Statistical Principles

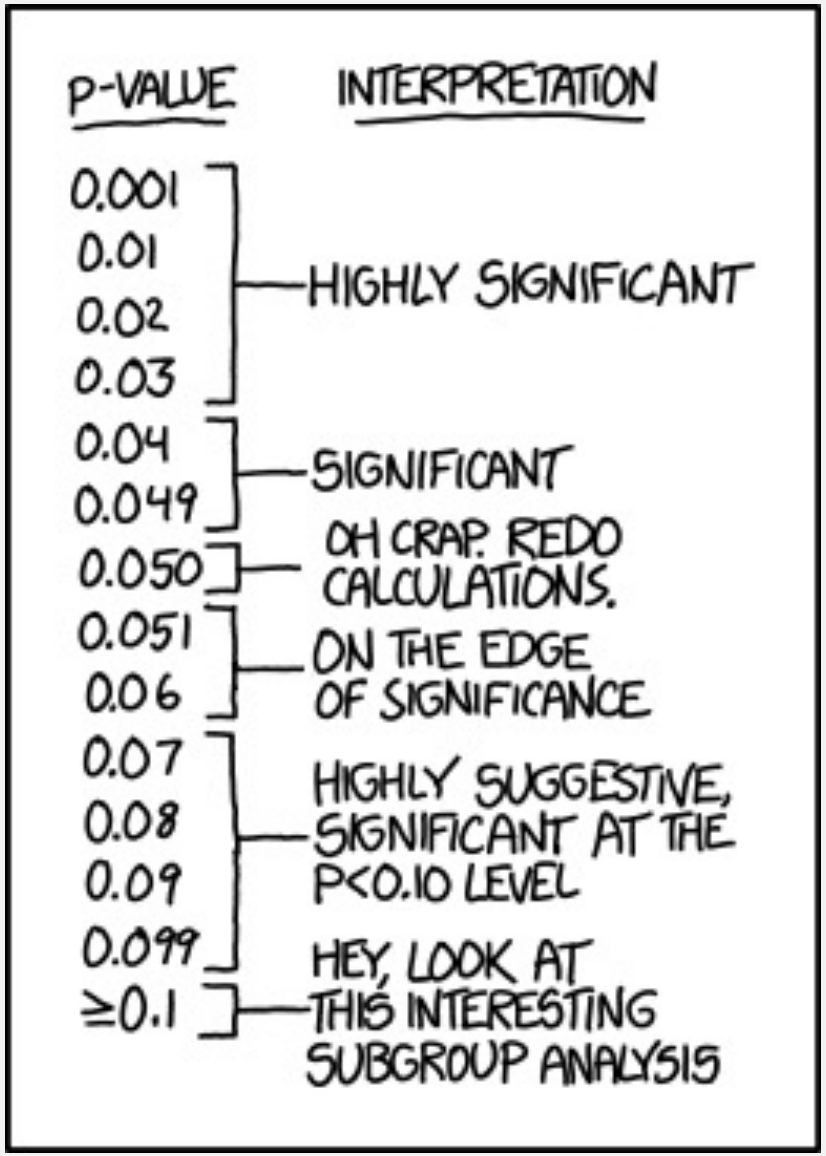
Biology 683

Heath Blackmon

Example test questions:

Describe the differences and advantages of vector vs raster images.

What is a sample vs a population.



Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.

Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

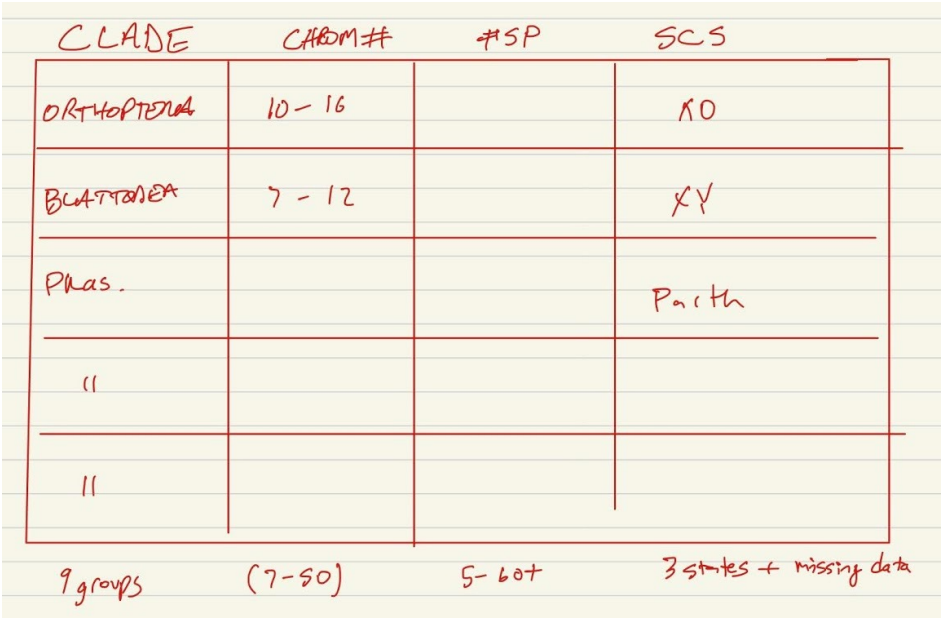
Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.

Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.



Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.

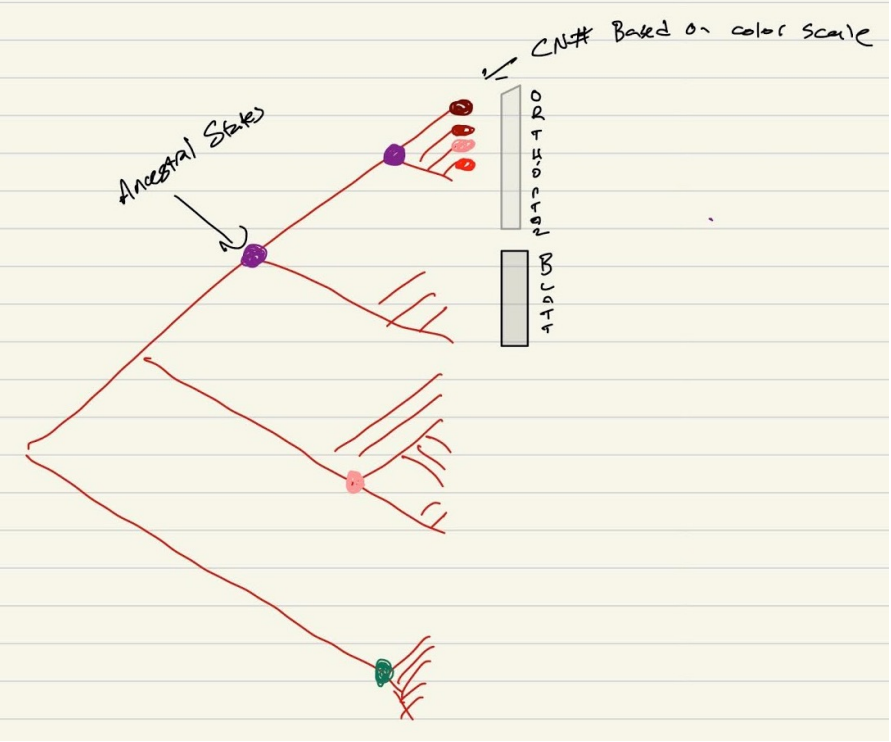
Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.



