

1. Vorlesung Statistik I

Daten in der Psychologie



We are happy to share our materials openly:

The content of these [Open Educational Resources](#) by [Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München](#) is licensed under [CC BY-SA 4.0](#). The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

Herzlich Willkommen!



Dr. Philipp Sckopke
philipp.sckopke@psy.lmu.de



Philipp Sterner
p.sterner@psy.lmu.de



Dr. Florian Pargent
florian.pargent@psy.lmu.de



Thomas Reiter
T.Reiter@psy.lmu.de



Prof. Dr. Felix Schönbrodt
felix.schoenbrodt@psy.lmu.de

Organisatorisches

Material

- Sie finden die meisten Materialien die für dieses Modul notwendig sind auf unserer zentralen Übersichtsseite <https://shiny.psy.lmu.de/oer>.
- Dort finden Sie
 - Vorlesungsfolien
 - Aufzeichnungen der Vorlesung vom letzten Jahr (inhaltsgleich)
 - Übungsblätter für die UKs mit Lösungen
 - Videos zur Lösung der Übungsblätter
 - Ein Videotutorial für den Einstieg in die Software R (UKs der ersten beiden Wochen)
 - Ein schriftliches Tutorial für die Verwendung von R in allen unseren Lehrveranstaltungen im BSc und MSc.

UKs (Unterricht in Kleingruppen)

- Wir versuchen, die UKs auf Ihre Bedürfnisse abzustimmen. Es gibt zu jedem Vorlesungsinhalt **Übungsblätter**, die zur Vertiefung und praktischen Anwendung der gelehrteten Inhalte konzipiert sind.
Es stehen Ihnen für Ihre Vorbereitung direkt auch **Lösungen** zu den Übungsblättern zu Verfügung, sowie **Videos** in denen die Lösung auch noch mal durchgesprochen wird. In den **UK-Terminen vor Ort** haben wir dann Gelegenheit, auf alle Fragen einzugehen, die bei der Bearbeitung der Übungsblätter angefallen sind.
- Natürlich beantworten wir dort gerne auch Fragen die sich noch auf frühere Vorlesungsinhalte beziehen.
- Idealerweise kommen Sie also **vorbereitet** ins UK:
 - Entweder mit konkreten Fragen zur Lösung der Übungsblätter oder
 - mit Fragen zu den Vorlesungsinhalten
 - (oder beides)
- Der Hauptfokus eines UKs wird zunächst erstmal die Vorlesung der jeweils vorangegangenen Woche sein, danach wird alles andere besprochen

Prüfung

- Die Prüfung bezieht sich auf die **Inhalte der Vorlesung und des Seminars** (Übungsaufgaben sind prüfungsrelevant!).
- Die Prüfung besteht aus Fragen mit **offenem** Antwortformat.
- Der **Prüfungstermin** im **Wintersemester 2023/24** ist am **Freitag, 9.2.24 um 14.00 Uhr**.
- Die Zuteilung zu den verschiedenen Klausurräumen erfolgt erst kurz vor dem Prüfungstermin wenn alle Anmeldungen erfolgt sind.
- Zur Prüfung ist als Hilfsmittel neben einem **Taschenrechner** auch ein so genanntes **Cheat Sheet** erlaubt. Hierauf dürfen Sie alles notieren, was Sie gerne in die Prüfung mitnehmen möchten.
- Das Cheat Sheet umfasst dabei maximal eine DIN A4–Seite (doppelseitig) und ist von Ihnen selbst handschriftlich (also Stift + Papier, keine digitale Anfertigung) verfasst.

