# RSCABS:

# An R Package for Performing the Rao-Scott Adjusted Cochran-Armitage Trend Test by Slices

Joe Swintek Badger Technical Services

April 30, 2020

# Contents

Introduction	3
Running RSCABS by GUI	3
Starting Histopath	3
Loading Data	5
Specifying Identification Fields	7
Running RSCABS	8
Interpreting the Results	10
Running Other Analyses	12
Running RSCABS by Command Line	14
Plotting	15
The Plotting Window	15
Plotting by Command Line	17
Acknowledgments	18
References	18

## Introduction

RSCABS[3] (<u>Rao-S</u>cott adjusted <u>C</u>ochran-<u>A</u>rmitage trend test <u>By</u> <u>S</u>lices) is a modification to the Rao-Scott[5] adjusted Cochran-Armitage trend test[1, 2] that allows for testing at each individual severity score often seen in histopathological data. The test was originally developed and implemented in SAS<sup>TM</sup> by John Green<sup>1</sup> as part of the <u>M</u>edaka <u>E</u>xtended <u>O</u>ne <u>G</u>eneration <u>R</u>eproduction <u>T</u>est[4] (MEOGRT) with the purpose of testing the effects of endocrine disruptors on histopathological endpoints. The RSCABS package allows for easy use of the RSCABS analysis with the capability of using both command line and graphical user interface (GUI) driven operations.

The RSCABS analysis is specificity designed to analyze histopathological results from standard toxicology experiments, for example the MEOGRT. These experiments typically have a number of organisms (e.g. medaka) contained in various amounts in different holding apparatuses (e.g. fish tanks). Multiples of each holding apparatuses are exposed to a either an experimental control (plain water) or one of several different concentrations of a chemical. At the end of the experiment several histopathological endpoints are evaluated on every organism. A severity score is assigned to every histopathological endpoint, which is typically an integer ranging from 0 (no effect) to 5 (an incredibly severe effect). A greater severity score indicates a more severe effect but the differences in severity scores are not consistent. For example the difference in severity of a score of 2 to a score of 1 is not the same as the change in severity moving from a score of 5 to of 4. Thus, even though severity scores have an order to them, they must be treated as categorical variables.

To develop an analysis of severity score data, several steps must be completed. The first step is to develop a basic test which tests a set of organisms for an increase in the presences (score > 0) or absence (score = 0) of an effect with an increase in the dose concentration of the treatments. The Cochran-Armitage (CA) trend [1,2] test was developed to test for this increase. However, it is common for group of organisms to be contained within the same holding apparatus. This could lead to organisms within the same apparatus having results that more closely resemble each other, then organisms in other apparatuses. The Rao-Scott (RS) adjustment controls for this by calculating an adjustment to the CA test statistic from correlation of organisms within each apparatuses. The by slices (BS) portion of the test allows for testing at each severity score instead of just presences or absence. By slices works by splitting the the severity scores associated with an endpoint into two groups based on the severity score being tested. One group contains all severity score less then the target severity scores and the other group contains all severity scores equal to or greater than the target severity score. The RSCA test statistic is calculated based on these two groups instead of just a presences (score  $\geq 1$ ) and absence (score < 0). For example testing at a severity score of 2 would involve splitting the data into a group of severity scores < 2 and a group of all severity scores  $\geq 2$ . RSCABS is a step down anylsis so, if the test statistic is calculated to be significant (typically, p-value  $\leq 0.05$ ) then highest treatment level is removed from the analysis and the RSCA test statistic is recalculated. The process is repeated until the test statistic is not significant or there are no treatment levels left. This step-down procedure is repeated for each unique score assigned to an endpoint. Further details an examples of RSCABS can be found in [3].

# **RSCABS GUI**

#### **Starting Histopath**

For ease of operation the RSCABS package has two ways of performing the RSCABS analysis, a command line function runRSCABS or a GUI front-end called Histopath. To call the GUI simply type the following into the console:

<sup>&</sup>lt;sup>1</sup>DuPont Applied Statistics Group

<pre>&gt; install.packages(RSCABS)</pre>	#Install RSCABS from CRAN
> library(RSCABS)	#Load the RSCABS library
> Histopath()	#Calls the GUI for RSCABS

RSCABS is depended on the RGtk2 package which uses gtk+ and can be found at http: //www.gtk.org/. If gtk+ is not installed, using "Histopath()" will cause R to produce an error message and then prompt for the installation of gtk+.

R Input

100% downloaded: R	gui.exe - System Error	×
Compute	ram can't start because libatk-1.0-0.dll is missing from r. Try reinstalling the program to fix this problem.	your
		ОК
	ed GTK+ ? (Restart R after i I <mark>nstall GTK+</mark> Do not install GTK+ OK Cancel	

The instalation of gtk+ can be accomplished within R by selecting the "Install GTK+" option. After gtk+ is finished installing, R will need to be restarted before continuing. After restarting R using the "Histopath()" command should produce this window:

R Histopath		- • •
Main		
Load Data	variables	•

# Loading Data

After Histopath has been successfully called, data can be loaded into the program by clicking on the [Load Data] button. This will open a dialogue box where the data file can be selected. Due note, this browser window cannot navigate through short cuts and will give an error if tried.

