Springer Proceedings in Mathematics & Statistics

Regina Bispo
Lígia Henriques-Rodrigues
Russell Alpizar-Jara
Miguel de Carvalho *Editors*

Recent Developments in Statistics and Data Science

SPE2021, Évora, Portugal, October 13–16

Springer

Springer Proceedings in Mathematics & Statistics

Volume 398

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including data science, operations research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Regina Bispo · Lígia Henriques-Rodrigues ·
Russell Alpizar-Jara · Miguel de Carvalho
Editors

Recent Developments in Statistics and Data Science

SPE2021, Évora, Portugal, October 13–16

Springer

Editors
Regina Bispo
NOVA School of Science
and Technology
Caparica, Portugal

Russell Alpizar-Jara
University of Évora
Évora, Portugal

Lígia Henriques-Rodrigues
University of Évora
Évora, Portugal

Miguel de Carvalho
University of Edinburgh
Edinburgh, UK

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-031-12765-6 ISBN 978-3-031-12766-3 (eBook)
https://doi.org/10.1007/978-3-031-12766-3

Mathematics Subject Classification: 62-06, 62P10, 62P25, 62R07

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Organization

SPE 2021 was organized by the University of Évora and by the Portuguese Statistical Society (SPE).

Executive Commitee

Russell Alpizar-Jara (President), University of Évora (PT)
Dulce Gomes, University of Évora (PT)
Lígia Henriques-Rodrigues, University of Évora (PT)
Patrícia A. Filipe, ISCTE, University Institute of Lisbon (PT)

Scientific Committee

Miguel de Carvalho (President), University of Edinburgh (UK)
Fátima Ferreira, University of Trás-os-monte e Alto Douro (PT)
João Andrade e Silva, University of Lisbon (PT)
Luís Meira-Machado, University of Minho (PT)
Marco Costa, University of Aveiro (PT)
Maria Eduarda Silva, University of Oporto (PT)
Marília Antunes, University of Lisbon (PT)
Paula Brito, University of Oporto (PT)
Regina Bispo, Nova University of Lisbon (PT)
Rosário Oliveira, University of Lisbon (PT)
Russell Alpizar-Jara, University of Évora (PT)

Partners and Sponsoring Institutions

International Statistical Associations:

- Bernoulli Society
- Brazilian Statistical Association
- CWS (Caucus for Women in Statistics)
- FENStatS (Federation of the European National Statistical Societies)
- ISBA (International Society for Bayesian Analysis)
- RBras (Brazilian Region International Biometric Society)
- SGaPEIO (Sociedade Galega para a promoción da Estatística e da Investigación de Operacions)

National Statistical Associations:

- CLAD (Portuguese Association for Classification and Data Analysis)

Industry and Official Statistics:

- GADES (Data Analysis Solutions)
- INE (Statistics Portugal)
- PSE—Your Data Specialists
- PORDATA—Estatísticas sobre Portugal e Europa

Referees for Proceedings

Ana Freitas
Andy Lynch
Carlos A. Braumann
Clara Cordeiro
Dulce Gomes
Filipe Marques
Inês Sousa
Isabel Pereira
Jessica Silva Lomba
Kamil Turkman
Lisete Sousa
Luís Machado
M. Filomena Teodoro
Manuela Neves
Marco Costa
Maria Antónia Turkman
Maria de Fátima Ferreira
Maria Ivette Gomes
Maria Polidoro

Marília Antunes
Maurizio Sanarico
Nuno Sepúlveda
Patrícia de Zea Bermudez
Patrícia Filipe
Paula Brito
Paulo C. Rodrigues
Paulo M. M. Rodrigues
Pedro Campos
Raquel Menezes
Rita Sousa
Soraia Pereira
Tiago Marques
Vanda Lourenço

Welcome Message from the Editors

XXV Congresso
Sociedade Portuguesa
de Estatística
2021 Évora

Dear authors, referees, and readers of

Recent Developments in Statistics and Data Science,

It is a great pleasure to welcome you to the proceedings of the XXV Congress of the Portuguese Statistical Society—**SPE 2021**—held during 13–16 October 2021 at Évora, Portugal. This was the first-time, ever, online SPE conference, and it gathered more than 200 delegates from all over the world.

The meeting was hosted by University of Évora, Portugal, in collaboration with the Portuguese Statistical Society, and we had a fantastic program including 4 plenary lectures, 31 sessions, and 22 posters. A variety of societies had virtual rooms at SPE 2021 including *Bernoulli Society*, *Brazilian Statistical Association*, *Caucus for Women in Statistics*, and the *International Society for Bayesian Analysis*—just to name a few. Institutional members of the Portuguese Statistical Society were also represented (e.g. Statistics Portugal, Banco de Portugal, PORDATA). For more details on the meeting please, see *www.spe2021.uevora.pt/en/*.

Recent Developments in Statistics and Data Science highlights some selected contributions that were presented at SPE 2021. This volume covers a broad range of topics lying at the interface between Statistics and Data Science, such as applied statistics, computational statistics, extremes and outliers, medical statistics, modeling time series and stochastic processes, and data visualization, among others.

And speaking of visualization, Fig. 1 depicts a word cloud of all the titles and abstracts in this volume. While this chart is not a substitute for a table of contents, it does not summarize the order by which the articles appear in this issue! it offers a visual roadmap of what is to be found ahead. Given the broad scope of topics covered, we have opted for clustering articles according to the similarity of topics.

