Probability and statistical hypothesis testing

Holger Diessel

holger.diessel@uni-jena.de

Probability

Two reasons why probability is important for the analysis of linguistic data:

- Joint and conditional probabilities are used to analyze corpus data
- Probability plays an important role in statistical hypothesis testing

Simple probability:

If you toss a dice with six number (i.e. 1,2,3,4,5,6) what is the probability that you will toss a 6?

$$P(6) = 1/6 = 0.1666$$

Probability

- Probability values range from 0 to 1.
- Adding all probabilities of the sample yields 1.
- If two events are independent, the probability is the sum of their individual probabilities.
- Two events A and B are independent if knowing that the occurrence of A does not change the probability of the occurrence of B.

The law of large numbers

When an experiment is conducted many times, the relative frequency (empirical probability) of an event can be expected to be close to the theoretical probability of the event. This approximation will improve as the number of replications is increased.

Joint probability

$$P(A,B) = P(A) \times P(B)$$

$$P(5,6) = (0.166) \times (0.166)$$

$$= 0.0277$$

Conditional probability

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

Conditional probability

In a corpus including 12.000 nouns and 3.500 adjectives, 2.000 adjectives precede a noun.

(1) What is the likelihood that a noun occurs after an adjective?

(2) What is the likelihood that an adjective precedes a noun?

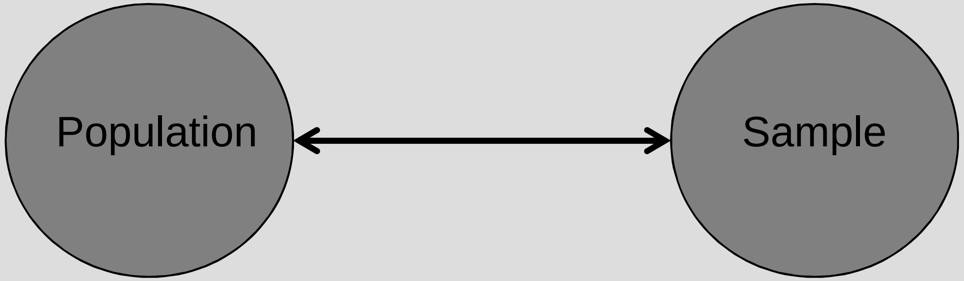
Conditional probability

$$P(ADJ \mid N) = \frac{P(ADJ \cap N)}{P(N)}$$

$$P(ADJ \mid N) = \frac{P(2000)}{P(12000)} = 0.1666$$

$$P(N \mid ADJ) = \frac{P(2000)}{P(3500)} = 0.5714$$

Statistical hypothesis testing



A researcher has collected 56 relative clauses from a corpus. 27 relative clauses are attached to an animate head noun, 29 relative clauses are attached to an inanimate head noun. 33 relative clauses are subject relatives, 23 relative clauses are object relatives:

(1) Der Mann, der dich gesehen hat. 21
(2) Der Mann, den du gesehen hast. 6
(3) Der Film, der dir gefallen hat. 12
(4) Der Film, den du gesehen hast. 17

Statistical hypothesis testing

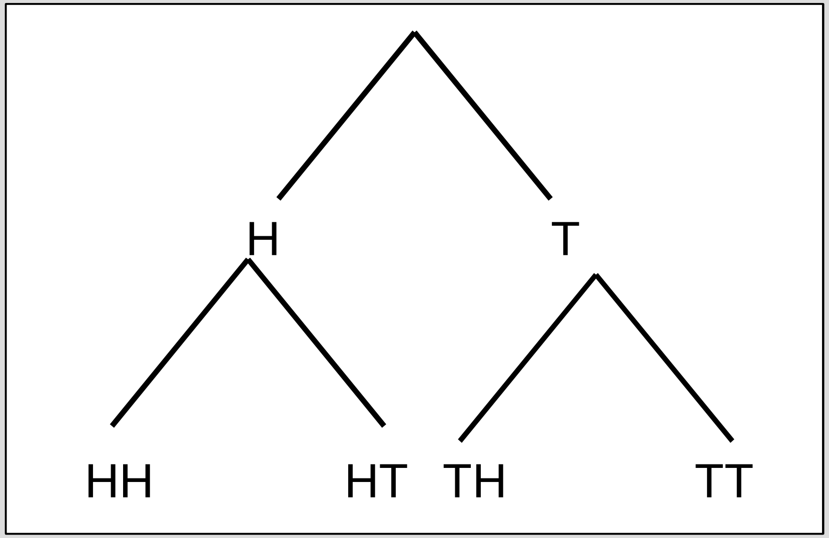
	Animate head	Inanimate head
Subject	21	12
Object	6	17

- Null hypothesis: There is no relationship between the animacy of the head and the syntactic role of the relative pronoun.
- Alternative hypothesis: There is a relationship between the animacy of the head and the syntactic role of the relative pronoun.

