Paper 250-2008

Understanding Multiple, Repeated Animal Measurements with the Help of PROC GPLOT

Martin J. Zuidhof, Alberta Agriculture and Food, Edmonton, AB, Canada
Robert A. Renema, University of Alberta, Edmonton, AB, Canada
Frank E. Robinson, University of Alberta, Edmonton, AB, Canada

ABSTRACT

In Animal Science, repeated measurements on individual animals are affected by many factors, including the age of the animal, feed intake, and metabolic, reproductive, and health status. In experiments involving multiple individuals, major physiological events such as puberty occur asynchronously, complicating the interpretation of age-based representation of data. PROC GPLOT can be used to visualize complex data for an initial qualitative analysis. We employ macros to incorporate ancillary data summarizing pertinent characteristics of individuals. Using this approach, we have identified trends in our data otherwise masked by the complex biological system.

Creating a graph of multiple longitudinal variables was somewhat complex, but once the initial graph was created for one subject, SAS allowed us to implement the code effortlessly for any number of individuals. In our standard experiments, we have come to rely on data visualization to understand unexpected data points and gain new insights into nutritional dynamics. In this paper, we illustrate how we generate complex visual representations of the responses of 288 individual broiler breeder hens to a variety of management regimes. This methodology can be applied readily to other areas of animal science, economics, or any discipline where multiple input and output variables are measured serially.

INTRODUCTION

SAS is capable of advanced graphical rendering of data. Simple summary data can easily be graphed in a spreadsheet such as MS Excel, but a large database of raw data on individual subjects poses a significant challenge. Our experience is that when multiple graphs are required to understand the behavior of individual subjects, it is well worth the investment of time to design a basic graph that characterizes traits of interest for an individual, and use SAS' powerful programming language to output multiple graphs for individual subjects. This poster describes the process of creating such a system of graphs.

BROILER BREEDER BASICS

A few basic concepts about broiler breeder management are provided to aid those not familiar with the process of hatching egg production. Broiler breeders are the parents of broiler (meat-type) chickens. If allowed to express their potential for growth, they become overweight and reproduce poorly. Therefore they are given restricted amounts of feed - effectively a diet that helps them to follow a target body weight curve at which they are more physically fit. The broiler breeder hens in this study were weighed twice weekly, at which time feed allocation was adjusted to ensure adherence to the target BW curve.

Once consumed by the chicken, the fate of nutrients is genetically determined. An ideal type of bird for the chicken meat industry would be a highly efficient parent that would reproduce well, and have progeny with rapid growth and high efficiency. Thus it is a useful exercise to track the fate of the nutrients that are provided to the breeder hens. There is a lot of variation in this biological system. The timing of sexual maturity, competition for feed, nutrient partitioning priorities, ability to maintain a consistent rate of lay, and health status are among the variables that often affect performance of individuals. It is very useful therefore to visualize our data. Visualization allows us to make qualitative observations which can be directly incorporated into our knowledge of the animals, and often leads to development of theories and statistical tests of those theories.

DATA

The data we will use in this example were collected on individual subjects at multiple ages. Body weight (BW), average daily feed intake (ADFI), and average daily gain (ADG) data were collected twice weekly, and egg weight (EggWt) was measured daily. Since groups of broiler breeders are fed a restricted amount of feed to optimize egg production, the target BW is also important.

