More on Applying Principal Components & Discrimination Analysis to Covid-19

Jorge Luis Romeu, Ph.D.
https://www.researchgate.net/profile/Jorge_Romeu
http://web.cortland.edu/romeu/
Email: romeu@cortland.edu

Introduction

The present analysis is a continuation of our paper "*Multivariate Statistics in the Analysis of Covid-19 Data*" that was written in May 15, 2020, and can be found in my ResearchGate page: https://www.researchgate.net/publication/341385856_Multivariate_Stats_PC_Discrimination_in_the_Analysis_of_Covid-19.

In said paper we used metrics and data from New York State (NYS) Regions, to illustrate the use and power of multivariate statistical analysis in the fight against the Coronavirus Pandemic. We implemented Principal Components (PC) and Discrimination Analyses (DA) to explore the Regional impact of Covid-19 metrics, and to assess which metrics do a better job in separating the Regions into those that can safely open their economy, and those that should still wait.

In the present paper **we pursue further this exploration**, now using data **at the County level**[1]. We again implement Principal Components and Discrimination Analyses **to explore** the impact of county Covid-19 data **and (1) assess which metrics better separate** the 62 NYS counties, **and to (2) obtain a DA equation** that is able **to classify all NYS counties** into high and low risk groups, **according to their Covid-19 metrics**.

We use, as classification variables, (1) percent positives per 10K (of county population) and (2) percent deaths per 10K. As we do not have population density per county, we use (3) subjective predominantly Urban/Rural status for each county.

We start our analysis by **implementing Principal Components/Factorial analyses**, to establish (1) **which variables most significantly differentiate** among high and low infection counties, as well as to obtain PC scores as yet another variable to differentiate between counties. Then, using such variables, we (2) **develop several Discrimination Functions and classify counties into** two (**high and low risk) groups**, according to their Covid-19 metrics.

We present, in Table 1, NYS county data from Syracuse Post Standard NY Coronavirus Tracker

Table 1: Description of the NYS County metrics (Urban=1; Rural=0):

Row	County	Positives	Per 10K	Deaths	10K 1	Urban
1	Bronx	46052	321.6	3259	22.8	1
2	Brooklyn	57260	221.7	4934	19.1	1
3	Manhattan	27217	167.1	2040	12.5	1
4	Queens	63097	276.9	4969	21.8	1
5	Staten Island	13727	288.3	735	15.4	1

[1] The Syracuse Post Standard NY Coronavirus Tracker: https://www.syracuse.com/coronavirus-ny/

6	Albany	1978	64.4	115	3.7	1
7	Allegany	52	11.2	2	0.4	0
8	Broome	624	32.6	54	2.8	1
9	Cattaraugus	95	12.4	5	0.7	0
10	Cayuga	104	13.5	1	0.1	0
11	Chautauqua	101	7.9	6	0.5	0
12	Chemung	137	16.3	3	0.4	0
13	Chenango	135	28.4	5	1.1	0
14	Clinton	104	12.9	4	0.5	0
15	Columbia	414	69.1	42	7.0	1
16	Cortland	45	9.4	0	0.0	0
17	Delaware	85	19.1	4	0.9	0
18	Dutchess	4006	136.4	144	4.9	1
19	Erie	6531	71.0	549	6.0	1
20	Essex	53	14.2	0	0.0	0
21	Franklin	114	22.7	0	0.0	0
22	Fulton	227	42.4	25	4.7	1
23	Genesee	208	36.2	5	0.9	0
24	Greene	247	52.0	13	2.7	1
25	Hamilton	5	11.3	1	2.3	0
26	Herkimer	140	22.6	4	0.6	0
27	Jefferson	78	7.0	0	0.0	0
28	Lewis	26	9.8	0	0.0	0
29	Livingston	121	19.1	7	1.1	0
30	Madison	354	50.0	15	2.1	1
31	Monroe	3203	43.1	236	3.2	1
32	Montgomery	100	20.2	4	0.8	0
33	Nassau	40947	301.4	2145	15.8	1
34	Niagara	1116	53.0	74	3.5	1
35	Oneida	1170	51.0	61	2.7	1
36	Onondaga	2271	49.2	151	3.3	1
37	Ontario	221	20.1	28	2.5	1
38	Orange	10523	275.5	444	11.6	1
39	Orleans	256	63.0	43	10.6	1
40	Oswego	115	9.8	3	0.3	0
41	Otsego	75	12.6	5	0.8	0
42	Putnam	1451	146.7	62	6.3	1
43	Rensselaer	532	33.4	29	1.8	1
44	Rockland	13340	409.6	655	20.1	1
45	St. Lawrence	209	19.3	2	0.2	0
46	Saratoga	504	21.9	15	0.7	0
47	Schenectady	712	45.8	33	2.1	1
48	Schoharie	54	17.4	2	0.6	0
49	Schuyler	14	7.8	0	0.0	0
50	Seneca	61	17.8	2	0.6	0
51	Steuben	270	28.2	42	4.4	1
52	Suffolk	40377	272.6	1935	13.1	1
53	Sullivan	1417	187.7	35	4.6	1
54	Tioga	144	29.7	21	4.3	1
55	Tompkins	171	16.6	0	0.0	0
56	Ulster	1748	97.9	80	4.5	1
57	Warren	257	40.0	33	5.1	1
58	Washington	240	39.2	14	2.3	1
59	Wayne	125	13.9	4	0.4	0
60	Westchester	34000	351.4	1395	14.4	1
61	Wyoming	90	22.5	5	1.2	0
62	Yates	43	17.3	7	2.8	0

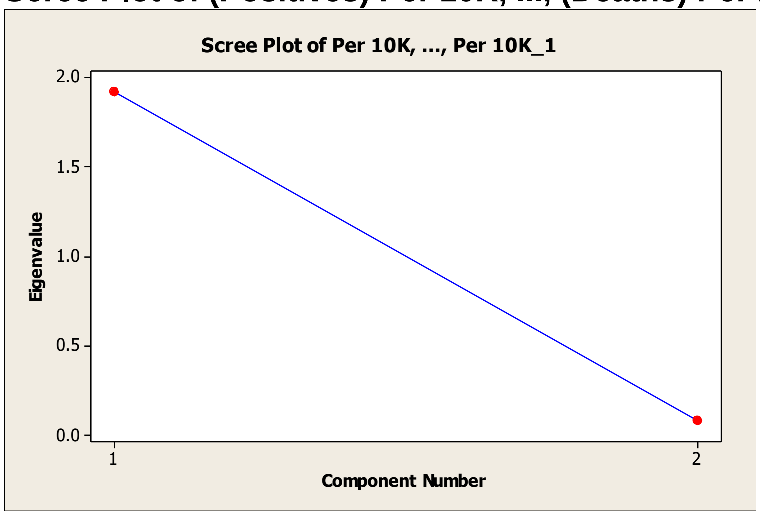
Principal Component Analyses to Assess Variables/Counties Relationship

We start by implementing a PC on ALL 62 NYS counties, for variables: (1) percent positives per 10K (of county population), denoted **Per 10K**, and (2) percent deaths per 10K, (also of a county population), denoted **Per 10K_1.**

Eigenanalysis of the Correlation Matrix

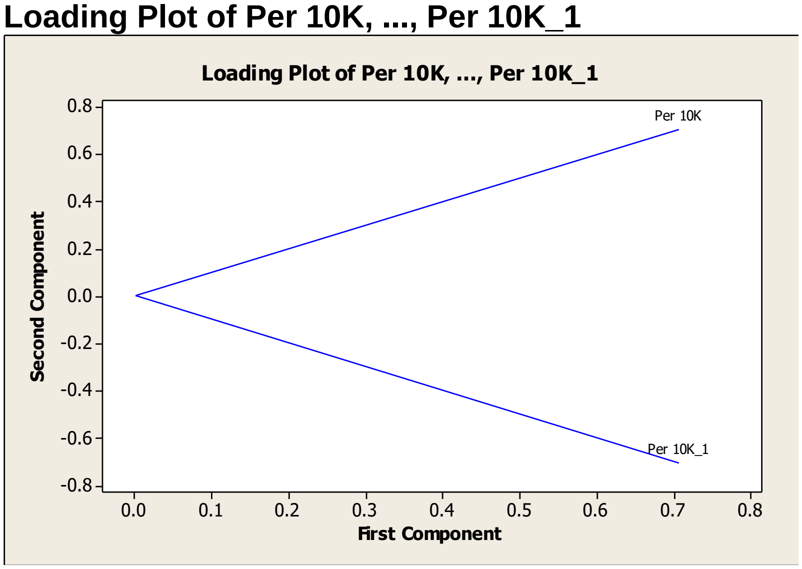
Eigenvalue	1.9212	0.0788
Proportion	**0.961**	**0.039**
Cumulative	0.961	1.000

Scree Plot of (Positives) Per 10K, ..., (Deaths) Per 10K_1



The above analysis for Two Principal Components and two variables shows how 96% of the Total explanation is given by the First Component, 4% by the Second, as corroborated by the Scree Plot. We decide to continue with the First Component only, given that its 96% percent variance explanation is overwhelmingly strong.

Loading Plot of Per 10K, ..., Per 10K_1



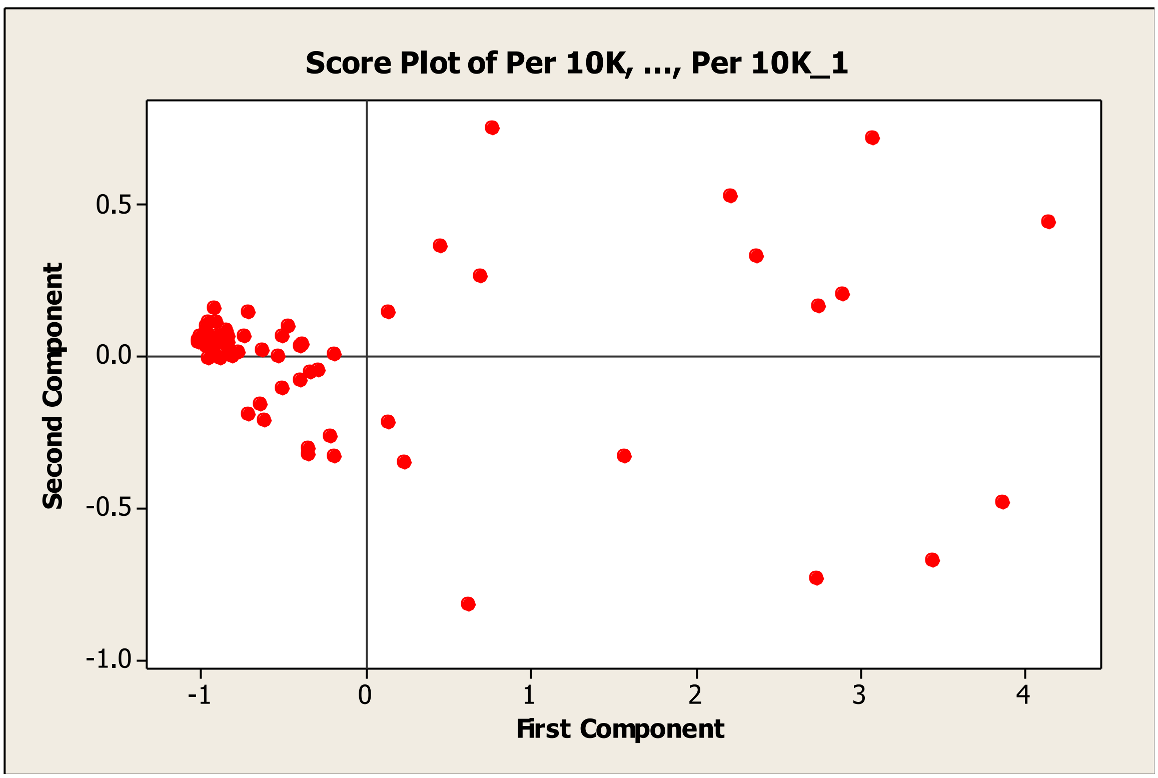
Loading Plot shows how variables Positives and Deaths are well differentiated by the two Principal Components.

