

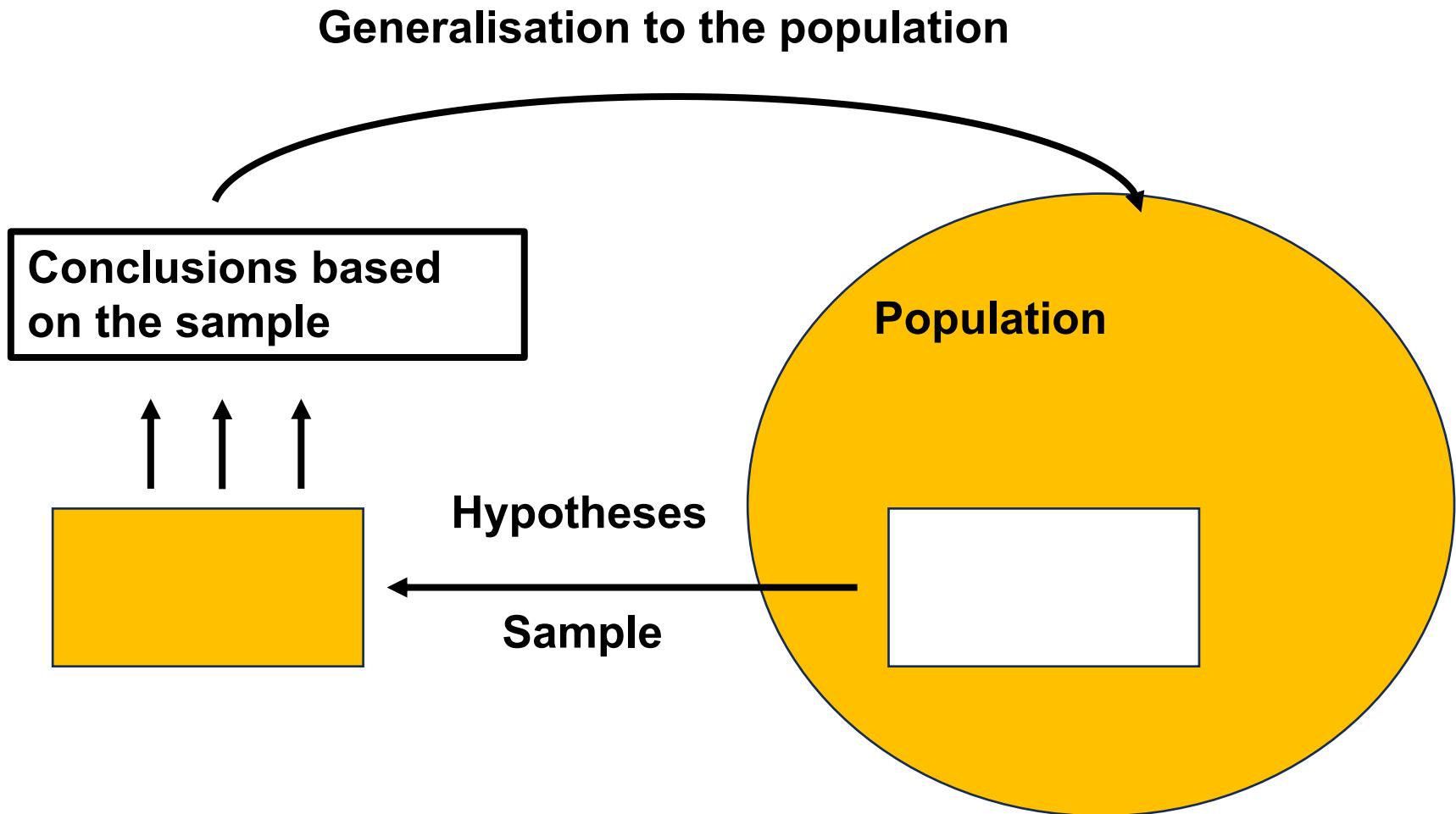
Repeatability in Evolutionary Ecology

Nicolas Rode

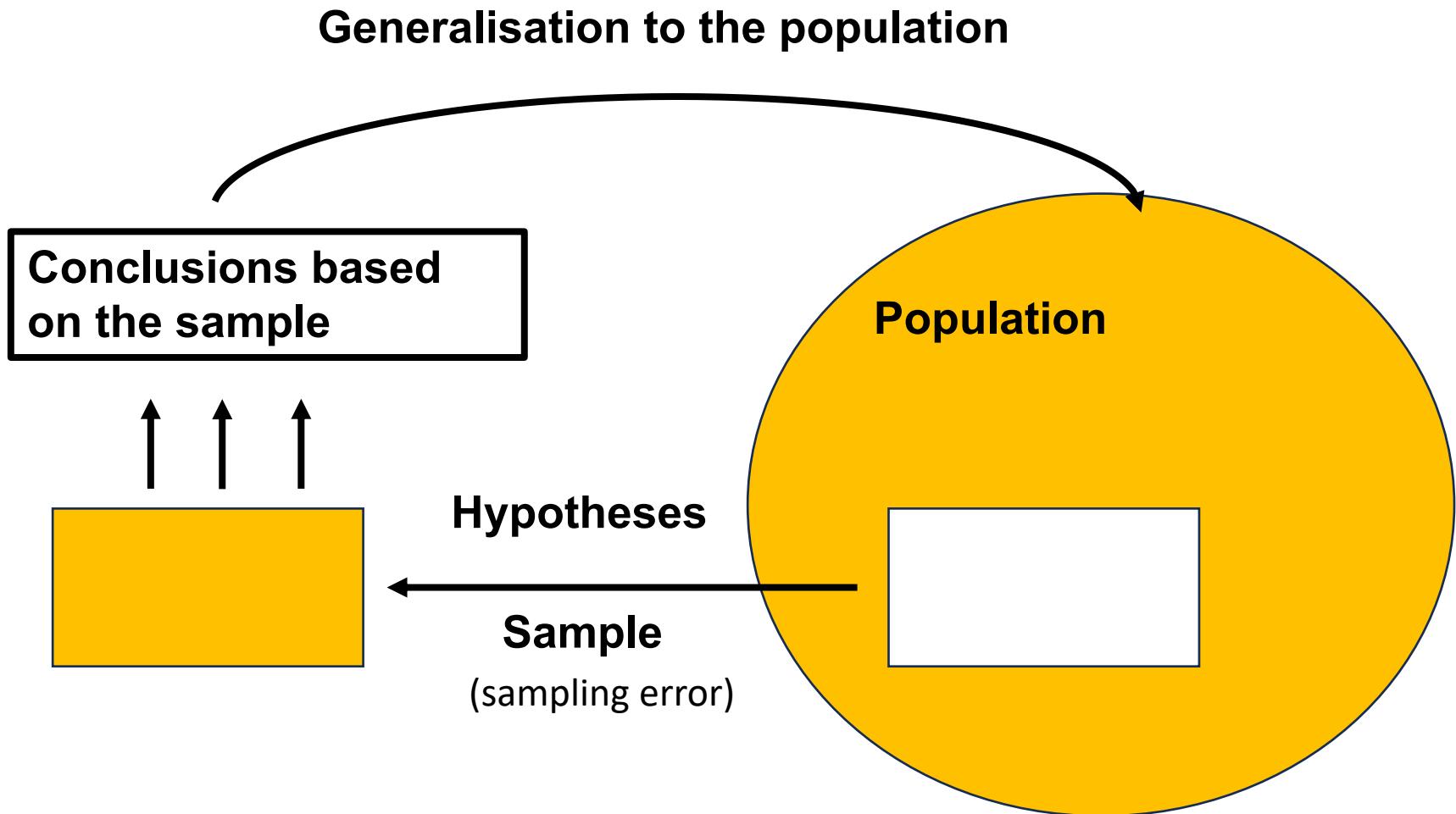


JC CBGP, 13 Feb. 2024

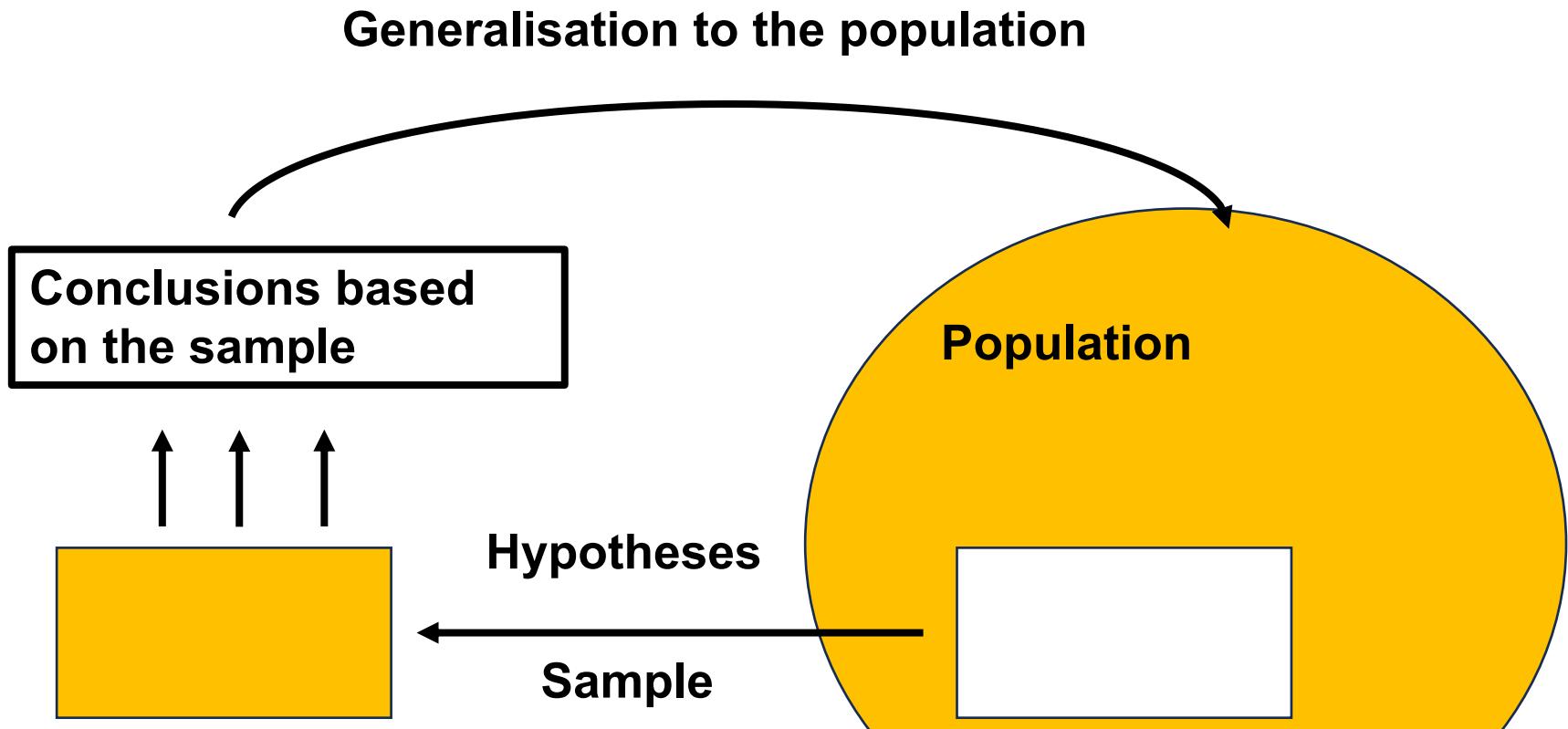
Statistical inference



Statistical inference



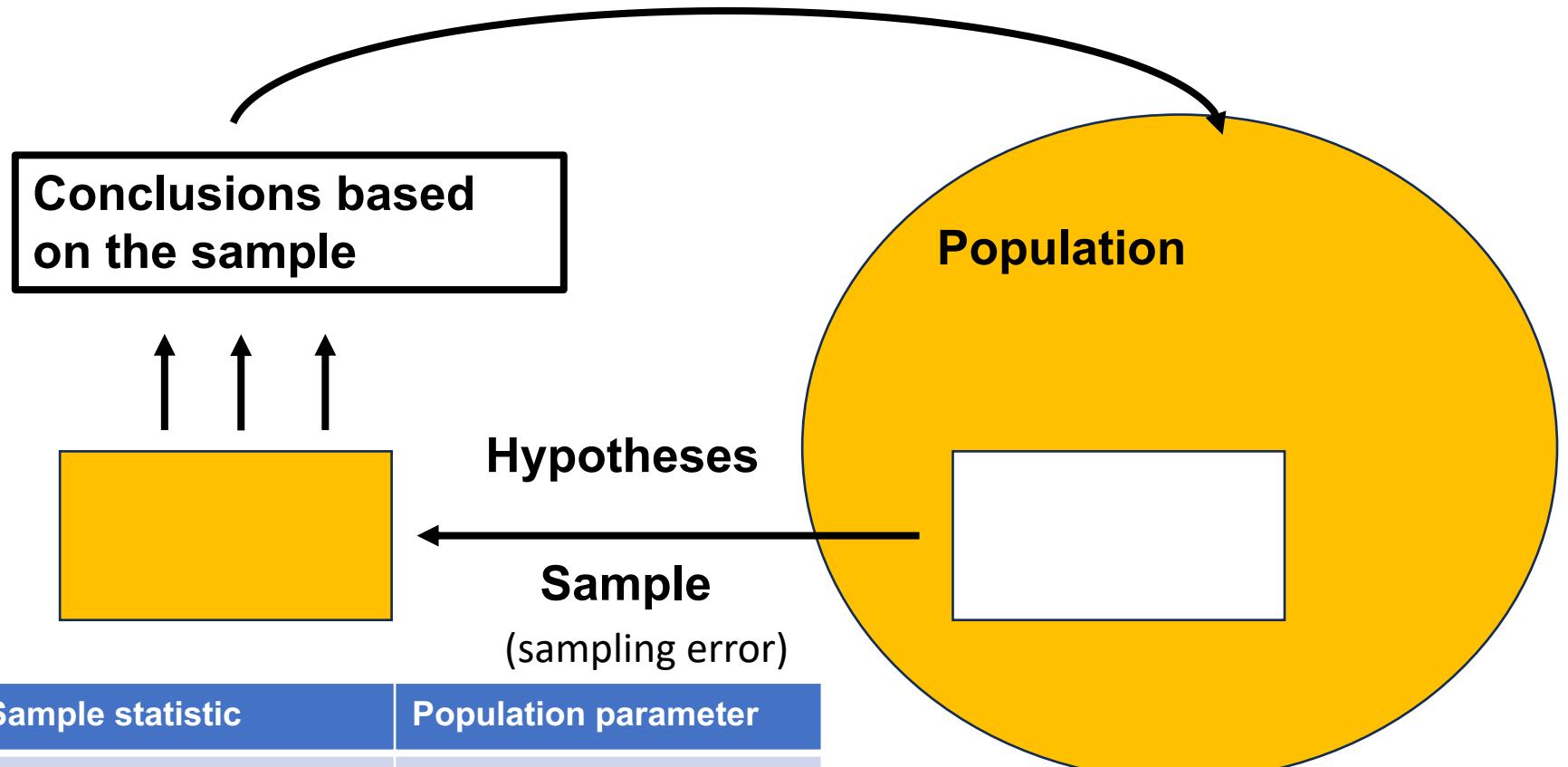
Statistical inference



Sample statistic	Population parameter
Mean - \bar{x}	Mean - μ
Variance - s^2	Variance - σ^2
Standard deviation - s	Standard deviation - σ

