L2

Basic Statistics for Economists

Spring 2020

Department of Statistics

Stockholm University

Today: Descriptive statistics

Univariate data = one variable at a time

Describe univariate data (a variable) by means of,

- ***Tables***
  - Distributions, frequency distributions (frequency = number of)
- ***Diagrams, graphics***
  - One variable at a time, univariate
  - Categorical and numerical variables
- ***Numerically***
  - Location ("where are the values typically located")
  - Variation, dispersion ("how spread out are they")

Frequency distribution – one variable

- ***Number of*** objects/observations that share the same property

$$n_k = \textbf{no. of objects with property } k$$

- The entire set of all n_k over all possible k is called the ***frequency distribution*** (sv. *frekvensfördelningen*)

- If there are in total n objects and in total C different categories we have

$$\sum_k n_k = n_1 + n_2 + \cdots + n_C = n$$

- ***Relative frequencies*** (%): $100 \cdot \frac{n_k}{n}$

Works for categorical and discrete numerical variables – but how do we deal with continuous variables?

Stockholm University

Frequency tables – count the numbers

- Categorical – nominal or ordinal
- Numerical – discrete or ***class-divided*** continuous

Count the number that fall into each defined category:

Nominal scale

Flavor	Frequency	Relative frequency %
Chocolate	70	35,0 %
Vanilla	50	25,0 %
Strawberry	45	22,5 %
Raspberry	30	15,0 %
Licorice	5	2,5 %
Sum	200	100 %

$\boxed{n_k}$ $\boxed{100 \cdot \frac{n_k}{n}}$

Largest first!

Pareto, NCT s. 32-33

Smallest last!

Note! Frequencies are always numerical but the variable is not necessarily numerical!

Stockholm University

Frequency tables, cont.

Categorical
ordinal scale

Grade	Frequency	Relative frequency %
A	30	15,0 %
B	56	28,0 %
C	80	40,0 %
D	20	10,0 %
E	14	7,0 %
Sum	200	100 %

Arrange in order of ranking!
Pareto not recommended!

Discrete
ratio scale

No. of points	0	1	2	3	4	Sum
Frequency	7	42	98	63	70	280
Relative frequency %	2,5	15,0	35,0	22,5	25,0	100

Arrange in order of numerical magnitude!
Pareto not recommended!

Stockholm University

Class separated continuous variable

- Continuous numerical variables (or discrete with many values) may be grouped into ***classes, bins*** – i.e. ***intervals***
  - categorization (sv. *klassindelning*)

- ***Classes and class widths*** must be defined
  ex. (0-4,99) (5-9,99) (10-19,99) (20 -) ⟵ (*open class,* ≥20)

- Same class width or varying width? What does **NCT** say?

- Summarize in a table - ordered by magnitude

Stockholm University

Class separated continuous variable

- **Cumulative** – indicates the total number of observations whose values are (e.g.) ***less than the upper limit*** of each class

Income per month, tkr	Frequency	Relative freq.	Cumulative freq.	Cumulative relative freq.
< 20	20	10,0 %	**20**	**10,0 %**
20 – 40	40	20,0 %	**60**	**30,0 %**
40– 60	74	37,0 %	**134**	**67,0 %**
60 – 80	40	20,0 %	**174**	**87,0 %**
80 – 100	18	9,0 %	**192**	**96,0 %**
100 – 120	6	3,0 %	**198**	**99,0 %**
120 – 140	2	1,0 %	**200**	**100,0 %**
≥ 140	0	0,0 %	**200**	**100,0 %**
Summa	200	100 %		

8 bins

Accumulated relative frequencies

Stockholm University

Graphical presentation – one variable

Frequencies (absolute or relative)

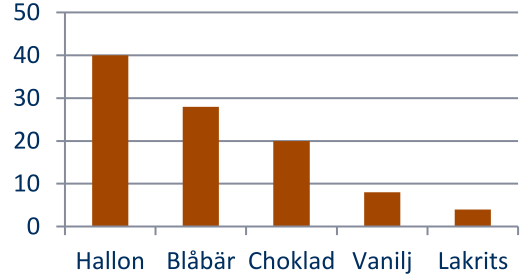
- ***Bar charts*** *(sv. stapeldiagram)*
  - categorical, nominal and ordinal; discrete numerical
  - ordered in the same way as we did for freq. tables
- ***Pie charts***
  - categorical, nominal
- ***Histogram*** (adjacent bars, intervals)
  - class separated continuous variable, discrete many values

