

Concours Ingénieur d'études- IFSTTAR

29 septembre 2015

Durée : 3 heures

Les candidats devront répondre aux trois exercices suivants. Pour le 1^{er} exercice, les réponses sont à indiquer dans le tableau proposé.

1. Examen pratique sur SAS.

Cette partie est destinée à évaluer vos connaissances pratiques du logiciel SAS. Il s'agit d'une série de questions pour lesquelles il faut choisir une unique bonne réponse parmi les réponses proposées.

Les réponses sont à indiquer dans le tableau proposé.

1.1 SAS Fundamental Concepts

1. Which of the following LIBNAME statements has the correct syntax?
 - a. **libname monthly 'c:\monthly';**
 - b. **libname 'c:\monthly' monthly;**
 - c. **libname monthly='c:\monthly';**
 - d. **libname 'c:\monthly'=monthly ;**
2. Which of the following statements is true?
 - a. A libref can be 32 or less characters.
 - b. A variable name can be 32 or less characters.
 - c. Numeric variables are stored as 32 bytes by default.
 - d. Character variables are stored as 32 bytes by default.
3. A program was submitted and the SAS log is shown below.

```
169 data work.sales;
170   set sashelp.orsales;
171   drop quarter year;
172 run;
NOTE: There were 912 observations read from the data set SASHELP.ORSALES.
NOTE: The data set WORK.SALES has 912 observations and 6 variables.
173 /*
174 proc contents data=work.sales;
175 run;
176 */
177 proc print data=work.sales;
178 run;
```

NOTE: There were 912 observations read from the data set WORK.SALES.

Which of the following is true regarding the CONTENTS procedure?

- a. The PROC CONTENTS step failed execution.
 - b. The PROC CONTENTS step did not execute.
 - c. The second NOTE applies to the PROC CONTENTS step.
 - d. The last NOTE applies to the PROC CONTENTS step and the PROC PRINT step.
4. The following program is submitted:

```
proc contents data=temp.sales;
run;
```

Which is the result?

- a. A report showing the data portion of the temporary data set **temp.sales**
 - b. A report showing the data portion of the permanent data set **temp.sales**
 - c. A report showing the descriptor portion of the temporary data set **temp.sales**
 - d. A report showing the descriptor portion of the permanent data set **temp.sales**
5. Which of the following statements is true concerning SAS date values?
- a. The SAS date value for 05JAN1960 is 4.
 - b. The SAS date value for 03/24/1952 is a positive number.
 - c. A SAS date value represents the number of days between January 1, 1950, and a specified date.
 - d. The SAS date value for September 16, 1999, can be written as the SAS date constant '09/16/1999'd.

1.2 Working with SAS Data Sets

1. The following program is submitted:

```
data work.newsales;
  set work.sales;
  sales=price*quantity;
  <insert statement here>
run;
```

Which SAS statement will output observations with **product** equal to the character value Shorts and **sales** less than one million?

- a. **if product eq Shorts and sales<1000000;**
 - b. **if product='Shorts' and sales lt 1000000;**
 - c. **where product=Shorts and sales lt 1000000;**
 - d. **where product eq 'Shorts' and sales<1000000;**
2. Which of the following is true regarding the SORT procedure?
 - a. The SORT procedure requires the BY statement.
 - b. The SORT procedure has the ability to create a report or a new data set.
 - c. The SORT procedure can only sort values based on character variables.
 - d. The SORT procedure sorts values by descending order unless the ASCENDING option is specified.
 3. Which items are potentially created at compile time of a DATA step?
 - a. input buffer, data values, and report
 - b. raw data file, program data vector, and report
 - c. raw data file, data values, and descriptor information
 - d. input buffer, program data vector, and descriptor information
 4. Given the input SAS data set **salary1**:

IDNUM	SALARY
12649	52000
49255	75000

Given the input SAS data set **salary2**:

ID	SALARY
56391	89000
88376	66000

The following program is submitted:

```
data salaryall;
  <insert statement here>
run;
```

Given the desired output SAS data set **salaryall**:

ID	SALARY
12649	52000

49255	75000
56391	89000
88376	66000

Which statement will produce the desired output data set?

- a. `set salary2 salary1(rename id=idnum);`
- b. `set salary1 rename=(idnum=id) salary2;`
- c. `set salary1(rename=(idnum=id)) salary2;`
- d. `set salary1 salary2(rename=(id=idnum));`

5. The following program is submitted:

```
data work.firsthalf work.thirdqtr work.misc;
  set sashelp.retail;
  if 1<=month<=6 then output work.firsthalf;
  else if 7<=month<=9 then output work.thirdqtr;
run;
```

Which of the following statements is true regarding the previous program with an observation having **month** equal to 12?

- a. The observation will be output to the **work.firsthalf** data set.
- b. The observation will be output to the **work.thirdqtr** data set.
- c. The observation will be output to the **work.misc** data set.
- d. The observation will not be output to any data set.

6. The following SAS program is submitted:

```
proc sort data=sashelp.class new=sortdata;
  by name descending age;
run;
```

What is the result?

- a. The program fails execution due to a syntax error with the NEW= option.
- b. The program fails execution due to a syntax error with the DESCENDING option.
- c. The program runs without errors and creates a new data set **work.sortdata** with the sorted observations.
- d. The program runs with warnings and overwrites the original data set **sashelp.class** with the sorted observations.