I	🥂 Choose a CSV file			×
		Data		
	<u>P</u> laces	Name Size	Modified	<u>~</u>
	🔍 Search	🖬 Histology Example Data.csv 100.0 KB	12/3/2015	
	Recently Used			
	🛅 Net MyDocu			
	🛅 Desktop			
	In the second se			
	🥔 DVD Drive (D:)			
	IVD RW Drive			
	🧼 System Reserv			
	🧼 System (G:)			
	🥪 Removable Di			Ξ
	🛫 Data (\\AA\O			
	🛫 Data (\\AA.AD			
	<u>A</u> dd <u>R</u> emove			Ŧ
		Cancel	<u>O</u> pen	
			_,	

Below is a screen shot of the example histology data in comma separated value (csv) provided with in the RSCABS package.

FILE	E HOME		PAGE LAYOUT	F FORMULAS	DATA	REVIEW VIEW	ESRI MAPS	ogy Example Data.csv - Exce					- 🗗 🕽
aste	Ba Conv -	Calibri	- 1		_	≫ - 🔐 Wrap Text	General		Normal Bad		The second seco		
aste •	e ؇ Format Pai	inter B I	<u>u</u> - 🖽 -	<u>◇</u> • <u>A</u> • ≡	= =	🖅 🖅 🔛 Merge & C	Center - \$ - % + %	Conditional Format Formatting * Table	as Good Neutral	Thisert Delete Fo	v de Clear v Filter v Select		
	Clipboard	6	Font	G.		Alignment	rs Number	G	Styles	Cells	Editing		
J1	Ψ	XV	fx Kid	Glomer Epithe	Hypertrop	phy							
al	Α	В	С		E	F	G	н	ī	i i	К	L	м
				Genotypic Se			Kid Tub Epithe Eosino		Gon Incr Spermatagonia	Gon Testicular Degen	Gon Interstitial Cell HH	Gon Tes	
	FO		1 A	Female	16 wk			1 NA	Con_inci_operindicagonic			NA	a con_re
	FO		1 A	Male	16 wk			3 NA	0			NA	
	FO	1	1 B	Female	16 wk			1 NA	C	C	) C	NA	
	FO	1	1 B	Male	16 wk			2 NA	C	C	) (	NA	
	FO	1	1 C	Male	16_wk	1		3 NA	C	C	C	NA	
'	FO	1	1 D	Female	16_wk	5		2 NA	C	C	0	NA	
	FO	1	1 D	Male	16_wk	1		2 NA	C	C	0	NA	
	FO	1	1 E	Female	16_wk	5		1 NA	C	C	) C	NA	
	FO		1 E	Male	16_wk			2 NA	C			NA	
_	FO		1 F	Female	16_wk			1 NA	C			NA	
_	FO		1 F	Male	16_wk			3 NA	C			NA	
-	FO		2 A	Female	16_wk			2 NA	C	-		NA	
	FO		2 A	Male	16_wk			3 NA	C			NA	
	FO		2 B	Female	16_wk			1 NA	C			NA	
	FO		2 B	Male	16_wk			3 NA	C			NA	
	FO		2 C	Female	16_wk			2 NA	C			NA	
	FO		2 C	Male	16_wk			3 NA	C			NA	
	FO		2 D	Female	16_wk			1 NA	C			NA	
	FO		2 D 2 E	Male	16_wk			3 NA	0			NA	
	FO		2 E 2 E	Female Male	16_wk			1 NA 3 NA	0			NA	
	F0 F0		2 E 2 F	Female	16_wk			3 NA 1 NA	C			NA NA	
	FO		2 F 2 F	Male	16_wk 16_wk			1 NA 3 NA				) NA ) NA	
	FO		2 F 3 A	Female	16_wk			1 NA	0			NA NA	
	FO		3 A	Male	16 wk			3 NA	0	-		NA NA	
_		istology Exam		÷	10_WK	1			: •				F

Data sets must be in a csv format. Each column indicates contains information used to identify a specimen or the severity scores of an endpoint while each row is a different specimen, which in this example is a fish. The exception is the first row which is the header row and contains the names of the fields. As with any csv file imported into R, missing data are indicated by either NA or a blank cell. Character entries (names etc...) may contain spaces, underscores ("\_"), or dots (".") to separate words, however R will convert all spaces to dots upon importing the file.

Histopath separates the data into three types of fields; identification, pathology endpoints, and ignored fields. The identification fields include a field for each of: gender, generation, age, treatment, and replicate. Of the identification fields treatment is mandatory while gender, generation, age, and replicate are not. However, if the replicate field is not included in the data set, Histopath will assume each specimen is independent and not apply the Rao-Scott adjustment. Any identification fields must have a value for every entry otherwise the row associated with that entry will be removed when RSCABS is ran. The treatment field must only contain integers, with 0 indicating controls and each larger number indicating a larger dose. The other identification fields do not have this restriction and can contain any combination of number and letters.

