

Methoden der medizinischen Wissenschaft (SSM2)

Grundlagen der Biostatistik und Versuchsplanung Standards

Vorlesung
Lernziele

1. Quellen der Variabilität

- Art und Gewinnung von Daten
- Biologische Variabilität
- Inter- und intraindividuelle Variabilität
- Variabilität zwischen Beurteilern
- Messfehler
- Zufällige und systematische Fehler

- Datenqualität
- Validität und Reliabilität
- Datenschutz

- Vollerhebung und Stichprobe
- Erhebung - Experiment
- Schließen unter Unsicherheit

Die Studierenden sollen anhand von Beispielen in die unterschiedlichen Arten von medizinischen Daten eingeführt werden.

Die Variabilität von Daten aus wiederholten Messungen am selben Individuum sowie aus gleichen Messungen an verschiedenen Individuen soll als die Regel (und nicht als die Ausnahme) für biologische Phänomene verstanden werden. Die verschiedenen Arten der Variabilität sollen definiert werden (Biologische Variabilität, Inter- und intraindividuelle Variabilität, Beurteilervariabilität, Messfehler). Damit verbunden sollen die möglichen Ursachen der Variabilität anhand von Beispielen angesprochen werden.

Es wird anhand eines Beispiels mit Heilungsraten auf die Schwierigkeiten hingewiesen, die sich beim Auftreten von fehlenden Werten für die statistische Auswertung ergeben. Mit der grundsätzlichen Unterscheidung in medizinische Beobachtungsstudien, im Rahmen derer gängige medizinische Praxis erhoben wird, oder experimentelle Studien, bei denen der Untersucher die Behandlung den Individuen zuordnet (z.B. zufällig), soll hier anhand von Beispielen an die Lehrveranstaltung „Grundlagen der Epidemiologie“ (Block 6) angebunden werden. Eine kurze Diskussion der aus diesen unterschiedlichen Studientypen zu

erwartenden Evidenz schließt auch den Bogen zu der vorangegangenen Lehrveranstaltung „Evidence Based Medicine“ (Block 7).

Anhand eines Beispiels über den Vergleich von zwei Messmethoden wird exemplarisch die statistische Auswertung demonstriert. Schließlich wird an einem unterhaltsamen Beispiel die Frage gestellt, wie aus unsicheren Daten Schlüsse gezogen werden sollen (und somit eine Überleitung zum nächsten Abschnitt gebildet): Sie sitzen mit einem Kollegen am Abend an einem Ort, an dem gegen bescheidenes Entgelt Erfrischungen gereicht werden. Bisher wurden siebenmal Getränke an Sie und Ihr Gegenüber gereicht, wobei jeweils der Wurf der Münze (durch den Kollegen) entschieden hat, wer die „Runde“ an Getränken bezahlen musste. Sie haben bisher jedes Mal (also insgesamt siebenmal) bezahlen müssen. Bei der kommenden (zu Beginn des Abends als letzte ausgemachten) Runde fällt das Los des Bezahlers wieder auf Sie. Was halten Sie davon? Welche Konsequenzen ziehen Sie daraus?

2. Wahrscheinlichkeit als Basis statistischen Schließens

- Relative Häufigkeiten als Surrogat für Wahrscheinlichkeiten
- Andere Interpretationen des Begriffs der Wahrscheinlichkeit
- Regeln für das Rechnen mit Wahrscheinlichkeiten
- Wahrscheinlichkeitsbaum
- Wahrscheinlichkeitsverteilungen
- Das Entstehen einer Normalverteilung
- Bedingte Wahrscheinlichkeiten

Um einen entspannten Zugang zu diesem Abschnitt zu ermöglichen, sollte zunächst darauf hingewiesen werden, dass der Begriff Wahrscheinlichkeit im Laufe der Geistesgeschichte dem Menschen stets erhebliche Schwierigkeiten bereitet hat. In einem pragmatischen Zugang soll der Begriff als Grenzwert relativer Häufigkeiten interpretiert werden: Wenn ein Experiment sehr oft wiederholt wird, dann kann die relative Häufigkeit der Ausgänge des Experiments als ihre Wahrscheinlichkeit interpretiert werden. Am Beispiel der Bestimmung der Blutgruppe 0 bei einer großen Zahl von Österreicherinnen und Österreicher soll der Begriff der Wahrscheinlichkeit der Blutgruppe 0 für („zufällig herausgegriffene“) Österreicherinnen oder Österreicher erklärt werden. Lernziel dieser Einführung ist ein Grundverständnis der Begriffe Wahrscheinlichkeit und Odds („Chance“, siehe auch „Einführung in die Epidemiologie“, Block 6).

