Basic Statistics for Biologists

Part I

Prof. Dr. Meinhard Kieser
Institute of Medical Biometry

Aims and scope of the course

- As annonced: this is a course of „**Basic** Statistics for Biologists“
- “You will be able to...
  - explain basic statistical concepts like description, inference, testing, estimation and the effect of sample size.
  - interpret statistical measures like mean, median, standard deviation, confidence intervals, p-value, reference intervals and measures of diagnostic accuracy.
  - interpret common visualization methods like bar charts, boxplots, histograms and scatterplots
  - select the appropriate statistical test for basic applications, covering the chi-square test, t test and non-parametric tests
  - …

Aims and scope of the course – today

- As annonced: this is a course of „**Basic** Statistics for Biologists“
- “You will be able to...
  - explain **basic statistical concepts** like description, inference, testing, estimation and the effect of sample size.
  - **interpret statistical measures** like mean, median, standard deviation, confidence intervals, p-value, reference intervals and measures of diagnostic accuracy.
  - **interpret common visualization methods** like bar charts, boxplots, histograms and scatterplots.
  - select the appropriate statistical test for basic applications, covering the **chi-square test**, t test and non-parametric tests.”
  - ...

Introduction and overview

- „Theory is inspired by practice“: the hamster experiment – motivation and starting point
- „What to do with all these data?“: descriptive analysis
- „Is this really true?“: confirmatory analysis – principle and illustration with the binomial test
- „How large is the effect?“: point estimation and confidence intervals – principle and illustration for the case of rates
- „This is not the end of the story“: another statistical test – the chi-square test

Learning goals

- You know the aims of descriptive data analysis as well as the most important measures of scale and variation and types of graphics. You are able to apply the related techniques and to interpret the results appropriately.
- You know the aims of confirmatory analysis and the principle of statistical tests. You are able to apply the binomial test and the chi-square test and to interpret the results appropriately.
- You know the concept of point estimation and confidence intervals. You are able to compute the quantities for rates and differences of rates and to interpret the results appropriately.

The starting point and the experiment

The starting point – a conversation at a walking tour

Foto: IMBI

The starting point – conversation at a walking tour (2)

- „When I talked with my sister recently, she said that there are more male hamsters than female ones.“
- „This is complete nonsens! I‘ve seldomly heard something such crazy as this statement.“
- „Initially, I thought the same. But my sister convinced me by showing me an article in a journal. She is a very serious person; now I believe her.“
- „Was this a scientific journal? And even if yes: Was this a peer-reviewd journal?“
- „I don‘t know. But who cares?“

The starting point – conversation at a walking tour (3)

- „At least me! And it should also be important for you! As a future biologist, you should address such claims with scientific methods!“
- „You‘re right! So what to do?“
- „Let‘s do an experiment to learn more about hamsters. Especially whether or not male hamsters are more frequent than female ones.“
- „That‘s great! We ask all our friends who have hamsters as pets and record data on them.“
- „When I think by myself: I know very little about hamsters. This is a big chance to change this!“

The starting point – conversation at a walking tour (4)

- „So let‘s not only ask for the gender of the hamsters but also for their age, weight, and length. Anything else?“
- „Yes, I would also be interested in the number of farrows during one year and the total number of pubs during one year. Furthermore, I‘m not sure which fur colors hamsters have and which are the most frequent ones. And we should also record the activity of the hamsters“
- „Let‘s start!“

The experiment

- Friends (and friends of friends) of the two guys had in total 37 hamsters at home
- The two guys collected the data and included them in an Excel-sheet:

A2 Identification number

	A	B	C	D	E	F	G	H	I
1	id	sex	age	weight	length	litters	pups	fur	activity
2	*Identification number*	*Sex*	*Age [months]*	*Weight [g]*	*Head-torso length [cm]*	*No. of litters in the last year*	*No. of pups in the last year*	*Main fur colour*	*Activity level*
3	1	m	24	395	28			red brown	very active
4	2	f	36	224	21	3	14	yellow	active
5	3	m	20	261	25			red brown	very active
6	4	m	45	422	30			yellow	active
7	5	m	22	292	27			light yellow	active
8	6	m	32	392	29			dark brown	very active
9	7	m	25	564	33			red brown	very active
10	8	m	36	264	26			black	active
11	9	f	34	261	24	2	6	dark brown	very active
12	10	m	57	368	28			dark brown	active

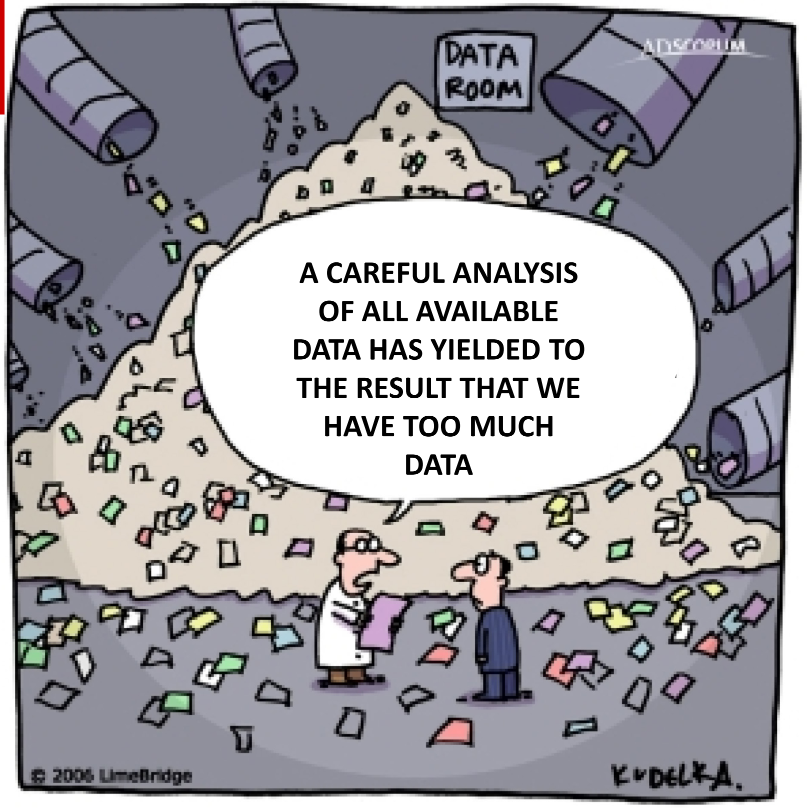
The experiment (2)

- „Whow! Much information. How can we get an overview on the results?“
- „If I remember rightly from my basic statistics course, this is the topic of descriptive analysis.“

Descriptive analysis

Descriptive analysis – what to do with all these data?

By means of descriptive analysis, the most important characteristics of the data are summarized and illustrated by graphics.



Quelle: Jon Kudelka www.limebridge.com

Descriptive analysis – describing the data

- In studies and experiments, **big data sets** are commonly collected (many objects with many characteristics)
- In our experiment:
  - age
  - sex
  - weight
  - length
  - activity
  - fur color

  …

Example of data in hamster experiment for one characteristic

age of hamsters (in months):

24, 36, 20, 45, 22, 32, 25, 36, 34, 57, 11, 35, 46, 14, 33, 25, 33, 21, 15, 15, 36, 44, 35, 48, 23, 38, 41, 36, 28, 24, 35, 44, 31, 34, 43, 34, 50

Example of data in hamster experiment for one characteristic (2)

age of hamsters (in months):

24, 36, 20, 45, 22, 32, 25, 36, 34, 57, 11, 35, 46, 14, 33, 25, 33, 21, 15, 15, 36, 44, 35, 48, 23, 38, 41, 36, 28, 24, 35, 44, 31, 34, 43, 34, 50

This is the complete information!

But is it also informative?

Descriptive analysis

- Data set is **reduced to a few meaningful measures**.
- Figures are supplemented by **graphics**.

➢ **condensed presentation of essential information**

Example of data in hamster experiment for one characteristic (3)

age of hamsters (in months):

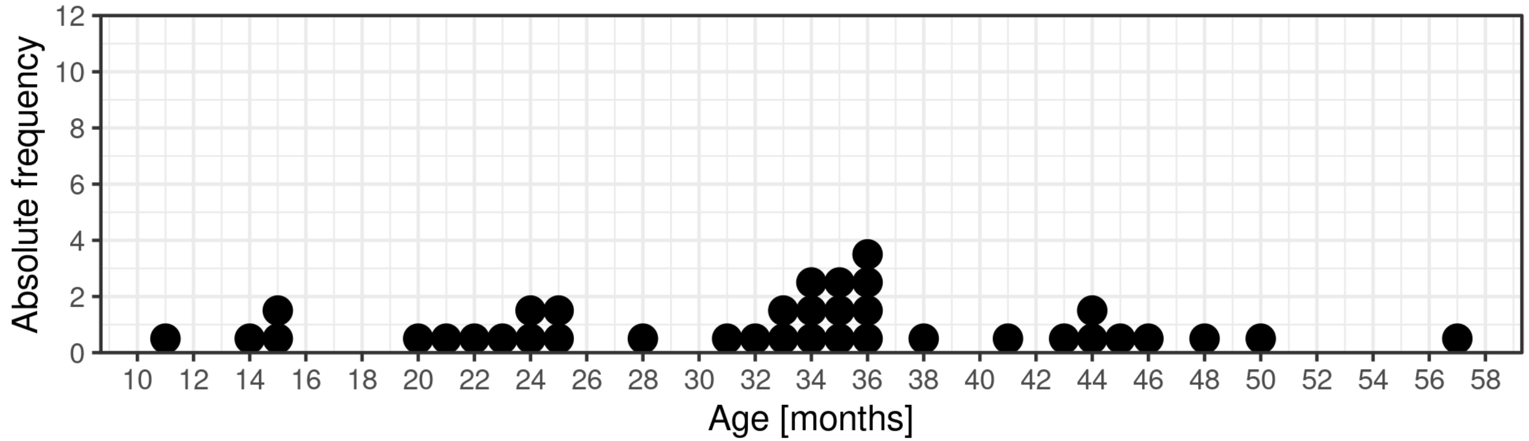
11, 14, 15, 15, 20, 21, 22, 23, 24, 24, 25, 25, 28, 31, 32, 33, 33, 34, 34, 34, 35, 35, 35, 36, 36, 36, 36, 38, 41, 43, 44, 44, 45, 46, 48, 50, 57 (ordered)

Example of data in hamster experiment for one characteristic (4)

age of hamsters (in months):

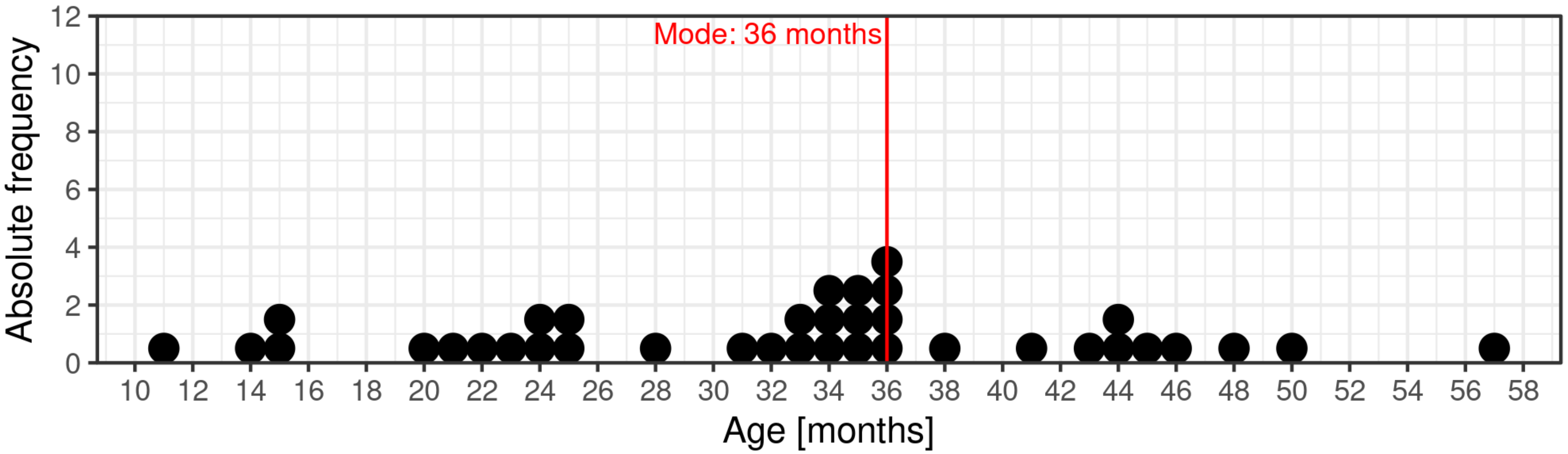
11, 14, 15, 15, 20, 21, 22, 23, 24, 24, 25, 25, 28, 31, 32, 33, 33, 34, 34, 34, 35, 35, 35, 36, 36, 36, 36, 38, 41, 43, 44, 44, 45, 46, 48, 50, 57 (ordered)

distribution of values:



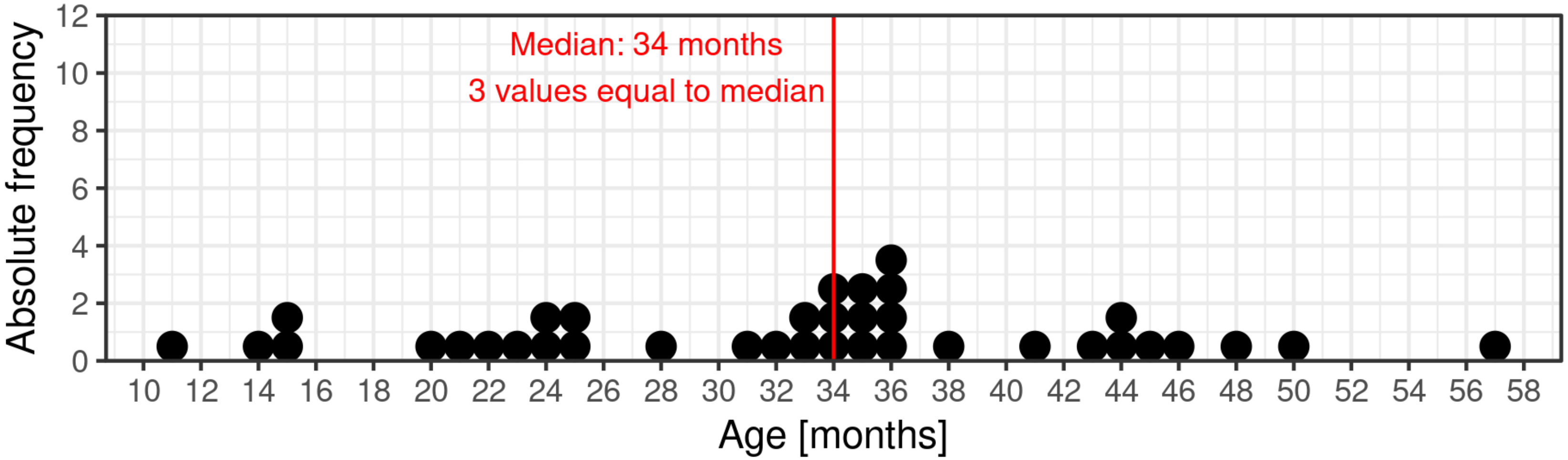
Measures of location – mode

The **mode** is the value that occurs
most frequently.



Measures of location – median

The **median** divides the ordered sample
in **two even halves**.

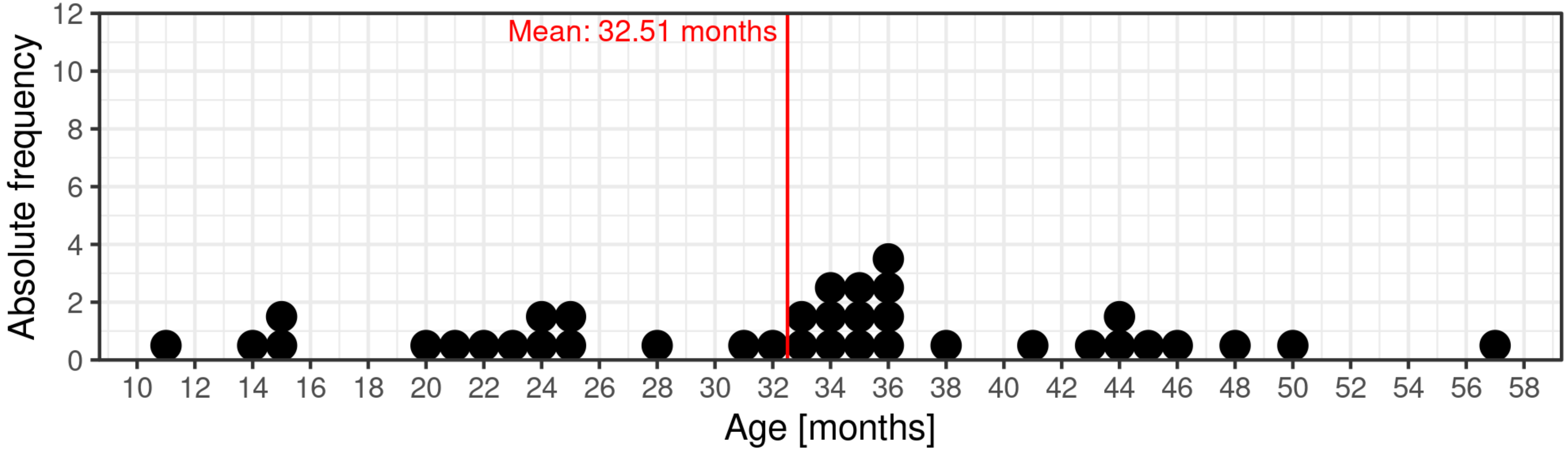


17 values smaller than median:
11, 14, 15, 15, 20, 21, 22, 23, 24,
24, 24, 25, 25, 28, 31, 32, 33, 33

17 values greater than median:
35, 35, 35, 36, 36, 36, 36, 38, 41,
41, 43, 44, 44, 45, 46, 48, 50, 57

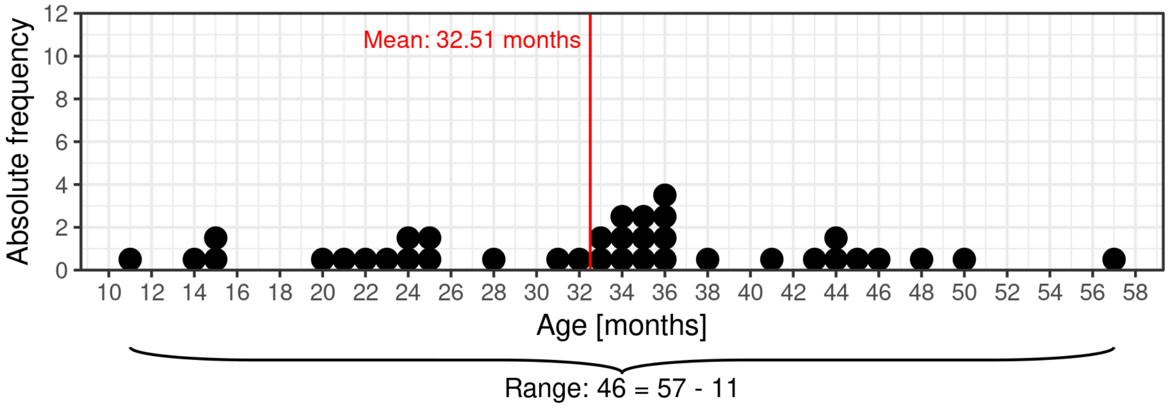
Measures of location – mean

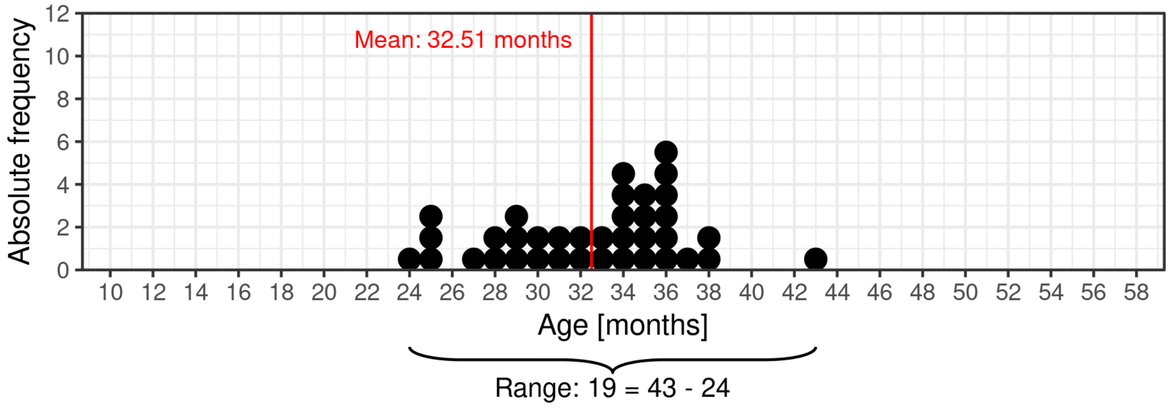
The **mean** (= average) corresponds to the balance point of the distribution.



Measures of variation – range

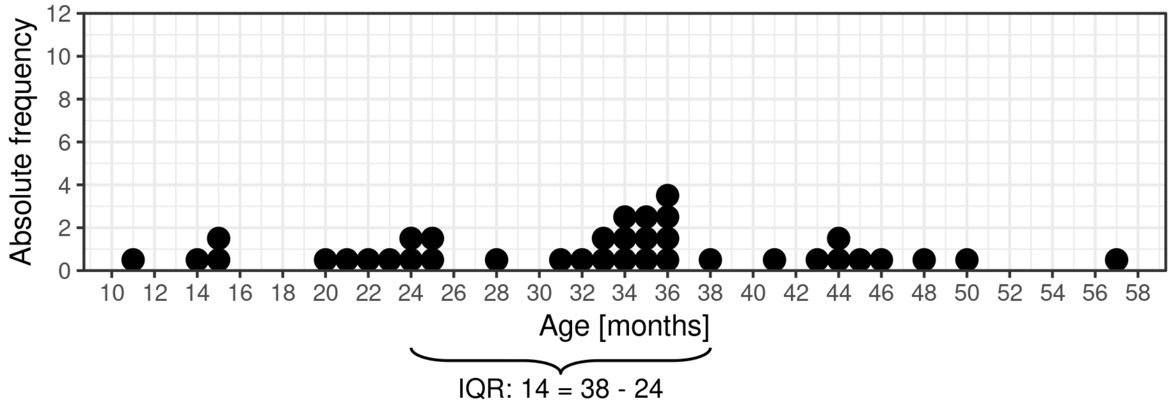
- Distributions may not only differ by their **location** but also by their **variation (variability)**.
- The most simple measure of variation is the **range**, i.e. the difference between the largest and smallest value

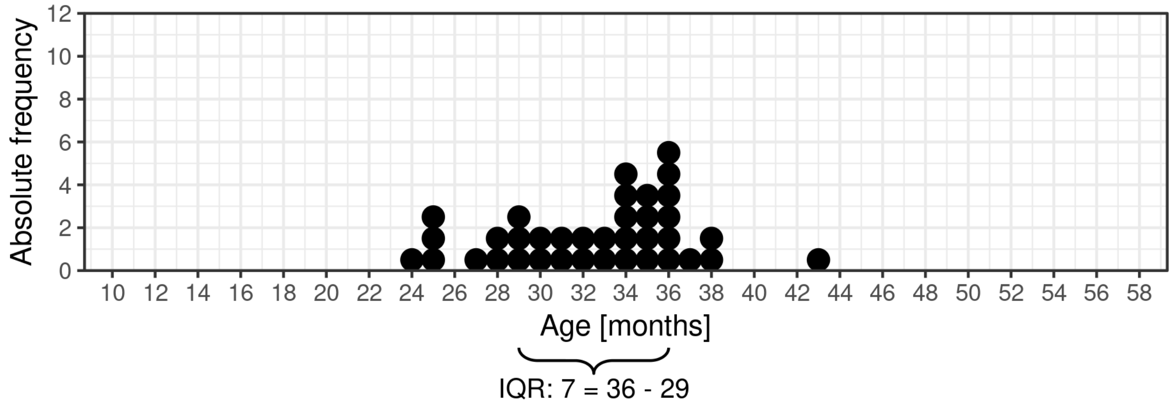




Measures of variation – interquartile range

- The **interquartile range (IQR)** shows, in which range the middle 50% of values lie.



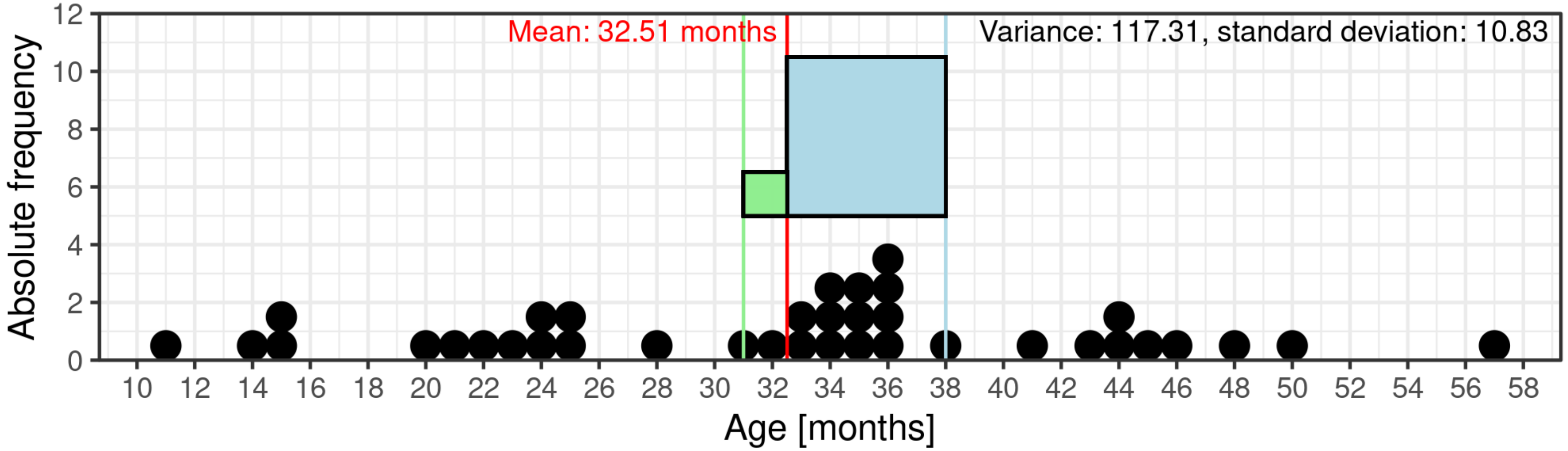


Measures of variation – standard deviation and variance

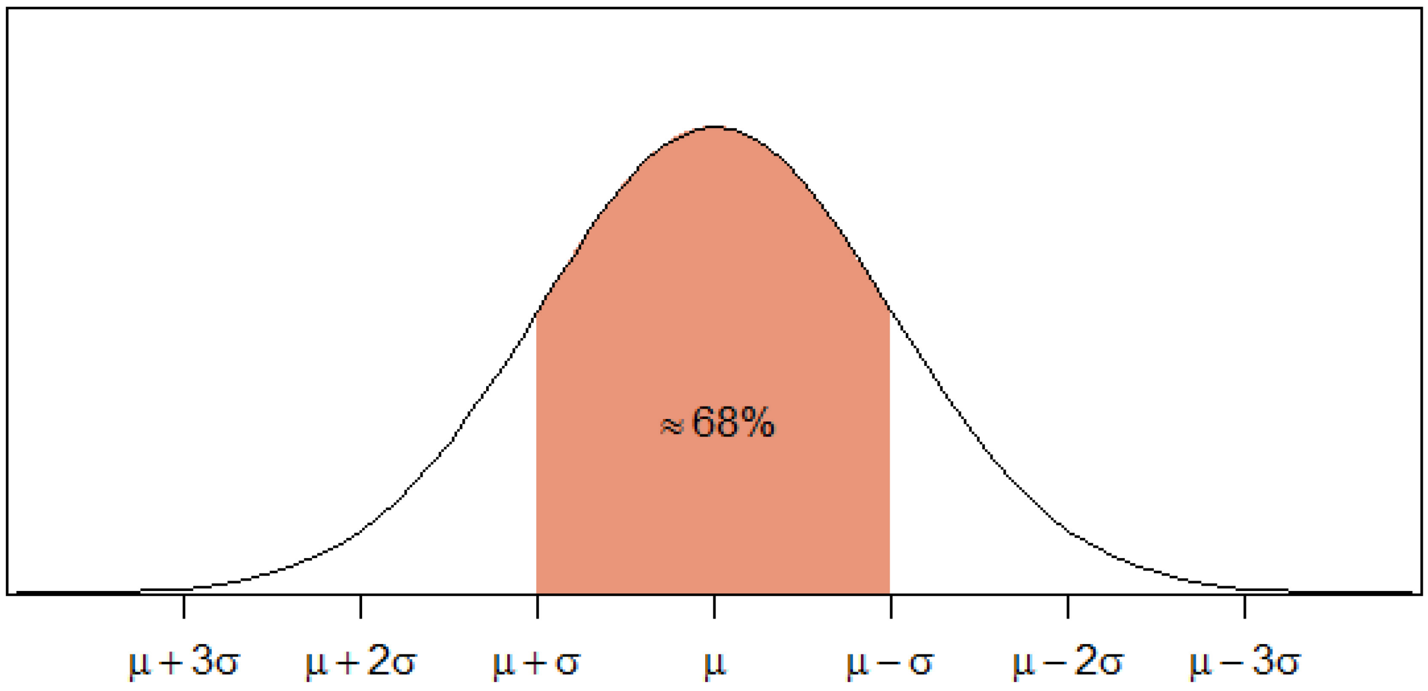
- The **variance** is the mean quadratic deviation of the single values from their mean. The **standard deviation** is the square root of the variance.
- The farther the value is away from the mean, the larger is its contribution to the variance.

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

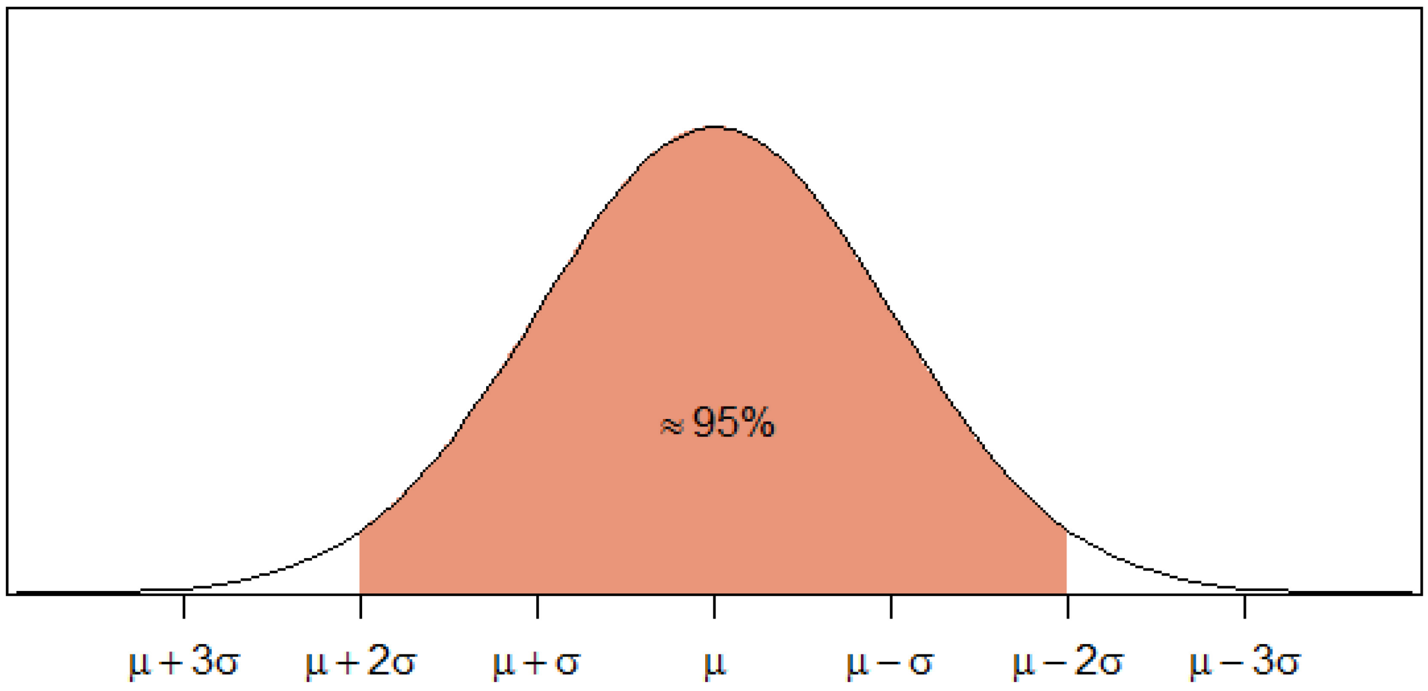
standard deviation $$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$



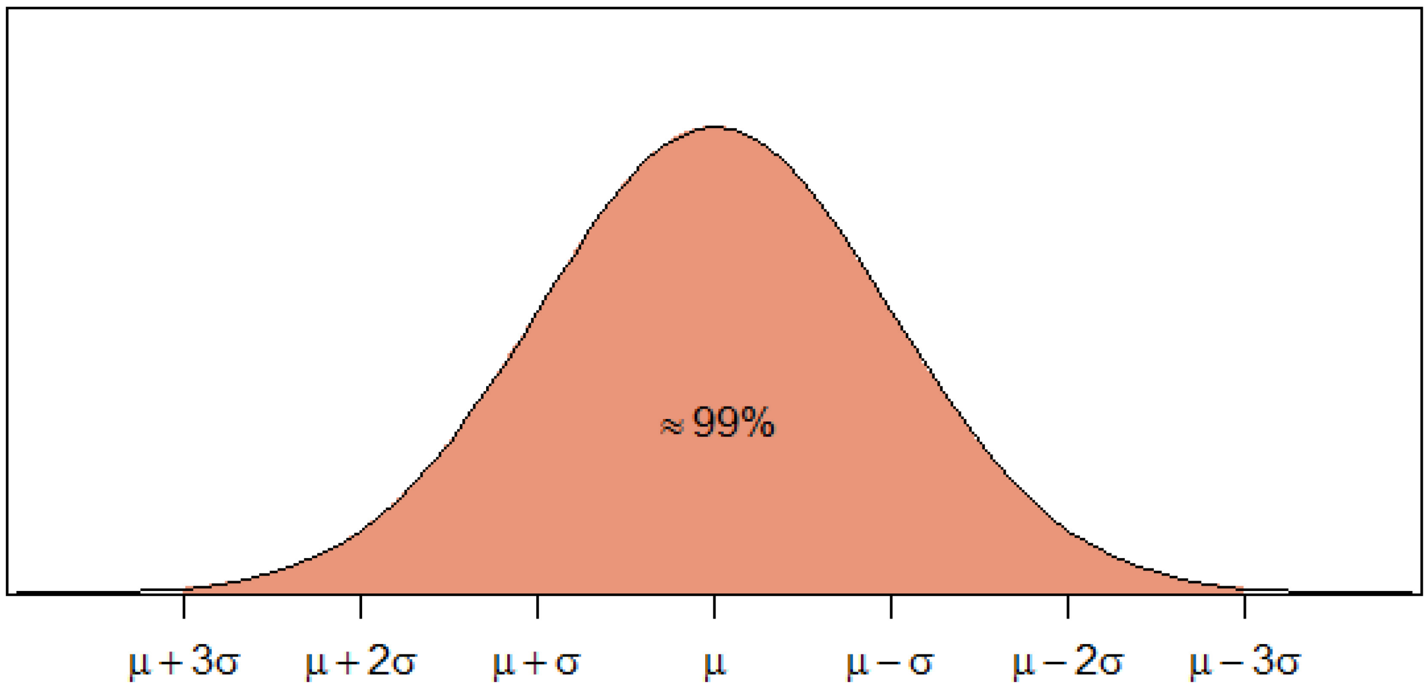
Measures of variation – standard deviation in case of normally distributed data



