

REPORT OF ACTIVITIES

Statistical Methodology and Computing Service

1 September 2018 to 31 August 2019



2018-2019

Statistical Methodology and Computing Service
Report of Activities
1 September 2018 to 31 August 2019

Contents

Introduction	4
News	5
Organization	5
Personnel	5
Statistical Software	5
Active watch for tenders	5
People	6
Management and Administration	6
Consultants	6
Associated Consultants	7
Professors	7
Consultants at other institutes	7
Researchers and Research Assistants	7
Interns	7
Workshop Instructors	7
Training	8
Consulting	11
Servers and Statistical Software	13
Statistical Computing Server	13
UDS	13
Available Software	13
Usage Statistics	14
Help for students	14
Larger Projects	15
Survey Support	15
External projects	15
General interest Projects – Gender project	15
Research Project - Erasmus+OLS (2015-2021)	15
Research project - CluePoints (2018-2020)	15
Research project - First spin-off (2018-2020)	15
Research Project - Macular degeneration (2017-2019)	15
Project linked to education - Development of teaching materials	15
Course Module - FOREM (2011-...)	16
University Certificate - Junior Data Analyst (2016-...)	16
Summer School - Collaboration with the CENTAL	16
Your research project? SMCS can be your partner!	16

Contents

Appendix.....	17
List of Trainings and Seminars 2018-2019.....	17
Trainings	17
Publications and Presentations by SMCS members	18
Publications	18
Seminars and Presentations	19
Poster.....	19

Introduction

Introduction

The SMCS continues its activity as a diversified service provider in practical statistics. Our team of 5.5-6.5 full time equivalents of consulting staff and our network of more than 20 associated specialists assist UCLouvain researchers and external clients in data collection, processing, analysis, and interpretation. This report gives an overview of the evolution of the activities over the past three academic years focusing on 2018/19.

The consulting activity has been stable at 420-450 projects per year. The training activity grew to 52 courses, representing 172 days of training. These courses, focusing on practical statistics and the use of statistical software, were attended by about 750 participants (2425 person days). The statistical computing and software-sharing servers had about 50 registered users each.

As recommended by the users committee, the SMCS continues improving its visibility towards potential new users which as doctoral researchers graduate and new candidates start out. As every year, we distributed our newsletter explaining our goals, expertise, and services via the three sectors.

With respect to larger projects, the support of the language testing and training of the Erasmus+ Online Linguistic Support (OLS) continues and will last until 2021. The SMCS also continues working with the Croix-Rouge de Belgique on the quality assurance of blood products.

In Fall 2018, we taught the third edition of the new certificate “Junior Data Analyst. A modular version will be launched in January 2020. This version targets working participants and is organized on Fridays and Saturdays.

Our collaboration with the Walloon career center, FOREM, continues. For several years, we have been teaching an intensive course in biostatistics for people with degrees or experience in the biosciences in need of additional qualifications. In collaboration with the LSBA we are now working on converting this course into two university certificates.

In July 2019, the SMCS organized its first Summer School, a new type of training in theoretical and practical aspects of an advanced methodology given by invited specialists.

The first session on text mining with Python (in French) was organized in collaboration with the Cental (Centre de traitement automatique du langage, UCLouvain). The week before the Summer School, a new classical SMCS training ‘Introduction to Python’ was launched.

As part of a project coordinated by the MiiL and supported by a Wernaers fund, two members of our team participated in the creation of a video. It concerns the impact of digital communication technology on our society and will be made available to journalists and the general public on Youtube. For more details, please see at http://ezcast.uclouvain.be/ezplayer?action=view_album_assets&album=1080-Cequenditlarech-pub&token=VYMASFCI.

For 2019/20 we are considering several new initiatives. Ideas include the second edition of our Summer school focusing on new topics. This is under discussions with the Cental.

As planned, our team answered several new calls for tender and developed new collaborations with external partners as well as inside UCLouvain.

Since the evolution of a more data-driven research society will lead many research institutes to upgrade their data collection and analysis capabilities, we foresee a need to extend our capabilities in the area of data technology and science. Meeting this challenge must be part of our long-term plan such that we can provide appropriate guidance, advice and assistance. We will stay in contact with the different UCLouvain entities concerned by that question such as the UCLouvain committee in charge of DMP questions and will work with them.

A related important challenge of the coming years will be handling growth as more visibility and the natural increase in data driven research will generate more demands for our services. The group of consultants and the expert network will grow. Maintaining good integration and communication will therefore require close attention.

As usual, your input, assistance, support, and curiosity will be greatly appreciated.

Enjoy reading, Christian Ritter - President of the management committee.

News

News

Organization

The management committee (comité de gestion) met twice. The Fall meeting has focused on long term planning whereas the Spring meeting has concentrated on finances, project accounting, and short-term planning. A continuing problem is uniting all members on these two dates.

With respect to the user committee, the format of an afternoon with public presentations of case studies organized on the same day as the official meeting restricted to the official representatives has again been very well-received and will continue in 2020.

Personnel

The team is stable and has focused on managing all the projects and on preparing new initiatives such as the Summer school or the new JDA version (with modules).

Statistical Software

The Big5-Stat initiative is still operated by the central computing service (SGSI) and the SMCS only maintains the role of advisor and tester.

The SGSI has made the transition to UDS during the year. This new tool provides softwares through a kind of remote desktop connection and is very useful for softwares we do not use often or with shared costs. During the last year, extensive testing was implemented, and this tool was considered good enough to be on production this after. UDS appears to be very practical, easy to use and even used in some courses.

Active watch for tenders

Since September 2017, we have been actively monitoring research and service opportunities, and this has led us to respond to new kinds of calls and requests.

