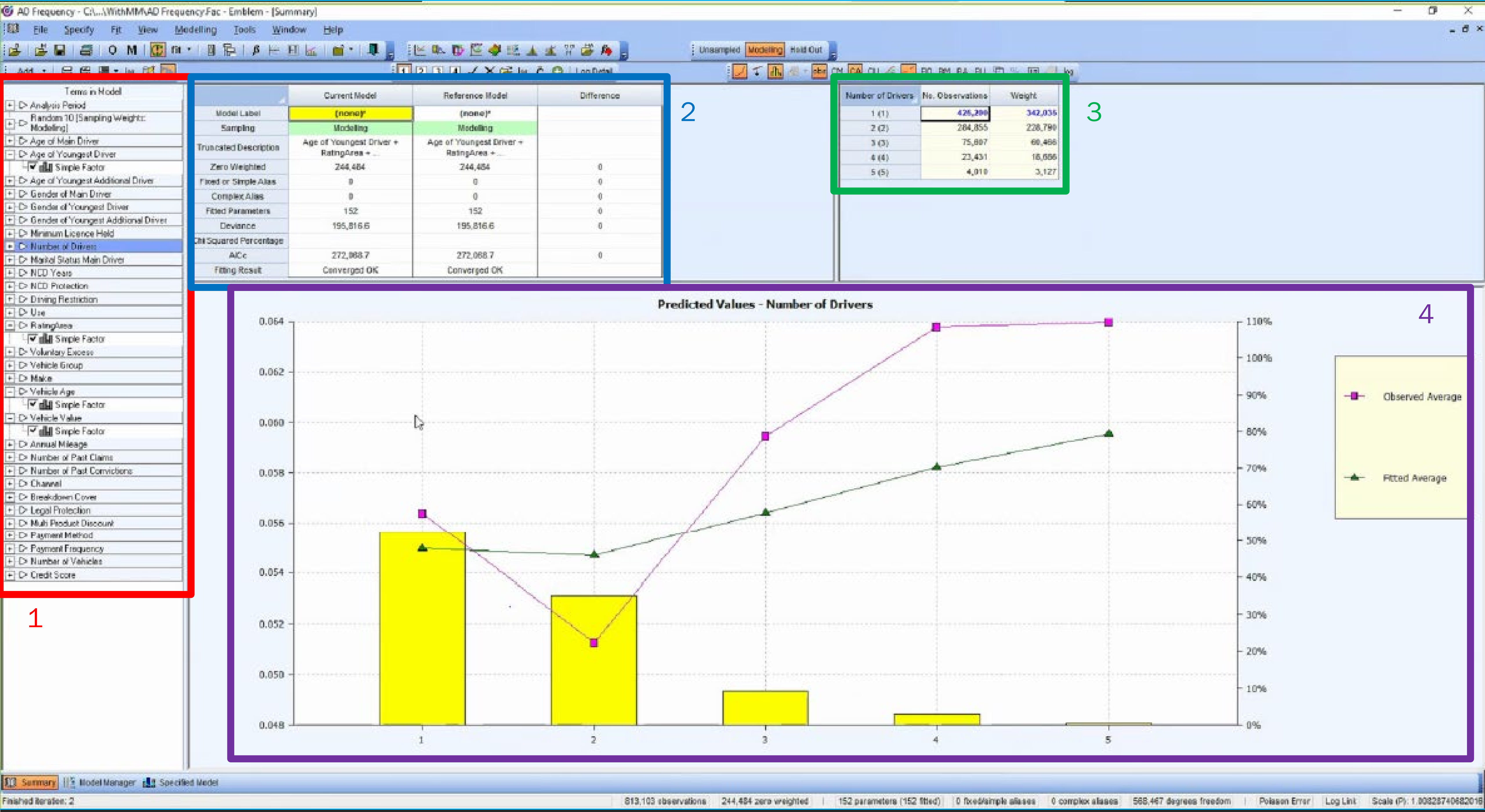
GLM Concepts in EMBLEM

SAM KLOESE, ACAS, CSPA

8/25/2020

Background

- EMBLEM is a prevalent predictive modeling tool in the insurance industry
- Benjamin Williams from Towers Watson presented a brief intro to EMBLEM on 4/28
- This is a further deep dive to really dig into GLM concepts using EMBLEM screenshots
- Most pictures are direct screenshots from Benjamin William’s presentation
- Augmented with other sources
  - 2019 GIS Regulator Advanced Modeling Training
  - Simpler theoretical examples built in Excel to demonstrate concepts
  - References to the CAS GLM monograph

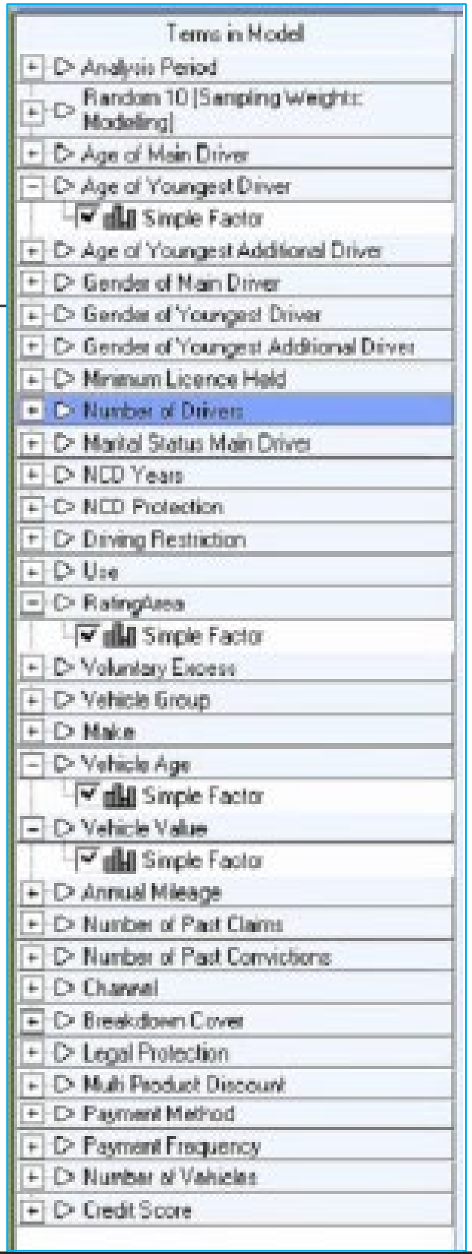

AD Frequency - C:\...\WithMM\AD Frequency.Fac - Emblem - [Summary]
File Specify Fit View Modelling Tools Window Help
Unsampled Modelling Hold Out
Terms in Model
Analysis Period
Random 10 [Sampling Weights: Modelling]
Age of Main Driver
Age of Youngest Driver
Simple Factor
Age of Youngest Additional Driver
Gender of Main Driver
Gender of Youngest Driver
Gender of Youngest Additional Driver
Minimum Licence Held
Number of Drivers
Marital Status Main Driver
NCD Years
NCD Protection
Driving Restriction
Use
RatingArea
Simple Factor
Voluntary Excess
Vehicle Group
Make
Vehicle Age
Simple Factor
Vehicle Value
Simple Factor
Annual Mileage
Number of Past Claims
Number of Past Convictions
Channel
Breakdown Cover
Legal Protection
Multi Product Discount
Payment Method
Payment Frequency
Number of Vehicles
Credit Score
1
Current Model
Reference Model
Difference
Model Label
(none)*
(none)*
Sampling
Modelling
Modelling
Truncated Description
Age of Youngest Driver + RatingArea + ...
Age of Youngest Driver + RatingArea + ...
Zero Weighted
244,484
244,484
0
Fixed or Simple Alias
0
0
0
Complex Alias
0
0
0
Fitted Parameters
152
152
0
Deviance
195,816.6
195,816.6
0
Chi Squared Percentage
AICc
272,068.7
272,068.7
0
Fitting Result
Converged OK
Converged OK
2
Number of Drivers
No. Observations
Weight
1 (1)
426,200
342,035
2 (2)
284,855
228,790
3 (3)
75,807
60,466
4 (4)
23,431
18,686
5 (5)
4,010
3,127
3
Predicted Values - Number of Drivers
0.064
0.062
0.060
0.058
0.056
0.054
0.052
0.050
0.048
110%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
2
3
4
5
Observed Average
Fitted Average
4
Summary
Model Manager
Specified Model
Finished iteration: 2
813,103 observations
244,484 zero weighted
152 parameters (152 fitted)
0 fixed/simple aliases
0 complex aliases
568,467 degrees freedom
Poisson Error
Log Link

```
anorex.1 <- glm(Postwt ~ Prewt + Treat + offset(Prewt), family = gaussian, data = anorexia)
```

1. Terms in Model

- This is a list of candidate Independent variables
  - There are 32 options here
  - These would all have been columns in the input data
  - The columns with metrics are not listed here
    - During data input step, you’d specify the target metric fields
- EMBLEM is much more “point and click” than R or Python
  - Put a check mark next to the variables you want to include
  - Then click “fit”
  - R would require typing out the names of the columns
    - Example:

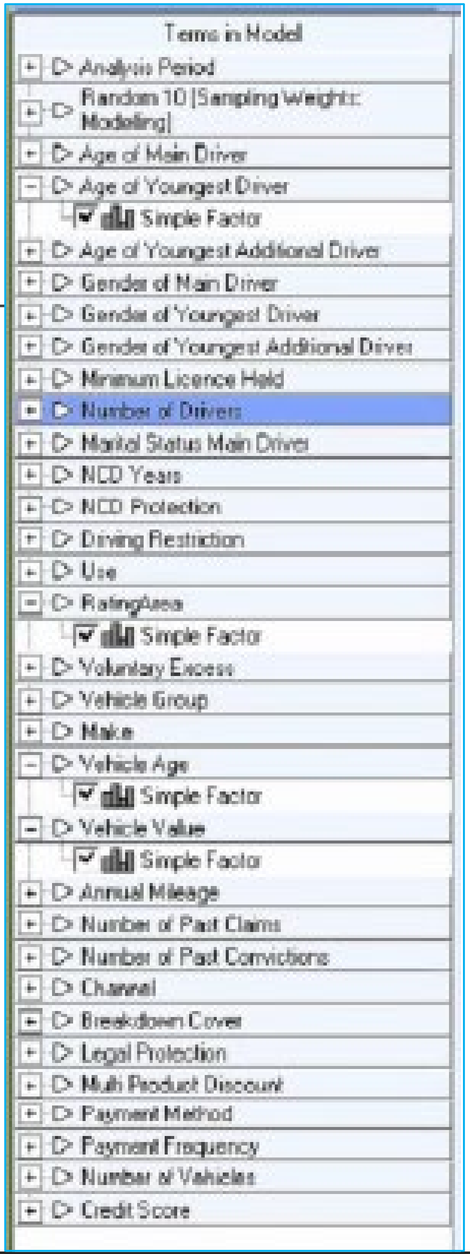
```
Freq_model <- glm(Frequency ~ Age_Youngest + RatingArea + VehicleAge + VehicleValue),
                  family = poisson(link = “log”),
                  data = mydata)
```



```
anorex.1 <- glm(Postwt ~ Prewt + Treat + offset(Prewt), family = gaussian, data = anorexia)
```

