FACTOR ANALYSIS ASSIGNMENT
By Daniel Lagos

The codebook for the 2009 California Health Interview Survey
The dataset for the 2009 CHIS - should be named adult2009 on your remote desktop
Base SAS and SAS/STAT software

SAS Enterprise Guide is optional

1. DATA QUALITY EVALUATION: Identify 10-20 variables in the adult2009 that you think represent 1-3 factors. Run a PROC UNIVARIATE, PROC MEANS and/or PROC FREQ to identify any problems with data such as excessive missing data, items not coded in order, extreme outliers. (2 points)

After conducting a keyword search in codebook, the following 29 variables were identified as candidates:

AB24 AB25 AB51 AB52 AB109 AB111 AB115 AB117 AB118 AB119 AB122 AB123 AB124 AF63 AF64 AF65 AF66 AF67 AF68 AB120 AB22 AB34 AB1 AB29 AB30 AB81 AB99 AH12 SRSEX

Keywords used for search: Health, Diabetes, Heart, and Feel

PROC COR was conducted with best=20 option to identify variables with the strongest correlations. This list was refined to the following 22 variables.

AB22 AB24 AB25 AB34 AB51 AB52 AB109 AB111 AB115 AB117 AB118 AB119 AB120 AB122 AB123 AB124 AF63 AF64 AF65 AF66 AF67 AF68

Findings after conducting PROC UNIVARIATE, PROC MEANS, and PROC FREQ:

- No missing values in selected variables. n = 47,614
- Surprisingly, PROC UNIVARIATE provided nothing that would indicate issues regarding extreme values or outliers.
- 20 of the 22 variables had values of -1, codebook indicates -1 = inapplicable.
- Of the 20 variables 1 of them (var AB120) had values of -1 and -2.
    - Codebook indicates that -1 = inapplicable, while -2 = proxy skipped.
    - 283 of the 47,614 observations contained AB120(-2).

2. FIX DATA QUALITY ISSUES Solve any data problems you found in Step #1. (1 point).

   Since none of the variables were coded as 0, the following was conducted to clean up the data, that is addressing coded values of -2 and -1:

   - All values of -1 were recoded as 0 – this is merely the researcher's preference.
   - Since AB120(-2) = 283 observations, that is 0.59% of the data, these observations were removed from dataset.

3. FACTOR ANALYSIS Perform at least one factor analysis of your data. Specify rotation, number of factors and (optionally) communality estimates. (3 points).

The following procedures were conducted.

- Initial factor analysis was conducted using 22 factors, that is every variable selected.
- Rotation = verimax, fuzz = 0.3 used on all factor analysis.
- Analysis of scree plot, Eigenvalue correlation matrix, and proportion of variance used to identify the number of factors. Criteria used:
  - Eigenvalue > 1.
  - Proportion of variance over 0.70
  - Initial indications suggested 3 factors, to be explained in more detail below.
- One more factor analysis was conducted using 3 factors.
  - Eigenvalue criteria: Eigenvalue > 1
  - Analysis of scree plot conducted
  - Proportion of variance analyzed.
  - Analysis of path diagram conducted.
  - Results discussed below.

4. DISCUSSION Explain your results. State the number of factors retained by the criterion used, amount of variance explained by those factors, variables with the highest loading on each factor after rotation. Interpret your factors, e.g., "The first factor, with substantial loadings of items measuring asthma symptoms in the past 12 months, use of daily medication and having an asthma episode in the last month appears to be a measure of chronic asthma symptoms. The second factor, with loadings of ER and hospital visits appears to be a measure of asthma severity." (4 points)

Findings after first factor analysis – 22 factors:

- Scree plot and Eigenvalue correlation matrix indicate that 3 factors exist with Eigenvalues of 1.

