

# ANALYSE DISCRIMINANTE

## QUALITE DES VINS DE BORDEAUX

### (Logiciel STATGRAPHICS)

L'analyse discriminante est présentée à l'aide des données du tableau ci-dessous.  
On cherche à expliquer la qualité des vins de Bordeaux par des variables météorologiques.

La variable à expliquer qualité prend 3 modalités. 1= très bon 2= bon 3= moyen

Les variables explicatives de la qualité sont :

**STEMP** : Somme des températures moyennes journalières (°C)

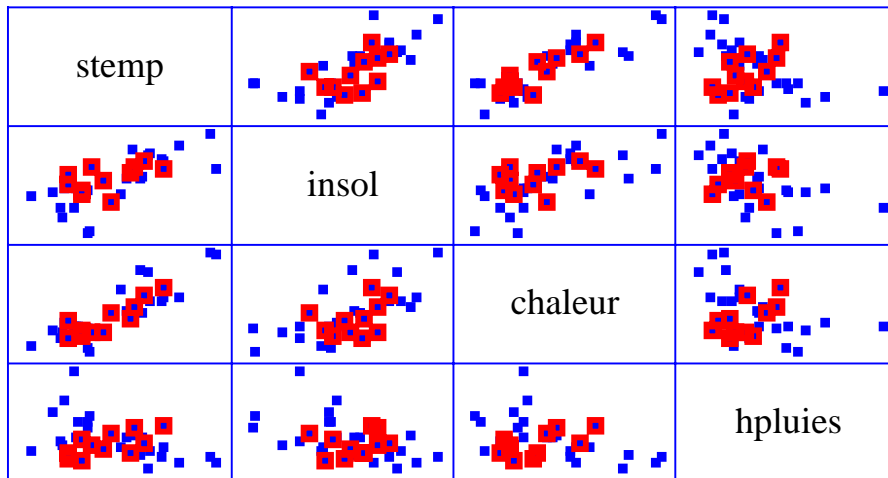
**INSOL** : Durée d'insolation (h)

**CHALEUR** : Nombre de jours de grande chaleur

**HPLUIES** : Hauteur des pluies (mm)

ANNEE	STEMP	INSOL	CHALEUR	HPLUIES	QUALITE
1924	3064	1201	10	361	2
1925	3000	1053	11	338	3
1926	3155	1133	19	393	2
1927	3085	970	4	467	3
1928	3245	1258	36	294	1
1929	3267	1386	35	225	1
1930	3080	966	13	417	3
1931	2974	1189	12	488	3
1932	3038	1103	14	677	3
1933	3318	1310	29	427	2
1934	3317	1362	25	326	1
1935	3182	1171	28	326	3
1936	2998	1102	9	349	3
1937	3221	1424	21	382	1
1938	3019	1230	16	275	2
1939	3022	1285	9	303	2
1940	3094	1329	11	339	2
1941	3009	1210	15	536	3
1942	3227	1331	21	414	2
1943	3308	1366	24	282	1
1944	3212	1289	17	302	2
1945	3361	1444	25	253	1
1946	3061	1175	12	261	2
1947	3478	1317	42	259	1
1948	3126	1248	11	315	2
1949	3458	1508	43	286	1
1950	3252	1361	26	346	2
1951	3052	1186	14	443	3
1952	3270	1399	24	306	1
1953	3198	1259	20	367	1
1954	2904	1164	6	311	3
1955	3247	1277	19	375	1
1956	3083	1195	5	441	3
1957	3043	1208	14	371	3

## Analyse à plusieurs variables



### Statistiques résumées

	stemp	insol	chaleur	hpluies
Effectif	34	34	34	34
Moyenne	3157,88	1247,32	18,8235	360,441
Médiane	3140,5	1253,0	16,5	342,5
Variance	19933,0	16033,4	100,332	8354,25
Ecart-type	141,184	126,623	10,0166	91,4016
Minimum	2904,0	966,0	4,0	225,0
Maximum	3478,0	1508,0	43,0	677,0
Etendue	574,0	542,0	39,0	452,0
1er quartile	3043,0	1175,0	11,0	302,0
3ème quartile	3252,0	1331,0	25,0	414,0
Etendue inter-quart	209,0	156,0	14,0	112,0

### Corrélations

	stemp	insol	chaleur	hpluies
stemp		0,7124 ( 34) 0,0000	0,8651 ( 34) 0,0000	-0,4096 ( 34) 0,0161
insol	0,7124 ( 34) 0,0000		0,6464 ( 34) 0,0000	-0,4734 ( 34) 0,0047
chaleur	0,8651 ( 34) 0,0000	0,6464 ( 34) 0,0000		-0,4011 ( 34) 0,0187
hpluies	-0,4096 ( 34) 0,0161	-0,4734 ( 34) 0,0047	-0,4011 ( 34) 0,0187	

### Le StatAdvisor

Ce tableau donne les corrélations de Pearson entre chaque paire de variables. Ces coefficients de corrélation varient entre -1 et +1 et mesurent la force de la relation linéaire entre les variables. Entre parenthèses sont indiqués les nombres de paires de données utilisées pour calculer ces coefficients. Le troisième nombre dans chaque case du tableau est la valeur de la probabilité qui teste la signification statistique des corrélations estimées. Des valeurs des probabilités au-dessous de 0.05 indiquent des coefficients de corrélation significativement différents de 0 au niveau de confiance de 95%.

# ANOVA à un facteur - stemp par qualite

Analyse de variance

Source	Somme des carrés	Ddl	Carré moyen	F	Proba.
Inter-groupes	420067,0	2	210034,0	27,39	0,0000
Intra-groupes	237722,0	31	7668,46		
Total (Corr.)	657790,0	33			