7. Given the input data set **products**:

CODE	PRODUCT
A123	Sandal
A234	Slipper
B345	Boot
B456	Sneaker

Given the input data set **costs**:

CODE	COST
A123	19.99
A234	9.99
B456	25.99

The following program is submitted:

```
data prodcost;
  merge products(in=p) costs(in=c);
  by code;
  if p and c;
run;
```

Which is the result?

- a. The program fails execution because of invalid IN= syntax.
- b. The program fails execution because the subsetting IF statement is incomplete.
- c. The program runs without errors or warnings and produces a data set with three observations and three variables.
- d. The program runs without errors or warnings and produces a data set with four observations and three variables.

8. The following program is submitted:

```
data work.orsales;
  set sashelp.orsales(firsttobs=500 obs=700);
run;
```

How many observations are in the output data set **work.orsales**?

- a. 200
- b. 201
- c. 1199
- d. 1200

9. The following program is submitted:

```
data work.sales;
  set sashelp.orsales;
  drop quarter year;
run;
```

Which of the following statements is true regarding the previous program?

- a. The variables **quarter** and **year** will not be in **work.sales**.
- b. The output data set will contain the variables **_N_** and **_ERROR_**.
- c. The data set **work.sales** is the input data set and **sashelp.orsales** is the output data set.
- d. The output data set will have no observations because there is no OUTPUT statement.

1.3 Working with Raw Data and Microsoft Excel Files

1. Given the raw data file **address.dat**:

```
Sue      Smith
123 Main Street
San Diego CA 92625
Julie    Johnson
456 Monroe Road
Birmingham AL 35235
```

The following SAS program is submitted:

```
data work.address;
  infile 'address.dat';
  <insert statement(s) here>
run;
```

Given the desired output data set **work.address**:

FIRST	LAST	STREET
Sue	Smith	123 Main Street
Julie	Johnson	456 Monroe Road

Which statement(s) will produce the desired output data set?

- a. **input first \$ 1-9 last \$ 10-20;**
input street \$ 1-15;
- b. **input first \$ 1-9 last \$ 10-20 / street \$ 1-15 /;**
- c. **input first \$ 1-9 last \$ 10-20 @ street \$ 1-15 @@;**
- d. **input #1 first \$ 1-9 last \$ 10-20 #2 street \$ 1-15;**

2. Given the raw data file **2005pop.dat**:

```
----|----10---|----20---|----30
8,143,197 1 New York     NY
3,844,829 2 Los Angeles  CA
2,842,518 3 Chicago      IL
```

The following SAS program is submitted:

```
data work.population;
  infile '2005pop.dat';
  input  @2 POPULATION comma9.
         @12 RANK 1.
         @14 CITY $12.
         @27 STATE $2;
run;
```

Which is the output data set **work.population**?

- a.

POPULATION	RANK	CITY	STATE
8,143,197	1	New York	NY
3,844,829	2	Los Angeles	CA
2,842,518	3	Chicago	IL

- b.

POPULATION	RANK	CITY	STATE
8,143,197		New York	NY

3,844,829	Los Angeles	CA
2,842,518	Chicago	IL

c.

POPULATION	RANK	CITY	STATE
8143197	1	New York	8
3844829	2	Los Angeles	3
2842518	3	Chicago	2

d.

POPULATION	RANK	CITY	STATE
8143197	1	New York	NY
3844829	2	Los Angeles	CA
2842518	3	Chicago	IL

3. The Excel workbook **customers.xls** contains a worksheet named **Females** and a worksheet named **Males**.

Which program will read the **Males** worksheet to create a SAS data set?

a.

```
libname customer 'customers.xls';

data work.males;
  set customer.males;
run;

libname customer clear;
```

b.

```
libname customer 'customers.xls';

data work.males;
  set customer.males.worksheet;
run;

libname customer clear;
```

c.

```
libname customer 'customers.xls';

data work.males;
  set customer.males$;
run;

libname customer clear;
```

d.

```
libname customer 'customers.xls';

data work.males;
  set customer.'males'$'n';
run;

libname customer clear;
```

4. Which of the following is true regarding the DSD option when reading raw data files?
 - a. The DSD option sets the delimiter to a blank.
 - b. The DSD option treats two consecutive delimiters as a missing value.
 - c. The DSD option belongs in the INPUT statement after a forward slash.
 - d. The DSD option removes any delimiters located inside a set of quotation marks.
5. Given the raw data file **info.dat**:

John Louisville KY Lorna Columbia MO 65203

The following SAS program is submitted:

```
data info;  
    infile 'info.dat';  
    input name $ city $ state $ zipcode $;  
run;
```

Which is the result?

- a. The program runs without errors or warnings and produces a data set with two observations and four variables.
- b. The program produces a warning of invalid data for **zipcode** in line 1 and produces a data set with two observations and four variables.
- c. The program produces a note that SAS went to a new line when the INPUT statement reached past the end of a line and produces a data set with one observation and four variables.
- d. The program produces an error that SAS went to a new line when the INPUT statement reached past the end of a line and produces an error that SAS stopped processing the step because of errors.

6. Given the raw data file **revenue.dat**:

Jan \$13,000 above
Feb \$900 below
Mar \$27,000 above

The following SAS program is submitted:

<pre>data work.target; infile 'revenue.dat'; <insert statement here> run;</pre>

Given the desired output data set **work.target**:

MONTH	REVENUE	TARGET
Jan	13000	above
Feb	900	below
Mar	27000	above

Which statement will produce the desired output data set?