Measures of variation – standard deviation in case of normally distributed data

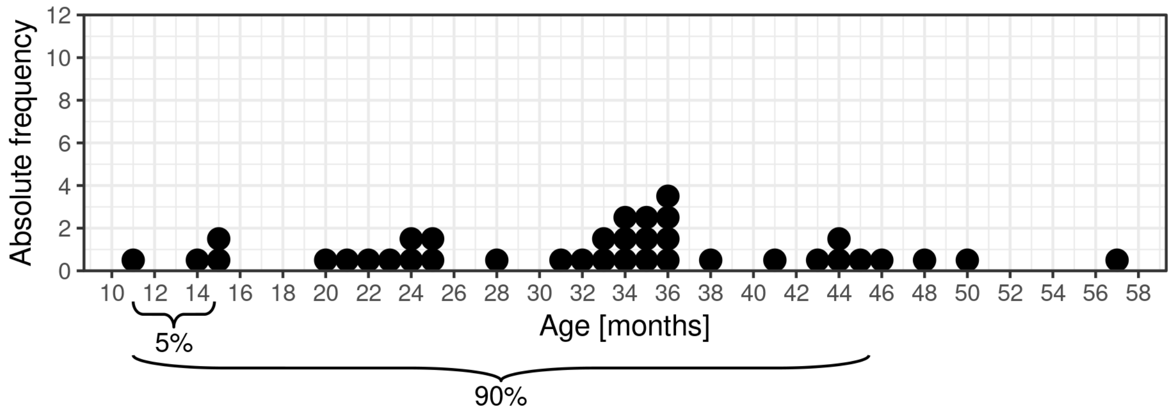


Measures of variation – standard deviation in case of normally distributed data

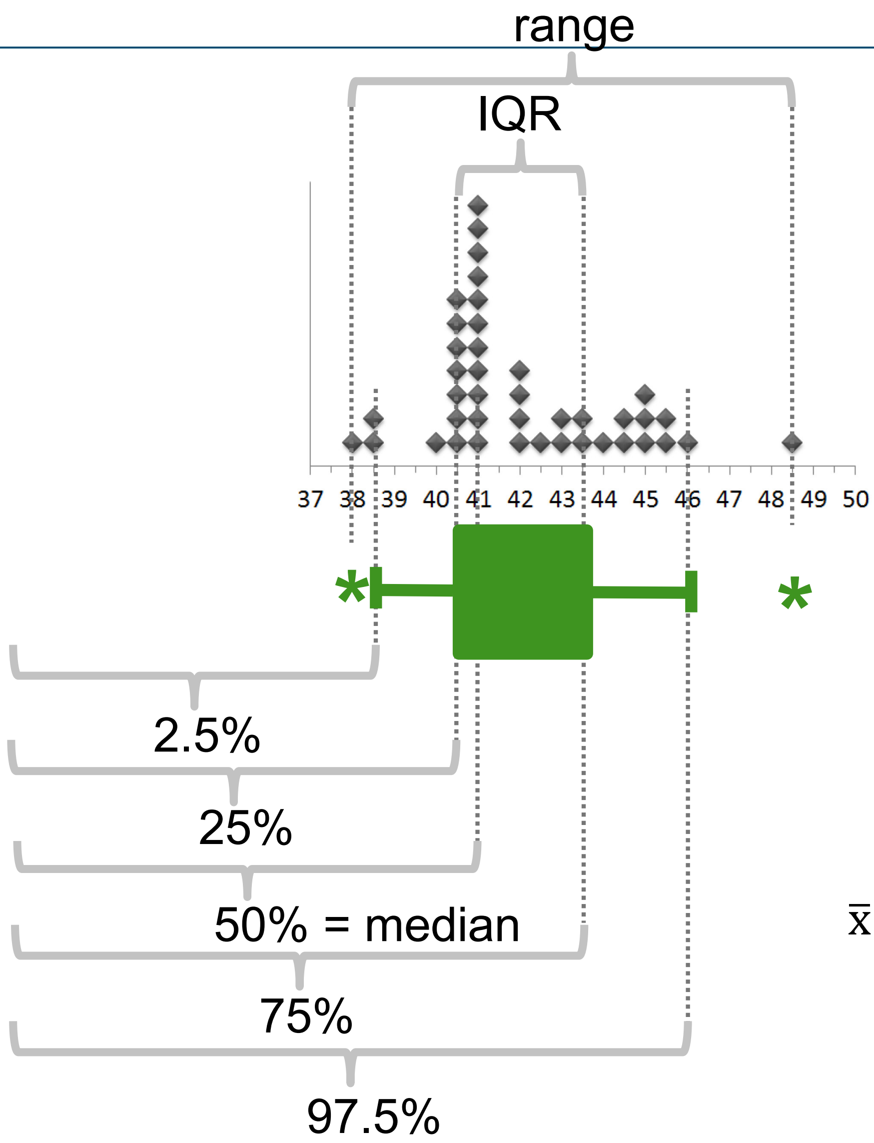


Excursus: percentiles

- The **quartiles** (lower quartile (Q1), median, upper quartile (Q3)) divide the ordered sample in 4 quarters.
- By **percentiles**, the ordered sample can be divided in arbitrary parts: x% of the values are smaller than the xth percentile.
- **Example:** The smallest 5% or the largest 10% (= smallest 90%) of the values, respectively, are bounded by the 5th or 90th percentile.



Graphics – box-whisker plot

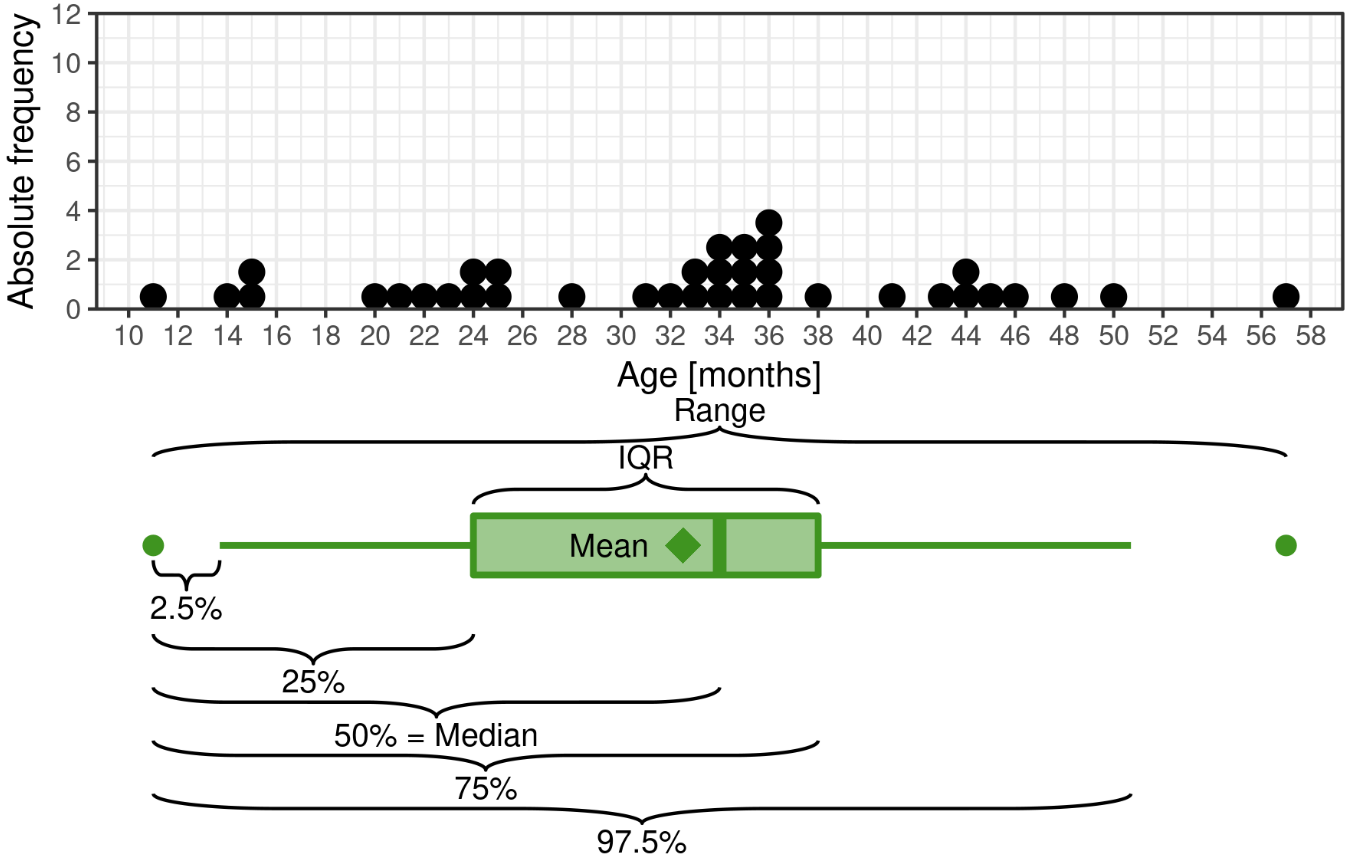


- summarizes measures of location and variation of values
- includes a huge amount of information about location, variation, and shape of the distribution

$\bar{x}$ = mean

For our example of data from hamster experiment

boxplot for age of hamsters (in months):



„No measure of location without measure of variation!“

When describing data, **both measures for location and variation** should be presented (whenever possible),

for example

- mean $\pm$ standard deviation
- median and quartiles / interquartile range

In addition, the underlying **sample size** should be reportet.

Which measures (and graphics) are reasonable for which kind of data? – Scale levels of measurements

- Measurements have different information content.
- **Example:** the activity of the hamsters can be measured by
  - the number of turns of the rat race during 24h, or
  - by rating the activity as „inactive", „active", or „very active"

➢ different **scale levels**

➢ require **different methods for description of data** by measures and graphics

Scale levels of measurements

scale level	characteristics	examples
nominal scale	values can be declared as equal or unequal	gender, fur color

Scale levels of measurements

scale level	characteristics	examples
ordinal scale	additionally: values can be ordered	activity („inactive“, „active“, „very active“)
nominal scale	values can be declared as equal or unequal	gender, fur color

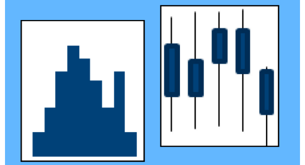
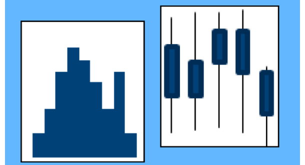
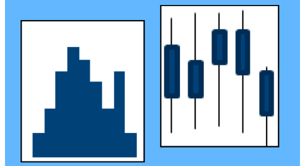
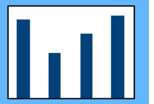
Scale levels of measurements

scale level	characteristics	examples
interval scale	additionally: differences between values can be interpreted	date of birth
ordinal scale	additionally: values can be ordered	activity („inactive“, „active“, „very active“)
nominal scale	values can be declared as equal or unequal	gender, fur color

Scale levels of measurements

scale level	characteristics	examples
ratio scale	additionally: ratios between values can be interpreted	age, weight, length
interval scale	additionally: differences between values can be interpreted	date of birth
ordinal scale	additionally: values can be ordered	activity („inactive“, „active“, „very active“)
nominal scale	values can be declared as equal or unequal	gender, fur color

Which measures (and graphics) are reasonable for which kind of data?

scale	statements	location	variation	graphics	examples
ratio	$= \neq < >$ $+ - \bullet :$	mean median	Q1 & Q3 min & max variance SD IQR range		age, weight, length
interval	$= \neq < >$ $+ -$	mean median	Q1 & Q3 min & max variance SD IQR range		date of birth
ordinal	$= \neq < >$	median mode	Q1 & Q3 min & max (IQR[1] range[1])		scores
nominal	$= \neq$	mode			gender, fur color

[1] Strictly, IQR and range require interval scaled data as they are based on differences. However, they are sometimes nevertheless shown in practice.

Application: descriptive analysis of data from hamster experiment – „the famous Table I“

Table 1: Description of baseline variables

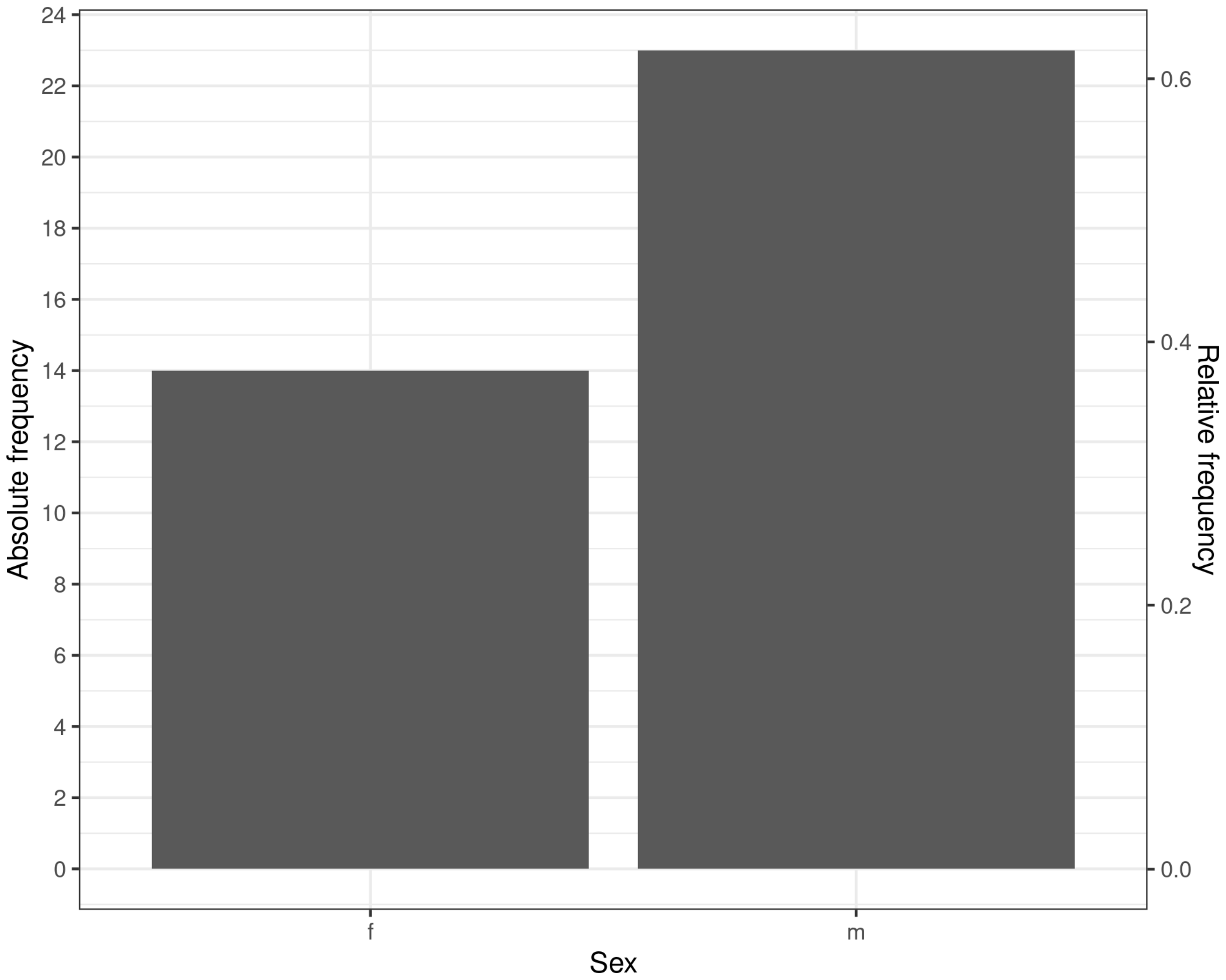
Variables	male (N=23)	female (N=14)	Total (N=37)
Age [months]			
N	23	14	37
mean	33	32	33
sd	11	11	11
median	34	35	34
Q1 - Q3	24 – 41	23 – 38	24 – 38
min - max	15 – 57	11 – 48	11 – 57
Weight [g]			
N	23	14	37
mean	366	235	316
sd	83	40	95
median	349	242	318
Q1 - Q3	318 – 395	217 – 267	261 – 361
min - max	261 – 604	139 – 285	139 – 604
Head-torso length [cm]			
N	23	14	37
mean	28	22	26
sd	2.3	1.6	3.6
median	28	22	27
Q1 - Q3	27 – 29	21 – 24	23 – 28
min - max	25 – 35	19 – 25	19 – 35
No. of litters in the last year			
N	0	14	14
Nmiss	23 (100%)	0 (0%)	23 (62%)
mean	NA	2.2	2.2
sd	NA	1.5	1.5
median	NA	2	2
Q1 - Q3	NA – NA	1 – 4	1 – 4
min - max	NA – NA	0 – 4	0 – 4

Application: descriptive analysis of data from hamster experiment – „the famous Table I“ (contd.)

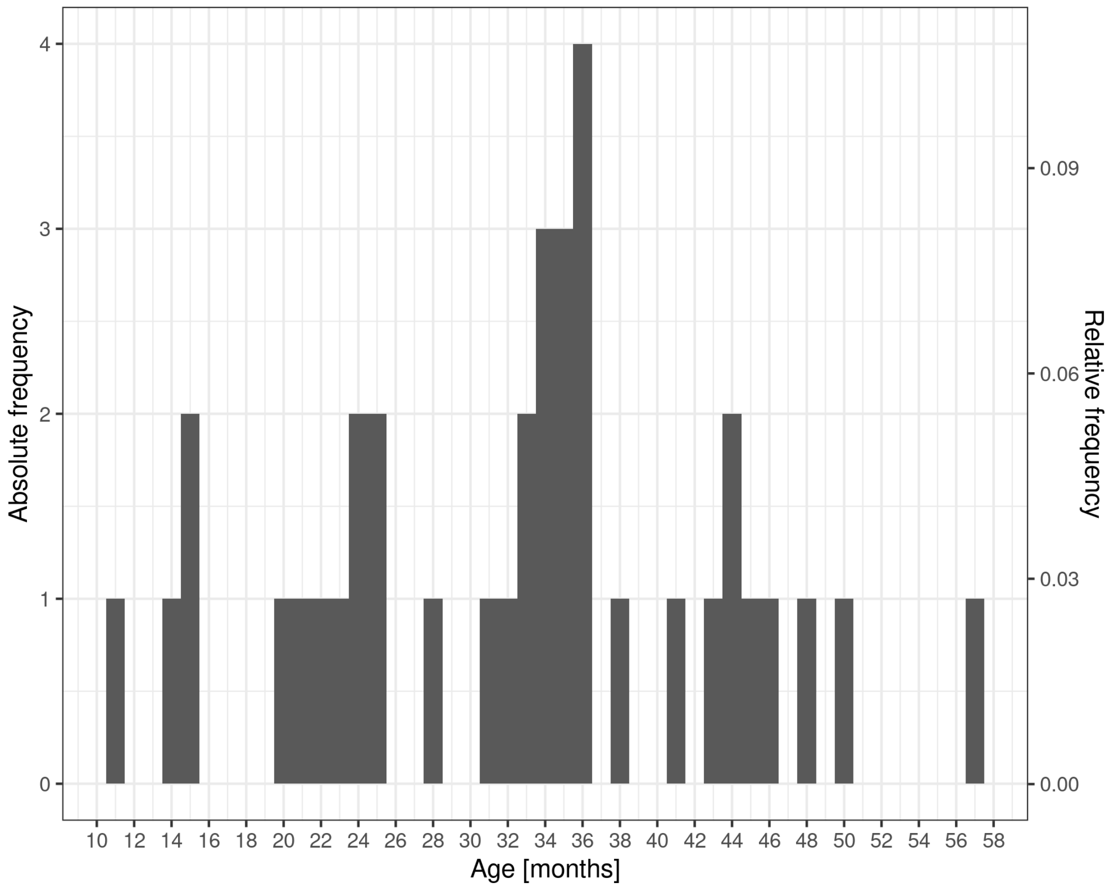
Table 1: Description of baseline variables *(continued)*

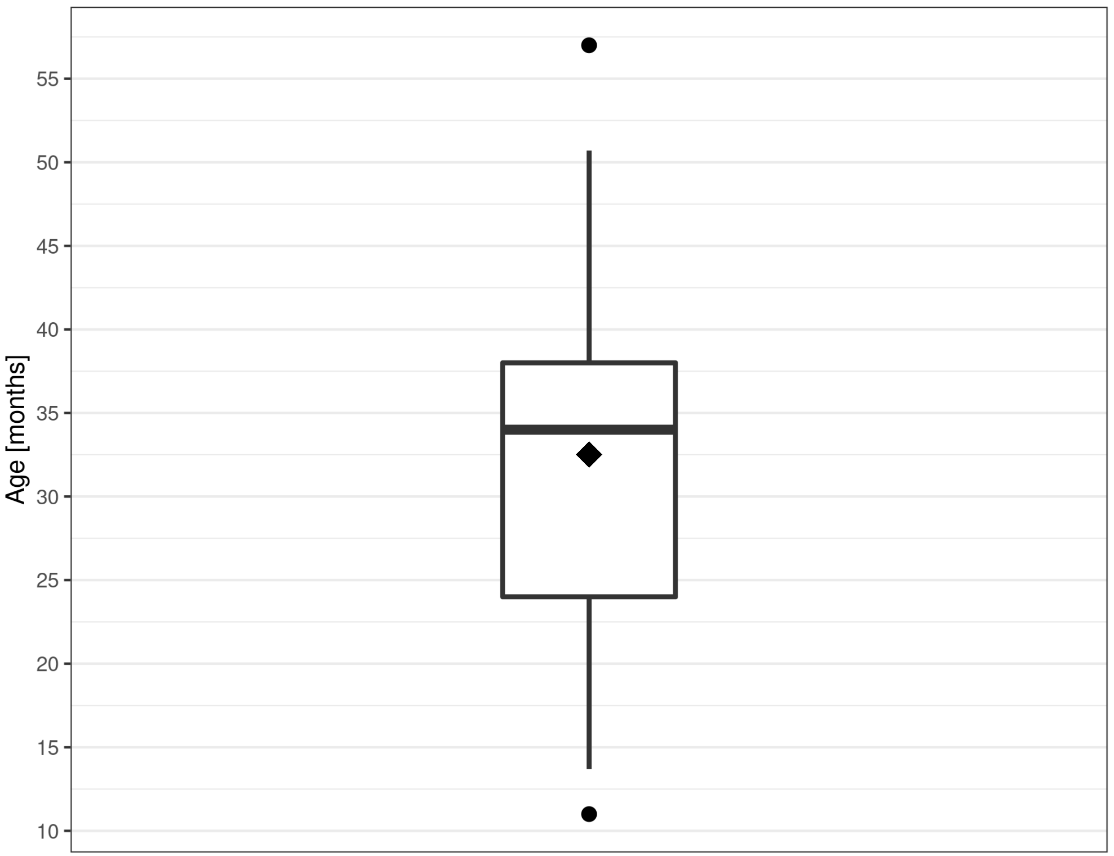
Variables	male (N=23)	female (N=14)	Total (N=37)
No. of pups in the last year			
N	0	14	14
Nmiss	23 (100%)	0 (0%)	23 (62%)
mean	NA	7.3	7.3
sd	NA	6.8	6.8
median	NA	5.5	5.5
Q1 - Q3	NA – NA	2 – 12	2 – 12
min - max	NA – NA	0 – 23	0 – 23
Main fur colour			
red brown	6 (26%)	5 (36%)	11 (30%)
yellow	5 (22%)	5 (36%)	10 (27%)
light yellow	6 (26%)	1 (7%)	7 (19%)
dark brown	4 (17%)	2 (14%)	6 (16%)
black	2 (9%)	1 (7%)	3 (8%)
Activity level			
inactive	4 (17%)	4 (29%)	8 (22%)
active	11 (48%)	7 (50%)	18 (49%)
very active	8 (35%)	3 (21%)	11 (30%)

Application: descriptive analysis of data from hamster experiment – variable „sex"

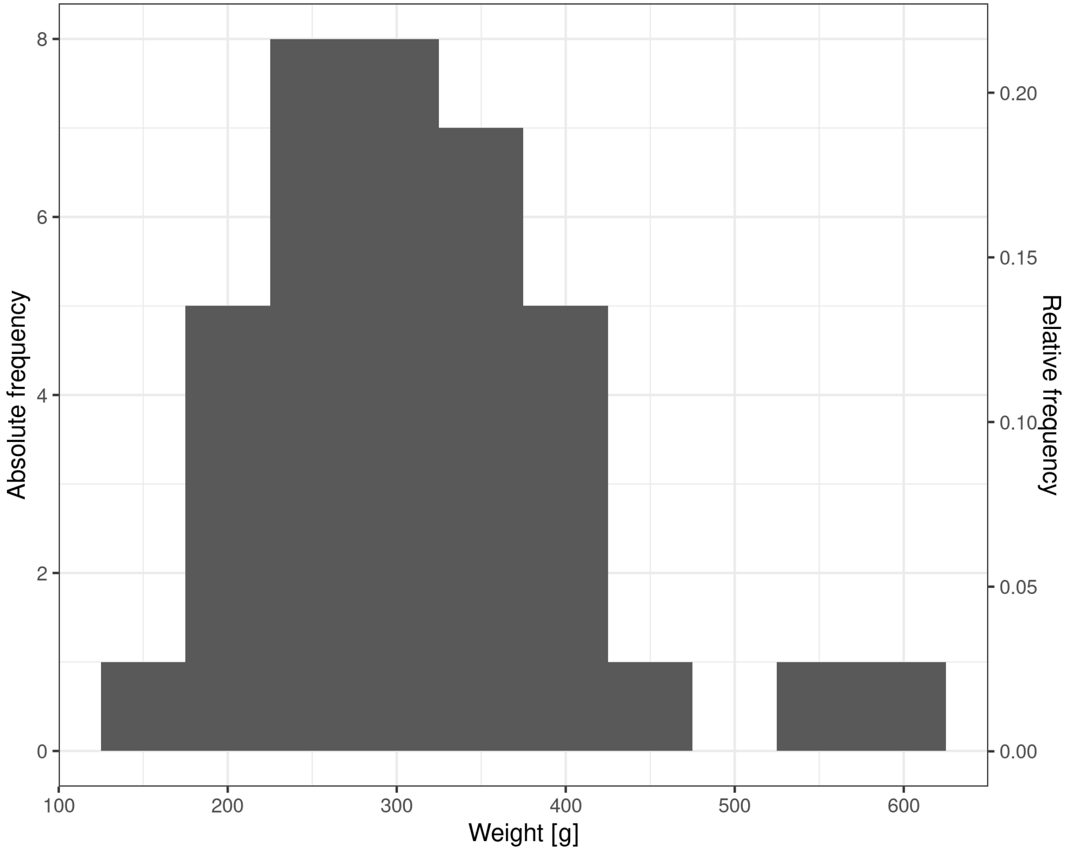


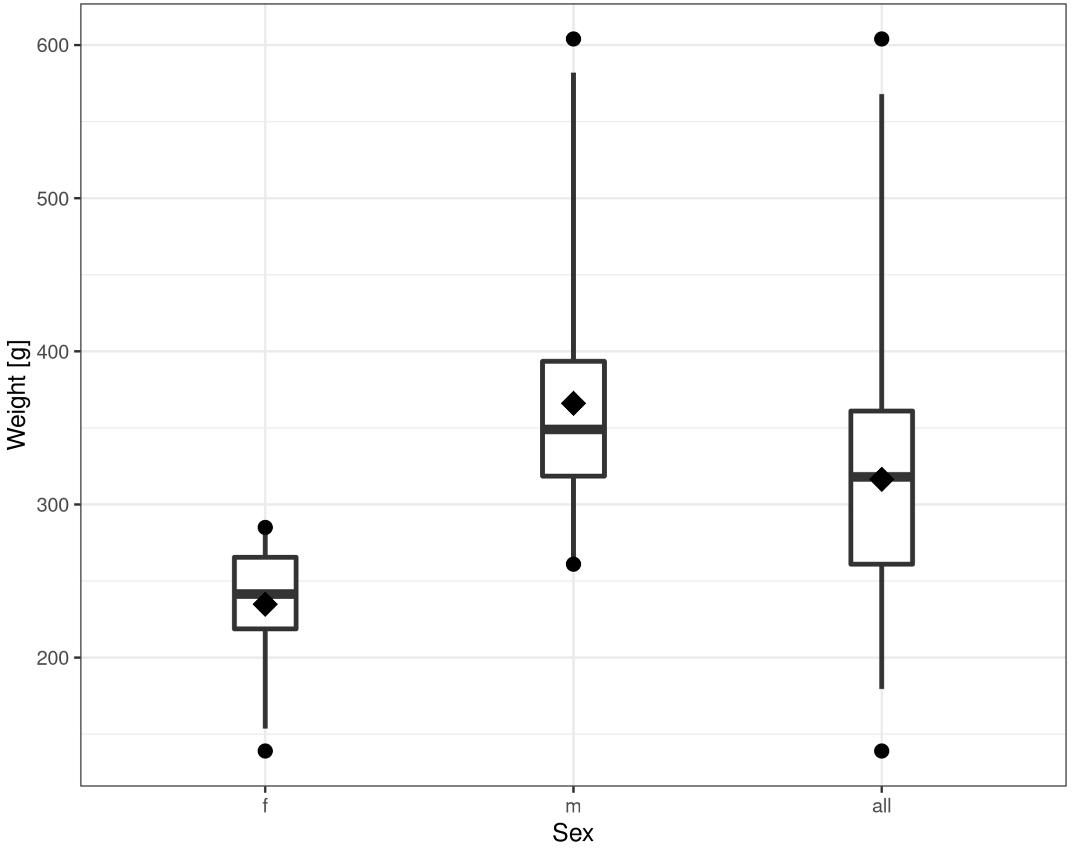
Application: descriptive analysis of data from hamster experiment – variable „age“



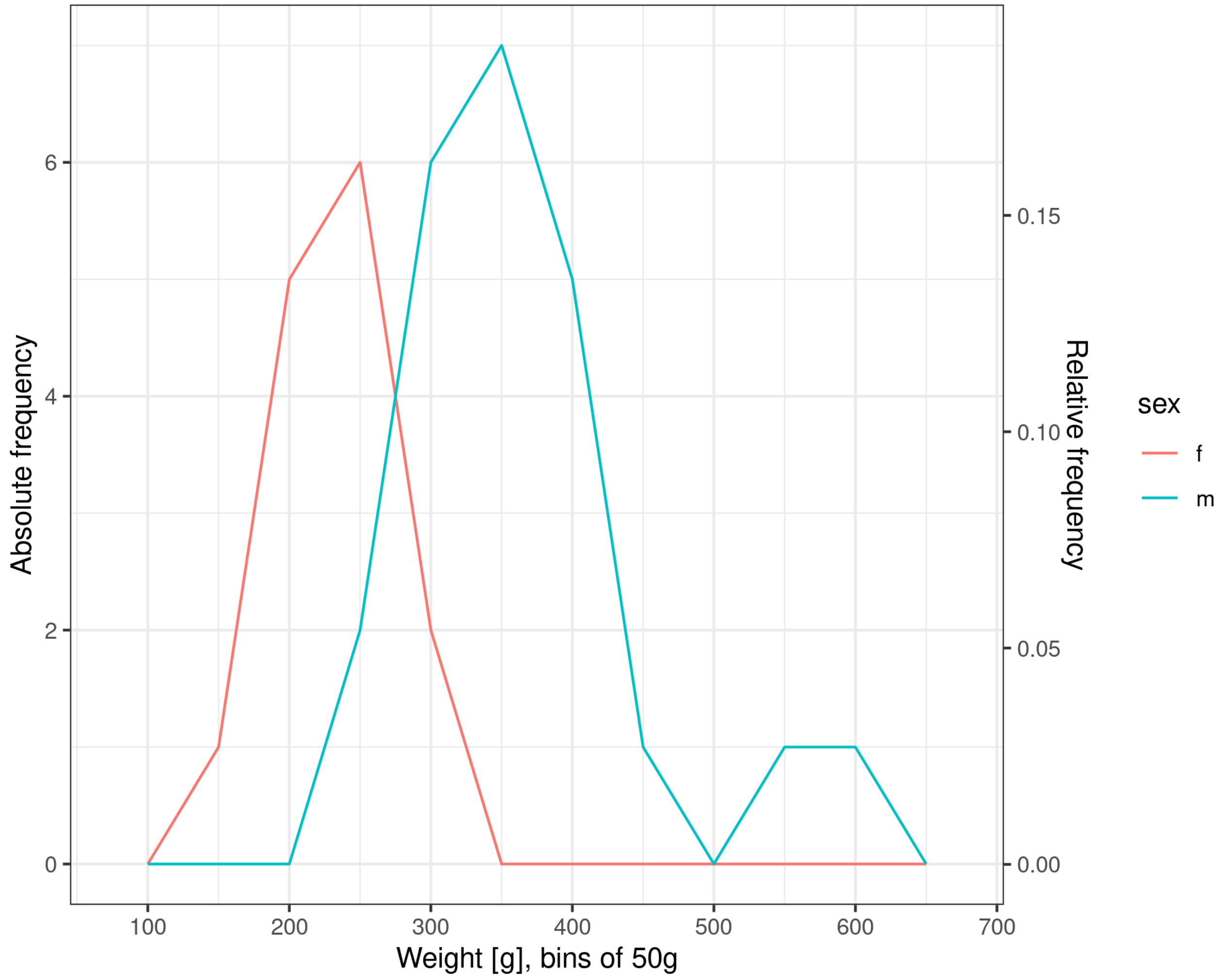


Application: descriptive analysis of data from hamster experiment – variable „weight“

