UNIVERSITÄT BERN

Computer Aided Archaeology

07 - Basic Statistics

Martin Hinz

Institut für Archäologische Wissenschaften, Universität Bern

10/11/21

Flavours of statistics

Descriptive statistics

- Summary and description of data by using parameters (mean, standard deviation etc.)

(graphical display)

- Summary and description of data by using graphs (bar charts, pie charts etc.)
- Useful for pattern detection and description, therefore intermediate position

Explorative statistics

- Summary and description of data for pattern detection (e.g. correspondence analysis)

Statistical inference or statistical induction

- testing of hypothesis on data (e.g. chi-squared test)

Sample and Population

Population:

- Amount of all items of relevance for an analysis.

Sample

- Selection of items on basis of certain criteria (e.g. representativity) which
- will be analysed instead of the population

Example opinion poll

- Population: all federal citizens who have a meaning
- Sample: the citizens who are polled by the polling organization

complete record of all the values ↔ *sampling*

In archaeology only sampling is possible! The population can never be investigated!

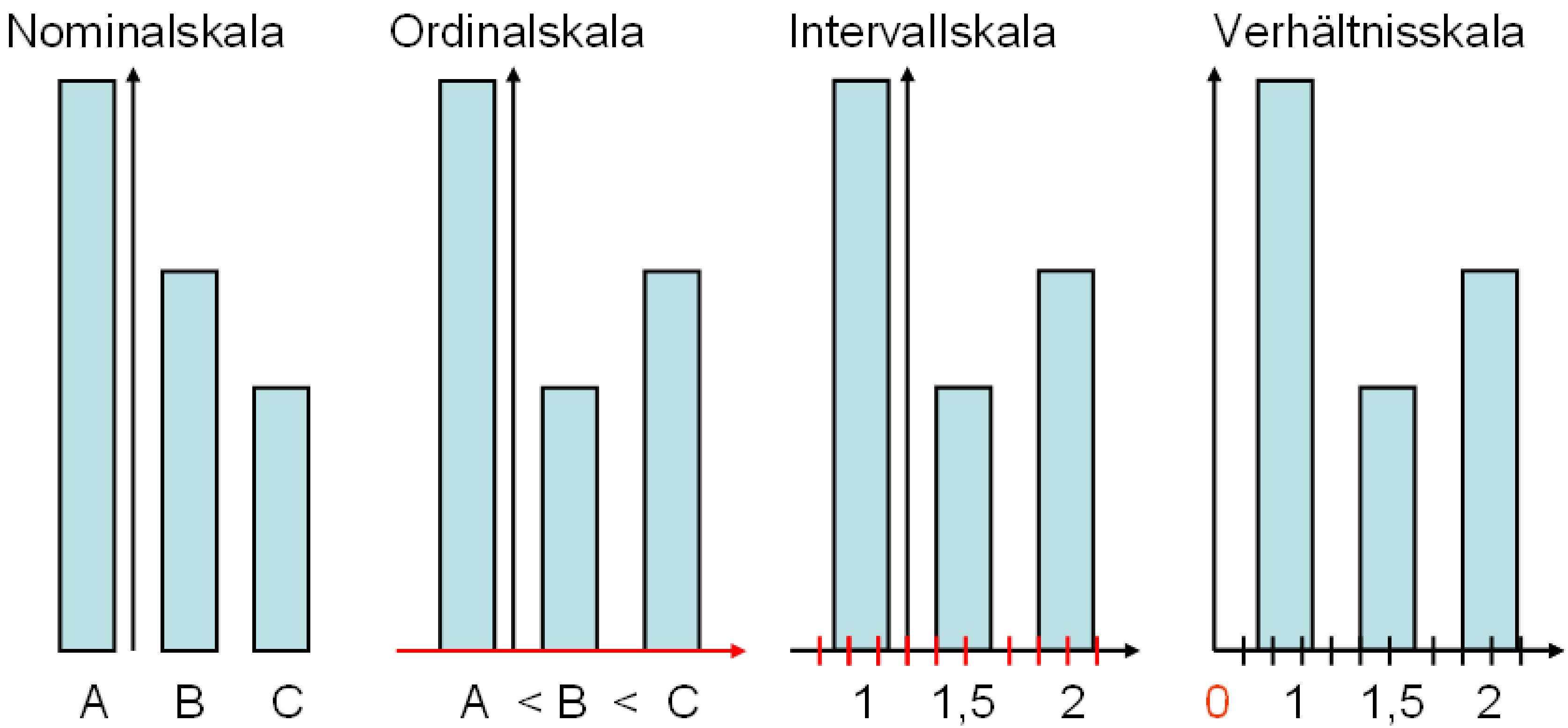
Levels of measurement

- nominal:
  - Categories which do not have a defined relationship among each other, only counting is possible (e.g. sex)
- ordinal:
  - Categories which are comparable and differ from each other in their characteristic [size/power/intensity]; their rank is determinable (e.g. preservation conditions – bad, medium, good)
- metric:
  - Variable has a defined system of measurement, all calculations are possible. To distinguish are

1. interval: The variable has an arbitrary choosen neutral point (°C)
2. ratio: The variable has an absolute neutral point (°K)

- Sometimes also used: absolut scale
  - counts (number of inhabitans)

Levels of measurement



Levels of measurement

scale	Meaningful statements	Examples
nominal	equality, inequality	Telephon numbers, illnesses, ceramic types
ordinal	bigger-smaller-relationship	Wind forces, academic ranks, classes of wealth, stratigraphic relations
interval	Equality of differences	Temperature in °C, calender age
ratio	Equality of ratios	Measurement of lengths, weight, height of a vessel

after Bortz 2005

Inductive statistics or statistical inference

Is used to draw conclusions about (unknown) parameters of the population on basis of a sample
The results are always statistical ;-)

i.e. all statements are true with a certain probability but could be also false with a certain probability

