④ Grundlagen der induktiven Statistik*

Zoltán Zomotor

Versionsstand: 31. März 2015, 09:29

Die nummerierten Felder bitte während der Vorlesung ausfüllen.

Bitte hier notieren, was beim Bearbeiten unklar geblieben ist:

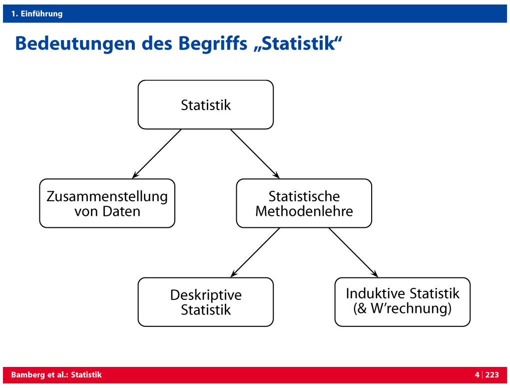
Inhaltsverzeichnis

*[Pap11, Teil III, Kapitel 3.2]

1 Einleitung

1.1 Bedeutungen des Begriffs „Statistik“

1



Beispiel: 12 Beschäftigte werden nach der Entfernung zum Arbeitsplatz (in km) befragt. Antworten: 4, 11, 1, 3, 5, 4, 20, 4, 6, 16, 10, 6

2

- deskriptiv:
  - durchschnittliche Entfernung: 7.5 km
  - Klassenbildung:

Klasse	[0,5)	[5,15)	[15,30)
Häufigkeit	5	5	2

3

- induktiv:
  - Schätze mittlere Entfernung aller Beschäftigten
  - Prüfe, ob mittlere Entfernung <10 km ist.

1.2 Induktive Statistik

- Vollerhebung oft unmöglich, deshalb:

4

  - Ziehe Stichprobe

  $\Rightarrow$ Schließe auf Grundgesamtheit

- Beispiel:
  - Warensendung 1000 Stück
  - darunter M Stück Ausschuss
  - M ist unbekannt

Stichprobe:

5

- zufällig Entnahme von $n = 30$ Stück
- darunter 2 Stück Ausschuss

Denkbare Zielsetzungen:

6

- Schätze M durch eine Zahl, z.B. $\frac{2}{30} \cdot 1000 = 66.67 \approx 67$
- Schätze ein Intervall für M, z.B. $P(M \in [58, 84]) = 0.95$
- Teste die Hypothese, dass $M \leq 50$ ist

2 Grundbegriffe

Merkmalsträger: 7 Untersuchte statistische Einheit

Merkmal: 8 Relevante Eigenschaft des Merkmalträgers

Ausprägung: 9 Konkret beobachteter Wert des Merkmals

Grundgesamtheit G:

10 Menge aller relevanten Merkmalsträger.

Verteilung der Grundgesamtheit:

11 $F(x) = P(X \leq x)$

$= $ W'keit für Auswahl eines MM-Trägers mit Ausprägung $\leq x$

Stichprobenumfang n:

12 Anzahl der Merkmalsträger in der Stichprobe

Uneingeschränkte Zufallsauswahl:

13 Jeder MM-Träger hat die selbe Chance, ausgewählt zu werden.

Einfache Stichprobe:

Uneingeschränkte Zufallsauswahl und 14 unabhängige Ziehung.

$\Rightarrow$ Alle Stichprobenvariablen $X_1, \ldots, X_n$ sind

15 independent, identically distributed i.i.d.

Stichprobenergebnis:

16 Tupel der Realisationen $(x_1, \ldots, x_n)$ der Stichprobenvariablen.

Stichprobenraum:

17 Menge aller möglichen Stichprobenergebnisse.

Likelihoodfunktion:

18

- Verteilungsklasse der Grundgesamtheit ist bekannt.
- Verteilungsparameter ϑ aber unbekannt, z.B. $N(\mu, \sigma)$
- Einfache Stichprobe $X_1, \ldots, X_n$

$\Rightarrow$ Likelihoodfunktion: $f(x_1, \ldots, x_n | \vartheta)$

- Gemeinsame Dichte / W'funktion von $X_1, \ldots, X_n$.
- Hängt von ϑ ab.

Beispiel:

- $G \sim B(1, p) \Rightarrow f_i(x_i) =$ 19 $p^{x_i}(1-p)^{1-x_i}$
- p ist unbekannt
- Einfache Stichprobe mit $n = 2$

⇒ Likelihoodfunktion:

20

$$\begin{aligned} f(x_1, x_2|p) &= f_1(x_1) \cdot f_2(x_2) \quad \text{(wegen Unabhängigkeit)} \\ &= p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \\ &= p^{x_1+x_2}(1-p)^{2-x_1-x_2} \end{aligned}$$

- Stichprobenergebnis $(0, 1)$ ⇒ 21 $f(0, 1|p) = p(1-p) = p - p^2$
- Welcher Wert p passt „am besten“ zu $(0, 1)$?