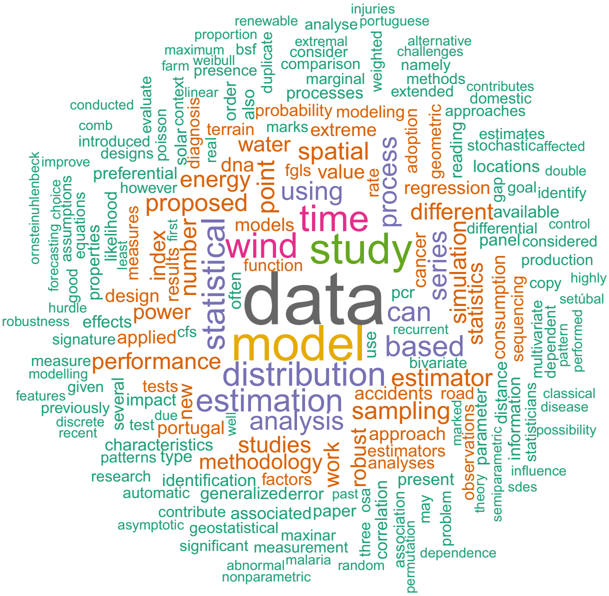
injuries
renewable analyse portuguese
proportion extremal alternative
maximum bsf consider challenges
farm weibull comparison namely
presence marginal methods contributes
linear processes extended domestic
conducted probability modeling approaches
comb terrain marks extreme estimates
introduced water stochastic affected
designs spatial locations double
improve preferential dna fgls value
however energy using regression goal identify
proposed time different available
models wind study differential control
panel considered
function production
data copy highly
pcr setúbal
design power can
hurdle
robustness effects recurrent
signature applied cfs model based
measure performance bivariate
modelling distribution estimator
features given tests accidents road classical
previously impact estimation sampling disease
discrete due test analysis
recent portugal approach possibility
characteristics studies estimators
patterns type methodology analyses influence
research identification factors present sdes
automatic generalized error past
contribute associated paper
asymptotic geostatistical maxinar
significant measurement dependence
abnormal malaria random
nonparametric
duplicate weighted evaluate poisson solar context order also diagnosis real
adoption geometric reading point process rate
ornsteinuhlenbeck forecasting choice assumptions good properties equations likelihood least measures
index results first number statistical often
cancer series simulation statistics consumption sequencing gap
multivariate dependent pattern performed use
several new well work robust
marked observations parameter theory distance semiparametric information statisticians
osa three correlation association permutation may problem

Fig. 1 Word cloud summarizing all titles, keywords, and abstracts of contributions included in this issue

And, as the tag cloud in Fig. 1 reveals, *data* are the common denominator across most contributions.

We are indebted to many people. First, we would like to thank the authors for their contributions and to everyone involved in the peer-review process who did a superb job on meeting tight deadlines in a thoughtful manner. They all worked hard so that the community keeps breaking new ground, and should be proud of their achievements. We are also indebted to Springer for their excellent collaboration on the production of this issue, and to the Scientific Committee and keynote speakers for their contributions to the meeting. Last but not the least, our words of thanks go to the organizers of SPE 2021 for their outstanding work, and to the community of the Portuguese Statistical Society for their continuous, and yet unbounded! support.

March 2022

Regina Bispo
Lígia Henriques-Rodrigues
Russell Alpizar-Jara
Miguel de Carvalho

Contents

How to Increase the Visibility of Statisticians in the Modern World of Dataism?

Nuno Sepúlveda

Abstract In the view of the historian Yuval Noah Harari, current human thought can be characterized by a deep belief in data, whether big or small, as the main vehicle to understand and control the world. This belief is referred to as Dataism. Notwithstanding their key role as guarantors of high-quality statistical exercises and data curators, statisticians typically remain in the shadow of big decisions in multidisciplinary and highly collaborative environments. This situation can be overcome by operating a change in the mindset of statisticians from *shoe clerks* to *statistical leaders*. Under the assumption that a statistician has already achieved a certain level of statistical proficiency, this paper aims to discuss useful skills, such as active listening, networking, and effective communication, which can foster statistical leadership and increase recognition and merit by non-statisticians inside and outside academia.

Keywords Impact · Interpersonal skills · Leadership · Statistical career

1 Introduction

We are living in a world in which data are an integral part of our daily experience as human beings. Take the example of our little good friend smartphone, which can collect data on the number of daily steps. With these data, we can judge whether we should be more active and, if so, we can get other data-collection apps to help us tracing that increase in activity. Data and the respective collection are so engrained in our lives that the famous historian Yuval Noah Harari writes in his best-selling book *Homo Deus: a Brief History of Tomorrow* about Dataism, a kind of new religion deeply rooted on the general belief that data and its flow are all that matters to understand and hold control of the world [1]. This religion made us to invest in

N. Sepúlveda (✉)
Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland
e-mail: N.Sepulveda@mini.pw.edu.pl

CEAUL - Centro de Estatística e Aplicações da Universidade de Lisboa, Lisbon, Portugal

R. Bispo et al. (eds.), *Recent Developments in Statistics and Data Science*,
Springer Proceedings in Mathematics & Statistics 398,
https://doi.org/10.1007/978-3-031-12766-3_1

innovative technologies in data acquisition, storage, and management [2]. As a result, current data can be big, huge, humungous!

In this brave new world of big data, our privacy, autonomy, and individuality are often given away to ensure our essential role as data providers. At the same time, such a profound and often blind belief on Dataism makes us all too vulnerable to unscrupulous politicians who use fake news or incomplete data to convince us to join their malicious cause. In this scenario, we statisticians can be seen as whistleblowers of data abuses, rationale and neutral players that can denounce disinformation, misinformation, or statistical malpractices [3]. A kind of justice league members of this data world. But is anyone out there who is willing to listen to us given that even the use of our *friend* p-value in Science is under debate and controversy [4]?

Unfortunately, the deep faith in Dataism did not make statisticians more visible in the society over the years. In fact, this reduced visibility can be traced back to the time of the great statisticians of the past. In 1938, Fisher [5] famously wrote in his usual cut-throat style that:

> To consult the statistician after an experiment finished is often merely to ask him to conduct a postmortem examination. He can perhaps say what the experiment died of.

This sentence is an open criticism to the still-prevailing attitude of seeking a statistician as a last resort. It also subtlety hints that statisticians are somehow invisible to their peers from the non-statistical world. In the past, statisticians were hidden disciples of Mathematics and now are simply *data crunchers* or *p-value providers*. To make things worse, statisticians are currently in direct competition with data scientists, bioinformaticians, and mathematical modellers in terms of their contribution to multidisciplinary, and above all, cutting-edge research. The fundamental question that we all statisticians face at the moment is then how to increase our visibility and value in this wondrous world of big data.

