Final semester project

Lu Wang

Background

In this research, we will investigate the diagnose diseases using immunosignature, a customized non-natural peptide array development in the lab I am working at. On the peptide array, there are 330,000 non-natural peptides generated using program aimed at diversifying the peptide space coverage. And the way the technology works is to get a drop of blood from an individual and put it onto the array. The antibodies in the blood will then bind onto the array based on antibody-specific binding. And certain method can be used to illuminate the binding result and turn into intensity value. By investigating the binding result, we will be able to tell if the individual is healthy, or ill, and which disease does he/she have. And since the immune system is too complicated and we have too many features (each peptide serves as a feature), peptide selection is needed before the real diagnosis. In this research, we will test if a feature selection method is appropriate to select the features needed for diagnosing diseases.

Question to be investigated

In this research we want to test if using two tale T-Test comparing healthy samples and disease samples is a good feature selection method. Because by selecting peptides that are significantly different, it can either be the result of disease specific factor, or it can be the result of general inflammation that occurs in any disease. Which means in this research, we will test if there is unique factor underlying each set of feature selected from T-Test or there is a common factor underlying all of them.

Data collection

To test this, we will use three groups of samples, healthy people, dengue patients and malaria patients. Peripheral blood will be drawn from each individual. Same amount of blood will be put onto the array. Wait certain amount of time for the antibodies in the blood to bind to specific peptides and then wash the blood away. What are left are only the binded antibodies. Using secondary antibody attached with fluorescence, we can quantify the amount of antibody binded at each location. Signal intensity is gained for each peptide and used as a feature. So at last, for each sample, we will have a list of 330,000 features each with an intensity value. Then we will do a two tale T-Test between dengue and healthy, malaria and healthy samples. And select the top three features in each comparison based on p-value. Each feature is a variable, so

we will have ten variables in total. The dataset will use will consist of the ten variables and the observations will consist of all the samples from healthy, dengue and malaria patients.

Models to be tested

In this research, we will test two models.

<u>Two factor model</u>

The first model will be that each set of peptides selected from the same comparison group will have unique underlying factor. And because we want the factor to be disease specific, we need to eliminate the covariance of the underlying variables. The path model is shown below.
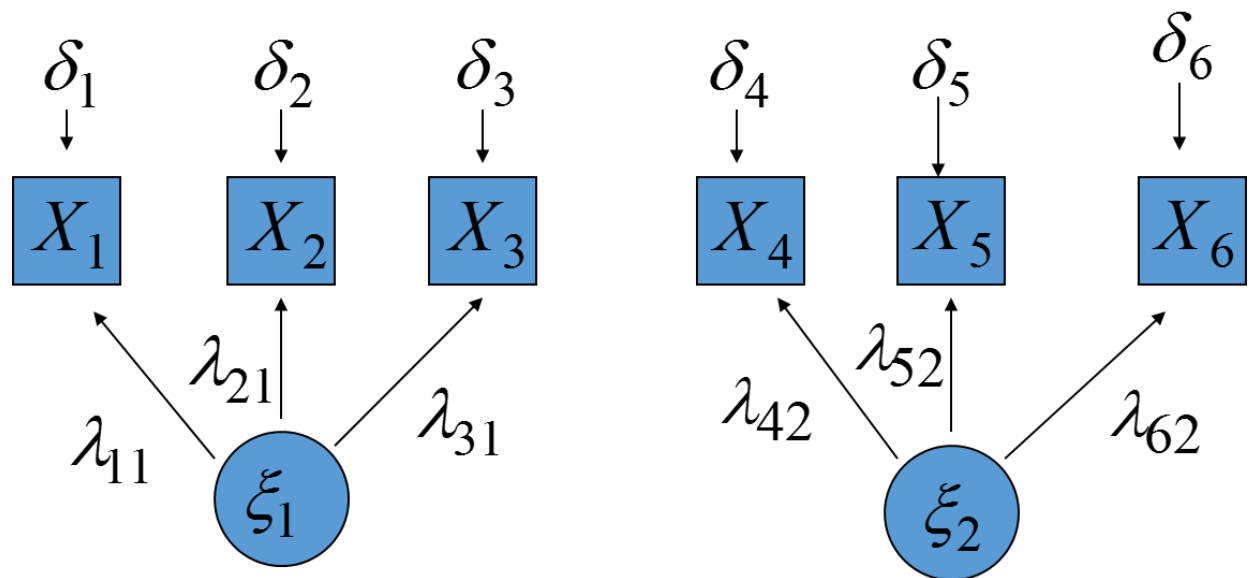
If this model holds, then we can conclude that the selection method is applicable to select disease specific peptides. And further analysis can be done using the peptides selected in this way.
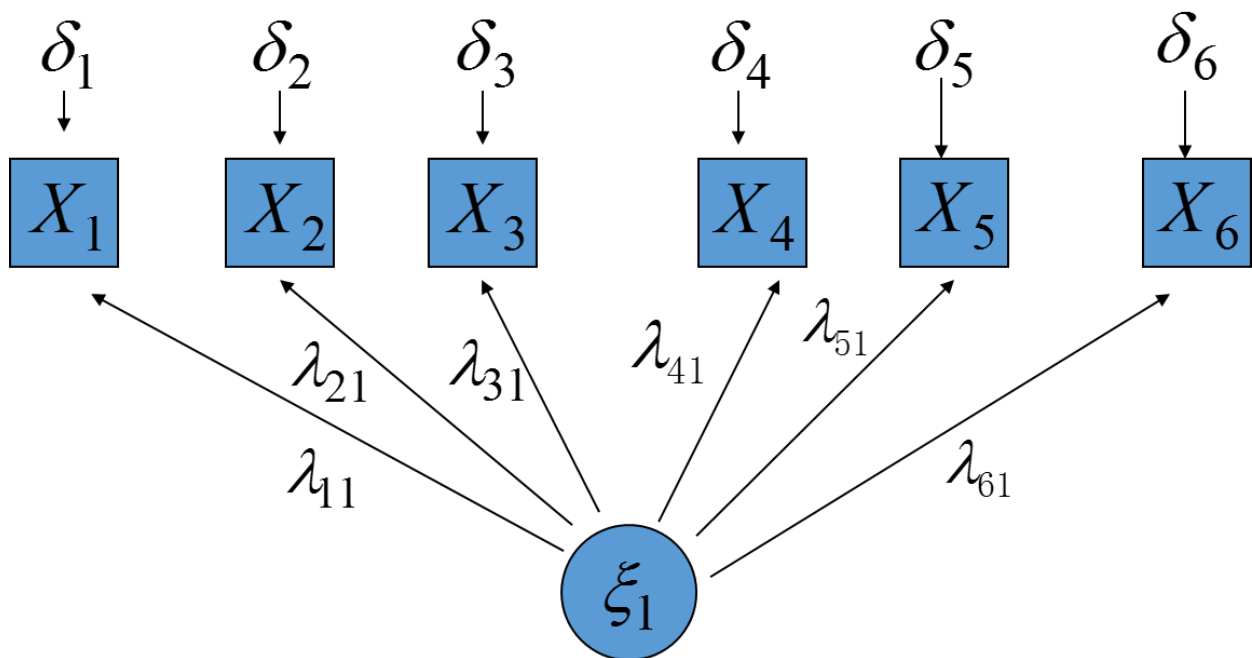
Concerning identification, according to three-indicator rule

