





Deskriptive Statistik Kapitel 2

Prof. Dr. Andrea Wirth

Gliederung



1. Einführung und Aufgaben

Data Science und Statistik, deskriptive Statistik, Wahrscheinlichkeitsrechnung, induktive Statistik, Vorgehensweisen und Beispiele

2. Grundbegriffe, statistische Untersuchung und Darstellung der Daten Datenquellen, Merkmale, Grundgesamtheit, Stichprobe, Messskalen, Vorgehensweise bei statistischen Untersuchungen Planung, Datenerhebung und –techniken, Datenaufbereitung, tabellarische und grafische Darstellung, Interpretation



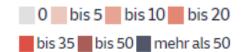
2. Eine Corona-Statistik

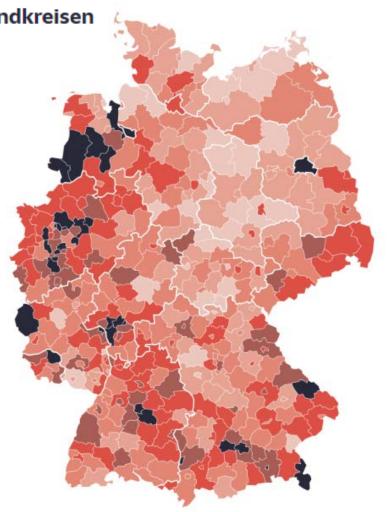


Bestätigte Coronafälle in Deutschlands Landkreisen

Was müssen Sie wissen, um die Grafik zu interpretieren?

Neue Fälle je 100.000 Einwohner in den vergangenen sieben Tagen:





Daten zuletzt aktualisiert: 13.10.20, 00:00 Uhr

Kartenmaterial: ⁰ GeoBasis-DE / BKG 2017

Quelle: Robert-Koch-Institut • Rohdaten herunterladen

2. Merkmalsträger und Grundgesamtheit



Der *Merkmalsträger* (auch Element, statistische Einheit und Untersuchungseinheit) ist der Gegenstand der statistischen Untersuchung, er ist der Träger der interessierenden statistischen Information.

Zum besseren Verständnis werden die Grundbegriffe zusätzlich am Beispiel "Altersstruktur der Mitarbeiter der Medicus-Klinik AG am 31.12.2017" illustriert: Merkmalsträger im Beispiel ist - zunächst grob gesagt - jeder einzelne Mitarbeiter der Medicus-Klinik AG am 31.12.2017. Der einzelne Mitarbeiter ist Gegenstand der Altersmessung bzw. Träger der statistischen Information Alter.

Die *Grundgesamtheit* ist die Menge aller Merkmalsträger, die übereinstimmende Abgrenzungsmerkmale besitzen.

Im Beispiel ist die Grundgesamtheit - auch hier zunächst grob gesagt - die Menge aller Mitarbeiter der Medicus-Klinik AG.

Aufgabe: wo sehen Sie bei diesen Definitionen Schwierigkeiten?

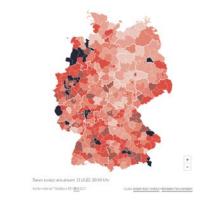


2. Anwendung auf die Corona-Statistik



Der *Merkmalsträger* (auch Element, statistische Einheit und Untersuchungseinheit) ist der Gegenstand der statistischen Untersuchung, er ist der Träger der interessierenden statistischen Information.

Die *Grundgesamtheit* ist die Menge aller Merkmalsträger, die übereinstimmende Abgrenzungsmerkmale besitzen.



2. Abgrenzung von Grundgesamtheiten



Abgrenzung von Grundgesamtheiten: Die Abgrenzung ist in sachlicher, räumlicher und zeitlicher Hinsicht vorzunehmen. Durch sie wird eindeutig geklärt, wer Merkmalsträger ist bzw. wie sich die Grundgesamtheit zusammensetzt.

Räumliche Abgrenzung:

Im Beispiel, die zum Wirkungskreis der Medicus Klink gehörenden Kliniken

 Für die zeitliche Abgrenzung ist ein Zeitpunkt oder ein Zeitraum festzulegen. Die Existenz des Merkmalsträgers an diesem Zeitpunkt bzw. in diesem Zeitraum ist entscheidend für die Zugehörigkeit oder Nicht-Zugehörigkeit zur Grundgesamtheit.

