

# Eine interaktive Einführung in die Medizinische Biometrie anhand der NANA-Studie und deren Umsetzung in SAS

Benjamin Mayer, Bettina Danner, Rainer Muche  
Institut für Epidemiologie und Medizinische Biometrie  
Universität Ulm  
Schwabstraße 13  
89075 Ulm  
benjamin.mayer@uni-ulm.de

## Zusammenfassung

Das Fach Medizinische Biometrie ist im Medizinstudium verankert, um den Studenten die Grundlagen der statistischen Planung, Durchführung, Auswertung und Interpretation klinischer Studien zu vermitteln. Diese Kenntnisse sind von großer Relevanz für die eigenständige Bewertung neuer Forschungserkenntnisse und deren inhaltlicher Bedeutung für den medizinischen Alltag. Im Rahmen dieses Beitrags möchten wir die von uns konzipierte NANA-Beispielstudie vorstellen, die im Rahmen des „Schülernachmittags“ der Universität Ulm entwickelt wurde und als aktivierende Lehrmethode in Vorlesungen und Seminare zur Medizinischen Biometrie integriert werden könnte. Die Implementation der Studie erfolgte in SAS. Es werden unter aktiver Beteiligung der Probanden die wesentlichen Aspekte einer klinischen Studie demonstriert. Ziel der Studie ist es herauszufinden, ob sich Teilnehmer, die eher zu Süßigkeiten tendieren („NAschkatzen“), im Vergleich zu Teilnehmern, die eher zu Knabberereien tendieren („NAgetiere“), hinsichtlich ihres Körpermasseindex (BMI) unterscheiden.

**Schlüsselwörter:** Lehrprojekt, Klinische Studie, %Windows, SAS-Makro

## 1 Einleitung

Das Fach Medizinische Biometrie ist im Medizinstudium verankert, um den Studenten die Grundlagen der statistischen Planung, Durchführung, Auswertung und Interpretation klinischer Studien zu vermitteln. Diese Kenntnisse sind von großer Relevanz für die eigenständige Bewertung neuer Forschungserkenntnisse und deren inhaltlicher Bedeutung für den medizinischen Alltag. In diesem Beitrag möchten wir die von uns konzipierte NANA-Beispielstudie vorstellen, die im Rahmen des „Schülernachmittags“ der Universität Ulm entwickelt wurde und als aktivierende Lehrmethode in Vorlesungen und Seminare zur Medizinischen Biometrie integriert werden könnte. Die Implementation der Studie erfolgte in SAS. Es werden unter aktiver Beteiligung der Probanden die wesentlichen Aspekte einer klinischen Studie demonstriert. Ziel der Studie ist es herauszufinden, ob sich Teilnehmer, die eher zu Süßigkeiten tendieren („NAschkatzen“), im Vergleich zu Teilnehmern, die eher zu Knabberereien tendieren („NAgetiere“), hinsichtlich ihres Körpermasseindex (BMI) unterscheiden und weist auf die NANA-Statue an der Universität Ulm hin (s. Abb. 1).

## 2 Die NANA-Studie

Die Studie umfasst eine Befragung der Teilnehmer inklusive Messung der für die Berechnung des BMI notwendigen Merkmale Körpergröße und –gewicht, sowie eine Datenein- und ausgabe über entsprechend konzipierte SAS-Programme. Die Teilnehmer sind aufgefordert, eine Süßigkeit oder eine Knabberlei zu wählen und anschließend an entsprechend eingerichteten Stationen (Personenwaage und Meterstab) ihre individuellen Daten zu erfassen, um sie nebst anderen in einen Fragebogen einzutragen (s. Abb. 3). Die ausgefüllten Erhebungsbögen werden von einem Mitarbeiter des Lehrteams oder von den Teilnehmern selber in eine SAS-Eingabemaske eingegeben. Nach Abschluss der Dateneingabe bekommt jeder Studienteilnehmer eine ID-Nummer auf seinem Erhebungsbogen notiert und das Auswertungsprogramm wird gestartet. Über die ID kann zuletzt jedem Teilnehmer ein Ausdruck der individuellen Ergebnisse zur Verfügung gestellt werden. Eine Übersicht über den Ablauf der NANA-Studie zeigt Abbildung 2.



