Lecture 5: Generalized Linear Models and Additive models

Wengang Mao (Marine Technology)
Department of Mechanics and Maritime Sciences,
Chalmers University of Technology,
Goteborg, Sweden

1

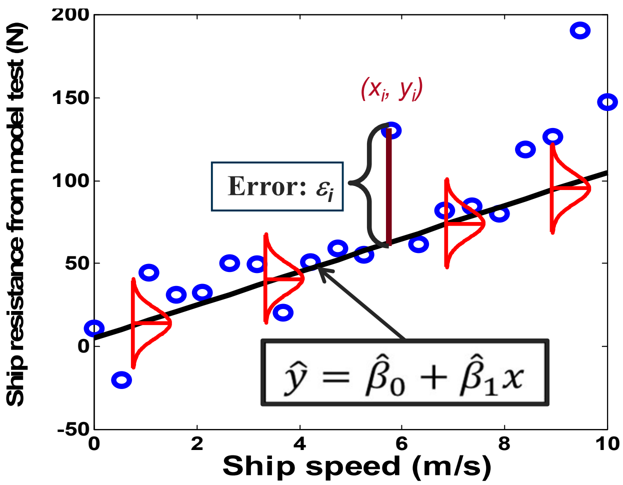
Contents of this lecture

- Assumptions in classical regression
- General formulation of GLM
- Regression of GLM
- Generalized additive models
- Mixed-effects model

2 2023-04-11

2

Assumptions in linear regression



Methods for the estimation

- Maximum likelihood
- (Ordinary) least square
- LASSO (L1 regularization)
- Ridge (L2 regularization)

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \boxed{+\varepsilon_i}$$

3 2023-04-11

3

Assumptions in linear regression

- **There are three assumptions that are required in order to perform a regression analysis.**
  - The error term ε must be zero mean
  - The error term ε must have a constant variance
  - The error term ε must be independent (uncorrelated)
  - (The error term ε is normally distributed or enough sample size is available to rely on large sample theory.)
- **How can we diagnose violations of these conditions?**
  - Residual (ε) Analysis
  - Examine the differences between the actual data points and those predicted by the linear equation
  - It is mainly examined by graphically plots

2023-04-11

4

General formulation of GLM

- Instead of assuming that ***Y*** is a normal distributed RV, ***Y*** can follow a general exponential family distribution (normal, exponential, poission, gamma, binary, logical, etc.),
- The actual meaning of regression is to fit/approximate the mean value of *Y* written as

$$\mu(x) = \mathbb{E}[Y|X = x]$$

- The linear predictor model is written as $\eta(\boldsymbol{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots$. For simplicity, let,

$$\eta(x) = \beta_0 + x \cdot \beta$$

General formulation of GLM

- Then it requires a link function g such that

$$\eta(x) = g(\mu(x)).$$

- Different distributions of $Y = \mu(x)$ will have different link function g between the predictor and mean of *Y*.
- Another important feature in GLM is the dispersion scale function *V* such that

$$\mathbb{V}[Y|X = x] = \sigma^2 V(\mu(x))$$

Motivation of GLM: the link functions

- The exponential distribution family with probability density function:

$$f_\theta(y) = \exp\left[\{y\theta - b(\theta)\}/a(\phi) + c(y,\phi)\right],$$

- Here $b'(\theta)$ is often refer to the mean of Y, $\alpha(\phi)$ is referred to the variance,
- For example, for the normal distribution, we have,

$$\begin{aligned} f_\mu(y) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \\ &= \exp\left[\frac{-y^2 + 2y\mu - \mu^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})\right] \\ &= \exp\left[\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})\right], \end{aligned}$$

which is of exponential form, with $\theta = \mu$, $b(\theta) = \theta^2/2 \equiv \mu^2/2$, $a(\phi) = \phi = \sigma^2$ and $c(\phi, y) = -y^2/(2\phi) - \log(\sqrt{\phi 2\pi}) \equiv -y^2/(2\sigma^2) - \log(\sigma\sqrt{2\pi})$. Table 3.1 gives a

Link functions to various distributions

- Then the link function between μ = *E[Y]* and linear predictor $\beta_0 + X\beta$ can be linked by function *g*(μ) through the parameter $\theta = g(\mu)$ in the probability density function as follows.

