bio
biologie

Matthias Rudolf
Wiltrud Kuhlisch

Biostatistik

Eine Einführung für Biowissenschaftler

Studentengetestet!

WWW.PEARSON-STUDIUM.DE
CWS
COMPANION WEBSITE

PEARSON
Studium

Matthias Rudolf
Wiltrud Kuhlisch

Biostatistik

Eine Einführung für Biowissenschaftler

PEARSON
Studium

ein Imprint von Pearson Education
München · Boston · San Francisco · Harlow, England
Don Mills, Ontario · Sydney · Mexico City
Madrid · Amsterdam

Prinzip von Monte-Carlo-Studien

Monte-Carlo-Studien können für sehr unterschiedliche Forschungsfragen eingesetzt werden, zum Beispiel bei der Vorhersage der 3D-Struktur von Proteinen (siehe Rohl et al., 2004). Ihr Grundprinzip soll am Beispiel der Untersuchung der Robustheit statistischer Tests bei nicht gegebenen Voraussetzungen veranschaulicht werden.

In der Situation von Anwendungsbeispiel 1 soll angenommen werden, dass die Ernteerträge in der Population keiner Normalverteilung unterliegen. Ernteerträge unter 95 t/ha kämen praktisch nicht vor, hohe Ernteerträge über 110 t/ha seien dagegen nicht selten. Die Ernteerträge von 20 Feldern könnten ausgewertet werden. Die Verteilung sei schief (linkssteil), die Parameter dieser schiefen Verteilung seien bekannt. Bei Gültigkeit der Nullhypothese $H_0 : \mu_{dD} = \mu_0 = 100\, t\,/\,ha$ würden die Ernteerträge in der Population also einer linkssteilen Verteilung mit Populationsmittelwert 100 unterliegen.

Bei der Anwendung der Monte-Carlo-Methode wird diese Verteilung simuliert. Aus der simulierten schiefen Verteilung werden sehr viele (in der Regel mindestens 1000–4000, selten mehr als 10000) Stichproben des Umfangs 20 gezogen. Vereinfacht ausgedrückt besteht das weitere Vorgehen darin, dass für jede der gezogenen Stichproben der Signifikanztest mit dem Signifikanzniveau α durchgeführt wird. Da in der zugrunde liegenden Population die Nullhypothese $H_0 : \mu_{dD} = \mu_0 = 100\, t\,/\,ha$ gültig ist, würde H_0 bei erfüllten Voraussetzungen in ca. $\alpha \cdot 100\%$ der Stichproben zurückgewiesen, die Wahrscheinlichkeit der Fehlers 1. Art wäre durch α begrenzt. In der Simulation, unter der Annahme der schiefen Populationsverteilung, wird die H_0 in $\gamma \cdot 100\%$ der Stichproben zurückgewiesen. Aus dem Vergleich von γ und α lassen sich die Eigenschaften des statistischen Tests für die untersuchte Verletzung der Voraussetzung ableiten:

- $\gamma < \alpha$: Der Test reagiert konservativ. Die H_0 wird mit einer geringeren Wahrscheinlichkeit als α abgelehnt, wenn sie in der Population zutrifft.
- $\gamma \approx \alpha$: Der Test reagiert robust. Die in der Population gültige H_0 wird trotz der untersuchten Verletzung der Voraussetzung ungefähr mit der Wahrscheinlichkeit α abgelehnt.
- $\gamma > \alpha$: Die H_0 wird mit einer größeren Wahrscheinlichkeit als α abgelehnt, obwohl sie in der Population zutrifft.