Die zeitliche Abgrenzung im Beispiel erfolgt z.B. über den Zeitpunkt 31.12.2017, 24.00 Uhr. Nur wer zu diesem Zeitpunkt Mitarbeiter war, wird in die Untersuchung einbezogen. Im Unterschied dazu spricht man beim Zeitraum von Bestands und Bewegungsmassen. Im Beispiel führen Veränderungen wie Einstellungen oder Entlassungen zur Veränderung der Bewegungs- und Bestandsmassen. Zur Bestandsmasse gehören Mitarbeiter B,C,E und F.

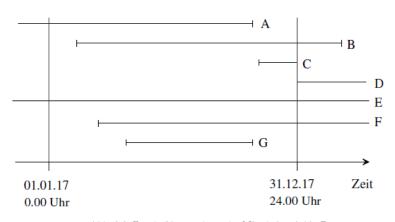


Abb. 1.1: Beschäftigungsdauer der Mitarbeiter A bis G

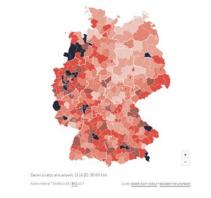
Quelle: Bourier: Beschreibende Statistik 2018

2. Abgrenzung von Grundgesamtheiten



Abgrenzung von Grundgesamtheiten: Die Abgrenzung ist in sachlicher, räumlicher und zeitlicher Hinsicht vorzunehmen. Durch sie wird eindeutig geklärt, wer Merkmalsträger ist bzw. wie sich die Grundgesamtheit zusammensetzt.

- **Räumliche Abgrenzung**:
- Für die zeitliche Abgrenzung ist ein Zeitpunkt oder ein Zeitraum festzulegen. Die Existenz des Merkmalsträgers an diesem Zeitpunkt bzw. in diesem Zeitraum ist entscheidend für die Zugehörigkeit oder Nicht-Zugehörigkeit zur Grundgesamtheit.



2. Merkmal und Merkmalswert



Die Eigenschaft des Merkmalträgers, die bei der statistischen Untersuchung von Interesse ist, wird als *Merkmal* (Prädikatsmerkmal, statistisches Merkmal, Untersuchungsvariable oder Variable) bezeichnet.

Im obigen Beispiel ist das Untersuchungsmerkmal "Alter" bei den Mitarbeitern der Medicus-Klinik AG von Interesse. - Weitere Beispiele für Merkmale der Mitarbeiter sind: Geschlecht, Jahreseinkommen, Art der Beschäftigung oder Familienstand

X = Alter des Mitarbeiters (Jahre)

Y = Geschlecht des Mitarbeiters

Z = Jahreseinkommen (€)

Der Wert, der bei der Beobachtung, Befragung, Messung oder durch einen Zählvorgang beim Merkmalsträger festgestellt wurde, heißt *Merkmalswert*.

Im obigen Beispiel ist das jeweilige Alter eines Mitarbeiters der Merkmalswert. Weitere Beispiele für Merkmalswerte sind

Merkmal	Merkmalswert
Geschlecht	männlich, weiblich
Jahreseinkommen (in €)	24.000, 61.235, 125.418,30
Familienstand	ledig, verheiratet, geschieden, verwitwet.

2. Anwendung auf die Corona-Statistik



Die Eigenschaft des Merkmalträgers, die bei der statistischen Untersuchung von Interesse ist, wird als *Merkmal* (Prädikatsmerkmal, statistisches Merkmal, Untersuchungsvariable oder Variable) bezeichnet.

Der Wert, der bei der Beobachtung, Befragung, Messung oder durch einen Zählvorgang beim Merkmalsträger festgestellt wurde, heißt *Merkmalswert*.



