

HW5

Jiangyan Feng (Netid: jf8)

Exercise 1

a)

We first fit the log-linear Poisson model of P4 as a function of Treat, BL, P1, and Age. The results are shown as below. Since both the scaled deviance (2.7003) and scaled Pearson X2 (2.5109) are larger than 1, there seems to be overdispersion.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	54	145.8141	2.7003
Scaled Deviance	54	145.8141	2.7003
Pearson Chi-Square	54	135.5868	2.5109
Scaled Pearson X2	54	135.5868	2.5109
Log Likelihood		591.2912	
Full Log Likelihood		-166.7912	
AIC (smaller is better)		343.5825	
AICC (smaller is better)		344.7146	
BIC (smaller is better)		353.9702	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	0.9023	0.3054	0.3036	1.5010	8.73	0.0031
Treat	1	-0.3100	0.1084	-0.5226	-0.0975	8.17	0.0043
BL	1	0.0195	0.0025	0.0146	0.0245	59.00	<.0001
P1	1	0.0039	0.0036	-0.0031	0.0109	1.20	0.2732
Age	1	0.0120	0.0088	-0.0051	0.0292	1.89	0.1696
Scale	0	1.0000	0.0000	1.0000	1.0000		

We then estimate the scale based on deviance. The results are shown as below. Now both the scaled deviance (1) and scaled Pearson X2 (0.9299) is close to 1.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	54	145.8141	2.7003
Scaled Deviance	54	54.0000	1.0000
Pearson Chi-Square	54	135.5868	2.5109
Scaled Pearson X2	54	50.2125	0.9299
Log Likelihood		218.9756	
Full Log Likelihood		-166.7912	
AIC (smaller is better)		343.5825	
AICC (smaller is better)		344.7146	
BIC (smaller is better)		353.9702	

According to the Pvalue in Maximum likelihood parameter estimates, Type 1 and 3 analyses, only BL is significant with Pvalue = <.0001 less than 0.05. Only for BL, 0 is not within the Wald 95% confidence limit. Together, only BL should be retained in the model.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	0.9023	0.5019	-0.0814	1.8861	3.23	0.0722
Treat	1	-0.3100	0.1782	-0.6593	0.0392	3.03	0.0819
BL	1	0.0195	0.0042	0.0113	0.0277	21.85	<.0001
P1	1	0.0039	0.0058	-0.0076	0.0154	0.44	0.5049
Age	1	0.0120	0.0144	-0.0162	0.0402	0.70	0.4032
Scale	0	1.6432	0.0000	1.6432	1.6432		

Note: The scale parameter was estimated by the square root of DEVIANCE/DOF.

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	476.2487						
Treat	473.0840	1	54	1.17	0.2838	1.17	0.2790
BL	149.6763	1	54	119.77	<.0001	119.77	<.0001
P1	147.6894	1	54	0.74	0.3948	0.74	0.3910
Age	145.8141	1	54	0.69	0.4083	0.69	0.4046

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Treat	1	54	3.06	0.0860	3.06	0.0803
BL	1	54	20.18	<.0001	20.18	<.0001
P1	1	54	0.45	0.5065	0.45	0.5037
Age	1	54	0.69	0.4083	0.69	0.4046

b)

We first fit the log-linear Poisson model of P4 as a function of BL. The results are shown as below. Since both the scaled deviance (2.8060) and scaled Pearson X2 (2.6526) are larger than 1, there seems to be overdispersion.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	57	159.9413	2.8060
Scaled Deviance	57	159.9413	2.8060
Pearson Chi-Square	57	151.2008	2.6526
Scaled Pearson X2	57	151.2008	2.6526
Log Likelihood		584.2276	
Full Log Likelihood		-173.8548	
AIC (smaller is better)		351.7097	
AICC (smaller is better)		351.9240	
BIC (smaller is better)		355.8648	

We then estimate the scale based on deviance. The results are shown as below. Now both the scaled deviance (1) and scaled Pearson X2 (0.9454) is close to 1.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	57	159.9413	2.8060
Scaled Deviance	57	57.0000	1.0000
Pearson Chi-Square	57	151.2008	2.6526

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Scaled Pearson X2	57	53.8851	0.9454
Log Likelihood		208.2074	
Full Log Likelihood		-173.8548	
AIC (smaller is better)		351.7097	
AICC (smaller is better)		351.9240	
BIC (smaller is better)		355.8648	

According to the Pvalue in Maximum likelihood parameter estimates, Type 1 and 3 analyses, BL is significant with Pvalue = <.0001 less than 0.05. For BL, 0 is not within the Wald 95% confidence limit. Together, BL is significant.

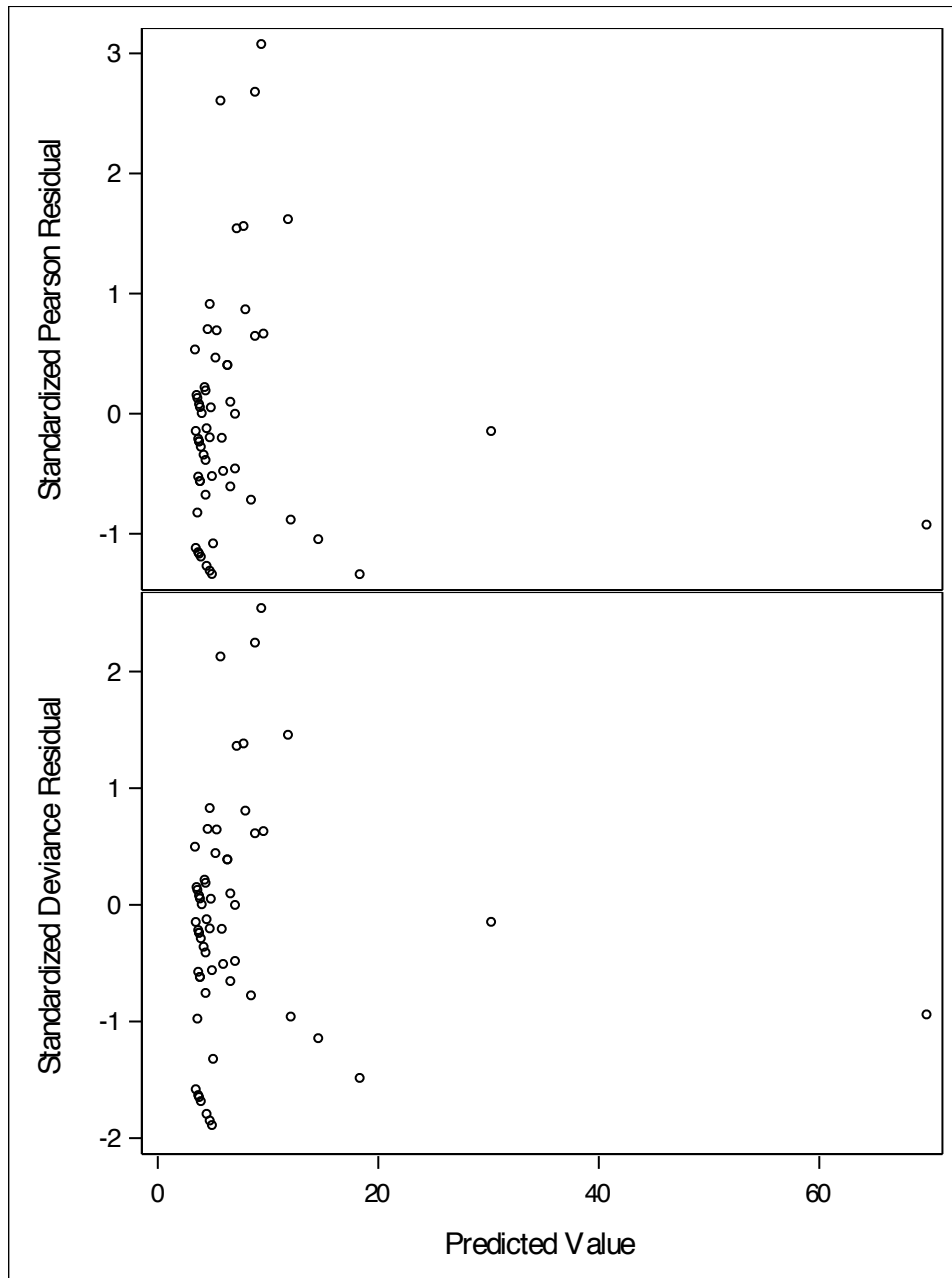
Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.0897	0.1295	0.8360	1.3435	70.84	<.0001
BL	1	0.0209	0.0017	0.0176	0.0242	156.27	<.0001
Scale	0	1.6751	0.0000	1.6751	1.6751		

