



Principle component analysis using JMP for better visualization of data

Rajesh Jha^{*}, PhD

Research Associate (Swine Nutrition)

Ruurd Zijlstra, PhD

Professor (Swine Nutrition)

Department of Agricultural, Food and Nutritional Science
University of Alberta

*rejesh.jha@ualberta.ca



Outlines of presentation

- Why factor or component analysis
- What is Principle Component Analysis (PCA)
- What the PCA can do
- JMP functions for PCA
- Illustration with an example



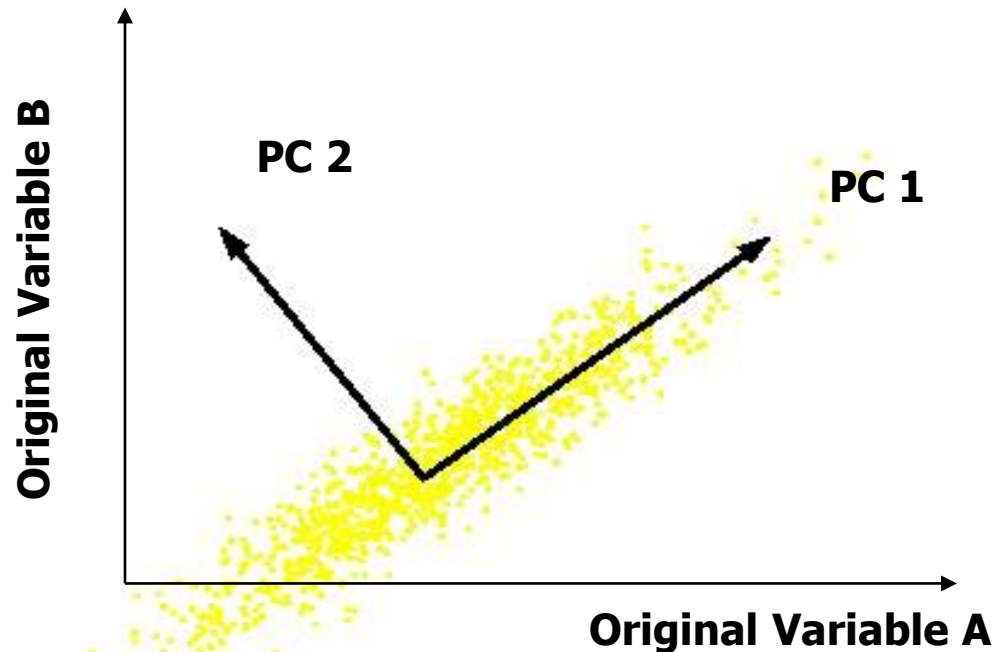
Why factor or component analysis?

- To discover and exploit unseen relationships
- In dataset with too many observations and dimensions
 - ❖ To reason about or obtain insights from
 - ❖ To visualize
 - ❖ Need to "reduce" them to a smaller set of factors
 - ❖ Better representation of data without losing much information
 - ❖ Can build more effective data analyses on the reduced-dimensional space: classification, clustering, pattern recognition



What are the new axes

- Each principal axis is a linear combination of the original two variables



- PC axes are the rigid rotation of the original variables
- Projections along PC1 discriminate the data most along any one axis



Principal Components

- First principal component is the direction of greatest variability (covariance) in the data
- Second is the next orthogonal (uncorrelated) direction of greatest variability
 - ❖ So first remove all the variability along the first component, and then find the next direction of greatest variability
- And so on ...



Principal Component Analysis

- PCA is a way of finding patterns in data
- Probably the most widely-used and well-known of the "standard" multivariate methods
- Invented by Pearson (1901) and Hotelling (1933)
- First applied in ecology by Goodall (1954) under the name "factor analysis" ("principal factor analysis" is a synonym of PCA).
 - Not sure exact date of its use in Animal science, probably not more than 2 decades.