Nuage de points par niveau

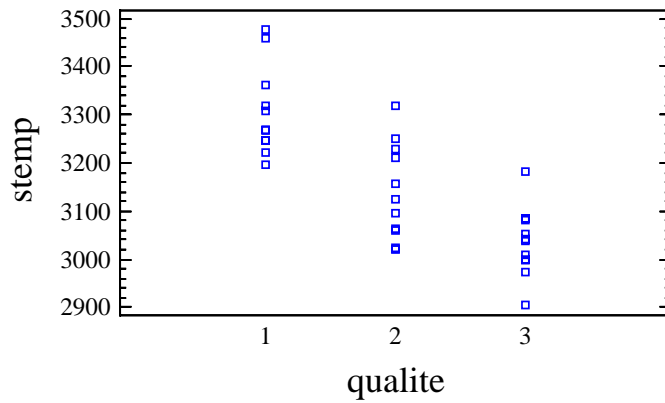
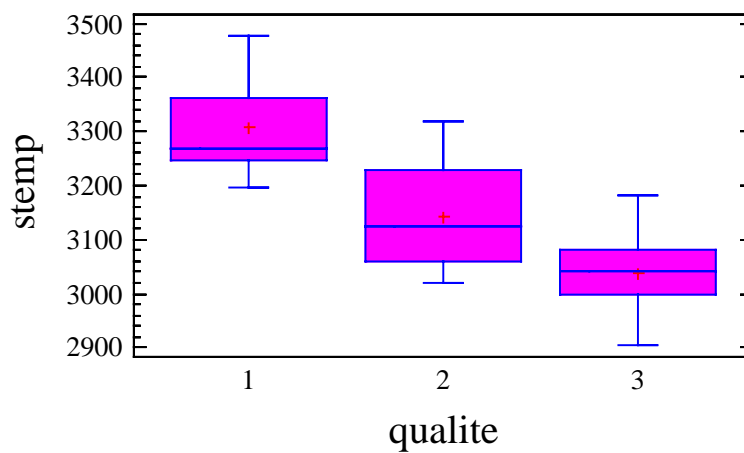


Tableau des moyennes pour stemp par qualite

qualite	Effectif	Moyenne	Erreur type (s agrégé)	Limite inf.	Limite sup.
1	11	3306,36	26,4033		
2	11	3140,91	26,4033		
3	12	3037,33	25,2792		
Total	34	3157,88			

Boîtes à moustaches



# ANOVA à un facteur - insol par qualite

Analyse de variance

Source	Somme des carrés	Ddl	Carré moyen	F	Proba.
Inter-groupes	326909,0	2	163455,0	25,06	0,0000
Intra-groupes	202192,0	31	6522,33		
Total (Corr.)	529101,0	33			

Nuage de points par niveau

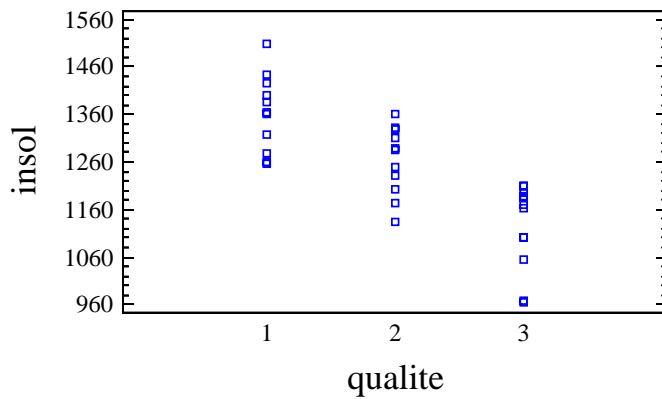
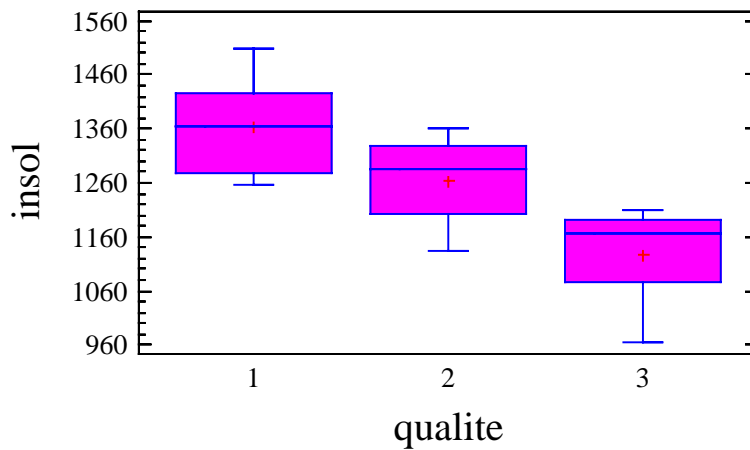


Tableau des moyennes pour insol par qualite

qualite	Effectif	Moyenne	Erreur type (s agrégé)	Limite inf.	Limite sup.
1	11	1363,64	24,3503		
2	11	1262,91	24,3503		
3	12	1126,42	23,3137		
Total	34	1247,32			

Boîtes à moustaches



# ANOVA à un facteur - chaleur par qualite

Analyse de variance

Source	Somme des carrés	Ddl	Carré moyen	F	Proba.
Inter-groupes	1646,57	2	823,285	15,33	0,0000
Intra-groupes	1664,37	31	53,6894		
Total (Corr.)	3310,94	33			

Nuage de points par niveau

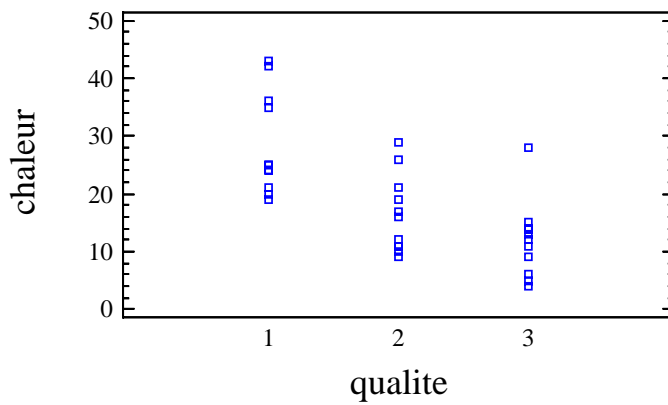
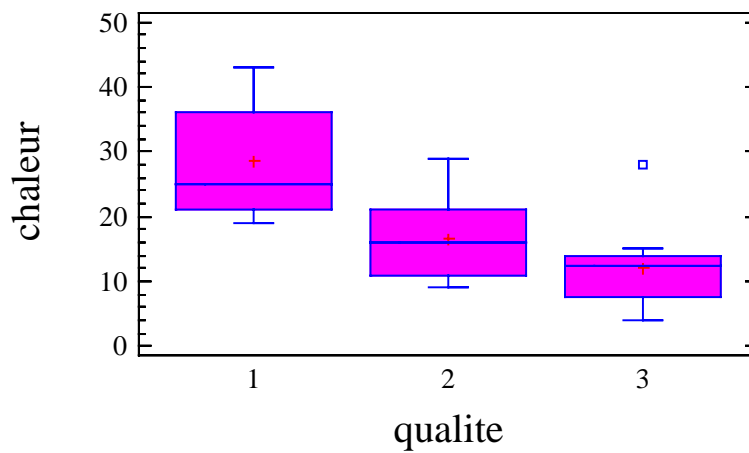