$$\Lambda = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \qquad \Phi = \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix}$$

All the conditions hold, as a result, this model is identified.

One factor model

The second model I will test here is that there is one common factor loads onto all the variables. The path model is shown below.

If this model holds, then it means that there is one common factor underlying all the peptides selected from different comparison, which indicates this T-Test selection method is not a good

way to find disease specific peptides. Others methods are needed to find disease specific signatures.

Concerning the identification, according to the three-indicator rule, this model is also identified.

Test of model fit

<u>Two factor model</u>

To evaluate the fit for the two factor model. I used the following model script

  MODEL: F1 BY x1 x2 x3;

                F2 by x4 x5 x6;

                F1 with F2@0;

When we look at the global fit indices below

```
Chi-Square Test of Model Fit

          Value                           117.100
          Degrees of Freedom                    9
          P-Value                          0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                         0.511
          90 Percent C.I.                  0.431  0.595
          Probability RMSEA <= .05         0.000

CFI/TLI

          CFI                              0.682
          TLI                              0.469

Chi-Square Test of Model Fit for the Baseline Model

          Value                           354.482
          Degrees of Freedom                   15
          P-Value                          0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                            0.478
```

The chi-square value is 117, not too bad. However, when we look at the other indices, the fitting seems to be poor. RMSEA has a 90% C.I. of (0.431, 0.595), indicating bad fit. CFI value is 0.682, also means bad fit. And SRMR is 0.478, far larger than 0.05, which also means bad fit. So overall, all the global fit indices indicates the model fits poorly, which means the two factor model is not an appropriate model for the data.

When look into the local fit information to find the aspect responsible for the bad fit.

```
Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr
                 X1              X2              X3              X4              X5

                _____      _____      _____      _____      _____
     X1             0.005
     X2             0.002          -0.002
     X3             0.003           0.002           0.002
     X4             4.222           4.245           4.414           0.000
     X5             4.215           4.452           4.342           0.000           0.003
     X6             4.090           4.450           4.461           0.000           0.003


         Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr
                 X6

                _____
     X6             0.005
```

The z-score for the covariance between x1-x3 and x4-x6 are all over 4, indicates the residual covariance are too big. That should be the part responsible for the lack of fit.

And from the modification index output below

```
Minimum M.I. value for printing the modification index      0.000

                          M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.

BY Statements

F1      BY X4             1.781    0.050    0.144       0.099
F1      BY X5             0.118    0.023    0.065       0.022
F1      BY X6             3.553    0.116    0.331       0.162
F2      BY X1             0.343    0.113    0.147       0.045
F2      BY X2             3.037    0.917    1.188       0.133
F2      BY X3             0.784    0.297    0.385       0.059

ON/BY Statements

F1      ON F2      /
F2      BY F1             37.494   2.079    0.943       0.943
F2      ON F1      /
F1      BY F2             37.494   0.428    0.943       0.943

WITH Statements

X4      WITH X1           0.167    0.072    0.072       0.072
X4      WITH X2           3.242   -0.860   -0.860      -0.317
X4      WITH X3           3.034    0.531    0.531       0.406
X5      WITH X1           0.528    0.224    0.224       0.252
X5      WITH X2           3.586    1.591    1.591       0.663
X5      WITH X3           3.844   -1.051   -1.051      -0.909
X6      WITH X1           1.604   -0.362   -0.362      -0.216
X6      WITH X2           0.132    0.282    0.282       0.062
X6      WITH X3           2.440    0.776    0.776       0.356
F2      WITH F1           37.494   3.486    0.943       0.943
```

By allowing F1 and F2 to correlate, the model will improve a lot, with a M.I. value of 37.494 and an E.P.C value of 3.486, both significantly larger than to loose other constraints.

