

## **Erfahrungen mit SAS bei statistischen Problemlösungen und Erarbeitung von Lehrmaterialien in der Humanbiologie**

Holle Greil (Universität Potsdam, Humanbiologie)

Andreas Baudisch (Humboldt Universität zu Berlin, Rechenzentrum)

Wolf F. Lesener (Humboldt Universität zu Berlin, Rechenzentrum)

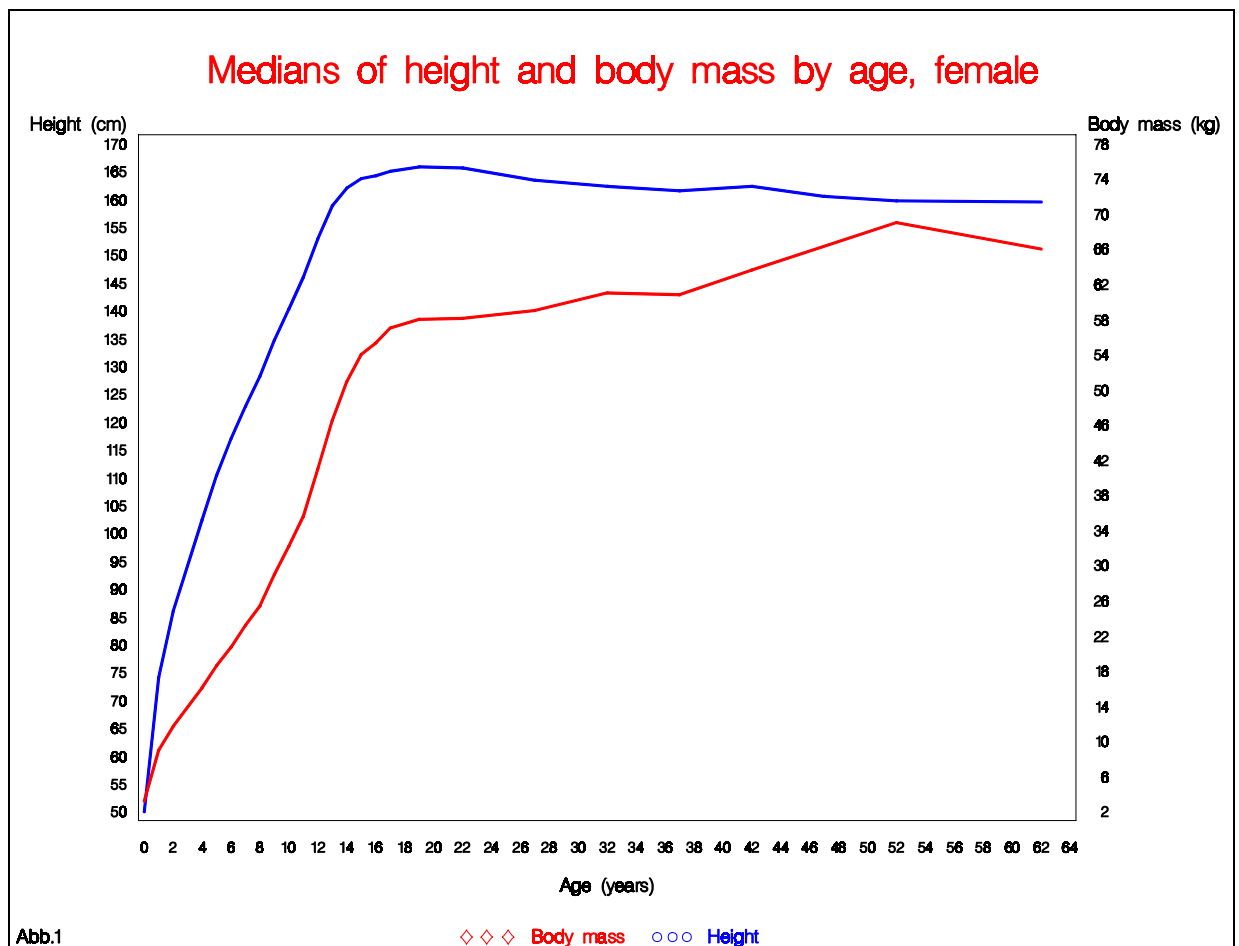
Arbeiten zur Erforschung der Variabilität der körperlichen Entwicklung von Menschen in Abhängigkeit von den unterschiedlichsten genetischen und umweltbedingten Einflüssen nehmen in der Humanbiologie einen breiten Raum ein. Wesentlichste Untersuchungsmethode ist dabei die Anthropometrie, die Menschenmessung. Im einfachsten Fall werden Körpermaße an den Probanden einer definierten Stichprobe mit einem standardisierten Instrumentarium bestimmt und statistisch ausgewertet, um zu allgemeingültigen Aussagen zu gelangen. Eine geeignete Archivierung der Daten und ein schneller Zugriff sind wesentlich, weil oft erst im Vergleich zu früher erhobenen Daten eine Kausalanalyse möglich ist. Am Rechenzentrum der Humboldt-Universität sind alle seit 1967 von mir und meinen Mitarbeitern erhobenen anthropometrischen Meßdaten archiviert. Die ersten DV-Auswertungen wurden an einem Zeiss Rechen-Automaten ZRA1 vorgenommen. Nach mehreren Zwischenstufen erfolgte sowohl im Rechenzentrum als auch im Institut für Humanbiologie der Übergang zu SAS Software. Die wesentlich verkürzten Rechenzeiten und die Vielfältigkeit des Systems erlauben seitdem ein deutlich höheres Auswertungsniveau. Auch bei der raschen Umsetzung von Forschungsergebnissen zur Präsentation in der Lehre hat sich SAS bewährt, teilweise direkt und teilweise durch die guten Exportmöglichkeiten in andere Grafiksysteme.

An einem klassischen anthropometrischen Beispiel - dem Zusammenhang zwischen der Körpermasse (Körpergewicht), der Körperhöhe (Körpergröße) und dem Alter im weiblichen Geschlecht - wird hier über Erfahrungen mit den SAS-Versionen 6.10, 6.11 und 6.12 auf SUN-Workstation und Pentium-PC (DESKPRO XL 5120) berichtet. Das Beispiel ist deshalb von breiterem Interesse, weil Übergewicht und Fettsucht in unserer Population ein durchaus ernst zu nehmendes Problem darstellen. Jährlich werden große Summen für ambulante Schlankheitsdiäten oder entsprechende kurative Behandlungen bzw. für die Behandlung fettsuchtbedingter Folgeerkrankungen ausgegeben, ohne daß dem Arzt oder dem betroffenen Patienten Tabellen zur Verfügung stehen, die auf wissenschaftlich fundierten und repräsentativen anthropometrischen Untersuchungen basieren und eine objektive Beurteilung der individuellen Körpermasse im Vergleich zu einem als gesündestes Gewicht definierten Optimalgewicht erlauben. Erste Schritte auf dem Weg zu bevölkerungsrelevanten Optimalgewichtstabellen sind statistische Analysen zur Klärung der Zusammenhänge zwischen Körpermasse, Geschlecht, Alter, Körperhöhe und Körperbautyp sowie die Erarbeitung spezifischer Normalgewichtstabellen. Hierbei hat sich in der Zusammenarbeit des Fachgebiets Humanbiologie der Universität Potsdam mit dem Rechenzentrum der Humboldt Universität zu Berlin die Nutzung von SAS bewährt, über die hier zu berichten ist.

