Using Factor Analysis and Quantile Regression in SAS to Investigate the Relationship Between Dietary Pattern and Children's Growth

Zhenzhen Zhang
School of Public Health
University of Michigan

Study Objective

- Nutrition data set collected from a study in Mexico City
- Investigate the relationship between children's BMI(body mass index) and dietary patterns
- 1062 subjects
- 32 food groups (high-fat dairy products, fruits, etc.)
- Food groups created from food questionnaire and taken as continuous variables

Analysis Plan

- Factor analysis to summarize the 36 food groups (e.g., factors that represent Mediterranean, Western, local diet, etc.)

- Using these dietary factors as covariates for subsequent regression to analyze the effect of dietary patterns on BMI

Factor Analysis

- $X_i - \mu_i = l_{i1}F_1 + l_{i2}F_2 + \cdots + l_{im}F_m + \varepsilon_i,$ $i = 1, \cdots, p.$
- $F_j =$ j-th unobserved factor
- $l_{i1} =$ loading of the i-th variable on the j-th factor

Food Groups	Factor 1	Factor 2	Factor 3
Green Vegetables	**0.633**	0.132	-0.150
Whole Grains	0.275	**0.415**	0.112
Refined Grains	0.020	**-0.526**	-0.018

Factor Analysis (cont.)

- $X_{p\times 1} - \mu_{p\times 1} = L_{p\times m} F_{m\times 1} + \varepsilon_{p\times 1}$.
  - F, ε independent
  - $E(F) = 0, Cov(F) = I$
  - $E(\varepsilon) = 0, Cov(\varepsilon) = \Psi$, Ψ is diagonal.
- $Cov(X) = S = LL' + \Psi$.
- Principal component method
  - Maximum likelihood method
- **proc factor** data=<-> method=principal scree;
  - var <->;
  - **run**;

Eigenvalues of the Correlation Matrix: Total = 32 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	2.92537367	1.29032866	0.0914	0.0914
2	1.63504500	0.00577704	0.0511	0.1425
3	1.62926796	0.25587643	0.0509	0.1934
4	1.37339154	0.00463010	0.0429	0.2363
5	1.36876144	0.07716569	0.0428	0.2791
6	1.29159575	0.07691878	0.0404	0.3195
7	1.21467696	0.04649159	0.0380	0.3574
8	1.16818537	0.04977351	0.0365	0.3939
9	1.11841187	0.03996491	0.0350	0.4289
10	1.07844696	0.00658473	0.0337	0.4626
11	1.07186223	0.01502792	0.0335	0.4961
12	1.05683431	0.05968511	0.0330	0.5291
13	0.99714920	0.04006253	0.0312	0.5603
14	0.95708667	0.02260401	0.0299	0.5902
15	0.93448266	0.02054759	0.0292	0.6194
16	0.91393507	0.02407577	0.0286	0.6480
17	0.88985930	0.02436324	0.0278	0.6758
18	0.86549606	0.01561143	0.0270	0.7028
19	0.84988463	0.03603550	0.0266	0.7294
20	0.81384913	0.00680099	0.0254	0.7548
21	0.80704813	0.04861071	0.0252	0.7800
22	0.75843743	0.00835604	0.0237	0.8037
23	0.75008139	0.00497569	0.0234	0.8272
24	0.74510570	0.03177851	0.0233	0.8504
25	0.71332719	0.02793731	0.0223	0.8727
26	0.68538988	0.01217668	0.0214	0.8942
27	0.67321320	0.00899511	0.0210	0.9152
28	0.66421809	0.05095990	0.0208	0.9360
29	0.61325819	0.01921837	0.0192	0.9551
30	0.59403981	0.10646954	0.0186	0.9737
31	0.48757027	0.13285530	0.0152	0.9889
32	0.35471497		0.0111	1.0000

