



Brandy Sinco, Brief Bio, Statistician Senior Center for Healthcare Policy and Outcomes (CHOP)

Brandy holds a bachelors degree in engineering, and masters degrees in mathematics and statistics.

She joined CHOP in September 2019, after working on central campus for over 20 years (Wolverine Tower, School of Public Health, School of Social Work). Worked at Irwin Magnetics, Med-Stat, and Ford before UM.

Outside of work, she enjoys playing the piano, composing music, singing in choir, yoga, tai chi, cardio karate, and making greeting cards.

Using Excel to Generate SAS and Stata Code and to Process Output



Brandy R. Sinco, BS, MA. MS



Outline

- 1) Useful Excel Functions and Applications (Display Missing + Asterisks).
- 2) Generate Attribute Statements in SAS and Properties in Stata.
- 3) SAS Macro Invocation, Commands for Stata .do file.
- 4) Generate Formulas for Computed Variables.
- 5) Estimate Statement.
- 6) Odds Ratio Plot.
- 7) Convert Stata syntax to SAS and vice versa
- 8) Closing Comments.



Useful Excel Functions, 1 of 2

- Let cell A1 = Ann Arbor.
- Len(A1). Number of characters of string in cell. Example $\text{len}(A1) = 9$.
- Find(find_text, within_text, start_num). Example $\text{find}(\text{"n"}, A1, 1) = 2$.
- Mid(A1, start, numchar). Sub-string from cell A1, beginning at "start" for character "numchar". Example $\text{mid}(A1, 2, 4) = \text{nn A}$.
- Left(A1, numchar). First numchar left characters. Example $\text{left}(A1, 2) = \text{An}$.
- Right(A1, numchar). Last numchar right characters. Example $\text{right}(A1, 3) = \text{bor}$.
- Trim(A1) removes extra spaces from right and left to the text.



Useful Excel Functions, 2 of 2

- **& = Concatenate.** Example `left(A1, 4) & right(A1, 5) = Ann Arbor.`
- `if(condition, true value, false value).`
- Examples: `if(B1<5, 1, 0)`, `if(C1<2, "low", "high")`.
- `Fixed(cell, number of decimals, thousands comma).`
- Default number of decimals = 2. `fixed(73.567) = 73.57.`
- Let `B1 = 10000.789.`
- `Fixed(B1, 1, 0) = 10,000.8.`
- `Fixed(B1, 2, 1) = 10000.79.`
- 0 enables thousands comma; 1 omits thousands comma.



Use Excel Functions to Extract N (%) Missing

TraumaIP	FPL_JWS	Frequency	Weighted Frequency
No	<0.50x FPL	292	100523
	0.50-1.49x FPL	778	272813
	1.50-3.99x FPL	1631	607989
Yes	<0.50x FPL	106	32856
	0.50-1.49x FPL	282	85197
	1.50-3.99x FPL	319	104354

- Frequency Missing = 543. Let cell A5 contain the total person count = 10,458 and cell I158 contain a cell count.
- `=FIXED(I158,0,0)&" ("&FIXED(100*I158/A5,1)&"%)” missing“` will generate 543 (5.2%) missing.
- Note: `Fixed(A1,0,0)` f1st 0 = 0 digits after decimal, 2nd 0 = commas after 000. Example: 1,025 (25.1%).



Use Excel to Check Variable and Label Lengths

Variable	Label	Length
MEDBILL	Problems paying medical bills	29
MEDBNOP	Unable to pay medical bills	27
FMEDBILL	Problems paying or unable to pay any medical bills	50

- Use len() to check widths of variable names and labels before writing attribute statements.
- In SAS and Stata, maximum variable name characters = 32.
- SAS, max label length = 40 characters.
- Stata, max label length = 80 characters.

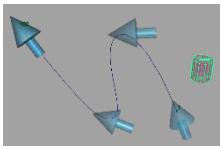


Use Excel to Generate Asterisks for P-Values

Table of Lumpectomy Re-Operation Rates by Facility Volume

2004%	2013%	2014%	2016%	Δ 2004-2013	Δ 2013-2014	Δ 2014-2016
19.1	17.6	14.9	15.1	-1.5**	-2.7***	0.1
19.4	18	15.8	15.3	-1.4***	-2.2***	-0.5
22.1	20.6	17.6	16	-1.5***	-2.9***	-1.7***

- Suppose % estimate is in cell C1 and p-value is in cell D1.
- To get % with *'s for p-values (*: <.05, **: <.01, ***: <.001),
- =fixed(C1,1) & "% " & if(D1<.05,"*", "") & if(D1<.01,"**", "") & if(D1<.001,"***", "").



