# Transkript zum Video Durchschnittsstatistik

aus der Reihe Daten Lesen Lernen (DaLeLe4all Homepage)

## Inhalt

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 – Durchschnittsstatistik

## Folientext

Datenlesen Lernen: Durchschnittsstatistik. Von Dr. Alexander Silbersdorff, Campus-Institut Data Science, Logo der Georg-August-Universität Göttingen.

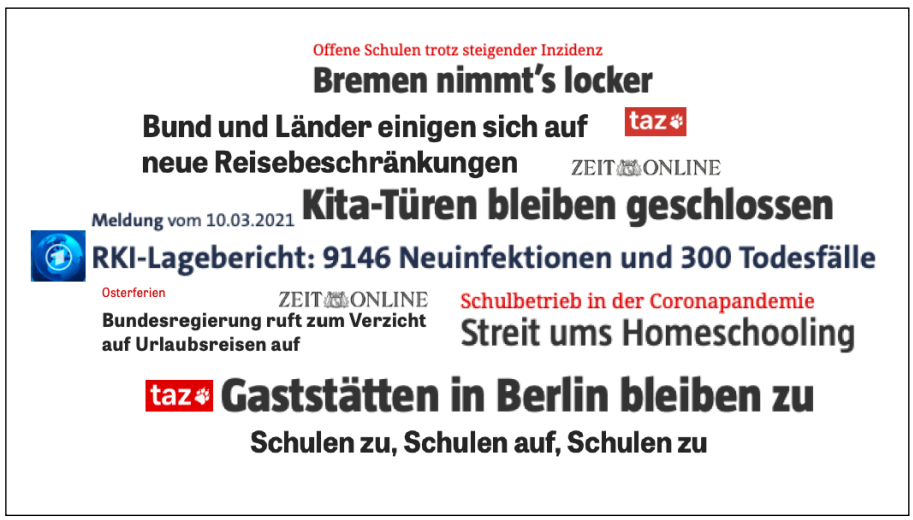
## Sprechtext

Herzlich willkommen zu diesem Daten Lesen Lernen Video zum Thema Durchschnittsstatistik. Mein Name ist Alexander Silbersdorff und was ich in diesem Video behandeln werde, ist die Anwendung von gängigen Durchschnittsmaßen zur Repräsentation von Datenstrukturen.

# Folie 2 – Datenfixierung zurzeit von Covid

## Folientext

- Abbildung: 8 Überschriften unterschiedlicher Zeitungen und Online-Medien als Beispiele der Berichterstattung zur Covid-Pandemie




## Sprechtext

Illustrieren möchte ich die Anwendung dieser Durchschnittsmaße anhand von Daten der Covid-19 Pandemie. Auf der hier dargestellten Collage von Nachrichtenüberschriften sind Beispiele für die gravierenden Einschnitte im Zusammenhang mit Covid-19 dargestellt, wie beispielsweise geschlossene Schulen und Kitas oder Einschränkungen der Reisefreiheit. Hinter diesen Überschriften bzw. den politischen Entscheidungen, welche diesen Überschriften zugrunde lagen, stand häufig eine Analyse von Maßzahlen, wovon die sieben Tage Inzidenz wohl die Prominenteste ist. Die Höhe dieser und anderer statistischer Maßzahlen entscheidet also bspw. darüber, ob Kinder in die Kitas und Schulen gehen können oder darüber, ob die Möglichkeit andere Orte und Personen zu besuchen grenzenlos gegeben oder nur eingeschränkt möglich ist. Und dies ist nur ein Beispiel wie das

alltägliche Leben vieler Menschen von dem Wert eines bestimmten Durchschnittsmaßes abhängig sein kann. Im Rahmen dieses Videos möchte ich ein Schlaglicht auf einige Grundgedanken und wichtige Aspekte statistischer Durchschnittsmaße werfen, welche, wie wir während der Covid-19 Pandemie gesehen haben, von so hoher Relevanz sein können.

## Folie 3 – Themen

### Folientext

1. Warum Maßzahlen nutzen
2. Durchschnittsmaße
3. Jenseits vom Durchschnitt

### Sprechtext

Wie ich das tun möchte ist, dass ich Ihnen zunächst ein paar Gedanken präsentieren werde, warum die Verwendung von Durchschnittsmaßen an vielen Stellen so populär und zuweilen zwangsläufig notwendig ist. Anschließend werde ich über ein paar technische Aspekte von zwei ausgewählten Durchschnittsmaßen reden. Und zuletzt möchte ich drei Aspekte aufzeigen, welche von Durchschnittsmaßen nicht, oder nur unzureichend berücksichtigt werden bzw. Ihnen exemplarisch drei Aspekte aufzeigen, wo eine kritische Reflektion dieser Durchschnittsmaße notwendig sein kann.

## Folie 4 – Lernziele

### Folientext

Nach diesem Video können Sie…

- Möglichkeiten und Limitierungen von Durchschnittsmaßen erkennen.
- das arithmetische Mittel und den Median selbstständig berechnen.
- Vor- und Nachteile des arithmetischen Mittels einschätzen.

### Sprechtext

Durch die Präsentation dieser Inhalte im Rahmen dieses Videos verfolgen wir vorrangig folgende Lernziele: Zum einen sollen Sie kennen lernen, welche Möglichkeiten die Nutzung von Durchschnittsmaßen bieten und welche Limitierungen mit der Verwendung von Durchschnittsmaßen einhergehen. Zum anderen sollen Sie für zwei ausgewählte Durchschnittsmaße - das arithmetische Mittel und den Median - das zugrundeliegende Konzept und die Form der Berechnung kennen lernen. Und zuletzt sollen Sie die Vor- und Nachteile und Voraussetzungen des Arithmetischen Mittels und des Median einschätzen können.

## Folie 5 – Thema 1

### Folientext

1. Warum Maßzahlen nutzen

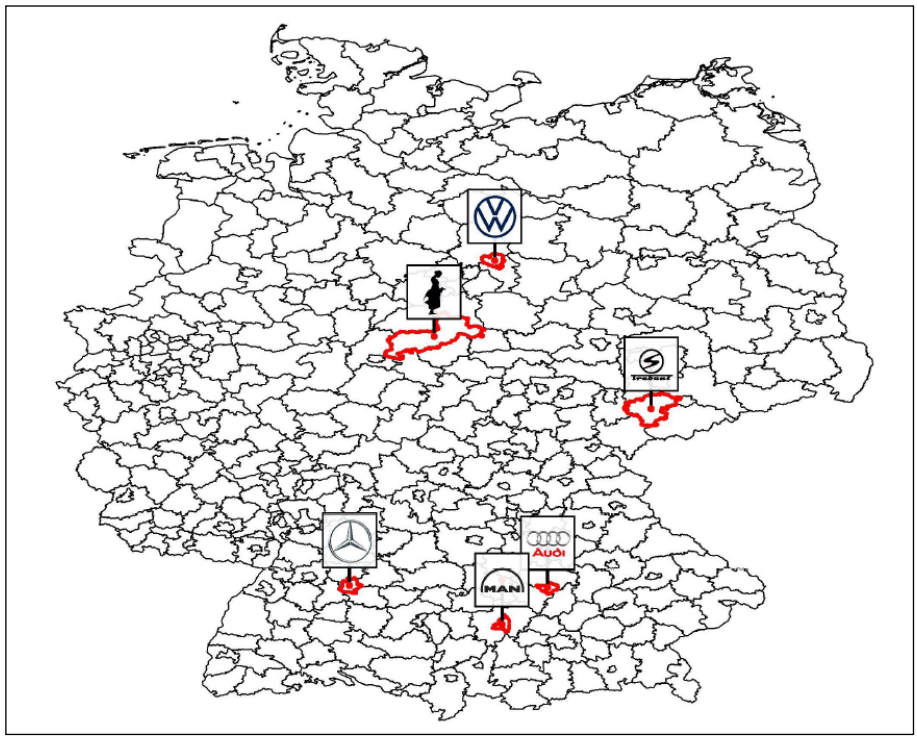
### Sprechtext

Zuerst also nun zu der Frage: Warum sind Durchschnittsmaße eigentlich wichtig und warum werden sie so häufig genutzt?

