STA305H/1004H Final Project Report

Spencer Y. Ki (1003165031)

2020-08-19

Description

Motivation

Valorant is a first-person-shooter video game where teams compete to complete various objectives. Billed as an "Esports ready" game, the mechanics of the game are non-randomised. Thus, the only uncontrolled covariate in-game is the players themselves. Toronto-based *Valorant* Team 98534 is undergoing an internal dispute. In question are two factors: 1.) the superiority of Player "AK" over Player "CS" with regard to general reflexes and game skill; and 2.) the superiority of the in-game "Phantom" rifle over the in-game "Vagrant" rifle with regard to usefulness in general gameplay. 'Usefulness' in this instance is understood to mean the factor by which the rifle simplifies the 'killing' of enemies; a core gameplay objective.

Methods

Factorial design is particularly appropriate in this scenario, as there is the possibility that one player is superior to the other with the use of one weapon but not the other. A 2^2 experiment was designed, with AK and CS as the two levels of the "Player" factor, and the Phantom and the Vagrant as the two levels of the "Gun" factor.

A single experimental trial involved fixing each factor at one level, and proceeding through the following procedure. *Valorant*'s "shooting range" mode was entered, which features standardised, computer-controlled "bots" to practice against in lieu of human opponents. In the set-up options, "Bot Armor", "Infinite Ammo", and "Bots Strafe" were activated, allowing the closest, controlled approximation to normal gameplay against human opponents. The subject then proceeded to play through the "Eliminate 100 Streak" mode: this involved the 'killing' of 100 bots in as short a time as possible. The time taken to complete this trial was recorded as the outcome of interest. The observations test $H_1 : \mu_a \neq \mu_b$ against $H_0 : \mu_a = \mu_b$ at a 90% confidence level, with μ_i representing mean time for trial completion at experimental condition i.

64 trials occured over 16 days, allowing a uniform 4 trials per day. Given a 2^2 design, this permitted 16 replications of each experimental condition. Trials were randomly ordered, to permit genuine replication. In the deviation coding matrix, AK was coded 1 and CS was coded -1 for the Players, with the Phantom coded 1 and the Vagrant coded -1 for the Guns.

Design Matrix

Given the large number of replications, it is infeasible to include a full design matrix with the spacing limitations on this report. Accordingly, the following presents the experimental factors and their levels alongside the mean time to complete a trial for each respective experimental condition:

Experimental_Condition	Player	Gun	Mean_Time
1	1	1	165.3125
2	1	-1	169.5625
3	-1	1	161.5625
4	-1	-1	165.4375