Statistical inference

Generalisation to the population



Sample statistic	Population parameter
Mean - \bar{x}	Mean - μ
Variance - s^2	Variance - σ^2
Standard deviation - s	Standard deviation - σ

Point estimate = $\hat{\mu}$
Interval estimate = $\hat{\mu} \pm 1.96 \times SE$

Bias and precision of the point estimate of the population mean



ChatGPT

Title: "Understanding Bias and Precision: Crucial Concepts in Statistical Estimation for Evolutionary Biologists"

Good morning, fellow researchers,

Today, I want to delve into a fundamental aspect of statistical inference that is often overlooked but crucial for any rigorous scientific analysis: bias and precision in statistical estimates. As graduate students in evolutionary biology, you're constantly making inferences and drawing conclusions from data, and understanding the biases and precision of your estimates is paramount to ensuring the validity and reliability of your findings.

Bias and precision of the point estimate of the population mean



ChatGPT

Title: "Understanding Bias and Precision: Crucial Concepts in Statistical Estimation for Evolutionary Biologists"

Good morning, fellow researchers,

Today, I want to delve into a fundamental aspect of statistical inference that is often overlooked but crucial for any rigorous scientific analysis: bias and precision in statistical estimates. As graduate students in evolutionary biology, you're constantly making inferences and drawing conclusions from data, and understanding the biases and precision of your estimates is paramount to ensuring the validity and reliability of your findings.

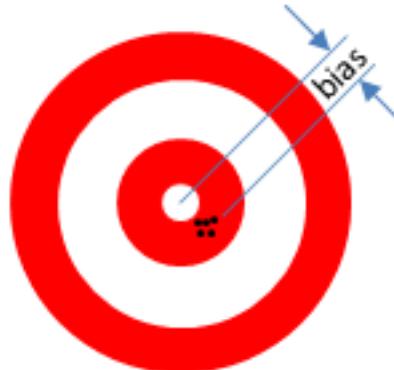
Bias and precision of the point estimate of the population mean

Bias: systematic error in an estimate that causes it to deviate from the true value
=> Over/underestimation of the true parameter of interest ($\hat{\mu}$)
=> Origin: sampling method, measurement error, or model misspecification

Bias and precision of the point estimate of the population mean



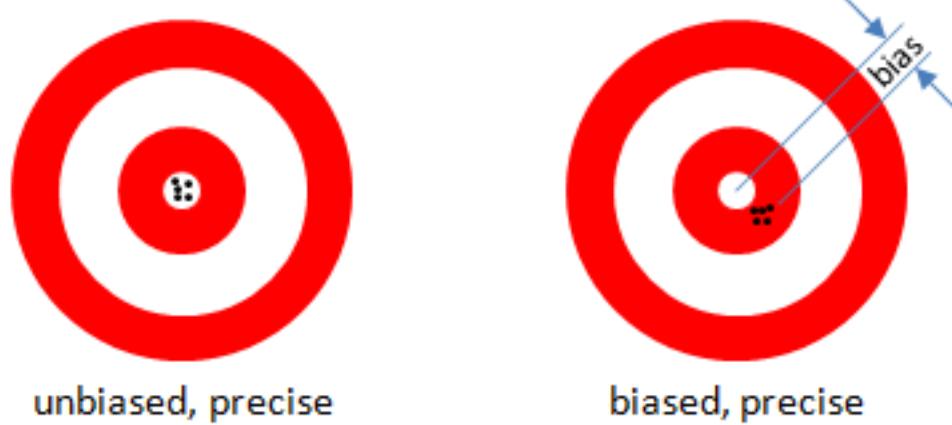
unbiased, precise



biased, precise

Bias: systematic error in an estimate that causes it to deviate from the true value
=> Over/underestimation of the true parameter of interest ($\hat{\mu}$)
=> Origin: sampling method, measurement error, or model misspecification

Bias and precision of the point estimate of the population mean

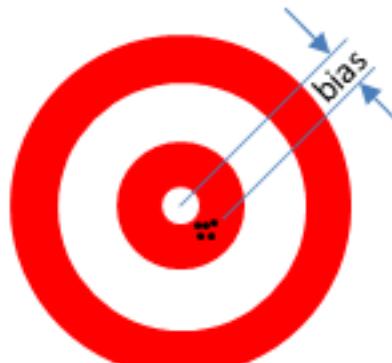


Precision: variability/uncertainty in your point estimate
=> reflects how close repeated point estimates would be to each other if you were to repeat the estimation procedure multiple times under identical conditions

Bias and precision of the point estimate of the population mean



unbiased, precise



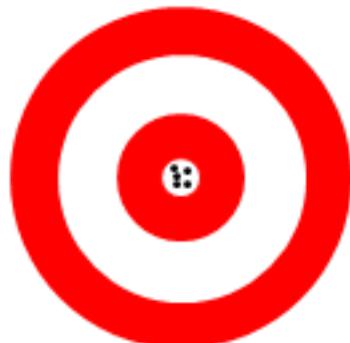
biased, precise



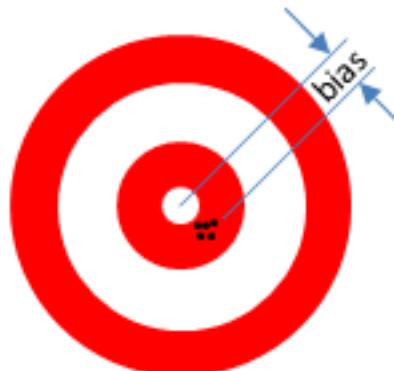
unbiased, imprecise

Precision: variability/uncertainty in your point estimate
=> reflects how close repeated point estimates would be to each other if you were to repeat the estimation procedure multiple times under identical conditions

Bias and precision of the point estimate of the population mean



unbiased, precise



biased, precise



unbiased, imprecise

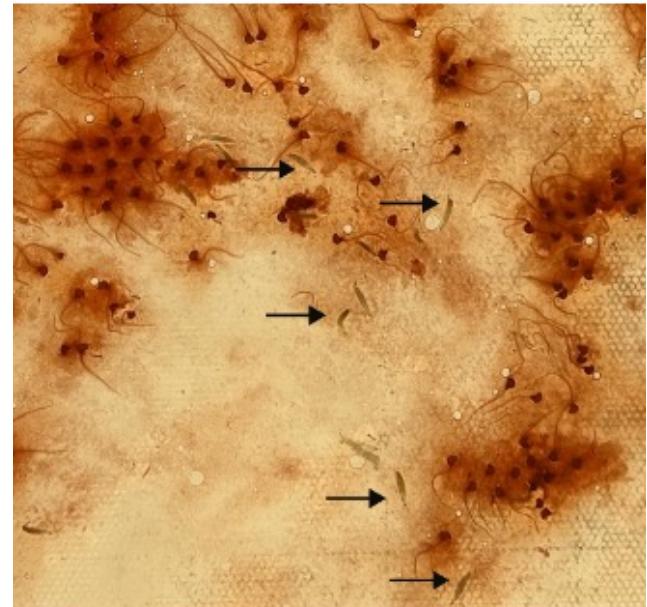


biased, imprecise



- Accurate parameter estimation
- Reliable hypothesis testing
- Interpretation of results

