1	1	9,4	546739	S
Mercedes E 200	4	5	480	1510	1998	100	209	9,3	6	1	1	1	9,3	534535	X
Mercedes S 320	4	5	504	1770	3199	165	240	8,2	6	1	1	1	11,5	1106640	X
Mercedes SL 280	2	2	447	1810	2799	150	232	9,7	2	1	1	1	11,4	1172064	S
Mercedes SLK 200	2	2	401	1364	1998	120	223	8,2	4	1	0	1	9,6	508636	S
Mercedes A 140	5	5	358	1095	1397	60	170	12,9	2	1	0	0	7,1	254146	K
Mini	2	4	305	770	1273	46	148	13	1	0	0	0	6,6	159000	K
Nissan Almera 1,5 16V	3	5	418	1205	1497	66	173	13,8	4	1	0	0	6,6	174900	C
Nissan Maxima 2,0 V6	4	5	477	1465	1995	103	201	9,9	4	1	1	1	9,9	384900	X
Nissan Micra 1,0 SMouse	3	5	372	915	998	40	150	16,4	4	1	1	1	6,1	169900	K
Nissan Primera 1,6	4	5	452	1240	1597	73	181	12	4	1	0	0	7,2	239900	M
Opel Astra 1,2 16V	3	5	411	995	1199	48	165	16	2	1	0	1	6,2	189400	C
Opel Corsa 1,0 Joy	3	5	373	885	973	40	150	18	2	0	0	0	5,6	159500	K
Opel Omega 2,2 16V	4	5	490	1605	2198	106	210	10,5	4	1	1	1	9,4	410000	X
Opel Tigra 1,4-16V	3	4	392	980	1389	66	190	11,5	2	1	0	0	7,3	213000	S
Opel Vectra 1,6	4	5	450	1198	1598	55	175	15,5	4	1	0	1	7,4	249900	M
Peugeot 106 1,0 ZEN	3	5	368	795	954	37	150	19,4	2	0	0	0	6,2	127500	K
Peugeot 306 1,4 Comfort	3	5	403	1010	1360	55	165	14,7	4	1	0	1	6,9	191300	C
Peugeot 406 1,8 SR	2	5	460	1350	1761	66	181	14,6	4	1	1	1	8,4	264900	M

Peugeot 607 2,2	4	5	488	1455	2230	116	220	9,6	6	1	1	1	7,2	455000	X
Porsche Boxster	2	2	432	1250	2687	162	250	6,6	4	1	0	0	9,9	692000	S
Renault Clio 1,2 Travel	3	5	377	880	1149	43	160	15	4	1	0	1	6,2	160300	K
Renault Laguna 1,6 16V	5	5	451	1370	1598	79	195	11,5	4	1	1	1	7,5	285000	M
Renault Megane 1,4 RT	5	5	416	1075	1390	55	170	13,8	4	1	1	1	6,6	221300	C
Renault Safrane 2,5 RXE	5	5	477	1630	2435	120	220	9,1	4	1	1	1	10,3	504000	X
Renault Twingo	3	4	343	820	1149	43	151	13,7	4	1	0	0	6	134600	K
Rolls Royce Silver Seraph	4	5	564	2450	5379	240	225	7,8	2	1	1	1	17,4	3829000	X
Rover 25 1,4i	3	5	399	1095	1396	62	180	12,9	2	1	1	1	6,6	188000	C
Seat Arosa 1,0i Stella	3	4	354	864	999	37	151	17,6	2	0	0	0	5,6	132700	K
Seat Cordoba 1,6i Signo	2	5	414	1018	1598	55	170	12,7	4	1	0	1	7,2	206800	K
Seat Leon 1,4 16V Stella	5	5	418	1210	1390	55	170	14,6	4	1	0	1	6,5	207700	C
Seat Toledo 1,6 Stella	4	5	444	985	1595	74	188	mv	4	1	0	1	7,7	230900	M
Skoda Fabia 1,0 Classic	5	5	396	1064	997	37	162	15,4	1	0	0	0	7,1	137000	K
Skoda Felicia 1,3 Viva	5	5	386	985	1289	40	151	15,5	1	0	0	1	6,4	119900	C
Skoda Octavia 1,6 LX	5	5	451	1160	1598	55	170	14,4	2	0	0	0	7,6	182000	M
Subaru Alto L	3	5	350	730	993	40	150	15,5	0	0	0	0	5,7	99900	K
Suzuki Baleno 1,3GL	3	4	390	910	1298	63	160	13,1	2	0	0	0	6,4	145900	C
Suzuki Swift 1,0 GLS	3	5	375	745	993	39	145	15,9	2	0	1	1	5,5	125900	K
Volvo C70 2,0T Coupé	2	4	472	1434	1984	120	210	9,3	4	1	1	1	9,8	518763	S
Volvo S40 1,6	4	5	448	1220	1587	80	190	12	4	1	0	0	8	288966	M
Volvo S70 2,4	4	5	472	1391	2435	103	200	9,9	4	1	1	1	8,6	422441	X
Volvo S80 2,4	4	5	482	1489	2435	103	205	10,5	6	1	1	1	8,5	455466	X
VW Beetle 1,6	3	4	409	1205	1595	74	178	11,7	4	1	0	0	8,1	232050	C
VW Bora 1,4	4	5	438	1149	1390	55	171	14,9	4	1	0	0	6,5	223750	M
VW Golf 1,4	3	5	415	1146	1390	55	171	13,5	4	1	0	0	6,4	203990	C
VW Lupo 1,0	3	4	353	830	999	37	152	17,9	2	0	0	0	5,6	134590	K
VW Passat 1,6	4	5	468	1250	1595	74	192	12,5	4	1	0	0	8,1	276450	M
VW Polo 1,0	3	5	374	982	999	37	151	18,4	2	1	0	0	5,7	149900	K

