Methods 3: The Philosophy of Multilevel Statistical Modeling and Machine Learning

Week 37: *Introduction*
September 9, 2024

Outline

1) What you already know (Methods 1&2)
2) Motivation for multilevel modelling
3) Motivation for machine learning
4) Presenting the instructor and myself
5) Academic regulations and exam

RATIONALE OF METHODS COURSES

Providing you <u>flexible</u> tools for estimating models

1) What you already know (Methods 1&2)

The foundations we build upon

← COURSE CATALOGUE

Methods 1: Introduction to Experimental Methods, Statistics, and Programming

Autumn semester 2023

The course includes an introduction to experimental methods (including experimental design and data collection methods), statistics, and basic programming (using for for example R).

Student's Sleep Data

Description

Data which show the effect of two soporific drugs (increase in hours of sleep compared to control) on 10 patients.

```
data(sleep)
print(sleep)
```

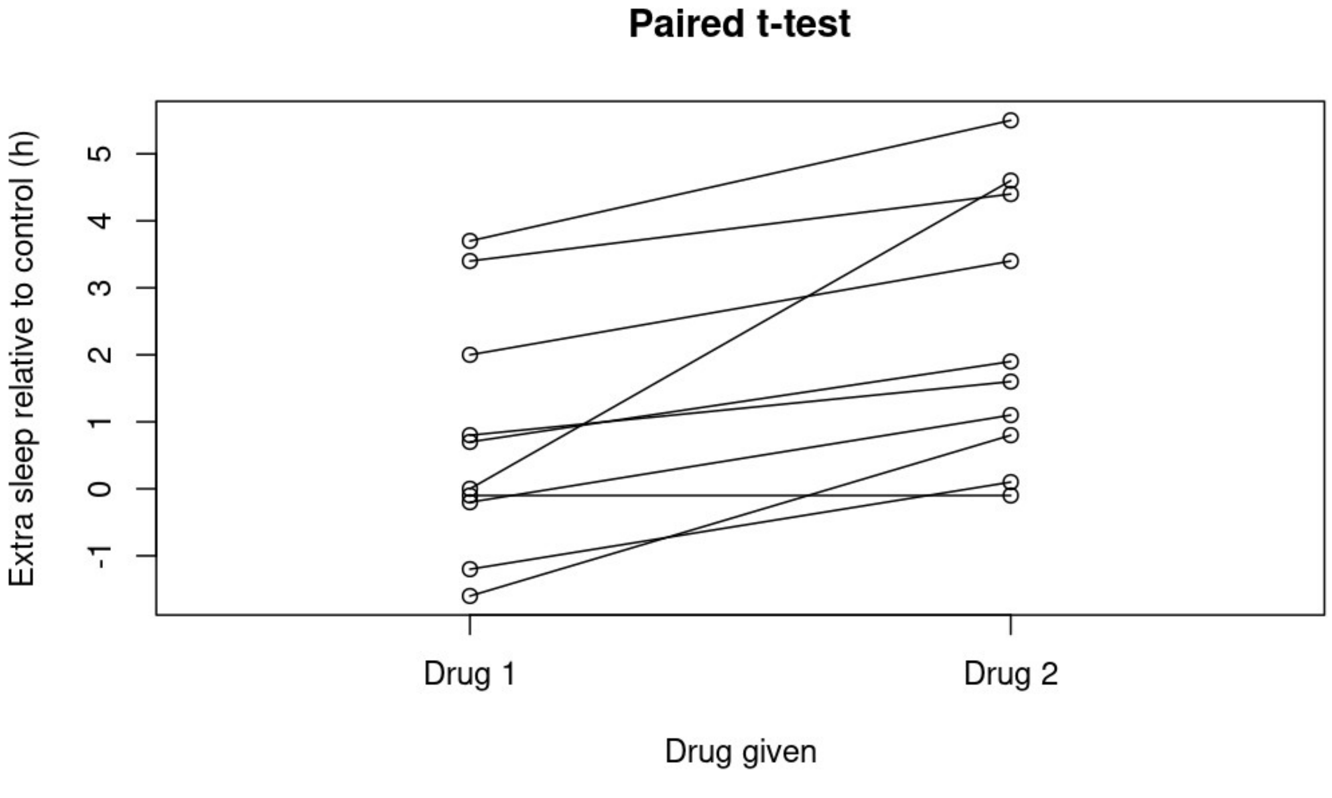
```
##    extra group ID
## 1    0.7     1  1
## 2   -1.6     1  2
## 3   -0.2     1  3
## 4   -1.2     1  4
## 5   -0.1     1  5
## 6    3.4     1  6
## 7    3.7     1  7
## 8    0.8     1  8
## 9    0.0     1  9
## 10   2.0     1 10
## 11   1.9     2  1
## 12   0.8     2  2
## 13   1.1     2  3
## 14   0.1     2  4
## 15  -0.1     2  5
## 16   4.4     2  6
## 17   5.5     2  7
## 18   1.6     2  8
## 19   4.6     2  9
## 20   3.4     2 10
```

The foundations we build upon

```r
print(t.test(sleep$extra[sleep$group == 1], sleep$extra[sleep$group == 2],
             paired = TRUE))
```

```
##
##  Paired t-test
##
## data:  sleep$extra[sleep$group == 1] and sleep$extra[sleep$group == 2]
## t = -4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.4598858 -0.7001142
## sample estimates:
## mean of the differences
##                   -1.58
```

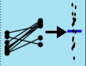
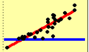
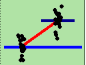
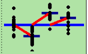
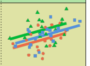
The foundations we build upon



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")`.

Jonas Kristoffer Lindeløv
https://lindeloev.net

Your Swiss Army Knife

After Methods 1

By Cutrofiano - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=6953153

The foundations we build upon

← COURSE CATALOGUE

Methods 2: The General Linear Model

Spring semester 2024

The course includes instruction on the General Linear Model. The course includes theoretical lectures on the GLM and practical exercises in programing.

This course builds on the Methods 1 course and provides the deeper knowledge of statistics required to acquire more advanced statistical methods. The course prepares students for multilevel modelling, machine learning, and Bayesian computational modelling in the Methods 3 and Methods 4 courses.

The General Linear Model is an extremely <u>flexible</u> tool

```r
data(mtcars)
print(mtcars)
```

RStudio: Notebook Output

	mpg	cyl	disp	hp	drat	wt
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Mazda RX4	21.0	6	160.0	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875
Datsun 710	22.8	4	108.0	93	3.85	2.320
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440
Valiant	18.1	6	225.0	105	2.76	3.460
Duster 360	14.3	8	360.0	245	3.21	3.570
Merc 240D	24.4	4	146.7	62	3.69	3.190
Merc 230	22.8	4	140.8	95	3.92	3.150
Merc 280	19.2	6	167.6	123	3.92	3.440
Merc 280C	17.8	6	167.6	123	3.92	3.440
Merc 450SE	16.4	8	275.8	180	3.07	4.070
Merc 450SL	17.3	8	275.8	180	3.07	3.730
Merc 450SLC	15.2	8	275.8	180	3.07	3.780
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250

Description

The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

Format

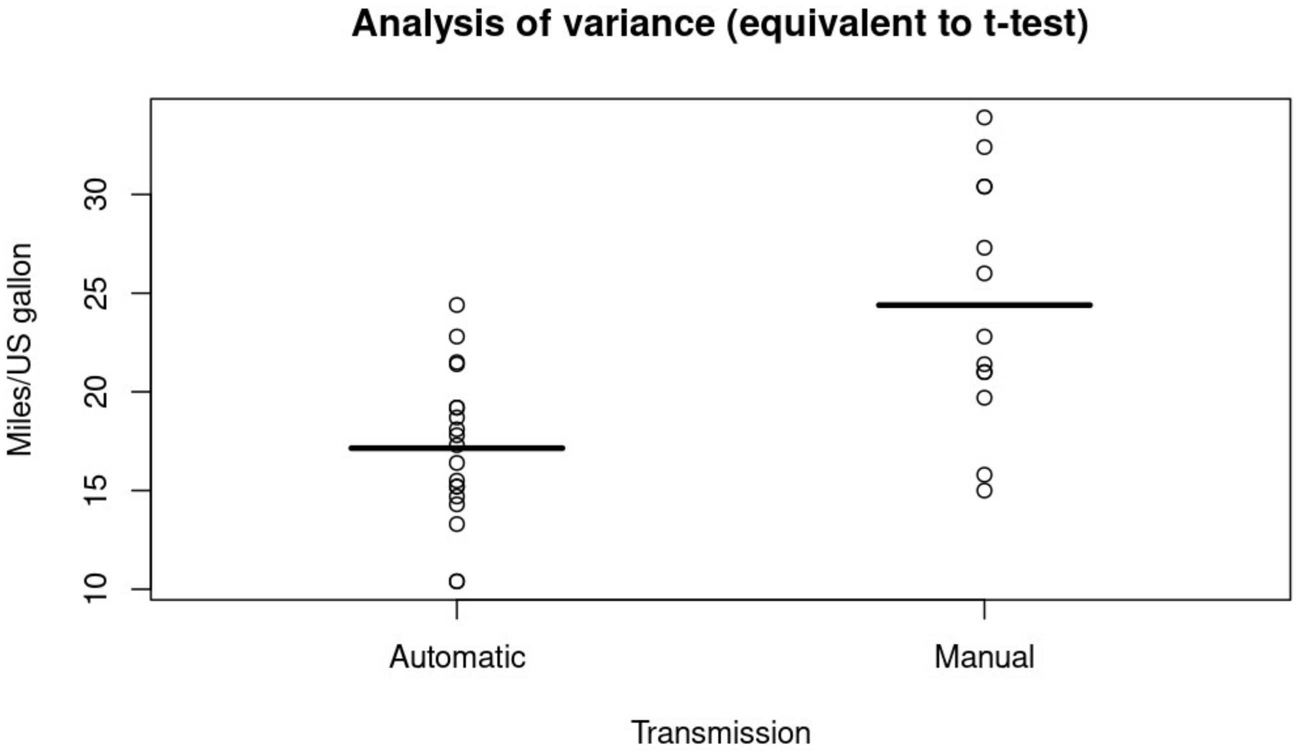
A data frame with 32 observations on 11 (numeric) variables.

