

# CS 215: Data Interpretation and Analysis

Fall 2015

Instructors:

Ajit Rajwade

&

Suyash Awate

# Where all do you analyze and interpret data?

|   |  |                                       |  |                                  |  |
|---|--|---------------------------------------|--|----------------------------------|--|
| Specimen #<br><b>253-814-1957-0</b>   |  | Control/Unit Number<br><b>0013332</b> |  | Pg <b>1</b>                      |  |
| Testing<br><b>N/A</b>   |  | Specimen Source<br><b>S / Final</b>   |  | Clinical Information             |  |
| Date Collected<br><b>09/10/07</b>   |  | Time Collected<br><b>09/11/07</b>     |  | Date Reported<br><b>09/11/07</b> |  |
| Patient ID Number<br><b>517.3</b>   |  | Patient Phone Number                  |  | Patient SSN                      |  |
| Patient Name  |  | Sex<br><b>F</b>                       |  | Date of Birth<br><b>03/21/83</b> |  |
| Patient Address   |  |                                       |  |                                  |  |
| Comments<br><b>PATN AGE: 024/05/20</b>  |  |                                       |  |                                  |  |
| Tests Requested <b>CBC With Differential/Platelet; Comp. Metabolic Panel (14); Testosterone, Serum;</b> |  |                                       |  |                                  |  |

