Modeling Medical Cost Differences Across the US

Kyle Smith & Jon Duea

4/29/2025

STAT 311 REGRESSION ANALYSIS, SPRING 2025

Introduction:

For our topic, we chose the research question: “Do the factors that influence medical costs remain consistent across different regions?”

This question is significant because it highlights the variability in medical costs billed by insurance across the United States, potentially driven by demographic and lifestyle differences. Understanding these variations can reveal whether people face cost disparities based solely on their region, and highlight the influence specific factors have on the costs people face around the U.S.

To answer this question, we use a publicly available dataset that records total annual medical costs billed by insurance (charges) for 1338 simulated beneficiaries (individuals with an insurance policy) together with six predictors (age, sex, BMI, children, smoker, region).

The approach we will take to answer this question is to first conduct exploratory data analysis (EDA) to inform our model building process, build regression models incorporating insights from our EDA, validate our model assumptions before making any conclusions, and then compare competing models for efficacy and to determine whether the effects of the predictors remain consistent across different regions.

Data Description:

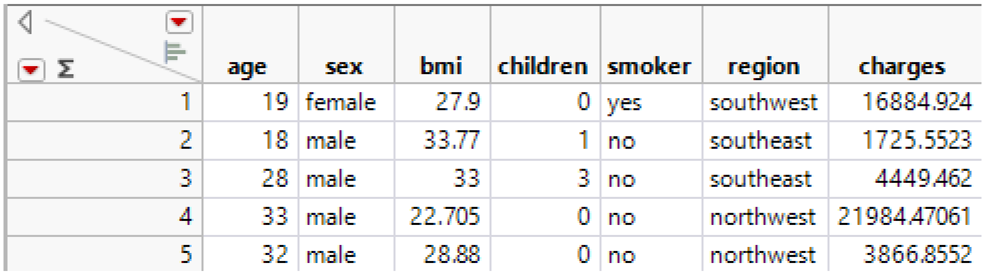
The dataset we will be using for our analysis was obtained from *Medical Cost Personal Datasets* (Choi, n.d.) available on Kaggle, which was originally curated for use in the book *Machine Learning with R* and was “created for this book using demographic statistics from the U.S. Census Bureau, and thus approximately reflect real-world conditions.” (Lantz, 2013). Although not explicitly stated in the book’s description of the dataset, it is assumed that it was generated based off 2013 demographic statistics from the U.S. Census Bureau.

Included in the dataset is complete information with no values missing for 1338 examples of beneficiaries currently enrolled in an insurance plan and covers the following variables:

- Charges (USD/year)
- Age (integer years, ≤ 64)
- Sex (male/female)
- BMI (Body-Mass Index, $\frac{kg}{m^2}$
- Children (number of dependents)
- Smoker (yes/no)
- Region (of United States: northeast, northwest, southeast, southwest)

This dataset represents a snapshot of individuals from varying demographic profiles (age, sex), health indicators (BMI, smoking status), family sizes (number of children), and geographic locations within the US. The associated charge for everyone reflects the medical expenses billed to their insurance and provides the basis for exploring how these factors influence medical costs collectively and individually.

Exploratory Data Analysis



	age	sex	bmi	children	smoker	region	charges
1	19	female	27.9	0	yes	southwest	16884.924
2	18	male	33.77	1	no	southeast	1725.5523
3	28	male	33	3	no	southeast	4449.462
4	33	male	22.705	0	no	northwest	21984.47061
5	32	male	28.88	0	no	northwest	3866.8552

Snapshot of first five records of the dataset.

Exploration of our datasets revealed many insights for us and highlighted some issues that will need to be accounted for when conducting our analysis. Appendix A contains the detailed results from our analysis, and the accompanying JMP workbook (insurance.jmp) can be found in our GitHub repository for this project (Duea & Smith, 2025) if you would like to run the analysis for yourself. All figures and tables included in this report and the appendices have an associated script in insurance.jmp of the same name that will reproduce them.

For a descriptive summary of the numeric variables in our data, see appendix A, figure A1. Below we highlight four key insights about our data uncovered during our analysis:

- **Charges**: Analysis of the distribution of charges shows outliers and a rightward skew, applying the log transformation transforms charges to a more normal distribution (see appendix A, figure A2). Additionally, when looking at charges by region and trying to visually estimate whether the assumption of homoscedasticity is satisfied we can see that it is not and we will have to take the necessary steps to address the heteroscedasticity in our modeling efforts (see appendix A, figure A3).
- **BMI**: The average BMI for an individual in our dataset (see appendix A, Figure A1) is 30.66, which classifies them as obese, according to info about adult BMI categories provided by the CDC (Centers for Disease Control and Prevention, 2024). Looking at charges across different BMI values (see appendix A, figure A4) we see the assumption of homoscedasticity violated here as well, with a noticeable increase in variance in charges at the 30 BMI mark, suggesting that individuals that would be classified as obese have increased variance in medical costs compared to those in healthy weight ranges.
- **Smoker:** The difference between charges between smokers and non-smokers is significant, with smokers facing over 3 times the medical costs of a non-smoker on average; furthermore, charges that would be considered outliers for non-smokers fall firmly within the range of normal values for smokers (see Appendix A, figure A5). Considering the average individual in our dataset falls into the obese category and research from the NIH indicating that individuals who are both obese and smoke face worse health outcomes than others (Stewart, Cutler, & Rosen, 2009), we looked to see if that interaction might be part of the reason we saw such a sharp increase in variance of charges for obese individuals (see appendix A, figure A6). Based on this we hypothesize the interaction between BMI and smoking seems to

be significant in influencing medical costs, with the interaction being more influential for obese individuals.

- **Region:** Average costs across the regions are similar, with one notable exception being the southeast region (see appendix A, figure A7). Looking at the prevalence of smokers and obese individuals across the different regions provides an explanation for why this might be, and we find that the southeast region has the highest prevalence of smokers (see appendix A, figure A8), and obese individuals (see appendix A, figure A9). With this in mind, we deemed it was important to account for the impacts of smoking and obesity before drawing any conclusions about how much a region influences the effect of our predictors on medical costs.

Based on these insights, we conclude with three actions we should take when preparing our models:

1. Taking the log transformation of charges is necessary to address the skew and unequal variances present in our data, so we will be interpreting our models as multiplicative models.
2. Incorporation of an indicator variable for BMI ≥ 30 will help us better model the effects of BMI and its interaction with smoking.
3. Incorporation of interactions between smoking and our BMI terms to account for the major effects they have on each other in influencing medical costs.

See appendix A, figure A10 and figure A11 notes for a detailed breakdown and supporting visualizations for this.

Methods:

We model medical charges using multiple linear regression, and because raw charges are highly skewed, violate constant-variance and normal-error assumptions, we fit multiplicative models with $y = \log(charges)$. Considering the nature of medical costs and the factors that influence them, it intuitively follows that percentages changes in charges are more appropriate than additive changes in charges. The accompanying JMP workbook (insurance.jmp) includes scripts for model fit along with their accompanying diagnostics so that, if desired, you can run the models yourself. We hypothesize three models, progressively adding more terms and comparing them for efficacy.

Model 1 (main effect terms, BMI x Smoker interaction):

$$\ln(\mathrm{y}) = \upbeta_0 + \overbrace{\beta_1 x_1}^{\substack{Main\ effect \\ age}} + \overbrace{\beta_2 x_2}^{\substack{Main\ effect \\ bmi}} + \overbrace{\beta_3 x_3}^{\substack{Main\ effect \\ children}} + \overbrace{\beta_4 x_4}^{\substack{Main\ effect \\ sex}} + \overbrace{\beta_5 x_5}^{\substack{Main\ effect \\ smoker}}$$

$$+ \overbrace{\beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8}^{\substack{Main\ effect \\ region}} + \overbrace{\beta_9 x_2 x_4}^{\substack{BMI\ x\ Smoker \\ Interaction}}$$

where:

$$y = mean\ charges \qquad x_1 = age \qquad x_2 = bmi$$

$$x_3 = children \qquad x_4 = \begin{cases} 1\ if\ male \\ 0\ if\ female \end{cases} \qquad x_5 = \begin{cases} 1\ if\ smoker \\ 0\ if\ not \end{cases}$$

$$x_6 = \begin{cases} 1\ if\ northwest \\ 0\ if\ not \end{cases} \qquad x_7 = \begin{cases} 1\ if\ southeast \\ 0\ if\ not \end{cases} \qquad x_8 = \begin{cases} 1\ if\ southwest \\ 0\ if\ not \end{cases}$$

Northeast is the base level for region

This model considers all the main effects of our predictors, along with the known interaction effect between BMI and smoking, providing a baseline for us to compare the other models against. The output from JMP for this model can be found in the script group Model 1 of the insurance.jmp workbook as well as in Appendix B.

Model 2 (extends model 1 with quadratic term for age and smoker interactions):

$$\ln(\mathrm{y}) = model\ 1 + \overbrace{\beta_{10} x_9}^{\substack{Main\ Effect \\ Obesity}} + \overbrace{\beta_{11} x_5 x_9}^{\substack{Obesity\ x\ Smoker \\ Interaction}} + \overbrace{\beta_{12} x_1^2}^{Age\ quadratic\ term} + \overbrace{\beta_{13} x_1 x_5 + \beta_{14} x_1^2 x_5}^{\substack{Age\ x\ Smoker \\ Interactions}}$$

where:

$y\ and\ x_1 \ldots x_8$ terms defined the same as model 1 with one added new qualitative variable.

$$x_9 = \begin{cases} 1\ if\ BMI \geq 30 \\ 0\ otherwise \end{cases}$$

This model is meant to test whether the terms (i.e. obesity indicator and smoker interactions) as well as adding the allowance for a curve in the relationship between age and charges will contribute significantly to the prediction of medical costs. The addition of the quadratic term is based on the hypothesis that the relationship between age and costs is not linear. In order to avoid issues of collinearity with the new quadratic term, age was standardized through JMP's standardize method which subtracts the group mean and divides by the standard deviation of the group. Testing whether the new terms added are statistically significant in contributing information for the prediction of charges will be assessed using a partial F-test. The output from JMP for this model can be found in the script group Model 2 of the insurance.jmp workbook as well as in Appendix C.

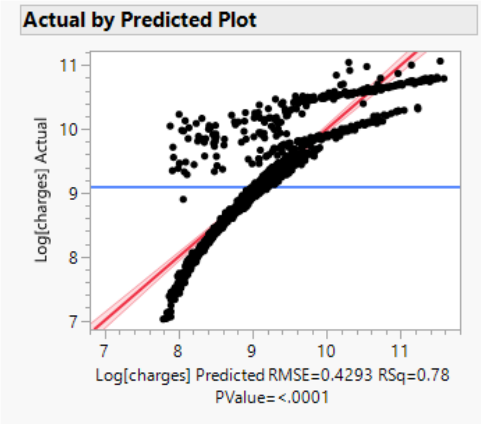
Model 3 (extends model 2 with region interactions over every previous term):

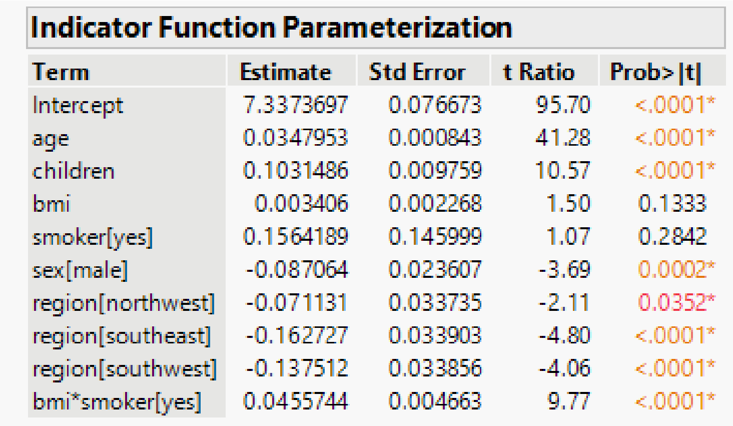
$$\begin{aligned}
\ln(y) = model\ 2 &+ \overbrace{\beta_{15}x_1x_6 + \beta_{16}x_1x_7 + \beta_{17}x_1x_8}^{\substack{age\ x\ region\\ interactions}} + \overbrace{\beta_{18}x_2x_6 + \beta_{19}x_2x_7 + \beta_{20}x_2x_8}^{\substack{children\ x\ region\\ interactions}} \\
&+ \overbrace{\beta_{21}x_3x_6 + \beta_{22}x_3x_7 + \beta_{23}x_3x_8}^{\substack{bmi\ x\ region\\ interactions}} + \overbrace{\beta_{24}x_4x_6 + \beta_{25}x_4x_7 + \beta_{26}x_4x_8}^{\substack{smoker\ x\ region\\ interactions}} \\
&+ \overbrace{\beta_{27}x_5x_6 + \beta_{28}x_5x_7 + \beta_{29}x_5x_8}^{\substack{sex\ x\ region\\ interactions}} + \overbrace{\beta_{30}x_3x_4x_6 + \beta_{31}x_3x_4x_7 + \beta_{32}x_3x_4x_8}^{\substack{bmi\ x\ smoker\ x\ region\\ interactions}} \\
&+ \overbrace{\beta_{33}x_9x_6 + \beta_{34}x_9x_7 + \beta_{35}x_9x_8}^{\substack{obesity\ x\ region\\ interactions}} + \overbrace{\beta_{36}x_4x_9x_6 + \beta_{37}x_4x_9x_7 + \beta_{38}x_4x_9x_8}^{\substack{smoker\ x\ obesity\ x\ region\\ interaction}} \\
&+ \overbrace{\beta_{39}x_1^2x_6 + \beta_{40}x_1^2x_7 + \beta_{41}x_1^2x_8}^{\substack{age^2\ \ x\ region\\ interactions}} + \overbrace{\beta_{42}x_1x_4x_6 + \beta_{43}x_1x_4x_7 + \beta_{44}x_1x_4x_8}^{\substack{age\ x\ smoker\ x\ region\\ interactions}} \\
&+ \overbrace{\beta_{45}x_1^2x_4x_6 + \beta_{46}x_1^2x_4x_7 + \beta_{47}x_1^2x_4x_8}^{\substack{age^2\ x\ smoker\ x\ region\\ interactions}}
\end{aligned}$$

This model is meant to provide us with the information needed to make an informed conclusion about whether the factors that affect medical costs remain consistent across the different regions of the US. We acknowledge that the more hierarchical and complete approach of taking all k-way interactions between our qualitative and quantitative predictors would allow us to have more certainty in our conclusion but due to the large number of terms (100+), interpretation of that model is not practically useful and a trade off between statistical robustness and practicality was made. To determine whether there are any significant interactions between region and our predictors from model 2, we'll perform a partial f-test to see if the new terms added are statistically significant in contributing information for the prediction of charges like we did for model 2. See the script group Model 3 of the insurance.jmp workbook as well as Appendix D for the outputs from JMP for this model.