Principal Components/Factor Loadings:

Variable	Meaning	PC1	PC2
PPer 10K	(Postvs/0000s)	0.707	0.707
DPer 10K_1	(Deaths/0000s)	0.707	-0.707

Using these PC coefficients we obtain the component scores to plot the different counties (Score Plot below) and assess how these counties are similar (clustered), or different, according to the two metrics (variables Positives and Deaths) analyzed.

Score Plot of Per 10K, ..., Per 10K_1



Notice how the counties with the lowest levels of infection (Positives) cluster to the left of the Origin (0,0). Those with the largest levels have high scores for their First Component. Below we show, per county, Positives and Deaths per 10K, PC scores and their urban/rural status.

Row	County	Pos/10K	D/10K	Score-1	Score-2	Urban
1	Bronx	321.6	22.8	3.85874	-0.483937	1
2	Brooklyn	221.7	19.1	2.72899	-0.735187	1
3	Manhattan	167.1	12.5	1.56809	-0.329056	1
4	Queens	276.9	21.8	3.43106	-0.674183	1
5	Staten Island	288.3	15.4	2.75008	0.164389	1
6	Albany	64.4	3.7	-0.18645	0.005787	1
7	Allegany	11.2	0.4	-0.94593	0.029835	0

8	Broome	32.6	2.8	-0.51310	-0.107167	1
9	Cattaraugus	12.4	0.7	-0.90202	0.002515	0
10	Cayuga	13.5	0.1	-0.96564	0.081347	0
11	Chautauqua	7.9	0.5	-0.95686	-0.004845	0
12	Chemung	16.3	0.4	-0.91068	0.065086	0
13	Chenango	28.4	1.1	-0.74394	0.065619	0
14	Clinton	12.9	0.5	-0.92230	0.029714	0
15	Columbia	69.1	7.0	0.23779	-0.353487	1
16	Cortland	9.4	0.0	-1.00585	0.064880	0
17	Delaware	19.1	0.9	-0.83196	0.025082	0
18	Dutchess	136.4	4.9	0.45366	0.360984	1
19	Erie	71.0	6.0	0.13221	-0.221639	1
20	Essex	14.2	0.0	-0.97268	0.098057	0
21	Franklin	22.7	0.0	-0.91393	0.156808	0
22	Fulton	42.4	4.7	-0.21980	-0.264989	1
23	Genesee	36.2	0.9	-0.71377	0.143275	0
24	Greene	52.0	2.7	-0.39088	0.038795	1
25	Hamilton	11.3	2.3	-0.71968	-0.195032	0
26	Herkimer	22.6	0.6	-0.84339	0.084888	0
27	Jefferson	7.0	0.0	-1.02244	0.048292	0
28	Lewis	9.8	0.0	-1.00309	0.067645	0
29	Livingston	19.1	1.1	-0.80822	0.001339	0
30	Madison	50.0	2.1	-0.47593	0.096200	1
31	Monroe	43.1	3.2	-0.39303	-0.082078	1
32	Montgomery	20.2	0.8	-0.83623	0.044556	0
33	Nassau	301.4	15.8	2.88811	0.207448	1
34	Niagara	53.0	3.5	-0.28899	-0.049265	1
35	Oneida	51.0	2.7	-0.39779	0.031883	1
36	Onondaga	49.2	3.3	-0.33900	-0.051787	1
37	Ontario	20.1	2.5	-0.63511	-0.157951	1
38	Orange	275.5	11.6	2.21049	0.527034	1
39	Orleans	63.0	10.6	0.62300	-0.823023	1
40	Oswego	9.8	0.3	-0.96747	0.032030	0
41	Otsego	12.6	0.8	-0.88876	-0.007974	0
42	Putnam	146.7	6.3	0.69105	0.265975	1
43	Rensselaer	33.4	1.8	-0.62628	0.017078	1
44	Rockland	409.6	20.1	4.14645	0.444838	1
45	St. Lawrence	19.3	0.2	-0.91368	0.109565	0
46	Saratoga	21.9	0.7	-0.83635	0.068178	0
47	Schenectady	45.8	2.1	-0.50496	0.067171	1
48	Schoharie	17.4	0.6	-0.87933	0.048946	0
49	Schuyler	7.8	0.0	-1.01691	0.053821	0
50	Seneca	17.8	0.6	-0.87656	0.051711	0
51	Steuben	28.2	4.4	-0.35356	-0.327523	1
52	Suffolk	272.6	13.1	2.36852	0.328917	1
53	Sullivan	187.7	4.6	0.77262	0.751177	1
54	Tioga	29.7	4.3	-0.35507	-0.305284	1
55	Tompkins	16.6	0.0	-0.95609	0.114646	0
56	Ulster	97.9	4.5	0.14007	0.142363	1
57	Warren	40.0	5.1	-0.18890	-0.329064	1
58	Washington	39.2	2.3	-0.52683	-0.002191	1
59	Wayne	13.9	0.4	-0.92726	0.048497	0
60	Westchester	351.4	14.4	3.06750	0.719243	1
61	Wyoming	22.5	1.2	-0.77285	0.012967	0
62	Yates	17.3	2.8	-0.61885	-0.212918	0

Description key:

Green: predominantly densely populated, urban counties

Yellow: predominantly low population, rural counties

No Color: intermediate (mixed Urban/Rural) counties

Variable Urban: (Valued One), if densely populated; (Valued Zero) if predominantly rural

Score Plot Clustering also reflects county population density. With respect to First Component predominantly Urban counties (in Green) are to the right hand side of the Plot (larger, positive scores), and they had the largest number of Covid-19 infections and fatalities. Predominantly Rural counties (Yellow) are clustered together (with negative values) in the left hand side of the Plot. Mixed counties fall between these two groups. Such clustering supports the propensity to have more infection and deaths, as the county density increases.

With respect to the Second Component, differentiation is not as clear (this component describes only 4% of the problem). For Urban counties, negative values of the Second Component point to the most densely populated (e.g. inner cities such as the Bronx, Queens, Brooklyn), and positive values point to suburban counties (e.g. Nassau, Orange, Westchester).

These groupings are in consonance with the current Covid-19 infection results.

Factor Analysis: (All counties) Per 10K, Per 10K_1

We now implement a Factor Analysis with the same data as above, perform a **Varimax rotation**, (that maximizes variable projections in the new axis) and compare their results.

Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
Per 10K	0.980	0.198	1.000
Per 10K_1	0.980	-0.198	1.000
Variance	1.9212	0.0788	2.0000
% Var	**0.961**	**0.039**	**1.000**

Rotated Factor Loadings and Communalities: Varimax Rotation

Variable	Factor1	Factor2	Communality
Per 10K	0.833	-0.553	1.000
Per 10K_1	0.553	-0.833	1.000
Variance	1.0000	1.0000	2.0000
% Var	0.500	0.500	1.000

Factor Score Coefficients

Variable	Factor1	Factor2
Per 10K	2.142	1.420
Per 10K_1	-1.420	-2.142

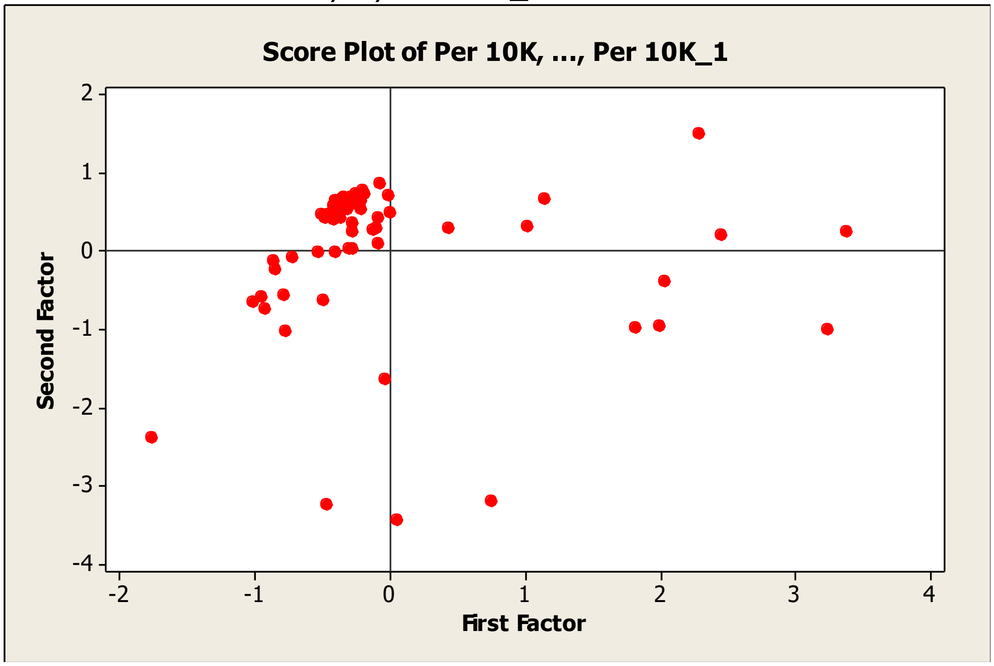
Factor1 describes 96% of the variance. **Communalities** are all unit, because there are only two variables and two factors (they can be less than unit if the number of variables is larger than that of extracted Factors). **Factor loadings** are correlation coefficients established between variables and **common factors,** and determine **factor** influence on each variable. **Loadings** close to -1 or +1 indicate strong **factor** influence. If they are close to 0, it shows a weak influence on the variable. Loading of 0.83 indicates a strong influence of **Factor1** on variable **Positives Per 10K**.

County Scores obtained with Rotated PC Coefficients:

Row	County	Per 10K	Per 10K_1	Scr-1	Scr-2
1	Bronx	321.6	22.8	0.74954	-3.18754
2	Brooklyn	221.7	19.1	-0.45968	-3.24408
3	Manhattan	167.1	12.5	-0.02890	-1.62883
4	Queens	276.9	21.8	0.05215	-3.44857
5	Staten Island	288.3	15.4	1.81704	-0.98887
6	Albany	64.4	3.7	-0.08054	0.10970
7	Allegany	11.2	0.4	-0.40741	0.55772
8	Broome	32.6	2.8	-0.53170	-0.00819
9	Cattaraugus	12.4	0.7	-0.45383	0.46650
10	Cayuga	13.5	0.1	-0.28772	0.69753
11	Chautauqua	7.9	0.5	-0.50035	0.47594
12	Chemung	16.3	0.4	-0.30064	0.62853
13	Chenango	28.4	1.1	-0.21423	0.54481
14	Clinton	12.9	0.5	-0.39567	0.54536
15	Columbia	69.1	7.0	-0.76909	-1.01171
16	Cortland	9.4	0.0	-0.34971	0.67656
17	Delaware	19.1	0.9	-0.36125	0.48761
18	Dutchess	136.4	4.9	1.14072	0.67785
19	Erie	71.0	6.0	-0.49084	-0.62574
20	Essex	14.2	0.0	-0.24921	0.74321
21	Franklin	22.7	0.0	-0.07125	0.86123
22	Fulton	42.4	4.7	-0.77962	-0.55535
23	Genesee	36.2	0.9	-0.00323	0.72503
24	Greene	52.0	2.7	-0.10168	0.29713
25	Hamilton	11.3	2.3	-0.85841	-0.12413
26	Herkimer	22.6	0.6	-0.21643	0.64408
27	Jefferson	7.0	0.0	-0.39996	0.64324
28	Lewis	9.8	0.0	-0.34134	0.68212
29	Livingston	19.1	1.1	-0.40894	0.41569
30	Madison	50.0	2.1	-0.00048	0.48512
31	Monroe	43.1	3.2	-0.40726	-0.00624
32	Montgomery	20.2	0.8	-0.31437	0.53884
33	Nassau	301.4	15.8	1.99592	-0.95083
34	Niagara	53.0	3.5	-0.27152	0.02334
35	Oneida	51.0	2.7	-0.12262	0.28324
36	Onondaga	49.2	3.3	-0.30339	0.04249
37	Ontario	20.1	2.5	-0.72187	-0.07386
38	Orange	275.5	11.6	2.45524	0.19987
39	Orleans	63.0	10.6	-1.75530	-2.39095
40	Oswego	9.8	0.3	-0.41288	0.57424
41	Otsego	12.6	0.8	-0.47349	0.43332
42	Putnam	146.7	6.3	1.02251	0.31743
43	Rensselaer	33.4	1.8	-0.27648	0.36252
44	Rockland	409.6	20.1	3.23583	-0.99481
45	St. Lawrence	19.3	0.2	-0.19013	0.74210
46	Saratoga	21.9	0.7	-0.25493	0.59840
47	Schenectady	45.8	2.1	-0.08841	0.42680
48	Schoharie	17.4	0.6	-0.32530	0.57188
49	Schuyler	7.8	0.0	-0.38321	0.65435
50	Seneca	17.8	0.6	-0.31693	0.57744
51	Steuben	28.2	4.4	-1.00537	-0.64463
52	Suffolk	272.6	13.1	2.03682	-0.37979
53	Sullivan	187.7	4.6	2.28631	1.49800
54	Tioga	29.7	4.3	-0.95012	-0.58785
55	Tompkins	16.6	0.0	-0.19897	0.77653
56	Ulster	97.9	4.5	0.43005	0.28715
57	Warren	40.0	5.1	-0.92525	-0.73251
58	Washington	39.2	2.3	-0.27428	0.26325
59	Wayne	13.9	0.4	-0.35088	0.59521
60	Westchester	351.4	14.4	3.37661	0.24682

61	Wyoming	22.5	1.2	-0.36161	0.42693
62	Yates	17.3	2.8	-0.85203	-0.22062

Score Plot of Per 10K, ..., Per 10K_1



The big cluster left of the origin (0,0) corresponds to the predominantly rural counties (no color). In Green are the predominantly urban counties (Bronx, Brooklyn), with larger negative scores. In Yellow are the suburbs (Nassau, Westchester), with larger positive scores. Below, we compare the factor coefficients for the two score plots shown above, and the loadings.

Comparison of Factor Score Coefficients: before/after rotation:

Row	Coef-1	Coef-2	Cf-1	Cf-2	Load-1	Load-2
Pos	0.707107	0.707107	2.14188	1.42041	0.833396	-0.552677
Dth	0.707107	-0.707107	-1.42041	-2.14188	0.552677	-0.833396

Analysis considering Tri-Groups (-1,0,1), Extreme Groups only and (Per 10K_2, Per 10K_1_1):

We now *create three county groups* denoted predominantly urban (1), predominantly rural (-1), and mixed (0). We *repeat the above analysis* using only the first two (-1,1), and *leave out mixed group (0)* to compensate for *missing population density data*. We are not sure of the placing of *counties in the mixed group*. Leaving mixed group out will *strengthen PC analysis* results. We will *use them to derive* discrimination equations to position the mixed (0) group, accordingly.

Principal Components Analysis

Eigenanalysis of the Correlation Matrix

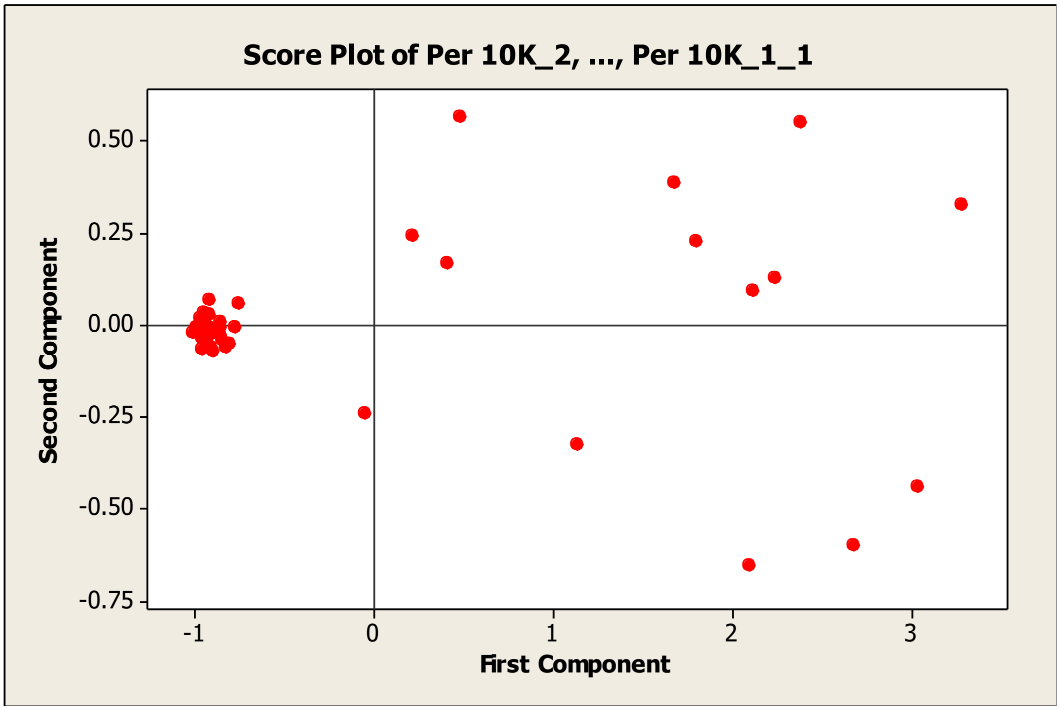
Eigenvalue	1.9428	0.0572
Proportion	**0.971**	**0.029**
Cumulative	0.971	1.000

Variable	PC1	PC2

```
Per 10K_2     0.707   0.707
Per 10K_1_1   0.707  -0.707
```

Comparing the results with those obtained using the entire 62 counties we observe that Percent Variance, for the First Component, has barely increased, from 96% to 97%. The PC Coefficients remain similar, Score plot (below) is also similar, just that clusters are now tighter.

Score Plot of Per 10K_2, ..., Per 10K_1_1



We also redo the Factor Analysis with the two extreme county groups, and compare its results.

Factor Analysis for only Extreme TriGroups (-1,1); with Variables (Per 10K_2, Per 10K_1_1):

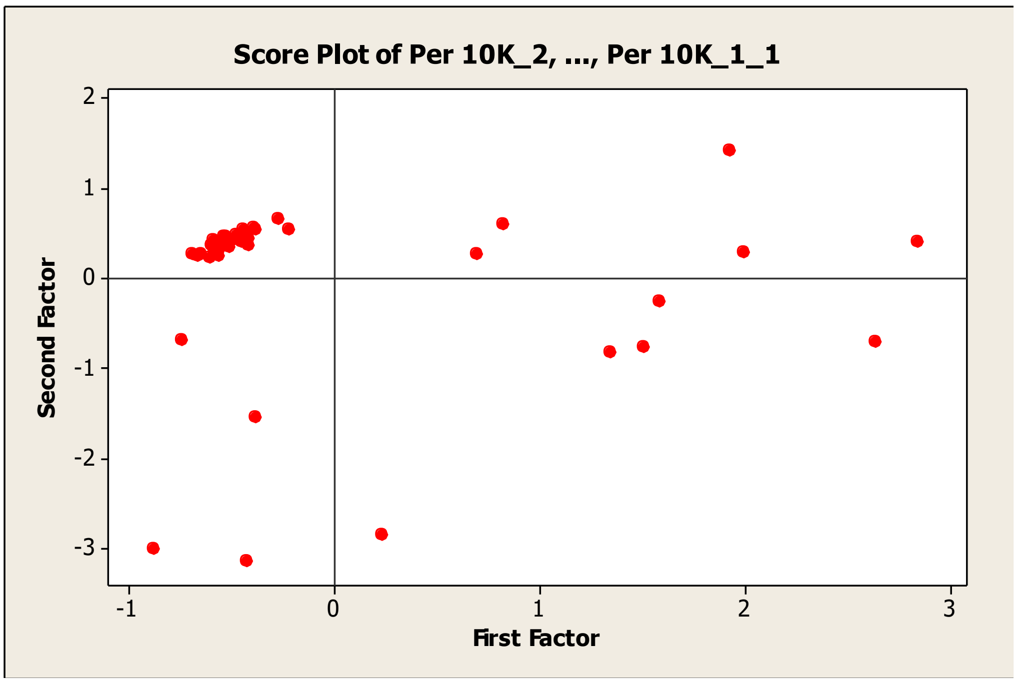
```
Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable      Factor1  Factor2  Communality
Per 10K_2       0.986    0.169        1.000
Per 10K_1_1     0.986   -0.169        1.000

Variance       1.9428   0.0572       2.0000
% Var           0.971    0.029        1.000

Rotated Factor Loadings and Communalities: Varimax Rotation

Variable      Factor1  Factor2  Communality
Per 10K_2       0.817   -0.577        1.000
Per 10K_1_1     0.577   -0.817        1.000

Variance       1.0000   1.0000       2.0000
% Var           0.500    0.500        1.000
```

Factor Score Coefficients

Variable	Factor1	Factor2
Per 10K_2	2.449	1.731
Per 10K_1_1	-1.731	-2.449

Comparing results with those obtained using all 62 counties: (1) Percent Variance, for First Component has barely increased from 96% to 97%; (2) PC Coefficients remain similar; (3) Score plot (below) is also similar, just now tighter. *The difference will be in the PC scores.*