## Folie 6 – Komplexitätsreduktion durch Maße

### Folientext

- Abbildung: Karte der Gebietseinheiten in Deutschland, von denen 6 markiert sind




### Sprechtext

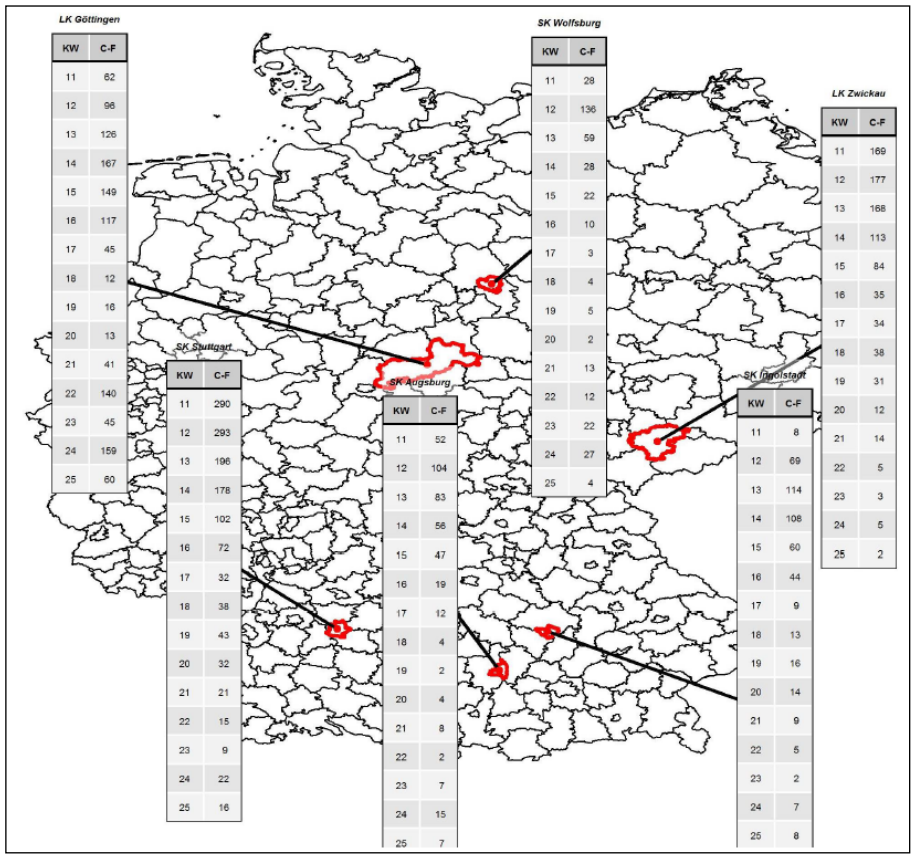
Das aus meiner Sicht wichtigste Argument für die Nutzung von Maßzahlen im allgemeinen und Durchschnittsmaßen im Besonderen ist die Reduktion von Komplexität, welche ich anhand des folgenden Beispiels illustrieren möchte. Auf der hier dargestellten Karte der Gebietseinheiten in Deutschland sehen Sie vier kreisfreie Städte und zwei Landkreise markiert: die kreisfreien Städte Wolfsburg, Ingolstadt, Augsburg und Stuttgart sind gekennzeichnet mit den vier historisch verbandelten Automarken – respektive VW, Audi, MAN und Mercedes. Die zwei Landkreise Zwickau und Göttingen sind respektive durch das Trabant Logo und, mangels nennenswerter Automobilindustrie, mit dem unmotorisierten Gänseliesel gekennzeichnet.

Für diese Städte und Kreise werde ich im Folgenden die Covid-19 Fallzahlen jener leidgeprägten Wochen des ersten Lockdowns von Mitte März bis Ende Juni 2020 präsentieren. Der einfachheitshalber werde ich anstelle von Inzidenzen, in Relation zur Einwohnerzahl, hier nur rohe Fallzahlen pro Woche betrachten.

## Folie 7 – Komplexitätsreduktion durch Maße 2

### Folientext

- Abbildung: Karte der Gebietseinheiten. Neben den sechs markierten Gebietseinheiten steht jeweils eine zugehörige Tabelle mit 15 Werten




### Sprechtext

Betrachten wir nun die Daten der ersten 15 Corona-Wochen für diese sechs Landkreise, welche Sie in der hier eingeblendeten Grafik sehen, so vermute ich, dass die meisten von uns von dem Umfang der Daten überrollt sind bzw. die Daten sind so umfangreich, dass sich die graphische Gestaltung schlichtweg als schwierig erweist. Bevor wir uns daher im Zahlenwirrwar verlieren, fokussieren wir uns vorerst auf nur zwei Städte.

## Folie 8 – Komplexitätsreduktion durch Maße 3

### Folientext

- Abbildung: Karte der Gebietseinheiten. Neben dem Stadtkreis Wolfsburg und der Stadt Ingolstadt steht jeweils die zugehörige Tabelle mit 15 Werten

### Sprechtext

Auf der linken Seite den Stadtkreis Wolfsburg und auf der rechten Seite den Stadtkreis Ingolstadt. Sie sehen also nun zweimal 15 Werte. Aber selbst bei nur zwei Tabellen mit je 15 Werten wäre meine Vermutung, dass die meisten von uns auf die Schnelle Schwierigkeiten hätten eine Frage, wie „welche dieser zwei Städte hat denn jetzt die härtere Corona Pandemie durchlitten in dem skizzierten Zeitraum“ zu beantworten.

Was dieses Beispiel also illustriert ist die Schwierigkeit komplexe Zahlenkonvolute zu fassen und vergleichen zu können. Selbst bei „nur" zwei Tabellen und „nur“ 15 Werten haben die meisten von uns wahrscheinlich schon Probleme sich ein Bild zu machen. Beim Vergleich von sechs Tabellen mit 15 Werten gilt das noch viel mehr und wenn wir für alle Landkreise und Stadtkreise in Deutschland alle 105 Tageswerte für den betrachteten Zeitraum direkt erfassen, ist die Komplexität schlichtweg enorm. Je nach unseren kognitiven Fähigkeiten und je nach dem Umfang der betrachteten Daten, brauchen wir früher oder später also zwangsläufig eine Reduktion der Komplexität, um von der

Menge an Daten nicht überfordert zu werden. Und eine solche Reduktion können eben Durchschnittsmaße liefern.

## Folie 9 – Komplexitätsreduktion durch Maße 4

### Folientext

- Abbildung: Karte der Gebietseinheiten. Über dem Stadtkreis Wolfsburg steht der Wert 25,00 und über der Stadt Ingolstadt der Wert 32,40.

### Sprechtext

In der nun hier aufgezeigten Karte sehen Sie nicht mehr je 15 Werte für unsere zwei betrachteten Städte, sondern jeweils nur noch ein Durchschnittsmaß, in diesem Fall das so genannte Arithmetische Mittel, zu welchem wir gleich nochmal kommen werden. Wir haben also 15 Zahlen pro Gebietseinheit auf eine einzige eindimensionale Zahl reduziert. Somit reduziert sich der Vergleich von zwei recht komplexen, hochdimensionalen Konstrukten von je 15 Zahlen, auf den Vergleich von zwei einzelnen Zahlen: 25 für Wolfsburg und 32 für Ingolstadt, sodass wir direkt erkennen können, dass im Durchschnitt die Corona Zahlen in Ingolstadt leicht höher waren als in Wolfsburg und somit Ingolstadt, dieser Metrik folgend, etwas härter getroffen war als Wolfsburg in Bezug auf Corona.