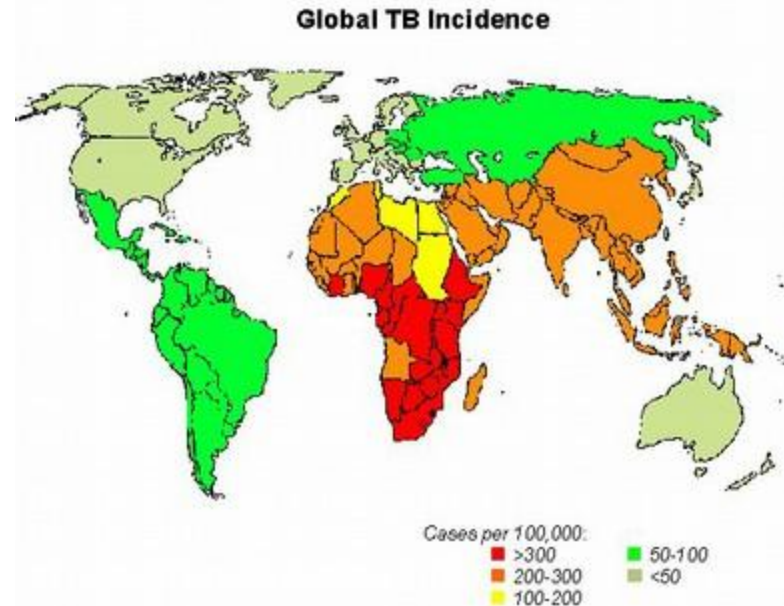
  

| TESTS                                 | RESULT | FLAG | UNITS    | REFERENCE INTERVAL | LAB |
|---------------------------------------|--------|------|----------|--------------------|-----|
| <b>CBC With Differential/Platelet</b> |        |      |          |                    |     |
| WBC                                   | 5.1    |      | x10E3/uL | 4.0 - 10.5         | 01  |
| RBC                                   | 5.34   | High | x10E6/uL | 3.80 - 5.10        | 01  |
| Hemoglobin                            | 15.7   | High | g/dL     | 11.5 - 15.0        | 01  |
| Hematocrit                            | 47.7   | High | %        | 34.0 - 44.0        | 01  |
| MCV                                   | 89     |      | fL       | 80 - 98            | 01  |
| MCH                                   | 29.4   |      | pg       | 27.0 - 34.0        | 01  |
| MCHC                                  | 32.9   |      | g/dL     | 32.0 - 36.0        | 01  |
| RDW                                   | 13.3   |      | %        | 11.7 - 15.0        | 01  |
| Platelets                             | 228    |      | x10E3/uL | 140 - 415          | 01  |
| Neutrophils                           | 65     |      | %        | 40 - 74            | 01  |
| Lymphs                                | 25     |      | %        | 14 - 46            | 01  |
| Monocytes                             | 7      |      | %        | 4 - 13             | 01  |
| Eos                                   | 2      |      | %        | 0 - 7              | 01  |
| Baso                                  | 1      |      | %        | 0 - 3              | 01  |
| Neutrophils (Absolute)                | 3.3    |      | x10E3/uL | 1.8 - 7.8          | 01  |
| Lymphs (Absolute)                     | 1.3    |      | x10E3/uL | 0.7 - 4.5          | 01  |
| Monocytes (Absolute)                  | 0.4    |      | x10E3/uL | 0.1 - 1.0          | 01  |
| Eos (Absolute)                        | 0.1    |      | x10E3/uL | 0.0 - 0.4          | 01  |
| Baso (Absolute)                       | 0.1    |      | x10E3/uL | 0.0 - 0.2          | 01  |
| <b>Comp. Metabolic Panel (14)</b>     |        |      |          |                    |     |
| Glucose, Serum                        | 83     |      | mg/dL    | 65 - 99            | 01  |
| BUN                                   | 26     |      | mg/dL    | 5 - 26             | 01  |
| Creatinine, Serum                     | 1.2    |      | mg/dL    | 0.5 - 1.5          | 01  |
| BUN/Creatinine Ratio                  | 22     |      |          | 8 - 27             | 01  |
| Sodium, Serum                         | 143    |      | mmol/L   | 135 - 148          | 01  |
| Potassium, Serum                      | 4.9    |      | mmol/L   | 3.5 - 5.5          | 01  |
| Chloride, Serum                       | 102    |      | mmol/L   | 96 - 109           | 01  |
| Carbon Dioxide, Total                 | 30     |      | mmol/L   | 20 - 32            | 01  |
| Calcium, Serum                        | 10.0   |      | mg/dL    | 8.5 - 10.6         | 01  |
| Protein, Total, Serum                 | 7.6    |      | g/dL     | 6.0 - 8.5          | 01  |
| Albumin, Serum                        | 4.6    |      | g/dL     | 3.5 - 5.5          | 01  |
| Globulin, Total                       | 3.0    |      | g/dL     | 1.5 - 4.5          | 01  |
| A/G Ratio                             | 1.5    |      |          | 1.1 - 2.5          | 01  |
| Bilirubin, Total                      | 0.3    |      | mg/dL    | 0.1 - 1.2          | 01  |
| Alkaline Phosphatase, S               | 64     |      | IU/L     | 25 - 50            | 01  |
| AST (SGOT)                            | 31     |      | IU/L     | 0 - 40             | 01  |
| ALT (SGPT)                            | 31     |      | IU/L     | 0 - 40             | 01  |
| Testosterone, Serum                   | 444    | High | ng/dL    | 14 - 76            | 01  |

517.3      253-814-1957-0 Seq# 0726 09-11-07 08:14AT

## (1) In Medicine: Examples

- Pathology reports,
- Epidemiology studies



<https://ethnomed.org/clinical/tuberculosis/firlan/d/epidemiology-of-tb>

# Where all do you analyze and interpret data?

| Leading ODI Run Scorers at Number 6 since 1 Aug 2009 |     |      |      |      |      |      |     |    |
|--|-----|------|------|------|------|------|-----|----|
| Player   | Mat | Inns | Runs | HS   | Ave  | SR   | 100 | 50 |
| MS Dhoni (India)                                     | 68  | 58   | 1960 | 139* | 50.3 | 84.6 | 1   | 15 |
| Umar Akmal (Pak)                                     | 59  | 54   | 1706 | 102* | 38.8 | 88.5 | 2   | 11 |
| AD Mathews (SL)                                      | 66  | 59   | 1571 | 89   | 34.2 | 81.3 | 0   | 11 |
| SK Raina (India)                                     | 47  | 41   | 1184 | 106  | 34.8 | 97.5 | 1   | 7  |
| Mushfiqur Rahim (Ban)                                | 40  | 37   | 897  | 86   | 30.9 | 79.3 | 0   | 6  |
| MEK Hussey (Aus)                                     | 29  | 27   | 875  | 79   | 39.8 | 93.8 | 0   | 5  |
| KA Pollard (WI)                                      | 39  | 36   | 843  | 119  | 24.8 | 85.1 | 2   | 2  |
| DA Miller (SA)                                       | 39  | 34   | 797  | 67   | 31.9 | 97.0 | 0   | 5  |
| RS Bopara (Eng)                                      | 31  | 27   | 715  | 101* | 31.1 | 84.5 | 1   | 3  |
| DJ Hussey (Aus)                                      | 23  | 20   | 684  | 74   | 42.8 | 95.3 | 0   | 6  |