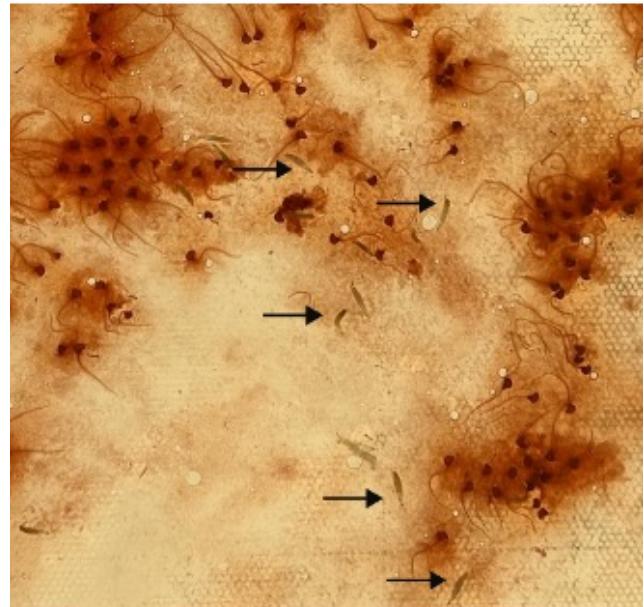
Real example



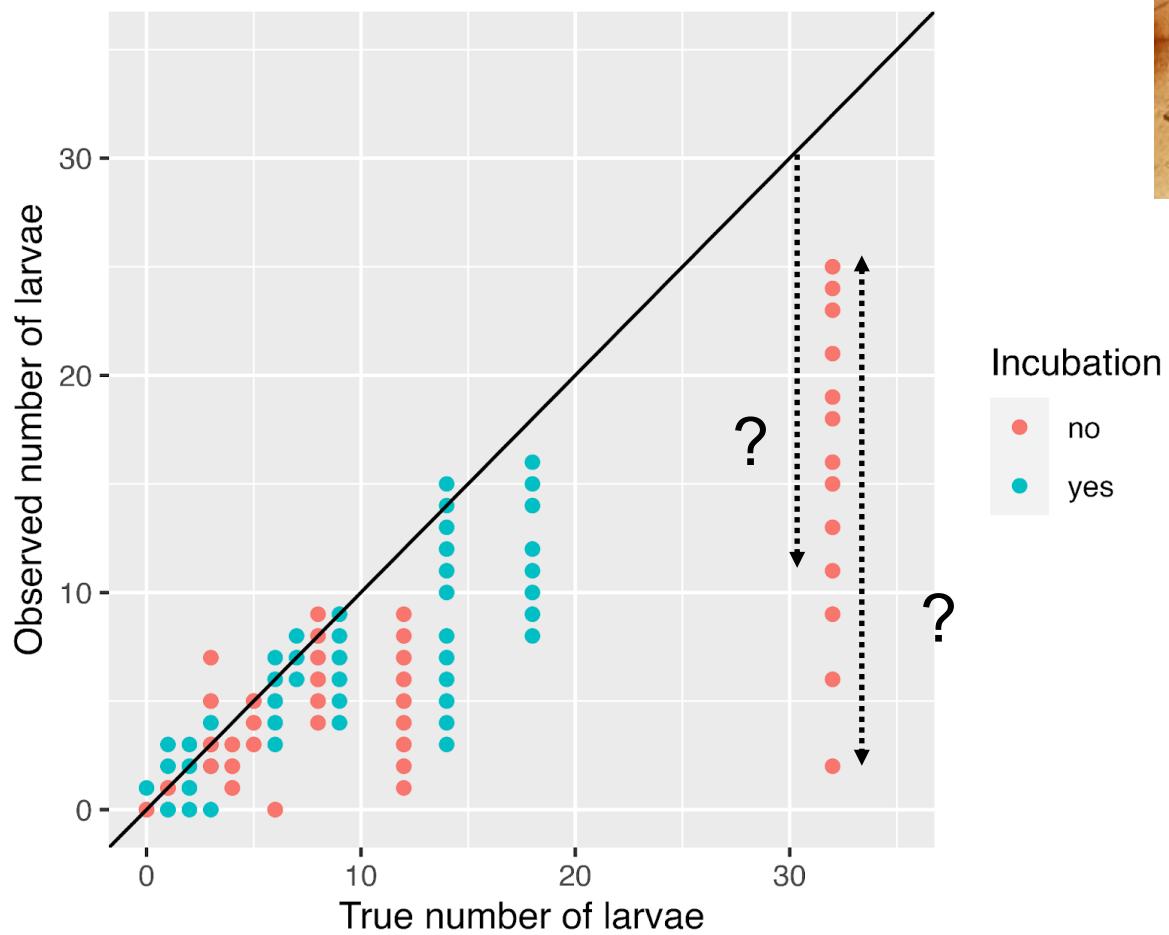
Zriki et al 2023 *J. Econ. Entom.*

n=30 strawberries
k= 20 replicates

Real example

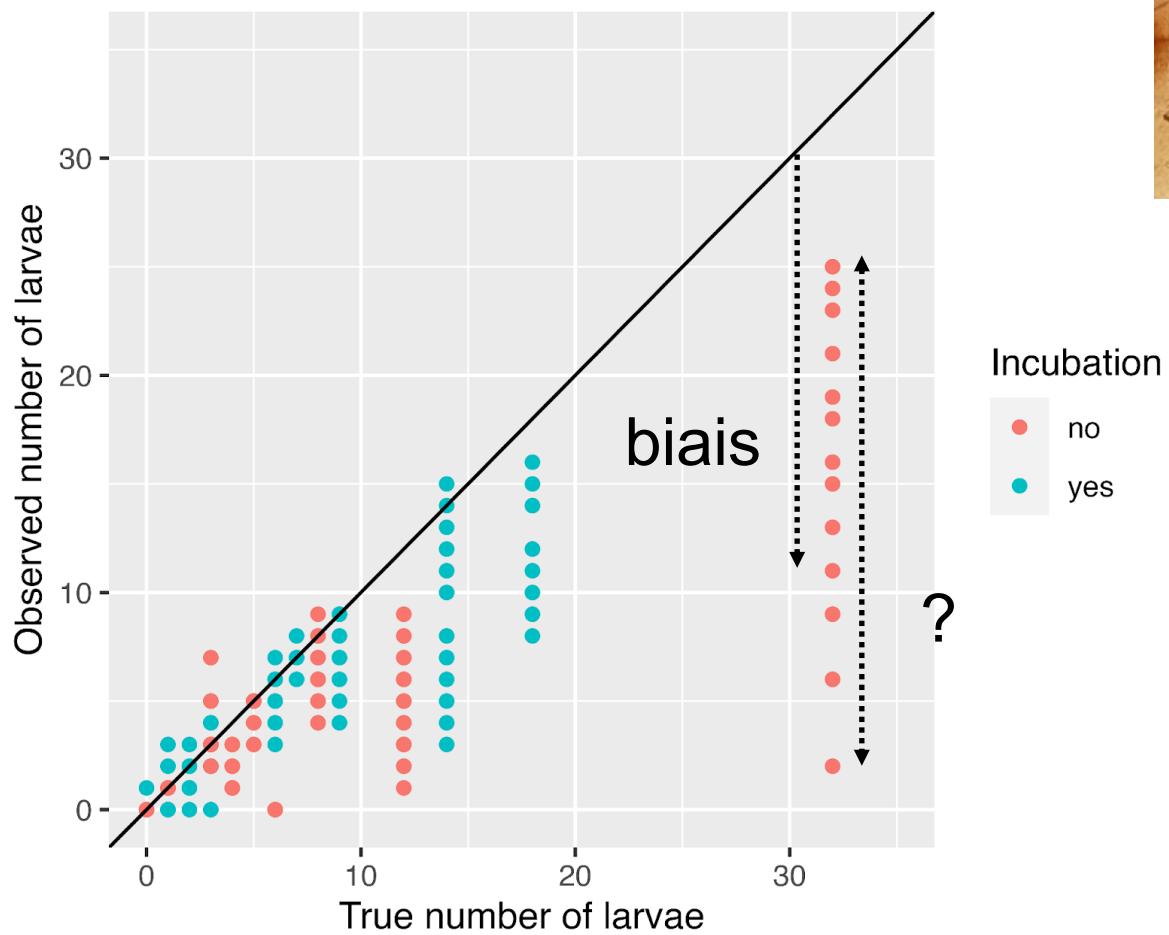
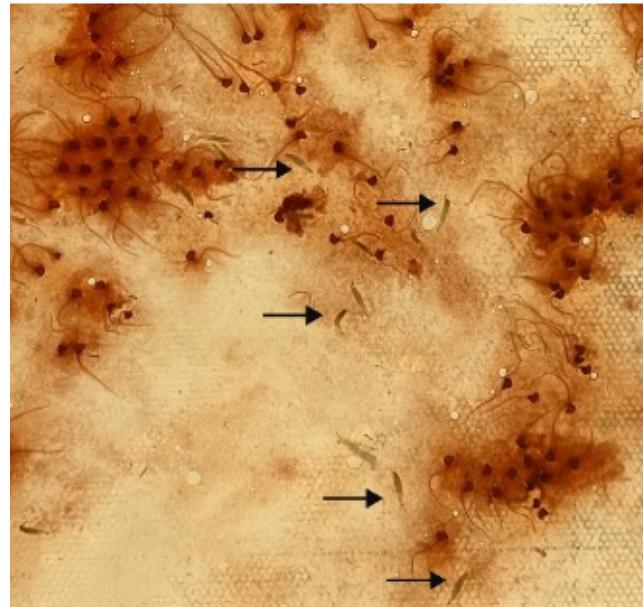


Zriki et al 2023 *J. Econ. Entom.*



$n=30$ strawberries
 $k=20$ replicates

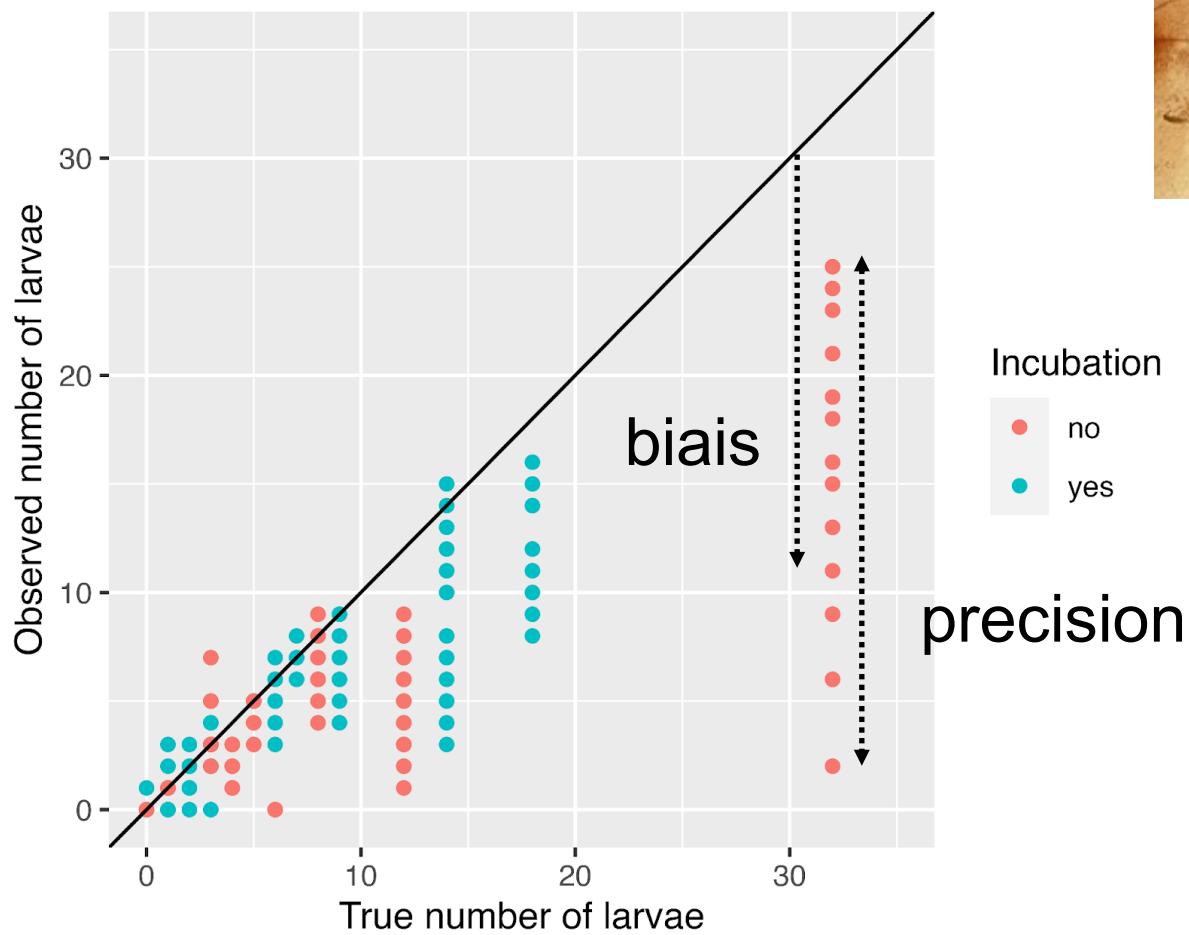
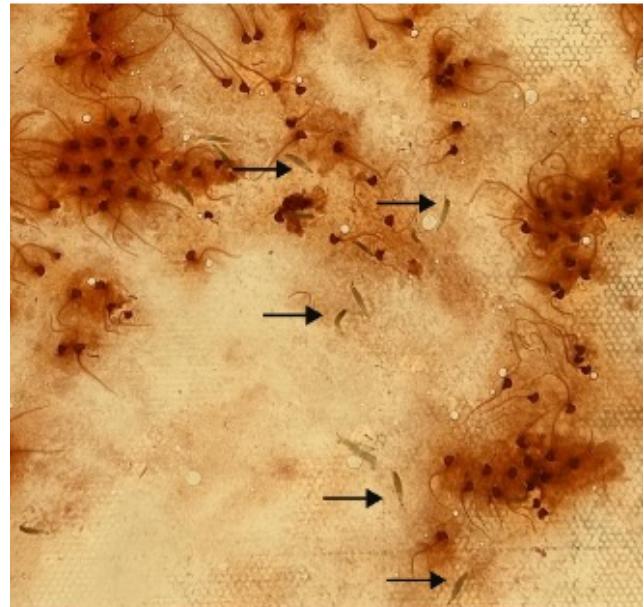
Real example