The following data step illustrates the (truncated) data structure for the first subject (individual hen within a cage):

```
Data sample;
  Input Cage Age BW TargetBW ADFI ADG EggWt;
  agewk=age/7;
Datalines;
1      143    2068   2216   85.0   2.8    .
1      146    2132   2278   91.7   21.3   .
1      150    2234   2369   99.5   25.5   .
1      153    2335   2440   105.7  33.7   .
1      157    2426   2540   108.8  22.8   .
1      160    2495   2615   113.0  23.0   .
1      164    2581   2711   119.5  21.5   .
1      167    2672   2778   123.0  30.3   .
1      171    2787   2862   124.0  28.8   .
1      173    .      .      .      .      50.1
1      174    2883   2921   124.0  32.0   52.3
1      176    .      .      .      .      51.1
1      177    .      .      .      .      53.3
1      178    2935   2997   124.0  13.0   53.1
1      179    .      .      .      .      53.2
1      180    .      .      .      .      52.1
1      181    2864   3052   124.7  -23.7  55.7
1      182    .      .      .      .      54.5
1      183    .      .      .      .      53.8
1      184    .      .      .      .      53.5
1      185    2868   3119   128.8  1.0    53.5
```

```
1       186    .       .       .       .       54.2
1       187    .       .       .       .       56.1
1       188    2870    3165    133.3   0.7     54.2
1       189    .       .       .       .       52.7
1       190    .       .       .       .       54.5
1       191    .       .       .       .       56.0
1       192    2918    3221    140.3   12.0    54.6
1       193    .       .       .       .       54.1
1       194    .       .       .       .       56.1
1       195    2951    3258    145.3   11.0    53.8
1       196    .       .       .       .       52.4
1       198    .       .       .       .       54.2
1       199    3031    3302    149.3   20.0    56.6
1       200    .       .       .       .       54.2
1       201    .       .       .       .       55.4
1       202    3063    3331    150.7   10.7    55.1
1       203    .       .       .       .       52.2
1       204    .       .       .       .       54.4
1       205    .       .       .       .       58.6
1       206    3042    3364    150.3   -5.3    57.0
1       207    .       .       .       .       56.8
1       208    .       .       .       .       58.9
1       209    3079    3385    150.7   12.3    57.3
1       210    .       .       .       .       56.9
1       211    .       .       .       .       57.9
1       212    .       .       .       .       58.7
1       213    3128    3400    151.0   12.3    58.9
1       214    .       .       .       .       57.4
1       215    .       .       .       .       57.8
1       216    3113    3408    149.7   -5.0    57.5
1       217    .       .       .       .       57.9
1       218    .       .       .       .       58.4
1       219    .       .       .       .       57.5
1       220    3146    3418    149.8   8.3     58.3
1       221    .       .       .       .       55.3
1       222    .       .       .       .       56.9
1       223    3156    3427    148.0   3.3     56.1
1       224    .       .       .       .       57.5
;
run;
```

GRAPHING THE DATA

One of the most challenging parts of visualizing complex data is the design of the graph. What should be incorporated? In our experience, an iterative design process often leads us to new insights; we learn as much in the process of data visualization as we do from the final collection of graphs. Our first challenge was to plot the various serial measurements on a single graph. To visualize change in parameter values, parameters with similar orders of magnitude should be grouped together. Appropriate symbols, lines, and legends make the graph easier to read and understand. This was accomplished with the following code:

```
/* reset graphics options, set the default text height and font and
   direct SAS to create a collection of GIF images */
goptions reset=all htext=1.5 ftext="arial/bo"
```

```
         device=gif goutmode=append;

/* close the regular output destination */
ODS Listing Close;
/* specify filenames for the ODS output and the location to
   store them */
ods html path = "C:\temp\SAS Global Forum"
             body="BirdGraphs.htm";

/* set up appropriate symbols and lines for each y-axis variable */
symbol1 v=dot  c=steel  i=none;
symbol2 v=none c=steel  i=join   line=1;
symbol3 v=none c=salmon i=join   line=2;
symbol4 v=none c=navy   i=join   line=3;
symbol5 v=none c=green  i=needle line=1;

/* define horizontal and left and right vertical axes */
axis1 order=20 to 30 by 1 minor=none;
axis2 order=0 to 4000 by 500 label=(angle=90 "BW (g)") minor=none;
axis3 order=-50 to 175 by 25 label=(angle=90 "Weight (g)") minor=none;

/* define legends for left and right axes */
legend1 noframe position=(inside left) label=none value=(h=1)
        mode=protect offset=(5 pt,0 pt);
legend2 noframe position=(inside right) label=none value=(h=1)
        mode=protect offset=(-5 pt,0 pt);

/* implement the gplot procedure to plot the five longitudinal
   variables simultaneously as a function of age */
proc gplot data=sample;
by cage; /* produce a plot for every subject */
label Agewk    = "Age (wk)"
      ADFI     = "Feed intake (/d)"
      ADG      = "Gain (/d)"
      Eggwt    = "Egg wt"
      TargetBW = "Target BW";
plot (BW TargetBW)*agewk / overlay noframe haxis=axis1 vaxis=axis2
      legend=legend1;
plot2 (ADFI ADG EggWt)*agewk / overlay vaxis=axis3 legend=legend2
      cvref=lig vref=52 ;
run;

/* close the ODS html output device and direct output to the default
   destination */
ods html close;
ods listing;

quit;
```

With the full data set, this produces a graph for every hen. The collection of graphs provides a picture over time of the characteristics of each subject (BW, relative to the target BW), together with inputs (ADFI), and outputs: average daily gain and discrete egg production events. One of the insights we gained immediately with this technique was that there a drop in BW (negative gain) commonly occurred around the time of puberty (first egg). Because feed intake was identical for birds within treatments, and the age at first egg is unique to each individual, the timing of the drop in BW varied slightly, and had never been identified before visualizing the data.

