Statistics I

population: all persons/... we are looking at

sample: subset of the population

- simple random sampling
- convenience sampling (e.g. on the street)
- systematic sampling (regular intervals in ordered list)
- stratified sampling (groups based on some factor, random samples from every group)
- quota sampling (groups, convenience sampling inside each group)

continuos —— vs —— discrete

(e.g. height) (IR) → measure

data

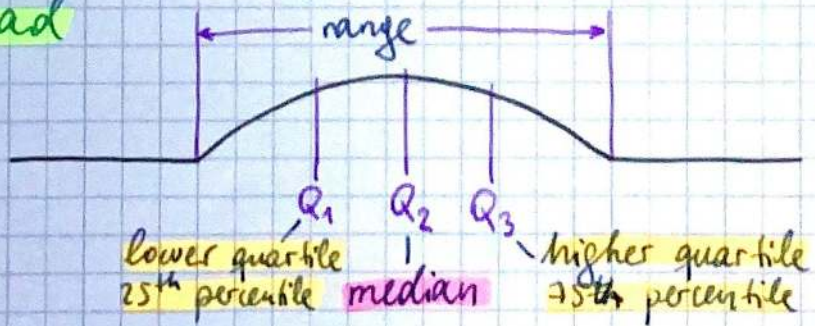
(number of sth.) (N) → count

outlier: extreme value

"center of data":

- mean $\bar{x}$
- median (in the middle), if two → calculate their mean
- mode (most frequent value), if two → bimodal

spread



$IQR = Q_3 - Q_1$

(interquartile range)

standard deviation σ

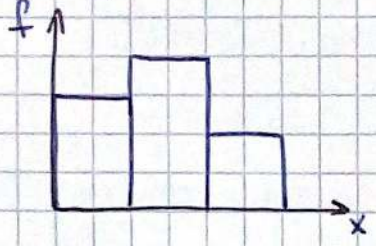
variance σ^2

outlier:

$x < Q_1 - 1{,}5 \cdot IQR$ or $x > Q_3 + 1{,}5 \cdot IQR$

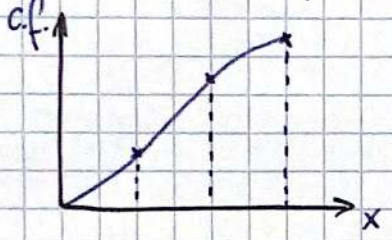
Statistics II

histogram



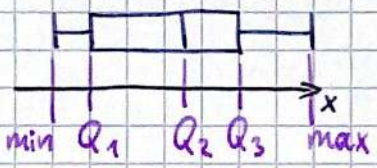
- bar = frequency of this range

cumulative frequency graph

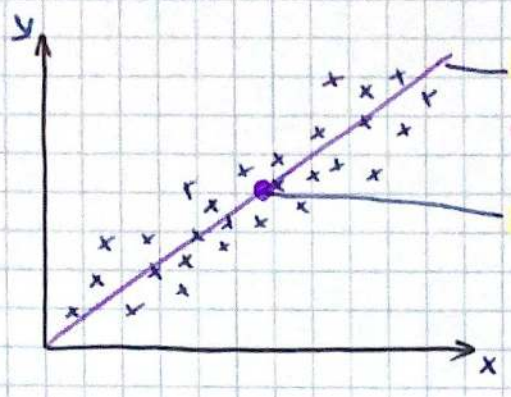


- section = frequency of this range AND ranges before = cumulative frequency

box-and-whiskers diagram



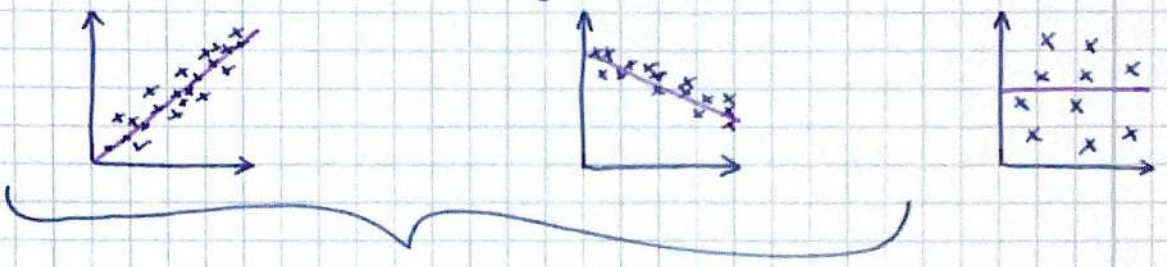
Correlation and regression



line of best fit always passes though the mean point $(\bar{x}, \bar{y})$

bivariate data = dataset with two variables

positive correlation | negative correlation | no correlation



points close to line: strong

points far from line: weak

→ Pearson's product-moment correlation r

-1 ...	0 ...	1 = r
strong negative	no	strong positive

spreadsheet → menu 4 1 2

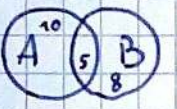
linear regression: menu 4 1 3

Probability

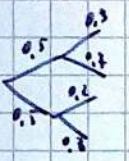
e.g. rolling a die → • outcomes: 1, 2, 3, 4, 5, 6 → sample space
• trial: rolling
• event: combination of outcomes

complementary events: $P(A') = 1 - P(A)$ $P(\bar{A})$

Venn diagram:



tree diagram:



sample space diagram:

coin 2 \ coin 1	front	back
front	0,25	0,25
back	0,25	0,25

→ or: sum of numbers on two dice

table of outcomes (it's rather a t.o.o.)

# pages \ type of book	A	B	C	D	total
0-60	..	...	...	...	..
60-100	..	..	..	..	..
100-	..	..	..	..	..
total	..	..	..	...	...

„Vierfeldertafel"

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

if mutually exclusive ($P(A \cap B) = 0$) $\Rightarrow P(A \cup B) = P(A) + P(B)$

for discrete random variables:

$$\sum_x P(X=x) = 1$$

$$E(X) = \sum_x x \cdot P(X=x)$$

↑ expected value

$$Var(X) = \sum_x x^2 \cdot P(X=x) - E(X)^2$$

binomial distribution: $X \sim B(n,p)$

- fixed # of trials
- two possible outcomes (success / failure)
- trials independent
- P(success) = const.

$$E(X) = n \cdot p$$

$$Var(X) = n \cdot p(1-p)$$

↑ variance σ^2

conditional probability:

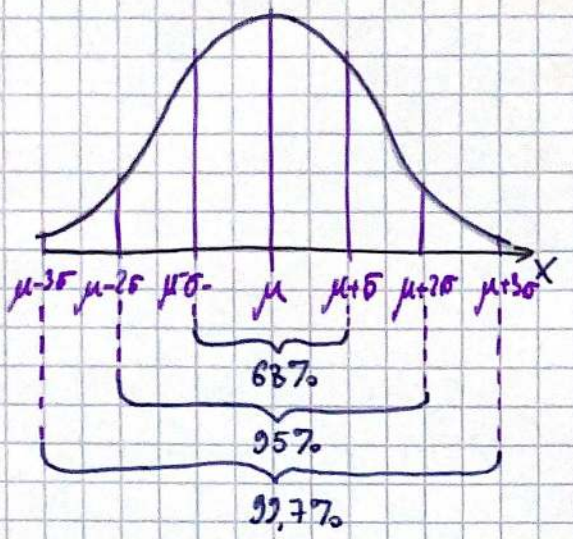
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

if $P(A|B) = P(A)$, then A & B are independent

or $P(A \cap B) = P(A) \cdot P(B)$

The normal distribution

$$X \sim N(\mu, \sigma^2)$$



standardization:

for a value x, the **z-value** gives the **number of standard deviations from the mean**:

$$z = \frac{x-\mu}{\sigma}$$

so for $X \sim N(\mu, \sigma^2)$ and $z = \frac{x-\mu}{\sigma}$, we get

$$Z \sim N(0, 1)$$

standard normal distribution