

HANSER

Michael Monka, Werner Voß, Nadine M. Schöneck

Statistik am PC

Lösungen mit Excel

ISBN-10: 3-446-41555-6

ISBN-13: 978-3-446-41555-3

Leseprobe

Weitere Informationen oder Bestellungen unter
<http://www.hanser.de/978-3-446-41555-3>
sowie im Buchhandel.

4 Mittelwerte

„Da, wo wir sind, da ist die Mitte!“

H. Kohl, ehemaliger Bundeskanzler, 1989

Die goldene Mitte

- 4.1 Wo bitte ist die Mitte?
- 4.2 Wozu braucht man Mittelwerte?
- 4.3 Das arithmetische Mittel
- 4.4 Der häufigste Wert (Modus)
- 4.5 Der Zentralwert (Median)
- 4.6 Das geometrische Mittel

4.1 Wo bitte ist die Mitte?

In einer Veröffentlichung des Dachverbandes der deutschen Krankenversicherer fand sich vor nicht allzu langer Zeit die Meldung, dass das mittlere Einkommen der niedergelassenen Ärzte in der Bundesrepublik Deutschland derzeit bei 125.200 Euro jährlich liegt. Kurze Zeit darauf wurde vom Chefstatistiker des deutschen Ärzteverbandes mitgeteilt, dass dem mitnichten so sei, vielmehr habe die Hälfte der niedergelassenen Ärzte pro Jahr weniger als 69.700 Euro, die andere Hälfte allerdings mehr – die Mitte der Einkommen läge also offensichtlich bei diesem Betrag.

„Was ist denn davon zu halten?“, fragte meine Schwiegermutter – die erste von beiden. „Die einen behaupten ein Durchschnittseinkommen, das fast doppelt so groß ist wie der Betrag, der von der anderen Seite angeführt wird. Ich denke Mittelwert ist Mittelwert?“

Es dauerte eine ganze Weile, bis ich ihr dargelegt hatte, dass es unterschiedliche Möglichkeiten gibt, die Mitte eines gegebenen Datenbestandes herauszufinden, und dass es deshalb nicht verwunderlich ist, wenn sich unterschiedliche Ergebnisse zeigen.

„Je nach Interessenlage“, sagte ich, „wird man sich für den einen oder für den anderen Mittelwert entscheiden. Natürlich wird dann aber der redliche Statistiker darauf hinweisen, wie sich die Unterschiede zwischen den Ergebnissen, die in der Tat manchmal recht gravierend sind, erklären lassen, sodass der zunächst verutzte Konsument der Daten weiß, was er mit ihnen anfangen kann.“

„Na, ob das so das Richtige ist“, sagte meine erste Schwiegermutter und zog sich in die Küche zurück.

4.2 Wozu braucht man Mittelwerte?

Mittelwerte werden auch als *Durchschnittswerte* bezeichnet. Aus der Verwendung des Plurals in diesem Zusammenhang können Sie schon entnehmen, dass es unterschiedliche Möglichkeiten gibt, die Mitte eines gegebenen Datenbestandes zu bemessen. Die Aufgabe dieser statistischen Maßzahlen ist es anzugeben, um welchen speziellen Merkmalswert herum sich die einzelnen Merkmalswerte einer Häufigkeitsverteilung konzentrieren.

Was aber ist eigentlich die Mitte? Stellt man sich einen Kreis vor, dann ist ziemlich klar, was mit dem Begriff der Mitte gemeint ist. Es ist der Punkt, von dem alle Punkte auf der Kreislinie gleich weit entfernt sind (siehe Abbildung 4.1).

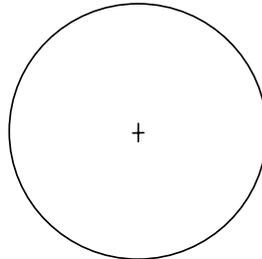


Abb. 4.1: Die Mitte eines Kreises

Bei einem Dreieck sieht das schon anders aus. Dort könnte man als Mitte den Schnittpunkt der drei Seitenhalbierenden bezeichnen (Geometriker bieten auch andere Lösungen an):

