Ask and Ye Shall Receive: Getting the Most from SAS-L

Howard Schreier
U.S. Dept. of Commerce

1

Purpose

This paper offers a number of suggestions, primarily directed at those asking questions on SAS-L. Following these suggestions may make questions clearer and easier to analyze, thus increasing the chances that prompt, appropriate, and useful responses will follow.

2

Access alternatives

- comp.soft-sys.sas
- sas-L@listserv.uga.edu (listserv@listserv.uga.edu)
- http://groups.google.com/groups ?group=comp.soft-sys.sas
- http://listserv.uga.edu/archives/sas-l.html
- others

3

Newsgroup and LISTSERV mechanisms

- Internet FAQ Consortium, *Usenet References*, http://www.faqs.org/usenet/
- L-Soft international, Inc., *General User's Guide to LISTSERV*, http://www.lsoft.com/manuals /1.8d/user/user.html

4

Typical exchange

- A problem
- A solution

5

Questions

- Have you done your homework?
- Is SAS-L the best place?
- Will it be noticed?
- Can it be understood?
- Will it be understood?
- Can people easily replicate your results, and do experiments?

6

Homework: Resources

- SAS software documentation
- SAS-L discussion archives
- SUGI (etc.) proceedings
- Technical and user-support materials at www.sas.com
- Your “laboratory”

7

Homework: Reality check

- You may not find the answer.
- Explain (briefly) where you looked and what you tried.

I looked at the examples in the PROC FREQ section of the Procedures Guide, but none of them showed how to left justify the output.

8

Questions

- Have you done your homework?
- **Is SAS-L the best place?**
- Will it be noticed?
- Will it be understood?
- Is the essence of the problem apparent?
- Can people easily replicate your results, and do experiments?

9

Is SAS-L the best place?

- Yes, if it’s a SAS question
- Otherwise, consider other lists and newsgroups.
- They may have bigger pools of expertise on non-SAS subjects.

10

Questions

- Have you done your homework?
- Is SAS-L the best place?
- **Will it be noticed?**
- Will it be understood?
- Is the essence of the problem apparent?
- Can people easily replicate your results, and do experiments?

11

Will it be noticed?

12

Will it be noticed?

- Subject:
- Subject: Please Help
- Subject: SAS Question
- Subject: PROC SUMMARY Options
- Subject: Counting Bluedoos
- Subject: Conditional Counting

13

Questions

- Have you done your homework?
- Is SAS-L the best place?
- Will it be noticed?
- **Can it be understood?**
- Will it be understood?
- Can people easily replicate your results, and do experiments?

14

Can it be understood?

- Define the environment.
- Describe the problem.
- Illustrate the problem.

15

Define the environment

- SAS version number
- SAS products licensed
- SAS system options in effect
- Host operating system
- Hardware
- Network configuration
- Specifics on non-SAS software products
- Constraints

16

Reality check

- It's highly unlikely that all of these are relevant.

17

Can it be understood?

- Define the environment.
- **Describe the problem.**
- Illustrate the problem.

18

Describe the problem

- Data
- Requirements
- Efforts made
- Difficulties encountered

Try to be consistent with SAS documentation in your use of terminology.

19

Can it be understood?

- Define the environment.
- Describe the problem.
- **Illustrate the problem.**

20

Illustrate the problem

- With real data
- With fabricated data

21

Illustrate the problem : Include

- Data
- SAS code
- Excerpts from the SAS log
- Results generated
- Results expected

22

Illustrate the problem: Good practice

- Copy and paste.
- Edit log excerpts.

23

Questions

- Have you done your homework?
- Is SAS-L the best place?
- Will it be noticed?
- Can it be understood?
- **Will it be understood?**
- Can people easily replicate your results, and do experiments?

24

Will it be understood?

- Miniaturize: make the datasets compact.
- "Zoom in": leave out upstream and downstream steps which are not immediately relevant.
- Abstract: minimize use of specialized terminology.

25

Trade-Offs Ahead

26

Miniaturize

- Fewer observations
- Fewer variables
- Fewer keys (BY variables)
- Fewer key value combinations (BY groups)

27

Miniaturize

- Use subset of "real" data
- Fabricate data

28

Miniaturize

- But don't lose generality
  - Example: BY groups
  - Example: missing values
- Note actual scale

The real data set has seven categorical variables, 40 response variables, and about 2 million observations. Two of the categorical variables and most of the response variables have some missing values.

29

Will it be understood?

- Miniaturize: make the presentation compact.
- **"Zoom in": leave out upstream and downstream steps which are not relevant.**
- Abstract: minimize use of specialized terminology.

30

Zoom in: From this

```
data ds1;
...
proc this data=ds1 out=ds2;
...
proc that data=ds2 out=ds3;
...
proc etc data=ds3 out=ds4;
...
data ds5;
set ds4;
...
```

31

Zoom in: To this

```
data ds4;
...
cards;
...
...
;
data ds5;
set ds4;
...
```

32

Zoom in: compensate for lost context

The input dataset is extracted from a remote transactions database by a scheduled job which stores a date-stamped flat file on our LAN early each morning. The summary file I'm trying to build will be used to generate a set of tables in our management monitor system. Each department likes to see its own day-to-day track but only needs current-year cumulatives for other departments.

33

Will it be understood?

- Miniaturize: make the presentation compact.
- "Zoom in": leave out upstream and downstream steps which are not relevant.
- **Abstract: minimize use of specialized terminology.**

34

Abstraction

- Use generic variable names; for example, "GroupID" rather than "SIC" (for "self-identified cohort").
- Note subject matter context.

The real data set is drawn from a credit card transactions database. The group ID variables are based on age brackets, as self-reported on a sweepstakes entry form.

35

Will it be understood?: The tradeoffs

- Miniaturization vs.
  - Generality
  - Scale
- Zoom-in vs. process context
- Abstraction vs. subject matter context

36

Questions

- Have you done your homework?
- Is SAS-L the best place?
- Will it be noticed?
- Can it be understood?
- Will it be understood?
- **Can people easily replicate your results, and do experiments?**

37

Can people experiment?: CPR helps

- Copy
- Paste
- Run

38

Can people experiment?: Non-CPR

Customer Num	Transaction Date	Amount
22	20030330	230
21	20030401	88
22	20030401	1000
21	20030402	335

39

Can people experiment?: CPR

```
data start_with;
input CustomerNum
 TransactionDate :date9. Amount;
format TransactionDate yymmddn8.;
cards;
22 30MAR2003 230
21 01APR2003 88
22 01APR2003 1000
21 02APR2003 335
;
```

40

Can people experiment?: data delivery

- In-line (CARDS;)
- Data generator
- E-mail
- Web site

41

Questions

- Have you done your homework? ✓
- Is SAS-L the best place? ✓
- Will it be noticed? ✓
- Can it be understood? ✓
- Will it be understood? ✓
- Can people easily replicate your results, and do experiments? ✓

42

Follow up

- No responses: restate and elaborate
- Responses going in the wrong direction: clarify
- Questions or request(s) for details
- Mid-course progress report
- Wrap up: share useful suggestions received via private e-mail.

But please don't change the subject.

43

Reality check

- Don't follow all of these suggestions.
- Recognize the trade-offs.
- You almost need to know the answer in order to frame the question optimally.
- Do your best.

44

About the speaker

Howard Schreier

U.S. Dept. of Commerce
Washington DC 20230

(202) 482-4180

Howard_Schreier@ita.doc.gov

45