Results:

Model 1:





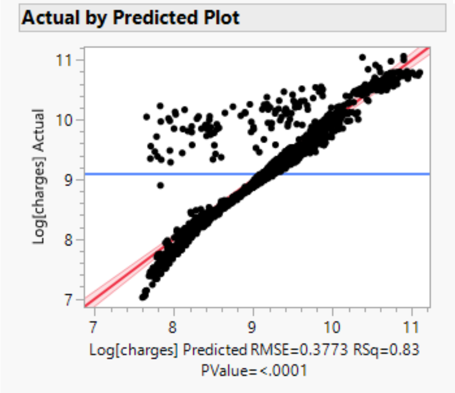
Indicator Function Parameterization

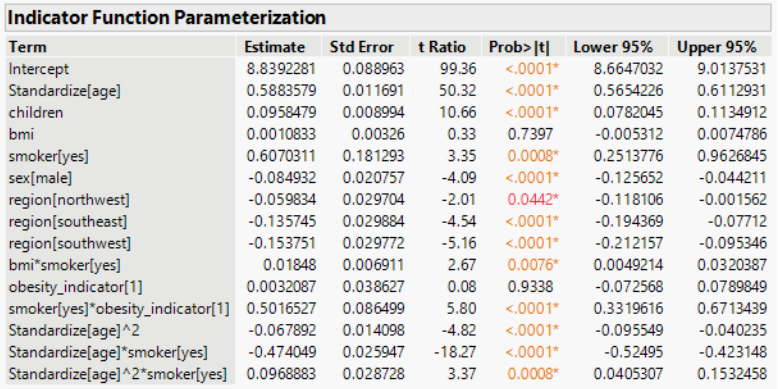
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.3373697	0.076673	95.70	<.0001*
age	0.0347953	0.000843	41.28	<.0001*
children	0.1031486	0.009759	10.57	<.0001*
bmi	0.003406	0.002268	1.50	0.1333
smoker[yes]	0.1564189	0.145999	1.07	0.2842
sex[male]	-0.087064	0.023607	-3.69	0.0002*
region[northwest]	-0.071131	0.033735	-2.11	0.0352*
region[southeast]	-0.162727	0.033903	-4.80	<.0001*
region[southwest]	-0.137512	0.033856	-4.06	<.0001*
bmi*smoker[yes]	0.0455744	0.004663	9.77	<.0001*

Appendix B, Figure B1 (above) and Appendix B, Figure B2 (right)

Our baseline model fits the data reasonably well, with nearly all terms being significant. Looking at appendix B, figure B1 we can see that our model does not account for all the relationships between our predictors and charges though. This model accounts for about 78% of the variance in charges after adjusting for our predictors (appendix B, figure B3) and is considered to be statistically useful for prediction of charges based on the global F-test (F=534, p < .001, appendix B, figure B4). Normal distribution and constant variance assumptions are satisfied to the best of our ability but studentized residuals show the presence of high leverage outliers (see appendix B figures B5-B7 for residual plots). Due to the nature of the data, and the robustness of regression to outliers no action is taken to exclude outliers at this point.

Model 2:





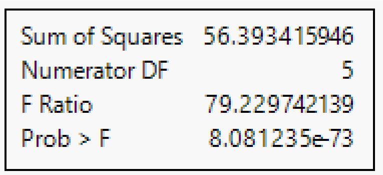
Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	8.8392281	0.088963	99.36	<.0001*	8.6647032	9.0137531
Standardize[age]	0.5883579	0.011691	50.32	<.0001*	0.5654226	0.6112931
children	0.0958479	0.008994	10.66	<.0001*	0.0782045	0.1134912
bmi	0.0010833	0.00326	0.33	0.7397	-0.005312	0.0074786
smoker[yes]	0.6070311	0.181293	3.35	0.0008*	0.2513776	0.9626845
sex[male]	-0.084932	0.020757	-4.09	<.0001*	-0.125652	-0.044211
region[northwest]	-0.059834	0.029704	-2.01	0.0442*	-0.118106	-0.001562
region[southeast]	-0.135745	0.029884	-4.54	<.0001*	-0.194369	-0.07712
region[southwest]	-0.153751	0.029772	-5.16	<.0001*	-0.212157	-0.095346
bmi*smoker[yes]	0.01848	0.006911	2.67	0.0076*	0.0049214	0.0320387
obesity_indicator[1]	0.0032087	0.038627	0.08	0.9338	-0.072568	0.0789849
smoker[yes]*obesity_indicator[1]	0.5016527	0.086499	5.80	<.0001*	0.3319616	0.6713439
Standardize[age]^2	-0.067892	0.014098	-4.82	<.0001*	-0.095549	-0.040235
Standardize[age]*smoker[yes]	-0.474049	0.025947	-18.27	<.0001*	-0.52495	-0.423148
Standardize[age]^2*smoker[yes]	0.0968883	0.028728	3.37	0.0008*	0.0405307	0.1532458

Appendix C, Figure C1 (above) and Appendix C, Figure C2 (right)

The addition of the quadratic term for age, obesity indicator, and smoking interactions for age and obesity have improved on model 2. Comparing figure B1 and figure C1 we can see that model 2 more accurately reflects the relationship between our predictors and charges, accounting for about 83% of the variance in charges after adjusting for our predictors (appendix C, figure C3). This model is statistically useful for prediction of charges based on the global F-test (f=472.341, p<.0001, appendix C, figure C4) as well. Assumptions around the error component are approximately satisfied by this model (see appendix D figures B5-B6) and looking at the studentized residual plot, we see that the residuals are now more tightly contained within ±2 of 0, but the presence of high leverage outliers is still there (see appendix C, figure C7).

Visually and based on the R_a^2 value we can see that adding the extra terms do significantly contribute information to the prediction of charges beyond the first models capability, this is validated by a partial F-test (see appendix C, figure C8 for full output).



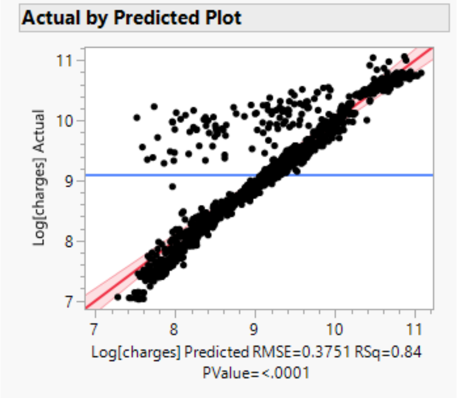
Sum of Squares	56.393415946
Numerator DF	5
F Ratio	79.229742139
Prob > F	8.081235e-73

Formally:

$$H_0: \beta_{10} = \cdots = \beta_{14} = 0 \; vs \; H_a: at\ least\ one\ \beta_i \neq 0; i = 10,11,12,13,14$$

Given an F-statistic of 79.23, and a p-value less than .001 (p < .001), the null hypothesis is rejected at the .01 significance level (α = .01). The inclusion of the new terms significantly improves the prediction of charges.

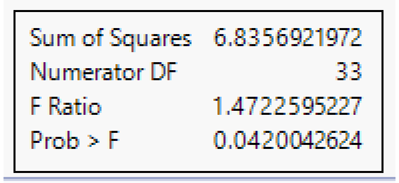
Model 3:



Appendix D, Figure D1 (above) and Indicator Function Parameterization can be found in Appendix D, Figure D2

This model appears to be slightly better at predicting charges at first glance when comparing figure D1 and figure C1 but looking at the indicator function parameterization most of the new terms show no statistical significance (see appendix D, Figure D2). Additionally, this model accounts for the same amount of variance in charges after adjusting for our predictors as model 2 (83%, see appendix D, Figure D3). This model is statistically useful for prediction of charges based on the global F-test (f=143.5069, p<.0001, appendix D, figure D4) and the assumptions about the error seem to be as close to satisfied as we could reasonably get with the data we had as the other 2 models (see appendix d, figures D5-D7).

The question we were hoping to answer with this model was whether region had a significant impact on our other predictors, that is, whether the factors that influence medical costs are consistent across regions. Looking at the effect summary (appendix D, figure D8) the answer would appear to be **yes** the factors that influence medical costs are consistent across different regions of the U.S. if you were to look at any one interaction individually; however the partial f-test was also performed across all the terms added, which signals that at least one of the added terms effects are significant (at the .05 significance level). (see appendix D, figure D9 for the full output).



Sum of Squares	6.8356921972
Numerator DF	33
F Ratio	1.4722595227
Prob > F	0.0420042624

From appendix D, figure D9, partial f-test statistics table

Formally:

$$H_0: \beta_{15} = \cdots = \beta_{47} = 0 \ vs \ H_a: at\ least\ one\ \beta_i \neq 0; 15 \leq i \leq 47$$

Given an F-statistic of 1.47, and a p-value less than .05 (p =.04), the null hypothesis is rejected at the .05 significance level (α = .05). The inclusion of the new terms significantly improves the prediction of charges, suggesting that at least one of the factors that effects medical costs does not remain consistent across the regions of the U.S.

Model comparison and recommendations:

Model	R_a^2	RMSE	BIC
1	0.78205	0.429282	1603.295
2	0.831639	0.377298	1288.831
3	0.833599	0.375096	1476.929

Summary table for all 3 models summary of fit statistics

Out of the three models fit, we would recommend using model 2 for the prediction of medical costs across the US. Model 2 and model 3 provide very similar predictive capabilities, with model 3 showing slightly higher R_a^2 and lower RMSE, but at the cost of increased complexity (15 terms in model 2 versus 47 terms in model 3). This is reflected in the BIC for the 2 models, which based on that criteria would put model 2 firmly in the best model spot.

Considering our selection of model 2 as our preferred model for predicting medical costs, this is practically equivalent as saying that the factors that influence medical costs are consistent across the different regions of the U.S. which is what (based on our data and model selection) is the practical answer to the question. It is important to note that the partial F-test for model 3 did suggest that at least one of the interaction terms for region and our predictors was significant but when assessing the effect of our parameters through the parameter effect summary none of them came back as significant.

Conclusions and Discussion:

Our project examined whether the factors driving medical costs are the same across different regions of the US, using data simulated from U.S. Census bureau statistics from around 2013. We looked at total annual medical charges alongside personal factors like age, sex, body mass index (BMI), number of children, smoking status, and region.

The initial analysis highlighted a few clear patterns. Notably, smokers and individuals with high BMI (in the obese range) tend to incur **significantly** higher medical charges on average compared to non-smokers and those at healthy weights. Regional differences in average costs were relatively small overall. The southeast region stood out with slightly higher average charges, but this was explained by that region having more smokers and obese individuals compared to the other regions. In other words, the higher costs seen in the southeast were likely to do with the higher prevalence of costly risk factors, rather than the region itself causing higher expenses.

Through model building and comparison, we ended up picking the model that balanced predictive accuracy and simplicity as the preferred model for predicting medical costs based on the factors available in our data. This model assumed that region had no impact on the other factors that effect medical costs, and performed comparably to one that did which lead us to the conclusion that, for the most part, and for all practical purposes region is not important in determining how much the impacts of other high risk factors such as obesity or smoking will contribute to average medical costs.

In the end, our analysis indicates that the major driver of higher medical costs – particularly smoking and obesity – are consistent nationwide. Any regional cost difference seems to stem from how common those risk factors are in each region, rather than the region altering how those factors impact medical expenses.

References

Centers for Disease Control and Prevention. (2024, 3 19). *Adult BMI categories*. Retrieved from CDC: https://www.cdc.gov/bmi/adult-calculator/bmi-categories.html

Choi, M. (n.d.). *Medical Cost Personal Datasets*. Retrieved from www.kaggle.com: https://www.kaggle.com/datasets/mirichoi0218/insurance/data

Duea, J., & Smith, K. (2025, 28 4). *Modeling-Medical-Cost-Differences-Across-the-US*. Retrieved from https://github.com: https://github.com/jmduea/Modeling-Medical-Cost-Differences-Across-the-US

Lantz, B. (2013). Example - predicting medical expenses using linear regression. In B. Lantz, *Machine Learning with R* (pp. 361-388). Packt Publishing. Retrieved from https://www.everand.com/book/253051595/Machine-Learning-with-R

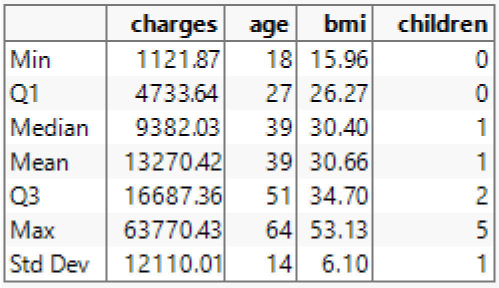
Stewart, S. T., Cutler, D. M., & Rosen, A. B. (2009, 12 3). Forecasting the Effects of Obesity and Smoking on U.S. Life Expectancy. *New England Journal of Medicine, 361*(23), 2252-2260. doi:10.1056/nejmsa0900459

Appendix A

Exploratory Data Analysis Figures and Tables

Figure A1

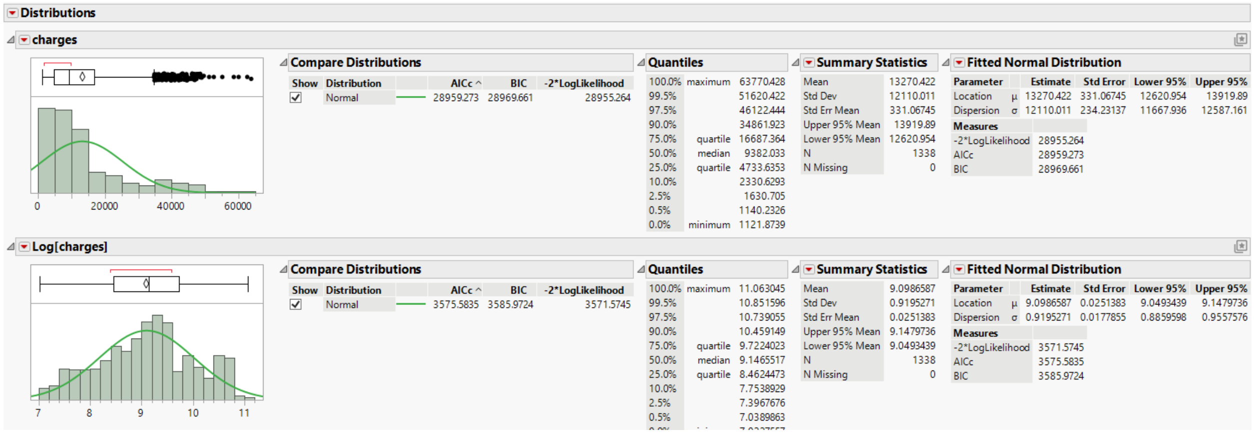
Descriptive summary of variables (n=1338)



	charges	age	bmi	children
Min	1121.87	18	15.96	0
Q1	4733.64	27	26.27	0
Median	9382.03	39	30.40	1
Mean	13270.42	39	30.66	1
Q3	16687.36	51	34.70	2
Max	63770.43	64	53.13	5
Std Dev	12110.01	14	6.10	1

Figure A2

Distribution of charges vs log(charges)



Distributions

charges

Compare Distributions

Show	Distribution		AICc ^	BIC	-2*LogLikelihood
☑	Normal	—	28959.273	28969.661	28955.264

Quantiles

100.0%	maximum	63770.428
99.5%		51620.422
97.5%		46122.444
90.0%		34861.923
75.0%	quartile	16687.364
50.0%	median	9382.033
25.0%	quartile	4733.6353
10.0%		2330.6293
2.5%		1630.705
0.5%		1140.2326
0.0%	minimum	1121.8739

Summary Statistics

Mean	13270.422
Std Dev	12110.011
Std Err Mean	331.06745
Upper 95% Mean	13919.89
Lower 95% Mean	12620.954
N	1338
N Missing	0

Fitted Normal Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Location	μ	13270.422	331.06745	12620.954	13919.89
Dispersion	σ	12110.011	234.23137	11667.936	12587.161

