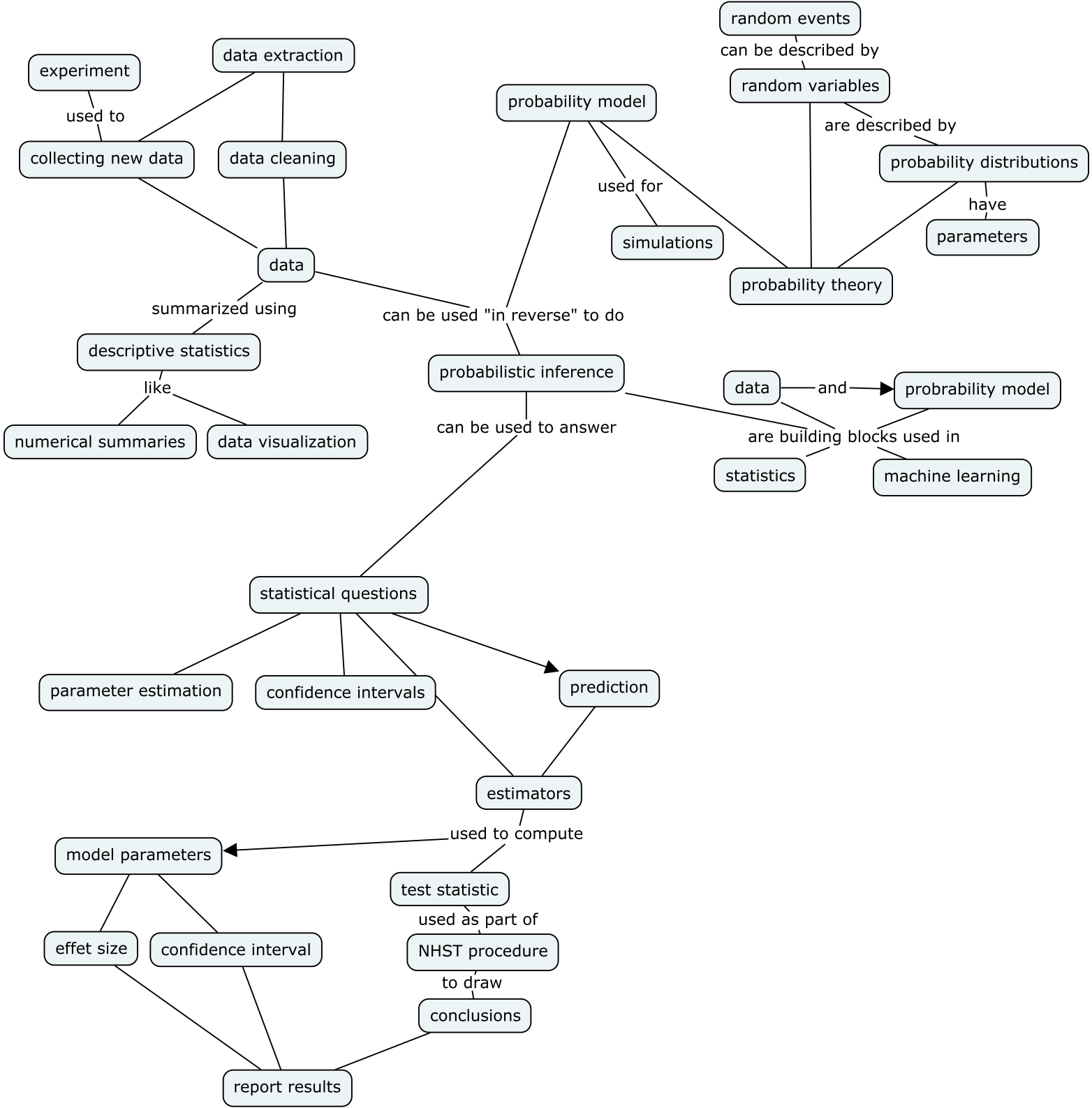
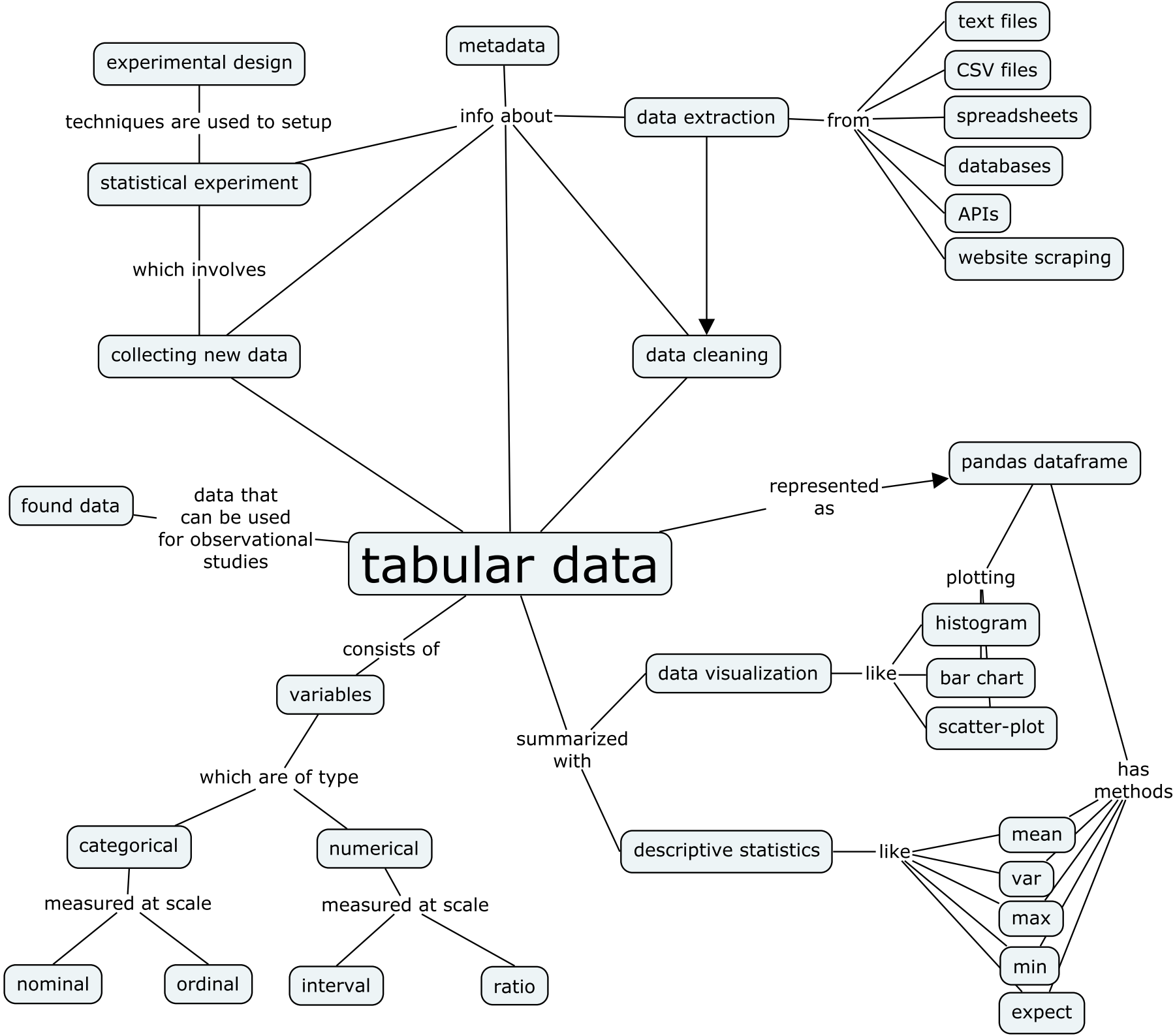
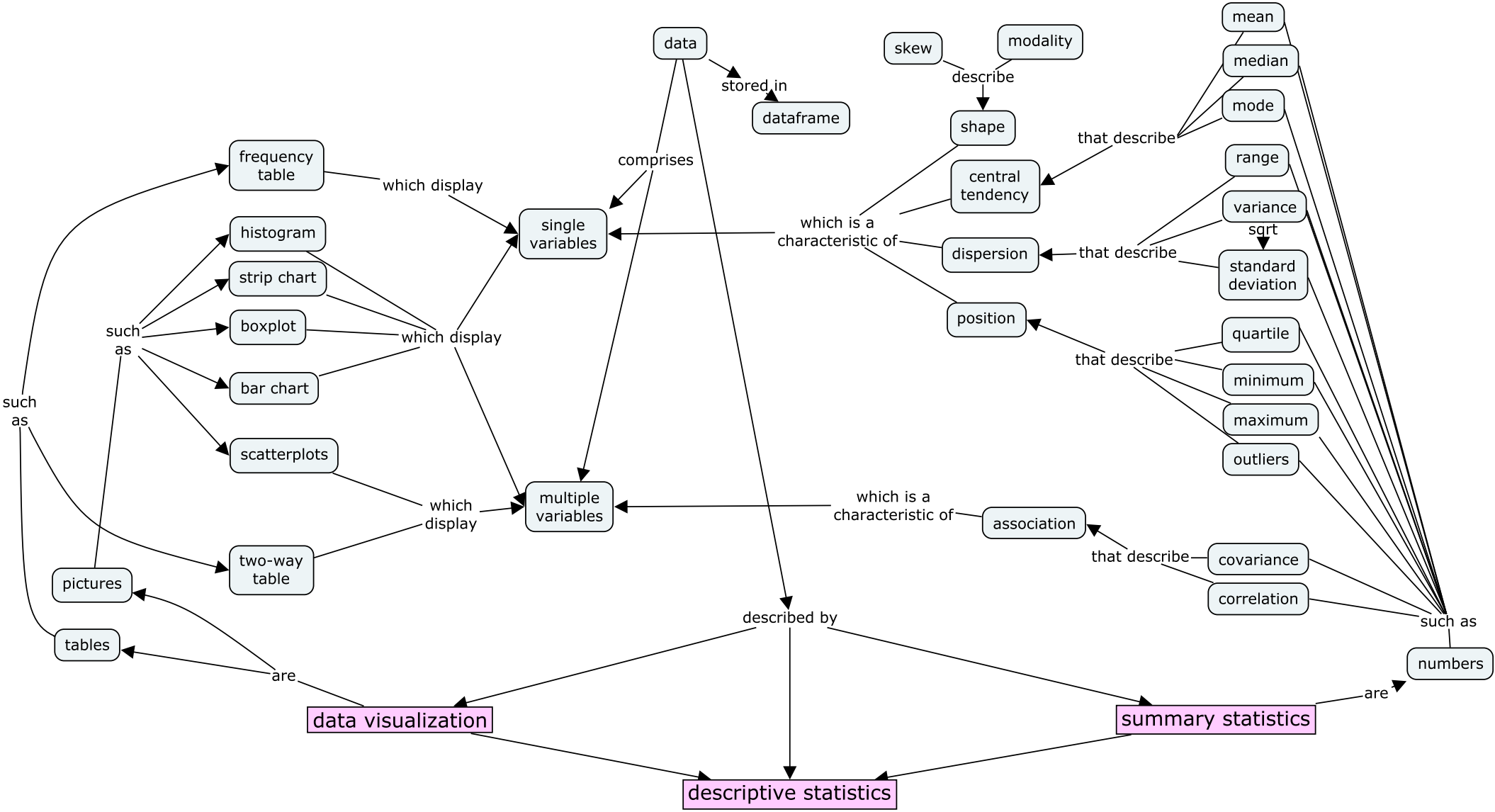
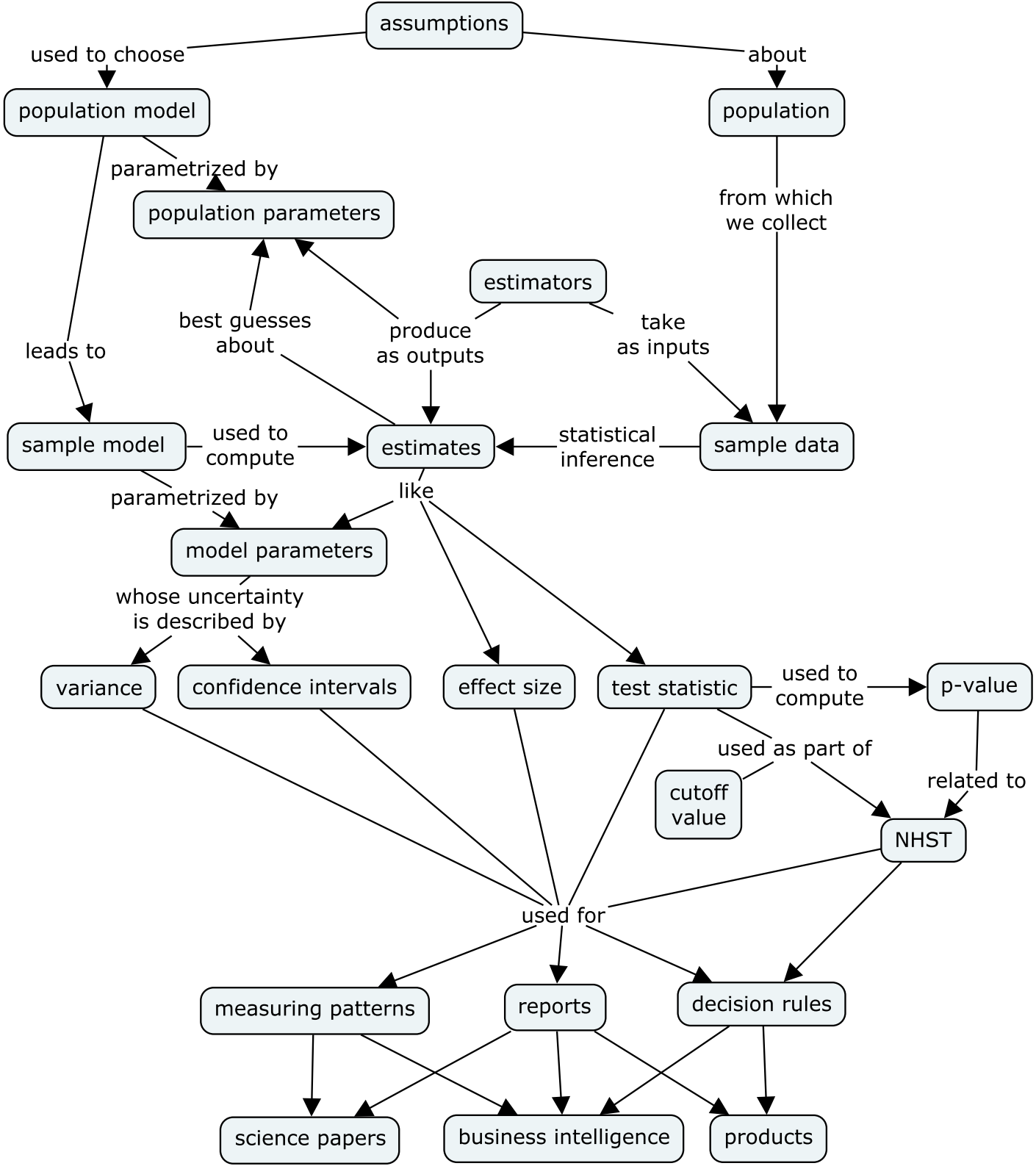
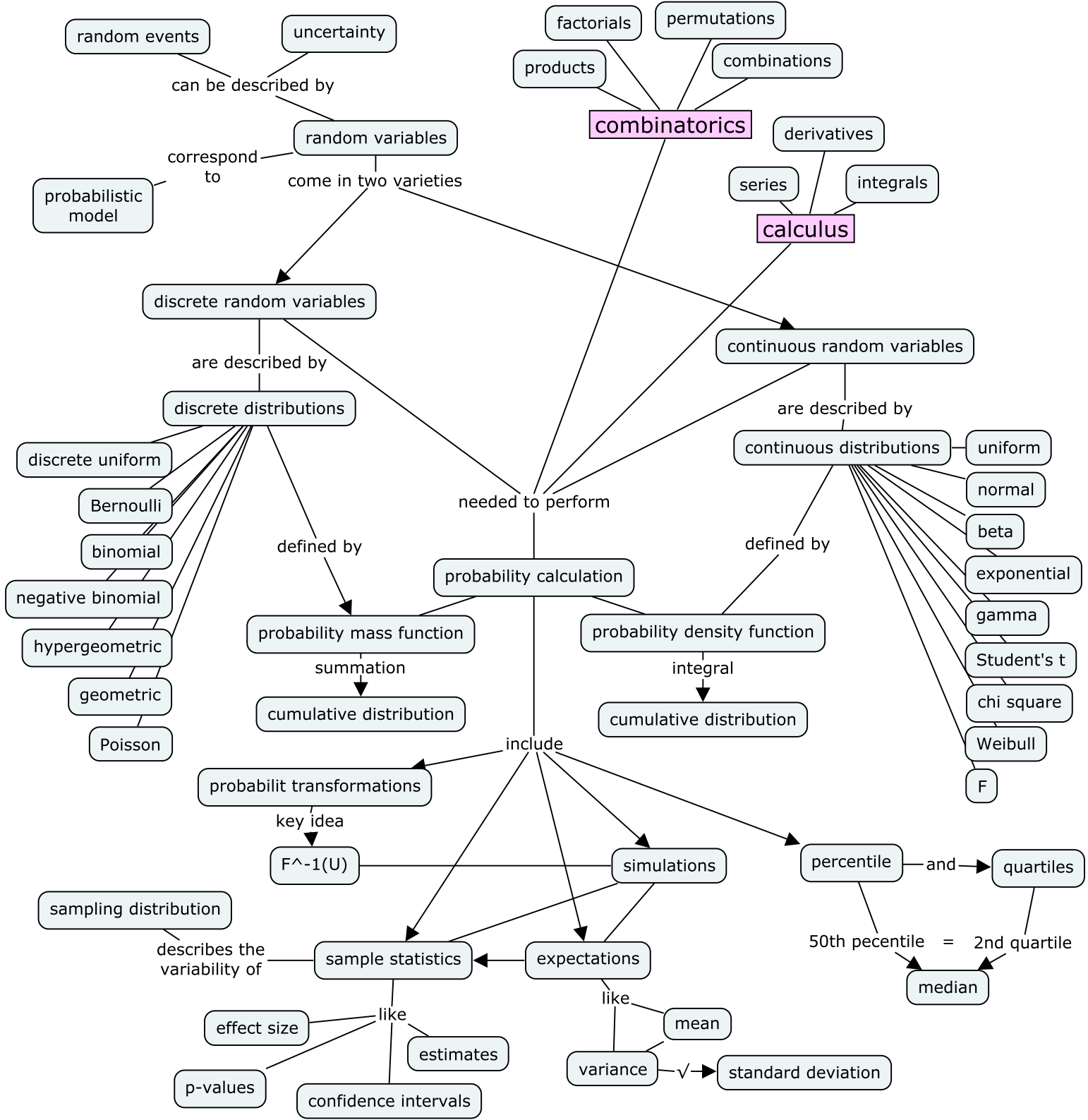
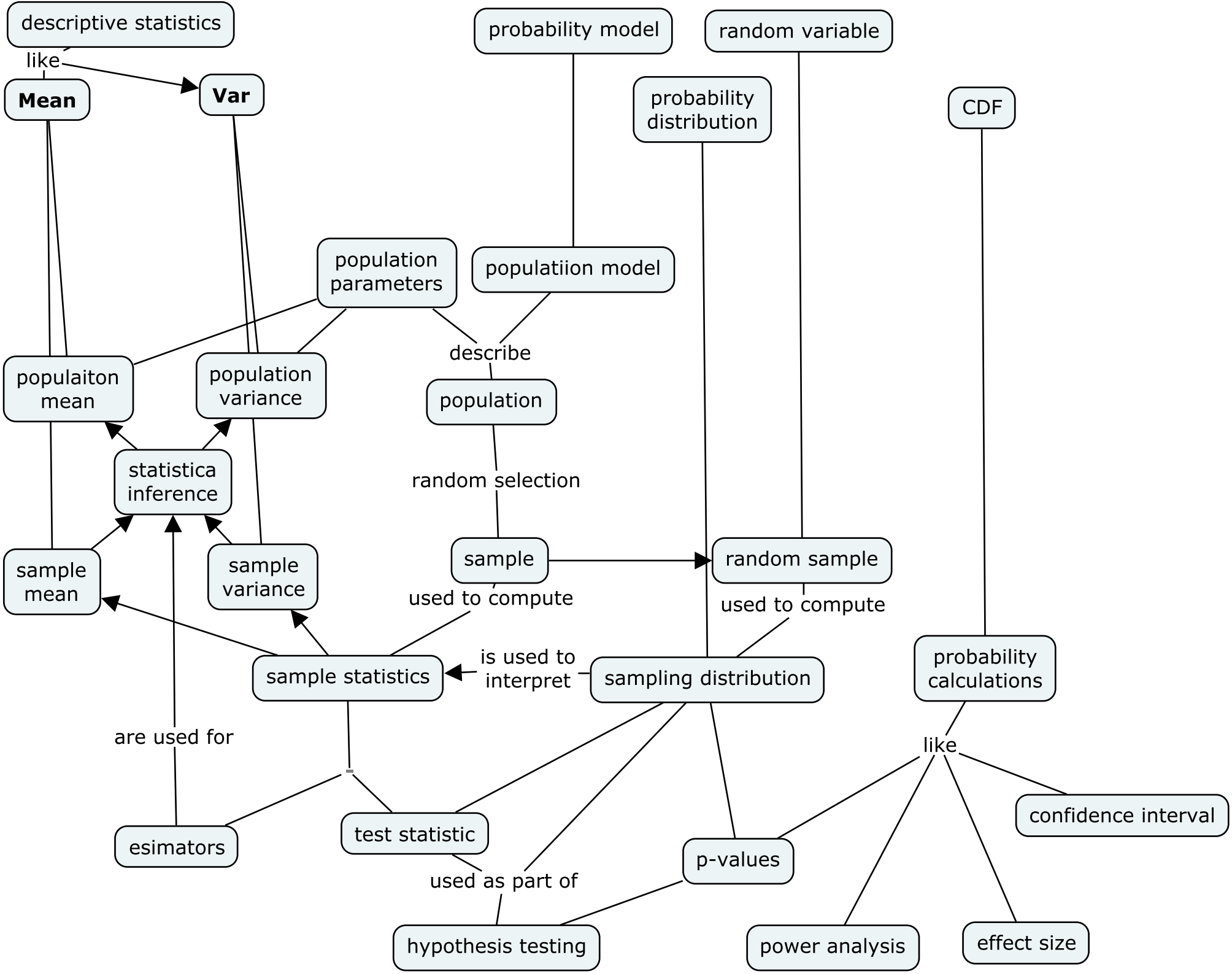

random events
can be described by
random variables
are described by
probability distributions
have
parameters
probability model
used for
simulations
probability theory
experiment
used to
data extraction
collecting new data
data cleaning
data
summarized using
descriptive statistics
like
numerical summaries
data visualization
can be used "in reverse" to do
probabilistic inference
can be used to answer
statistical questions
data
and
probrability model
are building blocks used in
statistics
