REPORT OF ACTIVITIES

Statistical Methodology and Computing Service

1 September 2019 to 31 August 2020

2019-2020

UCLouvain

LiDAM
Louvain Institute of Data Analysis and
Modeling in economics and statistics

sMCs
Plate-forme Technologique de Support en Méthodologie et Calcul Statistique

Statistical Methodology and Computing Service
Report of Activities
1 September 2019 to 31 August 2020

Contents

Contents

Introduction

Introduction

The SMCS continues its activity as a diversified service provider in practical statistics. Our team now consists of 7 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 2019/20.

The consulting activity has grown to about 500 projects per year. Due to the COVID-19 epidemic several courses had to be canceled. However, by quickly converting some courses to online versions, we were able to limit the disruptions. In the end, we taught 39 courses corresponding to 130 training days. 506 participants attended for a total of 1688 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 (researchers, administrators, staff at the university hospitals). As every year, we distributed our newsletter explaining our goals, expertise, and services via the three sectors. A new edition was printed and distributed in early 2020 but could not reach its audience due to the COVID-19 epidemic and the related transition to online work.

Several new projects started during the reporting period and their descriptions can be found in the section Larger Projects. They include three new certificate courses, a modular version of the Junior Data Analyst, and two versions, basic and advanced, of a new program called Data Analysis in Research and Development.

The SMCS has continued in assisting the UCLouvain administration. This included urgent assistance in the coordination of the online exams in June and August 2020 and ongoing support in data extraction and management for a study of student trajectories in undergraduate and graduate programs at UCLouvain (in collaboration with the "Service d'études" - SET).

Several external consulting projects started including analysis support for the Proof-Next Registry of Idiopathic Fibrosis Patients, a collaboration with the Belgian Respiratory Society.

Our support for surveys at UCLouvain and external entities has continued its growth. We now offer two to three training sessions per year and our expertise in working with survey tools helped in converting important institutional elections at UCLouvain to a fully online format.

Obviously, just like the Olympics, the second edition of our summer school planned for July 2020 had to be postponed. Anticipating continuing partial restrictions in 2021, we decided to postpone our summer school to 2022.

On the other hand, the adaptations forced on us by the epidemic gave us the opportunity to develop our expertise in e-learning and e-examination. We shall continue our support to adapting pedagogic resources to new realities. This includes a collaboration with EPHEC to help future students in Human and Social Sciences start their first courses of statistics in good condition. This is very important now as the new cohort of beginning students finished their secondary school under difficult circumstances and who now must cope with additional challenges due to sanitary restrictions.

We expect 2020/21 to be an atypical year characterized by uncertainty and rapid changes of learning and collaboration environments.

I am very impressed by the speed and professionalism by which our team succeeded its migration to remote working. Obviously, this is not the ideal environment for creating and maintaining our team-spirit, but the daily virtual coffee break has helped a lot in keeping us together. We shall have to continue like that since remote working will remain necessary for months to come. Adapting and seizing new opportunities will be the mission for the year.

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

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

People and organization

People and Organization

Management and Administration

Director: Nathalie Lefèvre
President of Management Committee: Christian Ritter
Administrative coordinator: Nancy Guillaume
Communications and Quality Manager: Marie-Pascale Cosse
Administrative support: Tatiana Regout

Management committee

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.

President: Christian Ritter
Director: Nathalie Lefèvre
SSH representative (LIDAM institute): Fabio Mariani
SSS representative (IRSS institute): Niko Speybroeck
SST representative (ELI institute): Claude Bragard
Representative of research activities (ISBA): Catherine Legrand
Representative of training activities (LSBA): Johan Segers
Associate consultant academic representative: Bernadette Govaerts
Consultant representative: Céline Bugli
With consultative voice:
Administrative coordinator: Nancy Guillaume
Representative of external activities: Antoine Froidure (St Luc)
Representative for each University entity with which the platform has IT collaborations: Thomas Keutgen for CISM and Laurent Buset for SGSI