## Folie 10 – Komplexitätsreduktion durch Maße 5

### Folientext

- Abbildung: Karte der Gebietseinheiten. Über dem Stadtkreis Wolfsburg steht der Wert 25,00. Über Göttingen der Wert 83,20; über Zwickau der Wert 59,33; über Stuttgart der Wert 90,60; über Ingolstadt der Wert 32,40 und über Augsburg der Wert 28,13.

### Sprechtext

Und weil wir recht gut geschult sind im Vergleich von eindimensionalen Zahlen – anstelle von dem Vergleich hochdimensionalen Zahlenkolonnen, mit 15 oder mehr Zahlen – ermöglicht der Einsatz von solchen eindimensionalen Maßzahlen auch, dass wir uns direkt ein Bild für alle sechs betrachteten Kreise machen können. Wo man jetzt hier auf der hier angezeigten Grafik sieht, dass Stuttgart und Göttingen mit durchschnittlichen Fallzahlen in der 90er bzw. 80er Region traurige Spitzenreiter sind, gefolgt von Zwickau mit etwa 60 Fällen im Schnitt und mit der Gruppe Ingolstadt, Augsburg, Wolfsburg mit den niedrigen Fallzahlen in der 20er und 30er Region.

## Folie 11 – Durchschnitt als Zusammenfassung

### Folientext

- Abbildung: Schwarz-weiß Foto des Autors W.E.B. Du Bois

- ,,When you have mastered numbers, you will in fact no longer be reading numbers, any more than you read words when reading books. You will be reading meanings.'' Zitat von W.E.B. Du Bois.

### Sprechtext

Was ich in diesem Kapitel aufzeigen wollte, ist also, dass Durchschnittszahlen es uns erlauben komplexe Zahlenkonstrukte auf ein überschaubares Maß zu reduzieren. Und diese Fähigkeit der Reduktion ist nicht nur grandios nützlich, sie ist gerade in einer Welt, die uns mit einer irrsinnigen Menge an komplexen Daten konfrontiert häufig dringend notwendig. Und um die Bedeutung der Nutzung von Durchschnittsmaßen noch einmal zu unterstreichen, möchte ich ein Zitat des amerikanischen Philosophen William Du Bois - den Sie hier abgebildet sehen - anführen, der grob übersetzt sagte: "Wenn ihr das Verständnis von Zahlen gemeistert habt, werdet ihr Zahlen so lesen wie Worte in Büchern, ihr werdet nicht Wort für Wort, sondern deren gemeinsame Bedeutung erkennen." Und in diesem Prozess der Erkenntnisgewinnung können Durchschnittsmaße eine enorme Bereicherung sein, da sie uns helfen können Bedeutung aus den Zahlen zu lesen und Verständnis zu schaffen, worauf es in der Praxis häufig natürlich am allermeisten ankommt. An dieser Stelle sei aber auch bereits schon auf eine der wichtigsten Problematiken von Durchschnittsmaßen hingewiesen, auf die ich im Folgenden noch etwas ausführlicher eingehen werde. Um in der Analogie von Du Bois zu bleiben sind Durchschnittsmaße in etwa vergleichbar mit Buchklappentexte von Büchern. Im Idealfall sind sie zusammenfassend informativ, aber sie können aufgrund der mit ihnen vorgenommenen Umfangs- und Komplexitätsreduktion niemals allumfassend informativ sein. Und so wie in Bezug auf Bücher gilt, dass es sich zuweilen durchaus lohnen kann ein Buch von vorne bis hinten zu lesen und eben nicht nur den Buchklappentext, lohnt es sich mit Zahlenmengen zuweilen auch, sich nicht nur auf deren Darstellung in Form von Durchschnittsmaßen zu beschränken oder zu verlassen.

## Folie 12 – Thema 2

### Folientext

2. Durchschnittsmaße

### Sprechtext

Im zweiten Teil möchte ich Ihnen nun zwei Durchschnittsmaße kurz vorstellen und diese miteinander kontrastieren. Zuerst werde ich das Arithmetische Mittel betrachten und anschließend den Median vorstellen.

## Folie 13 – Das Arithmetische Mittel

### Folientext

- Arithmetisches Mittel als ,,das Durchschnittsmaß''
- Berechnung: $\frac{\text{Summe der Werte}}{\text{Anzahl der Beobachtungen}}$

### Sprechtext

Das Arithmetische Mittel (AM) ist das bekannteste und meistgenutzte Durchschnittsmaß und der Ausdruck "Durchschnitt" wird häufig synonym für das AM verwendet.

Das Arithmetische Mittel, wie auch der Median, gehört zu der Klasse der so gennannten statistischen Lagemaße und wie es dieser Name es schon andeutet, versuchen diese Maße i.d.R. die Lage der Daten in einer einzelnen Maßzahl zu fassen. D.h. es wird eine Zahl gesucht, welche exemplarisch für die Vielzahl an Werten gesehen werden kann, welche man versucht zusammenzufassen. Das Arithmetische Mittel als eines dieser Lagemaße berechnen Sie indem Sie die Summe aller Werte, für

welche das Mittel gebildet werden soll, berechnen und diese Summe dann durch die Anzahl der in die Berechnung einfließenden Werte teilen.

## Folie 14 – Das Arithmetische Mittel 2

### Folientext

- Arithmetisches Mittel als ,,das Durchschnittsmaß''
- Berechnung: $\frac{\text{Summe der Werte}}{\text{Anzahl der Beobachtungen}}$
- Laufender Index: *i*
- Kalenderwoche: *KW_i*
- Fallzahl: *x_i*
- Abbildung: Tabelle der wöchentlichen Corona Fallzahlen für Wolfsburg

| ***i*** | ***KW_i*** | ***x_i*** |
|---|---|---|
| 1 | 11 | 28 |
| 2 | 12 | 136 |
| 3 | 13 | 59 |
| 4 | 14 | 28 |
| 5 | 15 | 22 |
| 6 | 16 | 10 |
| 7 | 17 | 3 |
| 8 | 18 | 4 |
| 9 | 19 | 5 |
| 10 | 20 | 2 |
| 11 | 21 | 13 |
| 12 | 22 | 12 |
| 13 | 23 | 22 |
| 14 | 24 | 27 |
| 15 | 24 | 4 |

### Sprechtext

Das Ganze möchte ich anhand der bereits kurz aufgezeigten wöchentlichen Corona Fallzahlen für Wolfsburg illustrieren, welche Sie hier links als Tabelle abgebildet sehen, wobei die erste Spalte eine laufende Indexzahl ist, die zweite entsprechende die Kalenderwoche anzeigt, und die dritte Spalte dann die Corona Fallzahlen für die jeweilige Woche darstellt. D.h. mein erster Coronafallzahlenwert mit der Indexzahl 1 ist für die 11. Kalenderwoche und beträgt 28 Fälle in dieser Woche. In der zweiten Kalenderwoche haben wir die Indexzahl 2 und haben 136 Fälle in Wolfsburg registriert. Für die folgenden Wochen haben wir dann respektive 59, 28, 22, 10, 3, 4, 5, 2, 13, 12, 22, 27 und 4 Fälle.

