Multivariate Analysis on AK, GA, MT, OH, OK, and TX Homelessness and Evictions Covariates 2017

By: Michael J. Carnival

Last Updated: 4/22/2025

Contents

Introduction

Group 1 were tasked to practice multivariate techniques in the homelessness and evictions covariates 2017 dataset by Professor Amin. This project aimed to conduct multivariate statistical techniques on six states where I and others pick variables to practice these techniques. I decided to conduct these analyses on states; AK, GA, MT, OH, OK, and TX with variables; primary_care_physician_rate (PCPR), pct_single_parent_households (PSPH), pct_smokers (PS), pct_obese (PO), pct_unemployed (PU), pct_high_school_graduation (PHSG), and violent_crime_rate (VCR). The analysis included normality testing, hoteling's T-square test statistics, one- and two-way MANOVA, profile analysis and plots, principal component analysis, and factor analysis. The interpretation of these techniques is described in the section where it is implemented.

/*Top code*/

```
data mydata;
do I=1 to 709;
 one=1; end;

input state $ year totalhomeless county_population eviction_filings renting_household_population
z_med_rent z_med_inc pct_white pct_african_american pct_latinx primary_care_physician_rate
pct_single_parent_households pct_smokers pct_obese pct_unemployed pct_high_school_graduation
violent_crime_rate z_air_rseihazard z_land_rseihazard z_water_rseihazard republican_voting_pct
pct_poverty pop_per_sq_mile lng lat@@;
datalines;
Arkansas 2017 4 19019 19 2895 0.180978261 0.166075121 0.7004524 0.2580556 0.03 0.115233069
0.390369331 0.204951914 0.338 0.043851287 0.845612245 0.285257918 0.000495963 0 0 0.66365999
0.929701877 4.81E-05 -91.37520643 34.29076839 …
;
ods html close;
ods html;
```

QQ plot for normality

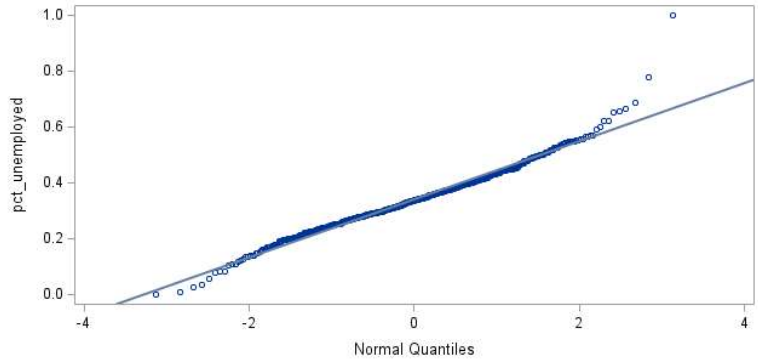
Test for normality: H_o: $normal$ | H_1: $non\ normal$

Overall

code

```
/*check normality for var*/
proc univariate normal plot;
var pct_single_parent_households; run;
proc rank normal=blom out=normals;
var pct_single_parent_households; ranks q;
data normals; set normals;
proc corr;
var pct_single_parent_households q; run;
```

single_parent_households rate



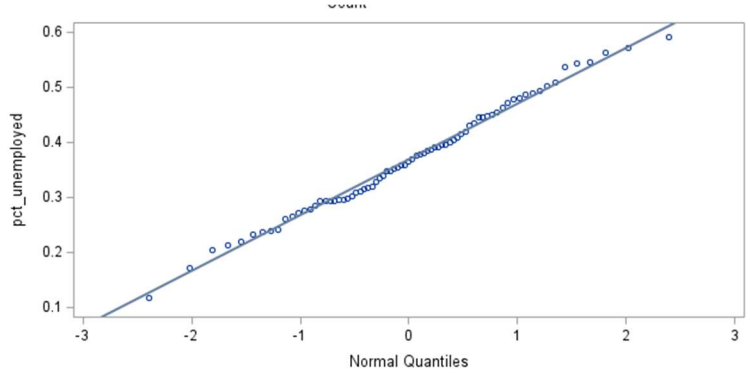
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.974777	Pr < W	<0.0001

Looking at the plot, some points do not lie on a straight line. Thus, the single parent household rates are nonnormal

Individual States

single_parent_households rate

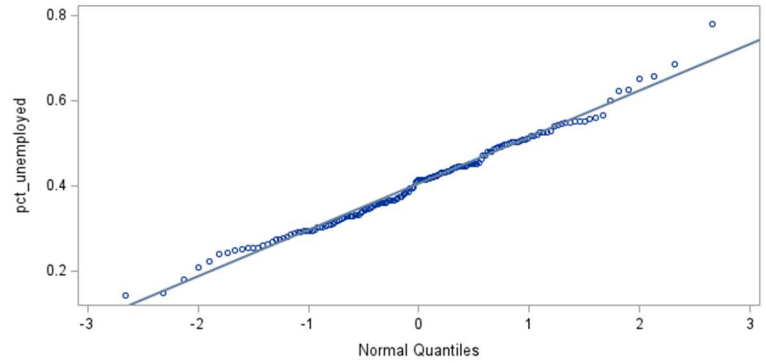
Arkansas



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.990904	Pr < W	0.8736

Looking at the plot, **most points do lie on a straight line**. Thus, the single parent household rates are **normal**

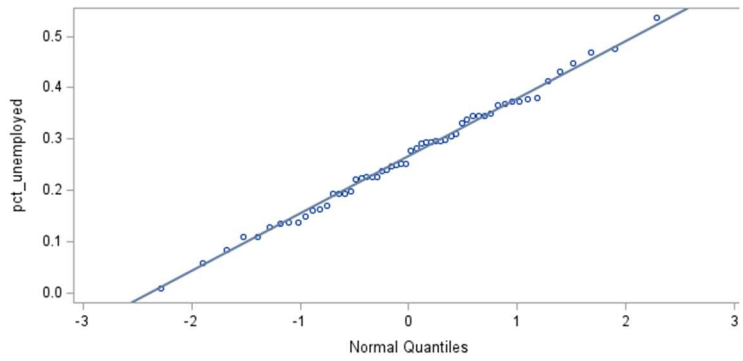
Georgia



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.988891	Pr < W	0.2428

Looking at the plot, **most points do lie on a straight line**. Thus, the single parent household rates **are normal**

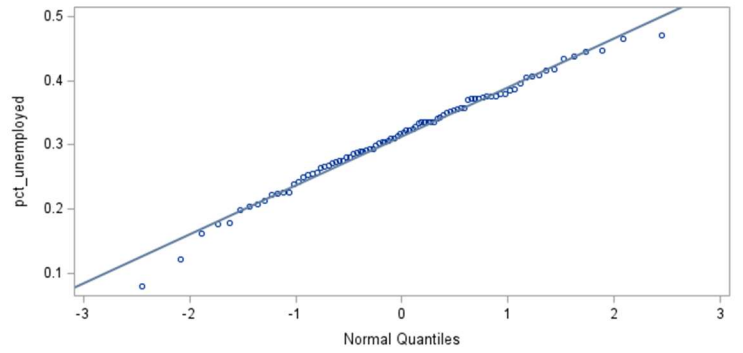
Montana



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.994036	Pr < W	0.9944

Looking at the plot, **most points do lie on a straight line.** Thus, the single parent household rates **are normal**

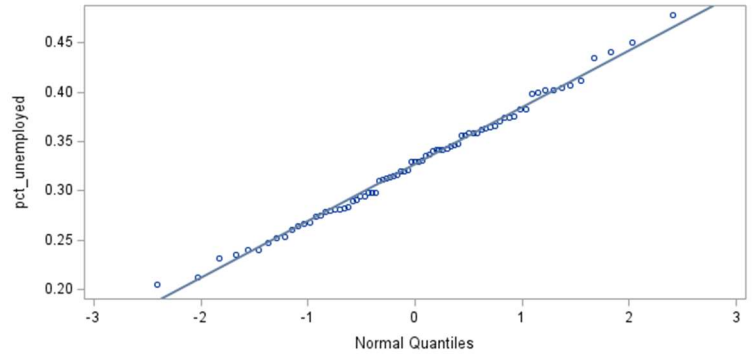
Ohio



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.987898	Pr < W	0.5932

Looking at the plot, **most points do lie on a straight line**. Thus, the single parent household rates **are normal**

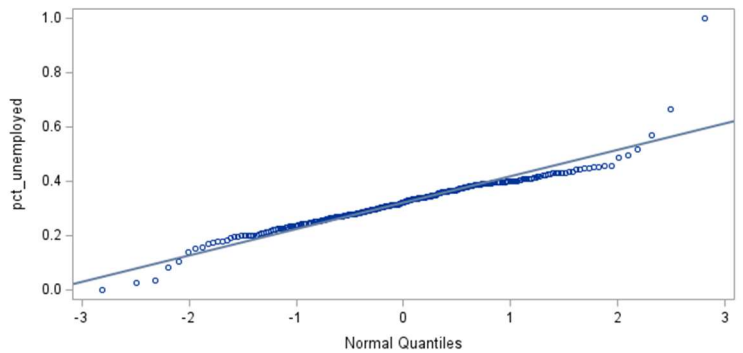
Oklahoma



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.992085	Pr < W	0.9196

Looking at the plot, **most points do lie on a straight line**. Thus, the single parent household rates **are normal**

Texas



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.909786	Pr < W	<0.0001

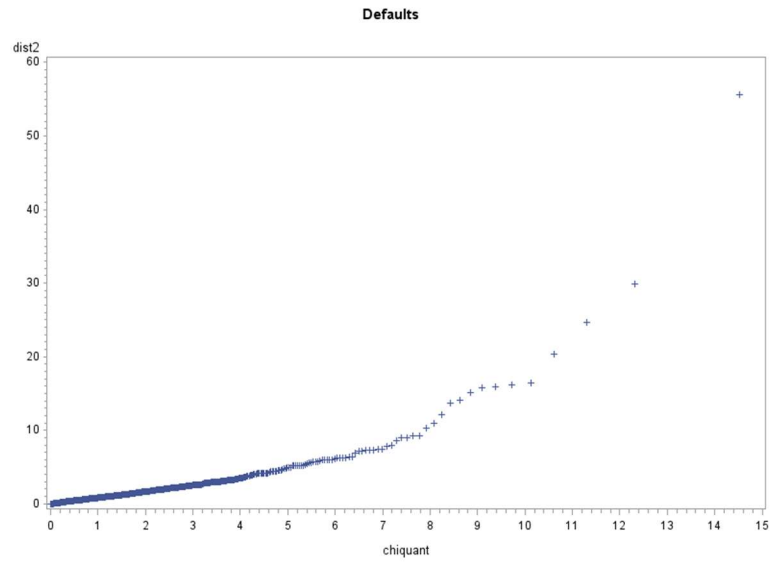
Looking at the plot, **most points do not lie on a straight line**. Thus, the single parent household rates **are nonnormal**

Bivariate Normal

CODE

```
/*check bivariate normality*/
proc princomp std out=pcresult;var var1 var2;run;
data mahal;set pcresult;dist2=uss(of prin1-prin2);run;
proc sort;by dist2;run;data plotdata;set mahal;
prb=(_n_ -.5)/709;chiquant=cinv(prb,2);run;
proc gplot;plot dist2*chiquant;run;
proc print; var dist2 chiquant; run;
proc corr; var dist2 chiquant; run;
```

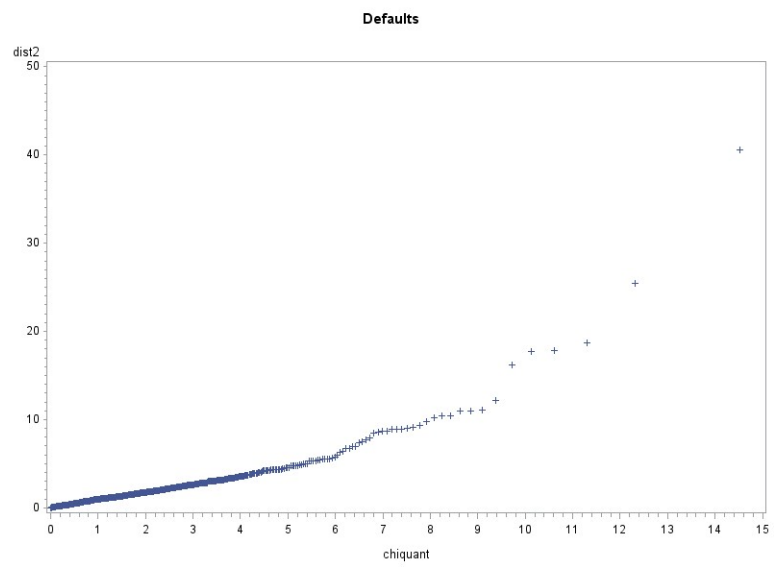
pct_single_parent_households, and pct_unemployed



Pearson Correlation Coefficients, N = 709 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.86422 <.0001
chiquant	0.86422 <.0001	1.00000

Ho: normal | H1: nonnormal. Since r_Q **=0.864 < 0.9953**, **we reject** the null. Thus, the data is **nonnormal.**

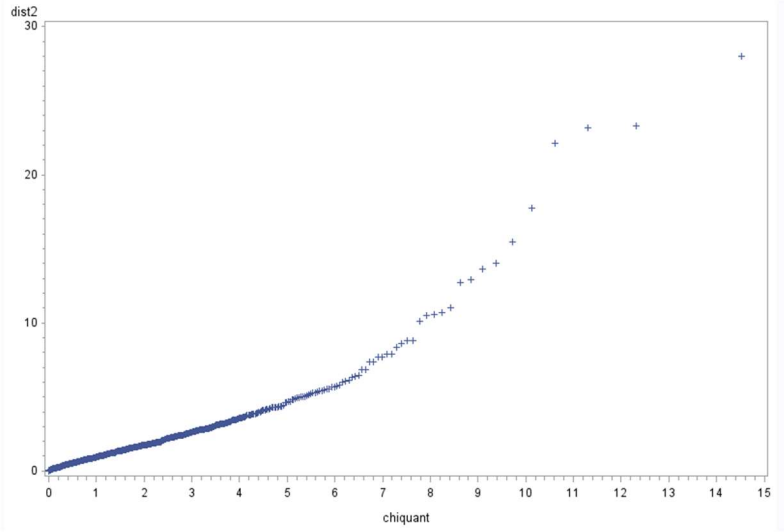
pct_single_parent_households, and pct_smokers



Pearson Correlation Coefficients, N = 709 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.91485 <.0001
chiquant	0.91485 <.0001	1.00000

Ho: normal | H1: nonnormal. Since r_Q **=0.91485 < 0.9953**, **we reject** the null. Thus, the data is **nonnormal.**

pct_unemployed, and pct_smokers



Pearson Correlation Coefficients, N = 709 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.93492 <.0001
chiquant	0.93492 <.0001	1.00000

Ho: normal | H1: nonnormal. Since r_Q **=0.93492 < 0.9953**, **we reject** the null. Thus, the data is **nonnormal.**

Looking at the pairs of bivariate plots, many points do not lie on a straight line. Thus, the **random vectors; pct_single_parent_households, pct_unemployed , and pct_smokers** do not follow a bivariate normal distribution.