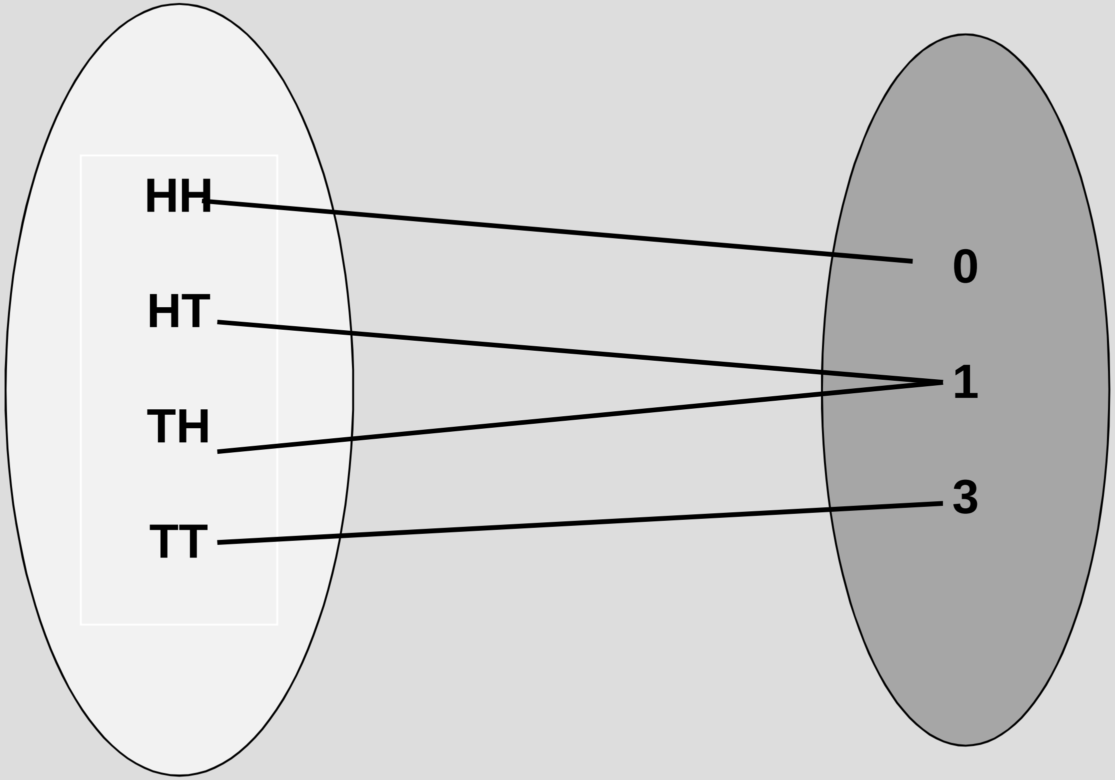
Probability distribution

What is the probability that you get two heads if you toss a coin twice?

