Graphical Display of Differences in Means and Proportions

Robin High
Department of Biostatistics
University of Nebraska Medical Center
Omaha, NE

University of Nebraska Medical Center

Outline

Preliminaries

- **Coding Data and SGPLOT Statements**
- **Graph Template Language (intro)**

_1_ Display Means (pvalues for differences)
_2_ T-tests: Difference in two mean
_3_ ANOVA: Differences among means
_4_ Polar Coordinates (introduction)
_5_ Display Proportions | Odds Ratios
_6_ Differences in Means with Splines

Data Names and Labels

Short variable names: 1 to 8 characters

avoid:

* SAS stat keywords:

n, t, min, max, mean, median, etc.

* index letters entered in ARRAYs: i, j, k

Add variable labels either in DATA step or PROC DATASETS:

```
LABEL  b = 'b Sex'
       d = 'd Race';
```

Apply text with formats (PROC FORMAT)

Code Categorical Data

Code categorical data with integers (1,2,3…) considering alphabetical or ordinal levels when relevant

Gender:	Outcome Observed?
Female =1	Yes = 1
Male =2	No = 0

Apply text with formats
(add initial letter to specify desired order)

```
PROC FORMAT;
VALUE gnd 1='a Female' 2='b Male';
VALUE ysn 0='b No'     1='a Yes'; *ordered;
RUN;
```

For SAS procedures, enter "order=formatted" either on PROC or CLASS statement option when order matters

Make a Format with LSMEANS

LsMeans Data set contents estimate

lvlx	a	b	estimate
1	0	0	5.2
2	0	1	6.1
3	0	2	6.5
4	1	0	5.7
5	1	1	6.0
6	1	2	7.2

Contents of the format data set

start	fmtname	type	label
1	lv	N	1 a= 0 | b= 0
2	lv	N	2 a= 0 | b= 1
3	lv	N	3 a= 0 | b= 2
4	lv	N	4 a= 1 | b= 0
5	lv	N	5 a= 1 | b= 1
6	lv	N	6 a= 1 | b= 2

lvlx: Plot LsMeans from two discrete factors to show interaction

Make a Format

```
DATA _Lfmt;
SET lsm(rename=(lvlx=start));
KEEP label fmtname type start;
LENGTH label $15;
RETAIN fmtname "lv" type 'N';
label = CAT(put(start,1.0),
            ' a=', put(a,2.0), ' | b=', put(b,2.0));

PROC PRINT DATA=_Lfmt;
ID start ;
VAR fmtname type label;
RUN;

PROC FORMAT cntlin=_Lfmt; RUN;

PROC PRINT DATA=lsm;  * print class data with format;
VAR lvlx estimate;
FORMAT lvlx lv. ;
RUN;
```

Apply Picture Format

```
PROC FORMAT;
PICTURE xx (round)
 .001 -< .01 = " 9.999"
  .01 -< .1  = " 9.99 "
   .1 -<  1  = " 9.9  "
    1 -<  10 = " 9.9  "
      .      = '      ';
run;
PROC PRINT ;
VAR a ab;
FORMAT ab xx. ;
RUN;
```

Obs	a	ab
1	0.00342	0.003
2	0.02630	0.03
3	0.28053	0.3
4	7.23253	7.2
5	.	

"a" has actual value

ab is rounded and printed with indicated number of decimals

Missing value printed as a blank

Format applied to print pvalues

```
PROC FORMAT;
PICTURE pval (round)
     0 -< .001 = "<.001 **" (NOEDIT)
  .001 -< .010 = "9.999 **"
  .010 -< .050 = "9.999 * "
  .050 -< .109 = "9.999   "
  .109 -< 1    = "9.99    "
          1    = "9.9     "
          .    = " ";
RUN;
```

Integer code 1, 2, 3 .. for Categorical (character) Data

```
PROC SORT DATA=dt; BY xx; RUN;

* add variable xxN coded as 1,2 ... ; * avoid 0 except binary;

DATA dt; SET dt;
BY xx;
RETAIN xxN . ; * need to add RETAIN with . in order to avoid 0
                in the numeric code assigned to missing values;
IF missing(xx)=0 AND first.xx THEN xxN+1;
run;

* To make a format begin with a file of character names;

ods output onewayfreqs=_onew(keep= xx );
ods listing close;
proc freq data=dt;
table xx / missprint;
run;
ods listing;

* Can make a format with code on previous slide;
```

Additional Numeric Variable

Obs	sexN	Sex
1	.	
2	.	
3	1	F
4	1	F
5	1	F
6	1	F
7	1	F
8	2	M
9	2	M
10	2	M
11	2	M
12	2	M

SGPLOT: XAXIS | YAXIS
Add Line Breaks in Long Variable Labels

Rather than enter this text through a format:

```
XAXIS < options >

valuesdisplay=(
 'I understand, physiologic properties, of conversation'
 'I feel comfortable, initiating conversation'
 'I feel confident, to ad lib conversation'
 'I feel confident, to change topics, of conversation')

  fitpolicy=splitalways splitchar="," splitjustify=center;
```

Means/Differences with Confidence Intervals

To distinguish differences with small or large pvalues note that group variable for linetype does not work with SCATTER statement; a solution is to assign two SCATTER statements with different variables:

Small pvalues (solid line):

```
SCATTER y=lvl  x=estimate / xerrorlower=lower xerrorupper=upper
                           errorbarattrs=(pattern=1 thickness=solid)
                           markerattrs=(symbol=circlefilled size=6)
                           datalabel=estimate datalabelattrs=(weight=bold size=7);
```

Large pvalues (medium dash line):

```
SCATTER y=lvl  x=estimate2 / xerrorlower=lower2 xerrorupper=upper2
                             errorbarattrs=(pattern=mediumdash thickness=1)
                             markerattrs=(color=black symbol= circlefilled size=6)
                             datalabel=estimate2 datalabelattrs=(weight=bold size=7);
```

In some situations can produce different linetpes with HIGHLOW statement:

```
STYLEATTRS datalinepatterns=(solid mediumdash);
HIGHLOW y=lvl low=lowercl high=uppercl / group=sgnf type=line lineattrs=(thickness=1)
                                         lowcap=serif highcap=serif ;
SCATTER y=lvl x=estimate / markerattrs=(symbol=circlefilled size=5 ) datalabel=estimate
                           datalabelpos=top datalabelattrs=(weight=bold size=6);
```

Attribute Maps
STYLE ATTRIBUTE STATEMENT

Attribute maps usually applied with graphs for data summaries, can also be necessary to correctly display attributes of means and differences

STYLEATTRS:

Style attributes, the order assigned is determined by which attribute appears first in the data file

```
STYLEATTRS datalinepatterns=(dash solid);
HIGHLOW y=variable low=lowerCL high=upperCL /
   group=sgnf type=line
    lineattrs=(color=black thickness=1)
    lowcap=serif highcap=serif ;
```

If the first comparison in row 1 has sgnf = 1, a dashed line assigned (when a solid line is desired)

XAXISTABLE | YAXISTABLE

Add summary statistics (pvalue, mean, oddsratio, percents, etc.) in top, bottom, left, or right margins of graph

COLUMN of pvalues on far right:

```
YAXISTABLE pvalue / title="pvalue" nolabel
        titleattrs=(color=red weight=bold size=6)
        STAT=sum location=inside position=right
        VALUEATTRS=(color=red weight=bold size=6)
        VALUEHALIGN=left
        PAD=(left=20 px);
```

XAXISTABLE | YAXISTABLE

Add summary statistics (mean or sum) for numeric data in top, bottom, left, or right margins of graph

For data summary graphs To print sample size, add a numeric variable assigned to 1 in the data set

```
_one=1;
```

for all non-missing values and enter stat=sum in XAXISTABLE or YAXISTABLE

Apply legend values without directly referring to a group= <var>

```
PROC SGPLOT DATA=plt;

SERIES y = y1 x=x / group=time;
SERIES y = y2 x=x / group=time;
SERIES y = y3 x=x / group=time ;

LEGENDITEM type=marker name="a1" / label="Time 1"
                                   markerattrs=(color=green symbol=diamond size=5);
LEGENDITEM type=marker name="a2" / label="Time 2"
                                   markerattrs=(color=red   symbol=square  size=5);
LEGENDITEM type=marker name="a3" / label="Time 3"
                                   markerattrs=(color=blue  symbol=triangle size=5);

KEYLEGEND "a1" "a2" "a3" / noborder title="Time Period"
                           titleattrs=(weight=bold size=8)
                           valueattrs=(weight=bold size=7)
                           across=1 down=3 location=inside position=topleft;

RUN;

NOTE: this approach can apply line type and marker symbol together
      in the printed legend
```

PVALUES: connected with differences

probF pvalues from model fixed effects table:

Main effects
Interactions

Individual pvalues for pairwise comparisons or contrasts of interest

Analysis procedure statements produce differences associated pvalues (adjusted if necessary)

LSMEANS: fixed effect with diffs or oddsratio
LSMESTIMATE: write specific contrasts
with non-positional parameters

Graph Template Resource

Plotting lsmeans and differences

www.lexjansen.com

Midwest SUGI conference, 2015

Paper PH-06-2015

Plotting LSMEANS and Differences in Generalized Linear Models with GTL

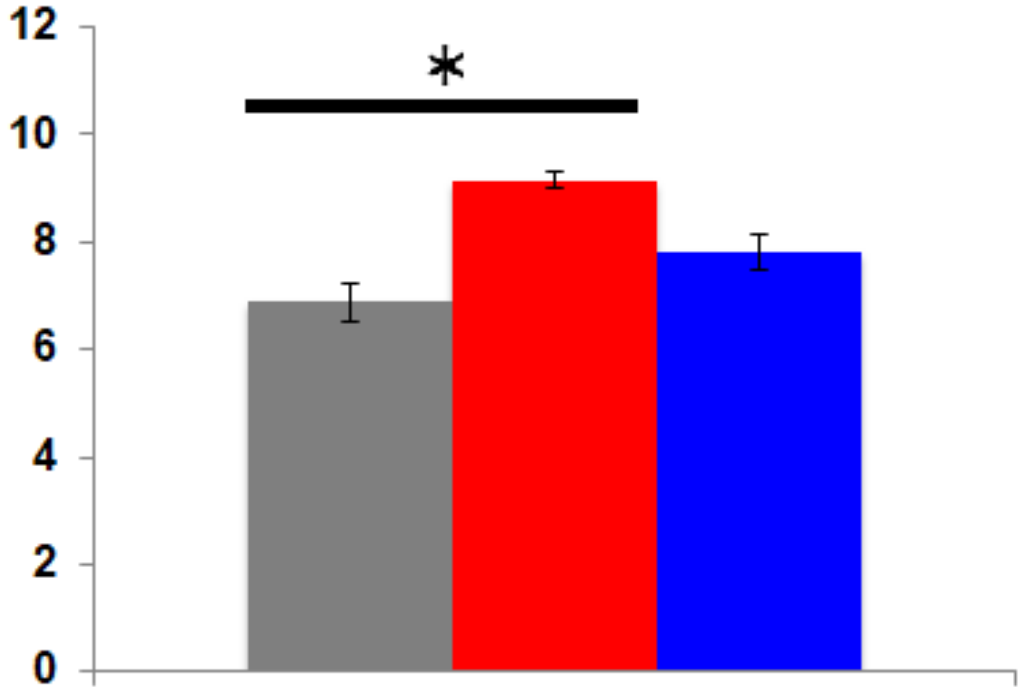
GRAPH TEMPLATE OUTLINE

```
PROC TEMPLATE;
DEFINE STATGRAPH anvmns;
BeginGraph;
DYNAMIC _x _xlbl _nlvls ; * others added to replace default values;

LAYOUT lattice / rows=2 rowweights=(.6 .4) /*row weights must add to 1*/
                 columns=1;

LAYOUT … ;
< specific CODE >
ENDLAYOUT;

LAYOUT … ;
< specific CODE >
ENDLAYOUT;

ENDGRAPH;
run;

Can generate template CODE for each layout code portion with SGPLOT:

PROC SGPLOT tmplout=anvmns.sas;
< enter code >
Run;
```

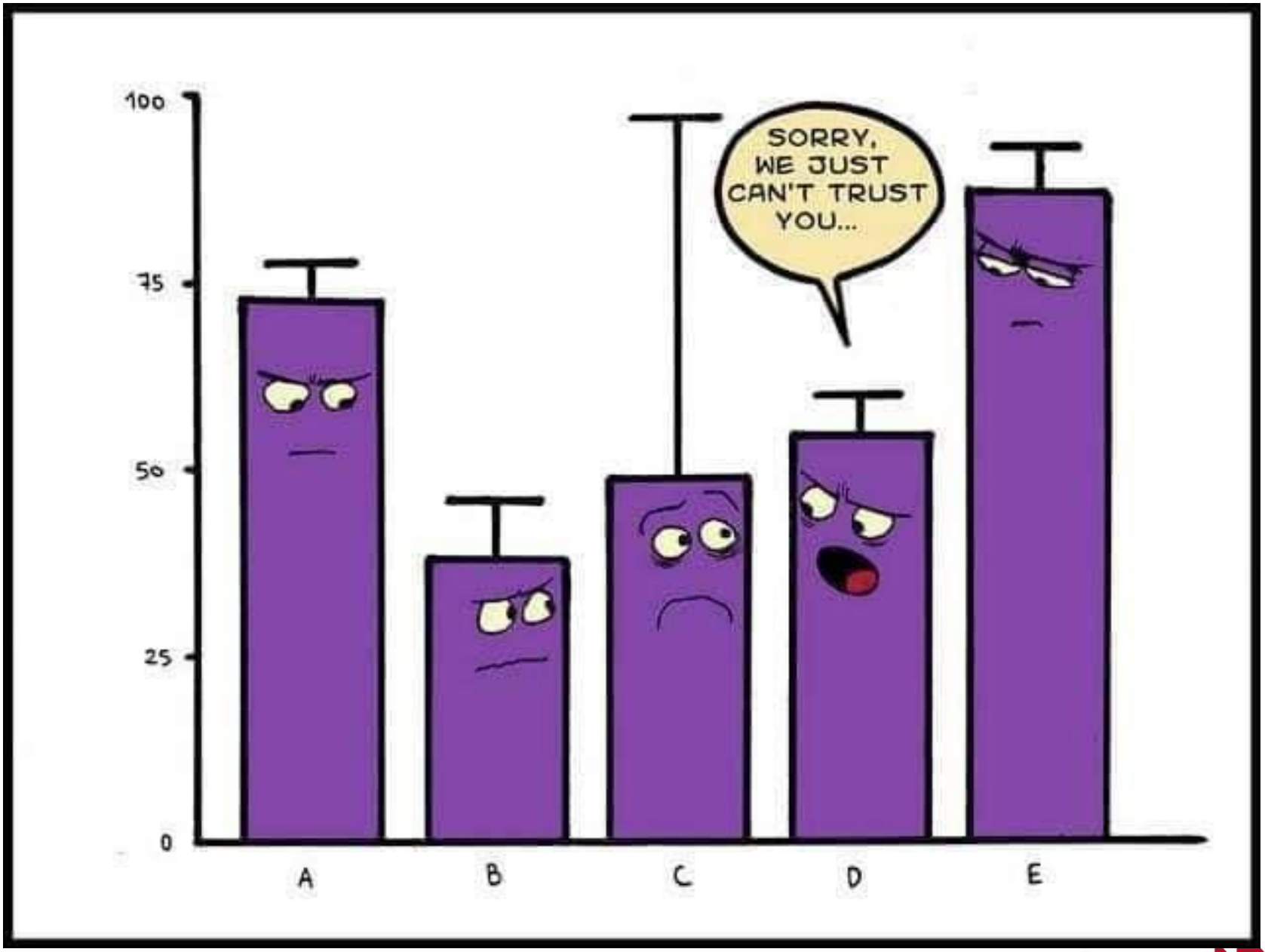
GRAPH TEMPLATE CODE

```
DYNAMIC variables:
  Template variables begin with underscore
  Values/Variable Names to send from data set
       placed within " "

PROC SGRENDER DATA=lsmdf TEMPLATE=anvmns;
DYNAMIC      _x = "lvl"
          _xlbl ="Factor Levels"
         _nlvls = "6" ;
FORMAT lvl lv. ;
RUN;
```