Generate Variable Attributes - SAS



Variable	Label	FORMAT
clnd	Completion Lymph Node Dissection	YNF
primary_site	Location of Primary Tumor	SiteF
ajcc8	AJCC 8th Edition Stage	StageF

SAS

```
attrib clnd label='Completion Lymph Node Dissection' format=YNF.;
```

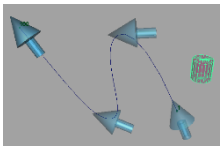
```
attrib primary_site label='Location of Primary Tumor' format=SiteF.;
```

```
attrib ajcc8 label='AJCC 8th Edition Stage' format=StageF.;
```

```
= "attrib "&TRIM(A2)&" label=""&TRIM(B2)&" format=""&C2&" .;"
```

```
= "attrib "&TRIM(A3)&" label=""&TRIM(B3)&" format=""&C3&" .;"
```

```
= "attrib "&TRIM(A4)&" label=""&TRIM(B4)&" format=""&C4&" .;"
```



Generate Variable Labels - Stata



Variable	Label	Values
clnd	Completion Lymph Node Dissection	YNF
primary_site	Location of Primary Tumor	SiteF
ajcc8	AJCC 8th Edition Stage	StageF
Stata Labels		

```
label variable clnd "Completion Lymph Node Dissection";
```

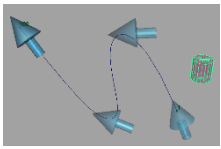
```
label variable primary_site "Location of Primary Tumor";
```

```
label variable ajcc8 "AJCC 8th Edition Stage";
```

```
= "label variable "&A2&" ""&TRIM(B2)&""&";"
```

```
= "label variable "&A3&" ""&TRIM(B3)&""&";"
```

```
= "label variable "&A4&" ""&TRIM(B4)&""&";"
```

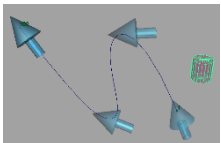


Generate Value Labels – Stata



Value Column A	Label Column B	label define fap001x	Excel Formula
10	Household	10 "Household"	=A2&" "&B2&"&"
20	Person	20 "Person"	=A3&" "&B3&"&"
30	Sample Adult	30 "Sample Adult"	=A4&" "&B4&"&"
40	Sample Adult	40 "Sample Adult"	=A5&" "&B5&"&"

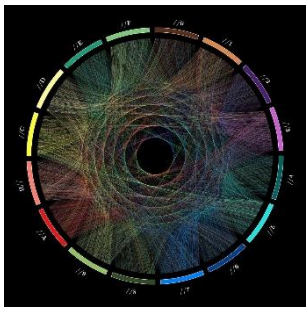
```
Label define fap001x  
10 "Household"  
20 "Person"  
30 "Sample Adult"  
40 "Sample Adult"  
; /* Add ; at end. */
```



Associate Variables with Label Definitions – Stata



Variable Name	Label Definitions	Syntax	Excel Formula
rectype	fap001x	label values rectype fap001x;	= "label values "&A2&" "&B2&" ;"
fint_y_p	fap006x	label values fint_y_p fap006x;	= "label values "&A3&" "&B3&" ;"
phoneuse	fap012x	label values phoneuse fap012x;	= "label values "&A4&" "&B4&" ;"

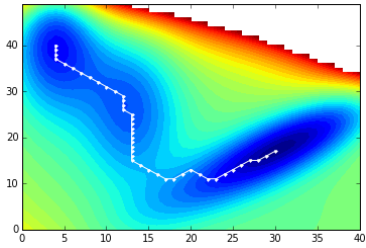


Macro Invocation

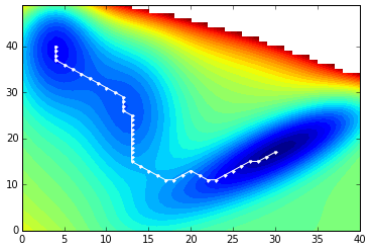
any_afford	%LogitUn(any_afford);	="%LogitUn(" & trim(I7) & ");"
any_forgone	%LogitUn(any_forgone);	="%LogitUn(" & trim(I8) & ");"
any_access	%LogitUn(any_access);	="%LogitUn(" & trim(I9) & ");"
any_debt	%LogitUn(any_debt);	="%LogitUn(" & trim(I10) & ");"
any_altered	%LogitUn(any_altered);	="%LogitUn(" & trim(I11) & ");"
any_food	%LogitUn(any_food);	="%LogitUn(" & trim(I12) & ");"
any_hardship	%LogitUn(any_hardship);	="%LogitUn(" & trim(I13) & ");"

- In Stata, want logit any_afford surg_any.
- Excel formula: "logit" & trim(I7) & " " & trim(J7)
- Copy and paste to .do file.

