Introduction/Motivation

The experiment (conducted in the T.G. Masaryk Water Research Institute in Prague in 2010) investigates a behavior of a group of mutually familiar catfish individuals when they encounter an unfamiliar fish (of the same species but of a different kind).

A group of regular pigmented catfish individuals (which were taken from the same shoal, therefore, familiar ones) was placed into an artificial indoor stream and an additional catfish (a fish unfamiliar with the group, i.e., a fish from another shoal from a different fish supplier) was placed into the same stream together with the group.

The idea behind the experiment is the following: the group of familiar catfish individuals should behave differently, depending on the type of the additional catfish added into the steam (either a regular pigmented individual same like the other members of the original group was added into the stream or an albino type fish – visually different from the group – was placed among the group).

Scientific Hypothesis

The group of familiar individuals should be more likely to accept and to get on with the regular pigmented individual. After some few hours the additional fish could be even treated as a group member. However, the same species fish, but an albino individual, is not expected to be accepted by the group of familiar individuals and, it is likely, that there will always be some distance kept between the original group and the additional albino fish. In some hyperbolic sense we could express this type of behavior as some form of exhibiting racist traits within the group of the regularly pigmented catfish. The following scientific hypothesis can be formulated:

1. The 'distance' between the original group and the additional catfish individual depends on the type of the additional fish (a pigmented or an albino catfish) and it is expected that the regular catfish individual will be more likely to join the group;

2. The albino individual should be more 'isolated' in the stream than an additional catfish with a regular pigmentation.

3. The group of the original catfish individuals should behave more 'cohesive' if there is an albino individual present in the stream;

Experiment Design

For the purpose of this experiment there were three different shoals from three distant fish suppliers in the Czech Republic considered (two shoals with pigmented catfish individuals and one shoal with albino species). The individuals within each shoal are considered to be familiar and individuals from different shoals had never seen each other, therefore, they were treated as being unfamiliar. The fishes from three different shoal were hold in three separate holding tanks prior the experiment and therefore the familiarity/unfamiliarity attribute is assumed to be preserved during the experiment.

The laboratory experiment was conducted in an artificial stream (beige color canal, 3.75 m long, 0.49 m wide, and 0.32 m deep). The stream was divided into 5 sub-units using 6 equidistant PIT antennae which were used to record the fish movements in the stream: each individual in the stream was injected with a transmitter and once some individual passed through some antenna to get to another sub-unit (or was very close to the antenna) the information on the fish specific transmitter number, the actual time of passing, and the ID of antenna was recorded and stored on an external memory.

Two different treatments were considered (adding a pigmented fish into the group of 8 familiar (pigmented) individuals or adding an albino fish, again into the group of 8 familiar (pigmented) individuals) and ten independent repetitions were considered for each treatment (controlling for various external effects: water pH, temperature, water brightness, external light conditions, food availability, individuals' weight and length, and many others). The treatments were rotated regularly (one albino experiment, one regular experiment) and every experimental trial lasted for about 24 h (beginning roughly at 8:00 a.m.). The follow up time of 24 hours was considered (by the biologist experts) to be enough for the catfish species to get familiar with the environment, existing predators, food sources, and other organisms being present in the same environment. Therefore, 24 hours of the follow up period was also considered to be enough for the group to get on with the additional fish if they would want to do so.

Data Description
The experimental data consists of two main parts: the data files from the experimental trials generated by the 6 antennae in the stream (directory *data*) and the manually created check files (directory *check files*), which can be used for a retrospective control (in order to get rid of some false records in the data files).

1. **data files** – for each experimental trial there is one specific data file (`data.txt`). In total, there are 20 data files (for 20 trial repetitions) located in the directory *data*, in sub-directories *albin* or *regular* (depending on the treatment), and additional sub-directories given by the date when the specific experiment took place (10 sub-directories for each treatment).

    Each data file (`data.txt`) consists of 10 columns (see the description below) and some number of observations (given in rows). However, the observations are split into two parts according to the last column. The reason for that is that one half of the artificial stream was covered by one set of antennae (so called *k-system*) and the second half of the stream was covered with the second set of antenna (*m-system*). These two systems are working independently: they both need to be turn on manually and they do not communicate with each other. Moreover, each system stops working once its memory is full (the other system can still run further while recording the fish activity in one half of the stream). Both systems are, however, perfectly synchronized using the radio transmitted time information. For more details about the antennae systems and the technical background behind see the Technical Issues session below.

    Each `data.txt` file consist of the following columns (covariates):

    V1 – date (format "mm:dd:yy") when the individual record took place;

    V2 – time the individual record took place (format "hh:mm:ss");

    V3 – technical information – only contains the word "Unit" (not relevant);

    V4 – technical information for recording errors – it is always left blank (not relevant);

    V5 – the antenna specific number provided in the hexadecimal code;

    V6 – fish specific identifier (fish transmitter ID);

V7 – technical information – only contains the word "ISO" (not relevant);

V8 – technical information – only contains the abbreviation "FDX-B' (not relevant);

V9 – technical information – only contains the word "Animal" (not relevant);

V10 – binary factor on whether the record comes from the *k-system* or the *m-system*;

2. **check files** – for each experimental trial there is also one check file (`checkTable_x.y.txt`, where `x.y` stands for the day when the specific trial ended – it is the date, when data were downloaded from the recording devise) located in the directory *check files*. The idea of the check files is to have some means for a backward control mechanism. The technical limitations of the experiment allowed for some accidental (false) records (for instance, a transmitter forgotten in the stream being constantly recorded, a fish with two transmitters, etc.) thus, it is very likely and quite common to have some fish specific ID records in the data, which are not meant to be the part of the experiment. In the check files there is always an information provided on which fish IDs should be included in the experiment and which ones are false. In addition, from the ID records in the raw data files one can not distinguish which IDs belong to the group members and which one is used to identify the additional individual. This information is also provided in the check files.

