Evoman Specialist Statistical tests - randomini

Meena Kerolos

01 October, 2021

Using Scheirer–Ray–Hare test

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

```
scheirerRayHare(gain ~ method + enemy, data = results)
```

```
##
## DV:  gain
## Observations:  1000
## D:  0.995
## MS total:  83417
```

```
##               Df   Sum Sq   H p.value
## method         1   124389   1  0.2209
## enemy          4 48122232 580  0.0000
## method:enemy   4  1318329  16  0.0032
## Residuals    990 33367831
```

There is indeed an interaction between method and enemy The enemy is already known to have an effect on the outcome We can't test the actual effect of the method because it will already have an effect through the interaction.

Separate experiments

Let's treat each enemy as a separate experimental setup The comparison is made between methods and we are interested in testing the difference between the populations of the set of outputs coming from each method. For that we will use Mann-Whitney test.

Enemy 1

```
enemy1_Static = results[(results$enemy == 1) & (results$method == 'EvomanAnnNeatStatic'),]$gain
enemy1_Dynamic = results[(results$enemy == 1) & (results$method == 'EvomanAnnNeatDynamic'),]$gain
z = wilcox.test(enemy1_Static, enemy1_Dynamic); z
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  enemy1_Static and enemy1_Dynamic
## W = 3962, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
```

```
p_value = z$p.value
```

This test is testing the null hypothesis H0: F = G which means that both sets are coming from the same population.

With a p-value of $0.009 < 0.05$, we can reject that hypothesis and conclude that there is indeed a treatment effect. Here are the mean and median of each set as an estimators of the location of each population:

- gain for Static method: mean=-29, median=-40
- gain for Dynamic method: mean=-7.7, median=-30

Considering the mean of each set as an estimator to the location of the population we see that the Dynamic method results in more gain than the Static method for enemy 1.

Enemy 2

```
enemy2_Static = results[(results$enemy == 2) & (results$method == 'EvomanAnnNeatStatic'),]$gain
enemy2_Dynamic = results[(results$enemy == 2) & (results$method == 'EvomanAnnNeatDynamic'),]$gain
z = wilcox.test(enemy2_Static, enemy2_Dynamic); z
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  enemy2_Static and enemy2_Dynamic
## W = 6738, p-value = 2e-05
## alternative hypothesis: true location shift is not equal to 0
```

```
p_value = z$p.value
```

This test is testing the null hypothesis H0: F = G which means that both sets are coming from the same population.

With a p-value of $1.981 \times 10^{-5} < 0.05$, we can reject that hypothesis and conclude that there is indeed a treatment effect. Here are the mean and median of each set as an estimators of the location of each population:

- gain for Static method: mean=26.58, median=40
- gain for Dynamic method: mean=-2.66, median=-10

Considering the mean of each set as an estimator to the location of the population we see that the Static method results in more gain than the Dynamic method for enemy 2.

Enemy 3

```
enemy3_Static = results[(results$enemy == 3) & (results$method == 'EvomanAnnNeatStatic'),]$gain
enemy3_Dynamic = results[(results$enemy == 3) & (results$method == 'EvomanAnnNeatDynamic'),]$gain
z = wilcox.test(enemy3_Static, enemy3_Dynamic); z
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  enemy3_Static and enemy3_Dynamic
## W = 6756, p-value = 2e-05
## alternative hypothesis: true location shift is not equal to 0
```

```
p_value = z$p.value
```

This test is testing the null hypothesis H0: F = G which means that both sets are coming from the same population.

With a p-value of $1.705 \times 10^{-5} < 0.05$, we can reject that hypothesis and conclude that there is indeed a treatment effect. Here are the mean and median of each set as an estimators of the location of each population:

- gain for Static method: mean=28.74, median=41
- gain for Dynamic method: mean=6.3, median=10

Considering the mean of each set as an estimator to the location of the population we see that the Static method results in more gain than the Dynamic method for enemy 3.

Enemy 6

```
enemy6_Static = results[(results$enemy == 6) & (results$method == 'EvomanAnnNeatStatic'),]$gain
enemy6_Dynamic = results[(results$enemy == 6) & (results$method == 'EvomanAnnNeatDynamic'),]$gain
z = wilcox.test(enemy6_Static, enemy6_Dynamic); z
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  enemy6_Static and enemy6_Dynamic
## W = 5600, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
```

```
p_value = z$p.value
```

This test is testing the null hypothesis H0: F = G which means that both sets are coming from the same population.

With a p-value of 0.088 > 0.05, we CANNOT reject the null hypothesis and conclude that both methods don't have different effects of the gain. Here are the mean and median of each set as an estimators of the location of each population:

- gain for Static method: mean=-93.8, median=-90
- gain for Dynamic method: mean=-95, median=-95

Enemy 7

```
enemy7_Static = results[(results$enemy == 7) & (results$method == 'EvomanAnnNeatStatic'),]$gain
enemy7_Dynamic = results[(results$enemy == 7) & (results$method == 'EvomanAnnNeatDynamic'),]$gain
z = wilcox.test(enemy7_Static, enemy7_Dynamic); z
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  enemy7_Static and enemy7_Dynamic
## W = 5100, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
```

```
p_value = z$p.value
```

This test is testing the null hypothesis H0: F = G which means that both sets are coming from the same population.

With a p-value of 0.806 > 0.05, we CANNOT reject that hypothesis and conclude that both methods don't have different effects of the gain. Here are the mean and median of each set as an estimators of the location of each population:

- gain for Static method: mean=-34.12, median=-60
- gain for Dynamic method: mean=-30.62, median=-60