

AS02



Experis®
ManpowerGroup

An Introduction to Obtaining Test Statistics and P-Values from SAS® and R for Clinical Reporting

Brian Varney, Experis



Agenda

- 1** Introduction
- 2** SAS® Approach Using Output Statements w/ PROC FREQ
- 3** SAS® Approach Using ODS TRACE w/ PROC FREQ
- 4** SAS® Approach Examples w/ PROC GLM
- 5** R Approach Using List from `chisq.test()` Function
- 6** R Approach Using Broom Package Functions on Results from `chisq.test()` Function
- 7** R Approach Examples w/ `stats::aov()` and `sasLM::GLM()` function
- 8** Conclusions & Q&A

Introduction

- SAS uses procedures to obtain the statistics required for clinical output reporting.
- There are a couple of methods for getting those results in a usable form.
 - Using output data set options on procedure
 - Using ODS TRACE and ODS OUTPUT
- R uses functions to obtain the statistics required for clinical output reporting.
- There are a couple of methods for getting those results in a usable form.
 - Using default list output by the function.
 - Using broom package functions on the list output by the function.

SAS Approach Using Output Statements with PROC FREQ

```
proc freq data=heart;  
    table sex*diabetes / list chisq fisher cmh;  
    output out=freqstats chisq fisher cmh;  
quit;
```

VIEWTABLE: Work.Freqstats

	N	_PCHI_	DF_PCHI	P_PCHI	_LRCHI_	DF_LRCHI	P_LRCHI	_AJCHI_	DF_AJCHI	P_AJCHI	_MHCHI_	DF_MHCHI	P_MHCHI	XPL_F ^
1	299	7.4387003651	1	0.0063835972	7.4030522611	1	0.0065113304	6.7838527579	1	0.009198613	7.4138217685	1	0.0064724687	*****
2	299	.	.	.	.	.	.	.	.	.	.	.	.	.

VIEWTABLE: Work.Freqstats

	XPL_FISH	XPR_FISH	XP2_FISH	_PHI_	_CONTGY_	_CRAMV_	_CMHCOR_	DF_CMHCO	P_CMHCOR	_CMHRMS_	DF_CMHRM	P_CMHRMS	_CMHGA_
1	0.0046720481	0.9977744115	0.0070598512	-0.157729504	0.1558033268	-0.157729504	.	.	.	.	.	.	.
2	.	.	.	.	.	.	7.4138217685	1	0.0064724687	7.4138217685	1	0.0064724687	7.4138217685

VIEWTABLE: Work.Freqstats

	DF_CMHGA	P_CMHGA	_MHOR_	L_MHOR	U_MHOR	_LGOR_	L_LGOR	U_LGOR	_MHRRC1_	L_MHRRC1	U_MHRRC1	_LGRRC1_	L_LGRRC1
1	.	.	.	.	.	.	.	.	.	.	.	.	.
2	1	0.0064724687	0.5131964809	0.3168548262	0.8312028295	0.5131964809	0.3168548262	0.8312028295	0.7450076805	0.5938520854	0.9346375261	0.7450076805	0.5938520854

VIEWTABLE: Work.Freqstats

	U_LGRRC1	_MHRRC2_	L_MHRRC2	U_MHRRC2	_LGRRC2_	L_LGRRC2	U_LGRRC2
1	.	.	.	.	.	.	.
2	0.9346375261	1.4517006803	1.1177571541	1.8854138911	1.4517006803	1.1177571541	1.8854138911

SAS Approach Using ODS TRACE w/ PROC FREQ

```
ods trace on;  
proc freq data=heart;  
    table sex*diabetes / list chisq fisher cmh;  
quit;
```

- As you will see on the next slide, the log will contain the available pieces of output that can be requested as a SAS Data Set.

SAS Approach Using ODS TRACE w/ PROC FREQ (continued)

Output Added:

Name: CrossTabFreqs
Label: Cross-Tabular Freq Table
Template: Base.Freq.CrossTabFreqs
Path: Freq.Table1.CrossTabFreqs

Output Added:

Name: ChiSq
Label: Chi-Square Tests
Template: Base.Freq.ChiSq
Path: Freq.Table1.ChiSq

Output Added:

Name: FishersExact
Label: Fisher's Exact Test
Template: Base.Freq.ChisqExactFactoid
Path: Freq.Table1.FishersExact

Output Added:

Name: CMH
Label: Cochran-Mantel-Haenszel Statistics
Template: Base.Freq.CMH
Path: Freq.Table1.CMH

Output Added:

Name: CommonRelRisks
Label: Common Relative Risk Estimates
Template: Base.Freq.CommonRelRisks
Path: Freq.Table1.CommonRelRisks

SAS Approach Using ODS TRACE w/ PROC FREQ (continued)

- Once you have chosen, the statistics you would like in a SAS Data Set, you can use ods output statements to request them.

```
ods output chisq=chisqout;  
ods output fishersexact=fishersexactout;  
ods output cmh=cmhout;
```

```
proc freq data=heart;  
    table sex*diabetes / list chisq fisher cmh;  
quit;
```

NOTE: The data set WORK.CMHOUT has 3 observations and 6 variables.

