Distributional Regression

Sydney Informatics Hub Masterclass series

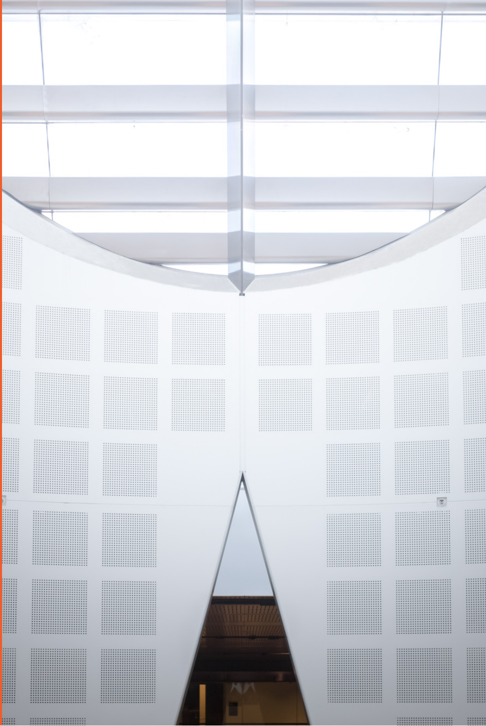
May 25, 2023

Presented by

Stanislaus Stadlmann

Sydney Informatics Hub

THE UNIVERSITY OF SYDNEY



Overview

About SIH masterclasses and this talk

How did we get here?

Distributional Regression

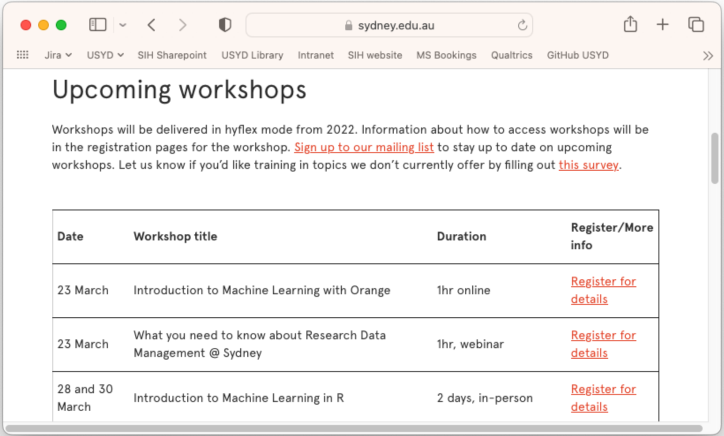
1 About SIH masterclasses and this talk

THE UNIVERSITY OF SYDNEY

SIH masterclasses

About

- Standalone workshops on a particular topic
- Third thursday of the month
- Check out our ”upcoming workshops” on the SIH website



Upcoming workshops

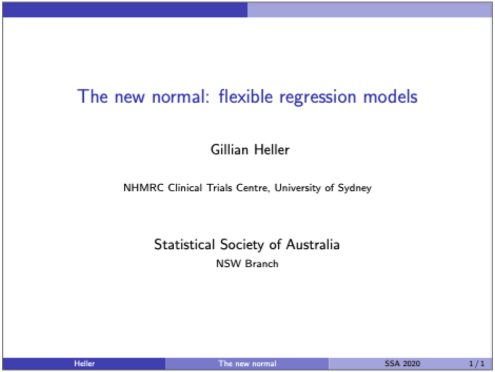
Workshops will be delivered in hyflex mode from 2022. Information about how to access workshops will be in the registration pages for the workshop. Sign up to our mailing list to stay up to date on upcoming workshops. Let us know if you'd like training in topics we don't currently offer by filling out this survey.

Date	Workshop title	Duration	Register/More info
23 March	Introduction to Machine Learning with Orange	1hr online	Register for details
23 March	What you need to know about Research Data Management @ Sydney	1hr, webinar	Register for details
28 and 30 March	Introduction to Machine Learning in R	2 days, in-person	Register for details

This talk

More info

- Gillian Heller's talk at JB Douglas awards 2020
- "The new normal: distributional regression"
- Re-written for new audience
- Recorded



2 How did we get here?

THE UNIVERSITY OF SYDNEY

Regression Analysis: A brief history

How long is a metre?

The last original metre in Paris, France

Regression Analysis: A brief history

How long is a metre?

