



Major Assignment

STAT 5002

Student Name: _____

Student Number: _____

Parul Jaiswal, a Humber College graduate who always had only perfect marks in statistics, was hired by the Etobicoke Commercial Bank (ECB). Parul is assigned to conduct statistical analysis of data and provide recommendations to the management. As Parul is on probation, please help her to complete the following tasks. **All solutions should be supported by Excel output.** For your convenience the data are given in Major Assignment Data file. You can also find useful information on the Blackboard in Excel Instructions folder in the files Confidence Intervals and Hypothesis Testing with Excel parts I, II, and III.

1. The ECB managers wonder whether omitting the annual credit card fee for customers who charge at least \$5000 in a year would increase the amount charged on ECB credit cards. On September 1, 2018 the bank made this offer to 75 randomly selected credit card holders. Parul is asked to compare how much these customers charged this year (from September 1, 2018 to August 31, 2019) with the amount they had charged last year (from September 1, 2017 to August 31, 2018). The data are provided. Is there significant evidence that the mean amount charged increases under the no-fee offer? Use Data Analysis *t*-Test: Paired Two Sample for Means. Provide Excel output.

a. Write the null and alternative hypotheses for this test using $d = \text{Last Year Amount} - \text{This Year Amount}$:

H_0 :

H_1 :

b. What is the value of the *t*-test statistic?

c. What is the associated *P*-value?

d. State the conclusion using $\alpha = 0.05$. Do it using both P -value and critical value.

e. Can we draw the same conclusion at $\alpha = 0.01$? Why?

2. Long queues plague every industry, banks included. They can inspire customer's irritation — even disgust. As the number of ECB clients is rapidly increasing, the waiting time problem is becoming more and more serious. This year, ECB is trying a new queue management system at a number of bank branches. However, some managers are sceptical and do not believe that the new approach can have any influence on the waiting times.

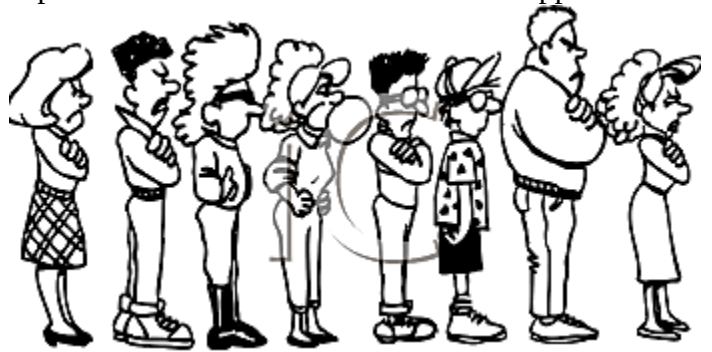
Parul is asked to take random samples of waiting times at two EBC branches (one of them adopted the new method) and check whether the new system has any effect on the average waiting times. The data provided.

Please calculate sample variances for both files and decide what function is most

appropriate: Data Analysis t -Test: Two-

Sample Assuming Equal Variances or Data Analysis t -Test: Two-Sample Assuming Unequal

Variances. It is known that the waiting times are approximately normally distributed. Please help Parul Jaiswal to conduct the test. Provide Excel output.



a. Write the null and alternative hypotheses for this test:

$H_0 :$

$H_1 :$

- b. How do you think, should Parul use t -test for equal or unequal variances? What is the value of the t -test statistic?
- c. What is the associated P -value?
- d. State the conclusion using $\alpha = 0.02$. Do it using both P -value and critical value.

3. ECB provides insurance services to its clients. All insurance companies, large and small, take risks. When they insure a property or a life, they must price the policy in such a way that their expected profit enables them to survive. They can base their projections on actuarial tables, but reality of the



insurance business often demands that they discount policies to a variety of customers and situations. Managing this risk is made even more difficult by the fact, that until the policy expires, the company won't know if it has made a profit, no matter what premium it charges.

One of Parul's managers wants to see how well her insurance sales representatives are doing, so she asked Parul to select a random sample of 60 matured policies that has been sold by the sales reps and compute the (net) profit (premium charged minus paid claims) for each policy. The manager wants Parul to construct a confidence interval for the mean profit of the policies sold by the sales reps. Please help Parul. Use the data provided in the data file. Use t -distribution as all necessary conditions are satisfied. Provide Excel output.

