# Homework 8

Xinyi Lin

10/12/2019

Question 1

Weighted histogram of the Satisfaction with Government is shown above.

```
libname esshw '/folders/myfolders/health_survey';

proc contents data=esshw.ess6hw8;
run;

/*Question 1*/
title "Weighted Histogram of the Satisfaction with Government";
proc sgplot data=esshw.ess6hw8 noautolegend;
    histogram stfgov / WEIGHT=pspwght ;
    yaxis grid offsetmin=0.05 label="Weighted Proportion";
run;
```
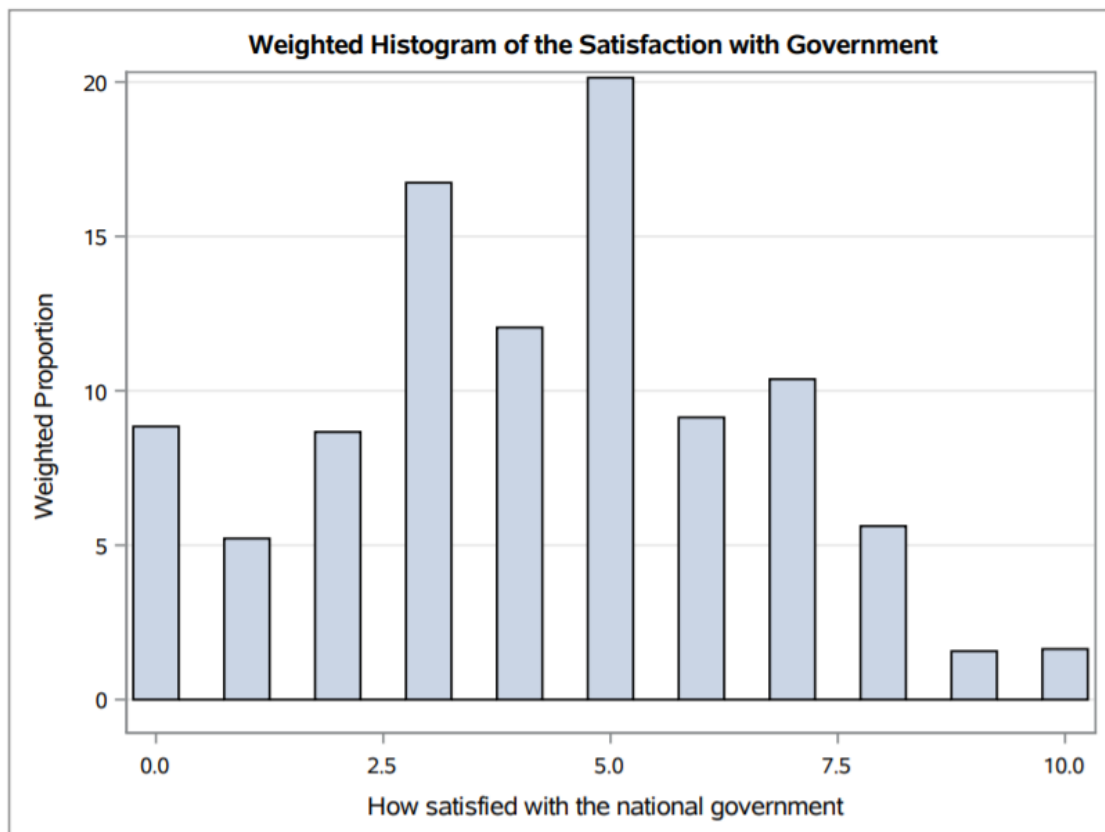
The distribution of variable STFGOV is similar to a normal distribution but not match around 5.

This variable can function as a continuous outcome because: 1. The variable is ordinal which means larger numbers indicates higher levels of satisfaction. 2. The increase of satisfaction can be continuous.

Question 2

1) Political scale

```
proc surveyreg data=esshw.ess6hw8;
strata stratify; cluster psu; weight pspwght;
class lr3cat;
model stfgov=lr3cat / solution;
run;
```

The SURVEYREG Procedure

Regression Analysis for Dependent Variable STFGOV

Data Summary

Number of Observations	1771
Sum of Weights	1765.1
Weighted Mean of STFGOV	4.38351
Weighted Sum of STFGOV	7737.3

Design Summary

Number of Strata	8
Number of Clusters	183

Fit Statistics

R-Square	0.02825
Root MSE	2.2795
Denominator DF	175

Class Level Information

CLASS Variable	Label	Levels	Values
lr3cat	1=left 2=middle 3=right	3	1 2 3

Tests of Model Effects

Effect	Num DF	F Value	Pr > F
Model	2	10.63	<.0001
Intercept	1	1967.39	<.0001
lr3cat	2	10.63	<.0001

Note: The denominator degrees of freedom for the F tests is 175.

