**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2015

**Homework #3**

January 23, 2015

**Written problems:** To be submitted as a MS-Word compatible file to the class Catalyst dropbox by 9:30 am on Monday, February 2, 2014. See the instructions for peer grading of the homework that are posted on the web pages.

*On this (as all homeworks) Stata / R code and unedited Stata / R output is* ***TOTALLY*** *unacceptable. Instead, prepare a table of statistics gleaned from the Stata output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*

***Unless explicitly told otherwise in the statement of the problem, in all problems requesting “statistical analyses” (either descriptive or inferential), you should present both***

* ***Methods: A brief sentence or paragraph describing the statistical methods you used. This should be using wording suitable for a scientific journal, though it might be a little more detailed. A reader should be able to reproduce your analysis. DO NOT PROVIDE Stata OR R CODE.***
* ***Inference: A paragraph providing full statistical inference in answer to the question. Please see the supplementary document relating to “Reporting Associations” for details.***

This homework considers pregnancy outcomes in an observational study of women attending a prenatal clinic in South Africa. Questions in this homework focus most closely on association with delivery of babies that are small for gestational age (SGA). The data can be found on the class web page (follow the link to Datasets) in the file labeled pregout.txt (you will not need any of the longitudinal measurements in the file preglong.txt). Documentation is in the file pregnancy.pdf.

1. Provide suitable descriptive statistics relevant to this analysis.

Method: In order to study the association with delivery of babies that are small for gestational age (SGA), I divided the total sample into two groups, namely, SGA group and normal group. We are interested in the association between maternal smoking behaviors and delivery of babies that are SGA and the association between maternal age and delivery of babies that are SGA. In addition, I suspect the number of prior delivery of mother might be confounders and sex of infant might be effect modifier. Therefore the descriptive statistics for maternal smoking behaviors, maternal age, sex of infant and number of prior delivery are presented for each groups and overall. For binary variables, including maternal smoking behaviors and sex of infant, the percentage of smokers and females are presented. For ordinal variables, including maternal age and prior delivery, mean, standard deviation, minimum and maximum are presented. Moreover, since the maternal smoking behavior is the predicator of interests, the sample odds and the probability of delivering a SGA infant separately for smokers and nonsmokers are presented.

Results: There were 755 subjects in the study. 105 of delivered babies were SGA and 650 of delivered babies are normal. There are 4 missing data in sex of infant; one of them was in SGA group and the other 3 were in normal group. There are 4 missing data in maternal smoking behavior; one of them was in SGA group and the other 3 were in normal group. As shown in the table, 43.3% of mothers who smoked gave birth to babies that are in SGA group and 28.6% of non-smoking mothers gave birth to babies that are in normal group. The probability of nonsmoker mother delivering a SGA infant was 0.113 and the odds of nonsmoker mother delivering a SGA infant 0.1280. The probability of smoking mother delivering a SGA infant was 0.195 and the odds of smoking mother delivering a SGA infant was 0.242. The average maternal age of SGA group is 24.9 years and the average maternal age of normal group is 23.9 years. 57.7% of babies in SGA group are female while 47.6% of babies in normal group are female. Average number of prior deliveries is 0.9 in SGA group while the average number of prior deliveries is 1.13 in normal group. Therefore, based on the descriptive statistics, smoking and young mothers tend to delivery babies that are SGA. Also, mothers with less number of prior deliveries tend to give birth to babies that are SGA and there are more female babies that are SGA in the sample.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Nmiss\* | SGA Yes (N=105) | Normal (N=650) | Overall (N=755) |
| Sex of infant (Female) | 4 | 57.7% | 47.6% | 49.0% |
| Maternal smoking behavior (Yes) | 4 | 43.3% | 28.6% | 30.8% |
| Maternal age (Year)\*\* | 0 | 23.9 (4.90; 16 - 35) | 24.9 (5.45; 14 - 43) | 24.8 (5.39; 14 - 43) |
| Number of prior deliveries\*\* | 0 | 0.90 (1.11; 0 - 6) | 1.13 (1.23; 0 - 6) | 1.10 (1.21; 0-6) |

\* There are 4 missing data in sex of infant; one of them was in SGA group and the other 3 were in normal group. There are 4 missing data in maternal smoking behavior; one of them was in SGA group and the other 3 were in normal group.