Die für die Anwendung des statistischen Tests zu ziehenden Schlussfolgerungen sind im vorhergehenden Abschnitt ausführlich beschrieben. Ausdrücklich muss darauf hingewiesen werden, dass eine Monte-Carlo-Studie nur eine Auskunft über die Eigenschaften des Tests bezüglich der konkret betrachteten Abweichung von der Voraussetzung liefern kann. Wenn im Beispiel die zu unterstellende Abweichung von der Normalverteilung nicht in der Schiefe der Verteilung, sondern in einer Zweigipfligkeit bestehen würde, wäre eine spezielle Monte-Carlo-Studie erforderlich, die diesen Verteilungstyp simulieren würde. Deshalb ist für viele Tests eine zusammenfassende Bewertung von allen möglichen Verletzungen der Voraussetzungen nicht kompakt verfügbar. In entsprechender Fachliteratur findet man aber Ergebnisse von Simulationen für sehr viele Tests und für unterschiedlichste Verletzungen von Voraussetzungen. Dabei wird im allgemeinen Fall die Robustheit nicht nur für einen Parameter, sondern für alle Parameter untersucht (zum Beispiel für alle möglichen Parameter μ_0).

5.5.2 Die Bootstrap-Technik

Die Bootstrap-Technik (zurückgehend auf Efron, 1979 und 1982) ist das wichtigste Resampling-Verfahren (siehe Manly, 2007). Sie wird vor allem mit dem Ziel eingesetzt, Parameterschätzungen oder statistische Entscheidungen aus sehr kleinen Stichproben abzuleiten. Die Bootstrap-Technik kann hier nur mit ihren Grundideen einführend dargestellt werden.

Kleine Stichproben entstehen in den Biowissenschaften vor allem dann, wenn die Erfassung eines Merkmals sehr hohen technischen Aufwand erfordert, oder wenn aus anderen Gründen nur sehr wenige Untersuchungsobjekte zur Verfügung stehen (seltene Tier- oder Pflanzenarten). Die Bootstrap-Technik verwendet Informationen einer gegebenen Stichprobe mit dem Ziel, Vorstellungen über die Variabilität bzw. die Verteilung des zu untersuchenden Stichprobenkennwerts zu gewinnen. Das prinzipielle Vorgehen soll anhand von **Anwendungsbeispiel 1** in groben Zügen skizziert werden. Dabei soll davon ausgegangen werden, dass nur die Ernteerträge von sieben Feldern zur Verfügung stehen und weitere Daten nicht erhoben werden können. Über die Verteilung des Merkmals in der Population sei nichts bekannt, von Normalverteilung könne nicht ausgegangen werden. Die Messwerte seien

99, 104, 104, 98, 101, 103, 102.

Das Prinzip der Bootstrap-Methode besteht darin, aus dieser Stichprobe eine große Anzahl (üblicherweise mindestens 1000, manchmal auch mehr als 10000) Bootstrap-Stichproben mit Zurücklegen zu ziehen. Dabei ergeben sich für $n = 7$ insgesamt $7^7 = 823543$ theoretisch mögliche unterschiedliche Stichproben. Mögliche Bootstrap-Stichproben sind zum Beispiel

101, 99, 101, 103, 102, 101, 99 oder
102, 102, 102, 102, 102, 102, 102 oder
102, 101, 102, 99, 98, 98, 98.

Für jede dieser Bootstrap-Stichproben kann der arithmetische Mittelwert $\bar{x}$ (oder in anderen Beispielen ein anderer interessierender statistischer Kennwert) berechnet werden. Nach der Berechnung des Standardfehlers der Mittelwerte kann im einfachsten Fall ein Konfidenzintervall für den unbekannten Populationsmittelwert berechnet werden. Wenn das Konfidenzintervall den Wert $\mu_0 = 100$ nicht enthält, ist die statistische Nullhypothese aus Anwendungsbeispiel 1 abzulehnen (siehe Kapitel 6). Bei der Interpretation ist zu beachten, dass das so ermittelte Konfidenzintervall nur für die untersuchte Stichprobe und nicht für Stichproben aus anderen vergleichbaren Untersuchungen gültig ist.

