Intermediate SAS

Hsueh-Sheng Wu

CFDR Workshop Series

Spring 2011

BGSU
Center for Family and Demographic Research

Outline

- Validity and Reliability of A Scale
  - Different types of reliability and validity
  - Cronbach's Alpha (Inter-item Reliability)
  - Factor Analysis (Construct Validity)
- Inference Analyses for Continuous Dependent Variables
  - T-test
  - Anova
  - Ordinary Least Squares (OLS) regression
- Inference Analysis for Categorical Dependent Variables
  - Chi-square test
  - Testing proportions
  - Logistic regression
  - Ordered logistic regression
  - Multinomial logistic regression

Validity and Reliability of A Scale

- Reliability:
  - Inter-item reliability (internal consistency)
  - Inter-observer reliability
  - Test-retest reliability
  - Alternate-forms reliability (split-halves reliability)

- Validity:
  - Face validity
  - Criterion validity: concurrent validity and predictive validity
  - Construct validity: convergent validity and discriminant validity
  - Content validity

Cronbach's Alpha (Inter-Item Reliability)

- **SAS Command:**

```
Proc contents data =in.bg2;
Run;
```

- **SAS Output:**

The SAS System 13:30 Thursday, February 3, 2011 79

The CONTENTS Procedure

Data Set Name	IN.BG2	Observations	568
Member Type	DATA	Variables	7
Engine	V9	Indexes	0
Created	Thu, Feb 03, 2011 07:01:49 PM	Observation Length	52
Last Modified	Thu, Feb 03, 2011 07:01:49 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label	bg2 dataset written by Stat/Transfer Ver. 10.1.1773.1104		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
2	bg2cost1	Num	8	Best health care is expensive
3	bg2cost2	Num	8	Cost is a major consideration
4	bg2cost3	Num	8	Determine cost of tests first
5	bg2cost4	Num	8	Monitor likely complications only
6	bg2cost5	Num	8	Use all means regardless of cost
7	bg2cost6	Num	8	Prefer unnecessary tests to missing tests
1	clinid	Num	4	Physician identifier

- **SAS Command:**

```
proc corr data= in.bg2;
var  bg2cost1 bg2cost2 bg2cost3 bg2cost4 bg2cost5 bg2cost6;
run;
```

- **SAS Output:**

Pearson Correlation Coefficients, N = 568
Prob > |r| under H0: Rho=0

	bg2cost1	bg2cost2	bg2cost3
bg2cost1	1.00000	0.09200	0.05395
Best health care is expensive		0.0283	0.1992
bg2cost2	0.09200	1.00000	0.32816
Cost is a major consideration	0.0283		<.0001

Pearson Correlation Coefficients, N = 568
Prob > |r| under H0: Rho=0

	bg2cost4	bg2cost5	bg2cost6
bg2cost1	-0.03800	0.23799	0.24314
Best health care is expensive	0.3659	<.0001	<.0001
bg2cost2	0.14200	-0.13937	-0.06709
Cost is a major consideration	0.0007	0.0009	0.1102

The SAS System 13:30 Thursday, February 3, 2011 81
The CORR Procedure

Pearson Correlation Coefficients, N = 568
Prob > |r| under H0: Rho=0

	bg2cost1	bg2cost2	bg2cost3
bg2cost3	0.05395	0.32816	1.00000
Determine cost of tests first	0.1992	<.0001	
bg2cost4	-0.03800	0.14200	0.26765
Monitor likely complications only	0.3659	0.0007	<.0001
bg2cost5	0.23799	-0.13937	-0.05500
Use all means regardless of cost	<.0001	0.0009	0.1906
bg2cost6	0.24314	-0.06709	-0.10749
Prefer unnecessary tests to missing tests	<.0001	0.1102	0.0104