Diagram types – qualitative, categorical

Bar charts

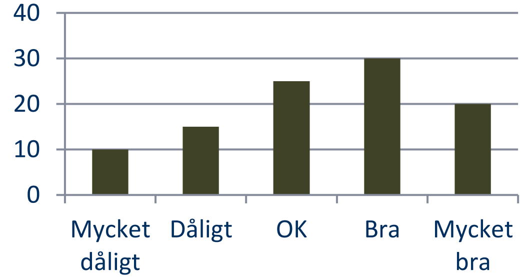
Nominal scale

Pareto ordered



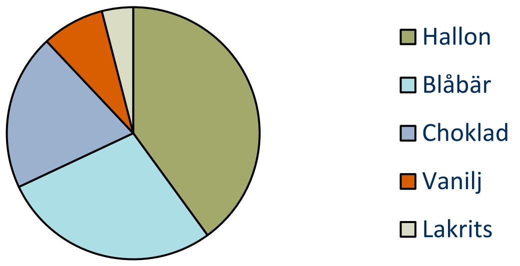
Ordinal scale

Not Pareto ordered



Pie charts

Nominal scale

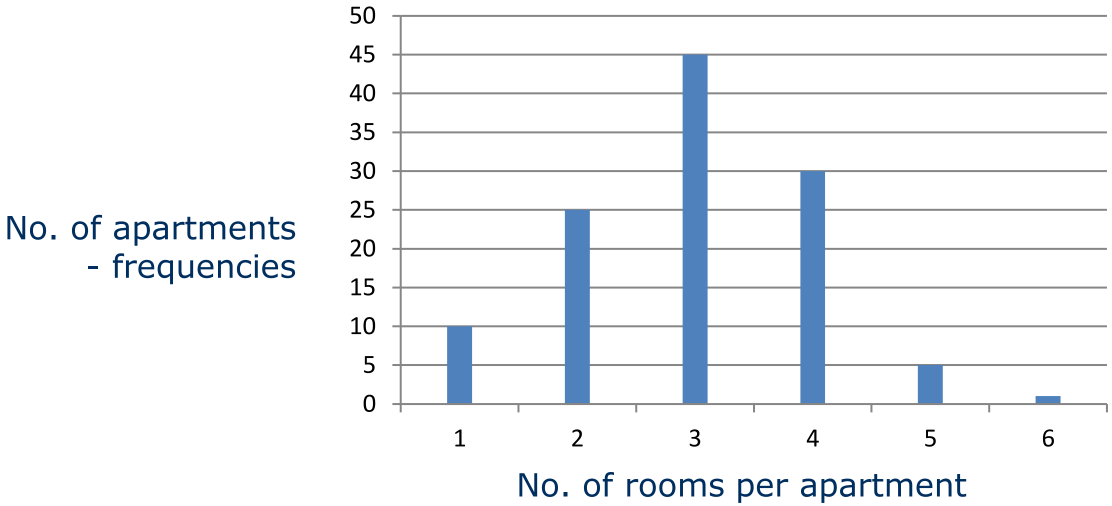


Start at 12 o'clock and move clockwise, starting with the largest, second largest and so on …

Stockholm University

Diagram types - numerical discrete

Bar chart (sv. stolp- el. stapeldiagram)



Ordered by magnitude

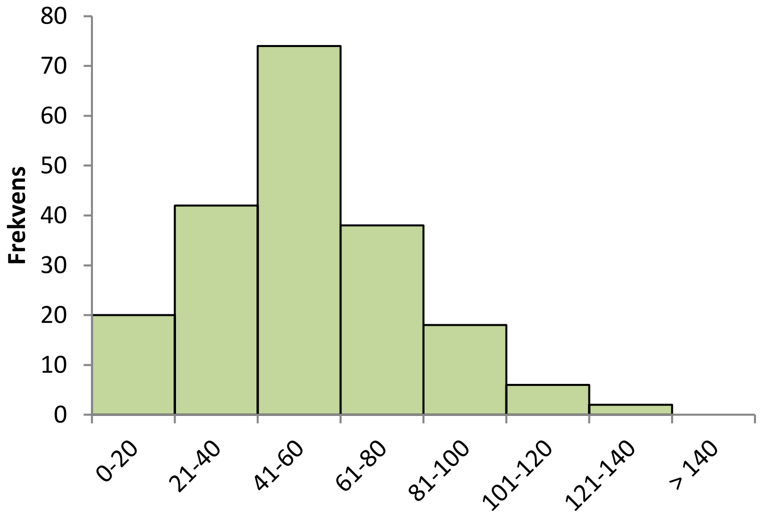
Histogram - numerical continuous

- Histogram are used for ***continuous*** variables
- ***Class widths*** are defined (bins)
- Inclusive and non-overlapping – each observation belongs to one class
- The ***frequency*** of a class is represented by ***area of the bar*** not its height (se NCT sid 52-53)
  - *however, if class widths are the same for all bins the heights are proportional to the areas and thus the frequencies*
- Open classes (e.g. >65) are indicated with e.g. dotted lines
  - *we don’t know where it ends and thus nor the area!*

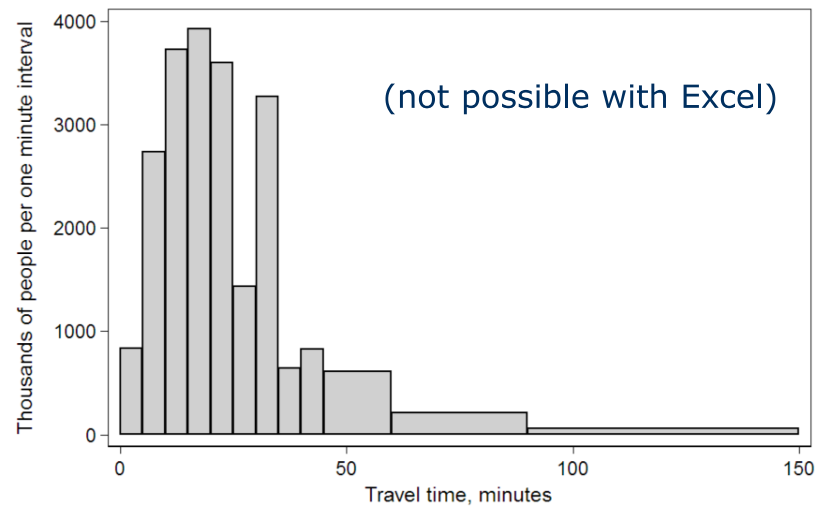
Histogram

How is the scale of the y-axis used?