- **Prinzipiell:** Stoff der Prüfung sind die **Inhalte der Vorlesung und der Seminare**. (Das ist auch der Grund, warum unsere Folien so ausführlich und deshalb manchmal gar nicht mal so schön sind.)
- **Zur Vertiefung:** Gehen Sie in die Bibliothek (sofern möglich, sonst online) und schauen Sie sich dort die Statistik-Lehrbücher an. Finden Sie heraus, mit welchem Sie am besten zurechtkommen, d.h. welche der Lehrbücher **auf Ihre Fragen, Ihre Grundkenntnisse und Ihren Zugang zur Statistik am besten zugeschnitten sind**. Auch im Internet gibt es viel qualitativ hochwertiges Material zu allen Inhalten der Statistik.
- **Zum Arbeiten mit R:** Zu R als primäre Software für statistische Analysen gibt es im Internet unzählige Hilfen. Speziell für die Lehrveranstaltungen unseres Lehrstuhls haben wir ein Dokument erstellt, das bei den meisten notwendigen Schritten zur Datenverarbeitung und –analyse Hilfestellung gibt: <https://shiny.psy.lmu.de/tutorials/>

Sprechstunde

- Wir haben einmal pro Woche eine Stunde Sprechstunde (Termine siehe Homepage).
- Gerne Beantworten wir **konkrete Verständnisfragen** zu den inhaltlichen Themen.
- Für Studierende mit Behinderung oder chronischer Erkrankung hat die LMU zusätzlich eine eigene Beratungsstelle. Informieren Sie sich dort, ob Sie ein **Recht auf Nachteilsausgleich** in der Klausur haben.

Schieben der Prüfung

- Ein Schieben der Prüfung ist erfahrungsgemäß fast nie eine gute Idee. Das Verstehen von Statistik I klappt am allerbesten im ersten Semester. Auch wenn es schwer ist, aber in den Folgesemestern machen es die Umstände häufig nur noch schwerer.
Bitte sprechen Sie uns an, bevor Sie eine solche Entscheidung treffen. Vielleicht können wir mit Ihnen eine bessere Idee erarbeiten.

Einleitung

Einige Antworten:

- Um Ergebnisse **wissenschaftlicher Studien verstehen, interpretieren und bewerten** zu können, da diese die Grundlage aller psychologischen Anwendungsfächer sind.
 - Um zu bewerten, ob sich Ergebnisse von wissenschaftlichen Studien erfolgreich therapeutisch umsetzen lassen.
 - Um die Wirksamkeit klinischer Interventionen einschätzen zu können.
 - ...
- Einige psychologische Studien sind methodisch sehr schlecht und deren Ergebnisse und Schlussfolgerungen nicht haltbar, werden aber trotzdem gelehrt.
- Wenn Sie sich mit diesen Studien nicht selbst kritisch auseinandersetzen können, müssen Sie immer alles glauben was Ihnen als die Wahrheit präsentiert wird.
- Damit tragen Sie implizit zur weiteren Verbreitung nicht belastbarer Studienergebnisse sowohl innerhalb der Psychologie als auch in der allgemeinen Öffentlichkeit bei.

Datenbeschreibung und **Daten**analyse:

- Module Statistik I und Statistik II (erstes und zweites Semester)

Datengenerierung: Theorie und Praxis des Messens von psychologischen Phänomenen

- Diagnostik I (drittes Semester)
- Diagnostik II (viertes Semester)

Erhebungsmethoden

- Seminar Methoden der **Daten**erhebung (erstes Semester)

Anwendung (Lehrveranstaltungen auch von anderen Lehrstühlen):

- Empirisch-psychologisches Praktikum I (zweites Semester) und II (fünftes Semester)
- ...
- Bachelorarbeit (sechstes Semester)

Daten

VP	Klasse	Sprint [s]	Anzahl Liegestütz	Rumpfbeuge [cm]
VP1	5a	4,56	12	9
VP2	5a	4,34	20	-1
VP3	5a	4,60	13	-9
VP4	5a	4,54	13	2
VP5	5a	4,22	16	12
VP6	5a	4,29	21	8
VP7	5a	4,18	19	-1
VP8	5a	3,94	15	5
VP9	5a	4,00	17	12
VP10	5a	4,19	20	12
VP11	5a	4,42	16	11
VP12	5a	4,15	15	-4
VP13	5a	4,31	15	-4
VP14	5a	4,53	16	5
VP15	5a	4,50	15	4

Vorläufige Definition:

Von Daten spricht man immer dann, wenn die Wirklichkeit mithilfe eines Erhebungsinstrumentes abgebildet wird. Daten sind das Ergebnis - der Output - dieser Erhebungsinstrumente.

