Virginia Journal of Science
Volume 67, Issue 3 & 4
Fall & Winter 2016
doi: 10.25778/SETP-H375

Note: This manuscript has been accepted for publication, and is online ahead of print. It will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form.

Exploring New Models for Seatbelt Use in Survey Data

Mark K. Ledbetter[1], Norou Diawara[2], Bryan E. Porter[3]

ABSTRACT

Problem: Several approaches to analyze seatbelt use have been proposed in the literature. Two methods that has not been explored are the use of unweighted and weighted logistic regression model and the use of item response theory (IRT) or the Rasch model. Since accurate methods to predict seatbelt use behavior based upon observed data must include a built-in design method and model, and overcome computation challenges, weighted and IRT method deem to be other options for an observational survey of seat belt use in the state of Virginia.

Method: The observed data from 136 sites within the Commonwealth of Virginia over two years was collected in a two stage systematic stratified proportional to size sampling plan. The data is analyzed using a weighted Rasch model.

Results: A relationship between seatbelt use of drivers weighted for county aggregate population size and length of the road segment observed and the factors of vehicle type and gender standardized using a standardized scale is confirmed using logistic regression model selection and AIC analysis. IRT model was considered and was found highly significant.

Practical Application: The addition of socio-economic measures, measure of road and driving difficulty, and data from other states may allow the prediction of seatbelt use with a in a new methodology: the models provide tools for policy decision-making.

Keywords: Seatbelt use; logistic regression, item response theory; structure design; sampling weight

[1]Mark Ledbetter, mledb001@odu.edu, Mathematics and Statistics Department, Old Dominion University, 4700 Elkhorn Ave., Norfolk, VA 23529
[2] Corresponding author, Norou Diawara, ndiawara@odu.edu, Mathematics and Statistics Department, Old Dominion University, 4700 Elkhorn Ave., Norfolk, VA 23529
[3] Bryan E. Porter, Department of Psychology, Old Dominion University, Norfolk, VA 23529

INTRODUCTION

The goal of this manuscript is to apply unweighted and weighted logistic regression models and the item response theory (IRT), specifically the Rasch model to analyze seatbelt use data for the years 2012, 2013, and 2014 from the Commonwealth of Virginia, under a sampling weight estimation methodological model.

The sampling methodology proposed has a two stage design associated with primary sampling unit (PSU) strata from 15 counties and secondary sampling units (SSU) from 136 road segments within the counties, under National Highway Transportation Safety Authority (NHTSA) guidelines. Because of that, the design is called higher order. If sampling weights are ignored, then the model parameter estimates can be biased (Lohr, 1999). In fact, since the sample is collected from a two stage stratified sampling design, standard underlying assumptions of parametric statistical models may be violated, and guidelines based on the statistical design cannot be ignored. Thomas and Heck (2001) provide guidelines for data analysis under weighted and designed data. Hahs-Vaughn (2005) provides examples of analysis weighting and design effects. Korn and Graubard (1995) provide regression models under weighted and unweighted schemes with incorrectly specified models and conclude that results of analyses can be misleading if the weighting is ignored in the large population of interest. Weighting allows us to measure the impact of factors when a simple random data sample is not collected.

Several studies have explored the factors affecting seatbelt usage. Preusser et al. (1991), Pickrell and Ye (2009), and Vivoda et al. (2004) concluded that females are more likely to wear seatbelts than males. The relationship between vehicle type and seatbelt use has been explored by Eby et al. (2002), Glassbrenner and Ye (2006), and Boyle and Vanderwolf (2004) who concluded that seatbelt use in pickup trucks is lower than other passenger vehicles. Nambisan and Vasudevan (2007) suggested that passenger and driver use are related. Shults and Beck (2012) assert that the seatbelt use is increased in those states within the United States that have primary seatbelt enforcement laws and actively enforce seatbelt use. Studies have also explored relationships between race, socio-economic status, age, rural/urban environments, law enforcement type (primary, secondary), the amount of fines, and the type of road traveled (primary, secondary, tertiary). Molnar, et al. (2012) employed a multivariate approach using the aforementioned factors along with cultural variables to explain the differences in seatbelt use between states using self-reported information, direct observation, and crash reports. However, Özkan, et al. (2012) demonstrated that the validity of self-reported seatbelt use in surveys is questionable compared to observed seatbelt usage. Our data which is from the direct observations of drivers eliminates self-reporting bias. Initially, we explore bivariate and also multivariate relationships between factors such as driver gender, vehicle type, traffic volume, road segment length, weather conditions, driver cellphone use, passenger presence, lane, and passenger seatbelt use.

Others authors (Hardouin and Mesbah, 2004, and Bartolucci, 2007) have proposed adding a score variable due to the measurement of concern. Those authors have incorporated latent traits of data in a score function. Models built under that structure fall in the class of Item Response Theory (IRT), and the Rasch model (Rasch, 1961) is a version of IRT. Suitable measurement methods and variables are used to reflect meaningful information that can be translated into quantitative measures. We consider that case of the Rasch model. However, because of the multidimensional nature of the data, and since the Rasch model is based on unidimensionality, a standard version of score function will be built.

Moreover, ignoring weights may lead to imperfection in the sample (as departing from the reference population) and serious bias in latent variable models (Kaplan and Ferguson, 1999). To avoid that problem, we apply a weight which is a function of the location strata size, and we account for potentially significant factors such as age, race, political leaning and religious affiliation which are not quantifiable using direct observations. Molnar et al. (2012) cautioned about the use of other factors to develop more effective countermeasures for increasing seatbelt use. We propose the weighted logistic model with IRT after variable selections and compare the finding with the logistic model under unweighted and weighted constraints. Analysis is performed on seatbelt use data in the Commonwealth of Virginia in the years 2012, 2013 and 2014. The manuscript is organized as follows. In Section 2, we present background on the selection probabilities at two different stages of sampling and present the data, then build the model in Section 3. In Section 4, the weighting scales are built into the models. The IRT is also reviewed. In Section 5, we use the available tools in SAS® software version 9.4 to perform the analyses under weighted and unweighted schemes and compare the results.

THE SAMPLING PLAN AND DATASET

In this paper, we analyze data collected in the summers of 2012, 2013, and 2014 for Virginia seat belt use. Sampling and data collection are in accordance with the final rule for 23 CFR Part 1340: Uniform Criteria for State Observational Surveys of Seat Belt Use. The rule is published in the Uniform Criteria for State Observational Surveys of Seat Belt Use (2011). The sampling design is a two-stage stratified systematic probability proportional to size (PPS) sampling plan. Primary sampling units (PSU) are county aggregates and were stratified using the five-year average annual VMT (vehicle miles traveled) in millions. Out of 97 total county aggregates, 57 account for 87.2 percent of passenger vehicle crash related fatalities. The 57 eligible county aggregates were grouped by VMT into three strata: low, medium, and high. Within each stratum, five PSU's were selected with PPS where the measure of size (MOS) was the five-year average annual VMT. The PSU sampling weights are calculated by taking the inverse of the five year average annual VMT, and varied from approximately 0.089 to approximately 0.967. Secondary sampling units (SSU) are road segments. Road segments were stratified by type (primary, secondary, and local) and by segment length (short, medium and long) within each county. The eligible SSU were then selected by PPS with segment length as the MOS resulting in 136 selected road sites for observation. The SSU weights are calculated by taking the inverse of the segment length and varied from approximately 0.0001 to approximately 0.1657.

