

Les problèmes des beaux mercredis

Saison 2
juillet 2012 - juin 2013

Ou,

Le cahier de vacances de www.sas-sr.com

Créée en septembre 2011, « Les beaux mercredis » est une section du site support de mon ouvrage d'introduction à SAS qui me permet de présenter, au travers d'exercices qui vous sont proposés, des exemples de programmes un peu "avancés" mais utilisant généralement les outils de programmation évoqués dans mon ouvrage.

La section des beaux mercredis fonctionne de la manière suivante : chaque mercredi, je propose

- 1- soit un nouvel exercice et vous donne une semaine pour y réfléchir,
- 2- soit la solution (totale ou partielle) à l'exercice proposé précédemment.

Les exercices proposés peuvent quelque fois sembler "amusants" : soyez cependant conscient que, pour résoudre ces petits exercices "amusants", je mobilise les mêmes instruments de programmation que ceux que vous utilisez pour résoudre des questions, plus réelles et concrètes et très certainement nettement moins amusantes... Je suis convaincu (et j'espère vous en convaincre !) qu'on peut vite progresser dans son usage quotidien de SAS en tentant de répondre aux petites questions posées par mes exercices.

Ce document reprend l'ensemble des sujets qui ont été proposés entre juillet 2012 et juin 2013. Il prend la forme d'un « cahier de vacances » puisque les solutions ne sont pas reprises dans le présent document. Si, au cours de cette année, vous n'avez pas été un lecteur attentif des beaux mercredis, emportez ce cahier de vacances avec vous sur la plage et réfléchissez aux programmes que vous pourriez rédiger afin de répondre aux questions posées !

Si vous souhaitez consulter les solutions, il vous suffit de vous connecter avec votre nom d'utilisateur et votre mot de passe, vous trouverez alors les solutions dans la section « les beaux mercredis – l'intégrale ».

Si vous ne vous identifiez pas sur www.sas-sr.com, vous ne pouvez consulter que les sujets repris dans le présent document – vous n'aurez pas accès aux solutions.

Vous avez perdu votre mot de passe ou votre identifiant ?

Rendez-vous ici : <http://www.sas-sr.com/login.php?action=forget>

Et indiquez l'email avec lequel vous vous êtes inscrit (c'est l'email que j'utilise pour vous contacter) – le serveur peut mettre du temps à vous envoyer vos informations de connexion... surveillez aussi votre boîte de courriers indésirables... et contactez moi si vous restez bloqué.

Vous n'avez aucune idée du contenu de mon ouvrage ?

Consultez ce document qui reprend la préface, l'introduction générale et le plan détaillé de mon ouvrage :

http://www.sas-sr.com/SAS_Introduction_Ringuede_seconde_edition.pdf

Vous souhaitez savoir comment vous procurer mon ouvrage ?

Consultez ce post : <http://www.sas-sr.com/viewtopic.php?id=500>

C'est collé !

Problème proposé le 26/06/2013 – solution mise en ligne le 02/07/2013

Un utilisateur de mon ouvrage m'a contacté pour me faire part de son petit problème et comme c'est un problème amusant, je le transforme en sujet des beaux mercredis.

Il dispose dans une table SAS d'une variable qui reprend le prénom et le nom de clients mais au sein de cette variable unique, prénom et nom sont "collés" et il voudrait bien pouvoir disposer d'une variable PRENOM et d'une variable NOM...

le programme suivant crée la table sur laquelle vous allez devoir réfléchir :

Code:

```
data test;
input colle $15.;
cards;
AndréDupont
PierreDubois
MaxLamenace
;
```

Vous devez obtenir le résultat suivant :

Code:

Obs.	colle	prenom	nom
1	AndréDupont	André	Dupont
2	PierreDubois	Pierre	Dubois
3	MaxLamenace	Max	Lamenace

et comme dans la vraie vie, les choses sont toujours plus compliquées que dans les exercices d'application, vous réfléchirez aussi à comment obtenir un résultat similaire avec les deux jeux de données suivants :

Code:

```
data test2;
input colle $24.;
cards;
DUPUISJean-Louis
DUPONTPierre
DUPUIS-DUBOISLuc
DUPUIS-DUBOISJean-Pierre
;
```

```
data test3;
input colle $24.;
cards;
Jean-LouisDupuis
PierreDupont
LucDupuis-Dubois
Jean-PierreDupuis-Dubois
;
```

Quel est l'impact du durcissement du quotient familial ?

Problème proposé le 05/06/2013 – solution mise en ligne à partir du 12/06/2013

Vous n'êtes pas sans savoir que le gouvernement, pour équilibrer les comptes de la branche famille de la sécurité sociale, vient de décider que la réduction maximale sur les impôts sur le revenu apportée par demi-part supplémentaire allait passer de 2000 euros à 1500 euros.

Les impôts sur le revenu des ménages avec enfants les plus aisés devraient donc augmenter et je vous propose aujourd'hui de mesurer l'impact de cette mesure (qui est difficile à apprécier - même si vous avez des enfants et si vous payez des impôts (c'est mon cas...) - j'ignore au jour d'aujourd'hui si mes propres impôts augmenteront... vous devinez donc que je n'ai pas encore préparé la solution à ce problème des beaux mercredis de www.sas-sr.com...)

Le calcul de l'impôt sur le revenu n'est pas une chose simple mais on peut déjà consulter la notice explicative du calcul de l'impôt sur le revenu 2012 ici : [http://www2.impots.gouv.fr/documentatio ... index.html](http://www2.impots.gouv.fr/documentatio...index.html)

la notice fait 382 pages... mais les pages qui vont nous intéresser sont les pages 313 à 317 (les points 1 à 7 de la fiche de calculs facultatifs).

