

STA 517: CATEGORICAL DATA ANALYSIS PROJECT

(100 Points)

This is an observational cohort study of infants born in the state of Florida in 1993. It is an exploratory epidemiological evaluation of a secondary database merging birth vital statistics (BVS), Florida's Death Data source supplied by the Florida Department of Health, Medicaid eligibility and enrollment data files supplied by Florida's Agency of Health Care Administration; Women, Infants and Children (WIC) Nutritional Supplement Program certification files supplied by the Florida WIC office; and the Florida Healthy Start prenatal risk screen score data file supplied by Florida Department of Health. It contains singletons only.

The data is aggregated by 10 explanatory variables - smoking, drk, BI, edu, ms, mrace, bsex, kotel, med, momage. Variable "Total" is the total number of infants, while variable "vlbw" is the total number of infants with vlbw and "ideath" is the total number of infants dead within 1st year of his/her life, within the combination of 10 explanatory variables. Your study's outcome is **vlbw**.

Response (Outcome) variable:

vlbw	Very Low Birth Weight (<1500g, Yes, No)
ideath	Infant Death (dead within first year of life)

Explanatory variables:

smoking	Smoking (Yes, No)
drk	Drinking (Yes, No)
BI	Pregnancy Interval (<=15 mo, First Birth, >15mo)
edu	Mothers Education (<HS HS >HS)
ms	Marital Status, married? (No Yes)
mrace	Mother's Race (Black Other White)
momage	Age of Mom (<20, >34, 20-34)
bsex	Sex of Baby (Male Female)
kotel	Adequate prenatal care? (No Yes)
med	Medicaid (Yes, No)

code	smoking	drk	BI	edu	ms	mrace	momage	bsex	kotel	wic	med
1	Yes	Yes	<=15 mo	<HS	No	Black	<=20	Male	No	Yes	Yes
2	No	No	First Birth	HS	Yes	Other	>34	Female	Yes	No	No
3			>15 mo	>HS		White	20-34				

Objective: To identify perinatal and sociodemographic risk factors associated with the vlbw.

Plan and conduct the data analysis as appropriate, with the 10 explanatory variables and corresponding two-factor interactions. Model selection (stepwise) entering and removing p-values are set as 0.05. Provide a report on the analysis. Your report should be less than 10 pages, well written and well organized. Brevity, clarity, organization and completeness are all desirable attributes in a report, as is legibility! Please do not simply put SAS output together. For the formats or contents, you may refer to the paper “Child Abuse & Neglect 28 (2004) 1253–1264” which can be downloaded at our class web page. Link <http://www.buffalo.edu/~cxma/STA517/childabusepaper.pdf>

At least, in your report you need to include

1. The sample size (total singletons), how many vlbw in your cohort, and percentage of vlbw. (5 Points)
2. A table with raw rates by explanatory variables (15 Points, similar to Table 1 in the abuse paper, but you only have one outcome)
3. The model building and final model (10 Points)
4. A table with adjusted main odds ratios and 95% CI (15 Points)
5. A table with adjusted odds ratios for interactions and 95% CI (15 Points) [Note: calculating for the first two interactions are enough to get this credit]
6. Interpretation of ORs (10 Points)
7. Conclusion (10 Points)
8. Please attach the R / SAS code in Appendix. (10 Points, it does not count for the 10 pages)
9. Others, i.e. writings etc. 10 Points

In the report, please use the real level instead of the code. For example, use smoking Yes vs. No; not smoking 1 vs. 2.