



Downloading and distribution via your company's intranet of the following article in accordance with the terms and conditions hereinafter set forth is authorized by SAS Institute Inc. Each article must be distributed in complete form with all associated copyright, trademark, and other proprietary notices. No additional copyright, trademark, or other proprietary notices may be attached to or included with any article.

THE ARTICLE CONTAINED HEREIN IS PROVIDED BY SAS INSTITUTE INC. "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. RECIPIENTS ACKNOWLEDGE AND AGREE THAT SAS INSTITUTE INC. SHALL NOT BE LIABLE FOR ANY DAMAGES WHATSOEVER ARISING OUT OF THEIR USE OF THIS MATERIAL. IN ADDITION, SAS INSTITUTE INC. WILL PROVIDE NO SUPPORT FOR THE MATERIALS CONTAINED HEREIN.

Graphical Techniques for Displaying Multivariate Data Using SAS/GRAPH Software

James R. Schwenke and Brian J. Fergen

James R. Schwenke, Ph.D., is the Associate Director of Biostatistics for Covance Periapproval Services, Inc. in Radnor, Pennsylvania. His areas of expertise include base SAS, SAS/IML, SAS/GRAPH, and SAS/STAT software, as well as SAS macros and works in the PC environment. Jim has a M.S. degree in math education from the University of South Florida and a Ph.D. degree from Kansas State University. Jim has been using SAS software since it was first introduced 24 years ago.

Brian J. Fergen, Ph.D., is a research scientist in biometrics working in the Animal Health Group of Pfizer Inc. His areas of expertise include base SAS, SAS/GRAPH, SAS/IML, and SAS/STAT software. Brian has a B.S. degree in Mathematics from Northern State University as well as M.S. and Ph.D. degrees in Statistics from Kansas State University. Brian has been a SAS user for over 10 years.

Note: This article was completed before Brian Fergen joined Pfizer, Inc.

Abstract

When measuring several response variables, multivariate statistical techniques, such as multivariate analysis of variance, are often more powerful in detecting differences among populations than traditional univariate techniques. The increased power of multivariate techniques is achieved by utilizing the correlation among the various response variables measured on a single experimentation unit. In recent years, a number of graphical techniques and computer software packages have been developed for displaying multivariate data. These techniques use various combinations of color, shape and movement in attempting to describe the multidimensional relationship of the response data. However, these techniques do not always transfer easily to the printed page for use in reports or research documents.

This paper is a review of two traditional graphical techniques, the profile and Andrews plots, which have been used extensively for displaying multidimensional data. The pinion plot is introduced as an alternative 2-dimensional graphical technique for displaying multivariate data. The pinion plot is compared to the profile and Andrews plots for describing differences among populations and as a graphical tool for detecting multivariate outliers. SAS code is provided to construct each graph using PROC GPLOT in SAS/GRAPH software.

Contents

- Introduction
- Profile Plot
- Andrews Plot
- Pinion Plot
- Multivariate Outliers
- Canonical Correlation
- Summary
- References

Introduction

Multivariate data analysis techniques are appropriate when more than one response is measured on an experimentation unit. Traditionally, multivariate data analysis techniques are considered when a variety of response variables are measured on individual experimentation units which together quantify some general characteristic. For example, heart rate, blood pressure, and cholesterol levels may be measured on each subject in a clinical trial to quantify the risk of cardiovascular disease. To investigate the differences in mean response among a set of treatments, univariate statistical techniques could be employed, where each response variable is statistically analyzed separately. However, the univariate analysis approach ignores the potential correlation among response variables.

Multivariate techniques use information in the correlation structure among response variables, which often increases the power of the statistical analysis to detect treatment difference as compared to the univariate counterpart. In addition, multivariate techniques maintain the nominal level of significance where a series of univariate tests, without further adjustment, on individual response variables may demonstrate some degree of multiplicity.

Longitudinal and repeated measures data can be considered as multivariate data where each response is considered as an individual response variable. Here again, the correlation among response variables can be utilized to potentially increase the power of detecting differences above univariate statistical procedures.

Although the benefit of multivariate statistical procedures over univariate procedures can be quite dramatic, it is somewhat difficult to accurately visualize results because of the multidimensional nature of the problem. With multivariate data, each response variable adds another dimension to the analysis problem. Primarily, summary reports and research documents are still physically restricted to the 2-dimensional boundaries of the printed page. However, current research is providing graphical techniques and computer software to better display multidimensional data. These techniques often use color, shape, size, movement, and even 3D glasses to give a visual sense of more than two dimensions.

This paper is a discussion of three graphical techniques for easily displaying multivariate data in 2-dimensions. Two standard techniques will be reviewed; the profile plot and the Andrews plot. The pinion plot is introduced as an alternative graphical technique for displaying multidimensional data in two dimensions. The benefit of these techniques is that each extends easily to any number of response variables and can accurately represent the multidimensional nature of the data in two dimensions. Examples are presented demonstrating each graphical technique and SAS 6.12 code is provided for easy implementation.

Profile Plot

A profile plot (Rencher, 1995) is a generalization of a traditional 2-dimensional plot, using a series of vertical axes presented consecutively along the base (x-axis) of the plot. Any number of response variables can be considered with varying scales of measurement. The response variables are arranged along the base of the plot, similar to discrete data. Each experimentation unit's set of response data is plotted on the corresponding vertical axis, with the plotted data connected by a line. Each line defines an experimentation unit's "profile" of response. Color and line type can be used to discriminate among populations or treatment groups. Of course, statistics such as the sample means associated with treatment groups can be plotted instead of or in addition to the individual experimentation unit's observed data.

Example 1

The data presented in Table 1 are taken from Rencher (1995), originally presented by Kleiner and Hartigan (1981). The data are the percentage of republican votes cast in presidential elections. Data for six southern states were collected from six selected election years. Here, the six southern states represent the sampling units with the six election years representing the response variables. One objective of a multivariate analysis and graphical display of these data is to contrast the relationship of the election years among states.

Figure 1 is a profile plot of the data presented in Table 1. The profile plot shows the relationship in voting preference among election years for the various states. The voting profile for Kentucky, Maryland and Missouri is more uniform over the selected years, as compared to the rise in percentage of people voting republican in 1964 for the other states. It is this relationship among profiles that defines the apparent differences between the two voting patterns. Univariate plots or summaries of individual election year voting preferences would not demonstrate these differences as effectively.

The SAS code for producing the profile plot in Figure 1 is given in Table 2. GOPTION statements are provided assuming that the figures are to be displayed on a color monitor. The font type and size for titles and text are specified to allow for a reasonable sized plot image. Color and line type are specified in the GPLOT statements allowing for use of either a color monitor or printer. If color is not available, the color requests will default to black.

The response data are read into DATA A, defining a character variable for the state names (STATE) and the election years (Y32, Y36, Y40, Y60, Y64 and Y68). DATA A will be more useful when constructing the Andrews and pinion plots that are discussed below. DATA B is used to define individual variables for the election year (YEAR) and the percent republican vote (VOTE) for construction of the profile plot. PROC GPLOT is used to construct the profile plot, treating the election years as a discrete variable. Color and line type are used in symbol statements to define the individual state's voting profile.

