SCR Data Summary

Loading objects:  
 data\_xlsx

Loading objects:  
 data\_no\_repeat\_all

Loading objects:  
 data\_combine\_ext

Loading objects:  
 data\_no\_missing

## Data Cleaning

### Raw data data\_xlsx

The raw data data\_xlsx loaded from ./data/raw/2024-07-16\_FSARS\_EDA\_Long\_V2\_101-269.xlsx contains 124 subjects.

length(unique(data\_xlsx[["id"]]))

[1] 124

### Cleaned data data\_no\_repeat\_all

I first clean the dataset. I found that some trial numbers are repeated. One has reaction and variable values, and the other has missing values. I keep the trial that has values. The data set that removes all repeated trials is saved in data\_no\_repeat\_all.

data\_no\_repeat\_all

# A tibble: 19,840 × 11  
 id run condition trial cs\_stim\_time cs\_scl cs\_latency cs\_amplitude  
 <chr> <chr> <fct> <int> <dbl> <dbl> <dbl> <dbl>  
 1 101 Learning fear 1 95.9 9.91 2.20 0.735  
 2 101 Learning fear 2 259. 7.68 2.26 2.31   
 3 101 Learning fear 3 NA NA NA NA   
 4 101 Learning fear 4 NA NA NA NA   
 5 101 Learning fear 5 NA NA NA NA   
 6 101 Learning fear 6 NA NA NA NA   
 7 101 Learning fear 7 NA NA NA NA   
 8 101 Learning fear 8 NA NA NA NA   
 9 101 Learning fear\_safety 1 NA NA NA NA   
10 101 Learning fear\_safety 2 NA NA NA NA   
# ℹ 19,830 more rows  
# ℹ 3 more variables: cs\_rise\_time <dbl>, cs\_size <dbl>, cs\_onset <dbl>

Note that each subject has 4 runs, 5 conditions, and 8 trials, so we have rows.

Also, we usually use snake\_case to name variables, and set character values. So the variable condition has values “safety”, “reward\_safety”, “fear\_safety”, “reward”, “fear”. The variable run will later change its values to “learn”, “run1”, “run2”, and “run3”.

The variable condition is of type factor having levels

class(data\_no\_repeat\_all$condition)

[1] "factor"

levels(data\_no\_repeat\_all$condition)

[1] "safety" "reward\_safety" "fear\_safety" "reward"   
[5] "fear"

Levels: safety reward\_safety fear\_safety reward fear

### Cleaned data data\_combine\_ext

Based on data\_no\_repeat\_all, with Jacklynn’s comments, for the run variable, I further merge Extincti and Run3, renaming all Ectinctis as Run3. The saved data set is data\_combine\_ext. Also note that the variable run’s value has been changed to learn, run1, run2, run3.

table(data\_no\_repeat\_all$run)

Extincti Learning Run1 Run2 Run3   
 2720 4960 4960 4960 2240

table(data\_combine\_ext$run)

learn run1 run2 run3   
 4960 4960 4960 4960

### Cleaned data data\_no\_missing

The data further cleaned is data\_no\_missing that removes all rows with no reaction values or physiological response. There are 2229 trials among 224 subjects that have physiological responses.

data\_no\_missing

# A tibble: 2,229 × 11  
 id run condition trial cs\_stim\_time cs\_scl cs\_latency cs\_amplitude  
 <chr> <chr> <fct> <int> <dbl> <dbl> <dbl> <dbl>  
 1 101 learn fear 1 95.9 9.91 2.20 0.735  
 2 101 learn fear 2 259. 7.68 2.26 2.31   
 3 101 learn fear\_safety 6 484. 9.05 1.34 3.66   
 4 101 learn reward 1 82.8 10.7 2.19 0.371  
 5 101 learn reward 4 365. 8.56 3.40 1.71   
 6 101 run1 fear 4 979. 10.7 2.35 1.15   
 7 101 run1 fear\_safety 3 1030. 10.9 4.18 0.911  
 8 101 run2 fear\_safety 6 1762. 13.1 2.53 0.528  
 9 101 run2 reward 5 1718. 9.58 1.48 3.61   
10 101 run2 reward\_safety 7 1875. 10.2 3.40 0.789  
# ℹ 2,219 more rows  
# ℹ 3 more variables: cs\_rise\_time <dbl>, cs\_size <dbl>, cs\_onset <dbl>

In the experiment, ID 238 and 239 do not have any reaction values, or physiological response to any trials. Both participants are removed.

setdiff(unique(data\_xlsx$id), unique(data\_no\_missing$id))

[1] 238 239

Later we use data\_no\_missing for analysis. We can always go back to other data sets or the raw data when we need to. The code for cleaning data is saved in 01-data.R.

## Data Summary

### Frequency

The frequency table of run is

data <- data\_no\_missing  
(freq\_react\_run <- data |> select(run) |> table() )

run  
learn run1 run2 run3   
 823 534 451 421

**Findings**: (Just observed results from data, not formal statistical inference)

* **Learning round has more reactions, and frequency of reaction decays with time passed.**

We can check the contingency table of run and condition.

safety reward\_safety fear\_safety reward fear freq\_react\_run  
learn 140 120 149 137 277 823  
run1 61 91 87 104 191 534  
run2 92 83 70 82 124 451  
run3 80 73 90 72 106 421  
 373 367 396 395 698 2229

safety reward\_safety fear\_safety reward fear freq\_react\_run  
learn 0.06 0.05 0.07 0.06 0.12 0.37  
run1 0.03 0.04 0.04 0.05 0.09 0.24  
run2 0.04 0.04 0.03 0.04 0.06 0.20  
run3 0.04 0.03 0.04 0.03 0.05 0.19  
 0.17 0.16 0.18 0.18 0.31 1.00

**Findings**: (Just observed results from data, not formal statistical inference)

* **Participants have more physiological responses to condition fear.**
* **For other conditions, their number of responses are not very different.**
* **The number of responses with fear and reward decreases faster. (Check their percentage).**

I then check how condition and trial are associated with the response frequencies. Frequencies with and without Learning run are both considered.

**Findings**: (Just observed results from data, not formal statistical inference)

* **Participants tend to have reactions in the first trial, especially for “fear” condition.**
* **The number of physiological response is decreasing with trial orders when the learning run is included.**
* **The decreasing pattern is not obvious for other condition types when the learning run is not included.**
* **The number seems to increase back a little bit in 7th and 8th trial.**