1. Terms in Model

- Types of Terms in model
  - Potential Rating Variables
    - Variables that you would consider putting in your rating plan
  - Control Variables
    - Variables we would not use in rating, but we include so certain effects do not influence our potential rating parameter estimates
      - (CAS GLM Paper Section 5.1.3)
    - Examples are easier to explain
      - Year as a control variable when undeveloped losses are used
      - State as a control variable in a countrywide model because loss level varies by state
  - Offset Variables
    - Variables with pre-determined factor that we want our GLM to “work around”, not recalculate
      - Limits & Deductibles (CAS GLM Paper Section 9.1)
      - Territory (CAS GLM Paper Section 9.2)



2. Quick Model Comparison

- Here, Curr Model = Ref Model
  - No difference given for “Truncated Description”
  - Goodness of Fit statistics are equal
- This table compares the latest fitted model with a reference model
- You can save up to 4 reference models
- Example of a nested model comparison on upcoming slide

	Current Model	Reference Model	Difference
Model Label	(none)*	(none)*	
Sampling	Modelling	Modelling	
Truncated Description	Age of Youngest Driver + RatingArea + ...	Age of Youngest Driver + RatingArea + ...	
Zero Weighted	244,484	244,484	0
Fixed or Simple Alias	0	0	0
Complex Alias	0	0	0
Fitted Parameters	152	152	0
Deviance	195,816.6	195,816.6	0
Chi Squared Percentage			
AICc	272,068.7	272,068.7	0
Fitting Result	Converged OK	Converged OK	

2. Quick Model Comparison

- Here, Curr Model = Ref Model

- No Aliases!
  - Simple Alias: 2 columns provide the same info (or extremely correlated)
    - Example:
      - Driver Birth year and Driver Age
      - Married Indicator and Single Indicator
  - Complex Alias: 2 or more columns together provide the same info as another column
    - Number of Vehicles and Number of Drivers together explain "Vehicle Driver Ratio"

	Current Model	Reference Model	Difference
Model Label	(none)*	(none)*	
Sampling	Modeling	Modeling	
Truncated Description	Age of Youngest Driver + RatingArea + ...	Age of Youngest Driver + RatingArea + ...	
Zero Weighted	244,484	244,484	0
Fixed or Simple Alias	0	0	0
Complex Alias	0	0	0
Fitted Parameters	152	152	0
Deviance	195,816.6	195,816.6	0
Chi Squared Percentage			
AICc	272,068.7	272,068.7	0
Fitting Result	Converged OK	Converged OK	

2. Quick Model Comparison

- Here, Curr Model ≠ Ref Model
- Difference column tells us many things
  - We added # Drivers
  - We added 4 parameters
    - (n – 1) parameters for a categorical field with n levels.
    - We had 1, 2, 3, 4, or 5 drivers, so 4 parameters.

	Current Model	Reference Model	Difference
Model Label	[none]*	(none)*	
Sampling	Modelling	Modelling	
Truncated Description	Age of Youngest Driver + Number of Drivers + ...	Age of Youngest Driver + RatingArea + ...	+ Number of Drivers
Zero Weighted	244,484	244,484	0
Fixed or Simple Alias	0	0	0
Complex Alias	0	0	0
Fitted Parameters	156	152	4
Deviance	195,727.8	195,816.6	-88.8367
Chi Squared Percentage		Sub-Model	0.0%
AICc	272,944.2	272,968.7	-24.52389
Fitting Result	Converged OK	Converged OK	

Note: EMBLEM considers all fields “Categorical” until you tell it otherwise!

2. Quick Model Comparison

- Here, Curr Model ≠ Ref Model
- Difference column tells us many things
  - Deviance went down
    - Always does when adding parameters
  - Chi Squared Percentage is low
    - Implies the larger model is better
    - Chi Squared Percentage is blank if this is not a nested model
  - AICc went down
    - Penalized measure of deviance.
    - Better than deviance to look at

	Current Model	Reference Model	Difference
Model Label	(none)*	(none)*	
Sampling	Modelling	Modelling	
Truncated Description	Age of Youngest Driver + Number of Drivers + ...	Age of Youngest Driver + RatingArea + ...	+ Number of Drivers
Zero Weighted	244,484	244,484	0
Fixed or Simple Alias	0	0	0
Complex Alias	0	0	0
Fitted Parameters	156	152	4
Deviance	195,727.8	195,816.6	-88.8367
Chi Squared Percentage		Sub-Model	0.0%
AICc	272,044.2	272,068.7	-24.52389
Fitting Result	Converged OK	Converged OK	

Note: BIC is available in EMBLEM but not shown here

3. Volume Summary

- No. Observations is the # rows in input data
- Weight is the sum of the weight metric in our input data
- Weight is the denominator of the target variable
- Target (dependent) variable
  - Frequency: Claim Count / Earned Exposures
  - Severity: Loss Dollars / Claim Count
  - Pure Premium: Loss Dollars / Earned Exposures

Number of Drivers	No. Observations	Weight
1 (1)	426,296	342,035
2 (2)	284,855	228,790
3 (3)	75,807	60,466
4 (4)	23,431	18,688
5 (5)	4,018	3,127

Note: This is likely a frequency model, which means weight is likely earned exposures

3. Volume Summary

- Possible explanation for Weight < No. Observations
  - Perhaps each row is one policy year
  - Not all policies were insured for the entire policy year
  - Therefore the Earned Exposure column is between 0 and 1 for each row
- Weight does not have to be less than Observations
  - Data could be aggregated up to unique class level
  - Weight should exceed row count in this situation

Number of Drivers	No. Observations	Weight
1 (1)	426,290	342,035
2 (2)	284,855	228,790
3 (3)	75,807	60,466
4 (4)	23,431	18,688
5 (5)	4,018	3,127

Note: This is likely a frequency model, which means weight is likely earned exposures

3. Volume Summary

- How the data was aggregated impacts No. Observations
- Consider the following 2 ways to arrange the same claims experience

Split by Policy

Policy	Gender	Age	Claim	Earned Exposure
1	M	16	1	1
2	F	16	0	0.5
3	M	16	0	0.25
4	F	16	0	1
5	M	25	0	1
6	F	25	0	1
7	M	25	0	1
8	F	25	0	1
9	F	60	0	0.75
10	F	60	0	1
11	F	60	0	1
12	F	60	0	1

No. Observations	12
Earned Exposures	10.5

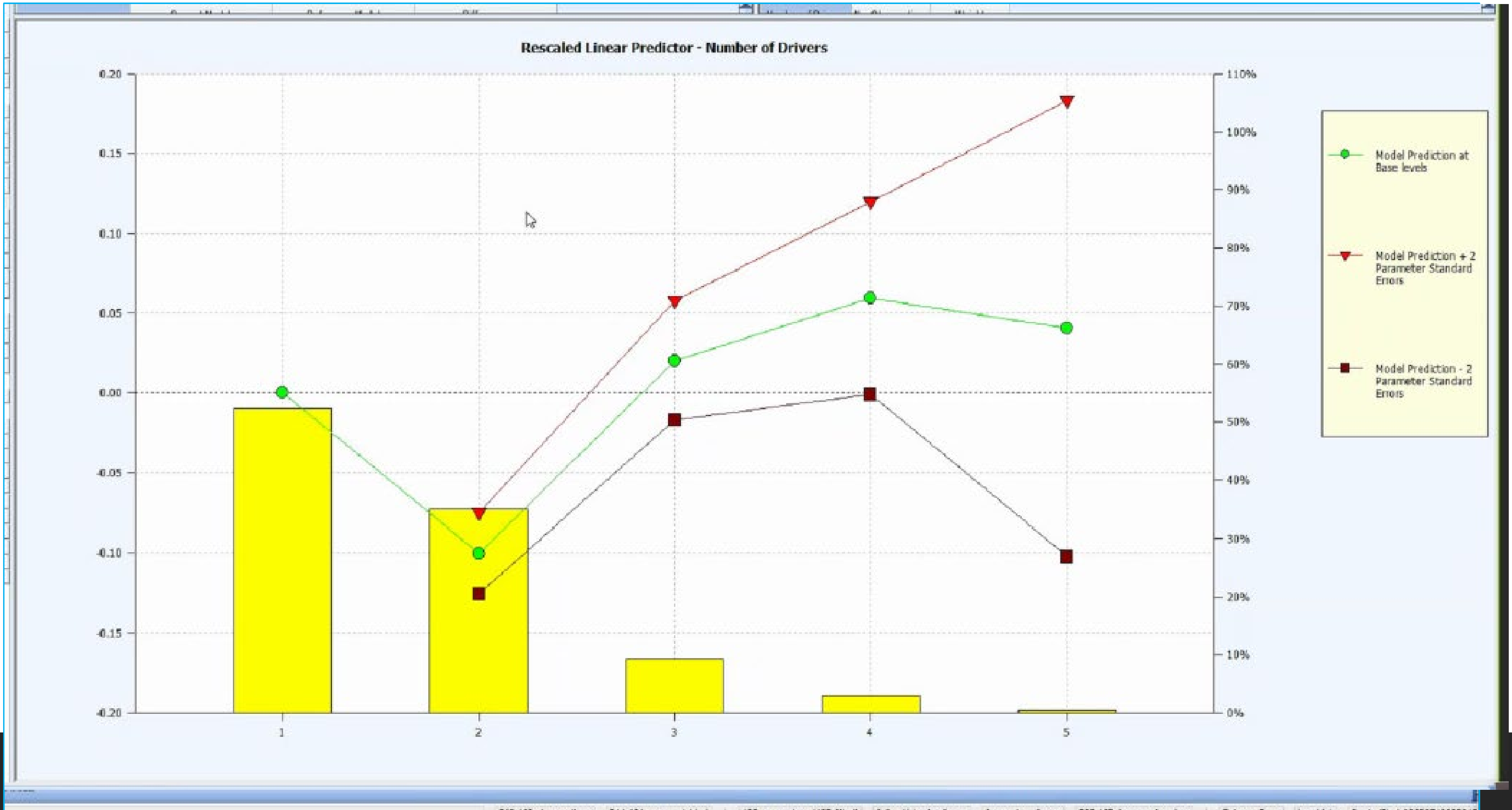
Summarized to Rating Class

Gender	Age	Claim	Earned Exposure
M	16	1	1.25
F	16	0	1.5
M	25	0	2
F	25	0	2
F	60	0	3.75

