

Transkript zum Video E3.2

Doppelsummen

aus der Vorlesung Mathematik für Wirtschaftswissenschaften

Inhalt

Folie 1 – Summen und Logik.....	1
Folie 2 – Thema E3.2	1
Folie 3 – Lernziele.....	2
Folie 4 – Tabellarische Datenstrukturen – Illustrationsbeispiel	2
Folie 5 – Tabellarische Datenstrukturen – Illustrationsbeispiel 2	3
Folie 6 – Tabellarische Datenstrukturen	3
Folie 7 – Zeilensummen.....	4
Folie 8 – Zeilensummen – Illustrationsbeispiel	4
Folie 9 – Spaltensummen	5
Folie 10 – Spaltensummen – Illustrationsbeispiel.....	5
Folie 11 – Spaltensummen – Illustrationsbeispiel 2.....	6
Folie 12 – Gesamtsumme	6
Folie 13 – Gesamtsumme 2	7
Folie 14 – Gesamtsumme – Illustrationsbeispiel.....	8
Folie 15 – Rechnerleistung nutzen: Umsetzung in R – Doppelsummen	8
1. Benötigte Datenstruktur	9
2. Illustrationsbeispiel 1: Beispielobjekt Umsatztabelle	9
3. Einschub: apply()-Befehl	12
Folie 19 – Vielen Dank für die Aufmerksamkeit	12

Hinweis zur Schreibweise

Im Folgenden werden (sofern vorhanden) hochgestellte Zahlen oder Buchstaben durch $^$ ($A^2 = A^2$) und tiefgestellte Zahlen oder Buchstaben durch $_$ ($a_j = a_j$) markiert.

Folie 1 – Summen und Logik

Folientext

Einführung III: Summen und Logik. Alexander Silbersdorff, Wirtschaftswissenschaftliche Fakultät und Campus-Institut Data Science der Georg-August-Universität Göttingen, Logo der Georg-August-Universität Göttingen.

Folie 2 – Thema E3.2

Folientext

Themen

E3.2 Doppelsummen

Sydsæter et al. (2018): Unterkapitel 1.11

Sprechtext

Herzlich Willkommen zu diesem Lehrvideo zu Kapitel E 3.2 zum Thema Doppelsummen im Rahmen der Vorlesung Mathematik für Wirtschaftswissenschaften an der Georg-August-Universität Göttingen. Falls Sie die Inhalte dieses Kapitels noch mal anhand eines Buches nacharbeiten möchten, verweisen wir auf Sydsæter et al. (2018), Unterkapitel 1.11.

Folie 3 – Lernziele

Folientext

Im Rahmen dieses Unterkapitels lernen Sie über...

- Zeilen- und Spaltensummen
- Doppelsummen
- Umsetzung am Rechner

Sprechtext

Im Rahmen dieses Unterkapitels werden Sie lernen, wie Sie Summen über Zeilen und Spalten von tabellarisch strukturierten Datenmengen bilden können und wie diese Summen mit Hilfe des Summenzeichens darstellbar sind. Anschließend betrachten wir die Summierung dieser Zeilen- und Spaltensummen mittels Doppelsummen. Und zuletzt zeigen wir, wie die Berechnung von Zeilen- und Spaltensummen sowie von Doppelsummen mittels der in dieser Veranstaltung genutzten Software umgesetzt werden kann.

Folie 4 – Tabellarische Datenstrukturen – Illustrationsbeispiel

Folientext

Umsatz in 3 Verkaufsregionen über 4 Quartale:

- Abbildung: Tabelle der Umsatzzahlen als Unbekannte

	1. Quartal: Q1	2. Quartal: Q2	3. Quartal: Q3	4. Quartal: Q4
1. Region: NordGö	U_N1	U_N2	U_N3	U_N4
2. Region: SüdGö	U_S1	U_S2	U_S3	U_S4
3. Region: Umland	U_U1	U_U2	U_U3	U_U4

Sprechtext

Lassen Sie mich einsteigen mit einem Illustrationsbeispiel zu tabellarischen Datenstrukturen. Betrachten wir eine Göttinger Firma, die ihre Umsatzzahlen entlang drei Verkaufsregionen und vier Quartalen strukturiert. Hier aufgeführt haben wir drei Verkaufsregionen in den Zeilen: Nordgöttingen, Südgöttingen und das Göttinger Umland. Hinsichtlich der Quartale, dargestellt in den Spalten, betrachten wir alle vier Quartale eines beliebigen Jahres. Entsprechend ergeben sich pro Verkaufsregion vier Umsatzzahlen, nämlich eine pro Quartal, sodass wir insgesamt zwölf Umsatzzahlen für alle drei Verkaufsregionen vorliegen haben. In der auf dieser Folie gezeigten tabellarischen Darstellungsform sind die Umsätze für die Region Göttingen Nord in der ersten Zeile dargestellt, wobei der erste Eintrag U mit dem Index N1 den Umsatz für das erste Quartal darstellt, während die folgenden Einträge der Zeile die Umsätze für die folgenden Quartale von U_N2 bis U_N4 darstellen. Analog sind dann in der zweiten Zeile die vier Umsatzeinträge für die Region Südgöttingen zu finden, welche respektive U mit den Indizes S1 bis S4 benannt sind. Und zuletzt haben wir für das Umland vier Umsatzeinträge, welche als U für Umsatz mit den Indizes U1 bis U4 notiert sind. Und

