Statistik für Wirtschaftsinformatiker
Einführung

Dr. Wolfgang Böhme

Oktober 2010

Gliederung

- **1. Einführung**
- 2. Deskriptive univariate Verteilungsanalyse
- 3. Deskriptive Korrelations- und Regressionsanalyse
- 4. Deskriptive Zeitreihenanalyse
- 5. Verhältniszahlen / Indexzahlen
- 6. Einführung in ausgewählte multivariate Analysemethoden

- Statistiksoftware: SPSS

einige andere Ansichten zur Statistik

1. Traue keiner Statistik, die Du nicht selbst gefälscht hast!- Churchill
2. Statistik ist die Wissenschaft der systematischen Lüge.-Klaus Brunnstein, *1937; deutscher Informatiker
3. Wenn man den Kopf in der Sauna hat und die Füße im Kühlschrank, sprechen Statistiker von einer angenehmen mittleren Temperatur.- Franz Josef Strauß
4. Die Statistik ist wie ein Bikini: Was sie enthüllt ist vielversprechend, was sie verbirgt ist wesentlich! - Aaron Levenstein; New Yorker BWL-Professor
5. Manche (Politiker, Manager, etc...) benutzen die Statistik, wie der Betrunkene die Straßenlaterne: zum Festhalten, nicht zur Erleuchtung!
6. Statistik ist ein Verfahren, das gestattet, ungewisse Daten mit der Genauigkeit von Hundertstel-Prozent auszudrücken.
7. Wenn ein Jäger einmal an einem Hasen links und rechts vorbeischießt, dann ist der Hase im Durchschnitt tot.

1. Einführung - Gliederung

- **1.1 Aufgaben der Statistik**
- 1.2 Grundbegriffe
- 1.3 Etappen einer statistischen Untersuchung

Aufgaben und Gebiete der Statistik

Methoden und Verfahren für die Erhebung, Aufbereitung und Analyse der Daten zu entwickeln und anzuwenden und die aus der Anwendung resultierenden Ergebnisse zu interpretieren

Gebiete:

- Deskreptive Statistik (beschreibende Statistik)
- Induktive Statistik (schließende Statistik)
- (Wahrscheinlichkeitsrechnung)

deskreptive Statistik

- Vollständige Erhebung der Daten des Untersuchungsobjektes
- Aufbereitung des erhobenen Datenmaterials; übersichtliche aggregierte Form (Tabellen, Grafiken...)
- Analyse der aufbereiteten Daten (Berechnung von Kennzahlen, Erkennen von Gesetzmäßigkeiten ...)
- Sachbezogene Interpretation der Analyseergebnisse

induktiven Statistik

Situation

- Erhebung der Daten des Untersuchungsobjektes ist unvollständig (Teilerhebung)
- Informationen über das Untersuchungsobjektes liegen nicht vollständig vor, sondern nur für eine Teilgesamtheit

Ziel:

Aus den Daten über die Eigenschaften in der Teilgesamtheit sind Rückschlüsse auf die Eigenschaften in der übergeordnete Gesamtheit zu ziehen.

Problem:

Rückschluß ist mit Fehlerrisiko behaftet.

Fehlerrisiko kann unter bestimmten Voraussetzungen mit Erkenntnissen aus der Wahrscheinlichkeitsrechnung quantifiziert werden.

induktiven Statistik

Teilgebiete der induktiven Statistik:

- Entwicklung und Anwendung von Verfahren der Stichprobenziehung
- Entwicklung und Anwendung von Methoden zur Schätzung unbekannter statistischer Parameter
- Entwicklung und Anwendung von Methoden zum Test von Hypothesen
- Entwicklung und Anwendung von Methoden zur Prognose

Wahrscheinlichkeitsrechnung

Der Begriff Stochastik stammt aus dem Griechischen und heißt soviel wie „Kunst des Mutmaßens“. Mathematische Stochastik beschäftigt sich mit der Beschreibung und Untersuchung von Zufallsexperimenten.
Zufallsexperiment:

- Es gibt mehrere mögliche Ergebnisse (Ausgänge) des Experiments.
- Das Ergebnis bei einer Durchführung ist nicht mit Sicherheit vorhersehbar.
- Das Experiment ist unter den gleichen Randbedingungen wiederholbar.

