

## Fallzahlplanung mit der Statistiksoftware SAS: Möglichkeiten und Limitierungen

|                          |                         |
|--------------------------|-------------------------|
| Katrin Häußler           | Rainer Mucbe            |
| Hochschule Ulm           | Institut für Biometrie  |
| Albert-Einstein-Allee 55 | Schwabstraße 13         |
| 89081 Ulm                | 89075 Ulm               |
| kahaeuss@mail.hs-ulm.de  | rainer.mucbe@uni-ulm.de |

### Zusammenfassung

Zu den Neuerungen in SAS Version 9.1 gehören die beiden Prozeduren PROC POWER und PROC GLMPOWER zur Fallzahlschätzung und Powerberechnung. PROC POWER beinhaltet die geläufigsten statistischen Testverfahren, wohingegen mit PROC GLMPOWER Fallzahlschätzungen für die mehrfaktorielle Varianzanalyse durchgeführt werden können. Aufbauend auf beiden Prozeduren ist darüber hinaus die Benutzeroberfläche SAS Power and Sample Size Application (PSS) zu nennen, deren Aufruf über einen Webbrowser getätigt wird. Sie eignet sich zur menügesteuerten Fallzahlschätzung ohne Programmierkenntnisse. Zusätzlich ist auf eine weitere Benutzeroberfläche in SAS, den SAS-Analyst, hinzuweisen, mit dem ein kleiner Umfang an Fallzahlschätzungen durchgeführt werden kann.

Eine Ergänzung der Möglichkeiten der Prozeduren PROC POWER bzw. PROC GLMPOWER ist durch Beispielprogramme des Statistiklehrbuches „Bestimmung des Stichprobenumfangs“ gegeben. Zusätzlich findet man bei einer Recherche im Internet eine Vielzahl an SAS-Makros zur Bestimmung von Fallzahlen. Allerdings besteht bei den Beispielprogrammen und Makros das Problem, dass diese nicht oder fraglich validiert sind und daher nicht ohne weiteres zur Fallzahlschätzung für klinische Studien nach den GCP/ICH-Richtlinien verwendet werden können.

Zu Demonstrationszwecken werden die Fallzahlschätzungen für eine sich in der Planung befindliche Therapiestudie anhand einer stetigen Zielgröße mit dem Zweistichproben-Test durchgeführt. Insgesamt wird die Vielzahl der Möglichkeiten zur Fallzahlplanung und Powerberechnung und deren Erweiterung mit SAS aufgezeigt.

**Schlüsselwörter:** Fallzahl- und Powerschätzung, PROC POWER, PROC GLMPOWER, PSS

## 1 Einleitung

Zur Überprüfung der Wirksamkeit, Sicherheit und Verträglichkeit von Medikamenten sowie zur Erforschung von Erkrankungen bisher unbekannter Ursache wird Jahr für Jahr weltweit eine große Anzahl von Studien durchgeführt. In der Planungsphase von Studien ist der Fallzahlschätzung eine große Bedeutung beizumessen. Der Stichprobenumfang sollte so frühzeitig wie möglich geschätzt werden, damit es beim Ablauf einer Studie nicht zu Problemen kommt. Soll z.B. ein kleiner Effekt in der Therapie einer sehr seltenen Erkrankung überprüft werden, würde dafür eine sehr große Fallzahl benötigt

werden. Die Rekrutierung einer großen Anzahl von Patienten ist aber bei sehr seltenen Erkrankungen nicht immer möglich.

Auch kann es durchaus immer wieder vorkommen, dass ein Sponsor eine Studie durchführen möchte, aber seine finanziellen Mittel dafür auf Grund der großen benötigten Fallzahl nicht ausreichen. Es wäre fatal, wenn die Fallzahl nicht im Vorfeld berechnet würde und erst bei der Auswertung der Studie auffiele, dass die Fallzahl nicht ausreichend war, die Power somit zu gering ist und eventuell tatsächlich bestehende Effekte im statistischen Test nicht nachweisbar sind. Somit bliebe die Investition von Zeit, Geld und Ressourcen des Sponsors in die Studie ohne Ergebnis.

Die Fallzahl sollte hoch genug gewählt werden, um eine verlässliche Antwort auf die vor Beginn der Studie formulierten Fragestellungen zu erhalten. Dennoch darf die Fallzahl nicht beliebig hoch gewählt werden, damit Ressourcen wie Zeit, Personal und Kapital nicht verschwendet werden [11] und ethische Grundsätze nicht verletzt werden. So dürfen bei Tierversuchen nicht unnötig viele Tiere zum Teil erheblichen Qualen ausgesetzt werden.

Sollte zur Durchführung einer Studie eine enorm große Anzahl an Patienten benötigt werden, deren Rekrutierung einen langen Zeitraum in Anspruch nähme, bietet sich die Beteiligung mehrerer Prüfzentren an der Studie an.

Ein zu geringer Umfang könnte zur fälschlichen Verwerfung einer Substanz führen und damit Therapiechancen vergeben.

Die Größe der Fallzahl wird letztendlich im Wesentlichen von vier Faktoren beeinflusst [11]:

- Bei Vergrößerung des zu untersuchenden Effektes sinkt die Fallzahl
- Die Fallzahl wird bei steigender Varianz  $\sigma^2$  größer
- Mit genauerer Irrtumswahrscheinlichkeit  $\alpha$  vergrößert sich die Fallzahl
- Eine höhere Genauigkeit ( $1-\beta$ ) führt zu einem größeren Stichprobenumfang

In der Statistiksoftware SAS Version 9.1 wurden erstmals die beiden Prozeduren PROC POWER [13] und PROC GLMPOWER [12] zur Fallzahlplanung und Powerberechnung eingeführt. Gleichzeitig wurde die Benutzeroberfläche SAS Power and Sample Size Application [4] mitgeliefert, die auf den beiden oben genannten Prozeduren aufbaut und mit der ohne Programmieraufwand menügesteuert die Bestimmung des Stichprobenumfangs erfolgen kann. Diese Applikation kann über einen Webbrowser aufgerufen werden. Zusätzlich werden die Fallzahlschätzungen in SAS-Analyst, einer standardmäßig seit Version 8 in SAS enthaltenen Benutzeroberfläche, berücksichtigt und der SAS Enterprise Guide bezüglich seiner Möglichkeiten zur Fallzahlplanung untersucht. Die in PROC POWER und PROC GLMPOWER enthaltenen statistischen Prozeduren sollen durch dem Statistiklehrbuch zur Fallzahlplanung von Bock [2] auf Diskette beigelegte SAS-Beispielprogramme sowie durch zum kostenlosen Download im Internet bereitgestellte Makros zur Fallzahlplanung ergänzt werden.

Zur Durchführung der Fallzahlplanung wird zur Anschauung ein Beispiel aus einer sich in der Planung befindlichen Studie herangezogen. Die Fallzahlplanung für die Studie

wird mit dem Zweistichproben-t-Test durchgeführt. Die Irrtumswahrscheinlichkeit ist nicht, wie üblicherweise, mit  $\alpha = 0,05$  gegeben, sondern wird aus Gründen, deren Erläuterung hier nicht relevant ist, auf  $\alpha = 0,03$  festgelegt. Es wird eine Power von 80% vorausgesetzt. Für den Zweistichproben-t-Test kann man bei der Kontrollgruppe von einem Verlust der Variablen X von 0,8 Einheiten ausgehen, wohingegen bei der Therapiegruppe von einem Stillstand der Progredienz auszugehen ist, d.h. von einem Verlust in X von 0 Einheiten. Das heißt, es wird die Mittelwertdifferenz  $\Delta = 0,8$  Einheiten von X zwischen Therapie- und Kontrollgruppe als Zielgröße für den Zweistichproben-t-Test festgelegt. Es ist ungefähr von einer mittleren Standardabweichung von  $\sigma = 0,8$  Einheiten auszugehen. Die Fallzahl für den Zweistichproben-t-Test wird mit Hilfe von Formel 1.1 nach Bock [2] berechnet als:

$$N \approx \frac{[t_{1-\frac{\alpha}{2},df} + t_{1-\beta,df}]^2}{c^2}, \quad \text{wobei} \quad c = \frac{\Delta}{2\sigma}.$$

Für den Zweistichproben-t-Test sind  $n_1 + n_2 - 2$  Freiheitsgrade zu berücksichtigen, d.h.,  $df = N - 2$ . Der Nichtzentralitätsparameter  $c$  berechnet sich aus dem Quotienten der Mittelwertdifferenz  $\Delta$  und der Standardabweichung  $\sigma \cdot 2$ . Die t-Quantile können Tabellen, z.B. aus Statistiklehrbüchern, entnommen werden. Hierbei ist der für den Nichtzentralitätsparameter  $c$  berechnete Wert unter  $N - 2$  und einer zweiseitigen Fragestellung abzulesen, vgl. Tabelle T5 bei Bock [2].

