

# EPG Data Analysis 101

A brief introduction to Backus 2.0

- by
- T.A. Ebert
- M.E. Rogers

EPG  
Work-  
shop



# Introduction

(The program and citation are on the website)

- Used about as frequently as the Sarria Workbook.
- Includes a statistical analysis.
- Is easy to use.
- Calculates many variables that the other programs do not calculate
- Fails to calculate many variables calculated by the other programs.

EPG  
Work-  
shop



# Outline

- The analysis using SAS
  - The alternative is to use Enterprise Guide.
    - Enterprise Guide is a user interface for SAS.
    - Enterprise Guide is harder to use in combination with Backus 2.0.
    - We will use Enterprise Guide when we use Ebert 1.0.
    - Do not use Enterprise Guide in this section.
- Looking at residuals.

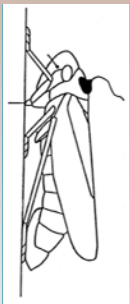
EPG  
Work-  
shop



# Backus Program in SAS

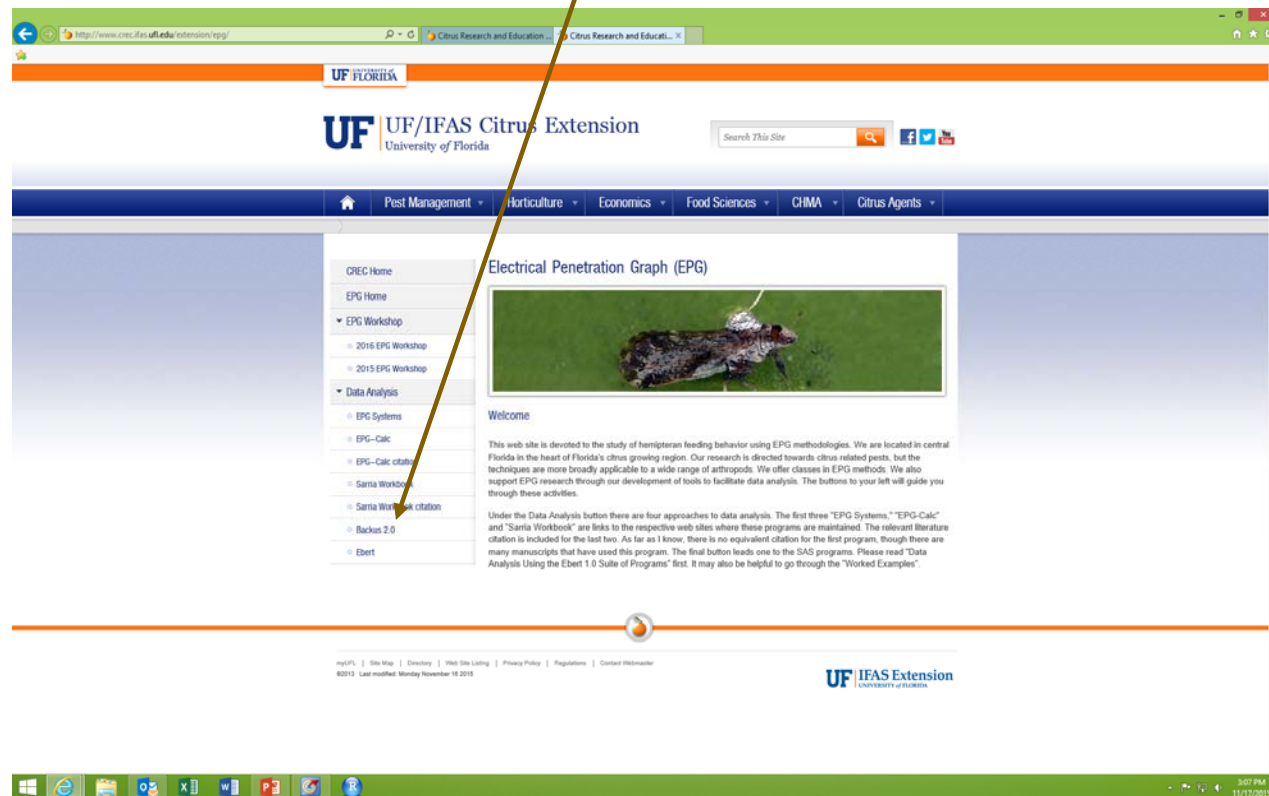
- It is assumed that:
  - You have a text file with all the data.
  - The data file has three columns of numbers.
    - Insect number
    - Waveform
    - Duration
  - The file is error free.
  - Completing lectures 1 through 8 is strongly suggested.
    - We will use the data file for psyllids that you should have following lectures 1-8 for windaq.

