Statistische Bedingungsanalysen

Sebastian Jeworutzki

25.06.2009

Rev: –revision–

Gliederung

Bedingte Verteilungen

Zweidimensionale statistische Variablen

$$(X, Y) : \Omega \rightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$$

Eine Variable X wird als unabhängige Variable bezeichnet, Y als abhängige Variable.

Bedingte Verteilungen

Bedingte Verteilungen

$$\left.P[\overbrace{Y}^{\text{Variablenname}} \mid \underbrace{X_1 \geq 12, X_2 = 1}_{\text{Bedingungen}}]\right\}\text{bedingte Häufigkeitsverteilung von Y}$$

Bedingte Verteilungen

Ω	X	Y
ω_1	0	21
ω_2	1	22
ω_3	1	21
ω_4	1	24
ω_5	0	23
ω_6	0	25
ω_7	1	24
ω_8	0	26
ω_9	1	27
ω_{10}	1	23

Beispieldaten:
$(X, Y) : \Omega \rightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ Daten über 10 Studierende. X bezeichnet das Geschlecht (0 männlich, 1 weiblich) und Y das Alter in Jahren.

Statistische Regressionsrechnung

Statistische Regressionsrechnung

Die weiteren Überlegungen beginnen mit der Definition einer allgemeinen Regressionsfunktion.

Allgemeine Regressionsfunktion

Allgemeine Regressionsfunktionen

$$\underbrace{\tilde{x} \longrightarrow \overbrace{(\tilde{y} \longrightarrow P[Y|X = \tilde{x}](\tilde{y}))}^{\text{bedingte Verteilung}}}_{\text{allgemeine Regressionsfunktion}}$$

Diese Funktionen „zweiter Ordnung“ ordnet jedem $X = \tilde{x}$ die bedingte Verteilung $P[Y|X = \tilde{x}](\tilde{y})$ zu, die die bedingte Häufigkeit für alle $\tilde{y} \in \tilde{\mathcal{Y}}$ liefert.

Wie lässt sich eine solche Regressionsfunktion einfacher darstellen?

Spezielle Regressionsfunktionen

Spezielle Regressionsfunktionen

$$\tilde{x} \longrightarrow \text{Charakterisierung von } P[Y|X = \tilde{x}]$$

Diese Funktion „erster Ordnung“ liefert für jedes $X = \tilde{x}$ eine Zahl, die $P[Y|X = \tilde{x}]$ charakterisiert.

Spezielle Regressionsfunktionen

Arten von Charakterisierungen

- Bedingte Mittelwerte: $\tilde{x} \longrightarrow M(Y|X = \tilde{x})$
- Bedingte Quantile: $\tilde{x} \longrightarrow Q_p(Y|X = \tilde{x})$
- Bedingte Häufigkeiten: $\tilde{x} \longrightarrow P(Y|X = \tilde{x})(\tilde{y})$

Beispiel: Autofahrer an einer Ampel

X	Y	Anzahl
0	0	47
0	1	0
1	0	3
1	1	50

X erfasst ob die Ampel rot ist (1) oder nicht (0), Y erfasst ob ein Auto anhält (1) oder nicht (0)

Beispiel: Autofahrer an einer Ampel

Regressionsfunktion: Bedingte Häufigkeiten

$$P[Y|X = 0](\tilde{Y} = 1) = 0$$
$$P[Y|X = 1](\tilde{Y} = 1) = 0,94$$

Wichtig! Regressionsfunktionen sind keine Abbildungen in den Merkmalsraum der abhängigen Variable:

Im Ampelbeispiel: $P[Y|X = 1](\tilde{Y} = 1) = 0,94$ aber $\tilde{\mathcal{Y}} = \{0, 1\}$.

Beispiel: Beruf & Einkommen

Ω	X	Y
ω_1	1	1500
ω_2	2	4800
ω_3	3	1000
ω_4	1	1800
ω_5	2	3700
ω_6	3	900
ω_7	1	1400
ω_8	3	1050
ω_9	2	5600
ω_{10}	1	1650

X erfasst den Beruf, Y das Einkommen in €

$x \in \tilde{\mathcal{X}}$

1 Bestattungsfachkraft
2 Hausarzt
3 Gymnastiklehrer

Welche Werte liefert eine Medianregression?

Beruf	$Q_{0.5}(Y \mid X = x)$
Bestattungsfachkraft	1575
Gymnastiklehrer	1000
Hausarzt	4800

Beispiel: Größe & Gewicht

Ω	X	Y
ω_1	180	73
ω_2	185	73
ω_3	165	61
ω_4	180	65
ω_5	175	67
ω_6	165	57
ω_7	180	60
ω_8	185	61
ω_9	150	52
ω_{10}	165	65
ω_{11}	175	67
ω_{12}	180	60

X erfasst die Größe einer Person in cm und Y erfasst das Gewicht dieser Person in kg.

Welche Werte liefert eine Mittelwertregression?