Zriki et al 2023 *J. Econ. Entom.*

$n=30$ strawberries
 $k= 20$ replicates

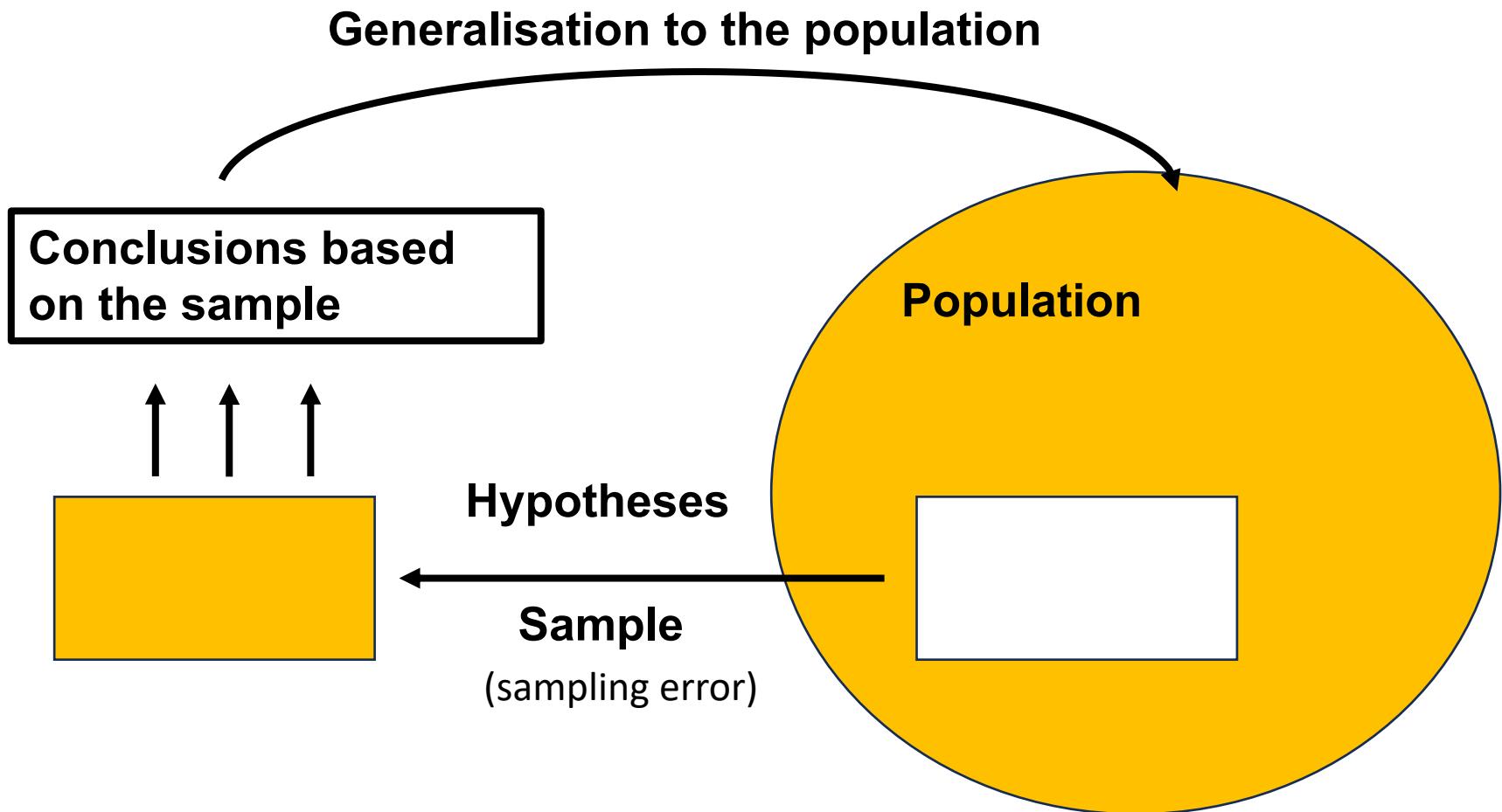
Real example



Zriki et al 2023 *J. Econ. Entom.*

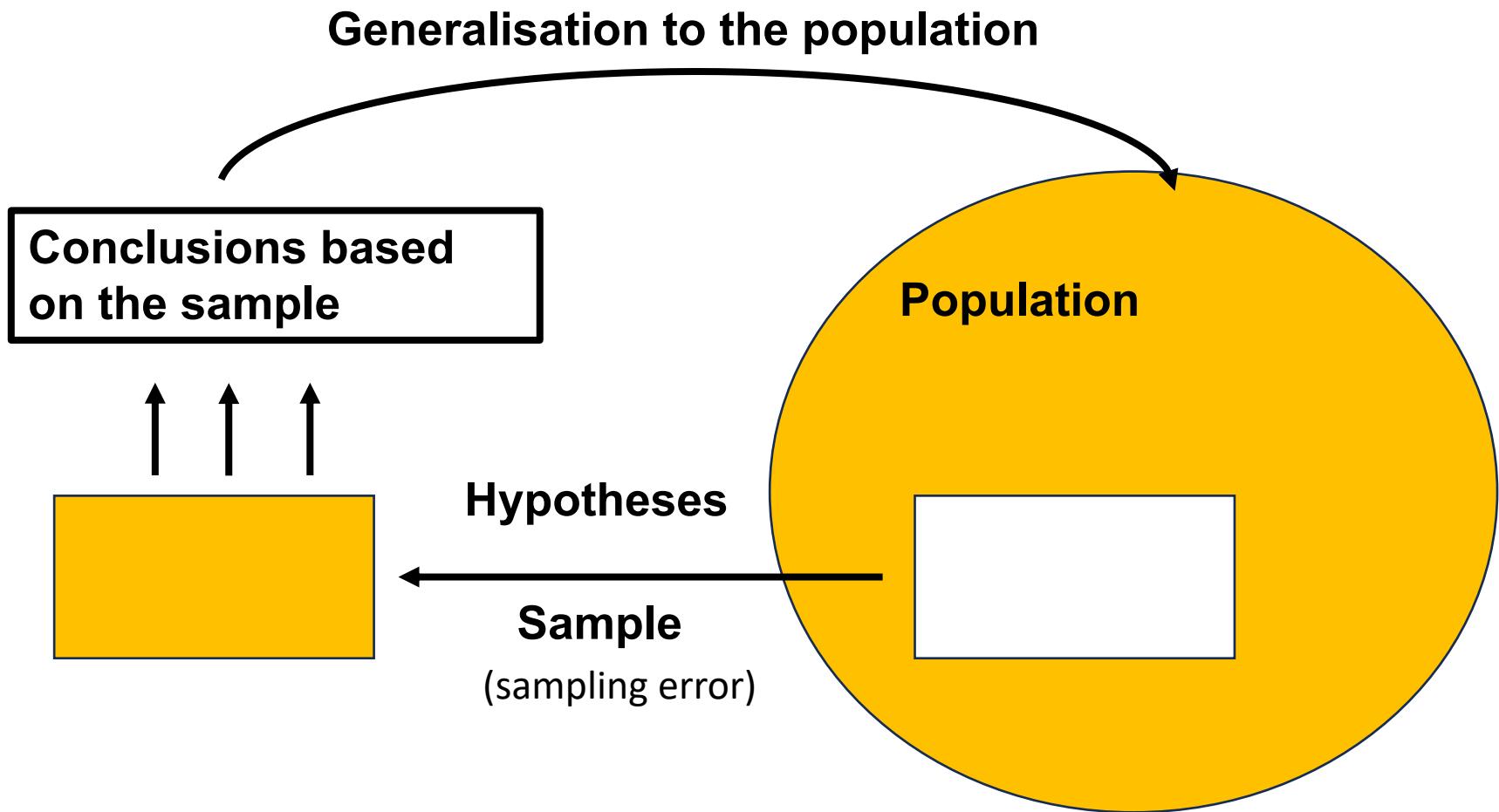
n=30 strawberries
k= 20 replicates

Repeatability of the measurement of a trait



How can I ensure that my results are reproducible?

Repeatability of the measurement of a trait



How can I ensure that my results are reproducible?
How can I ensure that my measurement are repeatable?

Repeatability of the measurement of a trait

UNREPEATABLE REPEATABILITIES:
A COMMON MISTAKE

C. M. LESSELLS¹ AND PETER T. BOAG²

¹Department of Zoology, University of Sheffield, Sheffield S10 2TN, England, and

²Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

ABSTRACT.—Repeatability is a useful tool for the population geneticist or genetical ecologist, but several papers have carried errors in its calculation. We outline the correct calculation of repeatability, point out the common mistake, show how the incorrectly calculated value relates to repeatability, and provide a method for checking published values and calculating approximate repeatability values from the *F* ratio (mean squares among groups/mean squares within groups). Received 6 February 1986, accepted 25 August 1986.



C. M. Lessells

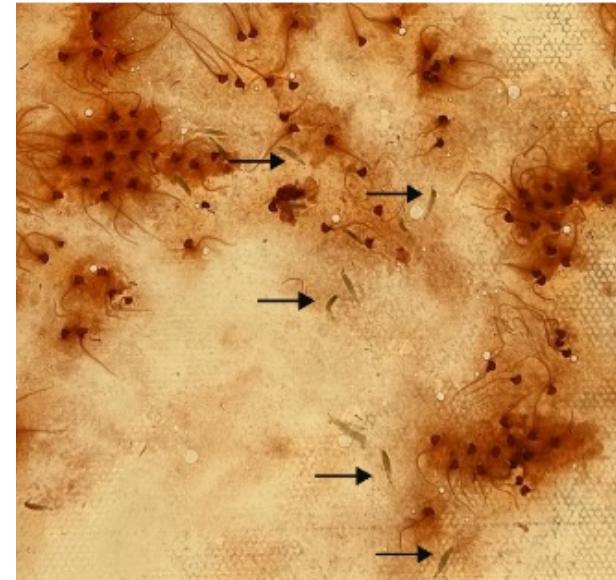
P. T. Boag

Repeatability: proportion of the total variation that is reproducible among repeated measurements of the same subject or group
=> intra-class correlation coefficient (ICC) of a one-way ANOVA

One-way ANOVA

n_group: 30 strawberries

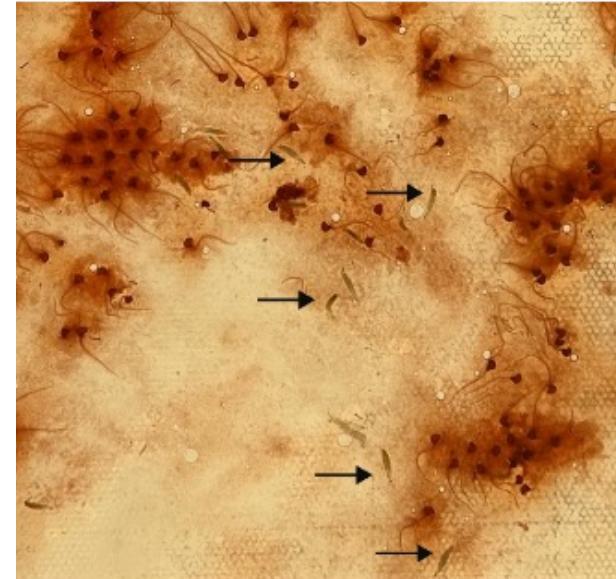
n0: 2 replicates / strawberry



Zriki et al 2023 *J. Econ. Entom.*

One-way ANOVA

n_group: 30 strawberries
n0: 2 replicates / strawberry



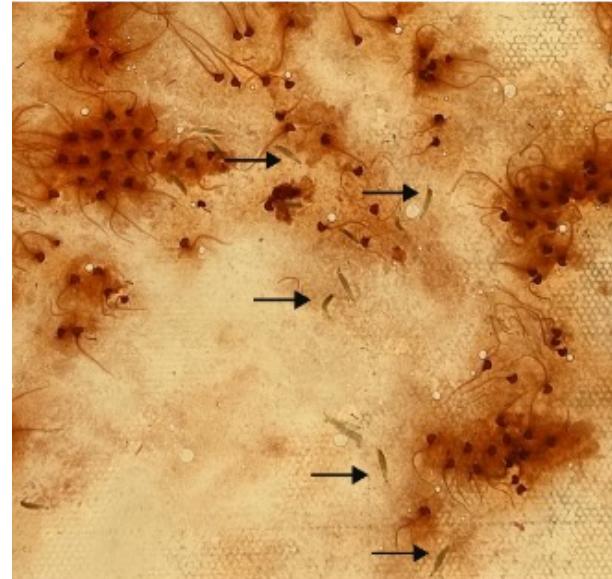
	Model	Number of parameters (mean +variance)	Residual degree of freedom, df_{res} (number of ε_{ij})
Reduced model m0	$Y_{ij} = \mu + \varepsilon_{ij}$	2 (1+1)	59
Complete model m1	$Y_{ij} = \mu + \delta_i + \varepsilon_{ij}$	31 (1+29+1)	30

With $i=1, \dots, n_{group}$
with $j=1, \dots, n_0$

Zriki et al 2023 *J. Econ. Entom.*

One-way ANOVA

n_group: 30 strawberries
 n0: 2 replicates / strawberry



	Model	Number of parameters (mean +variance)	Residual degree of freedom, df_{res} (number of ε_{ij})
Reduced model m0	$Y_{ij} = \mu + \varepsilon_{ij}$	2 (1+1)	59
Complete model m1	$Y_{ij} = \mu + \delta_i + \varepsilon_{ij}$	31 (1+29+1)	30

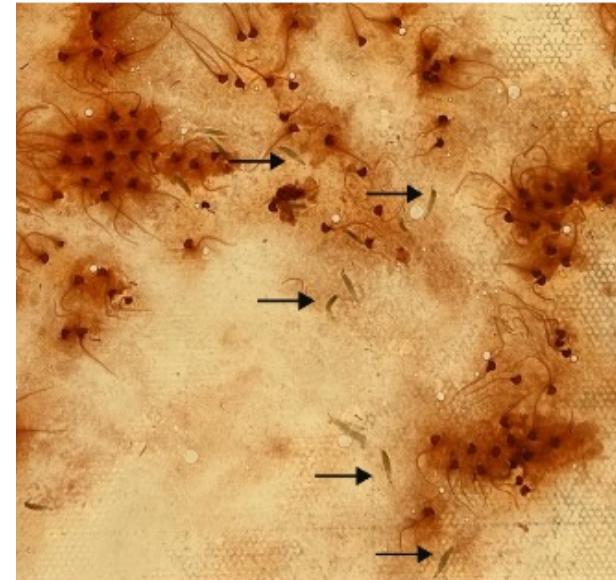
With $i=1, \dots, n_group$
 with $j=1, \dots, n_0$

Zriki et al 2023 *J. Econ. Entom.*

Source of variation	Model	Mean Square	Expected mean square	Observed mean square
Among groups	$df_1 = n_group - 1$ $= df_{res}(m0) - df_{res}(m1)$	MS_A	$k_{rep} \times \sigma_A^2 + \sigma_W^2$	$n_0 \times S_A^2 + S_W^2$
Within groups	$df_2 = n_group \times (n_0 - 1)$ $= df_{res}(m1)$	MS_W	σ_W^2	S_W^2

One-way ANOVA

n_group: 30 strawberries
 n0: 2 replicates / strawberry



	Model	Number of parameters (mean +variance)	Residual degree of freedom, df_{res} (number of ε_{ij})
Reduced model m0	$Y_{ij} = \mu + \varepsilon_{ij}$	2 (1+1)	59
Complete model m1	$Y_{ij} = \mu + \delta_i + \varepsilon_{ij}$	31 (1+29+1)	30

With $i=1, \dots, n_group$
 with $j=1, \dots, n_0$

Zriki et al 2023 *J. Econ. Entom.*

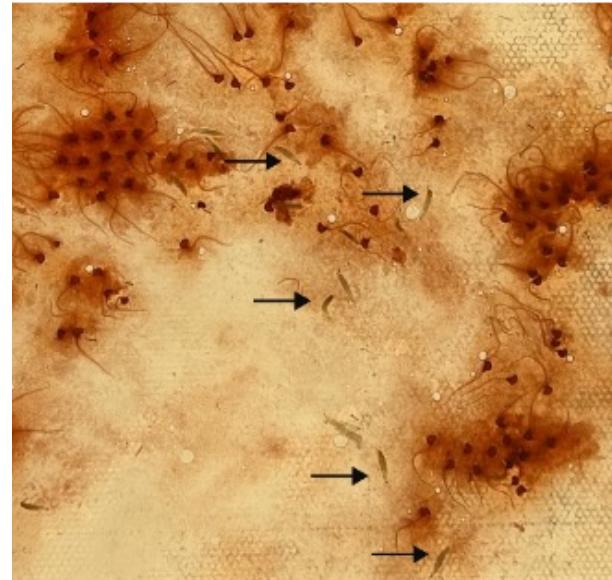
Source of variation	Model	Mean Square	Expected mean square	Observed mean square
Among groups	$df_1 = n_group - 1$ $= df_{res}(m0) - df_{res}(m1)$	MS_A	$k_{rep} \times \sigma_A^2 + \sigma_W^2$	$n_0 \times S_A^2 + S_W^2$
Within groups	$df_2 = n_group \times (n_0 - 1)$ $= df_{res}(m1)$	MS_W	σ_W^2	S_W^2

$$Repeatability(ICC) = \frac{S_A^2}{S_A^2 + S_W^2}$$

↑
 Among individuals Within individuals

One-way ANOVA

n_group: 30 strawberries
 n0: 2 replicates / strawberry



	Model	Number of parameters (mean +variance)	Residual degree of freedom, df_{res} (number of ε_{ij})
Reduced model m0	$Y_{ij} = \mu + \varepsilon_{ij}$	2 (1+1)	59
Complete model m1	$Y_{ij} = \mu + \delta_i + \varepsilon_{ij}$	31 (1+29+1)	30