Score Plot of Per 10K_2, ..., Per 10K_1_1



Data Display

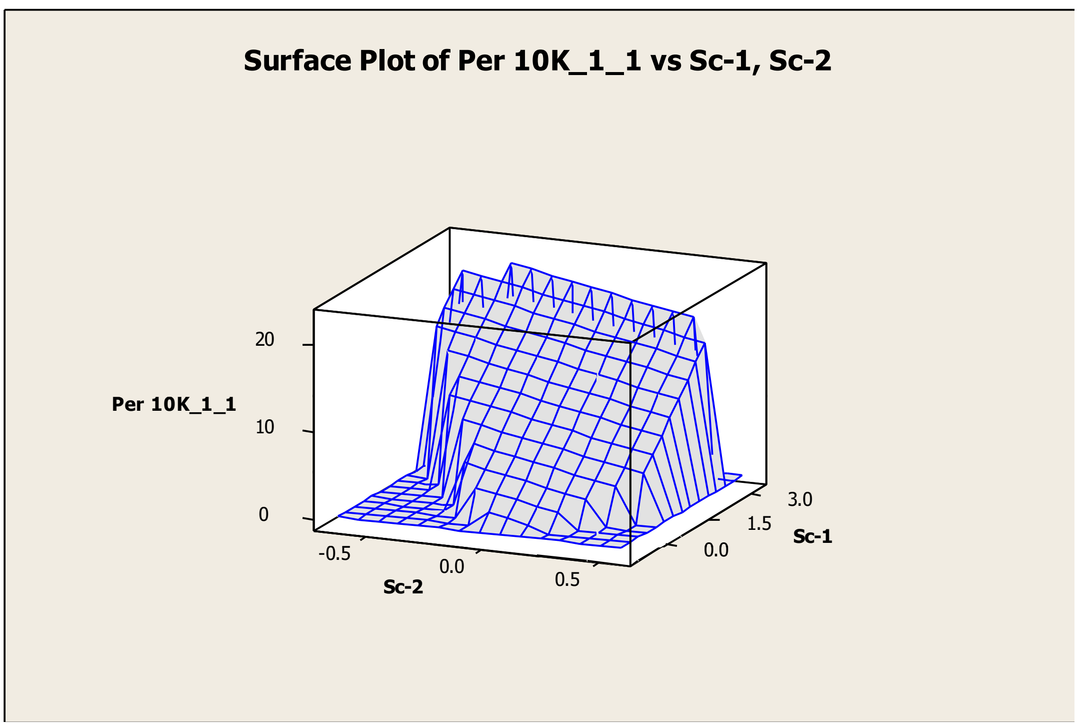
Row	County_1	Per 10K_2	Per 10K_1_1	Sc-1	Sc-2	RScr-1	RScr-2
1	Bronx	321.6	22.8	3.03319	-0.442175	0.23185	-2.84569
2	Brooklyn	221.7	19.1	2.09626	-0.655477	-0.87391	-3.00082
3	Manhattan	167.1	12.5	1.13654	-0.324408	-0.38226	-1.53542
4	Queens	276.9	21.8	2.67807	-0.601721	-0.41987	-3.13710
5	Staten Island	288.3	15.4	2.11786	0.089746	1.33967	-0.80915
6	Allegany	11.2	0.4	-0.94419	-0.038683	-0.59334	0.36467
7	Cattaraugus	12.4	0.7	-0.90795	-0.061111	-0.64124	0.27999
8	Cayuga	13.5	0.1	-0.96029	0.003894	-0.47566	0.49867
9	Chautauqua	7.9	0.5	-0.95341	-0.067460	-0.68307	0.28429
10	Chemung	16.3	0.4	-0.91483	-0.009323	-0.49166	0.43655
11	Chenango	28.4	1.1	-0.77672	-0.008116	-0.41803	0.37005
12	Clinton	12.9	0.5	-0.92463	-0.038675	-0.58339	0.35476
13	Cortland	9.4	0.0	-0.99367	-0.009931	-0.53345	0.47475
14	Delaware	19.1	0.9	-0.84982	-0.042097	-0.55555	0.30670
15	Dutchess	136.4	4.9	0.21661	0.242039	0.82527	0.60549
16	Erie	71.0	6.0	-0.05232	-0.242029	-0.74190	-0.68881
17	Essex	14.2	0.0	-0.96604	0.017702	-0.43776	0.54240
18	Franklin	22.7	0.0	-0.91710	0.066636	-0.26830	0.66221
19	Genesee	36.2	0.9	-0.75138	0.056346	-0.21464	0.54772
20	Herkimer	22.6	0.6	-0.85901	0.007388	-0.41395	0.45762

21	Jefferson	7.0	0.0	-1.00749	-0.023747	-0.58130	0.44092
22	Lewis	9.8	0.0	-0.99137	-0.007628	-0.52548	0.48039
23	Livingston	19.1	1.1	-0.83026	-0.061655	-0.60343	0.23897
24	Montgomery	20.2	0.8	-0.85327	-0.025986	-0.50968	0.35607
25	Nassau	301.4	15.8	2.23239	0.126047	1.50507	-0.75996
26	Orange	275.5	11.6	1.67258	0.387650	1.99428	0.29724
27	Oswego	9.8	0.3	-0.96203	-0.036964	-0.59730	0.37880
28	Otsego	12.6	0.8	-0.89702	-0.069739	-0.66119	0.24895
29	Putnam	146.7	6.3	0.41281	0.164433	0.69543	0.27658
30	Rockland	409.6	20.1	3.27577	0.328459	2.63265	-0.69102
31	St. Lawrence	19.3	0.2	-0.91712	0.027505	-0.38397	0.54656
32	Saratoga	21.9	0.7	-0.85326	-0.006421	-0.45185	0.41389
33	Schoharie	17.4	0.6	-0.88894	-0.022548	-0.51762	0.38433
34	Seneca	17.8	0.6	-0.88664	-0.020245	-0.50964	0.38997
35	Suffolk	272.6	13.1	1.80256	0.224274	1.57734	-0.25158
36	Sullivan	187.7	4.6	0.48261	0.566705	1.91982	1.43015
37	Tompkins	16.6	0.0	-0.95222	0.031519	-0.38991	0.57623
38	Wayne	13.9	0.4	-0.92865	-0.023140	-0.53951	0.40272
39	Westchester	351.4	14.4	2.38333	0.550795	2.83705	0.41887
40	Wyoming	22.5	1.2	-0.80091	-0.051860	-0.55959	0.25303

In Green are the *high density (urban)* counties (group 1); the *rural counties* are in No Color.

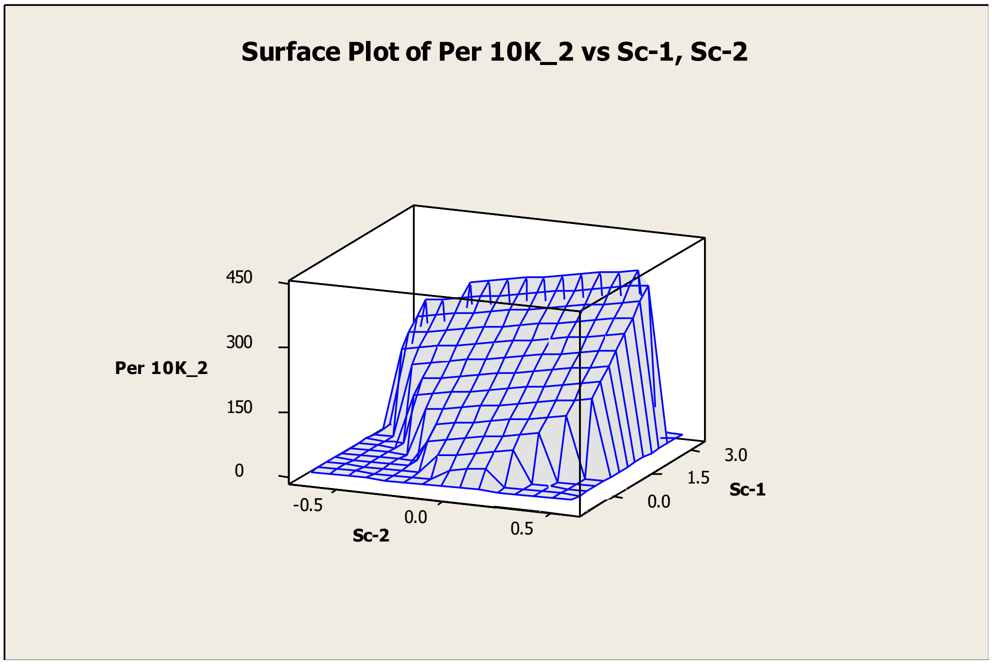
We show below three 3D surface plots that illustrate the relationship between the *variables* (Positives and Deaths) and all the *counties* respective scores, (i.e. the two factor components).

Deaths vs. Principal Component Scores one and two



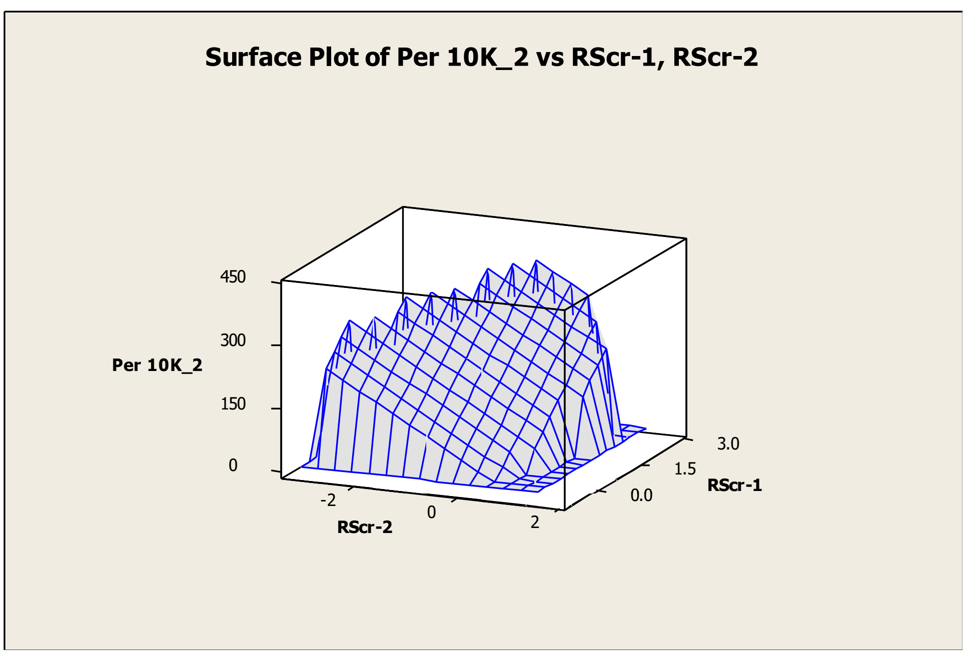
In the graph above we observe how Positive tests peak for scores greater than 1.5 on the First Component, and decrease as these are smaller than 1.5. There are two regions, in the corners of the graph, where there are no Positives.

Deaths vs. Principal Component Scores one and two



As in the graph above we observe how Deaths peak for scores greater than 1.5 on the First Component, and decrease as they are smaller than 1.5. There are two regions, in the corners of the graph, where there are no deaths.

Deaths vs. Principal Component Rotated Scores one and two



Deaths behavior is similar, when using scores obtained from the Varimax rotation.

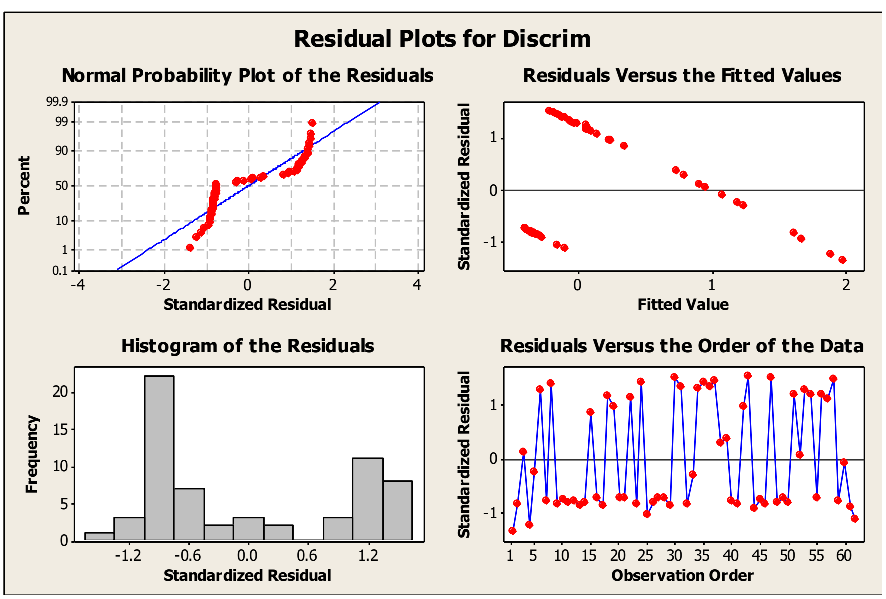
<u>**Discriminant Analysis: high v. low population density, and vars Positives and Deaths**</u>

We divided all 62 <u>NYS counties into two groups</u>, *highly infected*, *urban* (e.g. Manhattan, Bronx) and *less infected*, *predominantly rural* (e.g. Lewis, Thompkins) *based upon Positives and Deaths per 10K residents and our* own experience. This division allows implementing a Discrimination analysis to Covid-19. It could be *better if* it was *based upon county density* (that we do not have).