diese Umsatzzahlen können wir entweder, wie auf dieser Folie hier dargestellt, grundsätzlich als Unbekannte betrachten, oder wir betrachten konkrete Umsatzzahlen, welche auf der nächsten Folie gezeigt sind.

Folie 5 – Tabellarische Datenstrukturen – Illustrationsbeispiel 2

Folientext

Umsatz in 3 Verkaufsregionen über 4 Quartale:

- Abbildung: Tabelle der konkreten Umsatzzahlen

	1. Quartal: Q1	2. Quartal: Q2	3. Quartal: Q3	4. Quartal: Q4
1. Region: NordGö	53	32	61	7
2. Region: SüdGö	92	92	46	57
3. Region: Umland	22	28	89	3

Sprechtext

Hier hätten wir beispielsweise in der Verkaufsregion Göttingen Nord für das erste Quartal einen Umsatz von 53 und in den Folgequartalen dann Umsätze von 32, 61 und 7. Für die Region Göttingen Süd in der zweiten Zeile hätten wir entsprechend Umsätze von zweimal 92 fürs erste und zweite Quartal und dann 46 im dritten und 57 im letzten Quartal. Und für das Umland hätten wir respektive Umsätze von 22, 28, 89 und 3.

Folie 6 – Tabellarische Datenstrukturen

Folientext

- Abbildung: Tabellarische Darstellung von Daten mit entsprechender Notation

	n Spalten					
m Zeilen	a ₁₁	a ₁₂	...	a _{1j}	...	a _{1n}
	a ₂₁	a ₂₂	...	a _{2j}	...	a _{2n}
	⋮	⋮		⋮		⋮
	a _{i1}	a _{i2}	...	a _{ij}	...	a _{in}
	⋮	⋮		⋮		⋮
	a _{m1}	a _{m2}	...	a _{mj}	...	a _{mn}

- $a_{ij} \in \mathbb{R}$ mit $1 \leq i \leq m$ und $1 \leq j \leq n$
- (zeitweilige Einblendung: Doppelsummen können nur auf entsprechend strukturierte Objekte angewandt werden.)

Sprechtext

Dieses Illustrationsbeispiel verallgemeinernd werden wir im Folgenden Summen für Objekte bilden, welche eine tabellarische Struktur aufweisen, das heißt, unsere Daten sind in Zeilen und Spalten geordnet. Die jeweiligen Einträge notieren wir grundsätzlich allgemein abstrakt im Rahmen dieser Veranstaltung mit klein a und einem Index, welcher zuerst die dem Eintrag entsprechende Zeile notiert und dann die Spalte. Der Eintrag a₁₂ kennzeichnet also den Eintrag in der ersten Zeile in der zweiten Spalte. Und grundsätzlich gehen wir davon aus, dass alle Einträge entweder bekannte oder aber unbekannte reelle Zahlen sind, was wir hier auf den Folien mit der Notation a_{ij} ist Element von \mathbb{R} kennzeichnen, also Element der reellen Zahlen. Und dabei ist i die jeweilige Zeile des Eintrages und nimmt die ganzzahligen Werte zwischen 1 und m an, und j die jeweilige Spalte des Eintrages mit den ganzzahligen Werten zwischen 1 und n. Bevor wir nun mit dem Bilden von Summen starten, möchte

ich an dieser Stelle noch einmal festhalten, dass diese tabellarische Struktur mit Zeilen und Spalten eine essenzielle Voraussetzung für das Bilden von Doppelsummen ist, so wie wir sie im Rahmen dieser Vorlesung betrachten. Mit dieser tabellarischen bzw. matrixartigen Struktur werden wir uns in dem Kapitel ‚Lineare Algebra 1‘ noch ausführlicher beschäftigen.