Als Ausgangsmaterial für die statistischen Analysen dienten die anthropometrischen Daten einer geschichteten Stichprobe von 17.802 Mädchen und Frauen aus allen neuen Bundesländern im Altersbereich von der Geburt bis zu 65 Jahren. Die Datensätze entsprachen den untersuchten Probanden. Je Datensatz waren mindestens 81 Merkmale dokumentiert, davon mindestens 34 Körpermaße. Zur Anwendung kamen

Rechenoperationen mit SAS-BASE (proc univariate) und SAS-STAT, Grafik mit SAS-GRAPH (proc gplot, proc g3d, proc gedit), Grafik mit SAS-INSIGHT (scatter-plot), Tabellen mit SAS-FSP (proc fsvview) und Texte mit SAS-GRAPH (proc gslide). Mit SAS-INSIGHT erstellte Scatter-Plots erwiesen sich neben anderen Prüfverfahren als besonders geeignet zur Beurteilung und gegebenenfalls Revision von Outsider-Werten.

Die Körpermasse verändert sich altersabhängig. Sie nimmt im weiblichen Geschlecht im Durchschnitt von der Geburt bis zu etwa 52 Jahren mit unterschiedlicher Geschwindigkeit zu. Erst nach Beendigung des fortpflanzungsfähigen Lebensalters beginnt mit etwa 52 Jahren bei Frauen eine langsame Massenabnahme, die bis zum Lebensende andauert. Im Unterschied zur Körpermasse ist das Wachstum der Körperhöhe im weiblichen Geschlecht bereits mit etwa 18 bis 20 Jahren beendet und mit 40 bis 45 Jahren beginnt eine altersbedingte Körperhöhenreduktion, weil die Zwischenwirbelscheiben sich verändern. Abb. 1 zeigt anhand der Medianwerte diesen unterschiedlichen ontogenetischen Verlauf der beiden Körpermaße.



Das Diagramm dokumentiert diese Abnahme der mittleren Körperhöhe mit zunehmendem Lebensalter bereits von 20 Jahren an. Dies ist ein Effekt der säkularen Akzeleration, der für unsere Bevölkerung bei jeder anthropologischen Querschnittstudie auftritt. Bei der Datenaufnahme wurden unterschiedlich alte Probandinnen relativ gleichzeitig gemessen. Da die durchschnittliche Körperhöhe in