<u>We used Regression approach to Discrimination Analysis</u>. We regressed the 62 County variables vs. -1 and +1 (refer to Table 1) according to which of the two groups each had been assigned to. Additional Covid-19 metric/variables could be included, if available to the investigator.

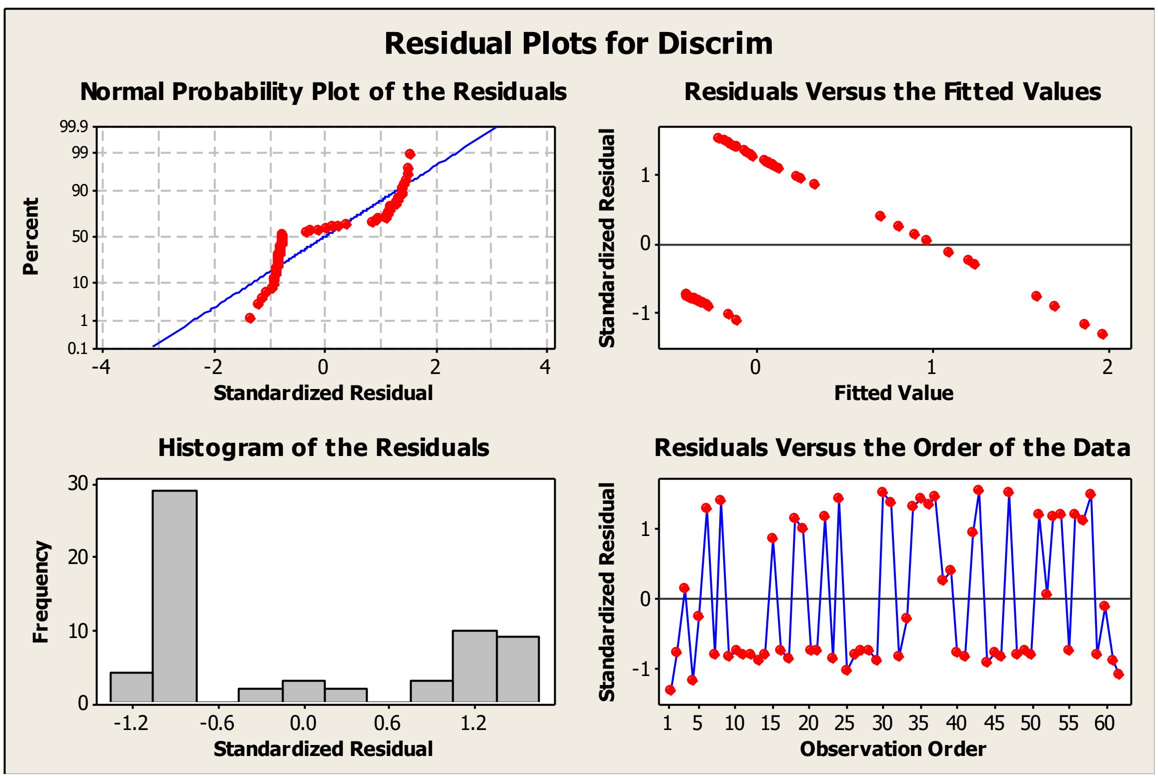
Regression Analysis: Discrim (All counties) versus Per 10K, Per 10K_1

```
The regression equation is:

Discrim = - 0.405 - 0.00022 Per 10K + 0.108 Per 10K_1

Predictor        Coef    SE Coef      T      P
Constant      -0.4046     0.1295  -3.12  0.003
Per 10K     -0.000217   0.002588  -0.08  0.933
Per 10K_1     0.10775    0.04445   2.42  0.018

S = 0.804503   R-Sq = 38.2%   R-Sq(adj) = 36.1% (not very explanatory of the problem)

Unusual Observations (counties that are out of "expected" values)

Obs  Per 10K  Discrim    Fit  SE Fit  Residual  St Resid
  1      322    1.000  1.982   0.352    -0.982    -1.36 X
  2      222    1.000  1.605   0.353    -0.605    -0.84 X
  4      277    1.000  1.884   0.370    -0.884    -1.24 X
 39       63    1.000  0.724   0.322     0.276     0.37 X
 44      410    1.000  1.672   0.363    -0.672    -0.94 X
 60      351    1.000  1.071   0.363    -0.071    -0.10 X

X denotes an observation whose X value gives it large influence.
```



Regression Analysis: Discrim (All counties) versus Per 10K_1 (Positives)

```
The regression equation is: Discrim = - 0.406 + 0.104 Per 10K_1

Predictor      Coef  SE Coef      T      P
Constant    -0.4059   0.1275  -3.18  0.002
Per 10K_1   0.10432  0.01715   6.08  0.000

S = 0.797818   R-Sq = 38.1%   R-Sq(adj) = 37.1% (not very explanatory of the problem)

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  23.551  23.551  37.00  0.000
Residual Error  60  38.191   0.637
Total           61  61.742

Unusual Observations

       Per
Obs  10K_1  Discrim    Fit  SE Fit  Residual  St Resid
  1   22.8    1.000  1.973   0.330    -0.973     -1.34 X
  2   19.1    1.000  1.587   0.270    -0.587     -0.78 X
  4   21.8    1.000  1.868   0.313    -0.868     -1.18 X
 44   20.1    1.000  1.691   0.286    -0.691     -0.93 X
```



<u>Residual plots for Discriminant analysis functions</u>, are *usually non-compliant* with regression assumptions of normality of residuals, etc. *Most concerning* is that *Mode is away from Zero*.

Regression Analysis: Discrim (All counties) versus Score-1 (of the PC analysis)

The regression equation is: Discrim = 0.065 + 0.438 Score-1

Predictor	Coef	SE Coef	T	P
Constant	0.0645	0.1027	0.63	0.532
Score-1	**0.43813**	**0.07471**	**5.86**	**0.000**

S = 0.808765   R-Sq = 36.4%   R-Sq(adj) = 35.4% (not very explanatory of the problem)

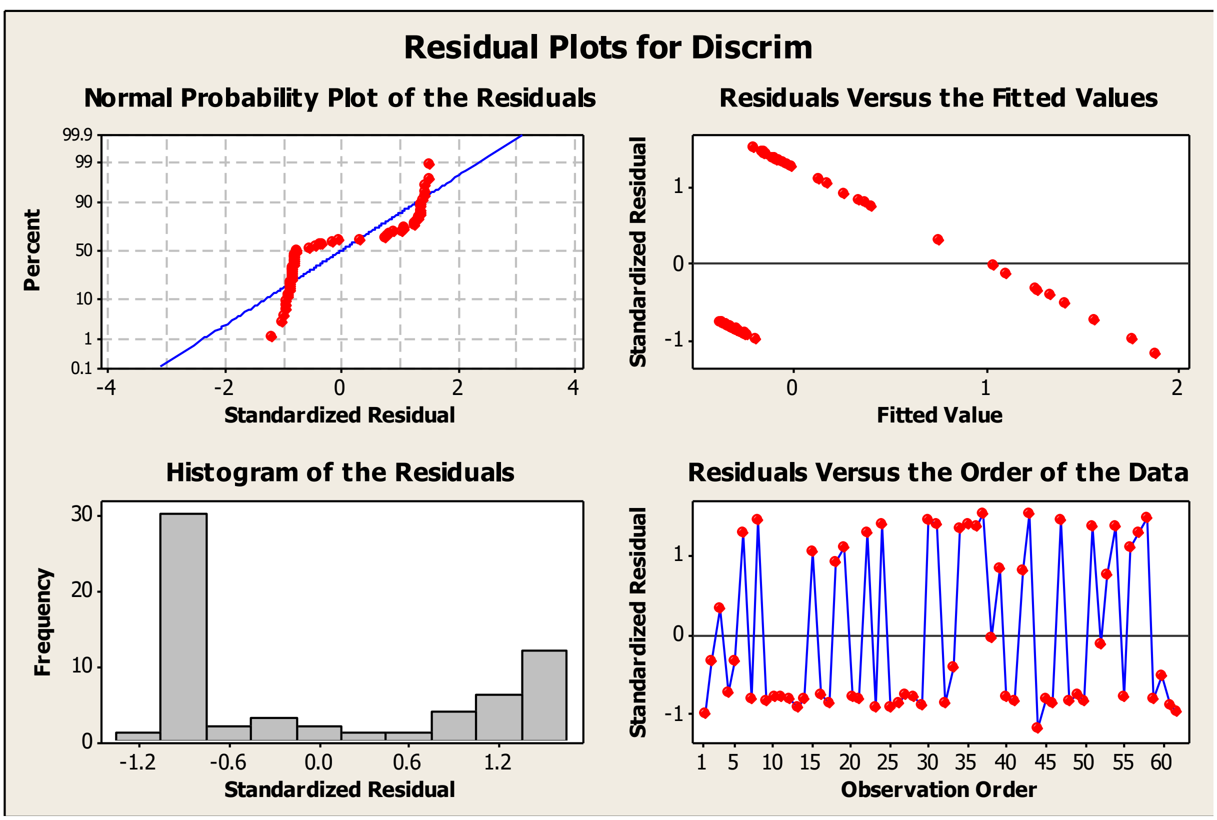
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	22.496	22.496	34.39	0.000
Residual Error	60	39.246	0.654		
Total	61	61.742			

Unusual Observations

Obs	Score-1	Discrim	Fit	SE Fit	Residual	St Resid
1	3.86	1.000	1.755	0.306	-0.755	-1.01 X
4	3.43	1.000	1.568	0.276	-0.568	-0.75 X
44	4.15	1.000	1.881	0.326	-0.881	-1.19 X

X denotes an observation whose X value gives it large influence.



This Graphical Residual analysis is not very encouraging, either.

Analysis for the Extreme Counties [only (-1, 1); 0s or mixed counties are out]:

Discriminant Analysis: TriGrp versus Per 10K, Per 10K_1

Row	County_1	Per 10K_2	Per 10K_1_1	TriGrp_1	Sc-1	Sc-2
1	Bronx	321.6	22.8	1	3.03319	-0.442175
2	Brooklyn	221.7	19.1	1	2.09626	-0.655477
3	Manhattan	167.1	12.5	1	1.13654	-0.324408
4	Queens	276.9	21.8	1	2.67807	-0.601721
5	Staten Island	288.3	15.4	1	2.11786	0.089746
6	Allegany	11.2	0.4	-1	-0.94419	-0.038683
7	Cattaraugus	12.4	0.7	-1	-0.90795	-0.061111
8	Cayuga	13.5	0.1	-1	-0.96029	0.003894
9	Chautauqua	7.9	0.5	-1	-0.95341	-0.067460
10	Chemung	16.3	0.4	-1	-0.91483	-0.009323
11	Chenango	28.4	1.1	-1	-0.77672	-0.008116
12	Clinton	12.9	0.5	-1	-0.92463	-0.038675
13	Cortland	9.4	0.0	-1	-0.99367	-0.009931
14	Delaware	19.1	0.9	-1	-0.84982	-0.042097
15	Dutchess	136.4	4.9	1	0.21661	0.242039
16	Erie	71.0	6.0	1	-0.05232	-0.242029
17	Essex	14.2	0.0	-1	-0.96604	0.017702
18	Franklin	22.7	0.0	-1	-0.91710	0.066636
19	Genesee	36.2	0.9	-1	-0.75138	0.056346
20	Herkimer	22.6	0.6	-1	-0.85901	0.007388
21	Jefferson	7.0	0.0	-1	-1.00749	-0.023747
22	Lewis	9.8	0.0	-1	-0.99137	-0.007628
23	Livingston	19.1	1.1	-1	-0.83026	-0.061655
24	Montgomery	20.2	0.8	-1	-0.85327	-0.025986
25	Nassau	301.4	15.8	1	2.23239	0.126047
26	Orange	275.5	11.6	1	1.67258	0.387650
27	Oswego	9.8	0.3	-1	-0.96203	-0.036964
28	Otsego	12.6	0.8	-1	-0.89702	-0.069739
29	Putnam	146.7	6.3	1	0.41281	0.164433
30	Rockland	409.6	20.1	1	3.27577	0.328459
31	St. Lawrence	19.3	0.2	-1	-0.91712	0.027505
32	Saratoga	21.9	0.7	-1	-0.85326	-0.006421
33	Schoharie	17.4	0.6	-1	-0.88894	-0.022548
34	Seneca	17.8	0.6	-1	-0.88664	-0.020245
35	Suffolk	272.6	13.1	1	1.80256	0.224274
36	Sullivan	187.7	4.6	1	0.48261	0.566705
37	Tompkins	16.6	0.0	-1	-0.95222	0.031519
38	Wayne	13.9	0.4	-1	-0.92865	-0.023140
39	Westchester	351.4	14.4	1	2.38333	0.550795
40	Wyoming	22.5	1.2	-1	-0.80091	-0.051860

