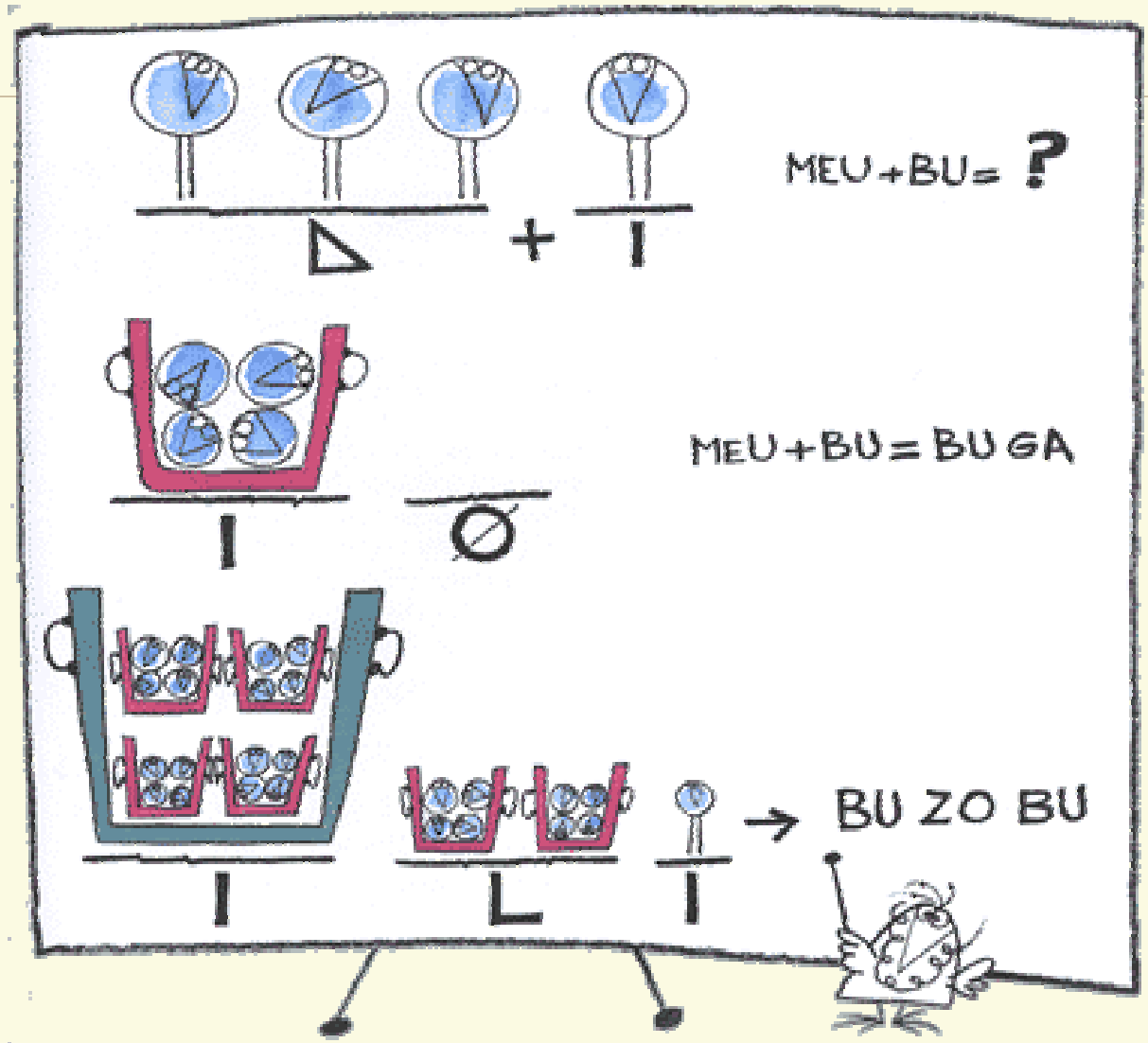
STATISTICIENS: ATTENTION, LOGICIELS !

JMS 16-17 décembre 2002

Dominique Ladiray

Statistique Canada


MEU+BU= ?
MEU+BU= BU GA
BU ZO BU

Introduction

- Critères de choix d'un logiciel statistique:
  - Fonctionnalités
  - Ergonomie, « jolis graphiques »
  - Documentation, Formation
  - Maintenance, Portabilité
  - Prix, etc.
- Et la précision des calculs ?
  - « *les entreprises de logiciels statistiques font subir des tests intensifs à leurs produits pour s'assurer que les algorithmes mis en œuvre font des calculs précis* » (SAS)

Autrefois ….