Measures	
-2*LogLikelihood	28955.264
AICc	28959.273
BIC	28969.661

Log[charges]

Compare Distributions

Show	Distribution		AICc ^	BIC	-2*LogLikelihood
☑	Normal	—	3575.5835	3585.9724	3571.5745

Quantiles

100.0%	maximum	11.063045
99.5%		10.851596
97.5%		10.739055
90.0%		10.459149
75.0%	quartile	9.7224023
50.0%	median	9.1465517
25.0%	quartile	8.4624473
10.0%		7.7538929
2.5%		7.3967676
0.5%		7.0389863

Summary Statistics

Mean	9.0986587
Std Dev	0.9195271
Std Err Mean	0.0251383
Upper 95% Mean	9.1479736
Lower 95% Mean	9.0493439
N	1338
N Missing	0

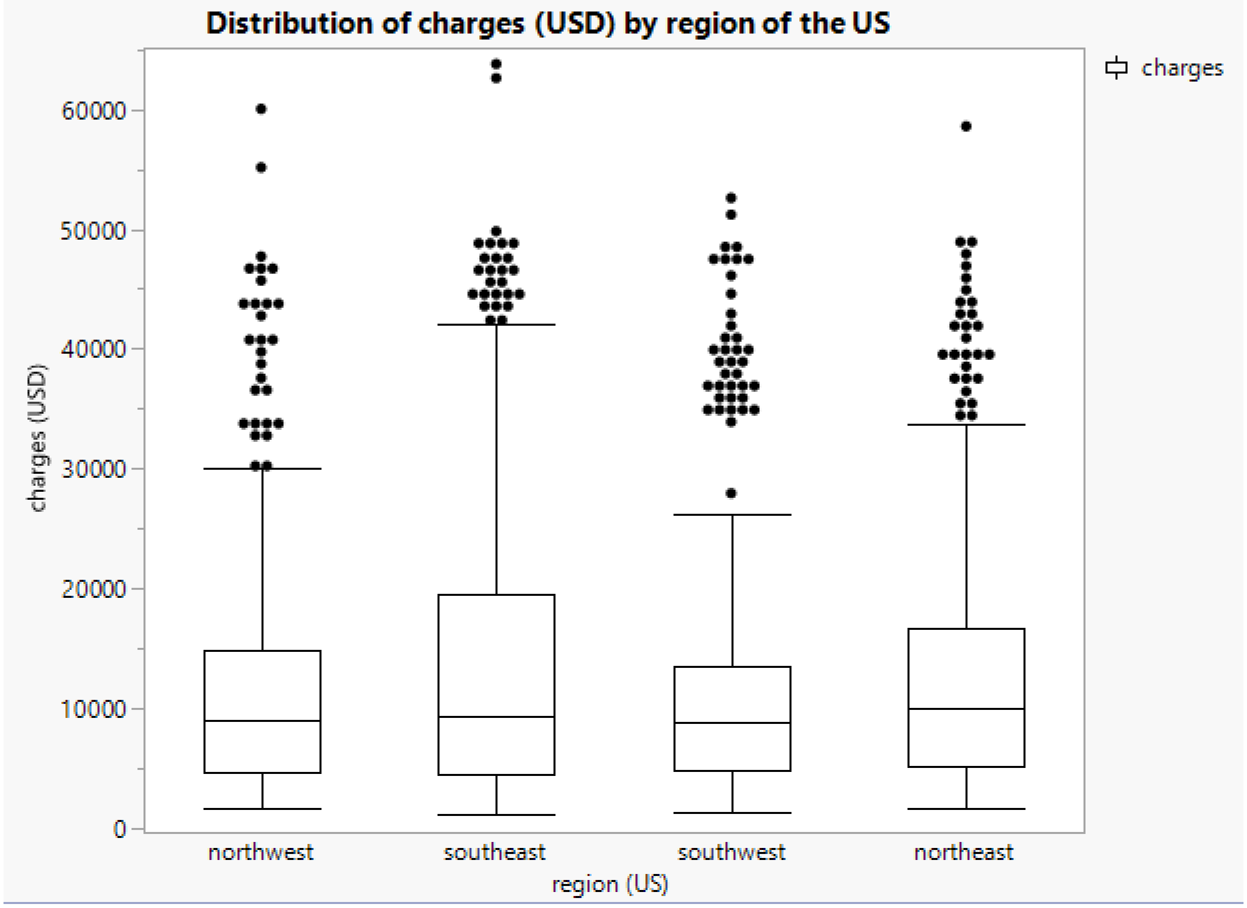
Fitted Normal Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Location	μ	9.0986587	0.0251383	9.0493439	9.1479736
Dispersion	σ	0.9195271	0.0177855	0.8859598	0.9557576

Measures	
-2*LogLikelihood	3571.5745
AICc	3575.5835
BIC	3585.9724

Figure A3

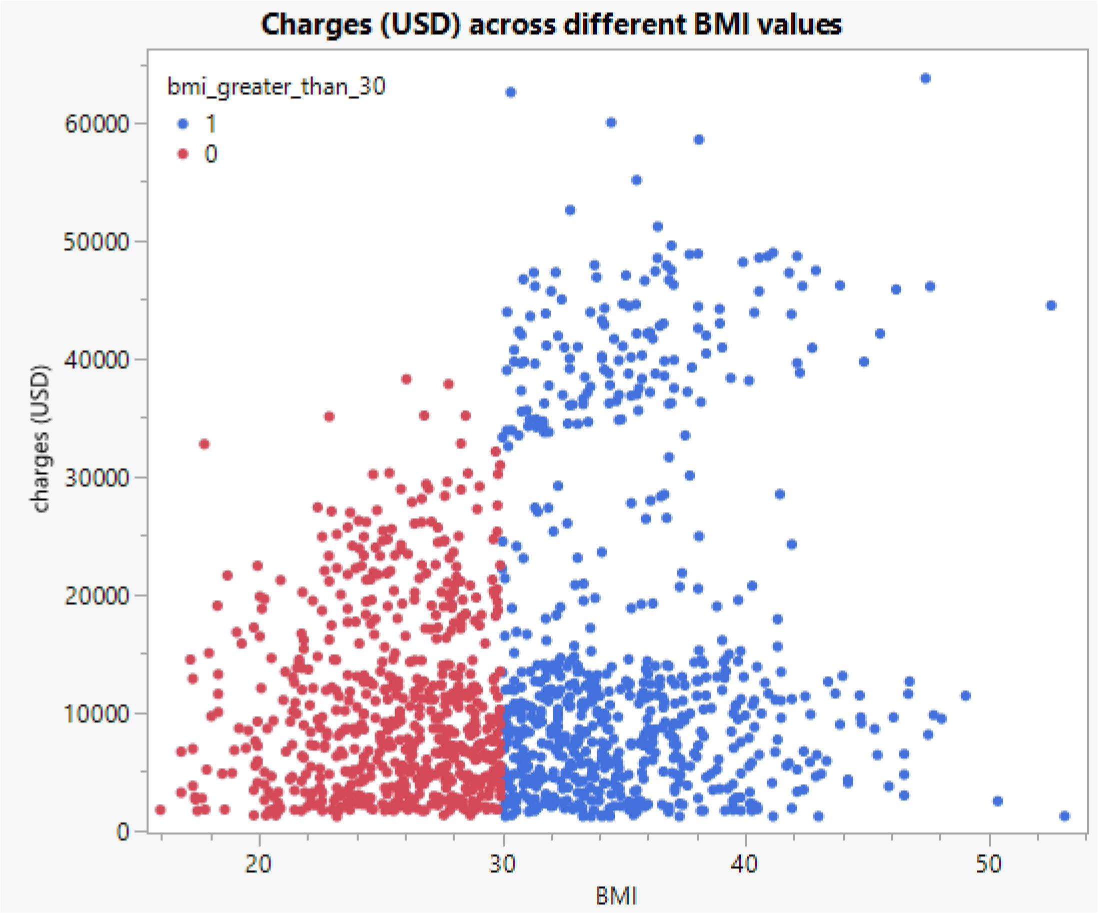
Distribution of charges (USD) by region of the US



Note. *Variance in charges across the four different regions of the US are not constant, with the most pronounced differences between the southeast and southwest regions.*

Figure A4

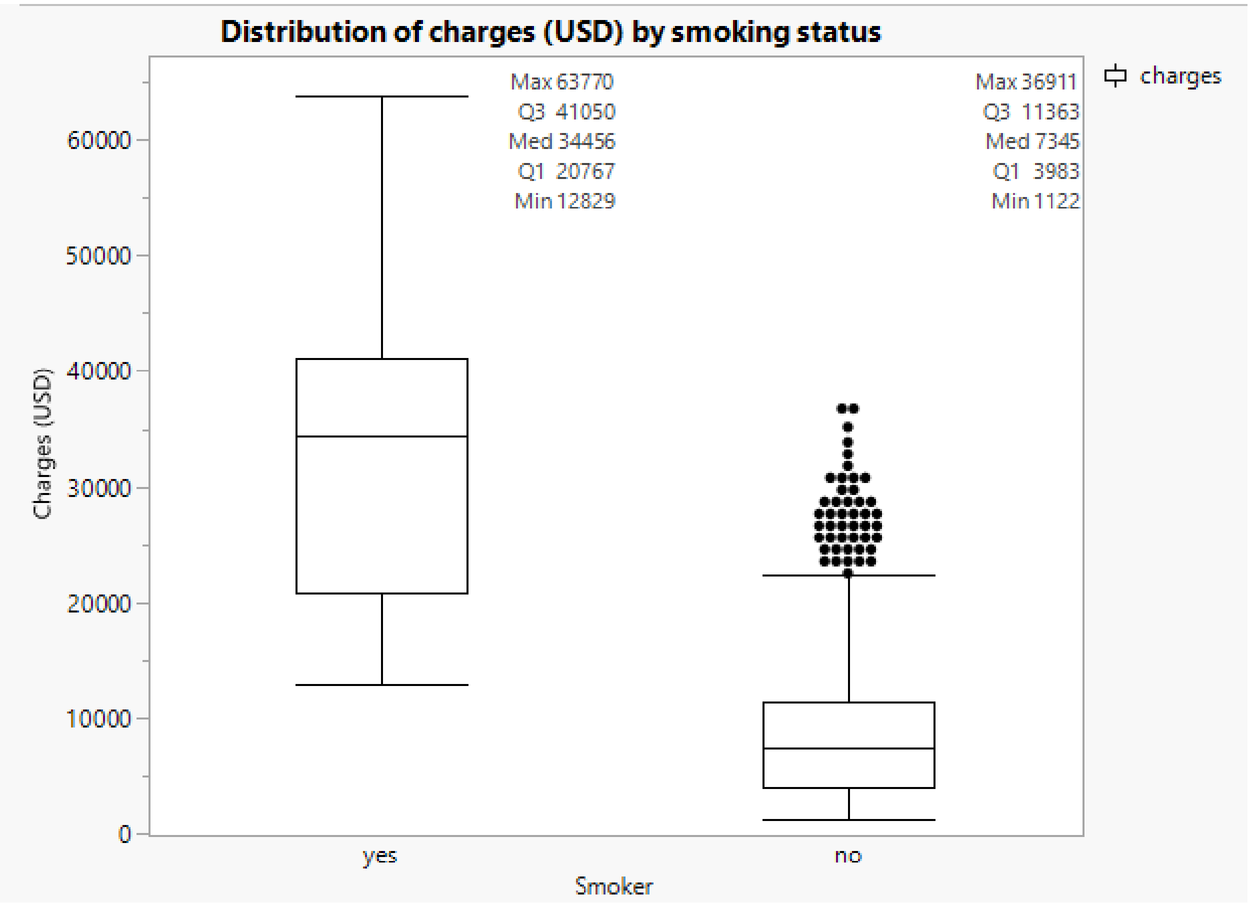
Charges (USD) across different BMI values



Note. *Variance in charges nearly doubles at the 30 BMI mark, research into the interaction of obesity (BMI>30) and smoking (Stewart, Cutler, & Rosen, 2009) provides some context on why we might be seeing this in our data.*

Figure A5

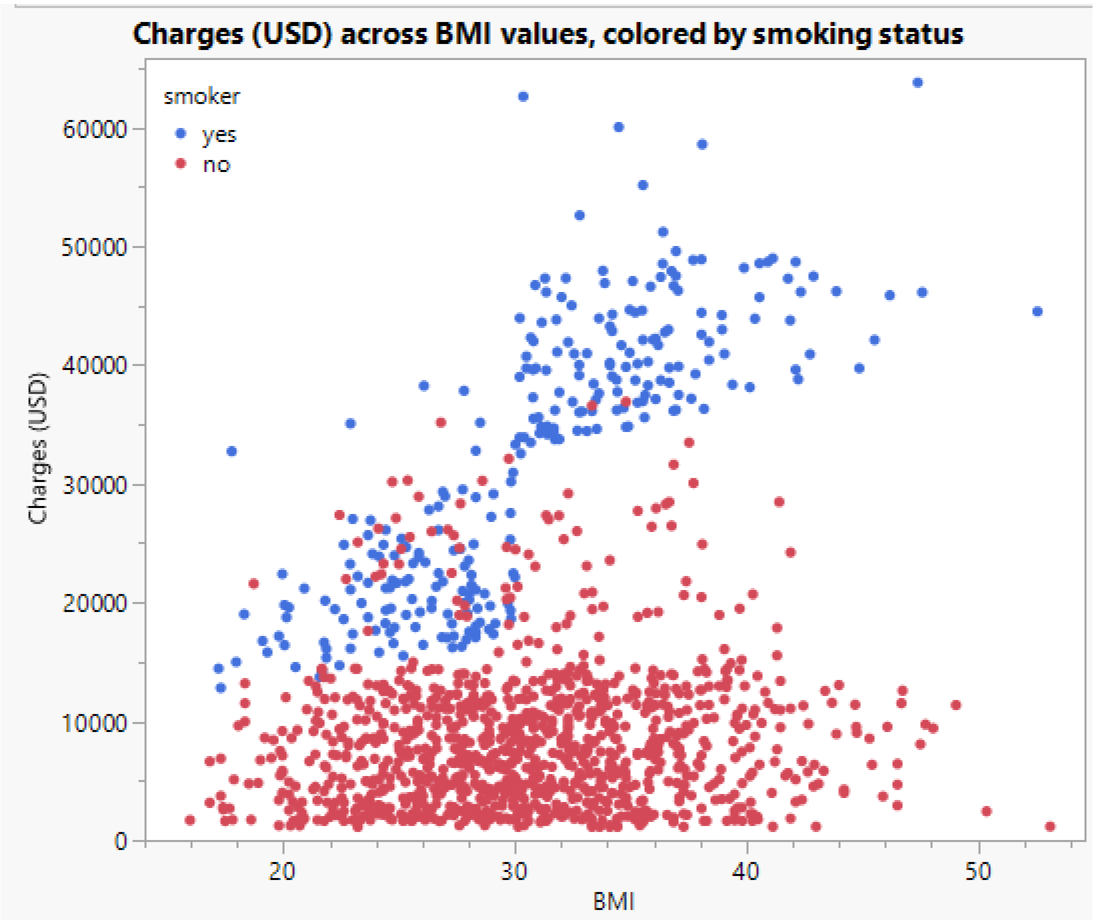
Distribution of charges (USD) by smoking status



Note. The cost differences between smokers and non-smokers is significant, as well as the variance in costs for the two groups.