[, 1]	mpg	Miles/(US) gallon
[, 2]	cyl	Number of cylinders
[, 3]	disp	Displacement (cu.in.)
[, 4]	hp	Gross horsepower
[, 5]	drat	Rear axle ratio
[, 6]	wt	Weight (1000 lbs)
[, 7]	qsec	1/4 mile time
[, 8]	vs	Engine (0 = V-shaped, 1 = straight)
[, 9]	am	Transmission (0 = automatic, 1 = manual)
[,10]	gear	Number of forward gears
[,11]	carb	Number of carburetors

```
Call:
lm(formula = mpg ~ factor(am),
data = mtcars)
Coefficients:
(Intercept)  factor(am)1
     17.147        7.245
```

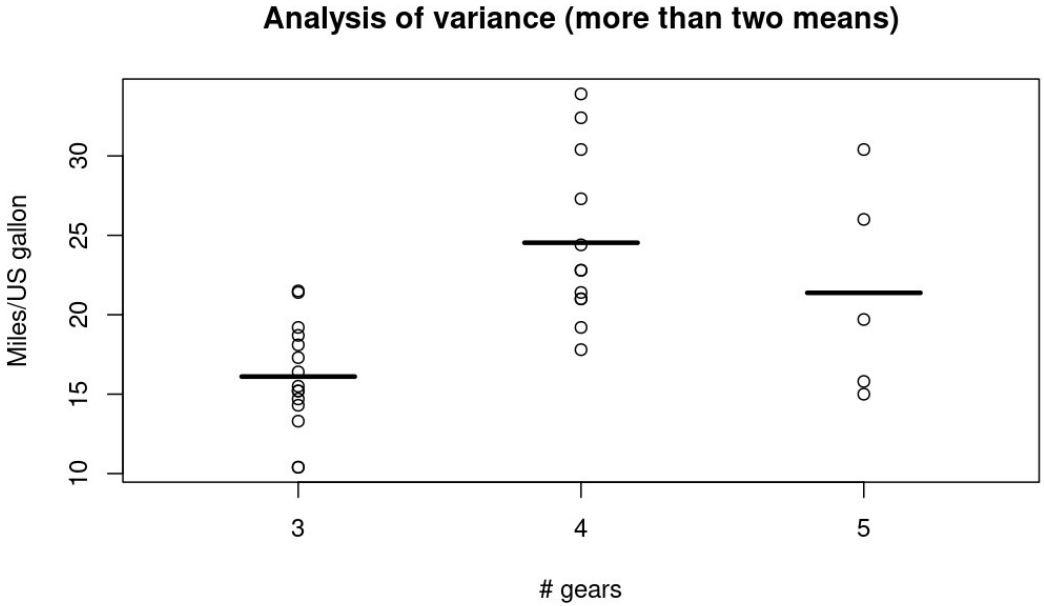
```
Call:
lm(formula = mpg ~ factor(am), data = mtcars)
Coefficients:
(Intercept)  factor(am)1
     17.147        7.245
```

We can do the same as in Methods 1



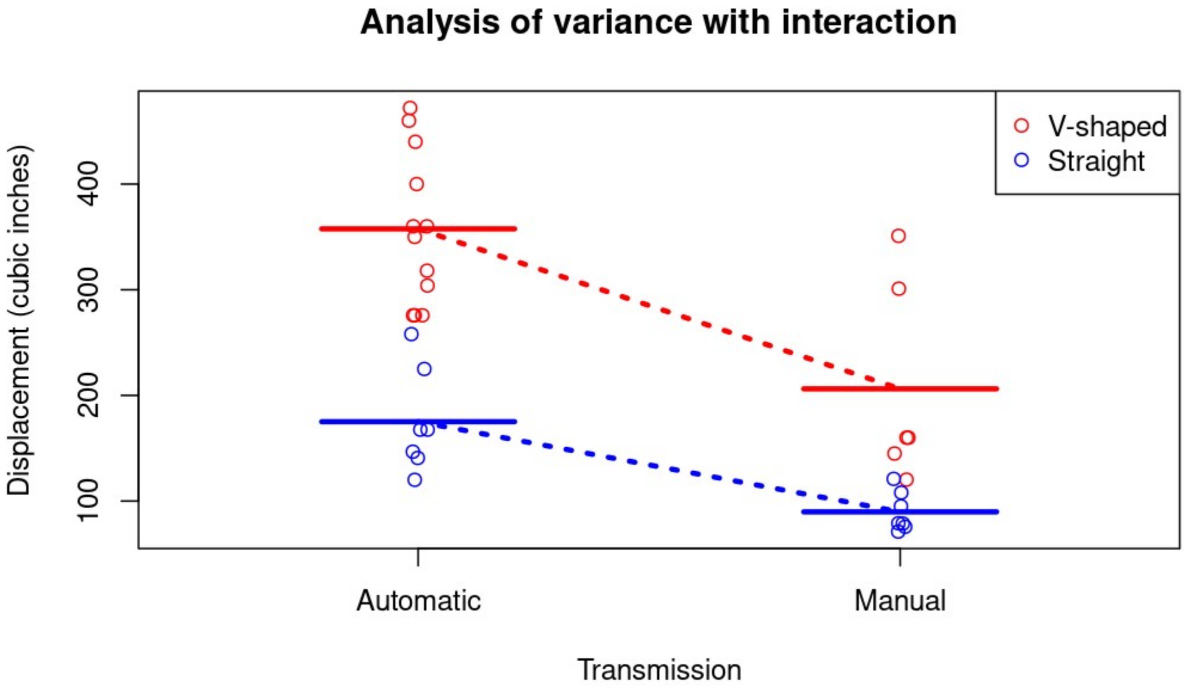
```
Call:
lm(formula = mpg ~ factor(gear), data = mtcars)
Coefficients:
  (Intercept)  factor(gear)4  factor(gear)5
       16.107          8.427          5.273
```

... but we can extend to factors (gear) with more than two levels (3, 4 or gears)



```
Call:
lm(formula = disp ~ factor(vs) * factor(am), data = mtcars)
Coefficients:
          (Intercept)             factor(vs)1              factor(am)1  factor(vs)1:factor(am)1
               357.62                 -182.50                  -151.40                    66.09
```

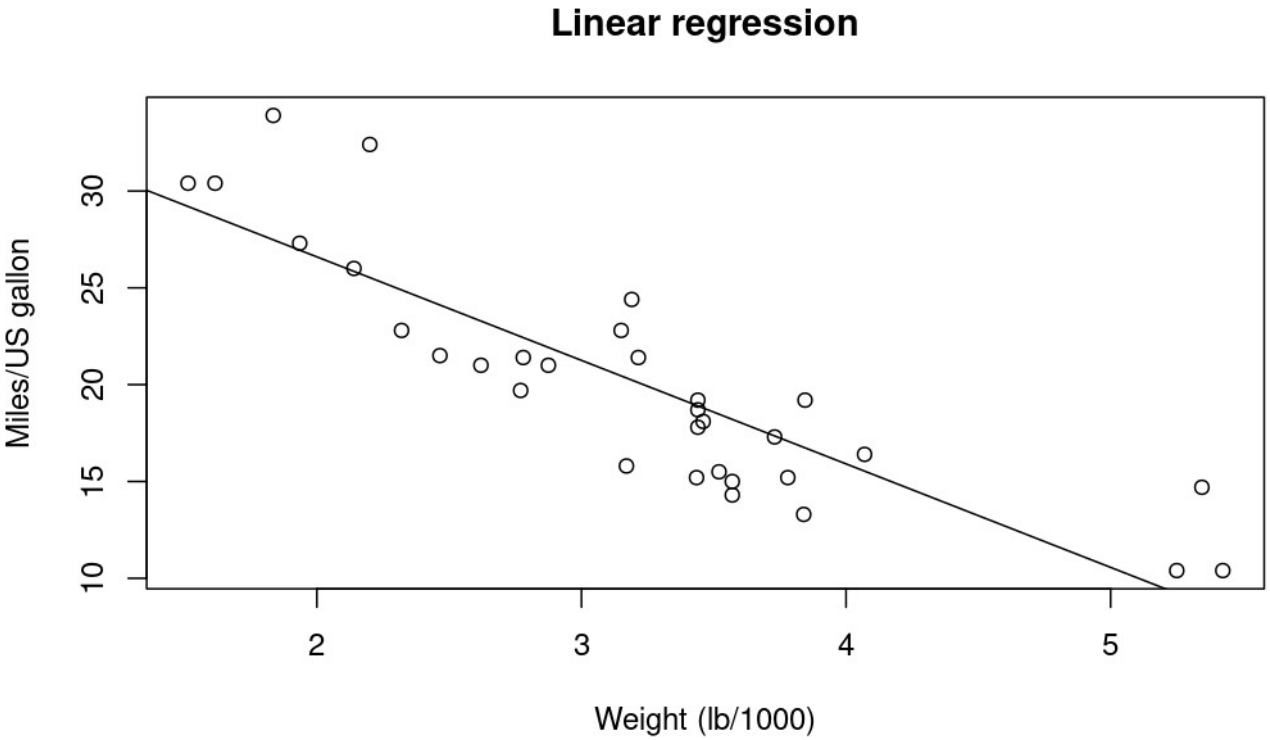
... and we can model interactions, (i.e. one independent variable (vs) depends on another independent variable (am))



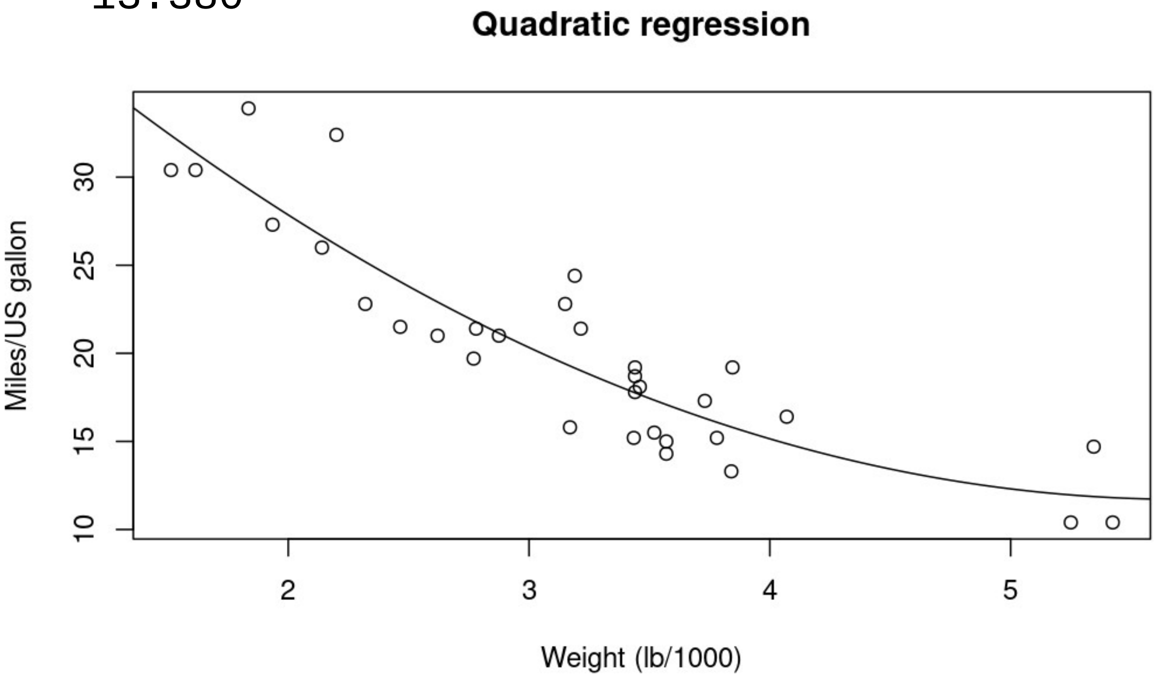
```
Call:
lm(formula = mpg ~ wt, data = mtcars)
Coefficients:
(Intercept)           wt
     37.285       -5.344
```