Offene Fragen:

- Was sind „Erhebungsinstrumente“?
- Welche „Wirklichkeit“ wollen wir abbilden?
- Was bedeutet „Wirklichkeit abbilden“?

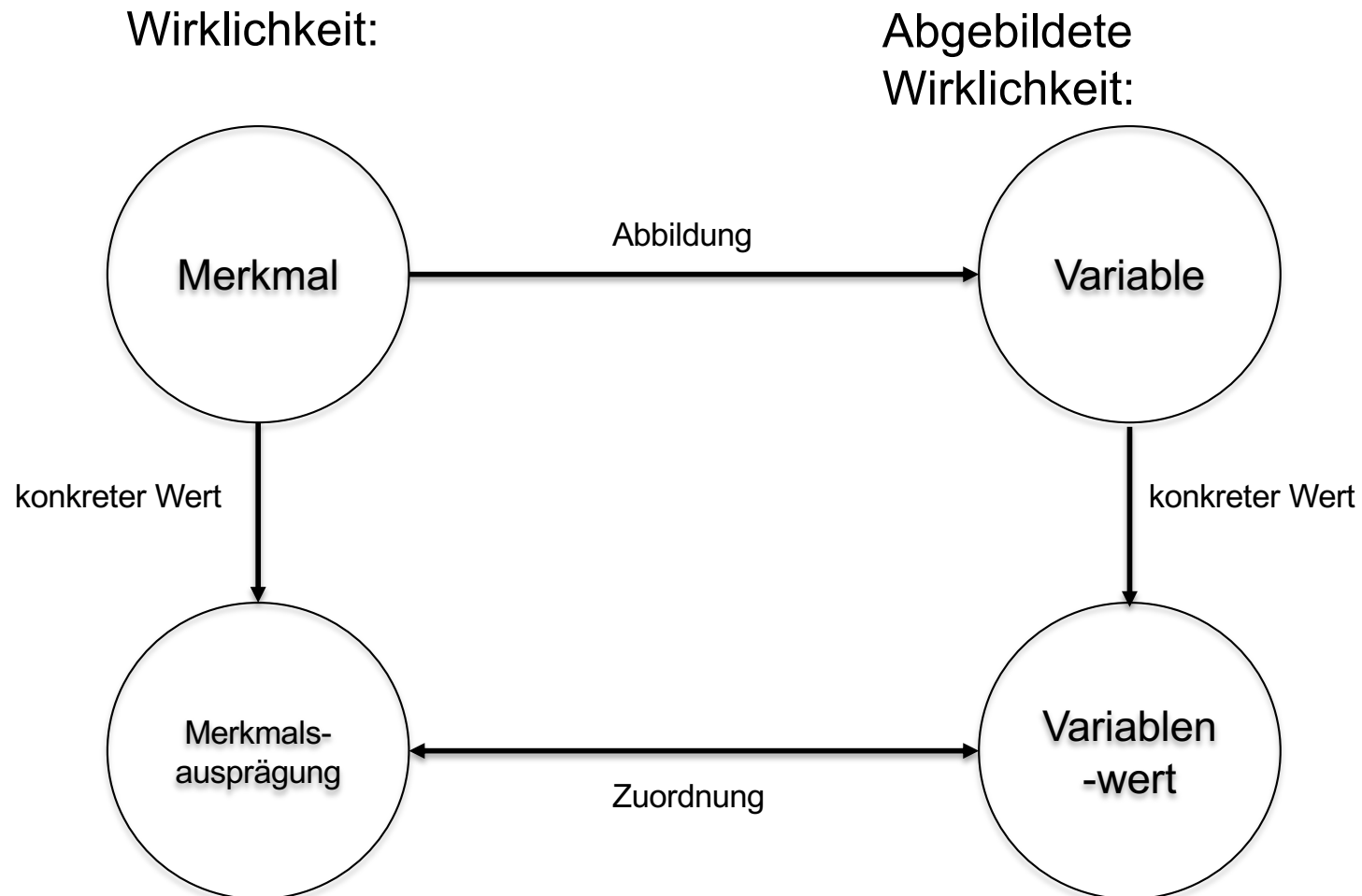
Was sind „Erhebungsinstrumente“?