Analysis for the Extreme (only -1, 1) Counties (0s or intermediate are out)

Regression Analysis: TriGrp_1 versus Per 10K_2, Per 10K_1_1

The two variable regression equation is:

TriGrp_1 = - 0.973 + 0.00571 Per 10K_2 + 0.0243 Per 10K_1_1

Predictor	Coef	SE Coef	T	P
Constant	-0.97341	0.08821	-11.03	0.000
Per 10K_2	**0.005707**	**0.001696**	**3.36**	**0.002**
Per 10K_1_1	0.02433	0.02882	0.84	0.404

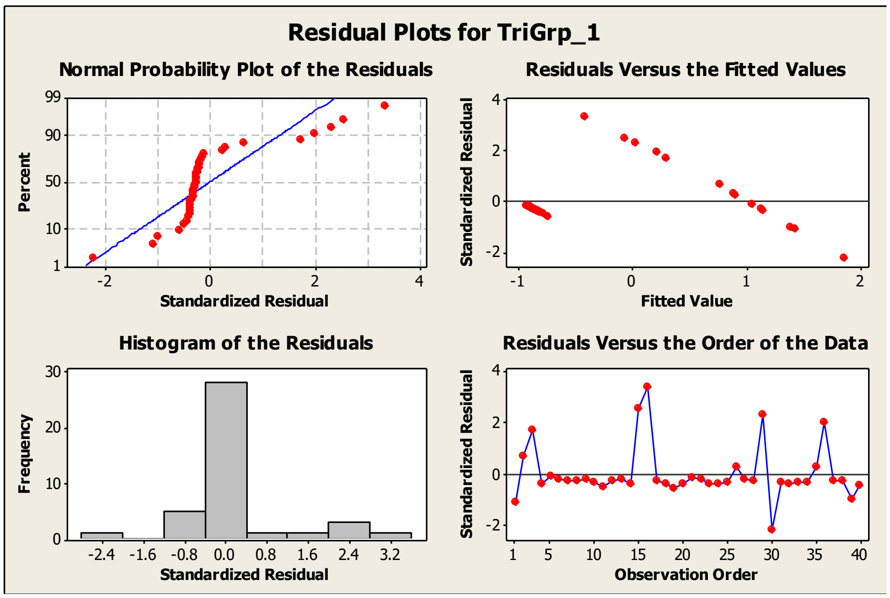
S = 0.433922 R-Sq = 80.9% R-Sq(adj) = 79.8% (improved % explanatory of problem)

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	29.433	14.717	78.16	0.000
Residual Error	37	6.967	0.188		
Total	39	36.400			

Unusual Observations

Obs	10K_2	TriGrp_1	Fit	SE Fit	Residual	St Resid
1	322	1.0000	1.4164	0.2099	-0.4164	-1.10 X
2	222	1.0000	0.7564	0.2277	0.2436	0.66 X
4	277	1.0000	1.1370	0.2304	-0.1370	-0.37 X
15	136	1.0000	-0.0758	0.0988	1.0758	2.55R
16	71	1.0000	-0.4223	0.0983	1.4223	3.37R
29	147	1.0000	0.0170	0.0861	0.9830	2.31R
30	410	1.0000	1.8529	0.2012	-0.8529	-2.22R
36	188	1.0000	0.2096	0.1799	0.7904	2.00R
39	351	1.0000	1.3822	0.2107	-0.3822	-1.01 X



We repeated the Regression Analysis: TriGrp_1 versus Per 10K_2 (for one variable):

The regression equation is: TriGrp_1 = - 0.981 + 0.00706 Per 10K_2

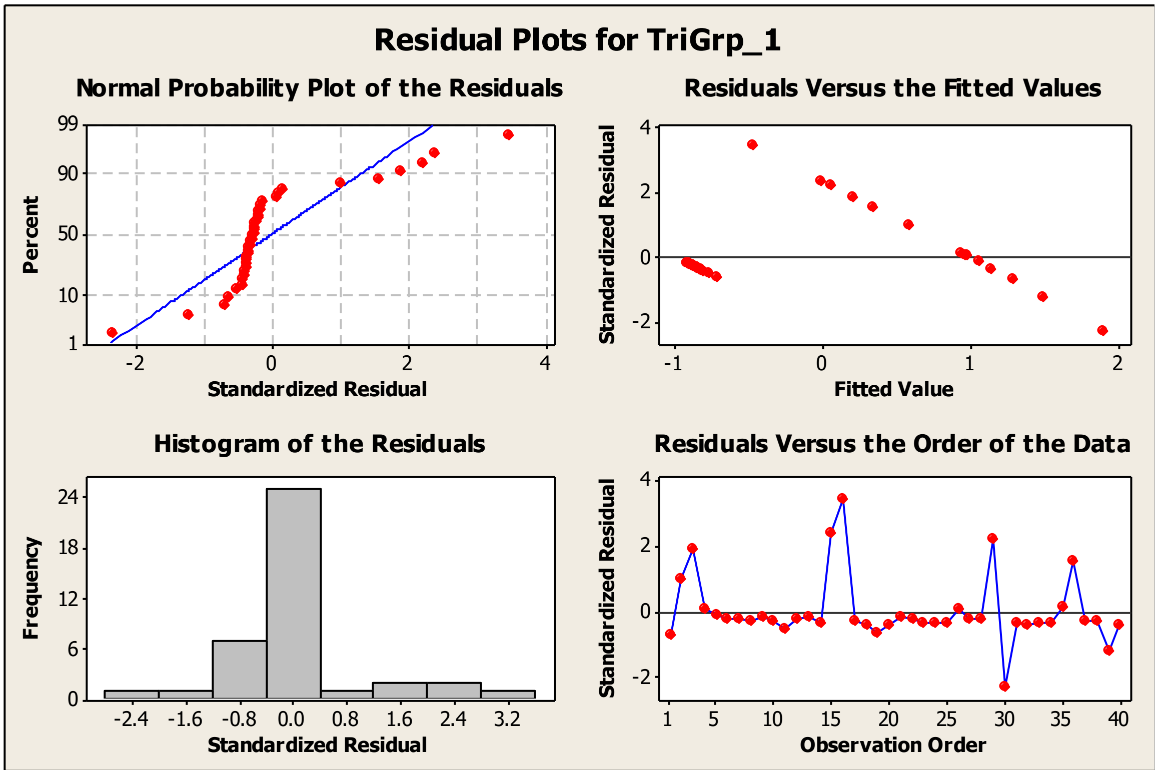
Predictor	Coef	SE Coef	T	P
Constant	-0.98143	0.08737	-11.23	0.000
Per 10K_2	0.0070567	0.0005636	12.52	0.000

S = 0.432278 R-Sq = 80.5% R-Sq(adj) = 80.0% **(improved % explanatory of problem)**

Unusual Observations

Obs	10K_2	TriGrp_1	Fit	SE Fit	Residual	St Resid
15	136	1.0000	-0.0189	0.0719	1.0189	2.39R
16	71	1.0000	-0.4804	0.0699	1.4804	3.47R
29	147	1.0000	0.0538	0.0740	0.9462	2.22R
30	410	1.0000	1.9090	0.1892	-0.9090	-2.34RX

R denotes an observation with a large standardized residual.



Mode of the Histogram of Residuals is now Zero, and residual v. order is also better.

We have obtained a number of different Discrimination Functions by regressing on different variables (Positives and Deaths per 10K ha.; PC Scores). We now have to assess these functions, rank them and find the best one. We do this comparison work, next.

Comparison of two Discriminant Functions using the Mahalanobis Distance:

Mahalanobis criteria: *the larger the distance between the two groups, and the smaller the Miss-Classification probability, the better that the Discriminant Function is***.**

For a Discriminant Function derived from the two (-1, 1) Extreme county groups (total 40):

```
   I) The regression equation is: TriGrp_1 = - 0.981 + 0.00706 Per 10K_2

Predictor        Coef    SE Coef       T      P
Constant     -0.98143    0.08737  -11.23  0.000
Per 10K_2  0.0070567  0.0005636   12.52  0.000

S = 0.432278   R-Sq = 80.5%   R-Sq(adj) = 80.0% (improved % explanatory of problem)
```

This Discrimination Function explains 80% of the problem and is able to correctly classify most counties in their respective group. The **Mahalanobis Distance** that separates these two Groups: (-1, 1), can be obtained in the following way:

For: n1= 26 (1); n2 = 14 (-1);
Lambda ^{2} = n1*n2/(n1+n2) = 26*14/40 = 9.1
Dp^2 = [(n1+n2-2)/Lambda ^{2}]*[R^2 / (1- R^2)] = (38/9.1)*(0.80/(1-0.80)) = 16.703
Dp = Sqrt (Dp^2) = 4.087 => Prob (- ½ Dp) = Prob (- 4.087) = 2.18 E-05

For a Discriminant Function derived from two groups (with all 62 counties):

```
   II)   The regression equation is: Discrim = 0.065 + 0.438 Score-1

Predictor      Coef  SE Coef     T      P
Constant     0.0645   0.1027  0.63  0.532
Score-1     0.43813  0.07471  5.86  0.000

S = 0.808765   R-Sq = 36.4%   R-Sq(adj) = 35.4% (not very explanatory of the problem)
```

This Discrimination Function explains 35.5% of the problem and is able to correctly classify many counties in their respective group. The **Mahalanobis Distance** that separates these two Extreme Groups (-1, 1), can be obtained in the following way:

For: n1= 33 (1); n2 = 29 (0);
Lambda ^{2} = n1*n2/(n1+n2) = 33*29/62 = 15.44
Dp^2 = [(n1+n2-2)/Lambda ^{2}]*[R^2 / (1- R^2)] = (60/15.44)*(0.355/0.645) = 2.139
Dp = Sqrt (Dp^2) = 1.463 => Prob (- ½ Dp) = Prob (- 1.463) = 0.072

Comparison of results:

Function	Mahalanobis Distance	Prob. Misclassification	Evaluation
I – Tri Group	16.703	2.18 E-05	Best function
II – All Counties	2.139	0.072	Not good

<u>Graphical</u> <u>Comparison of selected Discriminant Functions obtained using Fisher's method</u>

Several Discriminant Functions were obtained, using different regressors (variables). We need to compare them and select the best one. Toward this end, we have developed two types of graphs, representing the discrimination ordering of the NYS counties.