Nous allons considérer diverses situations ; couple mariés (ou pacsé) avec 1, 2, 3 et 4 enfants, vie maritale ou célibataire (pas de déclaration commune dans ce cas) avec 1, 2, 3 ou 4 enfants et regarder pour des niveaux de revenu annuels allant jusqu'à 200 000 euros (couple) ou 100 000 euros (célibataire) (par pas de 2500 euros) ce que seraient les impôts sur le revenu si la réforme était appliquée dès maintenant.

Vous devez donc calculer les impôts de ces personnes au moyen du mode de calcul appliqué actuellement et les calculer ensuite, toujours sur la base du mode de calcul de cette année, mais en modifiant le seuil du plafonnement du quotient familial (voir le point 7 de la fiche de calculs).

Considérez que nos ménages types ne perçoivent que des revenus salariaux (pas de revenus des valeurs et capitaux mobiliers, de revenus fonciers, de revenus des professions non salariées, ni de déficits des années antérieures - pas de CSG déductible) - pas de déduction de charges ni d'abattements spéciaux (voir point 2 de la fiche de calculs)

Vous devez calculer l'impôt de vos ménages types jusqu'à l'application de la décote expliquée au point 7.

Le mode de calcul est expliqué intégralement mais les choses semblent quelque fois compliquées... on devrait s'en sortir puisqu'on utilise SAS...

N'hésitez pas à utiliser les simulateurs d'impôts sur le revenu que vous trouverez en ligne par exemple ici : [http://www3.finances.gouv.fr/calcul_imp ... /index.htm](http://www3.finances.gouv.fr/calcul_imp.../index.htm)

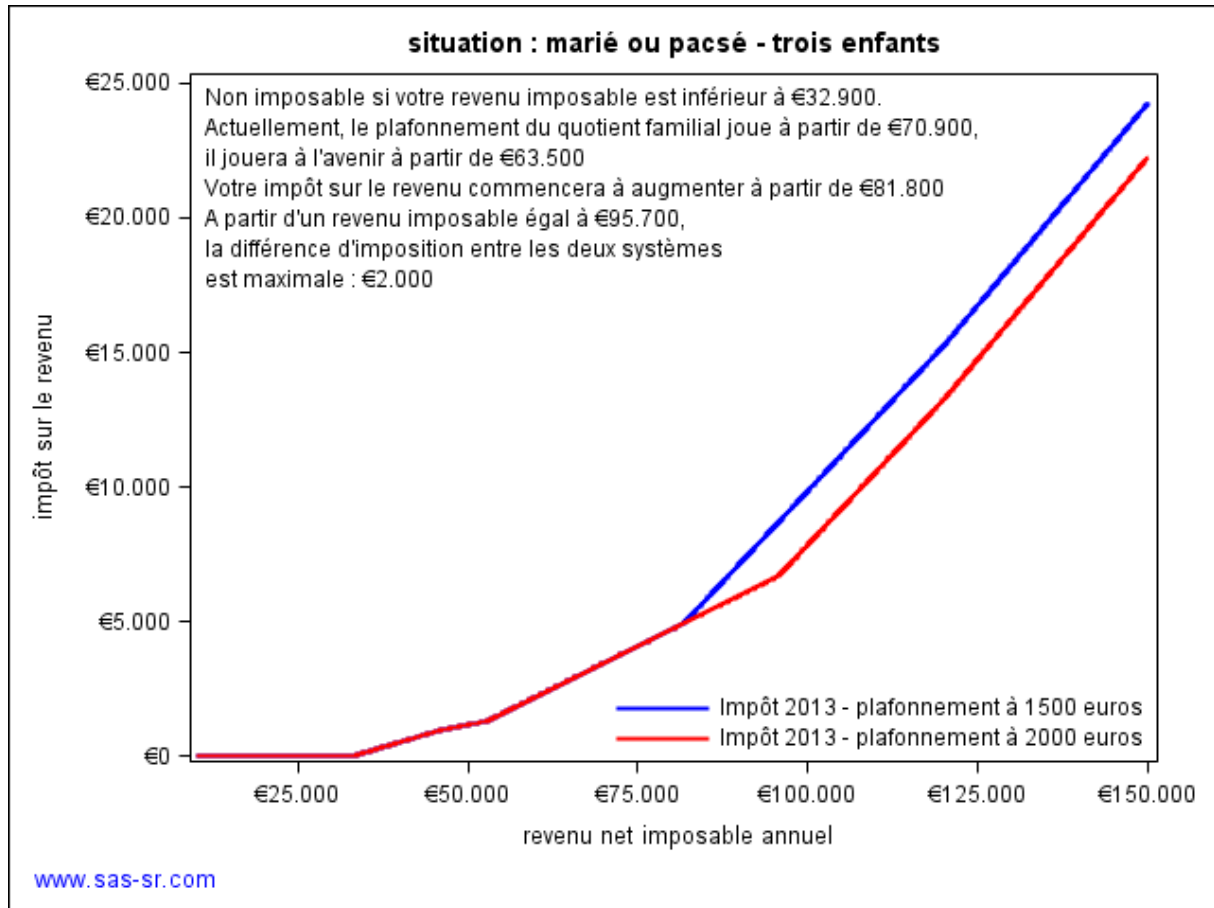
Ces simulateurs vous aideront à vérifier si votre programme calcule correctement les impôts sur le revenu dans les conditions actuelles

Quelques explications (mais vous en trouverez sûrement d'autres ailleurs) sur le fonctionnement du seuil du plafonnement du quotient familial peuvent être trouvées ici : [http://www.huffingtonpost.fr/2013/06/03 ... 76733.html](http://www.huffingtonpost.fr/2013/06/03...76733.html)

Mais il doit certainement y avoir sur Internet d'autres sites devant vous aider à comprendre comment les choses fonctionnent...

Une fois que vous avez vos impôts avant / après, réfléchissez à un moyen de représenter l'impact de cette modification du seuil du plafonnement du quotient familial.