... and we can subsume linear regression under the general linear model

$$y = ax + b$$



```
Call:
lm(formula = mpg ~ I(wt^2) + wt, data = mtcars)
Coefficients:
(Intercept)      I(wt^2)           wt
     49.931        1.171      -13.380
```



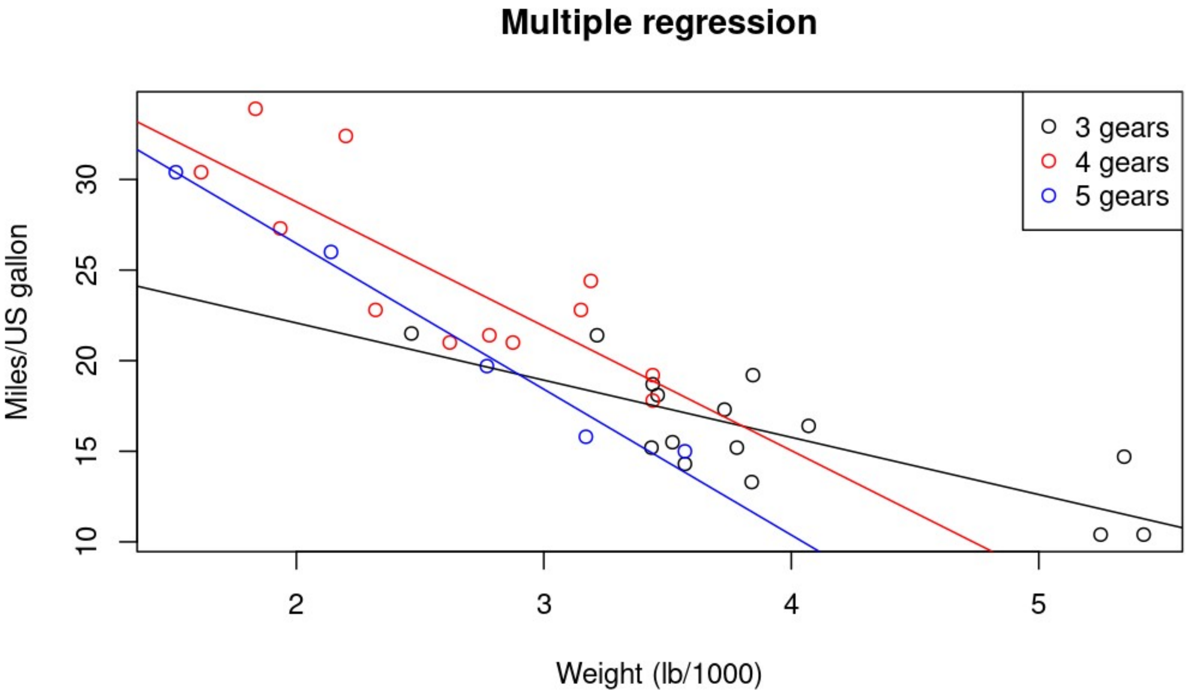
... and quadratic regression

$$y = ax^2 + bx + c$$

```
Call:
lm(formula = mpg ~ wt * factor(gear), data = mtcars)

Coefficients:
     (Intercept)                 wt     factor(gear)4     factor(gear)5  wt:factor(gear)4  wt:factor(gear)5
          28.395             -3.157            14.098            14.168            -3.707            -4.889
```

... and multiple regression



The beauty of the General Linear Model

$$Y = X\beta + \epsilon$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Y: a column vector with n observations (known)

X: the design matrix (known), size: $n \times p$

β: a column vector with p (unknown) model parameters

ϵ: a column vector with n residuals, $N(0, \sigma^2)$

The design matrix: *X*

what is this?

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \\ 1 & x_7 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \end{bmatrix}$$

what is this?

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \end{bmatrix}$$

$$Y = X\beta + \epsilon$$

https://en.wikipedia.org/wiki/Design_matrix

Your Swiss Army Knife

After Methods 2

By D-M Commons - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=20224699

Your Swiss Army Knife

After Methods 3

By CubeBubi (talk) (Uploads) - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=135762122

General multilevel linear model

$$Y = X\beta + Zb + \epsilon$$

$Y:$ a column vector with n observations (known)
$X:$ the first level design matrix (known), size: $n\ x\ p$
$\beta:$ unknown $p\ x\ 1$ column vector of the first level coefficients
$Z: n\ x\ q$ design matrix (known) for the q random effects
$b:$ unknown $q\ x\ 1$ column vector of the second-level coeefficients: $N(0, G)$
$G:$ the covariance matrix of the second-level effects (unknown): $q\ x\ q$
$\epsilon: N\ x\ 1$ column vector of the residuals: $N(0, \sigma^2 I)$

General multilevel linear model

$$Y = X\beta + Zb + \epsilon$$

Now we have two design matrices, *X* and *Z*, but no analytical solution. (What does it mean and entail that we have no analytical solutions?)

Your Swiss Army Knife

After Methods 4

By Slartibartfass - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=19875863

Why is the Swiss Army Knife an apt metaphor?

- Each tool is independent of everything else and works perfectly fine on its own
- Sometimes all you need is a blade (t-test)
- Flexibility comes at a cost (the Methods 4 knife is heavier to wield)

A misconception

Paraphrase of what I have overheard in conversations:

“*In Methods 4 we are going to learn the correct way, the other courses represent the wrong way*”

2) Motivation for multilevel modelling

The foundation we are building

← COURSE CATALOGUE

Methods 3: Multilevel Statistical Modeling and Machine Learning

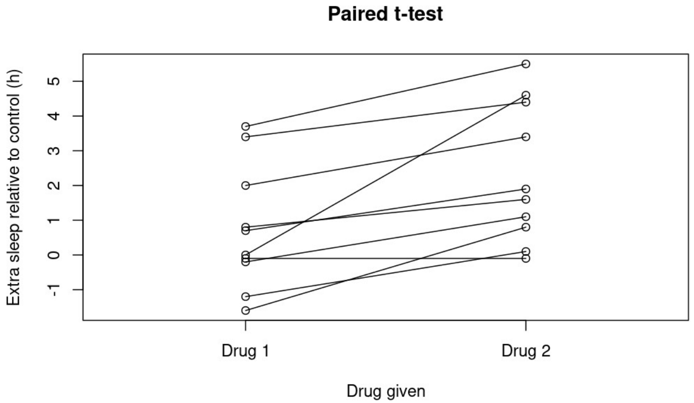
Autumn semester 2021

The course includes instruction on statistical techniques relying on the Generalised Linear Model, including multilevel modelling and basic machine learning procedures (e.g. predictive approaches and validation techniques). The course includes lectures on statistical and machine learning concepts as well as practical exercises in programming.

When might the General Linear Model allow for too little flexibility?

Advantages of multilevel modelling

```
Paired t-test
data:  sleep$extra[sleep$group == 1] and sleep$extra[sleep$group == 2]
t = -4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.4598858 -0.7001142
sample estimates:
mean of the differences
                  -1.58
```



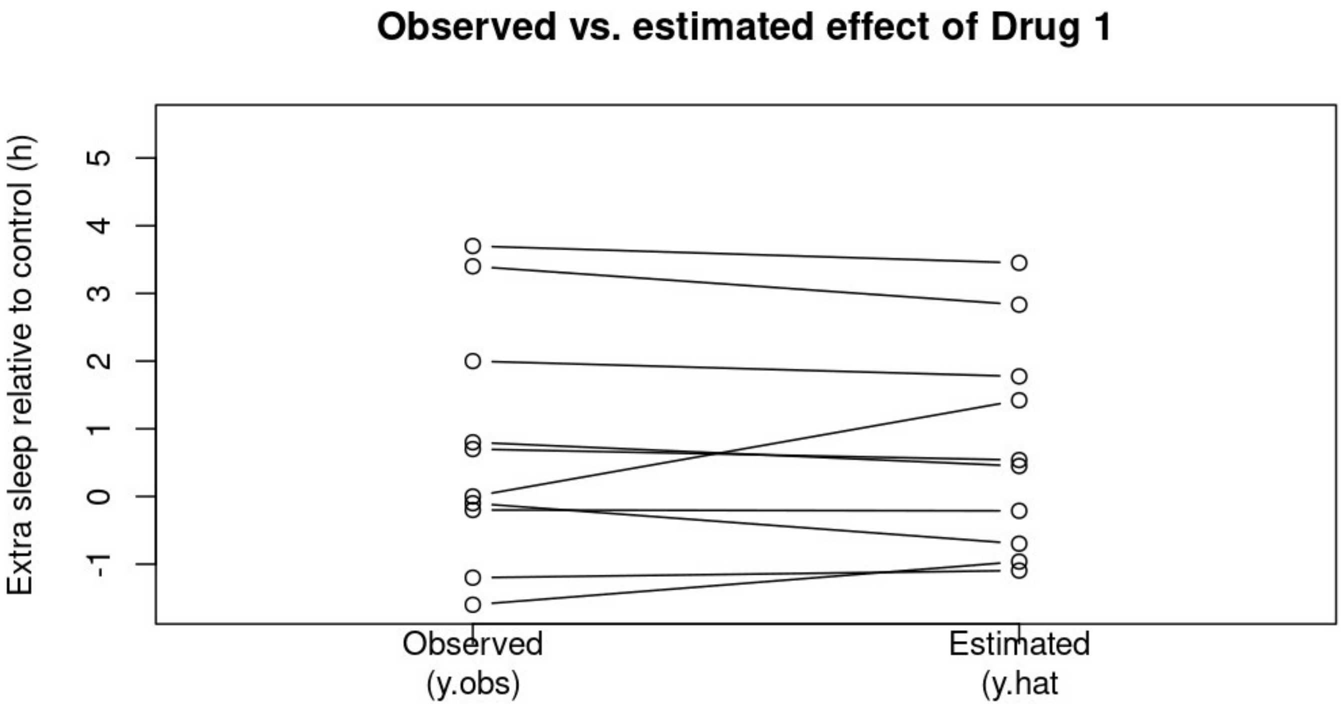

```
Call:
lm(formula = extra ~ factor(group), data =
sleep)

Coefficients:
   (Intercept)  factor(group)2
          0.75            1.58
```

Can explain more variance

```
Linear mixed model fit by REML ['lmerMod']
Formula: extra ~ factor(group) + (1 | ID)
   Data: sleep
REML criterion at convergence: 69.9559
Random effects:
 Groups   Name        Std.Dev.
 ID       (Intercept) 1.6877
 Residual             0.8697
Number of obs: 20, groups:  ID, 10
Fixed Effects:
   (Intercept)  factor(group)2
          0.75            1.58
```