Tableau des moyennes pour chaleur par qualite

qualite	Effectif	Moyenne	Erreur type (s agrégé)	Limite inf.	Limite sup.
1	11	28,5455	2,20927		
2	11	16,4545	2,20927		
3	12	12,0833	2,11521		
Total	34	18,8235			

Boîtes à moustaches



# ANOVA à un facteur - hpluies par qualite

Analyse de variance

Source	Somme des carrés	Ddl	Carré moyen	F	Proba.
Inter-groupes	97191,2	2	48595,6	8,44	0,0012
Intra-groupes	178499,0	31	5758,04		
Total (Corr.)	275690,0	33			

Nuage de points par niveau

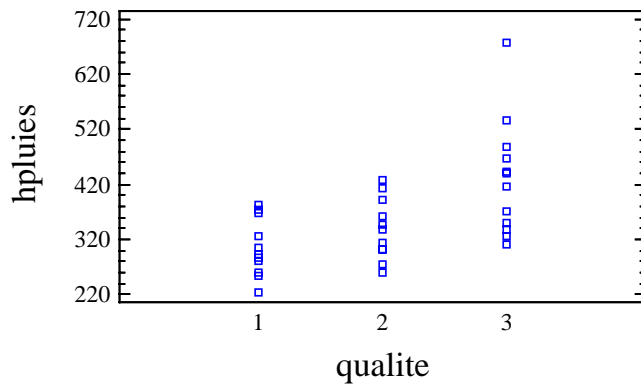
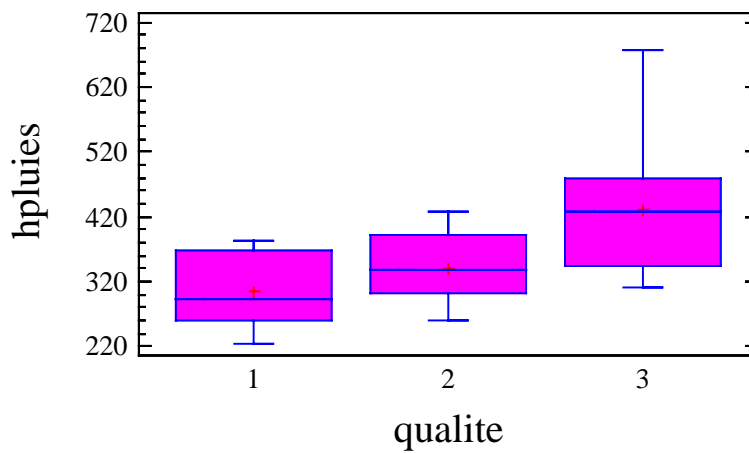


Tableau des moyennes pour hpluies par qualite

qualite	Effectif	Moyenne	Erreur type (s agrégé)	Limite inf.	Limite sup.
1	11	305,0	22,8792		
2	11	339,636	22,8792		
3	12	430,333	21,9052		
Total	34	360,441			

Boîtes à moustaches



# RESULTATS PREALABLES A L'ANALYSE DISCRIMINANTE

## Statistiques résumées par groupe

qualite	1	2	3	TOTAL
COMPTAGES	11	11	12	34
<b>MOYENNES</b>				
stemp	3306,36	3140,91	3037,33	3157,88
insol	1363,64	1262,91	1126,42	1247,32
chaleur	28,5455	16,4545	12,0833	18,8235
hpluies	305,0	339,636	430,333	360,441
<b>ECARTS-TYPES</b>				
stemp	92,0568	100,045	69,3389	141,184
insol	80,306	71,9409	88,3932	126,623
chaleur	8,80186	6,7285	6,30236	10,0166
hpluies	52,2934	54,9859	104,846	91,4016

## Statistiques intra-groupes pour qualite

Matrice des covariances intra-groupes				
	stemp	insol	chaleur	hpluies
stemp	7668,46	1880,15	461,321	430,236
insol	1880,15	6522,33	169,201	-158,001
chaleur	461,321	169,201	53,6894	-34,8231
hpluies	430,236	-158,001	-34,8231	5758,04
Matrice des corrélations intra-groupes				
	stemp	insol	chaleur	hpluies
stemp	1,0	0,26585	0,71896	0,0647463
insol	0,26585	1,0	0,285928	-0,0257823
chaleur	0,71896	0,285928	1,0	-0,0626304
hpluies	0,0647463	-0,0257823	-0,0626304	1,0

Le StatAdvisor

Cette fenêtre affiche les corrélations estimées entre les données des variables indépendantes dans chacun des groupes. L'information intra-groupes pour tous les groupes a été agrégée.

# ANALYSE DISCRIMINANTE

Variable de classement: qualite

Variabiles explicatives: stemp insol chaleur hpluies

Nombre d'observations complètes: 34

Nombre de groupes: 3

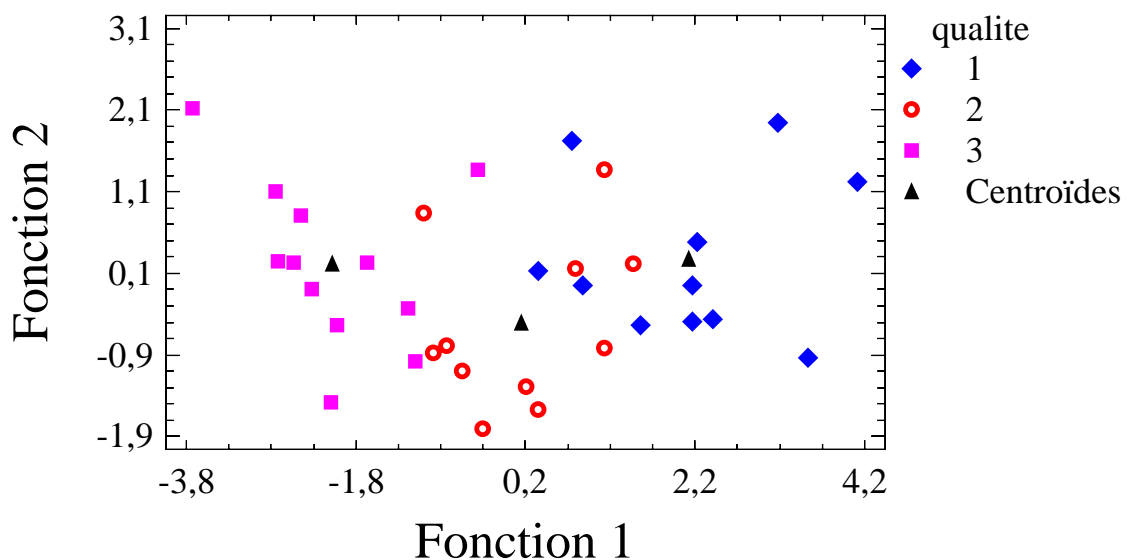
Fonction discriminante	Valeur propre	Pourcentage relatif	Corrélation canonique
1	3,27886	95,95	0,87538
2	0,138574	4,05	0,34887

