Bias Discussion 2.0 – Moving from Theory to Practice

CAS RPM
San Diego, California
15 March 2023

Christopher Cooksey, Guidewire Software
Senior Director of Advanced Analytics

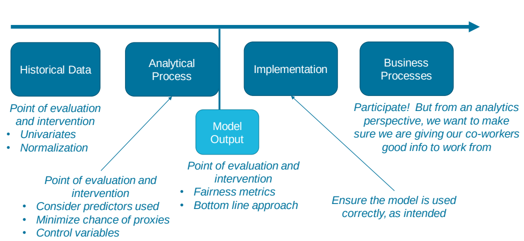
1

Agenda

1. Options for checking and intervening
2. Correlations are tricky
3. BOP example

2

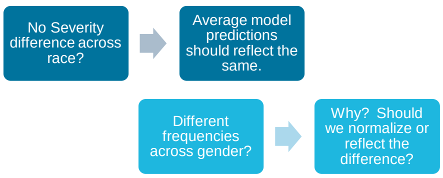
Where to Intervene?



3

Working from Knowledge

At the beginning of a modeling project, what is the univariate pattern across protected classes, given the target objective?



4

Effects of correlation

360 cells, each homo-geneous with respect to the target.

To model this, we identify attributes that usefully discriminate.

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

5

Effects of correlation

Attribute 1 identifies three equal-sized groups.

Averaging the experience gives a useful spread in the target.

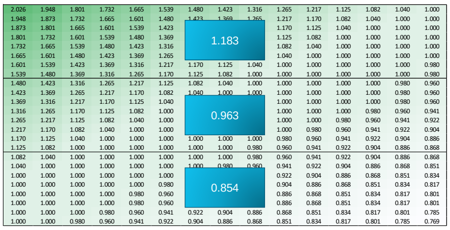
2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423				1.170	1.082				1.000	1.000				0.960
1.369		1.159		1.125	1.040		0.973		1.000	1.000		0.868		0.960
1.316				1.082	1.000				1.000	1.000				0.941
1.265				1.040	1.000				1.000	1.000				0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

6

Effects of correlation

Attribute 2 identifies three equal-sized groups, as well, but splits the data differently.

Averaging the experience gives a useful spread in the target.

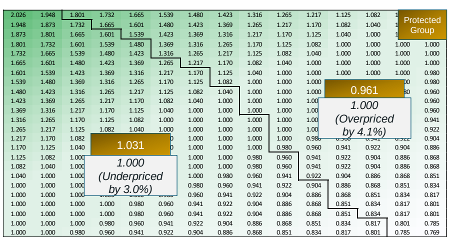


7

Effects of correlation

What if there is a protected class to keep track of?

And there is a difference in the data between groups.

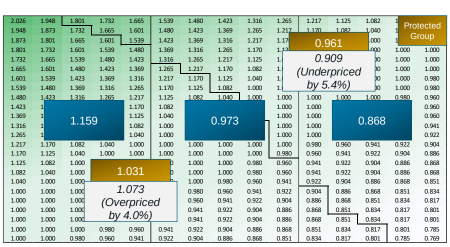


8

Effects of correlation

The uneven distribution changes the averages.

Which group is over or underpriced has flipped.

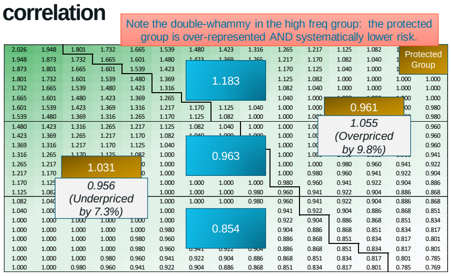


9

Effects of correlation

Attribute 2 is less correlated with the protected attribute.

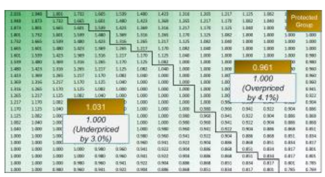
But only serves to distort the problem.



10

Effects of correlation – Lessons?