So when taking all the above discussions into consideration, the two factor model is a bad fit. And the bad fit is caused by constraining F1 not to correlate with F2 because the goal is to find disease specific peptides. And the constraint of F1 not correlating with F2 caused the residual covariance between x1-x3 and x4-x6 to be larger than normal, caused the bad fit.

One factor model

For the one factor model, the following model script in Mplus is used

MODEL: F1 BY x1 x2 x3 x4 x5 x6;

The global fit indices is followed:

```
Chi-Square Test of Model Fit

        Value                          22.817
        Degrees of Freedom                  9
        P-Value                        0.0066

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                       0.183
        90 Percent C.I.                0.091  0.278
        Probability RMSEA <= .05       0.015

CFI/TLI

        CFI                            0.959
        TLI                            0.932

Chi-Square Test of Model Fit for the Baseline Model

        Value                         354.482
        Degrees of Freedom                 15
        P-Value                        0.0000

SRMR (Standardized Root Mean Square Residual)

        Value                          0.022
```

The chi-square value is 22.817, significantly smaller than the two factor model, and the p-value=0.0066, larger than the former model too, providing the first set of evidence that this model is a better fit than the last one. The RMSEA has a 90% C.I. of (0.091, 0.278), still larger than 0.05 to be compared a good fit. But the improvement is significant compared with the last model. And the CFI is 0.959, indicating good fit. And SRMR value is 0.022, also indicates good fit. So all the global indices indicates the model is a good fit except the RMSEA indicates only a fair fit.

Up to this points, we can actually conclude that the one factor model is better fit and more logical than the two factors model.

And when looking at the local fit indices, we can see from the below output that all the residuals are fairly fitted

```
        Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr
            X1              X2              X3              X4              X5

            _____        _____        _____        _____        _____
X1          999.000
X2          -2.892          999.000
X3           0.780          999.000         999.000
X4           0.442          999.000           0.565           0.000
X5          -0.604            0.659         999.000           1.042           0.000
X6          999.000            0.928           0.793         999.000          -1.089



        Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr
            X6

            _____
X6            0.000
```

And then we can look at the modification indices to see how we can improve the model.

```
                    M.I.      E.P.C.    Std E.P.C.   StdYX E.P.C.

WITH Statements

X2      WITH X1     0.439     -0.656     -0.656      -0.121
X3      WITH X1     2.046      0.982      0.982       0.274
X3      WITH X2     1.531     -2.077     -2.077      -0.269
X4      WITH X1     0.310      0.099      0.099       0.096
X4      WITH X2     1.579     -0.526     -0.526      -0.237
X4      WITH X3     0.895      0.276      0.276       0.189
X5      WITH X1     0.151     -0.130     -0.130      -0.070
X5      WITH X2     1.432      0.963      0.963       0.240
X5      WITH X3     5.629     -1.344     -1.344      -0.511
X5      WITH X4     5.168      0.322      0.322       0.426
X6      WITH X1     1.181     -0.266     -0.266      -0.189
X6      WITH X2     3.802      1.130      1.130       0.372
X6      WITH X3     2.761      0.673      0.673       0.337
X6      WITH X4     7.292     -0.278     -0.278      -0.485
X6      WITH X5     0.200     -0.088     -0.088      -0.085
```

None of the M.I value is as significant as in the last model. However, if we use a common value of larger than 4 as a cutoff for needed modification, we need to lose the constraints between X3 with X5, X3 with X6 and X4 with X5. However, in factor analysis, that is usually not used, but can be considered as a factor.

And the parameter estimates for this model is listed as below

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F1	BY				
	X1	1.000	0.000	999.000	999.000
	X2	2.938	0.313	9.396	0.000
	X3	2.190	0.222	9.879	0.000
	X4	0.465	0.053	8.842	0.000
	X5	0.986	0.105	9.360	0.000
	X6	0.658	0.074	8.901	0.000
Intercepts					
	X1	4.230	0.476	8.896	0.000
	X2	14.652	1.317	11.121	0.000
	X3	9.072	0.966	9.394	0.000
	X4	2.250	0.215	10.473	0.000
	X5	4.560	0.443	10.294	0.000
	X6	3.095	0.302	10.247	0.000
Variances					
	F1	7.885	2.122	3.716	0.000
Residual Variances					
	X1	2.518	0.590	4.270	0.000
	X2	11.781	3.121	3.774	0.000
	X3	5.079	1.479	3.434	0.001
	X4	0.419	0.103	4.051	0.000
	X5	1.362	0.364	3.741	0.000
	X6	0.785	0.196	4.014	0.000

The loadings, intercepts, and residual variances for x1-x3 are all larger than x4-x6, probably because the raw data are larger. And the p-value for all parameters are significant, indicates the loading exist.

Discussion

Overall, after building the two models and analyzed the fitting result, I concluded that the one factor model is a much better fit than the two factor model. Since the two factors model implies that the T-Test can select disease specific peptides while the one factor model indicates that the T-Test can only found a general disease response that is not disease specific, we can actually conclude that this T-Test selection method is not a good way to find disease specific peptides.

The reason we want to use T-Test as a selection method is because it is widely used and our data meets all the assumption. And this method only need to take the disease of interest and normal and don't need to care of other disease group samples. However, we are worried that the most significant peptides found in the T-Test is not the disease specific ones, but instead, they are telling information that the person is ill or not. So if that is the case, there will only be one factor underlying the selected peptides across several disease-normal comparisons, which means the one factor model. And if the selected ones are really disease specific, we can find a unique factor underlying the peptides selected each comparison, which in this case I only used two disease groups to start with.

