A Trivariate Distribution for the Height, Weight, and Fat of Adult Men

David E. Burmaster
Alceon Corporation
PO Box 382669 Harvard Square Station
Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954
deb@Alceon.com

Donald M. Murray
Energy and Environmental Research
87 Fayerweather Street
Cambridge, MA 02138
tel: 617-868-3018

Keywords

body fat, body weight, height, trivariate Normal distribution, PBPK models

Abstract

Using exploratory data analysis, probability plots, scatterplots, and computer animations to rotate and visualize the data, we fit a trivariate Normal distribution to data for the height, the natural logarithm of body weight, and the body fat for 646 men between the ages of 50 and 80 years as reported by the medical staff of the US Veterans Administration's "Normative Aging Study" in Boston, MA. Although these data do not include any children, women, or young men, the measurements represent the best data that we could find through a 4-year search. We believe that these data are well measured and reliable for men in the specified age range and that these data reveal an interesting statistical pattern for use in probabilistic PBPK models.

Introduction and Data

Risk assessors and toxicologists often use physiologically-based pharmacokinetic (PBPK) models to analyze, understand, and predict the temporal relationships of the tissue concentrations of xenobiotic compounds in the human body (NRC, 1987). Such models can account for physiological changes (such as changes in blood flow or renal clearance for an individual subject), interspecies and intersubject differences, nonlinear metabolic kinetics, and difference in uptake mechanisms (Krewski et al, 1987). PBPK models typically group individual organs with similar blood flow, diffusion, and permeability properties into single compartments. In particular, many PBPK models group adipose tissue (fat) and bone marrow into a single group often called the "fat group" (Krewski et al, 1987).

 © *Alceon* ®, 1997

Some standard references (e.g., Snyder et al, 1984; Lentner, 1981) contain information on the body fat and lean body mass of children and adults. For example, Snyder et al. (1984) state that the male and female "Reference Adult" have 13.5 kg and 16.0 kg, respectively, of total body fat (a combination of nonessential and essential fat; page 43). More generally, in their Figure 33 on page 43, Snyder et al. show temporal trends for (typical) body fat as a function of postnatal age in males and females. This figure shows that a typical man between the ages of 35 and 60 years has a constant average body fat of ~26 percent of body weight.

Through an on-line literature search and discussions with knowledgeable people, we searched for data that could help us develop probability distributions for the body fat of adult men and women for use in probabilistic PBPK models. The medical staff of the US Veterans Administration's (VA) "Normative Aging Study" in Boston, MA, kindly gave us an anonymous dataset collected in 1993 from this long-term longitudinal cohort study that includes measurements of (i) wall height (Ht; nearest 0.1 inch); (ii) body weight (Wt; nearest integer pound); and (iii) body fat (F; nearest integer percent (pct) of total body weight) for 646 men, 50 yr ≤ age < 80 yr. Each man in this cohort served in a branch of the US Military at some time and was chosen, at random, for this study of "normal aging" so that the results could be applied to the general (healthy) population. The VA's medical staff measured the wall heights in the common fashion, body weights with a traditional beam-scale, and body fat with an electronic instrument [EndNote 1]. Before beginning the statistical analyses of these data, we converted the measurements of Ht to centimeters (cm) and the measurements of Wt to kilograms (kg).

Although these data from the "Normative Aging Study" do not include any children, women, or young men, these measurements from the Study represent the best data that we could find through telephone calls, on-line literature searches, and Internet inquiries spanning ~4 years. Until new research becomes available, we believe that these data are well measured and reliable for men in the specified age range and that these data reveal an interesting statistical pattern for use in probabilistic PBPK models [EndNote 2].

Statistical Methods and Results

We completed this analysis in 5 basic steps. First, we used the tools of exploratory data analysis to look for patterns in the data, to consider possible transformations of the data,

 © *Alceon* ®, 1997

and to prepare summary statistics for the data. Second, we examined scatterplots for the three main variables as a function of age. Third, we examined the three probability plots for the univariate marginal distributions of the data, and then we examined the pairwise bivariate scatterplots for the data. Finally, we analyzed the full trivariate structure of the data.