## Folie 15 – Das Arithmetische Mittel 3

### Folientext

- Arithmetisches Mittel als ,,das Durchschnittsmaß''
- Berechnung: $\frac{\text{Summe der Werte}}{\text{Anzahl der Beobachtungen}}$
- Anzahl der Werte: *n=15*
- Summe der Werte: $\sum_{i=1}^{n} x_i = 28+136+\ldots+4 = 375$
- Arithmetisches Mittel: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{375}{15} = 25$
- Abbildung: Tabelle der wöchentlichen Corona Fallzahlen für Wolfsburg (siehe Folie 14)

### Sprechtext

In dieser Tabelle haben wir also 15 Werte insgesamt und wenn wir alle 15 Werte aufsummieren, d.h. 28, 136, 59 usw. miteinander addieren, erhalten wir eine Summe von 375. Mit diesen zwei Zahlen lässt sich dann das Arithmetische Mittel berechnen, indem wir 375 – die Summe aller Werte – durch 15 – die Anzahl aller Werte – teilen. Dies ergibt in diesem Fall genau 25 als Wert für das Arithmetische Mittel. Folglich wird 25 als jene repräsentative Zahl gesehen, welche die Lage zusammenfassend beschreibt.

## Folie 16 – Das Arithmetische Mittel 4

### Folientext

- Idee: Jener Wert, zu dem sich die Differenzen auf beiden Seiten aufheben

### Sprechtext

Nun wäre meine Hoffnung, dass Sie als wissbegierige Menschen, diese Formel nicht einfach so hinnehmen, sondern sich fragen, warum gerade diese Formel eine repräsentative Zahl ausspucken soll, welche von so vielen Menschen als das Durchschnittsmaß schlechthin gesehen wird. Und auf diese Frage gibt es eine ganze Reihe schlauer Antworten. An dieser Stelle möchte ich mich auf eine kurze Antwort beschränken. Ein Grund für die Beliebtheit des Arithmetischen Mittels ist dessen Eigenschaft, dass sich die Abweichungen der Werte vom Mittel selbst die Waage halten und somit der Mittelwert in einem physikalischen Sinne als Schwerpunkt der Daten verstanden werden kann.

## Folie 17 – Das Arithmetische Mittel 5

### Folientext

- Idee: Jener Wert, zu dem sich die Differenzen auf beiden Seiten aufheben

| ***i*** | ***KW_i*** | ***x_i*** | $x_i - \bar{x}$ |
|---|---|---|---|
| 1 | 11 | 28 | 3 |
| 2 | 12 | 136 | 111 |
| 3 | 13 | 59 | 34 |
| 4 | 14 | 28 | 3 |
| 5 | 15 | 22 | - 3 |
| 6 | 16 | 10 | - 15 |
| 7 | 17 | 3 | - 22 |
| 8 | 18 | 4 | - 21 |
| 9 | 19 | 5 | - 20 |
| 10 | 20 | 2 | - 23 |
| 11 | 21 | 13 | - 12 |
| 12 | 22 | 12 | - 13 |
| 13 | 23 | 22 | - 3 |
| 14 | 24 | 27 | 2 |
| 15 | 24 | 4 | - 21 |

### Sprechtext

Lassen Sie mich diese Eigenschaft erneut an den Corona Daten aus Wolfsburg illustrieren. Ich habe zu der Tabelle nun die Differenz zwischen dem Wert der Beobachtung und dem Arithmetischen Mittel hinzugefügt, d.h. für den ersten Wert aus der KW11, haben wir einen Wert von 28 welcher abzüglich des Wertes vom Arithmetischen Mittel, von 25, eine Differenz von 3 ergibt. In der darauffolgenden Woche haben wir eine Differenz von 111 und so weiter und so fort.

## Folie 18 – Das Arithmetische Mittel 6

### Folientext

- Idee: Jener Wert, zu dem sich die Differenzen auf beiden Seiten aufheben

| *i* | *KW_i* | *x_i* | $x_i - \bar{x}$ |
|---|---|---|---|
| 1 | 11 | 28 | 3 |
| 2 | 12 | 136 | 111 |
| 3 | 13 | 59 | 34 |
| 4 | 14 | 28 | 3 |
| 5 | 15 | 22 | - 3 |
| 6 | 16 | 10 | - 15 |
| 7 | 17 | 3 | - 22 |
| 8 | 18 | 4 | - 21 |
| 9 | 19 | 5 | - 20 |
| 10 | 20 | 2 | - 23 |
| 11 | 21 | 13 | - 12 |
| 12 | 22 | 12 | - 13 |
| 13 | 23 | 22 | - 3 |
| 14 | 24 | 27 | 2 |
| 15 | 24 | 4 | - 21 |

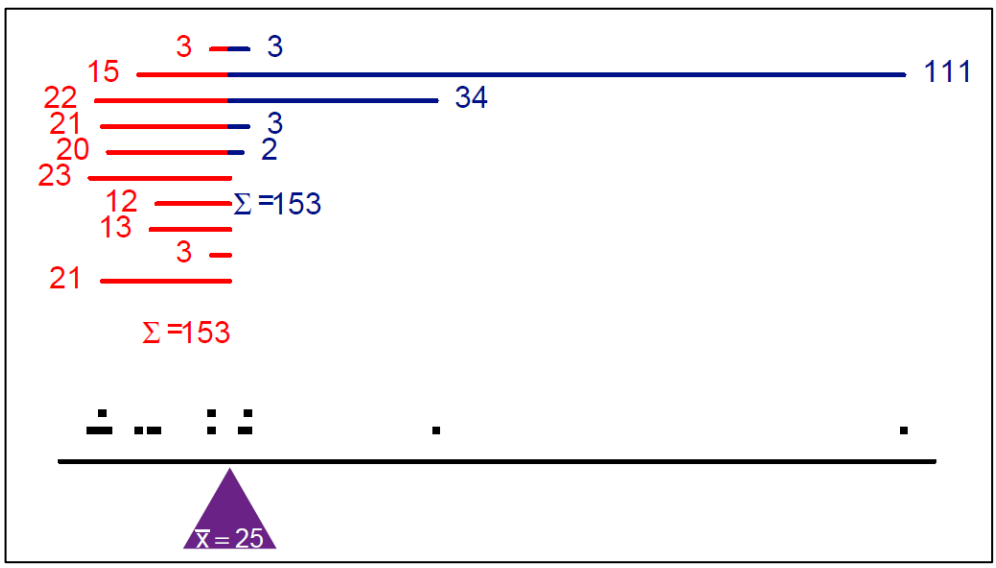
### Sprechtext

Manche Werte liegen über dem berechneten arithmetischen Mittel und deren Differenzen sind jetzt hier in Blau eingezeichnet. Die anderen Werte, deren Differenzen in Rot eingezeichnet sind, liegen unter dem berechneten Arithmetischen Mittel bzw. haben eine entsprechend negative Differenz.

## Folie 19 – Das Arithmetische Mittel 7

### Folientext

- Idee: Jener Wert, zu dem sich die Differenzen auf beiden Seiten aufheben
- Abbildung: Zwei grafische Abbildungen des Arithmetischen Mittels




### Sprechtext

Betrachten wir die Differenzen nun als Abstände von Mittelwerten, welche hier in dieser Grafik visualisiert sind, wo sie auf der linken Seite die Abstände zu Werten haben, die niedriger als 25 sind,

welche in Rot dargestellt sind, und auf der rechten Seite die Abstände zu den Zahlen, die größer waren als der Mittelwert von 25 und in Blau dargestellt sind. Wenn Sie sich diese Abstände angucken und diese Abstände aufsummieren, so erhalten Sie auf beiden Seiten eine Summe von 153. Das heißt die Summe auf der linken und der rechten Seite ist äquivalent. Das heißt das Arithmetische Mittel ist jener Punkt, der in der Mitte der Daten liegt in dem Sinne, dass die Abstände der Abweichungen auf beiden Seiten gleich sind.