Note: The scale parameter was estimated by the square root of DEVIANCE/DOF.

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	476.2487						
BL	159.9413	1	57	112.73	<.0001	112.73	<.0001

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
BL	1	57	112.73	<.0001	112.73	<.0001

According to the standardized pearson and deviance residual plots, the residuals are flat and there is no trend. Therefore, there is no problem with the model assumption.



According to the parameter estimates, estimate for BL is 0.0209. It suggests with 1 unit increase in BL, there will be $\exp(0.029) = 1.021$ fold increase in the seizure counts after four treatment periods.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.0897	0.1295	0.8360	1.3435	70.84	<.0001
BL	1	0.0209	0.0017	0.0176	0.0242	156.27	<.0001
Scale	0	1.6751	0.0000	1.6751	1.6751		

Exercise 2

a)

We first fit the log-linear Poisson model of P1 as a function of Treat, BL, and Age. The results are shown as below. Since both the scaled deviance (3.4069) and scaled Pearson X2 (3.6267) are larger than 1, there seems to be overdispersion.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	55	187.3802	3.4069
Scaled Deviance	55	187.3802	3.4069
Pearson Chi-Square	55	199.4665	3.6267
Scaled Pearson X2	55	199.4665	3.6267
Log Likelihood		908.6710	
Full Log Likelihood		-191.9378	
AIC (smaller is better)		391.8756	
AICC (smaller is better)		392.6163	
BIC (smaller is better)		400.1857	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.2411	0.2733	-0.7768	0.2946	0.78	0.3778
Treat	1	-0.1189	0.0926	-0.3005	0.0627	1.65	0.1994
BL	1	0.0257	0.0010	0.0238	0.0276	695.56	<.0001
Age	1	0.0465	0.0078	0.0312	0.0619	35.42	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

We then estimate the scale based on deviance. The results are shown as below. Now both the scaled deviance (1) and scaled Pearson X2 (1.0645) is close to 1.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	55	187.3802	3.4069
Scaled Deviance	55	55.0000	1.0000
Pearson Chi-Square	55	199.4665	3.6267

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Scaled Pearson X2	55	58.5476	1.0645
Log Likelihood		266.7139	
Full Log Likelihood		-191.9378	
AIC (smaller is better)		391.8756	
AICC (smaller is better)		392.6163	
BIC (smaller is better)		400.1857	

According to the Pvalue in Maximum likelihood parameter estimates, Type 1 and 3 analyses, only BL and Age is significant with Pvalue less than 0.05. Only for BL and Age, 0 is not within the Wald 95% confidence limit. Together, BL and Age should be retained in the model.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.2411	0.5045	-1.2299	0.7477	0.23	0.6328
Treat	1	-0.1189	0.1710	-0.4540	0.2163	0.48	0.4869
BL	1	0.0257	0.0018	0.0222	0.0293	204.16	<.0001
Age	1	0.0465	0.0144	0.0182	0.0748	10.40	0.0013
Scale	0	1.8458	0.0000	1.8458	1.8458		

Note: The scale parameter was estimated by the square root of DEVIANCE/DOF.

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	746.4361						
Treat	745.4460	1	55	0.29	0.5920	0.29	0.5898
BL	222.2496	1	55	153.57	<.0001	153.57	<.0001
Age	187.3802	1	55	10.23	0.0023	10.23	0.0014

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Treat	1	55	0.48	0.4896	0.48	0.4867
BL	1	55	163.79	<.0001	163.79	<.0001
Age	1	55	10.23	0.0023	10.23	0.0014

b)

We first fit the log-linear Poisson model of P1 as a function of BL and Age. The results are shown as below. Since both the scaled deviance (3.3755) and scaled Pearson X2 (3.5746) are larger than 1, there seems to be overdispersion.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	56	189.0289	3.3755
Scaled Deviance	56	189.0289	3.3755
Pearson Chi-Square	56	200.1801	3.5746
Scaled Pearson X2	56	200.1801	3.5746
Log Likelihood		907.8467	
Full Log Likelihood		-192.7621	
AIC (smaller is better)		391.5243	
AICC (smaller is better)		391.9606	
BIC (smaller is better)		397.7569	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.3815	0.2520	-0.8754	0.1123	2.29	0.1300
BL	1	0.0257	0.0010	0.0238	0.0276	694.89	<.0001
Age	1	0.0494	0.0075	0.0348	0.0641	43.60	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

We then estimate the scale based on deviance. The results are shown as below. Now both the scaled deviance (1) and scaled Pearson X2 (1.0590) is close to 1.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	56	189.0289	3.3755
Scaled Deviance	56	56.0000	1.0000
Pearson Chi-Square	56	200.1801	3.5746
Scaled Pearson X2	56	59.3036	1.0590

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		268.9505	
Full Log Likelihood		-192.7621	
AIC (smaller is better)		391.5243	
AICC (smaller is better)		391.9606	
BIC (smaller is better)		397.7569	

According to the Pvalue in Maximum likelihood parameter estimates, Type 1 and 3 analyses, BL and Age are significant with Pvalue less than 0.05. For BL and Age, 0 is not within the Wald 95% confidence limit. Together, BL and Age are significant.