And using the peptides selected from dengue-healthy and malaria-healthy comparison to fit the models, we have found that there is actually one common factor underlying all the peptides from both groups. This is actually a disappointing result because this indicates that the most significantly different peptides are not even close to disease specific. And if we want to really find disease specific peptides for dengue and malaria, maybe further selection is needed, like performing a T-Test between dengue and malaria. If that is the case, when it will make the future analysis very difficult. Because if by comparing healthy and specific disease, we can find the specific peptides, then for any disease, we only need to reference the healthy samples. So the comparison needed to differentiate diseases increase linearly with the number of diseases. However, since the find found that this method could not find the needed peptides, and disease-disease comparison is possibly needed, then the number of comparisons needed increase quadratically with the number of diseases needed to differentiate.

And a by-product of the modelling result is the fact that the most significant peptides are not disease specific. Because from prior knowledge, most academic researchers would assume when you get ill, the dominant antibodies in your blood will be the ones against the disease directly, which would make them easily stand out from the T-Test selection. However, from my result, that seems not to be the case. The most dominate antibodies reflected through the peptides are the ones that indicates a general response to illness while not disease specific. The disease specific information seems to be deeply imbedded in the immune system than previously thought. This is actually a first tentative experiment for an ongoing research I am currently working on in the lab, which is to redefine the way that antibody immune system works.

Mplus VERSION 7
MUTHEN & MUTHEN
05/05/2014   2:18 PM

INPUT INSTRUCTIONS

 TITLE: Lu Wang final project one common factor model
 DATA: FILE = "F:\Dropbox\class\Spring 2014\PSY533\final\data.txt";
 VARIABLE: NAMES ARE x1-x6;
           USEVARIABLES ARE x1-x6;
 ANALYSIS: TYPE = GENERAL;
       ITERATIONS=3000;
       ESTIMATOR=ML;
 MODEL: F1 BY x1 x2 x3 x4 x5 x6;
 OUTPUT: sampstat standardized residual mod(0);

INPUT READING TERMINATED NORMALLY

Lu Wang final project one common factor model

SUMMARY OF ANALYSIS

Number of groups   1
Number of observations   46

Number of dependent variables   6
Number of independent variables   0
Number of continuous latent variables   1

Observed dependent variables

 Continuous
  X1   X2   X3   X4   X5   X6

Continuous latent variables
  F1

Estimator   ML
Information matrix   OBSERVED
Maximum number of iterations   3000
Convergence criterion   0.500D-04
Maximum number of steepest descent iterations   20

Input data file(s)
 F:\Dropbox\class\Spring 2014\PSY533\final\data.txt

Input data format FREE

SAMPLE STATISTICS

Means

	X1	X2	X3	X4	X5
	_____	_____	_____	_____	_____
1	4.230	14.652	9.072	2.250	4.560

Means

	X6
	_____
1	3.095

Covariances

	X1	X2	X3	X4	X5
	_____	_____	_____	_____	_____
X1	10.403				
X2	22.735	79.847			
X3	17.848	49.780	42.900		
X4	3.739	10.452	8.181	2.124	
X5	7.687	23.357	16.395	3.814	9.025
X6	4.996	15.915	11.715	2.225	5.060

Covariances

	X6
	_____
X6	4.197

Correlations

	X1	X2	X3	X4	X5
	_____	_____	_____	_____	_____
X1	1.000				
X2	0.789	1.000			
X3	0.845	0.851	1.000		
X4	0.795	0.803	0.857	1.000	
X5	0.793	0.870	0.833	0.871	1.000
X6	0.756	0.869	0.873	0.745	0.822

Correlations

	X6
	_____
X6	1.000

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION

Number of Free Parameters   18

Loglikelihood

    H0 Value   -567.775
    H1 Value   -556.367

Information Criteria

    Akaike (AIC)   1171.550
    Bayesian (BIC)   1204.466
    Sample-Size Adjusted BIC   1148.027
     (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

    Value   22.817
    Degrees of Freedom   9
    P-Value   0.0066

RMSEA (Root Mean Square Error Of Approximation)

    Estimate   0.183
    90 Percent C.I.   0.091 0.278
    Probability RMSEA <= .05   0.015

CFI/TLI

    CFI   0.959
    TLI   0.932

Chi-Square Test of Model Fit for the Baseline Model

    Value   354.482
    Degrees of Freedom   15
    P-Value   0.0000

SRMR (Standardized Root Mean Square Residual)

    Value   0.022

MODEL RESULTS

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value

F1   BY

	Estimate	S.E.	Est./S.E.	P-Value
X1	1.000	0.000	999.000	999.000
X2	2.938	0.313	9.396	0.000
X3	2.190	0.222	9.879	0.000
X4	0.465	0.053	8.842	0.000
X5	0.986	0.105	9.360	0.000
X6	0.658	0.074	8.901	0.000

Intercepts

	Estimate	S.E.	Est./S.E.	P-Value
X1	4.230	0.476	8.896	0.000
X2	14.652	1.317	11.121	0.000
X3	9.072	0.966	9.394	0.000
X4	2.250	0.215	10.473	0.000
X5	4.560	0.443	10.294	0.000
X6	3.095	0.302	10.247	0.000