The weighting was added so that information from the whole population would be captured. If the selection mechanism is not informative, the parameter estimates will remain consistent regardless of the weights, and weights should be excluded from the model (Asparouhov et al. 2004). Moreover, if the strata sample sizes are large enough, the parameter estimates are unbiased. In sampling surveys, it is not always possible to determine whether the weights are informative. However, the observations should reflect the sampling weights to avoid biased sampling.

Trained data collectors recorded observations at their randomly assigned collection sites (SSU). Trained quality control (QC) monitors performed random and unannounced visits to the collection sites. More than 10% of all sites were visited by QC Monitors. Observations were performed on all days of the week between the hours of 7:00 am and 5:30 pm. Data collectors were assigned 90 minute intervals during which 50 minutes of active observation, and recording occurred. If inclement weather (rain which damages recording media) prevented collection of at least 25 minutes of data, then the site was rescheduled for observation for that interval.

The data collected includes the following observed binary data: driver seat belt use (yes, no), driver gender (female, male), passenger present (yes, no), passenger seatbelt use (yes, no), and visible driver cellphone use (yes, no). The other observed data is categorical: vehicle type (car, truck, SUV, van, or minivan), lane of the road (1-5, where lane 1 represents the lane furthest to the right and lane 5 denotes the fifth lane from the right in the direction of travel), and weather (sunny/clear, light rain, cloudy, fog, or clear but wet conditions). The VMT for each site observed is classified (Road Class) within each county aggregate as lower, average, and upper. Vehicle type was assigned in no particular order, and later we reclassified it to describe the size of the vehicle which crudely correlates to seatbelt use. Weather is also not ordered in its assignment, and we reclassify it based on severity and impediment of driving ability. The data set also includes the following continuous variables: VMT, road segment length, and selection probabilities determined in the sampling design stage.

Table 1 gives the summary measures of the variables and their basic proportions for the two years. The data shows a slight decrease in driver seat belt use and a slight increase in visible driver cell phone use. The number of observed drivers increased by 5.8% in 2013 compared to 2012. The number of drivers observed in 2014 decreased by 1.2% compared with 2012. When the seat belt information for driver or passenger is unknown, the data will be dropped. The weight for each driver is obtained by taking the product of the PSU and SSU weights. All calculations are performed using SAS® software version 9.4.

TABLE 1: Summary of 2012, 2013, and 2014 Virginia Seatbelt Data

Variables	Levels	2012 Sample	2013 Sample	2014 Sample	2012 Percent	2013 Percent	2014 Percent
Driver Seatbelt Use	Yes	10956	11396	10072	80.37	79.01	74.83
	No	2230	2581		16.36	17.89	19.61
	Unknown	446	447	749	3.27	3.10	5.56
Passenger Seatbelt Use	Yes	2817	2937	2669	73.86	72.88	67.98
	No	675	724	785	17.70	17.97	19.99
	Unknown	322	369	472	8.44	9.16	12.02
Passenger Present	Yes	9783	10145	9467	71.94	70.32	70.69
	No	3815	4282	3926	28.06	29.68	29.31
Driver Gender	Male	7990	8511	8022	60.47	60.28	61.32
	Female	5223	5607	5060	39.53	39.71	38.68
	Unknown	1	1	-	0.01	0.01	-
Driver Cellphone Use	Yes	782	840	712	5.80	6.4	5.32
	No	12703	12277	12671	94.20	93.60	94.68
Vehicle Type	Car	6800	7057	6471	49.92	48.95	48.08
	Truck	2401	2739	2490	17.63	19.00	18.50
	SUV	3130	3309	3370	22.98	22.95	25.04
	Van	342	354	305	2.51	2.46	2.27
	Minivan	948	958	823	6.96	6.64	6.11
Road Class (Vehicle Miles Traveled, VMT)	Lower	5066	5246	4758	37.23	36.36	35.34
	Average	4170	4523	4387	30.65	31.35	32.59
	Higher	4371	4658	4317	32.12	32.29	32.07
Lane Observed	1	10090	10225	9586	74.03	70.89	71.27
	2	3207	3820	3470	23.53	26.48	25.80
	3	282	270	279	2.07	1.87	2.07
	4	35	84	100	0.26	0.58	0.74
	5	16	25	16	0.12	0.17	0.12
Weather	Clear/Sunny	10999	9703	10302	80.83	73.31	76.54
	Light Rain	235	331	34	1.73	2.50	0.25
	Cloudy	2353	2902	3089	17.29	21.93	22.95
	Fog	20	-	20	0.15	-	0.15
	Clear/Wet	-	299	15	-	2.26	0.11
Road Segment Length	< 0.30	11805	12690	11826	86.60	87.96	87.85
	0.31 - 0.81	1827	1737	1636	13.40	12.04	12.15
Total	-	13632	14427	13462	-	-	-

STATISTICAL MODELS

Generalized linear models were considered in the investigation of the data. First, a classic linear model was suggested to obtain a general relationship between the response (driver seatbelt use) and predictive variables. However, use of a linear model on binary responses is not recommended (Kleinbaum, et al.) since predicted values may be outside of the domain of the response variable. From this point forward, a classic model also known as classical test theory (CTT) is considered. We consider fitting a logistic model to the data.

Models are fit using stepwise selection in PROC Logistic. Once the model is selected, the predictors that were not selected are removed and PROC Logistic is performed on the selected predictors without a selection process since they are known to be significant; results provided in this manuscript are from this final analysis.

Initial Proposed Model

In this model, p = P(Y=1) is the probability that the driver is wearing a seat belt, and 1 - p = P(Y=0) is the probability that the driver is not wearing a seatbelt. The initial model is:

$$\text{Log}\left[\frac{p}{1-p}\right] = \beta_0 + \beta_v X_v + \beta_r X_r + \beta_g X_g + \beta_s X_s + \beta_l X_l + \beta_c X_c + \beta_w X_w + \beta_{pp} X_{pp} + \beta_{ps} X_{ps}$$

where β_0 denotes the intercept of the model, X_v denotes Vehicle Type (car, truck, SUV, van, or mini-van), X_r denotes Road Classification for VMT (low, average, high), X_g denotes Driver Gender (male/female), X_l denotes Lane in which vehicle observed (right to left), X_c denotes Driver Cell Phone Use (yes/no), X_w denotes Weather (clear, light rain, cloudy, foggy, or clear but wet), X_{pp} denotes Passenger Present (yes/no), X_{ps} denotes Passenger Seatbelt Use (yes/no). This notation is used consistently throughout this manuscript. The weights w_{ij} are obtained as $p_{ij} = p_i * p_{j(i)}$ where p_i is the selected probability of the selected county, and $p_{j(i)}$ is the selection probability of the jth road type selected within the ith county; $i = 1, 2 \ldots, 15$, and $j = 1, 2, \ldots, n_i$.

