æ

# MULTIVARIATE STATISTICS HOMEWORK II.

1. (*1 point*) Let $X_1, \dots, X_n$ be i.i.d. Gaussians with 0 expectation. Verify the decomposition

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

where

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

Apply the Fisher–Cochran theorem for this decomposition, and find out whether $Q_1$ and $Q_2$ are independent $\chi^2$-distributed or not.

2. (*1.5 point*) With the ANOVA notation prove that in the two-way ANOVA without interaction

$$Q = Q_1 + Q_2 + Q_3$$

holds true, 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. (*1.5 point*) The test results of 3 students (on a 0–5 scale) are the following:

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

Perform one-way ANOVA to answer whether the average performance of the students differs significantly at level $\alpha = 0.1$.

4. (*2 point*) 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 sign. 0.1).

b. The oven position does not influence significantly the density (with sign. 0.1).

5. (*1 point*) 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. (*2 point*) The dependence of the blood-pressure ($Y$) of 13 men on their age ($X_1$ years) and body-weight ($X_2$ kg) is investigated. The sample means and the empirical cocariance matrix of the three variables are as follows:

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

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

In the two-variate regression model, estimate the model parameters and calculate the multiple correlation coefficient between $Y$ and $(X_1, X_2)$!

Further, investigate the hypothesis

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

with significance 0.1 and explain the result.

7. (*1 point*) With the notation of the linear model, assume that the rank of the $n \times p$ matrix $\mathbf{X}$ is $p$ ($p < n$). Let $\hat{\mathbf{a}}$ denote 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. (*2 point*) For the eye- and hair-color of 200 persons we get the following cross-tabulation.

| | blond hair | brown hair | black hair |
|---|---|---|---|
| blue eye | 42 | 28 | 3 |
| brown eye | 17 | 89 | 21 |

a. Are hair- and eye-colors independent (with sign. 0.01)?

b. Perform correspondence analysis on the above table. Find the non-trivial factors together with the corresponding singular values. Make a picture illustrating the closeness of the hair- and eye-color categories! Draw conclusions, interpret the results!

9. (*2 point*) At our university, the male students have average hight 172.6 cm and weight 72.3 kg; whereas, the female students 165.3 cm and 65.4 kg. Further the

male:female proportion is $3:1$. Assume that the body-hight and weight follow 2-dimensional normal distribution with covariance matrix

$$\mathbf{C} = \begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix}.$$

We forgot the name and gender of a student, we only know the hight 170 cm and weight 68 kg. Using the Fisher's discriminating function, to which group (male or female) this student could be classified?

10. (*1 point*) The Stirling number (of second type) $\left\{ n \atop k \right\}$ denotes the number of possible partitions of an $n$-element set into $k$ non-empty disjoint subsets ($k \leq n$). Prove the following:

$$\left\{ n \atop k \right\} = \left\{ n-1 \atop k-1 \right\} + k \left\{ n-1 \atop k \right\}.$$