Aufgabe:

Ausmaß der Sicherheit, mit der ein bestimmter Ausgang eines Zufallsexperiment eintritt (Wahrscheinlichkeit), zu quantifizieren.

1. Einführung - Gliederung

- 1.1 Aufgaben der Statistik
- **1.2 Grundbegriffe**
- 1.3 Etappen einer statistischen Untersuchung

1.2. Grundbegriffe

- **1.2.1 Statistische Einheit, statistische Gesamtheit**
- 1.2.2 Statistische Merkmale, Merkmalsausprägung
- 1.2.3 Messen, Messwerte
- 1.2.4 Statistische Skalen
- 1.2.5 Transformation von Messwerten

Statistische Einheit, statistische Gesamtheit

Eine **statistische Einheit** ist das kleinste Element in der Statistik. Eine statistische Einheit ist der Träger von Informationen bzw. Eigenschaften, die für eine statistische Untersuchung von Interesse sind.

Bezeichnung: Z_i

Synonyme: Merkmalsträger, Untersuchungseinheit, Element

Eine endliche Menge $Z = \{Z_i, i = 1, 2, ..., n\}$ wohl unterschiedener, sachlich, örtlich und zeitlich gleich abgegrenzter statistischer Einheiten Z_i heißt **statistische Gesamtheit** Z vom Umfang n.

Synonyme: statistische Masse, Grundgesamtheit

1.2. Grundbegriffe

- 1.2.1 Statistische Einheit, statistische Gesamtheit
- **1.2.2 Statistische Merkmale, Merkmalsausprägung**
- 1.2.3 Messen, Messwerte
- 1.2.4 Statistische Skalen
- 1.2.5 Transformation von Messwerten

Statistische Merkmale, Merkmalsausprägung

Eine Eigenschaft einer statistische Einheit Z_i , die Grundlage bzw. Gegenstand einer statistische Untersuchung ist, heißt **statistisches Merkmal**.

Bezeichnung: X_i, Y_i, Z_i *bzw.* X, Y, Z

Synonyme: Variable, Eigenschaft

Eine Aussage über ein Merkmal bzw. über eine Eigenschaft einer statistische Einheit Z_i heißt **Merkmalsausprägung**.

Bezeichnung: x_i, y_i, z_i *bzw.* x, y, z

Synonyme: Realisation, Datum, Beobachtung

1.2 Grundbegriffe

- 1.2.1 Statistische Einheit, statistische Gesamtheit
- 1.2.2 Statistische Merkmale, Merkmalsausprägung
- **1.2.3 Messen, Messwerte**
- 1.2.4 Statistische Skalen
- 1.2.5 Transformation von Messwerten

Messen, Messwerte

Unter **Messen** versteht man das Zuordnen von Zahlen (Messwerten) zu Merkmalsträgern nach einer bestimmten Vorschrift.

Ein Merkmal heißt **stetig**, wenn zwischen je zwei Messwerten stets ein weiterer Messwert liegt. Ist das nicht der Fall, heißt das Merkmal **diskret**.

Der **tatsächliche Messwert** ist der exakte Zahlenwert, der mit einem idealen Messinstrument, das unendlich genau misst, dem Merkmalsträger zugeordnet wird. Der **beobachtete Messwert** ist der gerundete Zahlenwert, der mit einem realen Messinstrument, das ungenau und fehlerbehaftet misst, dem Merkmalsträger zugeordnet wird.

1.2. Grundbegriffe

- 1.2.1 Statistische Einheit, statistische Gesamtheit
- 1.2.2 Statistische Merkmale, Merkmalsausprägung
- 1.2.3 Messen, Messwerte
- **1.2.4 Statistische Skalen**
- 1.2.5 Transformation von Messwerten

Statistische Skalen

Die Menge $M = \{m_j, j = 1, 2, ..., m\}$ aller m theoretisch möglichen bzw. aller m empirisch beobachteten und wohl voneinander unterschiedlichen Merkmalsausprägungen x_i eines statistischen Merkmals X, die an den n Merkmalsträgern Z_i einer Statistischen Gesamtheit $Z = \{Z_i, i = 1, 2, ..., n\}$ erhoben werden können bzw. erhoben wurden, heißt **Zustandsmenge M** des statistischem Merkmals X.

