

```

ods rtf file='data visualization @20231215.rtf';

DATA chemometric;
  INFILE 'final.csv' firstobs=2 dsd;
  INPUT molecule group wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100
wv1200 wv1300 wv1400 wv1500 wv1600;
  RUN;

ODS STARTPAGE = NO;

proc capability data=chemometric NORMATEST;
  qqplot wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
  run;

proc corr data=chemometric nomiss plots=matrix(histogram NVAR=all);
  var wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
  run;

proc corr data=chemometric COV plots(maxpoints=NONE)=matrix(histogram);
  var wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
  ods select MatrixPlot;
run;

proc corr data=chemometric COV plots(maxpoints=NONE)=matrix(histogram);
  var wv900 wv1000 wv1100 wv1200;
  ods select MatrixPlot;
run;
proc corr data=chemometric COV plots(maxpoints=NONE)=matrix(histogram);
  var wv1300 wv1400 wv1500 wv1600;
  ods select MatrixPlot;
run;

proc gplot data=chemometric;
  plot wv400*wv500;
run;

proc capability data=chemometric noprint graphics;
  probplot wv800/gamma(alpha=2 sigma=2 theta=0);
run;

/*Add SAS code for Homework 3 below this line and above green line of stars
below*/
PROC GLM;
  CLASS group;
  MODEL wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300
wv1400 wv1500 wv1600 = group;
  MANOVA H = group/PRINTE PRINTE MSTAT=EXACT;
  RUN;

/*********************************************/

```

```
ods rtf close;
```

```

ods rtf file='STAT764_PROJECT_ @20231215.rtf';

title color="green" height=25pt bold italic underlin=1 "Q1: group variance";

DATA chemometric;
  INFILE 'final.csv' firstobs=2 dsd;
  INPUT molecule group wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100
wv1200 wv1300 wv1400 wv1500 wv1600;
  RUN;

ODS STARTPAGE = NO;

*
* Q1: group variances
*
/*Add SAS code for Homework 3 below this line and above green line of stars
below*/
title color="purple" height=25pt bold italic underlin=1 "Chapter6 hw3";


title color="purple" height=15pt bold italic underlin=1 "Chapter6, hw3";
PROC GLM;
  CLASS group;
  MODEL wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300
wv1400 wv1500 wv1600 = group;
  MANOVA H = group/PRINTE PRINTE MSTAT=EXACT;
  RUN;

/*Add SAS code for Homework 4 below this line and above green line of stars
below*/
title color="purple" height=15pt bold italic underlin=1 "Chapter7,8, hw4";
PROC DISCRIM DATA=chemometric POOL=TEST;
  CLASS group;
  VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
  run;

PROC STEPDISC STEPWISE SIMPLE STDMEAN TCORR WCORR;
  CLASS GROUP;
  VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
  TITLE 'STEPWISE SELECTION';
  RUN;

*
* Q2: discriminant function
*
/*Chapter 9* QUADRATIC do not use poopled covariance matrix */

title color="purple" height=25pt bold italic underlin=1 "chapter9
classification";