- a. The 90% confidence interval for the average profit is:

- b. The 95% confidence interval for the average profit is:
- c. The 99% confidence interval for the average profit is:
- d. Please construct a histogram with bin values -500, 0, 500, 1000, 1500, 2000, 2500, 3000, and 3500 to make sure that the data come from an approximately normal distribution. Is the histogram roughly symmetric and bell-shaped?
4. ECB insurance department has to track life expectancy information to assist in determining the cost of life insurance policies. Life expectancy is a statistical measure of average time a person is expected to live, based on a number of demographic factors. Mathematically, life expectancy is the mean number of years of life remaining at a given age, assuming age-specific mortality rates remain at their most recently measured levels.

Last year the average life expectancy of all ECB policyholders in Ontario at age 65 was 22.3 years (meaning that a person reaching 65 last year was expected to live, on average, until 87.3). ECB insurance department wants to determine if their clients now have a longer average life expectancy, so they randomly sample some of their recently paid policies. ECB will only change their premium structure if there is evidence that people who buy their policies are living longer than before. The sample is given in the Major Assignment Data file. Help Parul to answer questions a.- h. Answers to questions d.- h. should be supported by Excel output.

- a. Write the null and alternative hypotheses for this test:

$H_0 :$

$H_1 :$

- b.** In this context, describe a Type I error possible. How might such an error impact ECB decision regarding the premium structure?
- c.** In this context, describe a Type II error possible. How might such an error impact EBC decision regarding the premium structure?
- d.** Construct a histogram for the given sample with bin values: 16, 18, 20, 22, 24, 26, 28, and 30. Does it look that the data come from a distribution that is unimodal and symmetric (approximately normal) and therefore t -model is applicable as a data analysis tool?
- e.** What is the value of the t -test statistic?
- f.** What is the associated P -value?
- g.** State the conclusion using $\alpha = 0.05$. Do it using both P -value and critical value.

h. Please reconsider the previous question at $\alpha = 0.01$. Is your conclusion different? Why?

5. It is known that in Canada females live longer than males, on average. As a matter of fact, females live longer everywhere, but the difference varies: say, about 1 year in India, 5 years in the UK, 7



help Parul to test the hypothesis. Provide Excel output.

years in the USA, 8 years in France, and 12 years in Russia (actually way more than 12 according to some independent Russian sources). ECB insurance is trying to estimate the difference for various social groups in different parts of Canada. Two random samples were taken in Southern Ontario, providing life expectancy data at 65 for males and females who worked in education (see the data file). Parul thinks that for this category of policy holders the average difference is about 4 years. She decided to use t -test for two means as both distributions are approximately normal. Please

- a. Write the null and alternative hypotheses for this test:

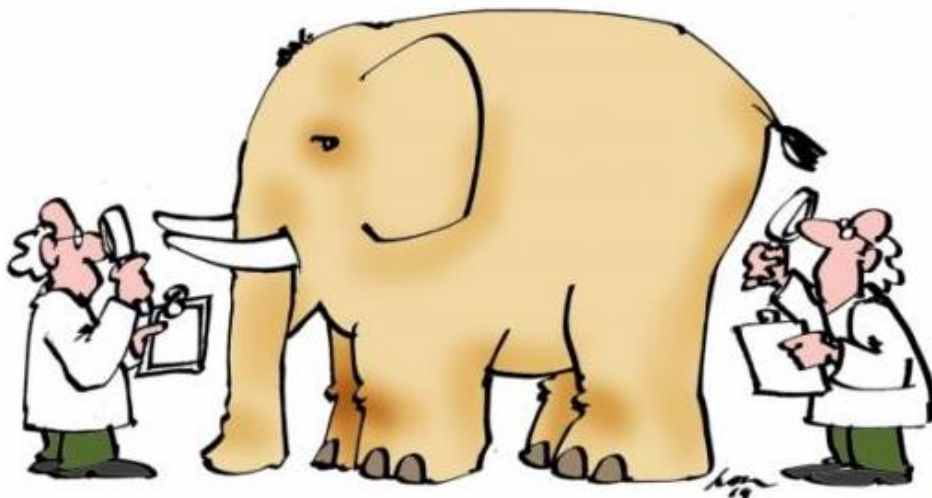
$H_0 :$

$H_1 :$

- b. How do you think, should Parul use t -test for equal or unequal variances? What is the value of the t -test statistic?

c. What is the associated P -value?

d. State the conclusion using $\alpha = 0.05$. Do it using both P -value and critical value.



"Statistics: The only science that enables different experts using the same figures to draw different conclusions."

Evan Esar