- Beispiele für Erhebungsinstrumente:
 - Beobachtung
 - Fragebogen
 - Psychologischer Test
 - Physikalische Messinstrumente
- Beispiele für Daten, die aus diesen Erhebungsinstrumenten resultieren könnten:
 - Schüler*in A meldet sich dreimal im Deutsch-Unterricht.
 - Schüler*in A gibt an, blaue Augen zu haben.
 - Das Ergebnis der Testung der Leistungsmotivation von Schüler*in A ergibt einen Testwert von 123 Punkten.
 - Schüler*in A ist 175 cm groß.

Welche „Wirklichkeit“ wollen wir abbilden?

- Wir interessieren uns in den meisten Fällen nur für bestimmte Aspekte der Wirklichkeit: Für Untersuchungsobjekte und deren Eigenschaften.
- Die Untersuchungsobjekte werden auch **Merkmalsträger*innen** genannt. Merkmalsträger*innen sind in der Psychologie meist (Versuchs-)Personen. Sie können aber z.B. auch Schulklassen, Tage oder Bundesländer sein.
- Für uns interessante Eigenschaften dieser Merkmalsträger*innen werden **Merkmale** genannt.
Beispiele: Schulabschluss, Größe, Intelligenz.
- Konkrete Ausprägungen von Merkmalen werden **Merkmalsausprägungen** genannt. Merkmalsausprägungen des Merkmals Schulabschluss sind z.B. kein Abschluss, Hauptschulabschluss, mittlerer Schulabschluss, Fachhochschulreife, Abitur.

- Was bedeutet „Wirklichkeit abbilden“?
- In unserem Kontext bedeutet dies Repräsentation der Merkmale durch mathematische Größen mithilfe von Erhebungsinstrumenten.
Beispiele: Kategorisierung des Merkmals Studienabschluss durch die Begriffe „Hauptschulabschluss“, „Abitur“, ... ; Messung des Merkmals Gewicht in der Einheit kg.
- Diese mathematischen Größen werden **Variablen** genannt.
- Konkrete Ausprägungen von Variablen werden **Variablenwerte** genannt.
Variablenwerte können entweder Begriffe oder Zahlen sein.
Konkrete Variablenwerte können konkreten Merkmalsausprägungen zugeordnet werden und umgekehrt.



- Wirklichkeit:
 - Merkmalsträger*in (Synonym: Untersuchungseinheit, -objekt)
 - Merkmal
 - Merkmalsausprägung (Synonym: Merkmalswert)
- Abgebildete Wirklichkeit:
 - Variable
 - Variablenwert (Synonym: Messwert)

Schulischer Alltag: Erfassung des Wissens von Schüler*innen mithilfe einer von einer Lehrkraft entworfenen Klausur.

- Die **Merkmalsträger*innen** sind in diesem Beispiel ...
- Das **Merkmal**, das abgebildet werden soll, ist ...
- Die **Merkmalsausprägung** ist ...
- Die **Variable** ist ...
- Der **Variablenwert** ist ...

Wichtige Begriffe: Beispiel II

- Die **Merkmalsträger*innen** sind in diesem Beispiel die Schüler*innen.
- Das **Merkmal**, das abgebildet werden soll, ist das Wissen der Schüler*innen.
- Die **Merkmalsausprägung** ist der Grad des Wissens einer Schüler*in.
- Die **Variable** ist die Note in der Klausur.
- Der **Variablenwert** ist die konkrete Note (z.B. 6) einer Schüler*in in der Klausur.