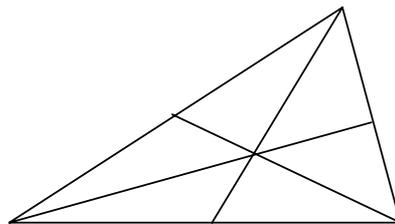


Abb. 4.2: Die Mitte eines Dreiecks

Sicherlich fallen Ihnen auch geometrische Figuren ein, bei denen die Mitte an einem Ort liegt, der seinerseits gar nicht zur Figur selbst gehört. Schauen Sie sich zum Beispiel den folgenden Ring an: Seine Mitte liegt in dem vom Ring umschlossenen leeren Bereich, trifft also auf einen Punkt, der selbst nicht zum Ring gehört (siehe Abbildung 4.3).

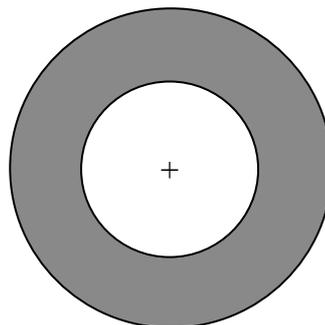


Abb. 4.3: Die Mitte eines Rings

Insbesondere dann, wenn mehrere Elemente in mehr oder weniger unregelmäßiger Anordnung versammelt sind, ist es manchmal nicht ganz einfach, die Mitte dieser Elemente festzustellen:

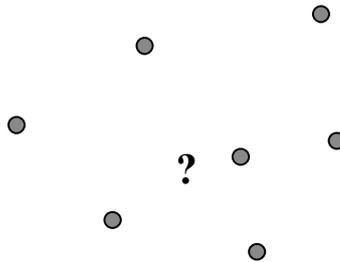


Abb. 4.4: Schwierige Mitte

Oder ist die Mitte einfach dort, wo die meisten Elemente sind? Ist die politische Mitte in Deutschland derzeit wieder bei der CDU, weil diese Partei vielleicht mehr Wählerstimmen auf sich vereinigt als eine der anderen Parteien?

Sie sehen, manchmal ist es gar nicht so einfach, darüber zu entscheiden, was mit dem Begriff Mitte gemeint ist. Glücklicherweise sind in der Statistik die Probleme nicht allzu gravierend. Wenn man die Mitte einer Häufigkeitsverteilung sucht, dann interessiert man sich für ihre *zentrale Tendenz*, wie es die Statistiker gern formulieren. Man geht dabei von der Überlegung aus, dass die Angabe dieser Mitte eine außerordentlich wichtige Detailinformation über den Datenbestand ist. Durch das Aufsuchen der Mitte aller einzelnen Merkmalswerte wird praktisch der gesamte vorliegende Datenbestand zu einer einzigen aussagekräftigen Zahl komprimiert. Die Zielsetzung der Beschreibung der zentralen Tendenz einer Häufigkeitsverteilung durch Komprimierung oder Verdichtung der Einzelinformationen (man spricht in diesem Zusammenhang auch von *Datenreduktion*) kann man auf mehreren Wegen, also unter Verwendung unterschiedlicher Maße der zentralen Tendenz, erreichen. Die wichtigsten dieser Maßzahlen sind das arithmetische Mittel, der häufigste Wert (Modus), der Zentralwert (Median) und das geometrische Mittel.

4.3 Das arithmetische Mittel

Wenn von Durchschnittsberechnungen gesprochen wird, versteht man meist die Verwendung des arithmetischen Mittels. Dieser Mittelwert ist so bekannt und gebräuchlich, dass oft mit dem Wort Durchschnitt genau diese Maßzahl gemeint ist. Das arithmetische Mittel ist ein *rechnerischer Mittelwert* (wir werden auch andere kennenlernen), weil bei seiner Bestimmung gerechnet werden muss.

4.3.1 Ungewogenes arithmetisches Mittel

Sicherlich ist Ihnen bekannt, dass das arithmetische Mittel in der Weise berechnet wird, dass man alle Merkmalswerte zusammenzählt und diese Summe dann durch die Anzahl der Merkmalswerte dividiert. Geht man so vor, spricht man von einem ungewogenen arithmetisches Mittel.