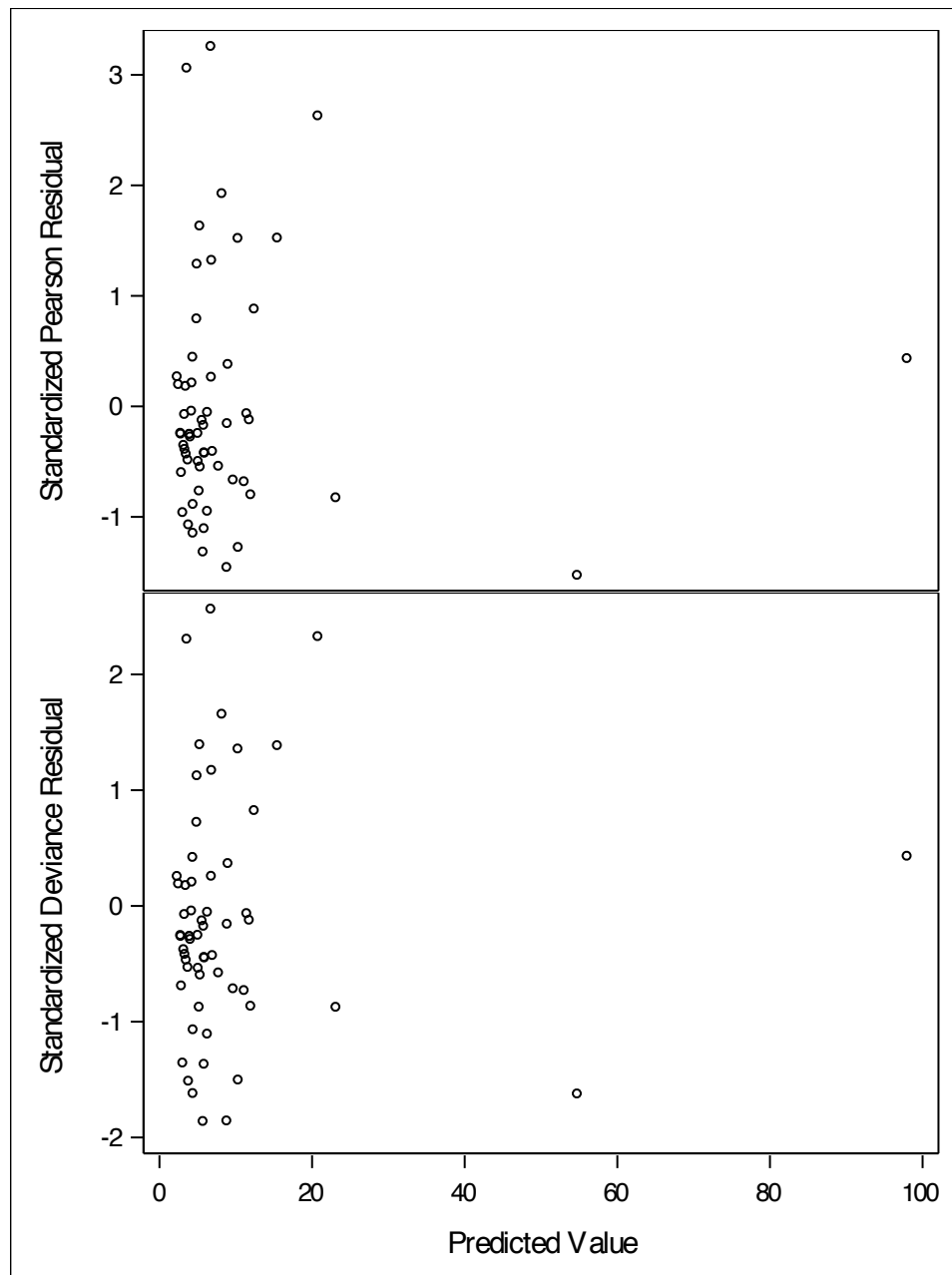
Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.3815	0.4630	-1.2890	0.5259	0.68	0.4099
BL	1	0.0257	0.0018	0.0222	0.0292	205.86	<.0001
Age	1	0.0494	0.0138	0.0225	0.0764	12.92	0.0003
Scale	0	1.8373	0.0000	1.8373	1.8373		

Note: The scale parameter was estimated by the square root of DEVIANCE/DOF.

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	746.4361						
BL	231.9037	1	56	152.43	<.0001	152.43	<.0001
Age	189.0289	1	56	12.70	0.0008	12.70	0.0004

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
BL	1	56	165.11	<.0001	165.11	<.0001
Age	1	56	12.70	0.0008	12.70	0.0004

According to the standardized pearson and deviance residual plots, the residuals are flat and there is no trend. Therefore, there is no problem with the model assumption.



According to the parameter estimates, estimate for BL and Age is 0.0257 and 0.0494. It suggests with 1 unit increase in BL, there will be $\exp(0.0257)$ and $\exp(0.0494)$ fold increase in the seizure counts after one treatment periods.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.3815	0.4630	-1.2890	0.5259	0.68	0.4099
BL	1	0.0257	0.0018	0.0222	0.0292	205.86	<.0001
Age	1	0.0494	0.0138	0.0225	0.0764	12.92	0.0003
Scale	0	1.8373	0.0000	1.8373	1.8373		

Similarities and differences:

Similarities: The seizure counts at the first and last visit all depend on BL. And the estimate for BL is similar for both the first and the last visit.

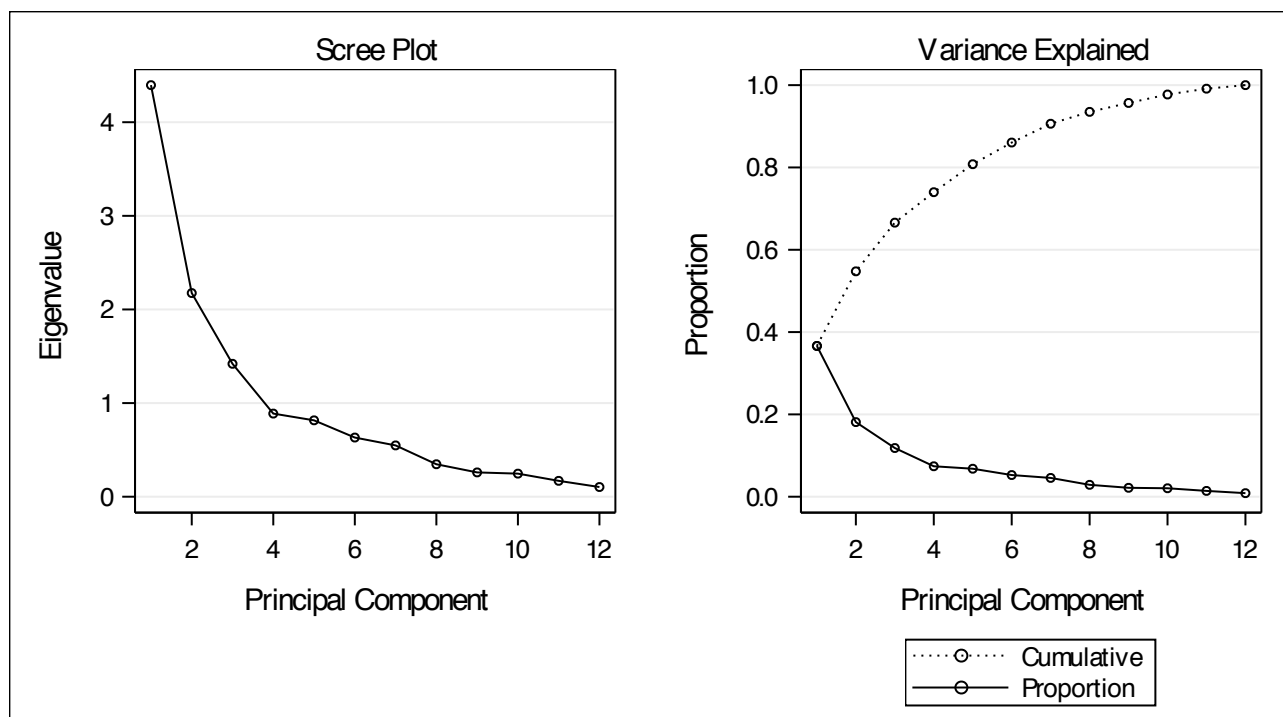
Difference: The seizure counts at the last visit only depend on BL. However, the seizure counts at the first visit also depend on the edge.

Exercise 3

a)

The results of principal component analysis are shown as below. We need to keep 4 components (cumulative=0.7399) to retain at least 70% of the variation. Based on average eigenvalue, the eigenvalue of the first 3 components is larger than 1. Based on the Scree plot, the elbow is 3. So we may choose 3 components based on average eigenvalue and Scree plot.