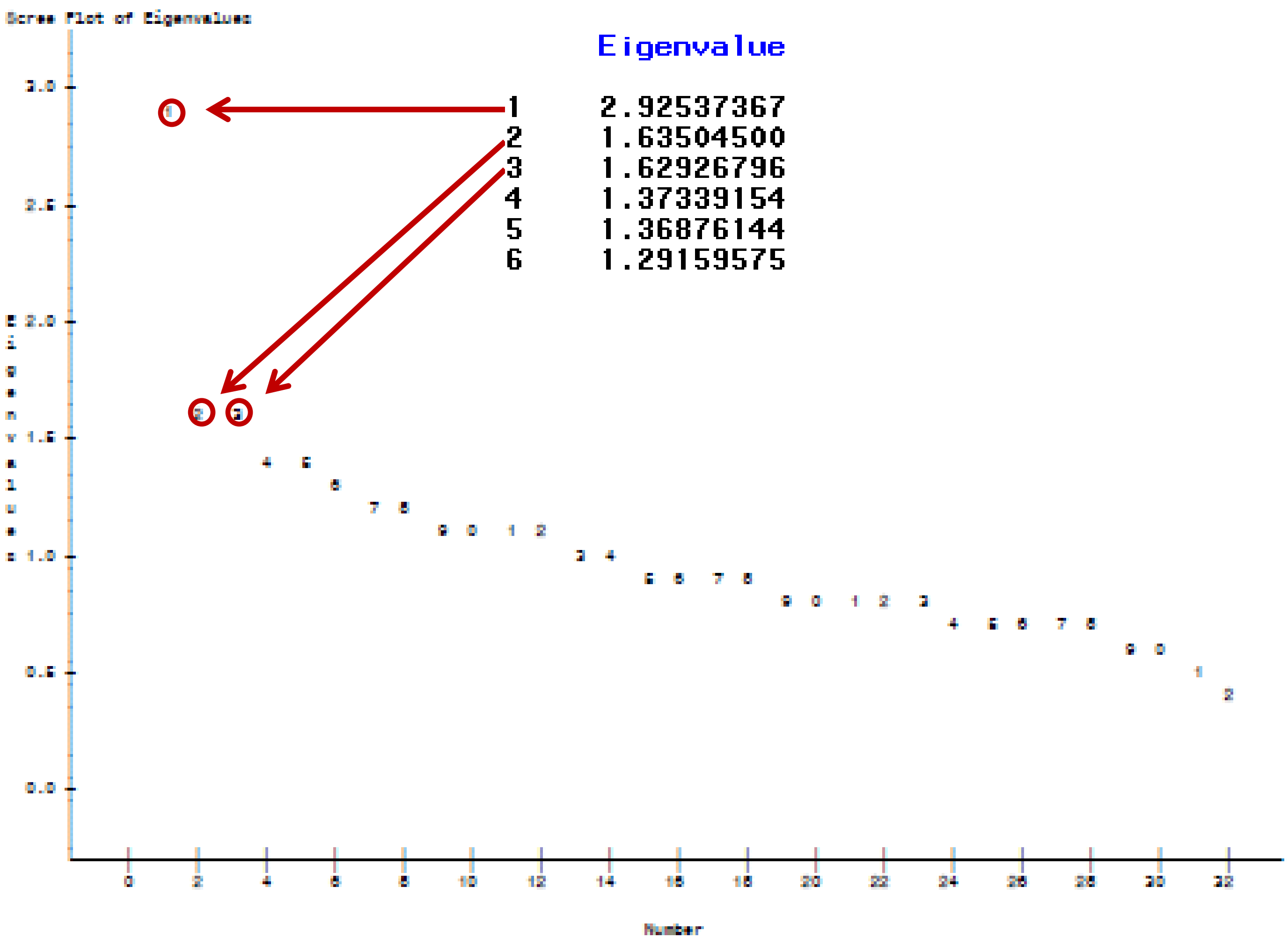
12 factors will be retained by the MINEIGEN criterion.

Variance Explained by each factor = $\frac{\text{jth eigenvalue}}{p}$.

