CDC Announces Plan To Send Every U.S. Household Pamphlet On Probabilistic Thinking

| Thursday 1:40PM | Alerts

Satire alert

https://www.theonion.com/cdc-announces-plan-to-send-every-u-s-household-pamphle-1848354068

"the onion", providing laugh therapy for decades!

The Onion

Tutorials ARE not simple exercises OR applications – they are also designed to improve and complement lectures.

In a way, they are additional teaching to strengthen the concepts covered in our lectures as well.

Tutorial 2: Statistical Hypothesis Testing

The principles of statistical hypothesis testing: generating evidence-based conclusions without complete biological knowledge!

"Statistical roads are paved with uncertainty and confidence"

Understanding these two concepts is critical for promoting long-term learning a practice of statistics

These concepts are key to inference & statistical hypothesis testing

Schematic summary of lecture 2

RESEARCH QUESTION

Humans are predominantly right handed. **Do other animals exhibit handedness as well?** Bisazza et al. (1996) tested this possibility on the common toad.

STATISTICAL QUESTION

Do right-handed and left-handed toads occur with equal frequency in the toad (statistical) population, or is one type more frequent than the other?

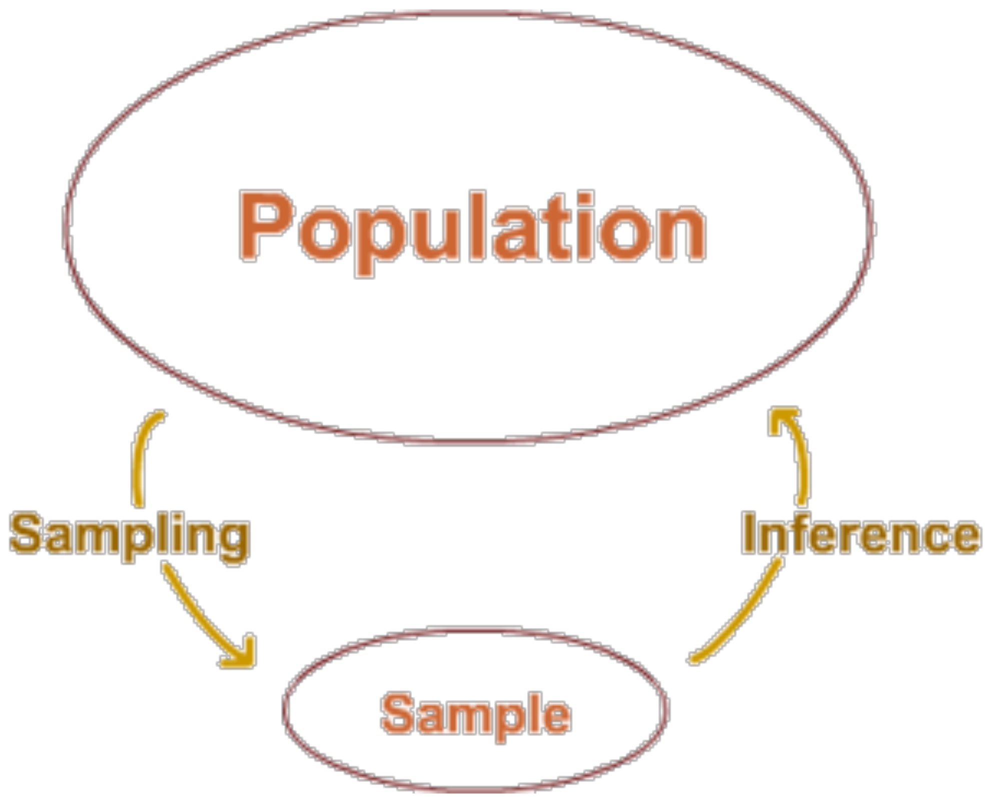
STATISTICAL INFERENTIAL APPROACH

→ **Statistical hypothesis testing.**

→ **Parameter estimation and confidence interval.**

Key statistical element → **Estimating *uncertainty* due to sampling variation**

Statistical inferential process: Estimating uncertainty due to sampling variation to infer from samples (incomplete knowledge or without complete certainty) to entire populations



Sampling variation is (in general) the main source of statistical uncertainty: two or more samples of toads from the same population will always differ in their proportion of right/left leg dominance from the true population value (parameter) due to chance alone!

Remember: random sampling minimizes sampling error, uncertainty & inferential bias (i.e., how close or far sample values for the statistic of interest are from the true population value for that statistic)

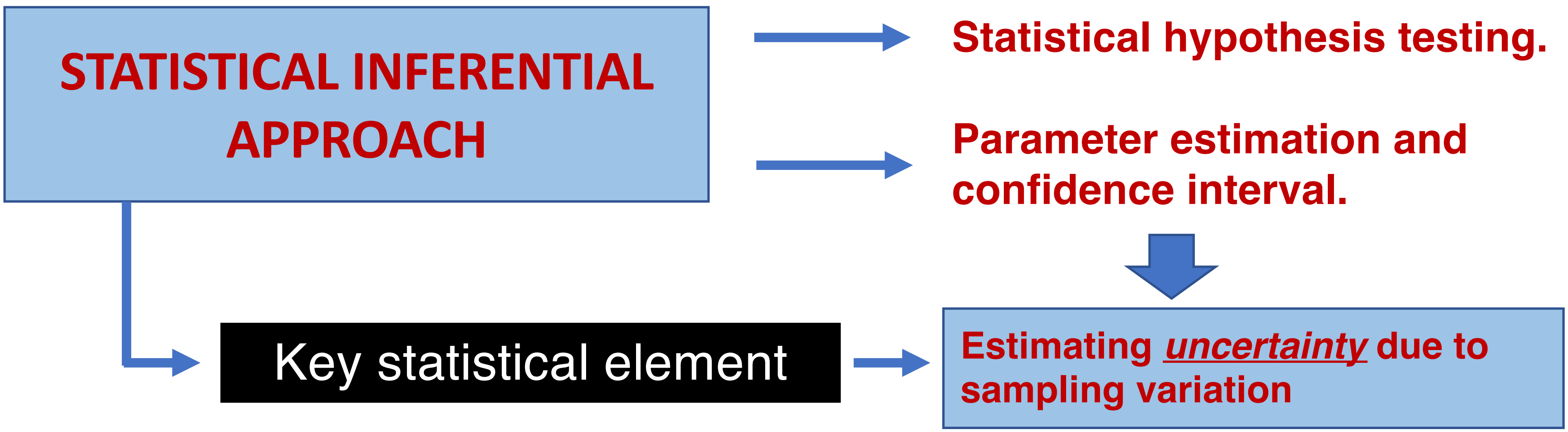
The common requirement for statistical inference is that data come from a **random sample**. A random sample is one that fulfills two criteria:

1) Every observational unit in the population (e.g., individual tree) have an **equal chance** of being included in the sample.

2) The selection of observational units in the population (e.g., individual tree) must be **independent**, i.e., the selection of any unit (e.g., individual tree) of the population must not influence the selection of any other unit.

Samples are biased when some observational units of the intended population have lower or higher probabilities to be sampled.

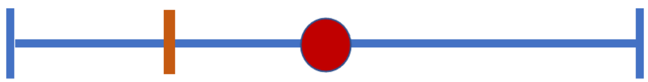
Statistical inferential process: we use **sampling theory** to determine the **sampling distribution** required to **estimate uncertainty** around a sample value of interest (e.g., number of right- and left-legged toads).



Estimation [& associated confidence intervals] and statistical hypothesis testing agree but have different interpretations

45% right-handed

65% right-handed



H_0

55%

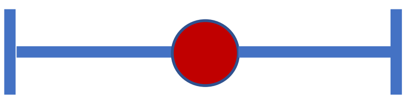
"Quantitative" statement

Don't reject H_0; $p > 0.05$

"Qualitative" statement

30% RH

48% RH



39%

"Quantitative" statement

Reject H_0; $p < 0.05$

"Qualitative" statement

Statistical inferential process: The role of sampling theory in parameter estimate and estimating confidence intervals

```
"The purpose of statistical inference is to develop
theory and methods to make inference on the unknown
  parameters based on observed data" (Hong, 2017)
```

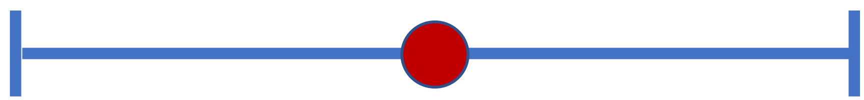
Statistical inferential process: The role of sampling theory in parameter estimate and estimating confidence intervals

"The purpose of statistical inference is to develop theory and methods to make inference on the unknown parameters based on observed data" (Hong, 2017)

we can state that we have some confidence (say 95%) that the true parameter of interest (say number of right-legged frogs is between two values

95% confidence interval for the proportion of right-legged toads

52.4% right-legged 93.6% right-legged



77.8%

(14 right- and 4 left-legged toads, i.e., 14/18 = 77.8%)

A large confidence interval (e.g., 95% or 99%) provides a most plausible range for a parameter of interest (true population value). Values lying within the interval are most plausible, whereas values outside are less plausible, based on the sample data alone.

Statistical inferential process

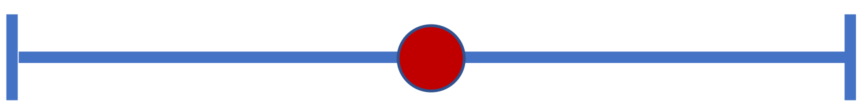
computationally easy but we need to understand this theory to become more comfortable with statistics

```
> binom.test(14,18,0.5,alternative="two.sided")

	Exact binomial test

data:  14 and 18
number of successes = 14, number of trials = 18, p-value = 0.03088
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.5236272 0.9359080
sample estimates:
probability of success
             0.7777778
```

95% confidence interval for the proportion of right-legged toads

52.4% right-legged

93.6% right-legged



77.8%

(14 right- and 4 left-legged toads, i.e., 14/18 = 77.8%)

Statistical inferential process: The role of sampling theory?

If sampling is random and distributional properties of the population are met (e.g., observations are dichotomous), then two critical statistical features will happen:

[1] 95% of the "infinite" (really large value) confidence intervals that could be built based on each possible sample from a given population will contain the true population value.

[2] Because of statement 1, we can be then 95% confident that the interval estimated based on a single sample contains the true population value of the population from where that sample was taken!

Why not a 100% confidence interval? There are some important reasons for that, but one is because intervals (smaller than 100%) become narrower with increased sample sizes (100% intervals would cover the entire possible range for the parameter which is not very useful for inference in general; and in many cases would go from $-\infty$ and $+\infty$).

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the “bag approach”

Take one observational unit (piece of paper) randomly at the time (close eyes and take a paper) out of the bag, write it down whether a left or right and return to the bag (i.e., sampling with replacement*. Repeat this 18 times (i.e., number of toads used by the toad study (Bisazza et al. 1996).