Andrews Plot

An Andrews plot (Everitt and Dunn, 1992) is based on a Fourier transformation of the multivariate response data. Basically, a Fourier transformation is an alternating sine-cosine functional representation of, in this case, the response data for each experimentation unit. The Fourier transform is defined as

$$f(t) = y_1/2^{1/2} + y_2*\sin(t) + y_3*\cos(t) + y_4*\sin(2t) + y_5*\cos(2t) + \dots$$

Each response variable in a multivariate data set is represented by an individual component in the sum of the Fourier transform. The observed value of the response variables are used to replace the corresponding y_i in each component of the transformation. Traditionally, t is varied between $-\pi$ and π to allow for an adequate representation of the data. The magnitude of each response variable for a particular experimentation unit's data affects the frequency, amplitude and periodicity of the combined sine-cosine wave, giving a unique representation of each experimentation unit's set of responses.

Example 1, *continued*

Figure 2 is the Andrews plot of the Republican voting preference data presented in Table 1. Here again, the similarities and differences among the states are clearly highlighted. However, it is not as obvious what the relationship is among the election years or which years define the differences among states.

The SAS code for constructing the Andrews plot in Figure 2 is given in Table 3. DATA C is constructed from DATA A to define the variables (F and T) for each state's Andrews curve. PROC GPLOT is used to construct the Andrews plot.

Pinion Plot

The pinion plot is an alternative to the profile and Andrews plots for displaying multidimensional data in two dimensions. (A pinion is a bird's wing and, in our opinion, the pinion plot is a very useful multidimensional graphical technique.) As with the profile and Andrews plots, the pinion plot can be constructed using standard graphics software. The basic concept of the pinion plot is to reuse the axes of a standard 2-dimensional plot to define the response for any number of variables. Some graphic software packages allow for the definition of a third axis, which

can be incorporated as a variation of the pinion plot defined here. Because each axes of the plot will be reused to define a response for more than one response variable, either the response variables have similar scales of measurement or each variable must be standardized to a common scale.

Let Y_1, Y_2, \dots, Y_p denote p response variables measured on a set of experimentation units. The standard 2-dimensional scatter diagram would plot Y_1 versus Y_2 , for example. This plot has limited usefulness because it is a projection of the p -dimension sample on to a 2-dimension space, which does not give a clear representation of the association between Y_1 and Y_2 with the other response variables. By allowing each axis of the standard 2-dimensional scatter plot to represent more than one response variable, the multidimensional relationship among the response variables can be displayed more effectively. For example, the axes of a 2-dimensional plot could first be used to plot Y_1 versus Y_2 . The axes are then reused to plot Y_3 versus Y_4 . This reuse of the axes continues until all response variables are represented in some pairing. If an odd number of response variables are measured, the final variable can be plotted on one of the axes as an individual variable. The points are then connected, giving a 2-dimensional representation of the multivariate data. The resulting pattern and grouping of lines are interpreted similar to profile or Andrews plots. For example, a concentrated cluster of lines indicates a uniform response in magnitude across the response variables.

The basic pinion plot can be enhanced by using various symbols to highlight which plotted point is associated with which pair of response variables. For example, a dot can be used to highlight the plotted point that is associated with the Y_1 - Y_2 pairing, a circle to highlight the plotted point that is associated with the Y_3 - Y_4 pairing, etc., with the final point defined by the end of the connecting line.

The pinion plot still represents a "projection" of the multivariate data in the sense that different pairings of response variables to axes will produce different "views" of the multidimensional data. It is suggested that the user investigate different possible pairings of the response variables to optimize the plot to meet specific goals.

Example 1, continued

Figure 3 is the pinion plot of the Republican voting preference data presented in Table 1, pairing consecutive election years. The SAS code for constructing this plot is given in Table 4. Because the response variables are on similar scales for this example, no standardization of the data was required. The pinion plot again shows the dramatic difference between the two groups of states. The dense cluster of the lines representing Kentucky, Maryland and Missouri shows a uniform voting record. The consistency of the pattern within each group shows the similarity among the states within each group.

The coordinates for the pinion plot are defined in DATA D using DATA A. A numeric code (CODE) is assigned to each state (STATE) to simplify the addition of the highlighted vertices. The reuse of the X-Y axes is accomplished by repeatedly defining an x-y pairing of the response data using OUTPUT statements. For the pinion plot in Figure 3, the 1932 data is paired with the 1936 first. To define the coordinates of the highlighted vertices, an x-y pairing is again defined using the 1932 and 1936 data for the first vertex. To allow for using a different symbol for plotting, the variable CODE is given the next largest value after the STATE codes. A format is defined to provide easy interpretation of the pinion plot and define the symbols used for vertex points. PROC GPLOT is again used to construct the pinion plot. Different lines are assigned to each state and symbols defined for the vertex points using the CODE variable and the SYMBOL statement.

Multivariate Outliers

An outlier is an extreme or aberrant data value, deemed to be uncharacteristic of the sampled population. Outliers can exhibit undue influence on summary statistics or test results. There are a number of univariate statistics developed to detect outliers. However, there are relatively few statistical procedures developed for detecting outliers in multivariate data. One method to detect multivariate outliers is based on the standardized multivariate distance D^2 (Rencher, 1995). The D^2 statistic is a measure of the squared multidimensional distance from an observation to the sample mean of the data set, standardized by the sample covariance matrix. Multivariate outliers are difficult to

visualize, again because a data point is an outlier with respect to the multidimensional characteristics corresponding to the set of response variables. In the next example, the profile, Andrews and pinion plots are used to highlight potential multivariate outliers.

Example 2

The data presented in Table 5 are taken from Rencher (1995), originally presented by Elston and Grizzle (1962). The length of the ramus bone of 20 boys was measured at four different ages; 8.0, 8.5, 9.0 and 9.5 years of age. For each subject, the D^2 statistic was computed to highlight potential outliers. Boys #9, 12 and 20 have relatively high values of D^2 , indicating potential outliers.

Figure 4 is a profile plot of the ramus bone data presented in Table 5. The profiles for boys #9, 12 and 20 are plotted as solid (red) lines, with all other boys plotted as dashed (blue) lines, to highlight the characteristics detected by the D^2 statistic. Although the number of lines plotted does make for a rather congested plot, it is possible to see that these boys have demonstrated an increased growth rate or growth spurt. It is interesting to note that there are other boys who also demonstrate similar growth spurts that were not designated as outliers using the D^2 statistic.

Figure 5 is an Andrews plot of the ramus bone data. Boys #9, 12 and 20 are again plotted with solid (red) lines. It is very difficult to determine any potential outlier characteristics because each boy's individual Andrews plot is within the mass for the rest of the boys.

Figure 6 is a pinion plot of the ramus bone data. Again, boys #9, 12 and 20 are plotted with solid (red) lines, with the remaining boys plotted with dashed (blue) lines. The pinion plot shows that boys #9, 12 and 20 are on the extreme of the sample. The remaining boys demonstrate an overall increasing pattern in response following a distinct linear trend. Clearly, boys #9, 12 and 20 are not characteristic of this same pattern of response. However, the increasing linear trend pattern is not consistent among the remaining boys in that boys with shorter ramus bone lengths demonstrate less variation about the linear trend than do boys with longer ramus bones. This observation may lead to the conclusion that there are two subgroups of boys represented in this data set.

Figure 7 is an Andrews plot constructed by first standardizing the response data for each age to have mean 0.0 and variance 1.0. Boys #9, 12 and 20 are again plotted using a solid (red) line. Here again, in addition to the boys denoted as outliers, two distinct subgroups of boys are observed. The two groups are defined by their pattern of response where one subgroup's Andrews plots increase in the center of the plot with the other subgroup decreasing. Two of the outliers, boys #9 and 12, have associated Andrews plots which are not that dissimilar to the "increasing" subgroup of boys. The remaining outlier, boy #20, has an Andrews plot that is somewhat unique, although boy #20 had the smallest D^2 of the three outliers.