Fonctions dérivées	Lambda de Wilks	Chi-carré	DDL	Proba.
1	0,205263	46,7122	8	0,0000
2	0,878292	3,8284	3	0,2806

Le StatAdvisor

-----  
 Cette procédure permet de définir un ensemble de fonctions discriminantes qui peuvent aider à prévoir qualite en se basant sur les valeurs d'autres variables quantitatives. 34 observations ont été utilisées pour bâtir le modèle permettant de discriminer entre les 3 niveaux de qualite. 4 variables prédictives ont été entrées. La fonction discriminante qui a une valeur de probabilité inférieure à 0.05 est statistiquement significative au niveau de confiance de 95%. Pour afficher les fonctions discriminantes, sélectionner Fonctions discriminantes dans la liste des options pour les graphiques. Pour prévoir de nouvelles observations, sélectionner Tableau de classement dans la liste des options pour les tableaux.

## Graphique des fonctions discriminantes





## Coefficients des fonctions discriminantes pour qualite

### Coefficients standardisés

	1	2
stemp	0,750126	-0,00405015
insol	0,547064	-0,430399
chaleur	-0,198237	0,935229
hpluies	-0,445097	0,468536

### Coefficients non standardisés

	1	2
stemp	0,00856605	-0,0000462506
insol	0,00677387	-0,00532929
chaleur	-0,0270545	0,127636
hpluies	-0,00586567	0,00617456
CONSTANTE	-32,8763	2,16528

### Le StatAdvisor

Cette fenêtre affiche les coefficients des fonctions utilisées pour discriminer entre les différents niveaux de qualite. Les coefficients standardisés sont particulièrement intéressants. La première fonction discriminante standardisée est :

$$0,750126 * \text{stemp} + 0,547064 * \text{insol} - 0,198237 * \text{chaleur} - 0,445097 * \text{hpluies}$$

A partir des grandeurs des coefficients dans cette équation, vous pouvez déterminer comment les variables indépendantes sont utilisées pour discriminer entre les groupes.

## Centroides des groupes pour qualite

	1	2
1	2,12196	0,271812
2	0,146307	-0,513104
3	-2,07925	0,221184

### Le StatAdvisor

Cette fenêtre affiche les valeurs moyennes de chacune des 2 fonctions discriminantes pour chacune des 3 valeurs de qualite. Les centroides sont inclus dans le graphique des fonctions discriminantes.

## Tableau de classement

Observé qualite	Groupe Taille	Prévu qualite		
		1	2	3
1	11	9 ( 81,82%)	2 ( 18,18%)	0 ( 0,00%)
2	11	2 ( 18,18%)	8 ( 72,73%)	1 ( 9,09%)
3	12	0 ( 0,00%)	2 ( 16,67%)	10 ( 83,33%)

Pourcentage d'observations bien classées: 79,41%

# SELECTION DE VARIABLES

## Régression pas à pas : Méthode: sélection ascendante

F en entrée: 1,0  
F en sortie: 1,0

Etape 0:

-----

0 variable(s) dans le modèle.

Etape 1:

-----

Ajout de la variable stemp avec un F en entrée = 27,3893

1 variable(s) dans le modèle.

Lambda de Wilks = 0,361395 F approximé = 27,3893 avec une probabilité = 0,0000

Etape 2:

-----

Ajout de la variable insol avec un F en entrée = 5,80425

2 variable(s) dans le modèle.

Lambda de Wilks = 0,260568 F approximé = 14,3853 avec une probabilité = 0,0000

Etape 3:

-----

Ajout de la variable hpluies avec un F en entrée = 2,74454

3 variable(s) dans le modèle.

Lambda de Wilks = 0,219098 F approximé = 10,9851 avec une probabilité = 0,0000

Modèle final sélectionné.

Fonction discriminante	Valeur propre	Pourcentage relatif	Corrélation canonique
1	3,22035	97,53	0,87353
2	0,0814678	2,47	0,27446

Fonctions dérivées	Lambda de Wilks	Chi-carré	DDL	Proba.
1	0,219098	45,5471	6	0,0000
2	0,924669	2,3496	2	0,3089

Le StatAdvisor

-----

En utilisant un algorithme de sélection pas à pas, il a été déterminé que 3 variables étaient des prédicteurs significatifs de qualité. La fonction discriminante qui a une valeur de probabilité inférieure à 0.05 est statistiquement significative au niveau de confiance de 95%.

## Régression pas à pas : Méthode: sélection descendante

F en entrée: 1,0  
F en sortie: 1,0

Etape 0:

-----

4 variable(s) dans le modèle.

Lambda de Wilks = 0,205263 F approximé = 8,45051 avec une probabilité = 0,0000

Etape 1:

-----

Suppression de la variable chaleur avec un F en sortie = 0,943617

3 variable(s) dans le modèle.

Lambda de Wilks = 0,219098 F approximé = 10,9851 avec une probabilité = 0,0000

Modèle final sélectionné.

## Modèle à trois variables

### Coefficients des fonctions discriminantes pour qualite

#### Coefficients standardisés

	1	2
stemp	0,619277	0,776937
insol	0,531197	-0,504444
hpluies	-0,426545	0,53192

#### Coefficients non standardisés

	1	2
stemp	0,00707181	0,00887221
insol	0,0065774	-0,00624614
hpluies	-0,00562118	0,00700985
CONSTANTE	-28,51	-22,7531

