Day 1 - Part I
Course introduction

1. Course Structure & Schedule, Course Materials, Exam
2. Motivations for the choice of Methods
3. What is OCBE?

Valeria Vitelli
Oslo Centre for Biostatistics and Epidemiology (OCBE)
Department of Biostatistics, UiO
valeria.vitelli@medisin.uio.no

MF9130E – Introductory Course in Statistics
08.04.2024

Course Structure

Structure Week 1: Flipped Classroom

- afternoon:
    - lecture on a topic + Q & A session
    - some free time for reflection / group work
- morning after: practical lab session on the same topic

Structure Week 2: More Traditional Approach

- Half-day blocks on a specific topic
- Structure of each block:
    - lecture on a topic
    - practical lab session on the same topic

Course Schedule

Week 1

	Monday	Tuesday	Wednesday	Thursday	Friday
AM		Lab 1 Lab 2	Distributions Lab 4	Lab 5 t-test	Lab 6 tables
PM	Intro Course Data & Descript Statistics	Intro Prob Diagnostics Lab 3	Inference part 1 t-test	Inference part 2 table analysis	

Week 2

	Monday	Tuesday	Wednesday	Thursday
AM	Non-parametrics Lab	Study designs	Regression 2 Lab	Logistic Lab
PM	Sample size Power Lab	Regression 1 Lab	Regression 3 lab Course Summary	Survival Lab

Overview for Day 1: "Data and Descriptive Statistics"

This afternoon: Lectures in flipped classroom style

- **Introduction** to this course
- **Data and statistics** in medicine: Introduction and motivation
- **Descriptive statistics**
  - Data presentation: Central measures & Measures of variation
  - Graphical presentation of data
- **Self-study and Group-work**

Tomorrow morning: Labs in flipped classroom style

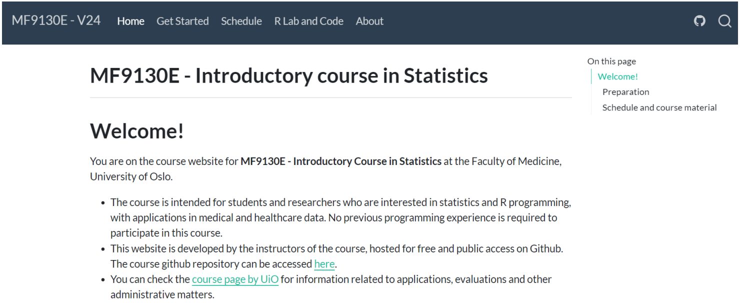
- Introduction to statistical computing with **R**
- Descriptive statistics with **R**

Course textbook chapters:

- Kirkwood and Sterne chapters 2-4
- Aalen chapters 1 and 2

Links and Course Material

- **Course webpages:** `https://ocbe-uio.github.io/teaching_mf9130e/`
- We will mainly use the course webpages for all information and access to material. The webpages will be continuously updated throughout the course.



MF9130E - Introductory course in Statistics

Welcome!

You are on the course website for **MF9130E - Introductory Course in Statistics** at the Faculty of Medicine, University of Oslo.

- The course is intended for students and researchers who are interested in statistics and R programming, with applications in medical and healthcare data. No previous programming experience is required to participate in this course.
- This website is developed by the instructors of the course, hosted for free and public access on Github. The course github repository can be accessed here.
- You can check the course page by UiO for information related to applications, evaluations and other administrative matters.

Links and Course Material – Alternatives

- **Canvas room:** We will not use the Canvas room a lot, but Canvas is used for **emails** and general communication. Please let us know asap, if you do not have access to Canvas!
- **Official UiO course pages** with schedule, literature and details on admission rules, exam etc: `https://www.uio.no/studier/emner/medisin/med/MF9130E/`