machine learning
parameter estimation
confidence intervals
prediction
estimators
used to compute
model parameters
test statistic
used as part of
NHST procedure
to draw
conclusions
effet size
confidence interval
report results

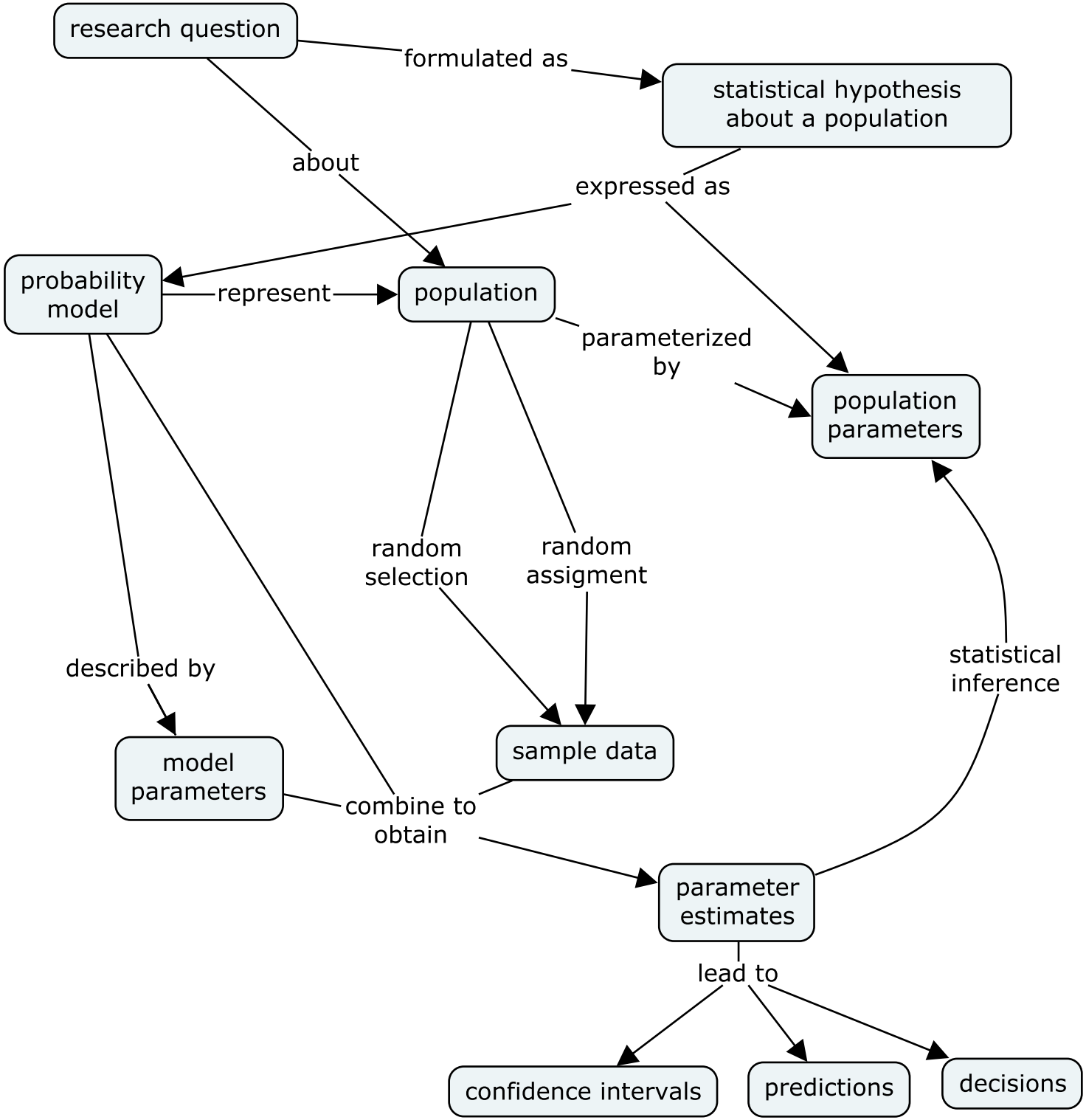

tabular data
metadata
info about
experimental design
techniques are used to setup
statistical experiment
which involves
collecting new data
data extraction
from
text files
CSV files
spreadsheets
databases
APIs
website scraping
data cleaning
found data
data that can be used for observational studies
represented as
pandas dataframe
plotting
has methods
consists of
variables
which are of type
categorical
numerical
measured at scale
measured at scale
nominal
ordinal