Note : le programme proposé par la solution permet la production de ce type de graphique :



Qui a eu la meilleure note ?

Problème proposé le 22/05/2013 – solution mise en ligne le 29/05/2013

Pour ce nouveau "beau mercredi", nous allons travailler sur la table créée par le programme suivant :

Code:

```
data exem;
input cours $ sophie pierre jean rene roger charlotte;
cards;
cours1 10 12 14 17 11 4
cours2 17 12 16 14 17 8
cours3 11 12 6 5 11 13
;
```

Je souhaite disposer dans cette table d'une variable supplémentaire qui me donnera le nom de l'étudiant qui a obtenu la meilleure note. Je veux donc obtenir le résultat suivant :

Code:

Obs.	cours	sophie	pierre	jean	rene	roger	charlotte	gagnant
1	cours1	10	12	14	17	11	4	rene
2	cours2	17	12	16	14	17	8	roger
3	cours3	11	12	6	5	11	13	charlotte

Le programme que vous allez écrire doit être indépendant du nom des étudiants pour l'instant saisi uniquement comme nom de variable... il est donc parfaitement exclu d'écrire un programme qui aurait des instructions prenant la forme suivante :

Code:

```
If ... then gagnant="charlotte";
```

Vous remarquez pour le cours 2 que roger et sophie ont eu la même note mais que seul roger apparaît comme gagnant... nous devons aussi régler ce petit problème...

Comprenez ce problème des beaux mercredis comme une invitation à fouiller dans l'aide SAS et découvrir peut être une fonction (ou deux) que vous ne connaissez pas encore...

528 fonctions sont citées dans cette page de support.sas.com :
<http://support.sas.com/documentation/cd ... h2ouzm.htm>

il y en a certainement au moins une qui pourrait vous être utile...

Vous aimez les longs week-ends ? Le prochain, c'est quand ?

Problème mis en ligne le 8/05/2013 – solution proposée le 15/05/2013

Quelle belle année que 2013 ! (et c'est l'actualité qui me pousse à proposer un second "beau mercredi" en parallèle à celui qui traite du rendement de l'éducation...voir le sujet suivant)

Cette année, le jeudi de l'ascension tombe le 9 mai, soit le lendemain du 8 mai (c'est implacable...) et vous bénéficiez d'un "pont" qui ressemble plus à un viaduc qu'à autre chose...

Mais en quelle année bénéficierons-nous du prochain "viaduc" ?

Dit autrement, en quelle année observerons-nous un 8 mai qui tombe un mercredi et un jeudi de l'ascension qui tombera le lendemain ?

Lorsque vous aurez la réponse à la question, vous pourrez aussi vous demander si un "mini-viaduc" ne va pas apparaître avant ?

Un mini viaduc c'est un jeudi de l'ascension qui tombe le 7 mai...

Pendant que nous y sommes, il est aussi possible que le jeudi de l'ascension tombe un 8 mai... en quelle année interviendra cette conjonction néfaste ?

Pour répondre aux petites questions posées ici, il va vous falloir peut être consulter cette page de wikipédia... http://fr.wikipedia.org/wiki/Ascension_%28f%C3%AAt%29

et bien entendu, rédiger un petit programme SAS...

Une autre conjonction intéressante serait d'avoir le 8 mai un lundi et d'avoir aussi comme jeudi de l'ascension le jeudi qui précède...

en quelle année observerons-nous cela ?

Si enfin, la "super conjonction"... imaginez... le 8 mai tombe un mardi et le jeudi qui suit est le jeudi de l'ascension !

En quelle année allons-nous observer cette super conjonction ?

La dernière fois que c'est arrivé, c'était en 1956... La prochaine fois, c'est pour bientôt...

Le rendement de l'éducation

Problème mis en ligne le 17/04/2013 – solution proposée à partir du 01/05/2013

Chaque année, l'OCDE fait paraître un rapport intitulé « regard sur l'éducation ». Il s'agit d'une source de premier ordre pour toute personne qui souhaite disposer d'indicateurs récents et comparables sur les résultats des systèmes éducatifs des pays de l'OCDE.

Pour ce beau mercredi, nous allons plus particulièrement nous intéresser à la rentabilité privée d'une formation dans le supérieur (formation tertiaire dans la terminologie l'OCDE).

Le rapport peut être consulté à partir de cette page : <http://www.oecd.org/fr/edu/rse2012.htm> (Cliquez sur la couverture pour le consulter) – les pages 168 à 188 vont plus particulièrement nous intéresser.

L'éducation peut être vue comme un investissement et comme tout investissement, il y a des coûts d'investissement, des bénéfices attendus et un taux de rendement interne.

Vous êtes actuellement étudiant et vous préparez un diplôme de l'enseignement supérieur. La préparation de votre diplôme est synonyme de coûts de deux ordres :

- 1- Des coûts directs (les frais de scolarité)
- 2- Un manque à gagner lié au fait que, pendant vos études, vous ne travaillez pas et ne percevez donc pas de salaire.

L'OCDE estime que le coût direct d'une formation dans le supérieur est en France d'environ 7868 dollars et que le manque à gagner est de 54888 dollars pour un garçon (52263 dollars pour une fille).