Caution must be exercised when interpreting standardized Andrews plots. A consequence of standardizing data is to create a set of Andrews plots that are centered about the zero axis for the transformed response values, typical of what is seen in Figure 7. The zero axis corresponds to the multivariate mean of the sample. In this example, the multivariate mean happens to separate the two suspected subgroups in the sample. Thus, in this example, the two subgroups of boys are represented by the increasing and decreasing sets of Andrews plots.

Figure 8 is the pinion plot using the standardized response data. As indicated by the pinion plot in Figure 6, the same two subgroups of boys are observed. Here again, boys #9 and 12 display similar characteristics to the subgroup of boys with longer ramus bone lengths. Boy #20 is somewhat separated from the rest of the data.

With this graphical evaluation, it is not clear that boys #9, 12 and 20 should be considered as outliers. These data could be indicating that there are two distinct subgroups of boys, with the subgroup of boys with longer ramus bone lengths observed in all years, including the highlighted "outliers" being a subgroup demonstrating a greater generalized variance. The D^2 statistic can then be interpreted as an indication that at least three boys do not fit the assumption of one coherent sampled population. These graphical displays would indicate that further investigation of the two subgroups is warranted.

Canonical Correlation

Canonical correlation can be described as an extension of multiple correlation to multivariate data, where canonical correlation measures the correlation among groups of variables. The number of canonical correlations that can be estimated is equal to the smallest number of variables in either group. Each canonical correlation assesses a dimension of the relationship between the sets of variables as defined by linear combination of variables. These linear combinations, called canonical variates, are defined such that the first set of canonical variates provides the maximum correlation possible among linear combinations of variables from each group. The second canonical variates are uncorrelated with the first set of canonical variates and have a correlation that is less than or equal to the correlation for the first canonical covariates.

Example 3

The data presented in Table 6 are taken from Rencher (1995), originally presented by Box and Youle (1955). The data resulted from 19 designed experiments involving a chemical reaction in which various combinations of temperature (X_1), concentration (X_2) and time (X_3) were used in the chemical reaction as the design variables. At the end of each experiment the percentage of unchanged starting material (Y_1), percentage converted to the desired product (Y_2) and percentage of unwanted by-product (Y_3) were measured as the response variables.

One approach to an interpretation of this data is to conduct a canonical correlation analysis. For the data in Table 6, the response variables (Y) and the design variables (X) form a natural grouping of the variables. To fully investigate the relationship between the design and response variables, the square and cross-product terms among the design variables were considered as well. Because there are only three response variables (versus now nine design variables in the full analysis), the maximum number of canonical variates is three.

The correlations for the three canonical variates are 0.9899, 0.9525 and 0.4625, respectively. The correlation for the third canonical variate is small enough to preclude its inclusion in any further discussion. The values of the standardized coefficients (the values of the coefficients of the linear combinations) for the first two canonical variates are displayed in a profile plot in Figure 9, with the scale for each set of coefficients represented on either side of the figure. The coefficients of the first canonical variate are displayed as solid (red) lines with the coefficients of the second canonical variate displayed as dashed (blue) lines. This display makes it possible to determine graphically which variables are important for a given canonical covariate and which variables are important in the canonical correlation analysis. The magnitude of the coefficients in absolute value is important as it indicates the contribution the corresponding variable has to the canonical covariate. For example, the first canonical variate indicates that Y_1 is related to X_1 , X_2 , $X_1 * X_2$, $X_1 * X_3$ and X_1^2 , with X_2 having more influence than the remaining X variables, as its coefficient is the largest in absolute value. This type of presentation of the standardized coefficients will also be valuable when more than two canonical variates are plotted on the profile plot, which can easily be accomplished by proportionally adjusting all coefficients to the same scale. This allows you to simultaneously see what the relationship is among the canonical covariates, and which variables are influencing each individual canonical covariate. In addition, it is easy to determine the variables from each group that have small coefficients (in absolute magnitude) for all canonical covariates, and thus are unimportant to the analysis.

Figure 10 is a residual plot based on a multivariate regression of Y_1 , Y_2 , and Y_3 on X_1 , X_2 , $X_1 * X_2$, $X_1 * X_3$ and X_1^2 which incorporates aspects of both the profile and the pinion plot. To make the plot easier to interpret, the predicted values for each response variable, $YHAT1$, $YHAT2$, and $YHAT3$, respectively, were standardized. The values of the standardized predicted responses for Y_2 and Y_3 were then increased by an appropriate amount so that each set of residuals could be viewed separately on a continuous scale. The range of the transformed scale does not have meaning with respect to the original scale of measurement, and thus is displayed on a unitless axis in Figure 10. Similar to a pinion plot, the residual values, $R1$, $R2$, and $R3$, respectively, are connected corresponding to each experiment. Vertical reference lines are used to separate residuals corresponding to each response variable. If the multivariate regression provides an appropriate fit to the data, the residuals corresponding to each response variable should demonstrate a random scattering about zero with constant variance within each group. In Figure 10, no

discernable pattern in the residuals can be detected. However, two potential extreme observations are highlighted corresponding to experiments #8 and #15, demonstrating the minimum residual values for Y_2 and the maximum residual values for Y_3 .

Summary

Three graphical procedures appropriate for multivariate data were presented. The traditional profile and Andrews plots were presented. The pinion plot was introduced as an alternative 2-dimensional procedure for plotting multidimensional data. The three graphical procedures were compared through three examples involving multivariate data. Each technique proved to be useful in displaying multivariate data, characterizing the differences between populations and in highlighting potential outliers.

References

- Box, G.E.P., and Youle, P.V. (1955), "The Exploration of Response Surfaces: An Example of the Link between the Fitted Surface and the Basic Mechanism of the System," *Biometrics*, 11, 287-323.
- Elston, R.C. and Grizzle, J.E. (1962), "Estimation of Time-response Curves and Their Confidence Bands," *Biometrics*, 18, 148-159.
- Everitt, B.S. and Dunn, G. (1992). *Applied Multivariate Data Analysis*, New York: Oxford University Press.
- Kleiner, B. and Hartigan, J.A. (1981), "Representing Points in Many Dimensions by Trees and Castles," *Journal of the American Statistical Association*, 76, 260-269.
- Rencher, A.C. (1995). *Methods of Multivariate Methods*, New York: John Wiley and Sons.
- SAS Institute Inc. 1990, *SAS/GRAPH Software: Reference, Version 6, First Edition, Volumes 1 and 2*, Cary, NC: SAS Institute Inc.

THE FOREGOING ARTICLE IS PROVIDED BY SAS INSTITUTE INC. "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. RECIPIENTS ACKNOWLEDGE AND AGREE THAT SAS INSTITUTE INC. SHALL NOT BE LIABLE FOR ANY DAMAGES WHATSOEVER ARISING OUT OF THEIR USE OF THE ARTICLE. IN ADDITION, SAS INSTITUTE INC. WILL PROVIDE NO SUPPORT FOR THE ARTICLE.

Modified code is not supported by the author or SAS Institute Inc.

Copyright© 2000 SAS Institute Inc., Cary, North Carolina, USA. All rights reserved.

This article, number OBSWWW22, is found at the following URL: www.sas.com/obs.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. IBM® is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.