0 heads	=	HH	25%
1 head	=	HT + TH	50%
2 heads	=	TT	25%



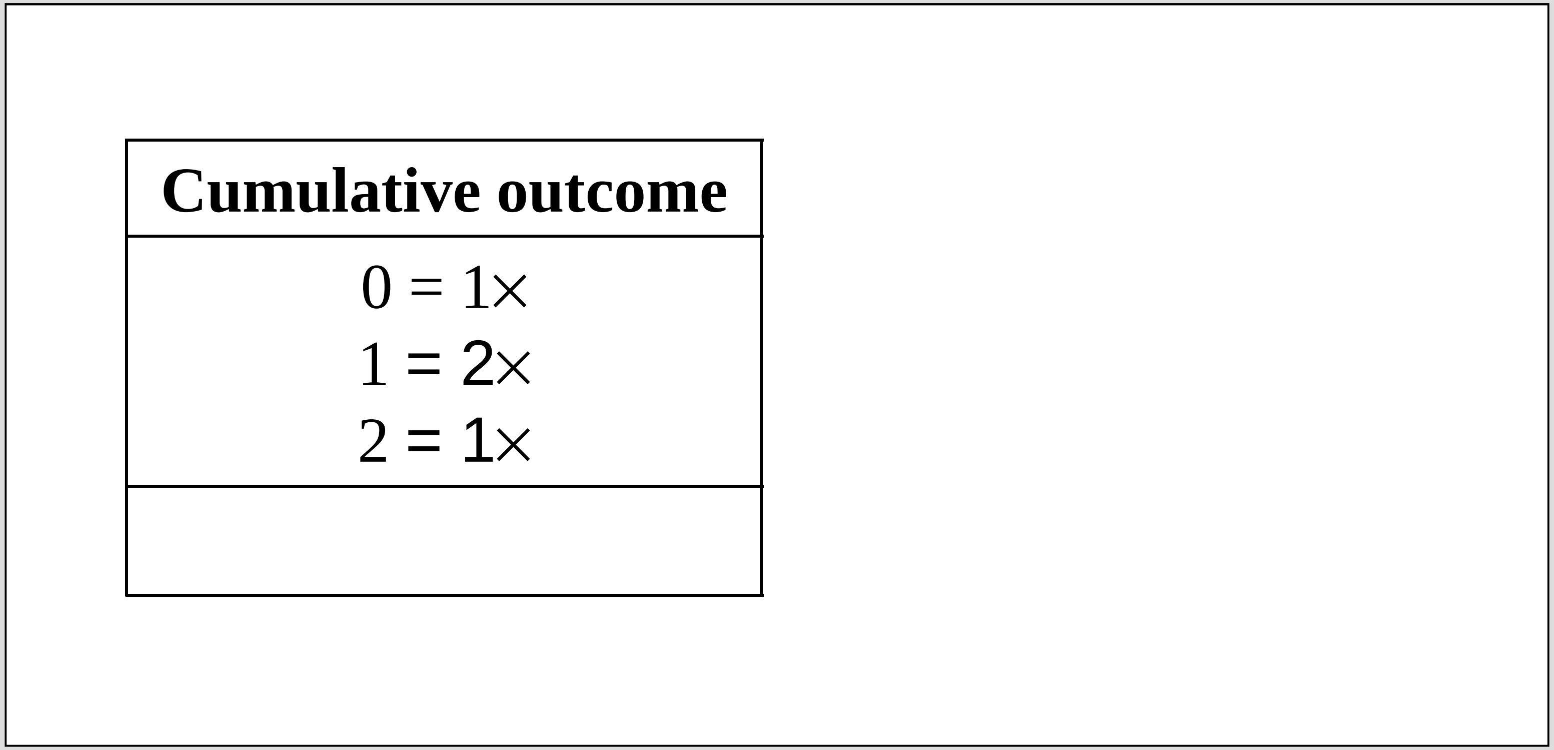
Probability distribution



Sample space

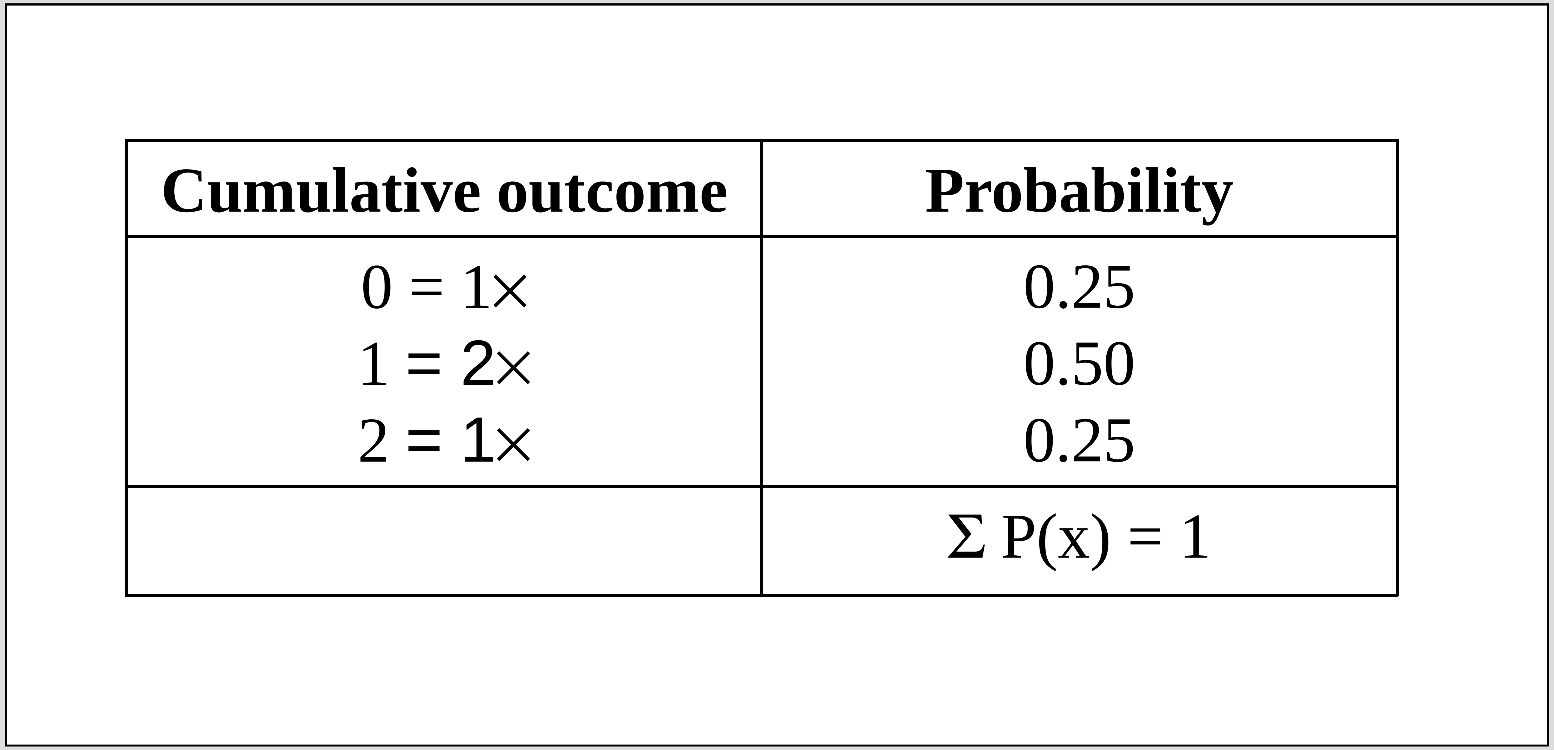
Random variable

Probability distribution



Cumulative outcome
0 = 1× 1 = 2× 2 = 1×

Probability distribution



Cumulative outcome	Probability
0 = 1× 1 = 2× 2 = 1×	0.25 0.50 0.25
	Σ P(x) = 1

Binomial distribution

Bernoulli trail:

- two possible outcomes on each trail
- the outcomes are independent of each other
- the probability ratio is constant across trails