Every field that is not an identification field is potentially a pathology endpoint. These fields may contain any entry, however, any entry that is not a **0** or a **positive number** is treated as missing data and is not included in the analysis. Due to how RSCABS groups severity scores non-integer numbers are treated as the next smallest integer, e.g. both 1.1 and 1.9 are treated as 1. Columns that are not identification fields, do not contain any number larger then **0**, or contain numbers larger then **20** will be ignored by Histopath.

#### **Specifying Identification Fields**

After the data set is loaded into Histopath, a [Specify Data] button will appear. Clicking the button will create a Data Specification tab.

1	Constituted	T	Deellerte 4	Constantia Con d	A	Can Dharatana (	Con See Devend 4
Load Data	F0	1	A A	Female	16_wk		Gon_Sex_Reversal   NA
pecify Data	FO	1	Ā	Male	16_wk		NA
	FO	1	В	Female	16_wk		NA
	FO	1	в	Male	16_wk		NA
	FO	1	c	Male	16_wk		NA
	FO	1	D	Female	16_wk		NA
	FO	1	D	Male	16_wk		NA
	FO	1	E	Female	16_wk		NA
	FO	1	E	Male	16_wk	1	NA
	FO	1	F	Female	16_wk	5	NA
	FO	1	F	Male	16_wk	1	NA
	FO	2	A	Female	16_wk		NA
	FO	2	Α	Male	16_wk	1	NA

<u>R</u> Histopath				-		×
Main Data specification						
Select Gender Variable	Gender Variable Not Selected	Select Gender Value	Gender Varial	ole Not S	Selected	
Select Generation Variable	Generation Variable Not Selected	Select Generation Value	Generation Va	ariable N	lot Select	ed
Select Age Variable	Age Variable Not Selected	Select Age Value	Age Variable	Not Sele	cted	
Select Treatment Variable	Treatment Variable Not Selected	Select Replicate Variable	Replicate Va	riable No	ot Selecte	:d
Select Endpoint(s) to Invert		Nothing Inverted				
	Confirm Selected Va	alues and Variables				

The form above is where all the identifiers for the data set are specified. The only entry in the form that must be specified is the **treatment** variable. However, if the **replicate** variable is not specified, Histopath will default to using SCABS (<u>S</u>tandard <u>C</u>ochran-<u>A</u>rmitage trend test <u>By</u> <u>Slices</u>) which is RSCABS without the RS correction. A warning; if a replicate structure was used in the experiment but is not specified in the analysis, pseudo-replication will occur which may lead to an increase in the number of falsely positive results. After all entry forms are filled out, click on the [Confirm Selected Values and Variables] button to set the selected variables into Histopath. After the selection is set, you can navigate back to the main tab to perform the RSCABS analysis. Note, that at any time you may navigate back to the Data specification tab to change a selection, just re-click the [Confirm Selected Values and Variables] button after a new selection is made to accept the change. Clicking on [Confirm Selected Values and Variables] button after a new selection is made to recognise the change in the variables.

Some endpoints are coded in such a way that a decreasing severity score indicates an increasing severity. For instance, medaka gonad phenotype is set to a scale of 1 to 5. Where a 1 is a complete male gonad and a 5 is a complete female gonad, with scores in between indicating some degree of intersex. For genotypic males the severity of the change increases with the score for gonad phenotype but for genotypic females the scale means the opposite with 5 indicating the least severe change and 1 indicating the most severe change. To account for this, as of version 0.90.3, RSCABS has an additional button in the specification form; [Select Endpoints to Invert]. Using this button you can select one or more endpoints for which RSCABS will treat the highest severity score as indicating the least severe condition and the lowest severity scores as indicating the most severe condition is selected a new column, called "Inverted", will be added to the results which will use a "Yes" to indicate that the endpoint was inverted.

#### **Running RSCABS**

After the data has been specified and the [Confirm Selected Values and Variables] has been clicked, the Histopath main tab will show the data set updated with the specifications and should be navigated back to.

lain Data specificati	on							
Load Data	Generation 4	Treatment 4	Replicate 4	Genotypic_Sex 4	Age 4	Gon_Phenotype 4	Gon_Sex_Rev	-
C	F2	1	A	Female	16_wk	5	0	
Specify Data	F2	1	В	Female	16_wk	5	0	
Run RSCABS	F2	1	D	Female	16_wk	5	0	
Run Other Analyses	F2	1	E	Female	16_wk	5	0	-
	F2	1	F	Female	16_wk	-2	-2	
	F2	2	Α	Female	16_wk	5	0	
	F2	2	В	Female	16_wk	5	0	L
	F2	2	D	Female	16_wk	5	0	
	F2	2	E	Female	16_wk	5	0	
	F2	2	F	Female	16_wk	5	0	
	F2	3	A	Female	16_wk	-2	-2	
	F2	3	в	Female	16_wk	5	0	
	F2	3	D	Female	16_wk	5	0	

Two buttons; [Run RSCABS] and [Run Other Analyses] will have appeared on the Histopath main tab. Clicking on the [Run RSCABS] will perform RSCABS (or SCABS if a replicate variable is not defined), on the data.