Equal class widths

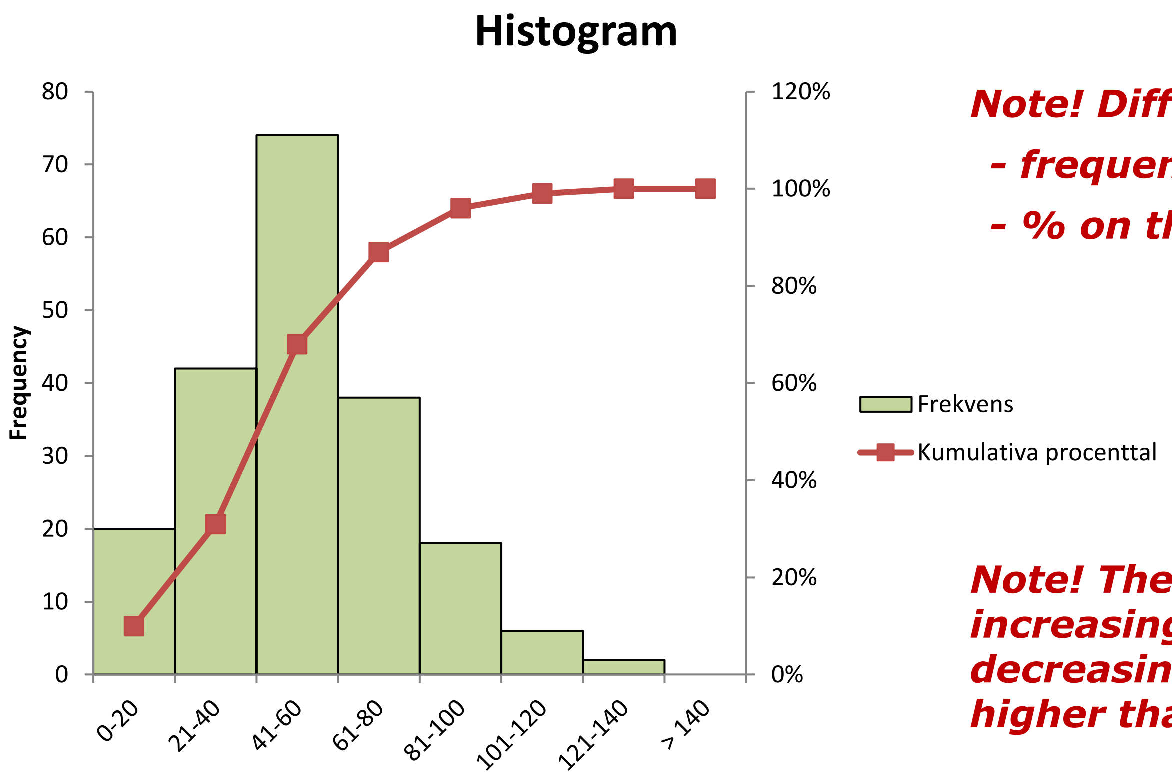


Unequal class widths



Height is proportional to the area

Cumulative relative frequency - Ogive



Note! Different axis scales!

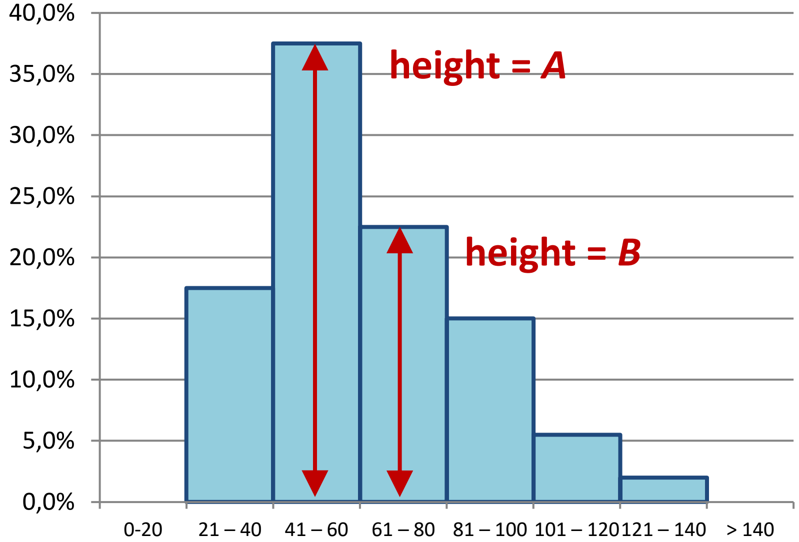
- ***frequency on the left side***
- ***% on the right side***

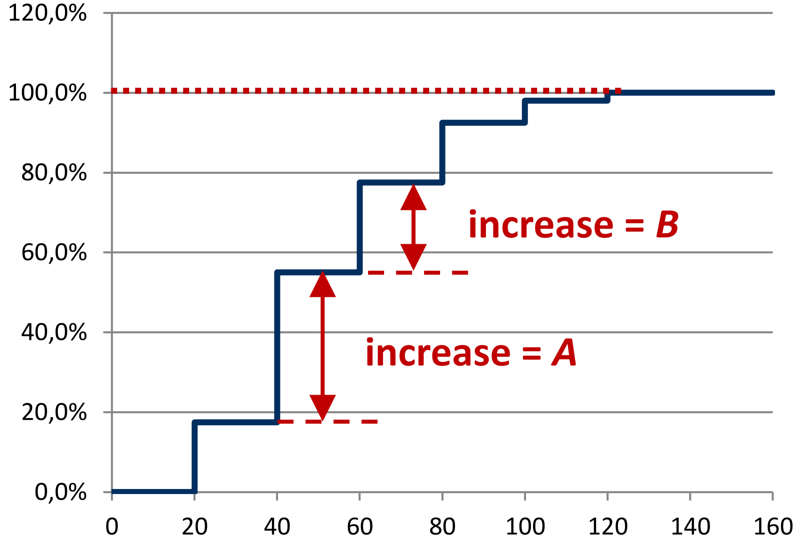
Note! The red ogive is an increasing function (never decreasing) and never goes higher than 100 %

Stockholm University

Cumulative rel. freq. – Step function

- The increase at each step is equal to the height of the corresponding bar in the histogram





Increases (never decreases) and never goes above 100 %

Stockholm University

Quick summary:

What type of diagram do we use to show a frequency distribution?