interval
ratio
summarized with
data visualization
like
histogram
bar chart
scatter-plot
descriptive statistics
like
mean
var
max
min
expect

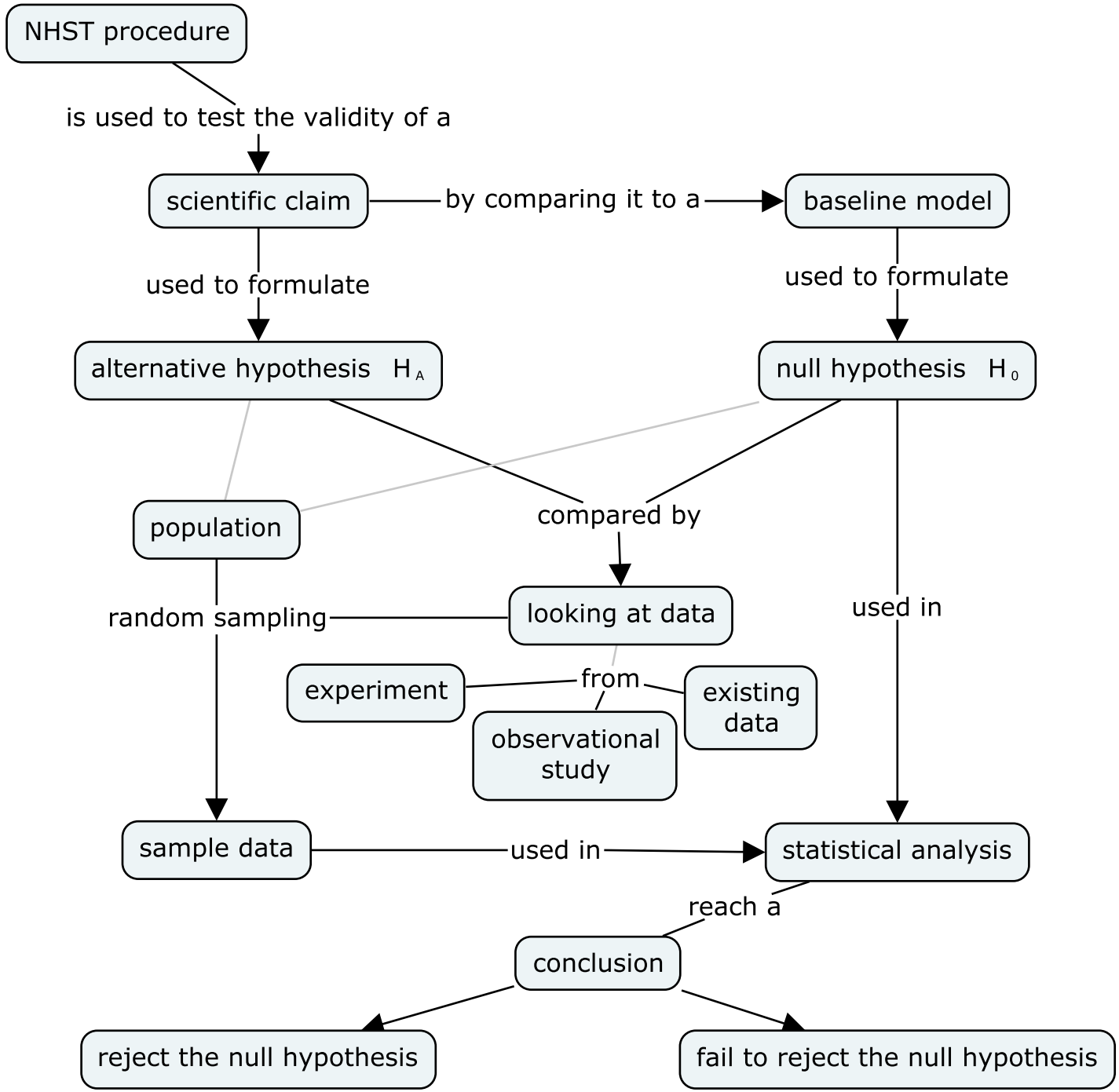

data
stored in
dataframe
comprises
skew
modality
describe
shape
central tendency
dispersion
position
which is a characteristic of
that describe
mean
median
mode
range
variance
sqrt
standard deviation
quartile
minimum
maximum
outliers
frequency table
which display
single variables
histogram
strip chart
boxplot
bar chart
scatterplots
such as
which display
multiple variables
which display
two-way table
such as
pictures
tables
are
association
covariance
correlation
such as
numbers
are
described by
data visualization
summary statistics
descriptive statistics

