

Statistik Formelsammlung

Tim Hilt

Emil Slomka

30. Dezember 2019

Inhaltsverzeichnis

I	Beschreibende Statistik	3
1	Kombinatorik und Wahrscheinlichkeitsrechnung	4
1.1	Kombinatorik	4
1.1.1	Permutation	4
1.2	Kombination und Variation	4
1.3	Wahrscheinlichkeitsrechnung	5
1.3.1	Wahrscheinlichkeitsverteilungen	5
II	Deskriptive Statistik	8

Teil I

Beschreibende Statistik

Kapitel 1

Kombinatorik und Wahrscheinlichkeitsrechnung

1.1 Kombinatorik

1.1.1 Permutation

Permutation ohne Wiederholung:	n
Permutation mit Wiederholung:	$\frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_s!}$

1.2 Kombination und Variation

Kombinationen werden verwendet, wenn **nur einige** der Elemente in einer Menge angeordnet werden sollen. Permutationen hingegen ordnen **alle** Elemente an.

Urnenmodell: Ziehen von k aus n	Anzahl Möglichkeiten	Name Vorlesung	typische Beispiele
ohne Zurücklegen; ungeordnet	$\binom{n}{k}$	Kombination verschiedene Elemente	a) Lotto: 6 aus 49 b) k Personen aus n (Arbeitsgruppe)
mit Zurücklegen; ungeordnet	$\binom{n+k-1}{k}$	Kombination Elemente mehrfach	a) Widerstände parallel b) 2 T-Shirts aus 5 Farben auswählen
ohne Zurücklegen; geordnet	$\frac{n!}{(n-k)!}$	Variation verschiedene Elemente Spezialfall: $n = k$ Permutation	a) Siegerpodest b) Rangreihenfolge Auswahl Studierende c) Zieleinlauf insgesamt
mit Zurücklegen; geordnet	n^k	Variation Elemente mehrfach	a) Binäre Ziffernfolge b) Wörter aus 7 Buchstaben

1.3 Wahrscheinlichkeitsrechnung

1.3.1 Wahrscheinlichkeitsverteilungen

Allgemeine Form

Dichtefunktion:

Funktion, bei der auf der x -Achse alle Elemente mit der auf der y -Achse aufgetragenen Wahrscheinlichkeit gezeichnet sind. Es ergibt sich ein Säulendiagramm.

Verteilungsfunktion:

$$F(x) = P(X \leq x) = \sum_k^x (k \cdot P(X = k))$$

Erwartungswert:

$$E(X) = \mu = \sum_k k \cdot P(X = k)$$

Varianz

$$\text{Var}(X) = \sigma^2 = \sum_k (k^2 \cdot P(X = k)) - \mu^2$$

Hypergeometrische Verteilung

Beschreibung: Ziehen ohne Zurücklegen \rightarrow Wahrscheinlichkeit verändert sich im Verlauf des Experiments

Es müssen folgende Variablen (bis auf k) gegeben sein:

N	Anzahl aller Elemente
M	Anzahl Elemente mit bestimmter Eigenschaft
n	Anzahl Elemente in der Stichprobe
k	Anzahl Elemente aus der Stichprobe, die das Merkmal aufweisen sollen

Dichtefunktion:

$$X \sim H(n, N, M)$$

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

Verteilungsfunktion:

$$F(x) = P(X \leq x) = \sum_{k=0}^x H(k, n, N, M)$$

Erwartungswert:

$$E(X) = \mu = n \cdot \frac{M}{N}$$

Varianz:

$$\text{Var}(X) = \sigma^2 = n \cdot \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$$

Binomialverteilung

Beschreibung: Ziehen **mit** zurücklegen \rightarrow Wahrscheinlichkeit bleibt während dem Experiment gleich
Es müssen folgende Variablen gegeben sein:

p	Anteil der Elemente/ Wahrscheinlichkeit beim Ziehen eines Elementes aus der Grundgesamtheit
n	Anzahl Elemente in der Stichprobe
k	Anzahl Elemente aus der Stichprobe, die das Merkmal aufweisen sollen

Dichtefunktion:

$$X \sim B(n, p)$$

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Verteilungsfunktion: Hier müssen die einzelnen Dichtefunktionen berechnet werden

$$F(x) = P(X \leq x) = \sum_{k=0}^x B(k, n, p)$$

Erwartungswert:

$$E(X) = \mu = n \cdot p$$

Varianz:

$$\text{Var}(X) = \sigma^2 = n \cdot p \cdot (1 - p)$$

Annäherung der Hypergeometrischen Verteilung durch Binomialverteilung:

Falls $\frac{n}{N} \leq 0.1$ kann der Parameter p durch $p = \frac{M}{N}$ angenähert werden.