\*\*Descriptive statistics presented are the mean (standard deviation; minimum – maximum)

|  |  |  |
| --- | --- | --- |
|  | Sample Prob. of delivering SGA infant | Sample Odds of delivering SGA infant |
| Smoking mother | 0.195 | 0.242 |
| Non-smoking mother | 0.113 | 0.128 |

1. Perform a statistical regression analysis evaluating an association between the odds of delivery of infants who were small for gestational age (SGA) and maternal smoking behavior. (Only give a formal report of the inference where asked to.)
   1. Give full inference regarding the association between SGA and maternal smoking.

Methods: An indicator of maternal smoking behavior (1=smoking, 0=non-smoking) and an indicator of delivering SGA infant were created. A simple logistic regression of maternal smoking behavior (smoking and non-smoking) as the predictor and odds of SGA as the response was used. Assuming approximate normality of regression estimates, Wald - based estimates were computed for two-sided p-value and 95% confidence internal. Subjects missing data for SGA and maternal smoking behavior were excluded from the analysis.

Results: Of the 231 mothers who smoked, the odds of delivering SGA babies was 0.242, while of 520 mothers who did not smoke, the odds of delivering SGA babies was 0.128. The point estimates of odds ratio of SGA for the comparison of the smoker and non-smoker mothers was 1.89. Based on a 95% confidence interval, this observed odds ratio is not usual if the true odds ratio was between 1.23 and 2.88. With high confidence (two-sided p-value=0.003), we reject the null hypothesis that the odds of delivering SGA babies is not associated with maternal smoking behavior and in favor of the alternative hypothesis that odds of delivering SGA babies is associated with maternal smoking behavior.

* 1. Use the regression model parameter estimates to provide estimates of both the odds and the probability of delivering a SGA infant separately for smokers and nonsmokers. How do these estimates compare with simple descriptive statistics as you might have reported in problem 1. Explain any differences or similarities.

The estimated slope was 0.6368 and the estimated intercept was -2.0559. Thus the odds of nonsmoker mother delivering a SGA infant was exp(-2.0559) = 0.1280 and the probability of nonsmoker mother delivering a SGA infant was 0.1280/(0.1280+1) = 0.113. The odds of smoking mother delivering a SGA infant was 0.1280\*exp(0.6368)=0.242 and the probability of smoking mother delivering a SGA infant was 0.242/(1+0.242)=0.195.

These numbers was consistent of descriptive statistics in problem 1. It is not surprise since the model is saturated (number of parameters estimated = number of groups =2). Thus, the estimated odds and probability of the smoking / nonsmoking mother delivering SGA infant are same as corresponding sample odds and simple probability.

* 1. There were actually four regression analyses that could have been used to answer this question. I am betting that all students would have fit a regression model with SGA as response and the indicator of maternal smoking as the predictor. Presuming that you did indeed fit that model, explain the similarities and differences between the estimates and inference you would have obtained for the following three additional models (You do not need to run these analyses, if you can tell me how they differ without doing so. It is of course okay to run the analyses if it will help you recognize the more general principles.):
     1. You create an indicator NONSMOKER that the mother was a nonsmoker, and you fit a logistic regression model of response SGA on predictor NONSMOKER.