Principal Component Analysis

➤ Uses:

- ❖ Data Visualization
- ❖ Data Reduction
- ❖ Data Classification
- ❖ Trend Analysis
- ❖ Factor Analysis
- ❖ Noise Reduction



Principal Component Analysis

Table: Physico-chemical properties and fermentation characteristics of DDGS

	IVDMD	Vf	T	T/2	μ	Ac	Pr	Bu	BCFA	Total VFA	Ether extract	Crude fiber	CP	ADF	NDF	Total DF	Total NSP	Insoluble NSP	Soluble NSP	Starch	Viscosity
WH-Nenz	67.7	170	2.05	19.55	0.042	3.01	1.26	0.50	0.16	5.04	8.56	7.84	35.45	10.48	39.9	30.81	24.76	20.38	4.38	2.23	1.3
WT-Nenz	72.1	155	1.15	19.18	0.041	2.94	1.22	0.45	0.10	4.78	4.57	7.12	43.42	12.5	27.89	26.59	22.48	16.03	6.45	0.51	1.6
Corn-Nenz	59.6	200	2.49	23.21	0.045	3.87	1.60	0.56	0.25	6.44	12.04	6.69	29.03	8.92	32.43	30.87	23.04	21.96	1.08	6.40	1.1
WH-CFD	67.7	180	0.03	14.28	0.042	3.17	1.46	0.47	0.19	5.43	8.56	7.84	35.45	10.48	39.9	30.81	24.76	20.38	4.38	2.23	1.3
WT-CFD	72.1	159	0.00	14.37	0.039	2.89	1.33	0.38	0.11	4.82	4.57	7.12	43.42	12.5	27.89	26.59	22.48	16.03	6.45	0.51	1.6
Corn-CFD	59.6	211	0.21	18.41	0.041	3.53	1.61	0.55	0.27	6.11	12.04	6.69	29.03	8.92	32.43	30.87	23.04	21.96	1.08	6.40	1.1
WH-CFDP	67.7	123	0.86	6.99	0.087	2.11	1.37	0.40	0.29	4.28	8.56	7.84	35.45	10.48	39.9	30.81	24.76	20.38	4.38	2.23	1.3
WT-CFDP	72.1	111	0.54	7.10	0.079	2.01	1.28	0.33	0.25	4.02	4.57	7.12	43.42	12.5	27.89	26.59	22.48	16.03	6.45	0.51	1.6
Corn-CFDP	59.6	152	0.23	12.52	0.047	2.73	1.52	0.52	0.39	5.35	12.04	6.69	29.03	8.92	32.43	30.87	23.04	21.96	1.08	6.40	1.1



Principal Component Analysis

A screenshot of the JMP software interface. The window title is "Untitled 3 - JMP". The menu bar includes File, Edit, Tables, Rows, Cols, DOE, Analyze, Graph, Tools, View, Window, and Help. The Edit menu is open, showing options such as Undo (Ctrl+Z), Redo (Ctrl+Y), Cut (Ctrl+X), Copy (Ctrl+C), Copy As Text, Copy With Column Names, Paste (Ctrl+V), Paste With Column Names (highlighted), Clear, Select All (Ctrl+A), Save Selection As..., Run Script (Ctrl+R), Stop Script, Submit to SAS (FB), Search, Go to Line..., Reformat Script, Journal (Ctrl+J), and Layout (Ctrl+L). The main workspace is a large grid. On the left, there are panels for "Un", "Co", "Co", "Ro", "All row", "Select", "Excluded", "Hidden", and "Labelled", with values 0, 0, 0, 0, 0, 0, 0, 0, and 0 respectively. The Windows taskbar is visible at the bottom, showing the time as 3:38 PM.



Principal Component Analysis

PCA dataset 2 March 2012 - JMP

File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help

PCA dataset 2 March 20...