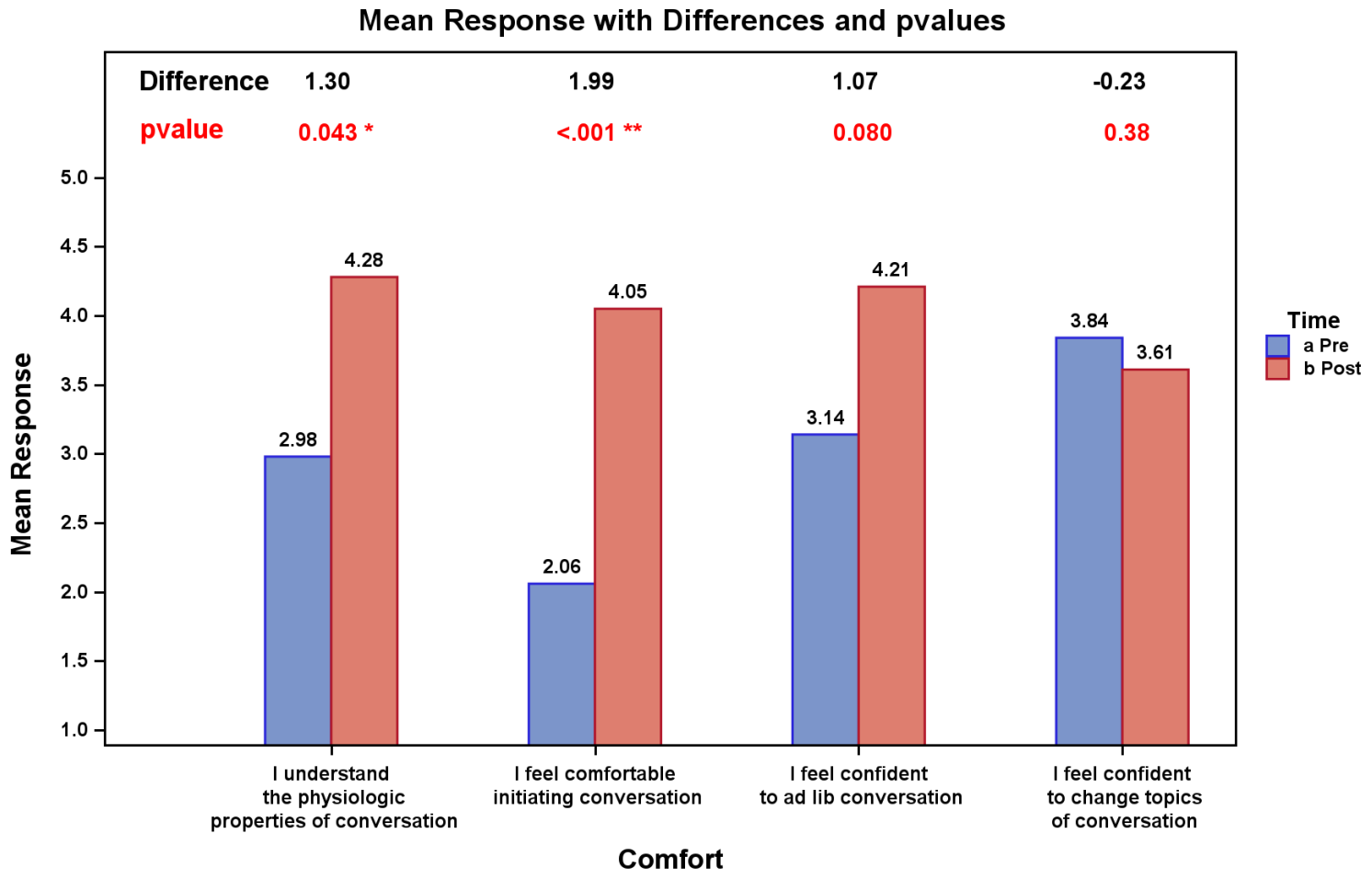
_1_ Bar Graphs



* Significant at 0.05 level

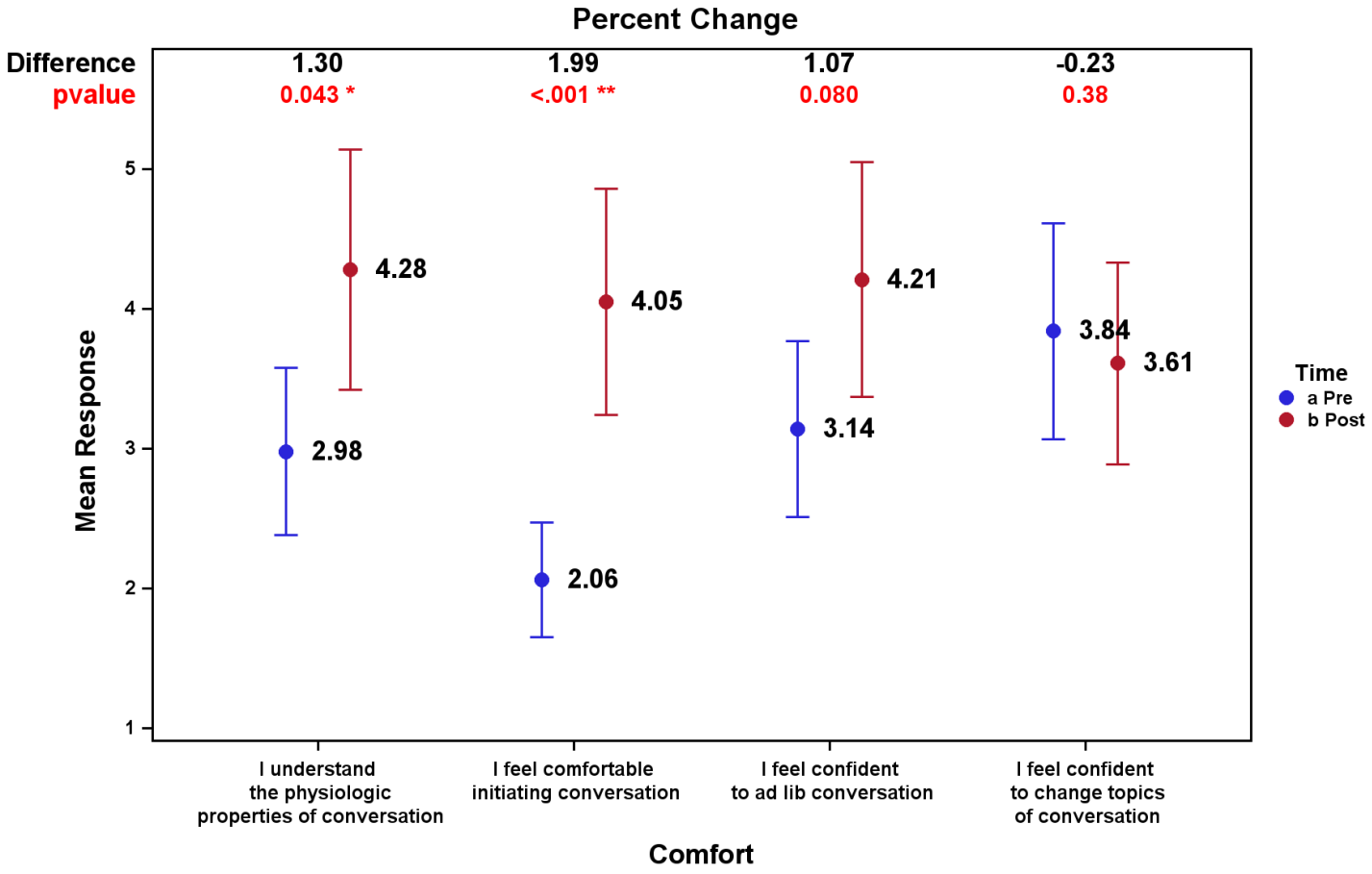

100
75
50
25
0
A
B
C
D
E
SORRY, WE JUST CAN'T TRUST YOU...

Vertical Bar Graph with Differences and Pvalues



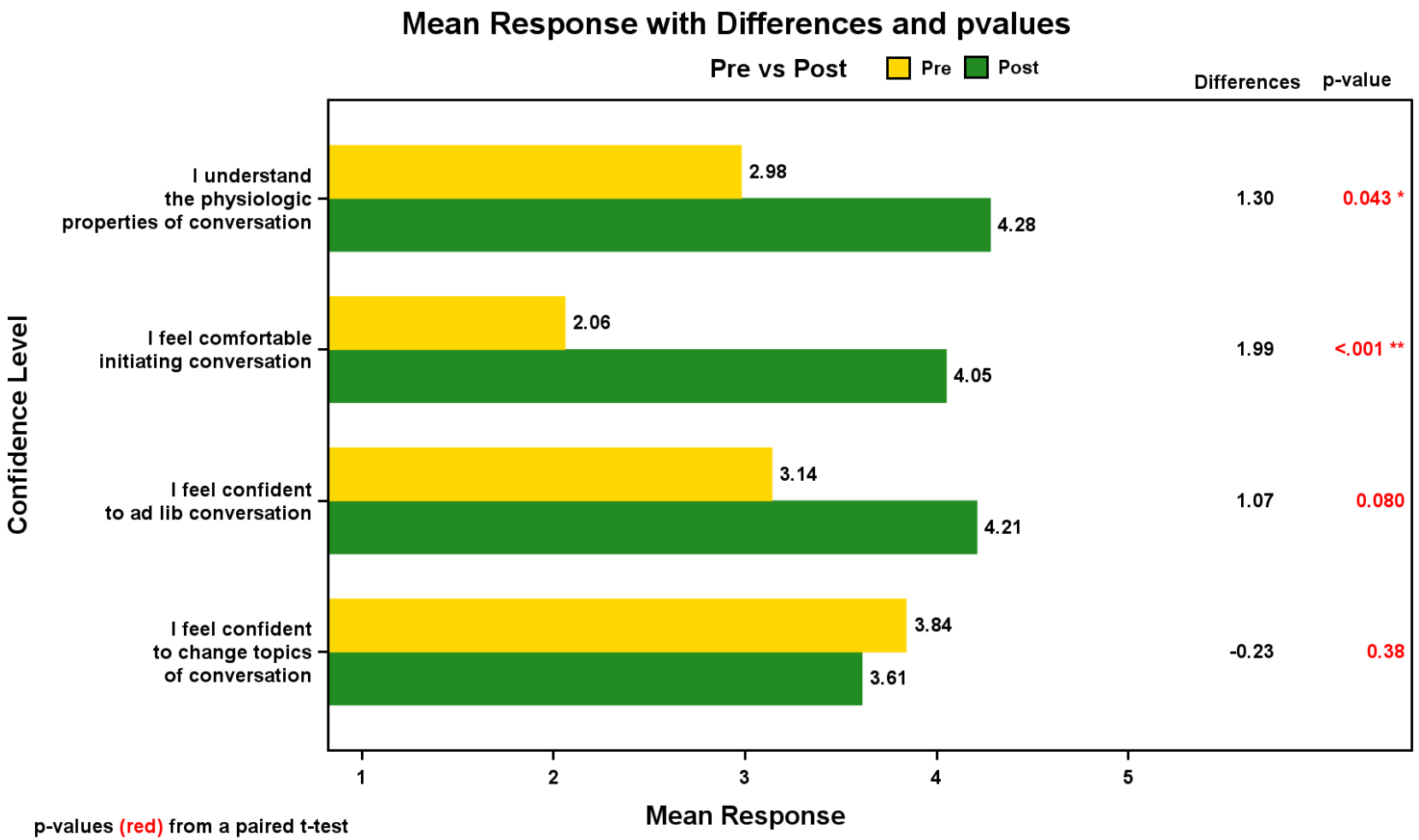
p-values (red) from a paired t-test

Vertical lines showing confidence limits with differences and pvalues

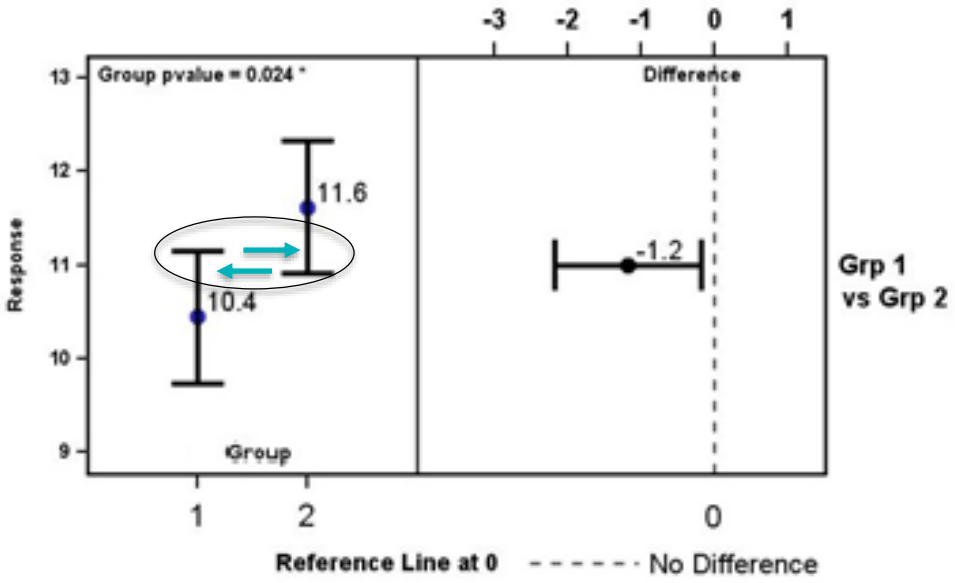


p-values (red) from a paired t-test

Horizontal Bar Graph with Differences and Pvalues



_2_ T-Test
Plots of Means: can be deceptive



Graph displays two means with overlapping 95% confidence intervals, yet t-test reveals a 95% confidence interval for the difference of interest (p=0.024)