1 sample: 16 R & 2 L
2 sample: 15 R & 3 L
.
.
.
Large number of samples (~Infinite)

sampling distribution for the test statistic of interest for the theoretical statistical population

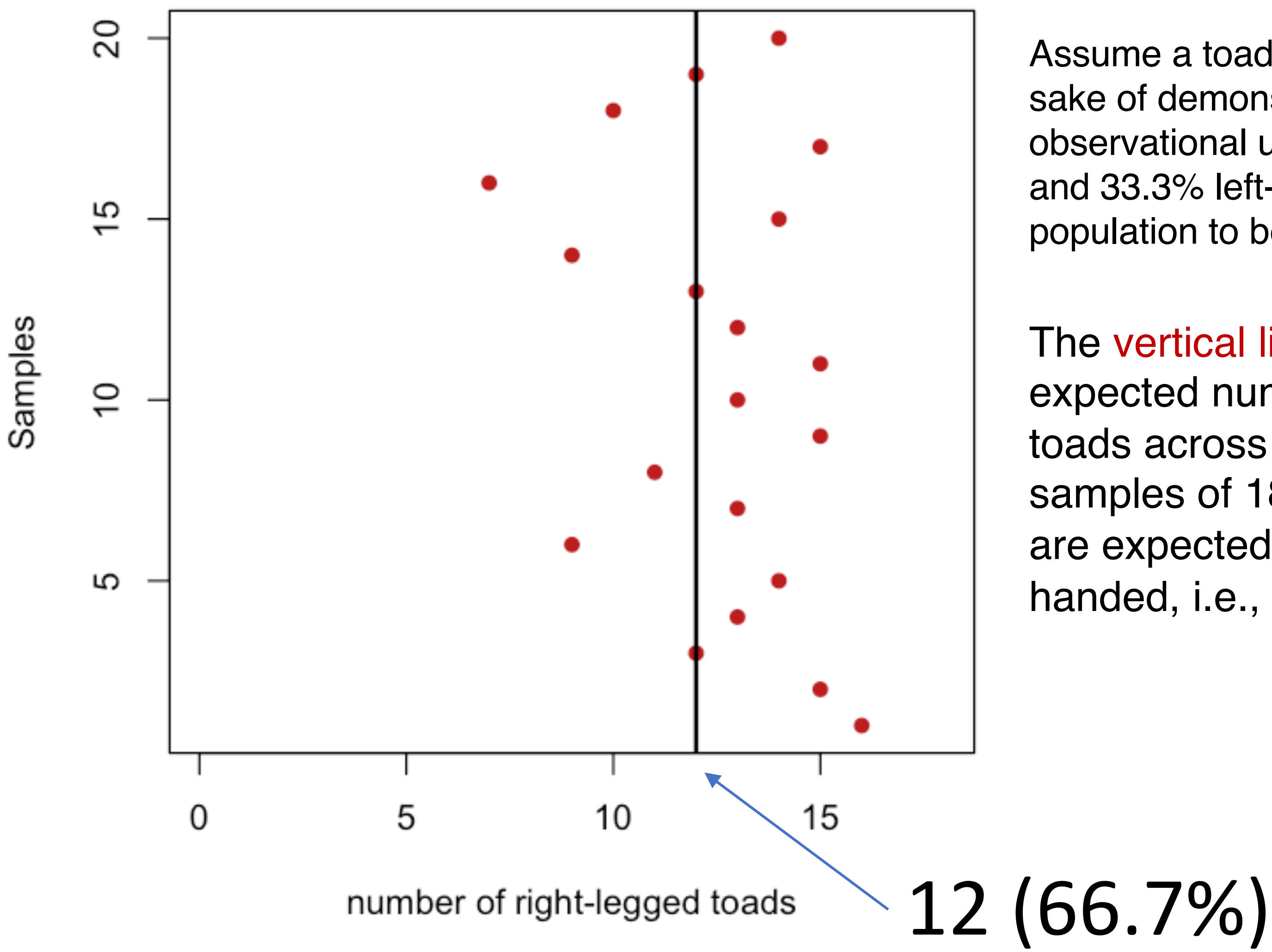
Assume a toad population in which (for the sake of demonstration) 66.7% (12/18) of observational units (toads) are right-legged and 33.3% (6/18) left-handed. Assume this population to be mathematically infinite.

*Resampling (i.e., replacement) is important to assure that the selection of observational units in the population (e.g., individual piece of paper here) must be independent, i.e., the selection of any unit (e.g., L or R) of the population must not influence the selection of any other unit.

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the "bag approach"

Let's imagine 20 samples of 18 toads, each using the "bag approach"



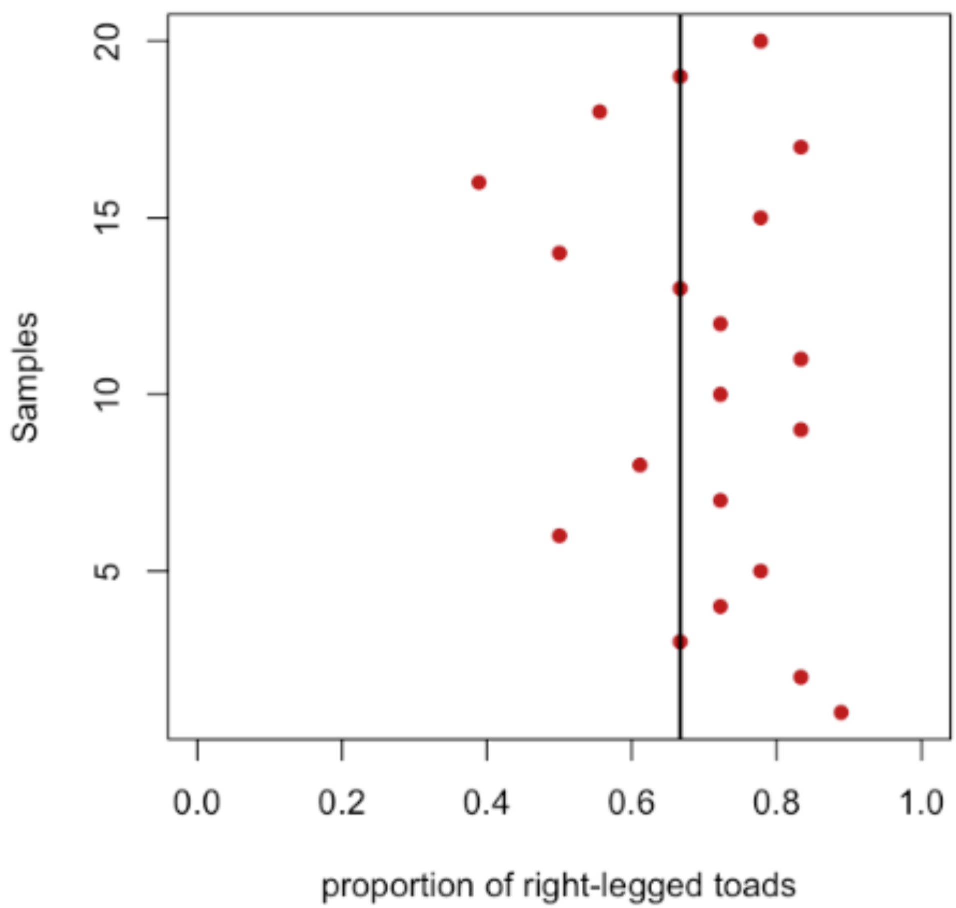
Assume a toad population where (for the sake of demonstration) 66.7% of observational units (toads) are right-legged and 33.3% left-legged. Assume this population to be mathematically infinite.

The vertical line line is the expected number of right-handed toads across all possible infinite samples of 18 toads (i.e., 12 toads are expected in average to be right-handed, i.e., 0.66667*18=12)

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the "bag approach"

More traditionally we use the proportion instead of the numbers (easier to generalize mathematically)



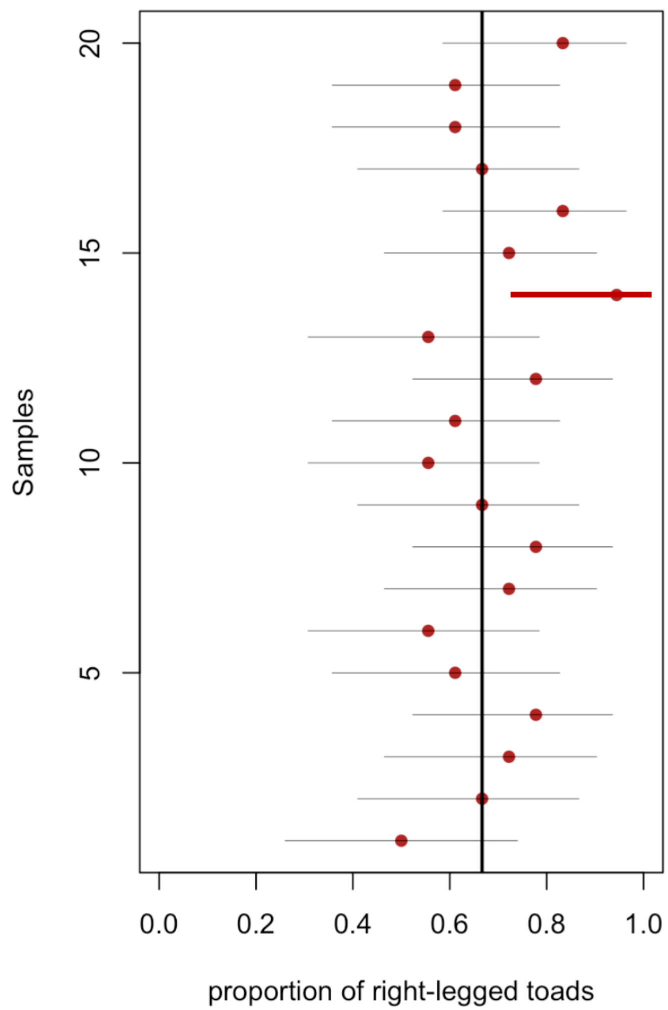
Assume a toad population where (for the sake of demonstration) 66.7% of observational units (toads) are right-legged and 33.3% left-legged. Assume this population to be mathematically infinite.

The vertical line line is the expected number of right-handed toads across all possible infinite samples of 18 toads (i.e., 12 toads are expected in average to be right-handed, i.e., 0.66667*18=12)

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the “bag approach”

More traditionally we use the proportion instead of the numbers
(easier to generalize mathematically)



It is expected that 1 sample confidence interval over 20 does not include the true population proportion (i.e., 95%). This interval is in red.

