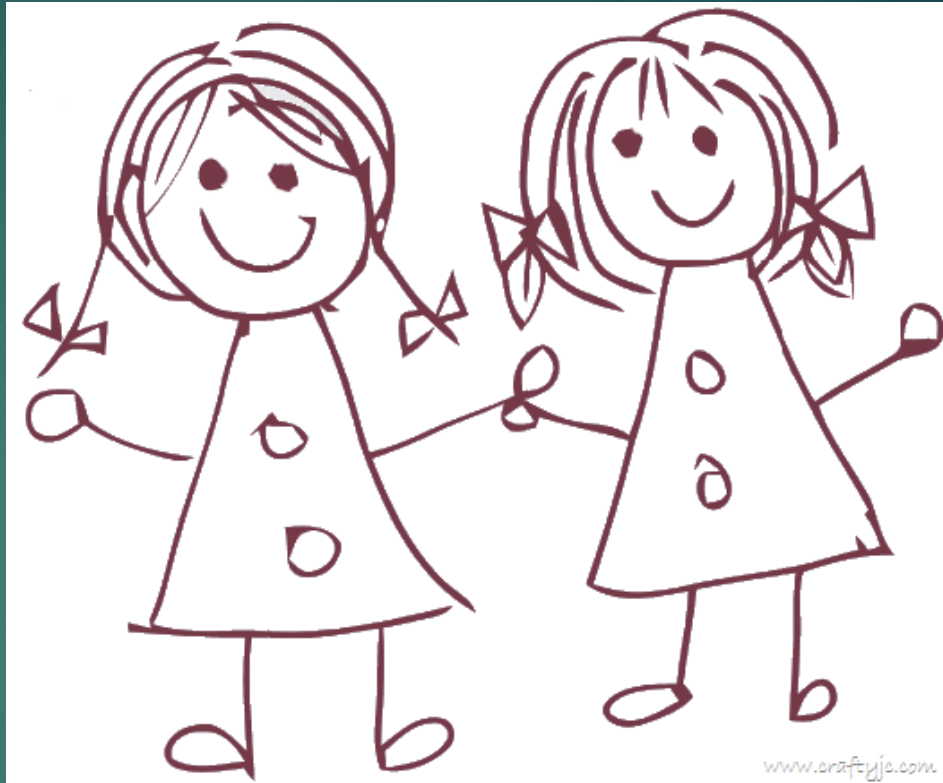
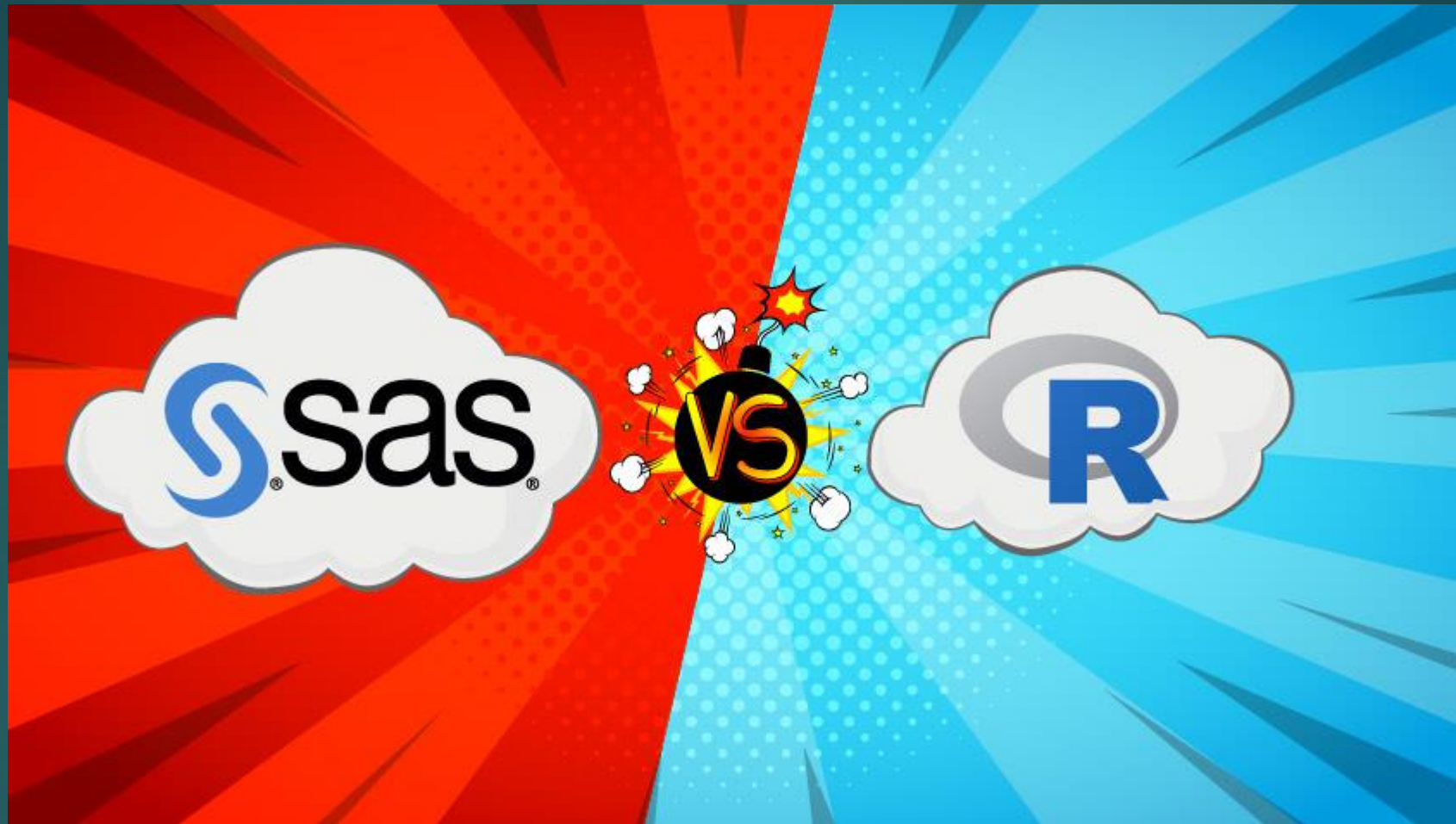


