

ARTIFICIAL INTELLIGENCE

Lab6. Two populations comparison

CAUTION! Most of the data available in *data_twoPop.csv* file

1. The researcher verified the percentage content of cellulose in a particular type of wood which appears in two different regions of Poland. In the first region a sample of 8 branches of wood was randomly chosen while for the second region a sample of 21 wood branches was measured. All the data are available in relevant file. Assuming normality of the distribution of cellulose content and 2% significance level:
 - (a) verify the assumption about homogeneity of variances; $F=0.479, p\text{-val}=0.32$, no reason to reject H_0
 - (b) does the data confirm that the mean content of cellulose in the first region differs significantly from the mean content of cellulose in the second region? $t=-1.54, p\text{-val}=0.135$, no reason to reject H_0
 - (c) estimate with 98% of confidence the true difference of means and compare the result with (b).
(-13.52; 3.15)
2. An engineer suspects that the time of house building by traditional method is longer than the time with the use of a new technology. Therefore, respectively 10 and 11 houses built with traditional or new technology were randomly chosen and the time of their construction was measured. All the results are available in relevant file. Assuming normality of the time of building houses verify the engineer's claim. Assume 10% of significance.
homogeneity of variances: $F=0.54, p\text{-val}=0.36$, no reason to reject H_0
hypothesis about means: $t = -0.29, p\text{-val}=0.61$, no reason to reject H_0
3. A bank wants to check which method of getting money (from public or private sources) gives bigger funds. Therefore, a sample of 11 companies with loans from public sources and 16 companies with loans from private sources were chosen. All the results are available in relevant file. Assuming that the value of a loan is normally distributed, is it reasonable to expect that the loans from public sources are smaller (in average) than the loans which come from private sources? Assume significance level 0.1.
homogeneity of variances: $F=0.336, p\text{-val}=0.087$, reject H_0
hypothesis about means: $t = -2.1, p\text{-val}=0.023$, reject H_0
4. To compare stability of two sportsmen in a given season, 12 results of the first player and 9 results of second player were randomly measured (in meters). All the results are available in respective file. Assuming 5% significance level verify whether the first player has more stable results.
 $F = 0.5978, p\text{-val}=0.21$, no reason to reject H_0
5. The effect of two different medicines for headache was tested to verify whether the first drug works longer. Thus, the sample of 25 patients was chosen, and the first medicine was served to 10 of them and the second one to the remaining patients. Then the duration of drugs action was measured (the results are available in respective file). Is it reasonable to expect that the mean duration of the first medicine is significantly longer than of the second one? Assume 10% of significance and normality of duration time.
homogeneity of variances: $F=0.73, p\text{-val}=0.64$, no reason to reject H_0
hypothesis about means: $t = 1.435, p\text{-val}=0.08$, reject H_0

6. The research conducted in Poland and USA it follows, that among 1200 Poles 78% are satisfied with their job, while among 2000 of Americans 20% were upset with a job.

(a) Compare the proportion of satisfied Poles and Americans using 90% confidence interval. Use the formula from the lecture and compare the result with `prop.test` function in R.

lecture: (-4.46%; 0.46%), R: (-4.53%; 0.53%)

(b) Is there enough evidence, that the percentage of satisfied Poles is smaller than the proportion of satisfied Americans. Assume 10% of significance.

$z=-1.35$, reject H_0 / $p\text{-val}=0.096$, reject H_0

7. One wants to verify whether the frequency of malaria patients depends on a region. In a particular tropical region of Asia 313 of type A malaria patients and 28 of type B malaria patients were noted. For comparison, in a specified tropical region in Africa 145 patients have type A malaria and 56 patients – malaria of type B.

(a) Does the frequency of type A malaria depend on the region? Assume 5% significance level.

$z=6.11$, reject H_0 / $p\text{-val}=0$, reject H_0

(b) Verify hypothesis from (a) using confidence interval.

(12.8%;26.5%) / (12.4%, 26.9%)

8. An engineer wishes to compare two alternative designs for a satellite electronics system. Design A is a maximum-redundancy one involving many inexpensive, low-reliability components in parallel. Design B is a high-quality one involving a minimum number of highly reliable but expensive components. He will select the second design unless the first is found to have a significantly smaller failure rate. Thus, a 1000-mission Monte Carlo simulation was conducted under each design. The results provided 47 failures with design A and 62 with design B. At the 5% significance level, which design should the engineer select?

Design B

9. A medicine doctor suspects that particular type of drugs changes the values of a specific biochemical parameter. Thus, the level of the parameter was measured for 9 patients before and after serving the drug:

Before:	15	4	9	9	10	10	12	17	14
After:	14	4	10	8	10	9	10	15	14

Assuming 5% significance level verify the doctor's suspicion. Assume normality of a level of biochemical parameter.

$t=2$, $p\text{-val}=0.08$, no reason to reject H_0

10. A pH of water in 8 lakes was measured on two different depths (15 cm and 100 cm). The results are presented in the following table:

depth	1	2	3	4	5	6	7	8
15cm	6.55	5.98	5.59	6.17	5.92	6.18	6.43	5.68
100cm	6.78	6.14	5.80	5.91	6.10	6.01	8.18	5.88

(a) Does the pH of water depend on the depth? Assume 10% of significance and normality of pH distribution.

$t=-1.31$, $p\text{-val}=0.23$, no reason to reject H_0

(b) Verify (a) using 90% confidence interval.

(-0.71; 0.13)