## 2 Die SAS Power and Sample Size Application (PSS)

Zu den Neuerungen in SAS Version 9.1 gehören die beiden Prozeduren PROC POWER und PROC GLMPOWER sowie die Benutzeroberfläche SAS Power and Sample Size Application (PSS). Die PSS baut auf diesen beiden Prozeduren auf. Alle Funktionen, die unter Verwendung der Benutzeroberfläche möglich sind, lassen sich auch direkt mit der Statistiksoftware SAS mittels genannter Prozeduren durchführen. Der Vorteil der PSS ist, dass sie von Personen, die über keine umfangreichen Programmierkenntnisse der Statistiksoftware SAS verfügen, genutzt werden kann.

Die SAS Power and Sample Size Application ist eine Web-Applikation. Das bedeutet, dass nur eine Webserver-Software, z.B. PSS Tomcat 4.1 Server, sie zum Laufen bringt. Die Ergebnisdarstellung ist als Tabelle und/oder Graphik und/oder erklärender Text, so genannte Narratives, möglich. Alle durchgeführten Analysen kann man zu einem späteren Zeitpunkt nochmals unter *Projects* betrachten.

Es ist auch möglich, den SAS-Code und das Log-File einzusehen und den SAS-Code zum selbstständigen Programmieren mit PROC POWER bzw. GLMPOWER in SAS zu übertragen.

Wenn die Verbindung zum Server aufgebaut ist (über Programme  $\rightarrow$  SAS  $\rightarrow$  Start Tomcat), kann der Link <http://localhost:8080/pss/Power> über den Microsoft Internet Explorer aufgerufen und somit die PSS gestartet werden.

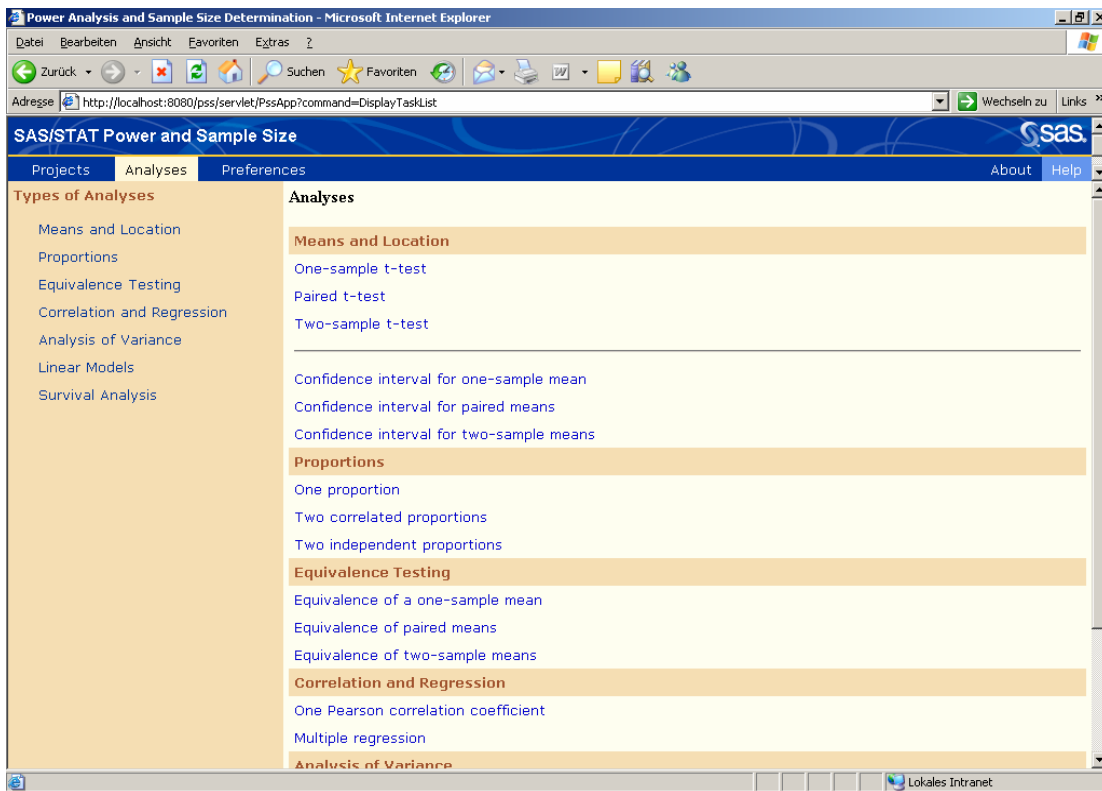


Abbildung 1: Die Auswahlmaske in der SAS PSS - Oberfläche

Die Vorteile der SAS Power and Sample Size Application liegen auf der Hand. Es wird Personen mit ausreichenden Statistik-, jedoch ohne Programmierkenntnisse, möglich, Fallzahl- und Powerberechnungen mit geringerem Aufwand durchzuführen. Da der für die PSS benötigte Server Tomcat 1.4 standardmäßig auf der Midtier-Components-CD enthalten ist und dessen Installation keine große Schwierigkeit darstellt, empfiehlt es sich für jeden SAS-Nutzer, der gelegentlich oder häufig Fallzahlplanungen durchführt, die PSS zu installieren. Allerdings ist zur Erzeugung von übersichtlichen, schönen Graphiken auf PROC POWER bzw. GLMPOWER zurückzugreifen. Auch zur automatischen Erzeugung von SAS-Datensätzen mittels ODS muss direkt programmiert werden.

## 2.1 Der Zweistichproben-t-Test in der PSS

Bei der Studie ist von einer mittleren Differenz der Zielgröße von  $\Delta = 0,8$  Einheiten zwischen der Behandlungs- und der Kontrollgruppe auszugehen. Die Standardabweichung wird als  $\sigma = 0,8$  Einheiten geschätzt und als Irrtumswahrscheinlichkeit ist  $\alpha = 0,03$  vorgegeben. Es wird ein zweiseitiger t-Test verwendet (unter Means and Location  $\rightarrow$  Two sample t-Test) und von Normalverteilung ausgegangen. Die erwartete Differenz der Mittelwerte beträgt 0,8 Einheiten und als Standardabweichung wird 0,8 geschätzt. Diese Werte sind in Abbildung 2 einzutragen. Es wird eine Power von 80% vorausgesetzt, die in Abbildung 3 einzutragen ist. Die Auswahl der Ergebnisdarstellung erfolgt auf drei verschiedene Arten: als Ergebnistabelle, als Graphik und als Narrative. Mit der PSS wird, wie in Abbildung 4 ersichtlich, eine Fallzahl von jeweils 20 Patienten pro Gruppe berechnet.