Load Data	Response	Treatment 4	Rscore 4	T-Value 4	P-value 4	Signif	•
Crasif Data	Gon_Incr_Oocyte_Atresia1P	5	1	1.33037	0.0917		
Specify Data	Gon_Incr_Oocyte_Atresia2P	5	2	1.23116	0.10913		
Run RSCABS	Gon_Incr_Oocyte_Atresia3P	5	3	1.23116	0.10913		
Run Other Analyses	Gon_Incr_Oocyte_Atresia4P	5	4	1.23116	0.10913		
Save Result	Gon_Proteinaceous_fluid1P	5	1	-1.41029	0.07923		
Save Result	Gon_Asynch_Dev1P	5	1	2.62202	0.00437	**	
	Gon_Asynch_Dev2P	5	2	2.62202	0.00437	**	
	Kid_Tub_Mineralization1P	5	1	0.3581	0.36013		
	Kid_Tub_Mineralization2P	5	2	-0.07038	0.47194		
	Kid_Tub_Dilation1P	5	1	-0.07038	0.47194		
	Kid_Tub_Dilation2P	5	2	-0.04896	0.48048		
	Kid_Tub_Regeneration1P	5	1	-1.37478	0.0846		

After the analysis on the data is ran, you may save the results by using the [Save Result] button, which will create a dialogue box that will prompt the saving of the results from the

RSCABS analysis as a csv file. Clicking the [Run Other Analyses] button will create a new window with options to perform SCABS, or get further details on a response.

The results themselves appear in the box on the right hand side of the window. The **Response** is the endpoint that is being tested, **Treatment** is the treatment level, **R-Score** is the severity score, **Statistic** is the test statistic corresponding to that row's endpoint, treatment level, and R-Score, with **P-Value** as the corresponding p-value. **Signif** is the significance flag where "." is a p-value > 0.05, "\*" is a 0.01 < p-value  $\le 0.05$ , "\*\*" for 0.001 < p-value  $\le 0.01$ , and "\*\*\*" for p-value  $\le 0.001$ .

#### Interpreting the Results

The example below better illustrates how to interpret the results generated by RSCABS.