CHAFEA Consumer Market Studies

The SMCS is a subcontractor of the Ernst & Young-led consortium who was selected, among others, for the CHAFEA Consumer Market Studies framework. CHAFEA is the Consumers, Health, Agriculture and Food Executive

Agency of the European Union and its studies aim at better understanding consumer conditions and opinions in different product and service markets.

Asynchronous e-learning and e-examination

Since September 2018, SMCS has been involved as a co-promotor in a project financed for 2 years by a "Fonds de Développement Pédagogique". This project consists in developing tools and procedures for electronic and asynchronous formative and certificate evaluations of students in statistics.

Preparing students for higher education

In collaboration with the EPHEC, our team has received financial support for a project related to the preparation of new students for higher education. This project will consist in collecting and creating resources to help future students in Human Sciences in the Pole Louvain who start their first course in statistics in the best conditions. This project started in July 2019 for one year.

Gender statistics

Four people of our team have been involved in a project for the SPRB ("Service Public Régional de Bruxelles"). This project consisted in collecting gendered official statistics about the competencies of the SPRB and publishing them in a brochure, which is now available at <https://servicepublic.brussels/brochure-inegalites-de-genre/>.

People

People

Management and Administration

Director: Nathalie Lefèvre

President of Management Committee: Christian Ritter

Administrative coordinator: Nancy Guillaume

Communications and Quality Manager: Marie-Pascale Cosse (on leave)

Administrative support: Tatiana Regout

Consultants



Aurélié Bertrand (UCLouvain/LIDAM/SMCS – 1 FTE)

Main areas of activity: Humanities Sector (SSH), Medical Sector (SSS), specific projects in collaboration with the Institute of Statistics, Biostatistics and Actuarial Sciences.

Doctorat en Sciences – Statistique (UCL, 2017)

Master en statistiques, orientation générale, à finalité spécialisée (UCL, 2011)

Bachelier en Ingénieur de gestion (UCL, 2009)

Vincent Bremhorst (UCLouvain/LIDAM/SMCS – 1 FTE - since September 2019)

Main areas of activity: Humanities Sector (SSH), Science and Technology Sector (SST), bayesian support.

Doctorat en Sciences – Statistique (UCL, 2017)

Master en statistique (ULB, 2011)

Bachelier en mathématique (ULB, 2009)

Céline Bugli (UCLouvain/LIDAM/SMCS – 0.8 FTE)

Main areas of activity: Medical Sector (SSS), services for external clients.

Doctorat en Sciences – Statistique (UCL, 2006)

Master (DEA) en statistique (UCL, 2003)

Ingénieur Civil en Mathématiques Appliquées (UCL, 2001)

Lieven Desmet (UCLouvain/LIDAM/SMCS – 0.75 FTE)

Main area of activity: Health Sciences Sector (SSS).

Doctor of Sciences, orientation statistics (KUL, 2009)

Master in Mathematics (VUB 2004)

Master in Applied computer sciences (VUB, 1999)

Alain Guillet (UCLouvain/LIDAM/SMCS – 1 FTE)

Main areas of activity: Medical Sector (SSS), and statistical computing.

Master professionnel Sciences, Santé, Technologie, mention Mathématiques et Applications, spécialité Ingénierie Mathématique (Université de Nantes, 2005)

DEA en informatique (Université de Nantes, 2004)

Séverine Guisset (UCLouvain/LIDAM/SMCS – 1 FTE - since September 2017)

Main areas of activity: Humanities Sector (SSH), Medical Sector (SSS), survey support, services for external clients.

Master en statistique (UCL, 2017)

Master in European Studies (King's College London, 2007)

Diplôme de l'Institut d'Etudes Politiques de Strasbourg (Université Robert Schuman, 2006)

Nathalie Lefèvre (UCLouvain/LIDAM/SMCS – 0.5 FTE and UCLouvain/IPSY – 0.5 FTE)

Main areas of activity: Humanities Sector (SSH) and University Administration.

Master en Statistique (UCL, 2007)

Doctorat en Sciences psychologiques (UCL, 2003)

Licence en Psychologie et en Sciences de l'Education, orientation psychologie expérimentale (UCL, 1992).

Catherine Rasse (UCLouvain/LIDAM/SMCS – 0.8 FTE)

Main areas of activity: Science and Technology Sector (SST).

Master (DEA) en Statistique (Gembloux, 2007).

Ingénieur agronome, orientation Génie rural (Gembloux, 2001).

People

Associated Consultants

Associated consultants are professors or consultants who collaborate with the SMCS on trainings and consulting projects but who are not formally SMCS staff.

Professors

Michel Denuit (LIDAM/ISBA)
Sébastien de Valleriola (ICHEC and prof. Invité UCLouvain)
Thomas François ILC/PLIN)
Laurent Gatto (DDUV)
Bernadette Govaerts (LIDAM/ISBA)
Marie-Paule Kestemont (IACCHOS)
Vincent Lorant (IRSS)
Catherine Legrand (IDAM/ISBA)
Bruno Masquelier (IACCHOS)
Eugen Pircalabelu (LIDAM/ISBA)
Christian Ritter LIDAM/LSBA)
Bruno Schoumaker '(IACCHOS)
Nicolas Schtickzelle (ELI)
Isabelle Thomas (LIDAM/CORE)
Vincent Vandenberghe (LIDAM/IRES)

Consultants at other institutes

Jonathan Dedonder (IACCHOS)
Sébastien Fontenay and Adam Levai (LIDAM/IRES)
Nathalie Lefèvre (IPSY)
Jérôme Ambroise (IREC)
Aline van Maanen (Institut Roi Albert II - Cancérologie et Hématologie)

Researchers and Research Assistants

Some researchers and doctoral students of the ISBA participate in the consulting and training activities of the SMCS. This gives them an opportunity to gain valuable experience in practical statistics.

Interns

David Stern, Augustus-October 2018

Workshop Instructors

The following list gives the names of the people who contributed this year to the training and consulting activities of the SMCS. This includes SMCS and associated

consultants, other members of the ISBA, visitors, and external trainers.