User committee

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. In 2020, due to the Covid crisis, we will organize a remote meeting with the user representatives.
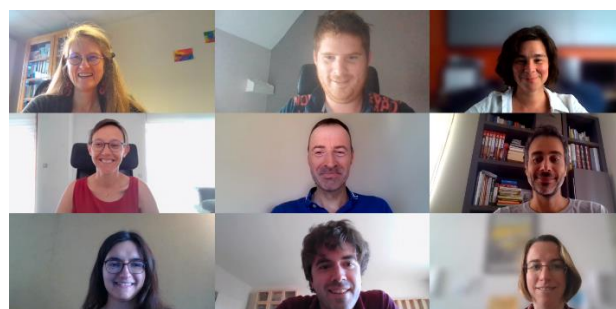
As a reminder, the user committee is a precious relay: relaying the needs of the users of their entities to the SMCS and relaying the services offered by the SMCS within their entities.

Bruno Schoumaker (IACCHOS), Magali Paquot (IL&C), Marco Saerens (LOURIM), Françoise Muller (INCAL), Michael Andres (IPSY), Cécile Bonmariage (ISP), Pierre Baudewyns (ISPOLE), Frédéric Vesentini (JUR-I), Michael Andres (IoNS), Annie Robert (IREC), Vincent Lorant (IRSS), Laure Elens (LDRI), Laurent Gatto (DDUV), Nicolas Schtickzelle (ELI), Jean-Charles Delvenne (ICTEAM), Tom Leyssens (IMCN), Vincent Destoop (IMMC), Andrea Giammanco (IRMP), Isabelle Donnay (LIBST), Isabelle Thomas (LIDAM), Frédéric Vrins (LIDAM-Mons), Françoise Docq (ADEF), Patrick Watrin (CENTAL), Damien François (CISM), Laurent Buset (SGSI)

People

Given the workload of the SMCS team, the management committee has validated the hiring of an additional consultant (1FTE) and financed by the SMCS funds. Benjamin Colling, PhD in Statistics, will start in September 2020. The network of data analysis consultants is generally expanding at UCLouvain in particular with the arrival of Gipsi Lima Mendez, hired in 2019 as an SST research logistician to provide precious support in bioinformatics.

Consultants

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

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

People and organization

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: Human Sciences 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 – 0.95 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 – 0.8 FTE)

Main areas of activity: Human Sciences 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: Human Sciences 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).

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 (SSH/IACCHOS), Sébastien Fontenay and Adam Levai (SSH/LIDAM/IRES), Nathalie Lefèvre (SSH/IPSY), Jérôme Ambroise (SSS/IREC), Aline van Maanen (Institut Roi Albert II - Cancérologie et Hématologie), Gipsi Lima Mendez (SST)

Researchers and Research Assistants

Some researchers and doctoral students of the ISBA participate in the consulting and training activities of the

People and organization

SMCS. This gives them an opportunity to gain valuable experience in practical statistics.

Interns
No intern this year.

Trainers and Consultants
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), Laure Cougnaud (Open Analytics), Lieven Desmet (SMCS), Laurent Gatto (DDUV), Bernadette Govaerts (ISBA), Alain Guillet (SMCS), Séverine Guisset (SMCS), Alexandre Jacquemain (ISBA), Eric Lecoutre (WeLoveDataScience!), Nathalie Lefèvre (SMCS), Adam Levai (IRES), Vincent Lorant (IRSS), Rebecca Marion (ISBA), Sophie Mathieu (ISBA), Gilles Mordant (ISBA), Emmanuel Niyigena (ISBA), Eugen Pircalabelu (ISBA), Catherine Rasse (SMCS), Christian Ritter (LSBA), Xavier Rouby (Pfease), Marco Saerens (LRIM), Antoine Soetewey (ISBA) 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. Six of the training sessions took place on the Woluwe campus. Nine training courses have been given in English. We will continue to propose every year more training courses in English but for very advanced trainings, participants generally prefer that they be given in their mother tongue.

Eleven training sessions (27 days) had to be cancelled, three of them because there were not enough participants (two of them concern SAS and one JMP). 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. The other eight cancelled trainings (20 training days) were reported because of the Covid-19 pandemic. Some Stata training courses which were cancelled have been replaced by videos available to the participants.

