æ

MULTIVARIATE STATISTICS HOMEWORK II.

1. Let $X_1, \dots, X_n$ be i.i.d., normally distributed random variables. Prove that the following decomposition is true:

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2,$$

where

$$Q_1 = \sum_{i=1}^{n-1} (X_i - \bar{X})^2 \quad \text{and} \quad Q_2 = (X_n - \bar{X})^2 + n\bar{X}^2.$$

Is the Fisher–Cochran theorem applicable here?

2. By means of the ANOVA notations, show that the two-way ANOVA without interaction satisfies the decomposition

$$Q = Q_1 + Q_2 + Q_3,$$

where

$$Q_1 = p \sum_{i=1}^{k} (\bar{X}_{i.} - \bar{X}_{..})^2$$

$$Q_2 = k \sum_{j=1}^{p} (\bar{X}_{.j} - \bar{X}_{..})^2$$

$$Q_3 = \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2$$

$$Q = \sum_{i=1}^{k} \sum_{j=1}^{p} (X_{ij} - \bar{X}_{..})^2$$

3. 3 students have the following test results (integers in the [0,5] interval):

1.student :	5, 3, 2, 2
2.student :	5, 0, 1
3.student :	2, 1, 0, 1

Perform one-way ANOVA to decide, whether the average performance of the three students are significanty different (with 90% level of significance).

4. 15 loaves of bread, 5-5-5 made according to a different recipe, are baked in 5 different oven positions at the same time. The densities (measured by a real number between 0 and 1) of the so baked loaves are the following:

	I.	*II.*	*III.*	*IV.*	*V.*
1.	0.95	0.86	0.71	0.72	0.74
2.	0.71	0.85	0.62	0.72	0.64
3.	0.69	0.88	0.51	0.73	0.44

Perform two-way ANOVA (without interaction) to test the following hypotheses:

a. The choice of recipe does not influence significantly the density (with 90% level of significance).

b. The oven position does not influence significantly the density (with 90% level of significance).

5. Trace back the following models to linear regression:

a. $Y = \dfrac{1}{(a+bX)^2}$ b. $\ln Y = a + \dfrac{b}{1+X}$ c. $Y = a_3X^3 + a_2X^2 + a_1X + a_0$

6. The blood pressure (Y) of 13 men is investigated as the function of the age (X_1 in years) and weight (X_2 in kilograms). The sample averages and the empirical covariance matrix of the three variables are the following:

$$\bar{Y} = 130 \qquad \bar{X}_1 = 48 \qquad \bar{X}_2 = 75.$$

$$\hat{\mathbf{C}} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}.$$

Estimate the coefficents of the two-variable linear regression model and find the multiple correlation between the dependent variable Y and independent variables X_1, X_2. Further test the zero-hypothesis

$$H_0: \quad a_1 = a_2 = 0$$

against the alternative that at least one of the coefficients a_1, a_2 differs from zero (with 90% level of significance). Interpret the results!

7. With the notation of the linear model suppose that the rank of the matrix $\mathbf{X}$ is p. Let $\hat{\mathbf{a}}$ be the solution of the Gauss' normal equations. Prove that

$$(\hat{\mathbf{a}} - \mathbf{a})^T \mathbf{X}^T \mathbf{X} (\hat{\mathbf{a}} - \mathbf{a}) / \sigma^2 \sim \chi^2_p!$$

8. 200 people are divided into the following groups according to their eye-colour and hair-colour:

	blond hair	brown hair	black hair
blue eyes	42	28	3
brown eyes	17	89	21

a. Decide whether the eye-colour is independent of the hair-colour (with 99% level of significance).

b. Perform correspondence analysis on the above table. Find the nontrivial factors and the corresponding singular values. Draw conclusions, interpret the results!

9. Let $(X_1, Y_1)^T$ és $(X_2, Y_2)^T$ be independent, identically distributed random vectors from the $\mathcal{N}_2(\mathbf{0}, \mathbf{C})$-distribution, where

$$\mathbf{C} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad (|\rho| < 1).$$

Define the following random variables:

$$X := \begin{cases} 1, & \text{if } X_1 < X_2, \\ -1, & \text{otherwise} \end{cases} \qquad Y := \begin{cases} 1, & \text{if } Y_1 < Y_2, \\ -1, & \text{otherwise.} \end{cases}$$

Prove that $\operatorname{Cov}(X, Y) = (2/\pi) \cdot \sin^{-1} \rho$.

10. Observations on the intelligence (X_1), weight (X_2), and age (X_3) of school children produced the following correlations:

$$corr(X_1, X_2) = 0.6162, \quad corr(X_1, X_3) = 0.8267, \quad corr(X_2, X_3) = 0.7321.$$

Find the partial correlation between X_1 and X_2 after eliminating the effect of X_3. Draw conclusion!