- 1789: French Revolution
- Desire to replace features of the Ancien Régime
- The *toise*: ”distance between the fingertips of the outstretched arms of a man” (Quinion, 2007)

Regression Analysis: A brief history

How long is a metre?

- 1789: French Revolution
- Desire to replace features of the Ancien Régime
- The *toise*: "distance between the fingertips of the outstretched arms of a man" (Quinion, 2007)

Definition

$m = \frac{d}{10{,}000{,}000}$, where d is the distance between Equator and North Pole

Regression Analysis: A brief history

How long is a metre?

- 1789: French Revolution
- Desire to replace features of the Ancien Régime
- The *toise*: ”distance between the fingertips of the outstretched arms of a man” (Quinion, 2007)

Definition

$m = \frac{d}{10{,}000{,}000}$, where d is the distance between Equator and North Pole

But: How long is this distance d?

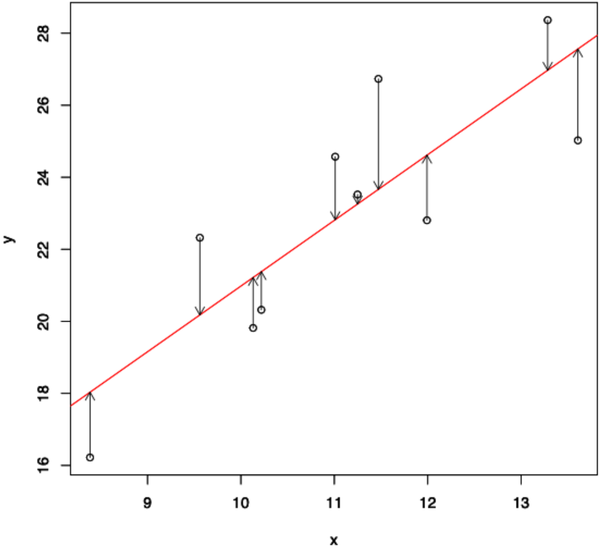
Regression Analysis: A brief history

How long is a metre?

- A portion of the quadrant to be surveyed within France
- Adrien-Marie Legendre tasked to combine multiple measurements
- His publication: Legendre (1805) comes from that
- “Invention” of the Least Squares method
- Spicy: Gauss later claimed he had already been using this technique since 1775.

Regression Analysis: A brief history

Linear Regression Analysis



Regression Analysis: A brief history

Linear Regression: Model Assumptions

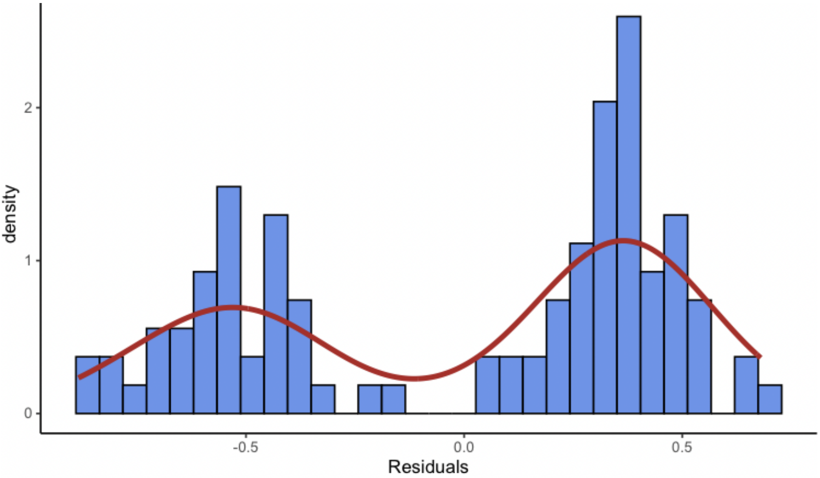
$$y_i = x_i^\top \beta + \epsilon_i$$
$$\epsilon_i \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2)$$

- normality of errors
- homoscedasticity of errors: $\mathbb{V}(y_i) = \sigma^2$
- independence

Regression Analysis: A brief history

Error distribution

But what if our errors are not normally distributed?