Abbildung 2: Auswahl der Eingabeparameter des Zweistichproben-t-Tests

Abbildung 3: Power und Ergebnisdarstellung

Two-sample t-test

Two-sample t Test for Mean Difference

| Fixed Scenario Elements |        |
|-------------------------|--------|
| Distribution            | Normal |
| Method                  | Exact  |
| Number of Sides         | 2      |
| Alpha                   | 0.03   |
| Mean Difference         | 0.8    |
| Standard Deviation      | 0.8    |
| Nominal Power           | 0.8    |
| Null Difference         | 0      |

| Computed N Per Group |             |
|----------------------|-------------|
| Actual Power         | N Per Group |
| 0.814                | 20          |

Abbildung 4: Summary Table

### 3 Die SAS-Prozeduren PROC POWER und PROC GLMPOWER

Mit der SAS-Version 9.1 wurden die beiden Prozeduren PROC POWER und PROC GLMPOWER zur Fallzahlplanung und Powerberechnung eingeführt. PROC POWER beinhaltet Fallzahlschätzungen für eine Vielzahl an statistischen Methoden und Testverfahren, wohingegen PROC GLMPOWER speziell für die Fallzahlplanung der mehrfaktoriellen Varianzanalyse geeignet ist. Die Abkürzung GLM in GLMPOWER steht für General Linear Models. Durch eine Trennung der geläufigsten statistischen Testverfahren und der mehrfaktoriellen ANOVA wurde die Übersicht der beiden Prozeduren erhalten, da die mehrfaktorielle ANOVA das aufwendigste Testverfahren ist, wofür sich eine eigene Prozedur anbietet. Alle anderen enthaltenen Fallzahlschätzungen wurden in PROC POWER zusammengefasst.

### 3.1 Die Prozedur PROC POWER

Die Prozedur PROC POWER beinhaltet eine Vielzahl an Fallzahlschätzungen für die unterschiedlichsten statistischen Testverfahren wie Mittelwertvergleiche, Überlebenszeitanalyse, Vergleich von Wahrscheinlichkeiten und Raten, Äquivalenztests und Korrelation und Regression. Auch die einfaktorische Varianzanalyse ist in PROC POWER enthalten [13]. Für die mehrfaktorische ANOVA wird auf PROC GLMPOWER verwiesen.

Die Syntax von PROC POWER soll hier vorgestellt werden. Für jede durchzuführende statistische Analyse wird ein spezielles Statement benötigt, dem Parameter mitgegeben werden können bzw. müssen.

```
PROC POWER <options>;
  MULTREG <options>;
  ONECORR <options>;
  ONESAMPLEFREQ <options>;
  ONESAMPLEMEANS <options>;
  ONEWAYANOVA <options>;
  PAIREFREQ <options>;
  PAIREDMEANS <options>;
  TWOSAMPLEFREQ <options>;
  TWOSAMPLEMEANS <options>;
  TWOSAMPLESURVIVAL <options>;
PLOT <plot-options> </graph-options>;
```

Nach der Auswahl eines Statements zur Fallzahlschätzung, z.B. TWOSAMPLEMEANS für den Zweistichproben-t-Test, wird dort für den zu berechnenden Parameter, entweder für die Gesamtfallzahl unter NTOTAL=, die Fallzahl pro Gruppe unter NPERGROUP= oder die Power unter POWER= ein fehlender Wert, d.h. ein Punkt, eingesetzt. Bei Auflösung nach der Fallzahl lautet die Syntax folglich NTOTAL=. oder NPERGROUP=. , die Berechnung der Power wird ermöglicht durch POWER=.

Graphiken können optional zusätzlich mit dem PLOT-Statement erstellt werden. Die ausschließliche Erzeugung einer Graphik funktioniert mit dem PLOTONLY-Statement als Option von PROC POWER.

Die Eingabe der Parameter ALPHA= (Defaultwert: 0,05) und SIDES= (Defaultwert: 2) ist nur erforderlich, falls ein anderer Wert als der Defaultwert benötigt wird. Für den Parameter SIDES= können vier verschiedene Werte eingetragen werden: Entweder eine 1 für einseitige Fragestellungen, eine 2 für zweiseitige Fragestellungen, ein U für den einseitigen Fall, dass der Wert der Alternativhypothese größer als der der Nullhypothese ist oder ein L für den Fall, dass der Wert der Alternativhypothese kleiner als der der Nullhypothese ist.

### 3.2 Fallzahlschätzung mit PROC POWER für die Beispielstudie

```

PROC POWER;
  TWOSAMPLEMEANS
  ALPHA = 0.03
  GROUPMEANS = (0 0.8)
  STDDEV = 0.8
  POWER = 0.8
  NPERGROUP = .;
  PLOT Y=POWER MIN=0.60 MAX=0.99;
RUN;

```

Die Irrtumswahrscheinlichkeit ist unter ALPHA= einzutragen. Die Mittelwerte in beiden Gruppen, bei der Studie 0 Einheiten in der Therapiegruppe und 0,8 Einheiten in der Kontrollgruppe, werden unter GROUPMEANS, eingeschlossen in eine runde Klammer, eingetragen. Die Standardabweichungen sind dem Statement STDDEV= zu übergeben. Es werden standardmäßig Gruppengrößen für ein balanciertes Design berechnet. Die Power ist unter POWER= einzutragen. Da für die Studie die Fallzahl zu berechnen ist, wird ein fehlender Wert, d.h. ein Punkt, für NPERGROUP= eingesetzt.

Äquivalent zum Ergebnis der PSS werden mit PROC POWER 20 Patienten pro Gruppe berechnet, d.h., insgesamt 40 Patienten, wie Abbildung 5 zeigt. In Abbildung 6 ist die mit Hilfe des PLOT-Statements erstellte Graphik zu sehen.

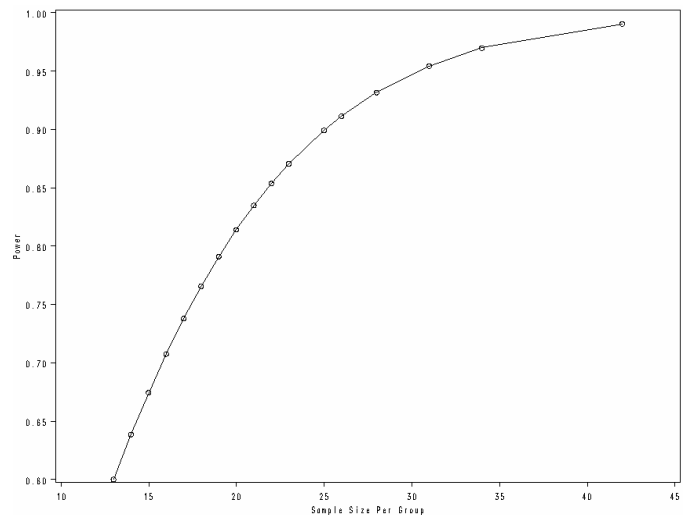
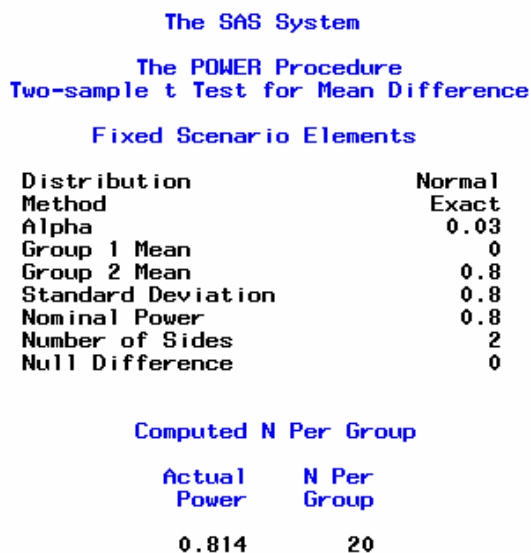