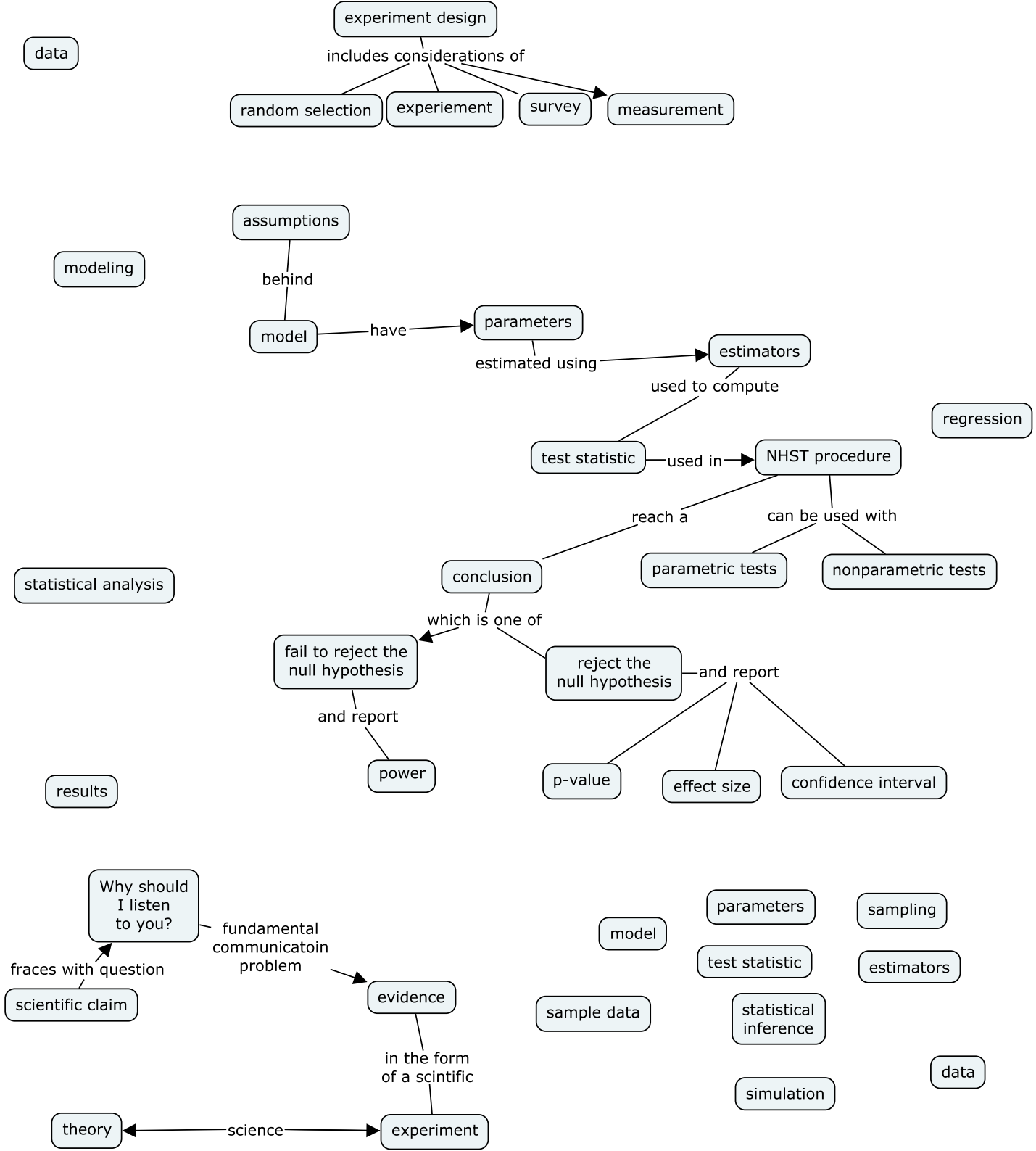

assumptions
used to choose
about
population model
population
parametrized by
population parameters
from which
we collect
estimators
best guesses
about
produce
as outputs
take
as inputs
leads to
sample model
used to
compute
estimates
statistical
inference
sample data
parametrized by
like
model parameters
whose uncertainty
is described by
variance
confidence intervals
effect size
test statistic
used to
compute
p-value
used as part of
cutoff
value
related to
NHST
used for
measuring patterns
reports
decision rules
science papers
business intelligence
products


random events
uncertainty
can be described by
random variables
correspond to
probabilistic model
come in two varieties