A first answer to this question can be found in the thought-provoking article entitled *The role of the Statistician: scientist or shoe clerk* by Irving Bross [6]. This author discusses the immediate and the long-term implications of a statistician adopting a posture similar of a shoe clerk whose primary objective is to please current and future customers. On the one hand, a shoe-clerk attitude has the advantage of neutrality and minimal hassle in moments of tension amongst team's members. It has also the advantage of fattening the resumé of applied statisticians (including the author of this paper) with a large number of middle-author publications; the underlying idea seems to be: minimal effort, maximum outcome, and a big boost of the ego. This advantage is in agreement with an increased number of middle authors in biomedical research [7], but it remains to be uncovered what is the contribution of statisticians to this trend. On the other hand, the hard truth is that, in the long-term, statisticians who solely act as shoe clerks will always be treated like one. Ultimately, the cordial, complacent but often neglecting treatment by their non-statistical peers suggests a certain *be a lamb* stereotype for the statistician as a professional. Finally, the quality of the statistical product itself could be also compromised, because it is intrinsically difficult for shoe-clerk-type statisticians to go against their customers who are typically in a standpoint of *I know what I want or need for my data* at the start of a collaboration.

More recently, Gibson [8] intertwines the concepts of visibility and the value of a statistician with leadership skills; however, the use of the word *value* has the unnecessary connotation that a statistician like a commodity can be sold up or down in the stock (or job) market. According to this author, the visibility of a statistician can be increased by creating a new culture around statistical leadership. This culture requires the acquisition of specific skills and the mindset of a leader, which will be discussed in Sect. 2. Such skills can be learned, practised, and improved. However, there is a limited number of universities offering courses on these leadership skills.

In this scenario, this paper is a collection of ideas and concepts scattered around the literature about leadership; a more personal account of this topic can be found elsewhere [9]. It is particularly directed to all the statisticians who wish to embrace a joyful journey towards a more impactful, fulfilling, and meaningful collaborations. Statistical leadership is above all a personal choice and not an authority, rank, or position. As such, it is accessible to everyone.

2 Statistical Leadership and Its Key Competences

According to Gibson [8], statistical leadership can be broadly defined as the use of influence without authority to guide the design, strategy, and decisions of a multidisciplinary team. The same author outlines three competences or soft skills essential for successful statistical leadership: (i) active listening; (ii) networking; and (iii) effective communication. These skills are not new and can be found in the popular book entitled *12 Rules for Life: Antidote to Chaos* by the clinical psychologist Jordan Peterson [10], but formulated as follows:

- Active listening: *Assume that the person you are listening to might know something you don't*;
- Networking: *Make friends with people who want the best for you*;
- Effective communication: *Be precise in your speech.*

A brief discussion on these skills will be presented in the next three subsections.

2.1 Active Listening

It is universally regarded that Nelson Mandela (1918–2013) was a great leader. He was the elected president in the first free democratic elections in South Africa after the end of the apartheid. One of the remarkable Mandela's leadership skills was his power of listening and using it for strategic and reflective questioning [11]. This power was developed by witnessing community meetings with his father who was the chief of his tribe [12]. He learned that everyone was seated in a circle and his father was always the last one to speak. The amazing capacity of listening to everyone before speaking and, more importantly, before any rushed judgement deeply resides the transformative power of active listening.

At the surface, one might think that active listening is just giving free and undivided attention to the speaker. However, it is more than that [13]. It involves (Table 1):

1. adopting an appropriate body language while listening;
2. reflecting in what is being heard;
3. understanding the consequences and implications of the information received.

As a consequence, active listening is able to generate mutual understanding, commitment between parties, and the joyful and fulfilling sensation that each side was heard. Ultimately, active listening builds trust and respect, which are necessary to maintain harmonious and sustainable collaborative environments.

We statisticians like medical doctors, nurses, and other professionals alike are required to develop such a listening skill due to our line of work. Unfortunately, this skill is taken for granted, because it is supposedly to be natural to have it. In the truth of the matter, it is not easy to master it without any effort and even more so in the modern world of constant distractions by smartphones, social media, amongst other factors. As suggested in the Introduction, this skill can be learned, trained, and improved. In the book entitled *How to be heard: Secrets for Powerful Speaking and Listening*, Julian Treasure [14] suggests simple exercises to improve one's listening capacities such as:

1. enjoy the sound of silence (or simply enjoy the song of Simon and Garfunkel or the cover by the Disturbed);
2. listen to mundane sounds like a bus passing by or a working dishwasher;
3. try to identify how many different sounds can be heard in a bar;
4. changing listening positions such as passive versus active or critical versus empathic;
5. follow RASA (Receive, Appreciate, Summary, and Ask) in a conversation.

The crisis in listening is so severe nowadays that the same author in his 2011 Ted Talk about this topic gathered more than 10 million views on YouTube since then [15]. Hence, it is time to sharpen our hearing and try to listening better.

2.2 Networking

Current scientific agenda aims to provide answers to complex societal problems, such as the impact of climate change in the world, the prediction of a new pandemic, or the reduction of social inequality. The complexity of these problems motivates the creation of large research teams, research consortia, or networks, in which people with different expertise converge. In this regard, statisticians are sought as strategic partners of these enterprises, because they can help with the design of a project and

Table 1 Active listening skills according to Robertson [13]

Attentive body language
Posture and gestures showing involvement and engagement
Appropriate body movement
Appropriate facial expressions
Appropriate eye contact
Non-distracting environment
Following skills
Interested *door openers*
Minimal verbal encouragers
Infrequent, timely and considered questions
Attentive silences
Reflective skills
Paraphrase (check periodically that you've understood)
Reflect back feelings and content
Summarize the major issues

deliver advanced statistical analysis that is typically out of reach of non-statistically-trained researchers. However, working in such multidisciplinary environments can be challenging and overwhelming for statisticians, because they need to interact with other researchers often enough to negotiate different strategic decisions for the course of a project. The development of networking skills is then necessary.