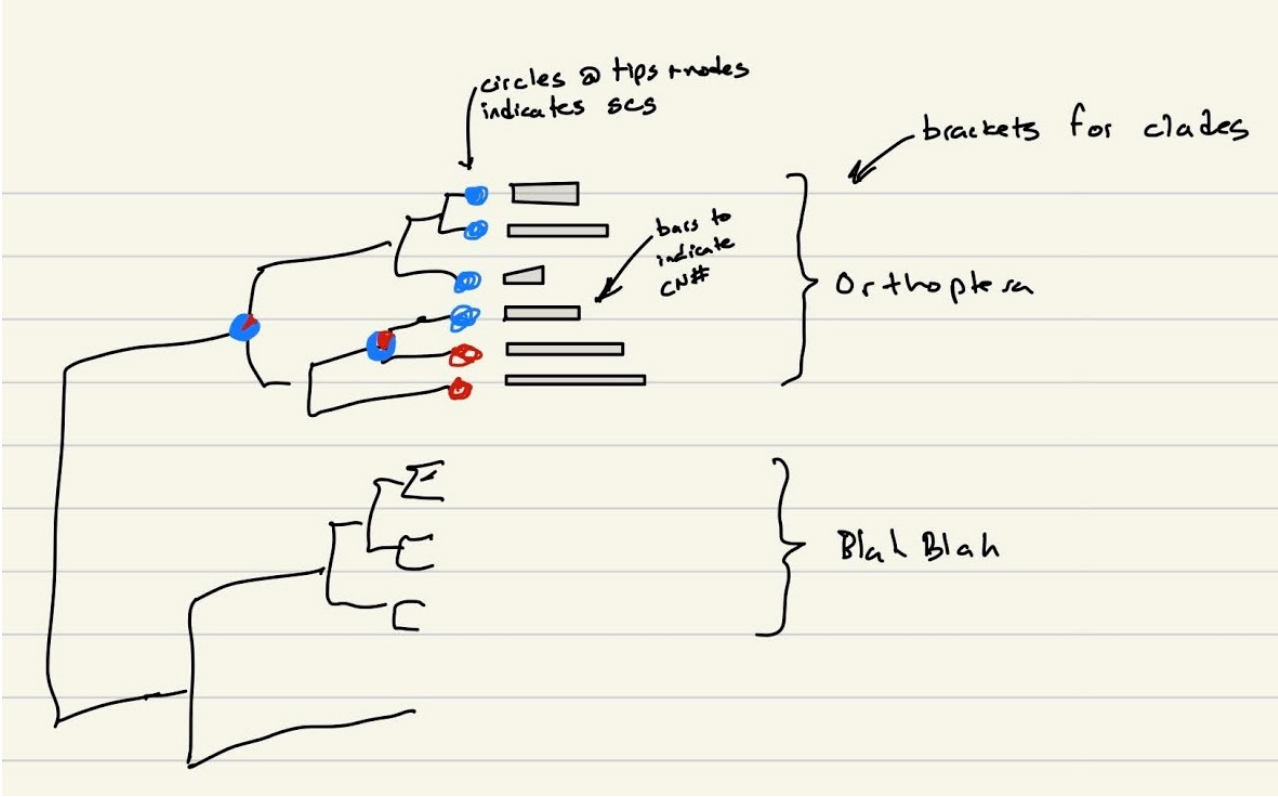
Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.



