

Example 1 - Results

Table 1: Number of policyholder who choose 'Other coverage' or 'Full coverage' as a function of predictors. Row percents indicate the proportion of policyholders. For quantitative regressors, the mean is shown.

	Other coverage $y = 0$	Full coverage $y = 1$	Total
Sex			
woman	498 (50.30%)	492 (49.70%)	990
man	2115 (70.27%)	895 (29.73%)	3010
Driving area			
rural	1906 (72.83%)	711 (27.17%)	2617
urban	707 (51.12%)	676 (48.88%)	1383
Vehicle use			
commercial	33 (84.62%)	6 (15.38%)	39
private	2580 (65.14%)	1381 (34.86%)	3961
Marital status			
single	467 (54.24%)	394 (45.76%)	861
married	2047 (68.85%)	926 (31.15%)	2973
other	99 (59.64%)	67(40.36%)	166
Age(years)	48.27	43.09	46.47
Seniority in company(years)	9.93	12.66	10. 88

Table 2: Logistic regression model results for `FullCoverage.csv` data set. Dependent variable is the choice of 'full coverage' ($y = 1$) versus 'other coverage' ($y = 0$).

Variable	Parameter estimate	Standard error	P-value	Odds-ratio
(Intercept)	-0.257	0.486	0.5959	0.773
men	-0.961	0.086	<0.001	0.382
urban	1.173	0.078	<0.001	3.230
private	1.065	0.469	0.0232	2.901
marital(married)	-0.083	0.096	0.3889	0.921
marital(other)	0.161	0.200	0.4212	1.175
age	-0.058	0.004	<0.001	0.944
seniority	0.133	0.007	<0.001	1.143
-2Log-Likelihood	$-2 \times (-2143.905)$			
Likelihood ratio test	875.483 (df=7, p-value<0.001)			
Pseudo-R ²	16.96%			

Table 3: Classification table for the logistic regression model in the `FullCoverage.csv` data set. Dependent variable is the choice of 'full coverage' ($y = 1$) versus 'other coverage' ($y = 0$).

	Predicted $y = 0$	Predicted $y = 1$
Observed $y = 0$	1858	755
Observed $y = 1$	313	1074

Table 4: Hosmer-Lemeshow test for the `FullCoverage.csv` data set. Dependent variable is the choice of 'full coverage' ($y = 1$) versus 'other coverage' ($y = 0$).

Group	Observed number of 'y=1'	Expected
1	22	20.83
2	16	44.45
3	30	67.28
4	31	90.20
5	128	114.98
6	225	140.23
7	219	166.60
8	245	198.92
9	237	241.17
10	234	302.33
X^2	Df	P(>Chi)
289.4653	8	<0.001

Figure 1: The ROC curve for the logistic regression model in the `FullCoverage.csv` data set. Dependent variable is the choice of 'full coverage' ($y = 1$) versus 'other coverage' ($y = 0$).