Tabelle 25

Anhang 2: Zellbezüge

Was ist ein Zellbezug?

Zellbezüge haben wir bereits bei der Berechnung verschiedenster Zahlen kennengelernt. Sind die Argumente einer Formel bereits in der Excel-Tabelle vorhanden, müssen sie nicht nochmals eingegeben werden. Es genügt, wenn man in der Formel den Ort (die Zelle bzw. den Bereich) angibt, in der sich die Argumente befinden.

Microsoft Excel - Maple1
Datei Bearbeiten Ansicht Einfügen

	A	B	C	D	E
1					
2		3			
3		5			
4		=B2+B3			
5					
6					
7					

Abbildung 114

Microsoft Excel - Maple1
Datei Bearbeiten Ansicht Einfügen

	A	B	C	D	E
1					
2		3			
3		5			
4		8			
5					
6					
7					

Abbildung 115

Microsoft Excel - Maple1
Datei Bearbeiten Ansicht Einfügen

	A	B	C	D	E
1					
2		9			
3		5			
4		14			
5					
6					
7					

Abbildung 116

Ein kleines Beispiel zeigt Abbildung 114. Die Formel „=B2+B3“ liefert das Ergebnis „8“. Excel nimmt den Inhalt der Zelle B3 („5“) und addiert ihn zum Inhalt der Zelle B2 („3“).

Der Vorteil dieser Formel wird dann klar erkennbar, wenn man etwa den Wert in Zelle B2 ändert: die Summe wird ebenfalls sofort aktualisiert (siehe Abbildung 116).

Relativer Zellbezug

Die Art von Zellbezug, die wir soeben kennengelernt haben, nennt sich *relativer Zellbezug*.

Microsoft Excel - Maple1
Datei Bearbeiten Ansicht Einfügen

	A	B	C	D	E
1					
2		1	2	=B2+C2	
3		7	8		
4		12	14		
5					
6					
7					

Abbildung 117

Microsoft Excel - Maple1
Datei Bearbeiten Ansicht Einfügen

	A	B	C	D	E
1					
2		1	2	3	
3		7	8		
4		12	14		
5					
6					
7					

Abbildung 118

Microsoft Excel - Maple1
Datei Bearbeiten Ansicht Einfügen

	A	B	C	D	E
1					
2		1	2	3	
3		7	8	15	
4		12	14	26	
5					
6					
7					

Abbildung 119

Betrachten wir dazu Abbildung 117, Abbildung 118 und Abbildung 119: Zunächst wird in Zelle D2 eine einfache Formel eingegeben: der Inhalt der Zellen B2 und C2 soll addiert werden. Excel versteht diese Formel folgendermaßen: Kopiere den Inhalt der Zelle links von Zelle D2 mit dem Inhalt der Zelle 2x links von Zelle D2.

Die Bedeutung wird klar, wenn man die Formel der Zelle D2 in die Zellen D3, D4, usw. kopiert: in Zelle D3 wird nun wieder der Inhalt links diesmal von D3 mit dem Inhalt 2x links von D3 addiert (also B3+C3) usw.

Absoluter Zellbezug

Es gibt Fälle, in denen der relative Zellbezug nicht verwendbar ist.

Beispiel: In Abbildung 120 soll etwa zu jedem Wert in Spalte C immer derselbe Wert addiert werden. Wir haben den Wert in Zelle B2 geschrieben. Wenn man nun die Formel aus Abbildung 117 verwenden würde, bekäme man nur für D2 das gewünschte Ergebnis. Wir müssen einen Teil der Formel „fixieren“, also zu einem absoluten Zellbezug umwandeln.

	A	B	C	D	E
1					
2		6			
3		7			
4		8			
5		9			
6					
7					

Abbildung 120

	A	B	C	D	E
1					
2		6		= \$C\$2 + C2	
3		7			
4		8			
5		9			
6					
7					

Abbildung 121

	A	B	C	D	E
1					
2		6		12	
3		7		13	
4		8		14	
5		9		15	
6					
7					

Abbildung 122

In Abbildung 121 haben wir die korrekte Formel eingegeben: Es soll immer B2 (die Fixierung erfolgt mit \$-Zeichen) jeweils mit dem linken Nachbarn der Zelle addiert werden.

Kopiert man diese Formel in die Zellen D3, D4 und D5, erhält man das gewünschte Ergebnis.