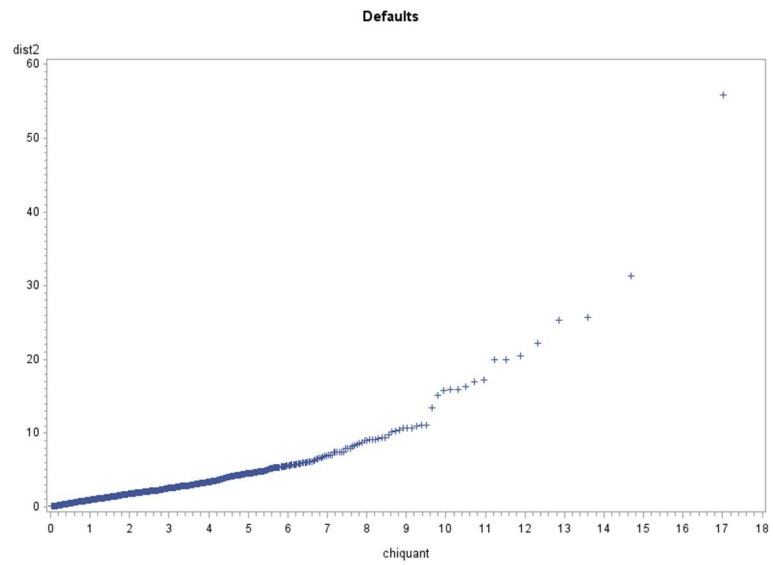
Gamma Plot

Code

```
proc princomp std out=pcresult;var pct_single_parent_households pct_smokers
pct_unemployed; run;
data mahal;set pcresult;dist2=uss(of prin1-prin3); by state; run;
proc sort;by dist2;run;data plotdata;set mahal;
```

```
prb=(_n_ -.5)/709;chiquant=cinv(prb,3);run;
proc gplot;plot dist2*chiquant;run;
proc corr; var dist2 chiquant; run;
```

Gamma plot for covariates, **pct_single_parent_households, pct_unemployed , and pct_smokers**



Pearson Correlation Coefficients, N = 709 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.89351 <.0001
chiquant	0.89351 <.0001	1.00000

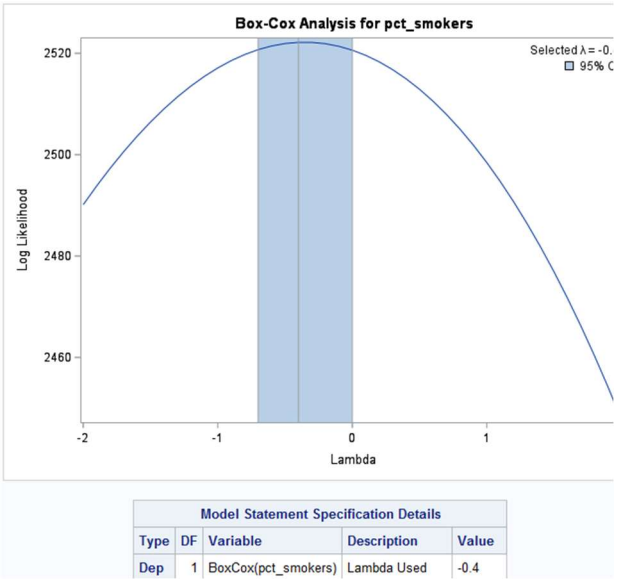
Ho: normal | H1: nonnormal. Since **r_Q = 0.89351 < 0.9953, we reject** the null. Thus, the data is **nonnormal.**

Looking at the gamma plot, many points do not lie on a straight line. Thus, the **random vectors; pct_single_parent_households, pct_unemployed , and pct_smokers** do not follow a multivariate normal distribution

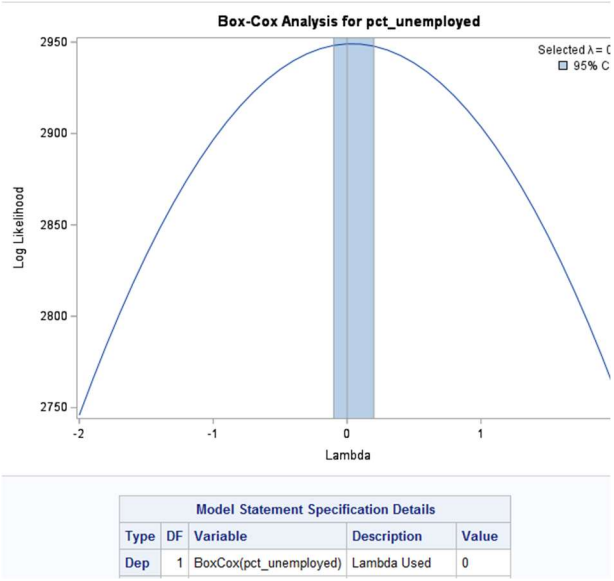
Boxcox transformation

Code

```
proc transreg data=mydata details;
      title2 'Defaults';
      model boxcox(pct_single_parent_households / lambda=-2 to 2 by 0.1) =
identity(one);
      run;
proc transreg data=mydata details;
      title2 'Defaults';
      model boxcox(pct_smokers / lambda=-2 to 2 by 0.1) = identity(one);
      run;
proc transreg data=mydata details;
      title2 'Defaults';
      model boxcox(pct_unemployed / lambda=-2 to 2 by 0.1) = identity(one);
      run;
```



Model Statement Specification Details				
Type	DF	Variable	Description	Value
Dep	1	BoxCox(pct_smokers)	Lambda Used	-0.4



Model Statement Specification Details				
Type	DF	Variable	Description	Value
Dep	1	BoxCox(pct_unemployed)	Lambda Used	0

The best transformations for variables of interest resulted in only two recommendations instead of three. These values are -0.4 and 0 (log transformation) for the **pct_smokers** and **pct_unemployed variables**.

Reassessing of transformed variables for **bivariate normal**

Single parent household vs unemployed

Pearson Correlation Coefficients, N = 709 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.91779 <.0001
chiquant	0.91779 <.0001	1.00000

Ho: normal | H1: nonnormal.
Since r_Q **= 0.9178 < 0.9953**, **we reject** the null. Thus, the data is **nonnormal.**

Single parent household vs smokers

Pearson Correlation Coefficients, N = 709 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.86871 <.0001
chiquant	0.86871 <.0001	1.00000

Ho: normal | H1: nonnormal.
Since r_Q **= 0.8687< 0.9953**, **we reject** the null. Thus, the data is **nonnormal.**

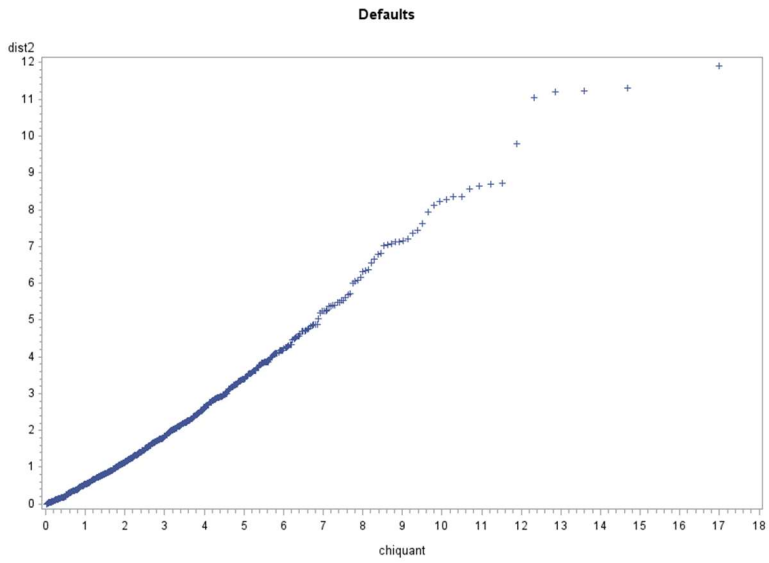
Unemployed vs smokers

Pearson Correlation Coefficients, N = 709 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.99770 <.0001
chiquant	0.99770 <.0001	1.00000

Ho: normal | H1: nonnormal.
Since r_Q **= 0.9977 > 0.9953**, **we do not reject** the null. Thus, the data is **normal.**

When the boxcox transformation were applied to the variables, **only the pair, pct_unemployed and pct_smokers resulted in bivariate normal** at alpha level = 0.5**.**

Reassessing of transformed variables for multivariate **normal test**



Pearson Correlation Coefficients, N = 708 Prob > |r| under H0: Rho=0		
	dist2	chiquant
dist2	1.00000	0.99447 <.0001
chiquant	0.99447 <.0001	1.00000

Looking at the plot, a decent amount of data points does not lie on the straight line. Thus, the **MVN is nonnormal**

The two-sample hoteling's T-square test statistics

$$H_0: \boldsymbol{\mu_1} = \boldsymbol{\mu_2} \mid H_1: \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$$

Code

```
proc iml;
  start hotel2;
    n1=nrow(x1);
    n2=nrow(x2);
    k=ncol(x1);
    one1=j(n1,1,1);
    one2=j(n2,1,1);
    ident1=i(n1);
    ident2=i(n2);
    ybar1=x1`*one1/n1;
    s1=x1`*(ident1-one1*one1`/n1)*x1/(n1-1.0);
/*    print n1 ybar1;*/
/*    print s1;*/
    ybar2=x2`*one2/n2;
    s2=x2`*(ident2-one2*one2`/n2)*x2/(n2-1.0);
/*    print n2 ybar2;*/
/*    print s2;*/
    spool=((n1-1.0)*s1+(n2-1.0)*s2)/(n1+n2-2.0);
/*    print spool;*/
    t2=(ybar1-ybar2)`*inv(spool*(1/n1+1/n2))*(ybar1-ybar2);
    f=(n1+n2-k-1)*t2/k/(n1+n2-2);
    df1=k;
    df2=n1+n2-k-1;
    p=1-probf(f,df1,df2);
    print t2 f df1 df2 p;
  finish;

  use mydata;
    read all var{PCPR PSPH PS PO PU PHSG VCR} where (state="Georgia") into x1;
    read all var{PCPR PSPH PS PO PU PHSG VCR} where (state="Arkansas") into x2;
  run hotel2;
  use mydata;
    read all var{PCPR PSPH PS PO PU PHSG VCR} where (state="Georgia") into x1;
    read all var{PCPR PSPH PS PO PU PHSG VCR} where (state="Montana") into x2;
  run hotel2;
....
```

2-Sample Hotellings T2 - states

t2	f	df1	df2	p
410.4454	57.118633	7	226	0

t2	f	df1	df2	p
270.65456	37.575784	7	207	0

t2	f	df1	df2	p
387.47061	53.997362	7	239	0

t2	f	df1	df2	p
624.48557	86.924732	7	228	0

t2	f	df1	df2	p
395.66886	55.696941	7	404	0

t2	f	df1	df2	p
268.46939	36.568921	7	123	0

t2	f	df1	df2	p
89.666256	12.332094	7	155	1.713E-12

t2	f	df1	df2	p
102.92488	14.115413	7	144	7.183E-14

t2	f	df1	df2	p
942.03703	132.09985	7	320	0

t2	f	df1	df2	p
127.67266	17.468291	7	136	2.22E-16

t2	f	df1	df2	p
157.67593	21.493448	7	125	0

t2	f	df1	df2	p
232.58724	32.577366	7	301	0

t2	f	df1	df2	p
109.28263	15.037137	7	157	5.995E-15

t2	f	df1	df2	p
675.51006	94.793447	7	333	0

t2	f	df1	df2	p
1051.0162	147.39861	7	322	0

We can **reject** the null hypothesis that the mean vector for the state1 equals the mean vector for the state2 given the evidence shown in the table below **P < 0.0001** for the states considered.