Folie 7 – Zeilensummen

Folientext

- Abbildung: Tabellarische Darstellung von Daten mit entsprechender Notation (siehe [Folie 6](#))

Zeilensummen:

$$\bullet \quad \sum_{j=1}^n a_{1j}, \quad \sum_{j=1}^n a_{2j}, \quad \dots, \quad \sum_{j=1}^n a_{ij}, \quad \dots, \quad \sum_{j=1}^n a_{mj}$$

Sprechtext

Wenn wir für die Tabelle die Zeilen einzeln betrachten, können wir grundsätzlich für jede Zeile die Summe entsprechend dem Vorgehen aus dem vorherigen Kapitel E 3.1 berechnen. Bei m Zeilen ergeben sich somit m einzeln zu bildende Summen, welche unter der Tabelle auf dieser Folie abgebildet sind. Die erste Summe, welche von j=1 bis n läuft und a_{1j} aufsummiert, stellt also die Summe der ersten Zeile dar. Ich kann das hier an dieser Stelle auf den Folien noch mal notieren. Das heißt, wir haben die erste Zeile, über die wir die Summe bilden, die ich hier auch noch mal rechts neben der Tabelle handschriftlich vermerken werde. Wir würden also die Summe von j=1 bis n über eben all die Elemente in der ersten Zeile, also a_i für eine gegebene j-te Spalte, berechnen. Und diesen Ausdruck finden Sie dann natürlich noch mal in lesbarer Form unten links hier aufgeführt. Analog gibt die zweite hier angegebene Summe die Summe über alle Elemente der zweiten Zeile. Das heißt, auch hier würde ich das ganze noch mal unterstreichen und mit einem Pfeil auf die zweite hier dargestellte Summe mit Hilfe des Summenzeichens verweisen. Und das können wir dann zeilenweise weiterführen bis zur m-ten und letzten Zeile, wo wir jetzt eben die Summe von j=1 bis n für a_{mj} laufen lassen, und diese letzte Zeile wäre hier entsprechend als letztes noch mal aufgeführt.

Folie 8 – Zeilensummen – Illustrationsbeispiel

Folientext

- Abbildung: Tabelle der konkreten Umsatzzahlen (siehe [Folie 5](#))

Zeilensummen:

$$\bullet \quad \sum_{q=1}^4 U_{Nq} = 53 + 32 + 61 + 7 = 153$$

$$\bullet \quad \sum_{q=1}^4 U_{Sq} = 92 + 92 + 46 + 57 = 287$$

$$\bullet \quad \sum_{q=1}^4 U_{Uq} = 22 + 28 + 89 + 3 = 142$$

- (zeitweilige Einblendung: U wird sowohl für Umland als auch für Untersummen genutzt. U_{Index} ist die Untersumme von der jeweiligen Zeile oder Spalte.)

Sprechtext

Für das vorhin eingeführte Illustrationsbeispiel der Umsätze pro Region und Quartal mit konkreten Zahlen würden diese drei Zeilen drei separate Zeilensummen implizieren, welche jeweils mit den Umsätzen für die jeweiligen Regionen über alle vier Quartale darstellen. Entsprechend ist die erste auf den Folien dargestellte Summe, welche von q=1 bis q=4 über U_{Nq} läuft, die Summe der

Umsätze für die Region Nordgöttingen über alle vier Quartale, also die Summe, welche ich hier auch noch mal mit dem Strich unterstreichen würde, also $53 + 32 + 61 + 7$ ist dann eben entsprechend gleich 153, also die hier in der ersten Zeile dargestellte Summe. Die zweite Summe wäre entsprechend das Pendant für Südgöttingen, während die dritte Summe für das Umland ist, und diese zwei Summen sind jeweils auch hier aufgeführt. Worum es mir geht, bzw. was ich hoffe, was Sie mitnehmen, ist, dass also eine Zeilensumme letztendlich einfach eine spezielle Form der Summe über eine bestimmte Teilmenge der Daten, nämlich die Summe über jene Teilmenge der Daten ist, die in einer bestimmten Zeile zu finden sind.

Folie 9 – Spaltensummen

Folientext

- Abbildung: Tabellarische Darstellung von Daten mit entsprechender Notation (siehe [Folie 6](#))

Spaltensummen:

$$\bullet \quad \sum_{i=1}^m a_{i1}, \sum_{i=1}^m a_{i2}, \dots, \sum_{i=1}^m a_{ij}, \dots, \sum_{i=1}^m a_{in}$$

Sprechtext

Analog zu den Zeilensummen können wir in tabellarischen Objekten natürlich auch die Elemente spaltenweise aufsummieren. Die resultierenden Spaltensummen haben wir hier nun erneut für das allgemeine Beispiel aufgeführt. Die erste Spaltensumme, welche von $i=1$ bis $i=m$ über a_{i1} läuft, ist entsprechend die Summe aller Elemente in der ersten Spalte, was ich auch hier noch mal mit dem Strich notieren würde und wo ich links neben die Tabelle am unteren Ende dieser ersten Spalte eben das Ganze auch noch mal handschriftlich aufschreiben würde, mit Hilfe des Summenzeichens. Also habe ich die Summe von $i=1$ bis eben m über a_{i1} . Das ist hier die Spaltensumme aus der ersten Spalte, welche entsprechend hier auch noch mal unten entsprechend aufgeführt ist. Und für die weiteren Spalten, welche hier auch auf der Folie entsprechend unten abgebildet sind, wird analog vorgegangen.