<http://i.dawn.com/primary/2015/02/54d32f884dfd0.jpg?r=1999182479>

## (2) In Sports

- Tournament data
- Player data
- Questions like: which is the best team?  
Which is the best batsman? Which is the best batsman from so and so age-group?

# Where all do you analyze and interpret data?

List by the International Monetary Fund (2014)

Rank Country/Region GDP (Millions of US\$)  
World

|    |                |                 |
|----|----------------|-----------------|
| 1  | United States  | 17,418,925      |
| 2  | China          | 10,380,380[n 2] |
| 3  | Japan          | 4,616,335       |
| 4  | Germany        | 3,859,547       |
| 5  | United Kingdom | 2,945,146       |
| 6  | France         | 2,846,889       |
| 7  | Brazil         | 2,353,025       |
| 8  | Italy          | 2,147,952       |
| 9  | India          | 2,049,501       |
| 10 | Russia         | 1,857,461[n 3]  |
| 11 | Canada         | 1,788,717       |
| 12 | Australia      | 1,444,189       |
| 13 | South Korea    | 1,416,949       |
| 14 | Spain          | 1,406,855       |
| 15 | Mexico         | 1,282,725       |
| 16 | Indonesia      | 888,648         |
| 17 | Netherlands    | 866,354         |
| 18 | Turkey         | 806,108         |
| 19 | Saudi Arabia   | 752,459         |
| 20 | Switzerland    | 712,050         |

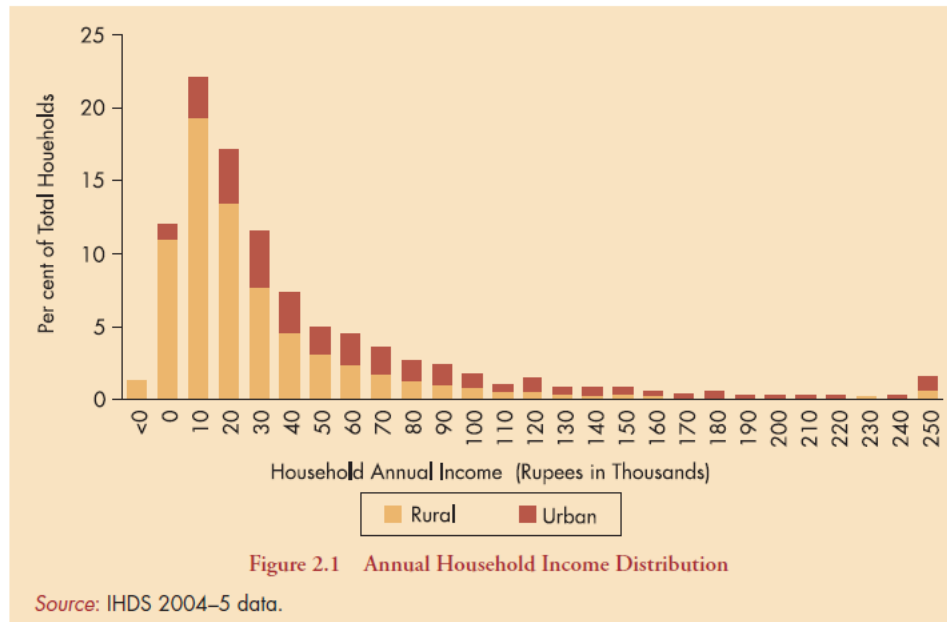
(3) In Economics and Finance:

- Country-wise data

**Gross Domestic Product (GDP)** is the broadest quantitative measure of a nation's total economic activity. More specifically, **GDP** represents the monetary value of all goods and services produced within a nation's geographic borders over a specified period of time.