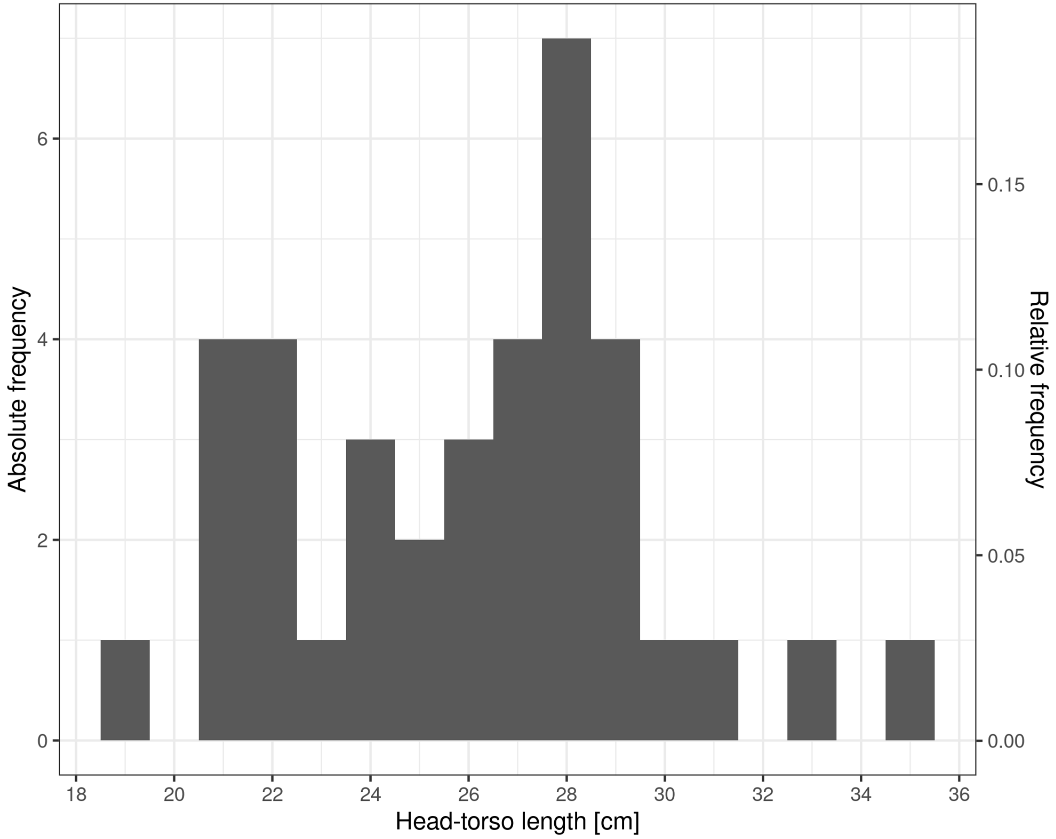


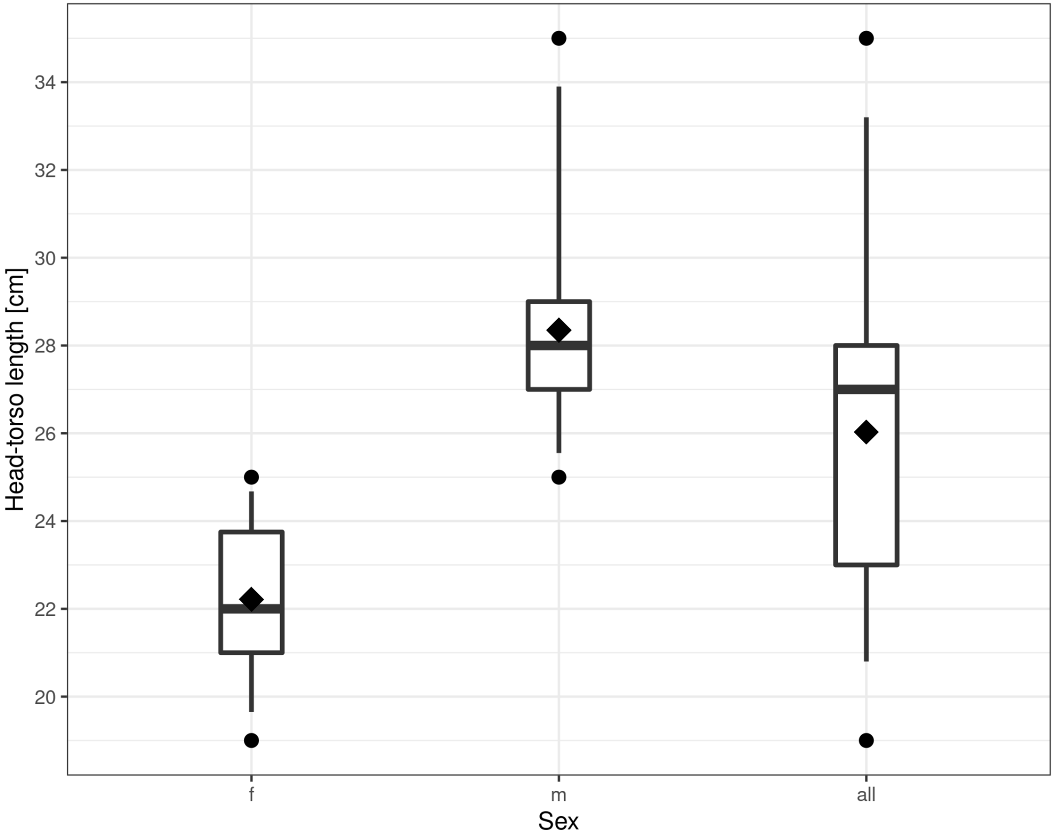


Application: descriptive analysis of data from hamster experiment – variable „weight“

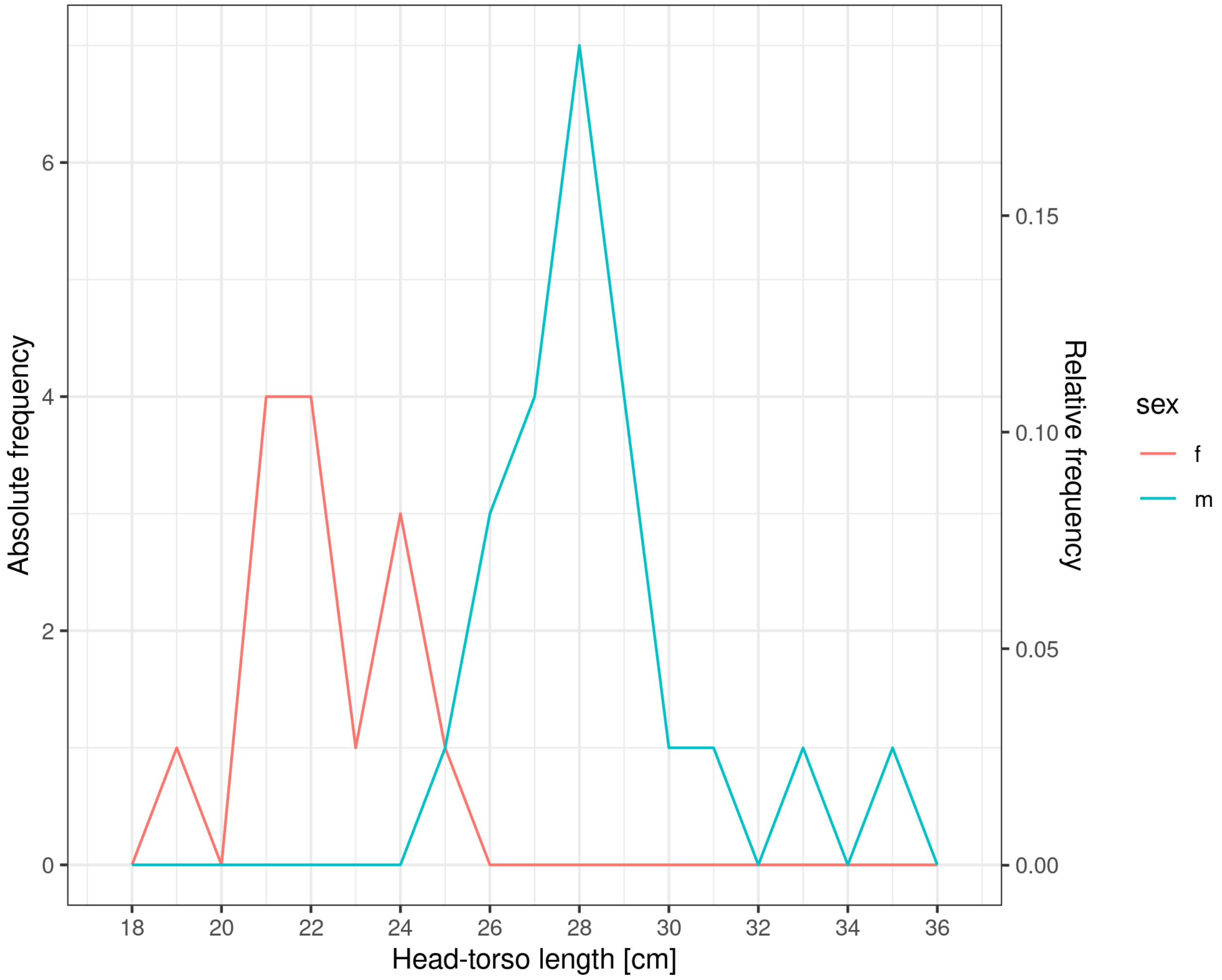


Application: descriptive analysis of data from hamster experiment – variable „length“

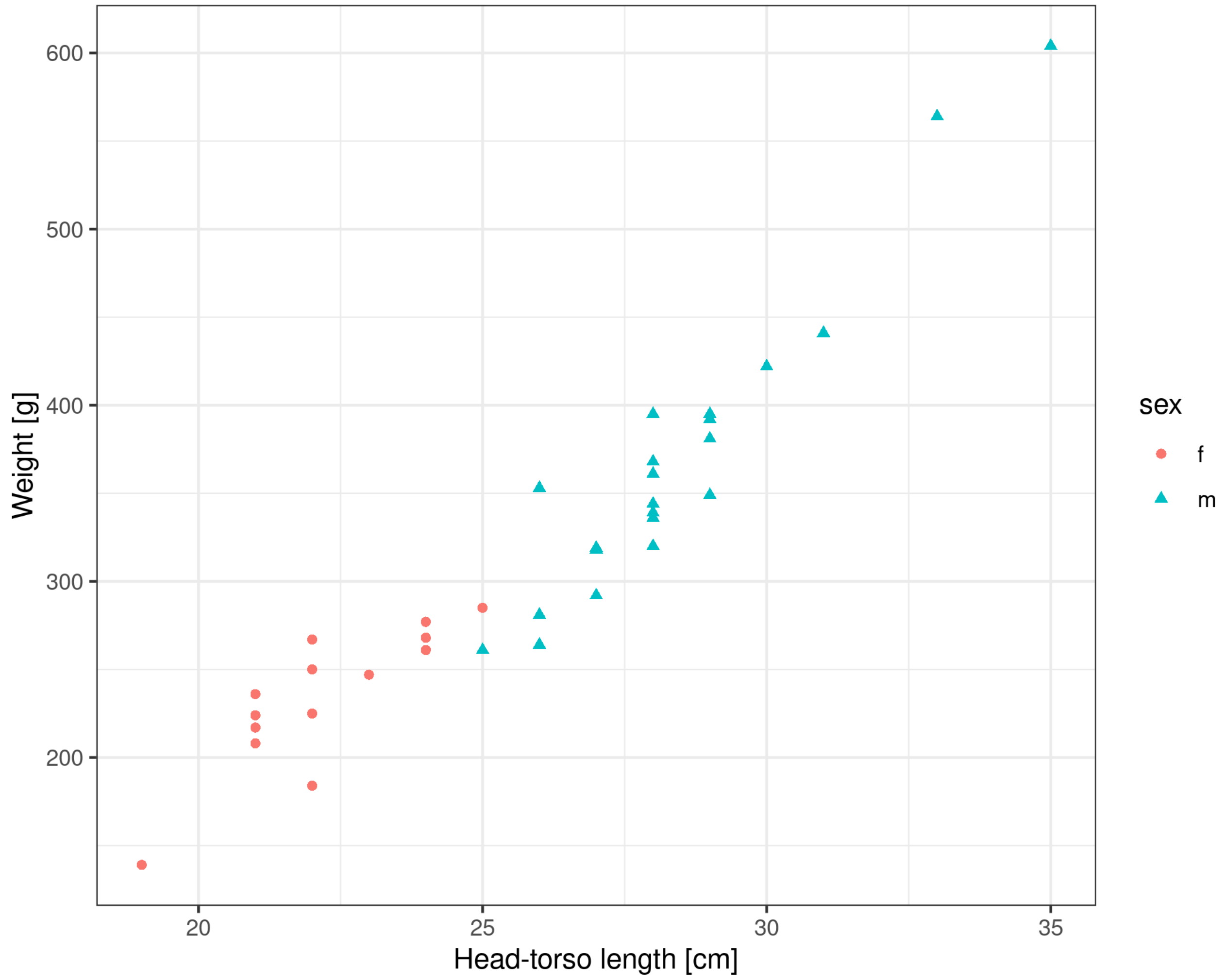




Application: descriptive analysis of data from hamster experiment – variable „length“



Application: descriptive analysis of data from hamster experiment – variables „length“ and „weight“



Take-Home-Messages I

(1) Descriptive analysis enables to condense the data by reporting a few meaningful summary measures and graphics.

(2) It is important to report both measures of location and measures of variation.

(3) The amount of information included in a variable is defined by its scale level. The scale level defines which measures and graphics are appropriate.

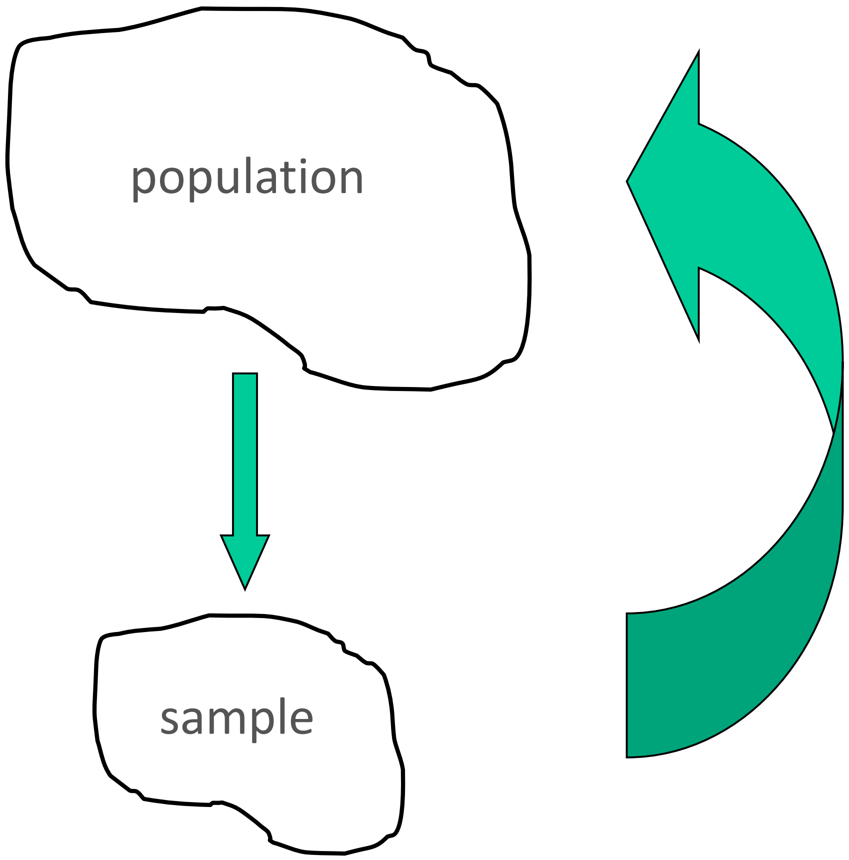
The conversation goes on …

- „It is really impressing how to get a quick overview on this bulk of data by looking at few measures and by inspecting appropriate graphics!“
- „Yes, indeed! And by the way: My sister was right. 23 of 37 hamsters are male, and only 14 are female. This is about 62% versus 38%, a clear majority for male hamsters.“
- „Oh no! This is only the result in our sample. In fact, the rate of male and female hamsters may actually be equal but I‘m sure that chance has lead to the observed result.“
- „How to decide whether the rate of male hamsters is actually larger than 50% or whether the observed result can be explained by chance?“

➢ This can be done by **confirmatory analysis / statistical tests**

Confirmatory analysis and statistical tests – principle and binomial test

The principle of confirmatory analysis / statistical tests



inference with
confirmatory analysis
(= statistical test)

descriptive analysis

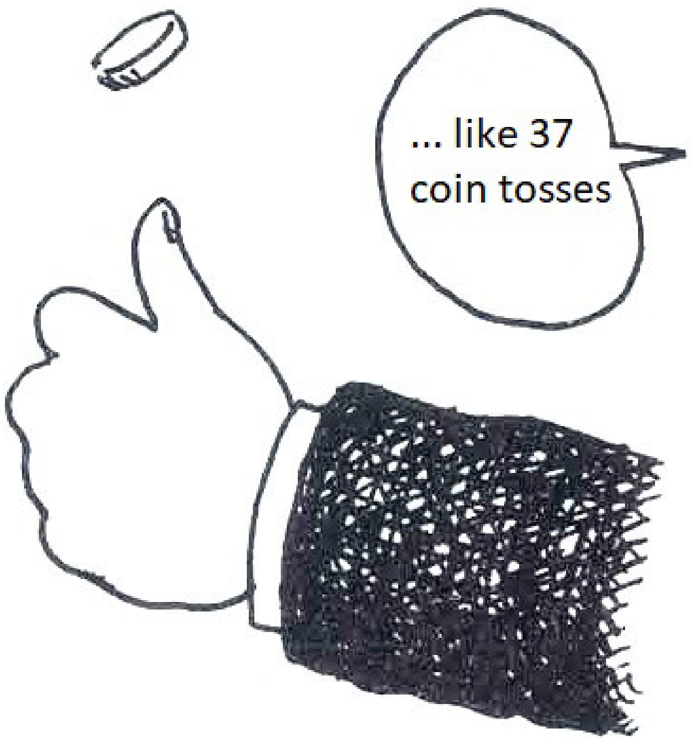
The hamster experiment

- We have 37 observations.
- Let us assume that the rate of male and female hamsters **in the population** is equal.
- In other words: The probability of a hamster to be male is equal to 50%.
- Then we would expect to observe about 18 or 19 male hamsters (and accordingly 19 or 18 female hamsters) **in the sample**.
- If we observe 20 or 21 male hamsters, would we then doubt on the equal rate of male and female hamsters?
- And what if we observe 23 male hamsters?

Confirmatory analysis / statistical tests give us rules when to decide that the pre-assumption (here: rate of male hamsters = 50%) is true and when to decide it is false (under control of the probability of making an errorneous decision; see below!)

The idea of statistical tests by means of the hamster experiment

- If the rate of male and female hamsters is equal in the population:
  What is the probability of observing 23 male hamsters in a sample of 37 hamsters?
- If this probability is too small, we don‘t believe any more in the premise (equal rates of male and female hamsters).



Larry Gonick und Woollcott Smith „The Cartoonguide to Statistics“

Ready, steady, go!

Quelle: www.gratis-bilder-download.de

What is the probability that …

… with **two tosses**, the coin shows **1 times „head“**?

- „head tail“ or „tail head“: 2 possibilities
- each possibility has probability $0.5 \cdot 0.5 = 0.25$

$\Rightarrow$ probability amounts to $2 \cdot 0.5 \cdot 0.5 = 0.5$

What is the probability that …

… with **three tosses**, the coin shows **2 times „head“**?

– „head head tail“ or „tail head head“ oder „head tail head“: 3 possibilities

– each possibility has probability $0.5 \cdot 0.5 \cdot 0.5 = 0.125$

➡ probability amounts to $3 \cdot 0.5 \cdot 0.5 \cdot 0.5 = 0.375$

What is the probability that …

… with **n tosses**, the coin shows **k times „head“**?

– number of possibilities „for n tosses k times „head“:

binomial coefficient

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots 3 \cdot 2 \cdot 1}{k! \cdot (n-k)!}$$

– each possibility has probability $p^k \cdot (1-p)^{n-k}$,

if for 1 toss the probability for „head“ is p

➡ probability for k times „head“ with n tosses („binomial probability“):

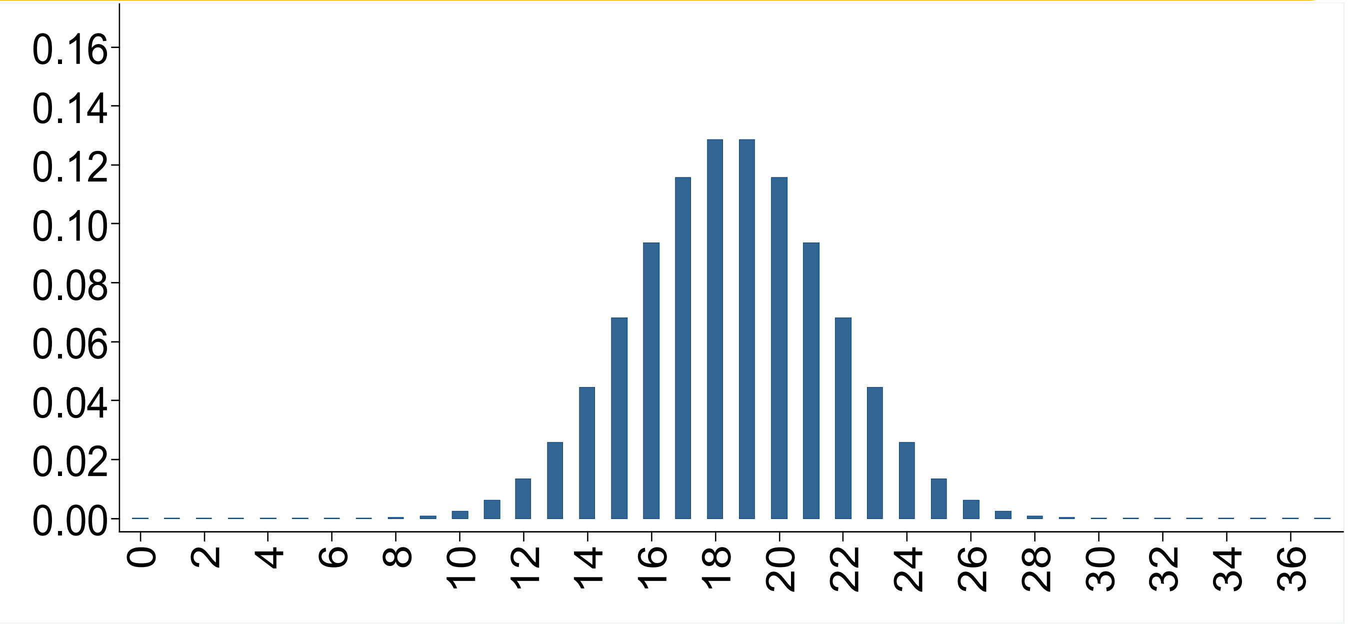
$$B(k|n,p) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

What is the probability that …

… with **37 tosses**, the coin shows **k times „head“** under the assumption that „head“ and „tail“ have equal probability, i.e. p=0.5?

binomial distribution for $p = 0.5$ and $n = 37$

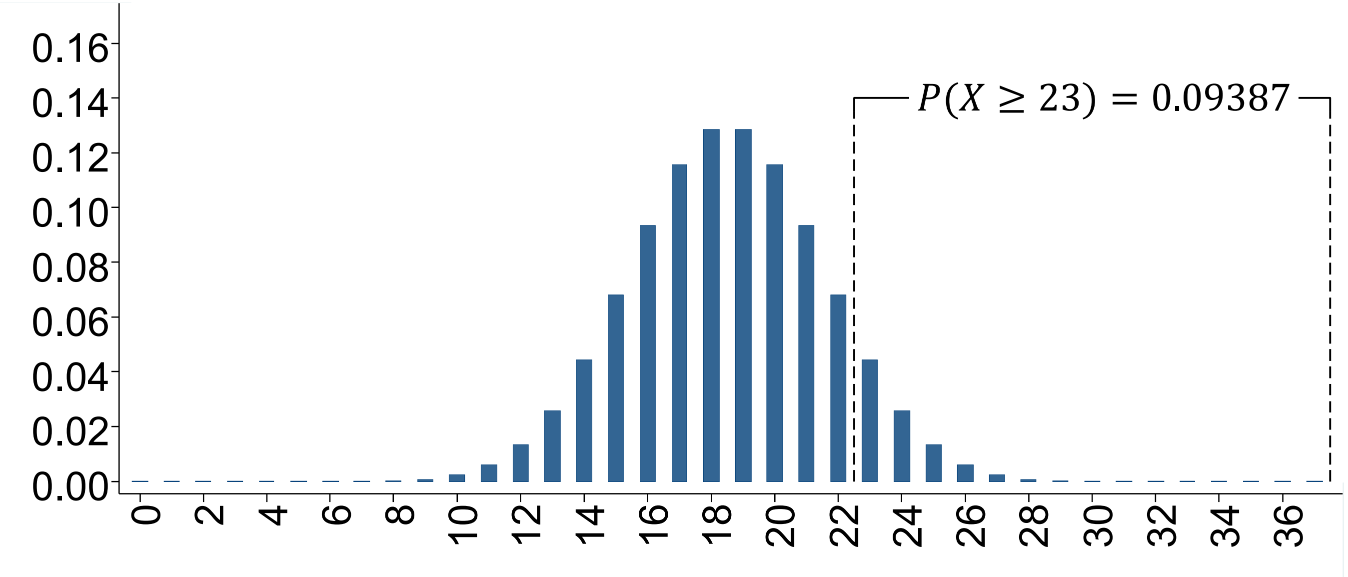
$$B(k|37, 0.5) = \binom{37}{k} \cdot 0.5^k \cdot (1-0.5)^{37-k} = \binom{37}{k} \cdot 0.5^{37}$$



What is the probability that ……

… with **37 tosses**, the coin shows **at least 23 times** (d.h. 23, 24, …, 36 or 37 times) „head“ (under the assumption that „head“ and „tail“ have equal probability, i.e. $p = 0.5$)?

➡ **probability=0.09387**



Result

The probability that, with 37 tosses, the coin shows at least 23 times „head“ – under the assumption that „head“ and „tail“ have equal probability – amounts to 0.09387.

Translated to the observation of 23 male hamsters in a sample of 37 hamsters:

The probability, that there are at least 23 male hamsters in a sample of 37 hamsters – under the assumption that male and female hamsters have equal probability – amounts to 0.09387.

Question

If we observe such a result …

… do we conclude

- that the ***assumption*** („male and female hamsters have equal probability“) ***is correct and the observation is due to chance***? (the probability of observing such a result under this assumption amounts to 0.09387)

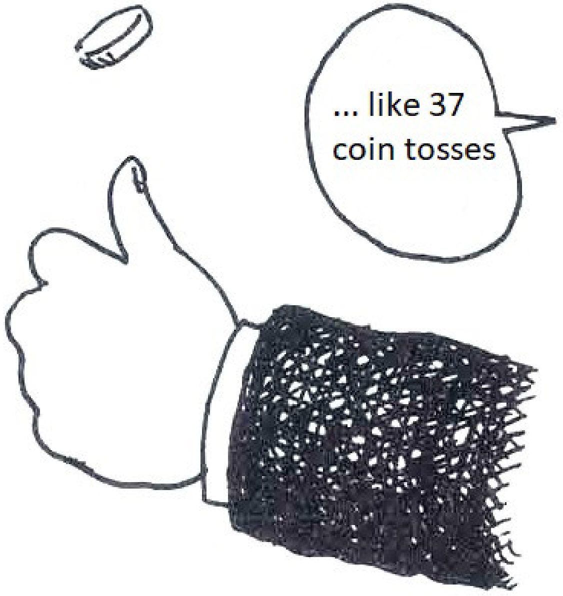
or do we conclude

- that the result is too unlikely and that we therfore ***do no longer believe in the assumption*** („male and female hamsters have equal probability“)?

Formally (in terms of hypotheses) – statistical test

null-hypothesis:

H_0: The probability of a hamster to be male is smaller or equal to 0.5.



alternative hypothesis:

H_1: The probability of a hamster to be male is larger than 0.5.

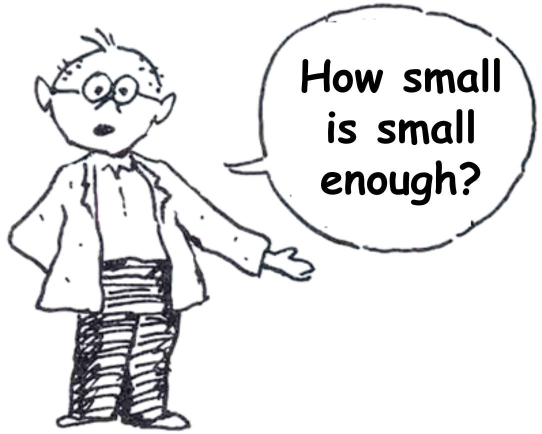
- If we assume that the null-hypothesis is true, …
- … but the probability of the observed result under this assumption is „too small“ …
- … then we do no longer believe in the assumption (that the null-hypothesis holds true) but decide for the alternative hypothesis.

Larry Gonick und Woollcott Smith „The Cartoonguide to Statistics“

Where is the threshold when we do no longer believe in the null-hypothesis?

This threshold is denoted by **significance level**.

If the null-hypothesis can be rejected, this is denoted as a **significant result**.



Larry Gonick und Woollcott Smith „The Cartoonguide to Statistics"

Where is the threshold when we do no longer believe in the null-hypothesis? (2)

In the scientific community, there is broad consensus to generally use **0.05** as a threshold („the magic 5%").

Therefore: If the **probability** (to measure the value observed in the sample or a value that points more extreme to the alternative hypothesis, given the null-hypothesis is true) **is smaller or equal to 0.05**, then the **null-hypothesis is rejected**.

Larry Gonick und Woollcott Smith „The Cartoonguide to Statistics"

The principle of statistical tests

step 1: formulation of null- and alternative hypothesis

H_0: The probability of a hamster to be male is smaller or equal to 0.5

H_1: The probability of a hamster to be male is larger than 0.5

step 2: collect the data

here: data of hamsters' gender

step 3: find the **appropriate probability distribution**

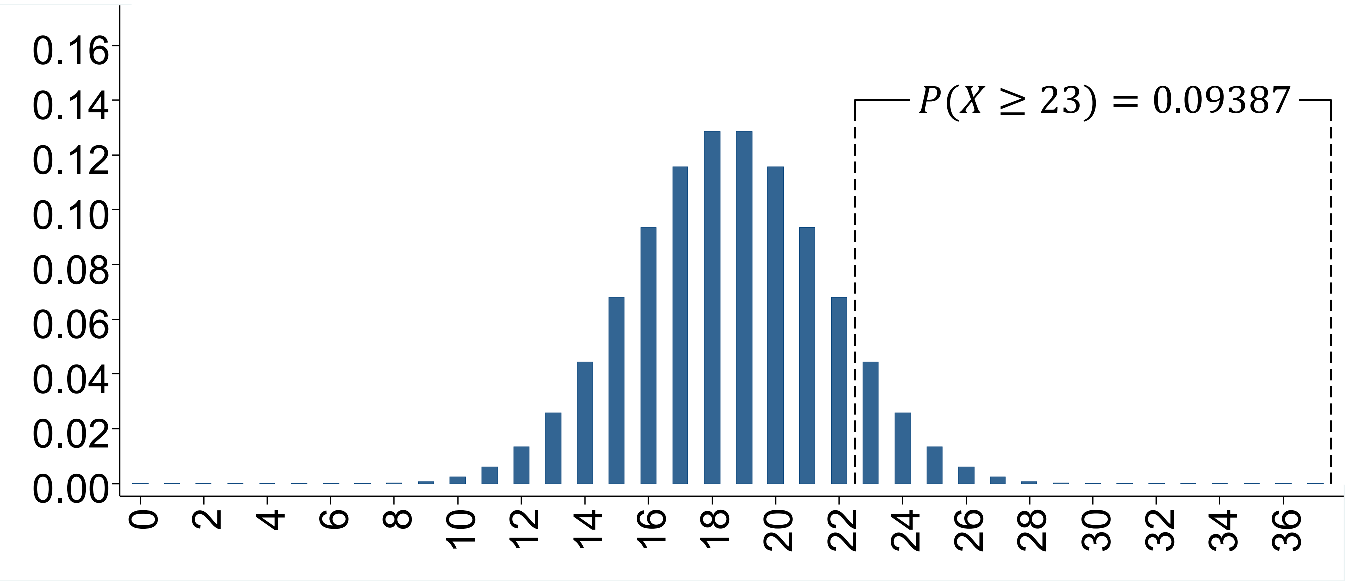
here: binomial distribution

step 4: compute the probability by using the distribution from step 3 (measuring the value observed in the sample or a value that points more extreme to the alternative hypothesis, given the null-hypothesis from step 1 is true).

here: = 0.09387 ***= p-value***

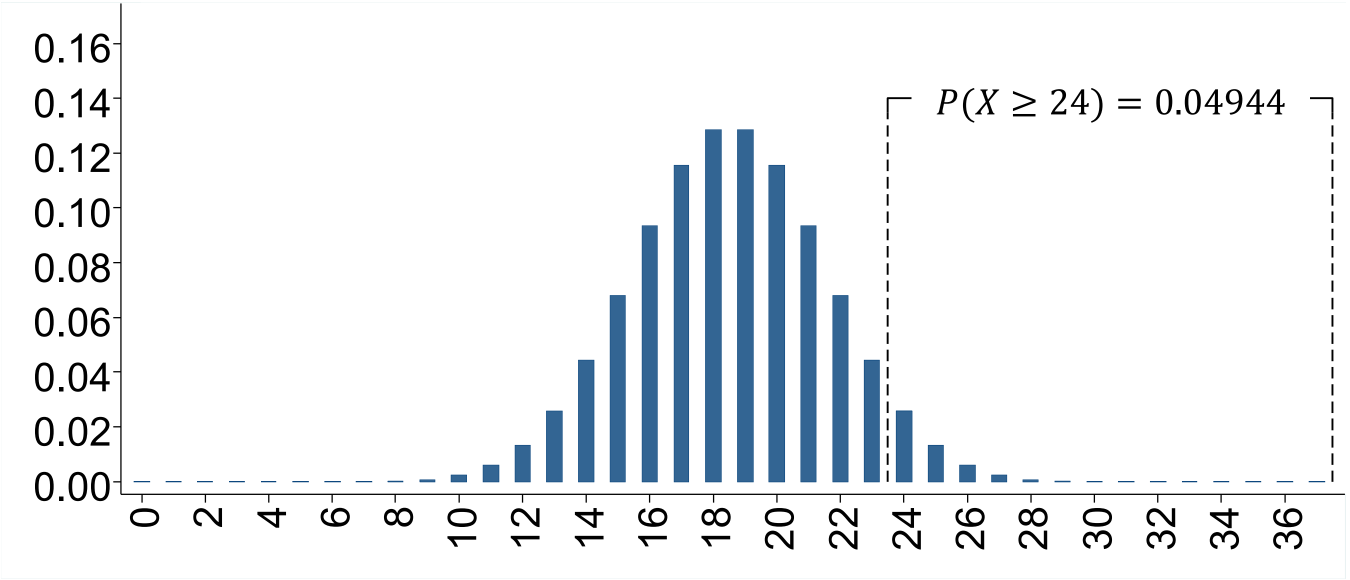
Step 4: test decision

Decide, whether the data (step 2) are compatible with the null-hypothesis (step 1) or not:



- p-value=0.09387 > 0.05
- ➢ **The null-hypothesis cannot be rejected**.
- ➢ „not enough evidence to decide in favour of a higher rate of male hamsters“

And what if we would have observed 24 male hamsters?