Variances

	Estimate	S.E.	Est./S.E.	P-Value
F1	7.885	2.122	3.716	0.000

Residual Variances

	Estimate	S.E.	Est./S.E.	P-Value
X1	2.518	0.590	4.270	0.000
X2	11.781	3.121	3.774	0.000
X3	5.079	1.479	3.434	0.001
X4	0.419	0.103	4.051	0.000
X5	1.362	0.364	3.741	0.000
X6	0.785	0.196	4.014	0.000

STANDARDIZED MODEL RESULTS

STDYX Standardization

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value

F1   BY

	Estimate	S.E.	Est./S.E.	P-Value
X1	0.871	0.039	22.529	0.000
X2	0.923	0.025	36.388	0.000
X3	0.939	0.022	43.358	0.000
X4	0.896	0.032	27.578	0.000
X5	0.921	0.026	35.265	0.000
X6	0.902	0.031	29.077	0.000

Intercepts

	Estimate	S.E.	Est./S.E.	P-Value
X1	1.312	0.201	6.522	0.000
X2	1.640	0.226	7.263	0.000
X3	1.385	0.206	6.711	0.000
X4	1.544	0.218	7.073	0.000
X5	1.518	0.216	7.018	0.000

	Estimate	S.E.	Est./S.E.	P-Value
X6	1.511	0.216	7.003	0.000

Variances

	Estimate	S.E.	Est./S.E.	P-Value
F1	1.000	0.000	999.000	999.000

Residual Variances

	Estimate	S.E.	Est./S.E.	P-Value
X1	0.242	0.067	3.598	0.000
X2	0.148	0.047	3.149	0.002
X3	0.118	0.041	2.911	0.004
X4	0.197	0.058	3.385	0.001
X5	0.151	0.048	3.134	0.002
X6	0.187	0.056	3.347	0.001

STDY Standardization

	Estimate	S.E.	Two-Tailed Est./S.E.	P-Value

F1   BY

	Estimate	S.E.	Est./S.E.	P-Value
X1	0.871	0.039	22.529	0.000
X2	0.923	0.025	36.388	0.000
X3	0.939	0.022	43.358	0.000
X4	0.896	0.032	27.578	0.000
X5	0.921	0.026	35.265	0.000
X6	0.902	0.031	29.077	0.000

Intercepts

	Estimate	S.E.	Est./S.E.	P-Value
X1	1.312	0.201	6.522	0.000
X2	1.640	0.226	7.263	0.000
X3	1.385	0.206	6.711	0.000
X4	1.544	0.218	7.073	0.000
X5	1.518	0.216	7.018	0.000
X6	1.511	0.216	7.003	0.000

Variances

	Estimate	S.E.	Est./S.E.	P-Value
F1	1.000	0.000	999.000	999.000

Residual Variances

	Estimate	S.E.	Est./S.E.	P-Value
X1	0.242	0.067	3.598	0.000
X2	0.148	0.047	3.149	0.002
X3	0.118	0.041	2.911	0.004
X4	0.197	0.058	3.385	0.001
X5	0.151	0.048	3.134	0.002
X6	0.187	0.056	3.347	0.001

STD Standardization

	Estimate	S.E.	Two-Tailed Est./S.E.	P-Value

F1   BY

X1	2.808	0.378	7.432	0.000
X2	8.250	1.005	8.211	0.000
X3	6.150	0.727	8.459	0.000
X4	1.306	0.168	7.792	0.000
X5	2.768	0.339	8.177	0.000
X6	1.847	0.235	7.876	0.000

Intercepts

X1	4.230	0.476	8.896	0.000
X2	14.652	1.317	11.121	0.000
X3	9.072	0.966	9.394	0.000
X4	2.250	0.215	10.473	0.000
X5	4.560	0.443	10.294	0.000
X6	3.095	0.302	10.247	0.000

Variances

F1	1.000	0.000	999.000	999.000

Residual Variances

X1	2.518	0.590	4.270	0.000
X2	11.781	3.121	3.774	0.000
X3	5.079	1.479	3.434	0.001
X4	0.419	0.103	4.051	0.000
X5	1.362	0.364	3.741	0.000
X6	0.785	0.196	4.014	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
X1	0.758	0.067	11.264	0.000
X2	0.852	0.047	18.194	0.000
X3	0.882	0.041	21.679	0.000
X4	0.803	0.058	13.789	0.000
X5	0.849	0.048	17.632	0.000
X6	0.813	0.056	14.539	0.000

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix   0.338E-02
   (ratio of smallest to largest eigenvalue)

RESIDUAL OUTPUT

ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED)