The intercept in this model will be the estimated log odds of smoker mother delivering SGA baby and the slope will be the estimated log odds ratio of delivering SGA baby for the comparison of the non-smoker and smoker mothers. After exponentiation and transformation, the estimates of the odds and probability of four events (smoker delivering SGA, non-smoker delivering SGA, smoker delivering non-SGA and non-smoker delivering non-SGA) will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

* + 1. You create an indicator NOTSGA that the infant was not small for gestational age, and you fit a logistic regression model of response NOTSGA on predictor SMOKER.

The intercept in this model will be the estimated log odds of non-smoker mother delivering non-SGA baby and the slope will be the estimated log odds ratio of delivering non-SGA baby for the comparison of the smoker and non-smoker mothers. After exponentiation and transformation, the estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

* + 1. You fit a regression model of response NOTSGA on predictor NONSMOKER.

The intercept in this model will be the estimated log odds of smoker mother delivering non-SGA baby and the slope will be the estimated log odds ratio of delivering non-SGA baby for the comparison of the non-smoker and smoker mothers. After exponentiation and transformation, the estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

1. Repeat problem 2, except consider a statistical regression analysis evaluating an association between the odds of delivery of infants who were small for gestational age (SGA) and maternal smoking behavior by evaluating the difference in probabilities for SGA across smoking groups.

a. Give full inference regarding the association between SGA and maternal smoking.

Methods: An indicator of maternal smoking behavior (1=smoking, 0=non-smoking) and an indicator of delivering SGA infant were created. I used a simple linear regression model of mean probability of delivering SGA infant as the response and maternal smoking behavior as the predicator. Without assuming equal variance, the Huber-white estimator was used to compute standard errors. Assuming approximate normality of regression estimates, Wald - based estimates were computed for two-sided p-value and 95% confidence internal. Subjects missing data for SGA and maternal smoking behavior were excluded from the analysis.

Results: Of the 231 mothers who smoked, the mean probability of delivering SGA babies was 0.195, while of 520 mothers who did not smoke, the mean probability of delivering SGA babies was 0.113. The point estimates of mean difference in probability for smoker and non-smoker mothers delivering SGA babies was 0.0813. Based on a 95% confidence interval, this observed difference in mean probability is not usual if the true difference in mean probability was between 0.02328 and 0.1394. With high confidence (two-sided p-value= 0.006), we reject the null hypothesis that the difference in mean probability of delivering a SGA infant from a smoking mother and non-smoking mother is zero.

b. Use the regression model parameter estimates to provide estimates of both the odds and the probability of delivering a SGA infant separately for smokers and nonsmokers. How do these estimates compare with simple descriptive statistics as you might have reported in problem 1. Explain any differences or similarities.

The estimated slope was 0.0813 and the estimated intercept was 0.113. Thus the probability of nonsmoker mother delivering a SGA infant was 0.113 and the probability of smoker mother delivering a SGA infant was 0.113+0.0813=0.195. The odds of smoking mother delivering a SGA infant was 0.195/(1-0.195)= 0.242 and the odds of nonsmoking mother delivering a SGA infant was 0.113/(1-0.113)=0.128.

These numbers was consistent of descriptive statistics in problem 1. It is not surprise since the model is saturated (number of parameters estimated = number of groups =2). Thus, the estimated odds and probability of the smoking / nonsmoking mother delivering SGA infant are same as corresponding sample odds and simple probability.

i.

The intercept in this model will be the estimated probability of smoker mother delivering SGA baby and the slope will be the estimated difference probability between nonsmoker mother and smoker mother delivering SGA baby. By addition of intercept and slope, we can get the estimate of probability of nonsmoker mother delivering SGA baby. By subtracting the probabilities from 1, we can get the estimates of probability of smoker and nonsmoker mother delivering non-SGA baby. After transformation from probability to odds, we can get the estimates of corresponding odds. The estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

ii.

The intercept in this model will be the estimated probability of nonsmoker mother delivering non-SGA baby and the slope will be the estimated difference probability between smoker mother and nonsmoker mother delivering non-SGA baby. By addition of intercept and slope, we can get the estimated probability of smoker mother delivering non-SGA baby. By subtracting the probabilities from 1, we can get the estimates of probability of smoker and nonsmoker mother delivering SGA baby. After transformation from probability to odds, we can get the estimates of corresponding odds. The estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

iii.

