

# Facebook Friends: An Activity in Statistical Analysis



*Facebook Friends*

**Developed by Matt Jones**  
**Presented by Loretta Griffy and Ellen Smyth**

Austin Peay State University  
Mathematics Department  
<http://apsu.edu/smythe/presentations.htm>

## Materials



- Internet Access
- Facebook Accounts and Friends
- TI-83 or 84

*Facebook Friends*

In order to complete this project, participants will need internet access and specifically access to Facebook. Several of the participants, but not necessarily all, will need to have pre-existing Facebook accounts with several Facebook friends, which we will refer to simply as friends from now on. Participants will also need to have a TI-83 or -84 calculator.



Split into groups of five, and choose as a group ten random Facebook friends.

How?

*Facebook Friends*

Divide the participants up into groups of a convenient size, and make sure at least one person in each group has a Facebook account with a couple of dozen friends or more. This group should collectively find ten of their Facebook friends. If every person in the group has a Facebook account, the group might be best choosing two friends from each participant. Once the

group decides how many friends each participant is choosing, these friends must be selected randomly from the total friends in the participant's network to create our random sample. The instructor may want to talk about why it is important to choose randomly and will certainly want to discuss how to choose randomly. There are many ways, but perhaps the most efficient way is to get the network of friends into a list, number the list, and randomly choose two (or five or ten, etc.) from that list using a random number generator. Since we are emphasizing the TI calculators, we will use the random number generator on the TI.

1. Select **Profile** and record your number of friends.
2. Go to **Account** and **Edit Friends**.
3. Select **Friends**.



Facebook Friends

To get the total number of a participant's friends, click on the profile link, which can be found in the upper left corner of Facebook. Then, on the Profile page, the number of total friends can be found along the left side.

To get the participant's network of friends into a list, click on the account link, right next to the profile link, and select "Edit Friends." Then, along the

left side menu, select the Friends icon. This action will list all of the friends in the participant's network with fifty friends per page – or at least that is what the action does in my account.

4. With the TI, Math > PRB > randInt(
5. randInt(1, [max#friends], [2 to 10])
6. From Facebook, use these random numbers to choose the friends, and record the number of friends that these random friends have.



Facebook Friends

To randomly select friends from the participant's network, we will use the random number generating function, randInt, on the TI. To get to this function, press Math and then scroll right to PRB. Then choose option 5 for the randInt function. This function takes in three things: the numbers you want as the minimum and maximum for your random range and the amount of

random numbers you want to generate, where the function looks like randInt(min, max, # of #s). If the participant wishes has 130 friends in her network and is supposed to choose five randomly, for example, she would do randInt(1, 130, 5) and press enter to get five random numbers between 1 and 130. If her numbers were 128, 37, 36, 16, and 7, then she would find her 128<sup>th</sup>, 37<sup>th</sup>, 36<sup>th</sup>, 16<sup>th</sup>, and 7<sup>th</sup> friends. Remember that there are fifty friends per page. Also note that, in some cases, the calculator will give duplicates, in which case you should throw out the duplicates and replace with new numbers using the randInt function again.

What is the shape of the distribution of the number of friends for Facebook users?



Facebook Friends

Have participants guess the distribution shape, and you may want to have them help you draw the various shapes. After you draw the shapes, talk about how you would label the axes with total number of friends on the x-axis and frequency of Facebook users on the y-axis.

Distribution



Most students guess normal, but is it?

Facebook Friends

Students will tend to guess normal, from our limited experience, because that is the distribution we talk the most about. But a normal distribution would mean no or very few Facebook users would have zero or just a handful or friends. It would also mean that the curve would be symmetrical around some specific mean. Discuss if that would be the case or not.

What is the average number of friends for all Facebook users?



Most students guess 200 to 400 average friends.



Facebook Friends

Have participants discuss what they would guess the average number of Facebook users to be. These guesses will likely be tainted by the data they've collected, so you may prefer to have them guess before gathering the data. From our limited experience, students tend to guess between 200 to 400 friends, and that range is a decent representation of their own number of friends.

What does Facebook claim to be the average number of friends?



<http://www.facebook.com/press/info.php?statistics>

Now, things can start to get interesting. Facebook keeps a lot of statistics on users, and they claim that the average Facebook user has just 130 friends in his or her network. If your guess is 300 and Facebook says it is 130, how do we explain that discrepancy? How can we use our data, our sample, to test whether Facebook's average of 130 is correct or not?



Is the claim from Facebook significantly different from your guess?