```

```

title color="green" height=25pt bold italic underlin=1 "Q2: discriminant
function";

title color="purple" height=15pt bold italic underlin=1 "chapter9
classification";

proc DISCRIM DATA=chemometric LIST CROSSVALIDATE POOL=TEST;
CLASS GROUP;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
RUN;

*
* PCA
*
/* Chapter 12 PCA: abstract information among the variables */
title color="purple" height=25pt bold italic underlin=1 "chapter12 PCA";

title color="green" height=25pt bold italic underlin=1 "PCA";

PROC PRINCOMP DATA=chemometric COV OUT=RESULTS;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
RUN;
PROC PRINT DATA=RESULTS;
RUN;
PROC PLOT DATA=RESULTS;
PLOT PRIN2*PRIN1;
RUN;

*
* Q3: Effect of PCA on Group Variance
*
/* Do Analysis Again by RESULTS which is the Principle Components */
/* Add SAS code for Homework 3 below this line and above green line of stars
below */
title color="purple" height=25pt bold italic underlin=1 "Chapter6 hw3 by
pca";

title color="green" height=25pt bold italic underlin=1 "Q3: Effect of PCA on
Group Variance";
title color="purple" height=15pt bold italic underlin=1 "Chapter6, hw3 by
pca";
PROC GLM DATA=RESULTS;
CLASS group;
MODEL PRIN1 PRIN2 PRIN3 PRIN4 PRIN5 PRIN6 PRIN7 PRIN8 PRIN9 PRIN10 PRIN11
PRIN12 PRIN13 = group;
MANOVA H = group/PRINT E MSTAT=EXACT;
RUN;

/*Add SAS code for Homework 4 below this line and above green line of stars
below*/
title color="purple" height=15pt bold italic underlin=1 "Chapter7,8, hw4 by
pca";
PROC DISCRIM DATA=RESULTS POOL=TEST;
CLASS group;

```

```

VAR PRIN1 PRIN2 PRIN3 PRIN4 PRIN5 PRIN6 PRIN7 PRIN8 PRIN9 PRIN10 PRIN11
PRIN12 PRIN13;
run;

PROC STEPDISC DATA=RESULTS STEPWISE SIMPLE STDMEAN TCORR WCORR;
CLASS GROUP;
VAR PRIN1 PRIN2 PRIN3 PRIN4 PRIN5 PRIN6 PRIN7 PRIN8 PRIN9 PRIN10 PRIN11
PRIN12 PRIN13;
TITLE 'STEPWISE SELECTION';
RUN;

/*Chapter 9* QUADRATIC do not use poopled covariance matrix */
title color="purple" height=15pt bold italic underlin=1 "chapter9
classification by pca";
proc DISCRIM DATA=RESULTS LIST CROSSVALIDATE POOL=TEST;
CLASS GROUP;
VAR PRIN1 PRIN2 PRIN3 PRIN4 PRIN5 PRIN6 PRIN7 PRIN8 PRIN9 PRIN10 PRIN11
PRIN12 PRIN13;
RUN;

*
* Q4: K-Means on 4 group overall dataset & PCA
*
/* Chapter 15 K- means cluterung: finding natural groups among sample units
*/
title color="purple" height=25pt bold italic underlin=1 "chapter15 K-means
cluterung ";
/*Standardize data set*/

title color="green" height=25pt bold italic underlin=1 "Q4: K-means
clustering ";

title color="purple" height=15pt bold italic underlin=1 "chapter15 K-Means";

proc standard data=chemometric out=chemometric0 mean=0 std=1;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
run;
/*Print standardized data set to make sure it matches data table on homework
assignment*/
proc print data=chemometric0;
run;
proc fastclus data=chemometric0 radius=3 maxc=4 replace=full maxiter=10
out=Clus_OUT;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
id molecule;
run;
proc sort data=Clus_OUT;
by cluster distance;
run;
proc print data=Clus_OUT;
var molecule cluster distance;
run;

/* Chapter 12 K-Means on PCA */