3 factors are chosen in this analysis mainly based on interpretability

Default criterion: eigenvalue > 1

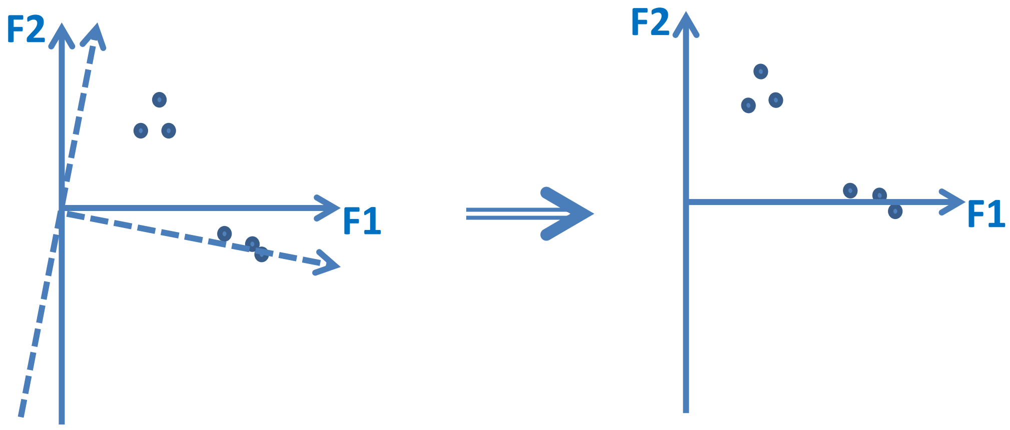
Scree plot


Scree Plot of Eigenvalues
Eigenvalue
1 2.92537367
2 1.63504500
3 1.62926796
4 1.37339154
5 1.36876144
6 1.29159575
Number

Factor Rotation

- T, an $m \times m$ orthogonal matrix, $TT' = I$.
- $X - \mu = LF + \varepsilon$

$= LTT'F + \varepsilon = (LT)(T'F) + \varepsilon = L^*F^* + \varepsilon$.



```
proc factor data=<- method=principal nfactor=3
fuzz=0.3 reorder rotate=varimax;
```

Rotated Factor Pattern

		Factor1	Factor2	Factor3
othveg	Other Vegetables	0.63362	.	.
yeveg	Dark-Yellow Vegetables	0.60724	.	.
lefveg	Green, Leafy Vegetables	0.59022	.	.
leg	Legumes	0.54864	.	.
root	Potato	0.54445	.	.
cruveg	Cruciferous Vegetables	0.42128	.	.
frefish	Fish	0.34660	.	.
toma	Tomato	0.32001	.	.
redmeat	Red Meat	.	.	.
chick	Chicken	.	.	.
fruit	Fruits	.	0.47004	.
whograin	Whole Grains	.	0.41487	.
lodai	Low-fat Dairy Products	.	0.36320	.
oj	Orange Juice	.	0.34386	.
avocado	Avocado	.	0.31852	.
soup	Cream Soup with Vegetables	.	0.30015	.
jam	Other Sweet Condiments	.	.	.
unsaoil	Vegetable Oils	.	.	.
butlard	Saturated Fats	.	.	.
mexfoods	Mexican Fried Foods	.	.	.
coffee	Coffee	.	-0.31365	.
chips	Chips	.	-0.38401	.
refgrain	Refined Grains	.	-0.52573	.
procmeat	Processed Meat	.	.	0.47812
hidai	High-fat Dairy Products	.	.	0.40812
dessert	Sweets and Desserts	.	.	0.36877
egg	Eggs	.	.	0.35313
spread	Spreads-unsaturated	.	.	0.32002
atole	Atole	.	.	.
chili	Chili Products	.	.	-0.31222
hidrink	Sugar-sweetened Beverages	.	.	-0.39192
maise	Maize-based products	.	.	-0.50578

Values less than 0.3 are not printed.

The FACTOR Procedure
Initial Factor Method: Principal Components

Variance Explained by Each Factor

Factor1	Factor2	Factor3
2.9253737	1.6350450	1.6292680

The FACTOR Procedure
Rotation Method: Varimax

Variance Explained by Each Factor

Factor1	Factor2	Factor3
2.6756253	1.8794025	1.6346589

Final Communality Estimates: Total = 6.189687

fruit	oj	egg	chick	redmeat	procmeat	frefish
0.27543124	0.11925556	0.17800331	0.13228798	0.09041763	0.29778516	0.21454422

- Rotation does not change the proportion of variance explained by all common factors.
- Communality denotes the part of variance of one variable that is explained by all common factors. It is not changed by rotation.
- $Var(X_i) = Var(l_{i1}F_1 + \cdots + l_{im}F_m) + Var(\varepsilon_i)$.