No. Observations	5
Earned Exposures	10.5

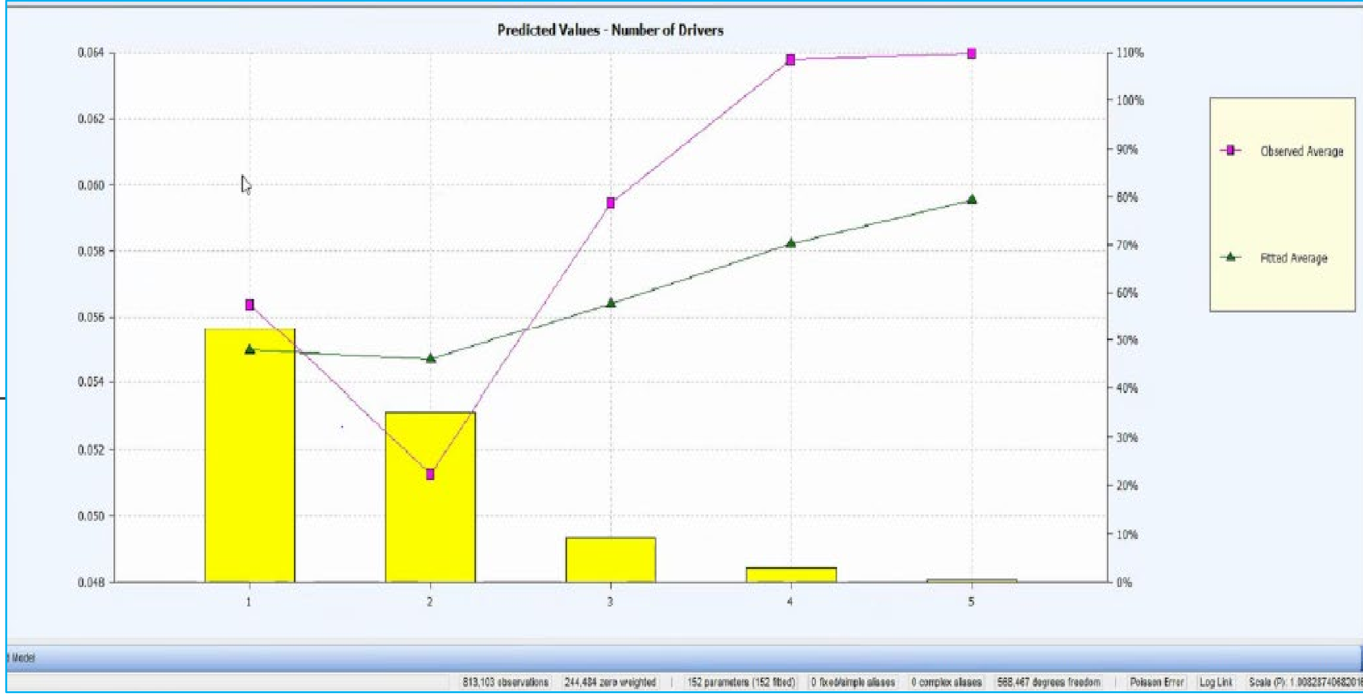
4. Graphs by Variable

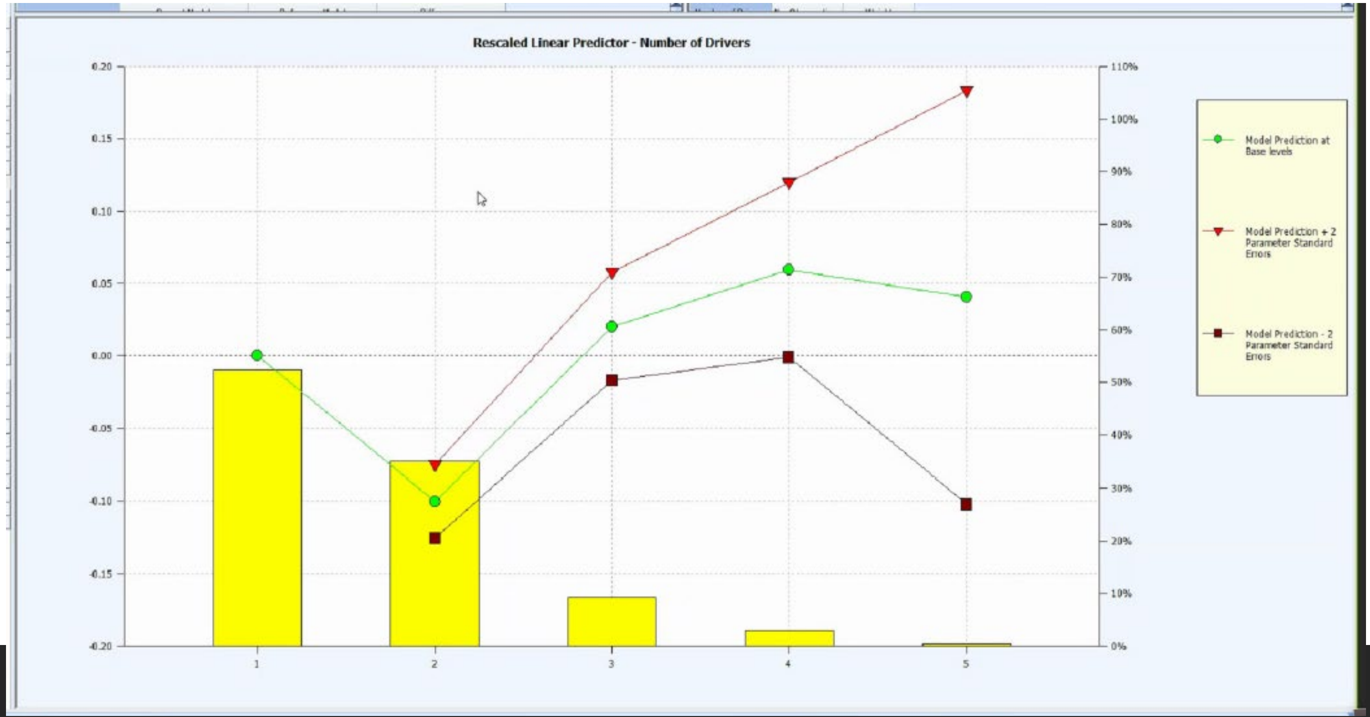
- Lines you can put on the graph
  - Observed Average
    - Univariate empirical average
  - Fitted Average
    - Univariate average of predictions
  - Parameter Info
    - Model Prediction at Base Levels (Point Estimate)
    - Model Prediction at Base Levels -2 SE (Lower Bound)
    - Model Prediction at Base Levels +2 SE (Upper Bound)



4. Graphs by Variable

- X Axis
  - Levels of the variable
- Primary Y Axis (Left Side)
  - 4 Target Variable Unit Options
    - Linear Predictor vs. Fitted Value
    - Unscaled vs. Rescaled
    - Examples on upcoming slide
  - “Rescaled Fitted Value” graphs indicated factor
- Secondary Y Axis (Right Side)
  - Weight Volume
    - Useful for identifying where the data is “thin”





4. Graphs by Variable

- Refresher on Linear Predictor with Log Link

Frequency at Base Level	5.0%	β (2.996)

Gender	Factor	β
Male	1.050	0.049
Female	1.000	-

Age	Factor	β
≤20	2.000	0.693
20-30	1.000	-
31-60	0.800	(0.223)
60+	1.200	0.182

	Predicted Value at Base Levels	Predicted Value Rescaled at Base Levls <Indicated>	Linear Predictor at Base Levels	Linear Predictor Rescaled at Base Levels	Assumption
Gender					
Male	5.3%	1.050	(2.947)	0.049	Assume Age 20-30
Female	5.0%	1.000	(2.996)	-	Assume Age 20-30
Age					
≤20	10.0%	2.000	(2.303)	0.693	Assume Gender = F
20-30	5.0%	1.000	(2.996)	-	Assume Gender = F
31-60	4.0%	0.800	(3.219)	(0.223)	Assume Gender = F
60+	6.0%	1.200	(2.813)	0.182	Assume Gender = F