Folie 10 – Spaltensummen – Illustrationsbeispiel

Folientext

- Abbildung: Tabelle der konkreten Umsatzzahlen (siehe [Folie 5](#))

Spaltensummen:

$$\bullet \quad \sum_{r=1}^3 U_{r1} = 53 + 92 + 22 = 167$$

$$\bullet \quad \vdots$$

Sprechtext

Für das verwendete Illustrationsbeispiel würde eine spaltenweise Aufsummierung bedeuten, dass wir für jedes Quartal die Summe über alle Regionen bilden. Das heißt, für das erste Quartal, was ich hier auch noch mal mit einem Strich markieren würde, hätten wir also entsprechend die Summe über die Umsätze aus den drei Regionen Nordgöttingen, Südgöttingen und das Göttinger Umland. Notationell könnten wir diese erste Spaltensumme über eine Indexierung über die jeweiligen Zeilen vornehmen. Das heißt, wir notieren die jeweilige Zeile des Eintrags mit klein r , sodass wir entsprechend als erste Spaltensumme die Summe von dem Index $r=1$ bis $r=3$ über U_{r1} haben. Und

dies entspricht dann den jeweiligen Einträgen, also die Summe von 53, 92 und 22, was dann zu einer Spaltensumme von 167 führt, was ich hier vielleicht der Sicherheit halber auch noch mal vermerke. Das heißt, hier unterhalb von der Tabelle würde ich schreiben, dass wir eben die Summe von $r=1$ über 3 für eben unsere Umsatzeinträge groß U_{r1} haben würde, was dann eben 167 ergibt. Und für das zweite Quartal würden wir dann entsprechend die Einträge von der zweiten Spalte aufsummieren. Das wird in dem Fall für diejenigen, die es nachrechnen, 152 ergeben, und für die dritte und vierte Spalte enthalten wir analog respektive 196 und 67.

Folie 11 – Spaltensummen – Illustrationsbeispiel 2

Folientext

- Abbildung: Tabelle der konkreten Umsatzzahlen (siehe [Folie 5](#))

$$\sum_R U_{R1} = \sum_{R=\{N,S,U\}} U_{R1} = \sum_{r=1}^3 U_{r1} = 53 + 92 + 22 = 167$$

- \vdots

Sprechtext

Für diejenigen von Ihnen, die sich optionalerweise, lobenswerterweise, noch mit alternativen Summennotationen auseinandersetzen möchten, sei an dieser Stelle kurz darauf hingewiesen, dass wir grundsätzlich auch von der im Rahmen dieser Veranstaltung üblicherweise verwendeten aufsteigenden Summenindexnotation abweichen können und Summen auch über Elemente von unstrukturierten Mengen bilden können, bzw. in unserem Fall die Menge aus N und S und U, welche ich hier an der Stelle beispielsweise dann unter der Menge r subsumieren könnte. Diese Summierung über die Elemente von Summen, die unstrukturierten Elemente von Summen, würden wir so darstellen, dass wir anstelle des Indexes, welche wir unter und über das Summationssymbol schreiben, einfach die Menge, über dessen Elemente wir summieren, unterhalb des Summenzeichens schreiben. Das können wir entweder ausführlich machen, wie das hier an zweiter Stelle auf der Folie notiert ist, was ich hier eben gerade unterstrichen habe. Das heißt, dass wir die Menge noch mal in aller Ausführlichkeit in ausformulierter Form unterhalb des Summationssymbols schreiben. Das heißt, dass ich schreibe, r ist gleich eben die Menge aus N und S und U, und dann eben die Summe über die U Elemente bilde für die jeweiligen Einträge zu den Regionen. Oder was eigentlich der Regelfall ist und was ich hier entsprechend noch mal umkringeln würde, ist, dass wir einfach die Menge, über deren Elemente wir summieren, dass ich ausschließlich diese Menge, ohne dass ich sie ausformuliere, unterhalb des Summenzeichens schreibe. Und was das dann implizit bedeutet, ist, dass ich über alle Elemente dieser Menge bzw. über alle Einträge, die diese Menge hat, und deren Einsetzen dann in unser Umsatzobjekt, was wir hier haben, dass dort eben entsprechend eine Aufsummierung über alle Elemente vorgenommen wird. Wichtig wäre mir aber grundsätzlich vorrangig, dass Sie mitnehmen, dass es Spaltensummen gibt, dass wir sie bilden können und dass letztendlich Spaltensummen auch wieder eine spezielle Form der Summe sind, wo Summen über Teilmengen von Daten gebildet werden. Diese sind, wie das der Name schon hergibt, dann eben die Summe von einer Spalte bzw. die Summen von mehreren Spalten.