Figure A6

Charges (USD) across BMI values, colored by smoking status



Note. Comparing the smoking and non-smoking groups, the spread of charges remains constant across all BMI values for non-smokers. Charges for smokers increase across BMI values, with a marked increase after the 30 BMI mark.

Figure A7

Average charges (USD) by region of the US

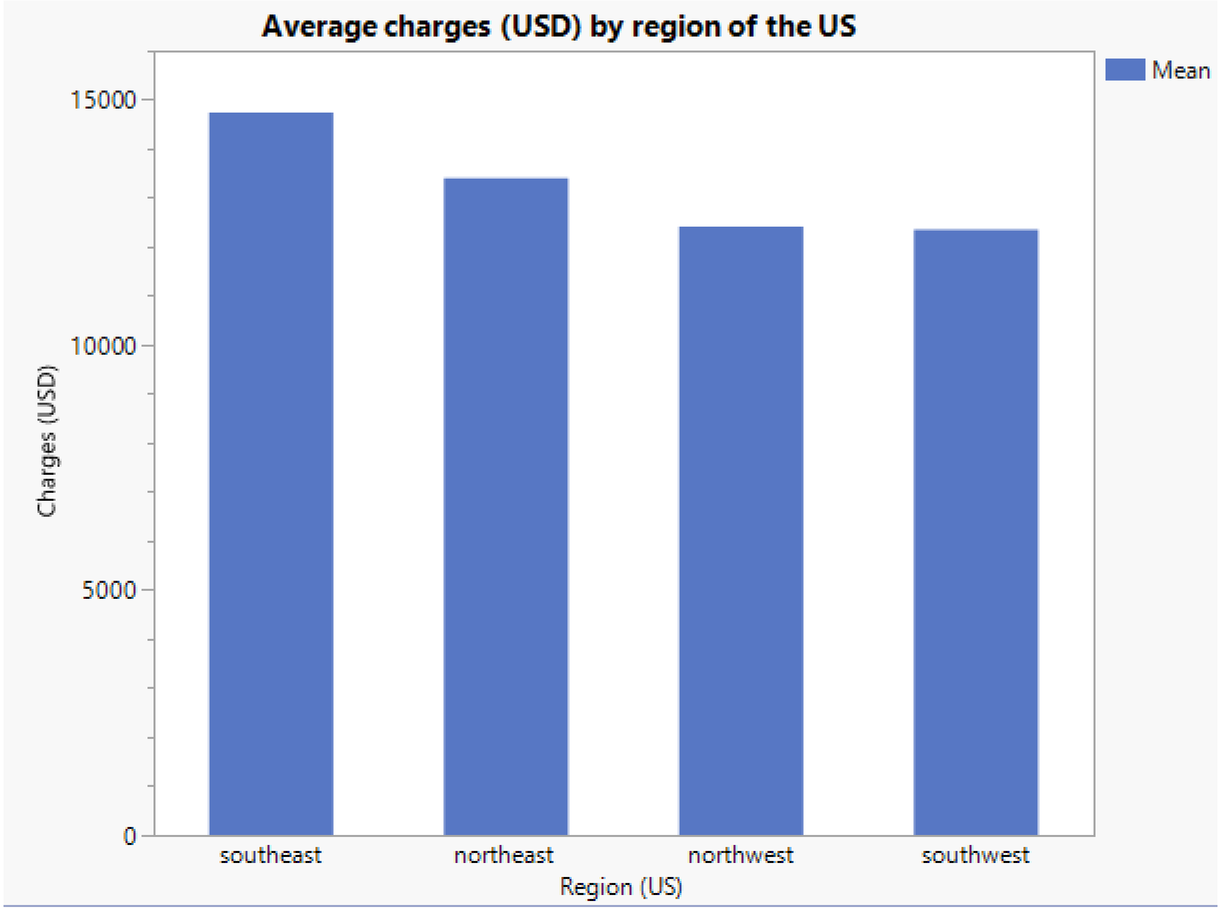


Figure A8

Counts of individuals by region, grouped by smoking status

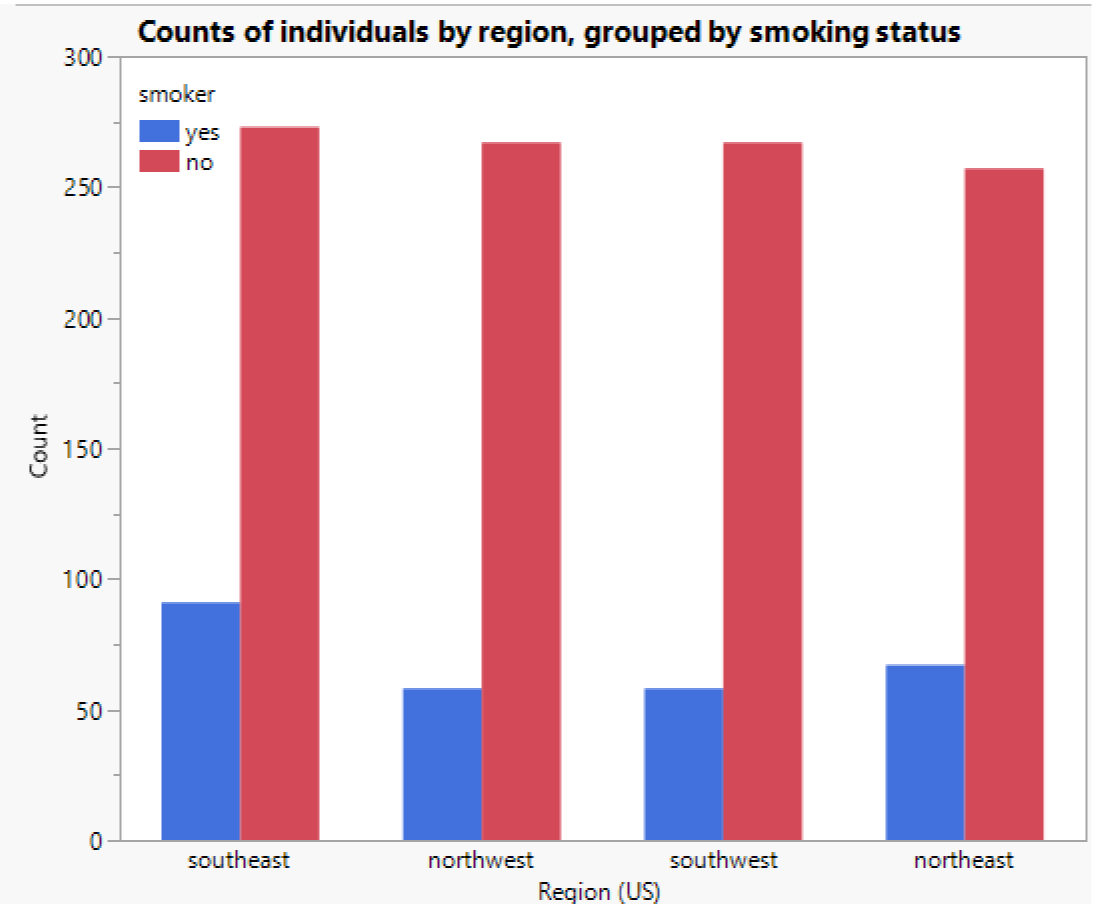


Figure A9

Counts of individuals by region, grouped by BMI > 30

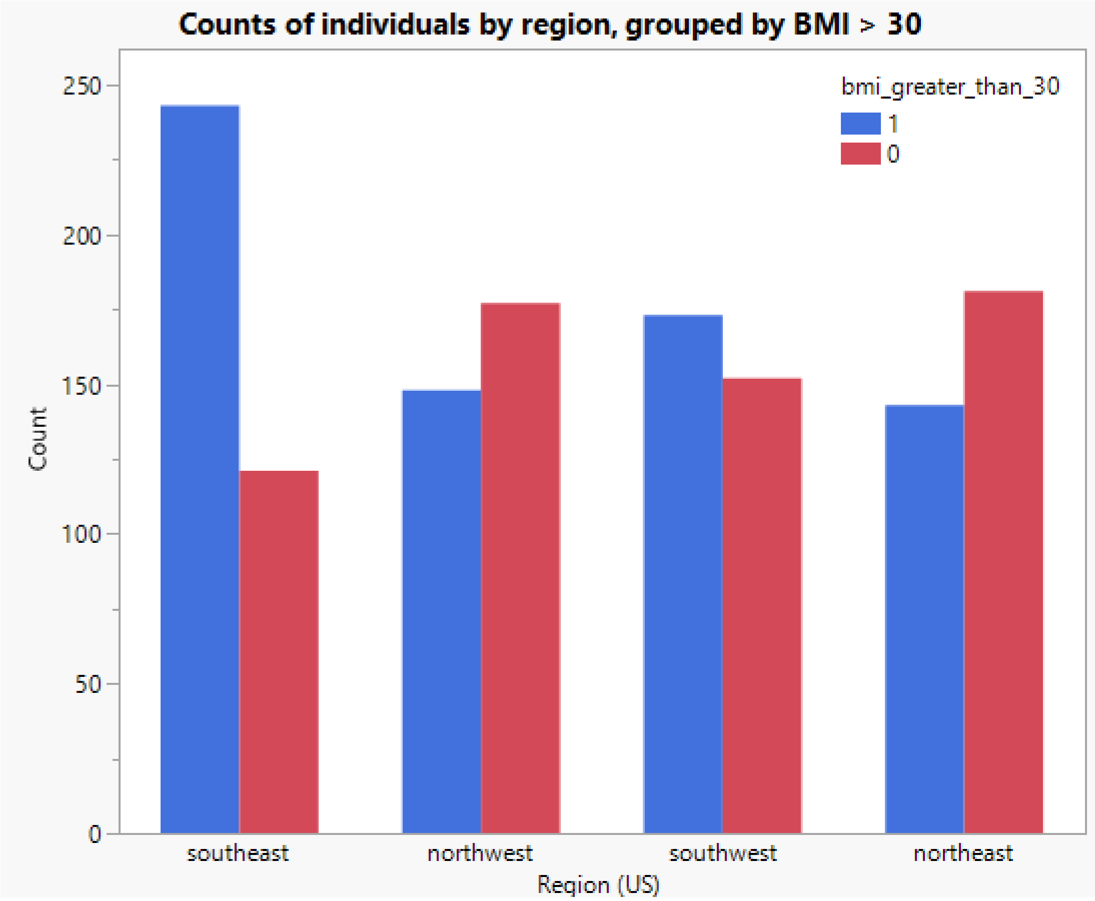
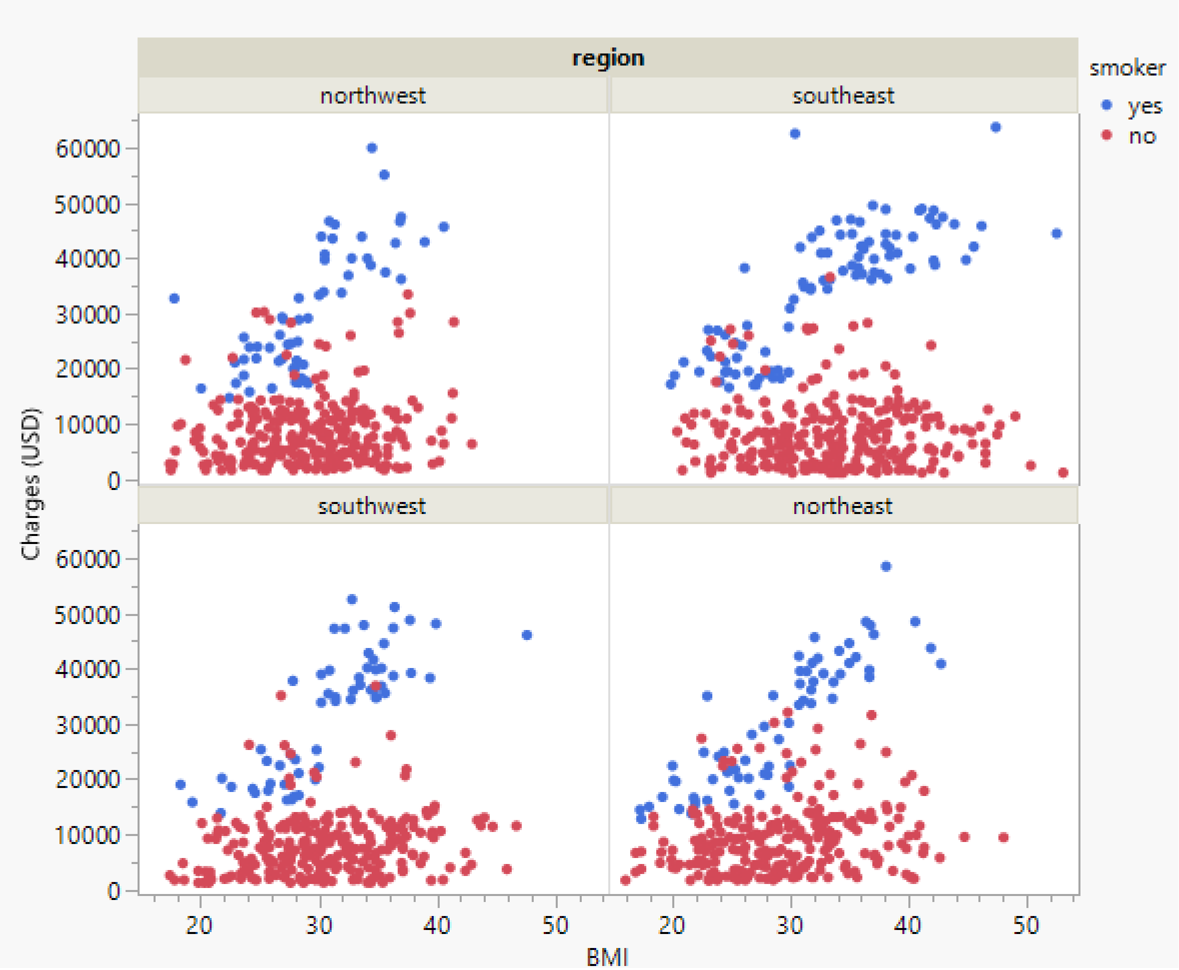


