LTAT.06.010 Pervasive Data Science
Lecture 4

Huber Flores, PhD
ASSOCIATE PROFESSOR

Tartu, Estonia 28/02/2022
Recap

• Each student presented their envisioned testbed
What to do with the data?

- Pre-processing
  - Cleaning the data, adjusting scales, representativeness and validity
- Analysis *(this lecture)*
- Model construction
  - Classical machine learning
# Analysis

<table>
<thead>
<tr>
<th>Goal</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Describe one group</strong></td>
<td><strong>Measurement (from Gaussian Population)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rank, Score, or Measurement (from Non-Gaussian Population)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Binomial (Two Possible Outcomes)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Survival Time</strong></td>
</tr>
<tr>
<td><strong>Describe one group</strong></td>
<td>Mean, SD</td>
</tr>
<tr>
<td><strong>Compare one group to a hypothetical value</strong></td>
<td>One-sample t test</td>
</tr>
<tr>
<td><strong>Compare two unpaired groups</strong></td>
<td>Unpaired t test</td>
</tr>
<tr>
<td><strong>Compare two paired groups</strong></td>
<td>Paired t test</td>
</tr>
<tr>
<td><strong>Compare three or more unmatched groups</strong></td>
<td>One-way ANOVA</td>
</tr>
<tr>
<td><strong>Compare three or more matched groups</strong></td>
<td>Repeated-measures ANOVA</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Quantify association between two variables</strong></td>
<td>Pearson correlation</td>
</tr>
<tr>
<td><strong>Predict value from another measured variable</strong></td>
<td>Simple linear regression or Nonlinear regression</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Predict value from several measured or binomial variables</strong></td>
<td>Multiple linear regression ** or Multiple nonlinear regression **</td>
</tr>
</tbody>
</table>

[source] https://www.graphpad.com/support/faqid/1790/
PENGUIN: Aquatic Plastic Pollution Sensing using Autonomous Underwater Vehicles (AUV)s

https://dl.acm.org/doi/10.1145/3396864.3399704
PENGUIN: Testbed

RIC

01 PET
02 PE-HD
03 PVC
04 PE-LD
05 PP
06 PS

Plastic bottles
Plastic bags
Shampoo bottles
Plastic cards
Six pack rings
Plastic trays
Respiratory masks
Diapers
Food packing
Car panel
PENGUIN: Testbed

Controlled plastic samples

Light sensor through glass

Testbed for plastic recognition underwater

Black cover

Plastic sample

Light

PPG sensor

Glass vessel
PENGUIN: Evaluation

**Result:** Green light reflectivity can characterize plastics in different environmental conditions

(a) Ambient light

(b) Dark environment
PENGUIN: Evaluation

Statistical analysis in R

```r
# Effect
# Kendall's W uses the Cohen's interpretation guidelines of 0.1 - < 0.3 (small effect), 0.3 - < 0.5 (moderate effect) and >= 0.5 (large effect).
right %>% friedman_effsize(fingerprint ~ material | id)
# W=0.54, magnitude = large effect

# pairwise comparisons
# A significant Friedman test can be followed up by pairwise Wilcoxon signed-rank tests for identifying which groups are different.
pwc <- right %>%
  wilcox_test(fingerprint ~ material, paired = TRUE, p.adjust.method = "bonferroni")
pwc
```
PENGUIN: Evaluation

Result: High accuracy to classify different types of plastics underwater

<table>
<thead>
<tr>
<th>Test</th>
<th>k-NN</th>
<th>Random forest</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross Validation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All conditions 6-folds</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Ambient 6-folds</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Darkness 6-folds</td>
<td>0.94</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>Model data → Predicted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient → All conditions</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Ambient → Darkness</td>
<td>0.69</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>Darkness → All conditions</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Darkness → Ambient</td>
<td>0.94</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>Average</td>
<td>89.0</td>
<td>88.7</td>
<td>88.9</td>
</tr>
</tbody>
</table>

Plastic classification accuracy in different experimental conditions.  
Model data -> Predicted
Next lecture

Phd students presenting projects (Part I)
Questions?

E-mail: huber.flores@ut.ee