

# SASにおける文芸的プログラミングへの試み

○吹谷 芳博<sup>1</sup>

(<sup>1</sup>株式会社エスアールディ)

## Attempt to Literate Programming Using SAS

Yoshihiro Fukiya

Data Science Department, SRD Co., Ltd.

## 要旨:

プログラミング開発においてチームの相互理解を高めるため関連文書の作成の重要だが、その反面作成に多くの負担をかけてしまう。

この解決策として「文芸的プログラミング」の導入を行った内容を報告する。

キーワード: 文芸的プログラミング, markdown, knitr, LaTeX

- 目的
- 背景
- 設定条件と方法
  - 処理フローの説明
  - R markdownの構造とchunkの指定
- 評価(結果)
- 結論と考察

## 目的(課題解決)

ソースコードとその関連文書の作成の効率化へ



文芸的プログラミングを用いて課題を解決したい



R言語のパッケージと併用することで実装可能



## 背景

## SASプログラム開発の効率化

開発時や運用時の手間をできるだけ最小限に!!  
可読性を高める工夫や情報共有など.



プログラムの管理・再利用(前回SASユーザー総会)  
- バージョン管理システムで対応(Gitなど)

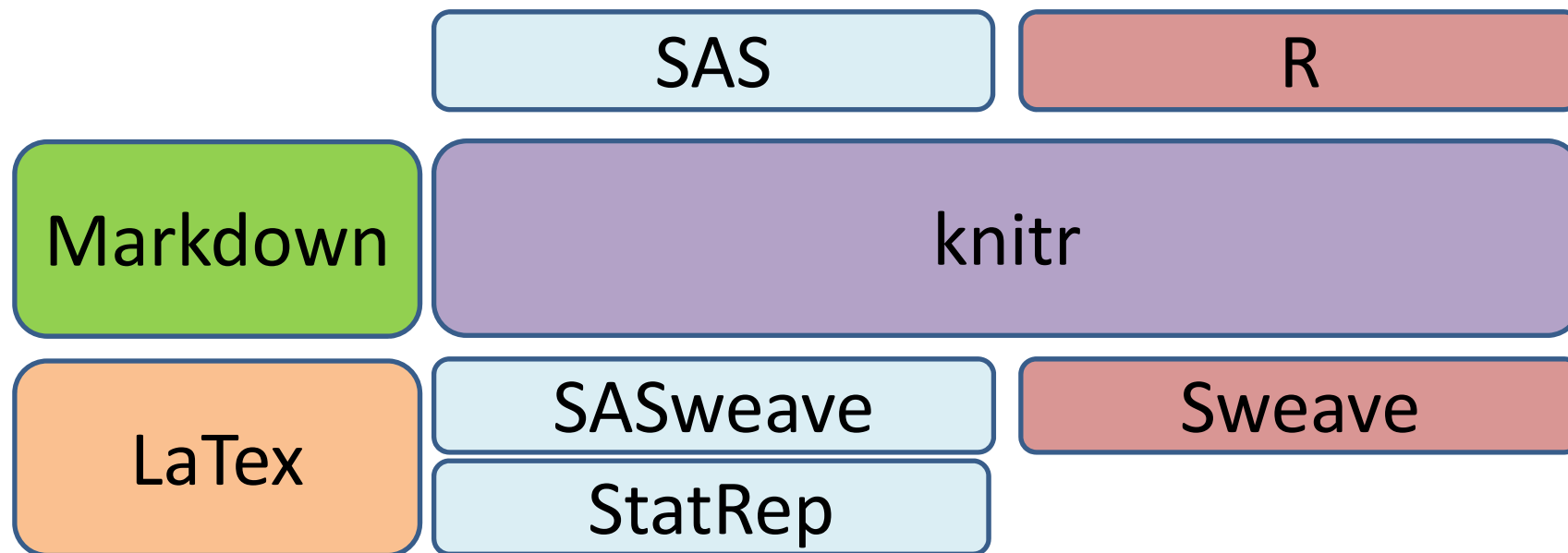
関連文書作成の軽減や整合性を担保(今回の発表)  
- **文芸的プログラミング**で対応

## 文芸的プログラミングとは？

ソースコードと関連文書（図表など）  
を一緒に生成させる方法.

1. 情報を一体化させることで整合性を高める.
2. プログラム開発以外の作業を軽減.