- a. **input MONTH \$ REVENUE TARGET \$;**
- b. **input MONTH \$ REVENUE \$ TARGET \$;**
- c. **input MONTH \$ REVENUE:dollar8. TARGET \$;**
- d. **input MONTH \$ REVENUE dollar7. TARGET \$;**

7. A DATA step was submitted and a portion of the SAS log is shown below.

<pre>601 data work.population; 602 infile '2000pop.dat'; 603 input population 1-7 604 rank 9 605 state 24-25 606 city \$ 11-21; 607 run; NOTE: Invalid data for state in line 1 24-25. RULE: -----1-----2-----3-----4-----5----- 1 8008278 1 New York NZ 25 population=8008278 rank=1 state=. city=New York _ERROR_=1 _N_=1 NOTE: Invalid data for state in line 2 24-25. 2 3694820 2 Los Angeles CA 25 population=3694820 rank=2 state=. city=Los Angeles _ERROR_=1 _N_=2 NOTE: Invalid data for state in line 3 24-25. 3 2896016 3 Chicago IL 25 population=2896016 rank=3 state=. city=Chicago _ERROR_=1 _N_=3</pre>
--

What is the cause of the notes about invalid data?

- a. NZ is not a valid **state** value.
- b. The variable **state** is not numeric data.
- c. The variable **state** is not located before **CITY**.
- d. The variable **state** is not located in positions 24 and 25.

8. Which LIBNAME statement will access the Excel workbook **products.xls**, which contains the worksheet **Children**?

- a. libname myexcel 'products.xls';
- b. libname myexcel children 'products.xls';
- c. libname myexcel 'products.xls' sheet='children';
- d. libname myexcel workbook='products.xls' worksheet='children';

9. Given the raw data file **1990pop.dat**:

New York!7,322,564!Los Angeles!3,485,398
Chicago!2,783,726!Houston!1,630,553

The following SAS program is submitted:

data work.population;
length CITY \$ 12;
<insert statements here>
run;

Given the desired output data set **work.population**:

CITY	POPULATION
New York	7322564
Los Angeles	3485398
Chicago	2783726
Houston	1630553

Which statements will produce the desired output data set?

- a. infile '1990pop.dat';

input CITY POPULATION:comma9. dlm='!' @;
- b. infile '1990pop.dat';

input CITY \$ POPULATION:comma9. dlm='!' @@;
- c. infile '1990pop.dat' dlm='!';

input CITY \$ POPULATION:comma9. @;
- d. infile '1990pop.dat' dlm='!';

input CITY POPULATION:comma9. @@;

1.4 Creating Variables

1. The following SAS program is submitted:

```
data work.class;
  set sashelp.class(keep=name age);
  if age>=13 then group='Teen';
  if 11<=age<=13 then group='Pre-Teen';
run;
```

What is the value of **group** in the data set **work.class** if an observation has a value of **age** equal to 13?

- a. missing
 - b. Teen
 - c. Pre-
 - d. Pre-Teen
2. A DATA step was submitted and a portion of the SAS log is shown below.

```
263 data new;
264   newvar = THIS IS A TEST;
      --
      388
      76
ERROR 388-185: Expecting an arithmetic operator.

ERROR 76-322: Syntax error, statement will be ignored.

265 run;

NOTE: The SAS System stopped processing this step because of
      errors.
WARNING: The data set WORK.NEW may be incomplete. When this
      step was stopped there were 0 observations and 1
      variables.
```

Which of the following actions resolves the error message?

- a. Put quotation marks around THIS IS A TEST.
 - b. Put parentheses around THIS IS A TEST.
 - c. Add commas between the words THIS IS A TEST.
 - d. Add a FORMAT statement declaring **newvar** as character.
3. Which statement must be added to the DATA step in order for SAS to create the temporary FIRST. and LAST. variables?
- a. assignment statement
 - b. BY statement
 - c. SET statement
 - d. subsetting IF statement
4. The following program is submitted:

```
data work.total;
  n1 = 4;
  n2 = .;
```

```

n3 = 10;
n4 = n1 + n2 + n3;
run;

```

What is the resulting value of **n4** in the data set **work.total**?

- a. .
- b. 14
- c. $4 + . + 10$
- d. $n1 + n2 + n3$

5. Given the SAS data set **Work.Employees**:

Name
Jeff
Dawn
Mary
Gene

The following SAS program is submitted:

```

data Work.EmployeeCount;
  set Work.Employees;
  Count=Count+1;
run;
proc print data=Work.EmployeeCount noobs;
run;

```

What is the result of the PRINT procedure?

a.

Name	Count
Jeff	.
Dawn	.
Mary	.
Gene	.

b.

Name	Count
Jeff	0
Dawn	0
Mary	0
Gene	0

c.

Name	Count
Jeff	1
Dawn	1
Mary	1
Gene	1

d.

Name	Count
Jeff	1
Dawn	2
Mary	3
Gene	4

6. The following program is submitted:

```
data personnel;
  hired='01MAR2003'd;
  name='William Smith';
run;
```

Which of the following is true regarding the variables created with the assignment statements?

- a. The variables **hired** and **name** are 8 bytes.
- b. The variables **hired** and **name** are character.
- c. The variable **hired** is 9 bytes and **name** is 13 bytes.
- d. The variable **hired** is numeric and **name** is character.