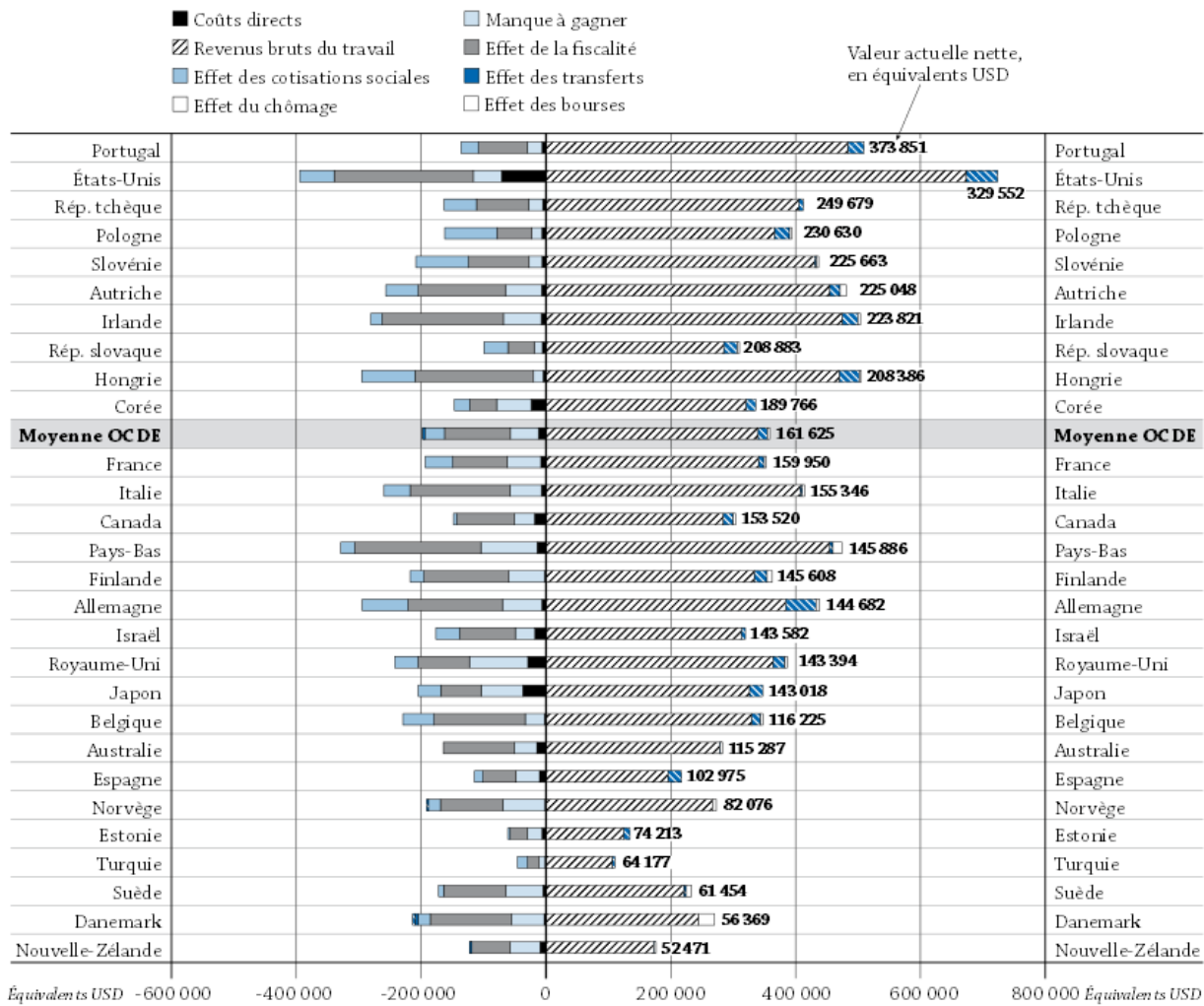
Quels sont les rendements attendus de cet investissement qui peut sembler conséquent ?

Ils sont composés de plusieurs éléments : le diplôme va déjà permettre d'obtenir des revenus salariaux plus élevés. On peut ajouter à ces revenus salariaux plus élevés les montants des bourses que vous recevez pour suivre vos études et le fait que votre diplôme vous évitera de passer un certain nombre de périodes au chômage. De l'autre côté, puisque vos revenus salariaux seront plus élevés, vous payerez aussi plus d'impôts et plus de cotisations sociales.

L'OCDE estime qu'en France, la valeur actualisée de l'ensemble des rendements attendus de votre investissement en formation est égal à 222 406 dollars pour un garçon et 170 561 dollars pour une fille (cette différence entre garçon et fille peut s'expliquer par le fait que filles et garçons ne suivent pas forcément le même type de formation, qu'ils n'exercent pas forcément le même type d'emploi dans les mêmes secteurs d'activité, qu'ils ne travailleront pas le même nombre d'années ni le même nombre d'heures par année et peut être aussi par des phénomènes de discrimination...).

Si vous vous rendez page 173, vous observerez un très intéressant graphique qui résume pour un ensemble de pays les caractéristiques des formations supérieures vues comme des investissements. Seules les données relatives aux garçons sont reproduites dans ce graphique.

Graphique A9.3. Composantes de la valeur actuelle nette privée de l'obtention d'un diplôme de fin d'études tertiaires (CITE 5/6) chez les individus de sexe masculin (2008 ou année de référence indiquée)



Votre mission (si vous l'acceptez) sera de reproduire avec SAS un graphique équivalent.