on								
Response •	R-Score 4	Treatment 4	Statistic 4	P-Value 4	Signif 4	Inverted		•
Oth_Thymus_Lymphoidhyperplasia3	3	б	-0.34264	0.36593				
Oth_Thymus_Lymphoidhyperplasia2	2	6	-0.34264	0.36593				
Oth_Thymus_Lymphoidhyperplasia1	1	6	-0.34264	0.36593				
Kid_Proteinaceous_Fluid3	3	6	1.8115	0.03503	*			
Kid_Proteinaceous_Fluid2	2	6	2.71278	0.00334	**			
Kid_Proteinaceous_Fluid2	2	5	1.79795	0.03609	*			
Kid_Proteinaceous_Fluid2	2	4	-0.45522	0.32447				
Kid_Proteinaceous_Fluid1	1	6	5.29442	0	***			
Kid_Proteinaceous_Fluid1	1	5	3.20582	0.00067	***			
Kid_Proteinaceous_Fluid1	1	4	1.22433	0.11041				
Gon_Phenotype1	1	6	-7.39499	0	***	Yes		
Gon_Phenotype1	1	5	-7.00647	0	***	Yes		
Gon_Phenotype1	1	4	-4.23705	1e-05	***	Yes		
Gon_Phenotype1	1	3	-4.09718	2e-05	***	Yes		
Gon_Phenotype1	1	2	-0.68577	0.24643		Yes		
Gon_Phenotype2	2	6	-4.81095	0	***	Yes		
Gon_Phenotype2	2	5	-3.73058	1e-04	***	Yes		
Gon_Phenotype2	2	4	-1.41091	0.07914		Yes		
Gon_Phenotype3	3	6	-3.95263	4e-05	***	Yes		
Gon_Phenotype3	3	5	-2.68911	0.00358	**	Yes		
Gon_Phenotype3	3	4	-1.41091	0.07914		Yes		
Gon_Phenotype4	4	6	-3.52407	0.00021	***	Yes		
Gon_Phenotype4	4	5	-2.51649	0.00593	**	Yes		
	Oth_Thymus_Lymphoidhyperplasia3 Oth_Thymus_Lymphoidhyperplasia3 Oth_Thymus_Lymphoidhyperplasia1 Kid_Proteinaceous_Fluid3 Kid_Proteinaceous_Fluid2 Kid_Proteinaceous_Fluid2 Kid_Proteinaceous_Fluid2 Kid_Proteinaceous_Fluid1 Kid_Proteinaceous_Fluid1 Kid_Proteinaceous_Fluid1 Gon_Phenotype1 Gon_Phenotype1 Gon_Phenotype1 Gon_Phenotype2 Gon_Phenotype2 Gon_Phenotype2 Gon_Phenotype3 Gon_Phenotype3 Gon_Phenotype3 Gon_Phenotype4	ResponseR-ScoreOth_Thymus_Lymphoidhyperplasia33Oth_Thymus_Lymphoidhyperplasia22Oth_Thymus_Lymphoidhyperplasia31Kid_Proteinaceous_Fluid33Kid_Proteinaceous_Fluid22Kid_Proteinaceous_Fluid22Kid_Proteinaceous_Fluid22Kid_Proteinaceous_Fluid22Kid_Proteinaceous_Fluid11Kid_Proteinaceous_Fluid11Kid_Proteinaceous_Fluid11Kid_Proteinaceous_Fluid11Gon_Phenotype11Gon_Phenotype11Gon_Phenotype11Gon_Phenotype22Gon_Phenotype22Gon_Phenotype33Gon_Phenotype44	ResponseR-Score Treatment Oth_Thymus_Lymphoidhyperplasia336Oth_Thymus_Lymphoidhyperplasia226Oth_Thymus_Lymphoidhyperplasia116Kid_Proteinaceous_Fluid336Kid_Proteinaceous_Fluid225Kid_Proteinaceous_Fluid224Kid_Proteinaceous_Fluid224Kid_Proteinaceous_Fluid116Kid_Proteinaceous_Fluid115Kid_Proteinaceous_Fluid115Kid_Proteinaceous_Fluid114Gon_Phenotype115Gon_Phenotype113Gon_Phenotype112Gon_Phenotype225Gon_Phenotype336Gon_Phenotype334Gon_Phenotype334Gon_Phenotype334Gon_Phenotype334	Response         R-Score         Treatment         Statistic           Oth_Thymus_Lymphoidhyperplasia2         2         6         -0.34264           Oth_Thymus_Lymphoidhyperplasia2         2         6         -0.34264           Oth_Thymus_Lymphoidhyperplasia1         1         6         -0.34264           Oth_Thymus_Lymphoidhyperplasia1         1         6         -0.34264           Oth_Thymus_Lymphoidhyperplasia1         1         6         -0.34264           Kid_Proteinaceous_Fluid2         2         6         2.71278           Kid_Proteinaceous_Fluid2         2         4         -0.45522           Kid_Proteinaceous_Fluid1         1         6         5.29442           Kid_Proteinaceous_Fluid1         1         5         3.20582           Kid_Proteinaceous_Fluid1         1         4         1.22433           Gon_Phenotype1         1         4         -4.23705           Gon_Phenotype1         1         4         -4.23705           Gon_Phenotype1         1         3         -4.09718           Gon_Phenotype2         2         5         -3.73058           Gon_Phenotype2         2         5         -3.73058           Gon_Phenotype3         3	Response         Ireatment         Statistic         P-Value           Oth_Thymus_Lymphoidhyperplasia3         3         6         -0.34264         0.36593           Oth_Thymus_Lymphoidhyperplasia2         2         6         -0.34264         0.36593           Oth_Thymus_Lymphoidhyperplasia1         1         6         -0.34264         0.36593           Oth_Thymus_Lymphoidhyperplasia1         1         6         -0.34264         0.36593           Kid_Proteinaceous_Fluid2         2         6         2.71278         0.00334           Kid_Proteinaceous_Fluid2         2         5         1.79795         0.3609           Kid_Proteinaceous_Fluid2         2         4         -0.45522         0.32447           Kid_Proteinaceous_Fluid1         1         6         5.29442         0           Kid_Proteinaceous_Fluid1         1         5         3.20582         0.00067           Kid_Proteinaceous_Fluid1         1         4         1.22433         0.11041           Gon_Phenotype1         1         5         -7.00647         0           Gon_Phenotype1         1         3         -4.09718         2e-05           Gon_Phenotype1         1         2         -0.68577         0.24643	Response         R-Score         Treatment         Statistic         P-Value         Signif           Oth_Thymus_Lymphoidhyperplasia         3         6         -0.34264         0.36593         .           Oth_Thymus_Lymphoidhyperplasia         2         6         -0.34264         0.36593         .           Oth_Thymus_Lymphoidhyperplasia         1         6         -0.34264         0.36593         .           Kid_Proteinaceous_Fluid3         3         6         1.8115         0.03503         *           Kid_Proteinaceous_Fluid2         2         6         2.71278         0.00334         ***           Kid_Proteinaceous_Fluid2         2         4         -0.45522         0.32477         .           Kid_Proteinaceous_Fluid1         1         6         5.29442         0         ****           Kid_Proteinaceous_Fluid1         1         4         1.22433         0.11041         .           Gon_Phenotype1         1         4         4.23705         1e-05         ****           Gon_Phenotype1         1         4         -4.23705         1e-05         ****           Gon_Phenotype1         1         2         -0.68577         0.24643         .           Gon_Phen	Response         Inverted         R-Score         Treatment         Statistic         P-Value         Signif         Inverted           Oth_Thymus_Lymphoidhyperplasia         3         6         -0.34264         0.36593         .         .           Oth_Thymus_Lymphoidhyperplasia         1         6         -0.34264         0.36593         .         .           Kid_Proteinaceous_Fluid3         3         6         1.8115         0.03303         *         .           Kid_Proteinaceous_Fluid2         2         6         2.71278         0.003304         **         .           Kid_Proteinaceous_Fluid2         2         5         1.79795         0.03609         *         .           Kid_Proteinaceous_Fluid2         2         4         -0.45522         0.32447         .         .           Kid_Proteinaceous_Fluid1         1         6         5.29442         0         ****         .           Kid_Proteinaceous_Fluid1         1         4         1.22433         0.11041         .         .           Gon_Phenotype1         1         4         -4.23705         1e-05         ****         Yes           Gon_Phenotype1         1         3         -4.09718         2e-05	Response         Inverted         Statistic         P-Value         Signif         Inverted           Oth_Thymus_Lymphoidhyperplasia2         2         6         -0.34264         0.36593         .           Oth_Thymus_Lymphoidhyperplasia2         2         6         -0.34264         0.36593         .           Oth_Thymus_Lymphoidhyperplasia1         1         6         -0.34264         0.36593         .           Kid_Proteinaceous_Fluid2         2         6         2.71278         0.00334         ***           Kid_Proteinaceous_Fluid2         2         6         2.71278         0.003609         *           Kid_Proteinaceous_Fluid2         2         4         -0.45522         0.32447         .           Kid_Proteinaceous_Fluid1         1         6         5.29442         0         ****           Kid_Proteinaceous_Fluid1         1         5         3.20582         0.00067         ****           Kid_Proteinaceous_Fluid1         1         4         1.22433         0.11041         .           Gon_Phenotype1         1         4         -2.243705         1e-05         ****         Yes           Gon_Phenotype1         1         3         -4.09718         2e-05         ****