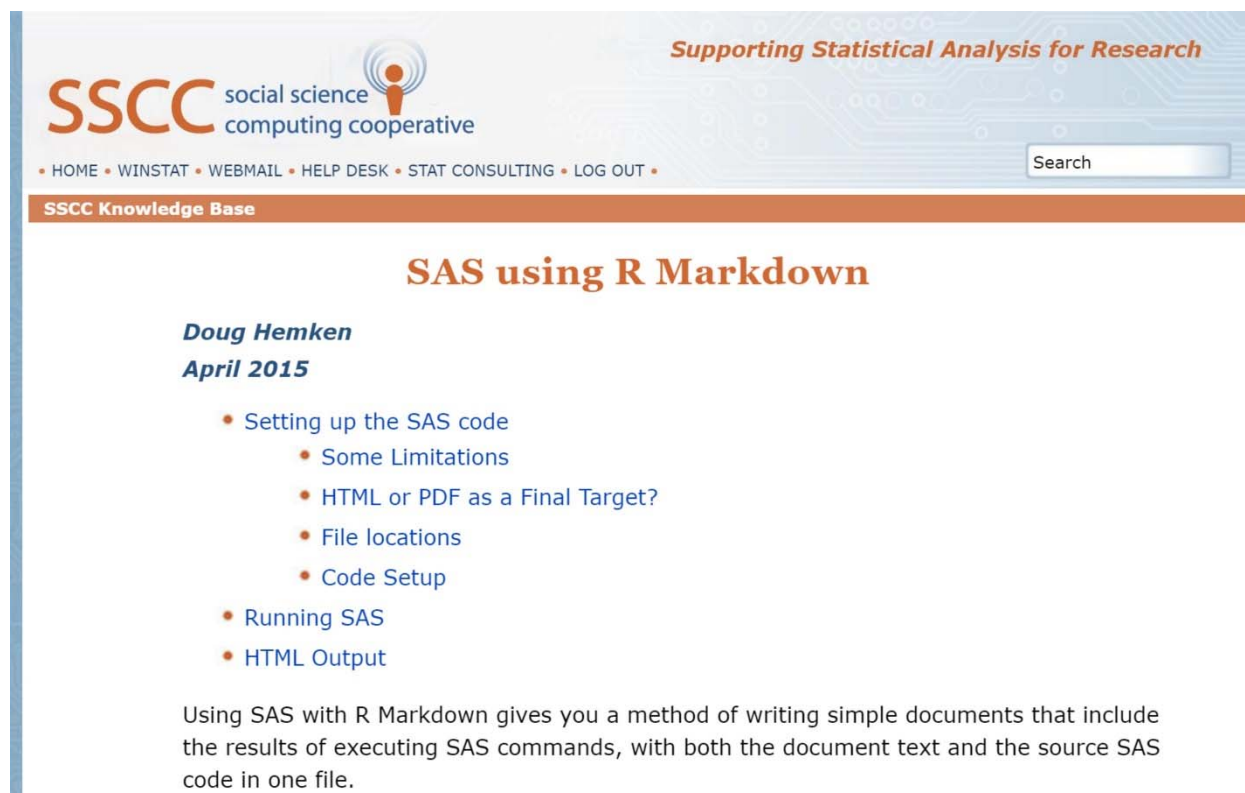
## 文芸的プログラミングパッケージの位置づけ



knitrはSASでも使用できるパッケージである.



## Dougがknitrを使用した文芸的プログラミング事例



SSCC social science computing cooperative  
Supporting Statistical Analysis for Research

• HOME • WINSTAT • WEBMAIL • HELP DESK • STAT CONSULTING • LOG OUT • Search

SSCC Knowledge Base

### SAS using R Markdown

**Doug Hemken**  
**April 2015**

- Setting up the SAS code
  - Some Limitations
  - HTML or PDF as a Final Target?
  - File locations
  - Code Setup
- Running SAS
- HTML Output