The intercept in this model will be the estimated probability of smoker mother delivering non-SGA baby and the slope will be the estimated difference probability between nonsmoker mother and smoker mother delivering non-SGA baby. By addition of intercept and slop, we can get the estimated probability of nonsmoker mother delivering non-SGA baby. By subtracting the probabilities from 1, we can get the estimates of probability of smoker and nonsmoker mother delivering SGA baby. After transformation from probability to odds, we can get the estimates of corresponding odds. The estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

1. Repeat problem 2, except consider a statistical regression analysis evaluating an association between the odds of delivery of infants who were small for gestational age (SGA) and maternal smoking behavior by evaluating the ratio of probabilities for SGA across smoking groups.

a. Give full inference regarding the association between SGA and maternal smoking.

Methods: An indicator of maternal smoking behavior (1=smoking, 0=non-smoking) and an indicator of delivering SGA infant were created. I used a simple Poisson regression model of probability of delivering SGA infant as the response and maternal smoking behavior as the predicator. Without assuming equal variance, the Huber-white estimator was used to compute standard errors. Assuming approximate normality of regression estimates, Wald - based estimates were computed for two-sided p-value and 95% confidence internal. Subjects missing data for SGA and maternal smoking behavior were excluded from the analysis.

Results: Of the 231 mothers who smoked, the mean probability of delivering SGA babies was 0.195, while of 520 mothers who did not smoke, the mean probability of delivering SGA babies was 0.113. The point estimates of mean probability ratio for smoker and non-smoker mothers delivering SGA babies was 1.769. Based on a 95% confidence interval, this observed mean probability ratio between smoker and nonsmoker mother delivering SGA baby was not usual if the true difference in mean probability ratio was between 1.2019 and 1.4627. With high confidence (two-sided p-value= 0.003), we reject the null hypothesis that the difference in mean probability of delivering a SGA infant from a smoking mother and non-smoking mother is zero.

b. Use the regression model parameter estimates to provide estimates of both the odds and the probability of delivering a SGA infant separately for smokers and nonsmokers. How do these estimates compare with simple descriptive statistics as you might have reported in problem 1. Explain any differences or similarities.

The estimated slope was 0.5405 and the estimated intercept was -2.1763. Thus the probability of nonsmoker mother delivering a SGA infant was exp(-2.1763)= 0.113 and the probability of smoker mother delivering a SGA infant was exp(-2.1763+0.5405 )=0.195. The odds of smoking mother delivering a SGA infant was 0.195/(1-0.195)= 0.242 and the odds of nonsmoking mother delivering a SGA infant was 0.113/(1-0.113)=0.128.

These numbers was consistent of descriptive statistics in problem 1. It is not surprise since the model is saturated (number of parameters estimated = number of groups =2). Thus, the estimated odds and probability of the smoking / nonsmoking mother delivering SGA infant are same as corresponding sample odds and simple probability.

i.

The intercept in this model will be the estimated log probability of smoker mother delivering SGA baby and the slope will be the estimated log probability ratio between nonsmoker mother and smoker mother delivering SGA baby. By exponentiation the intercept, we can get the estimated probability of smoker mother delivering SGA baby. By addition of intercept and slope and then exponentiation, we can get the estimated probability of nonsmoker mother delivering SGA baby. By subtracting the probabilities from 1, we can get the estimates of probability of smoker and nonsmoker mother delivering non-SGA baby. After transformation from probability to odds, we can get the estimates of corresponding odds. The estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

ii.