Computed Variables – Example 1, Percentages



- For logistic regression, suppose we want to compute formulas like:
- $\text{Percent_ALND} = 100 * \exp(\text{Estimate}) / (1 + \exp(\text{Estimate}))$;
- $\text{Percent_ALND_LCL} = 100 * \exp(\text{LCL}) / (1 + \exp(\text{LCL}))$;
- $\text{Percent_ALND_UCL} = 100 * \exp(\text{UCL}) / (1 + \exp(\text{UCL}))$;
- Where Estimate is a linear combination of logistic regression coefficients.



Computed Variables – Example 1, Percentages

Column A	Column B	SAS Code	Excel Code
EstimateFL7	PercentReOp7	<pre>PercentReOp7 = 100*exp(EstimateFL7)/ (1 + exp(EstimateFL7));</pre>	<pre>=B2&"= 100*exp("&A2&")/ (1 + exp("&A2&"));"</pre>
Estimate_LCL7	PercentReOp_LCL7	<pre>PercentReOp_LCL7 = 100*exp(Estimate_LCL7)/ (1 + exp(Estimate_LCL7));</pre>	<pre>=B3&"= 100*exp("&A3&")/(1 + exp("&A3&"));"</pre>
Estimate_UCL7	PercentReOp_UC7	<pre>PercentReOp_UCL7 = 100*exp(Estimate_UCL7)/ (1 + exp(Estimate_UCL7));</pre>	<pre>=B4&"= 100*exp("&A4&")/(1 + exp("&A4&"));"</pre>



Computed Variables – Example 2, Remove 99's

Variable	SAS Syntax	Excel
Melanoma	<pre>if Melanoma = 99 then Melanoma_No99=.; else Melanoma_No99 = Melanoma;</pre>	<pre>= "if "&A2&" = 99 then &A2&"_No99=.; else &A2&"_No99 = &A2&" ;"</pre>
Sarcoma	<pre>if Sarcoma = 99 then Sarcoma_No99=.; else Sarcoma_No99 = Sarcoma;</pre>	<pre>= "if "&A3&" = 99 then &A3&"_No99=.; else &A3&"_No99 = &A3&" ;"</pre>



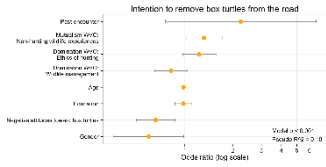
Estimate Statements for Time Series

To minimize typos, let Excel compute coefficients for estimate statements.

(Year-2006) Exponent	1	2	3	SAS Estimate Statement
2014	8	64	512	Estimate '2014-2006' yearover2006 8 yearover2006_2 64 yearover2006_3 512;
2015	9	81	729	Estimate '2015-2006' yearover2006 9 yearover2006_2 81 yearover2006_3 729;
2016	10	100	1000	Estimate '2016-2006' yearover2006 10 yearover2006_2 100 yearover2006_3 1000;

SAS Estimate Statements for Time Series Via Excel Formulas

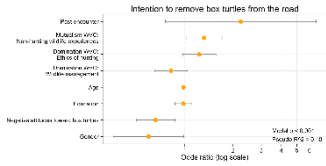
SAS Estimate Statement	Excel Formula
Estimate '2014-2006' yearsover2006 8 yearsover2006_2 64 yearsover2006_3 512;	="Estimate "&A2&"-2006"&" yearsover2006 "&B2&" yearsover2006_2 "&C2&" yearsover2006_3 "&D2&";"
Estimate '2015-2006' yearsover2006 9 yearsover2006_2 81 yearsover2006_3 729;	="Estimate "&A3&"-2006"&" yearsover2006 "&B3&" yearsover2006_2 "&C3&" yearsover2006_3 "&D3&";"
Estimate '2016-2006' yearsover2006 10 yearsover2006_2 100 yearsover2006_3 1000;	="Estimate "&A4&"-2006"&" yearsover2006 "&B4&" yearsover2006_2 "&C4&" yearsover2006_3 "&D4&";"