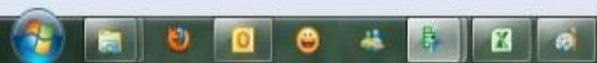
	Column 1	IVDMD	Vf	T	T/2	μ	Ac	Pr	Bu	BCFA	TotalVFA	Crude f
1	WH-Nenz	67.7	170	2.05	19.55	0.042	3.01	1.26	0.5	0.16	5.04	
2	WT-Nenz	72.1	155	1.15	19.18	0.041	2.94	1.22	0.45	0.1	4.78	
3	Com-Nenz	59.6	200	2.49	23.21	0.045	3.87	1.6	0.56	0.25	6.44	
4	WH-CFD	67.7	180	0.03	14.28	0.042	3.17	1.46	0.47	0.19	5.43	
5	WT-CFD	72.1	159	0	14.37	0.039	2.89	1.33	0.38	0.11	4.82	
6	Com-CFD	59.6	211	0.21	18.41	0.041	3.53	1.61	0.55	0.27	6.11	
7	WH-CFDP	67.7	123	0.86	6.99	0.087	2.11	1.37	0.4	0.29	4.28	
8	WT-CFDP	72.1	111	0.54	7.1	0.079	2.01	1.28	0.33	0.25	4.02	
9	Com-CFDP	59.6	152	0.23	12.52	0.047	2.73	1.52	0.52	0.39	5.35	

Columns (23/0)

- Column 1
- IVDMD
- Vf
- T
- T/2
- μ
- Ac
- Pr
- Bu
- BCFA
- TotalVFA
- Crude fiber
- CP
- ADF

Rows

- All rows 9
- Selected 0
- Excluded 0
- Hidden 0
- Labelled 0





Principal Component Analysis

PCA dataset 2 March 2012 - JMP

File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help

PCA dataset 2 March 2012

	Vf	T	T/2	μ	Ac	Pr	Bu	BCFA	TotalVFA	Crude f
	170	2.05	19.55	0.042	3.01	1.26	0.5	0.16	5.04	
	155	1.15	19.18	0.041	2.94	1.22	0.45	0.1	4.78	
	200	2.49	23.21	0.045	3.87	1.6	0.56	0.25	6.44	
			14.28	0.042	3.17	1.46	0.47	0.19	5.43	
			14.37	0.039	2.89	1.33	0.38	0.11	4.82	
			18.41	0.041	3.53	1.61	0.55	0.27	6.11	
			6.99	0.087	2.11	1.37	0.4	0.29	4.28	
			7.1	0.079	2.01	1.28	0.33	0.25	4.02	
			12.52	0.047	2.73	1.52	0.52	0.39	5.35	

Columns (23/0)

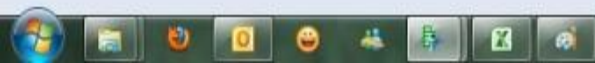
- Column 1
- IVDMD
- Vf
- T
- T/2
- μ
- Ac
- Pr
- Bu
- BCFA
- TotalVFA
- Crude fiber
- CP
- ADF

Rows

- All rows
- Selected
- Excluded
- Hidden
- Labelled

Analyze

- Distribution
- Fit Y by X
- Matched Pairs
- Fit Model
- Modeling
- Multivariate Methods
 - Multivariate
 - Cluster
 - Principal Components
 - Discriminant
 - PLS
 - Item Analysis
- Reliability and Survival





Principal Component Analysis

PCA dataset 2 March 2012 - JMP

File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help



PCA dataset 2 March 20...