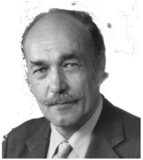
Regression Analysis: A brief history

Generalized Linear Models

"Theoretical and applied statistics were both convulsed by the publication of the GLM paper by Nelder and Wedderburn (1972)." (Aitkin, 2018)

J. R. Statist. Soc. A, (1972), **135**, *Part* 3, p. 370

370



Generalized Linear Models

By J. A. NELDER and R. W. M. WEDDERBURN

Rothamsted Experimental Station, Harpenden, Herts

SUMMARY

The technique of iterative weighted linear regression can be used to obtain maximum likelihood estimates of the parameters with observations distributed according to some exponential family and systematic effects that can be made linear by a suitable transformation. A generalization of the analysis of variance is given for these models using log-likelihoods. These generalized linear models are illustrated by examples relating to four distributions; the Normal, Binomial (probit analysis, etc.), Poisson (contingency tables) and gamma (variance components).

The implications of the approach in designing statistics courses are discussed.

Regression Analysis: A brief history

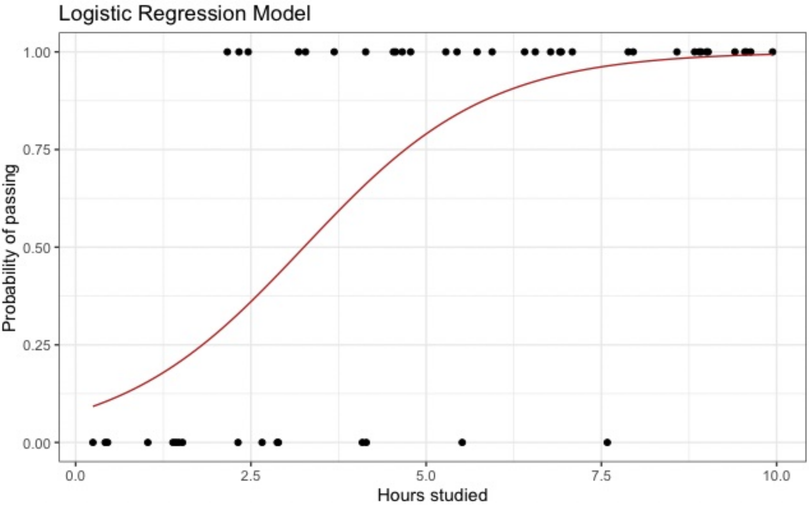
Generalized Linear Models

Model assumptions:

$$y_i \overset{\text{ind}}{\sim} \mathcal{D}(\mu_i, \phi)$$
$$\mathbb{E}(y_i) = \mu_i = h(x_i^\top \beta)$$

1. exponential family response distribution
   - normal
   - Poisson (count regression)
   - binomial (logistic model)
   - Gamma
   - inverse Gaussian
   - (negative binomial)
2. constant *dispersion parameter* ϕ
3. independence

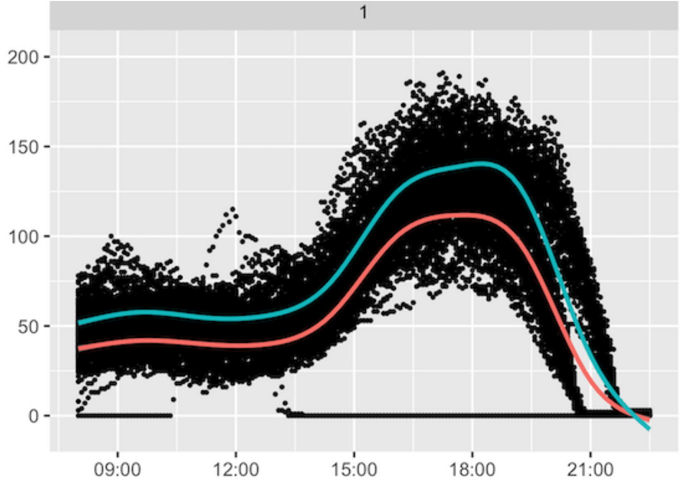
Regression Analysis: A brief history



Regression Analysis: A brief history

Linearity Assumption

What if our explanatory variables x_i don't have a linear connection with $E(y)$?