The basis of statistical inference is probability theory (stochastic)

Statistical hypothesis testing

Validation of an assumption about the population

A assumption (hypothesis) about the population is made and than its probability is checked against the sample.

Usual questions:

How probable is it that two or more samples descend from the different/the same population?

(eg. Is the custom of grave goods for man and women so different that two different social groups are visible?)

How probable is it that a given sample descend from a population with certain parameters?

(Is the amount of grave goods random or is a pattern visible?)

Null hypothesis [1]

Validation through falsification

In statistical tests most of the times not the statement is tested which one expects to be true but one tries to disprove the statement which one expects to be wrong: the null hypothesis.

This hypothesis states mostly, that a association do not exists or that there is no differences between the samples and the distribution of the observations is by chance.

Example: Is the composition of grave goods different between male and female deceased?

H_0: The compositionisthe same

H_1: The composition is different

Reason

1. It is (logical) easier to prove, that a statement is wrong (falsify) then to prove that a statement is true (verify).
2. Most of the times it is easier to formulate a null hypothesis (How exactly is the composition different?). It doesn't make a assumption about how the character of a association/difference exactly is.

Null hypothesis [2]

„Workflow“ of a statistical test

Construction of a alternative hypothesis:

The composition of the grave goods is different between male and female deceased.

Construction of the null hypothesis:

The composition of the grave goods is the same in male and female burials.

Test of the null hypothesis

If the result of the test is significant:

Rejection of the null hypothesis, choice of the alternativ hypothesis. The composition of the grave goods is different between male and female deceased. If the result of the test is not significant:

The null hypothesis could not be rejected.

We can not say if the composition of the grave goods is different between male and female deceased or not!

Stat. Significance

How true is true?

Statistical significance is effectively a measurement how probable a error is.

On basis of the significance the null hypothesis will be rejected and the alternative hypothesis will be choosen … or not.