Hat man zum Beispiel die folgenden fünf Körpergrößen in einer kleinen Untersuchung erhoben:

173 cm, 180 cm, 175 cm, 187 cm, 185 cm,

so ergibt sich das arithmetische Mittel aus diesen fünf Werten (ihr Durchschnitt) wie folgt:

$$\text{Arithmetisches Mittel} = \frac{173 + 180 + 175 + 187 + 185}{5} = \frac{900}{5} = 180(\text{cm})$$

Die Durchschnittsgröße der befragten fünf Personen erhält man also, indem man die Ausprägungen der Variablen Körpergröße (die Merkmalswerte) zusammenzählt und diese Summe durch fünf teilt. Die durchschnittliche Körpergröße ergibt sich dabei zu 180 cm. Der Rechenansatz lautet also allgemein:

$$\text{Arithmetisches Mittel} = \frac{\text{Summe aller Merkmalswerte}}{\text{Anzahl der Werte}}$$

Wenn wir dies, wie es in der Statistik leider üblich ist, etwas mathematischer formulieren, sieht das so aus:

Die Untersuchungsvariable (Körpergröße) bezeichnen wir mit X , ihre Ausprägungen mit x_i (wobei $i = 1, 2, 3, \dots$).

Die Gesamtzahl der Beobachtungen wird mit n bezeichnet und das arithmetische Mittel mit dem griechischen Buchstaben μ (gesprochen als mü). Somit ergibt sich:

$$\mu = \frac{\sum x_i}{n}$$

Natürlich bietet auch das Programm Excel die Möglichkeit, das arithmetische Mittel auszurechnen. Wenn Sie mit Excel so arbeiten wollen, wie gerade per Hand gerechnet wurde, benutzen Sie die Funktion SUMME, um zunächst die Summe aller Merkmalswerte zu berechnen, die dann durch 5 dividiert werden muss. Sie könnten also folgendermaßen vorgehen:

1. Geben Sie in die Zellen A1 bis A5 einer neuen Arbeitstabelle die obigen Zahlenwerte ein.
2. Klicken Sie die Zelle A6 an.
3. Klicken Sie im Register START, Bereich BEARBEITEN auf die Summenschaltfläche und dann auf das Häkchen vor der Editierzeile.

Jetzt erscheint die Summe (900) in Zelle A6.

4. Geben Sie in A7 ein: =A6/5 und RETURN.

In der Zelle A7 Ihrer Tabelle erscheint jetzt das arithmetische Mittel (180).

Ein schnellerer Weg besteht darin, eine andere Excel-Funktion zu verwenden, nämlich die Funktion MITTELWERT. Statt zuerst eine Summe zu bilden, die dann durch die Anzahl der Merkmalswerte dividiert werden muss, können Sie mit dieser Funktion in einem einzigen Rechenschritt das arithmetische Mittel ausrechnen (siehe Abbildung 4.6)..

Diese Funktion können Sie nach Anklicken der Zelle A6 durch Anklicken des Dreiecksschalters bei der Summenschaltfläche im Register START, Bereich BEARBEITEN an der zweiten Stelle der sich öffnenden Liste finden und auswählen. In Abbildung 4.5 sehen Sie noch einmal dieses Auswahlfenster.



Abb. 4.5: Auswahl von Funktionen

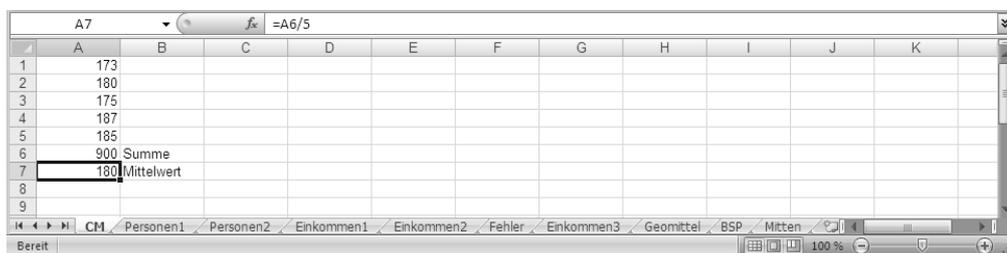


Abb. 4.6: Mittelwertberechnung (E04.XLSX, CM)

Klicken Sie dann auf das Häkchen vor der Editierzeile, wird das arithmetische Mittel in Zelle A6 ausgegeben. Oder Sie schreiben einfach in die Zelle A6:

```
=MITTELWERT(A1:A5)
```

