ARTIFICIAL INTELLIGENCE

Lab7. *ANOVA*

1. A chemical engineer wishes to assess the effect of different pressure settings on the mean yield of a final product from a given level of raw ingredients. The following sample data (in grams/liter) have been obtained:

Pressure	1	2	3	4	5	6	7	8	9	10
Low	28	26	29	30	28	31	26	32	25	29
Moderate	30	29	30	30	28	32	29	32	28	30
Strong	31	29	33	33	29	33	28	32	27	32
Very strong	29	27	30	31	27	32	27	32	27	30

   Does the pressure have an effect on the yield of final product? Assume normality of the yield of a final product and significance level 0.05. Verify the homogeneity of variance (using e.g. bartlett.test function available in R) before analysis of variance is performed.

2. A petroleum engineer for a large oil company is evaluating secondary recovery techniques for oil wells. The following sample data provide the daily increase (barrels) in pumped crude obtained from sample wells:

Secondary treatment	1	2	3	4	5	6	7	8
Explosive fracture	-5	11	13	-5	25	21	11	33
Water injection	11	14	22	0	35	24	15	30
Steam injection	15	10	25	2	40	28	19	28
Controlled pumping	8	10	18	1	28	27	17	32

   Can the engineer conclude that some secondary treatments might be better than others? Assume significance level 0.05 and normality of the increase in pumped crude. Verify homogeneity of variance before ANOVA is performed.

3. The following table contains the results of measuring a metal object by three different micrometers:

   Micrometer I: 4.5; 4.7; 4.8; 4.7
   Micrometer II: 4.7; 4.8; 4.5; 4.7; 4.4; 4.8
   Micrometer III: 4.8; 4.9; 4.8; 4.9; 4.8

   Can the choice of the micrometer have an effect on the measurement? Assume significance level 0.05 and normality of size of metal object. Verify assumption about homogeneity of variance before ANOVA is performed.

4. An industrial-engineering student helped an entomological research team evaluate various location strategies for gypsy moth scent-lure traps. The following data were obtained (in %):

	Trap location strategy				
observation	scattered	concentrated	host plant	aerial	ground
1	90	99	95	98	87
2	92	97	96	98	93
3	94	98	97	99	90
4	93	98	97	99	91
5	92	99	96	98	89

   The response variable is the estimated percentage of the native male population trapped.

   (a) Can the location strategy have an effect on the number of gypsy moth trapped? Assume significance level 0.05 and normality of proportion of trapped gypsy moths. Verify assumption about homogeneity of variance before ANOVA is performed;

   (b) Identify homogeneous groups using Tukey test; *two groups: (c,h,a), (s,g)*

   (c) Identify homogeneous groups using Fisher test; *one group: (a,c)*

   (d) Identify homogeneous groups using simultaneous confidence intervals (Bonferroni method). *two groups: (c,h,a), (s,g)*

5. The population of sportsmen was warned that smoking cigarettes can block their development. One of the measures of the effect of smoking is the sinus rhythm of a heart. The following table represents the sinus rhythm of 24 sportsmen 5 minutes after physical activity:

observation	Non-smoker	Light-smoker	Medium-smoker	Heavy-smoker
1	69	91	55	66
2	52	72	60	81
3	71	81	78	70
4	58	67	58	77
5	59	95	62	57
6	65	84	66	79

   (a) Assuming significance level 0.05 and normality of sinus rhythm of sportsmen verify whether smoking cigarettes can influence the sinus rhythm of a heart (begin with verification of homogeneity of variance);

   (b) Identify homogeneous groups using Tukey test; *two groups: (N,M,H), (L,H)*

   (c) Identify homogeneous groups using Fisher test; *two groups: (N,M,H), (L,H)*

   (d) Identify homogeneous groups using simultaneous confidence intervals (Bonferroni method).

   two groups: (N,M,H), (L,H)

6. The mining engineer used ore taken from 10 different major sites within the deposit complex. These were selected for their distinctive material composition, and the ores provided represent the wide range of types found at the mine. Each of the sites is referred to as a block. A random sample was then taken from each site of block (in each case three strains of bacteria were tested) – the results are available in *anova_bacteria2.csv* file. Does the type of bacteria strain has an effect on the final copper yield? Assume normality of data and significance level 5%.

7. In *anova_chemical2.csv* the sample results of an investigation to establish operating parameters for processing a chemical are presented. The mean yield of raw material (in grams per liter) is measured. Two treatment variables are considered: T1, which represents the pressure setting (in pounds per square inch), and T2, which represents the temperature setting (in degrees Fahrenheit). Does the data confirm, that the mean yield of raw material is the same under all pressure settings? Does the data confirm, that the mean yield of raw material is the same under all temperature settings? Assume normality of variables and significance level 5%.

8. R software contains many various data bases. Read the dataset *ToothGrowth*, in which the response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC). Does the data confirm, that the mean length of odontoblasts depends on the dose of vitamin? Does it depend on the delivery method? Does the data confirm, that the length of odontoblasts depend on the common effect of the dose of vitamin and delivery method? Assume normality of variables and significance level 5%.