Schedule of the course

Introduction to MATLAB

21.09.
28.09.
05.10.
12.10.

Flash Talks

19.10.
26.10.
02.11.
09.11.
16.11.
23.11.
30.12.
07.12.
14.12.

Modeling overview

Introduction to social-science modeling and simulations

Working on projects (seminar thesis)

Handing in seminar thesis and giving a presentation

2015-11-30
Final presentation schedule

- Project presentation 15’ + 5’ (for Q&A)
- All group members have to actively participate in the presentation
- Registration for final presentation is binding; if you do not want to obtain credits, do not register!
- There are 15 slots on two days:
  - Monday, 14 December: 16:20 – 18:20
  - Tuesday, 15 December: 16:20 – 19:00
- Sign up at: http://goo.gl/4psqsM
Final presentation schedule

- Reminder:
- Reports are due Friday **Dec 11 2015** at midnight
- Submission is through GIT:
  simply upload your report, source code, videos etc.
  before the deadline and we will retrieve it
Final presentations location

- Presentations take place in the seminar room of the offices of the Chair of Sociology, in particular Modeling and Simulation.

- The seminar room is located on the first floor, first door on the right. **Room C 1**

- Enter without ringing the bell, but please keep silence in the corridor.

- [http://www.soms.ethz.ch/box_feeder/howtoreach.pdf](http://www.soms.ethz.ch/box_feeder/howtoreach.pdf)
Statistical Tests

- When do we need statistical tests?
  - Determining the strength of your results (error-bars)
  - Quantify the agreement between simulated and empirical data
    - How similar are a simulated distribution and the corresponding empirical distribution?
    - Can we be “certain” that we have reproduced a stylized fact of the empirical data?
  - Optimizing your model for empirical data
    - Identify model parameters for which the agreement with empirical data is maximal
### Statistical Tests: Description and Comparison

<table>
<thead>
<tr>
<th>Goal:</th>
<th>Data:</th>
<th>Binominal (Two Possible Outcomes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe one group</td>
<td>Measurement (from Gaussian Population)</td>
<td>Proportion</td>
</tr>
<tr>
<td>Compare one group to a hypothetical value</td>
<td>Rank, Score, or Measurement (from Non-Gaussian Population)</td>
<td></td>
</tr>
<tr>
<td>Compare two unpaired groups</td>
<td>One-sample t test</td>
<td>Chi-square or Binomial test</td>
</tr>
<tr>
<td>Compare two paired groups</td>
<td>Unpaired t test</td>
<td>Fisher's test (chi-square for large samples)</td>
</tr>
<tr>
<td></td>
<td>Paired t test</td>
<td>McNemar's test</td>
</tr>
</tbody>
</table>

*Source: [http://www.graphpad.com/www/Book/Choose.htm](http://www.graphpad.com/www/Book/Choose.htm)*
# Statistical Tests: Association and Prediction

<table>
<thead>
<tr>
<th>Goal:</th>
<th>Data:</th>
<th>Goal:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantify association between two variables</td>
<td>Measurement (from Gaussian Population)</td>
<td>Rank, Score, or Measurement (from Non-Gaussian Population)</td>
</tr>
<tr>
<td></td>
<td>Pearson correlation</td>
<td>Spearman correlation</td>
</tr>
<tr>
<td></td>
<td>Simple linear regression or Nonlinear regression</td>
<td>Nonparametric regression</td>
</tr>
<tr>
<td></td>
<td>Multiple linear regression or Multiple nonlinear regression</td>
<td>Multiple logistic regression</td>
</tr>
</tbody>
</table>

**Source:** [http://www.graphpad.com/www/Book/Choose.htm](http://www.graphpad.com/www/Book/Choose.htm)
Statistical Tests

- Tests in MATLAB:
  - Many hypothesis tests already efficiently implemented:
    - **Pearson’s correlation**: `corr()` (as options: Spearman and Kendall)
    - **Kolmogorov-Smirnov test**: `kstest()` and `kstest2()` for one-sample and two-sample test respectively
    - **t-test**: `ttest()`

A comprehensive list of all hypothesis tests implemented in MATLAB may be found at [http://www.mathworks.ch/help/toolbox/stats/f31338.html](http://www.mathworks.ch/help/toolbox/stats/f31338.html)
Statistical Tests

- **Example: Power Law**
  - The occurrence of inter-state wars as a function of their size (casualties) follows a power-law

  \[
  \log P(S>s) = 1.27 - 0.41 \log s \\
  R^2 = 0.965 \quad N = 97
  \]

  "war size" in a model of inter-state wars should follow the same distribution!
  
  appropriate test here: **Kolmogorov-Smirnov (KS) test**

Statistical Tests

- Are statistical tests important for your projects?
  - If you are working with empirical data you may want to consider using a statistical test to quantify your agreement with the data…
  - … it is certainly not mandatory, though!
Scientific Writing

- Clear, easy to read, concise, and precise.
  - avoid metaphors; use examples instead
  - keep sentences short
- Main purpose: communicate your work so that others can replicate it, modify it and object to it.
- It is still a form of creative writing!
  - create a narrative
  - guide the reader through your project and results
The Hour Glass Structure

- General
- More specific
- Most specific
- More general
- Most general

Introduction
Materials & Methods
Results
Discussion
Conclusions
What are your research questions?