Regression Analysis: A brief history

Generalized Additive Models

Hastie and Tibshirani (1986)

$$y_i \overset{\text{ind}}{\sim} \mathcal{E}(\mu_i, \phi)$$
$$g(\mu_i) = \eta_i = s_1(x_{i1}) + \ldots + s_J(x_{iJ})$$

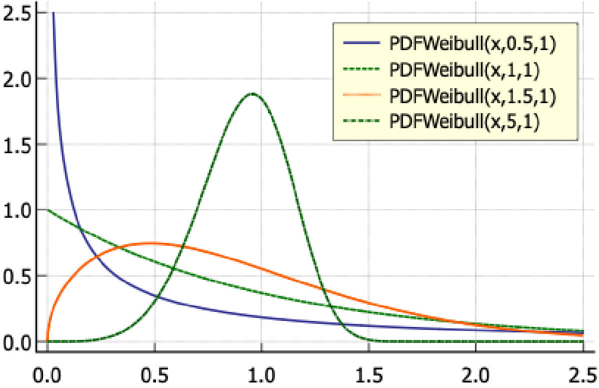
- $s_j(x_{ij})$ are parametric *or* smooth functions
- Smooth splines can be parametric or non-parametric (penalized)
- The $s_j(\cdot)$ can be
  - curves (e.g. growth curves)
  - spatial effects
  - varying coefficient terms
  - interaction surfaces of two continuous variables
  - random effects

3 Distributional Regression

Distributional Regression

Back to basics

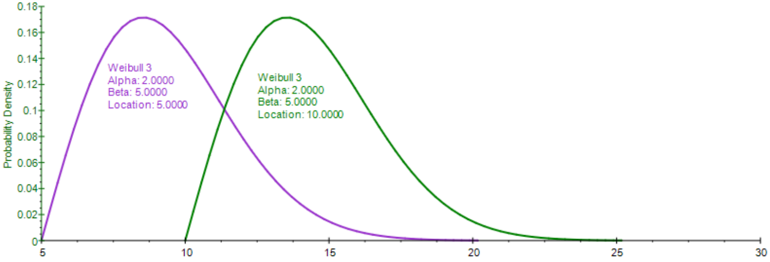
What are the three properties (moments) of a distribution?



Distributional Regression

Three properties

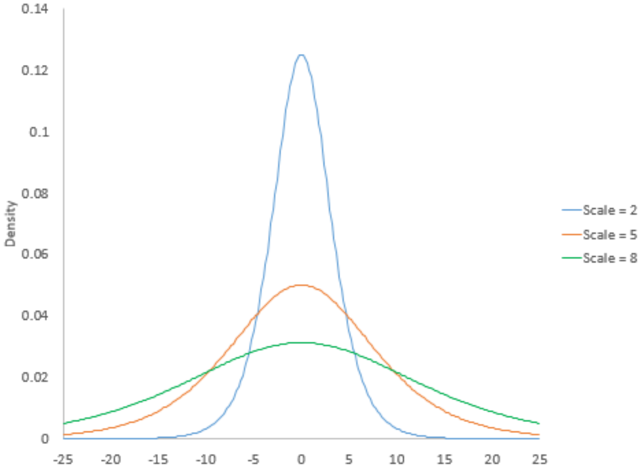
- Location



Distributional Regression

Three properties

- Scale

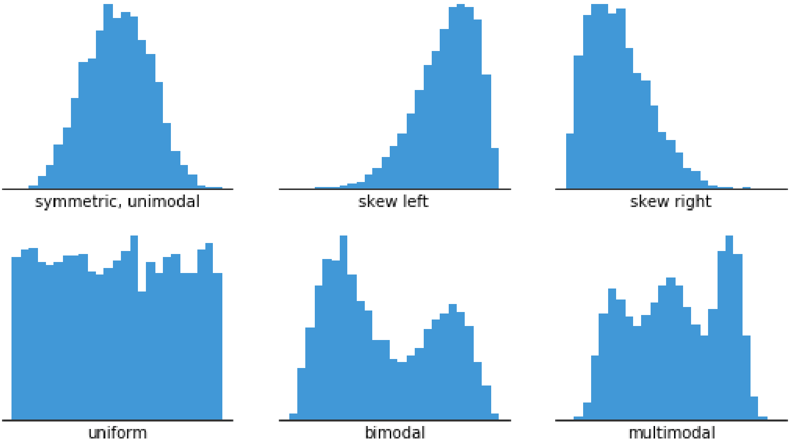


Distributional Regression

Three properties

Yi (2021)

- Shape



Distributional Regression

"It is difficult to understand why statisticians commonly limit their enquiries to Averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once."
- Sir Francis Galton (England, 1822-1911)

Distributional Regression

An expected result

"...most empirical modelling is firmly located in the lowlands of the conditional expectation of the distribution of the response, e.g. in generalized linear models."
(Kneib, Silbersdorff, & Säfken, 2023)

- Linear Models, GLM, GAM centered around finding the mean of a distribution
- What about all the other properties?