Andere Interpretationen für Wahrscheinlichkeiten (z.B. über die persönliche Einschätzung der Chancen) werden erwähnt. Anhand dieses Beispiels sollen die Regeln des Rechnens mit Wahrscheinlichkeiten pragmatisch erarbeitet werden. Unter der Annahme, dass diese Wahrscheinlichkeit 0.4 beträgt (gerundet), wird zunächst die Wahrscheinlichkeit berechnet, dass unter einer Gruppe von 8 Personen alle als Universalspender mit der Gruppe 0 (der Einfachheit ohne Beachtung des Rhesusfaktors) in Betracht kommen. Auch die Wahrscheinlichkeit für andere Ergebnisse (z.B. keine Person, genau eine Person, eine oder zwei Personen, mindestens 1 Person mit Blutgruppe 0 unter den acht Personen) werden berechnet. Die dabei spontan und naiv angewandten Rechenregeln werden reflektiert und ihre Voraussetzungen (Unabhängigkeit der Individuen) diskutiert. Dies endet mit einer verbalen Formulierung der wenigen Grundsätze der Wahrscheinlichkeitsrechnung.

Lernziel dieses Abschnitts ist ein Grundverständnis für die Regeln des Rechnens mit Wahrscheinlichkeiten.

Das Beispiel wird in natürlicher Weise erweitert, indem die Wahrscheinlichkeiten für sämtliche 9 Ausgänge, 0, 1, 2, ..., 8 Universalspender unter acht Personen, als Wahrscheinlichkeitsfunktion in Form eines Stabdiagramms dargestellt werden. Hier schafft ein Verweis auf Stabdiagramme für diskrete Daten den Bogen zu den Übungen. Mittel- oder Erwartungswert dieser Verteilung werden eingeführt und dem Konzept des Mittelwerts aus Stichprobendaten gegenübergestellt, wie es in den Übungen eingeführt wird. Mit dem Hinweis auf die Binomialverteilung wird die generelle Berechenbarkeit solcher Wahrscheinlichkeiten erwähnt. Lernziel ist das Verständnis von diskreten Wahrscheinlichkeitsverteilungen und eine Annäherung an das Konzept der Binomialverteilung unter Vermeidung formaler Hilfsmittel.

Das Beispiel wird fortgeführt, indem die Anzahl der Personen in der Gruppe von 8 auf 12, 16, 20, 40, 60, 80 und 100 erhöht wird. Dies demonstriert, dass sich die Wahrscheinlichkeitsverteilung der Anzahl der Personen mit der Blutgruppe 0 ausgehend von einer schiefen Verteilung bei der Gruppengröße von acht mit größer werdendem Gruppenumfang relativ rasch zu einer symmetrischen Glockenkurve entwickelt. Die Charakterisierung dieser symmetrischen Verteilung durch Mittelwert und Standardabweichung stellt die Verbindung zu den in den Übungen behandelten Lage- und Streuungsmaßen für Stichprobendaten her. Lernziel dieses Teils ist das Verständnis für die Rolle des Stichprobenumfangs bei Häufigkeitsverteilungen und dafür, dass die einfach charakterisierbare und am meisten in der Medizin verwendete statistische Verteilung, die Normalverteilung, als Verteilung der Summe einer „sehr großen“ Anzahl von (zufälligen) Größen resultiert.

Beendet wird diese Einheit mit dem Begriff der bedingten Wahrscheinlichkeit, der am Beispiel eines diagnostischen Tests an Gesunden und Kranken über die Begriffe Sensitivität, Spezifität, positiver und negativer prädiktiver Wert an einem Beispiel eingeführt wird. Lernziel ist das Verständnis bedingter Wahrscheinlichkeiten und ihrer Bedeutung für medizinische Entscheidungen, wobei auch auf „Grundlagen der Epidemiologie“ (Block 6) verwiesen werden kann.