First, based on the ideas of exploratory data analysis (Tukey, 1977), we transformed the data for Wt using the natural logarithm function (see also: Brainard & Burmaster, 1992; Burmaster & Crouch, 1997), but we did not transform the data for either Ht or F. The top section in Table 1 shows key summary statistics for the 646 measurements of Ht (cm), ln[Wt (kg)] and F (pct), including the arithmetic mean (AMean) and standard deviation (StdDev).

Second, based on the scatterplots of Ht, ln[Wt], and F as a function of age in Figures 1A, 1B, and 1C, respectively, we concluded that age is a such a weak explanatory variable in these data for 646 men that we do not consider it further. In Figures 1A, 1B, and 1C, the loess regression (Cleveland, 1993; Cleveland, 1994; Systat, 1992) reveals weak downward trends in Ht, ln[Wt], and F with increasing age, but the trends are small when compared to the overall scatter in the data [EndNote 3]. In each case, the trend accounts for less than a 5 percent change in the data across the ages, and in each case, the scatter in the data exceeds 15 to 20 percent across the ages. In Figures 1A, 1B, and 1C, some points plotted in the graphs represent multiple occurrences of the same values; we kept these repeated values in all calculations because they represent different individuals. In Figures 1A, 1B, and 1C, some data points align in rows and/or columns (or overlap) because the VA's medical staff reported age and percent fat as integers.

Third, based on the Normal probability plots of Ht, ln[Wt] and F in Figures 2A, 2B, and 2C, respectively, we concluded that a Normal (or Gaussian) distribution fits each of these three univariate marginal distributions well (D'Agostino & Stephens, 1986; Burmaster & Hull, 1997). In Figure 2A for Ht, the straight line fit by ordinary least squares (OLS) regression has an intercept equal to 174.204 and a slope equal to 6.499. The adjusted R^2 (adjR^2) for this straight line equals 0.997, indicating an excellent fit to the data. In Figure 2B for ln[Wt], the straight line fit by OLS regression has an intercept equal to 4.410 and a slope equal to 0.144. The adjR^2 for this straight line equals 0.996, also indicating an excellent fit to the transformed data. In Figure 2C for F, the straight line fit by OLS regression has an intercept equal to 21.050 and a slope equal to 4.984.

The adjR^2 for this straight line equals 0.994, again indicating an excellent fit to the data. As expected, the intercepts and slopes of these fitted lines correspond closely to the means and standard deviations of the (transformed) data (Table 1, top section). In Figure 2C, many of the data points follow a staircase pattern because the VA's medical staff rounded all measurements of F to the nearest integer percent. In Figures 2A, 2B, and 2C, some points in the graphs represent multiple occurrences of the same values; we kept these repeated values in all calculations because they represent different individuals.

Fourth, based on the bivariate scatterplots shown as a scatterplot matrix (SPLOM; Systat, 1992) in Figure 3, we concluded that correlated bivariate Normal distributions give a good fit to each of these bivariate marginal distributions (Keeping, 1995). In each panel in the SPLOM, the ellipse contains 90 percent (chosen for visual reasons) of the probability mass for the bivariate distribution fit to the data. While we did not formally test these bivariate marginal distributions for normality, the visual patterns in the SPLOM strongly support the inference (Hutchinson & Lai, 1990). Again, some points in the SPLOM represent multiple occurrences of the same values; we kept these repeated values in all calculations because they represent different individuals. Also, some points in the panels align in rows or columns (or overlap) because the medical staff rounded some measurements to the nearest integer.