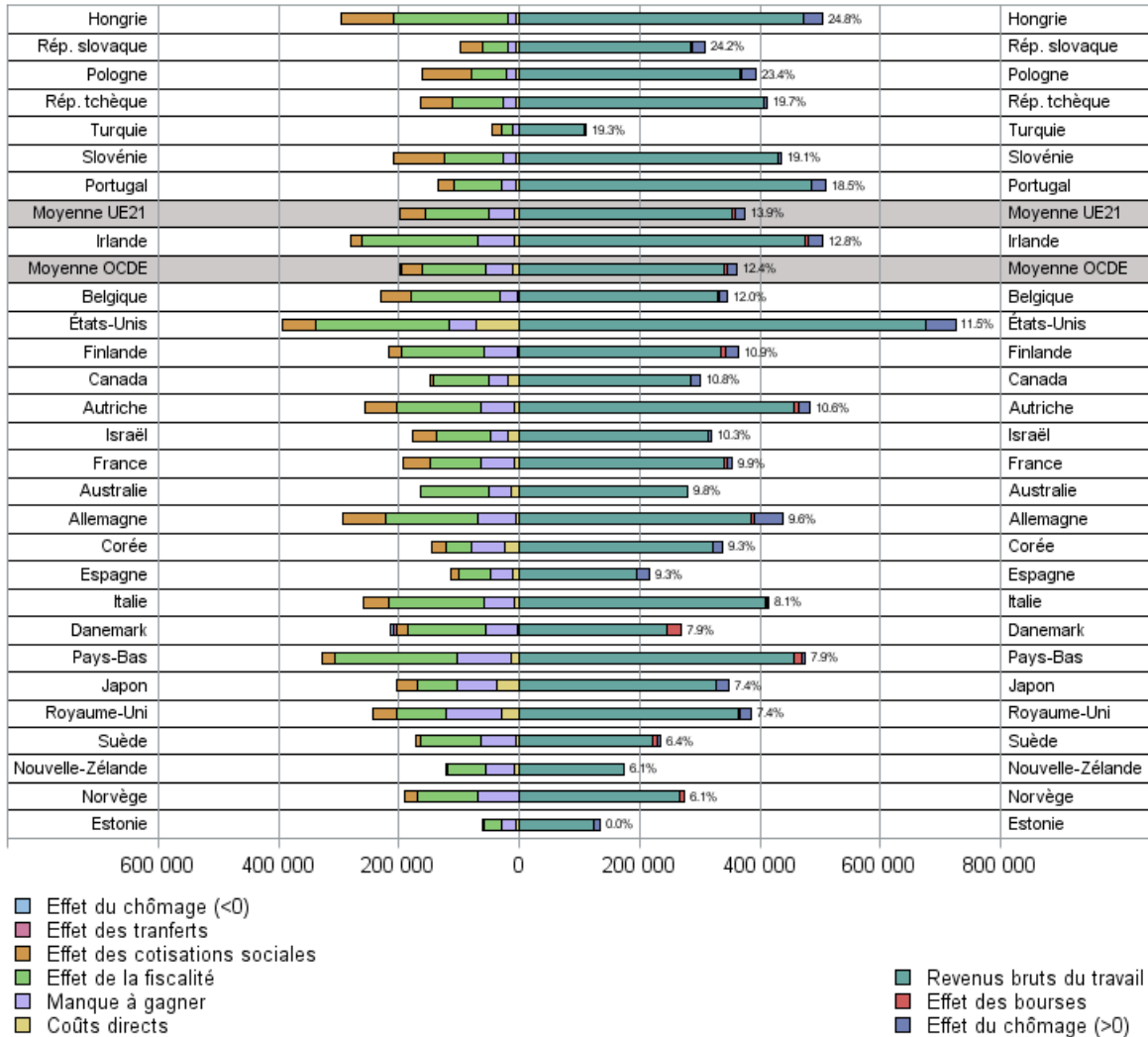
Les données peuvent être [téléchargées ici](#)

Et comme je serai en vacances la semaine prochaine, vous avez 15 jours pour réfléchir à ce problème...

Au fait, le taux de rendement interne d'une formation dans le supérieur en France est de 9.9% pour un garçon et de 9.4% pour une fille : vous ne trouverez pas, dans toute votre vie, d'investissement qui vous rapportera autant...

Note : nous aboutirons avec SAS au résultat suivant :

Composantes de la valeur actuelle nette privée de l'obtention d'un diplôme de fin d'études tertiaires chez les individus de sexe masculin



Source : *Regards sur l'éducation 2012 - OCDE*

Manipuler des données de calendrier avec SAS

Problème mis en ligne le 6/03/2013, solution proposée à partir du 13/03/2013

Il m'arrive assez souvent d'avoir des questions sur des tables dans lequel on suit l'évolution d'individus au moyen d'un calendrier. Voici un petit exemple de table :

Code:

ind	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
1	3	1	4	2	2	3
2	1	.	.	.
3	.	3	2	3	1	1	2	.	.	.
4	.	.	.	4	2	2	2	.	.	.
5	.	1	3	3	1	4	1	1	2	.

Les variables T1 à T10 vous donnent le statut de l'individu au cours des mois 1 à 10. L'individu 1 est « arrivé » le mois 5, avec un statut 3, il est resté dans la table jusqu'au mois 10, soit 6 périodes - le statut a connu des variations entre les deux dates.

L'individu 2 n'est resté qu'un mois dans la table (mois 7). L'individu 3 est arrivé le mois 2 et il est sorti de la table le mois 7.

Ces tables posent généralement aux utilisateurs de SAS des difficultés et nous allons examiner au moyen de quelques exercices.

La table que nous allons manipuler est créée par le programme suivant :

Code:

```
data test(drop=mlo mla i alea);
array t(200);
do ind=1 to 100000;
  mlo=ceil(200*ranuni(54));
  mla=ceil(200*ranuni(32));
  do i=1 to 200;
    t(i)=.;
    alea=ranuni(47);
    if i>min(mlo,mla) then t(i)=1+4*((alea>.90)+(alea>.95));
    if i>max(mlo,mla) then t(i)=.;
  end;
  output;
end;
run;
```

Nous allons donc travailler sur une table de 100 000 observations et 200 variables T. Les variables T, si elles ne sont pas en valeur manquante, peuvent prendre les valeurs 1 (dans 90% des cas), 5 (5% des cas) ou 9 (5% des cas). Nous allons considérer que ces modalités 1,5 et 9 décrivent le statut de l'individu à chaque période. Entre le moment où l'individu présente pour la première fois un statut et le moment où on dispose pour la dernière fois d'un statut, il n'y a pas de valeur manquante.

Vous allez me dire « mais on ne peut pas travailler sur moins d'observations et moins de variables ? ».