2. Qualitative und quantitative Merkmale



Ein *qualitatives Merkmal* liegt vor, wenn den Merkmalswerten lediglich Namen oder Klassenbezeichnungen zugeordnet werden können. Sie können nur verbal beschrieben werden und verschließen sich einer zahlenmäßigen/messtechnischen Erfassung.

abgestufte qualitative Merkmale

Merkmal	Merkmalswert	Merkmal	Merkmalswert
Beruf	Erzieherin, Lehrer, Professorin	Schulnote	Sehr gut, gut,, mangelhaft
Farbe	blau, orange, gelb grün	Vortragsweise	Langweilig,, sehr interessant
Familienstand	ledig, verheiratet, verwitwet,	Weingüte	Tafelwein, Qualitätswein,, Auslese.

Ein *quantitatives Merkmal* liegt vor, wenn es ein eine messbare Dimension besitzt oder in Mengeneinheiten ausgedrückt werden kann. Die Merkmalswerte werden also durch Zahlen ausgedrückt und können gemessen oder gezählt werden.

Merkmal	Merkmalswert
Alter (Jahre)	, 3, 19, , 55, , 85
Mitarbeiterzahl	, 6, 12, , 12.345, , 3.654.321
Gewinn (Mio.)	, 1,3, 4,6, , 12,98, , 38,21
Benzinverbrauch (Liter)	, 3,52,, 10,56,, 13,1,



2. Häufbare und nicht-häufbare Merkmale



Bei qualitativen Merkmalen sind die Merkmalsausprägungen primär durch verbale Ausdrücke charakterisiert. Ersetzt man diese durch Zahlen, d. h. führt man eine *Quantifizierung* durch, so wird ein qualitatives Merkmal formal zu einem quantitativen. Dies verwischt den Unterschied zwischen einem (echt) quantitativen und einem (qualitativen, aber) quantifizierten Merkmal jedoch nur äußerlich.

Überlegen Sie sich bitte hierfür ein Beispiel

Ein Merkmal, von dem ein Merkmalsträger mehr als einen Merkmalswert besitzen kann, heißt *häufbares Merkmal*. In einer statistischen Erhebung findet sich i.d.R. der Hinweis: Mehrfachnennungen möglich. Ein Merkmal hingegen, von dem ein Merkmalsträger nur genau einen Merkmalswert besitzen kann, heißt *nicht-häufbares Merkmal*.

Überlegen Sie sich bitte auch hierfür Beispiele

Quelle: Bourier: Beschreibende Statistik 2018

2. Diskrete und stetige Merkmale



Ein quantitatives Merkmal, das abzählbar viele Werte annehmen kann, heißt **diskretes Merkmal** (vom Nachbarn unterscheidbar). Dabei können nur bestimmte Werte und nicht alle Zwischenwerte als Merkmalswerte auftreten.

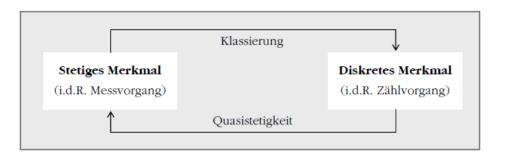
Beispiele: Kinderanzahl, Einwohnerzahl, KFZ-Bestand

Ein **stetiges oder kontinuierliches Merkmal** ist ein quantitatives Merkmal, welches überabzählbare Werte annehmen kann. Es kann in einem gegebene Intervall jeden beliebigen Wert, bzw. jedes reelle Zahl und hat überabzählbare Merkmalswerte.

Beispiele: Alter, Körpergröße, Benzinverbrauch, Gewicht, Geschwindigkeit

In der Praxis werden stetige Merkmale häufig wie diskrete Merkmale behandelt. Stetige Merkmale wie z.B. das Alter oder die Körpergröße werden oft aus messtechnischen Unzulänglichkeiten oder aus Ver-

einfachungsgründen wie diskrete Merkmale behandelt. So werden in der Regel das Alter in ganzen Jahren und die Körpergröße in vollen Zentimetern angegeben. Umgekehrt werden diskrete Merkmale manchmal wie stetige Merkmale behandelt. So werden bei Wechselkursangaben oder bei Benzinpreisen Bruchteile eines Cents angegeben.