EPG  
Work-  
shop



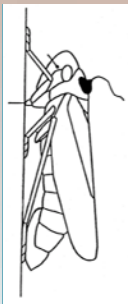
# Download Backus

- Go to:
  - <http://www.crec.ifas.ufl.edu/extension/epg/>
- Select Backus 2.0



The screenshot shows the UF/IFAS Citrus Extension website. The main content area is titled "Electrical Penetration Graph (EPG)" and features a photograph of a citrus thrips on a leaf. Below the photo is a "Welcome" message and a paragraph of text. The left sidebar contains a navigation menu with the following items: CREC Home, EPG Home, EPG Workshop (with sub-items for 2016 and 2015 EPG Workshops), Data Analysis (with sub-items for EPG Systems, EPG-Calc, EPG-Calc citation, Samia Workbook, Samia Workbook citation, Backus 2.0, and Ebert). A yellow arrow points from the "Backus 2.0" link in the sidebar to the "Backus 2.0" link in the main content area.

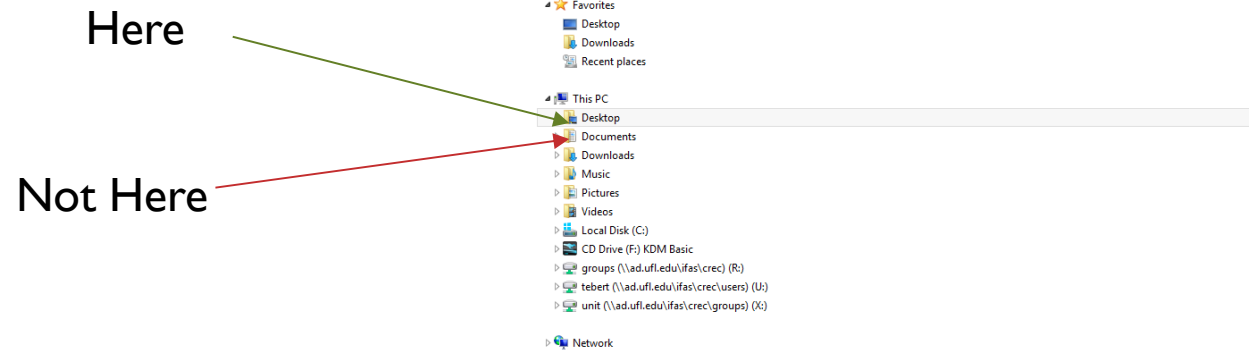
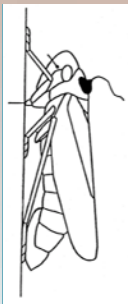
EPG  
Workshop