3. Der statistische Test

- Das Prinzip der Falsifizierung von Hypothesen
- Die „skeptische“ Ausgangshypothese (Nullhypothese)
- Das zu beweisende Gegenteil (Alternativhypothese)
- Welche Ergebnisse erwarten wir für unsere Stichprobe unter der skeptischen Ausgangshypothese
- Welche Ergebnisse haben wir in unserer Stichprobe tatsächlich erhalten
- Bleiben die Abweichungen unserer Ergebnisse von diesen erwarteten Ergebnissen in einem durch den Zufall erklärbaren Rahmen, so haben wir keine ausreichende Begründung von der skeptischen Ausgangshypothese abzugehen
- Weichen die Ergebnisse jedoch zu stark von diesen Erwartungen ab, so lassen wir uns von unseren Beobachtungen überzeugen (und gehen von der skeptischen Ausgangshypothese ab)
- Welche Fehler sind bei einer solchen Entscheidungsprozedur möglich
 - o Signifikanzniveau
 - o Power

- p-Wert
- Konfidenzintervalle
- Einige wichtige Testverfahren

Dieses Modul knüpft an das unterhaltsame Beispiel aus der ersten Vorlesung an. Die „skeptische“ Ausgangshypothese kann in diesem Fall mit „Der Kollege ist ein fairer Spieler“ formuliert werden. Dies kann mit der Nullhypothese „die Gewinnwahrscheinlichkeit pro Münzwurf für Sie und Ihren Kollegen ist gleich $\frac{1}{2}$ “ formalisiert werden. Die Alternativhypothese ist, dass die Münzwürfe nicht fair waren, also die Gewinnwahrscheinlichkeit für Sie von $\frac{1}{2}$ abweicht. Die (unvoreingenommene) Formulierung in beide Richtungen lässt dabei offen, dass der Kollege auch eine Münze verwenden könnte, die Sie bevorzugen (zweiseitige Sicht des Problems). Eine einseitige Sicht des Problems würde sich nur auf die Beantwortung der Frage ausrichten, ob Ihre Gewinnwahrscheinlichkeit geringer als $\frac{1}{2}$ ist, also der Kollege ein unfaires Spiel zu Ihren Ungunsten betreibt. Da in der medizinischen Forschungspraxis generell die zweiseitige Sicht forciert wird, wird diese auch in dem Beispiel weiter verfolgt.

In Analogie zur diskreten Wahrscheinlichkeitsverteilung für die Häufigkeit der Blutgruppe 0 unter acht Personen in der 2. Vorlesung wird die jetzt symmetrische Wahrscheinlichkeitsverteilung für die Anzahl von siegreichen Spielen unter acht Versuchen bei Voraussetzung eines fairen Spiels mit der Gewinnwahrscheinlichkeit von $\frac{1}{2}$ pro Spiel dargestellt. Die Wahrscheinlichkeit für die extremsten Ausgänge 0 oder 8 Siege wird berechnet, ebenso für die Ausgänge 1 oder 7 Siege.

Zwei Entscheidungsregeln werden betrachtet:

1. Wenn ich niemals oder achtmal bezahlen muss („kritischer Bereich“ für den Ausgang des Spiels), dann ist mein Zweifel an der Fairness der Münze zu groß, z.B. mit der Konsequenz, dass ich sie mir zeigen lasse (ich rücke von der Ausgangshypothese eines fairen Spiels ab). Sonst finde ich mich mit dem Ergebnis ab, ohne an der Fairness des Spiels zu zweifeln.
2. Ich rücke von der Ausgangshypothese eines fairen Spiels schon dann ab, wenn ich einmal oder siebenmal gewinne (natürlich zählt dann auch keinmal oder achtmal zum kritischen Bereich). Sonst finde ich mich mit dem Ergebnis ab, ohne an der Fairness des Spiels zu zweifeln.

Zwei Arten von Fehlern können bei derartigen Entscheidungsregeln auftreten. Der erste der beiden Fehler kann passieren, wenn der Kollege tatsächlich fair spielt (die Nullhypothese trifft also zu). Dann besteht für mich die Fehlentscheidung im Abrücken von der richtigen Annahme eines fairen Spiels (falsche Verdächtigung des Kollegen, irrtümliche Verwerfung der Nullhypothese). Die zweite Art des Fehlers kann passieren, wenn vom Kollegen tatsächlich eine verfälschte Münze verwendet wird (die Alternativhypothese trifft zu). Der Ausgang des Spiels ist jedoch nicht extrem genug (etwa 3 Siege für sie, 5 für den Kollegen), sodass ich mich irrtümlicherweise mit dem Ergebnis abfinde, da es auch bei einem fairen Spiel gut durch den Zufall erklärt werden kann. (In diesem Fall trifft die Fehlentscheidung mich selbst, da ich ein unfaires Spiel akzeptiere.) Die Entscheidungstafel mit der wahren Natur des Spiels (fair oder nicht fair) einerseits und den zwei Testentscheidungen aus den beobachteten Ergebnissen (Verwerfen oder Beibehaltung der Nullhypothese) andererseits wird in Verbindung mit den zugehörigen Fehlerwahrscheinlichkeiten diskutiert.