SAS & R



CAN THEY REALLY PLAY NICE TOGETHER?



Price Tag

- ▶ SAS has a large price tag particularly for smaller companies
- ▶ R is free to download for anyone
- ▶ SAS now has free University edition
- ▶ Not available to small companies, non-profit research organizations

Learning

- ▶ R has an abrupt learning curve
 - ▶ Programming language and straight forward task can require longer code
 - ▶ Flexibility and power for a programmer

- ▶ SAS viewed as less difficult to learn
 - ▶ Utilizes SQL programming

Technical Support

- ▶ R has no customer support but does have large online community
 - ▶ Nearly 15,000 programs in Comprehensive R Archive Network (CRAN)

- ▶ SAS has great customer support, complete instruction manual, and large online communities

<https://communities.sas.com/>,

<http://support.sas.com/software/products/stat/index.html>

SAS User groups and User Group websites, <http://www.lexjansen.com/>

Innovation

- ▶ SAS has slower roll out of new techniques
 - ▶ New releases are done in managing environment therefore less opportunity for mistakes

- ▶ R is open source so latest cutting edge techniques are released faster
 - ▶ Potential for more errors in the programming code

Handling of Data

- ▶ Server editions of SAS can handle terabytes of data
 - ▶ the size of data sets analyzed in SAS are generally bottlenecked by the size of the hard disk
- ▶ R allocated and handles memory by loading entire dataset in RAM
 - ▶ R are bottlenecked by the size of the RAM
- ▶ Both are physical hardware limitations of the machine

Debate Summary




- ▶ Both R and SAS are here to stay
- ▶ Why not work together to improve quality of analysis, results, etc

BUT: I already know how to use SAS and I am busy

WHY DO I CARE??