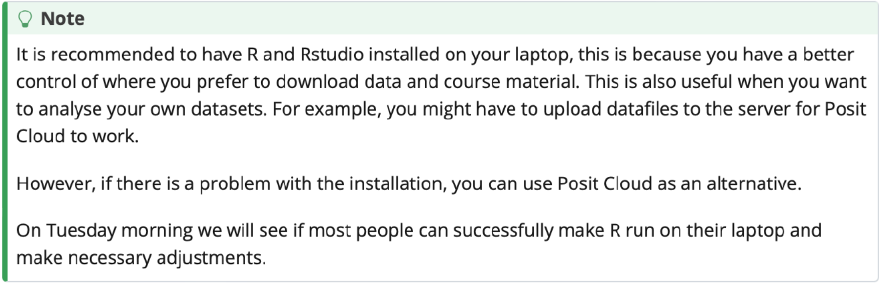
Computer exercises in R (starting tomorrow morning)

- You will need to have a laptop computer with access to R and RStudio for the labs.
- We advise that you install R/RStudio on your own laptop.
- Alternatively, you could register for a (free trial) account on a Posit Cloud server.
- See here for instructions: `https://ocbe-uio.github.io/teaching_mf9130e/get_started/get_started.html`

- Note: You can also access R/RStudio through the UiO Programkiosk: `https://www.uio.no/english/services/it/home-away/kiosk/`.

Homework for tomorrow morning

- Go through the instructions above to get working access to R and RStudio. There will be a detailed introduction to R and RStudio tomorrow morning.



Note

It is recommended to have R and Rstudio installed on your laptop, this is because you have a better control of where you prefer to download data and course material. This is also useful when you want to analyse your own datasets. For example, you might have to upload datafiles to the server for Posit Cloud to work.

However, if there is a problem with the installation, you can use Posit Cloud as an alternative.

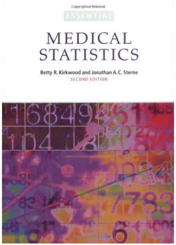
On Tuesday morning we will see if most people can successfully make R run on their laptop and make necessary adjustments.

Exam

- Take-home exam.
- Will be published via Inspera at the end of the course.
- To be submitted within a specified deadline (4 weeks after the end of the course).
- A passed exam is required to get the course approved.

- More details on the third day of week 2.

Main course textbook: Kirkwood and Sterne (2003)



- Betty R. Kirkwood and Jonathan A. C. Sterne. **Essential Medical Statistics**. Second edition, Blackwell Science Ltd, 2003
- `www.blackwellpublishing.com/essentialmedstats/`

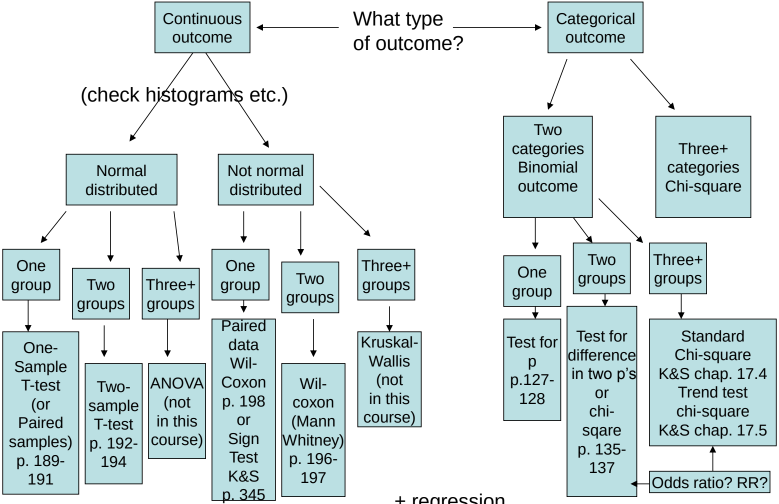
Norwegian alternative: Aalen (ed) *et al* (2006)

- Odd O. Aalen (red), Arnoldo Frigessi, Tron Anders Moger, Ida Scheel, Eva Skovlund, Marit B. Veierød. **Statistiske metoder i medisin og helsefag**. Gyldendal Akademisk 2006
- `www.med.uio.no/imb/studier/ressurser/statistikk/statistikkressurser-shs/aalen.html`

Methods in this course

Page numbers in Aalen
Except K&S=Kirkwood & Sterne

Diagram for test choice



What type of outcome?

Continuous outcome

(check histograms etc.)