Folie 12 – Gesamtsumme

Folientext

- Abbildung: Tabellarische Darstellung von Daten mit entsprechender Notation (siehe [Folie 6](#))

Gesamtsumme:

- Summe aller Elemente aus

- $\{a_{ij} \in \mathbb{R} : (i, j) \in \mathbb{Z}^2, 1 \leq i \leq m, 1 \leq j \leq n\}$

Sprechtext

Kommen wir nun abschließend zu Doppelsummen bzw. erst mal allgemeiner der Gesamtsumme für eine wieder tabellarisch geordnete Menge. Grundsätzlich gilt, dass die Gesamtsumme über alle Elemente in der Tabelle, dem Kommutativgesetz folgend, eigentlich über die Aufsummierung aller Elemente in beliebiger Reihenfolge gebildet werden kann, und dies wird an dieser Stelle auf der Folie unterhalb der Tabelle auch noch mal formal dargestellt. Dort steht, dass wir die Summe aller Elemente aus der Menge aller a_{ij} nehmen und aufsummieren, wobei a_{ij} jegliche reelle Zahl darstellen kann, oder ein Element der reellen Zahlen ist, und deren ganzzahlige Indizes bei i zwischen 1 und m liegen und bei j zwischen 1 und n . Und über all diese Elemente summieren wir eben auf, in beliebiger Reihenfolge, und das gibt uns dann entsprechend die Gesamtsumme. Gerade bei einer tabellarischen Anordnung bietet sich aber natürlich eine bestimmte Reihenfolge besonders an, oder zwei Reihenfolgen bieten sich besonders an, im Fall von in Zeilen und Spalten strukturierten Daten. Dort bietet es sich an, entweder zeilen- oder aber spaltenweise vorzugehen.

Folie 13 – Gesamtsumme 2

Folientext

Gesamtsumme via Zeilen-/Spaltensummen:

- $$\sum_{j=1}^n a_{1j} + \sum_{j=1}^n a_{2j} + \dots + \sum_{j=1}^n a_{ij} + \dots + \sum_{j=1}^n a_{mj}$$
- $$\sum_{i=1}^m a_{i1} + \sum_{i=1}^m a_{i2} + \dots + \sum_{i=1}^m a_{ij} + \dots + \sum_{i=1}^m a_{in}$$

Gesamtsumme via Doppelsumme:

- $$\sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \right) = \sum_{j=1}^n \left(\sum_{i=1}^m a_{ij} \right) = \sum_{j=1}^n \sum_{i=1}^m a_{ij}$$

Sprechtext

Das heißt, die Gesamtsumme kann unter anderem über die Aufsummierung aller Zeilensummen gebildet werden oder sie kann über die Aufsummierung aller Spaltensummen gebildet werden. Entsprechend kann die Gesamtsumme einer tabellarisch in Spalten und Zeilen geordneten Menge grundsätzlich mittels eines doppelten Summierungsvorgangs gebildet werden. Wie auf der Folie hier im oberen Block dargestellt, können wir diese Aufsummierung der Zeilen- bzw. Spaltensummen ganz einfach über die Verwendung des Plusoperators darstellen. Im ersten Fall würden wir für die m Zeilen entsprechend m Zeilensummen bilden und diese m Zeilensummen dann über den Plusoperator aufsummieren. Bzw. im Fall der Spaltensumme würden wir bei n Spalten n Spaltensummen mittels des Plusoperators aufsummieren, und bei beiden Aufsummierungen würden wir entsprechend dann die Gesamtsumme erhalten. Eine deutlich bündigere Notation, welche im unteren Block dargestellt ist, wäre die Nutzung von zwei verschachtelten Summenzeichen. Bei der Darstellung auf der Folie bzw. den drei Darstellungen auf der Folie, die hier abgebildet sind, ist zuerst die Gesamtsumme über das Summieren aller Zeilensummen dargestellt. Der zweite Ausdruck ist die Summierung aller Spaltensummen. Und bei diesen Ausdrücken wird über die auf der Folie angezeigten Klammern die Reihenfolge der Summierungsoperationen angezeigt. Der dritte und letzte hier aufgeführte Ausdruck ist die gängigste Darstellung von Doppelsummen und verwendet eben