- ▶ Competing Risk analysis with time dependent variables
 - ▶ Cancer relapse and mortality with transplant
 - R: Package 'cmprsk' – June 2014
 - SAS: SAS/STAT 13.1 ver 9.4 Fine and Gray's sub-distribution
- ▶ Joint models of Longitudinal and Survival data
 - ▶ Repeated hospital Admissions and mortality of patients
 - R: JM (2010), joineR, JMbayes & frailtypack,
 - SAS: JMFIT Macro (2016)

- 
- ▶ The SAS/IML Studio interface allows you to integrate R functionality with SAS/IML or SAS programs. You can also exchange data between SAS and R as data sets or matrices.
 - ▶ “This is just the first step,” said Radhika Kulkarni, Vice President of Advanced Analytics. “We are busy working on an R interface that can be surfaced in the SAS server or via other SAS clients. In the future, users will be able to interface with R through the IML procedure.”



▶ **R Interface Now Available in SAS/IML® Studio**

- ▶ If you already have SAS 9.2 or later (Base SAS®, SAS/STAT®, and SAS/IML®) installed, you can [download](#) the latest release of SAS/IML Studio. Installations of SAS 9.2 and later will include SAS/IML Studio 3.3 with SAS/IML from this point forward. Documentation can be found in the chapter "[Calling Functions in the R Language](#)" in *SAS/IML Studio 14.2 for SAS/STAT Users*.

Compatibility

Compatibility with R Releases				
SAS Version	PROC IML	SAS/IML Studio	Release Date	R Versions
9.2	N/A	3.2	Jul 2009	2.6.1 - 2.11.1
9.22	9.22	3.3	Nov 2010	2.9.1 - 2.11.1
9.3	9.3	3.4	Jul 2011	2.9.1 - 2.15.3
9.3m2	12.1	12.1	Aug 2012	2.9.1 - 2.15.3
9.4	12.3	12.3	Jul 2013	2.13.0 - 3.0.1
9.4m1	13.1	13.1	Dec 2013	2.13.0 - 3.2.5
9.4m2	13.2	13.2	Aug 2014	2.13.0 - 3.2.5
9.4m3	14.1	14.1	Aug 2015	2.13.0 - 3.2.5
9.4m4	14.2	14.2	Nov 2016	2.13.0 - 3.4.2
9.4m5	14.3	14.3	Sep 2017	2.13.0 - 3.4.2

- ▶ Locate your file "sasv9.cfg"
- ▶ 9.4 - C:\Program Files\SASHome\SASFoundation\9.4\sasv9.cfg

Step 1:

```
-RLANG
```

```
-config "C:\Program Files\SASHome\SASFoundation\9.4\nls\en\sasv9.cfg"
```

```
-SET R_HOME "C:\Program Files\R\R-3.1.3"
```

Step 2:

```
proc options option=rlang;
```


```
run;
```

```
SAS - [Log - (Untitled)]
File Edit View Tools Solutions Window Help
NOTE: There were 647 observations read from the data set CLARKE.HOSPITAL_VE
WHERE study_id<9999;
NOTE: The data set WORK.CLARK has 647 observations and 221 variables.
NOTE: DATA statement used (Total process time):
      real time           0.90 seconds
      cpu time            0.20 seconds

6574
6575   proc options option=rlang;
6576   run;

      SAS (r) Proprietary Software Release 9.4   TS1M3

      RLANG                Enables SAS to execute R language statements.
NOTE: PROCEDURE OPTIONS used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds
```



<https://www.datacamp.com/courses/r-for-sas-spss-and-stata-users-r-tutorial>

[Ken Kleinman Nicholas J. Horton](#) . SAS and R: Data Management, Statistical Analysis, and Graphics, Second Edition, Edition 2. July 17, 2014. CRC Press

