NATIONAL RESEARCH UNIVERSITY

On MOOCs Quality Estimation: a Case of Modern Nonparametric Superiority and Noninferiority Statistical Tests

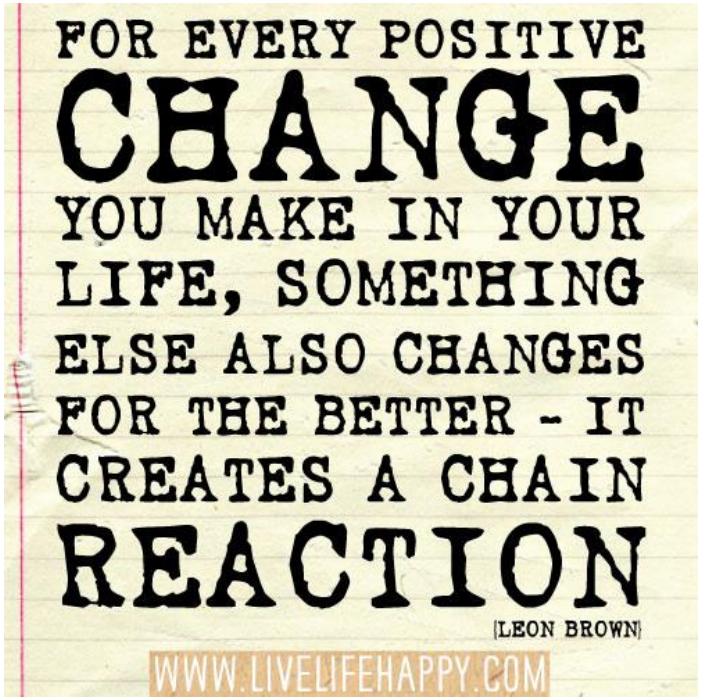
Olga Maksimenkova, SSE, ISSA Lab, FCS NRU HSE

Alexey Neznanov, SDAAI, ISSA Lab, FCS NRU HSE

Maxim Skryabin, Stepik

Motivation

- Everything is changing…
- How changes impact on the quality of a MOOC?



Our goal

- Estimation of dynamic course quality (DCQ) and adjacent topics
  - Dynamic course quality estimation project at ResearchGate (https://www.researchgate.net/project/Dynamic-Course-Quality-Estimation)

Research Frame

Data Science methodology

- Quality assurance (QA)
- Quality estimation
- Educational data mining (EDM)

Dynamic Course Quality (DCQ)

- How to evaluate the content of MOOC
  - Quality
  - Dramatic changes
  - Etc…

Levels of DCQ estimation

- **(first level) quantitative data** about learning actions and statistical analysis
- **(second level) full analysis of course content** in relation to learning outcomes

Methodology

- Non-inferiority analysis of students' progress
- Non-inferiority analysis of course results
- Superiority analysis of courses and audience

Framework

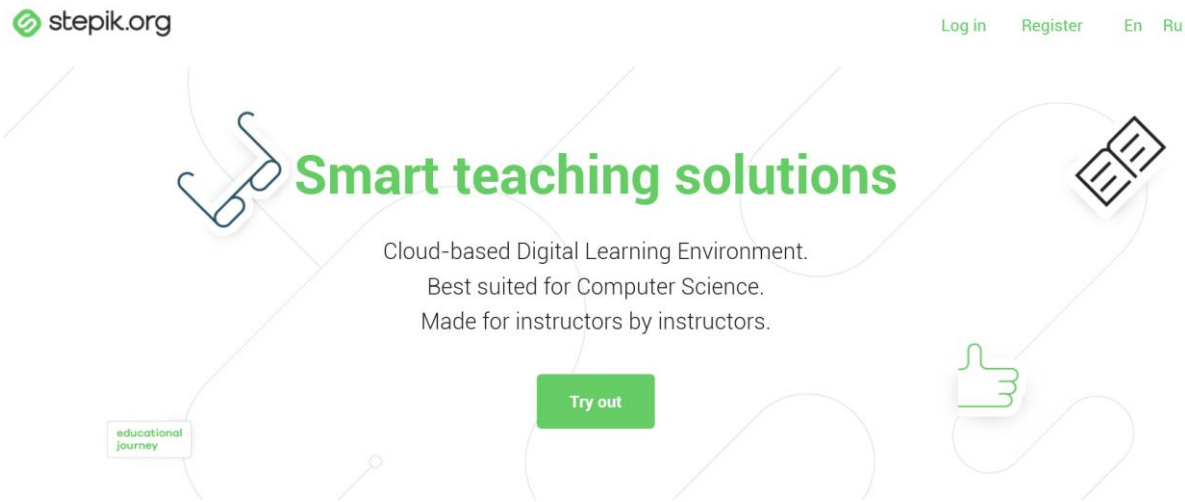
- Blinding
- Randomization
- Experts (teachers) calibration
- Estimating reference level of students' assessment
- Determining non-inferiority margin
- Prior distribution considerations
- Practical significance level