	Column 1	IVDMD	Vf	T	T/2	μ	Ac	Pr	Bu	BCFA	TotalVFA	Crude f
1	WH-Nenz	67.7	170	2.05	19.55	0.042	3.01	1.26	0.5	0.16	5.04	
2	WT-Nenz	72.1	155	1.15	19.18	0.041	2.94	1.22	0.45	0.1	4.78	
3	Com-Nenz	59.6	200	2.49	23.21	0.045	3.87	1.6	0.56	0.25	6.44	
4	WH-CFD	67.7	180	0.03	14.28	0.042	3.17	1.46	0.47	0.19	5.43	
5	WT-CFD	72.1	159	0	14.37	0.039	2.89	1.33	0.38	0.11	4.82	
6	Com-CFD	59.6	211	0.21	18.41	0.041	3.53	1.61	0.55	0.27	6.11	
7	WH-CFDP	67.7	123	0.86	6.99	0.087	2.11	1.37	0.4	0.29	4.28	
8	WT-CFDP	72.1	111	0.54	7.1	0.079	2.01	1.28	0.33	0.25	4.02	
9	Com-CFDP	59.6	152	0.23	12.52	0.047	2.73	1.52	0.52	0.39	5.35	

Columns (23/0)

- Column 1
- IVDMD
- Vf
- T
- T/2
- μ
- Ac
- Pr
- Bu
- BCFA
- TotalVFA
- Crude fiber
- CP
- ADF

Rows

- All rows 9
- Selected 0
- Excluded 0
- Hidden 0
- Labelled 0

Principal Components - JMP

Select Columns

- Column 1
- IVDMD
- Vf
- T
- T/2
- μ
- Ac
- Pr
- Bu
- BCFA
- TotalVFA
- Crude fiber
- CP
- ADF
- NDF