<http://www.investinganswers.com/financial-dictionary/economics/gross-domestic-product-gdp-1223>

# Where all do you analyze and interpret data?



<sup>1</sup> Some households reported negative incomes. These are usually farm households with partially failed production whose value did not fully cover the reported expenses. Other analyses show that these households do not appear especially poor: their consumption expenditures and household possessions resemble average households more than they do to other low-income households. Because of this anomaly, for income calculations in the remainder of the study, we exclude all households with income below Rs 1,000 (N = 837). The median income after this exclusion is Rs 28,721.

## (3) In Economics and Finance:

[http://ihds.umd.edu/IHDS\\_files/02HDinIndia.pdf](http://ihds.umd.edu/IHDS_files/02HDinIndia.pdf)

- Country-wise data

### GDP per person, Indian states

Purchasing power parity

US\$, 2011

Avg GDP per capita:  
India = \$5138  
World = \$12857  
Non-OECD = \$7219

**GDP per capita, PPP**

|                 |
|-----------------|
| \$8000 - \$9500 |
| \$7000 - \$8000 |
| \$6000 - \$7000 |
| \$5000 - \$6000 |
| \$4000 - \$5000 |
| \$3000 - \$4000 |
| \$2000 - \$3000 |
| \$1000 - \$2000 |

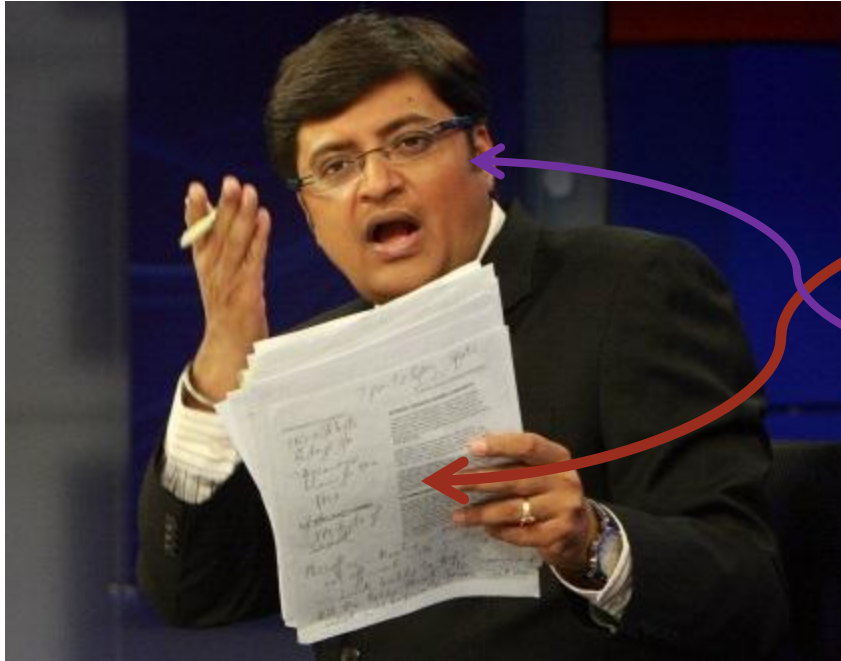
Indian states and territories shown on the map include: Jammu & Kashmir, Himachal Pradesh, Punjab, Chandigarh, Haryana, Uttar Pradesh, Rajasthan, Gujarat, Maharashtra, Goa, Karnataka, Kerala, Tamil Nadu, Andhra Pradesh, Chhattisgarh, Madhya Pradesh, Bihar, Jharkhand, West Bengal, Odisha, Sikkim, Arunachal Pradesh, Assam, Meghalaya, Tripura, Mizoram, Manipur, Nagaland, and the Andaman & Nicobar Islands. Major cities marked include Srinagar, Shimla, Dehradun, New Delhi, Jaipur, Lucknow, Patna, Ranchi, Bhubaneswar, Kolkata, Agartala, Shillong, Guwahati, Itanagar, Kohima, Imphal, Aizawl, Bhopal, Raipur, Hyderabad, Chennai, Pondicherry, Bengaluru, Thiruvananthapuram, Panaji, Silvassa, Daman & Diu, and Dadra & Nagar Haveli. Neighboring countries are Pakistan, China, Nepal, and Bangladesh. Surrounding water bodies are the Arabian Sea, Bay of Bengal, and Indian Ocean. A scale bar at the bottom indicates distances up to 600 km.