4.3.2 Gewogenes arithmetisches Mittel

Treten einzelne Merkmalswerte mehrfach auf, so lässt sich die Berechnung des arithmetischen Mittels vereinfachen: Statt alle Merkmalswerte zusammenzuzählen, können zunächst diejenigen, die mehrfach aufgetreten sind, mit der Häufigkeit ihres Auftretens multipliziert werden, bevor zusammengezählt wird. Die Merkmalswerte werden durch diese Multiplikation mit ihren Häufigkeiten gewogen oder, wie man auch sagt, gewichtet. Insbesondere dann, wenn schon gruppiertes Material vorliegt, also zum Beispiel eine diskrete Häufigkeitsverteilung, bietet sich diese Vorgehensweise an.

Betrachten Sie zur Illustration noch einmal die Verteilung der Personenzahlen in befragten Haushalten, die wir in Kapitel 3 vorgestellt haben (siehe Abbildung 4.7).

Um aus den Daten der Abbildung 4.7 das arithmetische Mittel auszurechnen, also die durchschnittliche Haushaltsgröße, hat es keinen Sinn, wie im Beispiel zuvor, etwa in der Zelle A10 mit der Funktion SUMME oder direkt mit der Funktion MITTELWERT zu operieren. Es würde dann ja unterstellt, dass die Personenzahlen 1, 2 usw. jeweils nur einmal aufgetreten seien – und genau das ist nicht der Fall. Vielmehr werden wir zunächst in der Spalte C jeden Merkmalswert (Spalte A) mit der Häufigkeit seines Auftretens (Spalte B) multiplizieren und erst dann die Summe dieser Produkte bilden. Diese muss dann durch $n = 120$ dividiert werden.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Personen	Häufigkeit										
2		1	17									
3		2	33									
4		3	34									
5		4	19									
6		5	10									
7		6	4									
8		7	2									
9		8	1									
10			120									

Abb. 4.7: Diskrete Häufigkeitsverteilung (E04.XLSX, Personen1)

Die entsprechende Berechnungsformel lautet also:

$$\mu = \frac{\sum x_i \cdot f_i}{n}$$

Unter Benutzung relativer Häufigkeiten könnten wir auch schreiben:

$$\mu = \frac{\sum x_i \cdot f_i'}{100}$$

wenn die f_i' als Prozentwerte gegeben sind bzw.

$$\mu = \sum x_i \cdot f_i'$$

wenn die f_i' als Dezimalzahlen (Bruchteile von 1) gegeben sind.

1. Schreiben Sie in die Zelle C1 den Begriff Produkte.
2. Rechnen Sie in C2 $=A2*B2$.
3. Ziehen Sie diese Berechnung nach unten bis zur Zelle C9.
4. Berechnen Sie in C10 die Summe der Produkte durch Anklicken der Summen-Schaltfläche.
5. Errechnen Sie in C11 das arithmetische Mittel durch Eingabe von $=C10/B10$.

Es ergibt sich als Durchschnitt der Wert 2,975. Die durchschnittliche Personenzahl bei den befragten 120 Haushalten liegt also knapp unter drei Personen.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Personen	Häufigkeit	Produkte									
2		1	17	17								
3		2	33	66								
4		3	34	102								
5		4	19	76								
6		5	10	50								
7		6	4	24								
8		7	2	14								
9		8	1	8								
10	Summe		120	357								
11	Mittelwert			2,975								

Abb. 4.8: Diskrete Häufigkeitsverteilung mit Mittelwertberechnung (E04.XLSX, Personen2)

Sie erkennen, das arithmetische Mittel kann ein Wert sein, der real gar nicht beobachtet worden ist und manchmal auch nie beobachtet werden könnte – oder haben Sie schon einmal einen Haushalt mit 2,975 Personen gesehen?

Auch diese Berechnung lässt sich vereinfachen: Statt die einzelnen Produkte zu berechnen, um diese dann aufzusummieren, können Sie diese benötigte Summe (es hatte sich der Wert 357 ergeben) auch direkt erzeugen, wenn Sie den Dreiecksschalter bei der Summenschaltfläche anklicken und dann WEITERE FUNKTIONEN ... auswählen. Wählen Sie dann aus der Kategorie MATH.&TRIGONOM. die Funktion SUMMENPRODUKT aus (siehe Abbildung 4.9).