# Folie 20 – Der Median

## Folientext

- Median: ein weiteres Durchschnittsmaß

## Sprechtext

Kommen wir nun zum zweiten Durchschnittsmaß, dem Median. Zwar ist der Median sicherlich nicht so gebräuchlich und landläufig auch nicht so bekannt, wie das Arithmetische Mittel, dennoch ist er in einigen Kontexten zumindest theoretisch ein deutlich geeigneteres Maß, als das Arithmetischen Mittel. Der Grundgedanke des Median ist, wenn man so will, tatsächlich noch ein wenig simpler als der des Arithmetischen Mittels. Die dem Median zugrundeliegende Logik ist es einfach jenen Wert als Lagemaß zu wählen, welcher in der Mitte der Daten steht, wenn die Werte der Größe nach sortiert sind. Lassen Sie mich dies erneut anhand von unserem Wolfsburger Fallbeispiel verdeutlichen.

# Folie 21 – Der Median 2

## Folientext

- Median: ein weiteres Durchschnittsmaß
- Berechnung: 1. Sortierung der Werte, 2. Auswahl des mittleren Wertes
- Abbildung: Tabelle der wöchentlichen Corona Fallzahlen für Wolfsburg (siehe Folie 14)

## Sprechtext

Hier sehen Sie jetzt erneut die ursprüngliche Tabelle der Fallzahlen, sortiert nach den Kalenderwochen. Für die Berechnung des Medians benötigen wir nun eine Sortierung der Werte in aufsteigender oder absteigender Reihenfolge. Das heißt unsere ursprüngliche Tabelle mit Fallzahlen wird bspw. so neu sortiert, dass die niedrigsten Fallzahlen ganz oben stehen und die größten Fallzahlen am Ende gelistet werden.

# Folie 22 – Der Median 3

## Folientext

- Median: ein weiteres Durchschnittsmaß
- Berechnung: 1. Sortierung der Werte, 2. Auswahl des mittleren Wertes

| *j* | *KW_j* | $\tilde{x}_j$ |
|---|---|---|
| 1 | 20 | 2 |
| 2 | 17 | 3 |
| 3 | 18 | 4 |
| 4 | 25 | 4 |
| 5 | 19 | 5 |
| 6 | 16 | 10 |
| 7 | 22 | 12 |
| 8 | 21 | 13 |
| 9 | 15 | 22 |

| *j* | *KW_j* | $\tilde{x}_j$ |
|---|---|---|
| 10 | 23 | 22 |
| 11 | 24 | 27 |
| 12 | 11 | 28 |
| 13 | 14 | 28 |
| 14 | 13 | 59 |
| 15 | 12 | 136 |

### Sprechtext

Diese Sortierung habe ich in der hier nun aufgezeigten Tabelle vorgenommen, so dass nun der niedrigste Wert 2 aus der 20. KW ganz oben in der Tabelle aufgeführt wird und dann die Werte der Größe nach aufsteigend bis zum höchsten Wert von 136 aus der 12. KW aufgelistet sind. Somit sind die sortierten Werte 2,3, dann zweimal die 4, 5, 10, 12, 13, zweimal die 22, 27, zweimal die 28, 59 und abschließend die 136.

## Folie 23 – Der Median 4

### Folientext

- Median: ein weiteres Durchschnittsmaß
- Berechnung: 1. Sortierung der Werte, 2. Auswahl des mittleren Wertes
- Anzahl der Werte: $n = 15$
- Mittlerer Wert bei $j = 8$
- Median: $x_{med} = \tilde{x}_8 = 13$

| *j* | *KW_j* | $\tilde{x}_j$ |
|---|---|---|
| 1 | 20 | 2 |
| 2 | 17 | 3 |
| 3 | 18 | 4 |
| 4 | 25 | 4 |
| 5 | 19 | 5 |
| 6 | 16 | 10 |
| 7 | 22 | 12 |
| **8** | **21** | **13** |
| 9 | 15 | 22 |
| 10 | 23 | 22 |
| 11 | 24 | 27 |
| 12 | 11 | 28 |
| 13 | 14 | 28 |
| 14 | 13 | 59 |
| 15 | 12 | 136 |

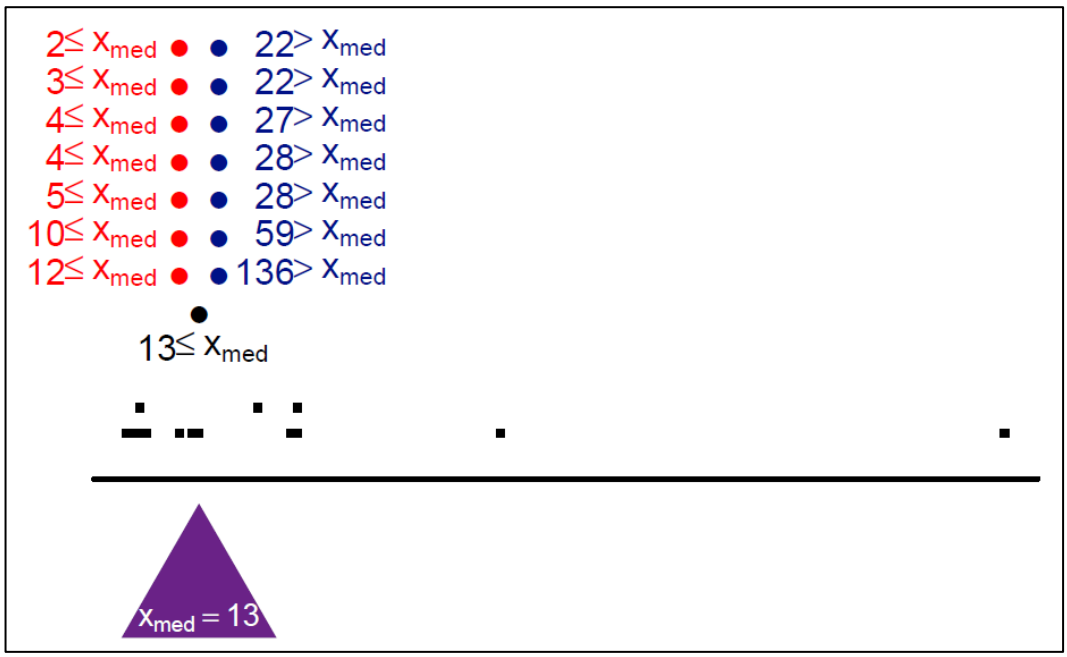
### Sprechtext

Aus dieser sortierten Reihenfolge gilt es nun den mittleren Wert auszuwählen, welcher bei 15 Werten der 8. wäre. Dieser 8. Werte stammt in unserem Fall aus der 21. KW und hätte den Zahlenwert 13. Und dieser Median Wert von 13 kann ebenso wie das Arithmetische Mittel, dessen Wert 25 war, als repräsentativ für die Lage der beobachteten Daten gesehen werden.

## Folie 24 – Der Median 5

### Folientext

- Median: ein weiteres Durchschnittsmaß
- Berechnung: 1. Sortierung der Werte, 2. Auswahl des mittleren Wertes
- Abbildung: Zwei grafische Darstellungen des Median