factorials
permutations
products
combinations
combinatorics
derivatives
series
integrals
calculus
discrete random variables
continuous random variables
are described by
discrete distributions
discrete uniform
Bernoulli
binomial
negative binomial
hypergeometric
geometric
Poisson
defined by
probability mass function
summation
cumulative distribution
needed to perform
probability calculation
are described by
continuous distributions
uniform
normal
beta
exponential
gamma
Student's t
chi square
Weibull
F
defined by
probability density function
integral
cumulative distribution
include
probabilit transformations
key idea
F^-1(U)
simulations
percentile
and
quartiles
50th pecentile
=
2nd quartile
median
sampling distribution
describes the variability of
sample statistics
expectations
like
effect size
estimates
p-values
confidence intervals
like
mean
variance
√
standard deviation


descriptive statistics
like
Mean
Var
probability model
random variable
probability distribution
CDF
population parameters
populatiion model
describe
populaiton mean
population variance
population
statistica inference
random selection
sample mean
sample variance
sample
random sample
used to compute
used to compute
sample statistics
is used to interpret
sampling distribution
probability calculations
are used for
like
esimators
test statistic
confidence interval
p-values
used as part of
hypothesis testing
power analysis
effect size


research question
formulated as
statistical hypothesis about a population
about
expressed as
probability model
represent
population
parameterized by
population parameters
random selection
random assigment
described by
statistical inference
model parameters
sample data
combine to obtain
parameter estimates
lead to
confidence intervals
predictions
decisions


NHST procedure
is used to test the validity of a
scientific claim
by comparing it to a
baseline model
used to formulate
used to formulate
alternative hypothesis H_A
null hypothesis H_0
population
compared by
random sampling
looking at data
used in
experiment
from
existing data
observational study
sample data
used in
statistical analysis
reach a
conclusion
reject the null hypothesis
fail to reject the null hypothesis


data
experiment design
includes considerations of
random selection
experiement
survey
measurement
modeling
assumptions
behind
model
have
parameters
estimated using
estimators
used to compute
regression
test statistic
used in
NHST procedure
reach a
can be used with
parametric tests
nonparametric tests
statistical analysis
conclusion
which is one of
fail to reject the null hypothesis
reject the null hypothesis
and report
and report
power
p-value
effect size
confidence interval
results
Why should I listen to you?
fundamental communicatoin problem
fraces with question
scientific claim
evidence
in the form of a scintific
theory
science
experiment
model
parameters
sampling
test statistic
estimators
sample data
statistical inference
data
simulation