The interpolation 'needle' provided a great way of representing discrete egg production events. The length of the 'needle' represents the weight of each egg. Another helpful addition was the vertical reference line at 52 g, which is the threshold weight at which local commercial hatcheries accept breeder eggs. With this new collection of graphs, we began to recognize differences in nutrient partitioning priorities in broiler breeders. Some were selfish, holding on to nutrients, and became heavier, often producing small eggs, while some tended to be martyrs, diverting a more substantial proportion of available nutrients to reproduction while drifting further and further below the target BW over time. Other, biologically efficient hens were able to both gain BW and produce large numbers of settable eggs. This subclass of hens had not been previously identified in experimental data sets. Because of their exceptional efficiency, these hens hold much promise for study in future research and commercial selection programs. These insights would have been very difficult to elucidate without visualizing these variables together for each hen.

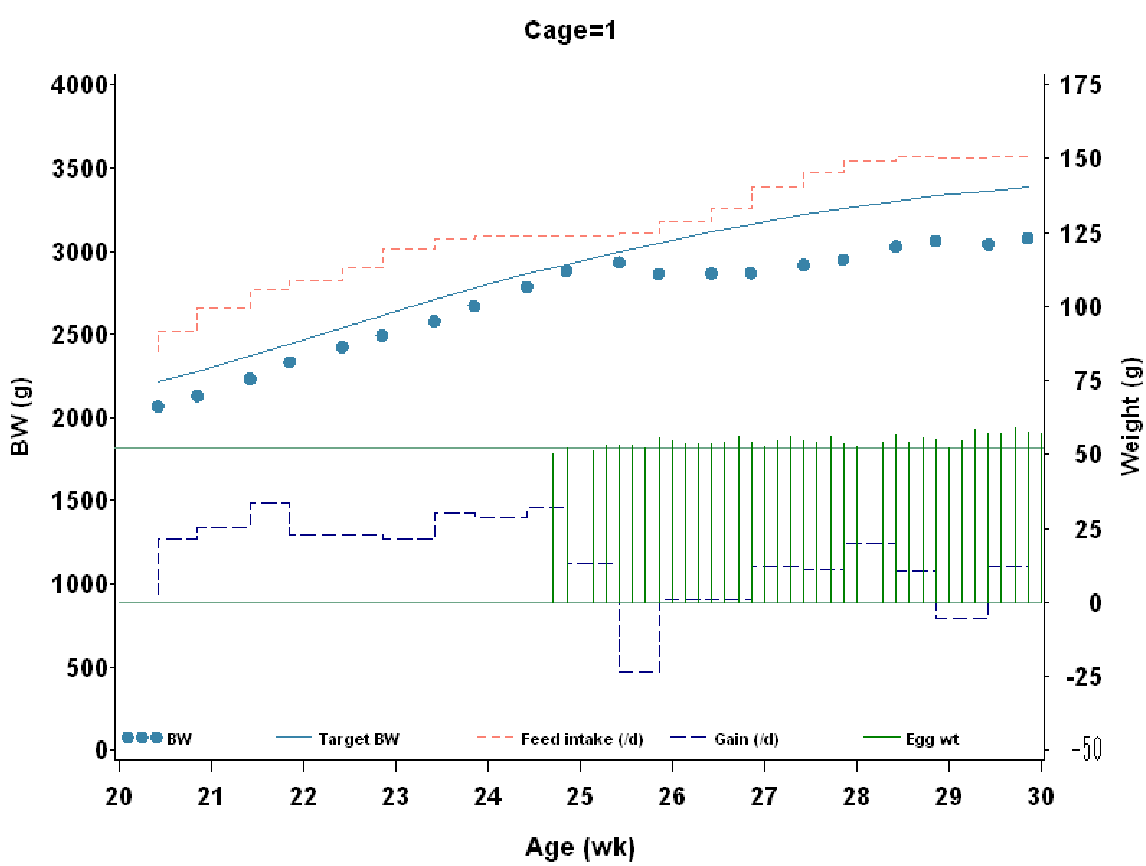

Figure 1. Graph of serial measures of multiple variables from an individual broiler breeder hen.