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	4.39595632	2.22065955	0.3663	0.3663
2	2.17529677	0.75572719	0.1813	0.5476
3	1.41956959	0.53169566	0.1183	0.6659
4	0.88787392	0.07206032	0.0740	0.7399
5	0.81581361	0.18385862	0.0680	0.8079
6	0.63195499	0.08407250	0.0527	0.8605
7	0.54788249	0.20115694	0.0457	0.9062
8	0.34672554	0.08719022	0.0289	0.9351
9	0.25953532	0.01307050	0.0216	0.9567
10	0.24646482	0.07696032	0.0205	0.9773
11	0.16950450	0.06608236	0.0141	0.9914
12	0.10342214		0.0086	1.0000



b)

For the component 1, the largest positive coefficient is nonflavanoid_phenols, the largest negative coefficient is from proanthocyanins. So prin1 may contrast nonflavanoid_phenols over proanthocyanins. With the increase in nonflavanoid_phenols and the decrease in proanthocyanins, the prin1 will increase.

For the component 2, the largest positive coefficient is hue, the largest negative coefficient is from od280_od315. So prin1 may contrast hue over od280_od315. With the increase in hue and the decrease in od280_od315, the prin2 will increase.

For the component 3, the largest positive coefficient is magnesium, the largest negative coefficient is from malic_acid. So prin1 may contrast magnesium over malic_acid. With the increase in magnesium and the decrease in malic_acid, the prin3 will increase.

For the component 4, the largest positive coefficient is ash, the largest negative coefficient is from total_phenols. So prin1 may contrast ash over total_phenols. With the increase in ash and the decrease in total_phenols, the prin4 will increase.

Eigenvectors								
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8
malic_acid	0.091117	0.513423	-.272161	0.056347	0.346100	-.307291	0.011071	0.476373
ash	-.270461	0.212854	0.053675	0.479292	-.374837	-.449505	0.424981	0.059783
alcalinity_ash	-.032286	0.368823	0.604043	-.111362	0.223412	-.198971	-.121932	-.178046
magnesium	-.234041	0.000809	0.622592	0.057773	-.149722	0.134509	-.323528	0.414439
total_phenols	0.115882	0.360852	0.079682	-.690601	-.411900	-.030540	0.339990	-.119475
flavanoids	0.401759	0.174439	0.101046	0.217309	0.092075	0.055670	-.012998	-.373440
nonflavanoid_phenols	0.437057	0.104214	0.116337	0.167173	0.058894	-.001610	-.049458	-.173616
proanthocyanins	-.311970	-.041107	0.211733	0.083123	0.519266	0.244948	0.601632	-.209902
color	0.324574	0.146536	0.095412	0.260981	-.269933	0.603280	0.313095	0.340714
hue	-.147665	0.537132	-.208820	0.029726	0.170214	0.359295	-.227546	-.018474
od280_od315	0.328153	-.256709	0.133869	-.273486	0.336884	-.142788	0.262469	0.465785
proline	0.405845	-.077350	0.158127	0.220377	-.028456	-.261483	-.029436	-.068693

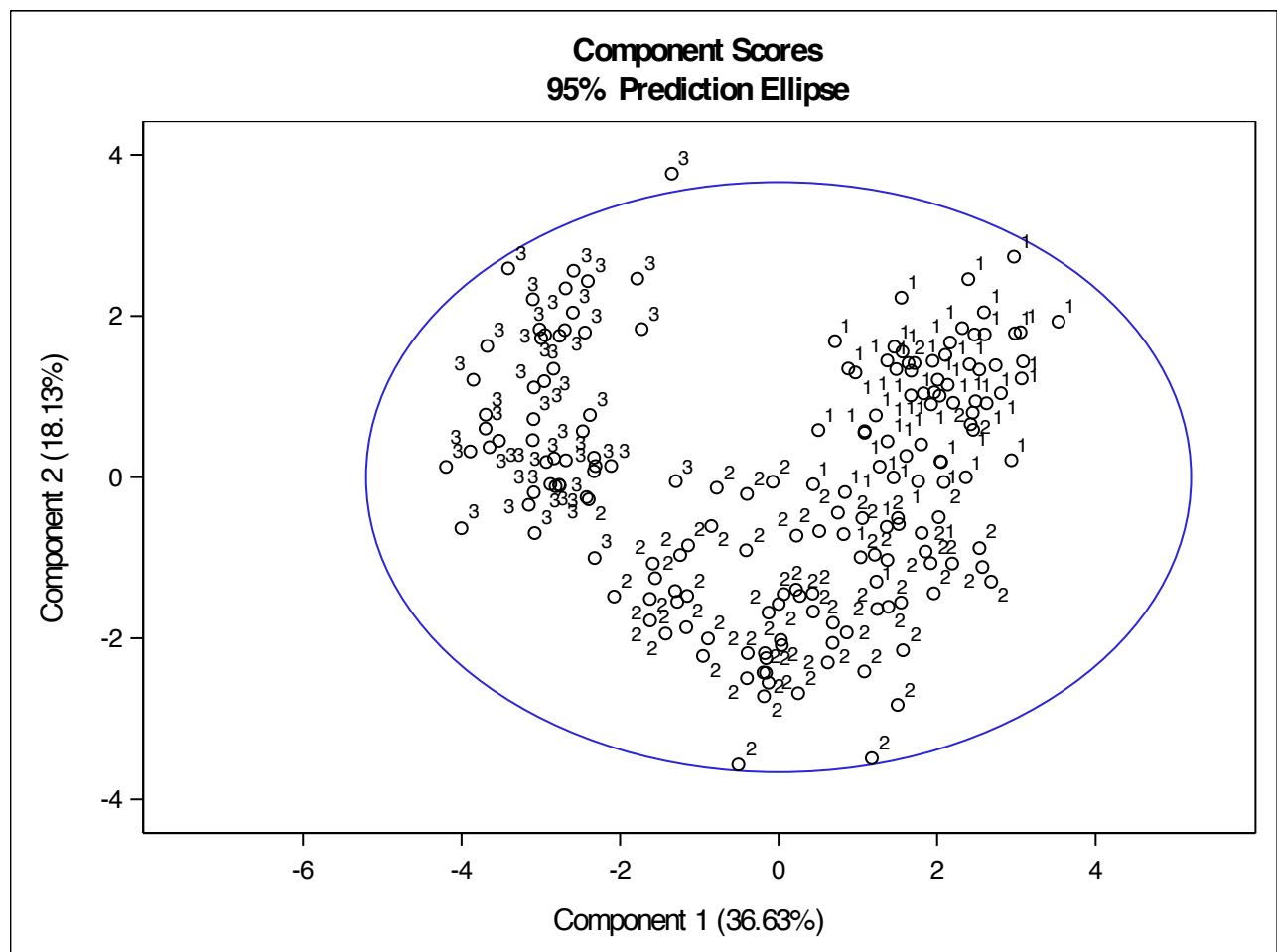
Eigenvectors				
	Prin9	Prin10	Prin11	Prin12
malic_acid	0.310651	-.152282	-.299147	0.019934
ash	-.201018	0.255290	0.126378	0.025478
alcalinity_ash	-.455102	-.342277	-.080059	-.137506
magnesium	0.389505	0.278728	-.020473	0.086627
total_phenols	0.248079	0.066514	0.056849	0.058457