7. Given the SAS data set **birth**:

NAME	STATE
Tim	CA
Sue	IN
Bill	NY

The following SAS program is submitted:

```
data birthregion;
  set birth;
  if state='CA' then do;
    region='West';
  end;
  else if state='NY' then do;
    region='East';
  run;
```

What is the result?

- a. The program fails execution because of invalid DO block syntax.
- b. The program fails execution because there is not a DO block for the **state** value of IN.
- c. The program runs without errors or warnings and produces a data set with two observations and three variables.
- d. The program runs without errors or warnings and produces a data set with three observations and three variables.

8. Which of the following is true regarding the sum statement?

- a. The sum statement can only be used for variables being read in from a SET statement.
- b. The sum statement initializes the variable to missing before the first iteration of the DATA step.
- c. The sum statement automatically retains the variable value without using a RETAIN statement.
- d. The sum statement produces an error if a missing value is added to the accumulator variable.

9. Which of the following is valid syntax for SELECT and WHEN statements?

- a.

```
select(salary);
```

```
when <100000 status='Non-Exec';
when >=100000 status='Exec';
end;
```

b.

```
select(salary);
when(<100000) status='Non-Exec';
when(>=100000) status='Exec';
end;
```

c.

```
select;
when salary<100000 status='Non-Exec';
when salary>=100000 status='Exec';
end;
```

d.

```
select;
when(salary<100000) status='Non-Exec';
when(salary>=100000) status='Exec';
end;
```

2. Analyse de données

On présente ci-après le nombre d'élèves admis au baccalauréat session 2014 sur l'ensemble du territoire français y compris DOM et Mayotte, issus de l'enseignement public et privé. *Source DEPP ministère de l'éducation nationale.*

Ces effectifs sont distingués d'une part selon la série du baccalauréat obtenu -baccalauréat général (Bac G), baccalauréat technologique (Bac T) et baccalauréat professionnel (Bac T)- et, d'autre part, selon les catégories socio professionnels de leurs parents. Les professeurs et les instituteurs font l'objet d'une rubrique spécifique.

Origine sociale	Baccalauréat général	Baccalauréat technologique	Baccalauréat professionnel
Agriculteurs exploitants	5 940	2 033	2 763
Artisans, commerçants, chefs d'entreprise	29 028	12 240	18 063
Professeurs et assimilés	15 554	2 099	888
Instituteurs et assimilés	4 695	874	481
Ouvriers	33 708	24 352	46 562
Autres personnes sans activité professionnelle	20 116	14 298	13 674
Cadres, professions intellectuelles supérieures, sans les professeurs et assimilés	91 066	17 338	12 590
Professions intermédiaires, sans les instituteurs et assimilés	42 828	17 546	15 560
Employés	48 811	25 406	22 417
Retraités	5 144	2 650	5 017
Indéterminé	8 777	10 374	52 758

On effectue une analyse factorielle des correspondances (AFC) de ces données.

Répondre aux questions suivantes.

1/ Qu'espère-t-on obtenir à l'aide de cette méthode ? Préciser sous quelle(s) condition(s) elle peut être mise en œuvre.

2/ Commenter les tableaux de profils colonne et profils ligne (annexe 1).

3/ Combien de dimensions peuvent être utilisées au maximum pour cette AFC ?

4/ Quelle distance est représentée lors d'une AFC entre les points représentant chacune des modalités des variables Série de Baccalauréat et Origine sociale ?

5/ Techniquement, quelle est la méthode sous-jacente utilisée pour effectuer une AFC ?

6/ Que représente la «masse» dans les tableaux de l'annexe 2 ? A quoi cet indicateur peut-il servir ?

7/ Que signifie l'indicateur « inertie » dans le tableau de l'annexe 3 ? Interpréter les valeurs obtenues.

8/ Quel test est-il effectué dans cette AFC (annexe 3) permettant de déterminer s'il est intéressant ou non de poursuivre l'AFC ? Interpréter le résultat obtenu.

9/ Quels enseignements tirer du tableau de l'annexe 3 ?

10/ On note CTR les contributions des modalités-individus et CO le cosinus carré (annexe 4). A quoi servent ces deux indicateurs ? Sont-ils à comparer avec un autre indicateur ?

11/ Indiquer les points modalités ligne et colonne qui seront finalement interprétables en précisant dans quelle dimension et pour quelle(s) raison(s).

12/ La proximité de deux points modalités lignes est-elle interprétable ? Si oui, préciser en quel sens. Même question pour deux points modalités colonne.

La proximité d'un point modalité ligne et d'un point modalité colonne est-elle interprétable ?

13/ Commenter la dernière représentation graphique de l'annexe 5 en utilisant l'ensemble des précautions prises grâce aux indicateurs préalablement étudiés.

14/ Rédiger une analyse synthétique de ces données accessible à un non spécialiste (20 lignes).