# Location

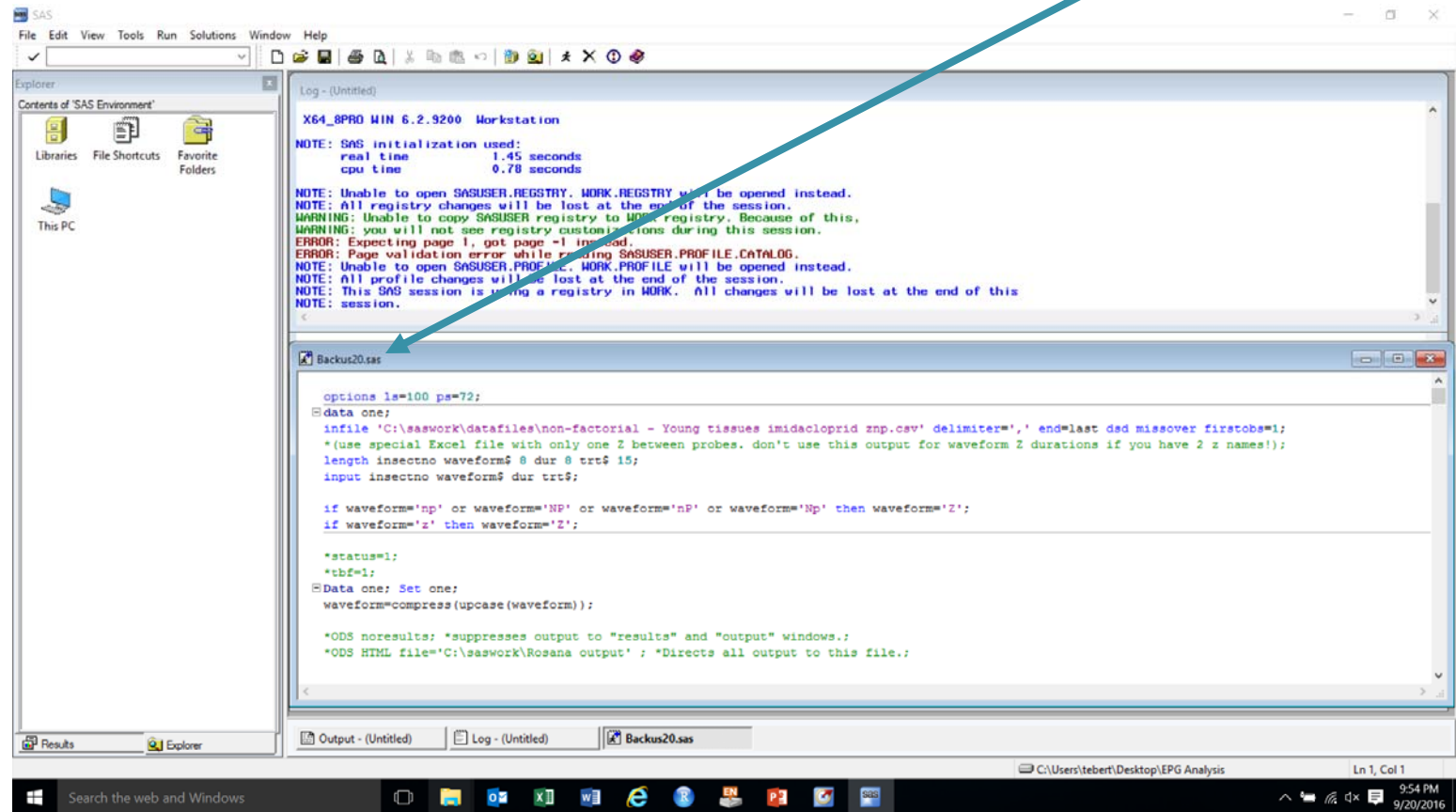
- **Use the folder that you created in Lecture 3:**
  - This should be a folder “EPG Analysis”
  - The folder should be on the desktop.

EPG  
Work-  
shop



# Open Backus 2.0

Open in SAS 9.4. Do not use Enterprise Guide.  
You should see this if you are in full screen mode.  
You will need to make several changes to the program in this window.



The screenshot shows the SAS 9.4 interface. The top menu bar includes File, Edit, View, Tools, Run, Solutions, Window, and Help. The Explorer window on the left shows the 'Contents of SAS Environment' with folders for Libraries, File Shortcuts, Favorite Folders, and This PC. The Log window (Log - (Untitled)) displays the following text:

```
X64_8PRO WIN 6.2.9200 Workstation
NOTE: SAS initialization used:
      real time      1.45 seconds
      cpu time       0.78 seconds

NOTE: Unable to open SASUSER.REGISTRY. WORK.REGISTRY will be opened instead.
NOTE: All registry changes will be lost at the end of the session.
WARNING: Unable to copy SASUSER registry to WORK registry. Because of this,
WARNING: you will not see registry customizations during this session.
ERROR: Expecting page 1, got page -1 instead.
ERROR: Page validation error while reading SASUSER.PROFILE.CATALOG.
NOTE: Unable to open SASUSER.PROFILE. WORK.PROFILE will be opened instead.
NOTE: All profile changes will be lost at the end of the session.
NOTE: This SAS session is using a registry in WORK. All changes will be lost at the end of this
NOTE: session.
```

The code editor window (Backus20.sas) contains the following SAS code:

```
options ls=100 ps=72;
data one;
infile 'C:\saswork\datafiles\non-factorial - Young tissues imidacloprid znp.csv' delimiter=';' end=last dsd missover firstobs=1;
*(use special Excel file with only one Z between probes. don't use this output for waveform Z durations if you have 2 z names!);
length insectno waveform$ 8 dur 8 trt$ 15;
input insectno waveform$ dur trt$;

if waveform='np' or waveform='NP' or waveform='nP' or waveform='Np' then waveform='Z';
if waveform='z' then waveform='Z';

*status=1;
*tbody=1;
Data one; Set one;
waveform=compress(upcase(waveform));

*ODS noresults; *suppresses output to "results" and "output" windows.;
*ODS HTML file="C:\saswork\Rosana output"; *Directs all output to this file.;
```

The taskbar at the bottom shows the Windows Start button, search bar, and various application icons. The system tray on the right shows the time as 9:54 PM on 9/20/2016.