(note. the variables used were primary_care_physician_rate pct_single_parent_households pct_smokers pct_obese pct_unemployed pct_high_school_graduation violent_crime_rate)

pairs	t2	f	df1	df2	p
GA vs AK	410.45	57.12	7	226	0
GA vs MT	270.66	37.58	7	207	0
GA vs OH	387.47	54.00	7	239	0
GA vs OK	624.50	86.92	7	228	0
GA vs TX	365.67	55.70	7	404	0
AK vs MT	268.47	36.69	7	123	0
AK vs OH	89.67	12.33	7	155	1.71E-12
AK vs OK	102.92	14.12	7	144	7.14E-14
AK vs TX	942.04	132.10	7	320	0
AK vs OH	127.67	17.47	7	136	2.23E-16
AK vs OK	157.68	21.49	7	125	0
AK vs TX	232.90	32.58	7	301	0
OH vs OK	109.28	15.04	7	157	6.00E-15
OH vs TX	675.51	94.79	7	333	0
OK vs TX	1051.02	147.39	7	322	0

One-way MANOVA

code

```
proc glm;
class state;
model PCPR PSPH PS PO PU PHSG VCR = state/ss3;
means state/tukey;
manova h=state/printe printh;
run;
```

p: variable = 7

g: group = 6

the hypothesis of no treatment effects

$$H_0: \tau_1 = \tau_1 = \cdots = \tau_g = 0$$

$$H_1: at\ least\ one\ \tau_l \neq 0\ is\ available$$

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall state Effect
H = Type III SSCP Matrix for state
E = Error SSCP Matrix

S=5 M=0.5 N=347

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.17208550	43.49	35	2930.2	<.0001
Pillai's Trace	1.29274799	34.87	35	3500	<.0001
Hotelling-Lawley Trace	2.59276208	51.46	35	1968.3	<.0001
Roy's Greatest Root	1.68681451	168.68	7	700	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Since **wilks' lambda** is **significant**, we reject H_0 at α =0.01 level and conclude that treatment difference exists. Then we look into each variable.

Primary care physician rate

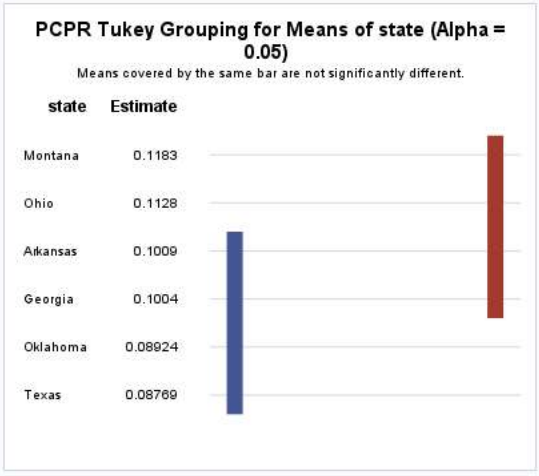
MANOVA

The GLM Procedure

Dependent Variable: PCPR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.07657801	0.01531560	4.98	0.0002
Error	702	2.15796379	0.00307402		
Corrected Total	707	2.23454180			

Primary care physician rate is significant, thus at least one state differs in **primary care physician rate.**



Primary care physician rate: MSE = 0.003; All means not covered by a common bar are significantly different. That is µMT are significantly different from µOK and µTX.

Percent of single parent household

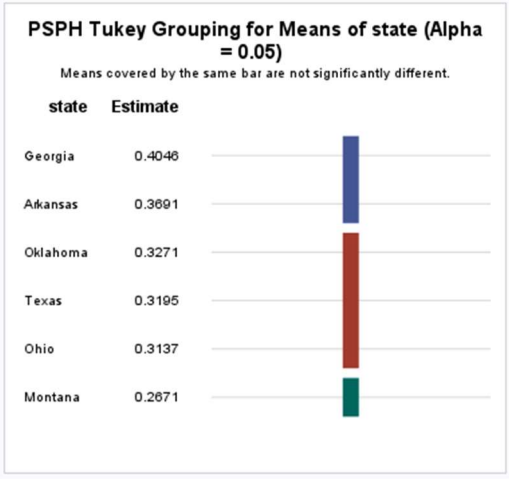
MANOVA

The GLM Procedure

Dependent Variable: PSPH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.20564434	0.24112887	28.18	<.0001
Error	702	6.00723407	0.00855731		
Corrected Total	707	7.21287840			

Percent of single parent household is significant, thus at least one state differs in **Percent of single parent household.**



Percent of single parent household: MSE = 0.0086; All means not covered by a common bar are significantly different. That is µGA are significantly different from µOK, µTX, µOH, and µMT.

Percent of smokers

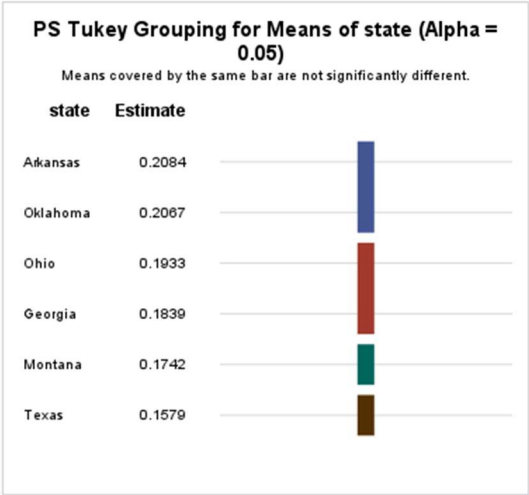
MANOVA

The GLM Procedure

Dependent Variable: PS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.25883150	0.05176630	103.53	<.0001
Error	702	0.35099368	0.00049999		
Corrected Total	707	0.60982518			

Percent of smokers is significant, thus at least one state differs in **percent of smokers**



Percent of smokers: MSE = 0.0086; All means not covered by a common bar are significantly different. That is µAK are significantly different from µOH, µGA, µMT and µTX.

Percent of obesity

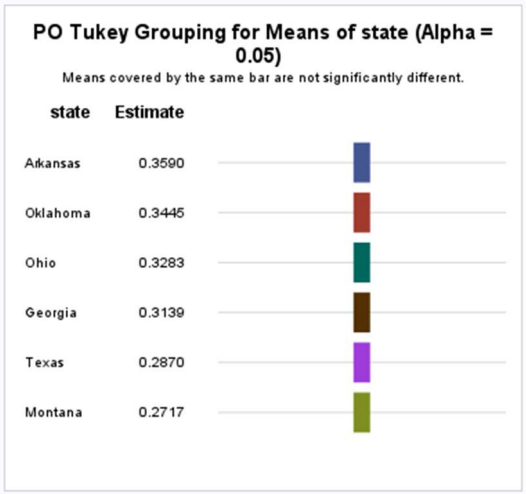
MANOVA

The GLM Procedure

Dependent Variable: PO

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.51871365	0.10374273	134.14	<.0001
Error	702	0.54294095	0.00077342		
Corrected Total	707	1.06165459			

Percent of obesity is significant, thus at least one state differs in p**ercent of obesity**



Percent of obesity: MSE = 0.00077; All means not covered by a common bar are significantly different. That is µAK are significantly different from the rest

Percent of unemployment

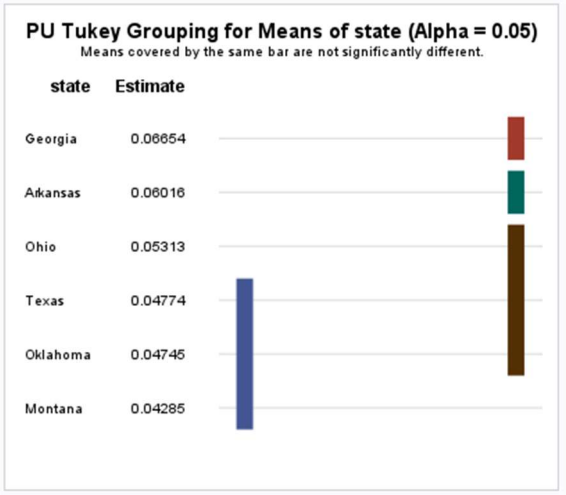
MANOVA

The GLM Procedure

Dependent Variable: PU

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.04793742	0.00958748	45.39	<.0001
Error	702	0.14828312	0.00021123		
Corrected Total	707	0.19622054			

Percent of unemployment is significant, thus at least one state differs in **percent of unemployment**



Percent of unemployment: MSE = 0.00021; All means not covered by a common bar are significantly different. That is µGA are significantly different from the rest.

Percent of high school graduate

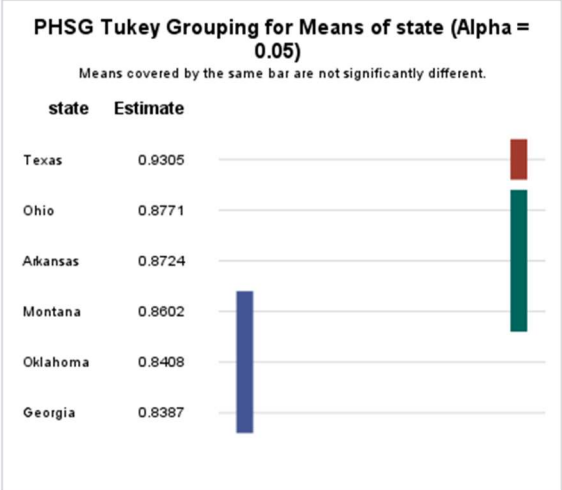
MANOVA

The GLM Procedure

Dependent Variable: PHSG

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.05963052	0.21192610	50.51	<.0001
Error	702	2.94534667	0.00419565		
Corrected Total	707	4.00497719			

Percent of high school graduate is significant, thus at least one state differs in **percent of high school graduate**



percent of high school graduate : MSE = 0.0041; All means not covered by a common bar are significantly different. That is µTX are significantly different from the rest.

Violent crime ratee

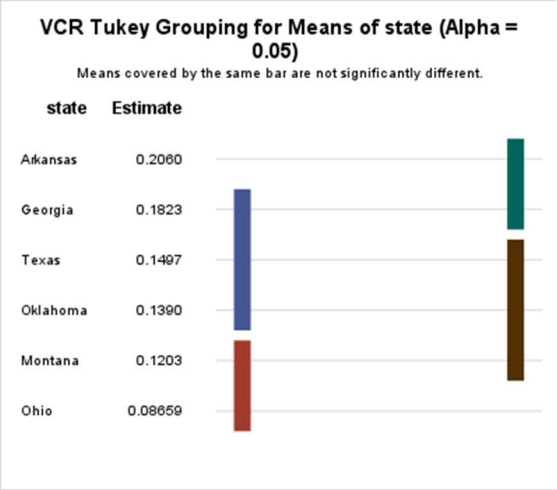
MANOVA

The GLM Procedure

Dependent Variable: VCR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.81204972	0.16240994	14.51	<.0001
Error	702	7.85688719	0.01119215		
Corrected Total	707	8.66893691			

Violent crime rate is significant, thus at least one state differs in v**iolent crime rate**



Violent crime rate
: MSE = 0.0011; All means not covered by a common bar are significantly different. That is µAK are significantly different from the µTX, µOK, µMT, and µOH.

Two-way ANOVA:

Main effect factor A: The mean across all levels of factor 1 are all equal

$$Equivalent: H_0^A: \mu_{A1} = \mu_{A2} = \cdots = \mu_{AK}$$

$$H_1^A: at\ least\ one\ \mu_{Ai} \neq \mu_{Aj}$$

Main effect factor B: The mean across all levels of factor 2 are all equal

$$Equivalent: H_0^B: \mu_{B1} = \mu_{B2} = \cdots = \mu_{BM}$$

$$H_1^B: at\ least\ one\ \mu_{Bi} \neq \mu_{Bj}$$

Interaction effect factor A and B: there is no interaction effect between the factors