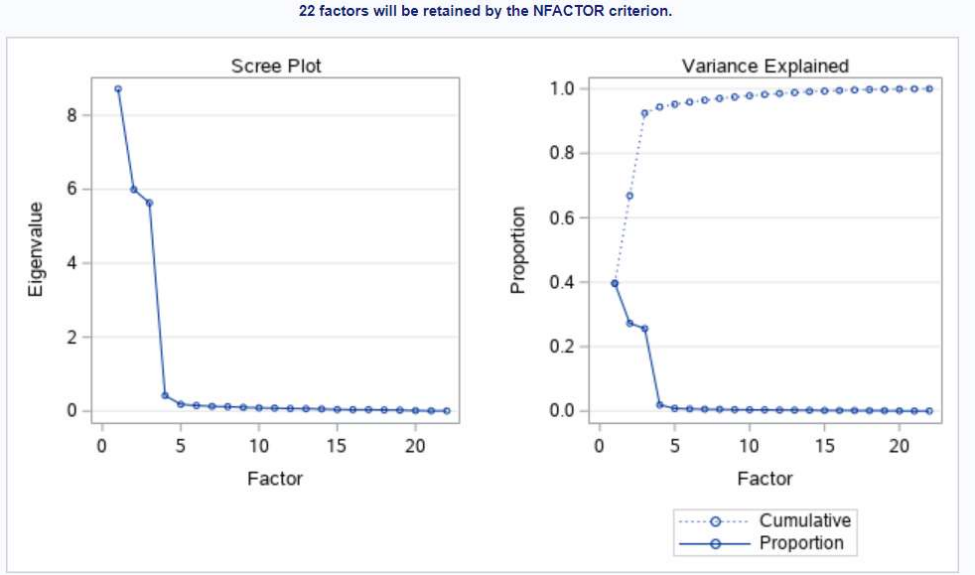
22 Factors

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 22 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	8.71731963	2.72392030	0.3962	0.3962
2	5.99339933	0.36133053	0.2724	0.6687
3	5.63206880	5.21173938	0.2560	0.9247
4	0.42032942	0.23807421	0.0191	0.9438
5	0.18225520	0.03174496	0.0083	0.9521
6	0.15051024	0.02193303	0.0068	0.9589
7	0.12857722	0.00862396	0.0058	0.9647
8	0.11995326	0.02095279	0.0055	0.9702
9	0.09900046	0.01180135	0.0045	0.9747
10	0.08719911	0.00567136	0.0040	0.9787
11	0.08152775	0.01261725	0.0037	0.9824
12	0.06891050	0.00593999	0.0031	0.9855
13	0.06297051	0.00565638	0.0029	0.9884
14	0.05731413	0.01503064	0.0026	0.9910
15	0.04228350	0.00392658	0.0019	0.9929
16	0.03835691	0.00029859	0.0017	0.9946
17	0.03805832	0.00731649	0.0017	0.9964
18	0.03074183	0.00652778	0.0014	0.9978
19	0.02421405	0.01126666	0.0011	0.9989
20	0.01294739	0.00570293	0.0006	0.9995
21	0.00724446	0.00242648	0.0003	0.9998
22	0.00481798		0.0002	1.0000

- First three factors contain eigenvalues greater than one. Note, lowest eigenvalue over 1 is 5.6.
- At the third factor 0.9247 of the variation is explained. This is also confirmed by Variance plot shown below.



- Scree plot indicates a steep drop in eigenvalue after the third factor.
- Variance plot indicates not much gain in variation as factors are increased after the third.
- Initial findings suggest 3 factors, therefore run factor analysis with 3 factors.

Findings after second factor analysis – 3 factors:

- Because may of the plots and diagrams are the same across all factor analysis, only those that have changed will be provided.
- Analysis of rotated factor pattern, chart seen below.