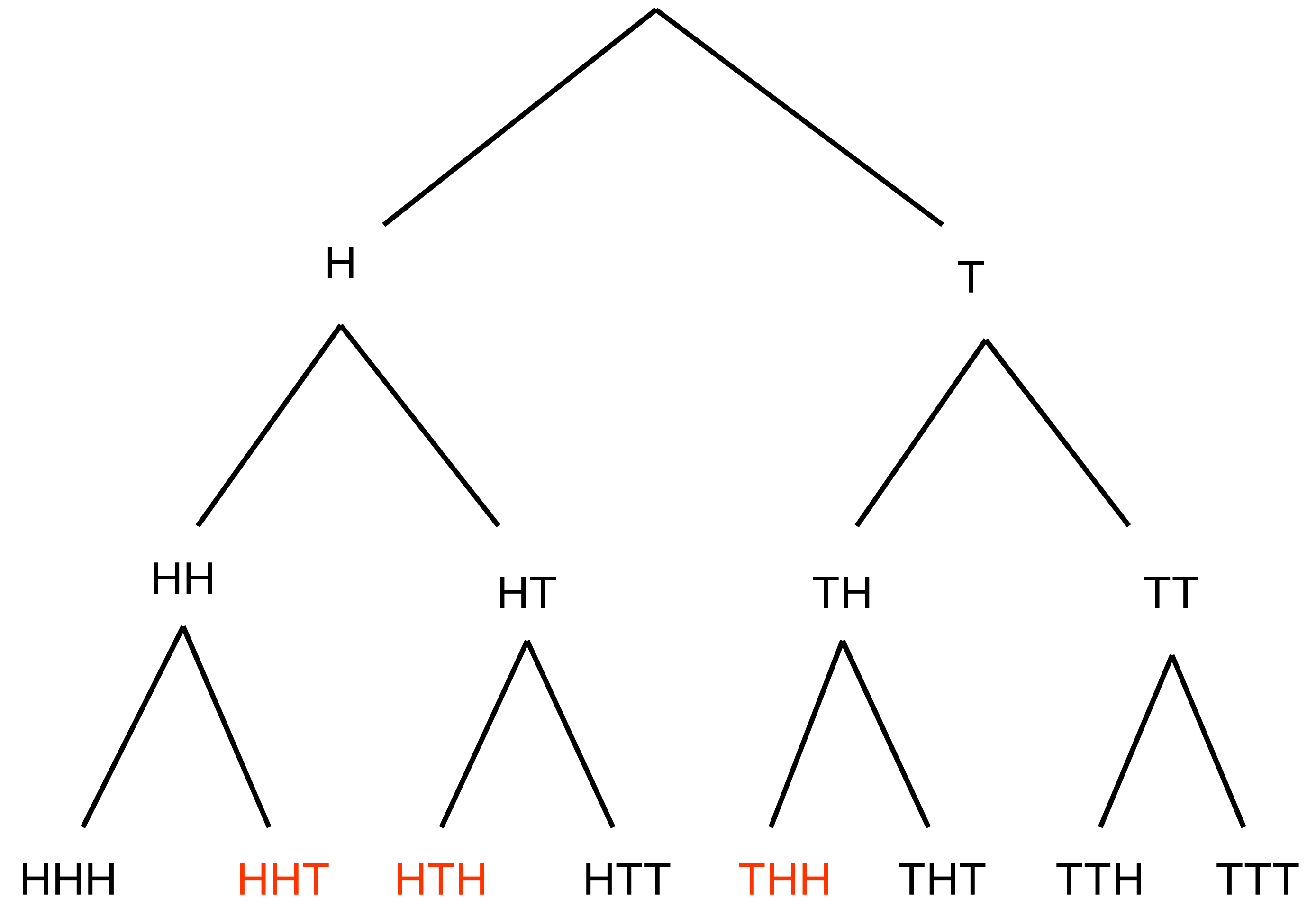
Binomial distribution

The binomial distribution has the following properties:

- It is based on categorical / nominal data.
- There are exactly two outcomes for each trail.
- All trials are independent.
- The probability of the outcomes is the same for each trail.
- A sequence of Bernoulli trails gives us the binomial distribution.

Binomial distribution

A coin is tossed three times. What is the probability of obtaining two heads?


H
T
HH
HT
TH
TT
HHH
HHT
HTH
HTT
THH
THT
TTH
TTT

Binomial distribution

Sample space:

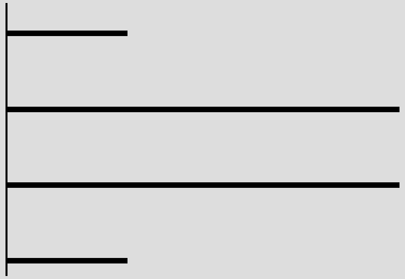
HHH
HHT
HTH
THH

TTT
TTH
THT
HTT

Random variables:

0 Head
1 Head
2 Heads
3 Heads

0 head:	1	/	8	=	0.125
1 head:	3	/	8	=	0.375
2 heads:	3	/	8	=	0.375
3 heads:	1	/	8	=	0.125

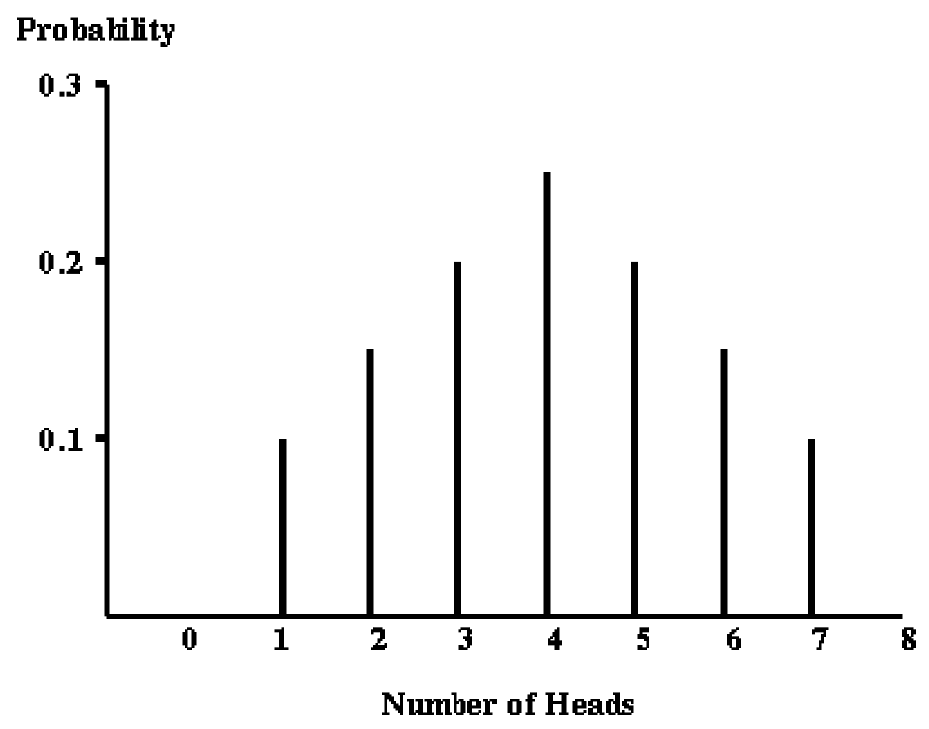


Binomial distribution

If you toss a coin 8 times what is the probability of obtaining a score of:

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

Binomial distribution



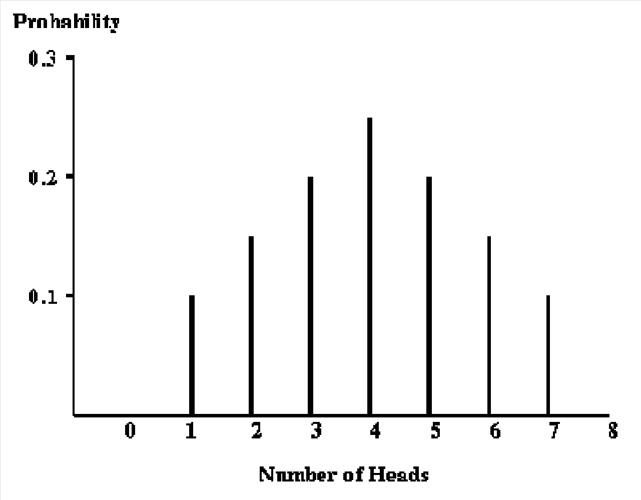
Binomial distribution

Example: Tossing a coin one 100 times, yielded 42 heads and 58 tails. Is this a fair coin?

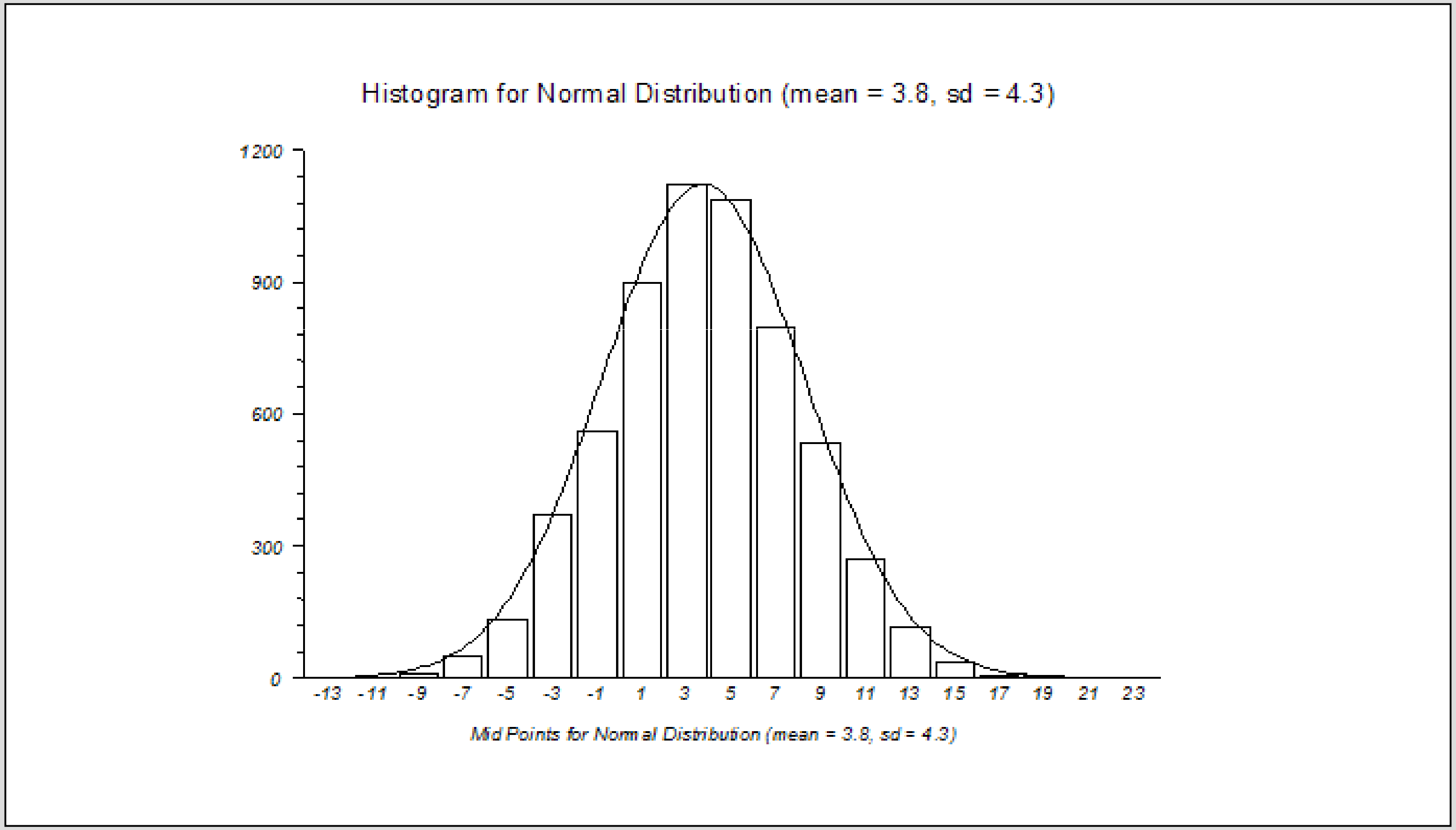
Heads:	42
Tails:	58

Expected: 50% - 50%

Sample error?