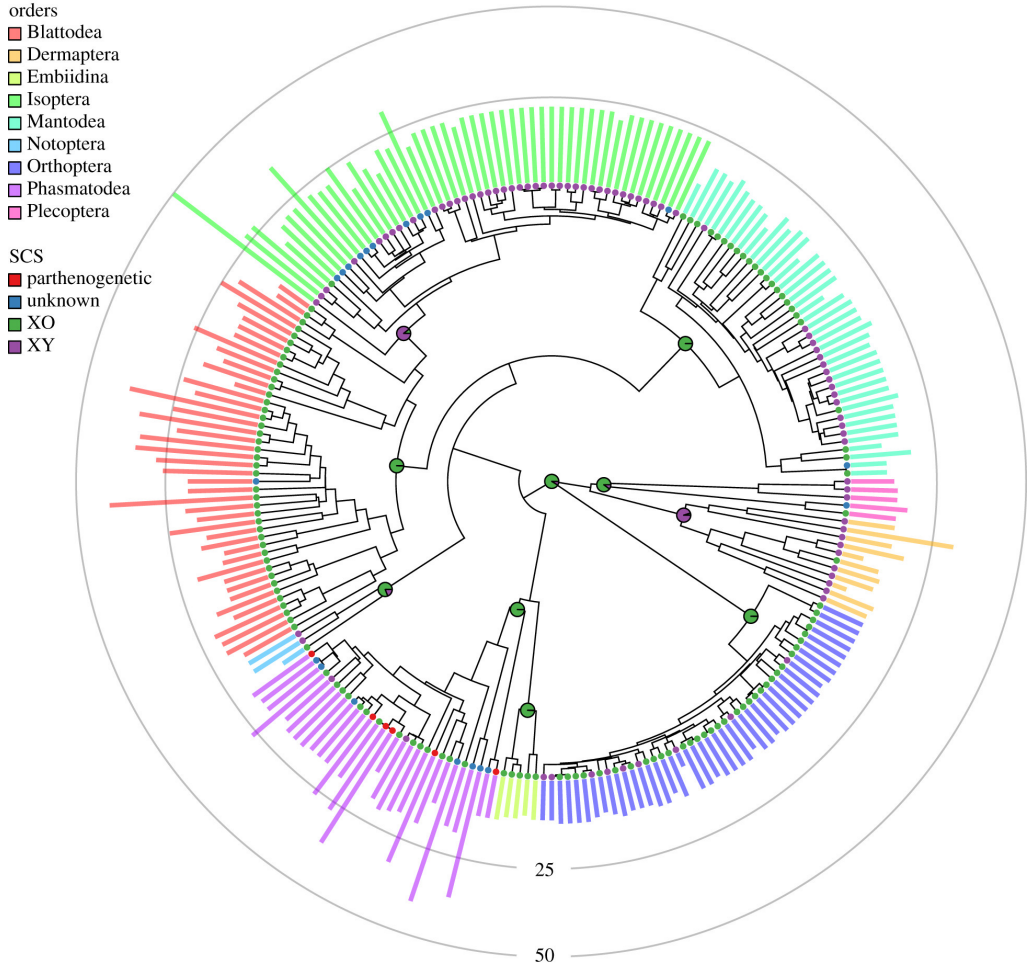
Steps in making a great figure

1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.



Central limit theorem

- Imagine that we sample from the same population many times, so we have a bunch of different, independent samples.

- Each sample will have a mean, but the means will be different due to chance. In principle, we could draw a histogram of these means.

- In general, you only have one sample from a given population, however, so what can you infer about the distribution of the means from your sample?

- The Central Limit Theorem states that regardless of the underlying population distribution of the variable of interest, the distribution of the population of means will be roughly normal.

Central limit theorem

Your estimate of the sample mean is an estimate of the mean of this distribution of means (that is, it's your best estimate of the population mean).

The hypothetical distribution of sample means has a standard deviation equal to s divided by the square root of n.

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

We call this standard deviation the standard error of the mean (SEM). The true population mean should be within $\bar{Y} \pm 1.96SE_{\bar{Y}}$ 95% of the time