Factor Scores

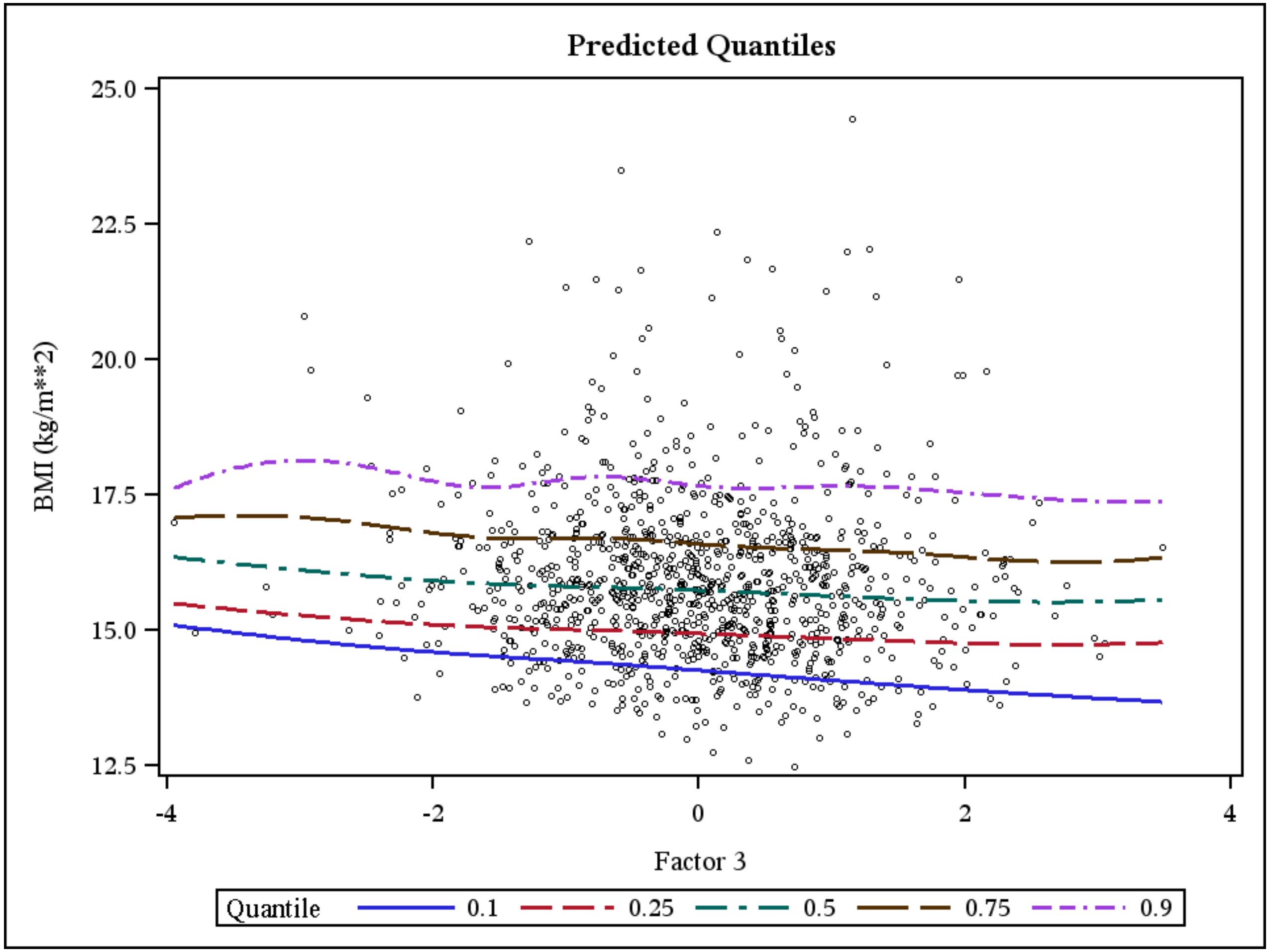
- For each subject, estimate the factor scores.
- **proc factor** data=input method=principal score outscore=out1;
- **proc score** data=input score=out1 out=out2;

ID	Factor 1	Factor 2	Factor 3
SF514	3.17	-0.48	0.22
BI1162	-2.44	1.37	-1.16

Quantile Regression

- Response: BMI
- Predictor: Dietary Factors (and other adjusting covariates)
- Purpose of quantile regression is to explore more detailed changes in the distribution quantiles besides the mean.
- In quantile regression the model predicts the specified quantile instead of the mean.