The estimated non-weighted seat belt use for each year is $\hat{p} = 0.8037\ for\ 2012, \hat{p} = 0.7901\ for\ 2013,\ and\ \hat{p} = 0.7483\ for\ 2014$. Table 2 below gives the estimates of the proportion of seat belt use by county aggregates.

TABLE 2: Seat Belt Use by County Aggregate for the Years 2012, 2013, and 2014

County Aggregate	2012 Unweighted Seatbelt Use	2012 Weighted Seatbelt Use	2013 Unweighted Seatbelt Use	2013 Weighted Seatbelt Use	2014 Unweighted Seatbelt Use	2014 Weighted Seatbelt Use
Alleghany	0.7106	0.7350	0.7101	0.7784	0.6594	0.7214
Carroll	0.7301	0.7890	0.8472	0.8477	0.6667	0.7225
Fairfax	0.9006	0.9032	0.8999	0.8974	0.8876	0.8910
Halifax	0.8297	0.8704	0.7813	0.8064	0.7338	0.7657
Henry	0.8531	0.8816	0.7260	0.7958	0.6729	0.7133
Loudoun	0.8012	0.7867	0.8548	0.8532	0.8595	0.8801
Mecklenburg	0.7273	0.7875	0.7376	0.7696	0.7013	0.7519
Prince George	0.8240	0.8355	0.7759	0.7796	0.7567	0.7582
Rockbridge	0.8128	0.7331	0.7647	0.7617	0.7627	0.7098
Shenandoah	0.7444	0.7402	0.7422	0.7616	0.7031	0.7056
Southampton	0.8482	0.8871	0.8258	0.8455	0.8189	0.7977
Southeast	0.8741	0.8781	0.8232	0.8400	0.8063	0.8131
Stafford	0.8399	0.8398	0.8599	0.8420	0.8714	0.8746
Tazewell	0.7394	0.7492	0.7414	0.7438	0.6419	0.5978
Washington	0.7569	0.8380	0.7773	0.7605	0.8115	0.8042

Model Fitting

The model is fit using the logistic procedure in SAS® 9.4 with stepwise selection at a p = 0.15 significance level for both entry into the model and retention in the model. The results are verified using forward selection with p=0.15 for entry into model and backward selection with p=0.15 for retention in the model. The three procedures produce the same results.

Analysis of the effects of weather on seatbelt use revealed inconsistent associations between seatbelt use and weather severity for the three years. Further, the selection process does not identify weather as significant for any combined data. Hence, weather has been removed from the model and the analysis repeated. Analysis of the predictor variables reveals a high correlation (Spearman's correlation coefficient, $r_s = 0.94, p - value < 0.0001$) between road segment length and road class which indicates a confounding condition. Other correlations are less than 0.15 and do not indicate the presence of other confounding effects. As a result, road segment length was removed from the model and the analysis performed again.

Table 3 provides the Wald Test for significance in the selected Model with variables as Vehicle type, Road class, driver gender, and so on. For 2012, all remaining predictors are significant at p=0.01, while passenger presence is removed due to a p-value > 0.15. For 2013, all predictors are significant at p=0.01. For the combined 2012 and 2013 data, five of the six remaining predictor variables have p-values <0.0001, however passenger presence is only significant at p=0.10. For the combined data for 2012, 2013, and 2014, all six of the remaining predictors are significant at p=0.05.

TABLE 3: Type 3 Analysis of Effects										
		2012		2013		Combined 2012 and 2013		Combined 2012, 2013, and 2014		
Effect	**DF**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**	
Vehicle Type	4	272.927	<.0001	238.353	<.0001	513.796	<.0001	773.573	<.0001	
Road Classification	2	58.9336	<.0001	14.0610	.0009	62.3872	<.0001	63.9249	<.0001	
Driver Gender	1	22.9354	<.0001	29.4874	<.0001	51.2621	<.0001	58.2420	<.0001	
Lane	4	24.0216	<.0001	32.8055	<.0001	52.3698	<.0001	57.5625	<.0001	
Driver Cell Phone Use	1	14.4898	0.0001	9.7584	.0018	25.6445	<.0001	49.5231	<.0001	
Passenger Present	1	-	-	12.2656	.0005	2.8089	.0937	5.3603	.0206	

The close agreement between the models may indicate that the aggregate data follows a standard model which also fits the individual data sets. The test of the global hypothesis of null model, shown in Table 4, of $\beta_i = \beta_j = 0 \; for \; i \neq j$ versus at least one $\beta_i \neq 0$ $(i, j = r, g, l, c, r, or\ pp$ depending upon the model) indicates significant evidence exists ($p < 0.0001$) to support the claim that the models are not explained solely by the intercept (i.e. the response is not a constant) for 2012, 2013, the combined 2012/2013 data, and for the combined 2012/2013/2014 data which is consistent with the Wald Test results in Table 3.

TABLE 4: Testing Global Null Hypothesis: β=0								
	2012		2013		Combined 2012 and 2013		Combined 2012, 2013, and 2014	
Test	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**
Likelihood Ratio	485.5235	<.0001	452.3717	<.0001	917.5154	<.0001	1299.7313	<.0001
Score	521.6322	<.0001	470.3904	<.0001	972.3826	<.0001	1377.5713	<.0001
Wald	488.6745	<.0001	446.5182	<.0001	918.1369	<.0001	1305.6547	<.0001
DF	12		13		13		13	

Computational efficiency is measured by Akaike Information Criterion (AIC) numbers, displayed in Table 5, which assess the goodness of fit of the model: smaller numbers indicate a better fit. AIC is defined as follows:

$$\mathrm{AIC} = 2\mathrm{p} + \mathrm{n}\log\left(\frac{\mathrm{SS}_r}{n}\right)$$

where p is the number of parameters in the model, SS_r is the residual sum of squares, and N is the number of observations in the dataset.

The results of the AIC for logistic regression performed on the significant variables identified during the selection process are in the 10 thousands. Since the intercept alone is not a sufficient explanation of the model, we use the values for intercept and covariance. The AIC numbers obtained for individual years are approximately 30% lower than those obtained by

Molnar (2012); however, the combined data is significantly higher. The significantly higher numbers for the combined data indicate a significant amount of variation in the model, or a less than optimum fit.