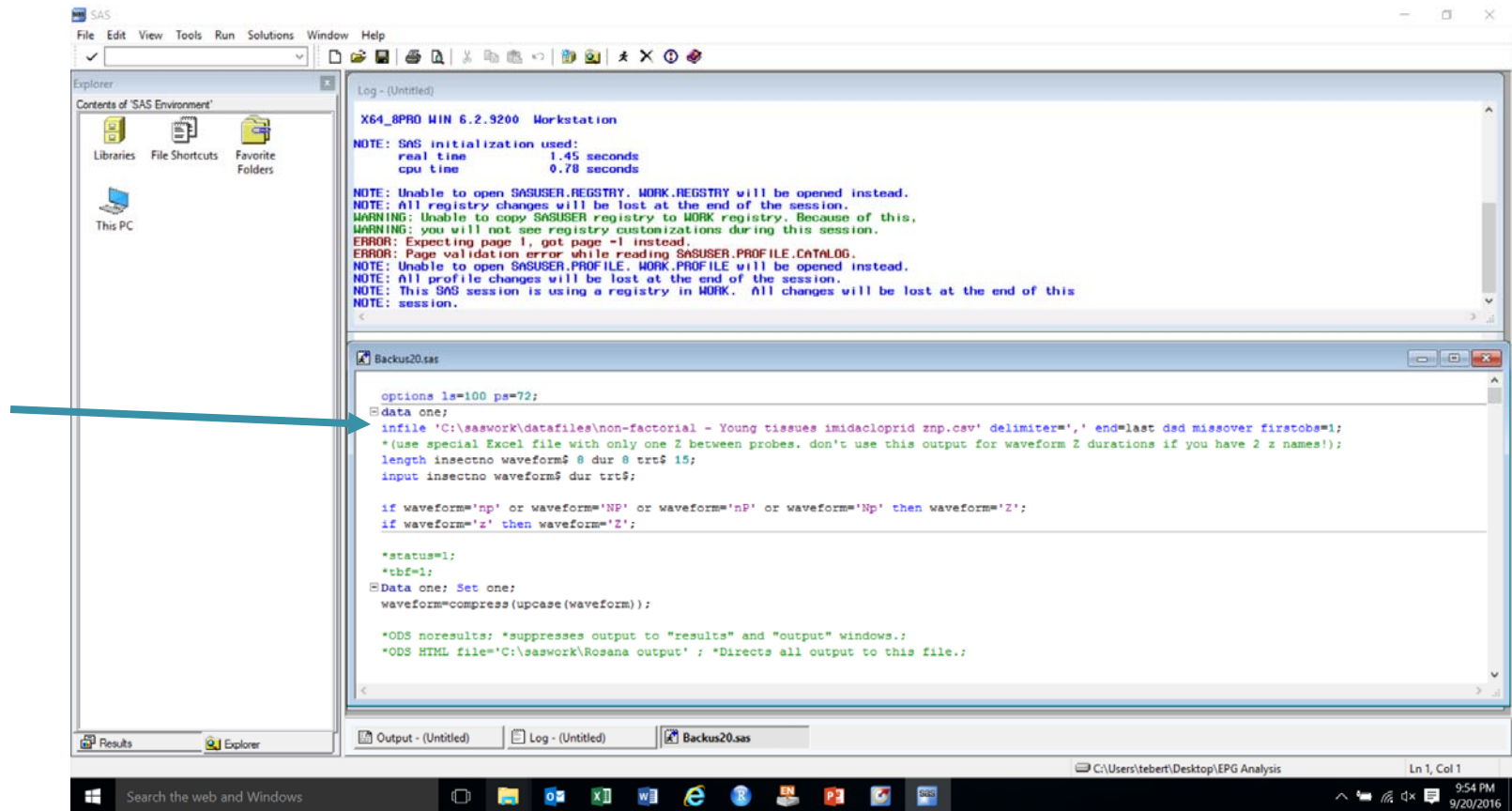
EPG  
Workshop



# Make Several Changes I

Change the file name.