2. Skalierung von Merkmalen



Skala	Relationen	Beschreibung	Beispiel
Nominalskala	= ≠	Einfache Aufzählung ohne Rangfolge	Geschlecht, Farbe
Ordinalskala	= ≠ < >	Natürlich Rangfolge ohne aussagekräftige Abstände	Rangplätze Bundesliga, Windstärke
Intervallskala	= \neq < > + -	Differenz zwischen 2 Werte aussagekräftig - ohne Nullpunkt	Geburtsdatum, Kalenderzeitrechnung
Verhältnisskala	= \(\neq \) \(\neq \) \(\neq \)	Verhältnis zwischen 2 Werten aussagekräftig - mit Nullpunkt	Alter, Größe, Gewicht, Einkommen

Bitte laden Sie die aktuelle Corona-Statistik über unseren Hochschule-Statistika-Account herunter:

https://de.statista.com/statistik/studie/id/71015/dokument/corona-pandemie-2019-20/

Beachten Sie, dass die Daten immer wieder aktualisiert werden



2. Skalierung von Merkmalen



Skala	Relationen	Beschreibung	Beispiel
Nominalskala	= ≠	Einfache Aufzählung ohne Rangfolge	
Ordinalskala	= ≠ < >	Natürlich Rangfolge ohne aussagekräftige Abstände	
Intervallskala	= ≠ < > + -	Differenz zwischen 2 Werte aussagekräftig - ohne Nullpunkt	
Verhältnisskala	= \(\neq \cdot > \\ + \ - \qquad * \end{array}	Verhältnis zwischen 2 Werten aussagekräftig - mit Nullpunkt	

Überlegen Sie sich bitte für die Mass-Skalen weitere Beispiele im Zusammenhang mit den Corona-Statistiken

2. Wiederholung - Vorgehensweisen in der Statistik



- **Planung:** exakte Formulierung des Untersuchungsziels und die Festlegung des Erhebungsprogramms.
- **Datenerhebung**: dient der Gewinnung des statistischen Datenmaterials.
- **Datenaufbereitung und -darstellung**: Verdichtung und Ordnung hin zu Tabellen und Schaubildern.
- **Datenanalyse und interpretation**: Anwendung mathematischer / statistischer Methoden. Die erhaltenen Ergebnisse werden interpretiert und zusammengefasst.



2. Vorgehensweisen in der Statistik – Schritt 1: Planung



- Planung: exakte Formulierung des Untersuchungsziels und die Festlegung des Erhebungsprogramms.
 - Die Planung erstreckt sich auf die Gestaltung und Durchführung der drei anschließenden Phasen. Es ist dabei festzulegen,
 - welche Merkmale bei welchen Merkmalsträgern mit welcher Technik zu erheben,
 - welche Aufbereitungsverfahren einzusetzen,
 - welche Formen der Darstellung zu wählen
 - und welche statistischen Analyseverfahren einzusetzen sind.
- 2. Datenerhebung: dient der Gewinnung des statistischen Datenmaterials.
 - Konkretisierung des Untersuchungsziels bzw. der statistischen Fragestellung

Beispiel:

Die Geschäftsleitung eines Kaufhauses erteilt den Auftrag, die Zufriedenheit der Kunden des Hauses in der vorweihnachtlichen Zeit festzustellen.

Ist diese Zielsetzung gut formuliert?



Quelle: Bourier: Beschreibende Statistik

2. Schritt 2: Datenerhebung - Untersuchungsziel



Konkretisierung des *Untersuchungsziels* bzw. der statistischen Fragestellung

Beispiel:

Die Geschäftsleitung eines Kaufhauses erteilt den Auftrag, die Zufriedenheit der Kunden des Hauses in der vorweihnachtlichen Zeit festzustellen.

Ist diese Zielsetzung gut formuliert?



2. Schritt 2: Datenerhebung - Erhebungstechnik



Erhebungstechniken:

Eine *Primärstatistik* (Primärerhebung) liegt vor, wenn für die aktuelle Untersuchung erstmalig (primär) Daten erhoben werden. Das interessierende Untersuchungsziel ist Basis für die Erhebung der Daten.