ANNEXES

Annexe 1

Tableau croisé Origine sociale * Série du baccalauréat

% dans Origine sociale

		Série du baccalauréat			Total
		Bac G	Bac P	Bac T	
Origine sociale	Agriculteurs exploitants	55,3%	25,7%	18,9%	100,0%
	Artisans, commerçants, chefs d'entreprise	48,9%	30,4%	20,6%	100,0%
	Autres personnes sans activité professionnelle	41,8%	28,4%	29,7%	100,0%
	Cadres, professions intellectuelles supérieures, sans les professeurs et assimilés	75,3%	10,4%	14,3%	100,0%
	Employés	50,5%	23,2%	26,3%	100,0%
	Indéterminé	12,2%	73,4%	14,4%	100,0%
	Instituteurs et assimilés	77,6%	8,0%	14,4%	100,0%
	Ouvriers	32,2%	44,5%	23,3%	100,0%
	Professeurs et assimilés	83,9%	4,8%	11,3%	100,0%
	Professions intermédiaires, sans les instituteurs et assimilés	56,4%	20,5%	23,1%	100,0%
	Retraités	40,2%	39,2%	20,7%	100,0%
Total		48,9%	30,5%	20,7%	100,0%

Tableau croisé Origine sociale * Série du baccalauréat

% dans Série du baccalauréat

		Série du baccalauréat			Total
		Bac G	Bac P	Bac T	
Origine sociale	Agriculteurs exploitants	1,9%	1,4%	1,6%	1,7%
	Artisans, commerçants, chefs d'entreprise	9,5%	9,5%	9,5%	9,5%
	Autres personnes sans activité professionnelle	6,6%	7,2%	11,1%	7,7%
	Cadres, professions intellectuelles supérieures, sans les professeurs et assimilés	29,8%	6,6%	13,4%	19,3%
	Employés	16,0%	11,8%	19,7%	15,4%
	Indéterminé	2,9%	27,7%	8,0%	11,5%
	Instituteurs et assimilés	1,5%	0,3%	0,7%	1,0%
	Ouvriers	11,0%	24,4%	18,8%	16,7%
	Professeurs et assimilés	5,1%	0,5%	1,6%	3,0%
	Professions intermédiaires, sans les instituteurs et assimilés	14,0%	8,2%	13,6%	12,1%
Total	Retraités	1,7%	2,6%	2,1%	2,0%
		100,0%	100,0%	100,0%	100,0%

Annexe 2

Profils lignes

Origine sociale	Série du baccalauréat			
	Bac G	Bac P	Bac T	Marge active
Agriculteurs exploitants	,553	,257	,189	1,000
Artisans, commerçants, chefs d'entreprise	,489	,304	,206	1,000
Autres personnes sans activité professionnelle	,418	,284	,297	1,000
Cadres, professions intellectuelles supérieures, sans les profe	,753	,104	,143	1,000
Employés	,505	,232	,263	1,000
Indéterminé	,122	,734	,144	1,000
Instituteurs et assimilés	,776	,080	,144	1,000
Ouvriers	,322	,445	,233	1,000
Professeurs et assimilés	,839	,048	,113	1,000
Professions intermédiaires, sans les instituteurs et assimilés	,564	,205	,231	1,000
Retraités	,402	,392	,207	1,000
Masse	,489	,305	,207	

Profils colonnes

Origine sociale	Série du baccalauréat			
	Bac G	Bac P	Bac T	Masse
Agriculteurs exploitants	,019	,014	,016	,017
Artisans, commerçants, chefs d'entreprise	,095	,095	,095	,095
Autres personnes sans activité professionnelle	,066	,072	,111	,077
Cadres, professions intellectuelles supérieures, sans les profs	,298	,066	,134	,193
Employés	,160	,118	,197	,154
Indéterminé	,029	,277	,080	,115
Instituteurs et assimilés	,015	,003	,007	,010
Ouvriers	,110	,244	,188	,167
Professeurs et assimilés	,051	,005	,016	,030
Professions intermédiaires, sans les instituteurs et assimilés	,140	,082	,136	,121
Retraités	,017	,026	,021	,020
Marge active	1,000	1,000	1,000	

Annexe 3

Récapitulatif

Dimension	Valeur singulière	Inertie	Khi-deux	Sig.	Proportion d'inertie	
					Représentation	Cumulé
1	,445	,198			,920	,920
2	,131	,017			,080	1,000
Total		,215	134 711,197	,000 ^a	1,000	1,000

a. 20 degrés de liberté

Annexe 4

Présentation des points de ligne^a

Origine sociale	Masse	Score de la dimension		Inertie	Contribution					
		1	2		Du point vers l'inertie de la dimension		De la dimension vers l'inertie du point			
					CTR1	CTR2	CO1	CO2	QLT	
Agriculteurs exploitants	,017	-,189	,108	,000	,001	,002	,913	,087	1,000	
Artisans, commerçants, chefs d'entreprise	,095	-,002	,001	,000	,000	,000	,868	,132	1,000	
Autres personnes sans activité professionnelle	,077	,059	-,617	,004	,001	,224	,030	,970	1,000	
Cadres, professions intellectuelles supérieures, sans les profe	,193	-,785	,392	,057	,268	,227	,932	,068	1,000	
Employés	,154	-,172	-,394	,005	,010	,183	,392	,608	1,000	
Indéterminé	,115	1,393	,496	,103	,501	,216	,964	,036	1,000	
Instituteurs et assimilés	,010	-,868	,380	,003	,016	,011	,947	,053	1,000	
Ouvriers	,167	,522	-,153	,021	,102	,030	,975	,025	1,000	
Professeurs et assimilés	,030	-1,022	,585	,015	,070	,078	,912	,088	1,000	
Professions intermédiaires, sans les instituteurs et assimilés	,121	-,311	-,183	,006	,026	,031	,907	,093	1,000	
Retraités	,020	,300	,013	,001	,004	,000	,999	,001	1,000	
Total actif	1,000			,215	1,000	1,000				