Central limit theorem

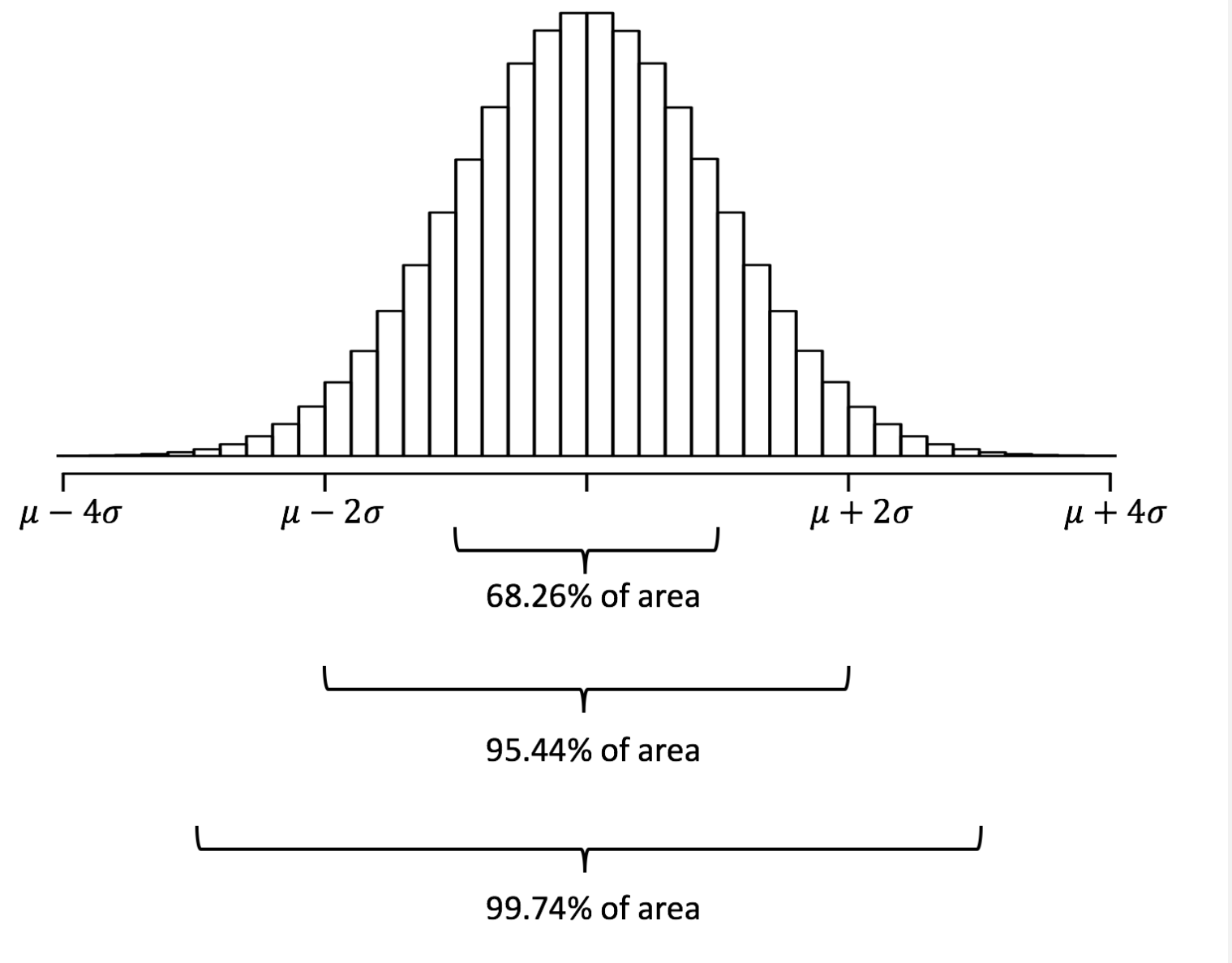
Lets try that

create a population with a known mean.

sample from it and calculate the mean and standard error and see if it includes the true mean.