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	4.9063498	0.16262245	30.17	<.0001
lr3cat 1	-1.0197408	0.22546454	-4.52	<.0001
lr3cat 2	-0.6769416	0.19616263	-3.45	0.0007
lr3cat 3	0.0000000	0.00000000	.	.

Note: The degrees of freedom for the t tests is 175.
Matrix X'WX is singular and a generalized inverse was used to solve the normal equations. Estimates are not unique.

2) Gender

```
proc surveyreg data=esshw.ess6hw8;
strata stratify; cluster psu; weight pspwght;
class gndr;
model stfgov=gndr / solution;
run;
```

The SURVEYREG Procedure

Regression Analysis for Dependent Variable STFGOV

Data Summary

Number of Observations	2386
Sum of Weights	2382.5
Weighted Mean of STFGOV	4.24591
Weighted Sum of STFGOV	10116.0

Design Summary

Number of Strata	8
Number of Clusters	184

Fit Statistics

R-Square	0.005728
Root MSE	2.3858
Denominator DF	176

Class Level Information

CLASS Variable	Label	Levels	Values
GNDR	Gender	2	1 2

Tests of Model Effects

Effect	Num DF	F Value	Pr > F
Model	1	9.14	0.0029
Intercept	1	1873.38	<.0001
GNDR	1	9.14	0.0029

Note: The denominator degrees of freedom for the F tests is 176.

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	4.4116868	0.10485046	42.08	<.0001
GNDR 1	-0.3634242	0.12023642	-3.02	0.0029
GNDR 2	0.0000000	0.00000000	.	.

Note: The degrees of freedom for the t tests is 176.
Matrix X'WX is singular and a generalized inverse was used to solve the normal equations. Estimates are not unique.

3) Self-rated health status

```
proc surveyreg data=esshw.ess6hw8;
strata stratify; cluster psu; weight pspwght;
class health;
model stfgov=health / solution;
run;
```

The SURVEYREG Procedure

Regression Analysis for Dependent Variable STFGOV

Data Summary	
Number of Observations	2372
Sum of Weights	2367.6
Weighted Mean of STFGOV	4.25536
Weighted Sum of STFGOV	10074.9

Design Summary	
Number of Strata	8
Number of Clusters	184

Fit Statistics	
R-Square	0.002867
Root MSE	2.3858
Denominator DF	176

Class Level Information			
CLASS Variable	Label	Levels	Values
HEALTH	Subjective general health	5	1 2 3 4 5

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	4	0.89	0.4705
Intercept	1	1161.38	<.0001
HEALTH	4	0.89	0.4705

Note: The denominator degrees of freedom for the F tests is 176.

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	4.04611397	0.34275837	11.80	<.0001
HEALTH 1	0.41548772	0.45014908	0.92	0.3573
HEALTH 2	0.35902707	0.35769830	1.00	0.3169
HEALTH 3	0.10071163	0.34164133	0.29	0.7685
HEALTH 4	0.17075890	0.35927701	0.48	0.6352
HEALTH 5	0.00000000	0.00000000	.	.

Note: The degrees of freedom for the t tests is 176.
Matrix X'WX is singular and a generalized inverse was used to solve the normal equations. Estimates are not unique.

4) Satisfaction of life

```
proc surveyreg data=esshw.ess6hw8;
strata stratify; cluster psu; weight pspwght;
model stfgov=stflife / solution;
run;
```

The SURVEYREG Procedure

Regression Analysis for Dependent Variable STFGOV

Data Summary	
Number of Observations	2366
Sum of Weights	2362.8
Weighted Mean of STFGOV	4.25412
Weighted Sum of STFGOV	10051.8

Design Summary	
Number of Strata	8
Number of Clusters	184

Fit Statistics	
R-Square	0.1059
Root MSE	2.2597
Denominator DF	176

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	1	130.27	<.0001
Intercept	1	138.95	<.0001
STFLIFE	1	130.27	<.0001

Note: The denominator degrees of freedom for the F tests is 176.

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2.31554312	0.19643557	11.79	<.0001
STFLIFE	0.33283616	0.02916106	11.41	<.0001

Note: The degrees of freedom for the t tests is 176.