Abbildung 1: NANA-Statue an der Universität Ulm

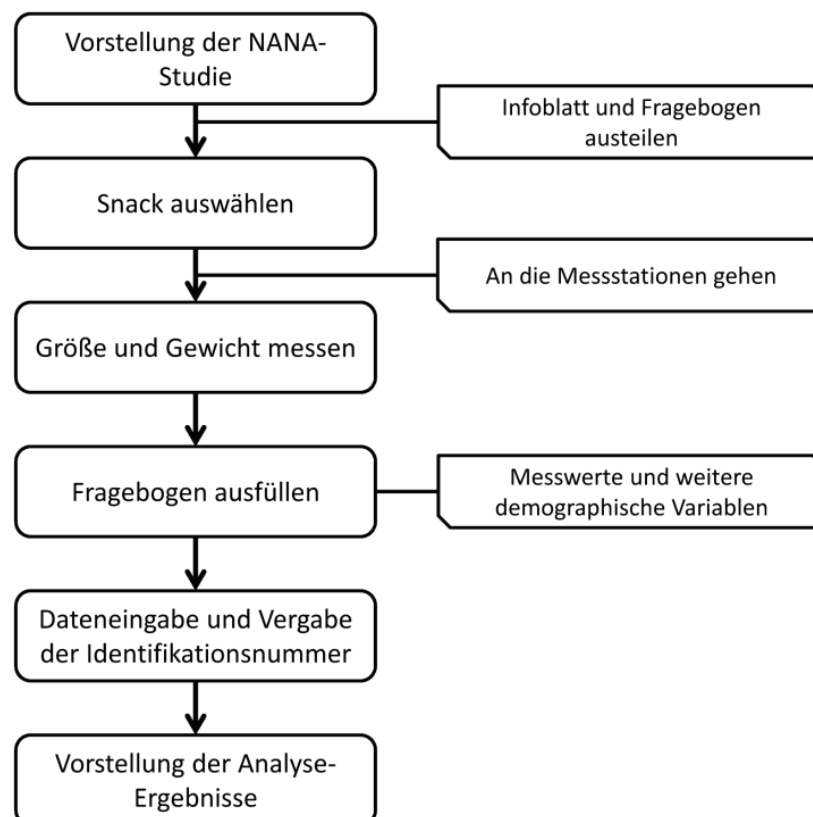


Abbildung 2: Ablauf der NANA-Studie

Der folgende Erhebungsbogen mit den wichtigsten Items wird in der Studie eingesetzt. Neben der Erfassung der „Therapie“ (Süßigkeit oder Knabberie) werden die Parameter für die Berechnung des BMI (Verhältnis zwischen Körpergewicht zu Körpergröße) erfasst. Geschlecht, Alter und Wohnort stehen dann für Untergruppenanalysen zur Verfügung. Gerade das Bundesland ist für einen Standort direkt an der Landesgrenze immer wieder für eine Diskussion mit den Teilnehmern über solche Untergruppenanalysen interessant. Die „anonyme“ Identifikationsnummer ist notwendig, um den Teilnehmern anschließend ihre individuellen Ergebnisse mitzuteilen, führt aber auch zur Problematik der Anonymität in klinischer Forschung.

\*\*\*\*\*

### Erhebungsbogen NANA-Studie

\*\*\*\*\*

Ich bin  weiblich  
 männlich

und wurde am   .   .     geboren.

Ich wohne in  Baden-Württemberg  
 Bayern  
 anderer Wohnort

Die Messungen ergaben

eine Körpergröße von    cm  
 und ein Gewicht von    .  kg.

Ich habe mich entschieden für

eine Süßigkeit  
 eine Knabberie

Meine persönliche Identifikationsnummer lautet:

**Abbildung 3:** Erhebungsbogen NANA-Studie

Die Erfassung der Parameter führen die Teilnehmer selbst aus, indem eine Personenwaage und ein Meterstab zur Verfügung gestellt werden. Hier wird auch oft über die Genauigkeit von Daten diskutiert.