While the visualization in Figure 1 was very helpful, we found that the qualitative analysis still required a great deal of cross referencing. The next hurdle was to find a way to include some summary information from data sets for other parameters as notes on each hen's graph. This information helped us to simultaneously observe multiple other potential energy reserviors, giving us a more complete picture of the way hens diverted nutrients – a picture that we had never seen before.

INCORPORATION OF ANCILLARY SUBJECT-SPECIFIC DATA

Our ancillary data consisted of treatment information, lifetime productivity summaries, and physical characteristics of the subject. For demonstration purposes, we will notate the graph with a small number of variables: genetic strain; experimental treatment; total eggs produced; average (g) and total (kg) egg weights; the length of the longest period of consecutive eggs (prime sequence); and the weights of the pectoralis (breast) muscle and fat at the end of production, expressed as a percentage of BW.

The sample ancillary data (for the first two subjects) were in the following format:

```
Data ancillary;
  Input cage strain$ treatment$ total_eggs breast_pct
        fat_pct prime_sequence average_eggwt total_eggwt;
  Datalines;
1 Ross508 Standard 187 0.178 0.214 31 61.4 11.5
2 HiY     Standard 158 0.173 0.241 17 52.1 9.8
;
run;
```

We employed SAS' macro language to uniquely notate each hen's graph. A macro named 'create' was developed, with two parameters: the first and last subjects to be graphed. To incorporate subject-specific values as notes on the graph, we used the CALL SYMPUT routine within a data step to assign data from the ancillary data set into macro variables, which we then inserted into footnotes. More specific details are provided as comments in the following snippet of code:

```
/* Turn off the byline option to optimize the custom title later */
options nobyline;

/* The macro statement begins the macro definition. Two parameters are
   Defined which represent the first and last subject to be graphed.
   The macro is run once per subject. */
%macro create(begin,end);
%do i=&begin %to &end;
  /* A null data step is used to extract subject-specific summary data
     into macro variables */
  data _null_;
    set ancillary;
    where cage=&i;
    call symput('Strain',trim(Strain));
    call symput('Treatment',trim(Treatment));
    call symput('total',trim(put(Total_Eggs,3.0)));
    call symput('breast',put(breast_pct,percent7.1));
```

```
    call symput('pctfat',put(fat_pct,percent7.1));
    call symput('Prime',trim(put(Prime_Sequence,3.0)));
    call symput('eggwt',put(average_eggwt,4.1));
    call symput('eggmass',put(total_eggwt,4.1));
  run;

  /* Summary information is inserted into footnotes and the title */
  footnote1 j=c h=1.2
            j=l "&breast. breast at 58 wk"
            j=r "&eggwt. g average egg wt";
  footnote2 j=left h=1.2 "&pctfat. carcass fat at 58 wk"
            j=c "&Prime. day prime sequence"
            j=right "Total: &eggmass. kg eggs";
  Title1 j=l "Strain: &Strain." j=c "Cage &i" j=r "Treatment:
            &treatment.";

  /* Implement the gplot procedure to plot the five variables */
  proc gplot data=sample;
  where cage=&i;
  label Agewk    = "Age (wk)"
        ADFI     = "Feed intake (/d)"
        ADG      = "Gain (/d)"
        Eggwt    = "Egg wt"
        TargetBW = "Target BW";
  plot (BW TargetBW)*agewk / overlay noframe haxis=axis1 vaxis=axis2
        legend=legend1;
  plot2 (ADFI ADG EggWt)*agewk / overlay vaxis=axis3 legend=legend2
        cvref=lig vref=52 ;
  run;
  quit;
%end;
%mend;

/* The following statement invokes the macro once. To generate a graph
   for 288 hens we would modify the statement to read %create(1,288) */
%create(1,1);

/* The following closes the ODS html engine, and returns to the default
   output listing. */
ods html close;
ods listing;
```

Note that the GPLOT syntax is almost identical. We simply needed to place the graphing procedure within a loop to collect subject-specific data from a related dataset. The resulting graphic is presented in Figure 2.