- Categorical, nominal **Bar chart, Pareto-ordered; Pie chart**
- Categorical, ordinal **Bar chart, ordered by rank (lowest - highest)**
- Numerical, discrete, few values **Bar chart, one bar for each discrete value**
- Numerical, discrete, many values **Histogram, divided into classes (approximates continuous)**
- Numerical, continuous **Histogram , divided into classes**
- Visualizing a cumulative frequency distribution?

Ogive or Step function (discrete and continuous)

Historically interesting!

Stem-and-Leaf Displays

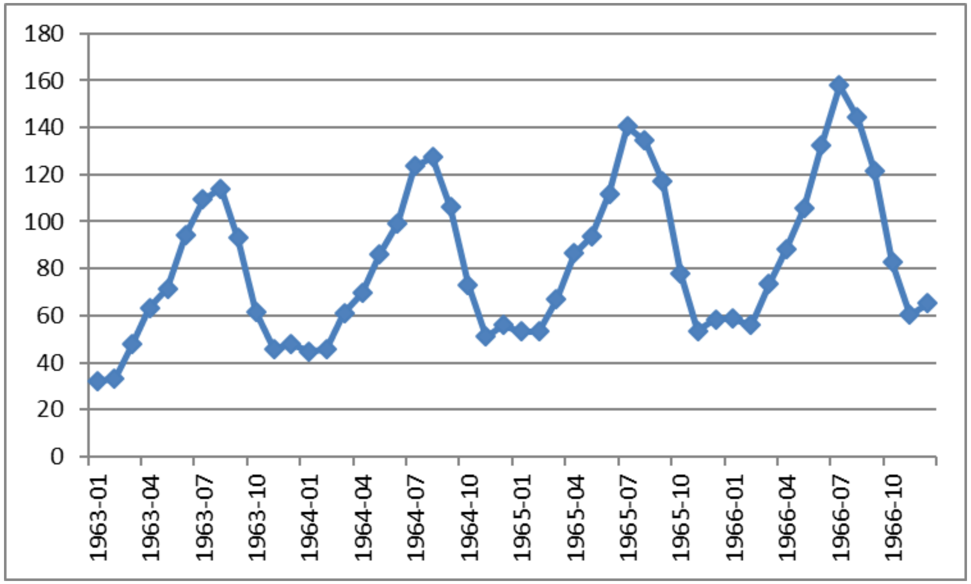
- sv. ***stambladdiagram***
- Provides exact values and visualizes the distribution
- In this example
  - stems = tens (10, 20, …)
  - leafs = ones (0, 1, 2,…, 9)
- *Not very common these days, before the era of graphical printing*

		8	8	
		7	3	
		6	3	
	8	5	2	
	6	5	2	
	5	4	1	
6	3	3	0	1
1	2	3	4	5

Stockholm University

Time series, observations over time

- **Line chart, time series plot**
  - visualizes changes over time (what types of change?)
  - time on the *x*-axis, observed values on the *y*-axis
  - points are connected



No. of passenger miles (Eng. miles), domestic flights U.K. Jan 1963 - Jan 1966

Month	Miles
1963-01	32,2
1963-02	33,1
1963-03	48,1
1963-04	63,2
1963-05	71,5
1963-06	94,1
1963-07	109,4
⋮	⋮

Stockholm University

Numerical measures - summaries

- Define numerical measures that summarize the most important properties of a set of observations
- ***Location*** – where are the observations?
  - Around 4, 25, 100 or 10 000?
  - **Measures of location, central tendency**
- ***Dispersion*** – how spread out are the observations from the central location?
  - about 2-8 or 4-500? Or in the interval 100±20?
  - Many close to the center? Or in the “tails” i.e. endpoints?
  - **Measures of variability**

The data and notation

- Denote the variable by $\boldsymbol{x}$ (or some other letter y, z, u, v, …)
- $\boldsymbol{n}$ observations, sample size ($\boldsymbol{N}$ population size)
- Denote by indexing with $\boldsymbol{i = 1, 2, 3, \ldots, n-1, n}$ ***(labels)***
- Value of the $\boldsymbol{i}^{\mathbf{th}}$ observation of the variable $\boldsymbol{x}$ is denoted by $\boldsymbol{x_i}$
- The entire set of observed values may be denoted as

$$\boldsymbol{(x_1, x_2, x_3, \ldots, x_{n-1}, x_n)}$$

Location: Mode

- Sv. ***typvärde***
- The most frequently occurring value; largest frequency

ex. $(4, 2, 3, 3, 5, 1, 3, 5) \Rightarrow$ $Mode$ = **3** ***Unimodal***

ex. $(5, 2, 3, 3, 5, 1, 3, 5) \Rightarrow$ $Mode$ = **3 and 5**

Bimodal

- Useful for categorical variables (nominal and ordinal scales)

ex. $(b, a, c, b, d, b, a, e) \Rightarrow$ $Mode$ = **b**

Stockholm University

Location: Median

Can be applied to **ordinal** data also

- The ***median*** separates a <u>numerical</u> dataset in half
- 50% of the observations lie on either side of the median

- Arrange the observations in <u>increasing</u> order, smallest-largest

  n even ⇒ median = mean of the two in the middle

  n odd ⇒ median = the middle value

  ex. (2, **3**, **4**, 5) ⇒ median = **3,5**

  ex. (2, **3**, **4**, 25) ⇒ median = **3,5**

  ex. (2, 3, **4**, 25, 135) ⇒ median = **4**

Not so sensitive to extreme values

Stockholm University

Location: Mean

- ***Arithmetic sample mean*** (sv. *medelvärde*)
- Sum all and divide by the number
- $\boldsymbol{x}$-bar
- Mean value is sensitive of extreme values:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

ex. (2, 3, 4, 5) $\Rightarrow$ $\bar{x} = 3{,}5$

ex. (2, 3, 4, **25**) $\Rightarrow$ $\bar{x} = 8{,}5$

ex. (22, 23, 24, 25) $\Rightarrow$ $\bar{x} = 23{,}5$

- ***Population mean*** often denoted $\boldsymbol{\mu}$ or $\boldsymbol{\mu_x}$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