	Normal	Poisson	Binomial	Gamma	Inverse Gaussian
$f(y)$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\}$	$\frac{\mu^y\exp(-\mu)}{y!}$	$\binom{n}{y}\left(\frac{\mu}{n}\right)^y\left(1-\frac{\mu}{n}\right)^{n-y}$	$\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^\nu y^{\nu-1}\exp\left(-\frac{\nu y}{\mu}\right)$	$\sqrt{\frac{\gamma}{2\pi y^3}}\exp\left\{\frac{-\gamma(y-\mu)^2}{2\mu^2 y}\right\}$
Range	$-\infty < y < \infty$	$y = 0, 1, 2, \ldots$	$y = 0, 1, \ldots, n$	$y > 0$	$y > 0$
θ	μ	$\log(\mu)$	$\log\left(\frac{\mu}{n-\mu}\right)$	$-\frac{1}{\mu}$	$-\frac{1}{2\mu^2}$
ϕ	σ^2	1	1	$\frac{1}{\nu}$	$\frac{1}{\gamma}$
$a(\phi)$	$\phi(=\sigma^2)$	$\phi(=1)$	$\phi(=1)$	$\phi\left(=\frac{1}{\nu}\right)$	$\phi\left(=\frac{1}{\gamma}\right)$
$b(\theta)$	$\frac{\theta^2}{2}$	$\exp(\theta)$	$n\log(1+e^\theta)$	$-\log(-\theta)$	$-\sqrt{-2\theta}$
$c(y,\phi)$	$-\frac{1}{2}\left\{\frac{y^2}{\phi} + \log(2\pi\phi)\right\}$	$-\log(y!)$	$\log\binom{n}{y}$	$\nu\log(\nu y) - \log\{y\Gamma(\nu)\}$	$-\frac{1}{2}\left\{\log(2\pi y^3\phi) + \frac{1}{\phi y}\right\}$
$V(\mu)$	1	μ	$\mu(1-\mu/n)$	μ^2	μ^3
$g_c(\mu)$	μ	$\log(\mu)$	$\log\left(\frac{\mu}{n-\mu}\right)$	$\frac{1}{\mu}$	$\frac{1}{\mu^2}$
$D(y,\hat\mu)$	$(y-\hat\mu)^2$	$2y\log\left(\frac{y}{\hat\mu}\right) - 2(y-\hat\mu)$	$2\left\{y\log\left(\frac{y}{\hat\mu}\right) + (n-y)\log\left(\frac{n-y}{n-\hat\mu}\right)\right\}$	$2\left\{\frac{y-\hat\mu}{\hat\mu} - \log\left(\frac{y}{\hat\mu}\right)\right\}$	$\frac{(y-\hat\mu)^2}{\hat\mu^2 y}$

104

The regression of GLM (1)

- First, we should get the data for the regression analysis
- Second, we can check/assume the distribution type of the data
  - Because of the nature of data, sometimes it is impossible to say that the data will strictly follow a certain distribution
  - Normal distribution is often a good assumption at the beginning
  - Of course, if the normal assumption of linear models lead to quite high errors and small R2, we can test other distributions and try to find the best linear model

The regression of GLM (2)

- After picking up the distribution type, we should find the link function and dispersion scale function for the estimation of the parameters in the linear models
  - Estimate the mean of the data *Y*, *i.e.*, $\boldsymbol{\mu}(\boldsymbol{X}) = E[Y]$
  - Through the link function to transfer mean to predictor,

    $$\eta(X) = g\big(\mu(X)\big) = g(E[Y])$$
  - Then use regression methods to estimate the coefficients in the GLM,

    $$\eta(X) = \beta_0 + X\beta$$
- The mathematically procedures are given as follows.

Mathematically regression procedures for the GLM

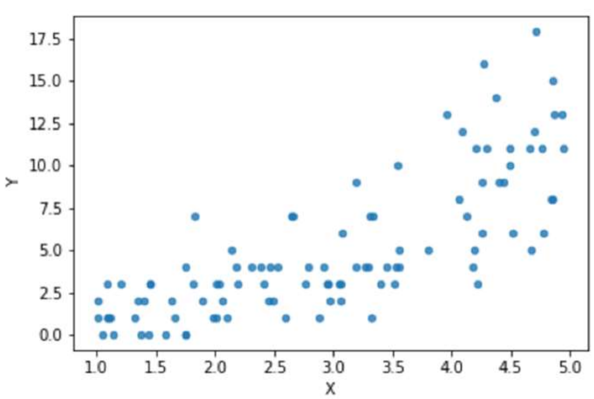
1. Get the data $(x_1, y_1), \ldots (x_n, y_n)$, fix link function $g(\mu)$ and dispersion scale function $V(\mu)$, and make some initial guesses β_0, β.
2. Until β_0, β converge:
   1. Calculate $\eta(x_i) = \beta_0 + x_i \cdot \beta$ and the corresponding $\widehat{\mu}(x_i)$
   2. Find the effective transformed responses $z_i = \eta(x_i) + (y_i - \widehat{\mu}(x_i))g'(\widehat{\mu}(x_i))$
   3. Calculate the weights $w_i = \left[(g'(\widehat{\mu}(x_i))^2 V(\widehat{\mu}(x_i))\right]^{-1}$
   4. Do a weighted linear regression of z_i on x_i with weights w_i, and set β_0, β to the intercept and slopes of this regression

2023-04-11

11

Examples of GLM: Poisson regression

- **For example, assume you need to predict the number of defect products (*Y*) with a sensor value (*X*) as the explanatory variable. The scatter plot looks like the figure below:**



- The relationship between *X* and *Y* does not look linear. It's more likely to be exponential.
- The variance of *Y* does not look constant wrt *X* (std(*Y)* increase as X).
- As *Y* is a positive and discrete variable → normal distribution used for linear regression may not work → linear regression may be not appropriate for this kind of data.