### Körperhöhen und altersspezifische Körpermasse, weiblich

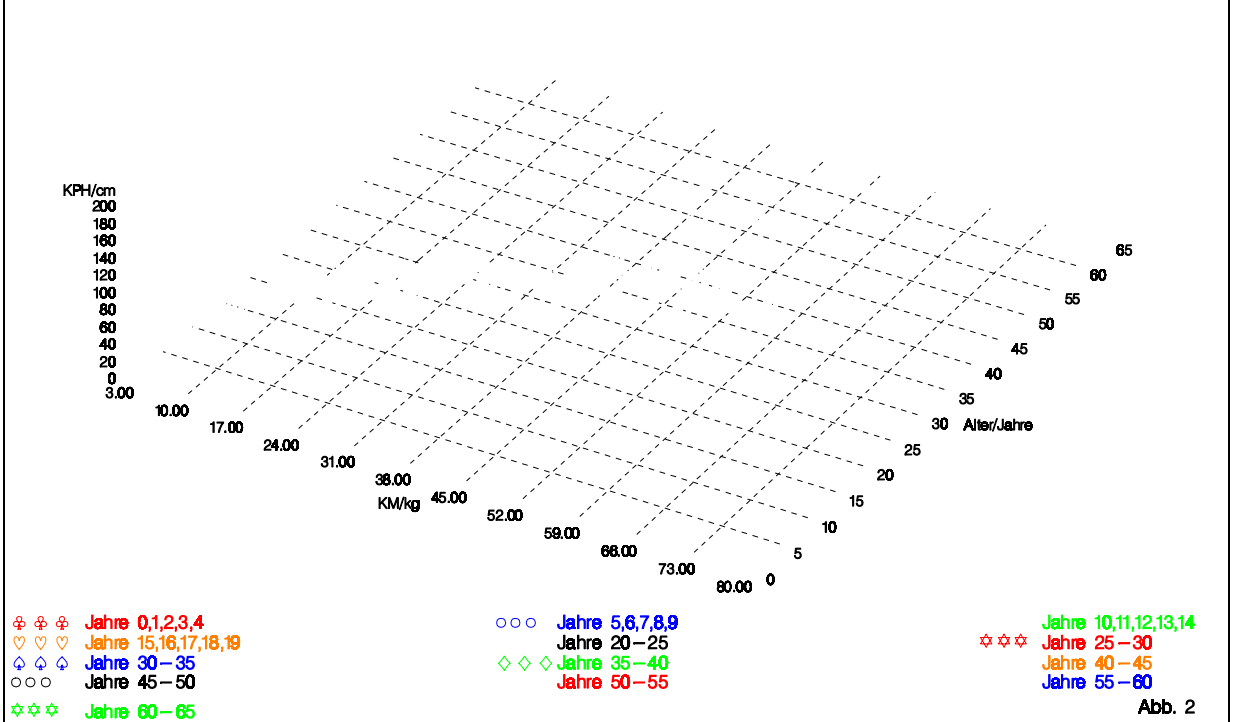


Abb. 2

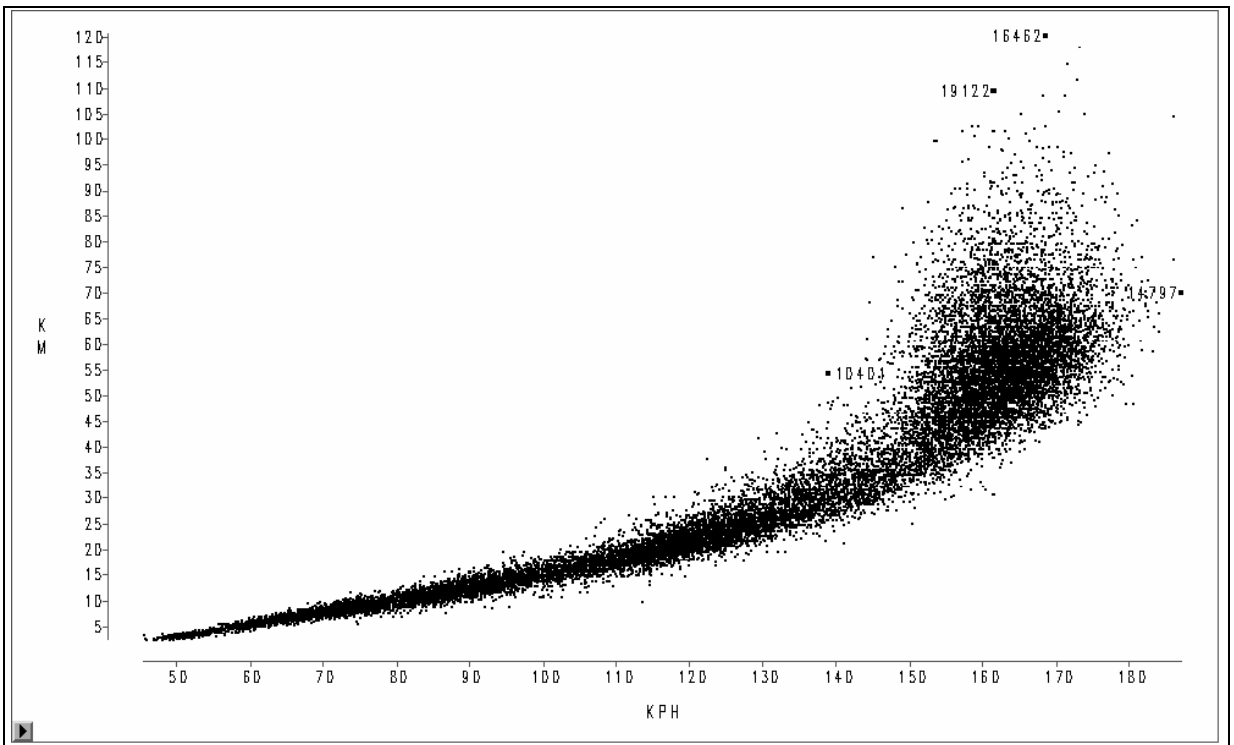


Abb3

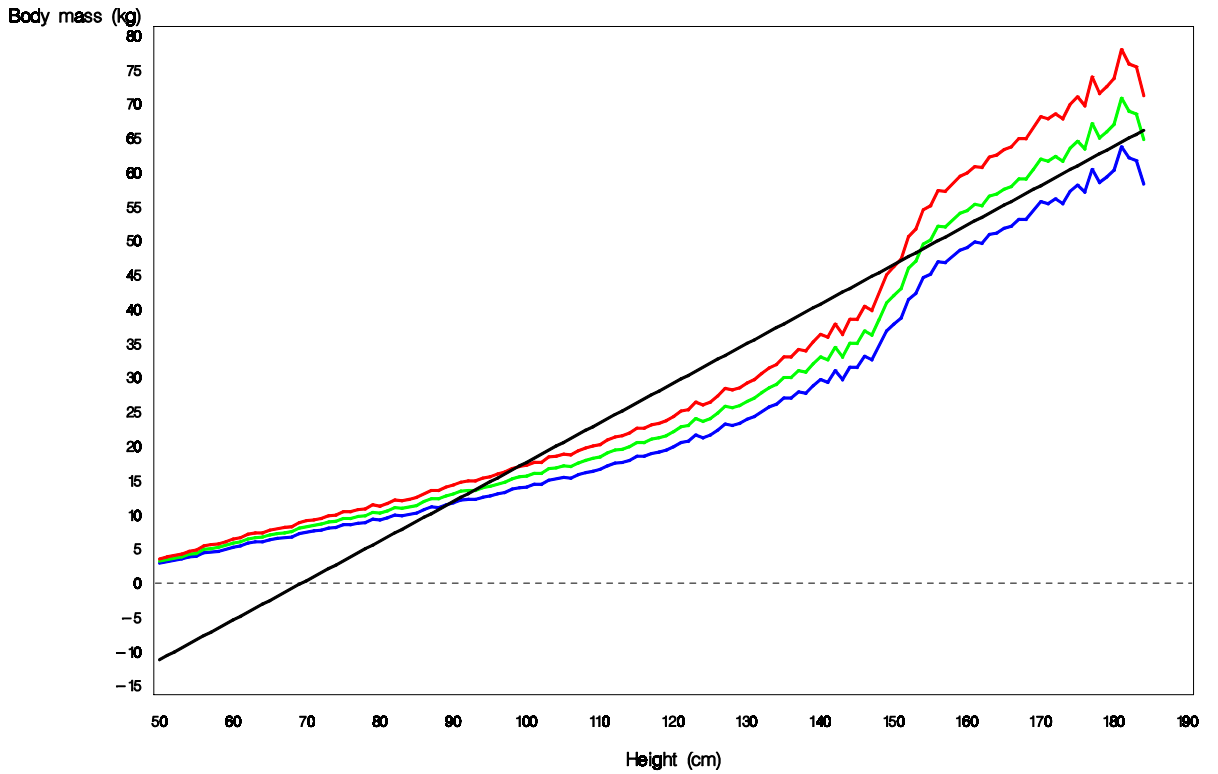
den Industrieländern seit mehr als 100 Jahren zugenommen hat, waren die heute älteren Frauen zur Zeit ihrer größten individuellen Körperhöhe in ihrem jungen Erwachsenenalter im Durchschnitt kleiner als junge Frauen heute. Dies ergibt in der Querschnittsanalyse einen Summationseffekt zwischen individueller Altersschrumpfung und säkularer Akzeleration.

Innerhalb eines jeden Altersbereichs ist die Körpermasse von der Körperhöhe abhängig. Im 3D-Needle-Plot (Abb.2) läßt sich diese Abhängigkeit durch gezielte Farbauswahl didaktisch gut darstellen. Deutlich sichtbar werden die breiten Überlappungsbereiche der Körpermasse sowohl in den Körperhöhengruppen als auch in den Altersgruppen. Die 2D-Darstellung der Körpermasse zur Körperhöhe in Form eines Scatterplots vermittelt durch die Projektion aller individuellen

Körperhöhen-Körpermassen-Beziehungen ebenfalls einen guten Eindruck von der Beziehung beider Merkmale zueinander und von der körperhöhenabhängigen Variabilität der Körpermasse (Abb. 3). Darüber hinaus besteht bei dieser Projektionsmethode die Möglichkeit, auffällige Datenpunkte anzuklicken und den gesamten Datensatz abzufragen. So läßt sich bei allen Extremwerten leicht überprüfen, ob es sich um einen Beobachtungs- oder Dokumentationsfehler handelt oder um real vorkommende Werte.