Data gathering – Requirements

1. A repetitive course with consistent endpoint
   - Ideal case is sessions of a course with *known start time* in terms of modern MOOC platforms
2. Data about content change (update events with timestamps)
   - We need to know the *state of content* for each student
3. Data about the student's life cycle (accesses to service with timestamps)
   - We need to be able to *link a student* with the *state of a course*
4. Data about students' progress at the endpoint
   - We need at least an *ordinal scale* for a final student mark and knowledge about conversion from raw scores (for every assignment) into a final mark
   - Standard form of such knowledge is *weighted summations with normalization*

Data gathering – Data

- **Stepik** platform (http://stepik.org)



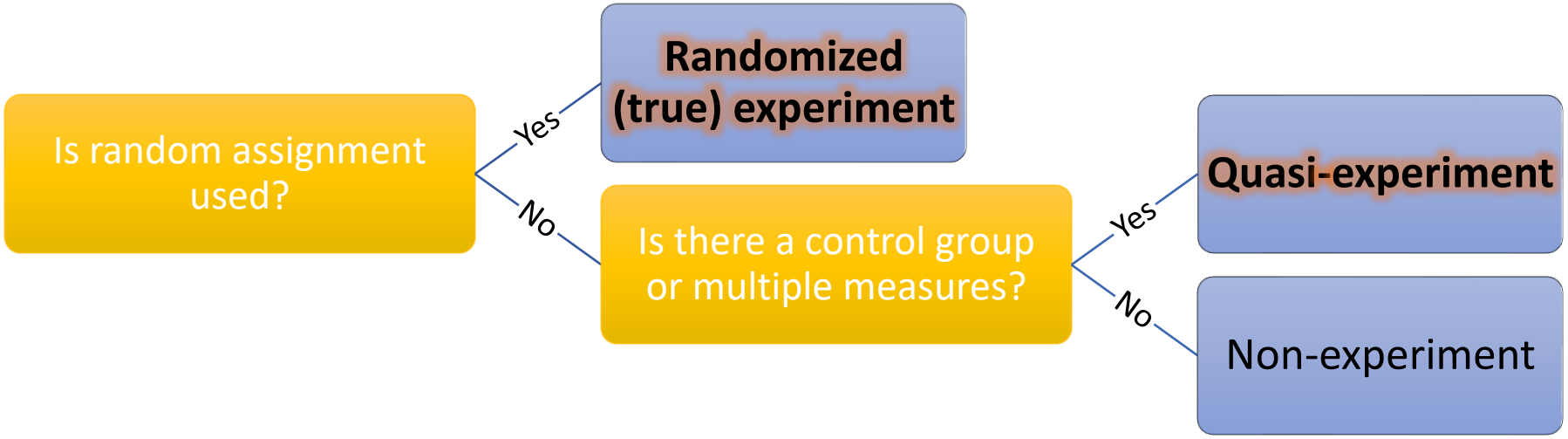
- Pilot data
  - "Introductory Statistics" course (http://stepik.org/course/76)
  - 2 years
  - 10626 students
  - 460179 assignments
  - 787980 submissions