TABLE 5: Model Fit Statistics				
	2012	**2013**	**Combined 2012 and 2013**	**Combine 2012, 2013, and 2014**
Criterion	**Intercept and Covariates**	**Intercept and Covariates**	**Intercept and Covariates**	**Intercept and Covariates**
AIC	11159.889	11856.530	23015.856	35333.162
SC	11256.770	11960.618	23129.764	35452.647
-2 Log L	11133.889	11828.530	22987.856	35305.162

Selected Model

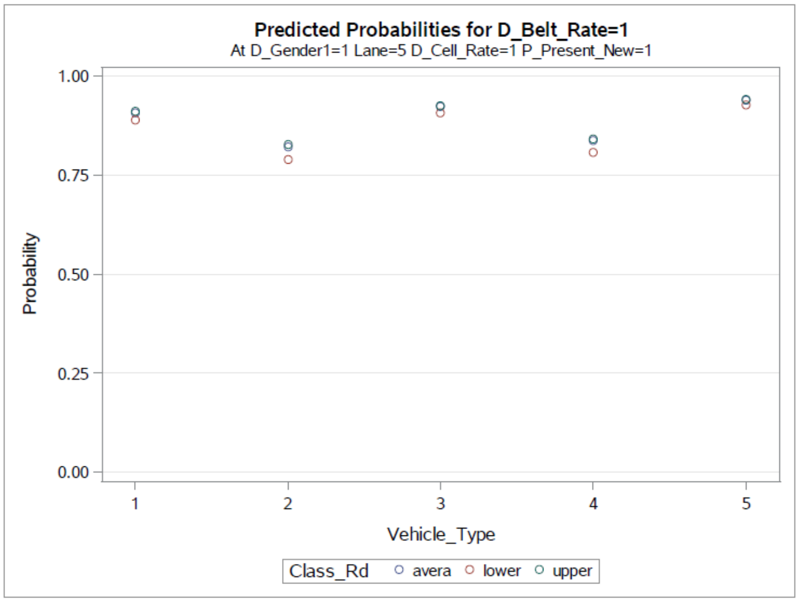
Since the models for 2012, 2013, and the combined data are very similar, and the combined data results in all predictors with $p<0.05$, we select the following model at this step of the analysis:

$$\text{Log}\left[\frac{p}{1-p}\right] = \beta_0 + \beta_v X_v + \beta_r X_r + \beta_g X_g + \beta_l X_l + \beta_c X_c + \beta_{pp} X_{pp}$$

Again, the test of the global hypothesis that the model is expressed by the intercept alone is not accepted with $p < 0.0001$.

Figure 1 displays the predicted probability of seat belt use (for female drivers in lane 5 using a cellphone with a passenger present) versus the vehicle type for each road class (VMT). The oscillating trend indicates that the model is not designed to predict an increase in seatbelt use for increasing values of the predictors in this subset of predictor variable levels. (Please note that the authors have only included one chart for this model due to the excessive space required to depict all 40 such combinations.) Therefore, we investigate other methods to improve the fit of the model.

FIGURE 1. Model 1: Multivariate Logistic Regression on Combined 2012-2014 Raw Data



The AIC model criterion is used to compare the models of the years 2012, 2013, and the combined years. As we see there is no evidence that the model based on the combined years is better than the individual year models. As discussed by Rosenthal (1978) there could be issues with the validity of the data collected in our sampling as well due to observer error of approximately 1%. Moreover, there could be carryover effects in the years, and as such, data from 2012 should be taken as baseline for 2013 and so on. It is satisfactory to see that variables selected in 2012 are also selected for 2013 and the combinations of the years, but then the question is how is the relative efficiency of the combined data models compared to the 2013 model and to the 2012 model. To partially answer the questions, we propose to review the variables at hand in the next section.

Variable Standardization and Reclassification

The goal of this manuscript is to develop a prediction model using the Rasch / IRT model. Such an IRT technique is built to control the source of sampling error using an unknown latent variable from the selected variables which we will call the score. The latent variables are described next. Since vehicle types are listed in no particular order, vehicle type is reclassified to indicate size of the vehicle which negatively correlates to driver seatbelt use: i.e. in general, the drivers of larger vehicles tend to wear seatbelts less often than drivers of smaller vehicles as suggested in Eby et al. (2002). Preliminary analysis of the data appears to support this hypothesis, so smaller vehicle types are given a larger value to indicate that the driver is more likely to wear a seatbelt. Table 6 below contains the reclassifications of vehicle type. The remaining five predictor variables have positive correlations to driver seatbelt use and reclassification is not necessary.

It is known that the variance is larger for population parameters where the measurement of interest is a large value than in population parameters where the measurements are small values. In order to make the variance between variables more homogenous and reduce the overall model variance, each variable of interest was standardized by dividing its value by its third quartile (Q3) in an approach similar to Illi et al. (2012) in which the weighted variables were standardized by their quartiles. Standardizing the variables may affect whether they are selected in the model, so all six of the predictors are standardized. The Q3 values of the variables after reclassification are listed in Table 7 below.

TABLE 6: Reclassification of Variables

Vehicle Type	Original Value	New Value for Size
Car	1	3
Truck	2	1
SUV	3	1
Van	4	1
Mini-Van	5	2

TABLE 7: Third Quartiles After Reclassification (No Weight)

Variable	2012: 75th Percentile (Q3)	2013: 75th Percentile (Q3)	2012-2013: 75th Percentile (Q3)	2012-2014: 75th Percentile (Q3)
Vehicle Type	3	3	3	3
Gender	1	1	1	1
Lane	2	2	2	2
Road Class	3	3	3	3
Cell Phone	1	1	1	1
Passenger Present	1	1	1	1

Model Fitting: Standardized and Reclassified Variables

The logistic selection process with $p = 0.15$ for entry and retention in the model is performed on the reclassified and standardized variables. The significant variables indicated prior to standardization in 3.2 above remain significant (Table 8). The model fit statistics are comparable to the previous analysis (Table 9). The global null hypothesis test indicates that the model is not sufficiently described solely by the intercept (Table 10). All variables selected are significant ($p < 0.01$) for all datasets analyzed. In this analysis, it is reasonable to select the model fit by the combined 2012, 2013, and 2014 data:

$$\text{Log}\left[\frac{p}{1-p}\right] = \beta_0 + \beta_v X_v + \beta_r X_r + \beta_g X_g + \beta_l X_l + \beta_c X_c + \beta_{pp} X_{pp}.$$

Figure 2 displays the predicted probability of seat belt use for female drivers in lane 5 using a cellphone with a passenger present) versus the vehicle type for each road class (VMT).

The general upward trend indicates that the reclassification has resulted in a model that is showing an increased probability of seatbelt use based on road classes with higher VMT for each vehicle size. Please note that the authors have only included one chart for this model due to the excessive space required to depict all 40 such combinations.