At the end of the infile statement there is “firstobs=1”. Change this to “firstobs=2”



The screenshot shows the SAS interface with a log window and a code editor window. The log window displays the following text:

```
X64_8PRO WIN 6.2.9200 Workstation
NOTE: SAS initialization used:
      real time      1.45 seconds
      cpu time       0.78 seconds

NOTE: Unable to open SASUSER.REGISTRY. WORK.REGISTRY will be opened instead.
NOTE: All registry changes will be lost at the end of the session.
WARNING: Unable to copy SASUSER registry to WORK registry. Because of this,
WARNING: you will not see registry customizations during this session.
ERROR: Expecting page 1, got page -1 instead.
ERROR: Page validation error while reading SASUSER.PROFILE.CATALOG.
NOTE: Unable to open SASUSER.PROFILE. WORK.PROFILE will be opened instead.
NOTE: All profile changes will be lost at the end of the session.
NOTE: This SAS session is using a registry in WORK. All changes will be lost at the end of this
NOTE: session.
```

The code editor window shows the following SAS code:

```
options ls=100 ps=72;
data one;
infile 'C:\saswork\datafiles\non-factorial - Young tissues imidacloprid znp.csv' delimiter=';' end=last dsd missover firstobs=1;
*(use special Excel file with only one Z between probes. don't use this output for waveform Z durations if you have 2 z names!);
length insectno waveform$ 8 dur 8 trt$ 15;
input insectno waveform$ dur trt$;

if waveform='np' or waveform='NP' or waveform='nP' or waveform='Np' then waveform='Z';
if waveform='z' then waveform='Z';

*status=1;
*cbf=1;
Data one; Set one;
waveform=compress(upcase(waveform));

*ODS noresults; *suppresses output to "results" and "output" windows.;
*ODS HTML file="C:\saswork\Rosana output" ; *Directs all output to this file.;
```

A blue arrow points to the 'firstobs=1' parameter in the SAS code.

EPG  
Work-  
shop





# Make Several Changes II

Change these two lines

```
length insectno waveform$ 8 dur 8 trt$ 15;  
input insectno waveform$ dur trt$;
```

To this by deleting the variable “trt”

```
length insectno waveform$ 8 dur 8;  
input insectno waveform$ dur;
```

EPG  
Work-  
shop

A screenshot of a SAS code editor window titled 'Backus20.sas'. The code is as follows:

```
options ls=100 ps=72;  
data one;  
infile 'C:\saswork\datafiles\non-factorial - Young tissues imidacloprid znp.csv' delimiter=';' end=last dsd missover firstobs=1;  
*(use special Excel file with only one Z between probes. don't use this output for waveform Z durations if you have 2 z names!);  
length insectno waveform$ 8 dur 8 trt$ 15;  
input insectno waveform$ dur trt$;  
  
if waveform='np' or waveform='NP' or waveform='nP' or waveform='Np' then waveform='Z';  
if waveform='z' then waveform='Z';  
  
*status=1;  
*cbf=1;  
Data one; Set one;  
waveform=cbf*status (inbase (waveform)) .;
```