1

Statistical Analysis

Verifying the independence and normality of the results is trivial due to the large number of observations. As can be seen in Figure 1, the subset of observations for each experimental condition are distribted in a manner that is sufficiently close to Gaussian for analysis. This conclusion is partially supported by the subsets' Shapiro-Wilk test p-values, three of which are greater than 0.10 ($p = 0.16$ for experimental condition 1, $p = 0.42$ for experimental condition 2, and $p = 0.68$ for experimental condition 3). Interestingly, the Shapiro-Wilk test p-value of experimental condition 4's obervations is $p = 0.045 < 0.05$, the possible interpretation of which is discussed in the conclusion below.
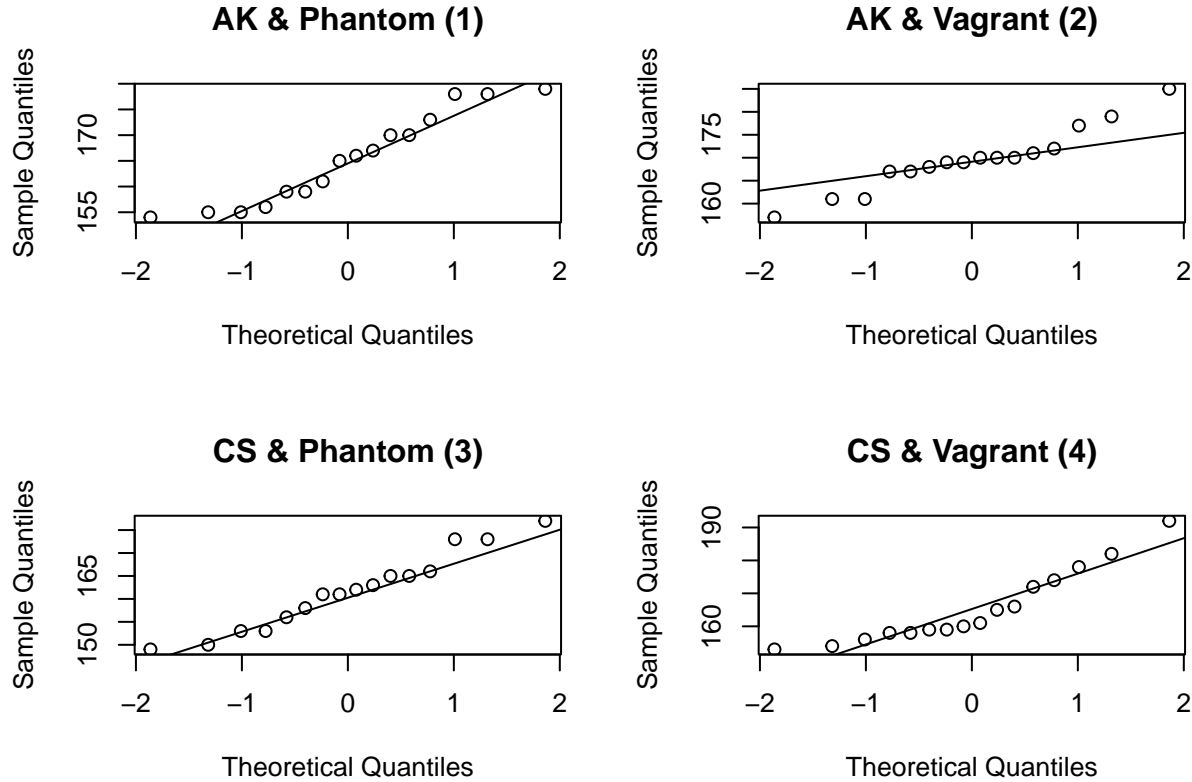

Figure 1: Q-Q plots of subsets of obesrvations by experimental condition

This normality permits the calculation and interpretation of the main and interaction factor effects. By considering the changes in mean trial completetion time between experimental conditions, one determines that the main effect of the Player factor is `3.94` (when changing from CS to AK) and the main effect of the Gun factor is `-4.06` (when changing from Vagrant to Phantom). When fixing the Player factor, factor level AK feaures a Gun effect of `-4.25` (when changing from Vagrant to Phantom), while factor level CS feaures a Gun effect of `-3.88` (when changing from Vagrant to Phantom). When fixing the Gun factor, factor level Phantom feaures a Player effect of `3.75` (when changing from CS to AK), while factor level Vagrant feaures a Player effect of `4.12` (when changing from CS to AK). These interactions can be observed graphically in Figure 2, and when combined yield a two-factor interaction effect of `-0.19`.