Aurélie Bertrand (SMCS), Vincent Bremhorst (SMCS), Céline Bugli (SMCS), Jonathan Dedonder (IACCHOS), Sébastien Fontenay (IRES), Thomas François (ILC), Bernadette Govaerts (ISBA), Alain Guillet (SMCS), Séverine Guisset (SMCS), Alexandre Jacquemain (ISBA), Eric Lecoutre (WeLoveDataScience!), Nathalie Lefèvre (SMCS), Adam Levai (IRES), Rebecca Marion (ISBA), Manon Martin (ISBA), Gilles Mordant (ISBA), Hubert Naets (Cental), Catherine Rasse (SMCS), Christian Ritter (ISBA), Xavier Rouby (Pfease), Marco Saerens (LRIM), Antoine Soetewey (ISBA), Vincent Vandenberghe (IRES), Patrick Watrin (Cental), SAS Institute trainers.

Training

Training

Training is one of our main missions. In general, we focus on hands-on training in practical statistics and the use of statistical software in ways that are not covered by regular university courses. Four of the training sessions took place on the Woluwe campus. Six training courses have been given in English. We will continue to propose every year more training courses in English.

Five training sessions had to be cancelled because there were not enough participants (two of them were planned in Woluwe). A solution of personalized coaching has been proposed to the participants who had registered. Based on this year experience, some of our training courses will only be proposed upon specific request.

For very advanced trainings, participants generally prefer that they be given in their mother tongue.

Ten new training programs have been developed.

The SMCS offers the following types of training:

- **Regular SMCS trainings:** 2-3-day courses (once or twice per year) on commonly used statistical and associated software at UCLouvain (SPSS, R, SAS, SAS Enterprise Guide, JMP, Stata, NVivo, Excel, Limesurvey).

The SMCS offers many trainings about R, some as an initiation to the R language and others more advanced. Some trainings are organized to respond to users' requests. Some trainings organized to respond to user requests thereafter become Regular SMCS trainings. For example, the R training, "*Pratique de la statistique avec R*".

- **Customized trainings for specific audiences** (UCLouvain and external). Examples: "*Introduction aux modèles multi-niveaux avec SPSS*" or "*Introduction to the basic concepts of Bayesian statistics*" (organized in collaboration with IPSY but open to everyone), "*Créez votre liste de randomisation avec le logiciel JMP*" (in collaboration with IREC but open to everybody).
- **In partnership: courses taught in collaboration with SAS.** The courses "*Formation SAS : programmation de*

base" and "*SAS advanced programming*" belong to this category and were partially subsidized until end of the academic year 2018-2019.

- **University certificate.** In collaboration with the LSBA, the SMCS can also be involved in and/or coordinate university certificates. Example: *the third edition of the certificate "Junior Data Analyst"* - Teaching this course implied working with several external experts active in the data science business and we learned a lot about current practices.
- **Summer School:** Summer school. A one-week in-depth training on an advanced topic given by invited speakers who are specialists in the field. The first edition took place in July 2019. It was about Text mining with the Python language (in French) and was organized in collaboration with the Cental (Centre de traitement automatique du langage, UCLouvain). The specificity of the program was to propose, besides courses, some social activities which have been greatly appreciated by the participants.
- **Seminars and Lunch sessions (Midis du SMCS):** short (1-3h) informal demonstrations of statistical methods or software of particular practical interest.

Table 1 shows that the SMCS organized 52 courses in the academic year 2018-2019 for 172 teaching days. This stays stable compared to the previous year (small increase of teaching days).

Table 1. Global statistics on SMCS trainings.

	2016/17	2017/18	2018/19
Courses	40	57	52
Course Days	135	168.5	172
Participants			
UCLouvain	537	660	493
Person-Days*			
UCLouvain	1199	1517	1425
Participants Ext	128	224	253
Person-Days* Ext	680	737.5	1000

*Total effective number of training days: for instance, 9 people attending 2 training days is equivalent to 18 person-days.

Training

Table 2 shows the number of participants from the three sectors. Table 3, Table 4 and Table 5 show the origin of participants from respectively SSH, SSS and SST sectors.

Table 2. Number of participants per sector

	2016/17	2017/18	2018/19
SSH	173	277	177
SSS	122	130	97
SST	122	90	54
Students	120	159	165
Central services		4	

The observed fluctuations between years and sectors reflect natural variation and the organization of special events. For example, the spike for SST in 2016/17 corresponds to a special session “*La statistique, outil du chercheur et de l'ingénieur* » organized in April 2017.

Compared to previous years, there is a decrease in all sectors. This can be explained by the fact that training courses (1/2 day format) have been organized in 2017/18 such as Text mining, Life Trajectory and Bayesian Statistics. These were some kind of specific introduction seminars organized in large rooms (with no computer) allowing to accept a larger number of participants. Another reason to explain the decrease is that last year, we had received some specific requests from sectors (especially SSH and more particularly IPSY, PSP, IACS and ESPO) to organized specific training courses and the attendance had been very high.

The statistics training program reaches the entire SSH sector. The high participation rate of IPSY-PSP stems from the fact that the IPSY Institute finances the organization by the SMCS of specific trainings. The high participation of IACS-SPLE-ESPO and LIDAM are explained by specific seminars or training requested by these entities. The concerning trainings are often organized in collaboration with associated consultants coming from these entities. However, these trainings remain open to all.

Table 3. Origin of participants from the humanities sector (SSH)

	2016/17	2017/18	2018/19
SSH	214	321	247
IACS-SPLE-ESPO	61	87	54
ILC	7	19	18
LOURIM-LSM	16	15	13
LIDAM	8	40	37
INCAL-FIAL	5	10	5
IPSY-PSP	71	98	48
JURI-DRT	5	3	1
RSCS-TECO		5	1
Students	41	44	70

In the health sciences sector courses were also followed by participants from a wide range of institutes. Many clinical researchers find it difficult to attend longer trainings due to their appointment schedules at the university clinics. We updated the training format to take this particularity into account.