Abbildung 5: Output von PROC POWER

Abbildung 6: Graphik mit PROC POWER

**Tabelle 1:** Statements in PROC POWER (aus [5] )

| Themengebiet                        | Beschreibung  | Statement         |
|-------------------------------------|---|-------------------|
| Mittelwerte und Konfidenzintervalle | Einstichproben-t-Test<br>Konfidenzintervalle  | ONESAMPLEMEANS    |
|                                     | Zweistichproben-t-Test<br>Konfidenzintervalle   | TWOSAMPLEMEANS    |
|                                     | Gepaarter t-Test<br>Konfidenzintervalle   | PAIREDMEANS       |
| Varianzanalyse (ANOVA)              | Einfaktorielle Varianzanalyse   | ONEWAYANOVA       |
| Überlebenszeitanalyse               | Log-rank-Test<br>Gehan-Test<br>Tarone-Ware-Test   | TWOSAMPLESURVIVAL |
| Wahrscheinlichkeiten / Raten        | Einstichproben-Binomialtest<br>Exakter Gauß-Test<br>Exakter Gauß-Test mit Stetigkeitskorrektur<br>Approx. Gauß-Test<br>Approx. Gauß-Test mit Stetigkeitskorrektur | ONESAMPLEFREQ     |
|                                     | Exakter und approximativer Mc Nemar-Test  | PAIREFREQ         |
|                                     | Exakter Test von Fisher<br>$\chi^2$ -Test von Pearson<br>Likelihood-Ratio- $\chi^2$ -Test   | TWOSAMPLEFREQ     |
|                                     | Einstichproben-Vergleiche   | ONESAMPLEMEANS    |
|                                     | Zweistichproben-Vergleiche  | TWOSAMPLEMEANS    |
|                                     | Gepaarte Stichproben  | PAIREDMEANS       |
| Korrelation und Regression          | Korrelationskoeffizient von Pearson   | ONECORR           |
|                                     | Multiple lineare Regression   | MULTREG           |

### 3.3 Die Prozedur PROC GLMPOWER

Die Prozedur PROC GLMPOWER dient prospektiven Power- und Fallzahlberechnungen für univariate lineare Modelle. Die multivariate ANOVA wird noch nicht unterstützt. Die statistischen Analyseverfahren beinhalten Typ III-Tests und Kontraste in univariaten linearen Modellen. Optional können stetige oder diskrete Kovariablen berücksichtigt werden. Die Syntax wird im Folgenden vorgestellt [12].

```
PROC GLMPOWER <options>;
  CLASS variables;
  MODEL dependent-variables = effects;
  WEIGHT variable;
  CONTRAST 'label' effect values <effect values> </options>;
  POWER <options>;
PLOT <plot-options> </graph-options>;
```

Zur Berechnung der Fallzahl bzw. der Power wird innerhalb des POWER-Statements der Option NTOTAL bzw. POWER ein fehlender Wert zugewiesen, d.h. NTOTAL= . oder POWER= . .



### 3.4 In PROC POWER und PROC GLMPOWER enthaltene zusätzliche Optionen

PROC POWER und PROC GLMPOWER enthalten folgende drei zusätzliche Optionen, deren Funktionen nicht in der PSS enthalten sind:

- OUTPUTORDER
- NFRACTIONAL
- Automatische Erzeugung von SAS-Dateien mit dem Output Delivery System

Die Verwendung der Option OUTPUTORDER ist bei jedem Statement möglich. OUTPUTORDER hat die möglichen Ausprägungen *internal*, *syntax*, und *reverse*. *Internal* bezeichnet den Defaultwert und steht für die von SAS festgelegte Ausgabereihenfolge der Ergebnisparameter. *Syntax* gibt die Ergebnisse in der vom Programmierer festgelegten Reihenfolge aus und *reverse* in der zu *internal* umgekehrten Reihenfolge. In der Benutzeroberfläche PSS dagegen gibt es keine Möglichkeit, die Ausgabereihenfolge der Parameter selbst zu bestimmen.

Die Option NFRACTIONAL, die jeder Analyse mitgegeben werden kann, gibt nicht nur die aufgerundete Fallzahl aus, sondern auch die ursprünglich berechnete mit sechs Nachkommastellen.

Mit dem Output Delivery System (ODS) können Ergebnisse automatisch in temporäre oder permanente SAS-Dateien geschrieben werden. Diese Option besteht in der PSS nicht explizit, doch alle Ergebnisse werden dort unter *Projects* gespeichert.

Drei Schlüsselwörter spielen in ODS eine wichtige Rolle:

- Mit FixedElements kann die Fixed Scenario Elements-Tabelle abgespeichert werden
- Output dient dem Speichern von Ergebnisparametern
- PlotContent wird zum Speichern von Daten, die in Graphiken enthalten sind, benötigt. Hierbei wird nicht die Graphik selbst abgespeichert, sondern alle Daten, die erforderlich sind, um sie zu erzeugen

### 3.5 Erstellung von professionellen Graphiken für Publikationen

Durch die vary-Option des Plot-Statements wird für jede Irrtumswahrscheinlichkeit eine Graphik in einer eigenen Farbe und für jede Standardabweichung eine Kurve mit einem eigenen Symbol gekennzeichnet. Diese Möglichkeiten bestehen in der PSS nicht. Graphiken können dort nicht farbig erzeugt werden. Will man die Graphik so erzeugen, dass die Fallzahl auf der x-Achse und die Power auf der y-Achse abgebildet wird, ist dies zwar möglich, führt aber zu folgender Fehlermeldung:

```
ERROR: The argument axis is the same as the result.
```

SAS wählt zur Darstellung der Ergebnisvariablen standardmäßig die y-Achse aus, was allerdings nicht zwangsläufig Sinn macht. In der PSS können Titel für Graphiken erzeugt werden, was mit PROC POWER nicht möglich ist. Die PSS nutzt hierfür ein vorgefertigtes ODS-Statement. Der Titel wird dadurch nicht direkt in die Graphik eingefügt, sondern über die Graphik geschrieben.

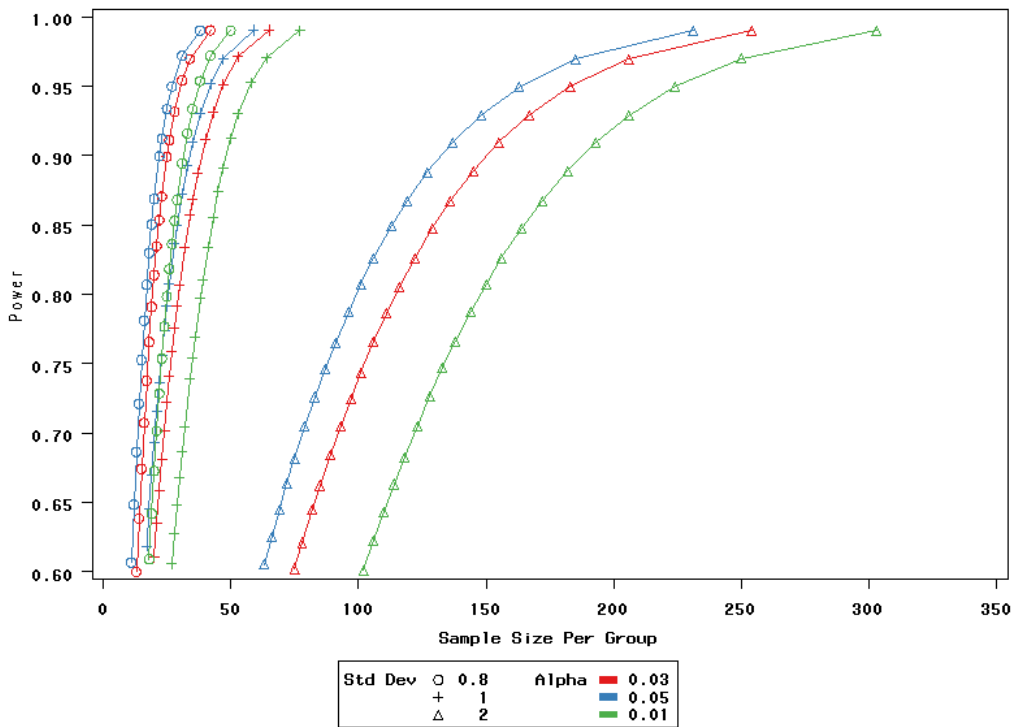


Abbildung 7: Graphik mit PROC POWER

Fallzahlplanung für die MigAN-Studie (t-Test)

