Harshaanth Thiyagaraja Kumar

An analyst of an Insurance company notices that the number of accident claims has increased. He decides that it would be better to analyze the key factors that lead to claim and use them to evaluate the risk of claim for new contracts in the future. Using the pattern developed with past data, he is able to predict the risk of claim for new contracts considering the profile of the subscriber and the car insured.

Use the original dataset, please transform categorical variables into dummies (use man, no childen, agency-customer, no multiple car, sedan, manual and petrol as baseline), split data into training and validation datasets, and use stepwise for variable selection.

1. Choose the last model among models recommended by variable selection. Report the selected model's
   a) Regression Model (write down the model in three forms and paste the coefficients table),

Predictor	Estimate	Confidence Interval: Lowe	Confidence Interval: Upp	Odds	Standard Erro	Chi2-Statisti	P-Value
Intercept	-1.6473989	-1.938755362	-1.356042506	0.19255	0.14865397	122.8130553	1.53E-28
Gender_Woman	-0.7866305	-0.990497192	-0.582763838	0.455377	0.104015522	57.19331541	3.95E-14
Children_1	-0.660674	-0.950086255	-0.371261672	0.516503	0.147662046	20.01873465	7.67E-06
Children_2	-0.6184466	-0.868785481	-0.368107734	0.538781	0.127726262	23.4446642	1.29E-06
Children_equalormore	-1.1362748	-1.409091308	-0.863458299	0.321013	0.139194652	66.63795402	3.26E-16
Car category_SUV	-0.6348516	-0.906894975	-0.362808298	0.530014	0.138800172	20.92013475	4.79E-06
Car category_Sport	0.44278089	0.221578755	0.663983017	1.557031	0.112860304	15.39200869	8.74E-05
Fuel_Diesel	0.45538301	0.224852181	0.685913845	1.576777	0.117619933	14.9896606	0.000108

Logit (claim=1)=-1.65 -0.79woman-0.66children1-0.62children2-1.14childrenmorethan4 -0.63SUV+0.44sport+0.46diesel

Odds(claim=1)=e^(-1.65 -0.79woman-0.66children1-0.62children2-1.14childrenmorethan4 -0.63SUV+0.44sport+0.46diesel)

Odds(claim=1)=0.19*(0.46)^woman*(0.52)^children1*(0.54)^children2* *(0.32)^childrenmorethan4 *(0.53)^SUV*(1.56)^sport**(1.58)^diesel

P(claim=1)=1/(1+ e^-(-1.65 -0.79woman-0.66children1-0.62children2-1.14childrenmorethan4 -0.63SUV+0.44sport+0.46diesel))

   b) Explain the coefficients (all of them)

Women are 0.46 times less likely to make a claim than men holding other variables constant.

A customer with one child is 0.52 times less likely to make a claim compared to a customer with either no children or three children holding other variables constant.

A customer with two children is 0.54 times less likely to make a claim compared to a customer with either no children or three children holding other variables constant.

A customer with equal or more than 4 children is 0.32 times less likely to make a claim compared to a customer with either no children or three children holding other variables constant.

A customer owning an SUV is 0.53 times less likely to make a claim compared to a customer owning a sedan, with all other variables held constant.

A customer with a sports car is 1.56 times more likely to make a claim than a customer with a sedan, when other variables are controlled for.

A customer with a diesel car is 1.58 times more likely to make a claim than a customer with a petrol car, with all other variables constant.

c) In general, which types of customers are more likely to claim?

Men, no children or 3 children, sport car, diesel

2. What's the odds that a man with one child and a petrol sedan would claim? Show the formulas you use to answer this question. What is the classification of this man? (using cutoff value of p=0.5)

Logit (claim=1)=-1.65 -0.79woman-0.66children1-0.62children2-1.14childrenmorethan4 -0.63SUV+0.44sport+0.46diesel
=-1.65 -0.79*0-0.66*1-0.63*0+0.44*0+0.46*0=-2.31
Odds(claim=1)=0.099
-2.31<0
The man would not claim.

3. Using the cutoff value of 0.05 to answer question 2

Odds(claim=1)=0.099
Cutoff of logit: logit=ln(0.05/(1-0.05))=-2.94
-2.31>-2.94 The man would claim