Educating Professional Statisticians: Some thoughts from Brazil

Pedro Luis do Nascimento Silva
Pedro-luis.silva@ibge.gov.br

Escola Nacional de Ciências Estatísticas
(ENC/IBGE)
The Profession & The Challenges

**Big data** and the associated revolution in data availability, accessibility, speed of production and pressure to ‘do something’ with the available data.
**Volume**

- Scale of Data
- 40 Zettabytes (40 trillion gigabytes) of data will be created by 2020, an increase of 300 times from 2005
- 6 billion people have cell phones

**Velocity**

- Analysis of Streaming Data
- By 2016, it is projected there will be 18.9 billion network connections—almost 2.5 connections per person on earth
- The New York Stock Exchange captures 1 TB of trade information during each trading session
- Modern cars have close to 100 sensors that monitor items such as fuel level and tire pressure

**Variety**

- Different forms of data
- As of 2011, the global size of data in healthcare was estimated to be 150 exabytes (150 billion gigabytes)
- 4 billion+ hours of video are watched on YouTube each month
- 30 billion pieces of content are shared on Facebook every month
- 400 million tweets are sent per day by about 200 million monthly active users

**Veracity**

- Uncertainty of Data
- In one survey, 27% of respondents were unsure of how much of their data was inaccurate
- By 2011, 4.4 million IT jobs will be created globally to support big data, with 1.9 million in the United States

**Source:**

http://www-01.ibm.com/software/data/bigdata/
The Profession & The Challenges

Modern approaches where the ‘art and science’ of data analysis, modeling and inference are being built up into software:

– Machine learning;
– Artificial intelligence;
– Data mining;
– Etc.
The Profession & The Challenges

Rapidly evolving software, which aims to be less dependent on the skills of the users;

Source:
The Profession & The Challenges

• **Increasing costs** of traditional in-person education.

• **Fast evolution** of methodology and technology, demanding time and effort to keep up with the developments.

• Ever **wider areas of application** of Statistics with their own dialects and promotion of ‘DIY’ by users:
  – Often lacking broader statistical education,
  – but using complex and specialized statistical tools (models, methods and software) relevant to their fields.
Professional Education (in Brazil)

Three levels of higher education:
- BSc (4 years) – 34 institutions
- MSc (2 years) – 9 institutions
- Doctorate (4 years) – 7 institutions

Two main routes for statistical education:
- BSc. in Statistics
  - Undergraduate degree in other areas + graduate degree in Statistics

- Scientific culture
- Communication skills
- Design & planning for data collection & measurement
- Data organization and processing
- Numerical and graphical synthesis of data
- Statistical modelling and analysis
- Proposing decisions based on data analysis
- Critical capacity & ability to work in multidisciplinary teams
- Managerial ability

Professional Education (in Brazil)

BSc degrees, offered mostly by public institutions (federal or state universities).

‘Traditional’ educational approach:

– First teach (lots of) maths;
– Then probability & mathematical statistics;
– Then teach modelling & methods (linear models, design of experiments, time series, multivariate analysis, etc.);
– Not enough computing; not enough practice...
Table 1 – Some data for Statistics Undergraduate Degrees – Brazil, 2013

<table>
<thead>
<tr>
<th>Item</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Places offered</td>
<td>1.798</td>
</tr>
<tr>
<td>Demand</td>
<td>13.667</td>
</tr>
<tr>
<td>Intake</td>
<td>1.572</td>
</tr>
<tr>
<td>Enrollment</td>
<td>4.796</td>
</tr>
<tr>
<td>Graduates</td>
<td>448</td>
</tr>
<tr>
<td>Demand / Place</td>
<td>7,6</td>
</tr>
<tr>
<td>Graduates / Place</td>
<td>24,9%</td>
</tr>
</tbody>
</table>

Source: INEP
Undergraduate Assessment

• Exam carried out in 2012 with BSc. graduates.
• Exam covered core curriculum topics described on national undergraduate guidelines.
• 20 institutions participated.
• 17 out of 20 institutions got average grades below 60%!
ASA Guidelines for Undergrad. Majors