NOTE: The data set WORK.FISHERSEXACTOUT has 6 observations and 5 variables.

NOTE: The data set WORK.CHISQOUT has 7 observations and 5 variables.

SAS Approach Using ODS TRACE w/ PROC FREQ (continued)

- The resulting SAS Data Sets from the ods output statements are below.

chisqout SAS Data Set

VIEWTABLE: Work.Chisqout (Chi-Square Tests)					
	Table	Statistic	DF	Value	Prob
1	Table sex * diabetes	Chi-Square	1	7.4387	0.0064
2	Table sex * diabetes	Likelihood Ratio Chi-Square	1	7.4031	0.0065
3	Table sex * diabetes	Continuity Adj. Chi-Square	1	6.7839	0.0092
4	Table sex * diabetes	Mantel-Haenszel Chi-Square	1	7.4138	0.0065
5	Table sex * diabetes	Phi Coefficient	—	-0.1577	—
6	Table sex * diabetes	Contingency Coefficient	—	0.1558	—
7	Table sex * diabetes	Cramer's V	—	-0.1577	—

fishersexactout SAS Data Set

VIEWTABLE: Work.Fishersexactout (Fisher's Exact Test)					
	Table	Name1	Label1	cValue1	nValue1
1	Table sex * diabetes	Cell1_FREQ	Cell (1,1) Frequency (F)	50	50.000000
2	Table sex * diabetes	XPL_FISH	Left-sided Pr <= F	0.0047	0.004672
3	Table sex * diabetes	XPR_FISH	Right-sided Pr >= F	0.9978	0.997774
4	Table sex * diabetes				.
5	Table sex * diabetes	P_TABLE	Table Probability (P)	0.0024	0.002446
6	Table sex * diabetes	XP2_FISH	Two-sided Pr <= P	0.0071	0.007060

cmhout SAS Data Set

VIEWTABLE: Work.Cmhout (Cochran-Mantel-Haenszel Statistics)						
	Table	Statistic	AltHypothesis	DF	Value	Prob
1	Table sex * diabetes	1	Nonzero Correlation	1	7.4138	0.0065
2	Table sex * diabetes	2	Row Mean Scores Differ	1	7.4138	0.0065
3	Table sex * diabetes	3	General Association	1	7.4138	0.0065

R Approach Using List from chisq.test() Function

```
chisq <- stats::chisq.test(heart$sex, heart$diabetes)
```

Name	Type	Value
chisq	list [9] (S3: htest)	List of length 9
statistic	double [1]	6.783853
parameter	integer [1]	1
p.value	double [1]	0.009198613
method	character [1]	'Pearson\'s Chi-squared test with Yates\' continuity correction'
data.name	character [1]	'data_heart\$sex and data_heart\$diabetes'
observed	integer [2 x 2] (S3: table)	124 50 70 55
expected	double [2 x 2]	112.9 61.1 81.1 43.9
residuals	double [2 x 2] (S3: table)	1.05 -1.42 -1.23 1.68
stdres	double [2 x 2] (S3: table)	2.73 -2.73 -2.73 2.73

R Approach Using List from `chisq.test()` Function (continued)

- It is easy to pick off the test statistic and p-value from the `chisq` list object as shown below.

```
chisq_teststat <- chisq$statistic  
chisq_pvalue <- chisq$p.value
```

values	
chisq_pvalue	0.00919861301978445
chisq_stat	Named num 6.78

- There are separate functions in R for other tests such as Fisher's Exact Test (`fisher.test()`).

