# Evoman Generalist Statistical tests

Meena Kerolos

17 October, 2021

## Separate experiments

We will consider training against enemy group `7_8` as a different experimental setup from training against enemy group `3_7_8`. Therefore, training against one enemy group contains a set of experiments to compare between two EA instances. Each EA instance is tested against each of the 8 enemies. The experimental setup is factorial block design where the method is considered the treatment factor, the enemy is considered the block factor and the gain is the outcome.

Scheirer–Ray–Hare test will be used as it is a non-parametric test (doesn't have assumptions on data).

## Enemy Group 3_7_8

```
results_3_7_8 = results[(results$enemy_group == '3_7_8'),]
z = scheirerRayHare(gain ~ method + enemy, data = results_3_7_8); z
```

```
##
## DV:  gain
## Observations:  1600
## D:  0.993
## MS total:  213467
```

```
##                Df   Sum Sq    H p.value
## method          1 6.98e+05    3   0.070
## enemy           7 2.62e+08 1236   0.000
## method:enemy    7 1.53e+06    7   0.409
## Residuals    1584 7.47e+07
```

With a p-value of $0.409 > 0.05$, there is no significant interaction between method and enemy and we can look at the effect of each of them. It is known that enemy effect is significant and p-value of $0 < 0.05$ for the enemy effect supports it. However, a p-value of $0.07 > 0.05$ for the method effect is not significant. Therefore, GA and DE doesn't give a significantly different gain when trained on enemy group `3_7_8` and tested against all enemies.

## Enemy Group 7_8

```
results_7_8 = results[(results$enemy_group == '7_8'),]
z = scheirerRayHare(gain ~ method + enemy, data = results_7_8); z
```

```
##
## DV:  gain
## Observations:  1600
## D:  0.992
## MS total:  213467
```

```
##                Df   Sum Sq    H p.value
## method          1 1.38e+06    7  0.0106
## enemy           7 2.28e+08 1076  0.0000
## method:enemy    7 1.99e+06    9  0.2247
## Residuals    1584 1.07e+08
```

With a p-value of $0.225 > 0.05$, there is no significant interaction between method and enemy and we can look at the effect of each of them. It is known that enemy effect is significant and p-value of $0 < 0.05$ for the enemy effect supports it. A p-value of $0.011 < 0.05$ for the method effect is significant. Therefore, GA gives a higher gain than DE when trained against enemy group `7_8` and tested against all enemies.

Here are the mean and median estimators of the gain for each method:

- gain for GA method trained against enemy group `7_8`: mean=-28.501, median=-50
- gain for DE method trained against enemy group `7_8`: mean=-32.721, median=-50