- On n’y croyais pas !
  - Jeux d’essai : Longley (1967), Wampler (1970), Wilkinson (1985)
  - Et pourtant contraintes : mémoire, temps de calcul
  - Donc : algorithmique, un statisticien derrière chaque logiciel ….
- Et aujourd’hui ? Nos logiciels sont-ils « bons » ?
- Regain d’intérêt pour les tests et jeux d’essai

Tester un logiciel

- ✓ Comme n'importe quel autre produit
- ✓ Jeux d'essai, faciles et difficiles, données réelles et/ou « étranges »
- ✓ Plusieurs aspects:
  - Estimations
  - Générateurs de nombres aléatoires,
  - Lois statistiques (tests, p-values)

Un exemple « simple »

- Le « vilain fichier » de Wilkinson

X	Big	Little	Huge	Tiny	Round
1	99999991	0.99999991	1E12	1E-12	0.5
2	99999992	0.99999992	2E12	2E-12	1.5
3	99999993	0.99999993	3E12	3E-12	2.5
4	99999994	0.99999994	4E12	4E-12	3.5
5	99999995	0.99999995	5E12	5E-12	4.5
6	99999996	0.99999996	6E12	6E-12	5.5
7	99999997	0.99999997	7E12	7E-12	6.5
8	99999998	0.99999998	8E12	8E-12	7.5
9	99999999	0.99999999	9E12	9E-12	8.5

- Toutes les variables se déduisent linéairement l’une de l’autre. Corrélations égales à 1 ?

SPAD 5

- Problèmes pour lire *Little* (enlever un 9)

	X	Big	Little modifiée	Huge	Tiny	Round
X	1.00					
Big	0.69	1.00				
Little modifiée	1.15	0.79	1.00			
Huge	1.00	0.69	1.15	1.00		
Tiny	1.00	0.69	1.15	1.00	1.00	
Round	1.00	0.69	1.15	1.00	1.00	1.00

- Beaucoup de programmes en simple précision auront ce type de problème

Méthodologie

- McCullough (1998)
- Jeux d'essai du NIST
  - 9 statistique descriptive, 11 analyse de variance, 11 régression linéaire, 27 régression non linéaire
  - Valeurs certifiées.
- Tests DIEHARD du RNG uniforme (Marsaglia)
  - 18 tests d'indépendance
- Tests ELV des distributions (Knüsel)

NIST régression : données de Longley

Y Emploi total	x1 Déflateur du PIB	x2 PIB	x3 Chômage	x4 Effectif des forces armées	x5 Population de 14 ans et plus	x6 Année
60323	83.0	234289	2356	1590	107608	1947
61122	88.5	259426	2325	1456	108632	1948
60171	88.2	258054	3682	1616	109773	1949
61187	89.5	284599	3351	1650	110929	1950
63221	96.2	328975	2099	3099	112075	1951
63639	98.1	346999	1932	3594	113270	1952
64989	99.0	365385	1870	3547	115094	1953
63761	100.0	363112	3578	3350	116219	1954
66019	101.2	397469	2904	3048	117388	1955
67857	104.6	419180	2822	2857	118734	1956
68169	108.4	442769	2936	2798	120445	1957
66513	110.8	444546	4681	2637	121950	1958
68655	112.6	482704	3813	2552	123366	1959
69564	114.2	502601	3931	2514	125368	1960
69331	115.7	518173	4806	2572	127852	1961
70551	116.9	554894	4007	2827	130081	1962

NIST régression : données de Longley

- Valeurs certifiées (15 chiffres)

Estimation des paramètres :

Paramètre	Valeur	Ecart-type
B0	-3482258.63459582	890420.383607373
B1	15.0618722713733	84.9149257747669
B2	-0.358191792925910E-01	0.334910077722432E-01
B3	-2.02022980381683	0.488399681651699
B4	-1.03322686717359	0.214274163161675
B5	-0.511041056535807E-01	0.226073200069370
B6	1829.15146461355	455.478499142212
Ecart-type des résidus	304.854073561965	
R2	0.995479004577296	