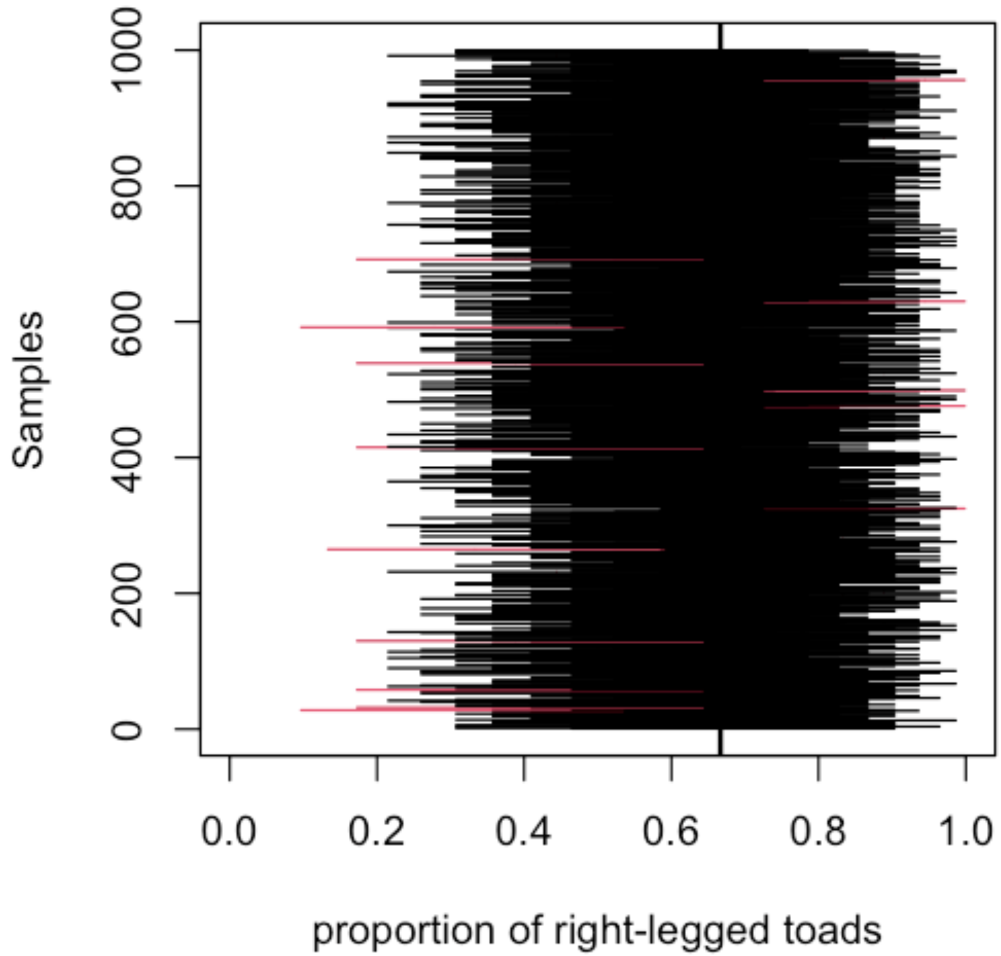
There are different ways to estimate the confidence interval for proportions (here we used the *exact* method which can produce non-symmetric intervals and tend to be wider than the asymptotic estimation).

For now, the estimation method is not important; what’s important is the rationale. We will see some more details later in this lecture.

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the “bag approach”

Now 1000 estimates for the confidence intervals. The point here is that we can produce a really large number of intervals & 95% of them will contain the true population value



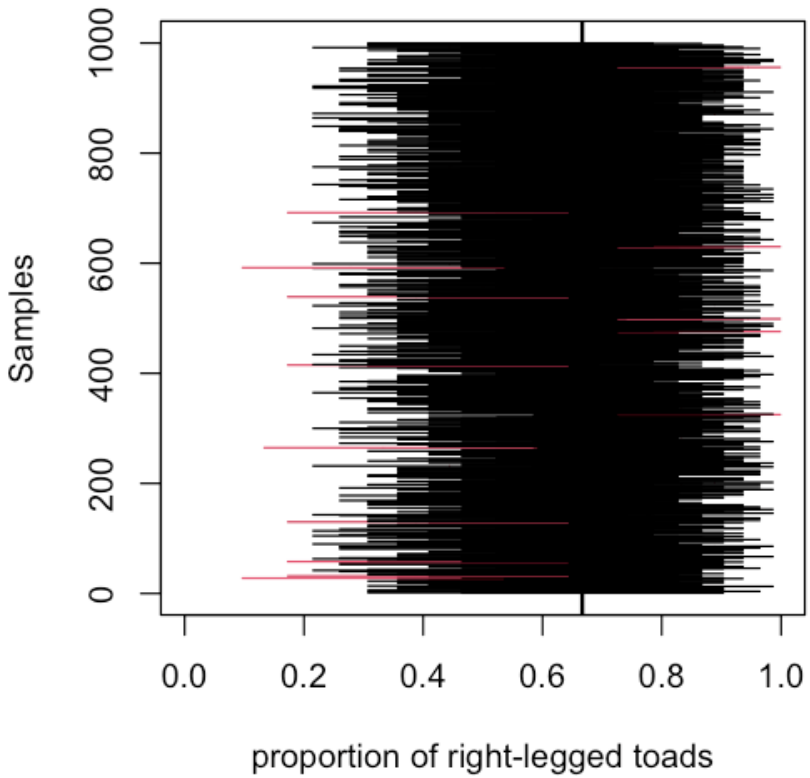
intervals not containing the true parameter are plotted in red (i.e., 5% of the intervals).

The role of sample size in increasing confidence and decreasing uncertainty

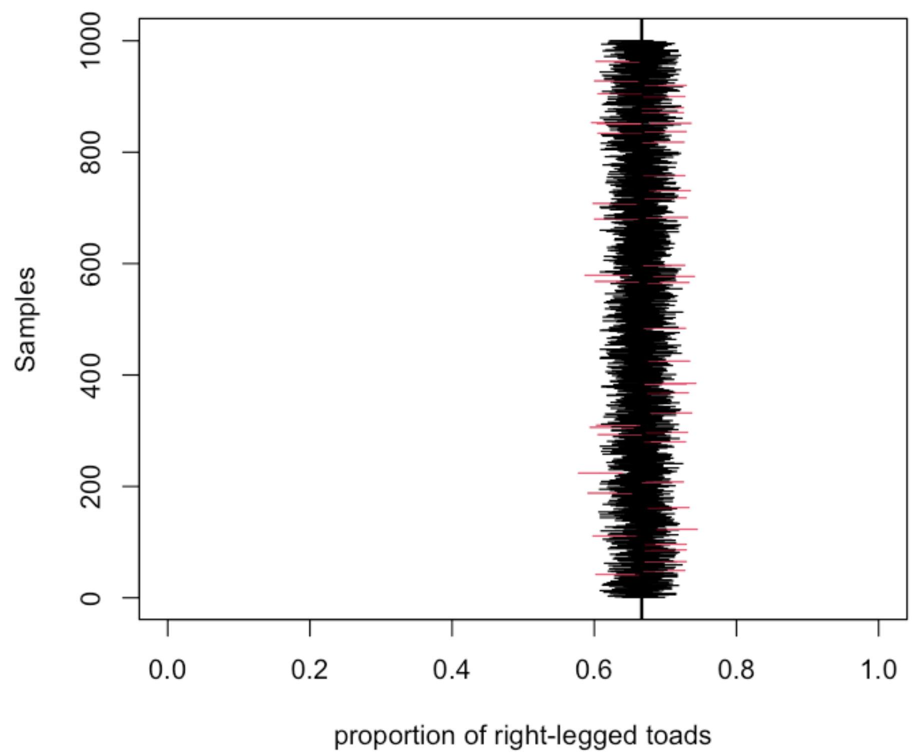
One way to improve estimation is to increase sample size.

Remember though that increasing sample may be extremely timely and financially consuming, and even ethically irresponsible (e.g., collecting and manipulating too many individuals that can put biological populations and species at risk).

Sample size = 18



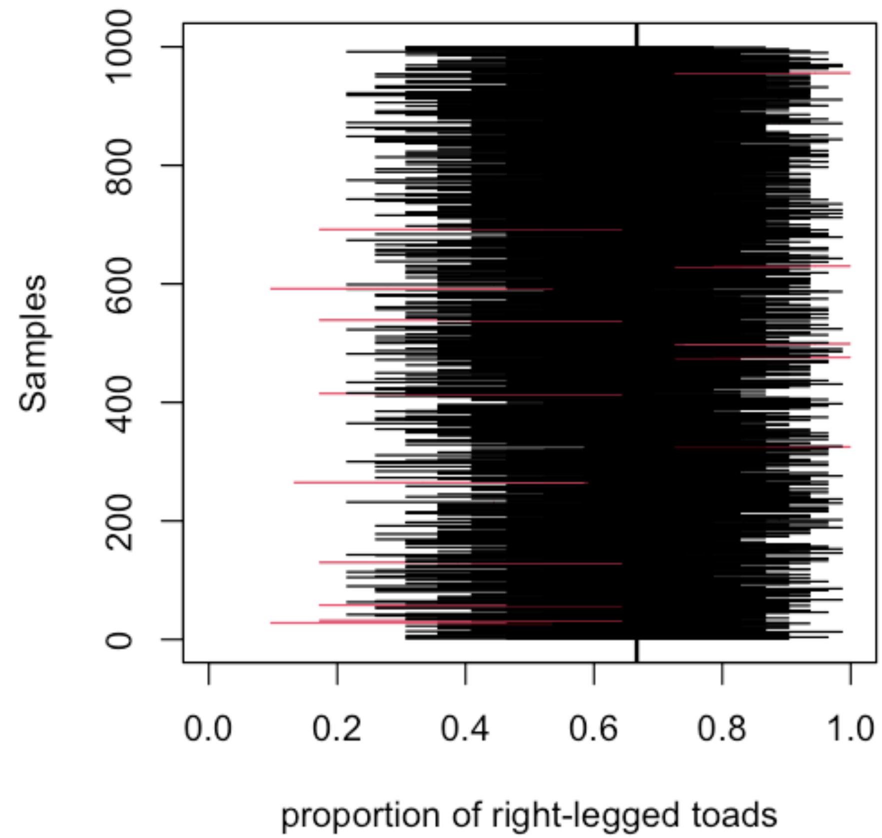
Sample size = 1000



Very important & often confusing!

For any given sample confidence interval, we can state that "we are 95% confident that the true population mean lies between the lower and upper limits of the interval".

We cannot say that "there is a 95% probability that the true population mean lies within the confidence interval". Either the parameter is within the interval or not!



Confidence intervals are not well grasped by a large number of users of statistics!