Variablen

VP	Klasse	Sprint [s]	Anzahl Lie
VP1	5a	4,56	15
VP2	5a	4,34	20
VP3	5a	4,60	15
VP4	5a	4,54	15
VP5	5a	4,22	10
VP6	5a	4,29	20
VP7	5a	4,18	15
VP8	5a	3,94	15
VP9	5a	4,00	15
VP10	5a	4,19	20
VP11	5a	4,42	10
VP12	5a	4,15	15
VP13	5a	4,31	15
VP14	5a	4,53	10
VP15	5a	4,50	15

Merkmalsträger*innen

Variablenwerte

- Mithilfe der eingeführten Begriffe können wir nun den Begriff Daten etwas genauer definieren:
- Daten = Variablenwerte einer Gruppe von Merkmalsträger*innen auf einer oder mehreren Variablen.

- Bislang:
 - Wichtige Begriffe, die im Zusammenhang mit der Erhebung von Daten verwendet werden
- Jetzt:
 - Arten von Variablen

Arten von Variablen

Man kann unterscheiden zwischen

- diskreten und kontinuierlichen (stetigen) Variablen
- kategorialen und metrischen Variablen
- unabhängigen und abhängigen Variablen

Die Unterscheidung zwischen diskreten und kontinuierlichen (synonym: stetigen) Variablen bezieht sich auf die **Anzahl an möglichen Variablenwerten**:

- **Diskrete Variablen**
 - können eine (theoretisch gesehen) **endliche (oder unendliche) Anzahl an unterschiedlichen Variablenwerten** annehmen.
 - Beispiele: Studienfach, Augenfarbe, Anzahl an Fehltagen.
- **Kontinuierliche (stetige) Variablen**
 - können eine (theoretisch gesehen) **unendliche Anzahl an unterschiedlichen Variablenwerten** (mit unendlich feinen Abstufungen) annehmen.
 - Beispiele: Größe, Gewicht, Temperatur.

Die Unterscheidung zwischen kategorialen und metrischen Variablen bezieht sich auf die in den Variablenwerten **abgebildeten Beziehungen zwischen den Merkmalsträger*innen hinsichtlich des Merkmals:**

- **Kategoriale Variablen** bilden **Unterschiede** oder **Rangreihen** ab.
- **Metrische Variablen** bilden **Differenzen**, **Verhältnisse** oder **Anzahlen** ab.

- Einige Merkmale können durch Bildung von **unterschiedlichen Kategorien** abgebildet werden.
- Beispiele:
 - Erkrankung: gesund vs. erkrankt
 - Haarfarbe: blond vs. braun vs. schwarz
- Man sagt auch, dass Variablen dieser Art **Nominalskalenniveau** aufweisen
- Die Variablenwerte von Variablen mit Nominalskalenniveau sind meist Begriffe, welche die unterschiedlichen Kategorien bezeichnen.
- Möglich ist jedoch auch die Abbildung in Zahlen. Man spricht in diesem Zusammenhang auch von **Kodierung**.
 - Beispiel: gesund = 0, erkrankt = 1
 - Welcher Zahlenwert welcher Kategorie zugeordnet wird, ist beliebig, solange die Unterschiedlichkeit der Kategorien durch die Zuordnung erhalten bleibt
 - Merkmalsträger*innen mit unterschiedlichen Merkmalsausprägungen sollten also unterschiedliche Zahlen zugeordnet werden und Merkmalsträger*innen mit denselben Merkmalsausprägungen dieselben Zahlen.

- Einige Merkmale können unter Erhaltung von **Rangreihen** abgebildet werden.
- Beispiele:
 - Platzierung in einem Wettkampf: 1. Platz, 2. Platz, 3. Platz, etc.
 - Schulabschluss: Hochschulabschluss, Abitur, Mittlere Reife
- Man sagt auch, dass Variablen dieser Art **Ordinalskalenniveau** aufweisen
- Die Variablenwerte von Variablen mit Ordinalskalenniveau können Begriffe oder Zahlen sein (aber Zahlen sind in der Regel praktischer).
- Welcher Zahlenwert welchem Rang zugeordnet wird, ist beliebig, solange die Rangreihen durch die Zuordnung erhalten bleiben.
- Merkmalsträger*innen mit höheren Merkmalsausprägungen sollten also größere Zahlen zugeordnet werden.
- Beispiele: Hochschulabschluss = 3, Abitur = 2, Mittlere Reife = 1 oder
Hochschulabschluss = 579, Abitur = 385, Mittlere Reife = 187