Odds Ratio Plot, 1 of 7

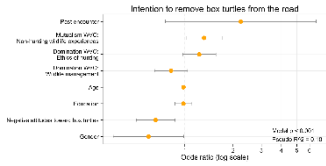
- Colleague gave me this Stata output from logistic regression.
- Need to convert entries like 2.65 (2.00-3.52) to three separate columns with entries OR 2.65, OR_LCL 2.00, and OR_UCL 3.52.
- Use Excel string functions MID() and Find().

	<u>Income</u>		
	<50% FPL	50-149% FPL	150-399% FPL
General Financial Stress	2.65 (2.00-3.52)	4.54 (3.81-5.40)	2.05 (1.86-2.27)
normal bills	3.30 (2.72-4.01)	3.81 (3.32-4.38)	2.46 (2.22-2.73)



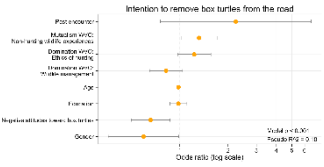
Odds Ratio Plot, 2 of 7, How to Process String

	<50% FPL
General Financial Stress	2.65 (2.00-3.52)
2.65	=MID(B7,1,FIND(" ",B7,1)-1). Get the OR by finding the first space and taking the substring up to the space.
2.00	=MID(B7,FIND("(" ,B7,1)+1,4). Extract the LCL by finding (and selecting 4 characters.
3.52	=MID(B7,FIND("-" ,B7,1)+1,FIND(")",B7,1)-FIND("-" ,B7,1)-1). Obtain the UCL by finding) and the number of characters between) and -.



Odds Ratio Plot, 3 of 7, Delimited Data

Ord	Outcome Label	OR	OR_LCL	OR_UCL
1	# General Financial Stress ="# " & TRIM(G7)	# 2.65 ="# " & TRIM(H7)	# 2.00	# 3.52
2	# Normal bills ="# " & TRIM(G8)	# 3.30 ="# " & TRIM(H8)	# 2.72	# 4.01
3	# Housing ="# " & TRIM(G9)	# 2.88 ="# " & TRIM(H9)	# 2.34	# 3.53

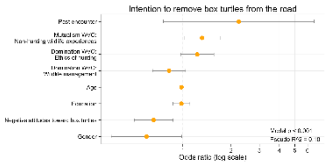


Odds Ratio Plot, 4 of 7, SAS Delimited Data

```

Data LogitReg_FPL;
infile datalines dlm='#';
Length Outcome $32;
Format Ord VarTitleF. Outcome $char32. FPL1_OR FPL1_LCL FPL1_UCL FPL2_OR FPL2_LCL FPL2_UCL
FPL3_OR FPL3_LCL FPL3_UCL 7.2;
Input Ord Outcome FPL1_OR FPL1_LCL FPL1_UCL FPL2_OR FPL2_LCL FPL2_UCL FPL3_OR FPL3_LCL
FPL3_UCL;
Datalines;
1 # General Financial Stress # 2.65 # 2.00 # 3.52 # 4.54 # 3.81 # 5.40 # 2.05 # 1.86
# 2.27
2 # Normal Bills # 3.30 # 2.72 # 4.01 # 3.81 # 3.32 # 4.38 # 2.46 # 2.22 # 2.73
3 # Housing # 2.88 # 2.34 # 3.53 # 3.22 # 2.79 # 3.70 # 2.23 # 1.99 # 2.50
4 # Credit Cards # 3.52 # 2.60 # 4.77 # 3.50 # 2.90 # 4.22 # 2.23 # 1.93 # 2.57
5 # Medical # 1.88 # 1.57 # 2.25 # 2.32 # 2.04 # 2.64 # 1.91 # 1.75 # 2.08
6 # Food* # 11.3 # 9.44 # 13.58 # 13.66 # 11.92 # 15.64 # 3.99 # 3.54 # 4.50
7 # Affordability # 1.56 # 1.26 # 1.91 # 2.26 # 1.96 # 2.59 # 1.91 # 1.74 # 2.09
8 # Medical Debt # 1.50 # 1.28 # 1.76 # 2.33 # 2.06 # 2.63 # 2.13 # 1.95 # 2.31
;
Run;