R Approach Using Broom Package Functions on Results from `chisq.test()` Function

Example Situation R Code:

For the sake of this paper, we will just look at the `tidy()` function as we just want to save the test statistic and p-value into a tibble; which is easier to work with versus a list.

```
tidy_stats_chisq <- tidy(stats_chisq)
```

Resulting tibble named `tidy_stats_chisq`

	statistic	p.value	parameter	method
1	6.783853	0.009198613	1	Pearson's Chi-squared test with Yates' continuity correction

SAS Approach Using Output Statements with PROC GLM

- There is an option in PROC GLM called `outstat` which will output the test statistics and p-values to the designated output data set.

```
proc glm data=heart outstat=glmoutstat;  
  class sex;  
  model platelets = sex anaemia sex*anaemia;  
quit;
```

VIEWTABLE: Work.Glmoutstat

	NAME	_SOURCE_	_TYPE_	DF	SS	F	PROB
1	platelets	ERROR	ERROR	295	2.7915049E12	.	.
2	platelets	sex	SS1	1	44626047643	4.7159809164	0.0306801795
3	platelets	anaemia	SS1	1	8905783533.7	0.9411432876	0.3327787142
4	platelets	anaemia*sex	SS1	1	5532509892.7	0.5846632729	0.4451014602
5	platelets	sex	SS3	1	44766901095	4.7308659942	0.0304200658
6	platelets	anaemia	SS3	1	12391677511	1.3095247675	0.2534083666
7	platelets	anaemia*sex	SS3	1	5532509892.7	0.5846632729	0.4451014602

SAS Approach Using ODS TRACE w/ PROC GLM

After running the PROC GLM with the ods trace on, you will notice a couple of ods trace output elements of interest which contain test statistics and p-values.

Output Added:

Name: OverallANOVA
Label: Overall ANOVA
Template: stat.GLM.OverallANOVA
Path:
GLM.ANOVA.platelets.OverallANOVA

Output Added:

Name: ModelANOVA
Label: Type I Model ANOVA
Template: stat.GLM.Tests
Path:
GLM.ANOVA.platelets.ModelANOVA

SAS Approach Using ODS TRACE w/ PROC GLM (continued)

- The resulting SAS Data Sets from the ods output statements are below.

Overall ANOVA Resulting Data Set

VIEWTABLE: Work.Overallanovaout (Overall ANOVA)

	Dependent	Source	DF	SS	MS	FValue	ProbF
1	platelets	Model	3	59064341069	19688113690	2.08	0.1029
2	platelets	Error	295	2.7915049E12	9462728631.4	—	—
3	platelets	Corrected Total	298	2.8505693E12	—	—	—

Model ANOVA Resulting Data Set

VIEWTABLE: Work.Modelanovaout (Type I Model ANOVA)

	Dependent	HypothesisType	Source	DF	SS	MS	FValue	ProbF
1	platelets	1 sex		1	44626047643	44626047643	4.72	0.0307
2	platelets	1 anaemia		1	8905783534	8905783534	0.94	0.3328
3	platelets	1 anaemia*sex		1	5532509893	5532509893	0.58	0.4451
4	platelets	3 sex		1	44766901095	44766901095	4.73	0.0304
5	platelets	3 anaemia		1	12391677511	12391677511	1.31	0.2534
6	platelets	3 anaemia*sex		1	5532509893	5532509893	0.58	0.4451

R Approach Using List from aov() Function

```
aov1 <- aov(platelets ~ sex + anaemia + sex:anaemia, data = data_heart)
```

Name	Type	Value
aov1	list [13] (S3: aov, lm)	List of length 13
coefficients	double [4]	256160 35033 -4510 -18164
residuals	double [299]	8840 7198 -94160 -41650 58481 -47650 ...
effects	double [299]	-4553886 211249 -94370 -74381 49418 -52695 ...
rank	integer [1]	4
fitted.values	double [299]	256160 256160 256160 251650 268519 251650 ...
assign	integer [4]	0 1 2 3
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	295
contrasts	list [2]	List of length 2
xlevels	list [2]	List of length 2
call	language	aov(formula = platelets ~ sex + anaemia + sex:anaemia, data = data_heart)
terms	formula	platelets ~ sex + anaemia + sex:anaemia
model	list [299 x 3] (S3: data.frame)	A data.frame with 299 rows and 3 columns

- Trying to find and pick off the elements off of the list from the aov() and similar functions is cumbersome. I would not recommend it.