CANADIAN JOURNAL OF SCIENCE, MATHEMATICS AND TECHNOLOGY EDUCATION, *14*(1), 23–34, 2014
Published with license by Taylor & Francis
ISSN: 1492-6156 print / 1942-4051 online
DOI: 10.1080/14926156.2014.874615

Routledge
Taylor & Francis Group

Confidence Trick: The Interpretation of Confidence Intervals

Colin Foster
School of Education, University of Nottingham, Nottingham, United Kingdom

Often stated by students and practitioners of statistics: "there is a 95% probability that the true population mean lies within the confidence interval". We can't state that; either the parameter is within the interval or not! So, no probability attached to this condition.

Let’s take a break - 2 minutes

Confidence intervals are not well grasped by a large number of users of statistics! RECAP

Confidence intervals is a concept based on sampling theory.

Here, sampling theory relates to repeated sampling making certain assumptions about the statistical population.

We use the principle of repeated sampling to model the expectations of sampling variation. Under repeated sampling, if we were to estimate a confidence interval for each sample, 95% of them would contain the true population parameter.

As such, we can be confidence that one single sample confidence interval (i.e., we usually only have one sample) will most likely contain the true population value.

A large confidence interval (e.g., 95% or 99%) provides a most plausible range for a parameter (true population value). Values lying within the interval are most plausible, whereas values outside are less plausible, based on the sample data alone.

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the “bag approach”

Let’s go back to the original results of Bisazza et al. (1996) in which 14 toads were 14 right-legged and 4 were left-legged (i.e., 77.8% right- and 22.2% left-legged). Let’s estimate its confidence interval based on the “bag approach”

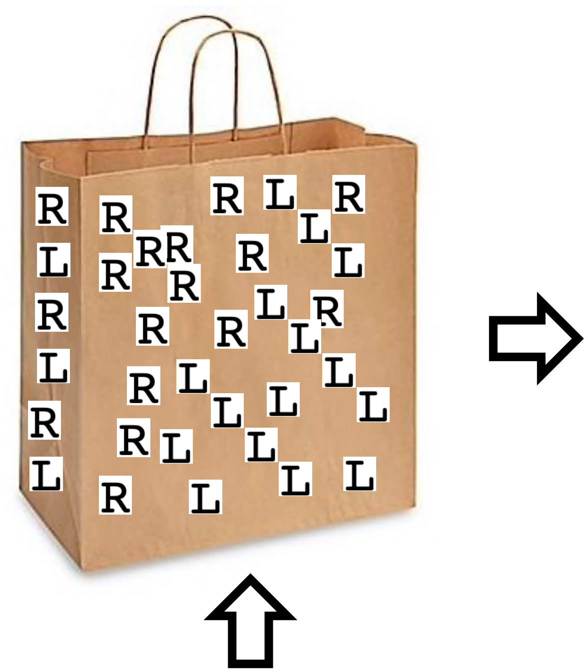
BUT FIRST REMEMBER:

[1] 95% of the “infinite” (really large value) confidence intervals that could be built based on each possible sample from a given population will contain the true population value.

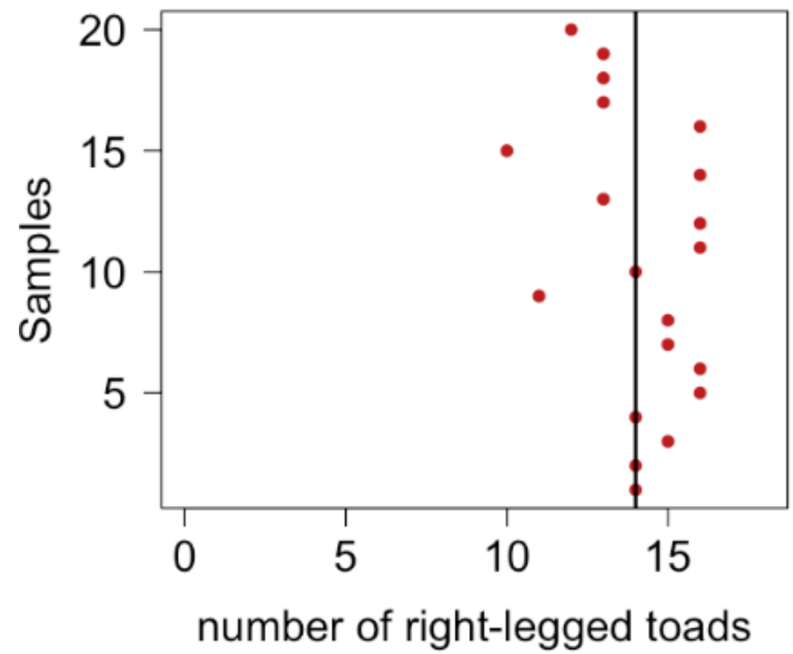
[2] Because of statement 1, we can be then 95% confident that the interval estimated based on a single sample contains the true population value of the population from where that sample was taken!

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the "bag approach"



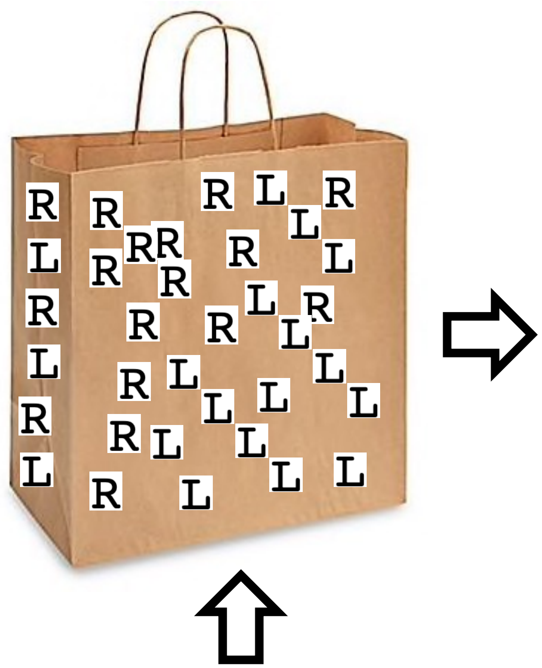
20 random samples

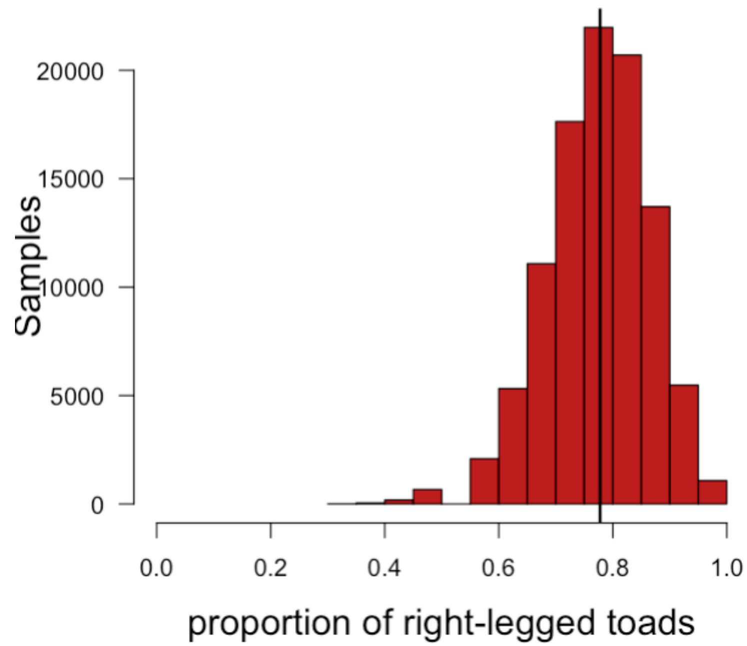


Assume that 77.7% (14/18 individuals) of toad population is right-legged and 22.2%(4/18 individuals) left-handed. Assume this population to be mathematically infinite.

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the “bag approach”





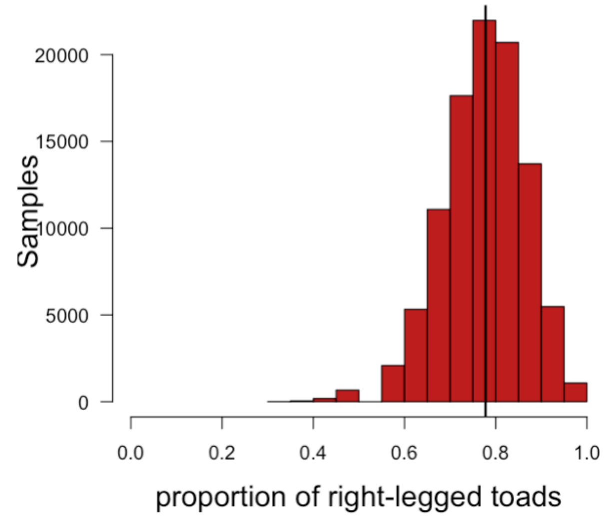
in each of the 100000 samples

Assume that 77.7% (14/18 individuals) of toad population is right-legged and 22.2%(4/18 individuals) left-handed. Assume this population to be mathematically infinite.

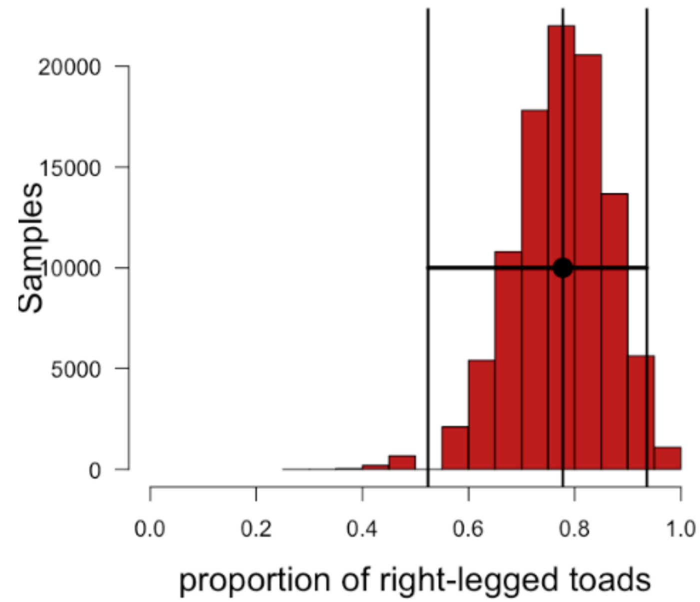
How are confidence intervals computationally derived?

Building knowledge about sampling theory using the "bag approach"

100000 random samples



Assume that 77.7% (14/18 individuals) of toad population is right-legged and 22.2%(4/18 individuals) left-handed. Assume this population to be mathematically infinite.