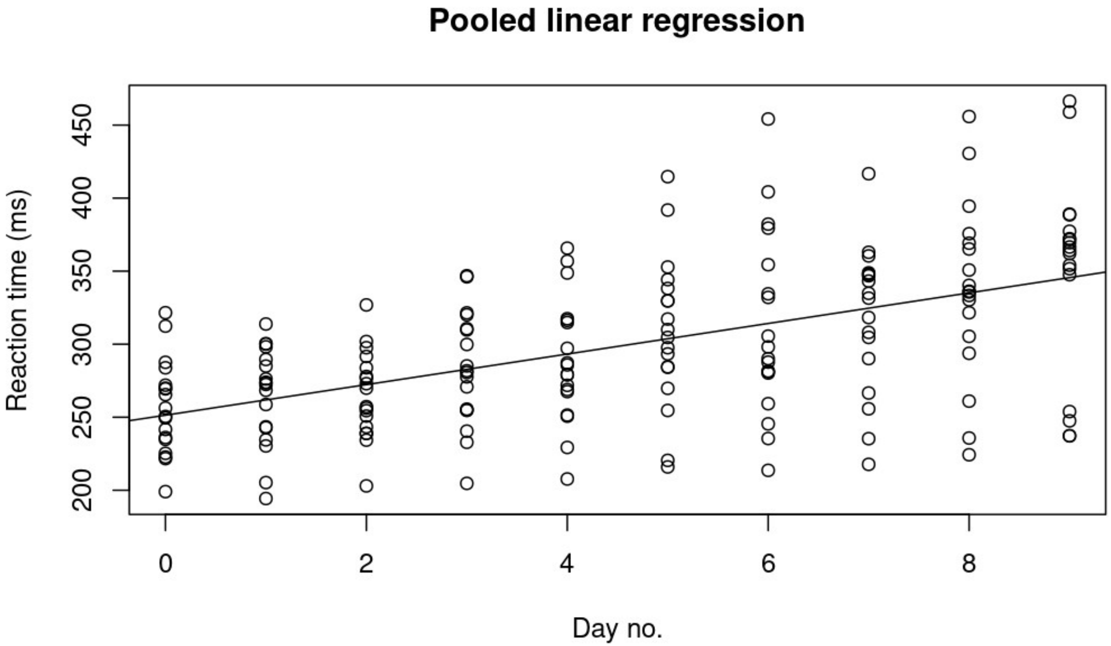
```
> sigma(lm(extra ~ factor(group), data=sleep))
[1] 1.898625
```



Can estimate single subjects

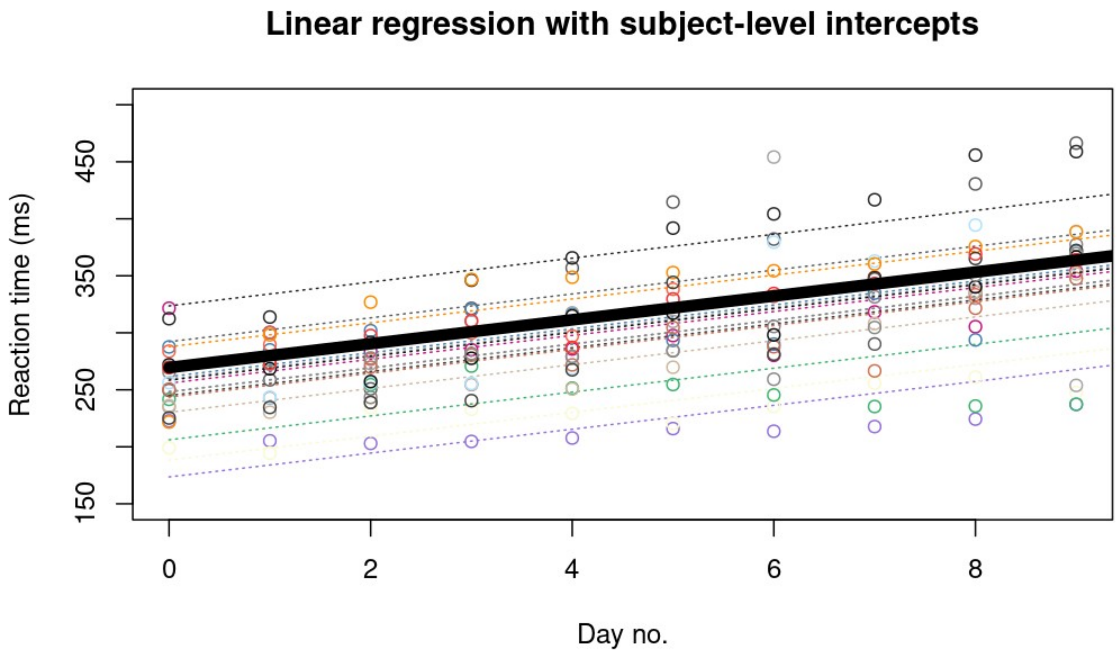
```
> print(str(sleepstudy))
'data.frame':   180 obs. of  3 variables:
 $ Reaction: num  250 259 251 321 357 ...
 $ Days    : num  0 1 2 3 4 5 6 7 8 9 ...
 $ Subject : Factor w/ 18 levels "308","309","310",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
Call:
lm(formula = Reaction ~ Days, data = sleepstudy)
Coefficients:
(Intercept)         Days
     251.41        10.47
```



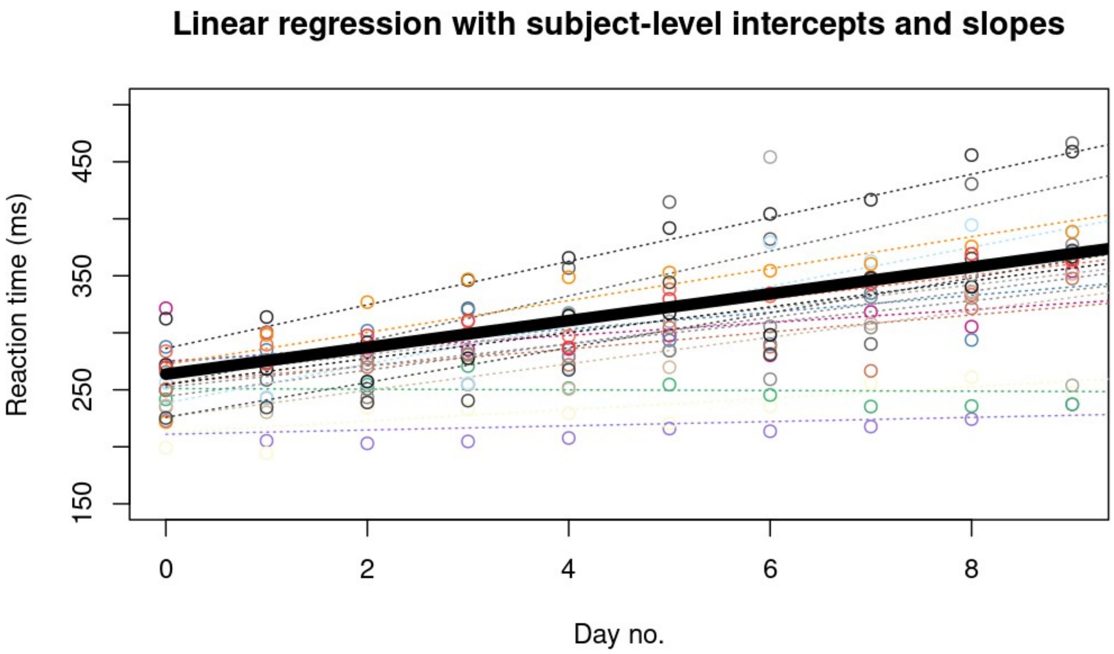
```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (1 | Subject)
   Data: sleepstudy
REML criterion at convergence: 1786.465
Random effects:
 Groups   Name        Std.Dev.
 Subject  (Intercept) 37.12
 Residual             30.99
Number of obs: 180, groups:  Subject, 18
Fixed Effects:
(Intercept)        Days
     251.41       10.47
```

Respects data structure

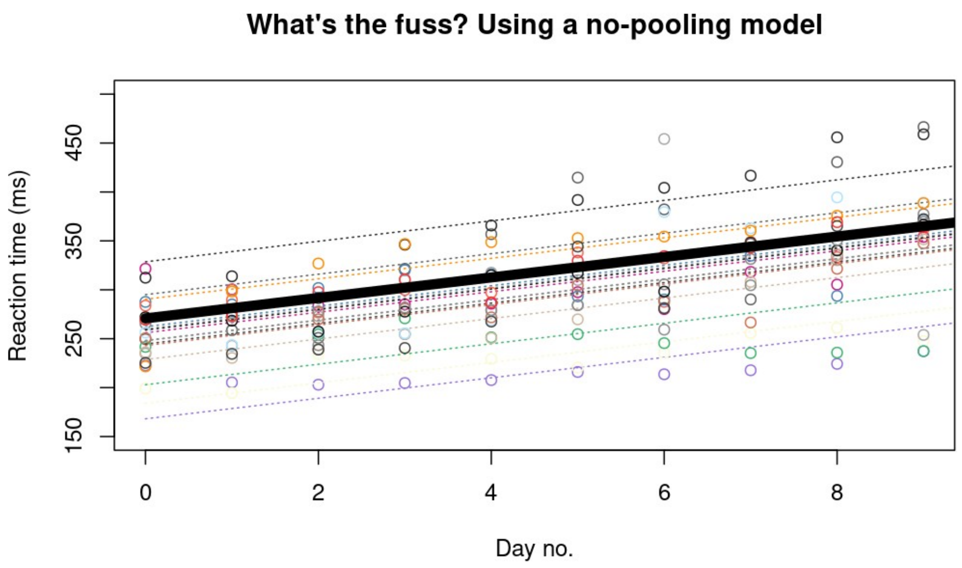


```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (Days | Subject)
   Data: sleepstudy
REML criterion at convergence: 1743.628
Random effects:
 Groups   Name        Std.Dev. Corr
 Subject  (Intercept) 24.741
          Days         5.922   0.07
 Residual             25.592
Number of obs: 180, groups:  Subject, 18
Fixed Effects:
(Intercept)         Days
     251.41        10.47
```

Respects data structure



```
Call:
lm(formula = Reaction ~ Days + Subject - 1, data = sleepstudy)
Coefficients:
      Days  Subject308  Subject309  Subject310  Subject330  Subject331  Subject332  Subject333  Subject334  Subject335  Subject337  Subject349  Subject350
     10.47      295.03      168.13      183.90      256.12      262.33      260.20      269.06      248.20      202.97      328.62      228.73      266.50
Subject351  Subject352  Subject369  Subject370  Subject371  Subject372
    242.99      290.32      258.93      244.60      247.88      270.78
```




```
> mean(model$coefficients[2:length(model$coefficients)])
[1] 251.4051
```

```
Call:
lm(formula = Reaction ~ Days * Subject - 1, data = sleepstudy)
Coefficients:
           Days       Subject308       Subject309       Subject310       Subject330       Subject331       Subject332       Subject333       Subject334       Subject335
         21.765          244.193          205.055          203.484          289.685          285.739          264.252          275.019          240.163          263.035
     Subject337       Subject349       Subject350       Subject351       Subject352       Subject369       Subject370       Subject371       Subject372  Days:Subject309
        290.104          215.112          225.835          261.147          276.372          254.968          210.449          253.636          267.045          -19.503
Days:Subject310  Days:Subject330  Days:Subject331  Days:Subject332  Days:Subject333  Days:Subject334  Days:Subject335  Days:Subject337  Days:Subject349  Days:Subject350
        -15.650          -18.757          -16.499          -12.198          -12.623           -9.512          -24.646           -2.739           -8.271           -2.261
Days:Subject351  Days:Subject352  Days:Subject369  Days:Subject370  Days:Subject371  Days:Subject372
        -15.331           -8.198          -10.417           -3.709          -12.576          -10.467
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (Days | Subject)
   Data: sleepstudy
REML criterion at convergence: 1743.628
Random effects:
 Groups   Name        Std.Dev. Corr
 Subject  (Intercept) 24.741
          Days         5.922   0.07
 Residual             25.592
Number of obs: 180, groups:  Subject, 18
Fixed Effects:
(Intercept)         Days
     251.41        10.47
```

```
> sigma(lm(Reaction ~ Days * Subject - 1, data=sleepstudy))
[1] 25.59182
```

So what’s the fuss?

