Skript

Statistik für Ökonomen BLE

Thiemo M. Kessel & Johannes Sauter

Februar 2010

mathekurse.ch

Vorwort

Dieses Skript beginnt in den beiden Teilen „Beschreibende Statistik“ und „Wahrscheinlichkeitstheorie“ mit einer Einführung in die Statistik. Beide Teile dienen als Grundlage für den dritten Teil über „Schliessende Statistik“. Alle für die Wirtschaftswissenschaften relevanten statistischen Tests werden darin behandelt und im tabellarischen Anhang in Form von Flussdiagrammen schematisch zusammengefasst. Als Grundlage dienen die Lehrbücher

a) J. Schira, *Statistische Methoden der VWL und BWL*. 2. Auflage, Pearson Studium 2005

b) R.M. Weiers, *Introduction to Business Statistics*. 6. Auflage, Thomson South-Western 2007

Es wurde zudem darauf geachtet, die mathematischen Hilfsmittel aus der Assessment-Stufe gezielt einzusetzen. Dabei wird formal auf den Skripten *Mathematik I und II für Ökonomen* von *mathekurse.ch* aufgebaut und auf entsprechende Kapitel verwiesen.

An dieser Stelle möchten wir uns bei Thomas Mettler für die Unterstützung bei den Beispielen sowie vor allem bei David Pumberger für seine Hilfe und die gute Zusammenarbeit bei *mathekurse.ch* bedanken.

Zürich, im Februar 2010

Thiemo M. Kessel
Johannes Sauter

Inhaltsverzeichnis

Teil I
Beschreibende Statistik

Die erste Aufgabe der Statistik besteht darin, die durch die Erhebung einer grossen Anzahl von Einzeldaten gewonnene Datenmenge übersichtlich zu „beschreiben“.

1 Statistische Variablen

Es wird von einer zu beschreibenden Menge Ω ausgegangen, welche als *Grundgesamtheit* bezeichnet wird. Jedem Element $\omega \in \Omega$ wird dann ein bestimmtes *statistisches Merkmal* $M(\omega)$ zugeordnet, wobei eine Unterscheidung zwischen vier Arten von statistischen Merkmalen erfolgt – auch Skalenniveaus genannt.

1. **Nominalskala**: Das statistische Merkmal gehört zu genau einer von k ungeordneten Kategorien.
   Beispiele sind Geschlecht, Beruf, Wohnort, Postleitzahl.

2. **Ordinalskala**: Wie die Nominalskala, jedoch besteht zwischen den verschiedenen Kategorien eine Ordnung. Es lassen sich aber keine Abstände definieren.
   Beispiele sind Noten, Rangierungen, Beurteilungen.

3. **Intervallskala**: Neben der Ordnung sind nun auch (messbare) Abstände, jedoch kein absoluter Nullpunkt definiert.
   Beispiele sind Uhrzeit, Temperatur in $°C$, Jahrgang.

4. **Verhältnisskala**: Wie die Intervallskala, nur ist zusätzlich ein absoluter Nullpunkt definiert, welcher erlaubt, auch sinnvoll Verhälntisse zwischen verschiedenen Merkmalen zu bilden.
   Beispiele sind Grösse, Gewicht, Temperatur in $°K$, Einkommen.

Es besteht also eine aufsteigende Ordnung zwischen den verschiedenen Skalenniveaus. Die Nominalskala ist das tiefste, primitivste Niveau, während die Verhältnisskala dem höchsten, entwickeltsten Niveau entspricht. Dies führt zur folgenden, alternativen Unterscheidung.

a) **Qualitative Merkmale**: Merkmale der Grundgesamtheit Ω können nur qualitativ miteinander verglichen werden. Dazu gehören die Nominal- und Ordinalskala.

b) **Quantitative Merkmale**: Merkmale der Grundgesamtheit Ω können quantitativ miteinander verglichen werden, da relative Messungen möglich sind. Es wird auch von metrischen Merkmalen gesprochen. Dazu gehören die Intervall- und Verhältnisskala.

Eine *statistische Variable* X ordnet dann jedem Element $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zu und kann daher als Funktion auf der Menge Ω aufgefasst werden.