### Centroïdes des groupes pour qualite

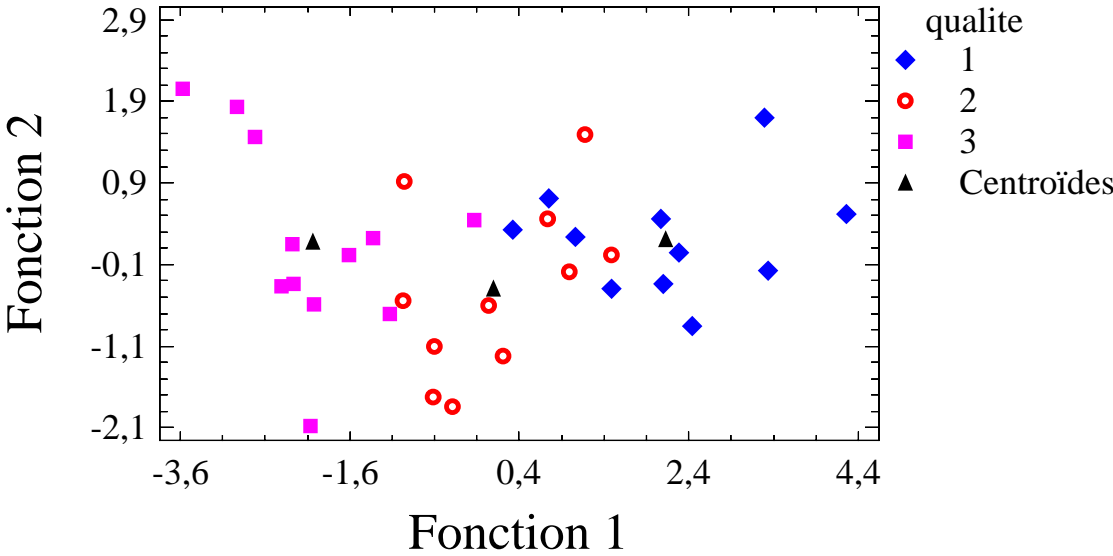
	1	2
1	2,12671	0,202217
2	0,0994283	-0,393779
3	-2,04063	0,175598

### Tableau de classement

Observé qualite	Groupe Taille	Prévu qualite		
		1	2	3
1	11	9 ( 81,82%)	2 ( 18,18%)	0 ( 0,00%)
2	11	2 ( 18,18%)	8 ( 72,73%)	1 ( 9,09%)
3	12	0 ( 0,00%)	1 ( 8,33%)	11 ( 91,67%)

Pourcentage d'observations bien classées: 82,35%

# Graphique des fonctions discriminantes



## PROGRAMME SAS : VINS DE BORDEAUX

```
options ls=75;
DATA;
INPUT ANNEE $ TEMP SOLEIL CHAL PLUIES QUALITE;
CARDS;
1924      3064      1201      10      361      2
1925      3000      1053      11      338      3
1926      3155      1133      19      393      2
1927      3085      970       4      467      3
1928      3245      1258      36      294      1
1929      3267      1386      35      225      1
1930      3080      966       13      417      3
1931      2974      1189      12      488      3
1932      3038      1103      14      677      3
1933      3318      1310      29      427      2
1934      3317      1362      25      326      1
1935      3182      1171      28      326      3
1936      2998      1102      9       349      3
1937      3221      1424      21      382      1
1938      3019      1230      16      275      2
1939      3022      1285      9       303      2
1940      3094      1329      11      339      2
1941      3009      1210      15      536      3
1942      3227      1331      21      414      2
1943      3308      1366      24      282      1
1944      3212      1289      17      302      2
1945      3361      1444      25      253      1
1946      3061      1175      12      261      2
1947      3478      1317      42      259      1
1948      3126      1248      11      315      2
1949      3458      1508      43      286      1
1950      3252      1361      26      346      2
1951      3052      1186      14      443      3
1952      3270      1399      24      306      1
1953      3198      1259      20      367      1
1954      2904      1164      6       311      3
1955      3247      1277      19      375      1
1956      3083      1195      5       441      3
1957      3043      1208      14      371      3
;
proc print;
var ANNEE -- QUALITE;
proc candisc all;
class QUALITE;
var TEMP--PLUIES;
proc discrim list crossvalidate;
class QUALITE;
var TEMP--PLUIES;
proc stepdisc fw ;
var TEMP--PLUIES;
class QUALITE;
proc stepdisc sw ;
var TEMP--PLUIES;
class QUALITE;
proc stepdisc bw;
var TEMP--PLUIES;
class QUALITE;
RUN;
```

# PROCEDURE CANDISC

## Canonical Discriminant Analysis

34 Observations	33 DF Total
4 Variables	31 DF Within Classes
3 Classes	2 DF Between Classes

### Class Level Information

QUALITE	Frequency	Weight	Proportion
1	11	11.0000	0.323529
2	11	11.0000	0.323529
3	12	12.0000	0.352941

## Canonical Discriminant Analysis

### Within-Class Correlation Coefficients / Prob > |R|

#### QUALITE = 1

Variable	TEMP	SOLEIL	CHAL	PLUIES
TEMP	1.00000 0.0	0.45867 0.1559	0.72233 0.0121	-0.56276 0.0715
SOLEIL	0.45867 0.1559	1.00000 0.0	0.23077 0.4948	-0.31816 0.3403
CHAL	0.72233 0.0121	0.23077 0.4948	1.00000 0.0	-0.66503 0.0256
PLUIES	-0.56276 0.0715	-0.31816 0.3403	-0.66503 0.0256	1.00000 0.0

#### QUALITE = 2

Variable	TEMP	SOLEIL	CHAL	PLUIES
TEMP	1.00000 0.0	0.49207 0.1242	0.87342 0.0004	0.68951 0.0189
SOLEIL	0.49207 0.1242	1.00000 0.0	0.38889 0.2372	0.23057 0.4952
CHAL	0.87342 0.0004	0.38889 0.2372	1.00000 0.0	0.61162 0.0455
PLUIES	0.68951 0.0189	0.23057 0.4952	0.61162 0.0455	1.00000 0.0

#### QUALITE = 3

Variable	TEMP	SOLEIL	CHAL	PLUIES
TEMP	1.00000 0.0	-0.16374 0.6111	0.53436 0.0735	0.03123 0.9232
SOLEIL	-0.16374 0.6111	1.00000 0.0	0.28257 0.3735	-0.00645 0.9841
CHAL	0.53436 0.0735	0.28257 0.3735	1.00000 0.0	-0.03871 0.9049
PLUIES	0.03123 0.9232	-0.00645 0.9841	-0.03871 0.9049	1.00000 0.0

**Pooled Within-Class Correlation Coefficients / Prob > |R|**

Variable	TEMP	SOLEIL	CHAL	PLUIES
TEMP	1.00000 0.0	0.26585 0.1414	0.71896 0.0001	0.06475 0.7248
SOLEIL	0.26585 0.1414	1.00000 0.0	0.28593 0.1126	-0.02578 0.8886
CHAL	0.71896 0.0001	0.28593 0.1126	1.00000 0.0	-0.06263 0.7335
PLUIES	0.06475 0.7248	-0.02578 0.8886	-0.06263 0.7335	1.00000 0.0