It flouts the assumption of independent observations!

We only had 18 subjects

```
Call:
lm(formula = Reaction ~ Days, data = sleepstudy)

Residual standard error: 47.71 on 178 degrees of freedom
Multiple R-squared:  0.2865,    Adjusted R-squared:  0.2825
F-statistic: 71.46 on 1 and 178 DF,  p-value: 9.894e-15
```

The GLM does not handle unbalanced datasets well

```
Call:
lm(formula = Reaction ~ Days, data = sleepstudy.NA)
Coefficients:
(Intercept)         Days
     252.69        10.31
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (1 | Subject)
   Data: sleepstudy.NA
REML criterion at convergence: 1593.928
Random effects:
 Groups   Name        Std.Dev.
 Subject  (Intercept) 36.82
 Residual             31.43
Number of obs: 160, groups:  Subject, 18
Fixed Effects:
(Intercept)         Days
     248.95        10.86
```

Reason: GLM weights each group (subject in this case), equally despite different sample sizes

Experiments have many repetitions

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

ELSEVIER

Check for updates

Visual expectations change subjective experience without changing performance

Lau Møller Andersen[a, b, *], Morten Overgaard[b], Frank Tong[c]

[a] *NatMEG, Department of Clinical Neuroscience, Karolinska Institutet, Nobels väg 9, 171 77 Stockholm, Sweden*
[b] *Cognitive Neuroscience Research Unit (CNRU), Center of Functionally Integrative Neuroscience (CFIN), Aarhus University, Aarhus, Denmark*
[c] *Psychology Department, Vanderbilt University, Nashville, TN, USA*

Classical strategy – Aggregation

```
str(data)
'data.frame':   25044 obs. of  18 variables:
 $ trial.type     : chr  "experiment" "experiment" "experiment" "experiment" ...
 $ pas            : Factor w/ 4 levels "1","2","3","4": 1 1 1 3 3 4 1 1 1 3 ...
 $ trial          : int  0 1 2 3 4 5 6 7 8 9 ...
 $ jitter.x       : num  -0.0608 -0.4088 0.0546 -0.4474 -0.2459 ...
 $ jitter.y       : num  0.1038 0.0166 0.4703 -0.4389 -0.4155 ...
 $ odd.digit      : Factor w/ 4 levels "3","5","7","9": 3 3 2 3 3 3 1 3 2 3 ...
 $ target.contrast: num  0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 ...
 $ target.frames  : int  3 2 1 5 6 4 3 1 2 6 ...
 $ cue            : Factor w/ 36 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ task           : chr  "quadruplet" "quadruplet" "quadruplet" "quadruplet" ...
 $ target.type    : chr  "odd" "odd" "even" "even" ...
 $ rt.subj        : num  1.007 0.879 0.735 1.135 1.055 ...
 $ rt.obj         : num  0.606 0.69 1.029 0.695 0.849 ...
 $ even.digit     : Factor w/ 4 levels "2","4","6","8": 3 3 4 2 3 1 2 1 4 2 ...
 $ seed           : int  93764 93764 93764 93764 93764 93764 93764 93764 93764 93764 ...
 $ obj.resp       : chr  "o" "e" "e" "e" ...
 $ subject        : Factor w/ 29 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ correct        : logi  TRUE FALSE TRUE TRUE TRUE TRUE ...
```

aggregated.data <- aggregate(rt.obj ~ subject + correct, data=data, mean)

Why is the mean used?

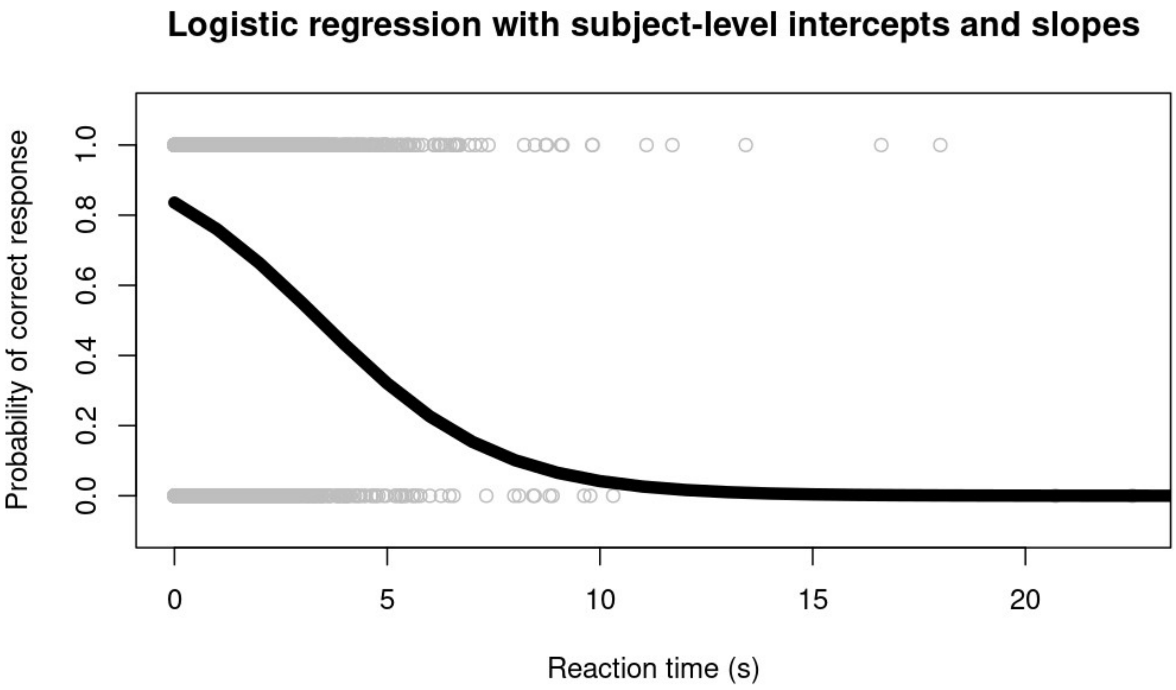
Aggregated data

```
> str(aggregated.data)
'data.frame':   58 obs. of  3 variables:
 $ subject: Factor w/ 29 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ correct: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ rt.obj : num  0.673 0.586 0.882 0.785 1.178 ...
```

```
Call:
lm(formula = rt.obj ~ correct, data = aggregated.data)
Coefficients:
(Intercept)  correctTRUE
     0.9338      -0.1305
```

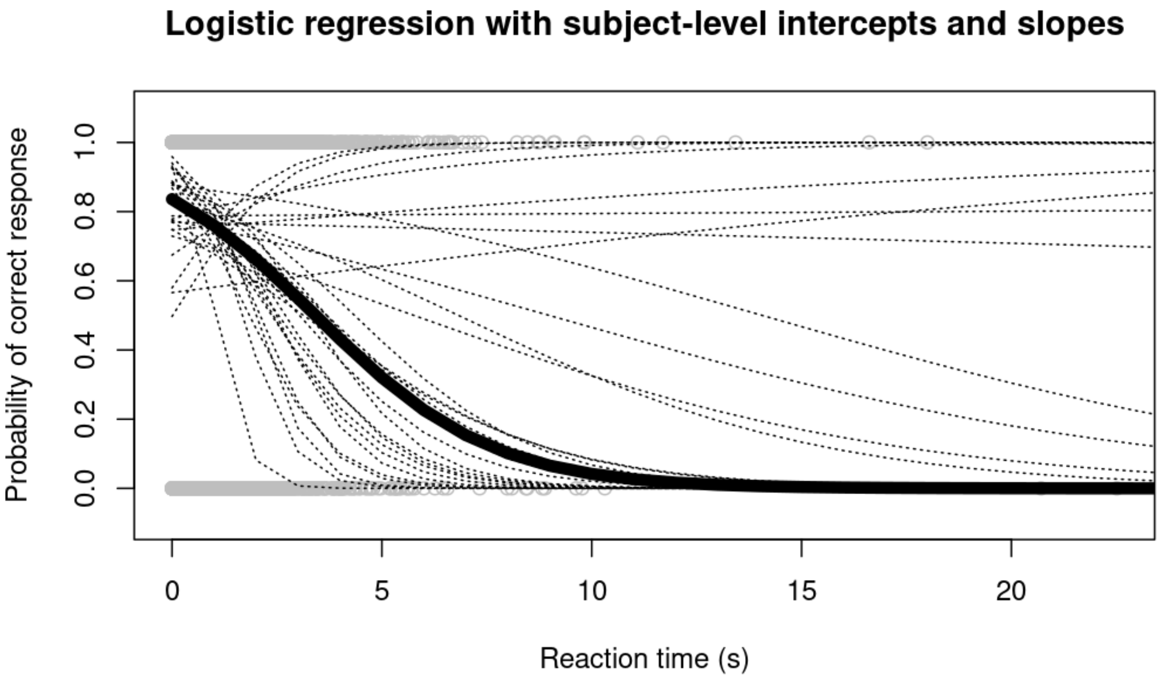
```
Linear mixed model fit by REML ['lmerMod']
Formula: rt.obj ~ correct + (1 | subject)
   Data: data
REML criterion at convergence: 51330.79
Random effects:
 Groups   Name        Std.Dev.
 subject  (Intercept) 0.3046
 Residual             0.6721
Number of obs: 25044, groups:  subject, 29
Fixed Effects:
(Intercept)  correctTRUE
     0.9049      -0.1042
```

What about responses that are not continuous?