a. Normalisation symétrique

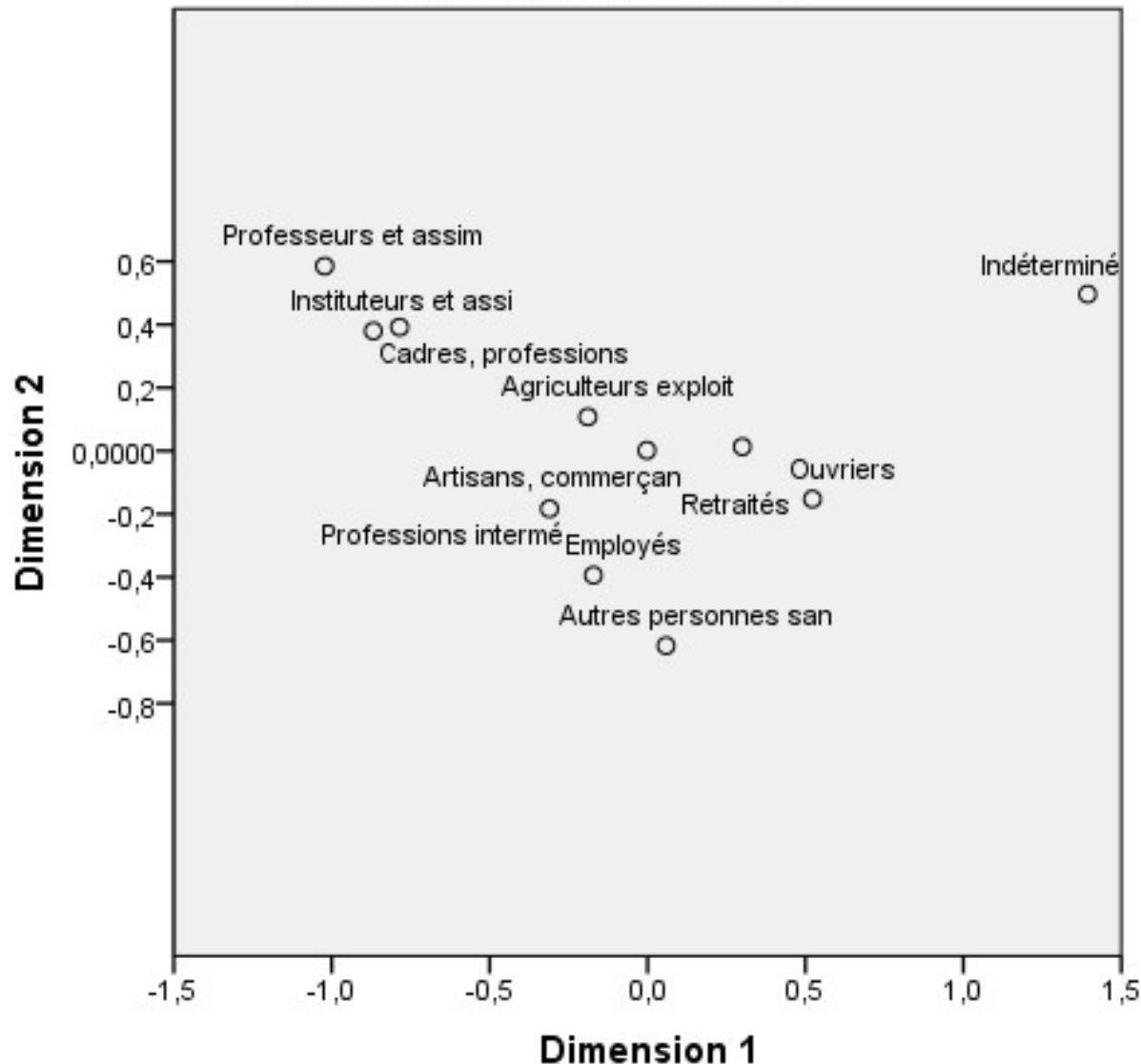
Présentation des points de colonne^a

Série du baccalauréat	Masse	Score de la dimension		Inertie	Contribution					
		1	2		Du point vers l'inertie de la dimension			De la dimension vers l'inertie du point		
					CTR1	CTR2	CO1	CO2	QLT	
Bac G	,489	-,601	,176	,081	,396	,115	,975	,025	1,000	
Bac P	,305	,938	,199	,121	,603	,092	,987	,013	1,000	
Bac T	,207	,036	-,709	,014	,001	,793	,009	,991	1,000	
Total actif	1,000			,215	1,000	1,000				