Tableau d'analyse de la variance:

	Degrés de liberté	Somme des carrés	Moyenne des carrés	Statistique F
Régression	6	184172401.944494	30695400.3240823	330.285339234588
Résidus	9	836424.055505915	92936.0061673238	

Méthodologie

- Valeurs certifiées: 15 chiffres pour problèmes linéaires, 11 chiffres pour non linéaires
- Indicateur du nombre de chiffres significatifs exacts :

$$LRE = -\log_{10}\left[\left|q-c\right|/\left|c\right|\right] \text{ si c est non nul et } LAR = -\log_{10}\left[\left|q\right|\right] \text{ sinon}$$

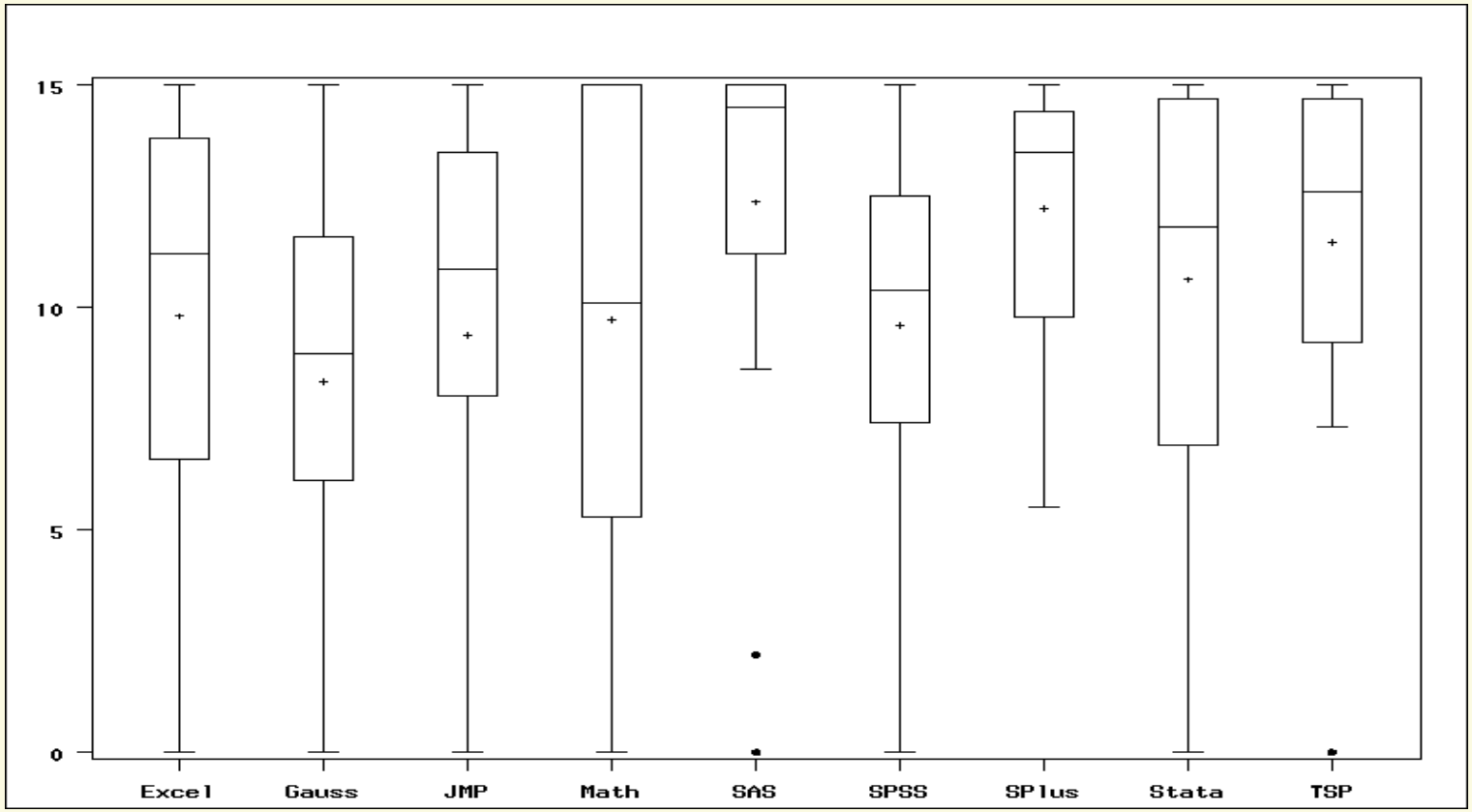
- Pour les problèmes de régression, on prend la valeur minimum du LRE pour les paramètres, les écart-types.
- Tests pour Excel, Gauss, JMP, SAS, SPSS, Splus, TSP et Mathematica (score parfait !)