Cast Selected Columns into Roles

Y, Columns: IVDMD, Vf, T, T/2

Weight: optional numeric

Freq: optional numeric

By: optional

Action: OK, Cancel, Remove, Recall, Help

Estimation Method: Default

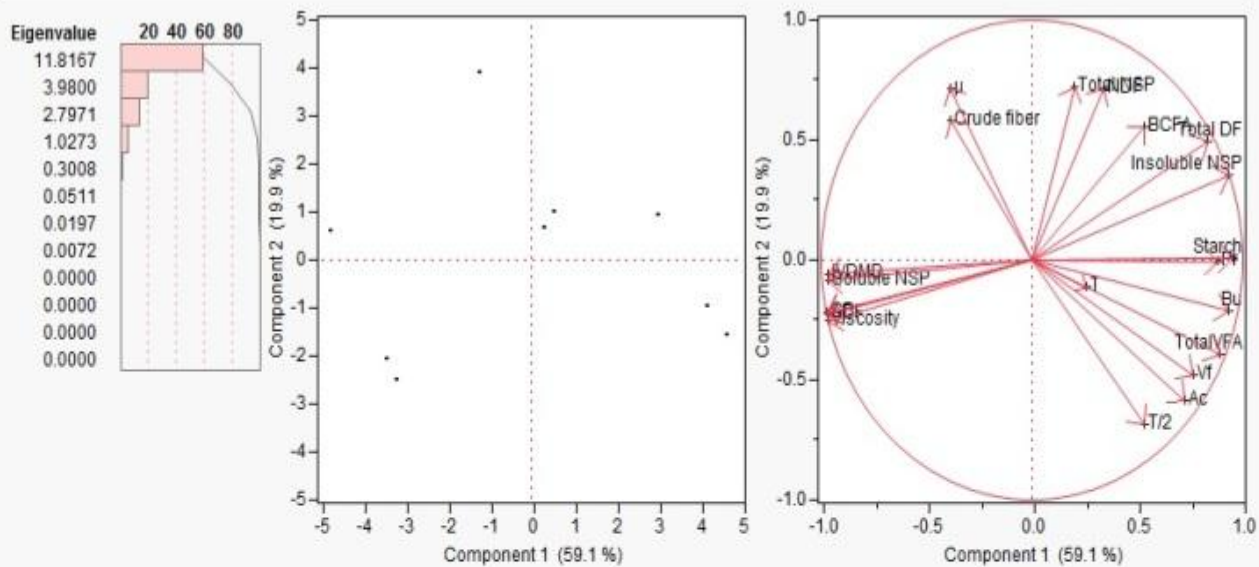


Principal Component Analysis

PCA dataset 2 March 2012 - Principal Components - JMP

Principal Components: on Correlations

Summary Plots



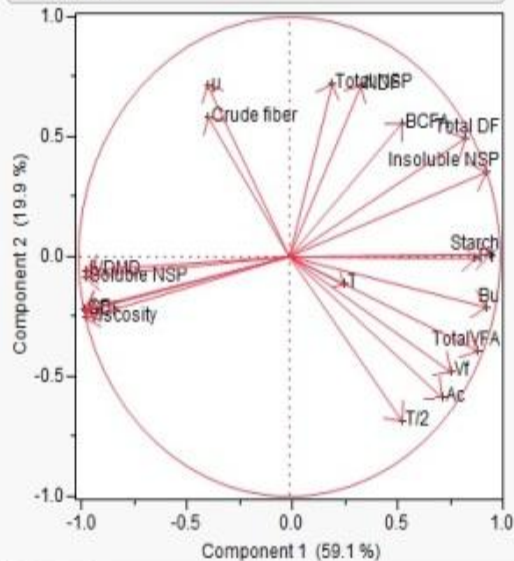


Principal Component Analysis

PCA dataset 2 March 2012 - Principal Components - JMP

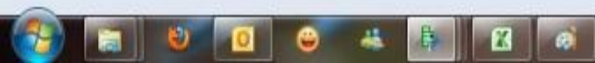
Principal Components: on Correlations

