

# ***Chapter 1:***

## ***Statistics and samples***

1

### ***Biostatistics***

- **Statistics:** the study of methods for measuring aspects of populations from samples and for quantifying the uncertainty of the measurements

2

## *Sampling populations*

- A **population** is all the individual units of interest, whereas a *sample* is a subset of units taken from the population

**text like this** means that this term/definition appears in the textbook as an important note highlighted in a blue box

e.g., Boston College class of 2023

3

## *Sampling populations*

- **Variables** are characteristics that differ among individuals

4

## Sampling populations

- A **parameter** is a quantity describing a population, whereas an *estimate* or *statistic* is a related quantity calculated from a sample
- The *parameter is the truth*, whereas the *estimate/statistic is an approximation of the truth* that is subject to error
- Average height of a population is a parameter that can be approximated by an estimate/statistic of a sample

5

## Sampling populations

- **Estimation** is the process of inferring an unknown quantity of a population using sample data
- How would you sample height data to form an estimate?

6

## *Sampling populations*

- In a **random sample**, each member of a population has an equal and independent chance of being selected
  - Minimizes bias and makes it possible to measure the amount of sampling error
- A **sample of convenience** is a collection of individuals that are easily available to the researcher

7

## *Sampling populations*

- **Sampling error** is the difference between an estimate and the population parameter being estimated caused by chance
- **Bias** is a systematic discrepancy between the estimates we would obtain, if we could sample a population again and again, and the true population
- **Volunteer bias**: resulting from systematic differences between the pool of volunteers and the population to which they belong
- How might your sample of heights contain error/bias?

8

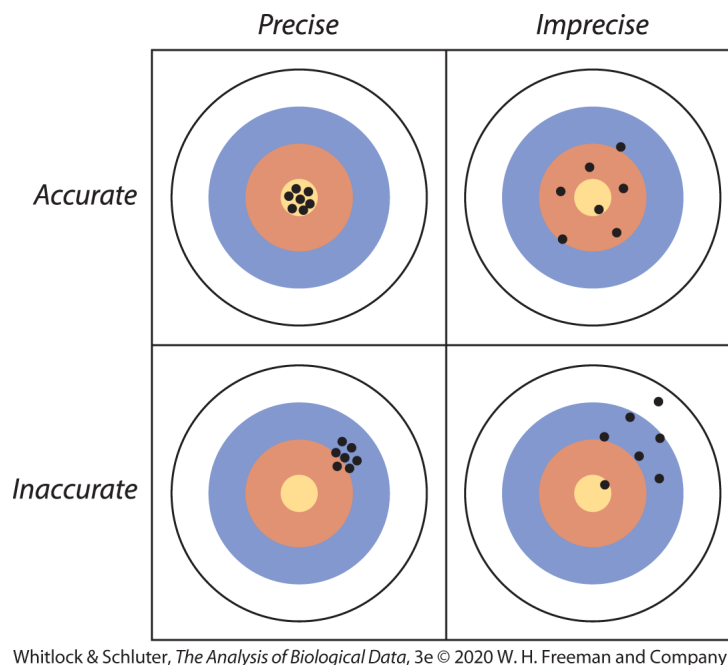


Fig 1.2-2

9

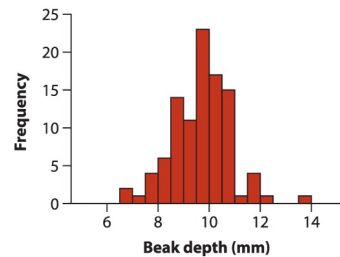
## Types of data

- **Categorical data** are qualitative characteristics of individuals that do not have magnitude on a numerical scale
  - e.g., present or absent; language (English, Spanish, etc.); life stage (egg, larvae, adult)
  - Can be **nominal** (no inherent order) or **ordinal** (inherent order)
- **Numerical data** are quantitative measurements that have magnitude on a numerical scale
  - e.g., age, number of mates, heartrate, length
  - Can be **discrete** (indivisible units) or **continuous** (any real number)

10

## Frequency distributions

- The **frequency distribution** describes the number of times each value of a variable occurs in a sample



<https://www.youtube.com/watch?v=mcM23M-CCog>

The large-beaked ground finch on the Galápagos Islands.

Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company  
Dolph Schluter

11

## Types of studies

- A study is **experimental** if the researcher assigns treatments randomly to individuals, whereas a study is **observational** if the assignment of treatments is not made by the researcher

12

## *Types of variables*

- One common use of statistics is to relate one variable to another (i.e., explore associations between variables)
- How well an **explanatory variable** can predict or affect the **response variable**
  - In graphs, it is typical to have the **explanatory variable on the x-axis** and the response variable on the y-axis
  - [https://www.youtube.com/watch\\_popup?v=gxSUqr3ouYA](https://www.youtube.com/watch_popup?v=gxSUqr3ouYA)
- A **confounding variable** is a variable that masks or distorts the causal relationship between measured variables in a study