Stichprobenfunktion:

22 Funktion der Stichprobenvariablen: $V = g(X_1, \ldots, X_n)$

3 Wichtige Stichprobenfunktionen

- Gegeben:
  - Einfache Stichprobe 23 $X_1, \ldots, X_n$
  - Beliebige Verteilung mit 24 $\mathrm{E}(X_i) = \mu$ und $\mathrm{Var}(X_i) = \sigma^2$
- Besonders wichtige Zusammenhänge:
  - Der Erwartungswert der mittleren quadratischen Abweichung bezüglich μ ist die Varianz σ^2 der Stichprobe wegen:

25

$$\begin{aligned} E(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2) &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}((X_i-\mu)^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}n\sigma^2 = \sigma^2 \end{aligned}$$

  - Die Varianz lässt sich schätzen, indem der (theoretische und im allgemeinen unbekannte) Erwartungswert μ mit Hilfe des Stichprobenmittels $\bar{X}$ geschätzt wird. Wegen

26

$$\mathrm{E}(\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2) = \frac{n-1}{n}\sigma^2$$

muss für eine erwartungstreue[1] Schätzung der Stichprobenvarianz S^2 der Korrekturfaktor $\frac{n}{n-1}$ berücksichtigt werden:

27

$$S^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

– Verschiebungssatz für S^2:

28

$$S^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}^2\right)$$

$$\Rightarrow S^2 = \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 - \frac{n}{n-1}\cdot\bar{X}^2$$

– Wegen der Jensenschen Ungleichung gilt

29

$$\mathrm{E}(S) = \mathrm{E}(\sqrt{S^2}) \leq \sqrt{\mathrm{E}(S^2)} = \sqrt{\sigma^2} = \sigma$$

$$\Rightarrow \mathrm{E}(S) \leq \sigma$$

[1] Definition folgt später.

30

Stichprobenfunktion	Bezeichnung	Erwartungswert	Varianz
$\sum_{i=1}^{n} X_i$	Merkmalssumme	$n\mu$	$n\sigma^2$
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	Stichprobenmittel	μ	$\frac{\sigma^2}{n}$
$\frac{\bar{X}-\mu}{\sigma}\sqrt{n}$	Gauß-Statistik	0	1
$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$	mittlere quadratische Abweichung bezgl. μ	σ^2	
$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$	mittlere quadratische Abweichung	$\frac{n-1}{n}\sigma^2$	
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$	Stichprobenvarianz	σ^2	
$S = \sqrt{S^2}$	Stichproben-Standardabweichung		
$\frac{\bar{X}-\mu}{S}\sqrt{n}$	t-Statistik		

4 Prüf- und Testverteilungen [Pap11, Teil II, Kapitel 8]

4.1 Chi-Quadrat-Verteilung

- Sind $X_1, \ldots, X_n \sim N(0,1)$ i.i.d, so wird die Verteilung von

  31

  $$Z = \sum_{i=1}^{n} X_i^2$$

  als *Chi-Quadrat-Verteilung* mit n Freiheitsgraden bezeichnet.

- Kurzschreibweise: 32 $Z \sim \chi^2(n)$

- Es gilt: 33 $\mathrm{E}(Z) = n$ und $\mathrm{Var}(Z) = 2n$

- Fraktile (Quantile):
  - Tabelle 1, siehe auch [Pap11, Tabelle 3, Seite 744]
  - Ab $n > 30$ Näherung: 34 $x_\alpha = \frac{1}{2}(\tilde{x}_\alpha + \sqrt{2n-1})^2$
    $\tilde{x}_\alpha$ ist das α-Fraktil der $N(0,1)$-Verteilung

- Beispiel: $x_{0.975}$ aus

35

$$\chi^2(30) \quad x_{0.975} = 47.0$$
$$\chi^2(110) \quad \tilde{x}_{0.975} = 1.96 \Rightarrow x_{0.975} = \frac{1}{2}(1.96 + \sqrt{219})^2 = 140.9166$$

exakt:140.4255

4.2 t-Verteilung

- Ist $X \sim N(0,1)$, $Z \sim \chi^2(n)$, X, Z unabhängig, so wird die Verteilung von

36

$$T = \frac{X}{\sqrt{\frac{1}{n}Z}}$$

als t-Verteilung mit n Freiheitsgraden bezeichnet.