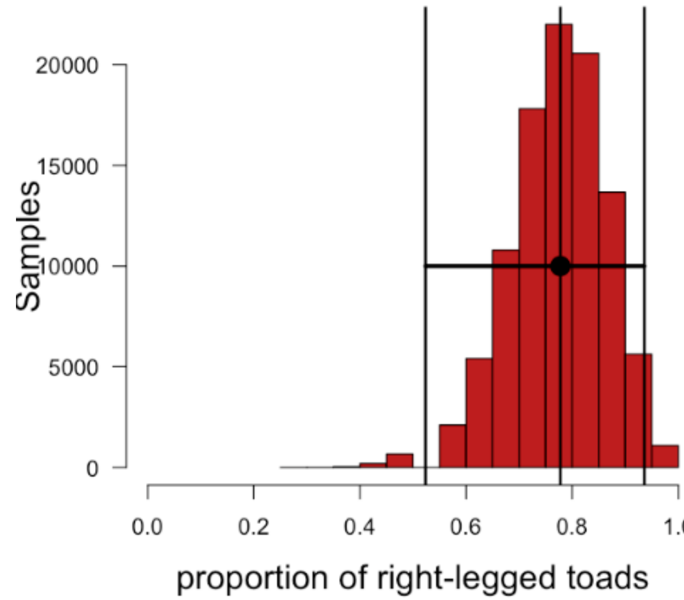


Calculate the 2.5% percentile and 97.5% percentile as the confidence Interval.

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the “bag approach”

100000 random samples



Calculate the 2.5% percentile and 97.5% percentile as the confidence Interval.

Assume that 77.7% (14/18 individuals) of toad population is right-legged and 22.2%(4/18 individuals) left-handed. Assume this population to be mathematically infinite.

In this way, we can be then 95% confident that the interval estimated based on a single sample contains the true population value of the population from where that sample was taken!

How are confidence intervals computationally derived?

Building knowledge about sampling theory using the “bag approach”

For binomial distributions, i.e., distributions that have two possible outcomes (here right- and left-legged individuals), there are a few different ways to estimate confidence intervals – and they differ somewhat, particularly when sample sizes are smaller.

```
> binconf(14,18,method = "all")
            PointEst     Lower     Upper
Exact      0.7777778 0.5236272 0.9359080
Wilson     0.7777778 0.5478542 0.9099907
Asymptotic 0.7777778 0.5857194 0.9698362

> binconf(778,1000,method = "all")
           PointEst     Lower     Upper
Exact         0.778 0.7509431 0.8034091
Wilson        0.778 0.7512054 0.8026670
Asymptotic    0.778 0.7522419 0.8037581
```

This may happen because different methods are used to approximate the distribution of a variable.

Some confidence intervals for some distributions can be modelled exactly (e.g., normal) whereas others only approximated (e.g., binomial).

The exact method is better but can be computationally intense (impractical for large sample sizes) compared to the asymptotic (approximation via a normal approximation); the Wilson is better for small sample sizes than the asymptotic.

PREVIOUS SLIDE INTO "SOLID" WORDS

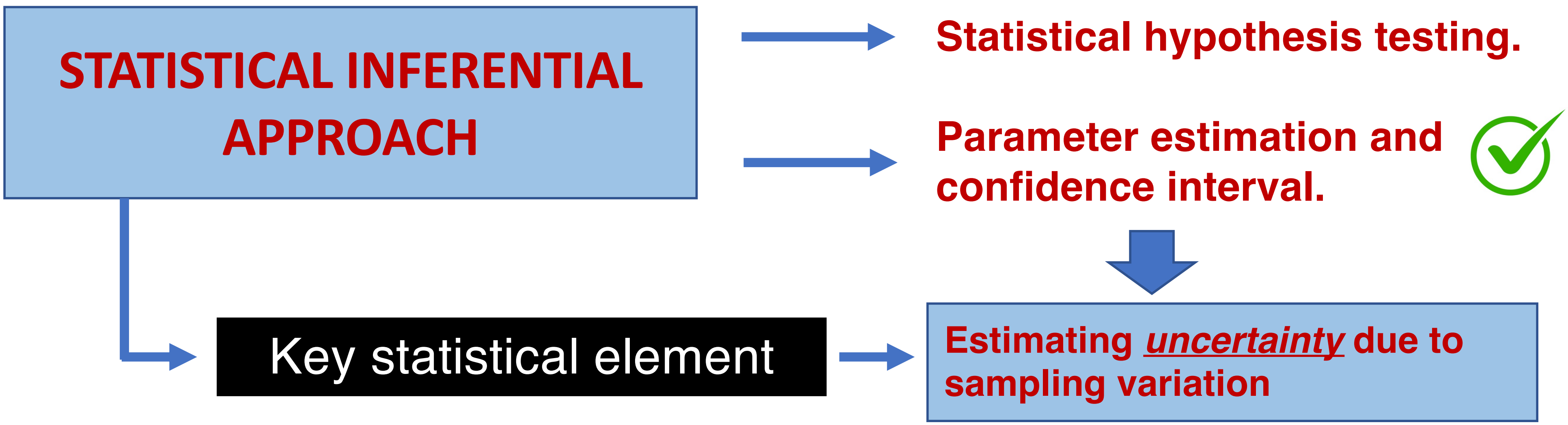
Statistical inferential process: The role of sampling theory?

Estimating uncertainty from sample-based values (information) that allows generalization to entire populations

The basic idea of statistical inference is to assume that the observed data (e.g., 77.8% of frogs were right-legged) is generated from a probability distribution (all possible sample values for the population of interest; e.g., frogs) which is modelled by a function in the form of a probability distribution (e.g., Exact, Wilson, Asymptotic).

Sampling theory is applied to predict sampling uncertainty from sample estimates that is then used to estimate uncertainty. The prediction is made by making assumptions about certain aspects of the sample or populations to estimate the sampling distribution for the value of interest (e.g., number of right- and left-legged toads).

Statistical inferential process: we use sampling theory to determine the sampling distribution required to estimate uncertainty around a sample value of interest (e.g., number of right- and left-legged toads).



Let’s take a break - 2 minutes

statistical hypothesis testing is an intimate stranger!!

Most users know how to implement and interpret it, but they don't really understand its philosophy and how it really works.

Tackling research hypotheses using the framework of statistical hypothesis testing

The **statistical hypothesis framework** (most often involving statistical tests) is a quantitative method of statistical inference that allows to generate evidence for or against a research hypothesis.

CONFUSING: BUT ONLY GENERATES SUPPORT AGAINST THE STATISTICAL NULL HYPOTHESIS (NOT FOR). It also doesn't generate support for (or against) the alternative hypothesis.

But by building support AGAINST a statistical null hypothesis, one builds support FOR research hypothesis.

A small p-value makes us reject the null hypothesis of equal proportion of limb usage and therefore provides support to the research hypothesis of handedness.

The intuition behind the framework of statistical hypothesis testing

You can generate evidence for or against a hypothesis (handedness) using a computational approach (the “bag approach”). All you need is to assume a particular hypothesis as true (null hypothesis) and then reject it (or not) in support of an alternative hypothesis!

Null hypothesis (H_0): the proportion of right- and left-handed toads in the population ARE equal.

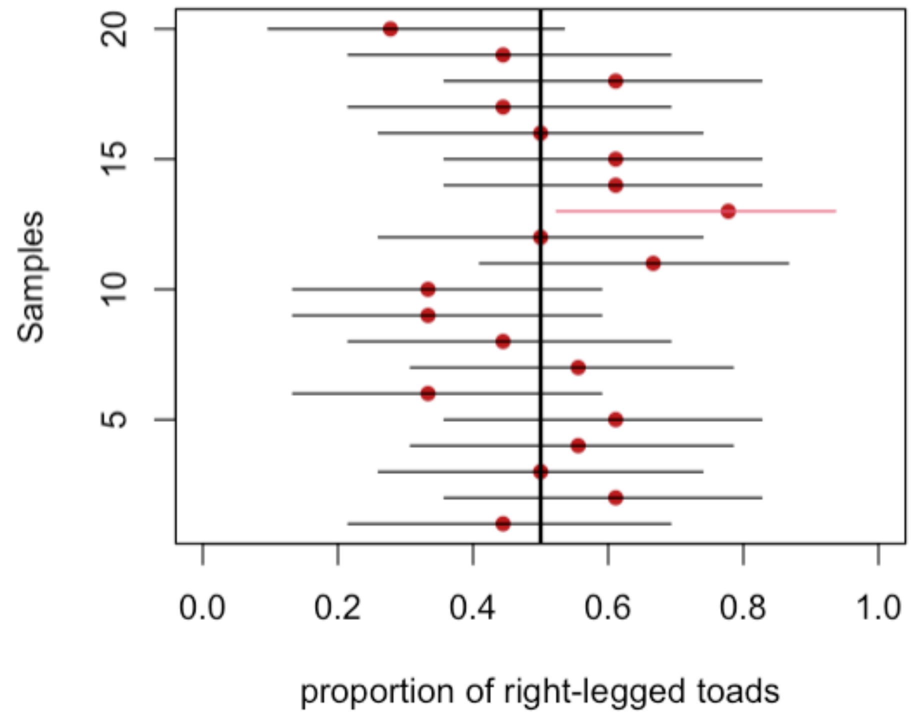
Alternative hypothesis (H_A): the proportion of right- and left-handed toads in the population ARE NOT equal.

It estimates a p-value (seen often as more “quantitative”) than when contrasted to a threshold (alpha, i.e., significance level) it “forces” us into a yes (reject H_0) or no (don’t reject H_0) answer.

Right	Left

Statistical hypothesis testing can be also understood as building the confidence interval for samples that come from a population of “no interest”

20 random samples



Assume that 50% of toad population is right-legged and 50% left-handed. In other words, assume that the null hypothesis is true.

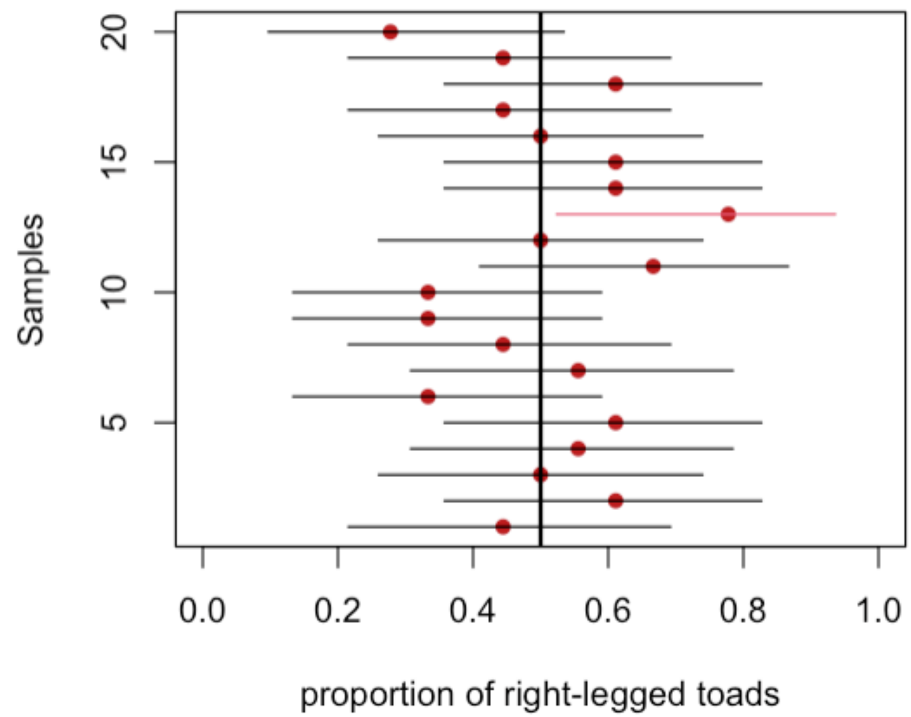
Assume this population to be mathematically infinite.