Table 4. Origin of participants from the health sciences sector (SSS)

	2016/17	2017/18	2018/19
SSS	133	155	112
ASS	1		
DDUV-Ludwig Inst.	9	13	5
IONS	21	23	13
IREC-MEDE-FSM	36	40	44
IRSS-FSP	28	42	17
LDRI-FASB	20	10	14
Clinics	7	2	4
Students	11	25	15

Training

Table 5. Origin of participants from the science and technology sector (SST)

	2016/17	2017/18	2018/19
SST	190	180	134
AGRO	7	9	6
ELI	39	32	26
ICTEAM-EPL	9	5	2
IMCN	2	1	
IMMC	32	7	
IRMP	9	5	1
LIBST	16	20	17
LOCI	6	7	
SC	2	4	2
Students	68	90	80

The SMCS is also regularly invited to sessions organized by other entities and these presentations are not included in the training statistics.

Overall, the training activity of the SMCS occupies about 3.1 full time staff equivalents (FTE). 1.8 FTE of this effort is directly oriented towards the UCLouvain research and teaching units, the remainder towards external clients, such as the FOREM and some private companies and public administrations.

The trainings given to external clients fulfill three objectives. First, they are a part of our mission of service to society. Second, they allow us to develop new course modules for internal use, such as the FOREM training which allowed us to bring together an entire course portfolio in the area of bioinformatics. These course modules can then be offered to members of the life and health sciences groups in SST, SSH or SSS. Third, but not least, they allow us to generate financial resources to have enough funds for our staff members not financed by UCLouvain. We of course have to be cautious in the choice and management of such trainings to make sure there is a positive return (financial and others).

Finally, members of the SMCS have made a presentation in the framework of the Applied statistical workshop. They presented a specific methodology but also their work as a consultant in some specific projects.

Consulting

Consulting

Consulting is another of our main missions. We help researchers, administrators, and external clients in collecting meaningful data and in interpreting them.

Whenever possible, we provide methodological support to enable our clients to solve their problems by themselves. Our support can just be a piece of advice or short-term assistance. This type of consultancy is pre-paid by the financial contribution of the sectors up to 1 FTE and requires about 1.7 FTE.

On occasion, the complexity or the amount of the statistical work requires that the SMCS consultant spend a larger amount of time on a project. The consultancy thereby evolves into a “*collaboration*”. This requires additional funding by the client. Moreover, if the input of the statistical consultant yields new insight, common publishing rules require that he/she should become a co-author of publications based on the work. This type of support was particularly important this year and corresponds to more or less 2200 hours (800 hours for external projects and 1400 for internal one’s).

Overall, the consulting activity of the SMCS occupies about 3.8 full time staff equivalents (FTE). 3 FTE of this effort is directly oriented towards the UCLouvain research, teaching and administrative units, the remainder towards external clients.

Table 6. Origin of consulting projects

Client split	2016/17	2017/18	2018/19
Research	384	401	425
Educational project		2	9
Admin. services	5	8	9
External clients	11	9	10
Total	400	420	453

From 1 September 2018 to 31 August 2019, the SMCS worked on 453 projects of which 443 came from UCLouvain. As expected, most of the projects are linked to university research. Some larger projects involving the university administration or the schools or institutes and

external clients are presented in the Larger Projects section.

The following tables present the breakdown of the project origins in the three university sectors. They clearly show that the SMCS services reach almost all areas of research at UCLouvain. For the third year, we include in our statistics a cross-disciplinary support to each sector. This consists in statistical computing support for all the sector community. Even if this type of support counts for one unit in the table, this can sometimes represent up to 50 hours of work. This support consisted this year in the information collection, identification, implementation and sharing of good data protection practices (a GDPR project).

Table 7 (SSH) shows that the number of requests from the humanities sector usually fluctuating around 130 requests per year, increased this year to nearly 170. As in previous years, the institute and the school of psychology had by large the highest share. Therefore these units have a 0.5 FTE statistical advisor (Nathalie Lefèvre, who in the remainder of her time acts as the director of the SMCS). AIACCHOS and IRES/ECON also have an internal specialist for data questions matters (Jonathan Dedonder for IACCHOS, Sébastien Fontenay and Adam Levai for IRES/ECON), who are also associated members of the SMCS. Note that Sébastien and Adam shared their time half for master students to help them get data and analyze them and the other half is for PhD students and professors. Besides IPSY with particularly high activity, requests came from almost all units of this sector.

Table 7. Origin of consulting requests from the humanities sector (SSH)

	2016/17	2017/18	2018/19
SSH	133	133	169
Cross-disciplinary	1	1	2
IACCHOS-ISPOLE-ESPO	12	18	16
ILC-INCAL-FIAL	6	20	16
LOURIM-LSM	9	7	14
LIDAM	10	12	26
IPSY-PSP	94	73	93
JURI-DRT	1	2	2

Consulting

Table 8 (SSS) shows that on average there are 150 requests per year from the health sciences sector, with an increase this year to nearly 170 files. The areas related to clinical research, Health and Society Institute and the medical school account for about 60% of this, other requests are split among the remaining institutes. The number of requests increase in IRSS probably thanks to the presence of two of our consultants regularly in the institute.

Table 8. Origin of consulting requests from the health sciences sector (SSS)

	2016/17	2017/18	2018/19
SSS	150	151	167
Cross-disciplinary	1	1	1
DDUV-Ludwig Inst.	10	8	7
FASB	6	3	3
IONS	25	26	18
IREC-MEDE-FSM	92	63	87
IRSS-FSP	8	30	39
LDRI	5	8	3
Clinics - Albert II	3	12	9

Table 9 (SST) shows that the number of requests from the science and technology sector fluctuates around 100-130 requests per year, with a decrease this year to nearly 100 files. Requests came from all units of this sector.