Um praxiswirksam zu werden, muß die im Scatterplot gezeigte Variabilität auf körperhöhenabhängige Normalgewichte zurückgeführt werden. Dabei ist unter Normalgewicht der jeweils gruppenspezifische Medianwert der körperhöhenabhängigen Körpermasse zu verstehen. Für eine Darstellung dieses körperhöhenabhängigen Normalgewichts bieten sich die beiden in Abb. 4 wiedergegebenen Vorgehensweisen an. Man kann eine lineare Regression zwischen Körperhöhe und Körpermasse bilden oder man kann die Körperhöhe äquidistant skalieren, beispielsweise in Zentimeterabschnitte, und für jede entstehende Körperhöhengruppe den Medianwert der Körpermasse bilden. Sind anstelle von körperhöhen-spezifischen Normalgewichten Normalgewichtsbereiche erwünscht, so lassen sich diese in beliebiger Breite prozentual bilden. In Abb. 4 wurde ein Bereich von  $P 50 \pm 10 \%$  gewählt. Möglich wäre auch ein Bereich von beispielsweise P40 bis P60. Die Wahl von Art und Breite der Normalgewichtsbereiche ist kein statistisches Problem. Sie muß fachspezifisch humanbiologisch entschieden werden. Die durch derartige Gruppierungen erhaltenen Normalgewichtsbereiche vermitteln einen guten Eindruck von der Realität. Es ergeben sich jedoch ohne Berücksichtigung des Alters besonders in den höheren Körperhöhengruppen durch das gemeinsame Vorkommen von hoch-schlankwüchsigen Kindern und kleinen bis mittelgroßen korpulenten Erwachsenen in ein und derselben Körperhöhengruppe sehr unregelmäßige Perzentilkurven, wie dies aus Abb. 4 ersichtlich ist. Die Einführung des Alters als zusätzliche Variable ist demnach auch bei dieser Auswertungsmethode unumgänglich. Bei der Bildung einer linearen Regression zwischen Körperhöhe und Körpermasse ohne Berücksichtigung des Alters wird die Unsinnigkeit Vorgehensweise auch biologische besonders deutlich. Im Bereich der unteren Grenze der Körperhöhenkala ist das Regressionmodell kaum verwendbar. Im Körperhöhenbereich um 70 cm geht das Regressionsgewicht gegen Null, und noch kleineren Babies wird sogar eine negative Körpermasse zugewiesen.

### Height – specific body mass, female



$P50 \cdot 0.9$   $P50$   $P50 \cdot 1.1$   
Regression line body mass by height

Abb. 4

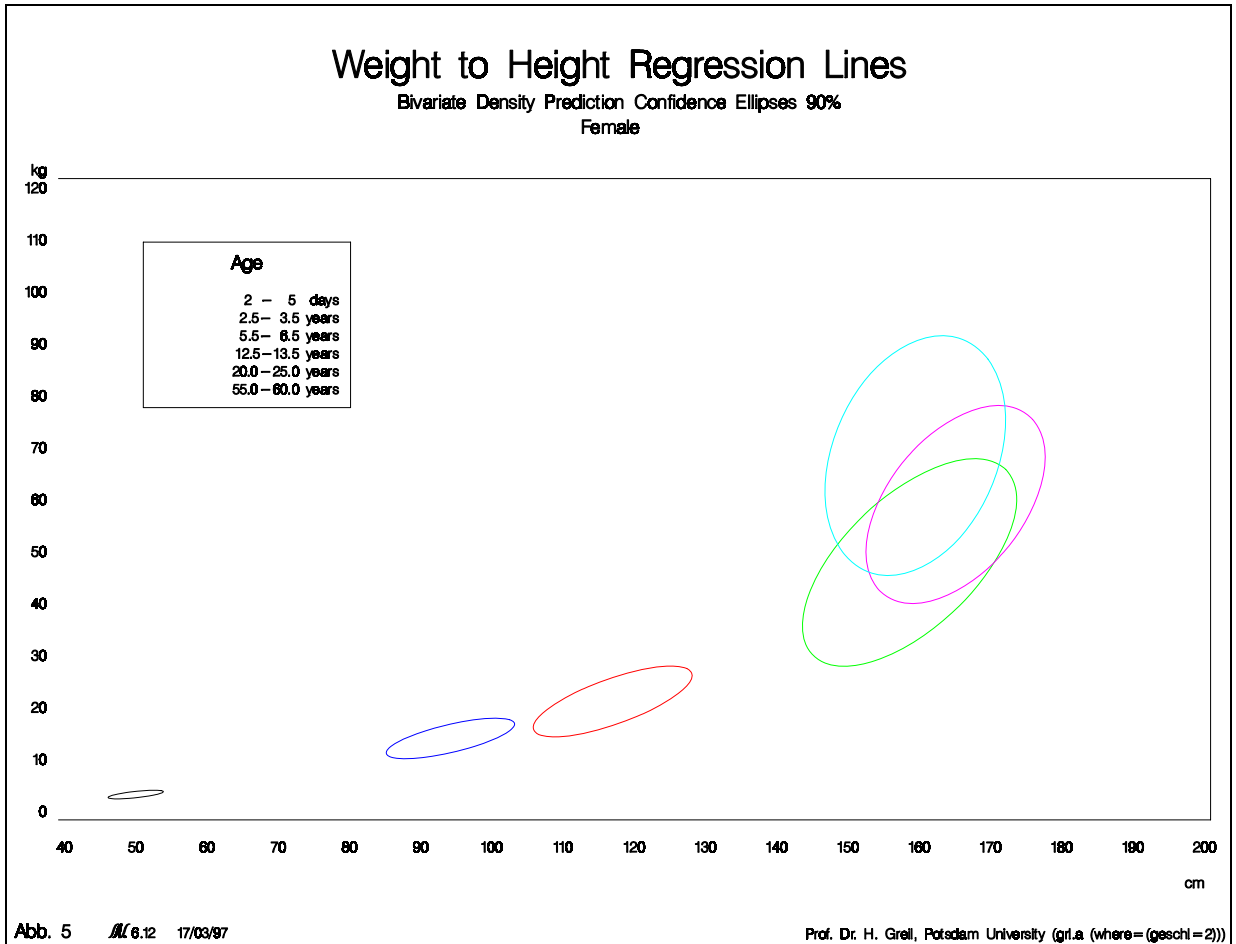


Abb. 5 *ML* 6.12 17/03/97

Durch die in Abb. 4 mit SAS-GRAPH (proc gplot) gegebene Darstellungsweise lässt sich die Forderung nach altersspezifischen Regressionen nachdrücklich demonstrieren. Eine Auswahl alterspezifischer Regressionsgeraden mit bivariaten Konfidenzellipsen mit einem Vertrauensbereich von 90 % verdeutlicht in Abb. 5 die Zusammenhänge zwischen den 3 Variablen Körpermasse, Körperhöhe und Alter. Während des Wachstumsalters nimmt der Steigungswinkel der Regressionsgeraden ständig zu, und auch im Erwachsenenalter bestehen deutliche Unterschiede zwischen den Gruppen der jüngeren und der älteren Frauen.