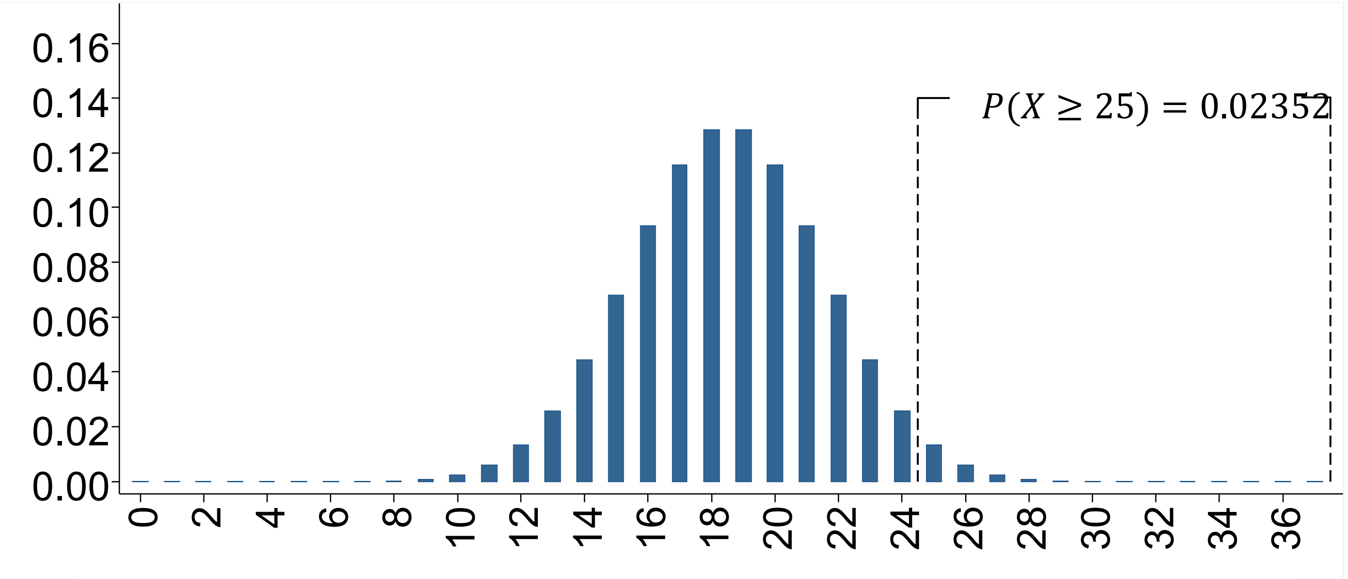


- p-value=0.04944 < 0.05

➢ **The null-hypothesis can be rejected**.

➢ „enough evidence to decide in favour of a higher rate of male hamsters“

And what if we would have observed 25 male hamsters?



- p-value=0.02352 < 0.05
- ➢ **The null-hypothesis can be rejected**.
- ➢ „enough evidence to decide in favour of a higher rate of male hamsters“

Calculating the p-values with the (free) software R

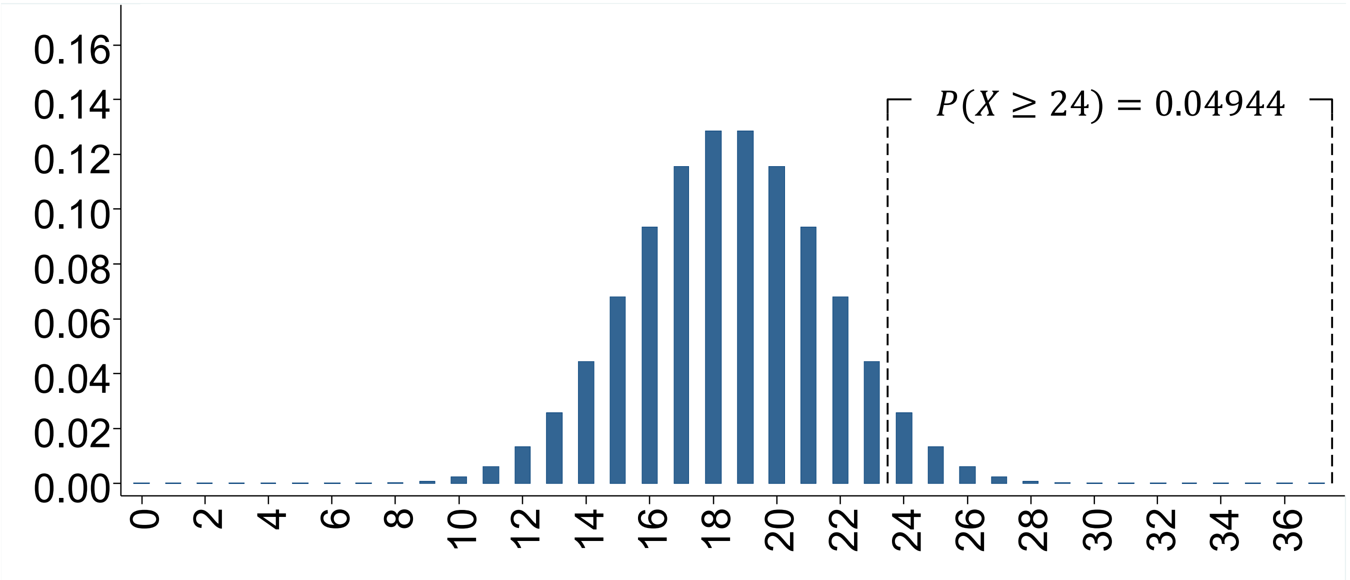
The R-function „pbinom“ calculates the probability P(X>q). Therefore, you need to subtract 1 from the boundary q in order to obtain P(X>=q) to reproduce the above results:

```
> # P(X >= 23) = P(X > 22)
> pbinom(q=22,size=37,prob=0.5,lower.tail=FALSE)
[1] 0.09387078
> # P(X >= 24) = P(X > 23)
> pbinom(q=23,size=37,prob=0.5,lower.tail=FALSE
[1] 0.04943587
> # P(X >= 25) = P(X > 24)
> pbinom(q=24,size=37,prob=0.5,lower.tail=FALSE)
[1] 0.02351551
```

Can false decisions („false alarm“) occur?

Clipart http://office.microsoft.com

Type I error



- If the null-hypothesis is in fact true, we observe with probability 0.04944 at least 24 times „head“ with 37 tosses. Then, we (falsely!) reject the null-hypothesis.
- In other words: If the null-hypothesis is true, we decide with probability 0.04944 falsely that the alternative hypothesis is true.
- In other words: The probability for a type I error is equal to 0.04944.

Type I and type II error

„fire brigade“	no fire	fire
no alarm	no error	type II error
alarm	type I error	no error

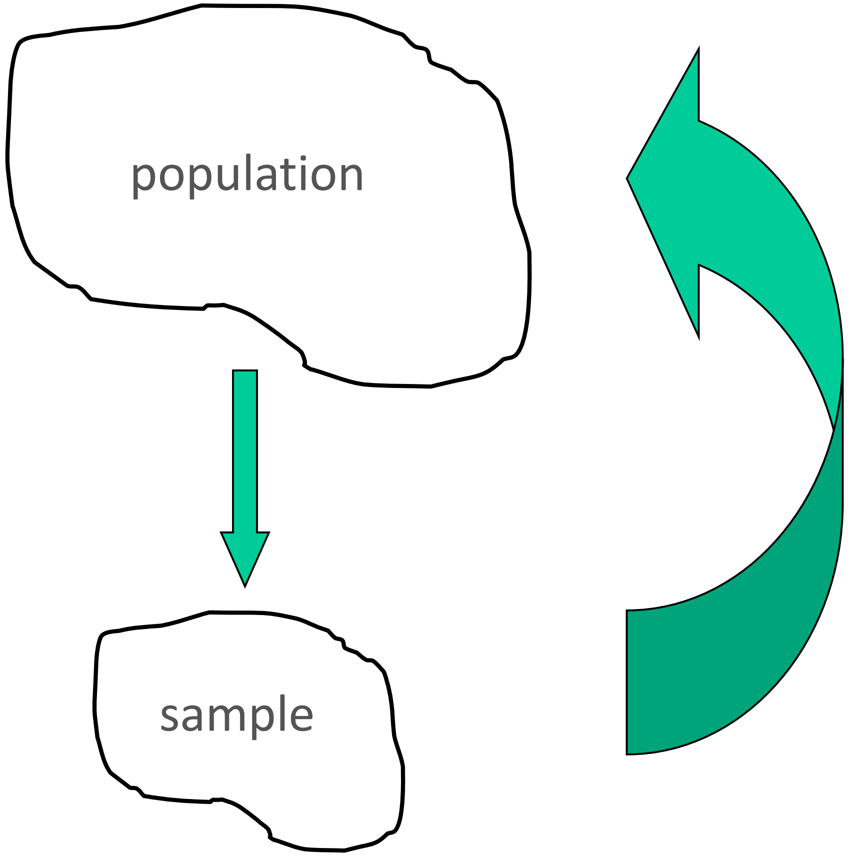
Type I and type II error

„fire brigade“	no fire	fire
no alarm	no error	type II error
alarm	type I error	no error

statistical test	null-hypothese is true	alternative hyp. is true
decision for H_0	no error	type II error
decision against H_0	type I error	no error

imbi

The principle of confirmatory analysis / statistical tests



inference with
confirmatory analysis
(= statistical test)

descriptive analysis

Take-Home-Messages II

(1) Confirmatory analysis enables to decide about hypotheses (which are formulated for the *population*) based on data observed in a *sample*.

(2) By use of a statistical test, it can be judged whether or not a result is compatible with chance.
If the observed result is „too unlikely" to be judged as chance (assuming that the null-hypothesis is true), one rejects the null-hypothesis and decides for the alternative hypothesis.

(3) The null-hypothesis is rejected, if the p-value of the result is smaller or equal than the pre-specified type I error rate (commonly chosen as 0.05).

Exercise: octopus Paul

- **Is octopus Paul able to predict the future?**

Quelle: https://www.ibtimes.com/octopus-made-better-world-cup-predictions-goldman-sachs-photos-1613882

Exercise: octopus Paul (2)

In the Sea Life Centre in Oberhausen, Paul acted as octopus oracle for predicting the outcomes of games during soccer world championship 2010. Some days before each game, two equal boxes out of acrylic glass were put into the aquarium. The boxes contained water and a blue mussel. The sides of the observer were laminated with the ensigns of the soccer national teams of the respective play. Paul‘s choice of the food was taken as the prediction of the winner. Eight times in series, Paul chose the correct ensign. (translated from Wikipedia, 17.11.2010).

competitor	round	prediction	result (German view)	correctness of „prediction“
Australia	group phase	Germany	4:0	yes
Serbia	group phase	Serbia	0:1	yes
Ghana	group phase	Germany	1:0	yes
UK	round of 16	Germany	4:1	yes
Argentina	quarter-final	Germany	4:0	yes
Spain	half-final	Spain	0:1	yes
Uruguay	game for third	Germany	3:2	yes
			World Cup final	
Netherlands-Spain		Spain	0:1	yes

Exercise: octopus Paul (3)

- **Is octopus Paul able to predict the future?**
  - assumption: the number of guesses was pre-fixed (n=8)
  - null- and alternative hypothesis?
  - statistical test at type I error rate $\alpha = 0.05$?
  - result?

$\rightarrow$ hint: for $X \sim B(8, 0.5)$ it holds true that

$P(X = 0) = P(X = 8) = 0.00391$; $P(X = 1) = P(X = 7) = 0.03125$

$P(X = 2) = P(X = 6) = 0.10938$; $P(X = 3) = P(X = 5) = 0.21875$

$P(X = 4) = 0.27344$

Solution to exercise: octopus Paul

- null-hypothesis: $p \leq 0.5$
- alternative hypothesis: $p > 0.5$ (Paul is better than „guessing“)
- statistical test at type I error rate $\alpha = 0.05$:
  reject the null-hypothesis, if $X \geq 7$
  (as $P(X \geq 7) = 0.03125 + 0.00391 = 0.03516 < 0.05$)
- alternatively: the p-value is given by $P(X \geq 8) = 0.00391 < 0.05$
- result: the null-hypothesis can be rejected.
  hence: „Paul is able to predict the future!“ (or a type I error has occurred).

$\rightarrow$ remember: for $X \sim B(8, 0.5)$ it holds true that

$P(X = 0) = P(X = 8) = 0.00391; P(X = 1) = P(X = 7) = 0.03125$

$P(X = 2) = P(X = 6) = 0.10938; P(X = 3) = P(X = 5) = 0.21875$

$P(X = 4) = 0.27344$

Point estimation and confidence intervals – principle and the case of rates

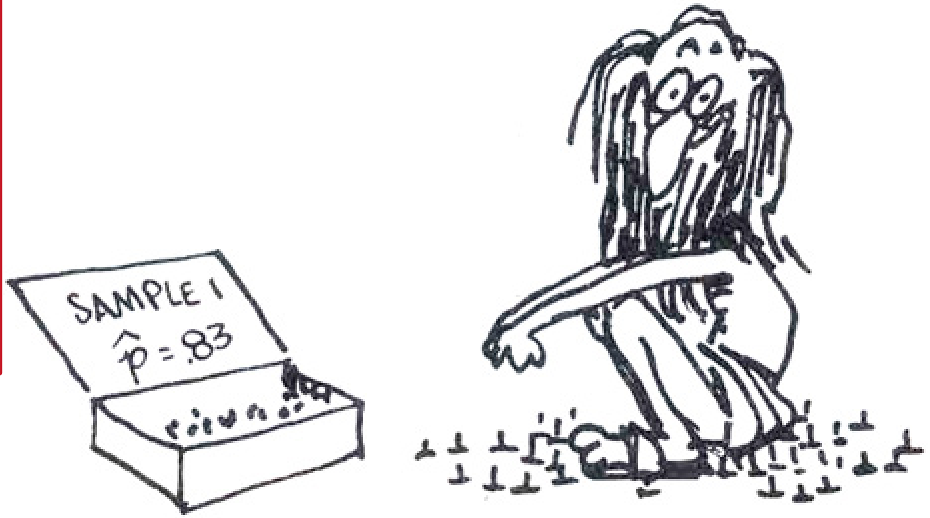
Point estimation

Statement: We now know from the result of the statistical test that the probability for a hamster to be male is presumably not larger than 0.5 (more precisely: it is either in fact not larger than 0.5 or a type II error has been committed).

Question: What is the value of the probability for a hamster to be male?

Point estimation:
Based on the data of the sample („hamster experiment“) we obtain an estimate for p:

$$\hat{p} = \frac{23}{37} = 0.62$$



Larry Gonick and Woollcott Smith „The Cartoonguide to Statistics“

What would happen, if …

… we repeat the experiment with another sample of 37 hamsters?

… we repeat the experiment with another 3700 hamsters?

Would we again observe 23 male hamsters (or 2300, respectively)?

Larry Gonick and Woollcott Smith „The Cartoonguide to Statistics“

A random experiment

- Let us assume that the actual probability for a hamster to be male is 0.5, i.e. $p = 0.5$
- We randomly repeat the „hamster experiment“ with samples of 37 hamsters and obtain for each experiment a „point estimate“:

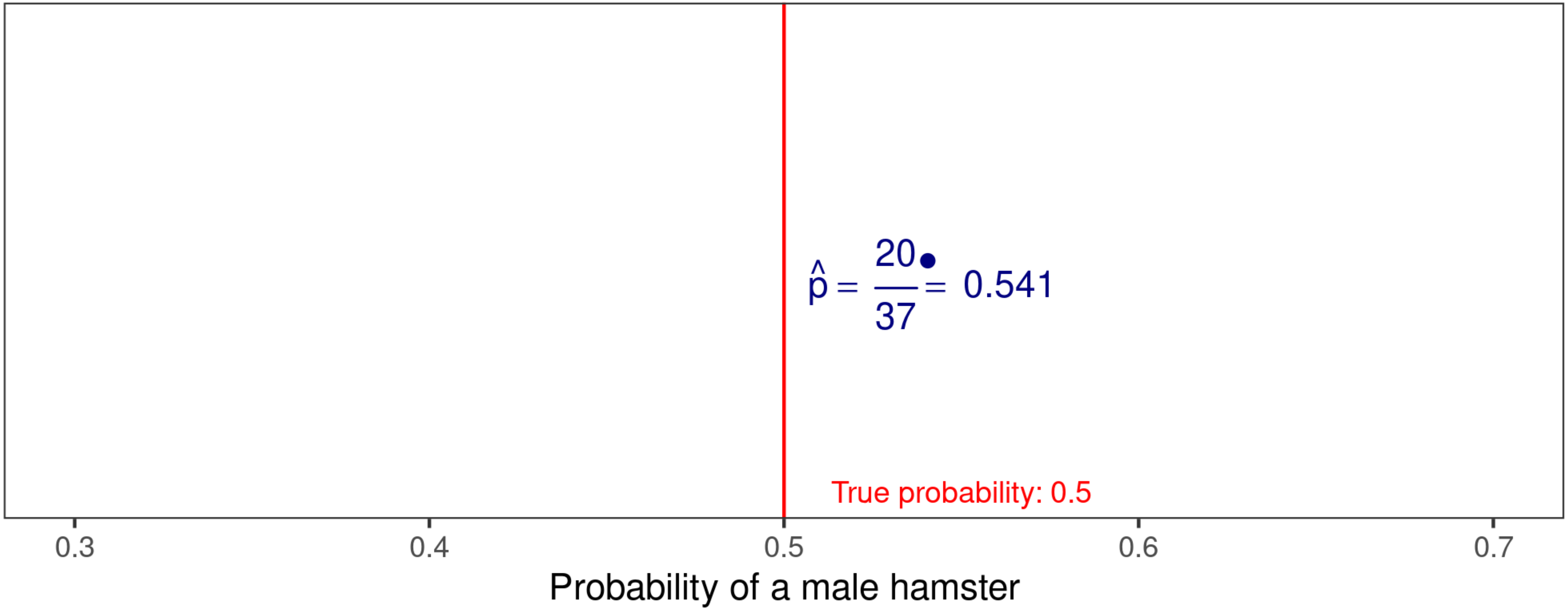
$$\hat{p} = \frac{\#male\ hamsters}{37}$$

Larry Gonick and Woollcott Smith „The Cartoonguide to Statistics“

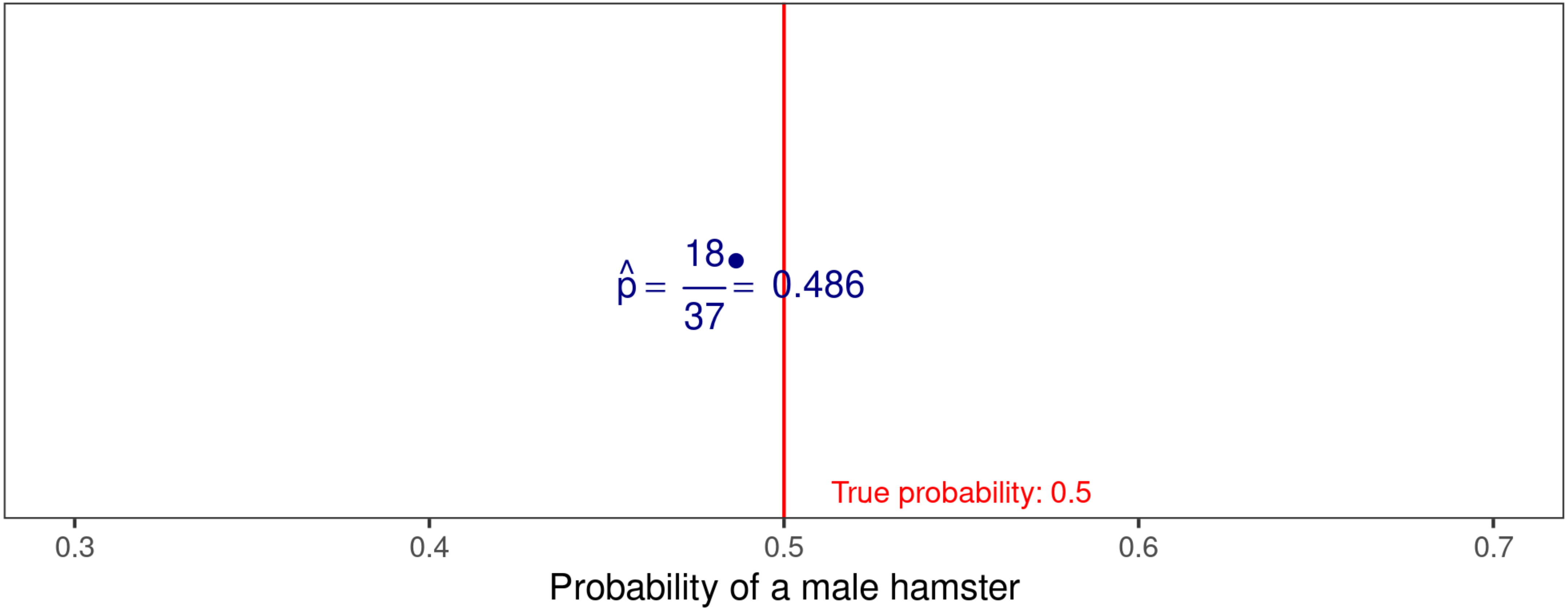
Ready, steady, go!

Quelle: www.gratis-bilder-download.de

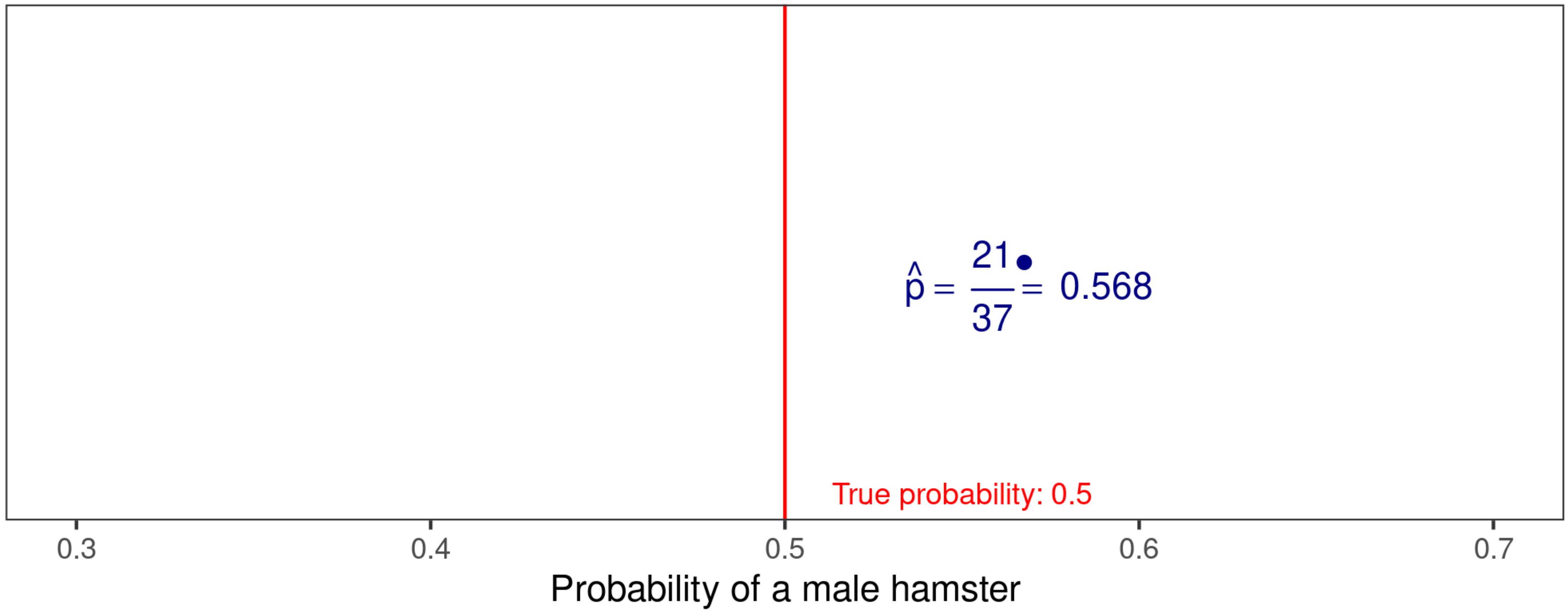
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 1



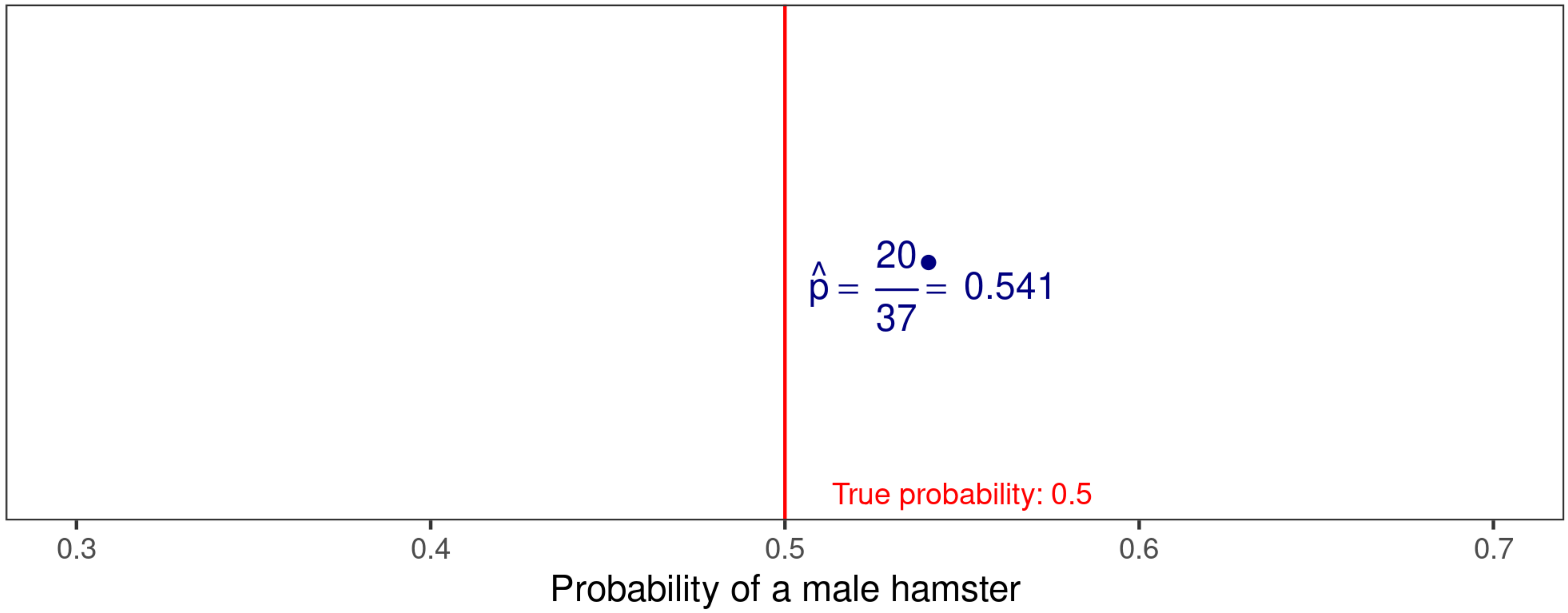
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 2



A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 3

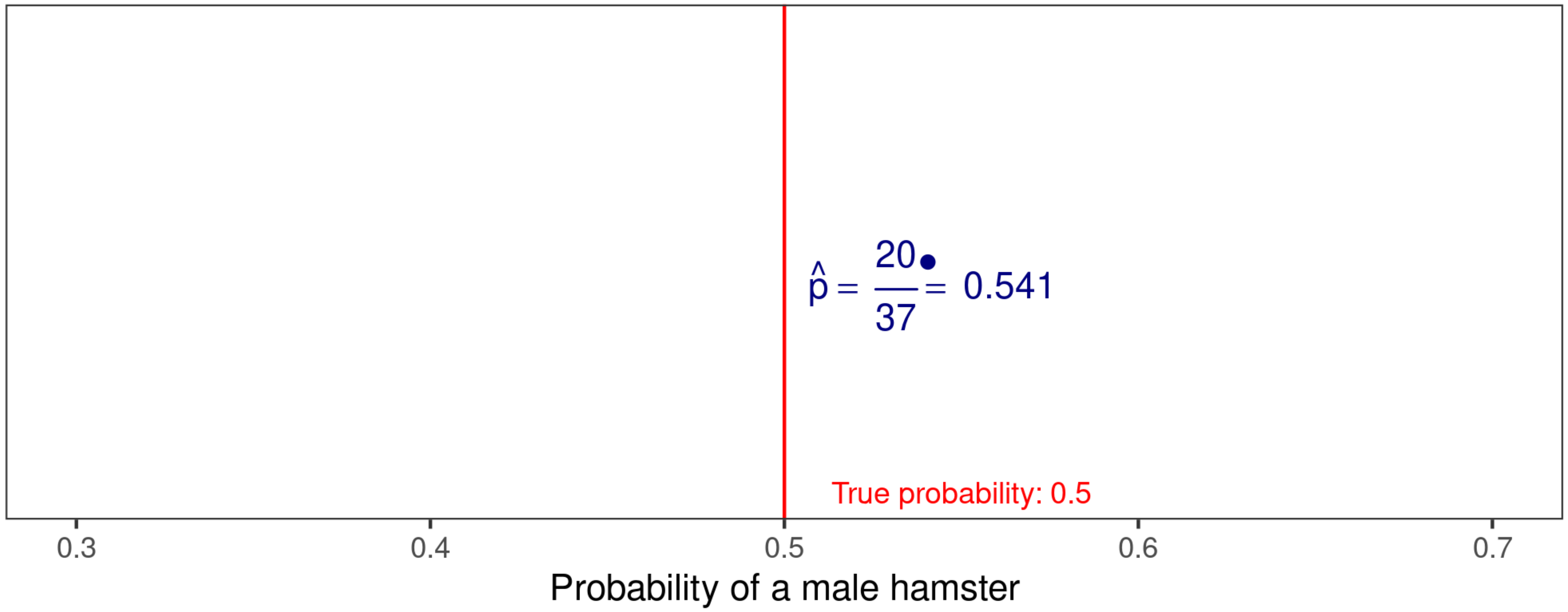


A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 4

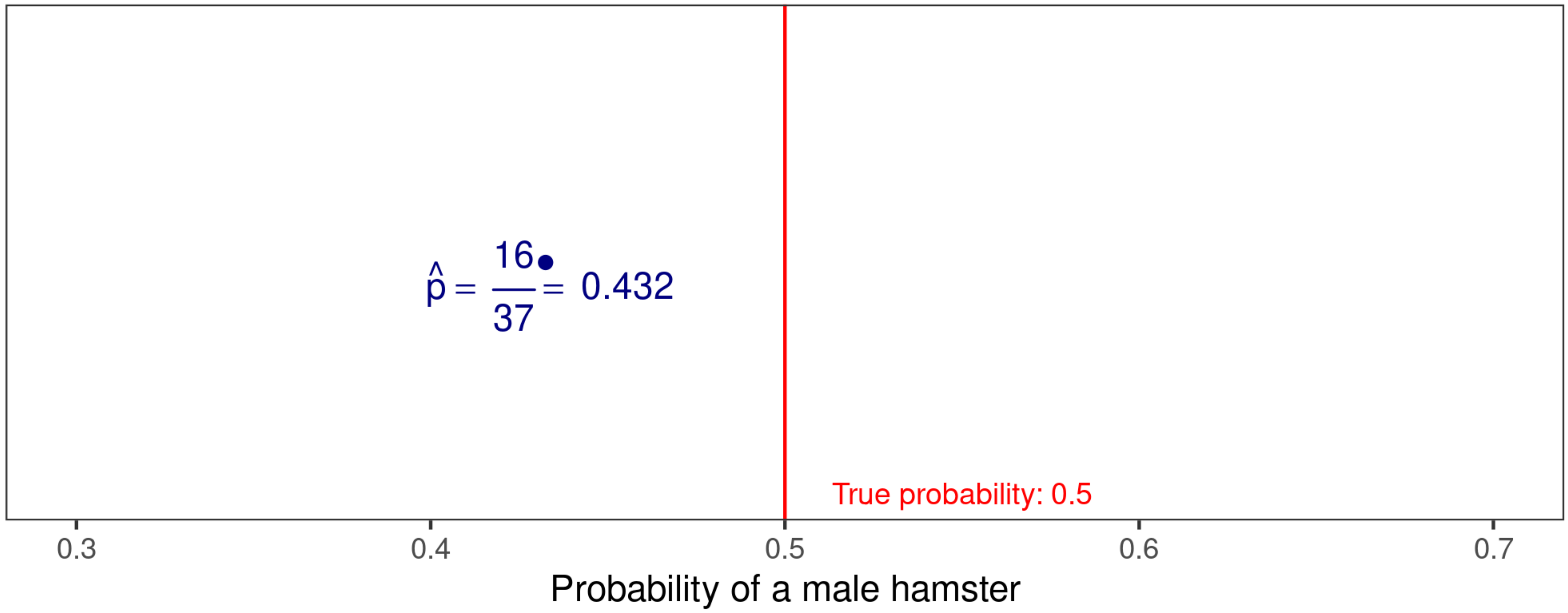


imbi

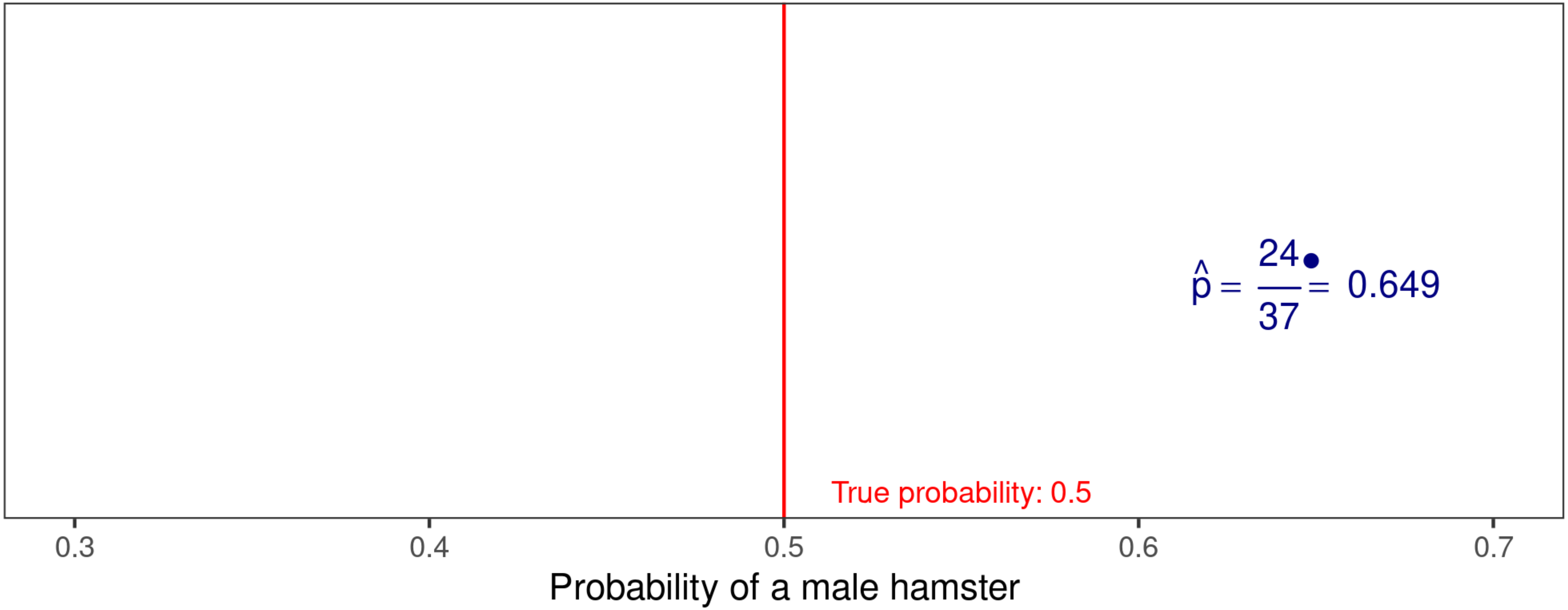
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 5



