

Statistische Beratung am Zentrum für Klinische Studien (ZKS) Universitätsklinikum Regensburg

Karolina Müller und Florian Zeman

Allgemeine Hinweise

Die statistische Beratung wissenschaftlicher Arbeiten ist eine interne Serviceleistung des ZKS für Angehörige des Klinikums und der Medizinischen Fakultät der Universität Regensburg.

Für Interessierte wird eine Erstberatung durchgeführt. Diese findet nach vorheriger Anmeldung statt (siehe online Anmeldebogen). Eine weitergehende Beratung und eventuelle Auswertung kann nur durchgeführt werden, wenn Fragestellung, Studienplan, Studiendurchführung und Datenqualität angemessen und methodisch überzeugend sind.

Die statistische Beratung sollte bereits in der Planungsphase der Arbeit in Anspruch genommen werden, da Fehler bei der Planung erhobene Daten nutzlos machen können.

Die statistische Beratung einschließlich der Benutzung der EDV-Anlagen ist kostenlos, sofern es sich um die Betreuung einer rein wissenschaftlichen Arbeit handelt. Wird die zu beratende Arbeit von einer wissenschaftlichen Förderungsgesellschaft (z. B. DFG, BMBF) unterstützt, sollte dies bei der Anmeldung zum ersten Beratungsgespräch mitgeteilt werden. Die Unterstützung der Auswertung von pharmazeutischen Studien bedarf besonderer Absprachen.

Umfang der Beratung

„Schnellberatungen“ im Sinne einer Notfallambulanz – insbesondere telefonische Anfragen – können nicht erbracht werden, da fundierte Auskünfte ohne genaue Kenntnisse des Versuchsplans und der Daten unmöglich sind.

Das Angebot des ZKS Regensburg beinhaltet hauptsächlich die Hilfestellung bei der Planung, Durchführung und Auswertung („how to do“). Die notwendigen praktischen Arbeiten (z. B. Dateneingabe, Anwendung von Standard-Statistik-Software zur Datenauswertung etc.) sind jedoch selbstständig zu erbringen („learning by doing“).

Beratung von Promotionsarbeiten

Im Allgemeinen kann das ZKS aufgrund geringer Personalkapazitäten nicht die Beratung von Promotionsarbeiten als Serviceleistung übernehmen.

Primäres Ziel ist daher, die Arbeitsgruppenleiter von klinischen Studien in den einzelnen Kliniken bei ihrer wissenschaftlichen Tätigkeit zu unterstützen. Wenn im Zuge dessen eine Promotionsarbeit entsteht, so ist damit zumindest eine indirekte Promotionsberatung gegeben. In jedem Fall ist vor Beginn dieser Kooperation ein Beratungsgespräch zwischen Betreuer und Doktoranden und dem ZKS erforderlich. Die Betreuung der Promotionsarbeit liegt weiterhin beim Arbeitsgruppenleiter der Klinik als primären Ansprechpartner für das ZKS.

Voraussetzungen

Als Basis für eine effektive Beratung wird erwartet, dass zusätzlich zum Anmeldebogen die zu untersuchende Fragestellung und das Ziel der Untersuchung inklusive Operationalisierung (verwendete Variablen und Skalenniveau) in präziser Formulierung schriftlich vorliegen.

Die Beratung und Unterstützung bei der Auswertung setzt auf Seiten der Interessenten Grundkenntnisse in SPSS (Umfang: Kurs Statistische Datenanalyse mit SPSS, Rechenzentrum Universität Regensburg) und der statistischen Methodik (Umfang: Vorlesung Epidemiologie,) voraus. Weitere statistische Basiskenntnisse können mit Hilfe einführender Lehrbücher z. B. [1] erworben werden.

Datenerfassung

Zur Vorbereitung der Dateneingabe werden vom ZKS Hilfestellungen gegeben (siehe auch Anforderungen an die Daten). Die Dateneingabe selbst liegt in der Hand der Interessenten.