- Region-wise data within a country

- Region-wise data within a country



# Where all do you analyze and interpret data?




(4) In Journalism:

Data: that's what those white papers contain! 😊

And he analyzes those data big time!

[Image source](#)



# Where all do you analyze and interpret data?

(5) In many other fields:

- Weather forecasting
- Psephology
- Stock markets
- Industrial testing
- Market research (eg: in industry and storehouses)



# So what's this course all about?

- Sounds like everything under the



[http://www.clipartpanda.com/clipart\\_images/clipart-sun-rays-clipart-1587813](http://www.clipartpanda.com/clipart_images/clipart-sun-rays-clipart-1587813)

# What's this course all about?

- A beginning course on probability and statistics
- A very useful base for future courses in machine learning, data mining, statistics, image processing and computer vision.

# What's this course all about? Three sections

- **Data analysis:** Process of gathering, displaying/visualizing and summarizing the data
- **Probability:** The “chance” that something happens
- **Statistical Inference:** The science of drawing precise inferences from the data gathered using tools from probability

# Example in Toxicology

- Imagine I invent two new medicines (say) to reduce blood pressure.
- I test the two medicines on two groups of rats – A and B – respectively.
- I will then periodically measure BP of rats in groups A and B.
- And seek to determine which medicine is “better”.

# Example in Toxicology: Data Analysis

- What should be the size of A and B?
- How should I pick the members of A and B? Example: can A be all males, B be all females? Can A be all white rats and B be all black rats?
- Once I acquire the BP measurements, how do I display them succinctly? How do I compute averages?

# Example in Toxicology: Data Interpretation (or Statistical Inference)

- Let's say the average BP of A was much lower than that of B after feeding the two drugs.
- Does this mean the first medicine is more effective?
- Or was this just a matter of chance? (Example: If I flip an unbiased coin 50 times, I could land up with 30 heads – just by chance!)

# Course Information

- Instructors: Ajit Rajwade (first half) and Suyash Awate (second half)
- Lecture venue: LH-101, timings: Tue and Fri, 2:00 to 3:25 pm (i.e. post lunch - and strong coffee ☺)
- Course webpage (for the first half):  
[http://www.cse.iitb.ac.in/~ajitvr/CS215\\_Fall2015/](http://www.cse.iitb.ac.in/~ajitvr/CS215_Fall2015/)



# Course Information

- Grading scheme:
  - ❖ 25% midterm (closed notes, most formulae will be provided)
  - ❖ 25% cumulative final exam (closed notes, most formulae will be provided)
  - ❖ Two quizzes: 15% total
  - ❖ Team-based solving of programming and written assignments: 35% (about 5 assignments)
- Attendance mandatory. Students with less than 80% may get a DX.
- We will all adhere to principles of academic honesty. Penalties for violation will be severe.

# Course Information

- We will make extensive use of MATLAB – in and out of class.
- Assignments will be posted on moodle.
- Course textbook: Introduction to Probability and Statistics for Engineers and Scientists: Fourth Edition
- The material will cover lots of examples for each concept! I will cover many examples from medicine, social studies and image processing!

# Course information

- Other rules to follow:
  - ❑ Come to class on time
  - ❑ Submit homeworks on time
  - ❑ Ask questions – in and out of class (after class, during office hours, over email, on moodle's discussion forum, etc.)