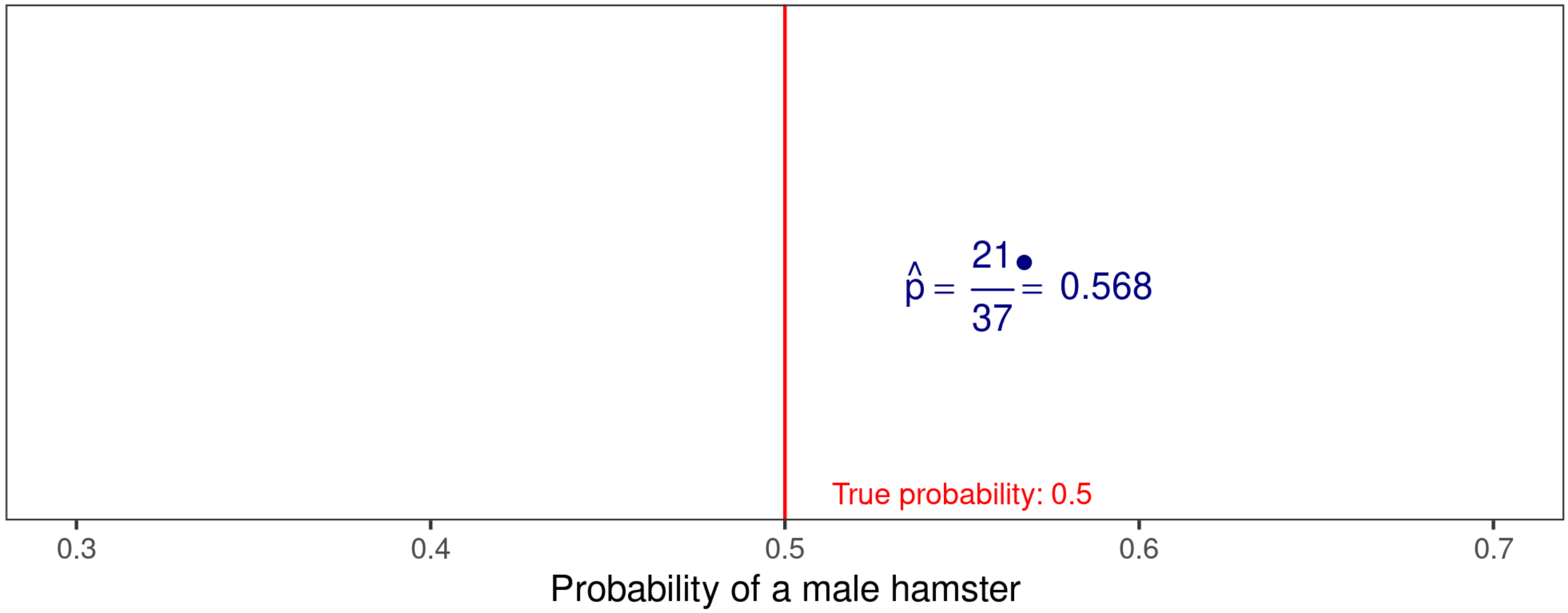
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 6



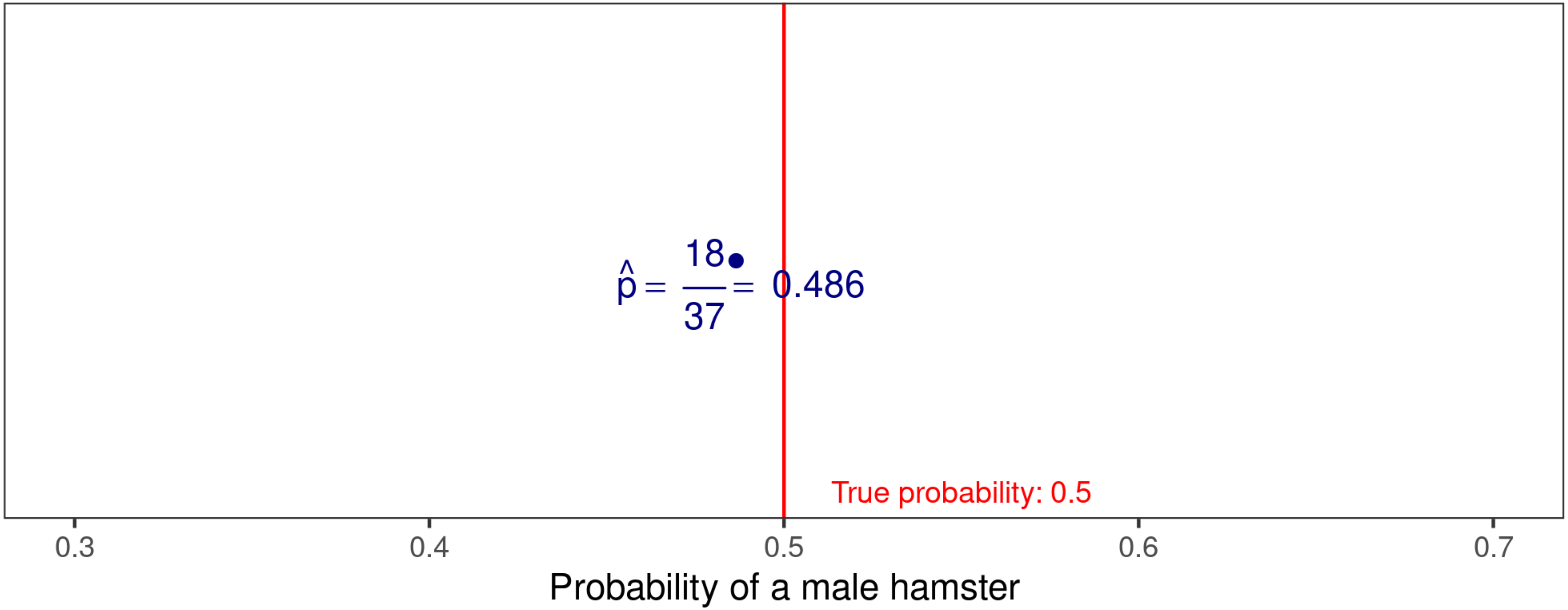
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 7



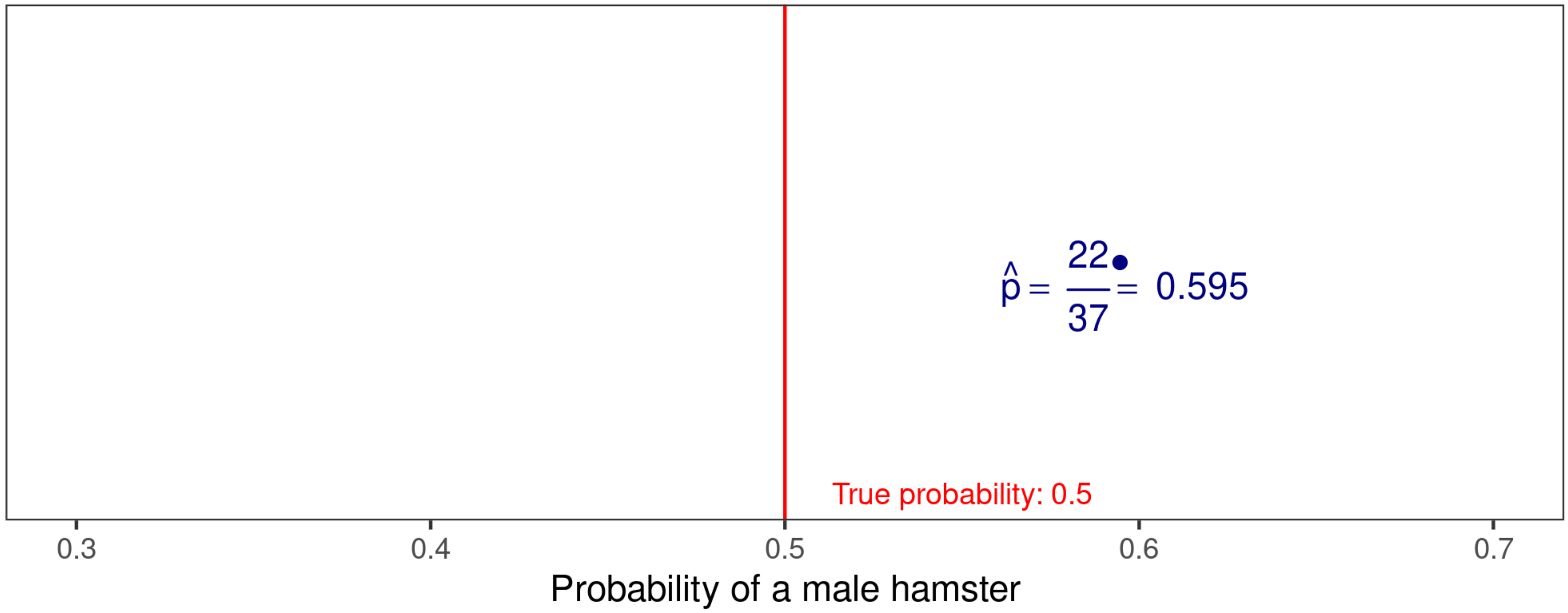
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 8



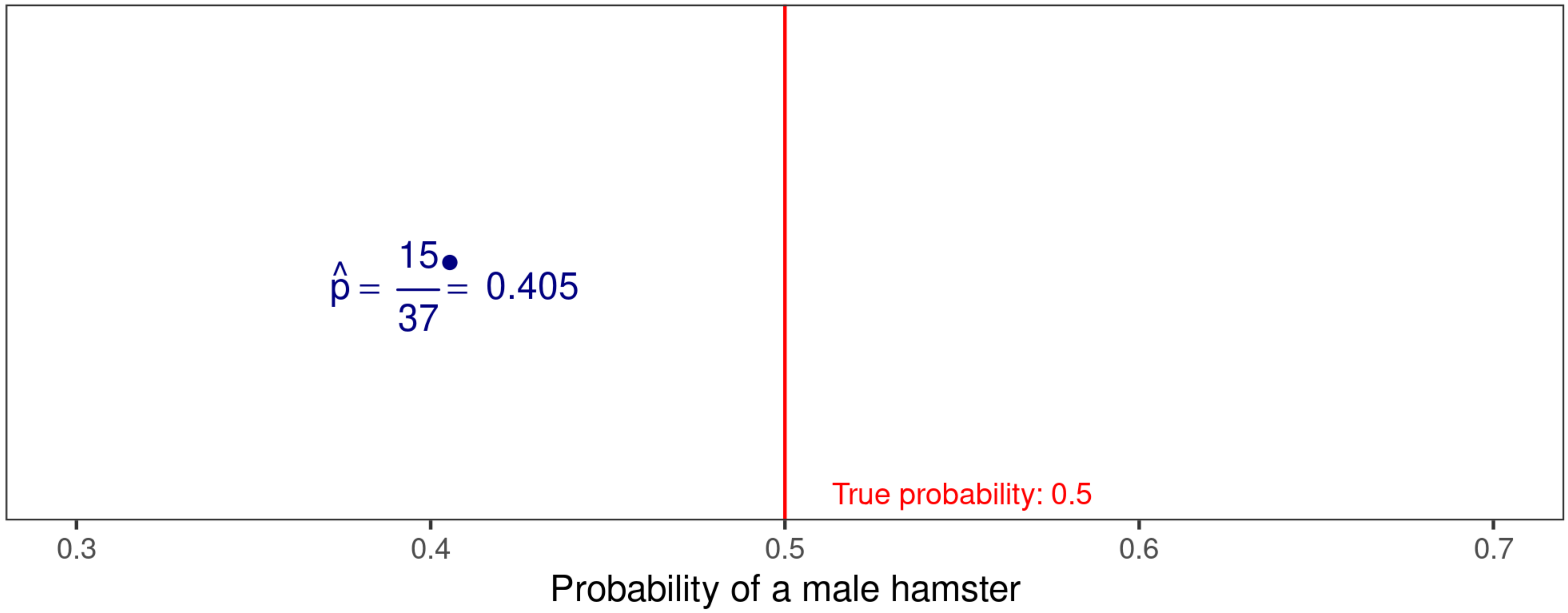
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 9



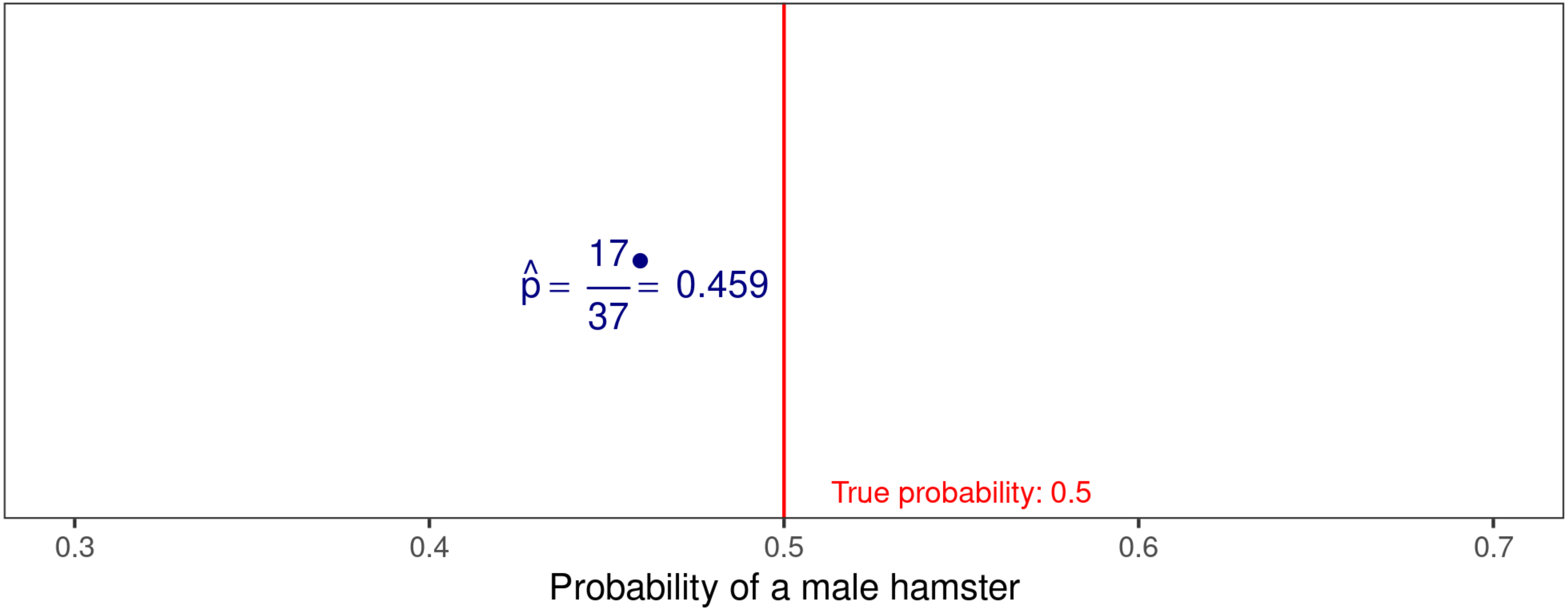
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 10



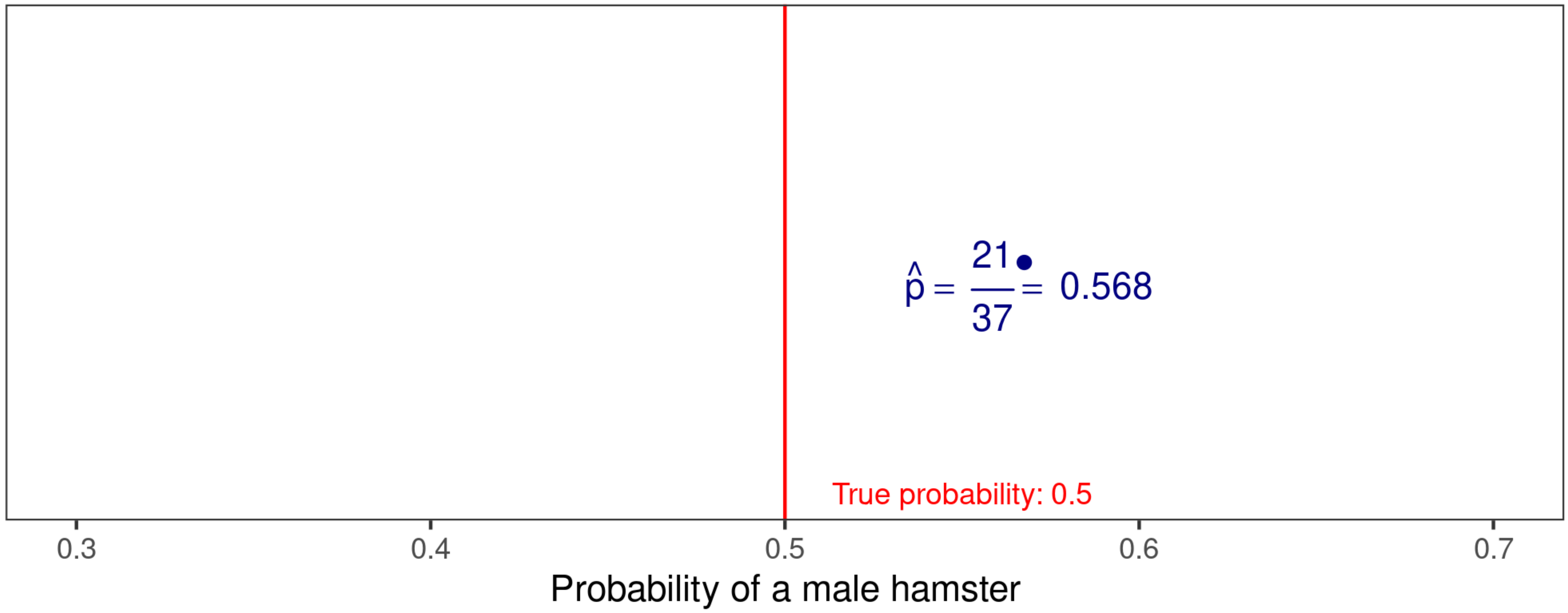
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 11



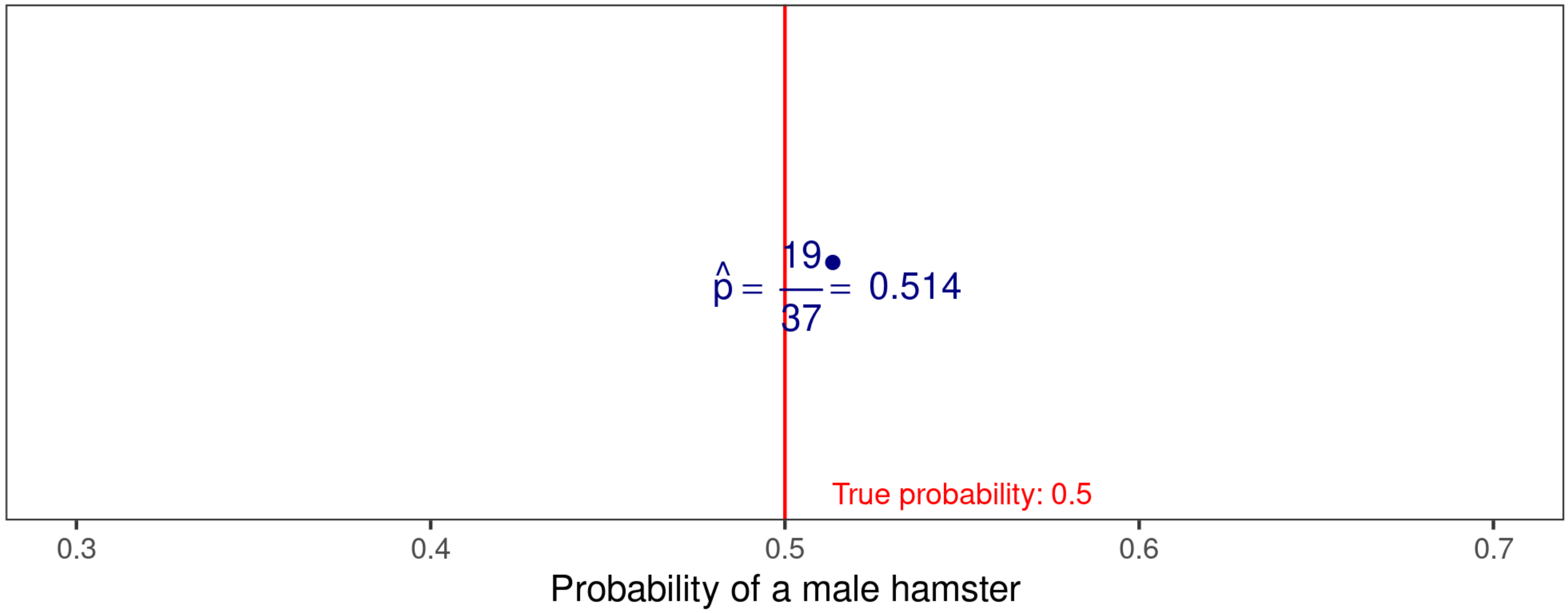
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 12



A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 13

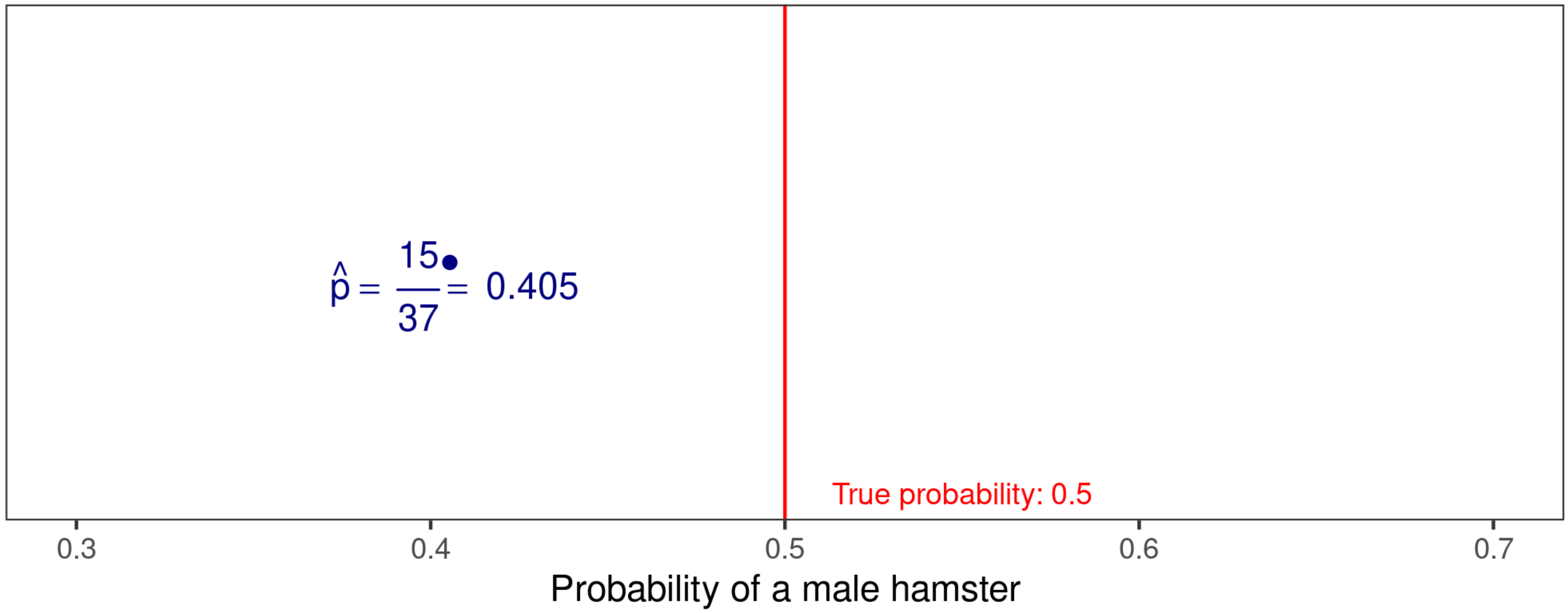


A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 14

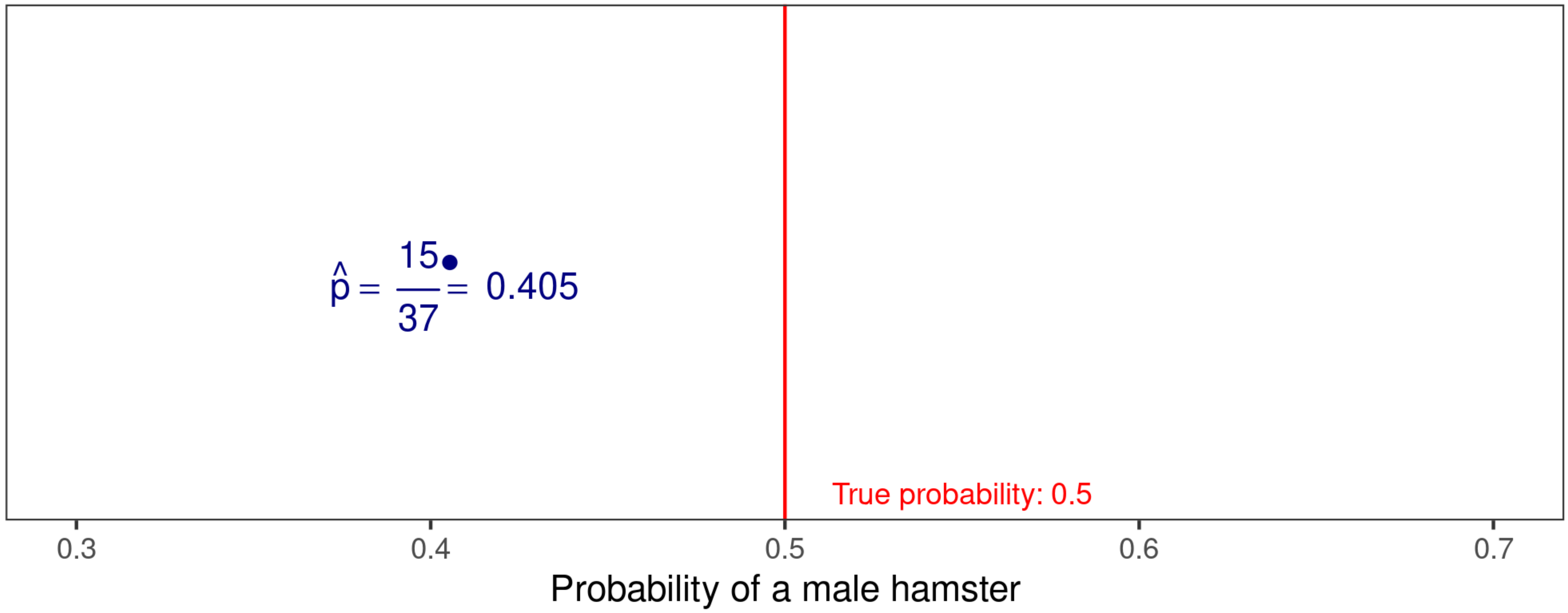


imbi

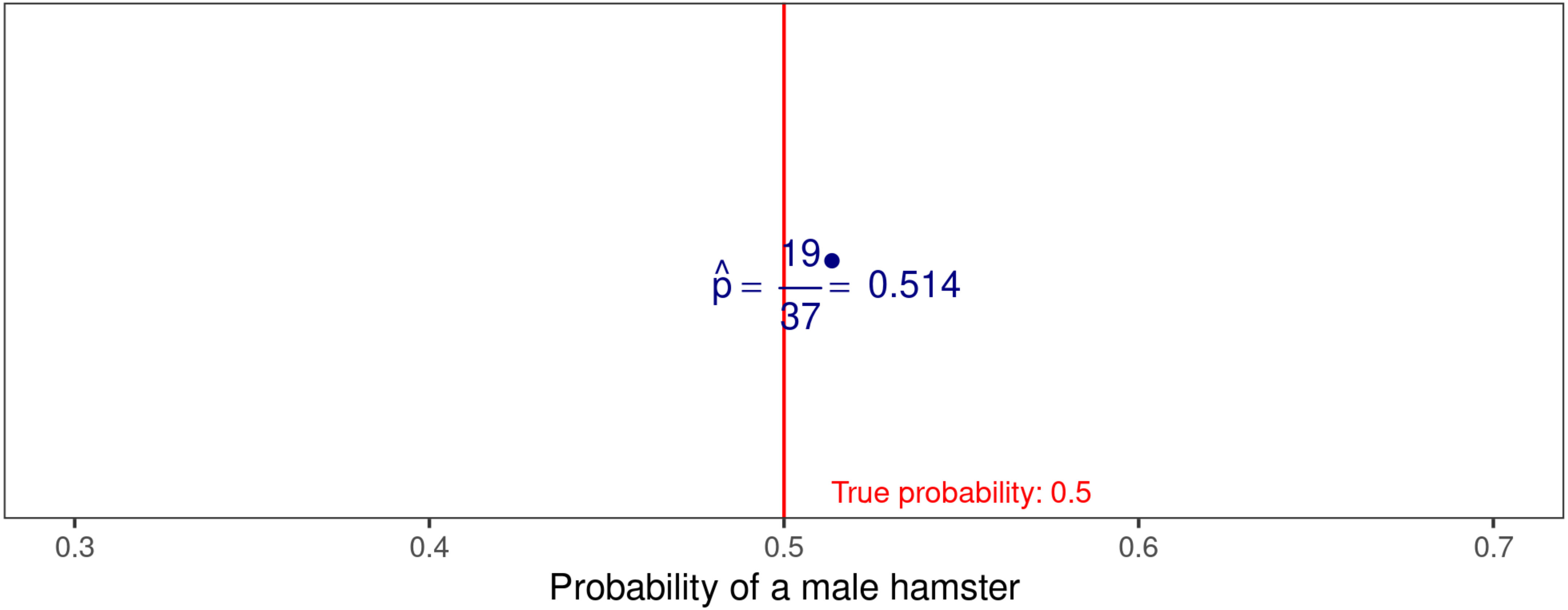
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 15



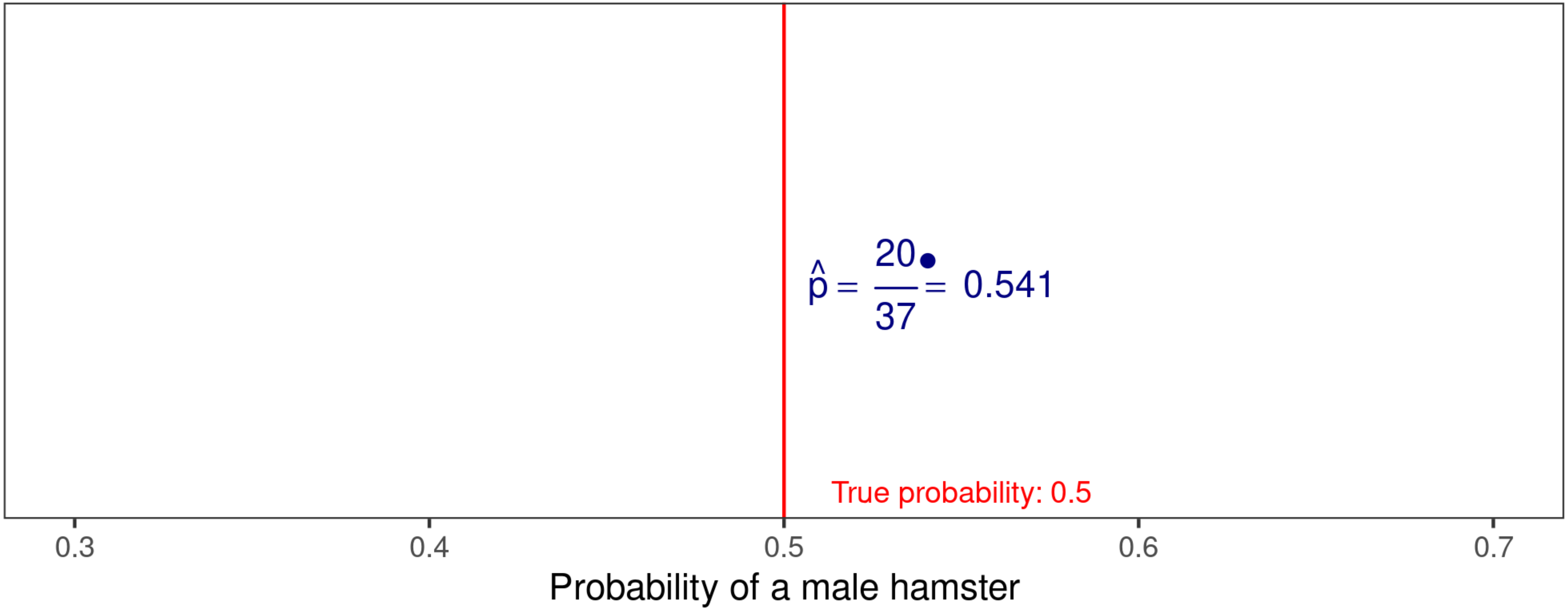
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 16



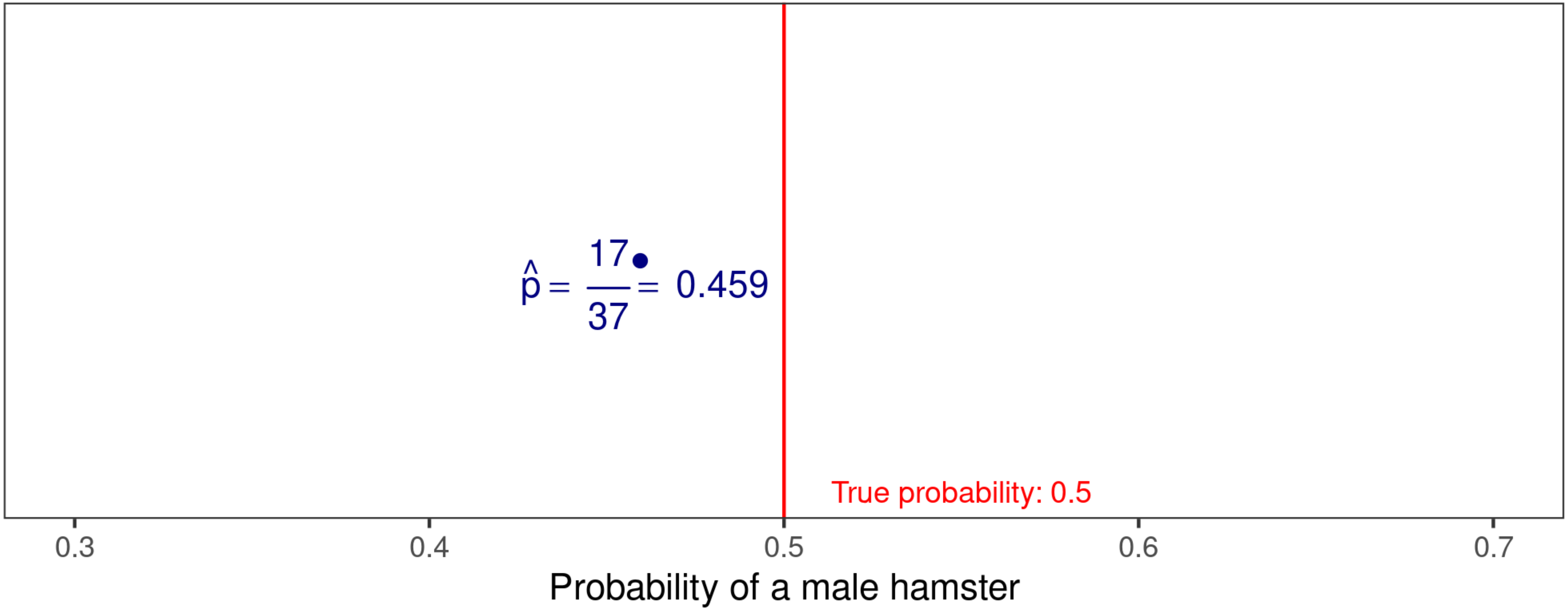
A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 17



