

Indian Institute of Technology Madras
Department of Data Science and Artificial Intelligence
DA5000: Mathematical Foundations of Data Science
Tutorial VII

Problem

1. Suppose the blood pressure in a population is normally distributed with a mean of $\mu = 120$ mmHg and a standard deviation of $\sigma = 10$ mmHg. A researcher wants to determine if a new diet affects blood pressure levels. She recruits 25 participants to follow the diet for one month and records their blood pressure levels at the end of the month:

Participant	Blood Pressure (mmHg)
1	118
2	122
3	115
4	117
5	121
6	119
7	116
8	123
9	114
10	120
11	119
12	118
13	117
14	122
15	115
16	116
17	121
18	120
19	118
20	117
21	123
22	114
23	119
24	120
25	118

Perform a one-sample z-test at a significance level of $\alpha = 0.05$ to decide whether to reject or fail to reject the null hypothesis based on the p-value and significance level.

2. Two random samples were drawn from two normal populations, and their values are:

A	65	66	73	80	82	84	88	90	92		
B	64	66	74	78	82	85	87	92	93	95	97

Test whether the two populations have the same variance at the 5% significance level.

3. A county environmental agency suspects that the fish in a particular polluted lake have elevated mercury level. To confirm that suspicion, six striped bass in that lake were caught and their tissues were tested for mercury. For the purpose of comparison, five striped bass in an unpolluted lake were also caught and tested. The fish tissue mercury levels in mg/kg are given below.

Sample 1 (from polluted lake)	Sample 2 (from unpolluted lake)
0.610	0.429
0.703	0.391
0.582	0.570
0.591	0.573
0.621	0.497
0.551	

- Construct the 95% confidence interval for the difference in the population means based on these data.
 - Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that fish in the polluted lake have elevated levels of mercury in their tissue.
4. We intend to compare several methods for predicting the shear strength for steel items. Data for two of these methods, the M1 and M2, when applied to nine specific items, are shown in Table 1. Determine if there is a significant difference in shear strength predictions between methods M1 and M2 at a significance level of $\alpha = 0.05$. Address the following while arriving at the conclusion:
- Formulate the Null and Alternate Hypothesis
 - Calculate the test statistic and compare with the critical value
 - Also compute the p value corresponding to the test statistic

Item	Method 1	Method 2
1	1.186	1.061
2	1.151	0.992
3	1.322	1.063
4	1.339	1.062
5	1.200	1.065
6	1.402	1.178
7	1.365	1.037
8	1.537	1.086
9	1.559	1.052

Table 1: Comparison of predictions of Method 1 and Method 2 for 9 items