$$Equivalent: H_0^{AB}: \mu_{AiBj} = \mu_{AkBl}$$

$$H_1^B: at\ least\ one\ \mu_{AiBj} \neq \mu_{AkBl}$$

code

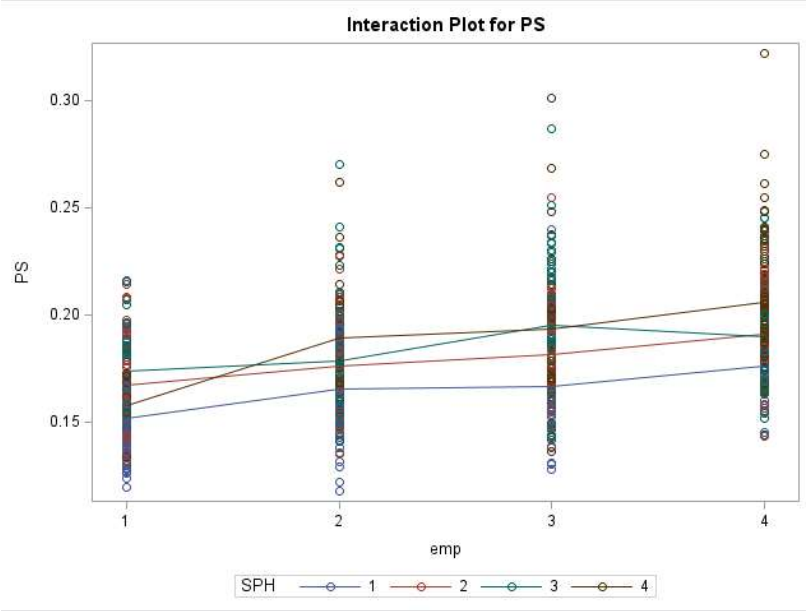
```
/*top*/
data mydata;
/*primary_care_physician_rate pct_single_parent_households pct_smokers
pct_obese pct_unemployed pct_high_school_graduation violent_crime_rate*/
input state $ PCPR PSPH PS PO PU PHSG VCR;

if PU le 0.0416618 then emp ='1';
if PU gt 0.0416618 and PU le 0.0508919 then emp = '2';
if PU gt 0.0508919 and PU le 0.0635383 then emp = '3';
if PU gt 0.0635383 then emp ='4';


if PSPH le 0.277101 then SPH ='1';
if PSPH gt 0.277101 and PSPH le 0.337355 then SPH = '2';
if PSPH gt 0.337355 and PSPH le 0.398567 then SPH = '3';
if PSPH gt 0.398567 then SPH ='4';
*/
…
/*proc univariate; var PCPR PSPH PS PO PU PHSG VCR; run;*/
proc glm;
class emp SPH;
model PS = emp SPH emp*SPH;
manova h = emp SPH emp*SPH/printe printh;
run;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall emp*SPH Effect
H = Type III SSCP Matrix for emp*SPH
E = Error SSCP Matrix

S=1 M=3.5 N=345

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.97497709	1.97	9	692	0.0398
Pillai's Trace	0.02502291	1.97	9	692	0.0398
Hotelling-Lawley Trace	0.02566513	1.97	9	692	0.0398
Roy's Greatest Root	0.02566513	1.97	9	692	0.0398

Looking at the wilks' lambda, there is a **significant effect on the interaction between unemployment level** and **level of single parent household** on **smoking rate**.

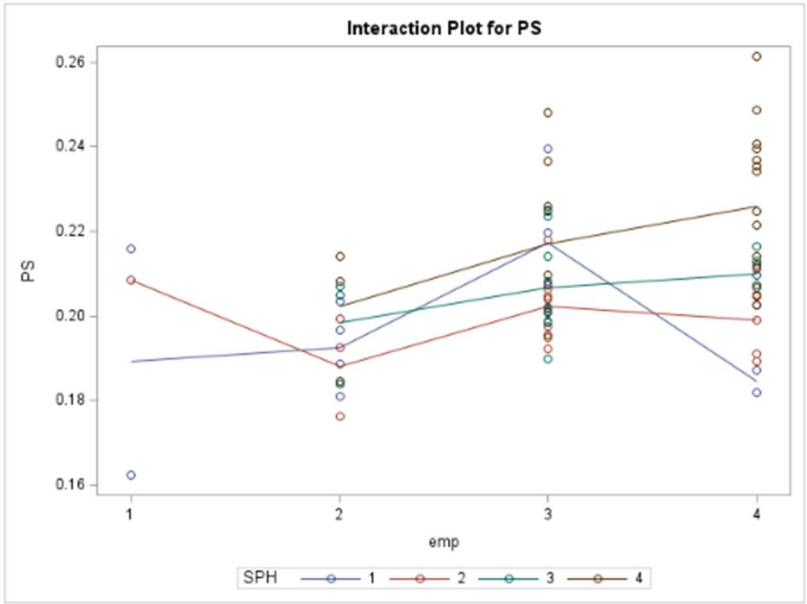


The difference in PS for the different level of SPH depends on the level of unemployment level.

- For level of single parent household, low level of unemployment has a lower percent of smoking

Arkansas

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.00219394	0.00073131	3.48	0.0212
SPH	3	0.00313640	0.00104547	4.97	0.0038
emp*SPH	7	0.00245357	0.00035051	1.67	0.1343

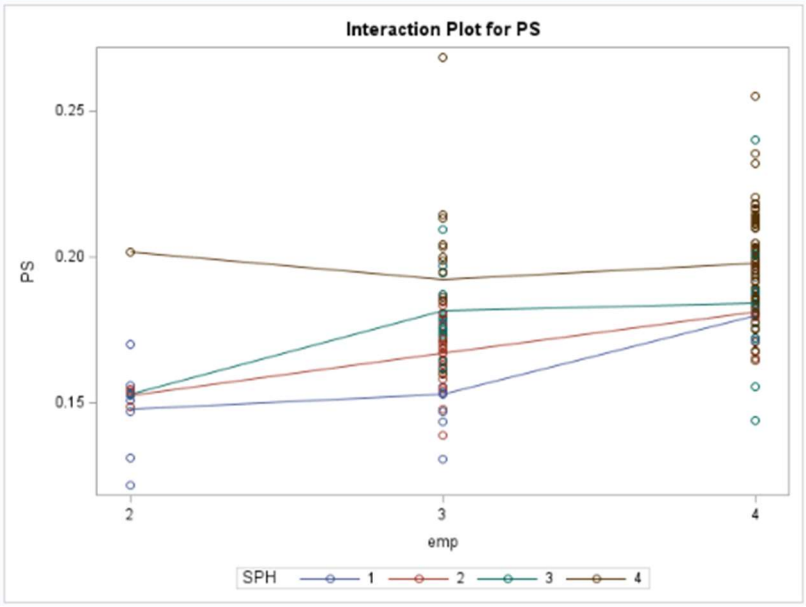


There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects.**

- percent of smoker depends on level of single parent household level
- percent of smokers depend on unemployment level

Georgia

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	2	0.00331251	0.00165626	5.40	0.0055
SPH	3	0.00743418	0.00247806	8.08	<.0001
emp*SPH	6	0.00174503	0.00029084	0.95	0.4627

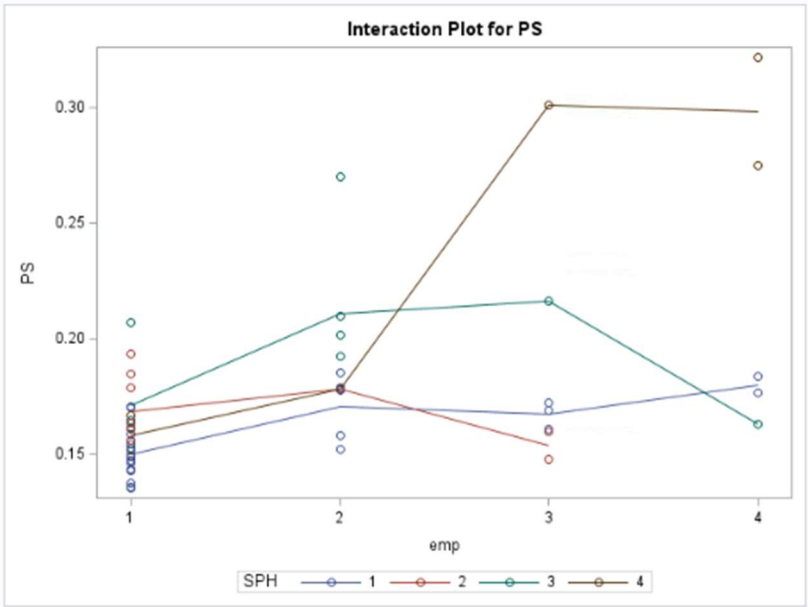


There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects.**

- percent of smoker depends on level of single parent household level
- percent of smokers depend on unemployment level

Montana

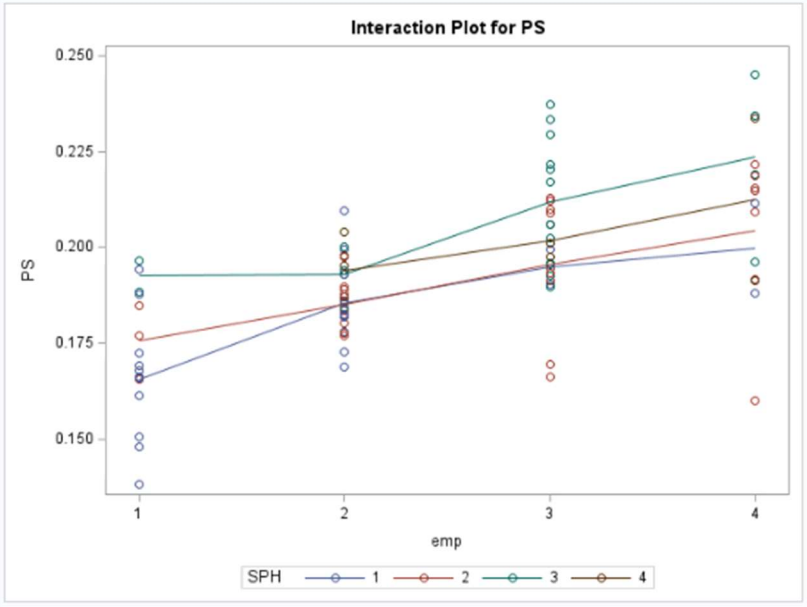
Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.01387948	0.00462649	15.08	<.0001
SPH	3	0.01831479	0.00610493	19.90	<.0001
emp*SPH	8	0.02400390	0.00300049	9.78	<.0001



The difference in **percent of smoking** for the different level of **single parent household depends** on the **levels of unemployment**.

- For level of single parent household, low level of unemployment has a lower percent of smoking

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.00661677	0.00220559	10.00	<.0001
SPH	3	0.00301064	0.00100355	4.55	0.0056
emp*SPH	8	0.00064013	0.00008002	0.36	0.9367

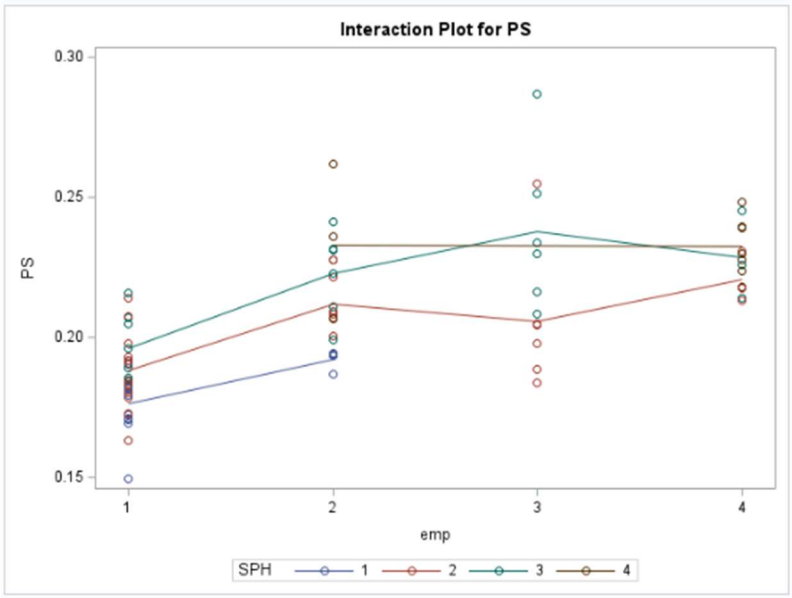


There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects.**

- percent of smoker depends on level of single parent household level (high smoker= high single parent household)
- percent of smokers depend on unemployment level (high smoker= high unemployment)

Oklahoma

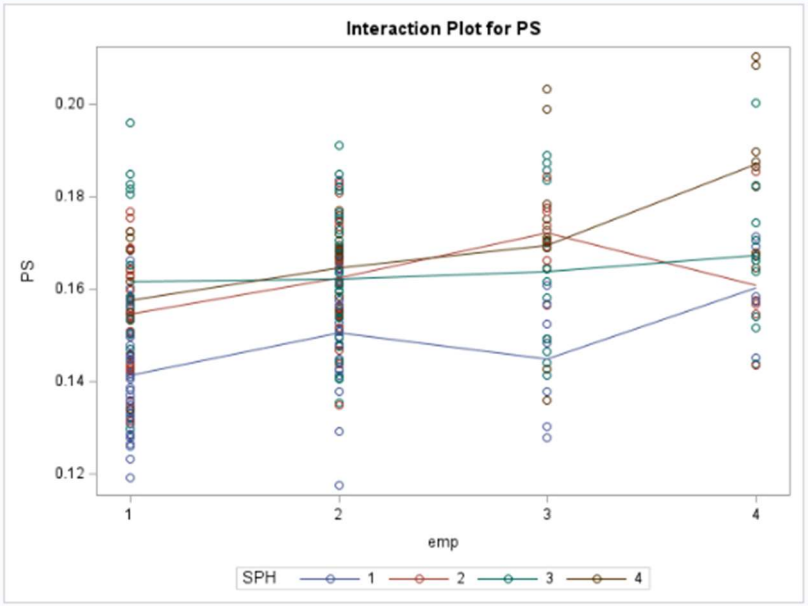
Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.00964962	0.00321654	12.70	<.0001
SPH	3	0.00729057	0.00243019	9.60	<.0001
emp*SPH	5	0.00155794	0.00031159	1.23	0.3051



There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects.**

- percent of smoker depends on level of single parent household level
- percent of smokers depend on unemployment level

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.00487769	0.00162590	8.59	<.0001
SPH	3	0.00858485	0.00286162	15.11	<.0001
emp*SPH	9	0.00321180	0.00035687	1.88	0.0550



There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects.**

- percent of smoker depends on level of single parent household level
- percent of smokers depend on unemployment level

Two-way MANOVA:

Overall

$$H_0: \gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = 0$$
$$H_1: at\ least\ one\ interaction\ \gamma \neq 0$$