Table 9. Origin of consulting requests from the science and technology sector (SST)

	2016/17	2017/18	2018/19
SST	101	117	98
Cross-disciplinary	1	2	2
AGRO	9	7	10
ELI	37	36	28
ICTEAM-EPL	3	3	5
IMCN			1
IMMC	8	14	7
LIBST	35	45	36
LIDAM			1
LOCI	4	4	4
SC	3	6	4
Commission doctorale	1		

Our main consultant for the science and technology sector (SST), Catherine Rasse, also continues to be a member of an ethics committee related to experimentation on animals.

Finally, **nine projects** also originated from the central administration (AC, LLL, SGSI, BIUL).

The statistics shown above do not present in details large-scale internal and external projects documented in a later section of this report.

Computing

Servers and Statistical Software

The SMCS now has two servers dedicated to statistical computing and no software-sharing server since UDS replaces it (see below). Moreover, the SMCS works together with the SGSI and the CISM on several activities to reduce barriers to using appropriate statistical software for university members. The computing activities require about 0.5 FTE.

Statistical Computing Server

We renewed the system four years ago and it will be replaced the next academic year. Our current two performance servers are based on a dual Intel Xeon E5-2667 processor with 128 GB of RAM and an SSD disk for the SAS temporary folder. The CISM (UCLouvain's high performance computing platform) hosts this server system and its storage and is in charge of the operating system and the user management. The SMCS takes care of the statistical software on these machines. Currently, SAS, R and Stata-MP are available.

The statistical computing server has about 65 users. People from LIDAM and IRSS are the main users of this service (the table on the following page provides more details about the users of the two servers). We can divide users in two categories: people who want to use SAS for free in the Big5-STAT project and people in need of a high-performance computer mainly for the three installed software packages (large RAM and multi-processors version). The average CPU charge was 21 % on smcs3 and 6 % on smcs4 with spikes (100% for smcs3 and 100% for smcs4) occurring when users execute paralleled programs in Stata or R.

More data analysts at UCLouvain should consider using this server system. It offers more performance, storage, and data security than local PC-based systems and is a good first step before using bigger HPC infrastructures like Manneback or Lemaitre3.

UDS

UDS is a service provided by the SGSI that allow users who asked for an access to use software without having it on their computer. It works through a web browser. Note that Ucilla disappeared and is now fully replaced by UDS.

It is possible to add software to UDS. Who pays for it will depend on many criteria so you should send an email to 8282@uclouvain.be to have more details.

Available Software

Via the Big5-STAT initiative, any UCLouvain member can request the installation of or the access to one or several of the five most used statistical software packages, R, JMP, SAS (via Enterprise Guide), SPSS, and Stata. The request should be made via the central computing Service desk a.k.a. 8282 or even better via the "Catalogue de service". Many UCLouvain members used this service but there still are too many PhD students who don't know about it and buy a software they could get for free.

High performance solutions in SAS, R and Stata (multi-processor) are available on the two statistical computing servers.

Finally, UDS already offers the Big5-STAT software and gives an access to HLM, MPlus, PASS and Simca. The SGSI paid all the software but Simca that was paid by the professor B. Govaerts.

The access conditions and procedures for the statistical computing server and the software-sharing server are available under the Servers tab of our [//www.uclouvain.be/smcs](http://www.uclouvain.be/smcs) or you can ask for them at the Service desk.

Computing

Usage Statistics

Table 10. Users of the statistical computing server

	2016/17	2017/18	2018/19
Total	54	52	65
SSH	30	30	35
IACS-SPLE-ESPO	4	3	8
ILC	1	2	1
LIDAM	23	24	24
IPSY-PSP	2	1	2
SSS	17	14	12
IONS	2		
IREC-MEDE	1		1
IRSS-FSP	13	14	10
LDRI	1		1
SST	7	8	18
ELI	4	6	5
ICTEAM-EPL	1	1	1
IMCN			1
IMMC			5
IRMP			1
LIBST			1
SC	2	1	4

Table 11. Users of the software-sharing server (UDS)

	2016/17	2017/18	2018/19
Total	25	50	27
SSH	19	25	17
IACCHOS	1		3
ILC		1	
LIDAM	9	9	4
IPSY-PSP	9	12	8
LOURIM-LSM		3	2
SSS	2	17	10
DDUV		4	
IONS-FSM	1	2	4
IREC	1	5	4
IRSS		2	1
LDRI		4	1
SST	4	8	0
AGRO	1	2	
ICTEAM			
IMMC		1	
LIBST		2	
SC	3	3	

Usage of UDS

The most used software was Mplus (81 connections from 10 different users), followed by PASS (61 connections from 12 users), SPSS Premium (12 connections from 6

users) and HLM (9 connections from 3 users) during the year. Surprisingly, no user came from SST whereas they need a software like PASS to justify the sample size of their experiment when they enter a new project that the Ethical committee needs to validate. More generally, it is surprising to observe a decrease of PASS since researchers need more and more to justify the chosen sample size.

Help for students

Students sometimes need help installing a statistical software package on their own computers to prepare exercises or to work on a project. The SMCS helps them to do so since the course assistants or professors do not always know what went wrong during the installation.

In the academic year 2018-2019, the platform helped 54 students, half of last year, of which 28 came for assistance with SPSS. We suspect the decrease of need for help comes from a bigger use of UDS among students since they all have access for free to it without asking and that this doesn't require anything but an internet browser.

Larger Projects

Larger Projects

Survey Support

Free survey support consulting was provided for a wide range of applications in 2018-2019 and included advice on survey and questionnaire design, programming in LimeSurvey, sampling, data management, weighting and analysis.