Model Estimated Means/Intercepts/Thresholds

```
          X1         X2         X3         X4         X5
        _____   _____   _____   _____   _____
   1      4.230     14.652      9.072      2.250      4.560


    Model Estimated Means/Intercepts/Thresholds
      X6
        _____
   1      3.095


    Residuals for Means/Intercepts/Thresholds
      X1         X2         X3         X4         X5
    _____   _____   _____   _____   _____
   1    0.000      0.000      0.000      0.000      0.000


    Residuals for Means/Intercepts/Thresholds
      X6
        _____
   1      0.000


    Standardized Residuals (z-scores) for Means/Intercepts/Thresholds
      X1         X2         X3         X4         X5
    _____   _____   _____   _____   _____
   1    0.000      0.000      0.000      0.000      0.000


    Standardized Residuals (z-scores) for Means/Intercepts/Thresholds
      X6
        _____
   1      0.000


    Normalized Residuals for Means/Intercepts/Thresholds
      X1         X2         X3         X4         X5
    _____   _____   _____   _____   _____
   1    0.000      0.000      0.000      0.000      0.000


    Normalized Residuals for Means/Intercepts/Thresholds
      X6
        _____
   1      0.000


    Model Estimated Covariances/Correlations/Residual Correlations
      X1         X2         X3         X4         X5
    _____   _____   _____   _____   _____
X1    10.403
X2    23.167     79.847
X3    17.269     50.738     42.900
```

```
X4      3.667    10.774     8.031     2.124
X5      7.773    22.838    17.024     3.615     9.025
X6      5.187    15.239    11.359     2.412     5.113


        Model Estimated Covariances/Correlations/Residual Correlations
          X6


          _____
X6         4.197


        Residuals for Covariances/Correlations/Residual Correlations
          X1        X2        X3        X4        X5


        _____  _____  _____  _____  _____
X1        0.000
X2       -0.431     0.000
X3        0.579    -0.959     0.000
X4        0.072    -0.322     0.150     0.000
X5       -0.086     0.519    -0.630     0.199     0.000
X6       -0.191     0.676     0.356    -0.187    -0.053


        Residuals for Covariances/Correlations/Residual Correlations
          X6


          _____
X6         0.000


        Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr
          X1        X2        X3        X4        X5


        _____  _____  _____  _____  _____
X1      999.000
X2       -2.892   999.000
X3        0.780   999.000   999.000
X4        0.442   999.000     0.565     0.000
X5       -0.604     0.659   999.000     1.042     0.000
X6      999.000     0.928     0.793   999.000    -1.089


        Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr
          X6


          _____
X6         0.000


        Normalized Residuals for Covariances/Correlations/Residual Correlations
          X1        X2        X3        X4        X5


        _____  _____  _____  _____  _____
X1        0.000
X2       -0.080     0.000
X3        0.142    -0.085     0.000
X4        0.081    -0.131     0.081     0.000
X5       -0.047     0.099    -0.167     0.232     0.000
```

X6   -0.156   0.189   0.136   -0.341   -0.045

   Normalized Residuals for Covariances/Correlations/Residual Correlations
      X6

      _____
X6   0.000

MODEL MODIFICATION INDICES

NOTE: Modification indices for direct effects of observed dependent variables
regressed on covariates may not be included. To include these, request
MODINDICES (ALL).

Minimum M.I. value for printing the modification index   0.000

              M.I.   E.P.C. Std E.P.C. StdYX E.P.C.

WITH Statements

		M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
X2	WITH X1	0.439	-0.656	-0.656	-0.121
X3	WITH X1	2.046	0.982	0.982	0.274
X3	WITH X2	1.531	-2.077	-2.077	-0.269
X4	WITH X1	0.310	0.099	0.099	0.096
X4	WITH X2	1.579	-0.526	-0.526	-0.237
X4	WITH X3	0.895	0.276	0.276	0.189
X5	WITH X1	0.151	-0.130	-0.130	-0.070
X5	WITH X2	1.432	0.963	0.963	0.240
X5	WITH X3	5.629	-1.344	-1.344	-0.511
X5	WITH X4	5.168	0.322	0.322	0.426
X6	WITH X1	1.181	-0.266	-0.266	-0.189
X6	WITH X2	3.802	1.130	1.130	0.372
X6	WITH X3	2.761	0.673	0.673	0.337
X6	WITH X4	7.292	-0.278	-0.278	-0.485
X6	WITH X5	0.200	-0.088	-0.088	-0.085

Mplus VERSION 7
MUTHEN & MUTHEN
05/05/2014 11:40 AM

INPUT INSTRUCTIONS

 TITLE: Lu Wang final project two factor model
 DATA: FILE = "F:\Dropbox\class\Spring 2014\PSY533\final\data.txt";
 VARIABLE: NAMES ARE x1-x6;
          USEVARIABLES ARE x1-x6;
 ANALYSIS: TYPE = GENERAL;
      ITERATIONS=3000;
      ESTIMATOR=ML;
 MODEL: F1 BY x1 x2 x3;
       F2 by x4 x5 x6;
                  F1 with F2@0;
 OUTPUT: sampstat standardized residual mod(0);

INPUT READING TERMINATED NORMALLY

Lu Wang final project two factor model

SUMMARY OF ANALYSIS

Number of groups   1
Number of observations   46

Number of dependent variables   6
Number of independent variables   0
Number of continuous latent variables   2

Observed dependent variables

  Continuous
   X1   X2   X3   X4   X5   X6

Continuous latent variables
   F1   F2

Estimator   ML
Information matrix   OBSERVED
Maximum number of iterations   3000
Convergence criterion   0.500D-04
Maximum number of steepest descent iterations   20

Input data file(s)

F:\Dropbox\class\Spring 2014\PSY533\final\data.txt

Input data format FREE

SAMPLE STATISTICS

  SAMPLE STATISTICS

        Means
        X1   X2   X3   X4   X5

        _____   _____   _____   _____   _____
    1   4.230   14.652   9.072   2.250   4.560

        Means
         X6

        _____
    1   3.095

        Covariances
        X1   X2   X3   X4   X5

        _____   _____   _____   _____   _____
  X1   10.403
  X2   22.735   79.847
  X3   17.848   49.780   42.900
  X4   3.739   10.452   8.181   2.124
  X5   7.687   23.357   16.395   3.814   9.025
  X6   4.996   15.915   11.715   2.225   5.060