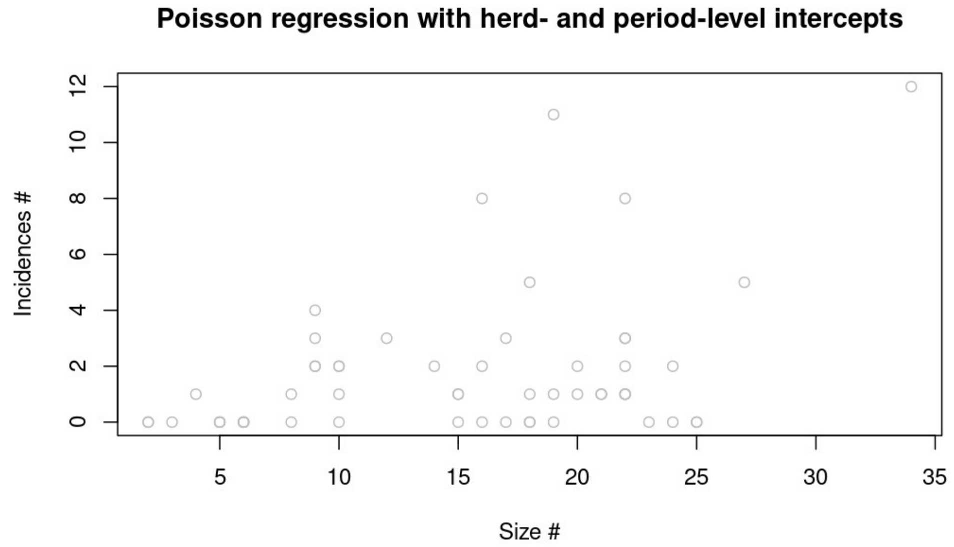

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: correct ~ rt.obj + (rt.obj | subject)
```

The average may mask considerable differences



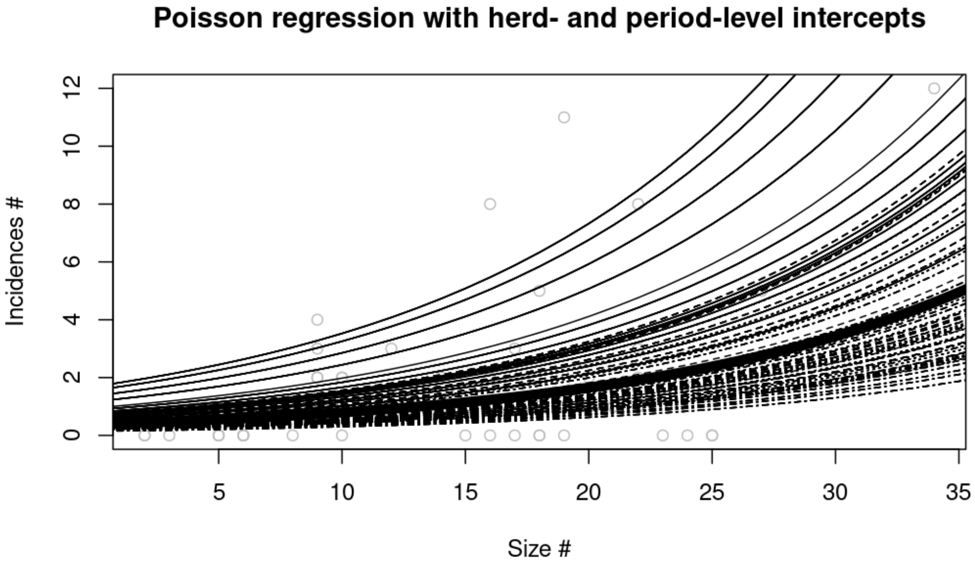
What about responses that are not continuous?

```
> print(str(cbpp))
'data.frame':   56 obs. of  4 variables:
 $ herd      : Factor w/ 15 levels "1","2","3","4",..: 1 1 1 1 2 2 2 3 3 3 ...
 $ incidence: num  2 3 4 0 3 1 1 8 2 0 ...
 $ size      : num  14 12 9 5 22 18 21 22 16 16 ...
 $ period    : Factor w/ 4 levels "1","2","3","4": 1 2 3 4 1 2 3 1 2 3 ...
```



What about responses that are not continuous?

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: poisson  ( log )
Formula: incidence ~ size + (1 | herd) + (1 | period)
   Data: cbpp
     AIC      BIC   logLik deviance df.resid
201.2789 209.3804 -96.6395 193.2789       52
Random effects:
 Groups Name        Std.Dev.
 herd   (Intercept) 0.4953
 period (Intercept) 0.5137
Number of obs: 56, groups:  herd, 15; period, 4
Fixed Effects:
(Intercept)         size
   -0.96037      0.07312
```



What is common to all approaches covered so far is that they *maximise* likelihood, $P(O \mid M)$, of the observations, O, given the model, M, that you are fitting

Interim summary

- Multilevel modelling handles unbalanced data well
- Multilevel modelling handles non-normal data well
- Multilevel modelling allows one to estimate effects on all levels, e.g. *subject* and *group*
- Multilevel modelling allows for using all the data, not just aggregates, e.g. *mean*, *median* or *mode*

3) Motivation for machine learning

Machine learning

Half-way into the course, we are going to turn away from *estimation* towards *prediction*

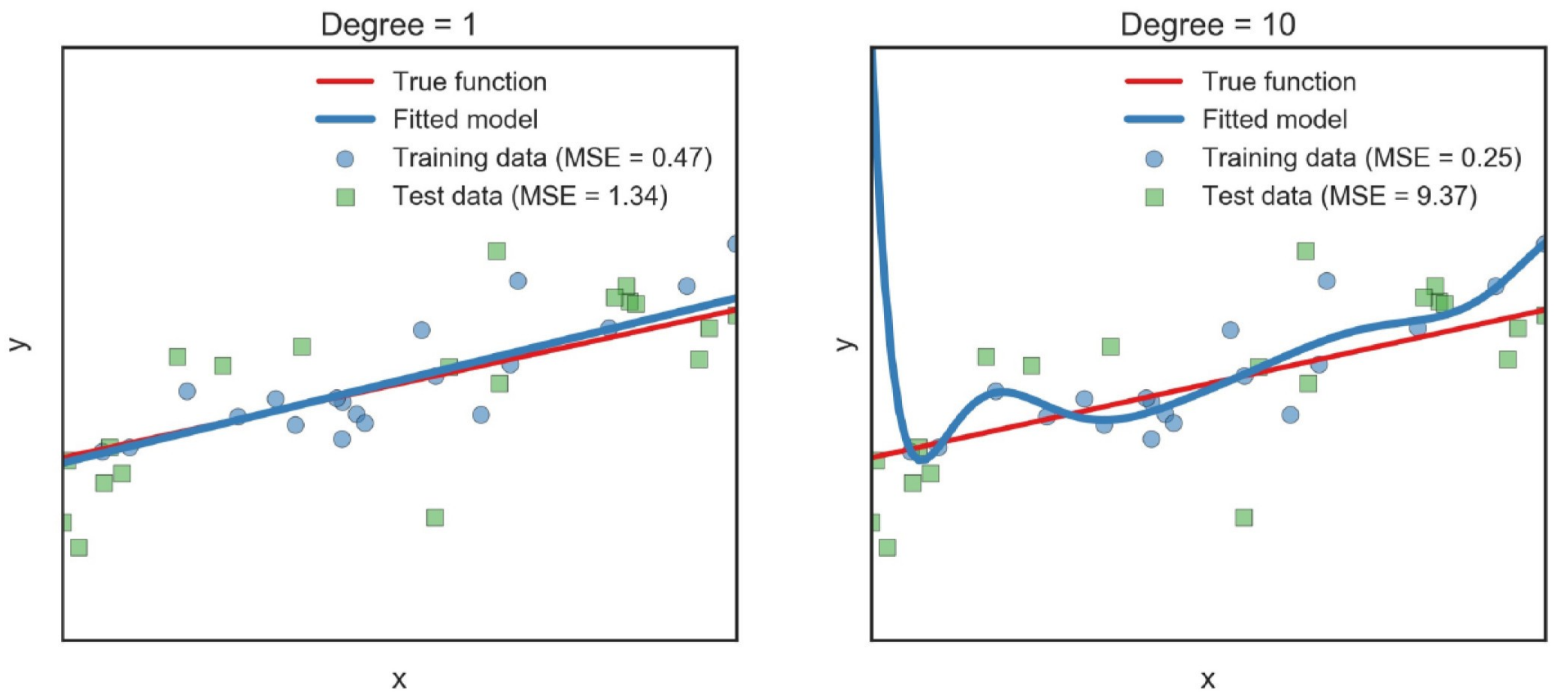
but most of the underlying mechanics are the same, as we are focusing on linear and logistic regression

The spectre of overfitting

"The tendency for statistical models to mistakenly fit sample-specific noise as if it were signal is commonly referred to as overfitting*."*

Yarkoni T, Westfall J (2017) Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci 12:1100–1122.
https://doi.org/10.1177/1745691617693393

The spectre of overfitting



Yarkoni T, Westfall J (2017) Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci 12:1100–1122.
https://doi.org/10.1177/1745691617693393

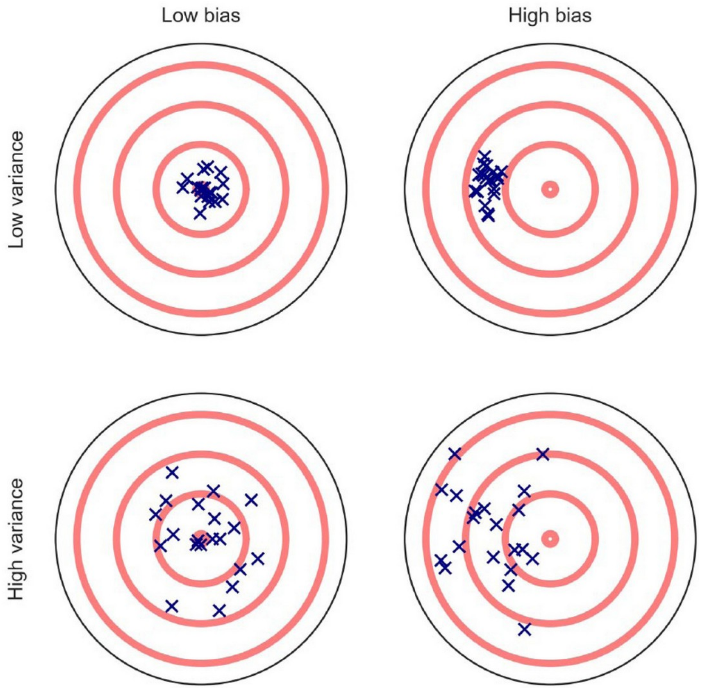
Bias-variance decomposition of error

Accurate and precise

Inaccurate and precise

Accurate and imprecise

Inaccurate and imprecise

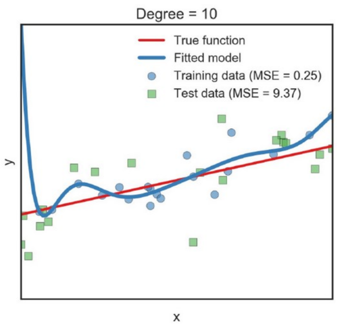


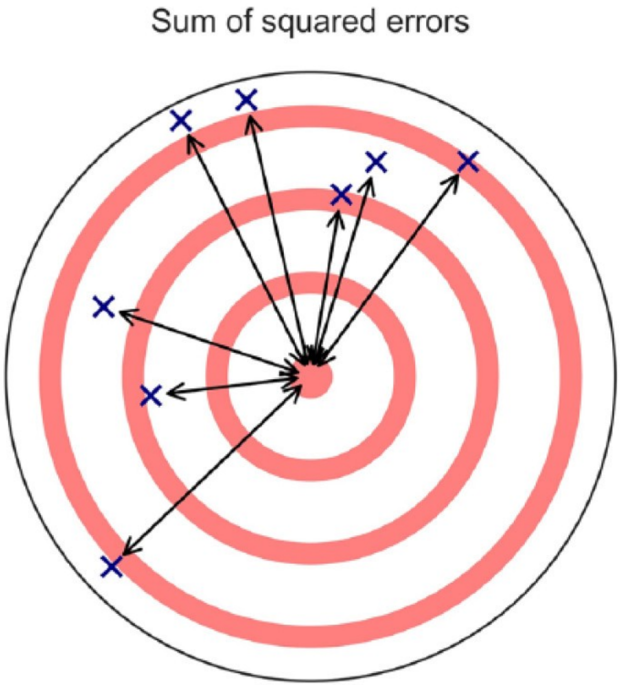
Yarkoni T, Westfall J (2017) Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci 12:1100–1122. https://doi.org/10.1177/1745691617693393