Analyse de la variance

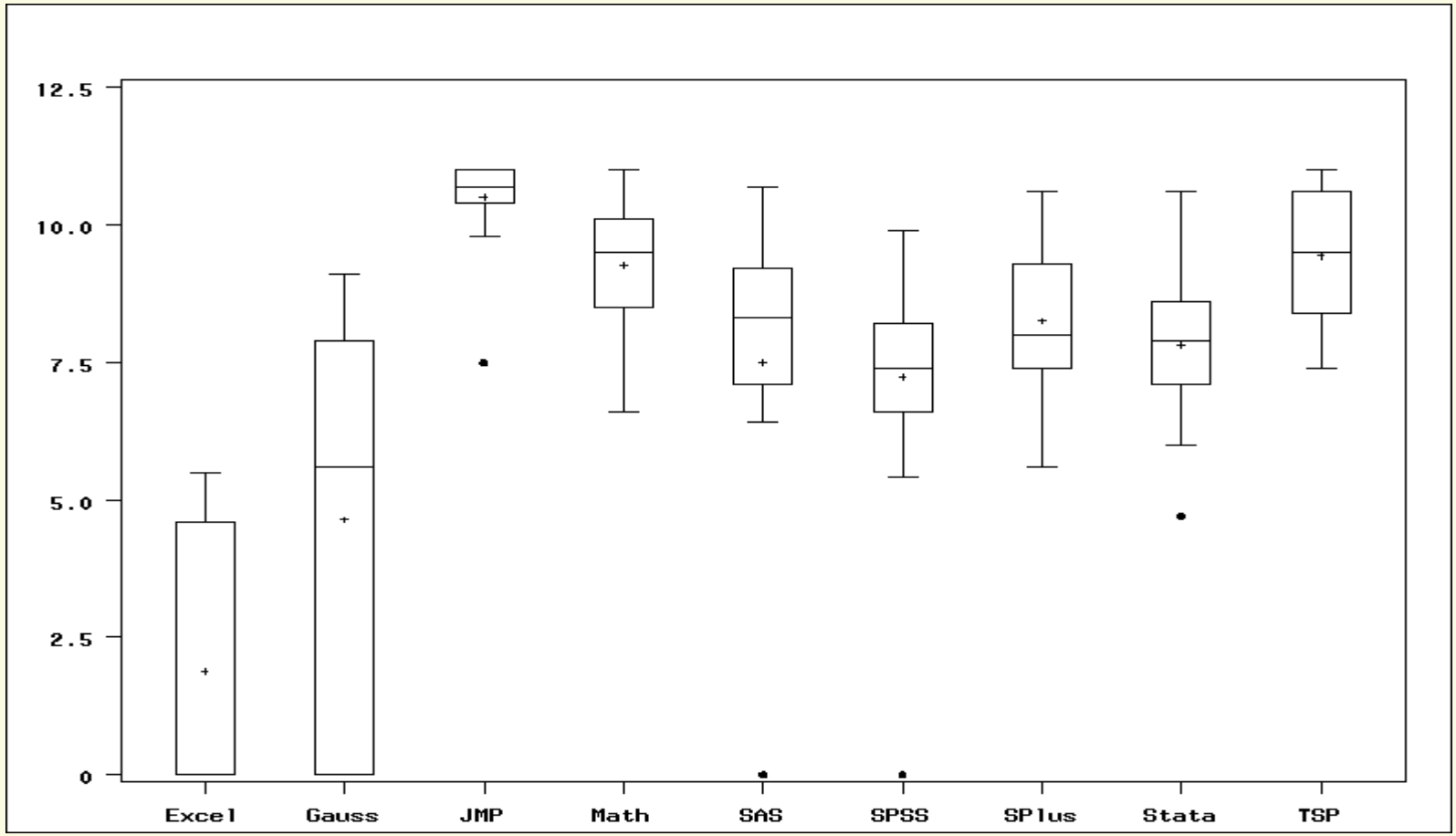
- Statistique de Fisher

Données	Logiciel	Excel XP	GAUSS 3.2.37	JMP 5.0	Mathematica 4.0 (B)	SAS 8.2	S-Plus 4.0	SPSS 7.5	Stata 6	TSP 4.4
SiRstv	Facile	8.5	12.4	12.4	15.0	12.7	13.3	9.6	13.1	13.1
SmLs01	Facile	14.3	14.5	14.0	15.0	15.0	14.5	15.0	14.4	14.6
SmLs02	Facile	12.5	14.1	13.4	15.0	13.9	14.3	15.0	13.3	14.7
SmLs03	Facile	12.6	12.7	12.4	15.0	12.7	12.9	12.7	14.7	12.3
AtmWtAg	Moyen	1.8	8.5	8.4	15.0	8.8	9.7	miss	10.2	10.2
SmLs04	Moyen	1.7	8.5	8.2	15.0	10.4	10.4	0.0	10.4	10.4
SmLs05	Moyen	1.1	8.3	8.0	15.0	10.2	10.2	0.0	10.2	10.2
SmLs06	Moyen	0	6.5	6.2	15.0	10.2	10.2	0.0	10.2	10.2
SmLs07	Difficile	0	2.7	2.4	15.0	4.4	4.6	0.0	4.4	4.6
SmLs08	Difficile	0	2.2	1.9	15.0	4.2	2.7	0.0	4.4	1.9
SmLs09	Difficile	0	0	0.3	15.0	4.2	0.0	0.0	4.2	0.8

Régression linéaire



Régression non linéaire



D’une commande à l’autre

- Différents algorithmes au sein d’un même logiciel. Ex : SAS