        Covariances
         X6

        _____
  X6   4.197

        Correlations
        X1   X2   X3   X4   X5

        _____   _____   _____   _____   _____
  X1   1.000
  X2   0.789   1.000
  X3   0.845   0.851   1.000
  X4   0.795   0.803   0.857   1.000
  X5   0.793   0.870   0.833   0.871   1.000
  X6   0.756   0.869   0.873   0.745   0.822

        Correlations
         X6

```
          _____
X6         1.000
```

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION

Number of Free Parameters   18

Loglikelihood

        H0 Value   -614.917
        H1 Value   -556.367

Information Criteria

        Akaike (AIC)   1265.834
        Bayesian (BIC)   1298.749
        Sample-Size Adjusted BIC   1242.310
          (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

        Value   117.100
        Degrees of Freedom   9
        P-Value   0.0000

RMSEA (Root Mean Square Error Of Approximation)

        Estimate   0.511
        90 Percent C.I.   0.431 0.595
        Probability RMSEA <= .05   0.000

CFI/TLI

        CFI   0.682
        TLI   0.469

Chi-Square Test of Model Fit for the Baseline Model

        Value   354.482
        Degrees of Freedom   15
        P-Value   0.0000

SRMR (Standardized Root Mean Square Residual)

        Value   0.478

MODEL RESULTS

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
F1	BY				
X1		1.000	0.000	999.000	999.000
X2		2.789	0.315	8.847	0.000
X3		2.190	0.222	9.884	0.000
F2	BY				
X4		1.000	0.000	999.000	999.000
X5		2.275	0.222	10.264	0.000
X6		1.327	0.166	8.013	0.000
F1	WITH				
F2		0.000	0.000	999.000	999.000
Intercepts					
X1		4.230	0.476	8.896	0.000
X2		14.652	1.317	11.121	0.000
X3		9.072	0.966	9.394	0.000
X4		2.250	0.215	10.473	0.000
X5		4.560	0.443	10.294	0.000
X6		3.095	0.302	10.247	0.000
Variances					
F1		8.152	2.160	3.775	0.000
F2		1.677	0.443	3.787	0.000
Residual Variances					
X1		2.251	0.632	3.562	0.000
X2		16.437	4.752	3.459	0.001
X3		3.821	2.179	1.753	0.080
X4		0.447	0.131	3.403	0.001
X5		0.350	0.485	0.722	0.470
X6		1.246	0.307	4.062	0.000

STANDARDIZED MODEL RESULTS

STDYX Standardization

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
F1	BY				
X1		0.885	0.039	22.513	0.000
X2		0.891	0.038	23.353	0.000
X3		0.954	0.028	34.042	0.000
F2	BY				

	Estimate	S.E.	Est./S.E.	P-Value
X4	0.889	0.040	22.454	0.000
X5	0.980	0.028	35.426	0.000
X6	0.839	0.050	16.939	0.000

F1   WITH

	Estimate	S.E.	Est./S.E.	P-Value
F2	0.000	0.000	999.000	999.000

Intercepts

	Estimate	S.E.	Est./S.E.	P-Value
X1	1.312	0.201	6.522	0.000
X2	1.640	0.226	7.263	0.000
X3	1.385	0.206	6.711	0.000
X4	1.544	0.218	7.073	0.000
X5	1.518	0.216	7.018	0.000
X6	1.511	0.216	7.003	0.000

Variances

	Estimate	S.E.	Est./S.E.	P-Value
F1	1.000	0.000	999.000	999.000
F2	1.000	0.000	999.000	999.000

Residual Variances

	Estimate	S.E.	Est./S.E.	P-Value
X1	0.216	0.070	3.109	0.002
X2	0.206	0.068	3.027	0.002
X3	0.089	0.054	1.664	0.096
X4	0.211	0.070	2.994	0.003
X5	0.039	0.054	0.715	0.475
X6	0.297	0.083	3.574	0.000

STDY Standardization

		Two-Tailed		
	Estimate	S.E.	Est./S.E.	P-Value

F1   BY

	Estimate	S.E.	Est./S.E.	P-Value
X1	0.885	0.039	22.513	0.000
X2	0.891	0.038	23.353	0.000
X3	0.954	0.028	34.042	0.000

F2   BY

	Estimate	S.E.	Est./S.E.	P-Value
X4	0.889	0.040	22.454	0.000
X5	0.980	0.028	35.426	0.000
X6	0.839	0.050	16.939	0.000

F1   WITH

	Estimate	S.E.	Est./S.E.	P-Value
F2	0.000	0.000	999.000	999.000

Intercepts

	Estimate	S.E.	Est./S.E.	P-Value
X1	1.312	0.201	6.522	0.000
X2	1.640	0.226	7.263	0.000
X3	1.385	0.206	6.711	0.000
X4	1.544	0.218	7.073	0.000
X5	1.518	0.216	7.018	0.000
X6	1.511	0.216	7.003	0.000

Variances

F1	1.000	0.000	999.000	999.000
F2	1.000	0.000	999.000	999.000

Residual Variances

X1	0.216	0.070	3.109	0.002
X2	0.206	0.068	3.027	0.002
X3	0.089	0.054	1.664	0.096
X4	0.211	0.070	2.994	0.003
X5	0.039	0.054	0.715	0.475
X6	0.297	0.083	3.574	0.000

STD Standardization

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
F1	BY				
X1		2.855	0.378	7.549	0.000
X2		7.963	1.044	7.630	0.000
X3		6.251	0.731	8.553	0.000
F2	BY				
X4		1.295	0.171	7.574	0.000
X5		2.945	0.329	8.941	0.000
X6		1.718	0.248	6.933	0.000
F1	WITH				
F2		0.000	0.000	999.000	999.000