- Kurzschreibweise: 37 $T \sim t(n)$

- Es gilt: 38 $\mathrm{E}(T) = 0$ und $\mathrm{Var}(T) = \dfrac{n}{n-2}$

- Fraktile:
  - Tabelle 2, siehe auch [Pap11, Tabelle 4, Seite 746]
  - Ab $n > 30$ Näherung 39 durch $N(0,1)$
  - Achtung: 40 Nur $\alpha \geq 0.9$ vertafelt $(p = \alpha)$
    Ggf. Symmetrie ausnutzen $x_\alpha = -x_{1-\alpha}$
  - Beispiel: Bestimme folgende Fraktile für $t(10)$:

41

$$x_{0.9} = 1.372$$
$$x_{0.5} = 0$$
$$x_{0.1} = -x_{0.9} = -1.372$$

5 Verteilungen wichtiger Stichprobenfunktionen

Gegeben: Einfache Stichprobe $X_1, \ldots, X_n$ aus $N(\mu, \sigma)$-Verteilung:

Stichprobenfunktion	Verteilung
$\sum_{i=1}^{n} X_i$	$N(n\mu, \sigma\sqrt{n})$
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$N(\mu, \frac{\sigma}{\sqrt{n}})$
$\frac{\bar{X}-\mu}{\sigma}\sqrt{n}$	$N(0,1)$
$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$	$\chi^2(n)$
$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{n-1}{\sigma^2}S^2$	$\chi^2(n-1)$
$\frac{\bar{X}-\mu}{S}\sqrt{n}$	$t(n-1)$

Beliebige Verteilung $\Rightarrow \frac{\bar{X}-\mu}{\sigma}\sqrt{n}$ und $\frac{\bar{X}-\mu}{S}\sqrt{n} \sim N(0,1)$ (näherungsweise)

6 Übungsaufgaben

X_i (mit $i = 1, \ldots, n$) seien unabhängige, jeweils $N(\mu, \sigma)$-verteilte Zufallsvariable. Geben Sie für jede der folgenden Größen die Zahl $x_{0.95}$ an, die mit 5 % Wahrscheinlichkeit überschritten wird (zum Beispiel für (a): $P(x_{0.95} < A) = 0.05$).

(a) $A = \frac{1}{\sigma^2}\sum_{i=1}^{7}(X_i-\bar{X})^2$

(b) $B = \frac{1}{\sigma^2}\sum_{i=1}^{137}(X_i-\bar{X})^2$

(c) $C = \frac{1}{\sigma^2}\sum_{i=1}^{12}(X_i-\mu)^2$

(d) $D = \frac{\sum_{i=1}^{16} X_i - 16\mu}{4S}$ (mit S^2: Stichprobenvarianz von X)

Literatur

[Pap11] Lothar Papula. *Mathematik für Ingenieure und Naturwissenschaftler*. 6. Auflage. Bd. 3. Vieweg + Teubner, 2011.

Tabelle 1: Fraktile der χ^2-Verteilung

	p									
n	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0	0	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.01	0.02	0.051	0.103	0.211	4.605	5.991	7.378	9.21	10.6
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.35	12.84
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.236	11.07	12.83	15.09	16.75
6	0.676	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	1.69	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.18	2.733	3.49	13.36	15.51	17.54	20.09	21.95
9	1.735	2.088	2.7	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.94	4.865	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	17.27	19.68	21.92	24.73	26.76
12	3.074	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.3
13	3.565	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69	29.82
14	4.075	4.66	5.629	6.571	7.79	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	22.31	25	27.49	30.58	32.8
16	5.142	5.812	6.908	7.962	9.312	23.54	26.3	28.84	32	34.27
17	5.697	6.408	7.564	8.672	10.09	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.39	10.87	25.99	28.87	31.53	34.8	37.16
19	6.844	7.633	8.907	10.12	11.65	27.2	30.14	32.85	36.19	38.58
20	7.434	8.26	9.591	10.85	12.44	28.41	31.41	34.17	37.57	40
22	8.643	9.542	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.8
24	9.886	10.86	12.4	13.85	15.66	33.2	36.41	39.36	42.98	45.56
26	11.16	12.2	13.84	15.38	17.29	35.56	38.88	41.92	45.64	48.29
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
30	13.79	14.95	16.79	18.49	20.6	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.8	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.5	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.4	79.08	83.3	88.38	91.95
70	43.27	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.2	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Tabelle 2: Fraktile der t-Verteilung

	p				
n	0.9	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
22	1.321	1.717	2.074	2.508	2.819
24	1.318	1.711	2.064	2.492	2.797
26	1.315	1.706	2.056	2.479	2.779
28	1.313	1.701	2.048	2.467	2.763
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
100	1.290	1.660	1.984	2.364	2.626
200	1.286	1.653	1.972	2.345	2.601
500	1.283	1.648	1.965	2.334	2.586
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
∞	1.282	1.645	1.960	2.326	2.576