Statistik I - Exercise session 1
28.4.2014 & 5.5.2014

Info

- Classroom: SPA1 220

- Time: Mondays, 16:15 - 17:45

- in English

- Assignments on webpage (lvb>staff>PB)

Contact: Petra Burdejova
             petra.burdejova@hu-berlin.de
             Office: SPA1 R400 (upon agreement)

Schedule:

Date	Week	Exercises
28.04.14	E1	1-2, 1-3 (even), 1-10
05.05.14	E1	1-2, 1-3 (even), 1-10
12.05.14	E2	1-20, 1-22, 1-32
19.05.14	E2	1-20, 1-22, 1-32
26.05.14	E3	TBA
02.06.14	E3	TBA
09.06.14	–	–
16.06.14	E4	TBA
23.06.14	E5	TBA
30.06.14	E5	TBA
07.07.14	E6	TBA
14.07.14	E6	TBA

Review

- week 1 & 2

- Slides: Introduction, Descriptive Statistics 1-27(30)

Statistics
science; information itself, methods/processes to work with this information

Descriptive Statistics
quantitive description/summary of sample
Statistical Inference
obtaining conclusion (based on theory of probability)

Steps of Stat.Survey (plan, survey, analysis,interpretation)

Statistical unit/individual object or process, which carries specific information/characteristics
set of individuals is called population

Variable (X)
characteristic of unit
having values $x_1, x_2, \ldots, x_n$

Sample
subset from st.population

Scales / Levels of measurement

- Qualitative

  - Nominal (gender,nationality,... !binary)
  - Ordinal (can be sorted)

- Quantitative

 - Interval

 - Ratio   - Continuous

 - Absolute   or   - Discrete

Notation
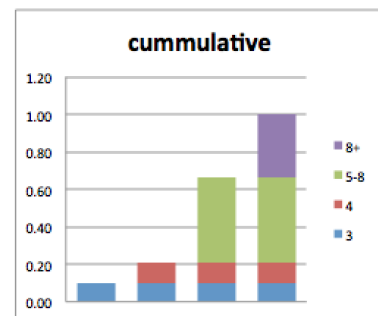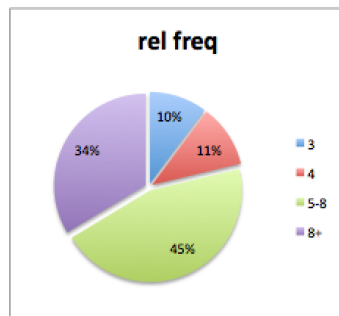Random variable X having n numbers of observations $x_1, \ldots, x_n$
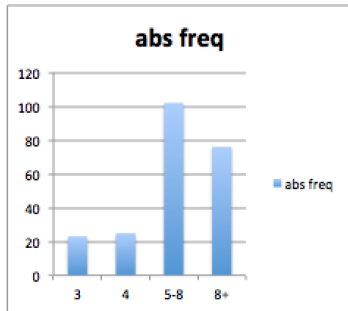
Frequency

$$h(X = x_j) = \sum_{i=1}^{n} \mathrm{I}(x_i = x_j) \tag{1}$$

with indicator function I.
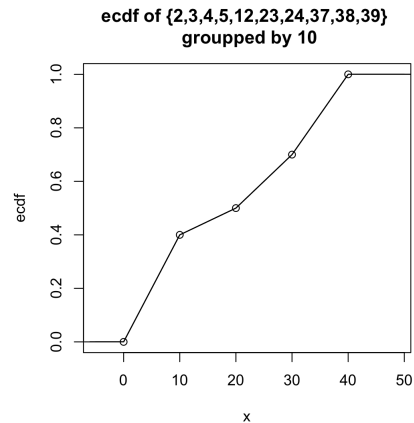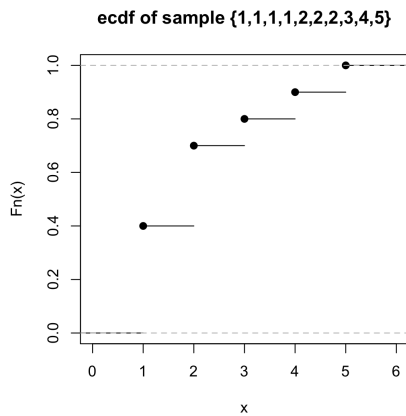
- Absolute