With $i=1, \dots, n_group$
 with $j=1, \dots, n_0$

Zriki et al 2023 *J. Econ. Entom.*

Source of variation	Model	Mean Square	Expected mean square	Observed mean square
Among groups	$df_1 = n_group - 1$ $= df_{res}(m0) - df_{res}(m1)$	MS_A	$k_{rep} \times \sigma_A^2 + \sigma_W^2$	$n_0 \times S_A^2 + S_W^2$
Within groups	$df_2 = n_group \times (n_0 - 1)$ $= df_{res}(m1)$	MS_W	σ_W^2	S_W^2

$$S_W^2 = MS_W$$

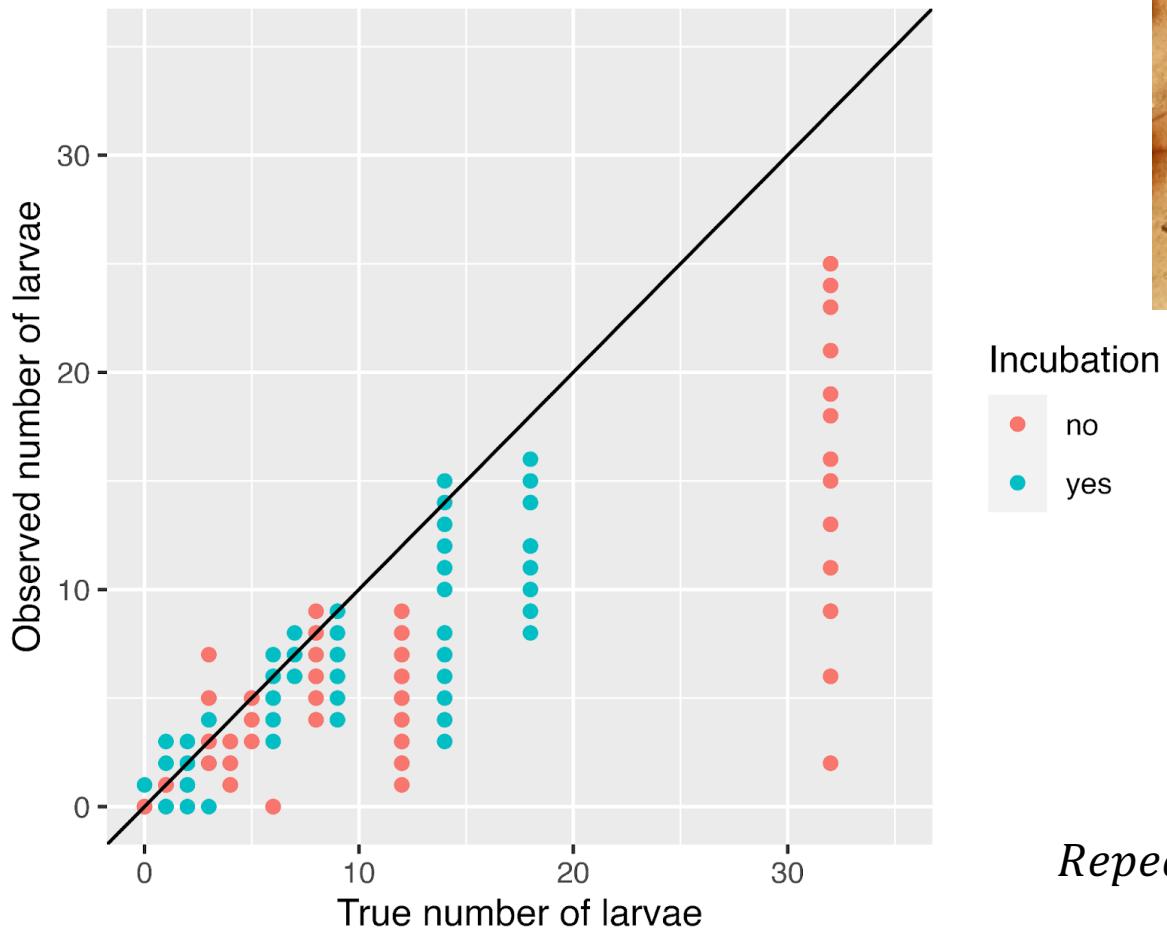
$$S_A^2 = \frac{MS_A - MS_W}{n_0}$$

$$Repeatability(ICC) = \frac{S_A^2}{S_A^2 + S_W^2}$$

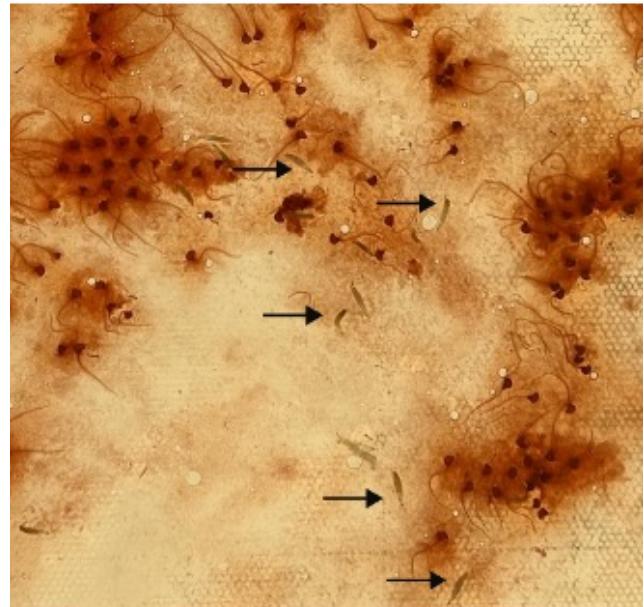
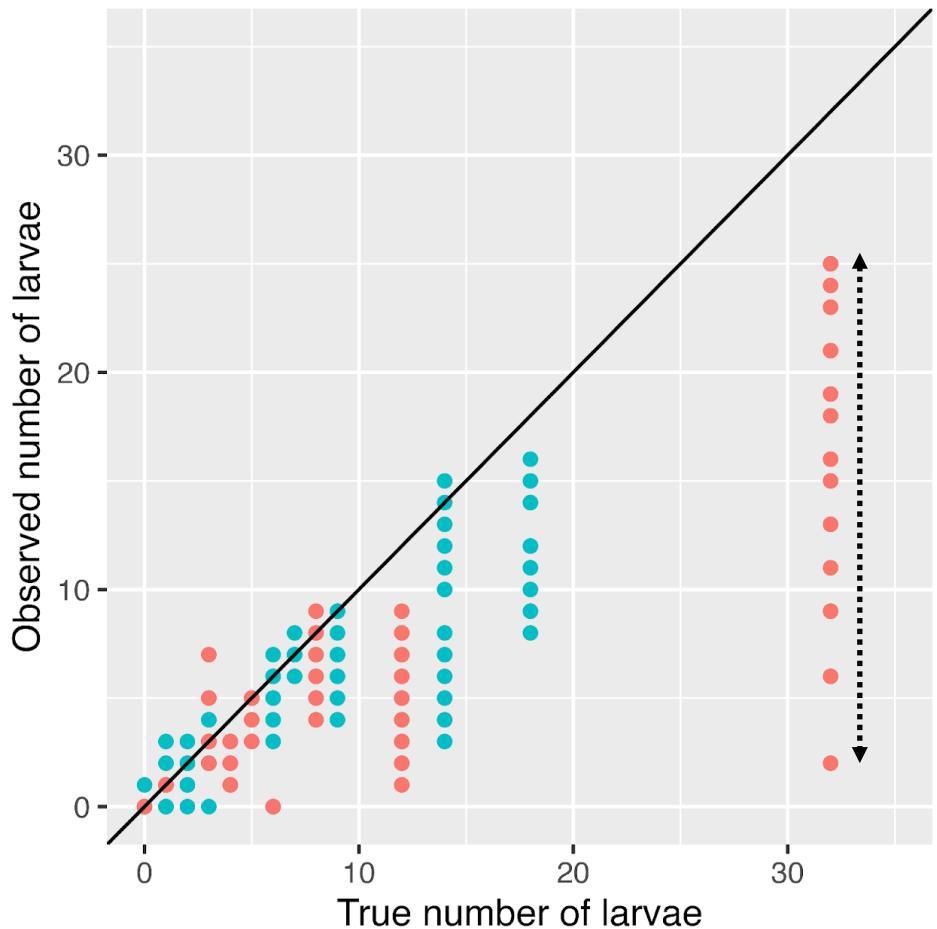
Among individuals

Within individuals

Real example



Real example



Incubation

- no
- yes

Zriki et al 2023 *J. Econ. Entom.*

?

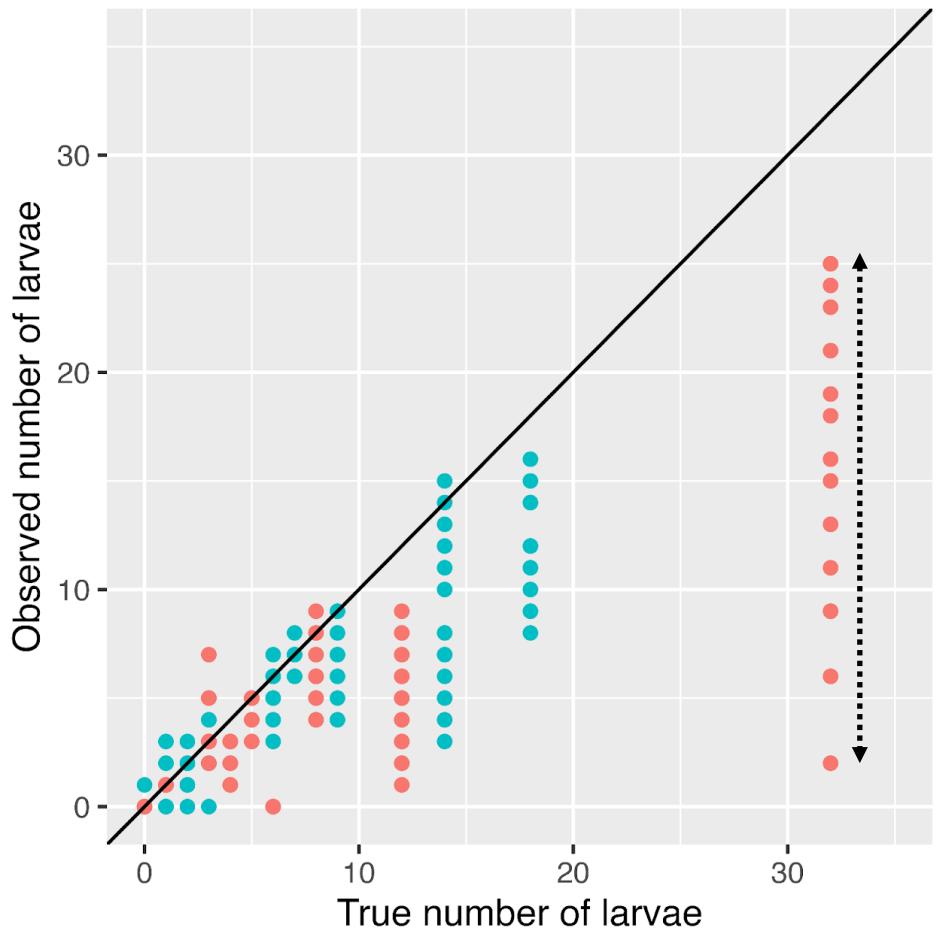
$$\text{Repeatability(ICC)} = \frac{S_A^2}{S_A^2 + S_W^2}$$

?