```
/*unbalanced -> sample sizenot equal */
proc glm;
class emp SPH;
model PCPR PS PO PHSG VCR = emp SPH emp*SPH;
manova h = emp SPH emp*SPH/printe printh;
lsmeans emp SPH/pdiff;
run;
```

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall emp*SPH Effect
H = Type III SSCP Matrix for emp*SPH
E = Error SSCP Matrix

S=5 M=1.5 N=343

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.92693875	1.17	45	3080.7	0.2031
Pillai's Trace	0.07496567	1.17	45	3460	0.2036
Hotelling-Lawley Trace	0.07678798	1.17	45	2140.6	0.2032
Roy's Greatest Root	0.03593898	2.76	9	692	0.0035

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Wilks' lambda results conclude **no significance** in **interactions**.

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall SPH Effect
H = Type III SSCP Matrix for SPH
E = Error SSCP Matrix

S=3 M=0.5 N=343

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.85463119	7.42	15	1899.7	<.0001
Pillai's Trace	0.14767759	7.14	15	2070	<.0001
Hotelling-Lawley Trace	0.16739577	7.67	15	1294.3	<.0001
Roy's Greatest Root	0.14942656	20.62	5	690	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

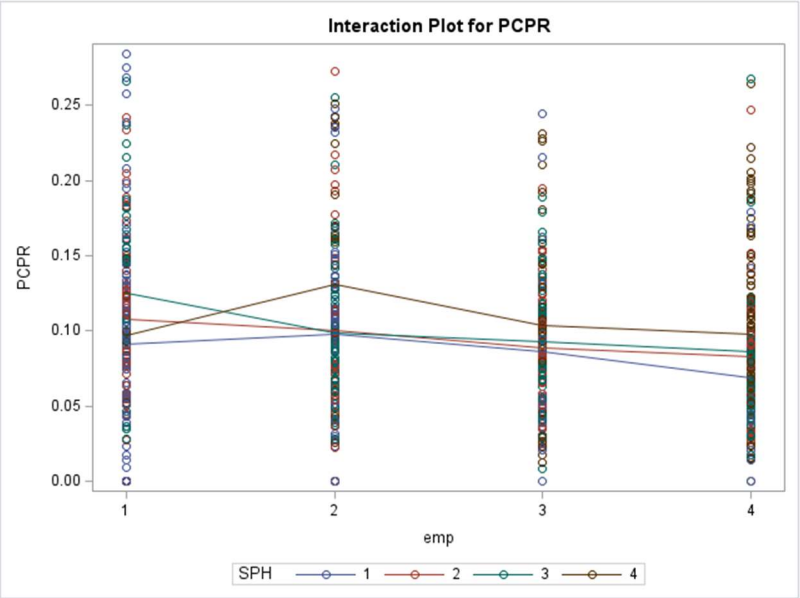
Wilks' lambda results conclude significance for **single parent household levels** in **the six states**

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall emp Effect H = Type III SSCP Matrix for emp E = Error SSCP Matrix S=3 M=0.5 N=343					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.86597477	6.78	15	1899.7	<.0001
Pillai's Trace	0.13593566	6.55	15	2070	<.0001
Hotelling-Lawley Trace	0.15256621	6.99	15	1294.3	<.0001
Roy's Greatest Root	0.13672100	18.87	5	690	<.0001
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					

Wilks' lambda results conclude significance for **unemployment levels in the six states.** Thus, interaction plots are not required for interpreting the results of the two-way multivariate analysis of variance.

Primary care physician rate (PCPR)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.04367663	0.01455888	4.71	0.0029
SPH	3	0.02629774	0.00876591	2.83	0.0375
emp*SPH	9	0.02562609	0.00284734	0.92	0.5067

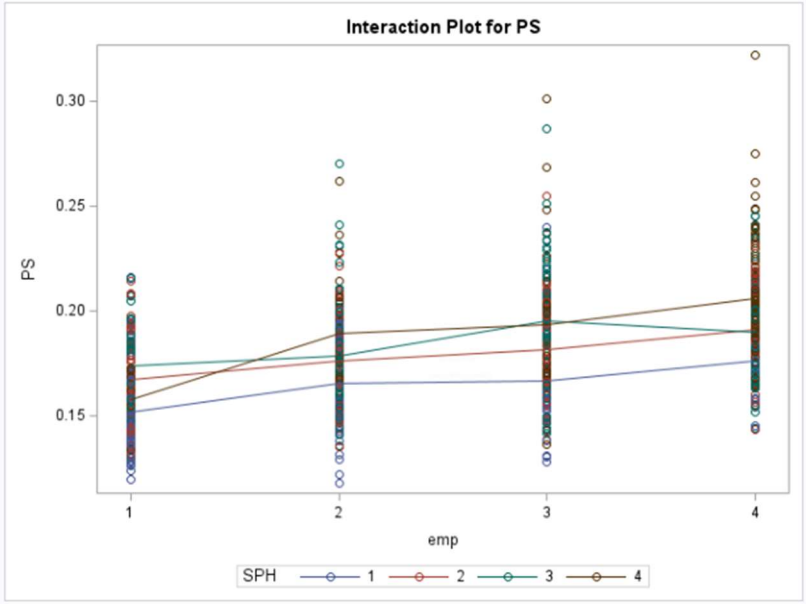


There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects α =0.01.**

- Primary care physician rates depend on level of single parent household level
- Primary care physician rates depend on unemployment level

Percent of smoking (PS)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.04564454	0.01521485	25.47	<.0001
SPH	3	0.03093145	0.01031048	17.26	<.0001
emp*SPH	9	0.01060975	0.00117886	1.97	0.0398

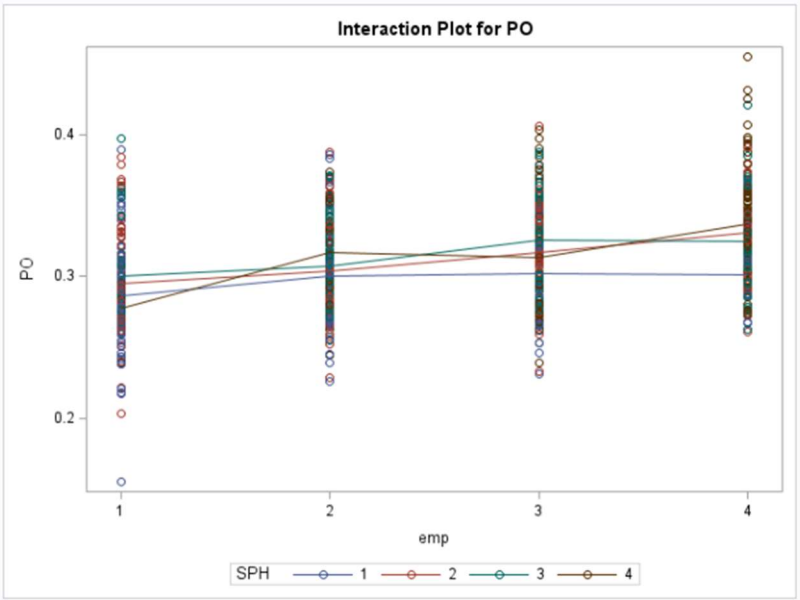


There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects at α =0.01.**

- Percent of smoking depend on level of single parent household level
- Percent of smoking depend on unemployment level

Percent of obesity

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.06441384	0.02147128	17.10	<.0001
SPH	3	0.02140281	0.00713427	5.68	0.0008
emp*SPH	9	0.01937526	0.00215281	1.71	0.0820

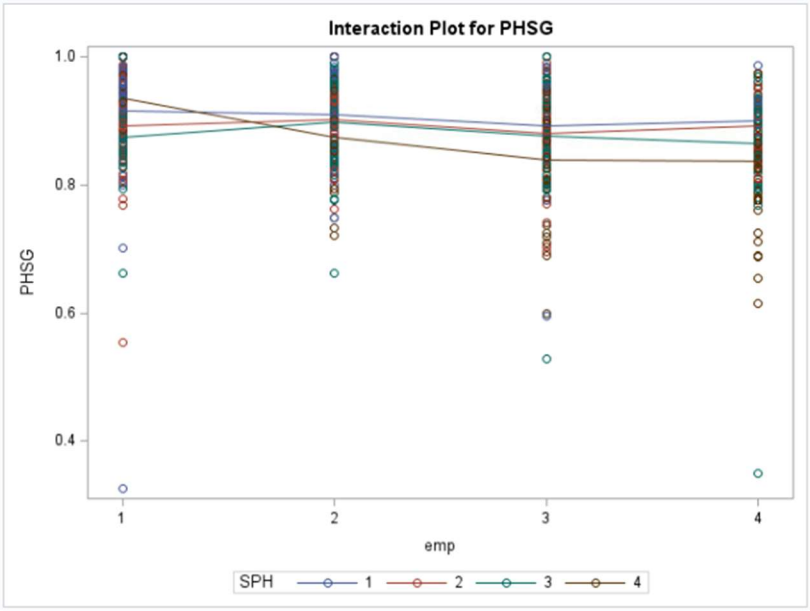


There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects at α =0.01.**

- Percent of obesity depend on level of single parent household level
- Percent of obesity depend on unemployment level

Percent of High School Graduate (PHSG)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.10007036	0.03335679	6.62	0.0002
SPH	3	0.06872489	0.02290830	4.55	0.0036
emp*SPH	9	0.07947811	0.00883090	1.75	0.0741

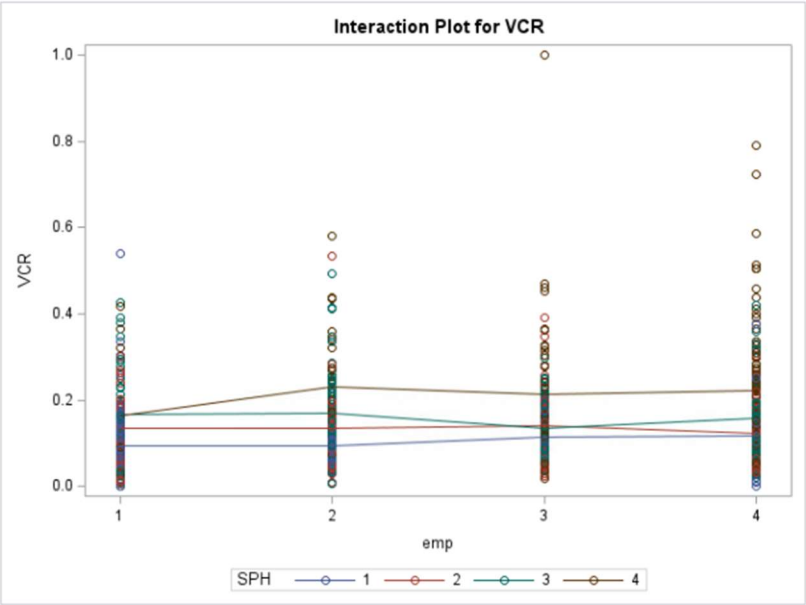


There are **two main effects**; the unemployment levels and single parent household level. **No interaction effects at α =0.01.**

- Percent of high school graduate depend on level of single parent household level
- Percent of high school graduate depend on unemployment level

Violent crime rate (VCR)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
emp	3	0.02128722	0.00709574	0.68	0.5648
SPH	3	0.58989690	0.19663230	18.83	<.0001
emp*SPH	9	0.08095006	0.00899445	0.86	0.5598



There are **one main effects**; the unemployment levels and single parent household level. **No interaction effects at α =0.01.**

- Violent crime rate graduate depends on level of single parent household level

Individual state

AK

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall emp*SPH Effect H = Type III SSCP Matrix for emp*SPH E = Error SSCP Matrix S=5 M=0.5 N=27.5					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.59500729	0.91	35	242.21	0.6193
Pillai's Trace	0.47563018	0.92	35	305	0.6085
Hotelling-Lawley Trace	0.56906586	0.91	35	150.03	0.6218
Roy's Greatest Root	0.25570365	2.23	7	61	0.0439
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					

Wilks' lambda results conclude **no significance** in **interactions**.

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall SPH Effect
H = Type III SSCP Matrix for SPH
E = Error SSCP Matrix

S=3 M=0.5 N=27.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.58234843	2.28	15	157.75	0.0062
Pillai's Trace	0.44444456	2.05	15	177	0.0143
Hotelling-Lawley Trace	0.67182733	2.51	15	102.55	0.0034
Roy's Greatest Root	0.59900620	7.07	5	59	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Wilks' lambda results conclude significance for **single parent household levels AK**

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall emp Effect
H = Type III SSCP Matrix for emp
E = Error SSCP Matrix

S=3 M=0.5 N=27.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.66675827	1.66	15	157.75	0.0635
Pillai's Trace	0.36969640	1.66	15	177	0.0631
Hotelling-Lawley Trace	0.44641372	1.67	15	102.55	0.0689
Roy's Greatest Root	0.27297522	3.22	5	59	0.0123

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Wilks' lambda results conclude no significance for **unemployment levels** in **AK**

GA:

Wilks' lambda results conclude **no significance** in **interactions 0.10 > α = 0.05.**

Wilks' lambda results conclude **significance** for **single parent household levels 0.008 < α = 0.05**.

Wilks' lambda results conclude **significance** for **unemployment levels** 0.03 **< α = 0.05**

MT:

Wilks' lambda results conclude **significance** in **interactions 0.0001< α = 0.05.**

Wilks' lambda results conclude NO **significance** for **single parent household levels**

Wilks' lambda results conclude NO **significance** for **unemployment levels**

OH

Wilks' lambda results conclude **NO significance** in **interactions 0.82 < α = 0.05.**

Wilks' lambda results conclude **significance** for **single parent household levels 0.0001 < α = 0.05**.

Wilks' lambda results conclude **significance** for **unemployment levels** 0.0008 **< α = 0.05**

OK

Wilks' lambda results conclude **NO significance** in **interactions 0.82 < α = 0.05.**

Wilks' lambda results conclude **significance** for **single parent household levels 0.0018 < α = 0.05.**

Wilks' lambda results conclude **significance** for **unemployment levels 0.001 < α = 0.05.**

TX

Wilks' lambda results conclude NO **significance** in **interactions 0.1798 < α = 0.05.**

Wilks' lambda results conclude significance for **single parent household levels 0.0001 < α = 0.05.**

Wilks' lambda results conclude **significance** for **unemployment levels 0.0002 < α = 0.05.**

Profile Analysis

Hypothesis testing

1. Are the profiles **parallel**?

   Equivalently: Is $H_{01}: \mu_{1i} - \mu_{1i-1} = \mu_{2i} - \mu_{2i-1}, i = 2,3, \dots, p, acceptable$

Code

```
proc glm;
class state;
model PCPR PSPH PS PO PU PHSG VCR = state/nouni;
manova h=state m=PCPR-PSPH, PSPH-PS, PS-PO, PO-PU, PU-PHSG, PHSG-VCR/ printe;
run;
```

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall state Effect on the Variables Defined by the M Matrix Transformation
H = Type III SSCP Matrix for state
E = Error SSCP Matrix

S=5 M=0 N=347.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.21204256	44.05	30	2790	<.0001
Pillai's Trace	1.16040065	35.31	30	3505	<.0001
Hotelling-Lawley Trace	2.21855957	51.45	30	1847.6	<.0001
Roy's Greatest Root	1.47922392	172.82	6	701	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

We **reject** the H_0 and conclude the profiles are **not parallel**.

2. Assuming that the profiles are *parallel*, are the profiles ***coincident***?

   Equivalently: Is $H_{02}: \mu_{1i} = \mu_{2i}, i = 1,2,3, \dots, p, acceptable$

code

```
proc glm;
class state;
model PCPR PSPH PS PO PU PHSG VCR = state/nouni;
manova h=state m=PCPR+PSPH+PS+PO+PU+PHSG+VCR/ printe;
run;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall state Effect on the Variables Defined by the M Matrix Transformation
H = Type III SSCP Matrix for state
E = Error SSCP Matrix

S=1 M=1.5 N=350

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.84496022	25.76	5	702	<.0001
Pillai's Trace	0.15503978	25.76	5	702	<.0001
Hotelling-Lawley Trace	0.18348767	25.76	5	702	<.0001
Roy's Greatest Root	0.18348767	25.76	5	702	<.0001

We **reject** the H_0 and conclude the profiles **are not coincident**.

