Stochastik

FH Aachen – Studienstandort Aachen
5. Dezember 2025

Prof. Dr. Florian Heinrichs

Klassierte Daten

Einleitung

Manchmal[1] haben wir eine große Menge stetiger Daten und wollen diese zu Klassen zusammenfassen. Das Ergebnis sind *klassierte Daten*.

Beispiel 136

Der Kundendienst eines Unternehmens misst die Wartezeiten, die Anrufer in einer Warteschleife verbringen mit folgendem Ergebnis:

i	Klasse	Dauer (min)	h_i abs. Häufigkeit	r_i rel. Häufigkeit	Klassenbreite
1	A_1	$0 \leq \cdot < 2$	1650	0.3300	2
2	A_2	$2 \leq \cdot < 4$	1111	0.2222	2
3	A_3	$4 \leq \cdot < 6$	720	0.1440	2
4	A_4	$6 \leq \cdot < 8$	508	0.1016	2
5	A_5	$8 \leq \cdot < 10$	332	0.0664	2
6	A_6	$10 \leq \cdot < 15$	427	0.0854	5
7	A_7	$15 \leq \cdot \leq 30$	252	0.0504	10
Σ			5000	1.0000	

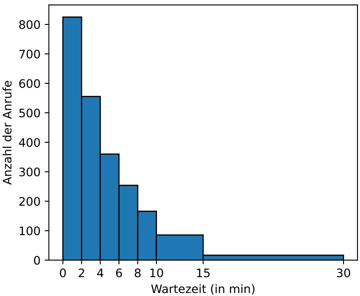
[1] In der Klausur.

Einleitung

Beispiel 136

Der Kundendienst eines Unternehmens misst die Wartezeiten, die Anrufer in einer Warteschleife verbringen mit folgendem Ergebnis:

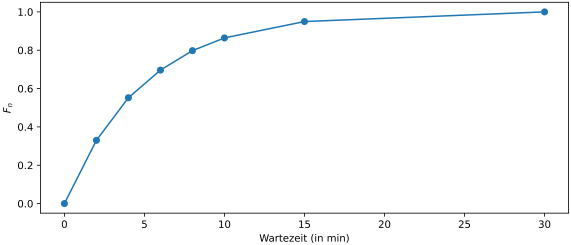
i	Dauer	h_i	r_i	KB
1	$[0, 2)$	1650	0.3300	2
2	$[2, 4)$	1111	0.2222	2
3	$[4, 6)$	720	0.1440	2
4	$[6, 8)$	508	0.1016	2
5	$[8, 10)$	332	0.0664	2
6	$[10, 15)$	427	0.0854	5
7	$[15, 30]$	252	0.0504	10



Achtung: Fläche der Rechtecke **proportional** zu h_i!

Verteilungsfunktion

- Wie können wir für klassierte Daten eine Verteilungsfunktion F_n zeichnen?
- Für F_n sollte gelten: $F_n(x) \approx \mathbb{P}(X \leq x)$
- Wir wissen nur für die oberen Klassengrenzen, wie viele Beobachtungen kleiner oder gleich der Grenze sind.
- Zeichne diese Punkte ein und interpoliere linear zwischen den Punkten.



Mittelwert

- Wie können wir für klassierte Daten den Mittelwert $\bar{x}_n$ berechnen?
- Sei α_i die Klassenmitte der Klasse A_i mit absoluter Häufigkeit h_i und relativer Häufigkeit r_i.
- Dann ist

$$\bar{x}_n \approx \tilde{x}_n := \frac{1}{n}\sum_{i=1}^{k} h_i\alpha_i = \sum_{i=1}^{k} r_i\alpha_i.$$

- Im Beispiel: $\tilde{x}_n \approx 5.23$

i	Dauer	h_i	r_i	α_i
1	$[0,2)$	1650	0.3300	1
2	$[2,4)$	1111	0.2222	3
3	$[4,6)$	720	0.1440	5
4	$[6,8)$	508	0.1016	7
5	$[8,10)$	332	0.0664	9
6	$[10,15)$	427	0.0854	12.5
7	$[15,30]$	252	0.0504	22.5

Empirische Varianz

Wie können wir für klassierte Daten die empirische Varianz s^2 berechnen?