Um die NANA-Beispielstudie durchführen zu können, müssen im Vorfeld noch verschiedene EDV-Materialien bereitgestellt werden. Es werden ein Laptop und ein Drucker benötigt. Auf dem Laptop sollte SAS<sup>®</sup> in der Version 9 installiert sein, da die Da-

teneingabe und –auswertungsprogramme dies voraussetzen. Für den Ausdruck der Analyseergebnisse wird ein Drucker mit ausreichend Papier benötigt. Jeder Ausdruck ist probandenspezifisch und beinhaltet zum einen die persönlichen Daten, die abgegeben wurden, sowie eine zusammenfassende Analyse des Gesamtkollektivs. Über die zuvor vergebene Identifikationsnummer können die Ausdrücke den jeweiligen Probanden zugeordnet werden.

### 3 Umsetzung in SAS

Die probandenspezifischen Daten werden in einer SAS<sup>®</sup>-Datei gespeichert und in Abhängigkeit des jeweiligen Merkmalstyps zunächst deskriptiv mit absoluten und relativen Häufigkeiten für qualitative Daten bzw. Mittelwert, Standardabweichung, Median, Minimum und Maximum für metrisch skalierte Daten ausgewertet. Ein Vergleich der medianen BMI-Werte von „Naschkatzen“ und „Nagetieren“ erfolgte mit dem unverbundenen Wilcoxon-Test bei einem zweiseitigen Signifikanzniveau von 5%. Eine zusätzliche Prüfung auf Häufigkeitsunterschiede zwischen den beiden Gruppen hinsichtlich der jeweiligen BMI-Kategorie „leicht“-, „normal“- und „schwergewichtig“ wurde mit dem exakten Test nach Fisher durchgeführt.

Die Umsetzung der Lehrstudie basiert auf 6 SAS-Programmen und 2 Datensätzen. Für die Dateneingabe werden einfache SAS-Makros, hauptsächlich auf Basis des `%WINDOW` Statements, verwendet. Ein `ODS` Befehl wird zur Ausgabe der Outputs genutzt. Voraussetzung ist SAS Version 9.x aufgrund der Verwendung indirekter Pfade, zudem sollten alle Dateien in einem Verzeichnis sein. *Nana.sas* legt die notwendigen Pfade und Verzeichnisse an, sowie beim erstmaligen Aufruf die Datei *Alle.sas7bdat* zur Speicherung aller Studienteilnehmer. Über das Programm *U\_nana.sas* kann die Dateneingabe gestartet werden, da dieses wiederum auf das Programm auf *U\_nana\_windows.sas* zugreift, in dem über das `%WINDOW` Statement das Eingabefenster implementiert wurde. Die Daten werden über *U\_nana\_macros.sas* in zwei Dateien abgelegt, um eine ID-gesteuerte Zuordnung der Auswertungsergebnisse pro Teilnehmer zu ermöglichen.

Die folgende skizzierte Vorgehensweise gibt die Reihenfolge der Aufrufe und Nutzung der Programme und Dateien wieder. Die Programme können vom Erstautor gerne angefordert werden, um die genauere Programmierung ansehen zu können.

#### Vorgehen:

##### 1) Datenerfassung und Speicherung

- Nana.sas
- U\_nana\_formate.sas
- U\_nana\_windows.sas, U\_nana.sas
- Alle.sas7bdat
- Auswertung.sas7bdat

## 2) Auswertung und rtf-Ausgabe

- U\_nana\_macros.sas
- U\_nana\_formate.sas
- U\_nana\_auswertung.sas
- Auswertung.sas7bdat

## 3) Aufruf der Programme

1. Nana.sas (ruft U\_nana\_formate.sas und U\_nana\_windows.sas auf, legt Pfade und Verzeichnisse an, benötigt Alle.sas7bdat)
2. U\_nana.sas (mind. 1x pro Tag: über dieses Fenster wird die Abfrage von Teilnehmerdaten ermöglicht; legt Daten in Alle.sas7bdat und Auswertung.sas7bdat ab)
3. U\_nana\_macros.sas (ausführen bei jeder Aktualisierung der Datei Auswertung.sas7bdat)
4. U\_nana\_auswertung.sas

## 4 Dateneingabe und Ergebnisdarstellung

Der Probanden-spezifische Ausdruck umfasst drei Bereiche: Zunächst erfolgt eine Ausgabe der Basisdaten des teilnehmenden Probanden (s. Abb. 4). Es werden sowohl die im Fragebogen erhobenen Merkmale nochmals ausgegeben, als auch der auf der Basis dieser Angaben berechnete BMI des Probanden.