Non

En vous proposant une table « importante » (mais ce n'est pas grand-chose par rapport à ce que vous pouvez rencontrer dans la vie réelle), je vais aussi vous inviter à programmer « efficacement » et donc à adopter de bonnes pratiques... Il existe toujours plusieurs manières d'obtenir un même résultat mais certains programmes peuvent être particulièrement « gourmands » et prendre des heures à être exécuté alors qu'au moyen d'une programmation astucieuse, il ne prendra que quelques secondes (ou minutes...).

Donc, pour la semaine prochaine, vous allez devoir construire cinq nouvelles variables :

Variable DEB : le mois d'arrivée de chaque individu dans la table (5 pour le premier individu de ma table exemple présentée plus haut),

variable STA1 : le statut de l'individu au moment où il rentre dans la table (3 dans le cas de ma table exemple pour le premier individu),

variable OUT : le mois de sortie de chaque individu (10),

variable STA2 : la statut de l'individu au moment où il sort de la table (3),

variable PER : le nombre de périodes pendant lesquelles on dispose pour l'individu d'un statut (6).

Note : de nombreuses autres questions ont été ajoutées à ce sujet proposé le 6/03/2013 et traité entre le 13/03/2013 et le 10/04/2013

PROC TABULATE et instruction BY-comment retirer variableby=modalité ?

Problème mis en ligne le 20/02/2013 – solution proposée le 27/02/2013

Je n'avais pas beaucoup d'inspiration cette semaine et puis hier, on m'a posé cette question... comme elle est amusante, je la recycle en sujet d'un "beau mercredi"....

alors voilà - partons du programme suivant :

Code:

```
proc sort data=sashelp.shoes out=t;
by region;
run;

proc tabulate data=t ;
class product;
var sales;
tables (product all)*sales;
by region;
run;
```

Si vous regardez le résultat, vous observez qu'au dessus de chaque tableau (un tableau par modalité de REGION), vous observez REGION=modalité :

Code:

```
Region=Africa
...
Region=Asia
...
Region=Canada
...
```

Mon interlocuteur me dit : "cette reprise systématique de REGION= , c'est vraiment trop laid, est-il possible de ne voir s'afficher que le nom de la région en question ?"

Et comme tout est possible avec SAS, je lui ai dit "oui"

(maintenant, au moment où j'écris ces lignes, je n'ai pas encore la solution... juste quelques idées...)

Alors, que pourriez-vous proposer ?

edit : bon, c'est en fait très simple... il suffit de le savoir... (Je l'ai su et je l'avais oublié...)

Recoder sans trier (et sans fusion...)

Problème mis en ligne le 06/02/2013 – solution proposée le 12/03/2013

Il m'arrive le week-end, quand je ne sais pas trop quoi faire, d'aller sur le forum SAS de developper.com pour répondre à des questions qui n'ont pas reçu de réponses.

Le 23 janvier dernier, ça a été le cas :

<http://www.developpez.net/forums/d13009...der-trier/>

Le problème posé par [ibri42](#) est amusant. Il dispose d'une table énorme, impossible à trier mais il souhaite pouvoir mettre à jour les modalités d'une variable appelée CODE au moyen d'une table contenant les correspondances entre anciens codes et nouveaux codes.

Je lui donne une piste mais j'ai dû être trop évasif parce qu'il me demande de préciser (rassurez vous, je le contacterai dans quelques minutes pour lui donner la solution). En attendant, aidé par les explications nébuleuses données sur le forum de developper.com, pouvez-vous rédiger le programme qui va permettre, sans aucun tri (et donc sans MERGE – ni fusion au moyen de PROC SQL...) d'effectuer cette mise à jour des modalités de la variable CODE ?

Il s'agit donc de trouver la programmation la plus économe en ressources...

Imaginons que cette première table construite par le programme suivant est la table énorme impossible à trier :

Code:

```
data enorme;
input code $ valeur;
cards;
AAA1 1
CCD2 2
AAB1 3
CCD2 4
AAA0 5
CCD2 6
AAA1 7
;
```

Imaginons maintenant que la seconde table vous donne les correspondances entre anciennes modalités de code (CODE) et nouvelles modalités de code (NCODE) :

Code:

```
data modifcode;
input code $ ncode $;
cards;
AAA1 OOOO
ZZZ2 PPPP
AAB1 ZZZZ
AAA0 RRRR
CCD2 SSSS
;
```

Les nombres fraternels

Problème mis en ligne le 16/01/2013 – solution proposée à partir du 23/01/2013

J'aime bien aller de temps en temps sur le site www.diophante.fr.

On y trouve plein de petits problèmes mathématiques amusants et si des utilisateurs de ce site y répondent au moyen d'une simple feuille de papier, d'un crayon et de leur cerveau, je tente de mon côté de trouver les solutions en utilisant SAS...

Voici un problème amusant :

Un entier est appelé nombre fraternel si sa représentation décimale est faite de la juxtaposition de deux nombres consécutifs de n chiffres chacun.

Par exemple 256 257 et 12 341 235 sont deux nombres fraternels mais 98099 ne l'est pas en raison du zéro inséré entre 98 et 99.

Trouver tous les nombres fraternels de 4, 6, 8 et 10 chiffres qui sont en même temps des carrés parfaits.

Parmi ces chiffres, vous devez retrouver par exemple $715716=846^2$ mais aussi 82428241 puisqu'il répond à la définition (et qu'il est égal à 9079^2).

Il faudra trouver un programme court et rapide...

Nous profiterons aussi de ce petit exercice pour revoir le fonctionnement de PROC FCMP (voir page 135 et suivantes).