A questionnaire design project for Bruxelles-Formation was completed in 2018-2019, while a weighting tool was developed for a MiiL project for the RTBF.

Two to three trainings sessions linked to surveys are given as standard each year, with additional training sessions tailored to specific requests or audiences.

In 2020, the feasibility of a database gathering information on surveys conducted by UCLouvain will be assessed, in partnership with several academics with an interest in surveys. The main objective is to provide a clearer picture of survey activity within the university, in order to foster information and best practices exchange, as well as potential partnerships.

External projects

The monitoring of public tenders linked to statistics and statistical training continued on 2018-2019 but did not lead to the submission of major tenders. This work will be pursued in 2019-2020.

General interest Projects – Gender project

We completed one general interest project, on gender statistics for the SPRB. This led to the creation and publication of a publicly available brochure on gender statistics for Brussels: <http://equal.brussels/brochure-genre-2>

This type of project helps strengthen relationships with other entities, brings forward internal expertise and allows SMCS consultants to develop new skills while providing additional budget to cover staff costs.

Research Project - Erasmus+OLS (2015-2021)

The European project Erasmus+OLS is a seven-year project to provide online language support to several hundred thousand participants in Erasmus+ programs, via a language assessment and an online course.

Our role in this project is to provide statistical expertise, in particular to monitor the assessment, and design and analyse surveys. We keep on using and refining text mining methods to analyze open survey answers.

Research project - CluePoints (2018-2020)

In collaboration with CluePoints' R&D department, the SMCS is involved in activities of research and validation of solutions for Central Statistical Monitoring of Clinical Trials.

Research project - First spin-off (2018-2020)

The SMCS helped a research unit from IREC with the elaboration of a first UCLouvain spin-off funding application (Service public de Wallonie). The SMCS is responsible for programming an algorithm for quantifying a new biomarker.

Research Project - Macular degeneration (2017-2019)

The participation in a research consortium on Macular degeneration ends this year.

Project linked to education - Development of teaching materials

Since September 2018, SMCS has been involved as a co-producer in a project financed for 2 years by a "Fonds de Développement Pédagogique" (project promoter: Prof. Marie-Paule Kestemont). As part of this project, our platform dedicated a 33% EFT throughout the academic year to the coordination and development of pedagogical materials for statistical learning. This project is an example of the projects carried out by our platform to support statistical education activities. In the same vein, in July 2019, we have started a project to develop a kit of

Larger Projects

activities in statistics in preparation for higher education. We also participated in the production of a popular video on statistics in the digital world. We were also asked by LLL to reflect with them on alternatives to the negative points in the MCQs to be proposed to UCLouvain teachers.

Course Module - FOREM (2011-...)

Since 2011/12, the SMCS provides a course in bioinformatics/biostatistics for the FOREM. This course is aimed at people who have bio-sciences degrees and difficulties finding a suitable job. The course volume is 33 days (corresponding to more than 200 hours of lectures and supervised exercise sessions).

The results are quite encouraging as most of the former participants quickly found work suited to their abilities and training. Since the volume of this course is substantial, we will submit a proposal to associate it with an accredited certificate.

University Certificate - Junior Data Analyst (2016-...)

In collaboration with the LSBA, the SMCS has launched a new certificate called "Junior Data Analyst" in 2016, an intensive course for graduates in humanities in practical data management, analysis, and interpretation. The courses were given for the third edition from October 2018 to December 2018 and is now self-financing. The fourth edition will start in October 2019 and we have the project to launch beginning of 2020 a modular JDA program in which participants can choose to follow one or more modules (up to five). This modular program is more compatible with a professional activity because half of the courses are organised on Saturdays.

Summer School - Collaboration with the CENTAL

The SMCS has launched a new type of training, a Summer school, organized in collaboration with trainers of the Cental (Centre de traitement automatique du langage), UCLouvain. The idea is to focus on an advanced topic which is not a core competence of the SMCS and to invite specialists in the field. The objective is to organize such an event once a year, in July. The topic of the first edition,

which took place in July 2019, was Text mining with Python.

Your research projects? SMCS can be your partner!

As you can see from this report, we are ready to become partners of research projects and grant applications. Do not hesitate to contact us.

We can provide an early estimation of the amount of statistical work needed in your project, plan important stages of this work, and give an outline of the statistical methodology which would be employed.

The advantages of involving us in the early stages of a project include a more thorough study design, efficient allocation of resources and methodologies, and data that are easier to analyze and interpret.

The benefits are higher research efficiency and more credibility towards the granting authorities and scientific journals.