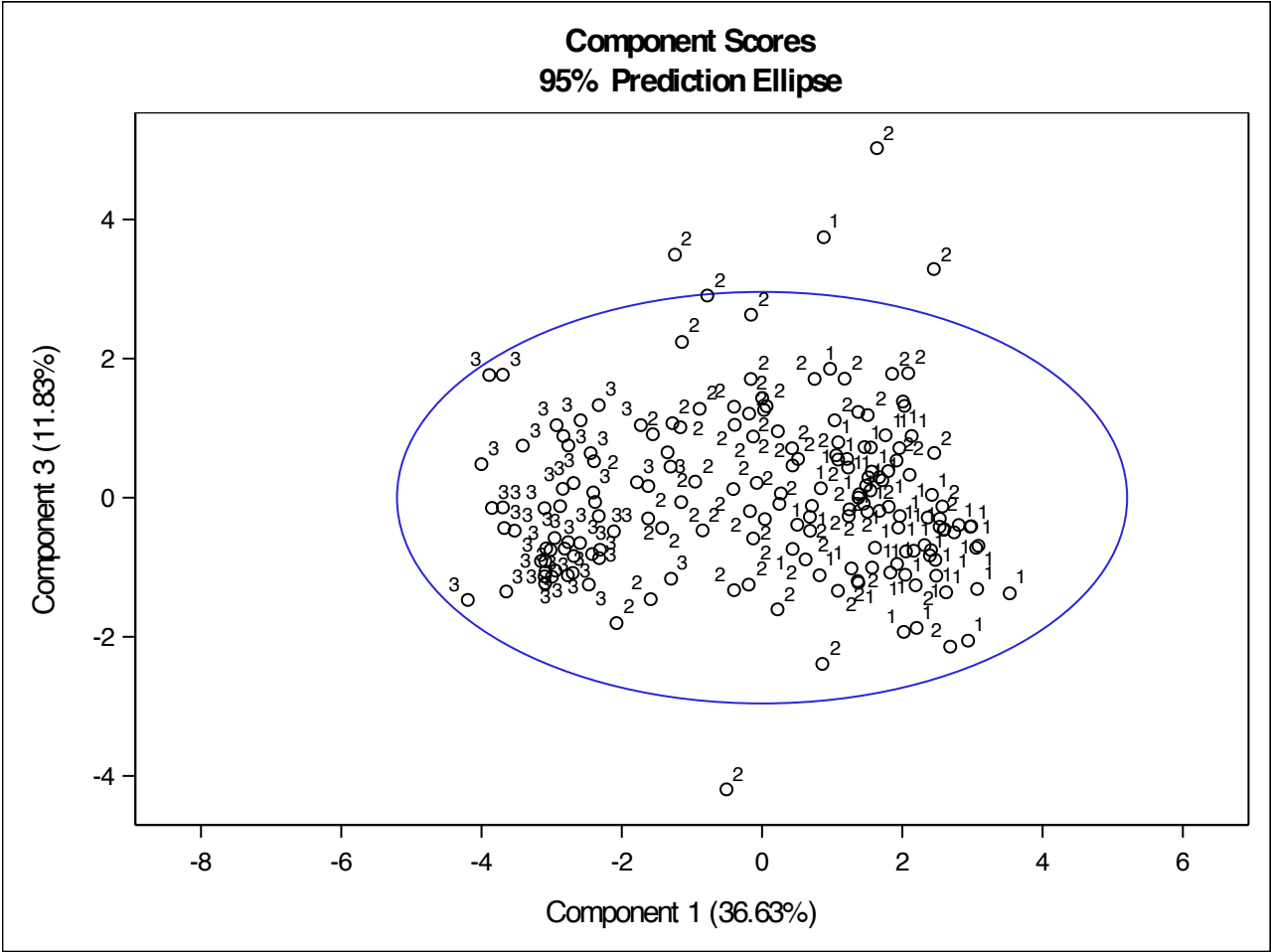
Eigenvectors				
	Prin9	Prin10	Prin11	Prin12
flavanoids	0.167235	0.517817	-.295946	-.464584
nonflavanoid_phenols	-.066495	0.135231	-.053219	0.833690
proanthocyanins	0.310269	-.046205	0.047901	0.113542
color	-.176352	-.299390	-.103626	-.116475
hue	-.134160	0.249965	0.595363	-.008470
od280_od315	-.316105	0.382656	0.245096	-.086924
proline	0.406245	-.353937	0.605393	-.155437

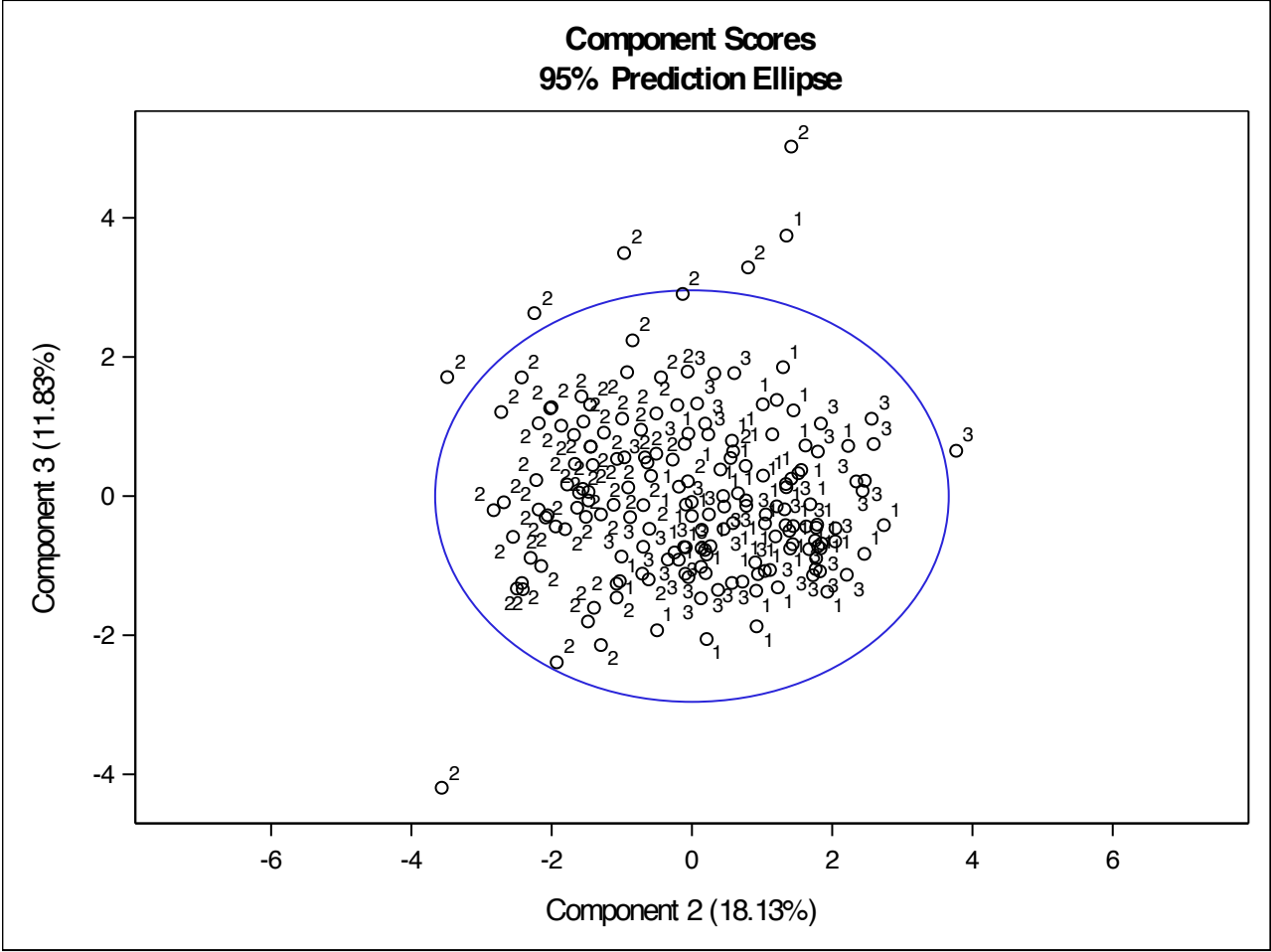
c)

The score plots are shown below for prin1 to prin4. Only prin1-2 plot provides clear associations.

