Shafi Consultancy

Quick Guide to Good Programming Practice

1 Getting started:

- Create an initialisation program e.g. **autoexec.sas** (called automatically) at the start of all programs:
  - Define **all libnames** and standard **options**, including **NOFMTERR** and **MSGLEVEL=I**
  - No libnames should be defined in individual programs.
  - Define all **study level macro variables**
  - Include **global** and **study level macros** and programs containing macros
  - Include a **format program** containing all study formats, e.g. **studyfmt.sas**
- Create **Abbreviations** for header, comment and section boxes for consistency.

2 Programming style:

- Use a **standard program header**. It should contain names of program, programmer and date first completed, purpose and revision history.
- When writing program **keep in mind** that **it will need to be changed**. Can you follow it after one year?
- Use different Comment style to split programs into **sections** (input, process, output) so they stand out in the program. Write comments in the form * **Comments** *; and **avoid** /* comments */, ideally within a box. If an existing program has /* comments */ then use **%MACRO donotuse**; and %MEND donotuse; to comment out sections of the program.
- Explain algorithm and logic in a comment box just before data steps and procedures.
- **Avoid inserting comments inside DATA steps** or PROC SQL steps, put them in a box above the data step. It makes the program easier to read.
- Set up program into INPUT, PROCESS and OUTPUT sections.
- Read in all external data ONLY ONCE. Do this at the top of the program in the INPUT section.
- Use **"KEEP="** to list variable names when reading (INPUT) and writing (OUTPUT) external datasets.
- Only write **ONE** SAS statement per line.
- Use **different meaningful names** for each temporary dataset.
- Use indentation. **Indent 2 characters** (or one tab), but remain consistent throughout the program.
- Use **Uppercase** for SAS KEYWORDS to make programs easier to read.

3 Check LOG:

- If it contains **Cartesian product** after a PROC SQL statement, check the dataset produced very carefully. It will most likely have more observations than expected, and therefore **may be incorrect**.
- Check for **ERROR, WARNING, uninitialized** and **repeats of BY values**.
- "Repeats of BY values" appears as a NOTE, but it **must be treated as an ERROR** and the code changed.
- **Variables overwritten during MERGE** (appears when MSGLEVEL=I is used) should also be corrected.

www.shaficonsultancy.com

Tel: +44 770 288 7219 E-Mail: info@shaficonsultancy.com