

NAME: Jillian Edrington

Answer the following questions showing all work. Full credit will not be given to answers without work shown. If you use Minitab Express or StatKey include the appropriate output (copy + paste). If you do hand calculations show your work using the Word equation editor. Clearly identify your final answers. Output without explanation will not receive full credit and answers with no output or explanation will not receive full credit. Round all answers to 3 decimal places. If you have any questions, post them to the course discussion board.

1. Market researchers working for a large fitness company want to know if male and female patrons differ in terms of how many hours per week they spend at the gym. At their gym, customers have ID cards that they use to sign in and out of the gym, so the researchers are able to obtain data concerning how long each customer was in the gym. They obtain data from a random sample of **50** male patrons and **50** female patrons. [30 points]

A. What is the independent variable? Gender is the independent variable

B. What is the dependent variable? Hours spent in the gym

C. Is the dependent variable categorical or quantitative? Explain why.

Quantitative because you can compute mathematical calculations using the findings.

D. Are the two groups (males and females) independent or paired? Explain why.

Independent. They are classified but each is handled and observed differently.

E. Write the null and alternative hypotheses that the researchers should be testing. Use the Equation Editor in Word and the correct notation.

Null Hypotheses: $h_o = \mu_m - \mu_f = 0$

Alternative Hypotheses: $h_a = \mu_m - \mu_f \neq 0$

F. What type of graph could the researchers make to display their data? Explain why.

Dotplot or histogram.. because one variable is categorical and the other is quantitative..

G. What method of constructing a sampling distribution should the researchers use (e.g., randomization, t distribution, z distribution)? Explain why.

T distribution. N is larger than 30 for both groups.

2. At the beginning of the Summer 2017 semester a random sample of Penn State students enrolled in online sections of STAT 200 were surveyed and asked for their employment status and primary campus. These data are presented in the contingency table below. [35 points]

	Primary Campus	
Employment Status	World Campus	Physical Campus
Full Time	27	12
Less Than Full Time	11	34

- A. What proportion of World Campus students in this sample were working full time?

$$27/38 = 0.711$$

- B. What proportion of students in this sample who typically attend a physical campus were working full time?

$$12/46 = 0.261$$

- C. Use the five-step hypothesis testing procedure to determine if the proportion of World Campus students who were working full time is different from the proportion of students from physical campuses who were working full time. If assumptions are met, use the normal approximation method. Do not do any calculations by hand. Use Minitab Express and remember to include all relevant output.