(”mu”, sv. ”my”)

More on means – grouped values

- $\boldsymbol{n}$ observations, distributed such that we have $\boldsymbol{n_k}$ of them sharing the same value $\boldsymbol{x = k}$
- E.g. $\boldsymbol{n_0}$ zeroes $(\boldsymbol{x = 0})$, $\boldsymbol{n_1}$ ones $(\boldsymbol{x = 1})$, $\boldsymbol{n_2}$ twos $(\boldsymbol{x = 2})$, … etc. up to $\boldsymbol{n_C}$ with value $\boldsymbol{x = C}$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{k=0}^{C}(n_k \cdot k) = \sum_{k=0}^{C}\left(\frac{n_k}{n} \cdot k\right)$$

Proportion with value *k*

- Ex. 0, 0, 0, 1, 1, 2, 2, 2, 3, 3, 3, 3 yields

We have 0,1,2, and 3 i.e. four categories

$$\bar{x} = \sum_{k=0}^{3}\left(\frac{n_k}{n}k\right) = \left(\frac{3}{12} \cdot 0 + \frac{2}{12} \cdot 1 + \frac{3}{12} \cdot 2 + \frac{4}{12} \cdot 3\right) = \frac{20}{12} = \frac{5}{3} = 1{,}6667$$

Stockholm University

Not required
- but useful knowledge

Geometric mean

- Monthly interest on an investment over 6 months:

  10%, 7%, -2%, 11%, 8%, 4%

- Total growth over the entire period is:

  $1{,}10 \cdot 1{,}07 \cdot 0{,}98 \cdot 1{,}11 \cdot 1{,}08 \cdot 1{,}04 = 1{,}4381$

  i.e. **+43,81%**

- ***Geometric mean***: $\bar{x}_g = \sqrt[n]{x_1 x_2 \cdots x_3}$

- Here $\bar{x}_g = \sqrt[6]{1{,}4381} = 1{,}0624$

  i.e. average interest per month = **6,24%**

Stockholm University

Variability: nominal and ordinal data

- Not much we can do
- However, the number observed classes/categories = ***C*** is a kind of measure of variability, although a simple measure
- Sample:

  (A, B, A, C, D, A, C, D, A, A, A, C) $\Rightarrow$ ***C* = 4**

  (A, B, C, D, E, F, G, H) $\Rightarrow$ ***C* = 8**

A measure very much out of scope for this course is based on information theory and entropies

- Entropy = 0 $\Leftrightarrow$ all observations are in the same single category, smallest possible dispersion
- Entropy = ${}^{2}\log(C)$ $\Leftrightarrow$ equal number in each category, maximal dispersion

Stockholm University

Variability: Range

Numerical variables

- The size of the observed range of values, the size of the interval where all observations lie
- Difference between the largest and smallest values

  Range = Max – Min

  ex. (2, 3, 4, 5) $\Rightarrow$ $\text{range} = 3$

  ex. (2, 3, 4, **25**) $\Rightarrow$ $\text{range} = 23$

  ex. (22, 23, 24, 25) $\Rightarrow$ $\text{range} = 3$

- Sensitive of extreme values

Variability: Quartiles – Quartile Range

Arrange the observations in increasing order:

- ***Q_1*** = ***1st quartile***

  25% of observations below, 75% above

- ***Q_3*** = ***3rd quartile***

  75 % of observations below, 25 % above

Can be applied to **ordinal** data also

- ***IQR*** = Inter Quartile Range **= $Q_3 - Q_1$**

  (sv. *kvartilavstånd)*

But not IQR!

- *50 % of the observations lie in an interval that is IQR wide*

Stockholm University

See NCT p. 64-67
Formulary p. 2

Percentiles

Let $\boldsymbol{x}_{(\mathbf{1})}, \boldsymbol{x}_{(\mathbf{2})}, \ldots, \boldsymbol{x}_{(\boldsymbol{n})}$ denote the *ordered* sample, ordered by size from the smallest value $\boldsymbol{x}_{(\mathbf{1})}$ to the largest $\boldsymbol{x}_{(\boldsymbol{n})}$

- Let $\boldsymbol{a} =$ **integer part** of $(n+1)\frac{p}{100}$
- Låt $\boldsymbol{b} =$ **decimal part** of $(n+1)\frac{p}{100}$
- $\boldsymbol{p}^{\text{th}}$ **percentile** $= x_{(\boldsymbol{a})} + \boldsymbol{b} \cdot (x_{(\boldsymbol{a}+1)} - x_{(\boldsymbol{a})})$

Ex. (11, 12, 14, 15, 17, 18, 20, 21, 21, 23, 30, 40), $\boldsymbol{n} = \mathbf{12}$

40th percentile: $(n+1)\frac{40}{100} = (12+1) \cdot 0{,}4 = \mathbf{5{,}2} \Rightarrow a = \mathbf{5}$ och $b = \mathbf{0{,}2}$

40th percentile $= x_{(\mathbf{5})} + \mathbf{0{,}2} \cdot \left(x_{(\mathbf{5}+1)} - x_{(\mathbf{5})}\right) = 17 + 0{,}2 \cdot (18 - 17) = \mathbf{17{,}2}$

Median = 50th precentile

- $\boldsymbol{n}$ **= 12** observations ordered by size:

  (11, 12, 14, 15, 17, **18, 20**, 21, 21, 23, 30 and 40)