***	Vielen Dank für Ihre Teilnahme an der NANA-Studie			***
*	Studienteilnehmer insgesamt:	218	und am heutigen Tag:	2
Ihre Daten				
Ihre Identifikationsnummer: 2				
Sie sind: fehlend				
wurden am: 10.06.1963 geboren				
und sind heute, am: 12.01.2010, 46.6 Jahre alt.				
Sie kommen aus: Baden-Württemberg,				
wählten eine: Knabbererei und gehören somit in die Gruppe der > Nagetiere <				
Bei einem Gewicht von: 52.4 kg				
und einer Größe von: 1.61 m				
sind Sie mit einem BMI von 20.2 kg/m <sup>2</sup> normalgewichtig.				

**Abbildung 4:** Ausgabe der Basisdaten eines Probanden

Im zweiten Teil wird die Hauptfragestellung der NANA-Studie beantwortet. Von primärem Interesse ist, ob sich Naschkatzen (Süßigkeit gewählt) von Nagetieren (Knabbererei gewählt) hinsichtlich des BMI unterscheiden. Die Auswertung umfasst zwei Ansätze: Zum einen werden die medianen BMI-Werte in beiden Gruppen mit dem Wilcoxon-Rangsummen-Test verglichen, zum anderen erfolgt eine Klasseneinteilung der BMI-Werte nach leicht-, normal- und schwergewichtig, um die entsprechenden Häufigkeiten anschließend mit dem exakten Test nach Fisher zu vergleichen (s. Abb. 5). Für beide Auswertungsansätze wird eine separate Deskription der beiden Kollektive durchgeführt und jeweils der beim Probanden berechnete BMI-Wert mit angegeben.

Nach 218 Studienteilnehmern ist hier zu erkennen, dass sich die Mehrheit (65%) für Süßigkeiten entschieden hat.

Auswertung 1:										
Gruppe	Ihr Wert	BMI [kg/m <sup>2</sup> ]								
		Anzahl	Mittelwert	Standardabweichung	Minimum	Median	Maximum			
> Naschkatzen <	-	-	142	21.8	4.1	11.7	22.1	40.8		
> Nagetiere <	20.2	normal	76	22.7	4.5	13.7	22.9	33.0		

**Bewertung:**  
 Der Unterschied zwischen den Gruppen der > Naschkatzen < mit n = 142 Teilnehmern und der > Nagetiere < mit n = 76 Teilnehmern ist für den BMI mit 22.1 kg/m<sup>2</sup> bzw. 22.9 kg/m<sup>2</sup> mit einem p-Wert von 9.8% (für den zweiseitigen Wilcoxon-Test) als eher zufällig zu bewerten.

Auswertung 2:										
Gruppe	Ihr Wert	BMI (klassiert)								
		leicht		normal		schwer		Gesamt		
		Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	
> Naschkatzen <	-	-	28	12.8	90	41.3	24	11.0	142	65.1
> Nagetiere <	20.2	normal	15	6.9	39	17.9	22	10.1	76	34.9

**Bewertung:**  
 Für den klassierten BMI ergibt sich mit einem p-Wert von 10.4% (für den zweiseitigen Fisher-Test) ein eher zufälliger Unterschied.

**Abbildung 5:** Zwei-Gruppen-Vergleich des medianen BMI mit dem Wilcoxon-Test bzw. der Häufigkeiten in BMI-Klassen mit dem exakten Test nach Fisher

Der mediane Unterschied zwischen Naschkatzen (22.1) und Nagetieren (22.9) liegt bei 0.8 Punkten auf der BMI-Skala. Bei den Naschkatzen sind 63% (90 von 142) normalgewichtig, während bei den Nagetieren nur 51% (39 von 76) im Normalbereich des BMI liegen. Beide Auswertungsansätze können einen signifikanten Unterschied zwischen Naschkatzen und Nagetieren nicht nachweisen, auch wenn die Beschreibung der Daten darauf hinweist, dass Naschkatzen tendenziell einen niedrigeren BMI haben und seltener als schwergewichtig eingestuft werden. Die insgesamt eher niedrigen medianen BMI-Werte in beiden Gruppen sind auf das überwiegend junge Alter der bisherigen Studienteilnehmer zurückzuführen.

