Humboldt-Universität zu Berlin

Institut für Physik

Winter Semester 2010/2011

Dr. U. Husemann, V. Ferrara

Statistical Methods of Data Analysis
Problem Set #2

Due Date: Discussion on Monday, November 1, 2010, during the exercise

Problem 1:

The one-dimensional histogram named `temperatures` in the data file `ProblemSet2.root` that you can download from the course web page displays the distribution of the monthly global temperature anomaly of the 20th century.

i) Normalize this histogram to unit area. Is what you get a PDF? Why?

ii) Plot the cumulative distribution and find the median $T_{1/2}$.

Tip: Once the root file is open, retrieve the histogram object with the instruction:
`TH1D *h1 = (TH1D*)file->Get("ProblemSet2.root")`
where file is the name of the `TFile` object you have defined. `h1` is now a pointer to a `TH1D` object. This implies that you now have to use an arrow (i.e. `->`) to apply methods to the pointed object, e.g. `h1->Draw()`

Problem 2:

The scatter plot named `ScatterPlot` in the same data file is a two-dimensional histogram $h(x,y)$.

i) Normalize this histogram to unit area.

ii) Plot the marginal distributions $h_x(x)$ and $h_y(y)$.

iii) Plot the conditional distribution for $h(y|x < 5)$.

iv) What is the probability of $x < 2$ given that $y > 6$?

Problem 3:

A psychologist wants to study the relation of the cleverness between father and son. He measures the IQ of ten father-son pairs and obtains the following results:

father	74	83	85	96	98	100	106	107	120	124
son	76	103	99	109	111	107	91	101	120	119

Do you think this data suggests a correlation of the cleverness between father and son?

URL: `http://www-zeuthen.desy.de/~husemann/teaching/2010_ws/datenanalyse/uebungen.html`