Pearson Correlation Coefficients, N = 56
Prob > |r| under H0: Rho=0

	bg2cost4	bg2cost5	bg2cost6
bg2cost3	0.26765	-0.05500	-0.10749
Determine cost of tests first	<.0001	0.1906	0.0104
bg2cost4	1.00000	-0.05670	-0.13294
Monitor likely complications only		0.1772	0.0015
bg2cost5	-0.05670	1.00000	0.35237
Use all means regardless of cost	0.1772		<.0001
bg2cost6	-0.13294	0.35237	1.00000
Prefer unnecessary tests to missing tests	0.0015	<.0001	

Cronbach's Alpha (Continued)

- **SAS command:**

```
proc corr data= in.bg2 alpha;
var  bg2cost1 bg2cost2 bg2cost3 bg2cost4 bg2cost5 bg2cost6;
run;
```

- **SAS Output:**

Cronbach Coefficient Alpha

Variables	Alpha
Raw	0.326356
Standardized	0.326356

Cronbach Coefficient Alpha with Deleted Variable

	Raw Variables		Standardized Variables		
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha	Label
bg2cost1	0.239235	0.219188	0.239235	0.219188	Best health care is expensive
bg2cost2	0.139199	0.292869	0.139199	0.292869	Cost is a major consideration
bg2cost3	0.194650	0.252679	0.194650	0.252679	Determine cost of tests first
bg2cost4	0.069402	0.341218	0.069402	0.341218	Monitor likely complications only
bg2cost5	0.132442	0.297658	0.132442	0.297658	Use all means regardless of cost
bg2cost6	0.111547	0.312318	0.111547	0.312318	Prefer unnecessary tests to missing tests

Center for Family and Demographic Research

Cronbach's Alpha (Continued)

SAS command:

```
data in.bg3;
set in.bg2;
r_bg2cost2 = bg2cost2 *-1 ;
r_bg2cost3 = bg2cost3 *-1 ;
r_bg2cost4 = bg2cost4 *-1 ;
run;
proc corr data= in.bg3 alpha;
var  bg2cost1 r_bg2cost2 r_bg2cost3 r_bg2cost4 bg2cost5 bg2cost6;
run;
```

SAS output:

Cronbach Coefficient Alpha

Variables	Alpha
Raw	0.483148
Standardized	0.483148

Cronbach Coefficient Alpha with Deleted Variable

	Raw Variables		Standardized Variables	
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha
bg2cost1	0.129552	0.496766	0.129552	0.496766
r_bg2cost2	0.208335	0.456315	0.208335	0.456315
r_bg2cost3	0.254904	0.431425	0.254904	0.431425
r_bg2cost4	0.228639	0.445553	0.228639	0.445553
bg2cost5	0.310129	0.400939	0.310129	0.400939
bg2cost6	0.335662	0.386482	0.335662	0.386482

Cronbach's Alpha (Continued)

SAS command:

```
proc corr data= in.bg3 alpha;
  var  r_bg2cost2 r_bg2cost3 r_bg2cost4 bg2cost5 bg2cost6;
run;
```

SAS output:

Cronbach Coefficient Alpha

Variables	Alpha
Raw	0.496766
Standardized	0.496766

Cronbach Coefficient Alpha with Deleted Variable

Deleted Variable	Raw Variables: Correlation with Total	Raw Variables: Alpha	Standardized Variables: Correlation with Total	Standardized Variables: Alpha
r_bg2cost2	0.277524	0.436114	0.277524	0.436114
r_bg2cost3	0.315385	0.410763	0.315385	0.410763
r_bg2cost4	0.242667	0.458868	0.242667	0.458868
bg2cost5	0.244510	0.457679	0.244510	0.457679
bg2cost6	0.269901	0.441138	0.269901	0.441138

