Carina Rammelkamp
May 9, 2013
CS185 Spring 2013

Results

For this User UI Evaluation, I timed three individuals on how long it took them to each enter in 15 total NBA games to two different applications. Before doing the experiment, I would have each individual practice entering in information to each of the apps using a sample list of 10 or so NBA data points. Then, for the experiment, each individual would be first be timed entering in 15 data points to the first application, then the same individual would be timed entering in those same 15 data points to the second application. Three separate 15-point data sheets were used, one for each individual. Additionally, the data was randomized. From the structure of this experiment, I was able to perform a paired t-test because I was comparing two different treatments where the measurements were applied to the same subjects. The two different treatments were the use of the two applications, the multi-screen interface and single-screen interface. Before delving into the results of this experiment, it is imperative to mention that the sample size was nearly negligible, so even if the results of this experiment yielded statistical significance, we would need a greater sample size to discover anything of prominence.

Looking at the paired t-test result, I obtain a p-value of 0.847 indicating that the results are not statistically significant. Thus, we are **unable** to refute the null hypothesis; the null hypothesis states that the time for a user to enter in 15 data points to each the multi-screen and single-screen applications is the same across applications.

Because I was one of the individuals timed in this experiment, we saw a variety of results. Firstly, when analyzing the pre-test surveys, we see that the other two individuals were around age 55. Because I am around 22, we see a large age difference in the individuals, which may have been a big contributor to the outcome of the experiment. Additionally, the difference in age for the three individuals may indicate a sample that better reflects the general population. Another factor that may have contributed to the experiment outcome is that I have more experience using smartphone applications, while the other individuals were not as experienced. Having subjects with little experience may have contributed to discovering the true usability of each of the apps, although it likely brought up the respective user times to complete data entry. Both subjects with little smartphone experience were adamant about mentioning the faults of the multi-screen interface when the experiment was over, and this is reflected in their post-study survey. What they mention here is that the single-screen interface was better, but only because of the difficulty posed when entering the second team name of the multi-screen interface. Overall, I believe it would be best to repeat this experiment with a larger sample size and a better formatted multi-screen interface.