



Master Environmental Sciences and Engineering

Applied Ecology

Data analysis

10/09/2024

Schedule of the lectures

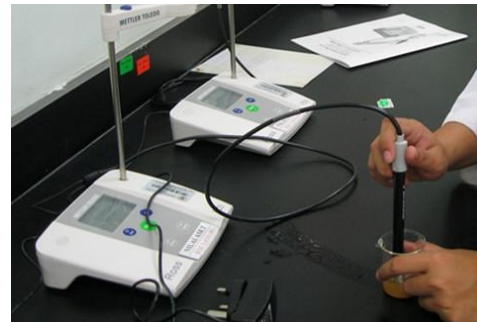
Autumn semester 2025					
Date	Time	Topic	Week	Location	Note
9/9/25	8-12h	Introduction	1	CHB 330	
16/9/25	8-12h	Presentation of cases 1, 2 and 3	2	CHB 330	
23/9/25	8-12h	Field work #1	3	Field	
30/9/25	8-12h	Field work #2	4	Field	
7/10/25	8-12h	Excursion to Lausanne	5	Field	Deadline to provide case study topic: 1/2 page summary
14/10/25	8-12h	Lab work	6	GR B2 423	Deadline to upload your data on Moodle
HOLIDAYS					
28/10/25	8-12h	Presentation by Spaak Piet (Eawag) at 9h15 Work on data and report	7	CHB 330	
4/11/25	8-12h	Excursion to Suchy	8	CHB 330	Deadline to submit the written report
11/11/25	8h30-10h 10-12h	Presentation of cases 4, 5, and 6	9	CHB 330	
18/11/25	09-11h 8-12h	Excursion EPFL Energy plateforme (1.5 h) Work on Presentation	10	CHB 330	
25/11/25	09-11h 8-12h	Excursion EPFL Energy plateforme (1.5 h) Work on Presentation	11	CHB 330	
2/12/25	8-12h	Work on Presentation	12	CHB 330	
9/12/25	8-12h	Student presentations I	13	CHB 330	Groups 1-5
16/12/25	8-12h	Student presentations II	14	CHB 330	Groups 6-10



Objectives of this lecture

(2) Compare the ecosystem services provided by a monoculture plantation vs. a semi-natural forest

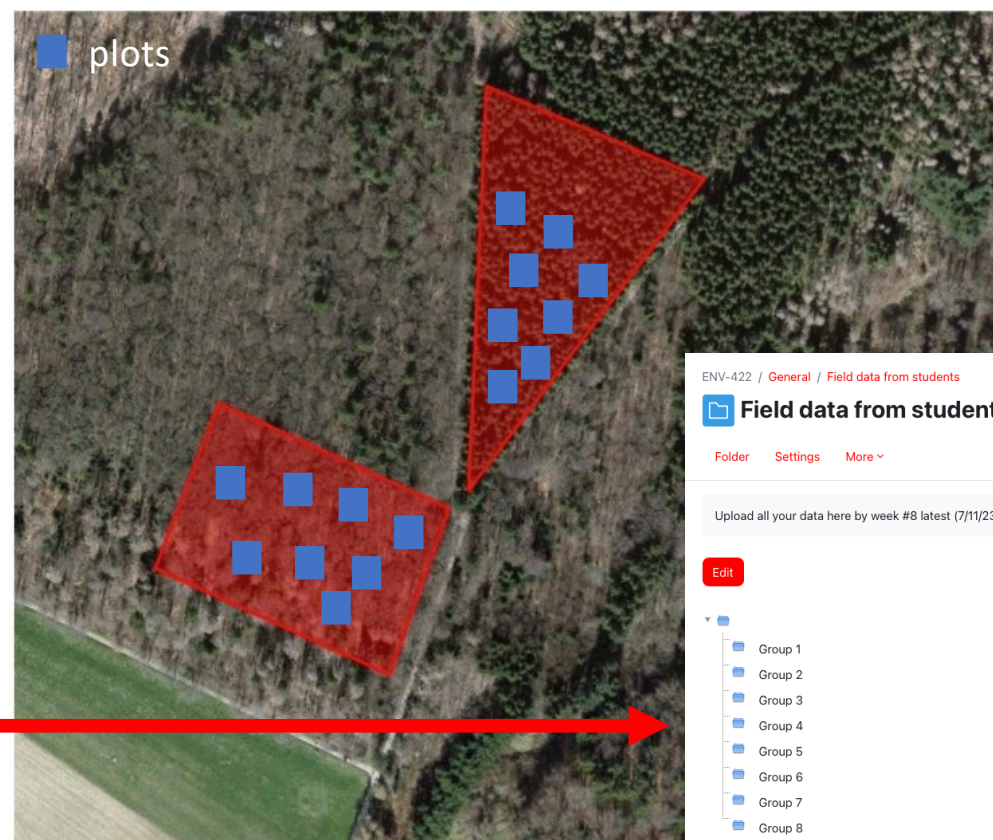
- Potential biodiversity index (IBP)
- Shannon index (H)
- Diameter at breast height (DBH), tree height, and tree density
- Presence of dendro-microhabitats
- Soil pH
- Plot Leaf Area Index (LAI)
- Aboveground wood production



