Question 1

Before evaluating the effect of an experiment, it is important to make sure that the experiment was executed correctly. Check whether the test and control groups are probabilistically equivalent on their observables?

a. More specific, compare the averages of the income, gender and gamer variables in the test and control groups. You should also report the % difference in the averages. Compute its statistical significance.

(Calculations are in R code)

For income level of experiment group users and control group user, income level is $0.2277 higher in experiment group and the two sample t test shows that the difference is not statistically significant. So we can conclude that income level in two groups are comparable.

For proportion distribution of gender, there are 64.8% users are male in control group and 64.72% male in exp group. The difference is slight. And the two-sample proportion test indicates that the gender proportion are comparable.

For proportion distribution of gamer or non-gamer, there are 60.18% users are gamers in control group and 60.13% gamers in exp group. The difference is slight. And the two-sample proportion test indicates that the gamer/nongamer proportion are comparable.

Since the three observable variances are all comparable in both groups, we can conclude that the experiment setting is reasonable.

b. Briefly comment on what these metrics tell you about probabilistic equivalence for this experiment.

For probabilistic equivalence regarding income, we will look at the mean value of income. The large number theory and Central Limit Theorem tells us that the sampling distribution of mean will follow normal distribution. By calculating the standard error, we can easily detect whether the income level of experiment group lies outside the 95% confidence interval and then tell whether they have a significant difference or not.

For gender and gamer proportion, since the value of “male” and “gamer” are 1 and the other side value is 0, it’s also a mean value; or considering the proportion, the CLT also tells us that the sampling distribution of proportion value will follow normal distribution as long as $n*p>5$ and $n*(1-p)>5$, so the story is the same here.

c. If you had run this type of analysis BEFORE executing an experiment and found a large difference between test and control groups, what you should do?

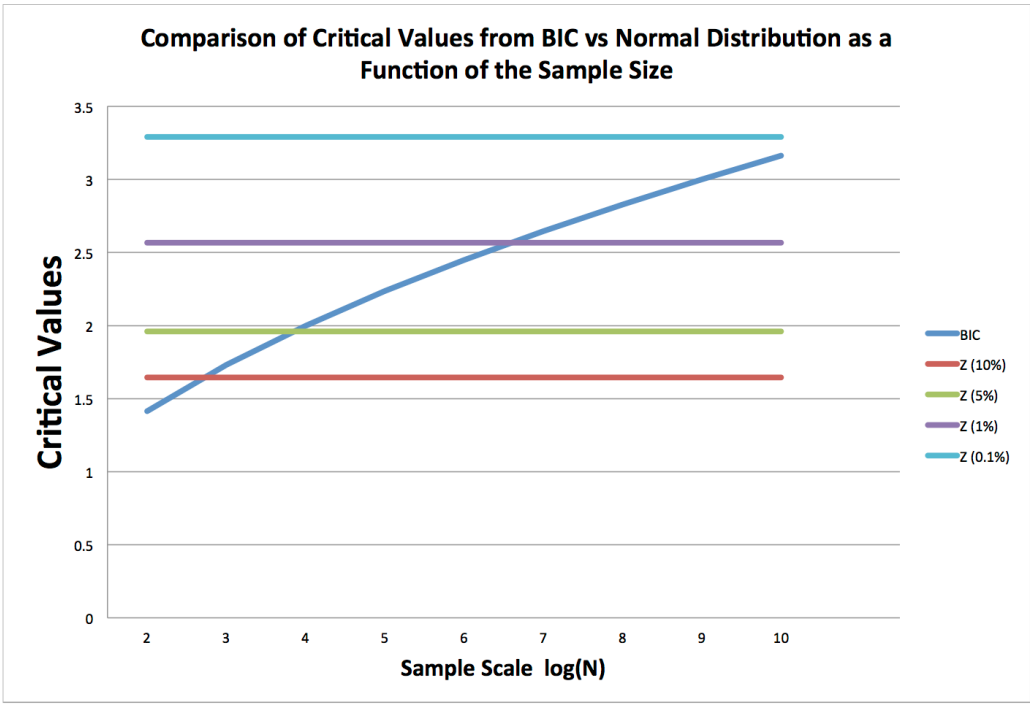
If I found a difference between test and control groups, I can think of three options regarding how to move on:

1) Confirm with the engineer team that all the experiment setting and how they randomly assign the units of test. If setup is wrong, then fix it and do a new round of test. (This is more probable!)

