

## A Comparison of SAS and Stata for Survey Corrected Poisson Regression

Pat Berglund

January 20, 2015

No SAS SURVEY procedure is currently available (as of SAS 9.4) for design-based Poisson regression. Instead, we present a SAS JRR macro that is modified to perform Poisson regression. The results are then compared with the Stata svy: poisson command. The results indicate very minor differences between the two programs.

### SAS Code for Poisson Regression with a JRR variance estimation approach:

```
***** ;
*Poisson regression* ;
*Berglund, updated 11 17 214 ;
***** ;
libname d 'f:\brahms\summerclasses\' ;
options compress=yes nofmterr symbolgen ;
options macrogen mprint;
data d.ncsdxdm3_poisson ;
    set d.ncsdxdm3 ;
* create a count of accidents during past month ;
num_accidents_month=floor(ranuni(456)*100/10) ;
if num_accidents_month in (1,2,3) then num_accidents_month=0 ; else if
num_accidents_month > 3 then num_accidents_month=num_accidents_month - 3 ;
* create an offset variable set to 30 days per month ;
offset_30=30 ;
ln_offset_30=log(offset_30) ;

proc freq ;
    tables ln_offset_30 offset_30 num_accidents_month ;
run ;

%macro jackpoisson (ncluster,weight,depend,preds,indata);
%let nclust=%eval(&ncluster);

data one;
    set &indata;

%macro wgtcal ;
    %do i=1 %to &nclust ;
        pwt&i=&weight;
        if str=&i and secu=1 then pwt&i=pwt&i*2 ;
        if str=&i and secu=2 then pwt&i=0 ;
    %end;
%mend;

%wgtcal ;

%macro base ;
ods output parameterestimates=parms (keep=parameter estimate ) ;
title "Example of Proc GLM with log link : no design correction" ;
proc genmod data=ONE ;
    model &depend=&preds / dist=poisson link=log offset=ln_offset_30 ; * use
distribution of interest here: this is POISSON but could be another choice ;
    weight &weight ;
run ;
proc sort ;
```

```

        by parameter ;
run ;

%mend base ;
%base ;

%macro reps ;
    %do j=1 %to &nclust ;
        ods output parameterestimates=parms&j (keep=parameter estimate
rename=(estimate=estimate&j )) ;
        *ods listing close ;

        proc genmod data=ONE ;
            model &depend=&preds / dist=poisson link=log offset=ln_offset_30 ;
            weight pwt&j ;
        run ;
        proc sort ;
            by parameter ;
        %end ;
    %mend reps;
%reps ;

data rep ;
merge parms
    %do k=1 %to &nclust;
        parms&k
    %end;;
by parameter ;
proc print ;
run ;

ods listing ;
data calculate ;
set rep ;
%macro it ;
    %do j=1 %to &nclust ;
        sqdiff&j=(estimate-estimate&j)**2;
    %end;

    sumdiff=sum(of sqdiff1-sqdiff&nclust);
    stderr=sqrt(sumdiff) ;

%mend it ;
%it;
run ;

proc print ;
title "Results from JRR for Poisson regression" ;
var parameter estimate stderr ;
run ;

%mend jackpoisson ;

%jackpoisson(42,p2wtv3,num_accidents_month,sexf ag25 ag35 ag45,d.ncsdxdm3_poisson) ;

```

SAS Results:

**Results from JRR for Poisson regression**

<b>Obs</b>	<b>Parameter</b>	<b>Estimate</b>	<b>stderr</b>
<b>1</b>	<b>AG25</b>	<b>-0.0084</b>	<b>0.058798</b>
<b>2</b>	<b>AG35</b>	<b>0.0165</b>	<b>0.069871</b>
<b>3</b>	<b>AG45</b>	<b>-0.0252</b>	<b>0.065057</b>
<b>4</b>	<b>Intercept</b>	<b>-2.6881</b>	<b>0.042559</b>
<b>5</b>	<b>SEXF</b>	<b>0.0954</b>	<b>0.039754</b>
<b>6</b>	<b>Scale</b>	<b>1.0000</b>	<b>0.000000</b>

## Stata Syntax

\* comparison to JRR presentation 2013 for Sampling Workshop

```
use "f:\brahms\summerclasses\ncsdxdm3_poisson.dta", clear
```

```
svyset secu [pweight=P2WTV3], strata(str)
```

```
svy: poisson num_accidents_month sexf AG25 AG35 AG45, exposure (offset_30)
```

## Stata Results

```
. * comparison to JRR presentation 2013 for Sampling Workshop
.
. use "f:\brahms\summerclasses\ncsdxdm3_poisson.dta", clear
.
. svyset secu [pweight=P2WTV3], strata(str)

      pweight: P2WTV3
             VCE: linearized
Single unit: missing
  Strata 1: str
      SU 1: secu
      FPC 1: <zero>

. svy: poisson num_accidents_month sexf AG25 AG35 AG45, exposure (offset_30)
(running poisson on estimation sample)
```

Survey: Poisson regression

Number of strata	=	42	Number of obs	=	5877
Number of PSUs	=	84	Population size	=	5877.0004
			Design df	=	42
			F( 4, 39)	=	1.40
			Prob > F	=	0.2519

num_accidents_month	Linearized			P> t	[95% Conf. Interval]	
	Coef.	Std. Err.	t			
sexf	.0954446	.0398672	2.39	0.021	.0149894	.1758998
AG25	-.0084081	.0591208	-0.14	0.888	-.1277187	.1109024
AG35	.0164695	.0696628	0.24	0.814	-.1241158	.1570548
AG45	-.0252236	.06505	-0.39	0.700	-.1564998	.1060526
_cons	-2.688079	.0426563	-63.02	0.000	-2.774163	-2.601995
ln(offset_30)	1	(exposure)				