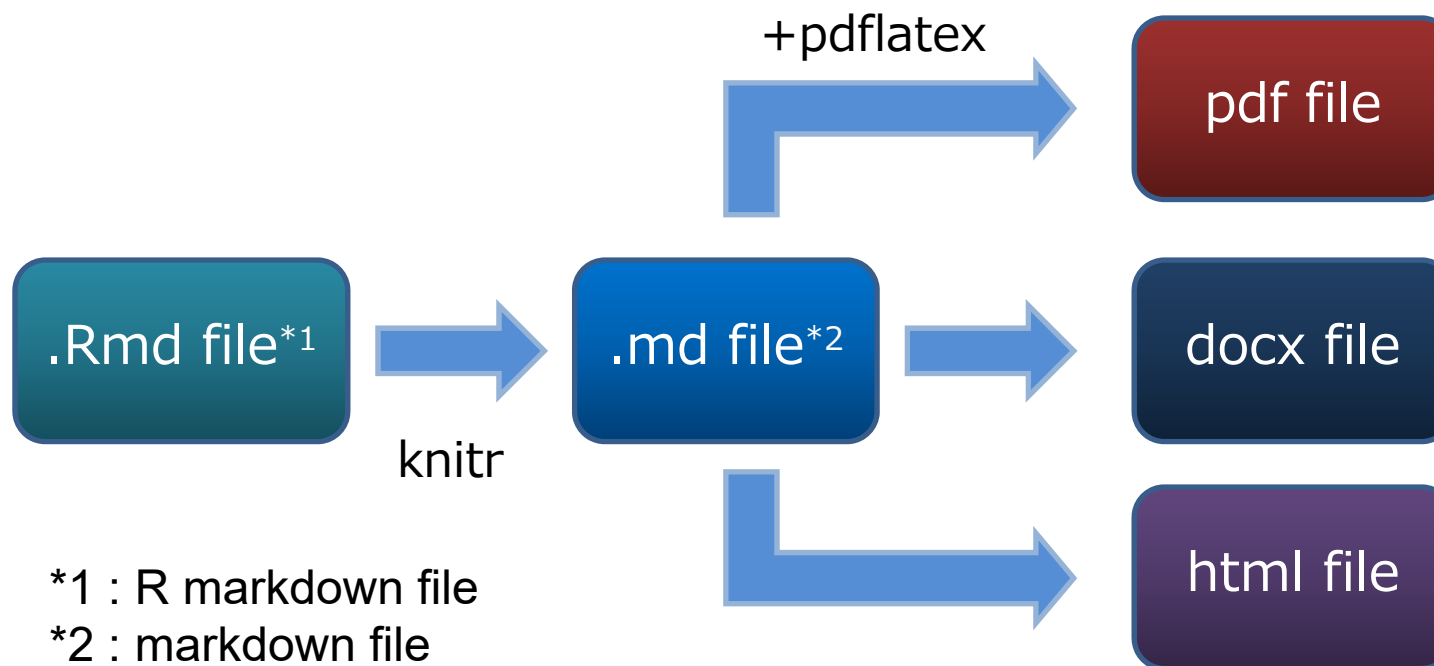
Using SAS with R Markdown gives you a method of writing simple documents that include the results of executing SAS commands, with both the document text and the source SAS code in one file.

この方法を利用してword形式への出力を検討.



## 設定条件と方法 - 処理フローの説明 -

Knitrを使った文書変換は様々な形式に出力可能!!