TABLE 8: Type 3 Analysis of Effects for Standardized and Reclassified Variables

		2012		2013		Combined 2012 and 2013		Combined 2012, 2013, and 2014	
Effect	**DF**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**
Vehicle Type	2	86.0490	<.0001	72.5569	<.0001	158.944	<.0001	198.594	<.0001
Road Classification	2	56.2572	<.0001	16.5001	.0003	63.0613	<.0001	62.6709	<.0001
Driver Gender	1	78.1552	<.0001	89.9920	<.0001	167.328	<.0001	227.771	<.0001
Lane	4	32.1929	<.0001	39.7865	<.0001	67.3511	<.0001	76.9267	<.0001
Driver Cell Phone Use	1	15.9983	<.0001	9.5138	.0020	25.9062	<.0001	48.904	<.0001
Passenger Present	1	-	-	16.8325	<.0001	7.5306	.0061	14.047	.0002

TABLE 9: Model Fit Statistics for Standardized and Reclassified Variables

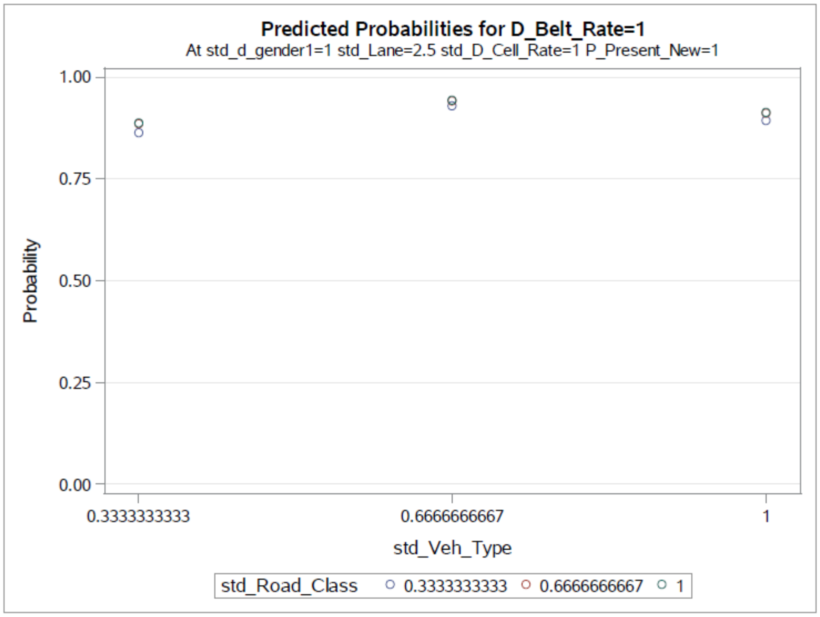
	2012		2013		Combined 2012 and 2013		Combined 2012, 2013, and 2014	
Test	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**
Likelihood Ratio	307.1437	<.0001	291.3205	<.0001	575.7315	<.0001	741.4629	<.0001
Score	300.7339	<.0001	281.3758	<.0001	560.1672	<.0001	722.7031	<.0001
Wald	291.7697	<.0001	273.5035	<.0001	544.4533	<.0001	704.7331	<.0001
DF	10		11		11		11	

TABLE 10: Global Null Hypothesis: **β**=0 for Standardized and Reclassified Variables

	2012	2013	Combined 2012 and 2013	Combine 2012, 2013, and 2014
Criterion	**Intercept and Covariates**	**Intercept and Covariates**	**Intercept and Covariates**	**Intercept and Covariates**
AIC	11334.269	12013.581	23353.640	35887.431
SC	11416.246	12102.799	23451.276	35989.846
-2 Log L	11312.269	11989.581	23329.640	35863.431

The AIC and SC numbers remain undesirably large and indicate that reclassification and standardization are not sufficient actions to improve model fit. Therefore it is reasonable to investigate the cause for the poor model fit.

FIGURE 2. Model 2: Multivariate Logistic Regression on Reclassified and Standardized Variables (2012-2014 Data)



In all the previous sections, the AIC, BIC and log likelihood have been used as best measures of goodness fit for the most parsimonious models. They turn out to be high, which is an evidence of over-dispersion, which could be an indication there is more variability in the data than expected from the fitted model, which is an indication of a poor fit. Over-dispersion is quite evident for count data under the Poisson regression model. Since the sample size is large, the corrected AIC would not lead us to better improvements. Variables have been selected for each dataset and the selection process results in about the same model. We will use these criteria as comparisons when adding the weights to the models we will consider in the next section.

WEIGHTED STATISTICAL MODELS

Weights

In all of the above analyses, the weights associated with the data were ignored. However, driver seat belt behavior is intricate and quite certainly involves non-collected data. Ignoring sample weights leads to inflated standard errors and biased estimates (Lohr, 1999). Thomas and Heck (2001) provide guidelines for data analysis under weighted and designed data which reduces bias that would result in over sampled strata. The weights are under stratum size and length of road segments. The inclusion of weights results in a significantly different model than selected in section 3 above as inferred by Korn and Graubard (1995). Additionally, the goodness of fit criteria is significantly reduced. The sampling plan for the data in this manuscript was developed as a joint effort between two of the authors (N. Diawara and B.E. Porter) and NHTSA. Therefore, in order to correct for bias due to stratum size and length of road segment, we included the weight designed for this analysis in our model, in accordance with NHTSA requirements as:

$$Weight = (Road\ Segment\ Length)\ x\ (County\ Selection\ Probability).$$

In this section, we will compare the results of the analysis based on the sampling weights and validate the appropriateness of the uses of the weights.

WEIGHTED LOGISTIC MODELS

Model Fitting: Weighted Logistic Regression

Prior to performing analysis on the reclassified and standardized variables, the 75th percentiles for the weighted reclassified variables is determined, see Table 11, and the predictors are standardized using the weighted 3rd quartiles.

TABLE 11: Weighted Third Quartiles for Reclassified Variables

Variable	**2012: 75th Percentile (Q3)**	**2013: 75th Percentile (Q3)**	**2012-2013: 75th Percentile (Q3)**	**2012-2014: 75th Percentile (Q3)**
Vehicle Type	3	3	3	3
Gender	1	1	1	1
Lane	2	2	2	2
Road Class	3	3	3	3
Cell Phone	1	1	1	1
Passenger Present	1	1	1	1

The selection process using the weighted logistic regression model and the surveylogistic procedure resulted in two significant predictors at p=0.15: driver gender and vehicle type for 2012.The selection process for both the 2013 data and the combined 2012-2013 data also indicates that passenger presence is significant at p=0.15. In addition, the combined data for 2012, 2013, and 2014 indicates the significance of passenger presence (Table 12). In the aggregate data for 2012-2014, the selection process results in four significant variables at p=0.15. The model is significant as indicated by the global null hypothesis test in Table 13.