Despite the 20 days of training cancelled because of the Covid-19 pandemic, it was also an opportunity to innovate and offer distance training sessions. No less than 6 trainings were proposed totally online, via Teams and 3 trainings partly at distance. This represents 30 days of distance training in total. Some video supports were also created in this context.

Six new training programs have been developed including 2 new university certificates.

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: "*Modèles multi-niveaux avancés avec SPSS*" or "*Basic concepts and practical applications of Bayesian statistics*" (organized in collaboration with IPSY but open to everyone), "*Analyse de données Omics sur base de techniques multivariées avec R*" (organized for and funded by GSK but open to UCLouvain researchers).

- **University certificate**. In collaboration with the LSBA, the SMCS is also involved in and/or coordinate university certificates. This year, 2 new certificates have been introduced and were partially organized remotely since March: "University Certificate: Applied Data Analysis in R&D - basic level" and the certificate "Junior Data Analyst" organized in a modular program more compatible with a professional activity throughout the calendar year 2020. Teaching this course implied working with several external experts active in the data science business and we have 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. Unfortunately, the 2020 edition, on a similar theme, had to be postponed to 2022 because of COVID.

- **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 39 courses in the academic year 2019-2020 for 130 teaching days (+1 training currently running: the 10 training days organized in 2019-2020 for the modular JDA are not included in the

Training

report)[1]. The decrease observed is mainly due to the pandemic. There is essentially a strong decrease in external Person-Days. This is due to the fact that the training organized in collaboration with the Forem is now a certificate, the participants are now listed as UCLouvain students and no longer as external. It should be noted that the number of participants in the trainings has decreased because some trainings could not be given remotely (such as the Summer school with about 30 participants during 5 days), but also because we have reduced the maximum number of participants for our online trainings in order to allow an easier management.

Table 1. Global statistics on SMCS trainings.

	2017/18	2018/19	2019/20
Courses	57	52	39
Course Days	168.5	172	130
Participants UCLouvain	660	493	386
Person-Days* UCLouvain	1517	1425	1426
Participants Ext	224	253	120
Person-Days* Ext	737.5	1000	262

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

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

	2017/18	2018/19	2019/20
SSH	277	177	162
SSS	130	97	82
SST	90	54	71
Students	159	165	70
Central services	4		1

The observed fluctuations between years and sectors reflect natural variation and the organization of special events.

Compared to previous years, there is an overall decrease in the number of UCLouvain participants, which is explained by a lower number of trainings. However, there was a reassuring increase in SST, which had a very low number of participants last year.

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 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 Human Sciences Sector (SSH)

	2017/18	2018/19	2019/20
SSH	**321**	**247**	**188**
IACS-SPLE-ESPO	87	54	51
ILC	19	18	21
LOURIM-LSM	15	13	3
LIDAM	40	37	16
INCAL-FIAL	10	5	
IPSY-PSP	98	48	64
JURI-DRT	3	1	1
RSCS-TECO	5	1	6
Students	44	70	26

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.

---

[1] Participants in the Modular JDA 2020 are not included in the report because the training continues beyond August 2020.

Training

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

	2017/18	2018/19	2019/20
SSS	**155**	**112**	**94**
ASS			1
DDUV-Ludwig Inst.	13	5	16
IONS	23	13	15
IREC-MEDE-FSM	40	44	29
IRSS-FSP	42	17	9
LDRI-FASB	10	14	10
Clinics	2	4	2
Students	25	15	12

In the science and technology sector courses were also followed by participants from a wide range of institutes. Nevertheless, more than 60% of the SST staff participating in SMCS training comes from the ELI Institute.

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

	2017/18	2018/19	2019/20
SST	**180**	**134**	**103**
AGRO	9	6	2
ELI	32	26	47
ICTEAM-EPL	5	2	
IMCN	1		1
IMMC	7		9
IRMP	5	1	
LIBST	20	17	5
LIDAM			3
LOCI	7		3
SC	4	2	1
Students	90	80	32

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 2.3 full time staff equivalents[2] (FTE). 2 FTE of this effort is directly oriented towards the UCLouvain research and teaching units, the remainder towards external clients, such as the Luxemburg Institute of Health 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 and to propose new accredited certificates. 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.