By default RSCABS will sort the results by decreasing **R-Scores** within reverse alphabetical order of each endpoints name. This allows for quick scanning for lowest observed effect concentrations (LOEC) for each severity on every endpoint. Going down the list above and using an alpha level of 0.05 we see that this test; did not show a response in Oth\_Thymus\_Lymphoidhyperplasia, showed a LOEC of treatment 6 for Kid\_Proteinaceous\_Fluid at severity grade 3, a LOEC of treatment 5 for Kid\_Proteinaceous\_Fluid at both severity grades 1 and 2, a LOEC of treatment 3 for Gon\_Phenotype at severity grade 1, and LOEC of treatment 5 for Gon\_Phenotype at severity grade 1, not show a response in other scale (as indicated by the "Yes" in the

column "Inverted") so for this analysis a severity grade of 1 is considered to be more severe then a score of 5.

RSCABS sometimes produces exactly the same test statistic and p-value for different endpoints or **R-Scores**. In fact, when data is sparse RSCABS will commonly have exactly the same test statistic (and by extension p-value) for different **R-Scores** of the same endpoint, as shown with Oth\_Thymus\_Lymphoidhyperplasia in the image above. This is due to having 0 observations with some severity score for that endpoint. In the example above, for Oth\_Thymus\_Lymphoidhyperplasia there were two fish with a severity score of 3 (one in treatment 4 and one in treatment 2) and no fish with severity scores of 1 or 2. This causes RSCABS to produce the exact same test statistic for Oth\_Thymus\_Lymphoidhyperplasia severity score 1 and 2 as it does for severity score 3. The images below show how this information can be attained. Noting the frequency table from each image, the first image shows that the there is one fish in treatment 2 and one fish in treatment 4 with a severity score of  $\geq 1$ , while the second image shows the same values of fish in each treatment for a severity score of  $\geq 3$ . From this it can be inferred that there are no fish with an severity score of 1 or 2 within treatment 2 and 4.

Oth_	Thymus_Lyn	npho	oidhy	perpla	asiaP		th_	Thymu	s_L	LymphoidhyperplasiaP2	Oth_Thymus_LymphoidhyperplasiaP3	
				Chi-	Squa	ared H	lete	rogenei	ity	Check of Between-Repl	icate Variances	
/alue 4	Treatment	•	Dosel	lum	• p-	value	•	Doses	•	Signif		
4.21	2		5		0.	51952		б		•		
11.48	4		5		0.	04268		6		*		
								F	Fre	equency Table		
Score/Tr	eatment 4	1 •	2 4	3 (	4 <	5 4	6	Tota				(   )
<1		21	19	6	23	29	5	103				
>=1		0	1	0	1	0	0	2				J.
Total		21	20	6	24	29	5	105				•
							Ra	io-Scot	t C	Cochran-Armitage Test		
Freatme	nt • R-Sco	re (	Stat	istic	• P-	Value	•	Signif				
6	1		-03	34264	់ត	36593						

-								_	
Other	Histology	y Analysis						— 🗆	×
• Oth_	Thymus_l	Lymphoid	hyperpl	asiaP1	Oth	n_Thymus	_LymphoidhyperplasiaP2	Oth_Thymus_LymphoidhyperplasiaP3	•
			Chi	Sauar	ad Hat	erogeneit	y Check of Between-Replic	rate Variances	
	- ·								
Value 4		ent 🖣 Dos	eNum				Signif		•
4.21	2	5		0.5	1952	6	•		
11.48	4	5		0.04	4268	6	*		
Г ———									
						Fi	requency Table		
Score/Tr	reatment	• 1 • 2	4 3 4	4 • 9	5 4 6	<ul> <li>Total</li> </ul>		•	$\wedge$
<3		21 1	9 6	23	29 5	5 103			
>=3		0 1	0	1	0 0	) 2			
Total		21 2	06	24	29 5	i 105			~
		2. 2							
					I	Rao-Scott	Cochran-Armitage Test		
Treatme	nt 🖣 R-S	Score 4 S	tatistic	< P-V	alue 4	Signif			•
6	3		0.34264		5593				
×	5		0.04204	0.50					

The tables above can be obtained by using the [Get Details on a Response] button in the "Other Analyses" section of RSCABS.