These skills consist in developing an interpersonal intuition on how different members of a research team fit together in order to understand team dependencies, responsibilities, and dynamics. For example, in a large epidemiological study, statisticians are typically asked to join forces with epidemiologists, mathematical modellers, and bioethical experts. Statisticians can increase visibility by talking to each of these colleagues in order to decide on the best study design. In a sentence, higher visibility comes when a statistician is a team player and sets the team's goal as his/her top priority.

Networking skills are also mandatory for choosing collaborators wisely. Like ice-creams, collaborators come in different flavours and, therefore, statisticians who intend to be treated as equal should create a network of collaborators who share the same principles, attitudes, and ambitions. When it comes to evaluating the success of a given collaboration, statisticians should weight the immediate research output (i.e., a high-impact paper or a funded project) against the sense of mutual respect, harmony, and sustainability in the long term. One cannot forget that, given the high demand for statistical services in academia and elsewhere, statisticians have all the autonomy and power to choose and embrace only durable and harmonious collaborations with their non-statistical peers.

The important question is then to know how to improve networking skills. Besides taking formal training, statisticians can also join a professional society such as the

Portuguese Statistical Society in Portugal, the Royal Statistical Society in the United Kingdom (UK), or the International Biometrical Society. Active citizenship in these societies allows statisticians to find and connect easily with other professionals with the same research interests. Alternatively, statisticians can make an effort to seek networking opportunities outside the field of Statistics. For example, being a member of COST actions funded by the European Union is a unique opportunity for statisticians to increase their network of collaborators across Europe. In the UK, the Academy of Medical Sciences and the Royal Society offer specific funding for creating new networks between UK-based and overseas researchers.

2.3 Effective Communication

The primary objective of any act of verbal communication is to create understanding from what is being said and heard. The same objective is also expected when communication takes the form of the written word. Effective communication goes beyond this basic objective by aiming to create impact, to generate action, or to motivate change.

In the case of applied statisticians, effective communication is likely to come in the shape of presenting or writing the results of a statistical analysis to a non-specialized audience. In this scenario, impactful communication should not be understood as the delivery of catchy and simple soundbites or keywords, or just speaking to the audience's emotions, or even more so sacrificing technical accuracy and rigour. Impact should be seen in a broader sense in which the target audience understands the results and the respective implications clearly. Impactful communication also sets the scene for statisticians to manage expectations and negotiations that might occur during the lifetime of a project or collaboration. Unsurprisingly, there is no magical solution for effective and impactful communication. However, some of the tips below are extremely useful for a scientist in general to learn, practice, and improve.

In the current work and scientific culture of frequent meetings and conferences, effective and impactful verbal communication is intimately related to delivering a good talk or presentation. In this regard, the TED curator Chris Anderson provides a set of tricks for public speaking [16]. According to this author, delivering a decent talk is at the reach of everyone's hand. Delivering a talk in the format of a story is always a very compelling way to fuel people's imagination. Stories are also easy to follow and natural for all of us given that we learn life through stories since childhood. Simplify the message and never underestimate the power of rehearsing are two other tips for effective talks. The Indian Yoga's master Sadhguru [17] also provides a very useful advice for speaking in general:

> *See if you can articulate the same things that you are saying with half the number of words. Suddenly you will become extremely conscious of everything.*

When preparing slides supporting a presentation, the humorous and bold David Phillips [18] advises five tips *to avoid death by powerpoint*:

1. one message per slide to increase focus of the audience and avoid distraction amongst competing content;
2. avoid the use of text to reduce the mental strain of listening and reading simultaneously;
3. increase the size of key objects to maximize their readability and interpretability by the audience;
4. use contrast of colours to guide people's attention;
5. use a maximum of 6 objects per slide to minimize the time for the audience to grasp what is on each slide.

From the above five tips, avoiding text should be the mantra for any public speaker including a statistician. In fact, slides with insane amounts of text might be one of the deadliest sins in public speaking. It can give the impression that the text is not there for the audience to read, but for the speaker not forget what to say. As a consequence, one might feel that the speaker is neither prepared, nor confident, nor comfortable in his/her shoes. While lack of preparation suggests some sort of disrespect for the audience, lack of confidence might generate empathy to some listeners; after all, we all have been out there, exposed in front of the audience with sharp eyes, but it ultimately generates pity rather impact. Reducing the amount of text has the benefit of creating the right motivation for a speaker to be brief and simple, and to rehearse the presentation. We should never forget that the speaker and what he/she is saying are the main focus of a talk. By logic, if the speaker wishes the audience to read from slides, why is he/she there?

For effective writing in scientific papers and reports, Ehrenberg [19] suggests the following guidelines:

1. to start at the end (or focus on findings first);
2. be prepared to revise;
3. cut down the long words;
4. be brief;
5. think of the reader.

Presenting and discussing the results first is the writing format that some scientific journals such as the Nature-branded and PLoS journals are adopting nowadays; the Materials & Methods section where statisticians feel more comfortable, is typically placed at the end of the paper or, in some extreme cases, buried in an online supplementary material. This writing format might be challenging for statisticians given their natural inclination and enthusiasm for methodological issues. However, at the same time, this inclination and enthusiasm should not be totally silenced, because providing detailed information about the statistical methodology is a moral and ethical obligation that promotes scientific replication and reproducibility [3].

To be prepared to revise is the joyful art for some or the painstaking task for others of making tweaks and adjustments to the text for better readability. This task is intimately related to be brief and cut down unnecessary jargon which is typically encapsulated in long words. It takes an underestimated number of iterations, specially, by students and early-career statisticians. The revision of a paper written in English might be challenging for non-native speakers. In this case, one should operate in a benevolent regime of *practice makes perfect*; that and a lot of patience. Finally, thinking of the reader helps deciding the level of (statistical) detail that statisticians can dive in a report or paper.