X	$M(Y\|X = x)$
150	52
165	61
175	67
180	64,5
185	67

Darstellung der Mittelwertregression

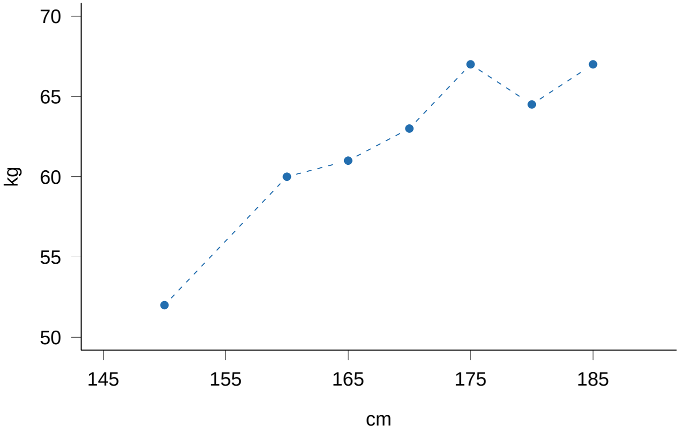

Abb. Darstellung der Mittelwertregression

Statistische Regressionsmodelle

Statistische Regressionsmodelle als Hilfsmittel ...

... zur einfacheren Darstellung von Regressionsfunktionen

... zur Berechnung von Regressionsfunktionen mit unvollständigen Daten

... zur Berechnung und Beurteilung von theoretisch vermuteten Regressionsfunktionen

... zur Berechnung von Schätzwerten bei bestimmten Werten der unabhängigen Variablen

Parametrische Regressionsmodelle

Ausgangspunkt ist eine spezielle Regressionsfunktion $g : \tilde{\mathcal{X}} \to \mathbf{R}$ wobei $g(x)$ eine Charakterisierung der bedingten Verteilung $P[Y|X = x]$ liefert.

parametrische Regressionsmodelle: $\tilde{g}(x; \theta)$

Bei parametrischen Verfahren wird g durch eine einfachere Modellfunktion $\tilde{g} : \tilde{\mathcal{X}} \to \mathbf{R}$ repräsentiert, die ihrerseits von weiteren Parametern θ abhängt.

Berechnungsverfahren für Modellfunktionen

Berechnungsverfahren für Modellfunktionen

1. Spezifizierung der parametrischen Modellfunktion $\tilde{g}(x;\theta)$
2. Festlegung eines Verfahres zu Bestimmung von θ
3. Berechnung von θ mit den vorhandenen Daten, so dass ein bestimmte Modellfunktion $\tilde{g}(x;\hat{\theta})$ entsteht

Modellfunktionen für Mittelwertregressionen

1. Modellfunktion für die lineare Mittelwertregression

$$\tilde{g}(x, \alpha, \beta) := \alpha + \beta x$$

2. Zielfunktion: „Ordinary Least Squares"-Kriterium

$$f_{LS}(\alpha, \beta) := \sum_{x \in \tilde{X}} (M(Y|X = x) - \tilde{g}(x; \alpha, \beta))^2 P(x)$$

Vergleich der Näherung der Modellfunktion mit den bedingten Mittelwerten $M(Y|X = x)$

Darstellung der Mittelwertregression

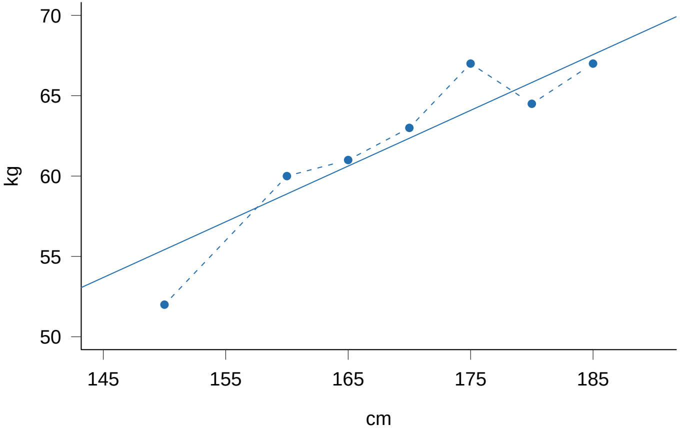

Abb. Darstellung der Mittelwertregression und Regressionsmodell ($\tilde{g}(x, \alpha, \beta) := 3.4098 + 0.3467x$)

Beispiel: Ausgaben privater Haushalte für Nahrungsmittel

		Ausgaben für Nahrung	
Einkommensklasse	Mittelwert	in DM	in %
unter 1 800	1 383	269	19.45
1 800 – 2 500	2 196	341	15.53
2 500 – 3 000	2 788	391	14.02
3 000 – 4 000	3 543	473	13.35
4 000 – 5 000	4 566	584	12.79
5 000 – 7 000	6 057	677	11.18
7 000 – 10 000	8 422	775	9.20
10 000 – 35 000	13 843	894	6.46