Fifth, based on the 3D scatterplot shown in Figure 4 and on computer animations (Wolfram, 1991; Wickham-Jones, 1994), we concluded that a correlated trivariate Normal distribution gives a good fit to the 646 measurements for Ht, ln[Wt], and F. Figure 4 shows the trivariate scatterplot and the three bivariate marginal scatterplots for the data projected orthogonally on to the "walls" of the bounding box. Using computer animation, we earlier examined these data in continuous and stepped 3D rotation (Pickover & Tewksbury, 1994). The human eye can see anomalies in patterns far better than any statistical test yet devised (Tufte, 1983; Tufte, 1990; Wolff & Yaeger, 1993). Since we did not see any anomalous voids or clusters in the rotating or stepped 3D scatterplots, we concluded that a correlated trivariate Normal distribution provides a good fit to the data for Ht, ln[Wt], and F. In particular, the top section of Table 1 provides the arithmetic means and standard deviations for Ht, ln[Wt], and F, and the second and third sections of Table 1 provide the Pearson (product-moment) correlation coefficients and the Variance-Covariance matrix to complete the specification of the trivariate Normal distribution (Anderson, 1958; Evans et al, 1993; Rose & Smith, 1996). We also

 © *Alceon* ®, 1997

calculated the Spearman (rank) correlation coefficients (not shown here) and found that they differ from the Pearson correlation coefficients by an inconsequential tolerance.

Discussion, Conclusions, and Limitations

First, we believe that these data are well measured and reliable for men in the specified age range and that these data reveal an interesting statistical pattern for use in probabilistic PBPK models. Considering the additional information in Snyder et al. (1984) that these physiological variables change little from age 35 years and older , we believe it is appropriate to extend the applicable age range down to 35 years of age.

Second, univariate Normal distributions provide excellent fits ($adjR^2 \geq 0.994$) to the univariate marginal data for Ht, ln[Wt], and F [EndNote 4]. Based on the visual patterns, we conclude that bivariate Normal distributions provide strong fits to the bivariate marginal data. Because the univariate and bivariate analyses strongly support the inference, and because dynamic and stepped computer animations reveal no visual anomalies in the 3D scatterplot, we believe that a trivariate Normal distribution (parameterized by the values in Table 1) characterizes these data reasonably well.

Third, these results for Ht and ln[Wt] are consistent with earlier findings. Working with data for men in a broader age range (18 - 74 years), Brainard and Burmaster (1992) report a similar mean value for Ht, a somewhat larger mean value for ln[Wt], and a slightly smaller correlation coefficient between Ht and ln[Wt]. Burmaster and Crouch (1997) report a similar mean and similar standard deviation for ln[Wt] of men in the same age range (50 yr $\leq$ age < 80 yr).

Fourth, for routine PBPK modeling, we suggest that most researchers will not need a trivariate distribution since few include Ht in a PBPK model. In such a situation, the researcher can reduce the dimensionality of the results in Table 1 to a bivariate Normal distribution for ln[Wt] and F by using the corresponding 2×1 AMean vector and corresponding 2×2 Variance-Covariance matrix as parameters in the bivariate Normal distribution (Hutchinson & Lai, 1990). In the sense of Smith et al. (1992), the correlation between Ht and F is negligible.

Finally, we note that this dataset contains no information on the distributions of Ht, ln[Wt] or F for girls, boys, women, or young men.

 © *Alceon* ®, 1997

Acknowledgments

We thank Pantel S. Vokonas, MD, and Kenneth Ward of the VA's "Normative Aging Study" in Boston, MA, for providing the data, and we thank Russell Stinebring for many helpful introductions and suggestions throughout this research. We thank Richard M. Dudley (Professor of Mathematics (Statistics) at the Massachusetts Institute of Technology) and Adolfo J. Quiroz (Professor of Computer Science and Statistics at the Simon Bolivar University) for help with goodness of fit tests for the multivariate Normal distribution. We also thank Andrew M. Wilson and three anonymous reviewers for many suggestions. Alceon Corporation funded this research.

Dedication

We dedicate this in memory of Margaret L.A. MacVicar.