Eine **Sekundärstatistik** (Sekundärerhebung) liegt vor, wenn bei der statistischen Untersuchung auf vorliegendes Datenmaterial zurückgegriffen wird. Es werden Daten, die primär für andere Zwecke gewonnen wurden, ein zweites Mal (sekundär) verwendet.

Träger der amtlichen Statistik sind zum Beispiel das Statistische Bundesamt der BRD, die Statistischen Landesämter, die Bundesbank und die kommunalen statistischen Ämter. Bekannte Veröffentlichungen sind das "Statistische Jahrbuch für die Bundesrepublik Deutschland", die Zeitschriften "Wirtschaft und Statistik" und "Monatsbericht der Deutschen Bundesbank".

Träger der nicht-amtlichen Statistik sind zum Beispiel Wirtschaftsforschungsinstitute (DIW, IFO, ...), Markt- und Meinungsforschungsinstitute (GfK, Infratest, ...), Unternehmensverbände und Unternehmen.

	Primärstatistik	Sekundärstatistik
Kosten	hoch	niedrig
Zeitaufwand	hoch	niedrig
Zielbezug	stark	evtl. eingeschränkt
zeitliche Nähe	aktuell	evtl. weniger aktuell

Quelle: Bourier: Beschreibende Statistik

2. Schritt 2: Datenerhebung - Erhebungsumfang + und -art



Erhebungsumfang:

Eine Vollerhebung (Totalerhebung, Totalstatistik) liegt vor, wenn sämtliche Merkmalsträger der Grundgesamtheit erfasst werden. Mit der Erfassung eines jeden Merkmalsträgers erfolgt die vollständige Information über die Grundgesamtheit.

Eine **Teilerhebung** (**Stichprobe**, Teilstatistik) liegt vor, wenn nur ein Teil der Merkmalsträger der Grundgesamtheit erfasst wird.

Was eignet sich für welchen Erhebungsumfang? Brenndauer Glühlampen, Stromverbrauch privater Haushalte, Corona-Statistik, Reißfestigkeit Material, Zufriedenheit Mitarbeiter, Zufriedenheit Kunden

Erhebungsart:

Bei der **Beobachtung** erfolgt die Erfassung der Daten per Augenschein durch das Erhebungspersonal oder mit Hilfe von Messgeräten oder Aufzeichnungsvorrichtungen.

Die **Befragung** kann auf mündlichem oder schriftlichem Weg erfolgen. Die Befragung richtet sich an eine Auskunftsperson oder -stelle, die mit dem Merkmalsträger nicht identisch sein muss.

2. Schritt 3: Datenaufbereitung und -darstellung —



- Datenaufbereitung und -darstellung: Verdichtung und Ordnung hin zu Tabellen und Schaubildern.
 - Zu Beginn der Datenaufbereitung oder auch schon während der Erhebungsphase müssen die Daten *geprüft bzw. kontrolliert* werden. Die Kontrolle erstreckt sich auf die Vollständigkeit der Erfassung und der Beantwortung sowie auf die Glaubwürdigkeit bzw. Plausibilität der erfassten Daten.
 - Nach der Erhebung liegen die Daten bzw. Merkmalswerte (Urwerte, Urdaten) zunächst in Form einer sogenannten *Urliste* (*statistische Reihe*) vor. In der Urliste sind die Merkmalswerte und eventuell auch die zugehörigen Merkmalsträger nacheinander aufgereiht.
 - In der Strichliste werden alle in der Urliste enthaltenen Merkmalswerte aufgelistet. Die Anordnung der Merkmalswerte ist vom Skalenniveau abhängig.
 - Zur Erstellung der Häufigkeitstabelle werden in der Strichliste die Striche ausgezählt und dem jeweiligen Merkmalswert als Häufigkeit zugeordnet. Die *Häufig*keitstabelle gibt also die Häufigkeitsverteilung eines Merkmals wieder, d.h. man kann aus ihr ersehen, wie sich die Merkmalsträger auf die verschiedenen Merkmalswerte verteilen.

2. Schritt 3: Datenaufbereitung und -darstellung



Beispiel: Zahl der Kinder der 20 Beschäftigten der Firma Maier KG

Zahl der Kinder	Anzahl der Beschäftigten
0	7
1	6
2	4
3	2
4	1

Was ist was?