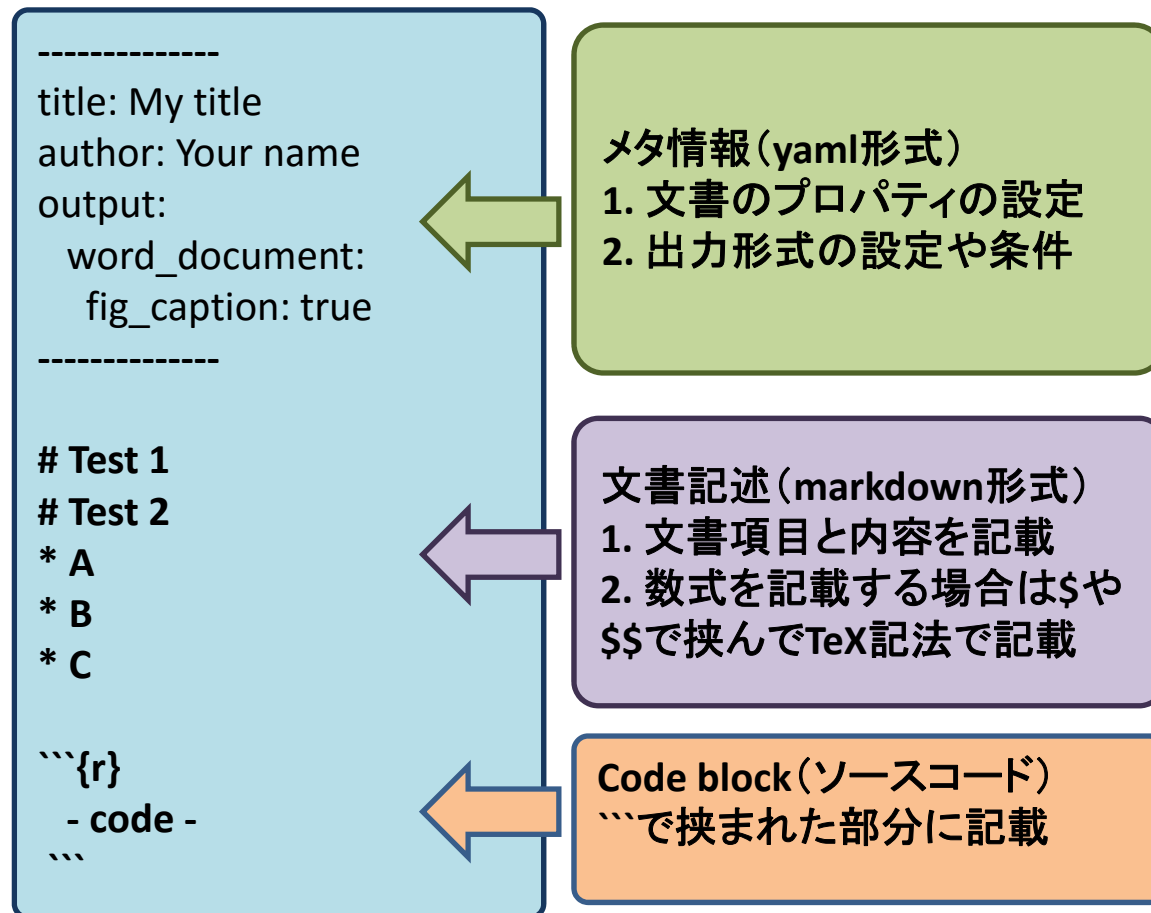


実際は.md fileからの変換はpandocが行っている.

## 設定条件と方法

- R MARKDOWNの構造とCHUNKの指定 -

## Markdownを理解することで容易に記載可能!!



## SASコードを実行するには事前の設定が必要.

# R言語の時のChunkの使い方。

```
``{r}  
- r code -  
``
```

# SASの時のChunkの使い方。

```
``{r, message=FALSE}  
saspath <- "C:/.../sas.exe"  
sasopts <- "-nosplash -log 'C:/.../' -print 'C:/.....' -ls 75"  
opts_chunk$set(engine="sas", engine.path=saspath, engine.opts=sasopts,  
comment="", results="hide")  
``  
``{r}  
- sas code -  
``
```



## 評価(結果)

Dougの方法でword形式へ出力した結果.

	html形式	word形式
数式(Tex記法)	○	○*1
リストの表示	○	△
図の表示	○	×

\*1: officeのバージョンにより文字化けが発生 (office 2007).

## 改善策

1. リストの取り込み: プログラムを取り込む方法を応用
2. 図の取り込み : markdownで図を取り込む



## 改善後のプログラムの内容について.

run.r [rmd2word.rmdを読み込んで実行するプログラム]

```
1 library(knitr)
2 library(rmarkdown)
3 render("rmd2word.Rmd", "word_document") # MS Wordレポート
```

rmd2word.Rmd [sasコードを実行して]

```
1 ---
2 title: 生存時間解析 (SAS)
3 author: "@Yoshihiro Fukiya"
4 output:
5   word_document:
6     fig_width: 2
7     fig_height: 1.5
8     fig_caption: true
9     pandoc_args: [
10      "-s", "--mathml"
11    ]
12 ---
13
14 ### 【統計手法】
15 #### Kaplan-Meier法
16
17 $$ \hat{S}(t)=(1-d_1/n_1)\times(1-d_2/n_2)\times\cdots=\prod_{t_i<t}^{(1-d_i/n_i)} $$
18
19 * $t_i$: イベント (死亡) が起こった時点.
20 * $d_i$: 時点$t_i$での死亡数.
21 * $n_i$: 時点$t_i$直前のリスク集合の大きさ.
```