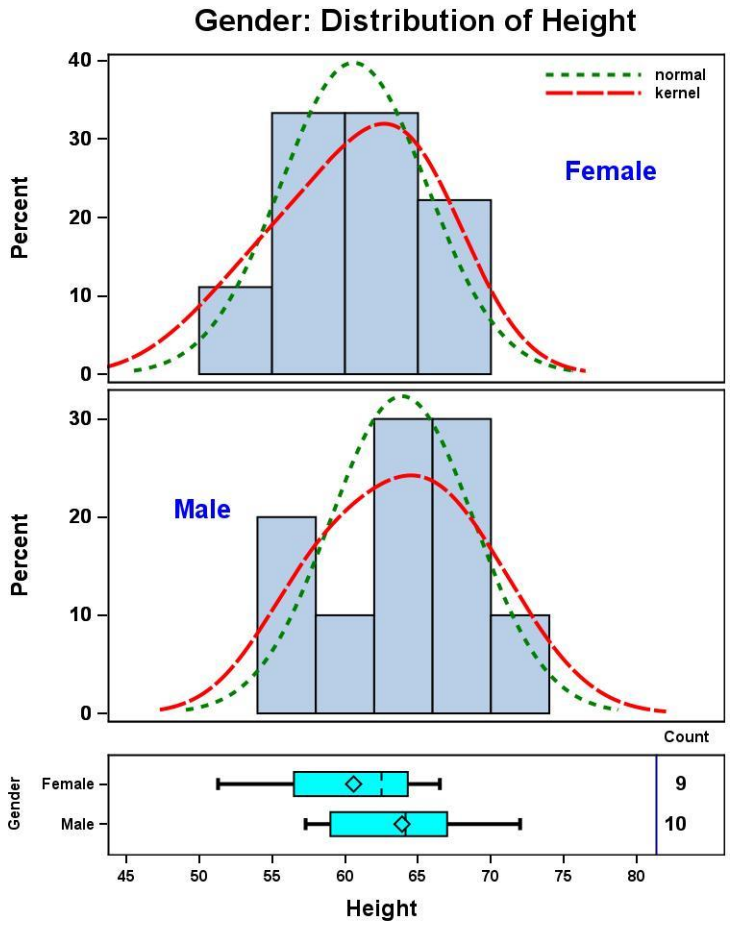
Means and Differences _2_ T-Test

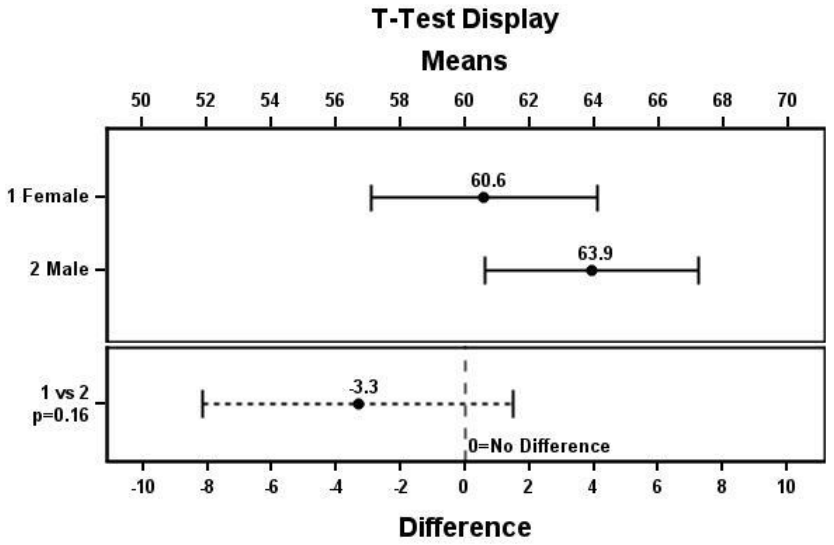
Compute T-Test means and Differences with the MIXED or GLIMMIX (the GLM or TTEST procedures also possible)

Save lsmeans, difference, confidence limits for both, and the pvalue for the difference in ODS data sets

Append the lsmeans and difference file into one data set (block diagonal form), change common variable names in one of them (such as estimate, lower, upper in the difference file)

T-test: Histogram and Boxplots
Advantages of Template to show means and difference





Means (top panel)

Difference (bottom panel)

Distribution:

Histogram and Boxplot

Means and Differences
_3_ ANOVA

Compute means and differences

- MIXED or GLIMMIX procedures
- process illustrated here also works with correlated data models

Adjust confidence intervals for differences and pvalues as needed:

- ANOVA: tukey
- Compare with a control: DUNN
- All others: Simulate

Plot Means and Differences

```
ODS graphics on / reset = all height= 7 in width=7 in ;
ODS LISTING image_dpi=300
            gpath="u:\sas\mixed\plots" ;
ODS GRAPHICS / imagename = "lsmeans_diffs" ;

PROC GLIMMIX DATA=cc176;
CLASS current ;
MODEL wt_d = current | wt_n ;
LSMEANS current / diff cl at means
          adjust=tukey plot=diff(noabs);
FORMAT current cr. ;
RUN;
ODS GRAPH OFF;
```

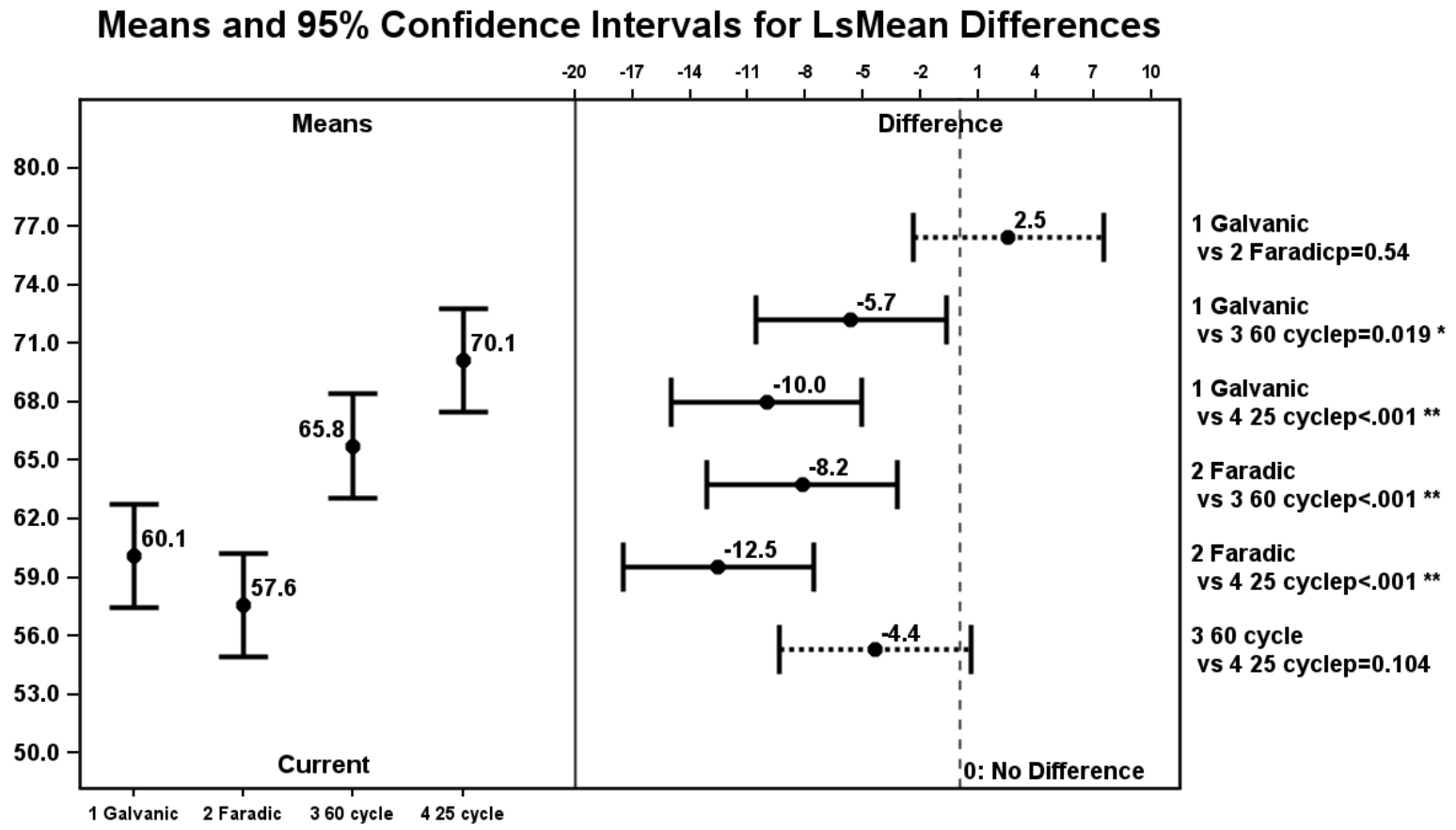
Means and 95% Conf Intervals	Mean	Lower	Upper
Current			
1 Galvanic	60.1	57.5	62.8
2 Faradic	57.6	54.9	60.2
3 60 cycle	65.8	63.1	68.4
4 25 cycle	70.1	67.5	72.8

Differences in Means
Effect=current wt_n=117.8 (stats computed at mean weight)

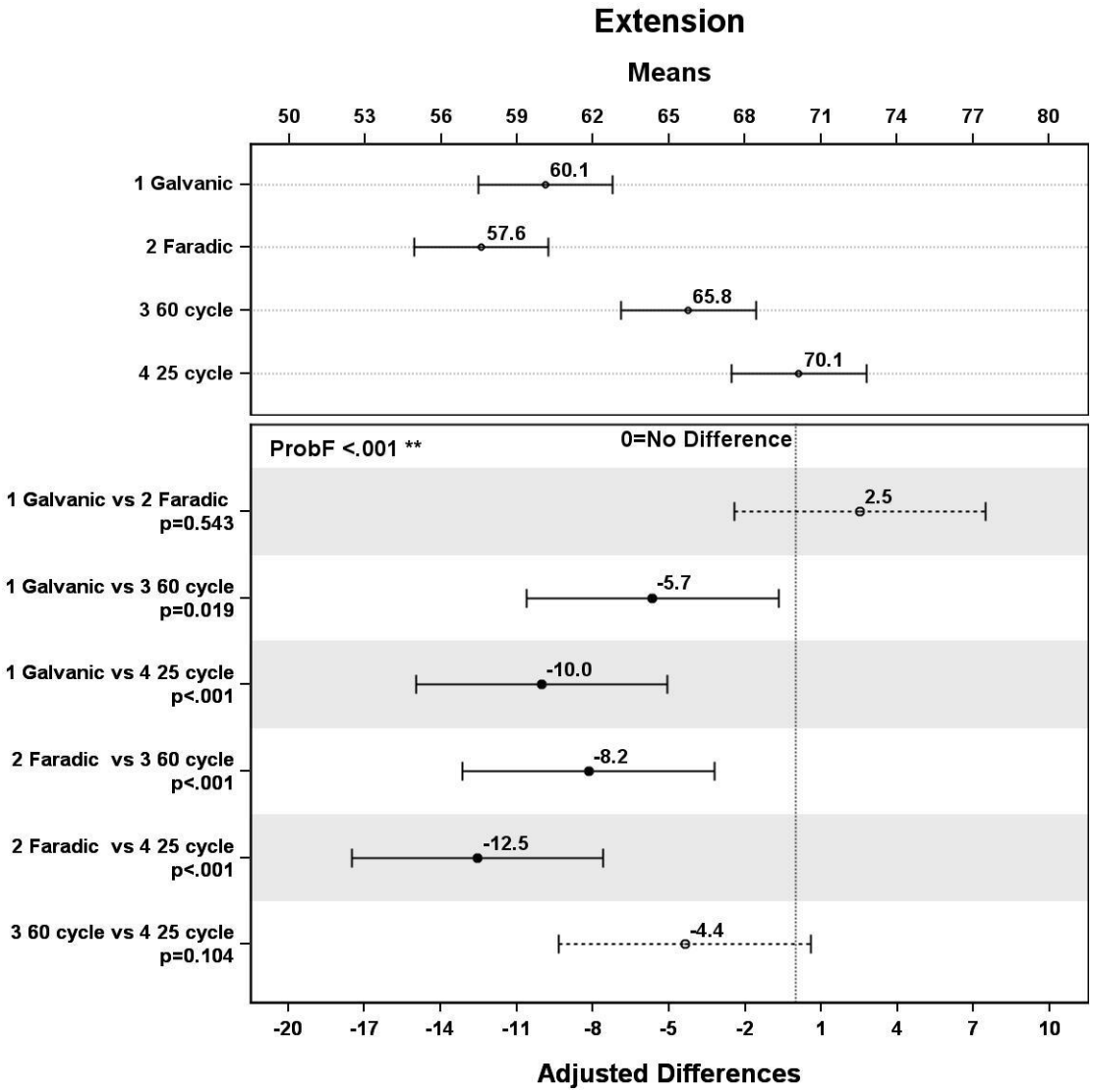
Six pairwise differences in means

current	_current	Estimate	Adj Lower	Adj Upper	Adjp
1 Galvanic	2 Faradic	2.5	-2.4	7.5	0.543
1 Galvanic	3 60 cycle	-5.7	-10.6	-0.7	0.019
1 Galvanic	4 25 cycle	-10.0	-15.0	-5.1	<.001
2 Faradic	3 60 cycle	-8.2	-13.1	-3.2	<.001
2 Faradic	4 25 cycle	-12.5	-17.5	-7.6	<.001
3 60 cycle	4 25 cycle	-4.4	-9.3	0.6	0.104

Means (left panel)
Differences (right panel)