**Between-Class Correlation Coefficients / Prob > |R|**

Variable	TEMP	SOLEIL	CHAL	PLUIES
TEMP	1.00000 0.0	0.97677 0.1375	0.99132 0.0839	-0.92931 0.2408
SOLEIL	0.97677 0.1375	1.00000 0.0	0.94013 0.2214	-0.98685 0.1033
CHAL	0.99132 0.0839	0.94013 0.2214	1.00000 0.0	-0.87269 0.3247
PLUIES	-0.92931 0.2408	-0.98685 0.1033	-0.87269 0.3247	1.00000 0.0

**Total-Sample Correlation Coefficients / Prob > |R|**

Variable	TEMP	SOLEIL	CHAL	PLUIES
TEMP	1.00000 0.0	0.71235 0.0001	0.86510 0.0001	-0.40962 0.0161
SOLEIL	0.71235 0.0001	1.00000 0.0	0.64645 0.0001	-0.47340 0.0047
CHAL	0.86510 0.0001	0.64645 0.0001	1.00000 0.0	-0.40114 0.0187
PLUIES	-0.40962 0.0161	-0.47340 0.0047	-0.40114 0.0187	1.00000 0.0

**Canonical Discriminant Analysis      Simple Statistics**

**Total-Sample**

Variable	N	Sum	Mean	Variance	Std Dev
TEMP	34	107368	3158	19933	141.18433
SOLEIL	34	42409	1247	16033	126.62297
CHAL	34	640.00000	18.82353	100.33155	10.01656
PLUIES	34	12255	360.44118	8354	91.40161

**QUALITE = 1**

Variable	N	Sum	Mean	Variance	Std Dev
TEMP	11	36370	3306	8474	92.05680
SOLEIL	11	15000	1364	6449	80.30601
CHAL	11	314.00000	28.54545	77.47273	8.80186
PLUIES	11	3355	305.00000	2735	52.29340

**QUALITE = 2**

Variable	N	Sum	Mean	Variance	Std Dev
TEMP	11	34550	3141	10009	100.04544
SOLEIL	11	13892	1263	5175	71.94088
CHAL	11	181.00000	16.45455	45.27273	6.72850
PLUIES	11	3736	339.63636	3023	54.98595

**QUALITE = 3**

Variable	N	Sum	Mean	Variance	Std Dev
TEMP	12	36448	3037	4808	69.33887
SOLEIL	12	13517	1126	7813	88.39319
CHAL	12	145.00000	12.08333	39.71970	6.30236
PLUIES	12	5164	430.33333	10993	104.84563

**Total-Sample Standardized Class Means**

Variable	1	2	3
TEMP	1.051683850	-0.120220577	-0.853841333
SOLEIL	0.918576089	0.123086366	-0.954857250
CHAL	0.970584854	-0.236506650	-0.672905020
PLUIES	-0.606566749	-0.227619767	0.764670974

**Pooled Within-Class Standardized Class Means**

Variable	1	2	3
TEMP	1.695577954	-0.193825701	-1.376606232
SOLEIL	1.440210930	0.192983828	-1.497095195
CHAL	1.326807817	-0.323309056	-0.919873865
PLUIES	-0.730626016	-0.274174152	0.921066821



**Pairwise Squared Distances Between Groups**

$$D(i|j) = (\bar{X}_i - \bar{X}_j)' \text{COV}^{-1} (\bar{X}_i - \bar{X}_j)$$

**Squared Distance to QUALITE**

From QUALITE	1	2	3
F Statistics,	Num DF= 2	Den DF= 31	
Total	Pooled	Between	RSQ/
1	0	4.51931	17.65273
2	4.51931	0	5.49227
3	17.65273	5.49227	0

**F Statistics, NDF=4, DDF=28 for Squared Distance to QUALITE**

From QUALITE	1	2	3
1	0	5.61269	22.87675
2	5.61269	0	7.11761
3	22.87675	7.11761	0

**Prob > Mahalanobis Distance for Squared Distance to QUALITE**

From QUALITE	1	2	3
1	1.0000	0.0019	0.0001
2	0.0019	1.0000	0.0004
3	0.0001	0.0004	1.0000

**Canonical Discriminant Analysis**

Variable	Univariate Test Statistics			R-Squared	(1-RSQ)
	STD	STD	STD		
TEMP	141.1843	87.5697	136.1337	0.638605	1.7671
SOLEIL	126.6230	80.7610	120.0935	0.617857	1.6168
CHAL	10.0166	7.3273	8.5231	0.497312	0.9893
PLUIES	91.4016	75.8817	65.4816	0.352537	0.5445

**Univariate Test Statistics**

Variable	F	Pr > F
TEMP	27.3893	0.0001
SOLEIL	25.0607	0.0001
CHAL	15.3342	0.0001
PLUIES	8.4396	0.0012

Average R-Squared: Unweighted = 0.5265777  
 Weighted by Variance = 0.5769962

**Multivariate Statistics and F Approximations**

Statistic	S=2 M=0.5 N=13			Num DF	Den DF	Pr > F
	Value	F				
Wilks' Lambda	0.20526297	8.4505		8	56	0.0001
Pillai's Trace	0.88800132	5.7896		8	58	0.0001
Hotelling-Lawley Trace	3.41743451	11.5338		8	54	0.0001
Roy's Greatest Root	3.27886049	23.7717		4	29	0.0001

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

**Canonical Discriminant Analysis**

	Canonical Correlation	Adjusted Canonical Correlation	Approx Standard Error	Squared Canonical Correlation
1	0.875382	0.861944	0.040683	0.766293
2	0.348867	0.280587	0.152891	0.121708

**Eigenvalues of INV(E)\*H  
= CanRsq/(1-CanRsq)**

	Eigenvalue	Difference	Proportion	Cumulative
1	3.2789	3.1403	0.9595	0.9595
2	0.1386	.	0.0405	1.0000

**Test of H0: The canonical correlations in the current row and all that follow are zero**

	Likelihood Ratio	Approx F	Num DF	Den DF	Pr > F
1	0.20526297	8.4505	8	56	0.0001
2	0.87829160	1.3395	3	29	0.2808