STFGOV Fit Plot
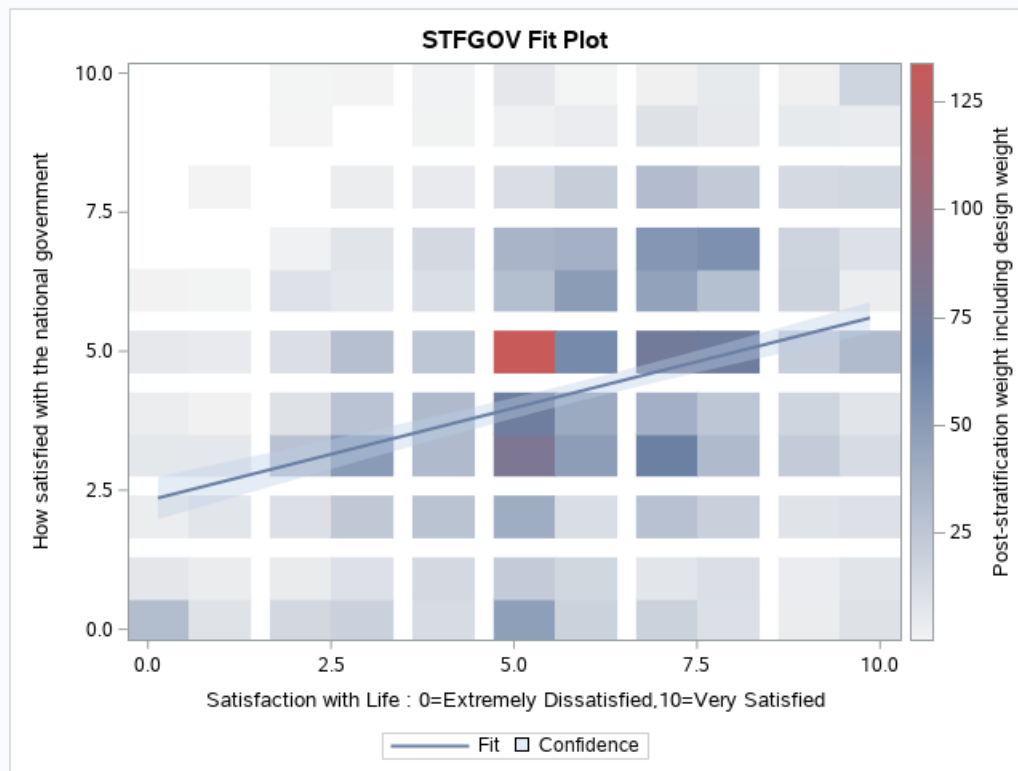
Conclusion: R-squared of above models are small and but predictor of model 1, 2, 4 are significant. This means those predictors are effective but not enough to predict dependent variable.

For R-squared: model 4 > model 2 > model 1 > model 3. This means if we fit bivariate models, satisfaction of life is the best predictor among these four predictors.

Question 3

The multivariate model:

```
proc surveyreg data=esshw.ess6hw8;
strata stratify; cluster psu; weight pspwght;
class lr3cat gndr health;
model stfgov=lr3cat gndr health stflife/ solution clparm deff;
output out=outdiag1 p=phat r=resid;
run;
```

The SURVEYREG Procedure

Regression Analysis for Dependent Variable STFGOV

Data Summary	
Number of Observations	1752
Sum of Weights	1747.9
Weighted Mean of STFGOV	4.39844
Weighted Sum of STFGOV	7688.0

Design Summary	
Number of Strata	8
Number of Clusters	183

Fit Statistics	
R-Square	0.1271
Root MSE	2.1574
Denominator DF	175

Class Level Information			
CLASS Variable	Label	Levels	Values
lr3cat	1=left 2=middle 3=right	3	1 2 3
GNDR	Gender	2	1 2
HEALTH	Subjective general health	5	1 2 3 4 5

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	8	19.18	<.0001
Intercept	1	156.86	<.0001
lr3cat	2	5.56	0.0045
GNDR	1	4.11	0.0441
HEALTH	4	1.22	0.3037
STFLIFE	1	118.68	<.0001

Note: The denominator degrees of freedom for the F tests is 175.

Estimated Regression Coefficients							
Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Interval		Design Effect
Intercept	3.2488385	0.36569260	8.88	<.0001	2.5271031	3.9705740	0.74
lr3cat 1	-0.6520633	0.20755234	-3.14	0.0020	-1.0616912	-0.2424354	1.95
lr3cat 2	-0.5064663	0.18162945	-2.79	0.0059	-0.8649324	-0.1480001	2.37
lr3cat 3	0.0000000	0.00000000	.	.	0.0000000	0.0000000	.
GNDR 1	-0.2383797	0.11754339	-2.03	0.0441	-0.4703648	-0.0063946	1.27
GNDR 2	0.0000000	0.00000000	.	.	0.0000000	0.0000000	.
HEALTH 1	-0.1794881	0.42704827	-0.42	0.6748	-1.0223159	0.6633397	0.88
HEALTH 2	-0.4584986	0.33557944	-1.37	0.1736	-1.1208024	0.2038051	0.66
HEALTH 3	-0.2505030	0.33077616	-0.76	0.4499	-0.9033270	0.4023209	0.66
HEALTH 4	0.0124950	0.35495033	0.04	0.9720	-0.6880394	0.7130294	0.69
HEALTH 5	0.0000000	0.00000000	.	.	0.0000000	0.0000000	.
STFLIFE	0.3266229	0.02998195	10.89	<.0001	0.2674501	0.3857957	1.58