2) Don't drop the current collected test-unit, use matching instead to generate functionally similar groups! I would make these groups similar based on these "observable", or in our case, "reported", characteristics: gamer/non-gamer, male/female and income.
3) In rare cases where there are lots of covariates in the tests and we cannot find good matching of people, we can match people with similar propensity scores

d. If you had millions of consumers, your "classic" statistical significance tests would not work (this is because the number of samples is used to compute those classic statistical tests). Do some research online and propose what significance test would you do in case you had "big data"?

One possibility for addressing this problem is to let the critical value be a function of the sample size. Cameron and Trivedi (2005) suggested in the book 'Microeconometrics: Methods and applications' that Bayesian information criterion(BIC) increases the penalty as sample size increases using a two-sided t-test critical value of $\sqrt{\ln(N)}$, whereas traditional hypothesis tests at a significance level such as 5% do not. For example, for nested models with q2 = q1 + 1 choosing the larger model on the basis of lower BIC is equivalent to using a two-sided t-test critical value of √ ln N, which equals 2.15, 3.03, and 3.72, respectively, for N = 10^2,10^4 , 10^6.



Question 2

Evaluate the average purchase rates in the test and control for the following groups. For each comparison, report the average purchase rate for the test, average purchase rate for the control and the absolute difference (not the % difference) between the test and control.

a. Comparison 1: All customers **(calculations are in R code)**

control purchase rate: 3.62%
experiment purchase rate: 7.68%
exp-control: 4.06%
The difference is statistically significant.

b. Comparison 2: Male vs Female customers **(calculations are in R code)**

male control: 3.71%
male exp: 7.46%
male exp-control: 3.74%
The difference is statistically significant.

female control: 3.44%
female exp: 8.09%
female exp-control: 4.65%
The difference is statistically significant.

c. Comparison 3: Gamers vs Non-Gamers Customers **(calculations are in R code)**

gamer control: 3.54%
gamer exp: 10.45%
gamer exp-control: 6.91%
gamer difference is statistically significant.

nongamer control: 3.74%
nongamer exp: 3.51%
nongamer exp-control: -0.23%
The difference is not statistically significant.

d. Comparison 4: Female Gamers vs Male Gamers **(calculations are in R code)**

male gamer control: 3.71%
male gamer exp: 10.14%
male gamer exp-control: 6.41%
The difference is statistically significant.

female gamer control: 3.20%
female gamer exp: 11%
female gamer exp-control: 7.81%
The difference is not statistically significant.

Question 3

a. Comparison 1: All customers **(calculations are in R code)**

Expected revenue of all control group is $1.36 per person and expected revenue of all

experiment group is $0.96 per person.

b. Comparison 2: Female Gamers vs Male Gamers (calculations are in R code)

Expected revenue of male control group is $1.4 per person and expected revenue of male experiment group is $1.27 per person.

Expected revenue of all control group is $1.2 per person and expected revenue of all experiment group is $1.38 per person.

Question 3

a. Based on your previous answers, provide a brief recommendation to your management team summarizing the expected financial outcome for Game-Fun.

According to the t-test on purchase rate of different segmentation of user base, we can find that the campaign is especially successful to the female gamer experiment group, with 7.81% lift with the purchase after the campaign. When the advertisements are applied to female gamer, we can also anticipate an increase of K-factor cause gamers usually have their special gamer circle where gamer can recommend other player gamefun's games. That will be the resources of great organic and free customers.

Considering the expected revenue per user, implementing the advertisement on the whole user group is not worth it. Further, with the whole user base sample size, there may be "big data curse" in a way that it's not a real significance.

Overall, I recommend the company offer the promotion targeting on female gamers.

References:

Cameron, C. A., and P. K. Trivedi (2005): Microeconometrics: Methods and applications. New York, NY: Cambridge University Press.