There are classic boundary values for significance (significance levels):

0.05: significant, with 95% probability the decision is right.

0.01: very significant, with 99% probability the decision is right.

0.001: highly significant, with 99,9% probability the decision is right.

Often named with p-value or α.

Nonparametric tests

Parametric vs. nonparametric

Parametric: The distribution of the values have to be in a certain form (e.g. normal distribution); assumptions about the distribution of the population are needed

non-parametric: no assumptions about the distribution of the sample and the population are needed

Nonparametric tests, advantages and disadvantages:

Advantage: Also appropriate if no statements about the distribution are possible or the distribution fits no for parametric tests.

Also smaller samples are possible.

Disadvantages: Tests have general a lesser power.

χ^2 test [1]

Possible Questions

Do settlements tend to be situated on rather good soil or is the distribution random?

Conclusions about settlement behaviour and economy would be possible

Do older individuals have more shoe-last celt as grave goods than younger?

If shoe-last celt would be signs of social rank than this situation would make conclusions possible about heredity or acquisition of social rank during life time.

Tests for nominal scaled variables are possible!

Therefore of particular value for archaeology because we have often to deal with such data.

Independent – dependent variable

Independent Variable:

- The assumed cause of a relationship

Dependent variable:

- The assumed effect of the independent variable in a relationship

example:

- Number of pearls in a grave (Dependent) vs.
- sex of the deceased (independent)
- Hypothesis: The number of pearls in a grave depends on the sex of the deceased

Can (have to be) not always to be defined

- e.g.: volume and height of a vessel...

χ^2 test [2]

Test for independence of two distributions

Requirements: at least 1 nominal scaled variable (one sample case) and 1 nominal scaled grouping variable (two sample case)

Procedure with one sample: observed values are compared with expected values given a certain distribution, no expected value should be < 5; n should be > 50

Procedure with two samples: observed values of both distributions are compared with expected values if the samples would be even distributed, no expected value should be < 5; n should be > 50

Test statistics: χ^2

Significance depend on degree of freedom (df)

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	female	total
cremation			201
inhumation			197
total	216	182	398

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	female	total
cremation	123		201
inhumation			197
total	216	182	398

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	female	total
cremation	123	78	201
inhumation	93	104	197
total	216	182	398

df=1: if one value is chosen all other can be calculated with the help of the margins

(number of columns – 1)*(number of rows – 1)

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	female	uncertain	total
cremation				201
inhumation				197
total	196	179	23	398

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	**female**	**uncertain**	**total**
cremation		78		201
inhumation				197
total	196	179	23	398

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	**female**	**uncertain**	**total**
cremation	113	78		201
inhumation				197
total	196	179	23	398

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	female	uncertain	total
cremation	113	78	10	201
inhumation	83	101	13	197
total	196	179	23	398

df=2: if two values are chosen all other can be calculated with the help of the margins

(number of columns – 1)*(number of rows – 1)

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	**female**	**uncertain**	**total**
cremation				201
inhumation				197
other				30
total	201	187	40	428

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	**female**	**uncertain**	**total**
cremation		78		201
inhumation	83		13	197
other		8		30
total	201	187	40	428

Excursus degree of freedom

Number of slots free to vary given the margin sums

	male	female	uncertain	total
cremation	113	78	10	201
inhumation	83	101	13	197
other	5	8	17	30
total	201	187	40	428

χ^2 test [3]

Test for one sample (example after Shennan)

Numbers of neolithic settlements by soil type in eastern france

Soil type	Number of settlements
Rendzina	26
Alluvial	9
Brown earth	18
total	53

Question: Is there a significant preference for a soil type?

We calculate two versions:

1. even distributed

2. even distributed with consideration of the proportion of the soil types on the total area

χ^2 test [4]

Version 1: even distributed

Soil type	Number of settlements	Proportion of soil type	expected number of settlements
Rendzina	26	1/3	17.6667
Alluvial	9	1/3	17.6667
Brown earth	18	1/3	17.6667
total	53	1	53

H_0: The settlements are evenly distributed on all soil types.