Distributional Regression

An example

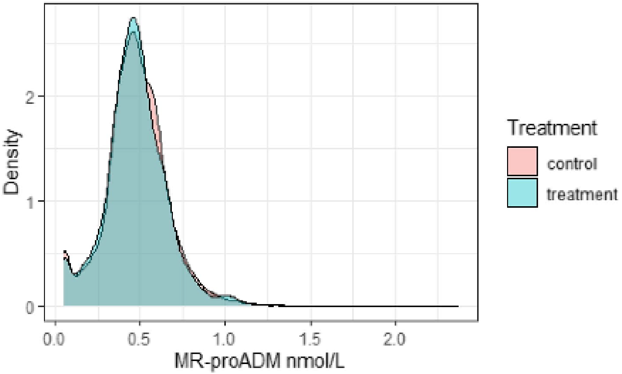


Figure: Kernel Density estimate of MR-proADM (a biomarker used on coronary heart disease), as seen in Heller et al. (2022)

- Does *Pravastatin* change the distribution of *MR-proADM*?

Distributional Regression

MR-proADM biomarker

- Four scenarios:
  1. Normal (reduced)
  2. BCT (reduced)
  3. Normal (extended)
  4. BCT (extended)

Distributional Regression

MR-proADM biomarker

- Four scenarios:
  1. Normal (reduced)
  2. BCT (reduced)
  3. Normal (extended)
  4. BCT (extended)

		Reduced model					
		BCT			NO		
Parameter	Coefficient	Estimate	SE	p	Estimate	SE	p
μ	(Intercept)	−0.129	0.007	< 0.001	0.668	0.005	< 0.001
μ	baseline	0.813	0.009	< 0.001	0.236	0.004	< 0.001
μ	sex						
μ	treatment	−0.014	0.005	0.004	−0.005	0.004	0.224
σ	(Intercept)	−1.959	0.019	< 0.001	−1.818	0.009	< 0.001

Treatment effect. Is that it?

Distributional Regression

MR-proADM biomarker

Scenario 3 & 4 (extended)

		Extended model					
		BCT			NO		
Parameter	Coefficient	Estimate	SE	p	Estimate	SE	p
μ	(Intercept)	−0.088	0.006	< 0.001	0.745	0.008	< 0.001
μ	baseline	0.879	0.008	< 0.001	0.319	0.008	< 0.001
μ	sex				−0.017	0.006	0.003
μ	treatment	−0.018	0.005	< 0.001	−0.005	0.004	0.166
σ	(Intercept)	−2.796	0.050	< 0.001	−1.890	0.025	< 0.001
σ	baseline	−1.426	0.039	< 0.001	−0.230	0.017	< 0.001
σ	sex	−0.158	0.037	< 0.001	−0.154	0.024	< 0.001
σ	treatment	0.073	0.030	0.014			

- Treatment effect on the scale parameter
- From unsuccessful to successful

Distributional Regression

Overview

Let $y \sim D(\alpha_1, \ldots, \alpha_K)$. With distributional regression...

- Every parameter α_i can be modeled using a (different) set of predictors
- Predictors can take different forms, e.g. non-linear, spatial, random effects
- Parameters are connected to the predictors using link-functions that uphold the support of the parameter

Model Equations

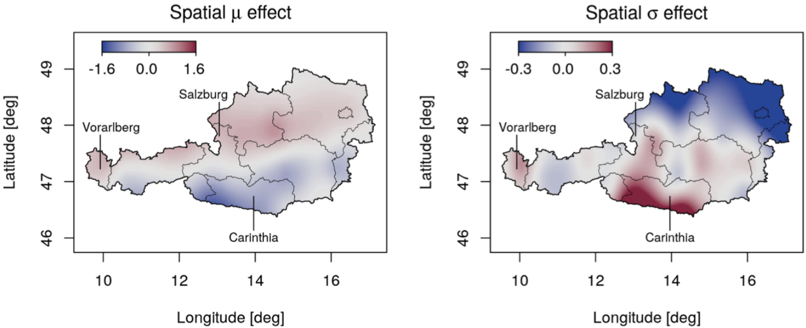
$\Rightarrow$ This allows for extremely flexible model equations:

$$g_l(\boldsymbol{\alpha}_l) = f_{1l}(\mathbf{X}_{1l}; \boldsymbol{\theta}_{1l}) + \ldots + f_{Q_l l}(\mathbf{X}_{Q_l l}; \boldsymbol{\theta}_{Q_l l}) \tag{1}$$