Zusammenfassung

Um mathematisch begründete Entscheidungen über Forschungshypothesen treffen zu können, müssen die zu untersuchenden inhaltlichen Hypothesen in statistische Hypothesen über Verteilungen von Merkmalen in Populationen bzw. über deren Parameter überführt werden. Dabei ergibt sich die statistische Alternativhypothese H_1 unmittelbar aus der inhaltlichen Hypothese, während mit der statistischen Nullhypothese H_0 behauptet wird, dass die zu H_1 komplementäre Aussage richtig sei.

Beim Testen dieser statistischen Hypothesen wird auf der Grundlage von Wahrscheinlichkeitsaussagen zur Nullhypothese eine Entscheidung darüber getroffen, ob die Alternativhypothese (und damit die inhaltliche Forschungshypothese) angenommen werden kann oder nicht. Da damit auf der Grundlage von Ergebnissen aus Stichproben Aussagen über die Populationen getroffen werden, sind Fehler bei den Entscheidungen möglich. Die inhaltlichen Konsequenzen der möglicher Fehlentscheidungen können unterschiedlich gravierend sein und müssen im Rahmen der jeweiligen Forschungsfrage beurteilt werden.

Zur statistischen Prüfung der Nullhypothese H_0 wird der p-Wert als die Wahrscheinlichkeit berechnet, mit der bei Gültigkeit der Nullhypothese der in der Stichprobe berechnete Schätzwert oder ein noch mehr der H_0 widersprechender Schätzwert ermittelt wird. Wenn dieser Wert kleiner als ein vorgegebenes Signifikanzniveau α ist, kann die H_0 abgelehnt und die H_1 akzeptiert werden. Alternativ ist die Testentscheidung über den Vergleich des Schätzwerts der Teststatistik mit dem kritischen Wert der Prüfverteilung möglich. Bei der Interpretation des Testergebnisses muss die Frage beachtet werden, ob gefundene statistisch signifikante Effekte auch praktisch bedeutsam sind.

Moderne rechenintensive Methoden können zur Untersuchung der Robustheit von Tests oder zur Gewinnung von Aussagen über die Verteilung von Stichprobenkennwerten bei kleinen Stichproben hinzugezogen werden.

Übungsaufgaben

Aufgabe 5.1

Die (bekannte) mittlere Anzahl gelegter Eier bei Hühnern soll erhöht werden, wozu ein Mineralstoffpräparat entwickelt wurde. In einer Untersuchung, für die nur 15 Hühner zur Verfügung stehen, soll die Wirksamkeit des Mittels, das ohne unerwünschte Nebenwirkungen für die Tiere ist, getestet werden. Man kann davon ausgehen, dass sich die Zahl gelegter Eier in der Population durch die Einnahme des Präparats nicht verringern wird.

Welche statistischen Hypothesen sind zu untersuchen? Welche inhaltlichen Folgen haben in diesem Beispiel ein möglicher Fehler 1. Art und ein möglicher Fehler 2. Art? Welche Fehlermöglichkeit sollte aus Sicht des am Gewinn orientierten Bauern möglichst gering gehalten werden, wenn das Hormonpräparat keine nennenswerten Mehrkosten verursacht?

Aufgabe 5.2

In einer Untersuchung möge unter Annahme der H_0 ein Populationsmittelwert von $\mu_0 = 70$ erwartet werden. In der empirischen Untersuchung ergab sich ein arithmetischer Mittelwert von $\bar{x} = 73.2$. Die Abweichung sei bei zweiseitigem Test auf dem 5%-Niveau nicht signifikant. Wäre die gleiche Abweichung auch bei einseitigem Test ($H_1 : \mu > \mu_0$) nicht signifikant ($\alpha = 0.05$)?

Aufgabe 5.3

In einer Untersuchung möge unter Annahme der H_0 ein Populationsmittelwert von $\mu_0 = 70$ erwartet werden. In der empirischen Untersuchung ergab sich jedoch ein arithmetischer Mittelwert von $\bar{x} = 74.2$. Die Abweichung sei bei einseitigem Test ($H_1 : \mu > \mu_0$) auf dem 1%-Niveau signifikant. Wäre die gleiche Abweichung bei zweiseitigem Test auf dem 5%-Niveau signifikant?