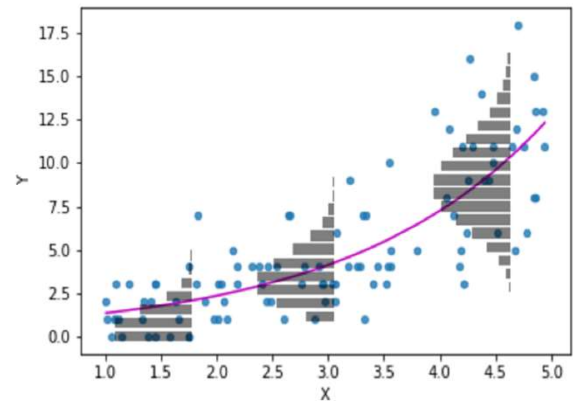
2023-04-11

12

Poisson regression

- **Now, we assume that Y follows Poisson distribution: Y~Poisson(λ)**
  - The mean value of Y is: $E[Y] = \lambda$. For each data point, $E[y_i|X] = \lambda_i$
  - The linear predictor is: $b_0 + b_1 X$
  - The link function is: $g(\lambda) = \ln(\lambda)$
  - Then GLM regression model becomes:

$$\ln \lambda_i = b_0 + b_1 x_i$$
$$\Leftrightarrow \lambda_i = \exp(b_0 + b_1 x_i)$$



2023-04-11

13

Examples of GLM: Normal and Binomial regressions

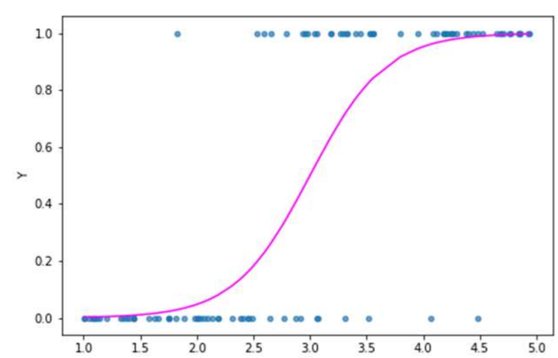
- **We have data (x_i, y_i), for normal case, the link function is the identify function**
- **while for the binomial distribution, the link function is the logit function (more will be given in Logical regression)**

$$\mu_i = b_0 + b_1 x_i$$
$$y_i \sim \mathcal{N}(\mu_i, \varepsilon)$$

$$y_i = b_0 + b_1 x_i$$

$$\log \frac{q_i}{1 - q_i} = b_0 + b_1 x_i$$
$$y_i \sim \mathrm{Binom}(q_i)$$

$$z_i = b_0 + b_1 x_i$$
$$q_i = \frac{1}{1 + \exp(-z_i)}$$



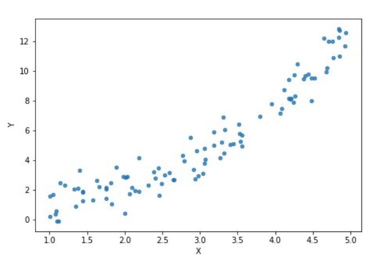
14

GLM: customize your own Link function (1)

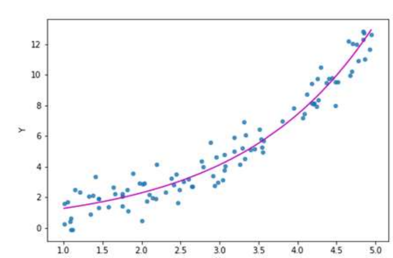
- Custom GLM
- The models I've explained so far uses a typical combination of probability distribution and link function. In other words, all the models above use the **canonical link function.**
- This is the list of probability distributions and their canonical link functions.
- Normal distribution: identity function
- Poisson distribution: log function
- Binomial distribution: logit function
- However, you don't necessarily use the canonical link function. Rather, the advantage of statistical modeling is that you can make any kind of model that fits well with your data.

GLM: customize your own Link function (2)

- This looks similar to the data I prepared for Poisson regression. However, if you see the data carefully, it seems the variance of *y* is constant with regard to *X*. Besides, *y* is continuous, not discrete.
- Therefore, it's appropriate to use normal distribution here. As the relationship between *X* and *y* looks exponential, you had better choose the log link function.
- ***NB:** you need to specify the link function as the default link for Gaussian distribution is the identity link function. The prediction result of the model looks like this.*



$$\ln \mu_i = b_0 + b_1 x_i$$
$$y_i \sim \mathcal{N}(\mu_i, \varepsilon)$$



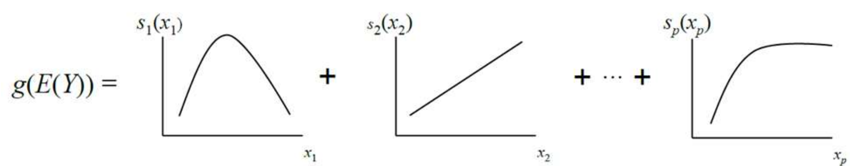
Generalized additive model (GAM)

17

GAM: Generalized additive models

- GAM based on an appealing and simple mental modelling techniques
  - Relationships between individual predictors X_i, i =1,2,… and dependent variable Y follow smooth patterns (linear/nonlinear)
  - These smooth relationship can be estimated simultaneously and then predict the $g(E[Y])$ (GLM) by simply adding them together
- The advantages of GAM can be summarized as
  - Easy to interpret, i.e., how individual predictors Xi can affect the values of $g(E[Y])$ (effect)
  - Flexible predictor functions f(Xi) can be of any forms (e.g., Spline) that can uncover hidden patterns in the data
  - Regularization of predictor functions help avoid overfitting