exp(-2.996+.182) = .06

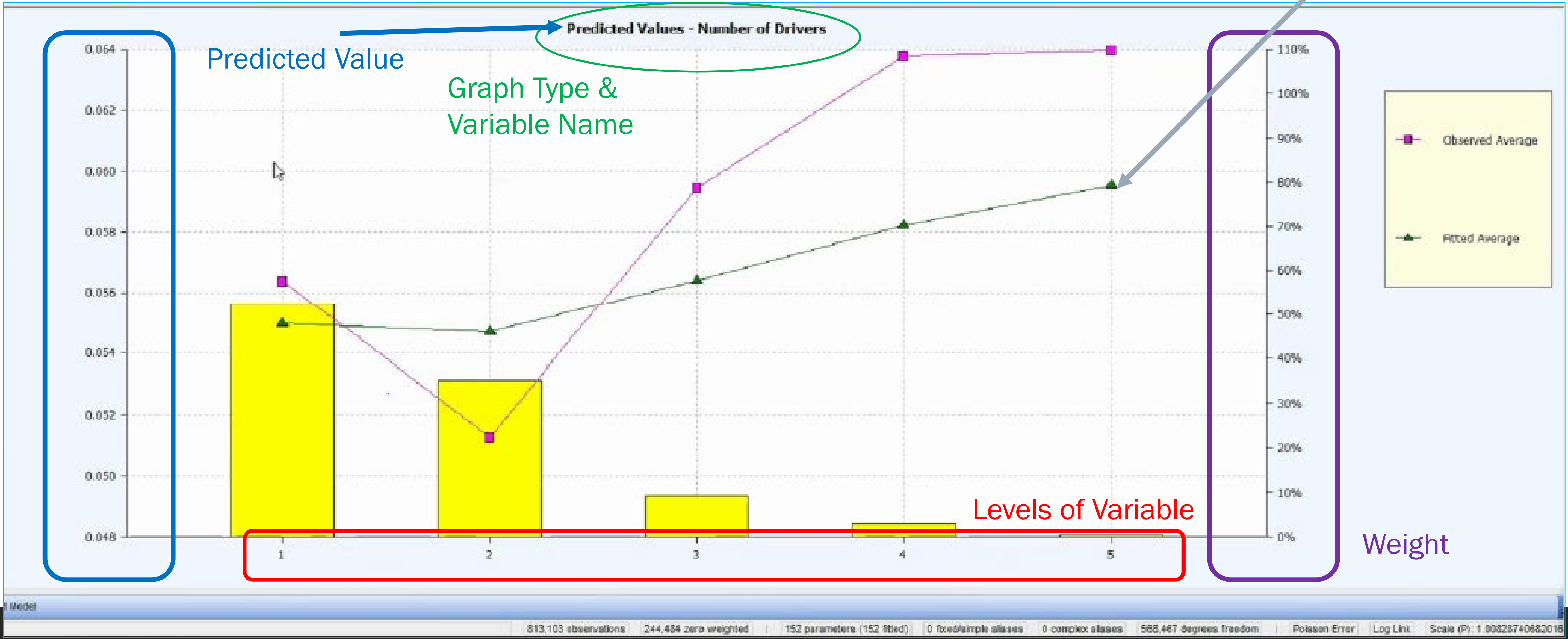
exp(.182) = 1.2

-2.996+.182 = -2.813

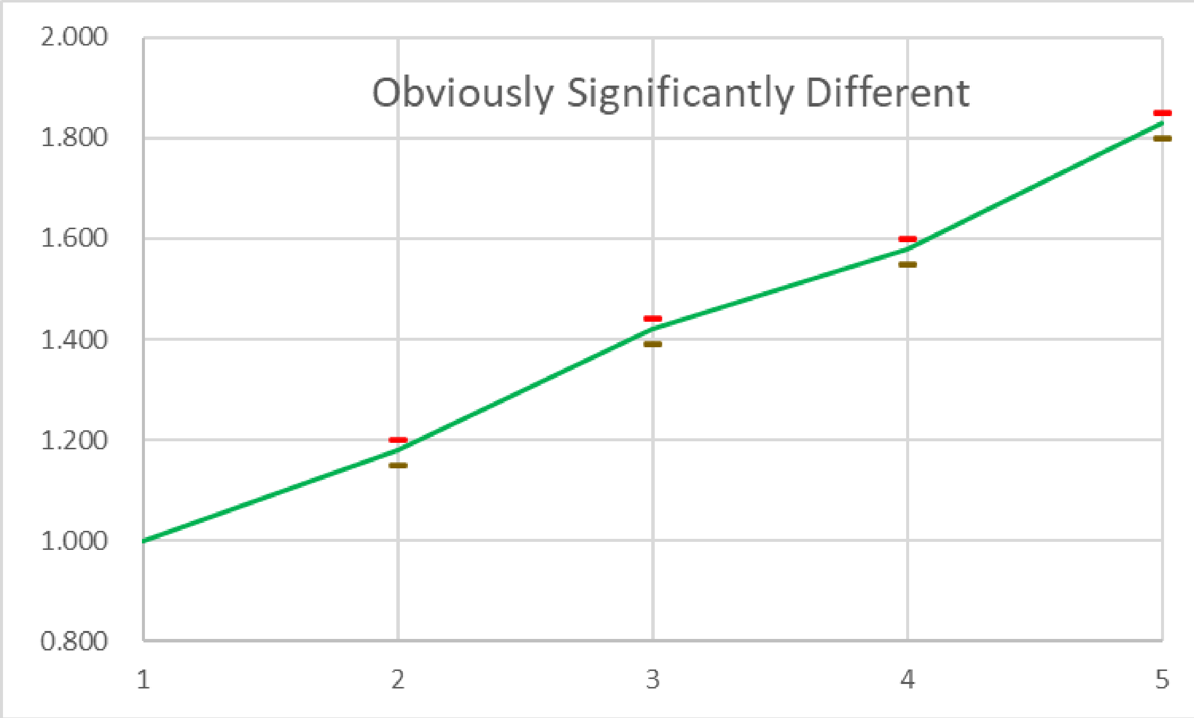
β = .182

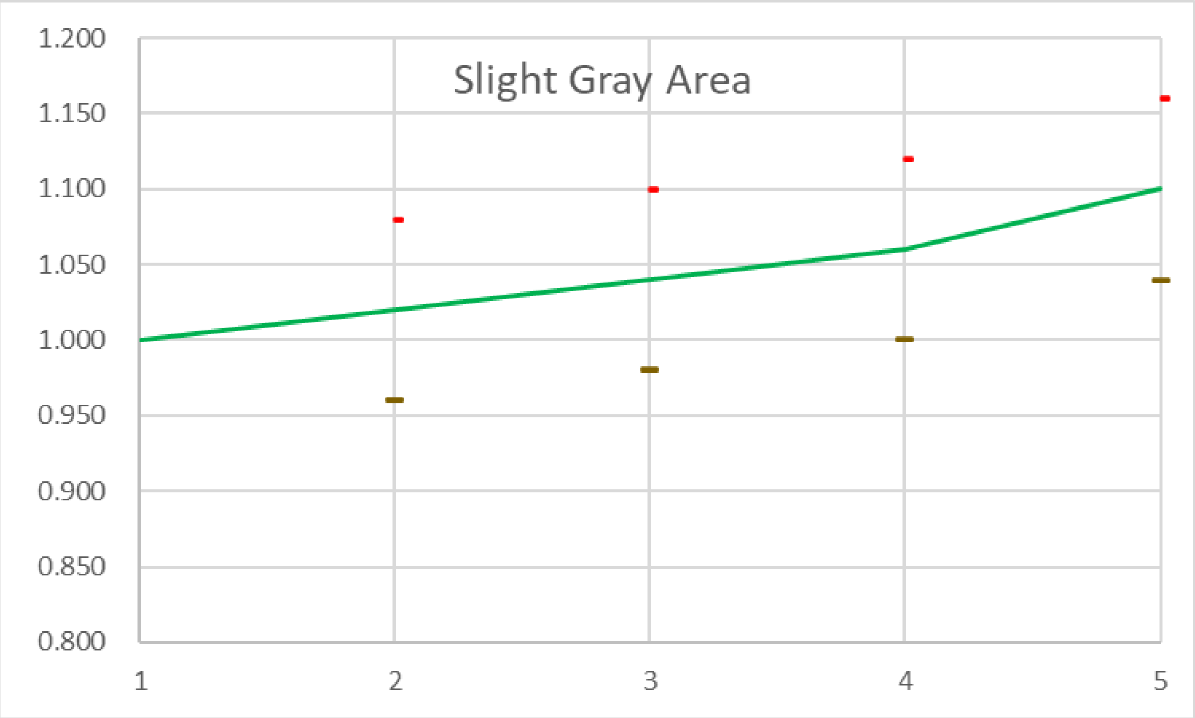
4. Graphs by Variable

Note: Fitted Average takes into consideration ALL modeled factors



4. Graphs by Variable

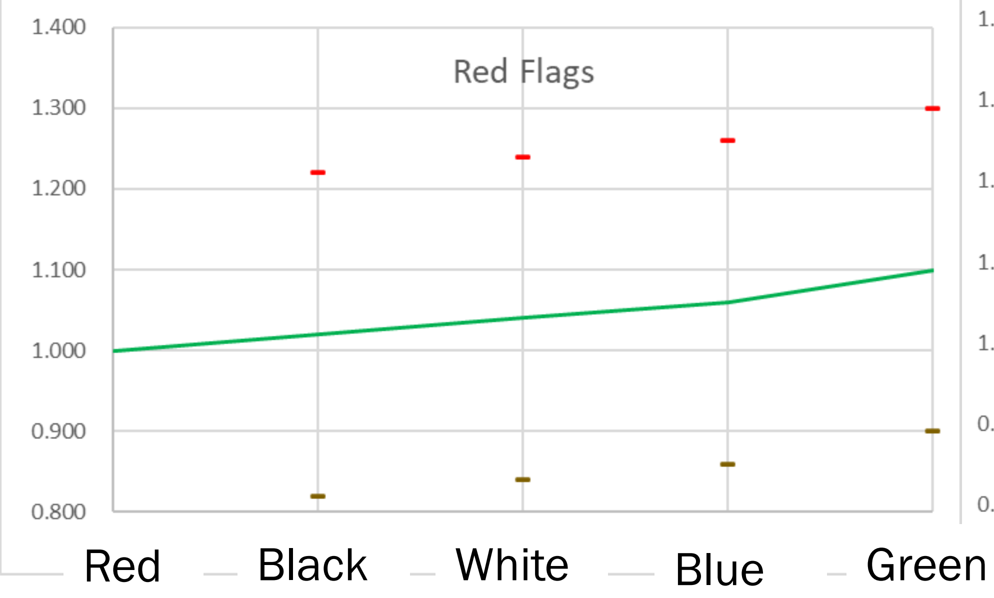


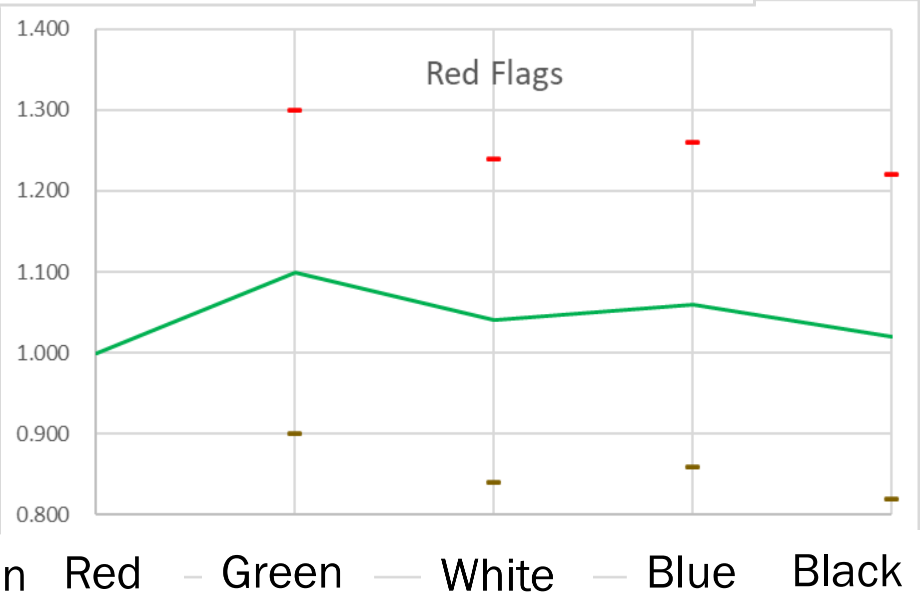


Analyzing graphs of confidence intervals by level can be illuminating

Red flags

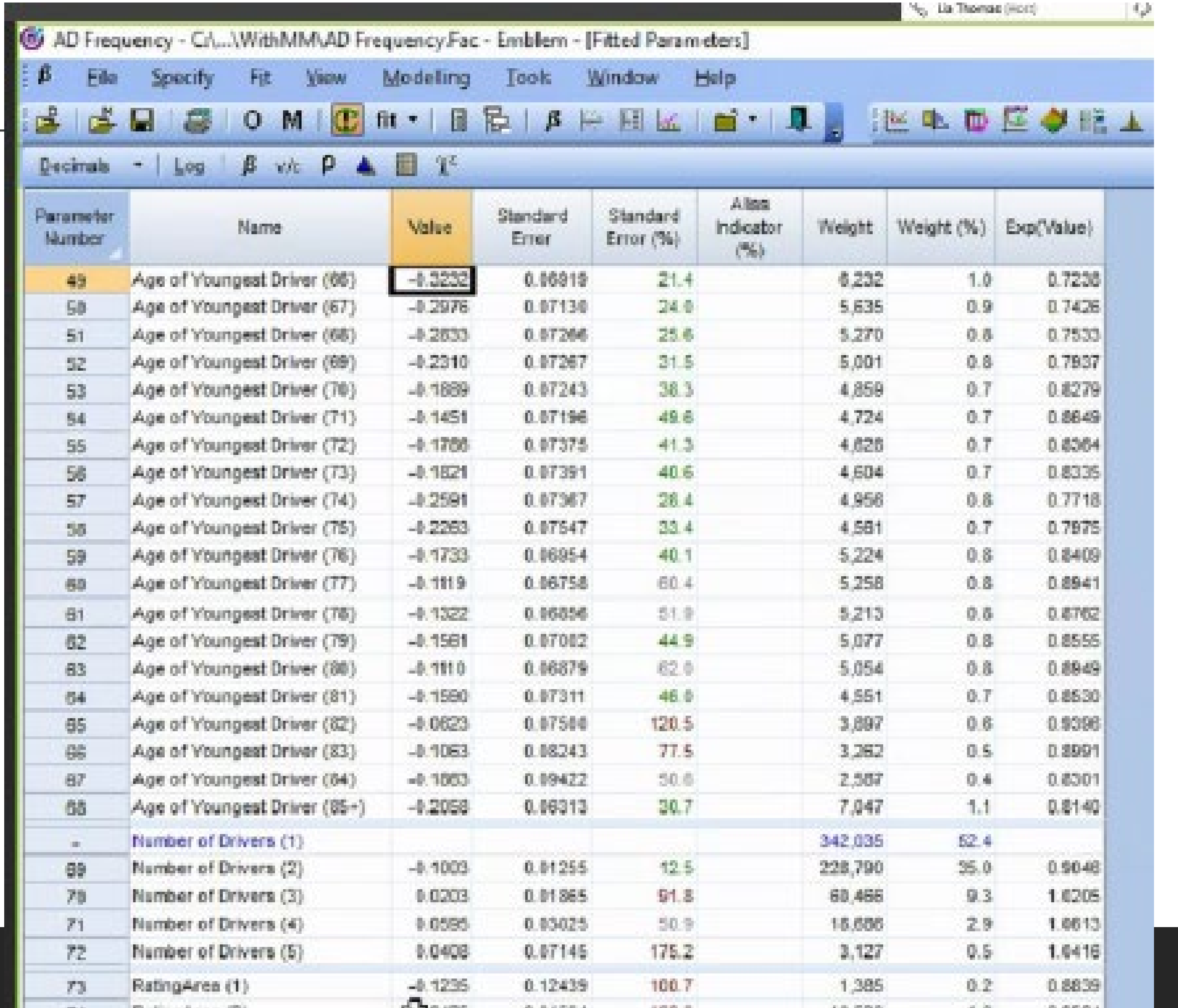
- Every β point estimate is within the confidence interval of the other levels
- The ordering of X axis is completely arbitrary