tally results and see if it worked about 95% of the time

Estimating with uncertainty



Confidence Interval vs Credible Interval

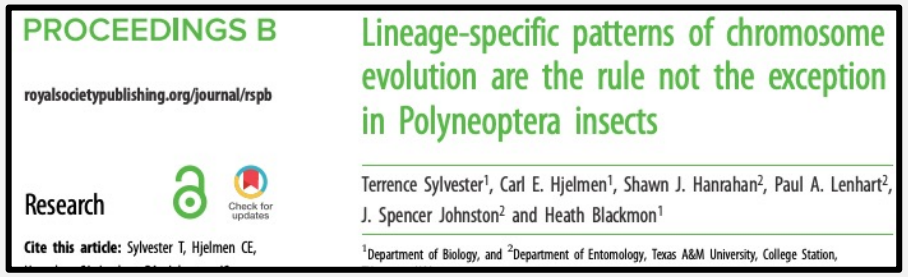
$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

z = 1.65 fo 90%
= 1.96 for 95%
= 2.58 for 99%

natural choice for things we go and measure in biological entities and we are interested in what the “true” mean value of the population is

Credible intervals are often used in Bayesian approaches. In these methods we often run an MCMC which yields an arbitrarily large number of estimates of our parameter of interest. It is not sensible to talk about the CI of a parameter estimate like this because it can always be narrowed to a point estimate with sufficient sample size.

Confidence Interval vs Credible Interval



PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Sylvester T, Hjelmen CE,

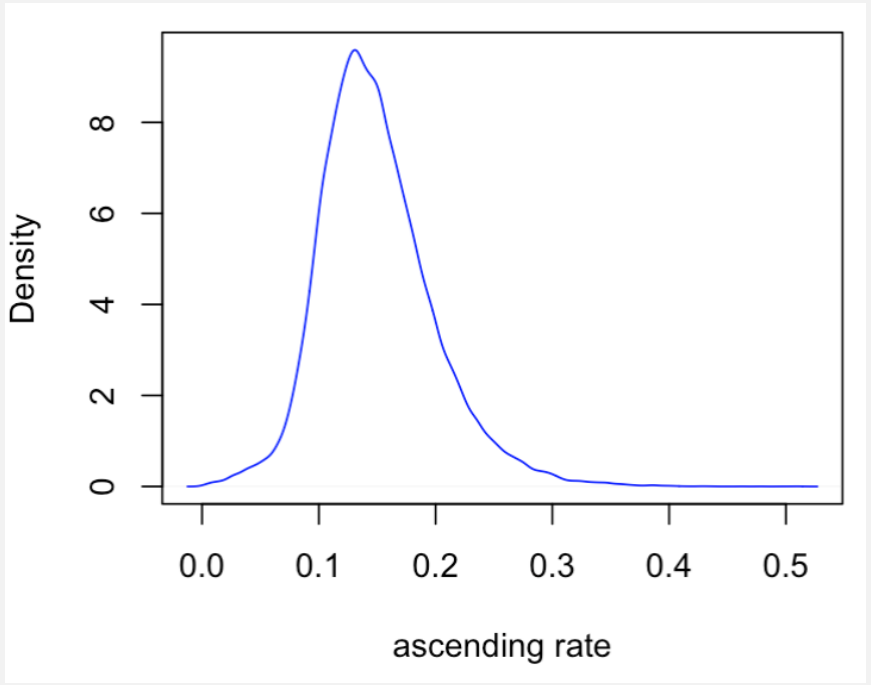
Lineage-specific patterns of chromosome evolution are the rule not the exception in Polyneoptera insects

Terrence Sylvester[1], Carl E. Hjelmen[1], Shawn J. Hanrahan[2], Paul A. Lenhart[2], J. Spencer Johnston[2] and Heath Blackmon[1]

[1]Department of Biology, and [2]Department of Entomology, Texas A&M University, College Station,

column 2: numeric with range 0 - 0.55	desc1	pol1	p
0.158975292	0.13567824	0.0012022928	-267.9407
0.141901449	0.17588564	0.0011763734	-269.2247
0.138279931	0.13940276	0.0021394110	-268.2814
0.123512205	0.11179867	0.0028047752	-268.0275

10,000 rows



Frequentist 95% CI
0.149-0.150

95% HPD (credible interval)
0.06-0.26

Some Experimental Design Considerations

Why do I need a control?