The intercept in this model will be the estimated log probability of nonsmoker mother delivering non-SGA baby and the slope will be the estimated log probability ratio between smoker mother and nonsmoker mother delivering non-SGA baby. By exponentiation the intercept, we can get the estimated probability of smoker mother delivering non-SGA baby. By addition of intercept and slope and then exponentiation, we can get the estimated probability of smoker mother delivering non-SGA baby. By subtracting the probabilities from 1, we can get the estimates of probability of smoker and nonsmoker mother delivering SGA baby. After transformation from probability to odds, we can get the estimates of corresponding odds. The estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

iii.

The intercept in this model will be the estimated log probability of smoker mother delivering non-SGA baby and the slope will be the estimated log probability ratio between nonsmoker mother and smoker mother delivering non-SGA baby. By exponentiation the intercept, we can get the estimated probability of smoker mother delivering non-SGA baby. By addition of intercept and slope and then exponentiation, we can get the estimated probability of nonsmoker mother delivering non-SGA baby. By subtracting the probabilities from 1, we can get the estimates of probability of smoker and nonsmoker mother delivering SGA baby. After transformation from probability to odds, we can get the estimates of corresponding odds. The estimates of the odds and probability of the four events will be the same as part in b. The inference and estimated value will be the exactly the same since the re-parameterization will not change the inference of the model.

1. How do the analyses performed in problems 2-4 compare to that that would be obtained in a simple two sample comparison of SGA by smoking status (i.e., using methods covered in Biost 517/514.) Explicitly mention where they would be similar or different?

Method: Without assuming equal variance, a T-test with two-sided p-values is used to test difference in mean probability of delivering a SGA infant from a smoking mother and non-smoking mother. Also we construct a 95% confidence interval in difference in mean probability of delivering a SGA infant without assuming equal variance. Welch-Satterthwaite approximation is used to calculate the degree of freedom. Subjects missing data for SGA and maternal smoking behavior were excluded from the analysis.

Results: Of the 231 mothers who smoked, the mean probability of delivering SGA babies was 0.195, while of 520 mothers who did not smoke, the probability of delivering SGA babies was 0.113. The point estimate of mean difference in probability for smoker and non-smoker mothers delivering SGA babies was 0.0813. Based on a 95% confidence interval, this observed difference in mean probability is not usual if the true difference in mean probability was between 0.0231 and 0.140. With high confidence (two-sided p-value= 0.006), we reject the null hypothesis that the difference in mean probability of delivering a SGA infant from a smoking mother and non-smoking mother is zero.

The estimated probabilities and odds are the same since the models are all saturated. The conclusions are the same. The slight differences are the standard error, confidence interval and p-value. The difference comes from following facts: in this homework setting, the logistic regression uses the classic estimated standard error; the linear regression and Poisson regression use Huber-White estimator for estimating standard errors; t-test without assuming equal variance uses pooled sample variance to estimate standard errors and Welch approximation for degree of freedom.

1. Perform a regression analysis of the distribution of the prevalence of SGA infants across groups defined by the continuous measure of maternal age. In all cases we want formal inference. (Note: In problem 7, I am asking you to plot the estimated probabilities of SGA infants from each of these regression models. Hence, you will want to make sure you estimate those fitted values following each regression.)
   1. Evaluate associations using risk difference (RD: difference in probabilities).

Methods: An indicator of delivering SGA infant was created. I used a simple linear regression model of mean probability of delivering SGA infant as the response and maternal age as the predicator. Without assuming equal variance, the Huber-white estimator was used to compute standard errors. Assuming approximate normality of regression estimates, Wald - based estimates were computed for two-sided p-value and 95% confidence internal. Subjects missing data for SGA and maternal age were excluded from the analysis.

Results: Mean maternal age for SGA group (N= 105) was 23.9 years and mean maternal age for normal group (N=650) was 24.94 years. The fitted slope was -0.00452. Based on a 95% confidence interval, this observed tendency of 0.00452 higher mean probability of delivering SGA baby among one year younger mother would not be judged unusual if the true mean probability difference of delivering SGA was anywhere between 0.00029 and 0.0087 higher among one year younger mother. Using a t test without assuming variances, this observation is statistically significant at a 0.05 level of significance (two-sided P-value=0.036). Thus, we reject the null hypothesis that the prevalence of SGA infants was not associated continuous measure of maternal age.