Figure A10

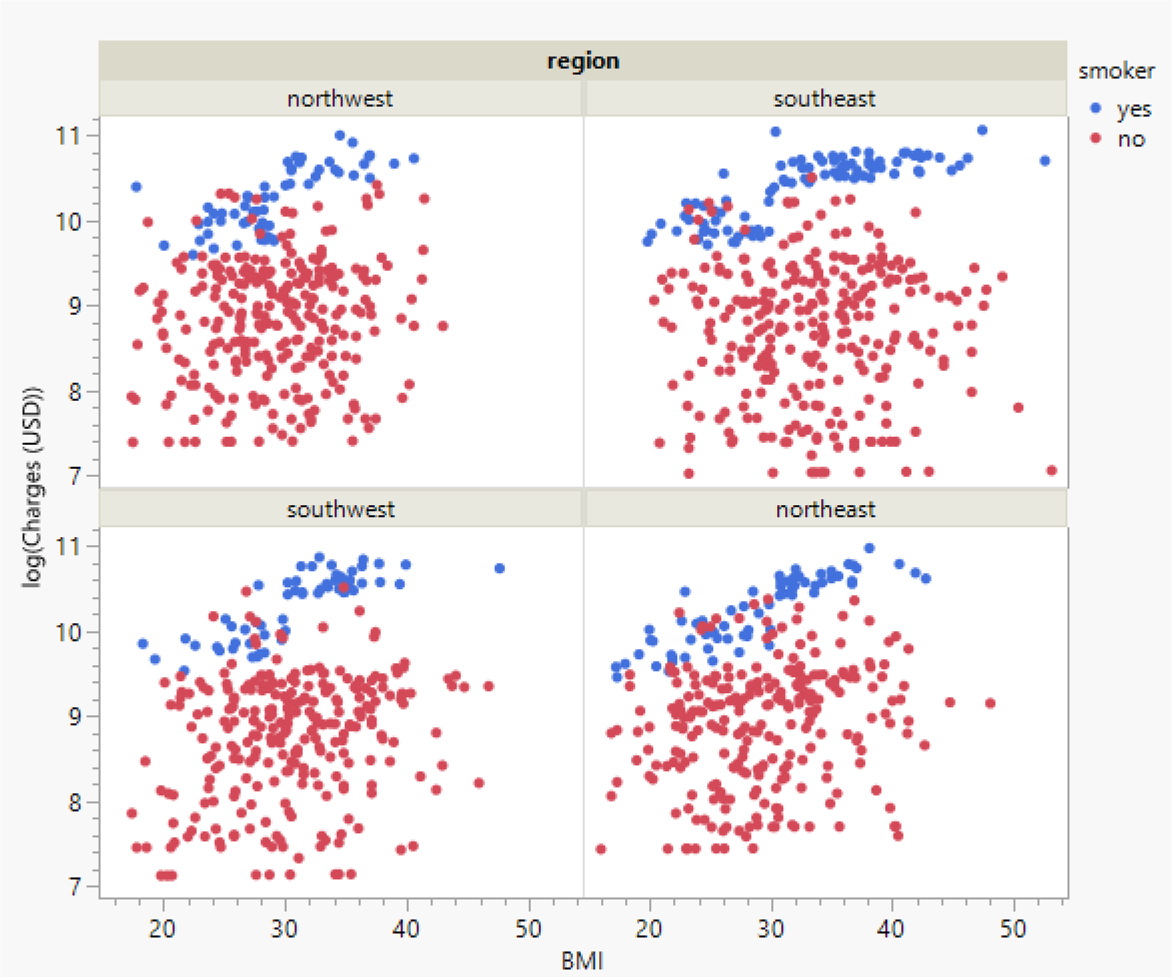
Raw Charges (USD) vs BMI, faceted by Region and colored by smoker status



Note. This figure pulls together the four insights from our EDA. In all four regions blue points (smokers) sit consistently higher than red points (non-smokers), confirming the significant difference in costs for smokers. Each region also shows the spread of charges increasing past the obesity threshold (BMI>30), with obese smokers seeing the highest jump in charges. The southeast region is clearly an outlier as well, with the cloud of high BMI smokers being larger than the other 3 regions.

Figure A11

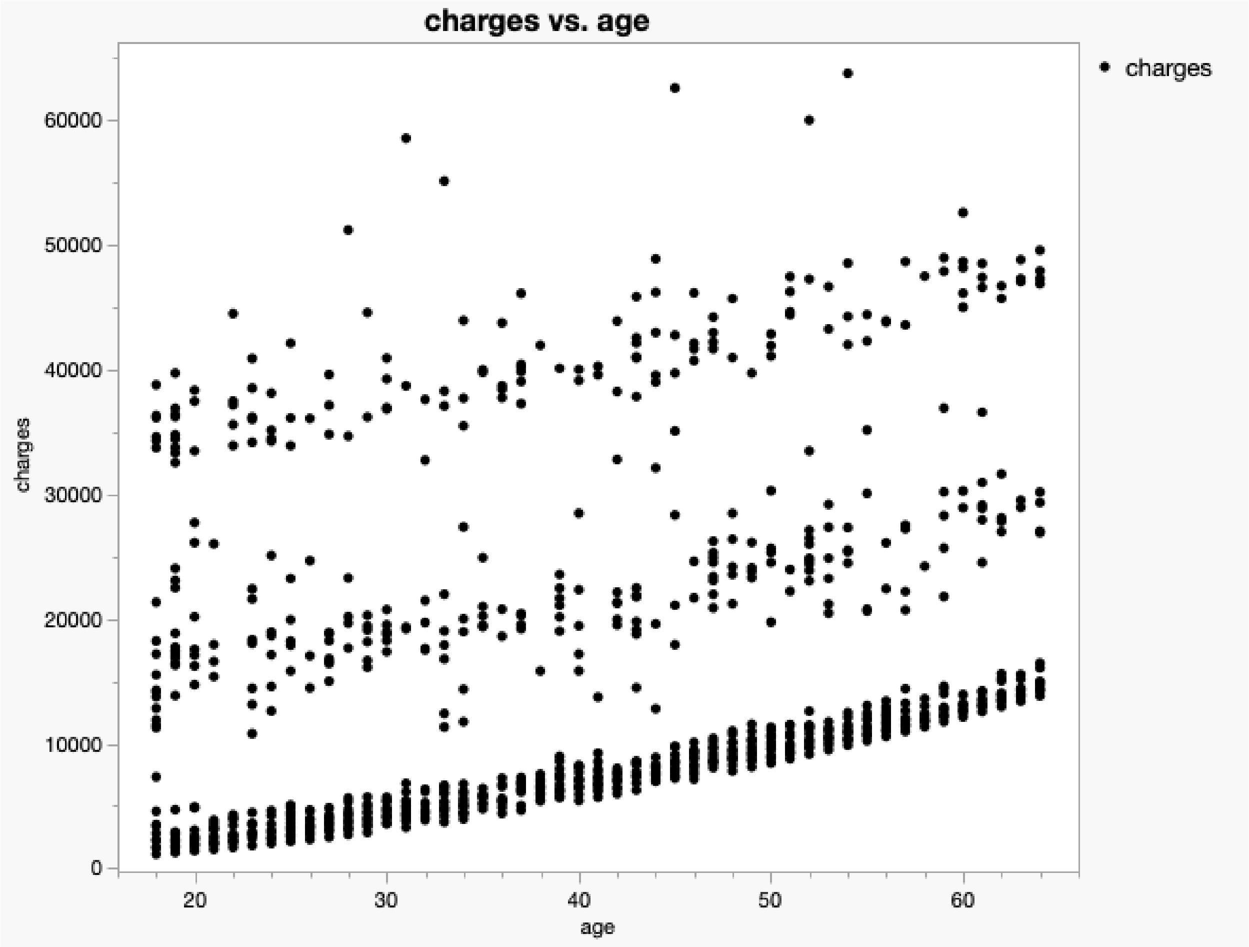
Log(Charges) (USD) vs BMI, faceted by Region and colored by smoker status



Note. This figure shows the same relationship as Figure A10, but on the logarithmic scale. The log transformation effectively evens out the spread for each region and satisfies the constant-variance assumption. The prevalence of high BMI smokers in the southeast region is still noticeable as well. This figure provides reassurance that the log transformation stabilizes variance without masking our key interactions identified.

Figure A12

Charges in relation to age.


charges vs. age
● charges
charges
60000
50000
40000
30000
20000
10000
0
20
30
40
50
60
age

Appendix B

Model 1 Fit and Diagnostics

Figure B1

Model 1 Actual by predicted plot

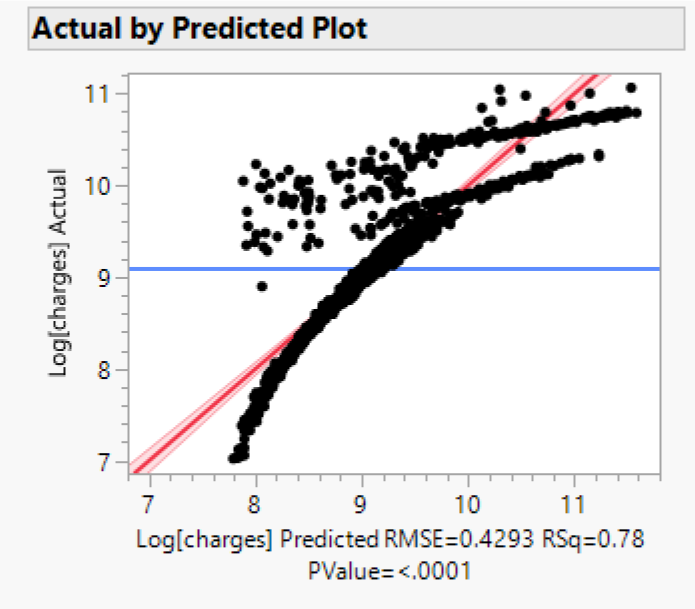
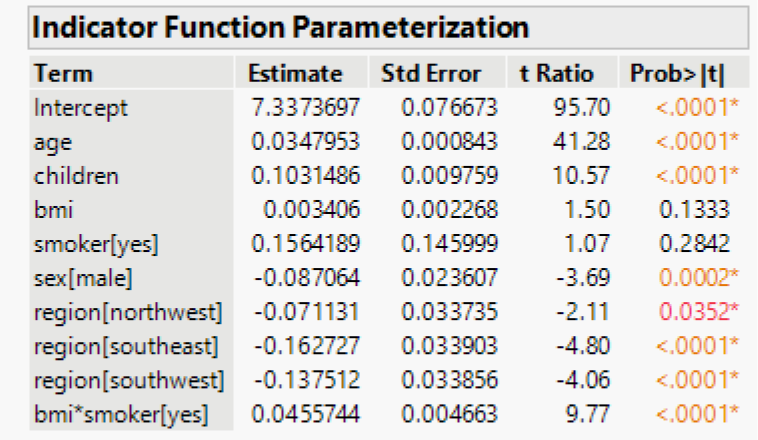


Figure B2

Model 1 Indicator Function Parameterization

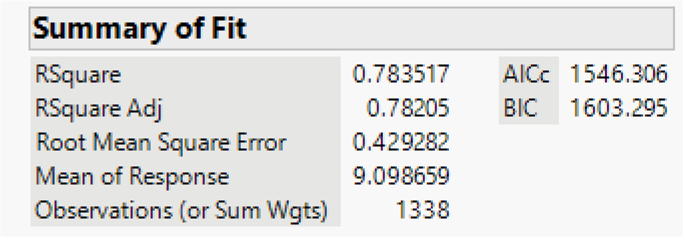


Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.3373697	0.076673	95.70	<.0001*
age	0.0347953	0.000843	41.28	<.0001*
children	0.1031486	0.009759	10.57	<.0001*
bmi	0.003406	0.002268	1.50	0.1333
smoker[yes]	0.1564189	0.145999	1.07	0.2842
sex[male]	-0.087064	0.023607	-3.69	0.0002*
region[northwest]	-0.071131	0.033735	-2.11	0.0352*
region[southeast]	-0.162727	0.033903	-4.80	<.0001*
region[southwest]	-0.137512	0.033856	-4.06	<.0001*
bmi*smoker[yes]	0.0455744	0.004663	9.77	<.0001*

Figure B3

Model 1 Summary of Fit

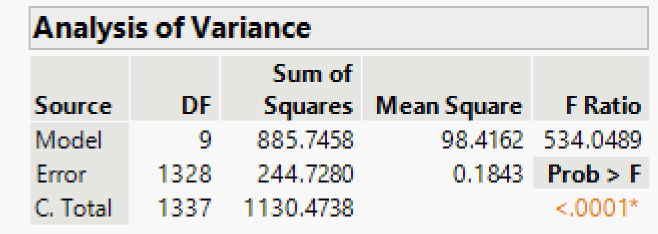


Summary of Fit

RSquare	0.783517	AICc	1546.306
RSquare Adj	0.78205	BIC	1603.295
Root Mean Square Error	0.429282		
Mean of Response	9.098659		
Observations (or Sum Wgts)	1338		

Figure B4

Model 1 Analysis of Variance



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	885.7458	98.4162	534.0489
Error	1328	244.7280	0.1843	**Prob > F**
C. Total	1337	1130.4738		<.0001*

Figure B5

Model 1 Residual Normal Quantile Plot

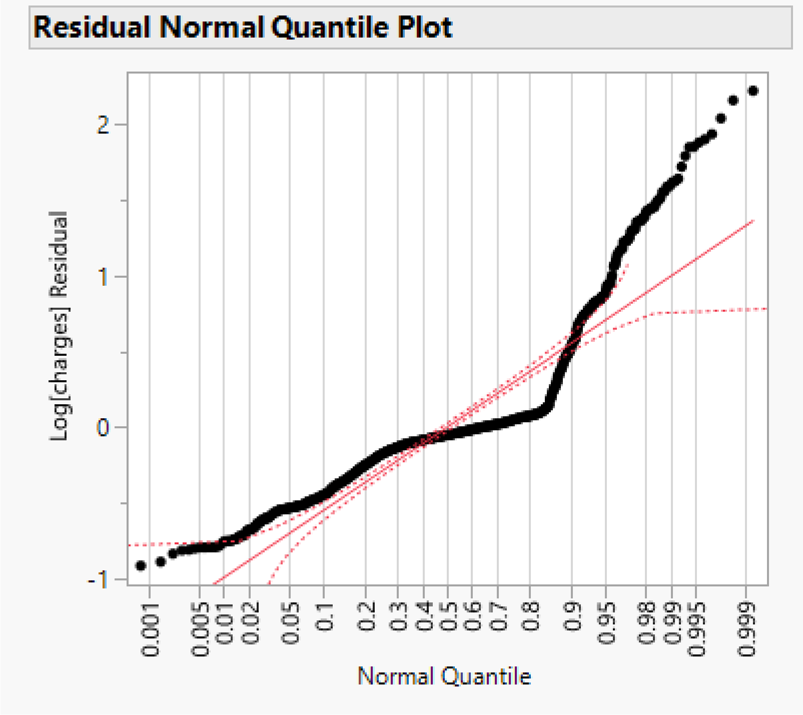


Figure B6

Model 1Residual by Predicted plot

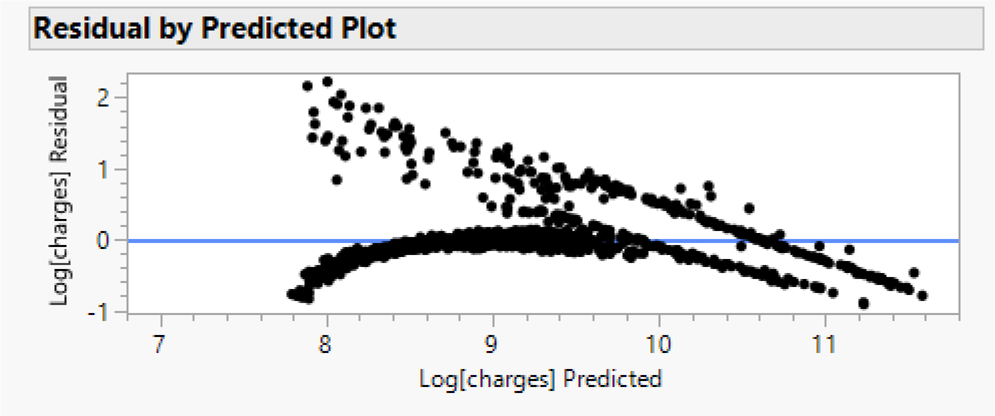
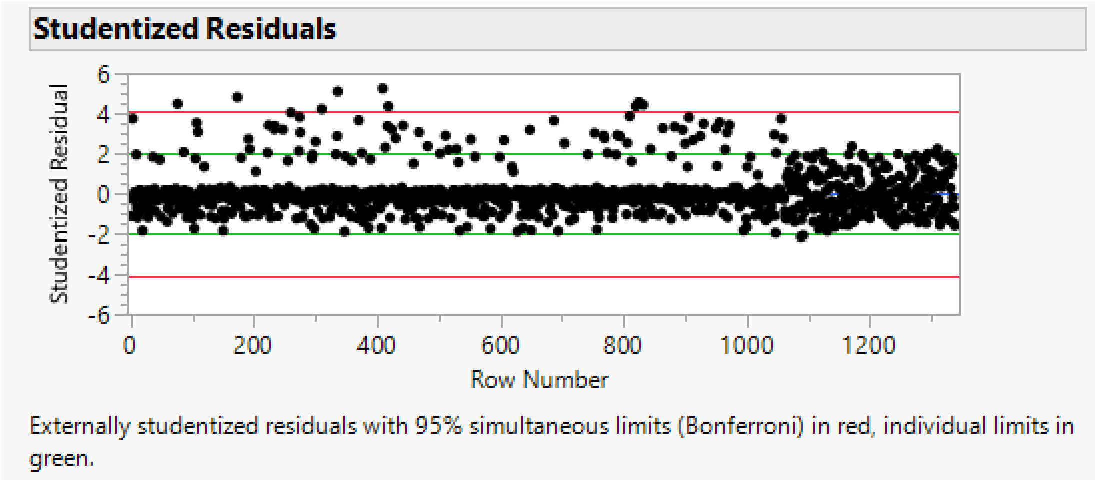


Figure B7

Model 1 Studentized Residuals



Externally studentized residuals with 95% simultaneous limits (Bonferroni) in red, individual limits in green.