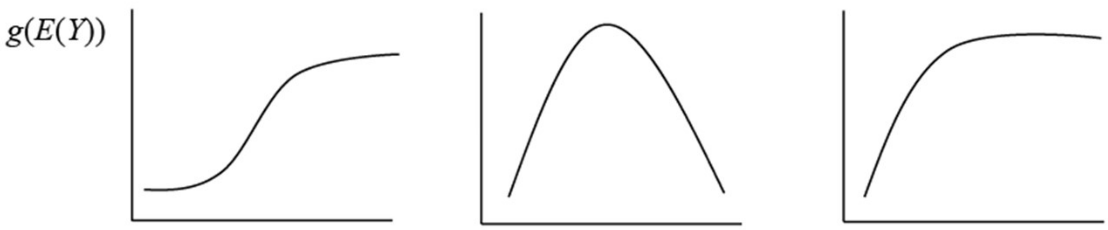
$$g(E(Y)) = \alpha + s_1(x_1) + \cdots + s_p(x_p)$$



18

GAM: interpretability

- When a regression model is additive, the interpretation of the marginal impact of a single variable (features, such as $f(x_i)$, or $f(x_i, x_j)$) does not depends on the values of other variables in the model.
- GAM is able to control the smoothness of the predictor functions. For each predictor, its effect is modelled through $f(x_i)$, often by Spline function that is smooth and easy controlled.
- This is beneficial if we have a prior knowledge/belief that the predictive models/relationships are smooth in nature rather than the noise data leading to very wiggly and nonsensical relationship because of different noises in the data.



GAM: mathematical models

- **Single predictor**

$$y = \beta_0 + x_1\beta_1 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \rightarrow y = \beta_0 + f(x_1) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\ln(y) = \beta_0 + \beta_1 x_1 + \varepsilon$$

$$\varepsilon \sim Poisson\ (\lambda)$$

$$\rightarrow$$

$$ln(y) = \beta_0 + f(x_1) + \varepsilon, \quad \varepsilon \sim Poisson(\lambda)$$

$$ln(y) = \beta_0 + f(x_1) + z_1\gamma + \varepsilon, \quad \varepsilon \sim Poisson(\lambda), \quad \gamma \sim N(0, \Sigma)$$

- **Multiple predictors, for simplification we assume the error is Gaussian RV**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ,\ldots, +\beta_n x_n + \varepsilon$$

$y = \beta_0 + f(x_1) + f(x_2)+, \ldots, +f(x_n) + \varepsilon$ without effect of tensor products

$y = \beta_0 + f(x_1) + f(x_2)+, \ldots, +f(x_n) + te(x_i, x_k) + \varepsilon$ with effects of tensor products

$f(x_i)$ is often models as a spline function.

GAM: an example

- **Overall** Science Score (average score for 15 years)
  - **Interest** in science
  - **Support** for scientific inquiry
  - **Income** Index
  - **Health** Index
  - **Edu**cation Index
  - **H**uman **D**evelopment **I**ndex (HDI)

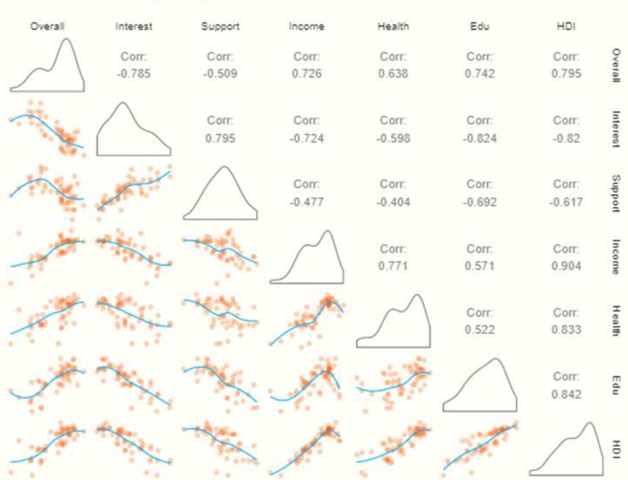
Correlation reflect how strong two variables are linearly dependent.