keine Klammern, und den würde ich an dieser Stelle hier auch noch mal unterstreichen. Das heißt, es werden keine Klammern verwandt, um die Reihenfolge explizit anzugeben. Wenngleich die Reihenfolge der Aufsummierungsoperationen implizit sich aus der Reihenfolge der Summenzeichen ergibt, spielt diese Reihenfolge, wie gesagt, aufgrund der kommutativen Eigenschaft der Summe für das Ergebnis keine Rolle und kann eigentlich dementsprechend auch vernachlässigt werden. Und entsprechend ist es eigentlich Geschmackssache, ob mensch zuerst als innere Summe die Zeilensumme bildet und diese dann aufsummiert oder ob man über die Aufsummierung der Spaltensummen geht. Hier auf dieser Folie bilden wir als innere Summe die Spaltensummen, welche dann anschließend mittels der äußeren Summe über die einzelnen Spalten aufsummiert werden. Wichtig ist mir an der Stelle vor allem, dass Sie mitnehmen, dass tabellarische Strukturen über Doppelsummen aufsummiert werden können und dass dies notationell üblicherweise entsprechend über ein doppeltes Summenzeichen notiert wird, welches klassischerweise ohne die Nutzung von Klammern geschrieben wird. Entsprechend hoffe ich, dass Sie mit solchen Doppelsummen umgehen können, und würde noch einmal abschließend diese Anwendung von Doppelsummen an unserem Beispiel der Umsatzzahlen illustrieren wollen.

Folie 14 – Gesamtsumme – Illustrationsbeispiel

Folientext

- Abbildung: Tabelle der konkreten Umsatzzahlen (siehe [Folie 5](#))

Gesamtsumme:

$$\sum_{q=1}^4 \sum_{r=1}^3 U_{rq} = 167 + 152 + 196 + 67 = 582$$

Sprechttext

Für das Bild in der Gesamtsumme in unserem Illustrationsbeispiel können wir entsprechend die Spaltensummen berechnen, indem wir U_{rq} über alle drei Zeilen aufsummieren und anschließend die resultierenden vier Spaltensummen dann wiederum im Rahmen einer doppelten Aufsummierung aufsummieren. Diese doppelte Summenbildung würde notationell über die hier auf der Folie genutzte Doppelsumme dargestellt, wo wir die Summe von $q=1$ bis $q=4$ über die Summe von $r=1$ bis $r=3$ über U_{rq} bilden. Das Resultat, diese Doppelsumme, wäre entsprechend die Summierung von 167, was der ersten Spaltensumme entspricht und was ich hier an der Stelle vielleicht noch mal aufschreiben würde. Also das ist hier die erste Spaltensumme, plus 152, das hier wäre die zweite Spaltensumme, abgekürzt mit Sp. Sigma, damit ich nicht so viel schreiben muss. Die dritte Spaltensumme wären die 196, das heißt auch hier Sp. Sigma für die dritte Spaltensumme, und die vierte Spaltensumme wären die 67, das heißt, hier hätten wir Sp. Sigma für die vierte Spaltensumme. Und die Aufsummierung dieser vier Spaltensummen ergibt dann letzten Endes die Doppelsumme oder die Gesamtsumme von 582.

Folie 15 – Rechnerleistung nutzen: Umsetzung in R – Doppelsummen

Folientext

- Zeilensummen mittels `rowSums()`
- Spaltensummen mittels `colSums()`
- Gesamtsumme mittels `sum()`

Sprechtext

Lassen Sie mich jetzt abschließend zu diesem Kapitel noch kurz auf die Bildung von Zeilen- und Spaltensummen sowie der Bildung von Doppelsummen in R eingehen.

1. Benötigte Datenstruktur

Sprechtext

Um Zeilen- und Spaltensummen bzw. Doppelsummen zu berechnen, benötigen wir grundsätzlich eine in Zeilen und Spalten geordnete Datenstruktur. In R existieren diverse solcher tabellarischen Datenstrukturen. Hier im Rahmen dieses Videos betrachten wir ein Matrixobjekt in Form eines Data Frames. Insbesondere Matrixobjekte werden wir noch ausführlicher in den zwei Kapiteln zu linearer Algebra betrachten.

2. Illustrationsbeispiel 1: Beispielobjekt Umsatztabelle

Text in R

Scriptfenster	<pre># Daten n <- 4; m <- 3 Umsatz <- as.data.frame(matrix(c(53,92,22,32,92,28,61,46,89,7,57,3),nrow=m)) row.names(Umsatz) <- c("RegionN", "RegionS", "RegionU") colnames(Umsatz) <- paste("Quartal",1:n,sep="") Umsatz</pre>
Ausgabe Konsole	<pre>> # Daten > n <- 4; m <- 3 > Umsatz <- as.data.frame(matrix(c(53,92,22,32,92,28,61,46,89,7,57,3),nrow=m)) > row.names(Umsatz) <- c("RegionN", "RegionS", "RegionU") > colnames(Umsatz) <- paste("Quartal",1:n,sep="") > Umsatz Ausgabe der Werte in Form der folgenden Tabelle</pre>

	Quartal1	Quartal2	Quartal3	Quartal4
RegionN	53	32	61	7
RegionS	92	92	46	57
RegionU	22	28	89	3