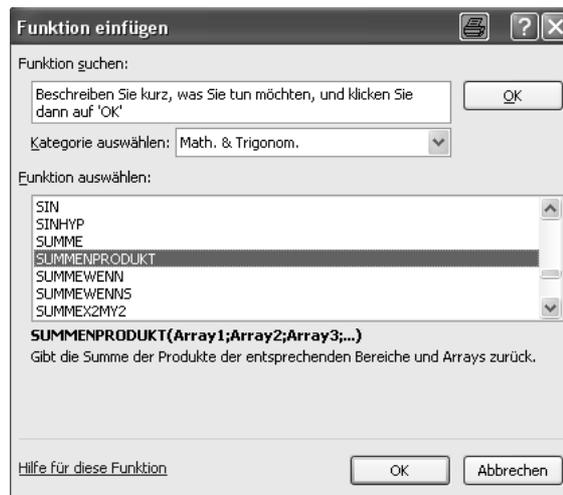


Abb. 4.9: Auswahl der Funktion SUMMENPRODUKT

Klicken Sie dann auf OK, gelangen Sie zur Eingabe der Funktionsargumente (siehe Abbildung 4.10).

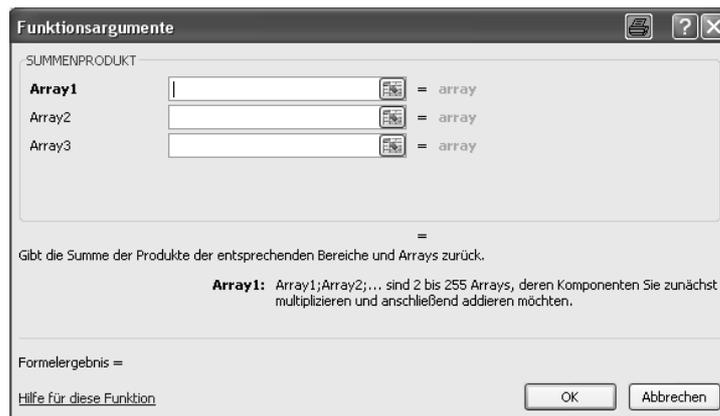


Abb. 4.10: Eingabe der Argumente bei der Funktion SUMMENPRODUKT

Hier geben Sie bei ARRAY1 ein: A2:A9
 und bei ARRAY2: B2:B9

Excel berechnet Ihnen dann nach dem Anklicken von OK sofort die Summe der Produkte, die dann durch $n = 120$ noch dividiert werden muss, um das arithmetische Mittel der gegebenen Verteilung zu finden.

Bei klassifiziertem Datenmaterial einer stetigen Untersuchungsvariablen ist die Berechnung eines arithmetischen Mittels etwas umständlicher. Betrachten wir noch einmal unser Ausgangsbeispiel zur Verteilung der Einkünfte Erwerbsloser:

	A	B	C	D	E	F	G	H	I	J	K
1	Klassenobergrenze	Anzahl									
2	1000	29									
3	1500	47									
4	2000	16									
5	2500	10									
6	3000	13									
7	3500	3									
8	4000	2									
9											
10											

Abb. 4.11: Stetige Einkommensverteilung (E04.XLSX, Einkommen1)

Hier ist es ja nun leider so, dass wir die Merkmalswerte gar nicht mehr kennen, die wir brauchen, um sie bei der Berechnung des arithmetischen Mittels zusammenzählen zu können – es sei denn, wir würden wieder auf die Urliste zugreifen und so verfahren, wie es oben bei der Mittelung der fünf Körpergrößen vorgeführt wurde. Uns stehen hier nur noch Klassen von Merkmalswerten zur Verfügung, wobei in der Tabelle sogar nur die Klassenendpunkte zu sehen sind (Spalte A). Man geht in einem solchen Fall so vor, dass man die mittleren Werte der einzelnen Merkmalsklassen als stellvertretende Merkmalswerte ansieht und mit ihrer Hilfe das gewogene arithmetische Mittel berechnet.