A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 18

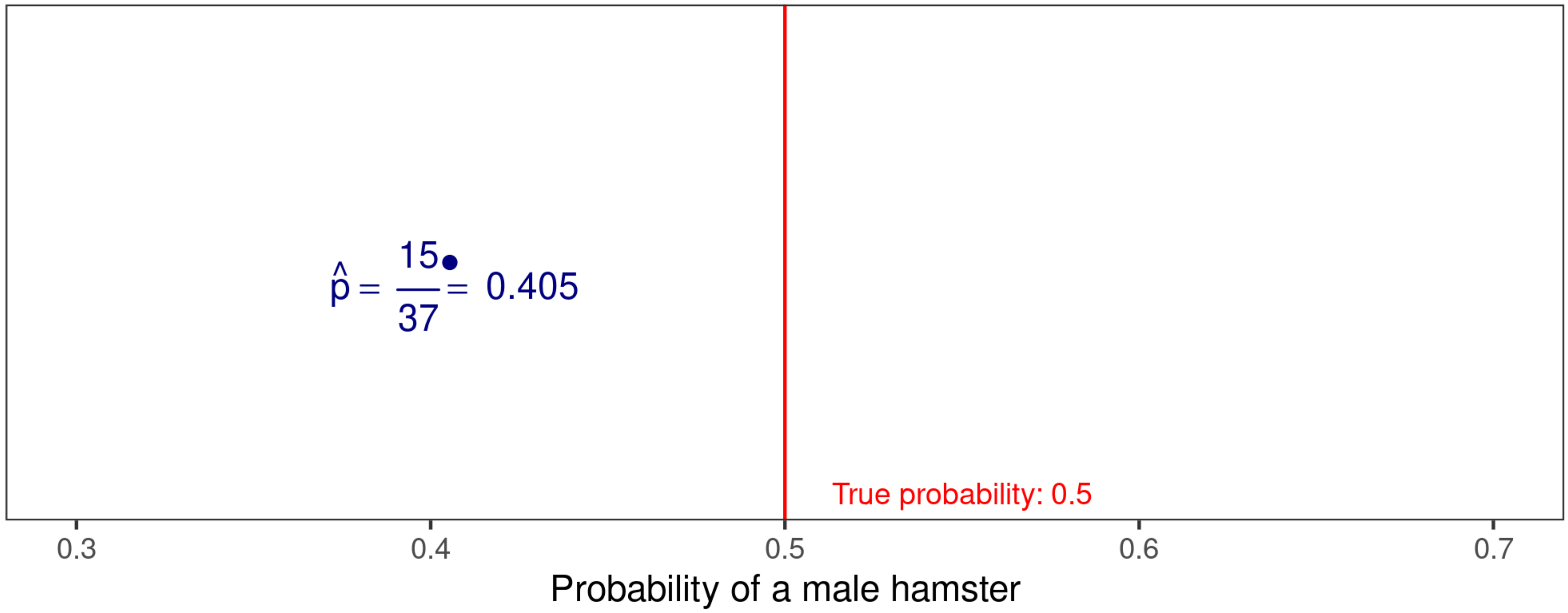


A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 19

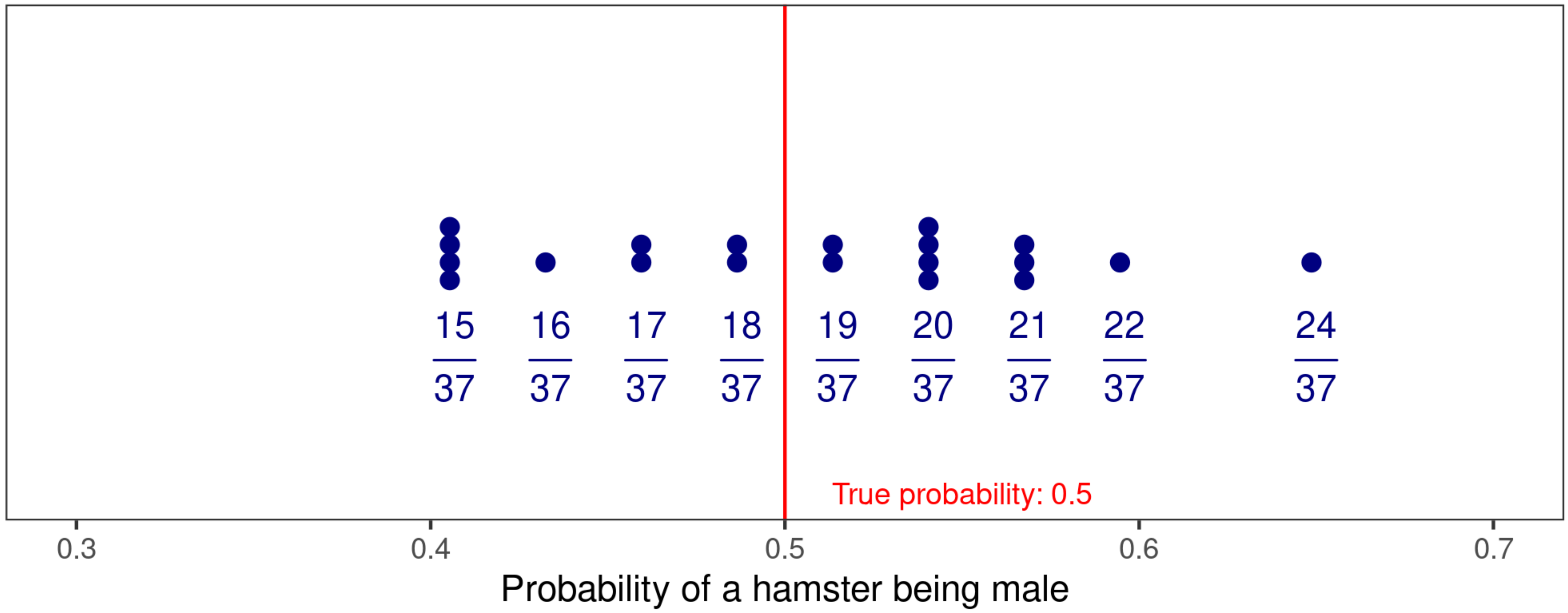


imbi

A random experiment – 37 hamsters, each with probability 0.5 to be male – experiment 20



A random experiment – 37 hamsters, each with probability 0.5 to be male – all experiments



The reality

In reality, we have a single estimate $\hat{p}$ for the probability for a hamster to be male.

Question: How can we derive from this information statements about the true (population) probability p for a hamster to be male?

→ **concept of confidence interval**

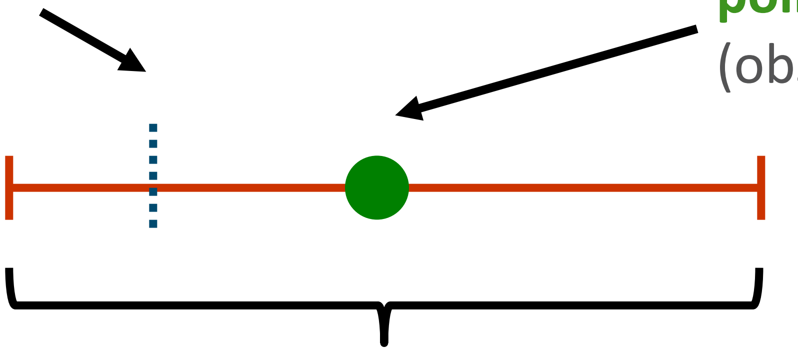
Interval estimation / confidence interval

Statement: For each new **sample**, we expect to observe a slightly different **rate estimate**.

Wanted: An interval, that covers the **true rate** in the **population** with a defined probability.

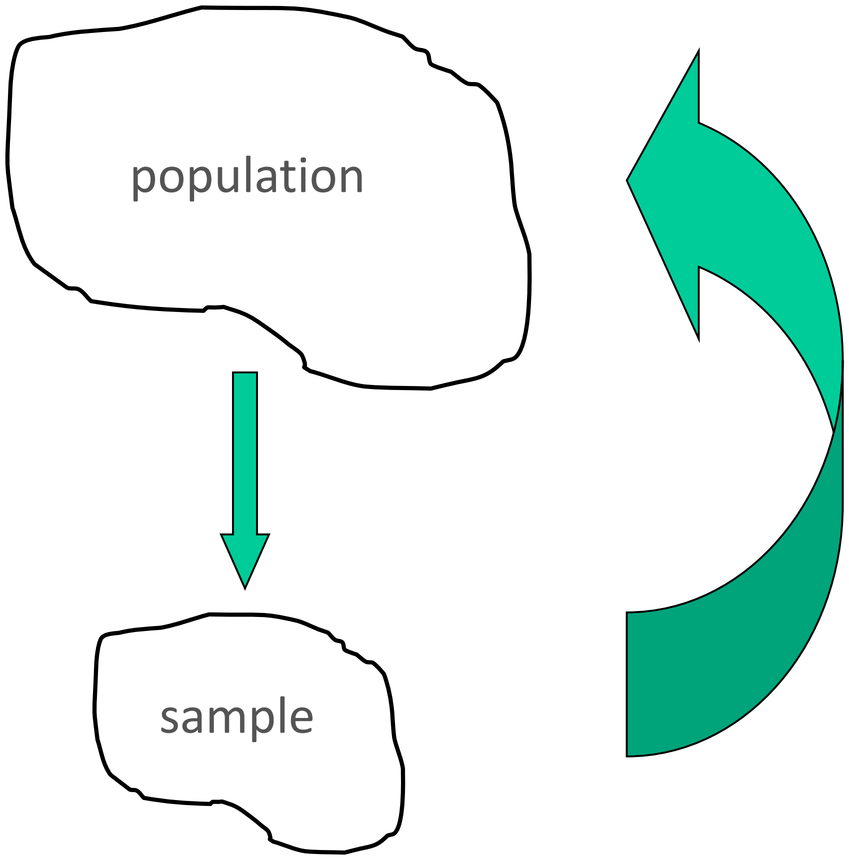
true rate
(in population, unknown!)

point estimate
(observed rate in sample)



interval estimation:
confidence interval

The same principle



inference with *interval estimation*
(= confidence interval)

point estimation

Confidence interval for rates with coverage probability 0.95

(approx.) 0.95-confidence interval $= \hat{p} \pm 1.96 \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$

for our example: $= 0.62 \pm 1.96 \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{37}}$

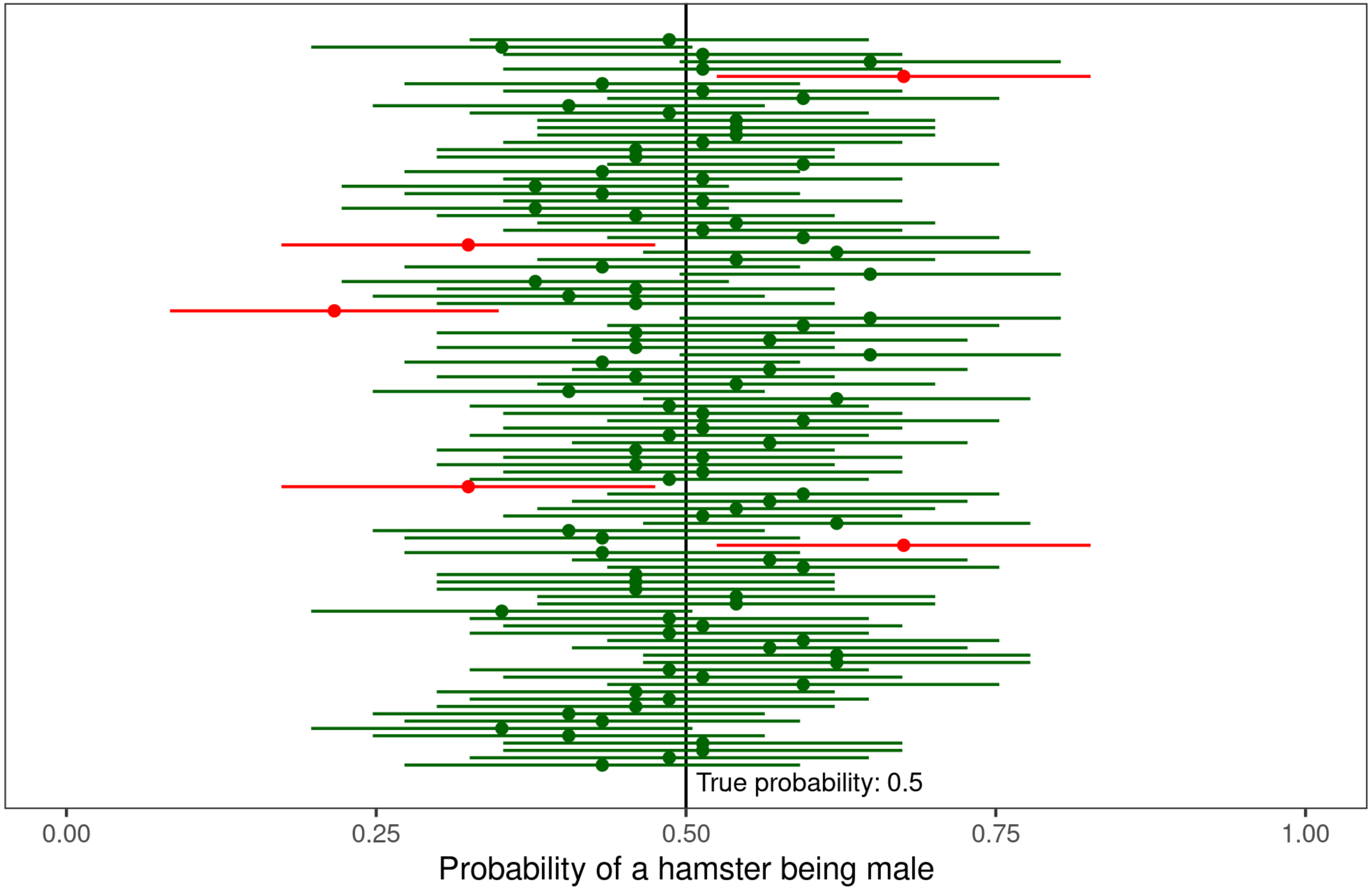
$= [0.46; 0.78]$

The width of the confidence interval is $2 \cdot 1.96 \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$

The width depends on …

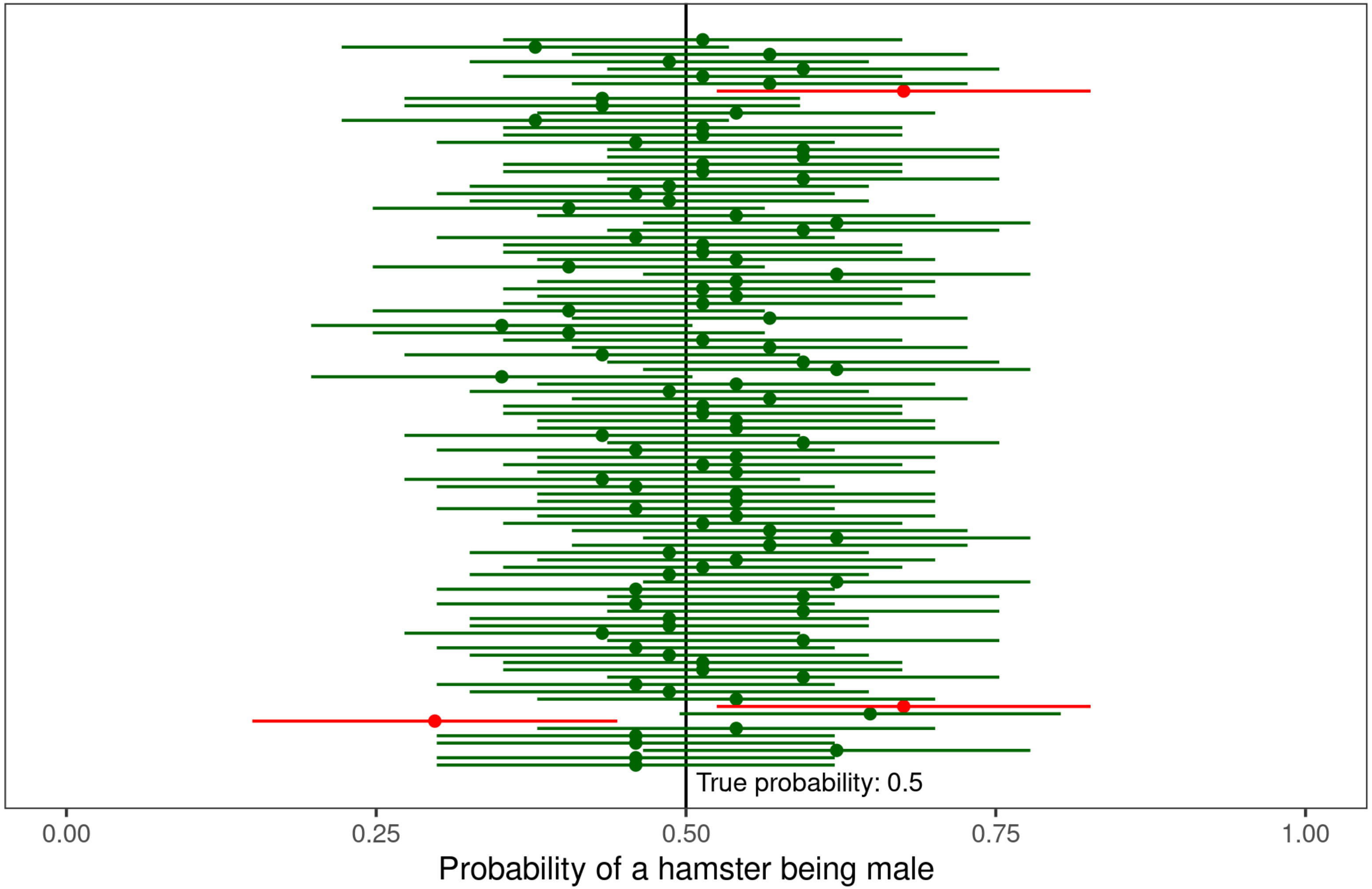
- the sample size (4-fold sample size = half width)
- the coverage probability (the larger, the wider)
- the rate estimate obtained from the sample (largest width for rate 0.5)

A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 1)

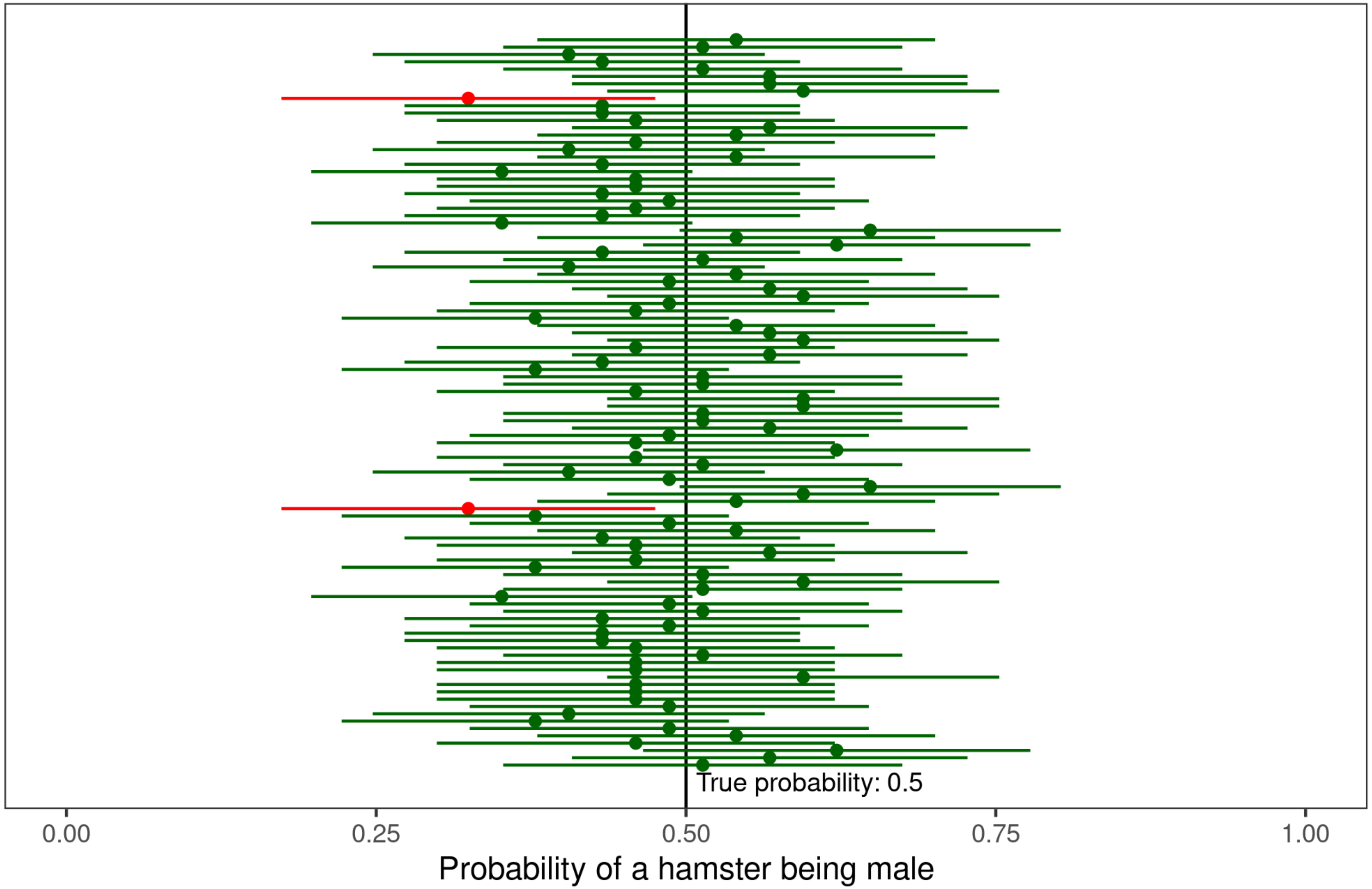


imbi

A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 2)

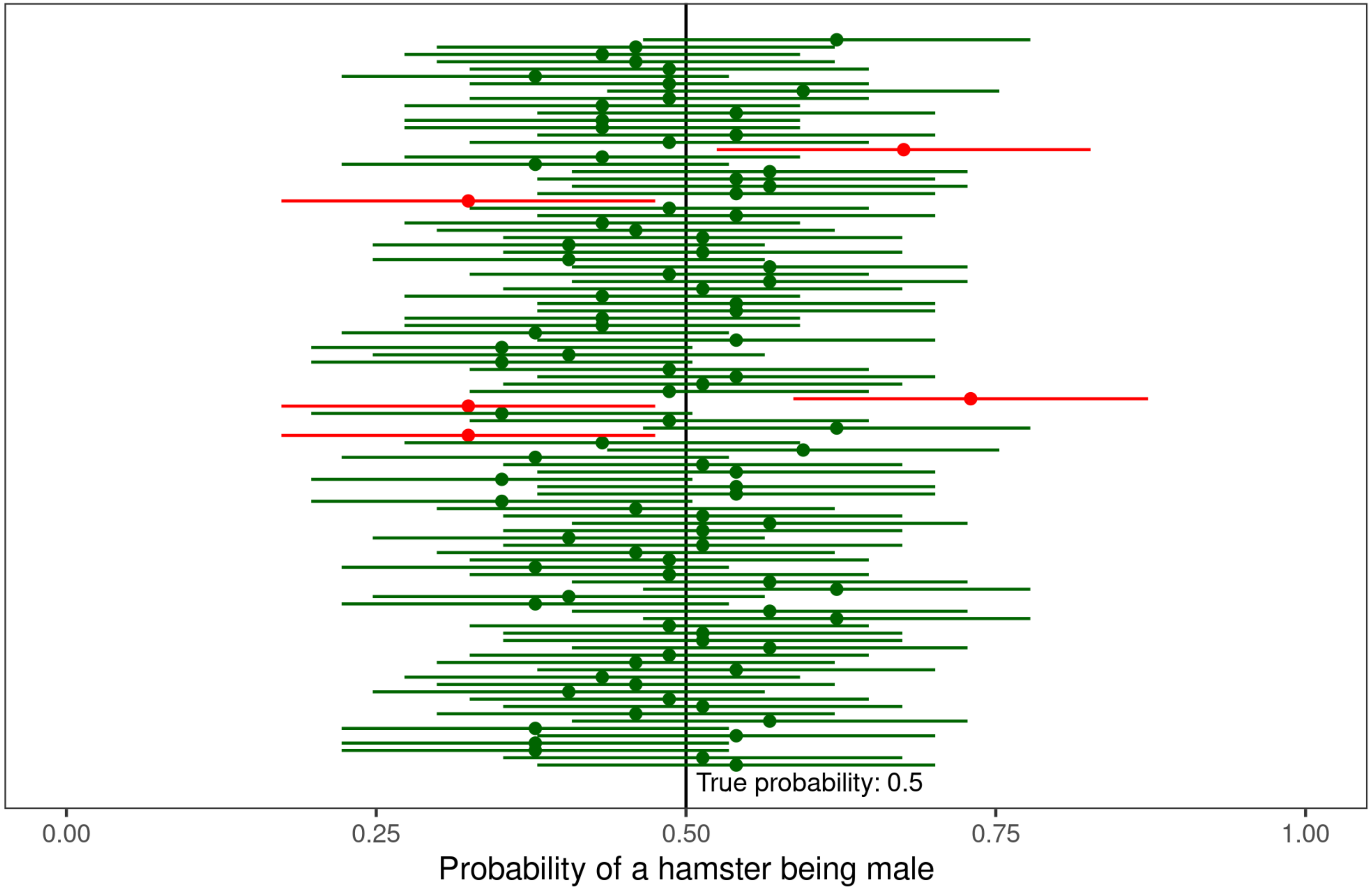


A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 3)

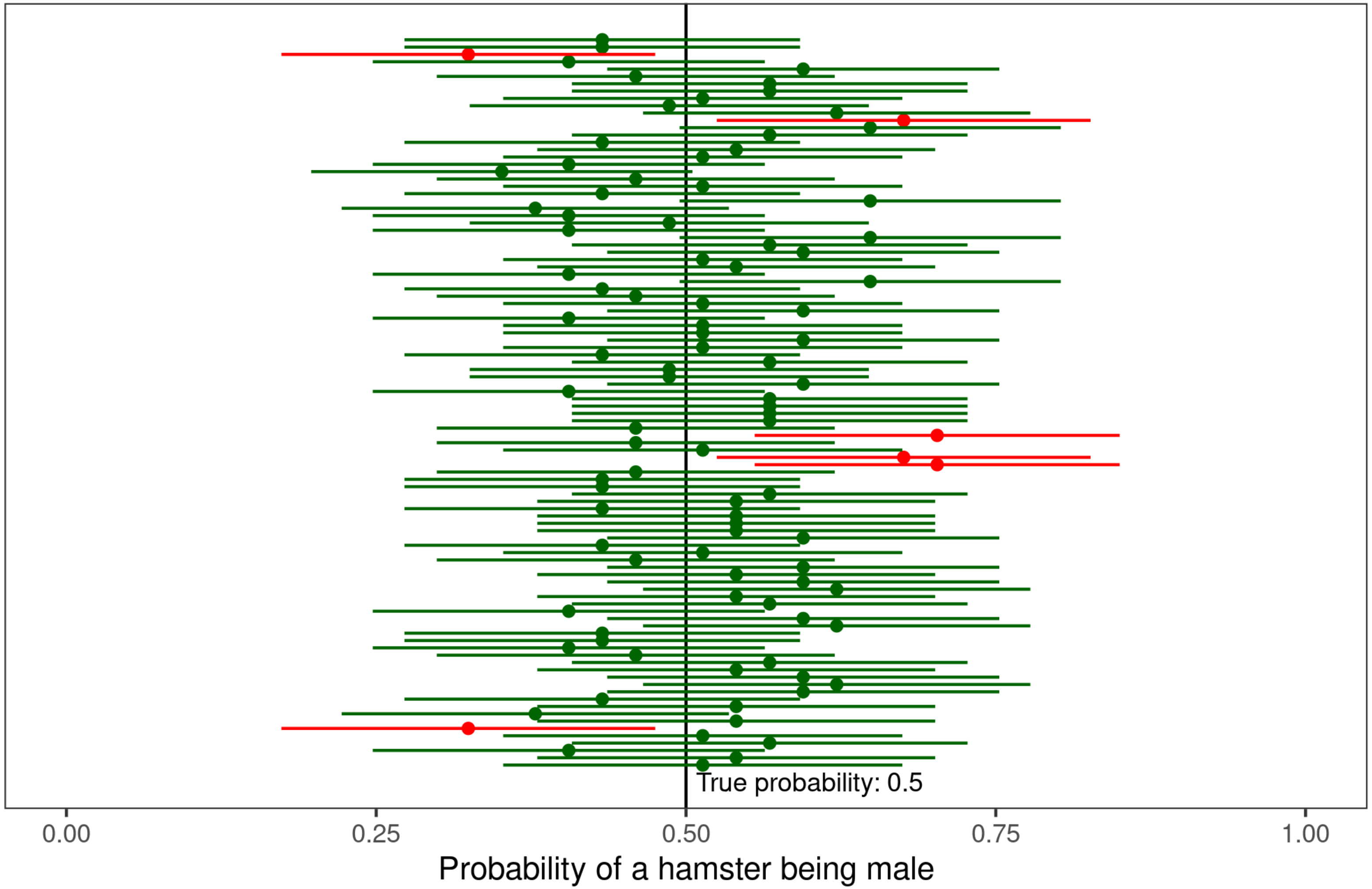


imbi

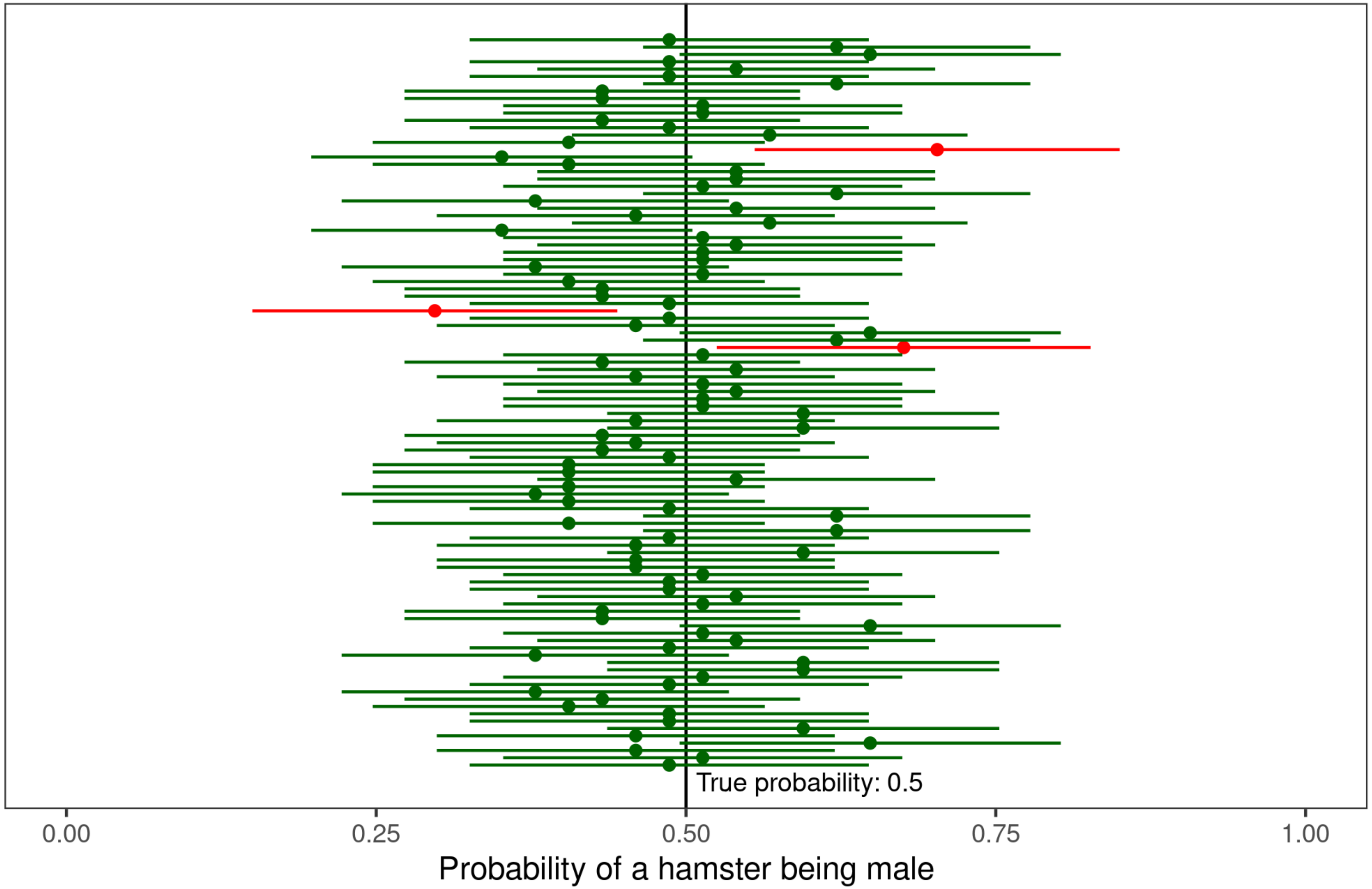
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 4)



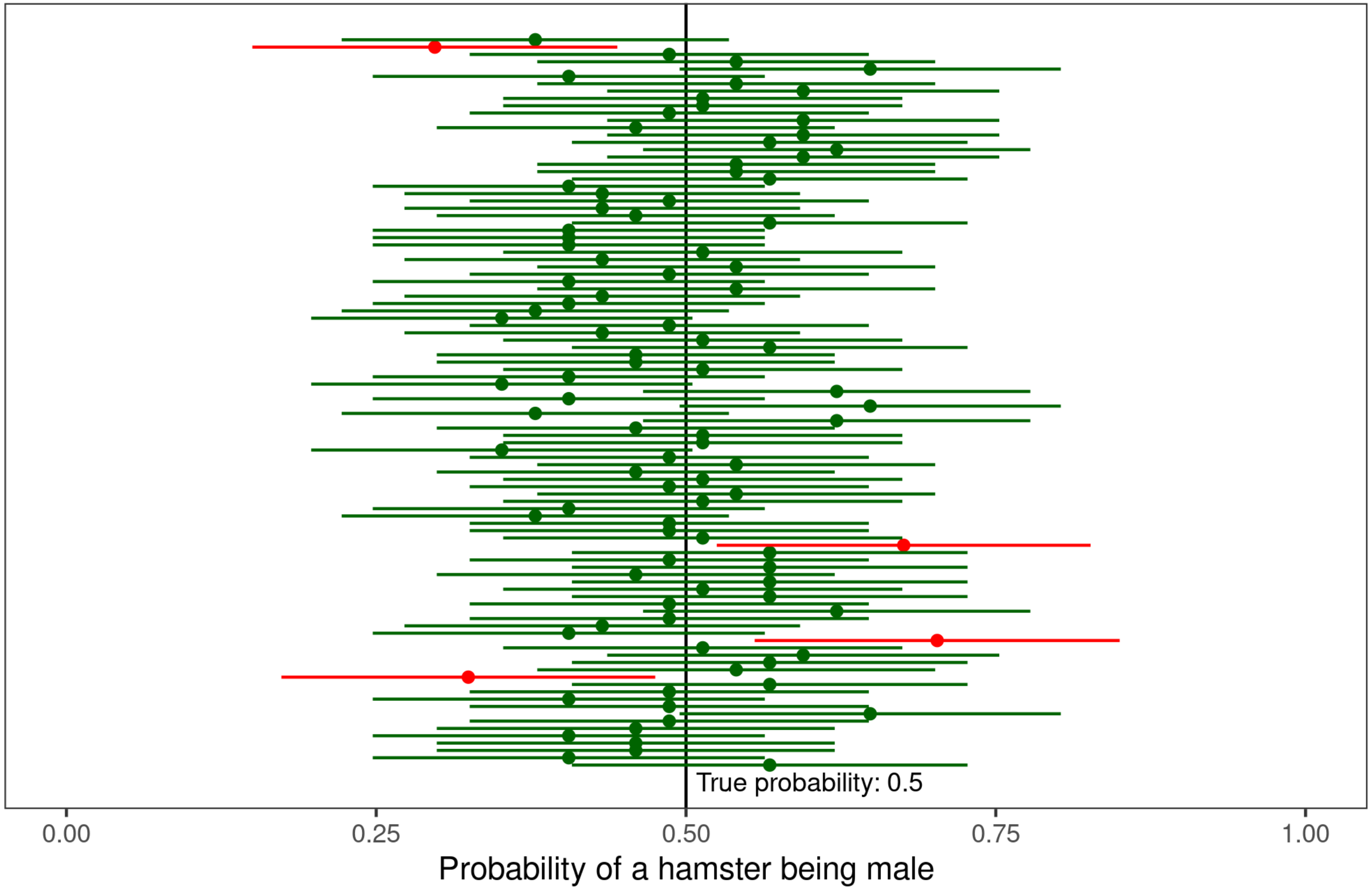
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 5)