Données		REG		ORTHOREG	
		$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$	$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$
Norris	facile	12.3	11.8	11.9	13.8
Pontius	facile	11.5	8.6	12.1	12.3
NoInt1	moyen	14.7	14.0	14.7	15.0
NoInt2	moyen	15.0	15.0	15.0	14.6
Filip	difficile	0.0	0.0	0.0	0.0
Longley	difficile	8.6	10.3	13.6	14.6
Wampler1	difficile	6.6	15.0	10.2	15.0
Wampler2	difficile	9.6	15.0	13.2	15.0
Wampler3	difficile	6.6	11.2	9.8	13.6
Wampler4	difficile	6.6	11.2	8.1	13.6
Wampler5	difficile	6.6	11.2	6.1	13.6

Générateurs de Nombres Aléatoires

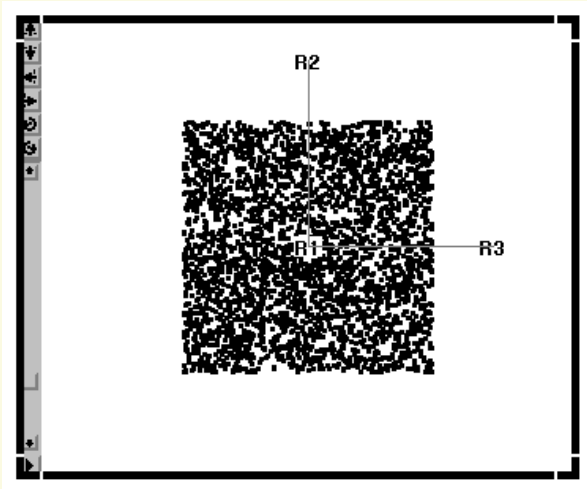
- ✓ Pièce essentielle : simulations, bootstrap etc.
- ✓ Exemple : étude McKinnon sur les tests de racine unité, 100 milliards de nombres aléatoires.
- ✓ Nécessite un RNG de période très grande
  - Knuth: n << p/1000 soit ici p > 10^{14}
- ✓ Tester le générateur « uniforme » suffit.
- ✓ Générateurs usuels : LCG