ANOVA: Plot Means and Differences

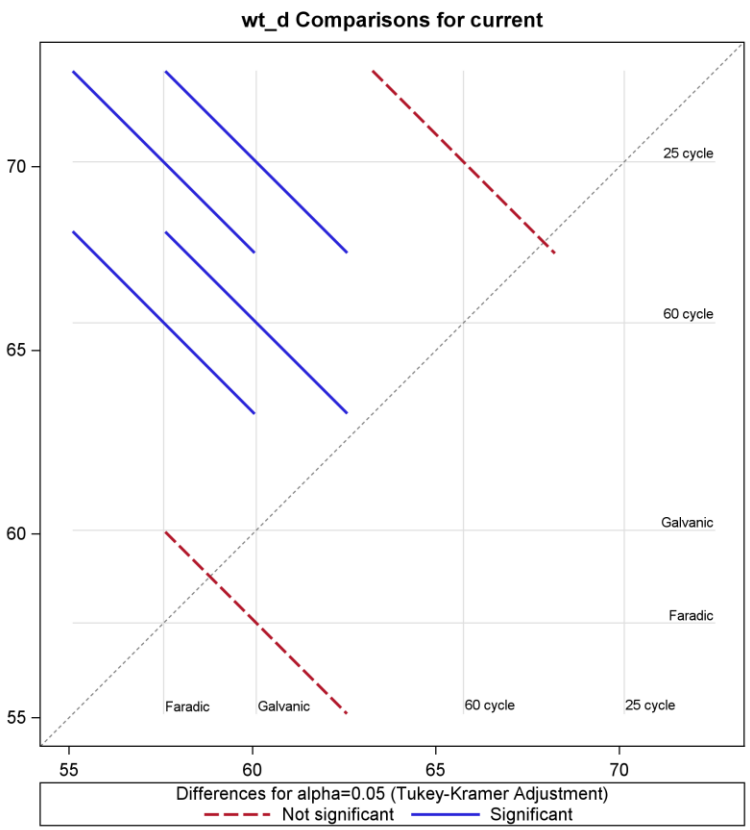


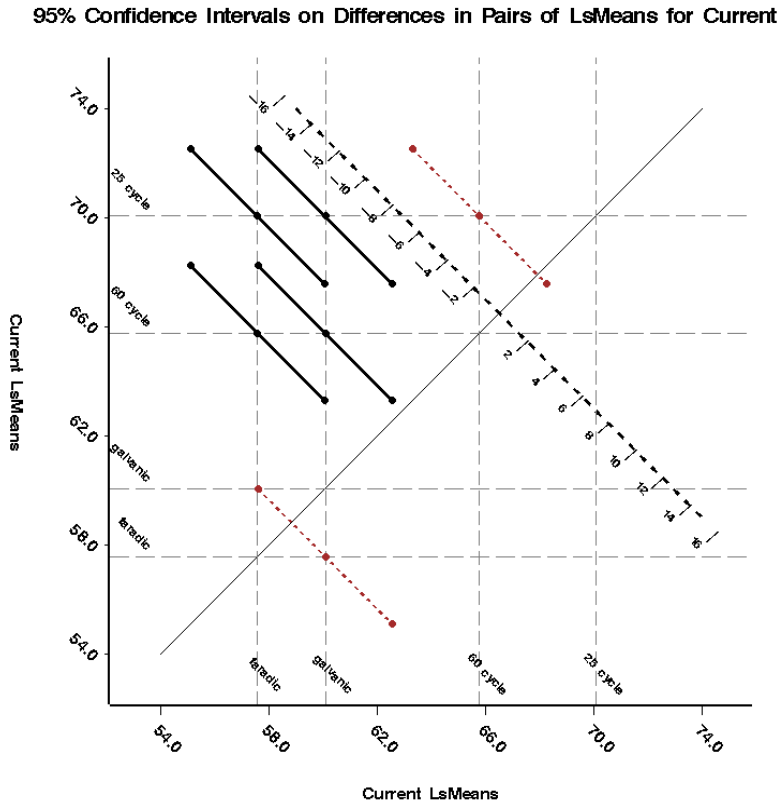
ANOVA template

MACRO variables enter to modify appearance of contents

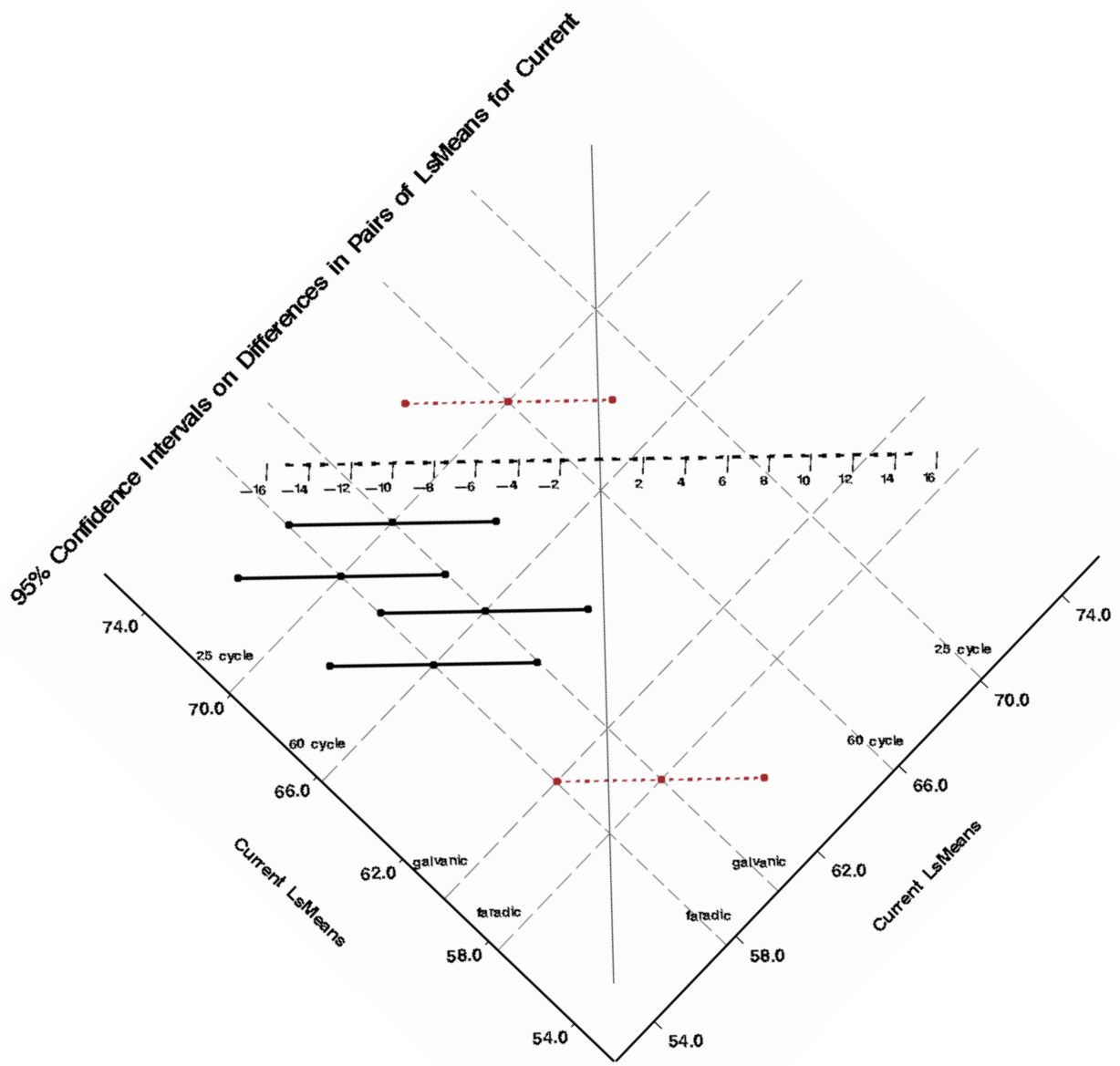
DYNAMIC variables enter within PROC SGRENDER based on specific contents of data file to graph

ANOVA: Mean Mean Scatter Plot

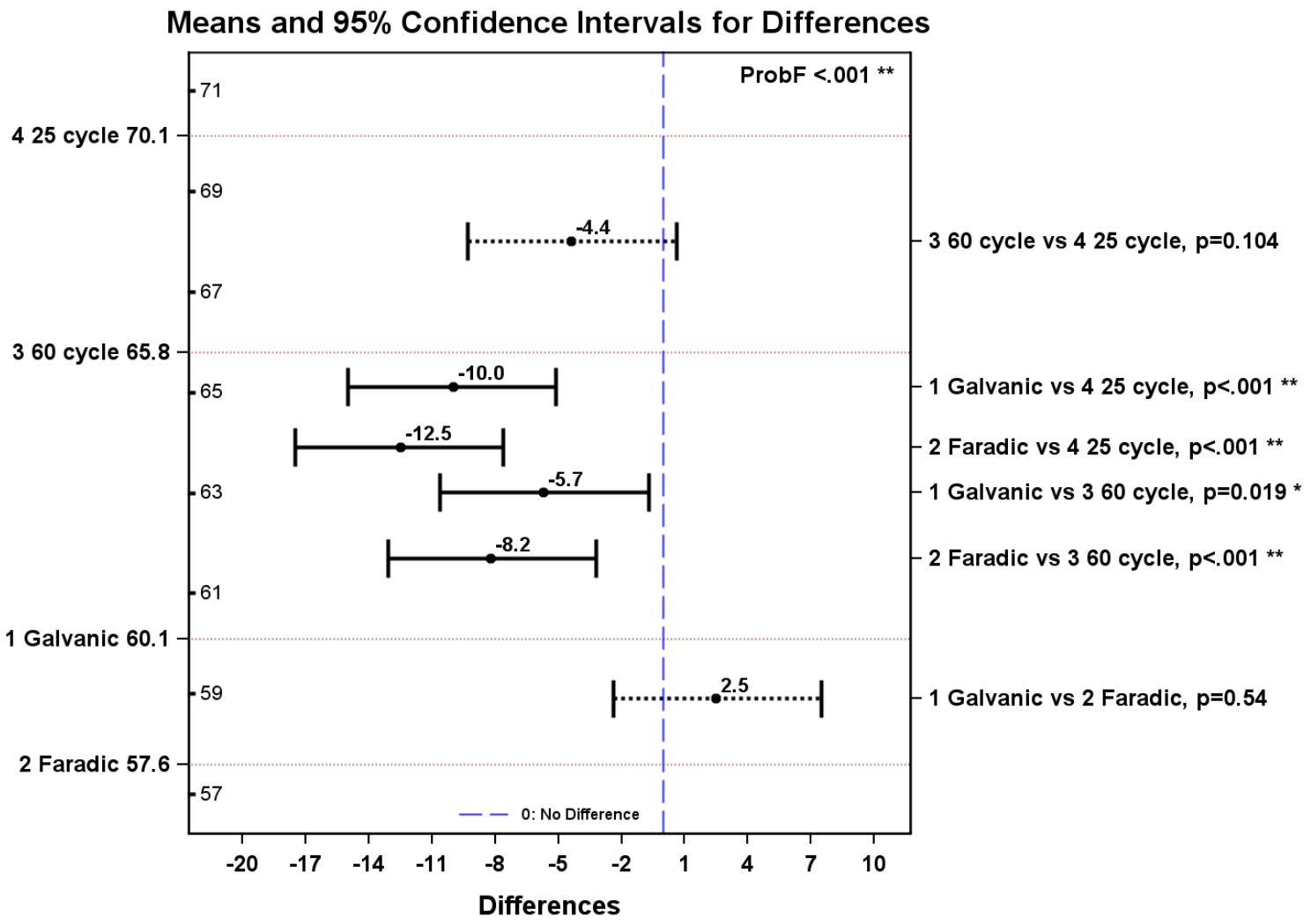




ANOVA: Mean Mean Scatter Plot



ANOVA: Mean Mean Scatter Plot



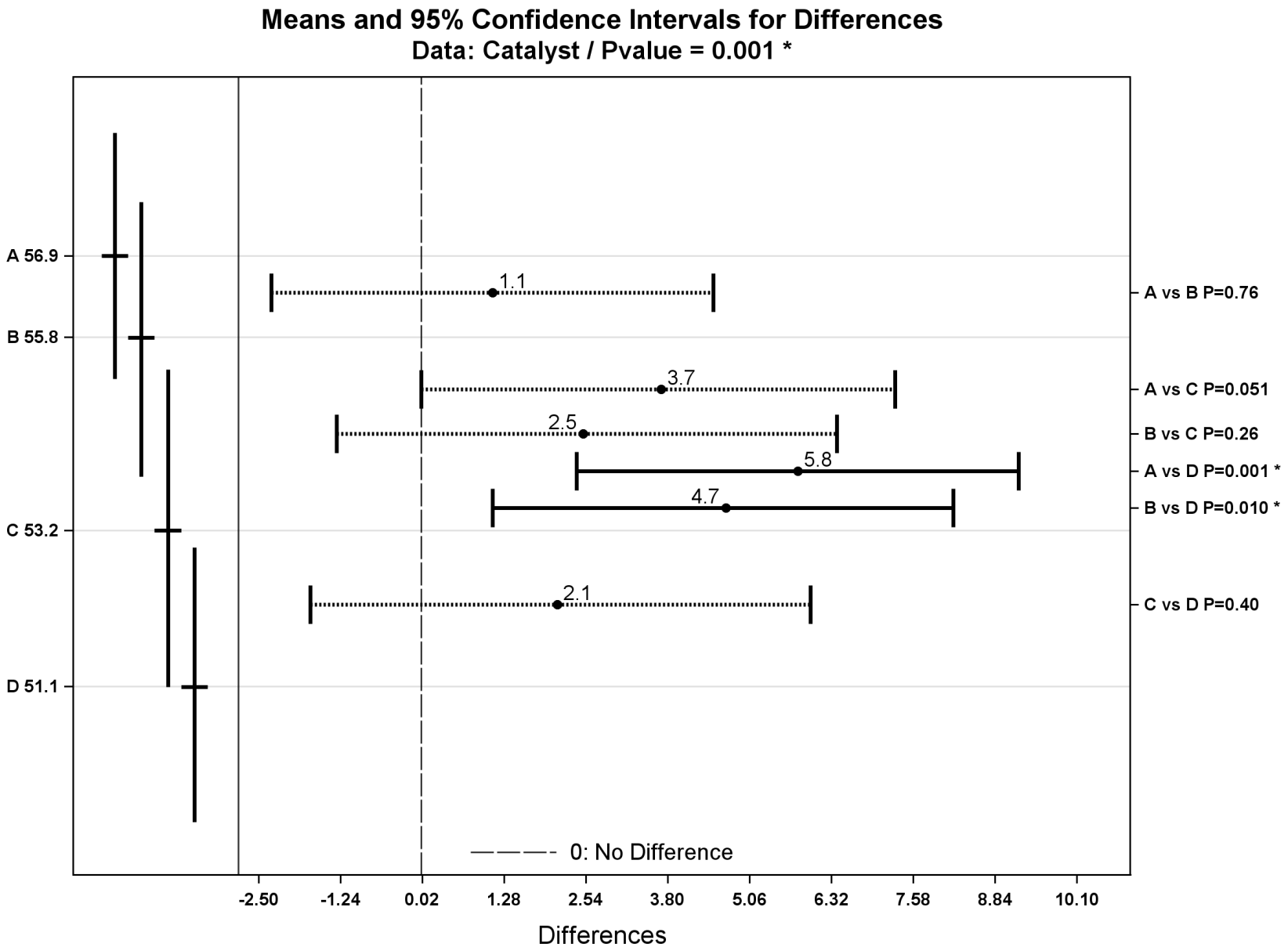
XAXIS | YAXIS
axis ticks at specific values

```
Rather than enter range

values = (57 to 71 by 2)

specify where to add tick marks and print specific value

YAXIS ..

 min=57 max=71
 values= (57.6 60.1 65.8 70.1)
 valueshint
;
```