These findings are supported by modelling the observations linearly. An R-constructed linear model of Mean Trial Completion Time as a function of observed Player and Gun levels yields an estimated Player variable coefficient of `1.97` ($p = 0.08$), an estimated Gun variable coefficient of `-2.03` ($p = 0.07$), and an estimated interaction variable coefficient of `-0.09` ($p = 0.93$). These correspond to estimated factorial effects of `3.94`, `-4.06`, and `-0.19` respectively, highly similar to the ones calculated above.
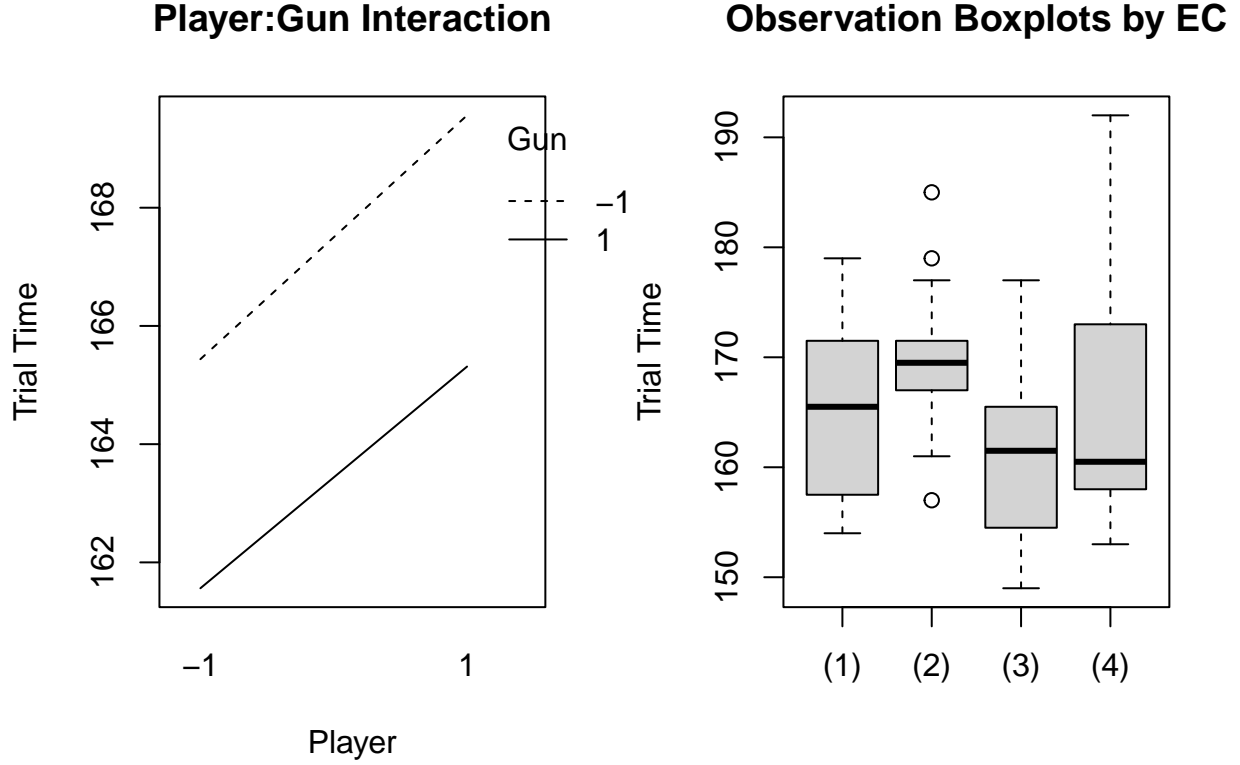
Figure 2: The main factors are largely non-interacting. Figure 3: Observations by experimental condition

Given that each experimental condition underwent 16 replications, each condition's sample variance may be calculated instead of estimated. Together, these condition variances yield a total pooled sample variance of `79.22`, with 4 degress of freedom. Since each factorial effect is a difference between 2 averages of 16 observations, the variance of factorial effect can also be calculated to be `2.48`, resulting in a standard error of factorial effect of `1.57`.

As this analysis is dependent on the assumption that the observations are independent and normally distributed, this standard error allows one to normalise the factorial effect for comparison to a t-distribution: $effect/se_{effect} \sim t_4$. As such, $effect \pm t_{4,0.1/2} * se_{effect}$ allows one to construct a 90% confidence interval for each factorial effect.

