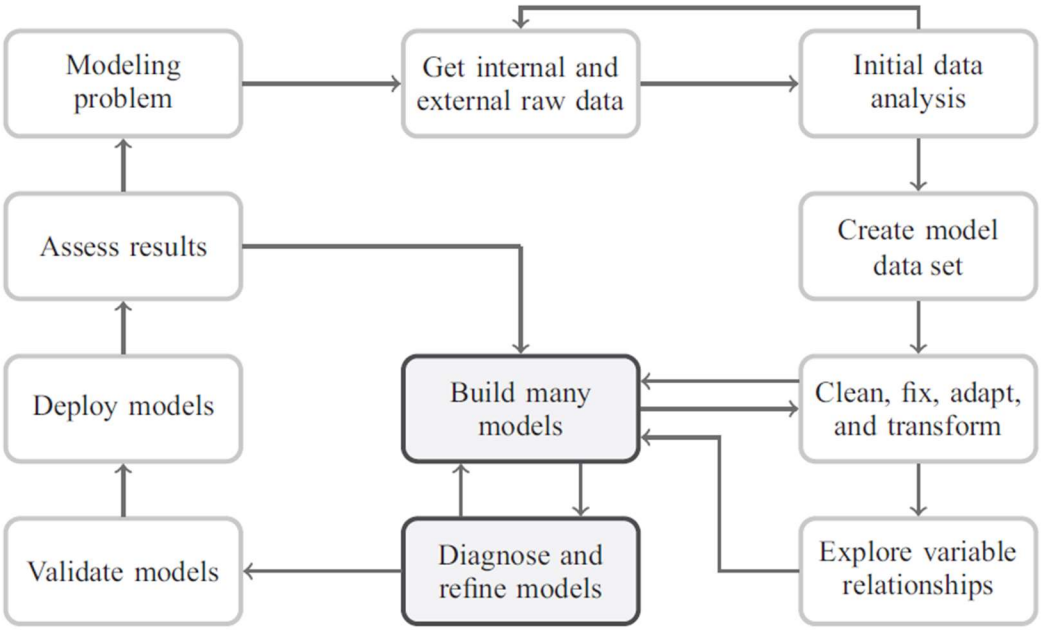

Fig. 1.1. Overall project cycle.

Frees, Meyers & Derrig (2016) eds. Predictive Modeling Applications in Actuarial Science Volume 2. Case Studies in Insurance

Datos:

https://instruction.bus.wisc.edu/jfrees/jfreesbooks/PredictiveModelingVol1/glm/v2-chapter-1.html

Notar que en el proceso de modelaje a partir de un ejemplo:

1. Replica (local) de código
    a. Diferentes formas de obtener el mismo resultado
    b. Simulación (set.seed(ZXY)) para validación
2. Significado de las variables
    a. `clm.count`: número de reclamos para una póliza.
    b. `clm.incurred`: Representa el acuerdo final, suma de reclamos individuales.
    c. `year`: Año calendario
    d. `exposure`: mide el tiempo que el carro estuvo expuesto durante el año calendario
    e. `age, driver.gender, marital.status, yrs.licensed, ncd.level nb.rb` (nueva poliza o renovación) `prior.claims`. Variables del conductor
    f. Variables del vehículo

Table 1.1. *Available Variables in Our Dataset*

Control	Driver	Vehicle	Geographic	Response
year	age	body.code	region	clm.count
exposure	driver.gender	driver.age		clm.incurred
row.id	marital.status	vehicle.value		
	yrs.licensed	seats		
	ncd.level	ccm		
	nb.rb	hp		
	prior.claims	length		
		width		
		height		
		fuel.type		

Descriptivas

Continuas

Table 1.2. *Summary Statistics for Continuous Variables*

Variable	Mean	Standard Deviation	Min.	Median	Max.
exposure	0.51	0.27	0.08	0.50	1.00
driver.age	44.55	10.78	18.00	44.00	93.00
yrs.licensed	3.21	1.89	1.00	3.00	10.00
vehicle.age	3.26	2.59	0.00	3.00	18.00
vehicle.value	23.50	8.89	4.50	22.10	132.60
ccm	1,670.69	390.12	970.00	1,560.00	3,198.00
hp	86.38	19.62	42.00	75.00	200.00
weight	1,364.22	222.01	860.00	1,320.00	2,275.00
length	4.32	0.36	1.80	4.28	6.95
width	1.78	0.10	1.48	1.74	2.12
height	1.81	0.09	1.42	1.82	2.52
prior.claims	0.83	1.33	0.00	0.00	21.00
clm.count	0.08	0.30	0.00	0.00	5.00
clm.incurred	66.52	406.23	0.00	0.00	11,683.58