Les vœux 2013 de www.sas-sr.com

Problème mis en ligne le 02/01/2013 – solution proposée le 09/01/2013

les inscrits sur le site support de mon ouvrage ont reçu le 2 janvier ce mail :

Bonjour

Je vous présente mes meilleurs vœux pour cette année 2013 qui débute et comme chaque année depuis 2010, j'ai aussi le plaisir de vous offrir un petit cadeau, entièrement réalisé avec SAS bien entendu ;-)

<http://www.sas-sr.com/calendrier2013.pdf>

Retrouvez à partir du 8 janvier prochain dans la section « les beaux mercredis » le programme qui permet la réalisation de ce calendrier et grâce auquel vous pourrez réaliser vous-même vos propres calendriers, forcément différents de celui que vous pouvez télécharger dès maintenant.

Cordialement

Sébastien Ringuédé

Le programme qui génère ce calendrier sera mis en ligne mercredi prochain !

Pour mémoire, le petit cadeau pour la nouvelle année 2012 peut encore être vu ici :

<http://www.youtube.com/watch?v=8qkB8X4iUec>

les explications sont là : <http://www.sas-sr.com/viewtopic.php?id=627>

le petit cadeau pour la nouvelle année 2011 est ici :

<http://www.sas-sr.com/img/table.html>

explications : <http://www.sas-sr.com/viewtopic.php?id=428>

le petit cadeau pour la nouvelle année 2010 est ici :

<http://www.sas-sr.com/img/voeux2010.png>

explications : <http://www.sas-sr.com/viewtopic.php?id=248>

Manipulation d'une matrice de var-covar avec une étape DATA

Problème proposé le 12/12/2012, solution mise en ligne le 19/12/2012

J'ai reçu une question amusante de la part d'une étudiante la semaine dernière et je vais vous en faire profiter ;-)

Son problème est le suivant : elle dispose dans une table d'une matrice de variances covariances et souhaiterait disposer dans une seconde table uniquement des covariances. Comme une matrice de variances covariances est symétrique, elle ne veut voir dans sa table que les covariances présentées au-dessus de la diagonale. Ces covariances doivent être reprises dans une variable unique.

Créons un fichier test (qui n'est pas une matrice de variances covariances mais si votre programme fonctionne avec ces données, il fonctionnera avec une vraie matrice de variances covariances) :

Code:

```
data test;
input x1-x6;
cards;
  0    1    2    3    4    5
-1    0    6    7    8    9
-2   -3    0   10   11   12
-4   -5   -6    0   13   14
-7   -8   -9  -10    0   15
-11 -12 -13 -14 -15    0
;
```

Et on doit donc arriver au résultat suivant :

Code:

Obs	cov
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15

Elle m'a demandé si on devait impérativement passer par IML ou si une étape DATA pouvait permettre d'obtenir ce résultat.

C'est bien entendu possible au moyen d'une étape DATA...

A vous de rédiger maintenant le programme qui permet l'obtention de ce résultat !

Révisons PROC TABULATE et apprenons de nouvelles choses !

Problème proposé le 21/11/2012 – solution mise en ligne à partir du 28/11/2012

J'aime beaucoup PROC TABULATE... cette procédure est en effet tout à fait intéressante dans des phases d'exploration de vos données mais les utilisateurs de SAS rechignent quelque fois à l'utiliser la jugeant d'une syntaxe un peu étrange...

Je vais vous donner l'occasion de réviser le fonctionnement de PROC TABULATE et aussi d'apprendre de nouvelles choses que vous ne savez pas forcément (et qui ne sont pas dans mon ouvrage).

Commençons par construire deux tables...

Code:

```
data ptabu;
input Produit $ Jugement $ Ligne n;
cards;
a bon 1 1
a mauvais 1 2
b bon 1 3
b mauvais 1 4
c bon 1 5
c mauvais 1 6
a bon 2 7
a mauvais 2 8
b bon 2 9
b mauvais 2 10
c bon 2 11
c mauvais 2 12
a bon 3 13
a mauvais 3 14
b bon 3 15
b mauvais 3 16
c bon 3 17
c mauvais 3 18
;

data ptabu2;
  set ptabu;
  do i=1 to n;
    output;
  end;
run;
```

Question 1 :

Construisez au moyen de la table PTABU2 les différents tableaux reproduits ci-dessous :

	Produit												Tout		
	a				b				c						
	Jugement				Jugement				Jugement						
	bon		mauvais		bon		mauvais		bon		mauvais				
	???	???	???	???	???	???	???	???	???	???	???	???			???
Ligne															
1	1	0.58	2	1.17	3	1.75	4	2.34	5	2.92	6	3.51	21	12.28	
2	7	4.09	8	4.68	9	5.26	10	5.85	11	6.43	12	7.02	57	33.33	
3	13	7.60	14	8.19	15	8.77	16	9.36	17	9.94	18	10.53	93	54.39	
Tout	21	12.28	24	14.04	27	15.79	30	17.54	33	19.30	36	21.05	171	100.00	

	Produit												Tout	
	a				b				c					
	Jugement				Jugement				Jugement					
	bon		mauvais		bon		mauvais		bon		mauvais			
	???	???	???	???	???	???	???	???	???	???	???	???		
Ligne														
1	1	4.76	2	8.33	3	11.11	4	13.33	5	15.15	6	16.67	21	12.28
2	7	33.33	8	33.33	9	33.33	10	33.33	11	33.33	12	33.33	57	33.33
3	13	61.90	14	58.33	15	55.56	16	53.33	17	51.52	18	50.00	93	54.39
Tout	21	100.00	24	100.00	27	100.00	30	100.00	33	100.00	36	100.00	171	100.00

	Produit												Tout	
	a				b				c					
	Jugement				Jugement				Jugement					
	bon		mauvais		bon		mauvais		bon		mauvais			
	???	???	???	???	???	???	???	???	???	???	???	???		
Ligne														
1	1	4.76	2	9.52	3	14.29	4	19.05	5	23.81	6	28.57	21	100.00
2	7	12.28	8	14.04	9	15.79	10	17.54	11	19.30	12	21.05	57	100.00
3	13	13.98	14	15.05	15	16.13	16	17.20	17	18.28	18	19.35	93	100.00
Tout	21	12.28	24	14.04	27	15.79	30	17.54	33	19.30	36	21.05	171	100.00

Question 2 : sauriez-vous réaliser les mêmes tableaux au moyen de la table PTABU ? (je vous aide, la réponse est « oui », maintenant, dites moi « comment »...)