---

[2] The calculation of this time is based on the duration of the training courses multiplied by a coefficient usually used to take into account the preparation time of the training courses. This year, this coefficient should have been increased because the organization of remote learning courses took a lot of time due to the preparation of specific materials.

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 1.6 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 still particularly important this year and

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 fourth year, we include in our statistics cross-disciplinary supports to each sector. This consists in statistical computing support for all the sector community or projects that target several research or teaching entities, such as a website project to bringing together statistical resources to help future students in Human Sciences Sector, or a project of collection, identification, implementation and sharing of good data protection practices (a GDPR project). Even if this type of support counts for one unit in the table, this can sometimes represent up to 50 hours of work.

Table 7 (SSH) shows that the number of requests from the Human Sciences Sector usually fluctuating around 130 requests per year, increased the two last years to now nearly 190 files (1572 hours of work). As in previous years, the institute and the school of psychology had by large the

corresponds to more than 3000 hours (660 hours for external projects and 2420 for internal one's).

Overall, the consulting activity of the SMCS occupies about 4.4 full time staff equivalents (FTE). 3.8 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	2017/18	2018/19	2019/20
Research	401	425	450
Educational project	2	9	14
Admin. services	8	9	23
External clients	9	10	18
Total	**420**	**453**	**505**

From 1 September 2019 to 31 August 2020, the SMCS worked on 505 projects of which 487 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

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). IACCHOS 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.

Consulting

Table 7. Origin of consulting requests from the Human Sciences Sector (SSH)

	2017/18	2018/19	2019/20
SSH	**133**	**169**	**187**
Cross-disciplinary	1	2	4
IACCHOS-ISPOLE-ESPO	18	16	26
ILC-INCAL-FIAL-ISP	20	16	20
LOURIM-LSM	7	14	7
LIDAM	12	26	20
IPSY-PSP	73	93	105
JURI-DRT	2	2	5

Table 8 (SSS) shows that on average there are 150 requests per year from the health sciences sector, with an increase the two last years to now nearly 200 files (864 hours of work). The areas related to clinical research, Health and Society Institute and the medical school account for about 65% of this, other requests are split among the remaining institutes.

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

	2017/18	2018/19	2019/20
SSS	**151**	**167**	**198**
Cross-disciplinary	1	1	1
DDUV-Ludwig Inst.	8	7	6
FASB	3	3	7
IONS	26	18	24
IREC-MEDE-FSM	63	87	104
IRSS-FSP	30	39	30
LDRI	8	3	13
Clinics - Albert II	12	9	13

Table 9 (SST) shows that the number of requests from the science and technology sector fluctuates around 90-120 requests per year, with a decrease this year to 92 files (644 hours of work). ELI and LIBST account for about 70% of this, other requests came from almost all units of this sector.

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

	2017/18	2018/19	2019/20
SST	**117**	**98**	**92**
Cross-disciplinary	2	2	2
AGRO	7	10	7
ELI	36	28	36
ICTEAM-EPL	3	5	3
IMCN		1	
IMMC	14	7	6
LIBST	45	36	29
LIDAM		1	1
LOCI	4	4	2
SC	6	4	6

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.

A research support project, **transversal to all sector** has to be added: this support consisted in the replacement of the SMCS server and installation of SAS Viya.

Finally, **nine projects** also originated from the central administration (AC, LLL, SGSI, BIUL). This year, support to central administration and administrative services in general was particularly high, with a total of 1051 hours of work. This is explained first of all by the establishment of a new collaboration between SET and SMCS on an institutional project to monitor success in BAC and MASTER. It is also explained by the support provided by SMCS on issues arising from the pandemic (remote elections and management of the exam organization). 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 57 users. People from LIDAM and IACCHOS 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 7% on smcs3 and 1% on smcs4 with spikes (100% for smcs3 and 50% 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 and PASS.

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 or you can ask for them at the Service desk.