Definition 1. *Sei Ω eine Grundgesamtheit. Eine Funktion*

$$\begin{array}{rccl} X: & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & x = X(\omega) \end{array} \tag{1}$$

heisst statistische Variable.

Es ist klar, dass im Fall eines quantitativen Merkmals die Wahl einer statistischen Variable meist obsolet ist, da das Merkmal bereits deren Rolle übernimmt — es wird daher in der Regel nicht zwischen dem Merkmal und der dazugehörigen Variable unterschieden. Besteht zum Beispiel Interesse am Merkmal „Einkommen pro Einwohner der Europäischen Union“ mit $\Omega =$„Einwohner der Europäischen Union“, so kann jedem Einwohner $\omega \in \Omega$ bereits sein Einkommen als Zahl $x = X(\omega) = M(\omega)$ zugeordnet werden. Im Fall qualitativer Merkmale kommt der Wahl einer statistischen Variable eine viel grössere Bedeutung zu, da erst die Zuordnung von Zahlenwerten eine mathematische Untersuchung überhaupt ermöglicht.

In der Regel ist es zu aufwändig, den Wert einer statistischen Variable $x = X(\omega)$ für jedes Element $\omega \in \Omega$ der Grundgesamtheit zu erfassen. Aus diesem Grund werden eine kleine Teilmenge[1] $\Omega' \subset \Omega$ ausgewählt, und lediglich die Werte $x = X(\omega')$ aller $\omega' \in \Omega'$ erfasst. Die Teilmenge Ω' wird dann als *Stichprobe* bezeichnet. Die Stichprobe kann entweder dem Zufall überlassend *zufällig* oder *repräsentativ*, d.h. mit einer ähnlichen Struktur wie die Grundgesamtheit, ausgewählt werden. Der *Umfang* der Stichprobe Ω' ist die Anzahl n der darin enthaltenen Elemente. Die n Werte von X auf Ω', nämlich $x_1 = X(\omega'_1), x_2 = X(\omega'_2), \ldots, x_{n-1} = X(\omega'_{n-1})$ und $x_n = X(\omega'_n)$, werden auch *Beobachtungsreihe* von X genannt.

Die in der Beobachtungsreihe vorkommenden Werte können sich mehrmals wiederholen. Daher kann jedem Wert x_i mit $1 \leq i \leq n$ seine *absolute Häufigkeit*

$$\mathrm{absH}\,(X = x_i) = n_i$$

zugeordnet werden. In Worten besagt n_i, wie oft der Wert x_i in der Beobachtungsreihe vorkommt. Die *relative Häufigkeit*

$$h_i = \mathrm{relH}\,(X = x_i) := \frac{n_i}{n}$$

[1] Die Bezeichnung $\subset$ kommt aus der Mengenlehre und bedeutet „Teilmenge von“ (siehe Kapitel 3.2).

setzt dann die absolute Häufigkeit n_i in Proportion zum Umfang n der gesamten Stichprobe. Es ist nun gebräuchlich, mittels der relativen Häufigkeit die so genannte *Häufigkeitsfunktion* und *empirische Verteilungsfunktion* der statistischen Variable X zu definieren.

Definition 2. *Die Häufigkeitsfunktion der statistischen Variable X ist definiert durch*

$$h(x) := \begin{cases} h_i & \textit{falls } x = x_i \\ 0 & \textit{sonst}\ . \end{cases} \tag{2}$$

Definition 3. *Die empirische Verteilungsfunktion der statistischen Variable X ist definiert durch*

$$H(x) := \sum_{x_i \leq x} h(x_i)\,. \tag{3}$$

Beispiel 1: Stelle die Häufigkeitsfunktion und Verteilungsfunktion folgender Beobachtungsreihe graphisch dar:

$1 \leq i \leq 10$	1	2	3	4	5	6	7	8	9	10
x_i	1.6	3.0	3.0	4.1	4.1	4.1	4.1	5.0	5.0	5.0

Lösung: Die relativen Häufigkeiten $h_i = n_i/n$ sind wegen $n = 10$ gegeben durch