Beispiel: Zahl der Kinder der 20 Beschäftigten der Firma Maier KG

Zahl der Kinder	Anzahl der Beschäftigten
0	H# II
1	HH I
2	
3	
4	

Beispiel: Familienstand, Zahl der Kinder und Tarifgruppe der 20 Beschäftigten der Firma Maier KG

Nr.	Name, Vorname	Fam.stand	Zahl der Kinder	Tarifgruppe
01	Amberger, Heinz	ledig	0	II
02	Bauer, Regine	verheiratet	2	I
03	Bertram, Günther	geschieden	1	II
04	Dünnes, Rita	ledig	0	I
05	Engel, Erika	verheiratet	1	II
06	Frühauf, Ernst	verwitwet	1	III
07	Frisch, Anton	verheiratet	3	II
08	Gillhuber, Erwin	geschieden	0	III
09	Hell, Marion	ledig	0	II
10	Jahn, Josef	verheiratet	2	II
11	Kaps, Wolfgang	verwitwet	0	III
12	Lechner, Ernst	verheiratet	4	II
13	Maier, Waltraud	ledig	0	П
14	Mayer, Elisabeth	ledig	1	I
15	Pagler, Fritz	ledig	1	IV
16	Polzer, Herrmann	verheiratet	2	IV
17	Rabe, Armin	verheiratet	3	III
18	Reiser, Gabriele	geschieden	2	II
19	Schmidt, Heinz	verheiratet	1	IV
20	Wenisch, Willy	verheiratet	0	I

Quelle: Bourier: Beschreibende Statistik

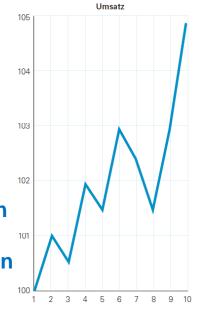
2. Schritt 3: Datenaufbereitung und -darstellung







Platz sparen durch Achsenbeschneidung

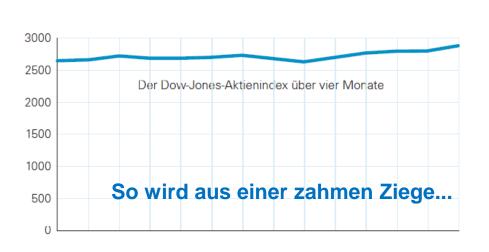




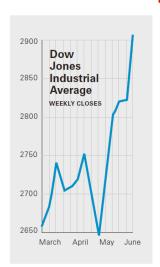
Strecken und **Stauchen**

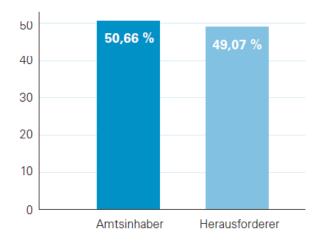
2. Schritt 3: Datenaufbereitung und -darstellung





... ein Racing Bull



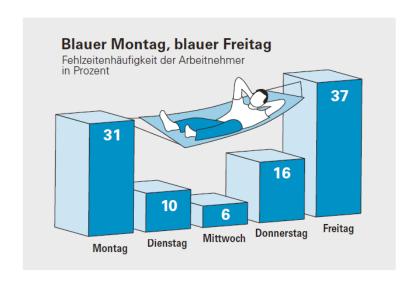




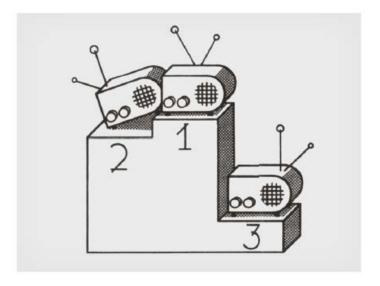
Knappes Ergebnis der Präsidentenwahl in Venezuela... Erdrutschsieg statt Zitterpartie

2. Schritt 3: Datenaufbereitung und -darstellung —





Grafik mit eingebauter Bewertung



Liegt 2 dichter an 1 als an 3?