- Einige Merkmale können unter Erhaltung von **Differenzen** abgebildet werden.
- Beispiele:
 - Temperatur
 - Intelligenz?
- Man sagt auch, dass Variablen dieser Art **Intervallskalenniveau** aufweisen
- Die Variablenwerte von Variablen mit Intervallskalenniveau sind stets Zahlen.
- Welcher Zahlenwert welcher Merkmalsausprägung zugeordnet wird, ist beliebig, solange die Größenordnung der Differenzen durch die Zuordnung erhalten bleiben. Diese Beliebigkeit ist durch die Beliebigkeit des Nullpunkts und der Maßeinheit bedingt.
- Beispiel:
 - Der Unterschied zwischen Person A und Person B in der Merkmalsausprägung ist größer als der Unterschied zwischen Person C und Person D.
 - Person A bekommt den Wert 10, Person B den Wert 5, Person C den Wert 8 und Person D den Wert 7 zugewiesen
 - Die Differenz der Werte von Person A und B ($10 - 5 = 5$) ist größer als die Differenz der Werte von Person C und D ($8 - 7 = 1$)

- Einige Merkmale können unter Erhaltung von **Verhältnissen** abgebildet werden.
- Beispiele:
 - Größe
 - Gewicht
- Man sagt auch, dass Variablen dieser Art **Verhältnisskalenniveau** aufweisen
- Die Variablenwerte von Variablen mit Verhältnisskalenniveau sind stets Zahlen.
- Welcher Zahlenwert welcher Merkmalsausprägung zugeordnet wird, ist beliebig, solange die Verhältnisse durch die Zuordnung erhalten bleiben. Diese Beliebigkeit ist durch die Beliebigkeit der Maßeinheit bedingt.
- Beispiel:
 - Person A springt dreimal so weit wie Person B
 - Person A bekommt den Wert 540(cm) und Person B den Wert 180(cm) zugewiesen ($540 / 180 = 3$)
 - Alternativ z.B.: Person A bekommt den Wert 5.4(m) und Person B den Wert 1.8(m) zugewiesen ($5.4 / 1.8 = 3$)

- Einige Merkmale können unter Erhaltung von **Häufigkeiten** abgebildet werden.
- Beispiele:
 - Anzahl an Fehltagen
 - Anzahl an Büchern zuhause
- Man sagt auch, dass Variablen dieser Art **Absolutskalenniveau** aufweisen
- Die Variablenwerte von Variablen mit Absolutskalenniveau sind stets Zahlen.
- Welcher Zahlenwert welcher Merkmalsausprägung zugeordnet wird, ist eindeutig festgelegt und entspricht der jeweiligen Häufigkeit.
- Beispiel:
 - Person A fehlt drei Tage, Person B fehlt acht Tage.
 - Person A bekommt den Wert 3 und Person B den Wert 8 zugewiesen.

- Das Skalenniveau einer Variable muss mathematisch und empirisch nachgewiesen werden!
- Mit den mathematischen und empirischen Bedingungen, unter denen psychologische Merkmale durch metrische Variablen erfasst werden können, beschäftigt sich die repräsentationale Messtheorie und teilweise auch die psychologische Testtheorie (VL im dritten Semester).
- Im Allgemeinen gehen Psycholog*innen jedoch einfach ohne Nachweis davon aus, dass psychologische Variablen, die mithilfe von psychologischen Tests (z.B. IQ-Tests) erfasst wurden, Intervallskalenniveau aufweisen.
- Ob es tatsächlich möglich ist, psychologische Merkmale durch metrische Variablen zu erfassen, ist jedoch eine bis heute weitestgehend ungeklärte Frage.