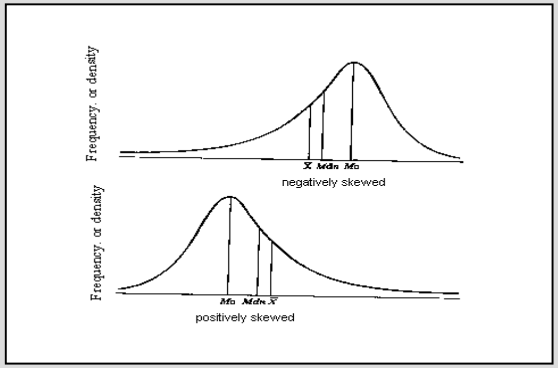


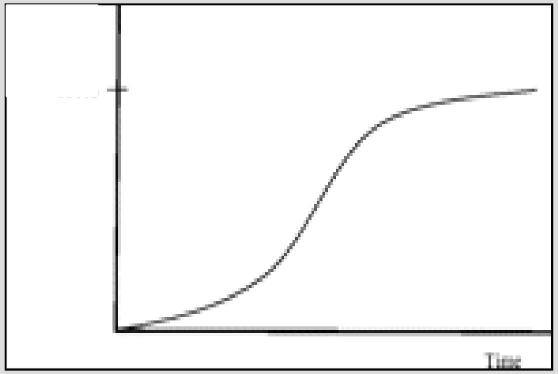
Normal distribution

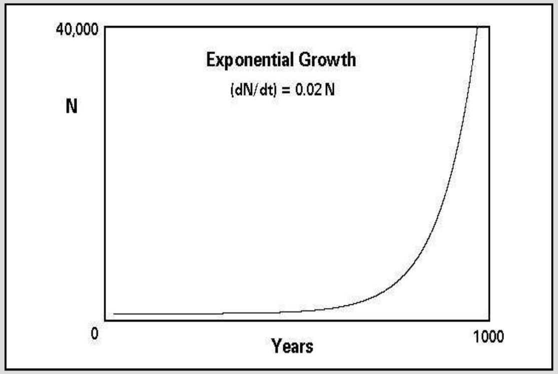


Normal distribution

- The center of the curve represents the mean, median, and mode.
- The curve is symmetrical around the mean.
- The tails meet the x-axis in infinity.
- The curve is bell-shaped.
- The total under the curve is equal to 1 (by definition).