If one seeks to master the art of effective (oral) communication to inspire others, Simon Sinek [20] proposes a simple but useful concept: the golden circle (Fig. 1). The way that we communicate on the daily basis is by progressing from *what*, *how*, and *why*. For example, if one aims to present the content of this paper in a conference, the traditional way to start the presentation could be the following:

> Today I will share with you some tips that I have learned about statistical leadership. These tips are related to active listening, networking, and effective communication. I will first define what they mean and then tell you how you can improve them. I hope all of these tips are useful for you and your future career.

This is the natural way of communicating for most of us, because we start from the most precise to the most vague piece of information. That is, (almost) everyone knows what he/she is supposed to do, some really know how to do it, but only a few know the reason for what they are doing. This way of communication is not necessarily ineffective per se, but fails to deliver impact to the audience; we listen to the same communication format over and over again and, therefore, boredom might set in with these repetitions.

According to Simon Sinek, simply reversing the order of the information given generates more impact on the listener. Let's come back to the above example. After some tweaking, one could alternatively start the presentation like this:

> Big data are the brave new world. However, we statisticians remain hidden in the shadow of this world of wonder. We are simply seen as data crunchers or p-value providers. We

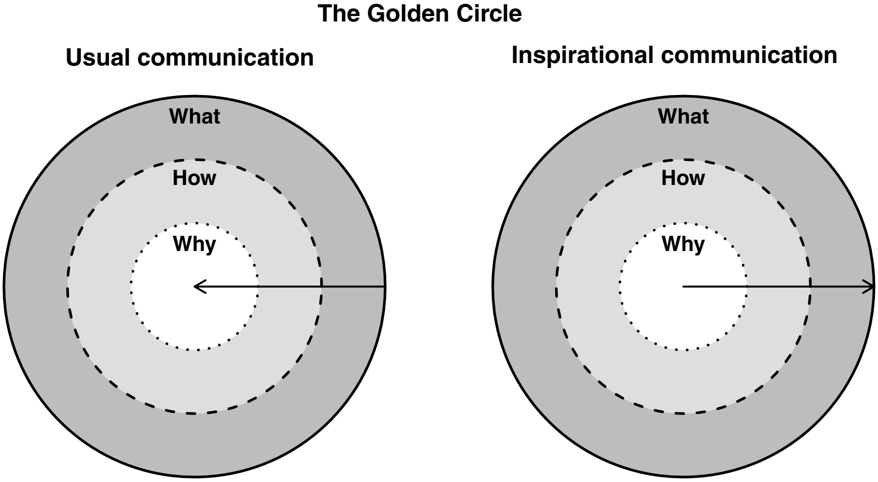

Fig. 1 The golden circle of communication by Simon Sinek [20]: usual communication travels from what, how, and why, while inspirational communication does the opposite

> should change this narrow view of our profession. How to operate this change? By thinking of statistical leadership and how we can develop it. Today I will give you some tips about active listening, networking, and effective communication, which we can use routinely in our profession. In the end, if we are little ambassadors of statistical leadership, everyone wins.

The clear articulation of *why* talks to our emotions and, as we all know it, they are capable to convince us to do wonderful things. However, for this communication approach to work, there is the challenge of knowing the reason of what we are doing. In this regard, the philosophical discussion about the role of a statistician, the duality between a scientist and shoe clerk as smartly put it by Bross [6], helps to solve this challenge. If a statistician acts like a shoe clerk, it is difficult to inspire anyone by simply pleasing the customer, getting a salary at the end of the month, or climbing up the academic ladder. In contrast, if a statistician considers him/herself as a scientist first of all, it is much easier to find a purpose for being part of a given project. After all, a scientist is a curious person about the world and an eternal chaser of the truth.

3 Increasing Visibility in Academia

In theory, the academic recognition and visibility of a statistician should be in a direct correlation with the publication record and the amount of funding awarded. Hence, any attempt to increase the number of publications and funding awarded are straightforward steps towards a higher recognition and visibility of statisticians in academia. In practice, there are other factors that one must consider.

Firstly, increasing the number of publications might require to extend the number of collaborators and projects involved. Managing different collaborators and projects might imply to become a *slave* of them. This can dramatically reduce the time dedicated to pursuit personal research interests. Therefore, applied statisticians should find their optimal balance between their own projects and statistical consultancy activities.

Secondly, funding opportunities for the development of statistical methodologies are scarce and, when available, are often shared with mathematical modellers, mathematicians, bioinformaticians, and data scientists. Given this scenario and the multidisciplinary nature of the statistical exercise, applied statisticians could try to widen their research interests beyond statistical methodology; genetics and climate changes are just two examples of scientific areas where a deep knowledge of statistics and statistical modelling is a requirement. Such a widening of the research agenda increases the chance of getting a project funded and provides an opportunity for statisticians to lead a project. Leading a project increases the visibility of the respective leader irrespective of the scientific area. In this regard, we should follow the footsteps of the great statisticians of the past who made remarkable contributions outside the field of Statistics: Ronald Fisher-population and quantitative genetics; Karl Pearson-biometrics; Egon Pearson and Walter Shewhart-quality control; Francis Galton-psychometrics; amongst others.

There are also less conventional ways that statisticians can use to increase their visibility in academia. Statisticians can find inspiration in a recent review of global travelling and infectious diseases using James Bond's movies as case studies [21]. At the time of writing, this review gathered almost 3,500 likes and 2,000 mentions on Twitter. The infamous Christmas edition of the prestigious British Medical Journal also offers the publication of formal and rigorous scientific enquiries to quirky, light-hearted, or funny biomedical questions, including the estimation of teaspoons disappearance from shared kitchens in a research institute [22], risk estimation of neck and head injuries in heavy metal lovers [23], or the reporting of side-effects in sword-swallowing [24]. A final example comes from James Carlisle who gained the nickname of *data detective* [25]. This Englishman is a trained anesthesiologist but spends his part-time screening the biomedical literature for unusual statistical consistency, data fabrication, and statistical anomalies [26, 27]. Of course, his hobby does not make him particular popular amongst the targets of his investigations [28]. However, his sleuth efforts were not left unnoticed by the research community and hopefully, they served the purpose of raising awareness on the statistical problems in the existing literature while promoting better science and better use of statistical methodology.