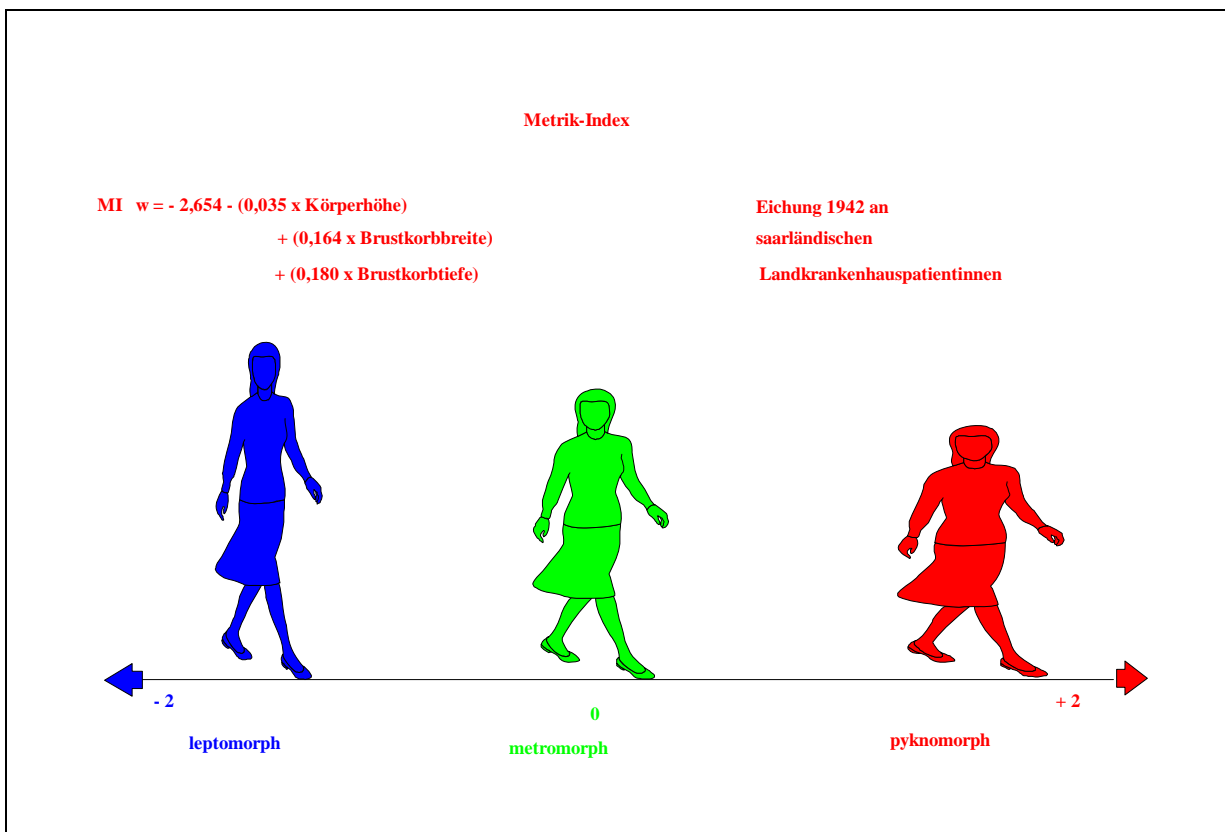


Abb. 6

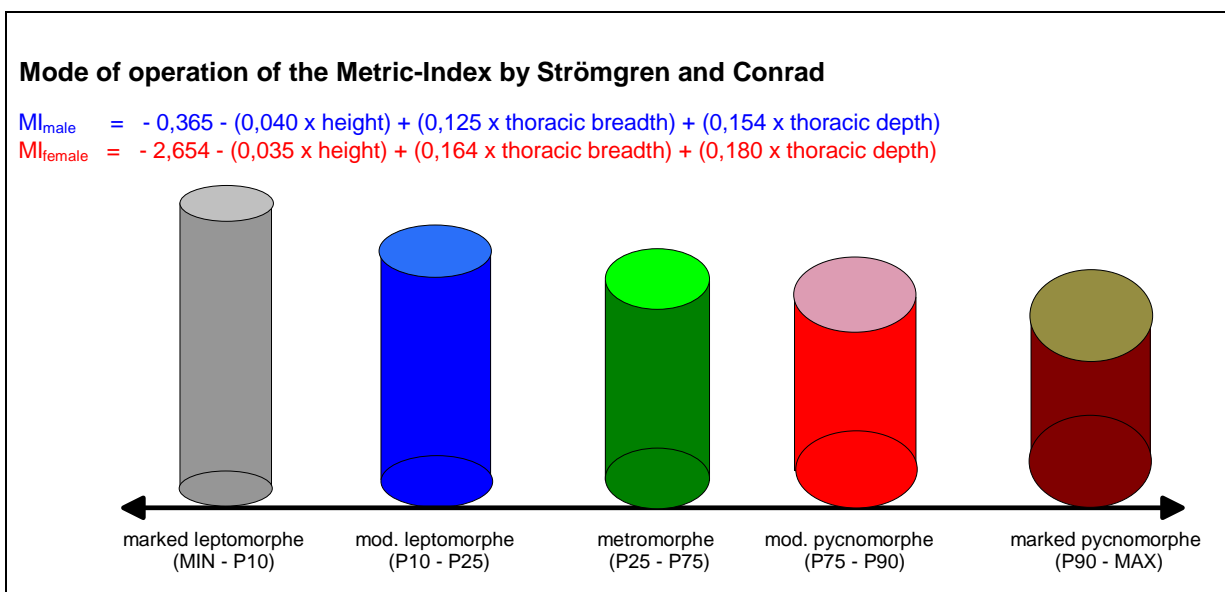


Abb. 7

Für eine gute Annäherung an eine optimale Körpermasse ist die Einführung des Körperbautyps als genetisch fundierter Körperstruktur mit Hilfe einer weiteren Variablen zusätzlich zu Körpermasse, Körperhöhe und Alter empfehlenswert. Hierzu gibt es unterschiedliche Verfahren. In den USA ist eine Einteilung des Körperbautyps nach der Skelettrobstizität üblich, gemessen am Verhältnis der Ellenbogenbreite zur Körperhöhe. Die deutsche Schule bevorzugt den sogenannten Metrik-Index, eine lineare Funktion der ebenfalls durch das Skelett geprägten Maße Körperhöhe, Brustkorbbreite und Brustkorbtiefe (Abb. 6). Nach definierten Perzentilbereichen des Metrik-Index läßt sich eine bipolare Variationsreihe der Körperform mit einer Spannweite von einem niedrig-breitwüchsigen pyknomorphen zu einem hochschlankwüchsigen leptomorphen Pol in beliebig viele Körperbautypen verteilungsunabhängig skalieren. Die mittlere Körperbauform wird bei dieser Typologie als metromorph bezeichnet. Praktikabel ist eine Einteilung der gesamten Spannweite in 3 Typen mit Hilfe der Metrik-Index-Perzentile 20 und 80 ( $> P20 = \text{leptomorph}$ ,  $P20 - P80 = \text{metromorph}$ ,  $< P80 = \text{pyknomorph}$ , s. Abb. 6) oder eine Einteilung in 5 Typen mit Hilfe der Perzentile 10, 25, 75 und 90 (Abb. 7)