VALUESHINT: specifies that the minimum and maximum axis values determined independently of the values specified in the values= option (1 to 10 specified as desired range)

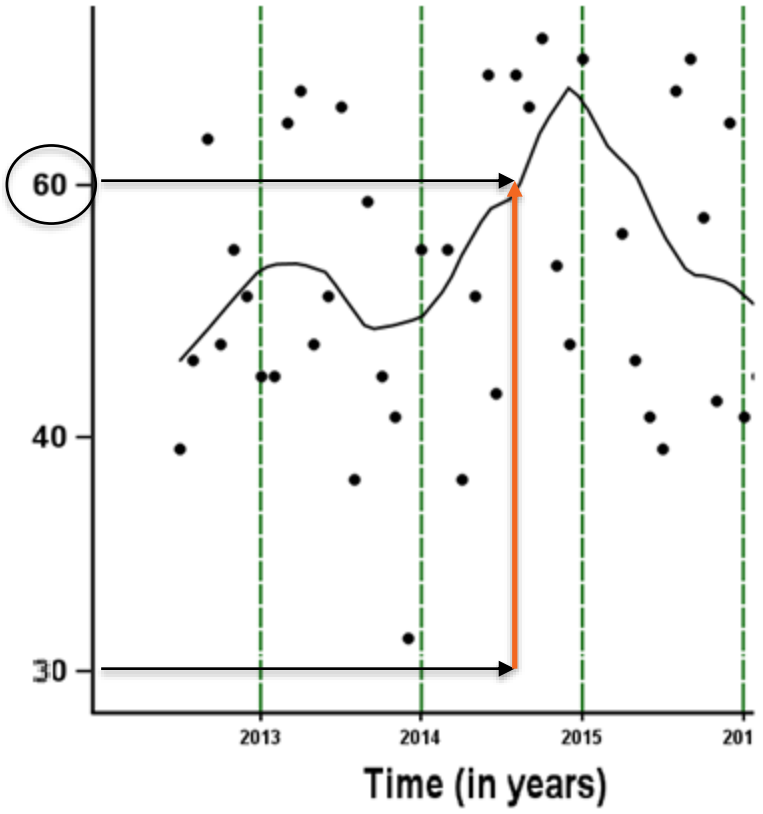
PROC SQL can automate production of a macro variable with desired values to plt (see MMC plot for example)

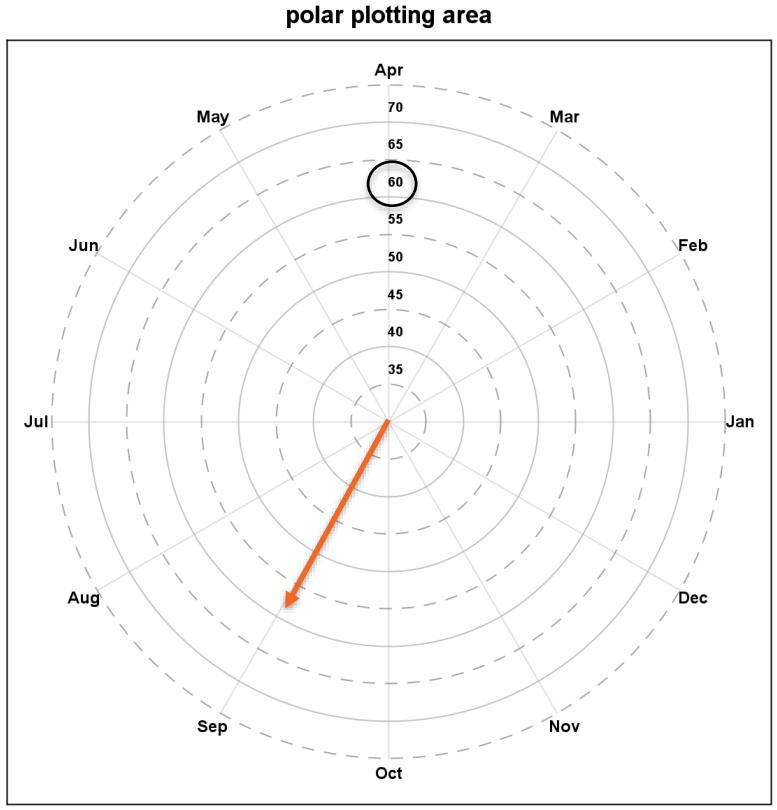
Add vertical bars to show means (different data)



_4_ Polar Coordinates

Horizontal axis variable repeats
(time: each loop around the circle is one year)
Vertical axis value is distance from center (=30)





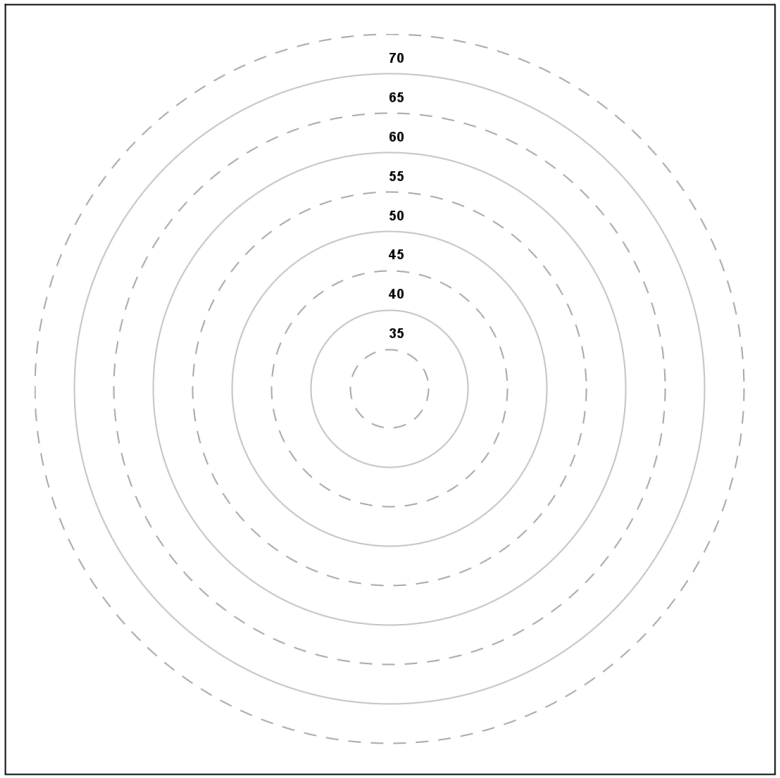
Convert days of the year (1-366) to degrees (1-360)

```
Date = sas date value and response collected over time

dayofyear = put(date,julday3.); * all dates assume day 1 of month;
          * converts date to number from 1 to 366;
day = 360*dayofyear/366 ;
          * convert range of days = 1 - 366 to degrees: 1 - 360;
Convert degrees (day) to radians:
Rads = (day * constant('pi') / 180) ; * angle formula in SAS code;

xCos = cos(rads);
ySin = sin(rads);

Polar coordinates for values (the response measured over time
x = value * xCos;    * lower bound assumed 0;
y = value * ySin;

value = sqrt(x**2 + y**2); * distance from origin;

< verify "value" is length of hypotenuse of a right triangle >
```

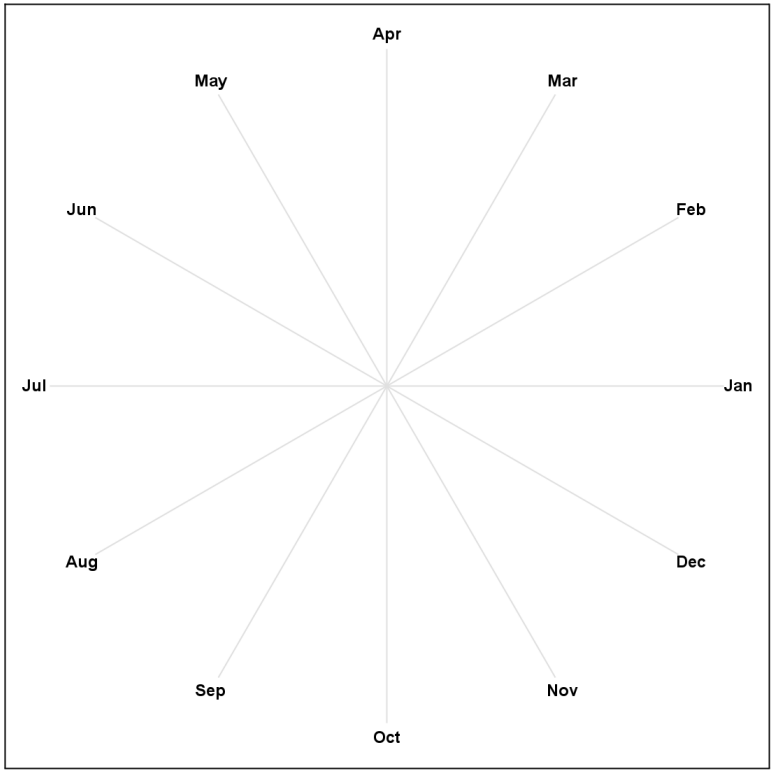
_4_ Polar Coordinates

Based on min and max values of the data, construct data sets to produce two background grids for the plot

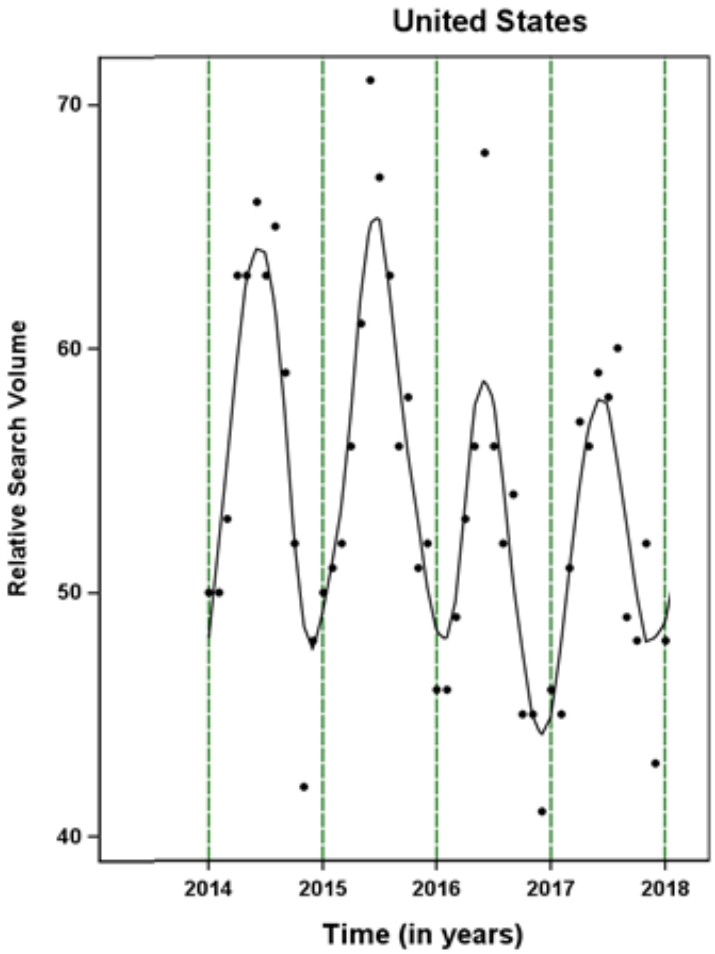
rgrid: circles and values (corresponds to vertical axis)

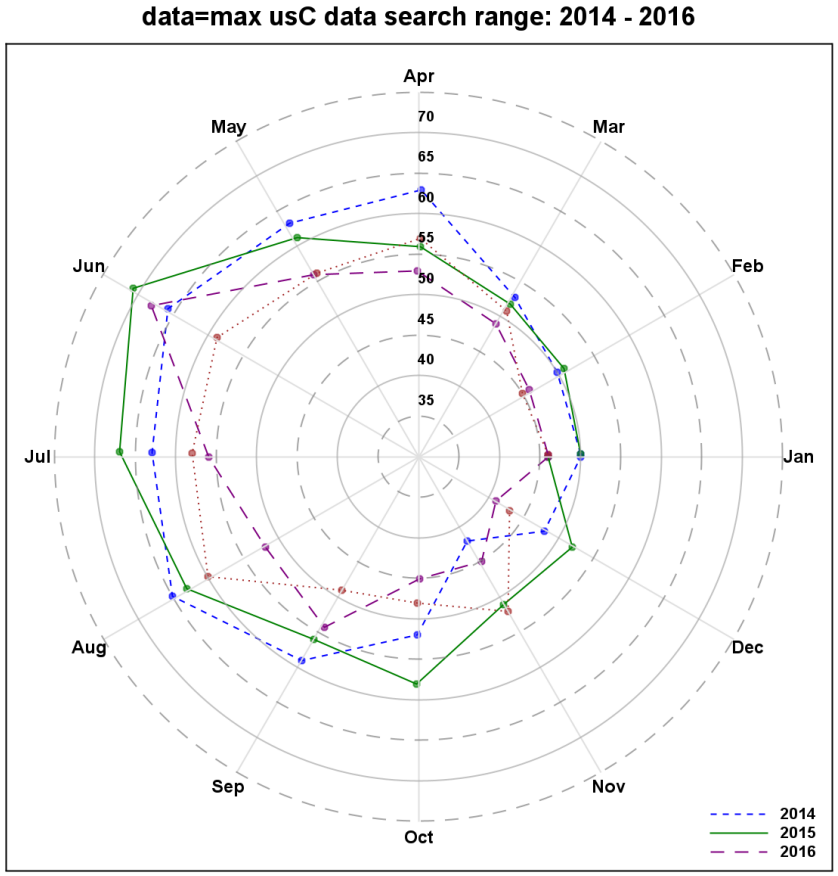


GRID: spokes and labels (corresponds to horizontal axis)