- Start with $(\boldsymbol{n}+\mathbf{1}) \;=\; \mathbf{13}$

0,5 because it's half (50%)

- $(n+1)\cdot 0{,}5 = 6{,}5$ i.e. between the 6th and 7th

- $\boldsymbol{md} = \boldsymbol{P}_{\mathbf{50}} = \boldsymbol{x}_{(\mathbf{6})} + \mathbf{0,5}\left(\boldsymbol{x}_{(\mathbf{7})} - \boldsymbol{x}_{(\mathbf{6})}\right) = \frac{\boldsymbol{x}_{(\mathbf{6})}+\boldsymbol{x}_{(\mathbf{7})}}{\mathbf{2}} = \frac{\mathbf{18+20}}{\mathbf{2}} = \mathbf{19}$

  = mean value of the 6th and the 7th observations

Stockholm University

Quartiles – 25th and 75th percentiles

- 1st quartile

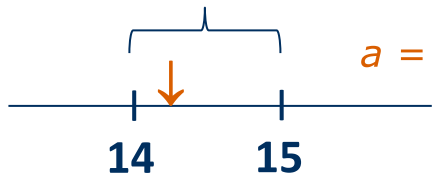
  25th percentile

  $(n+1)\cdot 0{,}25 = \boxed{3{,}25}$

  $\boldsymbol{Q_1} = 14 + 0{,}25\cdot\mathbf{1} = \mathbf{14{,}25}$

  Δ = 15-14 = 1

  a = 3, *b* = 0,25



  between 3rd and 4th

- 3:e quartile

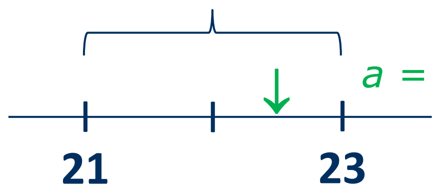
  75th percentile

  $(n+1)\cdot 0{,}75 = \boxed{9{,}75}$

  $\boldsymbol{Q_3} = 21 + 0{,}75\cdot\mathbf{2} = \mathbf{22{,}50}$

  Δ = 23-21 = 2

  a = 9, *b* = 0,75



  between 9th and 10th

- $\boldsymbol{IQR} = Q_3 - Q_1 = \mathbf{8{,}25}$

Stockholm University

With Excel

- Same data used in the example are in the cells A1–A12
- Write the following functions in any empty cell:

=MIN(A1:A12)

=QUARTILE.EXC (A1:A12;1)

=MEDIAN(A1:A12)

=QUARTILE.EXC (A1:A12;3)

=MAX(A1:A12)

	A	B	C	D
1	11			
2	12			
3	14			
4	15		Min	11
5	17		Q1	14,25
6	18		Md	19
7	20		Q3	22,5
8	21		Max	40
9	21			
10	23			
11	30			
12	40			

English and Swedish versions of Excel functions, see e.g.
http://www.exceldepartment.com/excelkurs/extramaterial/excelfunktioner-svenska-engelska/

Stockholm University

Box-and Whisker plots – visual summary

- We need:
  - smallest and largest values - ***min*** and ***max***
  - median, 1st and 3rd quartiles - ***Md***, $\boldsymbol{Q_1}$ and $\boldsymbol{Q_3}$

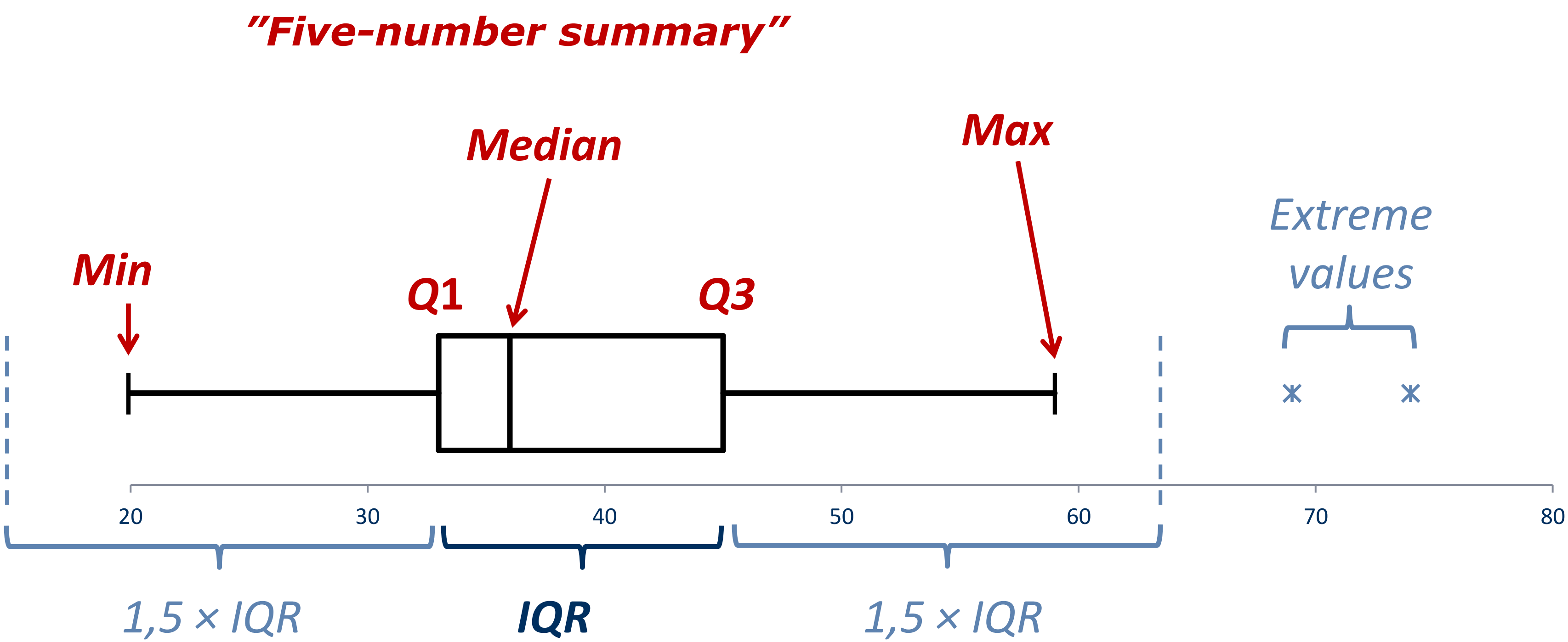
”Five-number summary” NCT p. 65

Definition of extreme values (according to Tukey):