Note: The degrees of freedom for the t tests is 175.
Matrix X'WX is singular and a generalized inverse was used to solve the normal equations. Estimates are not unique.

STFGOV Fit Plot
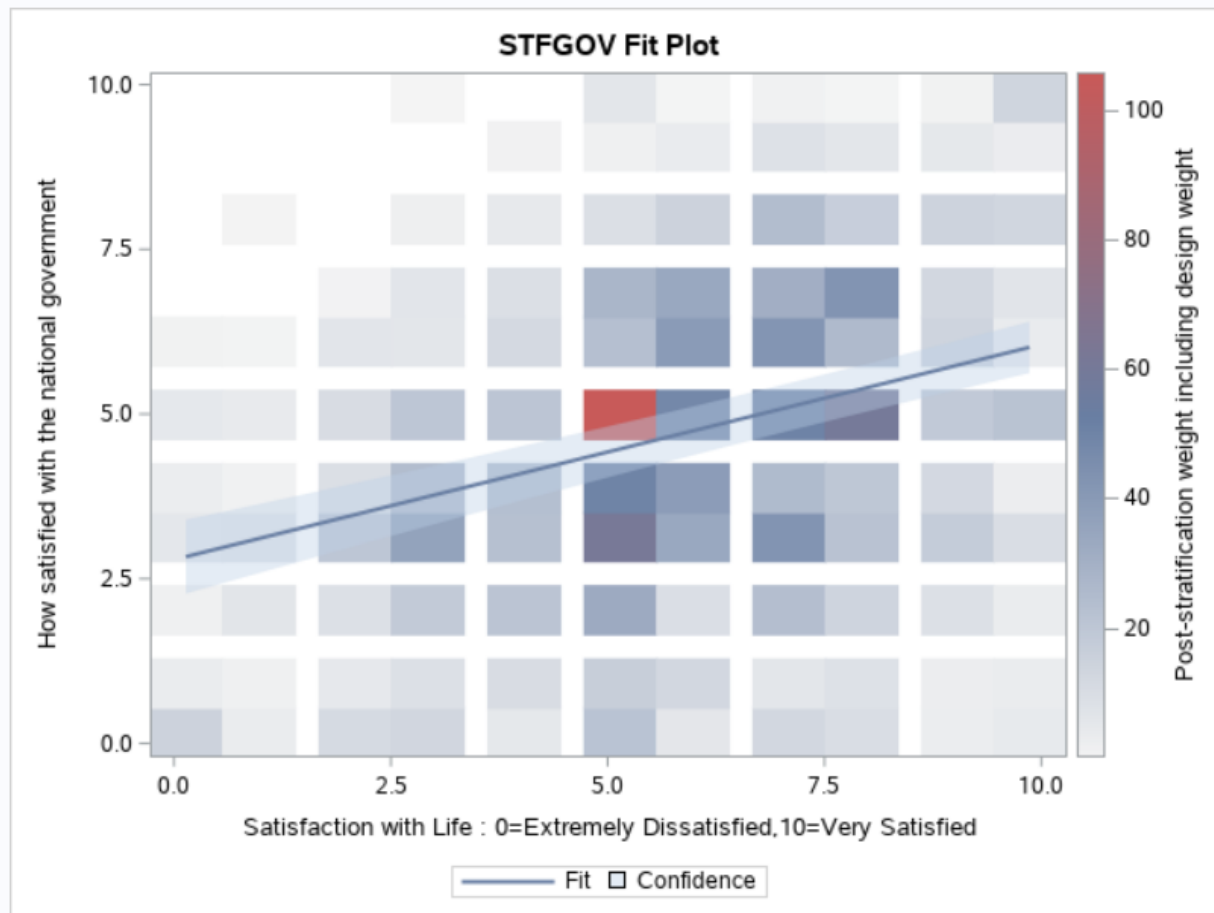
1)

According to results above, R-squared of this multivariate model is 0.1271.