Loading Plot



Eigenvalues

Number	Eigenvalue	Percent	Cum Percent	ChiSquare	DF	Prob>ChiSq
1	11.8167	59.083	59.083	1211.16	209.000	<.0001*
2	3.9800	19.900	78.984	1102.88	189.000	<.0001*
3	2.7971	13.986	92.969	1032.52	170.000	<.0001*
4	1.0273	5.136	98.106	911.237	152.000	<.0001*
5	0.3008	1.504	99.610	771.290	135.000	<.0001*
6	0.0511	0.256	99.865	609.816	119.000	<.0001*
7	0.0197	0.099	99.964	516.748	104.000	<.0001*
8	0.0072	0.036	100.000	406.146	90.000	<.0001*





Principal Component Analysis

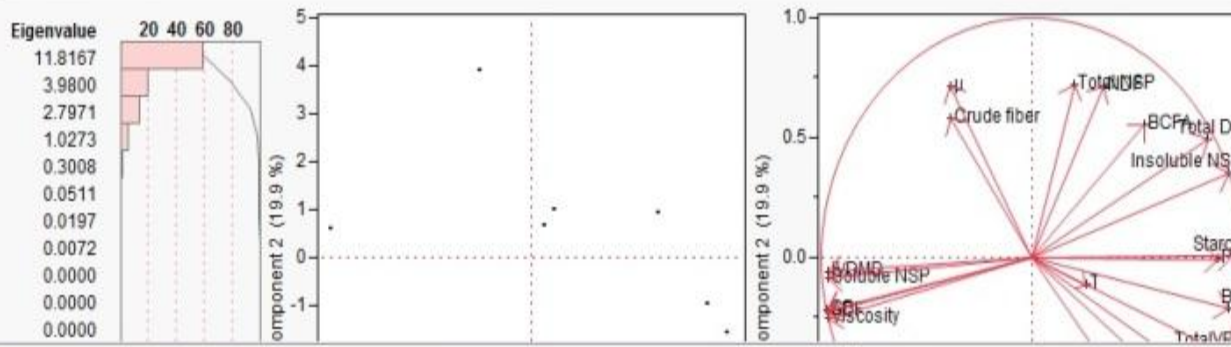
PCA dataset 2 March 2012 - Principal Components - JMP

