



## Using Factor Analysis and Multivariate Analysis of Variance to Explore Academic Achievement in the 2016 Monitoring the Future Study

Stefanie N. Kairs – NU Scholars Program, National University, San Diego, CA

### Introduction

The 2016 Monitoring the Future (MTF) 8<sup>th</sup>- and 10<sup>th</sup>-Grade surveys are part of an annual, long-term study of American adolescents and adult high school graduates conducted by the University of Michigan's Institute for Social Research. Study data have been collected since 1975, when the Institute of Social Research opened. The 2016 MTF survey involved 45,500 participants in grades 8, 10 and 12, and sampled from 372 secondary schools across the country<sup>1</sup>. With an estimated 500 variables per year, the survey covers a broad range of topics including drug use, attitudes and beliefs regarding drug use, and lifestyle choices and values. This secondary data analysis used factor analysis to extract the underlying structures for student achievement, their academic attitudes and environment, and at-risk behaviors or delinquency. The extracted factors were then used as the dependent variables in multivariate analysis of variance (MANOVA). The MANOVA analysis evaluated the relationship between the demographic independent variables for gender, ethnicity, and population density and the dependent factors for academic success, academic environment, and at-risk behaviors

### Methods

**FACTOR ANALYSIS** was performed in SAS<sup>TM</sup> Studio using the FACTOR procedure with varimax rotation. The REORDER option was used to sort variables by their factor loadings and the SCREE option was used to produce a scree plot. The parallel analysis criterion for factor retention was computed using a script previously published by Brian O'Connor<sup>2</sup>. Principal components analysis with a communality estimate of one was used for extracting factors.

**MANOVA** was performed using the GLM procedure. The independent variables were the categorical variables for gender (V7202), race (V1070) and population density (pop\_density). Gender has two levels, male (1) and female (2). Race had three levels, Black (1), White (2), and Hispanic (3). Population density had three levels, low-density (0), moderate-density (1), and high-density (2). For population density, we were primarily interested in the differences between high-density population centers and all other environments, so a contrast statement was used to code this comparison. All hypotheses were evaluated in the initial analysis. Post-hoc analysis was performed by Tukey's Studentized Range Test using a Bonferroni correction for multiple testing.

Factor Analysis: PCA varimax, parcels

The FACTOR Procedure  
Rotation Method: Varimax

	1	2	3
1	0.75478	0.52788	-0.38941
2	0.26225	0.30131	0.91875
3	0.80127	-0.79427	0.08999

Rotated Factor Pattern		Factor1	Factor2	Factor3
V7331	2016 B01 LSTYR DO BEST WK F1234	0.77090	-0.15485	-0.14447
V7221	2016 B01 R HS GRADE=1 F1234	0.80434	0.35468	-0.12423
V7226	2016 B09 R WL DO 4YR CLO F1234	0.53258	0.38945	-0.06788
V7223	2016 B01 R WL GRADUATE HS F1234	0.44488	0.34229	-0.06058
V7334	2016 B01 LSTYR WK NT DONE F1234	-0.74302	-0.11839	0.12143
parent_ed	Parental education	0.07411	0.88284	-0.04722
V7222	2016 B01 R15 HS PROGRAM F1234	0.31810	0.45219	0.07579
rem_school	Remedial schooling	-0.08895	-0.81594	0.12585
V7107	2016 A01c WK DRINKLAST30DAY F1234	-0.09548	0.12558	0.72841
V7102	2016 A01c WIGS SMKD30DAY F1234	-0.04807	-0.04178	0.87041
V7251	2016 B08 DA4W SKP CLASS F1234	-0.13283	-0.11259	0.53885
V7253	2016 B04 FRNDS DROP OUT F1234	-0.05417	-0.38135	0.50345

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
2.2550528	1.8583737	1.8041477

Figure 1. Rotated Factor Pattern

```
Title "Factor Analysis: PCA varimax, parcels";
proc factor data=fa_data parcels rotate=varimax
reorder scree nfact=3
out=fa_score;
var V7331 V7334 V7221 V7222 V7223 V7226
V7231 V7253 V7331 V7334 parent_ed
rem_school;
run;
Data remove_data;
set fa_score;
if factor1 ne .;
label factor1="achievement"
factor2="environment"
factor3="at-risk behavior";
run;
```

Figure 2. SAS Code for Factor Analysis

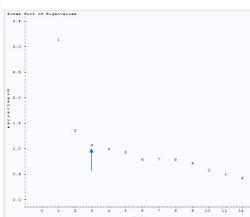


Figure 3. Scree Plot

### Results

**FACTOR ANALYSIS** Three factors relating to achievement, environment, and at-risk behaviors were extracted from 14 input variables. The rotated factor pattern is shown in Figure 1 and the scree plot is shown in Figure 3.

**MANOVA** A statistically significant overall effect was found for each independent variable: gender (Wilks'  $\lambda = 0.96$ ,  $F = 273.23$ ,  $df = (3, 17711)$ ,  $p < 0.0001$ ); race (Wilks'  $\lambda = 0.86$ ,  $F = 479.83$ ,  $df = (6, 35422)$ ,  $p < 0.0001$ ); population density (Wilks'  $\lambda = 0.98$ ,  $F = 65.31$ ,  $df = (6, 35422)$ ,  $p < 0.0001$ ). Race explains approximately 14% of the variance in academic achievement, academic environment, and at-risk behaviors (Pillai's trace = 0.145,  $p < 0.0001$ ). Gender explains 4% of the variance and population density only 2%.

The post-hoc analysis results are summarized in Table 1. Only the comparisons with a statistically significant difference at the  $\alpha = 0.0024$  level are shown.

Dependent Variable	Independent Variable	Comparison	Difference Between Means	99.76% CI
Achievement	Gender	Female – Male	0.420	0.375, 0.464
Achievement	Race	White – Hispanic	0.132	0.072, 0.192
Environment	Race	White – Black	0.457	0.384, 0.530
		White – Hispanic	0.853	0.796, 0.910
		Black – Hispanic	0.396	0.313, 0.479
Environment	Pop. Density	High – Moderate	0.064	0.009, 0.118
		High – Low	0.158	0.091, 0.221
		Moderate – Low	0.092	0.032, 0.152
At-risk behavior	Pop. Density	Moderate – Low	-0.090	-0.154, -0.026

**Table 1.** Tukey's Post Hoc Analysis Output from MANOVA. Only the comparisons significant at  $\alpha = 0.0024$  are shown. While statistically significant, the effect of population density on environment and at-risk behavior is very small.

### Conclusions

- Complete data were available for 17,719 2016 Monitoring the Future respondents from grades 8 and 10.
- 3 Factors were extracted from 14 input variables. These factors relate to academic achievement (1), environment (2), and at-risk behaviors (3).
- Student environment varies significantly by race. This factor was comprised of variables for parental education, college prep school programs, and remedial schooling (negative loading).

### References

- Johnston, L. D., et al. (2017). *Monitoring the Future: A Continuing Study of American Youth (8th- and 10th-Grade Surveys)*, 2016. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor].
- O'Connor, B. P. (2000). *SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments & Computers*, 32(3), pp. 396-402.