- read the files on Moodle
- Come prepared for fieldwork

Objectives of this lecture

- All groups will select 1 plot of 10 x 10 m in each forest type (2 plots per group in total)
- Measurements will take place over two days
- At the end of the work, all students will upload their data on Moodle so that all students can use them for the analyses (to increase sample size)



Objectives of this lecture

(2) Compare the ecosystem services provided by a monoculture plantation vs. a semi-natural forest

- **Grading of the report (50% of the total grade with 50/50 ratio for the whole group and your individual contribution to the work):** You will be evaluated based on the description of the context (literature review), the presentation of the methods and results (methods and results are separated parts) and the discussion (use the literature to place the results in a wider context by explaining the consequences of the results).
- The report should not exceed 7 pages, excluding the references, and should not have more than 4 figures.
- In the report, students need to explain their individual contribution.

Deadline: 4.11.2025, you need to send your reports to the teachers.

Structure your data

Example Excel sheet:

Group	Site	pH	LAI	IBP_score	IBP_numeric	Shannon index	Tree density
1	Pure	4.3	4.07	faible	1
1	Mixed	4.8	2.74	assez_forte	4		
2	Pure	4.2	3.27	faible	1		
2	Mixed	5.1	2.9	assez_forte	4		
3	Pure	5.3	6.8	faible	1		
3	Mixed	4.7	5.6	assez_forte	4		

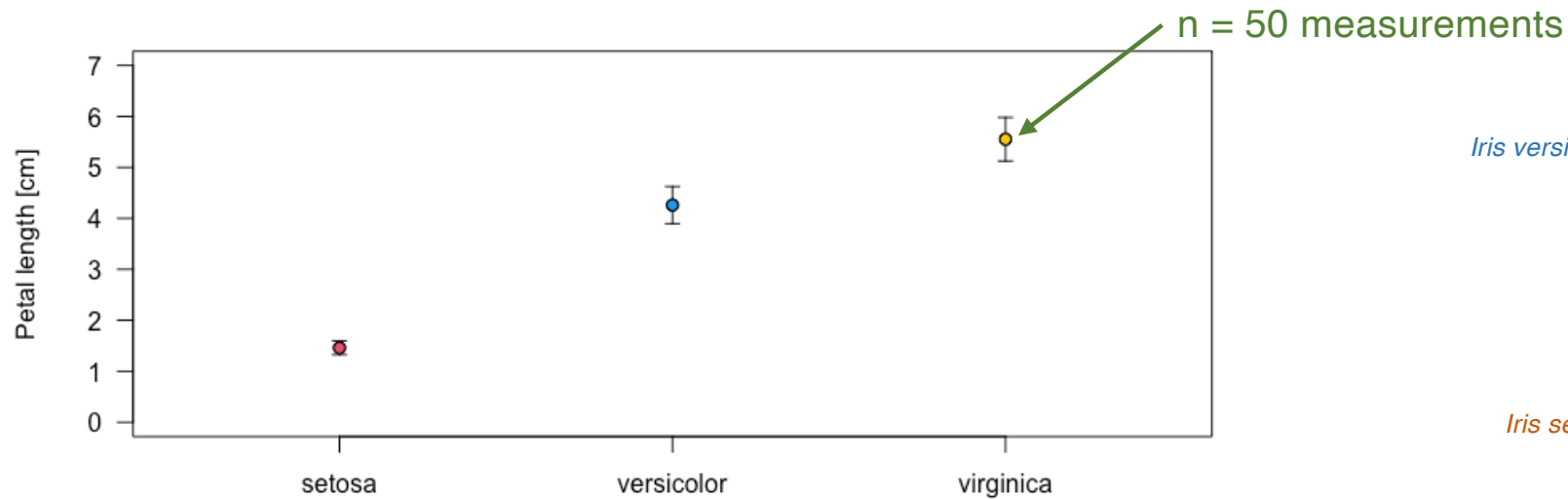
Structure your data

Example Excel sheet:

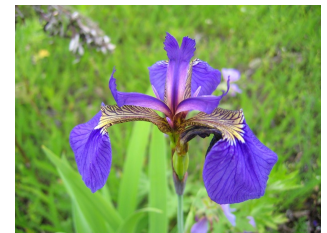
Group	Site	Tree_ID	Species	Height	DBH	Aboveground biomass
1	Pure	1	<i>Picea abies</i>	17.5	38	...
1	Pure	2	<i>Picea abies</i>	24.5	45	
1	Pure	3	<i>Castanea sativa</i>	23.0	34	
1	Mixed	1	<i>Fagus sylvatica</i>	13.2	29	
1	Mixed	2	<i>Abies alba</i>	26.6	53	
1	Mixed	3	<i>Picea abies</i>	21.3	37	

Descriptive statistics

Don't show all data points, but means and standard errors:



Iris versicolor



Iris setosa



Iris virginica

→ s.e. provides a "gut feeling" of differences, but is not a statistical test

Descriptive statistics

Don't show all data points, but means and standard errors:

For instance, to summarise by plot and site

Use the summarise function from the dplyr package with the following code:

```
install.packages("dplyr")
library(dplyr)

my_data_agg <- my_data %>%
  group_by(Plot, Site) %>%
  summarise(n = n(),
            height_m = mean(height, na.rm=T),
            height_se = se(height, na.rm=T)/sqrt(n),
            DBH_m = mean(DBH, na.rm=T),
            DBH_se = se(DBH, na.rm=T)/sqrt(n))
```

Group comparisons

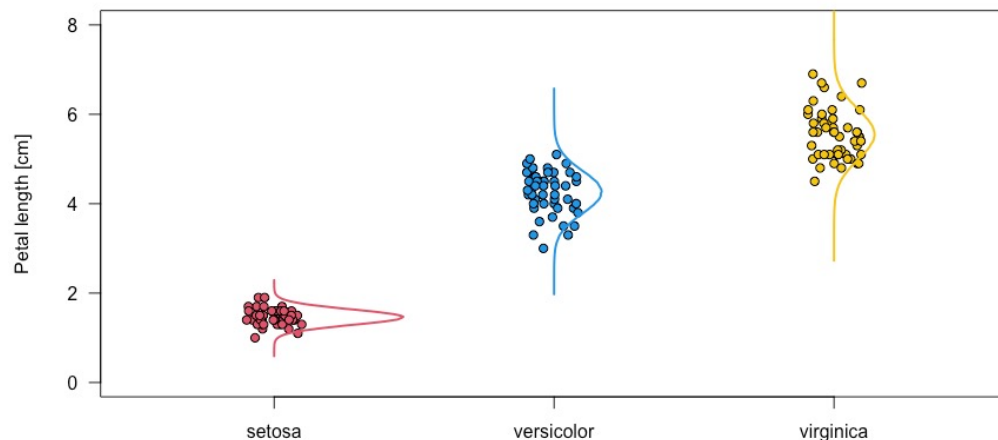
Use ANOVA to compare two or more groups
(e.g. mixture vs. monoculture or species)

```
> mymodel <- aov(Petal.Length ~ Species, data=iris)
> summary(model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Species	2	437.1	218.55	1180	<2e-16 ***
Residuals	147	27.2	0.19		

The probability that the measured difference between the species is observed by chance is very small.

So, the species probably differ. But all of them? We don't know with this anova.



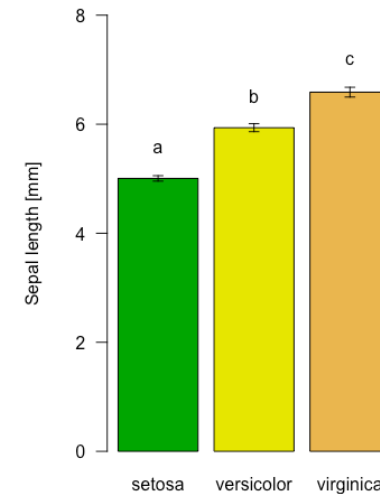
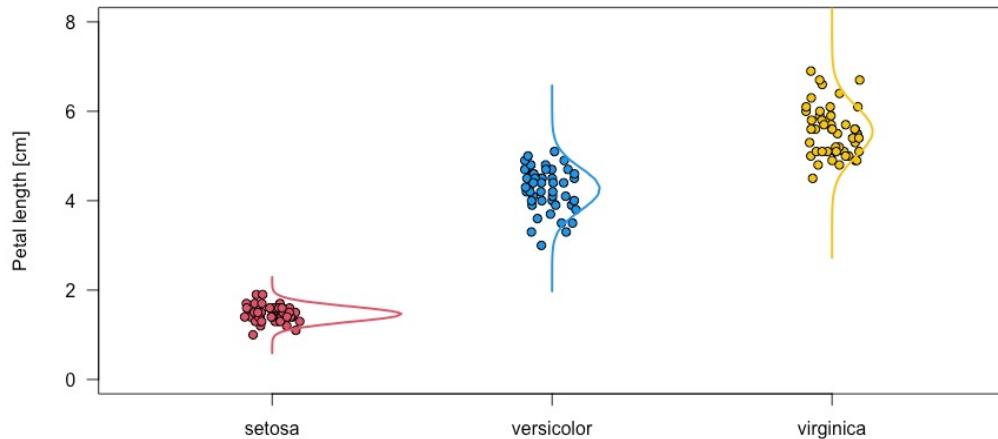
Group comparisons

Pairwise comparison: Tukey test (e.g. for species comparison)