- Relative $f(x_j) = \frac{h(x_j)}{n}$

  Shown as: Table, Histogram, etc..

group	abs freq	rel freq	perc	cummulative
3	23	0,10177	10%	0,10176991
4	25	0,110619	11%	0,21238938
5-8	102	0,451327	45%	0,66371681
8+	76	0,336283	34%	1
TOTAL	226	1	100%	

Empirical cumulative distribution function

$$F(x) = \begin{cases} 0 & x < x_1 \\ \sum_{i=1}^{j} f(x_i) & x_j \leq x < x_{j+1} \quad j = 1, \ldots, n-1 \\ 1 & x \geq x_n \end{cases}$$

ecdf of sample {1,1,1,1,2,2,2,3,4,5}

ecdf of {2,3,4,5,12,23,24,37,38,39} groupped by 10

Empirical cumulative distribution function for grouped variable

$$F(x) = \begin{cases} 0 & x \leq x_1^{down} \\ \sum_{i=1}^{j-1} f(x_i) + \frac{x - x_j^{down}}{x_j^{up} - x_j^{down}} f(x_j) & x_j^{down} < x \leq x_j^{up} \quad j = 1, \ldots, n \\ 1 & x > x_n^{up} \end{cases}$$

Exercises

Exercise 1-2

In a German city K saving behavior of workers is examined in March 1992. By what objective, spatial and time characteristics is examined population identified?

Exercise 1-3 (even)

Specify the scale of the following characteristics:

2) temperature in Celsius

4) number of children

6) notes at school

8) standard deviation

10) subscribed newspapers

12) election result of a party

14) fares (for ticket)

16) number of books in a library

18) speed

20) difficulty level (of climbing tour)

22) tariff class (for car liability insurance)

24) price of any commodity

26) income

28) vocation learnt

30) number of pages (of book)

32) annual turnover

34) field of study

36) quality class (for fruit)

38) place of residence

40) aggressiveness

42) intelligence

44) financing (fundings) of studies

46) number of semesters

Exercise 1-10

25 students of the X-University in X-town were asked in June 2012 about their field of study, number of siblings and income. The outcome was as follows:

No.	Name	Studies	No. of Siblings	Income
1	Martin A.	Economics	0	924
2	Ute A.	Social S.	1	789
3	Wilhelm A.	Business	0	1365
4	Kurt B.	Business	1	683
5	Sylvia B.	Polit. S.	1	744
6	Elke D.	Polit. S.	2	640
7	Klaus D.	Social S.	2	631
8	Theo E.	Economics	1	814
9	Jean F.	Polit. S.	1	778
10	Elvira G.	Business	0	1062
11	Karl H.	Business	0	1230
12	Andreas K.	Economics	1	700
13	Thomas K.	Business	0	850
14	Chris L.	Social S.	3	641
15	Uwe L.	Polit. S.	2	640
16	Axel M.	Business	0	850
17	Maria M.	Business	1	683
18	Ruth M.	Social S.	0	616
19	Bärbel N.	Business	1	683
20	Armin R.	Business	2	683
21	Christa R.	Economics	1	660
22	Bernd S.	Business	1	1440
23	Claudia S.	Social S.	3	794
24	Udo T.	Economics	0	660
25	Clausia W.	Polit. S.	1	640

a) What is the statistical population and units in this survey?
   What characteristics of identification can you define in this population?

b) How is the variable/category "Field of study" scaled?
   Calculate its absolute and relative frequency. Plot the results.

c) How is the variable/category "Number of siblings" scaled?
   Calculate its absolute and relative frequency.
   Calculate empirical cumulative distribution function.
   Plot the results.

d) How many student have at most 2 siblings?

e) What percentage of students has at least two siblings?

f) What percentage of students has 1 or 2 sibling?

g) How is the variable/category "Income" scaled?
   With respect to following grouping: $[600, 650), [650, 700), [700, 900), [900, 1200), [1200, 1450]$
   Calculate its absolute and relative frequency.
   Calculate empirical c.d.f.
   Plot the results.

h) Taking the results of previous task g) compute:

   - What percentage of students has income from 750 to 1300 € ?
   - What percentage of students has income more than 800 € ?
   - What is the highest income of the 50% of the students with the lowest income?
   - What is the smallest income of the 20% of the students with the highest income?