

Lab 01 – Excel

adult data set

<http://archive.ics.uci.edu/ml/datasets/Adult>

adult.data

adult.names

Import data

- open Excel
- open adult.data in notepad
- copy paste to Excel
- data/text to columns - comma delimited

Insert column labels

- copy paste adult.names to another excel sheet
- keep only rows with attribute labels
- data/text to columns - colon delimited
- insert row on top of data worksheet
- paste special (values, transpose)
- window horizontal split, show first 5 rows and freeze pane

Summary statistics for numerical attributes

- count, average, min, max, stdev, skew, countblank, countif(..., " ?")
- select non adjacent ranges
- ctrl+R to copy formulas to all numeric attributes
- countA counts the number of non-empty cells

Select a random sample of the records

- in the first column insert a row_id counter in order to be able to recuperate the initial order
- in the last column insert rand() function, fill-down for all records
- sort records by the new column
- select the first 1000 records
- copy sample to another sheet

Filters

- Data/filter
- check distinct values of categorical attributes
- filter by $20 \leq \text{age} \leq 25$
- subtotal function

Charts

- Histogram for age (bin 20,25,...,90); data/data analysis, histogram, select input and bin ranges. In worksheet select Bin-Frequency and go to Insert/column 2D. We need then do properly set the data series, right click on chart and select data
- Histogram for work-class (nominal attribute). Data/advanced filter list range=work-class and make sure you set 'unique records only'.
- copy paste the resulting list
- use countif(range, criteria) and create column chart

Array Functions

- histogram for hours-per-week
- bins from 0 to 90 on column Q
- select destination range of cells (plural) and type the formula
- =FREQUENCY(M:M,Q4:Q12) ctrl+shift+enter
-
- average age for people above 40 - =AVERAGE(IF(A:A>40,A:A,FALSE))
- average hour-per-week for people above 40 - =AVERAGE(IF(A:A>40,M:M,FALSE))

Pivot tables

- Insert\Pivot table
- Drag education to row labels and education to values
- Drag sex to column labels
- click on count of education, select value field settings, show value as: % of row
- add race to column label
- check sorting and filtering options
- create a chart based on table