Statistik I - Exercise session 4
09.05.2017

Preliminary schedule:

Date	Week	Exercises
18.04.2017	E1	1-1-c,e,g,i,k ; 1-2-a,b,d ; 1-3 ; 1-4 ; 1-5-a,b ; 1-6
25.04.2017	E2	1-7-C,D ; 1-8 ; 1-9-B; 1-10-A(a..e), 1-10-B(a,b,c) ; 1-11-a,b
02.05.2017	E3	2-1; 2-2; 2-3; 2-7
09.05.2017	E4	2-9; 2-19, 2-20
16.05.2017	E5	TBA
23.05.2017	E6	TBA
30.05.2017	E7	TBA
06.06.2017	E8	TBA
13.06.2017	E8	TBA
20.06.2017	E9	TBA
27.06.2017	E10	TBA
04.07.2017	E11	TBA
11.07.2017	E12	TBA
18.07.2017	E13	TBA

Review

- Slides: Descriptive Statistics 75-105/788

Frequency

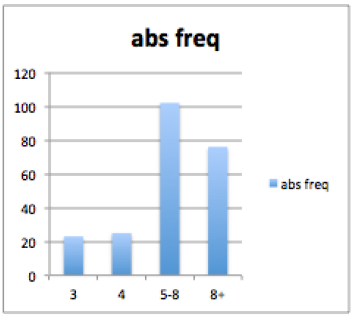
$$h(X = x_j) = \sum_{i=1}^{n} \mathrm{I}(x_i = x_j) \tag{1}$$

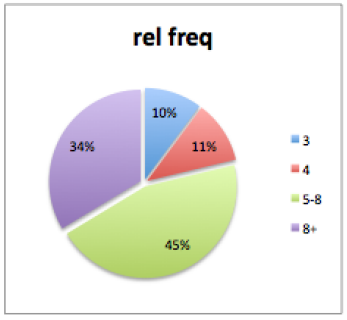
with indicator function I.

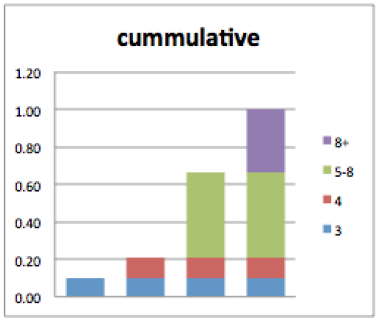
- Absolute
- Relative $f(x_j) = \frac{h(x_j)}{n}$

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%	

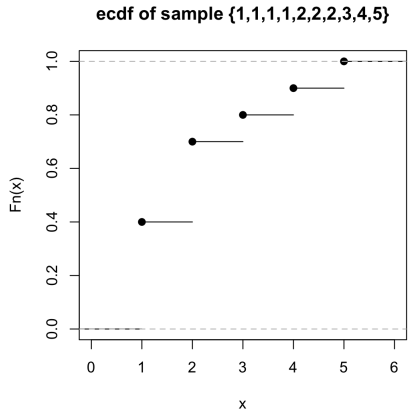


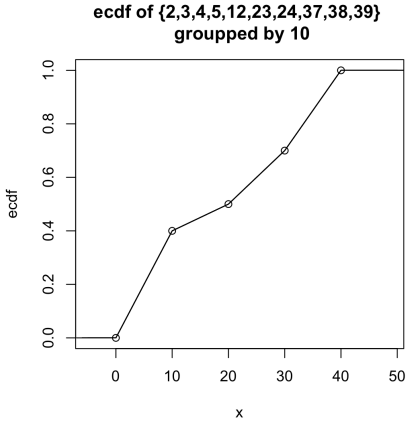




Empirical cumulative distribution function

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





Empirical cumulative distribution function for grouped variable

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

Mean

average value, calculated as a sum of sampled values divided by the number of observations (items in the sample)

Mode

value having the highest frequency (i.e. appears most often in sample)

Median

middle value, separating lower half of ordered sample from upper half

Mean

ungrouped variables $\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

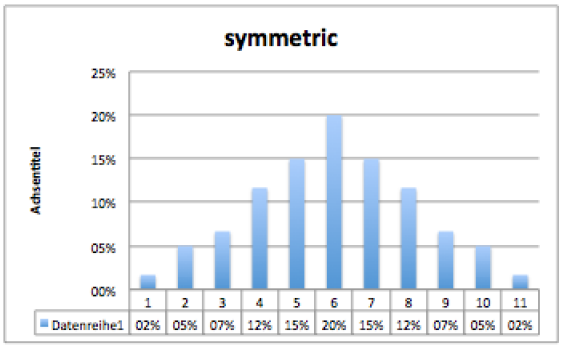
classified/grouped and discrete variable $\overline{x} = \frac{1}{n} \cdot \sum_{j=1}^{k} x_j \cdot h(x_j) = \sum_{j=1}^{k} x_j \cdot f(x_j)$

weighted $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i \cdot g_i}{\sum_{i=1}^{n} g_i}$