β Pages

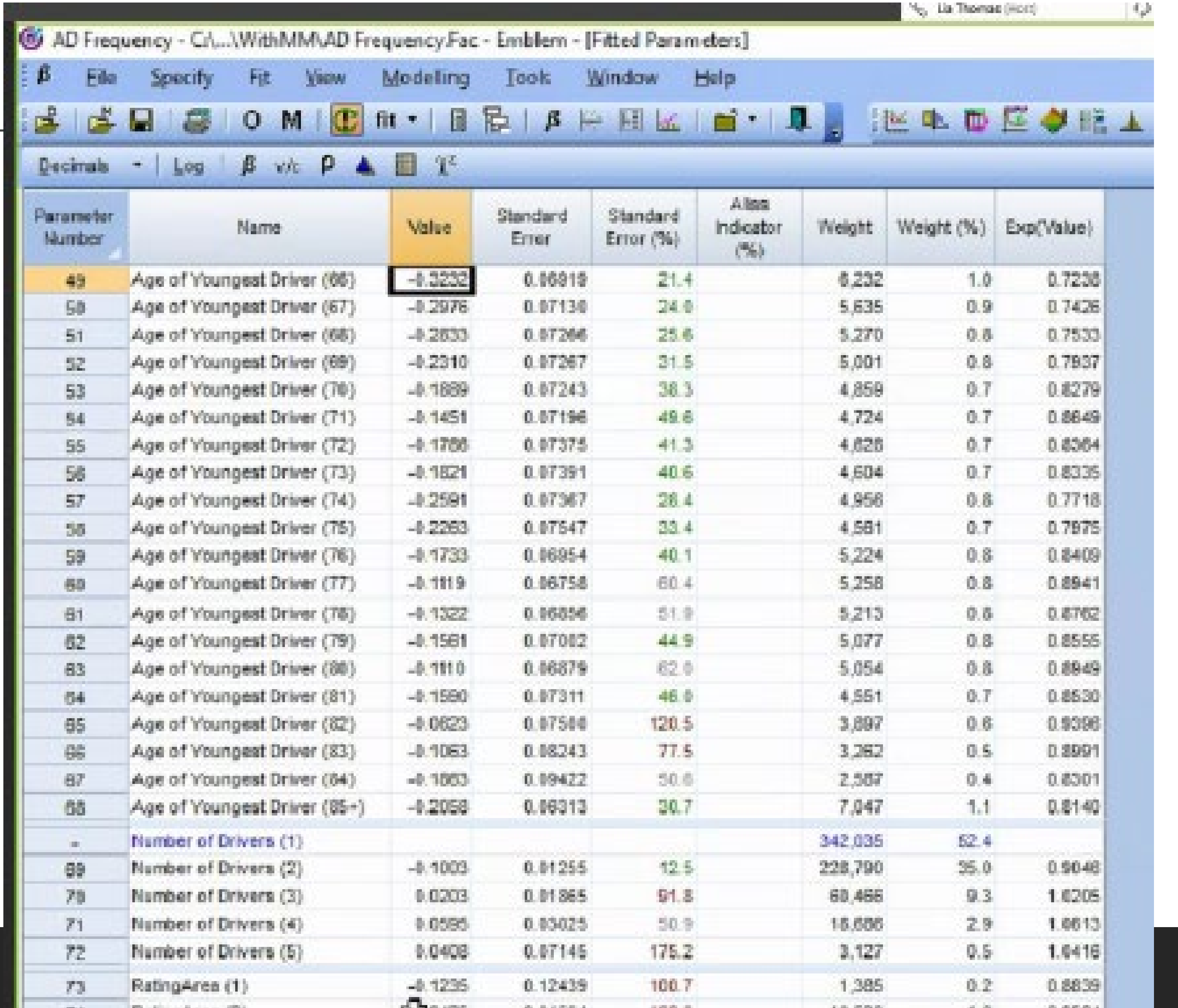
- Parameter Number
  - Each non-base level of a categorical variable is a parameter
  - Curve fit continuous variables will have a number based on coefficients in the curve fit
- Value is fitted β
- Standard Error helps describe size of confidence interval
  - Standard Error % is SE / abs(Value)
    - Low % green
    - High % red



Parameter Number	Name	Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
49	Age of Youngest Driver (66)	-0.3232	0.06919	21.4		6,232	1.0	0.7238
50	Age of Youngest Driver (67)	-0.2976	0.07136	24.0		5,635	0.9	0.7426
51	Age of Youngest Driver (68)	-0.2833	0.07266	25.6		5,270	0.8	0.7533
52	Age of Youngest Driver (69)	-0.2310	0.07267	31.5		5,001	0.8	0.7937
53	Age of Youngest Driver (70)	-0.1889	0.07243	38.3		4,859	0.7	0.8279
54	Age of Youngest Driver (71)	-0.1451	0.07196	49.6		4,724	0.7	0.8649
55	Age of Youngest Driver (72)	-0.1786	0.07375	41.3		4,826	0.7	0.8364
56	Age of Youngest Driver (73)	-0.1821	0.07391	40.6		4,604	0.7	0.8335
57	Age of Youngest Driver (74)	-0.2591	0.07367	28.4		4,956	0.8	0.7718
58	Age of Youngest Driver (75)	-0.2263	0.07547	33.4		4,581	0.7	0.7975
59	Age of Youngest Driver (76)	-0.1733	0.06954	40.1		5,224	0.8	0.8409
60	Age of Youngest Driver (77)	-0.1119	0.06758	60.4		5,258	0.8	0.8941
61	Age of Youngest Driver (78)	-0.1322	0.06856	51.9		5,213	0.8	0.8762
62	Age of Youngest Driver (79)	-0.1561	0.07082	44.9		5,077	0.8	0.8555
63	Age of Youngest Driver (80)	-0.1110	0.06879	62.0		5,054	0.8	0.8949
64	Age of Youngest Driver (81)	-0.1590	0.07311	46.0		4,551	0.7	0.8530
65	Age of Youngest Driver (82)	-0.0623	0.07506	120.5		3,897	0.6	0.9396
66	Age of Youngest Driver (83)	-0.1063	0.08243	77.5		3,262	0.5	0.8991
67	Age of Youngest Driver (84)	-0.1863	0.09422	50.6		2,567	0.4	0.8301
68	Age of Youngest Driver (85+)	-0.2058	0.06313	30.7		7,047	1.1	0.8140
-	Number of Drivers (1)					342,035	52.4	
69	Number of Drivers (2)	-0.1003	0.01255	12.5		228,790	35.0	0.9046
70	Number of Drivers (3)	0.0203	0.01865	91.8		60,466	9.3	1.0205
71	Number of Drivers (4)	0.0595	0.03025	50.9		18,686	2.9	1.0613
72	Number of Drivers (5)	0.0408	0.07145	175.2		3,127	0.5	1.0416
73	RatingArea (1)	-0.1235	0.12439	100.7		1,385	0.2	0.8839

β Pages

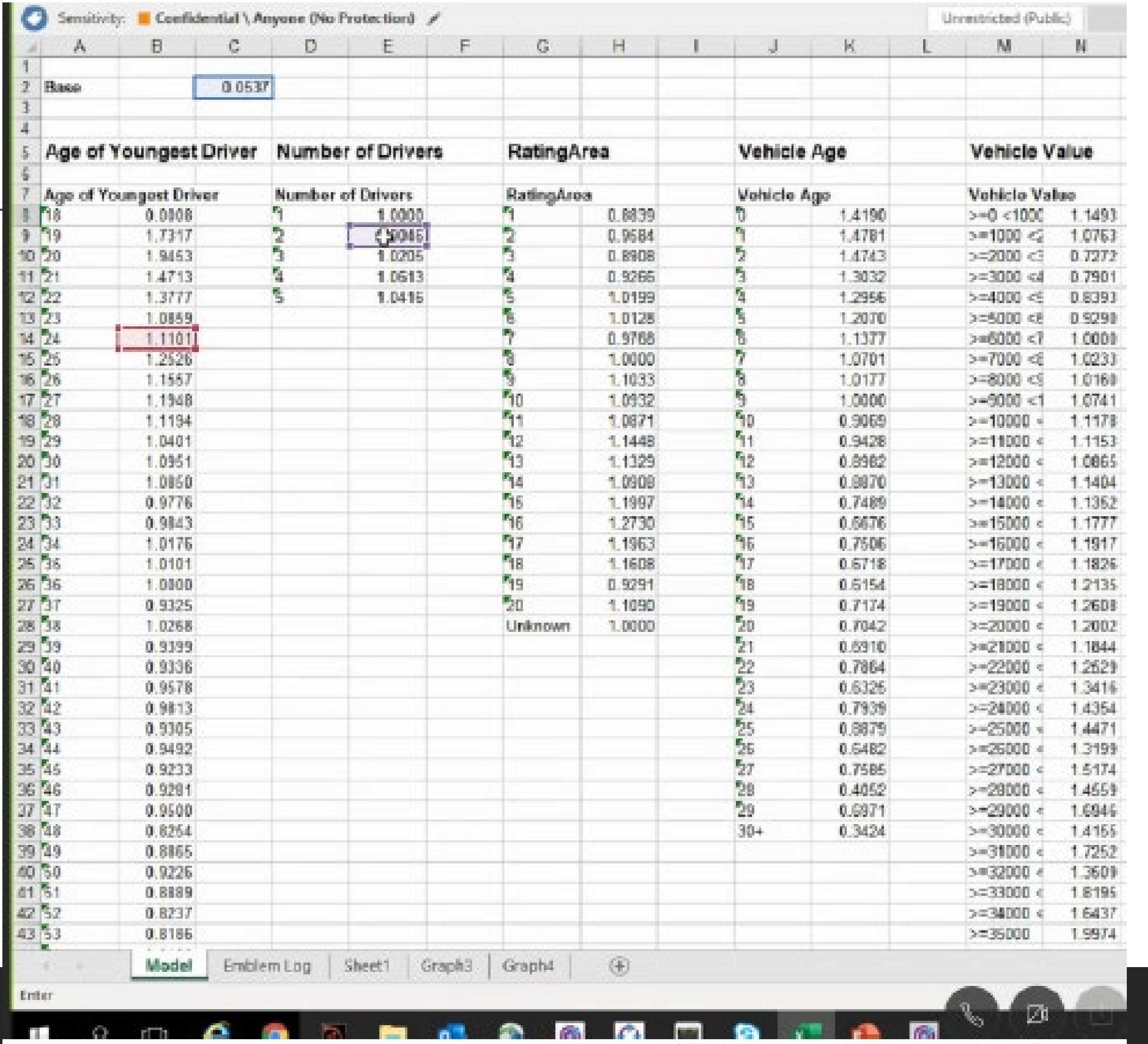
- Alias Indicator (%)
  - Higher percent means more likely
- Weight
  - Denominator of Target
  - Weight (%) is the weight in that level
- Exp(Value)
  - Prediction at base level