Aufgabe 5.4

In einer Untersuchung möge unter Annahme der H_0 ein Mittelwert von $\mu_0 = 90$ erwartet werden. In der empirischen Untersuchung ergab sich ein arithmetischer Mittelwert von $\bar{x} = 93.2$. Die Abweichung sei bei zweiseitigem Test auf dem 1%-Niveau signifikant. Wäre die gleiche Abweichung auch bei einseitigem Test ($H_1 : \mu > \mu_0$) signifikant ($\alpha = 0.01$)?

Aufgabe 5.5

In einer Untersuchung möge unter Annahme der H_0 ein Populationsmittelwert von $\mu_0 = 82$ erwartet werden. In der empirischen Untersuchung ergab sich jedoch ein arithmetischer Mittelwert von $\bar{x} = 75.2$. Die Abweichung sei bei zweiseitigem Test auf dem 5%-Niveau signifikant. Wäre die gleiche Abweichung auch bei einseitigem Test ($H_1 : \mu < \mu_0$) auf dem 1%-Niveau signifikant?

Aufgabe 5.6

In einer Untersuchung soll nachgewiesen werden, dass die durchschnittliche Lebensdauer von Tieren bei (für den Tierhalter aufwendiger) Freilandhaltung größer ist als die mittlere Lebensdauer der Tiere bei (weniger aufwendiger) Stallhaltung. Da die Freilufthaltung artgerechter ist, kann eine Verkürzung der mittleren Lebensdauer durch die Freilufthaltung ausgeschlossen werden.

Welche statistischen Hypothesen sind zu untersuchen? Welche inhaltlichen Folgen haben ein möglicher Fehler 1. Art und ein möglicher Fehler 2. Art? Welche Fehler-wahrscheinlichkeit sollte aus der Sicht des Tierschutzes möglichst gering gehalten werden?

Aufgabe 5.7

Beurteilen Sie folgende Aussage: Ein p-Wert von 0.02 bedeutet, dass die Nullhypothese, die der Berechnung des p-Werts zugrunde lag, mit einer Wahrscheinlichkeit von 0.98 (98 Prozent) korrekt ist.

Ausführliche Lösungen sowie weitere Aufgaben finden Sie auf der Companion Website zum Buch unter **http://www.pearson-studium.de**

Ausgewählte statistische Tests

6

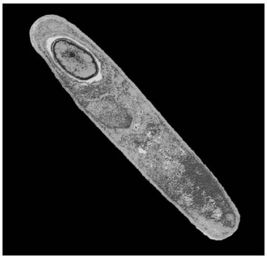
ÜBERBLICK

In diesem Kapitel werden ausgewählte statistische Tests vorgestellt, die einerseits in der biostatistischen Praxis besonders häufig angewendet werden und die sich andererseits jeweils durch unterschiedliche Vorgehensweisen auszeichnen. Neben der detaillierten und anwendungsorientierten Beschreibung der konkreten Durchführung der Tests steht die Diskussion der notwendigen Voraussetzungen im Mittelpunkt. Dabei wird zwischen parametrischen und nichtparametrischen Tests sowie zwischen Tests für unabhängige bzw. für verbundene Stichproben unterschieden. Mit der gewählten Form der Darstellung sollen die grundlegenden Prinzipien der behandelten Tests veranschaulicht werden. Für weitergehende Übersichten über verfügbare Signifikanztests wird auf entsprechende Literatur verwiesen.