Trademarks

Mathematica ® is a registered trademark of Wolfram Research, Inc: http://www.wri.com
Systat ® is a registered trademark of SPSS, Inc: http://www.spss.com
Alceon ® is a registered trademark of Alceon Corporation: http://www.alceon.com

EndNotes

1. According to the VA's medical staff, percent body fat was estimated by total body impedance using RJL model BIA-101A (RJL Systems, 33955 Harper Avenue, Clinton Twp, MI 48035), following instructions provided by the manufacturer. The measurement was taken with the participant lying relaxed on an exam table, using tetra polar electrode placement, with electrodes placed on the dorsal surfaces of the right hand and foot, at the distal metacarpals and metatarsals, respectively, and between the distal prominences of the radius and the ulna at the wrist, and of the medial and lateral malleoli at the ankle. An excitation current of 800 μA at 50 KHz is introduced into the participant at the distal electrodes of the hand and foot. Resistance is measured by detection of voltage drops at the proximal electrodes. Total body water is estimated with equations provided by the manufacturer which relate height, weight, and resistance. Total body water is proportional to fat free mass. Percent body fat is calculated by subtracting fat free mass from total body mass. Equations are similar to those reported by Lukaski and colleagues (Lukaski, H.C., Johnson, P.E., et al., American Journal of Clinical Nutrition, 1985; 41:810 - 817; Lukaski, H.C., Bolonchuk W.W., et al., Journal of Applied Physiology, 1986; 60:1327 - 1332).

2. In the last 3 years, the NHANES program has collected some measurements similar to these on a broader cross-section of the US population, but the results will not become public in this format until 4Q1998 or 1999.

 © *Alceon* ®, 1997

3. Loess regression is a form of nonparametric regression (Cleveland, 1993; Cleveland, 1994; Systat, 1992) in which a computer fits a family of linear or quadratic splines to data.

4. If a calculation involves only one of the three random variables, then the univariate (marginal) distribution for that variable can be used alone. However, if a calculation involves more than one of the three random variables, then the correlation structure becomes important.

© *Alceon* ®, 1997

References

Anderson, 1958

Anderson, T.W., An Introduction to Multivariate Statistical Analysis, Wiley, New York, NY

Burmaster & Crouch, 1997

Burmaster, D.E. and E.A.C. Crouch, 1997, Lognormal Distributions for Body Weight as a Function of Age for Males and Females in the United States, 1976 - 1980, Risk Analysis, Volume 17, Number 4, pp 499 - 506

Brainard & Burmaster, 1992

Brainard, J. and D.E. Burmaster, 1992, Bivariate Distributions for Height and Weight of Men and Women in the United States, Risk Analysis, 1992, Volume 12, Number 2, pp 267 - 275

Burmaster & Hull, 1997

Burmaster, D.E. and D.A. Hull, 1997, Using LogNormal Distributions and LogNormal Probability Plots in Probabilistic Risk Assessment, Human and Ecological Risk Assessment, Volume 3, Number 4, pp 235 - 255

Cleveland, 1993

Cleveland, W.S., 1993, Visualizing Data, AT&T Bell Laboratories, Hobart Press, Summit, NJ

Cleveland, 1994

Cleveland, W.S., 1994, The Elements of Graphing Data, AT&T Bell Laboratories, Hobart Press, Summit, NJ

D'Agostino & Stephens, 1986

D'Agostino, R.B. and M.A. Stephens, 1986, Goodness-of-Fit Techniques, Marcel Dekker, New York, NY

Evans et al, 1993

Evans, M., N. Hastings, and B. Peacock, 1993, Statistical Distributions, Second Edition, John Wiley & Sons, New York, NY

Hutchinson & Lai, 1990

Hutchinson, T.P. and C.D. Lai, 1990, Continuous Bivariate Distributions, Emphasizing Applications, Rumsby Scientific Publishing, Adelaide, South Australia