Two-sample t Test for Mean Difference

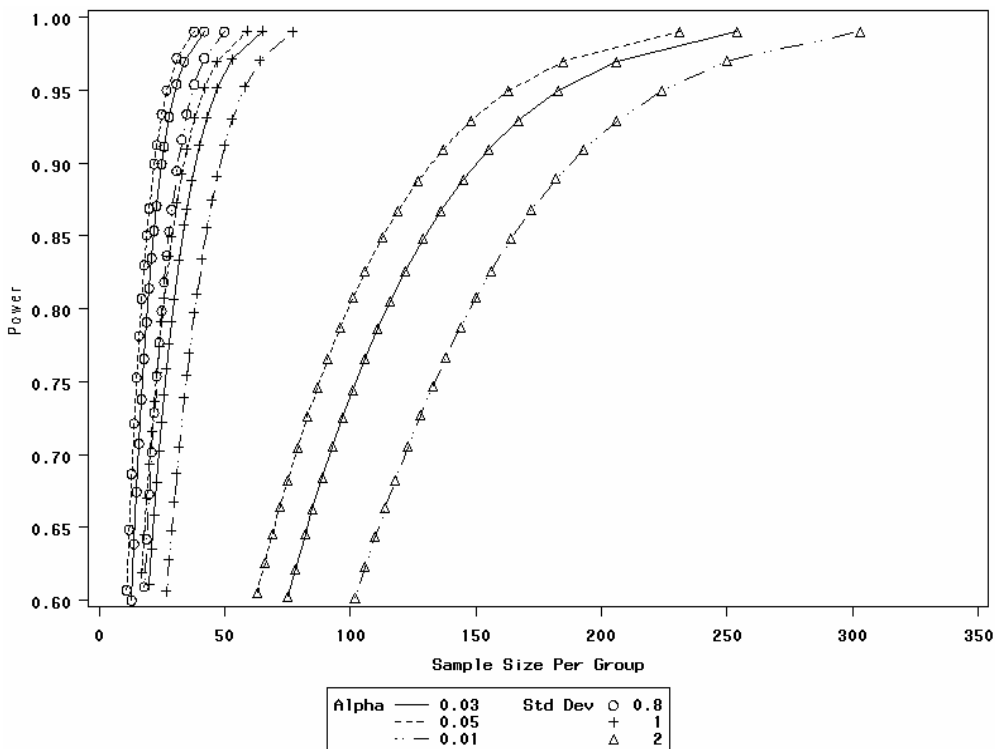


Abbildung 8: Graphik mit der PSS

## 4 Fallzahlplanung mit der Benutzeroberfläche SAS-Analyst

Der seit Version 8 standardmäßig in SAS enthaltene SAS-Analyst bringt den Vorteil mit sich, dass er nicht extra installiert werden muss, sondern ohne zusätzlichen Aufwand direkt genutzt werden kann. Da es sich um eine Benutzeroberfläche handelt, kann auch in diesem Fall ohne Programmieraufwand die Berechnung der Power und der Fallzahl durchgeführt werden. Er ist im Menü der englischen SAS-Version über den Pfad **Solutions** → **Analysis** → **Analyst**, in der deutschen Version über **Lösungen** → **Analyse** → **Analyst** [8] aufzurufen.

Die zweite in diesem Abschnitt zu nennende Benutzeroberfläche in SAS, der Enterprise Guide 4.1, wird auch auf ihre Möglichkeiten bezüglich der Fallzahlplanung hin untersucht. Laut Slaughter und Delwiche [14] und Ortseifen [10] sind aber im Enterprise Guide keinerlei Fallzahlschätzungen oder Powerberechnungen enthalten. Daher muss für Fallzahlschätzungen auf den SAS-Analyst oder die SAS PSS zurückgegriffen werden.

### 4.1 Bestimmung der Fallzahl mit dem SAS-Analyst

Die im SAS-Analyst enthaltenen Fallzahlschätzungen sind bei weitem nicht so umfangreich wie diejenigen aller anderen bereits vorgestellten und noch im folgenden erläuterten Möglichkeiten, aber eine Betrachtung ist wegen der Vollständigkeit der Möglichkeiten in SAS und auf Grund der einfachen Bedienbarkeit des SAS-Analyst trotzdem sinnvoll. In der Analyst-Oberfläche sind nur Fallzahlschätzungen für stetige Variablen in t-Testverfahren enthalten. Fragestellungen mit dichotomen Zielgrößen oder Fallzahlschätzungen für Wahrscheinlichkeiten und Raten [8] sind nicht implementiert.

Äquivalent zur PSS ermöglicht der SAS-Analyst es, das Log-File und den Programmcode, bzw., falls die Berechnungen unter Zuhilfenahme eines Makros stattfinden, den Makrocode einzusehen. Das Log-File ist, falls es nicht schon in SAS geöffnet ist, über View → Log zu sehen.

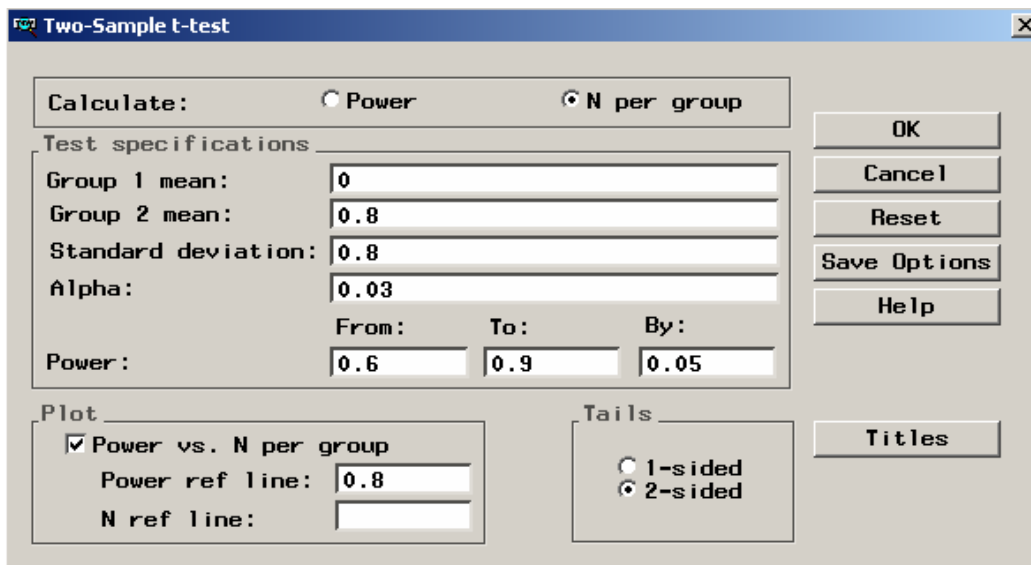


Abbildung 9: Zweistichproben-t-Test in SAS-Analyst

**Fallzahlplanung mit dem Zweistichproben t-Test  
für die MlgAN-Studie**

**Two-Sample t-Test**  
Group 1 Mean = 0    Group 2 Mean = 0.8  
Standard Deviation = 0.8    Alpha = 0.03    2-Sided Test

| Power | N per Group |
|-------|-------------|
| 0.600 | 13          |
| 0.650 | 15          |
| 0.700 | 16          |
| 0.750 | 18          |
| 0.800 | 20          |
| 0.850 | 22          |
| 0.900 | 26          |

**Abbildung 10:** Output der Fallzahlschätzung in SAS-Analyst

## **5 Die SAS-Beispielprogramme des Statistiklehrbuchs *Bestimmung des Stichprobenumfangs von Bock***

Eine Möglichkeit zur Ergänzung der Fallzahlplanungsmöglichkeiten in SAS stellt das ausführlichste deutschsprachige Buch zur Fallzahlplanung - *Bestimmung des Stichprobenumfangs* von Bock [2] - dar. Dem Buch ist eine Diskette beigelegt, auf der 22 SAS-Beispielprogramme gespeichert sind.