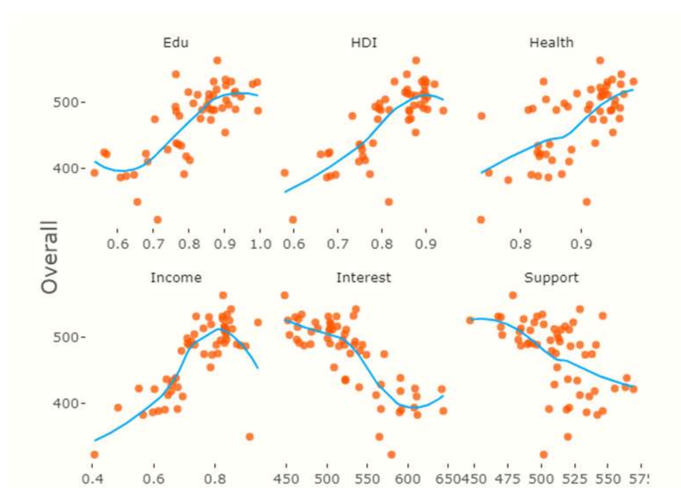
Variable	N	Mean	SD	Min	Q1	Median	Q3	Max	Missing
Overall	57	473.1	54.6	322.0	428.0	489.0	513.0	563.0	8
Issues	57	469.9	53.9	321.0	427.0	489.0	514.0	555.0	8
Explain	57	475.0	54.0	334.0	432.0	490.0	517.0	566.0	8
Evidence	57	469.8	61.7	288.0	423.0	489.0	515.0	567.0	8
Interest	57	528.2	49.8	448.0	501.0	522.0	565.0	644.0	8
Support	57	512.2	26.1	447.0	494.0	512.0	529.0	569.0	8
Income	61	0.7	0.1	0.4	0.7	0.8	0.8	0.9	4
Health	61	0.9	0.1	0.7	0.8	0.9	0.9	1.0	4
Edu	59	0.8	0.1	0.5	0.7	0.8	0.9	1.0	6
HDI	59	0.8	0.1	0.6	0.7	0.8	0.9	0.9	6

21

GAM: an example

- **Overall** Science Score (average score for 15 years)
- **Display their correlation structures**

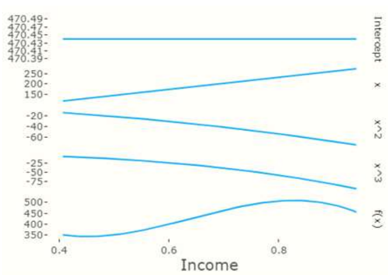




22

GAM: Single predictor

- **Linear regression (a special case of GAM, i.e., Gaussian noise + linear term additive)**
  - $Overall = \ b_0 + b_1 \times income$ (b0: intercept, b1: slope for income)
  - Let's write as $g(\mu) = b_0 + b_1 x$
  - In the following, we assume normal noise, i.e., identical link function



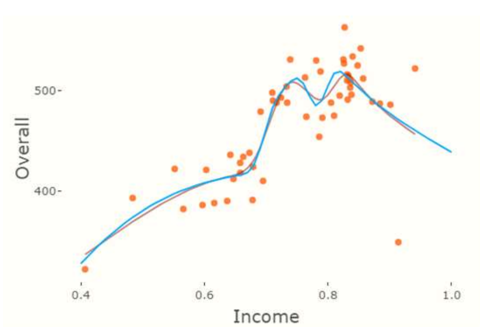
Models for effects of different $b_j(x)$

```
Family: gaussian
Link function: identity

Formula:
Overall ~ Income

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   204.32      35.37   5.777 4.32e-07 ***
Income        355.85      46.79   7.606 5.36e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.518   Deviance explained = 52.7%
GCV = 1504.5  Scale est. = 1448.8    n = 54
```

23

GAM: Single predictor

- **Spline regression (additive of polynomials)**
  - $g(\mu) = \sum_{j=1}^{q} b_j(x)\gamma_j$: $b_j(x)$ basis function of B-Spline function
  - In previous lectures, there are specific formulas for $b_j(x)$
  - In the GAM, it is to fit each $b_j(x)$ and add them together
  - The "fake" illustration and results are given as follows



Two models of the spline fits

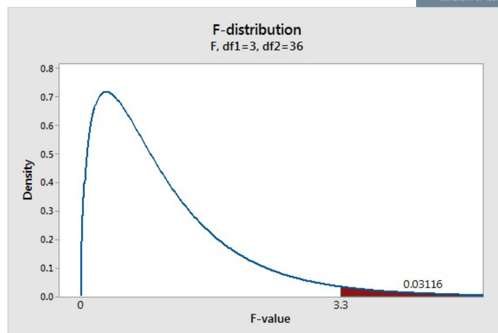
```
Family: gaussian
Link function: identity

Formula:
Overall ~ s(Income, bs = "cr")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  470.444      4.082   115.3   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
            edf Ref.df     F  p-value
s(Income) 6.895  7.741 16.67 1.59e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =    0.7   Deviance explained = 73.9%
GCV = 1053.7  Scale est. = 899.67    n = 54
```

2023-04-11

24

GAM: Multiple predictors (1)

- Linear model regression
  - $Overall = b_0 + b_1 \times income + b_2 \times edu + b_3 \times health$
  - Let's write as $g(\mu) = b_0 + b_1x_1 + b_2x_2 + b_3x_3$
  - $g(\mu)$ is additive of all linear models of various predictors




```
Family: gaussian
Link function: identity

Formula:
Overall ~ Income + Edu + Health

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   121.18      78.97   1.535   0.1314
Income        182.32      85.27   2.138   0.0376 *
Edu           234.11      54.78   4.274 9.06e-05 ***
Health         27.01     134.90   0.200   0.8421
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.616   Deviance explained = 63.9%
GCV = 1212.3  Scale est. = 1119       n = 52
```

Hypothesis of the significant code:
- If the model of predictor x_j, i.e., $s(x_j)$, should be taken away from the GAM.
- What is the probability to not reject the hypothesis.