Table 1

Percentage of People Voting Republican in Presidential Elections

<u>State</u>	<u>Year</u>					
	<u>1932</u>	<u>1936</u>	<u>1940</u>	<u>1960</u>	<u>1964</u>	<u>1968</u>
Missouri	35	38	48	50	36	45
Maryland	36	37	41	46	35	42
Kentucky	40	40	42	54	36	44
Louisiana	7	11	14	29	57	23
Mississippi	4	3	4	25	87	14
South Carolina	2	1	4	49	59	39

Table 2

SAS Code for Profile Plot Using Republican Vote Data

```
goptions reset=all goutmode=replace;
goptions htitle=1.0 ftitle=simplex htext=1.0 ftext=simplex;

data a;
input state $14. @16 y32 y36 y40 y60 y64 y68;
cards;
Missouri          35 38 48 50 36 45
Maryland          36 37 41 46 35 42
Kentucky          40 40 42 54 36 44
Louisiana         7 11 14 29 57 23
Mississippi       4 3 4 25 87 17
South Carolina    2 1 4 49 59 39
run;

data b; set a;
year=1932; vote=y32; output;
year=1936; vote=y36; output;
year=1940; vote=y40; output;
year=1960; vote=y60; output;
year=1964; vote=y64; output;
year=1968; vote=y68; output;
drop y32--y68;
run;

proc sort data=b;
by state year;
run;

proc gplot data=b;
title1 'Figure 1';
title3 'Percentage of People Voting Republican in Presidential Elections';
title5 'Profile Plot';
axis1 label=(a=90 'Percent of People Voting Republican')
width=1 major=(w=1) minor=(n=3 w=1)
order=0 to 100 by 20;
axis2 label=('Election Year')
width=1 major=(w=1) minor=none offset=(2)
order=1932 1936 1940 1960 1964 1968;
legend1 label=('State:') across=2;
```

```
plot vote*year=state / vaxis=axis1 haxis=axis2 legend=legend1
    href=1932 1936 1940 1960 1964 1968 lhref=2;
symbol1 v=none i=join l=1 c=red w=1;
symbol2 v=none i=join l=2 c=red w=1;
symbol3 v=none i=join l=1 c=blue w=1;
symbol4 v=none i=join l=2 c=blue w=1;
symbol5 v=none i=join l=1 c=green w=1;
symbol6 v=none i=join l=2 c=green w=1;
run;
quit;
```

Table 3

SAS Code for Andrews Plot Using Republican Vote Data

```
data c; set a;
pi=3.14159265;
inc=2*pi/100;
do t=-pi to pi by inc;
f=y32/sqrt(2)+sin(t)*y36+cos(t)*y40+sin(2*t)*y60+cos(2*t)*y64+sin(3*t)*y68;
output;
end;
run;

proc gplot data=c;
title1 'Figure 2';
title3 'Percentage of People Voting Republican in Presidential Elections';
title5 'Andrews Plot';
axis1 label=(a=90 'f(t)')
      width=&w major=(w=&w) minor=(n=1 w=&w)
      order=-150 to 200 by 50;
axis2 label=('t')
      width=&w major=(w=&w) minor=(n=1 w=&w)
      order=-3.2 to 3.2 by 0.8;
legend1 label=('State:') across=2;
plot f*t=state / vaxis=axis1 haxis=axis2 legend=legend1;
symbol1 v=none i=join l=1 c=red w=1;
symbol2 v=none i=join l=2 c=red w=1;
symbol3 v=none i=join l=1 c=blue w=1;
symbol4 v=none i=join l=2 c=blue w=1;
symbol5 v=none i=join l=1 c=green w=1;
symbol6 v=none i=join l=2 c=green w=1;
run;
quit;
```

Table 4

SAS Code for Pinion Plot Using Republican Vote Data

```
data d; set a;
if state='Kentucky'      then code=1;
if state='Louisiana'    then code=2;
if state='Maryland'     then code=3;
if state='Mississippi'  then code=4;
if state='Missouri'     then code=5;
if state='South Carolina' then code=6;
x=y32; y=y36; output;
x=y40; y=y60; output;
x=y64; y=y68; output;
code=7; x=y32; y=y36; output;
code=8; x=y40; y=y60; output;
keep state x y code;
run;

proc format;
value state 1='Kentucky' 2='Louisiana' 3='Maryland'
            4='Mississippi' 5='Missouri' 6='South Carolina'
            7='1932 vs 1936' 8='1940 vs 1960';
run;

proc gplot data=d;
title1 'Figure 3';
title3 'Percentage of People Voting Republican in Presidential Elections';
title5 'Pinion Plot';
axis1 label=(a=90 'Election Year 1936/1960/1968')
       width=1 major=(w=1) minor=(n=3 w=1)
       order=0 to 100 by 20;
axis2 label=('Election Year 1932/1940/1964')
       width=1 major=(w=1) minor=(n=3 w=1)
       order=0 to 100 by 20;
legend1 label=('State:') across=2;
plot y*x=code / vaxis=axis1 haxis=axis2 legend=legend1;
symbol1 v=none i=join l=1 c=red w=1;
symbol2 v=none i=join l=2 c=red w=1;
symbol3 v=none i=join l=1 c=blue w=1;
symbol4 v=none i=join l=2 c=blue w=1;
symbol5 v=none i=join l=1 c=green w=1;
symbol6 v=none i=join l=2 c=green w=1;
symbol7 v=dot i=none c=black;
symbol8 v=circle i=none c=black;
format code state.;
run;
quit;
```

Table 5

Length of Ramus Bone for Boys at Four Ages

<u>Subject</u>	<u>Age in Years</u>				<u>D²</u>
	<u>8.0 Years</u>	<u>8.5 Years</u>	<u>9.0 Years</u>	<u>9.5 Years</u>	
1	47.8	48.8	49.0	49.7	0.76
2	46.4	47.3	47.7	48.4	1.30
3	46.3	46.8	47.8	48.5	1.76
4	45.1	45.3	46.1	47.2	3.85
5	47.6	48.5	48.9	49.3	0.87
6	52.5	53.2	53.3	53.7	2.81
7	51.2	53.0	54.3	54.5	4.29
8	49.8	50.0	50.3	52.7	7.99
9	48.1	50.8	52.3	54.4	11.03
10	45.0	47.0	47.3	48.3	8.35
11	51.2	51.4	51.6	51.9	2.83
12	48.5	49.2	53.0	55.5	10.57
13	52.1	52.8	53.7	55.0	2.59
14	48.2	48.9	49.3	49.8	0.66
15	49.6	50.4	51.2	51.8	0.32
15	50.7	51.7	52.7	53.3	0.83
17	47.2	47.7	48.4	49.5	1.11
18	53.3	54.6	55.1	55.3	4.36
19	46.2	47.5	48.1	48.4	2.11
20	46.3	47.6	51.3	51.8	10.09

Table 6**Chemical Reaction Data**

Experiment Number	Response Variables			Design Variables		
	Unchanged Material	Converted Product	By-Product	Temperature	Concentration	Time
	<u>Y₁</u>	<u>Y₂</u>	<u>Y₃</u>	<u>X₁</u>	<u>X₂</u>	<u>X₃</u>
1	41.5	45.9	11.2	162	23.0	3.0
2	33.8	53.3	11.2	162	23.0	8.0
3	27.7	57.5	12.7	162	30.0	5.0
4	21.7	58.8	16.0	162	30.0	8.0
5	19.9	60.6	16.2	172	25.0	5.0
6	15.0	58.0	22.6	172	25.0	8.0
7	12.2	58.6	24.5	172	30.0	5.0
8	4.3	52.4	38.0	172	30.0	8.0
9	19.3	56.9	21.3	167	27.5	6.5
10	6.4	55.4	30.8	177	27.5	6.5
11	37.6	46.9	14.7	157	27.5	6.5
12	18.0	57.3	22.2	167	32.5	6.5
13	26.3	55.0	18.3	167	22.5	6.5
14	9.9	58.9	28.0	167	27.5	9.5
15	25.0	50.3	22.1	167	27.5	3.5
16	14.1	61.1	23.0	177	20.0	6.5
17	15.2	62.9	20.7	177	20.0	6.5
18	15.9	60.0	22.1	160	34.0	7.5
19	19.6	60.6	19.3	160	34.0	7.5