Step 1: Check assumptions and write hypotheses

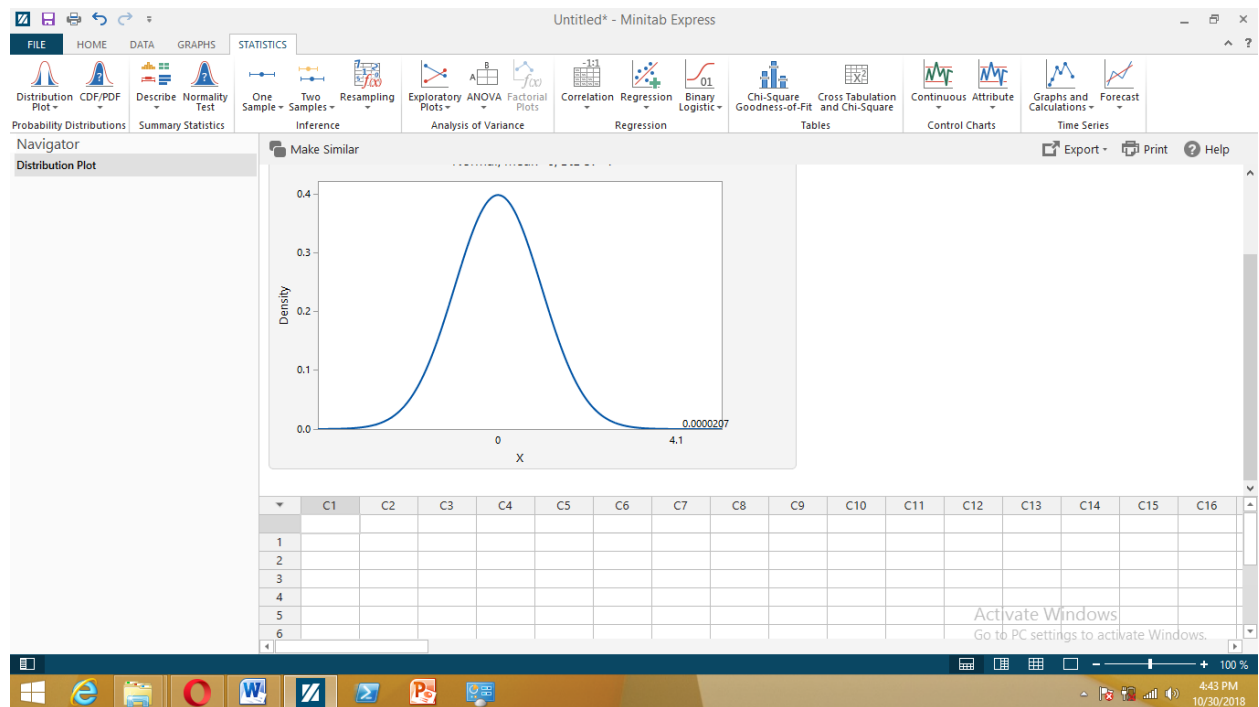
$$p_f - p_p = 0$$

$$p_f - p_p \neq 0$$

Step 2: Calculate the test statistic

4.1

Step 3: Determine the p-value



P value smaller than 0.001

Step 4: Decide between the null and alternative hypotheses

Reject null hypothesis

Step 5: State a real world conclusion

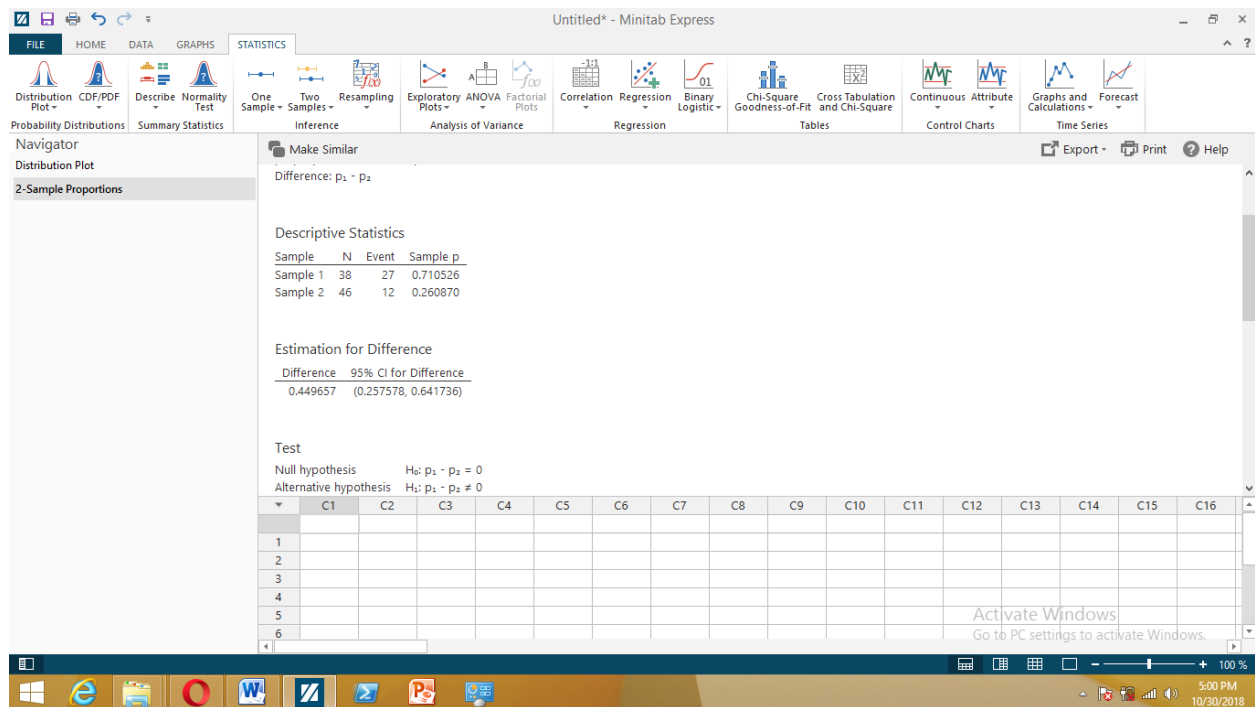
There is evidence that in the population, the proportion of World Campus students who were working full time is different from the proportion of students from physical campuses who were working full time.

D. Are your results in part C practically significant? Explain why or why not.

Not enough information to determine how big the difference is.