Accordingly, the calculated Player factor effect was found to have a 90% confidence interval of [`0.58` , `7.29`], the calculated Gun factor effect was found to have a 90% confidence interval of [`-7.42` , `-0.71`], and the calculated interaction factor effect was found to have a 90% confidence interval of [`-3.54` , `3.17`]. These intervals are highly similar to the 90% confidence intervals that can be retrieved from the previously-reference, R-constructed linear model:

	5 %	95 %
(Intercept)	327.2200521	334.6549479
Player	0.2200521	7.6549479
Gun	-7.7799479	-0.3450521
Player:Gun	-3.9049479	3.5299479

Conclusion

The formal result of the analysis is that the null hypothesis may be rejected, and that there is evidence that both player and gun factor level alter the time required to complete an experimental trial at a 90% confidence level. Given the very small and statistically insignificant interaction factor effect – calculated to be `-0.19` and estimated by the linear model to be `-0.19` ($p = 0.93$) – main factor interaction can be considered to be negligible and each main factor should be interpretted seperately. Accordingly, the Gun factor effect – calculated to be `-4.06` and estimated by the linear model to be `-4.06` – indicates that trials where the Phantom is used complete 4.06 seconds faster than trials where the Vagrant is used. Similarly, the Player factor effect – calculated to be `3.94` and estimated by the linear model to be `3.94` – indicates that trials performed by AK complete 3.94 seconds slower than trials performed by CS. Neither the Player factor effect's nor the Gun factor effect's 90% confidence intervals include 0, and the Player and Gun variable coefficients estimated by the R-computed linear model have p-values of `0.08` and `0.07` respectively. Therefore, the main effects can be judged to be statistically significant at the 90% confidnce level. Calling back to the original purpose of the experiment, one may conclude that the Phantom is a better rifle than the Vagrant, and that Player CS has superior game-skill to Player AK.

Thorough measures have been employed in-game to mitigate the effects of confounding and interacting covariates. The selected in-game options permit the bots to move semi-randomly, though their slow speed assures that any effect this has on outcome can be rendered negligible with replication. The player is also not required to reload their weapon, while "bot armour" requires more shots for a 'kill'. These options promote an outcome dependent solely on the player's reflexes and control of the weapon. The pre-programmed nature of the shooting range additionally means that there should be no difference in set-up for the experiment.

In reality, both players scheduled their daily two trials immediate and consecutive to one another, dependent on the assignment of the design matrix. To assure similar levels of 'practice', trials were scheduled immediately after the team had played at least one game for the day, to ensure that both subjects had experienced similar quantities of gameplay time in the near past. It is worth noting that since *Valorant* was only released to the public on the 2nd of June, 2020. With Team 98534 pre-dating the game's release, both subjects in question had near identical experience with the game prior to the experiment.

There are two principal covariates that may have resulted in bias. While both players have comparable computer systems, Player CS has a significantly superior GPU (an Nvidia GeForce RTX 2080 in comparison to AK's Nvidia GeForce GTX 1060) as well as a computer monitor with a refresh rate of 240 Hz (in comparison to AK's 119 Hz monitor). Considering the importance these components have to Esports-type video games, it would not be unreasonable to expect that they played a role in CS's evident superiority.

Finally, it is worth considering that genuine replication is not possible with these types of experimental trials. To interpret that insight: there exists the possibility that Player CS was improving their game-skill over the 16 day experimental period, while Player AK's had stagnated. This is suggested by the Shapiro-Wilk test p-value for experimental condition 4 that was highlighted in the beginning of the statistical analysis. A Shapiro-Wilk test p-value of `0.045` indicates that the subset of observations that belong to experimental condition 4 are not drawn from a Gaussian distribution. Additionally, one may fit a nonlinear least squares exponential decay model in R with the subset's observations as the dependent variable and the run number as the independent. This yields a decay coefficient with a p-value of `0.11`; not conclusive at a 90% confidence interval, but suggestive. One should note that experimental condition 4 is a combination of Player CS and the Gun Vagrant. Anecdotally, CS apparently solely uses the Vagrant in general gameplay, and this could be a cause for their apparent improvement in its use over the duration of the experiment. However, further experimentation would need to be carried out before analysis can be performed.