Figure 1
Percentage of People Voting Republican in Presidential Elections
Profile Plot

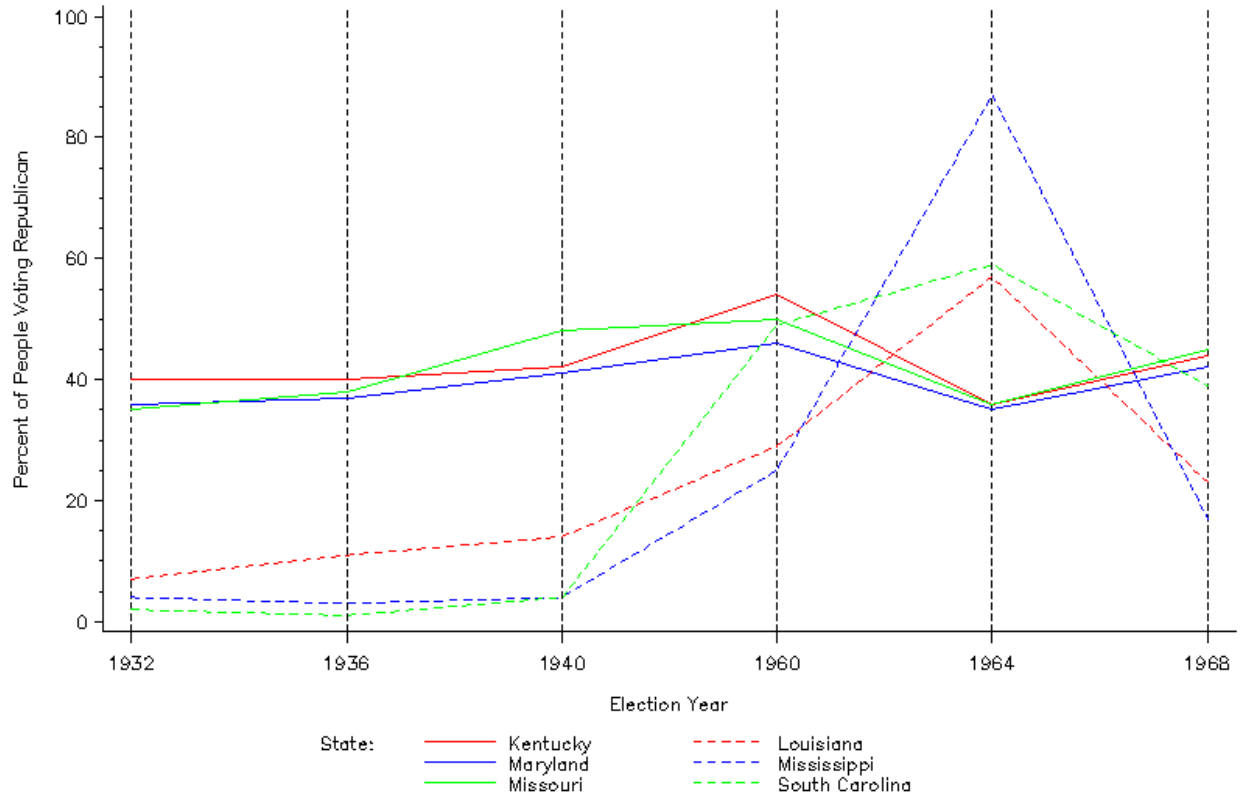


Figure 2

Percentage of People Voting Republican in Presidential Elections

Andrews Plot

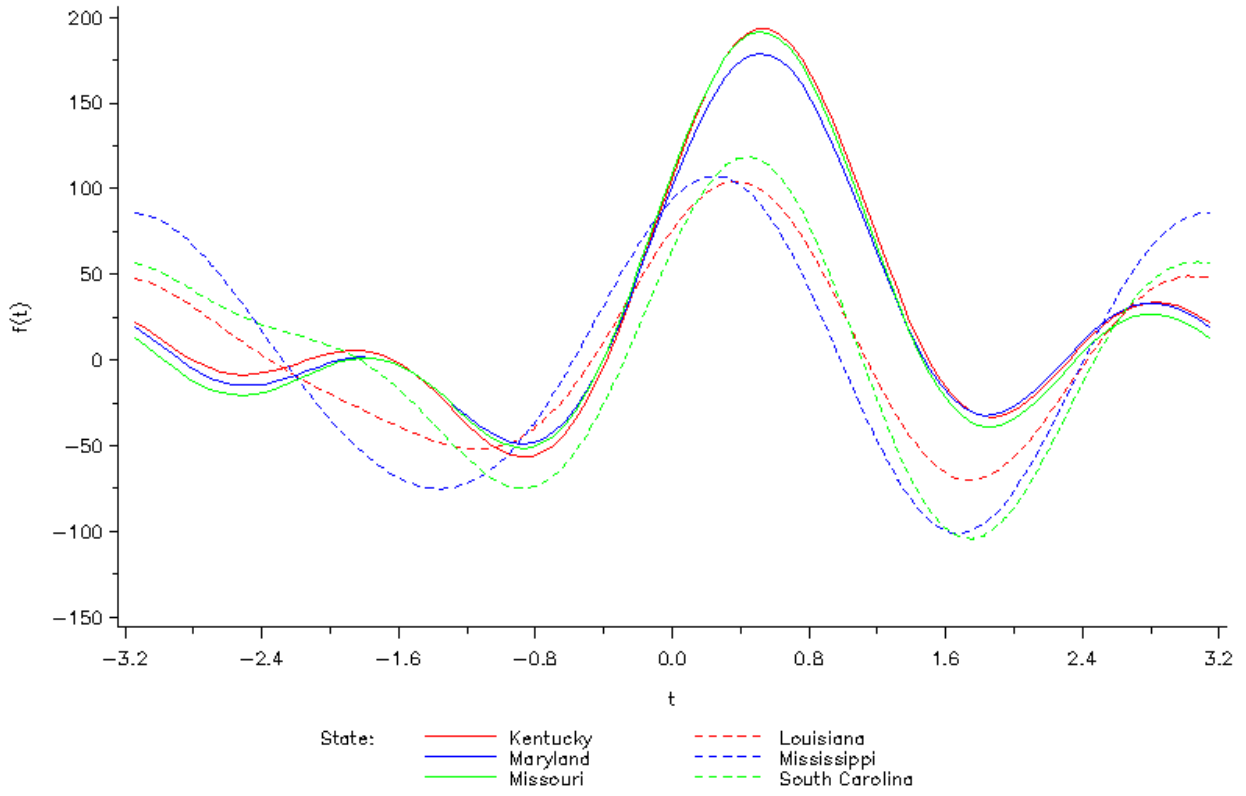