SAS/IML(R) 13.1 User's Guide: Calling Functions in the R Language


```
proc options option=rlang;  
run;
```

```
data test;  
  set clark;  
  keep study_id startage gender death time_risk;  
run;
```

```
proc iml;  
  use work.test;  
  read all var {study_id startage gender death time_risk};  
  close work.test;
```

```
run ExportDataSetToR("work.test", "Class");  
submit / R;  
  Model <- lm(time_risk ~ GENDER, data=Class, na.action="na.exclude")  
  summary(Model);  
  boxplot(time_risk ~ death, Class, xlab="death", ylab="Time to Last followup", las=1)  
  ttResult <- t.test(time_risk ~ death, Class, var.equal=TRUE);  
  ttResult;  
endsubmit;  
quit;
```

```
Call:
lm(formula = time_risk ~ GENDER, data = Class, na.action = "na.exclude")
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-961.94 -491.53  -88.94  405.27 1744.06
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  963.941     41.848   23.034  <2e-16 ***
GENDERM       3.587     52.634    0.068   0.946
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 645.6 on 645 degrees of freedom
```

```
Multiple R-squared: 7.2e-06, Adjusted R-squared: -0.001543
```

```
F-statistic: 0.004644 on 1 and 645 DF, p-value: 0.9457
```

Two Sample t-test

```
data: time_risk by death
```

```
t = 5.4487, df = 645, p-value = 7.227e-08
```

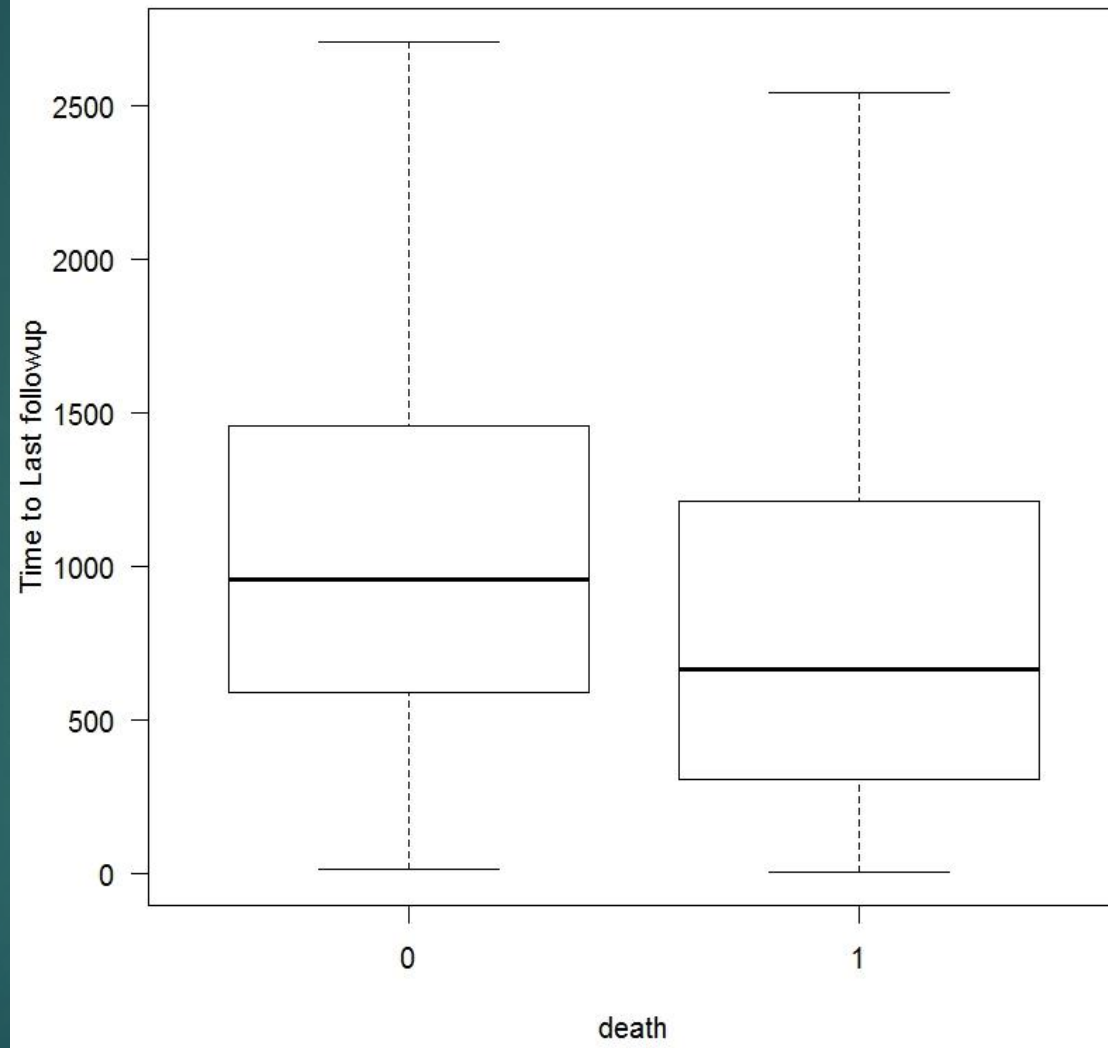
```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
 177.2345 376.9596
```

```
sample estimates:
```

```
mean in group 0 mean in group 1
```