Die Programme wurden mit dem Buch im Jahre 1998 vom Oldenbourg-Verlag herausgegeben. Zu dieser Zeit war dies eine der wenigen Möglichkeiten, mit der Statistiksoftware SAS die Fallzahlplanung durchzuführen, da die beiden Prozeduren PROC POWER und PROC GLMPOWER noch nicht existierten. Es besteht bei der Nutzung der Beispielprogramme jedoch das Problem, dass diese nicht validiert sind und daher nicht zur Fallzahlplanung im Rahmen GCP-ICH-konformer Studien verwendet werden sollten.

Zu Beginn beschreibt der Autor die Notwendigkeit der Fallzahlplanung, was die Anforderungen an Sicherheit und Genauigkeit einer klinischen Studie einschließt. Es folgt eine umfangreiche Erläuterung der Fallzahlplanungsmöglichkeiten für unterschiedlichste statistische Testverfahren wie Mittelwertvergleiche, Überlebenszeitanalyse, Äquivalenztests, Varianzanalyse, Vergleich von Wahrscheinlichkeiten und Raten, Korrelation und Regression etc. Am Ende jedes Kapitels wird unter Schlussfolgerungen und Empfehlungen alles Wesentliche zusammengefasst, was zum besseren Verständnis des Buches beiträgt und Denkanstöße für die Lösung eigener Problemstellungen gibt.

**Tabelle 2:** Fallzahlschätzungen bei Bock [2]

| Themengebiet                        | Beschreibung   |
|-------------------------------------|--|
| Mittelwerte und Konfidenzintervalle | Zweistichproben-t-Test   |
|                                     | gepaarter t-Test   |
|                                     | Schätzung eines Populationsmittels mit vorgegebener Güte         |
| Varianzanalyse (ANOVA)              | Einfaktorielle Varianzanalyse                                    |
|                                     | Zweifaktorielle Varianzanalyse                                   |
|                                     | Arzneimittelinteraktionsstudie                                   |
| Überlebenszeitanalyse               | Nach Schoenfeld - mit und ohne Schichten                         |
|                                     | Nach Freedman, Lachin und Foulkes, Bernstein-Lagakos             |
|                                     | Nach dem Markoff-Modell von Lakatos bzw. Log-rank-Test           |
| Wahrscheinlichkeiten/Raten          | $\chi^2$ -Test mit Stetigkeitskorrektur                          |
| Äquivalenzttests                    | Einseitiger Äquivalenzttest für Erfolgsraten                     |
|                                     | Äquivalenzttest für Populationsmittel nach der Westlake-Prozedur |
| Korrelation und Regression          | Vergleich von Korrelationskoeffizienten mit Null                 |
|                                     | Korrelationskoeffizient von Pearson                              |
|                                     | Parallelitätstest von Regressionsgeraden                         |
|                                     | Linearitätstest von Regressionsgeraden                           |
| Sonstiges                           | Schätzung einer Wahrscheinlichkeit mit vorgegebener Güte         |
|                                     | Vorzeichentest   |

Die dem Buch auf Diskette beigelegten Programme sind relativ einfach geschrieben und nicht unter einer Oberfläche zusammengefasst, da der Algorithmus vorgestellt werden soll und dadurch die Möglichkeit besteht, die Programme in eigene SAS-Programme einzubinden. Jedes der SAS-Programme wird anhand eines Beispiels aus dem Bereich der Medizin, der Biologie oder der Landwirtschaft vorgestellt. Eine Übertragung in andere Programmiersprachen ist gegeben, falls Unterprogramme für die verwendeten Funktionen zur Verfügung stehen.

### 5.1 Beispiel der Fallzahlplanung mit N\_TTEST

Für den Zweistichproben-t-Test kann bei der Studie von einer mittleren Differenz der Variablen X von  $\Delta = 0,8$  Einheiten zwischen der Behandlungs- und der Kontrollgruppe und einer Standardabweichung von  $\sigma = 0,8$  Einheiten ausgegangen werden. Als Irrtumswahrscheinlichkeit wird  $\alpha = 0,03$  vorausgesetzt. Der berechnete Wert des Nicht-zentralitätsparameters c wird in den Programmkopf von N\_TTEST eingesetzt:

$$c = \frac{\Delta}{2\sigma} = \frac{0,8}{2 \cdot 0,8} = 0,5$$

```

*****
**PROGRAMM N_TTEST: Berechnet den exakten Gesamtumfang fuer **
**den t-Test
*****
%let fname=MIGAN;          *Name=Name des Ergebnisfiles
***** VORGABEN (beliebig viele Daten_Zeilen) *****
data &fname;
seitig  alpha  c          DF_      power0;
2       0.03   0.5        2        0.8
;run;
*****
** alpha=vorgeg.Irrtumswkt;   ** seitig=1 od. 2,;
** c=Nichtzentralitaetsparameter/sqrt(N)
** power0=vorgeg. Power
** Waehle DF_ so, dass Freiheitsgrade DF= N - DF_
*****

```

**Abbildung 11:** Eingabe der Parameter in N\_TTEST

Exakter Gesamtumfang fuer den t-Test

| seitig | alpha | c   | DF_ | power0 | df | power   | nexakt |
|--------|-------|-----|-----|--------|----|---------|--------|
| 2      | 0.03  | 0.5 | 2   | 0.8    | 37 | 0.80293 | 39     |

**Abbildung 12:** Ergebnis des Programms N\_TTEST

Der Gesamtumfang beträgt 39 Patienten, was in etwa der mit der PSS berechneten Fallzahl von 20 Patienten pro Gruppe entspricht.

## 6 Makros zur Fallzahl- und Powerschätzung im Internet

Die Benutzeroberfläche PSS und die Beispielprogramme des Statistiklehrbuches *Bestimmung des Stichprobenumfangs* von Bock decken zwar schon eine Vielzahl der von in der Fallzahlplanung benötigten Testverfahren ab, doch eine Ergänzung der Möglichkeiten in SAS durch SAS-Makros zur Fallzahlplanung im Internet könnte der Vervollständigung dienen.

Die Richtlinien zur Durchführung von Studien der EMEA [3] legen fest, dass zur Berechnung aller Einflussgrößen einer Studie nur validierte Software zum Einsatz kommen darf. Nichtsdestotrotz wird die Webrecherche durchgeführt, da hierbei die Möglichkeit besteht, validierte Makros zur Fallzahlplanung zu finden. Allerdings ist bei keinem der gefundenen Makros bekannt, ob es validiert ist. Die Lauffähigkeit jedes einzelnen beschriebenen Makros wird getestet, wobei anzumerken ist, dass die Mehrzahl der Makros durch die zumeist mitgelieferten Beispiele problemlos zum Laufen gebracht werden kann. Nur insgesamt zwei Makros sind nicht lauffähig.