Computing

Usage Statistics

Table 10. Users of the statistical computing server

	2017/18	2018/19	2019/20
Total	**52**	**65**	**57**
SSH	**30**	**35**	**35**
IACS-SPLE-ESPO	3	8	15
ILC	2	1	
LIDAM	24	24	20
IPSY-PSP	1	2	
SSS	**14**	**12**	**10**
IONS			
IREC-MEDE		1	2
IRSS-FSP	14	10	8
LDRI		1	
SST	**8**	**18**	**12**
ELI	6	5	3
ICTEAM-EPL	1	1	2
IMCN		1	2
IMMC		5	4
IRMP		1	1
LIBST		1	
SC	1	4	

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

	2016/17	2017/18	2019/20
Total	**50**	**27**	**47**
SSH	**25**	**17**	**28**
IACCHOS		3	4
ILC	1		2
IPSY-PSP	12	8	12
JURI			1
LIDAM	9	4	7
LOURIM-LSM	3	2	2
SSS	**17**	**10**	**16**
DDUV	4		1
IONS-FSM	2	4	6
IREC	5	4	6
IRSS	2	1	1
LDRI	4	1	2
SST	**8**	**0**	**3**
AGRO	2		
ELI			2
IMMC	1		
LIBST	2		
LOCI			1
SC	3		

Usage of UDS

The most used software was PASS (20 different users), followed by SPSS Premium (17 users), MPlus (8 users) and HLM (2 users) during the year. Surprisingly, only one user of PASS came from SST whereas they need such a software to justify the sample size of their experiment when they enter a new project that the Ethical committee needs to validate. More generally, PASS should be more known 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 2019-2020, the platform helped 96 students. The decrease observed the previous academic year was still true during this one until May. After this, students began to ask for SPSS and Stata for their online exams (they were afraid of being disconnected from UDS during their exams) and so the demand for help burst.

Larger Projects

Larger Projects

This section summarizes the largest projects in which the SMCS was involved in 2019-2020. Some of these projects are research projects coordinated by academic members of UCLouvain or other universities. Other projects consist in providing support: regarding surveys, teaching activities, but also data management and data extraction (for the UCLouvain administration). Finally, some university certificates also belong to this category.

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 and provide regular reports to the project team.

Research project - CluePoints (2018-2022)

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 - PDR (2020-2024)

The SMCS is involved in research project METNATPAR for the identification of the metabolic and/or the biochemical pathways that are triggered by promising natural antiparasitic derivatives in Trypanosoma brucei and Plasmodium falciparum. Our involvement has just begun this year with the training of the researchers involved in

the project and assistance in the design of the data collection.

Collaborative research project - Survey among book authors in Fédération Wallonie-Bruxelles

This survey will be implemented within the framework of a cooperation agreement between SMCS and Maison des Auteurs asbl, with a grant from Fédération Wallonie-Bruxelles. The main objective is to study the socio-economic standing of authors in Wallonia and Brussels, and more specifically the diversity of their professional activities and income (type, structure and amounts). The SMCS will take part in the questionnaire design, programming and data analysis. The study will be completed by qualitative interviews and will lead to a final report presenting the findings.

Survey Support

Survey support consulting was provided for a wide range of applications in all three sectors in 2019-2020 and included advice on survey and questionnaire design, programming in LimeSurvey, sampling, data management, weighting and analysis.

Two external survey projects on the car and real estate markets in Belgium were conducted in 2020 following the Covid lockdown.

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 November 2019, two training sessions on surveys and research reports were provided to CEFORA as part of a broader training program.

Over the second quarter of 2020, several elections for presidents of institutes, deans and vice-rectors took place. These elections, usually organised on paper, had to move to a digital format this year. The SMCS contributed to the implementation and supervision of secure election processes via LimeSurvey for each of these. This process was implemented 10 times with success, including during the election of the Hainaut's vice-rector, involving over 2000 voters and a weighting system based on the voters' status.

Larger Projects

The project of a database gathering information on surveys conducted by UCLouvain, scheduled for 2020, was put on hold and will be on the agenda for 2021.

Data management support - Online exam sessions

The SMCS was involved in the coordination of the online exam sessions in June and September 2020. Several lists of students and summary tables were provided to different stakeholders, such as the service desk, the faculties, the professors, the cleaning service…, to ensure a proper organization of the exam sessions.