```



Odds Ratio Plot, 5 of 7, SAS Data Processing

Data FPL1;

```
set LogitReg_FPL(Keep=Ord Outcome FPL1_OR FPL1_LCL FPL1_UCL);
Rename FPL1_OR=FPL_OR FPL1_LCL=FPL_LCL FPL1_UCL=FPL_UCL;
FPL_JWS=1; Run;
```

Data FPL2;

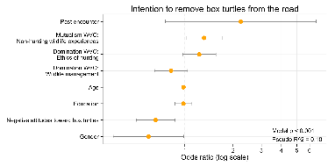
```
set LogitReg_FPL(Keep=Ord Outcome FPL2_OR FPL2_LCL FPL2_UCL);
Rename FPL2_OR=FPL_OR FPL2_LCL=FPL_LCL FPL2_UCL=FPL_UCL;
FPL_JWS=2; Run;
```

Data FPL3;

```
set LogitReg_FPL(Keep=Ord Outcome FPL3_OR FPL3_LCL FPL3_UCL);
Rename FPL3_OR=FPL_OR FPL3_LCL=FPL_LCL FPL3_UCL=FPL_UCL;
FPL_JWS=3; Run;
```

Data FPL4;

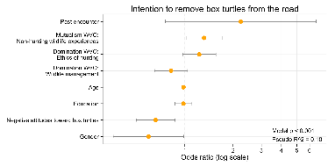
```
set LogitReg_FPL(Keep=Ord Outcome);
FPL_OR=1; FPL_LCL=.; FPL_UCL=.;
FPL_JWS=4; Run;
```



Odds Ratio Plot, 6 of 7, SAS Combine Datasets

```
Data FPL;
set FPL1 FPL2 FPL3 FPL4;
attrib FPL_JWS label="Income/FPL Ratio"
format=FPLJWSF.;
Run;
```

```
Proc Sort Data=FPL;
by Ord FPL_JWS;
Run;
```