R-squared is a statistical measure of how close the data are to the fitted regression line. R-square equals to 0.1271 indicates that the model explains 12.71% of the variability of the response data around its mean.

2)

The design effect for the slope associated with gender is 1.27. This means the ratio of true variance of this design in gender and the variance if data gathered using simple random sampling is 1.27.

3)

```
/*residual plot*/
title "Scatter Plot, Residual by Satisfaction of Life";
proc sgplot data=outdiag1;
scatter y=resid x=stflife;
run;
```

Scatter Plot, Residual by Satisfaction of Life
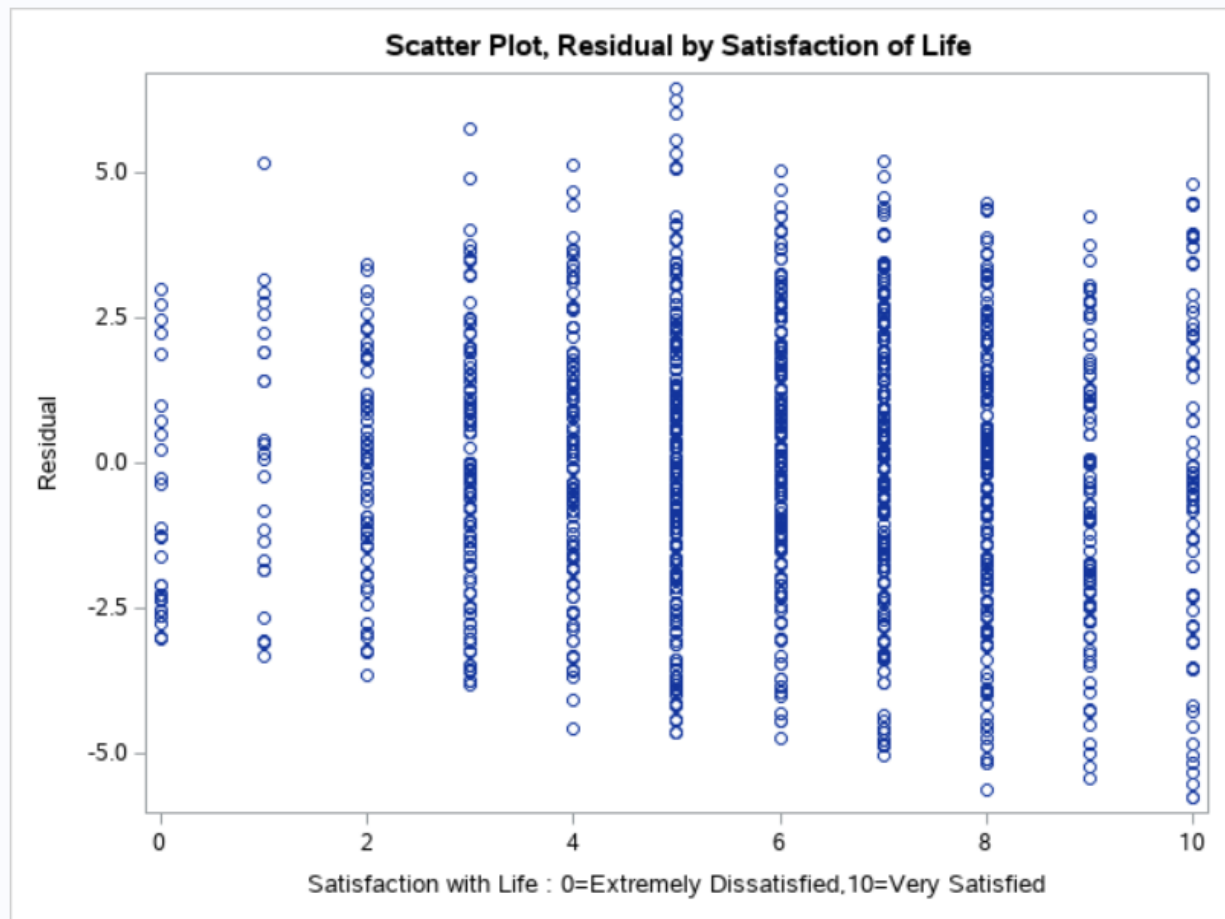
According to above plot, we can find that residuals do not random and constant across satisfaction with life. This means the model does not fit data well.

4)

①Comparing to those with LR3CAT equaling to 3, those with LR3CAT equaling to 1 and 2 have less satisfaction with government. ②Comparing to female, male have less satisfaction with government. ③Among people with different self-rated health status, people with bad self-rated health status have more satisfaction with government. ④With satisfaction of life increases, satisfaction of government increases.