Parameter Number	Name	Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
49	Age of Youngest Driver (66)	-0.3232	0.06919	21.4		6,232	1.0	0.7238
50	Age of Youngest Driver (67)	-0.2976	0.07136	24.0		5,635	0.9	0.7426
51	Age of Youngest Driver (68)	-0.2833	0.07266	25.6		5,270	0.8	0.7533
52	Age of Youngest Driver (69)	-0.2310	0.07267	31.5		5,001	0.8	0.7937
53	Age of Youngest Driver (70)	-0.1889	0.07243	38.3		4,859	0.7	0.8279
54	Age of Youngest Driver (71)	-0.1451	0.07196	49.6		4,724	0.7	0.8649
55	Age of Youngest Driver (72)	-0.1786	0.07375	41.3		4,626	0.7	0.8364
56	Age of Youngest Driver (73)	-0.1821	0.07391	40.6		4,604	0.7	0.8335
57	Age of Youngest Driver (74)	-0.2591	0.07367	28.4		4,956	0.8	0.7718
58	Age of Youngest Driver (75)	-0.2263	0.07547	33.4		4,581	0.7	0.7975
59	Age of Youngest Driver (76)	-0.1733	0.06954	40.1		5,224	0.8	0.8409
60	Age of Youngest Driver (77)	-0.1119	0.06758	60.4		5,258	0.8	0.8941
61	Age of Youngest Driver (78)	-0.1322	0.06856	51.9		5,213	0.8	0.8762
62	Age of Youngest Driver (79)	-0.1561	0.07082	44.9		5,077	0.8	0.8555
63	Age of Youngest Driver (80)	-0.1110	0.06879	62.0		5,054	0.8	0.8949
64	Age of Youngest Driver (81)	-0.1590	0.07311	46.0		4,551	0.7	0.8530
65	Age of Youngest Driver (82)	-0.0623	0.07506	120.5		3,897	0.6	0.9396
66	Age of Youngest Driver (83)	-0.1063	0.08243	77.5		3,262	0.5	0.8991
67	Age of Youngest Driver (84)	-0.1863	0.09422	50.6		2,587	0.4	0.8301
68	Age of Youngest Driver (85+)	-0.2058	0.06313	30.7		7,047	1.1	0.8140
-	Number of Drivers (1)					342,035	52.4	
69	Number of Drivers (2)	-0.1003	0.01255	12.5		228,790	35.0	0.9046
70	Number of Drivers (3)	0.0203	0.01865	91.8		60,466	9.3	1.0205
71	Number of Drivers (4)	0.0595	0.03025	50.9		18,686	2.9	1.0613
72	Number of Drivers (5)	0.0408	0.07145	175.2		3,127	0.5	1.0416
73	RatingArea (1)	-0.1235	0.12439	100.7		1,385	0.2	0.8839

β Exporter

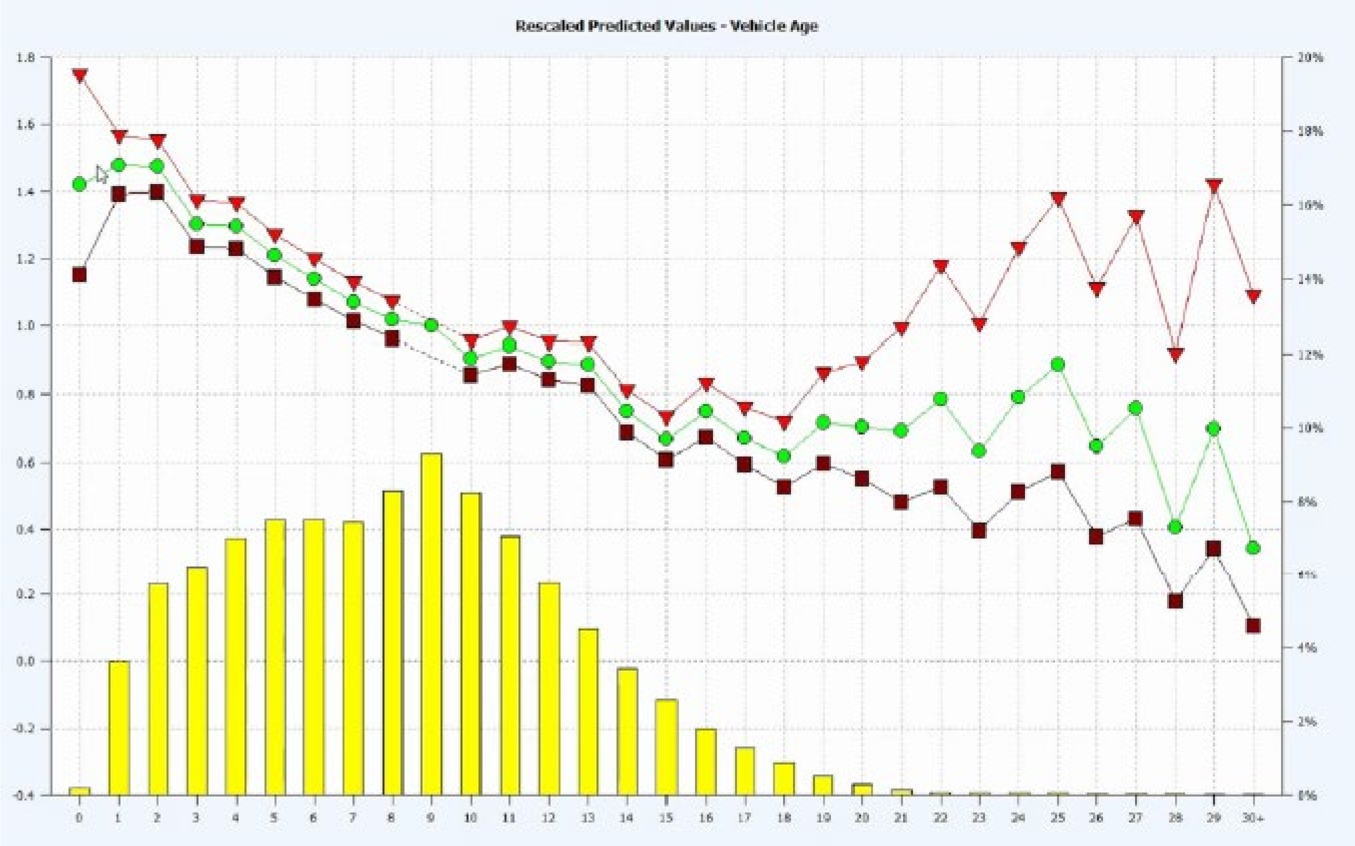
- Same options as graph options
  - Linear Predictor vs. Fitted Value
  - Unscaled vs. Rescaled
- Exports comprehensive indicated list to Excel
  - Base appears at the top (intercept term)
  - Simple factors appear below the base
  - Interaction factors appear below the simple factors
- Example is "Rescaled fitted value"
  - Every base level shows 1.000



Base	0.0537								
Age of Youngest Driver		**Number of Drivers**		**RatingArea**		**Vehicle Age**		**Vehicle Value**	
Age of Youngest Driver		Number of Drivers		RatingArea		Vehicle Age		Vehicle Value	
18	0.0808	1	1.0000	1	0.8839	0	1.4190	>=0 <100C	1.1493
19	1.7317	2	[illegible]	2	0.9684	1	1.4781	>=1000 <2	1.0763
20	1.9453	3	1.0205	3	0.8908	2	1.4743	>=2000 <3	0.7272
21	1.4713	4	1.0613	4	0.9266	3	1.3032	>=3000 <4	0.7901
22	1.3777	5	1.0416	5	1.0199	4	1.2956	>=4000 <5	0.8393
23	1.0859			6	1.0128	5	1.2070	>=5000 <6	0.9298
24	1.1101			7	0.9768	6	1.1377	>=6000 <7	1.0000
25	1.2526			8	1.0000	7	1.0701	>=7000 <8	1.0233
26	1.1567			9	1.1033	8	1.0177	>=8000 <9	1.0160
27	1.1948			10	1.0932	9	1.0000	>=9000 <1	1.0741
28	1.1194			11	1.0871	10	0.9069	>=10000 <	1.1178
29	1.0401			12	1.1448	11	0.9428	>=11000 <	1.1153
30	1.0951			13	1.1329	12	0.8982	>=12000 <	1.0865
31	1.0850			14	1.0908	13	0.8870	>=13000 <	1.1404
32	0.9776			15	1.1997	14	0.7489	>=14000 <	1.1352
33	0.9843			16	1.2730	15	0.6676	>=15000 <	1.1777
34	1.0176			17	1.1963	16	0.7506	>=16000 <	1.1917
35	1.0101			18	1.1608	17	0.6718	>=17000 <	1.1826
36	1.0000			19	0.9291	18	0.6154	>=18000 <	1.2135
37	0.9325			20	1.1090	19	0.7174	>=19000 <	1.2608
38	1.0268			Unknown	1.0000	20	0.7042	>=20000 <	1.2002
39	0.9399					21	0.6910	>=21000 <	1.1844
40	0.9336					22	0.7864	>=22000 <	1.2629
41	0.9578					23	0.6326	>=23000 <	1.3416
42	0.9813					24	0.7939	>=24000 <	1.4354
43	0.9305					25	0.8879	>=25000 <	1.4471
44	0.9492					26	0.6482	>=26000 <	1.3199
45	0.9233					27	0.7585	>=27000 <	1.5174
46	0.9281					28	0.4052	>=28000 <	1.4659
47	0.9500					29	0.6971	>=29000 <	1.6946
48	0.8254					30+	0.3424	>=30000 <	1.4155
49	0.8865							>=31000 <	1.7252
50	0.9226							>=32000 <	1.3609
51	0.8889							>=33000 <	1.8195
52	0.8237							>=34000 <	1.6437
53	0.8186							>=35000	1.9974