- In the introduction of your reports, clearly specify the research questions, e.g.:
  - How does crime influence population growth?
  - Will the outcome change if mechanism A is changed to mechanism B in the model? How?
  - Are the results of the model reproducible if the complicated function \( f(x) = \exp(x+y) + y^2 \) is replaced by the simpler \( g(x) = x + y \) ?

- In the “Results” section come back to your research questions.
Interpretation of Results

- Remember: Correlation does not imply causation
- Do not over-interpret your results
- Propose an underlying mechanism

Credit: xkcd.com/552
Correlation vs. Causality

- Correlation:

  \[
  \text{corrcoef}(A,B) = \begin{pmatrix} 1.0000 & 0.8390 \\ 0.8390 & 1.0000 \end{pmatrix}
  \]

- Causality: A correlates with B, but what is the causality?

  ![Diagram of A correlated with B](image)

  ![Diagram showing possible causal relationships](image)
Reproducibility of Results

- It is an important criterion for the quality of scientific work that your results are fully reproducible.
- Make sure to **set and store** the random seed of every simulation run to ensure reproducibility.
- `RandStream.getGlobalStream` returns default random number stream with its seed value.
- `RandStream.setGlobalStream(RandStream('mt19937ar','seed',Seed))` sets the random seed with the (new) value `Seed`.
Data Dump – storing your results

- Make sure that the results of your simulations are stored at the end of a simulation run.
- You can for example simply save the results in form of the array structure (matrix, cell array) of your simulation using MATLAB’s `save()` command.
- Or you may export them to comma separated files using `csvwrite()` (and `csvread()` to recover data).
Visualization of Results

- Graphical formats: use PDF or EPS in the print version
- Increase the font sizes for readability:
  \[
  \text{set(gca, 'FontSize', 16);}
  \]
- ALWAYS use labels on axes:
  \[
  \text{xlabel('Time (s)', 'FontSize', 16);}
  \]
  \[
  \text{ylabel('Number of Agents', 'FontSize', 16);}
  \]
Visualization of Results (2)

- Put a box around the figure: `box on`
- If you want to show grid: `grid on`
- Use thicker lines:
  ```matlab
  plot(x, y, 'LineWidth', 2);
  ```
- Set axis limits:
  ```matlab
  xlim(x0, x0); ylim(y0, y0);
  ```
Example

- As an example: Let us say that we have a project about simulating happiness of agents.
- We have the result of 10 different simulation runs in \( y \), and the corresponding time vector in \( x \).

\[
x = t0:dt:t1;
y = [\text{happiness values from simulation 1};
    \text{happiness values from simulation 2};
    \ldots
    \text{happiness values from simulation 10}];
\]
plot(x, y(1,:))

Figure 4: Simulation results.
plot(x,y(1,:))
xlabel('Time(days)')
ylabel('Fraction of happy agents')

Figure 4: Simulation results.
plot(x,y)
xlabel('Time(days)')
ylabel('Fraction of happy agents')

Figure 4: Simulation results.
plot(x,y)
set(gca, 'FontSize', 16)
xlabel('Time(days)')
ylabel('Fraction of happy agents')

Figure 4: Simulation results.
```matlab
errorbar(x, mean(y), std(y))
set(gca, 'FontSize', 16)
xlabel('Time (days)')
ylabel('Fraction of happy agents')
```

Figure 4: Simulation results.
errorbar(x, mean(y), std(y), 'k', 'LineWidth', 2)
set(gca, 'FontSize', 16)
xlabel('Time (days)')
ylabel('Fraction of happy agents')
xlim([0 10])
ylim([0 .5])

Figure 4: Simulation results.
errorbar(x,mean(y),std(y),'k','LineWidth',2)
set(gca,'FontSize',16)
xlabel('Time(days)')
ylabel('Fraction of happy agents')
xlim([0 10])
ylim([0 .5])

add caption in LaTeX!

Figure 4: Fraction of happy agents as a function time, for 10 different simulation runs. The average value (solid line) and one standard deviation (error bars) are shown. The parameter values are $a=1.0$ and $b=0.7$. 
Figure 4: Fraction of happy agents as a function time, for 10 different simulation runs. The average value (solid line) and one standard deviation (error bars) are shown. The parameter values are $a=1.0$ and $b=0.7$. 
Phase Diagrams

- To show how the result depend on two different parameters, phase diagrams can be used:

![Phase Diagram](image)

D. Helbing and W. Yu, PNAS (2009)
Phase Diagrams

- code snippet that uses `surf()`: here matrix $M$ has as entry $(i,j)$ the model fit for every combination of two parameters (here: $a$, $b$)

```matlab
[a,b]=meshgrid(0:0.01:1);
hold on;
view(0,90);
surf(a,b,M,'EdgeColor','none');
colorbar;
set(gca,'FontSize',14)
xlabel('parameter a');
ylabel('parameter b');
```
Phase Diagrams
Normalized Parameters