Poissonverteilung

Beschreibung: Gegeben ist ein Durchschnittswerts (*Erwartungswert*) λ pro einer gewissen Einheit (z.B. im Durchschnitt 3 Anrufe in 5 Minuten). Die Poissonverteilung soll berechnen, wie groß die Wahrscheinlichkeit ist einen anderen Wert k als Ergebnis zu erhalten.

Dichtefunktion:

$$X \sim \text{Po}(\lambda)$$

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Verteilungsfunktion:

$$F(x) = P(X \leq x) = \sum_{k=0}^x \text{Po}(k, \lambda)$$

Erwartungswert:

$$E(X) = \mu = \lambda$$

Varianz:

$$\text{Var}(X) = \sigma^2 = \lambda$$

Annäherung der Binomialverteilung durch Poissonverteilung:

Wenn n groß und p klein ist ($n \geq 30, p \leq 0.1$) kann der Parameter λ durch $\lambda = n \cdot p$ angenähert werden.

Geometrische Verteilung

Beschreibung: Gegeben ist die Wahrscheinlichkeit für einen Erfolg (p , Misserfolg $q = 1 - p$). Die geometrische Verteilung berechnet die Wahrscheinlichkeit dafür, dass genau k Versuche benötigt werden um zum ersten Erfolg zu kommen; dass man also **beim** k -ten Versuch Erfolg hat.

Dichtefunktion:

$$P(X = k) = p \cdot (1 - p)^{k-1}$$

Verteilungsfunktion:

$$P(X \leq k) = 1 - (1 - p)^k$$

Erwartungswert:

$$E(X) = \mu = \frac{1}{p}$$

Varianz:

$$\text{Var}(X) = \sigma^2 = \frac{1 - p}{p^2}$$

Quantile und Zufallsstrebereich der Normalverteilung

Quantile berechnen einen bestimmten Prozentsatz der Fläche unter der Verteilungsfunktion einer normalverteilten Zufallsvariable.

Bsp. das 95% Quantil wird geschrieben als

$$q_{0.95} = \mu + z_{0.95} \cdot \sigma$$

Das bedeutet, dass 95% der Werte unterhalb des Wertes $q_{0.95}$ liegen.

Die Werte für z_m sind tabelliert, können jedoch auch mit dem Taschenrechner (mit $\mu = 0$ und $\sigma = 1$) berechnet werden:

m	0.8	0.9	0.95	0.975	0.99	0.995	0.999
z_m	0.842	1.282	1.654	1.960	2.326	2.576	3.090

Der Zufallsstrebereich (ZSB) ist ein Intervall, das zwei Quantile berechnet. ZSBs können nach unten, nach oben oder zweiseitig beschränkt sein. Der Zufallsstrebereich kann für ein gegebenes μ oder für den arithmetischen Mittelwert \bar{X} eines gegebenen Datensatzes berechnet werden.

Es sei p die gegebene Wahrscheinlichkeit/die gewünschte Fläche unter der Verteilungsfunktion für das Quantil oder den ZSB. Wir definieren α als den Kehrwert $\alpha = 1 - p$ von p .

Soll nun ein ZSB berechnet werden so passiert dies über die Formeln:

Nach oben beschränkt	$(-\infty ; q_{1-\alpha}]$	=	$(-\infty ; \mu + z_{1-\alpha} \cdot \sigma]$
Nach unten beschränkt	$[q_\alpha ; \infty)$	=	$[\mu - z_{1-\alpha} \cdot \sigma ; \infty)$
Beidseitig beschränkt	$[q_{\frac{\alpha}{2}} ; q_{1-\frac{\alpha}{2}}]$	=	$[\mu - z_{1-\frac{\alpha}{2}} \cdot \sigma ; \mu + z_{1-\frac{\alpha}{2}} \cdot \sigma]$

Auch hier kann der Taschenrechner eingesetzt werden (InvNormal). Hier sind die Werte **Area**, μ und σ verlangt. μ und σ sind meist in der Aufgabenstellung gegeben, für Area muss:

einseitiger Grenzwert:		
	nach oben beschränkt:	p
	nach unten beschränkt:	α
zweiseitiger Grenzwert:		
	untere Grenze:	$\frac{\alpha}{2}$
	obere Grenze:	$1 - \frac{\alpha}{2}$

Ist ein Datensatz mit n Elementen gegeben, so ändern sich die Formeln zu:

Nach oben beschränkt	$(-\infty ; q_{1-\alpha}]$	=	$(-\infty ; \mu + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}]$
Nach unten beschränkt	$[q_\alpha ; \infty)$	=	$[\mu - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} ; \infty)$
Beidseitig beschränkt	$[q_{\frac{\alpha}{2}} ; q_{1-\frac{\alpha}{2}}]$	=	$[\mu - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} ; \mu + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}]$

Nun muss selbstverständlich für σ in der InvNormal-Funktion $\frac{\sigma}{\sqrt{n}}$ eingegeben werden.

Teil II

Deskriptive Statistik