MM 217: Data analysis and interpretation

M P Gururajan and Hina Gokhale

1 August, 2019

Welcome

Welcome to MM 217: Data analysis and interpretation

Contact information

Email id: guru.mp at iitb.ac.in
Please mention MM 217 in the subject.
My office is in third floor, new wing – room number 326.
Office phone number: 7631.

Some administrative stuff

- Instructors: Professor M P Gururajan and Professor Hina Gokhale – along with a team of TAs (Head TA: Mr Sushil Kumar)
- Professor Gokhale is a professional statistician with several decades of experience in dealing with materials data.
- We have to identify a time slot for on-computer hands-on assignments and tutorials. We will do that soon and let you know.
- For simplicity's sake, all administrative aspects of the course: will be handled by Professor M P Gururajan.
- TAs: an essential part of teaching this course; they should be accorded the same treatment as is due to any teacher / instructor; please follow their instructions without fail.

Moodle

- Extensively used;
- All announcements will be made via moodle;
- A day's notice in moodle is assumed to be sufficient;
- Please check moodle announcements at least once a day;
- All material (including this presentation) will be shared on moodle.

Evaluation procedure

- Class participation: 10 points
- Two quizes and computational tutorials/sessions: 20 points
- Mid-semester examination: 20 points
- Final examination: 50 points

Grading policy

- Both AA and FR are based on absolute performance;
- You need 85 points or more to get AA;
- You need 40 points or more to pass the course;
- AB-DD grades will be relative.

Make-up examination

- Only one make-up examination – at the end of the semester;
- The syllabus for the make-up examination will be the entire course material;
- No make-ups for tutorials and other class participation activities.

Attendance policy

- Please attend classes regularly, participate in the class activities and make the classroom environment vibrant and enjoyable.
- Students who participate in class activities can earn upto 10 points.
- Please do not arrive late for the class; be here in class at 10:35 on Mondays, 11:35 on Tuesdays and 8:30 on Thursdays.

Class participation activities

- Will be based on assignment sheets;
- Each student or group of students will be asked to solve a problem or two based on the assignment sheets in the class;
- You have to submit these solved problems to me at the end of the class;
- They will be evaluated towards your class participation.
- Please bring your calculator and pen to every class.

Assignments

- Please solve all the assignment problems;
- If you need any assistance with solving assignment problems, please contact me.
- I can, for example, set-up office hours or tutorial hours.

Cheating, plagiarism, malpractices

- Please do not copy tutorials or assignments; do not plagiarise; the material that you submit to me for evaluation should be your own work.
- If, at any time during the course, you are caught either cheating (copying, plagiarising, misappropriating somebody else's code or presentations, and so on) or indulging in malpractices (proxy attendance, forging medical certificates and so on), it would mean an FR grade.
- In addition, depending on the gravity of the incident, you will also be reported.

Miscellaneous stuff

- Please do not bring your cell phones to the class;
- If you happen to bring, I expect that you keep it switched off (or, at worst, in silent mode);
- In any case, use of cellphones (including as calculators and web-browsers) in the classroom is strictly prohibited.
- During the exams, you should not have a cell phone on you; even if by mistake you have brought your cell phone to the examination hall, leave it near the podium along with your bags, books and notebooks.
- Please make sure that you bring your identity card to the examination hall for all the quizes and examinations.
- No borrowing of calculators, rulers, and other material during the examinations and quizes are allowed.

Course calendar

- Third/fourth week of August: Quiz 1
- Third/fourth week of October: Quiz 2

Objectives

- To understand the data collection methodology
- To analyse data using commercial (excel) and open source (R) software
  R: open source software for statistical computing and graphics
- To draw conclusions from data and to know the errors / assumptions involved

Lecture-wise break-up

- Tentative listing; includes tutorials. For all these problems, we will also do R / excel tutorials using real world data wherever possible
- **Introduction** (3 hours)
- **Descriptive statistics** (3 hours)
- **Elements of probability** (3 hours)
- **Random variables and expectation** (3 hours)
- **Special random variables** (3 hours)

Lecture-wise break-up

- **Distributions of sampling statistics** (3 hours)
- **Parameter estimation** (3 hours)
- **Hypothesis testing** (3 hours)
- **Regression** (3 hours)
- **Analysis of variance** (3 hours)
- **Goodness of fit test** (3 hours)
- **Nonparametric hypothesis tests** (3 hours)

Recommended textbook

- Sheldon M Ross, Introduction to probability and statistics for engineers and scientists

Questions / comments / clarifications?

A bit of history of this course

- A core course for all students of the Institute: used to be taught by mathematics department
- Practical application and subject matter specific issues: better dealt at Department level
- Some departments have stopped; MEMS does not. Why?

Why data analysis and interpretation?

- An important skill
- Live in an age of data
- Dealing with data: collection, analysis, drawing conclusions, etc.: ubiquitous
- Finance, management, engineering, humanities (?), ...

Some terms...

- Data mining
- Data analytics
- Machine learning
- Artificial intelligence
- Big data