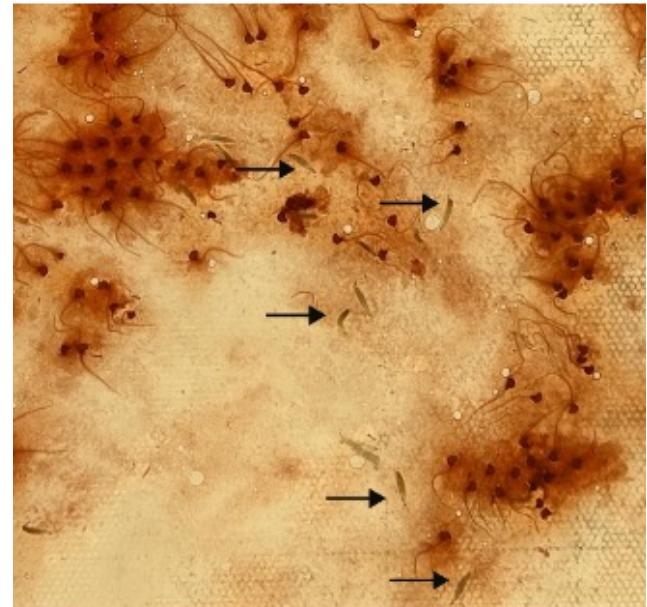
Among individuals

Within individuals

Real example



Variance among strawberries (S_A^2)



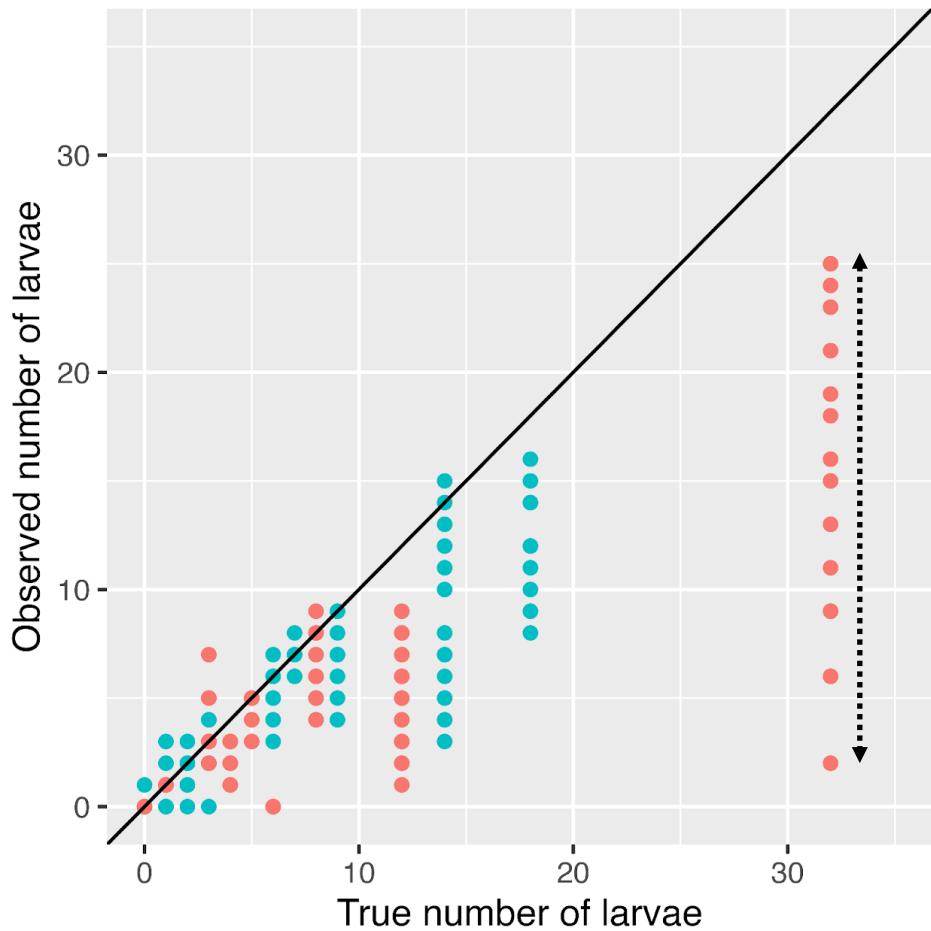
Incubation

- no
- yes

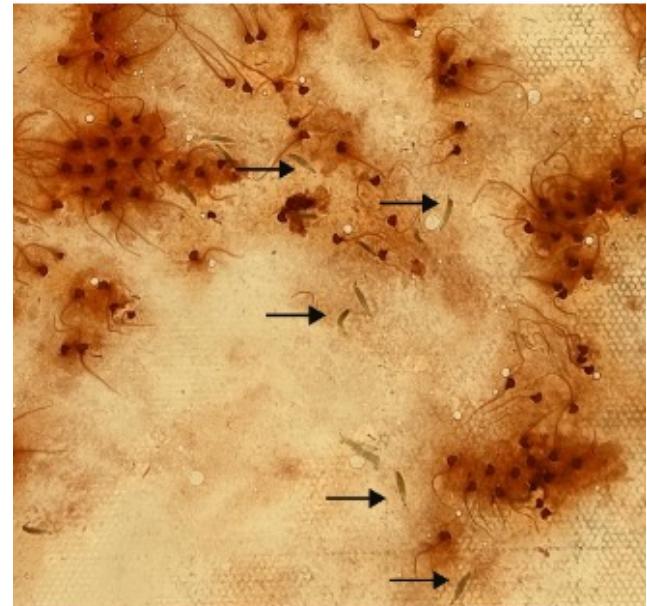
Zriki et al 2023 *J. Econ. Entom.*

Variance within each strawberry (S_W^2)

Real example



Variance among strawberries (S_A^2)



Incubation

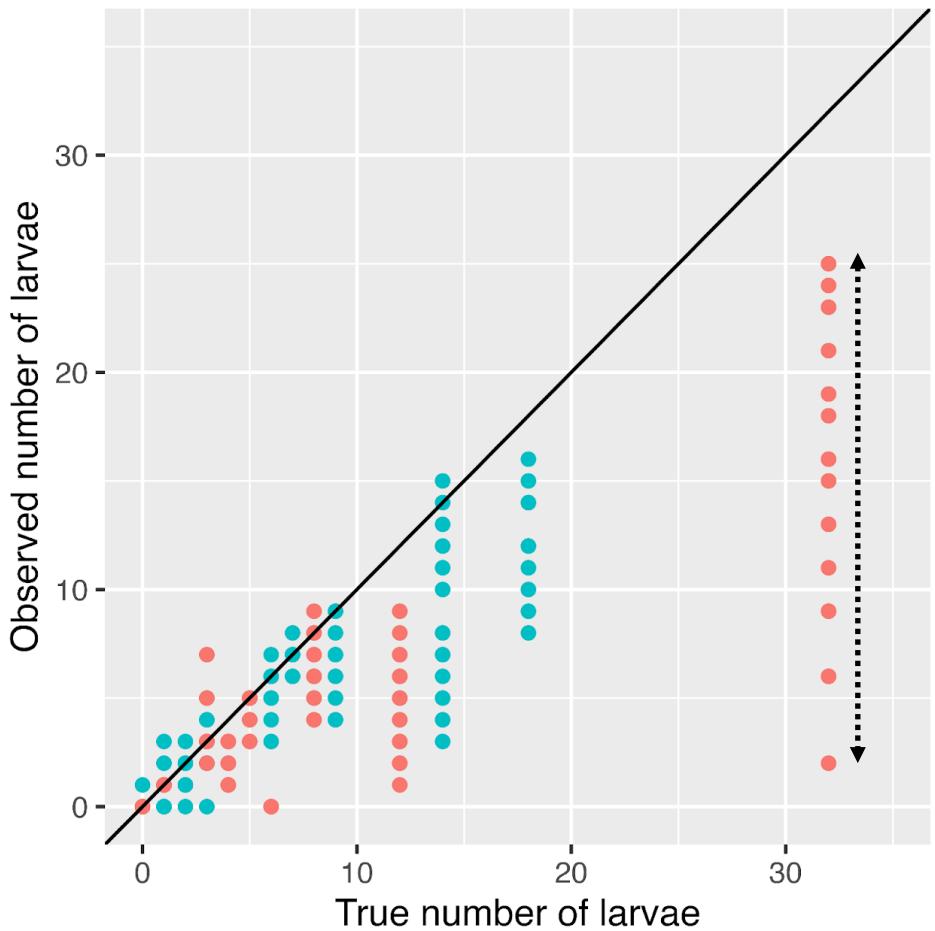
- no
- yes

Zriki et al 2023 *J. Econ. Entom.*

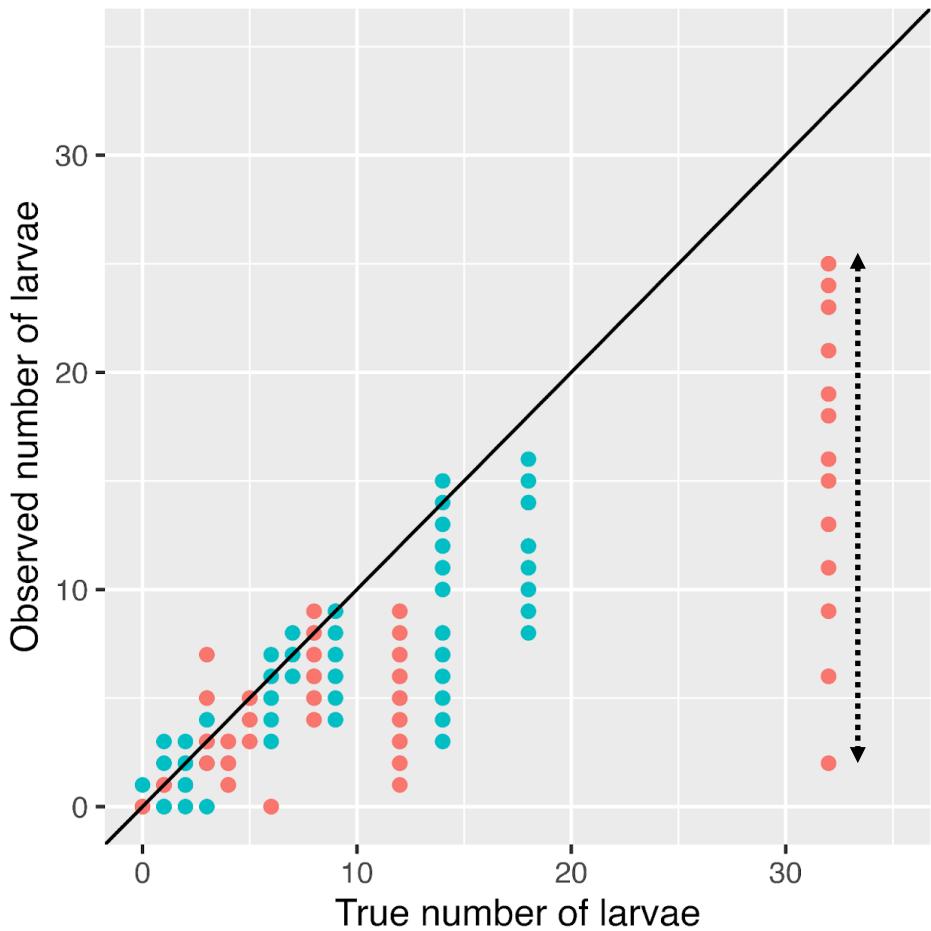
Variance within each strawberry (S_W^2)

$$\log(\text{Number of larvae} + 1)_{ij} = \mu + \delta_i + \varepsilon_{ij}$$

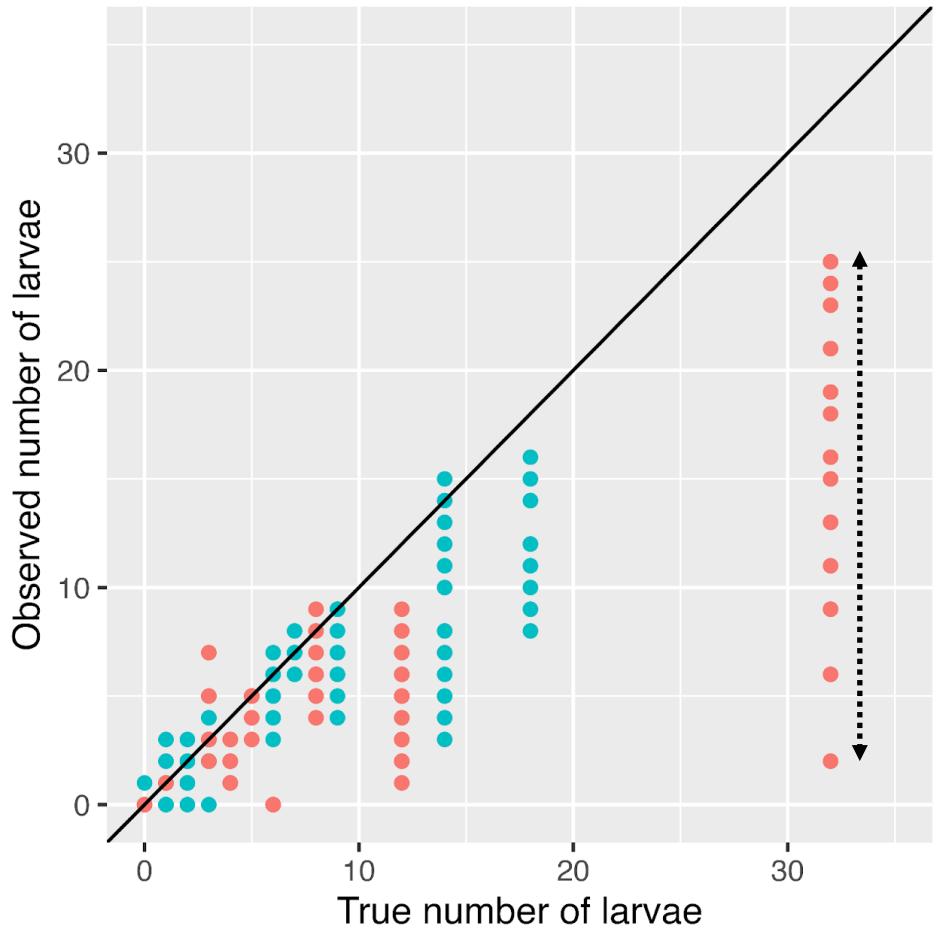
Real example



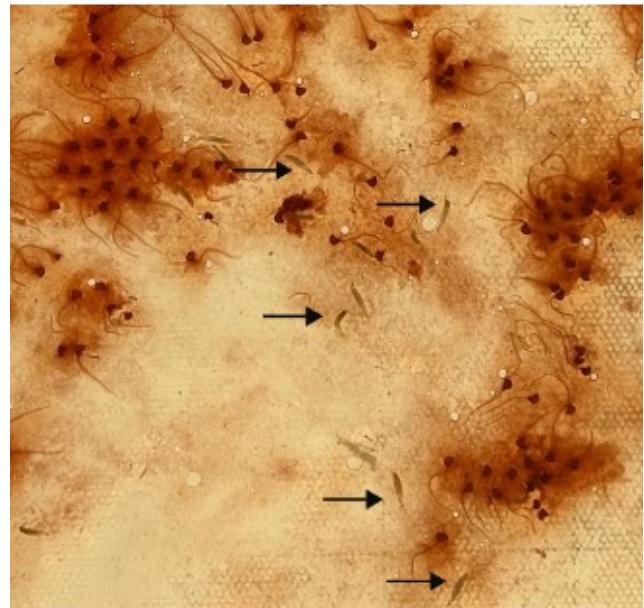
Real example



Real example



Variance among strawberries (S_A^2)