In der obigen Tabelle müssten also in der Spalte C zunächst die Klassenmittelpunkte notiert werden, bevor dann so verfahren werden kann wie im Beispiel zuvor.

1. Schreiben Sie in Zelle C1 das Wort Klassenmitten
2. Rechnen Sie in C2 $=A2-250$
3. Ziehen Sie diese Berechnung bis zur Zelle C8.

Erinnern Sie sich daran, dass wir Klassen der Breite 500 (Euro) gebildet hatten.

Die jeweiligen Klassenmittelpunkte erhält man also in der Tat, indem man von den Klassenendpunkten (Spalte A) den Wert 250 abzieht. Hätten Sie andere Klassenbreiten benutzt, müssten Sie entsprechend reagieren. Bei unterschiedlichen Klassenbreiten bleibt Ihnen nichts anderes übrig, als die Klassenmitten einzeln per Hand einzugeben.

4. Klicken Sie C9 an, und wählen Sie die oben erwähnte Funktion SUMMENPRODUKT, um aus den Werten (C2:C8) und (B2:B8) den Wert 184000 auszurechnen.
5. In C10 können Sie jetzt das arithmetische Mittel der Verteilung ausrechnen lassen, indem Sie eingeben: $=C9/B9$.

Es ergibt sich als Wert der durchschnittlichen Einkünfte der 120 befragten Personen der Wert 1533,33 Euro.

	A	B	C	D	E	F	G	H	I	J	K
1	Klassenobergrenze	Anzahl	Klassenmitten								
2	1000	29	750								
3	1500	47	1250								
4	2000	16	1750								
5	2500	10	2250								
6	3000	13	2750								
7	3500	3	3250								
8	4000	2	3750								
9		120	184000								
10			1533.333333								
11											

Abb. 4.12: Stetige Einkommensverteilung und Mittelwertberechnung (E04.XLSX, Einkommen2)

Man sieht bei diesem Beispiel sehr deutlich, dass mit einer einzigen statistischen Zahl, nämlich mit dem arithmetischen Mittel als Durchschnittswert, eine wesentliche Aussage über die zugrunde liegende Häufigkeitsverteilung gemacht werden kann. Die Angabe, dass das Durchschnittseinkommen bei ca. 1530 Euro liegt, ist ein wichtiges Charakteristikum dieser Verteilung.

An dieser Stelle muss mit Nachdruck darauf hingewiesen werden, dass das arithmetische Mittel nur bei metrischen Daten berechnet werden sollte, weil bei Ordinal- oder bei Nominaldaten eine inhaltliche Interpretation dieser Maßzahl nicht möglich ist. Anders ist es hingegen mit den folgenden beiden Mittelwerten.

Eine weitere Anmerkung ist für den Fall notwendig, dass aus klassifiziertem Material – wie eben – ein arithmetisches Mittel berechnet werden soll: Hat man *offene Flügelklassen*, so ist es nicht möglich, für diese Klassenmittelpunkte anzugeben; dann aber kann das arithmetische Mittel nicht berechnet werden – es sei denn, man schließt die offenen Klassen künstlich.

4.4 Der häufigste Wert (Modus)

Der häufigste Wert ist derjenige Merkmalswert in einem gegebenen Datenbestand, der am häufigsten aufgetreten ist. Auch er ist ein charakterisierender Mittelwert, weil er ohne Zweifel eine Aussage über die Mitte einer vorliegenden Häufigkeitsverteilung macht. Allerdings ist dies kein rechnerischer Mittelwert, sondern er wird als *lagetypischer Mittelwert* bezeichnet, weil er ausschließlich durch die Lage (die Position) der einzelnen Merkmalswerte bestimmt wird.

Offensichtlich ist der Modus viel leichter aufzufinden als ein arithmetisches Mittel. Er kann aber nur bei solchen Häufigkeitsverteilungen bestimmt werden, bei denen die Merkmalswerte mit unterschiedlichen Häufigkeiten aufgetreten sind. Bei dem Beispiel der fünf verschiedenen Körpergrößen, das wir als erstes Beispiel in diesem Kapitel verwendet haben, ist jeder Merkmalswert nur einmal genannt worden; ein häufigster Wert existiert hier gar nicht. Anders hingegen bei der Verteilung der Haushaltsgrößen der Abbildung 4.13.