

S-PLUSの企業ユーザー事例

SASからS-PLUSへ



栄研化学株式会社

情報システム部

稲葉 弥一郎

目次

- 何故S - PLUSを選んだのか
- S - PLUSの良いところ
- どの様にシステム化したか
- どの様にSPLUSに移行したか

何故S - PLUSを選んだか

- SASのユーザーであった
- SASは不要な情報の出力が多かった
 - とにかく紙の量が多い
- グラフ作成が煩雑で面倒だった
 - とにかく面倒くさい
- SASのプロシジャーで決められた順にしか出力できなかった

S-PLUSの良いところ 1

- 必要な情報のみをセレクトして出力できる
 - 紙の量が少なくすむ(わかりやすい表)
- 解析の順番として、全体から細部へのブレイクダウンがやりやすい
 - 解析結果をObjectとして保存ができる
- 複数のグラフパターンを、同一用紙・画面に簡単に出せる
 - 全体を俯瞰し易い形に出来る

S-PLUSの良いところ 2

- ユーザー関数が自由自在に作れる
 - SASはマクロであって関数ではない
 - 見やすい出力関数を作り他の仕事に応用
 - 組み合わせて使う検定を一つの関数に纏める
- データの互換性が高い
 - UNIX ↔ Windows 移行が簡単
 - もっとも、データがCSVなら似たようなもの
 - S-PLUSのソースは、UNIX, Windows完全互換

S-PLUSの良いところ 3

- GUIでは探索的な解析・グラフ化は簡単
 - 但し、定型化したものはスクリプト化する事
 - GUIで処理したものをスクリプトとして保存できる。(うまく使うと便利)
- 基本は、スクリプトベースと心得る
 - スクリプトは処理の再現性を保つ最良の方法
 - データ加工は、スクリプトの方が遥かに簡単

どの様にシステム化したか 1

- データのレイアウトを確定しておく
 - 当社はCSV形式と決めている
 - どんなシステムでも読み込める
- 変数名を、分かり易い名前を決める
 - 会社や業界で銘々規約は違う
- レイアウトの定義表をファイル化しておく
 - レイアウト定義体として使えるように、CSV形式にしておくこと

どの様にシステム化したか 2

- カテゴリカルデータは全てマスター化する
 - 全てテキストファイル化
 - 例 1, 1: Mail 2, 2: Femail
- レイアウト定義体とマスターを使って、データ読み込みのスク립トを自動生成する
 - PerlやAWKを使えば簡単にできる
- スクリプトの銘々規約を作る
 - 後でどの様なスク립トを作ったかがわかる

どの様にシステム化したか 3

- スクリプト毎に簡単な説明書を作成する
 - 後々何をしているかが分かるようにする
- スクリプト全体の実行手順書を作成する
 - 処理の再現性を保つ為
 - バリデーションとして、絶対に必要
- 手順書通りに実行するスクリプトを作る
 - 簡単に作れる

どの様にSPLUSに移行したか

- 新しい事を覚えることは、嫌がる人が多い
 - 特に別の方法を習熟していると尚更である
 - チャレンジ精神が必要
- 第一ステップ
 - S-PLUSを言語として覚える
 - データ加工技術を磨く
 - 業務として使えるレベルまで
 - 解析手法のスキプトの使い方を覚える
 - 統計が分かっていたら、何とかなる

どの様にSPLUSに移行したか

● 第二ステップ

- SASプログラムをS-PLUSに書き換える
 - 出来たら、業務プログラム一式を対象とする
- そのまま書き換えるのは駄目
 - 解析結果は見やすい順番に
 - 解析結果は必要な物だけを出す
 - グラフも見やすい形、順番に出す
- SASの結果と比較する
- S-PLUSを選んだ意味がこれでわかる

現在及び今後

● 現在

- ネットワークトラフィックを解析する予定
- 臨床評価研究会 (ACE) での勉強会
 - 実務者向けのマニュアルが無いので結構大変です

● 今後 (実力が伴えばやりたいこと)

- データ・マイニングツールとして使う
 - 販売情報のDWHが構築完了後
- 販売予測に使いたい (実力が有れば)
 - 生産計画策定に使えるレベルまで
- その他の依頼された統計解析