13

*Biologists use statistics all the time!!!*

14

# *Links between diet and probability of seizures*



Plasma glucose

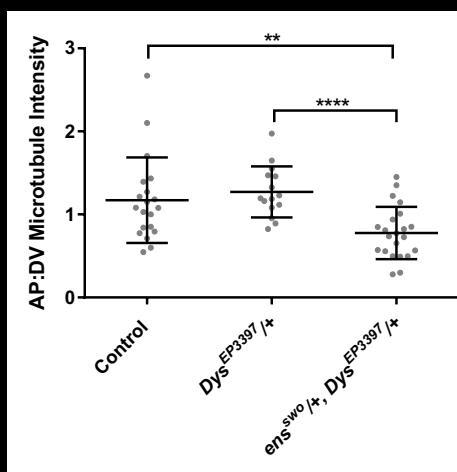
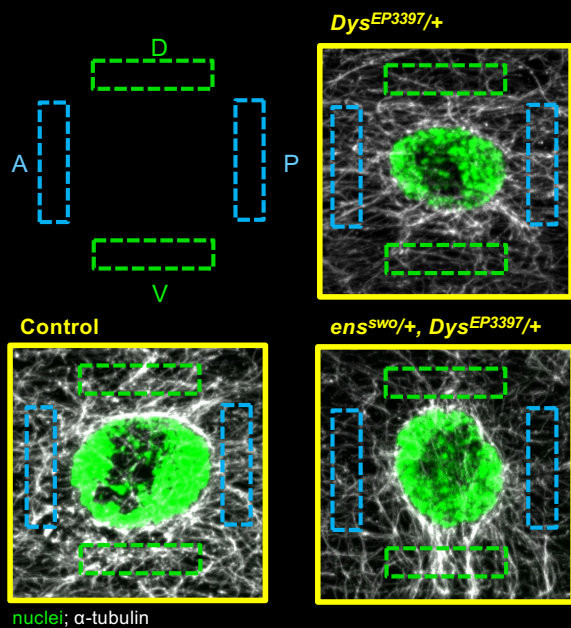
	Low	Med	High
Seizure	8	44	70
No seizure	226	190	164



Thomas Seyfried

15

## Microtubule organization disrupted in double heterozygotes



Eric Folker

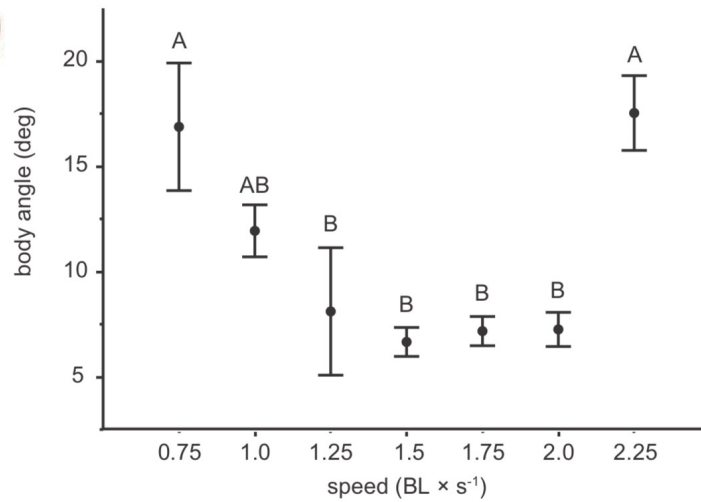


Alex Burgess

16



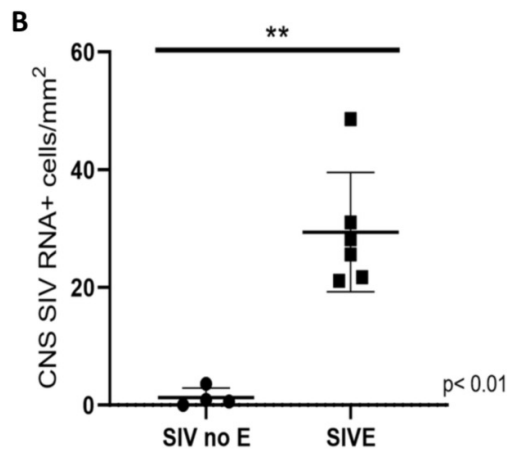
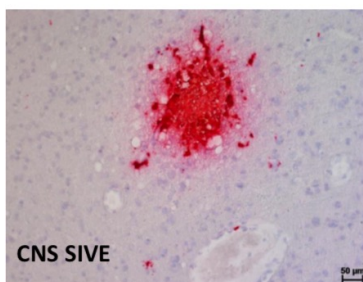
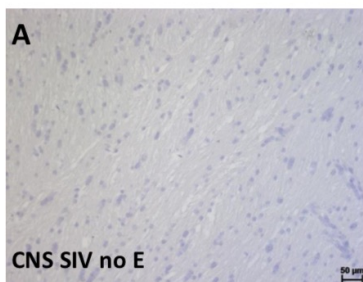
## Association between swimming speed and body position



Chris Kenaley

17

## Association between SIV infection of central nervous system and encephalitis



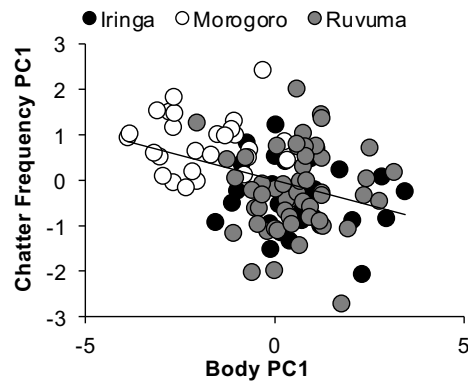
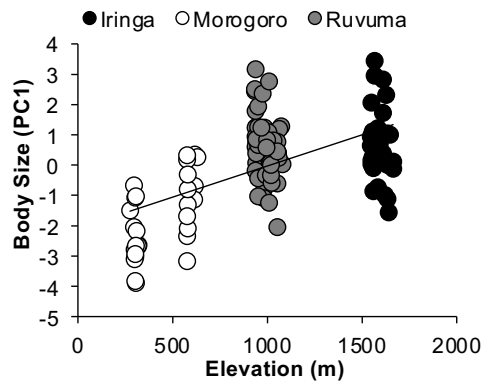
Kenneth Williams



Kevin White

18

## Correlations between elevation, body size, and song frequency



Jeff DaCosta