```
> my_model <- aov(Petal.Length ~ Species, data=iris)
> TukeyHSD(my_model)
```

	diff	lwr	upr	p adj
versicolor-setosa	2.798	2.59422	3.00178	0
virginica-setosa	4.090	3.88622	4.29378	0
virginica-versicolor	1.292	1.08822	1.49578	0

All species differ from each other.



Group comparisons

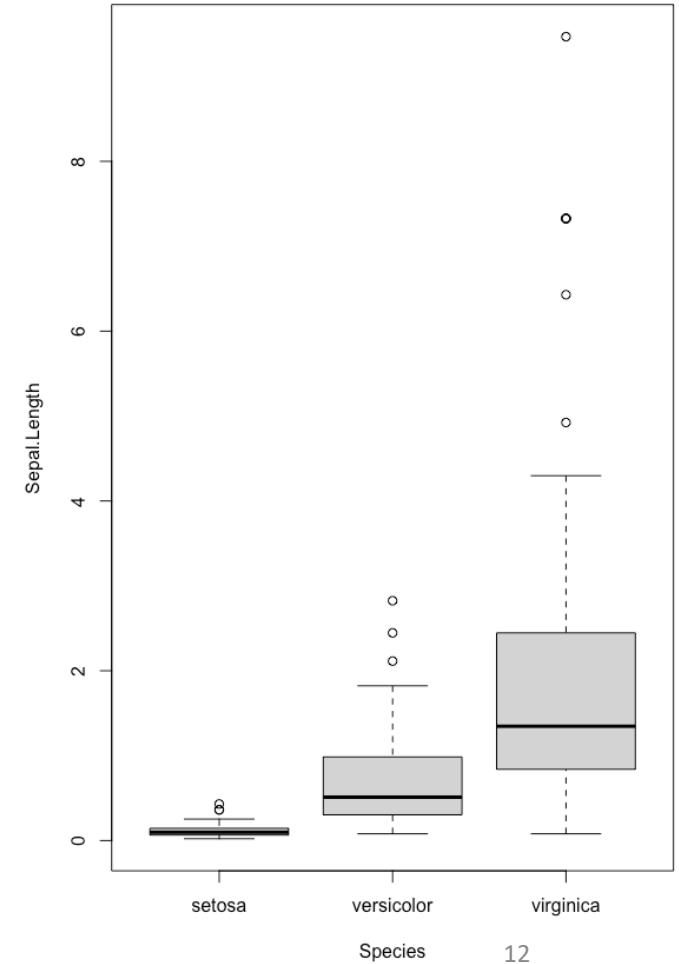
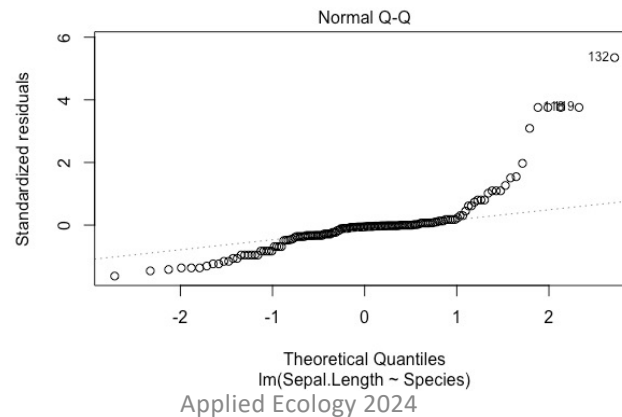
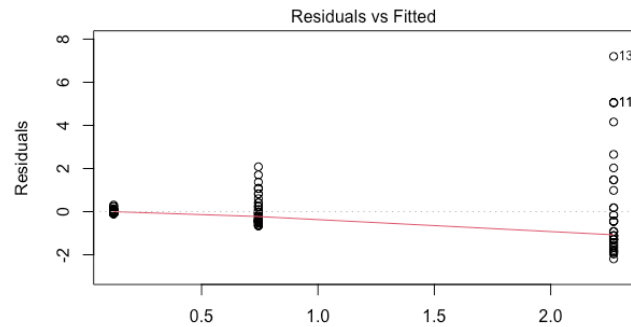
What to do if the data is not normally distributed?

→ transform the data: log-transformation or sqrt-transformation

