# Assignment 2

This retrospective experimental study is on analysing the effect of EPO in the endurance of athletes. There are two type of data recorded in the experiment.

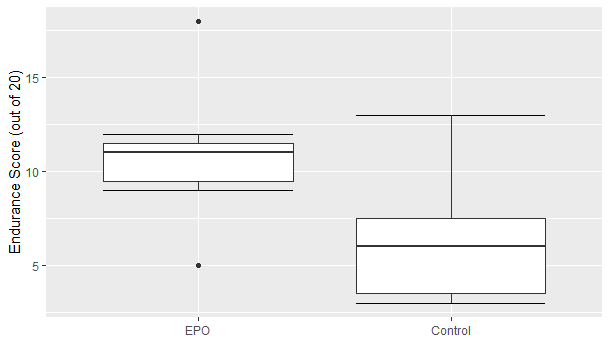
1. Whether an athletes took EPO on under Control (Qualitative)
2. Endurance core of each athlete when they took EPO and under control (Quantitative, Scale)

Since the experiment are carried out with a same set of athletes under two conditions (EOP and Control) this can be considered as related samples.

Sample size of this experiment is 7.

This analysis will validate the claim that EPO will improve the endurance of athletes.

**Boxplot**



It can be observed that in the boxplot of athletes who has taken EPO there are two outliers (5 and 18).

More than 50% of the data in Control sample lines below the first quartile of the EPO sample.

EPO boxplot is shifted towards up than the control boxplot suggests that there’s an improvement of the endurance but by overserving that only we cannot come into a solid conclusion.

|  |  |  |
| --- | --- | --- |
| **Statistics** | **EOD** | **Control** |
| sample size | 7 | 7 |
| Mean | 10.86 | 6.29 |
| 95% CI (lower) | 7.68 | 3.11 |
| 95% CI (upper) | 14.04 | 9.46 |
| Standard Deviation | 3.89 | 3.55 |
| Median | 11 | 6 |
| Min | 5 | 3 |
| Max | 18 | 13 |
| Skewness | 0.3829 | 0.7217 |
| Normally distributed | Yes | Yes |

Shapiro Wilk test suggest that both the distributions are normally distributed. (Skewness between -1 and 1, mean and median are roughly equal. These will support that claim)

Since the samples are normally distributed a best measure to describe centrality is mean and variance is the standard deviation.

Since the both samples are normally distributed and the experiment data can be paired (dependent) the correct statistical test to validate the claim is **Related T test.**

Based on the sample we are 95% confident that the true mean of the population of EOD will lies between 7.68 and 14.04 where true mean of the population Control will lie between 3.11 and 9.46.

**Statistical inference**

|  |  |
| --- | --- |
| Mean difference | 4.57 |
| 95% CI (lower) | 2.81 |
| 95% CI (upper) | 6.33 |
| p-value | 0.0007 |
| Effect size | 1.23 |

Null Hypothesis: There’s no difference in the population mean when athletes take EOP and when the same athletes takes Control.

Alternative Hypothesis: A difference exists in the population means when athlete take EOP and when the same athletes takes Control.

p-value (0.0007) < level of significance (0.05)

So we can reject the null hypothesis where a probability of making a type 1 error is less than 0.05.

This suggest that a mean difference of 4.57 was found to be statistically significant. (ES: 1.23, 95% CI [2.81,6. 33]).

Based on a sample of 7 athletes who took EOP and same athletes who took a control, we are 95% confident that the mean difference (EOP – Control) in the endurance lies between [2.81,6. 33]. The interval is always positive. This suggest that on average athletes taking EOP will have a higher endurance than the athlete who take control.

The difference between two quantitative samples is known as effect size. When the difference is high the effect size also will be higher and when the difference is low the effect size will also be lower.