1. Averages do funny things as the distribution changes.
2. Excluding information and being blind to the relevant attributes does not help.
3. It is not obvious ahead of time the group impact for a given predictor.



4. Simple models distort more. It is instructive to see what happens if we look at each attribute as a single-variable model, and then do a simple ensemble.

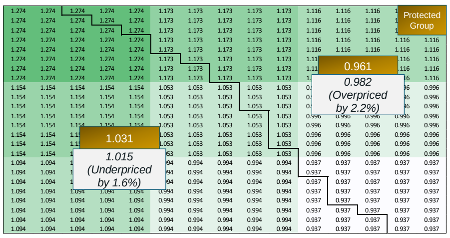
11

Effects of correlation – Lessons?

1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937

12

Effects of correlation – Lessons?

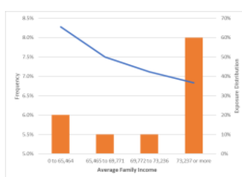


13

BOP Example *(Property coverage; Frequency target)*

Using the Zip-level field "Average Family Income" as our measure of bias. Lower incomes areas show higher property frequencies.

1. **Naïve GLM** – 10 predictors; no use of an income field.
2. **Controlled GLM** – added income field to control for its effect.
3. **Normalized GLM** – first normalized the data to remove the variation in frequency across income; then fit the 10-predictor GLM.



Note: these results are specific to one book of business for one target and should not be seen as generalizable. These GLMs were created quickly as illustrations, not as the best possible models of the data.

14

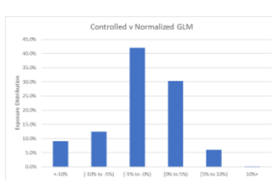
BOP Example *(Property coverage; Frequency target)*

Interestingly, the Controlled and Normalized GLMs give surprisingly different results.

In the graph we compare model predictions record by record.

Remember, normalizing data gives 100% credibility to the univariate pattern.

Avg Family Income	Exposure Distribution	Univariate Freq Rel	Controlled GLM Freq Rel
0 to 65,464	20%	1.212	1.130
65,465 to 69,771	10%	1.098	1.059
69,772 to 73,236	10%	1.042	0.993
73,237 or more	60%	1.000	1.000



15

BOP Example *(Property coverage; Frequency target)*

Does the Controlled GLM indicate a problem with the Naïve GLM?

One approach is to see if other fields lost predictive power.

Predictors 9 and 7 had the largest drop, but nothing eye-popping.

When included, Income had a low influence on the model.

There's always more than just the numbers. Do you get more curious when you note Pred7 is Protection Class?

	Naïve GLM	Controlled GLM	
Features	**Mean Absolute SHAP Values**	**Mean Absolute SHAP Values**	**Shift**
Predictor9	0.0393	0.0396	0.0002
Predictor2	0.0162	0.0157	-0.0005
Predictor1	0.0096	0.0095	-0.0001
Predictor4	0.0070	0.0070	0.0000
Predictor5	0.0060	0.0060	0.0000
Predictor10	0.0051	0.0051	0.0000
Predictor7	0.0045	0.0042	-0.0003
Predictor3	0.0041	0.0042	0.0001
Predictor8	0.0039	0.0043	0.0004
Ave Family Income	na	0.0031	na
Predictor6	0.0028	0.0028	0.0000

16

BOP Example *(Property coverage; Frequency target)*

Does the Controlled GLM indicate a problem with the Naïve GLM?

Another approach is look at the changes in the fitted coefficients.

Again, nothing dramatic pops, but let's look more closely at the two predictors already identified – a Building/Contents Indicator and Protection Class.