```
> my_data$log_measurement <-
  log(my_data$measurement)

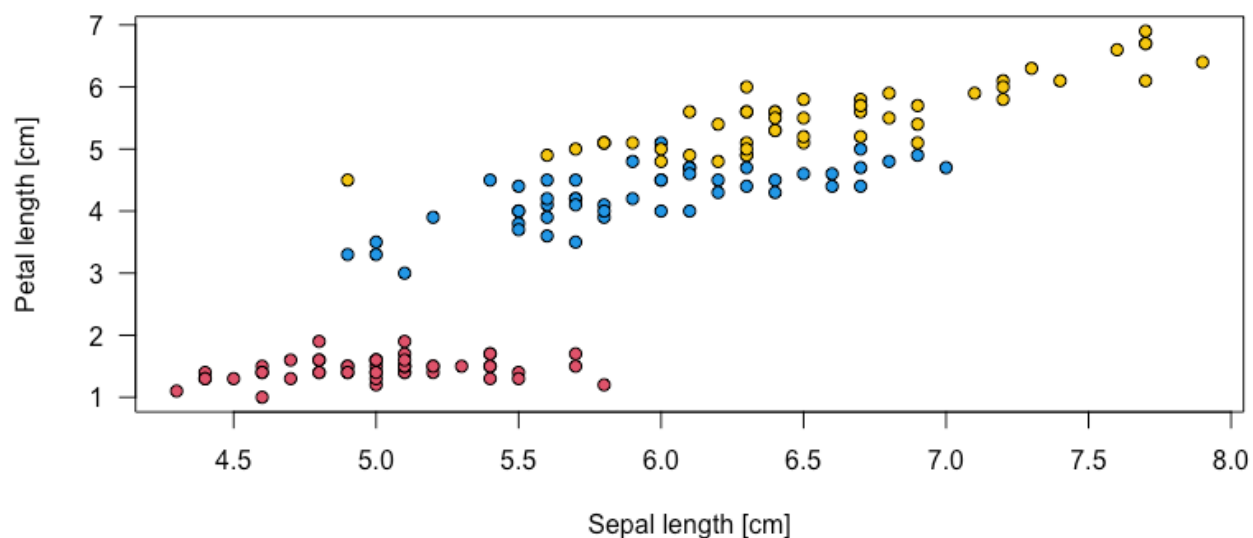
> my_data$sqrt_measurement <-
  sqrt(my_data$measurement)
```

→ Don't hesitate to transform if you think it improves the analyses



Correlation of measurements

Show relationships between measurements