H_1: The settlements are **not** evenly distributed on all soil types.

χ^2 test [5]

Version 1: even distributed

Soil type	Number of settlements	Proportion of soil type	expected number of settlements
Rendzina	26	1/3	17.6667
Alluvial	9	1/3	17.6667
Brown earth	18	1/3	17.6667
total	53	1	53

Formula for χ^2:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

O_i: number of **observed** cases

E_i: number of **expected** cases

χ^2: symbol for the test statistic chi-squared

χ^2 test [6]

Procedure: Calculation of the X²-value

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Soil type	Number of observed cases	Number of expected cases	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
Rendzina	26	17.6667	8.3333	69,4444	3.9308
Alluvial	9	17.6667	-8.6667	75,1117	4.2516
Brown earth	18	17.6667	0.3333	0.1111	0.0063
total	53	53			**8.18868**

Look up in a table (e.g. Shennan): Df=2 (2 colums (expected, observed), 3 categories)

Level of significance: 0.05

Boundary value: 5.99145

Significant result: The distribution is uneven!

χ^2 test [7]

Version 2: even distributed with consideration of the proportion of the soil types on the total area

Soil type	Number of settlements	Proportion of soil type	expected number of settlements
Rendzina	26	32%	16.69
Alluvial	9	25%	13.25
Brown earth	18	34%	22.79
total	53	1	53

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

χ^2 test [8]

Procedure: Calculation of the X²-value

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Soil type	Number of observed cases	Number of expected cases	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
Rendzina	26	16.69	9.04	81.7216	4.8185
Alluvial	9	13.25	-4.25	18.0625	1.1363
Brown earth	18	22.79	-4.79	22.9441	1.007
total	53	53			**7.1885**

Look up in a table (e.g. Shennan): Df=2 (2 colums (expected, observed), 3 categories)

Level of significance: 0.05

Boundary value: 5.99145

Significant result: The distribution is uneven also if we consider the proportions of the soil types!

χ^2 test [10]

Two sample case (Test for independence)

(example after Hinz, beautified)

Comparison of amber in graves and settlements

Classic 2x2 situation

Type of site	amber		total
	+	-	
settlement	6	18	24
grave	132	44	176
total	138	62	200

Is amber primary a grave good?

df=1

Level of significance = 0.05

χ^2 test [11]

Procedure: Calculation of the expected values

Multiply the margins and divide the result by the total number

Type of site	**amber**		**total**
	+	-	
settlement	24*138/200 = 16.56	24*62/200=7.44	24
grave	138*176/200=121.44	62*176/200=54,56	176
total	138	62	200

χ^2 test [12]

Procedure: Calculation of the expected values

Multiply the margins and divide the result by the total number

Type of site	amber		total
	+	-	
settlement	O=6 vs. E=16.56	O=18 vs. E=7.44	24
grave	O=132 vs. E=121.44	O=44 vs. E=54.56	176
total	138	62	200

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

χ^2 test [13]

Procedure: Calculation of the expected values

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Type of site	**amber**		**total**
	+	-	
settlement	(6-16.56)^2/16.56=6.73	(18-7.44)^2/7.44=14.99	24
grave	(132-121.44)^2/121.44=0.92	(44-54.56)^2/54.56=2.04	176
total	138	62	200

Is amber primary a grave good?

Df=1, Level of significance = 0.05;

χ^2=24,68; boundary value (df=1 and p=0.05): 3.84146

The difference in the distribution is significantly not by chance. Both variables are associated!

χ^2 test in Libre Office Calc

we need:

- Command CHITEST
- table of observed values
- table of expected values (from the marginal sums)

UNIVERSITÄT BERN

Any questions?

You might find the course material (including the presentations) at

https://berncodalab.github.io/caa

You can contact me at

martin.hinz@iaw.unibe.ch