Weitere Anmerkungen:

- Es ist möglich, Variablen zu erstellen die weniger differenziert sind als es das zugrunde liegende Merkmal eventuell erlauben würde, z.B. eine Gendervariable mit zwei Ausprägungen „weiblich“ und „männlich“. Ob dies statistisch und inhaltlich sinnvoll ist, hängt immer vom Einzelfall der jeweiligen wissenschaftlichen Fragestellung ab. Die Verwendung einer solchen Variable impliziert nicht unbedingt, dass man davon ausgeht, dass Geschlecht ein diskretes, binäres Merkmal darstellt, sondern eher, dass Sie die Modellierung des Merkmals als binäre Variable für ausreichend halten.
- Es ist jedoch nicht möglich, eine Variable zu erstellen die differenzierter ist, als es das zugrunde liegende Merkmal erlaubt. Wird dies übersehen, entsprechen (numerische) Unterschiede zwischen Variablenwerten keinen echten Unterschieden im Merkmal.
- Die Diskussion, welche Struktur bestimmte Merkmale aufweisen, ist für Statistik I und II wenig relevant, da wir die entsprechenden Variablen wie bereits erwähnt pragmatisch als gegeben voraussetzen. Wir gehen also implizit davon aus, dass die Variablen die wir verwenden im entsprechenden Kontext sinnvoll sind.

- Variablen mit Nominalskalenniveau: Kategorien
 - Variablen mit Ordinalskalenniveau: Rangfolgen
 - Variablen mit Intervallskalenniveau: Differenzen
 - Variablen mit Verhältnisskalenniveau: Verhältnisse
 - Variablen mit Absolutskalenniveau: Häufigkeiten
- } kategoriale Variablen
- } metrische Variablen

- Die Unterscheidung zwischen unabhängigen und abhängigen Variablen bezieht sich auf ihre **jeweilige Rolle in einer Hypothese**:
- **Unabhängige Variablen (kurz: UV)**
 - sind Variablen, von denen man vermutet, dass sie **andere Variablen beeinflussen**.
- **Abhängige Variablen (kurz: AV)**
 - sind Variablen, von denen man vermutet, dass sie **von anderen Variablen beeinflusst werden**.
- Beispiel:
 - Hypothese: „Alkoholkonsum auf Parties steigert die Extraversion“
 - UV: Alkoholkonsum
 - AV: Extraversion
- Wichtig: Es hängt von der konkreten Hypothese ab, ob eine Variable unabhängig oder abhängig ist.

- Die Unterscheidungen zwischen
 - diskreten und kontinuierlichen Variablen
 - kategorialen und metrischen Variablen
 - unabhängigen und abhängigen Variablensind weitestgehend unabhängig voneinander.
- So können Variablen zum Beispiel
 - diskret, kategorial und unabhängig sein (z.B. Studienfach in einer geeigneten Hypothese)
 - diskret, metrisch und abhängig sein (z.B. Anzahl an Fehltagen in einer geeigneten Hypothese)
 - usw.

- Diese Vorlesung:
 - Wichtige Begriffe, die im Zusammenhang mit der Erhebung von Daten verwendet werden
 - Arten von Variablen
- Nächste Vorlesung:
 - Einführung in die Deskriptive Statistik

Zusammenfassung

- Daten sind die Ergebnisse von Beobachtungen, Fragebögen, psychologischen Tests oder physikalischen Messinstrumenten mit deren Hilfe die Ausprägung von Merkmalen untersuchter Merkmalsträger abgebildet werden soll.
- Unterschiedliche Merkmale werden dabei durch unterschiedliche Variablen in den Daten repräsentiert.
- Variablen lassen sich basierend auf ihren Eigenschaften auf unterschiedlichen Dimensionen klassifizieren, z.B. als diskrete vs. kontinuierliche, kategoriale vs. metrische oder abhängige vs. unabhängige Variablen.
- Diese Klassifizierung kann dann dabei hilfreich sein, sich für adäquate Methoden der Statistik zu entscheiden (Ausblick auf folgende Vorlesungen).