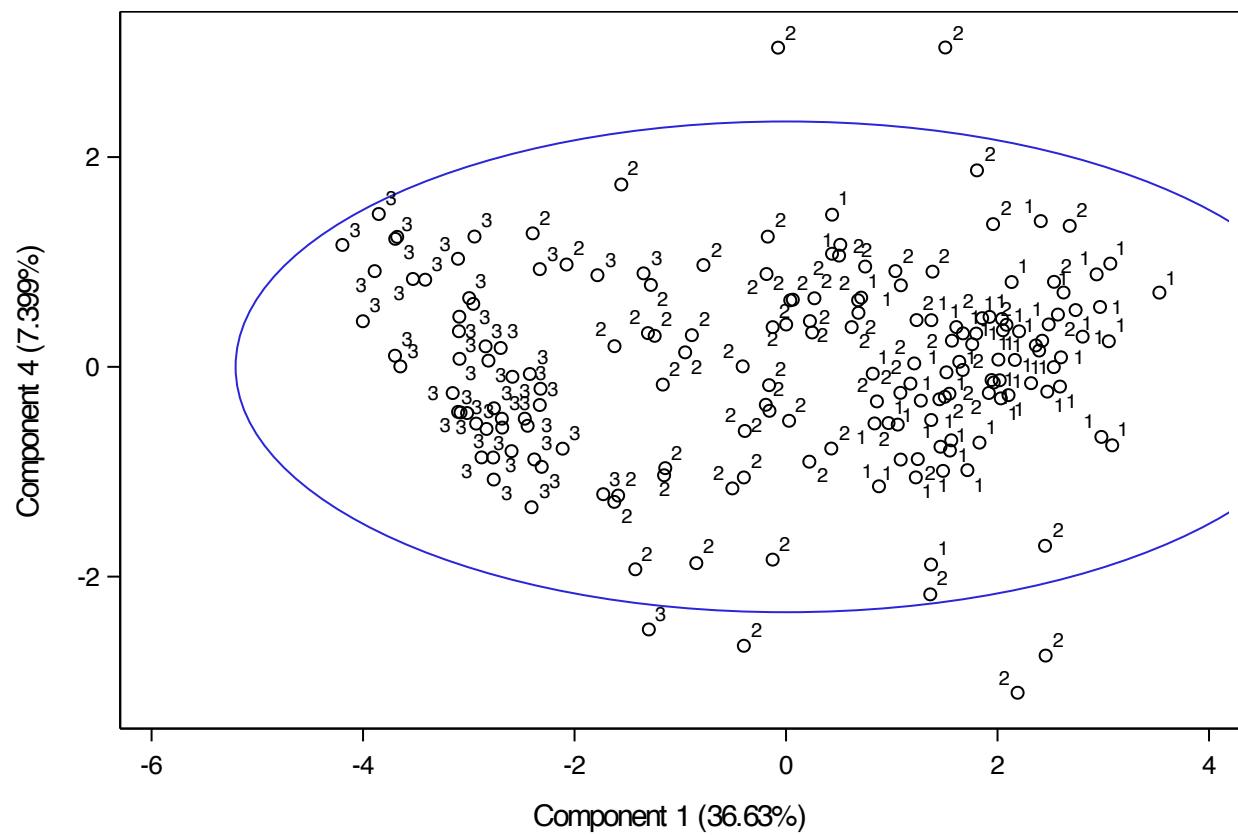
According to prin1-2, alcohol=1 seems to have larger prin1 and prin2 value whereas alcohol=3 seems to have the smaller prin1 and larger prin2. Based on the eigenvectors in part b, alcohol=1 may have the largest amount of nonflavanoid_phenols and hue. Alcohol=3 may have the lowest amount of nonflavanoid_phenols and the largest amount of hue.