- ***Outliers***: values that lie more than 1,5×*IQR* below Q_1 or above Q_3
- ***Extreme outliers***: 3×*IQR*

Stockholm University

Box plots, cont.



Stockholm University

Variability: Variance

- ***Average squared distance to the mean***

- Sample and population variances:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$ *("sigma")*

  - **Note!** For samples, divide by $\boldsymbol{n-1}$ rather than $\boldsymbol{n}$
  - Unit of measurement is transformed to square units

- ***Standard deviation***:
  - Restores unit of measurement
  - sv. *standardavvikelse*

$$s_x = \sqrt{s_x^2} \qquad \sigma_x = \sqrt{\sigma_x^2}$$

- ***Coefficient of Variation:*** read on you own in NCT p. 75

Stockholm University

Variance – alternative formulas

- **Sample variance**

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

Alternative formula sometimes easier to use

Excel: '=VAR.S(...)'

- **Population variance**

$$\sigma_x^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = \frac{\sum_{i=1}^{N} x_i^2 - N\mu^2}{N}$$

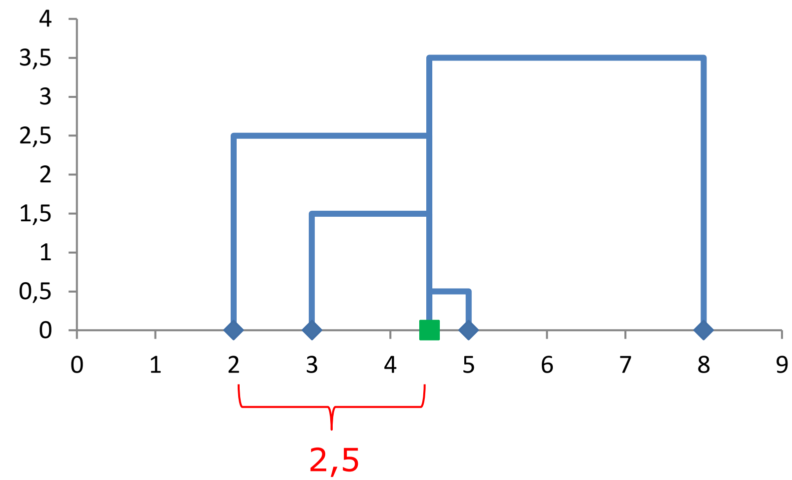
Excel: '=VAR.P(...)'

Stockholm University

Variance

- Four observations (2, 3, 5, 8); mean $\bar{x} = 4{,}5$
- Distance to mean $(x_i - \bar{x})$, square them and sum:

x_i	2	3	5	8	**18**
$(x_i - \bar{x})$	-2,5	-1,5	0,5	3,5	0
$(x_i - \bar{x})^2$	6,25	2,25	0,25	12,25	**21**
x_i^2	4	9	25	64	**102**



- Calculate the variance:
  - divide by $n - 1 = 3$ or $N = 4$?

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\mathbf{21}}{3} = 7 \qquad s_x^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{\mathbf{102} - 4 \cdot 4{,}5^2}{4-1} = \frac{21}{3} = 7$$

Properties of the variance

Think about it …

- Variances can never be negative (always ≥ 0)
- Add the same number $\boldsymbol{a}$ to every observed value

  $\Rightarrow$ the variance is unchanged

- Multiply every observed value by the same number $\boldsymbol{b}$

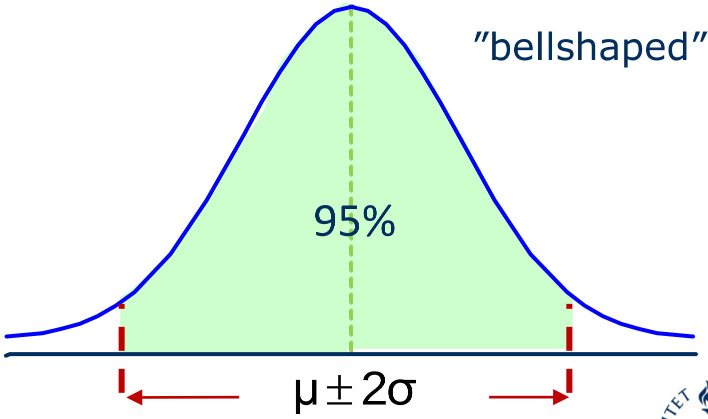
  $\Rightarrow$ the variance is $\boldsymbol{b}^2$ times larger

- Why is this interesting, to add or multiply?