Folgende Tests werden ausführlich beschrieben:

- Parametrische Tests für normalverteilte Populationen:
  - Eine Stichprobe: t-Test gegen eine Konstante.
  - Zwei unabhängige Stichproben: t-Test für unabhängige Stichproben.
  - Zwei verbundene Stichproben: t-Test für verbundene Stichproben.
  - Äquivalenztest für verbundene Stichproben.
- Tests zum Prüfen von Voraussetzungen:
  - Varianzhomogenität: Levene-Test.
  - Normalverteilung: Kolmogorov-Smirnov-Test (mit bzw. ohne Lillefors-Korrektur).
  - Normalverteilung: Grafische Überprüfung durch Q-Q-Plot.
- Tests für ordinalskalierte Merkmale:
  - Zwei unabhängige Stichproben: Exakter und asymptotischer U-Test.
  - Zwei verbundene Stichproben: Exakter und asymptotischer Wilcoxon-Test.
- Tests für nominalskalierte Merkmale:
  - Zwei unabhängige Stichproben: Exakter und asymptotischer Test.
  - Zwei verbundene Stichproben: Exakter und asymptotischer Test.

Anwendungsbeispiel



Bacillus thuringiensis mit insektizidem Protein

Der Maiszünsler ist einer der wirtschaftlich bedeutendsten Maisschädlinge in Deutschland. Die Larven dieses Kleinschmetterlings fressen zuerst an den Blättern und bohren sich später in die Kolben der Maispflanze. Die Bekämpfung dieses Schädlings erweist sich auch mit chemischen Mitteln als schwierig. Erfolgreich ist zum Beispiel die biologische Bekämpfung mit der Schlupfwespe *Trichogramma*, deren Larven die Eier des Maiszünslers parasitieren. Ein klassisches Konzept (allerdings mit geringerem Wirkungsgrad) ist die Verwendung von Präparaten mit Kulturen des Bakteriums *Bacillus thuringiensis*, welches ein giftiges Protein bildet, das die Darmwand bestimmter Fraßinsekten zerstört. Andere Insektenarten (und Schädlinge), so zum Beispiel auch Blattläuse, sind von dieser Wirkung nicht betroffen.

Die Wirkung eines *Bacillus-thuringiensis*-Präparats soll überprüft werden. Dazu wurden zehn Felder mit dem Präparat behandelt, zehn Felder blieben unbehandelt. Auf allen in die Untersuchung einbezogenen Feldern wurden zwei Tage nach Behandlung der durchschnittliche Befall mit Maiszünsler-Raupen und das Krankheitsbild (Bonitur) der Pflanzen untersucht. Nach weiteren fünf Tagen wurden Befall und Bonitur für die behandelten Felder erneut erhoben. Nach 14 Tagen wurde der Befallsstatus (befallen/nicht befallen) auf allen Feldern kontrolliert, diese Erfassung wurde nach weiteren 14 Tagen bei den behandelten Feldern wiederholt. Weiterhin wurde die Auswirkung der *Bacillus-thuringiensis*-Präparate auf den Befall mit Blattläusen untersucht, indem auf den behandelten Feldern die Anzahl der Blattläuse vor der Behandlung und 14 Tage danach ermittelt wurde. Der durchschnittliche Maisertrag beträgt bei unbehandelten Feldern etwa 50 dt/ha. Der Ertrag der behandelten Felder wurde drei Tage nach der Behandlung erfasst. In ►Tabelle 6.1, ►Tabelle 6.2 und ►Tabelle 6.3 sind die Variablen, die erhobenen Daten und die zu untersuchenden inhaltlichen Hypothesen zusammengestellt.

Merkmal	Skalenniveau	Erfassung	Erläuterungen
G: Behandlungsgruppe	nominal (alternativ)		2 Ausprägungen (1: behandelt; 2: unbehandelt)
M: Befall mit Maiszünsler nach 2 Tagen	metrisch	behandelte und unbehandelte Gruppe	Anzahl Raupen pro 100 Pflanzen
MW: Befall mit Maiszünsler nach 7 Tagen	metrisch	behandelte Gruppe	Anzahl Raupen pro 100 Pflanzen

Tabelle 6.1: Variablen im Anwendungsbeispiel.