Calling a R library

submit / R;

install.packages('cmprsk')

library(cmprsk)



```
proc iml;
  use work.final;
  read all var {study_id tstart tstop status frailtyscale death time_event gaptime} ;
  close work.final;
  run ExportDataSetToR("work.final", "frail");
  submit / R;
  install.packages('frailtypack')
  install.packages('survival')
  library(frailtypack)
  library(survival)

  modJoint.general <- frailtyPenal(Surv(gapTime,status) ~ cluster(Study_ID) + frailtyscale + terminal(death),
  formula.terminalEvent = ~ frailtyscale,
  data = frail, jointGeneral = TRUE, n.knots = 8,
  kappa = c(2.11e+08, 9.53e+11))
  modJoint.general
  endsubmit;
quit;
```

```
SAS - [Results Viewer - sashtml]
File Edit View Go Tools Solutions Window Help
9.53e+11))

General Joint gamma frailty model for recurrent and a terminal event processes
using a Penalized Likelihood on the hazard function

Recurrences:
-----
              coef exp(coef) SE coef (H) SE coef (HIH)      z      p
frailtyscale 0.126375  1.13471  0.0313849    0.0313574 4.02661 5.6588e-05

Terminal event:
-----
              coef exp(coef) SE coef (H) SE coef (HIH)      z      p
frailtyscale 0.322866  1.38108  0.0535692    0.0532372 6.02709 1.6694e-09

Frailty parameters:
theta (variance of u, association between recurrences and terminal event): 0.543441 (SE (H)
eta (variance of v, intra-subject correlation): 0.00500415 (SE (H): 0.000297851 ) p = 0

penalized marginal log-likelihood = -10055.07
Convergence criteria:
parameters = 6.42e-06 likelihood = 2.82e-06 gradient = 6.47e-10

LCV = the approximate likelihood cross-validation criterion
      in the semi parametric case      = 5.83945

n observations= 1726  n subjects= 564      ( 255  observation deleted due to missing)
n recurrent events= 1229
n terminal events= 214
n censored events= 497
number of iterations: 12
```

THANK YOU!


Kara Matheson

Research Biostatistician

Research Methods Unit

Nova Scotia Health Authority/ Dalhousie University

kara.thompson@dal.ca kara.Matheson@nshealth.ca

- 
- ▶ However, hard drive capacities have historically grown faster than support for RAM, giving R the bad rap of being limited. Two key developments make this topic moot. With the appearance of 64-bit operating systems that support far more RAM and with R connectivity to databases, R can be made to support any size data set that SAS can support, including those that contain billions of observations.