Chebyshev's theorem and the Empirical rule

- Provides a description of how spread out our observations that relates to the standard deviation (variance):

Rule	$\mu \pm \sigma$	$\mu \pm 2\sigma$	$\mu \pm 3\sigma$	
Chebyshev:	0 %	75 %	88,89 %	Guaranteed
Empirical:	ca 68 %	ca 95 %	ca 100 %	Under some conditions



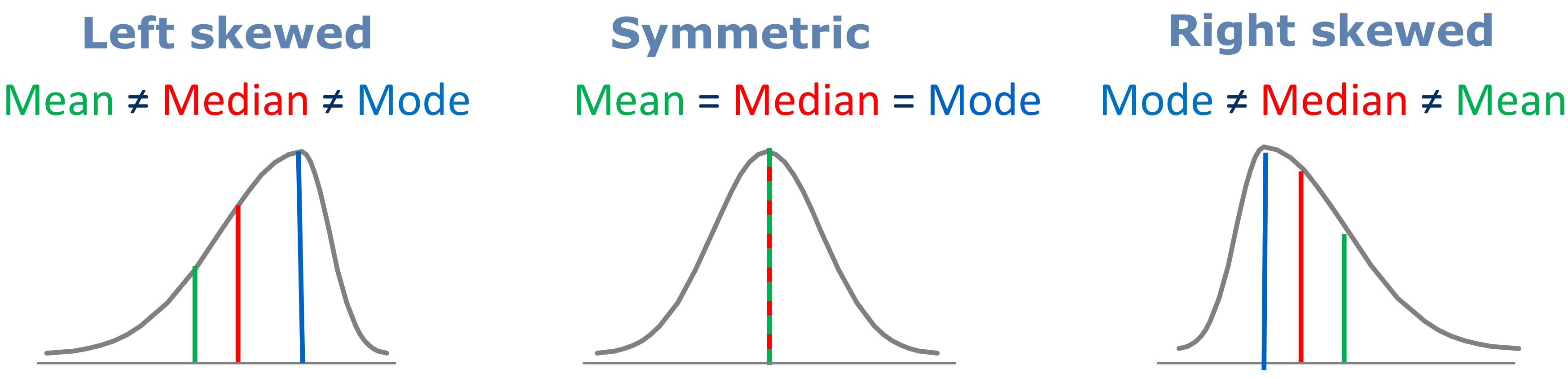
- **Compare to Q1, Q3 and IQR**

Stockholm University

NCT p. 91
- useful knowledge

Skewness – sv. snedhet

- If the distribution looks as if it has been "pulled out" to one side, we say the distribution is ***skewed*** (sv. *sned*)
- **Symmetric** if it equally distributed on both sides (non-skewed)



Stockholm University

Variable type & descriptive measures

Variables

Categorical

Numerical

Discrete

Continuous

Nominal- and ordinal scale

Location: mode

(for ordinal variables also median)

Variability: no. of observed categories/levels

(For ordinal variables Q1 and Q3 however not IQR!)

Interval and ratio scale

Location: mode, median (Q1 & Q3), mean

Variability: range, IQR, variance & standard deviation

Stockholm University

Next time …

… continue with descriptive statistics and discuss how to describe several variables at the same time (bivariate, multivariate):

- Tables and graphs etc.

Especially the relationship between two numerical variables

- Graphically
  - **scatter plots**
- Measures of Relationships between two variables:
  - **covariance** and **correlation coefficient**

Stockholm University

Exercise: CIY (check it yourself)

i	1	2	3	4	5	6	7	8	9	10	$\sum_i$
x_i	5	2	3	6	5	2	5	3	5	4	40

Mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(5 + 2 + \cdots + 4) = \frac{40}{10} = 4{,}0$

$x_i - \bar{x}$	1	-2	-1	2	1	-2	1	-1	1	0	0
$(x_i - \bar{x})^2$	1	4	1	4	1	4	1	1	1	0	18

Variance: $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{9}(1 + 4 + 1 + \cdots + 0) = \frac{18}{9} = 2{,}0$

Standard deviation: $s_x = \sqrt{s_x^2} = \sqrt{2{,}0} = 1{,}4142 \ldots$

Stockholm University

Exercise: CIY, cont.

i	1	2	3	4	5	6	7	8	9	10	$\sum_i$
x_i	5	2	3	6	5	2	5	3	5	4	40

Mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(5 + 2 + \cdots + 4) = \frac{40}{10} = 4{,}0$

x_i^2	25	4	9	36	25	4	25	9	25	16	178

alt. formula:

Variance: $s_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \frac{1}{9}(178 - 10 \cdot 4^2) = \frac{178-160}{9} = 2{,}0$

Mode $= 5$ Range $= Max - Min = 6 - 2 = 4$

Stockholm University

$n = 10 \Rightarrow n + 1 = 11$

Exercise: CIY, cont.

(i)	1	2	3	4	5	6	7	8	9	10	$\sum_i$
$x_{(i)}$	2	2	3	3	4	5	5	5	5	6	40

Median: 50% of $(\boldsymbol{n} + \mathbf{1}) = 5{,}5 \Rightarrow \quad a = \mathbf{5} \quad b = \mathbf{0,5}$

$$x_{(\mathbf{5})} + \mathbf{0,5} \cdot \left(x_{(\mathbf{6})} - x_{(\mathbf{5})}\right) = 4 + \mathbf{0,5}(5 - 4) = 4{,}5$$

Q1: 25% of $(\boldsymbol{n} + \mathbf{1}) = 2{,}75 \Rightarrow \quad a = \mathbf{2} \quad b = \mathbf{0,75}$

$$x_{(\mathbf{2})} + \mathbf{0,75} \cdot \left(x_{(\mathbf{3})} - x_{(\mathbf{2})}\right) = 2 + \mathbf{0,75}(3 - 2) = 2{,}75$$

Q3: 75% of $(\boldsymbol{n} + \mathbf{1}) = 8{,}25 \Rightarrow \quad a = \mathbf{8} \quad b = \mathbf{0,25}$

$$x_{(\mathbf{8})} + \mathbf{0,25} \cdot \left(x_{(\mathbf{9})} - x_{(\mathbf{8})}\right) = 5 + \mathbf{0,25}(5 - 5) = 5$$

Stockholm University