Reading: GLM.Validation GLM_LogisticLorenz (Problem 1)
Model: Source text
Problem Type: Produce a table which can be used to plot the Lorenz curve for the output of a logistic regression.

Given An actuary has built a logistic regression model on a training data set and has run the model on a holdout sample of 50 observations. The model produced the following data:

Table of Predictions		Associated Outcomes*	
0.219	0.740	FALSE	TRUE
0.342	0.080	FALSE	FALSE
0.860	0.798	TRUE	TRUE
0.429	0.086	TRUE	FALSE
0.803	0.023	TRUE	FALSE
0.634	0.960	TRUE	TRUE
0.432	0.325	FALSE	FALSE
0.284	0.994	FALSE	TRUE
0.888	0.384	TRUE	FALSE
0.110	0.943	FALSE	TRUE
0.772	0.421	TRUE	FALSE
0.695	0.391	TRUE	TRUE
0.318	0.424	FALSE	FALSE
0.375	0.405	TRUE	FALSE
0.521	0.830	TRUE	TRUE
0.963	0.652	TRUE	TRUE
0.329	0.489	FALSE	TRUE
0.524	0.264	FALSE	TRUE
0.496	0.384	FALSE	FALSE
0.072	0.790	FALSE	TRUE
0.328	0.523	FALSE	FALSE
0.199	0.448	FALSE	FALSE
0.640	0.673	FALSE	FALSE
0.553	0.554	TRUE	TRUE
0.139	0.164	FALSE	FALSE

* TRUE means the event occurred while FALSE means it didn't.

Find Produce a table which can be used to plot the Lorenz curve for the output of a logistic regression.

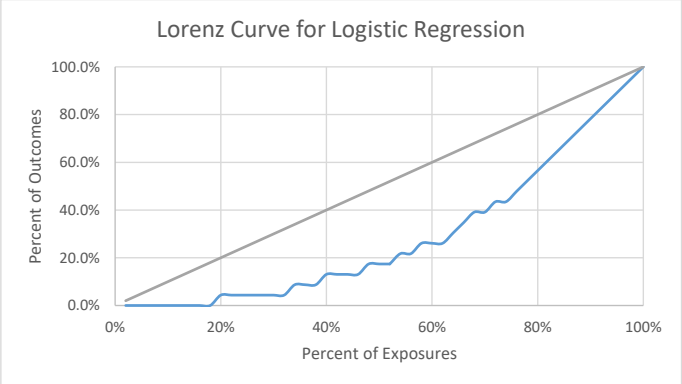
Solution GLM_LogisticLorenz (Solution 1)

We must order the risks by increasing prediction and then plot the cumulative exposure on the x-axis and the cumulative percentage of true outcomes on the y-axis.

Ranked Outcomes	
9	39
16	3
45	42
24	4
43	1
34	48
25	13
11	50
46	18
5	47
40	22
38	20
12	23
17	21
29	44
49	36
15	27
31	10
28	19
2	41
14	30
8	26
35	37
32	33
6	7

			Cumulative	
Rank	Prediction	Outcome	Exposure	TRUE
1	0.023	FALSE	2%	0.0%
2	0.072	FALSE	4%	0.0%
3	0.080	FALSE	6%	0.0%
4	0.086	FALSE	8%	0.0%
5	0.110	FALSE	10%	0.0%
6	0.139	FALSE	12%	0.0%
7	0.164	FALSE	14%	0.0%
8	0.199	FALSE	16%	0.0%
9	0.219	FALSE	18%	0.0%
10	0.264	TRUE	20%	4.3%
11	0.284	FALSE	22%	4.3%
12	0.318	FALSE	24%	4.3%
13	0.325	FALSE	26%	4.3%
14	0.328	FALSE	28%	4.3%
15	0.329	FALSE	30%	4.3%
16	0.342	FALSE	32%	4.3%
17	0.375	TRUE	34%	8.7%
18	0.384	FALSE	36%	8.7%
19	0.384	FALSE	38%	8.7%
20	0.391	TRUE	40%	13.0%
21	0.405	FALSE	42%	13.0%
22	0.421	FALSE	44%	13.0%
23	0.424	FALSE	46%	13.0%
24	0.429	TRUE	48%	17.4%
25	0.432	FALSE	50%	17.4%

			Cumulative	
Rank	Prediction	Outcome	Exposure	TRUE
26	0.448	FALSE	52%	17.4%
27	0.489	TRUE	54%	21.7%
28	0.496	FALSE	56%	21.7%
29	0.521	TRUE	58%	26.1%
30	0.523	FALSE	60%	26.1%
31	0.524	FALSE	62%	26.1%
32	0.553	TRUE	64%	30.4%
33	0.554	TRUE	66%	34.8%
34	0.634	TRUE	68%	39.1%
35	0.640	FALSE	70%	39.1%
36	0.652	TRUE	72%	43.5%
37	0.673	FALSE	74%	43.5%
38	0.695	TRUE	76%	47.8%
39	0.740	TRUE	78%	52.2%
40	0.772	TRUE	80%	56.5%
41	0.790	TRUE	82%	60.9%
42	0.798	TRUE	84%	65.2%
43	0.803	TRUE	86%	69.6%
44	0.830	TRUE	88%	73.9%
45	0.860	TRUE	90%	78.3%
46	0.888	TRUE	92%	82.6%
47	0.943	TRUE	94%	87.0%
48	0.960	TRUE	96%	91.3%
49	0.963	TRUE	98%	95.7%
50	0.994	TRUE	100%	100.0%



Alice: "As far as we know it's still unlikely you'll be asked to graph something in the exam environment. It's probably reasonable to create a table of data for a graph though."

"Also, this type of problem is a lot easier if you have the XLOOKUP function in Excel. However, not all versions of Excel have this functionality so we've shown a way of doing this another way. Pearson claims to have XLOOKUP in their software so you could practice with a small example in their online practice environment."