Data management support - Student path data (2019-…)

In collaboration with the "Service d'études (SET)", indicators related to the evolution of the duration of studies and the evolution of the percentage of students graduated on time are generated. Dashboards are developed to summarize such indicators numerically and to graphically illustrate their evolution since 2008. Up to now, only the first generation UCLouvain student cohort is considered.

Support to other platforms

The SMCS assists the CICN (Clinical Investigation Centre for Nutrition) platform in writing protocols for clinical studies in nutrition by reviewing study design, calculating sample sizes and creating randomization lists. The SMCS also assists in setting up statistical analyses.

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, this year, our platform followed up on the final steps in the coordination and development of pedagogical materials for statistical learning.

In the same vein, between July 2019 and August 2020, we took part, in collaboration with the EPHEC, in a project related to the preparation of new students for higher education. We created a website (https://aidealareussite.uclouvain.be/course/view.php?id=65) bringing together existing and new resources to help future students in Human Sciences Sector in the Pole Louvain who start their first course in statistics in the best conditions.

We also carried on providing support to LLL related to alternatives to the negative points in the MCQs to be proposed to UCLouvain teachers.

Finally, one member of our team is responsible for the maintenance of the Shiny applications of the project Shiny@UCLouvain (https://sites.uclouvain.be/RShiny/main.html).

Project linked to education - Dynamo project

This project aims to give tools to the teacher to evaluate how well a group of students works in using indices in Moodle. Students have to fill a short survey in from which we will give group performances. This index will then be implemented in Moodle interface.

The SMCS provides its statistical expertise in order to find the best possible index.

University Certificate - Analyse de données en R&D (2020-…)

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, the course is since January 2020 valued in 2 university certificates.

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.

Larger Projects

The courses were given for the fourth edition from October 2019 to December 2019 and is self-financing since the third edition. The fifth edition will start in October 2020.

In January 2020, the first edition of a modular JDA has been launched. This modular program is more compatible with a professional activity because half of the courses are organized on Saturdays. A second edition will start in January 2021.

Summer School - Collaboration with the CENTAL

Last year, the SMCS launched a new type of training, a Summer School, focused on an advanced topic and given by invited specialists. Last year edition was about Text Mining in Python, in collaboration with the Cental. Following this first successful edition, the second one, about Text mining and Social Network Analysis in Python (again in collaboration with the Cental), was planned for July 2020 but had to be cancelled due to the COVID pandemia. We hope to be able to reschedule it in January 2022.

External project - Training for journalists

In collaboration with the Royal Statistical Society of Belgium, the SMCS started the creation of material for two versions (one in Dutch, the other in French) of a training aimed at journalists about the correct reporting and interpretation of numbers and graphs. The SMCS will also be involved in the first session of the training in each language; they should take place in 2020-2021.

External project – Proof-Next (2019-2021)

In collaboration with the Belgian Respiratory Society, the SMCS is involved in the analysis of the Proof-Next registry of Idiopathic Pulmonary Fibrosis patients.

External Project - Baromètre de la propreté en Wallonie (Click4WaPP)

This project consists in the analysis of data collected from 2016 to 2019 in the framework of the Click4WaPP project, aiming at studying public cleanliness in Wallonia. Our role is to review existing analyses and methodologies and implement new analyses in order to better understand public cleanliness.

Your projects? SMCS can be your partner!

As you can see from this report, we are ready to become partners of various types of projects. 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 2019-2020