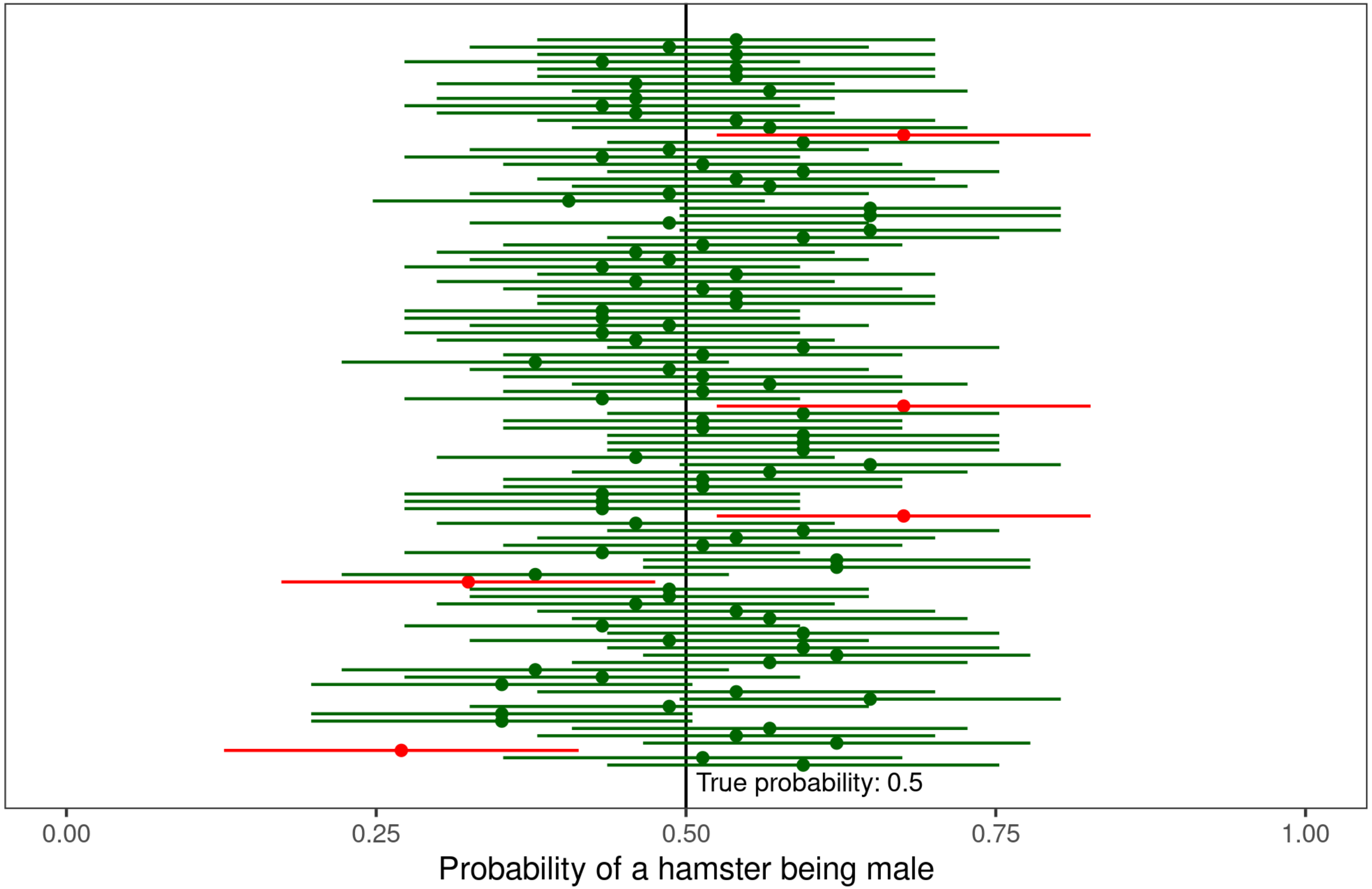
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 6)



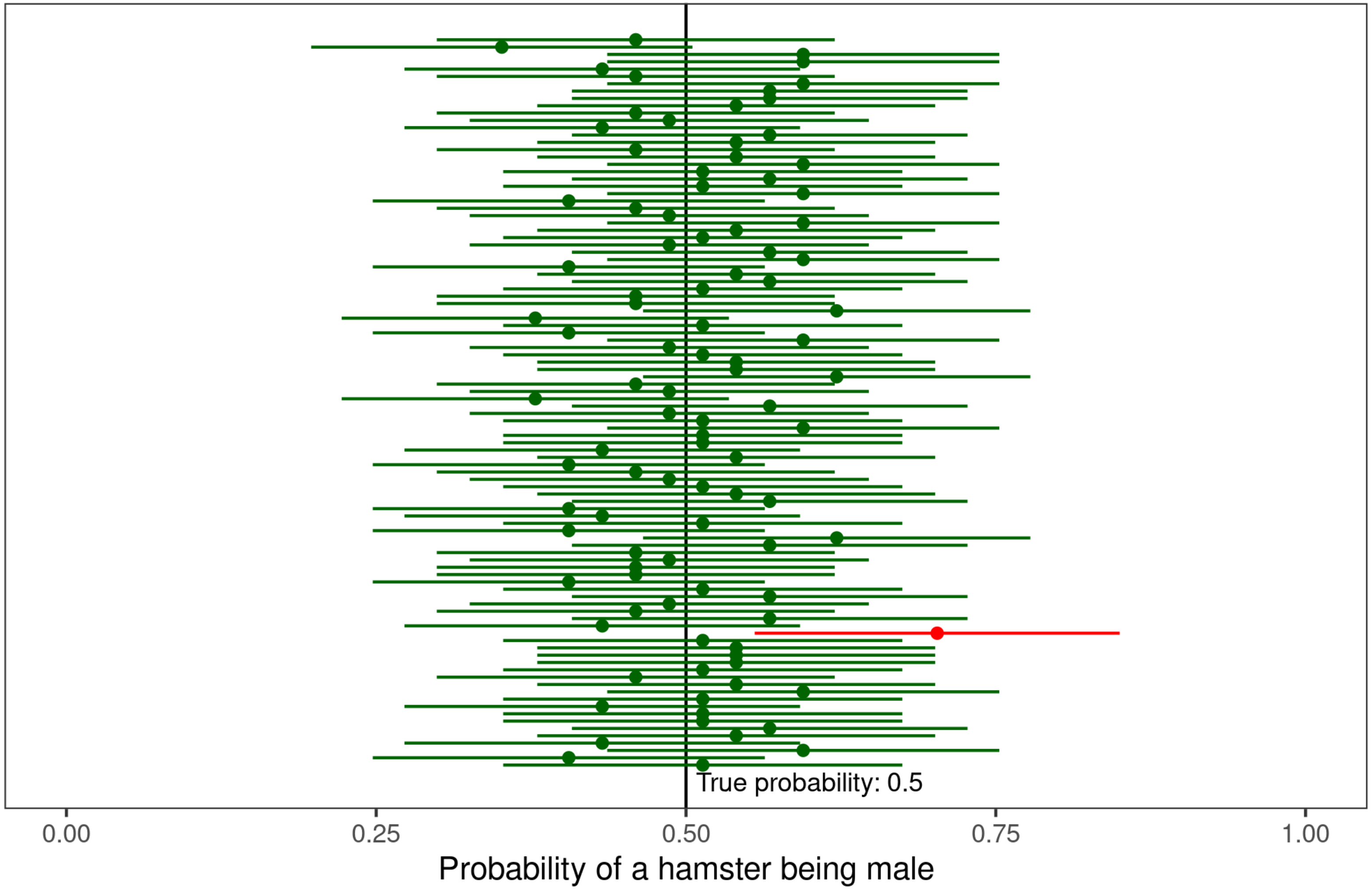
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 7)



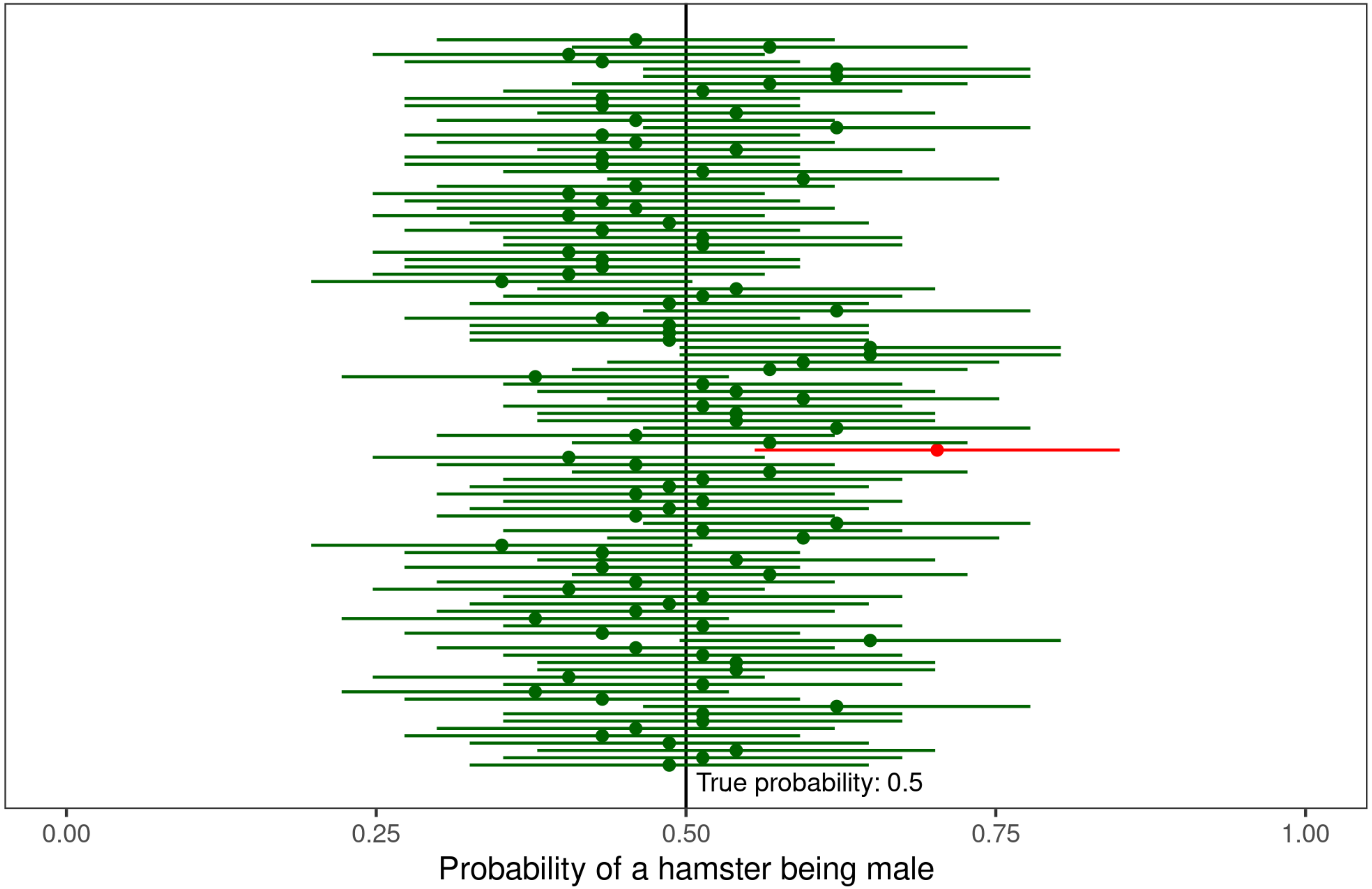
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 8)



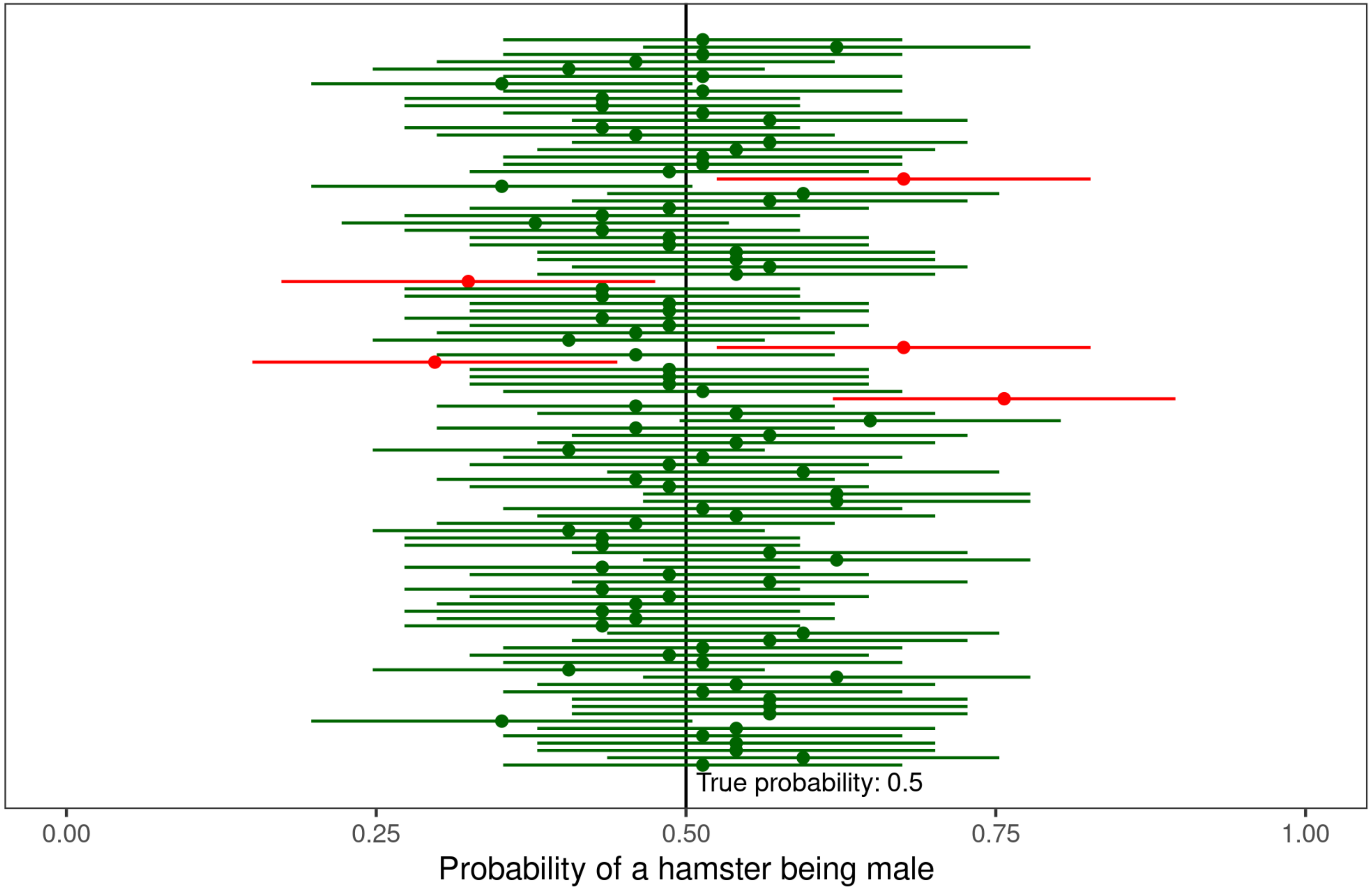
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 9)



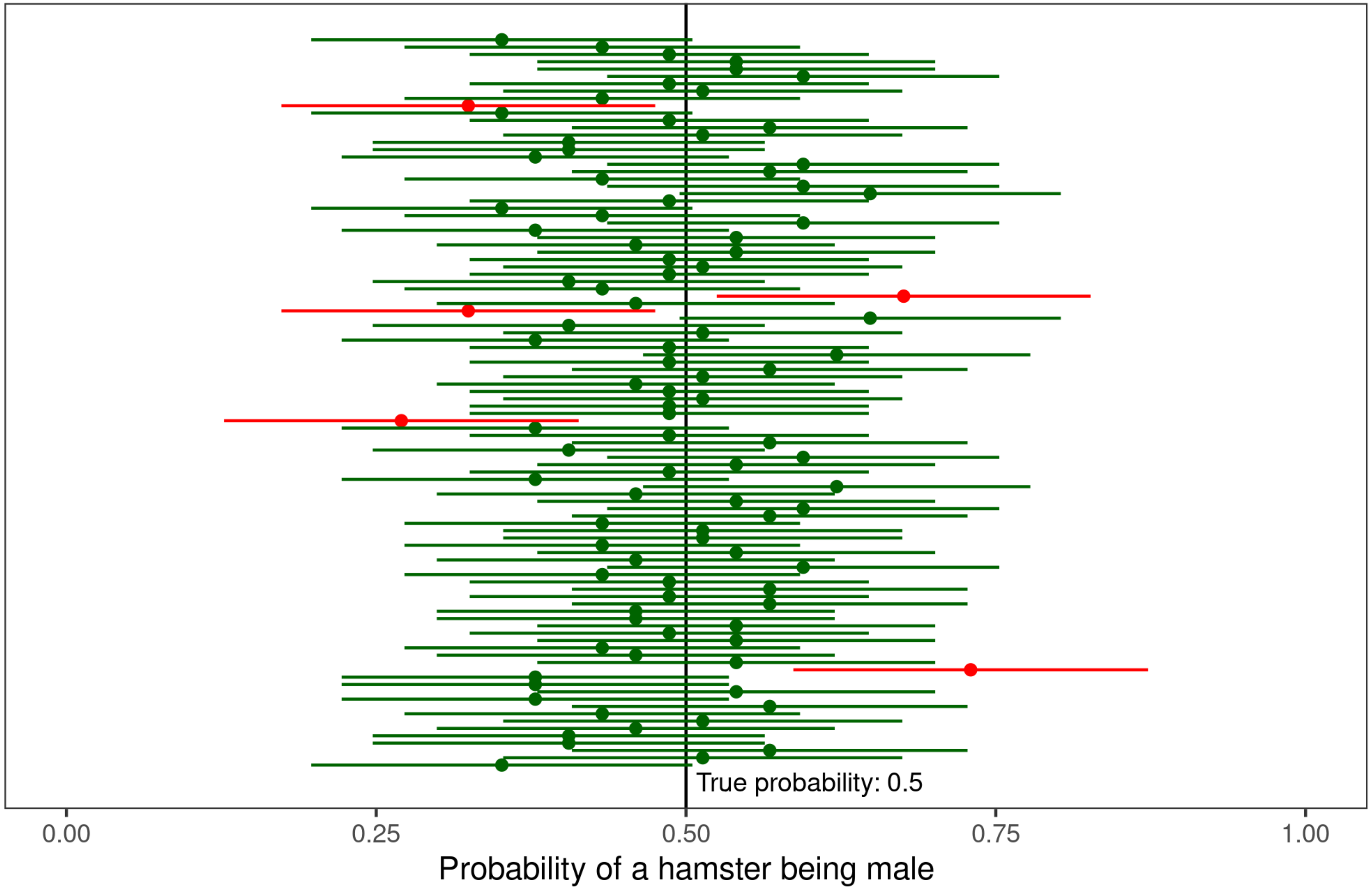
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 10)



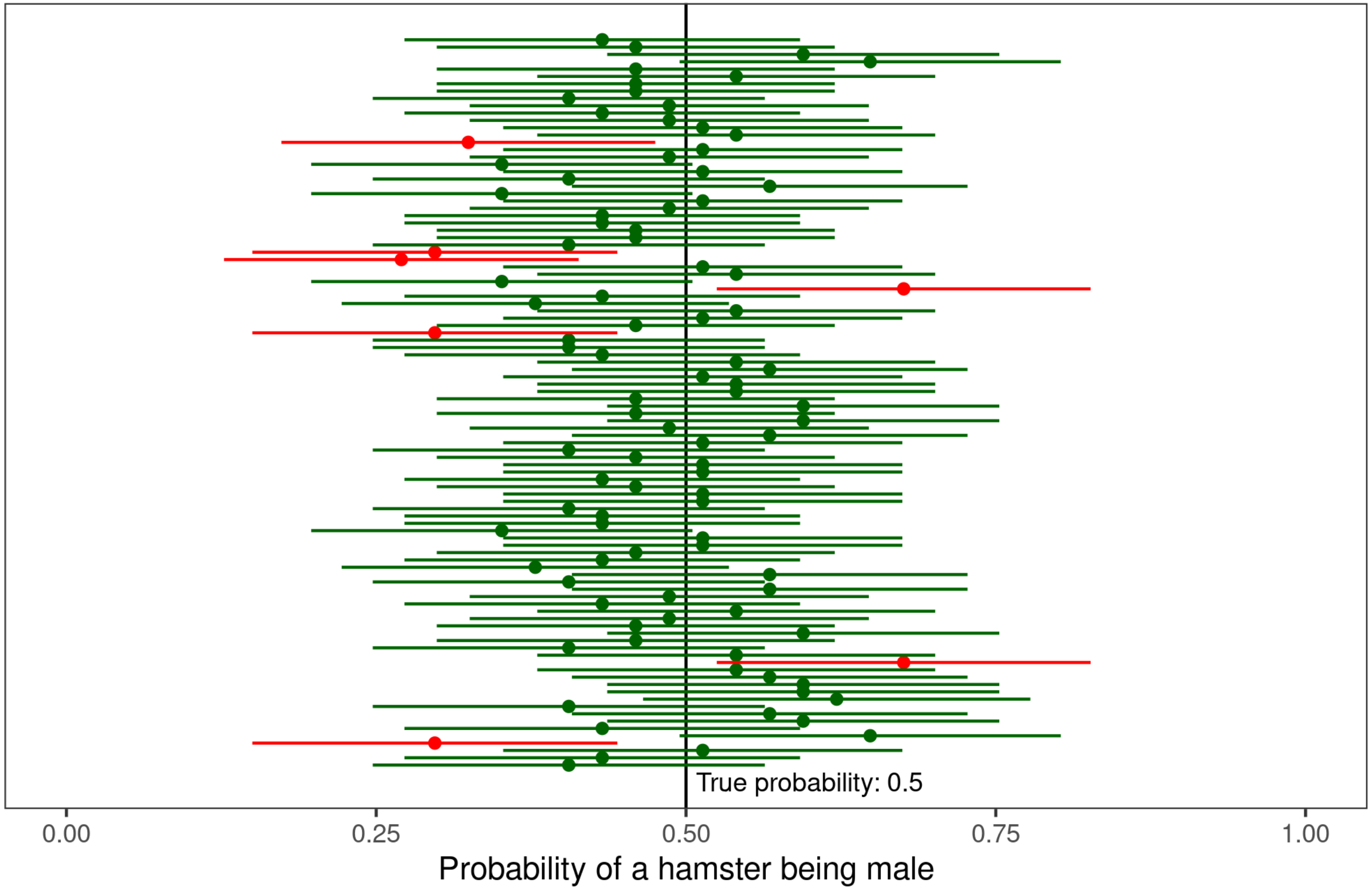
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 11)



A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 12)

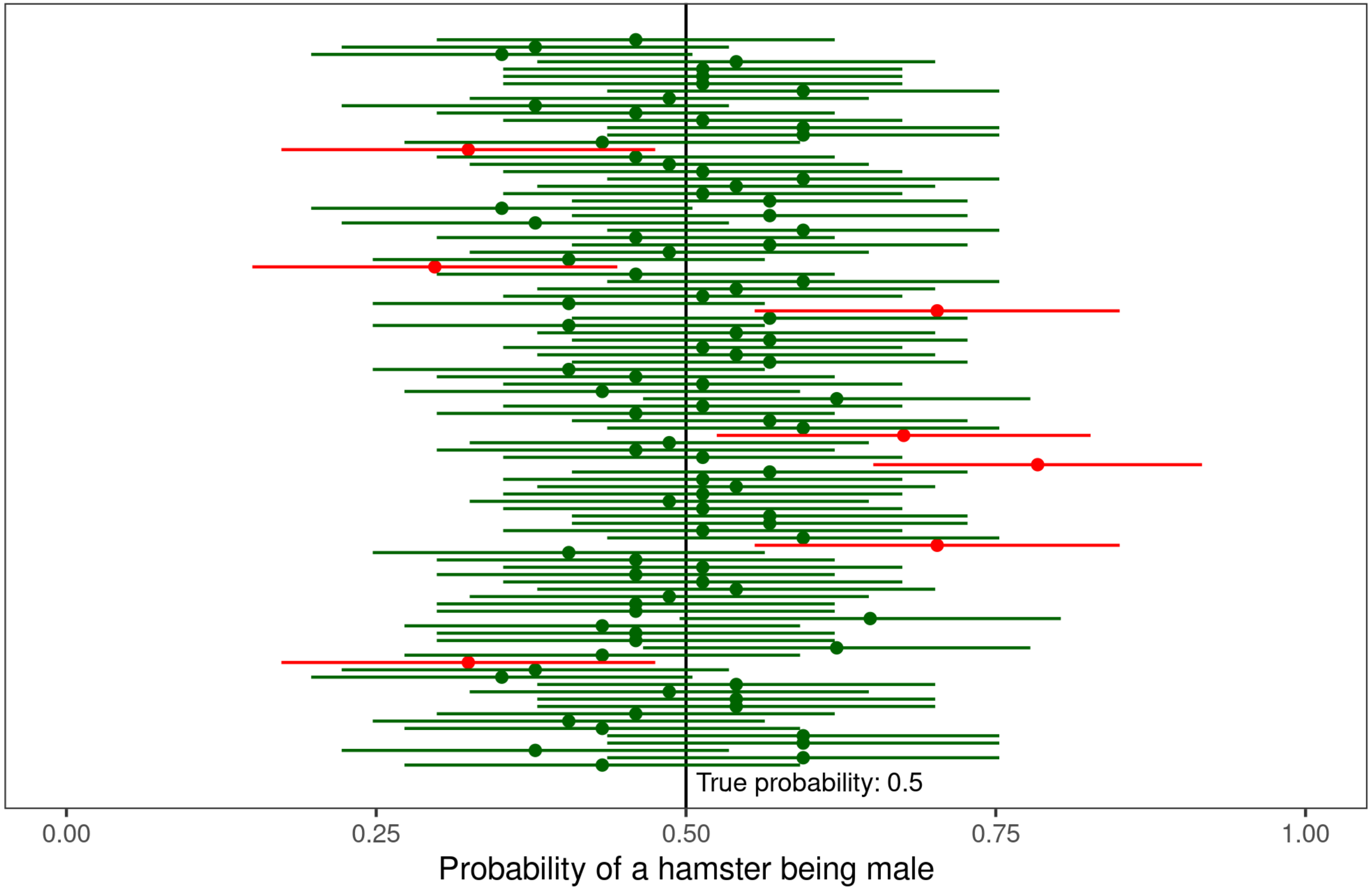


A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 13)

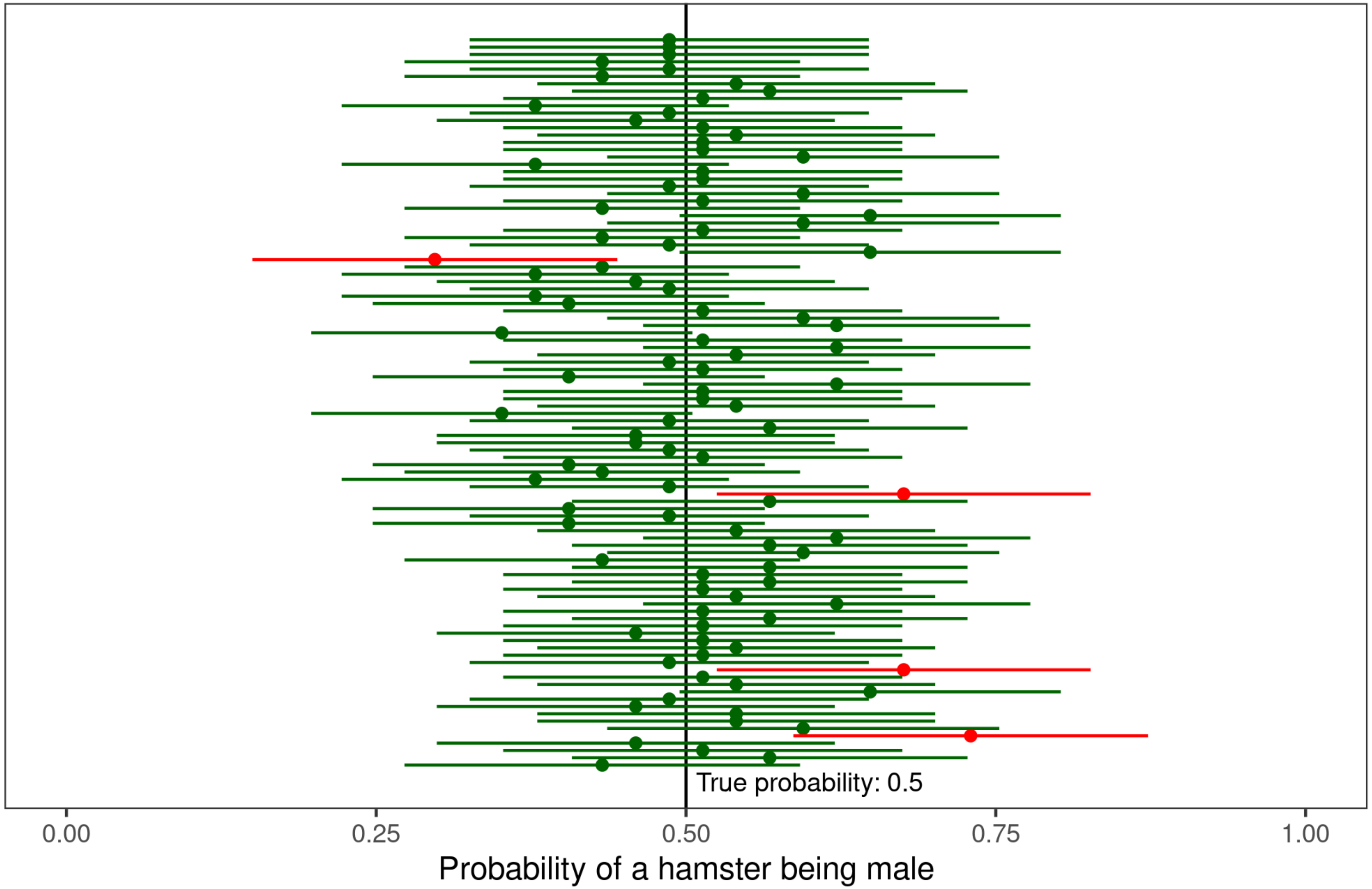


imbi

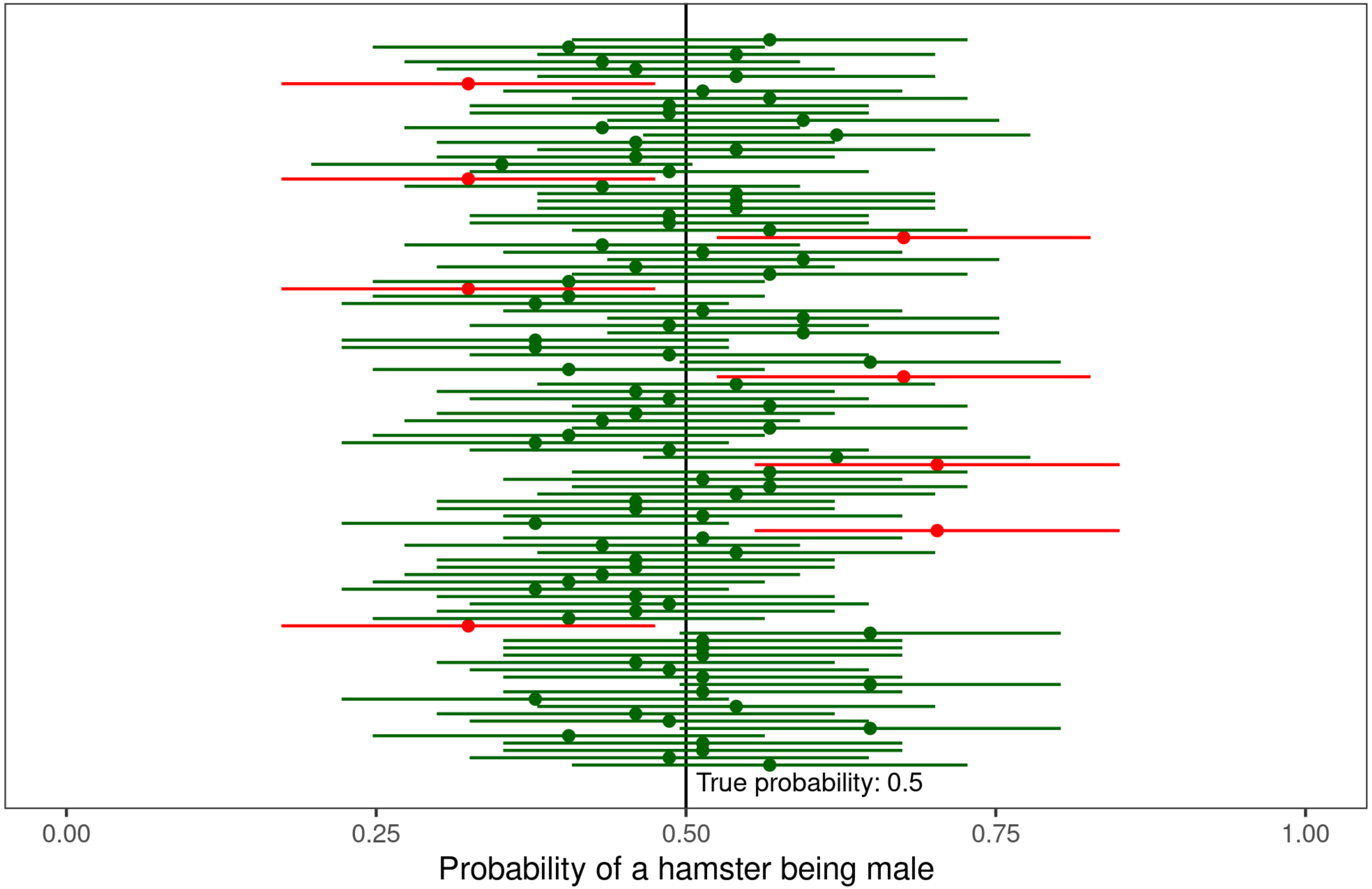
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 14)