Appendix

Appendix

List of Trainings and Seminars 2018-2019

Trainings

- **NEW** - Analyse avancée de données en évaluation de politiques publiques, du 3 septembre 2018 au 4 septembre 2018 (Vincent Vandenberghe)
- Introduction to R language, du 3 septembre 2018 au 5 septembre 2018 (Alain Guillet)
- Analyse des données de panel avec Stata, du 10 septembre 2018 au 11 septembre 2018 (Sébastien Fontenay)
- Gestion et analyse de données avec Stata, du 10 septembre 2018 au 11 septembre 2018 (Alain Guillet)
- Pratique de la statistique avec SPSS, du 10 septembre 2018 au 14 septembre 2018 (Nathalie Lefèvre)
- Initiation à la statistique avec applications en JMP (Formation intégrée dans un programme intensif plus large du FOREM), du 18 septembre 2018 au 25 septembre 2018 (Catherine Rasse, Céline Bugli)
- Collecter des données d'enquête sur tablette et smartphone, du 19 septembre 2018 au 20 septembre 2018 (Jonathan Dedonder)
- Introduction au calcul de taille d'échantillon, du 28 septembre 2018 au 28 septembre 2018 (Catherine Rasse, Céline Bugli)
- Certificat universitaire en Junior Data Analyst, du 1 octobre 2018 au 27 novembre 2018 (Aurélie Bertrand, Vincent Bremhorst, Jonathan Dedonder, Alain Guillet, Séverine Guisset, Eric Lecoutre, Nathalie Lefèvre, Catherine Rasse, Christian Ritter, Xavier Rouby, Marco Saerens et Antoine Soetewey)
- Initiation au logiciel NVivo, du 8 octobre 2018 au 10 octobre 2018 (Jonathan Dedonder)
- Cours d'initiation au langage R, du 15 octobre 2018 au 18 octobre 2018 (Vincent Bremhorst)
- Data management and analysis using Stata, du 22 octobre 2018 au 26 octobre 2018 (Alain Guillet, Adam Levai)
- **NEW** - Analyses multivariées avec R, du 5 novembre 2018 au 8 novembre 2018 (Vincent Bremhorst, Séverine Guisset)
- Initiation à la statistique avec applications en JMP (Formation intégrée dans un programme intensif plus large du FOREM), du 6 novembre 2018 au 9 novembre 2018 (Catherine Rasse, Céline Bugli)
- Pratique de la statistique avec SPSS, du 9 novembre 2018 au 30 novembre 2018 (Nathalie Lefèvre)
- Formation SAS : programmation de base, du 13 novembre 2018 au 20 novembre 2018 (Formateur SAS, Céline Bugli)
- Préparation et mise en place d'une enquête, du 26 novembre 2018 au 27 novembre 2018 (Jonathan Dedonder, Séverine Guisset)
- Analyse de données pré-cliniques et cliniques avec SPSS, du 30 novembre 2018 au 14 décembre 2018 (Aurélie Bertrand, Nathalie Lefèvre)
- Publier votre travail dans un package R, du 3 décembre 2018 au 4 décembre 2018 (Manon Martin, Vincent Bremhorst)
- Introduction à la réalisation d'enquêtes en ligne avec LimeSurvey, du 10 décembre 2018 au 11 décembre 2018 (Jonathan Dedonder, Nathalie Lefèvre)
- **NEW** - Analyse de données Omics sur base de techniques multivariées avec R, du 10 décembre 2018 au 12 décembre 2018 (Céline Bugli, Lieven Desmet)
- Analyser la médiation, la modération et la médiation modérée avec SPSS, du 18 décembre 2018 au 19 décembre 2018 (Vincent Bremhorst, Aurélie Bertrand)
- Pratique de la statistique avec SPSS, du 7 janvier 2019 au 11 janvier 2019 (Nathalie Lefèvre)
- Plans d'expérience avec JMP, du 10 janvier 2019 au 14 janvier 2019 (Catherine Rasse)
- Pratique de la statistique avec R, du 14 janvier 2019 au 18 janvier 2019 (Vincent Bremhorst)

Appendix

- Panel-data Analysis using Stata, du 21 janvier 2019 au 22 janvier 2019 (Alain Guillet)
- Introduction aux modèles multi-niveaux avec SPSS, du 21 janvier 2019 au 25 janvier 2019 (Aurélie Bertrand, Nathalie Lefèvre)
- Analyses multivariées en RCommander (FactoMineR), du 24 janvier 2019 au 25 janvier 2019 (Séverine Guisset)
- Initiation au logiciel NVivo, du 28 janvier 2019 au 30 janvier 2019 (Jonathan Dedonder)
- **NEW** - Formation aux outils d'analyse statistique en épidémiologie avec SPSS, du 29 janvier 2019 au 30 janvier 2019 (Nathalie Lefèvre)
- Formation en analyse de données et statistiques pour la recherche et le développement dans la bioindustrie, du 11 février 2019 au 26 juin 2019 (Aurélie Bertrand, Céline Bugli, Lieven Desmet, Bernadette Govaerts, Alain Guillet, Manon Martin, Catherine Rasse)
- Introduction to R language, du 12 février 2019 au 19 février 2019 (Lieven Desmet, Hassana Al hassan)
- Pratique de la statistique avec SAS, du 15 février 2019 au 22 février 2019 (Catherine Rasse, Céline Bugli)
- Formation SAS : SAS advanced programming, du 15 février 2019 au 26 avril 2019 (Formateur SAS)
- Analyses multivariées avec SPSS, du 25 février 2019 au 1 mars 2019 (Séverine Guisset, Nathalie Lefèvre)
- **NEW** - Concepts de base et pratique de la statistique bayésienne avec R et JASP, du 6 mars 2019 au 8 mars 2019 (Vincent Bremhorst)
- Initiation aux modèles d'équations structurelles (SEM) avec Stata, du 11 mars 2019 au 15 mars 2019 (Aurélie Bertrand)
- Inférence et modélisation statistique avec JMP, du 20 mars 2019 au 27 mars 2019 (Lieven Desmet)
- Gestion et analyse de données avec Stata, du 8 avril 2019 au 12 avril 2019 (Sébastien Fontenay, Alain Guillet)
- Utilisation avancée du logiciel Nvivo, du 15 avril 2019 au 18 avril 2019 (Jonathan Dedonder)
- Initiation aux modèles d'équations structurelles (SEM) avec Stata, du 6 mai 2019 au 7 mai 2019 (Aurélie Bertrand)
- **NEW** - Panel-data Analysis using Stata, du 9 mai 2019 au 10 mai 2019 (Sébastien Fontenay)
- Plans d'expérience avec JMP, du 9 mai 2019 au 14 mai 2019 (Catherine Rasse)
- **NEW** - Panel-data Analysis using Stata (advanced part), du 16 mai 2019 au 17 mai 2019 (Adam Levai)
- Pratique de la statistique avec R, du 20 mai 2019 au 24 mai 2019 (Antoine Soetewey)
- Cours d'initiation au langage R, du 3 juin 2019 au 7 juin 2019 (Alain Guillet)
- Introduction à l'analyse de Réseaux Sociaux avec Gephi, du 11 juin 2019 au 13 juin 2019 (Jonathan Dedonder)
- Introduction aux modèles mixtes avec R, du 17 juin 2019 au 21 juin 2019 (Alain Guillet)
- Plans d'expérience avec JMP, du 17 juin 2019 au 21 juin 2019 (Catherine Rasse)
- **NEW** - Plans d'expérience : mise en pratique, du 25 juin 2019 au 25 juin 2019 (Catherine Rasse)
- **NEW** - Initiation au langage Python, du 4 juillet 2019 au 5 juillet 2019 (Vincent Bremhorst, Sophie Mathieu)
- **NEW** - Ecole d'été sur le Text Mining en Python, du 8 juillet 2019 au 12 juillet 2019 (Patrick Watrin, Thomas François, Hubert Naets)