It is expected that 1 sample confidence interval over 20 does not include the true population proportion (i.e., 95% of intervals). This interval is in red.

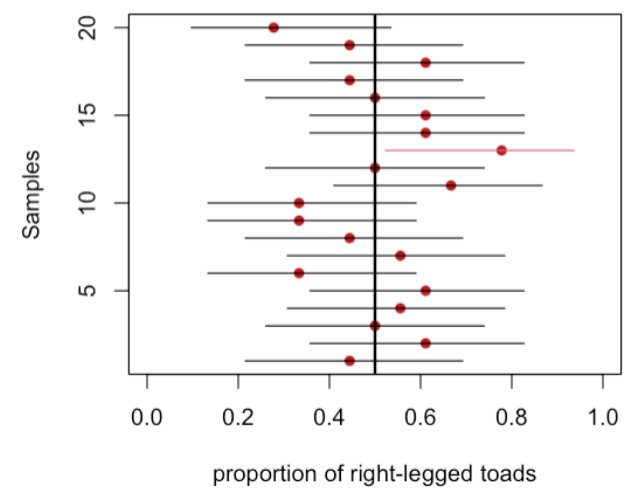
Statistical hypothesis testing can be also understood as building the confidence interval for samples that come from a population of “no interest”

It is expected that 1 sample confidence interval over 20 does not include the true population proportion (i.e., 95% of intervals). This interval is in red.

The principle of statistical hypothesis testing is that if a sample confidence interval covers the value underlying the null hypothesis (here 50%), then we should reject the null hypothesis.



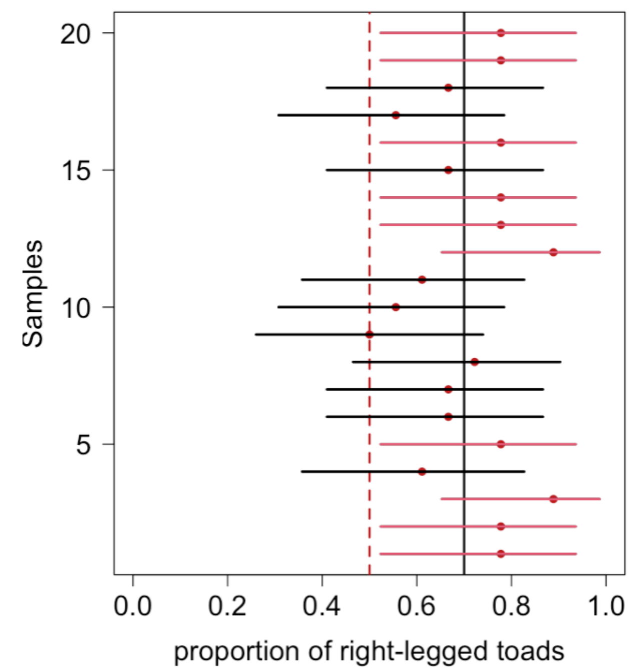
Statistical hypothesis testing can be also understood as building the confidence interval for samples that come from a population of "no interest"



H_0 true

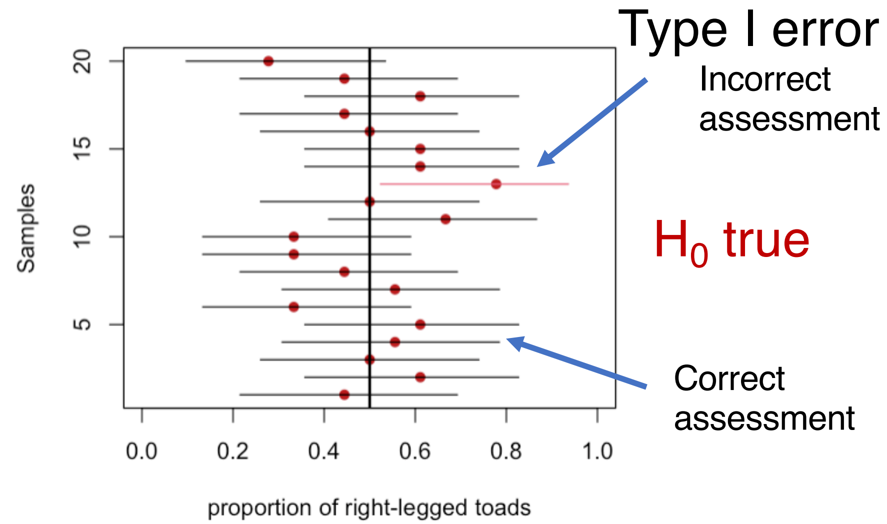
While only 5% of the confidence intervals based on the 50% right-legged population rejects H_0 when it should not (type I error); many intervals also do not reject when they should reject (type II error). Therefore, we can't state that a H_0 is true; all we can say is that we have evidence to reject it.

Type II errors (probability off not rejecting a H_0 that is false) will depend here on the sample size and the value of the true population.



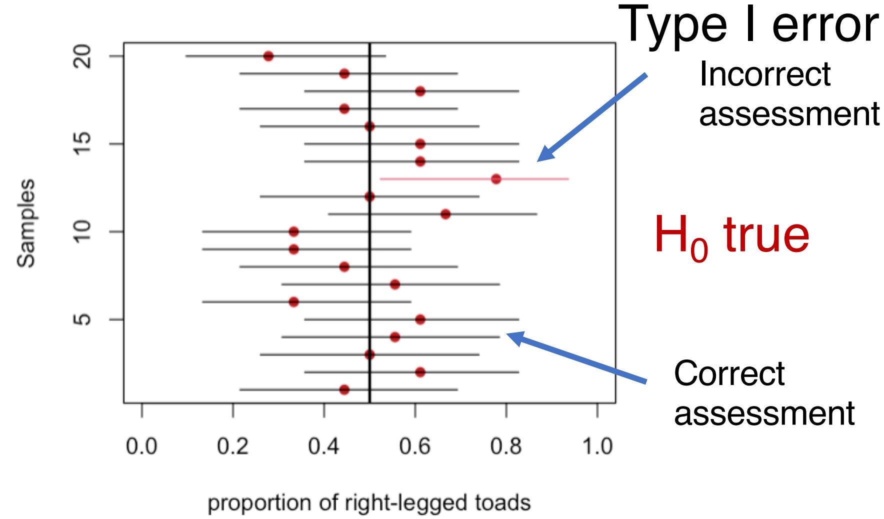
H_0 is false

Errors underlying statistical hypothesis testing – using confidence intervals to increase the ability to understand this notion

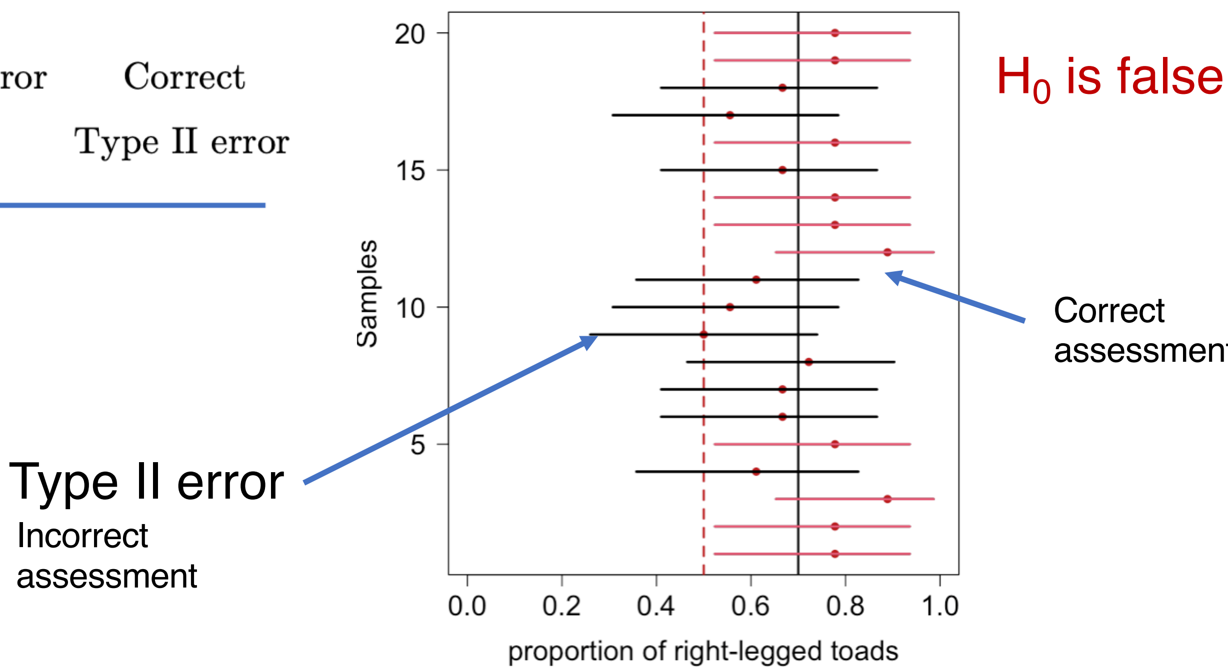


	Reality (unknown)	
Statistical decision	H_0 true	H_0 false
Reject H_0	Type I error	Correct
Do not reject H_0	Correct	Type II error

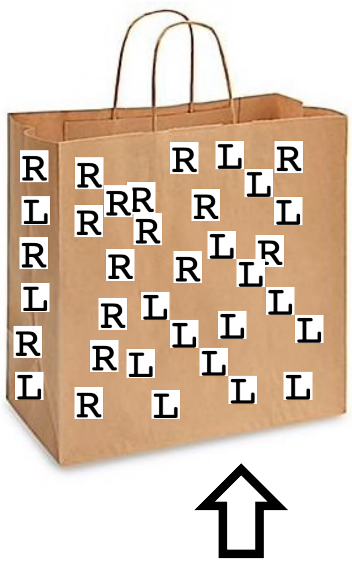
Errors underlying statistical hypothesis testing – using confidence intervals to increase the ability to understand this notion



Reality (unknown)		
Statistical decision	H_0 true	H_0 false
Reject H_0	Type I error	Correct
Do not reject H_0	Correct	Type II error



Statistical hypothesis testing also estimates p-values to support statistical conclusions (don't reject *versus* reject). In this case, one estimates the confidence interval for a sample value that is truly



10000000000 random samples (~infinite)

Assume that 50% of toad population is right-legged and 50% are left-handed. Assume this population to be mathematically infinite.