TABLE 12: Type 3 Analysis of Effects for Weighted, Standardized and Reclassified Variables										
		2012		**2013**		**Combined 2012 and 2013**		**Combined 2012, 2013, and 2014**	
Effect	**DF**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**	**Wald ChiSq**	**Pr > ChiSq**
Vehicle Type	2	4.0806	0.1300	5.2866	.0711	9.3692	.0092	11.2742	.0036
Driver Gender	1	4.2080	0.0402	6.0867	.0136	10.3672	.0013	12.5182	.0004
Driver Cell Phone Use	1	-	-	-	-	-	-	3.1076	.0779
Passenger Present	1	-	-	2.6228	.1053	2.1891	.1390	2.9222	.0874

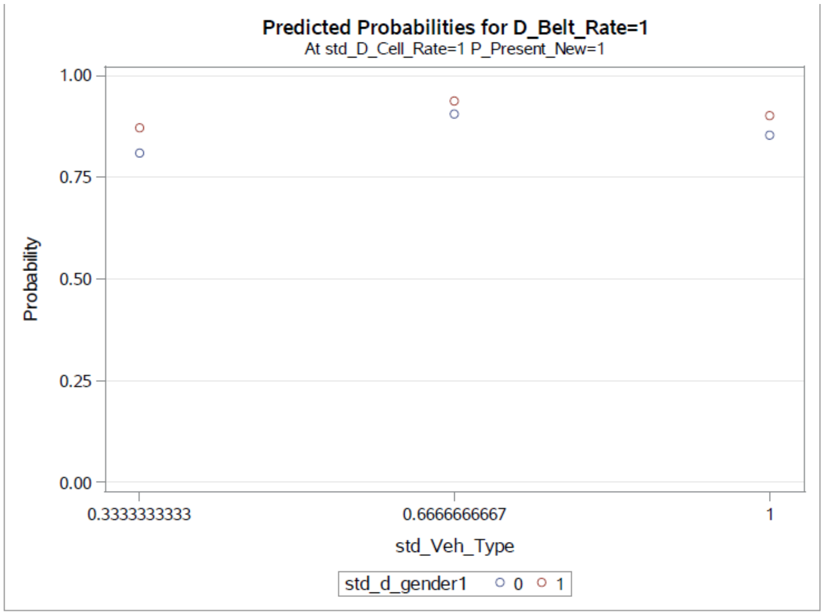
TABLE 13: Global Null Hypothesis: **β**=0 for Weighted, Standardized, and Reclassified Variables								
	2012		**2013**		**Combined 2012 and 2013**		**Combined 2012, 2013, and 2014**	
Test	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**	**Chi-Square**	**Pr > ChiSq**
Likelihood Ratio	10.3152	.0161	16.3853	.0025	26.3513	<.0001	35.1654	<.0001
Score	9.9895	.0187	15.6510	.0035	25.3421	<.0001	34.1321	<.0001
Wald	9.7190	.0211	15.1265	.0044	25.5806	<.0001	33.1892	<.0001
DF	3		4		4		5	

There is great gain in the AIC when the weights are added to the model, matching comments of Richards et al. (2011) that in the context of behavioral ecology a simple controlled model does not show all the complexity of the data. Table 14 contains the AIC and SC values, which are lower than the corresponding unweighted models by a factor of approximately 20.

TABLE 14: Model Fit Statistics for Weighted, Standardized and Reclassified Variables				
	2012	**2013**	**Combined 2012 and 2013**	**Combine 2012, 2013, and 2014**
Criterion	**Intercept and Covariates**	**Intercept and Covariates**	**Intercept and Covariates**	**Intercept and Covariates**
AIC	548.455	661.001	1201.336	1812.153
SC	578.296	698.644	1242.273	1863.365
-2 Log L	540.455	651.001	1191.336	1800.153

Figure 3 displays the predicted probability of seat belt use (for drivers using a cellphone with a passenger present) versus the vehicle type for each gender. The same general upward trend exists in the weighted model as that shown in the unweighted model but using less predictors. Please note that the authors have only included one chart for this model due to the excessive space required to depict all 24 such combinations.

FIGURE 3. Model 3: Multivariate Weighted Logistic Regression on Model with p=0.15 Selection (2012-2014 Data)



MODEL SELECTION: WEIGHTED LOGISTIC REGRESSION

The final model selected for the 2012-2014 aggregate data is

$$\text{Log}\left[\frac{p}{1-p}\right] = \beta_0 + \beta_v X_v + \beta_g X_g + \beta_c X_c \qquad + \beta_{pp} X_{pp}$$

where β_0, β_v, β_g, β_c, and β_{pp} are the estimates calculated using the weights.

As expected, the combination of the data results in an improvement in the significance of the predictors: vehicle type (0.0036), gender (0.0004), and passenger presence (0.0874). However, the models have different selected variables and two of the variables selected for the 2012-2014 combined data have p-values > 0.05 indicating the necessity for a different analytical method.

One suggestion is to develop an IRT model for prediction of seatbelt use, and it is advisable to include only very significant predictor variables. We explore a reduced model using a selection process with p=0.05 significance on the combined data. Vehicle type and gender are very significant predictors while passenger presence was excluded at p=0.05 which provides a consistent model with the preliminary analysis of 2012 and 2013 data individually. However, cellphone use is marginally significant at p=0.0439 suggesting that the combination of data from 2014 may have exposed an inherent association between cellphone and seatbelt use. Tables 18 through 20 contain the results of the selection at p=0.05 for the 2012-2014 data. The AIC and SC values are comparable to those of the four variable model. The final weighted logistic model using reclassified and standardized predictor variables with p=0.05 selection criteria is

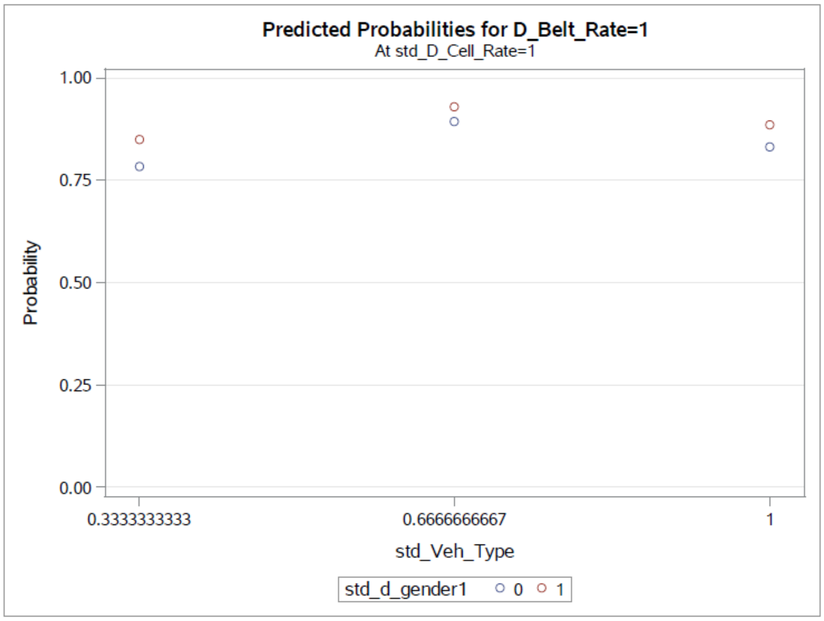
$$\text{Log}\left[\frac{p}{1-p}\right] = \beta_0 + \beta_v X_v + \beta_g X_g + \beta_c X_c.$$