Figure 3

Percentage of People Voting Republican in Presidential Elections

Pinion Plot

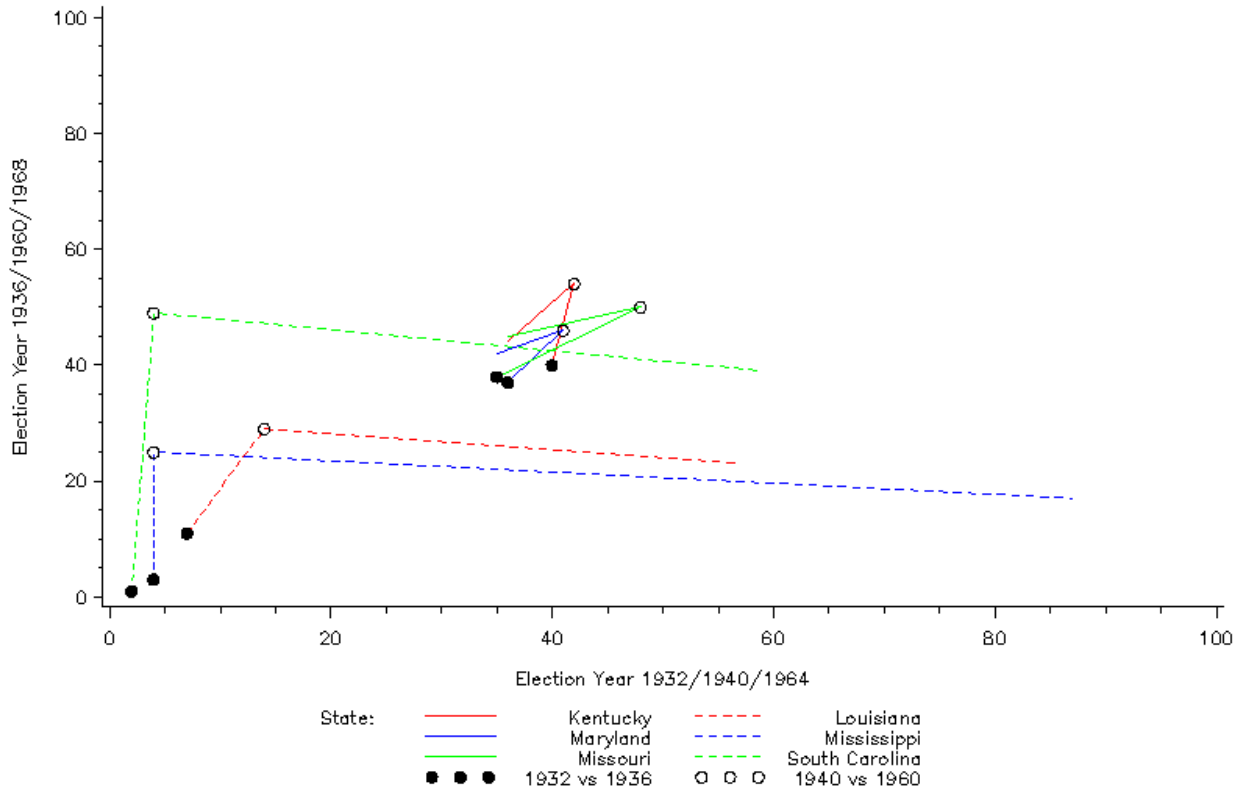


Figure 4
Length of Ramus Bone in Boys
Profile Plot

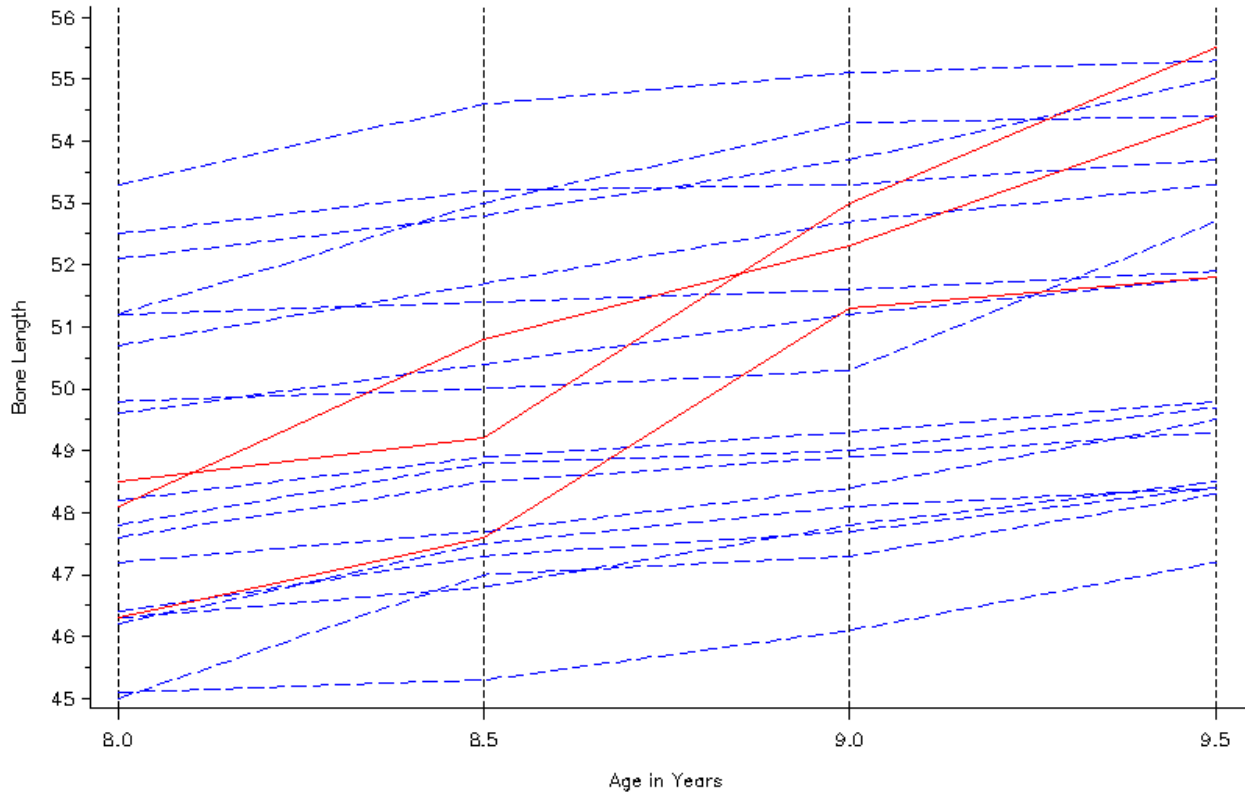


Figure 5
Length of Ramus Bone in Boys