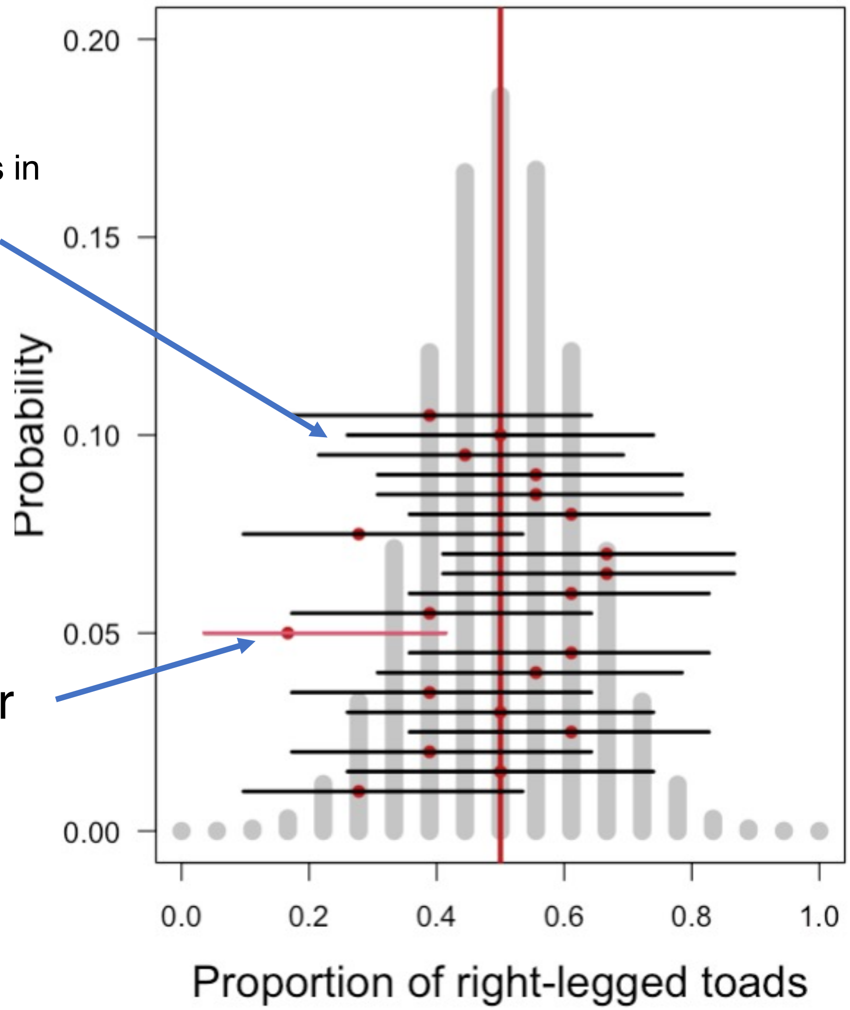
Number of right-handed toads	Probability of those samples
0	0.000004
1	0.00007
2	0.0006
3	0.0031
4	0.0117
5	0.0327
6	0.0708
7	0.1214
8	0.1669
9	0.1855
10	0.1669
11	0.1214
12	0.0708
13	0.0327
14	0.0117
15	0.0031
16	0.0006
17	0.00007
18	0.000004
Total	1.0

Contrasting the confidence intervals from samples from a population where H_0 is true against the null distribution assuming H_0 is true. Any wrong assessment is then obviously an error (wrong assessment).

Confidence intervals in black are correct assessments for those samples

Type I error

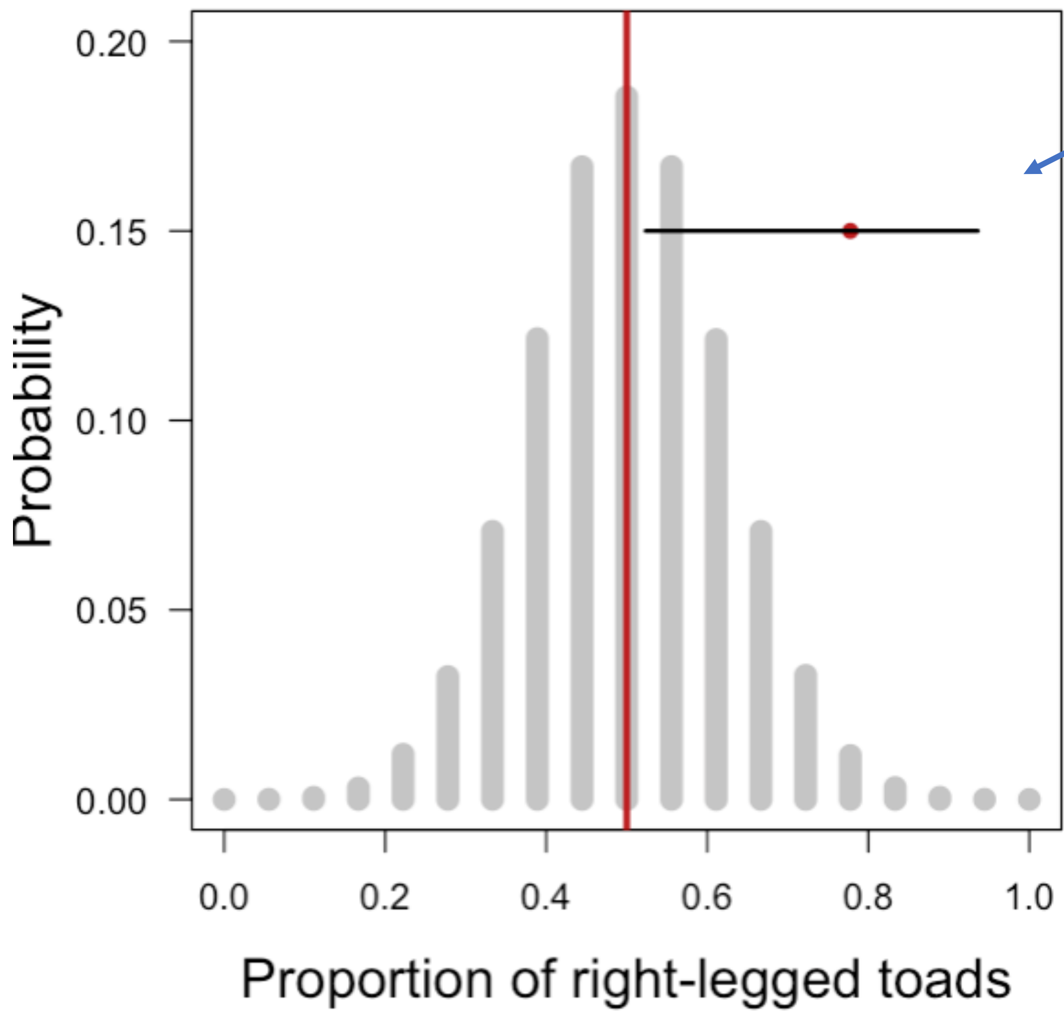
Incorrect assessment (in red)



We know which assessments are correct and incorrect here because we generated samples that also come from the population where H_0 is true.

In reality we don't know whether our samples are or are not from the population assumed to build H_0. This was obviously done for pedagogical purposes to show you the properties of confidence intervals and statistical hypothesis testing.

Contrasting the confidence intervals from the observed sample (14 right-legged) against the distribution of values under the H_0



Confidence interval for the observed value 14 right-legged (i.e., 77.8%)

Remember that we don't know reality because it's all based on a sample.

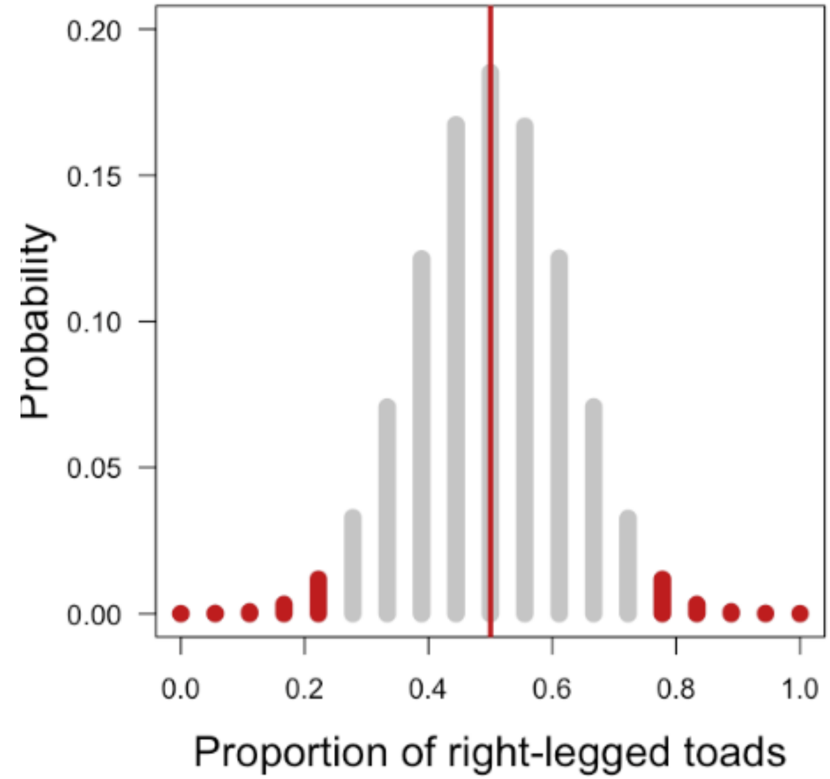
But we can say that we are 95% confident that this is NOT a type I error.

Because the answer (reject or not) always agree, we could estimate the p-value from the confidence interval for the sample and the null distribution. This is a bit advanced for our level right now.