Die Entscheidungen werden auch in Hinblick auf mögliche Konsequenzen analysiert (Streit mit dem Kollegen, Verlust der Freundschaft, Kosten des Spielverlusts usw.). Die Irrtumswahrscheinlichkeiten für eine Fehlentscheidung der ersten Art werden für beide Entscheidungsregeln berechnet. Das Signifikanzniveau als vorgegebene obere Grenze für diese Wahrscheinlichkeit wird eingeführt. Das Konzept des p-Werts wird ebenfalls an diesem Beispiel erklärt.

Lernziel ist das Grundverständnis dafür, dass Schließen unter Unsicherheit (aus zufällig schwankenden Größen) ganz allgemein mit der Möglichkeit des Irrtums konfrontiert ist. Statistische Methoden versuchen unter anderem, diese Irrtumswahrscheinlichkeiten zu quantifizieren und Entscheidungsregeln zu definieren, bei denen diese Wahrscheinlichkeiten kontrolliert oder begrenzt werden.

Der Bezug des Beispiels zu biologischen Phänomenen wird an einem Vergleich von zwei Therapien A und B am gleichen Patienten demonstriert (siehe auch Vorlesung 4). Dabei wird jeder Patient gefragt, welche der beiden erhaltenen Therapien er als wirksamer gegen seine Erkrankung einschätzt. Wenn die beiden Therapien völlig gleich wirksam sind (Nullhypothese), dann sind die Wahrscheinlichkeit für die Angaben der Präferenz „A ist besser als B“ oder „B ist besser als A“ für den einzelnen Patienten jeweils gleich $\frac{1}{2}$. Das Antwortverhalten in einer Stichprobe von 8 mit A und B behandelten unabhängigen Patienten würde dann exakt durch die Wahrscheinlichkeitsverteilung des obigen Münzwurfbeispiels beschrieben werden können. Die Konsequenzen für Fehlentscheidungen bei einem entsprechenden statistischen Test zum Vergleich von zwei medizinischen Behandlungen werden diskutiert.

Die graphische Darstellung des kritischen Bereichs für eine wachsende Anzahl von Spielen (8, 20, 50, 100, 1000) soll den Einfluss des Stichprobenumfangs auf die Entscheidungsregeln demonstrieren. Eine pragmatische Erklärung der Wahrscheinlichkeit für einen Fehler der zweiten Art und das Konzept der „Power“ eines Tests folgt mit dem Hinweis auf die entscheidende Rolle des Stichprobenumfangs (siehe auch Vorlesung 5). Schließlich wird am Beispiel des Vergleichs von zwei Stichproben ein bekannter statistischer Test, der Rangsummentest, vollständig hergeleitet.

4.5. Grundlagen der Versuchsplanung und Standards

Die Grundlagen der Versuchsplanung werden anhand internationaler Standards an Beispielen eingeführt:

- Studientypen [Level of Evidence Guidelines]
- Studienplan [Good Scientific Practice der Medizinischen Fakultät; Statistical Principles in Clinical Trials, ICH 9]
- Rationale Versuchsanordnung (Design)
- Kontrollgruppe [Choice of Control Group for Clinical Trials, ICH 10; Konvention von Helsinki, revidierte Version]
- Ein- und Ausschlusskriterien
- Vermeidung von „Bias“ - Randomisierung und Verblindung
- Primäre und sekundäre Zielkriterien
- Wahl des Stichprobenumfang
- Zwischenauswertungen
- Versuchsbegleitung [Good Clinical Practice, ICH 6]
- Fehlenden Werte

- Auswertungsstrategie
- Publikation [Consort Statement]

6. Wiederholung und Zusammenfassung

In dieser Einheit werden noch einmal die erlernten Inhalte in weitem Bogen zusammengefasst, wobei konkrete Ratschläge zur Vermeidung grundlegender Fehler bei der Umsetzung der Methoden in der Planung, Durchführung, Auswertung und Präsentation von Diplomarbeiten gegeben werden.