Andrews Plot

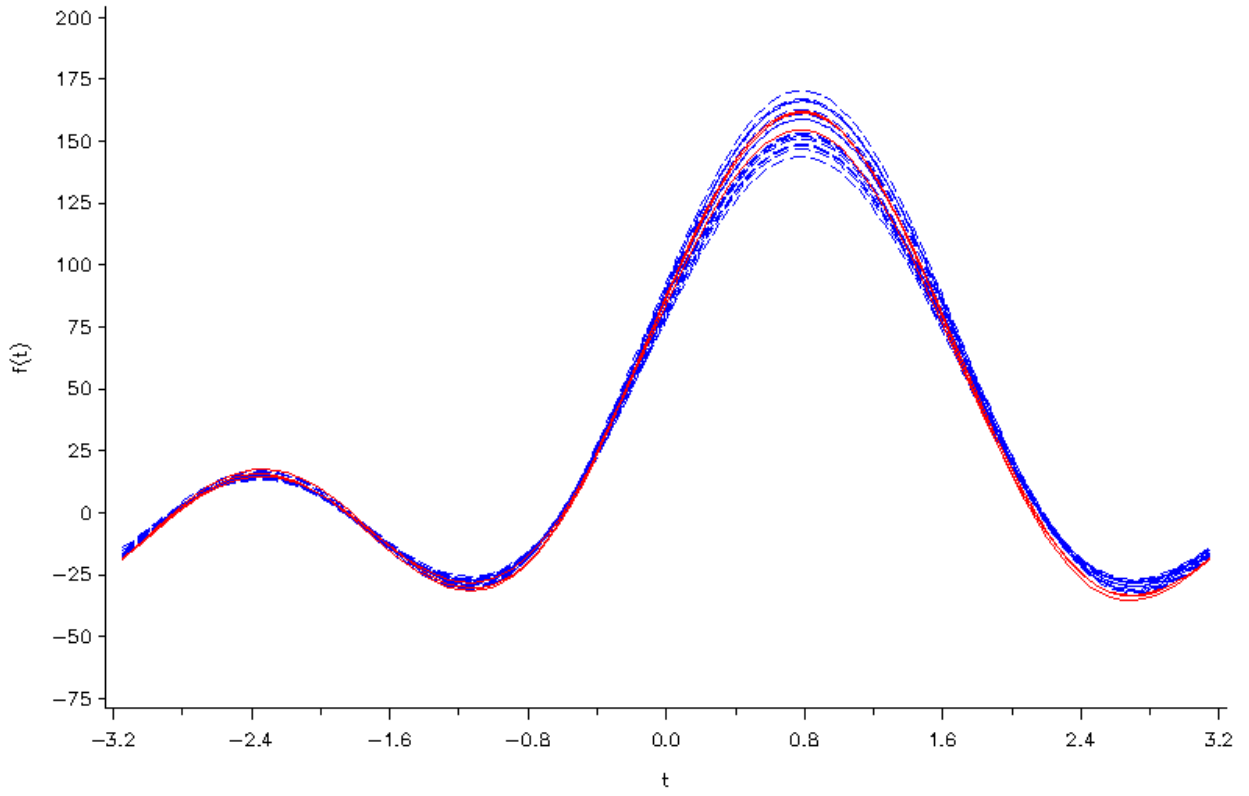


Figure 6
Length of Ramus Bone in Boys
Pinion Plot

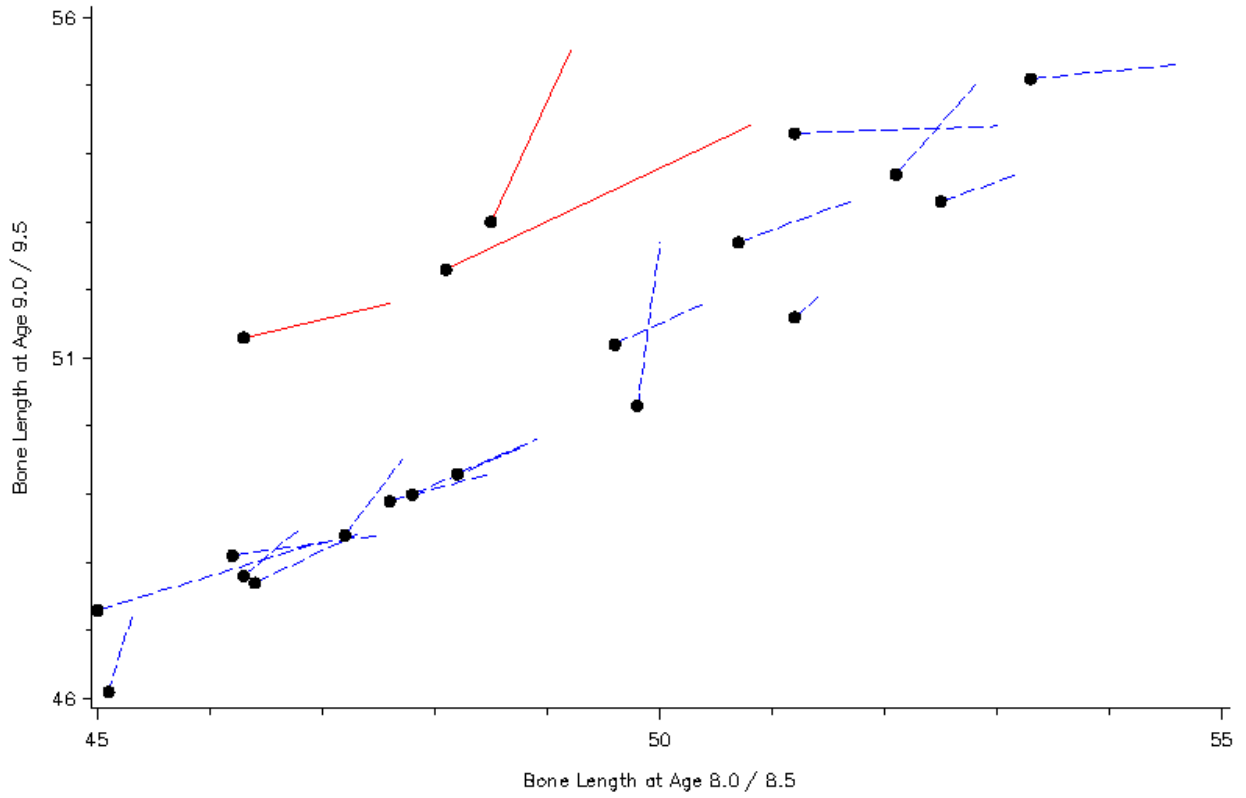


Figure 7
Length of Ramus Bone in Boys
Standardized Andrews Plot

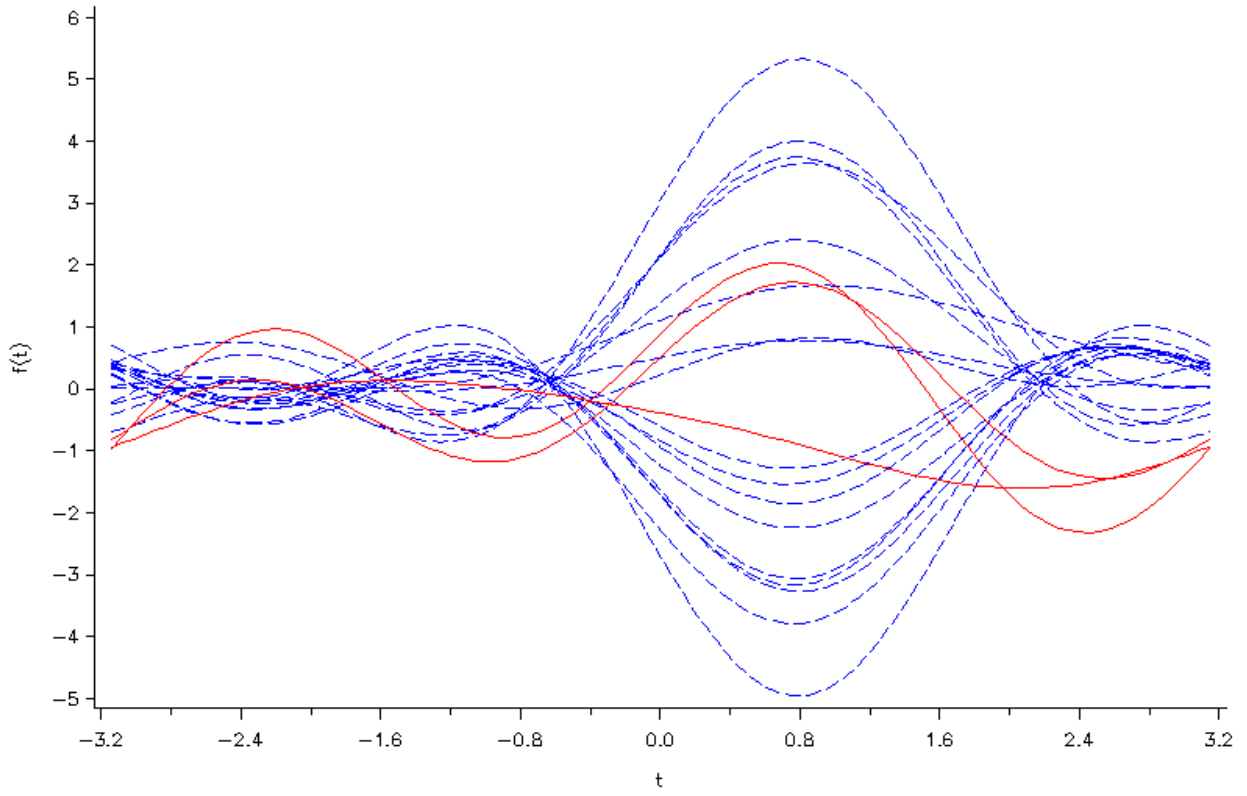


Figure 8