That depends on the guess and it also depends on what we mean by "significantly." Being a statistical activity, we should think about whether the difference is statistically significant at the 1%, 5%, or even 10% level, although we may just want to compute the p-value without choosing a particular significance level.

### Hypothesis Test

Test whether the average number of friends is different from the Facebook claim with your own data.

$H_0$ :

$H_a$ :



So, we are running the hypothesis test with our students, and the first steps should be to choose the null and alternative hypotheses. Our alternative hypothesis may be different depending on what guess we've given for the average, but it will probably be as follows.

$H_0$ : The average number of Facebook friends each user has is, as Facebook claims, 130.

$H_a$ : The average number of Facebook friends each user has is more than 130.

**AP Austin Peay State University** Assumptions



Are the necessary assumptions satisfied?

1. Simple random sample
2. Normal distribution or large sample size
3. Known  $\sigma$

*Facebook Friends*

Let the participants discuss each of these. The first assumption we have failed on, though we can hold off on bringing that to their attention. We have chosen a simple random sample of Facebook users among the friends of the participants, but this limited pool is hardly random to start with. We will hold off on discussing that for now. We also don't have a normal distribution, but we can let participants see that for

themselves when we look at the distribution our sample gives. We hopefully will have a large sample size so that this assumption actually will be met. We do not know sigma, and we should point that out. Even though sigma is unknown, we can still do a t-test.

**AP Austin Peay State University** Compile Data

Take data from all of the groups and add that data to L1.  
Stat > Edit > put data in L1



*Facebook Friends*

Now you will want to get all of the data together, write it on the board, and have participants add the data values into L1 in the calculators (Stat > Edit). Since each group is collecting 10 values, you will hopefully get 30 to 60 data values. If you have more than 6 groups, you may want to have participants collect only five values for each group.

### Histogram



2<sup>nd</sup> > STAT PLOT > ENTER  
 Highlight **On** and press ENTER  
 Type: histogram  
 xList: L1  
 Freq: 1  
 Zoom > 9

What is the shape?  
 Was your guess close?

*Facebook Friends*

Have participants generate a histogram with the sample data, walking them through with the instructions on the left. If students get errors, make sure this stat plot was entered correctly, all other stat plots are turned off, and the Y= graphs have been cleared.

The shape of the data will likely be right-skewed, and that is most probably the shape of the overall population even

though our sample is imperfect.

### t-Test



Stat > Tests > 2: T-Test...  
 Use the Data option  
 $\mu_0$ : 130  
 List: L1  
 Freq: 1  
 $\mu$ : > $\mu_0$   
 Calculate

*Facebook Friends*

Have participants do a t-test with the sample data using the steps on the left. The p-value you get will vary depending on your sample data and the number of data values you collect, but the p-value will probably be low, especially low in our experiences with college students.

### Test Results



What is the average number of Facebook friends from our sample data?  
 Is it close to our initial guess?  
 Is it close to Facebook's claim?

*Facebook Friends*

When we run the test, the calculator will give us  $\bar{x}$ , which is our sample average. We can compare the average of our sample to our guess and to Facebook's claim of 130.

- What is  $t$ ?
- What is  $p$ ?
- Can we reject  $H_0$  and accept  $H_a$ ?
- What is our official conclusion, then?



Facebook Friends

Record the  $t$ - and  $p$ -values your calculator gives, and discuss what that means for the null hypothesis. Discuss the strength of the evidence against the null. Decide whether to reject or not depending on  $p$ , though you will most likely want to reject. And then interpret what rejecting Facebook's claim really means.

- Should we email Facebook and let them know their stats are wrong?
- What do they have that we don't have?
- Why might our numbers be different?
- What assumptions might be erroneous in our sample?



Facebook Friends

So, what do we do now that we have statistically *proven* Facebook's claim to be false. Should we email them and tell them how wrong they are? Why are they so wrong, so different, to begin with? Perhaps they are not the ones who are wrong.

What would Facebook have access to that we don't? Well, Facebook has

access to the entire population whereas we only had a sample, which means Facebook could easily compute the average of the entire population leaving little room for error. Then why are we different? We didn't have the whole population, but just our sample, but that should be good enough to estimate... unless one of our assumptions is off. Which assumption might be wrong?