3. Assuming that the profiles are coincident, are the profiles level? That is, are all the means equal to the same constant

$$H_{01}: \boldsymbol{C\mu_1 = C\mu_2}$$

Code

```
class state;
model x1-x7 = state/nouni;
manova h=intercept m=x1-x2, x2-x3, x3-x4, x4-x5, x5-x6, x6-x7;
run;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Intercept Effect on the Variables Defined by the M Matrix Transformation
H = Type III SSCP Matrix for Intercept
E = Error SSCP Matrix

S=1 M=2 N=347.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00432737	26728.5	6	697	<.0001
Pillai's Trace	0.99567263	26728.5	6	697	<.0001
Hotelling-Lawley Trace	230.08711747	26728.5	6	697	<.0001
Roy's Greatest Root	230.08711747	26728.5	6	697	<.0001

We **reject the H_0** and conclude that the profiles are **not level.**

Profile Plot

all variables have the same units of measurement

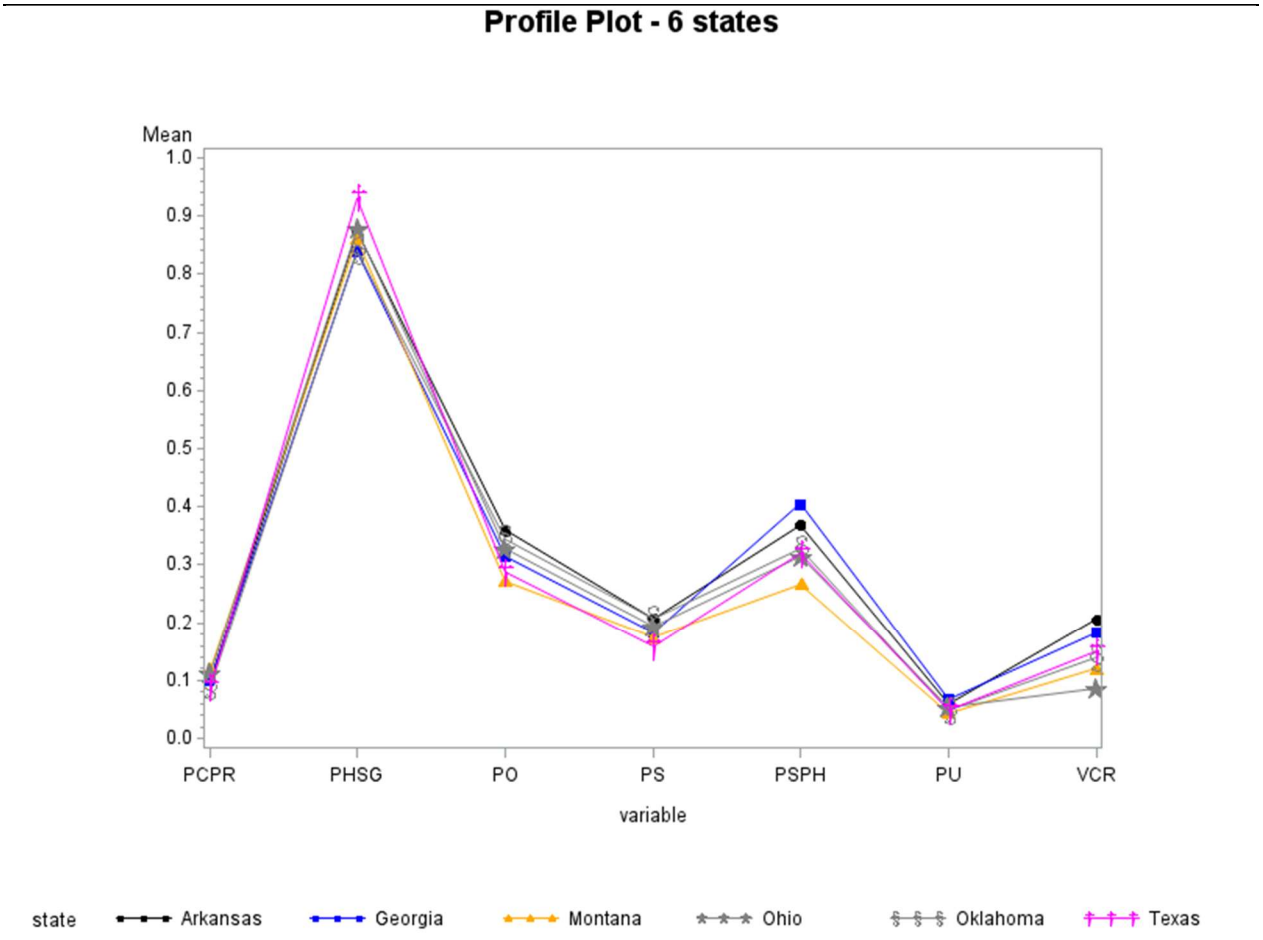
Top Code

```
data mydata;
  input state $ PCPR PSPH PS PO PU PHSG VCR;
  variable="PCPR  "; x=PCPR;   output;
  variable="PSPH  "; x=PSPH;   output;
  variable="PS    "; x=PS;     output;
  variable="PO    "; x=PO;       output;
  variable="PU    "; x=PU;     output;
  variable="PHSG  "; x=PHSG; output;
  variable="VCR   "; x=VCR;  output;

datalines;
Arkansas       0.115233069    0.390369331    0.204951914    0.338   0.043851287    0.845612245
       0.285257918…;
```

Bottom code

```
proc sort;
   by state variable;
  run;

 /* The means procedure calculates and saves mean for
  * each variable and saves the results in a new data set 'a'
  * for use in the steps below.
  */

proc means data=mydata;
  by state variable;
  var x;
  output out=a mean=xbar;
```

```
  run;

 /* The axis commands define the size of the plotting window.
  * The horizontal axis is of the variables, and the vertical
  * axis is used for the mean values.
  */

proc gplot;
  axis1 length=4 in label=("Mean");
  axis2 length=6 in;
  plot xbar*variable=state / vaxis=axis1 haxis=axis2;
  symbol1 v=J f=special h=2 i=join color=black;
  symbol2 v=K f=special h=2 i=join color=blues;
  symbol3 v=L f=special h=2 i=join color=browns;
  symbol4 v=M f=special h=2 i=join color=grays;
  symbol5 v=N f=special h=2 i=join color=olives;
  symbol6 v=O f=special h=2 i=join color=purples;
  run;
```



Between the states, the pairs of variables between **states seem** to **be parallel** to one another. However, In the previous profile analysis test, the null H_0 was rejected. Thus, concluding non parallel. Furthermore, the **test for interactions** led to the **rejecting of** H_0: no interactions. Concluding at least one interaction exist.

Principal Component Analysis

Overall

code

```
proc factor method=prin priors = one scree n = 6 out =new;
/*primary_care_physician_rate pct_single_parent_households pct_smokers
pct_obese pct_unemployed pct_high_school_graduation violent_crime_rate*/
var PCPR PSPH PS PO PU PHSG VCR;
run;
```

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.81689744	1.59457660	0.4024	0.4024
2	1.22232084	0.36391705	0.1746	0.5770
3	0.85840379	0.12877998	0.1226	0.6997
4	0.72962380	0.07473931	0.1042	0.8039
5	0.65488450	0.23917932	0.0936	0.8974
6	0.41570517	0.11354071	0.0594	0.9568
7	0.30216446		0.0432	1.0000

I decided to **take principal components** that were **above or close to 5% percent proportions.**

Factor Pattern						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
PCPR	0.06848	0.81649	-0.19741	0.51873	0.11176	0.08905
PSPH	0.75251	0.13480	0.41790	0.11646	-0.01508	-0.45841
PS	0.81968	-0.23571	-0.22749	0.13454	0.14013	-0.07376
PO	0.70705	-0.29811	-0.39855	-0.00850	0.38535	0.07848
PU	0.68462	-0.22998	0.43897	0.26889	-0.23392	0.39629
PHSG	-0.59940	-0.23080	0.42560	0.23076	0.59131	0.01906
VCR	0.49613	0.53563	0.24569	-0.55068	0.26394	0.16920

For pc 1: **percent single parent households, pct of smokers, percent of obesity, percent of unemployment** and **violent crime rate** are high. **Percent of high school graduate** are low.

For factor 2: **primary care physician rate, violent crime rate** are high. **Percent of smokers, Percent of obesity, percent of unemployment**, Percent **of high school graduate** are low.

For factor 3: **percent single parent households, percent of unemployment, percent of high school** are high. **Primary care physician rate Percent of smokers, Percent of obesity**, are low.

For factor 4: **primary care physician rate, percent of unemployment, percent of high school** are high. **Primary care physician rate and violent crime rate** are low.

For factor 5: **Percent of obesity, percent of high school** and **violent crime rate** are high. **Percent of unemployment and violent crime rate** are low.

For factor 6: **Percent of obesity, percent of high school** and **violent crime rate** are high. **Percent of unemployment and violent crime rate** are low.

Individual state

Code