Odds Ratio Plot, 7 of 7, SAS Proc SGPlot

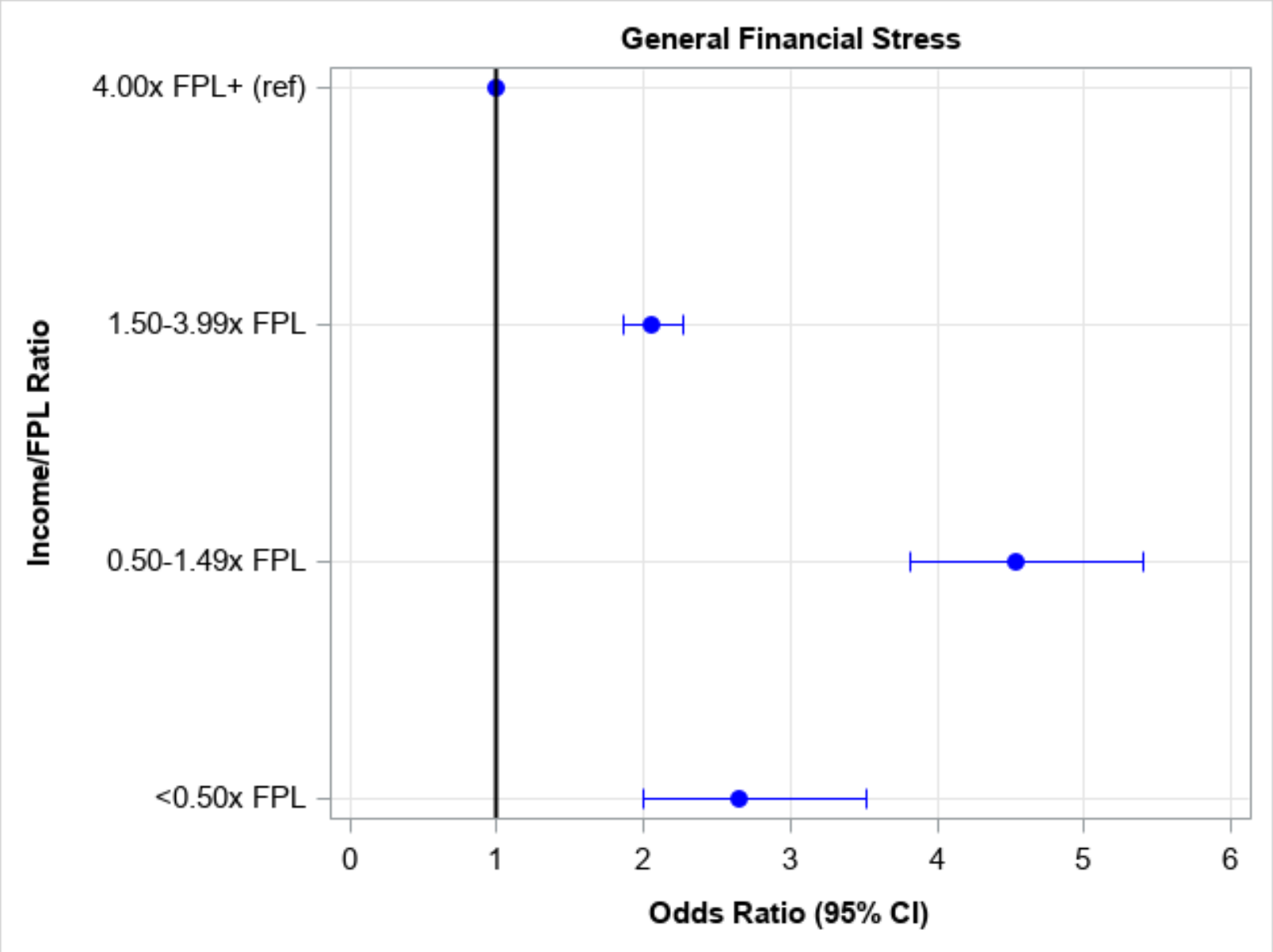
```

Proc SGPlot data=FPL;
Scatter X=FPL_OR Y=FPL_JWS / XerrorLower=FPL_LCL
XErrorUpper=FPL_UCL MarkerAttrs=OR(symbol=CircleFilled Size=10
color=blue)
ERRORBARATTRS=(color=blue);
Refline 1 / Axis=X LINEATTRS=(thickness=2 color=black) ;

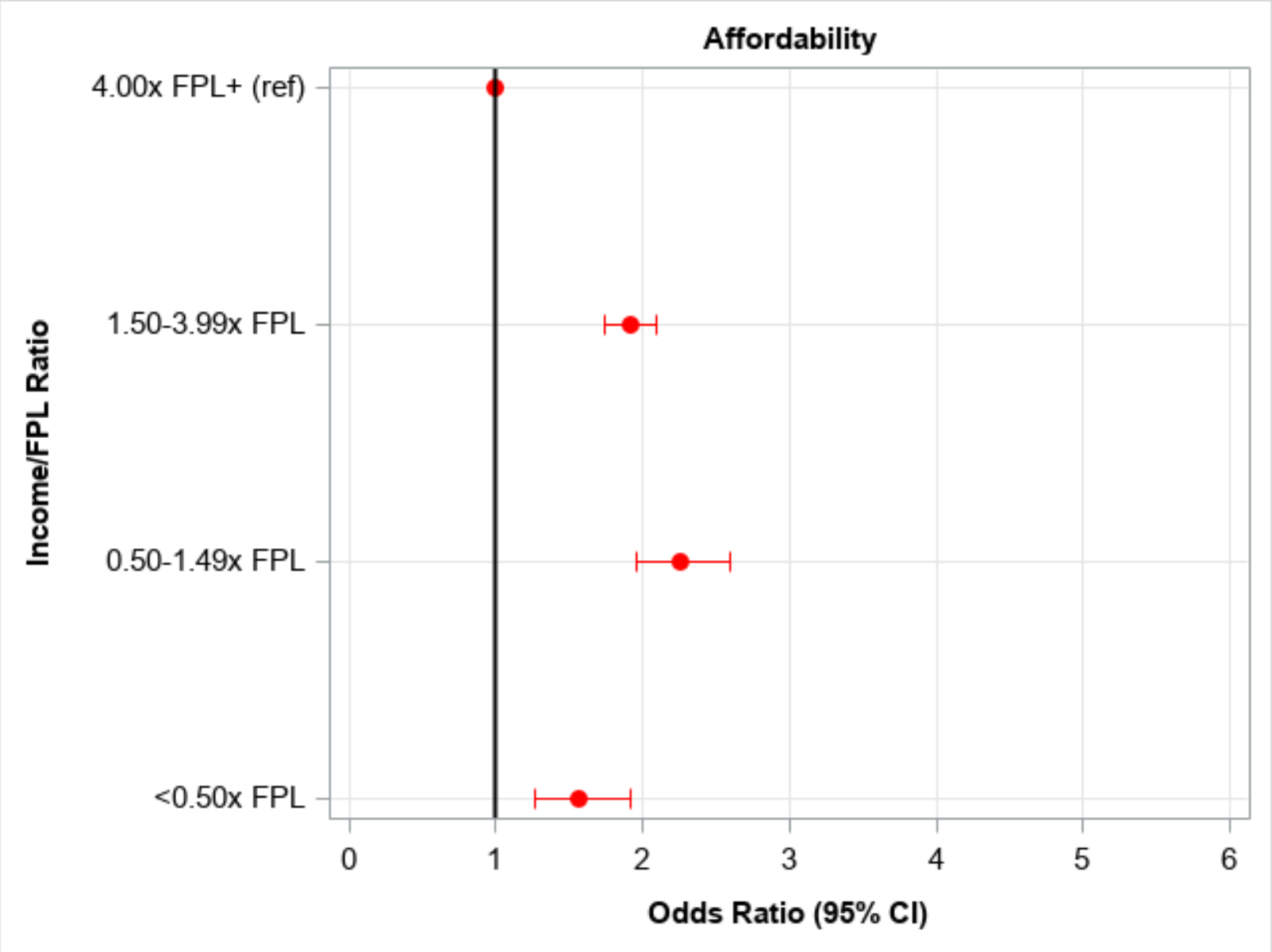
xaxis label='Odds Ratio (95% CI)' LABELATTRS=(Size=11 Weight=Bold)
VALUEATTRS=(Size=11) values=(0 to 6 by 1) grid;
yaxis LABELATTRS=(Size=11 Weight=Bold) VALUEATTRS=(Size=11)
DISCRETEORDER=DATA values=(1 to 4 by 1) grid;
Where Ord=1; /* Use first outcome variable */
Run;