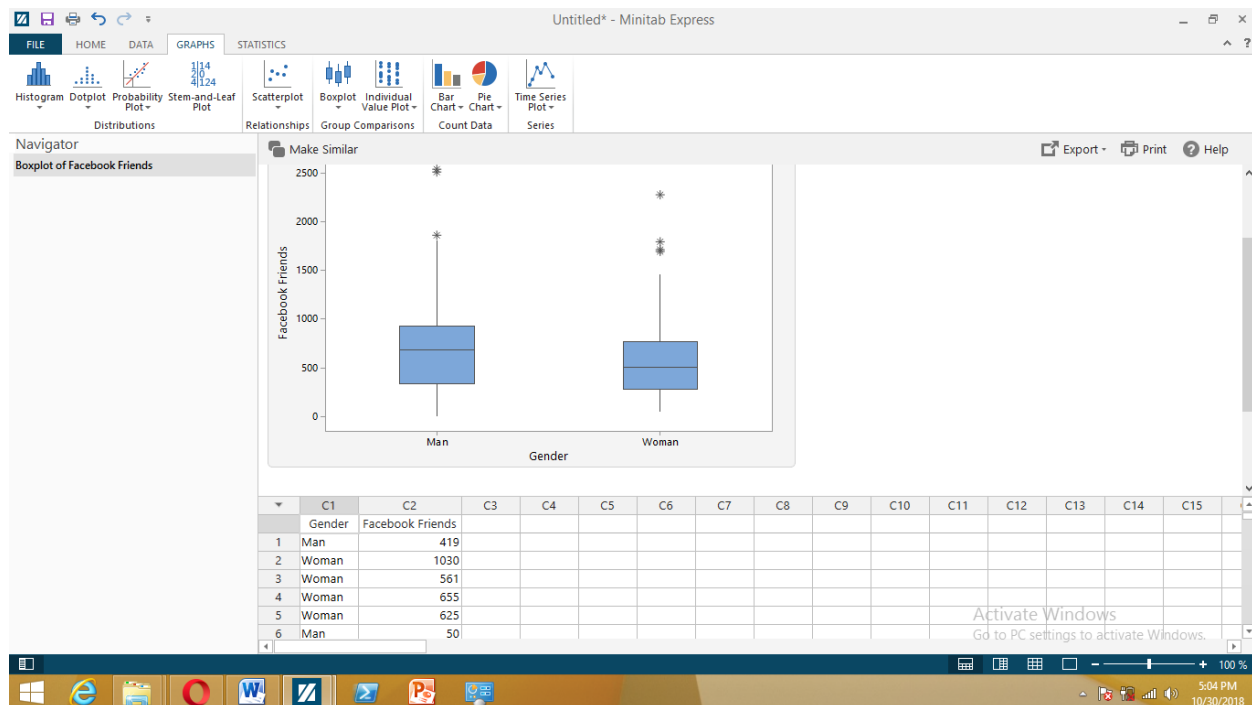
E. Construct a 95% confidence interval for the difference in the proportion of World Campus students and students who typically attend a physical campus who are working full-time. Show how you checked assumptions. If assumptions are met, use the normal approximation method. Do not do any calculations by hand. Use **Minitab Express** and remember to copy+paste all relevant output.

Assumptions check. np is greater than 10 and $n(1-p)$ is also greater than 10 for both groups.



3. Researchers are studying gender differences in Facebook users. They want to know if American men and women differ in terms of how many friends they have on average. Data were collected from a random sample of 130 Facebook users from the United States. Data concerning their genders and number of Facebook friends are in the file **FacebookGender.MTW**. [35 points]

A. Construct a graph to compare the number of Facebook friends that men and women in this sample have.



B. Use the five-step hypothesis testing procedure to determine if there is evidence that in the population of all American Facebook users men and women differ in terms of the number of Facebook friends they have on average. If assumptions are met, use a t distribution to approximate the sampling distribution. Do not do any calculations by hand. Use Minitab Express and remember to include all relevant output.