Die Verantwortung für die Daten (Korrektheit der Daten, Datensicherung, Datenschutz) liegt nicht beim ZKS. Ist eine Auswertung mit den am ZKS vorhandenen Softwareprogrammen vorgesehen, sind an Struktur und Format der Daten Minimalanforderungen zu stellen, die im Folgenden formuliert sind.

Anforderungen an das Programm zur Datenerfassung

Die häufigste Form der Datenerfassung sind Excel-Dateien. Davon raten wir ab! Da die meisten Auswertungen mit SPSS durchgeführt werden, ist es sinnvoll, bereits die Rohdaten mit SPSS zu erfassen.

Als Angestellter oder Student des UKR können Sie SPSS kostenlos auf Ihrem Rechner installieren und verwenden. Eine Anleitung zur Installation finden Sie hier: [SPSS am UKR](#)

Hinweise zum Datenschutz

Gemäß der Datenschutz-Grundverordnung (DSGVO) dürfen Datensätze, die im Rahmen der statistischen Beratung verwendet werden, **keine unmittelbar personenbezogenen Daten** enthalten.

Dazu gehören z.B.:

- Patientennamen / Namen von Therapeuten, Klinikmitarbeitern etc.
- vollständiges Geburtsdatum
- Anschriften
- Telefonnummern
- E-Mail-Adressen
- Versicherungsdaten
- sonstige direkt identifizierende Angaben.

Die Einhaltung der DSGVO ist in einer Dienstanweisung des UKR geregelt. Verstöße dagegen sind als Datenschutzpanne anzuzeigen.

Bitte verwenden Sie ausschließlich anonymisierte (einzelne Patienten sind nicht rückenschlüsselbar) oder pseudonymisierte (Patienten sind durch eindeutige Patientenidentifikationsnummern gekennzeichnet) Datensätze.

Sollte es notwendig sein, eine Schlüsseliste mit Patientenidentifikationsnummern und Patientennamen zu führen, so ist diese Schlüsseliste gesondert aufzubewahren.

Es ist nicht zulässig, Studiendaten auf einem ungesicherten USB-Stick ohne Passwortschutz mit sich zu führen!

Regeln für die Dateneingabe:

- 1 Patient bzw. 1 Beobachtungseinheit pro Zeile, auch bei Mehrfachmessungen
- 1 Variable pro Spalte
- 1 Variablenwert pro Zelle
- keine Leerspalten und -zeilen innerhalb der Tabelle
- keine farblichen Markierungen als Ersatz für Variablen
- wenn es mehrere Gruppen gibt (z.B. mehrere Behandlungen): nur eine SPSS-Datendatei anlegen, mit einer Spalte, in der die Gruppenzugehörigkeit steht (nicht pro Gruppe eine Datei)
- Regeln für die Einrichtung von Variablennamen: sinnvolle Abkürzungen verwenden; Variable muss mit einem Buchstaben beginnen; Variablenname darf außer Buchstaben nur die Ziffern 0 bis 9 und das Zeichen „_“ enthalten; weder Umlaute, Sonderzeichen, Leerzeichen noch mathematische Operatoren verwenden; Variablennamen nicht doppelt verwenden
- werden zu unterschiedlichen Messzeitpunkten gleiche Werte erhoben, so sollten diese mit gleicher Bezeichnung aber unterschiedlichen Suffixen angelegt werden (z.B. Blutdruck_base, Blutdruck_V1, Blutdruck_V2, Blutdruck_endofstudy)
- die Anzahl der Ausprägungen bei kategorialen Variablen sollte in einem sinnvollen Verhältnis zum Stichprobenumfang stehen
- alle Variablen (außer Freitexte) müssen numerisch angelegt sein. D.h. es darf kein Text oder Sonderzeichen dokumentiert werden (z.B. nicht: <0.01 bei Werten)
- Variablenkodierung: sinnvolle und einheitliche Wertelabels zuordnen (z. B. Ja/Nein Fragen mit 1=ja, 2=nein oder Geschlecht mit 1=männlich, 2=weiblich kodieren). Gleiche Ausprägungen sollten in gleicher Weise erfasst werden
- bei Mehrfachantworten bzgl. einer Fragestellung (z.B. Art der Komplikation) jede Antwortmöglichkeit als einzelne Variable anlegen und dann mit ja/nein kodieren (z.B. Komplikation_Blutverlust ja/nein, Komplikation_reop ja/nein, ...)
- Spalten, die sich durch Berechnungen aus anderen Spalten ergeben (z.B. Mittelwerte), sollten zunächst weggelassen werden. Diese können mit SPSS generiert werden
- Kalenderdaten nicht als Textfelder sondern als Datumsfelder definieren und eingeben; bei der Eingabe das Format TT.MM.JJJJ verwenden (z.B. 01.01.2006)
- Nachkommastellen mit Dezimalzeichen abtrennen
- bei fehlendem Wert Zelle leer lassen
- Freitext ist nicht unmittelbar auswertbar und sollte deshalb sinnvoll kodiert werden
- **KEINE** personenbezogenen Daten dokumentieren, insbesondere keine Personennamen und keine Geburtsdaten! Falls Zeitpunkt der Geburt für weitere Verrechnungen notwendig sein sollte, nur Monat und Jahr angeben, nicht den Tag der Geburt! (s. auch oben Hinweise zum Datenschutz)