Length of Ramus Bone in Boys
Standardized Pinion Plot

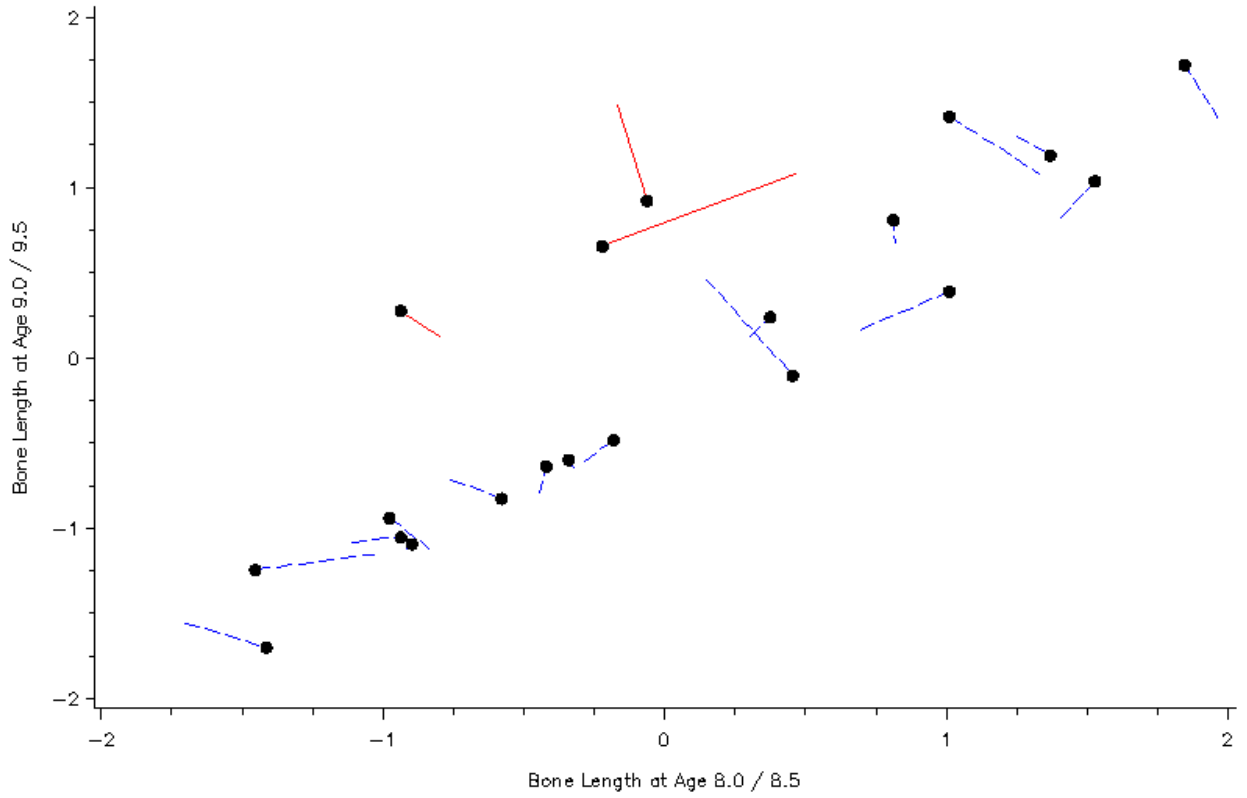


Figure 9

Standardized Canonical Coefficients

Profile Plot

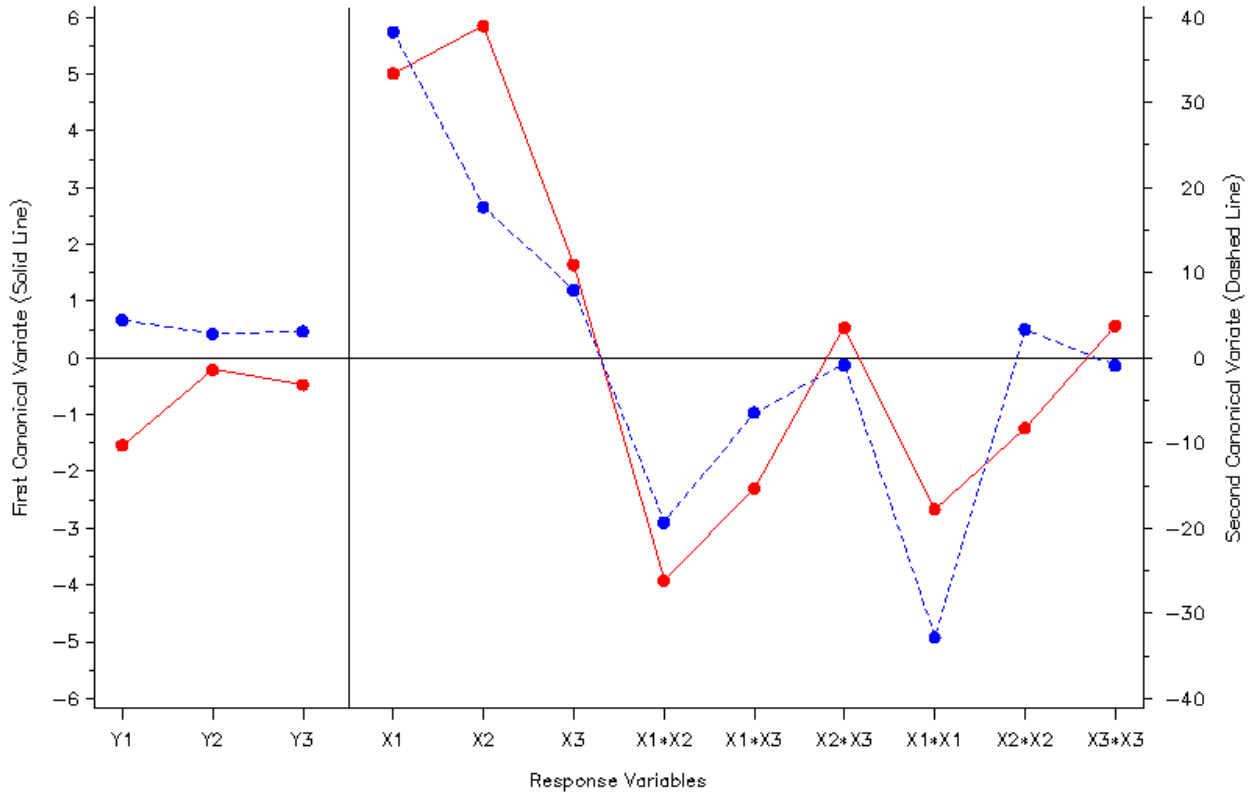


Figure 10
Multivariate Residual Plot
Profile Pinion Plot