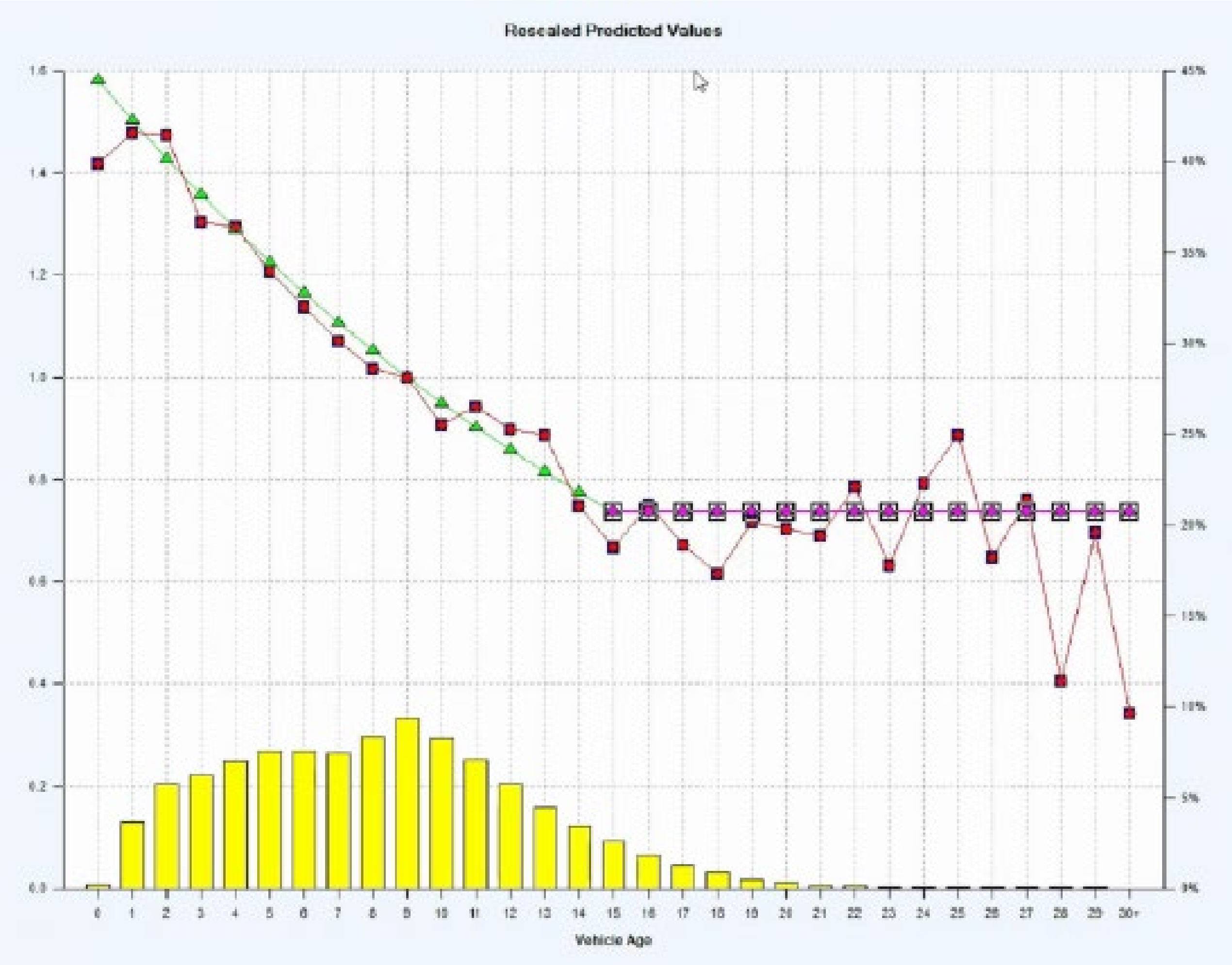
Curve Fitting (Before)

- Every level of Vehicle Age was it's own parameter
  - Essentially treated like a categorical variable
- Obvious pattern emerges on the left side
- Confidence interval expands on the right side
  - Caused by low data volume
  - Confidence intervals impacted by choice of base level
    - CAS GLM Paper Section 2.4.3
- "Rescaled Predicted Value" is essentially the indicated factor
  - Base = 1.000



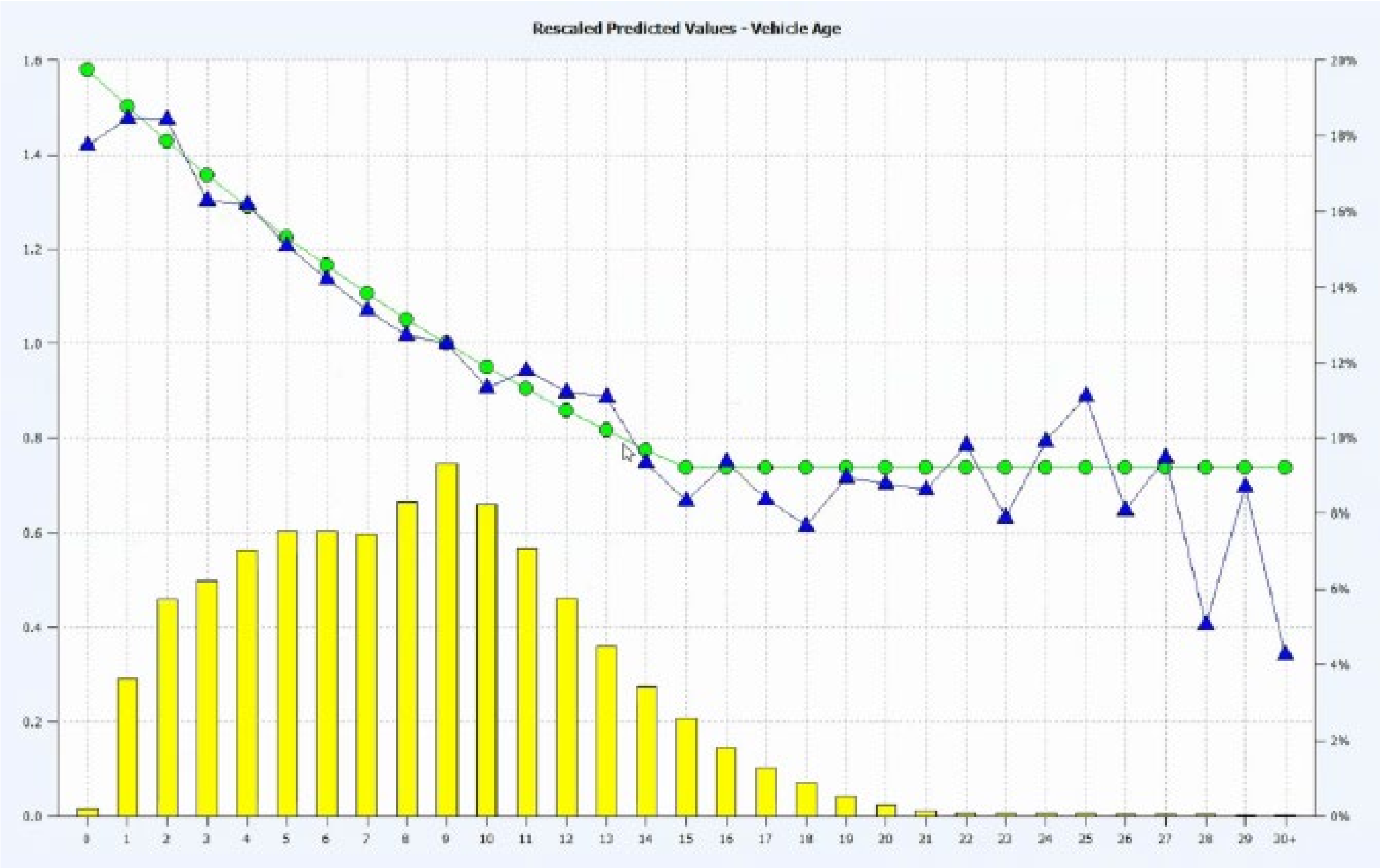
Curve Fitting (After)

- Number of parameters greatly reduced
  - # of degrees in polynomial fit
- Factors now reflect a smooth pattern
- Note, factors may appear slightly curved even on a 1 degree fit
  - The graph on the right is a 1 degree fit
  - Still appears slightly curved
  - This is because of the log link function
- Large continuous variables should be logged before fitting a curve
  - CAS GLM Paper (Section 2.4.1)



Curve Fitting (Before and After)

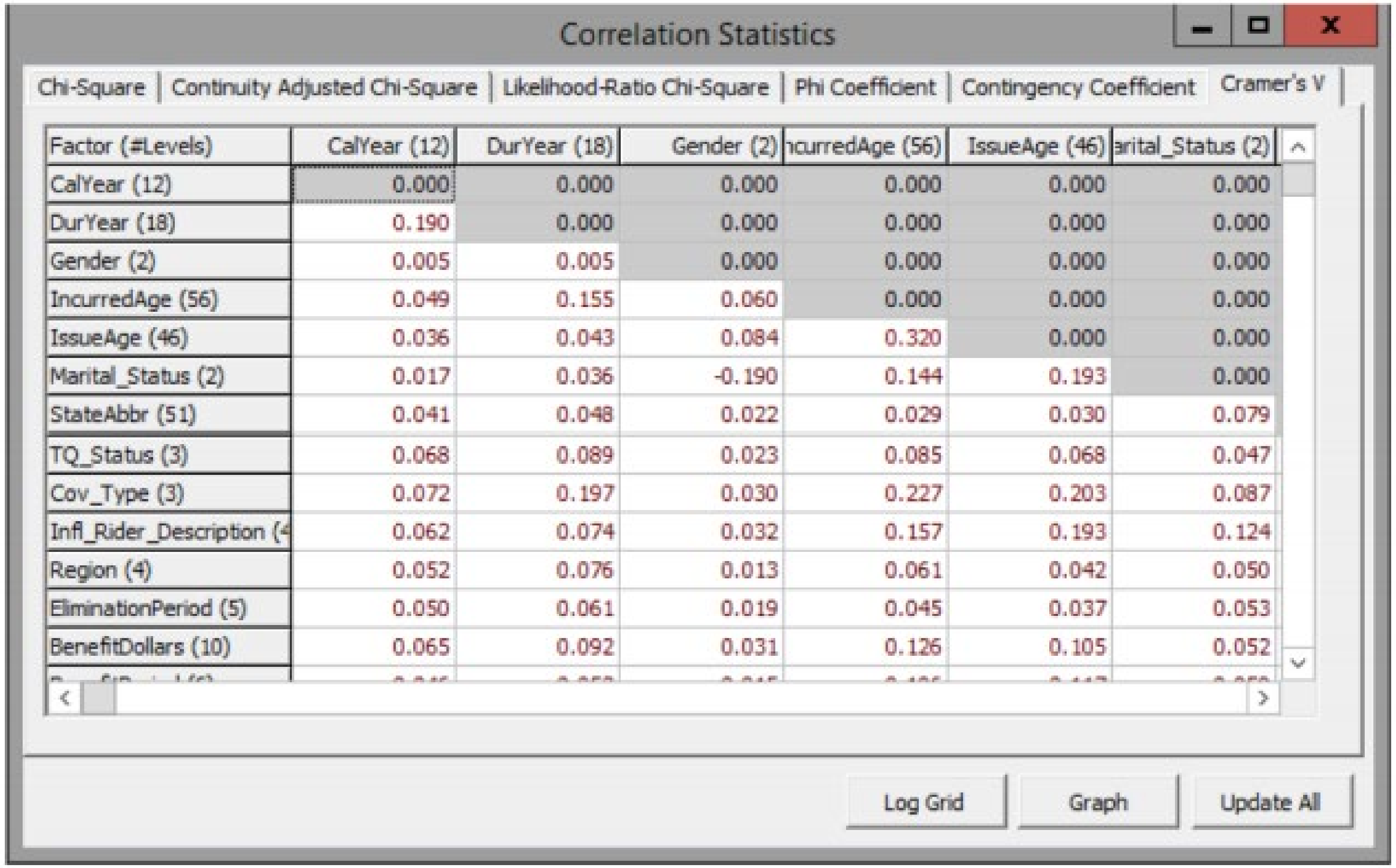
- You can compare fitted model to reference model at any time
- Reference Model
  - Blue Triangles
  - Before Curve Fitting
    - Lots of parameters
    - Not smooth at all
- Fitted Model
  - Green Circles
  - After Curve Fitting
    - Few parameters
    - Monotonic if desired
    - Smooth factors limit disruption