```

Odds Ratio Plot 1



Odds Ratio Plot 2





Translating Stata and SAS Syntax, 1 of 3

- A colleague sent me this Stata syntax to replace missing values.
- `replace InsureCat_JWS = 1 if InsureCat_JWS == . & HIKINDNA == 1`
- `replace InsureCat_JWS = 2 if InsureCat_JWS == . & HIKINDNB == 1`
- `replace InsureCat_JWS = 2 if InsureCat_JWS == . & HIKINDNC == 1`



Translating Stata and SAS Syntax, 2 of 3

- Use Excel string processing functions to extract key parts of Stata code.

=	New Value	Test Var	Test Val
23	1	HIKINDNA	1
=FIND("=",A2,1)	=MID(A2,B2+2,1)	=MID(A84, FIND("&",A2,1)+2,8)	=RIGHT(A2,2)
23	2	HIKINDNB	1
23	2	HIKINDNC	1



Translating Stata and SAS Syntax, 3 of 3

→ Translate to SAS.

```
if (InsureCat_JWS=. And HIKINDNA= 1) then InsureCat_JWS=1;  
="if (InsureCat_JWS=. And "&J2&"="&K2&") then InsureCat_JWS="&I2&";"
```

```
if (InsureCat_JWS=. And HIKINDNB= 1) then InsureCat_JWS=2;  
="if (InsureCat_JWS=. And "&J3&"="&K3&") then InsureCat_JWS="&I3&";"
```

```
if (InsureCat_JWS=. And HIKINDNC= 1) then InsureCat_JWS=2;  
="if (InsureCat_JWS=. And "&J4&"="&K4&") then InsureCat_JWS="&I4&";"
```



Closing Comments

- Key Excel string processing functions: & (concatenate), find(), fixed(), if(), left(), len(), mid(), right(), trim().
- Generate *'s for p-values and syntax for SAS estimate statements.
- Create SAS attribute statements and Stata variable properties.
- Invoke SAS macros and Stata commands.
- Manipulate output from procedures into format for SAS Proc SGPlot or other procedures.
- Convert syntax from Stata → SAS or from SAS → Stata, or to or from R.



Contact Information

Brandy R. Sinco, BS, MA, MS, Statistician Senior

Michigan Medicine

Center for Healthcare Outcomes and Policy (CHOP)

2800 Plymouth Rd.

Ann Arbor, MI 48109-2800

E-Mail: brsinco@med.umich.edu

Phone: 734-232-2252