Principal Components: on Correlations

Correlations

	IVDMD	Vf	T	T/2	μ	Ac	Pr	Bu	BCFA	TotalVFA	Crude fiber	CP	ADF	NDF	Total DF	Total NSP	Insoluble NSP	Soluble NSP	Starch	Viscosity
IVDMD	1.0000	-0.6200	-0.1760	-0.3758	0.2424	-0.5769	-0.9044	-0.8517	-0.7030	-0.8065	0.5212	0.9734	0.9705	-0.2128	-0.7772	-0.0686	-0.9078	0.9993	-0.9978	0.9599
Vf	-0.6200	1.0000	0.2036	0.8316	-0.7683	0.9671	0.6598	0.8396	-0.0920	0.9379	-0.3256	-0.6028	-0.6010	0.1291	0.4800	0.0396	0.5616	-0.6194	0.6188	-0.5943
T	-0.1760	0.2036	1.0000	0.5703	-0.0417	0.3459	-0.0518	0.3738	-0.1290	0.2677	0.0347	-0.2053	-0.2066	0.1823	0.2301	0.1600	0.2220	-0.1815	0.1658	-0.2105
T/2	-0.3758	0.8316	0.5703	1.0000	-0.8018	0.9085	0.2811	0.7510	-0.3578	0.7693	-0.3105	-0.3349	-0.3323	-0.0513	0.2075	-0.1083	0.2848	-0.3704	0.3838	-0.3230
μ	0.2424	-0.7683	-0.0417	-0.8018	1.0000	-0.7888	-0.2279	-0.6022	0.3942	-0.6530	0.2897	0.1921	0.1892	0.1355	-0.0680	0.1743	-0.1398	0.2350	-0.2546	0.1790
Ac	-0.5769	0.9671	0.3459	0.9085	-0.7888	1.0000	0.6085	0.8258	-0.1557	0.9430	-0.3841	-0.5391	-0.5363	0.0273	0.3868	-0.0579	0.4827	-0.5728	0.5821	-0.5263
Pr	-0.9044	0.6598	-0.0518	0.2811	-0.2279	0.6085	1.0000	0.7028	0.6108	0.8234	-0.5254	-0.8657	-0.8625	0.1306	0.6630	-0.0011	0.7945	-0.9013	0.9066	-0.8503
Bu	-0.8517	0.8396	0.3738	0.7510	-0.6022	0.8258	0.7028	1.0000	0.2856	0.9076	-0.3652	-0.8502	-0.8488	0.2714	0.7200	0.1505	0.8119	-0.8546	0.8437	-0.8434
BCFA	-0.7030	-0.0920	-0.1290	-0.3578	0.3942	-0.1557	0.6108	0.2856	1.0000	0.1790	-0.3338	-0.6930	-0.6915	0.1869	0.5704	0.0863	0.6542	-0.7039	0.6989	-0.6855
TotalVFA	-0.8065	0.9379	0.2677	0.7693	-0.6530	0.9430	0.8234	0.9076	0.1790	1.0000	-0.4872	-0.7671	-0.7639	0.0951	0.5775	-0.0228	0.6994	-0.8030	0.8099	-0.7522
Crude fiber	0.5212	-0.3256	0.0347	-0.3105	0.2897	-0.3841	-0.5254	-0.3652	-0.3338	-0.4872	1.0000	0.3116	0.3003	0.7230	0.1320	0.8157	-0.1153	0.4886	-0.5767	0.2610
CP	0.9734	-0.6028	-0.2053	-0.3349	0.1921	-0.5391	-0.8657	-0.8502	-0.6930	-0.7671	0.3116	1.0000	0.9999	-0.4312	-0.9008	-0.2955	-0.9798	0.9813	-0.9560	0.9986
ADF	0.9705	-0.6010	-0.2066	-0.3323	0.1892	-0.5363	-0.8625	-0.8488	-0.6915	-0.7639	0.3003	0.9999	1.0000	-0.4419	-0.9059	-0.3069	-0.9821	0.9790	-0.9524	0.9992
NDF	-0.2128	0.1291	0.1823	-0.0513	0.1355	0.0273	0.1306	0.2714	0.1869	0.0951	0.7230	-0.4312	-0.4419	1.0000	0.7802	0.9894	0.6029	-0.2495	0.1474	-0.4782
Total DF	-0.7772	0.4800	0.2301	0.2075	-0.0680	0.3868	0.6630	0.7200	0.5704	0.5775	0.1320	-0.9008	-0.9059	0.7802	1.0000	0.6811	0.9694	-0.8004	0.7337	-0.9225
Total NSP	-0.0686	0.0396	0.1600	-0.1083	0.1743	-0.0579	-0.0011	0.1505	0.0863	-0.0228	0.8157	-0.2955	-0.3069	0.9894	0.6811	1.0000	0.4806	-0.1062	0.0022	-0.3456
Insoluble NSP	-0.9078	0.5616	0.2220	0.2848	-0.1398	0.4827	0.7945	0.8119	0.6542	0.6994	-0.1153	-0.9798	-0.9821	0.6029	0.9694	0.4806	1.0000	-0.9230	0.8780	-0.9890
Soluble NSP	0.9993	-0.6194	-0.1815	-0.3704	0.2350	-0.5728	-0.9013	-0.8546	-0.7039	-0.8030	0.4886	0.9813	0.9790	-0.2495	-0.8004	-0.1062	-0.9230	1.0000	-0.8946	0.9698
Starch	-0.9978	0.6188	0.1658	0.3838	-0.2546	0.5821	0.9066	0.8437	0.6989	0.8099	-0.5767	-0.9560	-0.9524	0.1474	0.7337	0.0022	0.8780	-0.9946	0.8780	-0.9946
Viscosity	0.9599	-0.5943	-0.2105	-0.3230	0.1790	-0.5263	-0.8503	-0.8434	-0.6855	-0.7522	0.2610	0.9986	0.9992	-0.4782	-0.9225	-0.3456	-0.9890	0.9698	0.8780	-0.9946

