

# **Statistische Beratung am Zentrum für Klinische Studien (ZKS) Klinikum der Universität Regensburg**

(von Justine Rochon)

## **Allgemeine Hinweise**

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

Für Interessierte wird eine Erstberatung durchgeführt. Diese findet nach vorheriger Anmeldung statt (siehe 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 häufig zu nicht aussagekräftigen Ergebnissen führen und erhobene Daten nutzlos machen.

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 („keine Therapie ohne Diagnose“).

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

Umfang und Intensität der Beratung richten sich nach der Komplexität der erforderlichen Methoden. Nach Abschluss der Arbeit muss die Dauer der Aufbewahrungszeit der Daten und der Auswertungsprogramme in Abstimmung mit dem ZKS festgelegt werden.

## **Voraussetzungen**

Als Basis für eine effektive Beratung wird erwartet, dass zusätzlich zum Anmeldebogen die zu untersuchende Fragestellung und das Ziel der Untersuchung in präziser Formulierung schriftlich vorliegen. Diese kurze Beschreibung sollte die Form eines Abstracts haben.

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, Medizinische Biometrie und Medizinische Informatik) 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.

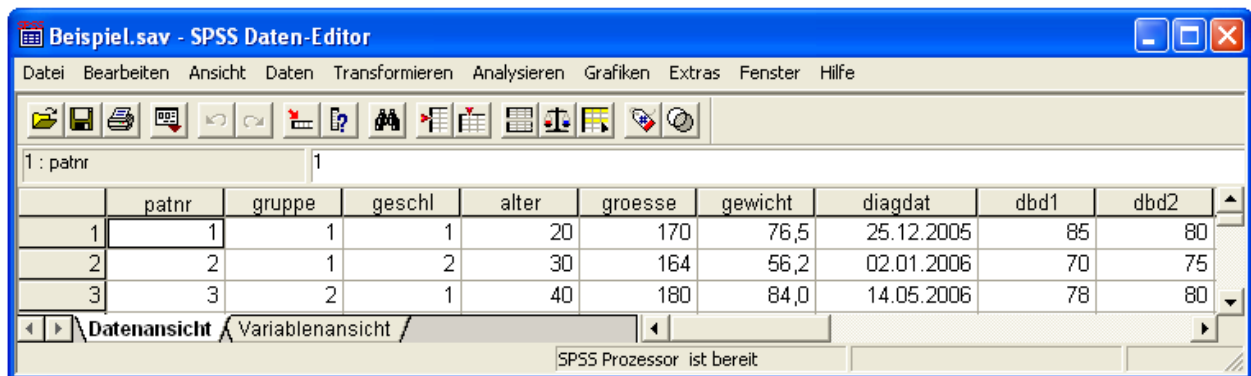
## **Anforderungen an die Daten**

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.

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
- wenn es mehrere Gruppen (z. B. mehrere Behandlungen) gibt: nur eine SPSS-Datendatei anlegen, mit einer Spalte, in der die Gruppenzugehörigkeit steht (nicht: pro Gruppe eine Tabelle)
- Regeln für die Einrichtung von Variablennamen: nicht mehr als 8 Zeichen; 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
- alle Variablen sollten sinnvoll und eindeutig gelabelt sein
- die Anzahl der Ausprägungen bei kategoriellen Variablen sollte in einem sinnvollen Verhältnis zum Stichprobenumfang stehen
- Variablenkodierung: sinnvolle und einheitliche Wertelabels zuordnen (z. B. Ja/Nein Fragen mit 1=ja, 2=nein oder Geschlecht mit 1=männlich, 2=weiblich kodieren). Eine numerische Kodierung ist einem Eintrag, der Buchstaben und/oder Sonderzeichen enthält, grundsätzlich vorzuziehen. Gleiche Ausprägungen sollten in gleicher Weise erfasst werden
- 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

Beispiel:



The screenshot shows the SPSS Daten-Editor window with the 'Datenansicht' (Data View) selected. The data is organized into a grid with columns for variables and rows for individual patients. The variables are: patnr, gruppe, geschl, alter, groesse, gewicht, diagdat, dbd1, and dbd2. The data points are as follows:

	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



The screenshot shows the SPSS Daten-Editor window with the 'Variablenansicht' (Variable View) selected. This view provides a detailed description of each variable, including its name, type, format, and labels.

	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)
- Personendaten müssen grundsätzlich anonymisiert werden, d.h. keine Patientennamen, sondern eindeutige Patientenidentifikationsnummern (Patient ID bzw. Patient Nummer) verwenden.

## Datentransfer

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

## **Datenauswertung**

Die Auswertung erfolgt nach Absprache mit dem ZKS.

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) 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.

## **Publikation**

Eine Veröffentlichung der Ergebnisse sollte grundsätzlich angestrebt werden.

Die Nennung des ZKS bei Veröffentlichungen der Ergebnisse bedarf dessen ausdrücklicher Zustimmung. 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.

## **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.

## **Referenzen**

1. Altman DG. *Practical Statistics for Medical Research*. London: Chapman & Hall, 1991.