Distributional Regression

Precipitation in Austria

Umlauf, Klein, and Zeileis (2018)



- Mean and variance spatial effect on precipitation

Distributional Regression

In Practice

Implementations in R:

- gamlss (Stasinopoulos, Rigby, Heller, Voudouris, & De Bastiani, 2017, Generalized Additive Models for Location, Scale and Shape)
- bamlss (Umlauf et al., 2018)
- VGAM (Yee, 2010, Vector Generalized Additive Models)

Distributional Regression

How to get started

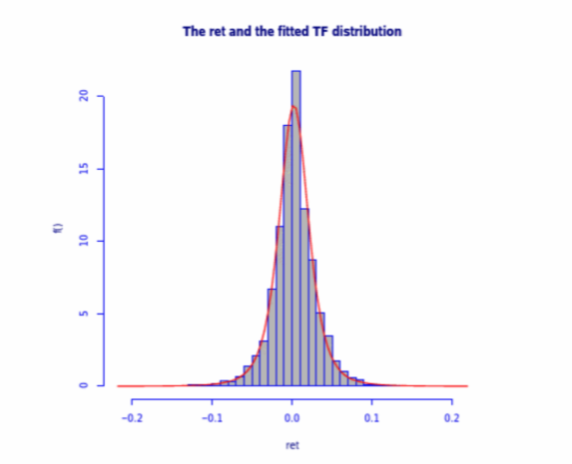


Figure: Turkish stock exchange returns, taken from Stasinopoulos et al. (2017)

Distributional Regression

How to get started

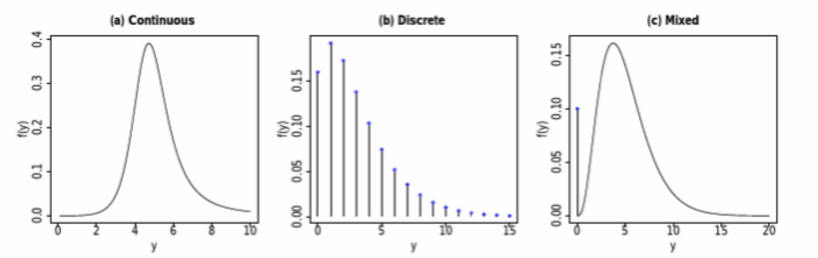
- `fitDist()` function can select the right distribution for your data
- Conditional distributions (as dependent on your data) can be fitted as well

```
Python                                                        ⌥⌘1
R> m5 <- fitDist(ret, data = tse, type = "realline", ncpus = 2)
  |======================================================================| 100%E
rror in solve.default(oout$hessian) :
  Lapack routine dgesv: system is exactly singular: U[3,3] = 0


R> m5$fits
     SEP2      SEP1      SEP3      SEP4        PE       PE2        GT    SHASHo
-12879.01 -12876.88 -12876.14 -12865.04 -12862.09 -12862.09 -12860.09 -12836.07
  SHASHo2     SHASH      JSUo       JSU       ST3        TF       ST2       ST5
-12836.07 -12833.92 -12803.68 -12803.68 -12790.80 -12789.56 -12788.83 -12788.20
      ST1       ST4       TF2       SST        LO       NET    exGAUS        NO
-12787.75 -12787.62 -12785.20 -12783.20 -12726.41 -12719.68 -12461.02 -12444.12
      SN2       SN1        GU        RG
-12442.92 -12442.12 -11578.60 -11305.11
```

Distributional Regression

How to get started

- `fitDist()` function can select the right distribution for your data
- Conditional distributions (as dependent on your data) can be fitted as well