Rotated Factor Pattern				
		Factor1	Factor2	Factor3
AB22	DOCTOR EVER TOLD HAVE DIABETES	-0.93628	.	.
AB24	CURRENTLY TAKING INSULIN	0.97068	.	.
AB25	CURRENTLY TAKING DIABETIC PILLS TO LOWER BLOOD SUGAR	0.94537	.	.
AB34	DOCTOR EVER TOLD HAVE ANY KIND OF HEART DISEASE	.	-0.99291	.
AB51	TYPE I OR TYPE II DIABETES	0.97777	.	.
AB52	EVER TOLD HAVE HEART FAILURE/CONGESTIVE	.	0.97087	.
AB109	VISITED ER FOR DIABETES IN PAST 12 MOS	0.98852	.	.
AB111	ADMITTED TO HOSPITAL OVERNIGHT OR LONGER FOR DIABETES PAST 12 MOS	0.98981	.	.
AB115	VISITED ER FOR HEART DISEASE IN PAST 12 MOS	.	0.98398	.
AB117	ADMITTED TO HOSPITAL OVERNIGHT/LONGER FOR HEART DX PAST 12 MOS	.	0.98484	.
AB118	MEDICAL PROVIDERS DEVELOPED HEART DISEASE PLAN	.	0.93497	.
AB119	HAVE WRITTEN COPY OF HEART DISEASE CARE	.	0.83229	.
AB120	CONFIDENCE TO CONTROL AND MANAGE HEART DIEASE	.	0.91206	.
AB122	TAKE MEDICINE TO LOWER CHOLESTEROL	0.95278	.	.
AB123	TAKE ASPIRIN REGULARLY TO REDUCE HEART ATTACK RISK	0.95392	.	.
AB124	TAKE OTHER MEDICATIONS TO REDUCE HEART ATTACK RISK	0.96995	.	.
AF63	FEEL NERVOUS WORST MONTH	.	.	0.95796
AF64	FEEL HOPELESS WORST MONTH	.	.	0.97797
AF65	FEEL RESTLESS WORST MONTH	.	.	0.96247
AF66	FEEL DEPRESSED WORST MONTH	.	.	0.97682
AF67	FEEL EVERYTHING AN EFFORT WORST MONTH	.	.	0.96421
AF68	FEEL WORTHLESS WORST MONTH	.	.	0.97747
Values less than 0.3 are not printed.				

- Variables contain values of over 0.9 (in terms of absolute values) with one exception. The only exception contains a value of 0.83. Note that values less than 0.3 have been omitted for clarity of chart.
- Number of factors retained: 3
- Variance explained by factors:

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
8.4082025	6.2939898	5.6405954

- Interpretation of factors – based on rotated factor pattern:
  - Taking Medication: The first factor with substantial loadings of items measuring diabetes and heart disease, with an emphasis on whether the patient is taking medication to mitigate these diseases.
  - Receive Treatment: The second factor with substantial loading of items measure heart disease with an emphasis on the patient receiving treatment for heart disease.
  - Mental State: The third factor with substantial loading of items measuring the patient's feelings and mental state.
- Excel spreadsheet provided for clarity

	Values less than 0.3 are not printed.	Factor1	Factor2	Factor3
AF64	FEEL HOPELESS WORST MONTH	.	.	0.97797
AF68	FEEL WORTHLESS WORST MONTH	.	.	0.97747
AF66	FEEL DEPRESSED WORST MONTH	.	.	0.97682
AF67	FEEL EVERYTHING AN EFFORT WORST MONTH	.	.	0.96421
AF65	FEEL RESTLESS WORST MONTH	.	.	0.96247
AF63	FEEL NERVOUS WORST MONTH	.	.	0.95796
AB117	ADMITTED TO HOSPITAL OVERNIGHT/LONGER FOR HEART DX PAST 12 MOS	.	0.98484	.

AB115	VISITED ER FOR HEART DISEASE IN PAST 12 MOS	.	0.98398	.
AB52	EVER TOLD HAVE HEART FAILURE/CONGESTIVE	.	0.97087	.
AB118	MEDICAL PROVIDERS DEVELOPED HEART DISEASE PLAN	.	0.93497	.
AB120	CONFIDENCE TO CONTROL AND MANAGE HEART DIEASE	.	0.91206	.
AB119	HAVE WRITTEN COPY OF HEART DISEASE CARE	.	0.83229	.
AB34	DOCTOR EVER TOLD HAVE ANY KIND OF HEART DISEASE	.	-0.99291	.
AB111	ADMITTED TO HOSPITAL OVERNIGHT OR LONGER FOR DIABETES PAST 12 MOS	0.98981	.	.
AB109	VISITED ER FOR DIABETES IN PAST 12 MOS	0.98852	.	.
AB51	TYPE I OR TYPE II DIABETES	0.97777	.	.
AB24	CURRENTLY TAKING INSULIN	0.97068	.	.
AB124	TAKE OTHER MEDICATIONS TO REDUCE HEART ATTACK RISK	0.96995	.	.
AB123	TAKE ASPIRIN REGULARLY TO REDUCE HEART ATTACK RISK	0.95392	.	.
AB122	TAKE MEDICINE TO LOWER CHOLESTEROL	0.95278	.	.
AB25	CURRENTLY TAKING DIABETIC PILLS TO LOWER BLOOD SUGAR	0.94537	.	.
AB22	DOCTOR EVER TOLD HAVE DIABETES	-0.93628	.	.