Trainings
- Introduction to Python, du 27 août 2019 au 3 octobre 2019 (Vincent Bremhorst)
- Introduction to data analysis with R, du 6 septembre 2019 au 13 septembre 2019 (Alain Guillet)
- Pratique de la statistique avec SPSS, du 9 septembre 2019 au 13 septembre 2019 (Nathalie Lefèvre)
- Certificat universitaire en Junior Data Analyst, du 30 septembre 2019 au 6 janvier 2020 (Aurélie Bertrand, Jonathan Dedonder, Alain Guillet, Christian Ritter, Séverine Guisset, Vincent Bremhorst, Marco Saerens, Formateurs externes)
- Initiation au logiciel NVivo, du 1 octobre 2019 au 2 octobre 2019 (Jonathan Dedonder)
- Analyse de données d'enquête, du 7 octobre 2019 au 8 octobre 2019 (Séverine Guisset)
- **NEW** - R object-oriented programming and package development, du 10 octobre 2019 au 10 octobre 2019 (Laurent Gatto)
- Data management and analysis using Stata, du 14 octobre 2019 au 18 octobre 2019 (Adam Levai)
- Initiation à la statistique avec applications en JMP (Formation intégrée dans un programme intensif plus large du FOREM), du 5 novembre 2019 au 8 novembre 2019 (Catherine Rasse, Lieven Desmet, Aurélie Bertrand)
- **NEW** - Introduction à l'analyse de Réseaux, du 6 novembre 2019 au 6 novembre 2019 (Séverine Guisset, Alain Guillet, Jonathan Dedonder, Vincent Lorant, Xavier Rouby)
- Pratique de la statistique avec SPSS, du 8 novembre 2019 au 29 novembre 2019 (Nathalie Lefèvre)
- Introduction to SAS programming, du 13 novembre 2019 au 15 novembre 2019 (Céline Bugli, Catherine Rasse, Lieven Desmet)
- Introduction au calcul de taille d'échantillon, du 14 novembre 2019 au 14 novembre 2019 (Catherine Rasse, Céline Bugli)
- Préparation et mise en place d'une enquête, du 25 novembre 2019 au 26 novembre 2019 (Jonathan Dedonder, Séverine Guisset)
- **NEW** - Programmation en syntaxe SPSS, du 26 novembre 2019 au 27 novembre 2019 (Vincent Bremhorst, Nathalie Lefèvre)
- Build interactive web applications with the R Shiny Package, du 3 décembre 2019 au 4 décembre 2019 (Formateurs externes)
- Analyse de données pré-cliniques et cliniques avec SPSS, du 6 décembre 2019 au 13 décembre 2019 (Nathalie Lefèvre)
- Introduction à la réalisation d'enquêtes en ligne avec LimeSurvey, du 9 décembre 2019 au 10 décembre 2019 (Jonathan Dedonder)
- Analyser la médiation, la modération et la médiation modérée avec SPSS, du 17 décembre 2019 au 18 décembre 2019 (Aurélie Bertrand)
- Pratique de la statistique avec SPSS, du 6 janvier 2020 au 10 janvier 2020 (Nathalie Lefèvre)
- Introduction à l'analyse de données avec R, du 10 janvier 2020 au 17 janvier 2020 (Alain Guillet)
- **NEW** - Certificat universitaire en Junior Data Analyst - Formule modulaire (*formation maintenue, partiellement à distance*), du 10 janvier 2020 au 24 octobre 2020 (Aurélie Bertrand, Jonathan Dedonder, Alain Guillet, Christian Ritter, Séverine Guisset, Vincent Bremhorst, Formateurs externes) - *Not counted in the statistics of this report because training is still ongoing in 2020-2021*
- Plans d'expérience avec JMP, du 13 janvier 2020 au 16 janvier 2020 (Catherine Rasse)