Publications and Presentations by SMCS members

Although the primary mission of SMCS members is to support other research projects, its members are encouraged to participate in scientific meetings, to present their work, and to publish.

Publications

Journal Articles

Andres, M. ; Salvaggio, S. ; Lefèvre, N. ; Pesenti, M. ; Masson, N. (in press). *Semantic associations between arithmetic and space : Evidence from temporal order judgements*. In: *Memory and Cognition*.

Appendix

Bachmann, R., Léonard, D., Nachit, M., Remue, C., Abbes O., Desmet, L., Faber, B., Danse, E., Trefois, P., Kartheuser, A. (2018). *Comparison between abdominal fat measured by CT and anthropometric indices as prediction factors for mortality and morbidity after colorectal surgery*. In: *Acta Gastro-Enterologica Belgica*, Vol. 81, no.4, p. 477-483.

Dormal V., Bremhorst V., Lannoy S., Lorant V., Luquiens A., Maurage P. (2018). *Binge drinking is associated with reduced quality of life in young students: A pan-European study*. In: *Drug & Alcohol Dependence*, Vol. 193, p.48-54.

Dormal, V., Larigaldie, N., Lefèvre, N., Pesenti, M., Andres, M. (2018). *Effect of perceived length on numerosity estimation: Evidence from the Müller-Lyer illusion*. In: *Quarterly Journal of Experimental Psychology*, Vol. 71, no. 10, p. 2142-2151.

Guisset, S., Martin, M., & Govaerts, B. (2019). Comparison of PARAFASCA, AComDim, and AMOPLS approaches in the multivariate GLM modelling of multi-factorial designs. *Chemometrics and Intelligent Laboratory Systems*, Vol. 184, p. 44-63 (2018).

Remacle, A., de Chamnourcy, L. and Lefèvre N. (in press). *Which teachers are most at risk for voice disorders? Individual factors predicting vocal load parameters*. In: *Journal of Speech, Language, and Hearing Research*.

Trotta L., Kabeya Y., Buyse M., Doffagne E., Venet D., Desmet L., Burzykowski T., Tsuburaya A., Yoshida K., Miyashita Y., Morita S., Sakamoto J., Praveen P. and Oba K. (2019). *Detection of atypical data in multicenter clinical trials using unsupervised statistical monitoring*. In: *Clinical Trials*, Vol. 16(5), p. 512-522; doi: 10.1177/1740774519862564.

Seminars and Presentations

Extra-muros seminars and video

Bertrand, A. et Guisset, S. Participation au MOOC Critix coordonné par le Louvain Learning Lab. *Vidéos sur l'échantillonnage et les sondages* à destination de tous les étudiants de BAC1, 1^{ère} utilisation en 2018-2019.

Bremhorst V. et Bertrand, A. Préparation et réalisation d'une vidéo sur les statistiques dans le monde numérique, vidéo de vulgarisation dans le cadre du fonds Wernaers, Été 2019.

Bremhorst V., *Introduction to the basic concepts of Bayesian statistics based on clinical applications*. Séminaire organisé chez GSK Vaccines, Rixensart, le 11 décembre 2018.

Bugli, C., *Design and Analysis of Clinical Trials*. Séminaire de formation continuée au sein du service de rééducation et réadaptation de Saint Luc, Woluwé, le 25 octobre 2018.

Desmet L., *Inlier/Outlier detection based on Mahalanobis distance*. Présentation par téléconférence lors du séminaire FDA CRADA, le 7 décembre 2018.

Desmet L. (presenter), Venet D., Doffagne E., Trotta L., Akhtar-Danesh N., Burzykowski T. & Buyse M. *Efficient detection of fraud in multicenter experiments*. Présentation lors la conférence ISCB40, Leuven, le 18 juillet 2019.

Presentations and applied statistics workshops

Belhocine R., et Guisset S., *Atelier d'introduction au text mining*. Présentation dans le cadre de l'atelier de statistique appliquée, Louvain-la-Neuve, le 30 novembre 2018.

Bremhorst V., *Fertility progression in Germany, An application using nonparametric cure survival models*. Intervention dans le cadre de l'Atelier de statistique appliquée, Louvain-la-Neuve, le 3 mai 2019.

Poster

Dormal, Valérie ; Larigaldie, Nathanaël ; Lefèvre, Nathalie ; Pesenti, Mauro ; Andres, Michael. *Effect of perceived length on numerosity estimation: Evidence from the Müller-Lyer illusion*. In: *Quarterly Journal of Experimental Psychology*, Vol. 71, no. 10, p. 2142-2151 (2018). <http://hdl.handle.net/2078.1/194300>

Remacle, Angélique, De Chambourcy, Laetitia, Lefèvre, Nathalie. *Quels facteurs individuels peuvent prédire la charge vocale des enseignants ?* Journées de Phonétique Clinique - 14-16 mai 2019, Mons, Belgium.



2018-2019