# **Running Other Analyses**

More details on each histopathological endpoint can be attained through the use of the [Run Other Analyses] button. This will bring up the Other Histology Analysis window.

- • •

On the Other Histology Analysis window the [Run SCABS] button will run a SCABS analysis on the data set. After the analysis is ran a table will appear with the results of the analysis. This table contains the same information as the results table from the RSCABS analysis.

Run SCABS	Response	Treatment 4	Rscore 4	T-Value 4	P-value 4	Signif
	Gon_Incr_Oocyte_Atresia1	5	1	1.34049	0.09004	
Get Details on a Response	Gon_Incr_Oocyte_Atresia2	5	2	1.27525	0.10111	
Get Details on all Responses	Gon_Incr_Oocyte_Atresia3	5	3	1.27525	0.10111	
Save	Gon_Incr_Oocyte_Atresia4	5	4	1.27525	0.10111	
	Gon_Proteinaceous_fluid1	5	1	-1.51815	0.06449	
	Gon_Asynch_Dev1	5	1	2.74448	0.00303	**
	Gon_Asynch_Dev2	5	2	2.74448	0.00303	**
	Kid_Tub_Mineralization1	5	1	0.39734	0.34556	
	Kid_Tub_Mineralization2	5	2	-0.08038	0.46797	
	Kid_Tub_Dilation1	5	1	-0.08038	0.46797	
	Kid_Tub_Dilation2	5	2	-0.05564	0.47782	
	Kid_Tub_Regeneration1	5	1	-1.44659	0.07401	

The [Get Details on a Response] button will supply three tables for the selected response; a table for the chi-squared  $(\chi^2)$  test for heterogeneity of between-replicate variances. A frequency table, which contains the total observations for each combination of treatment (shown in the columns) and slice of score (shown in the rows). There will also be an additional table showing the results of RSCABS for that treatment level. If there are several unique severity scores for an endpoint, results for the lowest severity score will be displayed in the main tab while results for additional severity scores will be added to the window in extra tabs. The [Save] button will save the current result displayed in the window, whether it is results from SCAB or the more detailed results.

	[									
Run SCABS	(	hi-Squa	ared He	terog	eneit	y Ch	eck of Betwe	een-Replicate Variances		
Get Details on a Response	Value   Treatm	nent 4	DoseN	um (	P-v	alue	Doses	Signif		•
Get Details on all Responses	6 5		5		0.3	0622	5			
Save										
Zane										
					F	requ	ency Table			
	Score/Treatmen	t • 1 •	2 4	3 4	4 •	5 4	Total		•	4
	<1	4	5	3	5	2	19			Ξ
	>=1	0	0	0	0	4	4			
	<b>T</b> 1 1			2		<u>,</u>	22			1
	Rao-Scott Cochran Armitage Test									
	Treatment • R	core 4	T-Val	ue (	P-Va	alue	<ul> <li>Signif</li> </ul>			
	5 1		2.622	02	0.00	437	**			

Finally, the [Get Details on all Responses] button will produce the three tables generated by [Get Details on a Response] for all responses. It does this by creating a new folder and populating that folder with HTML files containing the information.

# **RSCABS** Command Line

As an alternative to using the GUI, RSCABS can also be run through command line. This is done through the runRSCABS function.

R Input

>runRSCABS(Data, Treatment, Replicate, Effects ,test.type)

The **Data** variable is a data set in the same format needed for the **GUI**. **Treatment**, is the name (as a string) of the treatment variable, **Replicate** is the name (as a string) of the replicate variable, and **Effects** is a character vector of the endpoints exampleHistData.Sub tested. If **Effects** is not specified then the analysis will default to all columns that has at least one integer greater than 0 and no integers less then 20. The **test.type** input can be either "RS" or "CA" for either the RSCABS analysis or the SCABS analysis respectively. The **Replicate** input does not need to be specified, however if it is not specified **test.type** will default to "CA". An example of using runRSCABS is below. The code will produce the same results table as the example analysis in the **Running RSCABS** section.

```
#Take the subset corresponding to F0-females of 16 weeks of age
data(exampleHistData)
subIndex<-which(exampleHistData$Generation=='F2' &
        exampleHistData$Genotypic_Sex=='Female' &
        exampleHistData$Age=='16_wk' )
exampleHistData.Sub<-exampleHistData[subIndex, ]
#Run RSCABS
exampleResults<-runRSCABS(exampleHistData.Sub,'Treatment',
        'Replicate',test.type='RS')
```

# Plotting

#### Plotting by GUI

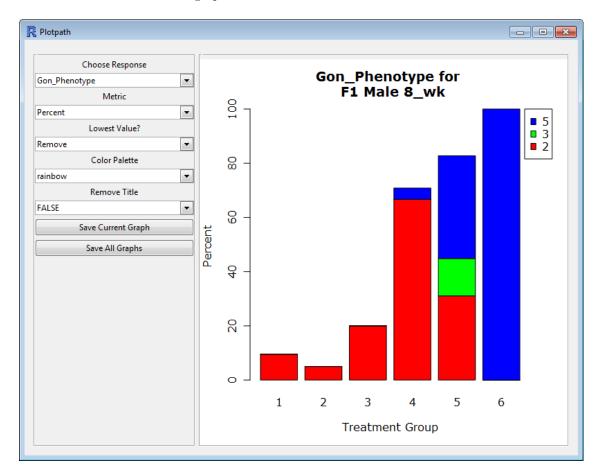
The plotting window should appear after the [Run RSCABS] button from the Histopath main window is pressed.

