Lab report 4

Fall 2021

CSE422L Data Analytics Lab

Submitted by: **Ayaz Mehmood**

Registration No.:**18PWCSE1652**

Section: **A**

"On my honor, as student of University of Engineering and Technology, I have neither given nor received unauthorized assistance on this academic work."

Student Signature: _____

Submitted to:

Engr. Mian Ibad Ali Shah

Last date of Submission:

12 December 2021

Department of Computer Systems Engineering

University of Engineering and Technology, Peshawar

OBJECTIVE:

The basic Objective of this lab is:

- To know about the z and t distribution
- To know how to find confidence interval for single and two means
- To know about dependent and independent samples and their formulas
- To work with different use cases of inferential statistics

TASKS

Task 1:

Task 1

Mean	Standard error	Margin of error	95% CI		Number of pairs, round up
2015-2016	15-'16	15-'16	15-'16		
2.17	0.39	0.82	1.35	2.98	3
1.58	0.34	0.71	0.88	2.29	2
1.33	0.34	0.70	0.63	2.04	2
2.33	0.41	0.85	1.48	3.19	3
4.79	0.60	1.24	3.55	6.03	6
7.88	0.94	1.96	5.92	9.83	10
16.33	1.26	2.62	13.72	18.95	19
25.58	1.77	3.66	21.92	29.25	29
18.79	1.33	2.75	16.04	21.54	22
14.96	1.02	2.12	12.84	17.08	17
7.54	0.72	1.49	6.05	9.03	9
5.33	0.58	1.21	4.13	6.54	7
3.08	0.58	1.20	1.88	4.29	4
1.21	0.26	0.53	0.68	1.74	2
1.96	0.30	0.62	1.34	2.58	3
0.54	0.19	0.39	0.15	0.94	1
0.00	0.00	0.00	0.00	0.00	0

	2015-2016
n	24
$t_{95\%,df}$	2.07

Interpretation:

I am 95% sure that the mean could lie between these interval. Like for the first data of size 6, the mean could lie between 1.35 and 2.98. By round up the upper mean limit, the mean value shows this much shoes of the given size should be in my order list per month. Here in the first case, the value is 3 so we can state 3 shoes of size 6 required to sell out per month.

Task 2:

Task 2

Number of pairs (In Class)	Difference Class - Exercise	Class Task					Difference	
		Mean	Std Er	ME	95%CI Lower	95%CI Upper	Lower	Upper
4	1	2.92	0.36	0.754264	2.16	3.67093	1.00	1.00
3	1	1.67	0.39	0.814003	0.85	2.48067	0.00	0.00
3	1	1.67	0.43	0.890477	0.78	2.557143	0.00	1.00
5	2	3.17	0.49	1.018504	2.15	4.185171	1.00	1.00
8	2	6.08	0.62	1.29494	4.79	7.378273	1.00	1.00
13	3	10.8	0.79	1.64584	9.10	12.39584	3.00	3.00
23	4	18.8	1.39	2.887356	15.95	21.72069	2.00	3.00
36	7	30.3	1.73	3.589248	26.74	33.92258	5.00	5.00
26	5	22.8	1.11	2.298512	20.45	25.04851	4.00	4.00
21	4	17.6	0.97	2.005774	15.58	19.58911	3.00	3.00
12	3	9.75	0.71	1.478986	8.27	11.22899	2.00	2.00
8	2	5.75	0.68	1.411339	4.34	7.161339	0.00	1.00
6	2	3.83	0.72	1.486769	2.35	5.320102	0.00	1.00
2	0	1.58	0.27	0.555174	1.03	2.138507	0.00	0.00
4	1	2.42	0.35	0.732343	1.68	3.14901	0.00	1.00
1	0	0.67	0.25	0.521112	0.15	1.187779	0.00	0.00
0	0	0	0	0	0.00	0	0.00	0.00

Interpretation:

Here the difference show the confidence interval in home task is less than the class task. As we know marginal error is inversely proportional to size of data. In lab task the size is 12 while in home task it is 24. That's why the interval in class task is greater than home task.

Task 3:

	Germany, GER1												Germany, GER2												Mean		Sample variance		Pooled variance	Margin of error	90% CI	
US	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	GER1	GER2	GER1	GER2				
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4.5	1	1	1	0	1	3	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0.75	0.08	0.75	0.08	0.42	0.45	0.21	1.29
5	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0.17	0.17	0.33	0.33	0.33	0.40	-0.40	0.74
5.5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	1	0	2	0	1	0.08	0.50	0.08	0.64	0.36	0.42	-0.84	1.00
6	0	2	0	0	0	0	0	0	0	0	0	0	0	1	3	1	2	0	0	0	0	0	0	0	0.17	0.58	0.33	0.99	0.66	0.57	-0.99	1.32
6.5	3	3	1	2	1	0	2	0	2	1	3	4	2	0	2	1	1	2	0	1	2	1	3	0	1.83	1.25	1.61	0.93	1.27	0.79	-0.21	3.87
7	0	3	3	4	1	0	1	0	2	0	0	1	0	0	0	4	1	3	1	1	1	3	1	4	1.25	1.58	2.02	2.27	2.14	1.03	-1.36	3.86
7.5	1	2	4	1	2	6	4	3	5	8	2	1	2	1	1	3	2	7	9	8	14	8	6	3	3.25	5.33	4.93	16.06	10.50	2.27	-4.35	10.85
8	6	10	3	9	1	3	6	8	3	12	3	9	13	6	5	13	5	3	11	6	6	9	8	3	6.08	7.33	12.27	12.24	12.25	2.45	-3.70	15.87
8.5	10	10	10	7	14	4	7	7	4	8	7	9	8	5	10	4	5	5	9	7	3	7	9	8	8.08	6.67	7.72	4.97	6.34	1.77	-0.35	16.52
9	1	3	8	6	3	1	4	4	0	2	4	2	5	2	2	9	3	1	1	7	2	1	4	2	3.17	3.25	5.06	6.57	5.81	1.69	-1.77	8.11
9.5	4	1	2	1	2	2	2	4	5	2	3	2	0	1	1	0	1	2	2	1	7	2	4	2	2.50	1.92	1.55	3.72	2.63	1.14	-0.55	5.55
10	0	1	1	1	1	1	3	1	0	0	0	1	0	1	1	0	0	0	2	3	0	2	0	0	0.83	0.75	0.70	1.11	0.91	0.67	-0.58	2.25
11	1	0	0	0	2	2	4	1	0	3	1	1	0	2	0	0	0	1	0	0	0	0	2	1	1.25	0.50	1.66	0.64	1.15	0.75	0.00	2.50
12	0	0	0	2	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	1	0	0	0	0	0.17	0.50	0.33	2.09	1.21	0.77	-1.11	1.44
Total	27	36	33	33	28	22	35	29	21	36	25	30	30	19	30	37	20	24	35	38	35	36	37	24								

	GER1	GER2
n	12	12
$t_{90\%,df}$	1.72	

Task 3: Estimate the 90% confidence interval for the same task as the lesson. ***What changes can you see?***

Lower confidence levels result in smaller intervals: 90% CI's are smaller than 95% CI's. The tradeoff here is that smaller intervals are less likely to contain the parameter we're after: 90% versus 95%