Keeping, 1995

Keeping, E.S., 1995, Introduction to Statistical Inference, Dover, New York, NY

Krewski et al, 1987

Krewski, D., D.J. Murdoch, and J.R. Withey, 1987, The Application of Pharmacokinetic Data in Carcinogenic Risk Assessment, in Pharmacokinetics in Risk Assessment, Volume 8 of Drinking Water and Health, National Academy Press, Washington, DC, pp 441 - 468

Lentner, 1981

Lentner, C., Ed., 1981, Geigy Scientific Tables, Volume 1, Eighth Edition, Ciba-Geigy, Basle, CH

NRC, 1987

National Research Council, 1987, Pharmacokinetics in Risk Assessment, National Academy Press, Washington, DC

Pickover & Tewksbury, 1994

Pickover, C.A. and S. K. Tewksbury, 1994, Frontiers of Scientific Visualization, Wiley & Sons, New York, NY

Rose & Smith, 1996

Rose, C. and M.D. Smith, 1996, The Multivariate Normal Distribution, Mathematica Journal, Volume 6, pp 32 - 37, Winter 1996

 © *Alceon* ®, 1997

Smith et al, 1992
Smith, A.E., P.B. Ryan, and J.S. Evans, 1992, The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating Population Distribution of Risk, Risk Analysis, Volume 12, Number 4, pp 467 - 474, December 1992

Snyder et al, 1975
Snyder, W.S., M.J. Cook, E.S. Nasset, L.R. Karhausen, and G.P. Howells, 1975, Report of the Task Group on Reference Man, International Commission on Radiological Protection Number 23, Pergamon Press, Oxford, UK

Systat, 1992
Systat, Inc., 1992, Users Manual, Evanston, IL

Tufte, 1990
Tufte, E.R., 1990, Envisioning Information, Graphics Press, Cheshire, CT

Tufte, 1983
Tufte, E.R., 1983, The Visual Display of Quantitative Information, Graphics Press, Cheshire, CT

Tukey, 1977
Tukey, J.W., 1977, Exploratory Data Analysis, Addison-Wesley, Reading, MA

Wickham-Jones, 1994
Wickham-Jones, T., 1994, Mathematica Graphics, Techniques & Applications, Springer-Verlag, Telos, Santa Clara, CA

Wolff & Yaeger, 1993
Wolff, R.S. and L Yaeger, 1993, Visualization of Natural Phenomena, Telos: Springer Verlag, Santa Clara, CA

Wolfram, 1991
Wolfram, S., 1991, Mathematica®, A System for Doing Mathematics by Computer, Second Edition, Addison- Wesley, Redwood City, CA

 © *Alceon* ®, 1997

proofed

Table 1
Summary Statistics (n = 646)

Statistic	Ht (cm)	ln[Wt (kg)]	F (pct)
•••••	•••••	•••••	•••••
Minimum	151.130	3.965	7.000
AMean	174.204	4.411	21.050
StdDev	6.509	0.145	4.999
Maximum	195.580	5.002	34.000

Pearson Correlation Coefficients	Ht (cm)	ln[Wt (kg)]	F (pct)
•••••	•••••	•••••	•••••
Ht (cm)	1	0.486	0.074 §
ln[Wt (kg)]	0.486	1	0.613
F (pct)	0.074 §	0.613	1

Variance Covariance Matrix	Ht (cm)	ln[Wt (kg)]	F (pct)
•••••	•••••	•••••	•••••
Ht (cm)	42.364	0.458	2.401 §
ln[Wt (kg)]	0.458	0.021	0.444
F (pct)	2.401 §	0.444	24.990

