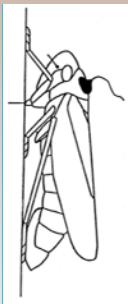


Error Checking: The Issues

- There are many ways problems can arise in the data. Here is a list of those problems for aphids.
 - The only behavior allowed after Np is A.
 - Unless A is recorded as part of C.
 - The only behavior allowed before E2 is E1.
 - No consecutive waveforms with the same behavior.
 - There can only be one non-probing behavior. There are work-arounds for this in some cases.
 - All recordings start with the non-probing behavior.
 - All durations are positive.

EPG
Work-
shop





Error Checking: outline

- The relevant program is “Error Checker”
- Problems are present when the program fails to produce the correct output, or gives an error message.
- The table that the program should produce helps identify problems with inappropriate transitions: e.g. Np going directly to E2.



Run the Program

- Open the program “Error Checker.”
- Change the file name in the infile statement.
- Make sure the *AphidData1.csv* file is not open in any other program.
- Run the program.
- These data have no errors that the program can detect.

The Output

- If there are no errors, then the output will have two tables at the top.

The screenshot displays the SAS Enterprise Guide interface. The main window shows the output of the FREQ procedure, which includes three tables. The first table is the 'Frequency Table of Waveform Event Transitions' (The FREQ Procedure), showing the frequency and percent of each waveform event. The second table is another 'Frequency Table of Waveform Event Transitions' (The FREQ Procedure), showing the frequency and percent of each transition between waveform events. The third table is 'Duration by waveform Output' (The MEANS Procedure), showing the mean and median duration for each waveform event.

waveform	Frequency	Percent	Cumulative Frequency	Cumulative Percent
C	1120	49.62	1120	49.62
E1	20	0.89	1140	50.51
E2	9	0.40	1149	50.91
F	6	0.27	1155	51.17
G	2	0.09	1157	51.26
NP	95	4.21	1252	55.47
PD	1005	44.53	2257	100.00

trans1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
C to E1	18	0.80	18	0.80
C to F	6	0.27	24	1.07
C to G	2	0.09	26	1.16
C to NP	87	3.87	113	5.02
C to PD	1005	44.69	1118	48.71
E1 to C	11	0.49	1129	50.20
E1 to E2	9	0.40	1138	50.60
E2 to C	5	0.22	1143	50.82
E2 to E1	2	0.09	1145	50.91
F to C	6	0.27	1151	51.18
G to C	2	0.09	1153	51.27
NP to C	93	4.14	1246	55.40
PD to C	1002	44.55	2248	99.96
PD to NP	1	0.04	2249	100.00

Insectno=a1 waveform=C				
Analysis Variable : dur				
N	Minimum	Maximum	Mean	Median
66	0.0600000	599.7700000	65.8074242	23.2000000

Insectno=a1 waveform=NP				
Analysis Variable : dur				
N	Minimum	Maximum	Mean	Median
45	0.8600000	1231.31	108.3911111	11.7900000

Insectno=a1 waveform=PD				
Analysis Variable : dur				
N	Minimum	Maximum	Mean	Median
21	4.0300000	719.7800000	47.1342857	5.5100000

Insectno=a2 waveform=C				
Analysis Variable : dur				
N	Minimum	Maximum	Mean	Median
66	0.0600000	599.7700000	65.8074242	23.2000000

Reading the output tables

- The first table

waveform	Frequency	Percent	Cumulative	Cumulative
			Frequency	Percent
C	1120	49.62	1120	49.62
E1	20	0.89	1140	50.51
E2	9	0.4	1149	50.91
F	6	0.27	1155	51.17
G	2	0.09	1157	51.26
NP	95	4.21	1252	55.47
PD	1005	44.53	2257	100

- Waveform is a list of all waveforms in the file. Capitalization matters, so NP is not the same as Np in this table.
- Frequency is the number of times that waveform appears in the data.
- Percent is the relative contribution of each waveform to the total of all observed behaviors.



The first table

- This table is useful for finding typos for Windaq users. Windaq will not consider waveform X as a mistakenly entered version of waveform C.

The second table

- Look carefully at all the transitions (first column).

transl	Frequency	Percent	Cumulative Frequency	Cumulative Percent
C to EI	18	0.8	18	0.8
C to F	6	0.27	24	1.07
C to G	2	0.09	26	1.16
C to NP	87	3.87	113	5.02
C to PD	1005	44.69	1118	49.71
EI to C	11	0.49	1129	50.2
EI to E2	9	0.4	1138	50.6
E2 to C	5	0.22	1143	50.82
E2 to EI	2	0.09	1145	50.91
F to C	6	0.27	1151	51.18
G to C	2	0.09	1153	51.27
NP to C	93	4.14	1246	55.4
PD to C	1002	44.55	2248	99.96
PD to NP	1	0.04	2249	100

- The first row shows that there are 18 cases where the aphid was in C and went directly to EI with no pd.



One last error

- This process does not check to make sure that all recordings start with N_p .
- This is a relatively minor error and will have minimal consequences.
- The fastest fix is to go through Excel.



Start with NP

- Open AphidData1.csv in Excel.
- Obviously the first insect has this problem. Do not correct it yet.
- In column D, cell 2 type in the formula
`=IF(A2=A1,"",IF(B2="C",1,""))`
- The number 1 is displayed in the cell.
- Move cursor to cell C2 and hit “End” and then ↓.
- Move cursor to cell D2258, hit “Shift End↑” to select all cells (you have to hold the shift key, the end key does not have to be held).



Finishing

- Fill down (Control D)
- In cell D1 type in =Sum(D2:D2275)
- The result is one, so only the first insect is a problem.
- Solution: If NP is not an important behavior to your research, I would suggest adding an NP to the first insect and selecting a duration of 0.5 seconds.
- Delete column D, and save the file.