\*\*\*\*\* 省略 \*\*\*\*\*

# SAS ユーザー総会 2016

\*\*\*\*\* 省略 \*\*\*\*\*

```
42 ▾ ### 【SASプログラム】
43 ▾ ```{r, message=FALSE, echo=FALSE}
44 saspath <- "C:/Program Files/SAS/SASFoundation/9.2/sas.exe"
45 sasopts <- "-nosplash -log 'C:/Users/y-fukiya/Documents/Markdown/rmdlog.log' -print
'C:/Users/y-fukiya/Documents/Markdown/rmdput.lst' -ls 75"
46 opts_chunk$set(engine="sas", engine.path=saspath, engine.opts=sasopts, comment="", results="hide")
47 ▾ ```
48
49 ▾ ```{r data}
```

\*\*\*\*\* 省略 \*\*\*\*\*

```
82 * ----- 出力設定;
83 ods graphics / reset=index imagename="Kaplan-Meier" imagefmt=png ;
84 ods html close;
85 ods listing gpath="C:/Users/y-fukiya/Documents/Markdown";
86
87 ods select survivalplot;
88 proc lifetest data=KM plots=s;
89   time TIME*CENSOR(0);
90   strata TRT;
91 run;
92
93 ods listing close;
94 ods graphics / reset=all;
95 ods listing;
96 ▾ ```
97
98 ```{r eval=FALSE, code=readLines("C:/Users/y-fukiya/Documents/Markdown/rmdput.lst")}
99 ```
100
101 ![proc lifetestの結果](C:/Users/Y-fukiya/Documents/Markdown/Kaplan-Meier.png "Kaplan Meier")
```

改善策1:リストの取り込み

改善策2:図の取り込み

改善後の出力結果(1) 参照:“実用SAS生物統計ハンドブック” p172-177

## 生存時間解析 (SAS)

@Yoshihiro Fukiya

### 【統計手法】

Kaplan-Meier 法

$$\hat{S}(t) = (1 - d_1/n_1) \times (1 - d_2/n_2) \times \dots = \prod_{t_i < t} (1 - d_i/n_i)$$

- $t_i$  : イベント (死亡) が起こった時点.
- $d_i$  : 時点 $t_i$ での死亡数.
- $n_i$  : 時点 $t_i$ 直前のリスク集合の大きさ.

生存関数の標準誤差は *Greenwood* の公式により計算される.