```
proc sort; by state;
proc factor method=prin priors = one n = 6 out=new; by state;
var PCPR PSPH PS PO PU PHSG VCR;
run;
```

Principal Component Analysis

The FACTOR Procedure
Initial Factor Method: Principal Components

state=Arkansas

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.98313062	1.49141892	0.4262	0.4262
2	1.49171170	0.61565940	0.2131	0.6393
3	0.87605230	0.32979972	0.1252	0.7644
4	0.54625257	0.10443030	0.0780	0.8424
5	0.44182227	0.05761592	0.0631	0.9056
6	0.38420635	0.10738217	0.0549	0.9605
7	0.27682418		0.0395	1.0000

6 factors will be retained by the NFACTOR criterion.

Factor Pattern						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
PCPR	0.15378	-0.68122	0.67317	0.00069	0.02655	-0.22025
PSPH	0.82031	-0.04292	0.09847	0.42396	-0.20504	0.02794
PS	0.77207	0.34486	-0.13012	0.32956	0.18654	-0.16270
PO	0.78362	0.23526	-0.05766	-0.40589	0.18080	-0.30076
PU	0.62675	0.48247	0.36794	-0.25970	-0.31653	0.22345
PHSG	-0.44496	0.62045	0.50333	0.13313	0.34783	0.10055
VCR	0.69686	-0.48353	-0.06499	-0.08935	0.33229	0.39742

For factor 1: **percent single parent households, pct of smokers, percent of obesity, percent of unemployment** and **violent crime rate** are high. **Percent of high school graduate** is low.

For factor 2: **primary care physician rate, percent of unemployment, Percent of smokers,** and **Percent of high school graduate** are high. **Violent crime rate** is low.

For factor 3: **primary care physician rate,** and **percent of high school** are high.

For factor 4: **primary care physician rate, percent of unemployment, percent of high school** are high. **Primary care physician rate and violent crime rate** are low.

For factor 5: **Percent of obesity, percent of high school** and **violent crime rate** are high. **Percent of unemployment and violent crime rate** are low.

For factor 6: **Percent of obesity, percent of high school** and **violent crime rate** are high. **Percent of unemployment and violent crime rate** are low.

Principal Component Analysis

The FACTOR Procedure
Initial Factor Method: Principal Components

state=Ohio

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.42611990	0.32613559	0.3466	0.3466
2	2.09998431	1.18003000	0.3000	0.6466
3	0.91995431	0.34705241	0.1314	0.7780
4	0.57290190	0.16181007	0.0818	0.8599
5	0.41109183	0.06738879	0.0587	0.9186
6	0.34370303	0.11745832	0.0491	0.9677
7	0.22624472		0.0323	1.0000

6 factors will be retained by the NFACTOR criterion.

Factor Pattern						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
PCPR	-0.02078	0.81212	0.28282	0.22959	0.45185	-0.05051
PSPH	0.83401	0.32563	0.21050	0.07795	-0.15838	0.01454
PS	0.81778	-0.28179	0.17239	-0.14137	0.03764	-0.42373
PO	0.48387	-0.47097	-0.49441	0.53522	0.11374	0.01300
PU	0.70134	-0.45605	0.26809	-0.18225	0.19910	0.38272
PHSG	-0.35865	-0.48940	0.66917	0.39247	-0.16680	-0.01005
VCR	0.45467	0.76531	-0.04279	0.14290	-0.31626	0.12104

Ohio

For factor 1: **percent single parent households, pct of smokers, percent of obesity,** and **percent of unemployment** are high. **Percent of high school graduate** is low.

For factor 2: **primary care physician rate, Violent crime rate,** and **Percent of high school graduate** are high. **percent of unemployment,** and **Percent of smokers** are low.

For factor 3: **percent of high school** are high.

For factor 4 **Percent of obesity,** and **percent of high school** are high. Percent of obesity are low.

For factor 5: **Percent of primary care physician rate** are high. **Violent crime rate** is low.

For factor 6: **percent of unemployment** are high. **Percent of smokers** is low.

Georgia

The FACTOR Procedure
Initial Factor Method: Principal Components

state=Georgia

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.82345984	1.62978831	0.4034	0.4034
2	1.19367153	0.21683114	0.1705	0.5739
3	0.97684040	0.25587108	0.1395	0.7134
4	0.72096931	0.07874994	0.1030	0.8164
5	0.64221937	0.24189106	0.0917	0.9082
6	0.40032831	0.15781708	0.0572	0.9654
7	0.24251123		0.0346	1.0000

6 factors will be retained by the NFACTOR criterion.

Factor Pattern						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
PCPR	-0.16582	0.81298	0.30975	-0.37955	0.24571	0.03175
PSPH	0.81875	0.12151	0.30978	-0.06178	-0.06705	-0.37516
PS	0.87705	-0.13754	0.18501	0.08801	-0.04602	-0.13405
PO	0.64058	-0.10071	-0.35042	0.04070	0.67057	0.05142
PU	0.74038	-0.23318	0.27813	-0.31727	-0.13846	0.43963
PHSG	-0.41366	-0.24648	0.73695	0.36947	0.29009	0.06029
VCR	0.47622	0.61138	-0.08638	0.57142	-0.14916	0.20263

For factor 1: **percent single parent households, pct of smokers, percent of obesity,** and **percent of unemployment** are high. **Percent of high school graduate** is low.

For factor 2: **primary care physician rate,** and **Violent crime rate** are high. **percent of unemployment,** and **Percent of smokers** are low.

For factor 3: **percent of high school** are high. **Percent of obesity** is low

For factor 4: **violent crime rate** and **percent of high school graduate** are high. **Percent of primary care physician rate,** and **percent of unemployment** is low.

For factor 5: **Percent of obesity** are high. **Violent crime rate** is low.

For factor 6: **percent of single parent household is** high. **Percent of unemployment** is low.

Montana

The FACTOR Procedure
Initial Factor Method: Principal Components

state=Montana

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	3.20978077	1.95913591	0.4585	0.4585
2	1.25064485	0.32297535	0.1787	0.6372
3	0.92766950	0.41873787	0.1325	0.7697
4	0.50893164	0.02834083	0.0727	0.8424
5	0.48059081	0.08946577	0.0687	0.9111
6	0.39112504	0.15986766	0.0559	0.9670
7	0.23125738		0.0330	1.0000

6 factors will be retained by the NFACTOR criterion.

Factor Pattern						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
PCPR	0.25958	0.83261	0.27144	-0.16880	-0.35982	0.08761
PSPH	0.74708	0.28964	-0.02775	-0.35797	0.44300	-0.15752
PS	0.89022	-0.18368	0.05675	-0.04222	-0.05764	-0.04227
PO	0.76883	-0.36731	-0.07834	-0.14416	-0.06205	0.47525
PU	0.54613	-0.08123	0.74184	0.34725	0.11915	-0.02731
PHSG	-0.77849	0.26121	0.22821	-0.02763	0.34523	0.34765
VCR	0.54440	0.47962	-0.49138	0.45649	0.11967	0.09686

For factor 1: **percent single parent households, percent of smokers, percent of obesity, percent of unemployment, and violent crime** are high. **Percent of high school graduate** is low.

For factor 2: **primary care physician rate,** and **Violent crime rate** are high. **Percent of smokers** is low.

For factor 3: **percent of unemployment** are high. **Violent crime rate** is low

For factor 4: **violent crime rate** and **percent of unemployment** are high. **Percent single household parents** is low.

For factor 5: **Percent single household parents** are high. **Primary care physician rate** is low.

For factor 6: **percent of obesity** high

Texas

Principal Component Analysis

The FACTOR Procedure
Initial Factor Method: Principal Components

state=Texas

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.38706999	1.14818248	0.3410	0.3410
2	1.23888751	0.30724873	0.1770	0.5180
3	0.93163878	0.13253264	0.1331	0.6511
4	0.79910614	0.20081002	0.1142	0.7652
5	0.59829613	0.04042988	0.0855	0.8507
6	0.55786625	0.07073104	0.0797	0.9304
7	0.48713521		0.0696	1.0000

6 factors will be retained by the NFACTOR criterion.

Factor Pattern						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
PCPR	0.01540	0.79806	-0.17093	0.51259	0.11494	0.23891
PSPH	0.72163	-0.02147	-0.24024	0.31639	-0.07209	-0.49066
PS	0.76114	-0.25037	-0.05386	0.17145	0.06712	-0.01191
PO	0.52930	-0.01998	0.72155	0.20692	-0.33028	0.18312
PU	0.64699	-0.40654	-0.30702	-0.02914	0.21160	0.45092
PHSG	-0.52620	-0.44804	0.31485	0.47192	0.43777	-0.08073
VCR	0.55777	0.41525	0.35746	-0.37480	0.47946	-0.12853

For factor 1: **percent single parent households, percent of smokers, percent of obesity, percent of unemployment, and violent crime rate** are high. **Percent of high school graduate** is low.

For factor 2: **primary care physician rate** is high. **Percent of unemployment** and **Percent of high school graduate** are low.

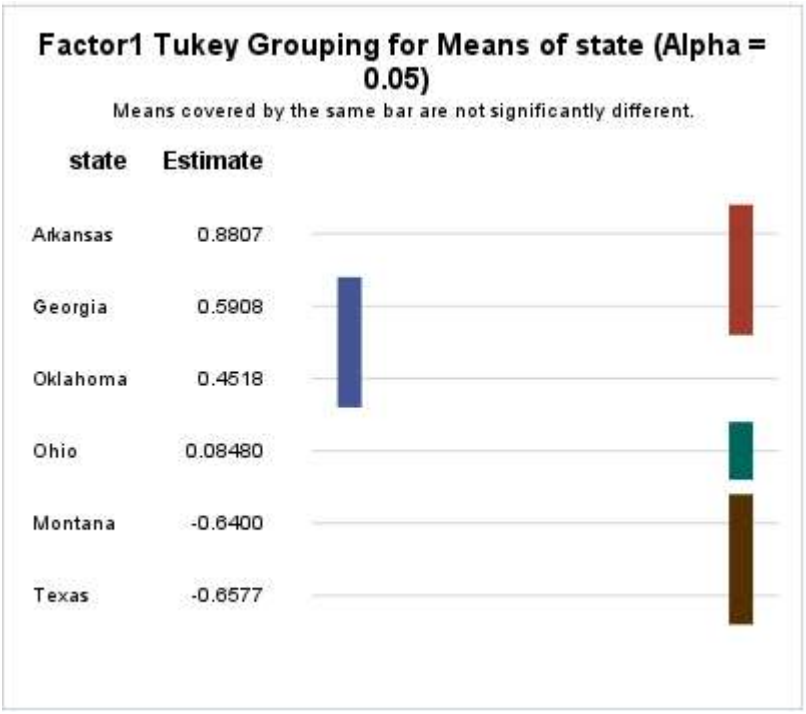
For factor 3: **Percent of obesity** are high. **Percent of unemployment** is low

For factor 4: **primary care physician rate**, and **percent high school graduate** are high. **Violent crime rate** is low

For factor 5. **Violent crime rate** is high. **percent of obesity** is low

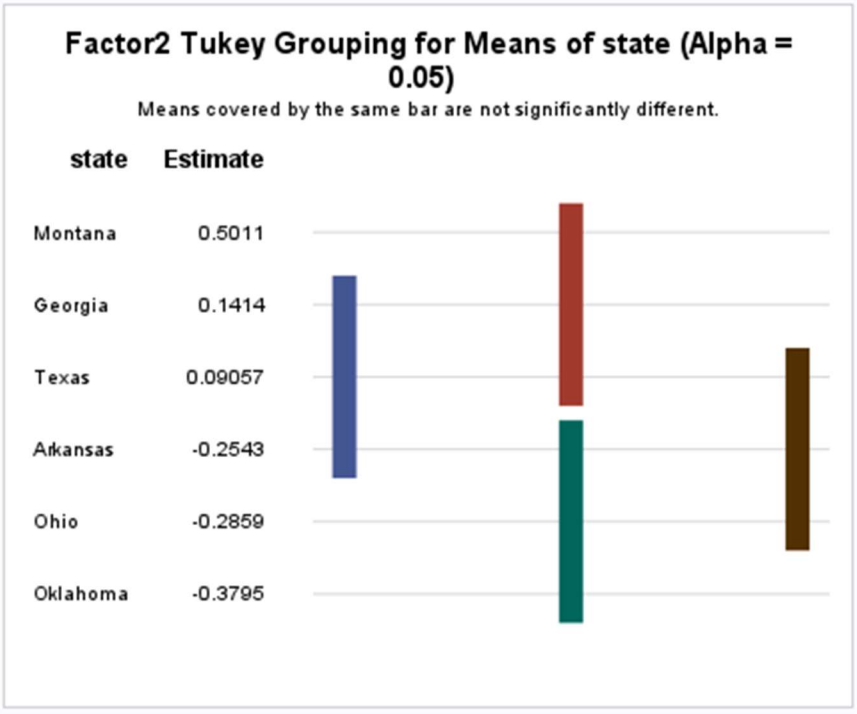
For factor 6: **percent of unemployment** high. **Percent single parent household is** low

Factor 1



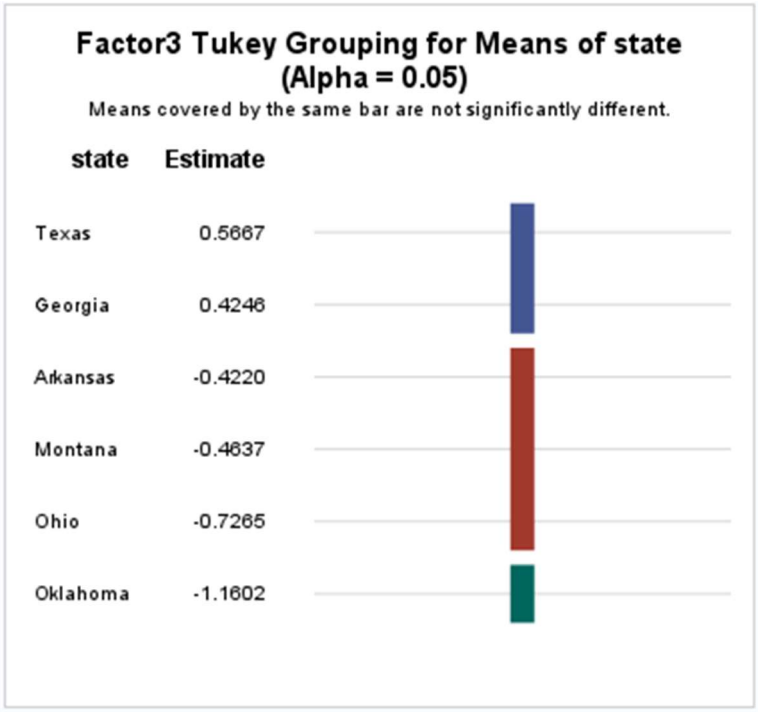
Factor1: MSE = 0.633; All means not covered by a common bar are significantly different. That is µAK are significantly different from µOK, µOH, µMT, µTX.

Factor 2



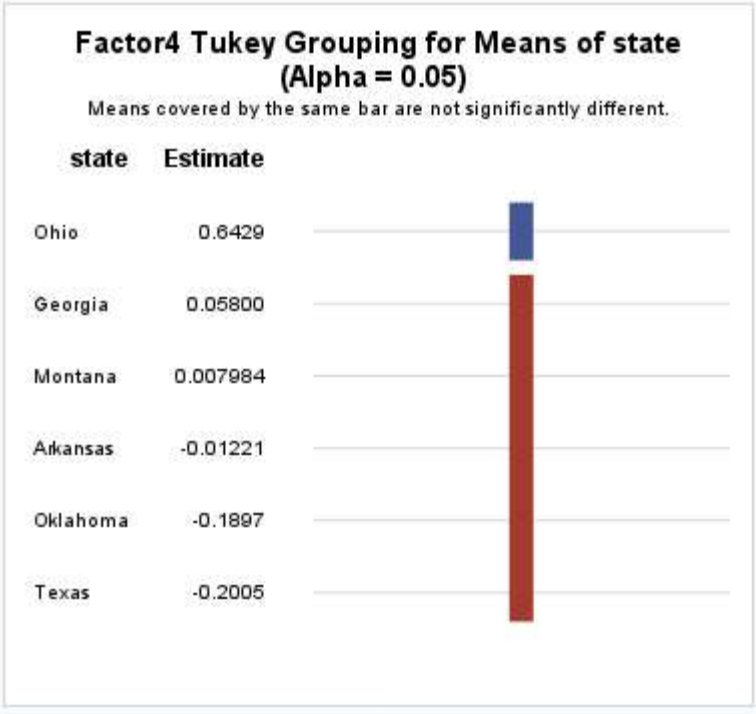
Factor2: MSE = 0.947; All means not covered by a common bar are significantly different. That is µAK are significantly different from µMT, µOH, µOK.

Factor 3



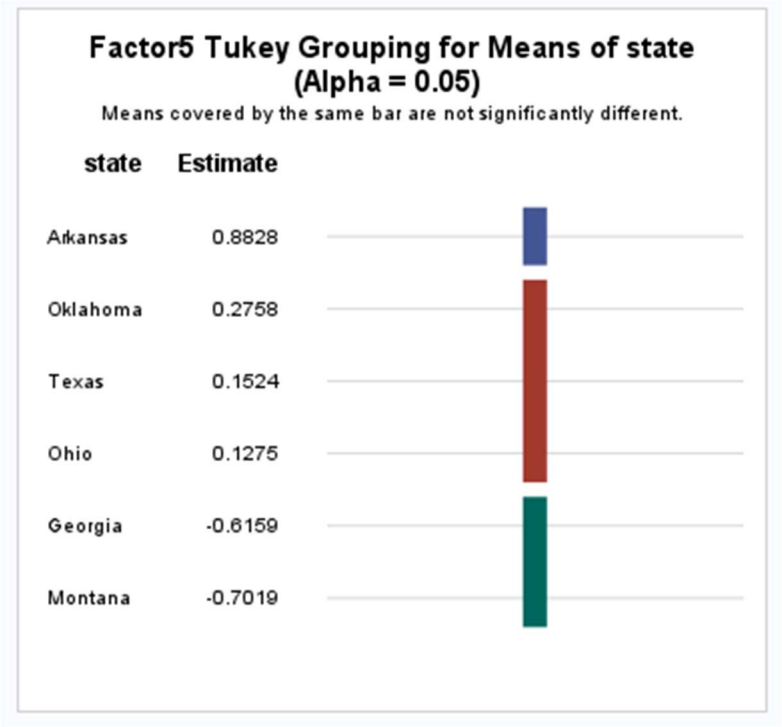
Factor 3: MSE = 0.601; All means not covered by a common bar are significantly different. That is µAK are significantly different from µGA, µTX, µOK.

Factor 4



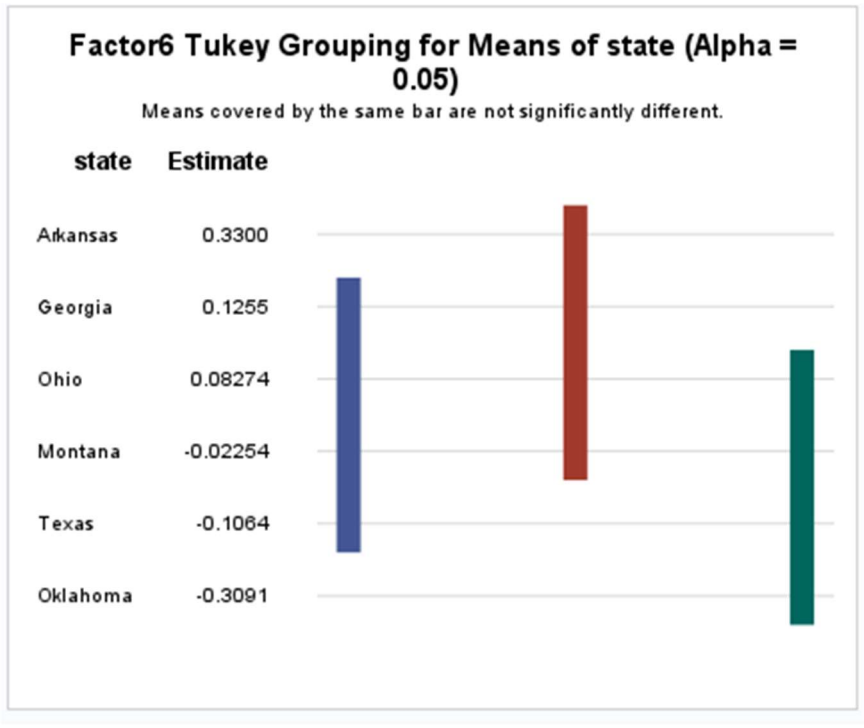
Factor 4: MSE = 0.9361; All means not covered by a common bar are significantly different. That is µAK are significantly different from µOH.

Factor 5



Factor 5: MSE = 0.7799; All means not covered by a common bar are significantly different. That is µAK are significantly different from µOK, µTX, µOH, µGA, µMT.

Factor 6



Factor 6: MSE = 0.9765; All means not covered by a common bar are significantly different. That is µAK are significantly different from µTX, and µOK.

Factor analysis

Goal: to describe the covariance relationship among variables in terms of a few factors which are unobservable

Principal Component Factor

top code

