University of California, Los Angeles
Department of Statistics

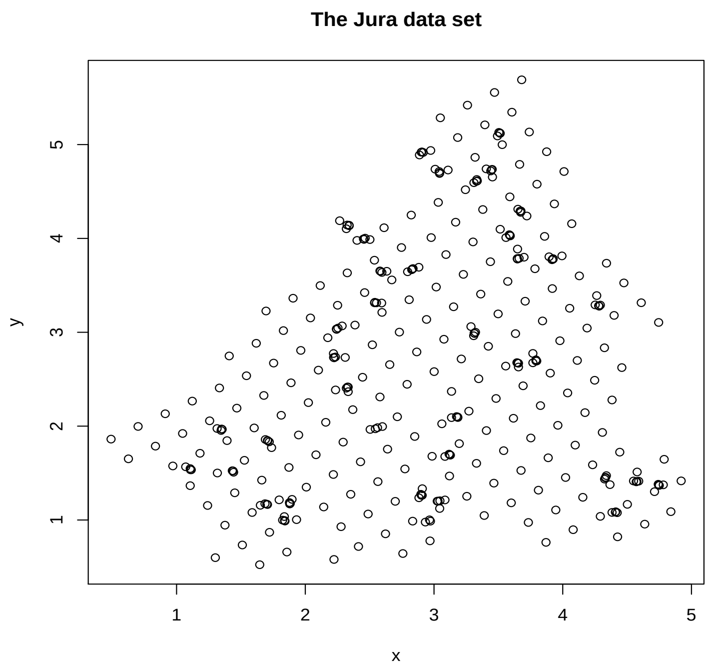
Statistics 100C **Instructor: Nicolas Christou**

Homework 17

Access the following data:

```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/jura.txt",
header=TRUE)
```

These Jura data were collected by the Swiss Federal Institute of Technology at Lausanne. See Goovaerts, P. 1997, "Geostatistics for Natural Resources Evaluation", Oxford University Press, New-York, 483 p. for more details. Data were recorded at 359 locations scattered in space (see figure below).



Concentrations of seven heavy metals (cadmium, cobalt, chromium, copper, nickel, lead, and zinc) in the topsoil were measured at each location. The type of land use and rock type was also recorded for each location.

```
> names(a)
 [1] "x"        "y"        "Landuse" "Rock"     "Cd"
 [6] "Co"       "Cr"       "Cu"       "Ni"       "Pb"
[11] "Zn"

y <- a$Pb

x1 <- a$Cd
x2 <- a$Co
x3 <- a$Cr
x4 <- a$Cu
x5 <- a$Ni
x6 <- a$Zn
```

The variables x, y are the coordinates. Landuse and Rock represent type of land use (forest, pasture, meadow, tillage) and rock type (Argovian, Kimmeridgian, Sequanina, Portlandian, and Quaternary). The other variables are concerntrations in ppm of the following chemical elements: `Cd`: Cadmium, `Co`: Cobalt, `Cr`: Chromium, `Cu`: Copper, `Ni`: Nickel, `Pb`: Lead, `Zn`: Zinc.

Answer the following questions:

a. Consider the full model (regression of y on $x_1, x_2, x_3, x_4, x_5, x_6$). Construct the $\mathbf{X}$ matrix.

b. Compute $\hat{\boldsymbol{\beta}}$, s_e^2, and $\mathbf{H}$ of the full model.

c. Test the overall significance of the model using the F test for the general linear hypothesis:

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})' \left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})}{m s_e^2},$$

d. Use the F test for the general linear hypothesis to test:

$$\begin{aligned} H_0 &: (\beta_1, \beta_3)' = \mathbf{0} \\ H_a &: (\beta_1, \beta_3)' \neq \mathbf{0} \end{aligned}$$

e. Test the hypothesis of question (d) using the extra sum of squares principle.

f. Consider the full model. Test the hypothesis that $\beta_4 = 0$ using the t statistic.

g. Consider the following two linear constraints:

$$\begin{aligned} H_0 &: \beta_1 + \beta_2 - 3\beta_5 = 2, \beta_3 + \beta_5 + \beta_6 = 3 \\ H_a &: \text{Not true} \end{aligned}$$

Under H_0 compute $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ and the unbiased estimator of σ^2 (make sure you use the constrained residuals).

h. Refer to question (g). Use the canonical form of the model to test the hypothesis with the method of extra sum of squares.

i. Repeat (h) using the F test for the general linear hypothesis.

j. Refer to question (g). Construct the variance covariance matrix of $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$.