

The following assignment is to be done in Microsoft Excel.

Download the **Excel Lab 4 Data** Excel file in D2L/Content/ Excel Lab 4. Open it.

1. Problem 1: Point Estimate, Confidence interval and Test of significance for proportion in Excel.

Make sure you are on **Problem 1** sheet.

The CDC reports vital statistics for all births in the US each year. We use a random sample of data from 2014 for two variables in this problem.

The **WIC** variable indicates whether the WIC food stamp program was used or not (N for No, Y for Yes) by the mother.

Note that these data are **categorical (qualitative)**. Further, there is not much of meaningful information by sorting those and hence these data are **nominal**.

The **Race** variable indicates the race of the mother.

Note that this is a label to given each person at birth and each person has one of the race labels. Therefore, these data are **categorical (qualitative)**. Further, there is not much of meaningful information by sorting those and hence these data are **nominal**.

The two-way (or pivot) table summarizes the number of mothers who used and did not use the WIC food stamp program by their racial group. Note that the total number of mothers belong to each racial group given on the last column of the table are not equal.

Different parts of this problem estimate or test about true proportion of mothers who used WIC within each racial group.

Please read each part carefully and provide answers on the Excel sheet.

For Parts c and d, you should change the numbers on **GREEN** highlighted cells and then you will see the values on the bellow cells change accordingly.

Please DO NOT change anything on those non-highlighted cells. There are formula written on those cells.

At the end, SAVE the Excel file with your answers.

2. Problem 2: Point Estimate, Confidence interval and hypothesis testing for mean in Excel.

The second sheet (**Problem 2**) in the Excel file has the data for the birth weight of babies based on a random sample of data obtained from CDC in 2014.

The **Weight** variable gives the weight (in grams) of baby at birth.

Note that the average or minimum weight of baby makes sense and therefore, these data are **numerical (quantitative)**. Further, these data are **ratio** because there is a true zero and division as well as difference of two ages makes sense. In addition, these are, in general, measured values and hence **continuous**.

Problem 2 Part (a) (ii)

First, we perform a hypothesis test about true (or population) mean birth weight of babies in US in 2014.

Note: if you are using **XLSTAT** in Excel, then please watch the video at the following link to perform a hypothesis test: <https://www.youtube.com/watch?v=G7CtNoX7gO8>

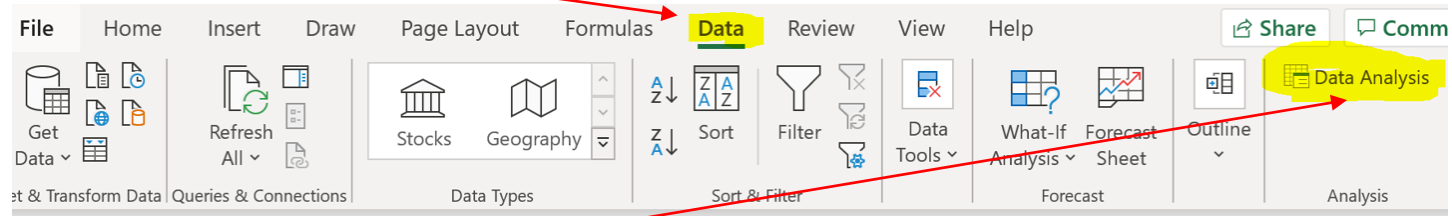
Others, please follow the instruction below.

There is no straight forward method in Excel to do significance test about one population mean. Therefore, we will need to trick it.

- a) Type the word **Dummy** in cell B1 of the Excel sheet (**Problem 2** sheet)
- b) Then enter zeros in column B corresponding to Weight data values starting from cell B2. (The number of zeros and number of weight data values must be same)

	A	B
1	Weight	Dummy
2	3572	0
3	1548	0
4	2787	0
5	3374	0
6	2955	0
7	2211	0
8	3600	0

- c) Next, click on **DATA** tab on the top.

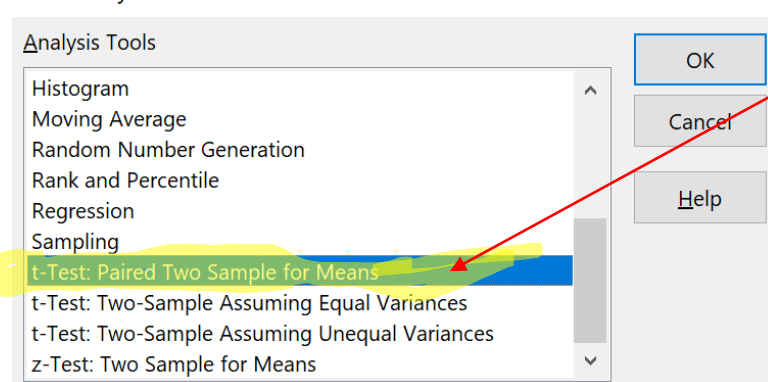


- d) If the **Data Analysis** button is not there, then please go to the course in **D2L** and click **Content** and **Excel Help**.