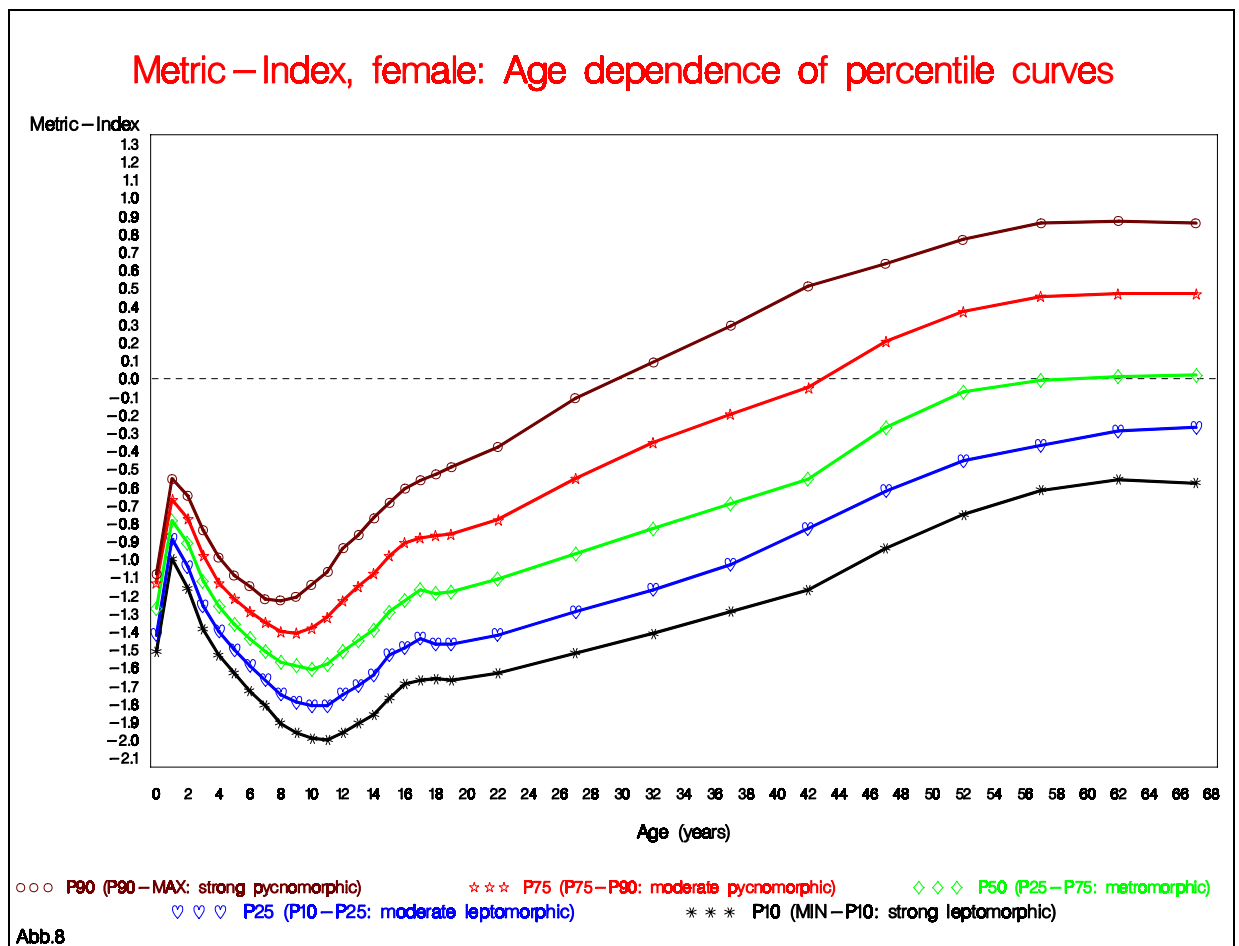
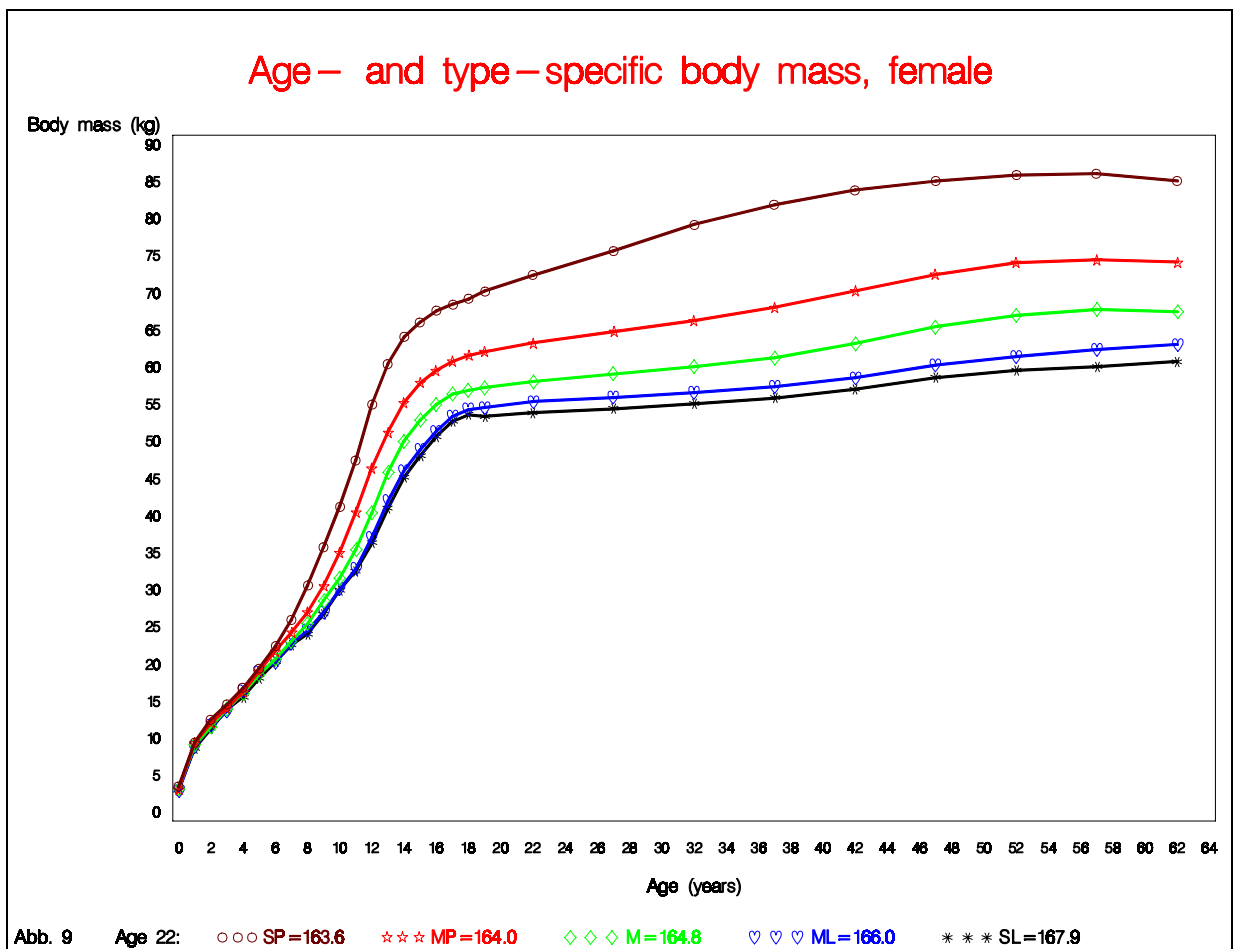