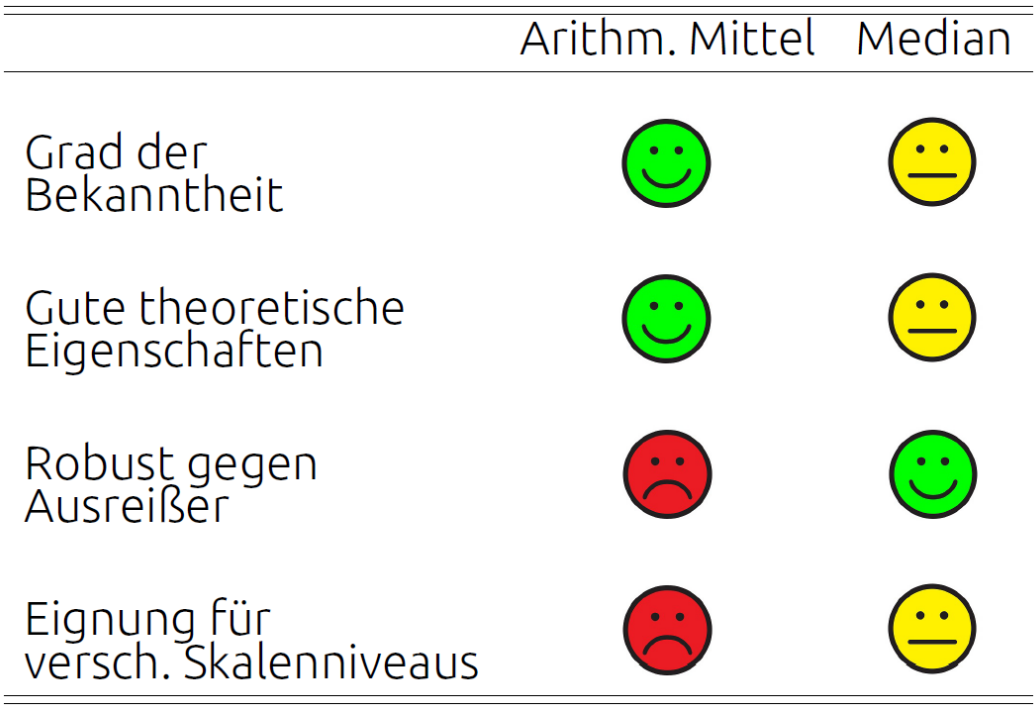

### Sprechtext

Und wenngleich die Werte für den Median häufig, aber nicht immer, ähnlich zu dem des Arithmetischen Mittels ist, ist die grundlegende Logik eine grundverschiedene, da es beim Median eben nicht um die Distanzen geht, sondern lediglich um das Ordnen und anschließende Aufteilen der Werte, was auf der hier aufgezeigten Grafik noch einmal visualisiert wird.

## Folie 25 – Wahl des Durchschnittsmaßes

### Folientext

- Abbildung: Bewertung der Eigenschaften des Arithmetischen Mittels und des Median




## Sprechtext

Hinsichtlich dieser unterschiedlichen Systematik möchte ich abschließend auf einige Vorteile und Nachteile des Arithmetischen Mittels bzw. des Median hinweisen. Zum einen ist es so, dass die meisten Menschen wahrscheinlich, ohne zweimal nachzudenken, das Arithmetische Mittel als Durchschnittsmaß akzeptieren würden und Sie somit keinerlei Erläuterungs- oder Überzeugungsarbeit dafür leisten müssen, dass dieses Durchschnittsmaß die Daten ordentlich repräsentiert, wie sie es vielleicht für den Median oder andere Durchschnittsmaße der Fall wäre. Zum anderen sei für diejenigen von Ihnen, die tiefergehend in die mathematischen Aspekte von Durchschnittsmaßen eintauchen wollen, an dieser Stelle erwähnt, dass das Arithmetische Mittel einige recht schöne mathematische Eigenschaften vorzuweisen hat, die in verschiedener Hinsicht vorteilhaft gegenüber dem Median und anderen Durchschnittsmaßen sind. Es gibt aber auch einige Nachteile des Arithmetischen Mittels im Vergleich zum Median und auch zu anderen Durchschnittsmaßen. Zum einen ist das Arithmetische Mittel sehr anfällig für Ausreißer, das heißt einzelne besonders große oder besonders kleine Zahlen können das Niveau des Arithmetischen Mittels sehr stark beeinflussen und dies kann in verschiedenen Situationen recht problematisch sein. Gegen solche Ausreißer ist der Median deutlich robuster und dementsprechend häufig das Mittel der Wahl. Ein weiterer tatsächlich häufig noch gravierender Nachteil des Arithmetischen Mittels ist die Problematik, dass das Arithmetische Mittel eigentlich nur für so genannte kardinalskalierte Variablen sinnvoll verwendet werden kann. Der Median hingegen ist etwas breiter anzuwenden, aber leider ist der Median auch nicht universell einsetzbar.

Grundsätzlich ist bei der Anwendung dieser zwei Maße – das heißt sowohl des Arithmetischen Mittels als auch des Median – zu hinterfragen, ob die dem Maß zugrundeliegende Logik, die ich vorhin vorgestellt habe, auf die jeweiligen Daten passt. Das heißt im Falle des Arithmetischen Mittels, ob dieses Ausbalancieren der Abstände zwischen den Werten nicht nur technisch möglich, sondern auch inhaltlich sinnvoll ist. Und inhaltlich sinnvoll kann es nur sein, wenn die Abstände zwischen den Werten auf deren Basis wir das Arithmetische Mittel berechnen, sinnvoll zu interpretieren sind und dies ist häufig schlichtweg nicht der Fall: Wenn Sie bspw. versuchen Mittelwerte aus Postleitzahlen der betrachteten Gebietseinheiten zu bilden, ist das zwar technisch möglich, aber es ist leider mitnichten sinnvoll, da die Abstände der Postleitzahlen inhaltlich nicht wirklich sinnvoll zu

interpretieren sind. Und folglich ist eine Anwendung des Arithmetischen Mittels hier auch nicht als sinnvoll anzusehen. Und auch der Median wäre in diesem Fall nicht wirklich sinnvoll anzuwenden, da beim Median die aufsteigende bzw. absteigende Ordnung der Werte möglich und sinnvoll sein muss – und auch dies wäre im Fall von Postleitzahlen nicht wirklich gegeben. Und gerade bei dieser Frage, ob die Maße, die wir berechnen für die inhaltliche Bedeutung der Daten wirklich sinnvoll ist, sind Sie als wissenschaftlich-kritisch denkende Menschen gefragt, da Sie in der Regel auf der inhaltlichen, interpretativen Ebene deutlich mehr Knowhow mitbringen als die zum Einsatz kommende Computersoftware, welche die technischen Berechnungen natürlich einfach umsetzen kann.