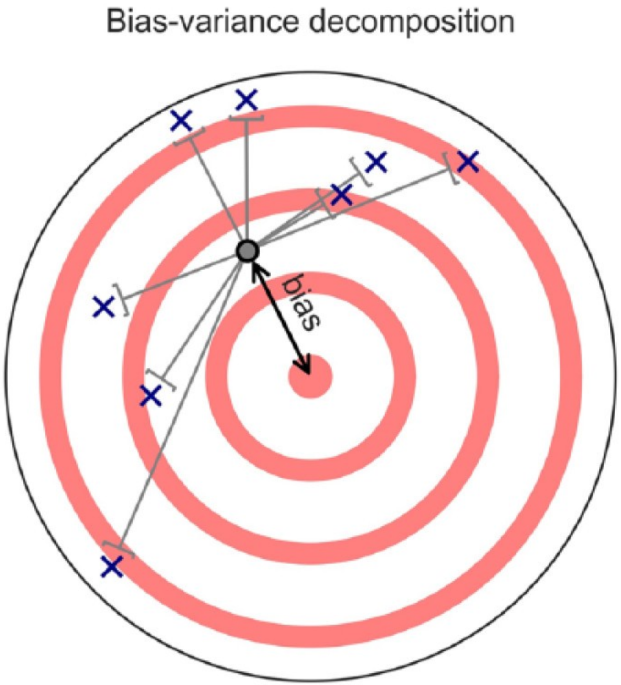
Wilfully introducing bias

WHEN WE DON'T KNOW THE TRUE FUNCTION

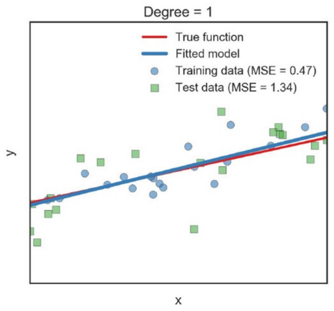
May not generalise well – due to overfitting







May generalise better – due to not fitting noise



Yarkoni T, Westfall J (2017) Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci 12:1100–1122. https://doi.org/10.1177/1745691617693393

Connection between machine learning and multilevel modelling

“[...] *multilevel modeling approaches to analyzing clustered data* [...] *improve on ordinary least squares (OLS) approaches to estimating individual cluster effects by deliberately biasing (through “shrinking” or “pooling”) the cluster estimates toward the estimated population average*”

Yarkoni T, Westfall J (2017) Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci 12:1100–1122. https://doi.org/10.1177/1745691617693393

hint: “ordinary least squares”: think “likelihood maximisation”

Surprising insight

Introducing bias can improve prediction

Interim summary

- Machine learning focuses on *prediction* over *estimation*
- Machine learning includes an estimation phase, fitting the training data, and a prediction phase, seeing how well the fit predicts test data
- In fitting the training data, introducing a bias may provide better prediction of the test data
  - Bias can be introduced by using what is called regularisation
- Multilevel modelling also introduces a bias on the second level, e.g. the subject level
  - This bias shrinks the second level coefficients, away from the solution that explains the most variance

4) Presenting the instructor and myself

Instructor and lecturer

Lydia Pauli Rambrand
BA; 5th semester student

Lau Møller Andersen
laumollerandersen.org

Learning your names

You

- My aim:
  - Every time I address you, I will do it with your name
  - Some times I will be wrong, and that might cause some humiliation both ways
  - Please correct me when I am wrong

Me as a lecturer

Me asking questions

- I will every now and then pick a student at random to answer questions
- I want to emphasise that I am only picking students that I judge are paying attention
  - i.e.: I am not trying to humiliate students not paying attention
- It is completely okay to say “I do not have an answer”; then I’ll just pass it on to someone else
- I will upload slides *after* the lecture

You asking questions

- Does such a thing as a stupid question exist?
- I don’t care if they do or not
- You ask a question because you want to know ...
- ... and I answer because I want to tell you

https://cryptpad.fr/doc/#/2/doc/edit/oEs7GSXz1hpN8IjSOdwS1gVe/

CryptPad

68

Online collaboration

METHODS 3 Q&A

WEEK 37 - what is it all about?

Q: (Lau) how do I ask questions here?

A: (Lau) you've just done it - well done!

Q: how do I ask questions anonymously

A: (Lau) you've just done it - well done!

Q: how do I answer questions?

A:(<student_name>): like this, or anonymously if you prefer. You can also write "student" to highlight that it is an answer given by a fellow student

WEEK 38: Multilevel linear regression

WEEK 39: Beyond linear regression — link functions

WEEK 40: Evaluating generalised linear mixed models — and the spectre of collinearity

WEEK 41: Explanation versus prediction

- Let's create a resource together
  - It will become whatever you make it
- So join in the collaboration