- Are the necessary assumptions satisfied?
1. Simple random sample
  2. Normal distribution or large sample size
  3. Known  $\sigma$

Facebook Friends

We don't need the third assumption since we used the  $t$ -test, and we should have had a sufficiently large sample. So that only leaves the first assumption. Even though our sample was randomly chosen among the friends of our participants, the friends of our participants are not random at all. We have severely limited our pool from all Facebook users to the friends of our participants, and so our results will

reflect not the general Facebook population but the friendship networks of our participants. Because this activity is geared for college students, their friends are much more likely to be college students. Facebook tends to be much more popular with college-aged users than the general population, and college students tend to be more social than the general population with wide and varying networks of friends. So this subpopulation from which we drew our sample typically gives us a much higher average number of Facebook friends than Facebook is reporting, but this difference should not be surprising since we aren't looking at the entire population.

Another possible explanation for the low Facebook number could be that Facebook is counting all user accounts, whether active or not, and there are probably thousands of inactive accounts with little or no friends pulling the average down.

Let participants discuss these explanations and draw their own conclusions. The 130 number could be wrong or outdated. Our sample is certainly flawed. How could we get a better sample? Discuss the possibility of truly choosing Facebook users at random. Do we have access to the random Facebook user, though? Is there anything we can do better with our limited access?

## Sources for Images

[http://www.masternewmedia.org/images/free\\_video\\_learning.jpg](http://www.masternewmedia.org/images/free_video_learning.jpg)  
<http://www.onlineuniversities-weblog.com/50226711/36725881.jpg>  
<http://www.davis.k12.ut.us/district/dsdonline/images/A89809963FB045ED9979A8B5B746E41E.jpg>  
<http://www.cel.sfsu.edu/images/programs/online-courses/web-based-learning-bed.jpg>  
<http://mbaforbetterfuture.com/wp-content/uploads/2008/02/online-learning2.jpg>  
<http://www.calnook.com/girl.jpg>  
<http://www.svcc.edu/UserMenu/employees/it/images/learning-styles.jpg>  
[http://www.appcon.com.au/Portals/0/images/11002000962\\_s.jpg](http://www.appcon.com.au/Portals/0/images/11002000962_s.jpg)  
<http://moodle.beauchamp.leics.sch.uk/moodle/file.php/1/images/Laptop4.jpg>  
[http://static.disaboom.com/content/images/articles/content/4976.10751.woman\\_onlineLearning.jpg](http://static.disaboom.com/content/images/articles/content/4976.10751.woman_onlineLearning.jpg)  
[http://www.onlinetrainingpartners.com/FCKeditor\\_files/Image/HomePage.jpg](http://www.onlinetrainingpartners.com/FCKeditor_files/Image/HomePage.jpg)  
[http://static.disaboom.com/content/images/articles/content/4931.10471.onlinelearning\\_large.jpg](http://static.disaboom.com/content/images/articles/content/4931.10471.onlinelearning_large.jpg)  
[http://www.pmpsucces.com/file.php/1/Online\\_Learning.jpg](http://www.pmpsucces.com/file.php/1/Online_Learning.jpg)  
[http://learn.intuit.com/files/images/66057\\_1\\_%281%29.JPG](http://learn.intuit.com/files/images/66057_1_%281%29.JPG)  
[http://www.findingdulcinea.com/docroot/dulcinea/fd\\_images/news/Americas/July-08/Critics-Battle-Over-Online-Learning/news/0/image.jpg](http://www.findingdulcinea.com/docroot/dulcinea/fd_images/news/Americas/July-08/Critics-Battle-Over-Online-Learning/news/0/image.jpg)  
[http://static.squidoo.com/resize/squidoo\\_images/-1/lens1374357\\_1244750526online-learning.jpg](http://static.squidoo.com/resize/squidoo_images/-1/lens1374357_1244750526online-learning.jpg)  
<http://www.bellevueuniversity.us/ccap/images/CreditGirl-sm.jpg>  
[http://www.dctc.edu/assets/pics/spring-2010/online1\\_big.jpg](http://www.dctc.edu/assets/pics/spring-2010/online1_big.jpg)  
[http://www.collegecrunch.org/wp-content/uploads/2008/12/engaged\\_student.jpg](http://www.collegecrunch.org/wp-content/uploads/2008/12/engaged_student.jpg)  
<http://www.pmi.org/AboutUs/PublishingImages/AdvocatingProfession/CaseStudyLibrary.JPG>  
<http://learning.unog.ch/Portals/0/images/skillport.JPG>  
<http://www.websitesandsoundbites.com/servic1.jpg>  
<http://awangshamsul.net/wp-content/uploads/2010/01/apple-lecture-hall.jpg>