Incubation

- no
- yes

Zriki et al 2023 *J. Econ. Entom.*

Variance within each strawberry (S_W^2)

$$\log(\text{Number of larvae} + 1)_{ij} = \mu + \delta_i + \varepsilon_{ij}$$

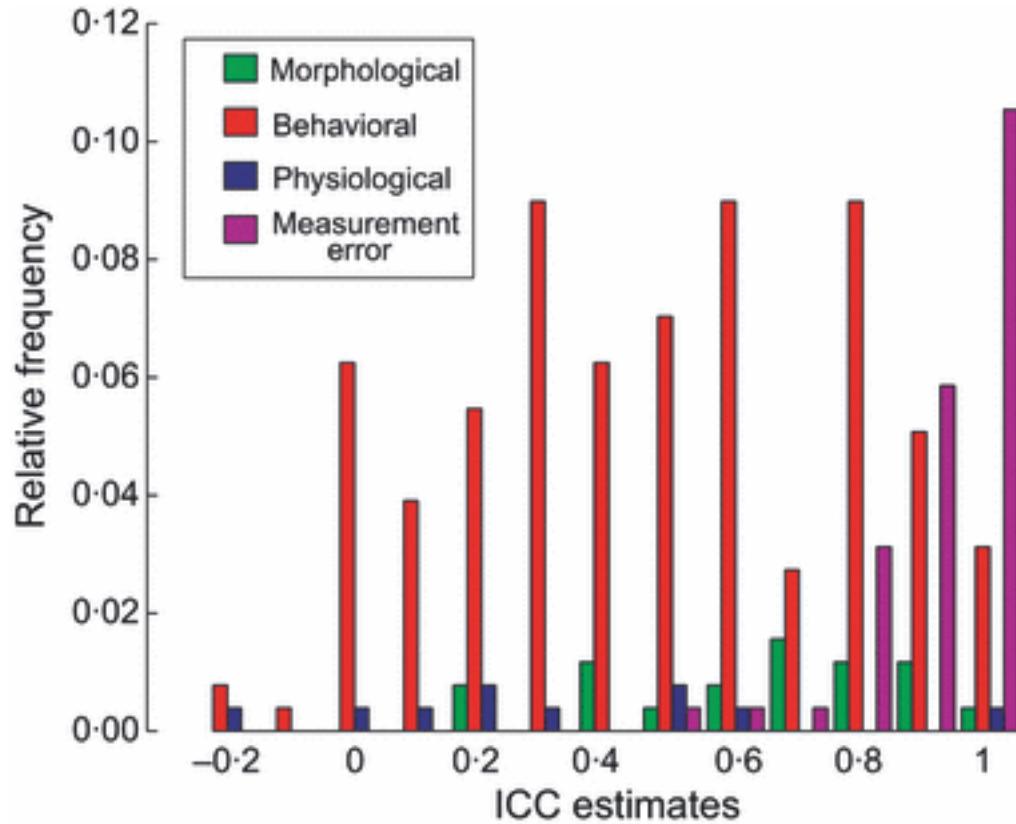
$$\mu = 0.92$$

$$S_A^2 = 0.81$$

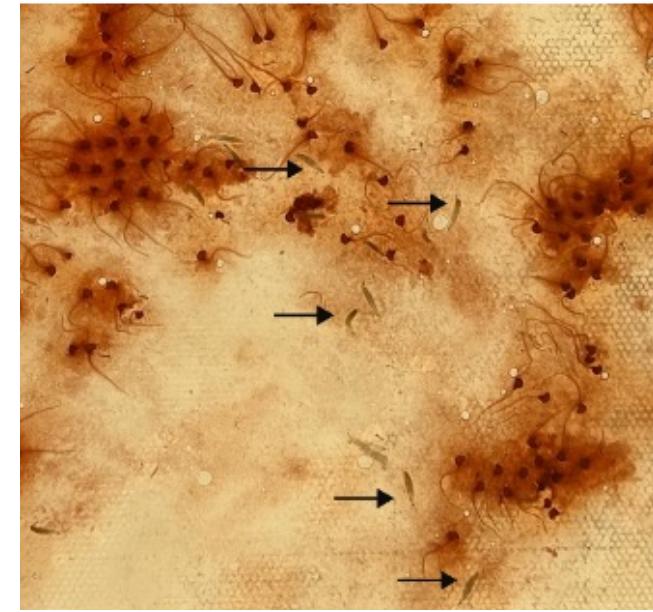
$$S_W^2 = 0.05$$

$$\text{Repeatability(ICC)} = \frac{S_A^2}{S_A^2 + S_W^2} = 0.94$$

Comparison with other datasets

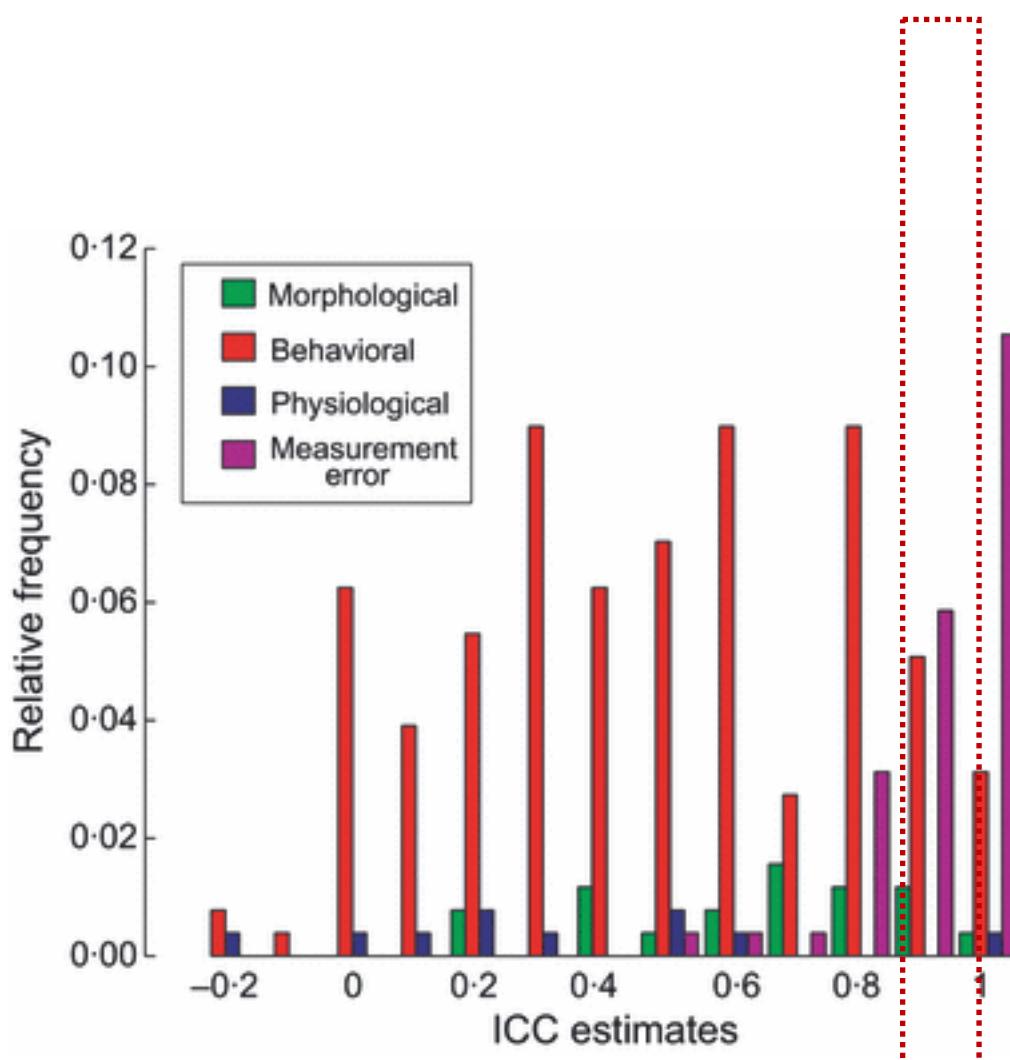


Wolak et al 2012 *Methods in Ecology and Evolution*

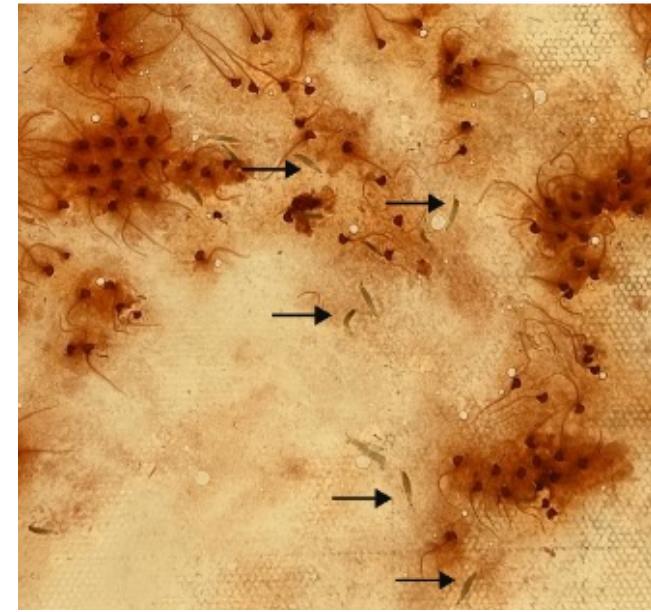


$$\text{Repeatability}(ICC) = \frac{S_A^2}{S_A^2 + S_W^2} = 0.94$$

Comparison with other datasets



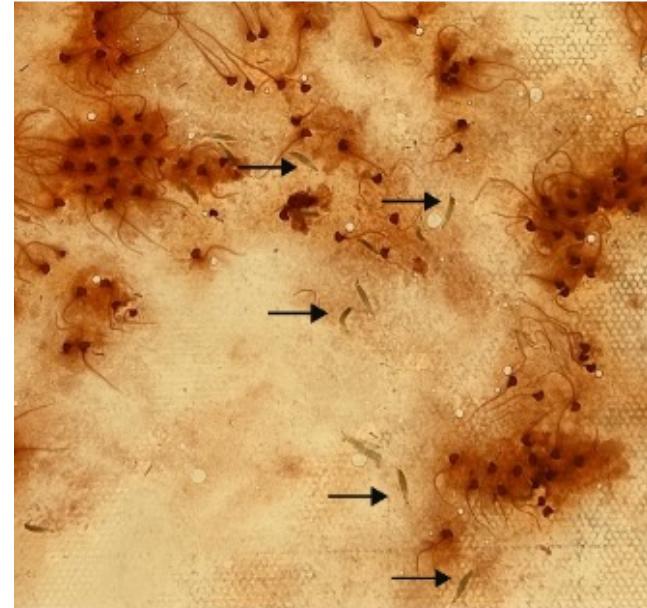
Wolak et al 2012 *Methods in Ecology and Evolution*