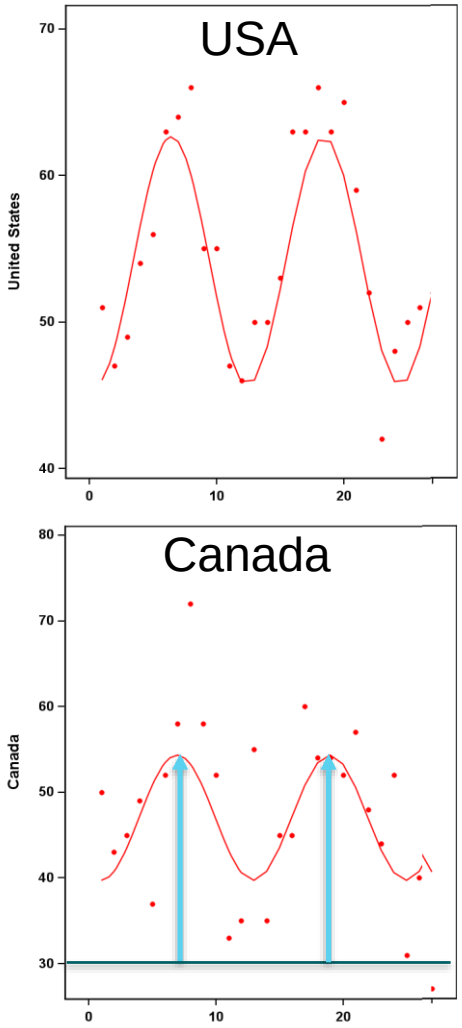


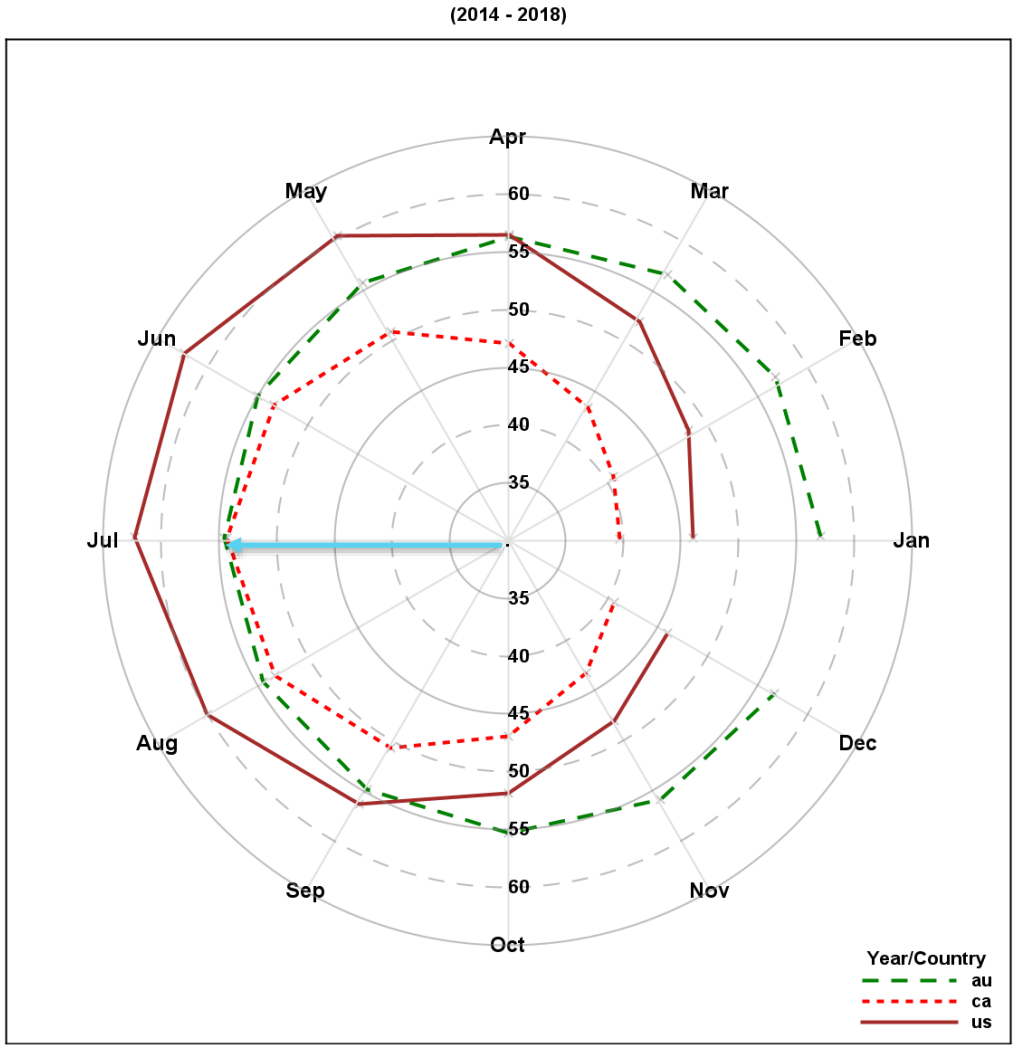
Polar Coordinates (time series)





Cosinar Analysis (SAS macro)
show seasonal differences (center=30)





_5_ Proportions and Odds Ratios

Graphs similar to plotting means and differences

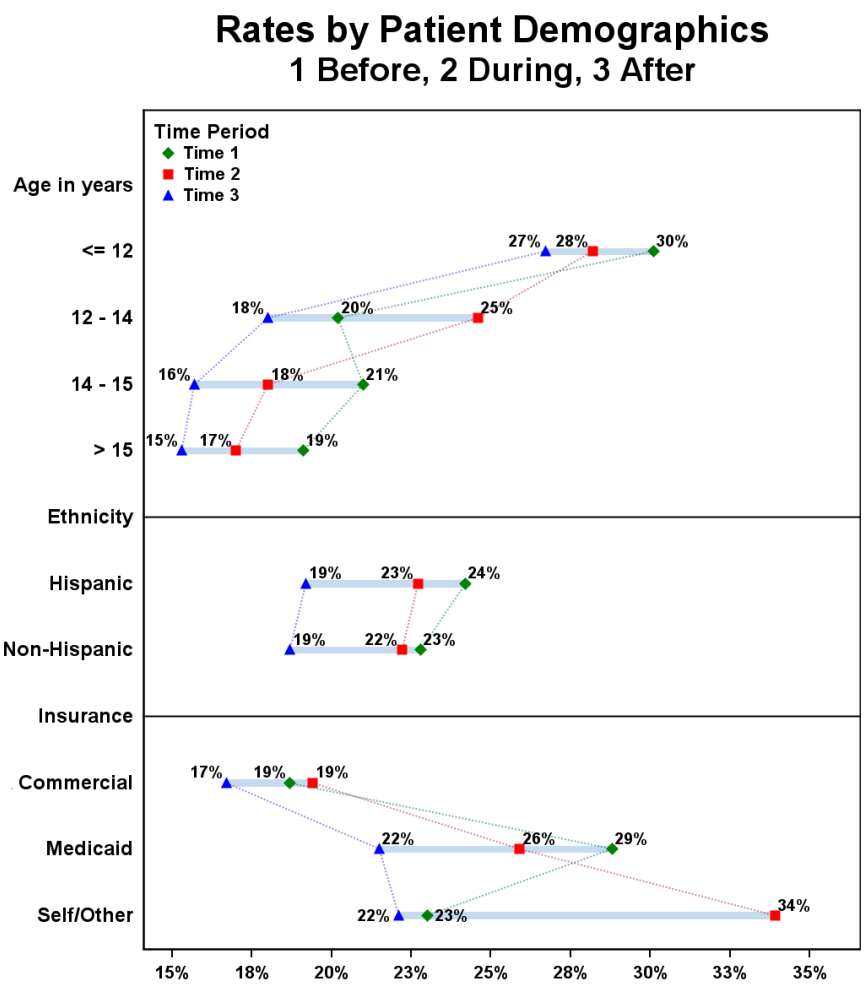
Results displayed in a table difficult to visualize patterns or comparisons

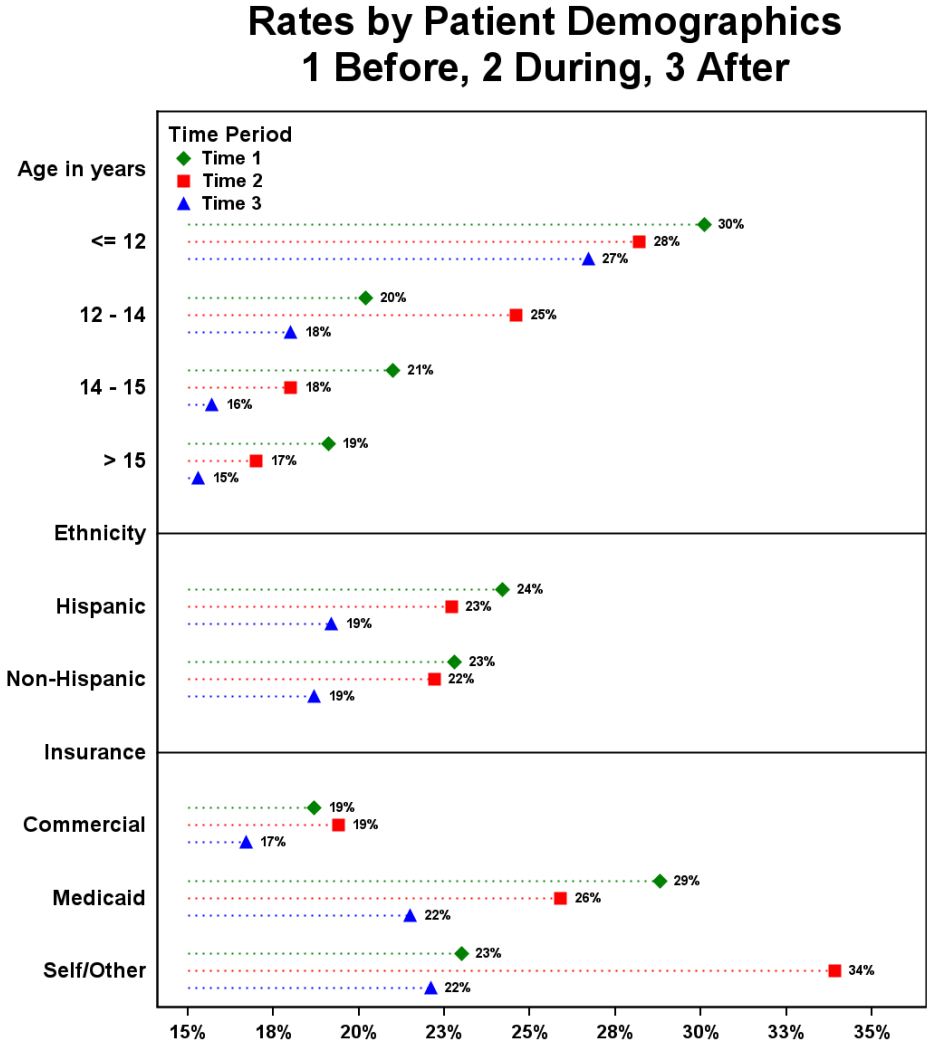
Graphs to visualize predicted proportions

Differences interpreted with:

- Odds ratios
- Risk ratios

_5_ Proportions of cases when surgery required for a medical condition





Odds Ratios: 2x2 tables

Berkley Admissions Data: Department A (of six)

PROC FREQ to compute percents and odds ratio

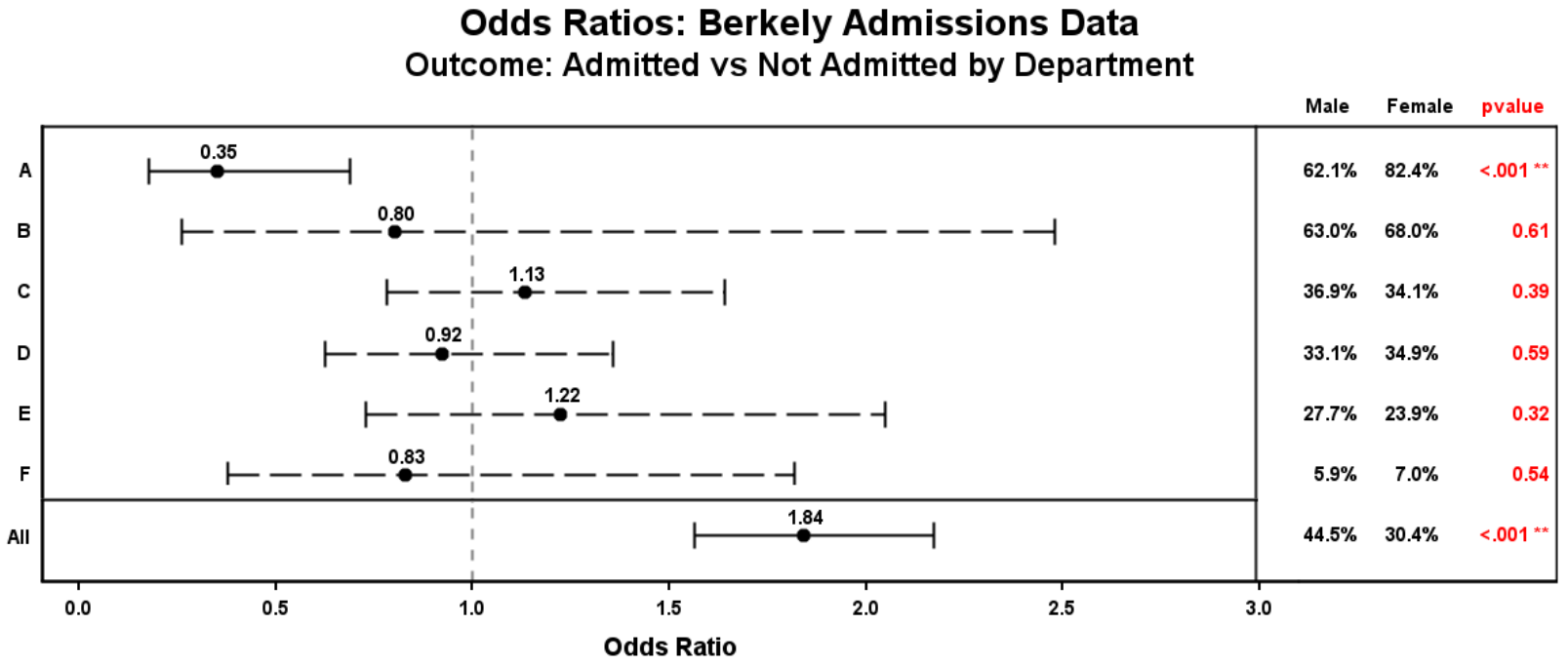
Table of sex by Admitted (Yes vs No)

sex Frequency|	Admitted? a Yes	 b No	 Total
a Male	512	313	825
	62.1	37.9	
b Female	89	19	108
	82.4	17.6	
Total	601	332	933

odds ratio =
62.1/37.9
--------- = 0.35
82.4/17.6

males less likely
admitted dept. A

Odds Ratios: 2x2 tables



Odds ratio GT 1 indicate males more likely admitted than females

Columns Male | Female contain percent admitted

Frequency| Row Pct	a Yes	b No	Total
a Male	1198 44.5	1493 55.5	2691
b Female	557 30.4	1278 69.6	1835
Total	1755	2771	4526