- Key points:
  - Increased importance of data science.
  - Real applications + more practice.
  - More diverse models and approaches.
  - Ability to communicate.

http://www.amstat.org/education/curriculumguidelines.cfm

- These imply the need to rethink the curriculum as well as the modes of teaching / learning.
‘Professional Statistician’

• Getting more and more difficult to define.
• Challenges from emerging ‘areas’ / trends:
• Implications for job naming:
  – Data Scientist;
  – Data Analytics Consultant;
  – Digital Data Analyst;
  – Research / Marketing Web Analyst;
  – Research Statistician and Data Manager;
  – Etc.
Professional Statisticians

• Why look at accreditation: for a definition of what is required to become a ‘Professional’ (US) or ‘Chartered’ (UK) Statistician.

• Accreditation is offered in some countries.

• US and UK adopt voluntary accreditation schemes.

• Brazil adopts compulsory accreditation.

• Accreditation based solely on completion on BSc. Degree in Statistics!
PStat® Scheme in the US

- **PStat® scheme** requires **peer recognition** that applicants have:
  - **Statistical training and knowledge** (MSc. or PhD.);
  - **Experience** in applying the expertise competently (> 5 years since graduating);
  - Maintenance of appropriate **professional development**;
  - Agreement to abide by professional **code of ethics**;
  - Ability to communicate effectively.

Source:
https://www.amstat.org/accreditation/pdfs/Guidelines_for_ASAVoluntary_Professional_Accreditation.pdf
PStat® Educational Requirements

• Advanced degree in Statistics or related quantitative field, with sufficient concentration of Statistics.

• Field may include mathematical or applied Statistics, or major application areas, such as Biostatistics.
GStat® Scheme in the US

- **GStat® scheme** is the entry level of accreditation created by ASA.

- A Graduate Statistician is an individual who has achieved the educational level required for full Professional Statistician status, but has not yet reached the level of experience and expertise required of a Professional Statistician (PStat®).

- **Source:**
  
  http://community.amstat.org/blogs/ronald-wasserstein/2014/03/24/asa-at-175-gstat-accreditation
Professional Statistical Experience

• Undertaking statistical analysis of data and reporting on the results;
• Having responsibility for the interpretation and presentation of statistical information;
• Leading projects with a substantial amount of statistical analysis or modelling;
• Teaching statistics based on practice;
• Work/consulting/collaboration and any resulting teaching of statistics for a field of application;
Professional Statistical Experience

• Working as a statistical consultant;
• Carrying out and implementing research to develop new methods to solve significant applied statistical problems;
• Taking responsibility for the design and analysis of statistically based surveys (or experiments);
• Managing a statistics section with work in one or more of the above areas;
• Generally recognized as having made a substantial contribution to the sound practice of statistics.
Professional Competence

Evidence that work as an applied statistician is of high quality, by demonstrating:

- Substantial positive impact on decision making;
- Appropriately applied methods and techniques;
- Adequate discussion of limitations of the data, methods, techniques;
- Proper attention to accuracy, reliability, relevance, reproducibility and transparency;
- Recognition of applied work by peers.
Some ideas for improving undergraduate programs
Directions for Change

- More flexible curricula, enabling stronger connections to particular fields of application;
- Less compulsory maths courses;
- Increased opportunities for statistical practice;
- Increased emphasis on computing and data management skills;
- Exposure to modern ‘Data Science’ topics not covered in traditional statistical education;
- Develop and practice communication skills.
After graduating, how to ‘keep fit’?
Some approaches to ‘keep fit’

- Read, listen, reach out
- Present and discuss your work
- Write-up your work
- Develop and teach courses
- Supervise, mentor junior colleagues
- Attend conferences
- Join professional or academic society
- Take courses
- Consider distance learning opportunities
- Consult & collaborate
Read, listen, reach out

- Lots going on ‘around’ Statistics.
- Read frequently.
- Read outside of your own area of expertise.
- Review work of others (refereeing, etc.).
- Attend seminars / talks / conferences in other areas, not only in ‘Statistics’.
- Join ‘study groups’ or ‘learning communities’.
Present and discuss your work

• Make sure that you present and discuss your work with peers and with ‘users’.