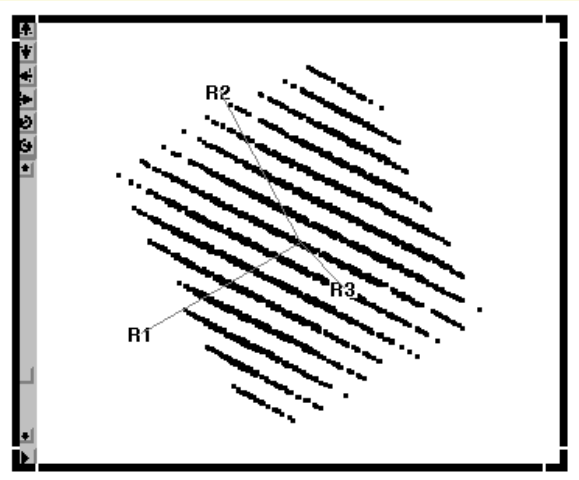
$$X_n \equiv aX_{n-1} + c \pmod{m}$$

- ✓ Période usuelle 2^{31}-1 ou 2^{32} (4.3 milliards)

Un problème usuel

- ✓ Exemple: le vieux (et pitoyable) RANDU





- ✓ Tester l'indépendance …
- ✓ DIEHARD de Marsaglia : 18 tests

Résultats des tests

p : pass, F : fail

Test	Excel	GAUSS	JMP	Mathematica	SAS	S-Plus	SPSS
Birthday Spacings Test	p	F	p	F	p	p	p
Overlapping 5-Permutation Test	p	F	p	p	p	p	p
Binary Rank For 31x31 Matrices	p	F	p	p	p	p	p
Binary Rank For 32x32 Matrices	p	F	p	p	p	p	p
Binary Rank For 6x8 Matrices	p	F	p	p	p	p	p
Bitstream Test (p values)	p	F	p	p	p	p	p
OPSO Test	F	p	p	p	p	p	p
OQSO Test	F	p	p	p	p	F	p
DNA Test	F	p	p	p	p	F	p
Count the Ones Test (stream of bytes)	F	p	p	p	F	F	F
Count the Ones Test (specific byte)	F	F	p	p	p	F	p
Parking Lot Test	p	F	p	p	p	p	p
Minimum Distance Test	p	F	p	p	p	p	p
3-D Spheres Test	F	F	p	p	p	p	p
Squeeze Test	F	p	p	p	p	p	p
Overlapping Sums Test	p	p	p	p	p	p	p
Runs Test	p	p	p	p	p	p	p
Craps Test	p	p	p	p	p	p	p
Echec	7	10	0	1	1	4	1

- ✓ JMP seul présente un très bon RNG, période 2^{19937} -1

Conclusions

- ✓ Algorithmes de qualité très variable
- ✓ Tests « faisables » : Mathematica réussit
- ✓ Résultats médiocres pour la version de GAUSS testée.
- ✓ Il est facile d'améliorer les algorithmes : ils existent !
- ✓ Et d'ailleurs ….. Les choses bougent

D’une version à l’autre

- ✓ Les fabricants font de plus en plus référence à ces jeux d’essais (JMP, Stata, TSP … SAS)
- ✓ Et améliorent leur produit. Ex: PROC ANOVA

		SAS 6.12	SAS 8.2
SiRstv	Facile	8.3	12.7
SmLs01	Facile	13.3	15.0
SmLs02	Facile	11.4	13.9
SmLs03	Facile	11.8	12.7
AtmWtAg	Moyen	0.9	8.8
SmLs04	Moyen	0.8	10.4
SmLs05	Moyen	0.0	10.2
SmLs06	Moyen	0.0	10.2
SmLs07	Difficile	0.0	4.4
SmLs08	Difficile	0.0	4.2
SmLs09	Difficile	0.0	4.2

Attention !!!!!

- ✓ Difficile (impossible) pour un amateur d'améliorer les algorithmes (« Add ins » : Nerlove, Vinod)
- ✓ Au mieux, on peut faire pire !

Données		REG		IML	
		$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$	$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$
Norris	facile	12.3	11.8	12.5	14.1
Pontius	facile	11.5	8.6	10.6	13.4
NoInt1	moyen	14.7	14.0	14.7	15.0
NoInt2	moyen	15.0	15.0	15.0	15.0
Filip	difficile	0.0	0.0	ns	ns
Longley	difficile	8.6	10.3	7.0	10.2
Wampler1	difficile	6.6	15.0	4.8	5.2
Wampler2	difficile	9.6	15.0	8.6	9.9
Wampler3	difficile	6.6	11.2	4.8	10.7
Wampler4	difficile	6.6	11.2	4.8	10.7
Wampler5	difficile	6.6	11.2	4.8	10.7