2023-04-11

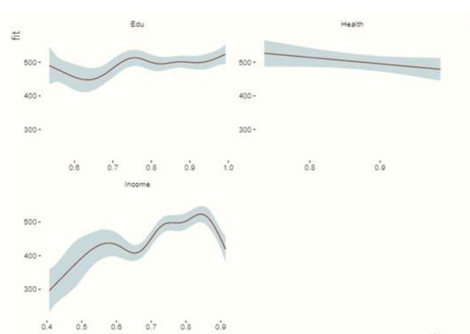
25

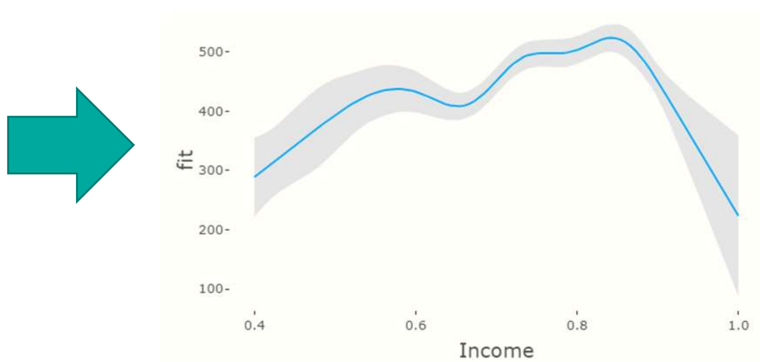
GAM: Multiple predictors (2)

```
Family: gaussian
Link function: identity

Formula:
Overall ~ s(Income) + s(Edu) + s(Health)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  471.154      2.772     170   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
            edf Ref.df     F  p-value
s(Income) 7.593  8.415 8.826 1.33e-07 ***
s(Edu)    6.204  7.178 3.308  0.00733 **
s(Health) 1.000  1.000 2.736  0.10661
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.863   Deviance explained = 90.3%
GCV = 573.83  Scale est. = 399.5      n = 52
```

- GAM: additive of $f(x_i)$. If $f()$ is spline, rewrite it as $s()$
  - $s(x_i) = \sum_{j=1}^{q} b_j(x_i)\gamma_j$: $b_j()$ basis function of B-Spline
  - $g(\mu) = b_0 + \sum_{i=1}^{3} s(x_i)$ (no effect on predictor interactions)
  - $g(\mu) = b_0 + \sum_{i=1}^{3} s(x_i) + \sum te(x_i, x_j)$

  The above is the model with effects on predictor tensor product





-11

26

Practical issues for GAM

- How can the GAM be regressed/estimated? Let write the GAM in a more general format

$$g(\mu) = \beta_0 + \sum s(x_i) + \sum te(x_i, x_j)$$
$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 +, \dots, +\beta_n x_n + \gamma_1 x_1^2 + \gamma_2 x_2^2 +, \dots$$
$$= X\beta + f(x_1) + f(x_2) +, \dots$$

where $f(x_i)$ contains second and high order of xi from the base functions

- The cost function of the ”penalized regression” is given as (thin plate Spline):

$$Loss = \sum (y - X\beta)^2 + \sum \lambda \int f''(x_i)^2 dx_i$$

- The value of λ is called the penalty coefficient that control the smoothness of the model
  - As $\lambda \to \infty$, the result is a linear fit because any wiggliness will add too much to the loss function.
  - As $\lambda \to 0$, any wiggliness is incorporated into the model.

2023-04-11

27

Estimation of the GAM

After getting the cost function, some optimization methods can be used to find the coefficient to minimize the cost.

- If the order of the base functions $f(x_i)$ is low, the least square method can be simply used to get the solution
- If the $f(x_i)$ is very smooth, i.e., with hight order of x_i, the backfitting method can be used.
  - The backfitting strategy is to first fit the Spline function for x_1, i.e., $f(x_1)$
  - Estimate residual between y and $f(x_1)$
  - Then fit another spline curve for the second predictor x_2
  - Continue the process for all the predictors. NB: there might be needs to adjust the previous Splines
  - The way of backfitting is quite similar as the mixed effect model (introduced later)

2023-04-11

28

Estimation of the GAM

- For each predictor $f(x_i)$, its effective degree of freedom (edf): number of terms in the model
- The value of edf can also determine the significance of the predictor in the GAM.
- After the modelling, it is important to run diagnostics to check, e.g., agreement between model and data, residual in terms of various predictors, preferably, compare different models such as GAM with smooth polynominal, B-splines, natural spline, different λ…

2023-04-11

29

Mixed-effects model

2023-04-11

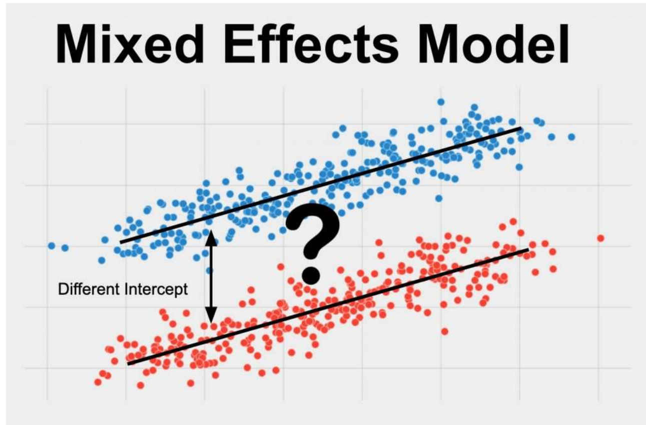
30

Mixed-effects model

Mixed effects model is to use two or more variables (predictors) to describe one output variable. For example:

- *To predict a person's weight in terms of hight, you may have a rough global relation/model or all the data.*
- *Then you may need to distinguish the males from the females.*
- *Or you may need to categorize the people into different countries.*

Summary: often it is a mixture of global trend added with local models for extra categorizes. But the global trend model may not be necessary for the mixed effects model



A Mixed Effects Model is sometimes also called Mixed Effects Regression, Multi-Level Model, Hierarchical Model, or Repeated Measures Linear Regression (from internet).

2023-04-11

31

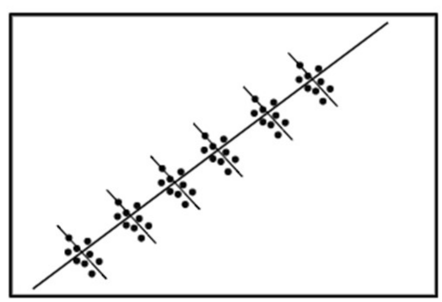
Mixed-effects Model: mathematical description

$$\mathbf{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

$$\overbrace{\mathbf{y}}^{\mathrm{N\,x\,1}} = \overbrace{\underbrace{\mathbf{X}}_{\mathrm{N\,x\,p}}\;\underbrace{\boldsymbol{\beta}}_{\mathrm{p\,x\,1}}}^{\mathrm{N\,x\,1}} + \overbrace{\underbrace{\mathbf{Z}}_{\mathrm{N\,x\,qJ}}\;\underbrace{\boldsymbol{u}}_{\mathrm{qJ\,x\,1}}}^{\mathrm{N\,x\,1}} + \overbrace{\boldsymbol{\varepsilon}}^{\mathrm{N\,x\,1}}$$

$$N = \sum_{j}^{J} n_j$$

$$\overbrace{\mathbf{y_j}}^{n_j\times 1} = \overbrace{\underbrace{\mathbf{X_j}}_{n_j\times 6}\;\underbrace{\boldsymbol{\beta}}_{6\times 1}}^{n_j\times 1} + \overbrace{\underbrace{\mathbf{Z_j}}_{n_j\times 1}\;\underbrace{\boldsymbol{u_j}}_{1\times 1}}^{n_j\times 1} + \overbrace{\boldsymbol{\varepsilon_j}}^{n_j\times 1}$$



J: number of groups/categories
p: number of features in global level
qJ: number of feature within each group
n_j: number of data within each group

2023-04-11

32

Mixed-effects Model: an example

CHALMERS UNIVERSITY OF TECHNOLOGY

- Data of interest
  - A survey of residents of 4 states with model mood=f(exercise)
- We will test the following models:
  - **Simple global Linear Regression (1 predictor)**
  - **A More Robust Linear Regression (2 predictor)**
  - **Mixed-Effect Models**

```
State      Exercise          Mood
A:120   Min.   :21.75   Min.   :0.700
B:120   1st Qu.:43.31   1st Qu.:4.200
C:120   Median :57.00   Median :5.500
D:120   Mean   :56.61   Mean   :5.288
        3rd Qu.:69.19   3rd Qu.:6.700
        Max.   :93.75   Max.   :9.400
```

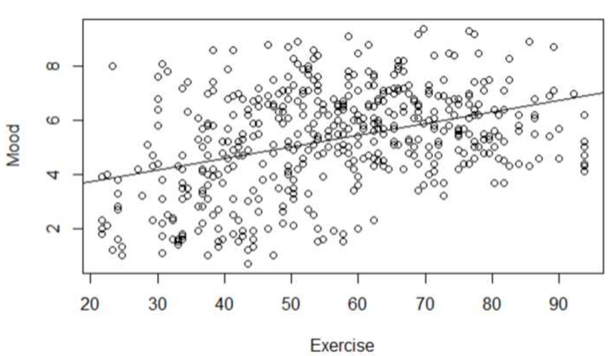
2023-04-11

33

Mixed-effects Model: an example

CHALMERS UNIVERSITY OF TECHNOLOGY

- **Let look the simple model: mood = intercept + slope * exercise**
- **Simple linear regression will give us the coefficients in the model**