• Good presentations always raise good questions and discussion, which can lead to improvement and learning.

• They also create opportunities for legitimate challenges to be identified.
Write-up your work

• Make sure that you write-up your work.
• Writing-up helps focus on the important ideas, concepts, methods and results.
• It is also essential to get it more thoroughly reviewed.
• Another good way of creating opportunities for legitimate challenges to be identified.
• Organizations may help with policies on writing and reviewing.
Develop and teach courses

- Teaching a course is a fantastic opportunity to learn / consolidate knowledge.
- It helps you to revisit the topics of the course, broaden your own knowledge.
- It also raises opportunities for questions, which you may use to learn more.
- Make sure your teaching is assessed and take notice of the outcome and comments.
- Organizations may assist with policies to encourage continued professional education.
Supervise, mentor, junior colleagues

- Supervising or mentoring the work of others also provides opportunities for learning.
- When possible, engage formally in such activities.
- Example: joint appointments.
- Mentoring is also great to keep abreast of new trends and perceptions prevailing with the younger generations.
Attend conferences

• Plan for and invest in attending professional conferences.
• They enable opportunities for concentrated exposure to what’s going on and novelties.
• Use them to plan your forthcoming personal development activities, by defining priority areas or topics.
• Exploit their networking opportunities to engage in the other learning activities.
Join a professional society

• Professional & academic societies facilitate all of the above mentioned activities.
• Yet not all professionals join or take part.
• The cost of joining societies is modest, for the services they provide.
• They offer a convenient and focused approach to networking and to be involved.
• They also provide many opportunities for continued education.
Take courses

- Formal courses are a good source of learning.
- They have the benefit of structure and a natural learning community.
- They may lack the precise focus you need.
- But they will help you focus on learning for a well defined period of time.
- They also help with record keeping.
Distance learning

• A powerful new approach for learning (and teaching) Statistics (and many other things).

• Challenge: Statisticians must be more pro-active in the learning process.

• Opportunity: Distance learning makes it possible to learn from the best.

• It enables continued education both at home or at the workplace.
Statistical software

- Professional statisticians must be proficient in at least one major statistical software.
- R seems the current best possible choice.
- Because it is free and open source, in theory you can use it ‘wherever you are’, ‘forever’.
- What if you don’t know R yet?
- Here is a good opportunity to start those plans and actions toward learning something.
Consult & collaborate

- Consulting and collaborating is another good approach to learn.
- You will face new problems.
- You will often need to learn something new to be able to tackle the problems.
- Even if you can solve a problem with the tools you know, you will have learned something about their use in each situation.
Thank you for your attention!
Professional Development

• PStat® accreditation requires maintaining competence in the chosen area of statistical practice.

• Must complete at least 60 hours of professional development each year.

• Must keep and submit records of professional development activities to apply for renewal of accreditation (every 5 years).

➤ Accreditation is NOT permanent.
Professional Development

• What counts towards professional development:
  – Work-based learning;
  – Professional activity;
  – Formal / educational activities;
  – Self-directed learning;
  – Other (??).

• Professionals must have activities in at least three of the categories above.
Save and invest in ‘keeping fit’

- Each professional should take control of their continued professional education.
- Saving and investing in your own continued development does not depend on others.
- Employers may help, but you should not rely solely on others for this.
- Programme your savings, keep to the plan.
- Then enjoy the ‘freedom’ to invest in what you find most profitable or effective.
Societies and what they can do

- Promote the idea that continued professional development is essential.
- Offer all the current activities and innovate to keep up with times.
- Facilitate engagement of professionals from the earliest stages of their education / career.
- Provide focal point for continued professional development.
Employers and what they can do

• Promote the idea that continued professional development is essential.

• Encourage / support activities that lead to continued professional development.

• Invest in the promotion of CPD for their staff.

• Engage with their staff in planning and implementing their CPD.
Teachers and what they can do

• Lead by example.
• Formally discuss ideas of how one can and should take charge of their own CPD.
• Demonstrate the importance of `keeping fit` for a profession that depends heavily on technology and methodology.
• Promote the attitude and develop the skills needed for independent learning.
Above All

• Develop an attitude to learn continuously.
• After all, who can say that what he/she learned at University will suffice for doing Statistics professionally after 20 years?