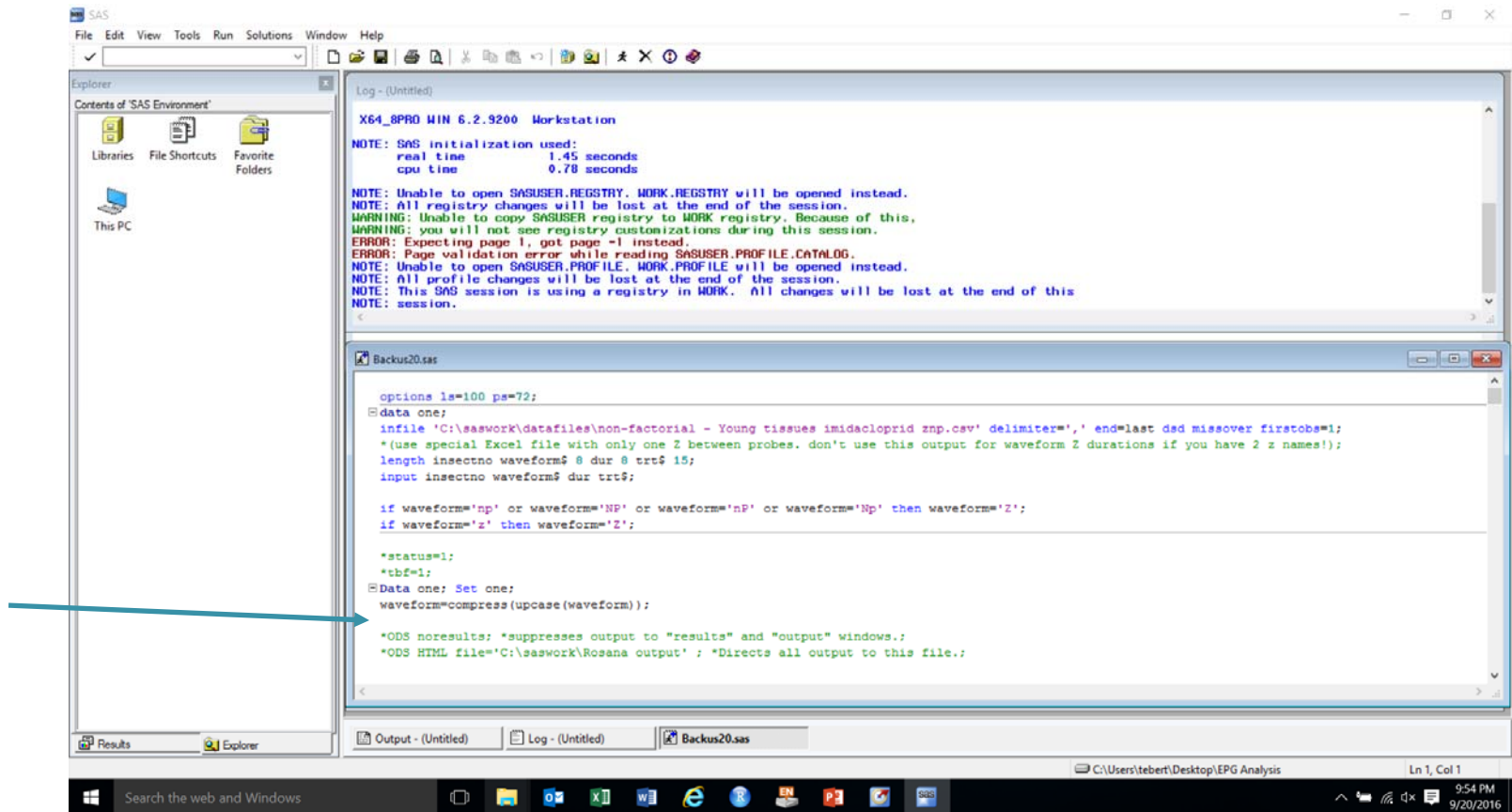
A blue arrow points from the left margin to the line `length insectno waveform$ 8 dur 8 trt$ 15;` in the code editor. Above the code editor, there is a log window with the following text:

```
ERROR: Page validation error while reading SASUSER.PROFILE.CATALOG.  
NOTE: Unable to open SASUSER.PROFILE. WORK.PROFILE will be opened instead.  
NOTE: All profile changes will be lost at the end of the session.  
NOTE: This SAS session is using a registry in WORK. All changes will be lost at the end of this  
NOTE: session.
```

# Last Change

Add the following at the location pointed to by the arrow.

```
trt=substr(insectno,1,1);
```



The screenshot shows the SAS software interface. The main window displays the SAS program 'Backus20.sas'. A blue arrow points to the location where the code 'trt=substr(insectno,1,1);' should be added. The program code is as follows:

```
options ls=100 ps=72;
data one;
infile 'C:\saswork\datafiles\non-factorial - Young tissues imidacloprid znp.csv' delimiter=';' end=last dsd missover firstobs=1;
* (use special Excel file with only one Z between probes. don't use this output for waveform Z durations if you have 2 z names!);
length insectno waveform$ 8 dur 8 trt$ 15;
input insectno waveform$ dur trt$;

if waveform='np' or waveform='NP' or waveform='nP' or waveform='Np' then waveform='Z';
if waveform='z' then waveform='Z';

*status=1;
*tbody=1;
Data one; Set one;
waveform=compress(upcase(waveform));

*ODS noresults; *suppresses output to "results" and "output" windows.;
*ODS HTML file="C:\saswork\Rosana output" ; *Directs all output to this file.;
```

EPG  
Work-  
shop



# An Explanation

The command “firstobs=2” tells SAS that it should start reading data in line 2. SAS then reads line 1 and ignores the result.

The data file has only three columns of data. There is no variable “trt” in the data.

Adding this command

```
trt=substr(insectno,1,1);
```

tells SAS to create a variable “trt” using the first character in the variable “insectno”. The command substr is short for sub-string. There are three arguments to this function. The first is the string. The second value is the starting character. In this case we start with the first character. The third value is the number of characters.