Intercepts

X1	4.230	0.476	8.896	0.000
X2	14.652	1.317	11.121	0.000
X3	9.072	0.966	9.394	0.000
X4	2.250	0.215	10.473	0.000
X5	4.560	0.443	10.294	0.000
X6	3.095	0.302	10.247	0.000

Variances

F1	1.000	0.000	999.000	999.000
F2	1.000	0.000	999.000	999.000

Residual Variances

X1	2.251	0.632	3.562	0.000
X2	16.437	4.752	3.459	0.001
X3	3.821	2.179	1.753	0.080
X4	0.447	0.131	3.403	0.001
X5	0.350	0.485	0.722	0.470
X6	1.246	0.307	4.062	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
X1	0.784	0.070	11.257	0.000
X2	0.794	0.068	11.676	0.000
X3	0.911	0.054	17.021	0.000
X4	0.789	0.070	11.227	0.000
X5	0.961	0.054	17.713	0.000
X6	0.703	0.083	8.469	0.000

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix   0.298E-02
   (ratio of smallest to largest eigenvalue)

RESIDUAL OUTPUT

ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED)

Model Estimated Means/Intercepts/Thresholds

	X1	X2	X3	X4	X5
1	4.230	14.652	9.072	2.250	4.560

Model Estimated Means/Intercepts/Thresholds

	X6
1	3.095

Residuals for Means/Intercepts/Thresholds

	X1	X2	X3	X4	X5
1	0.000	0.000	0.000	0.000	0.000

Residuals for Means/Intercepts/Thresholds

	X6
1	0.000

Standardized Residuals (z-scores) for Means/Intercepts/Thresholds

	X1	X2	X3	X4	X5
1	0.000	0.000	0.000	0.000	0.000

Standardized Residuals (z-scores) for Means/Intercepts/Thresholds
          X6
       _____
  1   0.000

Normalized Residuals for Means/Intercepts/Thresholds
        X1   X2   X3   X4   X5
     _____ _____ _____ _____ _____ _____
  1   0.000   0.000   0.000   0.000   0.000

Normalized Residuals for Means/Intercepts/Thresholds
          X6
       _____
  1   0.000

Model Estimated Covariances/Correlations/Residual Correlations
        X1   X2   X3   X4   X5
     _____ _____ _____ _____ _____
  X1   10.403
  X2   22.735   79.847
  X3   17.848   49.780   42.900
  X4   0.000   0.000   0.000   2.124
  X5   0.000   0.000   0.000   3.814   9.025
  X6   0.000   0.000   0.000   2.225   5.060

Model Estimated Covariances/Correlations/Residual Correlations
          X6
       _____
  X6   4.197

Residuals for Covariances/Correlations/Residual Correlations
        X1   X2   X3   X4   X5
     _____ _____ _____ _____ _____
  X1   0.000
  X2   0.000   0.000
  X3   0.000   0.000   0.000
  X4   3.739   10.452   8.181   0.000
  X5   7.687   23.357   16.395   0.000   0.000
  X6   4.996   15.915   11.715   0.000   0.000

Residuals for Covariances/Correlations/Residual Correlations
          X6
       _____
  X6   0.000

Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr

	X1	X2	X3	X4	X5
X1	0.005				
X2	0.002	-0.002			
X3	0.003	0.002	0.002		
X4	4.222	4.245	4.414	0.000	
X5	4.215	4.452	4.342	0.000	0.003
X6	4.090	4.450	4.461	0.000	0.003

Standardized Residuals (z-scores) for Covariances/Correlations/Residual Corr

	X6
X6	0.005

Normalized Residuals for Covariances/Correlations/Residual Correlations

	X1	X2	X3	X4	X5
X1	0.000				
X2	0.000	0.000			
X3	0.000	0.000	0.000		
X4	4.222	4.245	4.414	0.000	
X5	4.215	4.452	4.342	0.000	0.000
X6	4.090	4.450	4.461	0.000	0.000

Normalized Residuals for Covariances/Correlations/Residual Correlations

	X6
X6	0.000

MODEL MODIFICATION INDICES

NOTE: Modification indices for direct effects of observed dependent variables
regressed on covariates may not be included. To include these, request
MODINDICES (ALL).

Minimum M.I. value for printing the modification index   0.000

		M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.

BY Statements

		M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
F1	BY X4	1.781	0.050	0.144	0.099
F1	BY X5	0.118	0.023	0.065	0.022
F1	BY X6	3.553	0.116	0.331	0.162
F2	BY X1	0.343	0.113	0.147	0.045
F2	BY X2	3.037	0.917	1.188	0.133
F2	BY X3	0.784	0.297	0.385	0.059

ON/BY Statements

F1	ON F2	/				
F2	BY F1		37.494	2.079	0.943	0.943
F2	ON F1	/				
F1	BY F2		37.494	0.428	0.943	0.943

WITH Statements

X4	WITH X1	0.167	0.072	0.072	0.072
X4	WITH X2	3.242	-0.860	-0.860	-0.317
X4	WITH X3	3.034	0.531	0.531	0.406
X5	WITH X1	0.528	0.224	0.224	0.252
X5	WITH X2	3.586	1.591	1.591	0.663
X5	WITH X3	3.844	-1.051	-1.051	-0.909
X6	WITH X1	1.604	-0.362	-0.362	-0.216
X6	WITH X2	0.132	0.282	0.282	0.062
X6	WITH X3	2.440	0.776	0.776	0.356
F2	WITH F1	37.494	3.486	0.943	0.943