Appendix

- Basic concepts and practical applications of Bayesian statistics with R and JASP, du 20 janvier 2020 au 21 janvier 2020 (Vincent Bremhorst)
- Introduction aux modèles multi-niveaux avec SPSS, du 20 janvier 2020 au 24 janvier 2020 (Aurélie Bertrand, Nathalie Lefèvre)
- Analyses multivariées exploratoires en R, du 27 janvier 2020 au 28 janvier 2020 (Séverine Guisset)
- Initiation au logiciel NVivo, du 27 janvier 2020 au 28 janvier 2020 (Jonathan Dedonder)
- **NEW** - Modèles multi-niveaux avancés avec SPSS, du 27 janvier 2020 au 31 janvier 2020 (Aurélie Bertrand)
- Introduction to data analysis with R, du 11 février 2020 au 20 février 2020 (Lieven Desmet, Emmanuel Niyigena)
- Analyses multivariées avec SPSS, du 17 février 2020 au 21 février 2020 (Séverine Guisset)
- **NEW** - Certificat d'université : Analyse de données appliquée en R&D - niveau de base (*formation maintenue, partiellement à distance*), du 24 février 2020 au 12 juin 2020 (Aurélie Bertrand, Céline Bugli, Lieven Desmet, Bernadette Govaerts, Alain Guillet, Catherine Rasse)
- Concepts de base et pratique de la statistique bayésienne avec R et JASP, du 2 mars 2020 au 6 mars 2020 (Vincent Bremhorst)
- Initiation aux modèles d'équations structurelles (SEM) avec Stata (*partiellement dispensée à distance*), du 9 mars 2020 au 12 mars 2020 (Aurélie Bertrand)
- Initiation au logiciel NVivo (formation réservée aux employés de l'Agence pour une Vie de Qualité), du 12 mars 2020 au 13 mars 2020 (Séverine Guisset)
- Introduction au logiciel JMP - *Formation maintenue à distance avec TEAMS*, du 14 mai 2020 au 15 mai 2020 (Catherine Rasse)
- Inférence et modélisation statistique avec JMP - *Formation maintenue à distance avec TEAMS*, du 28 mai 2020 au 18 juin 2020 (Catherine Rasse)
- Introduction à l'analyse de données avec R - *Formation maintenue à distance via TEAMS*, du 5 juin 2020 au 12 juin 2020 (Vincent Bremhorst)
- Introduction à l'analyse de Réseaux Sociaux avec Gephi - *Formation maintenue à distance via TEAMS*, du 8 juin 2020 au 10 juin 2020 (Jonathan Dedonder)
- Initiation au langage Python - *Formation maintenue à distance via Teams*, du 2 juillet 2020 au 3 juillet 2020 (Vincent Bremhorst)
- Write and publish your work in an R-package - *Online training using Teams*, du 14 juillet 2020 au 14 juillet 2020 (Vincent Bremhorst)

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. (2020). *Semantic associations between arithmetic and space: Evidence from temporal order judgements*. In: *Memory and Cognition — Vol. 48, no. 3,* p. 361-369.

Darrigade, A.-S., Colmant, C., de Montjoye, L., Herman, A., Bugli, C., Tromme, I., & Baeck, M. (2020). *Atopic Dermatitis Score 7 (ADS7): A promising tool for daily clinical assessment of atopic dermatitis.* In: *Allergy — Vol. 75, no. 5,* p. 1264-1266.

Appendix

Friedel, M., Gilson, A., Bouckenaere, D. Brichard, B., Fonteyne, C., Wojcik, T., De Clercq, E., Guillet, A., Mahboub, A., Lahaye, M. & Aujoulat, I. (2019). *Access to paediatric palliative care in children and adolescents with complex chronic conditions: a retrospective hospital-based study in Brussels, Belgium.* In: *BMJ Paediatrics Open — Vol. 3, no.1,* p. e000547.

Harkemanne, E., Duyver, C., Leconte, S., Bugli, C., Thomas, L., Baeck, M., & Tromme, I. (2020). *Melanoma Diagnostic Practices of French-Speaking Belgian General Practitioners and the Prospective Study of Their Pigmented Skin Lesion Diagnostic Accuracy and Management.* In: *Journal of cancer education,* p. 1-9.

Lambert, P. & Bremhorst, V. (2020). Inclusion of time-varying covariates in cure survival models with an application in fertility studies. In: *Journal of the Royal Statistical Society. Series A, Statistics in society — Vol. 183, no. 1,* p. 333-354.

Seminars and Presentations

Extra-muros seminars and video

Bugli C., *Les statistiques dans votre mémoire: il vaut mieux y penser avant pour éviter les ennuis après...* Présentation aux mémorants FSM, Louvain-la-Neuve, le 28 octobre 2019.

Bugli C., *Comparaison des outils JMP et GraphPad Prism.* Présentation pour des membres de l'institut DDUV, le 17 juin 2020.

Desmet, L. et Bugli, C., *A gentle introduction to Mixed Models in biomedical and clinical research*. Présentation dans le cadre des séminaires IREC, Woluwé, le 25 novembre 2019.

Presentations and applied statistics workshops

Belhocine R. (Business & Decision), & Guisset S., *Atelier d'introduction au text mining*. Présentation dans le cadre de l'atelier de statistique appliquée, Louvain-la-Neuve, le 29 novembre 2019.

2019-2020