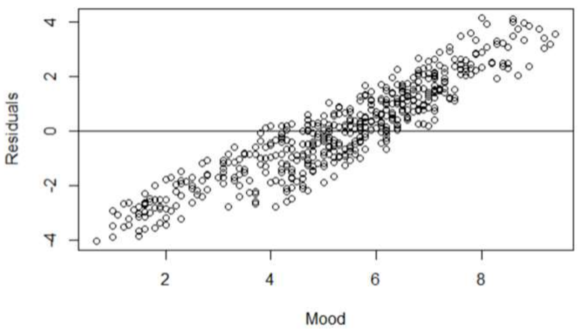

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.856399   0.275674  10.362   <2e-16 ***
Exercise    0.042951   0.004662   9.213   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.746 on 478 degrees of freedom
Multiple R-squared:  0.1508,    Adjusted R-squared:  0.149
F-statistic: 84.88 on 1 and 478 DF,  p-value: < 2.2e-16
```

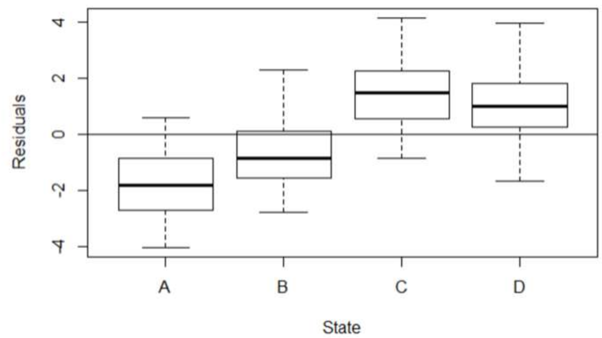
2023-04-11

34

Simple linear model: diagnostics

- Residual vs mood (predicted), Residual vs mood (data),Residual vs exercise
- Should be constant variance along the X-axis
- It is not cost variance, so we dig into the details of the residual for different states (different distributions of residual for those states)





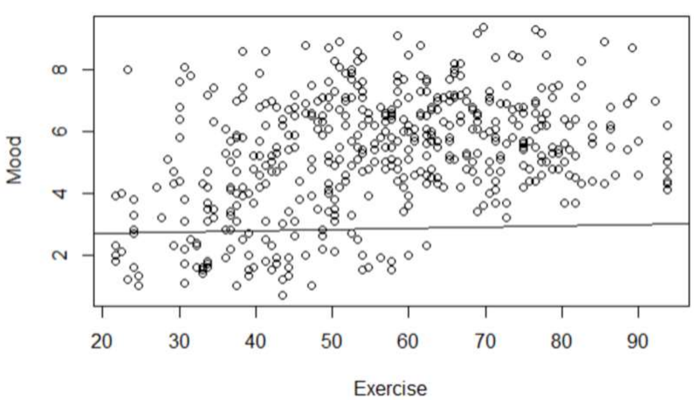
11

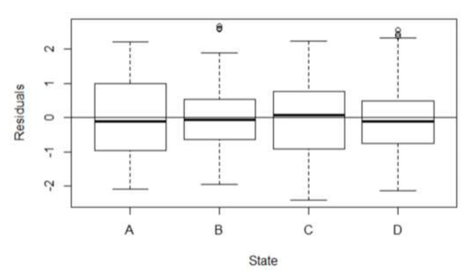
35

A More Robust Linear Regression

- **Now, we consider the state in the model:**
  - Mood = intercept + slope1* Exercise + slope2*State

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.633314   0.193123   13.63   <2e-16 ***
Exercise    0.003773   0.004238    0.89    0.374
StateB      2.236506   0.189464   11.80   <2e-16 ***
StateC      3.588067   0.141681   25.32   <2e-16 ***
StateD      3.938775   0.178897   22.02   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.052 on 475 degrees of freedom
Multiple R-squared:  0.6939,    Adjusted R-squared:  0.6913
F-statistic: 269.2 on 4 and 475 DF,  p-value: < 2.2e-16
```



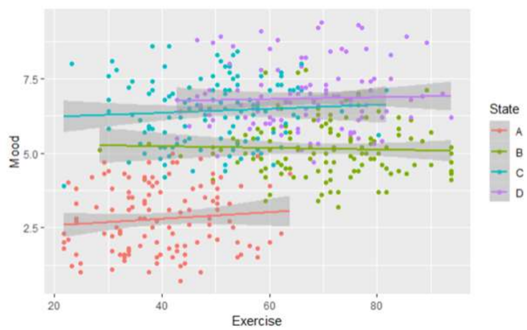


2023-04-11

36

A mixed effects model

- It is strange the above model that the mood is not dependent on the exercise. This is because the effect of states is dominating the value of mood
- Now we try to categorize states into different groups, and each group will have their own effect model between mood and exercise
- The model can be written as: Mood ~ Exercise + (1 + Exercise | State)



RMSE of the three models:

```
[1] 38.17183
[1] 22.91801
[1] 22.91435
```

```
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 State    (Intercept) 3.558e+00 1.886312
          Exercise    5.967e-06 0.002443 -1.00
 Residual             1.105e+00 1.051407
Number of obs: 480, groups:  State, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept) 5.068043   0.974318   5.202
Exercise    0.004157   0.004391   0.947

Correlation of Fixed Effects:
         (Intr)
Exercise -0.506
convergence code: 0
boundary (singular) fit: see ?isSingular

$`State`
  (Intercept)    Exercise
A    2.492797 0.007491521
B    4.822434 0.004474642
C    6.275295 0.002593187
D    6.681645 0.002066965
```

-04-11

37

A take home exercise

- **As a multi-level model**
- https://m-clark.github.io/mixed-models-with-R/random_intercepts.html

2023-04-11

38

CHALMERS
UNIVERSITY OF TECHNOLOGY

39