* 1. Evaluate associations between risk ratio (RR: ratios of probabilities).

Methods: An indicator of delivering SGA infant was created. I used a simple Poisson regression model of probability of delivering SGA infant as the response and maternal age as the predicator. Without assuming equal variance, the Huber-white estimator was used to compute standard errors. Assuming approximate normality of regression estimates, Wald - based estimates were computed for two-sided p-value and 95% confidence internal. Subjects missing data for SGA and maternal age behavior were excluded from the analysis.

Results: Mean maternal age for SGA group (N= 105) was 23.9 years and mean maternal age for normal group (N=650) was 24.94 years. The point estimates of mean probability ratio for one year older mother and one year younger mother delivering SGA babies was 0.9661. Based on a 95% confidence interval, this observed mean probability ratio between one year older mother and one year younger mother delivering SGA baby was not usual if the true difference in mean probability was between 0.9340 and 0.9995. With high confidence (two-sided p-value= 0.047), we reject the null hypothesis that mean probability of delivering a SGA infant was not associated with maternal age.

* 1. Evaluate associations using odds ratio (OR: ratios of odds)

Methods: An indicator of delivering SGA infant were created. A simple logistic regression of maternal age as the predictor and odds of SGA as the respond was used. Without assuming equal variance, the Huber-white estimator was used to compute standard errors. Assuming approximate normality of regression estimates, Wald - based estimates were computed for two-sided p-value and 95% confidence internal. Subjects missing data for SGA and maternal age were excluded from the analysis.

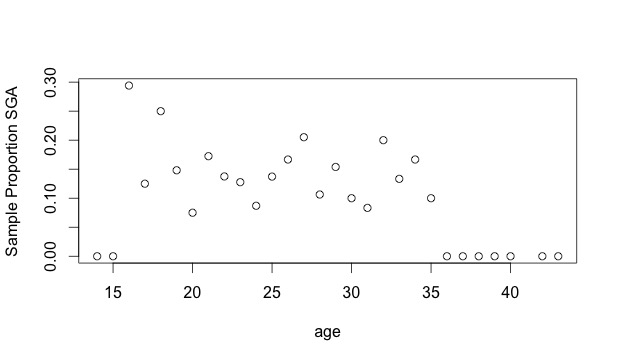
Results: Mean maternal age for SGA group (N= 105) was 23.9 years and mean maternal age for normal group (N=650) was 24.94 years. The point estimates of mean odds ratio for one year older mother and one year younger mother delivering SGA babies was 0.961. Based on a 95% confidence interval, this observed mean odds ratio between one year older mother and one year younger mother delivering SGA baby was not usual if the true difference in mean probability was between 0.9242 and 0.9993. With high confidence (two-sided p-value= 0.0461), we reject the null hypothesis that mean odds ratio of delivering a SGA infant was not associated with maternal age.

* 1. Using the regression parameter estimates from each of these regressions, provide an estimate of the probability that a 20 year old mother would have a SGA infant. Explain any similarities or differences these estimates might have when compared to the sample proportion of SGA infants among 20 year olds.