Data gathering – Tools

- *Python* scripts and several *R* packages, like *tidyr* by Wickham
- Microsoft Office 365 and other tools
  - *Access* and *SQL Server* DBMSs
  - *Excel* (especially Get & Transform features)
  - *PowerBI* dashboards for interactive data manipulating and visualization tasks

Methods – Basic experiment design

- We do not have real randomization but have control groups
  - We simulate randomization by random subsample generation



Methods – Trial setup

- Non-inferiority trial design with continuous outcome
  - Non-inferiority margin is defined by experts
- Superiority trial design with continuous outcome

Methods – Nonparametric statistics

- The set of algorithms suggested by Andreas Schulz and improved by Markus Neuhaeuser
  - Ideas from Brunner-Munzel test
  - R implementation
  - Different levels of precision – selection of different permutation tests
- ☑ We do not have any assumptions about distributions of raw scores and success limits!

Methods – Process

- Several intermediate cross-tables (content accesses, submissions' results, etc.) with or without timestamps
  - Any kind of course's event is called "step" ("step_id" is a unique step identifier in the context of a course)
  - We need to align identifiers between sessions in case of content change and build mappings of "step_id" in each session to components of the DCQ model
    - For example, a course's author may *add* a step or *change* an existing step
- Dedicated set of "step_id" with a predetermined weight function constitutes final course score (FCS) and forms the main component of the DCQ model

Methods – Example of cross-table (assessments' results)

user_id	step_id	LastSubmissionTime	Type	1	2	3	4	5
18	18085	14.02.2015 21:25:42	matching	correct				
18	18087	14.02.2015 21:33:51	matching	wrong	wrong	correct		
18	18088	14.02.2015 21:29:26	choice	correct				
18	18089	14.02.2015 21:35:32	choice	wrong	correct			
18	18090	14.02.2015 21:46:16	matching	wrong	correct			
18	18091	14.02.2015 21:43:43	choice	correct				
18	18095	14.02.2015 21:48:07	choice	correct				
18	18101	14.02.2015 21:50:45	choice	correct				
18	18119	14.02.2015 21:55:31	choice	wrong	wrong	correct		
18	18120	14.02.2015 22:11:26	choice	wrong	correct			
18	18121	14.02.2015 22:03:02	choice	correct				
18	18123	14.02.2015 22:12:19	choice	wrong	wrong	correct		
18	18124	14.02.2015 22:13:11	choice	correct				
18	18130	14.02.2015 22:03:23	choice	correct				
18	18145	14.02.2015 23:04:37	choice	correct				
18	18147	14.02.2015 23:15:09	number	wrong	wrong	wrong	correct	
18	18148	14.02.2015 23:15:32	choice	correct				
18	18149	15.02.2015 20:48:45	choice	wrong	wrong	wrong	wrong	correct
18	18174	15.02.2015 21:14:17	choice	wrong	wrong	wrong	wrong	wrong
18	18177	15.02.2015 21:19:54	choice	correct				
18	18181	15.02.2015 21:19:30	choice	correct				
18	18185	15.02.2015 21:11:31	choice	correct				
18	18192	15.02.2015 21:26:32	choice	wrong	wrong	wrong	wrong	wrong
18	18427	14.02.2015 22:48:07	choice	correct				

Preliminary results

- **Approving methodology** and **technology** of *dynamic course quality* (DCQ) estimation based on *quasi-experiments* and *nonparametric statistics*
- **Preparing** *data workflow* for processing courses from MOOCs
- **Confirming** quality of modern nonparametric statistical methods
- **Piloting** several *data visualizations* for experts
- **Discussing** various DCQ functions

Discussion

- Requirements for historical data
- Course structure -> data structure!!!
- Dynamic course quality, students, and teachers
  - DCQ definition
  - DCQ variations in context
  - Noncalibrated students and teachers
    - Calibration procedures!!!