If you have a **Mac** computer, then please follow the instruction in “**Video about adding Data Analysis tool on Mac**” and “**I can't find the Analysis ToolPak - Excel for Mac**” file.

If you have a **windows** operating system, please follow the instruction in “**Installing the Data Analysis ToolPak to perform complex data analysis - Excel**” file.

- e) Then click **DATA** tab and then **Data Analysis**. Next select **t-Test: Paired Two Sample for Means**



- f) Next, click on the white box next to **Variable 1 Range** and then select or highlight the **Weight** data given including the **Weight** label.

The screenshot shows an Excel spreadsheet with columns A (Weight) and B (Dummy). The 't-Test: Paired Two Sample for Means' dialog box is open. The 'Variable 1 Range' is set to '\$A\$1:\$A\$892', 'Variable 2 Range' is set to '\$B\$1:\$B\$892', 'Hypothesized Mean Difference' is set to '3250', 'Alpha' is set to '0.05', 'Labels' is checked, and 'Output Range' is set to '\$K\$1'. Red arrows point from the instructions to these specific fields.

- g) Then click on the white box next to **Variable 2 Range** and select zeros (including the **Dummy** name) in column B.
- h) Enter claimed value that is 3250 from the null hypothesis (H_0) into the box next to **Hypothesized Mean Difference**
- i) Then enter the value of **significance level** (α) which is 0.05 in the box next to **Alpha**
- j) Check the box next to **Labels** because you selected label when selecting data.
- k) Click on the radio button next to **Output Range:** and then click on the white box next to it.
- l) Now, click on cell K1 on the Excel sheet. (This cell must be empty so that output appears from that cell to down and right). Click OK.

You will see a table of output starting from cell K1 on the Excel sheet. If you did this correct, then you should see 891 for **Observations** and 3250 for **Hypothesized Mean Difference** in the output.

Part of the output is shown below. Your output must have numbers under **Weight**. The last column of the following table shows how to read the output.

t-Test: Paired Two Sample for Means

	Weight	Dummy	
Mean			The sample mean (\bar{x}) for each variable
Variance			The sample variance (s^2) for each variable
Observations			The sample size (n) for each variable
Pearson Correlation			
Hypothesized Mean Difference			
df			The degrees of freedom of t distribution (n-1)
t Stat			The t statistic
P(T<=t) one-tail			The p value for a one-sided test
t Critical one-tail			
P(T<=t) two-tail			The p value for a two-sided test
t Critical two-tail			

According to the statement you are asked to test in part (a) of Problem 2, decide if it is a one-tail (one-sided) or two-tail (two-sided) test.

Then select the correct p value and answer the rest of Part (a) questions on the Excel sheet.

Problem 2 Part (b) (ii)

- m) Next, we need to obtain the **margin of error** to compute 95% confidence interval for true (or population) mean birth weight of babies in US in 2014.

Since sample mean, sample variance, and sample size of weight data are given in the previous output table, we can use the **=CONFIDENCE.T()** function in Excel to compute the margin of error.

- n) To compute **margin of error** for 95% confidence interval based on the data, click cell G43 and type the following.

=CONFIDENCE.T(

- o) Then type 0.05 because we compute 95% confidence level. ($\text{Alpha} = 1 - 0.95 = 0.05$)
- p) Then type a comma and provide the value of sample standard deviation of weight data. **Note: you must give standard deviation and it is NOT variance.**
- q) Next, type a comma and provide the value of sample size of weight data.
- r) Hit enter key and then you will see the value for **Margin of Error (ME)** for 95% confidence interval in cell G43.
- s) Next, you can answer the Parts (iii) and (iv) of **Problem 2** based on your answers for point estimate and margin of error of this problem. Remember,

lower limit = point estimate – margin of error

upper limit = point estimate + margin of error

(show your computation of these two limits on the Excel sheet)

SAVE the Excel file as it is with all the answers.

Go to D2L/ Activities/ Assignments/ Excel Lab 4.

Upload Excel file using Add File button.

Then hit **Submit** button.