Step 1: Check assumptions and write hypotheses

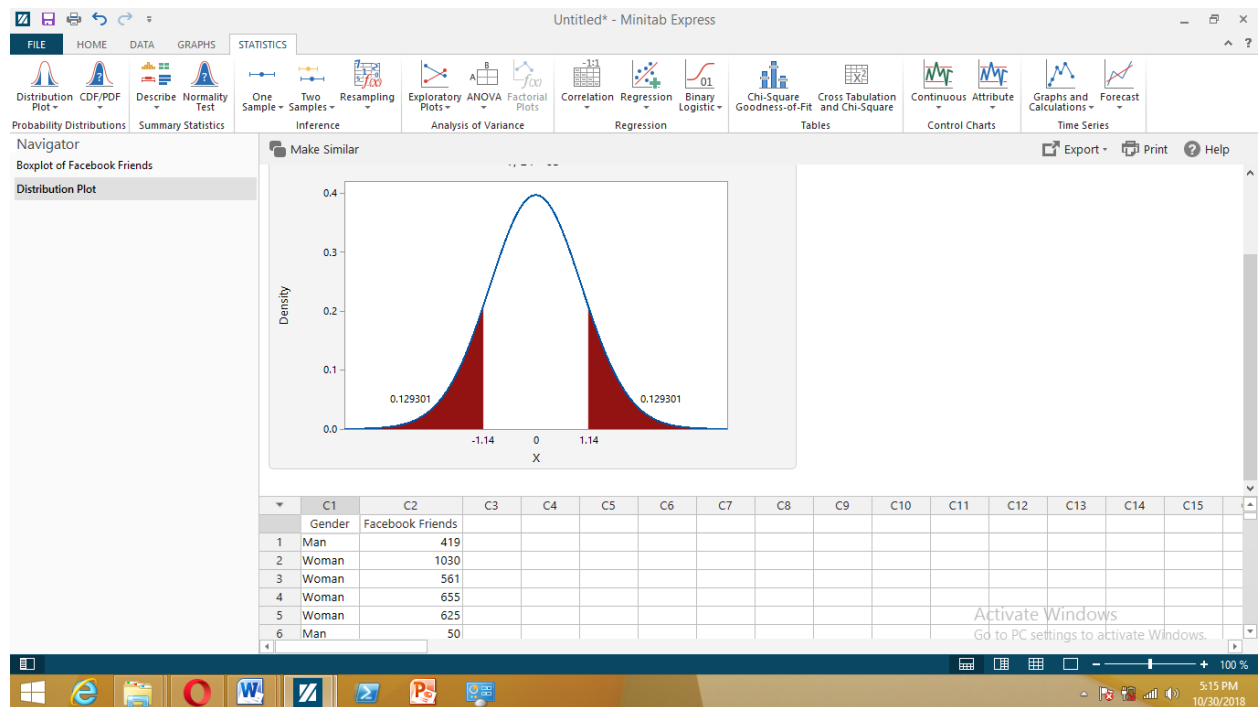
$$h_0 = \mu_m - \mu_f = 0$$

$$h_a = \mu_m - \mu_f \neq 0$$

Step 2: Calculate the test statistic

Test statistic is 1.14

Step 3: Determine the p-value



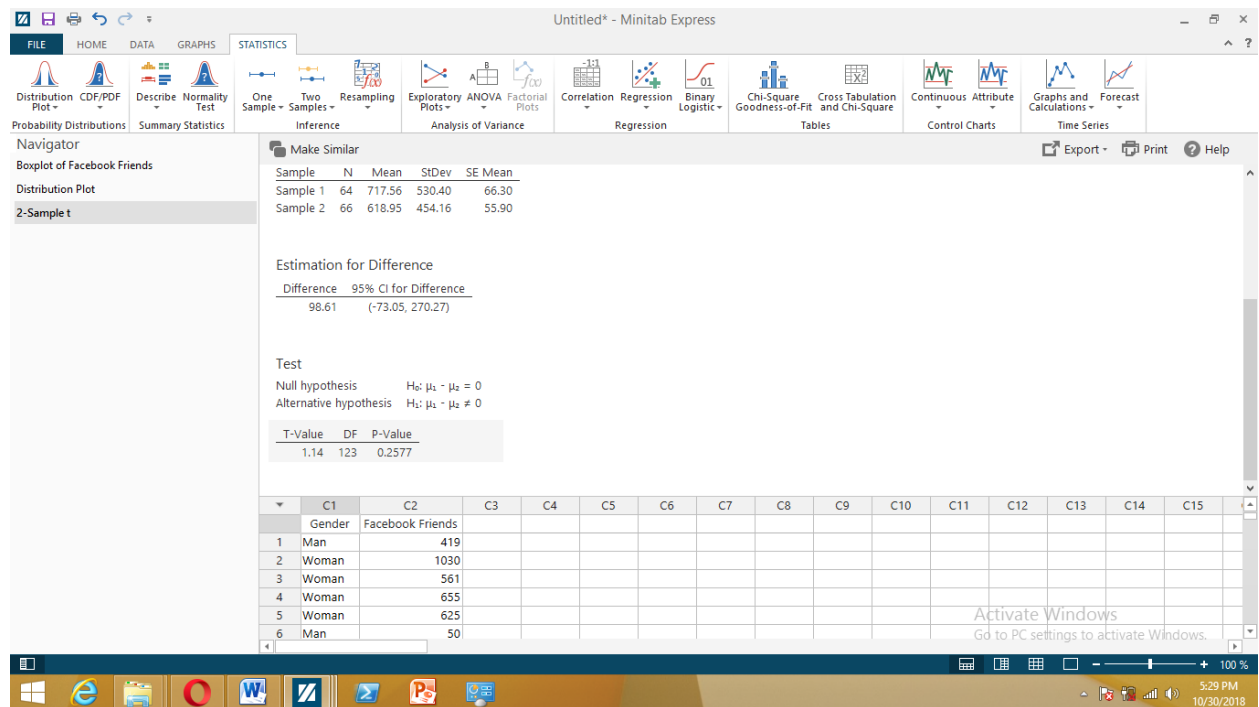
Step 4: Decide between the null and alternative hypotheses