Merkmal	Skalenniveau	Erfassung	Erläuterungen
BL: Befall mit Blattläusen vor der Behandlung	metrisch	behandelte Gruppe	Anzahl Blattläuse pro 100 Pflanzen
BLW: Befall mit Blattläusen nach 2 Wochen	metrisch	behandelte Gruppe	Anzahl Blattläuse pro 100 Pflanzen
BO: Bonitur nach 2 Tagen	ordinal	behandelte und unbehandelte Gruppe	skaliert zwischen 0 (keine Schadensymptome) und 9 (starke Schadensymptome)
BOW: Bonitur nach 7 Tagen	ordinal	behandelte Gruppe	skaliert zwischen 0 (keine Schadensymptome) und 9 (starke Schadensymptome)
B: Befallsstatus nach 2 Wochen	nominal (alternativ)	behandelte Gruppe	2 Ausprägungen (0: kein Befall; 1: Befall)
BW: Befallsstatus nach 4 Wochen	nominal (alternativ)	behandelte Gruppe	2 Ausprägungen (0: kein Befall; 1: Befall)
E: Ertrag nach 3 Tagen	metrisch	behandelte Gruppe	in dt/ha

Tabelle 6.1: Variablen im Anwendungsbeispiel (Fortsetzung).

Nummer des Feldes	Gruppe	*M*	*MW*	*BL*	*BLW*	*BO*	*BOW*	*B*	*BW*	*E*
1	behandelt	63	63	1000	900	0	0	0	0	52
2	behandelt	71	75	1100	1500	2	1	1	0	55
3	behandelt	62	60	1200	1100	1	4	0	1	48
4	behandelt	56	56	1400	1700	4	5	0	1	55
5	behandelt	64	62	1800	1600	2	6	0	0	51
6	behandelt	63	64	1400	1300	5	1	0	0	58
7	behandelt	56	61	3200	3400	3	1	0	1	56
8	behandelt	71	76	1400	1800	6	0	1	0	58
9	behandelt	58	63	800	600	4	4	0	1	62
10	behandelt	68	60	1000	900	4	3	0	0	55
11	unbehandelt	62				9		0		

Tabelle 6.2: Daten im Anwendungsbeispiel.

Nummer des Feldes	Gruppe	M	MW	BL	BLW	BO	BOW	B	BW	E
12	unbehan- delt	72				8		1		
13	unbehan- delt	70				6		1		
14	unbehan- delt	66				4		1		
15	unbehan- delt	74				7		1		
16	unbehan- delt	75				6		1		
17	unbehan- delt	70				5		0		
18	unbehan- delt	65				9		0		
19	unbehan- delt	66				3		0		
20	unbehan- delt	80				2		1		

Tabelle 6.2: Daten im Anwendungsbeispiel (Fortsetzung).

Inhaltliche Hypothese	Klassifizierung der Hypothese
A) Der mittlere Ernteertrag (Merkmal E) der behandelten Flächen ist nach drei Tagen höher als der bekannte mittlere Ernteertrag bei unbehandelten Flächen ($\mu_0 = 50\, dt\, /\, ha$). Für die Population kann eine Verminderung des mittleren Ernteertrags ausgeschlossen werden, da das eingesetzte Mittel keine bekannten schädlichen Nebenwirkungen hat.	Unterschiedshypothese gerichtet 1 Stichprobe metrische Daten → Abschnitt 6.1.1
B) Der mittlere Befall mit Maiszünsler-Raupen (Merkmal M) ist nach zwei Tagen in der behandelten Gruppe niedriger als in der unbehandelten Gruppe. Für die Population kann eine Erhöhung des Befalls in Folge der Behandlung ausgeschlossen werden, da das eingesetzte Mittel keine bekannten schädlichen Nebenwirkungen hat.	Unterschiedshypothese gerichtet 2 unabhängige Stichproben metrische Daten → Abschnitt 6.1.2

Tabelle 6.3: Klassifizierung der Hypothesen im Anwendungsbeispiel.