Quantile Regression



Quantile Regression

- **proc quantreg** data=<->;
  - model bmi = factor3 … / quantile=0.1 0.25 0.5 0.75 0.9;
  - output out=<-> / columnwise;
- run;

Quantile Regression

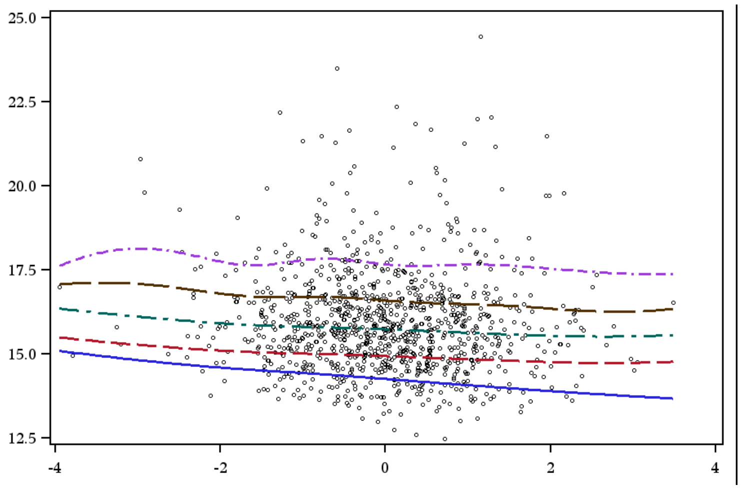
Quantile and Objective Function

Quantile	0.25
Objective Function	403.6618
Predicted Value at Mean	14.9431

Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > |t|
Intercept	1	11.1224	3.4820	4.2896	17.9552	3.19	0.0014
Factor3	1	-0.0934	0.1632	-0.4137	0.2269	-0.57	0.5673
factor3_school_t1	1	-0.0012	0.0166	-0.0337	0.0313	-0.07	0.9422
DELIVERY_SEX	1	0.1768	0.0853	0.0094	0.3441	2.07	0.0385
SCHOOL_T1	1	0.0236	0.0118	0.0005	0.0467	2.01	0.0452
DELIVERY_WEIGHT	1	0.4982	0.1103	0.2819	0.7146	4.52	<.0001
bmi_last	1	0.0352	0.0094	0.0168	0.0536	3.75	0.0002
child_age_months	1	0.0253	0.0725	-0.1170	0.1675	0.35	0.7272
BI	1	-0.3493	0.1342	-0.6126	-0.0860	-2.60	0.0094
PL	1	-0.4216	0.1454	-0.7069	-0.1362	-2.90	0.0038
SF	1	-0.1242	0.1346	-0.3883	0.1400	-0.92	0.3567

- **proc sort** data=quantileout; by factor3;
- **proc sgplot** data=quantileout;
  - scatter x=factor3 y=bmi / markerattrs=(size=**4**);
  - pbspline x=factor3 y=ht_pred / nomarkers group=QUANTILE;
  - yaxis label='BMI (kg/m**2)';
  - xaxis label='Factor 3';
- **run**;



	BMI	Factor3	bmi_pred	QUANTILE
1041	16.451478749	0.8097660475	13.949463469	0.1
1042	16.034985026	0.2112404794	14.116295528	0.1
1043	16.071221666	0.8584679653	14.29779635	0.1
1044	18.069972354	-0.054607074	14.350320175	0.1
1045	14.562767282	-0.315503198	14.745424724	0.25
1046	17.5	-1.809903623	15.427868336	0.25
1047	14.423077276	0.9939239088	14.791864587	0.25
1048	13.936947328	-1.235707327	14.58208764	0.25
1049	15.823583231	2.7638514295	14.71870942	0.25

- Thank You for Listening!
- Questions? Comments?