Appendix C

Model 2 Fit and Diagnostics

Figure C1

Model 2 Actual by predicted plot

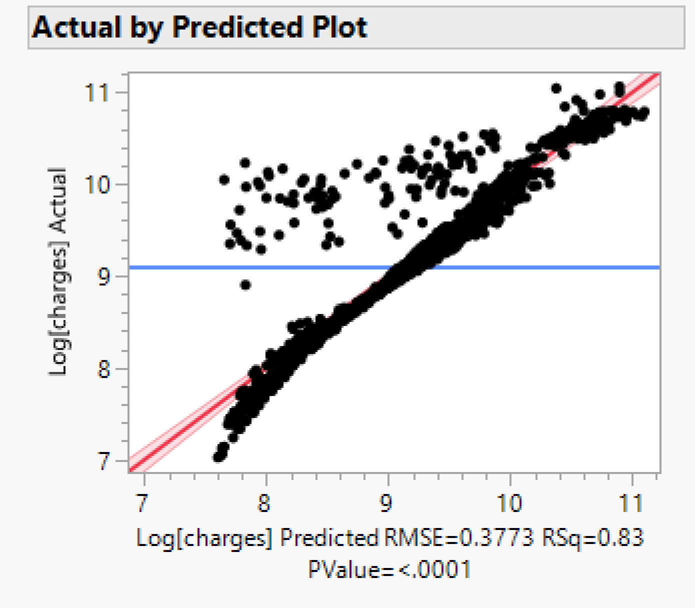
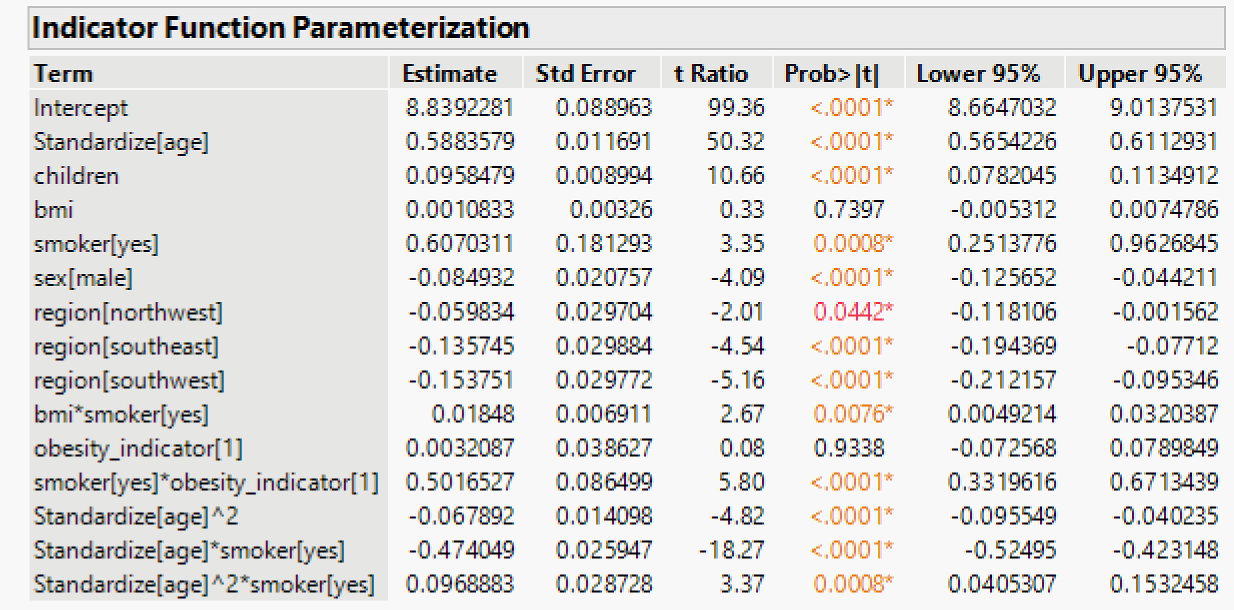


Figure C2

Model 1 Indicator Function Parameterization

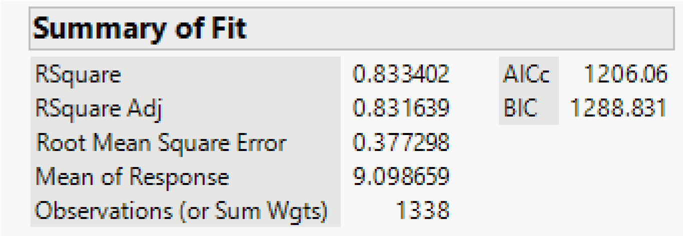


Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	8.8392281	0.088963	99.36	<.0001*	8.6647032	9.0137531
Standardize[age]	0.5883579	0.011691	50.32	<.0001*	0.5654226	0.6112931
children	0.0958479	0.008994	10.66	<.0001*	0.0782045	0.1134912
bmi	0.0010833	0.00326	0.33	0.7397	-0.005312	0.0074786
smoker[yes]	0.6070311	0.181293	3.35	0.0008*	0.2513776	0.9626845
sex[male]	-0.084932	0.020757	-4.09	<.0001*	-0.125652	-0.044211
region[northwest]	-0.059834	0.029704	-2.01	0.0442*	-0.118106	-0.001562
region[southeast]	-0.135745	0.029884	-4.54	<.0001*	-0.194369	-0.07712
region[southwest]	-0.153751	0.029772	-5.16	<.0001*	-0.212157	-0.095346
bmi*smoker[yes]	0.01848	0.006911	2.67	0.0076*	0.0049214	0.0320387
obesity_indicator[1]	0.0032087	0.038627	0.08	0.9338	-0.072568	0.0789849
smoker[yes]*obesity_indicator[1]	0.5016527	0.086499	5.80	<.0001*	0.3319616	0.6713439
Standardize[age]^2	-0.067892	0.014098	-4.82	<.0001*	-0.095549	-0.040235
Standardize[age]*smoker[yes]	-0.474049	0.025947	-18.27	<.0001*	-0.52495	-0.423148
Standardize[age]^2*smoker[yes]	0.0968883	0.028728	3.37	0.0008*	0.0405307	0.1532458

Figure C3

Model 2 Summary of fit

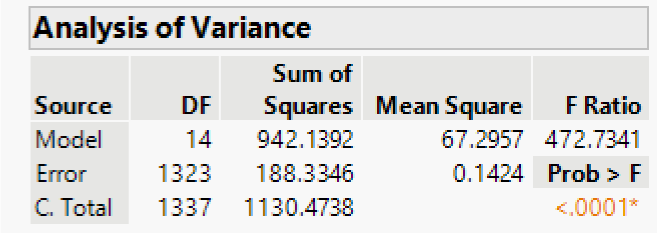


Summary of Fit

RSquare	0.833402	AICc	1206.06
RSquare Adj	0.831639	BIC	1288.831
Root Mean Square Error	0.377298		
Mean of Response	9.098659		
Observations (or Sum Wgts)	1338		

Figure C4

Model 2 Analysis of Variance



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	14	942.1392	67.2957	472.7341
Error	1323	188.3346	0.1424	**Prob > F**
C. Total	1337	1130.4738		<.0001*

Figure C5

Model 2 Residual Normal Quantile Plot

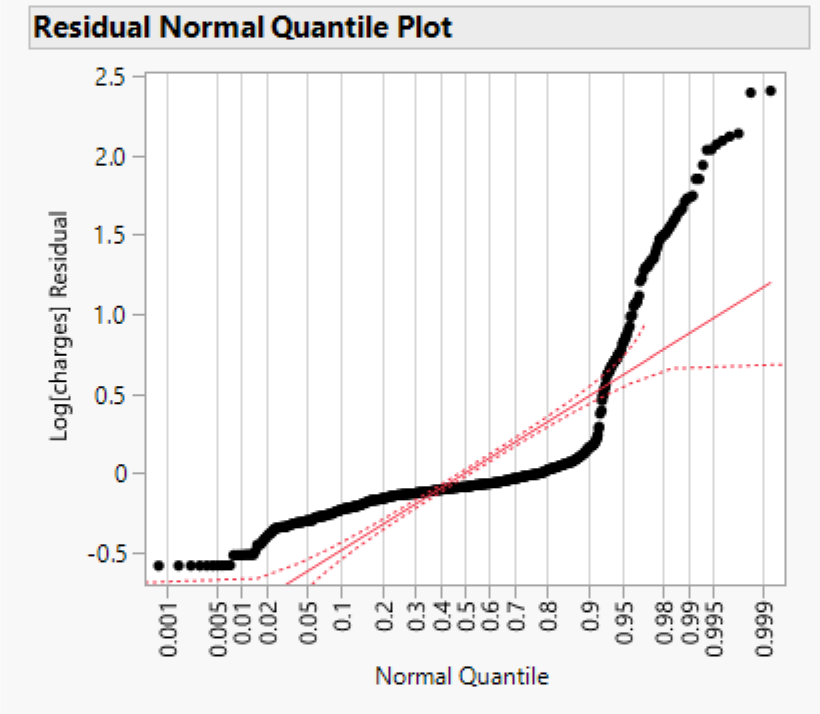


Figure C6

Model 2 Residual by Predicted plot

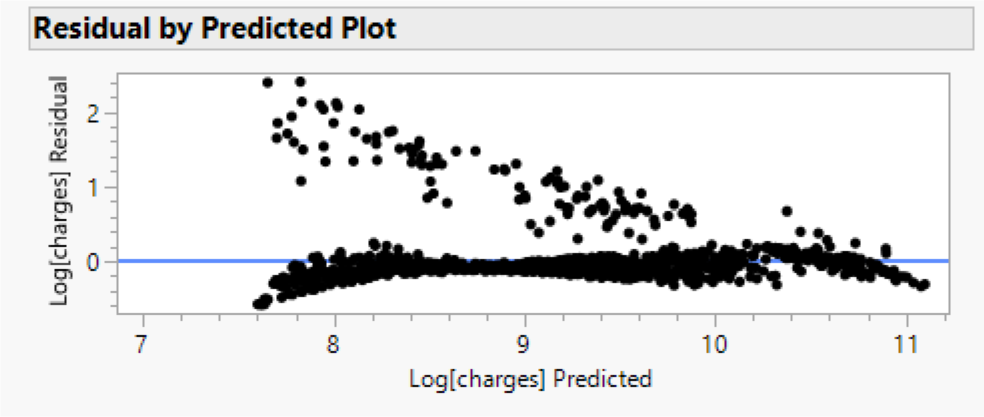
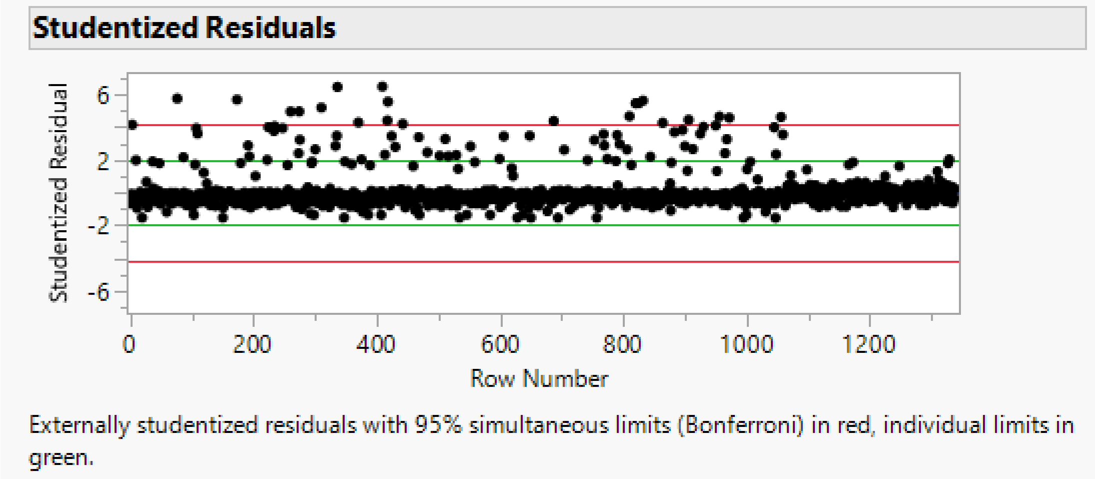


Figure C7

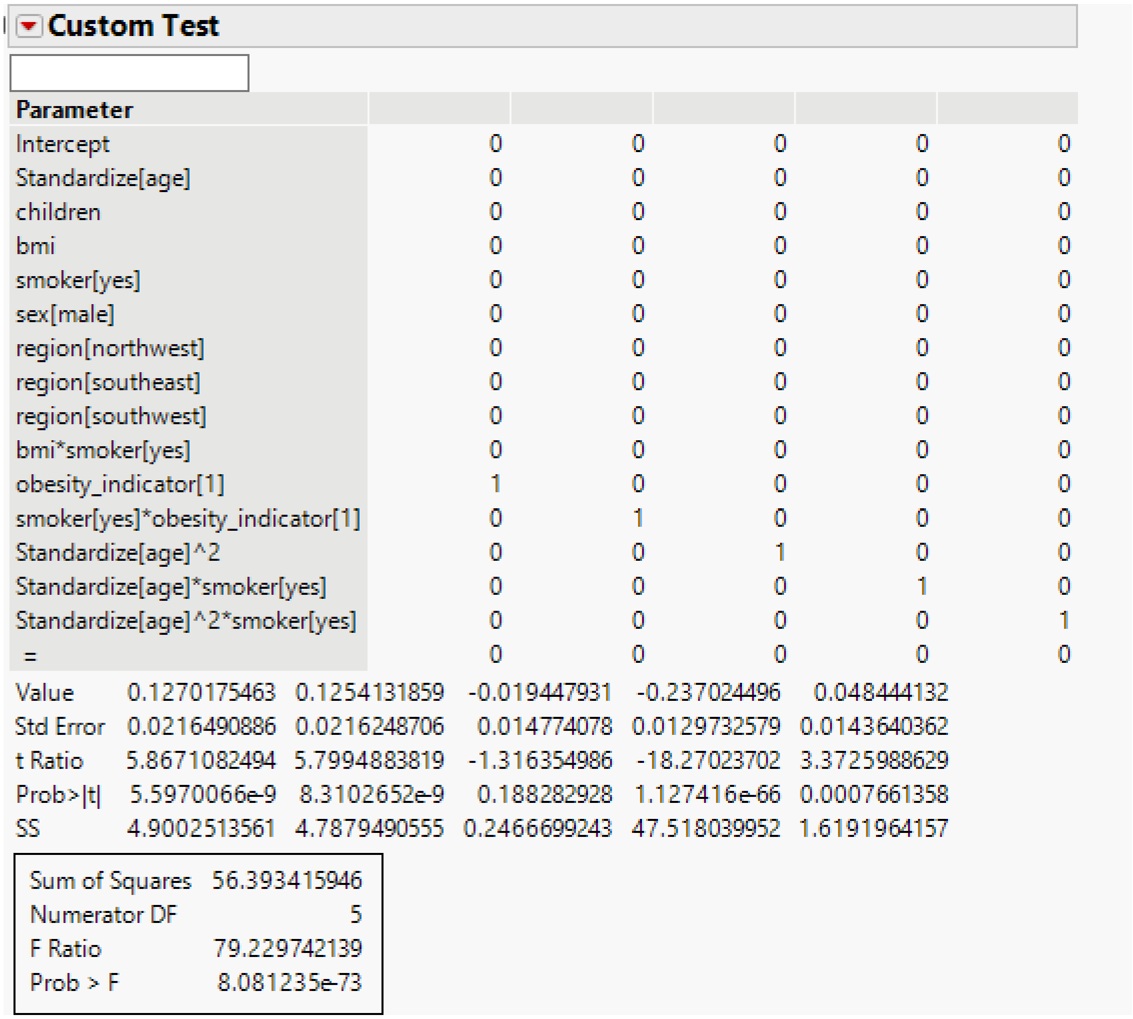
Model 2 Studentized Residuals



Externally studentized residuals with 95% simultaneous limits (Bonferroni) in red, individual limits in green.