If you want treatments to always be three character codes and you have insect numbers like ABC1,ABC2,TTS1,TTS2 then

EPG  
Work-  
shop



# Run the program

- You will now see something like this. The left side is a navigation aid to jump to whatever result you are interested in viewing.

EPG  
Work-  
shop



Navigation

Results

The screenshot shows the SAS Results Viewer window. The left pane is labeled 'Navigation' and contains a tree view of results. The main pane is labeled 'Results' and displays the following tables:

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	1	176	13.53	0.0003

trt Least Squares Means					
trt	Estimate	Standard Error	DF	t Value	Pr >  t
a	6.6384	0.1344	176	49.41	<.0001
b	7.5273	0.2009	176	37.47	<.0001

Differences of trt Least Squares Means					
trt_trt	Estimate	Standard Error	DF	t Value	Pr >  t
a b	-0.8889	0.2417	176	-3.68	0.0003

T Grouping for trt Least Squares Means (Alpha=0.05)  
LS-means with the same letter are not significantly different.

trt	Estimate	
b	7.5273	A
a	6.6384	B

# Navigating the results.

- To view specific output, double click at position indicated by arrows in the order (top to bottom) shown.

The screenshot shows the SAS Results Viewer interface. The left pane displays a tree view of results, and the right pane displays a summary table. Three blue arrows point to the following items in the tree view:

- Model Information
- Class Level Information
- Number of Observations

The summary table in the right pane contains the following information:

Number of Observations Read	6
Number of Observations Used	6
Dimensions	
Covariance Parameters	1
Columns in X	3
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	6
Optimization Information	
Optimization Technique	None
Parameters	3
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled
Fit Statistics	
-2 Res Log Likelihood	7.71
AIC (smaller is better)	13.71
AICC (smaller is better)	37.71
BIC (smaller is better)	11.87
CAIC (smaller is better)	14.87
HQIC (smaller is better)	9.67
Bayesian Priors	0.96

EPG  
Workshop



# Finishing this section

- Select tables in the results window, and then copy-paste them into a program of your choice.
- Until you become familiar with your arthropod, it is a good idea to examine most of the output.
  - Especially compare the results from transformed versus untransformed data.

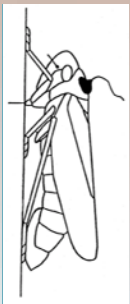
EPG  
Work-  
shop



# Examining Model Fit.

- The default in Proc GLIMMIX is a Gaussian (=Normal) distribution.
- You can change this to any of several options as described in the SAS user guides.
- Examining the residuals helps you decide if the model fits the data, and how to improve the statistical model.

EPG  
Work-  
shop



# Q-Q Plot

- Many statistical models require that the residuals are normally distributed.
- Did the transformation improve model performance relative to this issue?

EPG  
Work-  
shop





# Activating ODS graphics

- Enter “ODS graphics on;” here:

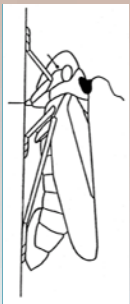
```
Data one; Set one;  
waveform=compress(upcase(waveform));  
trt=substr(insectno,1,1);
```

Add this → `ods graphics on;`

- Enter “ODS graphics off;” here towards the end of the program:

```
*ods html close;  
Add this → ods graphics off;  
run;  
quit;
```

EPG  
Work-  
shop



# Graphics

- You need the following command:

```
Plots=residualpanel
```

- It is entered every place where there is a `proc glimmix` statement.
- So you should now have this at multiple places in the program:

```
proc glimmix plots=residualpanel;
```

- Now run the program (F3 key)

EPG  
Work-  
shop

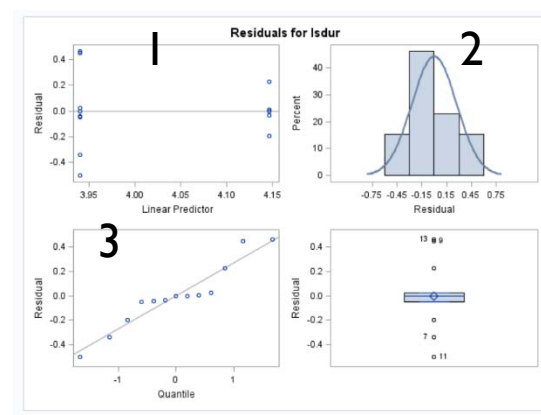
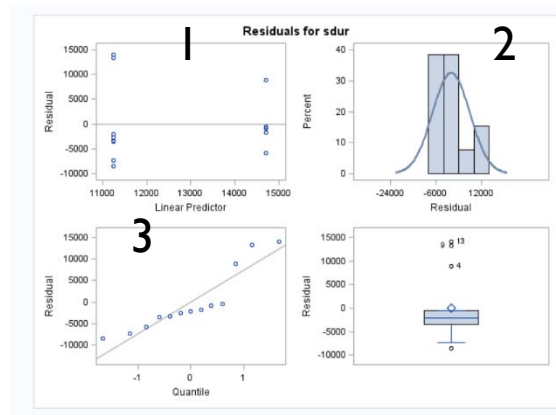