$1 \leq i \leq 10$	1	2	3	4	5	6	7	8	9	10
h_i	0.1	0.2	0.2	0.4	0.4	0.4	0.4	0.3	0.3	0.3

Daraus ergibt sich nach Definition 2 die Häufigkeitsfunktion $h(x)$ als

$$h(x) := \begin{cases} 0.1 & \text{falls } x = 1.6 \\ 0.2 & \text{falls } x = 3.0 \\ 0.4 & \text{falls } x = 4.1 \\ 0.3 & \text{falls } x = 5.0 \\ 0 & \text{sonst}\ , \end{cases} \tag{4}$$

und nach Definition 3 die Verteilungsfunktion $H(x)$ als

$$H(x) := \begin{cases} 0 & \text{falls } x < 1.6 \\ 0.1 & \text{falls } 1.6 \leq x < 3.0 \\ 0.3 & \text{falls } 3.0 \leq x < 4.1 \\ 0.7 & \text{falls } 4.1 \leq x < 5.0 \\ 1 & \text{falls } 5.0 \leq x. \end{cases} \tag{5}$$

Für eine graphische Darstellung sei auf Abbildung 1 verwiesen. □

Oft liegen einige Werte der Beobachtungsreihe so nahe beieinander, dass deren Unterschied zum Beispiel ausschliesslich durch einen Erhebungsfehler hätte entstanden sein können. In diesen Fällen ist es sinnvoll, die Werte von X in m *Klassen* zu unterteilen. Dafür werden $m+1$ *Klassengrenzen* $\xi_0 < \xi_1 < \cdots < \xi_m$ gewählt, welche den Wertebereich von X gruppieren. Die j-te Klasse mit $1 \leq j \leq m$ besitzt dann die Untergrenze ξ_{j-1} und die Obergrenze ξ_j. Für $1 \leq j \leq m$ wird die *Klassenhäufigkeit* h_j^K der j-ten Klasse definiert als

$$h_j^K := \mathrm{relH}\left(\xi_{j-1} < x_i \leq \xi_j\right),$$

also als die Anzahl der Werte x_i mit $\xi_{j-1} < x_i \leq \xi_j$ geteilt durch den Umfang n der Stichprobe. Analog zu den Definitionen 2 und 3 kann eine *Klassenhäufigkeitsfunktion* beziehungsweise eine *Klassenhäufigkeitsverteilung* definiert werden.

Beispiel 2: Es soll das Einkommen der Mitarbeiter eines grossen Konzerns untersucht werden. Dazu werden als Stichprobe zufällig $n = 100$ Mitarbeiter ausgewählt

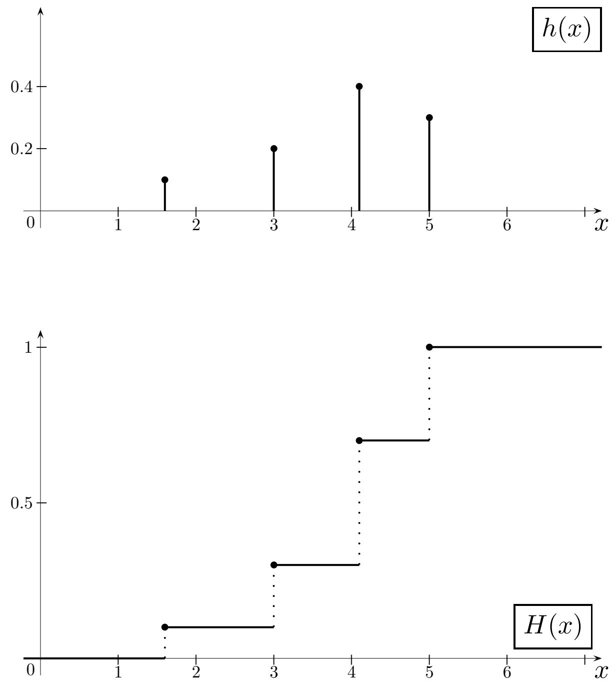

Abbildung 1: Häufigkeitsfunktion und Verteilungsfunktion.

Intervallschätzungen