Cronbach's Alpha (Continued)

```
. alpha bg2cost2 bg2cost3 bg2cost4 bg2cost5 bg2cost6
Test scale = mean(unstandardized items)
Reversed items:  bg2cost5 bg2cost6

Average interitem covariance:     .1648775
Number of items in the scale:            5
Scale reliability coefficient:      0.4968
```

```
. alpha bg2cost2 bg2cost3 bg2cost4 bg2cost5 bg2cost6,
    item
Test scale = mean(unstandardized items)
-                                                          average
                          item-test     item-rest       interitem
Item         |  Obs  Sign   correlation   correlation     covariance      alpha
-------------+------------------------------------------------------------------
bg2cost2     |  568    +       0.5821        0.2775        .1620242      0.4361
bg2cost3     |  568    +       0.6104        0.3154        .1484126      0.4108
bg2cost4     |  568    +       0.5552        0.2427        .1749139      0.4589
bg2cost5     |  568    -       0.5566        0.2445        .1742236      0.4577
bg2cost6     |  568    -       0.5762        0.2699        .1648135      0.4411
-------------+------------------------------------------------------------------
Test scale   |                                             .1648775      0.4968
```

Center for Family and Demographic Research

Factor Analysis (Construct Validity)

SAS command:

proc factor data = in.bg3 nfactors = 3 method = ml;
var bg2cost1 r_bg2cost2 r_bg2cost3 r_bg2cost4 bg2cost5 bg2cost6;
run;

SAS output:

The FACTOR Procedure

Initial Factor Method: Maximum Likelihood

Prior Communality Estimates: SMC

bg2cost1	r_bg2cost2	r_bg2cost3	r_bg2cost4	bg2cost5	bg2cost6
0.10542051	0.13702490	0.16371768	0.08658995	0.16709560	0.16833044

Preliminary Eigenvalues: Total = 0.97021123 Average = 0.16170187

	Eigenvalue	Difference	Proportion	Cumulative
1	1.00408133	0.37501044	1.0349	1.0349
2	0.62907089	0.60278365	0.6484	1.6833
3	0.02628724	0.19810690	0.0271	1.7104
4	-.17181966	0.04827133	-0.1771	1.5333
5	-.22009099	0.07722661	-0.2268	1.3064
6	-.29731759		-0.3064	1.0000

3 factors will be retained by the NFACTOR criterion.

WARNING: Too many factors for a unique solution.

BGSU Center for Family and Demographic Research

Factor Pattern

		Factor1	Factor2	Factor
bg2cost1	Best health care is expensive	0.23149	0.38402	0.16617
r_bg2cost2		0.41724	-0.38114	-0.25410
r_bg2cost3		0.44311	-0.46620	0.13542
r_bg2cost4		0.31331	-0.20144	0.25777
bg2cost5	Use all means regardless of cost	0.55638	0.36003	-0.17494
bg2cost6	Prefer unnecessary tests to missing tests	0.49482	0.27796	0.13153

Rotate the factor structure

SAS command:

```
proc factor data = in.bg3 nfactors = 3 method = ml rotate = varimax;
var bg2cost1 r_bg2cost2 r_bg2cost3 r_bg2cost4 bg2cost5 bg2cost6;
run;
```

SAS output:

Rotated Factor Pattern

		Factor1	Factor2	Factor3
bg2cost1	Best health care is expensive	0.44161	0.01890	-0.18248
r_bg2cost2		0.01644	0.24413	0.56925
r_bg2cost3		-0.00841	0.56607	0.33396
r_bg2cost4		0.09128	0.43985	0.05818
bg2cost5	Use all means regardless of cost	0.64034	-0.03596	0.24176
bg2cost6	Prefer unnecessary tests to missing tests	0.55207	0.18468	0.02272

Inference Analysis (Continuous Dependent Variables)

- t-test
- ANOVA
- Ordinary Least Squares (OLS) regression

t-test

SAS command:

```
proc ttest data=in.auto;
class foreign;
var mpg;
run;
```

SAS output:

The TTEST Procedure

Variable: mpg

foreign	N	Mean	Std Dev	Std Err	Minimum	Maximum
Domestic	52	19.8269	4.7433	0.6578	12.0000	34.0000
Foreign	22	24.7727	6.6112	1.4095	14.0000	41.0000
Diff (1-2)		-4.9458	5.3558	1.3622		

foreign	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Domestic		19.8269	18.5064	21.1475	4.7433	3.9751	5.8823
Foreign		24.7727	21.8415	27.7040	6.6112	5.0863	9.4478
Diff (1-2)	Pooled	-4.9458	-7.6612	-2.2304	5.3558	4.6059	6.3997
Diff (1-2)	Satterthwaite	-4.9458	-8.1201	-1.7716			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	72	-3.63	0.0005
Satterthwaite	Unequal	30.546	-3.18	0.0034

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	21	51	1.94	0.0549

ANOVA

SAS command:

```
PROC ANOVA DATA=in.auto;
CLASS foreign;
MODEL mpg = foreign  ;
RUN;
```

SAS output:

```
 The ANOVA Procedure

Dependent Variable: mpg
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	378.153515	378.153515	13.18	0.0005
Error	72	2065.305944	28.684805		
Corrected Total	73	2443.459459			

R-Square	Coeff Var	Root MSE	mpg Mean
0.154762	25.14788	5.355820	21.29730

Source	DF	Anova SS	Mean Square	F Value	Pr > F
foreign	1	378.1535154	378.1535154	13.18	0.0005

OLS Regression

SAS command:

```
proc reg data = in.auto2;
MODEL mpg = weight;
run;
```

SAS output:

The REG Procedure

Model: MODEL1
Dependent Variable: mpg
Number of Observations Read 74

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1591.99020	1591.99020	134.62	<.0001
Error	72	851.46926	11.82596		
Corrected Total	73	2443.45946			

Root MSE	3.43889	R-Square	0.6515
Dependent Mean	21.29730	Adj R-Sq	0.6467
Coeff Var	16.14707		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	39.44028	1.61400	24.44	<.0001
weight	1	-0.00601	0.00051788	-11.60	<.0001

Inference Analysis (Categorical Dependent Variables)

- Chi-square test
- Testing proportions
- Logistic regression
- Ordered logistic regression
- Multinomial logistic regression

Chi-square Test

SAS command:

```
PROC FREQ DATA=in.auto;
TABLES  foreign*rep78 / CHISQ ;
RUN;
```

SAS outcome:

The FREQ Procedure

Table of foreign by rep78

foreign	rep78					
Frequency						
Percent						
Row Pct						
Col Pct	1	2	3	4	5	Total
Domestic	2	8	27	9	2	48
	2.90	11.59	39.13	13.04	2.90	69.57
	4.17	16.67	56.25	18.75	4.17	
	100.00	100.00	90.00	50.00	18.18	
Foreign	0	0	3	9	9	21
	0.00	0.00	4.35	13.04	13.04	30.43
	0.00	0.00	14.29	42.86	42.86	
	0.00	0.00	10.00	50.00	81.82	
Total	2	8	30	18	11	69
	2.90	11.59	43.48	26.09	15.94	100.00

Frequency Missing = 5

Chi-square Test (Continued)

Statistics for Table of foreign by rep78

Statistic	DF	Value	Prob
Chi-Square	4	27.2640	<.0001
Likelihood Ratio Chi-Square	4	29.9121	<.0001
Mantel-Haenszel Chi-Square	1	23.8506	<.0001
Phi Coefficient		0.6286	
Contingency Coefficient		0.5322	
Cramer's V		0.6286	

WARNING: 40% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Effective Sample Size = 69
Frequency Missing = 5

Testing Proportions

```
proc freq data = in.auto;
tables foreign / binomial(p=.5);
exact binomial;
run;
```

SAS outcome

The FREQ Procedure

foreign	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Domestic	52	70.27	52	70.27
Foreign	22	29.73	74	100.00

Binomial Proportion for foreign = Domestic

Proportion (P)	0.7027
ASE	0.0531
95% Lower Conf Limit	0.5986
95% Upper Conf Limit	0.8068
Exact Conf Limits	
95% Lower Conf Limit	0.5852
95% Upper Conf Limit	0.8034

Test of H0: Proportion = 0.5

ASE under H0	0.0581
Z	3.4874
One-sided Pr > Z	0.0002
Two-sided Pr > |Z|	0.0005
Exact Test	
One-sided Pr >= P	3.214E-04
Two-sided = 2 * One-sided	6.428E-04

Sample Size = 74

BGSU Center for Family and Demographic Research

Logistic Regression

SAS command:

```
data in.auto3;
set in.auto;
if price >=10000 then r_price =1;
else if price <10000 then r_price =0;
run;

proc freq data in.auto3;
tables price*r_price;
run;

proc logistic data= in.auto3 descending;
 model r_price =  mpg rep78;
run;
```

SAS output:

The SAS System 13:30 Thursday, February 3, 2011 133

The LOGISTIC Procedure

NOTE: 5 observations were deleted due to missing values for the response or explanatory variables.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.5212	2	<.0001
Score	11.8495	2	0.0027
Wald	9.5668	2	0.0084

The SAS System 13:30 Thursday, February 3, 2011 134

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	9.6540	3.7191	6.7381	0.0094
mpg	1	-0.6179	0.2003	9.5137	0.0020
rep78	1	-0.2054	0.5039	0.1661	0.6836

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
mpg	0.539	0.364	0.798
rep78	0.814	0.303	2.186

Association of Predicted Probabilities and Observed Responses

Percent Concordant	90.4	Somers' D	0.817
Percent Discordant	8.7	Gamma	0.824
Percent Tied	0.9	Tau-a	0.188
Pairs	540	c	0.908

Ordered Logistic Regression

SAS command:

```
proc logistic data = in.auto descending;
model rep78 = mpg rep78;
run;
```

SAS output:

The LOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	195.384	12.104
SC	204.321	25.508
-2 Log L	187.384	0.104

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	187.2803	2	<.0001
Score	67.2508	2	<.0001
Wald	2.6058	2	0.2717

SAS output (continued):

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 5	1	-69.9116	43.9344	2.5321	0.1115
Intercept 4	1	-54.9250	36.8673	2.2195	0.1363
Intercept 3	1	-38.7175	30.0141	1.6640	0.1971
Intercept 2	1	-23.9620	26.8451	0.7967	0.3721
mpg	1	0.00643	0.9332	0.0000	0.9945
rep78	1	15.5802	9.8336	2.5102	0.1131

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
mpg	1.006	0.162	6.268
rep78	>999.999	0.025	>999.999

Multinomial Logistic Regression

SAS command:

```
proc catmod data = in.sysdsn1;
  direct age male nonwhite;
  response logits;
  model insure = age male nonwhite;
run;
```

SAS output:

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	2	13.64	0.0011
age	2	3.38	0.1847
male	2	8.52	0.0141
nonwhite	2	11.15	0.0038
Likelihood Ratio	1E3	1088.39	0.9967

Analysis of Maximum Likelihood Estimates

Parameter	Function Number	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	1.8112	0.5349	11.47	0.0007
	2	1.9679	0.5369	13.43	0.0002
age	1	0.00584	0.0114	0.26	0.6087
	2	-0.00535	0.0115	0.22	0.6418
male	1	-0.5102	0.3640	1.97	0.1610
	2	0.0638	0.3591	0.03	0.8591
nonwhite	1	-0.4333	0.4106	1.11	0.2913
	2	0.2980	0.4004	0.55	0.4568

Conclusion

- Among different types of reliability and validity, only Inter-item Reliability and Construct validity can be directly tested without using additional data.
- Confirmatory factor analysis is need for truly testing construct validity. SAS can perform this type of analysis by using Proc Calis statement.
- The measurement of the dependent variables contributes to what analyses to be conducted.
- Try to understand what different options in the SAS commands.
- Other useful resources for learning using SAS.
  - http://www.ats.ucla.edu/stat/sas/
  - http://help.pop.psu.edu/help-by-software-package/sas/
  - http://www.stat.ncsu.edu/sas/tutor/