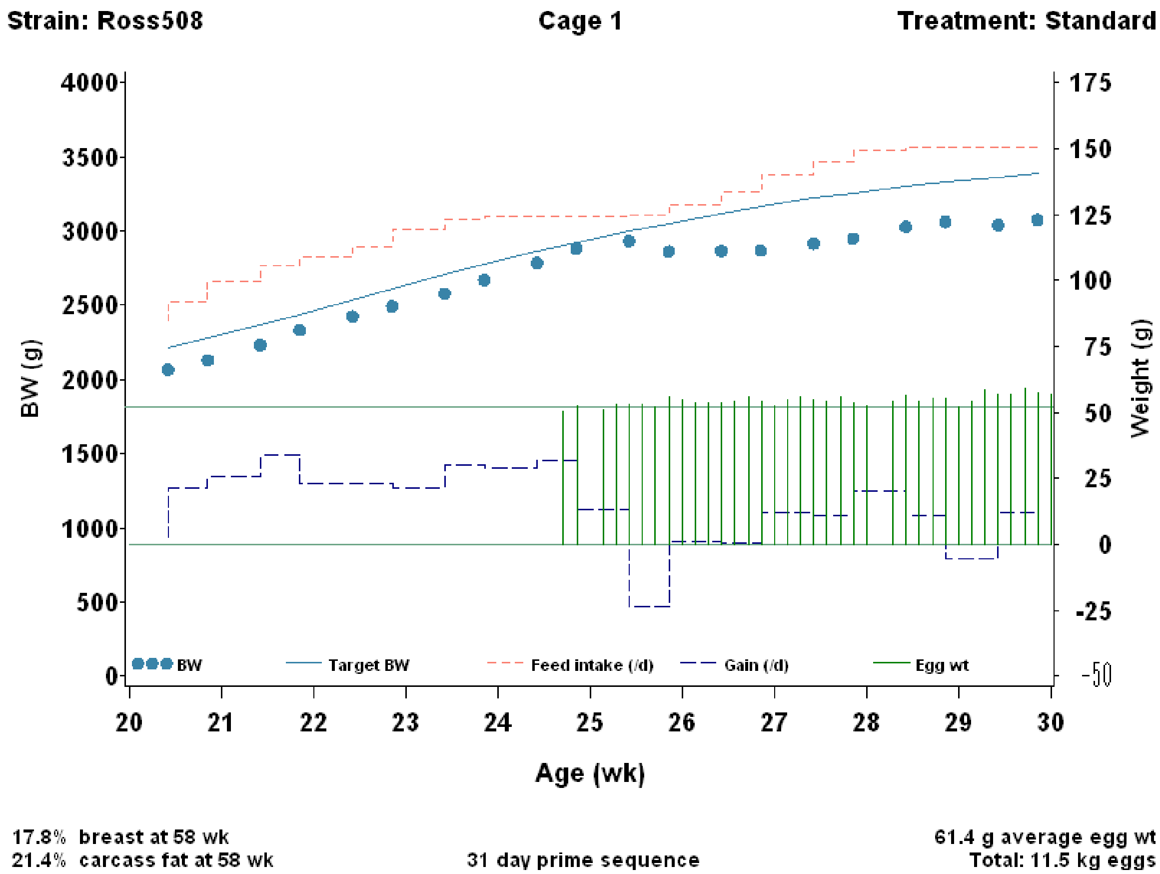

Figure 2. Graph of serial measures of multiple variables from an individual broiler breeder hen supplemented with summary data from an ancillary data set.

CONCLUSION

In the field of animal science, relatively complex relationships often reside in longitudinal data sets. Graphical representation of the data, particularly with subject-specific notation, is a powerful way to add value to initial qualitative analysis of the data. Using this approach, we have identified trends otherwise masked within a complex biological system. This methodology can be applied readily to other areas of animal science, economics, or any discipline where multiple input and output variables are measured serially.

REFERENCES

http://support.sas.com/onlinedoc/913/getDoc/en/graphref.hlp/gplot-plot.htm
http://support.sas.com/onlinedoc/913/getDoc/en/graphref.hlp/symbolchap.htm
http://support.sas.com/onlinedoc/913/getDoc/en/graphref.hlp/axischap.htm
http://support.sas.com/onlinedoc/913/getDoc/en/graphref.hlp/legendchap.htm
CALL SYMPUT:
http://support.sas.com/onlinedoc/913/getDoc/en/lrdict.hlp/a000127861.htm

CONTACT INFORMATION
Your comments and questions are welcome. Authors may be contacted at:

Martin Zuidhof PhD
Alberta Agriculture and Food
204, 7000 113 Street
Edmonton, AB T6H 5T6
Canada
Work Phone: (780)427-4608
Email: Martin.Zuidhof@gov.ab.ca

Robert Renema PhD
University of Alberta
Email: Robert.Renema@ualberta.ca

Frank Robinson PhD
University of Alberta
Email: Frank.Robinson@ualberta.ca