4 Increasing Visibility in Society

Imagination, creativity, and personal motivation are the only limits that can hold someone's back in the track of increased visibility in society. For example, statisticians can embrace the technological revolution in mass communication provided by the internet. Social media platforms such as Facebook or Twitter offer quick and cheap ways to disseminate research findings amongst collaborators, colleagues, family, and friends. These platforms also provide an informal forum of discussion between researchers and the general public who ultimately fund research through taxes. The production of podcasts dedicated to disseminate scientific ideas are also gaining popularity in different corners of science [29]. In this regard, the podcast called *The Effective Statistician* by Alexander Schacht [30] helps statisticians to improve efficiency at the workplace, to think more strategically about their career, and to appreciate leadership and negotiation skills.

An interesting opportunity to increase visibility amongst the youngsters is provided by the journal *Frontiers for Young Minds*. The journal publishes conceptual papers to be read by the young ones. The peer-review process is conducted by a young reviewer, but under the guidance of a professional scientist. The *modus operandi* of the journal offers the chance of disseminating statistical ideas and promoting their use amongst the young readers, as the case of Sendef and Robbins [31], who explored the concepts of population, statistics, and probability. The review was done by Joseph of 12 years of age with the help of Jonathan Montaño from the New Mexico State University.

5 Concluding Remarks

This paper discusses some useful skills with the potential of increasing the visibility and the potential of a statistician at the individual level. These skills require permanent and, if needed, formal training. Unfortunately, traditional statistical courses are mainly focused on the hardcore technical skills even if a successful statistician is required to master interpersonal skills given the translational nature of the statistical exercise. Therefore, there is a mismatch between the formal training of Statistics at the university and the prerequisites for a successful career in academia and elsewhere, namely, in the long-term. It is then advised for future or even established statisticians to seek opportunities for improving their interpersonal skills. The acquisition and practising of these skills will make them more prepared, more comfortable, and more confidence to go beyond the "shoe-clerk"-type mindset.

The underlying assumption of this discussion is that a statistician reached a certain *badge* of statistical proficiency when it is reasonable to think of leadership and visibility. This badge does not necessarily mean a world-class recognition of someone's achievements in terms of statistical methodology and modelling. It only means a level of understanding of what a statistical analysis is and what it entails. In other words, statistical leadership and visibility come naturally when a statistician understands not only the methodology, but also the *big picture* beyond the remit of a given statistical analysis. In this scenario, early-career statisticians would find themselves less inclined to invest time in developing leadership skills and taking the necessary steps towards a more impactful career. This comes more naturally to mid-career statisticians who were already involved in enough collaborations and projects, and therefore, have a better idea of the pros and cons of the statistical profession. However, it is important to emphasize again that leadership and visibility are personal choices and, as such, every one of us should make an introspective exercise at least once in a lifetime to answer the question whether statistical leadership is a sufficiently appealing or attractive journey to take. At the end of the day, a career of any professional should be joyful.

Statisticians with the intention to increase their (professional) influence should be aware of two possible psychological roadblocks. The first one is that leadership and increased visibility should be perceived as journeys rather than goals. These journeys require a great amount of patience, persistence, and resilience. These personal capacities typically clash with current culture of instant gratification and constant pursuit for impact. Anxiety might come along the way. If such happens, statisticians should make a step back and revaluate their situation. The second roadblock is the so-called impostor phenomenon. In this phenomenon, people express self-doubt on their accomplishments and skills, despite factual evidence or other people indicating otherwise. People who suffer from this phenomenon believe that their success is due to some kind of luck or error, and they live in constant fear of being unmasked as unintelligent or less capable. These impostor feelings can diminish career planning, career prospective, and the motivation to lead [32]. Therefore, it is possible that future highly visible statistical leaders should feel something similar. In that case, statisticians should embrace these feelings as a motivation and an opportunity for self-improvement and not for self-doubt.

The final remark is to make a clear distinction amongst individual, organizational, and policy levels of statistical leadership and visibility, as discussed by Gibson [8]. In this scenario, the present paper mainly focused the discussion at the individual level. This level relates to small research groups and day-to-day interactions between a statistician and his/her colleagues or collaborators. Statistical visibility and leadership at the organizational level is related to the situation where the influence of a (senior) statistician or a group of them aims to be felt at the level of a given institution, such as company or research consortium. This influence can take the form of trying to change a given statistical practice amongst all members of the same institution. In turn, statistical leadership and visibility at the policy level is operated by statisticians who sit at technical advisory committees representing different stakeholders. For example, statisticians together with epidemiologists, medical doctors, nurses, and other health staff might be put together to discuss with the national health authorities whether an existing policy needs to be change or whether a new policy needs to be created at the light of new data. An interesting example of statistical leadership at this level is the discussion around the salt consumption and health held by the Institute of Medicine (currently, named National Academy of Medicine) from the USA provided by Nancy Cook [33]. Another example is given by the statistician Mike Campbell who works on the NICE appraisal committee in the UK [34]; NICE is the agency that decides which new therapies should be allowed in the British National Health System. These two levels of statistical leadership are more challenging than the individual one, and require a deeper discussion of other skills (e.g., negotiation, conflict management, and mediation skills) that are beyond the scope of this paper. A more extensive discussion about these two levels of statistical leadership can be found in Gibson [8].

Acknowledgements The author would like to thank João Malato (Instituto de Medicina Molecular, Lisbon, Portugal), Emanuele Giorgi (Lancaster University, Lancaster, UK), Francisco Westermeier (FH Joanneum University of Applied Sciences, Graz, Austria) and Alexander Schacht (The Effective Statistician, Germany) for proof-reading the manuscript and providing insightful comments and suggestions. The author acknowledges funding from the Fundação para a Ciência e Tecnologia, Portugal (ref. UIDB/00006/2020), and the Polish National Agency for Academic Exchange, Poland (ref. PPN/ULM/2020/1/00069/U/00001).