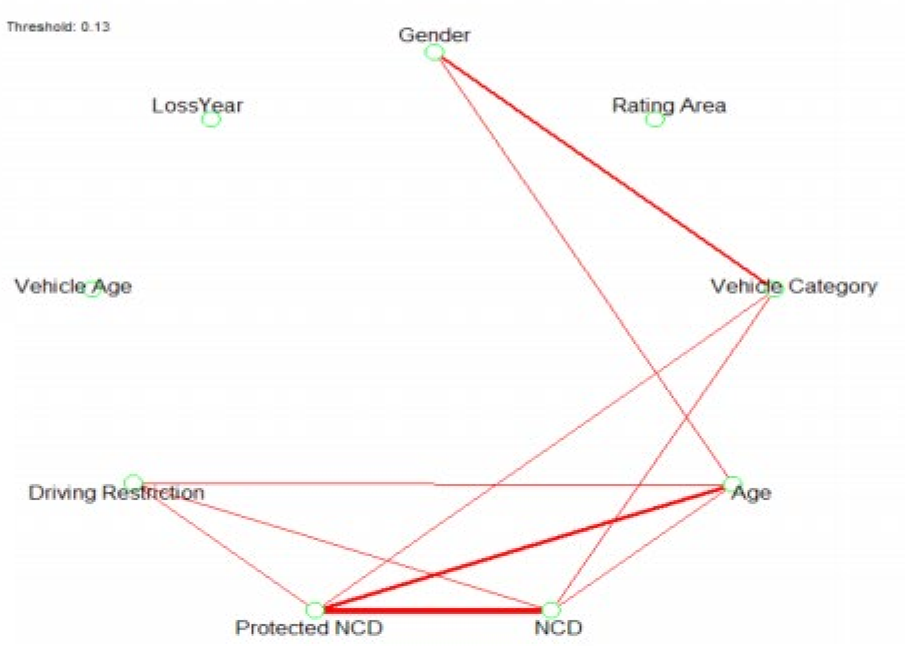
Correlation Tests

Determine an acceptable cutoff

Towers Watson
Emblem



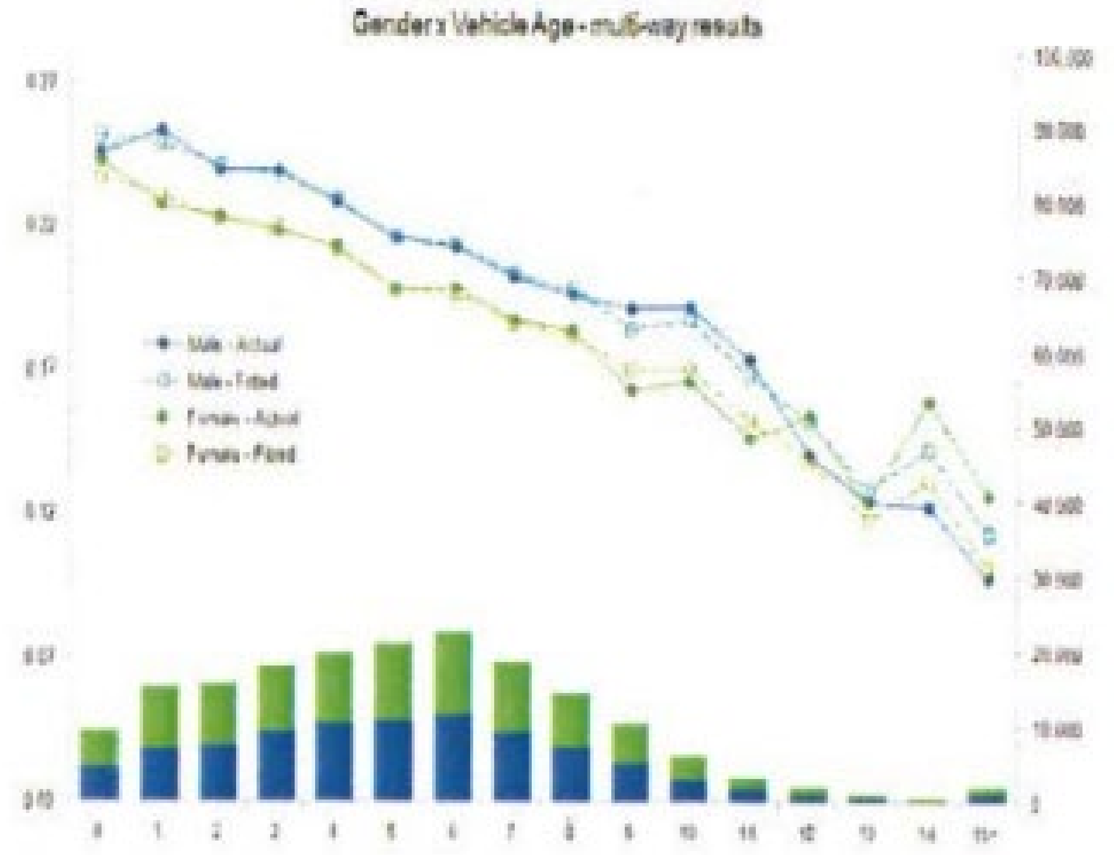
Factor (#Levels)	CalYear (12)	DurYear (18)	Gender (2)	ncurredAge (56)	IssueAge (46)	arital_Status (2)
CalYear (12)	0.000	0.000	0.000	0.000	0.000	0.000
DurYear (18)	0.190	0.000	0.000	0.000	0.000	0.000
Gender (2)	0.005	0.005	0.000	0.000	0.000	0.000
IncurredAge (56)	0.049	0.155	0.060	0.000	0.000	0.000
IssueAge (46)	0.036	0.043	0.084	0.320	0.000	0.000
Marital_Status (2)	0.017	0.036	-0.190	0.144	0.193	0.000
StateAbbr (51)	0.041	0.048	0.022	0.029	0.030	0.079
TQ_Status (3)	0.068	0.089	0.023	0.085	0.068	0.047
Cov_Type (3)	0.072	0.197	0.030	0.227	0.203	0.087
Infl_Rider_Description (4	0.062	0.074	0.032	0.157	0.193	0.124
Region (4)	0.052	0.076	0.013	0.061	0.042	0.050
EliminationPeriod (5)	0.050	0.061	0.019	0.045	0.037	0.053
BenefitDollars (10)	0.065	0.092	0.031	0.126	0.105	0.052



- Visualization shows thicker lines for stronger correlations
- Correlation Statistics appear in a table that can be easily exported

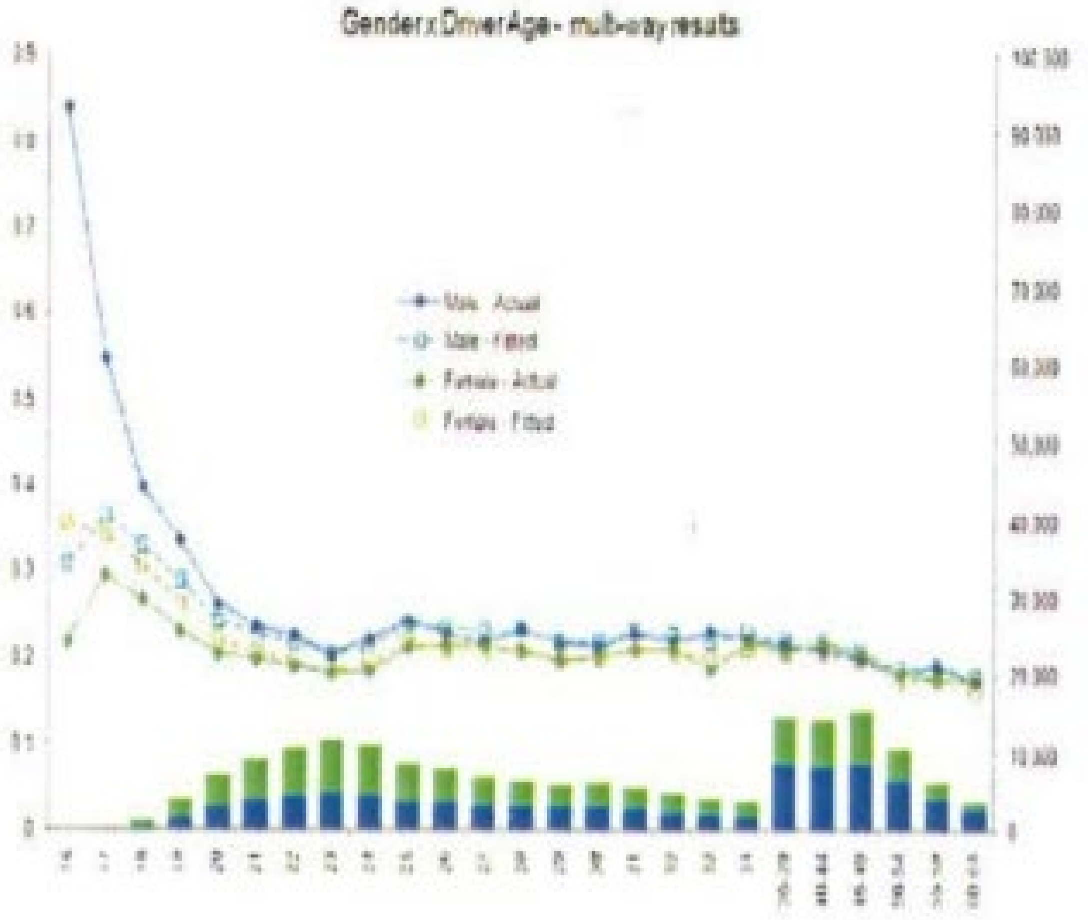
Multi-way Graphs
Checking for Interactions

- Multi-way test: Gender vs. Vehicle Age
- Colors
  - Blue for male, Green for female
  - Volume split nearly 50/50
- Looking at the graph WITHOUT interaction fitted in the model
  - Actual vs Expected for “Male” looks good
  - Actual vs Expected for “Female” looks good
- If our model fits both groups well without the interaction, it’s superfluous
- Chi-Square test comparing a model with the interaction and a model without doesn’t pass significance.



Multi-way Graphs
Checking for Interactions

- Multi-way test: Gender vs. Driver Age
- Colors
  - Blue for male, Green for male
  - Volume split nearly 50/50
- Looking at the graph WITHOUT interaction fitted in the model
  - Actual vs Expected biased low for young males
- Our model fails without an interaction
- Chi-Square test comparing a model with the interaction and a model without DOES pass significance test.



Other EMBLEM tools

- Backwards and Forwards Stepwise Regression Tests
  - Running a backwards stepwise regression on a final model highlights potentially insignificant terms
- Automated Interaction Tests
- Set Offset Factors
  - Coverage options (limits/deductibles) and territory are often better handled outside GLM
    - CAS GLM Paper (Sections 9.1 & 9.2)
- Convert model to a scoring model

Reference

- CAS GLM Paper
  - https://www.casact.org/pubs/monographs/papers/05-Goldburd-Khare-Tevet.pdf