Les questions 1 et 2 sont relativement faciles si vous avez bien compris la section consacrée à PROC TABULATE dans mon ouvrage (pages 290 à 307).

Corsons un peu les choses...

Question 3 : Sauriez-vous réaliser au moyen de PROC TABULATE ce tableau ?

	Produit												Tout	
	a				b				c					
	Jugement				Jugement				Jugement					
	bon		mauvais		bon		mauvais		bon		mauvais			
	???	???	???	???	???	???	???	???	???	???	???	???		
Ligne														
1	1	33.33	2	66.67	3	42.86	4	57.14	5	45.45	6	54.55	21	12.28
2	7	46.67	8	53.33	9	47.37	10	52.63	11	47.83	12	52.17	57	33.33
3	13	48.15	14	51.85	15	48.39	16	51.61	17	48.57	18	51.43	93	54.39
Tout	21	46.67	24	53.33	27	47.37	30	52.63	33	47.83	36	52.17	171	100.00

A la semaine prochaine...

Et si on construisait avec SAS une "jolie" table de Student ?

Problème proposé le 10/10/2012 – solution mise en ligne à partir du 17/10/2012

Il y a quelques temps, j'étais à la recherche d'une table de Student*.

Je suis allé sur Google et j'ai tapé "table de Student". Je suis tombé certes sur des tables de Student mais elles n'étaient pas très jolies... cela ressemblait essentiellement à des scans de photocopie de photocopie de photocopie de photocopie de stencils** de table de Student.

Et si, avec SAS, on réalisait une "jolie" table de Student ?

En travaillant bien, on peut arriver à [ce résultat](#)

Notre table de Student est composée de deux éléments, un graphique et un tableau. Pour la semaine prochaine, vous aurez à construire le tableau... vous aurez très certainement besoin de relire le chapitre 8 consacré à ODS... l'introduction dans ce tableau du signe alpha et du + l'infini nécessite de "*l'inline formatting*"... (voir page 360 et suivantes). Vous aurez aussi très certainement besoin de la fonction QUANTILE (voir [l'aide SAS](#))

Amusez vous bien...

* Mr Student n'a en fait jamais existé... son vrai nom était William Sealy Gosset et il travaillait comme chimiste pour la brasserie Guinness. La brasserie Guinness avait au début du 20ème siècle des difficultés avec la bière qu'elle produisait essentiellement parce que la qualité de la bière n'était pas constante. Gosset a développé pour régler ce problème que le théorème central limite ne permettait pas d'appréhender, les outils sur les échantillons de petite taille que nous connaissons maintenant. Guinness ne voulait cependant pas que Gosset publie sous son nom ses recherches (le précédent chimiste de Guinness avait publié sous son nom des articles scientifiques et avait été débauché par la concurrence...).

On ne sait pas trop s'il demanda alors à Guinness l'autorisation de publier anonymement en utilisant le pseudonyme "a Student" ou s'il prit de lui même cette initiative mais toujours est-il que pendant près de 30 ans, il continua à travailler pour Guinness tout en publiant sous le pseudonyme 'A Student' nombre d'articles scientifiques qui ont construit les outils utilisés tous les jours par les statisticiens et économètres.

Si ce genre d'histoire vous amuse, je vous invite à lire [The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century](#) de David Salsburg", qui retrace les temps héroïques de la statistique et dans lequel on apprend que Karl Pearson était un affreux bonhomme, que Ronald Fisher a passé une bonne partie de sa vie à montrer que la cigarette ne provoquait pas le cancer du poumon etc.

vous pouvez aussi (re)lire l'article de Student-Gosset [the probable error of the mean](#) Biometrika, vol 6, n°1, pp 1–25, March 1908

Une analyse critique des travaux de Fisher sur la causalité cigarette-cancer du poumon peut être [consultée ici](#)
Paul D. Stolley, "When Genius Errs: R. A. Fisher and the Lung Cancer Controversy", American Journal of Epidemiology, vol 33, n°5, pp 416-425

** Vous ignorez peut être ce que sont les stencils... dans les années 70, les photocopieuses n'étaient pas très répandues... les documents étaient alors dupliqués au moyen de stencils (ou duplication à l'alcool - pour plus de détails, voir [cette page wikipedia](#))

Les beaux mercredis sont en vacances! Un peu de "magie" en attendant ?

Problème mis en ligne le 12/07/2012 – solution mise en ligne le 5/09/2012

Comme vous l'indique le titre de ce post, les "beaux mercredis" sont en vacances (et moi aussi très prochainement...).

Retour des petits problèmes (que j'espère amusants...) début septembre !

En attendant, je vais vous faire un petit cadeau... un peu de magie pour illuminer vos vacances !

[un peu de magie \(en français\)](#)

[same magic thing... in english](#)

Bien entendu, c'est un programme SAS qui crée ce "tour de magie"... (Deux programmes en fait puisqu'il y a une version française et une version anglaise...)

Vous souhaitez consulter le programme ?

Mais vous oubliez qu'un magicien ne révèle jamais ses trucs !

(en tout cas, pas les mois de juillet et d'aout...pour vous laisser le temps de réfléchir à comment ce tour a pu être créé...)

bonnes vacances

Sébastien Ringuédé

PS : bien entendu, en cas "d'urgence", le forum de discussions reste ouvert !