```
> cor(my_data$my_measurements1,  
my_data$my_measurements2)
```

```
[1] 0.8717538
```

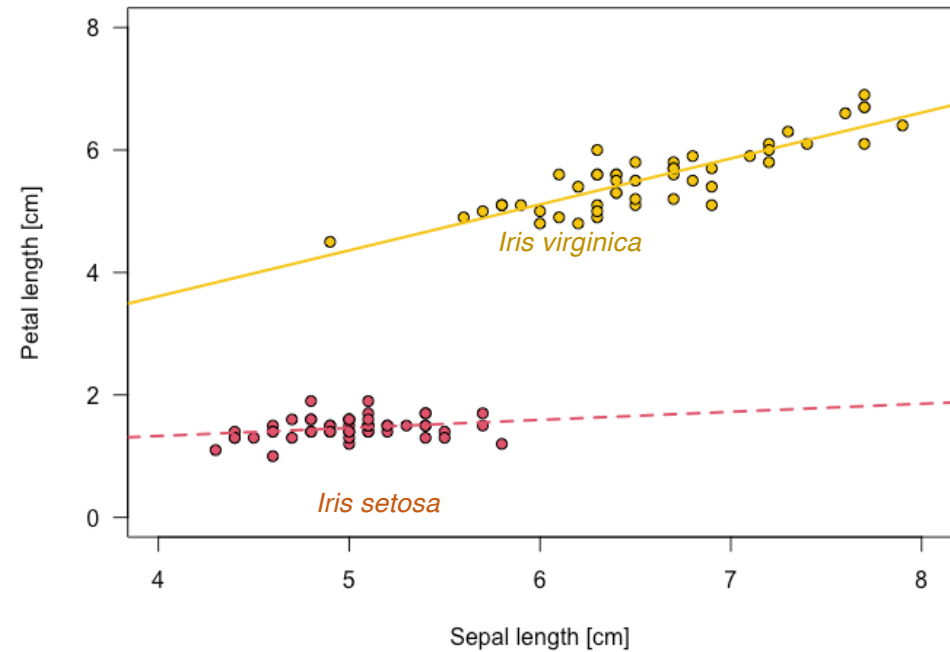
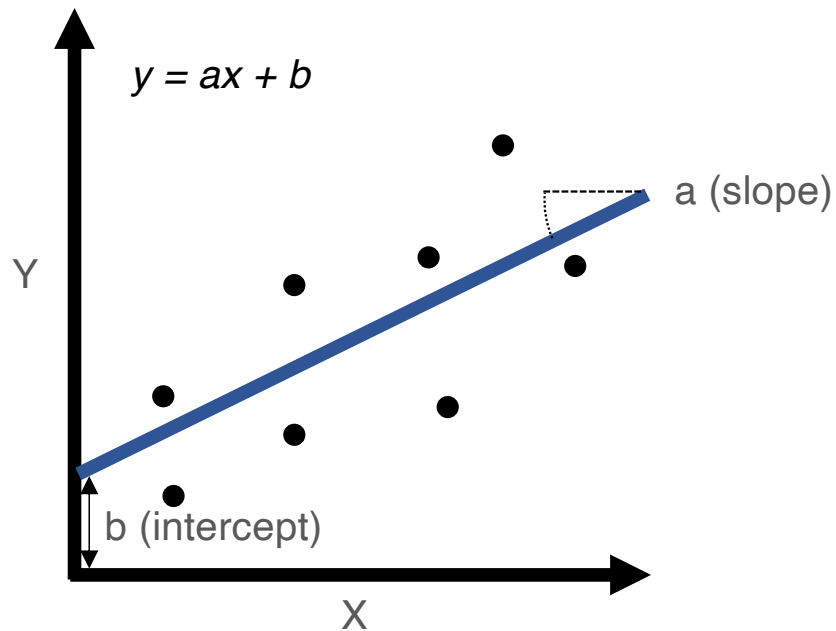
```
> cor.test(my_data$my_measurements1,  
my_data$my_measurements2,  
method = "pearson")
```

```
t = 21.646, df = 148, p-value < 2.2e-  
16
```

→ no causality inferred! Both measurements might be related to something else (fertilisation, species, etc.)

Linear regression

Test the linear effect of X on Y



```
> model1 <- lm(Petal.Length ~ Sepal.Length, data=iris[iris$Species=="setosa", ])
> summary(model1)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.80305	0.34388	2.335	0.0238 *
Sepal.Length	0.13163	0.06853	1.921	0.0607 .