Normal distributed

One group

One-Sample T-test (or Paired samples) p. 189-191

Two groups

Two-sample T-test p. 192-194

Three+ groups

ANOVA (not in this course)

Not normal distributed

One group

Paired data Wil-Coxon p. 198 or Sign Test K&S p. 345

Two groups

Wil-coxon (Mann Whitney) p. 196-197

Three+ groups

Kruskal-Wallis (not in this course)

Categorical outcome

Two categories Binomial outcome

One group

Test for p p.127-128

Two groups

Test for difference in two p's or chi-sqare p. 135-137

Three+ groups

Standard Chi-square K&S chap. 17.4 Trend test chi-square K&S chap. 17.5

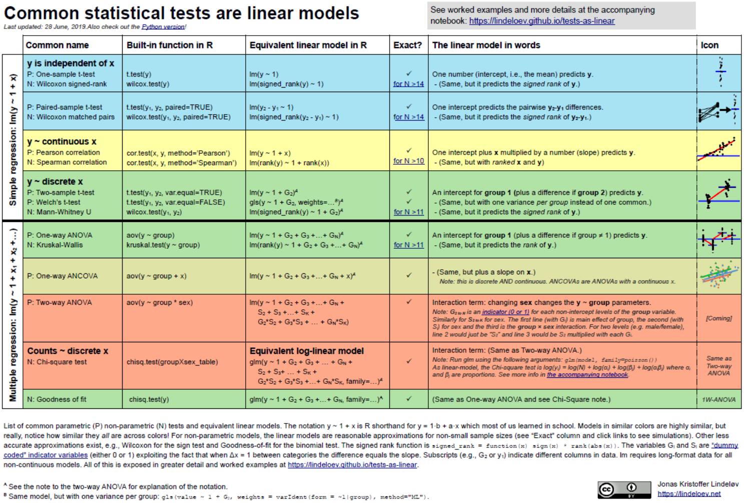
Odds ratio? RR?

Three+ categories Chi-square

+ regression…

Many of the methods we cover can be seen as **linear models**.

- `https://lindeloev.github.io/tests-as-linear/`
- Regression models as well as most statistical tests:



Common statistical tests are linear models

Last updated: 28 June, 2019.Also check out the Python version!

See worked examples and more details at the accompanying notebook: https://lindeloev.github.io/tests-as-linear

	Common name	Built-in function in R	Equivalent linear model in R	Exact?	The linear model in words	Icon

Simple regression: lm(y ~ 1 + x)	**y is independent of x**
	P: One-sample t-test
	N: Wilcoxon signed-rank
		t.test(y)
		wilcox.test(y)
			lm(y ~ 1)
			lm(signed_rank(y) ~ 1)
				✓
				for N >14
					One number (intercept, i.e., the mean) predicts **y**.
					- (Same, but it predicts the *signed rank* of **y**.)	

	P: Paired-sample t-test
	N: Wilcoxon matched pairs
		t.test(y_1, y_2, paired=TRUE)
		wilcox.test(y_1, y_2, paired=TRUE)
			lm(y_2 - y_1 ~ 1)
			lm(signed_rank(y_2 - y_1) ~ 1)
				✓
				for N >14
					One intercept predicts the pairwise **y_2-y_1** differences.
					- (Same, but it predicts the *signed rank* of **y_2-y_1**.)	

	y ~ continuous x
	P: Pearson correlation
	N: Spearman correlation
		cor.test(x, y, method='Pearson')
		cor.test(x, y, method='Spearman')
			lm(y ~ 1 + x)
			lm(rank(y) ~ 1 + rank(x))
				✓
				for N >10
					One intercept plus **x** multiplied by a number (slope) predicts **y**.
					- (Same, but with *ranked* **x** and **y**)	

	y ~ discrete x
	P: Two-sample t-test
	P: Welch's t-test
	N: Mann-Whitney U
		t.test(y_1, y_2, var.equal=TRUE)
		t.test(y_1, y_2, var.equal=FALSE)
		wilcox.test(y_1, y_2)
			lm(y ~ 1 + G_2)[A]
			gls(y ~ 1 + G_2, weights=…[B])[A]
			lm(signed_rank(y) ~ 1 + G_2)[A]
				✓
				✓
				for N >11
					An intercept for **group 1 (**plus a difference if **group 2**) predicts **y**.
					- (Same, but with one variance *per group* instead of one common.)
					- (Same, but it predicts the *signed rank* of **y**.)	