Beispiel:

	patnr	gruppe	geschl	alter	groesse	gewicht	diagdat	dbd1	dbd2
1	1	1	1	20	170	76,5	25.12.2005	85	80
2	2	1	2	30	164	56,2	02.01.2006	70	75
3	3	2	1	40	180	84,0	14.05.2006	78	80

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels
1	patnr	Numerisch	8	0	Patient Nummer	Kein
2	gruppe	Numerisch	8	0	Gruppe	{1, Behandlung A}
3	geschl	Numerisch	8	0	Geschlecht	{1, männlich}...
4	alter	Numerisch	8	0	Alter	Kein
5	groesse	Numerisch	8	0	Größe in cm	Kein
6	gewicht	Numerisch	8	1	Gewicht in kg	Kein
7	diagdat	Datum	10	0	Diagnosedatum	Kein
8	dbd1	Numerisch	8	0	Diastolischer Blutdruck bei Visite 1	Kein
9	dbd2	Numerisch	8	0	Diastolischer Blutdruck bei Visite 2	Kein

Außerdem gilt:

- die Daten sollten unter einem inhaltlich sinnvollen Namen in einem Ordner abgespeichert werden, der auch mühelos wiederzufinden ist. Die Dateinamenerweiterung .sav fügt SPSS automatisch hinzu
- neben der eigentlichen Rohdatendatei ist eine vollständige Liste der Variablen und des jeweiligen Wertebereichs notwendig (sog. Datenbeschreibung)
- wenn standardisierte Fragebögen verwendet werden: Reihenfolge der Variablen beibehalten, keine Fragen oder Antwortmöglichkeiten ändern, Fragebogen-Scores mit Hilfe des Manuals/der Syntax berechnen

Datentransfer

Die einfachste Form des Datentransfers ist die Versendung als E-Mail-Attachment oder USB-Sticks und ucrbox sind ebenfalls möglich.

Wichtig:

- keine privaten E-Mail-Adressen verwenden (weder zum Datentransfer noch für die Anmeldung zur Beratung oder zur weiteren Kommunikation)
- die Daten sind mindestens pseudonymisiert
- die Datei ist Passwort geschützt. Daten und Passwort müssen getrennt voneinander übermittelt werden (z.B. Daten per E-Mail und Passwort per Telefon)

Datenauswertung

Die Nutzung privater Endgeräte zur Verarbeitung von Patientendaten ist generell verboten.

Die Verarbeitung von Patientendaten zu Zwecken von Forschung oder Lehre muss am UKR erfolgen, solange ein Personenbezug herstellbar ist. Da durch Pseudonymisierung im Unterschied zur Anonymisierung der Personenbezug (wenn auch für einen eingeschränkten Personenkreis) grundsätzlich erhalten bleibt, gilt dies auch für pseudonymisierte Daten.

Die Auswertung erfolgt nach Absprache mit dem ZKS.