TABLE 18: Type 3 Analysis of Effects			
Combined 2012, 2013, and 2014: Reduced Model			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Vehicle Type	2	11.7957	.0027
Driver Gender	1	11.8360	.0006
Driver Cell Phone Use	1	4.0604	.0439

TABLE 19: Testing Global Null Hypothesis: BETA=0			
Combined 2012, 2013, and 2014: Reduced Model			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	32.4833	4	<.0001
Score	31.5075	4	<.0001
Wald	30.6313	4	<.0001

TABLE 20: Model Fit Statistics	
Combined 2012, 2013, and 2014: Reduced	
Criterion	**Intercept and Covariates**
AIC	1817.440
SC	1860.122
-2 Log L	1807.440

FIGURE 4. Model 4: Weighted Multivariate Logistic Regression on Combined 2012-2014 Data with p=0.05



WEIGHTED ITEM RESPONSE THEORY MODEL

Background

To analyze dichotomous events or polytomous level response data (as usually found in the quality of life field), the item response theory (IRT) model provides a complement to the classical test theory (CTT) as the behavior and characteristic of the driver is not directly understandable. The measurement of driver behavior is not suitable since it is based on qualitative indicators such as the type of vehicle used, and other ad hoc parameters that are not easy to translate into quantitative information to be used in a CTT statistical analysis. Because of that, IRT and its famous Rasch model have also been implemented to measure drivers' behaviors. The IRT model allows the inclusion of the latent factor common to all drivers that can be described by a score function. We applied such a model based on specified traits that reflect the dichotomy of the data such as gender, and made comparisons. We then compare the efficiency and effectiveness of the overall indicators by computing goodness of fit statistics.

Model

Because the model requires consideration of several conditions, the Rasch model is considered, as it provides a tool to analyze characteristics even when they are latent. Such a model can be included in the class of Item Response Theory in the framework proposed by Bartolucci (2007). Driving habits can be seen as a variable which depends on many factors. Our primary focus is on seat belt use and indicators which give additional information to evaluate seat belt use. We propose to extend the theory of logistic regression to include characteristics associated with driver seatbelt use which is translated into the driver's condition as an associated score. In such a context, the Rasch model (Rasch 1961, Samejima 1996) is an option where we can include each driver's behavior regarding seat belt use. One main concern is the associated measurement of the score. That score is based on the qualitative information to be translated into quantitative measure. Using ideas from Mesbah (2010), we develop a score function that can be used to build the sensitive attributes and behaviors of drivers. As mentioned in Beaumont et al. (2014), the bias reduction is achieved through appropriate weight adjustements.

A score function is built using a linear combination of significant predictor variables. The proposed score attempts to capture the features of vehicle type driven, driver gender, and driver cellphone use. Those features can alter the probability of seat belt use and they can be seen as sufficient statistics for the response (See Hardouin and Mesbah 2004). In our case, due to the logistic analysis on driver seat belt use, we propose to use a score function composed of driver gender, vehicle type, and handheld cellphone use as follows:

$$S = X_g + X_v + X_c$$

where X_g = driver gender (male = 0 and female = 1), X_v = size of vehicle driven standardized by the 3rd quartile (1/3 = SUV/Van/Truck, 2/3 = Minivan, and 1 = car), and X_c = driver cellphone use (no = 0 and yes = 1).

The final model is

$$\text{Log}\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 S\ .$$

RESULTS

The logistic regression analysis yields parameter estimates (standard error) $\widehat{\beta_0} = 0.6891$ (0.1855) and $\widehat{\beta_1} = 0.4691$ (0.0930) for the 2012-2014 combined data.

The AIC values are comparable to the AIC values in the traditional logistic analysis shown in 4.1.2 above indicating a satisfactory fit of the model. Figure 5 shows the regression line and 95% confidence limits for predicted probability of seatbelt use versus the weighted score function. The narrow confidence band and the linear upward trend also indicate a satisfactory fit of the model to the data. All such results conform with the findings by Beaumont et al. (2014) in the bias reductions even in the nonresponse situation, and provide an improvement on their suggested approach.

TABLE 21: Analysis of Maximum Likelihood Estimates

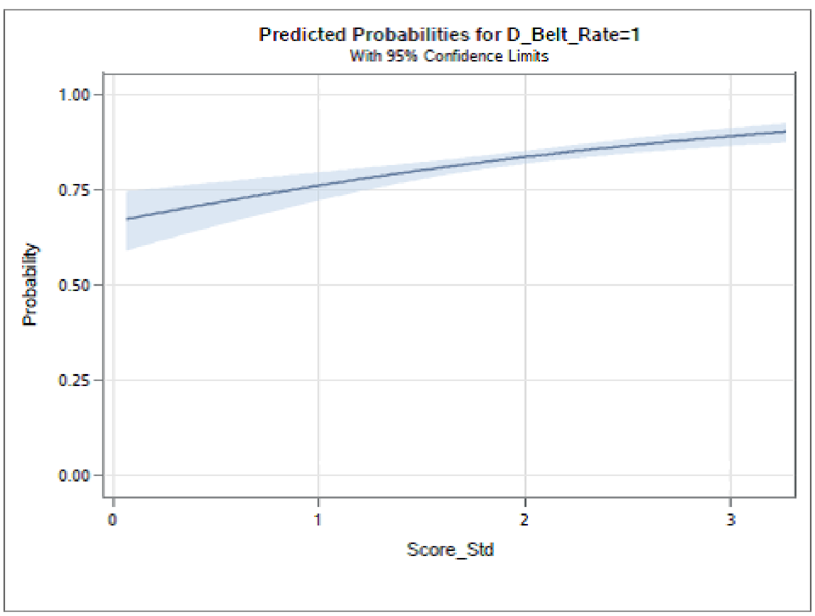
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq
Intercept	1	0.6891	0.1855	13.7984	0.0002
Score	1	0.4691	0.0930	25.4537	<0.0001

TABLE 22: Model Fit Statistics	
Combined 2012, 2013, and 2014	
Criterion	**Intercept and Covariates**
AIC	1817.789
SC	1834.862
-2 Log L	1813.789

TABLE 23: Testing Global Null Hypothesis: BETA=0			
Combined 2012, 2013, and 2014			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	26.1341	1	<0.0001
Score	25.8679	1	<0.0001
Wald	25.4537	1	<0.0001

TABLE 24: Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Score_Std_Reduced	1.599	1.332	1.918

FIGURE 5. Logistic Regression of Seatbelt Use versus Weighted Score



The present IRT model offers many more advantages than the classical test theory (CTT) methods developed in Section 3. The model is parsimonious and allows driver seat belt behavior to be easily estimated from scaled psychometric item measures under a weighted design model.