To interpret an experiment, we need to compare the experimental subjects to the correct reference group.

What about observational studies?

What is an appropriate control?

Ideal controls are identical to the experimental population, except for the one parameter being manipulated

The control population should be similar in all other respects to the experimental population

The control population should experience sham manipulations that simulate any manipulations applied to the experimental population

Sometimes you might need multiple different controls.

Avoiding Experimenter Bias

Experimenter bias is real

The results of your study can be influenced by your expectations

Randomization is key

1) HHHHHTHHHHHHHTTTTTHTT
2) THTHTHHTHTTHHTHTHHTTH
3) HTHTHTHTHTTHTHHTTHHTH

Avoiding Experimenter Bias

Experimenter bias is real

The results of your study can be influenced by your expectations

Randomization is key

Randomize assignment of subjects to controls and treatments (**use R or random.org**).

Humans are bad at recognizing and creating randomness.

Avoiding Experimenter Bias

Use a blind or double-blind experimental design

Blind: the subject doesn't know whether it's an experimental or control subject

Double-blind: neither the researcher nor subject know which subjects are experimental versus control

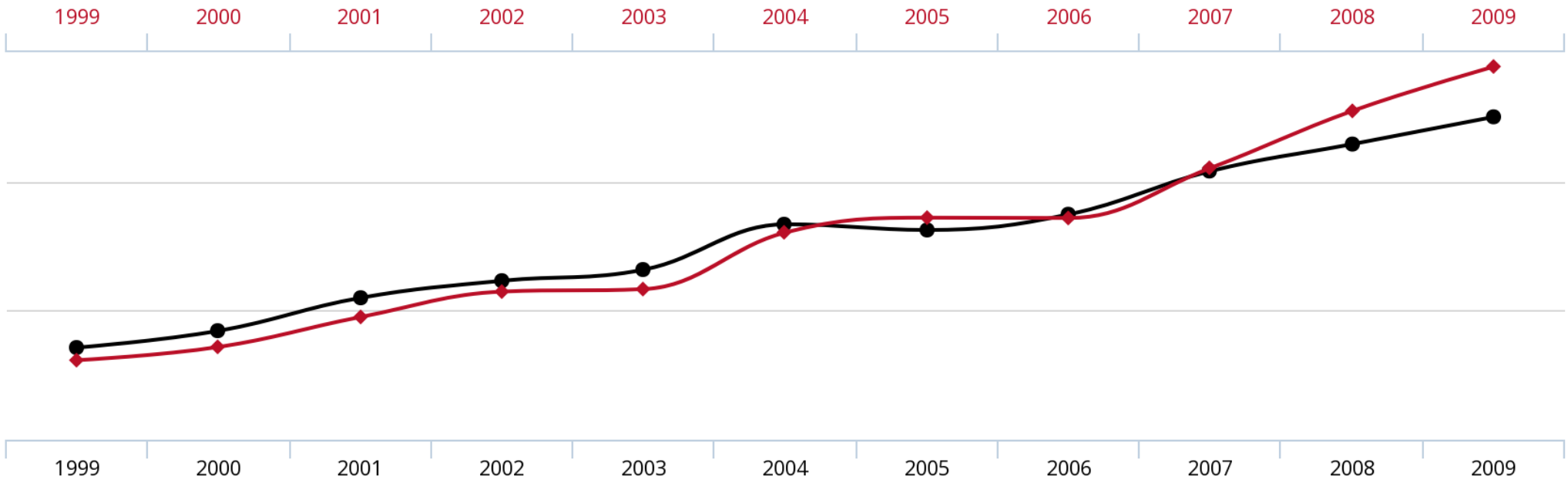
What is a more common connotation of blinding?

How can you apply this to your research?