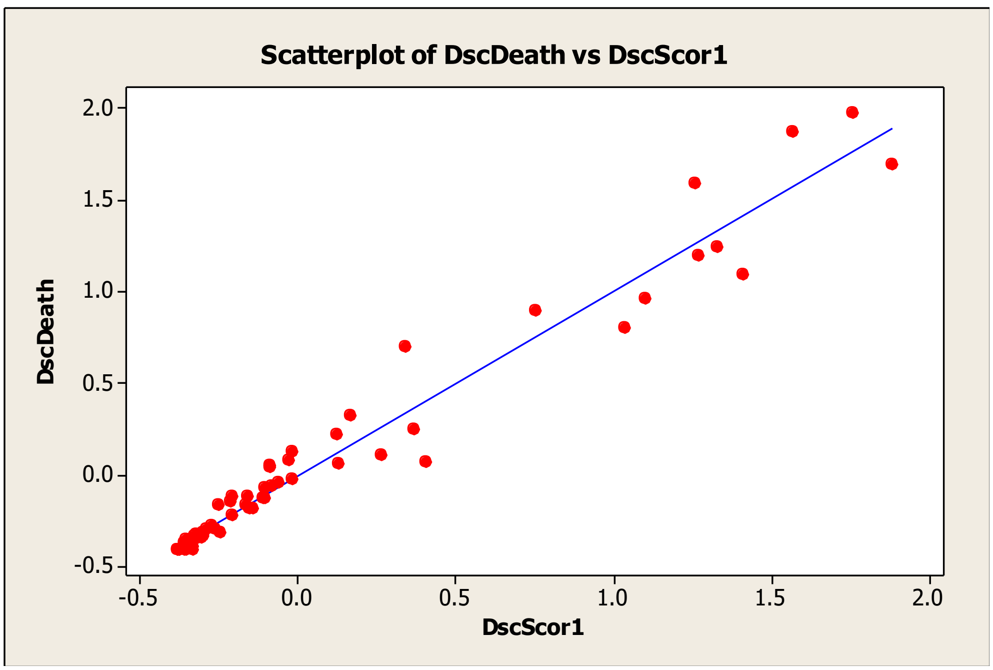
One type is scatter plots of pairs of DA functions for the NYS counties. If the points of their *orderings* are close to the line with *unit slope*, DA functions are similar. The second type is the juxtaposition of two dot plots of NYS county DA *ordering*. A better discrimination will separate the two groups further from boundary ZERO. Below are the notations for such graphs:

Using <u>Variable Death/0000s</u> as discriminator: <u>DscDeath</u>

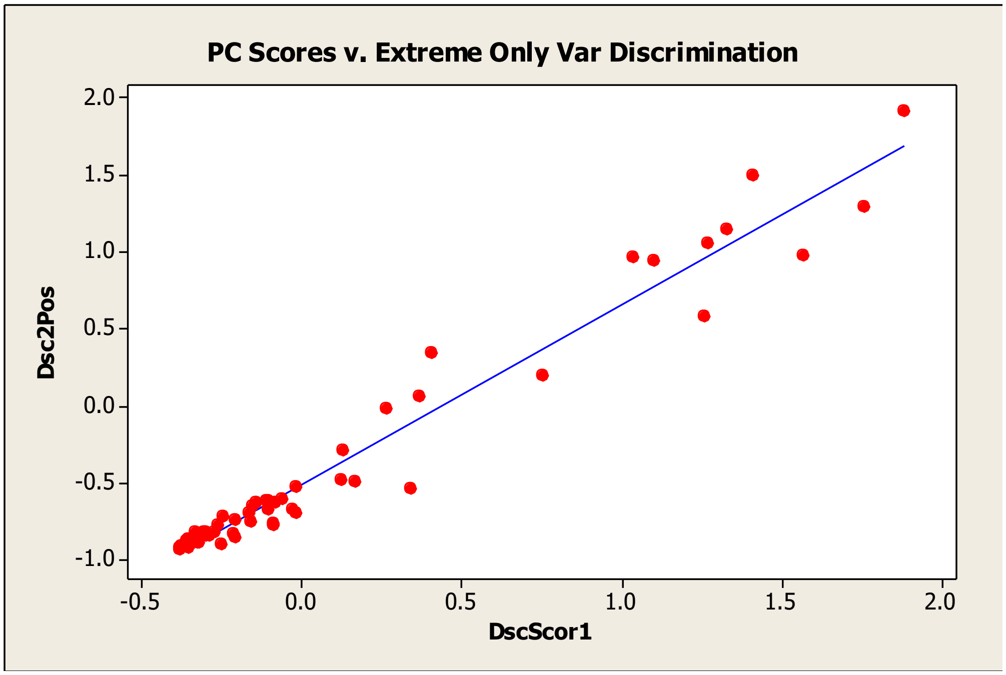
Using <u>Comp-1 Score</u> from the Principal Comp. Analysis: <u>DscScor1</u>

Using <u>Variable Pos/0000s</u> Extreme only, as discriminator: <u>Dsc2Pos</u>

<u>Comparison using Scatter plots</u>: do these equations perform similar job?

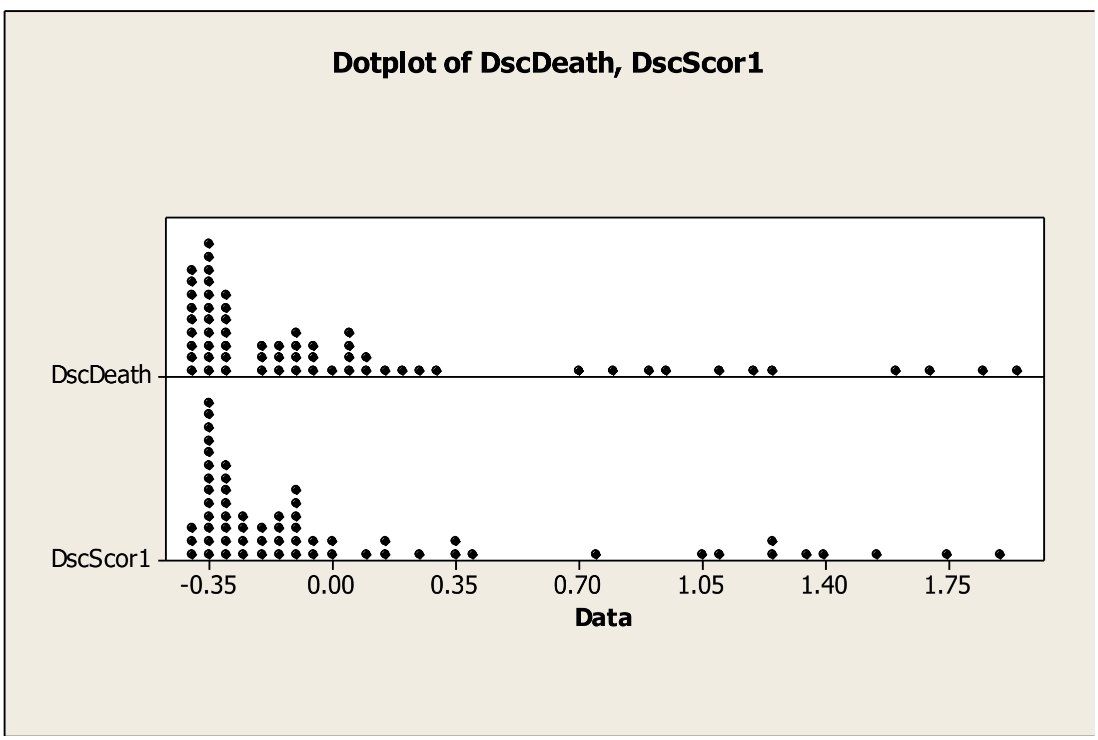


Both DA <u>functions are similar</u>, as their points are close to the <u>Line with slope Unit</u>.

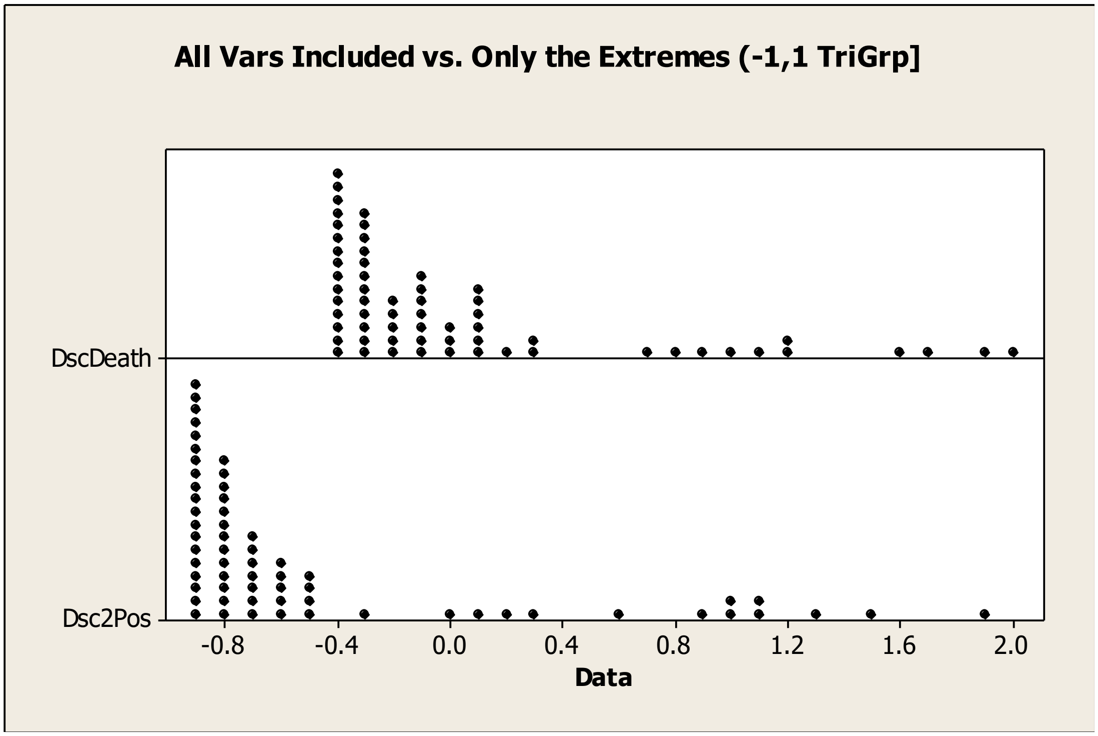


Both DA functions are similar. The Line with slope Unit has sharpened when DA uses, to build the Discriminant Function, Extreme counties Only. We can, implementing these DA functions with their variables, place the remaining counties (mixed) in their respective groups.

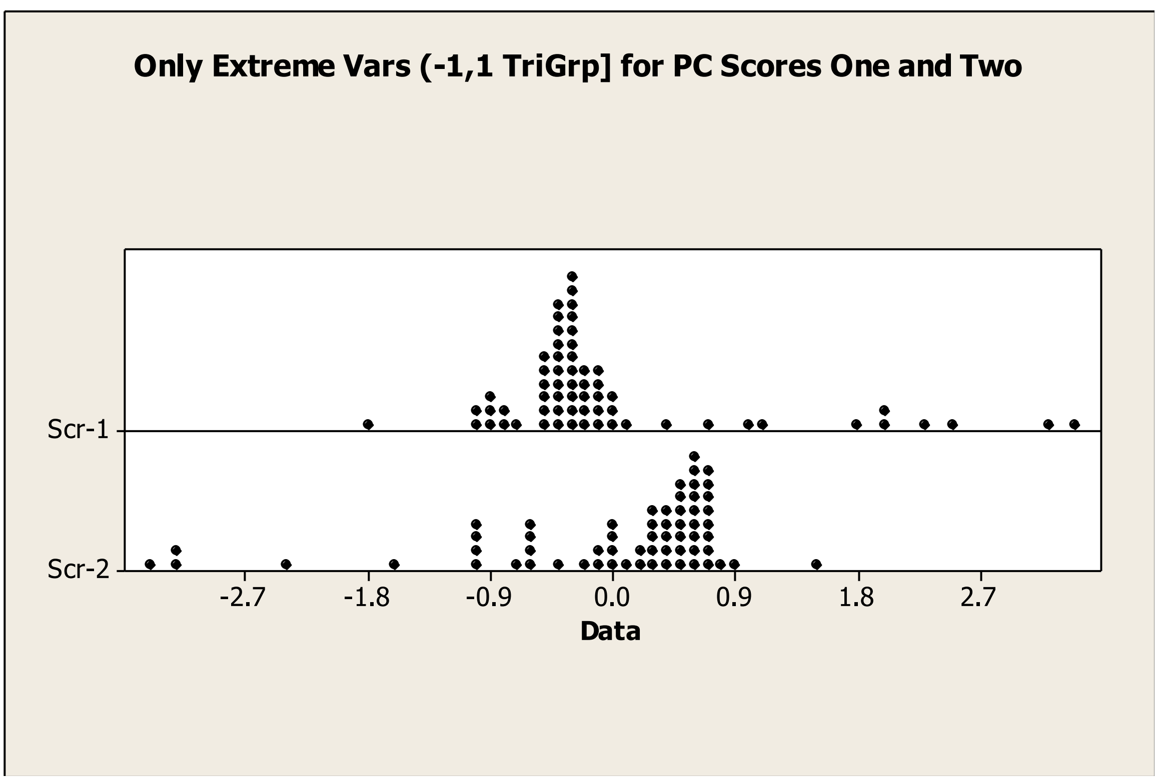
Comparison of Discriminant Functions that use a single variable or PC Score:



The above two discriminating variables (Death and PC Score1), obtained using All 62 counties, work similarly. Score1 seems to separate the groups somewhat better (further from Zero).