R Approach Using Broom Package Functions on Results from aov() Function

R Code:

```
aov1_tidy <-  
  aov1 %>%  
  tidy()
```

Resulting aov1_tidy tibble

	term	df	sumsq	meansq	statistic	p.value
1	sex	1	4.462605e+10	44626047643	4.7159809	0.03068018
2	anaemia	1	8.905784e+09	8905783534	0.9411433	0.33277871
3	sex:anaemia	1	5.532510e+09	5532509893	0.5846633	0.44510146
4	Residuals	295	2.791505e+12	9462728631	NA	NA

R Approach Using GLM() function from the sasLM() Package

```
sasLM_GLM <- GLM(platelets ~ sex + anaemia + sex:anaemia, data_heart)
```

```
$ANOVA
Response : platelets
      Df      Sum Sq    Mean Sq F value Pr(>F)
MODEL      3 5.9064e+10 1.9688e+10  2.0806 0.1029
RESIDUALS 295 2.7915e+12 9.4627e+09
CORRECTED TOTAL 298 2.8506e+12
```

```
$Fitness
Root MSE platelets Mean Coef Var   R-square   Adj R-sq
97276.56      263358    36.937 0.02072019 0.01076141
```

```
$`Type I`
      Df      Sum Sq    Mean Sq F value Pr(>F)
sex      1 4.4626e+10 4.4626e+10  4.7160 0.03068 *
anaemia  1 8.9058e+09 8.9058e+09  0.9411 0.33278
sex:anaemia 1 5.5325e+09 5.5325e+09  0.5847 0.44510
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
$`Type II`
      Df      Sum Sq    Mean Sq F value Pr(>F)
sex      1 4.8067e+10 4.8067e+10  5.0796 0.02494 *
anaemia  1 8.9058e+09 8.9058e+09  0.9411 0.33278
sex:anaemia 1 5.5325e+09 5.5325e+09  0.5847 0.44510
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
$`Type III`
      Df      Sum Sq    Mean Sq F value Pr(>F)
sex      1 4.5173e+10 4.5173e+10  4.7737 0.02968 *
anaemia  1 1.2392e+10 1.2392e+10  1.3095 0.25341
sex:anaemia 1 5.5325e+09 5.5325e+09  0.5847 0.44510
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Although the output looks much more SAS like, work is required to grab the desired statistics from the list.

R Approach Using GLM() function from the sasLM() Package

```
sasLM_GLM <- GLM(platelets ~ sex + anaemia + sex:anaemia, data_heart)
```

The list →

```
↓ sasLM_GLM      List of 5
 $ ANOVA      : 'anova' num [1:3, 1:5] 3.00 2.95e+02 2.98e+02 5.91e+10 2.79e+12 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:3] "MODEL" "RESIDUALS" "CORRECTED TOTAL"
 .. ..$ : chr [1:5] "Df" "Sum Sq" "Mean Sq" "F value" ...
 ..- attr(*, "heading")= chr "Response : platelets"
 $ Fitness    : 'data.frame':      1 obs. of  5 variables:
 ..$ Root MSE      : num 97277
 ..$ platelets Mean: num 263358
 ..$ Coef var      : num 36.9
 ..$ R-square      : num 0.0207
 ..$ Adj R-sq      : num 0.0108
 $ Type I     : 'anova' num [1:3, 1:5] 1.00 1.00 1.00 4.46e+10 8.91e+09 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:3] "sex" "anaemia" "sex:anaemia"
 .. ..$ : chr [1:5] "Df" "Sum Sq" "Mean Sq" "F value" ...
 $ Type II    : 'anova' num [1:3, 1:5] 1.00 1.00 1.00 4.81e+10 8.91e+09 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:3] "sex" "anaemia" "sex:anaemia"
 .. ..$ : chr [1:5] "Df" "Sum Sq" "Mean Sq" "F value" ...
 $ Type III   : 'anova' num [1:3, 1:5] 1.00 1.00 1.00 4.52e+10 1.24e+10 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:3] "sex" "anaemia" "sex:anaemia"
 .. ..$ : chr [1:5] "Df" "Sum Sq" "Mean Sq" "F value" ...
```

R Approach Using GLM() function from the sasLM() Package

```
sasLM_GLM <- GLM(platelets ~ sex + anaemia + sex:anaemia, data_heart)
```

```
# ANOVA tables of Type I, II, III
```

```
SSanova <- as.data.frame(unclass(sasLM_GLM$`ANOVA`))
```

```
fitness <- as.data.frame(sasLM_GLM$Fitness)
```

```
type1 <- as.data.frame(unclass(sasLM_GLM$`Type I`))
```

Conclusions

- Obtaining test statistics from SAS and R is not difficult.
- In SAS you work with procedures and ODS.
- In R you work with functions, lists, and the broom package.
- It is important to understand the default options and statistical methodologies each software is using so you can obtain the results you are expecting.

Thank you

Ask an Experis representative to set up a demonstration of Accel2R today!