Abb. 8 zeigt die Variabilität und die Altersabhängigkeit des Metrik-Index bei einer Einteilung nach 5 Typen. Dabei geben die Tiefstwerte des Index zu Beginn der Pubertät die Zeit des typspezifisch jeweils relativ am stärksten ausgeprägten Schlankwuchses an. Dieses Stadium der körperlichen Entwicklung wird körperbautypspezifisch von den akzeleriert reifenden ausgeprägt pyknomorphen Mädchen mit 8 Jahren am frühesten erreicht und von den langsamer reifenden ausgeprägt leptomorphen Mädchen mit 11 Jahren am spätesten. Diese in

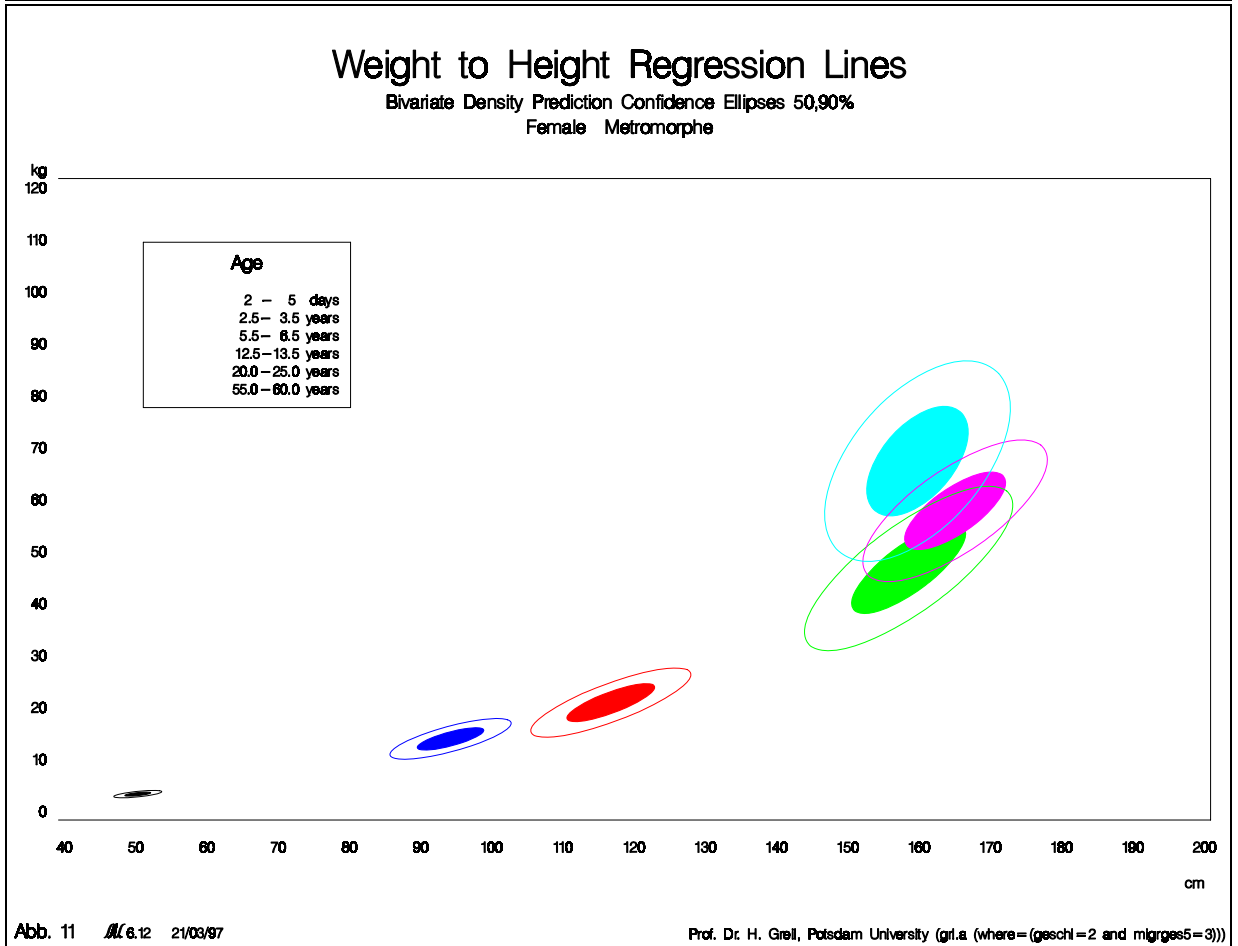
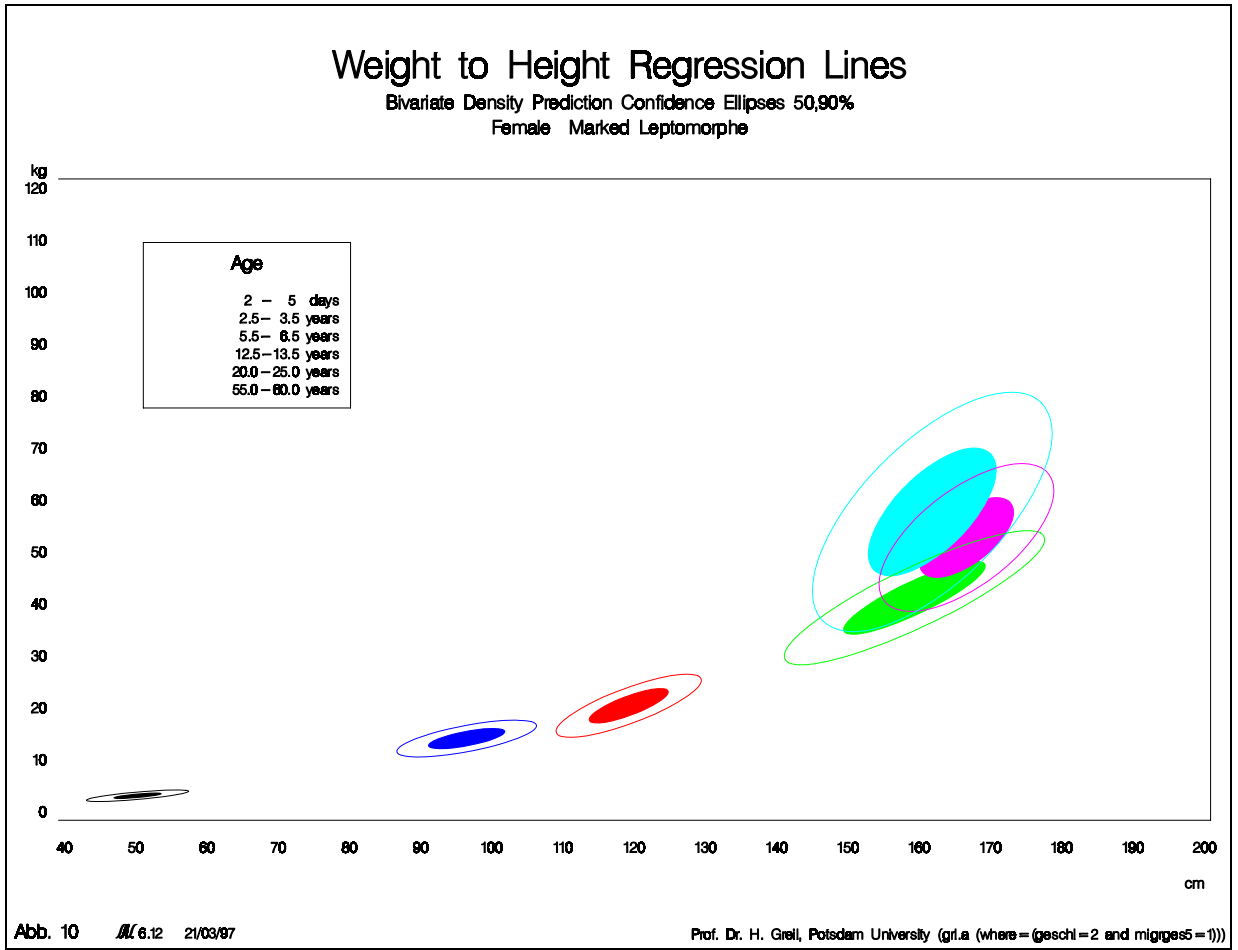
Abhängigkeit vom Körperbautyp unterschiedliche Geschwindigkeit der körperlichen Entwicklung wirkt sich auch auf die alters- und körperhöhen-spezifische Körpermasse aus.