**Total Canonical Structure**

	CAN1	CAN2
TEMP	0.900589	0.374779
SOLEIL	0.896744	-0.116190
CHAL	0.770513	0.590030
PLUIES	-0.662815	0.361294

**Between Canonical Structure**

	CAN1	CAN2
TEMP	0.986524	0.163614
SOLEIL	0.998669	-0.051569
CHAL	0.956452	0.291891
PLUIES	-0.977208	0.212284

**Canonical Discriminant Analysis**

**Pooled Within Canonical Structure**

	CAN1	CAN2
TEMP	0.724221	0.584256
SOLEIL	0.701280	-0.176148
CHAL	0.525372	0.779910
PLUIES	-0.398218	0.420797

**Total-Sample Standardized Canonical Coefficients**

	CAN1	CAN2
TEMP	1.209391427	-0.006529859
SOLEIL	0.857727422	-0.674810955
CHAL	-0.270993045	1.278475787
PLUIES	-0.536131215	0.564364371

**Pooled Within-Class Standardized Canonical Coefficients**

	CAN1	CAN2
TEMP	0.7501261906	-.0040501510
SOLEIL	0.5470642419	-.4303989053
CHAL	-.1982365052	0.9352290652
PLUIES	-.4450968911	0.4685360966

**Raw Canonical Coefficients**

	CAN1	CAN2
TEMP	0.0085660455	-.0000462506
SOLEIL	0.0067738690	-.0053292933
CHAL	-.0270544919	0.1276361644
PLUIES	-.0058656650	0.0061745562

**Class Means on Canonical Variables**

QUALITE	CAN1	CAN2
1	2.121962956	0.271812018
2	0.146306536	-0.513103518
3	-2.079247035	0.221183875

**PROCEDURE DISCRIM**

Discriminant Analysis

34 Observations	33 DF Total
4 Variables	31 DF Within Classes
3 Classes	2 DF Between Classes

Class Level Information

QUALITE	Frequency	Weight	Proportion	Prior Probability
1	11	11.0000	0.323529	0.333333
2	11	11.0000	0.323529	0.333333
3	12	12.0000	0.352941	0.333333

Discriminant Analysis Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
4	29.5185536

Discriminant Analysis

Pairwise Generalized Squared Distances Between Groups

$$D(i|j) = (\bar{X}_i - \bar{X}_j)' \text{COV}^{-1} (\bar{X}_i - \bar{X}_j)$$

Generalized Squared Distance to QUALITE

From QUALITE	1	2	3
1	0	4.51931	17.65273
2	4.51931	0	5.49227
3	17.65273	5.49227	0

Discriminant Analysis      Linear Discriminant Function

$$\text{Constant} = -.5 \sum_j \bar{X}'_j \text{COV}^{-1} \bar{X}_j \quad \text{Coefficient Vector} = \text{COV}^{-1} \bar{X}_j$$

	QUALITE		
	1	2	3
CONSTANT	-1350	-1284	-1212
TEMP	0.81768	0.80079	0.78169
SOLEIL	0.15409	0.14489	0.12590
CHAL	-7.00975	-7.05649	-6.90255
PLUIES	-0.04629	-0.03955	-0.02196

Discriminant Analysis

Classification Results for Calibration Data: WORK.DAT1

Resubstitution Results using Linear Discriminant Function

Generalized Squared Distance Function:

$$D_j(X) = (\bar{X} - \bar{X}_j)' \text{COV}^{-1} (\bar{X} - \bar{X}_j)$$

Posterior Probability of Membership in each QUALITE:

$$\text{Pr}(j|X) = \frac{\exp(-.5 D_j(X))}{\sum_k \exp(-.5 D_k(X))}$$

Obs	Posterior Probability of Membership in QUALITE:				
	From QUALITE	Classified into QUALITE	1	2	3
1	2	2	0.0069	0.6679	0.3252
2	3	3	0.0000	0.0447	0.9553
3	2	3 *	0.0098	0.3108	0.6794
4	3	3	0.0000	0.0147	0.9853
5	1	1	0.6434	0.3279	0.0287
6	1	1	0.9334	0.0665	0.0001
7	3	3	0.0000	0.0075	0.9925
8	3	3	0.0000	0.0226	0.9774
9	3	3	0.0000	0.0004	0.9996
10	2	1 *	0.7564	0.2368	0.0068
11	1	1	0.8924	0.1074	0.0001
12	3	2 *	0.0866	0.5131	0.4003
13	3	3	0.0002	0.1134	0.8864
14	1	1	0.6222	0.3768	0.0010
15	2	2	0.0108	0.7231	0.2661
16	2	2	0.0143	0.9228	0.0629
17	2	2	0.0622	0.9202	0.0176
18	3	3	0.0000	0.0179	0.9821
19	2	2	0.3791	0.6056	0.0153
20	1	1	0.9048	0.0952	0.0000
21	2	2	0.3669	0.6295	0.0036
22	1	1	0.9838	0.0162	0.0000
23	2	2	0.0137	0.8233	0.1630
24	1	1	0.9966	0.0034	0.0000
25	2	2	0.0590	0.9120	0.0289
26	1	1	0.9990	0.0010	0.0000
27	2	1 *	0.7143	0.2841	0.0016
28	3	3	0.0007	0.1355	0.8638
29	1	1	0.8521	0.1478	0.0001
30	1	2 *	0.1957	0.7553	0.0490
31	3	3	0.0001	0.1760	0.8239
32	1	2 *	0.3833	0.6059	0.0108
33	3	2 *	0.0036	0.5799	0.4165
34	3	3	0.0036	0.4143	0.5821

\* Misclassified observation

**Discriminant Analysis**

Classification Summary for Calibration Data: WORK.DAT1

Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function:

$$D_j(X) = (X - \bar{X}_j)' \text{COV}_j^{-1} (X - \bar{X}_j)$$

Posterior Probability of Membership in each QUALITE:

$$\text{Pr}(j|X) = \frac{\exp(-.5 D_j(X))}{\sum_k \exp(-.5 D_k(X))}$$

Number of Observations and Percent Classified into QUALITE:

From QUALITE	1	2	3	Total
1	9 81.82	2 18.18	0 0.00	11 100.00
2	2 18.18	8 72.73	1 9.09	11 100.00
3	0 0.00	2 16.67	10 83.33	12 100.00
Total	11	12	11	34
Percent	32.35	35.29	32.35	100.00
Priors	0.3333	0.3333	0.3333	

Error Count Estimates for QUALITE:

	1	2	3	Total
Rate	0.1818	0.2727	0.1667	0.2071
Priors	0.3333	0.3333	0.3333	

**Discriminant Analysis**

Classification Summary for Calibration Data: WORK.DAT1

Cross-validation Summary using Linear Discriminant Function

Generalized Squared Distance Function:

$$D_j(X) = (X - \bar{X}_j)' \text{COV}^{-1}(X - \bar{X}_j)$$

Posterior Probability of Membership in each QUALITE:

$$\text{Pr}(j|X) = \frac{\exp(-.5 D_j(X))}{\sum_k \exp(-.5 D_k(X))}$$

Number of Observations and Percent Classified into QUALITE:

From QUALITE	1	2	3	Total
1	7 63.64	4 36.36	0 0.00	11 100.00
2	2 18.18	8 72.73	1 9.09	11 100.00
3	0 0.00	2 16.67	10 83.33	12 100.00
Total	9	14	11	34
Percent	26.47	41.18	32.35	100.00
Priors	0.3333	0.3333	0.3333	

Error Count Estimates for QUALITE:

	1	2	3	Total
Rate	0.3636	0.2727	0.1667	0.2677
Priors	0.3333	0.3333	0.3333	

# Stepwise Discriminant Analysis

34 Observations            4 Variable(s) in the Analysis  
 3 Class Levels            0 Variable(s) will be included

The Method for Selecting Variables will be: FORWARD

Significance Level to Enter = 0.1500

## Class Level Information

QUALITE	Frequency	Weight	Proportion
1	11	11.0000	0.323529
2	11	11.0000	0.323529
3	12	12.0000	0.352941

## Stepwise Discriminant Analysis

### Forward Selection: Step 1

Statistics for Entry, DF = 2, 31

Variable	R**2	F	Prob > F	Tolerance
TEMP	0.6386	27.389	0.0001	1.0000
SOLEIL	0.6179	25.061	0.0001	1.0000
CHAL	0.4973	15.334	0.0001	1.0000
PLUIES	0.3525	8.440	0.0012	1.0000

Variable TEMP will be entered

The following variable(s) have been entered:  
 TEMP

## Multivariate Statistics

Wilks' Lambda = 0.36139542    F( 2, 31) = 27.389    Prob > F = 0.0001  
 Pillai's Trace = 0.638605    F( 2, 31) = 27.389    Prob > F = 0.0001

Average Squared Canonical Correlation = 0.31930229

### Forward Selection: Step 2

Statistics for Entry, DF = 2, 30

Variable	Partial R**2	F	Prob > F	Tolerance
SOLEIL	0.2790	5.804	0.0074	0.4926
CHAL	0.0348	0.541	0.5876	0.2516
PLUIES	0.2253	4.361	0.0217	0.8322

Variable SOLEIL will be entered

The following variable(s) have been entered:  
 TEMP            SOLEIL

## Multivariate Statistics

Wilks' Lambda = 0.26056848    F( 4, 60) = 14.385    Prob > F = 0.0001  
 Pillai's Trace = 0.776210    F( 4, 62) = 9.831    Prob > F = 0.0001

Average Squared Canonical Correlation = 0.38810478

**Forward Selection: Step 3**

Statistics for Entry, DF = 2, 29

Variable	Partial R**2	F	Prob > F	Tolerance
CHAL	0.0471	0.717	0.4966	0.2113
PLUIES	0.1592	2.745	0.0810	0.4529

Variable PLUIES will be entered

The following variable(s) have been entered:

TEMP SOLEIL PLUIES

Multivariate Statistics

Wilks' Lambda = 0.21909794 F( 6, 58) = 10.985 Prob > F = 0.0001  
 Pillai's Trace = 0.838383 F( 6, 60) = 7.217 Prob > F = 0.0001

Average Squared Canonical Correlation = 0.41919171

**Forward Selection: Step 4**

Statistics for Entry, DF = 2, 28

Variable	Partial R**2	F	Prob > F	Tolerance
CHAL	0.0631	0.944	0.4012	0.2113

No variables can be entered

No further steps are possible

Forward Selection: Summary

Step	Variable Entered	Number In	Partial R**2	F Statistic	Prob > F	Wilks' Lambda	Prob < Lambda
1	TEMP	1	0.6386	27.389	0.0001	0.36139542	0.0001
2	SOLEIL	2	0.2790	5.804	0.0074	0.26056848	0.0001
3	PLUIES	3	0.1592	2.745	0.0810	0.21909794	0.0001

Forward Selection: Summary

Step	Variable Entered	Number In	Average Squared Canonical Correlation	Prob > ASCC
1	TEMP	1	0.31930229	0.0001
2	SOLEIL	2	0.38810478	0.0001
3	PLUIES	3	0.41919171	0.0001



**Stepwise Discriminant Analysis**

34 Observations            4 Variable(s) in the Analysis  
3 Class Levels            0 Variable(s) will be included

The Method for Selecting Variables will be: **BACKWARD**

Significance Level to Stay = 0.1500

Class Level Information

QUALITE	Frequency	Weight	Proportion
1	11	11.0000	0.323529
2	11	11.0000	0.323529
3	12	12.0000	0.352941

**Backward Elimination: Step 0**

All variables have been entered

Multivariate Statistics

Wilks' Lambda = 0.20526297    F( 8, 56) = 8.451    Prob > F = 0.0001  
Pillai's Trace = 0.888001    F( 8, 58) = 5.790    Prob > F = 0.0001

Average Squared Canonical Correlation = 0.44400066

The SAS System

35

22:59 Thursday, June 19, 1997

**Stepwise Discriminant Analysis**

**Backward Elimination: Step 1**

Statistics for Removal, DF = 2, 28

Variable	Partial R**2	F	Prob > F
TEMP	0.2013	3.529	0.0430
SOLEIL	0.2293	4.165	0.0261
CHAL	0.0631	0.944	0.4012
PLUIES	0.1733	2.934	0.0697

Variable CHAL will be removed

The following variable(s) have been removed:

CHAL

Multivariate Statistics

Wilks' Lambda = 0.21909794    F( 6, 58) = 10.985    Prob > F = 0.0001  
Pillai's Trace = 0.838383    F( 6, 60) = 7.217    Prob > F = 0.0001

Average Squared Canonical Correlation = 0.41919171

**Backward Elimination: Step 2**

Statistics for Removal, DF = 2, 29

Variable	Partial R**2	F	Prob > F
TEMP	0.3125	6.590	0.0044
SOLEIL	0.2175	4.030	0.0286
PLUIES	0.1592	2.745	0.0810

No variables can be removed

No further steps are possible

**Stepwise Discriminant Analysis**

Backward Elimination: Summary

Step	Variable Removed	Number In	Partial R**2	F Statistic	Prob > F	Wilks' Lambda	Prob < Lambda
0		4	.	.	.	0.20526297	0.0001
1	CHAL	3	0.0631	0.944	0.4012	0.21909794	0.0001

Step	Variable Removed	Number In	Average Squared Canonical Correlation	Prob > ASCC
0		4	0.44400066	0.0001
1	CHAL	3	0.41919171	0.0001