```
data mydata;
/*primary_care_physician_rate pct_single_parent_households pct_smokers
pct_obese pct_unemployed pct_high_school_graduation violent_crime_rate*/
input state $ PCPR PSPH PS PO PU PHSG VCR;
```

bottom code

```
proc factor method = prin priors =one rotate = varimax scree n =5 out=new;
var PCPR PSPH PS PO PU PHSG VCR;
run;
proc glm data=new;
class state;
model factor1-factor5 = state;
run;
```

Eigenvalues of the Correlation Matrix: Total = 7 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.81689744	1.59457660	0.4024	0.4024
2	1.22232084	0.36391705	0.1746	0.5770
3	0.85840379	0.12877998	0.1226	0.6997
4	0.72962380	0.07473931	0.1042	0.8039
5	0.65488450	0.23917932	0.0936	0.8974
6	0.41570517	0.11354071	0.0594	0.9568
7	0.30216446		0.0432	1.0000

Since there is only TWO λ_i>1, we only **interpret TWO factor**

Unrotated pattern

Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
PCPR	0.06848	0.81649	-0.19741	0.51873	0.11176
PSPH	0.75251	0.13480	0.41790	0.11646	-0.01508
PS	0.81968	-0.23571	-0.22749	0.13454	0.14013
PO	0.70705	-0.29811	-0.39855	-0.00850	0.38535
PU	0.68462	-0.22998	0.43897	0.26889	-0.23392
PHSG	-0.59940	-0.23080	0.42560	0.23076	0.59131
VCR	0.49613	0.53563	0.24569	-0.55068	0.26394

For factor 1 there is a contrast between percent of **poor health** to **low education attainment** high smoking (PS), high unemployment (PU), low percent of high school graduate (PHSG)

For factor 2 there is a contrast between percent of **poor health** to **high primary care physician rate (PSPH)**

Rotated = varimax

Rotated Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
PCPR	-0.03878	-0.01177	0.08444	0.98963	-0.06118
PSPH	0.21555	0.76206	0.35104	0.12140	-0.08781
PS	0.78399	0.39462	0.03886	0.01298	-0.21186
PO	0.92902	0.10704	0.10048	-0.05507	-0.09240
PU	0.19197	0.88486	-0.02820	-0.10126	-0.10212
PHSG	-0.20030	-0.13696	-0.13251	-0.06809	0.95683
VCR	0.08821	0.13614	0.95792	0.08408	-0.12382

For factor 1: the **rotated factor pattern** suggests the **similar pattern** high smoking (PS), high unemployment (PU), low percent of high school graduate (PHSG)

For factor 2: the **rotated factor pattern** seems to have high unemployment and single parent household

ANOVA factors results

Code

```
proc glm data=new;
class state;
model factor1-factor5 = state;
run;
```

Principal Factor Analysis

The GLM Procedure

Dependent Variable: Factor1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	395.9078804	79.1815761	178.68	<.0001
Error	702	311.0921196	0.4431512		
Corrected Total	707	707.0000000			

Principal Factor Analysis

The GLM Procedure

Dependent Variable: Factor2

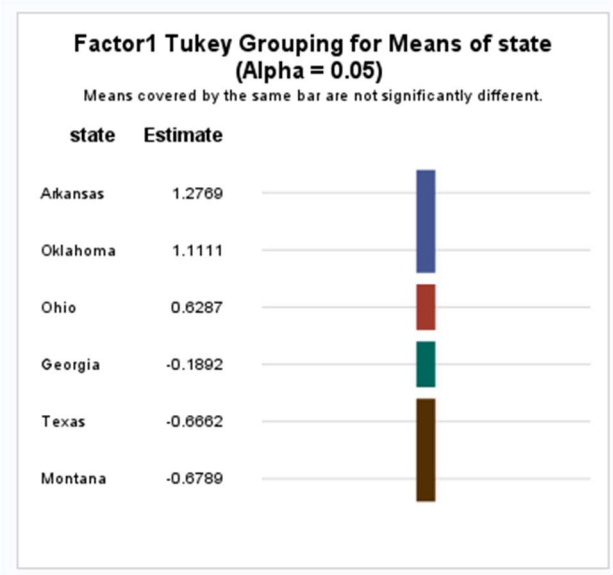
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	153.9529497	30.7905899	39.08	<.0001
Error	702	553.0470503	0.7878163		
Corrected Total	707	707.0000000			

The ANOVA procedure of the factor analysis by states for factor 1 and 2 **are significant.** Thus, at least one state differs in mean.

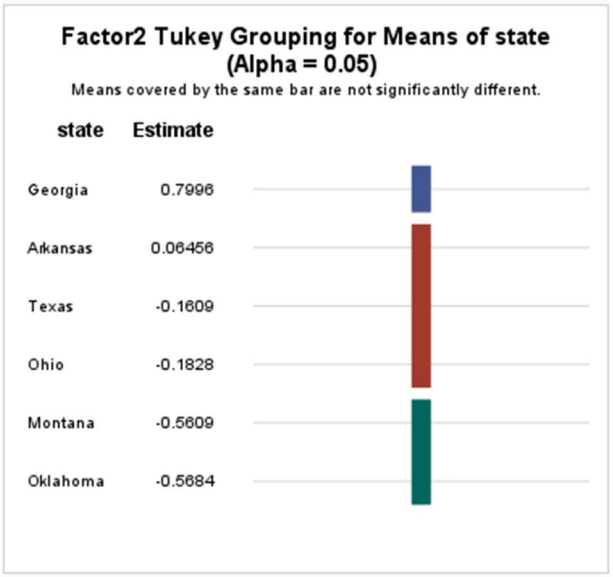
Post-hoc test

Code

```
proc glm data=new;
class state;
model factor1-factor2 = state;
means state/tukey lines;
run;
```



For factor 1, which consisted of **poor health** to **low education attainment, Arkansas** and **Oklahoma** leads.



For factor 2, which consisted of high unemployment and single parent household**, Georgia** leads.

Principal factor analysis

code

```
proc factor priors =smc rotate=promax n =5 out=new;
var PCPR PSPH PS PO PU PHSG VCR;
run;
```

Eigenvalues of the Reduced Correlation Matrix: Total = 2.2760317 Average = 0.32514739				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.22517681	1.78762481	0.9777	0.9777
2	0.43755200	0.22399591	0.1922	1.1699
3	0.21355609	0.28307691	0.0938	1.2637
4	-.06952082	0.02773249	-0.0305	1.2332
5	-.09725331	0.10317090	-0.0427	1.1905
6	-.20042421	0.03263064	-0.0881	1.1024
7	-.23305485		-0.1024	1.0000

Since there is only one λ_i>1, we only **interpret one factor**

Rotate = Promax

Rotated Factor Pattern (Standardized Regression Coefficients)			
	Factor1	Factor2	Factor3
PCPR	-0.04318	-0.08275	0.40993
PSPH	0.04817	0.63200	0.14564
PS	0.68619	0.18295	-0.04235
PO	0.73915	-0.01726	-0.05103
PU	0.08433	0.64144	-0.15214
PHSG	-0.38319	-0.02730	-0.28170
VCR	0.06108	0.20581	0.39500

the **rotated=promax factor pattern** suggests the **similar pattern** high smoking (PS), high unemployment (PU), low percent of high school graduate (PHSG)

MANOVA factor result

Code

```
proc glm data=new;
class state;
model factor1-factor5 = state;
manova h = da
run;
```

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall state Effect
H = Type III SSCP Matrix for state
E = Error SSCP Matrix

S=2 M=1 N=349.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.48422448	61.28	10	1402	<.0001
Pillai's Trace	0.57387613	56.50	10	1404	<.0001
Hotelling-Lawley Trace	0.94517098	66.19	10	1048.8	<.0001
Roy's Greatest Root	0.79406664	111.49	5	702	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

Since **wilks' lambda** is **significant**, we reject H_0 at α =0.01 level and conclude that state difference exists.

Conclusion

Through our practice in the multivariate analysis methods, we uncovered that **PSPH is non-normal** for combined samples. However, individual states AK, GA, MT, OH, and OK **are normal** besides Texas. Our bivariate normal analysis of the pairs of variables PSPH vs. PU, PSPH vs. PS, and PU vs. PS all resulted in non-normal data. Next, we tested for multivariate normal (MVN) using the gamma plot for the variables PSPH, PU, and PS, which led to non-normal data. In addition, when we applied the box-cox transformation, the bivariate normal and MVN tests still resulted in **non-normal.** The 2-sample Hotellings **test section** led to the **rejection of the mean vectors** for **all the state pairs**. One-way MANOVA on states concluded at least one treatment (PCPR, PSPH, PS, PO, PU, PHSG, VCR) difference exists. Two-way MANOVA on unemployment and single-parent house level resulted in no **interaction effect.** Only **the main effects were significant**. Analyzing each state for the two-way interactions resulted in state **MT** having interaction for the two factors, while other states had **only main effects**. Parallel analysis resulted in **rejecting H0: parallel profiles for all**. In the principal component analysis, we chose to analyze 6 principal components under the condition that the component be at least or close to 5 percent of the overall variance. Lastly, the principal component factor led to two factors to interpret using the varimax option while, for factor analysis using promax led to one factor to interpret.