Multiple regression: lm(y ~ 1 + x_1 + x_2 +…)	P: One-way ANOVA
	N: Kruskal-Wallis
		aov(y ~ group)
		kruskal.test(y ~ group)
			lm(y ~ 1 + G_2 + G_3 +…+ G_N)[A]
			lm(rank(y) ~ 1 + G_2 + G_3 +…+ G_N)[A]
				✓
				for N >11
					An intercept for **group 1** (plus a difference if group ≠ 1) predicts **y**.
					- (Same, but it predicts the *rank* of **y**.)	

	P: One-way ANCOVA	aov(y ~ group + x)	lm(y ~ 1 + G_2 + G_3 +…+ G_N + x)[A]	✓	- (Same, but plus a slope on **x**.) *Note: this is discrete AND continuous. ANCOVAs are ANOVAs with a continuous x.*	
	P: Two-way ANOVA	aov(y ~ group * sex)	lm(y ~ 1 + G_2 + G_3 +…+ G_N + S_2 + S_3 +…+ S_K + G_2*S_2 + G_3*S_3 + … + G_N*S_K)	✓	Interaction term: changing **sex** changes the **y ~ group** parameters. *Note: $G_{2 \text{ to } N}$ is an indicator (0 or 1) for each non-intercept levels of the **group** variable. Similarly for $S_{2 \text{ to } K}$ for sex. The first line (with G_i) is main effect of group, the second (with S_i) for sex and the third is the **group × sex** interaction. For two levels (e.g. male/female), line 2 would just be "S_2" and line 3 would be S_2 multiplied with each G_i.*	*[Coming]*
	Counts ~ discrete x N: Chi-square test	chisq.test(groupXsex_table)	**Equivalent log-linear model** glm(y ~ 1 + G_2 + G_3 + … + G_N + S_2 + S_3+ … + S_K + G_2*S_2 + G_3*S_3 +…+ G_N*S_K, family=…)[A]	✓	Interaction term: (Same as Two-way ANOVA.) *Note: Run glm using the following arguments: `glm(model, family=poisson())`* *As linear-model, the Chi-square test is $\log(y_i) = \log(N) + \log(\alpha_i) + \log(\beta_j) + \log(\alpha_i\beta_j)$ where α_i and β_j are proportions. See more info in the accompanying notebook.*	*Same as Two-way ANOVA*
	N: Goodness of fit	chisq.test(y)	glm(y ~ 1 + G_2 + G_3 +…+ G_N, family=…)[A]	✓	(Same as One-way ANOVA and see Chi-Square note.)	*1W-ANOVA*

List of common parametric (P) non-parametric (N) tests and equivalent linear models. The notation y ~ 1 + x is R shorthand for y = 1·b + a·x which most of us learned in school. Models in similar colors are highly similar, but really, notice how similar they *all* are across colors! For non-parametric models, the linear models are reasonable approximations for non-small sample sizes (see "Exact" column and click links to see simulations). Other less accurate approximations exist, e.g., Wilcoxon for the sign test and Goodness-of-fit for the binomial test. The signed rank function is `signed_rank = function(x) sign(x) * rank(abs(x))`. The variables G_i and S_i are "dummy coded" indicator variables (either 0 or 1) exploiting the fact that when Δx = 1 between categories the difference equals the slope. Subscripts (e.g., G_2 or y_1) indicate different columns in data. lm requires long-format data for all non-continuous models. All of this is exposed in greater detail and worked examples at https://lindeloev.github.io/tests-as-linear.

[A] See the note to the two-way ANOVA for explanation of the notation.

[B] Same model, but with one variance per group: `gls(value ~ 1 + G₂, weights = varIdent(form = ~1|group), method="ML")`.

Why do we need statistics?