BuildingContents Indic	Expo Dist	*GLM1-low*	GLM1 Rel	GLM2 Rel	*GLM1-high*	*% Decrease*
Building and contents	40.9%	*1.492*	1.573	1.549	*1.659*	*-1.5%*
Building only	13.0%	*0.958*	1.043	1.037	*1.135*	*-0.6%*
Contents only	44.9%	*1.000*	1.000	1.000	*1.000*	*0.0%*
No building or contents	1.2%	*0.468*	0.765	0.758	*1.250*	*-0.9%*

Protection Class	Expo Dist	*GLM1-low*	GLM1 Rel	GLM2 Rel	*GLM1-high*	*% Decrease*
unknown	1.1%	*0.553*	0.948	0.935	*1.626*	*-1.4%*
0-2 (best)	22.6%	*0.871*	0.928	0.921	*0.988*	*-0.7%*
3	27.1%	*1.000*	1.000	1.000	*1.000*	*0.0%*
4	18.6%	*0.971*	1.031	1.032	*1.094*	*0.1%*
5	10.8%	*1.080*	1.156	1.138	*1.237*	*-1.5%*
6+ (worst)	19.9%	*1.052*	1.120	1.096	*1.191*	*-2.1%*

Hmmm…areas with the worst protection class rating see higher predicted frequencies when we naively ignore income.

17

BOP Example *(Property coverage; Frequency target)*

Are the differences here large? No.

Are the differences statistically significant? Not even close.

Should you leave here today thinking I've shown some pattern that holds in general for other books of business? Please, no.

If you do this analysis and see something *like* this, is it worth thinking more about? I think so.

Protection Class	Expo Dist	*GLM1-low*	GLM1 Rel	GLM2 Rel	*GLM1-high*	*% Decrease*
unknown	1.1%	*0.553*	0.948	0.935	*1.626*	*-1.4%*
0-2 (best)	22.6%	*0.871*	0.928	0.921	*0.988*	*-0.7%*
3	27.1%	*1.000*	1.000	1.000	*1.000*	*0.0%*
4	18.6%	*0.971*	1.031	1.032	*1.094*	*0.1%*
5	10.8%	*1.080*	1.156	1.138	*1.237*	*-1.5%*
6+ (worst)	19.9%	*1.052*	1.120	1.096	*1.191*	*-2.1%*

Hmmm…areas with the worst protection class rating see higher predicted frequencies when we naively ignore income.

18

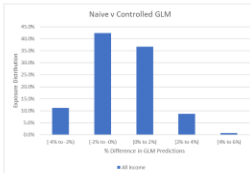
BOP Example *(Property coverage; Frequency target)*

Does the Controlled GLM indicate a problem with the Naïve GLM?

How about the model output? Do the predictions show anything?

The negative (left) end of the chart is when the Naïve GLM prediction is lower than the Controlled GLM prediction.

Overall, the predictions are quite similar. This is good.



19

BOP Example *(Property coverage; Frequency target)*

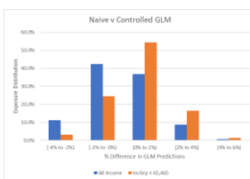
Does the Controlled GLM indicate a problem with the Naïve GLM?

How about the model output? Do the predictions show anything?

With model output, you can filter down to the group we are interested in checking.

This shows that lower income areas see a higher frequency prediction at a disproportionate rate when a naïve approach is used.

Big? No. There? Yes.
(in this book, this example…)



20

BOP Example *(Property coverage; Frequency target)*

Thoughts on working through this example:

- Checking the correlation with Income wouldn't have been sufficient.

Predictor	Correl	SHAP	SHAP Change	Change in Rels
Predictor8	14.8%	0.0039 (second lowest)	+0.0004 (biggest increase)	One level out of 10 saw a 2% move in relativity. Nothing systematic.
Protection Class	14.7%	0.0045	-0.0003 (second biggest decrease)	And yes, an interesting, though small, movement in the relativites.
State	11.0%	0.0028 (lowest)	No change	One state shifted 2%.
Building/ Contents	Not even top 3	0.0162 (second highest)	-0.0005 (biggest decrease)	One level dropped 2%.

- Checking multiple measures – correl, SHAP, model output distributions – gives much more context.
- In the end, this requires thought. Doubtful about using automatic measures.

21

Questions?

Christopher Cooksey
Senior Director of Advanced Analytics
ccooksey@guidewire.com

Guidewire Software

22