Beachte: Die Gesamtheit der beobachteten Merkmalsausprägungen x_i eines statistischen Merkmals kann kleiner als die Zustandsmenge des statistischen Merkmals X sein.

wichtig für den Skalenbegriff

Statistische Skalen

Eine relationstreue Abbildung von Merkmalsausprägungen eines Erhebungsmerkmales auf eine Zeichen- bzw. Zahlenmenge heißt **statistische Skala**.

Eine relationstreue Abbildung auf eine Zahlenmenge heißt **statistische Messskala**

Dabei muß gelten:

- Merkmalsträger, die bezüglich des Merkmales gleich sind, wird ein und dieselbe Zahl zugeordnet
- Merkmalsträger, die bezüglich des Merkmales unterschiedlich sind, werden verschiedene Zahlen zugeordnet

Statistische Skalen

Bemerkungen:

- Zahl oder Zeichen (ausgewählte Skalen benötigen keine Zahlendarstellung)⇒ Unterscheidung in qualitative und quantitative Merkmale
- qualitativen Merkmalen können Zahlenwerte zugeordnet werden.
- Wenn Zahlen, dann kann zwischen diskreten und stetigen Merkmalen unterschieden werden
- Verarbeitung von Merkmalen wird durch Zahlenwerte stark vereinfacht

Statistische Skalen

Relationen zwischen den Zahlen einer statistischen Skala:
(mathematische Operationen sind nur bei Zahlendarstellung möglich !!)

- Verschiedenheit [Vergleich auf Identität] (zulässige mathematische Operationen: $=, \neq$)
- Ordnung [Größenvergleiche] (zulässige mathematische Relationen: $<, >$)
- Abstand [benachbarte Zahlen haben gleiche Abstände] (zulässige mathematische Operationen: $+, -$)
- Verhältnis [benachbarte Zahlen haben gleiches Verhältnis] (zulässige mathematische Operationen: $*, /$)

Skalenniveaus:

- Nominalskala: Verschiedenheit
- Ordinalskala: Verschiedenheit, Ordnung
- Intervallskala: Verschiedenheit, Ordnung, Abstand
- Verhältnisskala: Verschiedenheit, Ordnung, Abstand, Nullpunkt

Statistische Skalen - Nominalskala

Eine statistische Skala, mit der lediglich die Gleichheit oder die Verschiedenheit von Merkmalsausprägungen zum Ausdruck gebracht werden kann, heißt **Nominalskala**

Beispiel: Geschlecht (männlich, weiblich)

statischische Gesamtheit: Personengruppe
statische Einheit: einzelne Person
Merkmal: Geschlecht
Merkmalsausprägung: männlich, weiblich oder eins von beiden
Zustandsmenge: männlich, weiblich

für mathematische Operationen ist eine Umkodierung notwendig:
0 - männlich, 1 - weiblich

Statistische Skalen - Ordinalskala

Eine statistische Skala, mit der sowohl die Gleichheit oder die Verschiedenheit als auch eine Rangfolge von Merkmalsausprägungen zum Ausdruck gebracht werden kann, heißt **Ordinalskala**

Beispiel: Notenvergabe: sehr gut, gut, ... oder 1,2,...

statischische Gesamtheit: Personengruppe
statische Einheit: einzelne Person
Merkmal: Note im Fach Statistik
Merkmalsausprägung: sehr gut, gut, ... oder 1,2,...
Zustandsmenge: sehr gut, gut, ... oder 1,2,...

Statistische Skalen- Messskalen

Eine statistische Skala, diemit Hilfe der reellen Zahlen sowohl die Gleichheit oder die Verschiedenheit als auch eine Rangfolge und die mess- und zählbaren Unterschiede (Abstand, Vielfaches) für Merkmalsausprägungen zum Ausdruck gbringen kann, heißt **metrische Skala**

Unterscheidungen metrischer Merkmalsausprägungen:

(a) nach Art der Operationen

- Intervallskaliertes Merkmal ($+, -$)
- Verhältnisskaliertes Merkmal ($*, /$)

(b) nach dem Charakter der Zustandsmenge

- diskretes Merkmal
- stetiges Merkmal

Statistische Messskalen - Beispiele

Beispiele (nach mathematischen Operationen):

- Intervallskala: Temperaturskala in °C
  Differenzenbildung möglich, aber nicht 20°C ist doppelt so warm als 10°C
- Verhältnisskala: Zinssätze in Prozent
  Verhältniszahlen können gebildet werden,
  z.B. 4 Prozent sind doppelt soviel wie 2 Prozent

