

Salmon Report

MSSP Consulting Group: Qian Xu, Zichun Liu, Zijian Han, TF mentor: Jun Li

Feb 1, 2016

1. Project Background

Our client David Minkoff is from BU Marine Program. In his project, Juvenile salmon were placed in a 2-choice maze to determine preference between 2 odors administered into the water column. Laminar flow was established through the maze, ensuring that no mixing of odors occurred from one side to the other. The location of the fish on one or the other side of the maze was observed every 5 seconds for a duration of 2 minutes, after which the sides on which odor was administered was switched. The location of the fish was again observed for an additional 2 minutes (number of data points per fish = 24 time points before odor side-switch + 24 time points after side switch = 48 total data points per fish). Three different odor choice scenarios were tested with n=36 fish for each scenario.

1.1 Research Question

Do fish prefer the odor to expose when they grow-up over the others?

Expectation: Fish prefer the odor to expose when they grow-up over the others.

1.2 Our client's plan and problem

Because of the relatively small sample size and non-normal distribution, our client's plan was to use a Wilcoxon signed rank test, however there are many cases in which the fish entered one side of the flume and didn't move, even after the side switch of odors (resulting in a tie: 24 observations in "odor #1" and 24 observations in "odor #2") as well as many cases in which there were 48 observations in "odor #1" and zero observations in "odor #2." The Wilcoxon signed rank test apparently doesn't work well with ties and with zeros, so our client need to find a good alternative.

1.3 Data Description

- (a) There are 3 kinds of odors: A, B, C
 - A: well/brook water with Arginine, for fish from well/brook water accordingly
 - B: odor given off by other fish
 - C: control, which is still water
- (b) raise fish in well water (2000), half treatment, half control
raise fish in brook water (2000), half treatment, half control
test different fish in maze at one time. Thus the fish are i.i.d.
- (c) Do 12 experiments, as follows. The rows show the average proportion of (turn to A)/(turn to A + turn to B), (turn to A)/(turn to A + turn to C), (turn to B)/(turn to B + turn to C),

| Tables | Well water treatment | Well water control | Brook water treatment | Brook water control |
|--------|----------------------|--------------------|-----------------------|---------------------|
| A vs B | 0.44, N=35 | 0.42, N=36 | 0.56, N=36 | 0.48, N=36 |
| A vs C | 0.44, N=36 | 0.44, N=36 | 0.52, N=36 | 0.50, N=33 |
| B vs C | 0.56, N=36 | 0.43, N=36 | missing, N=36 | 0.62, N=36 |

When we focus on the first two rows of this table: A vs B and A vs C, the treatment proportions are a little higher, which is a weak evidence supporting the expectation.

1.4 Summarization of the key factors

According to the research question and the data format, we choose to use if the fish turned to Arginine as response, which is binary when we decomposing the original data into independent measures. That is, take each measurement as an observation.

Predictors are:

- (a) Hatch in Brook water (True or False)
- (b) Hatch in Treatment (True or False)
- (c) Arginine on Left (True or False)
- (d) the other odor (B or C)

Note that if don't use the whole data, some predictors can't be added in. We will discuss it more in the following sections.

Random effect: fish ID

2. Exploratory Data Analysis Step

2.1 scatter plot for A percentage

We are interested in how fish (treatment or control) perform in A vs B or C condition, so we make the following plot:

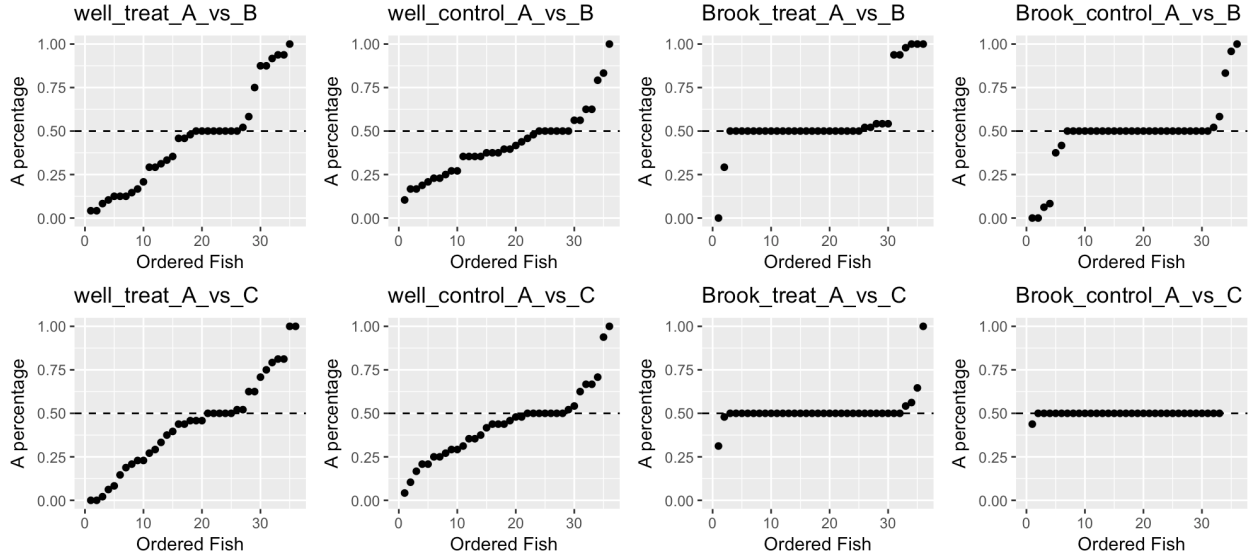


Figure 1: scatter plot for A percentage

Controlling other conditions the same and varying the fish grow-up condition (treatment or control), for example, comparing well treat A vs B and well control A vs B. We will always find the result always tends to support the expectation in Section 1.1, although the evidence still seems weak.