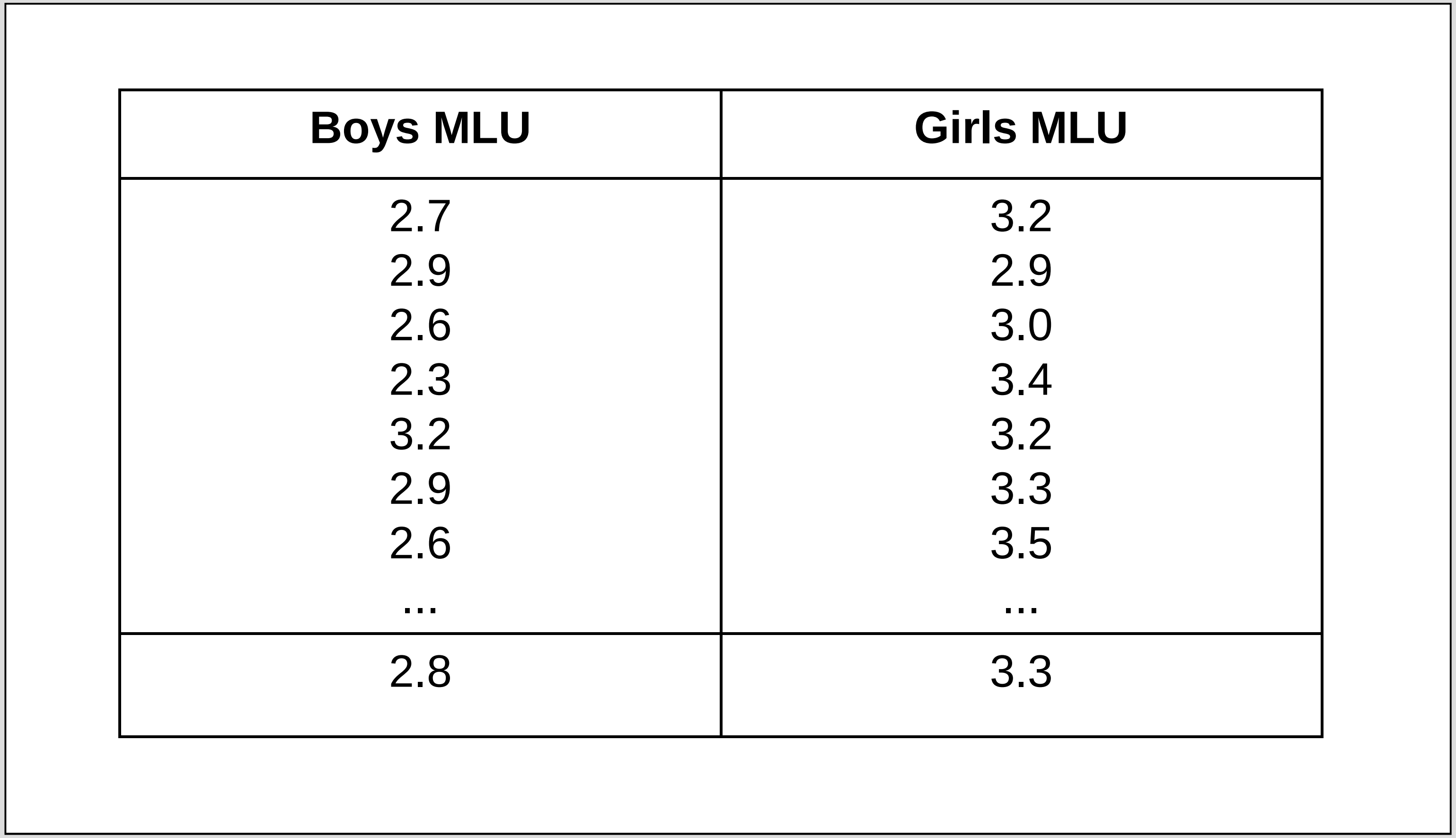


Normal distribution

A child language researcher wants to find out if there is a difference in the MLUs of English-speaking boys and girls. In order to answer this question, she collects a sample of utterances from 30 three-year-old boys and 30 three-year-old girls. The sample of utterances from the boys has an MLU of 2.8 words, while the sample of utterances from the girls has an MLU of 3.3 words. The sample means are different, but are they different enough to conclude that boys and girls produce utterances with different MLUs in the true population? To answer this question, we need a probability model. To find the right model, we need to inspect the data. There are two important aspects:

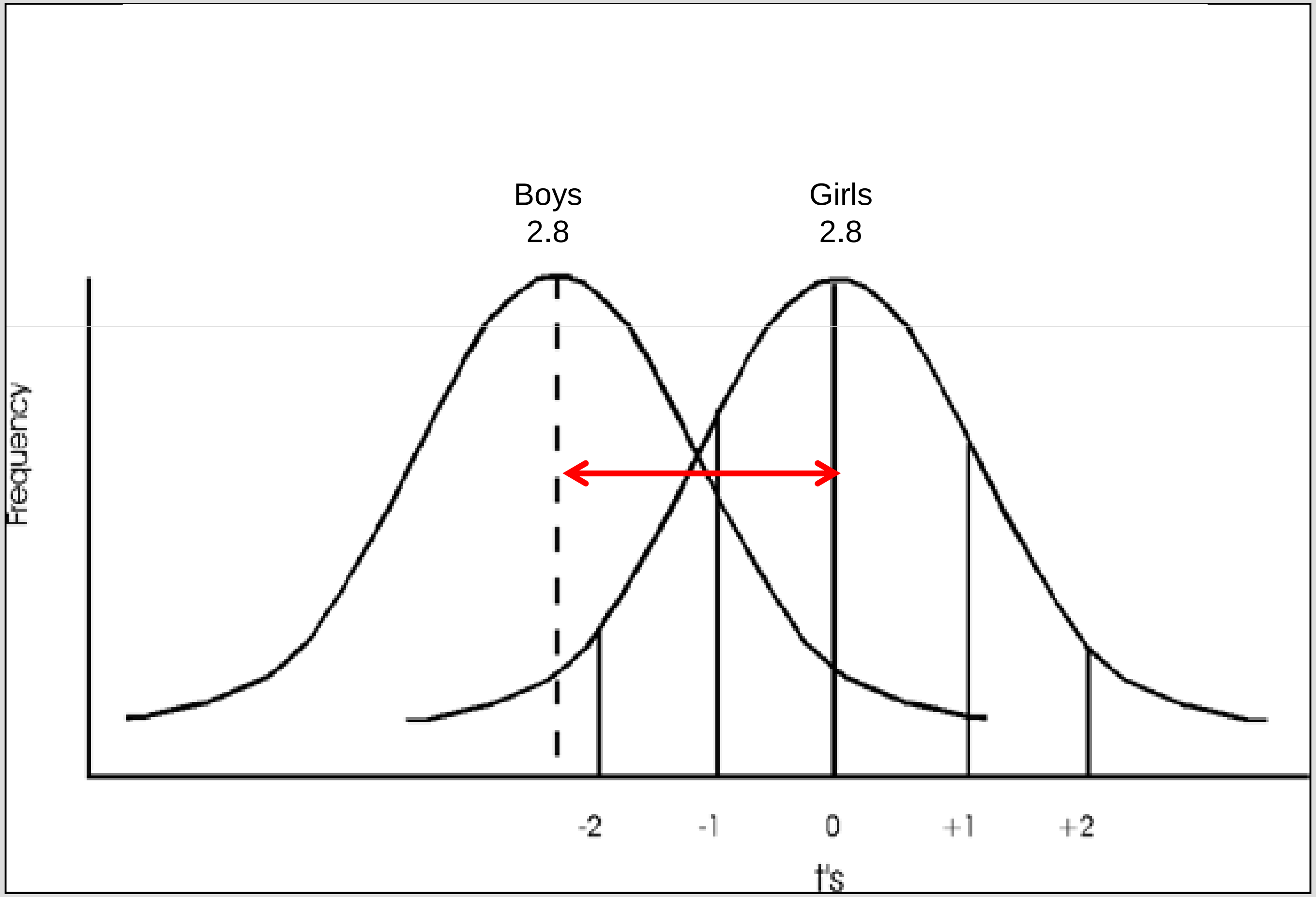
- The mean of utterance length is interval data
- The data of both groups is centered around a mean

Normal distribution



Boys MLU	Girls MLU
2.7	3.2
2.9	2.9
2.6	3.0
2.3	3.4
3.2	3.2
2.9	3.3
2.6	3.5
...	...
2.8	3.3

Normal distribution



p-value

The p-value indicates the probability that we will obtain the distribution in our sample given that there is no relationship between x and y in the true population.

The p-value is conditional on the hypothesis that the null hypothesis is true.