Source: my ICOTS7 presentation in Salvador.
Source:
http://reason.cs.uiuc.edu/jaesik/research/probabilistic-artificial-intelligence-lab/
Source:
Data Scientist
London, GB-LND | Competitive Salary
Guardian News & Media Ltd. is one of the most successful and innovative media organisations in the world, renowned for leading edge journalism, product and commercial development, and famous for nurturing talented individuals. The Guardian is Committee.

All jobs from: GUARDIAN NEWS AND MEDIA

Lead Data Scientist - Media Analytics
London | £36000 - £39000 per annum
One of the market leaders in media analytics is recruiting a lead data scientist to help them bring analytics to life for their clients and to eventually become a data guru for the agency.

Insight Analyst
London | £26,000 - £32,000 per annum + Excellent benefits
Are you a highly skilled Data Analyst with experience in delivering insights and analytical solutions? Do you use analytics to influence and improve marketing strategies?

Data Mining Manager - Weymouth
Weymouth, Dorset | £35,373 PA
Counter Fraud Skills are currently seeking a Data Mining Manager to work in a permanent role based in Weymouth.

Statistical/ Analytical Modeler
Dublin | Negotiable
Statistical Modeler - Based in central Dublin This position exists to provide consulting/advisory services and to conduct statistical analysis and build predictive models for a variety of performance outcomes such as risk, fraud, and collections for one o

All jobs from: LEXIS NEXIS
Research/ Marketing Web Analyst

Working for this international client to provide data, reports and analysis to inform the strategic decision making.

The role will focus on the provision and analysis of marketing, web and usage data and analytics.

The Guardian

http://jobs.theguardian.com/st/jobs-analytics.html
Captured on 6/Jul/2014.
## Job searches on Google

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Hits (Thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>job statistician</td>
<td>4,220</td>
</tr>
<tr>
<td>statistics jobs</td>
<td>548,000</td>
</tr>
<tr>
<td>jobs in statistics</td>
<td>746,000</td>
</tr>
<tr>
<td>analytics jobs</td>
<td>74,800</td>
</tr>
<tr>
<td>job analytics</td>
<td>187,000</td>
</tr>
<tr>
<td>jobs in analytics</td>
<td>171,000</td>
</tr>
<tr>
<td>jobs data mining</td>
<td>136,000</td>
</tr>
<tr>
<td>jobs in data science</td>
<td>479,000</td>
</tr>
<tr>
<td>jobs big data</td>
<td>837,000</td>
</tr>
</tbody>
</table>

Searched on 7/7/14.
Some tools which can be used

• PDCA cycle
  – Plan, develop, check/assess, adjust/review.

• Apply to personal development strategy.
<table>
<thead>
<tr>
<th>Period</th>
<th>Read, listen, reach out</th>
<th>Present and discuss your work</th>
<th>Write-up your work</th>
<th>Develop and teach courses</th>
<th>Supervise, mentor colleagues</th>
<th>Attend conferences</th>
<th>Join a professional or academic society</th>
<th>Take courses</th>
<th>Consult &amp; collaborate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td></td>
<td></td>
<td>Analysis of complex survey data</td>
<td>PhD Thesis A. Gutierrez</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td></td>
<td>Methodology Dept. Seminar</td>
<td>Survey on Cardiovascular Risk in Adolescents</td>
<td>Introd. Stats for Journalists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>Seminar São Carlos</td>
<td>Household Survey of Drug Consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td>Survey of lan houses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Jun</td>
<td></td>
</tr>
<tr>
<td>Jul</td>
<td>BSc Mon. H. Souza</td>
<td></td>
<td>MSc diss. R. Caumo</td>
<td></td>
<td></td>
<td></td>
<td>ISI, IASS, ABE, ASA, RSS,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td>Data Editing &amp; Imputation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep</td>
<td>Analysis of complex survey data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td>Designing Surveys for Policy Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td>ISI Regional Conf. Malaysia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>