Each check file consists of 6 columns and the same number of rows as the number of unique ID transmitters recorded in the experiment (optimally, there should be always 8 unique ID values for the group members and one additional ID for the extra fish).

For a detailed description of the information included in the check files see the list below.

- *originalID* – the fish specific ID, respectively the fish specific transmitter ID which was recorded during the experiment trial (is present in the corresponding raw data file);
- *newID* – fish specific ordering according to the transmitter ID;
- *absCounts* – the total number of records per each transmitter ID within the available follow-up period;
- *inExperiment* – an identifier on whether the recorded ID was supposed to be included in the experiment (value `yes`), or not - value `no` (e.g., an accidentally placed transmitter which was not meant to be there in the stream, but somehow got recorded in the data);
- *experimentType* – information about the treatment used in the given trial (`regular` if a pigmented fish was added into the stream or `albin` if an albino individual was placed into the stream instead);
- *albinFish* – identifier of the additional fish (value `yes` for the ID number which belongs to the additional fish – either a regular pigmented catfish or an albino individual);

Technical Issues

Below, there are briefly some specific technical issues described. These issues should be all properly taken into account in the data preparation step and the consecutive statistical modeling step.

- **Assembling the data** – firstly, each experimental trial has to be synchronized with respect to two recording systems used (*k-system* and *m-system*). The only relevant information about the fish activity in the stream is obtained if both systems are running and both are recording. This can be done by using the information about the time when individual records took place – the time is perfectly synchronized on both devices using some radio transmitted time information.

If the experiment trial started on Friday, the recording devices could technically run until Monday (or until the memory on some device got full). The follow-up period considered for the experiment is only 24 hours. On the other hand, it is also possible, that some system's memory got full even before the 24 hour time run out. In such cases, there is a shorter period of time available for the experiment run.

- **Fish position** – given the technical limitations of the experiment it is not possible to tell the position of some specific individual in the stream. The only information obtainable from the data is some backtracking sense – by tracking the fish movements using 6 antennae placed in the stream.

  There are 12 antennae all together (6 antennae for each system), however, there were only three from each system placed into the stream but the remaining ones had to stay active and they were placed aside. Therefore, it is also likely, that some false records were also recorded by these antennae. The artificial stream was defined by the edge antenna on the left hand side (antenna no. 15) and another edge antenna on the right hand side (antenna no. 6, which did not work reliably). In addition, there were four more antennae equidistantly placed in the stream, from left to right – antennae 17, 18, 1, and finally 2. Therefore, only records from these 6 antennae are reliable for the experiment. All other antennae records are clearly false records.

- **Multiple records** – again, due to the technical limitations of the experiment, there are many multiple records available in the data. The reason is that each antenna keeps recording the fish specific ID transmitter anytime the fish gets too close to the antenna. Thus, it is possible that some antenna makes multiple records but the fish is not moving anywhere – it just got stuck somewhere near by. Such multiple records bring no additional information about the fish position and they should not be considered. This is also the reason of such many observations in the data files in total – many of them are just multiple replications of the same fish position (or roughly the same location).

- **Missing records** – when tracking down the movements of some specific individual in the stream, it can happen that the fish is recorded by some antenna and the next record is made by some other antenna which is however not the neighboring antenna with the previous one. Thus, technically, there should be at least one more record made by some antenna in between but it is missing. This can happen because of some small time delay (roughly milli/micro seconds) which antenna takes before accepting two records. Thus, if there are two fish individuals swimming together and passing through antenna at (roughly) same time, only one transmitter will be detected by the antenna.

- **Distance from the group** – in the experiment we want to investigate how the group of familiar catfish individuals reacts when it encounters an additional fish. This behavior can be measured in terms of the distance the group keeps from the additional individual, or the coherence of the original group, or the isolation of the additional fish (or all three together). Therefore, one needs to find some reasonable representation of the "*distance*", "*coherence*", and '*isolation*" as there is no specific positions of any fish given in the data. This can be done in multiple ways using, for instance, the information provided in the data and combining it also with the information which is provided about the stream (its dimensions).

Statistical Analysis

The data created after the preparation step could form unbalanced longitudinal profiles where for each experiment trial we are interested in the time-development of the distance between the group and the additional individual (coherence of the group, or isolation of the additional individual). There are no fish specific characteristics included in the data sets: the experiment was designed in a way that all individuals included in the trial should have (roughly) the same weight and also (roughly) the same length (same size fish). Therefore, in the statistical analysis, we only control for the type of the additional fish (regularly pigmented vs. albino).

Some variance-covariance issues might be expected as well: the behavior of individuals within one experiment can not be assumed to be independent (catfish species are known for expressing some social interactions among them and preserving a form of hierarchy). On the other hand, the experimental trials can be considered to be independent among each other. Again, there are no additional covariates specifying the specific experimental trial – there is only an information whether the trial treatment was about adding a pigmented fish, or an albino one. The experiment took place in the artificial indoor stream and most other external conditions were carefully controlled and they were same for each experiment trial.

Prague, October 5th, 2023