Summary Plots





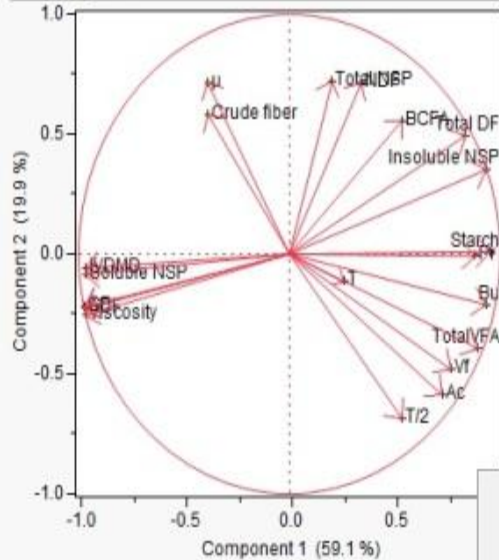
Principal Component Analysis

PCA dataset 2 March 2012 - Principal Components - JMP

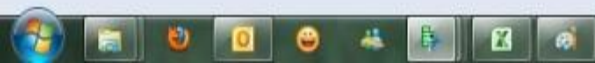
Principal Components: on Correlations

Summary Plots

Loading Plot

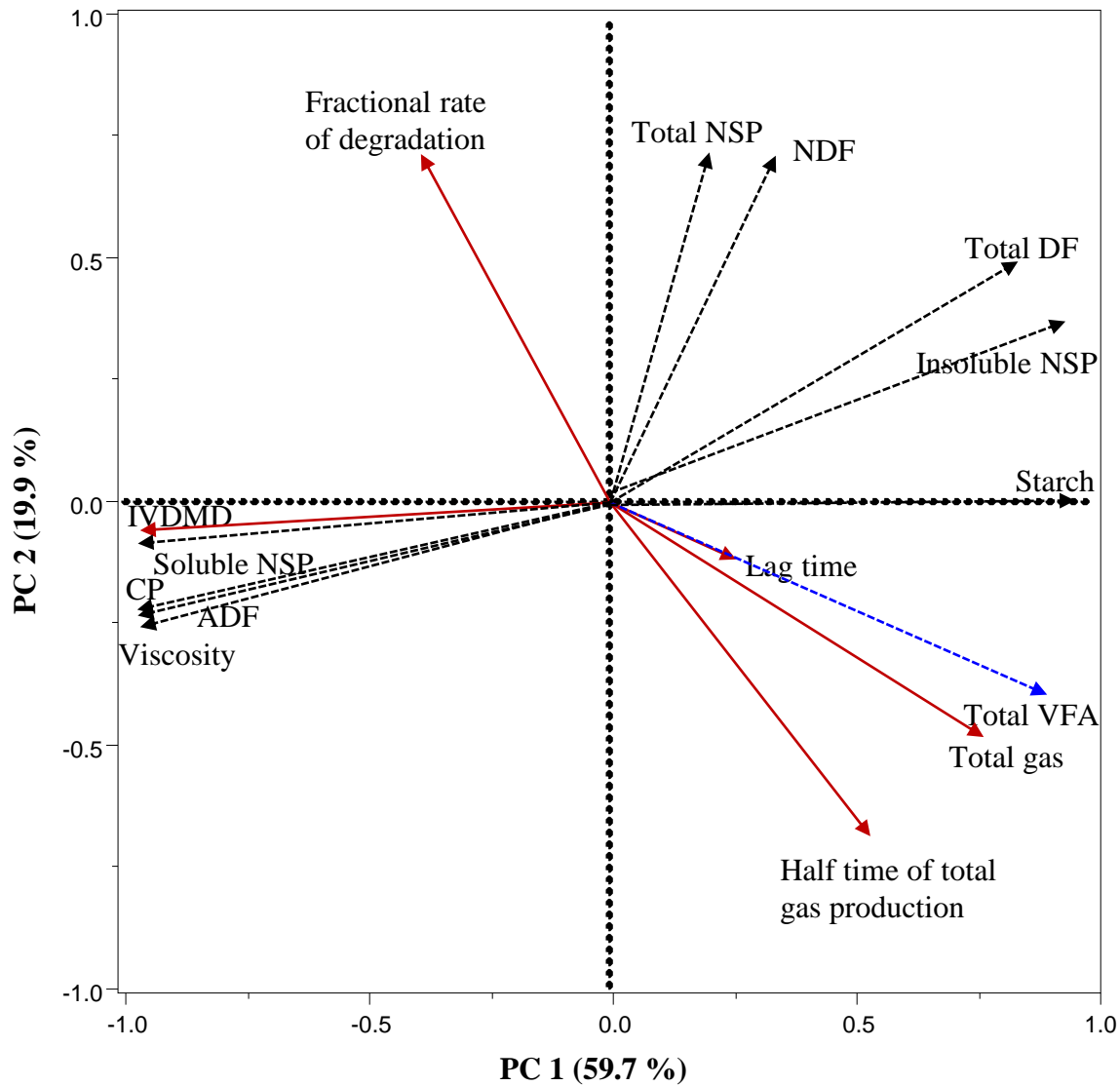


- Background Color...
- Marker Size
- Marker Drawing Mode
- Marker Selection Mode
- Line Width Scale
- Border
- Size/Scale
 - X Axis
 - Y Axis
 - Right Y Axis
 - Frame Size
 - Size to Isometric
- Transparency...
- Customize...
- Edit





Multivariate PCA- loading plot



Relationship between physico-chemical properties and fermentation characteristics of DDGS



Principle component analysis using JMP for better visualization of data

Rajesh Jha, PhD

Research Associate (Swine Nutrition)

Ruurd Zijlstra, PhD

Professor (Swine Nutrition)

Department of Agricultural, Food and Nutritional Science
University of Alberta