OK for measurement error

$$\text{Repeatability}(ICC) = \frac{S_A^2}{S_A^2 + S_W^2} = 0.94$$

Comparison with other datasets

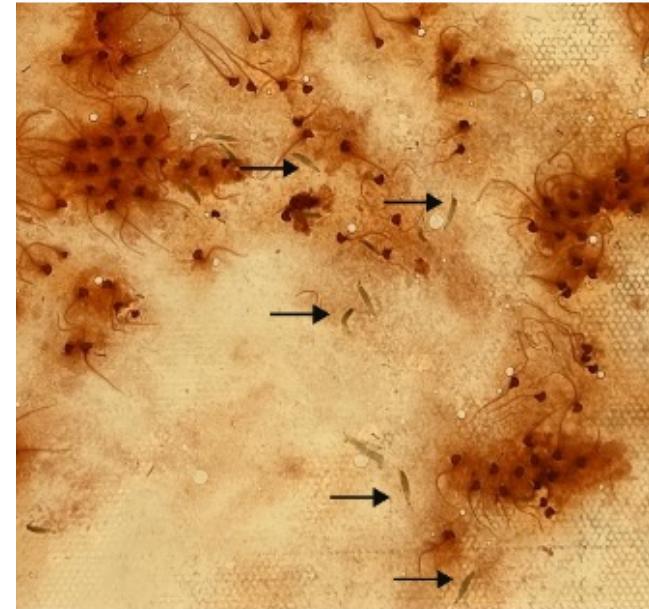
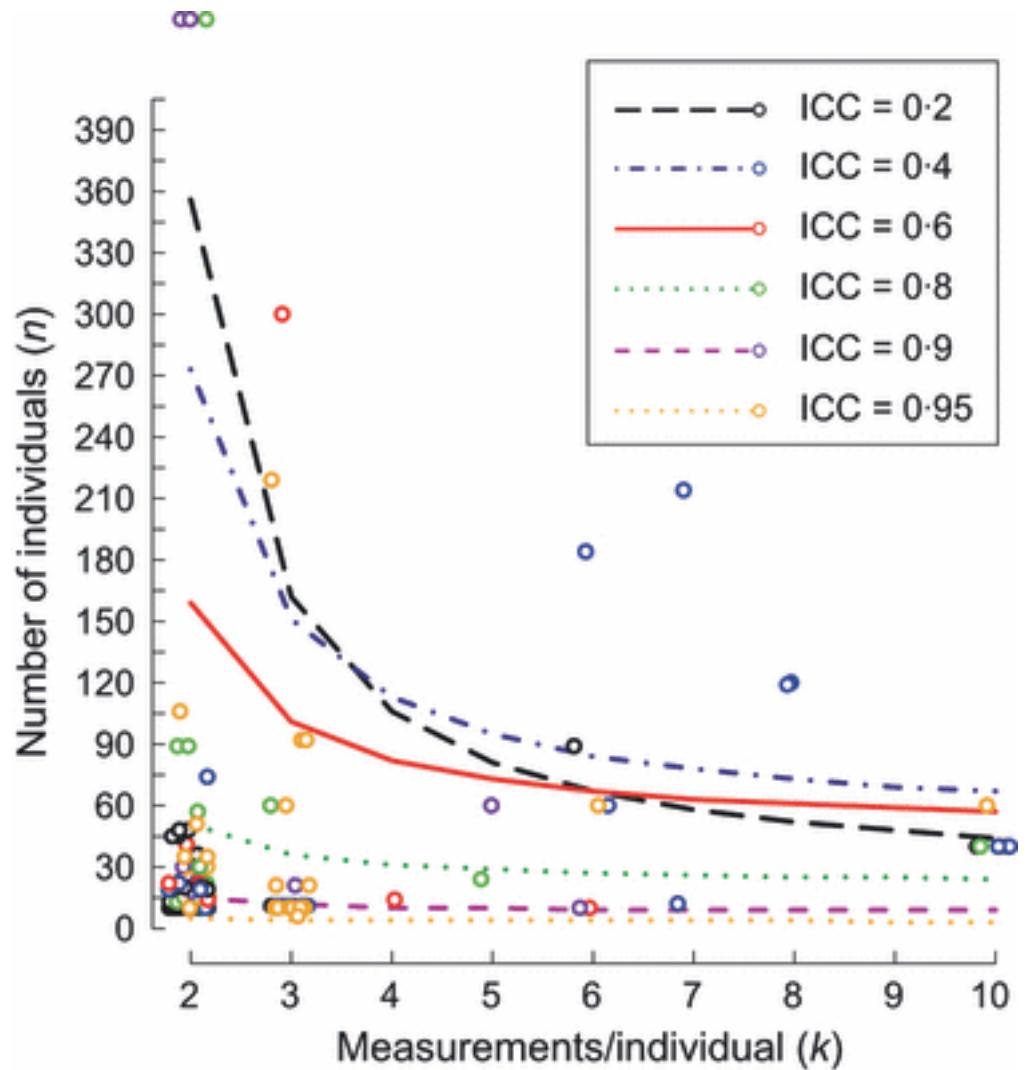


n=30 strawberries
k= 20 replicates

Wolak et al 2012 *Methods in Ecology and Evolution*

$$\text{Repeatability}(ICC) = \frac{S_A^2}{S_A^2 + S_W^2} = 0.94$$

Comparison with other datasets



$n=30$ strawberries
 $k=20$ replicates

Wolak et al 2012 *Methods in Ecology and Evolution*

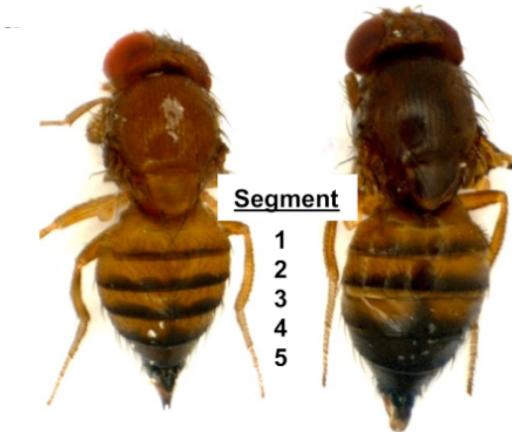
$$\text{Repeatability}(ICC) = \frac{S_A^2}{S_A^2 + S_W^2} = 0.94$$

Non-Gaussian traits

Measurement of abdominal melanization and wing length

Abdominal melanization was quantitatively rated using a visual scoring system under a stereomicroscope as described in [43, 72]. The thickness of each melanized band along the dorso-ventral line of abdominal segments 1–5 was estimated and scored on a scale from 0 (no melanization) to five (complete melanization) (Additional file 1: Figure S1). Melanization scores for the five melanized abdominal bands were compared between morphs for each sex.

Shearer et al 2016 *BMC Ecology*



Non-Gaussian traits

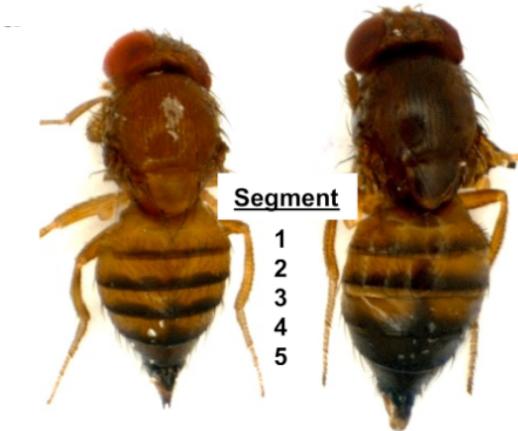
Measurement of abdominal melanization and wing length

Abdominal melanization was quantitatively rated using a visual scoring system under a stereomicroscope as described in [43, 72]. The thickness of each melanized band along the dorso-ventral line of abdominal segments 1–5 was estimated and scored on a scale from 0 (no melanization) to five (complete melanization) (Additional file 1: Figure S1). Melanization scores for the five melanized abdominal bands were compared between morphs for each sex.

Shearer et al 2016 *BMC Ecology*

=> GLMM for non-Gaussian traits

=> Interpretation difficult

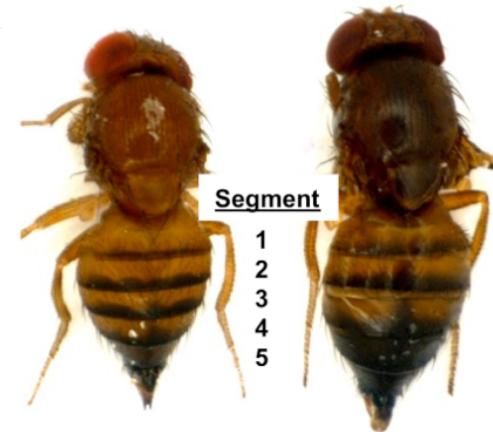


Non-Gaussian traits

Measurement of abdominal melanization and wing length

Abdominal melanization was quantitatively rated using a visual scoring system under a stereomicroscope as described in [43, 72]. The thickness of each melanized band along the dorso-ventral line of abdominal segments 1–5 was estimated and scored on a scale from 0 (no melanization) to five (complete melanization) (Additional file 1: Figure S1). Melanization scores for the five melanized abdominal bands were compared between morphs for each sex.

Shearer et al 2016 *BMC Ecology*



=> GLMM for non-Gaussian traits
=> Interpretation difficult

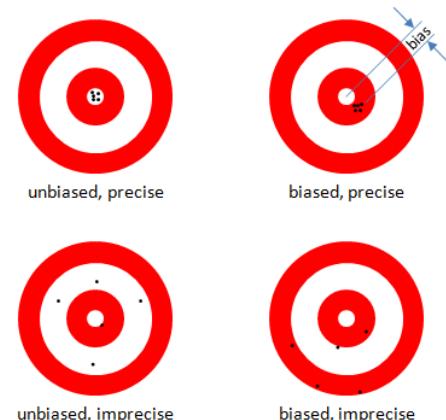
References

Lessells, C. M., & Boag, P. T. (1987). Unrepeatable repeatabilities: a common mistake. *The Auk*, 104(1), 116-121. https://www.researchgate.net/profile/Peter-Boag-2/publication/231424296_Unrepeatable_Repeatabilities_A_Common_Mistake/links/0912f5069dcf9a594b000000/Unrepeatable-Repeatabilities-A-Common-Mistake.pdf

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, 85(4), 935-956. <https://doi.org/10.1111/j.1469-185X.2010.00141.x>

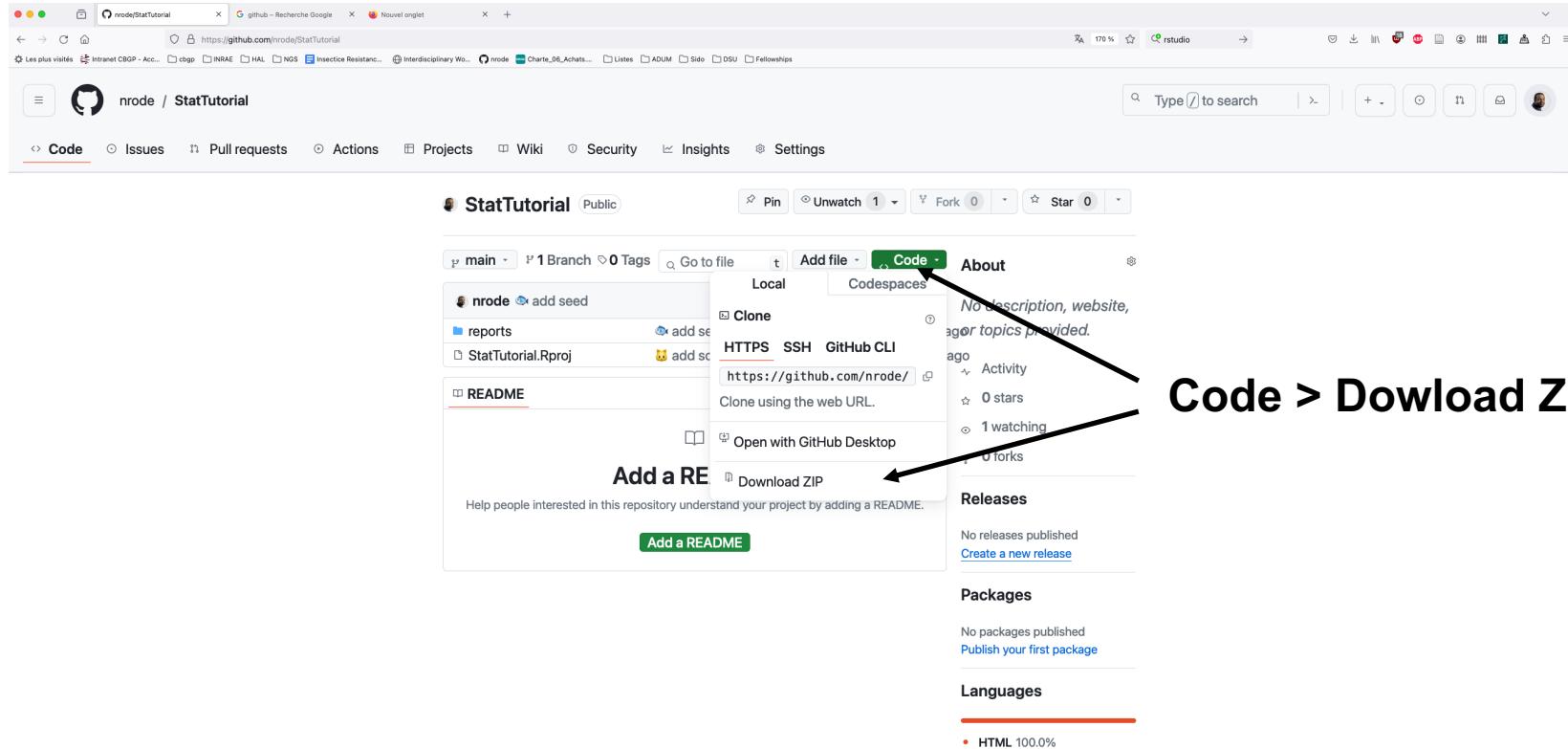
Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, 3(1), 129-137. <https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.2041-210X.2011.00125.x>

Zriki, G., Blois, R., Fournier, C., Tregot-Bertrand, L., Poupart, P. Y., Bardel, A., ... & Rode, N. O. (2023). A fast and reliable larval sampling method for improving the monitoring of fruit flies in soft and stone fruits, *Journal of Economic Entomology*, 2024;, toae001, <https://doi.org/10.1093/jee/toae001>



Travaux pratiques!

<https://github.com/nrode/StatTutorial>



Travaux pratiques!



The screenshot shows the R Studio interface with the following components:

- Source Pane:** Displays the R Markdown file `Compute_repeatability.Rmd`. The code includes YAML front matter and R code for generating simulated data and fitting models.
- Environment Pane:** Shows the global environment, which is currently empty.
- Files Pane:** Lists the contents of the `reports` folder, including the generated HTML file `Compute_repeatability.html` (1.3 MB) and the RMD file `Compute_repeatability.Rmd` (7.4 KB).
- Console Pane:** Displays the R command-line interface with the R version and working directory information.

```
1 ---  
2 title: "Compute repeatability with R"  
3 author: "Nicolas RODE"  
4 date: "`r format(Sys.Date(), '%d-%B-%Y')`"  
5 output:  
6   html_document:  
7     number_sections: yes  
8     theme: journal  
9     toc: yes  
10    toc_float:  
11      collapsed: no  
12      smooth_scroll: no  
13 editor_options:  
14   chunk_output_type: console  
15 ---  
16  
17 # Simulate data  
18 ## Enter values for simulations  
19 ``{r }  
20 library(ggplot2)  
21 library(dplyr)  
22 library(tidyr)  
23  
24 ## Number of individual measured  
25 n <- 20  
26 # Compute repeatability with R
```

R 4.3.1 · ~/Documents/Pro/Stat_Tutorials/StatTutorial/
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.