References

1. Harari, Y.N.: Homo Deus: A Brief History of Tomorrow. Vintage, London, London (2017)
2. Leonelli, S.: Data-from objects to assets. Nature **574**, 317–320 (2019)
3. Stark, P.B., Saltelli, A.: Cargo-cult statistics and scientific crisis. Significance **15**, 40–43 (2018)
4. Wasserstein, R.L., Lazar, N.A.: The ASA statement on *p*-values: context, process, and purpose. Am. Stat. **70**, 129–133 (2016)
5. Fisher, R.A.: Presidential address (1933–1960). Sankhya: The Indian J. Stat. **4**(1), 14–17 (1938)
6. Bross, I.D.J.: The role of the statistician: scientist or shoe clerk. Am. Stat. **28**, 126–127 (1974)
7. Mongeon, P., Smith, E., Joyal, B., Lariviére, V.: The rise of the middle author: Investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. PLoS One **12**, e0184601 (2017)
8. Gibson, E.W.: Leadership in statistics: increasing our value and visibility. Am. Stat. **73**, 109–116 (2018)
9. Sepúlveda, N.: Ser ou não ser um líder (estatístico)?. In: Boletim SPE Primavera, pp. 39–48, SPE Editions (2022)
10. Peterson, J.B.: 12 Rules for Life: An Antidote to Chaos. Penguin Allen Lane, London (2021)
11. Bunkers, S.S.: The power and possibility in listening. Nurs. Sci. Q. **23**, 22–27 (2009)
12. Mandela, N.R.: Long Walk to Freedom. Hachette Book Group, New York (1994)
13. Robertson, K.: Active listening: More than just paying attention. Aust. Family Phys. **34**, 1053–1055 (2005)
14. Treasure, J.: How to be Heard: Secrets for Powerful Speaking and Listening. Mango Publishing Group, Coral Gable (2017)
15. Treasure, J.: 5 ways to listen better. https://www.ted.com/talks/julian_treasure_5_ways
16. Anderson, C.: TED Talks: The Official TED Guide to Public Speaking. Houghton Mifflin Harcourt, Boston (2016)
17. Sadhguru: The importance of silence
18. Phillips, D.J.P.: How to avoid death by powerpoint. https://www.youtube.com/watch?v=KpMbR7WCLXI
19. Ehrenberg, A.S.C.: Writing technical papers or reports. Am. Stat. **36**, 326–329 (1982)
20. Sinek, S.: Start with Why: How Great Leaders Inspire Everyone to Take Action. Portfolio, London (2009)
21. Graumans, W., Stone, W.J., Bousema, T.: No time to die: An in-depth analysis of James Bond's exposure to infectious agents. Travel Med. Infect. Dis. **44**, 102175 (2021)
22. Lim, M.S.C., Hellard, M.E., Aitken, C.K.: The case of the disappearing teaspoons: longitudinal cohort study of the displacement of teaspoons in an australian research institute. BMJ **331**, 1498–1500 (2005)
23. Patton, D., McIntosh, A.: Head and neck injury risks in heavy metal: head bangers stuck between rock and a hard bass. BMJ **337**, a2825–a2825 (2008)
24. Witcombe, B., Meyer, D.: Sword swallowing and its side effects. BMJ **333**, 1285–1287 (2006)
25. Adam, D.: How a data detective exposed suspicious medical trials. Nature **571**, 462–464 (2019)
26. Carlisle, J.B.: The analysis of 168 randomised controlled trials to test data integrity. Anaesthesia **67**, 521–537 (2012)
27. Carlisle, J.B.: Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. Anaesthesia **72**, 944–952 (2017)
28. Fujii, Y.: The analysis of 168 randomised controlled trials to test data integrity. Anaesthesia **67**, 669–670 (2012)
29. Kwok, R.: How to make your podcast stand out in a crowded market. Nature **565**, 387–389 (2019)
30. Schacht, A.: The effective statistician podcast. http://theeffectivestatistician.com/podcast/
31. Sendef, J., Robbins, A.: How scientists use statistics, samples, and probability to answer research questions. Front. Young Minds **7**, 118 (2019)
32. Neureiter, M., Traut-Mattausch, E.: An inner barrier to career development: preconditions of the impostor phenomenon and consequences for career development. Front. Psychol. **7** (2016)

33. Cook, N.R.: Salt: how much less should we eat for health?: understanding the recent IOM report. Significance **10**, 6–10 (2013)
34. Campbell, M.: A statistician on a NICE committee. Significance **7**, 81–84 (2010)

A Robust Hurdle Poisson Model in the Estimation of the Extremal Index

Manuela Souto de Miranda, M. Cristina Miranda, and M. Ivette Gomes

Abstract In statistical extreme value theory, the occurrence of clusters of exceedances above a high threshold is related to the extremal index (*EI*), when that parameter exists. In such cases, the *EI* represents the reciprocal of the mean cluster dimension in the limit distribution. The set of observed cluster sizes may contain too many zeroes, depending on the scheme used in the identification of the clusters and posterior estimation process, as it happens with the Blocks estimator. We consider the estimation of the mean cluster size by modelling the clusters dimension with a hurdle zero truncated Poisson regression model. The goal is to find a robust estimator with a good performance along increasing quantiles and computationally user friendly. The paper highlights the importance of the last question also, since many statisticians use or do not use some methods, depending on the free software devoted to the method and respective confidence in their optimization procedures and results. A simulation study explores and compares different proposals.