## Folie 26 – Thema 3

### Folientext

3. Jenseits vom Durchschnitt

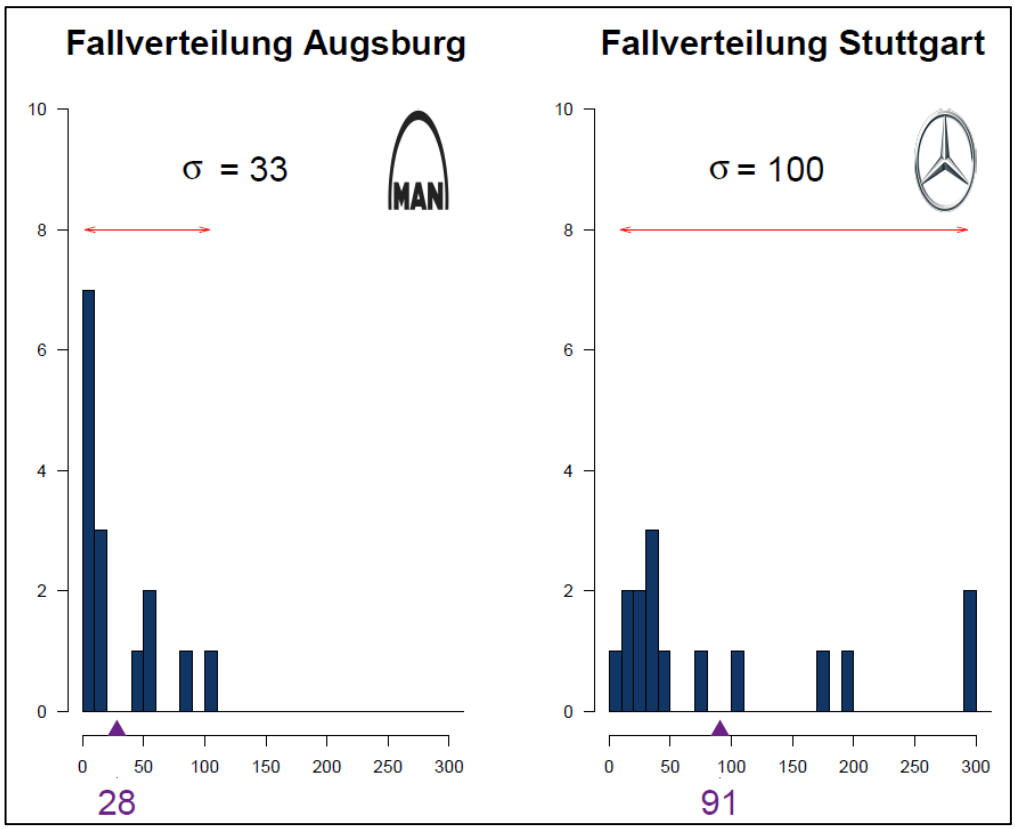
### Sprechtext

Kommen wir nun zum dritten und abschließenden Teil dieses Videos, wo ich kurz drei Aspekte jenseits der Berechnung von Durchschnittsmaßen beleuchten möchte.

## Folie 27 – Streuungsmaße

### Folientext

- Abbildung links: Säulendiagramm der Fallverteilung Augsburg
- Abbildung rechts: Säulendiagramm der Fallverteilung Stuttgart




### Sprechtext

Zuerst möchte ich auf den Aspekt der Streuung von Daten eingehen. Wie Sie bei der Berechnung der Durchschnittsmaße gesehen haben, gibt es in den Daten üblicherweise Streuung um das Arithmetischen Mittel, den Median oder ein anderes Durchschnittsmaß herum. Die Kategorie der Streuungsmaße, bspw. die Varianz oder Interquantilsabstände, geben ein Bild darüber, wie groß die

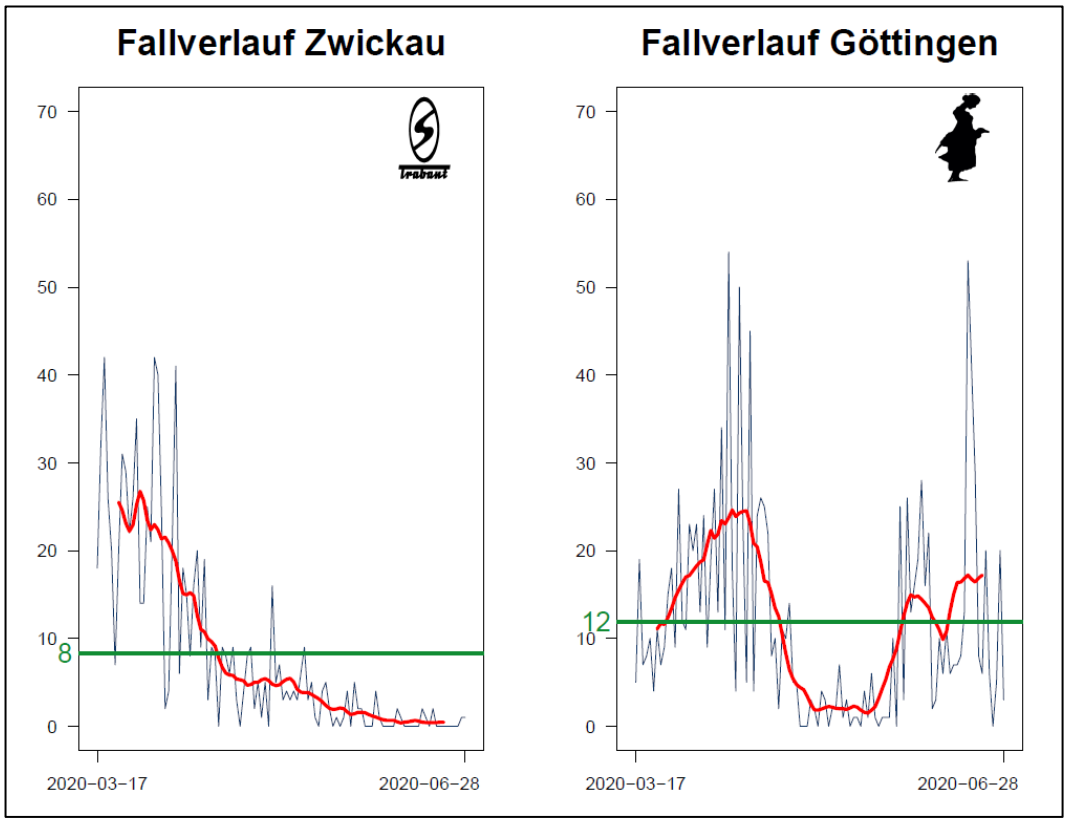
Abstände zu dem Lagemaß typischerweise sind. Um das ganze erneut an unseren Corona Fallzahlen zu illustrieren, betrachten Sie die folgende grafische Darstellung der Fallzahlen jetzt für die Städte Augsburg und Stuttgart, wobei in der Horizontalen die Fallzahlen pro Woche verortet sind, während die Höhe der Balken die Anzahl bzw. die Häufigkeit der jeweiligen Ausprägungen anzeigen.

Ohne weiter auf die Details dieser Darstellung einzugehen, ist hoffentlich zu erkennen, dass die Streuung für Augsburg auf der linken Seite geringer ist als die Streuung für Stuttgart auf der rechten Seite. Das heißt, während Augsburgs Fallzahlen recht kompakt um den Mittelwert von 28 herum fluktuieren, gibt es in Stuttgart häufiger gravierende Abweichungen vom Mittelwert. Und diese Streuung kann auch wieder durch Maßzahlen dargestellt werden, wie beispielsweise der erwähnten Varianz oder den Interquantilsabständen oder der hier auf dieser Grafik dargestellten Standardabweichung. Diese Standardabweichung beträgt für Augsburg 33 und für Stuttgart 100, sodass Sie an dieser Maßzahl auch wieder direkt ablesen können, dass die Streuung für Stuttgart höher ist als die für Augsburg. Und die Betrachtung von Maßzahlen für Streuung ist in vielen Kontexten tatsächlich von recht hoher Relevanz. In unserem Kontext bspw. ist sie dafür von Relevanz, um zu evaluieren, ob das Durchschnittsmaß die Daten recht präzise repräsentiert oder, ob es zu größeren Abweichungen davon kommen kann.

## Folie 28 – Trends und Tendenzen

### Folientext

- Abbildung links: Kurvendiagramm zum Fallverlauf Zwickau
- Abbildung rechts: Kurvendiagramm zum Fallverlauf Stuttgart




### Sprechtext

Den zweiten Aspekt, auf welchen ich eingehen möchte, ist die Berücksichtigung der zeitlichen Dimension von Daten bzw. etwaiger Trends in den Daten. Hier sehen Sie die täglichen Corona Fallzahlen aus Zwickau und Göttingen im Zeitverlauf. Dabei ist der Gesamt-Durchschnitt mit der grünen Linie eingezeichnet, welcher mit dem Wert 8 für Zwickau und 12 für Göttingen recht ähnlich

ist. Eine Betrachtung des zeitlichen Verlaufs via der in blau dargestellten Rohdaten oder des in rot dargestellten gleitenden Durchschnitts lässt hoffentlich jedoch erkennen, dass die Struktur der Verläufe sich stark unterscheidet. Während Zwickau zumindest für den betrachteten Zeitraum eine stetig abfallende Tendenz aufweist, sehen wir in Göttingen das wiederholte Aufflammen der Corona Zahlen, wie wir es im weiteren Verlauf der Pandemie leider allzu oft beobachten mussten. Diese Verläufe und Tendenzen werden durch eine ausschließliche Betrachtung des Mittelwertes verschleiert und gerade im Hinblick auf die Corona Pandemie spielen die Verläufe der Fallzahlen eine mindestens so wichtige Rolle wie deren allgemeines Niveau, welches durch ein Lagemaß repräsentiert wird. Das heißt die Betrachtung von Tendenzen, Trends und Verläufen ist an verschiedenen Stellen von hoher Relevanz und sollte auf keinen Fall vernachlässigt werden.