Figure C8

Model 2 Partial F-test



Custom Test

Parameter					
Intercept	0	0	0	0	0
Standardize[age]	0	0	0	0	0
children	0	0	0	0	0
bmi	0	0	0	0	0
smoker[yes]	0	0	0	0	0
sex[male]	0	0	0	0	0
region[northwest]	0	0	0	0	0
region[southeast]	0	0	0	0	0
region[southwest]	0	0	0	0	0
bmi*smoker[yes]	0	0	0	0	0
obesity_indicator[1]	1	0	0	0	0
smoker[yes]*obesity_indicator[1]	0	1	0	0	0
Standardize[age]^2	0	0	1	0	0
Standardize[age]*smoker[yes]	0	0	0	1	0
Standardize[age]^2*smoker[yes]	0	0	0	0	1
=	0	0	0	0	0

Value	0.1270175463	0.1254131859	-0.019447931	-0.237024496	0.048444132
Std Error	0.0216490886	0.0216248706	0.014774078	0.0129732579	0.0143640362
t Ratio	5.8671082494	5.7994883819	-1.316354986	-18.27023702	3.3725988629
Prob>|t|	5.5970066e-9	8.3102652e-9	0.188282928	1.127416e-66	0.0007661358
SS	4.9002513561	4.7879490555	0.2466699243	47.518039952	1.6191964157

Sum of Squares	56.393415946
Numerator DF	5
F Ratio	79.229742139
Prob > F	8.081235e-73

Appendix D

Model 3 Fit and Diagnostics

Figure D1

Model 3 Actual by predicted plot

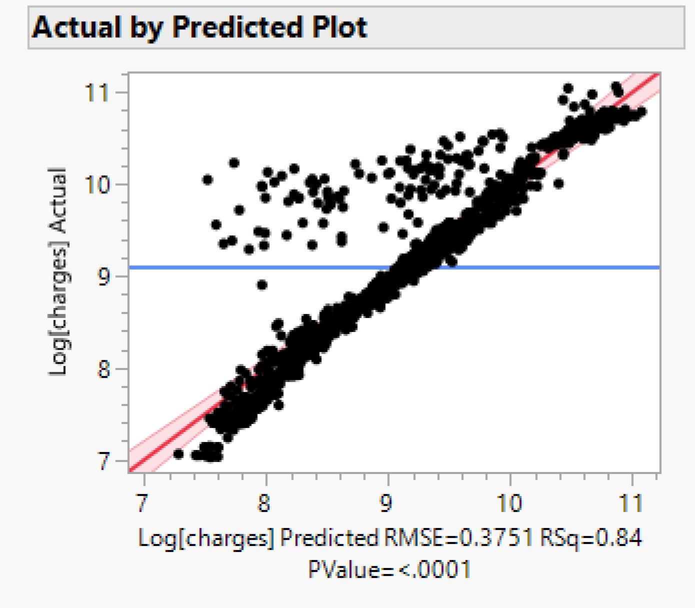
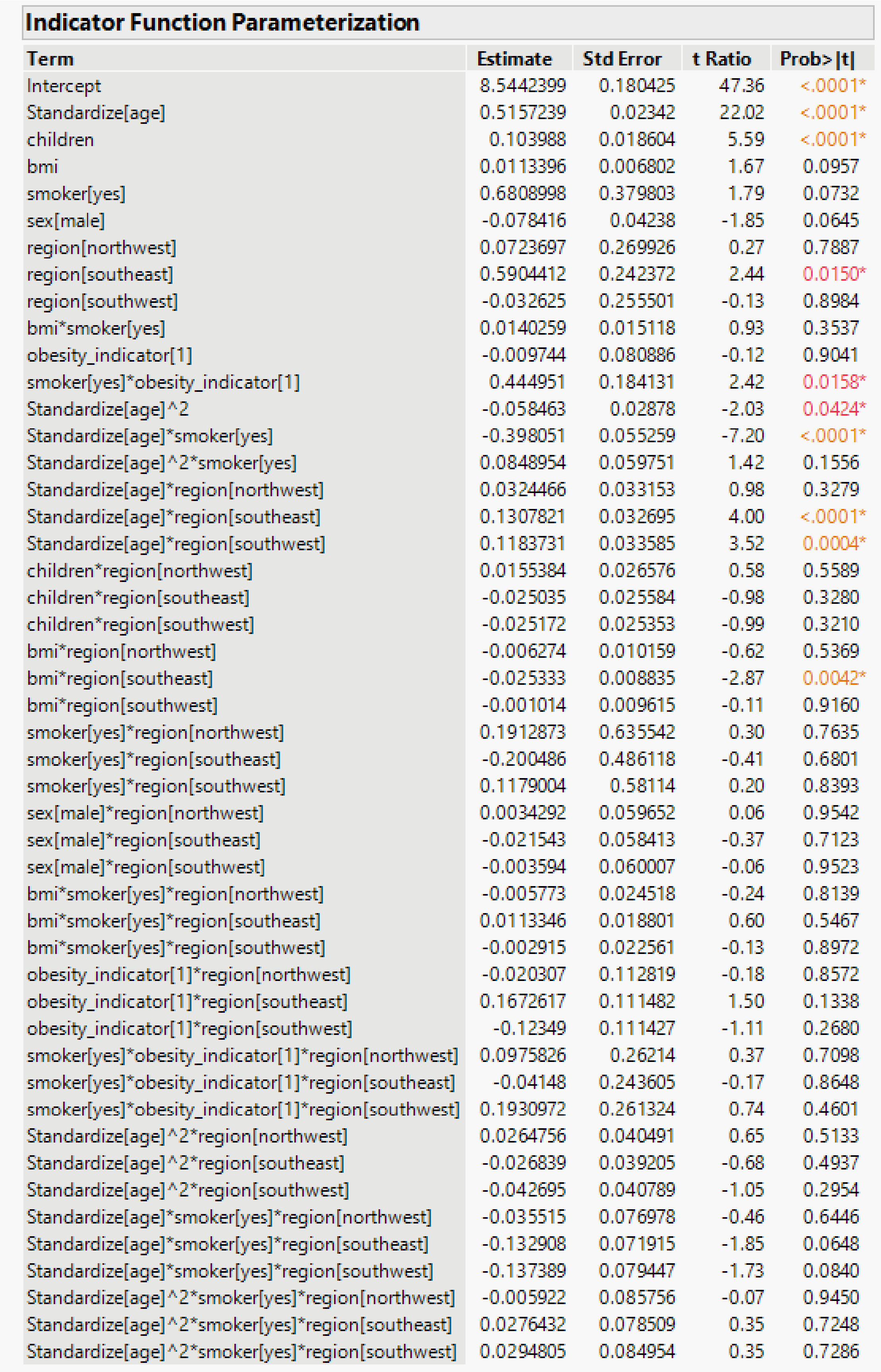


Figure D2

Model 3 Indicator Function Parameterization

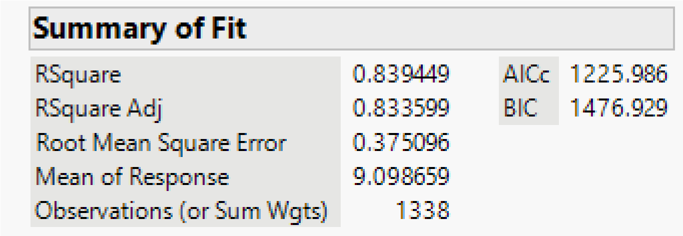


Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	8.5442399	0.180425	47.36	<.0001*
Standardize[age]	0.5157239	0.02342	22.02	<.0001*
children	0.103988	0.018604	5.59	<.0001*
bmi	0.0113396	0.006802	1.67	0.0957
smoker[yes]	0.6808998	0.379803	1.79	0.0732
sex[male]	-0.078416	0.04238	-1.85	0.0645
region[northwest]	0.0723697	0.269926	0.27	0.7887
region[southeast]	0.5904412	0.242372	2.44	0.0150*
region[southwest]	-0.032625	0.255501	-0.13	0.8984
bmi*smoker[yes]	0.0140259	0.015118	0.93	0.3537
obesity_indicator[1]	-0.009744	0.080886	-0.12	0.9041
smoker[yes]*obesity_indicator[1]	0.444951	0.184131	2.42	0.0158*
Standardize[age]^2	-0.058463	0.02878	-2.03	0.0424*
Standardize[age]*smoker[yes]	-0.398051	0.055259	-7.20	<.0001*
Standardize[age]^2*smoker[yes]	0.0848954	0.059751	1.42	0.1556
Standardize[age]*region[northwest]	0.0324466	0.033153	0.98	0.3279
Standardize[age]*region[southeast]	0.1307821	0.032695	4.00	<.0001*
Standardize[age]*region[southwest]	0.1183731	0.033585	3.52	0.0004*
children*region[northwest]	0.0155384	0.026576	0.58	0.5589
children*region[southeast]	-0.025035	0.025584	-0.98	0.3280
children*region[southwest]	-0.025172	0.025353	-0.99	0.3210
bmi*region[northwest]	-0.006274	0.010159	-0.62	0.5369
bmi*region[southeast]	-0.025333	0.008835	-2.87	0.0042*
bmi*region[southwest]	-0.001014	0.009615	-0.11	0.9160
smoker[yes]*region[northwest]	0.1912873	0.635542	0.30	0.7635
smoker[yes]*region[southeast]	-0.200486	0.486118	-0.41	0.6801
smoker[yes]*region[southwest]	0.1179004	0.58114	0.20	0.8393
sex[male]*region[northwest]	0.0034292	0.059652	0.06	0.9542
sex[male]*region[southeast]	-0.021543	0.058413	-0.37	0.7123
sex[male]*region[southwest]	-0.003594	0.060007	-0.06	0.9523
bmi*smoker[yes]*region[northwest]	-0.005773	0.024518	-0.24	0.8139
bmi*smoker[yes]*region[southeast]	0.0113346	0.018801	0.60	0.5467
bmi*smoker[yes]*region[southwest]	-0.002915	0.022561	-0.13	0.8972
obesity_indicator[1]*region[northwest]	-0.020307	0.112819	-0.18	0.8572
obesity_indicator[1]*region[southeast]	0.1672617	0.111482	1.50	0.1338
obesity_indicator[1]*region[southwest]	-0.12349	0.111427	-1.11	0.2680
smoker[yes]*obesity_indicator[1]*region[northwest]	0.0975826	0.26214	0.37	0.7098
smoker[yes]*obesity_indicator[1]*region[southeast]	-0.04148	0.243605	-0.17	0.8648
smoker[yes]*obesity_indicator[1]*region[southwest]	0.1930972	0.261324	0.74	0.4601
Standardize[age]^2*region[northwest]	0.0264756	0.040491	0.65	0.5133
Standardize[age]^2*region[southeast]	-0.026839	0.039205	-0.68	0.4937
Standardize[age]^2*region[southwest]	-0.042695	0.040789	-1.05	0.2954
Standardize[age]*smoker[yes]*region[northwest]	-0.035515	0.076978	-0.46	0.6446
Standardize[age]*smoker[yes]*region[southeast]	-0.132908	0.071915	-1.85	0.0648
Standardize[age]*smoker[yes]*region[southwest]	-0.137389	0.079447	-1.73	0.0840
Standardize[age]^2*smoker[yes]*region[northwest]	-0.005922	0.085756	-0.07	0.9450
Standardize[age]^2*smoker[yes]*region[southeast]	0.0276432	0.078509	0.35	0.7248
Standardize[age]^2*smoker[yes]*region[southwest]	0.0294805	0.084954	0.35	0.7286

Figure D3

Model 3 Summary of fit

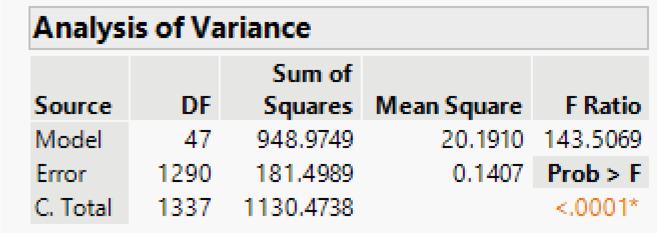


Summary of Fit

RSquare	0.839449	AICc	1225.986
RSquare Adj	0.833599	BIC	1476.929
Root Mean Square Error	0.375096		
Mean of Response	9.098659		
Observations (or Sum Wgts)	1338		

Figure D4

Model 3 Analysis of Variance



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	47	948.9749	20.1910	143.5069
Error	1290	181.4989	0.1407	Prob > F
C. Total	1337	1130.4738		<.0001*