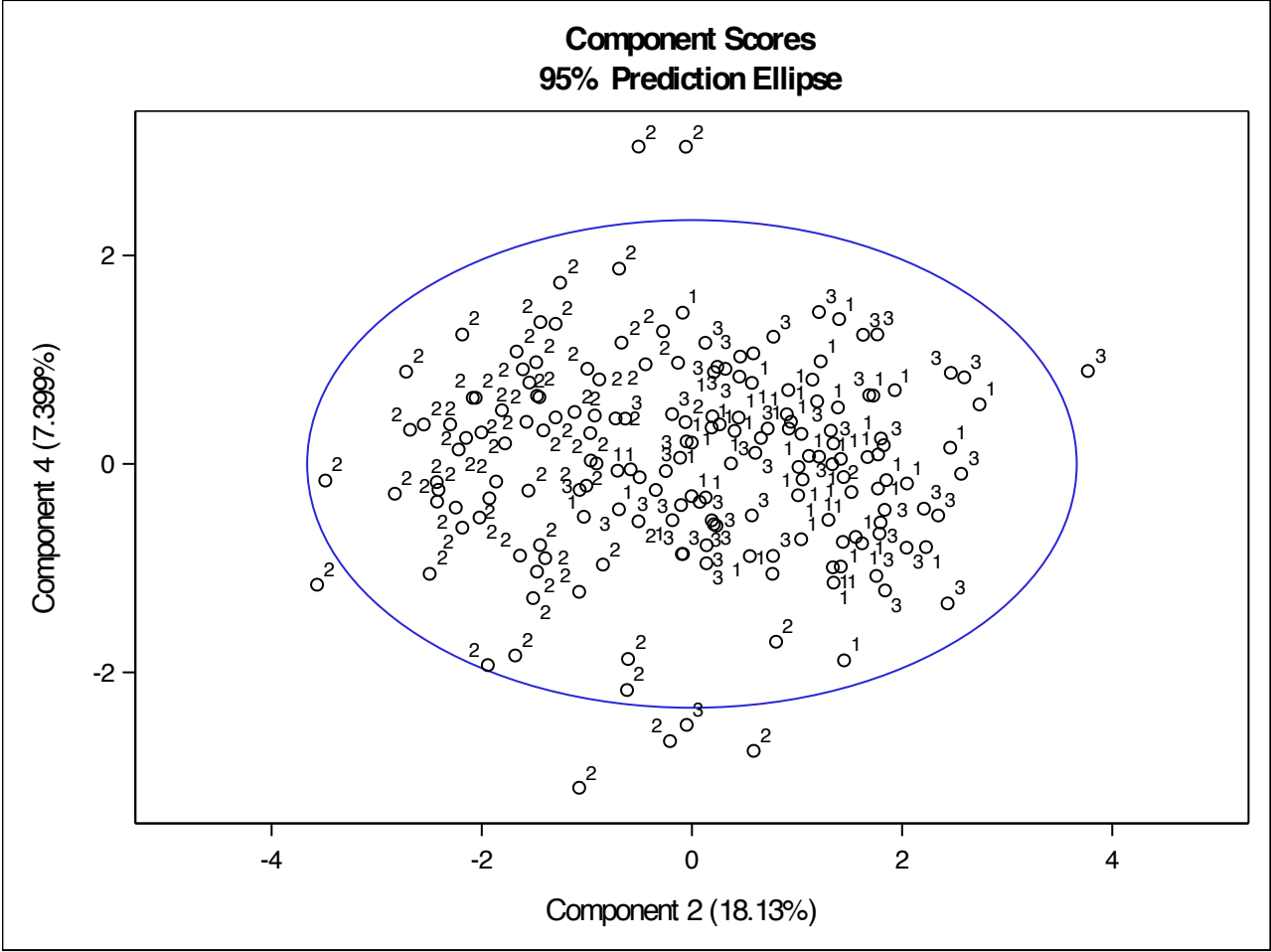




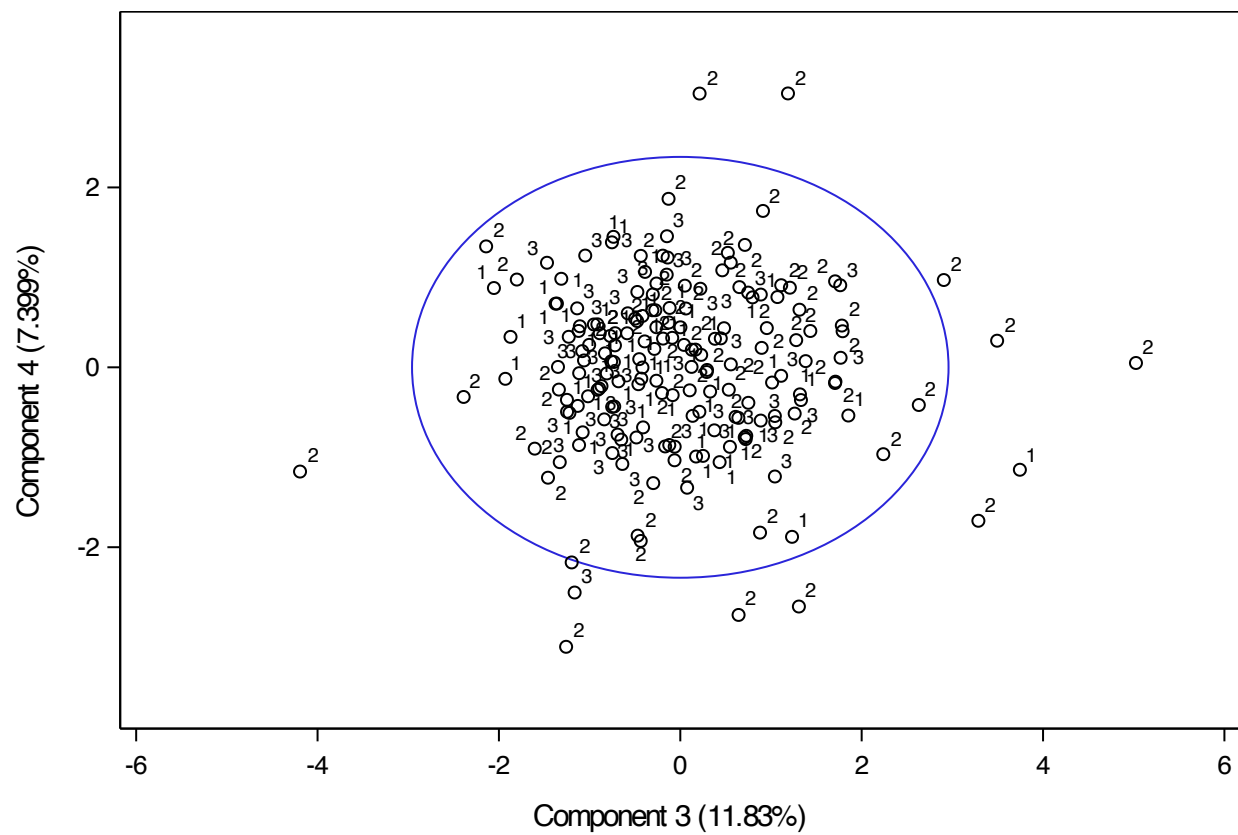


Component Scores
95% Prediction Ellipse





Component Scores
95% Prediction Ellipse

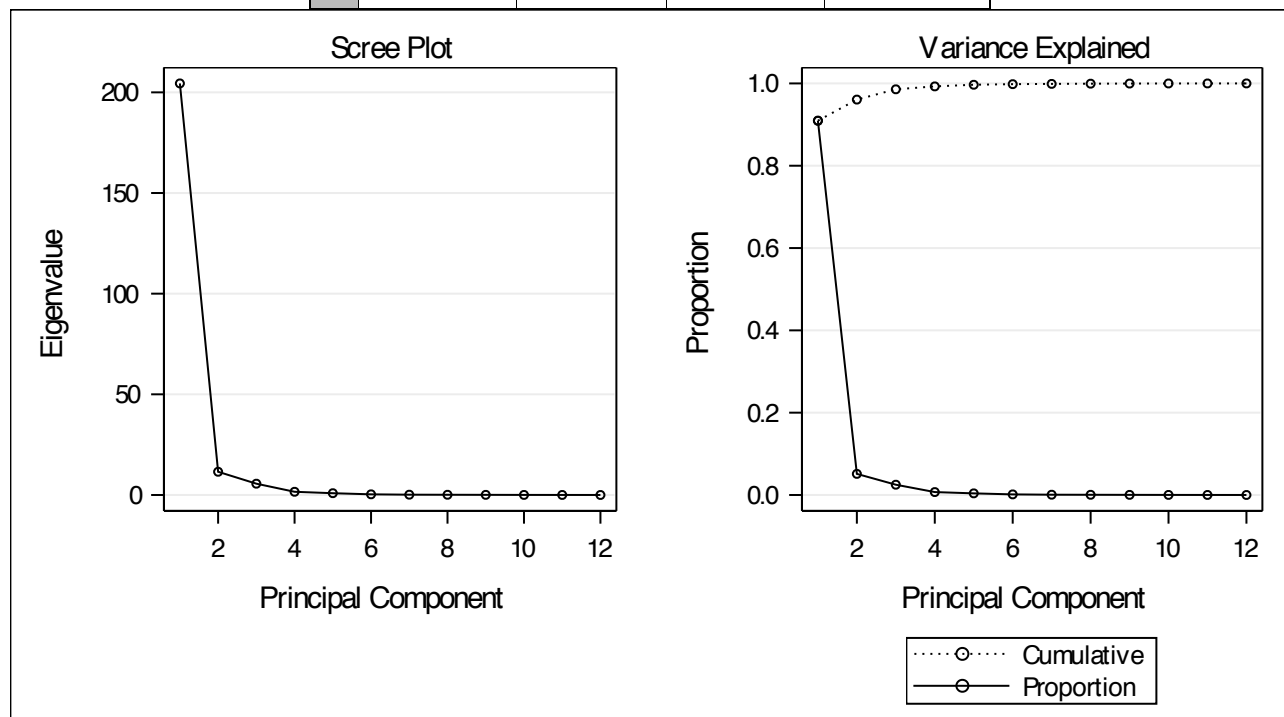


Exercise 4

a)

The results of principal component analysis are shown as below. We need to keep 1 component (cumulative=0.9094) to retain at least 70% of the variation. Based on average eigenvalue, the eigenvalue of the first 4 components is larger than 1. Based on the Scree plot, the elbow is 1. So we may choose 4 components based on the average eigenvalue and 1 component based on Scree plot.