# Folie 29 – Grundlage der Maßzahlen

## Folientext

- Abbildung: Landkarte der Region Göttingen

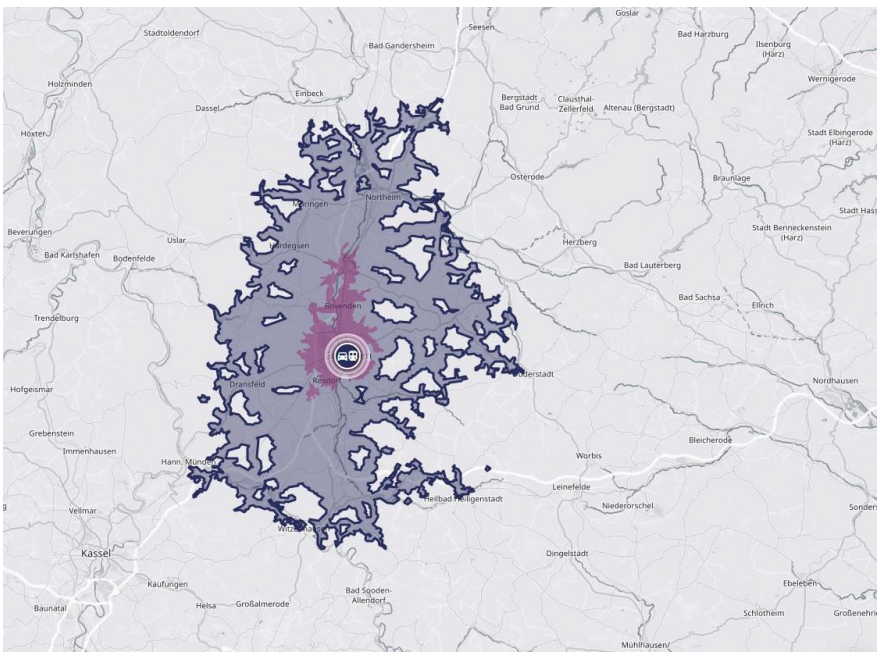
## Sprechtext

Als dritten und finalen Aspekte möchte ich ein Schlaglicht darauf werfen, was die Durchschnittsmaße eigentlich beschreiben und ob es wirklich das ist, was uns interessiert. Lassen Sie mich dies an folgendem Illustrationsbeispiel erläutern. Die Grafik, welche Sie hier auf der Folie sehen, zeigt einen klassischen Landkartenausschnitt um Göttingen herum, welcher potenziell herangezogen werden könnte, um die Distanz von Ihrem Standort zur Universitätsmedizin Göttingen zu eruieren, falls Sie beispielsweise im Falle einer Covid Infektion akute respiratorische Probleme kriegen und somit schnell in Krankenhaus müssen. Somit könnte, von leichten Verzerrungen aufgrund der Erdkrümmung einmal abgesehen, auf Grundlage Ihrer Aufenthaltsorte ermittelt werden, was ihre durchschnittlicher Luftlinien-Distanz zur Universitäts-Medizin Göttingen (UMG) ist. Es stellt sich jedoch die Frage, ob diese räumliche Abstandsmetrik das relevante Kriterium ist, wenn es darum geht zu eruieren, ob Sie im Falle einer Covid Infektion sich in hinreichender Nähe zur UMG befinden.

# Folie 30 – Grundlage der Maßzahlen 2

## Folientext

- Abbildung, Landkarte der Region Göttingen mit blauen und roten Flächenmarkierungen



### Deskription des Landkartenausschnitts

### Sprechtext

Denn im Falle von respiratorischen Problemen ist natürlich nicht die Distanz in Kilometern das entscheidende Maß, sondern viel mehr die Zeit, welche Sie benötigen, um in die UMG zu kommen und dort behandelt zu werden. Entsprechend wäre es an dieser Stelle wahrscheinlich sinnvoller, anstatt von räumlichen Distanzen, zeitliche Distanzen zu betrachten, welche auf dieser zweiten Karte in Form einer Iso-Chronischen Karte dargestellt sind, wo der blaue Bereich jene Orte anzeigt, welche nicht mehr als 30 Minuten von der UMG entfernt sind, während die rote Fläche jene Orte umspannt, welche 15 Minuten oder weniger von der UMG entfernt liegen. Bei der Betrachtung und Auswertung von Maßzahlen gilt es also nicht nur eine Reihe von Maßzahlen zu berechnen, sondern auch darum kritisch zu reflektieren, ob die Grundlage der Berechnung der Maßzahlen, die Daten auf deren Grundlage wir diese Maßzahlen berechnen, eigentlich das sind, was uns wirklich interessiert, und somit die Maßzahlen das repräsentieren, was für uns wichtig ist.

## Folie 31 – Zusammenfassung

### Folientext

In diesem Video haben Sie gelernt…

- wie Durchschnittsmaße helfen, Komplexität zu reduzieren.
- wie die beiden am häufigsten verwendeten Durchschnittsmaße berechnet werden.
- dass der sinnvolle Einsatz von Durchschnittsmaßen kontextabhängig ist.

### Sprechtext

Zuletzt möchte ich jetzt abschließend noch einmal die präsentierten Inhalte und die damit verbundenen Lernziele noch einmal zusammenfassen. Ich habe Ihnen präsentiert, wie Durchschnittsmaße zur Komplexitätsreduktion eingesetzt werden können und wie sie es uns erlauben komplexe Sachverhalte zusammenzufassen. Anschließend haben Sie das Arithmetische Mittel und den Median als zwei wichtige Durchschnittsmaße kennen gelernt. Und zuletzt habe ich über einige Aspekte jenseits von Durchschnittsmaßen, welche Sie nach Möglichkeit auch nicht aus den Augen verlieren sollten, gesprochen.

Ich hoffe mit den hier präsentierten Inhalten Ihnen einige Erkenntnisse mit auf den Weg gegeben zu haben, welche Ihnen nützen im heutigen Zeitalter der Daten den Überblick zu behalten.

## Folie 32 – Vielen Dank für Ihre Aufmerksamkeit

### Folientext

Inhalt und Gestaltung

- Dr. Alexander Silbersdorff, Dr. Jennifer Lorenz, Dr. Benjamin Säfken, Sina Ike

Barrierefreiheit und Gestaltung

- Dr. Nina-Kristin Pendzich, Katrin Lux, Thomas Finkbeiner, Susanne Martini

Unterstützung

- Dominik Becker, Nele Wolf, Christiane Gocke

Abbildungen grafischer Logos

- Logo des Sign Lab Göttingen
- Logo des Zentrums für Statistik Göttingen
- Logo des Campus-Institut Data Science Göttingen
- Logo des Projekt Daten Lesen Lernen
- Logo der Firma Yomma

- Logo der Georg-August-Universität Göttingen

### Sprechtext

Damit bin ich am Ende dieses Videos. Ich danke Ihnen für Ihre Aufmerksamkeit und danke auch den Unterstützer*innen, die bei der Erstellung dieses Videos mitgewirkt haben. Vielen Dank.

## Folie 33 – Anhang: Förderung und Copyright

### Folientext

Dieses Video wurde dank Unterstützung des Förderprogramms „Innovative Lehr- und Lernkonzepte: Innovation plus“ des Niedersächsischen Ministeriums für Wissenschaft und Kultur im Studienjahr 2021/22 umgesetzt.