### 6.1 Verschiedene Suchmaschinen

Es bietet sich die Suche mit einer Auswahl an mit verschiedenen Methoden arbeitenden Suchmaschinen an. Der erste Schritt besteht in der Einarbeitung in unterschiedliche von Suchmaschinen verwendete Konzepte zur Webrecherche.

Volltextsuchmaschinen bieten die Möglichkeit, den gesamten Text einer Website auszuwerten, wie z.B. google und yahoo. Andere Suchmaschinen begnügen sich bei der Indexierung mit den ersten Sätzen einer Website. Einige Suchmaschinen bedienen sich sogenannter Roboter bzw. Crawler, die aus tausenden von Computern bestehen, die fortlaufend Internetinhalte analysieren [7].

Metasuchmaschinen kombinieren die Ergebnisse mehrerer Suchmaschinen miteinander. Sie besitzen keine eigene Datenbank. Dadurch erhält man - je nach Einstellung der Suchmaschine - einen wesentlich größeren Umfang an Treffern, die allerdings unter Umständen weniger relevant sind oder bei Verwendung eines Filters relevantere Ergebnisse aus weniger Treffern.

Es wurde mit den Suchmaschinen abacho.de, suchnase.de, fireball.de, yahoo.de, google.de und metafinder.de gesucht.

Die Suchwörter für alle fünf verwendeten Suchmaschinen und die Metasuchmaschine waren stets folgende sechs:

- Stichprobenumfang SAS Makro
- Stichprobenumfang SAS
- Fallzahlplanung SAS Makro
- Fallzahlplanung SAS
- sample size SAS macro
- sample size SAS

## 6.2 Ergebnisse der Webrecherche

Es wurden 48 Makros näher untersucht, wovon 46 lauffähig waren. Hierbei stellte sich das Makro UnifyPow von O'Brien, einem Statistiker an der Cleveland Clinic Foundation, als mit Abstand interessantestes heraus, da es 25 Fallzahlschätzungen beinhaltet. Es wird unter Abschnitt 6.3 näher vorgestellt.

Besondere Auswertungssituationen bzgl. der Fallzahlschätzung sind unter anderem Tukeys-Trend-Test von Duncan (<http://www.lexjansen.com/pharmasug/2000/stats/st10.pdf>), die Power wiederholter Messwerte und geteilter graphischer Darstellungen von Chapman (<http://www.stat.colostate.edu/consulting/Statlab/Examples/powermeas.sas>), Überlebenszeitanalyse unter Beachtung der genauen Progredienzzzeiten von Chang (<http://www.lexjansen.com/~pharmasug/2006/statisticspharmacokinetics/sp07.pdf>), Äquivalenzstudien nach Benner und Ortseifen (<http://www.dkfz-heidelberg.de/biostatistics/benner/eqsurv.sas>) und die Überlebenszeitanalyse nach Hsieh und Lavori von Koehler (<http://www.public.iastate.edu/~kkoehler/stat565/sas/phpow.macro.sas>). Zusätzlich wurden unter [http://www.biostat.ufl.edu/~sample\\_programs](http://www.biostat.ufl.edu/~sample_programs) ein Makro zur Wahrscheinlichkeit der Exposition bei gegebenem Odds Ratio, ein Kontrollprogramm für binomiale Ansprechraten und ein Makro zur Berechnung einer asymptotischen Fallzahl zur Überprüfung einer auf der Log-Transformation beruhenden Korrelation von Shuster gefunden. Woodworth stellt sein Makro für ein Experiment mit einem Studienarm mit Bernoulli-Zielgröße unter <http://www.stat.uiowa.edu/~gwoodwor/BBI-Text/RatePrePost.sas> zum Download bereit. Die Power des Transmission-Disequilibrium-Tests aus der Genetik kann dank eines Makros von Knapp (<http://www.uni-bonn.de/~umt70e/tdtpower.zip>) berechnet werden.

## 6.3 Die vielfältige Nutzbarkeit des Makros UnifyPow

Das Makro UnifyPow kann nicht nur zur Fallzahlplanung für ein einzelnes statistisches Testverfahren wie alle anderen gefundenen Makros verwendet werden, sondern auch zur Realisierung der Fallzahlplanung für eine ganze Reihe statistischer Testverfahren. Das Makro UnifyPow wurde von O'Brien [9] entwickelt. Detailliertere Informationen über das Makro sowie hilfreiche Beispiele und Tipps zur Arbeit mit dem Makro werden unter <http://www.bio.ri.ccf.org/power.html> zur Verfügung gestellt.

Insgesamt ist das Makro sehr gut implementiert, weil es alle denkbaren Facetten der Fallzahlplanung abdeckt, allerdings ist es ohne tiefere Statistikkennntnisse nicht unbedingt anwendbar. Programmierkenntnisse sind für den Gebrauch des Makros zweitrangig, da nur relevant ist, an welcher Stelle des Programms die gewünschten Parameter einzutragen sind.

**Tabelle 3:** Fallzahlschätzungen des Makros UnifyPow [9]

| Themengebiet               | Beschreibung   |
|----------------------------|--|
| Mittelwerte                | Einstichproben-t-Test  |
|                            | Einstichproben-t-Test für verbundene Stichproben   |
|                            | Einstichproben-Wilcoxon-signed-rank-Test   |
|                            | Zweistichproben-t-Test   |
|                            | Zweistichproben-t-Test für verbundene Stichproben  |
|                            | Zweistichproben-Wilcoxon-test für verbundene Stichproben                                 |
|                            | Tests für Unterschiede zwischen verbundenen Stichproben                                  |
| Varianzanalyse (ANOVA)     | Einfaktorielle Varianzanalyse mit dem globalen F-Test                                    |
|                            | Zweifaktorielle Varianzanalyse   |
|                            | Mehrfaktorielle Varianzanalyse mit Wechselwirkungen                                      |
| Wahrscheinlichkeiten/Raten | Einstichproben-Binomialtest  |
|                            | $\chi^2$ -Test   |
|                            | Likelihood-Ratio-Test auf Zusammenhang   |
|                            | Globaler Likelihood-Ratio-Test   |
|                            | Globaler Likelihood-Ratio-Test für die multiple logistische Regression                   |
|                            | Genereller Likelihood-Ratio-Test für die logistische Regression oder log-lineare Modelle |
|                            | Exakter Test von Fisher  |
|                            | Mc Nemar-Test  |
|                            | Wald-type-Test   |
| Äquivalenztests            | Äquivalenztest von Blackwelder   |
| Korrelation und Regression | Einstichproben-Test für die Korrelation nach Pearson                                     |
|                            | Globaler F-Test für die Methode der kleinsten Quadrate im multiplen Regressionsmodell    |
|                            | Multiple partielle Korrelation von $q$ zusätzlichen Einflusswerten                       |
|                            | Lineare Kontrastypothesen unabhängiger Korrelationen nach Pearson                        |
|                            | Univariates und multivariates generelles lineares Modell                                 |

### 6.3.1 UnifyPow anhand des Zweistichproben-t-Tests

Auch für den Zweistichproben-t-Test wird man in der Beispielsammlung von O'Brien fündig. Zuerst wird die Makrovariable UnifyPow zur Speicherung des Makropfades angelegt, die anschließend über `%include` importiert werden muss.

```
%let UnifyPow =F:\Katrin\Bachelorarbeit\Webrecherche\Makros_Webrecherche\UnifyPowLatest.sas;
```

Anschließend trägt man die benötigten Parameter für die Studie ein.

```
%include "&UnifyPow";
title2 "Zweistichproben-t-Test für die Studie";
datalines;
```



```
mu 0 0.8
SD .8
alpha .03
power .8
%tables;
```

**Zweistichproben t-Test für die MigÄN-Studie**

Scenario: mu 0 .08  
and Effect: Compare two means

|       |           |                                 |
|-------|-----------|---------------------------------|
|       |           | Stand-<br>ard<br>Devia-<br>tion |
|       |           | 0.08                            |
|       |           | Minim-<br>um<br>Power           |
|       |           | 0.800                           |
|       |           | Total<br>N                      |
| Alpha | Type      |                                 |
| 0.030 | 2-sided t | 40                              |
|       | 1-sided t | 32                              |

**Abbildung 13:** Output von UnifyPow

Es wird automatisch die Fallzahl für ein- und zweiseitige Fragestellungen berechnet. Von Bedeutung für die Studie ist die zweiseitige Fragestellung. Auch hier wird, wie in Abb. 13 unter 2-sided t für eine zweiseitige Fragestellung abzulesen, eine Fallzahl von insgesamt 40 Patienten berechnet.