Sprechtext

An dieser Stelle werde ich entsprechend einfach den Befehl reinkopieren, welcher mir als Resultat in dem Objekt Umsatz die im Illustrationsbeispiel verwendete Umsatztabelle für drei Regionen in vier Quartalen wiedergibt. Das heißt, ich würde hier "# Daten" schreiben und dann den hier dargestellten Code einfach reinkopieren und einmal markieren und mit Strg+Enter umsetzen. Und in der Konsole können sie dann das Illustrationsbeispiel betrachten. Das heißt, wir haben die Umsatzzahlen, welche, wie gesagt, in dem Objekt Umsatz hinterlegt sind, und das ist geordnet in Spalten für die jeweiligen Quartale und in Zeilen für die jeweiligen Regionen, was dann entsprechend hier auch noch mal über die Zeilen- und Spaltennamen in R entsprechend vermerkt ist. Um in R jetzt für ein solches in Zeilen und Spalten geordnetes Objekt Spalten- und Zeilensummen zu berechnen, können Sie die im Folgenden aufgeführten und in der Folie eben schon gezeigten Befehle benutzen.

2.1. Zeilensummen mit dem rowSums()-Befehl

Text in R

Scriptfenster	# Zeilensumme rowSums(Umsatz)
Zeitweilige Einblendung	Wichtiger R-Befehl: rowSums()

Sprechtext

Zur Berechnung der Zeilensumme, welche ich auch hier noch mal über einen Hashtag vermerken werde, also "# Zeilensumme" habe ich jetzt eben gerade eingetippt, können Sie den Befehl rowSums() benutzen, das heißt, ich würde hier auch "rowSums()" eintippen. Und für die im Rahmen dieser Veranstaltung betrachteten Datenstrukturen, das heißt Data Frames oder Matrixobjekte, reicht es an dieser Stelle, dass wir nur das erste Argument dieses Befehls nutzen. Das heißt, wenn ich bei diesem Befehl rowSums() die Klammer aufmache, kann ich im Anschluss direkt das tabellarische Objekt, für welches eben die Zeilensummen berechnet werden sollen, eingeben. Das heißt, ich würde an dieser Stelle eben "Umsatz" hier in die Klammern nach dem Befehl schreiben. Was noch mal explizit zu beachten ist, ist, dass Sie bei diesem Befehl rowSums() das row, also der englische Begriff für Zeile, klein schreiben und dann bei dem Befehl Sums, englisch für Summen, das S von Sums groß schreiben.

2.2. Spaltensummen mit dem colSums()-Befehl

Text in R

Scriptfenster	# Zeilensumme rowSums(Umsatz) # Spaltensumme colSums(Umsatz)
Ausgabe Konsole	> # Zeilensumme > rowSums(Umsatz) Ausgabe der Werte in Form der ersten Tabelle unten > > # Spaltensumme > colSums(Umsatz) Ausgabe der Werte in Form der zweiten Tabelle unten
Zeitweilige Einblendung	Wichtiger R-Befehl: colSums()

Ausgabe zu >rowSums(Umsatz)

RegionN	RegionS	RegionU
153	287	142

Ausgabe zu >colSums(Umsatz)

Quartal1	Quartal2	Quartal3	Quartal4
167	152	196	67

Sprechtext

Analog können wir für Spaltensummen, welche ich jetzt hier noch mal berechnen würde und wo ich über "# Spaltensumme" noch mal vermerke, dass im Folgenden jetzt eben die Spaltensummen berechnet werden, könnten wir den Befehl colSums() nutzen. Das heißt, ich würde hier

"colSums(Umsatz)" eintippen und analog zu rowSums würde auch hier das col, was englisch abgekürzt für column, also Spalte, steht, kleingeschrieben werden, während das S von Sums auch hier an dieser Stelle wieder großgeschrieben würde. Und wenn wir diese zwei Befehle ausführen, über das Markieren und Strg+Enter drücken, dann sehen wir in der Konsole, dass die entsprechenden Zeilen- bzw. Spaltensummen gebildet wurden.

2.3. Vektorobjekte

Sprechtext

Wichtig wäre mir an der Stelle bereits vorzumerken für das Lineare-Algebra-Kapitel und für die Umsetzung, die wir gleich im Rahmen der Doppelsummen machen, dass diese Zeilen- und Spaltensummen über den rowSums()- und den colSums()-Befehl als Vektorobjekte gespeichert werden. Das heißt, wir haben hier zwei Objekte, die letzten Endes Vektoren sind, die dort hinterlegt sind, wo über den Vektoren noch mal die Zeilen- bzw. Spaltennamen markiert sind, was letzten Endes die Lesbarkeit vereinfacht. Aber wo eben wir Vektoren haben, die wir dann entsprechend für Doppelsummen, wie wir gleich sehen werden, einfach aufsummieren können, und wo wir sehen werden, wie wir mit Vektoren in R grundsätzlich umgehen können.