3. Modeling

We do the following modeling:

- (a) Model 1: combine the data corresponding to the first row of figure 1, fit a simplified model based on Section 1.4
- (b) Model 2: combine the data corresponding to the second row of figure 1, fit a simplified model based on Section 1.4
- (c) Model 3: combine the data corresponding to the whole matrix of figure 1, try to fit a full model based on Section 1.4

It is reasonable to not consider the interaction effect between “Hatch with Treatment”, “Arginine on the Left” and “the other odor type”.

Unfortunately, we only have “Brook Control A vs B”, “Brook Treatment A vs B”, “Brook Control A vs C”, “Brook Treatment A vs C”, so:

- (a) Model 1: combine “Brook Control A vs B”, “Brook Treatment A vs B”, predictors: “Hatch with Treatment”, “Arginine on Left”, random effect: “Fish ID”
- (b) Model 2: combine “Brook Control A vs C”, “Brook Treatment A vs C”, predictors: “Hatch with Treatment”, “Arginine on Left”, random effect: “Fish ID”
- (c) Model 3: combine all the 4 data sets, predictors: “Hatch with Treatment”, “Arginine on Left”, “the other odor”, random effect: “Fish ID”.

For following output, I only remain the fixed effect part.

3.1 Model 1

```
## fit1 <- glmer(data = dat1, formula = Turn_to_Arginine ~ as.factor(Hatch_With_Treatment) +
##               + as.factor(A_on_Left) + (1|FishID),family = binomial(link="logit"))
##
## Fixed effects:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.04321    0.23149  -0.187   0.8519
## as.factor(Hatch_With_Treatment)TRUE  0.55727    0.32426   1.719   0.0857 .
## as.factor(A_on_Left)TRUE           -0.16405    0.07444  -2.204   0.0275 *
```

The p-value and coefficient of “Hatch with Treatment” support the expectation, but only nearly significant.

3.2 Model 2

```
## fit2 <- glmer(data = dat2, formula = Turn_to_Arginine ~ as.factor(Hatch_With_Treatment) +
##               as.factor(A_on_Left) + (1|FishID),family = binomial(link="logit"))
##
## Fixed effects:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.26831    0.06167   4.351 1.36e-05 ***
## as.factor(Hatch_With_Treatment)TRUE  0.06908    0.07025   0.983   0.325
## as.factor(A_on_Left)TRUE           -0.55206    0.07019  -7.865 3.68e-15 ***
```

The p-value and coefficient of “Hatch with Treatment” support the expectation, but not significant.

3.3 Model 3

```
## fit3 <- glmer(data = dat_all, formula = Turn_to_Arginine ~ as.factor(Hatch_With_Treatment) +
##               as.factor(A_on_Left) + as.factor(the_other_odor) + (1|FishID),family = binomial(link="logit"))
##
## Fixed effects:
##
```

| | Estimate | Std. Error | z value | Pr(> z) |
|--|----------|------------|---------|--------------|
| ## (Intercept) | 0.17450 | 0.12705 | 1.373 | 0.1696 |
| ## as.factor(Hatch_With_Treatment)TRUE | 0.28180 | 0.14449 | 1.950 | 0.0511 . |
| ## as.factor(A_on_Left)TRUE | -0.37152 | 0.05110 | -7.271 | 3.58e-13 *** |
| ## as.factor(the_other_odor)C | -0.09534 | 0.14420 | -0.661 | 0.5085 |

The p-value and coefficient of “Hatch with Treatment” support the expectation, but only nearly significant. It seems that “the other odor” could be dropped, we try this model:

```
## fit31 <- glmer(data = dat_all, formula = Turn_to_Arginine ~ as.factor(Hatch_With_Treatment) +
##               as.factor(A_on_Left) + (1|FishID),family = binomial(link="logit"))
##
## Fixed effects:
##
```

| | Estimate | Std. Error | z value | Pr(> z) |
|--|----------|------------|---------|--------------|
| ## (Intercept) | 0.12808 | 0.10611 | 1.207 | 0.2274 |
| ## as.factor(Hatch_With_Treatment)TRUE | 0.27989 | 0.14461 | 1.935 | 0.0529 . |
| ## as.factor(A_on_Left)TRUE | -0.37152 | 0.05109 | -7.271 | 3.56e-13 *** |

The p-value and coefficient of “Hatch with Treatment” support the expectation, but only nearly significant. We could see the results from Section 1.3, Data Description, EDA and Modeling Part agree with each other: support the expectation, but only nearly significant or not significant.

4. Conclusion

Through EDA, we’ve identified there maybe effect of side where the Arginine was placed: Fixing the water type (well or brook) and the comparison type (A vs B or A vs C), the percentage that the fish goes to the Arginine side is always a little higher in the treatment group, where the fish are raised in Arginine environment.

After controlling for the side of the tank in the logistic regression model, even though we do see slightly effect: the coefficients of “Hatch with Treatment” are all positive in Model 1,2,3, the odor preference based on the method of how the fishes were raised become statistically insignificant.

Based on this analysis, the conclusion we’ve reached is that the fishes tend to prefer the right side of the tank because in all models, the coefficients of “A on the Left” is negative and significant, but there is no strong evidence that fish prefer the odor to expose when they grow-up over the others.