Discretas o categóricas

Table 1.3. *Summary Statistics for Categorical Variables*

Variable	No. of Levels	Base Level	Most Common	Sample Levels
year	4	2013	2012	2010, 2011, 2012, 2013
nb.rb	2	NB	NB	NB, RB
drv.age	74	38	38	18, 19, 20, 21, 22, 23, 24, 25, and others
driver.gender	2	Male	Male	Female, Male
marital.status	4	Married	Married	Divorced, Married, Single, Widow
yrs.lic	8	1	2	1, 2, 3, 4, 5, 6, 7, 8+
ncd.level	6	1	1	1, 2, 3, 4, 5, 6
region	38	17	17	1, 10, 11, 12, 13, 14, 15, 16, and others
body.code	8	A	A	A, B, C, D, E, F, G, H
veh.age	15	1	1	0, 1, 10, 11, 12, 13, 14+, 2, and others
seats	5	5	5	2, 3, 4, 5, 6+
fuel.type	3	Diesel	Diesel	Diesel, Gasoline, LPG

Exposición, reclamos (conteo y cantidad), frecuencia y severidad, por año calendario

Table 1.4. *Exposure, Claim Counts, Claim Amounts, Frequency, and Severity by Calendar Year for the Entire Dataset*

Year	Exposure	Claim Count	Claim Amount	Frequency	Severity
2010	3,661.9	422	287,869	0.115	682.2
2011	5,221.7	551	314,431	0.106	570.7
2012	6,527.2	1,278	1,021,152	0.196	799.0
2013	5,386.2	1,180	1,087,735	0.219	921.8
Total	20,797.1	3,431	2,711,187	0.165	790.2

Frecuencia por año calendario y renovación-inicio

Para la tabla de “entrenamiento.”

Table 1.5. *Frequency by Calendar Year and New/Renewal Business Indicator for the Training Dataset*

	Exposure		Claim Count		Frequency (%)	
Year	NB	RB	NB	RB	NB	RB
2010	1,504.5	684.6	193	67	12.8	9.8
2011	2,105.9	1,042.4	271	69	12.9	6.6
2012	2,643.5	1,278.5	551	189	20.8	14.8
2013	2,248.6	964.9	526	137	23.4	14.2
Total	8,502.5	3,970.4	1,541	462	18.1	11.6

Notice that over the training data, the frequency for new business is equal to 18.1%, and for renewal business, it is equal to 11.6%. This looks like a significant difference; thus this variable is a good candidate to include in our models.

Entre años la diferencia es aún mayor de los 6.5 puntos en 2013. Es más común observar una mayor variabilidad entre periodos.

Exposición y frecuencia por edad del conductor

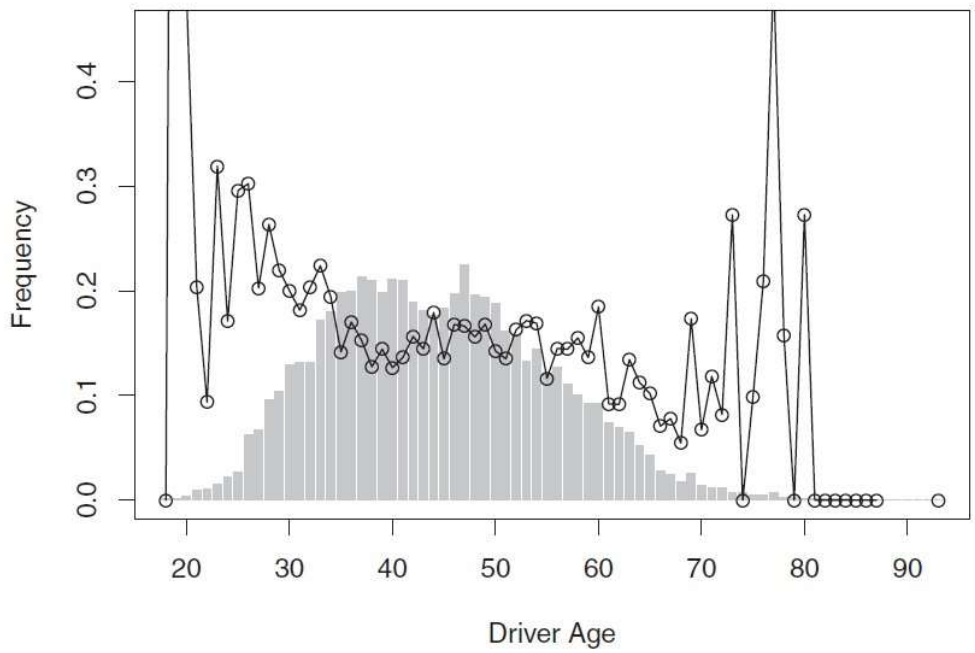

Fig. 1.2. Frequency and exposure by the driver age variable and for all three years of the training data. The y-axis has been restricted to the range [0, 0.45] to enhance the information shown. Four points have been omitted from the graph: (19, 184.6%), (20, 48.5%), (77, 50.3%), and (89, 92.3%).