Distributional Regression

How to get started

Once you are ready to fit your model...

```
Python
R> m0 <- gamlss (lborev1 ~ lboopen, data = film90, family = NO)
GAMLSS-RS iteration 1: Global Deviance = 15079.74
GAMLSS-RS iteration 2: Global Deviance = 15079.74

R> m1 <- gamlss(lborev1 ~ lboopen, sigma.formula = ~ lboopen, data = film90)
GAMLSS-RS iteration 1: Global Deviance = 14051.1
GAMLSS-RS iteration 2: Global Deviance = 13957.29
GAMLSS-RS iteration 3: Global Deviance = 13957.29
```

- μ term fitted just as in normal linear regression
- `sigma.formula` argument for σ

Distributional Regression

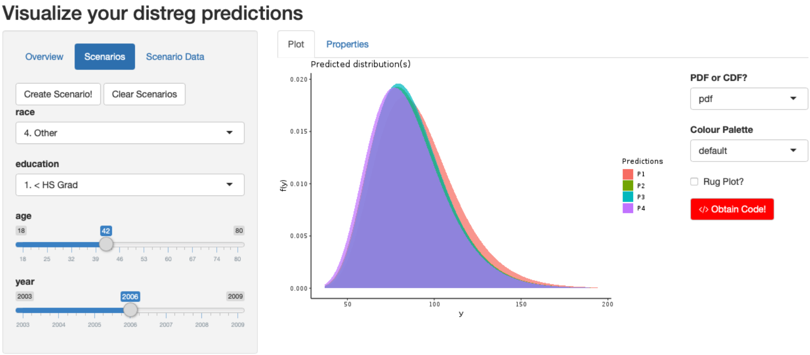
How to get started

```
R> summary(m1)
******************************************************************
Family:  c("NO", "Normal")

Call:  gamlss(formula = lborev1 ~ lboopen, sigma.formula = ~lboopen,
    family = NO, data = film90)

Fitting method: RS()

------------------------------------------------------------------
Mu link function:  identity
Mu Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.112446   0.106966    29.1   <2e-16 ***
lboopen     0.868785   0.007378   117.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

------------------------------------------------------------------
Sigma link function:  log
Sigma Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.18257    0.05408   40.36   <2e-16 ***
lboopen     -0.15872    0.00449  -35.35   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Distributional Regression

Post-model diagnostics

Stadlmann and Kneib (2022): `distreg.vis`



Thank you

Thank you!

4 References

THE UNIVERSITY OF SYDNEY

Aitkin, M. (2018). A history of the glim statistical package. *International Statistical Review*, *86*(2), 275-299. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12251` doi: https://doi.org/10.1111/insr.12251

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3), 297–310. Retrieved 2023-04-14, from `http://www.jstor.org/stable/2245459`

Heller, G. Z., Robledo, K. P., & Marschner, I. C. (2022). Distributional regression in clinical trials: treatment effects on parameters other than the mean. *BMC Medical Research Methodology*, *22*(1), 56. Retrieved from `https://doi.org/10.1186/s12874-022-01534-8` doi: 10.1186/s12874-022-01534-8

Kneib, T., Silbersdorff, A., & Säfken, B. (2023). Rage against the mean – a review of distributional regression approaches. *Econometrics and Statistics*, *26*, 99-123. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2452306221000824` doi: https://doi.org/10.1016/j.ecosta.2021.07.006

Legendre, A. M. (1805). *Nouvelles methodes pour la determination des orbites des cometes [microform] / par a.m. legendre* [Book, Microform]. F. Didot Paris.

Quinion, M. (2007). *World wide words: Toise.* Retrieved 2023-03-27, from `https://www.worldwidewords.org/qa/qa-toi1.htm`

Stadlmann, S., & Kneib, T. (2022). Interactively visualizing distributional regression models with distreg.vis. *Statistical Modelling*, *22*(6), 527-545. Retrieved from `https://doi.org/10.1177/1471082X211007308` doi: 10.1177/1471082X211007308

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). *Flexible regression and smoothing: Using GAMLSS in R*. Boca Raton, Florida: CRC Press.

Umlauf, N., Klein, N., & Zeileis, A. (2018). BAMLSS: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, *27*(3), 612–627.

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, *32*(10), 1–34.

Yi, M. (2021). *Histogram complete guide.* `https://chartio.com/learn/charts/histogram-complete-guide/`. (Accessed: April 19, 2023)