Tab. 15.1-1 Ausgaben privater Haushalte für Nahrungsmittel

Beispiel: Ausgaben privater Haushalte für Nahrungsmittel

Lineare Modellfunktion

$$\begin{aligned}\tilde{g}_1(x,\alpha,\beta) :=& \alpha + \beta x \\ =& 279.10 + 0.0507x\end{aligned}$$

Logit-Modellfunktion

$$\begin{aligned}\tilde{g}_2(x,\alpha,\beta,\gamma) :=& \gamma \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)} \\ =& 890.926 \frac{\exp(-1.390 + 0,000422x)}{1+\exp(-1.390 + 0,000422x)}\end{aligned}$$

Beispiel: Ausgaben privater Haushalte für Nahrungsmittel

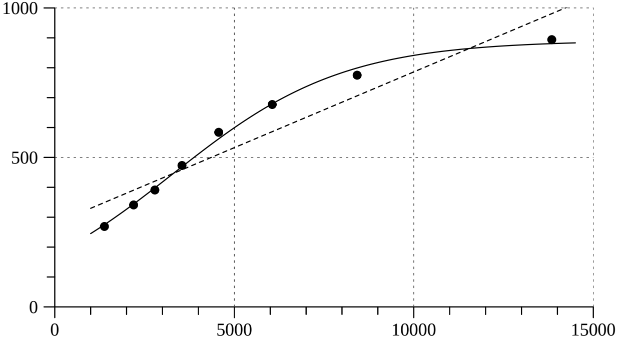

Abb. 15.1-2 Ausgaben für Nahrungsmittel (y-Achse: in DM) als Funktion des für Ausgaben verfügbaren Haushaltseinkommens (x-Achse: in DM) und Modellfunktionen $\tilde{g}_1$ (gestrichelt) und $\tilde{g}_2$ (durchgezogene Linie). Daten aus der EVS 1998 (Tab 15.1-1).

Statistische und substantielle Bedingungen

Statistische und substantielle Bedingungen

- Lassen sich die zur Konditionierung verwendeten Variablen auch im substantiellen Sinn als Bedingung verstehen?

Statistische und substantielle Bedingungen

Die Regressionsrechnung kann zeigen, wie statistische Strukturen (statistische Verteilungen!) von Werten statistischer Variablen abghängig sind.

Die Regressionsrechnung

... setzt die Existenz eines substantiellen Prozesses nicht voraus

... liefert keine begrifflichen Hilfsmittel zur Repräsentation substantieller Prozesse

$\Rightarrow$ substantielle Prozesse müssen jenseits der statistischen Begriffsbildungen und Rechnungen vorstellbar gemacht werden

Bezugnahme auf substantielle Prozesse

$$\begin{array}{ccc} \text{Substantieller Prozess} & \longrightarrow & \tilde{y} \\ \uparrow & & \\ \tilde{x} & & \end{array}$$

- Die Werte von $\tilde{y}$ kommen durch eine Prozess zustande, an dem Akteure Beteiligt sind
- Können die Werte von $\tilde{x}$ als Bedingungen für Verhaltensweisen und Tätigkeiten der Akteure interpretiert werden?

Allgemein: Können statistische Strukturen auch als Bedingungen für soziale Prozesse aufgefasst werden?

Strukturen als Handlungsbdingungen

Strukturen als Handlungsbedingungen

Es muss unterschieden werden:

- Verteilungen als Handlungsbedingungen
- Erfassung von Handlungsbedingungen durch statistische Verteilungen.

Strukturen als Handlungsbdingungen

Wie kann davon gesprochen werden, dass die Verteilung $P[X]$ eine Bedingung für einen substantiellen Prozess ist?

- $P[X]$ bezieht sich auf eine Situation die ein externer Akteur vorfindet.
  Bspw: Verteilung von BA/MA-Studierenden in einem Seminar als Handlungsbedingung für den Dozenten. Der Akteur ist kein Mitglied von Ω
- $P[X]$ als Charakterisierung der Situation eines Mitglieds von Ω.
  $P[X]$ kann auch eine Handlungsbedingung für die Individuellen Mitglieder von Ω sein. Bspw. Möglichkeit der Studierenden Lerngruppen zu bilden.
- Bezieht sich $P[X]$ nicht auf eine spezifische Situation, liefert P[X] keine Charakterisierung der Handlungsbedingungen für die Akteure.

Mikro-Relevanz statistischer Strukturen

Es muss also gezeigt werden, wie die statistische Struktur Handlungsbedingung für jedes Objekt in Ω ist.

$$
P[X] \begin{array}{c} \nearrow \\ \\ \searrow \end{array}
\begin{array}{ccc}
\omega_1 & \longrightarrow & Y(\omega_1) \\
\vdots & \vdots & \\
\omega_n & \longrightarrow & Y(\omega_n)
\end{array}
$$