2. Schritt 4: Datenanalyse und -interpretation

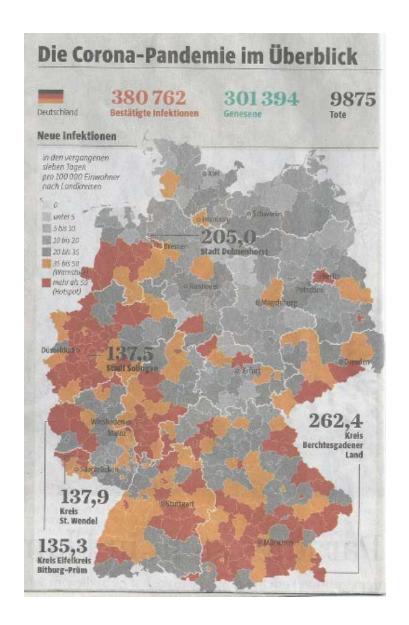


- **1. Planung**: exakte Formulierung des Untersuchungsziels und die Festlegung des Erhebungsprogramms.
- 2. Datenerhebung: dient der Gewinnung des statistischen Datenmaterials.
- 3. Datenaufbereitung und -darstellung: Verdichtung und Ordnung hin zu Tabellen und Schaubildern.
- 4. Datenanalyse und interpretation: Anwendung mathematischer / statistischer Methoden. Die erhaltenen Ergebnisse werden interpretiert und zusammengefasst.
 - 4.1. Häufigkeitsverteilung mit Kennzahlen und Parametern. Als Methoden werden hier *Mittelwerte*, Quantile oder *Streuungsparameter* eingesetzt
 - 4.2. Relationen von Zahlen und deren Gegenüberstellung, z.B. an *Verhältnis- und Indexzahlen*
 - 4.3. **Zeitreihenanalyse** mit den Methoden der Trendermittlung und der Ermittlung wiederkehrender Schwankungen
 - 4.4. Zusammenhang oder Abhängigkeit zwischen Merkmalen und ggf. deren Intensität. Eingesetzt werden hier die *Regressions- und Korrelationsanalyse*.

Und genau damit geht es jetzt in dieser Vorlesung weiter

2. Datenanalyse und -interpretation: Corona

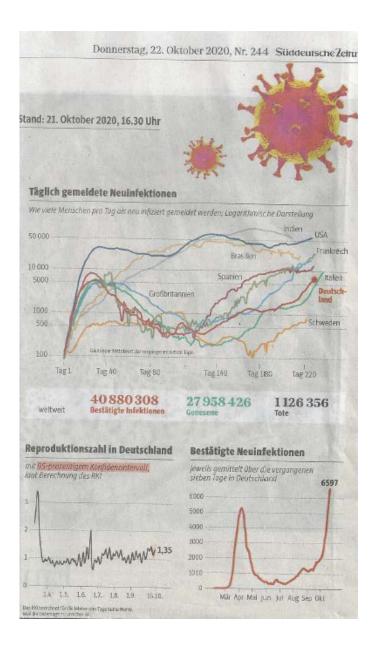






2. Datenanalyse und -interpretation: Corona







2. Datenanalyse und -interpretation: Corona







Was sollten Sie in diesem Kapitel mitgenommen haben? —

- Tieferes Verständnis der Grundbegriffe der Statistik.
- Verstehen, was Merkmalsträger, Grundgesamtheiten, Merkmale und Merkmalswerte sind.
- Die unterschiedlichen Skalierungen und Ausprägungen von Merkmalen verstehen und an Beispielen anwenden bzw. unterscheiden können.
- Die typischen statistischen Vorgehensweisen und Prozesselemente auf statistische Untersuchungen anwenden können.
- Klare Ziel-Definitionen statistischer Untersuchungen formulieren können.
- Die Datenerhebung mit Ihren unterschiedliche Erhebungstechniken verstehen und auf Ihre Fragestellung transferieren können.
- Die elementaren Methoden und auch Möglichkeiten der Datenaufbereitung verstehen und die Fallstricke interpretieren können.
- Die wesentlichen begrifflichen und methodischen Grundlagen für die deskriptive Statistik verstehen.