§ statistically indistinguishable from zero at p = 0.05

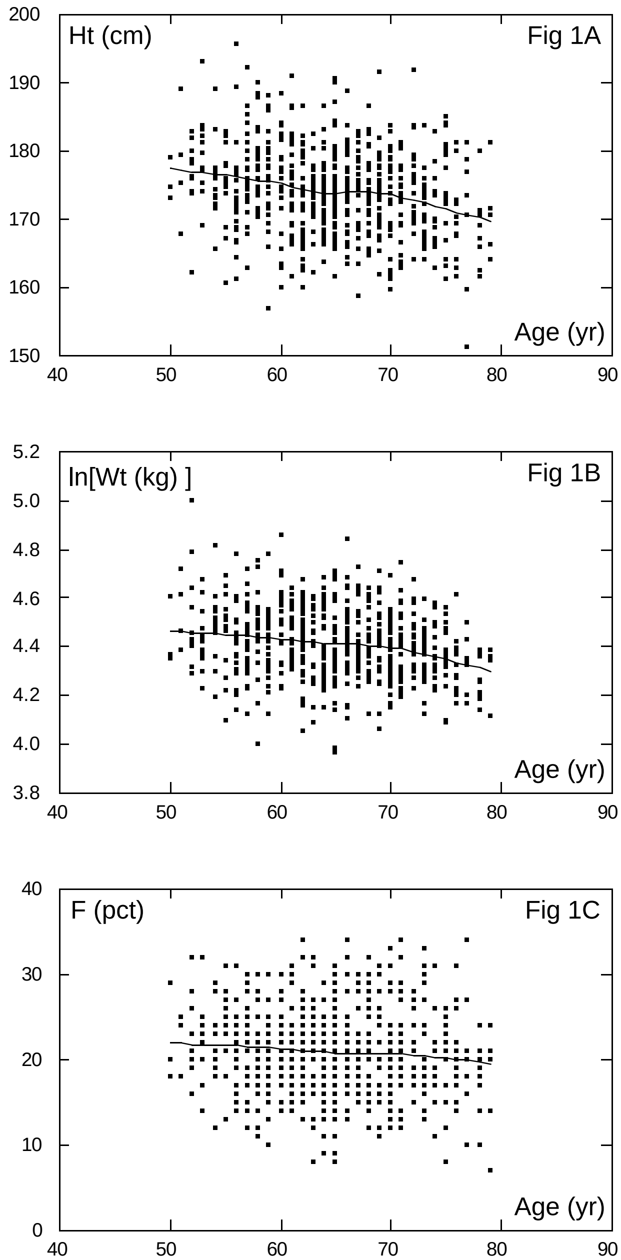

Figure 1
Scatterplots of Ht, ln[Wt], and F as a Function of Age
with splines fit by loess regression (n = 646)