- It is usually advantageous to rescale any parameter in your model to a uniform range, for example [0,1]

- Normalized parameters simplify the visualization and analysis of the model results

- Ideally, the parameter range of model parameters and their scaling to the uniform range is chosen such that all relevant variations of the model are visible within that range
Create Videos

- It is helpful to illustrate your simulations with videos during your presentation:

```matlab
% Get the handle of the figure
h = figure();

% Prepare the new file.
vidObj = VideoWriter('video.avi');
open(vidObj);

for i=1:100
    % plot something
    plot(x,y,...);
    % get the current frame
    currFrame = getframe(h);
    % add the current frame
    writeVideo(vidObj,currFrame);
end

% Close the video file.
close(vidObj);
```
Project Report

- Use the report template.
- Try to explain your results and your motivation behind the project.
- Upload your MATLAB code to GIT
- If you have any problems or concerns with your projects, we are here to help you!
Project Report

- There are a number of things that will certainly negatively affect your grade:
  - no clearly stated research question
  - no conclusion regarding the work you have completed and how it relates to your research question
  - graphs which have no legend, axis labels or caption, are not properly readable or the labels are too small
  - no references to the literature to validate and back-up your statements/model etc.
  - submitting you report late or incomplete
Project Report

- **References** are very important:
  - cite the sources you are using
  - be aware that *plagiarism* is taken very seriously at ETH, if you are unsure how to cite properly follow ETH’s plagiarism guidelines
  - we accept both natural science and social science citation conventions but please use them consistently
  - it is important that your references are correct and correctly formatted!
Project Report

- **Documentation** of your code is important

- You can, for example, use the following tool to automatically generate html files from comments in your code

  http://www.artefact.tk/software/matlab/m2html/
Final checklist

- A typical scientific report should include:
  - Title
  - Abstract
  - Introduction
  - Materials / Methods
  - Results
  - Discussion
  - References
  - Appendix (if needed)
Principles of a good presentation

- The audience wants:
  - To learn something new from you.
  - To trust your authoritativeness.

- The audience has limited attention.
  - You have it at the beginning. Use it wisely.
  - Humor or images recall attention (but don’t distract!)

- Your talk should be self-contained.
Principles of a good presentation

0) Have a **message** (not “what?” but “so what?”)

1) Adapt to your audience

2) Maximize signal/noise ratio

   Audience has limited attention

3) Use **effective redundancy** (different communication channels are complimentary)

   Source: J.L. Dumont
Principles of a good presentation

0) Have a **message** (not “what?” but “so what?”)

1) Adapt to your audience

2) Maximize signal/noise ratio

   Audience has limited attention

3) Use **effective redundancy** (different communication channels are complimentary)

Map out the story first. Point by point, slide by slide.

Source: J.L. Dumont
General structure of the talk

1. Start with a general motivation or story.
   - Connect to the specific context of this talk.

2. Introduction
   - Connect to past work in your field.
   - Explain what new insight you are adding.

3. Results
   - Approx. three main results is sufficient.
   - Time you spent presenting a result should not be proportional to time spent working on it!
General structure of the talk

4. Conclusion
   - Recap main points made.
   - Emphasize the “take-home” message.
   - Echo of introduction as a conclusion can create closure.

Tell the audience what you are going to tell them, tell them, and then tell them what you told them.
Crafting simple slides

- Use a slide to make one point and one point only.
- The title should capture the message of the slide.
  - Full sentences in the title are easy to grasp.
- Quickly convey information and message.

Source: NY Times and J.L. Dumont

General Mc Chrystal: “When we understand that slide, we’ll have won the war.”
Crafting simple slides

- Limit text – at most two lines per point.
  - The slides are not there to remind you of your points.
  - If the audience has to read or focus too hard on your slides they don’t listen to you!

- Use animations instead of pointing.
  - But remember that animations can distract from the message.
Crafting simple slides

- Limit text – at most two lines per point.
  - The slides are not there to remind you of your points
  - If the audience has to read or focus too hard on your slides they don’t listen to you!

- Use animations instead of pointing.

- Now that you have beautiful slides, do not just give the audience a “tour” of your slides!
Good delivery is critical

- Look at the audience. Take a confident pose.
- Practice, especially the introduction.
- Know your slides, anticipate.
- Use tone and volume:
  - Monotony makes the audience lose focus.
  - Lowering your volume can make the audience focus.
Good delivery is critical

- Look at the audience. Take a confident pose.
- Practice, especially the introduction.
- Know your slides, anticipate.
- Use tone and volume:
  - Monotony makes the audience lose focus.
  - Lowering your volume can make the audience focus.
- Plan to use about two-thirds of the allotted time.
  - Slide count “x of Total” can help structure time.
Answering questions

- Focus on the content, not the tone!
- Repeat the question to yourself and take time to think.
- Repeat the question to audience and speaker.
  - More time to respond.
  - Clarify that you understood.
- Never take it personally.
References

- J.-L. Dumont.


- How to give a good talk: [http://www.weizmann.ac.il/mcb/UriAlon/nurturing/](http://www.weizmann.ac.il/mcb/UriAlon/nurturing/)