The groups are separated further (Dsc2Po) when only Extreme groups (-1, 1) are considered.



These Discriminating functions were obtained with Extreme groups only.

For completeness, we implemented the Minitab SW DA Program to the data:

Minitab Discriminant Function: All 62 counties, one predictor and Two groups

Response: All 62 Counties; Predictors: Score-1

Group	0	1
Count	29	33

Summary of classification

	True Group	
Put into Group	0	1
0	29	16
1	0	17
Total N	29	33
N correct	29	17
Proportion	1.000	0.515

N = 62 N Correct = 46 Proportion Correct = 0.742

Squared Distance Between Groups

	0	1
0	0.00000	2.22811
1	**2.22811**	0.00000

Linear Discriminant Function for Groups

	0	1
Constant	-0.31561	-0.24374
Score-1	-0.71303	0.62660

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
6**	1	0	0	0.3933	0.544
			1	0.7491	0.456
8**	1	0	0	0.1116	0.649
			1	1.3425	0.351
22**	1	0	0	0.3567	0.555
			1	0.8018	0.445
24**	1	0	0	0.1969	0.611
			1	1.1004	0.389
30**	1	0	0	0.1350	0.638
			1	1.2663	0.362
31**	1	0	0	0.1952	0.612
			1	1.1044	0.388
34**	1	0	0	0.2864	0.578
			1	0.9169	0.422
35**	1	0	0	0.1914	0.613
			1	1.1134	0.387
36**	1	0	0	0.2403	0.594
			1	1.0049	0.406
37**	1	0	0	0.05040	0.685
			1	1.60828	0.315
43**	1	0	0	0.05402	0.683
			1	1.58825	0.317
47**	1	0	0	0.1165	0.647
			1	1.3257	0.353
51**	1	0	0	0.2277	0.599

```
                                       1     1.0312          0.401
         54**       1       0          0     0.2264          0.600
                                       1     1.0340          0.400
         57**       1       0          0     0.3906          0.545
                                       1     0.7529          0.455
         58**       1       0          0     0.1035          0.653
                                       1     1.3713          0.347
```

Results: Minitab uses all 62 counties, one predictor (Score) and two groups, for constructing the Discrimination Function. Resulting proportion (Probability) of correct placement is 0.74, and the (Mahalanobis) Distance between the two groups is 2.228. Compare the above results with the equivalent results, previously calculated for the Mahalanobis Distance.

Minitab Discriminant Function (All 62 Counties, two Predictors and Three Groups):

Response TriGrp and Predictors: Per 10K, Per 10K_1

```
Group         -1         0         1
Count         27        21        14

Summary of classification

                     True Group
Put into Group      -1       0       1
-1                  27       3       0
 0                   0      18       2
 1                   0       0      12
Total N             27      21      14
N correct           27      18      12
Proportion       1.000   0.857   0.857

N = 62           N Correct = 57          Proportion Correct = 0.919

Squared Distance Between Groups

         -1         0        1
-1   0.0000    1.0904  25.6871
 0   1.0904    0.0000  19.3567
 1  25.6871   19.3567   0.0000

Linear Discriminant Function for Groups

               -1       0        1
Constant   -0.076  -0.713  -14.609
Per 10K     0.011   0.006    0.105
Per 10K_1  -0.069   0.312    0.266

Summary of Misclassified Observations

                True   Pred           Squared
Observation     Group  Group  Group  Distance  Probability
         18**       1      0     -1     7.413        0.286
                                  0     6.430        0.468
                                  1     7.720        0.246
         19**       1      0     -1    3.0535        0.220
                                  0    0.5234        0.779
                                  1   14.8342        0.001
         30**       0     -1     -1    0.5444        0.527
                                  0    0.7596        0.473
                                  1   18.8641        0.000
         43**       0     -1     -1    0.1867        0.533
```

			0	0.4507	0.467
			1	21.8074	0.000
47**	0	-1	-1	0.4222	0.521
			0	0.5919	0.479
			1	19.5261	0.000

Results: Minitab uses three groups, all 62 counties, and two variables or predictors for constructing a Discrimination Function. The resulting proportion (*Probability*) of correct placement is 0.919, and the (*Mahalanobis*) Distance between the groups is 25.68. Compare these results with the ones calculated for the Mahalanobis Distance, where *we used* only two classification groups -and Minitab used three.

Minitab Discriminant Function (All counties, three groups and one predictor):

Response:Tri-Grp Counties; Predictors: Score-1

Group	-1	0	1
Count	27	21	14

Summary of classification

	True Group		
Put into Group	-1	0	1
-1	27	4	0
0	0	17	4
1	0	0	10
Total N	27	21	14
N correct	27	17	10
Proportion	1.000	0.810	0.714

N = 62 N Correct = 54 Proportion Correct = 0.871

Squared Distance Between Groups

	-1	0	1
-1	0.0000	0.8233	23.7777
0	0.8233	0.0000	15.7520
1	23.7777	15.7520	0.0000

Linear Discriminant Function for Groups

	-1	0	1
Constant	-0.9918	-0.1255	-6.0129
Score-1	-2.2009	-0.7830	5.4192

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
18**	1	0	-1	4.483	0.174
			0	1.464	0.789
			1	7.611	0.036
19**	1	0	-1	2.6083	0.257
			0	0.5008	0.738
			1	10.6355	0.005
25**	0	-1	-1	0.0805	0.538
			0	0.3889	0.462
			1	21.0907	0.000
37**	0	-1	-1	0.1730	0.509
			0	0.2415	0.491

			1	19.8944	0.000
42**	1	0	-1	6.192	0.116
			0	2.499	0.736
			1	5.702	0.148
43**	0	-1	-1	0.1847	0.505
			0	0.2281	0.495
			1	19.7715	0.000
53**	1	0	-1	6.842	0.095
			0	2.919	0.678
			1	5.110	0.227
62**	0	-1	-1	0.1948	0.503
			0	0.2172	0.497
			1	19.6683	0.000

Results: Minitab again uses all 62 counties for constructing the three-group Discrimination Function. The resulting proportion (*Probability*) of correct placement is 0.919, and the (*Mahalanobis*) Distance between the groups is 25.68. Compare these results with the ones we calculated for the Mahalanobis Distance, where we used only two classification groups and Minitab uses three.

Conclusions:

In this paper we have *developed* two powerful Multivariate Statistical methods: *Principal Components* and *Discriminant Analyses*, *to examine and assess NYS County Covid-19 me*trics. The data has been obtained from the *Syracuse Post Standard Web Site*[2]. *We thank Mike Dupras and Marie Morelli* for *facilitating the data.*

This is a preliminary analysis, subject to further validation and verification using subsequent and additional NYS county data**.** *The main objective of our analysis is to provide a detailed example of the use of Multivariate Analysis, especially Principal Components*, *Factor*, *and Discriminant Analyses* applied to the assessment of Covid-19 data. This is part of *our pro-bono collaboration to the American struggle against Covid-19*[3].

This framework can be used as a guide for the analysis of *similar data* from counties *of any state*, or provinces of *any country*. It can also *include more variables* (metrics), as they become available to researchers. It can *also help researchers generate new ideas for other types of analyses* that may be implemented *with the Covid-19 data*.

Specifically about this data set we can say: (1) *Principal Components* is *a good way to start*. It can *reduce* the *number of variables* by *absorbing them* into fewer Factors that explain a high percentage of the problem, as done here and in Romeu (2020); (2) Discriminant Functions based on all 62 counties were weak; (3) Discriminant Functions based on using two groups of Extreme counties (-1, 1), leaving the intermediate group (0) out of the analysis, as done in Romeu 1978, *separated much better the two populations*, yielding higher R-square (Index of Fit), larger Mahalanobis Distances, and smaller Miss-classification probabilities. The C*ounties* in the intermediate group (0) *can then be included* in their correct positions by using the Discriminant Function with said data.

[2] https://www.syracuse.com/, the Syracuse Post Standard web site.
[3] See (Bibliography) A Proposal for Fighting Covid-19 and its Economic Fallout.

Bibliography

Anderson, T.W. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons. New York. Second Edition. 1971.

Romeu, J. L. 2020a. *A Proposal for Fighting Covid-19 and its Economic Fallout*. ResearchGate. https://www.researchgate.net/publication/341282217_A_Proposal_for_Fighting_Covid-19_and_its_Economic_Fallout and https://web.cortland.edu/matresearch/Covid-19Proposal2020.pdf

Romeu, J. L. 2020b. *Multivariate Statistics in the Analysis of Covid-19 Data*. ResearchGate. https://www.researchgate.net/publication/341385856_Multivariate_Stats_PC_Discrimination_in_the_Analysis_of_Covid-19 and https://web.cortland.edu/matresearch/MultivarStatsCovid19.pdf

Romeu, J. L. *A Comparative Study of Goodness-of-Fit Tests for Multivariate Normality*. Journal of Multivariate Analysis. V. 46, No. 2. August 1993. 309--334.

Romeu, J. L. 1978. *Analisis y Clasificacion de Fuentes de Informacion, mediante métodos estadisticos*. Revista de Ingenieria Civil. Vol. 4. Havana, Cuba. Available in the Internet: https://web.cortland.edu/romeu/DiscrAnalClasFuentesGvnDap.pdf

Coronavarus Pandemic: *Leaders*, Going Global; *Briefing*, Flattening the Curve; *Graphic Detail*, Coronavarus Statistics. The Economist, February 29th, 2020. https://web.cortland.edu/matresearch/CoronavarusEconomist.pdf

About the Author:

Jorge Luis Romeu was, for sixteen years, a Research Professor at Syracuse University. He is currently an Adjunct Professor of Statistics. Romeu retired Emeritus from the State University of New York. He worked as a Senior Research Engineer, with IIT Research Institute. Romeu received seven Fulbright assignments: in Mexico (3), the Dominican Republic (2), Ecuador, and Colombia. Romeu holds a doctorate in Statistics/O.R., is a C. Stat Fellow of the Royal Statistical Society (RSS), a Member of the American Statistical Society (ASA) and of the American Society for Quality (ASQ). He is a Past ASQ Regional Director and holds Reliability and Quality ASQ Certifications. Romeu created and directs the *Juarez Lincoln Marti Int'l. Ed. Project* (https://web.cortland.edu/matresearch/) dedicated to support higher education in Ibero-America.