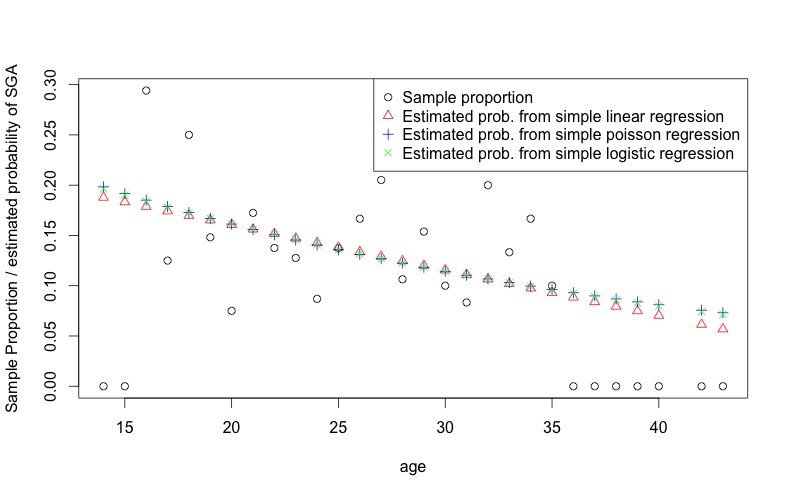
|  |  |  |  |
| --- | --- | --- | --- |
|  | slope | intercept | proportion / estimate of the probability  of SGA infants among 20 year olds. |
| sample | / | / | 0.075 |
| linear regression | -0.004515 | 0.250997 | 0.251-0.00451\*20=0.1608 |
| Poisson regression | -0.03442 | -1.13598 | exp(-1.13598-0.03442\*20)= 0.16131 |
| logistic regression | -0.03978 | -0.85316 | odds=exp(-0.85316-0.03978\*20)  odds/(1+odds)= 0.16138 |

The sample proportion of SGA infants among 20 year olds (N=40) was 0.075 however the estimates of the probability of three models are all about 0.16. Because this is not a saturated model, the estimates will not agree with sample proportion. The estimates from the model are maximum likelihood estimates. They are not only based on the sample of 20 years old mother but also borrow information from other age groups.

1. Produce a plot of the estimated probability of an SGA infant by age as derived by each of the following methods. Comment on the similarity and difference among the various fitted values form the various analyses performed in problem 6. (Note that Stata allows you to specify multiple Y variables for a single X variable: scatter y1 y2 y3 y4 age)
   1. Sample proportions within each unique age



* 1. Estimated probabilities for each age in the data as derived from each of the regression analyses.



As we can see from the graph, the estimated probabilities from three models are very similar to each other. They characterized the linear trend of association between age and probability of delivering SGA infant. Moreover, note that the estimates from Poisson regression and logistic regression are even more similar because that both regression use log as link function. On the other hand, these estimates are significantly different from sample proportion, especially for younger and older mother. The inconsistency may be due to the limited sample sizes in those groups.

1. Perform a logistic regression analyses of the distribution of the prevalence of SGA infants across groups defined by the logarithmically transformed maternal age.
   1. Provide formal inference for associations using odds ratio (OR: ratios of odds) and log transformed age.

Methods: An indicator of delivering SGA infant were created. A simple logistic regression of log-transformed maternal age as the predictor and odds of SGA as the respond was used. Without assuming equal variance, the Huber-white estimator was used to compute standard errors. Assuming approximate normality of regression estimates, Wald - based estimates were computed for two-sided p-value and 95% confidence internal. Subjects missing data for SGA and maternal age were excluded from the analysis.

Results: Mean maternal age for SGA group (N= 105) was 23.9 years and mean maternal age for normal group (N=650) was 24.94 years. The point estimates of mean odds ratio for log age one unit larger mother and log age one unit younger mother delivering SGA babies was 0.3853. Based on a 95% confidence interval, this observed mean odds ratio between one year older mother and one year younger mother delivering SGA baby was not usual if the true difference in mean probability was between 0. 1467and 1.0123. With high confidence (two-sided p-value= 0.053), we do not reject the null hypothesis that mean odds ratio of delivering a SGA infant was not associated with maternal age.

* 1. Why might it be reasonable or silly to have performed such an analysis rather than the analysis in problem 6c?

I think it might be silly to perform such an analysis because there is no scientific reason to do so. In addition, it makes the interpretation very hard and anti-intuitive.