Keywords Blocks estimator · Extremal index · Hurdle model · Robustness

M. Souto de Miranda (✉)
CIDMA, University of Aveiro, Aveiro, Portugal
e-mail: manuela.souto@ua.pt

M. C. Miranda
ISCA and CIDMA, University of Aveiro, Aveiro, Portugal
e-mail: cristina.miranda@ua.pt

CEAUL, University of Lisbon, Cidade Universitária, Campo Grande, Portugal

M. I. Gomes
Faculty of Science of Lisbon (FCUL/DEIO) and CEAUL, University of Lisbon, Cidade Universitária, Campo Grande, Portugal
e-mail: ivette.gomes@fc.ul.pt

R. Bispo et al. (eds.), *Recent Developments in Statistics and Data Science*,
Springer Proceedings in Mathematics & Statistics 398,
https://doi.org/10.1007/978-3-031-12766-3_2

1 The Extremal Index

1.1 Motivation

There is a great interest in modelling extreme values, particularly when they represent the exceedance of high thresholds. The theory is extensively developed for extremes in the independence framework. Nevertheless, many phenomena are more realistically modelled by the occurrence of clusters of extreme values than assuming a scenery of isolated independent ones. That is the case with heat or cold waves, extremely rainy days, price crashes in the stock market and so on. The duration of those phenomena can be traduced by a counting process that represents the cluster size, whose mean is related to the *EI*, when it exists. Thus, the *EI* estimation procedure deserves a great practical interest. But it is not enough to look for a method with good mathematical properties from a classical point of view. The procedure must be robust, in the sense that gives good estimates in the assumed model and, simultaneously, it is not very sensitive to small deviations from the model assumptions, for instance, gross error values or even the functional form of the cluster size distribution. Robust estimation theory has been widely investigated for location and regression models, particularly with continuous distributions. With respect to counting processes, the research is still very active nowadays.

Another important point of view is the existence of computation facilities that allow easy access to the *EI* estimates, either in individual case studies or in simulations. Computational techniques cover two main fields: the numerical questions, since many estimators depend on complex optimization problems that become more evident in simulation environments; and open access software tools, like the *R* platform with its packages, which are already programmed in a devoted way and well tested by investigators. It is also desirable that software is user friendly, so that it can be used by statisticians in general. Those aspects are essential to the success of the *EI* estimation process (and others), mainly outside the more popular Gaussian and independence scenarios.

1.2 Theoretical Introduction

Assume a strictly stationary sequence of random variables $\{X_n\}_{n\geq 1}$, from a *cumulative distribution function* (*CDF*) denoted by F, under general asymptotic and long-range dependence restrictions, like the long-range dependence condition **D** (see [1]) and the local dependence condition **D**" (see [2]). Let $\{X_{i:n}\}_{n\geq 1}$, $1 \leq i \leq n$, denote the associated sequence of ascending order statistics.

The stationary sequence $\{X_n\}_{n\geq 1}$ is said to have an *EI*, θ, with $(0 < \theta \leq 1)$, if for all $\tau > 0$, we can find a sequence of levels $u_n = u_n(\tau)$ such that, with $\{Y_n\}_{n\geq 1}$ the associated *independent, identically distributed* (*i.i.d.*) sequence (*i.e.*, an *i.i.d.* sequence from the same CDF F),

$$P\,(Y_{n:n} \le u_n) = F^n(u_n) \underset{n\to\infty}{\longrightarrow} e^{-\tau}$$

and

$$P\,(X_{n:n} \le u_n) \underset{n\to\infty}{\longrightarrow} \exp^{-\theta\tau}\,.$$

Since $0 < \theta \le 1$, there is thus a *'shrinkage'* of the values in the limit *CDF*, but after linearly normalized, $X_{n:n}$ has still an *extreme value* (*EV*) distribution, with a *CDF* with a functional form of the type

$$\mathrm{EV}_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\},\ 1+\xi x > 0, & \text{if } \xi \neq 0 \\ \exp(-\exp(-x)), x \in R, & \text{if } \xi = 0. \end{cases} \tag{1}$$

Under the two mixing conditions **D** and **D**”, the *EI* can also be defined as:

$$\theta = \frac{1}{\text{limiting mean size of clusters}} = \lim_{n\to\infty} P(X_2 \le u_n | X_1 > u_n),$$

with

$$u_n:\quad F(u_n) = 1 - \tau/n + o(1/n),\quad \text{as } n \to \infty,\ \text{with } \tau > 0,\ \text{fixed.} \tag{2}$$

The m-dependent (m-dep) processes are used here for illustration. It is known that for these processes the *EI* is given by $\theta = 1/m$. They may be based on *i.i.d.* Fréchet (ξ) random variables Y_i, $i \ge 1$, from a *CDF* $\Phi_\xi^{1/m}$, with $\Phi_\xi(x) = \exp\big(-x^{-1/\xi}\big)$, $x \ge 0$, the standard Fréchet *CDF*. They are then built upon the relation $X_i = \max_{i\le j\le i+m-1} Y_j$, $i \ge 1$. An illustration of clustering of high values with an asymptotic mean size equal to m, is presented in Fig.1 as illustrated in [12].

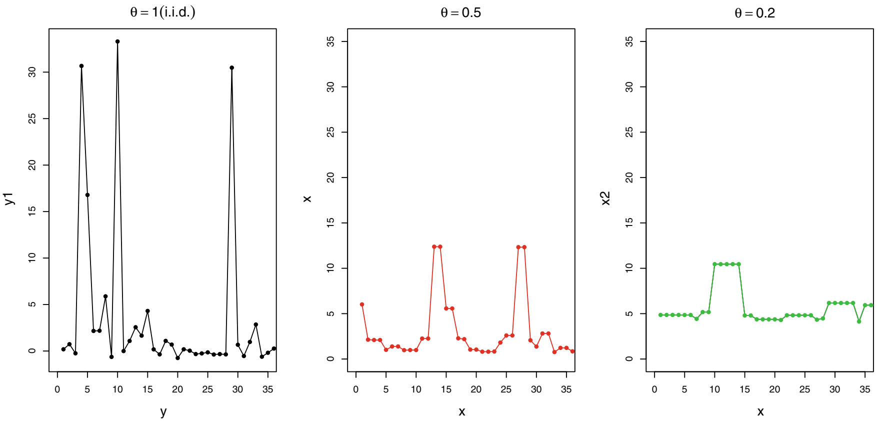

Fig. 1 Sample paths of an *i.i.d.* (left), 2-dep (center) and 5-dep (right) processes from the same underlying Fréchet ($\Phi_{\xi=1}$), but with *EI*s, respectively, equal to 1, 0.5 and 0.2 [12]