Number of right-handed toads	Probability
0	**0.000004**
1	**0.00007**
2	**0.0006**
3	**0.0031**
4	**0.0117**
5	0.0327
6	0.0708
7	0.1214
8	0.1669
9	0.1855
10	0.1669
11	0.1214
12	0.0708
13	0.0327
14	**0.0117**
15	**0.0031**
16	**0.0006**
17	**0.00007**
18	**0.000004**
Total	1.0

equal or smaller sum [P]=0.0155

equal or greater sum [P]=0.0155



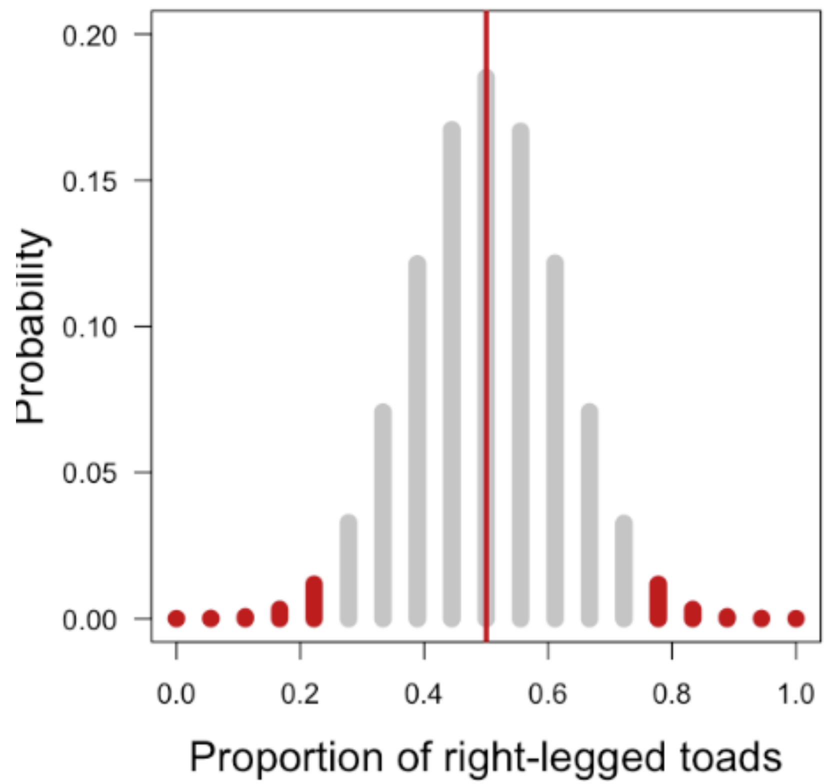
Pr[14 or more right-handed toads] = Pr[14] + P[15] + P[16] + P[17] + P[18] = 0.0155 x 2 (symmetric distribution) = **0.031**

OR: Pr[14 or more right-handed toads] +
Pr[4 or less right-handed toads] = 0.031

OR: Pr[14 or more left-handed toads] +
Pr[14 or less right-handed toads] = 0.031

The p-value is NOT the probability that the null hypothesis is true. IT IS the probability of observing a value of the test statistic that is as or more extreme than what was observed in the sample, assuming the null hypothesis is true.

The p-value is a measure of consistency between the sample data and the theoretical hypothesis assumed when stating the parameter for a theoretical population of no interest (null hypothesis, e.g., toads have equal number of individuals right and left-handed)



Pr[14 or more right-handed toads] =
Pr[14] + P[15] + P[16] + P[17] + P[18] =
0.0155 x 2 (symmetric distribution) = **0.031**

Let’s take a break - 2 minutes

Decision in statistical hypothesis testing – what do P-values represent?

The **p-value** is the probability of the observed sample data assuming that the null hypothesis is true.

The smallest the P-value, the stronger the evidence against the initial assumption (model) based on the parameter assumed for the theoretical population (i.e., null hypothesis).

That's not to say that handedness is true OR false but rather that we have strong evidence to say that lack of handedness (i.e., 50%/%50) is unlikely.

Decision in statistical hypothesis testing – what do P-values represent?

P = 0.031

AGAIN and VERY IMPORTANT and also “confusing”:

So we can say that we have evidence to reject the null statistical hypothesis BUT we cannot say that we have evidence to accept the alternative statistical hypothesis.

That’s because we made our decision based on the sampling distribution of values expected under chance alone from a population where the null hypothesis H0 is true (i.e., 50% right-legged and 50% left-legged).

BUT, by rejecting the statistical null hypothesis, we build evidence towards the research hypothesis and not towards accepting the alternative hypothesis (remember that the null distribution is built based on the null as there are infinite possible alternative hypothesis).

Decision in statistical hypothesis testing – what do P-values represent?

Research hypotheses cannot be proven right or wrong from the data. Hypotheses can be said to be either refuted (evidence is against the research hypothesis) or supported (evidence is in favour of the research hypothesis) by the data generated.

The p-value is NOT the probability that the null hypothesis is true. IT IS the probability of observing a value of the test statistic that is as or more extreme than what was observed in the sample, assuming the null hypothesis is true.

The p-value is a **measure** of **consistency** between the sample data and the theoretical hypothesis assumed when stating the parameter for a theoretical population of no interest (null hypothesis, e.g., toads have equal number of individuals right and left-handed)

The process of statistical hypothesis testing: SUMMARY OF critical details

Statistical hypothesis testing asks how unusual it is to get the observed value for the sample data within the distribution built assuming the null hypothesis as true.

Statistical hypotheses are about populations but are tested with data from samples.

Statistical hypothesis (usually) assumes that sampling is random.

The null hypothesis is usually the simplest statement, whereas the alternative hypothesis is usually the statement of greatest interest.

A null hypothesis is often specific (specific parameter for the theoretical population); an alternative hypothesis often is not.

What does the significance level (α level) represent?

There is disagreement among statisticians and users about whether to reject or not reject (referred as to thresholding) statistical hypotheses based on p-values.

i.e., whether to use α as a threshold for making a decision to state whether a p-value is non-significant (do not reject H_0) or a p-value is significant (reject H_0 in favour of H_A).

Although I agree with these arguments it is unlikely that radical changes will arrive in research behaviour any time soon!

Taylor & Francis

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://amstat.tandfonline.com/loi/utas20

Moving to a World Beyond "$p < 0.05$"

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019) Moving to a World Beyond "$p < 0.05$", The American Statistician, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

The don’ts about P values and hypothesis testing (Wasserstein et al. 2019)

1. P-values indicate how incompatible the observed data are with a specified statistical model (e.g., the one assumed under H_0).

2. P-values do not measure the probability that the studied research hypothesis is true.

3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold (alpha) – (even though they currently are)

4. A p-value, or statistical significance, does not measure the biological importance of a result.

- There are other important don’ts that we will see later in the course.


The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://amstat.tandfonline.com/loi/utas20

Moving to a World Beyond “$p < 0.05$”

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019) Moving to a World Beyond “$p<0.05$”, The American Statistician, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

Use p-values using "the language of evidence" against H_0

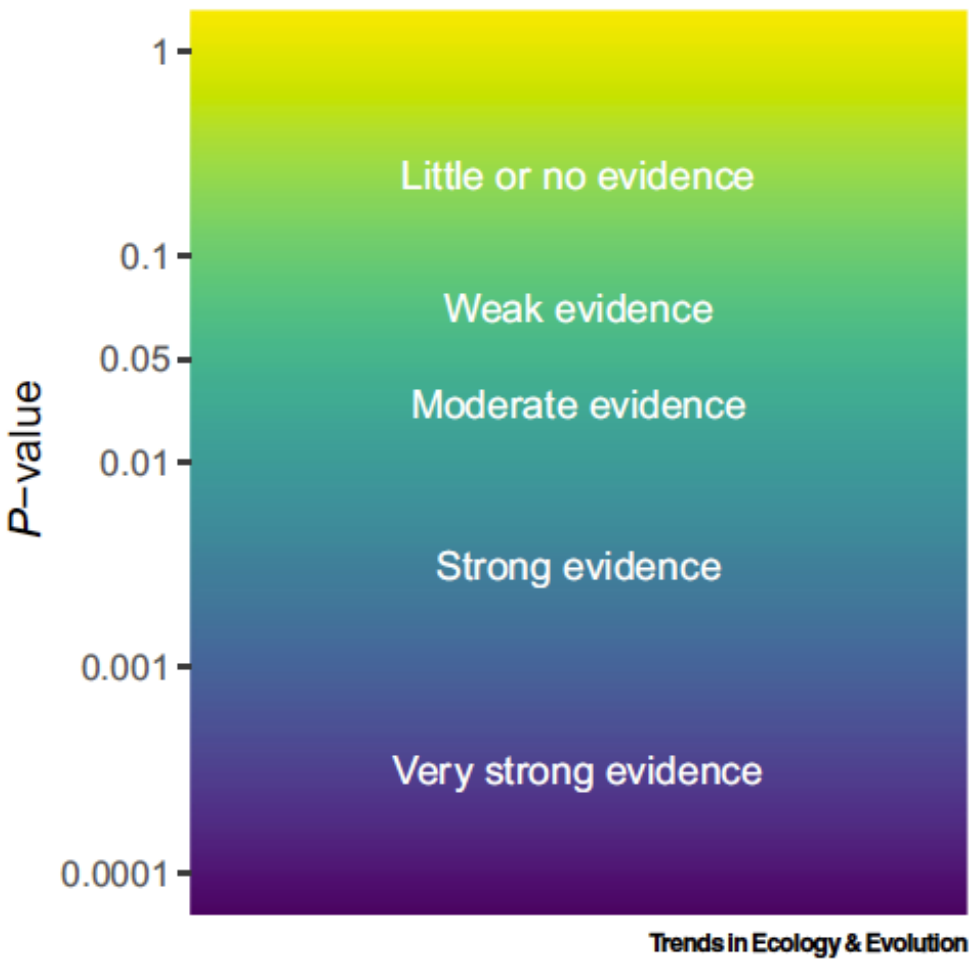


Figure 1. Suggested ranges to approximately translate the *P*-value into the language of evidence. The ranges are based on Bland (1986) [27], but the boundaries should not be understood as hard thresholds.

Note: because the p-value is based on the H_0, the evidence is against H_0 and not in favour of H_A. So, we have evidence to reject H0 (one fixed assumed parameter) but not accept HA (many potential parameters can fit HA (e.g., 55%/45%, 80%/20% right-handed, etc)

Stefanie Muff et al. 2022. Rewriting results sections in the language of evidence. Trends in Ecology and Evolution 3:203-210.

The don’ts about P values and hypothesis testing (Wasserstein et al. 2019)

Despite the limitations of p-values, we are not recommending that the calculation and use of p-values be discontinued. Where p-values are used, they should be reported as continuous quantities (e.g., $p = 0.08$) and not yes/no reject the null hypothesis.

The biggest push today is to abandon the idea of statistical significance. In other words, to abandon the almost universal and routine practice to state that if the probability is smaller than or equal to alpha, than we should state that the results are significant.

Abandoning significance is easily said than done. The majority of researchers do report results as significant or non-significant. We will try to guide you in a more nuanced ways in our course but it’s hard to get away from this common culture in the statistical applications in biology and in most other fields.

Statistical inferential process: we use sampling theory to determine the sampling distribution required to estimate uncertainty around a sample value of interest (e.g., number of right- and left-legged toads).