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	204.430157	192.898984	0.9094	0.9094
2	11.531173	5.918322	0.0513	0.9607
3	5.612851	4.019780	0.0250	0.9857
4	1.593071	0.703453	0.0071	0.9928
5	0.889618	0.563699	0.0040	0.9967
6	0.325918	0.174534	0.0014	0.9982
7	0.151385	0.039105	0.0007	0.9989
8	0.112279	0.040485	0.0005	0.9994
9	0.071795	0.032056	0.0003	0.9997
10	0.039738	0.018336	0.0002	0.9999
11	0.021403	0.013154	0.0001	1.0000
12	0.008249		0.0000	1.0000



b)

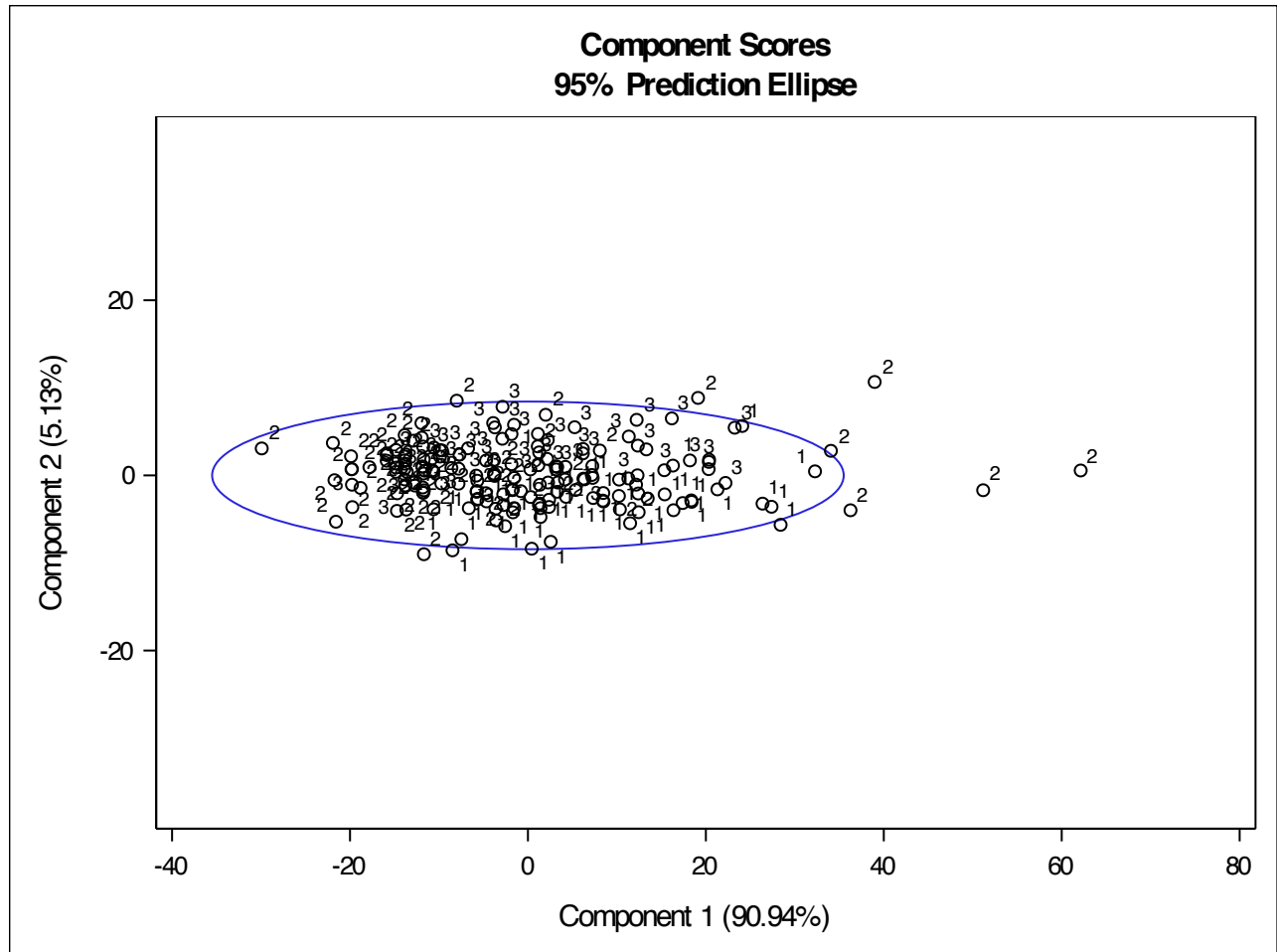
For the component 1, the largest positive coefficient is total_phenols (0.998890), the negative coefficients are all trivial. So prin1 may directly correlate with total_phenols. With the increase in total_phenols, the prin1 will increase.

Eigenvectors								
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8
malic_acid	0.015668	-.065065	0.187912	0.161677	0.240292	0.885632	0.287494	-.070489
ash	-.004332	0.113858	0.151458	-.430817	0.862211	-.157531	-.065092	-.018076
alcalinity_ash	0.005480	0.037488	0.016268	0.063172	0.049902	0.074794	-.149892	-.080382
magnesium	-.020748	0.974629	-.099771	0.184381	-.009645	0.054118	0.015088	-.002555
total_phenols	0.998890	0.022272	-.033182	-.023580	-.004593	-.002403	-.003512	0.002266
flavanoids	0.009507	-.067339	-.037169	0.368905	0.185038	-.083839	-.176913	-.250621
nonflavanoid_phenols	0.013847	-.120155	-.109412	0.615473	0.281703	-.131963	-.247855	-.374679
proanthocyanins	-.002245	0.013882	0.010282	-.031499	-.012403	0.005546	0.006342	-.038127
color	0.009540	-.037778	-.028309	0.264997	0.147812	-.355282	0.869821	0.047116
hue	0.033232	0.067341	0.944479	0.205120	-.111605	-.167790	-.081176	0.099885
od280_od315	0.000877	-.022772	-.054129	0.065496	-.027146	0.047273	-.003208	-.036197
proline	0.003303	-.073173	-.145601	0.342398	0.217849	0.034986	-.185593	0.877540

Eigenvectors				
	Prin9	Prin10	Prin11	Prin12
malic_acid	0.008002	-.057316	-.029191	0.012826
ash	-.019859	-.014003	0.070478	-.011416
alcalinity_ash	-.098151	0.942693	-.178441	-.164775
magnesium	0.003472	-.048573	0.009410	0.000902
total_phenols	0.000195	-.003957	0.000400	0.002390
flavanoids	0.849549	0.010319	0.002733	-.026033
nonflavanoid_phenols	-.514036	-.159034	-.038939	0.070461
proanthocyanins	0.035481	0.197010	0.129210	0.969669
color	0.006187	0.141199	-.017620	-.017009
hue	-.022595	-.011772	0.049075	-.003569
od280_od315	-.043031	0.143459	0.970181	-.155841
proline	0.016411	0.024607	0.007936	0.045369

c)

The score plots are shown below for prin1 to prin2. According to prin1-2, alcohol=2 seems to have larger prin1. Based on the eigenvectors in part b, alcohol=2 may have the largest amount of total_phenols.



Differences between correlation and covariance results:

For covariance analysis, only 1 component is used versus 4 in correlation analysis. And the covariance analysis identifies the total_phenols as the feature of different wines whereas the correlation analysis identifies different chemical components as features.