> If there is no relationship (difference) between X and Y in the true population, then there is a less than a 5% chance (i.e. 1 out of 20 chance) that we obtain the distribution in a given sample.

p-value

What does the p-value indicate?

- The probability of the null hypothesis to be true is 5%. false
- The probability of the alternative hypothesis to be true is 95%. false
- Given that the null hypothesis is true, there is a 5% chance of obtaining the distribution in a particular sample. true

The probability of the experimental hypothesis is not measured at all!

Type 1 and type 2 error

- Eine wahre Nullhypothese wird durch die Stichprobe bekräftigt. (p größer als .05 + es gibt keinen Zusammenhang zwischen x und y -> correct)
- Eine wahre Nullhypothese wird durch die Stichprobe fälschlicherweise entkräftet. (p ist kleiner als .05 obwohl es keinen Zusammenhang zwischen x und y gibt -> false)
- Eine falsche Nullhypothese wird durch die Stichprobe als falsch gekennzeichnet. (p größer als .05, es gibt aber einen Zusammenhang zwischen x und y -> correct)
- Eine falsche Nullhypothese wird durch die Stichprobe fälschlicherweise bestätigt. (p kleiner als .05 obwohl es gibt keinen Zusammenhang zwischen x und y -> false)

Type 1 and type 2 error

Two potential errors:

- Type 1 error: we reject a true null hypothesis
- Type 2 error: We accept a false null hypothsis

	p smaller than .05	**p larger than .05**
False null hypothesis	Correct	Type 2 error
True null hypothesis	Type 1 error	Correct

The p-value indicates the probability of making a type 1 error. It does not say anything about making a type 2 error!

Other statistical measures

Until recently the p-value was taken as a ‘holy’ cut-off point; but the value .05 is of course arbitrary. Therefore, researchers now consider the p-value in conjunction with other statistical measures:

- p-value
- confidence intervals
- effect size

Confidence intervals indicate the range within which the mean (or some other statistical measure) must lie, assuming a particular degree of certainty (e.g. 95%) in the true population.

The effect size indicates the degree to which difference in the dependent variable are due to changes in the independent variable.

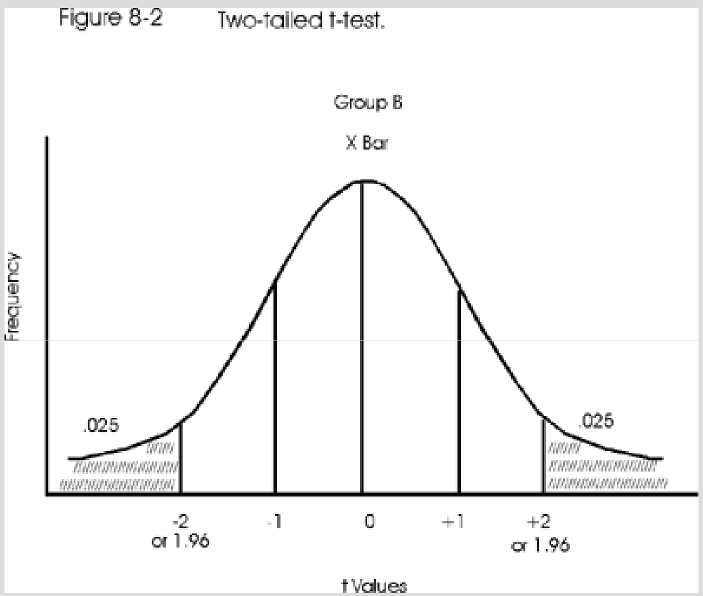
One-sided and two-sided tests

A researcher wants to find out if sex influences language development during childhood. He collected MLU values from a group of 3 year-old boys and 3 year-old girls.

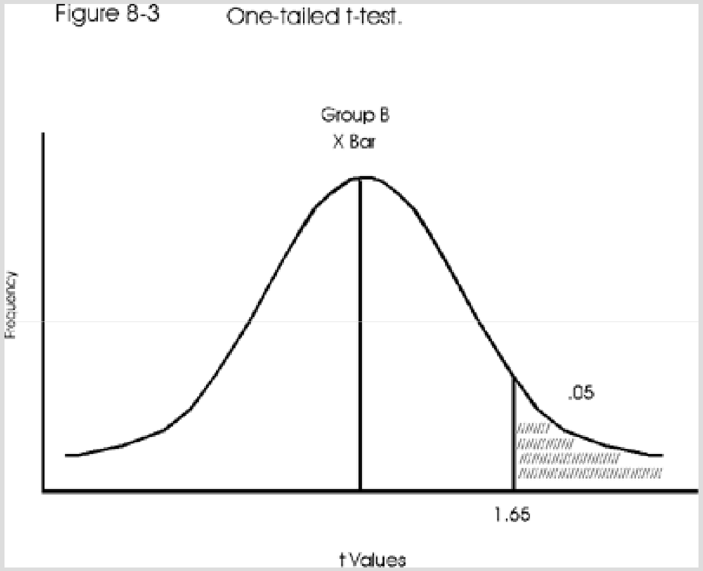
Hypotheses:

- Sex does not influence development (i.e. MLU)
- Sex influences development (i.e. MLU)
- Girls have a higher MLU
- Boys have a higher MLU

One-sided and two-sided tests



There is a difference between boys and girls



Boys have samller MLUs than girls (or vice versa)

Sex	**Weight**	**Height**
Male	178	75
Male	196	100
Male	145	60
Male	170	71
Male	180	80
Male	175	69
Male	185	78
Male	190	90
Male	183	70
Male	182	85
Female	163	60
Female	142	51
Female	150	55
Female	165	64
Female	160	53
Female	175	50
Female	182	72
Female	169	65
Female	162	62
Female	182	80