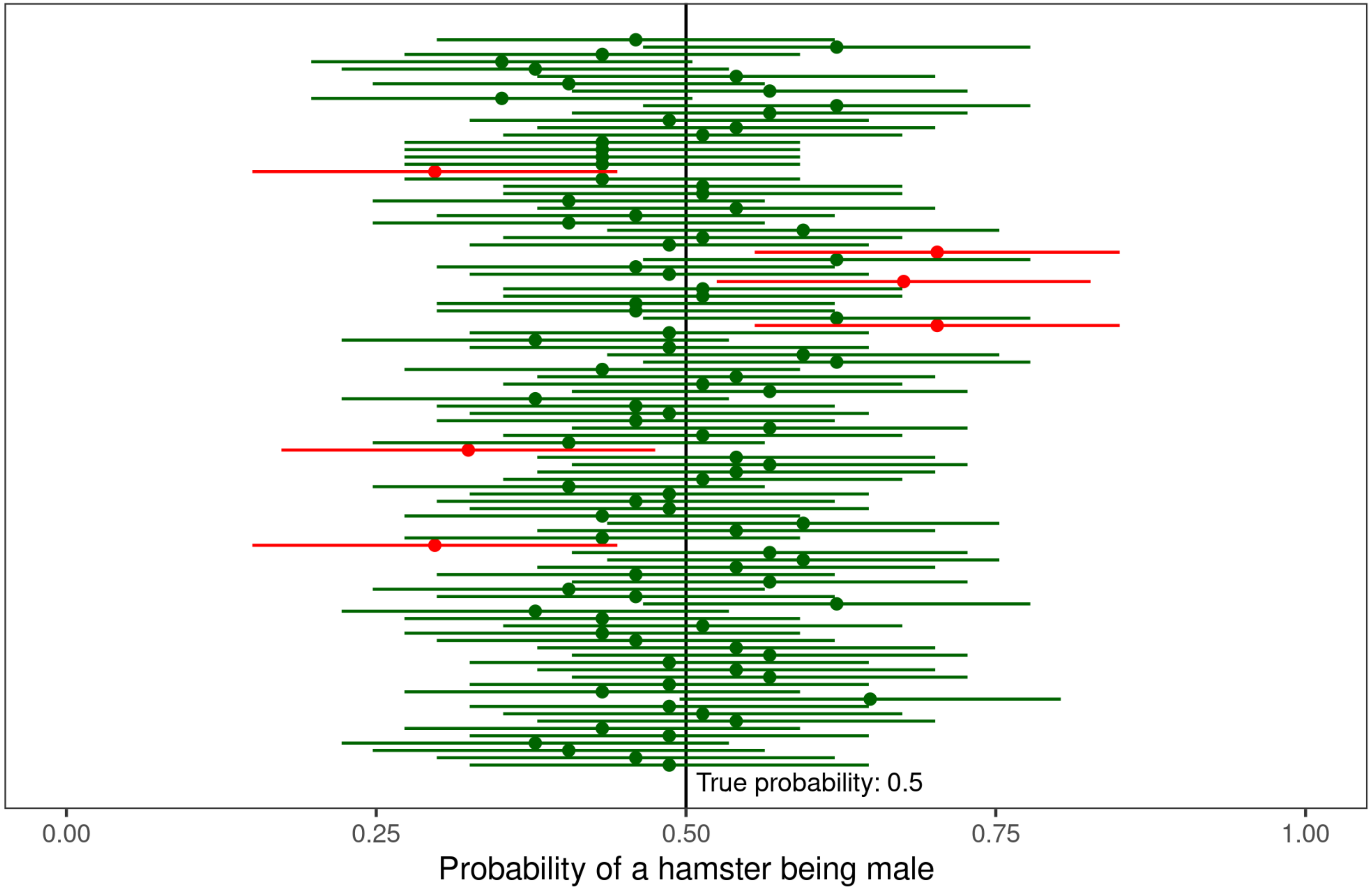
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 15)



A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 16)

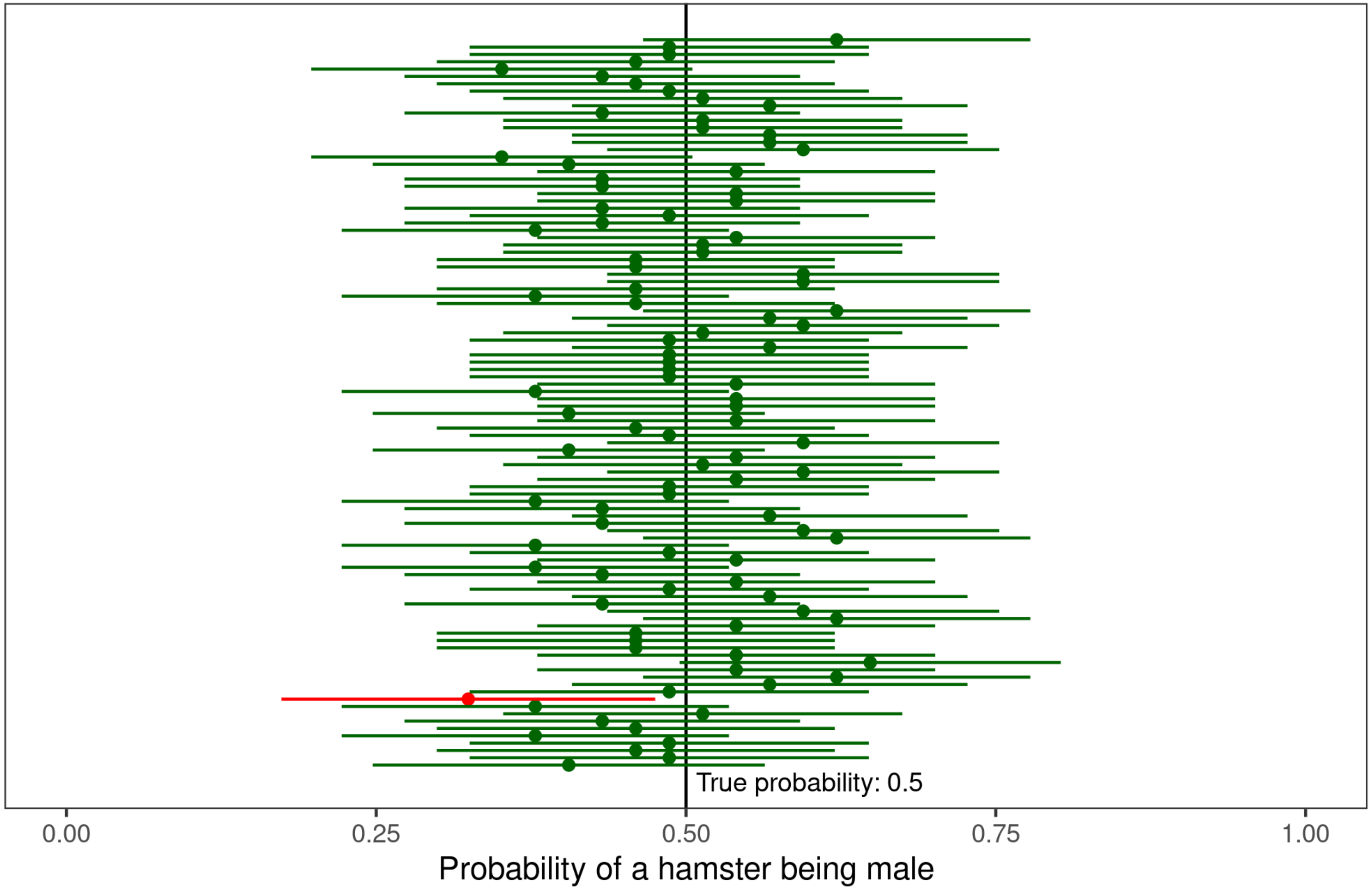


A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 17)

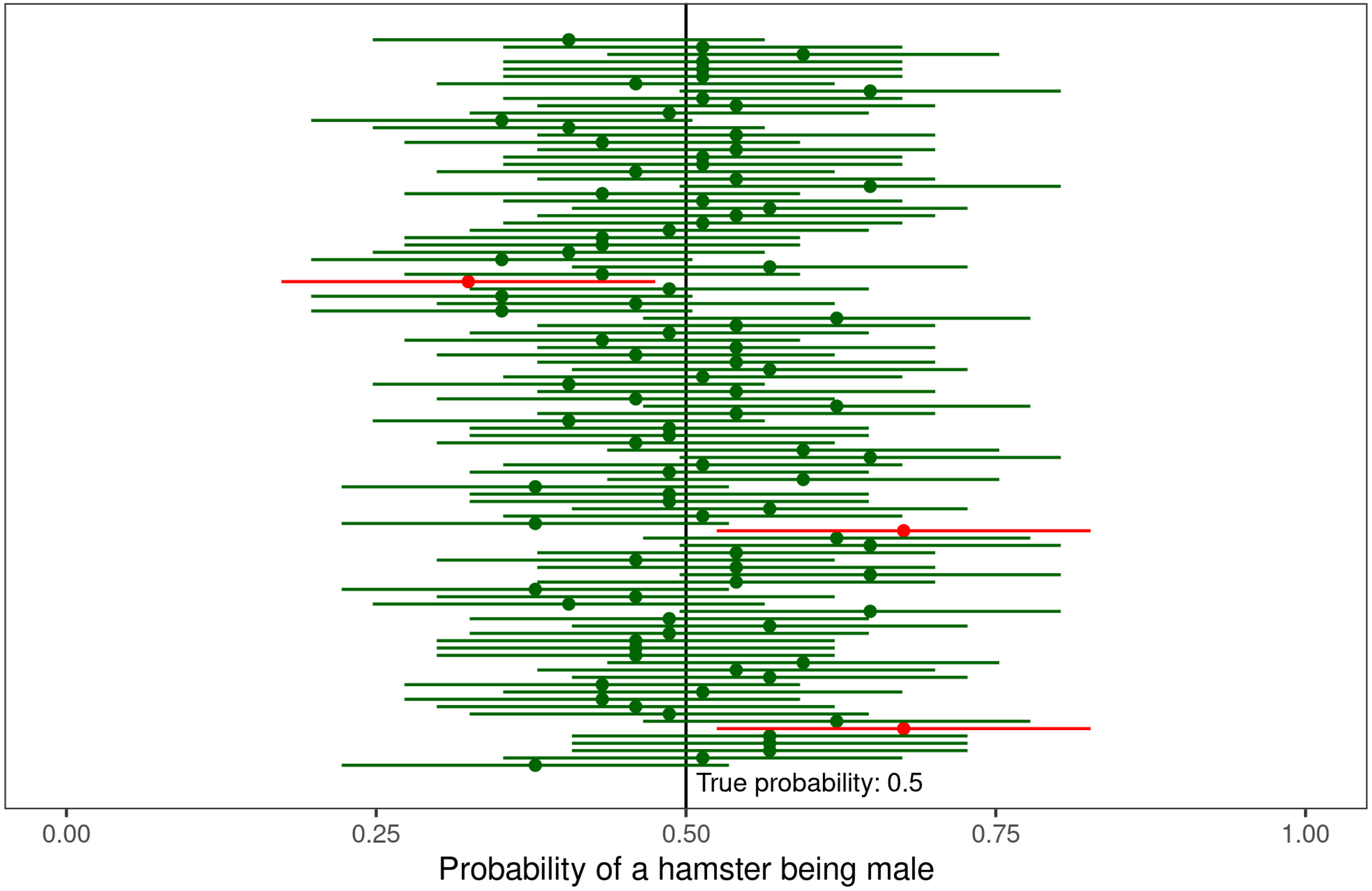


imbi

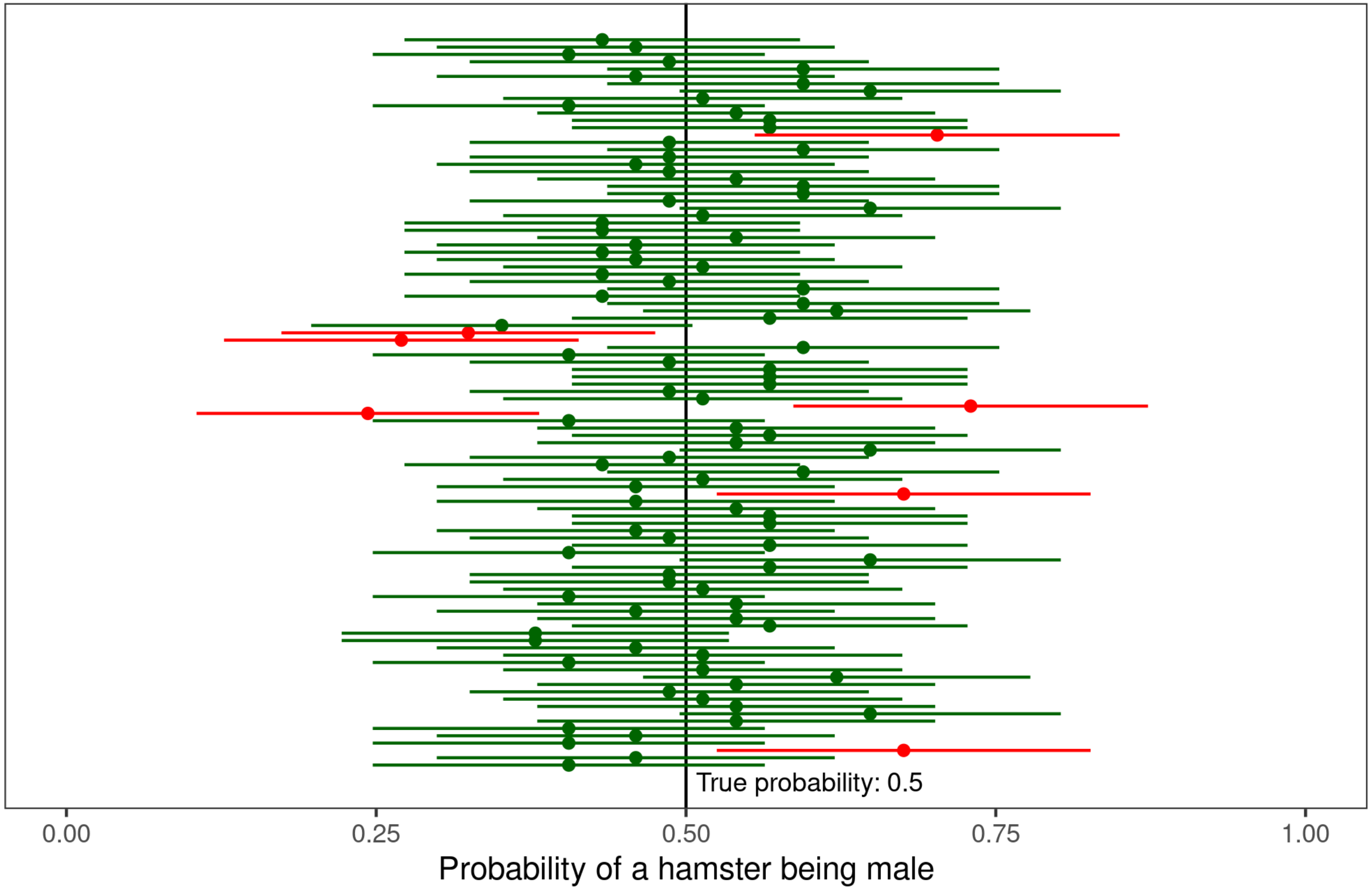
A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 18)



A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 19)

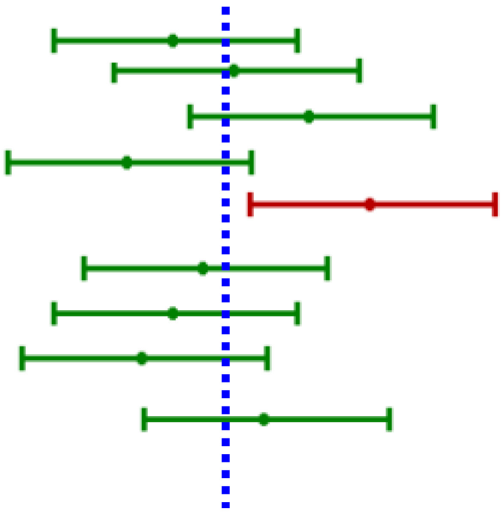


A random experiment – one hundred 0.95-confidence intervals for the data of 37 hamsters, each with probability 0.5 to be male (experiment 20)

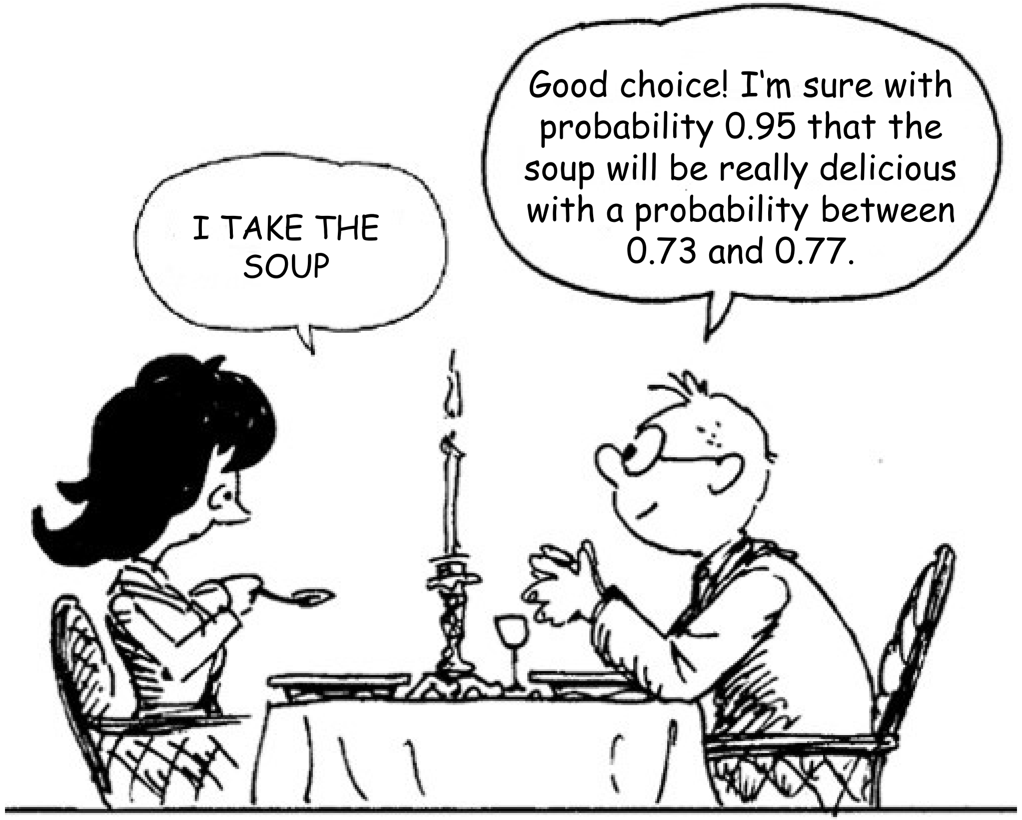


Interval estimation / confidence interval

The **random** **confidence interval** **covers the** **true rate** in the population **with a pre-specified probability** $1 - \alpha$.

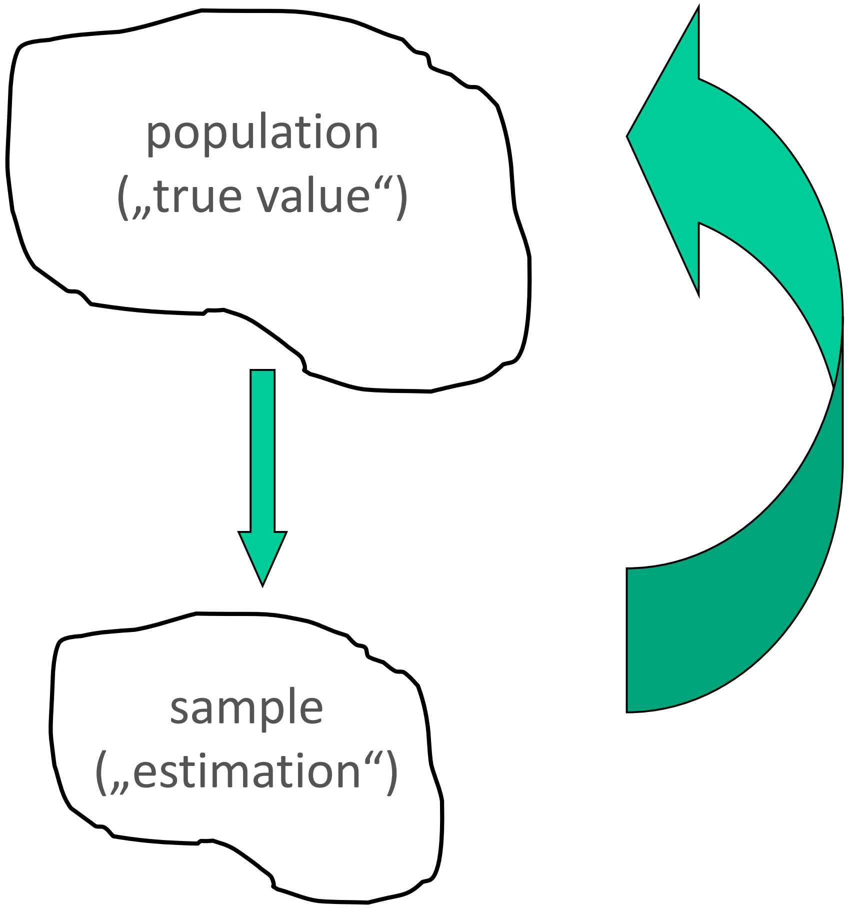


true value (unknown!)



Larry Gonick and Woollcott Smith „The Cartoonguide to Statistics“

Again: point estimation and confidence interval



confidence interval
(covers the true value
with probability $1-\alpha$)

Take-Home-Messages III

(1) In addition to the result of a statistical test, the size of the effect has to be reported.

(2) For this, the result observed in the sample should be given („point estimate“) as well as the interval that covers the true population value with a pre-specified probability (commonly 0.95) („confidence interval“).

Another conversation at another walking tour

- „When I talked with my sister recently, she said that among male hamsters there are more with black fur color than among female hamsters.“
- „Oh no, not that old chestnut again! Your sister is really annoying!“
- „This is in fact true! But the question is nevertheless interesting. And we learned in a recent course that making an experiment and performing a statistical test can give an evidence-based answer.“
- „We became familiar with the binomial test. But it considers just a single rate. Here, we would like to compare two rates: the ones for males and the one for females for having black fur color.“
- „So go on to visit the course. I will meanwhile collect further data on the fur color of male and female hamsters.“

Chi-square test and related confidence interval

The second hamster experiment

- **data of hamster experiment II:**

	fur black	fur not black	total
male hamster	20	80	100
female hamster	10	90	100
total	30	170	200

- **question:** Is the difference observed in the sample „significant“? I.e., are the true rates in the population different or not?

- **null- and alternative hypothesis:** $H_0 : p_x = p_y$ vs. $H_1 : p_x \neq p_y$

Chi-square test

- **What do we expect if $H_0: p_x = p_y$ is true?**

	black (n=30)	not black (n=170)	total (n=200)
male	20 $100 \cdot \frac{30}{200} = 15$	80 $100 \cdot \frac{170}{200} = 85$	100 100
female	10 $100 \cdot \frac{30}{200} = 15$	90 $100 \cdot \frac{170}{200} = 85$	100 100
total	30 30	170 170	200 200

Chi-square test (2)

- **general fourfold-table:**

Group	characteristics +	characteristics -	total
X	a	b	a+b
Y	c	d	c+d
total	m	n	m+n

Chi-square test (3)

- **What do we expect if $H_0: p_x = p_y$ is true?**

Group	characteristics +	characteristics -	total
X	a $n \cdot \frac{a+c}{m+n}$	b $n \cdot \frac{b+d}{m+n}$	n
Y	c $m \cdot \frac{a+c}{m+n}$	d $m \cdot \frac{b+d}{m+n}$	m
total	a+c	b+d	m+n

Chi-square test (4)

- **test statistic:**

$$X^2 = \sum (O - E)^2/E, \text{ where}$$

O: observed cell counts

E: expected cell counts, if H_0 is true

„The more E deviates from O, the more doubtful is that H_0 is true.“

- $\boldsymbol{(O - E)^2/E}$**:**

$(ad - bc)^2/(m+n)(a+b)m$	$(ad - bc)^2/(m+n)(c+d)m$
$(ad - bc)^2/(m+n)(a+b)n$	$(ad - bc)^2/(m+n)(c+d)n$

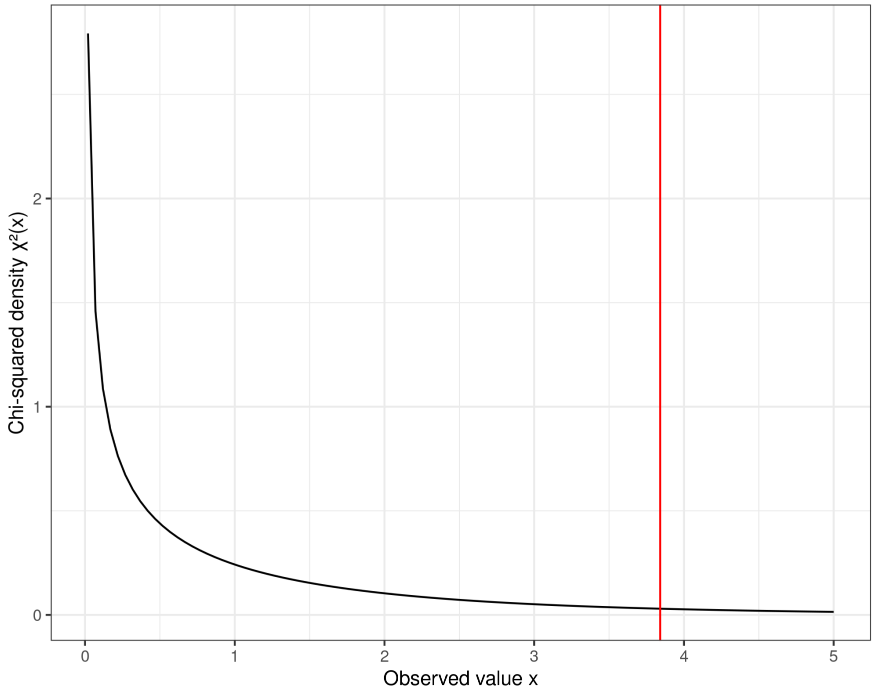
imbi

Chi-square test (5)

- **test statistic:**

$$X^2 = \frac{(n+m)(ad-bc)^2}{(a+b)(a+c)(c+d)(b+d)}$$

- If H_0 is true, X^2 is asymptotically chi-square-distributed with 1 degree of freedom:



Chi-square test (6)

- **test statistic:**

$$X^2 = \frac{(n+m)(ad-bc)^2}{(a+b)(a+c)(c+d)(b+d)}$$

- If H_0 is true, X^2 is asymptotically chi-square-distributed with 1 degree of freedom
- "asymptotically" means: "for large numbers"
- 0.95-quantile of chi-square-distributed with 1 degree of freedom:

$$\chi^2_{1,0.95} = 3.84$$

- (asymptotic) test of H_0 versus H_1 with type I error rate $\alpha = 0.05$:
  - "reject H_0 if $X^2 \geq 3.84$."
  - "do not reject H_0 if $X^2 < 3.84$."

Back to the second hamster experiment

- **data of hamster experiment II:**

	fur black	fur not black	total
male hamster	20	80	100
female hamster	10	90	100
total	30	170	200

- **question:** Is the difference observed in the sample „significant“? I.e., are the true rates in the population different or not?

Exercise: the hamster experiment II

What is the test decision by use of the chi-square test with type I error rate $\alpha = 0.05$ for the hamster experiment II?

Remember: test statistic $X^2 = \sum \frac{(O-E)^2}{E}$

Exercise: the hamster experiment II

	black (n=30)	not black (n=170)	total (n=200)
male	20 $100 \cdot \frac{30}{200} = 15$	80 $100 \cdot \frac{170}{200} = 85$	100 100
female	10 $100 \cdot \frac{30}{200} = 15$	90 $100 \cdot \frac{170}{200} = 85$	100 100
total	30 30	170 170	200 200

Solution to exercise:

test statistic:

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{(20-15)^2}{15} + \frac{(10-15)^2}{15} + \frac{(80-85)^2}{85} + \frac{(90-85)^2}{85}$$

$$X^2 = 3.92$$

As $X^2 \geq 3.84$, the p-value is smaller than $\alpha = 0.05$ (more exactly: the p-value is 0.0477) and the null-hypothesis can be rejected.

Alternative solution to exercise:

alternative solution (simpler, if expected cell counts were not already calculated as on slide 135):

test statistic:

$$X^2 = \frac{(n+m)\cdot(a\cdot d - b\cdot c)^2}{(a+b)\cdot(a+c)\cdot(c+d)\cdot(b+d)} = \frac{200\cdot(20\cdot 90 - 80\cdot 10)^2}{100\cdot 30\cdot 100\cdot 170}$$

$$X^2 = 3.92$$

As $X^2 \geq 3.84$, the p-value is smaller than $\alpha = 0.05$ (more exactly: the p-value is 0.0477) and the null-hypothesis can be rejected.

Solving the exercise with the (free) software R

```
> df_furcolor = data.frame(black = c(male=20, female=10),
                            not_black = c(male=80, female=90))
> chisq.test(df_furcolor, correct=FALSE)
```

```
Pearson's Chi-squared test data: df_furcolor X-squared =
3.9216, df = 1, p-value = 0.04767
```

Point estimation for difference of rates

- **point estimation:**

  $\hat{p}_x = \frac{20}{100} = 0.20, \hat{p}_y = \frac{10}{100} = 0.10$

  $\Rightarrow \widehat{\Delta} = \hat{p}_x - \hat{p}_y = 0.10$

- **But where lies $\boldsymbol{\Delta = p_x - p_y}$?** (= the difference in the population)

  $\Rightarrow (1-\alpha)$-confidence interval = interval that includes the true value

  $\Delta = p_x - p_y$ with probability $(1-\alpha)$

Confidence interval for the difference of rates

- **(approx.) two-sided $(\mathbf{1}-\boldsymbol{\alpha})$-confidence interval for $\boldsymbol{p_x}-\boldsymbol{p_y}$:**

  $$[\hat{p}_x - \hat{p}_y - z_{1-\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}};$$

  $$\hat{p}_x - \hat{p}_y + z_{1-\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}]$$

  where $\hat{p} = \frac{1}{n_x+n_y}\left(n_x\hat{p}_x + n_y\hat{p}_y\right)$ "overall rate" in the pooled sample

Confidence interval for the difference of rates (2)

- **width of confidence interval:**

  $$2 \cdot z_{1-\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})} \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

  $\Rightarrow$ The width of the confidence interval is the larger, the larger $(1-\alpha/2)$, the smaller the sample size, and the nearer $\hat{p}$ to 0.5.

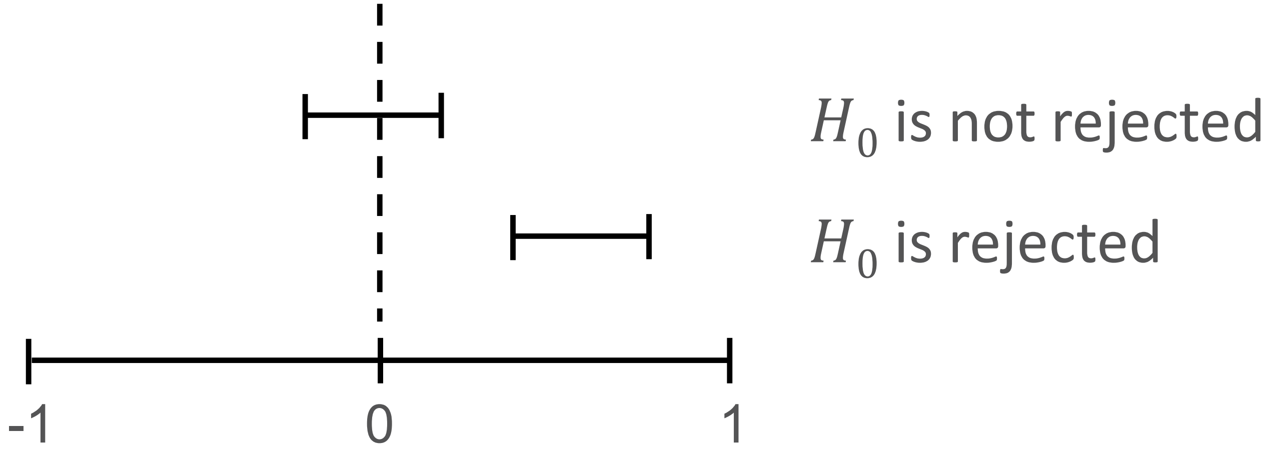
- **duality of test decision and confidence interval:**

  The following events are equivalent:

  – chi-square test rejects $H_0: p_x - p_y = 0$ at type I error rate α

  – value 0 is NOT included in the two-sided $(1-\alpha)$-confidence interval

Confidence interval for the difference of rates (3)



$\Rightarrow$ Knowing the confidence interval, the test decision can be deduced.

$\Rightarrow$ In addition: more precise statements on $\Delta = p_x - p_y$
(not only „$\Delta = 0$“ or „$\Delta \neq 0$“ as with statistical test)

Exercise: the hamster experiment II

Compute the two-sided 0.95 confidence interval for $\Delta = p_x - p_y$ for the hamster experiment II.

$\rightarrow$ hint: The 0.975-quantile of the standard normal distribution amounts to 1.96.

	fur black	fur not black	total
male hamster	20	80	100
female hamster	10	90	100
total	30	170	200

Solution to exercise:

$\hat{p}_x = \frac{20}{100}, \hat{p}_y = \frac{10}{100}, \hat{p} = \frac{1}{200}(20 + 10) = \frac{30}{200} = 0.15$

rate difference: $\frac{20}{100} - \frac{10}{100} = 0.10$

0.95-confidence interval:

$$[\frac{20}{100} - \frac{10}{100} - 1.96\sqrt{\left(\frac{30}{200}\left(1 \; - \frac{30}{200}\right)\right)}\sqrt{\frac{1}{100} + \frac{1}{100}};$$

$$\frac{20}{100} - \frac{10}{100} + 1.96\sqrt{\left(\frac{30}{200}\left(1 \; - \frac{30}{200}\right)\right)}\sqrt{\frac{1}{100} + \frac{1}{100}}]$$

$$= [0.001025; 0.198975]$$

0 is NOT included in 0.95-confidence interval.

This is consistent with the test decision („rejection of null-hypothesis“).

Take-Home-Messages IV

(1) The chi-square test enables confirmatory analysis when comparing two rates. It is valid for „sufficiently large“ sample sizes.

(2) Together with the result of the chi-square test, the point estimate of the rate difference as well as the related confidence interval should be reported.

(3) From the confidence interval, the test decision can be deduced. In addition, the confidence interval provides further information on the true rate difference in the population.

Learning goals

- You know the aims of descriptive data analysis as well as the most important measures of scale and variation and graphics. You are able to apply the related techniques and to interprete the results appropriately. ✓
- You know the aims of confirmatory analysis and the principle of statistical tests. You are able to apply the binomial test and the chi-square test and to interpret the results appropriately. ✓
- You know the concept of point estimation and confidence intervals. You are able to compute the quantities for rates and differences of rates and to interpret the results appropriately. ✓

Many thanks for your attention!

Have (also) fun with parts II and III!