CONCLUSION

Driver seatbelt use in the Commonwealth of Virginia may be satisfactory described using driver gender, vehicle type, and cellphone use in a multivariate logistic model using weights designed specifically for the dataset. However, prediction of seatbelt behavior is more appropriate using item response theory. As such, we have endeavored to build a score function considering driver gender, vehicle type driven, and cellphone usage by applying the Rasch model in logistic regression and through the application of weights within the model. The resulting models are greatly affected by the application of weights. Fitting a weighted model results in significant improvements in goodness of fit statistics, such as AIC numbers, by a factor of 18 to 20.

We suggest that a sound approach may be developed using a weighted IRT model which should also potentially include social factors such as education level, political affiliation, religion, education level, original background, race, socio-economic status, fine level, and current law enforcement policies. Such a model could be used to develop programs and interventions that can increase seatbelt use and save lives.

More research is needed in which the sample size and the time for data collection allocated to each site (based on the 50 minutes time frame) are not fixed in advance but depend on the outcomes of the successive observations being taken. Results from such sequential analysis in the discrete case will allow us to save resources.

REFERENCES

1. Uniform Criteria for State Observational Surveys of Seat Belt Use, 23 CFR Part 1340 (2011).

2. Asparouhov, T. (2006). General multi-level modeling with sampling weights. Communications in Statistics—Theory and Methods, Vol. 35(3), pp. 439-460.

3. Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. Psychometrika, Vol. 72, pp. 141-157.

4. Beaumont, J.F., Bocci, C. and Haziza, D. (2014). An Adaptive Data Collection Procedure for Call Prioritization. Journal of Official Statistics, Vol. 30 (3), pp. 607-621.

5. Boyle, J. M., Vanderwolf, P. (2004). 2003 Motor Vehicle Occupant Safety Survey: Volume 2. Safety Belt Report (DOT-HS-809-789). Washington, DC: National Highway Traffic Safety Administration.

6. Eby, D.W., Fordyce, T.A., and Vivoda, J.M. (2002). A comparison of safety belt use between commercial and noncommercial light-vehicle occupants. Accident Analysis and Prevention, Vol. 34, pp. 285–291.

7. Glassbrenner, D., and Ye, J. (2006). Seat belt use in 2006–Overall results. Traffic Safety Facts- Research Note (No. DOT HS 810 677). Washington, DC: National Center for Statistics and Analysis.

8. Gutman, R., Dede, G., Caplan, D., and Liu, J.S. (2011). Rasch Model and Its Extensions for Analysis of Aphasic Deficits in Syntactic Comprehension. Journal of the American Statistical Association, Vol. 106:496, pp. 1304-1316.

9. Hardouin, J.B. and Mesbah, M. (2004). Clustering binary variables in subscales using an extended Rasch model and Akaike information criterion. Communication in Statistics, Theory and Methods, Vol. 33, pp. 1277-1294.

10. Hahs-Vaughn, D.L. (2005). A Primer for Understanding Weights with National Datasets. The Journal of Experimental Education, Vol. 73(3), pp. 221-248.

11. Illi, S., Depner, M., Genuneit, J., Horak, E., Loss, G., Strunz-Lehner, C. and von Mutius, E. (2012). Protection from childhood asthma and allergy in Alpine farm environments—the GABRIEL Advanced Studies. Journal of Allergy and Clinical Immunology, 129(6), 1470-1477.

12. Kaplan, D., and Ferguson, A. J. (1999). On the utilization of sample weights in latent variable models. Structural Equation Modeling: A Multidisciplinary Journal, Vol. 6(4), pp. 305–321.

13. Kleinbaum, D., Kupper, L., Nizam, A., and Rosenberg, E. (2013). Applied regression analysis and other multivariable methods. Cengage Learning.

14. Korn, E. L., and Graubard, B. I. (1995). Examples of differing weighted and unweighted estimates from a sample survey. The American Statistician, Vol. 49(3), pp. 291–295.

15. Lambert-Bélanger, A. Dubois, S., Weaver, B., Mullen, N., and Bédard, M. (2012). Aggressive driving behaviour in young drivers (aged 16 through 25) involved in fatal crashes. Journal of Safety Research, Vol. 43, pp. 333-338.

16. Lohr, S. L. (1999), Sampling: Design and Analysis, Pacific Grove, CA: Duxbury Press.

17. Mesbah, M. (2010). Statistical Quality of Life, in N. Balakrishnan (Ed.), Methods and Applications of Statistics in Life and Health Sciences, New York, Wiley, p. 839-864.

18. Molnar, Lisa J.; Eby, David W.; Dasgupta, K.; Yang, Yang; Nair, Vijayan N.; Pollock, Stephen M. (2012). Explaining state-to-state differences in seat belt use: A multivariate analysis of cultural variables. Accident Analysis and Prevention, Vol. 47, pp. 78-86.

19. Nambisan, Shashi S., and Vasudevan, Vinod (2007). Is seat belt usage by front seat passengers related to seat belt usage by their drivers. Journal of Safety Research, Vol. 38, pp. 545-555.

20. Özkan, Türker; Puvanachandra, Prasanthi; Lajunen, Timo; Hoe, Connie; Hyder, Adnan. Accident Analysis and Prevention, Vol. 47, pp. 75-77

21. Pickrell, T.M., Ye, T.J., (2009). Traffic Safety Facts (Research Note): Seat Belt Use in 2008 – Demographic Results. National Highway Traffic Safety Administration/Department of Transportation, Washington DC.

22. Preusser, D.F., Lund, A.K., Williams, A.F., 1991. Characteristics of belted and unbelted drivers. Accident Analysis and Prevention, Vol. 23, pp. 475–482.

23. Rasch, G. (1961). On general laws and the meaning of measurement in psychology. Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4, p. 321-333.

24. Richards, S.A. and Whittingham, M.J., and P. A. Stephens (2011), "Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework," Behavioral Ecology and Sociobiology (2011) 65:77–89.

25. Rosenthal, R. (1978). How often are our numbers wrong? American Psychologist, 33(11), 1005.

26. Samejima, F. (1996), Evaluation of mathematical models for ordered polychotomous responses. Behaviormetrika, Vol. 23, 1996, p. 17-35.

27. Sangowawa, A.O., Alagh, B.T., Ekanem, S.E.U., et al. (2010). "An observational study of seat belt use among vehicle occupants in Nigeria," Injury Prevention Journal, Vol. 16, pp. 85-89.

28. Shults, Ruth A., and Beck, Laurie F. (2012). Self-reported seatbelt use, United States 2002-2010: Does prevalence vary by state and type of seatbelt law? Journal of Safety Research, Vol. 43, pp. 417-420.

29. Thomas, S. L., and Heck, R. H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. Research in Higher Education, 42(5), 517–540.

30. Vivoda, J.M., Eby, D.W., Kostyniuk, L.P., 2004. Differences in safety belt use by race. Accident Analysis and Prevention, 36, 1105–1109.