Im letzten Teil des Ausdrucks werden zwei Grafiken dargestellt, welche die Häufigkeiten der verschiedenen BMI-Klassen in Abhängigkeit des Geschlechts (s. Abb. 6) und des Wohnortes abbilden, jeweils getrennt für beide Gruppen der Naschkatzen und Nagetiere.

Geplant ist, die NANA-Studie regelmäßig einzusetzen, so dass sich über die Zeit der Datenbestand erhöht und die Fragestellung anhand der weiteren Daten genauer untersucht werden kann.

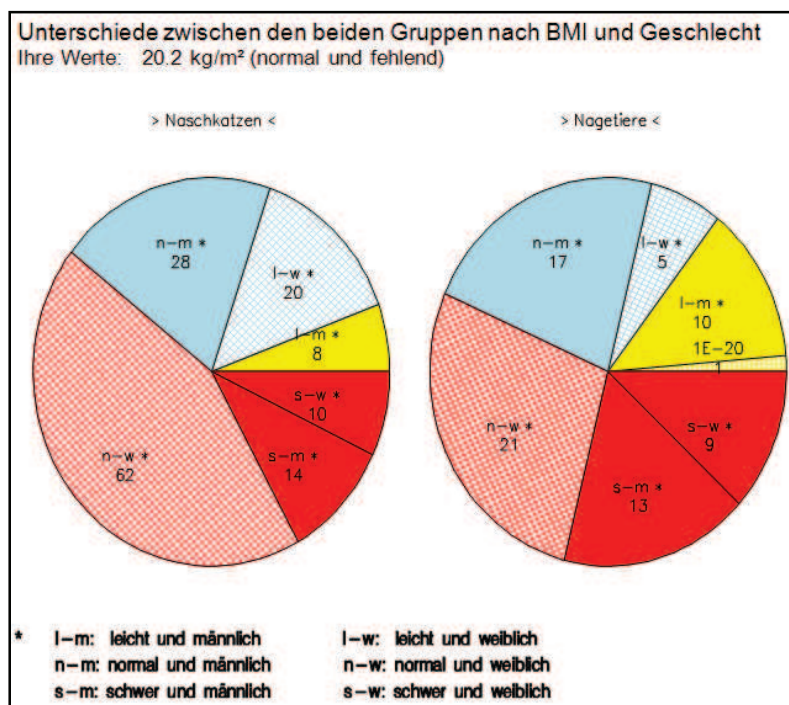


Abbildung 6: Grafische Darstellung nach BMI-Klasse und Geschlecht

## 5 Fazit

Durch die Nutzung einer einfachen Studie als aktivierendes Lehrkonzept (Lehre zum Mitmachen) kann den Teilnehmern die Bedeutung des Faches Medizinische Biometrie für die medizinische Forschung anschaulich vermittelt werden – zumindest deuten die Rückmeldungen der Teilnehmer darauf hin. Gerade in mathematisch-naturwissenschaftlichen Fächern ist es sehr wichtig, Alternativen zu einem klassischen Frontal-„Unterricht“ anzubieten, um die Hemmschwelle der Lernenden gegenüber den Lehrinhalten herabzusetzen. Eine interaktive Einführung in die Medizinische Biometrie ermöglicht es somit, dass die grundlegenden Begrifflichkeiten und Konzepte der Planung und Auswertung klinischer Studien, sowie die Erzeugung statistisch-begründeter, wissenschaftlicher Evidenz in einer lockeren und angenehmen Lernumgebung vermittelt werden.

Der Einsatz der Statistiksoftware SAS mit einfachen Komponenten (z.B. für die Datenerhebungsmaske über %Windows und einfachen Auswertungsprozeduren) hilft dabei, die Anwendungsseite der Forschung zu veranschaulichen. Wir lassen nicht hinter die Kulissen blicken – also die Programmierung zeigen – aber geben doch eine Einsicht, dass klinische Forschung mit Computeranwendungen kein Hexenwerk ist.

Die Integration der NANA-Studie in Vorlesungen und Seminare zur Medizinischen Biometrie könnte positiv dazu beitragen, dass die Hemmschwelle gegenüber unserem Fach herabgesetzt wird. Im Rahmen einer lockeren Lehr-Lern-Atmosphäre kann in Zusammenarbeit mit den Studienteilnehmern ein Grundverständnis zur Bedeutung und den Konzepten statistischer Auswertungen in der Medizin entwickelt werden.