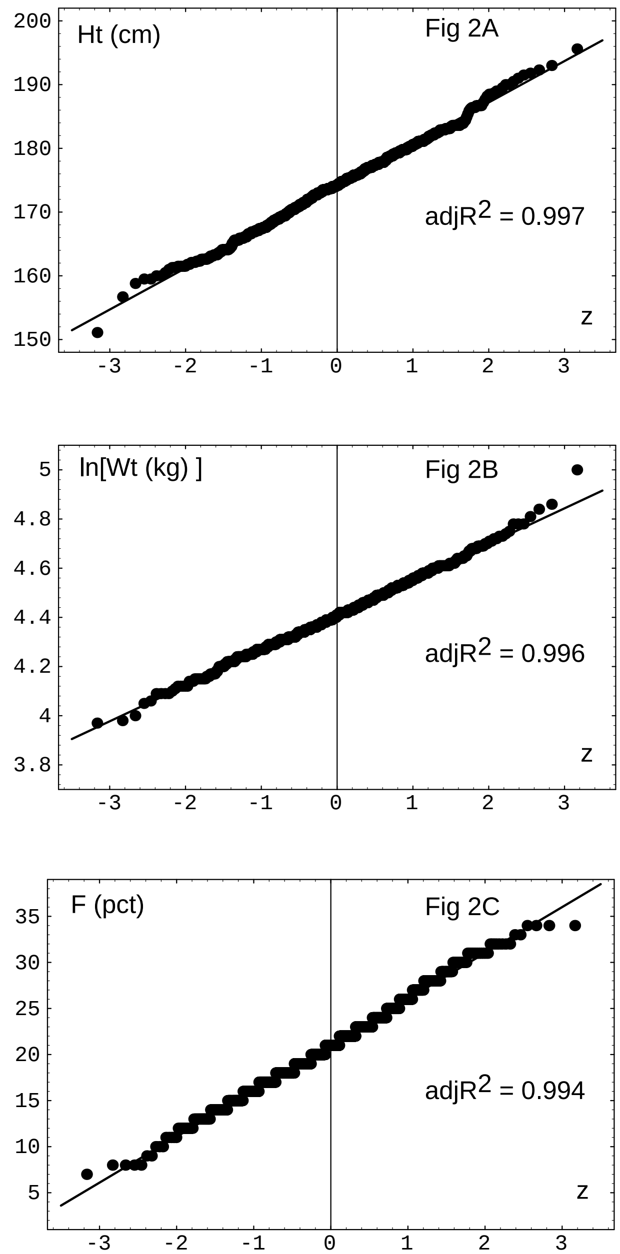

Figure 2
Probability Plots for Univariate Marginal Distributions
with straight lines fit by OLS regression (n = 646)