Languages

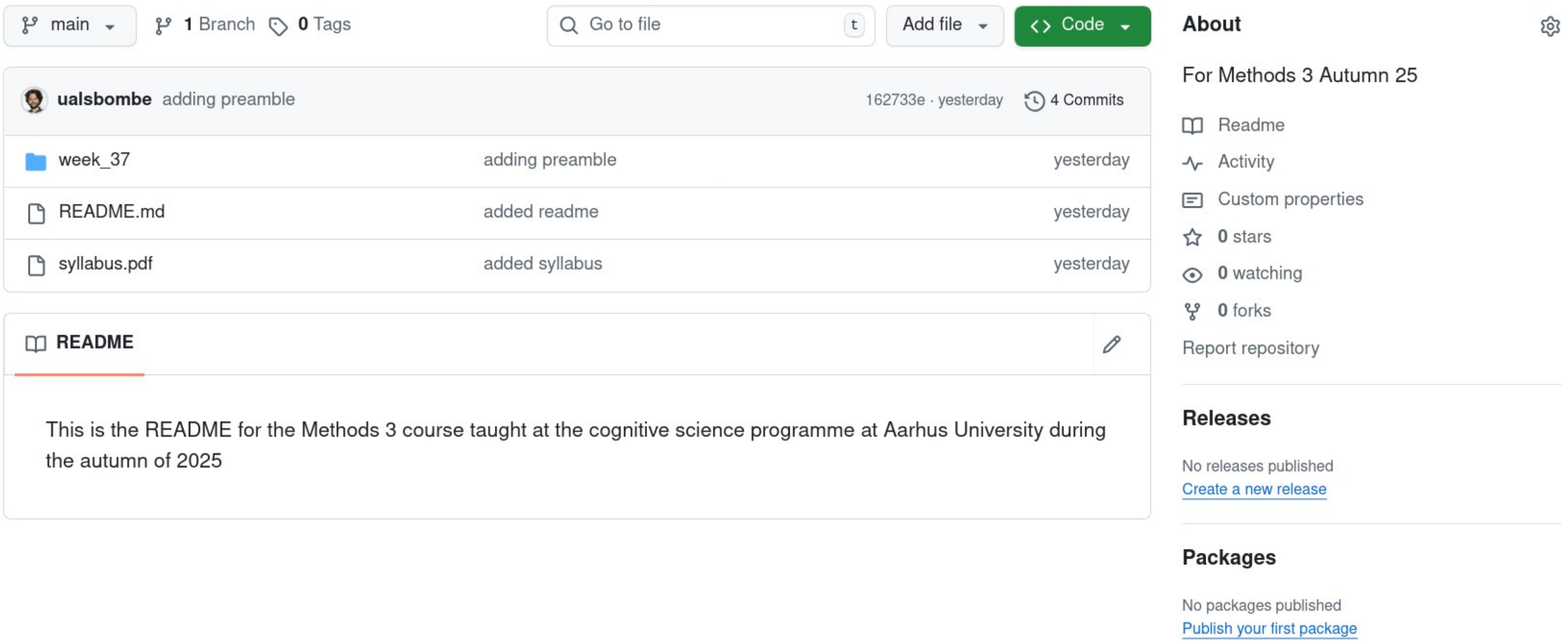
https://cran.r-project.org/mirrors.html

Multilevel models

My recommendation:
https://docs.conda.io/en/latest/miniconda.html

Machine learning

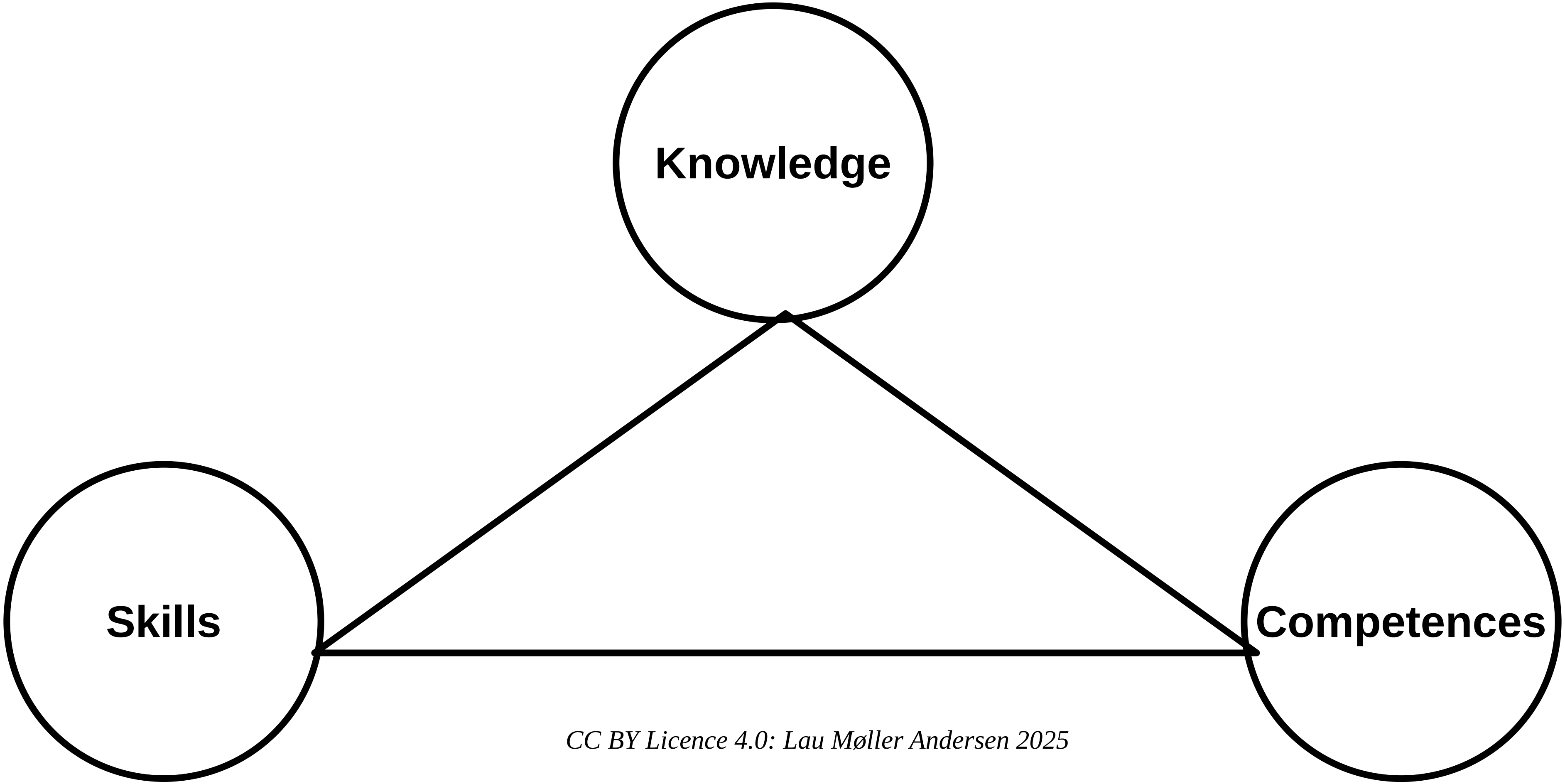
Keeping up to date



https://github.com/Methods-3/Methods-3-F25

5) Academic regulations and exam

Academic regulations – objectives



Academic regulations – objectives

Knowledge:

- demonstrate understanding of statistical techniques relying on the Generalised Linear Model

- demonstrate understanding of hierarchical modeling methods

- demonstrate understanding of basic machine learning concepts.

Academic regulations – KNOWLEDGE

You should thus be able to answer questions like:

What is logistic regression?

What are some naturally occurring hierarchies within research data?

What is cross-validation?

What is a multilevel model?

Academic regulations – objectives

Skills:

- build and evaluate models of hierarchically structured data

- integrate machine learning procedures in data analysis

- communicate analysis processes, results and interpretation.

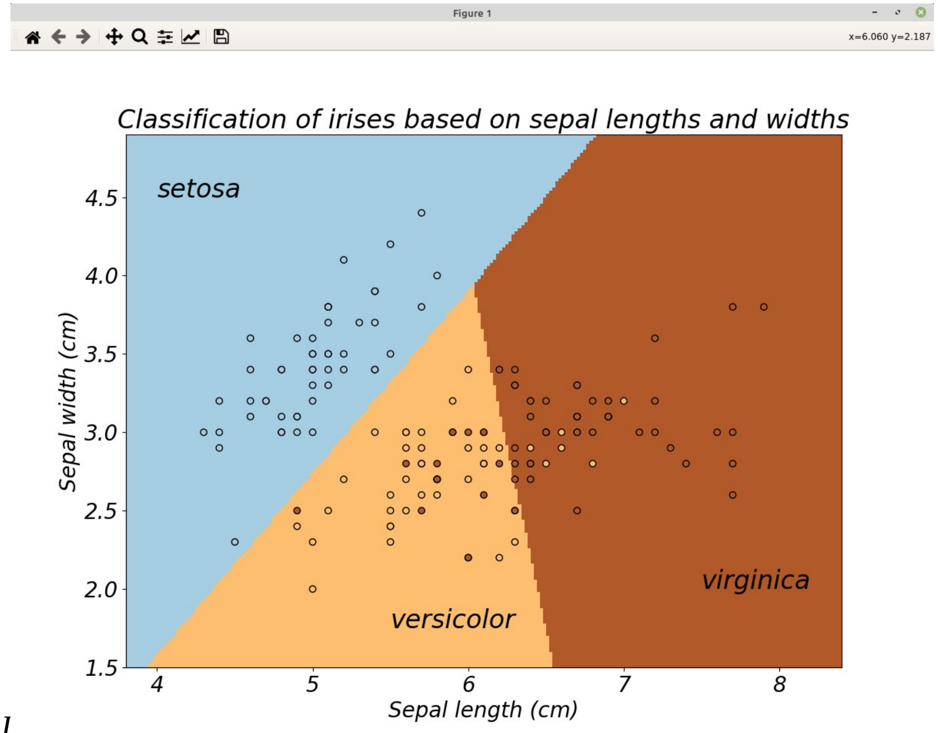
Academic regulations – SKILLS

build and evaluate models of hierarchically structured data

```
Linear mixed model fit by REML ['lmerMod']
Formula: Height ~ Gender + (1 | Family)
   Data: height
```

Academic regulations – SKILLS

integrate machine learning procedures in data analysis



Academic regulations – SKILLS

communicate analysis processes, results and interpretation

(Andersen et al., 2019)

Mixed model analyses (McCulloch & Neuhaus, 2005) were applied to investigate how top-down expectations (No. of Possible Targets) affected subjective experience and objective performance. We performed model comparisons between models that did or did not include the relevant fixed effects and interactions to find the best compromise between an explanatory and a parsimonious model. This was done using the log-likelihood ratio between two models because this ratio approximates a chi-square distribution. A chi-square test can thus be used to assess whether two models differ significantly, where the test statistic is the log-likelihood-ratio and the degrees of freedom is the difference in free parameters of the two models.

Academic regulations – objectives

Competences:

- independently decide on data analysis methods, given a data set and a research question

- justify decision making when pre-processing messy data for data analysis.

Difference between *skills* and *competences:*
Skills are specific
Competences are more generic

Academic regulations – COMPETENCES

independently decide on data analysis methods, given a data set and a research question

justify decision making when pre-processing messy data for data analysis.

EXAMPLE QUESTION: Can native Danish speakers tell apart soft d’s (ð) and l’s if they are embedded in English speech, e.g. maðfunction/malfunction

EXAMPLE DATA: Dataset with response times and discrimination responses

Exam; portfolio

- Ongoing assignments to be solved in study groups
  - I need **knitted** assignments; e.g. *RMarkdown*, *nbconvert*, *Quarto*; hand in the knitted pdf **<u>and</u>** the basis for the knitting, e.g. *.Rmd, .qmd, .ipynb*
- Final portfolio:
  - Revised assignments, handed in as short reports with reproducible code in GitHub Classroom
    - Assignment 0: Using mixed effects modelling to model hierarchical data (Winter & Grawunder, 2012)
    - Assignment 1: Mixed effects modelling of accuracy (Andersen et al., 2019)
    - Assignment 2: Using logistic regression to classify subjective experience from brain data (Sandberg et al., 2016)
    - Assignment 3: Dimensionality reduction; finding the signal among the noise

Re-examination

Will be done during the summer (tip: stay on schedule, such that you can enjoy your summer)

The structure of lectures

- Tuesdays 14.00-16.00 (NB: Academic quarter)
- A mix between the general theoretical framework and a few formulas
- Code will be incorporated
- Group discussions
- Please don’t hold back on your questions

The structure of a practical exercise

- Thursdays 08.00-10.00 and 10.00-12.00
- The practical exercises are led by either Lydia or me.
  - An introduction to the task
  - Data and an instructions file are shared
  - Work in study groups
  - Intermittent discussion of pressing issues
- Deadline:
  - On the day of the next class (23.59). i.e. a bit more than a week.
- Feedback policy
  - Detailed written feedback will be provided for ***Assignment 0 only***, with the aim of aligning expectations of the instructor and students regarding how assignments will generally be read and graded. This ***has to be*** handed in as a ***study group assignment*** – otherwise you do ***NOT*** get feedback. ***Oral feedback*** will be provided on subsequent portfolio assignments during in-class assignment preparation sessions, where students will have ample opportunity to ask the instructor questions on their assignments-in-progress.
  - Note for all assignments; feedback is only given if handed in on time

Student counsellors

Christian: Cognitive Science

Charlotte: Linguistics and Cognitive Semiotics

Sofia: Scandinavian Languages and Literature

Pernille: Experience Economy and Cultur Of Events

https://studerende.au.dk/en/studies/subject-portals/arts/counselling/student-counsellor/nordiskeng

The course plan

Week 37:

Lesson 0: What is it all about?

Class 0: Setting up R and Python — and recollection of the general linear model (Lau)

Week 38:

Lesson 1: Multilevel linear regression

Class 1: Modelling subject level effects – and how do they differ from group level effects? (Lydia)

Week 39:

Lesson 2: Beyond linear regression — link functions

Class 2: What to do when the response variable is not continuous? (Lau)

Week 40:

Lesson 3: Evaluating generalised linear mixed models — and the spectre of collinearity

Class 3: Code review (Lydia)

Week 41:

Lesson 4: Explanation versus prediction

Class 4: Summarising mixed models (Lydia)

Week 42:

no teaching (go fetch potatoes instead)

Week 43:

Lesson 5: Mid-way evaluation and machine learning introduction

Class 5: Getting Python running (Lau)

Week 44:

Lesson 6: Linear and logistic regression revisited (machine learning)

Class 6: Categorizing responses based on informed guesses (Lau)

Week 45:

Lesson 7: Model evaluation and hyperparameter tuning

Class 7: Code review (Lydia: To be moved)

Week 46:

Lesson 8: Dimensionality reduction — principled component analysis (PCA)

Class 8: What to do with very rich data? (Lau)

Week 47:

Lesson 9: Neural networks

Class 9: Code review (Lau)

Week 48:

Lesson 10: Unsupervised classification

Class 10: Create video explainers of central concepts (shared resource) (Lydia)

Week 49:

Lesson 0 again: What was it all about? and final evaluation

Class 11: Summarising machine learning (Lydia)

Week 50:

Student presentations:

Class 12: Work on portfolios: Ask anything! (Lydia & Lau)

Next time – multilevel linear regression

- Design matrix *Z*
- Pooling, no pooling and partial pooling
- How multilevel models are fit
- How bias, not necessarily a bad thing, is introduced in multilevel modelling

Reading questions

- Gelman and Hill:
  - Chapter 11
    - Do the motivations, section 11.5, apply to cognitive science research as you know it?
  - Chapter 12
    - What does pooling mean?
      - What is complete pooling?
      - What is no pooling?
      - What is partial pooling?
        - How is shrinkage achieved in partial pooling?
        - What does it mean that classical regression is a special case of multilevel modelling?
        - Have a good look at equations 12.1 and 12.4
    - How does the discussion about number of groups and number of observations per group relate to cognitive science research as you know it?
- Bates et al. 2015
  - You do not need to understand all the algebra of section 3, but have a look anyway
  - What does it mean that the random variable *Y* in eq. 2, is conditional on the instantiation, ***b***, of the random variable *B*?
  - How is the $||u^2||$ term in eq. 13 related to the bias, shrinkage, introduced in multilevel modelling? (see the machine learning summary above)?