R Input



The plotting window uses stacked bar plots for visual representation of histopathological data. The left side of the window contains the plotting controls. Starting from the top of the left side; the **Choose Response** category allows for selections of the endpoint to graph. The list of possible endpoints is generated from the endpoint fields in the data set. Next, **Metric** controls

the y-axis as it allows for graphing either total counts or for percent of the total observations for each severity score and treatment combination. The entry "Lowest Value?" allows for the removal of the lowest value, which is in most cases 0 and indicates that the endpoint was not observed for an organism. Selecting "Percent" as the metric and choosing to "Remove" the lowest value will allow for easer observation of treatment effects when they are present. The Color Pallet option allows for the use of any of the standard R color pallets (including grey scale) and Remove Title will remove the title from the graph.



The **Save Current Graph** will save the current graph being displayed in any of the default file formats available to R. Selecting "tiff" will generate high quality tiff images at 600 dpi. The **Save All Graphs** button will generate graphs using the current setting for all endpoints containing at least one severity score greater then zero.

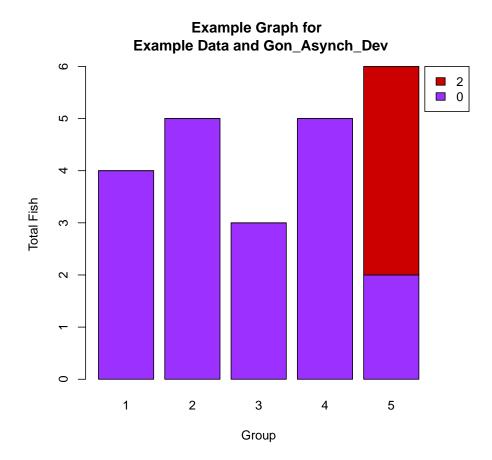
F	Rease select file format	
	Values pdf	•
	bmp	
	jpeg	
	png	
	tiff	
	select	

#### Plotting by Command Line

When greater control over the plot is desired, the plotting function can be called through the command line using the plotRSCABS function as shown in the example below.

R Input -

```
#Sub-set the data
require(RSCABS)
data(exampleHistData)
subIndex<-which(exampleHistData$Generation=='F2' &
    exampleHistData$Genotypic_Sex=='Female' &
    exampleHistData$Age=='16_wk')
exampleHistData.Sub<-exampleHistData[subIndex, ]
xlab<-'Group'
ylab<-'Total Fish'
main<-'Example Graph for \n Example Data and Gon_Asynch_Dev'
col<-c('purple1','red3')
plotRSCABS(Data=exampleHistData.Sub, Effect="Gon_Asynch_Dev",
        Treatment="Treatment", Metric="Total", Lowest = "Include",
        PlotParms =PlotParms, Format = NULL, File = NULL,
        xlab=xlab,main=main,ylab=ylab,col=col)
```



The arguments of the function are; **Data**, which is a standard data set used by RSCABS, **Effect** which is the name of the endpoint plotted, and **Treatment** which is the name of the treatment variable. **Metric** controls for plotting either total counts ("Total") or percent of total counts ("Percent"), while **Lowest** allows for the removal ("Remove") or inclusion ("Include") of non-effected responses. **Format** is the name of any file format R can save graphs in, including high resolution graphs which are saved using the 'tiff' format. **File** is the name of the file the graph is saved to. Lastly, the **plotRSCABS** function can also include any arguments used by the **barplot** function.

### Acknowledgments

I would like to acknowledge Rodney Johnson and Kevin Flynn for their guidance through this project, Kevin Flynn as a beta tester, and Tim Dawson as a reviewer of the documentation. The R version of RSCABS was built for and paid by the USEPA under Contract EPD—13—052.

# References

- Armitage, P. 1955. Tests for Linear Trends in Proportions and Frequencies. Biometrics 11(3): 417-451
- Cochran, WG 1954. Some methods for strengthening the common chi-squared tests. Biometrics 10(4): 417-451

- Green, John W. and Springer, Timothy A. and Saulnier, Amy N. and Swintek, Joe. 2014 Statistical analysis of histopathological endpoints. *Environmental Toxicology and Chemistry*, 33(5): 1108–1116
- 4. OECD Guidelines for the Testing of Chemicals, Section 2. Test No. 240. DOI10.1787/20745761
- Rao, J. N. K. & Scott, A. J. A. 1992. Simple Method for the Analysis of Clustered Data. Biometrics, 48: 577-586.