2.4. Anwendung sum()-Befehl in Verbindung mit Zeilen- und/oder Spaltensumme

Text in R

Scriptfenster	# Doppelsummen sum(rowSums(Umsatz)) sum(colSums(Umsatz))
---------------	--

Sprechtext

Wenn es mir jetzt darum geht Doppelsummen zu bilden bzw. Gesamtsummen, auch das würde ich noch mal hier über "#" in meiner Skriptkonsole notieren, können wir das Summieren dieser Zeilen- bzw. Spaltensummen dann entsprechend dem im vorherigen Kapitel demonstrierten Vorgehen, über die Anwendung des sum()-Befehls machen. Dieser sum()-Befehl, wie wir in E3.1 besprochen haben, summiert einfach alle Elemente von einem Vektorobjekt auf, und wir haben hier Vektorobjekte vorliegen. Das heißt, ich könnte für das Berechnen einer Doppelsumme über die Aufsummierung aller Zeilensummen wie folgt vorgehen. Ich würde "sum" schreiben, Klammer auf, "rowSums", Klammer auf und dann eben das Objekt, über das die Zeilensummen gebildet werden, "Umsatz". Was also passiert, ist, dass zuerst eigentlich die innere Summe gebildet wird, das heißt, es werden zuerst die Zeilensummen berechnet. Dort kriege ich dann eben für meine drei Zeilen die Zeilensummen und der sum()-Befehl anschließend würde dann über die drei Zeilensummen an dieser Stelle drüber summieren. Nach dem Kommutativgesetz entsprechend wäre ein zum identischen Resultat führendes Vorgehen, das ich jetzt die Spaltensummen bilde und diese aufsummiere, auch das könnte ich machen. Ich würde "sum(" jetzt "colSums(" eingeben, "Umsatz", also "sum(colSums(Umsatz))", und das wäre dann entsprechend eine Aufsummierung aller Spaltensummen.

2.5. Alleinige Anwendung des sum()-Befehls ohne Zeilen-/Spaltensumme

Text in R

Scriptfenster	# Doppelsummen sum(rowSums(Umsatz))
---------------	--

	<pre>sum(colSums(Umsatz)) sum(Umsatz)</pre>
Ausgabe Konsole	<pre>> # Doppelsummen > sum(rowSums(Umsatz)) [1] 582 > sum(colSums(Umsatz)) [1] 582 > sum(Umsatz) [1] 582</pre>

Sprechtext

Eine dritte Möglichkeit ist, dass wir, wie gesagt, alle Elemente aus dieser Tabelle aufsummieren. Das könnten wir machen, einfach über den `sum()`-Befehl, wenn wir dort ein matrixartiges Objekt eingeben. Dann würden einfach alle Elemente des matrixartigen Objektes aufsummiert werden und das würde letzten Endes auch zu dem identischen Ergebnis führen, bzw. uns die Gesamtsumme wiedergeben. Und diese drei Befehle würde ich jetzt an dieser Stelle hier entsprechend über `Strg+Enter` ausführen, und wie in der Konsole dargestellt ist, würden wir bei allen dreien Befehlen die identisch resultierende Doppelsumme bzw. Gesamtsumme sehen, von 582.

3. Einschub: `apply()`-Befehl

Sprechtext

Für diejenigen, die sich auch computational oder mit weiteren Möglichkeiten, mit weiteren Optionen der Berechnung von Spalten- und Zeilensummen und Doppelsummen in R befassen möchten, sei angemerkt, dass es natürlich grundsätzlich verschiedene Möglichkeiten gibt, um Zeilen- und Spalten- und auch Doppelsummen in R zu bilden. Insbesondere aus Gründen der numerischen Effizienz, der Geschwindigkeit, würde ich interessierten Studierenden anraten wollen, sich den oftmals numerisch recht vorteilhaften `apply()`-Befehl, ist ein recht generischer Befehl, den man auch für die Berechnung von Spalten- und Zeilensummen nutzen kann, einmal anzugucken. Aber auch das ist, wie gesagt, optional.

Folie 19 – Vielen Dank für die Aufmerksamkeit

Folientext

Inhalt und Gestaltung

- Dr. Alexander Silbersdorff

Barrierefreiheit und Gestaltung

- BaLLviHo: Dr. Nina-Kristin Meister, Katrin Lux, Thomas Finkbeiner, Kristina Schneider, Lea Dammann, Julia Berginski

Unterstützung

- Sina Ike, Miriam Panni

Abbildungen grafischer Logos

- Sign Lab Göttingen
- Zentrum für Statistik Göttingen
- Campus-Institut Data Science Göttingen
- Twillo
- Yomma
- Georg-August-Universität Göttingen

Angabe CC-Lizenz

- Folien und Videos sind unter CC BY (4.0) lizenziert - sofern nicht anderweitig angegeben.

Sprechttext

Damit bin ich am Ende dieses Lehrvideos und bedanke mich für die Aufmerksamkeit.