Fail to reject null hypothesis

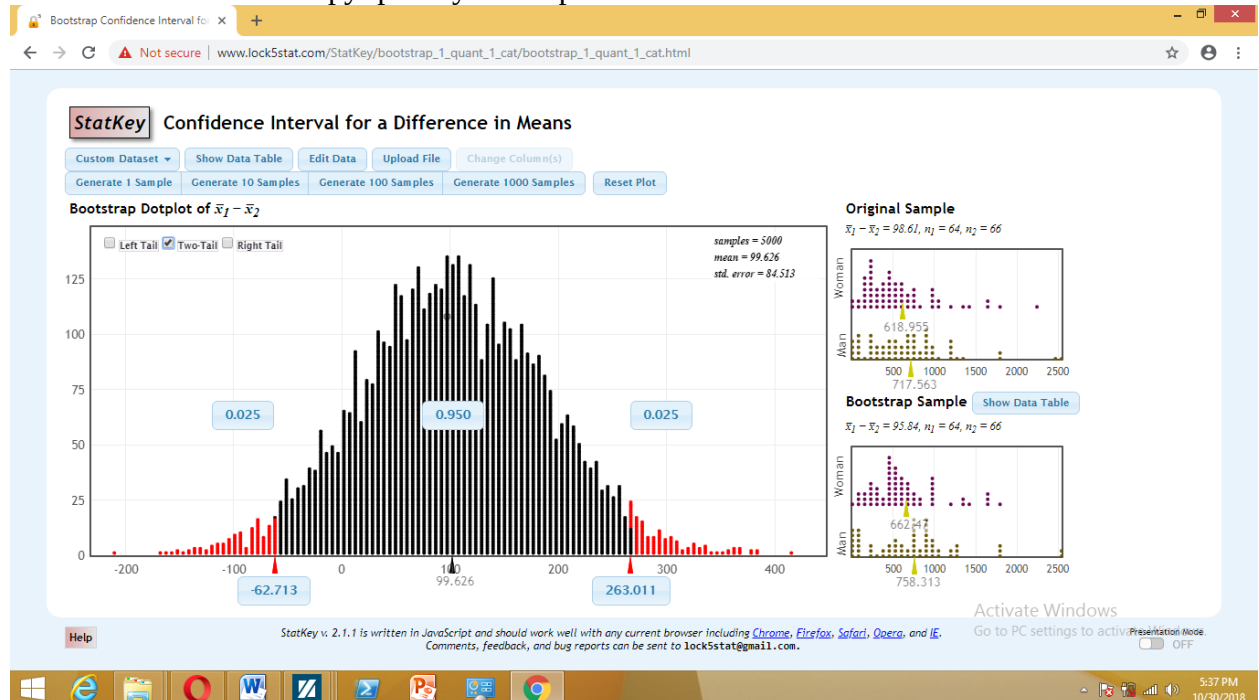
Step 5: State a real-world conclusion

There isn't evidence that men and women differ in number of friends that they have on facebook.

C. Construct a 95% confidence interval to estimate the mean difference in the number of Facebook friends that men and women have in the population of all American Facebook users. Do not do any calculations by hand. Use Minitab Express and remember to copy+paste all relevant output.



D. In **StatKey**, use bootstrapping methods to construct a 95% confidence interval for the differences in means. Copy+paste your output here.



E. Why are your results in parts C and D slightly different?

Because we resampled 5000 times using bootstrapping and replacement.