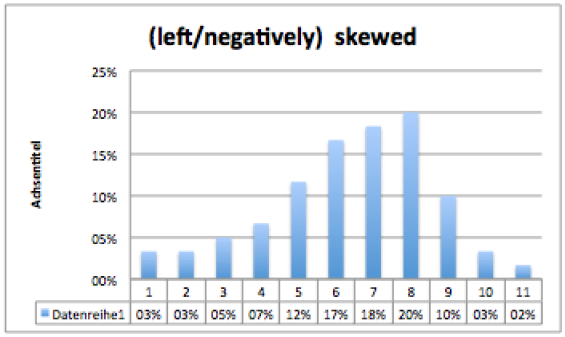
pooled $\overline{x} = \sum_{\ell=1}^{r} \dfrac{n_\ell}{n} \overline{x}_\ell$

n_ℓ observations and $\overline{x}_\ell$ middle value in group ℓ

value	frequency	rel.frequency		cumulative
1	1	0,017	1,7%	1,7%
2	3	0,050	5,0%	6,7%
3	4	0,067	6,7%	13,4%
4	7	0,117	11,7%	25,0%
5	9	0,150	15,0%	40,0%
6	12	0,200	20,0%	60,0%
7	9	0,150	15,0%	75,0%
8	7	0,117	11,7%	86,7%
9	4	0,067	6,7%	93,4%
10	3	0,050	5,0%	98,4%
11	1	0,017	1,7%	100,0%
TOTAL	60	1,000	100%	
MEAN	6			
MEDIAN	6			
MODE	6			



value	frequency	rel.frequency		cumulative
1	2	0,033	3,3%	3,3%
2	2	0,033	3,3%	6,6%
3	3	0,050	5,0%	11,6%
4	4	0,067	6,7%	18,3%
5	7	0,117	11,7%	30,0%
6	10	0,167	16,7%	46,6%
7	11	0,183	18,3%	65,0%
8	12	0,200	20,0%	85,0%
9	6	7,000	10,0%	95,0%
10	2	0,033	3,3%	98,3%
11	1	0,017	1,7%	100,0%
TOTAL	60	1,000	100%	
MEAN	6,4			
MEDIAN	7			
MODE	8			



Mode/Modus

unclassified data $x_D = \left\{ x_j \mid h_j = \max_{x_k} h_k \text{ or } f_j = \max_{x_k} f(x_k) \right\}$

classified data $x_D = x_j^u + \dfrac{\hat{f}(x_j) - \hat{f}(x_{j-1})}{2 \cdot \hat{f}(x_j) - \hat{f}(x_{j-1}) - \hat{f}(x_{j+1})} \cdot (x_j^{up} - x_j^{down})$

Median

unclassified data $x_{0,5} = x_{\left(\frac{n+1}{2}\right)}$ if n odd

$x_{0,5} = \frac{1}{2} \cdot \left\{ x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right\}$ if n even

classified data $x_{0,5} = x_j^{down} + \dfrac{0,5 - F(x_j^{down})}{f(x_j)} \cdot (x_j^{up} - x_j^{down})$

	Qualitative		Quantitative		
	Nominal	Ordinal	Interval	Ratio	Absolute
Mean	✗	✗	✓	✓	✓
Median	✗	✓	✓	✓	✓
Mode	✓	✓	✓	✓	✓

Exercises - Leftovers from 2.5.

Exercise 2-3
Specify the scale of the following characteristics:

26) income

27) marital status

28) vocation learnt

29) date(year) of the birth

30) number of pages (of book)

31) cause of death

32) annual turnover

33) land area

34) field of study

35) latitude (of the earth)

36) quality class (for fruit)

37) eye colour

38) place of residence

39) telephone number

40) aggressiveness

41) legal form of the company

42) intelligence

43) social status

44) financing (funding) of studies

45) production time

46) number of semesters

47) exam results - points

Exercises - New

Exercise 2-9

In a control point of rail line between the A and B the following time intervals between the train line were observed in minutes:

59; 43; 36; 63; 23; 4; 29; 41; 43; 31; 29; 69; 57; 36; 112; 43; 14; 11; 18; 77; 81;47; 12; 43; 44; 16; 80; 6; 52; 5; 5; 6; 21; 43; 44; 46; 51.

a) What is the variable which we measure/observe? What is the scale of that variable?

b) Determine the absolute and relative frequency of the variable. Split the variable into 30 minutes intervals. Show as a graph.

c) Give an empirical distribution function.

Exercise 2-19 Based on the interval data from exercise 2-9:

a) State and compute 3 meaningful location parameters.

b) What is the average time period based on the original data? How do you explain that it differs from result in part a) ?

Exercise 1-20

Mr. Meier own a wholesale of garden-dwarfs with three branches: Berlin, New York and Flensburg. At the end of financial year he would like to have a full overview of the business over the last year and therefore he requests the info about the orders from all three subsidiaries.

A) Berlin sent the following information:

Value of orders	Number of orders
0 - 20 000	15
20 000 - 50 000	30
50 000 - 150 000	45
150 000 - 300 000	10

a) What is the examined variable? How is it scaled?

b) For which statistical unit was this variable observed?

c) Calculate the average value per order.

B) New york replied to Mr. Meier briefly:

Y: ”the value (in $) per order” ; 95 orders ; $\bar{y} = 60000$ $.

For a better comparison, Mr. Meier wants to see this statement in €.

Calculate the average value per order when exchange rate 1 $ = 1.5 €.

C) From Flensburg Mr. Meier received the following data:

2 000 €, 12 000 €, 17 000 €, 12 000 €, 200 000 €.

a) What average value(s) can you state meaningfully?

Justify their advantages/disadvantages in this situation.

b) Per fax they corrected the value of the last order to 2 000 €.

Calculate the average value per order from corrected data.

D) Calculate the average value per order for the entire company.