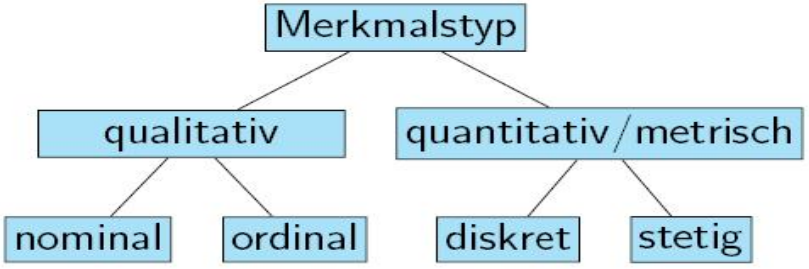
Statistische Messskalen - Beispiele

Beispiele (nach Zustandsmenge):

- Anzahl der Kinder in einer Familie
  Zustandsmenge: endlich, ganzzahlig
- diskretes Merkmal: Buchpreis in EUR und Cent
  Zustandsmenge: endlich, wird häufig als stetiges Merkmal approximiert.
- stetiges Merkmal: Zinssätze in Prozent
  in einem Intervall ist jede relle Zahl potentiell möglich.
  Zustandsmenge: unendlich viele Werte in einem Intervall.

Statistische Skalen

Klassifikation von Merkmalen: **Merkmalstypen**



Statistische Skalen

Bemerkungen:

- Nominal- und Ordinalskalen nicht unbedingt Abbildung auf Zahlenmengen (Qualitative Merkmale – Quantitative Merkmale)
- Nominal- und Ordinalskalen werden auch als kategoriale Skalen bezeichnet
- Intervall-, Verhältnisskalen werden auch als Kardinalskalen oder metrische Skalen bezeichnet.
- Merkmal mit nur zwei möglichen Ausprägungen: dichotomes Merkmal
- in SPSS werden folgende Skalen unterschieden: nominal, ordinal und metrisch (Skala)

Statistische Skalen - einfache Beispiele

Merkmal	*Auspraegung*	
Familienstand	*verheiratet*, *ledig*, ...	
Alter	$t \geq 0,\ t \in R$	
Alter	$t \geq 0,\ t \in N$	
Alter	$0 - 10, 11 - 20, \ldots$	
Farbe	*rot*, *gelb*, ...	
Farbe	*Wellenlaenge in* (μm)	
Klausurnote	$0, 1, \ldots, 15$(*Punkte*)	
Klausurergebnis	*bestanden*, *nicht bestanden*	

1.2. Grundbegriffe

- 1.2.1 Statistische Einheit, statistische Gesamtheit
- 1.2.2 Statistische Merkmale, Merkmalsausprägung
- 1.2.3 Messen, Messwerte
- 1.2.4 Statistische Skalen
- **1.2.5 Transformation von Messwerten**

Transformation von Messwerten

Definition: Eine Transformation ist eine eindeutige Zuordung von Variablen (Jedem Argument x wird durch die Funktion f geau ein Funktionswert y. bzw. $f(x)$ zugeordnet:

$$f : x \longrightarrow y = f(x)$$

Skalenniveau und zulässige Transformation:

Skalenniveau	**Transformationstyp**	**Eigenschaft der Transformation**
Nominalskala	*eineindeutige Transf.*	$x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$
Ordinalskala	*strengmonotone Transf.*	$x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$
Intervallskala	*positivlineare Transf.*	$\lvert x_1 - x_2\rvert = \lvert x_3 - x_4\rvert \Rightarrow$ $\lvert f(x_1) - f(x_2)\rvert = \lvert f(x_3) - f(x_4)\rvert$
Verhaeltnisskala	*Streckung*	$x_1 = 0 \Rightarrow f(x_1) = 0$

Transformation von Messwerten - Illustration

1. eineindeutige Transformation
2. streng monotone Transformation
3. positiv lineare Transformation
4. Streckung

1. Einführung - Gliederung

- 1.1 Aufgaben der Statistik
- 1.2 Grundbegriffe
- **1.3 Etappen einer statistischen Untersuchung**

Etappen einer statistischen Untersuchung

Beispiel:

Beispiel EU Vergleich.xls

1.3 Etappen einer statistischen Untersuchung

Gliederung

- **1.3.1 Untersuchungsplanung**
- 1.3.2 Datenerhebung
- 1.3.3 Datenaufbereitung
- 1.3.4 Datenanalyse

Untersuchungsplanung

Aufgaben

1. Definition der statistischen Einheit und Abgrenzung der statistischen Gesamtheit
2. Definition Erhebungsmerkmale
3. Formulierung Untersuchungshypothesen
4. Auswahl der Methoden zur Erhebung der Ausgangsdaten
5. Auswahl der Methoden für Datenaufbereitung und Analyse
6. rechentechnische Absicherung der Erhebung, Speicherung, Aufbereitung und Auswertung der Daten
7. organisatorische, finanzielle, personelle Absicherung der statistischen Untersuchung

Beispiel: Untersuchungsplanung

Einige ausgewählte Fragenstellungen

1. Formulieren einiger Aussagen über den Zustand der EU (Gesamtheit der betrachteten EU-Mitgliedsländer)
   - Durchschnittswerte
   - Abweichungen vom Durchschnitt
   - Lage der einzelnen Länder bezüglich dem Durchschnitt bzw. untereinander
2. Zusammenhänge zwischen den erhobenen Daten
3. weitere wünschenswerte Daten

1.3 Etappen einer statistischen Untersuchung

Gliederung

- 1.3.1 Untersuchungsplanung
- **1.3.2 Datenerhebung**
- 1.3.3 Datenaufbereitung
- 1.3.4 Datenanalyse

Datenerhebung

Vorgang der Ermittlung und Erfassung der Ausprägungen eines bzw. mehrerer statistischer Merkmale an den einzelnen Einheiten der definierten statistischen Gesamtheit.

Ergebnis der Datenerhebung: **Urliste**

$$X_{(n,k)} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \vdots & \vdots & \ddots & \\ x_{n1} & x_{n2} & & x_{nk} \end{pmatrix}$$

mit

x_{ij} — *Auspraegung des* $j-$ *ten Merkmals am* $i-$ *ten Merkmalstraeger*

j — *Index Merkmal* $(j = 1, 2, \ldots, k)$: $X_1, X_2, \ldots, X_k$

i — *Index Merkmalstraeger* $(i = 1, 2, \ldots, n)$

Beispiel: Urliste

$$X_{(n,k)} = \begin{pmatrix} 6861 & 73.0 & 6 & 21765.0 & 44.9 & 1.0 & 310.0 \\ 5150 & 72.5 & 7 & 28245.0 & 60.4 & 1.0 & 332.0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 3018 & 74.6 & 7 & 12201.0 & 37.1 & 4.2 & 104.0 \end{pmatrix}$$

mit

x_{ij} — *Auspraegung des* $j-$ *ten Merkmals am* $i-$ *ten Merkmalstraeger*

j — *Index Merkmal* $(j = 1, 2, \ldots, 8)$: *LAND*, *ENERGIE*, ..., *ZEITUNG*

i — *Index Merkmalstraeger* $(i = B,\ DK,\ \ldots,\ E)$

1.3 Etappen einer statistischen Untersuchung

Gliederung

- 1.3.1 Untersuchungsplanung
- 1.3.2 Datenerhebung
- **1.3.3 Datenaufbereitung**
- 1.3.4 Datenanalyse

Datenaufbereitung

1. Prüfung und Korrektur der erhobenen Daten (Vollständigkeit; Plausibilität; ....)
2. Datentransformation
3. Ordnung, Zusammenfassung und Darstellung des Datenmaterials (elementare statistische Auswertungen, numerische und / oder graphische Darstellung des Datenmaterials)

$\Rightarrow$ Datenimport und Datenaufbereitung in SPSS

1.3 Etappen einer statistischen Untersuchung

Gliederung

- 1.3.1 Untersuchungsplanung
- 1.3.2 Datenerhebung
- 1.3.3 Datenaufbereitung
- **1.3.4 Datenanalyse**

Datenanalyse

Anwendung der dem analytischen Problem und der Datenlage adäquaten statistischen Methode auf die erhobenen und aufbereiteten Daten zum Zwecke der Erkenntnisgewinnung.

Problembezogene Wertung und Nutzung der Analyseergebnisse für die Entscheidungsfindung.