Figure D5

Model 3 Residual Normal Quantile Plot

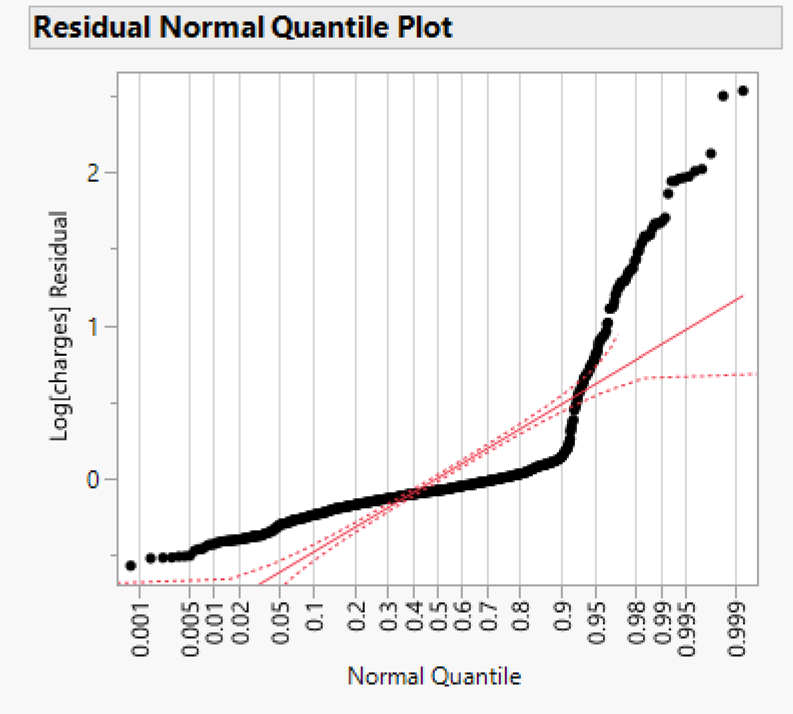


Figure D6

Model 3 Residual by Predicted Plot

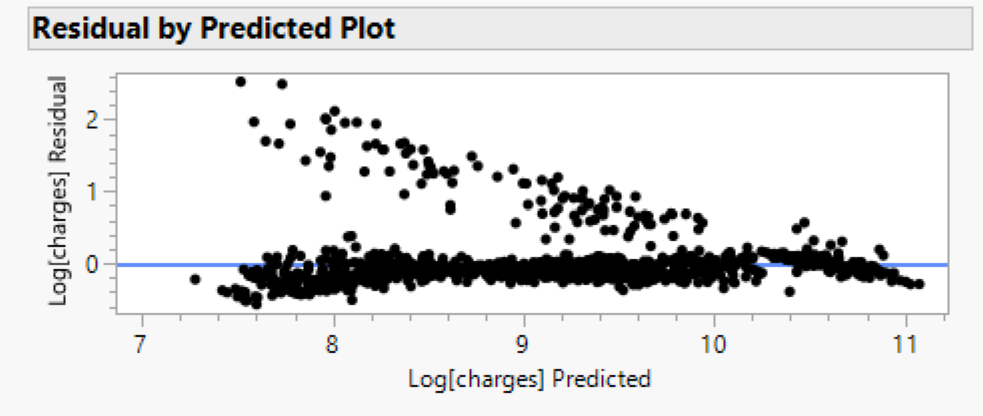
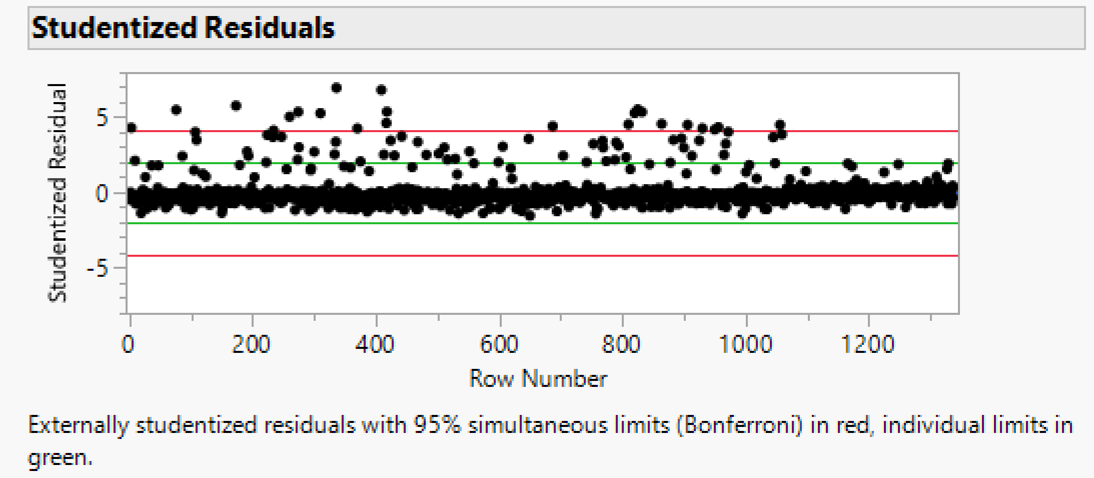


Figure D7

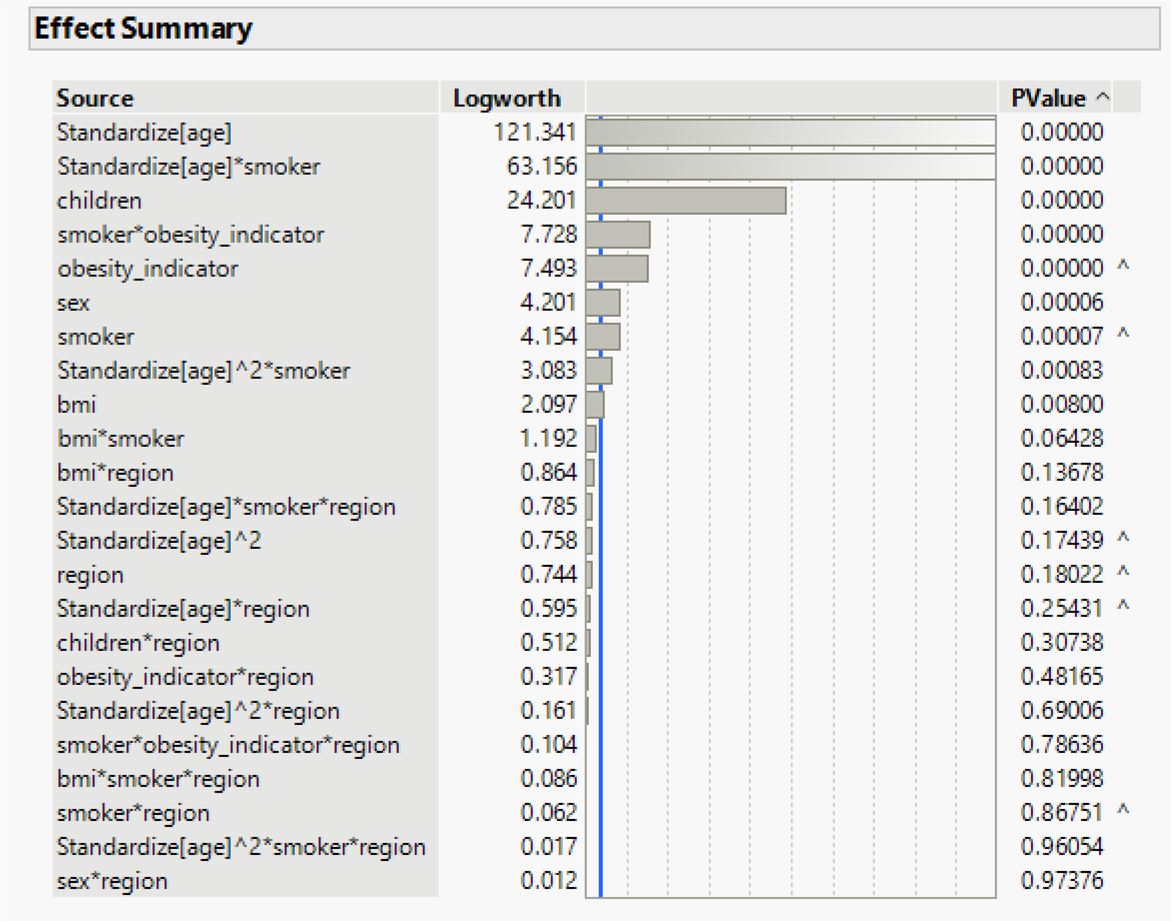
Model 3 Studentized Residuals



Externally studentized residuals with 95% simultaneous limits (Bonferroni) in red, individual limits in green.

Figure D8

Model 3 Effect Summary

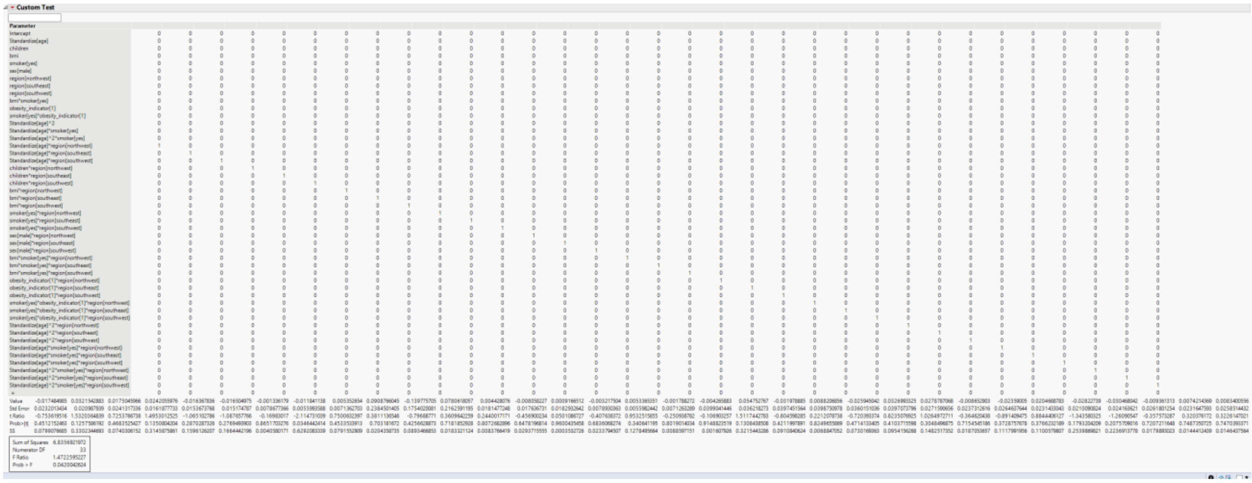


Effect Summary

Source	Logworth	PValue ^
Standardize[age]	121.341	0.00000
Standardize[age]*smoker	63.156	0.00000
children	24.201	0.00000
smoker*obesity_indicator	7.728	0.00000
obesity_indicator	7.493	0.00000 ^
sex	4.201	0.00006
smoker	4.154	0.00007 ^
Standardize[age]^2*smoker	3.083	0.00083
bmi	2.097	0.00800
bmi*smoker	1.192	0.06428
bmi*region	0.864	0.13678
Standardize[age]*smoker*region	0.785	0.16402
Standardize[age]^2	0.758	0.17439 ^
region	0.744	0.18022 ^
Standardize[age]*region	0.595	0.25431 ^
children*region	0.512	0.30738
obesity_indicator*region	0.317	0.48165
Standardize[age]^2*region	0.161	0.69006
smoker*obesity_indicator*region	0.104	0.78636
bmi*smoker*region	0.086	0.81998
smoker*region	0.062	0.86751 ^
Standardize[age]^2*smoker*region	0.017	0.96054
sex*region	0.012	0.97376

Figure D9

Model 3 Partial F-test



Custom Test

Parameter																																	
Intercept	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standardize[age]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
children	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
smoker[yes]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sex[male]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi*smoker[yes]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
obesity_indicator[1]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
smoker[yes]*obesity_indicator[1]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standardize[age]^2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standardize[age]*smoker[yes]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standardize[age]^2*smoker[yes]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standardize[age]*region[northwest]	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standardize[age]*region[southeast]	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Standardize[age]*region[southwest]	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
children*region[northwest]	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
children*region[southeast]	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
children*region[southwest]	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi*region[northwest]	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi*region[southeast]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi*region[southwest]	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
smoker[yes]*region[northwest]	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
smoker[yes]*region[southeast]	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
smoker[yes]*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sex[male]*region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sex[male]*region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sex[male]*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi*smoker[yes]*region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi*smoker[yes]*region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bmi*smoker[yes]*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
obesity_indicator[1]*region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
obesity_indicator[1]*region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
obesity_indicator[1]*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
smoker[yes]*obesity_indicator[1]*region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
smoker[yes]*obesity_indicator[1]*region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
smoker[yes]*obesity_indicator[1]*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Standardize[age]^2*region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Standardize[age]^2*region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Standardize[age]^2*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Standardize[age]*smoker[yes]*region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Standardize[age]*smoker[yes]*region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Standardize[age]*smoker[yes]*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Standardize[age]^2*smoker[yes]*region[northwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Standardize[age]^2*smoker[yes]*region[southeast]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Standardize[age]^2*smoker[yes]*region[southwest]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
=	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Value	-0.017484985	0.0321542883	0.0175045966	0.0242055976	-0.016367836	-0.016504975	-0.001336179	-0.011841138	0.005352654	0.0908766045	-0.139775705	0.0780618057	0.004428076	-0.008858227	0.0009166512	-0.003217504	0.0053365351	-0.001788272	-0.004265883	0.054752767	-0.031978885	0.0088206856	-0.025945042	0.0326993325	0.0278787068	-0.008652903	-0.02359005	0.0204688783	-0.02822739	-0.030468042	-0.009361313	0.0074214369	0.0083400556
Std Error	0.0232013434	0.020987939	0.0241317336	0.0161877733	0.0153673768	0.015174787	0.0078677366	0.0055983588	0.0071362703	0.2384501405	0.1754020081	0.2162591195	0.0181477248	0.017636731	0.0182992642	0.0078930963	0.0055982442	0.0071263289	0.0399041446	0.036218273	0.0397451564	0.0398750978	0.0360151036	0.0397073796	0.0271590656	0.0237312616	0.0264637644	0.0231433043	0.0210090824	0.024163621	0.0261801254	0.0231647598	0.0258514432
t Ratio	-0.753619516	1.5320364839	0.7253786738	1.4953012525	-1.065102786	-1.087657766	-0.16983017	-2.114731039	0.7500632397	0.3811136546	-0.79688771	0.3609642259	0.2440017171	-0.456900234	0.0501086727	-0.407638372	0.9532515855	-0.250938782	-0.106903257	1.5117442783	-0.804598285	0.2212078738	-0.720393374	0.8235076925	1.0264972711	-0.364620438	-0.891409475	0.8844406127	-1.343580325	-1.26090547	-0.357573287	0.320376172	0.3226147021
Prob>|t|	0.4512152463	0.1257586192	0.4683525427	0.1350804204	0.2870287326	0.2769489903	0.8651703276	0.0346442414	0.4533533913	0.703181672	0.4256628873	0.7181852908	0.8072682896	0.6478196814	0.9600435458	0.6836068274	0.340641195	0.8019014034	0.9148823519	0.1308438508	0.4211997891	0.8249655869	0.4714133405	0.4103715598	0.3048496875	0.7154545186	0.3728757678	0.3766232189	0.1793204209	0.2075709016	0.7207211648	0.7487350725	0.7470393371
SS	0.0799078685	0.3302344693	0.0740308152	0.3145875861	0.1598126207	0.1664442196	0.0040580171	0.6292083339	0.0791552909	0.0204358735	0.0893466853	0.0183321124	0.0083766419	0.0293715555	0.0003532726	0.0233794507	0.1278495664	0.0088597151	0.001607926	0.3215443286	0.0910840624	0.0068847052	0.0730169363	0.0954156288	0.1482517352	0.0187053657	0.1117991956	0.1100579807	0.2539869821	0.2236913778	0.0179893023	0.0144412439	0.0146437564

Sum of Squares	6.8356821972
Numerator DF	33
F Ratio	1.4722595227
Prob > F	0.0420042624