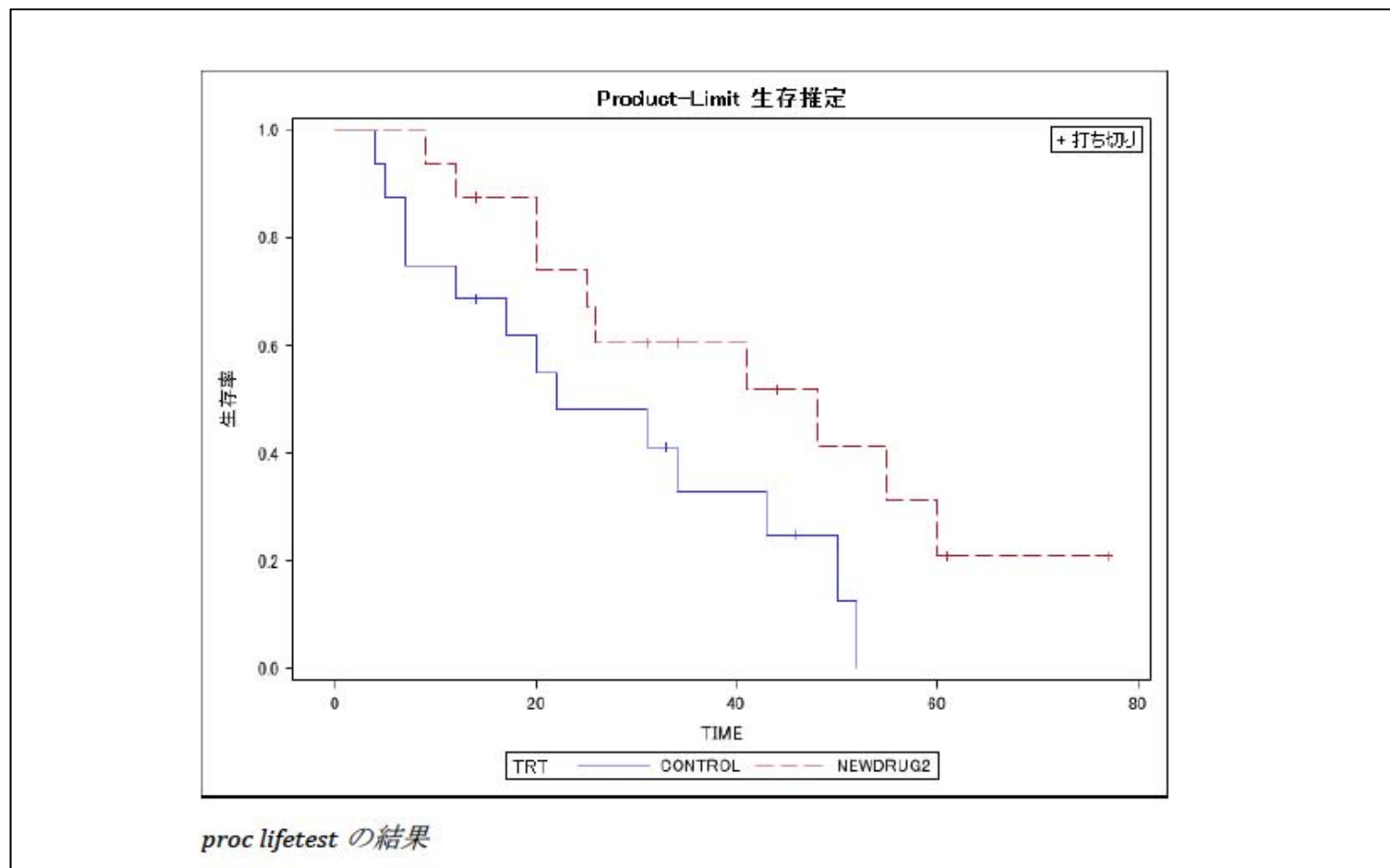
$$SE((\hat{S}(t))) \approx \hat{S}(t) \cdot \sqrt{\sum_{t_i < t} \frac{d_i}{n_i \cdot (n_i - d_i)}}$$

時点ごとのそれぞれの推定値 $\hat{S}(t)$ に対する正規近似した $1 - \alpha/2$ 信頼区間.

$$\text{信頼上限 : } UCL = \hat{S}(t) + z_{\alpha/2} \cdot SE((\hat{S}(t)))$$

$$\text{信頼下限 : } LCL = \hat{S}(t) - z_{\alpha/2} \cdot SE((\hat{S}(t)))$$

## 改善後の出力結果(2)





## 結果と考察

- 文書変換の評価についての考察
  - Dougの方法を改変することでword形式に出力可能.
  - SAS macroの仕様書作成などに応用ができる.
- 今後の検討課題
  - SAS university editionでは使用できない.
    - 文芸的プログラミングは情報共有に適しているため SAS university editionでの実装が望まれる.
  - SAS syntax highlightingへの対応も検討.

## 参考文献

- 1) Arnold, T and Kuhfeld, WF (2012) “Using SAS and LaTeX to Create Documents with Reproducible.”
- 2) Doug, H (2015) “SAS using R markdown”  
<http://www.ssc.wisc.edu/~hemken/SASworkshops/SASWindows/Markdown/SASmarkdown.html>
- 3) Knuth, Donald E (1984) “Literate programming.” The Computer Journal (1984) 27 (2): 97-111.
- 4) Lenth RV, Højsgaard S (2007) “SASweave: Literate Programming Using SAS.” Journal of Statistical Software, 19(8). 15 Page 16.
- 5) Leisch, F (2002) “Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis.” Compstat pp 575-580
- 6) Xie, Y (2012) “An Introduction to knitr.”
- 7) Xie, Y (2014) “knitr: a comprehensive tool for reproducible research in R.” Implement Reprod Res.
- 8) 浜田 知久馬 (2005) “実用SAS生物統計ハンドブック” サイエンス社, p172-177