- Klassenmitte: α_i
- Absolute Häufigkeit: h_i
- Relative Häufigkeit: r_i
- Mittelwert: $\tilde{x}_n$
- Dann ist

$$s^2 \approx \tilde{s}^2 := \frac{1}{n-1} \sum_{i=1}^{k} h_i(\alpha_i - \tilde{x}_n)^2.$$

- Im Beispiel: $\tilde{s}^2 \approx 27.83$

i	Dauer	h_i	r_i	α_i
1	$[0,2)$	1650	0.3300	1
2	$[2,4)$	1111	0.2222	3
3	$[4,6)$	720	0.1440	5
4	$[6,8)$	508	0.1016	7
5	$[8,10)$	332	0.0664	9
6	$[10,15)$	427	0.0854	12.5
7	$[15,30]$	252	0.0504	22.5

Empirische Quantile

Wie können wir für klassierte Daten Quantile berechnen?

1. Berechne aus rel. Häufigkeit r_i die kumulierte rel. Häufigkeit F_i.
2. Finde "Einfallsklasse" $A_i = [a_i, a_{i+1})$, sodass $F_i \geq p$ und $F_{i-1} < p$.
3. Innerhalb der Klasse: lineare Interpolation

$$q_p = a_i + \frac{p - F_{i-1}}{F_i - F_{i-1}}(a_{i+1} - a_i).$$

4. Was ist der Median im Beispiel?

$i = 2$: $A_2 = [2, 4)$

$$q_{0.5} = 2 + \frac{0.5 - 0.33}{0.5522 - 0.33}(4 - 2) \approx 3.53$$

Und $q_{0.8}$?

i	Dauer	h_i	r_i	F_i
1	$[0, 2)$	1650	0.3300	0.3300
2	$[2, 4)$	1111	0.2222	0.5522
3	$[4, 6)$	720	0.1440	0.6962
4	$[6, 8)$	508	0.1016	0.7978
5	$[8, 10)$	332	0.0664	0.8642
6	$[10, 15)$	427	0.0854	0.9496
7	$[15, 30]$	252	0.0504	1.0000

Modalklasse

1. A_i heißt *Modalklasse*, falls $h_i = \max_{i=1}^{k} h_j$.
2. Was ist der Modalklasse im Beispiel? A_1
3. Informell: Die Modalklasse ist die häufigste Klasse.

i	Dauer	h_i	r_i	F_i
1	$[0,2)$	1650	0.3300	0.3300
2	$[2,4)$	1111	0.2222	0.5522
3	$[4,6)$	720	0.1440	0.6962
4	$[6,8)$	508	0.1016	0.7978
5	$[8,10)$	332	0.0664	0.8642
6	$[10,15)$	427	0.0854	0.9496
7	$[15,30]$	252	0.0504	1.0000

Übung

Übung 86

Bei einer Zugverbindung von Jülich nach Aachen wurden in 100 Fahrten folgende Verspätungen (in Minuten) gemessen. Berechnen Sie aus der absoluten Häufigkeit h_i die relative Häufigkeit r_i, die kumulierte rel. Häufigkeit F_i, die Klassenbreite KB und die Klassenmitte α_i.
Zeichnen Sie die empirische Verteilungsfunktion F_n. Bestimmen Sie $\tilde{X}_n$, $\tilde{s}^2$, $q_{0.5}$, $q_{0.9}$ und die Modalklasse.

i	Dauer	h_i	r_i	F_i	α_i	KB
1	$[-2, 0)$	12				
2	$[0, 2)$	20				
3	$[2, 5)$	36				
4	$[5, 10)$	15				
5	$[10, 20)$	10				
6	$[20, 60)$	7				

Literatur I

- Bundesministerium für Gesundheit (2024).
  Infektionsradar.
  https://infektionsradar.gesund.bund.de/de/covid/inzidenz.
  Abgerufen: 2024-10-15.

- Dehling, H. and Haupt, B. (2006).
  Einführung in die Wahrscheinlichkeitstheorie und Statistik.
  Springer-Verlag.

Literatur II

Dinnes, J., Sharma, P., Berhane, S., van Wyk, S., Nyaaba, N., Domen, J., Taylor, M., Cunningham, J., Davenport, C., Dittrich, S., Emperador, D., Hooft, L., Leeflang, M., McInnes, M., Spijker, R., Verbakel, J., Takwoingi, Y., Taylor-Phillips, S., Van den Bruel, A., and Deeks, J. (2022).
Rapid, point-of-care antigen tests for diagnosis of sars-cov-2 infection.
Cochrane Database of Systematic Reviews, (7).

Henze, N. et al. (1997).
Stochastik für Einsteiger, volume 4.
Springer.