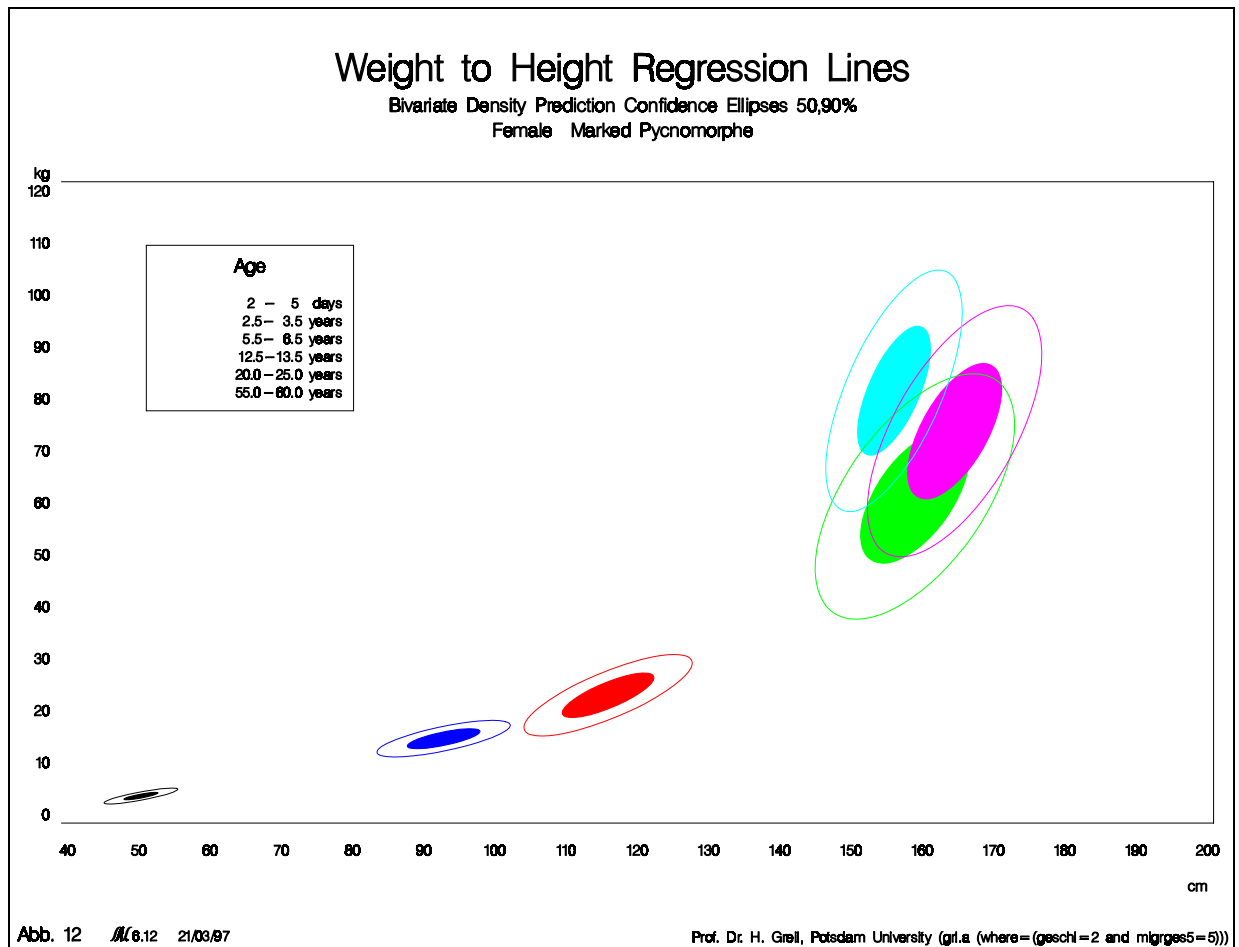


In Abb. 9 sind die Mittelwertskurven der körperbautyps-spezifischen altersabhängigen Körpermasse von der Geburt bis zu 65 Jahren dargestellt. Im Typenvergleich zeigt sich die deutlich schiefe Verteilung der Körpermasse. Auf diese Weise wird eine weitaus differenziertere Einschätzung von echtem und vermeintlichem Übergewicht möglich als dies ohne Berücksichtigung des Körperbautyps möglich ist. Wie in Abb. 1 ohne Berücksichtigung des Körperbautyps gezeigt, läßt sich auch unter seiner Einbeziehung die Körpermasse körperhöhen-spezifisch gruppieren.

Darüber hinaus besteht die Möglichkeit, alters- und körperbautyps-spezifische lineare Regressionen zwischen Körperhöhe und Körpermasse zu bilden. Entsprechende Diagramme unter Einbeziehung bivariater Konfidenzellipsen sind in Abb. 10 für stark leptomorphe, in Abb. 11 für metromorphe und in Abb. 12 für stark pyknomorphe Mädchen und Frauen dargestellt. Durch die Verwendung einer einheitlichen Achsen-skalierung kann demonstriert werden, wie die Steigung der Regressiongeraden sowohl mit dem Alter als auch vom hoch-schlankwüchsigen zum niedrig-breitwüchsigen Körperbautyp zunimmt.







Zusammenfassend möchten wir feststellen, daß SAS aufgrund seiner vielfältigen und nutzerfreundlichen Programmierungsmöglichkeiten hervorragend zur Anwendung bei humanbiologischen Auswertungen geeignet ist. Auch problembezogene Diagramme für eine unmittelbare und didaktisch individuell gestaltete Umsetzung neuester Forschungsergebnisse in die Lehre in Form von Overheadfolien bereitet keine Schwierigkeiten. Etwas erschwerend wirkt sich allerdings die Tatsache aus, daß bei der Herstellung von Farbdigrammen sowohl bei der Linienstärke als auch bei der Farbgestaltung gegenüber anderen Systemen, wie beispielsweise Excel, ein beträchtlicher Unterschied zwischen dem Monitorbild und dem Druckbild bestehen.