a. Normalisation symétrique

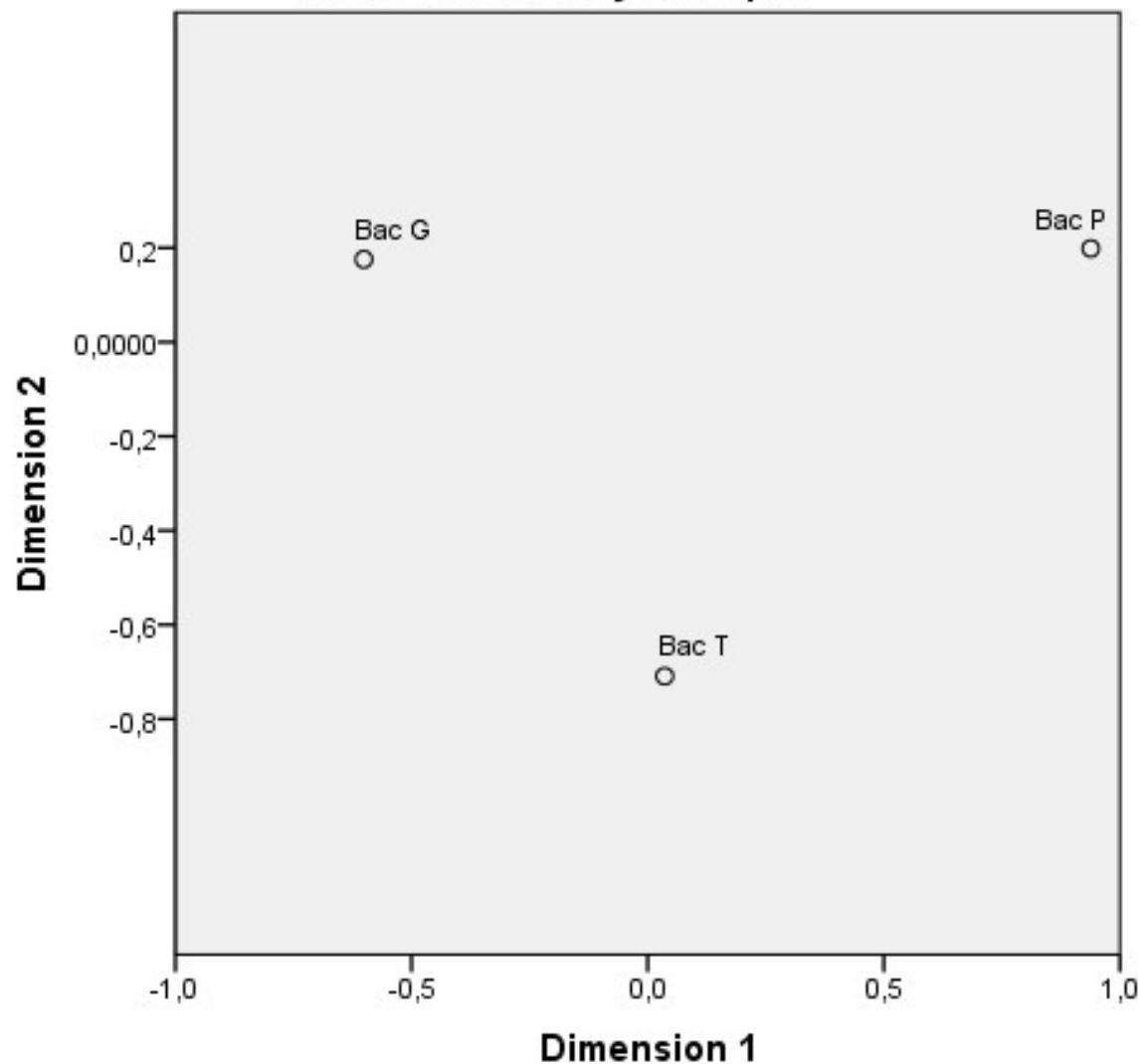
Annexe 5

Points de lignes pour Origine sociale
Normalisation Symétrique



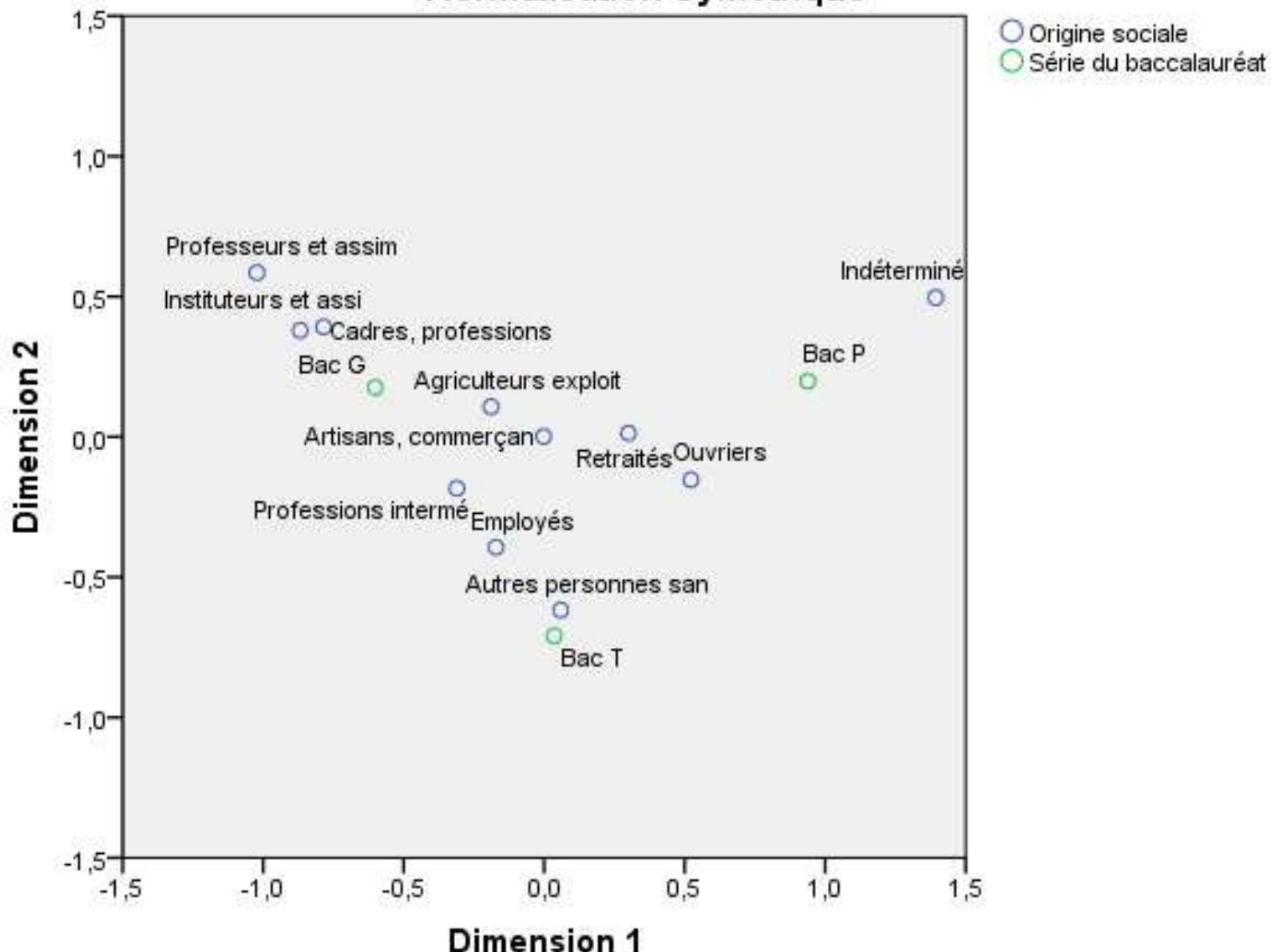
Points de colonnes pour Série du baccalauréat