“Statistics is the science of collecting, summarizing, presenting and interpreting data, and of using them to estimate the magnitude of associations and test hypotheses”

Kirkwood and Sterne p. 1

The build-up of a research project

- **Planning**
- **Design**
- **Execution** (data collection)
- **Data analysis**
- **Presentation**
- **Interpretation**
- **Publication**

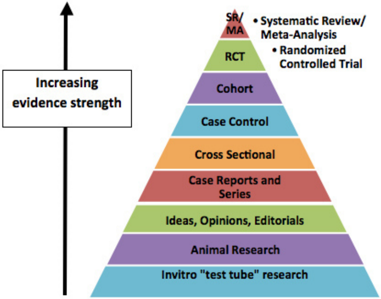
Statistics in all points

Critical reading of publications

- Research **design**
- **Inclusion and exclusion criteria**
- **Sample size**
- **Exposure** (risk factor) and **confounding factors**
- **Outcome** (response)
- Statistical **analysis**
- **Bias**
- **Interpretation** of results

Statistics in all points

Pyramid of evidence


Increasing evidence strength
SR/ MA
• Systematic Review/ Meta-Analysis
RCT
• Randomized Controlled Trial
Cohort
Case Control
Cross Sectional
Case Reports and Series
Ideas, Opinions, Editorials
Animal Research
Invitro "test tube" research

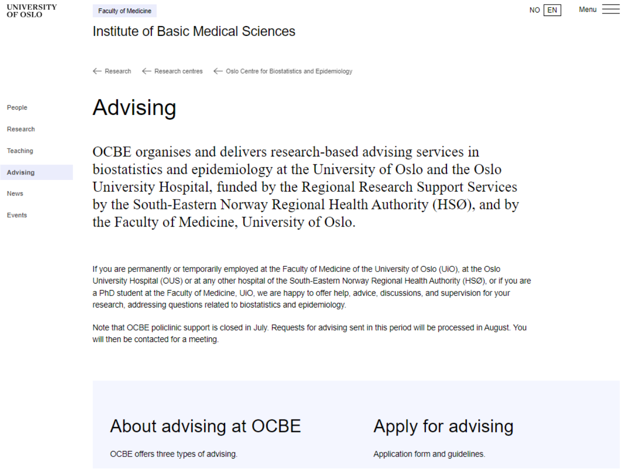
- **Grading the evidence** for practice guidelines after susceptibility of threats to internal validity
- **Health literacy guide** designed to help students find and assess sources of quality health information: `https://libraryguides.unh.edu/c.php?g=326606&p=2191225`

Oslo Centre for Biostatistics and Epidemiology (OCBE)

- ... is a joint centre of UiO (Department of Biostatistics, IMB) and OUS (Biostatistics and Epidemiology group at Forskningsstøtte). Approx. 80 people in total.

- **Research**: Methodological research in several areas, e.g.
  - Statistics for High-dimensional Data
  - Causal Inference
  - Clinical and Cancer Epidemiology
  - Hybrid Modeling / Mathematical Oncology
  - Infectious diseases
  - Clinical Trials Unit (CTU)
- **Statistical advising** for researchers at the Medical Faculty, OUS and Helse-Sør-Øst
- **Teaching** at MedFak: professional study programme for Medicine, Master's programmes in Clinical Nutrition and International Community Health, PhD courses

OCBE Statistical Advising Service



Institute of Basic Medical Sciences

Advising

OCBE organises and delivers research-based advising services in biostatistics and epidemiology at the University of Oslo and the Oslo University Hospital, funded by the Regional Research Support Services by the South-Eastern Norway Regional Health Authority (HSØ), and by the Faculty of Medicine, University of Oslo.

If you are permanently or temporarily employed at the Faculty of Medicine of the University of Oslo (UiO), at the Oslo University Hospital (OUS) or at any other hospital of the South-Eastern Norway Regional Health Authority (HSØ), or if you are a PhD student at the Faculty of Medicine, UiO, we are happy to offer help, advice, discussions, and supervision for your research, addressing questions related to biostatistics and epidemiology.

Note that OCBE policlinic support is closed in July. Requests for advising sent in this period will be processed in August. You will then be contacted for a meeting.

About advising at OCBE

OCBE offers three types of advising.

Apply for advising

Application form and guidelines.

https://www.med.uio.no/imb/english/research/centres/ocbe/advising/