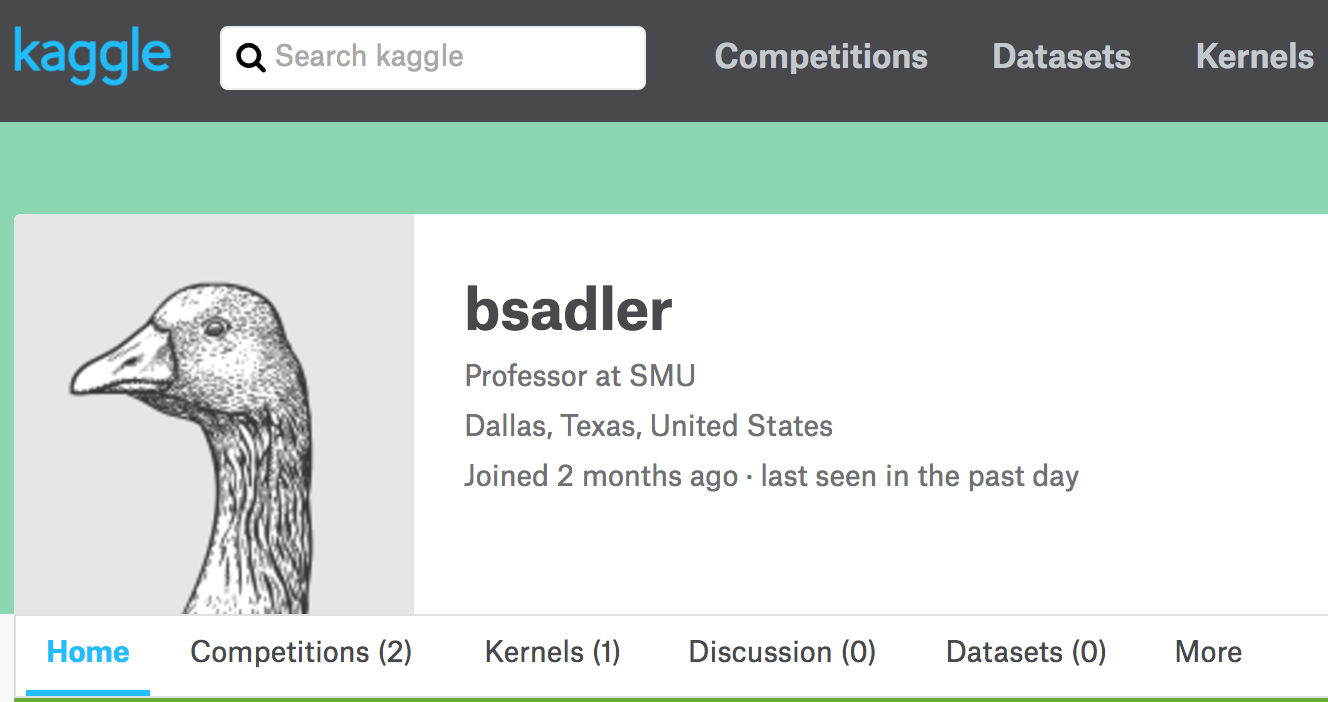
MSDS 6372: Homework 1

Chapter 23 Problem 18. Make sure and describe your experimental design including any treatments, factors, levels etc. Also remember to include a sample size estimate with respect to the given assumptions. Use SAS proc glmpower to get your estimated sample size and use proc power to verify. Provide your code and screenshots of the output for each proc. Finally, calculate an estimate of the sample size using the formula on page 709. Note that the textbook says that the ‘n’ on page 709 represents the sample size per group when it actually represents the total sample size of all the groups.

1. At least two of the projects we will do in this class will be Kaggle projects. Kaggle is a website that stages worldwide competitions and is really a community for data scientists. We asked the several recruiters over a few MSDS immersions and all of them indicated they wanted to see interest and work in data science outside of the coursework (HW, Tests, etc) and many of them specifically stated that performance in Kaggle competitions was a preferred method of showing this outside interest and work. For this reason we will make a large portion of the coursework Kaggle projects! In order to do this you will first need to go to Kaggle.com and register. No doubt many of you already have Kaggle accounts and if you do you may use your existing account.

For this assignment you will need to

1. Provide evidence of registering in Kaggle with a simple screen shot like the one below. If you don’t already have your Kaggle account, make sure your username has your last name in it.

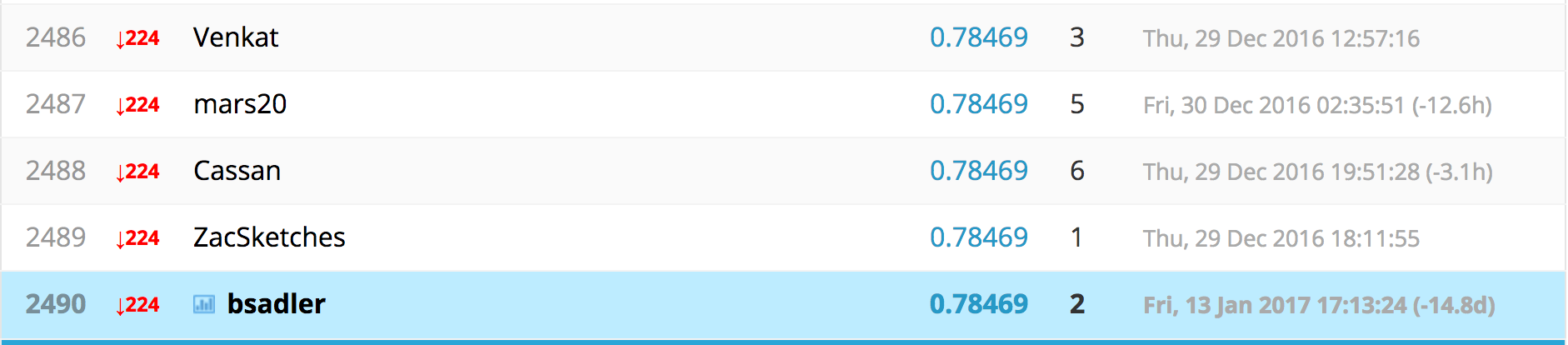


1. Learn the process of competing and submitting responses. To do this you will need to find the Titanic competition or simply navigate to this website: <https://www.kaggle.com/c/titanic>

Once here you will need to follow these steps:

1. Click on “Get the data”
2. download the train.csv and test.csv data sets.
3. Import both into SAS. Import the train.csv into a data set called titanictrain and the test.csv into a dataset called titanictest. It must be these names for it to work with the code I am going to give you.
4. Run the SAS code found in the HW1 folder on Canvas: SAS Kaggle Code Titanic.txt. This will produce an output data set called finalp.
5. Export finalp to your harddrive. Let’s call it submission.csv
6. Go back to the Kaggle website and now click on “Make a Submission”.
7. Submit the submission.csv file to Kaggle.
8. Take a screenshot of your rank in the list and paste it on the homework.

Example



That’s it for this week!