Normalisation Symétrique



Points de lignes et de colonnes

Normalisation Symétrique



3. Modélisation

En transport de fret, les marchandises sont expédiées en ensembles de quantité donnée, appelés *envois*, remis par un expéditeur à un transporteur pour être acheminés à un destinataire. Un modèle théorique simple, appelé *Economic Order Quantity* (EOQ), prédit que quand un expéditeur envoie Q tonnes par an ($t.an^{-1}$) de marchandises à un destinataire donné, et que la densité de valeur de ces marchandises en € par tonne (€.t^{-1}) est notée a , la taille d'envoi est proportionnelle à la racine carrée du ratio de Q sur a , à une constante près :

$$s = c^{te} \sqrt{\frac{Q}{a}} \quad (1)$$

La constante dépend du mode de transport. Théoriquement, plus le mode de transport est « lourd » (adéquat pour transporter de grandes quantités à chaque opération), plus la constante est élevée.

L'enquête ECHO, réalisée en 2004-2005, observe 10462 envois expédiés depuis la France dans le monde entier par 3002 entreprises. Pour environ 60% des observations, l'enquête ECHO contient également des informations sur le poids de l'envoi, la valeur de l'envoi, la quantité de marchandises échangée annuellement entre l'expéditeur et le destinataire, et le mode de transport.

Considérons une simple régression linéaire, estimant les paramètres du modèle suivant :

$$\ln s = \beta_Q \ln Q_{tot} + \beta_a \ln a_{dens} + \sum_{mode} \beta_{mode} X_{mode} + u \quad (2)$$

où \ln représente le logarithme népérien, et X_{mode} une variable qui vaut 1 si le mode de transport correspond à la variable et 0 sinon.

Les modes de transport sont les suivants :

- *common carrier* : transport routier pour compte d'autrui (un transporteur professionnel réalise le transport pour le compte de son client)
- *private carrier* : l'entreprise réalise elle-même le transport
- *rail* : transport ferroviaire
- *combined* : transport combiné (ferroviaire et routier, la marchandise étant placé dans un conteneur spécial appelé la caisse mobile)
- *waterway* : transport fluvial
- *sea* : transport maritime
- *air* : transport aérien

Lorsque deux modes de transport sont utilisés, le mode non routier est retenu (si la route et le fleuve sont utilisés, on suppose que le mode de transport est le fleuve.)

Les résultats de l'estimation sont donnés dans la Table 1.

- 1) Les résultats de l'estimation sont-ils cohérents avec le modèle théorique (1) ?

- 2) Commentez le diagnostic graphique de la Figure 1. Quelles informations supplémentaires souhaiteriez-vous avoir pour vérifier la qualité de l'estimation ?
- 3) Proposez des améliorations de la spécification du modèle (2), en expliquant vos choix.

TABLE 1 Estimation du modèle EOQ

Coefficients	Estimate	Std. Error	t-stat	
β_Q	0.50	0.01	73.27	***
β_a	-0.44	0.01	-37.57	***
$\beta_{common\ carrier}$	1.05	0.11	9.47	***
$\beta_{private\ carrier}$	1.46	0.11	12.87	***
β_{rail}	3.42	0.18	19.30	***
$\beta_{combined}$	2.09	0.20	10.31	***
$\beta_{waterway}$	4.37	0.33	13.05	***
β_{sea}	2.89	0.13	21.49	***
β_{air}	1.47	0.14	10.29	***
N	10,462			
NAs	4,741			
R^2	0.795			
Adjusted R^2	0.795			

Significance levels: ‘.’ at 10%; ‘**’ at 5 %; ‘***’ at 1%; ‘****’ at 0.1 %

FIGURE 1 Diagnostic graphique de l'estimation du modèle EOQ