Die Datenerhebung ist abgeschlossen und die Daten sind vollständig erfasst. D.h. alle notwendigen Variablen wurden soweit möglich vollständig erfasst, es kommen keine weiteren Patienten mehr dazu und es wurde bereits alles Mögliche getan um fehlende Werte zu füllen.

Eine deskriptive Analyse der Daten wurde bereits durchgeführt.

Die Auswertung im ZKS wird in der Regel mit Hilfe der Statistik-Software SPSS durchgeführt. Nach Absprache mit dem ZKS können weitere Statistikpakete (z.B. R und SAS) notwendig und in Folge benutzt werden. Von der Auswertung mit Excel wird grundsätzlich abgeraten.

Es werden nicht sämtliche möglichen Analyseschritte vorgenommen, sondern eine mit dem Beratungssuchenden vor Durchführung der statistischen Auswertung festgelegte, begrenzte Anzahl von Fragen wird gezielt mit geeigneten statistischen Methoden bearbeitet.

Abschlussarbeit und Publikation

Eine Veröffentlichung der Ergebnisse sollte grundsätzlich angestrebt werden.

Die Nennung des ZKS bei Abschlussarbeiten oder Veröffentlichungen der Ergebnisse bedarf es einer ausdrücklichen Zustimmung des Leiters des ZKS oder des betreuenden Mitarbeiters. Wenn in einer zu publizierenden Arbeit, in einem Vortrag oder auf einem Poster das ZKS erwähnt wird, muss das entsprechende Dokument vor der Einreichung rechtzeitig dem ZKS vorgelegt werden.

Werden im Rahmen der Beratung durch das ZKS Dienstleistungen (z.B. die Auswertung) erbracht, die zu wissenschaftlichen Publikationen führen, entspricht es wissenschaftlich korrektem Verhalten, die beteiligten Mitarbeiter des ZKS als Koautoren zu berücksichtigen. In diesem Fall ist es erforderlich, die Arbeit vor Einreichung zur Publikation den verantwortlichen ZKS Mitarbeitern zum Gegenlesen vorzulegen und alle Beteiligten über den Review-Prozess zu informieren.

Weiterführende Literatur (ohne Anspruch auf Vollständigkeit)

1. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1990.
2. Bühl A. SPSS 23: Einführung in die moderne Datenanalyse, 15. Auflage, 2016.
3. Brosius F. SPSS 24 für Dummies, Wiley-VCH Verlag, 2017.
4. Field A. Discovering Statistics Using IBM SPSS Statistics, 5. Auflage, SAGE Publications, 2017.
5. Fromm S. Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten, 2. Auflage, 2012.

Allgemeine SPSS Hilfen

SPSS bietet zahlreiche Hilfestellungen an und es sind eine Vielzahl von Handbüchern sowie begleitende Literatur zur Unterstützung bei der Anwendung von SPSS vorhanden: z.B.

- „IBM SPSS Statistics 25 Core-System-Benutzerhandbuch“ (grundlegende Informationen)
- ergänzende Beschreibung zu verschiedenen Modulen (z.B. „Advanced Statistics“, „Categories“)
- „IBM SPSS Statistics 25 Command Syntax Reference“: Referenzwerk zur Kommandosprache von SPSS

Dateneingabe

Hierzu gibt es sehr viele gute Anleitungen im Internet, z.B.

- [IBM SPSS Kurzanleitung](#)
- Diverse Tutorien von Universitäten, z.B. SPSS [Tutorial der FU Berlin](#)
- Diverse You-Tube Videos, z.B. [SPSS Tutorial 1](#)

Datenimport aus Excel in SPSS

Es gibt sehr viele gute Anleitungen im Internet, z.B. [Video](#)

Datenauswertung

Auch hier gibt es viele Anleitungen im Internet zu finden, z.B.

- Deskriptive Statistiken: [Video](#)
- t-Test für unabhängige Stichproben: [Video](#)
- Chi-Quadrat Test: [Video](#)

Zwei sehr gute Seiten sind auch die [Methodenberatung der Universität Zürich](#) und [StatistikGuru](#). Hier werden die Tests anhand von SPSS erklärt.