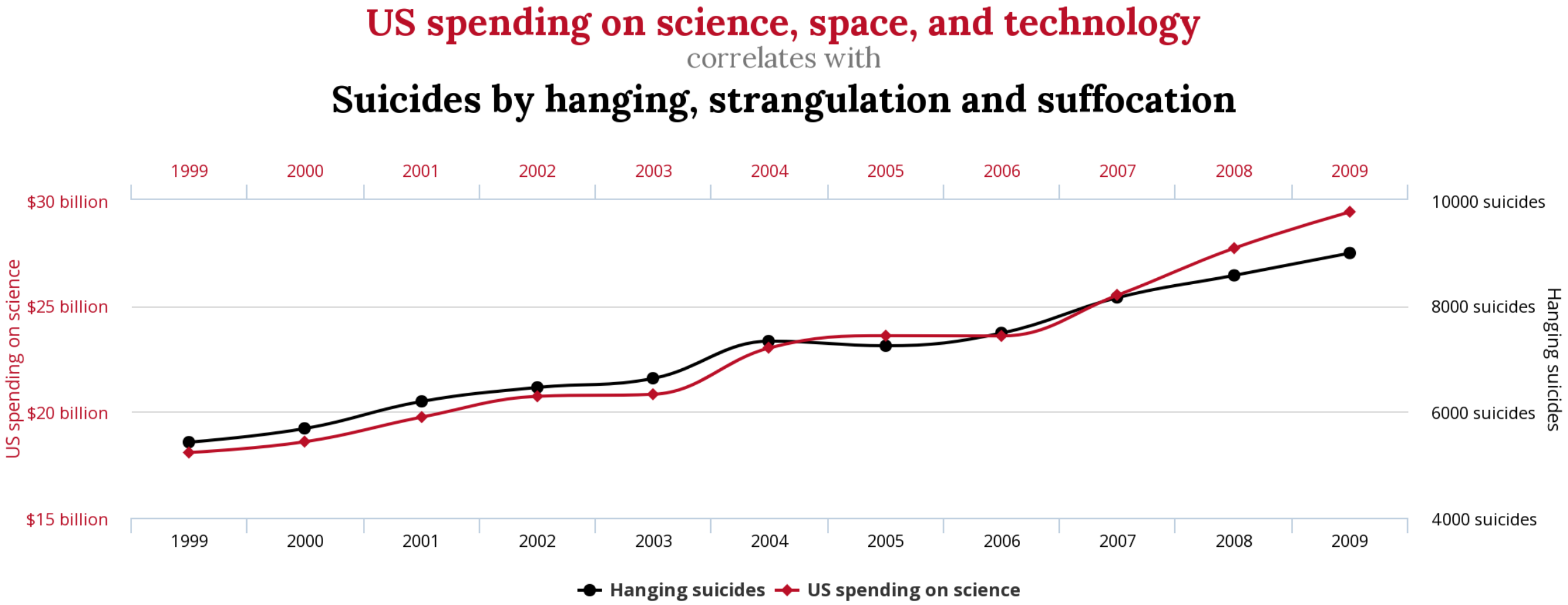
Confounding Variables

1. A difference between groups that the experimenter fails to account for
2. A hidden variable that creates an apparent causal relationship that isn't real
3. **An experiment with confounded variables can be impossible to interpret and impossible to fix**

Confounding Variables



Confounding Variables



Confounding Example

Study type	Confounding variable
Gene expression level	Tissue used
Diversification	unobserved traits
Lung cancer and coffee	coffee smoking correlation
Behavior	maternal effects
Effective population size	breeding system

Redesign the procedure

- Collect 750 beetles from a population cage.
- Create 30 new vials with 25 beetles each.
- Make the first 15 of these control vials and use food media A.
- Make the next 15 of these experiment vials and use food media B.
- Place in a rack as shown and place in the incubator.
- Measure growth at day 15.