## 7 Diskussion

Möglichkeiten der Fallzahlplanung mit der Statistiksoftware SAS schließen die in SAS 9.1 neu eingeführten Prozeduren PROC POWER bzw. PROC GLMPOWER, die Benutzeroberflächen PSS und SAS-Analyst, die Beispielprogramme von Bock [2] sowie bei der Webrecherche gefundene Makros zur Fallzahlplanung ein.

Die in SAS 9.1 mitgelieferte Benutzeroberfläche PSS ermöglicht es, Fallzahlschätzungen ohne Programmierkenntnisse menügesteuert auf der Basis statistischer Grundkenntnisse durchzuführen. Da sie und der zugehörige Server Tomcat 1.4 standardmäßig auf der SAS Installations-CD mitgeliefert werden, ist sie mit relativ kleinem Aufwand zu installieren. Zusätzlich ist es möglich, die PSS über einen zentralen Server zum Laufen

zu bringen. Dann müsste Tomcat 1.4 nicht auf jedem einzelnen PC, sondern einmal zentral installiert werden. Es besteht die Möglichkeit, den SAS-Code und das Log-File einzusehen und den SAS-Code zum selbstständigen Programmieren mit PROC POWER bzw. PROC GLMPOWER in SAS zu übertragen. In einigen Fällen, z.B. zur Erstellung von verschiedenfarbigen Kurven pro Graphik, der Ausgabe der Fallzahl ohne vorheriges Aufrunden, der Änderung der Reihenfolge der Ausgabeparameter oder der Erzeugung von SAS-Dateien mittels ODS-Statement ist direkt auf die beiden Prozeduren PROC POWER bzw. PROC GLMPOWER zurückzugreifen, ansonsten ist die Verwendung der PSS anzuraten.

Die zweite, standardmäßig seit Version 8 in SAS enthaltene Benutzeroberfläche SAS-Analyst kann zur Fallzahlplanung für Testverfahren mit stetigen Variablen verwendet werden. Ihre besondere Eignung liegt in der schnellen Fallzahlschätzung ohne zusätzlichen Installationsaufwand.

Die Ergänzung der Möglichkeiten der SAS PSS durch die Beispielprogramme von Bock [2] führt zu einer weiteren Anzahl von Fallzahlschätzungsmöglichkeiten. Jedes Programm wird anhand eines Beispiels im Buch anschaulich erläutert. Durch die Beispielprogramme von Bock besteht die Möglichkeit, die Fallzahlschätzungen der PSS durch umfangreichere Testverfahren zu ergänzen. Auf Grund der bestehenden Fehleranfälligkeit und des erhöhten Aufwandes bei der Berechnung der Eingabeparameter sollten die Beispielprogramme von Bock [2] nur für Fallzahlschätzungen angewendet werden, die nicht in der PSS enthalten sind.

Eine weitere Ergänzung der Fallzahlschätzungen in SAS stellt die Webrecherche nach SAS-Makros dar. Man entdeckt dabei sehr spezielle statistische Testverfahren, die nicht standardmäßig in der PSS bzw. PROC POWER/GLMPOWER enthalten sind. Das Makro UnifyPow von O'Brien [9] ist vielseitig einsetzbar und beinhaltet 25 verschiedene Fallzahlschätzungen, wobei alle anderen Makros nur eine Fallzahlschätzung enthalten. Eine sorgfältige Abwägung zwischen den Risiken und dem möglichen Nutzen der Verwendung der Makros zur Fallzahlplanung ist jeweils vor Gebrauch nötig.

Sowohl bei den Beispielprogrammen von Bock als auch bei den im Rahmen der Webrecherche gefundenen Makros besteht das Problem, dass nicht bekannt ist, ob diese validiert sind. Daher dürfen sie nicht ohne weiteres für Fallzahlschätzungen nach den GCP/ICH Richtlinien der EMEA [3] verwendet werden.

## Literatur

- [1] Altman, D.: Practical Statistics for Medical Research. London: Chapman& Hall, 1991
- [2] Bock, J.: Bestimmung des Stichprobenumfangs. Wien, München: Oldenbourg, 1998
- [3] Good Clinical Practice: Consolidated Guideline (ICH E9). EMEA, London, 1998.– Download unter: <http://www.emea.europa.eu/pdfs/human/ich/036396en.pdf> (zuletzt aufgerufen: 20.02.2008)

- [4] Getting Started with the SAS Power and Sample Size Application. SAS Institute Inc., Cary, NC, USA, 2004
- [5] Häußler, K.: Fallzahlplanung mit der Statistiksoftware SAS – Möglichkeiten und Limitierungen. Bachelorarbeit, Hochschule Ulm, 2007
- [6] Hartung, J.; Elpelt, B.; Klösener, K. H.: Statistik. 11. Auflage. München, Wien: Oldenbourg, 1998
- [7] Lönneker, B.: Konzeptframes und Relationen. Extraktion, Annotation und Analyse französischer Corpora aus dem World Wide Web. Funktionen von Suchmaschinen. – Online unter: [http://www1.uni-hamburg.de/lingkonnet/Dissertation/Konzept\\_frames/2\\_2\\_Funktionsweise\\_Suchmas.html](http://www1.uni-hamburg.de/lingkonnet/Dissertation/Konzept_frames/2_2_Funktionsweise_Suchmas.html) (zuletzt aufgerufen: 20.02.2008)
- [8] Muche, R.; Habel, A.; Rohlmann, F.: Medizinische Statistik mit SAS-Analyst. Berlin, Heidelberg: Springer, 2000
- [9] O'Brien, R.: A Tour of UnifyPow: A SAS Module/Macro for Sample-Size Analysis. SUGI, St. Louis, 1998. – Online unter: [http://www.bio.ri.ccf.org/Unify\\_Pow.all](http://www.bio.ri.ccf.org/Unify_Pow.all) (zuletzt aufgerufen: 20.02.2008)
- [10] Ortseifen, C.; Kuzhel, K.: Das Datenanalysestsystem SAS – Eine Einführung mit der Enterprise Guide Software 4 – Online unter: <http://www.urz.uni-heidelberg.de/statistik/kurse/EG-Selbststudium/SASEG41.pdf> (letzter Aufruf: 20.02.2008)
- [11] Rosner, B.: Fundamentals of Biostatistics. 5th Edition. Pacific Grove, CA, USA: Duxbury, 2000, S. 236–243
- [12] SAS OnlineDoc 9.1.3 - PROC GLMPOWER. SAS Institute Inc., Cary, NC, USA, 2006. – Online unter: <http://support.sas.com/91doc/getDoc/statug.hlp/glmin-dex.htm>, (zuletzt aufgerufen: 20.02.2008)
- [13] SAS OnlineDoc 9.1.3 - PROC POWER. SAS Institute Inc., Cary, NC, USA, 2006. – Online unter: <http://support.sas.com/91doc/getDoc/statug.hlp/powerindex.htm>, (zuletzt aufgerufen: 20.02.2008)
- [14] Slaughter, S.; Delwiche, L.: The Little SAS Book for Enterprise Guide 4.1. Cary, NC, USA: SAS Institute Inc., 2006