```

```

title color="purple" height=15pt bold italic underlin=1 "chapter15 K-Means on
PCA";
proc standard data=RESULTS out=RESULTSO mean=0 std=1;
VAR PRIN1 PRIN2 PRIN3 PRIN4 PRIN5 PRIN6 PRIN7 PRIN8 PRIN9 PRIN10 PRIN11
PRIN12 PRIN13;
run;
/*Print standardized data set to make sure it matches data table on homework
assignment*/
proc print data=RESULTSO;
run;
proc fastclus data=RESULTSO radius=3 maxc=4 replace=full maxiter=10
out=Clus_OUT;
VAR PRIN1 PRIN2 PRIN3 PRIN4 PRIN5 PRIN6 PRIN7 PRIN8 PRIN9 PRIN10 PRIN11
PRIN12 PRIN13;
id molecule;
run;
proc sort data=Clus_OUT;
by cluster distance;
run;
proc print data=Clus_OUT;
var molecule cluster distance;
run;

/*
* Q5: Factor Analysis: to find variable groups: 4 group overall + PCA
*/
/* Chapter 13 factor analysis: find natural groups among variables */
title color="purple" height=18pt bold italic underlin=1 "prin as
example13.3.1";

title color="green" height=25pt bold italic underlin=1 "Q5: Factor Analysis
";

title color="purple" height=15pt bold italic underlin=1 "prin as
example13.3.1";

proc factor data=chemometric method=prin nfactors=2 corr scree ev residuals
plot;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
run;

/* Chapter 13 factor analysis */
title color="purple" height=15pt bold italic underlin=1 "prin as
example13.3.1 by pca";
proc factor data=RESULTS method=prin nfactors=2 corr scree ev residuals plot;
VAR PRIN1 PRIN2 PRIN3 PRIN4 PRIN5 PRIN6 PRIN7 PRIN8 PRIN9 PRIN10 PRIN11
PRIN12 PRIN13;
run;

/*Read in original data set*/
title color="green" height=25pt bold italic underlin=1 "Q5: Factor Analysis
within each sample";

```

```

DATA chemometric_normal;
INFILE 'normal.csv' firstobs=2 dsd;
INPUT molecule group wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200
wv1300 wv1400 wv1500 wv1600;
RUN;

title color="purple" height=25pt bold italic underlin=1 "dataset";
proc print data=chemometric_normal;
run;

/* Chapter 13 factor analysis: find natural groups among variables */
title color="purple" height=18pt bold italic underlin=1 "prin as
example13.3.1 chemometric_normal";
proc factor data=chemometric_normal method=prin nfactors=2 corr scree ev
residuals plot;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
run;

/*Read in original data set*/
DATA chemometric_hp;
INFILE 'hp.csv' firstobs=2 dsd;
INPUT molecule group wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200
wv1300 wv1400 wv1500 wv1600;
RUN;

title color="purple" height=25pt bold italic underlin=1 "dataset";
proc print data=chemometric_hp;
run;

/* Chapter 13 factor analysis: find natural groups among variables */
title color="purple" height=18pt bold italic underlin=1 "prin as
example13.3.1 chemometric_hp";
proc factor data=chemometric_hp method=prin nfactors=2 corr scree ev
residuals plot;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
run;

/*Read in original data set*/
DATA chemometric_karzinom;
INFILE 'karzinom.csv' firstobs=2 dsd;
INPUT molecule group wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200
wv1300 wv1400 wv1500 wv1600;
RUN;

title color="purple" height=25pt bold italic underlin=1 "dataset";
proc print data=chemometric_karzinom;
run;

/* Chapter 13 factor analysis: find natural groups among variables */
title color="purple" height=18pt bold italic underlin=1 "prin as
example13.3.1 chemometric_karzinom";

```

```

proc factor data=chemometric_karzinom method=prin nfactors=2 corr scree ev
residuals plot;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
run;

/*Read in original data set*/
DATA chemometric_adenom;
INFILE 'adenom.csv' firsttobs=2 dsd;
INPUT molecule group wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200
wv1300 wv1400 wv1500 wv1600;
RUN;

title color="purple" height=25pt bold italic underlin=1 "dataset";
proc print data=chemometric_adenom;
run;

/* Chapter 13 factor analysis: find natural groups among variables */
title color="purple" height=18pt bold italic underlin=1 "prin as
example13.3.1 chemometric_adenom";
proc factor data=chemometric_adenom method=prin nfactors=2 corr scree ev
residuals plot;
VAR wv400 wv500 wv600 wv700 wv800 wv900 wv1000 wv1100 wv1200 wv1300 wv1400
wv1500 wv1600;
run;

ods rtf close;

```