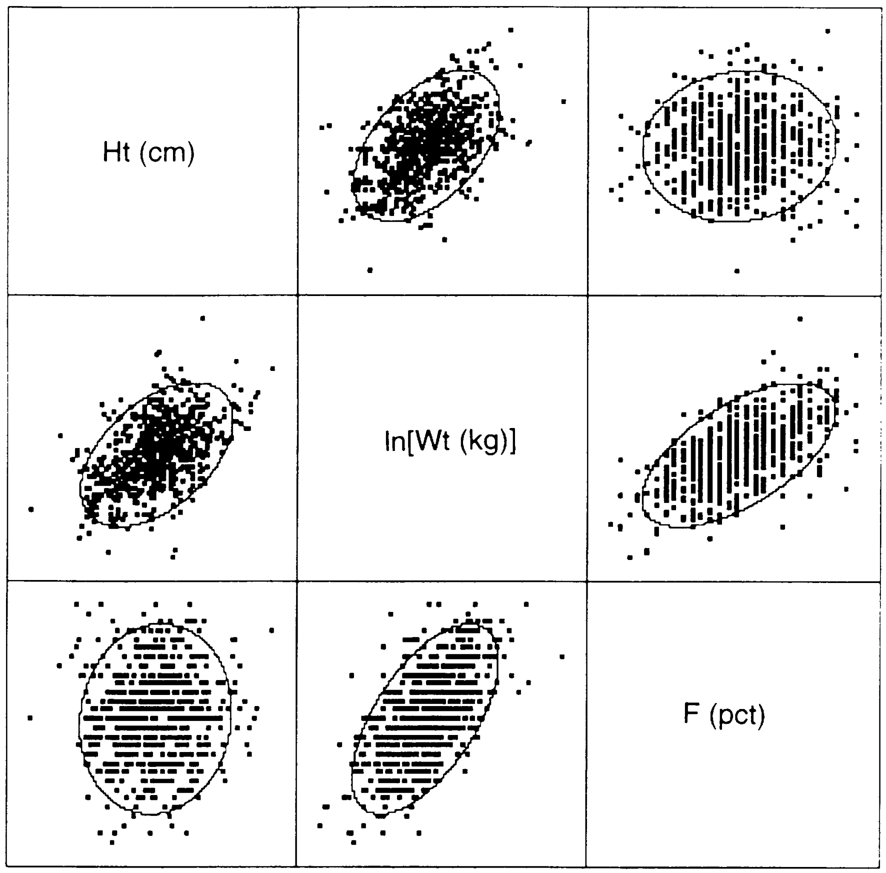

Figure 3: SPLOM for 646 Men, 50 yr ≤ Age < 80 yr

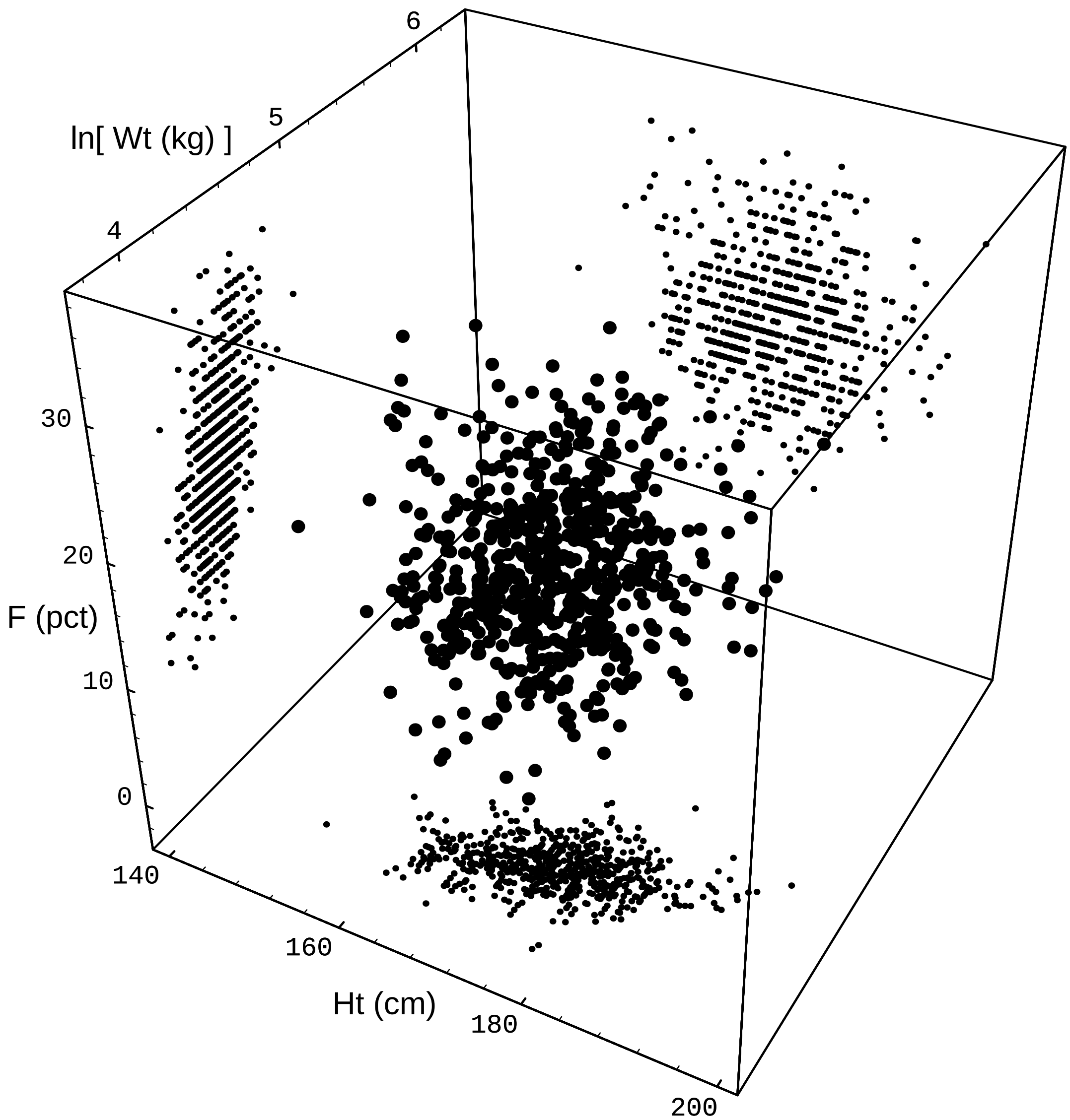

Figure 4
3D Scatterplot showing Bivariate Marginal Scatterplots (n = 646)