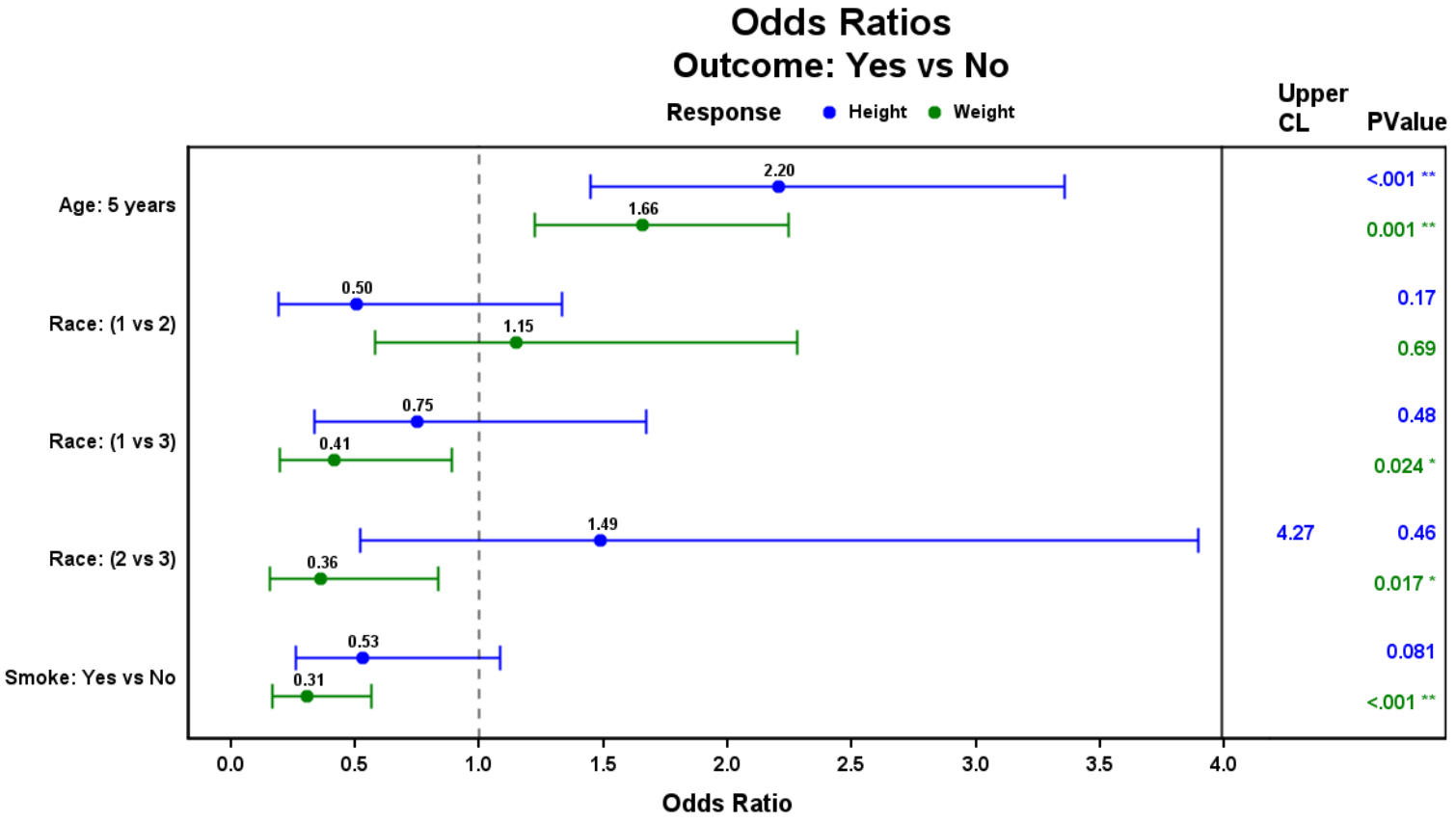
All Departments (A – F)

produces
odds ratio = 1.84

Unadjusted Odds ratios on two outcomes

Two binary outcome variables: height, weight
Three fixed effects

- **Age (continuous)**
- **Race (3 levels)**
- **Smoker (2 levels)**



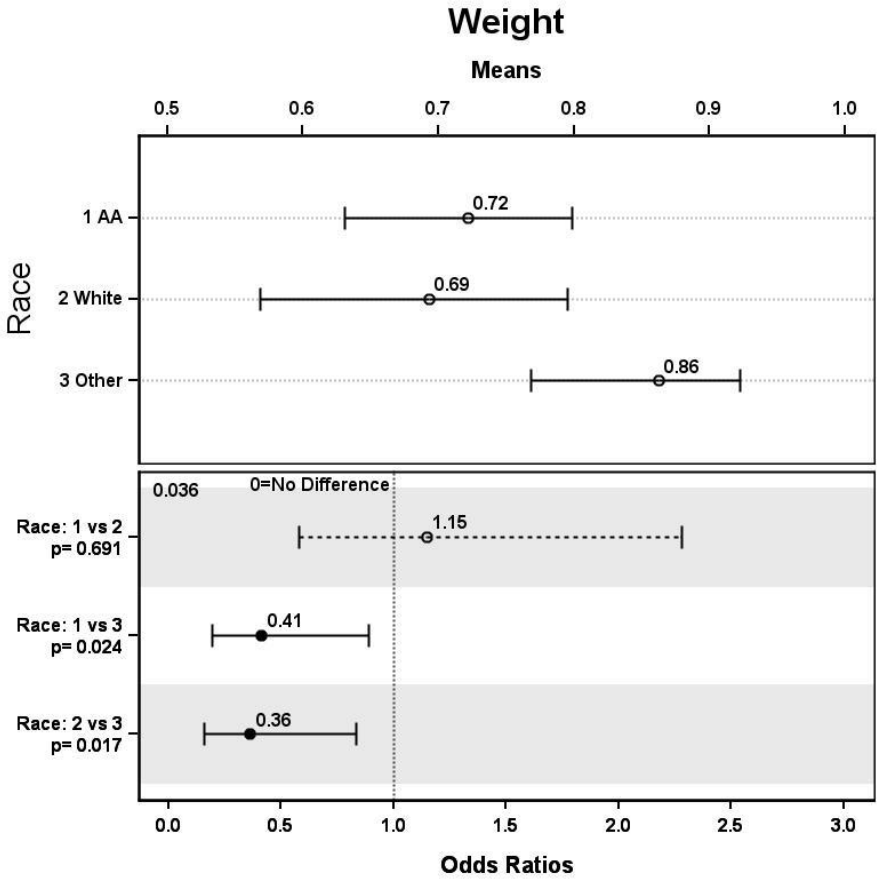
for categorical X data, odds ratio GT 1 indicates factor level on the left (xi vs xj) increases prob 'Yes'
for continuous X data, an odds ratio GT 1 indicates increase in numerical value increases prob 'Yes'

_5_ Proportion based on race with unadjusted Odds Ratios

Table of race by yB

race(Race) yB(Weight)

Frequency| Row Pct	a Yes	b No	Total
1 AA	78	30	108
	72.2	27.8	
2 White	43	19	62
	69.3	30.7	
3 Other	69	11	80
	86.2	13.8	
Total	190	60	250



PROCs LOGISTIC vs GLIMMIX

```
ODS OUTPUT oddsratiospl=odrtpl;

PROC LOGISTIC DATA=_dT desc;
CLASS race(ref='3') / param=ref;
MODEL race = race;
ODDSRATIO race / cl=pl;
run;
```

Can also enter:

```
CLASS race(ref='3') / param=glm;
```

options entered with PROC GLIMMIX to match PROC LOGISTIC results;

```
PROC GLIMMIX DATA=_dt
             method = mspl;
CLASS race (ref='3') ;
MODEL yB(desc) = race /
     dist=binary df = 99999;
LSMEANS race / ilink cl
             diff oddsratio;
RUN;
```

_6_ Spline

When to apply a spline?

Examine if a linear or curvilinear relationship is not practical with a binary or continuous response

- **Infant survival with birth weight**
- **Proportion of medical cases needing surgery over time**
- **Muscle strength/flexibility with leg angle**

_6_ Spline

Usually works best with large data sets

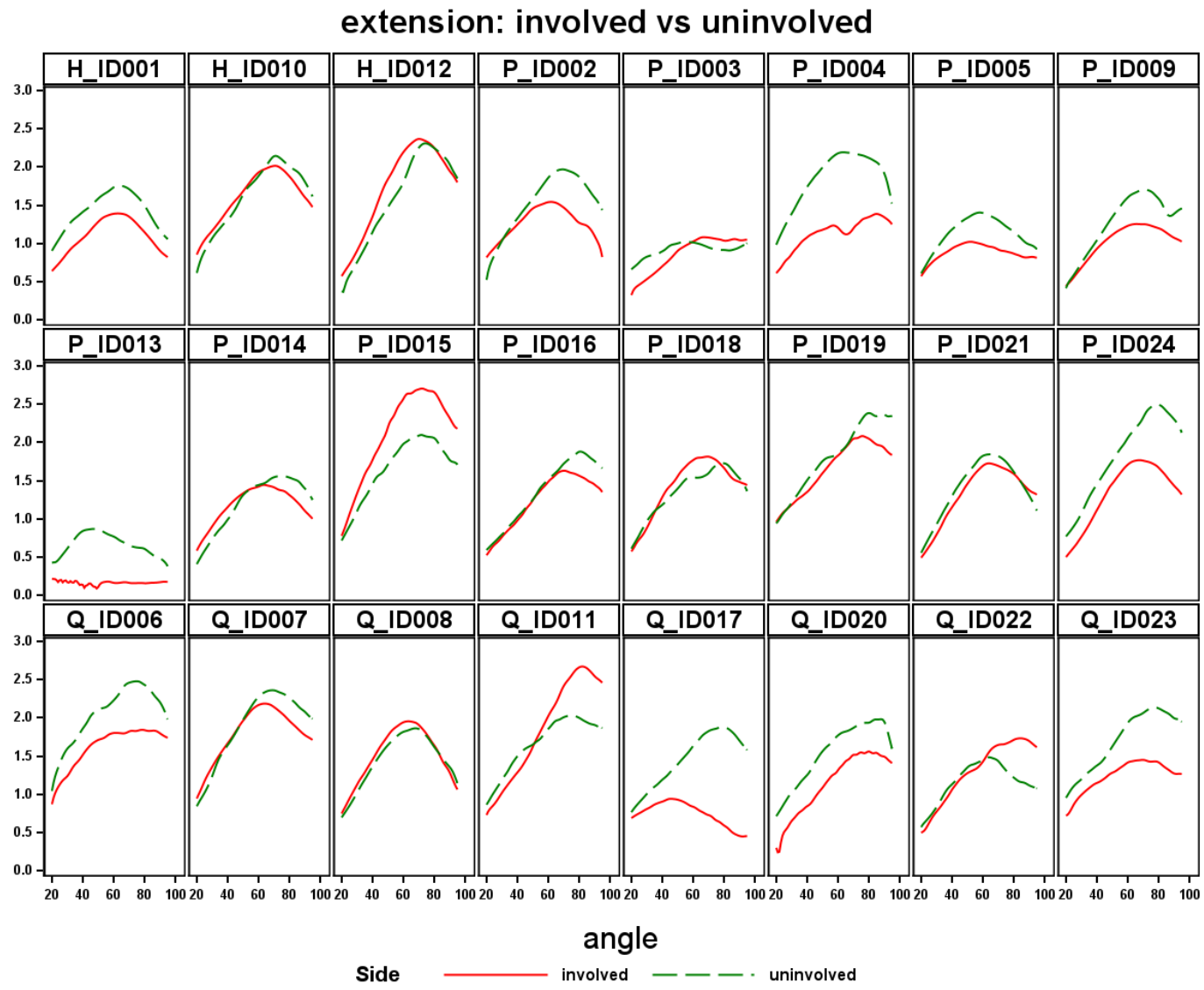
Appropriatness for a spline may be detected with an LOESS curve fit to the data

Generalized Linear Mixed Models, 2nd Ed.
Walter W. Stroup, Marina Ptukhina, and Julie Garai
Chapman Hall