# Things to look for

(This is WDP-E2 being used here)

- There are two treatments, so the residuals are in two bands (1).
- The transformation helped a little (2), but in a perfect world they would all align with the diagonal line (3).
- There were a total of 10 insects (5/treatment), but there are 13 data points (3). Why?

EPG  
Work-  
shop

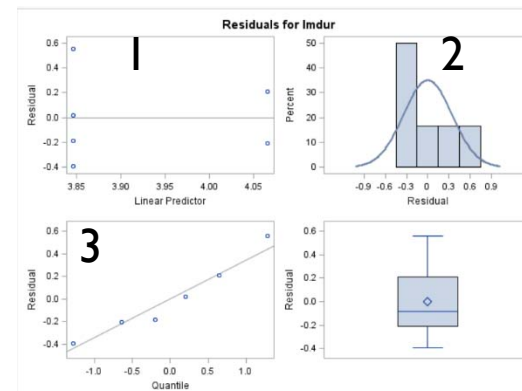
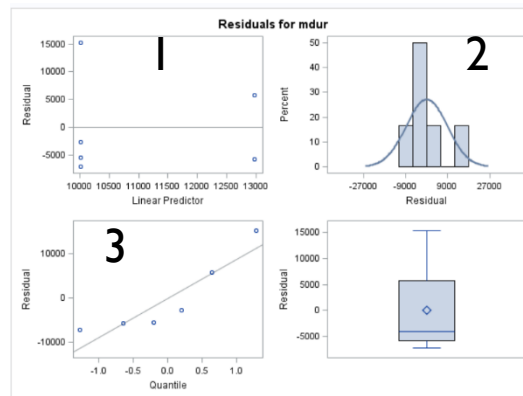


# Things to look for

(This is WDEI-E2 being used here)

- There are two treatments, so the residuals are in two bands (1).
- The transformation helped a little (2), and the result is a bit better than for WDP-E2.
- There were a total of 10 insects (5/treatment), but there are six data points (3). Why?

EPG  
Work-  
shop





# Critical Concepts

- Changing SAS code to read your file.
- Firstobs= command to adjust for headers in data file.
- Naming insects with treatments and having SAS get the treatment names.
- Turning on ODS graphics.
- Getting plots of residuals.
- Interpreting residual plots.



# Quiz

- For WDP-E2, why are there 3 data points but 10 insects?
- For WDEI-E2, why there are 6 data points but 10 insects?
- I want treatment names like ABC, and GMP. I will always have three character designations, so insect numbers are things like ABC1, ABC2, ABC3, ABC4, and so forth. How do I change “`trt=substr(insectno, 1, 1);`”?



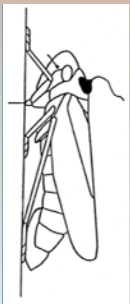
This slide is left blank

- Answers to questions follow

# The answers to the questions

- In the first case, probe is being treated as the experimental unit. So there are 13 probes with one or more E2 events.
  - One approach to dealing with pseudoreplication is to use insect number as a variable in your model.
  - Another approach is to use averaging to get “by insect” values.
  - A third approach would be to use repeated measures designs. This option is not currently supported.
- In the second case there are 4 insects that do not have an E2. These are treated as missing values, leaving 6 data points.
- `trt=substr(insectno, 1, 3) ;`

EPG  
Work-  
shop





# The end

- You should strongly consider working through the other lectures even if you decide to stick with the Backus program.
  - What sample size should I use?
  - Finding errors in the data.
  - Other analyses like discriminant analysis.

Because Backus and Ebert are both written in SAS, they are somewhat interchangeable. What one does, you can get the other to do as well. Thus, understanding how one gets Ebert 1.0 to do a discriminant analysis enables you to do the same type of analysis in Backus 2.0.

EPG  
Work-  
shop