Chapter 16: Smoothing Splines and Additive Models

SAS Code for the book available for download

Individual Profiles

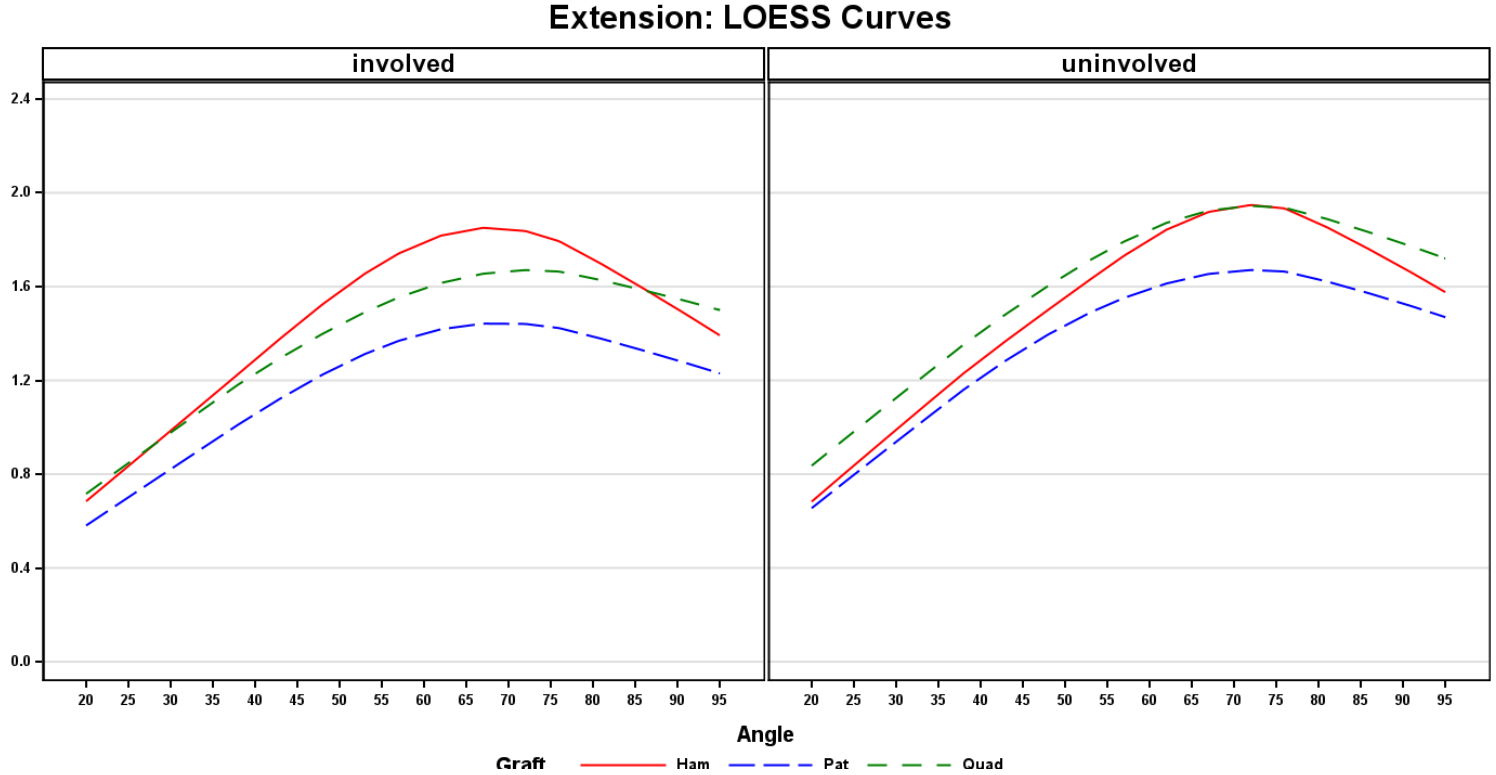


GLIMMIX code

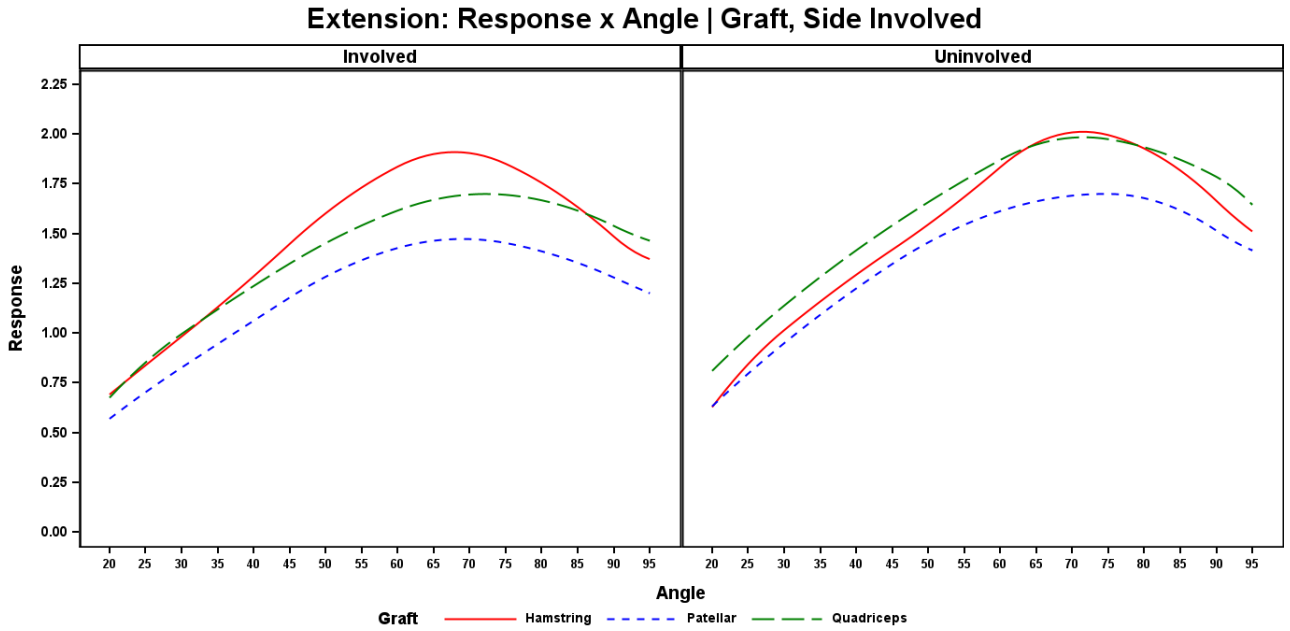
```
Angle is the variable name in the SAS data set

ODS OUTPUT       lsmeans = lsmsp
            lsmestimates = lsmEs;

PROC GLIMMIX DATA=isok  method=quad  initglm;
CLASS graft side id;

EFFECT ang = spline(angle / knotmethod=list(30 45 60 75 90)
                            degree=2 databoundary details);

MODEL extension = graft | side| ang / dist=normal solution cl;
RANDOM intercept / subject=shortid;
RANDOM intercept / subject=shortid*side;
RANDOM ang       / subject=shortid;

%lsmG(20,95,1);

%lsmESide(20,95,1);

RUN;
```

LSMEANS and LSMESTIMATE

```
%MACRO lsmG(lw,hg,inc);

%DO angl = &lw. %to &hg. %by &inc.;
   LSMEANS graft*side / cl at angle=&angl. ;
%END;
%MEND;

%MACRO lsmESide(lw,hg,inc);

%DO angl = &lw. %to &hg. %by &inc.;
  LSMESTIMATE graft*side
     "&angl  Ham: Inv vs UnInv" [1, 1 1] [-1, 1 2],
     "&angl  Pat: Inv vs UnInv" [1, 2 1] [-1, 2 2],
     "&angl Quad: Inv vs UnInv" [1, 3 1] [-1, 3 2] /
     cl at angle=&angl. adjust=simulate(nsamp=40000 seed=2233);
%END;
%MEND;

NOTE: SAS data set name entered in the LSMEANS and LSMESTIMATE
      statements
```



LOESS curves



LSMEANS from GLIMMIX with SPLINE

SGPLOT: BAND

show difference between two horizontal curves

From the file of LSMEAN differences, DATA=dfs, shade area between two boundary lines (space that lies between lower and upper confidence limits)

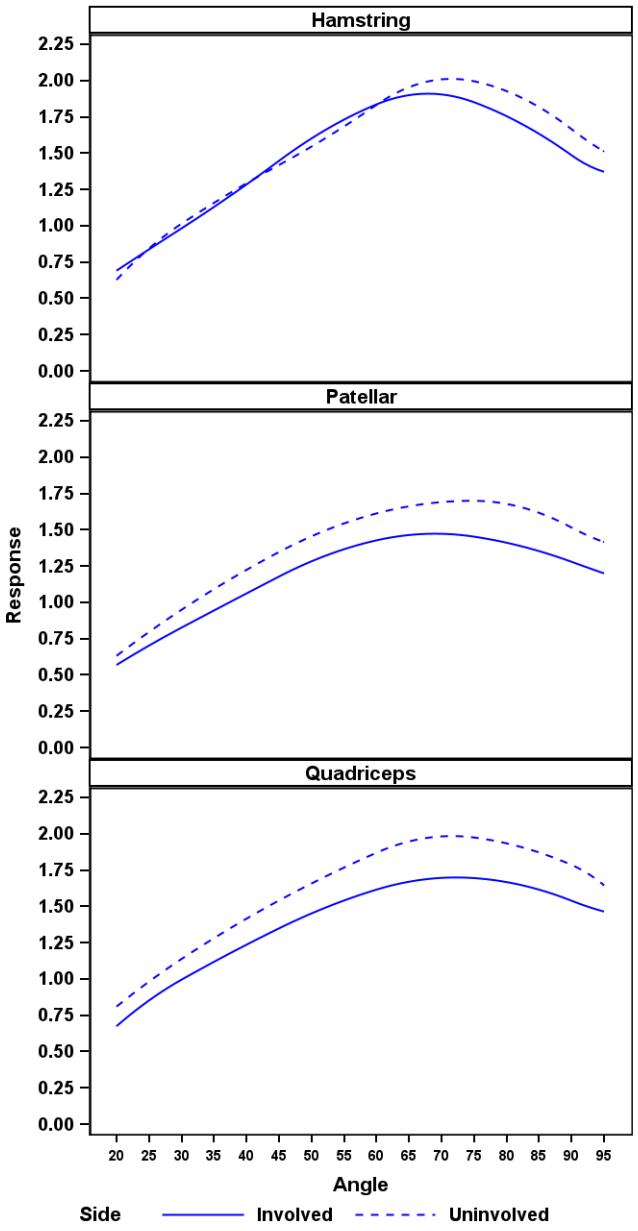
Where probt < 0.05, set _1 = lowerCL
_2 = upperCl

```
PROC SGPLOT DATA=dfs;
BAND x = covariate
     lower=_1  /* lower boundary */
     upper=_2  /* upper boundary */  / group= <var>
      fill
      fillattrs=(transparency=.85 color=blue) ;
RUN;
```

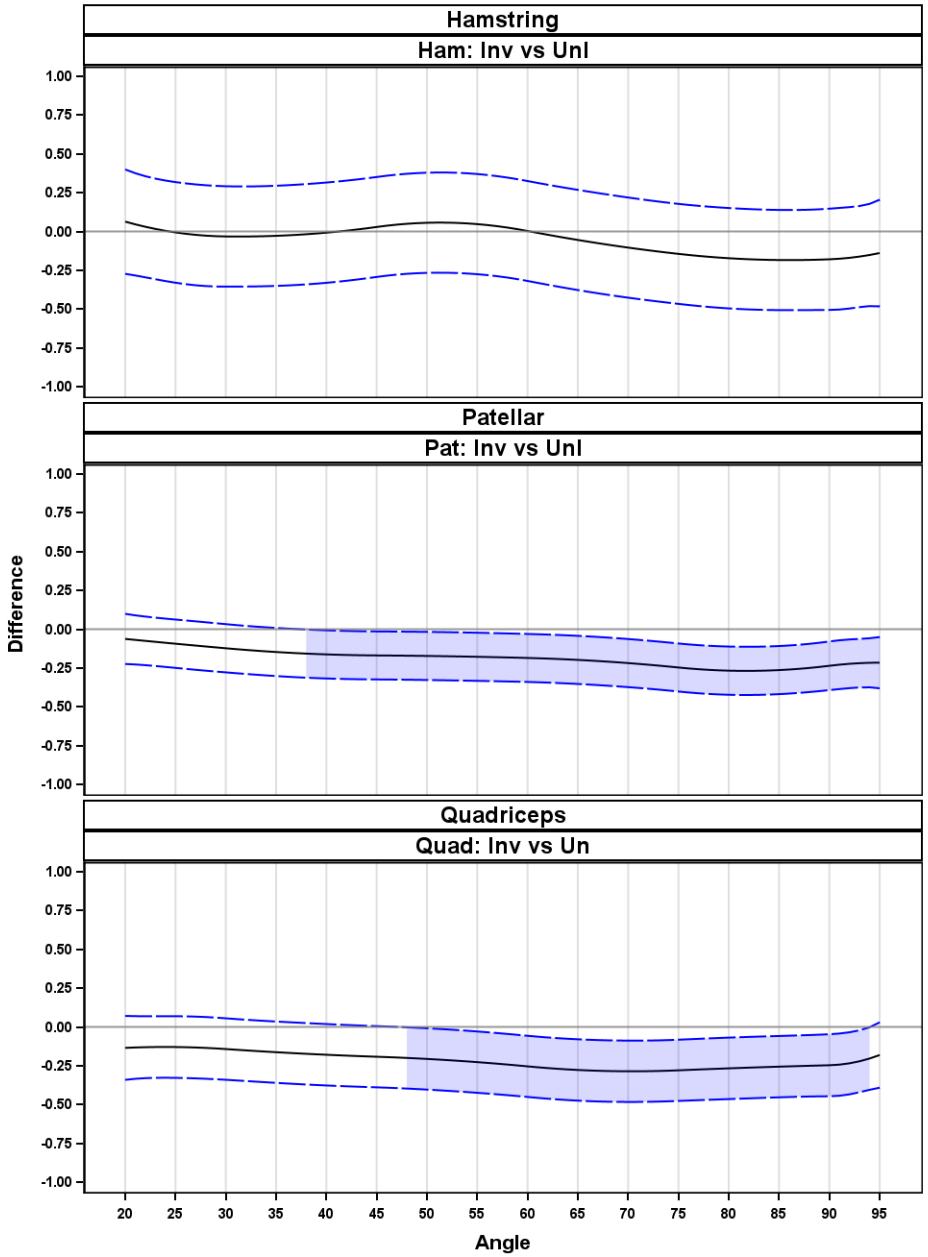
Show confidence intervals for differences in means (with small pvalues) for specific values of a continuous covariate, either with ANCOVA or with splines

Group=<var> allows for gaps in the shaded portions

Extension: Response x Angle | Graft, Side Involved

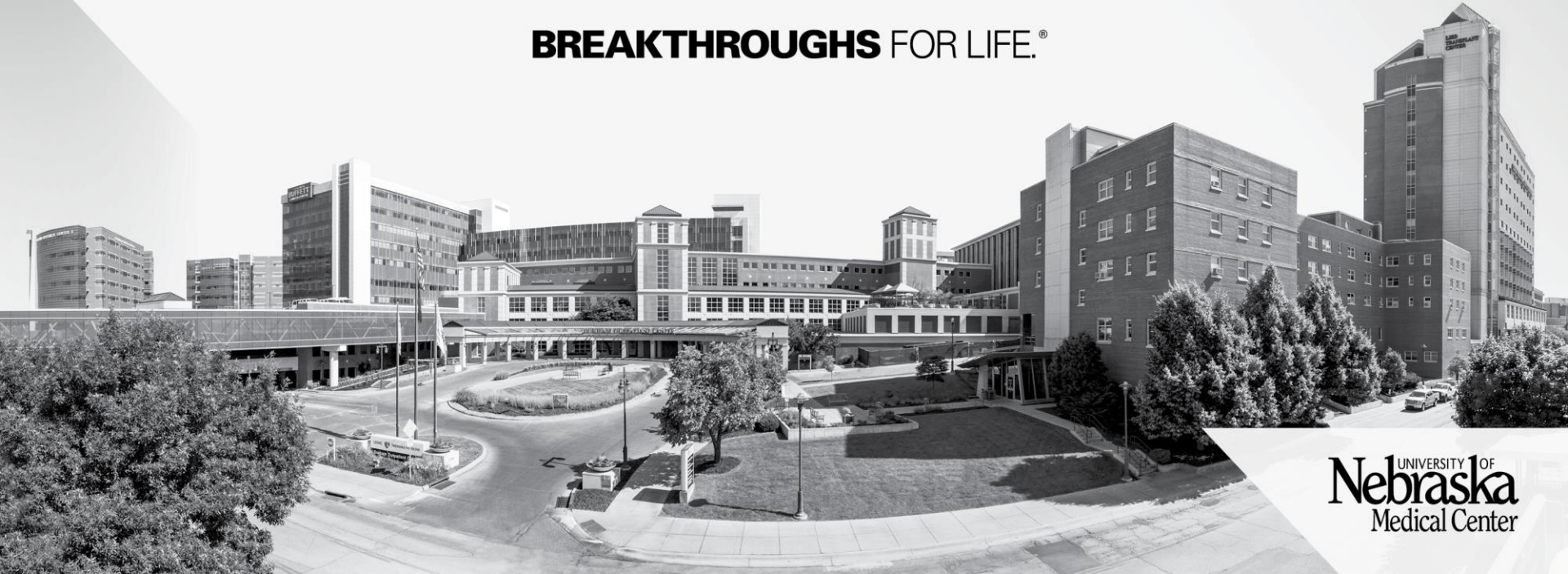


Extension: Response x Angle | Graft, Side Involved



University of Nebraska Medical Center℠

BREAKTHROUGHS FOR LIFE.®



University of Nebraska Medical Center